Minutes for Mobility

Investigating Municipal Differences in Travel Time Expenditures in the Netherlands

Master Thesis

D.A. Tromp

Faculty of Technology, Policy, and Management

TU Delft

Kennisinstituut voor Mobiliteitsbeleid

Minutes for Mobility

Investigating Municipal Differences in Travel Time Expenditures in the Netherlands

by

D.A. Tromp

Student Number: 4895479

To obtain the degree of **Master of Science**
in **Engineering and Policy Analysis**
at the Delft University of Technology

To be defended in public on August 30th, 2024

Thesis committee

1st supervisor:	dr.ir. M. Kroesen *(chair)*	TU Delft, TPM
2nd supervisor:	dr. O. Oviedo Trespalacios	TU Delft, TPM
External supervisor:	ir. R.M. Faber	KiM

Project Duration: February, 2024 - August, 2024
Faculty: Technology, Policy, and Management

Cover: Mensen in de metro, designed by Freepik (n.d.)
Style: TU Delft Report Style, with modifications by Daan Zwaneveld

TU Delft

Kennisinstituut voor Mobiliteitsbeleid

Preface

Before you lies my master thesis, a product of six months of research at the Netherlands Institute of Transport Policy Analysis (KiM). Even more so, it is the product and final part of six years of education at the TU Delft and all that I've learned during this period.

I want to start by thanking my supervisors for their support and input during the process of writing this thesis. Maarten, I have of course already worked with you for a while as TA for various courses, which was always very pleasant. Asking you as my first supervisor thus only seemed logical, and you immediately guided me towards the KiM when I expressed my desire for an internship, so thank you for this great referral as well. Oscar, thank you as well for your insights. Roel, thank you for your critical view and keeping me sharp during the project. I also want to thank my other colleagues and fellow interns at KiM, for making the whole thesis process more fun and less isolated.

Reflecting on my time in Delft, it now seems funny that my choice for studying 'Technische Bestuurskunde' came about due to chance. When attending the TU Delft Open Days seven years ago, I merely decided to attend the TB presentation in order to pass time between other presentations that I actually wanted to attend. The TB presentation however turned out to be very interesting (more so than the others), and given how much I have enjoyed studying and living in Delft, in hindsight attending this presentation might have been one of the best decisions I have made.

People often say that your student days are the best time of your life, and I fully understand why. I have greatly enjoyed all the various exciting opportunities and challenges that have come my way. The past six years have been filled with fun activities, parties, and lots of tennis at Obvius, some great trips, including an especially unforgettable semester abroad in the United States, and many, many more highlights. I am very grateful for all I have experienced, for which I also have many more people to thank.

Thank you to my study friends, for the many projects we successfully completed together, but even more for the fun outside of studying, like the weekend trips, including those (way too) fanatic games of bowling. Thank you to my housemates, for the fun but always relaxed living environment, and especially for making the COVID lockdown actually quite an enjoyable period in a way. A big thank you to all my friends from Obvius, especially my teammates, for the countless hours we spent on court together, but also for the clowning around outside of the tennis courts. All of you have made my time Delft truly unforgettable. Thank you to my high school friends as well. Though we do not see each other as often anymore, every time we do it is always still as fun as ever.

Finally, to my family: my parents, my sister, and my granddad. Your unconditional love and support in everything I do and have done, allows me to live a carefree life and enjoy it to the fullest extent. For this, I cannot thank you enough.

Though I am sad that my time as a student is over, I am excited for what the future holds. First up: a few months off to do some travelling. In the meantime, enjoy reading my thesis!

Daan Tromp
Delft, August 2024

Abstract

The mobility landscape is constantly changing and developing, with travel options becoming ever faster and more abundant over the years, due to the technological advancement and globalisation in modern times. With this, human mobility changes as well, as further away destinations become more easily reachable for many people. The BREVER-law however states that humans have a constant, intrinsic travel time budget that they tend to spend on travel. They adjust their trips and destinations to make their expenditures fit this budget, given the speed with which they can move around. This supposes that mobility developments have not led to considerable total travel time savings, but rather have urged people to travel longer distances in the same amount of time. This 'law' implies consistency over time, but also uniformity of travel time expenditures across space. There does however exist some dispute surrounding validity of this notion, as travel time expenditures mainly seem to be constant at high (national/international) levels of aggregation.

Within the Netherlands, though organisational decentralisation of infrastructure expenditure has taken place over the past years, most of the infrastructure expenditures has appeared to be fairly Randstad-oriented. However, given the recently published coalition agreement of the current Dutch administration, an attention shift towards rural accessibility projects seems upcoming. Various government reports allege there are critical differences in accessibility between the Randstad and more rural areas in the Netherlands, and this argument underlies the motivation for this attention shift. This narrative, combined with the supposed uniformity of travel times, begs the question whether alleged accessibility differences present themselves through measurable differences in travel time expenditures across the country. Therefore, insight into uniformity of travel time expenditures across the country can help determine whether infrastructure projects focusing on diminishing rural inaccessibility are the best use of infrastructure budgets.

Though travel behaviour research is a well-established field, intra-country patterns in travel time expenditures remain underexplored or unclear. Most research into this tends to consider inter-city comparisons, or studies differences between urban and rural regions. However, more granular disaggregate studies that also differentiate between rural regions are not common, and in general spatial coherence is seldom considered. On top of this, studies into travel time expenditures in the Netherlands specifically are often based on older data. This research aims to gain more up-to-date knowledge on spatial patterns in travel time expenditures across the Netherlands, and tries to answer the question:

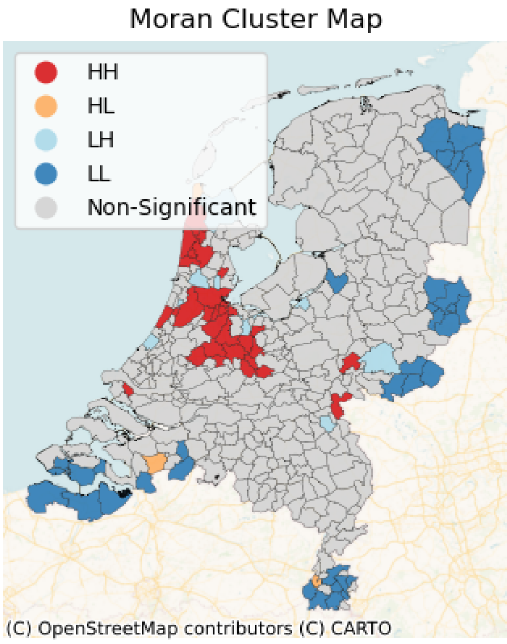
How do travel times expenditures vary spatially across the Netherlands, and is there evidence to support the concept of constant travel times at a disaggregated scale?

As a first step towards answering this question, a literature study was performed to distil the most important existing knowledge on travel time budgets, travel time expenditures, and their primary influencing factors. In this literature review, varying results were found on the effect of the spatial context on travel time expenditures. There is some evidence of travel behaviour varying with levels of urbanisation, but these relationships are often faced with conflicting research as well. From the literature it also became apparent that travel time expenditures do not only differ with the spatial context, but also with socio-demographic characteristics. In fact, socio-demographics are often considered to have a stronger, more significant, impact on travel time expenditures than the spatial context.

To investigate the travel time expenditures across the Netherlands, national travel survey (ODiN) data is used for analysis. The ODiN data contain diary-like travel records of a full day for each included individual, including detailed trip characteristics, as well as various other background variables on the

respondents. From this data, individuals' travel time expenditures, the factor of interest in this research, can be derived. The data is extensively preprocessed to make it usable for the analyses: two years of ODiN records (2018 and 2019) are joined, individual records are aggregated to municipal measures, and the data is enriched with a spatial element, amongst other preparatory actions. The preprocessed data used for the analyses in this research contain records of 104.818 individuals.

Using the municipal average travel time expenditures, spatial autocorrelation analysis has been performed to look for spatial patterns in the emergence of relatively high and/or low travel time expenditures. In general, the values vary quite a lot, opposing the claim of spatially constant travel time across the country. At the national level we found weak to moderate, but significant, spatial autocorrelation. Similar values of travel time expenditure do thus show spatial coherence and tend to appear near each other more often, though this relationship is not very strong. When zooming in and considering spatial coherence at the local level, it became clear that this spatial coherence is only prevalent in a few regions, while in most of the country no distinct spatial clustering can be identified. Regions with significantly higher travel time expenditure were found in the northern part of the Randstad, and the northern part of North-Holland, while regions with significant clusters of low travel times appeared in five distinct regions near the country border, far away from the Randstad, as the figure here shows.



Looking more closely at the background characteristics of the regions of interest, it showed that the high travel time regions housed more young adults, fewer car owners, and more high-income and highly educated individuals.

In search for further explanations for these emerging patterns, the data was subjected to various spatial and non-spatial regression analyses, to establish the explanatory power of various background variables regarding the travel time expenditures. The explanatory factors included socio-demographic, built environment, and distance measures. Consistent with the characteristics of the regions of interest from the spatial autocorrelation analysis, the regression models indicated primary importance of socio-demographic variables over built environment variables. Once again, high income and high education appeared to be related to higher travel time expenditures, and these factors seem to explain away the influence of the built environment or a 'Randstad effect'. Additionally, following the observation that the regions with lower travel time expenditures all appeared near borders, the spatial regression models with the best fit confirmed that border municipalities do show lower travel time expenditures. This could imply the existence of a border effect where the border, though easily crossable in reality, forms a psychological barrier preventing people from travelling across it, thus limiting their movements.

Based on these results, some remarks should be expressed regarding the intended infrastructure investments targeting regional accessibility. First, the highest travel time expenditures do show up in the allegedly well-connected Randstad region. Additionally, earlier research already showed that people living in the supposedly badly accessible regional areas do not show more dissatisfaction with their perceived accessibility, which begs the question whether investing in improving their accessibility would be worthwhile. Possibly residential self-selection plays a role here, if people tend to choose to reside in areas that fit their mobility preferences. Lastly, as not the urban environment, but rather socio-demographic background appears to be a more important determinant of travel behaviour, one might ask whether policies targeting population groups, instead of regions, would be more effective in achieving mobility goals. An example of this could be a subsidy programme to make travel more affordable for lower income groups, as they exhibit lower travel time expenditures. Given the results of this research, infrastructure investments should be carefully reconsidered and more elaborate research should be done into their expected effects, and how people would experience and react to improved accessibility. For this, it is important to gain a better understanding of the relationship between accessibility and total travel time expenditures.

This research has contributed to the scientific knowledge on travel time expenditures in a multitude of ways. It has provided an update on the outdated previous studies on the Dutch context, by using more recent (2018/2019) data for the performed analyses. It has also provided a new way of addressing regional differences in travel behaviour, by not only considering determining factors, but also the spatial coherence of similar values, which has allowed us to identify spatial patterns beyond the effect of the explanatory factors as well. We have also reconfirmed the notion that socio-demographic characteristics affect travel time expenditures more than the built environment does.

One of the limitations of this research is that it does not take into account the psychological aspect of how people experience their travel time either, or how much time they would be willing to spend. The analyses are based on real-life travel time expenditure, and though this is not totally unrelated to time budgets, it does not provide the full picture. The used data also contain some inherent limitations, as they depend on respondents accurately recalling and reporting their trips, it only containing records on one single day per respondent, as well as the fact that the applied aggregation methods blur individual-level details of the data at the municipal level. We must thus be wary of an ecological fallacy when translating the results of the analyses to individual-level insights.

Some of these limitations, as well as additional interesting avenues of inquiry that follow from this research, provide some recommendations for further research. First, more research should be performed into people's underlying travel preferences and their experiences, instead of only their exhibited behaviour. Additionally, the data used in this research does allow for way more detailed and additional analyses that can add to the understanding of travel behaviour, like analyses at the individual or neighbourhood level, or the inclusion of additional factors, like differences between modes of transport. It would also be interesting to see if the results that the 2018/2019 data yield can be derived from data from different years, or perform analyses to detect developments over multiple years. Finally, the identified effect of border vicinity decreasing travel time expenditures is interesting to investigate further, by closely tracking exact travel behaviour of inhabitants in these regions, to see if they really avoid crossing the border. Additionally, one might investigate whether this phenomenon also occurs on the other (Germany/Belgium) side of the border.

Contents

1 Introduction

With the vast technological advancements and rapid globalisation in modern times, barriers for travel are being lowered constantly. Better roads and more cars, as well as better public transport, have made it easier to get to the other side of a region or country, while bigger and faster aeroplanes have made it easier to get to the other side of a continent, or even the world. Today one can get from the U.K. to Australia in a day, a century ago this journey would have taken over a month (MacDonald, 2015). Initially, one might expect that an increase in travel speed would cause people to arrive at their destinations quicker, thus decreasing total travel time expenditure and leaving more time for other activities. This idea of travel time savings often is a central concept in transport economic analysis. However, this relationship appears to be not nearly as simple as this, which weakens the justification of using this measure for assessing policy options (Metz, 2008).

1.1. Background

In 1977 Geurt Hupkes first coined the BREVER[1]-law. This 'law' presents a notion in traffic engineering that humans have a constant travel time budget (Hupkes, 1977). This would mean that humans have a preset, fixed portion of our time that we are willing to spend on travelling. An implication of this would be, and Hupkes also mentions this in a later paper, that any increase in travel speed will not come to the benefit of saved travel time by leading to shorter or fewer trips. A constant travel time budget would imply that an increase in speed would merely lead to longer distances covered in the same amount of time (Hupkes, 1982). These findings by Hupkes were based on earlier work by Szalai, who set out the first large cross-national research into behavioural patterns (Szalai, 1966a, 1966b, 1975). Travel behaviour has been researched plenty more later on and this notion of a constant travel time budget has been widely accepted ever since, although there remains a lack of full consensus on the true existence of such a constant (Ahmed & Stopher, 2014). Different studies find different numbers, but people seem to spend on average about 60 to 90 minutes per day on travel[2]. These early studies on the BREVER-law imply both a consistency over time, as well as uniformity across space.

Travel behaviour is important to study for a multitude of reasons. First, as all things, roads and other infrastructure deteriorate as they are being utilised. Changes in travel behaviour have implications for traffic flows and therefore a proper understanding of travel behaviour is paramount for good infrastructure policy. To consider this, measured real-world travel behaviour is already embedded in Dutch travel models that support policy making in this area (Rijkswaterstaat, 2017). Travel is an important determinant of societal demand for infrastructure maintenance and related policies, though accurate demand forecasting proves to be difficult (Ivanchev et al., 2015; van Wee, 2007). This poses a challenge, as adequate and equitable access to proper infrastructure is paramount in fostering sustainable economic development (United Nations, 2015). In assuring regions of equitable development opportunities, their

[1]Behoud van Reistijd en Verplaatsingen (EN: Conservation of Travel Time and Trip rates)
[2]This idea of a constant travel time budget encompasses all forms of travel, including e.g. walking and undirected trips

(relative) accessibility thus must be considered. Because of this, insight into the uniformity of travel times over space are valuable.

To better understand where this value is derived from, it is important to understand the distinction between mobility and accessibility. Accessibility represents one's ability to reach a destination, as it considers what it takes to get somewhere. It is a combination of two factors: proximity and travel velocity (CROW, n.d.). Closer proximity to a facility (i.e. a shorter distance) and a higher travel velocity both lead to better accessibility, as they allow a person to more easily/quickly reach their destination. Mobility is considered to be the actual movement people exhibit, of which the travel time expenditure (TTE) can be a measure. It is expected that one's accessibility exhibits influence on the travel one actually engages in, e.g. their mobility (though the polarity of this relationship is somewhat ambiguous, as further unfolded in section 1.2). Because of this, knowledge on (differences in) mobility could give insight into accessibility (in)equality between regions. Figure 1.1 visually shows how these factors are considered to relate to each other in this research.

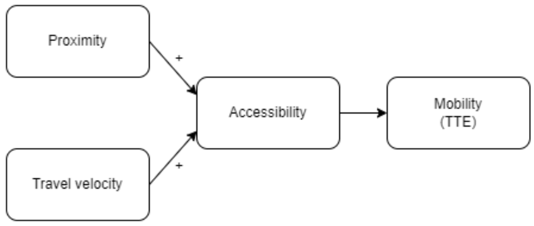


Figure 1.1: Relationship between proximity, velocity, accessibility, and mobility, as considered in this study

There is also value to insight into the BREVER-law's supposed consistency over time. Travel speeds and other aspects of travel patterns can have a variety of secondary effects in different systems. A first obvious example is the link with sustainability. As most modes of transport inevitably come with CO2-emissions, the length or frequency of commuting (or travel for any reason for that matter) does affect emissions (Walls & Safirova, 2004). Furthermore, in a paper published early on into the COVID-19 pandemic it was suggested these changes in commuting could also cause shifts in the housing market in the long term. Especially with the rapid increase of working-from-home opportunities during the pandemic, employees seem to become increasingly more likely to accept longer commute times (de Vos et al., 2018). Having to make the commute less often, it becomes less of a burden, a positive development recognised by employees (Kun et al., 2020). This development intuitively makes sense and it is imaginable that this could in the long-term even lead to re-ruralisation (Weening, 2020). This acceptance of longer, but less frequent trips can be considered part of this BREVER phenomenon, which also presents itself in the fact that the temporal gains of decreasing commute times during the pandemic appear to have been (partly) reallocated to other trips (Faber, Hamersma, de Haas, et al., 2023; Hook et al., 2023). Still, though temporal trends of travel time budgets are thus also of value through a variety of angles, this research focuses on the spatial uniformity aspect of the BREVER-law.

Some more important remarks should already be added to this BREVER concept here. Firstly, the BREVER-law is only considered to be applicable at a highly aggregated level and does not necessarily claim that all individuals have the same personal travel time budget (Hupkes, 1982). This means that the *average* travel time per person across the population appears to be stable and constant over time, but it does allow for variety across individuals (Wigan & Morris, 1981). Persons can differ in their personal travel time budget, which can vary with age, gender, location, and much more. However, for a single individual one's travel time remains constant over time, meaning the average value of travel times across all individuals is constant as well (Stopher et al., 2017). Taking this into account, it is important to consider these possible disaggregated differences, as just an average does not provide any information on the distribution across people and does not allow for full understanding of the travel times. With the aforementioned equitable accessibility across regions in mind, it becomes even more clear that there is valuable knowledge to be gained from identifying any regional differences in travel times.

Disaggregated research into travel behaviour in the Netherlands specifically has been performed in the past (Feng et al., 2013; Timmermans et al., 2002, 2003). Besides these specific research articles, there are also mentions of a variety of other Dutch-based researches in more (review) articles from this field (Ahmed & Stopher, 2014; Raux, Ma, Joly, & Cornelis, 2011; Raux, Ma, Joly, Kaufmann, et al., 2011). However, after looking at the analytical foundation of these papers, it appears that any rigorous researches into travel patterns in the Netherlands base their claims on older data from before 2010, or even from the previous century. There is thus value to be gained from a general reassessment, as there is a lack of substantiation of Dutch travel pattern in more recent times.

Also important to point out is the conceptual difference between the terminologies of 'travel time budget' and 'travel time expenditure'. When using the term 'budget', which is the term that is predominantly used in earlier papers surrounding BREVER, it tends to imply that people inherently have a specific amount of time that they are willing to spend on travelling (Hupkes, 1982; Szalai, 1966a; Zahavi, 1974). This can however differ from the actual observed time spent on travel, which is better captured by the term 'expenditure', as this tends to consider the real-world travel behaviour that people actually show (Gallotti et al., 2015; van Wee et al., 2006). This distinction is important to understand, as these two terms do not necessarily embody the same factors or values, even though they are sometimes used interchangeably. A truly inherent psychological value for a travel time budget might be very hard to derive, which leads to expenditures sometimes being used as proxies for the underlying budgets (Mokhtarian & Chen, 2004). In this thesis the same interpretation is used and the main focus of the analysis will be on travel time expenditures. In doing this it is assumed that, at the aggregate level, true expenditures can serve as a proxy for what people are willing to spend (budget). In this context expenditures will be used to look into the BREVER-law. When referring to travel times in this report, the terms 'budget' and 'expenditure' will sometimes be used interchangeably to refer to this general notion of (total) travel times.

1.2. Policy relevance

To further unfold the policy relevance of identifying spatial/regional differences in travel times, we dive deeper into how infrastructure provision is structured in the Netherlands and the recent trends that we see in the governments' financial resource allocation in this sector. Firstly, to understand the build-up of infrastructure investments, figure 1.2 provides a good overview. The three levels of Dutch government (national, provincial, municipal) all fund projects at their own level, with higher levels of government jointly funding projects at lowers levels as well[3]. In this funding structure, as the figure also shows, it makes sense that local governments take financial accountability for smaller local projects, while the national government takes care of national highways and other larger projects, as well as also helping out the local governments.

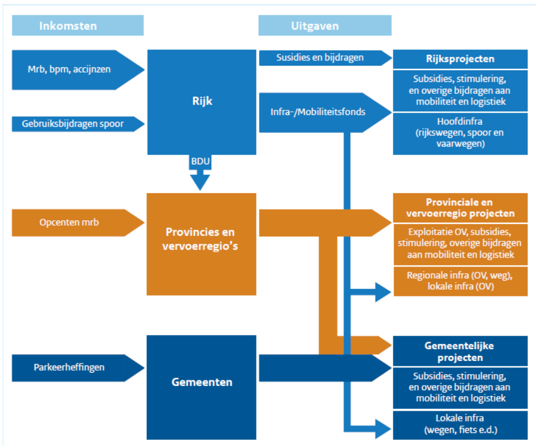


Figure 1.2: Infrastructure investment structure across levels of government in the Netherlands (Rienstra, 2022, p.26)

At the national government level, institutions are thus tasked with finding a balance in funds allocation between large national infrastructure projects and supporting local infrastructure. We also see this in the 2024 Ministry of Infrastructure & Water Management (I&W) budget, which contains cost items for both the main roads network (€3,75b) and regional and local infrastructure (€1,14b) (Kamerstuk 36410-XII, 2023). Over the period 2008-2019, the total amount of funding going to infrastructure has decreased from about 1,0% to about 0,7% of the GDP (Rienstra, 2022, p.10). Rienstra also shows that during this period, the funding division across the levels of government has changed as well, the share

[3]More generally, the financial resources of provinces and municipalities are already dependent on national government funds anyway (Ministerie van Financiën, n.d.)

of the national government decreasing from 59% in 2008 to 52% in 2019 (p.27). The share of province investment on the other hand increased from 17% to 22%, with municipal investments staying fairly stable at 25% to 26%. Interestingly, during this period the travel time loss (due to e.g. congestion) on the main roads network increased (KiM, 2022).

There thus appears to have been a shift towards decentralisation of infrastructure expenditure in governments, at least organisationally considering the levels of government as a whole. This does however not yet imply that either rural or urban regions receive more investments than the other, as therefor we need to look at any inter-municipal and inter-province differences in expenditures. At the provincial level we see that during 2014 to 2019, the infrastructure investments were highest in North- and South Holland, Gelderland, Utrecht, and North Brabant (Ecorys & VanBerkel Professionals, 2019). These being the provinces with the largest populations, this initially seems to make sense. However, converted to per-capita expenditure, North- and South-Holland actually spend the least on infrastructure, with Groningen spending the most according to numbers from Statistics Netherlands (CBS, 2023a). When looking at the municipal level, data shows that over the past seven years both road- and public transport per-capita expenditures are higher in larger municipalities[4] (CBS, 2024). Furthermore, under the previous administration, additional investments of the national government into 'local projects' were still mostly aimed at projects in or surrounding the Randstad or other larger cities (NU.nl, 2022). These numbers imply that, while budgets have been reallocated to different organisational levels, over the past years infrastructure expenditures were still fairly Randstad-oriented, though we cannot definitively conclude this without further insights into project-specific information.

More generally, the distribution of infrastructure expenditure has been a hot topic on the policy agenda in recent times, with a special focus on 'broad prosperity' (NL: brede welvaart) and accessibility in rural regions. The term accessibility here refers to the time, money and energy it takes to reach destinations (CPB & PBL, 2020, p.33). This report by CPB and PBL also explains the importance of it, as accessibility (to education, jobs, etc.) strongly influences one's social and economic development opportunities. An often repeated narrative is that the difference in accessibility is large between the Randstad and rural regions, and that this lacking accessibility in rural areas is detrimental to their development (Bastiaansen & Breedijk, 2022). This disbalance between the Randstad and regional areas was even discussed at length during a debate on the aforementioned budget of the Ministry of I&W (Eerste Kamer der Staten-Generaal, 2024). It is also one of the important points of attention in the 'Mobiliteitsvisie 2050' and was presented as one of the pillars of the coalition agreement for the current Dutch administration (Kabinetsformatie, 2024; Ministerie van Infrastructuur & Waterstaat, 2023). With support of this narrative, in a reaction to the aforementioned I&W budget the Association of Dutch Municipalities (VNG) advocates for (even) more cooperation from the national government in local mobility projects in regional areas, with the goal of increasing accessibility in rural regions (VNG, 2024). They do not only want more investment in infrastructure itself, but also direct their attention towards decreasing the necessity of having to cover distances in the first place by advocating for creating more facilities in these rural areas.

But here's the interesting twist: though it is true that rural regions might have lower accessibility compared to the Randstad and are more dependent on cars, the population in these regions does not seem to mind (Jorritsma et al., 2023; Pot, 2023). People living in rural areas often still judge their accessibility as satisfactory. This begs the questions whether additional infrastructure investments for these regions are really necessary. Would the local population truly benefit as much from these investments as is now assumed they will? A thorough understanding of this can help allocate infrastructural funding across government levels and projects in an adequate manner.

With figure 1.1 in mind, one could reason that the lower accessibility in rural regions is due to one or two factors: (1) fewer facilities in rural regions compared to cities makes destinations less proximal, and (2) less developed infrastructure networks (e.g. more rural roads and fewer highways, or fewer

[4]With exemption of the smallest (< 5.000 inhabitants) municipalities, which probably have high expenditures due to the 'default' expenditures being divided over few people

public transport options) lead to lower average travel velocities in rural regions. Being less proximal means one must cover larger distances, and combined with a lower speed one could expect this results in lower accessibility in rural areas. With this understanding, the aforementioned call by the VNG for not only additional infrastructure investment, but also for more essential facilities in rural areas, makes sense, as this second action would tackle the alleged accessibility problem from the proximity side. From this argument, it becomes clear that mobility in itself is not the objective, but accessibility rather is. Additionally, the plans of the new coalition to focus more on regional accessibility also suggest a shift from a general holistic social cost-benefit analysis mindset to a more equitability-focused approach.

Now, given the outlined relationships between these factors, one might expect that the lower accessibility of rural areas (due to less proximity and less developed infrastructure) would express itself in higher travel time expenditures (more mobility) in these regions, as it would take longer to reach the essential facilities that one has to get to. However, one must be careful with assuming this relationship. A closer proximity to essential facilities might also result in model shifts, as distance is of primary importance is mode choice (Scheiner, 2010). To give an example of this: imagine person A lives in a rural area, 10 kilometres from a doctor's office, a distance which would take them 15 minutes by car. Person B lives in a city, 1 kilometre from a doctor's office. However, because person B (who we assume does own a car) lives so close by, they decide to walk to the doctor, which takes them 15 minutes as well. They thus have the same travel time, though the doctor is way more proximal for person B. In this case, person B shows a modal shift from car to foot. Instances like these make that one should be careful with directly interpreting travel time expenditures as indicators of proximity or accessibility, and we will come back to this in the discussion (section 5.5).

The relationship between accessibility and travel time expenditures is thus not as clear-cut as one might have hoped. Still, to begin investigating this relationship further, one must start with assessing whether (what kind of) differences in travel time expenditures between regions can be identified in the first place. As described in the next section, literature often finds higher travel times in urban regions, so is this also the case in the Netherlands? This research aims to answer this question.

1.3. Scientific relevance

Travel behaviour research is an already very advanced field and plenty of research has been performed into travel time budgets, travel time expenditures, and other aspects of this field. As mentioned before, the BREVER-law only claims to apply at an (inter)nationally aggregated level, which has fuelled some to dive deeper into the concept of constant travel time, looking for more clarity on disaggregation and explanations of possible differences between individuals, groups, and/or regions. In general, these explanations tend to look into either one of two types of factors: socio-demographic variables, or spatial characteristics.

There exists plenty of research on the influence of socio-demographic characteristics on travel behaviour and it is widely accepted that differences between different groups exist (Timmermans et al., 2002, 2003). Travel time expenditures seem to be higher for e.g. men, persons between 15 and 65 years of age, and with high income (Koo et al., 2022; Mokhtarian & Chen, 2004; Raux, Ma, Joly, & Cornelis, 2011). Still, for most socio-demographic variables, some opposing evidence of their effect on travel can also be found, as Mokhtarian & Chen (2004) extensively outline. The true nature the relationships can be hard to identify, as socio-demographic variables also tend to interact with other variables and confound relationships due to interaction effects, caused by the complexity of their coherence (Jahanshahi et al., 2013; Zhou et al., 2022).

When considering spatial characteristics, some literature hints at variation of travel behaviour across different areas and urban contexts, usually stating that people tend to travel more in cities (Gallotti et al., 2015). However, this effect is disputed and the evidence in favour of such relationships is countered by papers that do not find significant relationships. Review articles by Van Wee et al. (2006) and Mokhtarian & Chen (2004) also come to the general conclusion that travel time increases with an increasing level of urbanisation, but also present a range of articles that did not find any significant effects of the built environment. Papers confirming the relationship often present research that was performed in different geographies (e.g. Gallotti et al. (2015) was an Italian study), and as they thus do not yield consistent results, it is hard to to translate findings to other contexts and take them as truth for the situation in the Netherlands. Additionally, most papers on this matter tend to focus on comparison either within a city (city centre v.s. suburban) or between cities (larger cities v.s. smaller cities) (Cats, 2023). Studies that do include rural regions as well often make a fairly binary comparison between rural and urban areas, but tend to not dive deeper into differences between multiple rural areas. The cited disaggregate studies in Chen & Mokhtarian (2008) also underline this. Research seldom draws conclusions at a high level of granularity, and therefore in this study we aim to further explore the country-level spatial variation within the Netherlands.

Another aspect of spatial factors that is seldom considered in research, is the spatial coherence that travel times may or may not show. If, for now, we assume that more urbanised areas cause higher travel times, one can reasonably assume that high values for travel times tend to cluster together in large agglomerations that contain multiple adjacent cities (e.g. the Randstad). Likewise, one could expect clusters of low travel times in more rural areas. Applying methods of spatial data analyses could help identify these patterns and assess spatial equity, as is also shown by Wismadi et al. (2014). In current research however, this self-contained use of analysing spatial relationships tend to be ignored or underlit, as a more common application of spatial data analysis tend to focus on controlling for spatial errors (Hong et al., 2014). This is a critical limitation of the current state of the art of knowledge in this area, as identifying regions with excessively high or low travel times could help assess the equity of infrastructural access, of which the importance was stressed earlier. These presented knowledge gaps will be explored and highlighted further in the literature research.

1.4. Research questions

This research aims to address the identified knowledge gap surrounding the spatial variation of travel time expenditures in the Netherlands, in the hope of providing a better understanding of regional travel behaviour and its explanatory factors, which could provide valuable insights for infrastructure funds allocation. The main research question is:

How do travel times expenditures vary spatially across the Netherlands, and is there evidence to support the concept of constant travel times at a disaggregated scale?

To compartmentalise answering this question, three subquestions have been formulated.

1. *What is the state of existing knowledge on the spatial and socio-demographic determinants of travel time expenditures?*

To be able to know what possible relationships to look for in the data and to assess the value of any found patterns later on in the study, it is paramount to have a proper understanding of the current knowledge on spatial and socio-demographic factors that might influence travel behaviour. Being aware of what relationships are currently implied by existing literature can help search for and identify relevant patterns in the data and assess whether they constitute valuable findings. This question lays the groundwork to be built on in the rest of the research and assures that results are able to be interpreted in the proper context.

2. *What spatial patterns in travel time expenditures across municipalities can be identified?*

For the second subquestion, by applying spatial data analysis to travel time data on a municipal level, we look for any interesting spatial patterns in the data. This helps not only to assess differences between rural and urban regions, but also to identify regions with clustering of either high or low values of travel time expenditures. Answering this question allows for highlighting regional differences in travel time and assessing whether travel time expenditures are significantly higher for some regions, compared to other regions in the country. The outcomes of this question also supply a starting point for later parts of the research, as the identified spatial patterns constitute what exactly we want to explain in the third subquestion.

3. *How can spatial patterns in travel time expenditure across municipalities be explained by the spatial context?*

The found patterns and regions of interest from the second subquestion do not in itself constitute a satisfactory outcome of this research. To understand these patterns better, regression models are fitted to the data to explain the found patterns with (spatial and other) background variables. Answering this third subquestion allows to look further than just identifying regions of interest, but helps in understanding and explaining why spatial differences might exist. Creating a proper understanding of this could eventually help in mitigating any inequities in infrastructure accessibility by implementing policies that target the relevant underlying factors.

1.5. Research methods

The overarching research design takes the form of a mixed methods approach, as the varying nature of the different research steps asks for different approaches to go about them, though the main analytical core of consists of spatial data analysis. A first assessment of the state of knowledge on travel time variations across regions and individuals is made by studying the literature. We aim to create a comprehensive understanding of the current knowledge in this field, also looking at specific applications at a disaggregated level, as that is what this study will focus on. Search factors include the BREVER-law, travel time budgets and expenditures, spatial characteristics, socio-demographic differences in travel behaviour, and other related areas of study. In the literature search particular attention is also dedicated to the method of analysing spatial autocorrelation, as this method will be applied later in this study to identify patterns in the data. More elaboration on the search strategy of this literature review is presented in chapter 2.

Following the literature search, the analytical body of this research consists of various methods of spatial data analysis. For this study the ODiN datasets from 2018 and 2019 are used (CBS, n.d.-a). These data contain a day's worth of travel information for tens of thousands of respondents, in the form of reported trips in a travel journal. The ODiN data collection process has been running for multiple years, and in this research the data from two years are combined to ensure sufficient data coverage. As the data contains travel data, but also data on socio-demographic characteristics, it can be used for a variety of analyses. This data is combined with geometric data maps from CBS, to allow for incorporation of spatial elements of the data. To ensure data quality, the data are subjected to rigorous data preparation, of which extensive documentation is presented in chapter 3.

To determine whether spatial differences in travel behaviour do exist (and with it, to test for regionally disaggregated applicability of the BREVER-law), it is important to compare cross-regional data on travel patterns. To this end, spatial autocorrelation is applied to assess whether travel data on a municipal disaggregated level exhibits spatial patterns, like clustering, and whether values differ significantly between regions. As mentioned earlier, studies on urban effects of travel time have yielded inconclusive results and this spatial analysis poses a new approach to this question. This spatial autocorrelation assessment of travel times is succeeded by a variety of additional spatial and non-spatial regression analyses. These regression analyses are used to find influential factors that can help explain the found spatial differences and/or patterns in travel times between/across regions. Here, both spatial and non-spatial regressions are applied to determine what additional explanatory power spatial factors might add to the models. All analyses are performed at the municipal level and focus on the municipality of residence. Further specification of the applied data analysis methods is presented in chapter 4.

1.6. Scope of the study

Firstly, as the research is performed for the Netherlands Institute for Transport Policy Analysis (KiM), a research institute within the Dutch Ministry of Infrastructure & Water Management, the geographical scope of this study is the Netherlands. This means that only travel data from the Netherlands is considered. More specifically, this concerns travel data from persons residing in the Netherlands. While this may include travel that has (partly) taken place outside of the Netherlands, all included data concerns subjects that reside in the Netherlands and whose travel behaviour on the recorded day at least partly took place in the Netherlands. This single-geography focus is rooted in the considerations that (1) the study aims to address disaggregated within-county differences, (2) Dutch-based research into travel times is lacking, and (3) the study aims be relevant for the Dutch Ministry of Infrastructure & Water Management. The first two considerations are supported by the literature presented in chapter 2.

Secondly, this study focuses on the notion of constant total travel times specifically. In the original formulation of the BREVER-law, both travel times and trip rates are deemed to be constant (Hupkes, 1977). This study has its focus on only the travel times aspect (the 'RE' in BREVER). This notion of travel times being constant includes walking and undirected trips and thus contains all forms of travel (Hupkes, 1982; Marchetti, 1994). Therefore, in this study the total travel time per person is considered the variable of analysis, taking into account all trips, regardless of their purpose or mode of transport. Any disaggregation into and possible differences between travel modes or purpose are considered to be outside the scope of this study. Because the focus is on total travel time, this also means that no recorded travel is excluded from the data beforehand. As also further explained in section 2.2, travel time expenditures are used to serve as proxies for the 'budgets' mentioned by Hupkes (Hupkes, 1982).

This study focuses specifically on spatial differences in travel times and therefore spatial disaggregation is necessary for the analyses. This research takes municipalities as the unit of analysis. The main reason for the decision to perform the analyses at the municipal level is linked to the policy relevance. For considering infrastructure investments, municipal measures of travel patterns give a better picture of the total activity in the region, which is more important to consider than individual movement patterns. This choice for the municipal level comes with additional data benefits as well. Even though the number of municipalities is decreasing over time due to merging, in 2019 there still were 355 municipalities in the Netherlands (CBS, n.d.-b). Additionally, recorded characteristics per municipality are plenty and readily available through CBS (2021), which allow for explanatory analyses using these characteristics as well. Though these characteristics are also available at a smaller (neighbourhood) geographical level, the municipal level is chosen to also better assure data sufficiency of the travel data. In order to draw meaningful conclusions from the analyses, sufficient data per area of analysis is needed, which is harder to satisfy when analysing smaller regions.

The main analyses in this study make use of Dutch national travel survey data from 2018 and 2019. Though more recent travel data from the Netherlands is available, travel behaviour from the period 2020-2022 has been impacted significantly by the COVID-19 pandemic. The pandemic has sparked an in increase in working-from-home, decreasing commute travel and thus causing a significant drop in travel times (Borkowski et al., 2021; Faber, Hamersma, Brimaire, et al., 2023). Because this societal shock of the pandemic seems to have (temporarily) altered travel behaviour, pre-pandemic data from 2018 and 2019 is taken to be the most recent data from a 'normal' situation without major distortions, and thus the most suitable for this research.

1.7. Outline of the report

Chapter 2 of this report presents an overview of the most important and relevant literature on travel time budgets and expenditures, possible explanatory factors and spatial data analysis. Chapter 3 describes the origin of the used data and the actions that have been performed to adequately preprocess the data for this research. This is followed by elaboration on the data analysis techniques used, explaining both spatial autocorrelation and the regression analyses in chapter 4. The results of these performed analyses are presented in chapter 5. Finally, in chapter 6 the conclusions are covered, as well as their implications and recommendations that follow from this research.

2 Literature

To assess the current state of knowledge on travel time budgets/expenditures, and specifically to gain more insight into this notion at disaggregated levels, a literature study has been performed. The subject of this literature review includes the wider notion of the BREVER-law and travel times, with a more specific focus on disaggregated studies (both in the Netherlands and other geographies) that consider socio-demographic and spatial factors.

The literature search was mainly performed using Google Scholar, with some additional papers found using Scopus. Searches were performed on a variety of queries, using combinations of keywords including, but not limited to: ’travel time budget’, ’travel time expenditure’, ’travel behaviour’, ’BREVER(-law)’, ’Netherlands’, ’disaggregated’, ’urbanism’, ’built environment’, ’spatial autocorrelation’, and ’socio-demographic’. A more extensive (but not all-encompassing) specification of used queries and yields is presented in table 2.1. This table presents the most important queries used to start off the literature search, but other search terms and minor adjustments to these queries have also been used to identify additional relevant material.

Table 2.1: Used queries and results in Google Scholar

Topic	Query	Article type	Results	Included
Disaggregation (general)	“constant travel time” AND disaggregated	Review articles	21	3
Disaggregation (general)	“travel time (budget OR expenditure)” AND disaggregated	Review articles	49	6
Netherlands	“travel time (budget OR expenditure)” AND Netherlands	Review articles	76	7
Netherlands	“constant travel time” AND Netherlands	Review articles	36	3
Built environment	“travel time (budget OR expenditure)” AND (“land use” OR “built environment” OR “urban environment” OR “urbanism”)	Review articles	115	8
Spatial autocorrelation	“travel time (budget OR expenditure)” AND “spatial autocorrelation”	All	44	2
Spatial autocorrelation	“travel behavior” AND “spatial autocorrelation”	Review articles	84	1
Socio-demographics	“travel time (budget OR expenditure)” AND ”(demographic OR socio-economic) variables”	All	134	5

For some queries, the number of results was so large, that the decision was made to focus only on research articles, which is indicated in the *Article type* column. This way, the number of papers was kept manageable, while still taking into account papers that present reviews on the larger body of literature. Based on the title and abstract, and sometimes scanning of the full paper, it was assessed whether papers were relevant for this research. The columns *Results* and *Included* present the number of papers the search yielded and the number of papers that are included in this report, respectively. It must be stated that there does exist some overlap in the included papers from the different queries. The total number of unique included papers that were identified through this process is 15. Additionally, backwards snowballing has been applied to already found papers to identify more relevant material in this field of knowledge that did not show up in the original search results.

2.1. Elaboration on the BREVER-law

As already introduced in the previous chapter, the BREVER-law embodies the concept of constant travel times, but this deserves some more elaboration. In a paper in which he aims to rectify some critiques on his earlier presentation of this concept, Hupkes offers two possible underlying mechanisms that can further explain this observed phenomenon: a bio-psychological approach and an utility-optimising approach (Hupkes, 1982, p.41). The bio-psychological approach basically reasons that humans are creatures of habit and resistant to change, which results in rigid behavioural patterns. The utility-optimising approach considers travel utility to be two-fold, consisting of intrinsic utility of the travel itself and the derived utility of being able to get somewhere. Hupkes argues that both aspects of travel utility are positive in the beginning, but decline later, meaning that the utility curve will have an optimum. As people tend to strive for this optimum, this would 'set' ones (constant) travel time. This would mean that after any changes in travel options and or speed, people adapt their behaviour to end up at the same optimum value of travel time of around 70 minutes (Mokhtarian & Chen, 2004). Plenty of literature points to an abundance of empirical evidence for this notion, which is most elegantly visualised in figure 2.1.

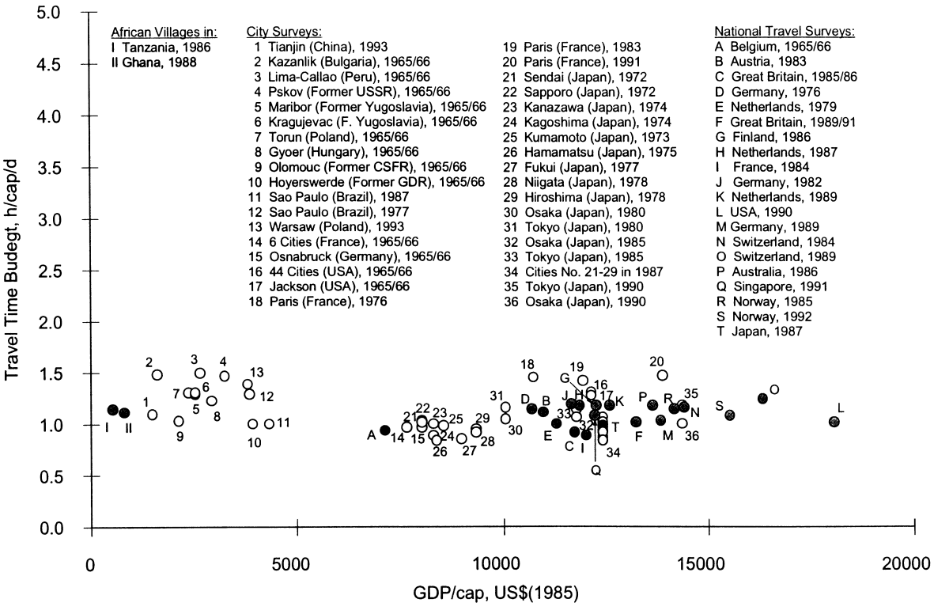


Figure 2.1: Average travel time budgets per person for different geographies and years (Schafer & Victor, 2000, p.175)

The idea of constant travel times is sometimes also presented as Marchetti's constant, referring to a paper on this matter by Italian physicist Cesare Marchetti (1994). Marchetti, quite like Hupkes' bio-psychological approach, also attributes travel times being constant to instinct and gives some interesting examples. He states that even in a prison, inmates tend to walk around for one hour per day (pp.75-76). He also presents the point that a set daily travel time, combined with a given travel speed, gives a 'territorial area'. Using figure 2.2, he demonstrates that this area has grown with an increase in speed, due to technological advancements. This aligns with the idea that travel speed improvements do not cause travel time savings, but rather increase distances covered and thus total travel volumes (Hupkes, 1982). Allowing people to travel quicker, thus makes them travel further but not shorter times. This would imply that any technological advancements might cause an increase in traffic volume, instead of a decrease, in a similar way to how widening a highway leads to increased traffic flows, but not less congestion, as it also attracts more travel (Metz, 2021).

