Data Exploration for Generative Design Research

Kun, Péter; Mulder, Ingrid; Kortuem, Gerd

DOI
10.21606/dma.2018.565

Publication date
2018

Document Version
Final published version

Published in
Proceedings of DRS 2018

DRS 2018 Design Research Society 2018 **Catalyst**

University of Limerick
25th – 28th June 2018

Data Exploration for Generative Design Research

KUN Peter*; MÜLDER Ingrid and KORTUEM Gerd

Delft University of Technology
*Corresponding author e-mail: p.kun@tudelft.nl
doi: 10.21606/dma.2018.565

The current work elaborates upon a Generative Data Exploration method, which is a design technique aiming at supporting designers in integrating data in their design activities. Digital data offers new opportunities in all sort of professional domains, yet existing approaches and tools to manipulate data are predominantly targeted at data experts. As access to data is becoming democratised, new types of techniques are needed to leverage the agency of designers and to empower them to utilise data in the design process. Designers without prior data experience can benefit from the techniques, know-how, best practices of experts, if such expert knowledge is codified in design methods and tools. The aims of a Generative Data Exploration method are two-fold. First, the method facilitates a learning curve on gaining holistic data literacy. Second, the method supports designing where digital data, exploration of data and sense-making of data is part of the process.

design methods; data exploration; generative design; fuzzy front-end

1 Introduction

The abundance of digital data has been gaining presence in all areas of life. This datafication trend has been quantifying and digitally describing everyday phenomena, from how individuals are connected to each other to complex sensor systems continuously collecting digital data about the physical world (Lycett, 2013). Under digital data, not aiming for a comprehensive list, we refer to quantitative data, sensor data, open data, data in databases and so forth. Access to such kinds of data is not limited anymore to data experts (analysts, engineers, developers, etc.), but oftentimes to be found in public (e.g., open data) or can be captured relatively easily by anyone (e.g., citizen science and collecting bottom-up environmental data).

The usage of various design methods, techniques or tools has been common in all genres of design for decades. Starting from the seminal “Design Methods” book (John Christopher Jones, 1970), many method and tool collections have appeared to support the conduction of the various steps of the design process. More recent design approaches, such as participatory design or co-design, have established an increasing number of tools and methods utilised at the early phase of design (Sanders & Stappers, 2008). Another area of tools are “Creativity Support Tools” (Shneiderman, 2007) that have made previously complex tasks much easier to be conducted within the design process and by designers (such as using CAD systems for form-giving of physical artefacts). These kinds of codification of knowledge as methods, techniques or tools supports designers of any level of experience to unlock advanced technology and to integrate new techniques into their toolbox.

Earlier, research on data exploration has been focused mainly on two strands; a mathematics-based direction using statistics to describe datasets and to create models to describe phenomena (e.g., Tukey, 1962; Tukey, 1977), and a cognition-based direction using information processes to support domain experts in their sense-making of datasets (e.g., Card, Mackinlay, & Shneiderman, 1999). Compared with these two directions, exploring data from a design practice perspective is still in its infancy. However, considering the trends of the growing ubiquity of digital data, for the future it is inevitable that designers will need to be able to integrate already existing data into their design process or be able to better collaborate with data experts to do so. Speed and Oberlander (2016) have recently presented a theoretical framework to distinguish *"designing from, with and by data"* to categorise existing data approaches. However, little work to date has been done to explore supporting designers with tools to be able to integrate data into the design process, and to link techniques and know-how from data science with design, especially via formats that can be integrated into the design process. In the current work, we present a design method – Generative Data Exploration – to support generative design research. We have designed this Generative Data Exploration method based on earlier work on design tools and creativity support tools. In the upcoming sections, first we position related work, and then introduce the design method we created. Afterwards, we report on an empirical study we conducted with novice designers using the method, and then we discuss the value the method provides for generative design research as well as the impact the integration of data into the design practice might mean.

2 Related work

2.1 Non-experts learning and using data

The field of data science has matured a lot in the past decade (Cao, 2017), and as a consequence, the foundations of teaching data competencies has also evolved: holistic approaches to teach data in undergraduate education has started to take place, teaching a full spectrum of tools to prepare students working with data in real settings (Baumer, 2015). This tactic helps to learn how to think with data, from asking a question that leads the data analysis inquiry and to communicate findings. Compared to this method from formal education, alternative approaches have appeared as well; Hill and colleagues (2017) present their experiences of teaching basic data science skills through community workshops *"democratizing data* science". Their approach is built on teaching basics of programming for the very aim of doing data science, namely to be able to ask questions from a data, acquire data from online sources and to be able to analyse and visualise such data. The approach by Hill and colleagues provides a flexible set of skills and tools, however with the price of a steep learning curve. D'Ignazio and Bhargava (2016) have approached this space from a different angle. They have created a set of learning tools for data literacy, that explicitly avoids programming, and targets data skill acquisition via tailored, single-purposed data tools – DataBasic – that nevertheless can be used with actual datasets and for actual visualisation and analysis. In another work, D'Ignazio (2017) adds to this work on her experiences with applying data literacy (and its teaching) put into creative work, such as design. Data directly applied in the design process, Bigelow and colleagues (2014) have explored how designers work with data to create visualisations, and Dove and Jones (2014) have shown ways how to inject visualised data (thus, a layer of abstraction over raw data) into co-design activities, and to stimulate creative thinking. These works indicate that visualisation contributes to the sense-making process with data, not necessarily as the outcome of the design process, but as interim thinking tools.

2.2 Toolkits

Sanders, Brandt and Binder (2010) provide an overview and categorization of the tools and techniques for participatory design. In their terms, *tools* are *"material components used in PD activities"*; *toolkit* is a collection of tools used in combination for a specific purpose; *technique* is a description how tools and techniques are put into action; *method* is a combination of tools, toolkits, techniques put together strategically towards a specific design research plan, and at last, *approach* is an overall mindset for conducting the design research plan. In the current work, we expand on this terminology with *know-how*: best practices, practical tacit knowledge turned into explicit that normally comes with experience in a domain.

