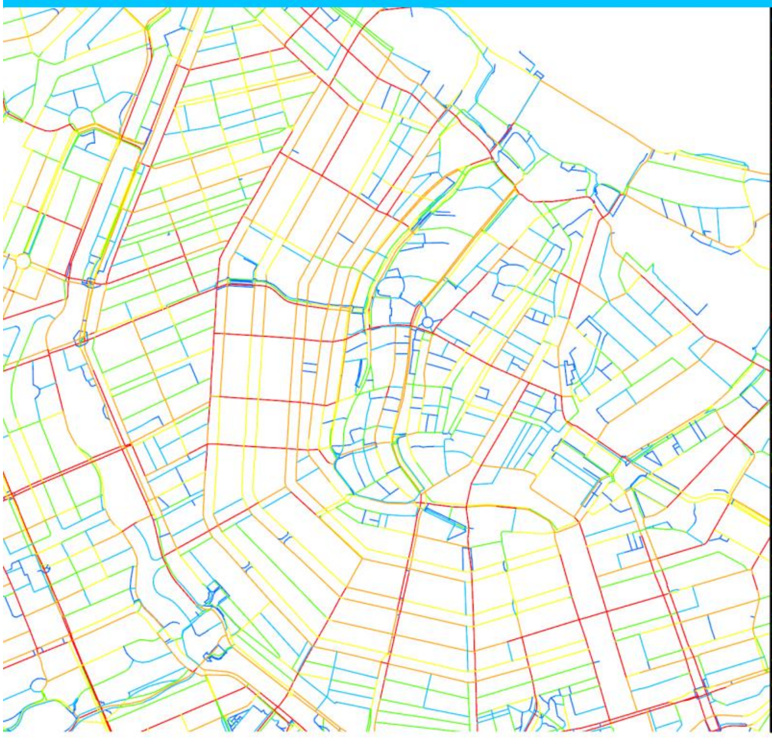
The search for cycling routes

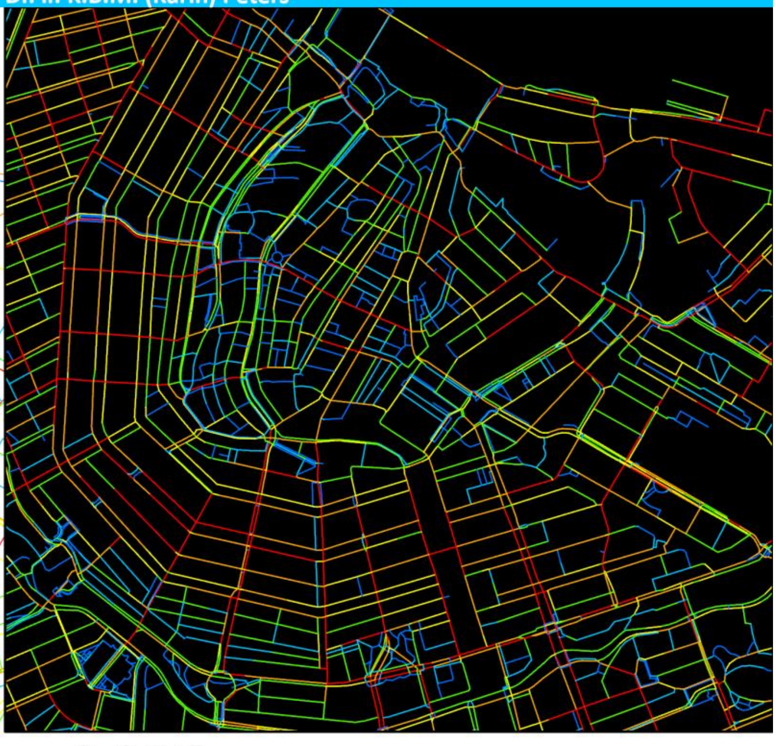
Analysing the influence of spatial characteristics on cycling route choices in Amsterdam

Author: **Simon Drolsbach**

Date: **13/12/2022**

Assessors: **Dr. ir. W.W.Y. (Wendy) Tan**
Dr. ir. D.C. (Dorine) Duives
Dr. ir. K.B.M. (Karin) Peters





TU Delft

WAGENINGEN UNIVERSITY & RESEARCH

AMS AMSTERDAM INSTITUTE FOR ADVANCED METROPOLITAN SOLUTIONS

MSc MADE

The search for cycling routes

Analysing the influence of spatial characteristics on cycling route choices in Amsterdam

MSc Metropolitan Analysis, Design & Engineering
Master Thesis

by

Simon Drolsbach
13/12/2022

E-mail:
simon.drolsbach@wur.nl
s.k.drolsbach@student.tudelft.nl

TU Delft student number:
 738330
WUR student number:
1048073

Supervisors:
Dr. ir. W.W.Y. (Wendy) Tan
Wageningen University & Research

Dr. ir. D.C. (Dorine) Duives
Delft University of Technology

Third Reader:
Dr. ir. K.B.M. (Karin) Peters
Wageningen Univesity & Research

Institutions:
AMS Institute
Delft University of Technology
Wageningen University & Research

Abstract

In recent years, cycling has attracted increasing attention as a sustainable alternative to private car use. In cities across the world, strategies are put in place to improve existing cycling infrastructure. This raises the question of how cyclists move through a city and what part of the road network attracts the most cyclists. Prior studies on cycling behaviour have revealed that cyclists prefer routes that require less turns, over separated bicycle paths, smoother street surface materials like asphalt or concrete. Furthermore, dense and mixed-use neighbourhoods seem to attract more active travel. This research will analyse correlations between spatial characteristics and cycling route choices in Amsterdam.

Space Syntax is an analysis method that studies the urban morphology of a city. Until recently, the implementation of Space Syntax has mainly focused on the analysis of pedestrian flows, with a limited number of studies applying the methodology to cyclist behaviour. This master thesis presents exploratory research into the application of Space Syntax – in combination with other built environment characteristics – to study cyclist route choice. GPS data from the 2016 Bicycle Counting Week shows the cycling counts of every street segment.

A linear regression analysis found that “through-movement potential” represented by Normalised Angular Choice (NACH) explained more than 22% of variance in cycling activity. The results indicate that Space Syntax is an interesting indicator to locate which street segments could potentially see large numbers of cyclists. More research encompassing multiple cities in a variety of different contexts is recommended, as Amsterdam is a city with a rich cycling culture that spans multiple decades, making it difficult to generalize any conclusions.

Keywords: *Cycling route choices, Linear regression, Space Syntax, Revealed preference data*

Acknowledgments

This thesis marks the end of a wonderful two-year-long MSc MADE programme at the AMS Institute. Following a master’s programme during the Covid-19 pandemic has not been easy. The staff at AMS Institute has made it possible for students to physically come to class as much as possible. I think the positive impact that this had on mental wellbeing among students cannot be emphasised enough.

I would like to thank both my supervisors for taking their time and putting their effort into helping me structure this research project from start to end. This thesis project took longer than originally intended. A lot has happened during this time, and I am happy that I still enjoy cycling through Amsterdam as much as I did before I started this thesis project.

I would also like to thank my friends and student-peers, with whom I have spent numerous hours every day in the university library. Without them, this journey would have been a very lonely one.

Last, but not least, I would like to thank my family for their support.

Contents

Terms & Acronyms	
BCW	Bicycle Countin Week (“Nationale Fietstelweek”) – Usually a week in September in which volunteers are asked to collect their own cycling data.
BGT	Basisregistratie Grootschalige Topografie – Geographical dataset containing geodata about almost all large objects in public space
BIC	Bayesian Information Criterion – statistical test, used to compare goodness of fit between models that predict the same phenomenon.
MLM	Multiple Linear Regression Model – A linear regression model, containing multiple independent variables.
NACH	Normalised Angular Choice – Space Syntax measure that depicts “through-movement potential” of each street segment in a network.
NAIN	Normalised Angular Integration – Space Syntax measure that depicts “to-movement potential of each street segment in a network.
RUDIFUN	“Ruimtelijke Dichtheden & Functiemenging” – Dataset that contains geodata about land use mix and urban density in the Netherlands.
Street segment	Part of a street between two junctions

1 Introduction

In cities and urban regions around the world, policymakers and planners have shown an increasing interest in cycling as a sustainable alternative to private car use. In European cities like Paris, London, and Berlin, numerous projects have been under construction to improve and expand cycling infrastructure. The benefits of cycling are ample. Compared to car usage, cycling requires less space, and almost no fuel – other than nutrition for the cyclist – and produces no excessive noise nor any greenhouse gas emissions (Heinen et al., 2010). Moreover, cycling is positively associated with life satisfaction(Ma & Ye, 2022), and the overall emotional well-being of people (Krizek, 2018).

In this research project, data from the Bicycle Counting Week will be used to analyse the route choices of cyclists in Amsterdam. More specifically, the characteristics of the built environment will be related to cycling activity. For policymakers and urban planners, it might be difficult to decide where to allocate sparse funding for cycling infrastructure improvements. Collecting cycling data in cities is known to be a costly and time-consuming endeavour. This necessitates the use of route choice models to be able to predict what parts of a street network are the most suitable for improvement. In the Netherlands, cycling has been an integral part of how people get around cities and regions for decades. The municipality of Amsterdam, for example, estimates the number of bicycles in Amsterdam to be around 880.000, which is more than its entire population. In recent years, both bicycle ownership and bicycle usage have increased, while car ownership has shown a trend of decline (Municipality of Amsterdam, 2019). The prominence of cycling activity in the Netherlands has meant that a lot of data on how people get around by bike is available. The national Bicycle Counting Week ("nationale fietstelweek") is one of many sources.

The increasing urgency for more sustainable transport modes has sparked an interest in cycling research. Various studies have tried to uncover the route choices of cyclists, either through Stated Preference (de Vries et al., 2010; Stinson & Bhat, 2003), or Revealed Preference data sources (Broach et al., 2012; Menghini et al., 2010; Prato et al., 2018; Sobhani et al., 2019). Stated preference data often comes in the form of travel diaries or surveys that are conducted among a pool of respondents. Revealed preference data is often comprised of GPS data showing the exact routes of cyclists in a geographical area. Recent technological advancements have made Revealed Preference data more easily obtainable. Various studies look at the preference of cyclists in terms of bicycle infrastructure, topology, and land-use variation (Prato et al., 2018).

Space Syntax is a method that is used to study movement patterns through space. It is used to analyse the connectivity of streets and travellers' cognitive understanding of road networks in cities. An important notion of Space Syntax as an analysis tool is that the spatial configuration of the urban grid is the most powerful determinant of urban movement (Hillier, 1996). This notion forms the basis of a central theory in Space Syntax:

"…movement largely dictates the configuring of space in the city, and in terms of the effects of spatial form, in that movement is largely determined by spatial configuration." (Hillier, 1996, p. 113).

Hillier adds that this is true for both pedestrian and vehicle movement. It is through the interplay between movement and the urban grid, that socio-economic factors shape the city. Well-functioning cities can therefore be seen as "movement economies".

The interplay between built environment factors and cyclists' behaviour has been studied extensively in the past decades. New GPS technologies have made it possible to analyse the route choices of cyclists even better than before. The connection between revealed cyclist behaviour and Space Syntax measures, although studied by some, remains rare. It is therefore interesting to see how the movement of cyclists through the city of Amsterdam is related to the spatial configuration of the urban grid. Prior studies on cyclist route choices suggest that cyclists prefer routes that require the least number of turns (Halldórsdóttir, 2015). As will be explained later, Space Syntax is based on the number of directional changes. It is interesting to see whether Space Syntax is useful in explaining cycling activity.

1.1. Research objective and research question

This research project is an attempt to unravel the influence of the built environment on cycling routes in urban areas, with a special focus on the use of Space Syntax to explain cycling activity.

Research question:

To what extent do spatial characteristics influence bicycle route choices of cyclists?

This research question will be supported by four sub-questions:

1. What (environmental) factors influence cycling patterns, according to existing literature?

2. What Space Syntax measures can be used to analyse cycling patterns in urban areas?

3. To what extent do Space Syntax measures explain cycling route choices in Amsterdam?

4. To what extent do land-use and transportation variables explain cycling counts per street segment in Amsterdam?

1.2. Thesis aim

The aim of this thesis research is to analyse and explain the spatial distribution of cycling activity, by comparing it with spatial characteristics. The novelty of this research lies in the inclusion of Space Syntax in the analysis of cyclist route choices. Over the years, a plethora of academic research has looked into the determinants of cyclist route choices, however, the use of Space Syntax in these analyses has been rare. In recent times, the urgency to rethink and reshape the way people get around in cities has been growing, with an increasing interest in active modes of transport. There are ample examples of the successful implementation of Space Syntax to create more pedestrian-friendly walking environments. Space Syntax has been used in various different contexts, all over the world. The outcome of this research could be used as a stepping stone for bicycle infrastructure master planning in urban areas where cycling infrastructure has not yet had its foothold.

Each sub-question that is mentioned above has its own objective. The objective of sub-question 1 is to construct an understandable theoretical framework of factors that influence cycling route choices. This framework will be formed, based on existing academic literature about cycling. The aim of sub-question 2 is to get a clear overview of the theory behind Space Syntax and to come to a measure that best explains with cycling activity. Sub-question 3 will dive further into the relationship between the Space Syntax measure from sub-question 2, to see how the distribution of cycling activity can be explained by Space Syntax. Finally, the objective of sub-question 4 is to find out what spatial characteristics explain the distribution of cycling activity the most.

1.3. Thesis outline

This thesis will try to answer the research question and sub-questions in the following way: Chapter 2 “Literature review” will first summarise existing academic literature on cycling route choices. The second half of the chapter will dive deeper into the theory behind Space Syntax, and the various instances that Space Syntax has been coupled with cycling analysis. The literature review will conclude with a conceptual framework. Chapter 3 “Methodology” will handle the data collection and data processing steps that have been undertaken in order to apply the existing theories in the study area of Amsterdam. Chapter 4 “Results” will answer sub-questions 2, 3, and 4. The thesis will be closed by a Discussion in chapter 5, and a Conclusion in chapter 6.

2. Literature review

Cyclist behaviour and Space Syntax have both been researched extensively, but mostly independently from each other. Studies that combine both subjects remain rare. In this chapter, an overview will be given of the existing literature about the various determinants that influence bicycle usage. This chapter will also give a brief explanation of Space Syntax, as it will play a significant role in the rest of this research.

Relevant sub-questions:

1. What environmental factors influence cycling patterns, according to existing literature?
2. What Space Syntax measures can be used to analyse cycling patterns in urban areas?

2.1. Utility (cost, travel time, safety, effort)

When people travel from A to B, they do so in a way that is most beneficial to them. The cost, travel time, safety, and effort to cycle somewhere are often represented as the "utility" of a mode or trip. An increase in travel cost, travel time, or effort of a mode means a decrease in its overall utility (Tan et al., 2015). A plethora of studies has been conducted in order to better understand the cycling route choices of people in cities. In general, it is believed that cyclists do not necessarily opt for the geographically shortest route available to them. Various factors influence the route choices of individual cyclists. For some people, the overall safety on the route might be more important than travel time. Other studies show that cyclists are willing to cycle 3.5 minutes longer on a more comfortable route, rather than the geographically shortest route (Annema, 2015). The overall utility of a route heavily depends on individual/household characteristics, trip purpose, and spatial characteristics.

2.2. Individual/household characteristics

The choice of whether to use a bicycle or not is firstly dependent on personal characteristics, abilities, and constraints, as well as the living situation of the individual. The physical ability to actually use active modes and know how to use a bicycle, is – naturally – a big factor. The possession of a bicycle and the accessibility of bicycle parking spaces at home are positively associated with cycling. Other determinants about the relationship between the socio-economic conditions of households and cycling mode choice remain ambiguous. For example, Heinen et al. (2010) point out how gender and age play a role in the choice of people to cycle. However, in countries where cycling makes up a significant part of travelling modes, women are found to be cycling more often than men, whereas, in countries where cycling is less pronounced, men are found to be cycling more. More recent studies have criticized this male/female division and argued for deeper research into identity and cycling (Ravensbergen et al., 2019). Younger people are generally believed to cycle more than older age groups (Waygood et al., 2015), although these results are also indecisive (Ton et al., 2019a). Household characteristics are also thought to influence the likeliness of people to travel by bicycle. Having no children, for example, increases the chance of cycling (Heinen et al., 2010). The effect of income on cycling mode choice remains unclear (Muñoz et al., 2016). Some studies relate lower incomes with lower bicycle ridership, while others show the opposite effect. Car (license) ownership is believed to have a negative effect on cycling trips, as households will naturally use the private car more, once they own a private vehicle (Muñoz et al., 2016).

Personal background also plays a factor in cycling mode and route choices in the Netherlands. People from foreign backgrounds are believed to cycle less, and more often opt for “easier” cycling routes with less traffic, and fewer turns (Heinen et al., 2010).

2.3. Trip purpose

Trip characteristics are represented in the context of the journey itself. For example, the motive of the journey makes a significant difference in whether cyclists opt for the shortest route or the most direct one. Cycling commuters might perceive a more intense time constraint than someone who cycles as a form of leisure activity. Furthermore, commuters might cycle the same route to work over and over again, therefore feeling constrained by the number of turns on the route (Broach et al., 2012).

2.4. Spatial characteristics

Spatial characteristics include variables that are specific to a place or a region. In this research, spatial characteristics are further divided into a macro-level and a micro-level. Macro-level characteristics are features that urban planners or traffic engineers have little to no influence on, such as the weather, climate, or topology. Urban planners and traffic engineers do have an influence on micro-level spatial characteristics. These include transport-related issues like the surface material of roads, availability of bicycle-dedicated infrastructure, and land use-related issues like land-use mix and urban density.

2.4.1. Macro level: Weather/Climate and natural environment

Compared to the private car or public transport, the bicycle is a transportation mode that is less protected against natural elements. Weather and climate are therefore seen as having a big influence on cycling as a mode choice (Mendiate et al., 2022). Occasional cyclists seem to be more affected by weather than regular cyclists (Jandari et al., 2020). Cycling also requires more physical effort than the private cars or public transport. Multiple studies point at the importance of topography in route choices, especially in areas with steep gradient changes (Menghini et al., 2010). Some route choice models suggest cyclists would prefer to cycle 1.72 miles – or 2.77 km – over flat terrain, then to cycle 1 mile – 1.60 km – on a 2% - 4% upslope (Broach et al., 2012). Environmental characteristics are very context-dependent, as some cities might have extreme slope changes, while other urban areas consist mostly of flat surfaces.

2.4.2. Micro level: Spatial characteristics

An extensive range of studies has been devoted to the effect of micro-level spatial characteristics on the way people get around in cities (Ewing & Cervero, 2010). Spatial characteristics are divided into two categories: “land use” and “transport”.

2.4.2.1. Land use characteristics: Land use mix and urban density

Due to the physical nature of cycling as a transportation mode, distance naturally has a more profound effect than it would on other modes, like personal car and public transport. Higher built densities, and a mixture of functions contribute to lower traveling distances and are therefore believed to encourage more active travel modes (Waygood et al., 2015). Ma & Ye (2017) found that the percentage of commercial land uses within a neighbourhood contributes to a higher cycling frequency. A study in the Netherlands also found that the density of addresses correlated positively with time spent walking and cycling (Fishman et al., 2015). Koohsari et al. (2020) combine both 24-hour travel surveys and Space Syntax measures to study the likelihood of adults using bicycles for transport. They found that adults living in higher-density neighbourhoods with high Walk Scores and higher street integration are more likely to opt for the bicycle as a transportation mode (Koohsari et al., 2020). More recently, a study on socioeconomic components, urban form, and street network morphology showed how Space Syntax "connectivity" measures are positively associated with cycling for commuting (Soltani, 2022).

2.4.2.2. Land use characteristics: Urban greenery & open water bodies

Research also shows a strong correlation between the accessibility of urban greenery and physical activity among its residents, both in parks (Kaczynski et al., 2009; Lachowycz & Jones, 2011), and along streets (Lu, 2019). Moreover, a positive relationship has been found between the number of trees in neighbourhoods and cycling (Mertens et al., 2017). However, longitudinal studies have not shown any significant correlation between the increase in urban greenery and the increase in active modes (Gubbels et al., 2016; Hogendorf et al., 2020). Research into cycling in other contexts, like sub-saharan Africa and Brazil, shows how cyclists prefer cycling in tree-covered streets during high heat periods (Mendiate et al., 2022; Segadilha & Sanches, 2014). In some studies, urban greenery was actually measured in combination with "aquatic areas"(Krenn et al., 2014; Snizek et al., 2013b). In these studies, urban greenery and open water bodies were both found to be contributing to a higher quality urban environment.

2.4.2.3. Transport characteristics: Separate bicycle paths & On-street parking space

Bicycle infrastructure ties into the notion of safety and comfort for cyclists. For example, parallel parking spaces on roads can pose dangerous situations for cyclists. Stated preference surveys reveal how cyclists perceive roads without parallel on-street parking as safer than roads with on-street parking(Stinson & Bhat, 2003; Winters & Teschke, 2010). The presence of bicycle paths seems to have varying results among studies (Heinen et al., 2010). The country in which the research has been undertaken seems to have a big influence on the importance of bike path availability. In countries where bicycle infrastructure is already adequate, the presence of bicycle paths does not seem to have a large effect on cycling as a mode choice.

2.4.2.4. Street surface material & Route directness

Heinen et al. (2010) mention how few studies have looked into the effect of street surface quality on cycling. Later, Hölzel et al. (2012) point out that while asphalt gives a bigger rolling resistance than concrete slabs, cycling on asphalt is far more comfortable than cycling on concrete slabs or cobblestones. Cyclists prefer cycling on separate bicycle paths over using the curb lanes (Heinen et al., 2010; Winters & Teschke, 2010). The number of turns at intersections also plays a role. As intersections can add quite a bit of waiting time on a journey, cyclists are believed to prefer routes with fewer turns and fewer intersections. A cycling route with a high number of turns is harder to remember and the chance of mistaking a route will be higher (Broach et al., 2012). More recent studies also suggest that cyclists aim to minimise wrong turns on a cycling route (Halldórsdóttir, 2015).

2.5. Space Syntax

Before diving into the subject of how Space Syntax methods have been used to analyse cycling behaviour in cities, the theories behind Space Syntax itself must be elaborated. Space Syntax is both a theory and a tool to analyse urban space. The basic principles of Space Syntax can be discovered in two books: "The Social Logic of Time and Space" and "Space as the Machine", both works of Bill Hillier, in 1984 and 1996 respectively. Space Syntax was developed in order to better understand the effect of the built environment on people's movement, and vice versa. The objective – according to Hillier himself – was to form a truly analytic and internal theory about architecture, one that is based on buildings and the built environment itself, instead of "borrowing" theories from other disciplines like engineering, or biology, as was the tendency in the past century (Hillier, 1996). Space Syntax is an instrument that can be used to better understand the effect of the built environment on everyday life. From pedestrian and vehicle movement patterns through the city, the tool has also been used to analyse the dispersion of crime and the development of property value. At first, Space Syntax was used to analyse spatial relationships in small-scale urban contexts and buildings. Later, advancements in computing power have made it possible to analyse more complex, urban regions (van Nes & Yamu, 2021).

The fundamental proposition of Space Syntax is that movement is what shapes socioeconomic factors in cities and that space is the biggest determinant of movement – both pedestrian and vehicular (Hillier,1996). Since its inception, the theory has been embraced by a plethora of academics, architects, and engineers in a wide range of urban contexts, ranging from individual buildings to larger metropolitan areas.

The Space Syntax tool is a method to quantitatively analyse urban spaces and their movement potential. This is done by drawing axial lines as the representation of the longest line of sight in a street. In this way, the number of connections of a street from all other streets – connectivity – can be calculated. With the connectivity of each street in mind, one can calculate the number of directional turns it takes to get from one street, to any other street in the network. This measure is called "integration" (Yamu et al., 2021). Figure 1 (a) shows a hypothetical map, with streets and buildings clearly visible. In Figure 1 (b) and 1 (c), the streets are represented as axial lines. In figure 1(d), the same network is visualised in a justified-graph form. The individual streets are represented by points, and the connections between the streets as lines. The justified graph allows a researcher to easily calculate the integration of an individual street.