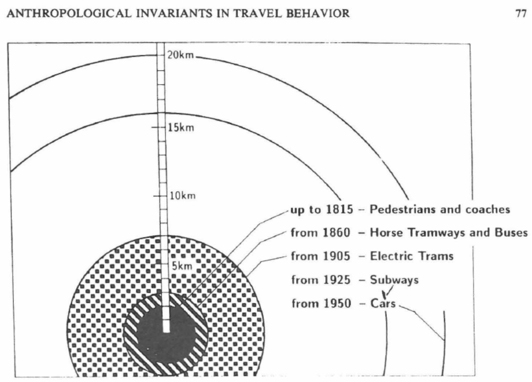


Figure 2.2: City dimensions and speed of transport: The case of Berlin. With every technological advancement, the city is able to increase in size, without increasing the time it takes to get to the centre (Marchetti, 1994, p.77)

It must be stressed, again, that the idea of constant travel times applies at a high level of aggregation and variations in results could be attributed to differences in the used geographical scales (Chen & Mokhtarian, 2008). In his clarifying paper, Hupkes himself said about the travel utility curves that "the curves will differ for each individual person" (Hupkes, 1982, p.42). People have different attitudes towards travel, so not everyone will derive the same utility from it. Even the report that introduced the general concept of a travel time budget in the first place, claims that such a budget can differ by person and vary with location (Zahavi, 1974). However, the idea is that across a larger population these variation average out to a constant value. While there thus exists variety between individuals, a single individual's budget across time appears to be roughly constant (Stopher et al., 2017). Still, both Hupkes and Zahavi deemed it imaginable that aggregate budgets can change over time, for example due to technological breakthroughs (Hupkes, 1982). Assuming that both these premises are true, this would imply that aggregate changes over time can change only with the speed of population change, as individuals will not be affected, but society as a whole will be. This could explain both constant budgets in the short run and changes in budget in the long run.

Like the original cross-national studies from the 1960s and 1970s by Szalai (1966a, 1966b, 1975), most studies supporting the notion use large-scale data and only claim any validity at a very high aggregation level (Ahmed & Stopher, 2014; Mokhtarian & Chen, 2004). Variations across geographies, individuals, and time are thus widely accepted as well. This leaves us with the question what exact aspect of this notion should be considered constant. The term 'constant' itself can imply constant values over both space and/or time (van Wee et al., 2006). The original studies by Szalai (1966a) and Zahavi (1974) mainly imply cross-sectional consistency at an aggregate (country) level across geographies. The general notion thus implies a rough consistency of the average values across large groups (e.g. whole countries), but does not rule out any significant variations within these larger groups between smaller regions (e.g. cities) or between subgroups (e.g. based on age, gender, occupation). Even though both Marchetti (1994) and Hupkes (1977) do originally mention a stable travel time over history, evidence does not necessarily support the claim that travel times are constant across time, but are subject to change (Gunn, 1979).

2.2. Travel time budget v.s. travel time expenditure

In the literature on travel behaviour, and more specifically travel times, the terminology used tends to be fairly inconsistent. The terms 'travel time budget' and 'travel time expenditure' are often confused with each other and are sometimes used interchangeably (Mokhtarian & Chen, 2004). However, to ensure we actually know what we are researching and interpret results correctly, it is important to very clearly distinguish between these two concepts. Even though they are related, they are definitely not the same. In order to avoid any confusion, we here clear up differences between them, how these terms are interpreted, and which terminology is used in this research.

Ahmed & Stopher frame the meaning of a budget very clearly based on a number of dictionary definitions (Ahmed & Stopher, 2014, p.609). The most striking terms here are "expected expense", "amount available", "plan for coordination of expenditures", and "restriction on expenditure"[5]. Bases on these definitions, the term travel time budget is hereby interpreted as: the amount of time one is *willing* to spend on travelling. This is an intrinsic personal variable and is very hard to measure in itself. We can however relate it to travel time expenditure, as a 'budget acts as a controlling measure over actual expenditures' (Ahmed & Stopher, 2014, p.609).

Travel time expenditure is considered to be the actual, measurable, real-world time that one spends on travel. While expenditures will probably not be identical to budgets, it can be expected that one's expenditure is at least in part determined by one's budget, with some deviations. Because of this, while thus not the same, travel time expenditures have in the past often been used as proxies for determining travel time budgets (Mokhtarian & Chen, 2004). One can assume that personal deviations in expenditure from the budget might average out over large samples, which could explain why aggregate studies have more often been able to conclude that a constant travel time budget exists, as opposed to disaggregate studies (Ahmed & Stopher, 2014; Stopher & Zhang, 2011). Still, as Ahmed & Stopher point out, measuring expenditures does not allow for estimation of any unused travel time budget. There might be more budget than people actually spend, or people could be forced to spend a bit more than their budget. However, as the budget in itself thus remains unmeasurable, in this research we will take expenditures as the variable of analysis and consider this a sufficient proxy to assess differences in travel behaviour and look into the BREVER-law. In doing this, we consider any individual positive and negative deviations between expenditures and budgets to cancel out at the group level, where expenditures will roughly equal budgets (Stopher & Zhang, 2011).

2.3. Disaggregated studies

The notion of constant travel times thus only applies at aggregated (national) scales. To dive deeper into possible explanations of differences, more literature on disaggregated studies is being considered. Literature on the effects of individual socio-demographic and spatial characteristics on travel behaviour is discussed.

2.3.1. The effects of socio-demographic variables

In a comparative studies between countries by Timmermans et al. (2002, 2003), it was presented that significant correlations exist between socio-demographic group membership and time allocation patterns, which includes travel time. Their study however based their analysis on predefined socio-demographic groups. Later on, more elaborate analyses have been performed that analyse the effect of socio-demographic variables separately and at an individual level (Raux, Ma, Joly, & Cornelis, 2011; Raux, Ma, Joly, Kaufmann, et al., 2011). From their analysis of travel surveys from eight European cities in three countries (France, Switzerland, & Belgium), the authors conclude that socio-demographic characteristics are indeed very important in one's time allocation decisions, which once again aligns with the idea that travel time budget stability only exists for broadly averaged groups (Wigan & Morris, 1981). Raux, Ma, Joly, & Cornelis (2011) find that marital status and gender are the two most important factors influencing (travel) time allocation.

Especially the differences by gender are often highlighted in other studies as well. Wigan & Morris (1981) find higher travel times for males in a 1977 Stockholm-based dataset. In UK data from 2003 to 2008 it was found that men tend to mainly spend more time on commuting, whereas their time expenditures for shopping are lower than for women (Jahanshahi et al., 2013). No general conclusion on total travel time is presented, but the numbers seem to imply little to no gender difference between genders for this cumulative value. When splitting up travel time into the categories 'maintenance', 'discretionary', and 'other', males mainly exhibited higher discretionary travel times only (Koo et al., 2022). Most studies tend to show that men travel more than women, though there are also some studies that find no significant difference or even an opposite relationship (Mokhtarian & Chen, 2004).

[5] Original definitions from: Dictionary.com (n.d.), Merriam-Webster (n.d.), Oxford English Dictionary (n.d.)

The presence of children in one's household is also often presented in literature as having a significant effect on travel times. Like the effect of gender, there is no real consensus on the direction of this relationship between having children and travel time expenditures. Raux, Ma, Joly, Kaufmann et al. (2011) find that persons from households with children under 12 appear to have lower travel time budgets. However, when once again taking into account gender differences, Wigan & Morris (1981) find that only women's travel times are lower when children are present, but a male's travel time tends to increase. More generally speaking, they state that ”timing constraints...are essentially a function of the stage in the family life cycle” (p.84). When differentiating between trip categories, maintenance travel specifically increases with the presence of children, in the form of more commuting to school and more shopping/personal business travel time (Koo et al., 2022). Results on the presence of children are thus not really unanimous.

When considering the factor age, people under 15 have the lowest travel time budgets, but there also exists a difference between the the age group 15 to 65, and 65 and above. The older population (>65 years of age) travel less that the population group of 15 to 65 years of age (Raux, Ma, Joly, & Cornelis, 2011). One may expect that this can be attributed to retirement at older age and thus fewer work-related travel requirements, but in the presented model work status is also included. These two factors thus both represent a separate significant effect on travel times. Even when distinguishing between maintenance and leisure travel, a decrease in both forms of travel can be found for the older age group (Koo et al., 2022). As a possible explanation of the decrease in leisure travel with age (where you could expect an increase due to more leisure time during retirement), they mention that travel itself also becomes more inconvenient with age (p.14). The broader literature study by Mokhtarian & Chen (2004) underlines the conclusion that persons in the 'working age' category travel the most.

Besides biological factors like age and gender, factors on a more economic ground are also important in assessing travel time differences. In general, working hours are positively correlated with commute times (Jahanshahi et al., 2013; Koo et al., 2022). This intuitively makes sense, though this relationship is partly confounded by gender (Mokhtarian & Chen, 2004). Both studies also show that an increase in working hours tends to decrease the time spent on non-work-related travel. While neither study explicitly mentions the effect on total travel time, this does point to some form of trade-off between work-related travel and shopping- and/or maintenance travel. Interestingly, a higher income is also correlated with more commute time (Jahanshahi et al., 2013). Koo et al. (2022) even find that a higher income has a positive effect on all kinds of travel, so also on travel for non-work-related purposes. The nature of one's job also has an impact, as commute times appear higher for households where the head of households has a professional or clerical job, compared to a manual or skilled manual job. Once again though, broader research on income and travel times yield varying outcomes (Mokhtarian & Chen, 2004). Also interesting to note is that based on a study in El Paso, Texas, the impact of the COVID-19 pandemic on travel varies with socio-economic status. Low-income and/or low-education households showed a smaller drop in non-work-related travel expenditure during the pandemic (Song et al., 2023).

Another interesting factor is car ownership. One can reasonably expect that those owning a car might travel more by car and consequently less by other means, but this does not necessarily imply anything regarding the total travel time expenditure. In the broad Dutch study mentioned earlier it is reported that people from non-car-owning household tend to have lower travel times, mainly during the work-week, but this effect is limited and not significant (van der Hoorn, 1979). This lack of a substantiated relationship between car ownership and total travel times is also more broadly supported by the cross-national/cross-city study by Raux et al. (2011). However, other research from the UK and Australia does point to a significant increase in travel times with car ownership (Jahanshahi et al., 2013; Wigan & Morris, 1981). Where the former only finds an increase for work-related travel, the latter presents an increase in general travel times. Once again, there is some evidence to support this relationship, but the literature is not unanimous on the matter, as contrary results have also been found (Mokhtarian & Chen, 2004).

A final thing to consider in this section, is that the influence of socio-demographic variables can root deeper than this. They might also exhibit indirect influences through other intermediate variables or confound effects of other factors, like the built environment (Zhou et al., 2022). The fact that these confounding interaction relationships exist is one of the main reasons why one must include socio-demographic variables in this otherwise spatially oriented research anyway. Socio-demographic variables play an important part in residential self-selection (Jahanshahi et al., 2013). As these spatial/residential variables on their turn can influence travel behaviour as well (as will be covered in the next section), they allow for indirect influence of socio-demographic variables on travel behaviour as well. Jahanshahi et al. (2013) also mention that these confounding effects are mainly strong for work-related travel and less so for shopping travel. Inclusion of both socio-demographic and spatial factors in analysis allows for correctly estimating the true, non-confounded, effects of factors on travel. This is especially important, as indirect effects can even outweigh direct effects (Koo et al., 2022). In this paper, Koo et al. also show the significance of a latent socio-demographic variable (vitality) they identified and stress the point that even not (directly) measurable variables can impact travel behaviour.

From the ample research on relationships between socio-demographic variables and travel times, it can be concluded that these variables do definitely impact travel patterns. It is clear that these factors are important to take into account, even though the exact influence (if any) of specific factors is not always unanimously agreed upon.

2.3.2. The effects of the spatial context

Earlier mentioned studies also present disaggregated differences in travel times when considering different regions and/or geographies. Van der Hoorn (1979) found that in the Netherlands, people living in dense urban areas do show the highest values for total travel time. This is especially true in the biggest three cities (Amsterdam, Rotterdam, The Hague). This variety does not only show up within research on this geography, but across multiple countries as well. In an Italian cross-city study, significant differences in travel time budgets between cities were also found (Gallotti et al., 2015). However, similar to the socio-demographic variables, the results are not so one-sided. A cross-national study that includes Dutch data concludes that ”travel patterns are largely independent from spatial setting, except for some extreme cases” (Timmermans et al., 2003, p.37).

Once again we refer to the review by Mokhtarian & Chen (2004) for an extensive overview. Most papers do point towards a positive relationship between city size and density and travel times. As possible explanation for this phenomenon, they pose that this could be due to ”variation in activity opportunities and transport services, as well as people’s lifestyles” (p.664). Still, they also mention a study that fails to yield a significant result and even one that finds longer travel times for outlying rural areas compared to an urban centre (Barnes & Davis, 2001; Downes & Morrell, 1981). Based on the broader body of literature, the wider accepted believe is that people tend have (slightly) higher travel times in cities, even though these effects of location appear to be smaller than the effects of personal and/or household characteristics (van Wee et al., 2006).

In an attempt to explain why people in larger cities appear to spend more time on travel, Van Wee et al. (2006) consider the broader development of urbanisation over time. They reason that with an the outward expansion of cities over time, new residential areas in the outskirts are being built ever further away from the city centre (p.115). This leaves people having to travel longer for the same locational utility. In the same paragraph they also mention a decrease in the number of service locations (e.g. hospitals), increasing the average distance to the nearest one. While this would explain an increase in travel times in cities over time, this would not exclude rural areas from suffering these consequences too. To test the application of this theory of Van Wee et al., a specific paper on the case of Shenzhen, China can be of use (Zhou et al., 2022). In this paper the the growth of the city and travel behaviour developments were analysed in unison, testing whether urban expansion indeed leads to higher travel times. Based on multi-year travel data, it was however concluded that while the city grew significantly, travel times stayed essentially stable. The researchers attributes this to the simultaneous enhancement of mobility in the city, which included car ownership growth and public transport development (p.14).

A different angle for explaining differences in travel behaviour between areas does not focus on city size or population density, but rather on the spatial structure of the environment itself (Mokhtarian & Chen, 2004). Gordon et al. (1989) provide an extensive description analysis of the difference between monocentric and polycentric urban areas. They state that in monocentric cities travel times tend to increase with with urban growth (in population size and density), mainly due to an increase in congestion. The same goes for larger urban land areas, as more distance will have to be covered. However, they state, for polycentric urban areas the relationships are not as clear-cut. Polycentrism could, in theory, lead to more evenly spread travel and thus lower congestion and travel times, even for larger cities. This relationship however is also not always inherently true, as this would depend on the home choices of the workers (Gordon et al., 1989, p.140). The transition from a monocentric to polycentric structure (due to excessive city growth) could potentially also drastically lower travel times, though there is little empirical proof of this (p.141). Changes in urban structure could also include improvements in infrastructure, which can decrease travel times (Martín-Barroso et al., 2022; Zhou et al., 2022).

Zhou et al. (2022) also mention that the effect of the built environment on travel behaviour might not be the same for all individuals. Persons with a predefined preference for a certain mode of transportation appeared to not be impacted as heavily by their environment, compared to people that do not have such a strong preference. Difference in travel attitudes thus matter for the effects of environment on travel behaviour. This relationship is highlighted more in-depth by Guan et al. (2020). They present the concept of residential self-selection, meaning that people are more likely to choose to reside in an area that best fits their (travel) preferences and attitudes. Still, even when controlling for this, they do find an independent effect of the built environment on travel, though they were not able to identify what role exactly attitudes play, due to the complexity of the relationships (p.14).

Besides differences in impact between individuals due to their characteristics or attitudes, the impact of spatial factors on travel behaviour can vary by location itself as well. This concept is called spatial heterogeneity (Bhat & Zhao, 2002; Zhou et al., 2022). While this is presented as a separate concept, there are some parallels to draw with the presented notion of attitudes and residential self-selection. After all, if persons with a certain attitude show similar sensitivity to the built environment while also residing in the same areas, as compared to persons with different attitudes, this would inherently result in spatial heterogeneity of the effects of the built environment.

One may notice that most of the presented literature so far has their primary focus on urban environments, or at best draw some conclusions on how travel behaviour in rural regions compare to behaviour in urban environments. Even a very recent extensive literature review on the new application of using smart card data for mobility analysis does not mention any literature on travel behaviour in rural areas (Cats, 2023). More detailed analyses of/between rural municipalities are thus not prominently represented, though they do exist. In an older granular research from the 1979 similar daily travel times were found across rural areas in the UK, which were also similar in values to urban areas, except for Greater London (Landrock, 1981). In a more recent estimative research in Spain based on census data from 2001 and 2011, the effects of municipalities of work and residence were extensively analysed (Martín-Barroso et al., 2022). They found, once again, that larger municipalities are associated with longer commute times. This applies to both the municipality of residence and the municipality of work, and also if one lives and works in the same municipality. Another less often investigated factor was also found to have an interesting effect, namely unemployment. Longer commute times are associated with high unemployment in the municipality of residence, but also with low unemployment in the municipality of work.

In general, though the body of literature is already vast, plenty papers state in their recommendations that further research into the effects of spatial factors on travel behaviour could be valuable (Cats, 2023; Timmermans et al., 2003; van Wee et al., 2006). To add a more specific insight from this literature assessment, additional research focusing more on non-urban municipalities with a higher level of granularity could help identify new patterns, as there is a lack of these detailed studies.

2.4. Research in the Netherlands

As mentioned, travel may differ by location and most presented literature covers research from a variety of different geographies. To focus more on our geography of interest, disaggregated intranational studies within the Netherlands and cross-national studies comparing Dutch travel data to other geographies can provide more insight into travel behaviour in the Netherlands specifically. In a study based on the 2008 Netherlands Mobility Survey the average travel time per day in the Randstad was found to be 70 minutes (Feng et al., 2013). This number, as aggregate value, is in line with earlier found values for western urban areas by Schafer & Victor (2000). In an older study based on 1975 data from the Dutch Social Cultural Planning Bureau, a yearly travel time budget of 425 hours was presented, which converted is also equal to 70 minutes per day (van der Hoorn, 1979). At the national level, travel times in the Netherlands thus seem to be constant. In studies comparing Dutch travel patterns (specifically the South-Rotterdam region) to other geographies in the USA, UK, Canada, and Japan, is it also underlined that there are significant similarities between these regions (Timmermans et al., 2002, 2003).

While these aforementioned numbers also imply the travel time being constant over time, van Wee et al. (2006) actually present significant increases in average travel time in the Netherlands. Based on two separate multi-year datasets, the travel time expenditure seems so have increased by from 7% between 1979 and 1998, to as much as 26% between 1975 and 2000. In the paper, they attribute the large difference in these results to a varying nature of the data, as well as the difference in time span analysed. It must be noted that these presented studies are based on fairly old data from the previous century, so one must be cautious with extrapolating to the modern day and drawing conclusions based on this, especially given the varying results that they present.

When looking more at disaggregated studies within the Netherlands, one discovers some individual and regional variations. When considering differences at the individual level, Van der Hoorn (1979) finds that schoolchildren, students, and working men and women have higher travel times than housewives. Feng et al. (2013) present more elaborate results and in this they differentiate between car-owning and non-car-owning households. In non-car-owning households in the Netherlands, women travel more than men and persons from a single household travel more than persons from a family household. In car-owning households, persons with a high education, high income, and high number of workers in one's household tend have higher travel times, while older people have lower travel times. In both households with or without a car, persons with a job tend to travel more than those without a job.

Some studies also comment on differences in travel times in the Netherlands based on spatial context. In their comparative study between the Netherlands and other regions, Timmermans et al. (2003) were not able to find significant variations in travel patterns between different spatial contexts. While they do mention a tendency of people living in suburban or poorly connected areas to chain more destinations, this relationship was not significant. They also do not make any mention of total travel times. Still, they do note that disaggregated intranational comparisons tend to show more variety than cross-national comparison between different geographies (Timmermans et al., 2002, p.92). These studies however only considered the South-Rotterdam region in the Netherlands. The broader Dutch study by Van der Hoorn (1979) did also yield the conclusion that trip rates are similar for all degrees of urbanisation, but also shows that travel time are actually higher in large cities than in rural areas. From the Randstad study by Feng et al. (2013) it was concluded that, while population density and distance to a city centre were not significant predictors for travel times, a better land use mix is associated with lower travel times.

While these few presented studies do offer some more insight into disaggregated individual and regional differences in the Netherlands, specific research into this geography appears to be fairly limited, thus supporting the case for more (intra)national research.

2.5. Applications of spatial analysis

Most papers on the effect of the spatial context on travel behaviour, including the papers mentioned in this chapter so far, do not touch upon the spatial relationships that might exist between travel behaviour across (adjacent) regions. Literature tends to present some form of non-spatial/regular analysis on differences across contexts, while not incorporating spatial coherence. As mentioned in the introduction, an effect of urbanism or density on travel times would likely result in specific spatial patterns (e.g. clustering) of values of travel time, like adjacent high values in urban areas and/or adjacent low values in rural areas. Looking at these patterns at the regional (or in this case, municipal) level could help identify patterns in the data that perhaps not immediately become apparent from analyses at individual levels (Rey et al., 2023, Ch.9)[6]. One of the main methods Rey et al. present for incorporating this is spatial autocorrelation, which essentially provides a measure for whether significant spatial patterns exists in the data. This method builds on Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970) This exact method will be further elaborated on in chapter 4.

There exist ample examples in a variety of fields of this self-contained application of spatial autocorrelation. In the field of health services spatial autocorrelation is very commonly used as a measure for spatial equity, according to an extensive literature review by Whitehead et al. (2019). Though not many, a few examples of this applications of spatial autocorrelation in the field of travel research were also found. In a specific application considering infrastructure resource allocation in Yogyakarta, Indonesia, spatial autocorrelation of travel times was used to assess spatial equity (Wismadi et al., 2014). In research into travel behaviour in Salt Lake City, UT, the identification of (negative) spatial autocorrelation was also used in drawing conclusions on the (in this case) non-clustering of active mode users (C. H. Wang et al., 2021).

Though these applications prove the added value of this method as presented by Rey et al., one must be wary of ecological fallacy when using it (Rey et al., 2023, Ch.9). While these spatial insights might be valuable, they remain aggregated and cannot directly support conclusions at the individual level. Moreover, the results of a spatial analysis can be dependent on the (ad-hoc) aggregation applied to the data (Pereira, 2019). This is known as the Modifiable Areal Unit Problem (MAUP) and it has been shown that both the method and the level of aggregation can significantly impact the outcome of spatial analyses (Apparicio et al., 2008; Tan & Samsudin, 2017). These limitations will have to be taken into account when drawing conclusions from the performed analyses.

Spatial autocorrelation can thus be used as a self-contained method for analysing spatial patterns in data. However, more often it is not exactly used in this manner, but rather as a method for error correction in (non-spatial) analyses. Standard regression analyses assume independence between observations and spatial dependence violates this fundamental prerequisite for a regression analysis (Pereira, 2019). Therefore, spatial patterns are often seen as problematic and as something that needs to be controlled for, as is done in Pereira's research on employment accessibility equity in Rio de Janeiro. Papers in the realm of travel behaviour more often apply spatial autocorrelation in this manner, as a way of correcting for errors in non-spatial analyses (Hong et al., 2014; Rodrigues & Targa, 2004; F. Wang, 2012). Even an extensive literature review on the applications of spatial data mining only highlights this error-correction application of spatial autocorrelation (Shekhar et al., 2011). There thus still is ground to be gained for the self-contained application of spatial analysis for identifying spatial patterns in travel behaviour.

2.6. Conceptual model based on the literature

Based on the information presented in this chapter, a conceptual model can be constructed to visualise what factors and relationships are highlighted in the literature. This conceptual model is shown in figure 2.3. The total travel time expenditure is the main variable of interest in this research and is therefore presented in the bold box. It is expected that there could be (internal) spatial coherence in high/low values of travel time (1), which is the main relationship of interest of this study. This coherence could potentially be explained by the spatial characteristics of the different regions that might affect travel times (2). However, the effect of spatial characteristics on travel times might also be confounded by

[6] Rey et al. apply this specifically to regional income inequality specifically, but the general idea of the method applies to a broader array of spatial differences

personal variables. This needs to be taken into account and controlled for. Spatial characteristics might be influenced by personal attitudes due to residential self-selection (3). Socio-demographics also seem to directly influence travel times (4). Lastly, due to individual and spatial heterogeneity there might be interaction effects, as the effect of spatial characteristics might not be the same for all individuals (5) or all regions (6). The conceptual relationships are supported by the identified literature. To summarise, a short explanation with supporting citations is provided for each of the six arrows shown in figure 2.3.

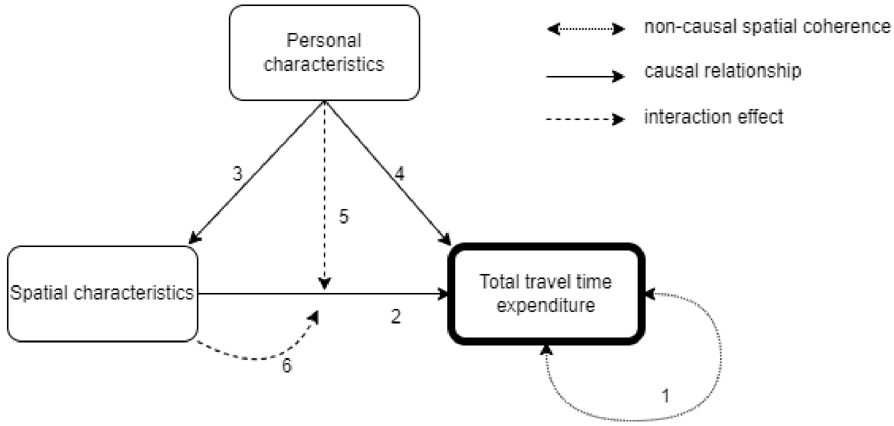


Figure 2.3: Conceptual model of the (groups of) variables and their relationships identified in the literature

(1) Based on Tobler's First Law of Geography, it could be expected that adjacent regions show more similarity in travel behaviour than further separated regions (Tobler, 1970). Identifying spatial patterns could help assess spatial equity of infrastructure (Wismadi et al., 2014).

(2) Spatial characteristics could explain regional differences in travel patterns, and thus possibly explain any spatial coherence as well. One of the most prominent effects is that people tend to travel a bit more in cities, compared to rural areas (Mokhtarian & Chen, 2004; van Wee et al., 2006; van der Hoorn, 1979).

(3) Due to the concept of residential self-selection, people tend to chose to reside in areas that best fit their preferences (Guan et al., 2020). Because of this, personal characteristics might, in a way, influence one's spatial context as well.

(4) Socio-demographic variables also directly influence travel times, as they tend to vary with an individual's or household's characteristics. While somewhat disputed, travel times seem to be higher for, among other factors, men, people between 15 and 65 years of age, and with higher income (Koo et al., 2022; Raux, Ma, Joly, & Cornelis, 2011; Wigan & Morris, 1981).

(5) Persons might also be more or less sensitive to spatial characteristics, bases on their predefined preferences regarding travel (Zhou et al., 2022). There is personal heterogeneity in relationship (2).

(6) The sensitivity to spatial characteristics also appear to vary by location itself, which is known as spatial heterogeneity (Bhat & Zhao, 2002; Zhou et al., 2022). This heterogeneity might be connected to residential self-selection (3) and personal heterogeneity (5).

Based on this literature review, we thus want to look further into the effect of both personal characteristics and spatial characteristics, based on our own data from the Netherlands. Regarding the personal characteristics, there are seven main factors of which (some) literature substantiates their influence on travel behaviour, and therefore we will consider these factors in our analyses. These factors are: sex, age, education, labour participation, presence of children in the household, car ownership, and income. Now, we ourselves will add one more factor to this list, namely migration background. Though the originally identified literature did not make mention of a relationship between one's migration status and their travel behaviour, it is generally known that immigrants tend to be more economically marginalised than natives. Lee et al. (2021) find a lower car ownership for immigrants in the United States, which could alter travel behaviour, so we want to see if one's migration background itself might also have an effect on travel time expenditures.

The spatial characteristics that we will consider further in this research, are a bit more vaguely derived from the literature. The four main urban environment factors that are mentioned in travel behaviour research are: urban density, city/municipality size, spatial structure, and (un)employment level. Urban density is vaguely, dually defined, as this could indicate both population density, or density in terms of buildings per area. These do not necessarily embody the same concept, so they will both be considered further. City/municipality we further consider to be represented by population numbers. Regarding the spatial structure, the Netherlands does not have enough cities of a large enough size to allow for proper comparison between mono- and polycentric cities, so unfortunately we cannot consider this factor further in this research. Lastly, (un)employment levels could be considered an indicator of economic activity, which we take as a more general notion of concept for this study.

How exactly these concepts of interest are further derived from the used data, what variables are used to represent them, and how they are incorporated in further analysis, is presented in sections 3.2.1, 3.3.1, and chapter 4.2. It must be noted that not all relationships shown in figure 2.6 are actually tested for in this research. In section 4.2.4, we further elaborate on the actually tested relationships.

3 Research Data

As presented in chapter 1, this research looks at geographical variety in travel time expenditures in the Netherlands at the municipal level. This chapter first elaborates on the used data sources for this research, methods for data preparation, and the justification of choices made in this preparation.

3.1. Data sources

As this research requires multiple types of data, multiple datasets have to be combined. On the one hand, travel data is required from a large number of people on a personal level. To be able to aggregate these data to municipal averages, these travel data should also contain the municipality of residence of each respondent. Additionally, to try to explain the travel behaviour by personal characteristics, additional background information on the respondents (e.g. socio-demographic characteristics) is required. These data requirements are satisfied by the ODiN data, provided by CBS (n.d.-a).

Besides these individual travel data, additional information on the Dutch municipalities is also required. For one, characteristics per municipality, like urbanism and population density, amongst other things, are needed for efforts to explain travel behaviour by spatial characteristics. Geographical data on municipalities is also required, so data can be assigned to specific municipalities and adjacency of regions can be used to allow for the identification of spatial patterns. These data requirements are satisfied by the geometric data files from CBS (CBS, 2021).

3.1.1. ODiN (travel data)

Every year, CBS has been collecting data on people's travel behaviour using widespread surveys of the Dutch population. Since 2018, this survey programme runs under the name 'Onderweg in Nederland', or ODiN for short. The survey is conducted through an online questionnaire and the target population is the *'The Dutch population aged 6 years and over living in private households, excluding residents of institutions, establishments and homes'* (CBS, n.d.-a).

Extensive research descriptions for each year of the ODiN project are publicly available, which tell us more about the data collection and processing and present extensively what data is included (CBS, 2019, 2020b). Respondents are selected using a stratified sampling technique. They are asked to record all their travel behaviour during a specific day of the week that they have been assigned. This includes all movements, for any reason, at any location, and with any mode. The data contains travel information per trip, including but not limited to: mode of transportation, start- and arrival times and locations, and purpose of the trip. This way, a full and exact picture of one's travel behaviour on the given day is recorded.

Through the survey additional information on the respondents is collected as well. This concerns information on car and E-bike ownership, frequency of use of travel modes, education, and societal position.

The respondents are also asked to check the collected personal information already available to CBS through other sources, like their socio-demographic background characteristics. All these factors are included in the final dataset as well. The data is also enriched with data from other government registries, to include more geographical and personal information. Besides the first general nation-wide data collection, additional data is collected in regions that might otherwise suffer from data insufficiency. Because of this, the data provides very broad and granular coverage and can also be used for regional analyses. The data is already pre-cleaned by CBS to filter out faulty data or trips that are not considered 'regular'/normal travel behaviour. Most importantly, with the focus on travel on Dutch soil in mind, this means that any trips made fully abroad are excluded from the data. Trips by aeroplane are also excluded from the data.

In this research, ODiN data from both 2018 and 2019 is used. With the 2018 dataset containing 57.260 respondents and the 2019 dataset containing 53.380 respondents, this supplies us with an original full dataset, with lots of background information and a day's worth of detailed travel behaviour, of 110.640 respondents.

3.1.2. Geometric data

The geometric data used in this study is also provided by CBS. Each year, extensive population statistics and spatial characteristics on all municipalities and neighbourhoods are published. These data are published together with geometric data per region, which allows for spatial mapping of characteristics. The geometrics in this data originate from the land registry (CBS, 2021). The data is available at three levels of detail: municipality, 'wijk' (part of a municipality), and 'buurt' (part of a 'wijk'). This research exclusively makes use of the municipal level data.

The geometries of the municipalities are formatted as ESRI Shape formats and contain the exact areas and boundaries of all Dutch municipalities. Because of this, they can be used to visually project data onto a map of the Netherlands and allow for spatial analysis, as adjacency of regions can be derived from this data. Additionally the included statistics (or 'kerncijfers' in Dutch) provide an abundance of information on each specific region. What exact statistics are included in this dataset is extensively covered on the CBS website (CBS, 2021).

These maps and statistics are readily available to use for multiple years. As municipal boundaries can change over the years, due to merging or boundary shifts, it is important to use geometric data from the corresponding year, to allow for combining it with the ODiN data. Because of this, the geometric dataset (including the additional statistics that come with it) from 2019[7] is used, which divides the Netherlands into 355 municipalities. This data comes with 194 more statistics per municipality (besides the geometrics), including variables on the built environment, economic factors, demographics, and more.

3.2. Preprocessing ODiN data

The collected ODiN data is already filtered and preprocessed by CBS as part of the larger ODiN project. Elaborate information on this preprocessing and what data is included and excluded is documented in the ODiN research descriptions as well (CBS, 2019, 2020b, Ch.6). Though this already partly boosts the quality of the data, this preprocessing is not done for any specific use. Therefore, for the application of the data as intended in this research, additional preprocessing is performed.

3.2.1. Selection of attributes

The ODiN datasets are very large and not all included data is relevant for this study. The raw data from 2018 contains 200 attributes, while the raw data from 2019 contains 203 attributes[8]. A lot of these attributes regard (for example) very detailed trip-specific information, which is not deemed relevant for

[7]The municipal labels in 2018 ODiN data therefore need to be recoded to 2019 municipalities to account for any differences. This is further covered in section 3.2.3

[8]These numbers are not the same, as for 2019 additional attributes regarding metropolitan region Rotterdam The Hague were included (CBS, 2020c)

this research, as the focus is on total travel time expenditure only. For the purpose of further data preprocessing and analyses to be performed later, initially the variables presented in table 3.1 are selected from the raw data.

Table 3.1: Selected ODiN variables

Variable	Description
OPID	Unique id for every individual
ResPC	Postal code (pc4) of residence
ResMun	Municipality of residence
HHComp	Household composition
MunUrb	Urbanism class of municipality of residence
MunPop	Population class of municipality of residence
Sex	Sex
Age	Age
MigBG	Migration background
SocPart	Social participation
Edu	Highest level of education completed
IncGroup	Standardised spendable income of household (10% groups)
CarOwn	Number of cars in household
PecDay	Peculiarities on day of reporting
ReasonNT	Reason for having no recorded travel in NL
AbrTripRem	Indicator of a completely abroad trip having been removed
AbrTripTT	Travel time per (part of the) trip made abroad
TotTT	Total recorded travel time in NL of individual

These variables are selected for a variety of reasons and uses. ***OPID*** allows for identification of specific individual records. While this research does not focus primarily on individual behaviour, including this unique identifier allows us to later gather additional (raw) data on the travel behaviour of specific persons, which is of use mainly for zooming in on outliers (appendix A) and justifying their exclusion. The variable ***PecDay*** is also included for further use in outlier detection (table 3.7). The variable ***ReasonNT*** allows us to filter out any persons that were abroad on the day of recording, and the variable ***AbrTripRem*** allows us to filter out any persons that had trips removed from the records due to travel abroad (section 3.2.4).

The central variable of interest is the total travel time per person per day, which is contained partly in ***TotTT*** and partly in ***AbrTripTT***, as elaborated on in section 3.2.2. As this research will focus this (average) value at the municipal level, including the variable ***ResMun*** allows for aggregation of travel times (and other characteristics) by grouping individuals by their municipality of residence and averaging their travel times to come to a municipal mean (section 3.4). ***ResPC*** is also included, as postal codes are necessary for properly joining the 2018 and 2019 data (section 3.2.3).

Other variables (***HHComp, Sex, Age, MigBG, SocPart, Edu, IncGroup, CarOwn***) provide information on the socio-demographic background of the respondents and thus represent the personal characteristics from section 2.6. They have a foundation in the literature from chapter 2, and will later on be used to explore possible effects of these factors on travel time expenditures. For this, these ODiN variables will be aggregated to municipal measures (section 4.2.1). The variables ***MunUrb*** and ***MunPop***, which present urbanisation characteristics of the municipality of residence of the respondent, are also included for comparison purposes in the data preparation, like for the cross-year comparison in section 3.2.3. While included here, these variables are also included in the in the CBS geometric data (section 3.3.1).

3.2.2. Recalculating total travel times

Before the data is further explored and cleaned, actual total travel times per person need to be recalculated. As mentioned before, trips that fully took place abroad are not included in the ODiN datasets. Respondents for which this occurs are removed from the data (section 3.2.4). However, trips that took place partly in the Netherlands and partly abroad are recorded, though only the travel time over the Dutch part of the trip is included in the *TotTT* variable (CBS, 2019, 2020b, App.A). If the parts of the trips made abroad are not counted towards the total travel time that is studied in this research, these values could become deflated. Therefore, the total travel times need to be recalculated, to also include the part of the travel that was made outside of national borders, but still recorded.

The variable *AbrTripTT* contains the part of the travel time of a border-crossing trip that was spent abroad. Unlike the *TotTT* variable, which contains a daily total, the *AbrTripTT* variable contains only this value per trip separately. Therefore, in order to correct the total travel time that is being studied, for each person the *AbrTripTT* of all their trips is summed and added to the original *TotTT* value (formula 3.1), to include foreign travel in this newly calculated total travel time. This new value thus includes all of a person's trips that have at least partly taken place in the Netherlands, but also include the time spent travelling across the border during those trips.

$$TotTT = TotTT + \sum AbrTripTT \tag{3.1}$$

3.2.3. Comparing and joining 2018 and 2019 data

To obtain a sufficient number of respondents, also at the municipality level, this research includes two years of ODiN data, 2018 and 2019, and considers this one dataset. To justify that this joining of multi-year data makes sense, the datasets are compared on the socio-demographic and built environment variables, and the travel times. If these variables show similar distributions for both years, they can be reasonably considered one single sample from the same population for the sake of this research.

The plots for comparison of the socio-demographic background variables are shown in figure 3.2. These plots show that for all variables, the relative distributions across categories are similar for both years. There are some minor differences in representation, but these differences are small, and in general we see the roughly the same relative frequencies in both years. When looking at the distribution of built environment characteristics (figure 3.1) a similar image appears. It can therefore be concluded that the samples of both years are drawn from very similar populations. This is also the the conclusion from the plausibility report by CBS (2020c).

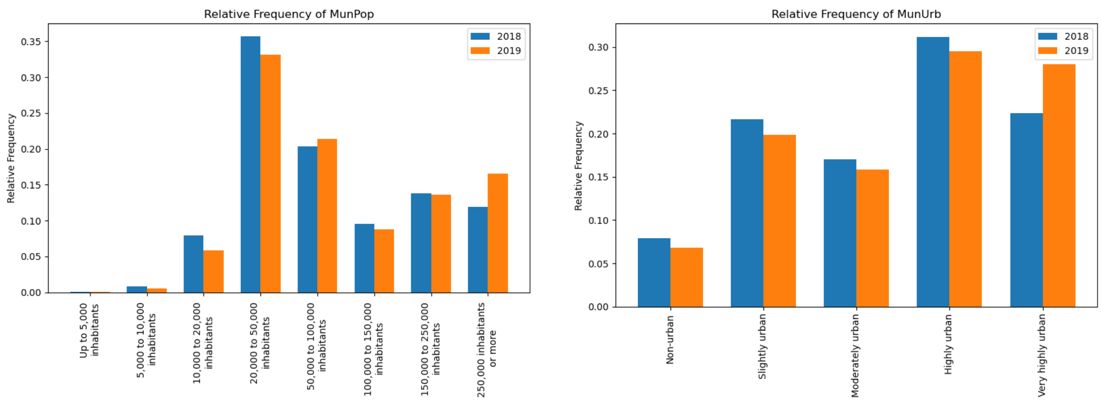


Figure 3.1: Relative distributions of built environment variables in ODiN 2018 and 2019 data

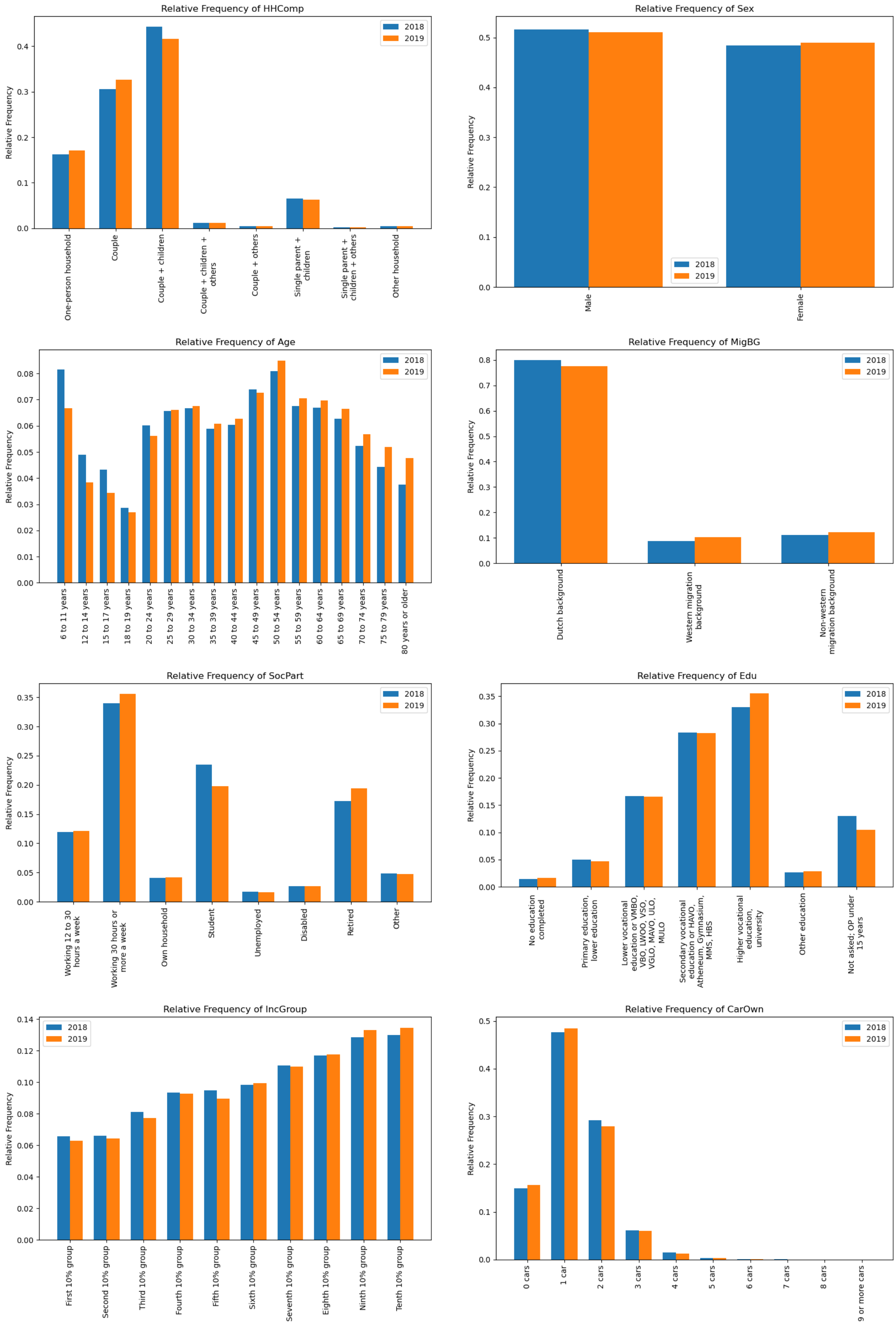


Figure 3.2: Relative distributions of socio-demographic variables in ODiN 2018 and 2019 data

Besides the similarity in socio-demographic and built environment characteristics of the data, the travel time values should also show a similar picture for both years. When comparing the *TotTT* over the two years, this indeed is the case. In figure 3.3 it is shown that this variable is also distributed roughly the same in both years, especially up to the value of 400 minutes, which is shown in the left plot. The right plot shows that more extreme values seem to appear slightly more often in 2018, though this difference can be marginalised, as it should be minded that the y-axis of this graph is of a much smaller scale. When determining the average value of total travel time per day in for this uncleaned data, the values are again very similar, the averages of both years showing only a 0,53% difference:

Average daily travel time expenditure	
2018:	78,59 minutes
2019:	78,18 minutes

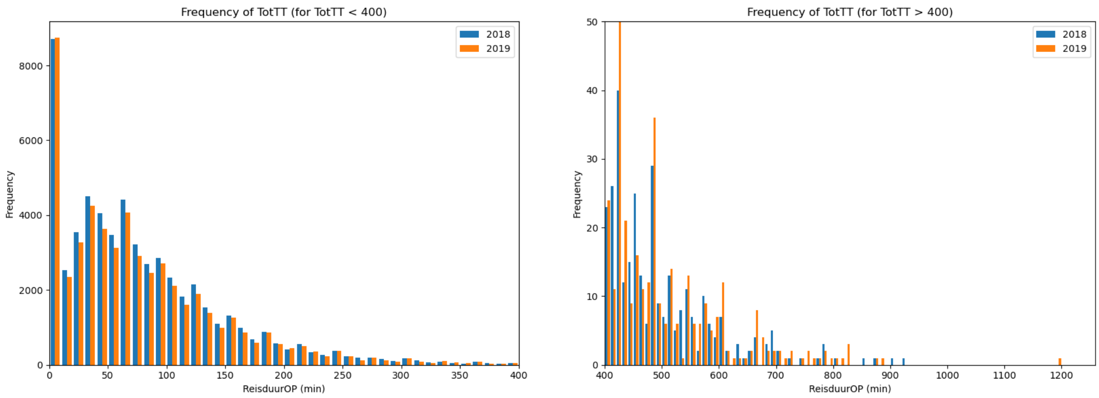


Figure 3.3: Distributions of daily travel times in ODiN 2018 and 2019 data (shown in two plots to enhance readability)

Now that has been shown that the 2018 and 2019 ODiN datasets are similar enough, they will be considered one dataset for further purposes in this research. As both datasets originate from the same source (CBS) and contain the same set of selected relevant variables (table 3.1), the data is fairly easily concatenatable. One issue presents itself though, regarding the municipality of residence (*ResMun*). As this study focuses on the municipal level, it is paramount that we are able to group values per municipality. However, the municipal division is not always the same for every year, as most years the number of municipalities decreases due to merging.