Data on its own is a rather generic material, and thus we primarily relate to toolkits that support generic processes or that can expand to varying levels of abstraction. Card-deck based tools seem to fit these criteria: card-decks have been effectively used in Information Visualisation – to learn about how to use data and design visualisations (He and Adar, 2016). Many other card-deck based tools are supporting the ideation phase of design in different ways: the Tango Cards serve as an example of bringing theoretical academic work into design practice (Deng, Antle, & Neustaedter, 2014); Hornecker's card brainstorming game turns a theoretical framework into a design tool (2010), and provides an account on how the use of design tools be used in facilitated setups, like a workshop.

To support the design of such tools, theory from Human-Computer Interaction can help: earlier work in Creativity Support Tools laid down design principles (Resnick, Myers, Nakakoji, & Shneiderman, 2005; Shneiderman, 2007), such as: *"Designing with low thresholds, high ceilings, wide walls".* This principle stands for a tool's desired attributes to be easy for novices to begin using, but provide functionalities that experts need, and if possible to provide additional functionalities to keep the number of tools involved in a workflow low.

The Generative Data Exploration method

This section presents our proposed Generative Data Exploration method. Following the terminology of Sanders, Binder and Brandt (2010), we present a method, consisting of a workshop methodology to conduct a time-pressured workshop, suggested *software tools*, and *design tools* we created to support the process. The aim of the method is to empower designers without expertise in data to be able to creatively use data in their design process and to ease the learning curve for gaining data literacy for design.

2.3 Rationale

Our method elaborates upon the four levels of creativity as defined by Sanders and Stappers (2008); *Doing, Adapting, Making, Creating*, which refer to an increasing order of expertise/interest necessary for each level:

- *Doing:* The level of *Doing* – being able to transform a dataset independent of a tool (thus having a sense of how to manipulate a dataset) is part of a generic technical literacy, at least through basic knowledge of spreadsheets software (e.g., Excel).
- *Adapting:* This is the level where appropriation of techniques starts to happen. This appropriation can be guided and inspired; novice data designers appropriating data thinking and data techniques into their processes.
- *Making:* The level of *Making* is *'asserting own ability or skill'*, which we see as the utilisation of data commonly in one's design practice.
- *Creating:* The level of *Creating* is the highest level of expertise/interest in this spectrum, addressing such cases that truly transforms the design practice intertwined with data.

Considering designers' relatively low level of data expertise, we assume that most designers today would be on the levels of *Doing* and *Adapting* to utilise data. Thus, in our design rationale, we mainly address the levels of *Doing* and *Adapting*; with the current work our aim is to create such a method for designers, that builds confidence for designers to *Do* with data, and *Adapt* it, appropriating data techniques for their design process.

2.4 *Design principles*

After revising related work and previous workshops we have held, we have concluded the following key design principles for data design tools for the fuzzy-front end:

- Data design tools should be open-ended; data can come in various shapes, formats, and topics, and the tools need to accommodate for this broad variety.
- Data design tools should integrate into a generic design process; the design process differs from person to person, thus the tools need to be familiar for designers and compatible with mainstream design tools.
- Data design tools should serve hands-on doing with data; as opposed to tools made for learning, the designed tools shall be used in real design situations.
- Data design tools should support for exploration; analytics tools for data support the process of deducting/inducting insights from data, but what designers need are support to find inspiration.
- Data design tools should generate outcomes that are valuable for the design process; the tools need to take real input into account (instead of requiring an over-abstracted input), and generate outcomes that can be actionable in the design process.
- Data design tools should help navigating through the complex world of data and data techniques; data has been black-boxed for designers, and the early learning curve is daunting. Thus, tools should help the early phase, showing designers a clear path to follow.

From previous workshops we we learnt that novice designers have foundational (or more) tacit knowledge on data and visualisation, however this knowledge needs to be made explicit. Designers have generally been exposed to visualisations (e.g., scatterplots, network graphs and more), but making sense of them might have been only an intuitive process that could be led by guiding questions.

It seemed that card decks as an approach is proven to be successful to these design principles, especially on the principles of open-endedness and suiting a generic design process. In the following section, we present the Generative Data Exploration method, including the workshop methodology and the various design tools we designed.

2.5 *Workshop methodology*

In keeping with the generic data process from Baumer (2015), the following workshop structure has been developed (see Table 1).

Related versions of this workshop process have been tested in one-day (n=20 and n=38) and three-days (n=26) workshop settings.

Throughout the workshop, we have selected the following software tools for certain tasks. The criteria for the tools are:

- Open source or publicly available for free;
- Working on the major computer platforms (or on the web);
- Easy to learn, providing a high ceiling on functionalities;
- Supporting a non-programmatic workflow with data.

Table 1 Workshop proceeding overview with the basic activities.

Workshop activity (in sequential order)	**Description**
Receiving the design brief	The participants receive the (design) brief. Depending on the available time and the scope of the workshop, a brief can be to 'find three valuable insights that would be interesting for designers to continue a design process with', related to the context of the dataset.
Opening or acquiring the dataset	Being able to open a dataset is an essential step to manipulate it later on. When the data comes in various formats, it may happen that additional steps (such as converting or extracting data from an API is necessary).
Setting direction	To set a direction for the inquiry, the participants are asked to first brainstorm and discuss the topic and formulate three initial research questions or data hypothesis.
Data transformation	Datasets most often require cleaning or steps of transformations based on the specific needs. Further data transformations involve various filtering or sorting and potentially deriving additional data (e.g., adding an additional column of time differences between two timestamps).
Data exploration	Data exploration is done by applying various data analysis techniques on the dataset to extract additional meaning. Such as, visual analytics can show relations between many data points, network analysis can show characteristics of relational data, and so forth. This step is 'messy'; explorative and looking for designerly inspiration.
Communicating the insights	The participants are asked to present their insights, preferably in a visual format. This provides focus and closure for the end of the workshop.

