MSc Interaction Technology
Final Project

# Thematic Map Making using Declarative Grammars

Ilse Schieven

Supervisor: Barend Köbben
Supervisor: Shenghui Wang
Examination committee chair: Dennis Reidsma

August, 2024

Human Media Technology (HMI)
Faculty of Electrical Engineering,
Mathematics and Computer Science.

Spatio-Temporal Analytics, Maps and Processing (STAMP)
Faculty ITC.

University of Twente

UNIVERSITY OF TWENTE.

# Contents

**Abstract**

This thesis explores the use of the high-level declarative visualisation language Vega-Lite to develop a thematic map-making tool guided by a specialised grammar. The grammar, formulated around user questions, ensures smooth integration between user input, Vega-Lite, and the resulting thematic maps. Using this grammar and Vega-Lite, a tool was created to generate thematic maps. A qualitative comparative usability test assessed the tool's suitability for thematic map making. Although there is room for improvement, the tool shows significant potential to become a user-friendly platform, enabling users to ask questions about their data sets and receive an answer in the form of a data visualisation.

# Acronyms

**CBS** Centraal Bureau Statistiek (Central Bureau of Statistics). 9, 13, 14, 17, 24, 39, 40, 92

**CSV** Comma Separated Value. 24

**GDI** Geo-Data Infrastructure. 13

**GDPR** General Data Protection Regulations. 14

**GIS** Geographic Information Systems. 3

**HTML** HyperText Markup Language. 11, 24, 38, 43, 44

**JSON** JavaScript Object Notation. 11, 24, 25

**PDOK** Publieke Dienstverlening Op de Kaart. 13, 17, 24, 25

**SDI** Spatial Data Infrastructure. 13

**UI** User Interface. 38, 39, 43, 44

# Chapter 1

# Introduction

## 1.1 Motivation and Problem Statement

Thematic maps, which visually illustrate statistical data on maps [18, 41], serve as essential media for conveying data linked to particular locations, often intertwined with socioeconomic themes, such as population density or level of education per inhabitant. These maps play a major role in rendering otherwise difficult-to-read data so that the data becomes accessible and comprehensible [44]. Cartographers utilise symbols and graphical encodings in both static and interactive maps to effectively communicate information, contributing to a richer understanding of complex data sets [3, 26]. The digital revolution has significantly transformed the landscape of map-making, providing cartographers with improved accessibility and ease of use regarding the making of maps [18]. The addition of high-level or declarative programming languages has emerged as a notable development in this domain and could potentially be very suitable for map making [24]. Unlike traditional programming languages, which require explicit instructions on how to accomplish a task, high-level languages operate at a more abstract level. In the context of data visualization on maps, this abstraction can be particularly advantageous. Declarative programming languages enable users to describe what they want to achieve, leaving the implementation details to the language itself [37]. This not only makes the process faster, since one does not have to study the programming language, but also makes it more intuitive for individuals who may not have advanced programming skills. The rules for the representation of information on maps are important when creating a formal grammar for data visualisation. These rules, established before the digitisation of map making [3, 25], guide cartographers through the steps of map creation. However, the advent of digital mapping platforms has introduced a need for the formalization of these rules [2, 35, 41]. Formalisation involves codifying the rules into a specific and consistent set, which essentially creates a grammar for map making [46].

This research explores the integration of formal grammar into high-level programming languages, such as Vega-Lite [34, 37]. The aim is to analyse whether these languages can provide a more seamless and standardised approach to data visualisation on maps. By formalising the rules governing graphical encoding, the process becomes not only more structured but also adaptable to various data sets and thematic representations. While map digitisation has opened the door to incorporating more data and accelerating map creation, it has also introduced challenges. The need for knowledge about data processing becomes crucial. The research into the suitability of high-level programming languages aims to lower the entry barrier eventually. The prospect is that creating a digital map may not require advanced programming skills but rather the availability of a relevant data set, since the understanding of the formalised rules and grammars are embedded in the chosen programming language. As the digital landscape continues to evolve, the integration of high-level programming languages and formal grammars into the field of cartography could make cartography more accessible to a broader audience. This, in turn, may potentially facilitate innovative and diverse representations of spatial data.

This thesis paper is a follow-up to the research topic paper called *Declarative Grammars for Visualisation Pipelines for Socio-Economic Data Visualisation*. That research topic paper has set the stage for subsequent experiments and in-depth investigations in the later stages of the research. As existing gaps are identified, an informed approach in the experiments will follow, contributing to the intersection of declarative programming languages, formal grammar, and thematic map making.

#### Current making of maps

Generally, there are three essential steps in creating a map: choosing data, fitting the data to constraints, and encoding the data using graphic parameters [10]. This applies both to topographic and to thematic maps. In recent years, there has been a move away from using a limited number of tools, such as Adobe Illustrator or ArcMap GIS software, towards using a combination of tools to produce a map [35], since the number of tools available is increasing. On top of that, the tools available come from different perspectives towards map making as well, for example some tools tend to have a software engineering rather than a cartography background [35]. This shift makes it more complex for new cartographers to get skilled with all tools involved in map making, and there is a lack of rules and therefore consistency between different ways of map making. On the other hand, it makes map making flexible, and a cartographer can create a map any way they like, within the rules and constraints given in symbolisation [3, 35]. Digital maps depend on Geographic Information Systems (GIS) data to create the maps. GIS enables the creation of maps from data and the description of space [16]. Reading GIS data, for example in a table, can be challenging. A graphic or visualisation makes it easier to read GIS data [42].

## 1.2 Research questions

This research focuses on the creation and evaluation of a grammar for data visualisation and a corresponding thematic map making tool which is made using a declarative language. The literature research in an earlier stage of the thesis resulted in the following research questions:

**RQ1** How can the implementation of a visualisation grammar into a high-level declarative programming language contribute to the effective creation of thematic maps from data?

- **RQ1a** How can a visualisation grammar facilitate the generation of effective thematic maps from given data and data specifications?
- **RQ1b** In what manner can such a grammar be structured for implementation within a high-level declarative programming language, such as Vega-Lite?

**RQ2** How can the user input be optimised to show the most effective map for the user's needs?

- **RQ2a** Which of the specifications should have degrees of freedom such that a user can enable the creation and exploration of meaningful thematic maps?

**RQ2b** Can a thematic map-making tool yield the intended output for a map based on a dataset given by a user?

**RQ3** How can the User Interface be created, such that it enables the user to articulate their data inquiries in a declarative format within the tool effectively?

## 1.3 Structure

After this introducing chapter, chapter 2 starts with background information, to present the most relevant literature found during the research topic. It will also introduce related work, similar to the research in this paper, to see how others have tackled this topic, their ways of working, and the points in their discussions left open for exploration. The methodology introduces a plan based on chapter 2 that explains how the research questions will be answered. The chapter after that, approaches, will give a summary on how the methodology was executed. Chapter 4 gives an overview on how the final product has been created and what things have been taken into account. This chapter ends with sharing the results of the user studies performed on the final product. This chapter is followed by a discussion chapter, where will be reflected upon the meaning and relevance of the results. In the final chapter, chapter 6, the research questions will be answered and suggestions will be made for future work based on this thesis.

# Chapter 2

# Background

This background section is introduced with general background information on thematic map making, such as data and data visualisation, after which it will dive deeper into the use of formal grammar for thematic map making. The chapter then includes relevant tools and literature, and concludes with an ethical statement and the chapter conclusion.

## 2.1 Thematic map making

This section summarises key elements from literature on data visualisation and thematic maps. The aim is to get familiar with the guidelines for data visualisation. The findings are presented through images, tables, and text. The purpose of this section is to provide a summary of the rules that guide the visualisation of socio-economic data, with a particular emphasis on maps; called grammar. This section will be introduced with the most important concepts of data visualisation and will then elaborate on more specific ideas towards map making. Since the section is based on concepts from data visualisation, "data" will come across quite a few times during this chapter. This refers to the data that will be used during the final thesis. If not specified, this "data" can be generalised to any (spatial) data one wants to visualise.

### 2.1.1 Visual variables

Visual variables are variations and combinations such that they are best suited for the data. They can consists of form, orientation, colour (hue), texture, value, size, transparency, crispness, and resolution. Visual variables allow people looking at the visualised information to perceive association, order and quantities related to the original data [3]. Some visual variables give certain perception of what is presented. What this perception is per visual variable is presented in table 2.1.

Table 2.1 is derived from the information on Kartoweb [27] and the article from White [45]. White, in his article, adds the visual variables transparency, crispness, and resolution; variables that can indicate certainty on maps. These variables are useful for example when displaying weather predictions; more detail comes across as a higher certainty.

### 2.1.2 Visual grammars for thematic maps

Besides perception properties, there are also different map types, that can be divided in topographic maps (accurate representation of earth topography) and thematic maps (one or more particular themes are emphasized, always some topographic content). Thematic maps use data (i.e. socio-economic data) to show information, using the perception properties

| | Perceived as ... data | | |
|---|---|---|---|
| Visual variable | Associative | Ordered | Quantitative |
| Colour | ■ | | |
| Position | ■ | ■ | ■ |
| Shape | ■ | | |
| Orientation | ■ | | |
| Texture | ■ | ■ | |
| Value | | ■ | |
| Size | | ■ | ■ |
| Transparency | | ■ | |
| Crispness | | ■ | |
| Resolution | | ■ | |

TABLE 2.1: Perception properties [14, 27, 45]

and visual variables as presented in table 2.1 [27], whereas topographic maps have the goal to show the surface of the earth. The thesis will not dive deeper into topographic maps, since the focus is on data visualisation for thematic maps.

Visualising maps is a form of data visualisation. The goal is to transfer information, however the type of information can be different. The goal of data visualisation is showing the information, a map is a specific tool to visualise information. Any graphic should be clear enough, including a map, that it can be extracted from its original context and be incorporated into any other set of information or document, and still be understandable [3]. Bertin [3] differentiates between the amount of components on a map, being one - a topographic map - or two - showing qualitative, ordered, or quantitive data together with a geographic component (a thematic map). The amount of components shown on a visualisation is a guideline Bertin [3] uses in his "standard schemas". This is a tool on how to visualise one or multiple components in an image or graphic. The schemas are rather complex, being created for people familiar with Bertin's work to construct visualisations that efficiently convey information through substantiated design choices, therefore not it is not necessary to dive deep into the details of the schemas. Most importantly to take away from it in the case for this thesis, is how many components can be added to a map for it to be a comprehensive image: two or less. Besides the schemas, Bertin has established a reading process of any visualisation [3]. The emphasis of the reading process is on the question to be answered with the visualisation, meaning the goal of the visualisation. The reading process can be a tool that can be used when being a reader of a visualisation, as well as when being the creator of one. When creating a visualisation, the tool can be used backwards so that the visualisation ensures to answer the question a reader might ask the visualisation. Bertin's reading process is as follows:

1. Identify the invariant and components involved in the information
2. Recognize by what visual variables each of the components is represented in the graphic
3. Understand the information itself
    - Formulate a question (conscious or unconscious) to become informed. With other words: find the information one was looking for.
    - Within a question, someone can read on three levels; elementary (ask for one element in the figure), intermediate (ask for a part of the figure), or overall or

| Visual variable | Note |
|---|---|
| Colour | Can be perceived as ordered, contains symbolism and association, take into account colour blindness |
| Shape | Contains symbolism, variation in shape is only applicable for elementary reading, variation is infinite but hard to focus on too many |
| Orientation | Limit number of orientations to a maximum of four to be able to differentiate |
| Texture | Can be perceived as ordered, can have a vibratory effect, can disappear when texture is too thin or thick, limited length |
| Value | Is ordered, can only be quantitative in combination with size |
| Size | Variation is limited, human eye can only differentiate up to 20 steps in size |

TABLE 2.2: Notes per visual variable to consider during designing process

global (the whole figure). On top of that, there are as many types of questions as components, and consequently any question can be defined by its level and its type.

Bertin [3] notes that in order to respond efficiently to all the types of questions, which can be generated by information having more than two components in maps, we must construct two types of graphics. If this is not done, not all questions can be answered as efficiently. If one were to construct a graphic with more than three components, it can be hard to focus on one of the variables. It will take a lot of time and effort to answer a formulated question, and it is an error-prone question to answer. In his book, Bertin [3] highlights important factors that may cause confusion related to visual variables, or that should be taking into account when visualising data. These are summarised in table 2.2.

Different types of data and perception properties generate different visualisations depending on the type of information [14]. The goal is to present clear and concise information with a logical flow and connections between data and visualisation. This can be done with visual components that can be linked, constructed, or varied, through different variables. This approach is applicable to standard data visualisation and cartography [13, 14, 36].

Building on the work of Bertin [3] is the DNA visualisation series [13, 14, 36]. The series calls itself a framework for a visual grammar, simply put a set of rules for visualisation, and can therefore be used as a guide for a thematic map-making grammar. The series presents itself as a potential tool for computer-based visualisation advice, which is the reason it is introduced in this research topic. DNA in the DNA visualisation papers refers to Heer et al. [19], who introduces the idea that all visualisations share a similar "DNA"; a set of visual variables combined to produce a visualisation. This metaphor is again used by Engelhardt and Richards [14] who have created a set of rules based on visual components as building blocks. They even take Heer's metaphor one step further, by arguing that each visualisation belongs to a "visualisation species" that share common DNA. For example, one scatter plot would share DNA with another scatter plot.

The series also present the work of Wilkinson [46] and the tool Vega-Lite [37] as grammars that can describe visualisations [36]. Wilkinson's Grammar of Graphics [46] contains some complex visualisations, many of which are out of scope for this project, but it shows what a grammar could look like; therefore, it will be elaborated more in the Grammar section (section 2.3). Wilkinson attempts to formalise the graphics as described in figure 2.1. The goal now is to create a simpler grammar specifically for thematic map making, so from socio-economic data. Engelhardt and Richards' DNA for Data Visualisation gives

| Data type | Example |
|---|---|
| *Qualitative data* | Descriptive and textual |
| Nominal/associative data | Data of all the same order; states of the US, types of land use |
| Ordinal/ordered data | Data that can be ordered, but not in quantitative terms; dry-medium-wet, city-village-town |
| *Quantitative data* | Data that can be counted or expressed numerically |
| Interval data | Can be ranked, does not have an absolute zero; temperature, altitude |
| Ratio data | Relative (percentage or density) or absolute data (single class of features that can be measured); inhabitants or inhabitant density |

TABLE 2.3: Data types [27]

a good foundation for such a grammar. Combining the theory on data visualisation and grammar of graphics with suitable tools, could result in such a grammar, which will be explored in the thesis.

## 2.2 Data

The components of geospatial data are location, attribute, and time [26]. The last one can be a time-stamp or an interval of time. Location can be for example a region or a location expressed with geographic coordinates. Attribute is about what or who is represented, like land use or inhabitant information.

Attributes can be divided into qualitative and quantitative data [26, 27]. Qualitative data are descriptive and textual, such as types of land, road names, or river names. Within qualitative data we distinct nominal and ordinal data. Nominal, or associative, data are different in nature or identity, but all information is of the same order, like provinces in the Netherlands (Overijssel, Utrecht, Drenthe, etc.). Ordinal, or ordered, data can be ordered or ranked in terms that are not quantitative, like climate (dry, medium, wet) or settlements (city, village, town). Quantitative data are data that can be counted or expressed numerically. It is usually objective data produced through a verifiable systematic process which can be reproduced and is not subject to interpretation. Quantitative data can be split up in interval and ratio data. Interval data can be ranked, and the differences between the data does make sense, like temperature or altitude. Ratio data are quantitative based upon an absolute zero, whereas interval data are not; temperature or altitude can also have negatives. Ratio data divides in relative and absolute data. Relative data are values derived from raw data, generally an absolute quantity divided by some other quantity (percentage, rate, ratio or density) [27]. Absolute data quantities concern a single class of features, expressed in absolute terms according to some measurements scale, like number of inhabitants (it can be measured or counted). For a concise overview, see table 2.3.

To create a thematic map, all these data types should be taken into account, and be able to be represented. For example, qualitative data to represent provinces, council parties, or road names, and quantitative data for population (density), area of water, or the number of houses. The data used for a map from the Netherlands could come from

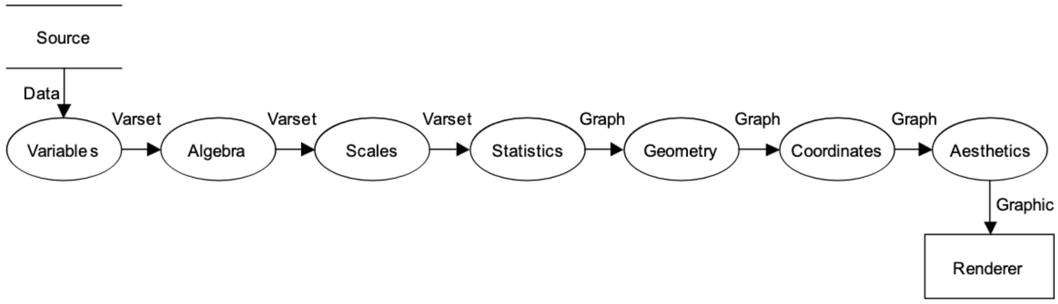


FIGURE 2.1: Wilkinson's process from data to graphic [46]

Centraal Bureau Statistiek (Central Bureau of Statistics) (CBS), since that is the official data source for publicly available data in the Netherlands [1]. Some datasets from CBS, the ones that are linked to location data, have all the components of geospatial data.

## 2.3 Grammar

In data visualization, a formal grammar, defined by Chomsky as a system of rules for lawful language statements [9], sets objective standards. These standards, a blend of technicalities on "how it works" and aesthetic considerations on "how it looks," enable the construction of comprehensible graphics [46]. Someone introducing a formal grammar for data visualisation is Wilkinson [46]. Wilkinson's book [46] focuses on rules for constructing graphs mathematically, from a statisticians perspective [12], and then representing them as graphics aesthetically. His image "data to graphic" is shown in figure 2.1, which broadly explains the process from data to graphic, with all steps in between. The goal of a grammar is to provide a reliable framework for presenting information in a particular manner, ensuring that the process as well as the outcome are both consistent. A formal specification is defined by van Lamsweerde [29] as the expression of a collection of properties some system should satisfy, in some formal language and at some level of abstraction. Formal according to [29] consists of three components: syntax, semantics, and rules for deriving useful information from the specification, called proof theory. The main advantages of using a formal grammar in data visualisation lie in facilitating component re-usability via specification matching and effectively verifying specification consistency and completeness.

A very basic example on visualisation grammars is given in figure 2.2 and figure 2.3. In these two figures can be compared how scatterplots are built up from the Visualisation DNA grammar [14, 36] and Wilkinson's Graphics Production Language [46]. Both rely on elements that explain how the different components of a visualisation are built up, rather than rules for a specific type of data.