This is also the case when 2018 and 2019 specifically are considered. On January 1st, 2019, a redivision of municipal boundaries, including merging of municipalities, was implemented. The total number of municipalities decreased from 380 in 2018, to 355 in 2019 (CBS, n.d.-b). To incorporate these changes in the ODiN data, the *ResMun* values of the merged municipalities in the 2018 data need to be recoded to the 2019 municipalities they merged into, to ensure compatibility with the geometric data as well, which contains geometrics of the 2019 municipalities. Table 3.2 present the 2018 municipalities that have been merged and the 2019 municipalities they have been converted into. After this conversion has been performed, the two datasets can be joined, resulting in one large ODiN dataset of 110.640 respondents, which will serve as the base for further data preprocessing and analysis.

Table 3.2: Mapping of 2018 municipalities to 2019 municipalities

Municipalities 2018	Municipality 2019
Bedum, Eemsmond, De Marne, Winsum[9]	Het Hogeland
Ten Boer, Groningen, Haren	Groningen
Grootegast, Leek, Marum, Winsum[9], Zuidhorn	Westerkwartier
Dongeradeel, Kolummerland en Nieuwkruisland, Ferwerderadiel	Noardeast-Fryslân
Geldermalsen, Neerijnen, Lingewaal	West Betuwe
Haarlemmerliede en Spaarnwoude, Haarlemmermeer	Haarlemmermeer
Leerdam, Vianen, Zederik	Vijfheerenlanden
Noordwijk, Noordwijkerhout	Noordwijk
Oud-Beijerland, Binnenmaas, Korendijk, Cromstrijen, Strijen	Hoeksche Waard
Giessenlanden, Molenwaard	Molenlanden
Aalburg, Werkendam, Woudrichem	Altena
Onderbanken, Nuth, Schinnen	Beekdaelen

3.2.4. Data cleaning

The original data is already preprocessed by CBS and therefore of high quality. Still, for the purpose of this research some additional data cleaning is performed to make the data as suitable as possible for the intended uses in this study. Additional checks are performed to identify missing values, incorrect entries, and outliers.

Missing values

The CBS has already filtered out most incomplete or illogical entries, as described in the research descriptions (CBS, 2019, 2020b, Ch.6). In the published data no direct missing values (where there was no data entry at all) were found, as any missing values were already relabelled 'Unknown' (NL: 'Onbekend'). When checking for 'Unknown' values for the variables where they could possibly appear, some but little were found. The number of 'Unknown' values per checked variable are shown in table 3.3.

Table 3.3: Number of 'Unknown' values found in the ODiN data

Variable	# instances of 'Unknown'
HHComp	0
MigBG	5
SocPart	0
Edu	0
IncGroup	0
CarOwn	159
PecDay	917
ReasonNT	0

In a total of 110.640 individuals, the numbers of missing values presented in table 3.3 are very low. Additionally, missing values in these specific variables do not automatically disqualify these data entries from being useful for the research. In the case of *MigBG* and *CarOwn*, though these individuals cannot be used in for explanatory regression analyses later on, their travel data is still very much useful and relevant for the analysis of averaged-out travel times per municipality. In the case of *PecDay*, this variable is used as a supporting indicator in outlier detection, but yields no true value in the further analyses themselves, meaning these missing values are also not detrimental to the research. Therefore, these unknown data entries are (for now) not removed from the data.

[9]Winsum was split up, partly over Het Hogeland and partly over Westerkwartier. In the data this mapping is based on postal codes (*ResPC*)

Incorrect travel time values

When exploring the travel time expenditures per person, it became apparent that a considerable portion of the population recorded no travel behaviour at all (*TotTT* = 0). This is clearly depicted in figure 3.3, with the far left bar of the histogram being by far the largest. In total over both years, 16.369 respondents are recorded to have not travelled at all. The variable *PecDay* has a 'Does not apply' (NL: 'Niet van toepassing') value as one of its possible entries, which should occur in case an individual did not do any travelling on the day on record. What this implies, is that for every individual for which *TotTT* is recorded to be 0, *PecDay* should show to take this 'Does not apply' value. A check on this was performed, from which it was shown that for 20 individuals this was not the case. As this implies a false 0-value for one's travel time, the records of these 20 individuals were removed from the data, leaving 110.620 respondents.

Additionally, the variable *ReasonNT* holds record of the reason why a person did not record any travel (in the Netherlands) on the day of recording. One of the possible reasons is that the person was abroad on this day. Though this means that this person indeed did not travel in the Netherlands, this does not necessarily mean that this person did not travel at all on this day. As their travel patterns abroad are not recorded and we thus do not know what travel behaviour they truly exhibited on this day, these individuals are also filtered out of the data. There are 1.355 instances of this and removing these leaves 109.265 respondents.

Lastly, some individuals may have travelled abroad during the day and also made a trip that took place completely outside of the Dutch country borders. These trips fully abroad have been omitted from the data and are not counted towards the recorded total travel time for the day, not even after the recalculation in formula 3.1. Therefore, the records of persons for which this is the case have a deflated total travel time, as part of their travel on the day is not included. The variable *AbrTripRem* allows us to filter out these individuals. It is important to understand clearly when this exclusion rule applies. The travel time of trips that have at least partly taken place in the Netherlands are still recorded in full, so the left situation in figure 3.4 does not cause any issues and is thus not excluded. In the right situation, trip 2 has taken place fully abroad and therefore the travel time of this trip is not recorded. Instances where this occurs are excluded, as we cannot possibly derive the total travel time expenditure, not knowing anything about trip 2. This operation ensures at least for all entries that we do keep, the total recorded travel time expenditure accurately reflects their full travel pattern on that day. There are a further 180 individuals for which this trip abroad issue arises. Removing them from the data leaves 109.085 respondents. In total, 1.555 incorrect entries were removed in this process.

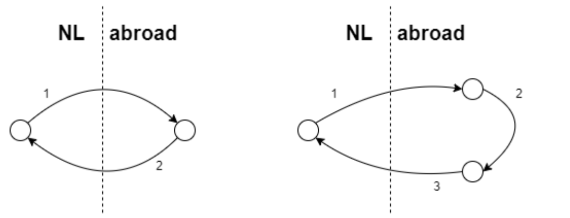


Figure 3.4: An illustration of when trips abroad are excluded. For the left situation, the entry is still included. For the right situation, the entry is excluded, due to the unrecorded details of trip 2

Outliers

As this research focuses on the municipal level and data will later be aggregated to municipal averages, it is important that the included data are an accurate representation of the travel time expenditures per municipality specifically. Therefore, outlier detection on *TotTT* values is performed on the municipality level. By looking for outliers based on these geographic subsets of the data, it is more accurately assessed whether an individual has recorded extraordinarily high or low travel time expenditures, taking into account the distribution of travel times in the municipality they are from. The outlier detection is performed using the Inter-Quartile Range[10] (IQR) method (formula 3.2), which calculates the thresholds for outliers as follows:

[10] Because the values of *TotTT* are not normally distributed (figure 3.3), the alternative z-score method for outlier detection is not as suitable.

$$
\begin{aligned}
&Q1 = first\ quartile \\
&Q3 = third\ quartile \\
&IQR = Q3 - Q1 \\
&Non - outlier\ range = [Q1 - 1,5 * IQR, Q3 + 1,5 * IQR]
\end{aligned} \tag{3.2}
$$

Based on the data of each municipality separately, the Inter-Quartile Range measures of *TotTT* values can be calculated and used to identify outliers. All values that fall outside of the $Non - outlier\ range$ of its municipality are considered statistical outliers. This outlier scan is applied to al municipalities that contain a sufficient sample size to base trustworthy statistics on, for which a threshold of 30 respondents is used. Five municipalities[11] do not meet this threshold, but applying this outlier detection to the other 350 sufficiently large municipalities yielded a total of 4.264 statistical outliers at municipality level, which is 3,9% of the data, though this percentage differs per municipality (figure 3.5). An example of the visualisation of these IQRs is shown in figure 3.6, which shows the boxplot of *TotTT* in the municipality of Aa en Hunze. It is interesting to note these outliers only contain values on the higher end of the spectrum, as due to the lower bound of 0 for travel times and the distribution of the data, outliers below the $Non - outlier\ range$ do not exist in this data.

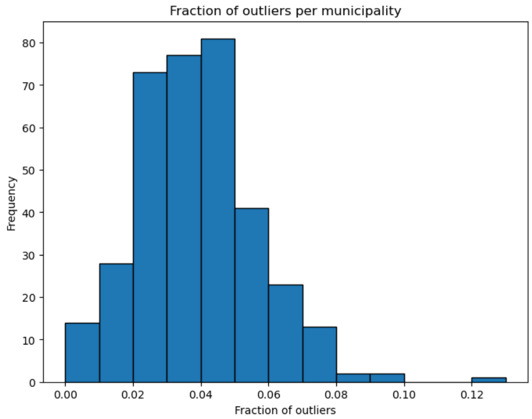


Figure 3.5: The fraction of outliers per municipality

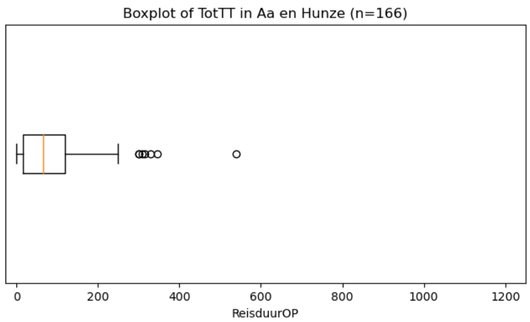


Figure 3.6: Boxplot of *TotTT* in Aa en Hunze. Values outside of the whiskers of the boxplot are considered statistical outliers

Now, these 4.264 values are statistical outliers, meaning that their extreme values alone set them apart from the rest of the data in the distribution. However, this does not automatically imply that these outliers contain false data. To determine how these outliers can best be handled, they are examined more rigorously. For this, firstly the variable *PecDay* is looked at. In this variable, respondents could flag whether the travel behaviour they showed on the day of record contained any peculiarities or unusual movements. The frequencies of the values of this variable in both the full ODiN data, as well as only the outliers, are presented in table 3.4. Figure 3.7 more clearly shows the respective relative frequencies in pie charts.

When comparing the outliers to the full dataset, the distribution of values for *PecDay* is clearly very different. In the full data, 44,3% state no peculiarities, which would imply a day like any other was recorded and the travel records could be considered representative of one's normal patterns. For the outliers, this concerns only 25,9%. If the value of *PecDay* is 'Yes', the record is not considered representative, and it can be seen that this portion of the data is almost twice as big for the outliers (38,6%) as it is for the full data (20,1%). Additionally, when a person has stated that the recorded day is different each week, it is inherently harder to determine whether this found outlier is an extreme value or not. Therefore, when taking the values 'Yes', 'No, this day of the week is always different', and

[11] Terschelling, Ameland, Rozendaal, Schiermonnikoog, Vlieland

'Unknown' together, we see that the share of observations of which the representativity is uncertain at best, is way higher in the outliers (74,1%) than it is in the full dataset (41,9%).

Table 3.4: Frequencies of *PecDay* values in the full data and outliers

PecDay value	Frequency in ODiN data	Frequency in outliers
No, no specific peculiarities	48289	1103
No, this day of the week is always different	22961	1495
Yes	21942	1646
Does not apply; person did not travel	14994	0
Unknown	899	20

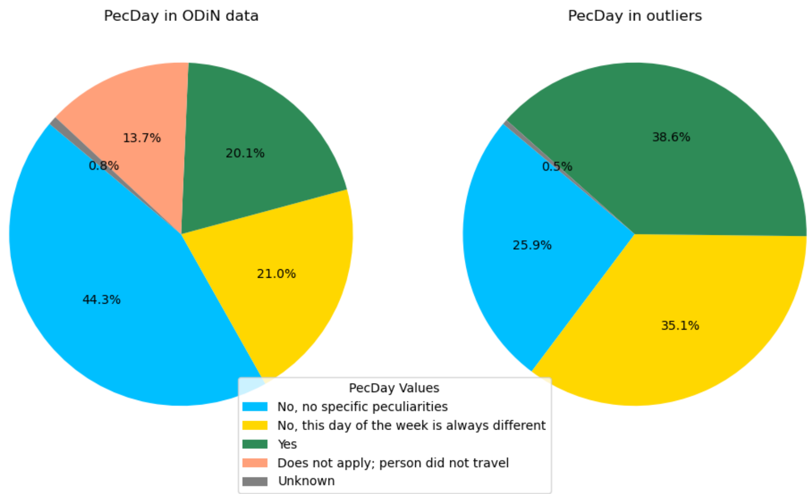


Figure 3.7: Pie charts of *PecDay* values in the full data and outliers

The representativity of travel times is thus way more ambiguous and less strongly supported for the outliers, compared to the full ODiN data, based on this *PecDay* variable. However, as more than a quarter of the outliers is still flagged as representative, we do not just want to throw out the baby with the bathwater, by excluding all outliers from this research. To find out if there are any specific groups that might be more likely to be part of the outliers, outlier characteristics are more extensively examined. However, when comparing the distributions of socio-demographic background variables and the built environment between the outliers and the full ODiN data (appendix A), no clear differences draw the eye. In the outliers men seem to be slightly overrepresented and children somewhat underrepresented, but in general the variables follow very similar distribution to the full dataset. Therefore we cannot exclude specific subgroups from the data based on their outlier representativity alone.

In a final effort to gain some meaningful insight into what type of outliers we are dealing with, some outliers are randomly selected and for a full examination of their raw ODiN data. As it is not feasible to case-by-case assess every single one of the 4.264 outliers this way, 40 outliers making up four subsets are selected for this deep-dive:

(1) The **10 most extreme** outliers with the **highest *TotTT* value**.

(2) The **10 least extreme** outliers with the **lowest *TotTT* values**.

(3) **10 randomly selected** outliers from the full set of outliers.

(4) **10 randomly selected** outliers that were **flagged as 'No, no specific peculiarities'**.

When looking extensively at the full travel records of the 10 most extreme outliers, it becomes clear that most of these extreme values for *TotTT* are the result of incorrect records, even though some of these records were also flagged as having no peculiarities. These outliers tend to show trip records that are very unrealistic and hard to rhyme with common sense (e.g. very long travel times for very short distances). Contrarily, most of the least extreme outliers do present travel behaviour that seems logical and therefore valid. When considering the 10 randomly selected outliers from the full set, a mixture of both more and less realistic records is found, where most questionable records were also already flagged in the *PecDay* variable as showing peculiarities. Finally, when examining the last sample of outliers (that were flagged as not peculiar), most records seemed realistic, though also here some records were found that are difficult to consider realistic. A more elaborate coverage of this deep-dive can be found in appendix A.

This extensive outlier examination leads to a double-sided conclusion. On the one hand, it is clear that not all outliers are necessarily false data points, as some statistical outliers do present travel behaviour that could still be considered realistic. On the other hand however, even some outliers that are flagged as having no peculiarities have been shown to contain data that is most likely false, thus eating away at the reliability of this (supposedly representative) subset of outliers, and therefore the reliability of the outliers as a whole. The only way to accurately separate the clearly false from clearly true data would be a case-by-case assessment of each outlier. However, as this is simply not feasible given the large number of outliers, it is decided to exclude all 4.264 statistical outliers from the analysis, leaving 104.821 respondents. This decision is supported by the following most important considerations:

- **The data quality of the outliers is too uncertain.** For almost three quarters of the outliers, their representativity is uncertain at best. Even in the rest of the data, some clear unrealistic values were found. The outliers are thus about 25% reliable, but possibly even less, and it is not possible to easily filter out the truly incorrect entries. Deleting all outliers thus comes at a lower data quality cost than keeping all outliers.
- **The outliers do not fairly represent the travel times of that municipality.** Though this is mainly a statistical argument, excluding these values that clearly have such a different travel time from others in the same municipality, will make sure that the aggregated average values more accurately represent the larger portion that is more similar in their travel behaviour.
- **The exclusion methods is applied to all municipalities similarly.** This method of outlier exclusion will deflate the average travel times per municipality towards a more representative value for the largest portion of its inhabitants. This will be the case for all municipalities. As this research mainly focuses on relative differences in travel times, it does not necessarily pose a problem that absolute values are being deflated, as long as relative differences are conserved.

When looking at the impact of removing these outliers on the average value of travel time expenditures per municipality, it can be seen that the impact does differ quite a bit still between municipalities (figure 3.8). While these may seem like large differences at first, this can be easily explained. Logically, larger municipalities with broader data coverage naturally are more robust against value changes. The municipalities with the most extreme drops tend to be smaller municipalities. Though a usual drop of 10% to 15% initially also feels large, this is also logical once you realise that removing one value of e.g. 500 minutes from a sample of 50 values, the average value will immediately decrease by roughly 10 minutes. Given that the data-wide average travel time expenditure for the uncleaned data was about 78 minutes, this would already be a decrease of over 12%.

For the too small municipalities that were not included in this round of outlier detection, additional detection is run after their merging with other regions, as described in section 3.3.2. With three more outliers being found there, the final cleaned ODiN data contains 104.818 respondents. With 1555 incorrect values and a total of 4267 outliers, 5822 entries (5,26%) were removed from the ODiN data.

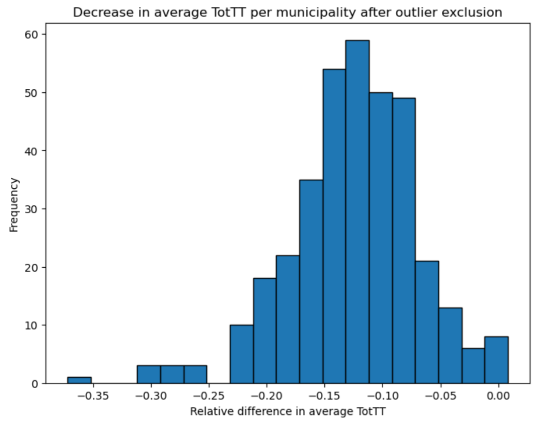


Figure 3.8: The impact of removing the outliers from the data on municipal averages of travel times

3.3. Preprocessing geometric data

The geometric data provided by CBS can be used to link attribute values to specific geographical regions. This enables the performance of spatial analysis of travel behaviour, but the geometric data also comes with additional regional data as well. To include the relevant additional statistics, as well as to make the geometric data compatible with the ODiN data, the geometric data also undergoes some additional data preparation.

3.3.1. Selection of attributes

The raw municipal geometric data, besides identifiers and the geometric shapes themselves, comes with 193 more attributes per municipality. Though abundant, not all attributes are useful or relevant for this study. Therefore, only a subset of these variables is included in further data preparation and analysis. The variables that are part of this subset are presented in table 3.5.

Table 3.5: Selected variables from the CBS geometric data

Variable	Description
AddDens	Address density
UrbClass	Urbanism class
PopDens	Population density
Pop	Population
NumBusi	Number of businesses
LabPart	Net labour participation
HouVal	Average house value (x1000EU)
Area	Land area
*Dist_*facility**	Average distance to *facility* (included for 32 facilities)
Geometry	Specification of the geometric shape

Most of these variables are selected because they provide background information on built environment attributes of the municipalities and thus represent the spatial characteristics from section 2.6. The literature presented in chapter 2 showed that levels of urbanism could possibly impact travel time expenditures, therefore these variables are deemed relevant for this research. This concerns the variables ***AddDens, UrbClass, PopDens***, and ***Pop***, that can serve as measures for urban density or city/municipality size. ***Area*** is also included, to serve as a weight factor for value aggregation across municipalities (appendix B). Additionally, ***LabPart*** and ***NumBusi*** can represent (un)employment, which was also mentioned as a possible impacting factor in the literature, and economic activity in general. Lastly, not necessarily based on the literature, ***HouVal*** is included, to act act as an indicator for welfare, and the 32 different ***Dist_*facility**** variables are included to later on be used to calculate distance measures to certain facility groups for the municipalities.

Lastly, and most importantly, the ***Geometry*** column contains the geometric shape of each municipality, coded as a polygon based on coordinates. This data can be used to geographically project municipalities onto a map. It can therefore also be used to determine adjacency of regions (section 4.1.1), which is critical for the to be performed spatial analysis. This geometric data thus is part of the foundation on which the spatial analyses in this research are built.

With regards to data cleaning, little more has to be done beforehand. The data is very complete and accurate, as for the built environment variables they consist of data based on official government records. Only the *HouVal* variable contains three missing values, but like in the ODiN data, this does not automatically render all data of these municipalities unusable. These entries will have to be handled when performing regressions on *HouVal*, but for the general spatial analysis missing values in background data do not impact the results, as long as the geometric data is complete and accurate.

3.3.2. Merging small municipalities

The main variable of interest in this study is the average travel time expenditure per municipality. To be able to reliably calculate this metric, we desire a sufficient number of respondents to base this average on. To assess this, we yield a threshold of 30 respondents per municipality. If a municipality has less than 30 respondents, it is considered to suffer from data insufficiency, on which one can not base aggregated statistics with enough confidence. Though joining two years of ODiN data already partly mitigates this issue, when grouping the cleaned ODiN data by municipality, six municipalities still do not contain the minimum required number of respondents:

Too small municipalities	
Baarle-Nassau:	29 respondents
Terschelling:	29 respondents
Ameland:	14 respondents
Rozendaal:	13 respondents
Schiermonnikoog:	5 respondents
Vlieland:	2 respondents

Based on the data of these municipalities alone, we can thus not reliably enough determine an average travel time. In order to solve this issue, these small municipalities are merged with neighbouring municipalities to create larger joint municipalities with sufficient data coverage. While doing so, attribute values will be combined through (weighted) aggregation, to ensure the statistics of the newly constructed regions describe the total area they encompass as accurately as possible. To prevent municipalities are being merged that are too dissimilar from each other, which would pollute their data, possible candidates for merging are first compared on their characteristics. While some differences in distributions of variables were found between to be merged municipalities, given the small sample sizes of there municipalities these were not considered critical. The full comparisons can be found in appendix B. After these comparisons, it was decided to perform the merges presented in table 3.6.

Table 3.6: Municipalities that are being merged

Municipality	**Merge with**	**New area (respondents)**
Baarle-Nassau	Alphen-Chaam	Alphen-Chaam + Baarle-Nassau (88)
Rozendaal	Brummen	Brummen + Rozendaal (105)
Terschelling, Ameland, Schiermonnikoog, Vlieland	*each other*	Friese Waddeneilanden (50)

Concretely, the application of this merging consists of three parts:

- In the ODiN data, for the respondents in the merging municipalities, the *ResMun* variable is changed to the newly formed region.
- In the geometric data, the shapes from the *Geometry* column are joined to form one area.

- In the geometric data, the other variables are (weightedly) aggregated to best present the newly formed region. A specification of this is included in appendix B.

Now, as mentioned in section 3.2.4, the at that time too small regions were not included in the outlier detection. Now that the issue of data insufficiency in these municipalities is resolved, the outlier detection and exclusion can be finished by applying the described IQR method from section 3.2.4 to the region Rozendaal + Brummen and the region Friese Waddeneilanden[12]. This resulted in three more detected outliers, bringing the size of the final cleaned ODiN dataset to 104.818 respondents.

3.4. Calculating the municipal average travel time

Now that sufficient data coverage for each municipality is assured by the described data preprocessing, the final variable of interest can be constructed. This variable (*AvTTE*) represents the average travel time expenditure per day per person, for each municipality. This measure is calculated by grouping all the 104.818 remaining observations by municipality, and taking the mean *TotTT* for all individuals in each group, as per formula 3.3

$$AvTTE_i = \frac{1}{n_i} \sum_h w_{ih} TotTT_h \tag{3.3}$$

$AvTTE_i$ *= average travel time expenditure per person in municipality i*; n_i *= number of persons in municipality i*; w_{ih} *= weight indicating if person h is from municipality i*; $TotTT_h$ *= travel time expenditure of person h*

[12]Baarle-Nassau had enough respondents to be included in the first round of outlier detection, but dropped under the threshold of n=30 due to the exclusion of outliers. Therefore, outlier detection is not run again for the joined region of Alphen-Chaam + Baarle-Nassau

4
Data Analysis Methods

In this chapter, we describe and elaborate on the the analyses that have been performed in this study. Firstly, to determine whether travel time expenditures (*AvTTE*) do show spatial coherence across municipalities in the Netherlands, analyses focussing on assessing (global and local) spatial autocorrelation have been performed. Subsequently, in attempts to explain the found spatial patterns, additional regressions are performed, incorporating built environment characteristics, as well as the socio-demographic background of the respondents.

4.1. Spatial autocorrelation

As described in section 2.5, spatial data analysis can be used as a self-contained method to identify interesting patterns in data that have to do with their spatial characteristics. Spatial patterns, like clustering of values or the (in)equitable distributions of values can be discovered by focussing specifically on what values appear where in a given (geographical) space. As a first step of data analysis in this research, the information on average travel times per municipalities undergoes such a spatial analysis, with the aim of identifying regions of interest, for example which municipalities and/or larger regions show excessively high or low travel times. In doing this, we address relationship 1 from the conceptual model in figure 2.6.

The above is accomplished by assessing the spatial autocorrelation of average travel time values in the data, on a spatial municipal level. Spatial autocorrelation essentially assesses how evenly values are spread, thus assessing how much equal values tend to appear next to/close to each other. In this research, the Moran's I will be used to measure the degree of spatial autocorrelation of travel times. The Moran's I exists in both a global and a local version[13]. The global Moran's I represents the spatial autocorrelation of the whole dataset in a single measure, assessing whether the full data as a whole shows any clustering. The local Moran's I represents a local measure for each municipality, assessing for each municipality whether it shows correlation with the values of its surroundings. The in this chapter presented methods for deriving these measures of spatial coherence are heavily bases on chapter 6 and 7 from the book by Rey et al. (2023).

4.1.1. Spatial weights

Firstly, in order to perform any form of spatial analysis, the underlying spatial/geometric structure needs to be accurately established. Though the data is already linked to specific municipalities, it is important to determine how these municipalities exactly spatially relate to each other. To do this, spatial weights are constructed based on the municipality's geographies, to establish which municipalities we can consider to be neighbouring each other for further analysis.

[13]Local Measures of Spatial Autocorrelation are sometimes also referred to using the acronym LISA

Rey et al. (2023, Ch.4) describe different methods of assigning spatial weights to geographical data, of which most fall into two most prominent categories: contiguity weights, and distance based weights. Contiguity weights are based on the concept of adjacency, where geographical units that share a border (or sometimes vertex) are considered neighbours. Alternatively, distance based weights assign weights based on how far apart (the centre of) two units are. In this research, any clustering of high-/low travel time values could appear due to spillover effects (Condeço-Melhorado et al., 2014): if some feature of region A results in high travel times, adjacent region B might also have higher travel times, as individuals from region B might also regularly travel through region A. Therefore, we want two municipalities to be considered neighbours when one can directly travel from one municipality into another.

Neighbours thus should be adjacent. Adjacency itself however can also mean multiple things. The two most used adjacency-based weight types are Queen and Rook (figure 4.1). Using Queen weights, areas that share a vertex are considered neighbours. Using Rook weights, areas have to share an edge to be considered neighbours. As we want a neighbour to imply a municipality one can directly travel into from a certain municipality, in this research Rook contiguity weights are applied to the data. In doing this, for simplicity's sake we assume that travelling directly across a vertex is not possible. Also, there are only four sets of non-Rook Queen neighbours[14], so this is not expected to have a big impact.

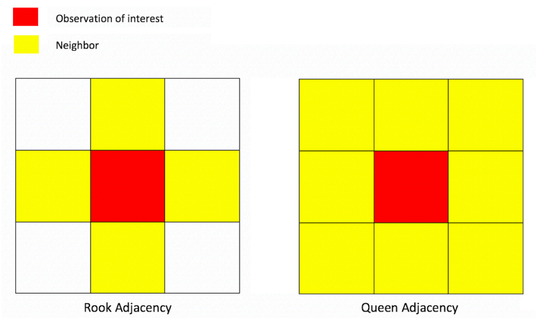


Figure 4.1: The difference between Queen and Rook adjacency (Brazil, 2023)

These spatial weights were assigned in a binary manner, meaning that two regions are either flagged as neighbours (weight = 1) or not neighbours (weight = 0). This means that in further analysis, each neighbour is considered to be of equal importance when calculating spatial coherence. Given the explanation of the choice for contiguity weights described in the previous paragraphs, all neighbours are considered to be equally important as one can easily cross a municipal border into any neighbouring municipality, independent of their size, population, or other factors.

When assigning these spatial weights, particular attention is paid to the the Wadden Islands in the north of the country. Even though these municipalities do not directly border other municipalities on land, to still include them in the analysis, they are considered to be neighbouring the municipalities that they are directly connected to through ferry services (WadgidsenWeb, 2024). This results in the addition of the neighbouring pairs in table 4.1. In the end, the weights are row-standardised per municipality, to ensure the weights of all neighbours of a municipality sum to 1, which is important for further calculations.

Table 4.1: Neighbouring municipalities of Wadden Islands municipalities

Municipality	Neighbours
Texel	Den Helder, Friese Waddeneilanden
Friese Waddeneilanden	Texel, Harlingen, Noardeast-Fryslân, Het Hogeland

4.1.2. Global spatial autocorrelation

With the spatial structure of the data clearly established, one can look for spatial patterns in the data. First, the data is analysed globally, which in this case means nation-wide. In doing this, we try to assess whether the data as a whole exhibits any spatial coherence, meaning that similar value are more (or for that matter, less) likely to be near each other. Figure 4.2 neatly shows what exactly this entails. If the data contained total spatial randomness (no spatial autocorrelation), we would not see

[14] Blaricum & Hilversum, Huizen & Laren, Krimpen aan den IJssel & Ridderkerk, Krimpenerwaard & Rotterdam

any patterns in values at all. If similar values are more likely to flock together, we speak of positive spatial autocorrelation. Negative spatial autocorrelation however would imply similar values being as evenly spread as possible across space.

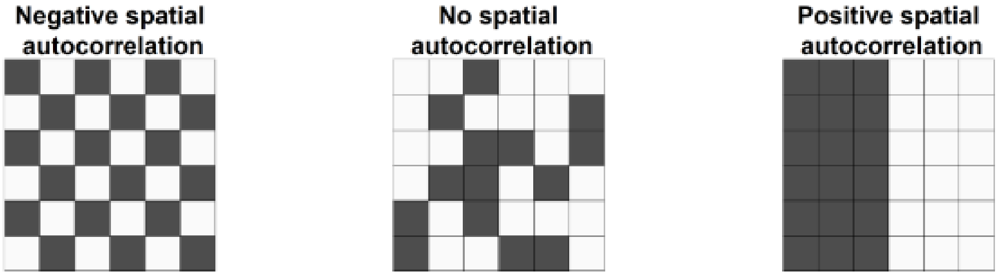


Figure 4.2: Examples of configurations with different types of spatial autocorrelation (Moraga, 2024, Ch.8)

Now, this research aims to address whether there is spatial autocorrelation in average personal travel time expenditure per municipality, across the Netherlands. In order to quantify the spatial autocorrelation, this analysis works towards calculating the global Moran's I, using the method lined out in the aforementioned book by Rey et al. (2023, Ch.7). Calculating the Moran's I (and its corresponding significance) consists of three main steps:

1. Calculating the spatial lag
2. Creating the Moran Plot and extracting the global Moran's I
3. Determining the significance of the found metric, by rerunning the analysis for multiple permutations

Calculating spatial lag

The first step in determining the Moran's I is calculating the spatial lag of each observation. The spatial lag represents the average value of all neighbouring observations, mathematically defined as formula 4.1. The spatial lag provides a measure for the values in an observation's surroundings. A high spatial lag here indicates a surrounding of high travel times, and a low spatial lag indicates a surrounding of low travel times.

$$y_{sl-i} = \sum_j w_{ij} y_j \tag{4.1}$$

y_{sl-i} *= spatial lag of observation i*; w_{ij} *= (row-standardised) weight of neighbour j for observation i*; y_j *= value of observation j*

Moran Plot and Moran's I

With the spatial lag of each observation having been calculated, a Moran Plot of the average travel times can be constructed. The Moran Plot is a scatterplot that shows the value of a variable against its spatial lag, for each municipality. It thus visualises how the variable of interest per municipality relates to the average values in its surroundings. Figure 4.3 shows example of such a plot.

To this scatterplot, a linear fit can be applied, which is shown in red in figure 4.3. The slope of this linear fit is equal to the value of the global Moran's I, and thus directly functions as a measure of global spatial autocorrelation. It is important to realise this measure does not consider absolute differences in travel times, but merely assesses whether relatively high and low values are more likely to be close to each other. It is thus purely a measure of coherence, not of (absolute) differences. The Moran's I can range from 1 (positive spatial autocorrelation) to -1 (negative spatial autocorrelation), with a value of 0 indicating no spatial autocorrelation at all. Mathematically, the global Moran's I is expressed as in formula 4.2.

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \tag{4.2}$$

I = global Moran's I; w_{ij} = (row-standardised) weight of neighbour j for observation i; z_i = standardised value of observation i; z_j = standardised value of neighbour j

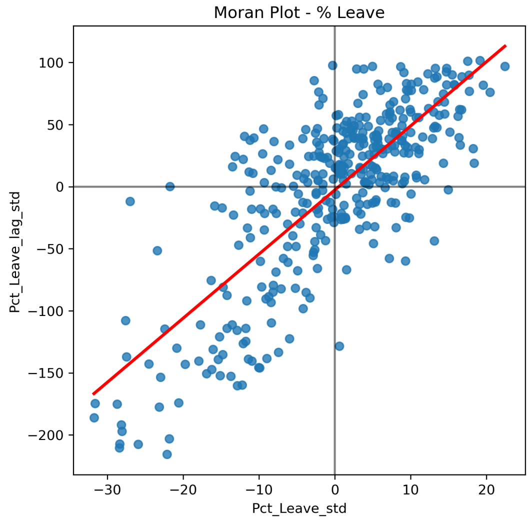


Figure 4.3: Example of a Moran Plot, with standardised values on the horizontal axis and standardised spatial lag on the vertical axis (Rey et al., 2023, Ch.6)

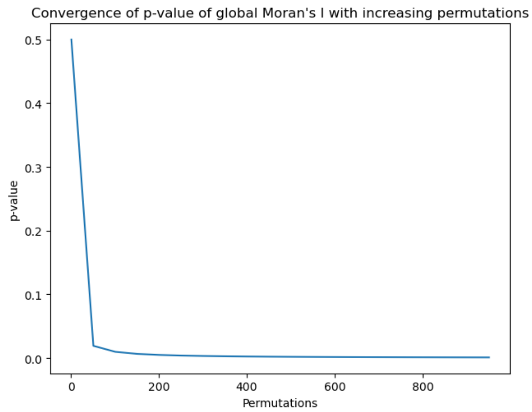


Figure 4.4: Convergence of the significance of the global Moran's I

Determining significance

From the Moran Plot plot and/or formula 4.2 it can thus be determined whether the average daily travel times per municipality show any spatial coherence. However, this measure in and of itself does not contain any information on the significance of the found spatial autocorrelation. To determine whether any found spatial patterns are truly different from what a random process would yield, significance of the Moran's I is based on simulating many more permutations of the same data.

More specifically, this is done by taking the same set of values for average daily travel times and randomly assigning them to municipalities for each permutation. This way, many more random maps are generated, using the same set of values that the originally found Moran's I is based on. For each permutation, the Moran's I of its randomly created map is calculated as well. The significance of the original Moran's I is based on how many permutations yield a random map with a more extreme Moran's I value than the original distribution. If the measure is deemed significant, it can thus be concluded that the distribution of values across space is unlikely to come from a random process.

To ensure that enough permutations are simulated for a trustworthy simulation-based p-value, the number of permutations is incrementally increased to test for the convergence of the p-value. As figure 4.4 shows, the p-value quickly converges after a few hundred permutations. For determining the final p-value of the found Moran's I, 999 permutations were run.

4.1.3. Local spatial autocorrelation

Now, while the global Moran's I is able to indicate whether the data as whole shows any spatial coherence, it does not tell us anything about where on the map certain values or patterns appear. For that, we need a local indicators of spatial autocorrelation (LISA). A local version of the Moran's I exists, which allows for assessing degrees of spatial autocorrelation for each observation in the data specifically. Once again, the book by Rey et al. (2023, Ch.7) lines out how to go about this. Identifying the type of spatial patterns found and calculating the local Moran's I (and its corresponding significance) also consists of multiple steps:

1. Classify all observations into spatial pattern types (quadrants)
2. Calculating the local Moran's I for all observation
3. Determining the significance of the metric for each observation, by rerunning the analysis for multiple permutations

Classifying observations into spatial pattern types

Firstly, to determine what type of spatial pattern each single observation shows, the plot from figure 4.3 is divided into four quadrants by the lines crossing 0 on both axes. This division results in four distinct groups of observations, as presented in table 4.2. This way, for each observation it is determined if their own values and spatial lag are relatively high or low compared to the rest of the data, and thus what type (positive or negative) of spatial autocorrelation they show.

Table 4.2: Quadrants in the Moran Plot

Quadrant	Relative values	Spatial autocorrelation	Label
Upper left	Low value, high spatial lag	Negative	LH
Upper right	High value, high spatial lag	Positive	HH
Bottom left	Low value, low spatial lag	Positive	LL
Bottom right	High value, low spatial lag	Negative	HL

Calculating local Moran's I

The classification of observations into quadrants gives a first impression of what kind of spatial patterns appear where in the data. However, this does not yet consider the strength of such patterns. To gain more insight into this, one must determine the local Moran's I for each observation, which is done using formula 4.3.

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j; \quad m_2 = \frac{\sum_i z_i^2}{n} \tag{4.3}$$

I_i = local Moran's I of observation i; w_{ij} = (row-standardised) weight of neighbour j for observation i; z_i = standardised value of observation i; z_j = standardised value of neighbour j; m_2 = variance of the data

As the formula shows, the calculation of the local Moran's I mainly consists of the multiplication of an observation's standardised value (z_i) and its standardised spatial lag ($\sum_j w_{ij} z_j$). Therefore, if an observation's value is similar to its spatial lag (both relatively high/positive or both relatively low/negative), the local Moran's I will be positive. If an observation's value is dissimilar to its spatial lag (high/positive value and low/negative spatial lag, or vice versa), the local Moran's I will be negative. Additionally, more extreme values in an observation's own value and/or its spatial lag will inflate its local Moran's I, becoming more extreme (more positive or more negative), and thus add to the strength of the spatial pattern.

A positive value of the local Moran's I thus implies positive local spatial autocorrelation, while a negative value implies negative local spatial autocorrelation, and its difference from 0 implies the relative strength of the autocorrelation. It is important to note here however, as table 4.2 shows, that based on only the local Moran's I itself, it cannot be concluded whether an observation and its surroundings shows high or low values. Positive spatial autocorrelation can point to both a HH cluster and a LL cluster, while negative spatial autocorrelation can point to both a HL and a LH cluster. In order to know both the direction of autocorrelation, as well as which type of spatial pattern a observation shows, both the local Moran's I and the assigned quadrant should be considered.

Determining significance

Though the combination of the quadrant classification and the local Moran's I contains information on the nature and strength of the local spatial coherence of the data, it does not yet contain any information on the significance of the found results. To test whether the local Moran's I values are different from what a random process would yield, once again simulating many more permutations of the data is used to determine the significance. Like when determining the significance of the global Moran's I, the same set of values is randomly assigned to municipalities for each permutation, creating many random maps. For each municipality and random map, the local Moran's I is calculated using formula 4.3. This creates a distribution of generated local statistics for each municipality, on which the significance of the true local Moran's I can be based.

To determine how many permutations should be run for reliable results, the number of permutations is again incrementally increased. As each run yields 350 distinct p-values (one for the statistic of each municipality), the number of significant municipalities (p-value < 0,05) is used to assess the convergence, rather than the p-values themselves. Figure 4.5 shows that after about 15.000 permutations, the results stabilise. For determining the final p-values and results of this local analysis, 24.999 permutation were run.

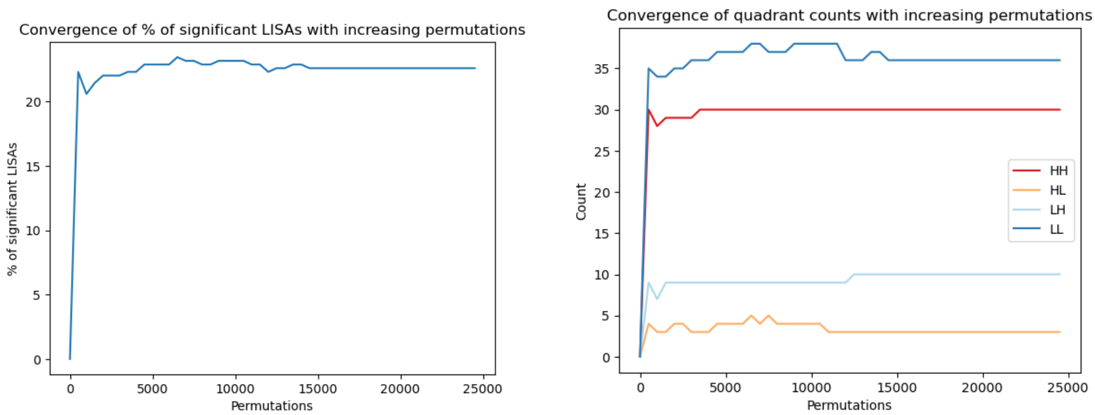


Figure 4.5: Convergence of the number of significant local Moran's I statistics

4.2. Regression analysis

After having performed the spatial analysis to identify the spatial patterns in travel time expenditures across the country, the second part of the data analysis in this research focuses on determining what background factors are important for explaining differences in these expenditures. To this end, travel time values are regressed on built environment and socio-demographic variables. The regressions in this research are, in consistency with the spatial autocorrelation analysis, run at the municipal level. The implementation of the various regression models as explained in this chapter, especially the ones considering spatial aspects, is heavily based on chapter 11 from the book by Rey et al. (2023).