The main recommended tools:

- **OpenRefine** (OpenRefine, 2017)**:** this tool provides functionalities to clean data and to do various data transformations on data, without programming knowledge. Spreadsheet software (i.e., Excel) could perform such functionalities as well, however not as robustly, especially on non-numerical data.
- **RAWGraphs** (Mauri, Elli, Caviglia, Uboldi & Azzi, 2017)**:** this tool provides visualisations beyond the typical charting options of spreadsheet software (such as bar charts, etc.). It is easy to use and the generated visualisations can be exported to vector formats for further editing and additional graphic design work.

Learning the software tools are not the focus of the workshop, thus they can be replaced with better or more appropriate software tools without any further change. Beyond OpenRefine and RAWGraphs, we encourage the use of a familiar spreadsheet software (e.g., Microsoft Excel, Google Sheets, Apple Numbers) and a text editor (e.g., Sublime Text, Atom) for "quick-and-dirty" text operations.

2.6 Design tools

Our Generative Data Exploration method utilises the card decks and booklets we designed to scaffold a variety of data know-how. We aimed to generate card decks and booklets that can be (and preferably be) tailored for certain datasets and situations. This may happen in ways to create additional cards, or to provide cards that dissect a dataset (e.g., different columns or rows as separate card decks). The general aim is to make the comprehension of a given dataset as simple as possible. Making it tangible and off-screen supports novices to better be able to think about it and get the process going, without data transformations and data visualisations.

The actual activities of how to use the card decks and the booklets are left *un-designed*. There are typical activities to do with design card decks, such as *forced pairing* of cards to trigger ideas, or combine the cards to *reverse engineer* and model existing projects. At the current stage, we find the need to have the card decks and booklets used in more settings to conclude suggested activities for their use.

The following of the section shows the card decks and booklets in detail.

2.6.1 Card decks

Basic data types and techniques: The basic data cards provide a quick overview of the basic types of data, and the most common techniques that can be applied on datasets (see Figure 1). They can be used as a reminder of alternative options, as well as a quick reference to navigate through a dataset. One part of the basic data card deck is the cards summarising the various types of data, such as: numerical data, geo-located data, categorical data, textual data, etc. – the most common types of data one can find describing everyday phenomena. Another part of the basic data card deck is the fundamental activities one can perform with data, such as: compare or identify. These activities are so common, that they go unnoticed in most cases. However, when someone is pursuing computational thinking, these activities become very obvious (such as selecting a datapoint - *identify*).

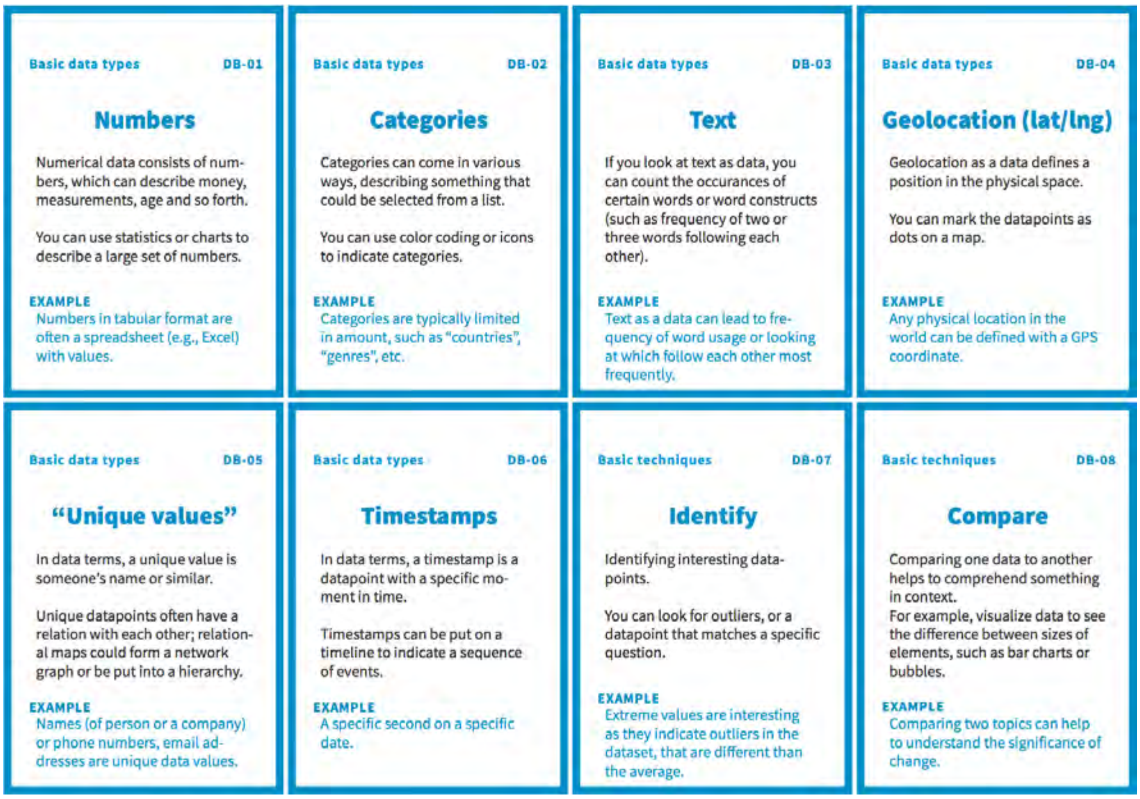

Figure 1. The Basics of data card deck summarises the most elementary data types and data techniques.

Data techniques: The data techniques card deck is a summary of the most typical techniques to apply on a dataset, in order to extract further meaningful information out of the data (see Figure 2). A typical data technique example is *map visualisation*, which can easily be done when there are e.g., GPS coordinates in the dataset. The related data technique card provides a basic overview of what kind of input(s) the technique requires (e.g., GPS coordinates, addresses). One explicit aim of the data techniques card deck is to trigger additional techniques to use for those that are more experienced with data, and in this way to stretch their boundaries. For novices, the techniques are a guided effort to follow their learning curve.