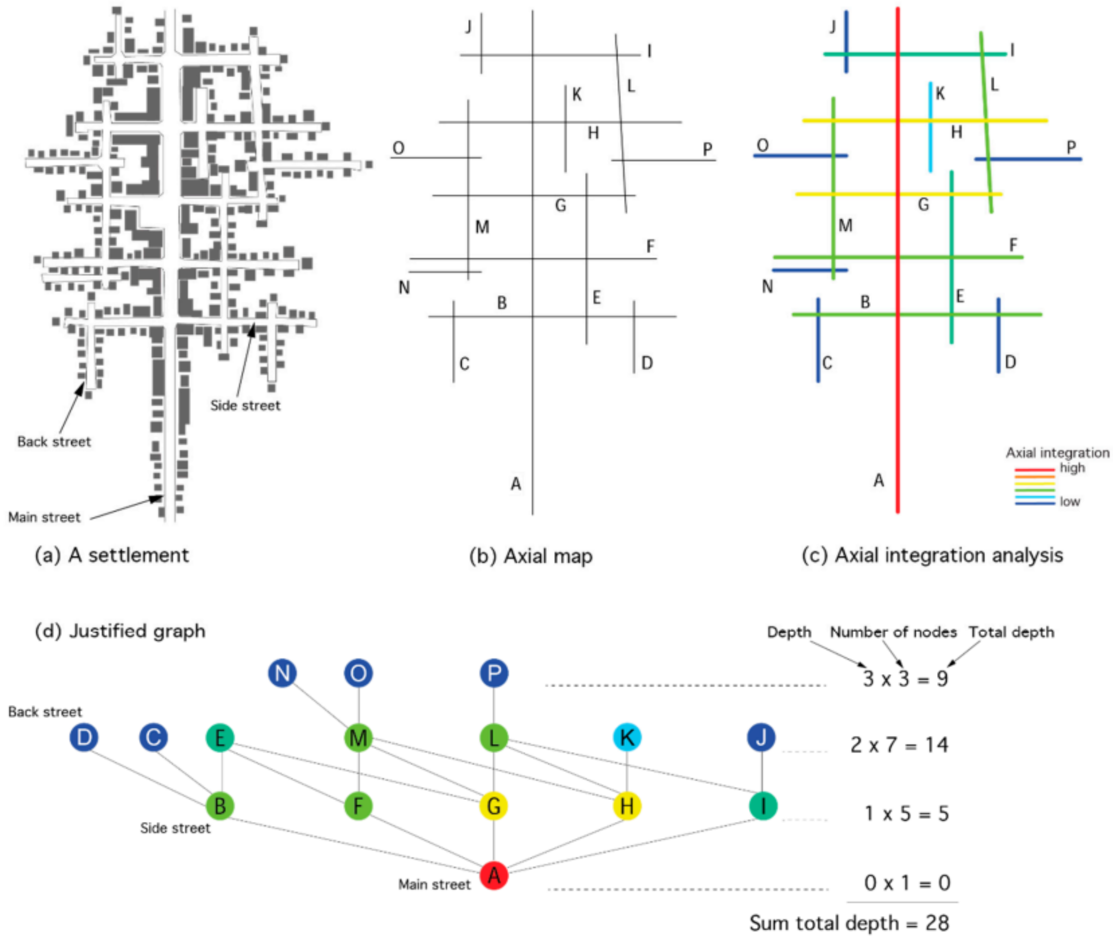


Figure 1: Example of an urban street network, and its axial representation (From Yamu & Van Nes, 2021).

One of the bigger criticisms on Space Syntax was the lack of consideration of metric distance (Ratti, 2004). Distance was only seen in terms of number of turns, instead of geographical distance. Later, software made it possible to break down an axial line into multiple segments, allowing the use of metric distances in segment analyses, and even considering the angle of each turn. The angular choice measure, as it became to be known, shows the "through-movement potential" of each segment in a street network**: It looks at the number of times each street segment is on the path of least angular deviance (or straightest route) between all pairs of segments within a given radius (r)** (van Nes & Yamu, 2021). Angular integration is a measure that depicts the "to-movement potential" of a street segment. It looks at the distance from each segment in a network to all other segments (Hillier et al., 2012). The "distance" refers to the topological distance, as the number of turns, and the angle of those turns is taken into account (Yamu & Van Nes, 2021).

Axial maps can be time-consuming to construct, especially when working on a higher, metropolitan scale (Van Nes & Yamu, 2021). That is why for some studies with larger study areas, the road-center lines are used as input for constructing maps. The usage of road-center lines was proven to be compatible with Space Syntax', angular choice and integration measures (Turner, 2009). Later, Hillier, Yang, and Turner (2012) proposed the normalisation of angular choice and angular integration measures.

Space Syntax and cyclist analysis

Space Syntax has been applied to study vehicle traffic on the city level, and pedestrian traffic at local neighbourhood levels. Global integration analysis tends to be used when researching vehicular traffic, for pedestrian analysis, local integration is mostly selected (Liu Ziqi Song et al., 2016). Cycling finds itself in the middle of those two types of traffic. Various studies have looked at different Space Syntax measures relating to cycling mode choices, as well as route choices (McCahill & Garrick, 2008; Raford et al., 2007; Rybarczyk & Wu, 2014; Soltani et al., 2022). Still, studies that relate cycling behaviour with Space Syntax measures remain rare. For one reason, few countries see a bicycle ridership that is as high as the Netherlands, therefore data is hard to come by. In fact, Liu et al. (2007) mention this as one of the big limitations in their research, analysing cyclists' route choices in Salt Lake City.

Raford et al. (2007) were among the first to use Space Syntax to compare street accessibility with the route choice of cyclists. In their study, axial analysis and angular segment analysis were used to analyse the street accessibility of the streets of London. Their study was divided into two parts: a part that analysed the total routes of individual cyclists in central London, and a multiple regression analysis that used actual gate counts as input. In the analysis of individual cyclists in London, Space Syntax was used to calculate the "fastest" path with the fewest turns. Next to this, a GIS tool was utilized to calculate the shortest path purely on distance. These two routes were compared with the actual route that was taken by a cyclist. The results show that both the shortest metric distance and the "fastest" route show little correlation with the actual route chosen. The sample size, however, only consisted of 46 routes. As for the analysis of the gate counts, Raford et al. (2007) found a strong correlation between global mean angular depth and route choice of cyclists. Liu et al. (2017) opt for the inclusion of local integration to analyse the bicycle route choice of people. McCahill & Garrick (2008) use the "choice method" and "angular segment analysis" to construct a model that predicts bicycle volumes through a network. A linear regression model is created, using population density, worker density, and Space Syntax "choice" measure, that accurately predicts cyclist volumes. Their research does include comparing the segment angular choice map to the actual routes as cycled through the city of Trondheim, but it does not result in a model as presented by McCahill & Garrick (2008). As cycling is not a very prevalent mode of transport in the case studies of the studies mentioned earlier, small datasets are a limitation that is mentioned in scientific studies.

Rybarczyk & Wu (2014) make use of a different Space Syntax measure – Visual Graph Analysis – to examine the impact of urban morphology on people's bicycle mode choice decisions. Unlike axial maps, Visual Graph Analysis is a raster-based analysis in which a space is subdivided into multiple rasters. Instead of analysing the street network of a city as a whole, their research focused on the micro-level of urban design.

A study in London, comparing cycling counts in 2003 and 2012, found that Space Syntax – in combination with other variables – could explain 65% of cyclist movement. Furthermore, direct and continuous routes were found to accommodate more cyclists than routes with better cycling infrastructure, that were less direct (Law et al., 2014). "Normalised Angular Choice" (NACH) was used to measure the "directness" of each route. A study in the city of Manta, Ecuador, concludes that the street network's structure – measured with Space Syntax – does indeed have an influence on cycling activity (Orellana et al., 2019). This study also used NACH, as well as NAIN to measure directness.

Conceptual model

Figure 2 visualizes the conceptual model of this theoretical framework. Starting at the very top, the utility of a cycling route is dependent on both the individual cyclist, as well as the spatial characteristics in which the route has to take place. Furthermore, the spatial characteristics can be divided into a macro-level and a micro-level. Macro-level spatial characteristics entail the climate, weather, or topological features within a given context. Naturally, urban planners and engineers have little influence over these place-dependent properties. On the other hand, the characteristics on a micro-level are variables that urban planners and engineers often do have an influence on. The focus of this thesis will lay on the variables that fall into that category. Based on prior research, NACH and NAIN will be used to measure route directness.

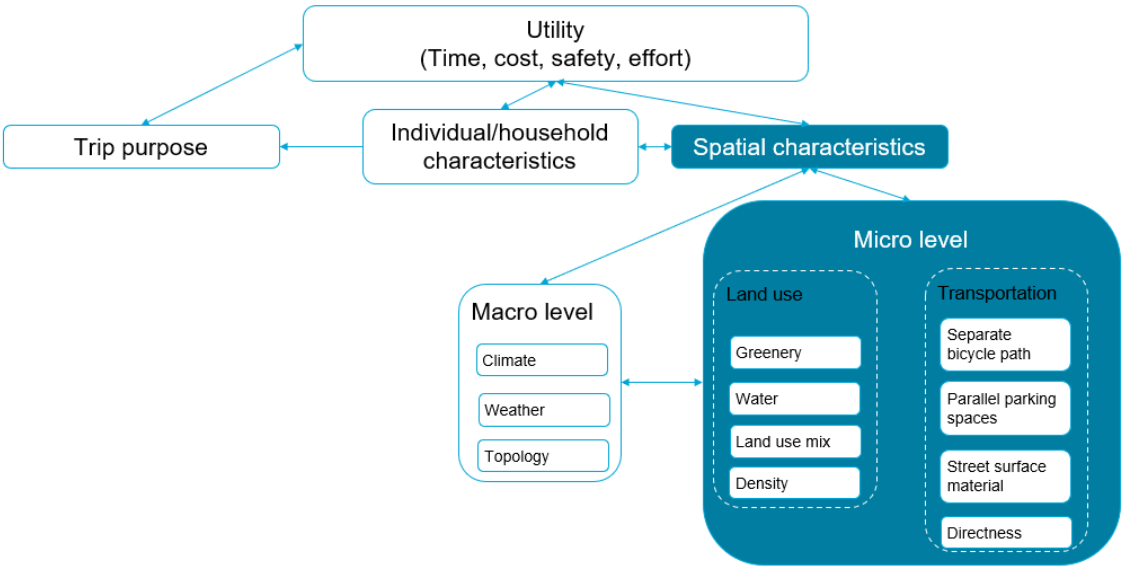


Figure 2: Conceptual model, the relevant variables for this research are shown in the blue box

Table 1: Variables considered in this research

Variable	Literature	Conclusion
Urban greenery	(Kaczynski et al., 2009; Lu, 2019; Mertens et al., 2017; Winters & Teschke, 2010)	Trees along streets attract physical activity (walking and cycling)
Open water bodies	(Krenn et al., 2014; Snizek et al., 2013)	Open water bodies are characterised as more attractive, and lead to more cyclists and pedestrians
Separate cycle paths	(Heinen et al., 2010; Winters & Teschke, 2010)	Cyclists prefer to cycle over protected, separated bicycle paths
Street surface material	(Hölzel et al., 2012)	Cyclists prefer to cycle over smooth surface material like asphalt, over other materials
On-street parking	(Winters & Teschke, 2010)	Cyclists tend to avoid streets with on-street parking facilities
Land-use mix	(Ma & Ye, 2022; Waygood et al., 2015)	A mixture of functions reduces the need to travel long distances, leading to more active modes
Urban density	(Fishman et al., 2015; Koohsari et al., 2020)	Higher urban density attracts more active travel, including cycling
Directness	(Halldórsdóttir, 2015; Law et al., 2014; McCahill & Garrick, 2008; Orellana et al., 2019)	Cyclists choose for direct routes through cities, and can be measured by Space Syntax

3. Methodology

The first sub-question of this thesis has been answered through a review of relevant academic literature. The empirical part of this research will be based on GPS data from the Bicycle Counting Week. The dataset of the BCW was required some data processing steps in order to perform the Space Syntax analysis, as well as. Individual linear regression was used to Multiple linear regression was used to examine the relation between Built Environment variables and the cycling activity per street segment. In this chapter, the data acquisition, data processing, and methodology of the data analysis steps will be elaborated on in more detail.

3.1. Literature Review

The acquisition of relevant academic literature was first carried out by relying on Scopus. The keywords to find literature about cycling and Space Syntax were “cycling” AND “route” AND “Choices” AND “Space” AND “Syntax”. The selection was narrowed down to only include literature from 2015 and 2022. This query gave four results. The second search included only “cycling” AND “Space” AND “Syntax”, resulting in 18 document results. In order to find the latest literature on cycling route choice modelling, the keywords “cycling” AND “route” AND “choice” AND “modelling” were used. The literature was further narrowed down to include only “Social Sciences”, “Engineering”, and “Environmental Sciences”. This resulted in 30 document results.

Other articles were found by making use of a “snowball” method, in which the references of academic literature that was found in the first phase were used. Two books were selected as a starting to better understand the theories and methodology behind Space Syntax: Bill Hillier’s “Space is the Machine” (1996), and the more recent “Introduction to Space Syntax in Urban Studies” By Yamu & Van Nes (2021). Especially the work of Yamu & Van Nes (2021) offers a clear explanation of the core concepts of Space Syntax and gives insight into the numerous developments that the method has gone through over the years.

3.2. (Multiple) Linear Regression

The goal of this research is to explain the number of cyclists per street segment. As elaborated in the literature review, this is done by looking at the relationship between cycling activity and spatial variables. In behavioural science and engineering, linear regression is often used to quantitatively explain or predict certain trends or phenomena in the real world (Myers, et al., 2010). Regression is used to examine the relationship between one or multiple independent continuous variables (X) and a dependent continuous variable (Y). Linear regression only describes a relationship and does not say anything about causality. However, through inference, one can reasonably assume causality. For example, in the case of a statistically significant positive relationship between cycling counts and urban greenery along a street segment, one can reasonably suggest that more greenery will lead to more cyclists opting for that route, rather than an increase in cyclists leading to more urban greenery. Of course, the choice of people to cycle through one street, instead of the other, is almost never dependent on just one variable. When two or more variables are used to explain a certain phenomenon Y, one speaks of Multiple Linear Regression.

Multiple Linear Regression is written as:

$$Y = \beta 0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

With Y as the dependent variable, $\beta 0$ as the constant, $X_1, X_2 \ldots X_n$ as the independent predictor variables, $\beta_1, \beta_2 \ldots \beta_n$ as the unknown predictor factors, and ε as the error term.

The null hypothesis states that the independent and dependent variables do not hold a statistically significant relation. If the p-value, representing the significance levels between two variables is under 0.01, the dependent and independent variable(s) do indeed hold a significant relation and the null-hypothesis is rejected.

Another aim of this research project is to see how "well" a variable is able to explain the number of cyclists passing a street segment. Through the use of the coefficient of determination (R^2 and adjusted R^2), one can determine how much of the variation of the dependent variable can be explained by the independent variable(s).

Multiple models can be compared with each other in order to see which model explains the dependent variable better. The Bayesian Information Criterion (BIC) was used to compare models. BIC does not show the extent of the correlation between independent and dependent variables but do give insight into which model can explain an dependent variable better, given the number of parameters in the model.

3.3. Data acquisition and processing

In order to answer the previously stated research questions, quantitative data has to be acquired and analysed. As mentioned before, data from the BCW was used to represent the cycling route choices. Data from the "Basis Registratie Grootschalige Topografie" (BGT) was used for geodata on urban vegetation, open water bodies, and street parking spaces. The "Ruimtelijke Dichtheden en Functiemenging in Nederland" (RUDIFUN) dataset was used for geodata on urban land use mix and open space ratio. All this data had to be prepared and analysed, using various different software programs. The preparation of BCW has been done in R Studio. The preparation of the built environment variables was done in ArcGIS Pro. Finally, QGIS, with its many plug-in options, was used to prepare the Space Syntax analysis. Table 2 shows the way in which the variables are measured. "Directness" will be measured by the Space Syntax value.

Table 2: The variables used in the empirical research

Variable	Source data	Source holder
Cycling counts	BCW	Fietstelweek.nl
Separate cycle path	BCW	Fietstelweek.nl
Trees	BGT (VegetatieObject_p)	Kadaster
Open water bodies	BGT (Waterdeel_v)	Kadaster
Parallel parking spaces	BGT (Wegdeel_v)	Kadaster
Surface material	BGT (Wegdeel_v)	Kadaster
Mixed-use index	RUDIFUN (MXI)	PBL
Open Space Ratio	RUDIFUN (OSR)	PBL

3.3.1. Description and preparation of BCW data

The "Nationale Fietstelweek", or BCW is held annually. Volunteers can track and share their cycling patterns by downloading an app on their smartphones. The GPS data is processed and transformed into "Bike PRINTS" that can later be used as input for policy making (Van de Coevering, de Kruijf & Bussche, 2014). In 2016, the BCW spanned from the 19th until the 25th of September. The collected data is fully anonymized by omitting the first and last 100 to 300 meters of a cycling trip. To prevent any irregularities, trips that were shorter than 500 meters were not included in the data (Van de Coevording, de Kruijf & Bussche, 2014). Also, no personal information can be traced back to any individual participant. Each cycling trip is therefore considered as unique. The collected GPS data is cleaned and coupled to the street network, available on OpenStreetMap (OSM). The OSM data contains information about the road type –

i.e. primary, secondary, living street, or cycleway. The resulting network is what was used in this research.

The BCW data came in the form of two different datasets: A shapefile representing all the individual links (street segments), with corresponding intensities of cyclists using that link in the counting week. The shapefile also stored information about the average speed and function of the link. The second dataset consisted of a CSV file with all the routes. This file included the route IDs, followed by the Link number of the links that made up the route, and the average speed, day, and time of the week. Each route consisted of multiple links, and each link could have multiple routes passing through them. Combining these two datasets turned out to be a more tedious task than expected. The two datasets were joined, using R Studio. A full join, based on "Linknummer" made it possible.

Upon close inspection of the data, some of the links were designated as "primary" roads, or "primary links". In the Netherlands, strict rules are in place for the placement of protected bicycle lanes next to roads (Ton et al., 2019). Cyclists are not allowed to be on the primary roads, and some secondary roads if there is a separate cycling path next to it. The links that were classified as "primary", "primary links", "secondary", and "service" were therefore deleted from the network, if within close distance to a parallel separate cycling path.

3.3.2. Data preparation steps for built environment variables

The ArcGIS Pro Model Builder was used to prepare the built environment variables. All the data from the BGT and RUDIFUN already had geography, eliminating the need for programs like R Studio and Python for data preparation. Furthermore, the data preparation would require the use of spatial joins, making the use of Geographic Information Systems a necessity. ArcGIS Pro was chosen because of the possibility to connect to various data sources through its "Live Atlas" interface and its overall robust model builder option, making every step in the process comprehensible and repeatable.

The BGT is a detailed, open-source map of the Netherlands. It contains information about roads, buildings, open water bodies, and vegetation. As was mentioned above, the data that was relevant for this research were the roads ("wegdeel_v") polygon dataset, the trees ("vegetatieObjecten_p") point dataset, and the open water bodies ("waterdeel_v") polygon dataset. BGT data can be downloaded from either RUDIFUN data from the PBL was used to measure the rate of land use mix (MXI) and the urban density (OSR).

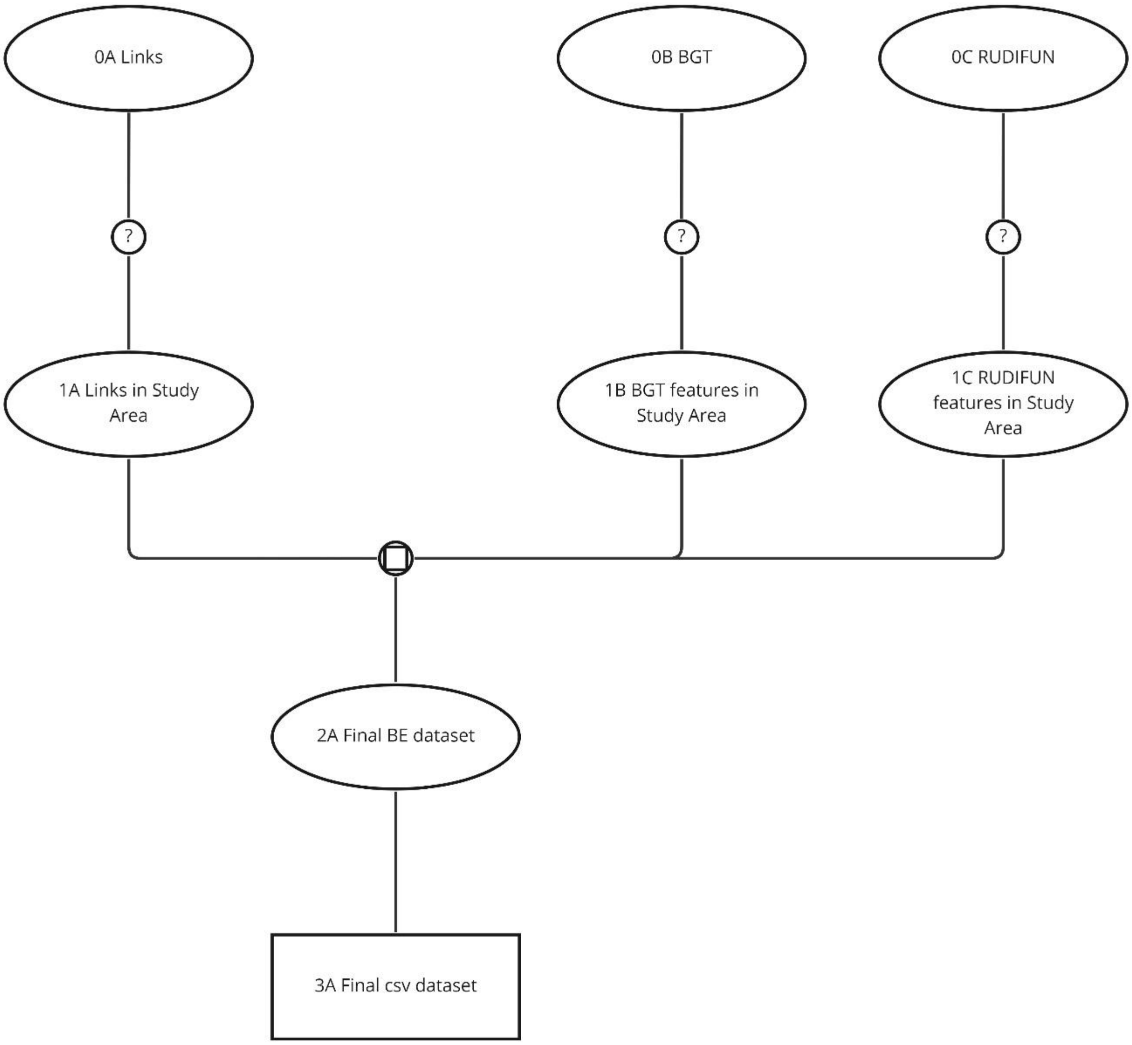


Figure 3: Flowchart of the data processing steps for the BCW, BGT, and RUDIFUN data

3.3.3. Preparing the "study area"

Before any analysis can be carried out, the study area has to be defined. Coming to a proper study area required multiple steps. First, an initial study area was constructed by taking the map of the Municipality of Amsterdam from 2016 as a polygon and defining the centroid. Then, a buffer was created around the centroid, with a radius of 15 kilometers. This radius of 15 kilometers around the centroid of the municipality was taken as the study area. Upon closer examination, it was found that there was a discrepancy between the precision with which the street segments in the city centre of Amsterdam were constructed versus the outer parts – Nieuw West, Noord, and Zuidoost – of the city. It was therefore decided to specifically look at the routes of cyclists that went through the city centre. To achieve this, the links of the Bicycle Counting Week were clipped to the borough of "Centrum" Amsterdam. The links in the city centre were then joined with the "routes" dataset, to find out what the average length of cycling routes through the city centre was. The average length of cycling routes through the city centre was 3.6 kilometres, this distance was taken as the radius for a new buffer around the centroid of the municipality of Amsterdam, minus the borough of Amsterdam Noord.

MXI & OSR

The Mixed Use Index (MXI) and Open Space Ratio (OSR) were used to measure the mixture of functions, and the urban densities along a route. All the data is part of the RUDIFUN dataset, provided by the PBL, or Dutch Living Environment Planning agency. The PBL provides RUDIFUN data on varying scale levels, from the level of individual building blocks to the level of municipalities. Furthermore, a distinction is made between "bruto" (gross) values and "netto" (net) values. In the values of "gross" densities, the space that takes up the streets is taken into account, whereas in the "net" densities, only the surface area of the buildings is considered.

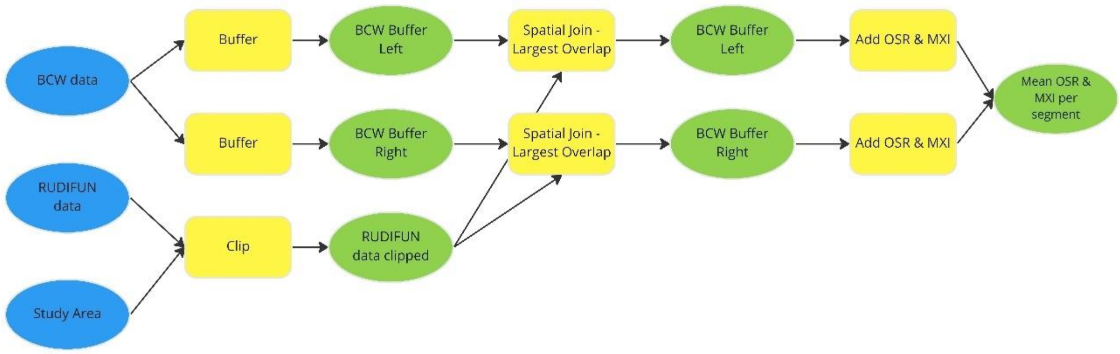
MXI represents the ratio of space meant for housing, relative to other functions in the building block. The MXI has a score between "0" and "1". An MXI of 0 means a building block has no space allocated for housing. An MXI of 1 means all of the floor space in a building is used for housing (Harbers et al., 2022).

$$MXI = \frac{Gross\ floor\ space\ for\ housing}{Gross\ floor\ space\ total}$$

The Open Space Ratio was taken as an indicator of the urban density around the cycling links. The indicator represents the area in a given building plot that is "empty".

$$OSR = \frac{Unbuilt\ terrain\ area}{Gross\ floor\ space\ area} = \frac{1 - GSI}{FSI}$$

Two buffers were made around the cycling links, one "left buffer", and one "right buffer". In many cases, a street segment would have a building block on two sides of the street. The "Spatial Join – Largest Overlap" tool, can only join one feature of the target data, to one feature of the input data, therefore ignoring the fact that there are two sides to each streets. This issue can be overcome by constructing two buffers for each link, one on the left side, and one on the right side. The values of the two buffers can later be added together in order to create a "mean MXI" and "mean OSR".