## 2.4 Tools

There are several data visualisation tools, many of which are suitable for map making. Well known ones are for example ArcGIS, Leaflet, Tableau, and various programming languages. The website of Keshif [22] collects data visualisation tools which allows comparison between tools. For example, it allows a user to select on tools that are web-based. One of the tools that presents itself with that filter is the declarative grammar Vega-Lite. The first part of the following subsection will delve deeper into Vega-Lite, while the second part introduces

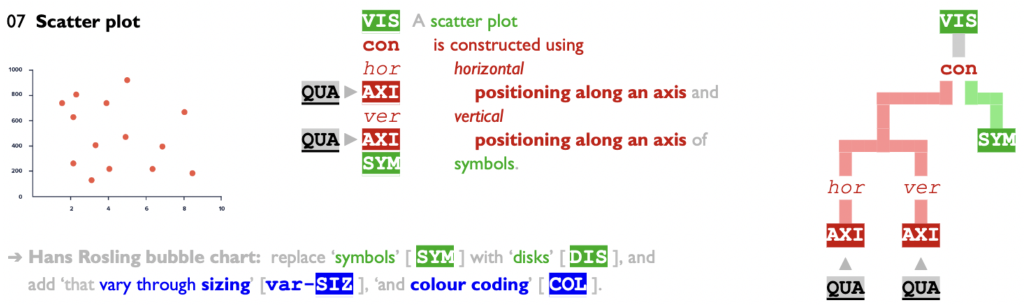


FIGURE 2.2: Scatter plot made up from Visualisation DNA [36]

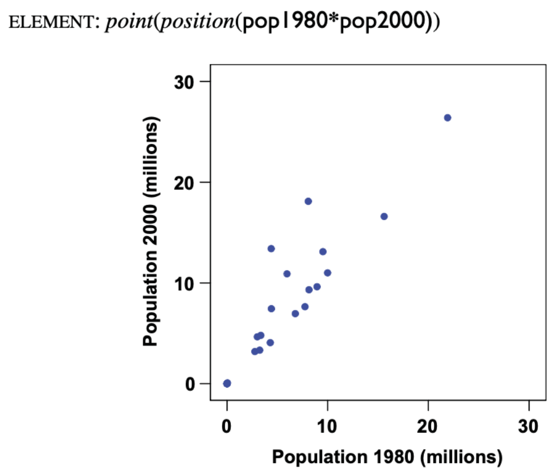


FIGURE 2.3: How to create a scatterplot with Wilkinson's Graphics Production Language [46]

the national atlas [24]. The national atlas is a digital and proof-of-concept tool that allows online thematic map making based on data sets.

### 2.4.1 Vega and Vega-Lite

The focus of this research will be on a tool that allows creating a grammar for thematic map making, and Vega-Lite could potentially be very suitable for that goal, hence the introduction for it here. Vega-Lite is a JavaScript Object Notation (JSON) syntax, based on Vega [38] and D3 [7] designed to create interactive graphics for data analysis and presentation. It promotes itself as a high-level grammar and declarative language. High-level meaning that it is designed to be a user-friendly programming language, and grammar refers to the rules the language is made up from [36, 37]. In order to understand if and why Vega-Lite is suitable, it is useful to understand the differences between imperative and declarative programming. Imperative programming is when one reaches the goal by stating the steps that explain how the program should run in order to reach the goal. Declarative programming is when you describe the goal and the program itself decides how to reach it. Languages such as Python and JavaScript support both, depending on how the user codes and what additional tools are used. Languages such as Haskell and HyperText Markup Language (HTML) are known to be declarative, while C and Java are imperative. In the case of data visualisation, a declarative way of visualisation would be "I want the population density per province of the Netherlands", while programming with an imperative language would look more like a step-by-step guide, first creating the list of provinces, followed by coupling a density to each province, after which one can state how the provinces are located et cetera. As Vega-Lite sates on their website, "specifications describe visualisations as encoding mappings from data to properties of graphical marks. [...] It determines default properties of these components based on a set of carefully designed rules" [34]. It is also mentioned however that these rules can be overruled by the user, in case the user does not agree with the set rules. This allows one to adapt Vega-Lite towards a suitable grammar for the specific application of the user [37].

Besides declarative and imperative, a language can also be formal, where formal generally means that there is a set of rules that make up the language. This is the case in linguistics, but also applies to programming languages like Vega-Lite. As described by [21], a grammar like Vega-Lite is able to describe the visualisation in its entirety, from the graphical marks to its interactions, rather than having to individually describe each part with another tool. Grammar-based visualisations have recently gained popularity, but tool support is lagging behind, since not all visualisations tools can read each other; Vega-Lite cannot be read in Jupyter Notebook without a module for example. Also exporting a visualisation to a declarative grammar is not easily possible yet [21].

After releasing Vega-Lite, it is later updated to support geographic visualisation [30], using the JavaScript library "Leaflet". This implementation is promising for thematic map making, however Lin and Paramasivan [30] also recognise the lack of interaction it yet provides. Zooming, hovering, and click handlers do not work optimally at the time of publishing of their article, and no updating research article or framework could be found. It could be that this problem is solved, but if not, a solution towards interactivity should be found elsewhere. Vega does include this interactivity, so perhaps it could be combined in order for the final product to work as wished.

Since Vega-Lite is built on top of the programming language D3, all components from D3 can be used withing Vega-Lite as well. Vega-Lite, in comparison to D3, is declarative and operates at a higher level of abstraction. It works with "unit specifications", that are blocks of elements that Vega-Lite can read and process. Satya-

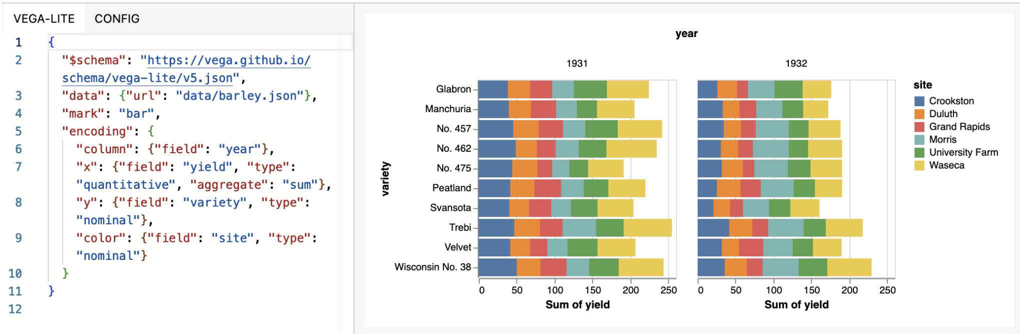


FIGURE 2.4: Example of the Vega-Lite syntax for horizontal bar charts [34]

narayan et al. [37] describe a unit as a single Cartesian plot. A Vega-Lite unit consists of the following: `unit = (data, transforms, mark-type, encodings)` [37]. The data refers to the inputted data(-set), so the data used as the input for visualisation. The transforms are any transforms performed over the data with formulas, such as the `lookup`-function, that allows connection between different datasets, or for example filtering the data. The mark-type, or mark, represents the geometric object used to visualise the data, such as a bar, area, or a geographic mark. Lastly, the encodings specify how data attributes are connected to the properties of visual marks, and is defined as `encoding = (channel, field, data-type, value, functions, scale, guide)`. The first parameter in the encoding-tuple is the (visual encoding) `channel`, which includes spatial position, colour, shape, size, and text. There are sub-types for channel, being order channel (determines the sequence of stacked elements/mark types), path order channel (determines the sequence in which points of a line or area mark are connected), and detail channel (includes group-by fields in aggregate plots). The second parameter of encoding is the `field` string, where one can declare the data attribute they want to visualise, with the `data-type` of that data attribute. `Value` can act a the data field as well, when the input is a constant literal. `Functions` can transform the input data with functions such as sorting or aggregation. `Scale` maps data to a visual range and a `guide` is an axis or legend that visualises the scale [37]. An example of the syntax can be seen in figure 2.4, which shows how the unit is used.

When a user would want to show multiple units using Vega-Lite, a composite view can be made using layer, concatenation, facet, or repeat. The signature for each of these are as following:

1. Layer:

$$layer([unit_1, unit_2, ...], resolve)$$

2. Concatenation:

$$hconcat([view_1, view_2, ...], resolve)$$

$$vconcat([view_1, view_2, ...], resolve)$$

3. Facet:

$$facet(channel, data, field, view, scale, axis, resolve)$$

4. Repeat:

$$repeat(channel, values, scale, axis, view, resolve)$$

Unit here refers to the unit as introduced before, view refers to any Vega-Lite specification, unit or composite. The units and how they can be set up specifically for visualisation, make the Vega-Lite language useful for this research and the creation of a thematic map tool later on in the research.

### 2.4.2 The National Atlas

The goal of the online national atlas of the Netherlands [32] is meant to make the National Geo-Data Infrastructure (GDI) accessible via user-friendly interactive cartographic interfaces. It is currently an experimental tool [24], but can still be used as an example. A National Atlas is defined as a set of maps that presents a nation, with geographically interested people as a target group [17, 40]. The most important aspect of National Atlases according to Köbben [24], is that the information presented is comparable, and comprehensive. The complete visualisation, a map, should be an improvement of the visualised data sets on their own. As stated above, there is a National Atlas of the Netherlands as well. The initiative was founded in 1929, and digitisation of the National Atlas has started around 2000 [23] as scanned documents of the paper version, which was hard to keep updated [24]. Since then, the digital National Atlas of the Netherlands has been updated to a more interactive platform, using Spatial Data Infrastructures (SDIs). What is missing in the National Atlas still, is the adaptability of the maps; such that a map can fit the specific needs for a user, which is exactly what this research focuses on. It suggests that this problem could potentially be solved *"using a formal, declarative language with degrees of freedom* [...] *that are influenced by user requirements or preferences"* [24]. The national atlas tool itself is not a formal grammar, but it does show what a formal grammar can result in. It shows different maps based on the information one selects, for example average income per inhabitant in 2013, shown by municipality. The GDI on itself is innovative, but not so user-friendly, whereas the printed national atlas cannot always be up-to-date, and has to be printed yearly to stay up-to-date. Also Noordhoff press, known from the Bos Atlas of the Netherlands - the most well known national atlas of the Netherlands - recognises this by partly digitising their student atlas [33]. The makers of the national atlas of the Netherlands emphasise the importance of having a flexible online map, since different visitors of the website would need different map types in order to receive the information they want. The example given by Köbben, Kraak, and Ormeling [28] is on population data; a student might want to have simple maps with which they can compare national developments over time, while a governance professional would want to know the exact growth rates on municipality level. Since the target group for the atlas of the Netherlands is so wide, the information on it should be complete and broad, and the visualisations should be flexible. The article also mentions that a digital Atlas of the Netherlands is useless without a user-friendly implementation of the system, making the most important requirements for the system flexibility and user-friendliness.

### 2.4.3 Cijfers op de Kaart

Another related tool is for example the website of CBS [15], however little documentation of this project can be found. Cijfers op de Kaart is a tool created by CBS an StatLine that visualises the public data from CBS [15]. It shows thematic maps, where the user can choose between different themes and on what level to show the themes. The tool lacks flexibility in use and it is not built on a grammar. What is further known, is that the maps are from Publieke Dienstverlening Op de Kaart (PDOK) and the data is retreived from StatLine. If this tool would be used to compare the final to against, the same dataset from

StatLine can be used.

## 2.5 Ethical considerations

It is important to acknowledge the ethical aspect about mapping data. There are vulnerabilities when mapping data, which have to be taken into account [20]. A map is a restricted tool, and it is necessary to be discerning of the veracity presented in a map. Good maps may contain inaccuracies, or "white lies" [31]. These inaccuracies may be simplifications or other choices made during the map design process [4]. On a map, it may be challenging to present the contexts and origins of used data and the mapping-process [20]. The data used by organisations like CBS are collected publicly and subsequently shared publicly by CBS, which makes them responsible for the data regarding General Data Protection Regulations (GDPR) at that moment and with their means of sharing the data [11]. That makes it easier as a cartographer to use public data rather than collecting it themselves. However, GDPR can be quite tricky as a processor of the data becomes responsible as well; meaning a cartographer also carries responsibility towards the transparency and fairness of the shared data. Therefore, when sharing a map, a cartographer should be transparent about where the data comes from, it is important to realise it is a communication tool [20].

## 2.6 Conclusion

In literature, there currently is a lack of formal grammars specifically for thematic map making. There are some articles that attempt towards it, for example the language Florence [35], but Florence is a new language, rather than the authors having tried to incorporate a formal grammar within an existing language. Also Tsorlini et al. [41] have designed a formalised tool for data visualisation for maps. These are good options if one is really quick in learning a new programming language, but for people less skilled in programming, it can be a hassle to learn a new language in order to create a map in a formalised manner. An option would be, in order to tackle this problem, to have a tool that automatically generates maps based on a formal grammar, with options to interact with the grammar. The bar should not be set high for this interaction with the grammar; a declarative and high-level language can be a good option for this, such as the Vega-Lite tool [37] presented in this chapter. This can be used to make flexible and user-friendly maps, taking the national atlas of the Netherlands [32] as an example.

# Chapter 3

# Methodology

The research consists of four phases: 1) literature research, 2) creation of grammar, 3) creation of the product, and 4) evaluation of the product. Phases 2 to 4 have their own methodology that is explained below. The first phase, literature research, is presented in the previous chapter. The findings of the literature research form the basis for this chapter. Section 2 describes how, with the findings of the literature, a grammar can be formed. Section 3 describes how the product will be formed based on the grammar, and the last section introduces how the product will be evaluated with usability tests. The combination of the four phases together will answer the research questions as introduced earlier, and which are visualised in figure 3.1.

## 3.1 Grammar

The first research phase after the literature research is the creation of the grammar. As introduced in section 2.3, a grammar is a set of rules or standards on how the data visualisation works and on how it looks, which combine to comprehensive graphics. The set of rules for this research are specifically tailored for a map making tool focused on data visualisation for thematic maps, and is based on relevant literature. The grammar is needed for this research, because it allows a more smooth connection between the data visualisations and the creation of the tool. It also ensures consistency between the different data visualisations and ensures the correctness and comprehensibility of them. The literature research has presented information regarding visual variables, visual grammar, data types, and tools. The summarising tables from the literature research, tables 2.3 and 2.1, have been used as the basis for the grammar, together with Bertin's reading process [3]. Bertin's reading process is about reading a graphic, but turning it around can create a 'writing process' for a graphic and has therefore been useful in creating the grammar. Turning the reading process around, and making it a writing process, gives the following:

1. Understand the information one is looking for by formulating (or in the case of a tool, receiving) a question on the elementary, intermediate, or global level.

2. Recognise by what visual variables each of the components should be represented in the graphic.

3. Visualise the invariant and components involved in the information.

The reverse-order reading process presented the variables needed to make the grammar. The visual variables that would be used in the grammar, the information in the form of

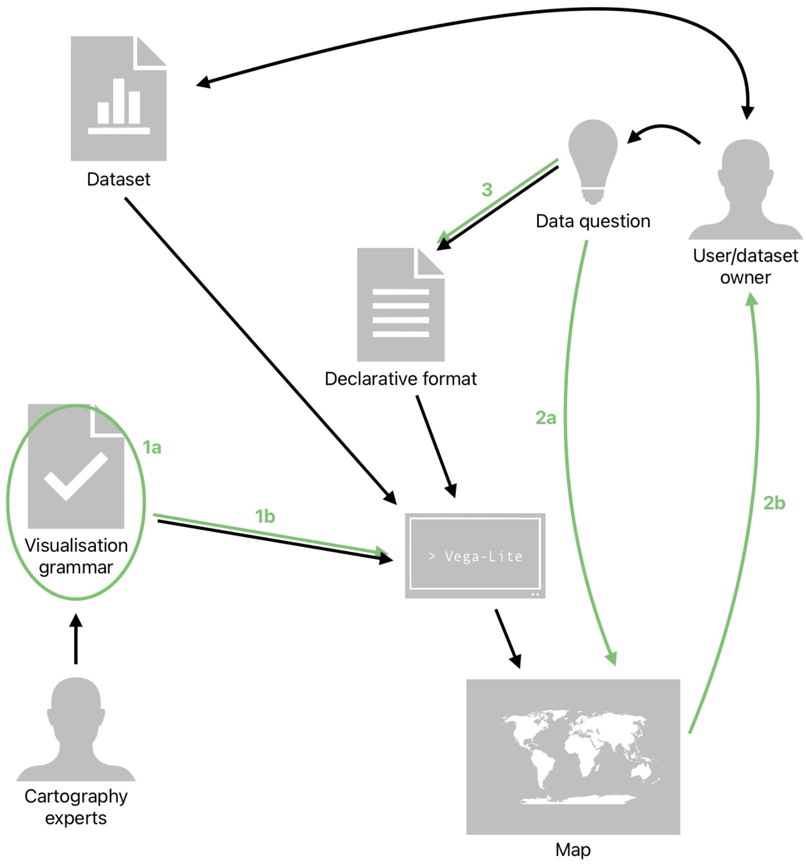


FIGURE 3.1: Visualisation pipeline of the final tool

data types, and the questions that could be received. On top of that, how the components, such as visual variables, are used in the graphics have been formalised in the grammar. The reading levels as presented by Bertin in the reading process have also been rewritten to make them more relevant for a 'writing process' for a graphic. These different variables have been combined to form a simple grammar, based on the literature available on making visualisation based on questions by Engelhardt and Richards [14]. The grammar created forms the basis for the product, which is created in the next phase.

## 3.2 Product

Based on the literature background and the grammar, it was decided that the product would be a proof-of-concept tool which makes thematic maps based on questions. The tool has been created using the high-level declarative programming language Vega-Lite, with public data sets from CBS for the socio-economic data, and with public data sets from PDOK for the geospatial data.

The grammar was implemented with the declarative programming language Vega-Lite, in such a way that the proof-of-concept tool could be tested by users, and compared with a similar tool, being Cijfers op de Kaart by CBS, such that it could be checked whether a declarative visualisation language such as Vega-Lite is suitable for thematic map making using a grammar. The workflow of the tool is visualised in figure 3.2. Because it to includes the grammar, it had to be a tool to which the users could ask a question based on data components, in order to make a data visualisation. Based on that, the tool had to return a thematic map while ensuring its correctness using the grammar. Since the proof-of-concept tool had to be tested against Cijfers op de Kaart, the number of datasets that could be used by the tool were limited to a subset of datasets used by Cijfers op de Kaart, to ensure comparability. In addition to that, the full-screen maps from Cijfers op de Kaart are limited to data sets shown on municipality level. A number of data-questions in the grammar could not be used in the proof-of-concept tool, also to ensure comparability.

## 3.3 Evaluation

The evaluation of the product was performed through usability testing. The goal of the evaluation was to get information on whether users find the product useful and if they think the product helps them to make informative visualisations that answer their data question. The evaluation was performed as qualitative comparative usability tests. The participants were in a private, quiet space with the researcher, which allowed the participant to use both Cijfers op de Kaart and the proof-of-concept tool, while thinking aloud. The think-aloud method as a usability tests makes sure that the participant can directly express their thoughts on a tool, while using it, but can feel as a bit unnatural. Still, it is a useful way of performing usability tests for both cartographic interfaces as well as user interfaces [43]. To ensure that the participants could also express themselves after the task, an interview was performed after using each of the tools. The interview consisted of a set number of questions, but the researcher could also ask clarifying questions based on the responses of the participants. The interviews were analysed to retrieve results.