Extendibility: These card decks are just initial decks; normally, they should be tailored to specific datasets or design situations. Such as, relational data, like metadata from a library's records, or open data containing the types of street artefacts and their locations, will probably need different data techniques to extract meaningful information out of them. Furthermore, bespoke card decks can support any layer of abstraction; a card deck of different visualisation charting options could be very valuable when the dataset is full of numbers and categorical data, providing a more detailed level of cards than the *Graph visualisation* card from Figure 2.

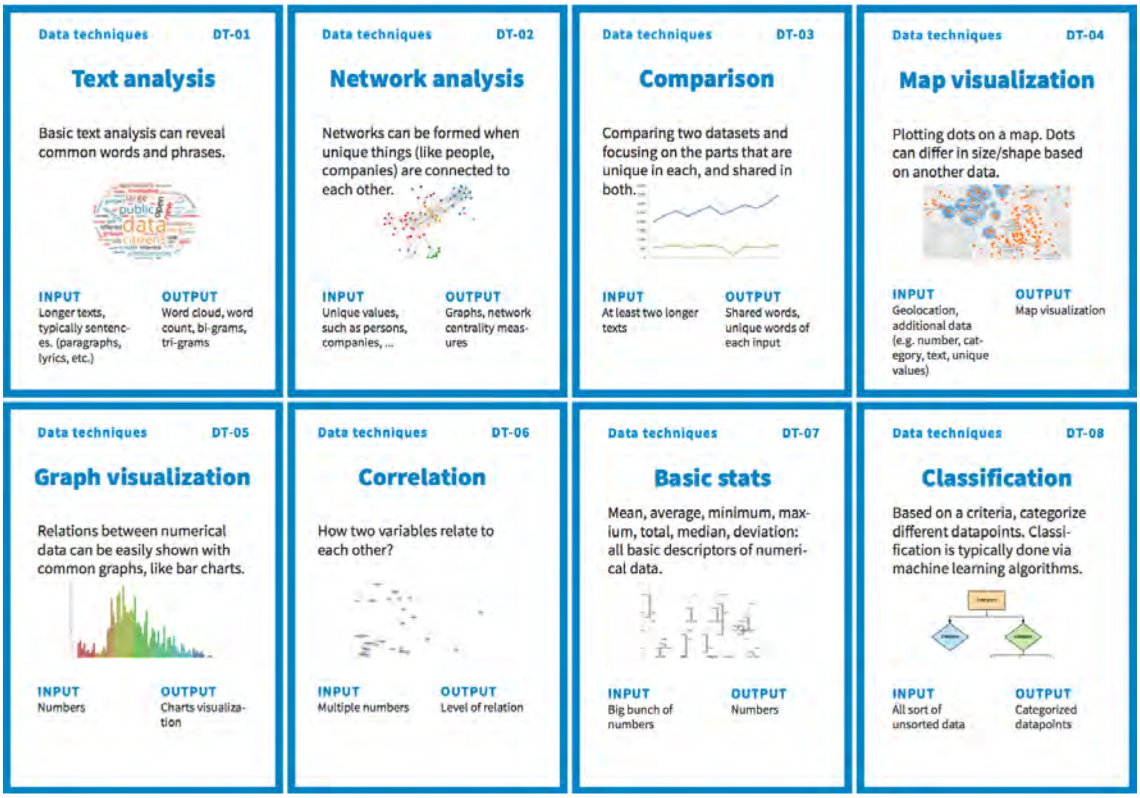

Figure 2. The Data techniques card deck summarises common techniques to extract information out of data.

2.6.2 Booklets

Questions for data: The "Questions for Data" booklet provides guidance for the users of the method to get them unstuck (see Figure 3). The booklet is based on the insight that at first, it is daunting to open a new dataset without knowing its content. The booklet contains triggering questions that can hint towards a successful strategy to process the dataset. Depending on the situation of being stuck, these questions attend the cases of:

- Looking at raw data and not knowing what is the next step;
- Looking at a visualisation and not knowing how to read it;
- Looking at data and not knowing how to extract further insights from it.

The questions in this booklet may state obvious ideas, but having these questions tangible, constantly available around data processing serves as a reminder that it is normal to be stuck, and in that case the way out is shifting the thinking process.

Working with data: The "Working with Data 101" booklet is a practical quick-start guide from opening a comma-separated value (CSV) file – a very typical format for datasets –, to doing more advanced data operations on it (see Figure 4). The booklet contains tips and tricks for the most typically conducted data operations, such as filtering or sorting data, in order to save time during the design process looking up how to do these operations, as well as to emphasise the right terminology in case the user wants to search for further information.

Extendibility: The booklets have been made 8-pages long to keep a concise format, as well as to be able to print and fold it easily. Similar to the card decks, the booklets can be tailored for specific datasets or design situations. The following section presents the empirical study conducted to assess the validity of the design of our Generative Data Exploration method.