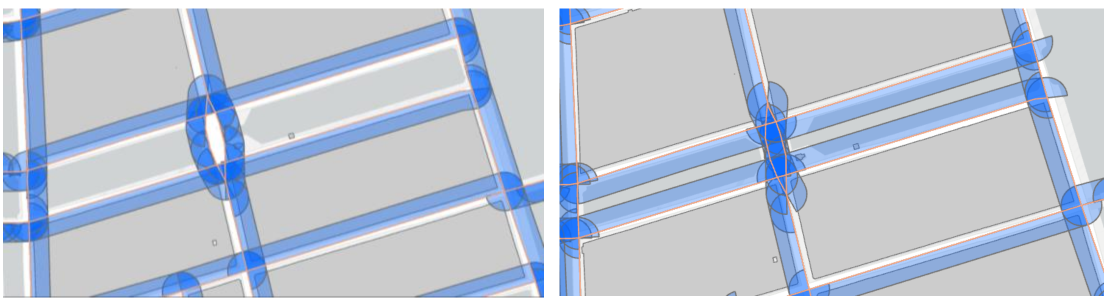
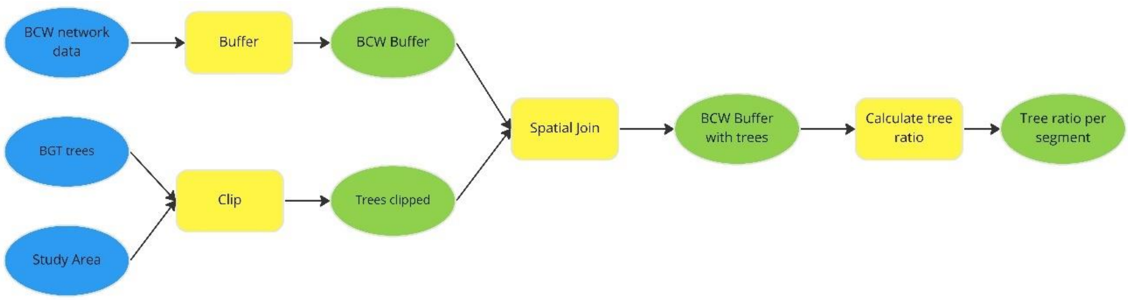


Figure 4: Left and right buffers for measuring mean OSR and MXI

Trees

The "VegetatieObject_p" data of the BGT dataset represents trees as point features in the Netherlands. This dataset was reached by connecting to the BGT geodatabase through ArcGIS Living Atlas. The first step is to clip the "VegetatieObject_p" data to the Study area. The objective of the data preparation was to come to a comprehensive dataset in which each link would have an attribute with the ratio of trees per street, divided by the length of the street. To achieve this, a buffer around the cycling links in the Study area was created. The buffers share the same fields as the corresponding links, therefore, the count of trees inside each buffer element could be divided by the "shape length", resulting in trees per segment length ratio. For the buffer size, the distance had to be large enough for cyclists to see the trees, but small enough to be considered next to the street segment. A buffer size of 15 meters was decided upon. The figure below shows the steps that were taken to come to the prepared data for vegetation objects in the study area

$$vegetation\ point\ ratio = \frac{trees\ along\ segment}{segment\ length}$$



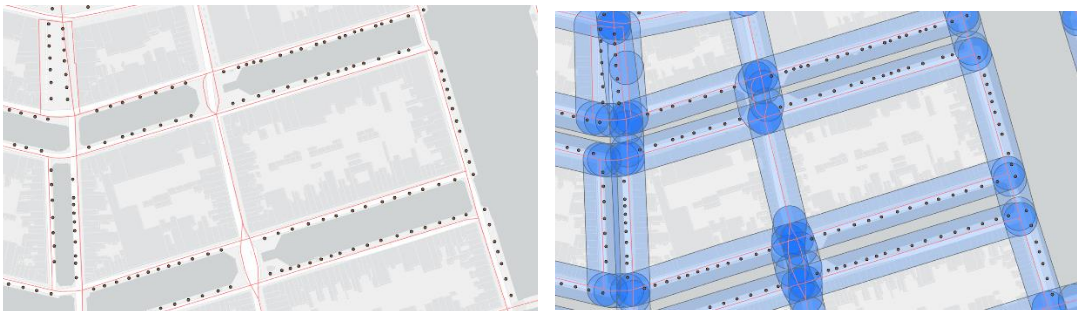
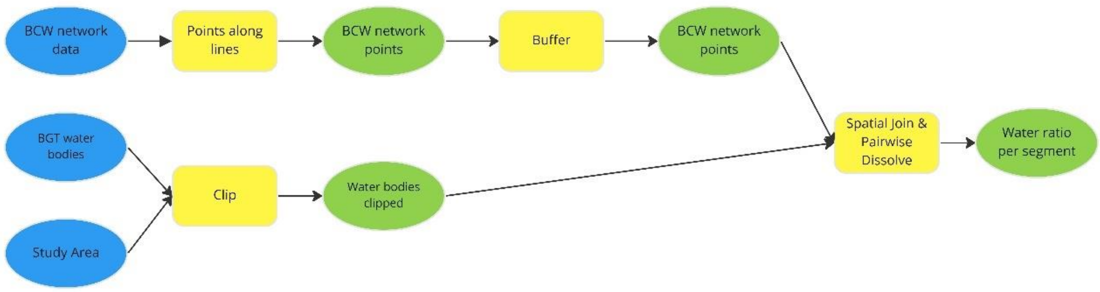


Figure 5: The trees were visualised as points and buffers were created around all street segments

Open water bodies

The "waterdeel_v" polygon data of the BGT dataset represents all the open water bodies in the Netherlands. As the water bodies were represented in the form of polygons, the analysis differed from the data preparation steps for the "vegetation objects". Instead of directly creating a buffer around each link, the links were divided into points using the "Points along Lines" tool in ArcGIS Pro. The points along the links were 5 meters apart from each other. Then, a buffer of 15 meters was created around each point. A new field was created in the point buffers, representing whether the buffer intersected with the water body or not. Then, the number of buffers that did intersect, could be divided by the total number of buffers per link, creating an "open water body ratio" per link.

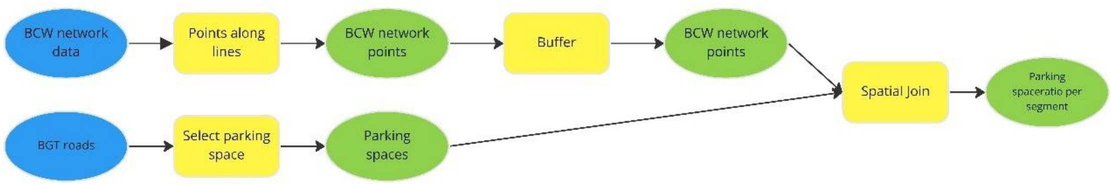
$$Open\ water\ body\ ratio = \frac{Points\ in\ proximity\ of\ waterbodies}{Total\ number\ of\ points}$$



On-street parking spaces

All public on-street parking spots in the Netherlands are represented in the BGT wegdeel_v dataset. For cyclists, only the parking spots that are right next to the cycling area pose a danger of cars crossing the cycling lane. The buffers have to be smaller than the buffers that were used for the data preparation of the open water bodies and urban greenery. Just as in the data preparation step of the open water bodies data, the cycling links were divided into points that were spaced 5 meters from each other. A radius of 3 meters was chosen for the buffers around the aforementioned points. From the BGT_wegdeel_v dataset, only the on-street parking spaces were selected using the “Copy Features” tool, resulting in a layer with all the on-street parking spaces in the study area. A new field was created in the layer of the parking spaces and the layer of the buffers, representing the count of each feature. The next step was to perform a spatial join between the layer of the parking spaces and the buffer features. In the same way as the open water bodies data, the number of buffers per link that did intersect with parking spaces was divided by the total number of buffers on that link, creating a parking space ratio field.

$$On\ street\ parking\ space\ ratio = \frac{parallel\ parking\ spaces}{segment\ length}$$



Separate cycling paths

The GPS data of the BCW were coupled to geodata of Open Street Map. The links of the BCW data, therefore already has a field with information on the function of a link area called “HIGHWAY”. The “HIGHWAY” field shows the function of the street segment that each link passes, for example, a cycleway, residential road, or footway. The frequency table shows that the most common function was “cycleway”, with 20.477 links represented in this category.

Surface material smoothness

One of the variables that are represented in the BGT wegdeel_v data is the surface material of every street in the Netherlands. A distinction can be made between four different types of surface level: unpaved (“Onverhard”); semi-paved (“half verhard”); open pavement (“open verharding”); and closed pavement (“gesloten verharding”). An example of unpaved street segments or paths is a sandy path or a path that has tree bark as its surface material. Half-paved paths have gravel, pebbles, or shells as surface material. Open pavement is the most used surface material type in the study area. It is mostly made up of street segments with clinker brick and tiles, which are seen in many cities in the Netherlands. Finally, the closed pavement category consists of asphalt and concrete.

The surface material data was prepared by making use of the "Spatial Join – Largest overlap" function in ArcGIS Pro. This is a Geoprocessing sample that can be downloaded separately from the ArcGIS hub, and used in ArcGIS Pro. As the name suggests, a spatial join is created for features that share the largest spatial overlap. The Clipped network data was taken as the target value, and the BGT wegdeel data was set as the join feature. The resulting shapefile was reclassified, to have two surface material types. The surface material types were then ranked from highest to lowest, the most comfortable surface material type ("closed pavement") having a score of 1, and the rest having the score of 0.

3.3.4. Data preparation for Space Syntax Analysis

The workflow of a Space Syntax analysis requires a different set of tools than the previously mentioned analyses. For the Space Syntax analysis, a combination of QGIS and various QGIS plug-ins was used. A more elaborate explanation of the steps conducted for this analysis will be presented below.

The analysis began by importing the clipped Bicycle Counting Week network into QGIS, as well as the clipped "wegdeel_v" polygon dataset of the BGT. As the Space Syntax tool only supports single-line features, the polylines (lines connecting more than two points) had to be turned into multiple single line features. The first step was therefore to use the "Explode lines" processing tool in QGIS to split the polylines into multiple single lines. Upon closer inspection of the data, one of the issues with the dataset was that some of the cycling routes also consisted of ferry rides connecting Amsterdam North with the rest of the city. A large portion of cyclists uses the Amsterdam ferries every day to get from the North to their destinations in the rest of the city and vice versa. As ferries are naturally not part of the street network of the city, these links had to be removed in order to conduct the Space Syntax analysis. Third, the BCW network also showed links that were located inside train stations and metro/tram stops. Moreover, some of the links were located inside areas that were not publicly accessible, like the Artis Zoo in the east of the city. These links were removed from the map as well. The fourth step was to identify the unlinks in the study area.

Unlinks represent the places where two street segments seem to cross each other, but are actually on different levels. An example is a tunnel for cyclists passing under a residential road. In Space Syntax, these instances have to be manually identified, in order to make sure that these crossings are not considered level crossings, wrongly impacting the results. The Place Syntax software program, developed by the Spatial Morphology Group of the Chalmers University of Technology, has a function to automatically import the unlinks corresponding to a dataset. However, in this research project, only the streets and roads accessible to cyclists were analysed, meaning some of the automatically generated unlinks were not useful for the BCW data. In the end, it was decided to manually construct the "unlink" points. To identify the unlinks, the "wegdeel_v" polygon data of the BGT dataset was used. This data presents the streets and roads of the Netherlands in the form of polygons. Bridges, tunnels, and viaducts can be identified when symbology is categorised for "subtype". On places that were "Niet vlakopdelend", and where two lines crossed, a point was put in place. The point features would function as the unlinks in the Space Syntax analysis.

Figure 6: Unlinks (in yellow) were identified at intersections marked as "niet-Vlakopdelend" (in red) in BGT data

The fifth step was to check the validity of both the network, as well as the unlinks in the DepthmapX SST Plug-in for QGIS (Gil et al., 2015). According to Gil et al. (2016), the verification tool in the DepthmapX plugin verifies layers on seven different attributes:

- "small line", when lines are below a minimum of 1 meter long
- "polyline line", when lines are made up of more than two nodes
- "coinciding points", when two points on a line are coinciding
- "duplicate geometry", when two lines share exactly identical geometries
- "short line", when a line has endpoints close to another line without intersecting it
- "orphan", when a line does not intersect with any other line
- "island", when a group of lines is completely separate from the main network map.

The verification toolkit is mainly used for the analysis of axial maps, but it can be a handy tool to identify "island" and "orphan" links in segment analyses as well. After running the verification, three different cases of "islands" were identified, these were deleted from the network.

For the actual Space Syntax analysis, the Place Syntax Tool (PST) plugin by the Chalmers University of Technology was used. This plugin is freely available from the School of Morphology Group (SMoG) of the Palms Chalmers University of Technology. Following the research of Orellano & Guerrero (2019), Normalised Angular Choice and Normalised Angular Integration were taken as Space Syntax measures. A selection of radii of 250, 500, 1250, and 2500 meters was taken to perform the analysis. Taking the average speed of cyclists as 15 kilometres an hour – or 250 meters per minute – the radii constitute 1-, 2-, 5-, and 10 minutes. As the PST plugin only handles "Linestring" geometry, "MultiLinestrings" of the study area had to be converted to "Linestrings" using the "Multipart to Singleparts" geoprocessing tool in QGIS.

3.4. Data modelling strategy

A total of 16.685 individual links were present in the study area. In order to reduce the number of data points, a random sample of 10% was selected to be analysed in the linear regression stage.

3.4.1. Modelling strategy

First, the Space Syntax measures for the different radii were all plotted against the cycling counts per street segment. This was done in order to distil which radius had the strongest correlation with cycling counts. The Space Syntax measure with the highest correlation with cycling counts was then selected to be included in the multiple linear regression model.

The other spatial variables were also each modelled individually against the cycling count per street segment. The variables that did not show a significant relation with cycling counts were filtered out. Then, the remaining significant variables were combined in pairs to come to a Multiple Linear Regression model. The Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) were used to compare models. Both BIC and AIC are measuresthat compare models for the goodness of fit to predict data. BIC tends to penalize the addition of an extra predicting variable in a model. The modelling strategy is visualized in figure 7. BIC and AIC are used to compare models that predict the same phenomenon for goodness of fit. However, both measures do not say anything about the *explanatory* power of a model. R^2 was used to uncover the explanatory power of each model. R^2, or the coefficient of determination explains how much of the variance in one variable, is due to the other variable(s).

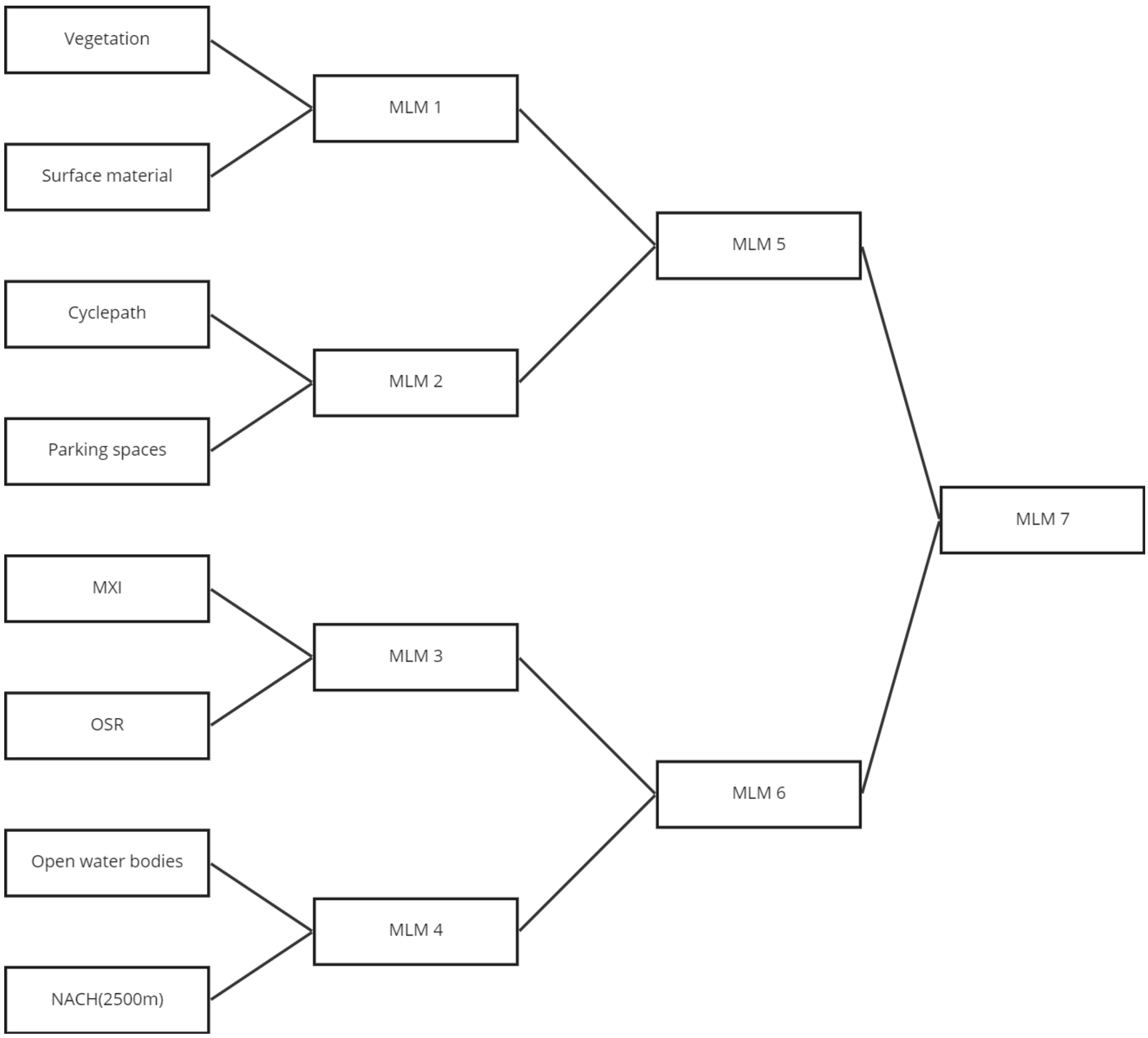


Figure 7: Schematic visualization of the modelling strategy

4. Results

This chapter will firstly touch upon the city of Amsterdam and trends related to cycling in this city. Facts and figures about mobility and accessibility in the municipality of Amsterdam are given in the “Amsterdamse Thermometer van de Bereikbaarheid” (ATB). Then, the results of the various analyses of this research project will be described in section 4.2. First going through the individual regression analyses, and then combining variables, in order to come to a multiple linear regression model.

Relevant sub-questions:

> 2. What Space Syntax measures can be used to analyse cycling patterns in urban areas?
> 3. To what extent do Space Syntax measures explain cycling route choices in Amsterdam?
> 4. To what extent do spatial characteristics explain cycling counts per street segment in Amsterdam?

4.1. Trends related to cycling in Amsterdam

In 2016, the municipality of Amsterdam had a total population of 834.713. The municipality comprises seven boroughs (stadsdelen). The city centre: “Centrum”, and the inner boroughs of “East”, “West”, “South” are roughly located inside of the highway A10, and the boroughs of “North”, “New West”, and “Southeast” are seen as outside of the highway A10. For this research, the study area covers the boroughs that are within the highway A10.

Since 2003, bicycle ownership has steadily increased among Amsterdam residents, while the percentage of Amsterdam inhabitants owning a private car has shown a decrease. Due to the overall population growth in the city, the number of private vehicles has still gone up since 2003. The borough with the most separate cycle paths was the borough of “Nieuw West”. While “Centrum” had the lowest number of kilometres of separate cycle paths, it did have the highest percentage of intensely used cycle paths in the whole city. According to the ATB (2019), the bicycle was the most used form of transport among residents of Amsterdam. Cycling accounted for 35% percent of all trips in the city. The modal split differs per borough, however, as the boroughs in the inner city show higher cycling percentages than boroughs outside. Furthermore, cycling adoption also differs among population groups, as highly educated inhabitants tend to cycle more than other groups.

The ATB also shows that on average, inhabitants who cycle on a regular basis feel healthier than inhabitants who do not. It is important to note that it is not clear whether cycling does indeed improve the health of people, or that people who do not feel well, are not able to cycle. Despite cycling making up the largest part of trips in Amsterdam, cyclists feel less safe compared to other modes of transport users.

In 2017, the municipality of Amsterdam has set three main goals for the future of its cycling infrastructure in the “Meerjarenplan Fiets 2017-2022”. The three goals are: “more comfortable routes for cyclists”, “easier bicycle parking”, and “attracting new cyclists”. Some of the measures include attached to these goals are: connecting existing cycling routes through the city, constructing more attractive cycling routes, and creating more space for cyclists on the busiest cycling routes.

4.2. Descriptive statistics

A total of 16.994 links were in the study area. The link with the highest cycling count recorded a total of 1.329 cyclists during the Bicycle Counting Week. The average number of recorded cyclists during the week in the study area was 84.

No. of links	16.685
Max. Count	1329
Mean Count	84
Median Count	37

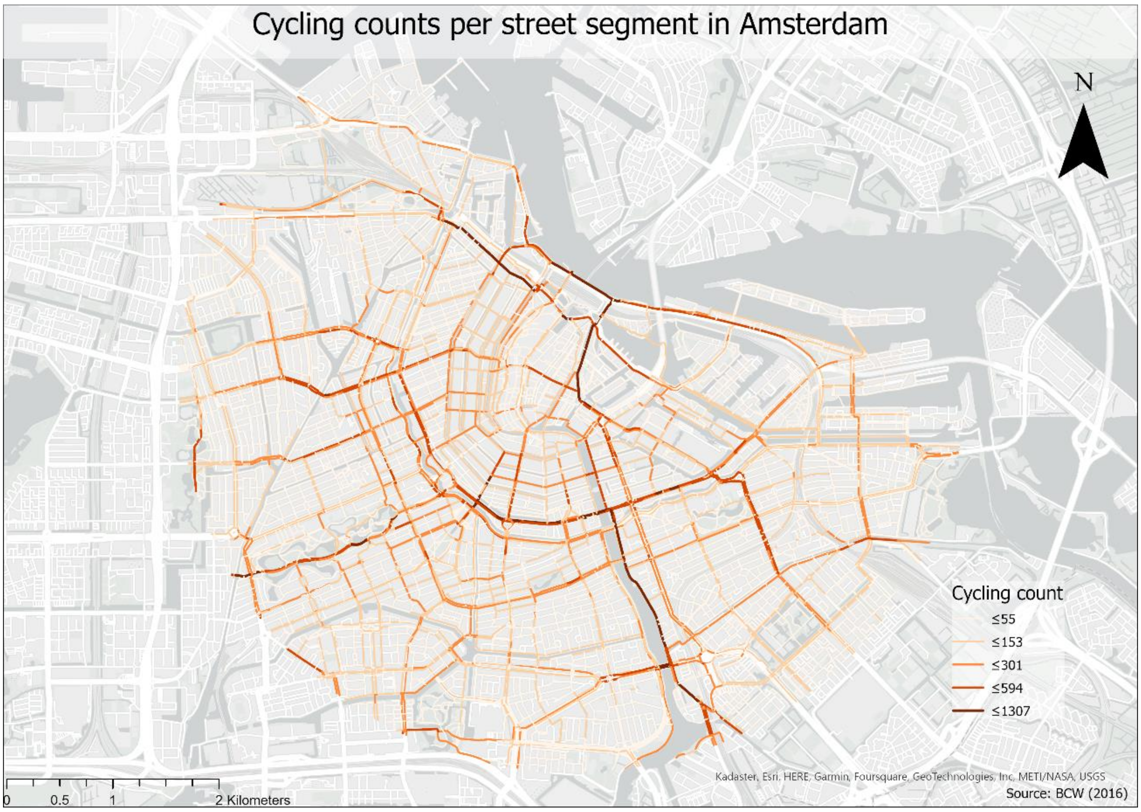


Figure 8: Map of the cycling counts per street segment in the study area

The map in figure 8 shows the activity counts for street segments in the study area of Amsterdam. The map is set so the main routes through the city are clearly visible. The thicker and darker coloured lines represent the busier street segments, while the thin and light lines represent the street segments with fewer cycling counts during the Bicycle Counting Week. The bigger roads radiating out of the city centre have a higher count than the smaller streets in between. The busiest cycling streets are along the eastern bank of the Amstel river, the Stadhouderskade south of the city centre, the Geldersekade, along with the Prins Hendrikkade right behind the Central station.

A correlation plot is visualised in the figure 9, on the next page. The correlation plot is used to indicate which variables might have high correlation levels. Positive correlations between predictor variables are shown in shades of blue, while negative correlations are shown in shades of red. From the diagram, a slight positive correlation between separate cycle paths and surface material smoothness is visible, as well as a positive correlation between the NACH variable and separate bicycle paths.