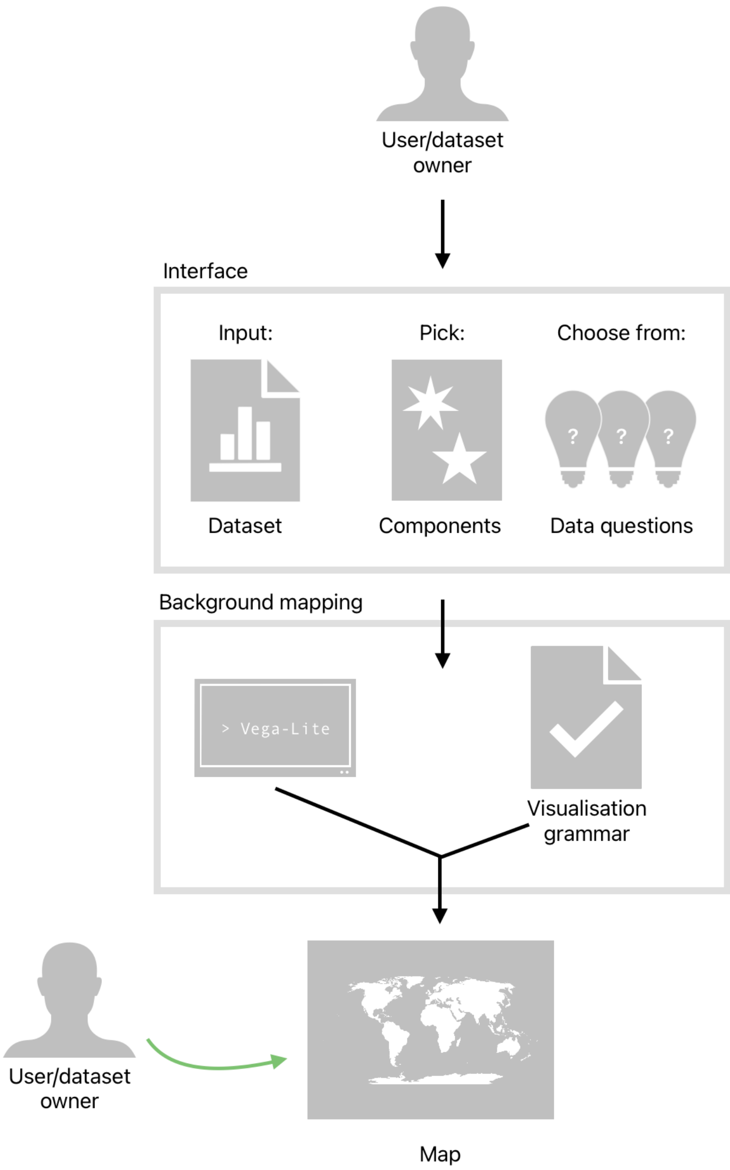


FIGURE 3.2: Workflow of the final tool

# Chapter 4

# Results

Similar to the methodology (chapter 3) this chapter is divided into the phases grammar, product, and evaluation. Each of the phases has its own section in this chapter, which describes the results obtained while performing the methodology.

## 4.1 Grammar

The grammar that has been created for this research has been based on visual variables, data components, and data-questions. As stated in the methodology, the reverse-order reading process from Bertin [3] has been used as a basis for the grammar, in combination with concluding tables from the literature research in section 2.3. These tables are repeated in this chapter as table 4.1 and table 4.2.

### 4.1.1 Writing process

The reverse-order reading process, as already introduced in the methodology, is as follows:

1. Understand the information one is looking for by formulating (or in the case of a tool, receiving) a question on the elementary, intermediate, or global level.

2. Recognise by what visual variables each of the components should be represented in the graphic.

3. Visualise the invariant and components involved in the information.

This reading process has formed the basis for the grammar created for this research. The reverse-order reading process has been assumed to be a correct-order writing process for visualisations. This means that the user should formulate a question, or be supported formulating a question, based on which the grammar "recognises" what visual variables should be represented in the graphic, and how the components should be visualised. This question, called "data-question", had to retrieve the components, which are the data (types), and the reading level. The reading level has been named the "creation level" in the case of the writing process and gives information about the number of components visualised, and how these are positioned. The number of components visualised together with a map has a maximum of two. If there are more components, it should be visualised in multiple figures [3]. A note about the creation level is that, for thematic maps, elementary level - when the user asks for one component - is excluded, since that would only return a map without any other data visualised.

| Data type | Example |
| --- | --- |
| *Qualitative data* | Descriptive and textual |
| Nominal/associative data | Data of all the same order; states of the US, types of land use |
| Ordinal/ordered data | Data that can be ordered, but not in quantitative terms; dry-medium-wet, city-village-town |
| *Quantitative data* | Data that can be counted or expressed numerically |
| Interval data | Can be ranked, does not have an absolute zero; temperature, altitude |
| Ratio data | Relative (percentage or density) or absolute data (single class of features that can be measured); inhabitants or inhabitant density |

TABLE 4.1: Data types [27], repetition of table 2.3

| | Perceived as ... data | | |
| --- | --- | --- | --- |
| Visual variable | Associative | Ordered | Quantitative |
| Colour | ■ | | |
| Position | ■ | ■ | ■ |
| Shape | ■ | | |
| Orientation | ■ | | |
| Texture | ■ | ■ | |
| Value | | ■ | |
| Size | | ■ | ■ |
| Transparency | | ■ | |
| Crispness | | ■ | |
| Resolution | | ■ | |

TABLE 4.2: Perception properties [14, 27, 45], repetition of table 2.1

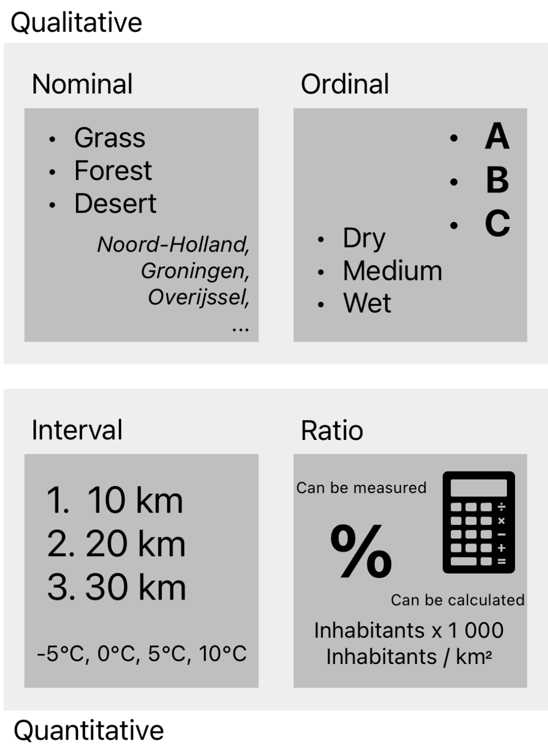


FIGURE 4.1: Types of data as in table 2.3

### 4.1.2 Grammar building blocks

Whereas the writing process formed the process of the grammar, there were visual building blocks needed in order to form a data visualisation grammar. The visual variables from table 4.1 are visualised in figure 4.1, and the data types from table 4.2 are visualised in figure 4.2. Since the visual variables are perceived as a certain data type, they have been used for that specific data type in the grammar. The last building block for the grammar has been figure 4.3. This figure has been derived from the fact that the grammar is based on questions, as can be concluded from the reverse-order reading process. It shows the questions that hold relevant information for data visualisation on maps, such that they encapsulate relevant information. The questions as in the figure do not stand on their own, but should also include (the name of) a data set. However, the question on its own does provide enough information for the grammar. The addition of the data set(s) is necessary for visualising the correct information.

### 4.1.3 Conclusion

In figure 4.4 all the building blocks of the grammar are combined into one figure, which visualised how the questions provide enough information for the visualisation. The figure shows the relationship between the components. In this figure, the visual variables are connected to the data types, and the entire figure is connected with questions from figure 4.3. Not all questions are visualised in figure 4.4, because not all questions were relevant to the final product. As stated in the methodology, the final product as introduced in section 4.2 will be compared to Cijfers op de Kaart. To ensure comparability, not all the questions as in figure 4.3 were relevant.

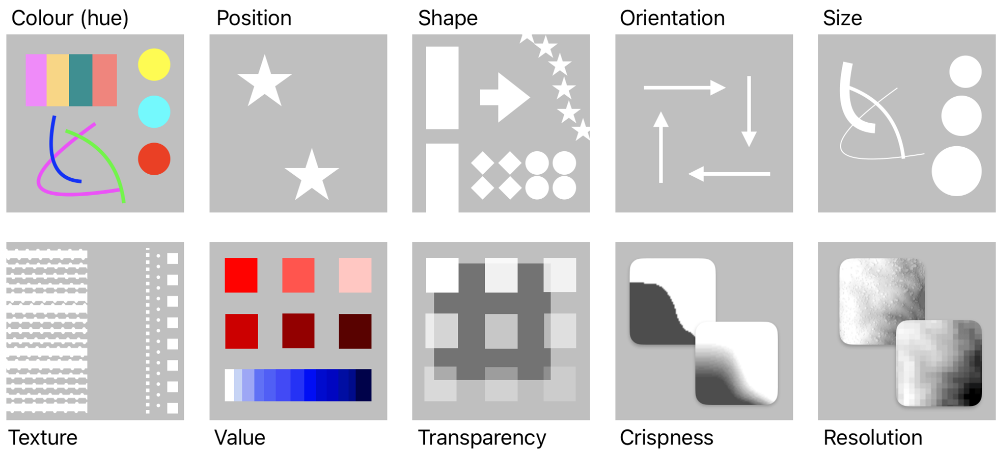


FIGURE 4.2: The visual variables

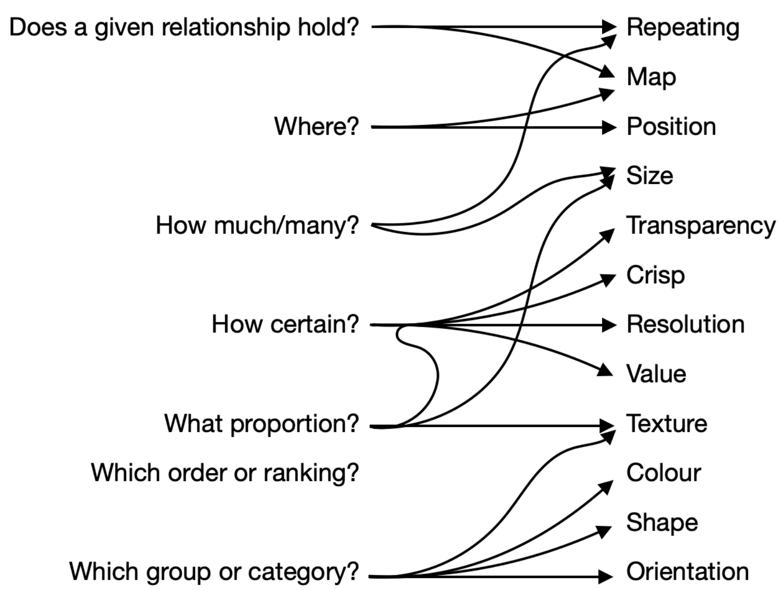


FIGURE 4.3: Questions and the connected visualisations

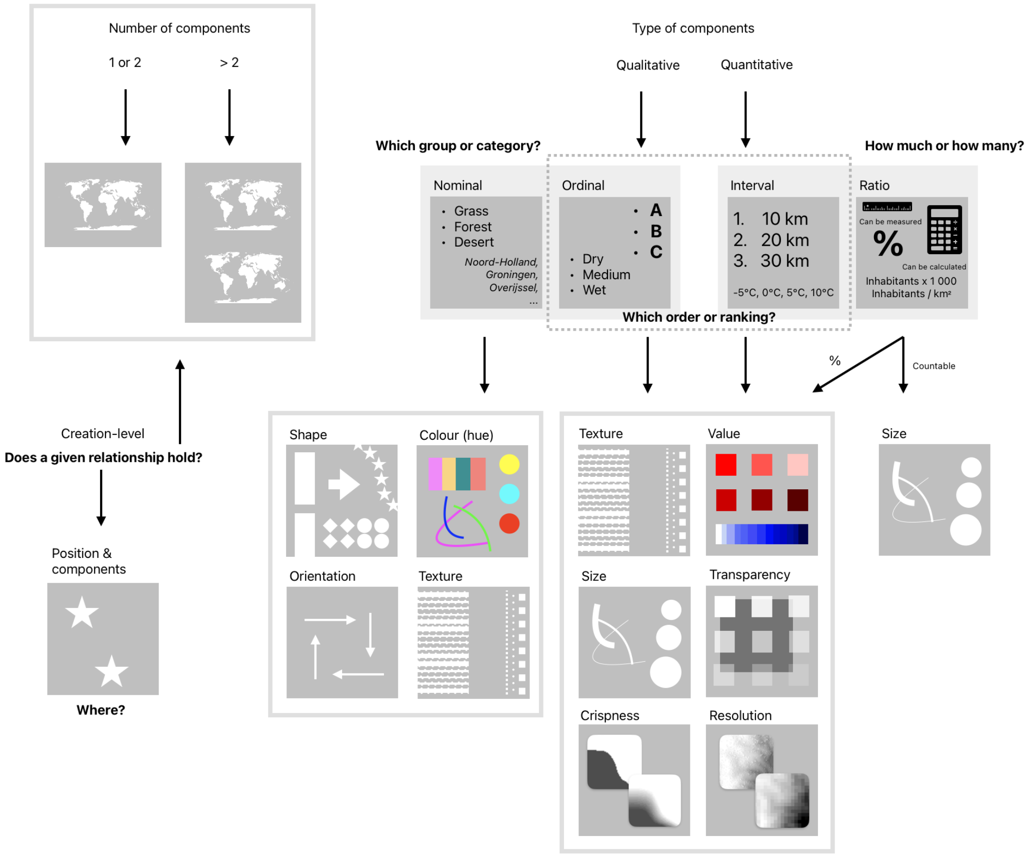


FIGURE 4.4: Basic grammar diagram based on question

## 4.2 Product

Based on the grammar and using Vega-Lite, the product was made. The product was made considering that it would be compared to Cijfers op de Kaart in a later phase. This limited which thematic data sets were used for the visualisations and which questions were selected from the grammar, since the same question had to be answered with Cijfers op de Kaart as well. This meant that the available thematic data sets in the product had to be a subset of the ones available within Cijfers op de Kaart. Regarding the questions, questions as "How certain?", "Which group or category?", and "Which order or ranking?" fell through because these could not be answered within Cijfers op de Kaart. In addition to that, there was also a limitation on which areas of the Netherlands to show in the tool. For example, making visualisation on province-level is not an option in Cijfers op de Kaart, and it cannot show district or neighbourhood on the full map, leaving municipality-level visualisations the only option for fair, full-map, comparison. Lastly, Cijfers op de Kaart is solely available in Dutch, making it a straightforward choice to make the product Dutch as well.

### 4.2.1 Data

Data sets from CBS have been used for thematic data, as these are also used by Cijfers op de Kaart, the tool against which this product was tested in the evaluation phase. The data sets, which were retreived as Comma Separated Value (CSV)-files from the CBS website, were formatted in such a way that they could be read by Vega-Lite. For geospatial data, a geo-JSON data set with municipal data of the Netherlands from PDOK was used, which is also the same as used by Cijfers op de Kaart. Both were connected within Vega-Lite, using the municipality code that was used by both data sets, such that each area displays the corresponding data. Finally, there was another geospatial data set from PDOK used that included the spatial coordinates of the middle of each municipality, which was useful for some of the data-visualisations.

### 4.2.2 Making the product

The product was programmed using HTML for the framework and Vega-Lite for the data visualisations. The framework in this case consisted of a title and an introduction followed by a dropdown menu where the user can input their data question. The brief introduction is as follows:

> This thematic map tool makes visualisation based on data questions. The visualisations can only be made on municipality-level. The datasets are from the public source "Central Bureau of Statistics" (CBS) and are as recent as possible (from 2022). You can insert your own data question below:

The introduction first included some examples of data questions, which were removed prior to executing the usability tests, so that the introduction was more compact and there was more space for the data visualisation on the screen. This resulted in the main page looking like figure 4.5, when the dropdown menu has been clicked. When one of the questions in the dropdown menu is selected, another dropdown menu appears with the thematic data to select from, which is different depending on the question selected. When a full question is selected, the visualisation will appear, which Vega-Lite creates based on the selected question and thematic data, using the mark `geoshape`, which in this case represents the map of the Netherlands with its municipalities, and for some of the questions


FIGURE 4.5: Product: Questions menu

the mark `circle` is layered upon the `geoshape`. The map of the Netherlands is based on the same PDOK geo-JSON file every time, ensuring that the map itself is consistent, both of which are processed by Vega-Lite. The full code of the product is given in Appendix C.

### 4.2.3 Using the product

The product is made on a blank page, with a title ("Thematische kaarten tool", which translates to "Thematic map tool"), an introduction, and a drop down menu. Considering comparability with Cijfers op de Kaart, four questions ended up in the proof-of-concept tool (see figure 4.5);

1. Where is the highest or lowest...?
2. What is the relationship between...?
3. How much/many...?
4. What is the percentage of...?

After clicking any of the questions, a new drowdown menu appears that shows the datasets the user can choose from, in order to finish the question. After selecting a data set, the user is shown the corresponding map.

#### "Where?" and "How much/many?"

Out of the four questions, two visualisations lead to the same visualisation of the data. The first question is derived from the "Where" question, as in figure 4.3. The question had to be altered in order to fit the functionalities of the proof-of-concept tool in combination with the datasets available. For example, a functionality of Cijfers op de Kaart is the fact that you can search for a location, such as the municipality Enschede, or villages or neighbourhoods within municipalities. The option to search for a place was not within the scope of the proof-of-concept tool as of now, so those kinds of "Where?"-questions cannot be asked to the tool. Another option could be questions similar to "Where are churches located?" or "Where are museums?" followed by a map with symbols to show location. The symbols could possibly be coloured-categorised, which could also answer the "Which group or category?"-question as shown in figure 4.3. However, these kinds of questions are not compliant with CBS datasets, nor the visualisations of Cijfers op de Kaart and did therefore not have the option to be compared during usability tests if they were to be included. Therefore, in the end, the "Where?" question was turned into "Where is the highest or lowest?". This still answered a question that results in map and position visualisations, and is similar to the "How much/many?" question, which results in the same visualisation; namely a map with a sized circle (size visualisation) in each municipality (location visualisation) on the map of the Netherlands (map visualisation). This means that two different questions lead to the same visualisation in the product.

**"Does a given relationship hold?"**

The second question which a user can choose from in the dropdown menu is "What is the relationship between?", which is derived from "Does a given relationship hold?" as in figure 4.3. This figure suggests the visualisations "Map" and "Repeating" for visualising a relationship. Both are used; the repetition is used in the form of a layered map, on which multiple visualisations are displayed. The layers consists of visualisation of sized circles for the absolute data and a value visualisation for the relative data, as displayed in figure 4.6. Layering two different visualisations allowed users to see two different datasets in one figure. Cijfers op de Kaart does not have the functionality of visualising different datasets in one figure. Since the datasets that are used for the relationship visualisation in the product are available within Cijfers op de Kaart, the datasets can be visualised in Cijfers op de Kaart as well. However, the only way to visualise these datasets is separately, resulting in two figures that are not on the same page. It is debatable whether or not this can answer the "Does a given relationship hold?"-question, but it can be argued that repetition can also work in the form of a repetition of multiple maps, even though they are not in the same figure or on the same webpage. Whether or not this gives a fair comparison between the tools in answering the same question, has been derived from the user tests in a later phase of this thesis.

**"What proportion?"**

The fourth question that a user can choose in the dropdown menu is "What is the percentage of?", which is derived from the "What proportion?"-question. For municipalities, it was best to visualise this as a value, since the other visualisation options were hard to show on municipalities or got cluttered, besides the fact that it was most suitable with the marks available within Vega-Lite. This meant that each municipality had a different value depending on the percentage; a lower percentage is closer to white than a higher percentage. The only data set available for this question was the population density, which caused that it is the only option to choose from in the product as well. This was also of influence on the "What is the relationship between?", since that visualisation uses both the value and circle visualisations, and "What is the percentage of?" is the only question that creates a value visualisation, meaning that all options within "What is the relationship between?" include the dataset of population density in combination with another dataset.