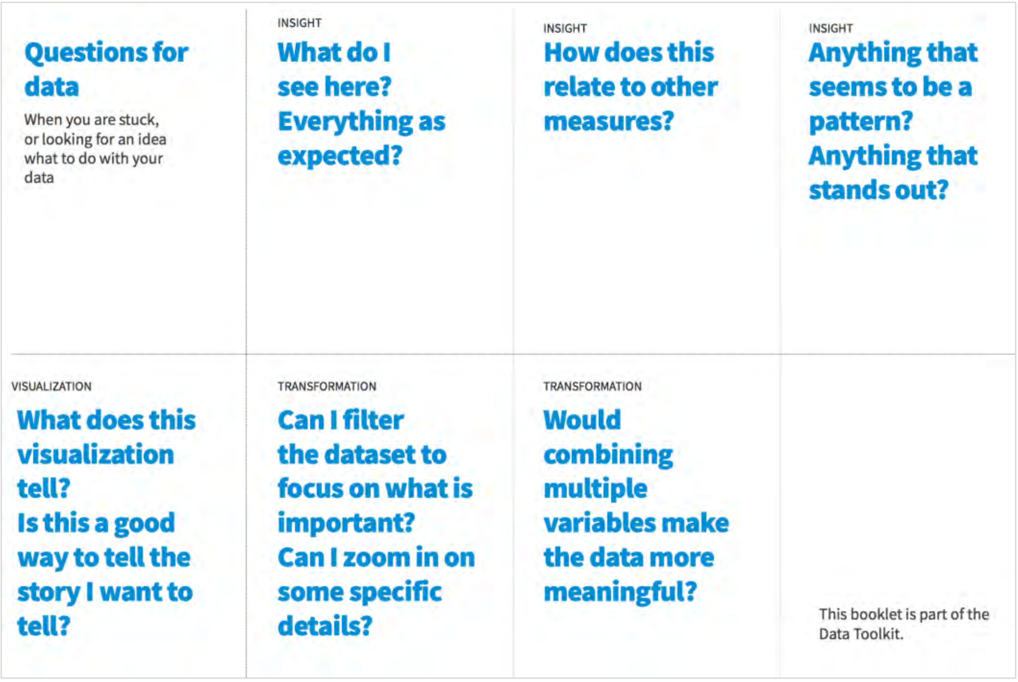

Figure 3. The Question for data booklet contains triggering questions to extract insight from a dataset or visualisation or to inspire next steps of the data transformation.

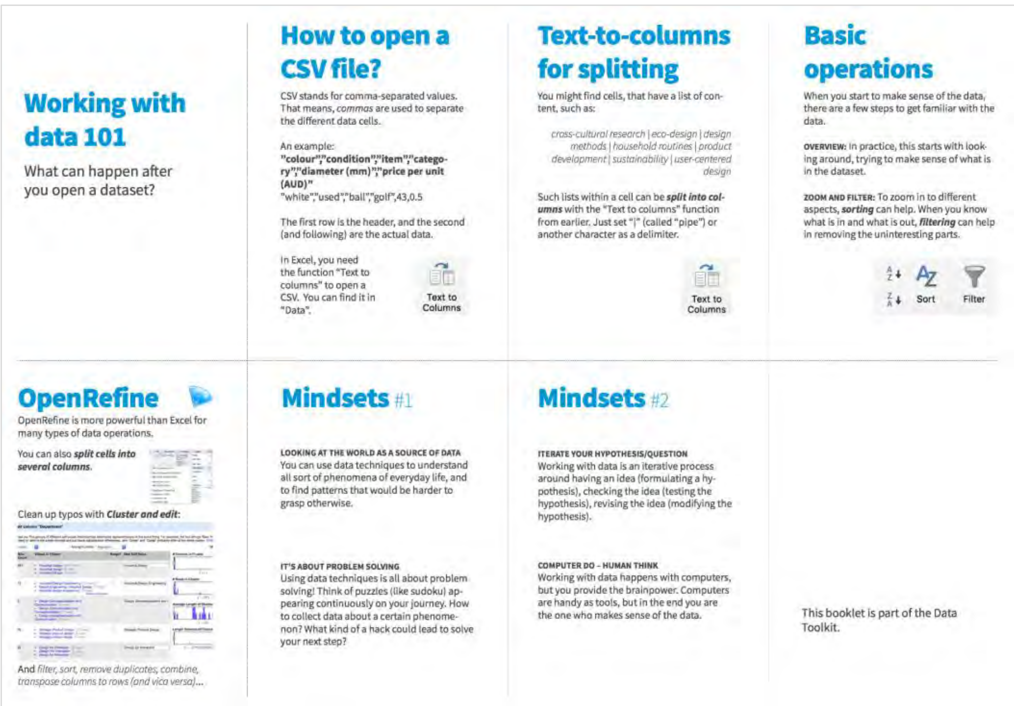

Figure 4. The Working with data 101 booklet contains practical knowledge how to open and manipulate a dataset in CSV format

3 Study setup

We conducted a pilot study to understand how our Generative Data Exploration method was useful in weaving data techniques into the design process and to assess the usability of the approach with novice designers (i.e., design students). Following the descriptions introduced in the previous section, this section details the setup and presents the methodology used. We assumed, that design students likely have tacit data knowledge that might inform their approach with data. Differently put, we expected the participants to have average familiarity with spreadsheet software (e.g., Excel) and familiarity with common visualisation techniques (e.g., charts, graphs).

Figure 5. Impressions from the workshop and the study setup.

Participants and setup: Thirteen students (7 female, 6 male) participated in the current study, as a one-day elective class. The students were first year master students in different orientations of design (strategic design, n= 1; interaction design/user research, n=5; industrial/product design, n=6), of a large, European industrial design faculty. The thirteen participants all had a bachelor degree in design. The study was offered as an elective workshop for the participants, without incentives (other than participating in a learning workshop). The participants' interest about the workshop was to learn more about data and to improve data skills to apply in their design practice. During the study, participants worked in groups (n=2-3). Prior to the workshop, the participants self-assessed their skills as following in Table 2 (for the assessment, see Data collection section).

Table 2 Overview of the study participants' skill self-assessment.

Programming skills (between 1-7, 7 highest)	**Data analysis skills** (between 1-7, 7 highest)	**Technical literacy** (between 1-7, 7 highest)
2.53 (SD: 1.80)	2.46 (SD: 1.05)	3.46 (SD: 2.18)

Apparatus: The participants were provided with a dataset, the Data Toolkit and suggested software tools to use. The dataset was a database of the participants' university's (Faculty of Industrial Design

Engineering, TU Delft) internal repository for master theses at the time of the study, containing 2040 rows and 6 columns of metadata, including the theses' *Title, Abstract, Mentors, Keywords*, etc. The provided materials were the Data basics and Data techniques card decks, and the Questions for Data and Working with Data 101 booklets.