4.2.1. Regression preparations

Before the municipal travel time averages can be regressed on the possible explanatory variables, two more data modification steps are performed: individual socio-demographic variables are aggregated to municipal values, and the distance measures in the data are combined to form fewer, more general measures.

Aggregating socio-demographic variables to municipal level

We do have individual-level data on the socio-demographic characteristics of the included respondents, but the regressions are to be performed at the municipal level. Therefore, some indicators at the municipal level are needed, that represents the socio-demographic characteristics of all of a municipality's inhabitants in a single measure per characteristic. One way to acquire these, is to use the statistics CBS publishes for each municipality, which also contains all of these measures (CBS, 2021). These socio-demographic statistics contain the true, real-life values/numbers of the full population, and are therefore very reliably representing these factors. However, as the dependent variable (the average travel time expenditure), is not based on this full population, but rather on a sample, a lack of representativity of this sample could still results in unreliable regression results.

An alternative way to derive the municipal-level socio-demographic indicators, is to aggregate them from the ODiN data that we use in this research. Using this sampled ODiN data as the base for the aggregated socio-demographic indicators provides us with less reliability regarding these aggregated measures themselves, given that they are based on a sample, instead of the whole population. However, they do more accurately represent the sample on which the average travel time expenditure is based as well. As they are thus more directly related to the base data for this variable of interest, we decide to use the ODiN data to base the aggregated socio-demographic variables on as well, for sake of the regression models. This might make the results a bit less generalisable, but will give more accurate regression results for at least the used samples.

For the sake of the spatial analysis in this research, the travel time expenditures were already aggregated to the municipal level by averaging the travel time expenditures over all persons per municipality. In order to also perform regression analyses at the municipal level, the socio-demographic ODiN variables thus also need to be aggregated to municipal measures. Where in aggregating the travel times, it being a numerical variable, the values could simply be averaged out, most socio-demographic variables are of a categorical nature. Therefore, different methods of aggregation need to be used. The newly constructed municipal measures are presented in table 4.3. The chosen methods of aggregation are explained here:

- **Sex.** The original variable is of binary categorical nature: male or female. This is converted to a percentage of males.
- **Age.** The original variable is of numerical nature. This is converted to an average value.
- **Migration background.** The original variable is of categorical nature: Dutch, western migration, or non-western migration. To see if people with any migration background do show differences will individuals with a fully Dutch background, this is converted to a percentage with (any) migration background.
- **Education.** The original variable is of categorical nature, with seven categories. This is converted to a percentage of highly educated individuals, as some literature indicates that high-education persons travel more (Feng et al., 2013).
- **Social participation.** The original variable is of categorical nature, with eight categories. This is converted to a percentage of people that have either a 30+ hours per week job or are a student. The reasoning behind this is that these people might have an obligation/commitment to travel to work/school every/most day(s), possibly causing more travel (Koo et al., 2022).
- **Household composition.** The original variable is of categorical nature, with eight categories. This is converted to a percentage of persons with children in their household, as some literature hints at lower travel times associated with (having) children (Raux, Ma, Joly, Kaufmann, et al., 2011).
- **Car ownership.** The original variable is of numerical nature, representing the number of cars one owns. This is converted to a percentage of people that do own a (at least one) car, as some literature hints towards an increase of travel time with car ownership (Wigan & Morris, 1981).
- **Income.** The original variable is of ordinal categorical nature, containing deciles of income. This is converted to an average income decile.

Table 4.3: Aggregation of socio-demographic variables to municipal metrics

Old variable	New variable	Description new variable
Sex	*Perc_Male*	Percentage in municipality that is male
Age	*Av_Age*	Average age in municipality[15]
MigBG	*Perc_ForeignBG*	Percentage in municipality with (western or non-western) migration background
Edu	*Perc_HighEdu*	Percentage in municipality with a higher professional education (HBO) or university degree
SocPart	*Perc_WorkStud*	Percentage in municipality that either (1) works >30 hours or (2) is a student
HHComp	*Perc_Children*	Percentage of persons in municipality with children in their household
CarOwn	*Perc_Car*	Percentage in municipality that owns a car
IncGroup	*Av_IncomeGroup*	Average income decile in municipality[16]

Calculating distance metrics

As presented in table 3.5, the geometric data contains 32 distance metrics, representing the shortest distance to a wide variety of facilities. For the purpose of the regressions, as regressing on 32 different distance measures is very excessive, these 32 measures are condensed into six measures, based on their facility category. For this, to give each facility an equal weight in the final distance metric, firstly all distance measures are scaled to a [0 , 1] range, using formula 4.4.

$$x_{scaled} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{4.4}$$

These rescaled distances are then averaged out per category of facilities, to create six separate distance metrics for each municipality. The constructed metrics, what they present, and how many locations they cover, are presented in table 4.4. Each of these metrics thus presents the relative average distance to facilities in that category. All values of these metrics also fall within the [0 , 1] range, where a high value implies a larger relative average distance to those facilities.

Table 4.4: Aggregation of distance metrics

Variable	Type of facilities included	Number of facilities included
Dist_Emergency	GP centre, hospitals, fire brigade	5
Dist_Stores	Pharmacy, supermarket, department store	4
Dist_Horeca	Bars, restaurants, hotels, cafes	4
Dist_Education	Daycare, primary and secondary schools, after-school care	6
Dist_Connection	Train stations, main road entry	3
Dist_Recreation	Library, swimming pool, cinema, theatre, museum, etc.	10

Handling missing values

Lastly, as stated in section 3.3.1, the variable *HouVal* contains missing values for three municipalities. Given the nature of the to be performed (spatial) regressions, omitting these municipalities from the data might have fairly large consequences for the data, as this would also impact the spatial lags

[15]When calculating the average age, all values of *Age* = '99 and older' are taken as 99

[16]For this, deciles are converted to number (1st decile = 1, 2nd decile = 2, etc.), to allow for calculating an average. This implicitly assumes linear steps between the deciles, which is not necessarily the case

of surrounding municipalities. Therefore these missing values will be handled by imputing a central value of *HouVal* for these entries. As the *HouVal*-variable is not perfectly symmetrically distributed in the data, the median value over all other municipalities of this variable is imputed to get rid of these missing values in the data.

4.2.2. Non-spatial regressions

To get a idea of what factors might be able to explain differences in travel times between regions, firstly several non-spatial regressions are performed, using different sets explanatory variables. These variables can be divided into three categories: (1) socio-demographic variables, presented in table 4.3, (2) distance metrics, presented in table 4.4, and (3) built environment variables. The built environment variables originally selected for use in regressing are *AddDens, PopDens, Pop, NumBusi, LabPart,* and *HouVal* (table 3.5).

Multicollinearity

Now, before the actual final regression models are run, within each subgroup of explanatory variables is it determined whether multicollinearity is present. Multicollinearity occurs when predictory variables are highly correlated amongst each other, and can undermine the significance of the coefficients that the model yields, as high multicollinearity inflates their standard errors, which is detrimental to the model (Kim, 2019). Because of this, model coefficients might not show up as significant, even if there is a clear link between a predictor and the dependent variable. This can occur when multiple predictory variables represent the same/a similar concept in the data and are strongly correlated (which is the case for the built environment- and distance variables, as shown later in figure 5.15). To determine whether multicollinearity exists within the groups of variables, we consider the often-used Variance Inflation Factor (VIF), of which the threshold value is traditionally considered to be between 5 and 10. Here, we consider the strict threshold of a VIF of >5 to indicate multicollinearity.

The sets of socio-demographic variables and distance metrics do not show any sign of problematic multicollinearity within their sets, the VIFs of all their variables remaining below 5. The group of built environment variables does however show problematic levels of multicollinearity, with VIFs going as high as 20. To solve this problem, we remove variables with high multicollinearity, until the set of variables left does not show multicollinearity anymore. This results in removal of the variables *AddDens*, and *NumBusi*. In hindsight, it makes sense that problematic collinearity showed up here, as *AddDens* is a density measure similar to *PopDens*, and *NumBusi* can be assumed to be fairly strongly correlated to *Pop*. Removal of these variables thus leaves us with a final set of four variables in the built environment group (without multicollinearity): *PopDens, Pop, LabPart*, and *HouVal*.

In other regression models where these groups of variables are combined, some variables' VIFs might still ever so slightly go over 5, but as they never go over 6, this limited multicollinearity in some of these models is not further considered to be problematic for this research.

Setup of non-spatial regression models

In all the performed regression analyses, the average travel time expenditure per person (*AvTTE*) is the dependent variable. This variable is regressed on all possible combinations of the three groups of explanatory variables, resulting in 7 different models, as presented in table 4.5. All these models are estimated using Ordinary Least Squares (OLS). All regressions also only fit linear models, according to formula 4.5, as the scatterplots of the explanatory variables with *AvTTE* variable do not provide any indication of possible quadratic or other higher order/non-linear relationships (appendix C). In running these regression models, we address relationships 2 and 4 from the conceptual model in figure 2.6.

$$y_i = \alpha + \sum_k x_{ik}\beta_k + \epsilon_i \tag{4.5}$$

y_i = dependent variable for observation i; α = regression constant; x_{ik} = value of explanatory variable k for observation i; β_k = coefficient of variable k; ϵ_i = error term for observation i

Important to note here, is that in constructing these regression models all observations enjoy the same importance when estimating the parameters. This means that the values of all municipality are considered in the same manner, not taking into account their size, population, or another magnitude factor. The constructed aggregated measures for the least represented municipality (Renswoude, n=42) and the most represented municipality (Rotterdam, n=4380) in the data thus have the same influence on the model. An alternative way of model construction (weighted regression) would take into account these differences, by assigning more importance to the larger municipalities, given that their values are based on more data points, and therefore can be considered more reliable. Though this does sound logical at first, a result of this would be that the model results would be most heavily based on the situation in the larger municipalities. As we are looking into spatial (in)equity, limiting the influence of the smaller municipalities is something we want to avoid, hence this method of weighing the observations is not applied.

Table 4.5: Independent variables in non-spatial regressions

Model	Variable groups included	Number of variables
1	Socio-demographic	8
2	Built environment	4
3	Distance	6
4	Socio-demographic, built environment	12
5	Socio-demographic, distance	14
6	Built environment, distance	10
7	Socio-demographic, built environment, distance	18

4.2.3. Spatial regressions

Now, to consider the spatial aspect of the data, additional, more advanced, spatial regressions models are built as well. There are various ways to incorporate space in regression analyses, that all focus on a different spatial aspect of the data. For this research, six spatial regression models are constructed, covering four district ways of including a spatial aspect. These four ways are:

(1) **Spatial lag model.** This model extends on the non-spatial model by also including the spatial lag of the dependent variable as a predictor.

(2) **Spatial lagged X model.** This model extends on the non-spatial model by also including the spatial lag of (some) explanatory variables as predictors.

(3) **Spatial fixed effects model.** This model extends on the non-spatial model by allowing the α to vary per group with a different (spatial) characteristic.

(4) **Spatial heterogeneity model.** This model extends on the non-spatial model by allowing both the α and the βs to vary per group with a different (spatial) characteristic.

For the spatial regressions described in this section, as base set of nine explanatory variables is included: *Perc_HighEdu, Av_IncomeGroup, Perc_Car, Perc_ForeignBG, Dist_Horeca, Dist_Connection, Dist_Recreation, HouVal*, and *PopDens*. These nine variables are selected, as they are the only ones that yielded a significant effect on the average travel time in at least one of the non-spatial regressions from table 4.5, as further described in section 5.3.1. All spatial regression models once again are unweighted, assigning equal importance to the situation in each municipality, regardless of their magnitude.

Spatial lag model

The first way of incorporating space into the regression, is by construction a spatial lag model. In this model, the dependent variable is regressed not only on the nine selected exogenous explanatory variables, but also on the spatial lag of the dependent variable itself. The reason for this, is that travel times in a certain area could have a spillover effect into nearby regions (Condeço-Melhorado et al., 2014). To measure this and account for this, the surrounding values (= the spatial lag) of travel time should also be regressed on. This results in regression formula 4.6.

$$y_i = \alpha + y_{sl-i}\rho + \sum_k x_{ik}\beta_k + \epsilon_i \tag{4.6}$$

y_i = dependent variable for observation i; α = regression constant; y_{sl-i} = spatial lag of the dependent variable for observation i; ρ = coefficient of the spatial lag of the dependent variable; x_{ik} = value of explanatory variable k for observation i; β_k = coefficient of variable k; ϵ_i = error term for observation i

However, this formulation of the model is not suitable for OLS regression. Because the spatial lag of the dependent variable is constructed using the dependent variable itself (formula 4.1), simultaneity between these two variables appears. The spatial lag (an explanatory variable) is therefore dependent on the dependent variable, which violates the exogeneity assumption of OLS regression models (Mustafa, 2024). Instead, the Two-Stage Least Squares (2SLS) regression is used for constructing the spatial lag model. In this method, first an instrumental variable is computed which contains estimated values of the spatial lag, based on other predictor values in the model. This instrumental variable is then used in fitting the regression model (IBM, 2024). Because this instrumental variable is no longer based on the values of the dependent variable, this solves the endogeneity problem.

Spatial lagged X model

In a similar way that travel times in certain regions might affect travel times in other nearby regions, other exogenous variables might also show spatial spillover effects. Travel times might not (just) depend on a municipality's size or density, but also on the size or density of the surrounding regions that people might often travel through or into. Therefore, we also want to consider the values of these variables at a wider scale, which can be done in an Spatial lagged X model. In this model, the dependent variable is not only regressed on the nine selected exogenous variables, but also on the spatial lag of some built environment variables, which thus represent these built environment characteristics for the wider surrounding region of a municipality. This results in the formulation of the model as in formula 4.7.

$$y_i = \alpha + \sum_k x_{ik}\beta_k + \sum_m x_{sl-im}\gamma_m + \epsilon_i \tag{4.7}$$

y_i = dependent variable for observation i; α = regression constant; x_{ik} = value of explanatory variable k for observation i; β_k = coefficient of variable k; X_{sl-im} = spatial lag of variable m for observation i; γ_m = coefficient of the spatial lag of variable m; ϵ_i = error term for observation i

$\{k\}$ = set of normally included explanatory variables; $\{m\}$ = set of explanatory variables for which the spatial lag is used

The model formulation shows that the regression is now built up of the summation over two groups of variables. The first summation considers the regular variables, and the second summations considers the spatial lag variables. For the Spatial lagged X model that is fitted in this research, {m} contains (the spatial lag of) *PopDens* and *Pop*. These variables are included as they could possibly have spillover affects due to the (travel) activity that the presence of people and/or businesses might generate. {k} contains, once again, the nine selected exogenous variables. This regression model can be fitted using OLS, as there are no exogeneity issues in this model formulation.

Spatial fixed effects model

A third way of incorporating space into the regression, is by allowing the fitted regression model to vary over space. This might allow the model to capture unobserved variance that is not already captured by the nine selected explanatory variables. If this unobserved variance differs across space (e.g. differences between cities compared to rural areas), allowing the model the model to vary with this spatial characteristic might help it explain the data better. In a spatial fixed effects model, only the intercept (α) of the model is allowed to vary with spatial characteristics. This results in a model formulation as in formula 4.8.

$$y_i = \alpha_r + \sum_k x_{ik}\beta_k + \epsilon_i \tag{4.8}$$

y_i *= dependent variable for observation i*; α_r *= regression constant for spatial subgroup r that i belongs to*; x_{ik} *= value of explanatory variable k for observation i*; β_k *= coefficient of variable k*; ϵ_i *= error term for observation i*

The way this is accomplished is by including an additional binary variable in the data that is used to distinguish the observations with different spatial characteristics from each other. Given the model formulation, when performing the OLS regression, only observations that fall into the same binary group are compared to each other. This results in the β_k representing the intra-group effect of a change in variable x_k. Inter-group differences are captured by the varying α, which can be different for each group. Conceptually, this is very similar to the inclusion of binary explanatory variable that represents the effect of being in one group, compared to the other. Only now in this model, this effect is represented in the difference between the the α values.

In this research, two spatial fixed effects model are constructed. In the first model, the intercept is allowed to vary by whether a municipality is considered urban or not. For this, the five-level urbanism class variable *UrbClass* is recoded into a binary variable, which indicates whether a municipality is considered urban (address density >= 1.500 addresses/km^2) or not urban (address density < 1.500 addresses/km^2). By allowing a different intercept for both groups, we aim to find out if there is a structural binary difference in travel times between urban and non-urban regions.

The second spatial fixed effects model aims to assess any structural differences between border municipalities and non-border municipalities. Most low clusters of travel times appear near a country border (section 5.2.2), which could be an indication of border effects (Ma et al., 2024). To assess whether there are truly any significant differences in travel time expenditures between border regions and non-border regions, a binary variable is constructed that indicates whether a municipality is on a country border. Once again, in the spatial fixed effects model, the intercept is allowed to vary between border and non-border groups.

Spatial heterogeneity model

The fourth and final spatial regression method extends on the spatial fixed effects model. While the fixed effects model only allows the intercept to vary between subgroups, the effects of the nine selected explanatory variables (the β coefficients) are forced to remain constant across groups. However, if the impact and importance of these variables is not equal across all regions, the data contains spatial heterogeneity. To estimate a model that does also consider spatial heterogeneity, we need to allow the β values to vary across subgroups as well. This results in a model formulation as presented in formula 4.9. Conceptually, this is very similar to the inclusion of interaction effects of the model, that would represent the β differences between groups. Only now in this model, this effect is represented by the allowing to fit separate β coefficients, one for each group.

$$y_i = \alpha_r + \sum_k x_{ik}\beta_{k-r} + \epsilon_i \tag{4.9}$$

y_i *= dependent variable for observation i*; α_r *= regression constant for spatial subgroup r that i belongs to*; x_{ik} *= value of explanatory variable k for observation i*; β_{k-r} *= coefficient of variable k for spatial subgroup r that i belongs to*; ϵ_i *= error term for observation i*

Once again, just like in the spatial fixed effects model, when using OLS to fit this model, only comparisons between observations from the same subgroup are used to estimate the α and β values. In this model, both the intercept and the coefficients can vary across subgroups, thus essentially fitting a separate regression model for each subgroup. This allows for addressing whether, besides differences in the intercept, there might also be differences in importance of the predictors in different spatial settings (e.g. perhaps car ownership has a bigger impact in rural areas compared to urban areas, as rural areas more often require a car for getting around). The spatial heterogeneity model is able to capture these differences.

In this research, two spatial heterogeneity models are constructed. Like for the spatial fixed effects models, one model considers the difference between urban and non-urban municipalities, while the other model considers the difference between border and non-border municipalities. This represents the interaction effects that these spatial characteristics may have both on the effect of other spatial characteristics (figure 2.6, relationship 6), and on the effect of the included personal characteristics, a possible interaction effect that did not emerge from the literature.

4.2.4. Conceptualisation of tested relationships

As mentioned in section 2.6, the conceptual model derived from the literature does not exactly correspond with the relationships that we have tried to test in these analyses. For sake of clarity, we revisit this here and conceptually present the relationships we did investigate in this research. This conceptualisation is shown in figure 4.6, where the left side shows the simplified model. The right side shows the elaborate model, where the spatial characteristics are divided into three factors: urban, border, and distance. This allows us to also show their relative coherence, and how and where they exactly have been used.

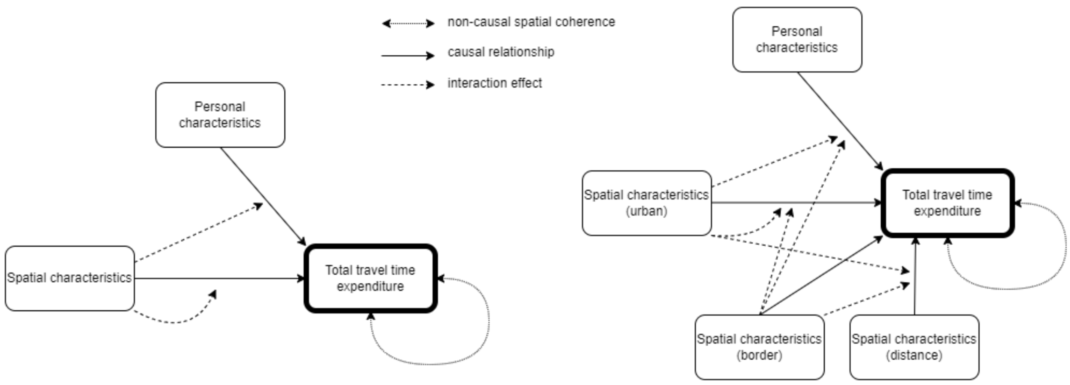


Figure 4.6: Conceptualisation of the relationships that were actually tested in the performed analyses. Both the simplified conceptualisation (l) and the elaborate conceptualisation (r) are shown

Considering the simplified model, we see that the spatial coherence of travel time expenditures is tested for. We do this by determining the measures of spatial autocorrelation, both globally and locally. We have test for the direct effects of personal characteristics and spatial characteristics on the travel time expenditures, by means of construction the various regression models. Lastly, we also consider possible differences in strength of the effects of certain factors on travel time expenditures, expecting that these strengths could vary with location. This constitutes an interaction effect of the spatial characteristics on the direct effects (of both groups of variables), and this is tested for in the spatial heterogeneity models, as explained earlier when elaborating on these models.

Now, we can distinguish three types of spatial aspects that we do consider in the end. To more clearly show how exactly each of these are included in the analyses, we split these up in the more elaborate right model in figure 4.6. The direct effect on travel time expenditures of all three types are considered. The urbanism and distance factors are included as normal regression factors, and the direct effect of the border is tested by its dummy-like inclusion in the spatial fixed effects model. Once again, the spatial heterogeneity models allow for estimation of interaction effects, which are thus estimated for both urbanism and the border factor. In these models, both these factors' effects on the direct influence of socio-demographic, urbanism, and distance factors are tested.

The (possibly confounding) influence of personal characteristics on spatial characteristics (figure 2.3, relationship 3) is not directly analysed in this research, as the relationship between these two groups of variables in itself is not a subject of investigation within the scope of this study. Still, by including both sets of variables in regression models, we do correct for these possible confounding effects. Additionally, we do not consider any confounding effect of personal characteristics on the influence of spatial characteristics (figure 2.3, relationship 5), but rather the other way around: a possible confounding effect of spatial characteristics of the influence of personal characteristics (which was not mentioned in the literature). This decision is based on the results of the spatial autocorrelation assessment (section 5.2) and non-spatial regression models (section 5.3.1) and has three main reasons:

- Firstly, the regression models show a more clear independent influence of personal characteristics on travel time expenditure, as opposed to spatial characteristics. Looking for variations in this clearly stronger effect seemed more logical.
- Secondly, the spatial autocorrelation already indicated an importance of border vicinity, begging the question whether this effect could have a confounding nature.
- Lastly, and more generally, as this research has a focus towards regional differences, any variation in importance of (significant) determinants inherently is relevant to consider.

5 Results

5.1. Descriptive statistics

Before the results of the analyses are presented, some general descriptive statistics are presented to gain more general insight into the data, which will help better understand and evaluate the outcomes of the analyses. For the variable of interest, the average travel time expenditure per municipality, these statistics are presented in table 5.1. The distribution of values is also visually presented in figure 5.1.

Table 5.1: Descriptive statistics of municipal averages of travel time expenditure (min)

Mean	68,18
Median	68,08
Standard deviation	7,62
Minimum	39,67
Maximum	95,45
Total range	55,78
25th percentile	63,35
75th percentile	73,06
Inter-quartile range	9,71

As one can see from the table and the figure, there is a fairly wide distribution of values, with the highest value being about 2,4 times as large as the lowest value. Still, with an IQR of 9,71, 50% of all observations fall within a margin of about 5 on either side of the mean, hinting at at least some consistency of travel times across municipalities, though the total range of values is quite large. The mean and median value are very similar, implying a symmetrical distribution, which is supported by the histogram in figure 5.1. This average value of 68,18 minutes of travel per day is also consistent with the most common values found in the literature, as was presented in chapter 2.

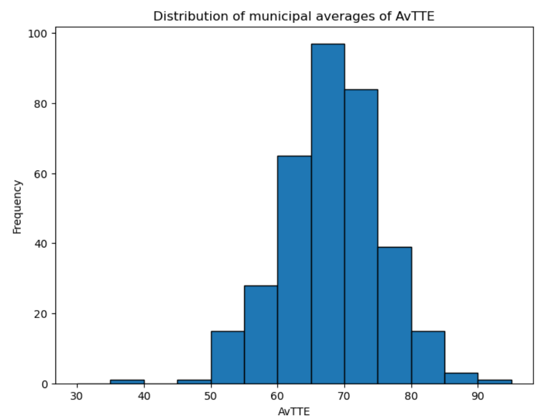


Figure 5.1: Distribution of municipal averages of travel time expenditure

The value distributions of used explanatory variables used in this research are not extensively covered here, but rather presented in appendix D. This appendix shows all distributions of (1) the explanatory variables at individual level derived from the ODiN data, (2) the built environment variables at municipal level from the CBS data, and (3) the socio-demographic variables aggregated to municipal level, as described in section 4.2. The socio-demographic variables (at individual level) inherently show a representative sample of the Dutch population, following from its very rigorous selection method (CBS, 2019, 2020b). The built environment variables are official, real-world values and thus also reliable.

5.2. Spatial autocorrelation

The descriptive statistics thus do indicate some limited consistency of travel times, but also still show quite some variability in the values. To assess how the differences and consistencies geographically relate to each other, spatial data analysis is applied. To show spatial correlation in the data, first the global spatial autocorrelation is assessed, to see whether any spatial coherence appears to exist at all. Following this, local spatial autocorrelation is also assessed, to identify more clearly what spatial patterns appear where in the spatial structure.

5.2.1. Global spatial autocorrelation

To get a first feeling for how relatively high and low average travel time expenditures are spatially distributed across the country, for both the average travel time per municipality and its spatial lag choropleth maps are created. These maps (figure 5.2) clearly show where certain relative values appear, by dividing the values into quintiles and colouring in the municipalities according to which quintile they belong to.

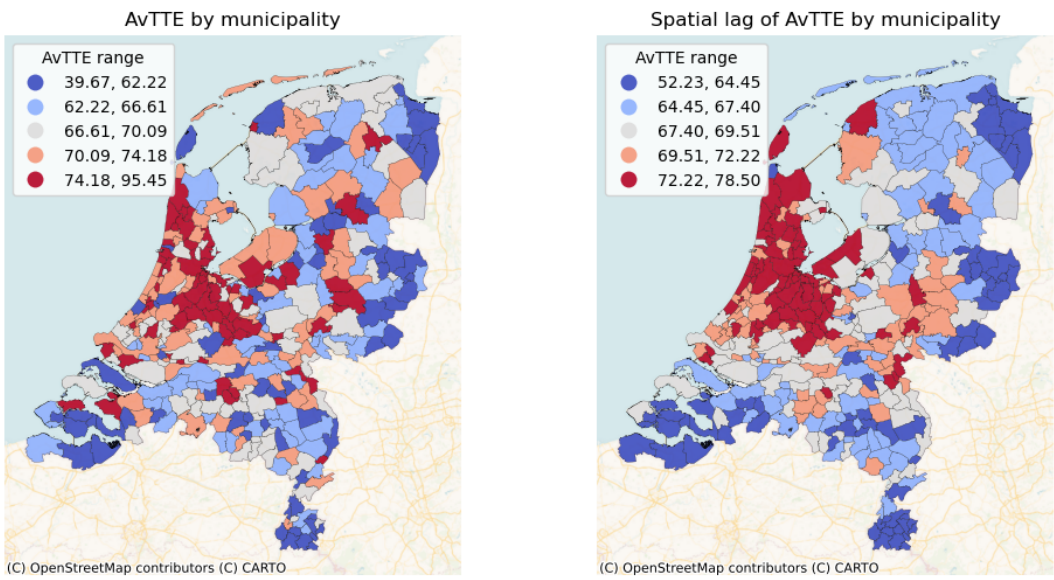


Figure 5.2: Quintile plots of the average travel time expenditure (left) and its spatial lag (right)

From these maps some first general insights about the spatial structure can be derived. Firstly, when looking at the left map, there appear to be several regions where similar values seem to cluster together. Mainly in the northern part of the Randstad (in Utrecht and North Holland) high travel times show up in plenty adjacent municipalities. Near the country border, multiple regions can be identified that consist of adjacent municipalities that all fall in the lower segment of travel time expenditures. Additionally, when also considering the right map, which shows the spatial lag, we see fairly similar patterns. High spatial lag seems to coincide with high values themselves. Therefore, this first look at these maps does show some indication of the existence of positive spatial autocorrelation in travel time expenditures.

More clear and conclusive insight into the amount of spatial autocorrelation in the data can be derived from the Moran Plot (figure 5.3). The plot does appear to show some positive correlation between the average travel time and spatial lag of the average travel time. The linear fit through the scatterplot, shown in red, also is an indication of this positive relationship. The cloud of data points is however also dispersed fairly widely around this linear fit, so this plot does not imply a very strong relationship. Calculating the global Moran's I (= the slope of the linear fit) using formula 4.2 gives a value of 0,27, which indicates a moderate amount of spatial autocorrelation.

To assess whether this amount of spatial autocorrelation can be considered significant, figure 5.4 shows the reference distribution of 999 more permutations that have been run. As the left plot shows, from this dataset, based on random allocation of values to locations, one would expect roughly no spatial autocorrelation. The true Moran's I of this dataset (indicated by the red line), at 0,27 falls far outside the references distribution of the random simulations. This process yields a p-value of 0,001 for our Moran's I, implying that the found spatial coherence is significantly different from the randomly generated maps. It can therefore be stated that the data shows weak to moderate, but very significant, spatial autocorrelation at a global level. The right pane of figure 5.4 once again shows the Moran plot with the fitted line, now in standard deviations.

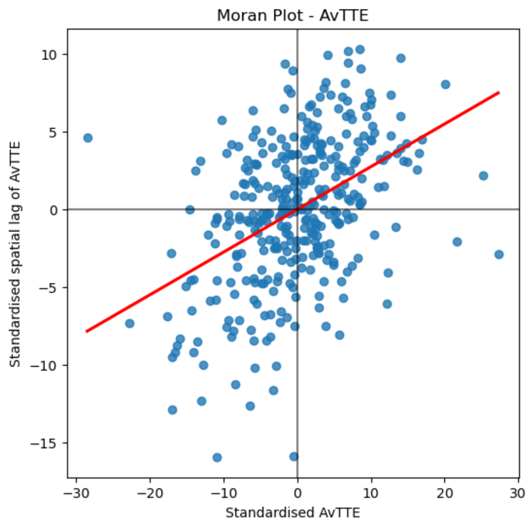


Figure 5.3: Moran plot of average travel time expenditures

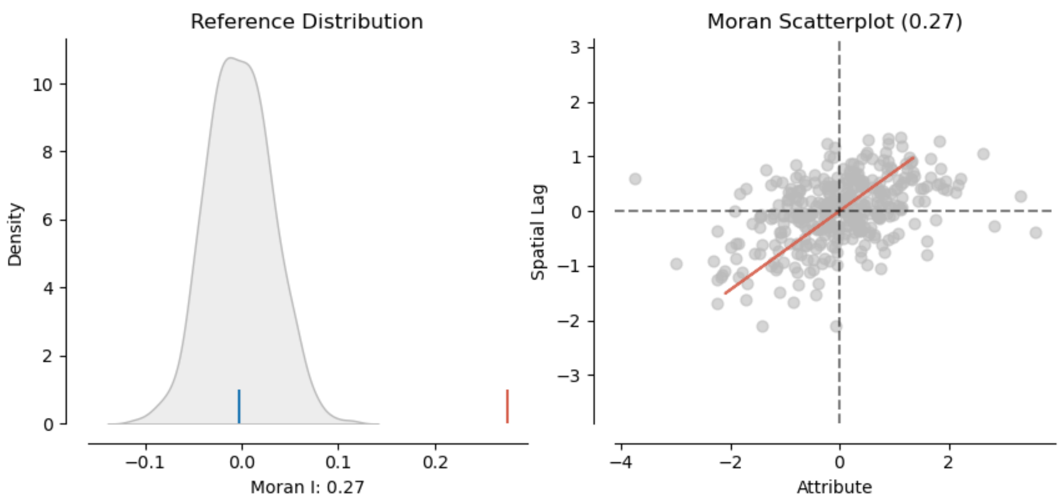


Figure 5.4: Reference distribution for determining the significance of the global Moran's I. In the left plot, the blue line indicates the expected value of the Moran's I, based on the permutations. The red line indicates the true Moran's I found.

5.2.2. Local spatial autocorrelation

For zooming in further and assessing the spatial autocorrelation at a local level, for each municipality it is determined what quadrant (HH, HL, LH, LL) it belongs to, as per table 4.2. This classification is based on the previously constructed Moran plot, as is illustrated in figure 5.5, where the data points are divided into four groups, based on their travel time value and spatial lag. Figure 5.6 shows this classification on the map, where the colouring of the data points in figure 5.5 thus matches the colouring of the corresponding municipalities in figure 5.6.

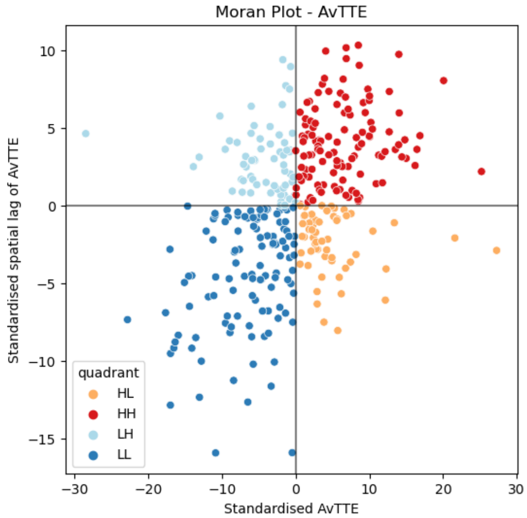


Figure 5.5: Moran plot with quadrant classification

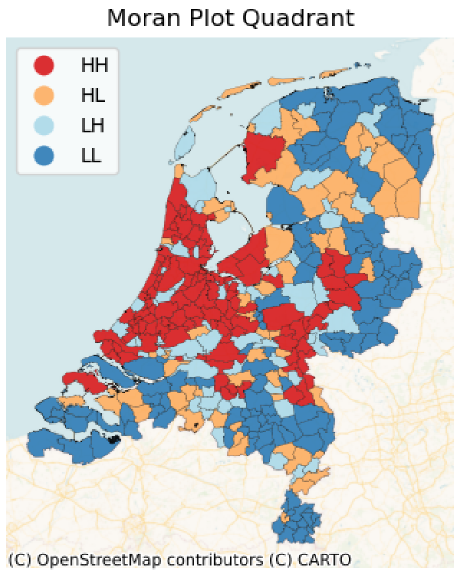


Figure 5.6: The quadrant classification shown for each municipality

Though this classification into clusters gives some first insight into where certain value concentration seems to appear, as explained in the methods (section 4.1.2), for assessing the significance of these labels we need to consider the local Moran's I statistic and its significance. For each municipality, the local Moran's I is calculated based on formula 4.3. The results of these calculations are once again spatially mapped and presented in figure 5.7. In this figure, green indicates positive spatial autocorrelation, and purple/pink indicates negative spatial autocorrelation. Again, positive spatial autocorrelation can imply the occurrence of HH or LL, and negative spatial autocorrelation can imply the occurrence of LH or HL.

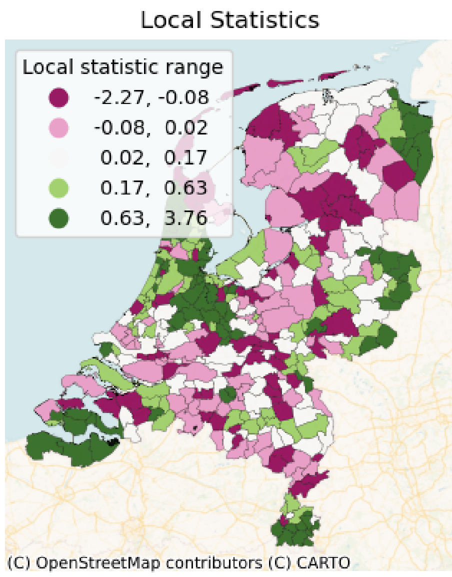


Figure 5.7: Local Moran's I per municipality

To test for significance of the local statistic, 24.999 more permutations have been run. Based on these permutations, a local statistic was deemed significant if the random permutations yielded a more extreme local Moran's I for a certain municipality less than 5% of the time. These significant municipalities are shows in pink in figure 5.8. Of these municipalities it can thus be said that the found spatial pattern (its value's relationship to the neighbouring values) is unlikely to arise from a random distribution of this dataset. These are thus deemed municipalities of interest.

When combining the significance results from figure 5.8 with the cluster assignment from figure 5.6, the areas of interest can clearly be highlighted. This is done so in figure 5.9, which only illustrates the significant areas and shows what quadrant they belong to. From this image, the preliminary insights that were derived from the global spatial autocorrelation assessment are confirmed: Utrecht and North Holland do show significant clustering of relatively high travel times over a larger region. More specifically, we distinguish two regions here: (1) the northern part of the Randstad (Amsterdam, Utrecht, and their surrounding regions) and (2) the northern part of North Holland (Alkmaar and the region north of it). Meanwhile, the clustering of relatively low travel times appears in five border regions in Groningen, Overijssel, Gelderland, Limburg, and Zeeland.

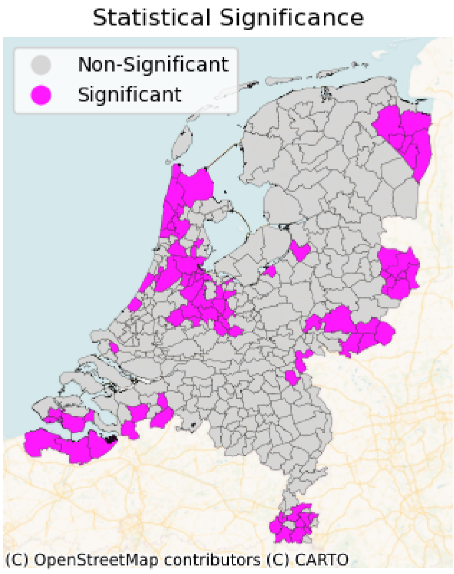


Figure 5.8: Significance of the local Moran's I per municipality

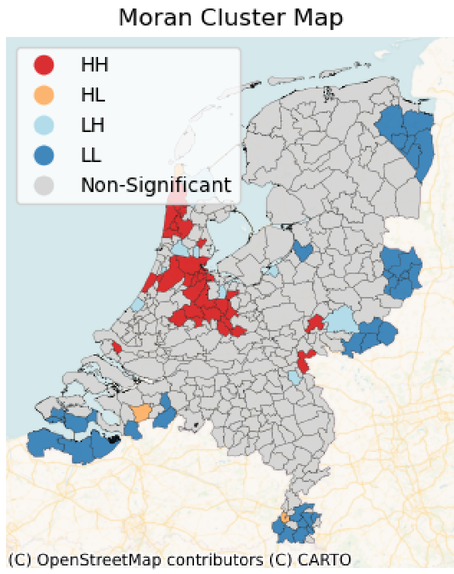


Figure 5.9: Quadrants of the municipalities with a significant local Moran's I

In total, 22,57% of the municipalities show some form of significant local spatial autocorrelation. The exact number of municipalities with each type of spatial pattern are presented in table 5.2. As the table shows, the most common spatial patterns are HH and LL (76 in total). HL and LH only appear a total of 13 times in the data. Positive spatial autocorrelation thus appears almost six times more often in the data than negative spatial autocorrelation. This is consistent with the earlier finding that the data as a whole globally contains positive spatial autocorrelation (global Moran's I = 0,27) as well.

Table 5.2: Number of significant municipalities in each quadrant

Quadrant	Number of municipalities
HH	30
HL	3
LH	10
LL	36

There are two more interesting things regarding figure 5.9 that are worth mentioning here. First, as presented in the literature review, most literature stated that travel times tend to be slightly higher in the big cities (Mokhtarian & Chen, 2004). Because of this, one could reasonably expect clustering of high travel times in the Randstad region, which contains most of the bigger urban areas in the country. While this clustering of high travel times does appear in the north part of the Randstad, and with it in and around two (Amsterdam and Utrecht) of the four big cities, this patterns does not show up around the other two (Rotterdam and The Hague).

Second, the larger clusters of low travel times all show up close to the border with Germany and/or Belgium. Most of these municipalities flagged as LL tend to be relatively small and rural. However Maastricht and Enschede, both strongly urban and part of the 25 largest cities in the country, also belong to this group. Additionally, there are way more small and rural areas in the country that do not show significantly low patterns, so the fact that these significant regions all appear on the edge of the country could perhaps be indicative of a border effect, where the border might form an invisible barrier limiting travel across it (Jin et al., 2021; Ma et al., 2024). If such a border effect does indeed apply here, this could pose a possible explanation for these results. Some caution is however appropriate when interpreting these results as definitive proof of a border effect, given the nature of the data and the exclusion of travel outside of the Netherlands. We come back to reflect on this is in section 5.5.

5.2.3. Characteristics of municipalities of interest

In the search for possible reasons why the found municipalities of interest show these clustering high-/low travel time values, the built environment and socio-demographic characteristics of these municipalities are studied. These characteristics are compared to the data-wide values to find out if anything sets these municipalities apart. For many variables, the groups of HH, HL, LH, or LL clusters do not appear to be very different from the rest of the data. The full comparison of clusters against the full data is presented by the figures in appendix E. Still, there are a few variables that do seem to be associated more with certain clusters in the data.

Firstly, when considering the average house value, a slight difference can be seen in the HH- and LL-municipalities compared to the full data, but also mainly compared to each other. This is shown in figure 5.10. Though both groups of observations mostly fall within the range of fairly high density, when comparing them to each other, the municipalities labelled HH appear more towards the middle to high-end of the house value spectrum, while the LL-municipalities are more common under the low-end to middle part. This might imply some form of positive relationship between travel times and house prices.

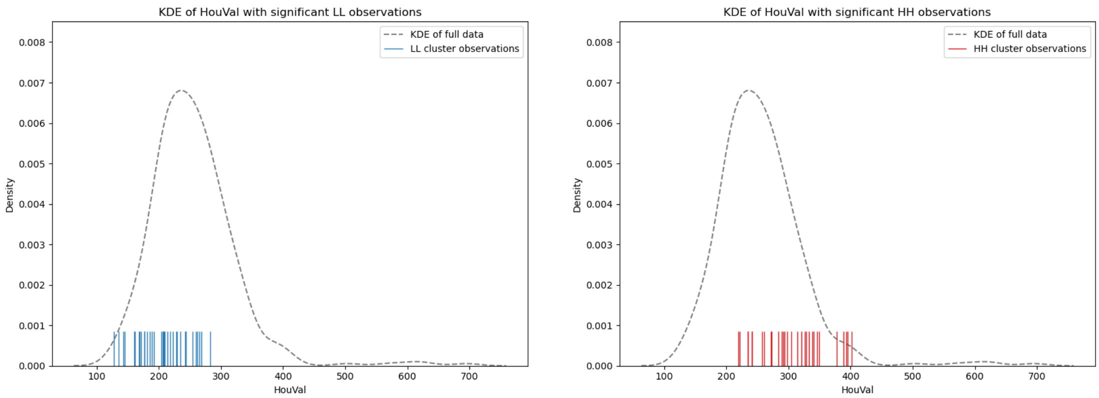


Figure 5.10: Kernel density function of the average house value, with the values for the LL and HH municipalities marked

When looking at the socio-demographic background of the respondents from the municipalities of interest, a few more characterisations stand out. As can be seen in figure 5.11, in the areas labelled as HH clusters, young adults (20-40 years old) are more strongly represented. Additionally, the other three clusters (HL, LH, LL) seem to show stronger representation of older people (65 years old and above). This does seem to point towards some relationship between age and travel time expenditure as well, with older people travelling less than younger people, though this relationship is not so clear-cut, as older people are also amply represented in the HL cluster, where the municipality itself thus shows high travel times as well. This apparent difference between age groups is consistent with the findings from other literature (Raux, Ma, Joly, & Cornelis, 2011).