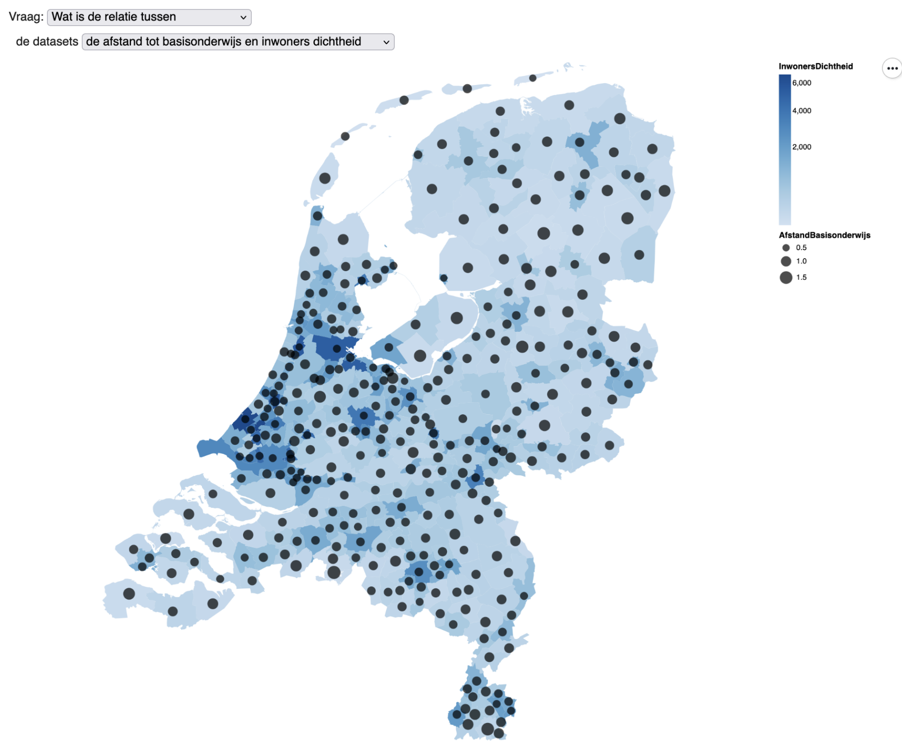


FIGURE 4.6: Visualisation: "What is the relationship between population density and distance to education?"

## 4.3 Usability evaluation

The usability evaluation consisted of qualitative comparative usability tests, in which participants compared the created product to Cijfers op de Kaart. The user tests have been performed with 18 Dutch-speaking participants, most of whom had some data visualisation experience, except for three. One of the participants had prior experience with "Cijfers op de Kaart", and one of the participants had prior experience with geodata visualisation tools such as Vega-Lite, D3, and ArcGIS. All participants got the task to point out the number of inhabitants of the municipality of Rotterdam, and whether or not that is the municipality with the highest amount of inhabitants in the Netherlands. The participants were divided into two groups, of which the group containing all odd participants (n1, n3, ..., n17) first used the product to answer this question, followed by using Cijfers op de Kaart to answer the question again. The second group contained all even participants (n2, n4, ..., n18) and first used Cijfers op de Kaart to answer the question, followed by the product. After the participant performed the task using each tool, the researcher asked a set of questions. This interview was followed by the participants filling out a question that they came up with themselves in both tools, always using the product first since that tool only contained a subset of the datasets that Cijfers op de Kaart includes. After this, another interview was held. For each participant, this process has been audio- and screen recorded with their permission, in order to be able to evaluate the results. For clarity, the product is called "Vega-Lite tool" in this chapter.

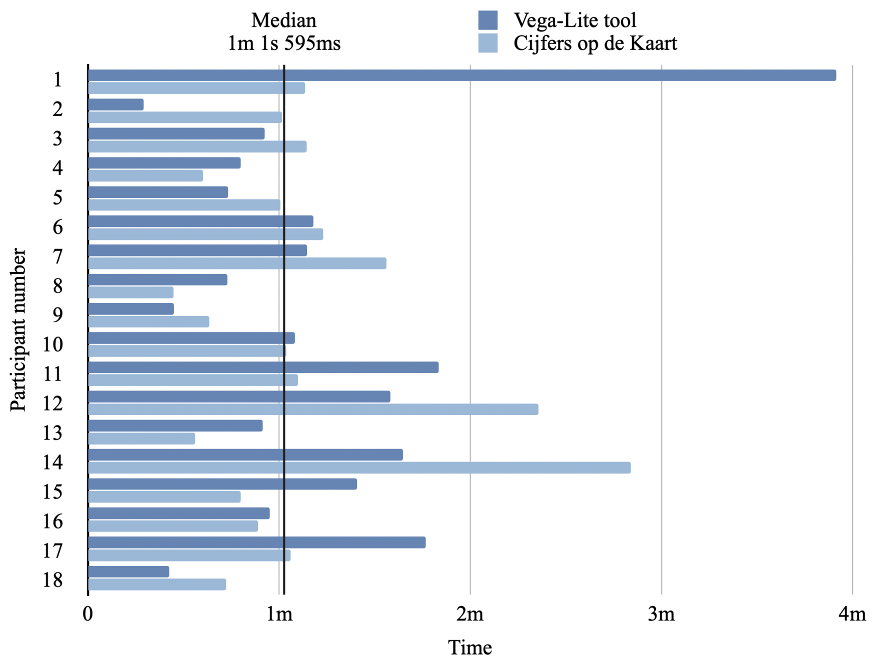


FIGURE 4.7: Time to complete task in Cijfers op de Kaart and the Vega-Lite tool

### 4.3.1 Quantitative results

The quantitative results in this research are the results that have been measured, such as the time it took participants to complete the task and interview questions that participants answered on a scale. Figure 4.7 shows the time each participant took to complete the task of pointing out how many inhabitants Rotterdam has and whether it is the largest municipality in both tools.

#### Time to complete the task

Time measurements were started when participants confirmed that they understood the task and were ready to start, for example by saying "Okay" or by moving the mouse to a clickable object. The time ended when the participant had completed the task and either gave a visual or textual confirmation of this. The median time it took to complete the task for all users with both tools is 1 minute and 1 second. For 9 of the participants, it took longer to complete the task in Cijfers op de Kaart than the Vega-Lite tool, and for the other 9 participants, it was the other way around. It is not always the case that the first tool used for the tasks takes more time than the second tool used for the task. This was only true for 10 of the 18 participants. The participants who took more time to complete the task with the second tool than the first tool were using Cijfers op de Kaart as second tool in 4 out of 8 cases, and in 4 out of 8 cases using the Vega-Lite second took longer. Figures 4.8 and 4.9 show the time per participant for both tools isolated.

As can be seen in the isolated figures, the median time it took the participants to complete the task, for both the Vega-Lite tool and Cijfers op de Kaart, is 1 minute and 1 second. The median time it took to complete the task with the first tool used by participants is 1 minutes and 9 seconds and 1 minute and 1 second for the Vega-Lite tool and Cijfers op de Kaart, respectively. Using the tools as the second tool came to a median time of 57 seconds for the Vega-Lite tool and 1 minute and 3 seconds for Cijfers op de

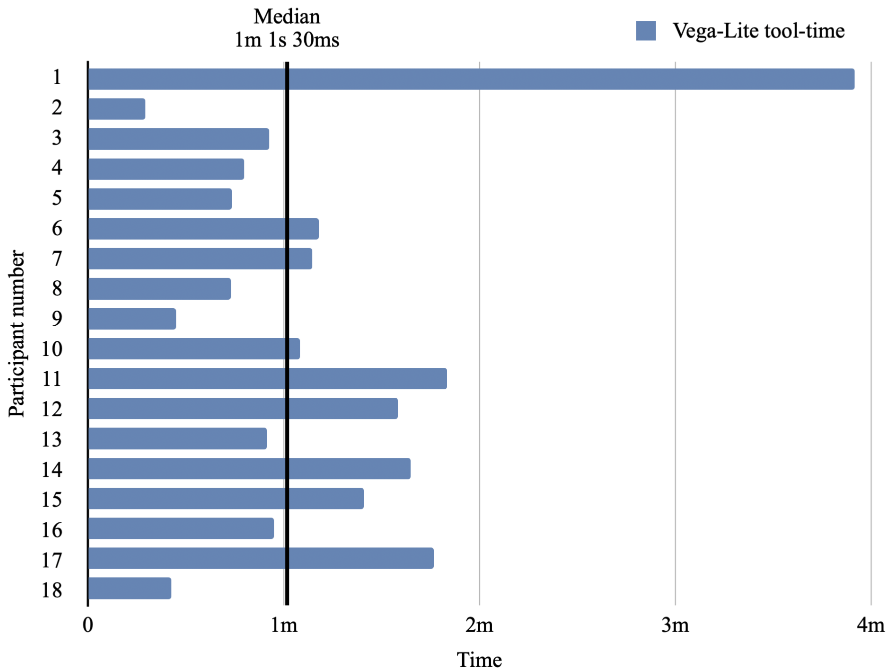


FIGURE 4.8: Time to complete task in the Vega-Lite tool

Kaart.

In figures 4.7 and 4.8 it can be seen that there is one obvious outlier, being participant 1. This participant used Vega-Lite first, but in the interview it became apparent that the task was not clear to the participant at first, which explains the outlier. Removing this outlier resulted in a median time of 1 minute for both tools in total and a median time of 57 seconds for the Vega-Lite tool. This participant is not disregarded of, since the answers at the interview have been useful as input for the research.

### Scaled interview questions

Although the answers to the interview questions that were answered on a scale are not numerical, the answers have been counted to allow comparison. The comments along with the results have been written in section 4.3.2. 10 out of 18 participants commented that they would use the Vega-Lite tool again, 7 would only use Cijfers op de Kaart again, and 1 participant would not use either again. Figures 4.10 to 4.13 show the results of interviews performed immediately after completion of the tasks in both tools. The figures have the concerned question that was asked to the participants in the figure title.

To the question how easy it was to answer the data question with both tools, 14 of the 18 participants rated the tool made with Vega-Lite "Very easy" or "Easy", and for Cijfers op de Kaart, this was 12 participants. All other participants rated the question with "A little bit easy" or "Neutral". This can also be seen in figure 4.10. 5 of the participants rated the Vega-Lite tool easier than Cijfers op de Kaart, 6 participants rated both tools the same, and 7 participants rated Cijfers op de Kaart easier than the Vega-Lite tool. The usability of both tools was mostly rated as "Good" or "Very good", with 13 participants for the Vega-Lite tool in total, and 15 for Cijfers op de Kaart. Others rated them as "Neutral", "A little bad", or "Bad", as can be seen in figure 4.11. Half of the participants rated the usability for both tools the same, 5 participants thought the usability of Cijfers

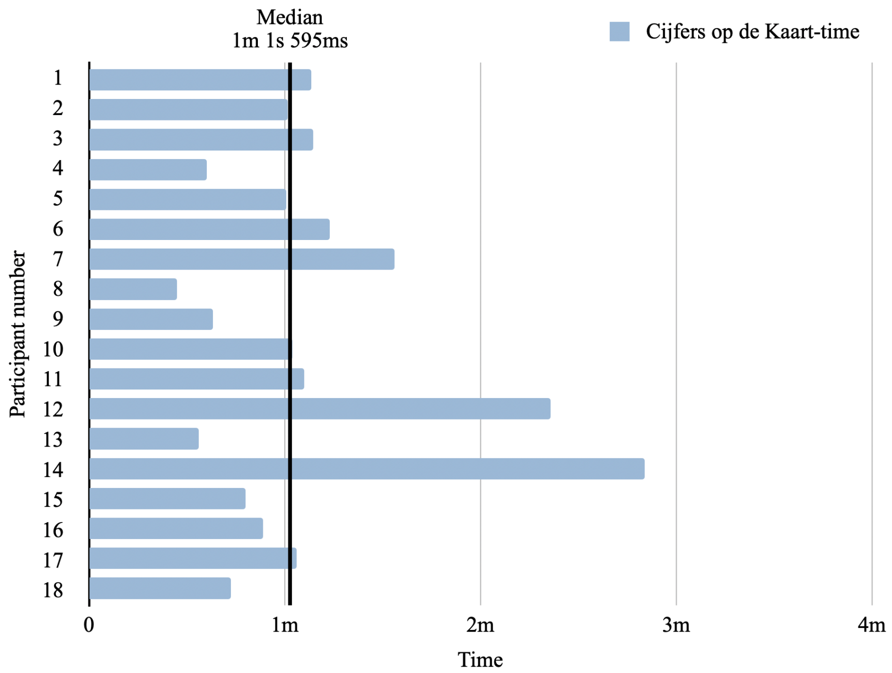


FIGURE 4.9: Time to complete task in the Cijfers op de Kaart tool

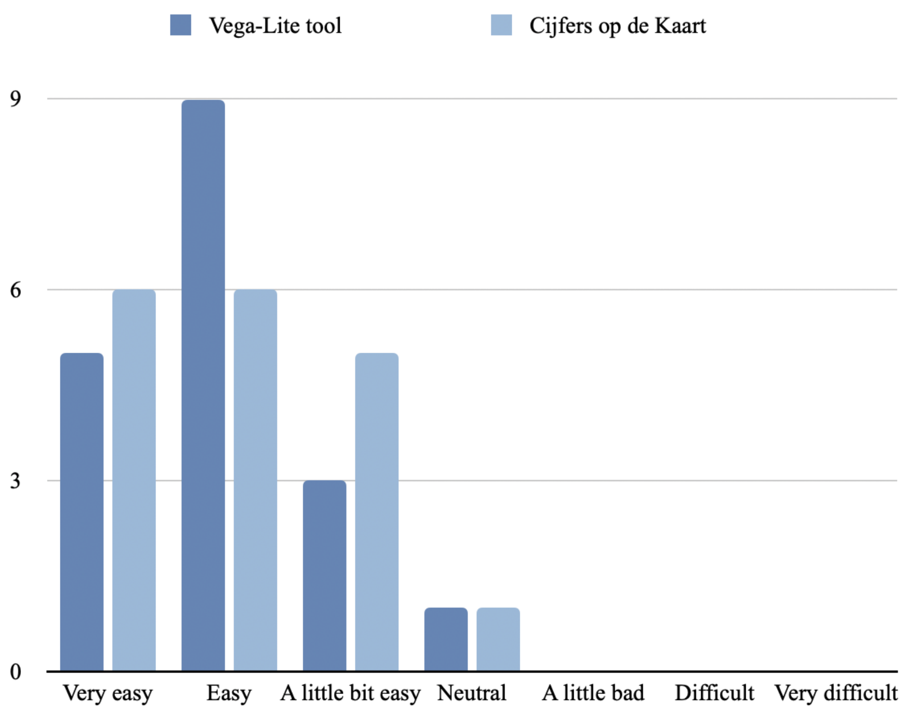


FIGURE 4.10: Question: How easy was it to answer the data question with this tool?

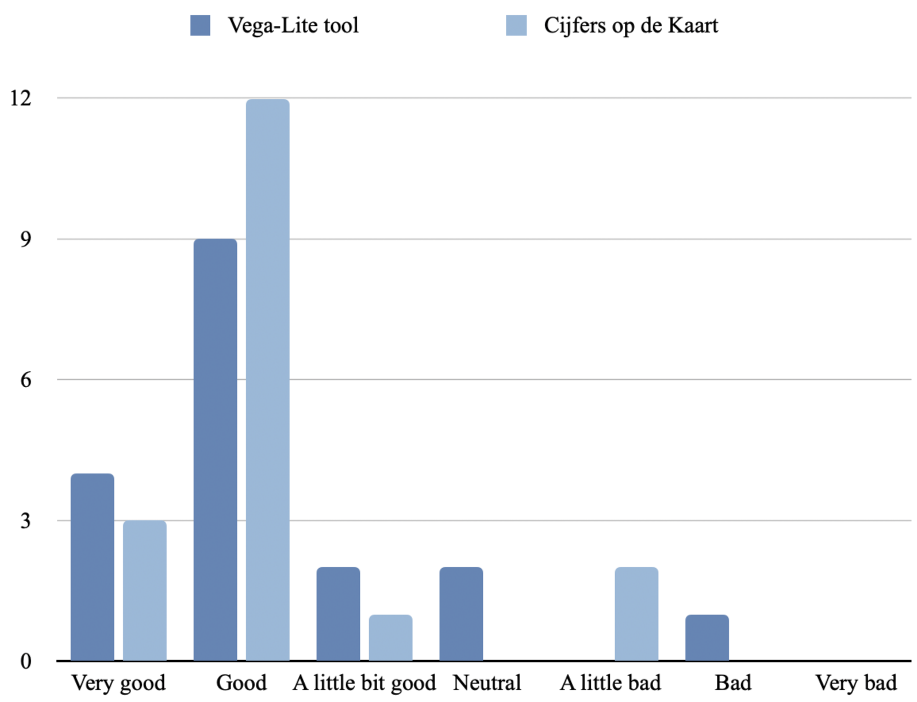


FIGURE 4.11: Question: How would you rate the usability of this tool?

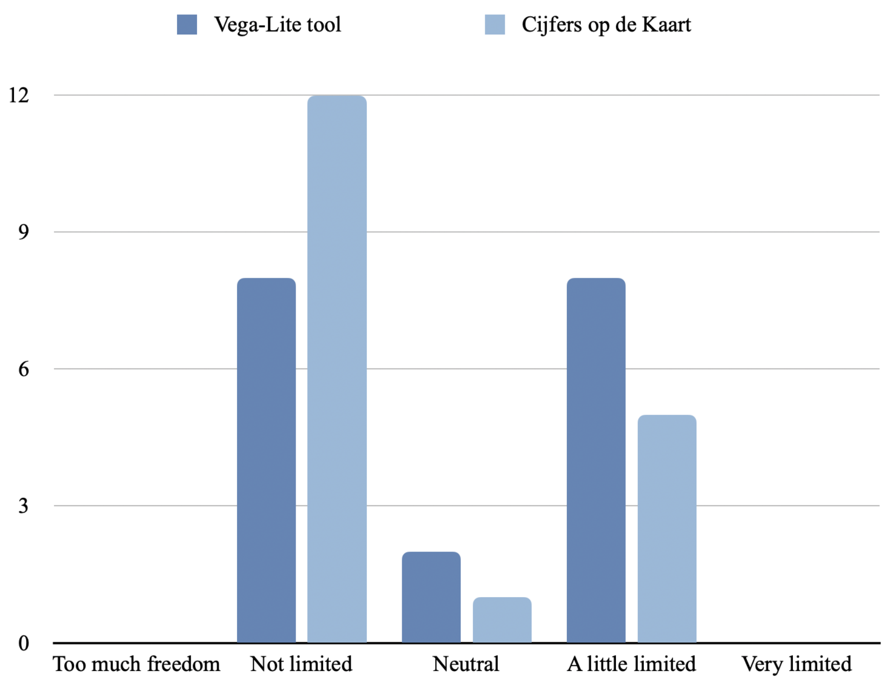


FIGURE 4.12: Question: Did you feel limited using this tool?