Procedure: The elective workshop was based on the earlier described Generative Data Exploration workshop methodology, facilitated by the first author. The elective workshop started with a basic introduction to using data in design and presenting a generic data workflow. After this, the participants were asked to form groups (n=2-3) and the groups received the dataset and the related design brief, and the card decks and the booklets.

- *Opening dataset and setting direction:* The initial activity during the study was to download and open the dataset and then to define at least three research questions to investigate with the data.
- *Data transformation:* The following activity was to immerse into the dataset, preferably by using OpenRefine as suggested software tool, and try to find answers for the research question. After providing some time for the participants to realise the problems with the data (such as cleaning is needed) and not knowing the various data transformations they could benefit from, a facilitator intervention happened, showing examples of powerful features of OpenRefine as well as RAWGraphs, the suggested visualisation tool.
- *Data exploration:* The following activity was to explore the dataset with OpenRefine and RAWGraphs for insights.
- *Communicating the insights:* For the end of the workshop, the groups needed to prepare a presentation out of their exploration process and the found insights, with the explicit task to make it visual (i.e., present visualisations). The presentations were audio-video recorded for further analysis.

The workshop ended with the participants filling up a reflection questionnaire and a Creativity Support Index (CSI) questionnaire (see Data collection session). Afterwards, an audio-recorded group discussion followed.

Data collection: Prior to the workshop, we asked the study participants to self-assess their related skills, using a Likert scale rating from 1 strongly disagree to 7 strongly agree (for results, see Table 2). The questions were as follows:

- My programming skills are great.
- My data analysis skills are great.
- I'm very technology literate.

For (research) data collection at the end of the workshop, we used online questionnaires and the Creativity Support Index (Cherry & Latulipe, 2014), a quantitative, psychometric tool to assess the workshop setup's assistance in creativity support in the design research process. Furthermore, observations were noted down throughout the workshop, and the presentations and the final reflective group discussion was audio-video recorded.

4 Results

Our observations of the participants' processes clearly showed that it is not straightforward for design students to start exploring a previously unknown dataset with the goal of concluding designerly insights. In general, the groups first defined some "interest directions" as research questions or data hypotheses, and then started with filtering and sorting the data. After seeing the struggles with the *Data transformation* activity, we intervened with a brief tutorial on tips and tricks with OpenRefine; it was important however, that first the participants realise what they don't know, instead of front-loading knowledge in the beginning as technical tutorials. After the initial confusion of how to use a new tool, they managed to "zoom in" on their interests in the dataset, with some groups going even further to deriving new data from the dataset (i.e., based on the raw data in the dataset, add additional data, such as counting the appearance of keywords). The participant groups commented that they needed to shift their thinking with transforming the data, indicating their general lack of practice with computational thinking. For *Data exploration*, the primary mean was visual inspection of the data, using RAWGraphs. The groups noted that RAWGraphs has many atypical charting options that they could use, but they lacked guidance on what charting works best for certain types of data to communicate.

4.1 Creativity support evaluation

The results from the CSI assessment indicates an average 73.85 (SD = 9.44) CSI score for our Generative Data Exploration method in this study (n=13).

Table 3 The CSI results from this study shows that participants rated Results Worth Effort and Exploration factors the most important, and the average weighted score for these two categories have been found highest.

Scale	**Avg. factor counts (SD)** (between 0-5, highest 5)	**Avg. factor score (SD)** (between 0-20, highest 20)	**Avg. weighted factor score (SD)** (between 0-100, highest 100)
Results Worth Effort	3.00 (1.78)	16.15 (1.47)	48.85 (30.92)
Exploration	3.85 (1.07)	14.62 (1.29)	55.85 (16.63)
Collaboration	2.08 (1.44)	14.15 (1.92)	28.46 (23.42)
Immersion	1.77 (1.42)	14.00 (2.38)	28.92 (28.15)
Expressiveness	2.31 (1.25)	13.54 (1.66)	30.46 (15.51)
Enjoyment	1.92 (1.44)	15.00 (1.27)	29.00 (21.94)

Table 3 presents the outcomes of the CSI survey. Following the example by Cherry and Latulipe (2014), we report the results with respect to average factor counts, factor score and weighted factor score. *Average factor counts* indicates the number of times participants chose a given factor important (between 0 and 5). *Average factor score* indicates how well the Generative Data Exploration method scored (between 0 and 20) for certain factors. The high rankings of *Exploration* and *Results Worth Effort* indicate that participants found these factors most important. The *average weighted factor scores* are most sensitive to factors that are marked more important, and in both Exploration and Results Worth Effort the Data Exploration workshop scored higher than the other factors. The outcomes of the CSI analysis confirm our design direction that exploration and generating meaningful outcomes that are worth the effort are of importance, and the method's direction is validated, however with room for improvement for future iterations.

5 Discussion and further work

We see the main contributions of our Generative Data Exploration method in empowering designers to discover meaningful insights from datasets, and to find inspiration that complements qualitative contextual research and informs the following steps in the design process, such as ideation and prototyping. Expanding the framework by Sanders and Stappers (2014), we place the Generative Data Exploration method primarily in the generative phase of design (see Figure 5).