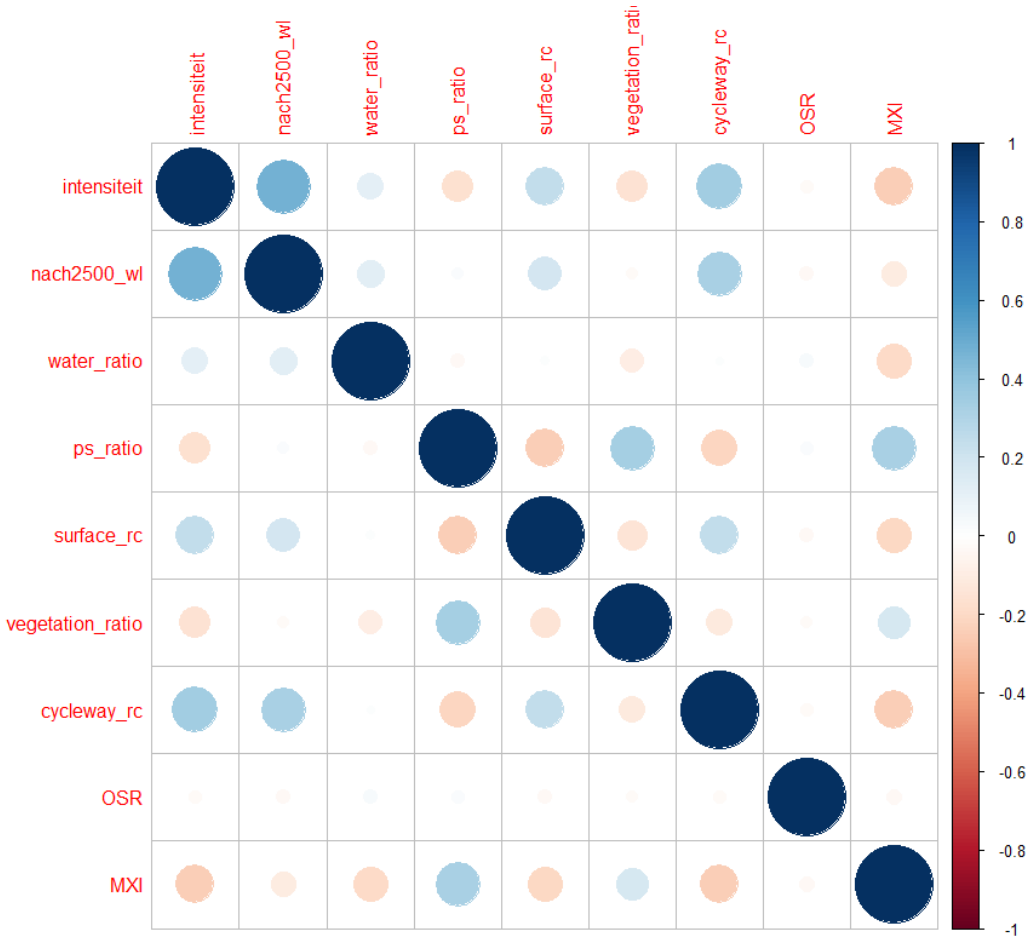


Figure 9: Correlation plot of all the relevant variables in this research

4.3. Individual correlations

The first step of the empirical research was to find out which Space Syntax measure – NACH or NAIN with radius 250m, 500m, 1250m, or 2500m – would be fit to analyse the distribution of cycling activity the best. Simple linear regression was used to see which of these Space Syntax values correlated the most with cycling activity. The graphs resulting from these linear regressions are visualised in figure 10 and 11. Once the Space Syntax measure was decided upon, the other spatial variables were also each individually modelled against cycling counts. The graphs for the spatial characteristics are visualised in figure 12 and 13. The individual models that were significant were kept and combined with other variables, while the individual models that did not show a significant relationship were dropped. Out of the built environment variables, only OSR showed an insignificant relationship with cycling counts. In the section below, all the individual variables will be discussed in more detail. To better understand the different variables, Google Streetview images were also included.

4.3.1. Directness (NACH & NAIN)

To answer the second sub-question: For the Space Syntax Normalised Angular Choice and Normalised Angular Integration analysis, a range of different radii was tested. For the two measures, the radius of 250, 500, 1250, and 2500 meters was chosen. Table 2 shows the results of the individual regression analyses of each radius. The maps that correspond with these measures can be found in the Appendix.

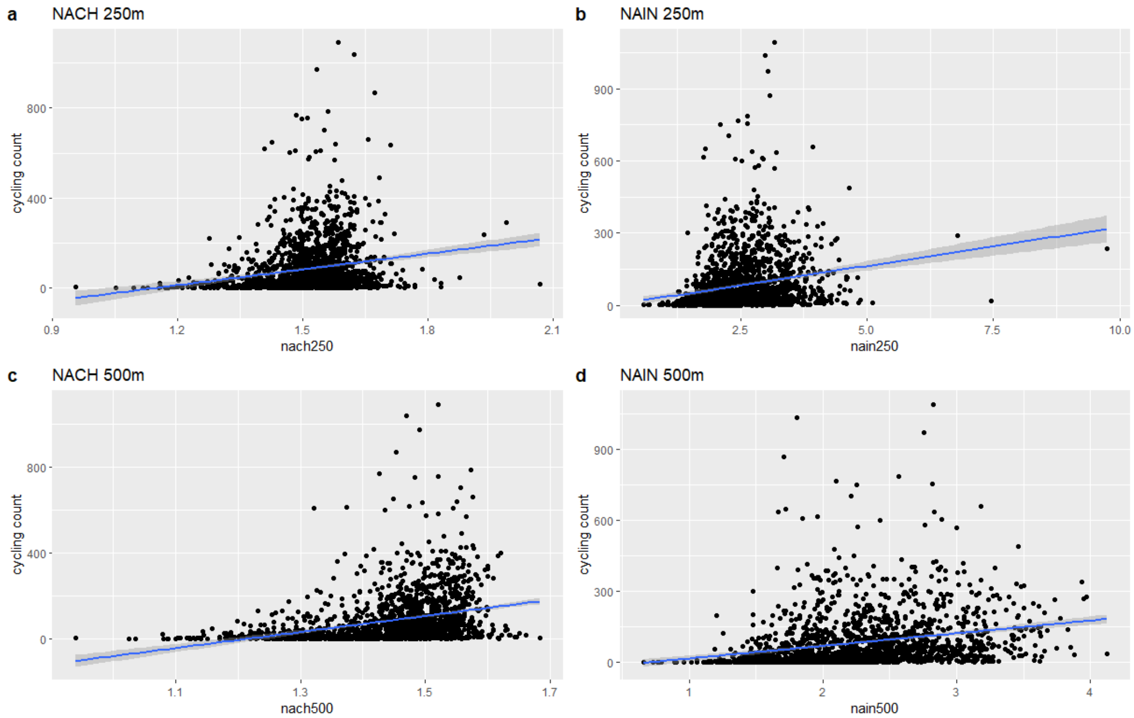


Figure 10: Scatterplot of NACH and NAIN radius 250 and 500 meters

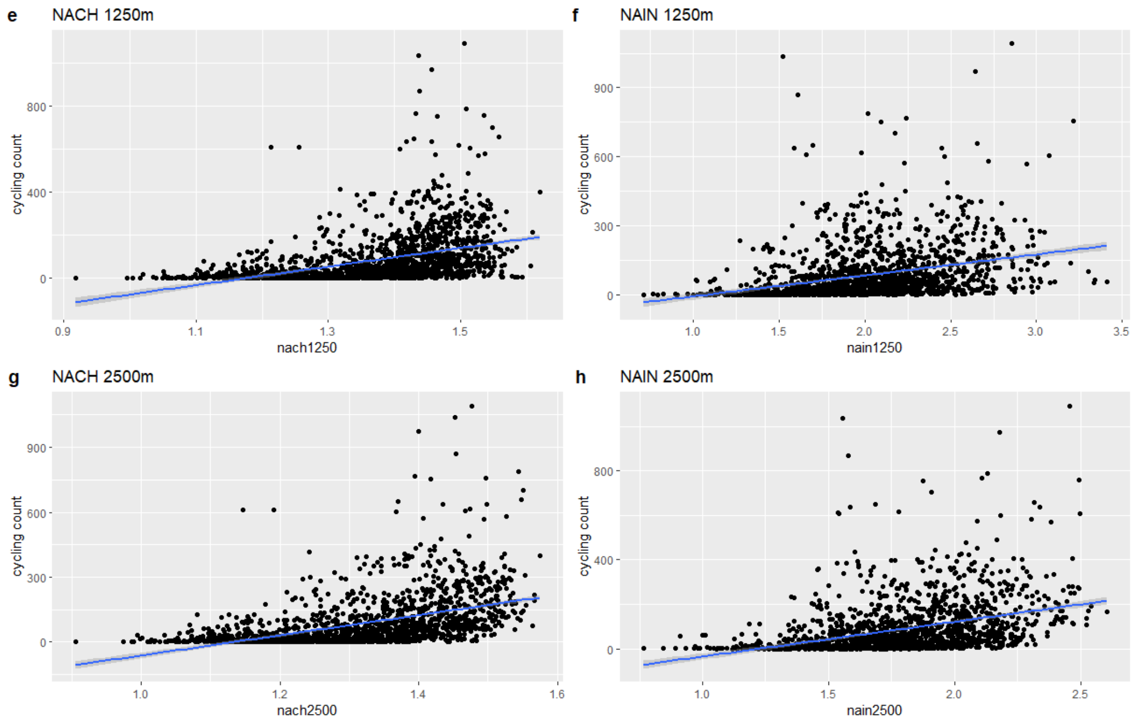


Figure 11: Scatterplot of NACH and NAIN radius 1250 and 2500 meters against cycling counts per street segment

Table 3: Correlation coefficients for the NACH & NAIN radii

NACH & NAIN values	Coefficient	Multiple R-squared	p-value	
NACH 250	198.24	0.03416	1.61e-14	***
NACH 500	363.87	0.09456	< 2e-16	***
NACH 1250	454.44	0.1826	< 2e-16	***
NACH 2500	**475.15**	**0.2244**	**< 2e-16**	*****
NAIN 250	15.228	0.02439	9.62e-11	***
NAIN 500	57.322	0.06889	< 2e-16	***
NAIN 1250	90.784	0.1014	< 2e-16	***
NAIN 2500	153.787	0.1468	< 2e-16	***

***p < 0.01, **p < 0.05, *p < 0.1

All of the NACH and NAIN values showed a significant correlation with cycling counts. Looking at the values of the Multiple R^2 tests in more detail, the NACH with a radius of 2500 meters had the highest R^2 (R^2 = 0.2244). This measure was then used for the multiple linear regression analysis. Compared to all other variables, the NACH measure had a far higher explanatory power.

The next section will elaborate on the individual regression between the micro-level spatial characteristics and cycling activity. The graphs that correspond with these variables are shown in figure 12 and 13.

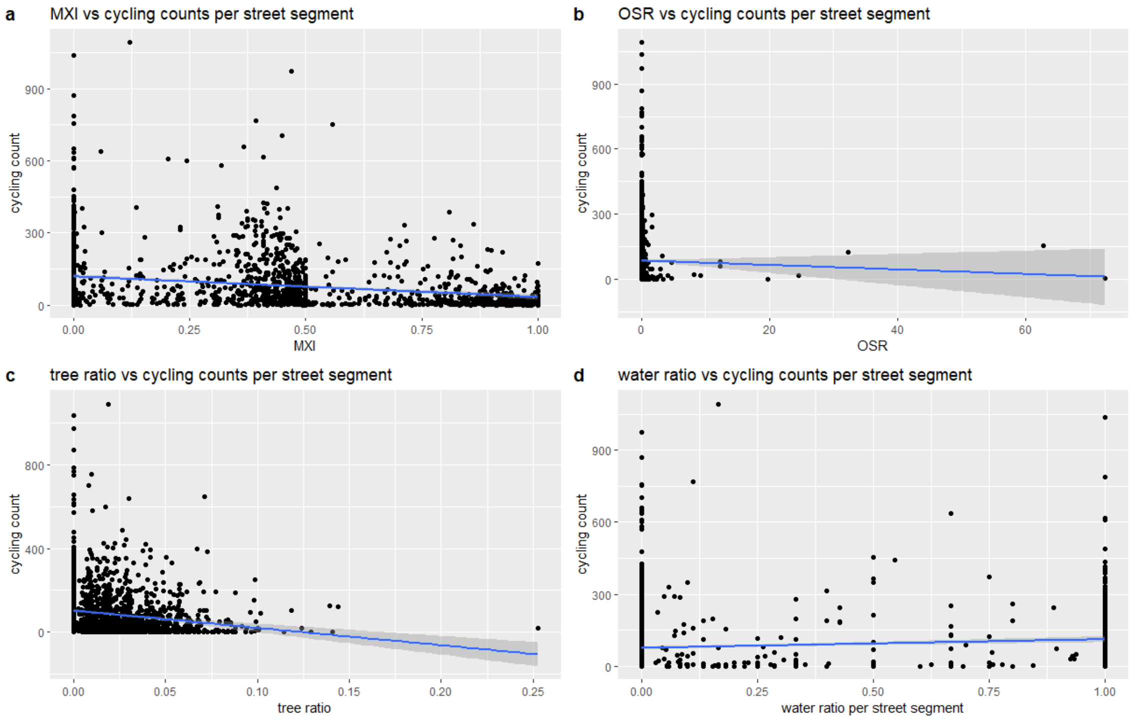


Figure 12: Scatterplot of MXI (a), OSR (b), tree ratio (c), and water ratio (d) against cycling counts per street segment

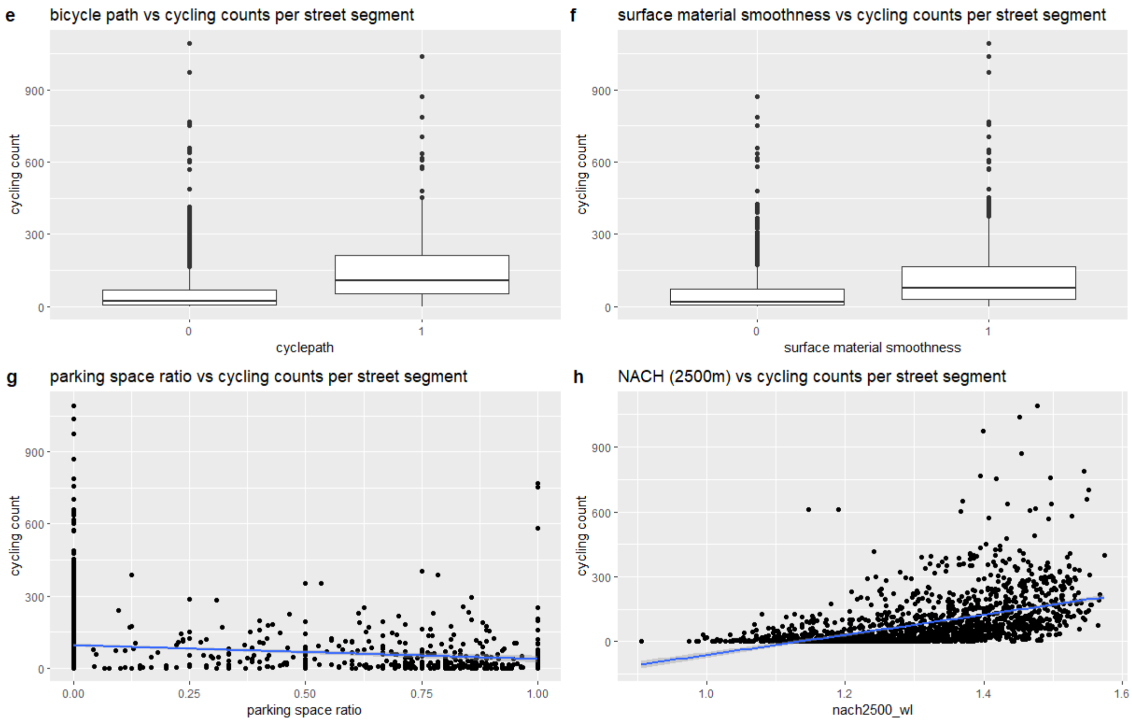


Figure 13: Scatterplot of separate bicycle paths (e), street surface material (f), parking space ratio (g), and NACH(2500m)

4.3.2. Land use variables

Function mix (MXI)

The MXI shows a negative significant relationship with cycling activity per street segment ($p < 2e-16$). R^2 is 0.0619, indicating only a small variance in cycling counts can be attributed to the MXI. The blue regression line in figure 10.a shows a slight negative slope. This indicates a tendency that a higher MXI score relates to lower cycling activity. Or, most residential streets show a lower cycling count than streets with more mixed functions. The scatterplot further indicates a concentration of data around the 0.50 mark on the x-axis. This poses some questions relating to the linearity of the relationship.

Figure 14: Streetview image of street segments with high MXI (left) and low MXI (right)

The map below visualizes the information from the scatterplot in a geographical context. MXI scores lay between 0 and 1, with a value of 0 meaning 0% of the building block is designated as “housing”. An MXI score of 1 on the other hand, means all the floor space of a building block is designated for “housing”. Street segments with a low MXI score are visualised in lighter colours, and the higher scores are with darker, thicker lines. It is important to note that MXI is only measured per building block, the lines that go through the urban parks of Amsterdam are therefore not visualised on the map. The street segments next to the canals in the city centre display a lighter colour than the more “residential” neighbourhoods, in the outer areas of the study area.

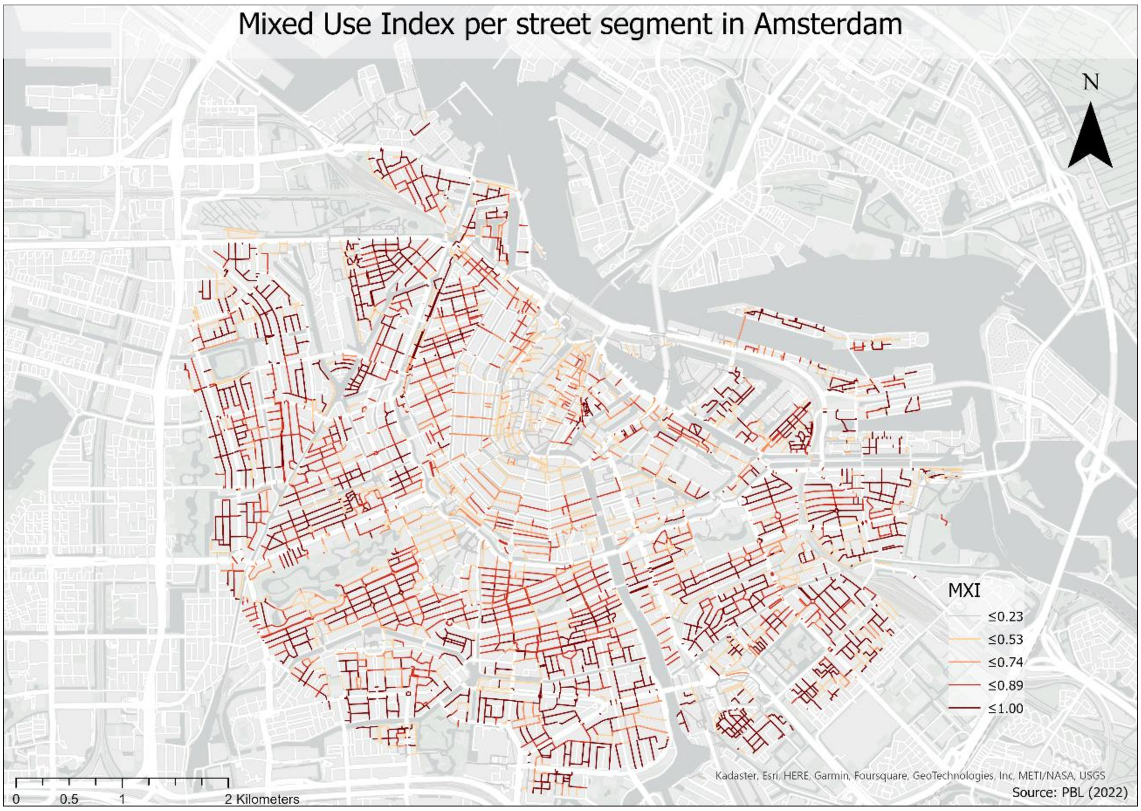


Figure 15: Map of Mixed Use Index per street segment

Map of Mixed Use Index per street segment

Urban density (OSR)

OSR was the only variable that showed no significant relation with cycling activity per street segment. Therefore, on the basis of this data, no conclusion can be drawn on the effect of empty spaces in the city on cycling counts. The reason why OSR had no significant relationship might be because the inner city of Amsterdam does not have a wide range of OSR values, to begin with. The built density in Amsterdam is rather high, both for busier streets and quiet streets. When the outer areas of the city are taken into account, this might be different. The graph of figure 10.b also shows how a great majority of streets have an OSR score that is low, while only a handful have larger OSR scores.


Figure 16: Streetview image of low OSR score (left) and high OSR score (right) (Source: Google Streetview, 2022)

In the map below, the small number of high OSR scores is immediately visible. The predominantly light street segments show that the inner city of Amsterdam has a high urban density.

Figure 17: Map of Open Space Ratio per street segment

Urban greenery

The graph in figure 10.c shows the plot of cycling counts versus tree ratio. The two variables show a significant negative correlation ($p < 2e\text{-}16$), meaning street segments with fewer trees have a higher cycling activity count. The R^2 value of 0.02 indicates that only 2% of the variance can be explained by the tree ratio. The negative relationship between the tree ratio and cycling activity seems to go against the existing theory provided in the literature review.

The number of trees per street segment shows a negative significant relation ($p = 1.43e\text{-}10$) with the cycling activity count per street segment. Although the low multiple R^2 of 0.02 means that only 2% of the cycling counts can be explained by the tree ratio. This is in contrast to the hypothesis that streets with more urban greenery would see more physical activity than streets that did not.

Figure 18: Streetview image of high tree ratio (left), and low tree ratio (right) (Source: Google Streetview, 2022)

The pictures above both show street segments with high cycling counts, the street segment in the left picture is located outside of the city centre, has broad sidewalks and a lot more place for greenery. The picture on the right is one of the busier streets in the city centre, but it has very little space for trees in public space.

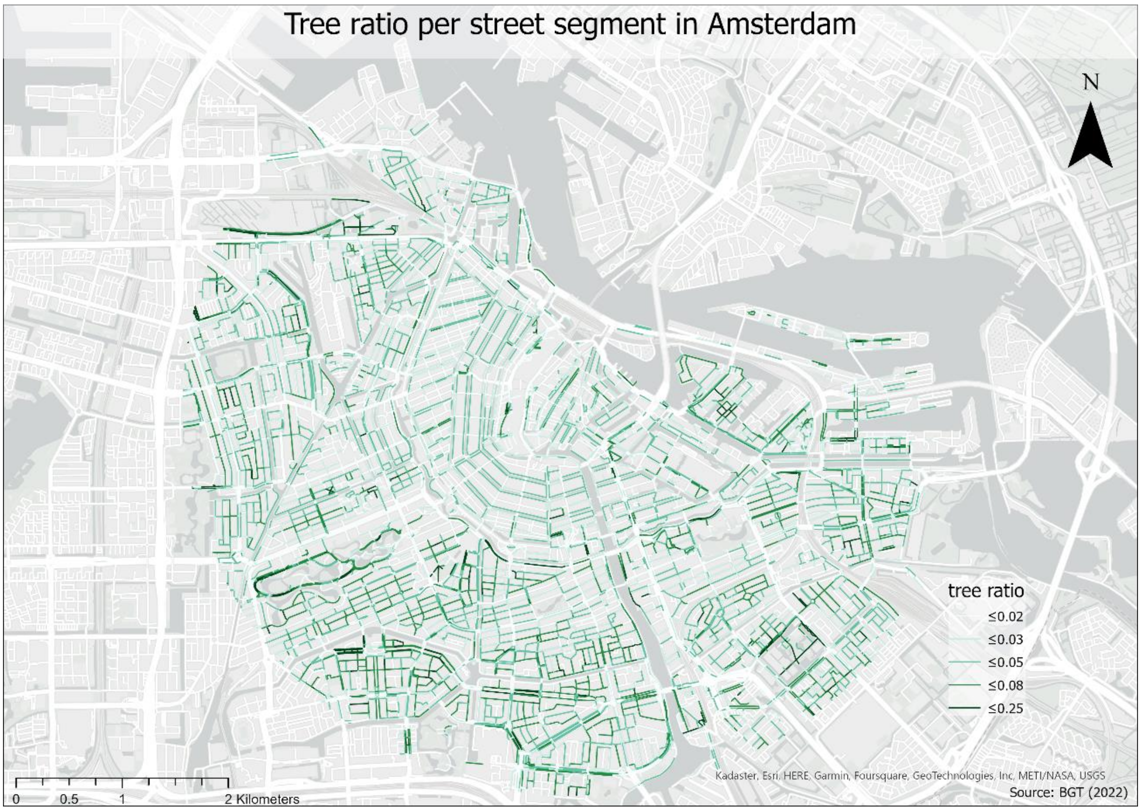


Figure 19: Map of Tree ratio per street segment

In the map above, the tree density per street segment is visualised. The street segments outside of the city centre seem to have a higher tree density than street segments inside the city centre. The segments passing through the urban parks (Vondelpark, Frankendael Park, and Westerpark have a higher-than-average tree density. Compared to the map in figure 1, the busier roads leading out of the city centre do not appear on this map.

Open water bodies

Open water bodies showed a significant positive relationship with cycling counts in the study area. Although the low R^2 value ($R^2 = 0.01365$) means that almost 98% of the variance in cycling count is due to other factors than open water bodies. Figure 10.d shows the data in the form of a scatterplot. In general, a tendency can be seen for street segments having a value of either 0, or 1. This is likely due to the measurement method utilised for this variable.

Figure 20: Map of water ratio per street segment

The map depicted in figure 20 shows the abundance of open water bodies in the centre of Amsterdam. The canals and the Amstel river are popular tourist destinations, giving Amsterdam its unique

Transportation variables

Parallel parking spaces

The ratio of parallel parking spaces along street segments shows a negative significant relationship with cycling count ($p = 1.2e\text{-}09$). The negative relationship falls in line with earlier research by Bhat & Stinson (2003), that cyclists tend to avoid streets with parallel parking. The Multiple R-squared of 0.02156 indicates that only 2% of cycling counts can actually be explained by the parking space ratio, and 98% of the variance is due to other factors.

Figure 21: Streetview image of low parallel parking space ratio (left), and high parking space ratio (right)

The negative relationship between parking space ratio and cycling counts could be due to Dutch road design practices, rather than active behaviour by cyclists. At busy arterial roads, the bicycle path will be fully separated from the main road and parking spaces, as is shown in the picture on the left. The streets where cyclists do need to cycle next to on-street parking spaces tend to be more residential, as is depicted in the picture on the right. The correlation plot presented in figure 9 also shows a slight positive correlation between the MXI and parking spaces. Meaning the street segments with higher parallel on-street parking numbers, also had more housing area per building block. Furthermore, when looking at the map below, it becomes clear that the street segments with higher number of on-street parking right next to cycling tend to be in residential areas.