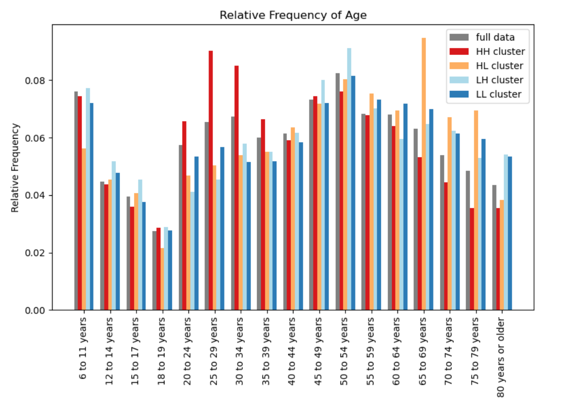


Figure 5.11: Age distribution of respondents across groups of clusters compared to the full data

When considering the education level of the respondents (figure 5.12), we especially see a difference in the level of highly educated (university and/or higher vocational education) individuals. In the clustering HH-municipalities, this group is way more prominently present. This could possibly be an indication of a relationship between education and travel time expenditure, as earlier mentioned in the literature review as well (Feng et al., 2013). Similar to education, higher incomes are overrepresented in the HH-municipalities (figure 5.13). This figure also shows strong underrepresentation of high incomes in LL-areas, thus hinting towards a positive relationship between income and travel time expenditures as well. Once again, this is not surprising given the existing literature already pointing out this relationship (Koo et al., 2022).

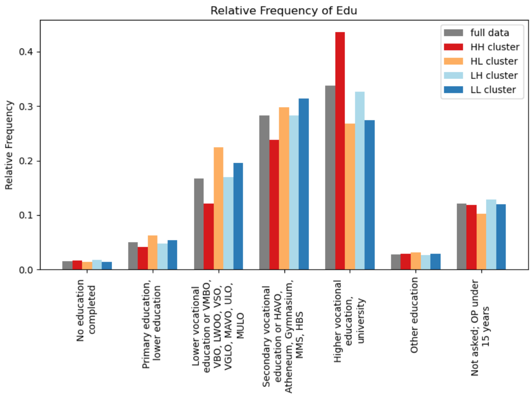


Figure 5.12: Education distribution of respondents across groups of clusters compared to the full data

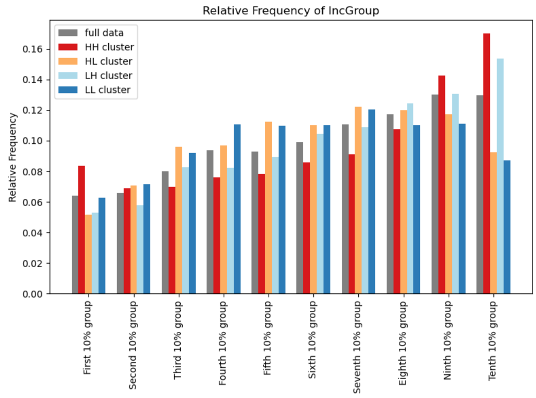


Figure 5.13: Income distribution of respondents across groups of clusters compared to the full data

Lastly, and more surprisingly, the representation of car ownership is covered, as shown in figure 5.14. In this figure we see an overrepresentation of people that do not own a car in the group of HH-municipalities. Strangely enough, regions where travel times are high thus seem to have lower car ownership. This goes against some literature that does find that car ownership is positively correlated with total travel time expenditure (Wigan & Morris, 1981). While it makes sense to assume that car owners are more mobile and therefore would travel more, these first findings do not seem to support this notion. One could also argue this the other way around, if car ownership would decrease travel time as one might be more mobile (faster) with a car, compared to using other modes for similar trips.

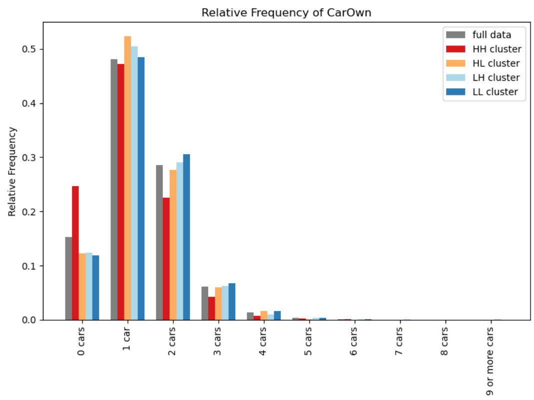


Figure 5.14: Car ownership distribution of respondents across groups of clusters compared to the full data

5.3. Regression analysis

To add to the understanding of why the found spatial patterns in travel time expenditure appear, we turn to the results of the regression analyses, in which the average travel time expenditure per municipality are regressed on the various sets of explanatory variables in the data.

5.3.1. Non-spatial regressions

Correlations between variables

Before the results of the regression models themselves are discussed, we take a first look at the plain correlations of all the explanatory variables with our dependent variable (table 5.3). Not a single variable shows strong ($|corr|$ > 0,7) or even moderate ($|corr|$ > 0,5) correlation with the dependent variable. There are three variables that do show a weak ($|corr|$ > 0,3) linear relationship, these variables being

Perc_HighEdu, *Av_IncomeGroup*, and *HouVal*. The correlation of these variables with the average travel time expenditure being the highest is consistent with the findings in the previous section, which showed income, education, and house prices to be higher in regions with higher travel times. All other variables do not show any clear linear relationship with *AvTTE*, their correlations all being lower than 0,3.

Table 5.3: Correlations of explanatory variables with average travel time expenditure (*excluded from the actual regression models, to avoid multicollinearity)

Variable	Correlation *AvTTE*	Variable	Correlation *AvTTE*
Perc_Male	-0,056	Pop	0,104
Av_Age	-0,091	NumBusi*	0,120
Perc_ForeignBG	0,091	LabPart	0,085
Perc_HighEdu	0,439	HouVal	0,322
Perc_WorkStud	0,250	Dist_Emergency	-0,148
Perc_Car	-0,139	Dist_Stores	-0,139
Perc_Children	0,027	Dist_Horeca	-0,057
Av_IncomeGroup	0,389	Dist_Education	-0,141
AddDens*	0,167	Dist_Connection	-0,177
PopDens	0,174	Dist_Recreation	-0,166

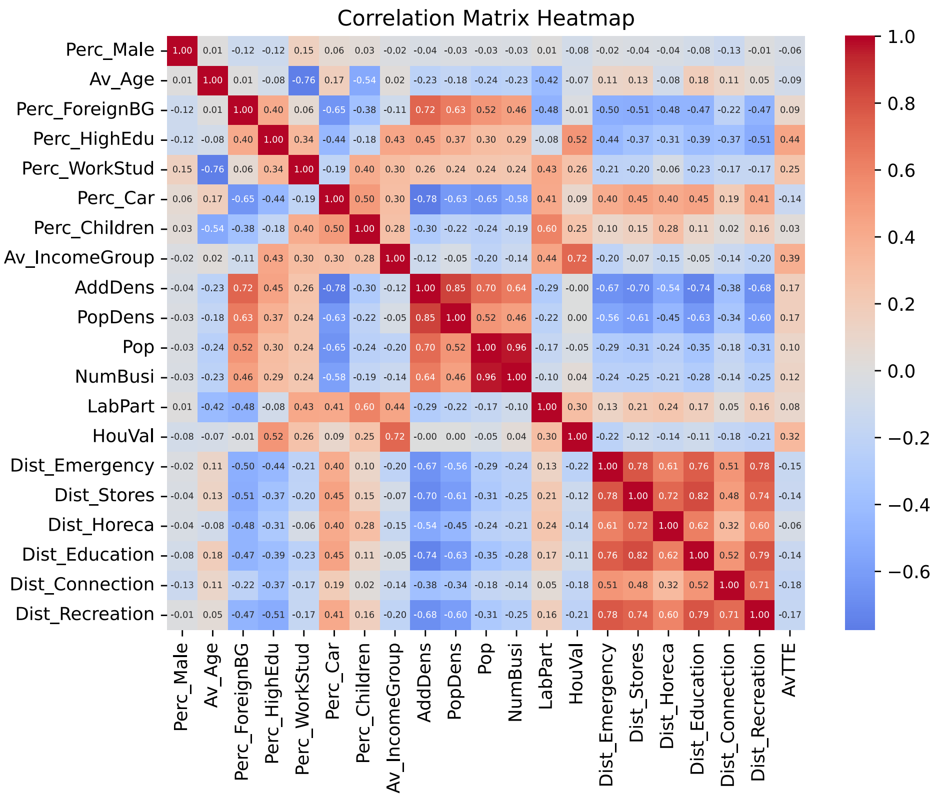


Figure 5.15: Correlation matrix of the explanatory variables amongst each other (and the dependent variable)

To be able to understand possible confounding effects that (groups of) variables might have on the relationships between other (groups of) explanatory variables and travel time expenditure, we also take a look at the correlation matrix of the explanatory variables (figure 5.15). Some takeaways from this matrix are:

- Distance measures are strongly positively correlated amongst each other, and negatively with variables representing address and population density. This intuitively makes sense, as in less dense areas, facilities are more likely to be further away on average.
- Car ownership percentage is negatively correlated with density and population. In larger, more dense municipalities, people are thus less likely to own a car.
- The foreign background percentage is positively correlated with density and population. Larger, more dense municipalities thus have a larger proportion of residents that have a foreign background.
- The average house value is positively correlated with the percentage of high education and strongly positively correlated with the average income. This also makes sense, as people with a higher income (who more often also have enjoyed a higher education) are better able to afford living in more expensive regions.

The high correlations amongst the distance measures might make one suspect issues regarding multicollinearity in the regressions, and the same applies to the correlations amongst the built environment variables. However, as already described in section 4.2.2, this was already accounted for by not including the variables *AddDens* and *NumBusi* in further analyses, solving the multicollinearity amongst the built environment variable group. The distance measures actually do not suffer from problematic multicollinearity, despite their high correlations.

Non-spatial regression outcomes

Model comparison To see how well the different sets of explanatory variables are able to explain the average travel time values derived from the data, we first turn to the results of the performed non-spatial regressions (as they were specified in section 4.2.2). A summary of the key insights gained from these regressions is presented in table 5.4, which shows for each model which variables show a significant relationship (p-value of the coefficient is <0,05) with the dependent variable, and measures for the model's fit to the data. For addressing the model fit the proportion of explained variance in the dependent variable is considered, using the R^2 and adjusted R^2. The full model outputs, including the coefficients and probabilities, can be found in appendix F.

Table 5.4: Results of the non-spatial regression models (Included: sd = socio-demographic variables, be = built environment variables, dis = distance measures)

Model	Included	Significant variables (p<0,05)	R^2	Adjusted R^2
1	sd	*Perc_HighEdu, Av_IncomeGroup, Perc_Car*	0,258	0,241
2	be	*PopDens, HouVal*	0,136	0,126
3	dis		0,040	0,024
4	sd, be	*Perc_HighEdu, Av_IncomeGroup, Perc_ForeignBG*	0,267	0,240
5	sd, dis	*Perc_HighEdu, Av_IncomeGroup, Perc_Car, Dist_Horeca, Dist_Connection, Dist_Recreation*	0,294	0,264
6	be, dis	*PopDens, HouVal*	0,151	0,126
7	sd, be, dis	*Perc_HighEdu, Av_IncomeGroup, Perc_Car, Dist_Horeca, Dist_Connection, Dist_Recreation*	0,304	0,267

These model results already provide us with some interesting insights into the dependence of the average travel time expenditure on the used explanatory variables. When only considering the socio-demographic variables (model 1, table F.1), we see that higher education and higher income have a positive effect on travel time expenditure. Car ownership negatively affects travel time expenditure, which goes against some of the literature (Wigan & Morris, 1981), but is consistent with figure 5.14.

In model 2, containing just the built environment variables, the population density and the average house value are significant (table F.2). It is interesting to note that when both the socio-demographic and the built environment variables are considered (model 4, table F.4), the built environment is no longer significant, but income and education still are. This may imply a confounding effect of income (and/or education) on the impact of the built environment on travel time expenditure (figure 2.6, relationships 3 and 4). The built environment on itself does not seem to impact travel time expenditure, but this is a by-product of both factors being related to income and/or education, which also showed up in figure 5.15. We also see that the negative influence of car ownership from model 1 is no longer significant when including built environment variables in model 4, while the percentage of people with a foreign background gains significance.

Another thing to notice is that when only including distance metrics (model 3), the distance to various locations does not significantly affect travel time expenditures (table F.3). Interestingly though, when also considering socio-demographic factors (model 5, table F.5) or even both socio-demographic and built environment factors (model 7, table F.7), thus singling out the pure effects, distances to horeca locations, recreational locations, and connection points do significantly affect travel times. Living further from horeca or recreation increases travel time expenditure, while living further from a connection point (thus being less accessible) decreases travel time.

When considering the explanatory power of the models (R^2), we see that on its own the socio-demographic model (model 1) is able to explain way more variance of the dependent variable than only the built environment model or only the distance model (model 2 and 3). We also see that the full model with all variables included (model 7) is able to explain the most variance of the dependent variable, at 30,4%. Even when penalising models for added complexity by including more variable (adjusted R^2), the full model still performs best, though the difference with model 5 is small.

Model interpretation By taking a closer look at the results of the 7^{th} model, more insight into the direction and strengths of the identified relationships can be gained. The full model results that include the (standardised) coefficients of model 7 are presented in table 5.5. By considering the unstandardised coefficients we can interpret what impact a change in socio-demographic distribution would have on travel time expenditures:

- If the percentage of highly educated people in a municipality is increasing by 10 percentage points, the average travel time expenditure would increase by 2,07 minutes.
- If the percentage of car owners in a municipality is increasing by 10 percentage points, the average travel time expenditure would decrease by 2,87 minutes.
- If the average income group in a municipality would increase by 1 (i.e. the whole population would on average shift a decile upwards in relative income), the average travel time would increase by 7,22 minutes.

Though significant, the real-world effects of these factors are fairly small. To compare their relative importance, the standardised coefficients are useful. Here we see that the average income group is the relatively most influential variable of the three, with a change of 0,49 standard deviations in travel time expenditure for a change of one standard deviation in this variable. All these socio-demographic variables are more influential than the distance metrics included in the model, all their standardised coefficients being stronger.

The interpretation of the distance metrics is different, given the different nature that the values of these variables have in the data. As explained earlier and shown in formula 4.4, these metrics represent the average relative distance for all locations included in this metric, ranging from 0 to 1. A value of 0 would imply that a municipality has the shortest distance (of all municipalities) to all included locations. A value of 1 would imply that a municipality has the longest distance (of all municipalities) to all included locations. The coefficients of these metrics should therefore be interpreted as follows: the difference in travel time expenditure between a hypothetical municipality that is closest to all facilities, and a hypothetical municipality that is furthest from all facilities. The value of the distance metric indicates where along this range a municipality lies. The distance coefficients therefore imply:

- In a municipality that would have the furthest distance to all the closest hotels, restaurants, and bars, the average travel time would be 11,61 minutes higher than in a municipality that would have the shortest distance to the closest hotels, restaurants, and bars.
- In a municipality that would have the furthest distance to all the closest connection points, the average travel time would be 13,39 minutes lower than in a municipality that would have the shortest distance to all the closest connection points.
- In a municipality that would have the furthest distance to all the closest recreational locations, the average travel time would be 22,76 minutes higher than in a municipality that would have the shortest distance to all the closest recreational locations.

Table 5.5: Results of non-spatial regression model 7 (*significant at the α=0,05 level)

R^2	0,304	**Adjusted R^2**	0,267
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	51,680*		0,005
Perc_Male	-0,087	-0,046	0,372
Av_Age	0,056	0,023	0,822
Perc_ForeignBG	-0,100	-0,108	0,157
Perc_HighEdu	0,207*	0,250*	0,002
Perc_WorkStud	0,012	0,010	0,918
Perc_Car	-0,287*	-0,225*	0,024
Perc_Children	0,059	0,056	0,497
Av_IncomeGroup	7,222*	0,492*	0,000
PopDens	0,001	0,107	0,158
Pop	0,000	0,029	0,650
LabPart	-0,179	-0,091	0,225
HouVal	-0,009	-0,084	0,257
Dist_Emergency	8,578	0,096	0,273
Dist_Stores	-8,732	-0,132	0,185
Dist_Horeca	11,606*	0,170*	0,022
Dist_Education	-6,443	-0,106	0,279
Dist_Connection	-13,389*	-0,156*	0,026
Dist_Recreation	22,764*	0,277*	0,011

Variable selection for spatial analysis For the more advance spatial regression analyses, the focus is on the variables that are shown to have a significant effect on travel time expenditure in at least one of the non-spatial regression models. This includes the variables *Perc_HighEdu, Av_IncomeGroup, Perc_Car, Perc_ForeignBG, Dist_Horeca, Dist_Connection, Dist_Recreation, PopDens*, and *HouVal*. We first fitted a last non-spatial model using only these variables, so later results from the spatial regressions can be compared to this base model. From the model outputs in table 5.6, we see an ever so slightly higher adjusted R^2 compared to model 7. It is also interesting to note that after excluding all other variables, *Dist_Horeca* loses its significance, while *PopDens* regains it again. Its coefficient here implies that with a density increase of 1.000 persons per km^2, the average travel time expenditure would increase by 1 minute.

Table 5.6: Results of non-spatial regression with selected variables (*significant at the α=0,05 level)

R^2	0,289	**Adjusted R^2**	0,270
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	43,707*		0,000
Perc_HighEdu	0,225*	0,271*	0,000
Av_IncomeGroup	6,086*	0,415*	0,000
Perc_Car	-0,254*	-0,199*	0,014
Perc_ForeignBG	-0,070	-0,075	0,271
Dist_Horeca	7,501	0,110	0,076
Dist_Connection	-12,377*	-0,144*	0,033
Dist_Recreation	17,207*	0,210*	0,020
HouVal	-0,007	-0,066	0,352
PopDens	0,001*	0,142*	0,046

5.3.2. Spatial regressions

Before the results of the spatial regressions are presented, two remarks need to be made on the inclusion and exclusion of certain elements of the results in this report. First, for the spatial lag model, the spatial fixed effects models, and the spatial heterogeneity models, no standardised coefficients are presented. This has to do with the nature of these analyses causing standardised coefficients to lose their original meaning and making them more difficult to interpret. For the spatial fixed effects models and the spatial heterogeneity models, this is due to the fact that models are fitted on subsets on the data, though the standardisation is performed centrally on the full data, before the models are fitted. Therefore, the standardised coefficients do not adequately represent the same concepts for these models as they do for other models.

For the spatial lag model, this is due to the fact that the model does not automatically standardise the spatial lag values as well, resulting in a coefficient for the spatial lag that is not directly comparable to the other coefficients in the standardised model. Additionally, the spatial lag model does not yield a normal R^2 and adjusted R^2 as model fit measures, as this model is not estimated using normal OLS, but rather with 2SLS. Instead, it yields a Pseudo R^2 and a Spatial Pseudo R^2. Here, the Pseudo R^2 is best suitable for cross-model fitting comparisons to non-spatial models, as it, just like the regular R^2, represents the squared correlation between the dependent variable's values and the model predictions ("spreg.GM_Lag", n.d.).

Spatial lag model (SL model)

When considering the spatial lag model (table 5.7), there are a few interesting results to point out. Firstly, this model that includes the spatial lag of *AvTTE* is able to explain 3,8 percentage points (32,7% compared to 28,9%) more of the variance in the average travel time expenditure across the country. Given the fact that the spatial autocorrelation analysis also identified significant correlation between the dependent variable and its spatial lag, this finding is not surprising.

Additionally, we see that the inclusion of the spatial lag has a major impact on the coefficients and the significance of the other predictor variables. For some variables, this inclusion results in an increase of their coefficients, while for other variables their coefficients decrease. Interestingly, *Perc_Car, PopDens* and *Dist_Connection* as well as the intercept, even lose their significance. Interestingly, *HouVal* does become significant again, where it was not before. Even when including spatial lag as a predictor, the most prominent predictors from the non-spatial regressions, *Perc_HighEdu* and *Av_IncomeGroup* remain significant. Given these various impacts on the different predictors, combined with the fact that the spatial lag of *AvTTE* is strongly significant as a predictor and has a fairly strong coefficient compared to other predictors, this model does imply that there is some spatial aspect to the travel time expenditures across regions that cannot be explained by the included predictor variables alone.

Table 5.7: Results of spatial lag model (*significant at the α=0,05 level)

Pseudo R^2	0,327	**Spatial Pseudo R^2**	0,280
Variable	**Coefficient**		**P-value**
CONSTANT	1,629		0,899
Perc_HighEdu	0,247*		0,000
Av_IncomeGroup	4,106*		0,001
Perc_Car	-0,107		0,302
Perc_ForeignBG	-0,078		0,200
Dist_Horeca	0,637		0,883
Dist_Connection	-6,587		0,248
Dist_Recreation	19,975*		0,005
HouVal	-0,019*		0,016
PopDens	0,001		0,105
W_AvTTE	0,651*		0,000

Spatial lagged X model (SLX model)

The results from the SLX model are presented in table 5.8. When comparing the model fit of this model to the model from table 5.6, the differences are very minor. The SLX model explains 0,7 percentage points more variance, but considering the added complexity of the model with the two additional predictors, the adjusted R^2 barely increased at all.

When further considering the coefficients, it also shows that considering the built environment variables for a wider region around the area of interest does not add any explanatory benefit to the model, as none of the included spatial lag variables are significant. Including these factors also has very little impact on the other predictors. Though the values of their coefficients do show minor adjustments, they all roughly stay constant in impact on the travel time expenditure. Additionally, there are no changes in significance of the predictors, when comparing this model to the model in table 5.6. One final interesting thing worth noticing here, is that the coefficient of the spatial lag of population density is opposite to the coefficient of the local spatial density itself, though we cannot conclude anything definitively from this, given the spatial lag coefficient is thus not significant.

Table 5.8: Results of SLX model (*significant at the α=0,05 level)

R^2	0,296	**Adjusted R^2**	0,273
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	40,719*		0,000
Perc_HighEdu	0,199*	0,240*	0,002
Av_IncomeGroup	6,729*	0,459*	0,000
Perc_Car	-0,257*	-0,202*	0,013
Perc_ForeignBG	-0,049	-0,052	0,457
Dist_Horeca	7,457	0,109	0,082
Dist_Connection	-13,502*	-0,157*	0,021
Dist_Recreation	16,267*	0,198*	0,028
HouVal	-0,005	-0,044	0,535
PopDens	0,001*	0,182*	0,015
W_PopDens	-0,002	-0,154	0,077
W_Pop	0,000	0,074	0,281

Spatial fixed effects model for urban/non-urban comparison (SFE urban model)

When considering the first spatial fixed effects model (table 5.9), that allows for a different intercept between urban and non-urban municipalities, once again the gain in explanatory power is not that large. By allowing for a varying intercept, the model gains 0,2 percentage points in its R^2 and 0 in its adjusted R^2. When considering the coefficients of the explanatory variables, most of them do not show any major differences and most of them maintain their (in)significance. The only exemption to this is that by allowing the intercept to vary between urban and non-urban regions, *PopDens* loses the significance of its coefficient again. This does make sense, as a distinction between urban and non-urban regions likely already incorporates differences in population density as well.

Table 5.9: Results of spatial fixed effects model for urban/non-urban comparison (*significant at the α=0,05 level)

R^2	0,291	**Adjusted R^2**	0,270
Variable	**Coefficient**		**P-value**
CONSTANT urban	42,841*		0,000
CONSTANT not urban	41,702*		0,000
Perc_HighEdu	0,229*		0,000
Av_IncomeGroup	6,027*		0,000
Perc_Car	-0,235*		0,026
Perc_ForeignBG	-0,077		0,228
Dist_Horeca	7,304		0,084
Dist_Connection	-12,735*		0,029
Dist_Recreation	18,970*		0,013
HouVal	-0,006		0,432
PopDens	0,001		0,122

When considering the varying intercept for the two groups, a slight difference does show up. This model estimates a slightly higher intercept (and therefore slightly higher travel time expenditures in general) for urban regions. The difference in the intercepts is about 1,14, which implies 1,14 minutes more average travel time expenditure in urban regions compared to non-urban regions. To determine whether these two separately fitted models are truly different from each other, a Chow test is performed. This basically tests whether the split regression is a better fit for the data than a singular regression on the full dataset (Chow, 1960). With the Chow test yielding a p-value of 0,358, the difference between the two regressions is not deemed significant at the 5% level, which means that we cannot conclude that there exists a definitive differences in the models (and in this case, the intercepts) between the two groups. This thus once again points to a lack of evidence that urban regions make for structurally higher travel times, just like the built environment variables were not significant in the non-spatial regressions when controlling for socio-demographic variables.

Spatial fixed effects model for border/non-border comparison (SFE border model)

The second spatial fixed effects model (table 5.10) allows for a different intercept between border and non-border municipalities. Here, the gain in explanatory power is larger than in the previously covered spatial models. Compared to the non-spatial model with the selected variables, this spatial fixed effects model explains 4,0 percentage points more of the variance in the dependent variable, and also gains 0,040 in its adjusted R^2. In this model, *PopDens* again loses its significance by allowing for a varying intercept, just as in the other spatial fixed effects model, and so does *Perc_Car*.

The difference in intercepts between the two distinguished groups of observations is larger than in the previous spatial fixed effects model. Here, it is indicated that in non-border municipalities, the intercept, and thus the travel time expenditure in general, is 4,58 minutes higher than in municipalities that border Germany or Belgium. This time, the Chow test yields a probability of 0,000, implying a statistically significant difference between the two groups. This model therefore implies that indeed people in border regions tend to travel less, even when controlling for the other factors in the model, which is consistent with the findings portrayed in figure 5.9.

Table 5.10: Results of spatial fixed effects model for border/non-border comparison (*significant at the α=0,05 level)

R^2	0,329	**Adjusted R^2**	0,310
Variable	**Coefficient**		**P-value**
CONSTANT border	35,921*		0,000
CONSTANT not border	40,503*		0,000
Perc_HighEdu	0,243*		0,000
Av_IncomeGroup	5,635*		0,000
Perc_Car	-0,175		0,087
Perc_ForeignBG	-0,013		0,838
Dist_Horeca	2,302		0,589
Dist_Connection	-13,334*		0,018
Dist_Recreation	22,689*		0,002
HouVal	-0,012		0,122
PopDens	0,001		0,167

Spatial heterogeneity model for urban/non-urban comparison (SHG urban model)
In the spatial heterogeneity models we not only allow the intercept to vary between the groups, but also the coefficients of all other predictors. The results of the spatial heterogeneity model varying with urbanisation, presented in table 5.11, show some interesting findings. Firstly, the model shows a better fit than the spatial fixed effects model for urbanism with a higher R^2, but the adjusted R^2 is only marginally higher, due to the model being way more complex.

When considering the coefficients for both submodels, it is interesting to note that there is quite some variety in significance between the two groups of observations. Where in urban regions *Perc_HighEdu, Av_IncomeGroup, Dist_Connection, Dist_Recreation*, and *PopDens* are significant predictors, in non-urban regions only *Av_IncomeGroup* and *Perc_Car* are significant. *Av_IncomeGroup* is thus the only variable that appears to have an impact in both subgroups of the data. These differences would imply that in urban regions the average travel time expenditure is influenced by different factors than in non-urban regions, which is interesting and perhaps somewhat unexpected. Especially surprising is that this model implies that even between rural (non-urban) regions, population density seems to have a slight impact on travel time expenditures.

Table 5.11: Results of spatial heterogeneity model for urban/non-urban comparison (*significant at the α=0,05 level)

R^2	0,316	**Adjusted R^2**	0,276	
Variable	**Coeff. (not urban)**	**P-value**	**Coeff. (urban)**	**P-value**
CONSTANT	35,445*	0,016	64,385*	0,000
Perc_HighEdu	0,256*	0,000	0,045	0,761
Av_IncomeGroup	5,645*	0,000	9,813*	0,001
Perc_Car	-0,190	0,221	-0,625*	0,004
Perc_ForeignBG	-0,040	0,614	-0,205	0,078
Dist_Horeca	6,415	0,162	23,489	0,054
Dist_Connection	-15,034*	0,018	-11,168	0,499
Dist_Recreation	26,692*	0,002	22,340	0,373
HouVal	-0,002	0,836	-0,025	0,250
PopDens	0,003*	0,047	0,001	0,289

For this model, once again a Chow test is performed to determine whether the differences between the models for both subgroups are significant, and whether the split in the models does create a better fit for the data. This Chow statistic is determined for the submodels as a whole, but also for each variable separately, as presented in table 5.12. This table shows that there is no significant difference between both submodels, not at a model-wide level, and also not for any specific variable. Therefore, like in the spatial fixed effects model for urbanism, from this model we cannot conclude there are clear differences in travel time expenditures or their determinants between urban and non-urban regions.

Table 5.12: Chow test statistics for the spatial heterogeneity model for urban/non-urban comparison (*significant at the α=0,05 level)

Variable	Chow statistic	P-value
Global test	12,828	0,233
CONSTANT	1,703	0,192
Perc_HighEdu	1,637	0,201
Av_IncomeGroup	1,646	0,200
Perc_Car	2,646	0,104
Perc_ForeignBG	1,394	0,238
Dist_Horeca	1,730	0,188
Dist_Connection	0,048	0,827
Dist_Recreation	0,027	0,870
HouVal	0,996	0,318
PopDens	1,800	0,180

Spatial heterogeneity model for border/non-border comparison (SHG border model)

The final constructed spatial model is the spatial heterogeneity model for comparing border municipalities to non-border municipalities (table 5.13). The R^2 of this model shows that this is the best model yet, being able to explain 35,1% of the variance in the dependent variable, though the adjusted R^2 is only ever so slightly (0,004) higher than for the SFE border model.

When considering the coefficients of both submodels, immediately it draws the eye that the model for border municipalities I does contain only one significant variable besides the intercept, namely *Perc_Car*. Even the so far always significant income and education levels are not significant in this submodel. In the non-border model, *Perc_HighEdu, Av_IncomeGroup*, and *Dist_Recreation* all show significant effects on the travel time expenditures. While this may imply a difference in importance of these variables between the two subgroups of observations, it must be stated that there are only 59 border municipalities, compared to 291 non-border municipalities. This difference in data availability could also have an effect on the significance of the variables in the fitted models, with relationships being less easily detectable in the smaller border group.

Table 5.13: Results of spatial heterogeneity model for border/non-border comparison (*significant at the α=0,05 level)

R^2	0,351	Adjusted R^2	0,314	
Variable	**Coeff. (border)**	**P-value**	**Coeff. (not border)**	**P-value**
CONSTANT	102,965*	0,000	32,627*	0,001
Perc_HighEdu	0,023	0,894	0,278*	0,000
Av_IncomeGroup	4,062	0,203	5,328*	0,000
Perc_Car	-0,726*	0,013	-0,081	0,470
Perc_ForeignBG	-0,099	0,457	0,039	0,624
Dist_Horeca	-8,960	0,432	2,382	0,610
Dist_Connection	-16,389	0,151	-11,669	0,091
Dist_Recreation	15,790	0,409	22,692*	0,005
HouVal	0,026	0,457	-0,015	0,060
PopDens	-0,005	0,072	0,001	0,143

When looking at the Chow test for model comparison for this model (table 5.14), we see that, like in the spatial fixed effects model for border/non-border comparison, the two submodels as a whole do significantly differ from each other. By allowing the coefficients to vary between the two subgroups, the model fit is significantly better than when fitting one linear model on the full dataset. *PopDens* and *Perc_Car* also seem to have significantly different coefficients, according to this test. The resulting coefficients for *PopDens* were however not significant in either submodel, so this insight remains ambiguous. The *Perc_Car* variable seems to be more impactful in border regions, compared to non-border regions.

Table 5.14: Chow test statistics for the spatial heterogeneity model for border/non-border comparison (*significant at the α=0,05 level)

Variable	**Chow statistic**	**P-value**
Global test	31,618*	0,001
CONSTANT	5,782*	0,016
Perc_HighEdu	1,976	0,160
Av_IncomeGroup	0,138	0,711
Perc_Car	4,257*	0,039
Perc_ForeignBG	0,792	0,373
Dist_Horeca	0,850	0,357
Dist_Connection	0,126	0,723
Dist_Recreation	0,110	0,740
HouVal	1,337	0,248
PopDens	4,138*	0,042

5.3.3. Overall regression results

Table 5.15 gives a final overview of the results of the six performed spatial regressions. As we see, the SHG border model performs the best, having the highest R^2 and thus being able to explain the most variance in the average travel time expenditure. This model also performs better than the best non-spatial model (table 5.4). When considering the complexity of the constructed models by looking at the adjusted R^2, the SHG border model still is the best, though it is only marginally better than the SFE border model. In general these border models have the highest explanatory power, while showing significant differences between border and non-border regions. The SFE urban and SHG urban models are not able to supply similarly strong results. This is interesting to say the least, as going into this research, a urban effect was expected, where a border effect was not considered.

Table 5.15: Results of the spatial regression models (x = significant variable; / = significant in only one submodel)

Model	***Perc_HighEdu***	***Av_IncomeGroup***	***Perc_Car***	***Perc_ForeignBG***	***Dist_Horeca***	***Dist_Connection***	***Dist_Recreation***	***HouVal***	***PopDens***	**(Pseudo) R^2**	**Adjusted R^2**
SL	x	x					x	x		0,327	[17]
SLX	x	x	x			x	x		x	0,296	0,273
SFE urban	x	x	x			x	x			0,291	0,270
SFE border	x	x				x	x			0,329	0,310
SHG urban	/	x	/			/	/		/	0,316	0,276
SHG border	/	/	/				/			0,351	0,314

When looking at the independent variables that were included in the spatial models, we see that the two variables that also appeared most relevant from the non-spatial regression, namely education and income, also remained relevant within each of the different spatial implementations. No matter what

[17] As explained earlier, there is no equivalent of the adjusted R^2 calculated for the SL model

spatial factors were considered, these variable remained of importance, solidifying their significance and their effect on the travel time expenditures. Car ownership also remained of some significance in most models. It is also interesting to see that the distance measures remain significant in so many spatial models (*Dist_Recreation* even remained significant in all models), given the fact that their added explanatory power to the (non-spatial) models was very minimal. Most often, the built environment did not show significant influence on travel time expenditures in the spatial models, most notably not in the best-fitting models.

5.4. Overall findings

Looking at all performed analyses together, we can derive some final general insights:

- There are significant differences in travel time expenditures between municipalities, but these differences can hardly be explained by the built environment alone. The effect of the population density on travel time expenditures is explained away by the inclusion of socio-demographic variables in most models.
- The socio-demographic factors of income and education are the most important explanatory factors for travel time expenditures. Both spatial autocorrelation analysis and all the (non-)spatial regression models indicate these variables to be of great relevance. Both an average higher income and higher education level make for higher average travel time expenditures in a municipality.
- Though not as important as income and education, car ownership and distance measures also show significant relationships with travel time expenditures. More car ownership is related to less travel time expenditure. Larger distances to recreational locations are related to more travel time expenditure, but larger distances to connection points lead to less travel time expenditure.
- There appears to be a border effect, where people living close to the border have lower travel time expenditures. The SFE border and SHG border models that take this into account are also the models that allow for the best model fit on the data.

5.4.1. Interpretation of identified relationships

This research was initiated to address differences in travel time expenditures between regions, as a means of exposing possible inequalities in infrastructural facilities. In political debate and literature, a supposed lack of connectivity of rural regions is frequently mentioned, and therefore it is surprising to find that the built environment does not seem to exert its own clear influence on travel time expenditures. Clear differences between regions are rather mainly explained by differences in socio-demographic build-up, which still is a notion the literature supports.

Given the fact that literature hinted at higher travel time expenditures in larger cities, it was expected that the Randstad region would yield high travel time expenditures. Therefore, initially it is surprising that significant clusters of high travel time expenditures only show up in and around Amsterdam and Utrecht, but less so around Rotterdam and The Hague, even though these are clearly also large cities. Now knowing the importance of the socio-demographic variables, these offer an explanation for this, as Amsterdam and Utrecht do contain a higher proportion of high education individuals and a higher average income, compared to Rotterdam and The Hague (CBS, 2020a, 2023b). In and of itself, this identified positive effect of income on travel time expenditures provides food for thought on future mobility developments. In a very general sense, the world as a whole is considered to get ever richer over time (O'Toole, 2013). If income/wealth is indeed a driver of mobility, one would expect a perhaps slow, but steady increase in travel time expenditures over time. However, the historic data showing this factor to remain fairly constant in the past, does not reconcile with such a supposed development.

Some of the other found relationships are also worth reflecting on further. Car ownership was found to have a significant effect on travel time expenditure in some (not all) models. While this is not surprising in itself (the literature makes mention of this as well), the direction of this relationship in the constructed models is opposite of what the literature would suggest. The models show that an increase in car ownership in a municipality is associated with a decrease in travel time expenditures. Though not consistent with the existing literature, one could reason that a relationship in this direction is not illogical. Not owning a car might make people less mobile in a general sense, forcing them to take slower modes of transportation, like a bicycle or public transport, which would increase travel time expenditures if we assume that one will still make the same trips.

For the distance measures regarding recreational facilities, the association with travel time expenditure is positive, which makes sense. If one lives more remote (and thus less proximal to certain locations), one will have to cover larger distances to get around. Interestingly, living further from infrastructure connection points (train stations, main road entries, etc.) is associated with lower travel times. One could expect that being less proximal to connection points would cause travel time expenditure to increase out of necessity, but this effect is opposite, which could imply that worse connectivity actually deters people from travelling altogether. This is an important relationship that needs careful consideration in any infrastructural equity assessment that might use travel times as a measure.

Lastly, we discuss the implication of the found difference between border and non-border municipalities. The fact that significantly lower travel time expenditures are found in municipalities on the edge of the country could indicate the existence of some form of a border effect. A border effect implies that, though the borders with Belgium and Germany are 'soft' and easily crossable without additional efforts or costs, still people are more reluctant to do so. Some internal blockade occurs that makes people more hesitant to travel across the border (Jin et al., 2021; Ma et al., 2024). If this is indeed the case, having a border on one side of a municipality would limit the travel in that direction, thus leaving only country-inwards travel options. Possibly this limitation in travel directions, and with it the possible destinations, causes an decrease in overall travel, which would explain the lower travel time expenditure in these regions. Additionally, the best-fitting SHG border model showed some heterogeneity between border and non-border regions and implies an interaction effect on border vicinity on the influence of other (socio-demographic) predictors.

A note must also be made on the found R^2 of the constructed regression models. As table 5.4 and table 5.15 show, no model is able to explain more than about a third of the variation in the municipal averages in travel time expenditures. Though this is not an negligible portion, it also means that, even using the best of these models, still about 65% of the observed variation cannot be explained using the included factors. Without undermining the importance of the found predictors, it must be stated that this does also imply that the biggest portion of the variance might be explained by characteristics that are not considered in this research.

5.4.2. Conceptualisation of identified relationships

In section 4.2.4 we have conceptually shown what relationships were actually tested in the performed analyses. We here revisit this conceptualisation, as we can now identify which relationships do indeed appear to exist in the data, based on the results presented in this chapter. The adjusted conceptualisation, showing the identified relationships, is visualised in figure 5.16.

From the spatial autocorrelation analysis, as well as the SL model, it has become clear that spatial coherence of travel time expenditure exists beyond the influence of other independent variables. Regarding the direct effect of the various independent variables on travel time expenditures, we do find a clear influence of personal characteristics and distance measures, both in the non-spatial and spatial regression models. Additionally, the direct effect of border vicinity is substantiated by the results of the SFE border and SHG border models. There also appears to exist an interaction effect of border vicinity, but only on the personal characteristics (specifically car ownership), which was also shown by the SHG border model.

A clear unambiguous direct effect of urbanism on travel time expenditures was not found. Especially in the non-spatial regressions, any apparent urban influence seems to be explained away by the socio-demographic variables, based on which we expect the personal characteristic to be confounding this relationship. Though not directly tested, we therefore expect that personal characteristics might also influence one's urban environment, possibly through residential self-selection. We indicate this relationship with a blue arrow in figure 5.16. The SHG urban model did also not provide any evidence of an interaction effect of the urbanism on other relationships.

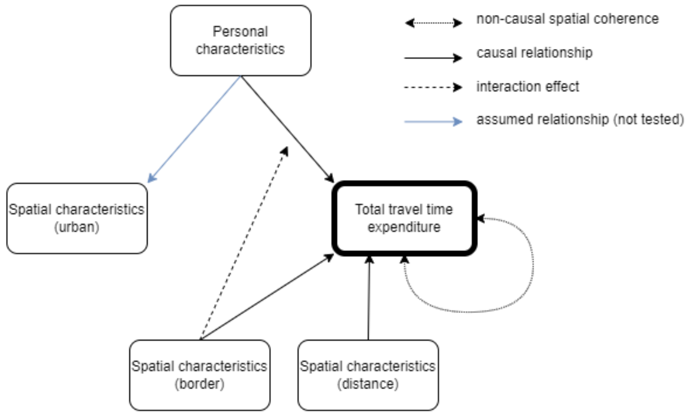


Figure 5.16: Conceptualisation of the relationships that were identified in the performed analyses

5.5. Discussion on the results

There are a few considerations to discuss regarding the general interpretations of the results of this research. Firstly, it is important to be aware of the relationship between the found patterns at municipal level and their counterparts at the individual level. Here, one must be aware that analysing these relationships at the municipal level can lead to ecological fallacy (Firebaugh, 2001). The effects of socio-demographic characteristics that have been identified at the municipal level cannot plainly be taken as applicable at the individual level as well. Firebaugh does present additional statistical methods that allow to derive unbiased individual level parameters from aggregated data. An important condition here is that "aggregated X ... is unrelated to Y, controlling for individual-level X" (p.4025). The aggregated dependent variable should thus not influence the dependent variable, outside of the effect that exists at the individual level. These methods are however not applied in this research.

To elaborate on this fallacy some more, we take the effect of income as an example. The found results imply that in a municipality with a higher average income level, the average travel time expenditure *of the municipal population as a whole* tends to be higher. This does not imply that for any individual the expenditure would increase with a higher income. For example, an individual might show high travel time expenditures as they live in an high-income municipality, though their own personal income might still be low. The analyses in this research thus only allow for drawing conclusions at an aggregated municipal level, but not for the individual level directly. Therefore, we can conclude that *a higher average income level of a municipality is related to higher average travel time expenditure*, but we cannot conclude that *a higher (personal) income is related to higher travel time expenditure*.

We need to consider more extensively the emergence of a border effect, it being one of the main, and definitely the most surprising, results from these analyses. As explained in section 3.2.4, persons that have made a full trip abroad are excluded from the data. Now, one may expect that this occurs more often in the actual border municipalities, which might impact these average travel time values more strongly, which would leave this result to be based on a data issue. Though this relatively is the case, still the actual portion of these occurrences in the total data is very small. This action removed only 0,51% of individuals' records in border municipalities (and 0,12% in non-border municipalities). Besides this, it is not necessarily the case that removing these individuals decreased the average travel time expenditures in those municipalities. Perhaps they only travelled a very short time abroad, leaving their total travel time expenditure to be vary low, but this we simply do not know for sure.

It is also important to stay aware of what exactly was measured in the performed analyses, which is the travel time expenditure. As explained in section 2.2, these expenditures can be considered proxies for ones travel time budgets. This means that the data used in this research does not necessarily exactly resemble one's true travel preferences, but it rather represents how much time one actually spent on travel, which is a result of one's (unknown) budget and the further influences that other factors might exert on travel behaviour. This is especially important to keep in mind when using these results in the context of accessibility (in)equity assessments, as high or low expenditures on their own cannot be used as pure indicators of this.