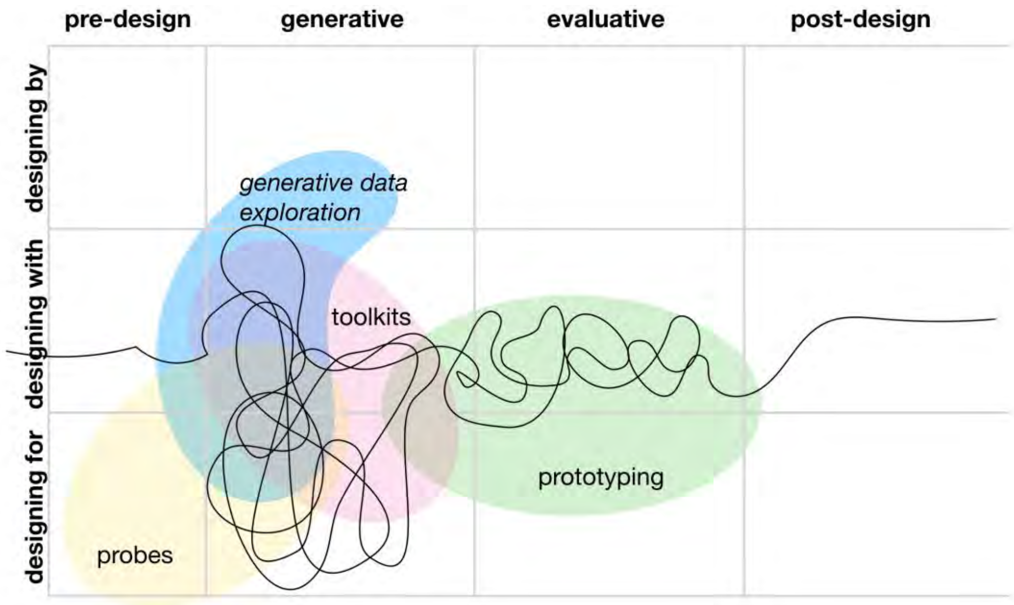

Figure 6. The Generative Data Exploration method placed in the co-design process (based on Sanders and Stappers (2014)

5.1 The value of digital data for generative design research

Our Generative Data Exploration method supported the participants in approaching and utilising already existing digital data in the fuzzy front-end. By using the method, designers managed to conclude designerly insights from data, which could be used as complimentary to traditional user and contextual research methods, such as contextmapping (Visser, Stappers, van der Lugt and Sanders, 2005) or design probes (Mattelmäki, 2006). Other researchers have explored complementing qualitative research in the fuzzy front-end with data collection by sensors: emerging examples, such as using everyday objects as data-collecting ethnographers to collect rich contextual insights (Giaccardi, Cila, Speed and Caldwell, 2016), or the use of data-collecting technology probes to augment rich and contextual data with sensor data (Bogers, Frens, van Kollenburg, Deckers and Hummels, 2016) have shown alternative paths to bring data techniques into the fuzzy front-end. However, both of these examples are technologically complex, often beyond the scope and resources available for a design team. Furthermore, these examples do not address how to utilise already existing data, for which we see the Generative Data Exploration method's main contribution. Exploring digital data, and used in complementary to the traditional qualitative methods, can provide *scale* – through making sense of large datasets –, and access to a digital footprint of human activity – such as networked interactions on a social network, that would not be accessible easily via qualitative methods.

5.2 Empowering designers with digital data for generative design research

The primary motivation for introducing our Generative Data Exploration method was to support the study participants to feel confident about utilising data in their design process. The study results indicate that the method and the contained tools indeed helped design students to make sense of a dataset and enabled them to successfully manipulate the dataset to extract insights. While doing, they gained confidence and familiarity with the basic mindset necessary to work with data. Data requires a specific skill set to be able to effectively transform and utilise it, and these skills are rarely included in design education (but common, though possibly addressed indirectly, in software engineering, business analysis, and similar). Our Generative Data Exploration method can be a valuable point of departure from the traditional design tools used to guide thinking in the design process, given the designers that use it are willing to approach research problems with a different mindset and by practicing different skills. A core of this is computational thinking, a skill that most tech-savvy designers possess, who have experience in programming. Computational thinking as a skill might not be practiced in design, however it is essential knowledge for data. In our experience, using non-programmable software tools in the beginning, the computational thinking problems for different data manipulations are not complicated, and designers with basic programming knowledge can get far enough to remain engaged. Thus, considering today's designers, a large set of people would be able to gain sufficient data skills (and learning about appropriate data tools) and contribute to the democratization of data, as far as guidance is provided how to do so.

5.3 Profiling the future types of data designers

Integrating digital data in the design practice will happen more and more, and this will transform how we do design (at least the design of interactive artefacts). We see that design does not only happen by expert professionals, but people applying design techniques and a designerly mindset for problem solving on a variety of problems in the world. Manzini (2015) describes this phenomenon as *expert design* and *diffuse design*, where diffuse design happens by people not trained in design, using their natural capacity for creativity and designerly thinking. Similarly, the best practices, know-how, tools, methods and so forth for data are a growing field as data science (Cao, 2017), but it is unlikely that data will remain a field that is limited to experts only. We hypothesise, that in the future, there will be designers that gain average-to-high level of expertise in data (that may exist today already with a niche expertise in data visualisation and similar), and there will be data experts that develop average-to-high level of expertise in design. These new intersections of the data and design will set the scene for new types of data tools for design, new types of design tools for data, and new types of designer and data expert profiles.

5.4 Limitations

A main limitation of this study is that our Generative Data Exploration method has been tested only through the study, in a facilitated workshop format, and thus not by independent designers. Furthermore, the target group of the Generative Data Exploration method is designers of all level of expertise, yet the study participants were master design students. Master-level design students are quite tech-savvy (and thus rather data literate already), which might not be representative for the whole design profession. It is also important to note, that the study's design brief and the provided dataset (metadata of library records) set up a limited problem space with its own properties, which does not model all sorts of potential design problems. With these caveats, it is difficult to assess whether the Generative Data Exploration is applicable in design research practice outside academia and in non-learning settings.

6 Conclusions

It can be concluded that our Generative Data Exploration empowers designers to utilise digital data in the fuzzy front-end. We developed two sets of card decks and booklets and a workshop methodology providing step-by-step guidance to utilise an existing dataset in the fuzzy front-end and to seek inspiring insights out of digital data. The design toolkit is tailorable and extendible for different datasets and different design situations. During the current study, the method has been proven useful in exploring data and in generating outcomes that are valuable for the design process. Furthermore, the method contributed to participants gaining confidence in utilising data in their design practice, mainly due to providing clear guidance while navigating through the workflow of data.