As with prior maps, the street segments with the highest parking space ratios are displayed in darker-coloured, thicker lines, whereas streets with lower parallel parking are displayed in lighter colours. The city centre of Amsterdam shows a lower parking space ratio than the neighbourhoods surrounding the centre. Naturally, the links going through the urban parks of Amsterdam, like the Vondelpark, Erasmuspark, Westerpark, and Oosterpark, do not have any parallel parking spaces.

Figure 22: Map of parking space ratio per street segment

Separate bicycle paths & surface material smoothness

Both the "Surface material" and "cycle path" variables were binary variables and showed a significant relationship. The boxplots for these variables are visible in figures 13.e and 13.f. Street segments with separated cycling paths had a significantly higher number of cycling activities, compared to street segments that did not. The same can be said about the surface material, street segments made of asphalt showed significantly more cycling activity as opposed to other street surface materials. The adjusted R-squared scores of 0.06 for Surface material and 0.13 for cycle paths mean that 6% and 13% of the higher cycling counts can be explained respectively. This is in line with the hypothesis that cyclists prefer cycling over separate cycle paths and smoother surfaces – like – asphalt, as opposed to cycling in mixed traffic and over less smooth materials like brick stone.

4.4. Multiple linear regression

After the individual linear regressions, the next step was to combine the various independent variables into a multiple linear model. The strategy to construct the Multiple Linear Regression model is further elaborated upon in chapter 3.3.

The method resulted in the following steps: First, the individual variables were combined in pairs of two. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used for model comparison. BIC penalizes the addition of an extra parameter more heavily than AIC. The BIC value of the combined models was then compared to the BIC value of the previous individual models. If the BIC value of the combined models was higher, or not significantly lower than the previous iteration, the variable would be dropped. Variables that showed an insignificant relationship during the combination process were also dropped. Figure 24 shows the schematic representation of the modelling process. The stars (***) represent the level of significance for each variable. Variables that did not show a significant relationship were not included in the next iteration.

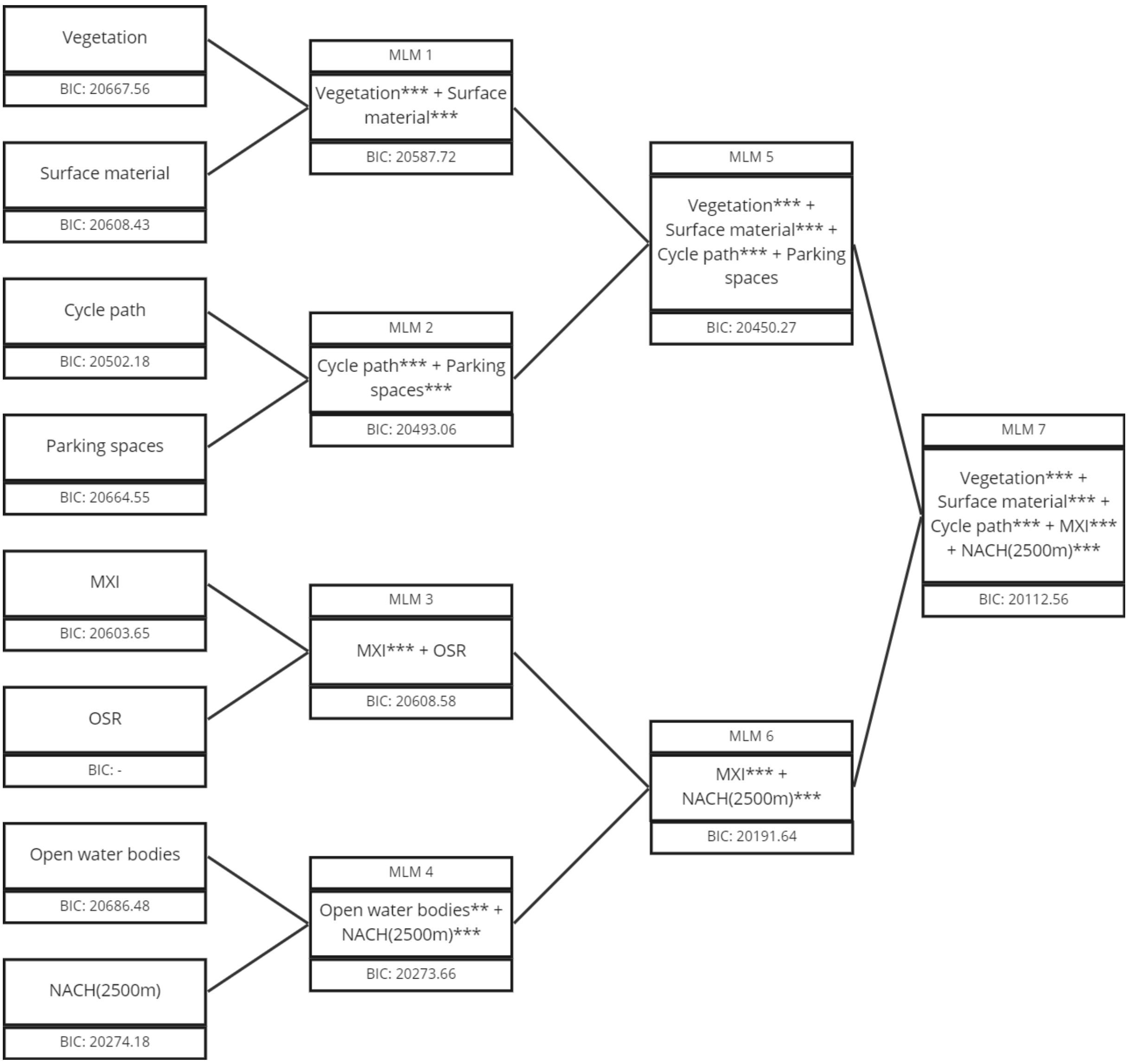


Figure 23: Schematic representation of modelling steps

The Open Space Ratio was not included in the Multiple Linear Regression, as it did not show a significant relationship with cycling counts in the first phase. The addition of Open water bodies in MLM 4 also did not show a significantly lower BIC score than NACH(2500m) on its own. Furthermore, the inclusion of parking space ratio did not result in a significantly better BIC, so parking space ratio was also left out of the Multiple Linear Regression model. The table below shows the number of parameters, Degrees of freedom, adjusted R^2, AIC, and BIC measures for each multiple regression model. Of all the models, model 7 has the lowest AIC and BIC scores.

Table 4: Statistics for every Multiple Linear Regression model

	MLM 1	MLM 2	MLM 3	MLM 4	MLM 5	MLM 6	MLM 7
No. of parameters	2	2	2	2	4	2	5
Df	4	4	4	4	6	4	7
Adjusted R^2	0.0739	0.125	0.06225	0.2328	0.1536	0.2695	0.3113
AIC	20566.04	20471.38	20586.90	20251.98	20417.75	20169.96	20074.62
BIC	20587.72	20493.06	20608.58	20273.66	20444.95	20191.64	20112.56

The variables of Model 7 are shown in more detail in the table below.

Table 5: MLM 7

	Estimate	Std. Error	t value	p-value
(Intercept)	-416.924	27.385	-15.224	< 2e-16 ***
Vegetation ratio	-504.126	108.689	-4.638	3.79e-06 ***
Surface material	22.189	5.169	4.293	1.87e-05 ***
Cycle path	38.386	5.874	6.535	8.44e-11 ***
MXI	-46.969	7.374	-6.370	2.44e-10 ***
NACH(2500m)	388.405	21.026	18.472	< 2e-16 ***

The final multiple regression model for cycling counts can be noted as follows:

$$Cycling\ count\ per\ street\ segment = -426.924 - 504.126 * X_{vegetation} + 22.189 * X_{surface\ material} + 38.386 * X_{cycle\ path} - 46.969 * X_{MXI} + 388.405 * X_{NACH}$$

MLM 7 gave the lowest overall BIC, and the highest adjusted R^2 score. With all the variables combined, the model could explain over 31% of the variance of cycling counts in the study area.

4.4.1. Checking for Multicollinearity

In multiple linear regression models, it is possible that the predictor variables show a degree of correlation. This is problematic for the accuracy of the model. Therefore, the model shown above was checked for multicollinearity with the Variance Inflation Factor (VIF). When predictor variables show a high degree of correlation among each other, one of the highly correlated predictor variable can be left out of the regression model.

As a rule of thumb, when the VIF value sits around the value of 1, the predictor variables are independent from each other, and multicollinearity is not an issue. When values for VIF are higher than 5, the level of collinearity that can be problematic. It is important to stress that these interpretations are rules of thumb, and are therefore not set in stone (source). The table below shows how all of the predictor variables show VIF values close to 1. None of the predictor variables had to be dropped.

Table 6: VIF scores for every independent variable

Variable	Vegetation ratio	Surface	Cycle path	MXI	NACH(2500m)
VIF	1.048997	1.115282	1.214281	1.110887	1.132014

5. Discussion

In the previous chapter, the results of the statistical and geographic analysis were presented. In this chapter, the interpretations and implications of those results for future research will be discussed in more detail. This chapter will also touch upon some of the limitations of this research that should be addressed in future studies.

5.1. Interpretation of the results

Some of the results from this research did line up with earlier expectations based on existing literature, while in other cases, the opposite pattern could be recognised. As the conceptual model suggests, the choice of someone to cycle a certain route depends on a plethora of small and large variables that cannot be explained fully. Spatial characteristics are just one of many factors that have an influence on cycling behaviour. Nonetheless, the prior review of the existing academic literature and the regression analysis suggests that characteristics of the built environment do correlate with cyclists' behaviour.

5.1.1. Land use characteristics

Mixture of functions

In regards to the land use mix, the data showed a negative correlation between cycling activity and MXI score. It should be stressed that MXI only shows the ratio of housing, relative to the total floor space per building block. This means that the street segments with less floor space dedicated to housing showed significantly higher numbers of cyclists than street segments with more housing. Also, MXI only distinguishes between "housing" and "other" land uses, without distinguishing between various commercial uses, for example. This makes the usability of MXI to measure land-use mix disputable.

Urban Density

Density was measured by calculating the Open Space Ratio around each street segment. A high OSR score translates to a high amount of open land per building block, and therefore a lower urban density. The analysis did not show any significant correlation between the OSR score around each street and cycling activity. This is different from prior research, which tends to suggest that higher urban densities attract more active travel modes and cyclists. The insignificant relationship between built density and cycling counts in this research can be accredited to the choice of the inner city of Amsterdam as the study area. Almost every street segment in the inner city of Amsterdam has low OSR scores – i.e. high urban densities. Future research could opt for a larger study area, with a larger variety in OSR per building block, which might give different results.

Urban greenery

The significant negative relationship between tree ratio and cycling activity is not in line with prior research presented in the theoretical framework. Greenery alongside street segments was measured by calculating a tree ratio within a distance of 15 meters around each street segment. The choice to only look at trees to measure urban greenery did probably have an influence on the results.

Prior research shows how urban greenery attracts more cycling in cities. On the basis of this research, this conclusion cannot be made. This could be due to the fact that for people who cycle on a daily basis – as is often the case for Amsterdam residents – urban greenery along the route becomes less of a priority than other factors.

Open water bodies

Contrary to the negative relationship between tree ratio and cycling counts, open water bodies did show a significant positive relationship with cycling counts. In prior studies, urban greenery and open water bodies were often both mentioned as measures for "attractive urban environment". The analysis performed in this research suggests that the two variables do not necessarily complement each other. The open water bodies were measured in a similar way as urban greenery. The choice of Amsterdam as the study area probably plays a significant role here, as the canals in Amsterdam attract a lot of visitors as a tourist destination. Overall, the city of Amsterdam has a lot of water in the city centre, making it difficult to compare with other contexts.

5.1.2. Transportation characteristics

Parallel on-street parking spaces

Existing literature suggests that the presence of on-street parallel parking poses a potential risk for cyclists, and therefore cyclists tend to avoid streets with high numbers of on-street parking spaces. The significant negative relationship between on-street parking spaces and cycling activity in this research is therefore in line with earlier academic studies. Here, context also plays a role, as Dutch road infrastructure tends to be designed

Separate bicycle paths

Prior academic literature suggests how separate bicycle paths attract more cyclists. Overall, a positive significant relationship was found between the number of cyclists and whether the street segment had a separate cycle path or not.

Street surface material smoothness

Street segments with asphalt as the surface material also showed a significantly higher number of cyclists than street segments with other surface materials. Literature seems to suggest that cyclists do prefer to cycle over smoother surfaces and separated cycle paths. However, the causality could also be the other way around: The municipality of Amsterdam identified the busiest cycling routes and improved the infrastructure on those street segments. On the basis of the available data and the utilised research methods, it is not possible to draw conclusions on whether cyclists prefer asphalt and separated cycle paths over other forms of infrastructure.

Directness

Normalised Angular Choice, with a radius of 2500 meters, was used to measure the "directness" of each street segment. This Space Syntax measure correlated most with cycling activity from the individual regression analysis. From the different radii that were used, the higher radii increasingly showed higher R^2 values. The radius of 2500 meters was the highest radius that was included in this research, this raises the question of whether a bigger radius would have resulted in an even stronger explanatory power of the model. A prior study that compared NACH and cycling activity suggested a radius of 4500 meters as having the highest correlation (Orellana et al., 2019). Due to the study area and the availability of data, this research was limited to 2500 meters.

The R^2 value of NACH 2500m exceeded that of other variables, meaning it explained the variance in cycling activity better than the other independent variables. This is in line with the earlier study in central London (Law et al., 2014). Space Syntax might therefore be used to determine which street segments could potentially see the most cycling activity. Also, Normalised Angular Choice was originally intended to compare cities from varying contexts with each other. It is interesting to see whether this also holds true when used for analysing cycling behaviour.

5.2. Research question & sub-questions

This research is centred around one main research question and four sub-questions. This thesis answers the sub-questions in the following way:

1. *What (environmental) factors influence cycling patterns, according to existing literature?*

The route choice of a cyclist depends on the utility that this route poses In this research, four land use variables, and four transportation variables were identified as having an influence on cycling route choices. Those include: "mixture of functions", "urban density", "urban greenery", "open water bodies", "parallel on-street parking spaces", "separate bicycle paths", "surface material smoothness", and "directness". The schematic model, shown in figure 2, displays the factors that influence cycling route choices, according to existing academic literature. Space Syntax was used to measure "directness".

2. *What Space Syntax measures can be used to analyse cycling patterns in urban areas?*

A dive into existing academic literature learned that "Normalised Angular Choice" and "Normalised Angular Integration" are both measures that represent "through-movement" and "to-movement" potential of street segments respectively. Following this, NACH and NAIN, with radii from 250 meters to 2500 meters were individually tested with cycling activity. From the individual regression models, NACH with a 2500 meter radius had the highest correlation with cycling activity. NACH 2500m was therefore selected to be included in the multiple linear regression analysis.

3. *To what extent do built environment characteristics explain cycling intensities per street segment?*

The multiple linear regression model that was constructed in this research could explain 31% of the variance in cycling count. Following the individual linear regression analyses performed in this research project, the results seem to suggest that urban greenery, the presence of open water bodies, street surface material smoothness, separate bicycle paths, the mixture of functions, and directness showed a significant relationship with cycling activity. The open space ratio – used to measure urban density – was the only variable that did not show a significant relationship with cycling activity.

4. *To what extent does the addition of Space Syntax measures improve route choice demand models?*

The results of the individual linear regression analyses indicate that the explanatory power of the Space Syntax measure (R^2=0.2244) was greater than that of the other built environment variables. For only one variable, this is a fairly large R^2 value. The separate bicycle path showed the second-highest explanatory power (R^2=0.13). Prior studies have shown that using bigger radii will lead to higher correlations, future research could look into the possibilities of higher radii for NACH and NAIN.

Main RQ: *To what extent do spatial characteristics influence bicycle route choices of cyclists?*

The simple answer is that 31% of the cycling activity could be explained by a model that only included spatial characteristics. More specifically, the number of trees along the street segment, the material of the street surface, separate bicycle paths, and land use mix, all accumulated to a fairly high R^2 value of the model. This answer is by no means final, and should be considered as a set up to analyse cycling routes through the lens of Space Syntax.

5.3. Implications

To the best of the researcher's knowledge, this has been the first time that Space Syntax has been used to analyse cycling behaviour in a context of high cycling activity, like the city of Amsterdam. Following the multiple linear regression model presented above, over 31% of cycling activity could be explained by spatial characteristics. Furthermore, this research has shown how that the Normalised Angular Choice – i.e. through movement potential – of street segments does show a significant relationship with cycling activity. The linear regression between NACH and cycling counts per street segment suggests that the explanatory power of NACH goes further than that of other spatial characteristics. This finding seems to suggest that: more than the level of quality of the cycling infrastructure, the street segment directness explained the variance in cycling activity the most. The exploratory nature of this research underlines the need for further research into Space Syntax and cycling activity.

This research has also shown that context is important when analysing the travel behaviour of people. Some of the insights from earlier academic studies were not observed in this research, like the negative relationship between trees and cycling activity. Most of those inconsistencies could be attributed to context specific characteristics of Amsterdam, like its narrow streets, and Dutch road design. In some countries, shade from trees might be an absolute necessity while cycling, while in countries with less developed cycling infrastructure, safety might be more important.

In the past, Space Syntax has been used to predict the effect of adding new infrastructure to the existing urban fabric. A further understanding of Space Syntax could make it possible to see what the effect of the new cycling infrastructure would be.

5.4. Limitations

The research presented in this thesis is of exploratory nature and therefore not without its limitations. The remainder of this chapter will touch upon the limitations that can be found within this research. The limitations relate to the research design, the methodological choices, and some unanticipated obstacles.

5.4.1. Research design

Linear regression was used to find correlations between spatial characteristics and cycling counts per street segment. While this has led to some interesting results discussed earlier, the analysis does not take into account the whole cycling route from origin to destination. Future research could therefore look into the whole cycling route, and compare the chosen route with alternatives.

The BCW data used in this research is revealed preference data that depicts cycling counts per street segment. Going back to the conceptual model presented in chapter 2, cycling route choices are affected by a whole array of variables that reach far further than the characteristics around the routes themselves. The data provided by the BCW is fully anonymized. Furthermore, the purpose of each counted cycling trip is also unclear. The data shows the outcome of certain choices by individuals, but not the reasoning that is behind those choices. As prior studies about cycling route choices point out, personal characteristics and trip purpose do indeed have a big influence on the route choice of people. A more complete research design would be to include both revealed preference data, as well as stated preference data. The time and resources needed to set up such a research project were beyond the scope of this thesis.

5.4.2. Methodological choices: Choice of the city of Amsterdam as a case

The city of Amsterdam – or any Dutch city in that manner – is well known for its high volume of daily cyclists. Cycling infrastructure has been in continuous development in the Dutch context for decades and is prioritized to a far greater extent than in other parts of the world. The growing attention to cycling infrastructure in the Netherlands can be dated back to the early seventies, with activists demanding more attention to fatal traffic accidents in which cyclists and pedestrians formed the majority of victims (Verkade & Brömmelstroet, 2020). Recent studies suggest that a quarter of daily trips in the Netherlands are made by (Harms & Kansen, n.d.). A benefit is that a lot of research on cyclist behaviour can be done, either through the use of GPS data, or travel surveys. On the other side, caution is needed when generalizing the outcomes of research in the Dutch context to other countries.

Regarding this research specifically, some of the variables that were analysed, such as the canals, have a unique effect in Amsterdam. However, the thought behind the Space Syntax methodology is for it to be an analysis method that can be performed in any part of the world, regardless of context (Hillier, 1996). Further research in other contexts in the Netherlands and abroad should give a better understanding of the utility of Space Syntax as a method to analyse cycling behaviour.

5.4.3. Unanticipated obstacles

As was elaborated in the methodology, the radius of the study area was chosen by calculating the average cycling distance of a route (3.6 kilometres) and taking that as the radius around the centroid of Amsterdam. The distance of 3.6 kilometres might be considered too small for a Space Syntax analysis on cycling. This is especially true when taking into account the "edge-effect" issue, also highlighted by Ratti (2009). The "edge effect" is a phenomenon in which the street segments at the edges of a study area appear to have a lower Integration or Choice value than it would have in reality. This is due to the fact that segments are "cut off" from their surroundings and have no further connecting street segments outside of the study area. Of course, in reality, the connections are still there, they just do not appear on the map. This issue is usually overcome by selecting a radius that is larger than is actually needed. However, the BCW data that was used for this research showed a number of inconsistencies and missing links outside of the core of the city. The missing data in the outer boroughs would have complicated the analysis.

The choice of 3.6 kilometres also meant that only the street segments in the inner city of Amsterdam were analysed. As the results for the OSR variables indicated, the urban density for almost all street segments was rather high.

5.5. Recommendations

This research has shown that over 31% of cycling activity per street segment can be explained by spatial characteristics. Furthermore, the simple linear regression analysis suggests that the "directness", or "through-movement potential" of a street segment explained its cycling count more than cycling infrastructure characteristics, like dedicated bicycle paths, or street surface material.

For future research

This research project can be seen as one of the first steps in which big data and Space Syntax is used to analyse cyclist route choices. The analysis presented in this research is limited to linear regression between cycling counts per street segment and spatial characteristics. Recommendations for future research include looking into whole cycling routes and comparing them with alternative routes. Additionally, bigger-scale research projects should combine revealed preference and stated preference data on cycling, to get a more holistic view of cycling behaviour, as well as the reasoning behind that behaviour.

For urban planners and traffic engineers

In a densely built city like Amsterdam, space is scarce. It is difficult to create new roads without drastically changing the urban fabric. For planners and traffic engineers, this means that dedicating new infrastructure for cyclists will most likely come at the expense of space for other transportation modes. Rather than constructing new bicycle paths along street segments that are not as direct, planners and engineers could look at ways to dedicate more space on existing streets to cyclists.

As was touched upon in the section above, cycling behaviour largely depends on context. In some countries, shade from trees might be an absolute necessity while cycling, while in countries with high numbers of cars, safety might be more important. These are all qualities of infrastructure that can be influenced by urban planners. When considering where to implement those improvements, this research suggests to consider the directness of the street segments to be used as a first indicator.

6. Conclusion

Climate change is increasingly affecting cities and regions around the globe. The urgency to change the way people get around is also growing. Rather than an ultra-modern new transportation method, part of the solution to reduce greenhouse gas emissions might be a transportation mode that has been around for ages. Cycling emits almost no emissions, and is far more space efficient than the private car. That is why cycling is attracting more and more attention as a sustainable and healthy alternative to motorized traffic.

To improve cycling infrastructure and attract more cyclists, it is important to know which parts of a street network are likely to attract large numbers of cyclists, and which are not. This research looked at the influence of spatial characteristics on cycling route choices. A couple of important results were found in this research:

First, a conceptual framework was drawn, based on existing literature about cycling route choices. Cycling route choices Second, the multiple linear model that was constructed in this research could explain over 31% of the variance in cycling counts per street segment in the study area. Moreover, simple linear regression showed how the directness of street segments – measured by Normalised Angular Choice – explains the cycling activity for 22%. This was more than any other spatial characteristic. This indicates that above dedicated cycling infrastructure, the directness of a street segment was most important in explaining the variance of cycling activity.

Other variables that did have a relationship in this study were the number of trees along the street segment, the material of the street surface, separate bicycle paths, and land use mix. The negative relationship between tree ratio indicates that urban greenery might not be the most important aspect of a cycling route for cyclists.

Some of the spatial characteristics had opposite relationships with cycling activity than what one would assume from academic studies. In part, this could be attributed to the unique spatial characteristics of Amsterdam. This also stresses the importance of local context when analysing cycling behaviour. A better understanding of cyclist distribution in Amsterdam was created by both statistical analysis and geographical visualisation.

As is elaborated in the Literature review, the route choice of a cyclist depends on a plethora of different variables that relate to living conditions of the individual, as well as the context. This research project explored whether spatial characteristics and Space Syntax could be used to explain cycling activity in the street network of Amsterdam.

The exploratory setup of this research means that there are possibilities for future research to analyse the relationship between Space Syntax and cycling behaviour in more detail. For example, by using a combination of stated and revealed preference data, or comparing the chosen route of a cyclist with alternative routes. An even more ambitious research project would be to draw axial maps of cities, and perform axial analyses, instead of using the Road Center Lines.

7. Literature

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

de Vries, S. I., Hopman-Rock, M., Bakker, I., Hirasing, R. A., & van Mechelen, W. (2010). Built environmental correlates of walking and cycling in dutch urban children: Results from the SPACE study. *International Journal of Environmental Research and Public Health*, *7*(5), 2309–2324. https://doi.org/10.3390/ijerph7052309

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Fishman, E., Böcker, L., & Helbich, M. (2015). Adult active transport in the Netherlands: An analysis of its contribution to physical activity requirements. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0121871

Gubbels, J. S., Kremers, S. P. J., Droomers, M., Hoefnagels, C., Stronks, K., Hosman, C., & de Vries, S. (2016). The impact of greenery on physical activity and mental health of adolescent and adult residents of deprived neighborhoods: A longitudinal study. *Health and Place*, *40*, 153–160. https://doi.org/10.1016/j.healthplace.2016.06.002

Halldórsdóttir, K. (2015). *General rights Behavioural models for cycling-Case studies of the Copenhagen Region*.

Harms, L., & Kansen, M. (n.d.). *Fietsfeiten Kennisinstituut voor Mobiliteitsbeleid | KiM*.

Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by bicycle: An overview of the literature. *Transport Reviews*, *30*(1), 59–96. https://doi.org/10.1080/01441640903187001

Hillier, B. (1996) Space Is the Machine: A Configurational Theory of Architecture. Cambridge University Press, Cambridge.