This last point is of primary importance. This research has identified high travel time expenditures in the Randstad regions with supposed good accessibility, and low travel time expenditure in border regions with supposed worse accessibility. In light of the originally represented rhetoric from section 1.2, which assumes worse accessibility would lead to higher travel time expenditures, these results are surprising. This would imply that the northern Randstad actually has worse accessibility. This cannot be due to proximity, which we know to be better in the Randstad, but in this line of reasoning this could also indicate a saturated infrastructure network due to busyness, leading to congestion. However, this argumentation does not hold when we extend it to the southern part of the Randstad, where we do not find this pattern in a still very busy region. Additionally, when we consider the possibility of modal shifts to slower modes of transport with increasing proximity, this direct connection between travel time expenditures and accessibility becomes even weaker.

For an alternative interpretation of the results, we need to make a distinction between travel time expenditures out of necessity (groceries, school/work, etc.) and for recreational purposes (any non-vital reasons). If we consider such a distinction, the higher travel in the northern Randstad could actually bear witness to better accessibility. This argument follows the line of reasoning that worse accessibility leads to lower travel time expenditures, as not having any options to (comfortably) reach destinations might deter people from travelling altogether. In the well-connected Randstad, recreation facilities are more proximal and thus people might spend more time on recreation travel besides their standard necessary travel, whereas in the country's outskirts people limit their travel to vital locations, as recreational facilities are not accessible enough. This line of reasoning is able to explain the found lower travel time expenditures in supposedly less accessible areas.

Now, even if we follow this second line of reasoning, one could argue that the accessibility of the rural regions is not as bad as is made out to be. The fact that the travel times are lower in these regions could be seen as a testimony to the notion that, at least for the necessary facilities, the accessibility is still sufficient, as people are able to visit all these vital locations within their exhibited lower travel time expenditure. It must be stated that from this research alone, it is not possible to determine which interpretation of the results in definitively correct. For that, additional travel information on trip purpose is needed to determine whether this possible distinction between necessary and recreational travel is able to explain the differences in travel time expenditure between regions.

It is also worth mentioning that some of the constructed regression models do suffer from heteroscedasticity, which means the error term of the regression does not have a constant variance across the range of values for the dependent variable. This is the case for the seven non-spatial models from table 5.4 and for the SFE urban model, as was indicated by their Breusch-Pagan test scores (Breusch & Pagan, 1979). The non-spatial selected variable model (table 5.6) and other spatial models did not suffer from this issue. Heteroscedasticity causes less reliability of the estimated coefficients, but does not directly affect the values of the coefficients or the model fit (R^2) themselves. One reason to deal with heteroscedasticity is by using robust standard errors. Though the results presented in this chapter do not incorporate this, a quick post-analysis check applying these robust standard errors to the affected models showed that this issue has not altered the general outcome of this research.

As a final note in this discussion, we come back to the fact that the regressions are unweighted, meaning every municipality enjoy equal importance, regardless of their size, as described in section 4.2.2. This does have as a consequence that measures on smaller municipalities are considered to be as reliable as those on larger municipalities, which could in theory skew results, if some aggregated measures are indeed not representative enough. Still, this choice was made to safeguard the importance of the smaller municipalities with this study's focus on spatial equity in mind. Additionally, the fact that a threshold of 30 respondents per municipality was used, already partly satisfies the requirement that the municipal measures should be based on a large enough sample, thus partly mitigating this unreliability risk.

6

Conclusion & Reflection

6.1. Answers to the research questions

The presented results allow us to formulate answers to the accompanying research questions that were presented in section 1.4. First, the main research question is answered, after which we answer each subquestions more elaborately as well.

How do travel times expenditures vary spatially across the Netherlands, and is there evidence to support the concept of constant travel times at a disaggregated scale?

Travel time expenditure do vary significantly across the (joint) municipalities in the Netherlands. The mean value of average travel time expenditures, at about 68 minutes, is consistent with previous findings (figure 2.1). However, differences between municipalities are large, values ranging from under 40 minutes to over 95 minutes per day. The results from the global spatial autocorrelation analysis show that these differences in average travel time expenditure do not appear totally randomly across spatial dimensions either. The maps in figure 5.2 show that, very generally speaking, municipalities with similar values in average travel time expenditure tend to be near each other a little more often than one would expect from a random spatial distribution. The value and significance of the Global Moran's I confirms this, the value of 0,27 implying weak to moderate spatial autocorrelation, but the p-value of 0,001 implying that the found pattern is definitely not random. From this we can thus conclude that there exist regions with higher travel time expenditures and regions with lower travel time expenditures. The range in values being this large, combined with the non-random spatial distribution of the values, does not support the concept of constant travel times at a geographically disaggregated scale.

1. *What is the state of existing knowledge on the spatial and socio-demographic determinants of travel time expenditures?*

This subquestion is answered with the literature search that is presented in chapter 2. Research into travel time budgets and/or expenditures is vast, and determinants of travel behaviour have been studies plenty already. The gist of the literature agrees that socio-demographic variables play an important part in this, with differences between groups being found in a multitude of studies. Mostly agreed upon relationships include higher travel times for males, people within the working age group (roughly 20 to 65 years of age), and persons with a higher income and/or education level. Higher travel time expenditures are also often found for car owners and people with more working hours. Still, there remains some lack of full consensus on the effects of the various factors.

When considering spatial factors, the results are more disputed. The studies that do find significant effects of the spatial context usually indicate (slightly) higher travel time expenditures in larger and more dense urban areas, as opposed to more rural areas. Within the Netherlands, this also points to higher travel time expenditures in the Randstad region. However, for these factors the opposing evidence is more abundant than for the socio-demographic determinants. Quite some literature concludes that there is no significant effect of the spatial context on travel times, or that these effects are much smaller than those of the socio-demographic factors.

2. *What spatial patterns in travel time expenditures across municipalities can be identified?*

Positive spatial autocorrelation was found in the data, which means that similar values of travel time expenditure tend to appear near each other more often. Though significant at a national level, the pattern is not very strong. In a large part of the country the spatial autocorrelation at a local level is not significant, but there are still some regions that do show significant distinct patterns. High values of average travel time expenditures appear a lot in the north Randstad region and the northern part of North-Holland. There are also five regions that show clustering of low values, which are all located near the county border. More generally, spatial regression models do show that the travel time expenditure is significantly different near the border, these expenditures being 4,74 minutes per day lower here according to the SFE border model. In the rest of the country, the travel time expenditure values do not show any clear patterns and appear to vary fairly randomly from municipality to municipality.

3. *How can spatial patterns in travel time expenditure across municipalities be explained by the spatial context?*

The short answer to this subquestion is: they cannot. When looking at characteristics of identified municipalities of interest, the main differences between them appeared not to be regarding the built environment, but rather socio-demographic. Regions with high travel time expenditure house more highly educated, high-income individuals, while regions with low travel time expenditure house fewer persons of from this educational background and income group. We also see more young adults in regions with high travel time expenditure, and more seniors in regions with low travel time expenditure. Finally, in regions with high travel time expenditures more people do not own a car.

Further support of the relationships between these socio-demographic variables and travel time expenditure is supplied by the regression analyses. When comparing the various non-spatial models we see that the predictive power of socio-demographic variables is way stronger than that of the included built environment variables. Only two built environment variable (the average house value, and population density) ever are significant predictors of travel time expenditure, but these influences are explained away when including socio-demographic variables as well. There appears to be a confounding effect of personal variables on the relationship between spatial characteristics and travel time expenditure, but a stand-alone effect of built environment factors on travel time expenditures is not found.

The spatial regression models substantiated once again the importance of the income and education predictors, as they remained significant in all regression models. Though not throughout all spatial models, still the distances to recreational facilities and connection points often also remained of some importance when including spatial elements in the models. Regarding the spatial aspects themselves, these models also allow us to derive some interesting conclusions. The SL model once again shows that there definitely is important spatial coherence in the data, but the SLX model shows that this cannot be found in the spatial aspect of the built environment variables. When considering these factors for the larger surrounding built environment, the variables indicating city size and density do not have a significant influence on travel time expenditures. In line with this, the results from the SFE urban model and SHG urban model do not support the concept of differences between urban and non-urban regions.

6.2. Policy implications

As presented in section 1.2, the majority of local infrastructure investments appears to have gone to the Randstad/urban regions over the past years. However, an attention shift seems to take place, allocating these investments more towards regional/local projects going forward, with the goal of improving accessibility in rural regions. In this light, the findings of this research are very interesting. The supposed difference in travel time expenditures between the Randstad and the rest of the Netherlands does show up in the data, but not necessarily in a way that one may have expected beforehand. Still, we do see significantly different travel time expenditures in the northern half of the Randstad compared to the rest of the country, just as we do in five corners of the country, located very far from the Randstad.

What is most interesting here, is that the well-connected Randstad region actually shows the highest travel time expenditures, while the supposedly badly connected outskirts of the country show the lowest travel time expenditures. From this it thus does not immediately become apparent that rural non-Randstad regions do suffer from lack of accessibility by making them spend more time on travel, which would be detrimental to their 'broad prosperity'. However, as covered in the discussion section 5.5, we cannot be sure of exactly how to interpret the found difference in travel time expenditures. From this research alone, it cannot be definitively concluded whether low travel time expenditures are an indicator of sufficient or insufficient infrastructure capacity.

Additionally, the satisfaction of the populations with their accessibility, as considered by Pot (2023), is also important to include in the interpretation. As discussed in section 1.2, perhaps worse connectivity of certain regions might not necessarily be that much of a disaster if the people living there do not have a problem with it. Possibly residential self-selection also plays a role here, if people tend to choose an area to reside in that fits with their mobility preferences. This is however also hard to distil, given the fact that true travel time budgets are hard to derive.

More generally, as we see a primary importance of socio-demographic factors, as opposed to built environment characteristics, we can question whether using the regional/geographical perspective is even the right way to consider infrastructure policy options. Alternatively, policy options could be considered that target certain socio-demographic groups instead of certain regions. By focussing on the personal characteristics instead of the spatial environment, the factors that travel behaviour appears to be rooted in are targeted more directly. For example, as this research finds a positive effect of a region's average income on its travel time expenditures, perhaps lower income groups travel less because they are not able to afford certain mobility options. If this is the case, a policy targeting travel affordability for low-income groups (e.g. through mobility subsidies) could be more effective than investing in additional infrastructure for low-income regions, as the infrastructure itself is not the bottleneck.

These results and uncertainties do have implications for the currently intended infrastructure policies. We cannot say that the lobbying of organisations like the VNG for more infrastructure investments towards local accessibility, and the resulting tendency to focus more on local infrastructure, is undoubtedly uncalled for. However, this research does cast a shadow of doubt on the true benefits that these investments would reap. It could very well be that this is the best way forward and that this will still lead to substantive prosperity gains, but these investments should first be carefully reconsidered. True welfare and travel time gains of infrastructure investments also still hinge on the applicability of the BREVER-law at the individual level across time. Depending on this, better accessibility would either lead to people having more facilities within their '(time) budget', or to people spending less time on travelling, not necessarily reaching more facilities.

There still is something to be said for rural infrastructure investment, as it remains a fact that their accessibility to facilities is lacking compared to the Randstad region, even though this does not express itself in higher travel time expenditure. Still, one could ask: if inhabitants apparently do not mind this low accessibility, what is the worth of targeting this 'problem' (if we can call it that)? Alternative view-points, like the socio-demographic perspective instead of the regional one, should be considered. For these reconsiderations, it is important to assess what the expected welfare gains and travel time expenditure changes would be for infrastructure budget allocations to different regions and/or policies focussing on socio-demographic subgroups. More elaborate research into true time budgets, travel time- and accessibility perceptions, and people's satisfaction with current and possible future infrastructure provision is necessary, before it can be adequately assessed how and where these funds are best spent.

6.3. Scientific contribution

This research contributes to the scientific knowledge on travel behaviour, and specifically travel time expenditures, in a multitude of ways. Firstly, it provides renewed insight into travel time expenditure in the Netherlands by basing the performed analyses on data from 2018 and 2019. Other identified papers on travel behaviour in the Netherlands based their findings on older (pre-2010, and often even pre-2000) data, so this research provides an update on these somewhat outdated previous studies.

This research has provides a new way of looking at geographical differences in travel behaviour, by focusing on the spatial coherence of travel time expenditures across municipalities, which has seldom been done before. By assessing measures of spatial autocorrelation (Moran's I), we have addressed whether differences in travel time expenditures have a spatial aspect to them, and it turns out they do. Values are not equally distributed across the country, but similar values tend to cluster together. We have also been able to identify larger regions with significantly higher or lower travel time expenditures. As this application allows for identifying specific regions of interest, it allows us to look deeper than, for example, a binary comparison between urban and rural regions. This method enables us to address even inter-rural differences, by considering spatial patterns in the data, instead of purely looking at the travel time expenditure values alone.

The results from this study once again confirm the primary importance of socio-demographic factors on travel behaviour, over the effect of the built environment. This is in itself not a new insight, but rather supports this notion that has been mentioned in the literature more often. Still, though disputed, some previous research did find a separate effect of urbanism on travel time expenditures, but we found that there is more to this. This relationship appears to be (partly) confounded by the influence of socio-demographic characteristics, and differences between various rural areas exist as well.

Lastly, the occurrence of lower travel time expenditure near country borders, and the possible border effect that this would suggest, has not previously been mentioned in the identified literature that chapter 2 is based on. Some research has been done on this phenomenon, but no previous research focusing on the Netherlands has been found. The supposed existence of this border effect near the country's border appears to be a determinant of travel behaviour that has not previously been considered, and this research thus provides a motive for further research into this matter.

6.4. Limitations of this research

There are some limitations that are rooted in the data sources and processing, the methods, and the scoping choices in this research that are important to be aware of. In this section we highlight these limitations and assess the possible effects they have on the outcomes of the analyses.

- The ODiN data used for this research is essentially survey data, and survey data inherently contains some unreliability. A respondent's interpretation of a question can alter the way they answer the question and thus introduce bias in the results. In behavioural research, survey data tends to come with underreporting (Yang et al., 2010). Reporting errors could very well be present in the ODiN data as well, as people are asked to recall all their movements of the day, and since the research description itself makes mention of "complex tables" in the survey (CBS, 2020b, p.16). In the extensive outlier inspection in appendix A.2 we however also see instances of reporting errors that lead to too high travel time values. Some of these errors are filtered out in the outlier detection, but we cannot be sure whether this has caught all errors. These reporting errors might therefore influence the measurements in both directions and with this the measured values become more unreliable, though the exact effect of this remains uncertain.

- A second limitation of the ODiN data is that it records the travel patterns of each individual for one single day only. If a person happens to exhibit very unusual travel behaviour on their selected day of record, not having data for multiple days eliminates the possibility of being able to correct for this. Having multi-day data would ensure that any one unusual day would be smoothed out by the data from the other days, making travel time expenditures determined by and averaged out over multi-day records more reliable. Though it is true that respondents had the opportunity to flag their entries as peculiar if they considered them to be abnormal, we cannot know for certain whether this is always properly done so. Similarly to the previous point, this limitation makes the data less reliable, but we cannot know in what way this might have impacted the results.
- Another, possibly even more vital limitation is the way that outliers were handled. As described in section 3.2.4, all individual values that fell outside 1,5 times the IQR range were excluded, though this does lead to some correct data points being thrown out with the incorrect/uncertain ones. As there were no outliers on the low side of the value distribution, due to the zero lower bound of possible values, only data of (relatively) high values were omitted. When averaging out the data per municipality, this method of data exclusion thus leads to deflated average travel time expenditures, compared to the possible real values. This effect might be stronger for municipalities that contained more extreme outliers, meaning that the total range of the average travel time values is compressed and brought closer together. As a result of this, found differences in average travel time expenditures might appear less pronounced than they truly are, which could decrease the apparent strength of the identified spatial patterns. Though we did assess how much the average travel time values were deflated (figure 3.8), no sensitivity analysis was performed that addresses the sensitivity of the spatial autocorrelation and regression analyses to this outlier detection method.
- The way that data is aggregated in this research also has some downsides to it. As mentioned earlier in section 5.5, the way that individual-level data is aggregated to municipal-level group data comes with the risk of ecological fallacy when interpreting results. In general, this way of aggregation causes data granularity and details to be compromised. The constructed municipality-wide measures of the various characteristics do not contain any information on the possible intra-group differences that exist. Any differences are more blurred/smoothed out at the group level (e.g. while individual ages range from 6 to 99+ years old, municipal age averages only range from 29 to 54 years old), which leads to a loss of information. This is the case for all included socio-demographic variables and could result in important individual-level effects no longer being detected at group levels, and thus possibly undermines the strength of any group-level relationships that we have been trying to identify.
- The construction of municipal-level socio-demographic measures based on the ODiN data, as done in this research, instead of using CBS statistics, leave these measures with more unreliability, as already reflected on in section 4.2.1. Still, by applying this method for gathering aggregated values provides measures that are more directly related to the dependent variable, hence this choice was made. This does leave us with the limitation that, though the regression results might be more accurate for this specific data set, one must be careful with generalising the results and taking them as truth for the full population. If for some municipalities the aggregated measures do not correctly represent their true demographic background, this could result in lack of generalisability of parts of the regression results.
- Similar to the previous points, but distinct in its own way, the way built environment variables are included in this research also brings its own limitations. For the non-spatial regression analyses, variables considering population and density are included as absolute values, looking for strict linear relationships between these factors and travel time expenditures. This does however not allow to test for more binary or group-based differences between, for example, urban v.s. suburban v.s. rural regions. Such a binary split is however incorporated in the SFE urban and SHG urban models, that assess differences between urban and non-urban municipalities (with a cut-off value of 1.500 addresses/km^2), though such a binary separation once again leaves us with large intra-group differences (e.g. Amsterdam and places like Hellevoetsluis are in the same urban group here, though their size and density vary massively). If there does exist an effect of the built environment on travel behaviour, using the right level of measurement and cut-off values for group construction might be paramount to finding such effects.

- Lastly, as explained in section 2.2 and also discussed earlier in the discussion in section 5.5, the data used in this research represent real travel time expenditures. Though this is the most easily measurable indicator of travel times, it does not directly consider underlying travel time budgets and relative personal travel time experiences. As long as we only have limited insight into these psychological travel time valuations, it will remain difficult to determine the exact impact of infrastructure policies (or other external developments for that matter) on travel time expenditures. Additionally, due to the aggregated nature of this research, factors like possible modal shifts/splits are not considered, though these can strongly blur the connection between measured travel time expenditures and true proximity and/or accessibility of facilities. The fact that no distinction is made between necessary and recreational travel also limits the interpretability of the results in the light of true accessibility differences between regions.

6.5. Recommendations for further research

The findings of this research provide some interesting directions for possible future research, that can build on the foundation laid by this work. These suggestions consider extensions on this work through delving even further into performed analyses, dealing with limitations, or exploring new avenues of inquiry that branch off from the results of this research.

- A first recommendation, in light of the last limitation being a crucial hurdle in travel behaviour research in general, would be to extend the use of travel time expenditures by incorporating some sort of valuation element regarding the time one spends on travel, in a way like Pot (2023) has done in his research. Knowing whether people judge their time expenditures as being satisfactory might help in estimating how they would react to possibilities to decrease these expenditures, and in addressing whether or not certain policies aimed at infrastructural improvement would increase their personal utility. More insights into differences in travel for varying purposes (necessary/recreational) would also be beneficial here, for understanding what currently motivates people to travel. The ODiN data also contains information on this and could thus be utilised further to this end.
- The recommendation above can be generalised some more, as there are various aspects of travel behaviour that we have not considered in this research, that could provide valuable insights. This research has only focused on the total travel time expenditure (the RE part of BREVER), but plenty other information is relevant for understanding the full picture of one's travel patterns and behaviour. Not only travel time, but also distances covered and number of trips (the VER part of BREVER) can be relevant, just as the distribution of travel time expenditures across travel modes, given each mode has a different speed. Taking these factors into account, as additional dependent variables, in further research can provide more detailed insight into differences in travel behaviour that do not become apparent from only considering total time expenditure.
- As the used ODiN data in its full capacity is so broad, detailed, and abundant, they also allow for plenty other additional analyses, of which we mention two especially interesting possibilities here. Firstly, in larger cities like Amsterdam, city districts might differ a lot from each other in various aspect. As ODiN data contains postal code information, analyses could also be performed at a neighbourhood level to detect more detailed intra-city differences, e.g. between city centres and outskirts. Secondly, as lower car ownership numbers were found in municipalities with high travel time expenditures, it could be interesting to see if relative expenditures with different travel modes also differ between regions. This would allow us to account for possible modal shift effects that pose a limitation to this research. The ODiN data does allow for this, as they also contain the travel modes of each trip. It would be interesting to see if using (not just owning) faster modes of transport (like cars) is also associated with lower travel time expenditures. More generally, the inclusion of various other predictive variables in regression models might add additional explanations for travel time expenditures, given the moderate R^2 of the models constructed in this research.
- More generally, as discussed in section 4.2.4, not all relationships that we mentioned in the literature (and included in figure 2.3) were actually tested in this research. Additional analyses could help complete the picture on how all (groups of) factors actually relate to each other. To gain more insight into the (now assumed) confounding effect of socio-demographic factors on the relationship between urban factors and travel time expenditures, the direct influence of socio-

demographic factors on one's urban environment could be more extensively researched, to perhaps more definitively find proof of residential self-selection. Additionally, the by the literature suggested, but not tested, existence of an interaction effect of personal characteristics on the urban influence could be further expanded on by using similar analyses that were used in this research.

- The absence of a sensitivity analyses also leaves room for improvement in the quality of this research, regarding multiple aspects of the methods used. Considering the effect of the outliers on the spatial autocorrelation and regression results, we currently do not know what the impact of removing these (partly correct) entries from the data has been on the results. Performing sensitivity analyses on this outlier removal method would be valuable, to test whether the results are greatly affected by this, or whether the results would still be valid without removing these outliers, or when using another (more accurate) outlier detection method. Besides sensitivity for the outlier exclusion, sensitivity to the used spatial weights (section 4.1.1) is also unknown. While the choice for binary Rook weights has been carefully considered, it could very well be that (slightly) different spatial weight assignments could impact the outcomes of the spatial autocorrelation analysis, as well as the SL and SLX regression models. Applying various weight systems could help assess the robustness of the results.
- As discussed in section 5.5, some of the constructed regression models suffer from heteroscedasticity, which makes some of the estimated models and their coefficients less reliable. As this issue is not sharply dealt with in this research, the quality of the performed analyses could be further improved by applying methods for dealing with heteroscedasticity properly. A mention of using heteroscedasticity-robust standard errors was made, but more sophisticated methods for handling this issue exist. This can be done by redefining the independent variables that have a large ranges (e.g. through log transformation) or by performing a weighted regression, where observations with a high variance have less influence on the outcome of the model (Frost, n.d.). Methods like these mitigate the negative effect heteroscedasticity can have on the regression results and therefore make the results more reliable.
- Another recommendation is to perform more extensive research into travel time expenditures at more granular and detailed levels. The spatial regression methods used in this research, that were based on Rey et al. (2023, Ch.11) do also allow for analysis at the individual level. By extending further into this level of analysis, the problem of ecological fallacy can be mitigated. It does also allow for incorporating the additional individual level information that was lost in this research, as described in the limitations. By refocusing on data at the individual level, influential factors for determining travel behaviour might be even more accurately identified. This individual deepening can be performed using the same ODiN data that was used for this research. Additionally, performing these analyses at more disaggregated (neighbourhood/individual) levels might also provide us with more insight into if the found results still hold at different levels of aggregation. This can create more clarity on the possibility of there being a Modifiable Areal Unit Problem (MAUP), a concept discussed in section 2.5.
- The BREVER-law and the concept of constant travel times supposedly also applies over time, so it would be interesting to see if the outcomes of the performed analyses on the 2018/2019 data will also be found for different years. The ODiN project is an ongoing, yearly performed, data collection process that thus supplies similar datasets for following years as well. Especially with the developments caused by the COVID-19 pandemic, like the rise of working-from-home, one may ask whether travel patterns have significantly been altered in the past few years. Performing similar analyses for the during- and post-pandemic ODiN data would give insights into the effects of the pandemic on travel behaviour and would help paint a more accurate picture of current behaviour.
- As the found lower travel time expenditures near the border were not expected beforehand, it would be interesting to investigate this phenomenon further. To determine whether we can actually speak of a border effect, more detailed insight is needed into the exact trips people living near a country border tend to make. Are they really mainly travelling country-inwards and do they avoid crossing the border? Carefully tracking their behaviour would help answer this question. Additionally, it would be interesting to see if this effect also appears on the other side of

the border. Would we find similar lower travel time expenditures near these borders in Belgium and Germany, compared to their national averages respectively? If results would indeed substantiate the existence of a border effect, this would birth the relevance for studying how people (psychologically) experience this and if there would be ways to take away (some of) this travel limitation.

- Lastly, it could be interesting to investigate further whether a 'Randstad-effect' might exist regarding travel time expenditure. The results showed that there is no clear effect of urbanism on travel time expenditures, but still we do see higher travel times in the (northern) Randstad region, compared to the rest of the country. While we have tested for an effect of municipality size and density, we have not tested for a possible effect of a municipality being in the Randstad region (regardless of its size/density). Perhaps the differences in travel time expenditure are not due to the urbanism level, but rather due to a municipality being within the Randstad or not. This could be tested in a similar way that we approached the border difference in this research, by constructing a spatial fixed effects or spatial heterogeneity model. For this however, one would have to construct clear boundaries regarding what municipalities are (not) considered part of the Randstad region, to allow for binary comparison between the Randstad and the rest of the country.

References

Ahmed, A., & Stopher, P. (2014). Seventy Minutes Plus or Minus 10 — A Review of Travel Time Budget Studies. *Transport Reviews*, *34*(5), 607–625. https://doi.org/10.1080/01441647.2014.946460

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, *7*(1), 7. https://doi.org/10.1186/1476-072X-7-7

Barnes, G., & Davis, G. (2001, July). *Land Use and Travel Choices in the Twin Cities, 1958-1990* (tech. rep.). University of Minnesota, Center for Transportation Studies. Minneapolis, MN. https://hdl.handle.net/11299/660

Bastiaansen, J., & Breedijk, M. (2022, October). *TOEGANG VOOR IEDEREEN? Een analyse van de (on)bereikbaarheid van voorzieningen en banen in Nederland* (tech. rep.). PBL. Den Haag. https://www.pbl.nl/uploads/default/downloads/pbl-2022-toegang-voor-iedereen-4932.pdf

Bhat, C., & Zhao, H. (2002). The spatial analysis of activity stop generation. *Transportation Research Part B: Methodological*, *36*(6), 557–575. https://doi.org/10.1016/S0191-2615(01)00019-4

Borkowski, P., Jażdżewska-Gutta, M., & Szmelter-Jarosz, A. (2021). Lockdowned: Everyday mobility changes in response to COVID-19. *Journal of Transport Geography*, *90*, 102906. https://doi.org/10.1016/j.jtrangeo.2020.102906

Brazil, N. (2023, February). Lab 7: Spatial Autocorrelation. https://crd230.github.io/lab7.html

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287. https://doi.org/10.2307/1911963

Cats, O. (2023). Identifying human mobility patterns using smart card data. *Transport Reviews*, *44*(1), 213–243. https://doi.org/10.1080/01441647.2023.2251688

CBS. (n.d.-a). Dutch National Travel survey. https://www.cbs.nl/en-gb/our-services/methods/surveys/brief-survey-description/dutch-national-travel-survey

CBS. (n.d.-b). Gemeentelijke indeling op 1 januari 2019. https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-jaar/indeling-per-jaar/gemeentelijke-indeling-op-1-januari-2019

CBS. (2019). *Onderweg in Nederland (ODiN) 2018, Onderzoeksbeschrijving* (tech. rep.). 12. Den Haag. https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksomschrijvingen/onderweg-in-nederland--odin---eindrapportage-2018

CBS. (2020a, April). Opleidingsniveau naar gemeenten, wijken en buurten [Dataset]. https://www.cbs.nl/nl-nl/maatwerk/2020/17/opleidingsniveau-naar-gemeenten-wijken-en-buurten

CBS. (2020b, July). *Onderweg in Nederland (ODiN) 2019, Onderzoeksbeschrijving* (tech. rep.). Den Haag. https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksomschrijvingen/onderweg-in-nederland--odin---onderzoeksbeschrijving-2019

CBS. (2020c, July). *Onderweg in Nederland (ODiN) 2019, Plausibiliteitsrapportage* (tech. rep.). CBS. Den Haag. https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksomschrijvingen/onderweg-in-nederland--odin---plausibiliteitsrapportage-2019

CBS. (2021, October). Toelichting Wijk- en Buurtkaart, 2019, 2020 en 2021. https://www.cbs.nl/nl-nl/longread/diversen/2021/toelichting-wijk-en-buurtkaart-2019-2020-en-2021?onepage=true

CBS. (2023a, March). Provinciale uitgaven per inwoner hoog in het Noorden. https://www.cbs.nl/nl-nl/nieuws/2023/09/provinciale-uitgaven-per-inwoner-hoog-in-het-noorden

CBS. (2023b, September). Inkomen per gemeente en wijk, 2020 [Dataset]. https://www.cbs.nl/nl-nl/maatwerk/2023/35/inkomen-per-gemeente-en-wijk-2020

CBS. (2024, March). Gemeentebegrotingen; baten en lasten naar regio en grootteklasse. https://opendata.cbs.nl/#/CBS/nl/dataset/83641NED/table

Chen, C., & Mokhtarian, P. (2008). *A Review and Discussion of the Literature on Travel Time and Money Expenditures* (tech. rep.). https://escholarship.org/uc/item/51d696jh

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, *28*(3), 591–605. https://doi.org/10.2307/1910133

Condeço-Melhorado, A., Tillema, T., de Jong, T., & Koopal, R. (2014). Distributive effects of new highway infrastructure in the Netherlands: the role of network effects and spatial spillovers. *Journal of Transport Geography*, *34*, 96–105. https://doi.org/10.1016/j.jtrangeo.2013.11.006

CPB & PBL. (2020). *Kansrijk mobiliteitsbeleid 2020* (tech. rep.). Den Haag. https://www.cpb.nl/sites/default/files/omnidownload/Kansrijk-mobiliteitsbeleid-2020.pdf

CROW. (n.d.). 1. Wat is bereikbaarheid. https://www.crow.nl/duurzame-mobiliteit/home/samenhang-1

de Vos, D., Meijers, E., & van Ham, M. (2018). Working from home and the willingness to accept a longer commute. *Annals of Regional Science*, *61*(2), 375–398. https://doi.org/10.1007/s00168-018-0873-6

Dictionary.com. (n.d.). Budget. http://dictionary.reference.com/browse/budget

Downes, J., & Morrell, D. (1981). Variation of travel time budgets and trip rates in reading. *Transportation Research Part A: General*, *15*(1), 47–53. https://doi.org/10.1016/0191-2607(83)90015-8

Ecorys & VanBerkel Professionals. (2019, December). *De financiering van lokale verkeer en vervoertaken door provincies en vervoerregio's* (tech. rep.). Rotterdam. https://zoek.officielebekendmakingen.nl/blg-916941.pdf

Eerste Kamer der Staten-Generaal. (2024, April). Debat samengevat: begroting Infrastructuur en Waterstaat 2024. https://www.eerstekamer.nl/nieuws/20240402/debat_samengevat_begroting

Faber, R., Hamersma, M., Brimaire, J., Kroesen, M., & Molin, E. J. (2023). The relations between working from home and travel behaviour: a panel analysis. *Transportation*. https://doi.org/10.1007/s11116-023-10401-4

Faber, R., Hamersma, M., de Haas, M., Krabbenborg, L., & Hoen, A. (2023). Estimating post-pandemic effects of working from home and teleconferencing on travel behaviour. *European Journal of Transport and Infrastructure Research*, *23*(1), 33–62. https://doi.org/10.18757/ejtir.2023.23.1.6733

Feng, J., Dijst, M., Prillwitz, J., & Wissink, B. (2013). Travel Time and Distance in International Perspective: A Comparison between Nanjing (China) and the Randstad (The Netherlands). *Urban Studies*, *50*(14), 2993–3010. https://doi.org/10.1177/0042098013482504

Firebaugh, G. (2001). Ecological Fallacy, Statistics of. In *International encyclopedia of the social & behavioral sciences* (pp. 4023–4026). Elsevier. https://doi.org/10.1016/B0-08-043076-7/00765-8

Freepik. (n.d.). Mensen in de metro [image]. https://nl.freepik.com/vrije-vector/mensen-in-de-metro_4543422.htm#fromView=search&page=1&position=34&uuid=76f45ce4-7fab-461f-8028-4574bd9878d9

Frost, J. (n.d.). Heteroscedasticity in Regression Analysis. https://statisticsbyjim.com/regression/heteroscedasticity-regression/

Gallotti, R., Bazzani, A., & Rambaldi, S. (2015). Understanding the variability of daily travel-time expenditures using GPS trajectory data. *EPJ Data Science*, *4*(1), 1–14. https://doi.org/10.1140/epjds/s13688-015-0055-z

Gordon, P., Kumar, A., & Richardson, H. W. (1989). The Influence of Metropolitan Spatial Structure on Commuting Time. *Journal of Urban Economics*, *26*, 138–151. https://doi.org/10.1016/0094-1190(89)90013-2

Guan, X., Wang, D., & Jason Cao, X. (2020). The role of residential self-selection in land use-travel research: a review of recent findings. *Transport Reviews*, *40*(3), 267–287. https://doi.org/10.1080/01441647.2019.1692965

Gunn, H. F. (1979). *Travel Budgets – A Review of Evidence and Modelling Implications*, Institute of Transport Studies, University of Leeds. https://eprints.whiterose.ac.uk/2406/

Hong, J., Shen, Q., & Zhang, L. (2014). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation*, *41*(3), 419–440. https://doi.org/10.1007/s11116-013-9462-9

Hook, H., De Vos, J., Van Acker, V., & Witlox, F. (2023). A comparative analysis of determinants, characteristics, and experiences of four daily trip types. *Travel Behaviour and Society*, *30*, 335–343. https://doi.org/10.1016/j.tbs.2022.10.013

Hupkes, G. (1977). *Gasgeven of afremmen*. Kluwer B.V.

Hupkes, G. (1982). The law of constant travel time and trip-rates. *Futures*, *14*(1), 38–46. https://doi.org/10.1016/0016-3287(82)90070-2

IBM. (2024, February). Two-Stage Least-Squares Regression. https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=regression-two-stage-least-squares

Ivanchev, J., Aydt, H., & Knoll, A. (2015). Spatial and Temporal Analysis of Mismatch between Planned Road Infrastructure and Traffic Demand in Large Cities. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2015-October*, 1463–1470. https://doi.org/10.1109/ITSC.2015.239

Jahanshahi, K., Jin, Y., & Williams, I. (2013). ANALYSING SOURCES OF VARIABILITY IN TRAVEL TIME USE IN A COMBINED FRAMEWORK USING EXTENDED STRUCTURAL EQUATION MODELS AND THE UK NATIONAL TRAVEL SURVEY DATA. *Transportation Research Board*. https://www.researchgate.net/publication/281464616

Jin, M., Gong, L., Cao, Y., Zhang, P., Gong, Y., & Liu, Y. (2021). Identifying borders of activity spaces and quantifying border effects on intra-urban travel through spatial interaction network. *Computers, Environment and Urban Systems*, *87*, 101625. https://doi.org/10.1016/j.compenvurbsys.2021.101625

Jorritsma, P., Jonkeren, O., & Krabbenborg, L. (2023, April). *De ontwikkeling van de mobiliteit en bereikbaarheid in stedelijk en ruraal Nederland* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid. Den Haag. https://www.kimnet.nl/publicaties/publicaties/2023/04/03/de-ontwikkeling-van-de-mobiliteit-en-bereikbaarheid-in-stedelijk-en-ruraal-nederland

Kabinetsformatie. (2024, May). Hoofdlijnenakkoord tussen de fracties van PVV, VVD, NSC en BBB. https://www.kabinetsformatie2023.nl/documenten/publicaties/2024/05/16/hoofdlijnenakkoord-tussen-de-fracties-van-pvv-vvd-nsc-en-bbb

Kamerstuk 36410-XII. (2023, September). Vaststelling van de begrotingsstaten van het Ministerie van Infrastructuur en Waterstaat voor het jaar 2024.

KiM. (2022, November). *Kerncijfers Mobiliteit 2022* (tech. rep.). Den Haag. https://www.kimnet.nl/publicaties/publicaties/2022/11/15/kerncijfers-mobiliteit-2022

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. https://doi.org/10.4097/kja.19087

Koo, J., Kim, J., Choi, S., & Choo, S. (2022). Identifying the Causal Relationship between Travel and Activity Times: A Structural Equation Modeling Approach. *Sustainability*, *14*(8), 4615. https://doi.org/10.3390/su14084615

Kun, A., Sadun, R., Shaer, O., & Teodorovicz, T. (2020). Where Did the Commute Time Go? *Harvard Business Review*. https://hbr.org/2020/12/where-did-the-commute-time-go

Landrock, J. N. (1981). SPATIAL STABILITY OF AVERAGE DAILY TRAVEL TIMES AND TRIP RATES WITHIN GREAT BRITAIN. *Transportation Research Part A: General*, *15*(1), 55–62.

Lee, S., Smart, M. J., & Golub, A. (2021). Difference in travel behavior between immigrants in the u.s. and us born residents: The immigrant effect for car-sharing, ride-sharing, and bike-sharing services. *Transportation Research Interdisciplinary Perspectives*, *9*, 100296. https://doi.org/10.1016/j.trip.2020.100296

Ma, Q., Liu, A., Chen, Y., & Tao, R. (2024). Border effects for domestic travel in China during COVID-19 pandemic. *Journal of Transport Geography*, *116*, 103857. https://doi.org/10.1016/j.jtrangeo.2024.103857

MacDonald, C. (2015, November). What travel looked like 100 years ago: Map shows how many DAYS it took to travel to the furthest corners of Britain's Empire in 1914. https://www.dailymail.co.uk/sciencetech/article-3339902/What-travel-looked-like-100-years-ago-Map-shows-DAYS-took-travel-abroad-1900s.html

Marchetti, C. (1994). *ANTHROPOLOGICAL INVARIANTS IN TRAVEL BEHAVIOR* (tech. rep. No. 7).

Martín-Barroso, D., Núñez-Serrano, J. A., Turrión, J., & Velázquez, F. J. (2022). Are workers' commutes sensitive to changes in the labour market situation? *Journal of Transport Geography*, *101*, 103352.

Merriam-Webster. (n.d.). Budget. https://www.merriam-webster.com/dictionary/budget

Metz, D. (2008). The Myth of Travel Time Saving. *Transport Reviews*, *28*(3), 321–336. https://doi.org/10.1080/01441640701642348

Metz, D. (2021). Economic benefits of road widening: Discrepancy between outturn and forecast. *Transportation Research Part A: Policy and Practice*, *147*, 312–319. https://doi.org/10.1016/j.tra.2021.03.023

Ministerie van Financiën. (n.d.). Jaarverslag. https://www.rijksfinancien.nl/visuals/2023/jaarverslag/uitgaven/incl-premies?graph=pie

Ministerie van Infrastructuur & Waterstaat. (2023, March). *Mobiliteitsvisie 2050 Hoofdlijnennotitie* (tech. rep.). Den Haag. https://www.rijksoverheid.nl/documenten/rapporten/2023/03/17/bijlage-hoofdlijnennotitie-mobiliteitsvisie-2050

Mokhtarian, P. L., & Chen, C. (2004). TTB or not TTB, that is the question: A review and analysis of the empirical literature on travel time (and money) budgets. *Transportation Research Part A: Policy and Practice*, *38*(9-10), 643–675. https://doi.org/10.1016/j.tra.2003.12.004

Moraga, P. (2024, December). *Spatial Statistics for Data Science: Theory and Practice with R* (1st). Chapman & Hall.

Mustafa, A. (2024, March). Understanding Endogeneity in regression analysis. https://medium.com/@akif.iips/understanding-endogeneity-in-regression-analysis-e53a13dfb16d

NU.nl. (2022, November). Kabinet investeert miljarden in infrastructuur, vooral veel geld naar het ov. https://www.nu.nl/economie/6236094/kabinet-investeert-miljarden-in-infrastructuur-vooral-veel-geld-naar-het-ov.html

O'Toole, K. (2013, March). The Big Trend: The World is Getting Wealthier. https://www.gsb.stanford.edu/insights/big-trend-world-getting-wealthier

Oxford English Dictionary. (n.d.). Budget. https://www.oed.com/search/dictionary/?scope=Entries&q=budget

Pereira, R. H. M. (2019). Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro Ex-ante evaluation of the accessibility impacts of transport policy scenarios: equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro. *Journal of Transport Geography*, *74*, 321–332. https://doi.org/10.1016/j.jtrangeo.2018.12.005

Pot, F. J. (2023, September). *The Extra Mile: Perceived accessibility in rural areas* [Doctoral dissertation, University of Groningen]. https://doi.org/10.33612/diss.737899635

Raux, C., Ma, T.-Y., Joly, I., & Cornelis, E. (2011). Daily and weekly time allocation to travel and activity in some European cities. *Conference on the Economics of the Family*. https://www.researchgate.net/publication/262840759_Daily_and_weekly_time_allocation_to_travel_and_activity_in_some_European_cities

Raux, C., Ma, T.-Y., Joly, I., Kaufmann, V., Cornelis, E., & Ovtracht, N. (2011). Travel and activity time allocation: An empirical comparison between eight cities in Europe. *Transport Policy*, *18*(2), 401–412. https://doi.org/10.1016/j.tranpol.2010.11.004

Rey, S., Arribas-Bel, D., & Wolf, L. J. (2023, June). *Geographic Data Science with Python* (1st). Chapman & Hall. https://geographicdata.science/book/intro.html#

Rienstra, S. (2022, January). *Nederlandse overheidsuitgaven en -inkomsten verkeer en vervoer* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid. Den Haag. https://www.kimnet.nl/publicaties/rapporten/2022/01/17/nederlandse-overheidsuitgaven-en--inkomsten-verkeer-en-vervoer

Rijkswaterstaat. (2017, December). *Het Landelijk Model Systeem* (tech. rep.). https://open.rijkswaterstaat.nl/publish/pages/34566/verkeer_en_vervoer_het_landelijk_model_sysyteem.pdf

Rodrigues, D., & Targa, F. (2004). Value of accessibility to Bogotá's bus rapid transit system. *Transport Reviews*, *24*(5), 587–610. https://doi.org/10.1080/0144164042000195081

Schafer, A., & Victor, D. G. (2000). The future mobility of the world population. *Transportation Research Part A: Policy and Practice*, *34*(3), 171–205. https://doi.org/10.1016/S0965-8564(98)00071-8

Scheiner, J. (2010). Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002. *Journal of Transport Geography*, *18*(1), 75–84. https://doi.org/10.1016/j.jtrangeo.2009.01.001

Shekhar, S., Evans, M. R., Kang, J. M., & Mohan, P. (2011). Identifying patterns in spatial information: A survey of methods. *WIREs Data Mining and Knowledge Discovery*, *1*(3), 193–214. https://doi.org/10.1002/widm.25

Song, Y., Lee, S., Park, A. H., & Lee, C. (2023). COVID-19 impacts on non-work travel patterns: A place-based investigation using smartphone mobility data. *Environment and Planning B: Urban Analytics and City Science*, *50*(1), 642–659. https://doi.org/10.1177/23998083221124930 spreg.GM_Lag. (n.d.). https://spreg.readthedocs.io/en/latest/generated/spreg.GM_Lag.html

Stopher, P., Ahmed, A., & Liu, W. (2017). Travel time budgets: new evidence from multi-year, multi-day data. *Transportation*, *44*(5), 1069–1082. https://doi.org/10.1007/s11116-016-9694-6

Stopher, P., & Zhang, Y. (2011, February). *Travel time expenditures and travel time budgets - Preliminary findings*, INSTITUTE of TRANSPORT and LOGISTICS STUDIES. https://ses.library.usyd.edu.au/bitstream/handle/2123/19360/ITLS-WP-11-04.pdf?sequence=1&isAllowed=y

Szalai, A. (1966a). Trends in Comparative Time-Budget Research. *American Behavioral Scientist*, *9*(9), 3–8. https://doi.org/10.1177/000276426600900901

Szalai, A. (1966b). The Multinational Comparative Time Budget Research Project. *American Behavioral Scientist*, *10*(4), 1–31. https://doi.org/10.1177/000276426601000401

Szalai, A. (1975). Women's time. *Futures*, *7*(5), 385–399. https://doi.org/10.1016/0016-3287(75)90017-8

Tan, P. Y., & Samsudin, R. (2017). Effects of spatial scale on assessment of spatial equity of urban park provision. *Landscape and Urban Planning*, *158*, 139–154. https://doi.org/10.1016/j.landurbplan.2016.11.001

Timmermans, H., Van Der Waerden, P., Alves, M., Polak, J., Ellis, S., Harvey, A. S., Kurose, S., & Zandee, R. (2003). Spatial context and the complexity of daily travel patterns: an international comparison. *Journal of Transport Geography*, *11*, 37–46. www.elsevier.com/locate/jtrangeo

Timmermans, H., van der Waerden, P., Alves, M., Polak, J., Ellis, S., Harvey, A. S., Kurose, S., & Zandee, R. (2002). Time allocation in urban and transport settings: an international, inter-urban perspective. *Transport Policy*, *9*(2), 79–93. https://doi.org/10.1016/S0967-070X(02)00003-3

Tobler, W. R. (1970). A COMPUTER MOVIE SIMULATING URBAN GROWTH IN THE DETROIT REGION. *Economic Geography*, *46*, 234–240. https://doi.org/10.2307/143141

United Nations. (2015, October). Transforming our world: the 2030 Agenda for Sustainable Development. https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf

van Wee, B. (2007). Large infrastructure projects: A review of the quality of demand forecasts and cost estimations. *Environment and Planning B: Planning and Design*, *34*(4), 611–625. https://doi.org/10.1068/b32110

van Wee, B., Rietveld, P., & Meurs, H. (2006). Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time. *Journal of Transport Geography*, *14*(2), 109–122. https://doi.org/10.1016/j.jtrangeo.2005.06.003

van der Hoorn, T. (1979). Travel behaviour and the total activity pattern. *Transportation*, *8*(4), 309–328.