Future work points at various directions. We aim to conduct studies with design research practitioners as well to ensure the validity of the current approach and to explore how do higher level of design expertise influence the outcomes. As a method designed to be extendible and tailored for different design situations, the method and the encapsulated tools could continuously develop if designers keep using the method. Understanding how data can be used creatively, such as what kind of mechanics lead to inspirational insights from data is still in its infancy. To better understand this, further studies are necessary based on research on creativity and sense-making. Furthermore, in this study we explored how designers incorporate data techniques, but how data scientists incorporate design techniques (and follow a design process) could lead to an additional perspective on combining designerly and data thinking.

Acknowledgements: The presented research is part of the Open4Citizens project, which is funded by the European Commission under grant agreement H2020-ICT-2015-687818.

The authors would like to express their appreciations for the members of the Open4Citizens consortium and especially Tomasz Jaskiewicz and Mercedes Leipoldt for the fruitful discussions leading up to the presented work, to the students participating in the study, and Ianus Keller, Hayagreev Narayan and the Teaching Lab at Delft University of Technology for the organizational support.

7 References

Baumer, B. (2015). A Data Science Course for Undergraduates: Thinking With Data. *The American Statistician*, *69*(4), 334–342. http://doi.org/10.1080/00031305.2015.1081105

Bigelow, A., Drucker, S., Fisher, D., & Meyer, M. (2014). Reflections on How Designers Design with Data. *Proceedings of AVI '14* (pp. 17–24), New York, NY, USA: ACM. http://doi.org/10.1145/2598153.2598175

Bogers, S., Frens, J., van Kollenburg, J., Deckers, E., & Hummels, C. (2016). Connected Baby Bottle: A Design Case Study Towards a Framework for Data-Enabled Design. *Proceedings of DIS 2016* (pp. 301–311), New York, New York, USA: ACM. http://doi.org/10.1145/2901790.2901855

Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys (CSUR)*, *50*(3), 43–42. http://doi.org/10.1145/3076253

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

Cherry, E., & Latulipe, C. (2014). Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *21*(4), 21–25. http://doi.org/10.1145/2617588

D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, *23*(1), 6–18. http://doi.org/10.175/idj.23.1.03dig

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics*, *12*(3).

Deng, Y., Antle, A. N., & Neustaedter, C. (2014). Tango cards: a card-based design tool for informing the design of tangible learning games. *Proceedings of DIS '14* (pp. 695–704). ACM. http://doi.org/10.1145/2598510.2598601

Dove, G., & Jones, S. (2014). Using data to stimulate creative thinking in the design of new products and services. *Proceedings of DIS'14* (pp. 443–452). http://doi.org/10.1145/2598510.2598564

Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016). Thing Ethnography: Doing Design Research with Non-Humans. *Proceedings of DIS'16* (pp. 377–387). http://doi.org/10.1145/2901790.2901905

He, S., & Adar, E. (2016). VIZITCARDS: A Card-Based Toolkit for Infovis Design Education. *IEEE Transactions on Visualization and Computer Graphics*, *23*(11), 561– 570. http://doi.org/10.1109/TVCG.2016.2599338

Hill, B. M., Dailey, D., Guy, R. T., Lewis, B., Matsuzaki, M., & Morgan, J. T. (2017). Democratizing Data Science: The Community Data Science Workshops and Classes. In S. A. Matei, N. Jullien, & S. P. Goggins (Eds.), Big Data Factories: Collaborative Approaches (pp. 115–135). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59186-5_9

Hornecker, E. (2010). Creative idea exploration within the structure of a guiding framework: the card brainstorming game. *Proceedings of TEI '10* (pp. 101–108). ACM. http://doi.org/10.1145/1709886.1709905

Jones, J. C. (1970). Design methods. John Wiley & Sons.

Lycett, M. (2013). “Datafication”: making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386. http://doi.org/10.1057/ejis.2013.10

Manzini, E. (2015). Design, when everybody designs. Cambridge, Massachusetts: The MIT Press.

Mattelmäki, T. (2006). Design probes. Aalto University.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. *Proceedings of CHItaly’17*. ACM. http://doi.org/10.1145/3125571.3125585

OpenRefine. (2017). OpenRefine. Retrieved November 14, 2017, from http://openrefine.org/

Resnick, M., Myers, B., Nakakoji, K., & Shneiderman, B. (2005). Design principles for tools to support creative thinking. *NSF Workshop Report on Creativity Support Tools (pp.* 25–35).

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-Design*, *4*(1), 5–18. http://doi.org/10.1080/15710880701875068

Sanders, E. B.-N., & Stappers, P. J. (2014). Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign*, *10*(1), 5–14. http://doi.org/10.1080/15710882.2014.888183

Sanders, E. B.-N., Brandt, E., & Binder, T. (2010). A framework for organizing the tools and techniques of participatory design. *Proceedings of PDC '10* (pp. 195–198). ACM. http://doi.org/10.1145/1900441.1900476

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, *50*(12), 20–32. http://doi.org/10.1145/1323688.1323689

Speed, C., & Oberlander, J. (2016). Designing from, with and by Data: Introducing the ablative framework. *DRS 2016 Conference Proceedings* (pp. 2991–3003). http://doi.org/10.21606/drs.2016.433

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, *33*(1), 1–67.

Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley.

Visser, F. S., Stappers, P. J., van der Lugt, R., & Sanders, E. B.-N. (2005). Contextmapping: experiences from practice. *CoDesign*, *1*(2), 119–149. http://doi.org/10.1080/15710880500135987

About the Authors:

Peter Kun is a PhD candidate at Faculty of Industrial Design Engineering, Delft University of Technology. In his research, he explores the appropriation of data techniques and the use of digital data in the research phase of the design process.

Ingrid Mulder is an Associate Professor of Design Techniques, Faculty of Industrial Design Engineering, Delft University of Technology. Her expertise is in transformative and social design, using strategic design and co-creation to empower people driving social change.

Gerd Kortuem is Professor of Internet of Things, Faculty of Industrial Design Engineering, Delft University of Technology. His research focuses on the Internet of Things as creative design material and explores connected products and services for a sustainable future.