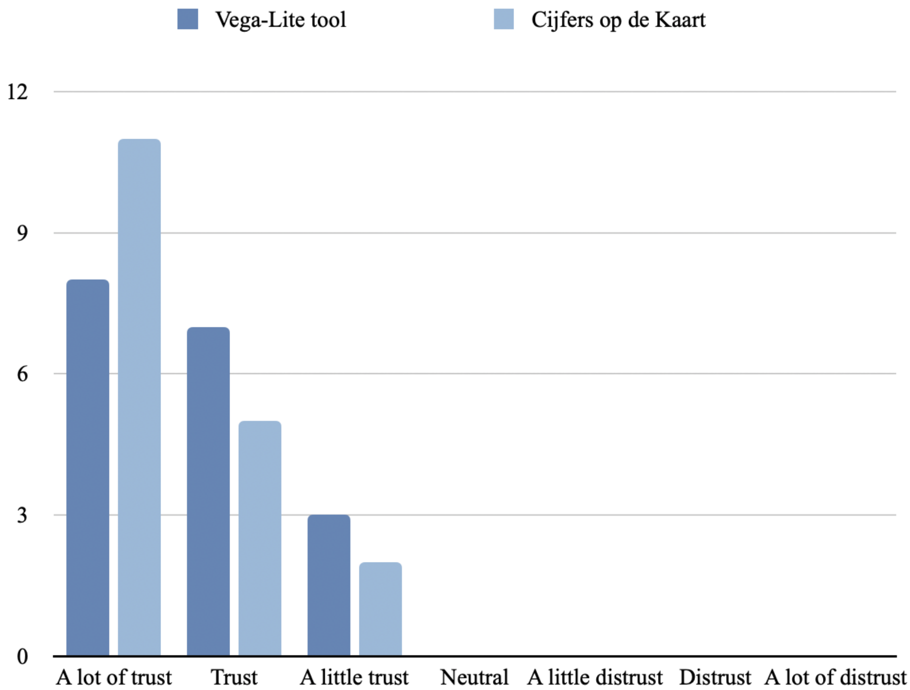


FIGURE 4.13: Question: How much do you trust that this tool makes a correct visualisation?

op de Kaart was better than the tool with Vega-Lite, and 3 participants rated the usability of the Vega-Lite tool better. Figure 4.12 shows how participants responded to the question if they felt limited or not using the tool. 6 participants felt more limited using the Vega-Lite tool rather than Cijfers op de Kaart, 1 participant felt less limited using the Vega-Lite tool, and for the other participants, both tools felt equally limiting. The last question for the interview after completing the task (figure 4.13), was about how much the participants trusted the tools to make the correct visualisations. All participants felt at least a little trust for both tools, though Cijfers op de Kaart was given "A lot of trust" most often. 5 out of 18 times, Cijfers op de Kaart was rated more trustful than the Vega-Lite tool, and the Vega-Lite tool was rated more trustful only once. The other 12 times, both tools were rated equally trusty for making visualisations.

### 4.3.2 Qualitative results

The qualitative results are derived from comments made by participants during the user test. The comments have been categorised to give an overview of the general opinions of the participants. General comments from the users are included in this section. Since the interviews were conducted in Dutch, the comments have been translated from Dutch to English.

#### Comments on topographic display

Thirteen participants commented that the tool made with Vega-Lite does not include location on the map or a location search function, making it hard to navigate. The option to search is a functionality of Cijfers op de Kaart, besides the fact that it shows location names on the map. The Vega-Lite tool did not do this, unless the user hoovers over the place, only allowing the user the view one place at a time. Comments about topography on the data-visualisation of the Vega-Lite tool are the following:

> *Participant 1*: Beforehand I did not expect to get a map, but just data. The map was useful, actually, but only if you knew what you were searching for. A bit of topography knowledge is useful. I found the Hague first, and Rotterdam only after that.
>
> *Participant 3*: I had expected a list with data, and personally that would have been my preference. The visualisation looks quite nice however, like this. I know where Rotterdam lies on a map, but if you would have picked a lesser known municipality, it would have taken me some searching.
>
> *Participant 4*: I have to search now. I now know where Rotterdam is located [after using Cijfers op de Kaart]. [...] With this tool it is easy to complete the task, but it helped that I now know where Rotterdam is located, approximately, and that it is one of the bigger circles. Because it does not say so on the map. If you would have mentioned a new place that I didn't know, I would have had to hoover over all of them.
>
> *Participant 6*: I could not use the search function to find the city I was looking for, so you would have to know where the city is that you are looking for. Luckily, that was not a problem for me in this case.
>
> *Participant 7*: I am really bad with topography. It is a miracle I found Rotterdam so quickly.
>
> *Participant 8*: [...] Then the map came, which I thought was quite good. But it was unfortunate I could not click the place, which might be my bad because my topographic knowledge is bad, so then I don't really know where I should be looking on the map.
>
> *Participant 11*: It is harder to find a place on this map, if you don't know where it is located. If I had to find a place I don't know, I would have to hoover over every municipality in order to find it.
>
> *Participant 12*: It is that I know where Rotterdam is located, that I could find it. I would have preferred to search for Rotterdam by typing it in a search bar, and that it would be emphasized then.
>
> *Participant 14*: With this tool, you are very dependent on your topographic knowledge.
>
> *Participant 15*: It is that I know where Rotterdam is located that I could find it, a search bar would have been easier. If you would have given me another municipality for the task, maybe I would find it more difficult to complete the task.
>
> *Participant 16*: You cannot search for a location. The circle visualisation for Rotterdam is large, but for a location with a smaller circle, maybe I would need to search for it on Google Maps for example, to see where it is.

> *Participant 17*: I though the first bit of Cijfers op de Kaart was easier [than the other tool], that I could search for Rotterdam and then immediately Rotterdam popped up.

> *Participant 18*: I do find this tool clear. But, if you wouldn't have topographic knowledge, it would be challenging to find Rotterdam. Then the task would be harder to complete.

**Comments on data transparency**

Six participants commented on the data list included with the Cijfers op de Kaart tool. Usually, the first impression of this list was good, but once they found out it could not be sorted, they were a bit disappointed. One of the participants even transferred the data to a spreadsheet tool so that they could order it themselves. Because this was included in Cijfers op de Kaart, some participants expected listed data, or commented that a list of data would be a nice addition to the tool made with Vega-Lite. 8 of the 18 participants made a comment on this.

> *Participant 1*: I think I expected tables, or something like that, to show up. Numbers, instead of a graphic representation.

> *Participant 2*: [With Cijfers op de Kaart,] I can view the data as a list. That is nice. [...] I did not see that option for the other tool.

> *Participant 3*: I did not expect a map. I expected a list with municipalities. [...] I would have really liked a list to answer this question.

> *Participant 4*: Okay, you can see everywhere how it works, but a clear list is not a thing, it seems like? I do have to get used to the overlayed map. But it is quite intuitive.

> *Participant 7*: The Cijfers op de Kaart "Data" options is useful, I can see everything in a row. [...] I like that that makes the data traceable. [...] It is easier for me to find data in a list, especially if I would be able to sort it.

> *Participant 12*: It makes me more content just looking at data, rather than looking on a map. [...] For the question "Where is the highest or lowest..." I would have expected ranked data, but the map stayed the same.

> *Participant 14*: When I clicked "Where is the highest or lowest...", I expected to see a list with data, or numbers. [...] How the map is displayed is good, but if you would have a list of data next to it that can be sorted, that would be more valuable for me.

> *Participant 17*: With Cijfers op de Kaart, it helps that you select "Data" and "Information", to go to the source-table. That is not possible for the other tool.

Another two participants did not necessarily say they required a list, but they did want a better, traceable source of data, such as a clickable link to the source somewhere near the tool.

### Map expectations

There were four participants who specifically mentioned that they expected different maps for different questions. The visualisation was the same for both the question "Where is the highest or lowest..." and "How much or many...". This is only specifically said by even-numbered participants, so participants that used Cijfers op de Kaart before using the Vega-Lite tool. Half of those participants would use the tool made with Vega-Lite again.

### Comparability of data

What most of the participants valued was the comparability of the Vega-Lite tool. Because of the larger visual contrast in the visualisations, data points were easier to compare. On top of that, one of the options in the tool is "What is the relationship between...", followed by datasets the users could pick. This option visualised two different datasets on one map and was received positively by the participants. There were four participants who chose to ask the question "What is the relationship between..." in the last round of their usability test. All four of these participants mentioned that they would like to use the Vega-Lite tool again. Other forms of comparability that was appreciated in that tool, was between municipalities. The participants valued the contrast between the visual marks in the visualisation. It was often mentioned that the visualisation of Cijfers op de Kaart was rather cluttered when zoomed out. The circle-visualisations sometimes were larger than the municipality itself, and with many larger circles in one area, the circles sometimes overlapped as well.

### Comments on using the tool again

Comments on why participants would use the Vega-Lite tool again are listed below.

> *Participant 4*: If I know where the place is, and the datasets would be both as complete, I would use [the Vega-Lite tool]. Scroll through all the questions, click all possibilities. I'd like that.
>
> *Participant 5*: I would use the tool recreational, out of curiosity, and I think for that purpose this tool would work better, since it gives suggestions for questions.
>
> *Participant 6*: I would like to use [the Vega-Lite tool] again, but I would have to get used to the way of navigation towards a visualisation. I like the visualisations better.
>
> *Participant 9*: I would like to use [the Vega-Lite tool] again if it is worked out more, although I find it a bit complicated to get to the visualisation through a question. Especially if more marks[1] would be implemented, the tool would be nicer than Cijfers op de Kaart.
>
> *Participant 10*: Now I know this tool exists, I would use it, probably for recreational questions. It can directly answer a question that I fill out. If the question that I have is not included, I can pick the most similar question.

[1] With marks, the Vega-Lite marks, as introduced in section 2.4.1, are meant. This participant had previous experience with Vega-Lite.

*Participant 13*: Navigating towards the visualisation is more obvious for [the Vega-Lite tool]. You have to get the hang of it, but we are now around 15 minutes in this conversation, and I already understand how it works; it's not rocket science.

*Participant 14*: [The Vega-Lite tool] does not seem completely worked out yet, it's not an eye-catcher so to say. That is better for Cijfers op de Kaart. However, if I would be specifically looking for something I would use [the Vega-Lite tool]. If the looks of both tools would be the same, I would use [the Vega-Lite tool].

*Participant 15*: I do not really care for which tool to use, it's the same data and I like gaining insight over the data, it is really valuable that it is on a map.

*Participant 16*: Cijfers op de Kaart cannot really make a one-view comparison between maps, the other tool can. If I had to pick one of the two, I would just check which makes the clearest visualisation. [...] Asking a question to the tool is almost like setting up a research question. It makes you think more about the question and information.

*Participant 17*: I would use this tool again, if it would allow me to edit the code for example, depending on the code. Making visualisation myself would not limit me to the datasets available. [...] Asking a question can be useful, but more often not so much, for example if my question would be different. A search bar for questions would be better.

The participants that would not use the Vega-Lite tool again, made the following comments:

*Participant 1*: I know CBS a bit, and it seems like Cijfers op de Kaart has more options available. Even if I did not know where to search, I would just look around. A lotten is written down, so it seems like more information is available.

*Participant 2*: Cijfers op de Kaart looks more trustworthy.

*Participant 3*: Even though the source is the same, Cijfers op de Kaart looks more trustworthy. And [the Vega-Lite tool] made me feel restricted, which was not a nice feeling. It feels like it only answers a specific question, it does not leave room for interpretation, and I think interpretation is really important for statistics and data.

*Participant 7*: For my own questions, I would prefer Cijfers op de Kaart, because I prefer looking to data and apparently my topographic knowledge is not very good.

*Participant 8*: I am already familiar with CBS, so I would use Cijfers op de Kaart again.

*Participant 11*: I don't think I would use something like this again, I never look on a map. I dropped geography [in highschool] for a reason. Both tools make choices that I wouldn't make myself. A different visualisation than a map could be better.

*Participant 12*: Cijfers op de Kaart makes more sense. The categories are more clear.

*Participant 18*: Cijfers op de Kaart has more options. I can see more maps, and I like that.

# Chapter 5

# Discussion

This chapter discusses the results found in chapter 4 on different topics. Considerations and interpretations for the research process are featured.

## 5.1 Comparing the tools

It has been mentioned in section 4.2 that it is unsure whether or not it is a fair comparison between the Vega-Lite tool and Cijfers op de Kaart when asking for "What is the relationship between?". During the usability tests there was one participant who said that they could not answer the relationship question in Cijfers op de Kaart, after they had chosen it in the proof-of-concept tool for the second part of the usability test. This participant tried to answer the question with Cijfers op de Kaart, but said that they would probably have come to a different conclusion with Cijfers op de Kaart than with the Vega-Lite tool, if they would not have known the correct conclusion beforehand. The other three participants who picked "What is the relationship between?" as a question, did not conclude that they could not answer the question properly. From this, it is hard to say whether or not this was a fair comparison. All participants agreed that it was a strong asset for the Vega-Lite tool and that they enjoyed seeing the layered visualisation. A possibility that none of the participants explored, is opening a second web-page to have a side-by-side view of the two Cijfers op de Kaart maps. This was not suggested by the researcher either, so the participants likely did not consider this as an option. Since it was a strong asset of the Vega-Lite tool, it is important that it was in the usability tests, to ensure that the participants could see the capabilities of the tool. Cijfers op de Kaart also had functionalities that could not be directly translated to the Vega-Lite tool, such as the "Search for a location"-function. With these factors, the comparison can be considered fair.

## 5.2 Declarative language

Not all of the Vega-Lite tool has been made with Vega-Lite, it is partly done in HTML. This makes it that not everything of the tool has been made with a declarative visualisation language. However, the use of HTML code is limited; only the introduction and drop-down menus have been written in HTML. Vega-Lite offers the possibility of incorporating a drop-down menu; however, given that the User Interface (UI) was initially constructed using HTML, the drop-down menu was also created using HTML rather than Vega-Lite. The creation of drop-down menus in HTML was also a more logical approach, as the insertion of two drop-down menus, as seen in the final proof-of-concept tool, was not a possibility within Vega-Lite. The visualisation results from the tool are completely from the declarative visualisation language Vega-Lite, so that the results are considered as results from a declarative visualisation language.

## 5.3 Dependency on Cijfers op de Kaart

As discussed in section 4.2, the Vega-Lite tool was dependent on the datasets and functionalities available within Cijfers op de Kaart. If Cijfers op de Kaart included more datasets, for example for percentage, more options would be available for "What is the percentage of?"-questions. At the moment, the only option is population density for both tools, but CBS also includes datasets such as percentages of land use, that Cijfers op de Kaart does not include. However, these data sets could potentially be visualised within both tools. Besides that, in section 2.3, some other questions have been introduced that are not in the tool and do have the potential of being visualised with Vega-Lite, but did not have any datasets within Cijfers op de Kaart to which they could be compared. These questions are "How certain?", "Which order or ranking?", and "Which group or category?". The latter two questions have qualitative, rather than numerical answers, hence why they can not be answered with Cijfers op de Kaart, that solely visualises numbers (hence the name, "Numbers on the Map"). Certainty could be expressed numerically, for example in percentages, but is not something CBS includes in their datasets either at this moment of time, and therefore Cijfers op de Kaart does not include certainty visualisations either.

## 5.4 Conflict of interest

The usability tests were conducted with participants who were all known to the researcher. This may have introduced a conflict of interest, as the UI of the Vega-Lite tool was relatively simple being a blank page with a visualisation, and most participants got aware that it was the tool created by the researcher. However, during the interviews, the participants were often asked to explain themselves when answering a question. When asked if the participants would use the tool again in the future, it was always followed by asking the participants to explain why they would use the tool in the future, and for what purpose. This ensured that all participants argued their answers. However, this cannot ensure that the participants subconsciously would rate the Vega-Lite tool more positively because of being familiar with researcher. Therefore, the times it took the participants to find an answer to the task with the tools was also measured. From these times, the median time was calculated in order to compare how long it took the participants to complete the task with each tool. For uniform datasets, the calculating the average would be better, but there were some outliers. These are kept in the collected data, because their input was still valuable, but had a relatively large influence on the average time. That is why it was chosen to go with median time, rather than the average time.

## 5.5 Think-Aloud method

The research made use of the "Think Aloud" method in order to collect findings. This method is useful for usability tests [43], but is rather unnatural for the participants. This was also noticed during the research. Participants often had to be reminded during the research that they had to think aloud, and would often start talking less during the use of the second tool. The first question asked by the researcher to the participants after using each tool was "Was there anything you noticed with this tool?" so that the participants could still explain their thoughts or say anything that they had not said yet during the think aloud. This worked well as an addition to the think-aloud protocol, although hindsight thoughts can still be different than thoughts at the moment of using it.

## 5.6 Participant expectations

Interpreting the results of the usability tests is that the Vega-Lite tool and Cijfers op de Kaart can be used for similar purposes, depending on the preferences of the users. Both tools are quick to learn and quick to use; most users had their answer to a question within a minute. During the interviews, many participants also gave feedback on the tools. Learning to use the Vega-Lite tool was not difficult, but some participants were surprised with the outcome. Not every participant expected to see a map, or knew what to expect when asking a question. For example, it was not clear that a second drop-down menu would pop up after clicking the first part of the question. All participants indicated that they had at least a little trust that the tool made a correct visualisation; often they indicated that since they saw CBS as a source for the data, they trusted the source, and that their trust for the tool was based on that. Other participants indicated that the maps checked out for them. For example, when visualising the inhabitants in the Netherlands, it made sense for the participants seeing a smallar number of inhabitants visualised in the North of the Netherlands, and a larger amount in the West, which made them trustful of the visualisation. Two participants indicated that it would have been better if the source had been traceable, in the form of a link to the dataset. Some other participants suggested that they would have liked to see the data somewhere within the tool, rather than a link to the data source. This could be in the form of a table or a bar chart within the tool. Often, this idea was from the tab "Data" from Cijfers op de Kaart, since that tab did include the raw data. This table however did not include the option to sort on data, only on municipalities, to the frustration of some participants.

An outcome that was not always expected by the participants was the result to "Where is the highest or lowest?", since it had the same output as the question "How much or many?". The first question was included, so that the "Where?" question could be included in the map, but the participants were setting expectations based on the last part of the question; "highest or lowest". This part probably hinted to be more similar to "Which order or ranking?", rather than "Where?", causing confusion in outcomes with the participants. Participants trying that option, indicated that they had expected a ranking, or a different visualisation from "How much or many?".

## 5.7 Question based grammar

The opinions of participants regarding making a visualisation based on a question were not consistent. Some participants liked being steered, and thought this as an easy and clear way to make the visualisation. Others felt limited by it, and mentioned that it felt weird that they were asking a question without getting a direct answer to it; they only got a visualisation, so they had to answer the question themselves. It was also mentioned that if the question they had would not be covered by the questions available within the tool, it would have been frustrating to still find an answer. Some participants concluded that asking questions in order to get to a visualisation was something that would be nice for people that would have a harder time understanding tools like Cijfers op de Kaart, such as students in high school or primary school. Different audience groups have not been tested; the participants mainly existed of (graduated) students, making the group of participants only relevant for a specific subgroup of people that make, or have an interest in making, data visualisations. On the other side, it was also mentioned by other participants that the Vega-Lite tool felt more academic, because it felt like asking a research question.

## 5.8 Interviews

The interviews were designed to take around 30 minutes, so that it would not take too much time of participants. This caused that the interviews were not as extensive as they could have been. For example, it can be seen that the interviews are error-prone looking at participant one. The interviews did not leave room for measuring the learning curve for both tools, or exploring the different options available within the Vega-Lite tool. This gave limited data for example on participants answering "What is the relationship between?" question, with only 4 out of 18 participants picking that question for the second task.

# Chapter 6

# Conclusions and Future Work

## 6.1 Conclusions

This section will answer each research question using the findings of this research.

### 6.1.1 Research Question 1

*How can the implementation of a visualisation grammar into a high-level declarative programming language contribute to the effective creation of thematic maps from data?*

The first research question is supported by two sub-questions:

**a** How can a visualisation grammar facilitate the generation of effective thematic maps from given data and data specifications?

**b** In what manner can such a grammar be structured for implementation within a high-level declarative programming language, such as Vega-Lite?

As stated in the methodology in chapter 3, this question can be answered with both literature research on a visualisation grammar, and user tests. Section 4.1 shows the result of a grammar based on the existing literature for visualisation grammars, and shaped in such a way that it is compatible for thematic maps and Vega-Lite. It takes into account the marks that are available within the high-level declarative data-visualisation language Vega-Lite. Just as Bertin's [3] reading process is based on questions that a user can ask to the visualisation, this grammar is based on questions a user has, on which a visualisation can be based. Having this grammar, that is specifically designed for thematic map making, does make it efficient to make thematic data visualisations on a map. This is also resulted from the usability tests performed; while users are not used to making visualisations with asking a question, the median time of making one and answering a question, was the same as for a more familiar tool. Majority of the participants (10 out of 18) would use the tool again, although it is not for everybody. Some participants felt restricting when having to ask a specific question, rather than selecting themes and datasets in order to get to a visualisation. To specifically answer the sub-questions:

**a** Having a grammar is useful since it ensures that the visualisations that are created are consistent. Besides that, the thematic map visualisation will also fit the dataset given, and will generate a visualisation depending on what type of data is the input. This can be derived from the data-question asked. This ensures that the user thinks about what they want to see before letting the tool make the corresponding visualisation. The tool does have to improve expectation management however; for example by adding example questions and a map without any thematic data visualisation on it, so that the user expects that their data question is answered with a map.