Hillier, B., Yang, T., & Turner, A. (2012). *The Journal of Space Syntax J S S Normalising least angle choice in Depthmap and how it opens up new perspectives on the global and local analysis of city space Normalising least angle choice in Depthmap and how it opens up new perspectives on the global and local analysis of city space*. http://www.journalofspacesyntax.org/

Hogendorf, M., Oude Groeniger, J., Noordzij, J. M., Beenackers, M. A., & van Lenthe, F. J. (2020). Longitudinal effects of urban green space on walking and cycling: A fixed effects analysis. *Health and Place*, *61*. https://doi.org/10.1016/j.healthplace.2019.102264

Jandari, A., Meulders, M., Desplenter, S., & Vandebroek, M. (2020). Attribute non-attendance in choosing the bike as a transport mode in Belgium. *European Journal of Transport and Infrastructure Research*, *20*(4), 127–151. https://doi.org/10.18757/ejtir.2020.20.4.4475

Kaczynski, A. T., Potwarka, L. R., Smale, B. J. A., & Havitz, M. F. (2009). Association of Parkland proximity with neighborhood and park-based physical activity: Variations by gender and age. *Leisure Sciences*, *31*(2), 174–191. https://doi.org/10.1080/01490400802686045

Koohsari, M. J., Cole, R., Oka, K., Shibata, A., Yasunaga, A., Hanibuchi, T., Owen, N., & Sugiyama, T. (2020). Associations of built environment attributes with bicycle use for transport. *Environment and Planning B: Urban Analytics and City Science*, *47*(9), 1745–1757. https://doi.org/10.1177/2399808319845006

Krenn, P. J., Oja, P., & Titze, S. (2014). Route choices of transport bicyclists: A comparison of actually used and shortest routes. *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1). https://doi.org/10.1186/1479-5868-11-31

Krizek, K. J. (2018). Measuring the wind through your hair? Unravelling the positive utility of bicycle travel. *Research in Transportation Business and Management*, *29*, 71–76. https://doi.org/10.1016/j.rtbm.2019.01.001

Lachowycz, K., & Jones, A. P. (2011). Greenspace and obesity: A systematic review of the evidence. *Obesity Reviews*, *12*(5). https://doi.org/10.1111/j.1467-789X.2010.00827.x

Law, S., Sakr, F. L., & Martinez, M. (2014). Measuring the changes in aggregate cycling patterns between 2003 and 2012 from a space syntax perspective. *Behavioral Sciences*, *4*(3), 278–300. https://doi.org/10.3390/bs4030278

Liu Ziqi Song, Z., Chen, A., & Ryu, S. (2016). *Exploring Bicycle Route Choice Behavior with Space Syntax Analysis FINAL REPORT*.

Lu, Y. (2019). Using Google Street View to investigate the association between street greenery and physical activity. *Landscape and Urban Planning*, *191*. https://doi.org/10.1016/j.landurbplan.2018.08.029

Ma, L., & Ye, R. (2022). Utilitarian Bicycling and Mental Wellbeing: Role of the Built Environment. *Journal of the American Planning Association*, *88*(2), 262–276. https://doi.org/10.1080/01944363.2021.1950036

McCahill, C., & Garrick, N. W. (2008). The applicability of space syntax to bicycle facility planning. *Transportation Research Record*, *2074*, 46–51. https://doi.org/10.3141/2074-06

Mendiate, C. J., Nkurunziza, A., Soria-Lara, J. A., & Monzon, A. (2022). Exploring Users' perceptions of factors influencing cycling route choice: A perspective from Quelimane, Mozambique. *Transportation Planning and Technology*, *45*(2), 119–137. https://doi.org/10.1080/03081060.2022.2088535

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, *44*(9), 754–765. https://doi.org/10.1016/j.tra.2010.07.008

Mertens, L., Compernolle, S., Deforche, B., Mackenbach, J. D., Lakerveld, J., Brug, J., Roda, C., Feuillet, T., Oppert, J. M., Glonti, K., Rutter, H., Bardos, H., de Bourdeaudhuij, I., & van Dyck, D. (2017). Built environmental correlates of cycling for transport across Europe. *Health and Place*, *44*, 35–42. https://doi.org/10.1016/j.healthplace.2017.01.007

Muñoz, B., Monzon, A., & López, E. (2016). Transition to a cyclable city: Latent variables affecting bicycle commuting. *Transportation Research Part A: Policy and Practice*, *84*, 4–17. https://doi.org/10.1016/j.tra.2015.10.006

Myers, R. H., Montgomery, D. C., Vining, G. G. & Robinson, T. J. (2010). *Generalized Linear Models: with Applications in Engineering and the Sciences*. Wiley.

Orellana, D., Guerrero, M., María, ;, & Guerrero, L. (2019). *The influence of Space Syntax on Cycling Movement. Proceedings of the 12 Space Syntax Symposium Proceedings of the 12 th Space Syntax Symposium THE INFLUENCE OF SPACE SYNTAX ON CYCLING MOVEMENT IN MANTA, ECUADOR*. https://www.researchgate.net/publication/334520638

Prato, C. G., Halldórsdóttir, K., & Nielsen, O. A. (2018). Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. *International Journal of Sustainable Transportation*, *12*(10), 770–781. https://doi.org/10.1080/15568318.2018.1437236

Raford, N., Chiaradia, A., Gil, J., & Org, E. (n.d.). *UC Berkeley Research Reports Title Space Syntax: The Role of Urban Form in Cyclist Route Choice in Central London Publication Date*.

Ratti, C. (2004). Space syntax: Some inconsistencies. *Environment and Planning B: Planning and Design*, *31*(4), 487–499. https://doi.org/10.1068/b3019

Ravensbergen, L., Buliung, R., & Laliberté, N. (2019). Toward feminist geographies of cycling. *Geography Compass*, *13*(7). https://doi.org/10.1111/gec3.12461

Rybarczyk, G., & Wu, C. (2014). Examining the impact of urban morphology on bicycle mode choice. *Environment and Planning B: Planning and Design*, *41*(2), 272–288. https://doi.org/10.1068/b37133

Segadilha, A. B. P., & Sanches, S. da P. (2014). Identification of Factors that Influence Cyclistś Route Choice. *Procedia - Social and Behavioral Sciences*, *160*, 372–380. https://doi.org/10.1016/j.sbspro.2014.12.149

Snizek, B., Sick Nielsen, T. A., & Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography*, *30*, 227–233. https://doi.org/10.1016/j.jtrangeo.2013.02.001

Sobhani, A., Aliabadi, H. A., & Farooq, B. (2019). Metropolis-Hasting based Expanded Path Size Logit model for cyclists' route choice using GPS data. *International Journal of Transportation Science and Technology*, *8*(2), 161–175. https://doi.org/10.1016/j.ijtst.2018.11.002

Soltani, A., Allan, A., Javadpoor, M., & Lella, J. (2022). Space Syntax in Analysing Bicycle Commuting Routes in Inner Metropolitan Adelaide. *Sustainability (Switzerland)*, *14*(6). https://doi.org/10.3390/su14063485

Stinson, M. A., & Bhat, C. R. (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transportation Research Record*, *1828*, 107–115. https://doi.org/10.3141/1828-13

Tan, R., Adnan, M., Lee, D. H., & Ben-Akiva, M. E. (2015). New path size formulation in path size logit for route choice modeling in public transport networks. In *Transportation Research Record* (Vol. 2538, pp. 11–18). National Research Council. https://doi.org/10.3141/2538-02

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice*, *123*, 7–23. https://doi.org/10.1016/j.tra.2018.08.023

Turner, A., (2000). Angular Analysis: A Method for the Quantification of Space. Centre for Advanced Spatial Analysis.

van Nes, A., & Yamu, C. (2021). Introduction to Space Syntax in Urban Studies. In *Introduction to Space Syntax in Urban Studies*. Springer International Publishing. https://doi.org/10.1007/978-3-030-59140-3

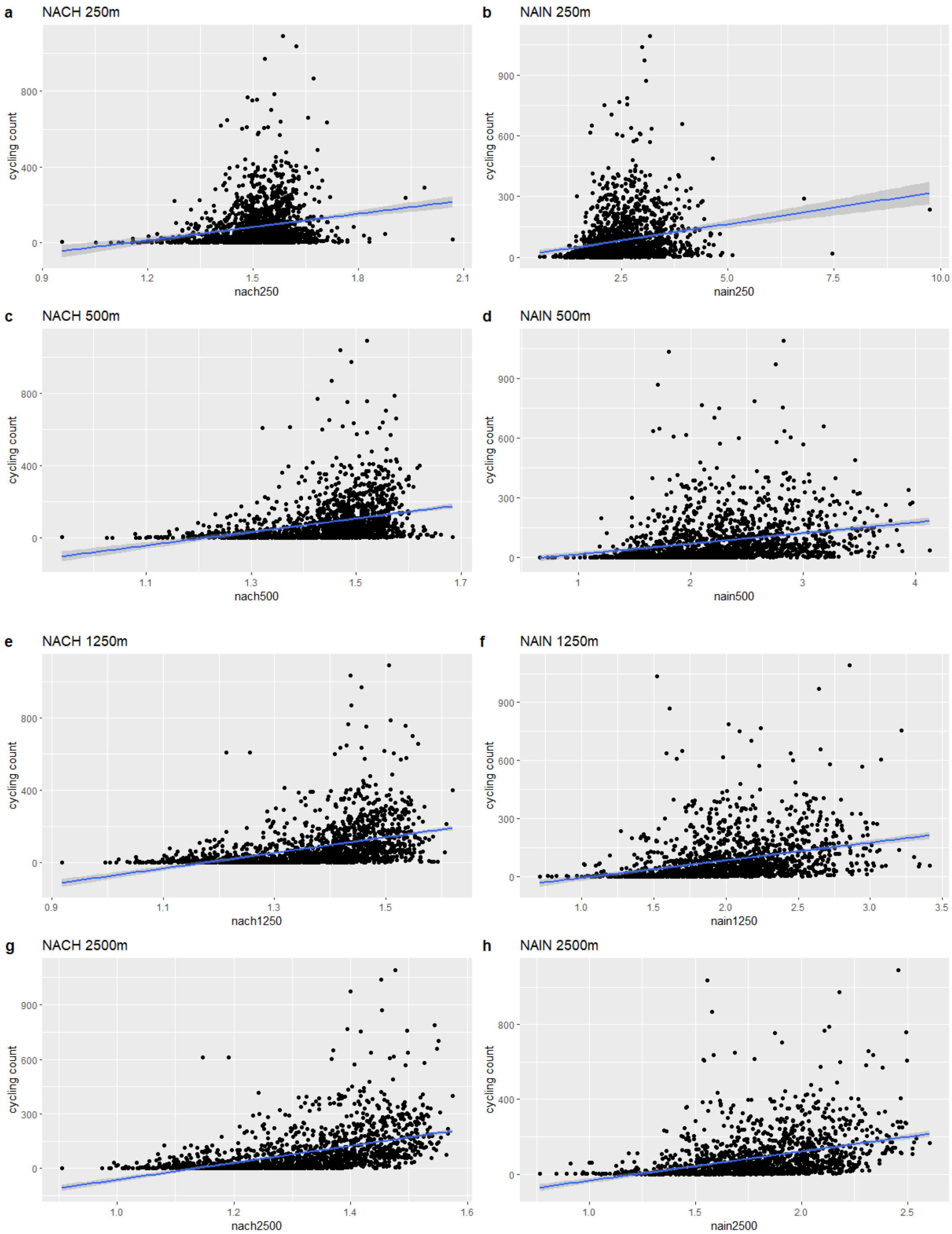
Waygood, E. O. D., Sun, Y., & Letarte, L. (2015). Active travel by built environment and lifecycle stage: Case study of Osaka Metropolitan area. *International Journal of Environmental Research and Public Health*, *12*(12), 15900–15924. https://doi.org/10.3390/ijerph121215027

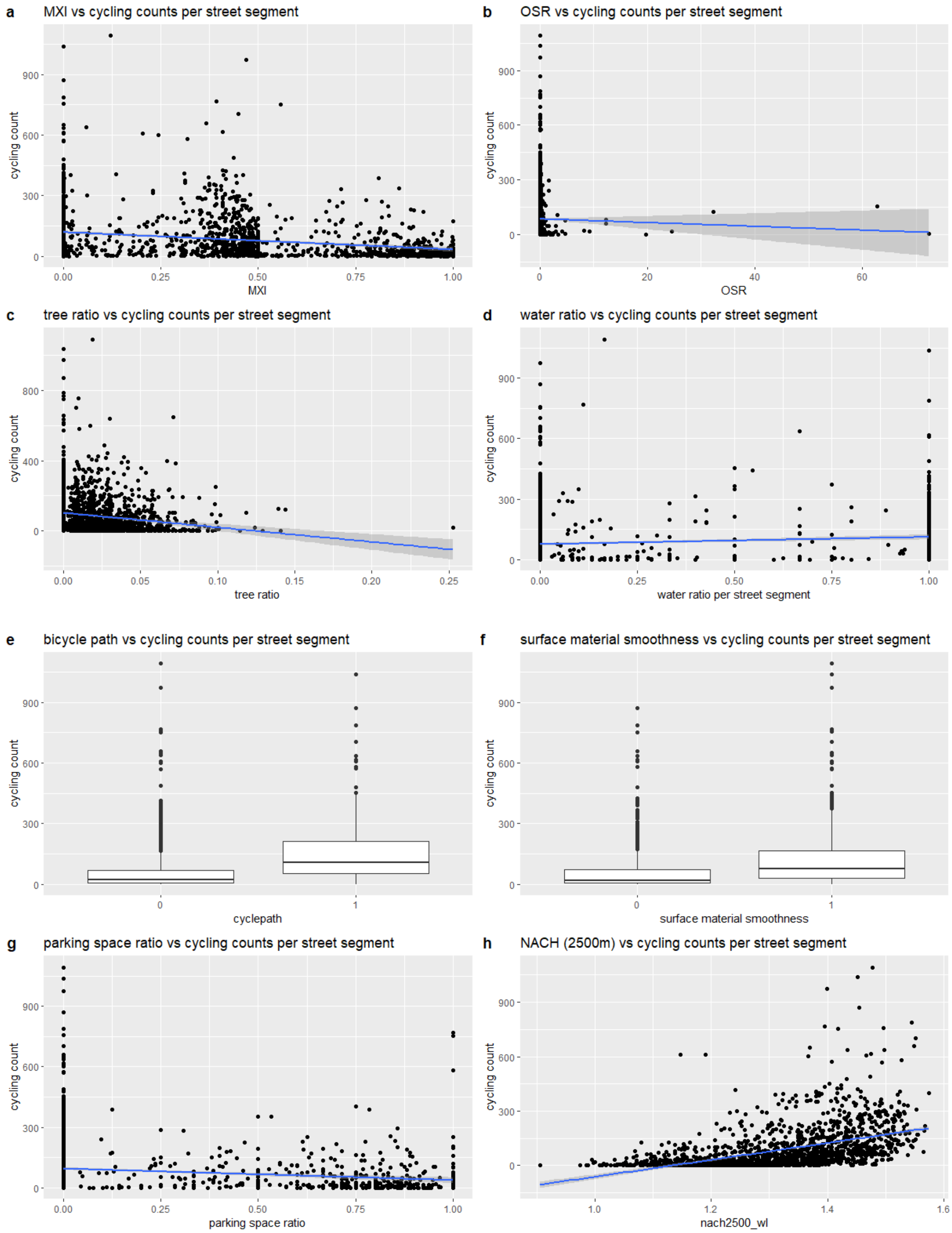
Winters, M., & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion*, *25*(1), 40–47. https://doi.org/10.4278/ajhp.081006-QUAN-236

Yamu, C., van Nes, A., & Garau, C. (2021). Bill hillier's legacy: Space syntax—a synopsis of basic concepts, measures, and empirical application. *Sustainability (Switzerland)*, *13*(6). https://doi.org/10.3390/su13063394

8. Appendix

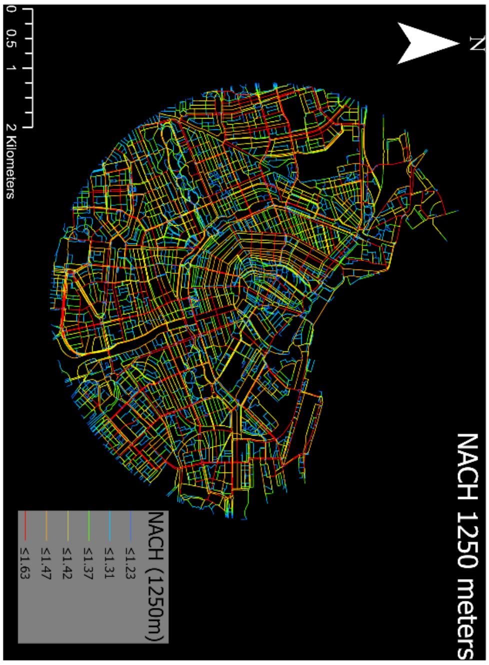
8.1. Scatterplots

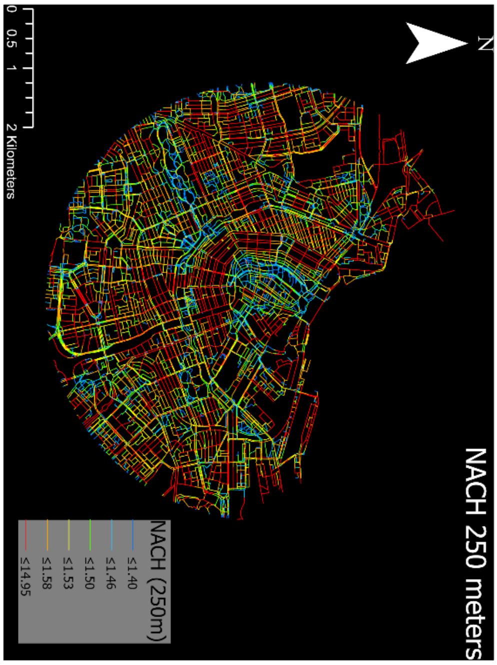


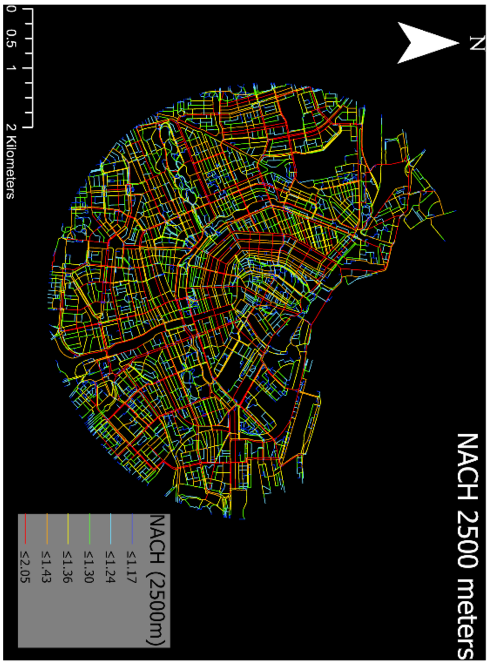


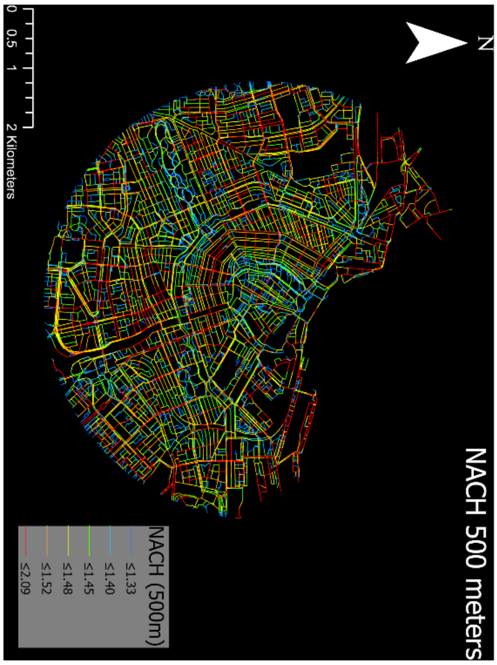
a
MXI vs cycling counts per street segment
cycling count
MXI
b
OSR vs cycling counts per street segment
cycling count
OSR
c
tree ratio vs cycling counts per street segment
cycling count
tree ratio
d
water ratio vs cycling counts per street segment
cycling count
water ratio per street segment
e
bicycle path vs cycling counts per street segment
cycling count
cyclepath
f
surface material smoothness vs cycling counts per street segment
cycling count
surface material smoothness
g
parking space ratio vs cycling counts per street segment
cycling count
parking space ratio
h
NACH (2500m) vs cycling counts per street segment
cycling count
nach2500_wl

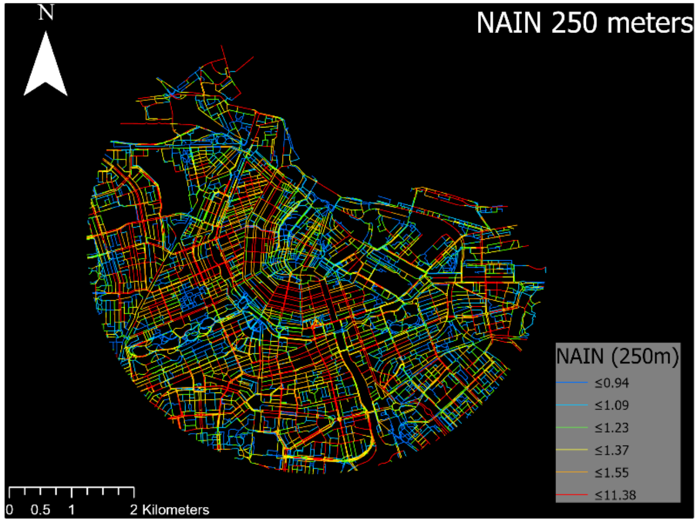
8.2. Space Syntax maps

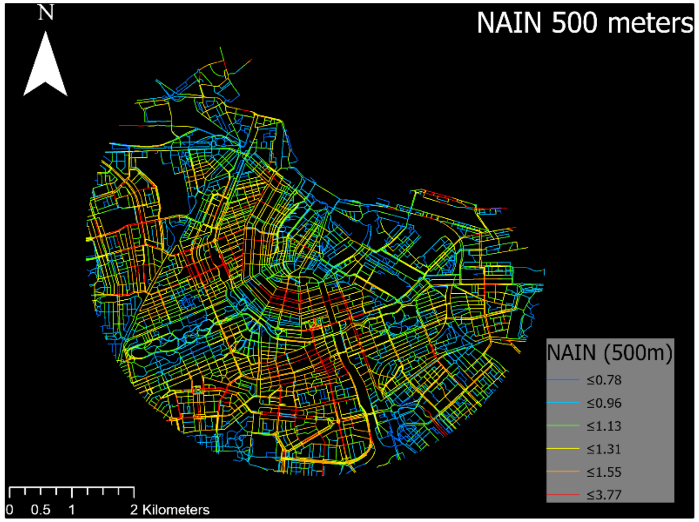


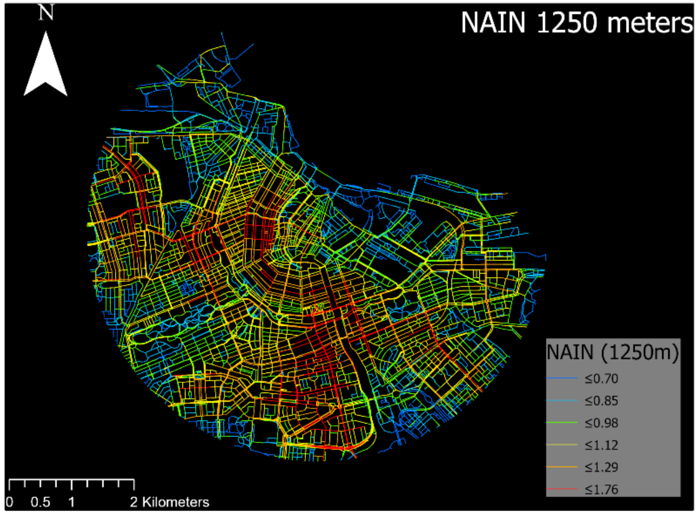


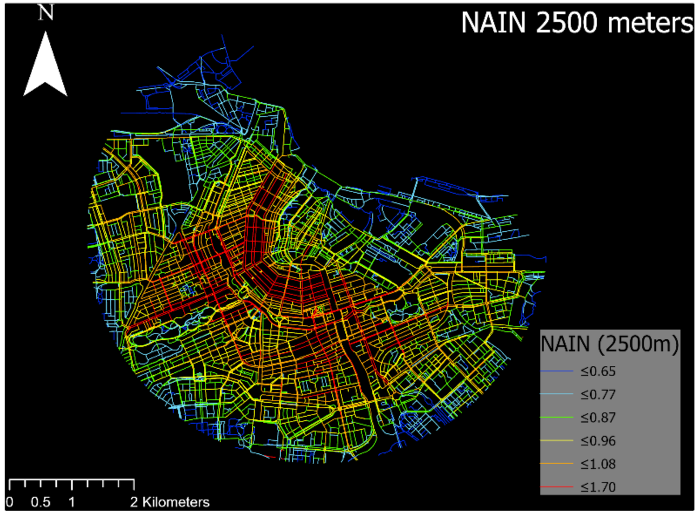




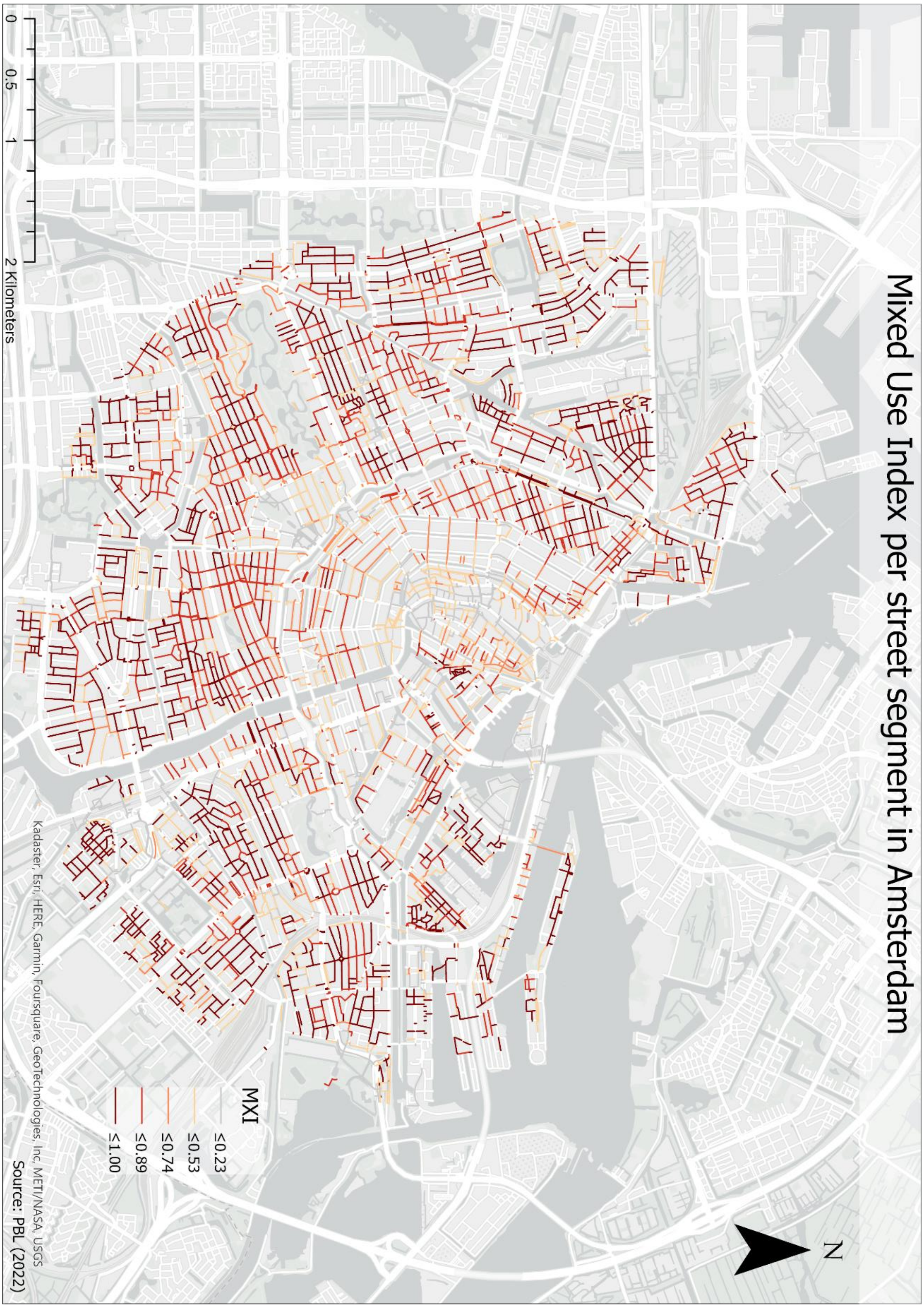

N
NAIN 250 meters
NAIN (250m)
≤0.94
≤1.09
≤1.23
≤1.37
≤1.55
≤11.38
0 0.5 1 2 Kilometers