VNG. (2024, January). Opmerkingen van de VNG op de begroting van het Ministerie van Infrastructuur en Waterstaat. https://vng.nl/sites/default/files/2024-01/20240119-brief-parlement-vng-positionpaper-begroting-min-ienw.pdf

WadgidsenWeb. (2024, January). Veerdiensten Wadden NL, BRD en DK 2024. https://www.wadgidsenweb.nl/veerdiensten-wadden.html

Walls, M., & Safirova, E. (2004, December). *A Review of the Literature on Telecommuting and Its Implications for Vehicle Travel and Emissions* (tech. rep.). Resources for the Future. Washington, D.C. https://ageconsearch.umn.edu/record/10492/

Wang, C. H., Chen, N., & Tian, G. (2021). Do accessibility and clustering affect active travel behavior in Salt Lake City? *Transportation Research Part D: Transport and Environment*, *90*. https://doi.org/10.1016/j.trd.2020.102655

Wang, F. (2012). Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Annals of the Association of American Geographers*, *102*(5), 1104–1112. https://doi.org/10.1080/00045608.2012.657146

Weening, M. (2020). The Re-Ruralization of the Economy: Covid-19 and Work-from-Anywhere [Post]. *LinkedIn*. https://www.linkedin.com/pulse/re-ruralization-economy-covid-19-work-from-anywhere-michael-weening/

Whitehead, J., L. Pearson, A., Lawrenson, R., & Atatoa-Carr, P. (2019). How can the spatial equity of health services be defined and measured? A systematic review of spatial equity definitions and methods. *Journal of Health Services Research & Policy*, *24*(4), 270–278. https://doi.org/10.1177/1355819619837292

Wigan, M., & Morris, J. (1981). The transport implications of activity and time budget constraints. *Transportation Research Part A: General*, *15*(1), 63–86. https://doi.org/10.1016/0191-2607(83)90017-1

Wismadi, A., Zuidgeest, M., Brussel, M., & van Maarseveen, M. (2014). Spatial Preference Modelling for equitable infrastructure provision: An application of Sen's Capability Approach. *Journal of Geographical Systems*, *16*(1), 19–48. https://doi.org/10.1007/s10109-013-0185-4

Yang, S., Zhao, Y., & Dhar, R. (2010). Modeling the Underreporting Bias in Panel Survey Data. *Marketing Science*, *29*(3), 525–539. https://doi.org/10.1287/mksc.1090.0536

Zahavi, Y. (1974, May). *Traveltime budgetes and mobility in urban areas* (tech. rep.). U.S. Department of Transportation. Washington D.C.

Zhou, M., Wang, D., & Guan, X. (2022). Co-evolution of the built environment and travel behaviour in Shenzhen, China. *Transportation Research Part D: Transport and Environment*, *107*. https://doi.org/10.1016/j.trd.2022.103291

A

Extensive outlier analysis

A.1. Value distributions of outliers

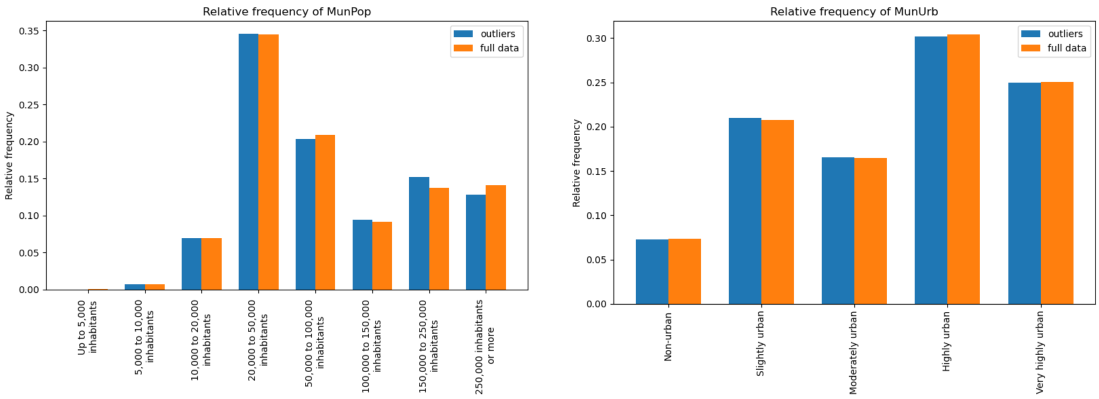


Figure A.1: Relative distributions of built environment variables in outliers and full ODiN data

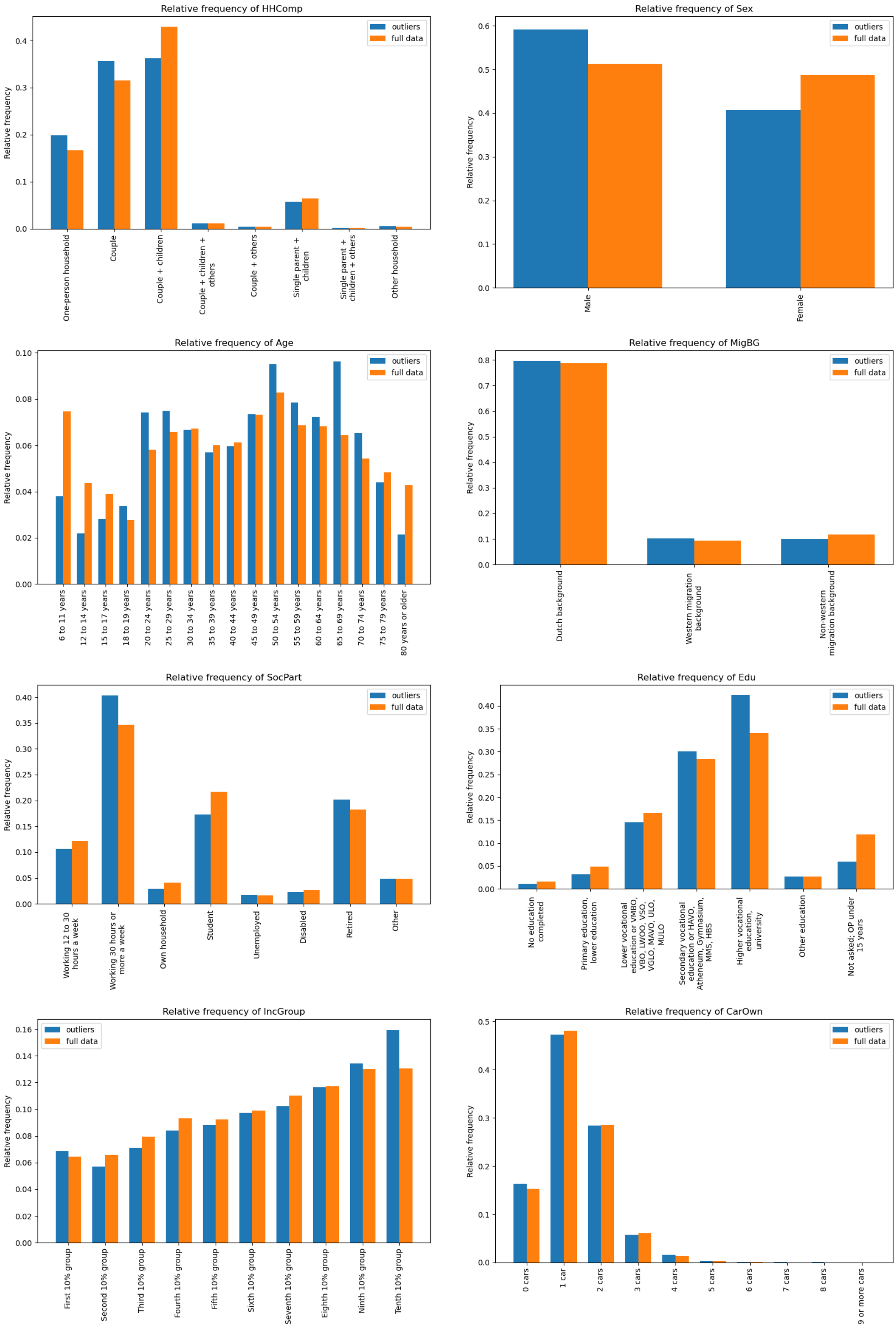


Figure A.2: Relative distributions of socio-demographic variables in outliers and full ODiN data

A.2. Deep-dive into specific outliers

In this part of the appendix, the full socio-demographic and travel profile of several individual outliers are more extensively looked at. The individuals' travel times and other potentially interesting characteristics of travel behaviours are shown in bullet points, with a verdict on the plausibility of their records being true/reliable entries.

10 MOST EXTREME OUTLIERS

OPID = 59275451478

- PecDay = Unknown
- TotTT = 1200 min (20h)
- 1 trip: Going home from work

VERDICT: Likely wrongly entered trip(s)

OPID = 56175178523

- PecDay = Yes
- TotTT = 920 min (15:20h)
- Multiple legs: public transport and by foot. Probably day of hiking
- Only covered 9,7 km in 9h of walking (very slow). Probably did not walk for 9h straight.

VERDICT: Implausible

OPID = 55868918828

- PecDay = No, this day of the week is always different
- TotTT = 905 min (15:05h)
- Covered 1100km in 15:05 by car
- Started in France. Perhaps returned from holiday

VERDICT: Plausible

OPID = 59089154312

- PecDay = No, this day of the week is always different
- TotTT = 880 min (14:40h)
- Covered 1200km in 14:40 by car
- Started in France. Perhaps returned from holiday

VERDICT: Plausible

OPID = 59058243127

- PecDay = Yes
- TotTT = 875 min (14:35h)
- Also flagged as peculiar in duration
- Transporting goods

VERDICT: Unsure, but at least not a normal day

<u>OPID = 55958214483</u>

- PecDay = No, no specific peculiarities
- TotTT = 870 min (14:30h)
- One trip: Touring on non-electric bike
- Only covered 60km in 14:30h of cycling (very slow). Probably did not cycle for 14:30h straight.

VERDICT: Implausible

<u>OPID = 56023475218</u>

- PecDay = Yes
- TotTT = 855 min (14:15h)
- Only covered 147km in 11h by car (very slow). Trip purpose was to visit someone.
- Possibly recorded leaving home -> arriving back home as one trip.

VERDICT: Likely wrongly entered trip(s)

<u>OPID = 59151190644</u>

- PecDay = No, no specific peculiarities
- TotTT = 825 min (13:45h)
- Covered 4 km by bike in 10:45h (very slow), going to work and home again.
- Possibly recorded leaving home -> arriving back home as one trip.

VERDICT: Likely wrongly entered trip(s)

<u>OPID = 58968570357</u>

- PecDay = Yes
- TotTT = 825 min (13:45h)
- Covered 16km by train in 12:15h, going to work and home again.
- Possibly recorded leaving home -> arriving back home as one trip.

VERDICT: Likely wrongly entered trip(s)

<u>OPID = 59161052240</u>

- PecDay = No, this day of the week is always different
- TotTT = 820 min (13:40h)
- Took 13:40h to get from The Hague to Zwolle (149km) by car to visit someone
- Possibly wrongly entered leaving time, now set at 1:30 am.

VERDICT: Likely wrongly entered trip(s)

10 LEAST EXTREME OUTLIERS

OPID = 58965201618

- PecDay = No, this day of the week is always different
- TotTT = 185 min (3:05h)
- Nothing special

VERDICT: Very plausible

OPID = 55837896136

- PecDay = Yes
- TotTT = 185 min (3:05h)
- Peculiarity: Different route(s) due to the weather

VERDICT: Very plausible

OPID = 56115100926

- PecDay = Yes
- TotTT = 182 min (3:02h)
- Peculiarity: Different address(es).

VERDICT: Very plausible

OPID = 55838187908

- PecDay = No, this day of the week is always different
- TotTT = 178 min (2:58h)
- Nothing special

VERDICT: Very plausible

OPID = 58934351720

- PecDay = No, this day of the week is always different
- TotTT = 174 min (2:54h)
- Covered 25km by car in 2,5h, going shopping
- Possibly recorded leaving home -> arriving back home as one trip.

VERDICT: Likely wrongly entered trip(s)

OPID = 55927425407

- PecDay = Yes
- TotTT = 172 min (2:52h)
- Peculiarity: Day off
- Nothing special

VERDICT: Very plausible

OPID = 56082252861

- PecDay = No, this day of the week is always different
- TotTT = 165 min (2:45h)
- Nothing special

VERDICT: Very plausible

OPID = 59213216002

- PecDay = Yes
- TotTT = 162 min (2:42h)
- Peculiarity: Different address(es) due to appointment(s)

VERDICT: Very plausible

OPID = 59182215940

- PecDay = Yes
- TotTT = 155 min (2:35h)
- Peculiarity: Different address(es) due to appointment(s)

VERDICT: Very plausible

OPID = 55989246568

- PecDay = Yes
- TotTT = 150 min (2:30h)
- Peculiarity: Different route(s)
- Cycled 54km in 2,5h

VERDICT: Very plausible

10 RANDOMLY SELECTED OUTLIERS

OPID = 55961500948

- PecDay = No, this day of the week is always different
- TotTT = 480 min (8:00h)
- Toured 173km on a motor in 8h (fairly slow)
- Possibly did not record stops during the day

VERDICT: Unsure

OPID = 55896213894

- PecDay = No, no specific peculiarities
- TotTT = 314 min (5:14h)
- Nothing special

VERDICT: Very plausible

OPID = 59129961596

- PecDay = No, this day of the week is always different
- TotTT = 330 min (6:30h)
- Cycled 10km in 2h by bike (very slow) to go to the store
- Walked 5km in 2,5h of walking (fairly slow)

VERDICT: Not very plausible

OPID = 56115634684

- PecDay = No, no specific peculiarities
- TotTT = 264 min (4:24h)
- Cycled 17km in 3:24h (slow) while touring
- Possibly did not record intermediate stops

VERDICT: Unsure

OPID = 55927143214

- PecDay = Yes
- TotTT = 383 min (6:23h)
- Peculiarity: Different address(es) due to day off
- Cycled 31,8km in 6:22h (slow)
- Possibly did not record intermediate stops

VERDICT: Unsure

OPID = 56115120549

- PecDay = Yes
- TotTT = 240 min (4:00h)
- Peculiarity: Different address(es) due to day off
- Visited someone in a different province

VERDICT: Plausible

OPID = 55868209796

- PecDay = Yes
- TotTT = 240 min (4:00h)
- Peculiarity: Different route(s) due to day off
- Covered 140km in 4h by car, while touring

VERDICT: Plausible

OPID = 56020126319

- PecDay = No, no specific peculiarities
- TotTT = 350 min (5:50h)
- Drove from Emmen to Zandvoort (200km) in about 3h, and back, for recreational activity

VERDICT: Plausible, but likely not something one would do every week

OPID = 59089128644

- PecDay = No,this day of the week is always differnt
- TotTT = 265 min (4:25h)
- Travelled from 145km by train to education/course and back

VERDICT: Plausible

OPID = 56116734967

- PecDay = Yes
- TotTT = 270 min (4:30h)
- Peculiarity: Different duration, due to traffic
- Travelled for visitation and work, nothing spatial

VERDICT: Plausible

10 RANDOMLY SELECTED 'NO SPECIFIC PECULIARITIES' OUTLIERS
OPID = 56054665589

- TotTT = 345 min (5:45h)
- Retired man
- 3:30h of cycling

VERDICT: Plausible

OPID = 56026495982

- TotTT = 225 min (3:45h)
- Retired man
- 30km of cycling in 3:45h (fairly slow)
- Possibly did not record intermediate stops

VERDICT: Unsure

OPID = 55958375032

- TotTT = 265 min (4:25h)
- Cycling, hobbies, and groceries
- Nothing special

VERDICT: Very plausible

<u>OPID = 56051294928</u>

- TotTT = 234 min (3:54h)
- 17km of cycling in 3:24h (fairly slow)
- Possibly did not record intermediate stops

VERDICT: Unsure

<u>OPID = 55958328935</u>

- TotTT = 440 min (7:20h)
- 60,2km in 7h by car (very slow) for going to the store
- Possibly recorded leaving home -> arriving back home as one trip

VERDICT: Likely wrongly entered trip(s)

<u>OPID = 55992880884</u>

- TotTT = 330 min (5:30h)
- Retired man, 3:30h of cycling
- Nothing special

VERDICT: Very plausible

<u>OPID = 56114328406</u>

- TotTT = 272 min (4:32h)
- Student, travelling 2+ hours to education and back
- Mostly public transport

VERDICT: Very plausible

<u>OPID = 55992707408</u>

- TotTT = 307 min (5:07h)
- 9 year old girl
- 5+ hours in the car for school, while staying in the same municipality (Eindhoven)
- Possibly recorded leaving home -> arriving back home as one trip

VERDICT: Likely wrongly entered trip(s)

<u>OPID = 59216633535</u>

- TotTT = 229 min (3:49h)
- 9 year old boy
- Covering 20km in 3:49h by bike (fairly slow)
- 3:49h of cycling by a 9 year old child

VERDICT: Implausible

OPID = 56144374908

- TotTT = 261 min (4:21h)
- 15km of cycling in 2:30h (fairly slow)
- Possibly did not record intermediate stops

VERDICT: Unsure

B

Merging of municipalities

B.1. Comparison of merging municipalities

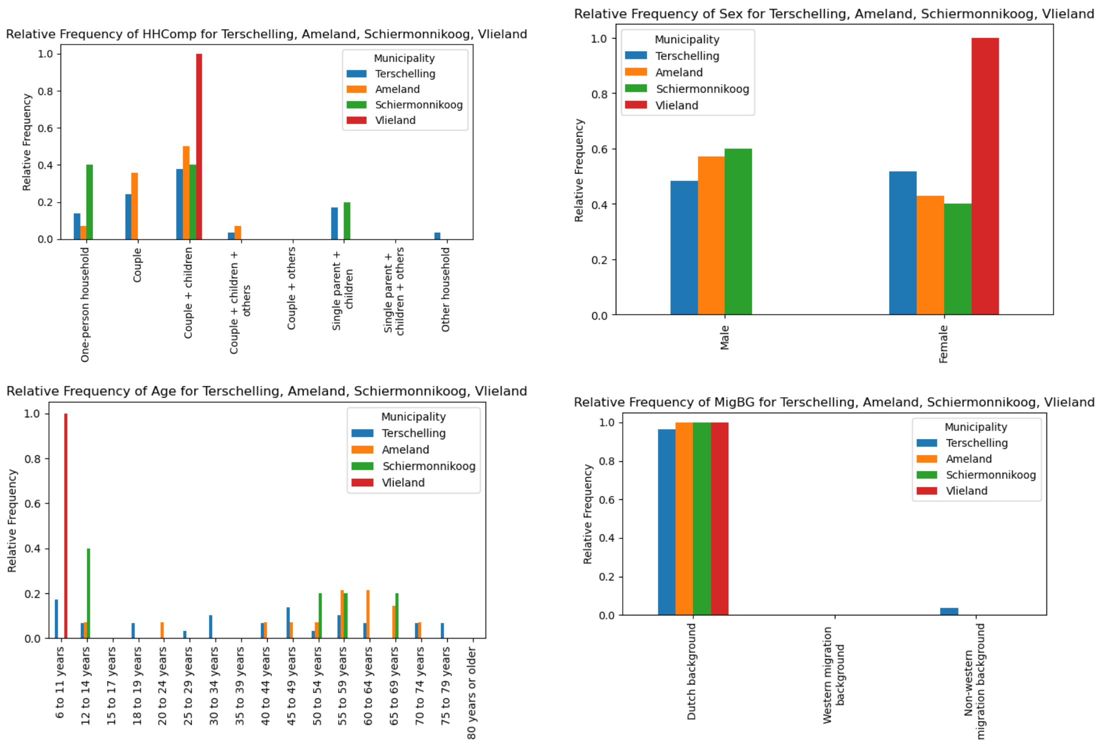


Figure B.1: Distributions of socio-demographic variables in Wadden Islands (1/2)

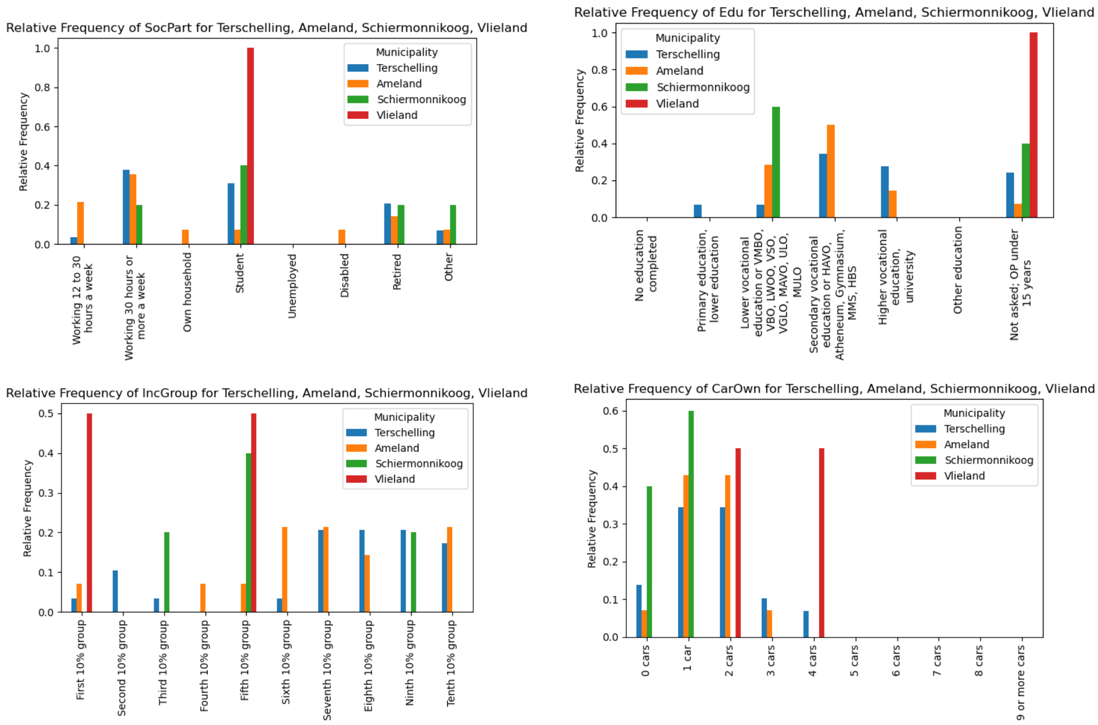


Figure B.2: Distributions of socio-demographic variables in Wadden Islands (2/2)

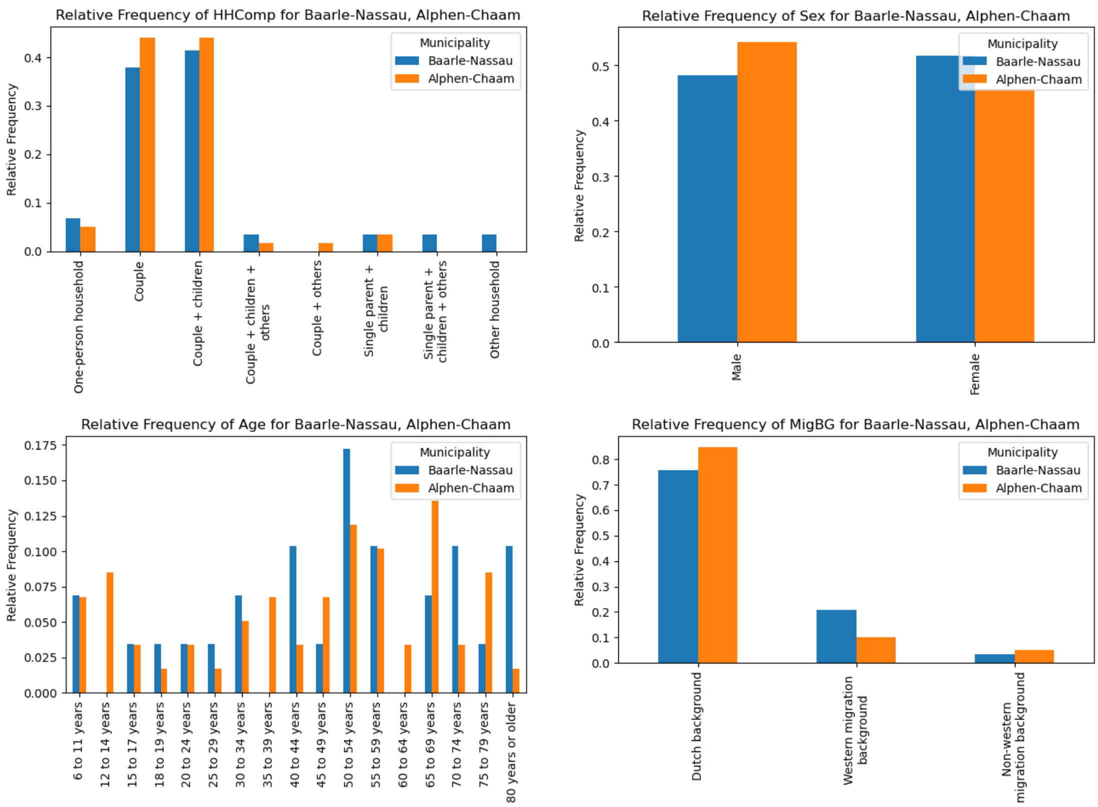


Figure B.3: Distributions of socio-demographic variables in Baarle-Nassau and Alphen-Chaam (1/2)

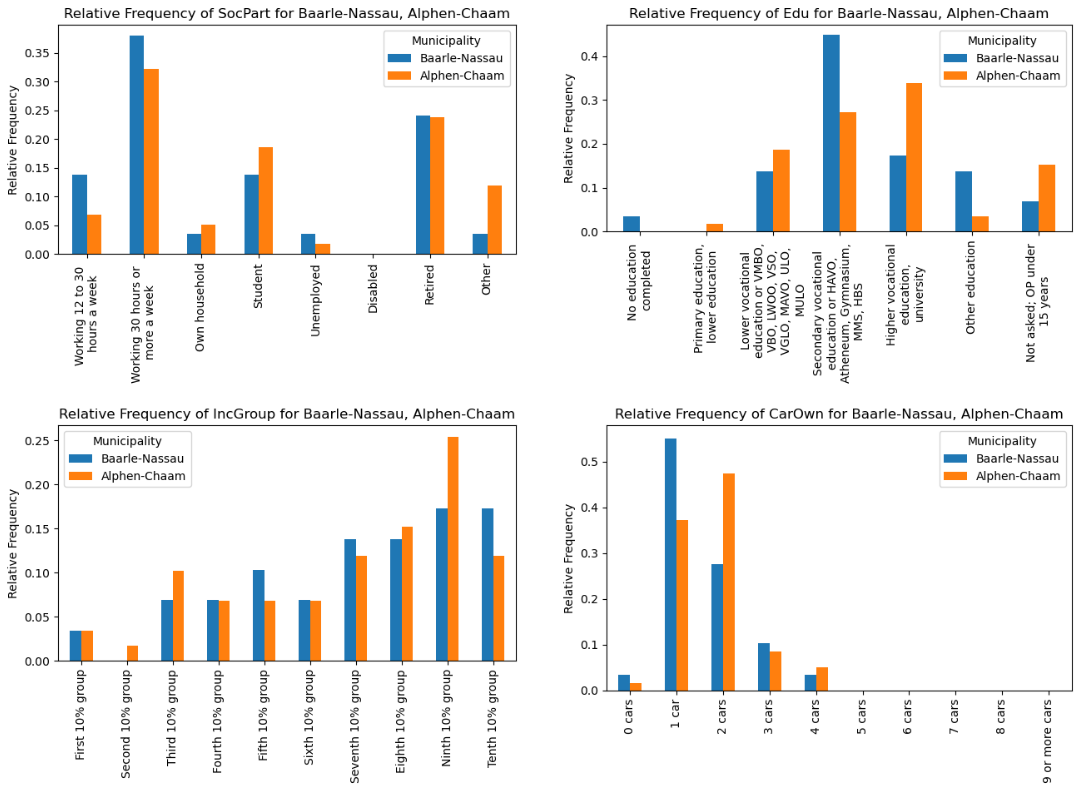


Figure B.4: Distributions of socio-demographic variables in Baarle-Nassau and Alphen-Chaam (2/2)

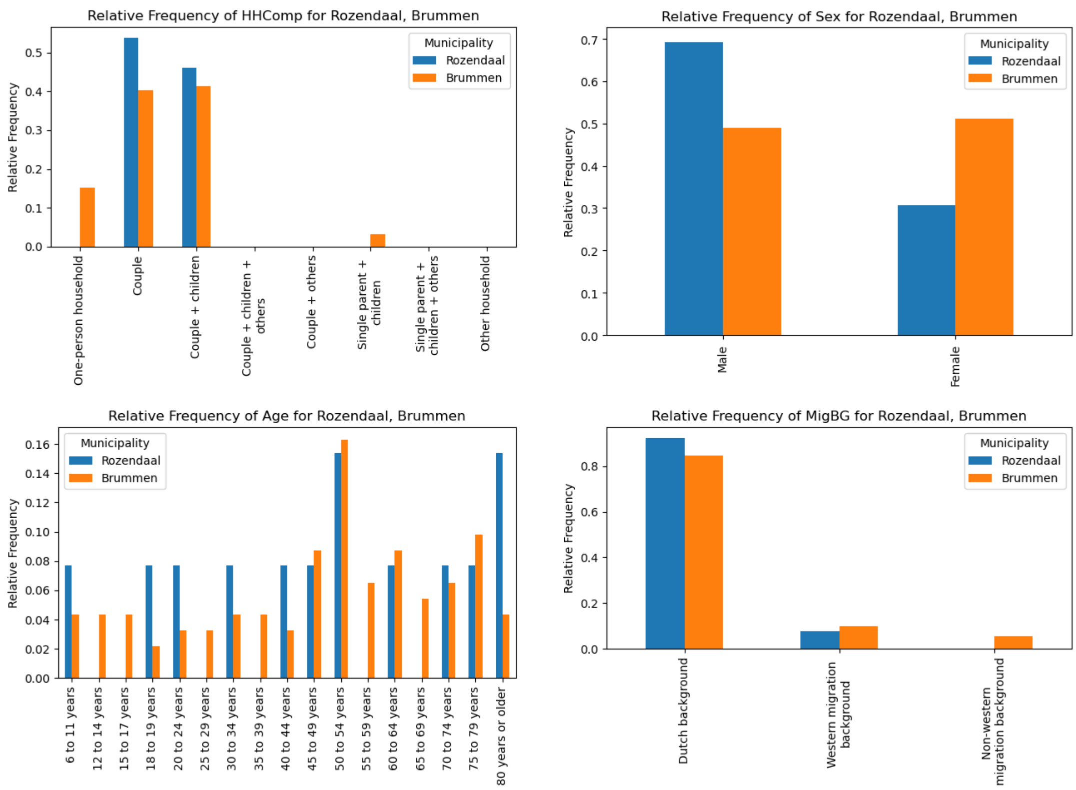


Figure B.5: Distributions of socio-demographic variables in Rozendaal and Brummen (1/2)

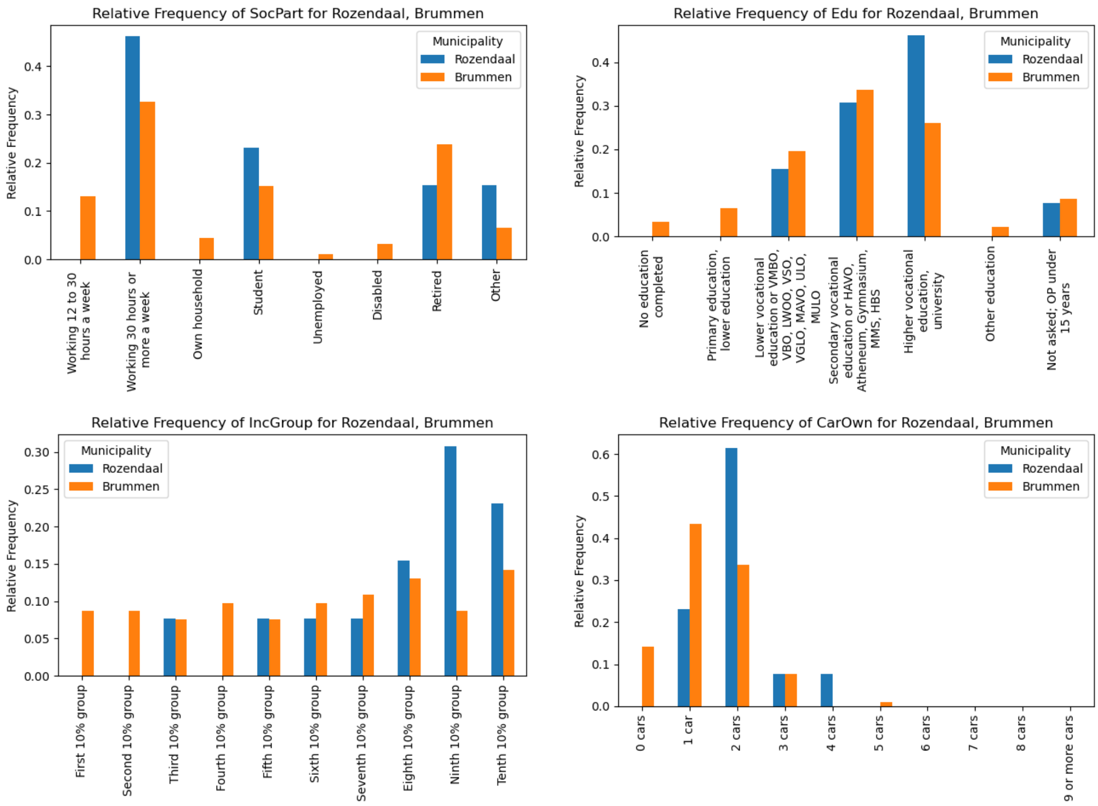


Figure B.6: Distributions of socio-demographic variables in Rozendaal and Brummen (2/2)

Table B.1: Built environment variables in Wadden Islands

Variable	Terschelling	Ameland	Schiermonnikoog	Vlieland
AddDens	248	253	317	204
UrbClass	Slightly urban	Slightly urban	Slightly urban	Slightly urban
PopDens	57	62	23	29
Pop	4890	3673	936	1138
NumBusi	745	540	130	195
HouVal	322	248	256	268

Table B.2: Built environment variables in Baarle-Nassau and Alphen-Chaam

Variable	Baarle-Nassau	Alphen-Chaam
AddDens	336	310
UrbClass	Non-urban	Non-urban
PopDens	90	109
Pop	6847	10149
NumBusi	935	1260
HouVal	274	311

Table B.3: Built environment variables in Rozendaal and Brummen

Variable	Rozendaal	Brummen
AddDens	935	763
UrbClass	Slightly urban	Slightly urban
PopDens	59	247
Pop	1654	20698
NumBusi	195	1785
HouVal	492	255

B.2. Value aggregation in geometric data

Table B.4: Methods of aggregation for variables in geometric data

Variable	Method of aggregation	Weight factor
AddDens	Weighted average	*Area*
PopDens	Weighted average	*Area*
Pop	Total	n/a
NumBusi	Total	n/a
HouVal	Weighed average	*WONINGEN*
Area	Total	n/a
*Dist_*facility**	Weighted average	*Pop*

C

Scatterplots of explanatory variables and average travel times

C.1. Socio-demographic characteristics

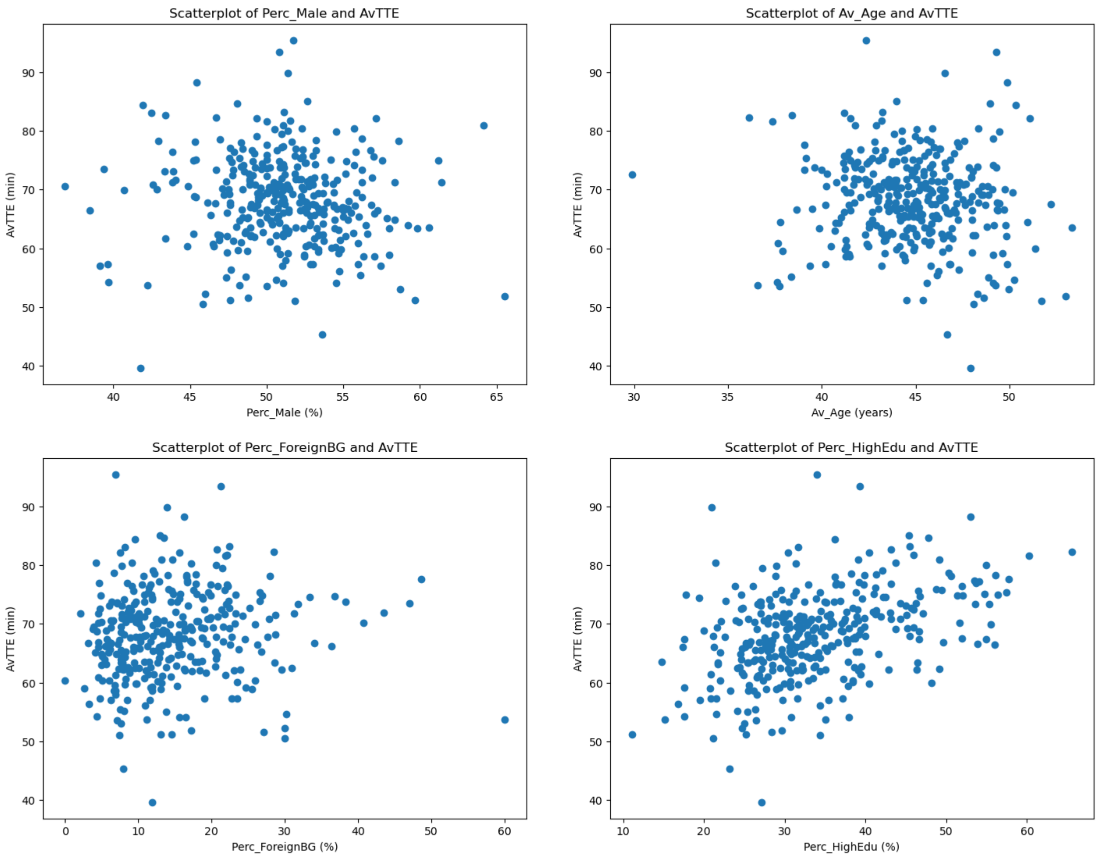


Figure C.1: Scatterplots of socio-demographic variables and average travel time expenditure per municipality (1/2)

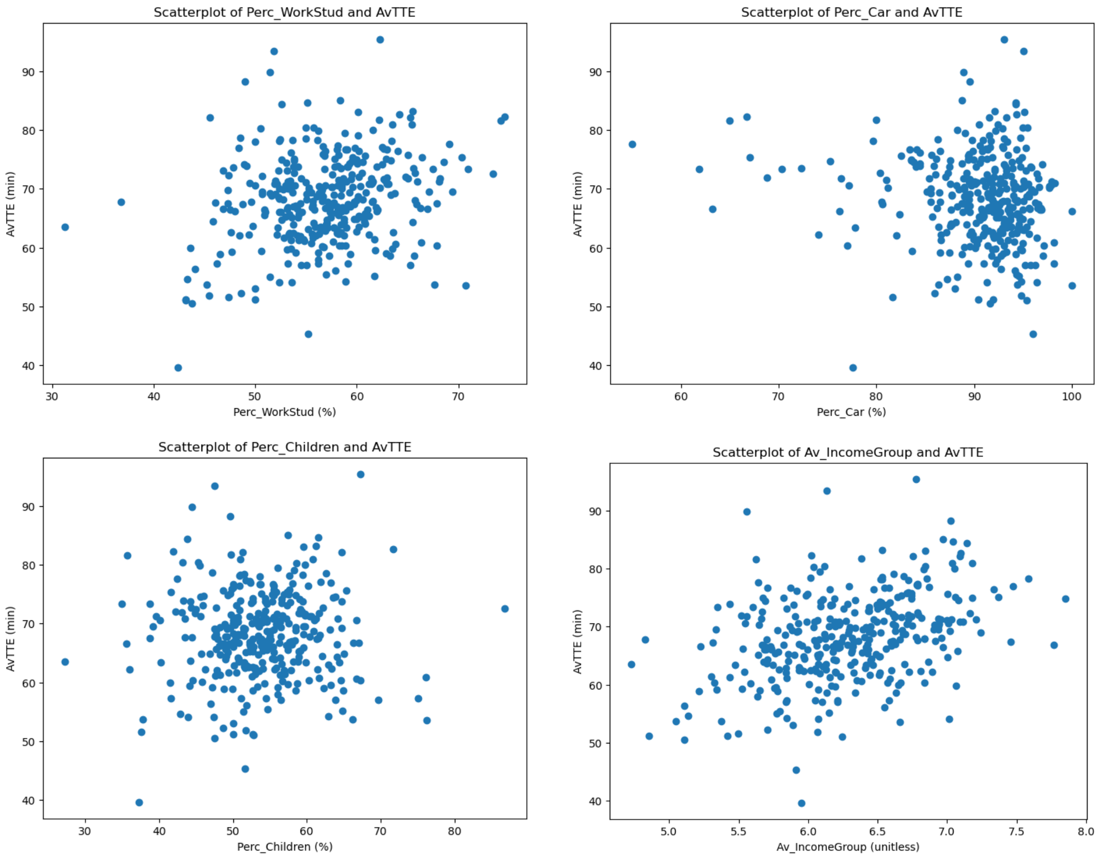


Figure C.2: Scatterplots of socio-demographic variables and average travel time expenditure per municipality (2/2)

C.2. Urban-characteristics

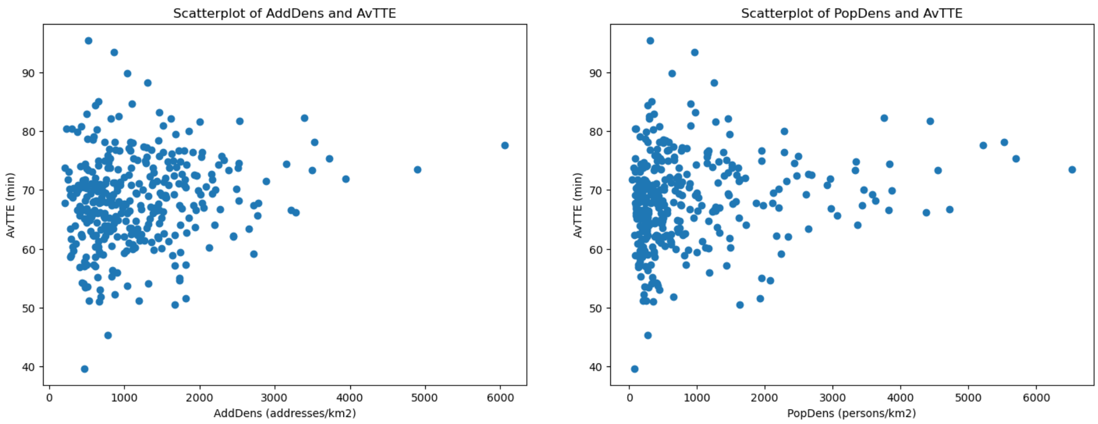


Figure C.3: Scatterplots of urban variables and average travel time expenditure per municipality (1/2)