**b** The grammar is structured in such a way that it is based on data-questions. This ensures easy incorporation into Vega-Lite. The data-questions can be asked to an HTML-interface, which makes that the correct information is passed onto Vega-Lite in a declarative format, which makes a visualisation from the information. The visualisations are based on questions, rather than a dataset. This makes it easier to generalise the tool on different datasets, and possibly datasets that are not yet incorporated in the tool. This is because the grammar for the visualisations is based on questions, so the visualisation is correct based on the question, and not the data set. This makes it easier to swap between different data sets.

### 6.1.2 Research Question 2

*How can the user input be optimised to show the most effective map for the user's needs?*

Research question two is, just as the first research question, supported by two sub-questions:

**a** Which of the specifications should have degrees of freedom such that a user can enable the creation and exploration of meaningful thematic maps?

**b** Can a thematic map-making tool yield the intended output for a map based on a dataset given by a user?

The UI of the proof-of-concept tool is kept relatively simple, with an explanation on what the tool is, but it does not include any examples or options besides asking a question. The limited screen space for the UI leaves more screen space for the data visualisation. The UI has one clickable option, a drop down menu, which gives four questions that they can choose from. Based on the question, a user can pick one thematic dataset from different dataset to complete the question. This makes it that all clicks are straightforward, guiding the user towards the visualisation. The only problem is that the user does not always know what to expect when selecting a question with a dataset.

**a** The degrees of freedom chosen for the tool used for usability testing are the questions and datasets. After usability testing, some users indicated they would like to be able to make some changes to the visualisation, or have the option to gain more insight in the data. This corresponds with the results in figure 4.12, where it can be seen the participants on average felt a bit more limited using the tool made with Vega-Lite, rather than Cijfers op de Kaart. When using Cijfers op de Kaart, users did not complain on the options to show the datasets in a different year, or make the area more specific. This makes it that a wider range of degrees of freedom is something that should be explored.

**b** The users can explore the thematic map with a tool-tip: a mouse-over function that shows the name and data-value when hovering over a municipality. This allows the user to answer their question asked to the tool, while using the map to answer it. The tool can not directly give a textual answer to the question asked by the user, since the questions the user asks the tool are often more general than the user has in mind. For example, a question could often be about one or two municipalities, but a municipality could not be inserted in the question for the tool. This is something that the users had to find on a map. For the usability tests, the thematic map-making tool could therefore yield the intended output for a dataset. However, this dataset was chosen by the user, not given. Because of the flexibility of the grammar, since it is based on questions, it is also possible to yield the intended output for a map based on a dataset *given* by a user, if the tool would be further developed.

### 6.1.3 Research Question 3

*How can the User Interface be created, such that it enables the user to articulate their data inquiries in a declarative format within the tool effectively?*

The UI has been created using HTML, which includes the title, an introduction of the tool, and a drop down menu. The drop down menu includes questions that are connected to corresponding data visualisations, which visualise the thematic data that the user can input in using a second drop down menu, which shows up once the user selects one of four questions available. These drop down menus steer the user towards creating a question suitable for making a data visualisation. The user is limited to the questions available within the tool, currently being four, resulting in three different visualisations. This means that the user can not come up with any question which they can put in the tool, but has to base their question on the ones available. The limitation of questions is caused by making the tool suitable for usability tests, and does have potential for incorporation of more questions, since the grammar does include more questions.

As already slightly covered in the answers of the previous research questions, the grammar is designed in such a way the visualisations are based on questions. This is something that can be seen back in the tool that was used for usability testing. The user asks a question to the tool, and the visualisations are based on the questions, rather than the datasets themselves. This makes it that the datasets are interchangeable within each question. The declarative format is based on the question asked by the user, which passes onto the declarative visualisation language Vega-Lite, that makes a visualisation out of the dataset chosen within the data-question. The only thing the user has to do for this, is selecting a question, followed by selecting a dataset. This means that the user does not have to understand declarative formats in order to be able to make a data visualisation that is based on a declarative format.

## 6.2 Future work

While the tool has shown potential, several areas require further exploration to improve its functionality and user experience. For example, implementing a search feature similar to "Cijfers op de Kaart" would allow users to locate specific places without the need to hover over multiple locations. This improvement would serve to optimise the user experience and enhance the tool's efficiency. Another functionality that could work for the tool is the possibility to upload a dataset when asking a question, which would increase the flexibility of the tool. This means that any dataset can be visualised, and the user is not limited to the ones available within the tool. There is also a possibility to incorporate a model that can recognize questions in natural language. This model could pass the question onto Vega-Lite in a declarative format such that Vega-Lite can return a visualisation. Something that participants asked for was a (sortable) list, table, or barchart that accompanies the map. These are features that are supported by Vega-Lite, and these features could possibly also make searching for a place easier if it would be interactive. For example, a bar chart sorted by the data appears alongside the map when asking the question. When a location on the barchart is clicked, it lights up on the map, so it can easily be found. This would give a quicker answer to the "Where is the highest or lowest?"-question. Examples of this, made with Vega-Lite, are already available, such as [5] and [6]. Unfortunately, they do not include the geoshape-mark, but could be used as an inspiration for a visualisation including a geoshape-mark. An example has been created incorporating the geoshape-mark, and can be found in Appendix F.

Customisation could be added for map making. As section 4.1 states, one question does not always have one way of visualisation in order to be correct. The tool could potentially let the user choose between visualisation options for each question, or allow the user to make changes to colour, so that the user can fit the visualisation to their liking. There was even one participant who specifically stated that they would like to edit the code if that was a possibility, so it could be explored whether this could potentially work for the right audience. Since all of the participants of the study were (graduated) students, most of whom were familiar with data visualisations, some of the participants mentioned that the tool could work for an audience less familiar with making data visualisations. This is something that could be explored by performing usability tests with for example primary- or high school students with an interest for data visualisation. A new round usability testing would also allow to perform more extensive usability testing, exploring the learning curve of the Vega-Lite tool, and the opinion of users on all the different functionalities of the tool.

# Bibliography


[1] How does cbs collect data? - the netherlands in numbers 2022 | cbs. URL: https://longreads.cbs.nl/the-netherlands-in-numbers-2022/how-does-cbs-collect-data/ [cited 1-10-2023].

[2] Melih Basaraner. Revisiting cartography: towards identifying and developing a modern and comprehensive framework. *Geocarto International*, 31(1):71–91, 2016. doi:10.1080/10106049.2015.1041560.

[3] J. Bertin. *Sémiologie Graphique*. Mouton, Paris/Den Haag, 1967.

[4] Jeremy Black. *Maps and politics*. University of Chicago Press, 2000.

[5] José Borges. Car sales in portugal - vega-lite. URL: https://observablehq.com/@jlmborges/vega-lite-car_sales [cited 01-06-2024].

[6] José Borges. Interactive dashboard for the analysis of the number of registered cars in portugal. URL: https://paginas.fe.up.pt/~jlborges/vis/cars_dashboard/db.html [cited 01-06-2024].

[7] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. $D^3$ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011. doi:10.1109/TVCG.2011.185.

[8] CBS. Cijfers op de kaart. URL: https://www.cbs.nl/nl-nl/visualisaties/cijfers-op-de-kaart [cited 29-05-2024].

[9] Noam Chomsky. Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124, 1956. ISSN: 2168-2712. doi:10.1109/TIT.1956.1056813.

[10] Sidonie Christophe, Bertrand Duménieu, Jérémie Turbet, Charlotte Hoarau, Nicolas Mellado, Jérémie Ory, Hugo Loi, Antoine Masse, Benoit Arbelot, Romain Vergne, Mathieu Brédif, Thomas Hurtut, Joëlle Thollot, and David Vanderhaeghe. Map Style Formalization: Rendering Techniques Extension for Cartography. In Pierre Bénard and Holger Winnemöller, editors, *Non-Photorealistic Animation and Rendering*. The Eurographics Association, 2016. ISBN: 978-3-03868-002-4. doi:10.2312/exp.20161064.

[11] Intersoft Consulting. General data protection regulation (GDPR). European Union. URL: https://gdpr-info.eu/ [cited 18-10-2023].

[12] Nicholas Cox. [review of] the grammar of graphics. *Journal of Statistical Software*, 17(1):1–7, 2007. ISSN: 1548-7660. doi:10.18637/jss.v017.b03.

[13] Yuri Engelhardt and Clive Richards. A framework for analyzing and designing diagrams and graphics. In *Diagrammatic Representation and Inference: 10th International Conference, Diagrams 2018, Edinburgh, UK, June 18-22, 2018, Proceedings 10*, pages 201–209. Springer, 2018.

[14] Yuri Engelhardt and Clive Richards. The dna framework of visualization. In *International Conference on Theory and Application of Diagrams*, pages 534–538. Springer, 2020.

[15] Karel Feenstra. Cijfers op de kaart: nieuwe cbs-visualisatie van regionale statistieken, 3 2023. URL: https://www.cbs.nl/nl-nl/corporate/2023/11/cijfers-op-de-kaart-nieuwe-cbs-visualisatie-van-regionale-statistieken [cited 01-10-2023].

[16] A.U. Frank and M.J. Egenhofer. Computer cartography for GIS - an object-oriented view on the display transformation. *Computers & Geosciences*, 18(8):975–987, 1992. ISSN: 0098-3004. doi:10.1016/0098-3004(92)90015-J.

[17] U. Freitag. National atlases: Production and use. *Modern Cartography for Navigating the Information Highway, ed. by Ormeling, F., Villanueva, K. and Tichelaar, T.R.*, pages 15–22, 1997.

[18] Sybil Adams Gertrud Schaab and Serena Coetzee. Conveying map finesse: thematic map making essentials for today's university students. *Journal of Geography in Higher Education*, 46(1):101–127, 2022. arXiv:https://doi.org/10.1080/03098265.2020.1850656, doi:10.1080/03098265.2020.1850656.

[19] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67, 2010.

[20] J. Heesen, D.F. Lorenz, M. Nagenborg, B. Wenzel, and M. Voss. Blind spots on achilles' heel: The limitations of vulnerability and resilience mapping in research. *International Journal of Disaster Risk Science*, 5:74–85, 2014. doi:10.1007/s13753-014-0014-5.

[21] Marius Hogräfer and Hans-Jörg Schulz. ReVize: A Library for Visualization Toolchaining with Vega-Lite. In Marco Agus, Massimiliano Corsini, and Ruggero Pintus, editors, *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference*. The Eurographics Association, 2019. doi:10.2312/stag.20191375.

[22] Keshif. Data visualisation tools. URL: https://gallery.keshif.me/VisTools [cited 31-08-2023].

[23] B.J. Köbben. Proof of concept national atlas. URL: https://geoserver.itc.nl/ [cited 6-12-2023].

[24] B.J. Köbben. Towards a National Atlas of the Netherlands as part of the National Spatial Data Infrastructure. *The Cartographic Journal*, 50(3):225–231, 2013. doi:10.1179/1743277413Y.0000000056.

[25] Menno-Jan Kraak. Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(5):390–399, 2003. Challenges in Geospatial Analysis and Visualization. ISSN: 0924-2716. URL: https://www.sciencedirect.com/science/article/pii/S0924271602001673, doi:10.1016/S0924-2716(02)00167-3.

[26] Menno-Jan Kraak, R. Roth, B. Ricker, A. Kagawa, and G. Le Sourd. *Mapping for a sustainable world*. United Nations and the International Cartographic Association, 2020. ISBN: 9789216040468. doi:10.18356/9789216040468.

[27] Barend Köbben. Cartographic Grammar Explained, 2019. URL: https://kartoweb.itc.nl/TMT/OER/index.html [cited 2023-06-07].

[28] Barend Köbben, Menno-Jan Kraak, and Ferjan Ormeling. Een nieuwe atlas van Nederland - toegangspoort tot de nationale geo-data infrastructuur. *Geografie*, 19:28–31, 2 2010.

[29] Axel van Lamsweerde. Formal specification: a roadmap. In *Proceedings of the Conference on The Future of Software Engineering*, ICSE '00, pages 147–159, New York, NY, USA, 2000. ACM. ISBN: 1-58113-253-0.

[30] Youying Lin and Vivek Paramasivam. Geographic support for vega-lite. Technical report, University of Washington, 2016. URL: http://bobocandys.github.io/fp-liny33-paramv/.

[31] Mark Monmonier. *How to lie with Maps*. University of Chicago Press, 1996. ISBN: 9780226534213. doi:10.7208/chicago/9780226029009.001.0001.

[32] Stichting Wetenschappelijke Atlas Nederland. De nationale atlas van nederland – 3e editie. URL: http://avn.geo.uu.nl/Editie3Index.html [cited 5-11-2023].

[33] De grote bosatlas flex. URL: degrotebosatlas.nl [cited 01-09-2023].

[34] University of Washington Interactive Data Lab. Vega-lite – a grammar of interactive graphics. URL: https://vega.github.io/vega-lite/ [cited 8-11-2023].

[35] Ate Poorthuis, Lucas van der Zee, Grace Guo, Jo Hsi Keong, and Bianchi Dy. Florence: a web-based grammar of graphics for making maps and learning cartography. *Cartographic Perspectives*, 96:32–50, 2020. URL: https://cartographicperspectives.org/index.php/journal/article/view/1645, doi:10.14714/CP96.1645.

[36] Clive Richards and Yuri Engelhardt. The DNA of Visualisation, 2023. URL: visdna.com [cited 8-11-2023].

[37] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350, 2016. doi:10.1109/TVCG.2016.2599030.

[38] Arvind Satyanarayan, Ryan Russell, Jane Hoffswell, and Jeffrey Heer. Reactive vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):659–668, 2016. doi:10.1109/TVCG.2015.2467091.

[39] Ilse Schieven. Thematic map project. URL: https://thematic-map-proj.onrender.com/ [cited 29-05-2024].

[40] R. Sieber and S. Huber. Atlas of switzerland 2 – a highly interactive thematic national atlas. *Multimedia Cartography ed. by Cartwright, W., Peterson, M. P. and Gartner, G.*, pages 161–182, 2007.

[41] Angeliki Tsorlini, René Sieber, Lorenz Hurni, Hubert Klauser, and Thomas Gloor. Designing a rule-based wizard for visualizing statistical data on thematic maps. *Cartographic Perspectives*, 86:5–23, 2017. URL: https://cartographicperspectives.org/index.php/journal/article/view/cp86-tsorlini-et-al, doi:10.14714/CP86.1392.

[42] Edward R. Tufte. Envisioning information. *Optometry and Vision Science*, 68(4):322–324, April 1991.

[43] Tomas Vanicek and Stanislav Popelka. The think-aloud method for evaluating the usability of a regional atlas. *ISPRS International Journal of Geo-Information*, 12(3):95, 2023. doi:10.3390/ijgi12030095.

[44] Colin Ware. *Information Visualization*. Interactive Technologies. Morgan Kaufmann, Boston, 3rd edition, 2013. ISBN: 978-0-12-381464-7. doi:10.1016/B978-0-12-381464-7.00020-X.

[45] Travis White. Symbolization and the visual variables. John P. Wilson (ed.), 2017. ISSN: 2577-2848. URL: https://gistbok.ucgis.org/bok-topics/symbolization-and-visual-variables, doi:10.22224/gistbok/2017.2.3.

[46] Leland Wilkinson. *The Grammar of Graphics*. Statistics and Computing. Springer, 2nd edition, 2005. ISBN: 0-387-24544-8.

## Appendix A

# Participant consent form (Dutch)

Beste deelnemer,

Het doel van dit experiment is om de geschiktheid van de programmeertaal “Vega-Lite” voor het maken van thematische kaarten, te testen. Met Vega-Lite is een proof-of-concept tool gemaakt, welke de gebruiker helpt bij het maken van data visualisaties op thematische kaarten. Tijdens het experiment zal de onderzoeken informatie verzamelen over de usability van de proof-of-concept tool, en of de tool waardevolle visualisaties maakt. Alles wat de deelnemer zegt kan worden meegenomen in het onderzoek of voor de verdere ontwikkeling van de proof-of-concept tool.
Het testen is een kwalitatieve usability evaluatie waar de deelnemers twee verschillende data-visualisatie zullen gebruiken en vergelijken: de proof-of-concept thematische kaarten programma, en een al eerder bestaand programma, zijnde “Cijfers op de Kaart” van Centraal Bureau Statistiek (CBS). De kwalitatieve evaluatie betekent dat het experiment een interview-achtige vorm aan neemt, waar de interviewer vragen stelt aan de deelnemer en meer vragen kan stellen op basis van de antwoorden. Dit interview neemt plaats na het gebruik van de programma’s. Daarnaast wordt de gebruiker gevraagd om hardop te denken tijdens het gebruiken van beide tools (think aloud methode). Tijdens het gebruik van de tools zal dan ook de audio (anoniem) i.c.m. het beeldscherm worden opgenomen. Er wordt niet gefilmd.
De deelnemer krijgt per programma twee taken toegewezen van de onderzoeker, welke hij of zij dan moet uitvoeren. De volgorde van het gebruik van programma’s wordt willekeurig toegewezen.