N
NAIN 500 meters
NAIN (500m)
≤0.78
≤0.96
≤1.13
≤1.31
≤1.55
≤3.77
0 0.5 1 2 Kilometers

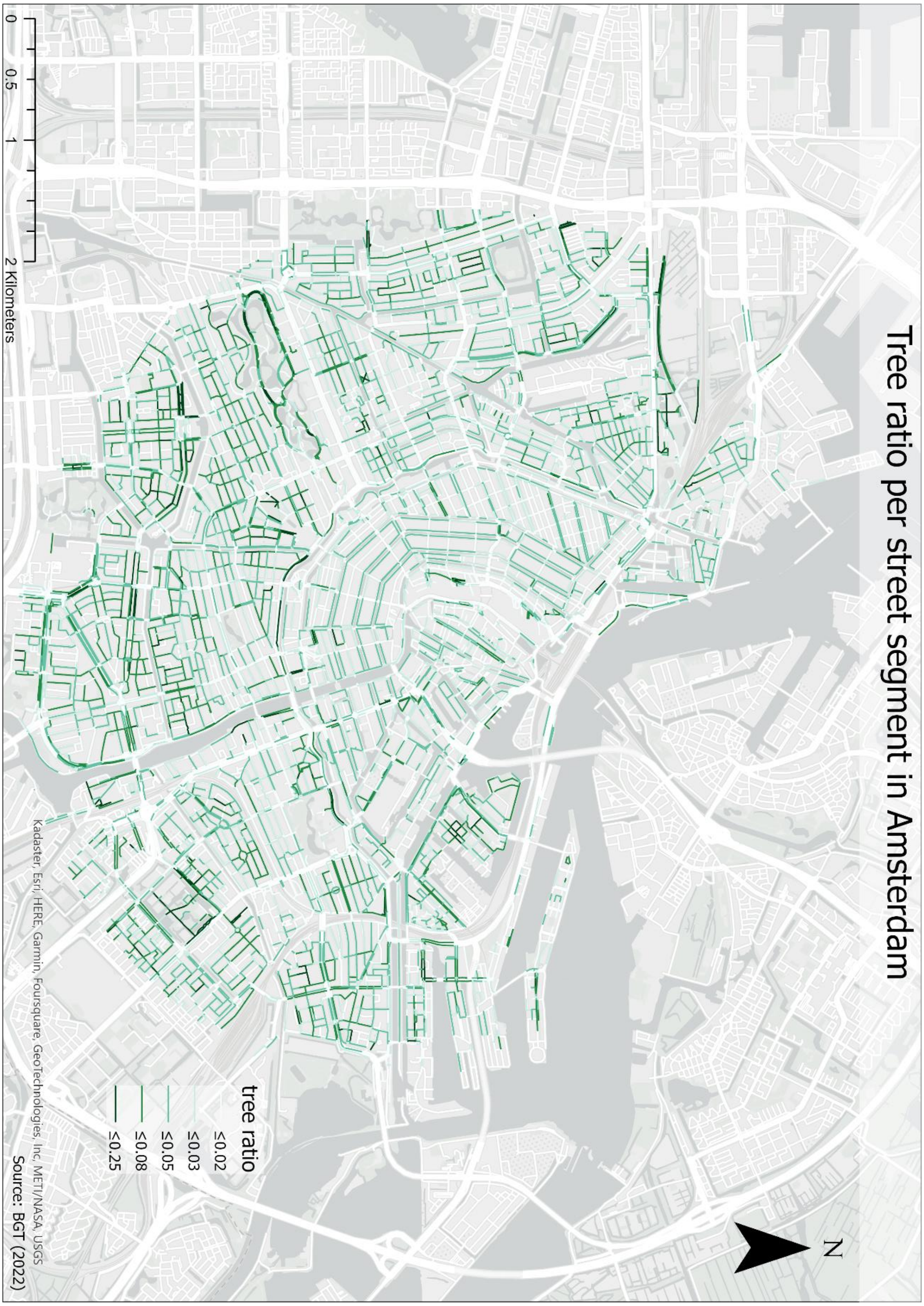

N
NAIN 1250 meters
NAIN (1250m)
≤0.70
≤0.85
≤0.98
≤1.12
≤1.29
≤1.76
0 0.5 1 2 Kilometers


N
NAIN 2500 meters
NAIN (2500m)
≤0.65
≤0.77
≤0.87
≤0.96
≤1.08
≤1.70
0 0.5 1 2 Kilometers

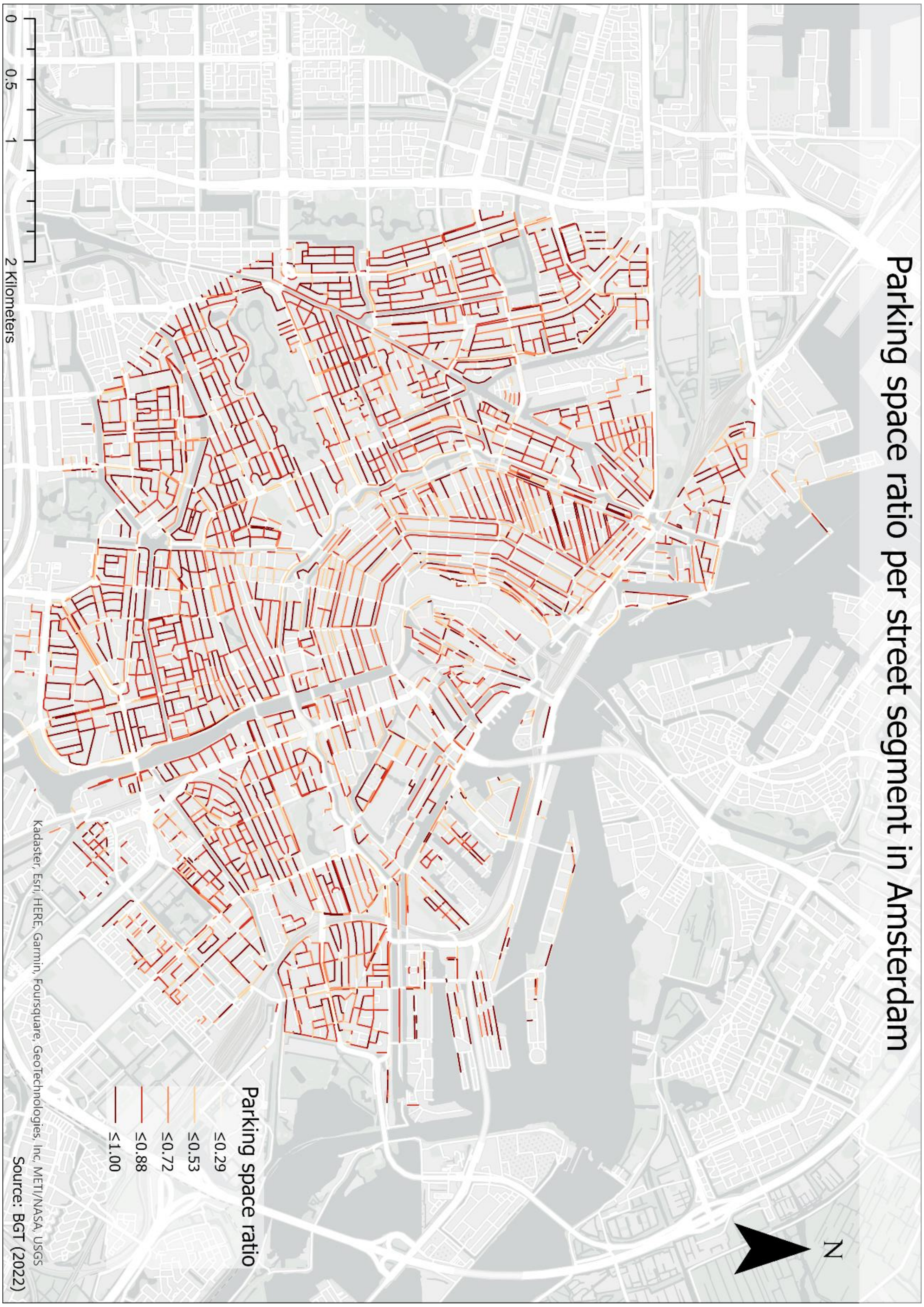
8.3. Other maps

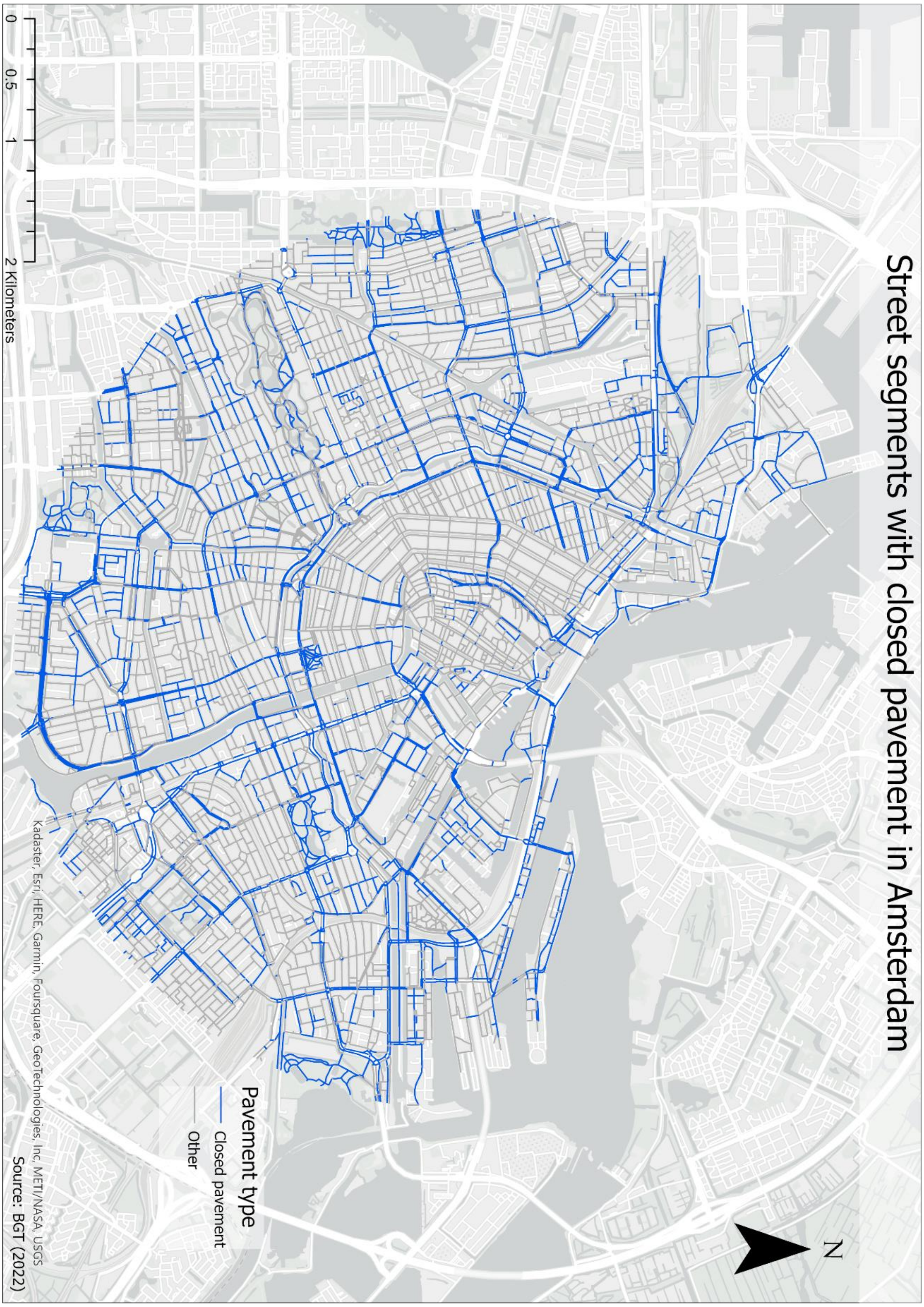

Mixed Use Index per street segment in Amsterdam
MXI
≤0.23
≤0.53
≤0.74
≤0.89
≤1.00
0
0.5
1
2 Kilometers
N
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: PBL (2022)

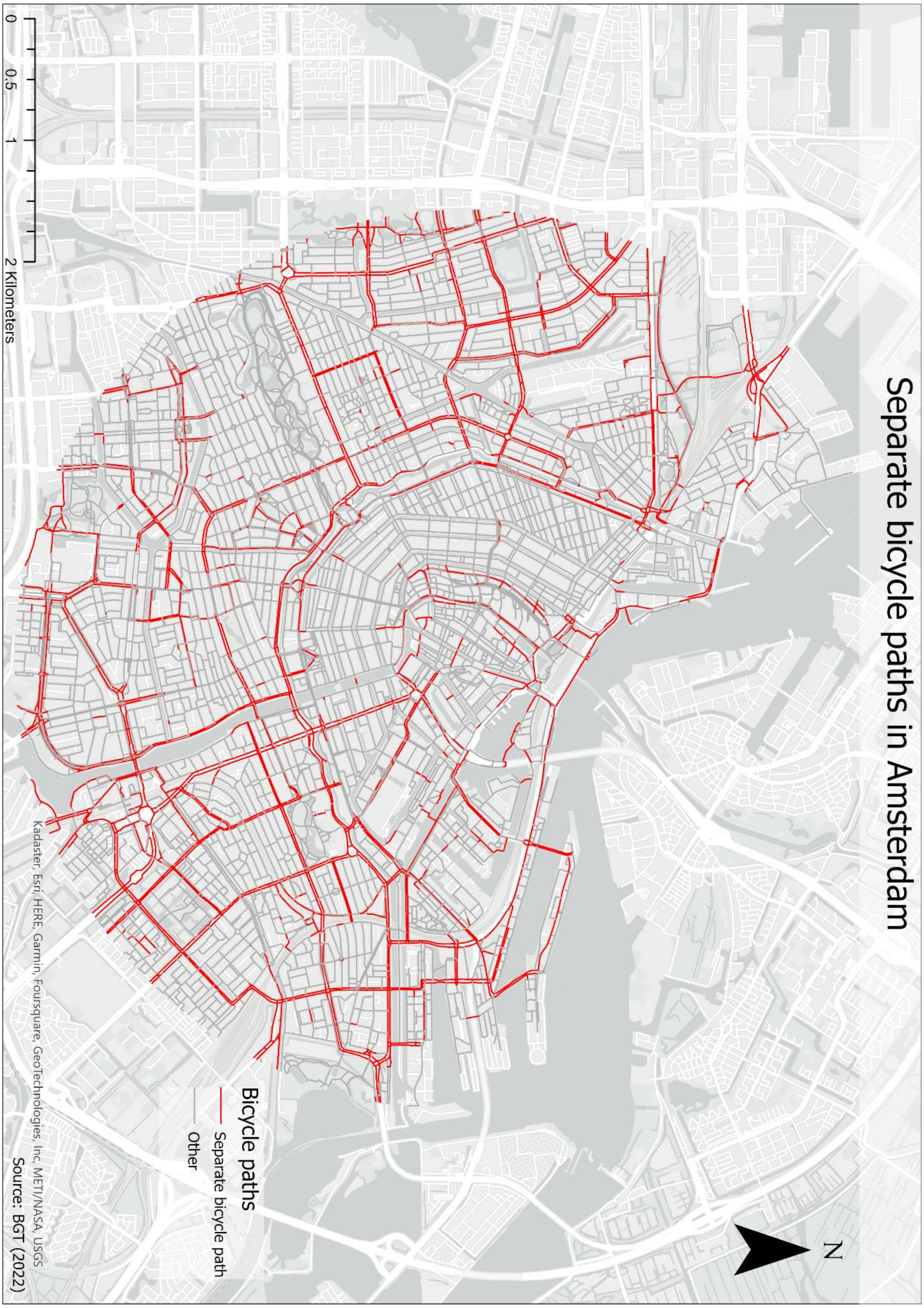

Open Space Ratio per street segment in Amsterdam
N
OSR
≤0.46
≤2.57
≤11.67
≤36.20
≤72.39
0
0.5
1
2 Kilometers
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: PBL (2022)

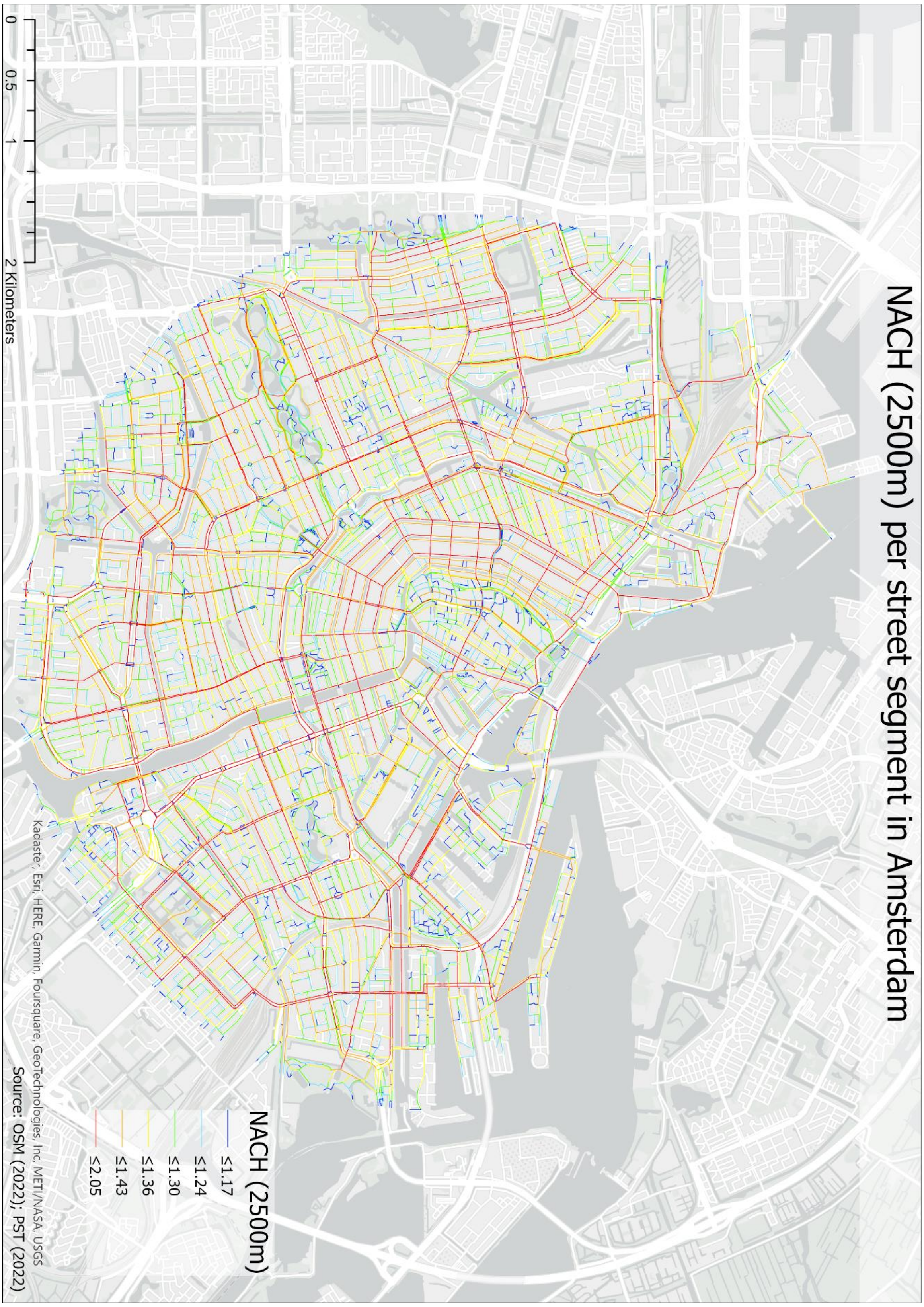

Tree ratio per street segment in Amsterdam
tree ratio
≤0.02
≤0.03
≤0.05
≤0.08
≤0.25
0
0.5
1
2 Kilometers
N
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: BGT (2022)


Water ratio per street segment in Amsterdam
N
Water ratio
0.00
≤0.09
≤0.29
≤0.56
≤0.84
≤1.00
water
0
0.5
1
2 Kilometers
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: BGT (2022)


Parking space ratio per street segment in Amsterdam
Parking space ratio
≤0.29
≤0.53
≤0.72
≤0.88
≤1.00
N
0
0.5
1
2 Kilometers
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: BGT (2022)


Street segments with closed pavement in Amsterdam
Pavement type
Closed pavement
Other
Source: BGT (2022)
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
0 0.5 1 2 Kilometers
N


Separate bicycle paths in Amsterdam
Bicycle paths
Separate bicycle path
Other
Source: BGT (2022)
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
0 0.5 1 2 Kilometers
N