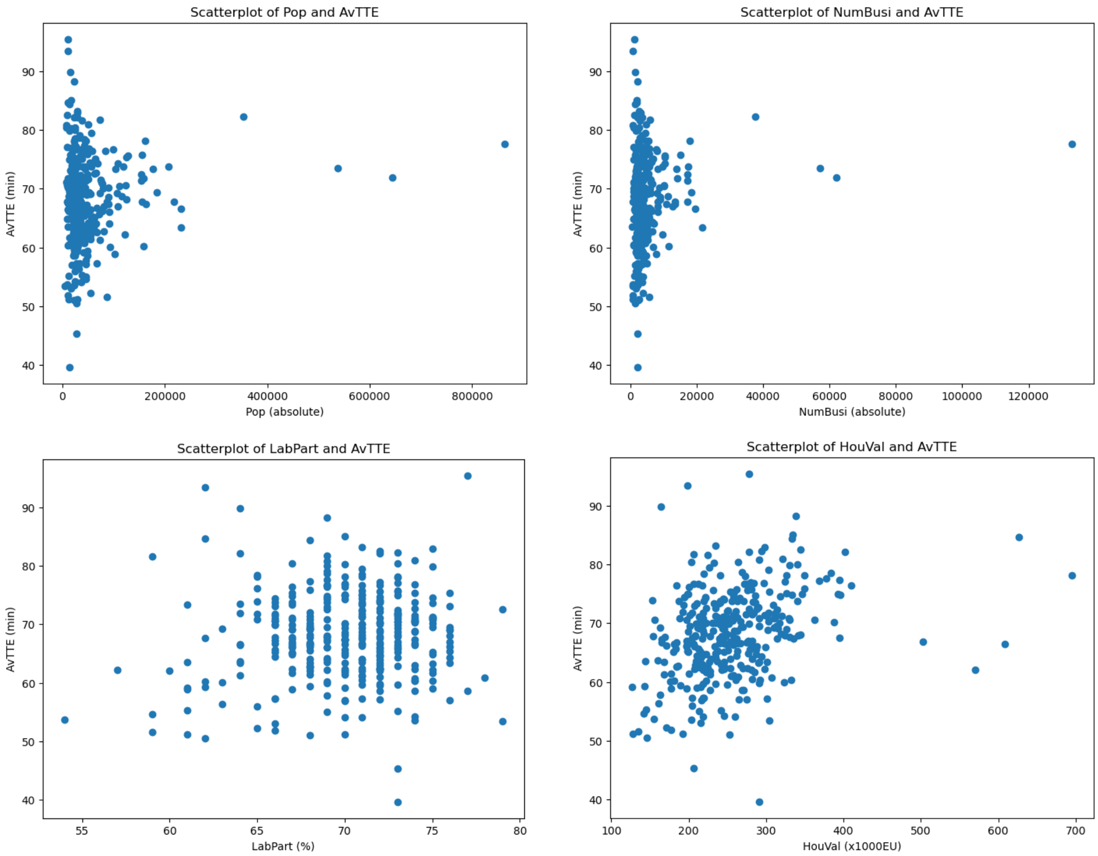


Figure C.4: Scatterplots of urban variables and average travel time expenditure per municipality (2/2)

C.3. Distance measures

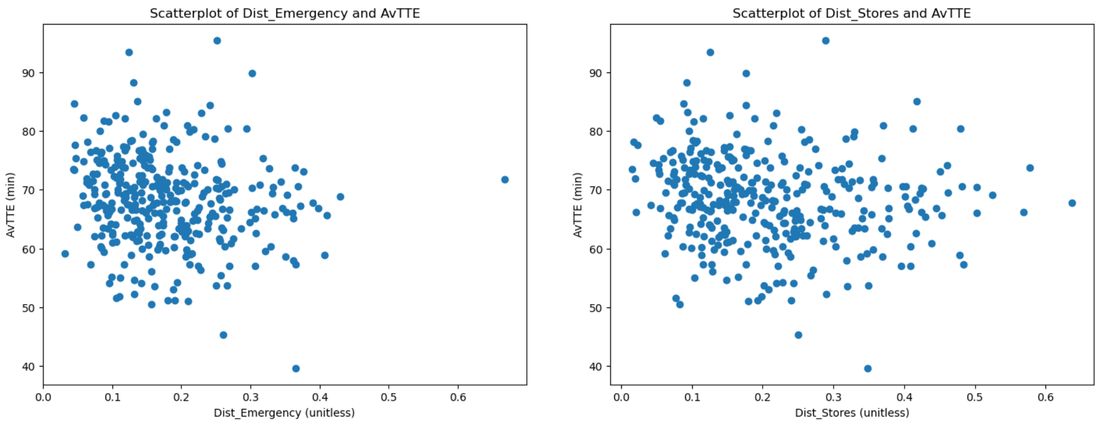


Figure C.5: Scatterplots of distance measures and average travel time expenditure per municipality (1/2)

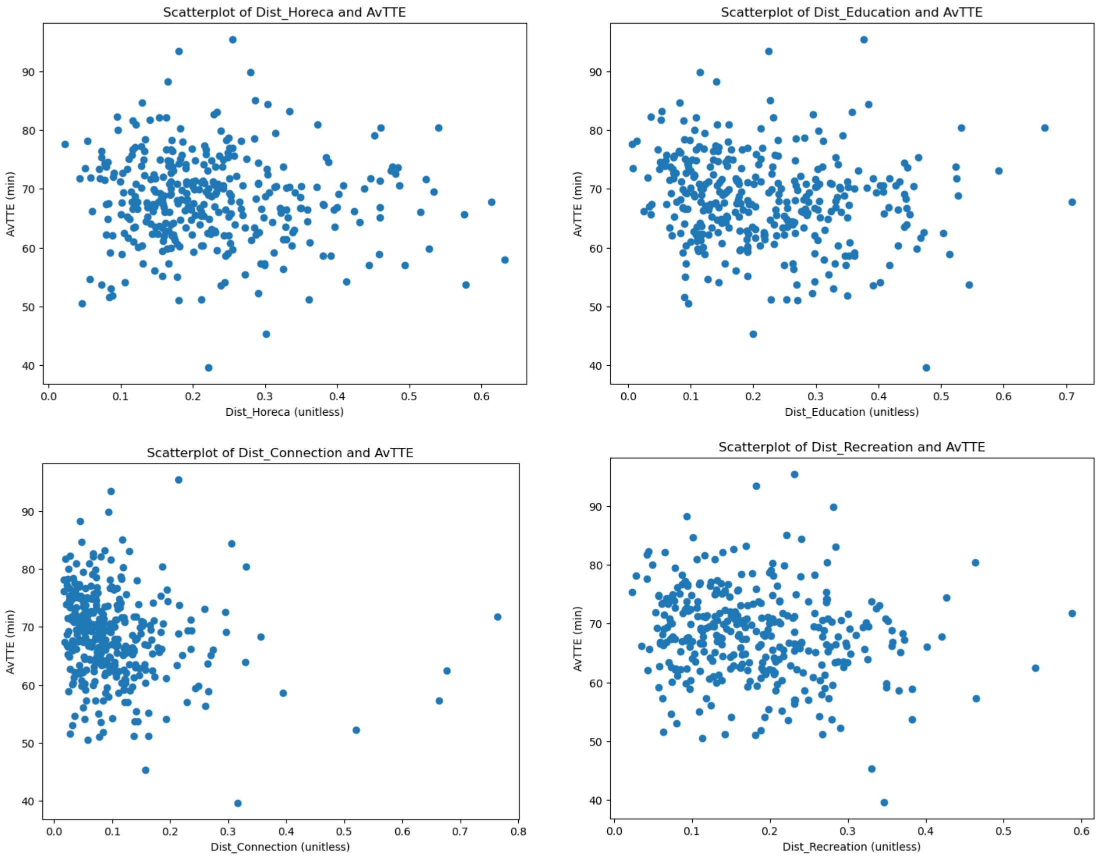


Figure C.6: Scatterplots of distance measures and average travel time expenditure per municipality (2/2)

D

Distributions of socio-demographic and built environment variables

D.1. Distributions of variables at individual level

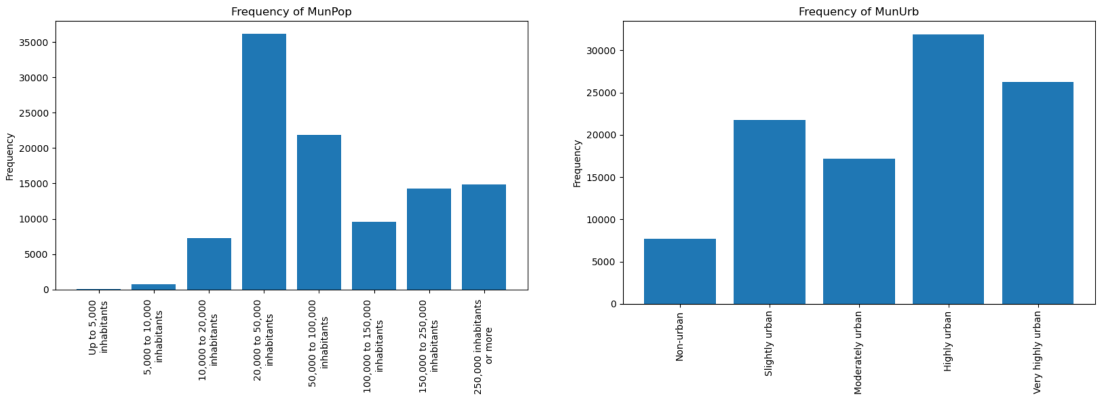


Figure D.1: Distributions of built environment variables in preprocessed ODiN data

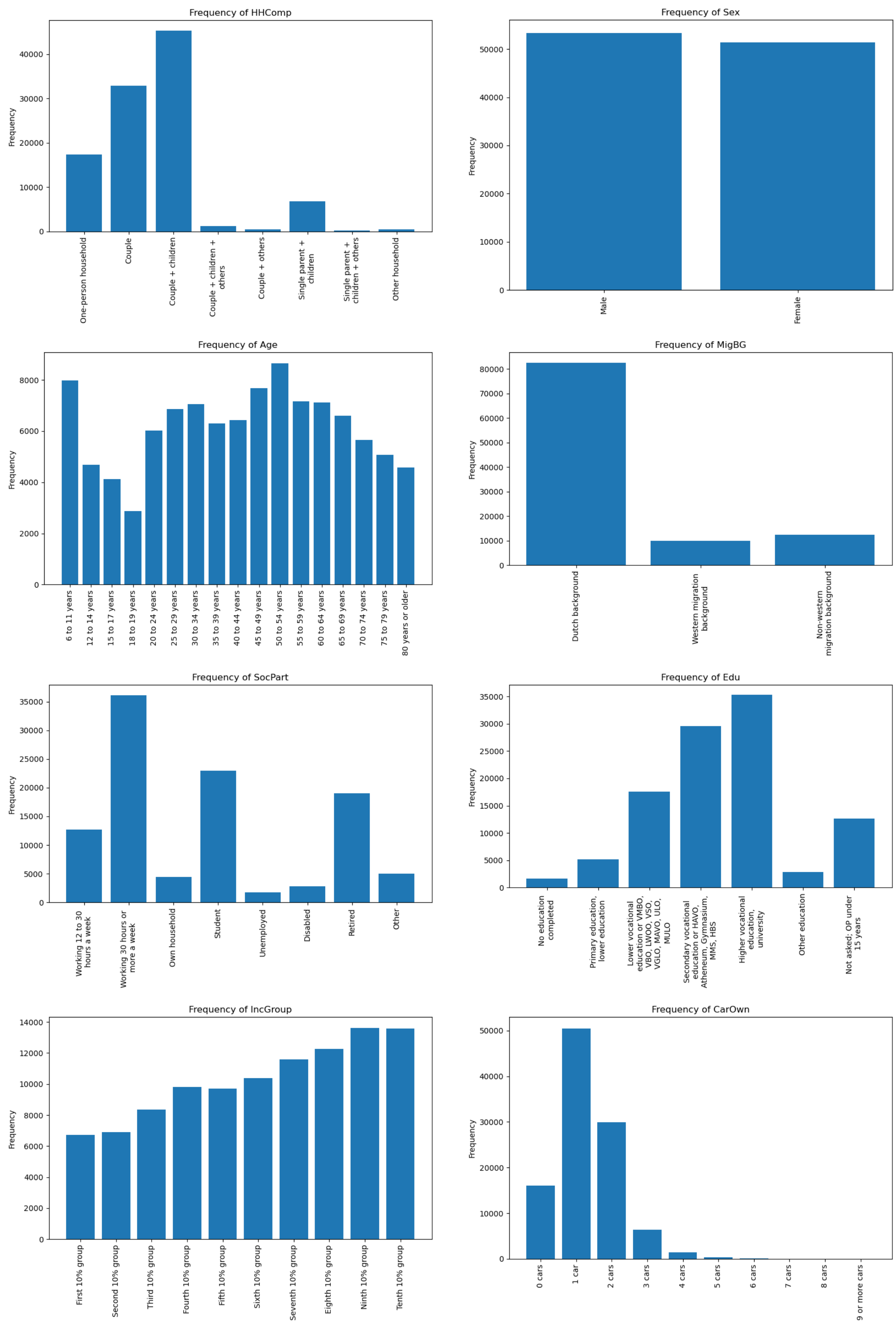


Figure D.2: Distributions of socio-demographic variables in preprocessed ODiN data

D.2. Distributions of socio-demographic variables at municipal level

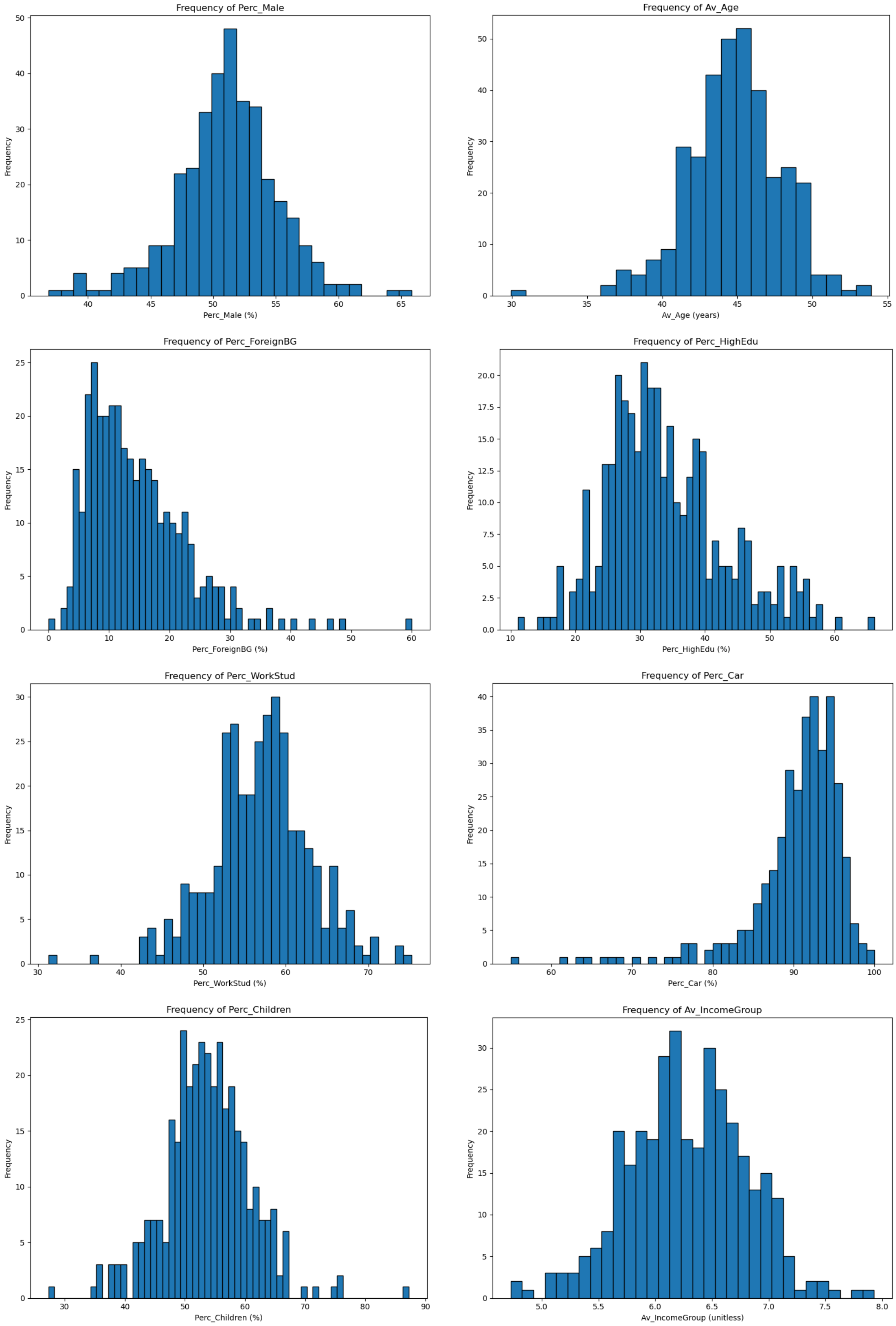


Figure D.3: Distributions of socio-demographic variables aggregated per municipality

D.3. Distributions of built environment variables at municipal level

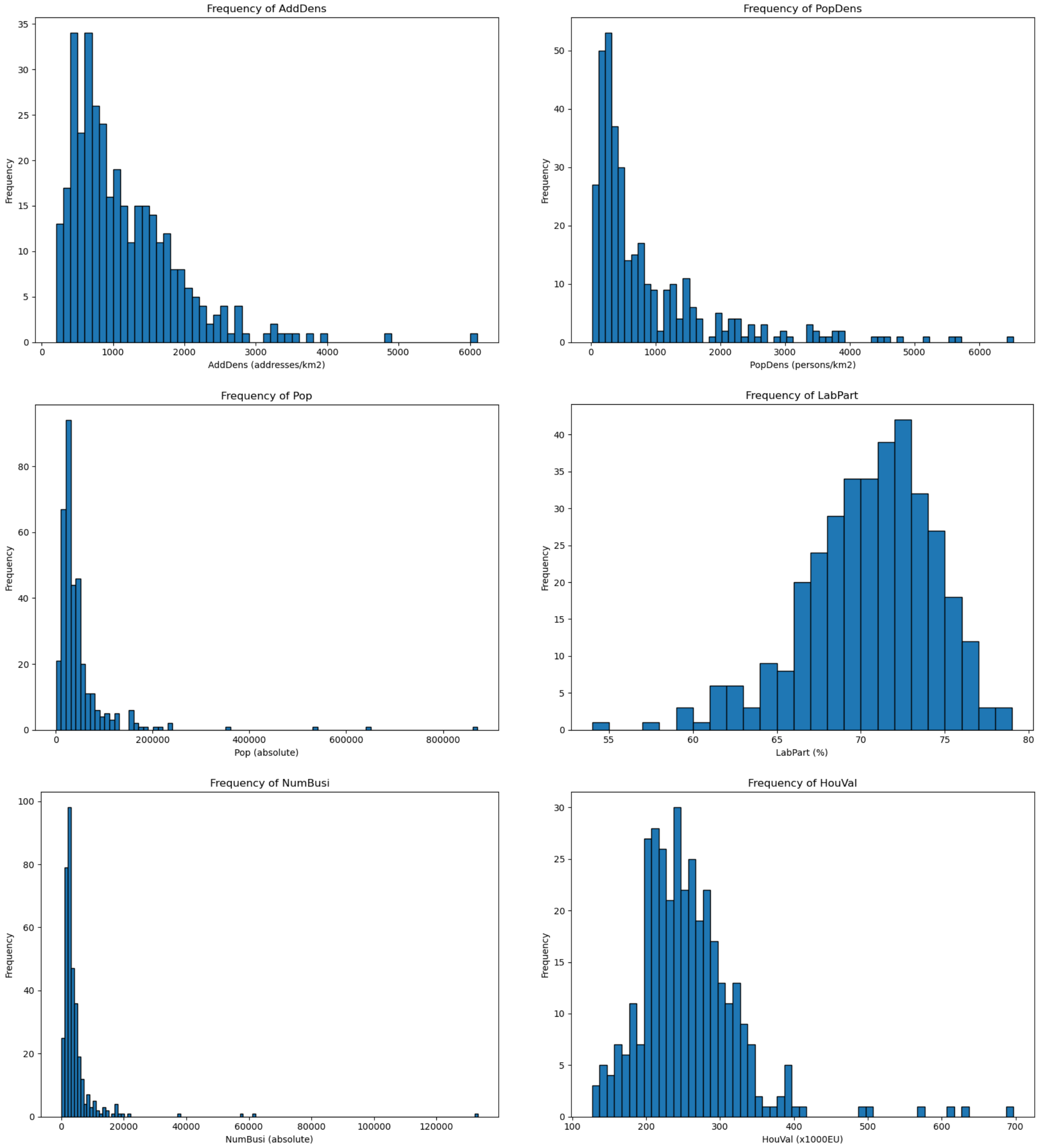


Figure D.4: Distributions of built environment variables in full CBS geometric data

E

Characteristics of municipalities of interest

E.1. Urban characteristics

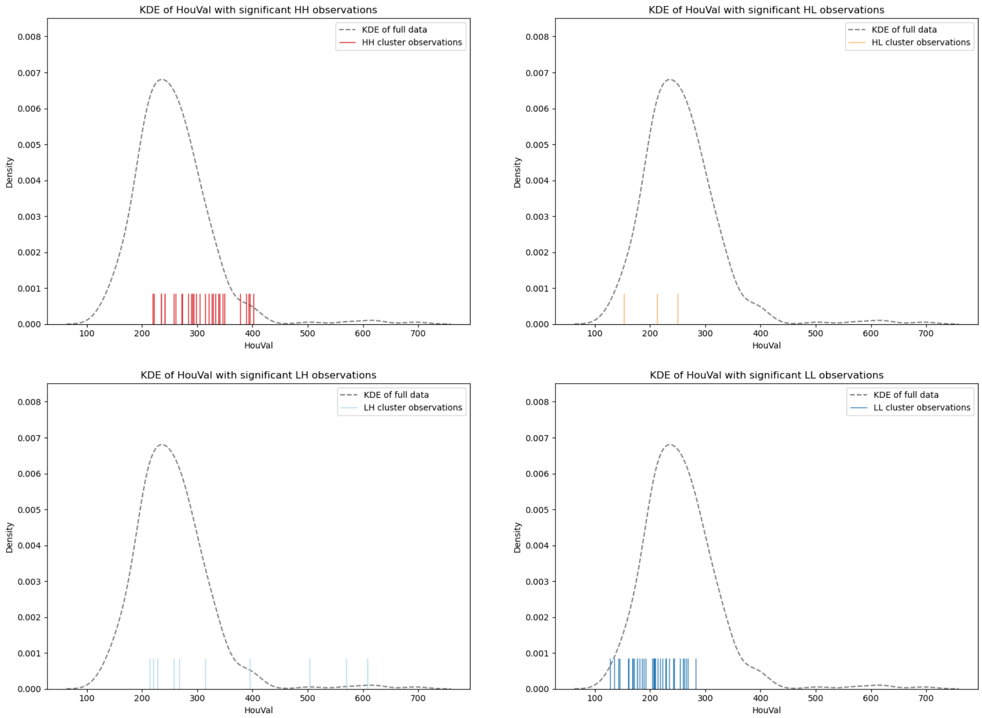


Figure E.1: Kernel density function of the average house value, with the values of the municipalities of interest marked

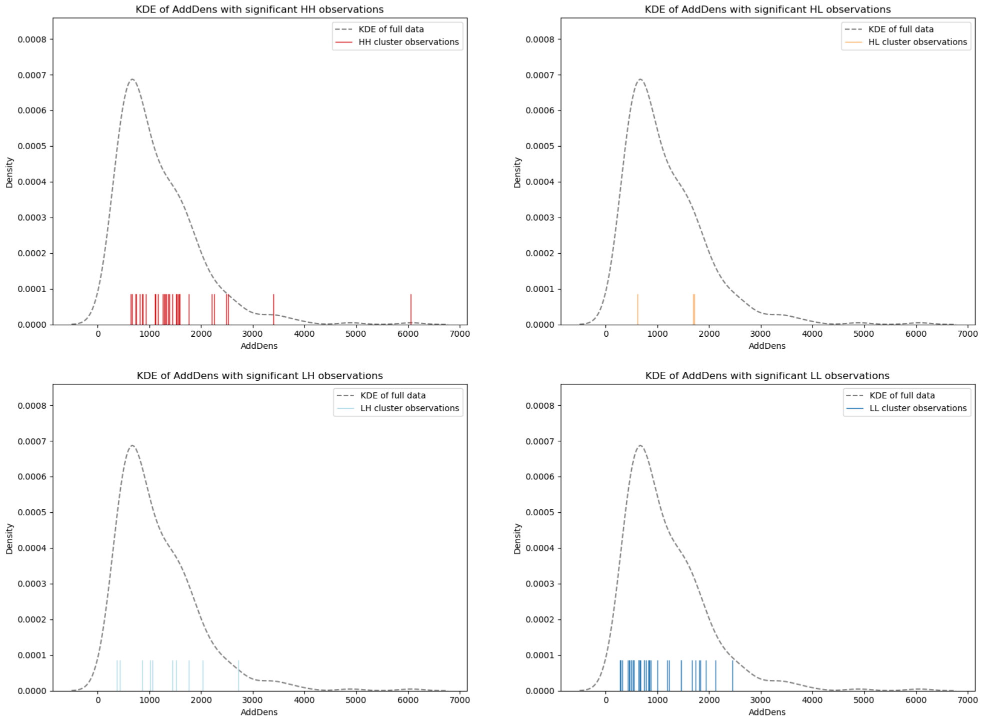


Figure E.2: Kernel density function of the address density, with the values of the municipalities of interest marked

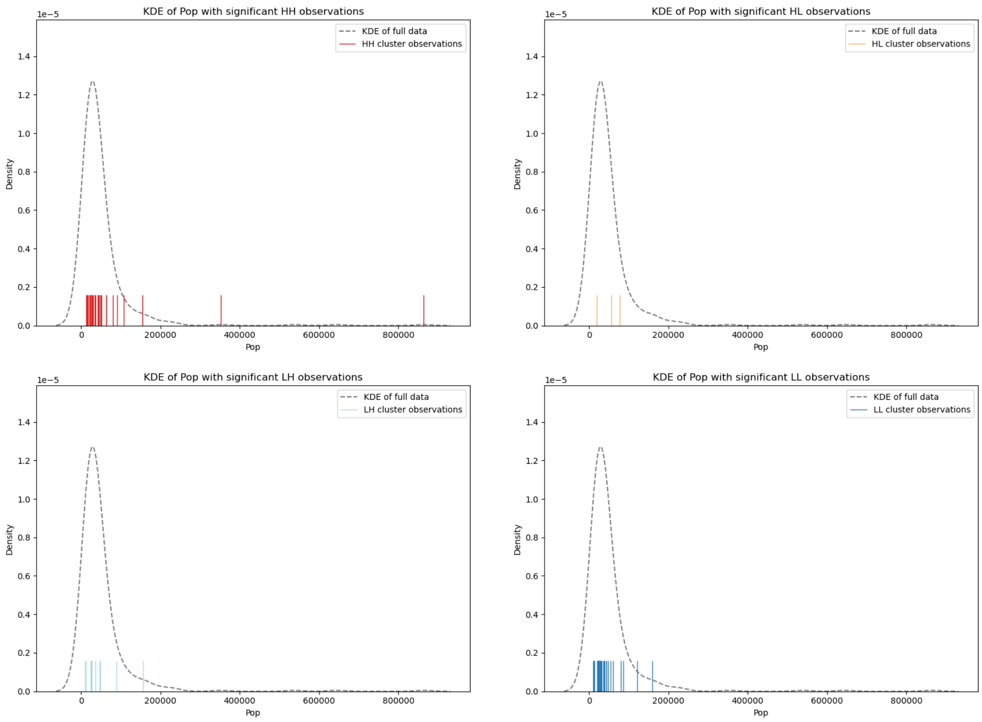


Figure E.3: Kernel density function of the population, with the values of the municipalities of interest marked

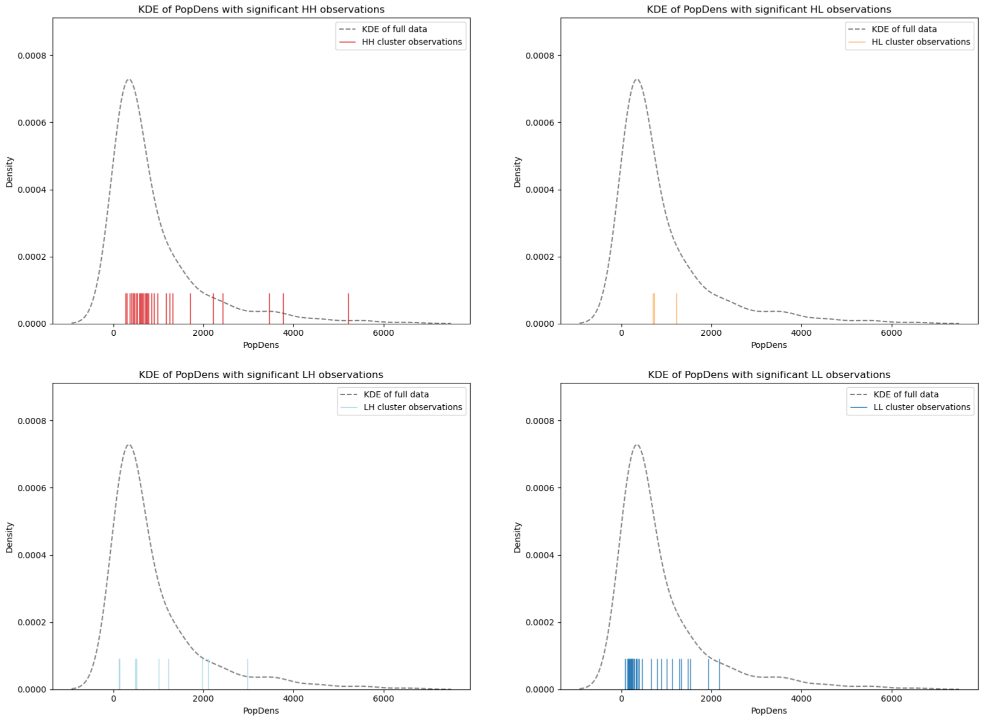


Figure E.4: Kernel density function of the population density, with the values of the municipalities of interest marked

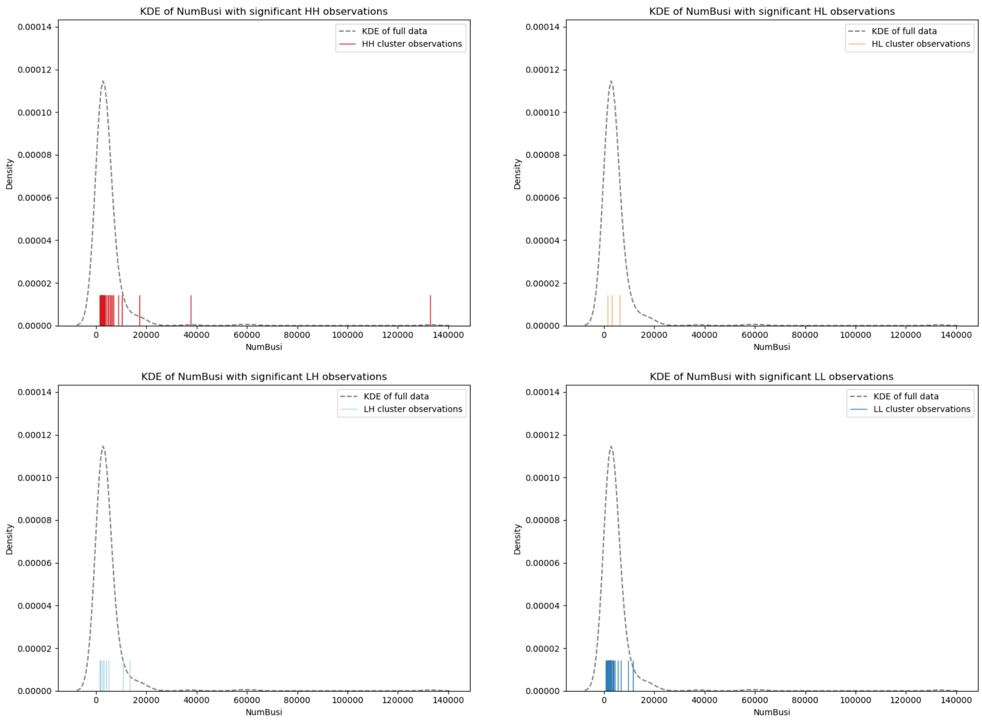


Figure E.5: Kernel density function of the number of businesses, with the values of the municipalities of interest marked

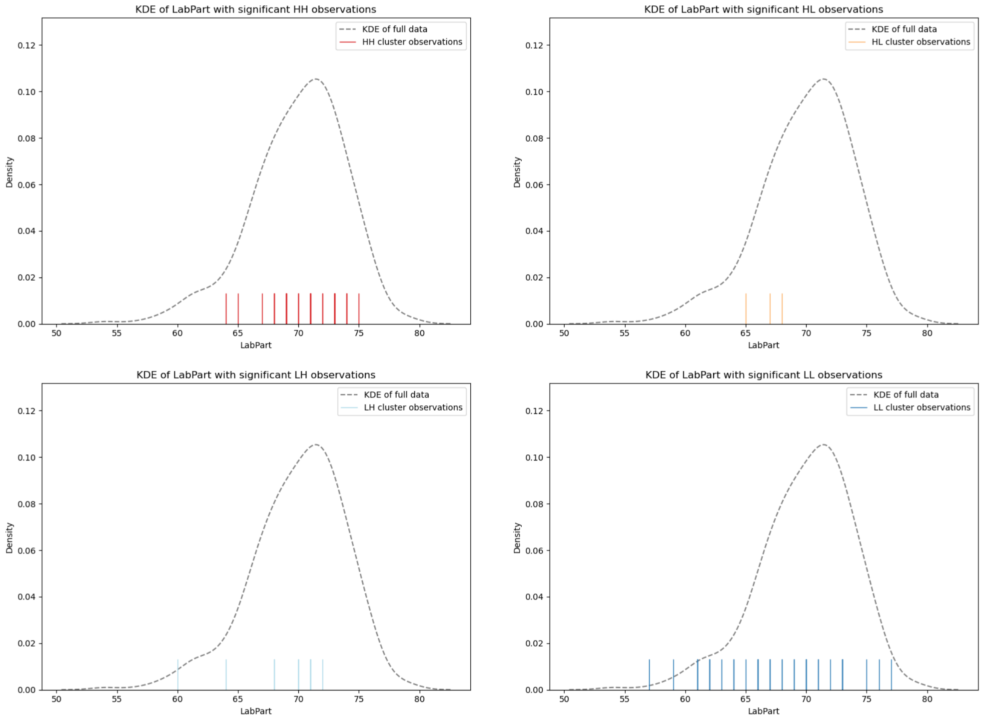


Figure E.6: Kernel density function of the number of net labour participation, with the values of the municipalities of interest marked

E.2. Socio-demographic characteristics

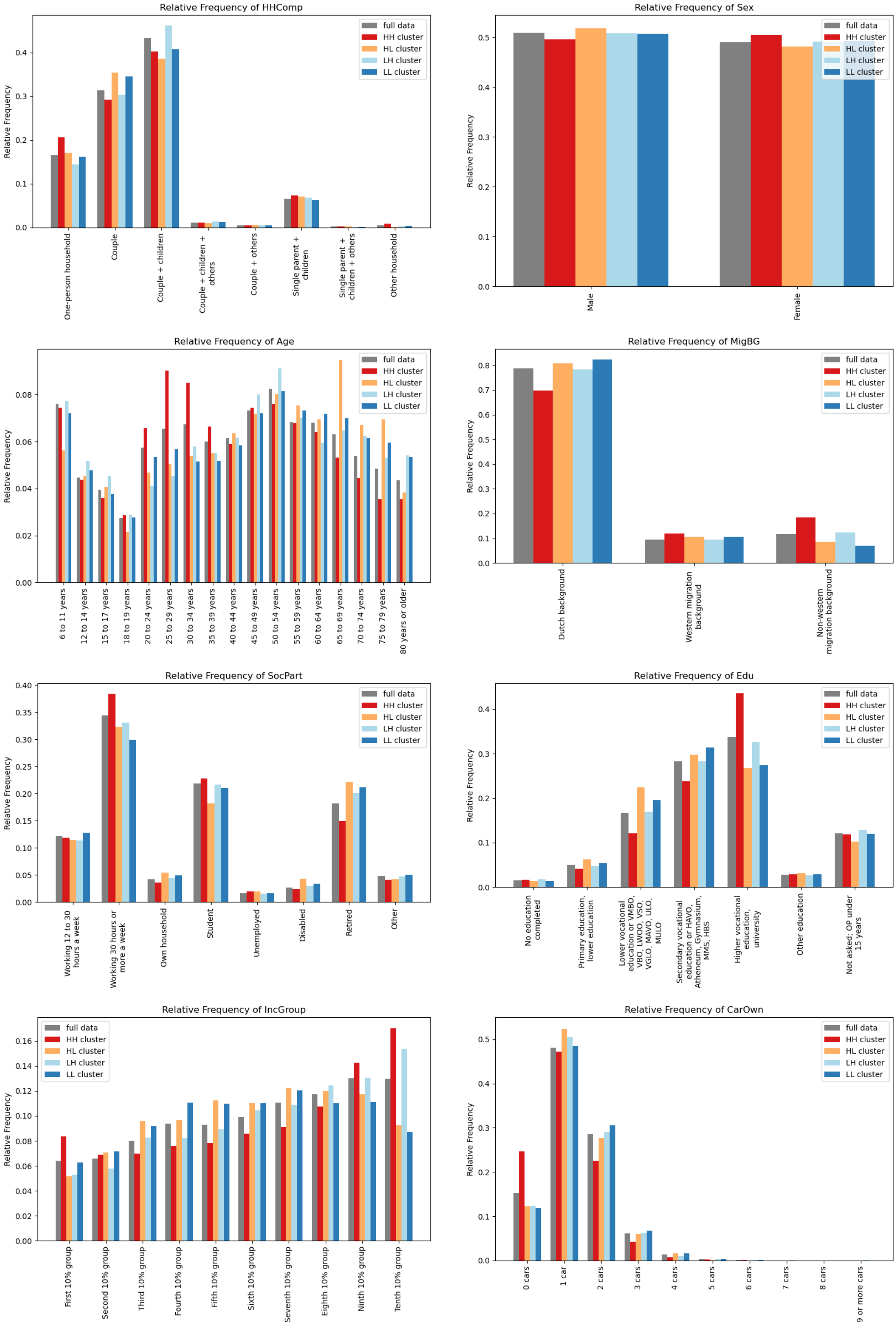


Figure E.7: Relative distributions of socio-demographic variables in municipalities of interest

F

Non-spatial regression models

Table F.1: Results of non-spatial regression model 1 (*significant at the α=0,05 level)

R^2	0,258	Adjusted R^2	0,241
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	54,242*		0,001
Perc_Male	-0,051	-0,027	0,592
Av_Age	-0,006	-0,002	0,982
Perc_ForeignBG	-0,086	-0,092	0,148
Perc_HighEdu	0,211*	0,255*	0,000
Perc_WorkStud	0,040	0,031	0,733
Perc_Car	-0,237*	-0,185*	0,038
Perc_Children	0,035	0,033	0,681
Av_IncomeGroup	4,471*	0,305*	0,000

Table F.2: Results of non-spatial regression model 2 (*significant at the α=0,05 level)

R^2	0,136	Adjusted R^2	0,126
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	53,166*		0,000
PopDens	0,001*	0,159*	0,008
Pop	0,000	0,043	0,463
LabPart	0,068	0,035	0,522
HouVal	0,035*	0,314*	0,000

Table F.3: Results of non-spatial regression model 3 (*significant at the α=0,05 level)

R^2	0,040	**Adjusted R^2**	0,024
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	70,303*		0,000
Dist_Emergency	-4,208	-0,047	0,628
Dist_Stores	-5,226	-0,079	0,478
Dist_Horeca	6,432	0,094	0,227
Dist_Education	-0,299	-0,005	0,963
Dist_Connection	-9,391	-0,109	0,152
Dist_Recreation	-3,832	-0,047	0,686

Table F.4: Results of non-spatial regression model 4 (*significant at the α=0,05 level)

R^2	0,266	**Adjusted R^2**	0,240
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	59,278*		0,002
Perc_Male	-0,071	-0,038	0,455
Av_Age	-0,050	-0,020	0,841
Perc_ForeignBG	-0,148*	-0,159*	0,035
Perc_HighEdu	0,220*	0,265*	0,001
Perc_WorkStud	0,035	0,028	0,765
Perc_Car	-0,190	-0,149	0,132
Perc_Children	0,061	0,058	0,488
Av_IncomeGroup	5,608*	0,382*	0,000
PopDens	0,000	0,062	0,358
Pop	0,000	0,039	0,554
LabPart	-0,176	-0,089	0,237
HouVal	-0,009	-0,077	0,306

Table F.5: Results of non-spatial regression model 5 (*significant at the α=0,05 level)

R^2	0,294	**Adjusted R^2**	0,264
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	47,740*		0,002
Perc_Male	-0,071	-0,038	0,461
Av_Age	0,111	0,045	0,650
Perc_ForeignBG	-0,035	-0,038	0,571
Perc_HighEdu	0,186*	0,224*	0,003
Perc_WorkStud	0,024	0,019	0,837
Perc_Car	-0,338*	-0,265*	0,004
Perc_Children	0,029	0,027	0,731
Av_IncomeGroup	6,050*	0,412*	0,000
Dist_Emergency	9,725	0,109	0,213
Dist_Stores	-10,302	-0,156	0,113
Dist_Horeca	12,818*	0,188*	0,011
Dist_Education	-7,603	-0,125	0,198
Dist_Connection	-12,454*	-0,145*	0,037
Dist_Recreation	19,393*	0,236*	0,027

Table F.6: Results of non-spatial regression model 6 (*significant at the α=0,05 level)

R^2	0,151	**Adjusted R^2**	0,126
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	54,213*		0,000
PopDens	0,001*	0,173*	0,021
Pop	0,000	0,044	0,459
LabPart	0,030	0,015	0,785
HouVal	0,037*	0,329*	0,000
Dist_Emergency	4,057	0,046	0,626
Dist_Stores	-5,686	-0,086	0,422
Dist_Horeca	7,650	0,112	0,135
Dist_Education	-0,967	-0,016	0,877
Dist_Connection	-9,747	-0,113	0,121
Dist_Recreation	5,965	0,073	0,519

Table F.7: Results of non-spatial regression model 7 (*significant at the α=0,05 level)

R^2	0,304	**Adjusted R^2**	0,267
Variable	**Coefficient**	**Standardised Coefficient**	**P-value**
CONSTANT	51,680*		0,005
Perc_Male	-0,087	-0,046	0,372
Av_Age	0,056	0,023	0,822
Perc_ForeignBG	-0,100	-0,108	0,157
Perc_HighEdu	0,207*	0,250*	0,002
Perc_WorkStud	0,012	0,010	0,918
Perc_Car	-0,287*	-0,225*	0,024
Perc_Children	0,059	0,056	0,497
Av_IncomeGroup	7,222*	0,492*	0,000
PopDens	0,001	0,107	0,158
Pop	0,000	0,029	0,650
LabPart	-0,179	-0,091	0,225
HouVal	-0,009	-0,084	0,257
Dist_Emergency	8,578	0,096	0,273
Dist_Stores	-8,732	-0,132	0,185
Dist_Horeca	11,606*	0,170*	0,022
Dist_Education	-6,443	-0,106	0,279
Dist_Connection	-13,389*	-0,156*	0,026
Dist_Recreation	22,764*	0,277*	0,011