De totale duur van het experiment is ongeveer 30 minuten. Er zijn geen risico’s met deelname aan dit experiment. De verzamelde data zal worden verwijderd na het einde van het onderzoek en is niet herleidbaar naar de deelnemers. De resultaten zijn niet herleidbaar naar de deelnemers. De verzamelde data zal alleen worden gebruikt voor dit onderzoek en zal niet worden gedeeld. De resultaten van dit onderzoek zijn wel publiek beschikbaar. De deelnemer is niet verplicht om deel te nemen aan dit experiment en kan er elk moment tijdens het experiment voor kiezen om te stoppen.
Mocht je nog vragen hebben over het onderzoek kun je contact opnemen met de onderzoeker, Ilse Schieven, via i.schieven@student.utwente.nl
Als je vragen hebt over je rechten als deelnemer aan dit onderzoek, als je meer informatie wil, vragen wil stellen, of je wil praten over zorgen betreft dit onderzoek met iemand anders dan de onderzoeker, kun je contact opnemen met de Secretaris van de Ethics Committee Information & Computer Science via: ethicscommittee-CIS@utwente.nl

- Ik heb heb de onderzoeksinformatie (gedateerd [____/____/________]) gelezen of voorgelezen gekregen en begrepen. Ik had de mogelijkheid tot vragen stellen over het onderzoek en mijn vragen zijn voldoende beantwoord. [ Ja / Nee ]
- Ik geef vrijwillig toestemming om een deelnemer te zijn in dit onderzoek en ik begrijp dat ik kan wijgeren om vragen te beantwoorden, en dat ik mijn deelname, zonder reden, op elk moment kan terugrekken. [ Ja / Nee ]
- Ik begrijp dat deelname aan dit experiment impliceert dat er audio en schermopnames gemaakt worden. [ Ja / Nee ]
- Ik begrijp dat de informatie die ik vergeef gebruikt kan worden voor een onderzoek thesis. [ Ja / Nee ]
- Ik geef toestemming voor audio-opnames tijdens dit onderzoek. [ Ja / Nee ]
- Ik geef toestemming voor schermopnames tijdens dit onderzoek. [ Ja / Nee ]
- Ik begrijp dat persoonlijke informatie die wordt verzameld over mij, welke mij kan identificeren, zoals audio-opnames, niet zullen worden gedeeld door de onderzoeker. [ Ja / Nee ]

| | | |
|---|---|---|
| ____________________ Naam van de deelnemer | ____________________ Handtekening | ________ Datum |
| I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting. | | |
| Ilse Schieven Naam van de onderzoeker | ____________________ Handtekening | ________ Datum |

# Appendix B

# Participant questionnaire (Dutch)

## Vragen van tevoren:

Ben je bekend met data visualisatie programma’s, zoals:

- Vega-Lite
- D3
- Tableau
- PowerBI
- Adobe Illustrator (als data visualisatie tool)
- Anders, namelijk…

Ben je bekend met geografische programma’s, zoals:

- Cijfers op de Kaart
- ArcGIS
- GeoPython
- ArcPy
- QGIS
- Anders, namelijk…

## **Vragen na tool met gesteld data question:**

| | | | | | | |
|---|---|---|---|---|---|---|
| Wat waren dingen die opvielen bij het gebruiken van het programma tijdens de taak? | | | | | | |
| | | | | | | |
| | | | | | | |
| Was de visualisatie zoals je had verwacht? (2. Als ja, hoe was het anders dan verwacht?) | | | | | | |
| | | | | | | |
| | | | | | | |
| Hoe gemakkelijk ging het om de data-vraag te beantwoorden met deze tool? | | | | | | |
| Erg gemakkelijk | gemakkelijk | Een beetje makkelijk | Neutraal | Een beetje moeizaam | Moeizaam | Erg moeizaam |
| Erg gemakkelijk | gemakkelijk | Een beetje makkelijk | Neutraal | Een beetje moeizaam | Moeizaam | Erg moeizaam |
| Hoe zou jij de usability van de tool beoordelen? De definitie van usability is hier: “Met de tool kan ik op een efficiënte en effectieve manier het doel halen.” | | | | | | |
| Erg goede usability | Goede usability | Een beetje goede usability | Neutraal | Een beetje slechte usability | Slechte usability | Erg slechte usability |
| Erg goede usability | Goede usability | Een beetje goede usability | Neutraal | Een beetje slechte usability | Slechte usability | Erg slechte usability |
| Hoe beperkt voelde je tijdens het maken van een visualisatie met deze tool? (Had je meer vrijheid willen hebben in het maken van de visualisatie) | | | | | | |
| Er was te veel vrijheid | | Niet beperkt | Neutraal | Een beetje beperkt | | Erg beperkt |
| Er was te veel vrijheid | | Niet beperkt | Neutraal | Een beetje beperkt | | Erg beperkt |
| Waarom en op welke manier? | | | | | | |
| | | | | | | |
| | | | | | | |
| Hoe veel vertrouw je dat deze tool een juiste en correcte visualisatie maakt? | | | | | | |
| Erg veel vertrouwen | Vertrouwen | Een beetje vertrouwen | Neutraal | Een beetje wantrouwig | Wantrouwig | Helemaal geen vertrouwen |
| Erg veel vertrouwen | Vertrouwen | Een beetje vertrouwen | Neutraal | Een beetje wantrouwig | Wantrouwig | Helemaal geen vertrouwen |

**Questions after using each tool with own data question:**

Hoe zou je de data vraag beantwoorden?
Waren er dingen die opvielen tijdens het gebruik van het programma?
Was de visualisatie zoals je van tevoren had verwacht?
Is je mening over dit programma veranderd nu je je eigen data-vraag hebt beantwoord? Hoe?
Zou je dit programma in de toekomst weer gebruiken? (Waarom, waarvoor, en welk alternatief?)