NACH (2500m) per street segment in Amsterdam
NACH (2500m)
≤1.17
≤1.24
≤1.30
≤1.36
≤1.43
≤2.05
0
0.5
1
2 Kilometers
Kadaster, Esri, HERE, Garmin, Foursquare, GeoTechnologies, Inc, METI/NASA, USGS
Source: OSM (2022); PST (2022)

```
library(dplyr)
library(ggplot2)
library(pscl)
library(readr)
library(MASS)
library(ggpubr)
library(RColorBrewer)
library(car)
library(corrplot)

#import the final_dataset_3_6km
be_variables <- read.csv("F:\\MSc
MADE\\Thesis\\Data\\final_netwerk_Clipped_3_6km.csv")

#Change the name of the linknummer variable
be_cleaned <- rename(be_variables, linknummer = LINKNUMMER)
be_cleaned <- rename(be_cleaned, OID = ï..OID_)
be_cleaned <- rename(be_cleaned, intensiteit = INTENSITEI)

#Reclassify "fysiekVoorkomen" (surface material) variable and "cycleway" variable
be_cleaned <- be_cleaned %>% mutate(surface_rc=recode(fysiekVoorkomen_reclassified,
'0'=0,

'1'=0,

'2'=1))
be_cleaned <- mutate(be_cleaned, cycleway_rc=ifelse(HIGHWAY == "cycleway", 1, 0))

#Surface material and cycleway are now numeric -> turn them into factors
be_cleaned$surface_rc <- as.factor(be_cleaned$surface_rc)
be_cleaned$cycleway_rc <- as.factor(be_cleaned$cycleway_rc)

#Change "NA's" to 0's in be_variables_cleaned dataset
be_cleaned["waterdeel_ratio"][is.na(be_cleaned["waterdeel_ratio"])] <- 0
be_cleaned["vegetatieObject_ratio"][is.na(be_cleaned["vegetatieObject_ratio"])] <-
0
be_cleaned["Parking_Space_ratio"][is.na(be_cleaned["Parking_Space_ratio"])] <- 0
be_cleaned["L_OSR_a"][is.na(be_cleaned["L_OSR_a"])] <- 0
be_cleaned["R_OSR_a"][is.na(be_cleaned["R_OSR_a"])] <- 0
be_cleaned["L_MXI_a"][is.na(be_cleaned["L_MXI_a"])] <- 0
be_cleaned["R_MXI_a"][is.na(be_cleaned["R_MXI_a"])] <- 0



#Change the names of the variables
be_cleaned <- rename(be_cleaned, water_ratio = waterdeel_ratio)
be_cleaned <- rename(be_cleaned, ps_ratio = Parking_Space_ratio)
be_cleaned <- rename(be_cleaned, vegetation_ratio = vegetatieObject_ratio)

#Mutate OSR and MXI variables, so that the average of left and right is taken
be_cleaned <- be_cleaned %>% mutate(OSR = (L_OSR_a + R_OSR_a)/2)
be_cleaned <- be_cleaned %>% mutate(MXI = (L_MXI_a + R_MXI_a)/2)

be_cleaned <- subset(be_cleaned, select = c("linknummer","intensiteit",
                                             "water_ratio",
                                             "ps_ratio",
                                             "surface_rc",
                                             "vegetation_ratio",
                                             "cycleway_rc",
                                             "OSR",
                                             "MXI"))

###This marks the end of the Built environment variables data processing###

###Bring in the NACH and NAIN data###
nach_nain <- read.csv("F:\\MSc
MADE\\Thesis\\Data\\Unlinks_testing\\nach_nain_edited_network.csv")
nach_nain <- rename(nach_nain, linknummer = LINKNUMMER)
nach_nain <- nach_nain %>% group_by(linknummer) %>%
summarise(nach1250_sl=mean(AC_sl1250_),
```

```
nach1250_wd=mean(AC_w1250_w),

nach2500_sl=mean(AC_sl2500_),

nach2500_wd=mean(AC_w1250_w),

nach5000_sl=mean(AC_sl5k_wl),

nach5000_wd=mean(AC_w5k_wl_),

nain1250_sl=mean(AI_sl1250_),

nain1250_wd=mean(AI_w1250_w),

nain2500_sl=mean(AI_sl2500_),

nain2500_wd=mean(AI_w2500_w),

nain5000_sl=mean(AI_sl5k_wl),

nain5000_wd=mean(AI_w5k_wl_),
                                                          na.rm=TRUE)
nach_nain <-nach_nain %>% group_by(linknummer) %>%
summarise(nach250_wl=mean(AC_w250_wl_NACH),

nach500_wl=mean(AC_w500_wl_NACH),

nach1250_wl=mean(AC_w1250_wl_NACH),

nach2500_wl=mean(AC_w2500_wl_NACH),

nach250_nw=mean(AC_w250_NACH),

nach500_nw=mean(AC_w500_NACH),

nach1250_nw=mean(AC_w1250_NACH),

nach2500_nw=mean(AC_w2500_NACH),

nain250_wl=mean(AI_w250_wl_NAIN),

nain500_wl=mean(AI_w500_wl_NAIN),

nain1250_wl=mean(AI_w1250_wl_NAIN),

nain2500_wl=mean(AI_w2500_wl_NAIN),

nain250_nw=mean(AI_w250_NAIN),

nain500_nw=mean(AI_w500_NAIN),

nain1250_nw=mean(AI_w1250_NAIN),

nain2500_nw=mean(AI_w2500_NAIN),
                                                         na.rm=TRUE)
###This marks the end of the NACH and NAIN variables data processing###

#Join the NACH & NAIN dataset with the BE dataset
nach_nain_be <- left_join(nach_nain, be_cleaned, by = "linknummer")
write.csv(nach_nain_be,"F:\\MSc
MADE\\Thesis\\Data\\Unlinks_testing\\nach_nain_be.csv")


sample_1 <- sample_n(nach_nain_be, 1669)

#Adjusting the MXI variable so that the relative distance to 0.5 is measured.
sample_1 <- mutate(sample_1, MXI_adjusted=abs(0.5-MXI))

#Time for the linear models
vegetation <- lm(intensiteit ~ vegetation_ratio, data = sample_1)
```

```
water <- lm(intensiteit ~ water_ratio, data = sample_1)
surface <- lm(intensiteit ~ surface_rc, data = sample_1)
cycleway <- lm(intensiteit ~ cycleway_rc, data = sample_1)
pp <- lm(intensiteit ~ ps_ratio, data = sample_1)
OSR <- lm(intensiteit ~ OSR, data = sample_1)
MXI <- lm(intensiteit ~ MXI, data = sample_1)
nach <- lm(intensiteit ~ nach2500_wl, data = sample_1)
nain <- lm(intensiteit ~ nain2500_nw, data = sample_1)

#Summary statistics of the individual independent variables
summary(vegetation)
summary(water)
summary(surface)
summary(cycleway)
summary(pp)
summary(OSR)
summary(MXI)
summary(nach)
summary(nain)
summary(MXI_alt)


#plots of all the individual variables
tree_plot <- ggplot(sample_1, aes(x=vegetation_ratio, y=intensiteit)) +
  geom_point() +
  labs(x="tree ratio", y= "cycling count",
       title= "tree ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

water_plot <- ggplot(sample_1, aes(x=water_ratio, y= intensiteit)) +
  geom_point() +
  labs(x="water ratio per street segment", y= "cycling count",
       title = "water ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

surface_plot <- ggplot(sample_1, aes(x=surface_rc, y=intensiteit)) +
  geom_boxplot() +
  labs(x="surface material smoothness", y="cycling count",
       title="surface material smoothness vs cycling counts per street segment")

cycle_plot <- ggplot(sample_1, aes(x=cycleway_rc, y=intensiteit)) +
  geom_boxplot() +
  labs(x="cyclepath", y="cycling count",
       title="bicycle path vs cycling counts per street segment")

ps_plot <- ggplot(sample_1, aes(x=ps_ratio, y=intensiteit)) +
  geom_point() +
  labs(x="parking space ratio", y="cycling count",
       title = "parking space ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

osr_plot <- ggplot(sample_1, aes(x=OSR, y=intensiteit)) +
  geom_point() +
  labs(x="OSR", y="cycling count",
       title = "OSR vs cycling counts per street segment") +
  geom_smooth(method = "lm")

mxi_plot <- ggplot(sample_1, aes(x=MXI, y=intensiteit)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x="MXI", y="cycling count",
       title = "MXI vs cycling counts per street segment")

nach_plot <- ggplot(sample_1, aes(x=nach2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach2500_wl", y="cycling count",
       title="NACH (2500m) vs cycling counts per street segment") +
  geom_smooth(method = "lm")


### Individual regression plots for all Space Syntax measures ###
```

```
nach250 <- ggplot(sample_1, aes(x=nach250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach250", y="cycling count",
       title = "NACH 250m") +
  geom_smooth(method = "lm")

nain250 <- ggplot(sample_1, aes(x=nain250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain250", y="cycling count",
       title = "NAIN 250m") +
  geom_smooth(method = "lm")

nach500 <- ggplot(sample_1, aes(x=nach500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach500", y="cycling count",
       title = "NACH 500m") +
  geom_smooth(method = "lm")

nain500 <- ggplot(sample_1, aes(x=nain500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain500", y="cycling count",
       title = "NAIN 500m") +
  geom_smooth(method = "lm")

nach1250 <- ggplot(sample_1, aes(x=nach1250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach1250", y="cycling count",
       title = "NACH 1250m") +
  geom_smooth(method = "lm")

nain1250 <- ggplot(sample_1, aes(x=nain1250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain1250", y="cycling count",
       title = "NAIN 1250m") +
  geom_smooth(method = "lm")

nach2500 <- ggplot(sample_1, aes(x=nach2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach2500", y="cycling count",
       title = "NACH 2500m") +
  geom_smooth(method = "lm")

nain2500 <- ggplot(sample_1, aes(x=nain2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain2500", y="cycling count",
       title = "NAIN 2500m") +
  geom_smooth(method = "lm")


figurenachnain1 <- ggarrange(nach250, nain250, nach500, nain500,
                         labels = c("a", "b", "c", "d"),
                         ncol = 2,  nrow = 2)
figurenachnain2 <- ggarrange(nach1250, nain1250, nach2500, nain2500,
                             labels = c("e", "f", "g", "h"),
                             ncol = 2, nrow = 2)

figure <- ggarrange(tree_plot, water_plot, surface_plot, cycle_plot, ps_plot,
osr_plot, mxi_plot, nach_plot,
                    labels = c("1", "2", "3", "4", "5", "6", "7", "8"),
                    ncol = 2, nrow = 4)

figure1 <- ggarrange(mxi_plot, osr_plot, tree_plot,water_plot,
                     labels = c("a", "b", "c", "d"),
                     ncol = 2, nrow = 2)
figure2 <- ggarrange(cycle_plot,surface_plot, ps_plot, nach_plot,
                     labels = c("e", "f", "g", "h"),
                     ncol = 2, nrow = 2)
figure1
figure2
figurenachnain1
figurenachnain2
```

```
#We can see that OSR has a non-significant relation with intensity, OSR will
therefore be left out.
# > This is probably because there is little difference in OSR in the city centre.

#Next step will be to combine the variables in a stepwise manner
mlm_1 <- lm(intensiteit ~ vegetation_ratio + surface_rc, data = sample_1)
summary(mlm_1)
BIC(mlm_1, vegetation)
BIC(mlm_1, surface)

mlm_2 <- lm(intensiteit ~ cycleway_rc + ps_ratio, data = sample_1)
summary(mlm_2)
BIC(mlm_2, cycleway)
BIC(mlm_2, pp)


mlm_3 <- lm(intensiteit ~ MXI + OSR, data = sample_1)
summary(mlm_3)
BIC(mlm_3, MXI)

mlm_4 <- lm(intensiteit ~ water_ratio + nach2500_wl, data = sample_1)
summary(mlm_4)
BIC(mlm_4, water)
BIC(mlm_4, nach)

mlm_5 <- lm(intensiteit ~ vegetation_ratio +
              surface_rc +
              ps_ratio +
              cycleway_rc, data = sample_1)
summary(mlm_5)

mlm_5a <- lm(intensiteit ~ vegetation_ratio +
               surface_rc +
               cycleway_rc, data = sample_1)
summary(mlm_5a)
BIC(mlm_5, mlm_5a)

mlm_6 <- lm(intensiteit ~ MXI +
              nach2500_wl, data = sample_1)
summary(mlm_6)
BIC(mlm_6, mlm_3)
BIC(mlm_6, mlm_4)

mlm_7 <- lm(intensiteit ~ vegetation_ratio +
               surface_rc +
               cycleway_rc +
               MXI +
               nach2500_wl, data = sample_1)
summary(mlm_7)
BIC(mlm_7, mlm_5)
BIC(mlm_7, mlm_6)

mlm_8 <- lm(intensiteit ~ surface_rc +
              cycleway_rc +
              MXI +
              nach2500_wl, data = sample_1)
BIC(mlm_7, mlm_8)


#Na het verhogen van de nach en nain waarden, wordt "waterdeel" stukje bij beetje
minder significant...
#It seems that nach is a better model

#Akaike Information Criterion
AIC(mlm_1, mlm_2, mlm_3, mlm_4, mlm_5, mlm_6, mlm_7)


### The next section is to check for multicollinearity
#First transform the two categorical variables to numeric
sample_num <- sample_1 %>% mutate_at(c('surface_rc', 'cycleway_rc'), as.numeric)
```

```
sample_num1 <- subset(sample_num, select =
c("intensiteit","nach2500_wl","water_ratio","ps_ratio","surface_rc","vegetation_rat
io","cycleway_rc","OSR","MXI"))
sample_num2 <- subs
library(dplyr)
library(ggplot2)
library(pscl)
library(readr)
library(MASS)
library(ggpubr)
library(RColorBrewer)
library(car)
library(corrplot)

#import the final_dataset_3_6km
be_variables <- read.csv("F:\\MSc
MADE\\Thesis\\Data\\final_netwerk_Clipped_3_6km.csv")

#Change the name of the linknummer variable
be_cleaned <- rename(be_variables, linknummer = LINKNUMMER)
be_cleaned <- rename(be_cleaned, OID = ï..OID_)
be_cleaned <- rename(be_cleaned, intensiteit = INTENSITEI)

#Reclassify "fysiekVoorkomen" (surface material) variable and "cycleway" variable
be_cleaned <- be_cleaned %>% mutate(surface_rc=recode(fysiekVoorkomen_reclassified,
'0'=0,

'1'=0,

'2'=1))
be_cleaned <- mutate(be_cleaned, cycleway_rc=ifelse(HIGHWAY == "cycleway", 1, 0))

#Surface material and cycleway are now numeric -> turn them into factors
be_cleaned$surface_rc <- as.factor(be_cleaned$surface_rc)
be_cleaned$cycleway_rc <- as.factor(be_cleaned$cycleway_rc)

#Change "NA's" to 0's in be_variables_cleaned dataset
be_cleaned["waterdeel_ratio"][is.na(be_cleaned["waterdeel_ratio"])] <- 0
be_cleaned["vegetatieObject_ratio"][is.na(be_cleaned["vegetatieObject_ratio"])] <-
0
be_cleaned["Parking_Space_ratio"][is.na(be_cleaned["Parking_Space_ratio"])] <- 0
be_cleaned["L_OSR_a"][is.na(be_cleaned["L_OSR_a"])] <- 0
be_cleaned["R_OSR_a"][is.na(be_cleaned["R_OSR_a"])] <- 0
be_cleaned["L_MXI_a"][is.na(be_cleaned["L_MXI_a"])] <- 0
be_cleaned["R_MXI_a"][is.na(be_cleaned["R_MXI_a"])] <- 0



#Change the names of the variables
be_cleaned <- rename(be_cleaned, water_ratio = waterdeel_ratio)
be_cleaned <- rename(be_cleaned, ps_ratio = Parking_Space_ratio)
be_cleaned <- rename(be_cleaned, vegetation_ratio = vegetatieObject_ratio)

#Mutate OSR and MXI variables, so that the average of left and right is taken
be_cleaned <- be_cleaned %>% mutate(OSR = (L_OSR_a + R_OSR_a)/2)
be_cleaned <- be_cleaned %>% mutate(MXI = (L_MXI_a + R_MXI_a)/2)

be_cleaned <- subset(be_cleaned, select = c("linknummer","intensiteit",
                                            "water_ratio",
                                            "ps_ratio",
                                            "surface_rc",
                                            "vegetation_ratio",
                                            "cycleway_rc",
                                            "OSR",
                                            "MXI"))

###This marks the end of the Built environment variables data processing###

###Bring in the NACH and NAIN data###
nach_nain <- read.csv("F:\\MSc
MADE\\Thesis\\Data\\Unlinks_testing\\nach_nain_edited_network.csv")
```

```
nach_nain <- rename(nach_nain, linknummer = LINKNUMMER)
nach_nain <- nach_nain %>% group_by(linknummer) %>%
summarise(nach1250_sl=mean(AC_sl1250_),

nach1250_wd=mean(AC_w1250_w),

nach2500_sl=mean(AC_sl2500_),

nach2500_wd=mean(AC_w1250_w),

nach5000_sl=mean(AC_sl5k_wl),

nach5000_wd=mean(AC_w5k_wl_),

nain1250_sl=mean(AI_sl1250_),

nain1250_wd=mean(AI_w1250_w),

nain2500_sl=mean(AI_sl2500_),

nain2500_wd=mean(AI_w2500_w),

nain5000_sl=mean(AI_sl5k_wl),

nain5000_wd=mean(AI_w5k_wl_),
                                                            na.rm=TRUE)
nach_nain <-nach_nain %>% group_by(linknummer) %>%
summarise(nach250_wl=mean(AC_w250_wl_NACH),

nach500_wl=mean(AC_w500_wl_NACH),

nach1250_wl=mean(AC_w1250_wl_NACH),

nach2500_wl=mean(AC_w2500_wl_NACH),

nach250_nw=mean(AC_w250_NACH),

nach500_nw=mean(AC_w500_NACH),

nach1250_nw=mean(AC_w1250_NACH),

nach2500_nw=mean(AC_w2500_NACH),

nain250_wl=mean(AI_w250_wl_NAIN),

nain500_wl=mean(AI_w500_wl_NAIN),

nain1250_wl=mean(AI_w1250_wl_NAIN),

nain2500_wl=mean(AI_w2500_wl_NAIN),

nain250_nw=mean(AI_w250_NAIN),

nain500_nw=mean(AI_w500_NAIN),

nain1250_nw=mean(AI_w1250_NAIN),

nain2500_nw=mean(AI_w2500_NAIN),
                                                           na.rm=TRUE)
###This marks the end of the NACH and NAIN variables data processing###

#Join the NACH & NAIN dataset with the BE dataset
nach_nain_be <- left_join(nach_nain, be_cleaned, by = "linknummer")
write.csv(nach_nain_be,"F:\\MSc
MADE\\Thesis\\Data\\Unlinks_testing\\nach_nain_be.csv")


sample_1 <- sample_n(nach_nain_be, 1669)

#Adjusting the MXI variable so that the relative distance to 0.5 is measured.
sample_1 <- mutate(sample_1, MXI_adjusted=abs(0.5-MXI))
```

```
#Time for the linear models
vegetation <- lm(intensiteit ~ vegetation_ratio, data = sample_1)
water <- lm(intensiteit ~ water_ratio, data = sample_1)
surface <- lm(intensiteit ~ surface_rc, data = sample_1)
cycleway <- lm(intensiteit ~ cycleway_rc, data = sample_1)
pp <- lm(intensiteit ~ ps_ratio, data = sample_1)
OSR <- lm(intensiteit ~ OSR, data = sample_1)
MXI <- lm(intensiteit ~ MXI, data = sample_1)
nach <- lm(intensiteit ~ nach2500_wl, data = sample_1)
nain <- lm(intensiteit ~ nain2500_nw, data = sample_1)

#Summary statistics of the individual independent variables
summary(vegetation)
summary(water)
summary(surface)
summary(cycleway)
summary(pp)
summary(OSR)
summary(MXI)
summary(nach)
summary(nain)
summary(MXI_alt)


#plots of all the individual variables
tree_plot <- ggplot(sample_1, aes(x=vegetation_ratio, y=intensiteit)) +
  geom_point() +
  labs(x="tree ratio", y= "cycling count",
       title= "tree ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

water_plot <- ggplot(sample_1, aes(x=water_ratio, y= intensiteit)) +
  geom_point() +
  labs(x="water ratio per street segment", y= "cycling count",
       title = "water ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

surface_plot <- ggplot(sample_1, aes(x=surface_rc, y=intensiteit)) +
  geom_boxplot() +
  labs(x="surface material smoothness", y="cycling count",
       title="surface material smoothness vs cycling counts per street segment")

cycle_plot <- ggplot(sample_1, aes(x=cycleway_rc, y=intensiteit)) +
  geom_boxplot() +
  labs(x="cyclepath", y="cycling count",
       title="bicycle path vs cycling counts per street segment")

ps_plot <- ggplot(sample_1, aes(x=ps_ratio, y=intensiteit)) +
  geom_point() +
  labs(x="parking space ratio", y="cycling count",
       title = "parking space ratio vs cycling counts per street segment") +
  geom_smooth(method = "lm")

osr_plot <- ggplot(sample_1, aes(x=OSR, y=intensiteit)) +
  geom_point() +
  labs(x="OSR", y="cycling count",
       title = "OSR vs cycling counts per street segment") +
  geom_smooth(method = "lm")

mxi_plot <- ggplot(sample_1, aes(x=MXI, y=intensiteit)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x="MXI", y="cycling count",
       title = "MXI vs cycling counts per street segment")

nach_plot <- ggplot(sample_1, aes(x=nach2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach2500_wl", y="cycling count",
       title="NACH (2500m) vs cycling counts per street segment") +
  geom_smooth(method = "lm")
```

```
### Individual regression plots for all Space Syntax measures ###
nach250 <- ggplot(sample_1, aes(x=nach250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach250", y="cycling count",
       title = "NACH 250m") +
  geom_smooth(method = "lm")

nain250 <- ggplot(sample_1, aes(x=nain250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain250", y="cycling count",
       title = "NAIN 250m") +
  geom_smooth(method = "lm")

nach500 <- ggplot(sample_1, aes(x=nach500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach500", y="cycling count",
       title = "NACH 500m") +
  geom_smooth(method = "lm")

nain500 <- ggplot(sample_1, aes(x=nain500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain500", y="cycling count",
       title = "NAIN 500m") +
  geom_smooth(method = "lm")

nach1250 <- ggplot(sample_1, aes(x=nach1250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach1250", y="cycling count",
       title = "NACH 1250m") +
  geom_smooth(method = "lm")

nain1250 <- ggplot(sample_1, aes(x=nain1250_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain1250", y="cycling count",
       title = "NAIN 1250m") +
  geom_smooth(method = "lm")

nach2500 <- ggplot(sample_1, aes(x=nach2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nach2500", y="cycling count",
       title = "NACH 2500m") +
  geom_smooth(method = "lm")

nain2500 <- ggplot(sample_1, aes(x=nain2500_wl, y=intensiteit)) +
  geom_point() +
  labs(x="nain2500", y="cycling count",
       title = "NAIN 2500m") +
  geom_smooth(method = "lm")


figurenachnain1 <- ggarrange(nach250, nain250, nach500, nain500,
                         labels = c("a", "b", "c", "d"),
                         ncol = 2,  nrow = 2)
figurenachnain2 <- ggarrange(nach1250, nain1250, nach2500, nain2500,
                             labels = c("e", "f", "g", "h"),
                             ncol = 2, nrow = 2)

figure <- ggarrange(tree_plot, water_plot, surface_plot, cycle_plot, ps_plot,
osr_plot, mxi_plot, nach_plot,
                    labels = c("1", "2", "3", "4", "5", "6", "7", "8"),
                    ncol = 2, nrow = 4)

figure1 <- ggarrange(mxi_plot, osr_plot, tree_plot,water_plot,
                     labels = c("a", "b", "c", "d"),
                     ncol = 2, nrow = 2)
figure2 <- ggarrange(cycle_plot,surface_plot, ps_plot, nach_plot,
                     labels = c("e", "f", "g", "h"),
                     ncol = 2, nrow = 2)
figure1
```

```
figure2
figurenachnain1
figurenachnain2

#We can see that OSR has a non-significant relation with intensity, OSR will
therefore be left out.
# > This is probably because there is little difference in OSR in the city centre.

#Next step will be to combine the variables in a stepwise manner
mlm_1 <- lm(intensiteit ~ vegetation_ratio + surface_rc, data = sample_1)
summary(mlm_1)
BIC(mlm_1, vegetation)
BIC(mlm_1, surface)

mlm_2 <- lm(intensiteit ~ cycleway_rc + ps_ratio, data = sample_1)
summary(mlm_2)
BIC(mlm_2, cycleway)
BIC(mlm_2, pp)


mlm_3 <- lm(intensiteit ~ MXI + OSR, data = sample_1)
summary(mlm_3)
BIC(mlm_3, MXI)

mlm_4 <- lm(intensiteit ~ water_ratio + nach2500_wl, data = sample_1)
summary(mlm_4)
BIC(mlm_4, water)
BIC(mlm_4, nach)

mlm_5 <- lm(intensiteit ~ vegetation_ratio +
              surface_rc +
              ps_ratio +
              cycleway_rc, data = sample_1)
summary(mlm_5)

mlm_5a <- lm(intensiteit ~ vegetation_ratio +
               surface_rc +
               cycleway_rc, data = sample_1)
summary(mlm_5a)
BIC(mlm_5, mlm_5a)

mlm_6 <- lm(intensiteit ~ MXI +
              nach2500_wl, data = sample_1)
summary(mlm_6)
BIC(mlm_6, mlm_3)
BIC(mlm_6, mlm_4)

mlm_7 <- lm(intensiteit ~ vegetation_ratio +
               surface_rc +
               cycleway_rc +
               MXI +
               nach2500_wl, data = sample_1)
summary(mlm_7)
BIC(mlm_7, mlm_5)
BIC(mlm_7, mlm_6)

mlm_8 <- lm(intensiteit ~ surface_rc +
              cycleway_rc +
              MXI +
              nach2500_wl, data = sample_1)
BIC(mlm_7, mlm_8)


#Na het verhogen van de nach en nain waarden, wordt "waterdeel" stukje bij beetje
minder significant...
#It seems that nach is a better model

#Akaike Information Criterion
AIC(mlm_1, mlm_2, mlm_3, mlm_4, mlm_5, mlm_6, mlm_7)
```

```
### The next section is to check for multicollinearity
#First transform the two categorical variables to numeric
sample_num <- sample_1 %>% mutate_at(c('surface_rc', 'cycleway_rc'), as.numeric)
sample_num1 <- subset(sample_num, select =
c("intensiteit","nach2500_wl","water_ratio","ps_ratio","surface_rc","vegetation_rat
io","cycleway_rc","OSR","MXI"))
sample_num2 <- subset(sample_num, select =
c("nach2500_wl","surface_rc","vegetation_ratio","cycleway_rc","MXI"))
corrplot(cor(sample_num2))
corrplot(cor(sample_num1))

#VIF
vif(mlm_7)et(sample_num, select =
c("nach2500_wl","surface_rc","vegetation_ratio","cycleway_rc","MXI"))
corrplot(cor(sample_num2))
corrplot(cor(sample_num1))

#VIF
vif(mlm_7)
```