# Appendix C

# Product code

Code for the final tool excluding the comments.

```
<!DOCTYPE html>
<html class="no-js" lang="">

<head>
  <meta charset="utf-8">
  <title>Thematic map tool</title>
  <meta name="description" content="">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <meta property="og:title" content="Thematic map tool">
  <meta name="theme-color" content="#fafafa">
  <link rel="stylesheet" href="css/normalize.css">
  <link rel="stylesheet" href="css/main.css">

  <style>
    #additionalText {
      display: none;
      top: 100%;
      left: 0;
      width: 100%;
      padding: 10px;
      z-index: 1;
    }
  </style>

</head>

<body>
<h1>Thematische kaarten tool</h1>

<p> Deze thematische kaarten tool maakt visualisaties gebaseerd op data vragen. <br>
  De visualisaties kunnen alleen op gemeente-level gemaakt worden.
  De datasets zijn van de publieke bron "Centraal Bureau voor Statistiek" (CBS) en
  zijn zo recent mogelijk (uit 2022). <br>
  Hier onder kun je je eigen data vraag invullen:<br>
  </p>
```

```
<form action="/action_page.php">
  <label for="qs">Vraag: </label>
  <select id="qs" name="qs">
    <option value="null">---</option>
    <option value="where">Waar is het hoogste of het laagste</option>
    <option value="relationship">Wat is de relatie tussen</option>
    <option value="many">Hoe veel</option>
    <option value="percentage">Wat is het percentage</option>
  </select>
</form>
```

```
<div id="additionalText"></div>
<script>
  document.getElementById('qs').addEventListener('change', function () {
    var selectedValue = this.value;
    var additionalText = document.getElementById('additionalText');
    switch (selectedValue) {
      case 'where':
        additionalText.innerHTML = `
          <label for="additionalDropdown"> </label>
          <select id="additionalDropdown-where">
            <option value="null-1">---</option>
            <option value="Inwoners">aantal inwoners</option>
            <option value="Mannen">aantal mannen</option>
            <option value="Vrouwen">aantal vrouwen</option>
            <option value="BesteedbaarInkomenPerHuishouden1000eu">
            gestandaardiseerde inkomen</option>
            <option value="NrBasisscholenBinnen3km">
            afstand tot basisonderwijs</option>
            <option value="WOZwaarde">gemiddelde WOZ-waarde</option>
            <option value="WaterOppervlakte">wateroppervlakte (km2)</option>
            <option value="Personenauto">aantal personenautos</option>
            <option value="Uitkeringsontvangers">
            aantal uitkerings ontvangers</option>
          </select>
        `;
        const whereDropdownOptions = (event) => {
          var selectedValue = event.target.value;
          switch (selectedValue) {
            case 'Inwoners':
              gInwonersFunction();
              break;
            case 'Mannen':
              gMannenFunction();
              break;
            case 'Vrouwen':
              gVrouwenFunction();
```

```
            break;
          case 'BesteedbaarInkomenPerHuishouden1000eu':
            gInkomenFunction();
            break;
          case "NrBasisscholenBinnen3km":
            gBasisscholenFunction();
            break;
          case "WOZwaarde":
            gWozFunction();
            break;
            case 'WaterOppervlakte':
            gWaterOppFunction();
            break;
          case 'Personenauto':
            gAutoFunction();
            break;
          case 'Uitkeringsontvangers':
            gUitkeringFunction();
            break;
          default:
            break;
        }
      }
      document.getElementById('additionalDropdown-where').addEventListener(
      'change', whereDropdownOptions);
      break;
    case 'relationship':
      additionalText.innerHTML = `
        <label for="firstDropdown">de datasets</label>
        <select id="firstDropdown-relationship">
          <option value="null-1">---</option>
          <option value="InwonersCompare">
          de gemiddelde WOZ-waarde en de inwoners dichtheid</option>
          <option value="BasisscholenCompare">
          de afstand tot basisonderwijs en inwoners dichtheid</option>
        </select>
      `;
const relationshipDropdownOptions = (event) => {
        var selectedValue = event.target.value;
        switch (selectedValue) {
          case 'InwonersCompare':
            gInwonersCompFunction();
            break;
          case 'BasisscholenCompare':
            gBasisscholenCompareFunction();
            break;

          default:
            break;
```

```
    }
  }

  document.getElementById('firstDropdown-relationship').addEventListener(
  'change', relationshipDropdownOptions);
  break;
case 'many':
  additionalText.innerHTML = `
    <label for="additionalDropdown"> </label>
    <select id="additionalDropdown-many">
      <option value="null-1">---</option>
      <option value="WaterOppervlakte">is het wateroppervlakte</option>
      <option value="InkomenGestandaardiseerdx1000">
      is het gestandaardiseerde inkomen</option>
      <option value="Personenauto">personenauto's zijn er</option>
      <option value="Uitkeringsontvangers">
      uitkerings ontvangers zijn er</option>
      <option value="Inwoners">inwoners zijn er</option>
      <option value="Mannen">mannen zijn er</option>
      <option value="Vrouwen">vrouwen zijn er</option>
      <option value="NrBasisscholenBinnen3km">
      is de afstand tot basisonderwijs</option>
      <option value="WOZwaarde">is de gemiddelde WOZ-waarde</option>

    </select>
  `;
  const manyDropdownOptions = (event) => {
    var selectedValue = event.target.value;
    switch (selectedValue) {
      case 'WaterOppervlakte':
        gWaterOppFunction();
        break;
      case 'InkomenGestandaardiseerdx1000':
        gInkomenFunction();
        break;
      case 'Personenauto':
        gAutoFunction();
        break;
      case 'Uitkeringsontvangers':
        gUitkeringFunction();
        break;
        case 'Inwoners':
        gInwonersFunction();
        break;
      case 'Mannen':
        gMannenFunction();
        break;
      case 'Vrouwen':
        gVrouwenFunction();
```

```
            break;
          case "NrBasisscholenBinnen3km":
            gBasisscholenFunction();
            break;
          case "WOZwaarde":
            gWozFunction();

          default:
            break;
        }
      }
      document.getElementById('additionalDropdown-many').addEventListener(
      'change', manyDropdownOptions);
      break;
    case 'percentage':
      additionalText.innerHTML = `
        <label for="additionalDropdown">van </label>
        <select id="additionalDropdown-percentage">
          <option value="null-1">---</option>
          <option value="InwonersDichtheid">inwoners</option>
        </select>
      `;
            const percentageDropdownOptions = (event) => {
        var selectedValue = event.target.value;
        switch (selectedValue) {
          case 'InwonersDichtheid':
            gInwonersDichtFunction();
            break;
          default:
            break;
        }
      }

</script>


<script src="https://cdn.jsdelivr.net/npm/vega@5.25.0"></script>
<script src="https://cdn.jsdelivr.net/npm/vega-lite@5.16.3"></script>
<script src="https://cdn.jsdelivr.net/npm/vega-embed@6.22.2"></script>

<script id="municipalityScript">

  function gInwonersFunction() {
    const dataspec = {
      $schema: "https://vega.github.io/schema/vega-lite/v5.json",

      "height": 800,
      "width": 900,
```

```
layer: [
  {
    data: {
      url: "gemeente_gegeneraliseerd.json",
      format: {
        type: "json",
        property: 'features'
      }
    },
    transform: [
      {
        lookup: "properties.statcode",
        from: {
          data: {
            url: "Regionale_kerncijfers_Nederland_correct.csv",
            format: {type: 'dsv', delimiter: ';'}
          },
          key: "statcode",
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
          "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
          "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        },
      },
    ],

    mark: 'geoshape',
    projection: {type: 'identity', reflectY: true},
    encoding: {
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Name'
      },
        {field: 'Inwoners',
        type: "quantitative",
        title: 'Inwoners'},
      ]
    }
  },
  {
    data: {
      url: 'gemeente-labelpoint.json',
      format: {
       type: 'json',
       property: 'features'
      }
    },
    transform: [
      {
```

```
            lookup: 'properties.statcode',
            from: {
              data: {
                url: 'Regionale_kerncijfers_Nederland_correct.csv',
                format: {type: 'dsv', delimiter: ';'}
              },
              key: 'statcode',
              fields: ["Inwoners"]
            }
          }
        ],
        mark: 'circle',
        projection: {type:'identity', reflectY: true},
        encoding: {
          longitude: {
            field: "geometry.coordinates[0]",
            type: "quantitative"
          },
          latitude: {
            field: "geometry.coordinates[1]",
            type: "quantitative"
          },
          size: {
            field: 'Inwoners',
            type: 'quantitative',
          },
          color: {
            value: 'black'
          },
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          },
            {
              field: 'Inwoners',
              type: 'quantitative',
              title: 'Inwoners'
            }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {actions: false});


}

function gWozFunction(){
```

```
const dataspec = {
  $schema: "https://vega.github.io/schema/vega-lite/v5.json",

  "height": 800,
  "width": 900,

  layer: [
    {
      data: {
        url: "gemeente_gegeneraliseerd.json",
        format: {
          type: "json",
          property: 'features'
        }
      },
      transform: [
        {
          lookup: "properties.statcode",
          from: {
            data: {
              url: "Regionale_kerncijfers_Nederland_correct.csv",
              format: {type: 'dsv', delimiter: ';'}
            },
            key: "statcode",
            fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
            "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
            "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
          },
        },
      ],

      mark: 'geoshape',
      projection: {type: 'identity', reflectY: true},
      encoding: {
        tooltip: [{
          field: 'properties.statnaam',
          type: 'nominal',
          title: 'Name'
        },
          {field: 'WOZwaarde',
          type: "quantitative",
          title: 'WOZ-waarde (x1000 eu)'}]
      }
    },
    {
      data: {
        url: 'gemeente-labelpoint.json',
        format: {
         type: 'json',
```

```
      property: 'features'
    }
  },
  transform: [
    {
      lookup: 'properties.statcode',
      from: {
        data: {
          url: 'Regionale_kerncijfers_Nederland_correct.csv',
          format: {type: 'dsv', delimiter: ';'}
        },
        key: 'statcode',
        fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
        "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
        "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
      }
    }
  ],
  mark: 'circle',
  projection: {type:'identity', reflectY: true},
  encoding: {
    longitude: {
      field: "geometry.coordinates[0]",
      type: "quantitative"
    },
    latitude: {
      field: "geometry.coordinates[1]",
      type: "quantitative"
    },
    size: {
      field: 'WOZwaarde',
      type: 'quantitative',

    },
    color: {
      value: 'black'
    },
    tooltip: [{
      field: 'properties.statnaam',
      type: 'nominal',
      title: 'Name'
    },
      {
        field: 'WOZwaarde',
        type: 'quantitative',
        title: 'WOZ-waarde (x1000 eu)'
      }]
  }
}
```

```
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});
}

function gMannenFunction() {
  const dataspec = {
    $schema: "https://vega.github.io/schema/vega-lite/v5.json",

    "height": 800,
    "width": 900,

    layer: [
      {
        data: {
          url: "gemeente_gegeneraliseerd.json",
          format: {
            type: "json",
            property: 'features'
          }
        },
        transform: [
          {
            lookup: "properties.statcode",
            from: {
              data: {
                url: "Regionale_kerncijfers_Nederland_correct.csv",
                format: {type: 'dsv', delimiter: ';'}
              },
              key: "statcode",
              fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
              "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
              "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
            },
          },
        ],

        mark: 'geoshape',
        projection: {type: 'identity', reflectY: true},
        encoding: {
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          },
            {field: 'Mannen',
            type: "quantitative",
            title: 'Mannen'}]
        }
```

```
},
{
  data: {
    url: 'gemeente-labelpoint.json',
    format: {
     type: 'json',
     property: 'features'
    }
  },
  transform: [
    {
      lookup: 'properties.statcode',
      from: {
        data: {
          url: 'Regionale_kerncijfers_Nederland_correct.csv',
          format: {type: 'dsv', delimiter: ';'}
        },
        key: 'statcode',
        fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
        "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
        "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
      }
    }
  ],
  mark: 'circle',
  projection: {type:'identity', reflectY: true},
  encoding: {
    longitude: {
      field: "geometry.coordinates[0]",
      type: "quantitative"
    },
    latitude: {
      field: "geometry.coordinates[1]",
      type: "quantitative"
    },
    size: {
      field: 'Mannen',
      type: 'quantitative',
    },
    color: {
      value: 'black'
    },
    tooltip: [{
      field: 'properties.statnaam',
      type: 'nominal',
      title: 'Name'
    },
      {
        field: 'Mannen',
```

```
            type: 'quantitative',
            title: 'Mannen'
          }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});
}

function gVrouwenFunction() {
   const dataspec = {
    $schema: "https://vega.github.io/schema/vega-lite/v5.json",

    "height": 800,
    "width": 900,

    layer: [
      {
        data: {
          url: "gemeente_gegeneraliseerd.json",
          format: {
            type: "json",
            property: 'features'
          }
        },
        transform: [
          {
            lookup: "properties.statcode",
            from: {
              data: {
                url: "Regionale_kerncijfers_Nederland_correct.csv",
                format: {type: 'dsv', delimiter: ';'}
              },
              key: "statcode",
              fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
              "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
              "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
            },
          },
        ],

        mark: 'geoshape',
        projection: {type: 'identity', reflectY: true},
        encoding: {
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
```

```
    },
      {field: 'Vrouwen',
      type: "quantitative",
      title: 'Vrouwen'}]
  }
},
{
  data: {
    url: 'gemeente-labelpoint.json',
    format: {
     type: 'json',
     property: 'features'
    }
  },
  transform: [
    {
      lookup: 'properties.statcode',
      from: {
        data: {
          url: 'Regionale_kerncijfers_Nederland_correct.csv',
          format: {type: 'dsv', delimiter: ';'}
        },
        key: 'statcode',
        fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
        "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
        "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
      }
    }
  ],
  mark: 'circle',
  projection: {type:'identity', reflectY: true},
  encoding: {
    longitude: {
      field: "geometry.coordinates[0]",
      type: "quantitative"
    },
    latitude: {
      field: "geometry.coordinates[1]",
      type: "quantitative"
    },
    size: {
      field: 'Vrouwen',
      type: 'quantitative',
    },
    color: {
      value: 'pink'
    },
    tooltip: [{
      field: 'properties.statnaam',
```

```
            type: 'nominal',
            title: 'Name'
          },
            {
              field: 'Vrouwen',
              type: 'quantitative',
              title: 'Vrouwen'
            }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});
 }

 function gBesteedbaarFunction(){
const dataspec = {
     $schema: "https://vega.github.io/schema/vega-lite/v5.json",

     "height": 800,
     "width": 900,

     layer: [
       {
         data: {
           url: "gemeente_gegeneraliseerd.json",
           format: {
             type: "json",
             property: 'features'
           }
         },
         transform: [
           {
             lookup: "properties.statcode",
             from: {
               data: {
                 url: "Regionale_kerncijfers_Nederland_correct.csv",
                 format: {type: 'dsv', delimiter: ';'}
               },
               key: "statcode",
               fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
               "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
               "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
             },
           },
         ],

         mark: 'geoshape',
         projection: {type: 'identity', reflectY: true},
```

```
      encoding: {
        tooltip: [{
          field: 'properties.statnaam',
          type: 'nominal',
          title: 'Name'
        },
          {field: 'InkomenGestandaardiseerdx1000',
          type: "quantitative",
          title: 'Gestandaardiseerd inkomen per huishouden (1000 euro)'},]
      }
    },
    {
      data: {
        url: 'gemeente-labelpoint.json',
        format: {
         type: 'json',
         property: 'features'
        }
      },
      transform: [
        {
          lookup: 'properties.statcode',
          from: {
            data: {
              url: 'Regionale_kerncijfers_Nederland_correct.csv',
              format: {type: 'dsv', delimiter: ';'}
            },
            key: 'statcode',
            fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
            "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
            "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
          }
        }
      ],
      mark: 'circle',
      projection: {type:'identity', reflectY: true},
      encoding: {
        longitude: {
          field: "geometry.coordinates[0]",
          type: "quantitative"
        },
        latitude: {
          field: "geometry.coordinates[1]",
          type: "quantitative"
        },
        size: {
          field: 'InkomenGestandaardiseerdx1000',
          type: 'quantitative',
        },
```

```
          color: {
            value: 'black'
          },
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          },
            {
              field: 'InkomenGestandaardiseerdx1000',
              type: 'quantitative',
              title: 'Gestandaardiseerd inkomen per huishouden (1000 euro)'
            }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});
}

function gBasisscholenFunction(){
 const dataspec = {
    $schema: "https://vega.github.io/schema/vega-lite/v5.json",

    "height": 800,
    "width": 900,

    layer: [
      {
        data: {
          url: "gemeente_gegeneraliseerd.json",
          format: {
            type: "json",
            property: 'features'
          }
        },
        transform: [
          {
            lookup: "properties.statcode",
            from: {
              data: {
                url: "Regionale_kerncijfers_Nederland_correct.csv",
                format: {type: 'dsv', delimiter: ';'}
              },
              key: "statcode",
              fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
              "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
              "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
            },
```

```
        },
      ],

      mark: 'geoshape',
      projection: {type: 'identity', reflectY: true},
      encoding: {
        tooltip: [{
          field: 'properties.statnaam',
          type: 'nominal',
          title: 'Name'
        },
          {field: 'AfstandBasisonderwijs',
          type: "quantitative",
          title: 'Afstand tot basisonderwijs (km)'},]
      }
    },
    {
      data: {
        url: 'gemeente-labelpoint.json',
        format: {
         type: 'json',
         property: 'features'
        }
      },
      transform: [
        {
          lookup: 'properties.statcode',
          from: {
            data: {
              url: 'Regionale_kerncijfers_Nederland_correct.csv',
              format: {type: 'dsv', delimiter: ';'}
            },
            key: 'statcode',
            fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
            "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
            "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
          }
        }
      ],
      mark: 'circle',
      projection: {type:'identity', reflectY: true},
      encoding: {
        longitude: {
          field: "geometry.coordinates[0]",
          type: "quantitative"
        },
        latitude: {
          field: "geometry.coordinates[1]",
          type: "quantitative"
```

```
          },
          size: {
            field: 'AfstandBasisonderwijs',
            type: 'quantitative',
             scale: { range: [0, 300] }
          },
          color: {
            value: 'black'
          },
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          },
            {
              field: 'AfstandBasisonderwijs',
              type: 'quantitative',
              title: 'Afstand tot basisonderwijs (km)'
            }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});
}

function gInwonersDichtFunction() {
  const dataspec = {
    $schema: "https://vega.github.io/schema/vega-lite/v5.json",

    "height": 900,
    "width": 1000,
    data: {
      url: "gemeente_gegeneraliseerd.json",
      format: {
        type: "json",
        property: 'features'
      }
    },
    transform: [
      {
        lookup: "properties.statcode",
        from: {
          data: {
            url: "Regionale_kerncijfers_Nederland_correct.csv",
            format: {type: 'dsv', delimiter: ';'}
          },
          key: "statcode",
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
```

```
                "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
                "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        },
      }
    ],
    mark: 'geoshape',
    projection: {type: 'identity', reflectY: true},

    encoding: {
      color: {
        field: 'InwonersDichtheid',
        type: 'quantitative',
        scale: {type: 'sqrt', scheme: 'blues'}
      },
      legend: {
        title: null,
        format: "%"
      },
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Municipality'
      },
        {
          field: "InwonersDichtheid",
          type: "quantitative",
          title: "Inwoners Dichtheid"
        }]
    },


  };
  vegaEmbed('#data', dataspec, {action: false});
  }

function gWaterOppFunction(){
       const dataspec = {
      $schema: "https://vega.github.io/schema/vega-lite/v5.json",

      "height": 800,
      "width": 900,

      layer: [
        {
          data: {
            url: "gemeente_gegeneraliseerd.json",
            format: {
              type: "json",
              property: 'features'
```

```
      }
    },
    transform: [
      {
        lookup: "properties.statcode",
        from: {
          data: {
            url: "Regionale_kerncijfers_Nederland_correct.csv",
            format: {type: 'dsv', delimiter: ';'}
          },
          key: "statcode",
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
          "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
          "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        },
      },
    ],

    mark: 'geoshape',
    projection: {type: 'identity', reflectY: true},
    encoding: {
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Name'
      },
        {field: 'WaterOppervlakte',
        type: "quantitative",
        title: 'Water oppervlakte (km2)'},]
    }
  },
  {
    data: {
      url: 'gemeente-labelpoint.json',
      format: {
       type: 'json',
       property: 'features'
      }
    },
    transform: [
      {
        lookup: 'properties.statcode',
        from: {
          data: {
            url: 'Regionale_kerncijfers_Nederland_correct.csv',
            format: {type: 'dsv', delimiter: ';'}
          },
          key: 'statcode',
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
```

```
                    "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
                    "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
                  }
                }
              ],
              mark: 'circle',
              projection: {type:'identity', reflectY: true},
              encoding: {
                longitude: {
                  field: "geometry.coordinates[0]",
                  type: "quantitative"
                },
                latitude: {
                  field: "geometry.coordinates[1]",
                  type: "quantitative"
                },
                size: {
                  field: 'WaterOppervlakte',
                  type: 'quantitative',
                   scale: { range: [0, 800] }
                },
                color: {
                  value: 'black'
                },
                tooltip: [{
                  field: 'properties.statnaam',
                  type: 'nominal',
                  title: 'Name'
                },
                  {
                    field: 'WaterOppervlakte',
                    type: 'quantitative',
                    title: 'Water oppervlakte (km2)'
                  }]
              }
            }
          ]
        };
        vegaEmbed('#data', dataspec, {action: false});
    }

    function gInkomenFunction(){
           const dataspec = {
          $schema: "https://vega.github.io/schema/vega-lite/v5.json",

          "height": 800,
          "width": 900,

          layer: [
```

```
{
  data: {
    url: "gemeente_gegeneraliseerd.json",
    format: {
      type: "json",
      property: 'features'
    }
  },
  transform: [
    {
      lookup: "properties.statcode",
      from: {
        data: {
          url: "Regionale_kerncijfers_Nederland_correct.csv",
          format: {type: 'dsv', delimiter: ';'}
        },
        key: "statcode",
        fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
        "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
        "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
      },
    },
  ],

  mark: 'geoshape',
  projection: {type: 'identity', reflectY: true},
  encoding: {
    tooltip: [{
      field: 'properties.statnaam',
      type: 'nominal',
      title: 'Name'
    },
      {field: 'InkomenGestandaardiseerdx1000',
      type: "quantitative",
      title: 'Gestandaardiseerd inkomen (1000 eu)'},]
  }
},
{
  data: {
    url: 'gemeente-labelpoint.json',
    format: {
     type: 'json',
     property: 'features'
    }
  },
  transform: [
    {
      lookup: 'properties.statcode',
      from: {
```

```
                data: {
                  url: 'Regionale_kerncijfers_Nederland_correct.csv',
                  format: {type: 'dsv', delimiter: ';'}
                },
                key: 'statcode',
                fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
                "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
                "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
              }
            }
          ],
          mark: 'circle',
          projection: {type:'identity', reflectY: true},
          encoding: {
            longitude: {
              field: "geometry.coordinates[0]",
              type: "quantitative"
            },
            latitude: {
              field: "geometry.coordinates[1]",
              type: "quantitative"
            },
            size: {
              field: 'InkomenGestandaardiseerdx1000',
              type: 'quantitative',
               scale: {range: [0, 250]}
            },
            color: {
              value: 'black'
            },
            tooltip: [{
              field: 'properties.statnaam',
              type: 'nominal',
              title: 'Name'
            },
              {
                field: 'InkomenGestandaardiseerdx1000',
                type: 'quantitative',
                title: 'Gestandaardiseerd inkomen (1000 eu)'
              }]
          }
        }
      ]
    };
    vegaEmbed('#data', dataspec, {action: false});
}

function gAutoFunction(){
           const dataspec = {
```

```
$schema: "https://vega.github.io/schema/vega-lite/v5.json",

"height": 800,
"width": 900,

layer: [
  {
    data: {
      url: "gemeente_gegeneraliseerd.json",
      format: {
        type: "json",
        property: 'features'
      }
    },
    transform: [
      {
        lookup: "properties.statcode",
        from: {
          data: {
            url: "Regionale_kerncijfers_Nederland_correct.csv",
            format: {type: 'dsv', delimiter: ';'}
          },
          key: "statcode",
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
          "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
          "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        },
      },
    ],

    mark: 'geoshape',
    projection: {type: 'identity', reflectY: true},
    encoding: {
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Name'
      },
        {field: 'Personenauto',
        type: "quantitative",
        title: "Aantal personenauto's"},]
    }
  },
  {
    data: {
      url: 'gemeente-labelpoint.json',
      format: {
       type: 'json',
       property: 'features'
```

```
        }
      },
      transform: [
        {
          lookup: 'properties.statcode',
          from: {
            data: {
              url: 'Regionale_kerncijfers_Nederland_correct.csv',
              format: {type: 'dsv', delimiter: ';'}
            },
            key: 'statcode',
            fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
            "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
            "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
          }
        }
      ],
      mark: 'circle',
      projection: {type:'identity', reflectY: true},
      encoding: {
        longitude: {
          field: "geometry.coordinates[0]",
          type: "quantitative"
        },
        latitude: {
          field: "geometry.coordinates[1]",
          type: "quantitative"
        },
        size: {
          field: 'Personenauto',
          type: 'quantitative',
           scale: { range: [0, 800] }
        },
        color: {
          value: 'black'
        },
        tooltip: [{
          field: 'properties.statnaam',
          type: 'nominal',
          title: 'Name'
        },
          {
            field: 'Personenauto',
            type: 'quantitative',
            title: "Aantal personenauto's"
          }]
      }
    }
  ]
```

```
    };
    vegaEmbed('#data', dataspec, {action: false});
  }
function gUitkeringFunction(){
           const dataspec = {
      $schema: "https://vega.github.io/schema/vega-lite/v5.json",

      "height": 800,
      "width": 900,

      layer: [
        {
          data: {
            url: "gemeente_gegeneraliseerd.json",
            format: {
              type: "json",
              property: 'features'
            }
          },
          transform: [
            {
              lookup: "properties.statcode",
              from: {
                data: {
                  url: "Regionale_kerncijfers_Nederland_correct.csv",
                  format: {type: 'dsv', delimiter: ';'}
                },
                key: "statcode",
                fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
                "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
                "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
              },
            },
          ],

          mark: 'geoshape',
          projection: {type: 'identity', reflectY: true},
          encoding: {
            tooltip: [{
              field: 'properties.statnaam',
              type: 'nominal',
              title: 'Name'
            },
              {field: 'UitkeringsOntvangers',
              type: "quantitative",
              title: 'Aantal uitkeringsontvangers'},]
          }
        },
        {
```

```
data: {
  url: 'gemeente-labelpoint.json',
  format: {
   type: 'json',
   property: 'features'
  }
},
transform: [
  {
    lookup: 'properties.statcode',
    from: {
      data: {
        url: 'Regionale_kerncijfers_Nederland_correct.csv',
        format: {type: 'dsv', delimiter: ';'}
      },
      key: 'statcode',
      fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
      "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
      "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
    }
  }
],
mark: 'circle',
projection: {type:'identity', reflectY: true},
encoding: {
  longitude: {
    field: "geometry.coordinates[0]",
    type: "quantitative"
  },
  latitude: {
    field: "geometry.coordinates[1]",
    type: "quantitative"
  },
  size: {
    field: 'UitkeringsOntvangers',
    type: 'quantitative',
    scale: { range: [0, 800] }
  },
  color: {
    value: 'black'
  },
  tooltip: [{
    field: 'properties.statnaam',
    type: 'nominal',
    title: 'Name'
  },
    {
      field: 'UitkeringsOntvangers',
      type: 'quantitative',
```

```
                title: 'Aantal uitkeringsontvangers'
              }]
          }
        }
      ]
    };
    vegaEmbed('#data', dataspec, {action: false});
}

function gInwonersCompFunction(){
        const dataspec = {
      $schema: "https://vega.github.io/schema/vega-lite/v5.json",

      "height": 900,
      "width": 1000,

      layer: [
        {
          data: {
            url: "gemeente_gegeneraliseerd.json",
            format: {
              type: "json",
              property: 'features'
            }
          },
          transform: [
            {
              lookup: "properties.statcode",
              from: {
                data: {
                  url: "Regionale_kerncijfers_Nederland_correct.csv",
                  format: {type: 'dsv', delimiter: ';'}
                },
                key: "statcode",
                fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
                "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
                "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
              },
            },
          ],

      mark: 'geoshape',
        projection: { type: 'identity', reflectY: true},

        encoding: {
          color: {
            field: 'InwonersDichtheid',
            type: 'quantitative',
            scale: {type: 'sqrt', scheme: 'blues'}},
```

```
      legend: {
        title: null,
        format: "%"
      },
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Municipality'},

        {
          field: "InwonersDichtheid",
          type: "quantitative",
          title: "InwonersDichtheid"
        },
        {
          field: "WOZwaarde",
          type: 'quantitative',
          title: 'Gemiddelde WOZ-waarde (1000 eu)'
        }]}
    },
    {
      data: {
        url: 'gemeente-labelpoint.json',
        format: {
         type: 'json',
         property: 'features'
        }
      },
      transform: [
        {
          lookup: 'properties.statcode',
          from: {
            data: {
              url: 'Regionale_kerncijfers_Nederland_correct.csv',
              format: {type: 'dsv', delimiter: ';'}
            },
            key: 'statcode',
            fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
            "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
            "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
          }
        }
      ],
      mark: 'circle',
      projection: {type:'identity', reflectY: true},
      encoding: {
        longitude: {
          field: "geometry.coordinates[0]",
          type: "quantitative"
```

```
          },
          latitude: {
            field: "geometry.coordinates[1]",
            type: "quantitative"
          },
          size: {
            field: 'WOZwaarde',
            type: 'quantitative',
            scale: { range: [0, 300] }
          },
          color: {
            value: 'black'
          },
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          },
            {
              field: 'InwonersDichtheid',
              type: 'quantitative',
              title: 'Inwoners dichtheid'
            },
            {
              field: 'WOZwaarde',
              type: 'quantitative',
              title: 'Gemiddelde WOZ-waarde (1000 eu)'
            }]
        }
      }
    ]
  };
  vegaEmbed('#data', dataspec, {action: false});


}
function gBasisscholenCompareFunction(){
        const dataspec = {
      $schema: "https://vega.github.io/schema/vega-lite/v5.json",

      "height": 900,
      "width": 1000,

      layer: [
        {
          data: {
            url: "gemeente_gegeneraliseerd.json",
            format: {
              type: "json",
```

```
        property: 'features'
      }
    },
    transform: [
      {
        lookup: "properties.statcode",
        from: {
          data: {
            url: "Regionale_kerncijfers_Nederland_correct.csv",
            format: {type: 'dsv', delimiter: ';'}
          },
          key: "statcode",
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
          "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
          "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        },
      },
    ],

mark: 'geoshape',
  projection: { type: 'identity', reflectY: true},

  encoding: {
    color: {
      field: 'InwonersDichtheid',
      type: 'quantitative',
      scale: {type: 'sqrt', scheme: 'blues'}},
    legend: {
      title: null,
      format: "%"
    },
    tooltip: [{
      field: 'properties.statnaam',
      type: 'nominal',
      title: 'Municipality'},
      {
        field: "AfstandBasisonderwijs",
        type: "quantitative",
        title: "Afstand tot basisonderwijs (km)"
      },
      {
        field: "InwonersDichtheid",
        type: 'quantitative',
        title: 'Inwoners dichtheid'
      }]}
  },
  {
    data: {
      url: 'gemeente-labelpoint.json',
```

```
      format: {
       type: 'json',
       property: 'features'
      }
    },
    transform: [
      {
        lookup: 'properties.statcode',
        from: {
          data: {
            url: 'Regionale_kerncijfers_Nederland_correct.csv',
            format: {type: 'dsv', delimiter: ';'}
          },
          key: 'statcode',
          fields: ["Inwoners","Mannen","Vrouwen","InwonersDichtheid",
          "WOZwaarde","InkomenGestandaardiseerdx1000", "UitkeringsOntvangers",
          "Personenauto", "AfstandBasisonderwijs","WaterOppervlakte"]
        }
      }
    ],
    mark: 'circle',
    projection: {type:'identity', reflectY: true},
    encoding: {
      longitude: {
        field: "geometry.coordinates[0]",
        type: "quantitative"
      },
      latitude: {
        field: "geometry.coordinates[1]",
        type: "quantitative"
      },
      size: {
        field: 'AfstandBasisonderwijs',
        type: 'quantitative',
        scale: { range: [0, 300] }
      },
      color: {
        value: 'black'
      },
      tooltip: [{
        field: 'properties.statnaam',
        type: 'nominal',
        title: 'Name'
      },
        {
          field: 'AfstandBasisonderwijs',
          type: 'quantitative',
          title: 'Afstand tot basisonderwijs (km)'
        },
```

```
                {
                  field: 'InwonersDichtheid',
                  type: 'quantitative',
                  title: 'Inwoners dichtheid'
                }]
            }
          }
        ]
      };
      vegaEmbed('#data', dataspec, {action: false});


  }


    function runMunicipalityScript() {
      const dataspec = {
        $schema: "https://vega.github.io/schema/vega-lite/v5.json",

        "height": 900,
        "width": 1000,

        data: {
          url: "gemeente_gegeneraliseerd.json",
          format: {
            type: "json",
            property: 'features'
          }
        },
        mark: 'geoshape',
        projection: {type: 'identity', reflectY: true},
        encoding: {
          tooltip: [{
            field: 'properties.statnaam',
            type: 'nominal',
            title: 'Name'
          }]
        }

      };
      vegaEmbed('#data', dataspec, {action: false});
    }

  </script>
  <div id="data"></div>
  </body>
  </html>
```

# Appendix D

# Product for usability tests

The following figures are screenshots from the product with which the usability tests are done. The tool can be found online via [39].


FIGURE D.1: Questions menu


FIGURE D.2: Menu: Where is the highest or lowest...?

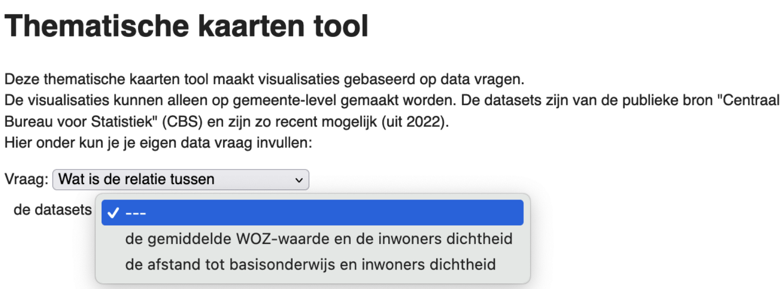


FIGURE D.3: Menu: What is the relationship between...?

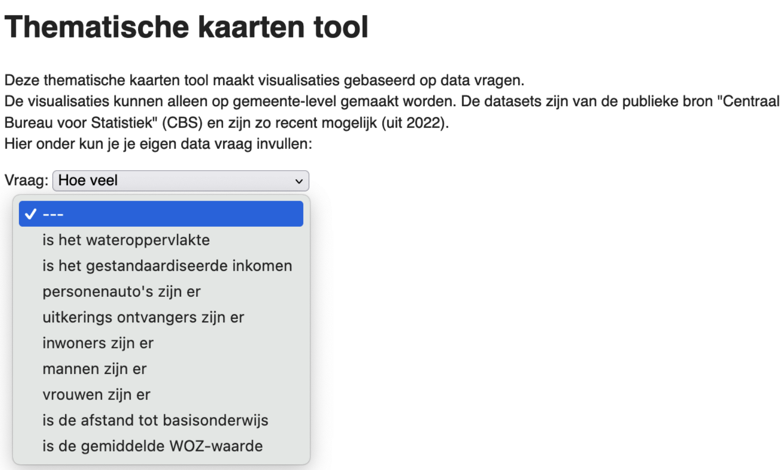


FIGURE D.4: Menu: How much/many...?


FIGURE D.5: Menu: What is the percentage of...?

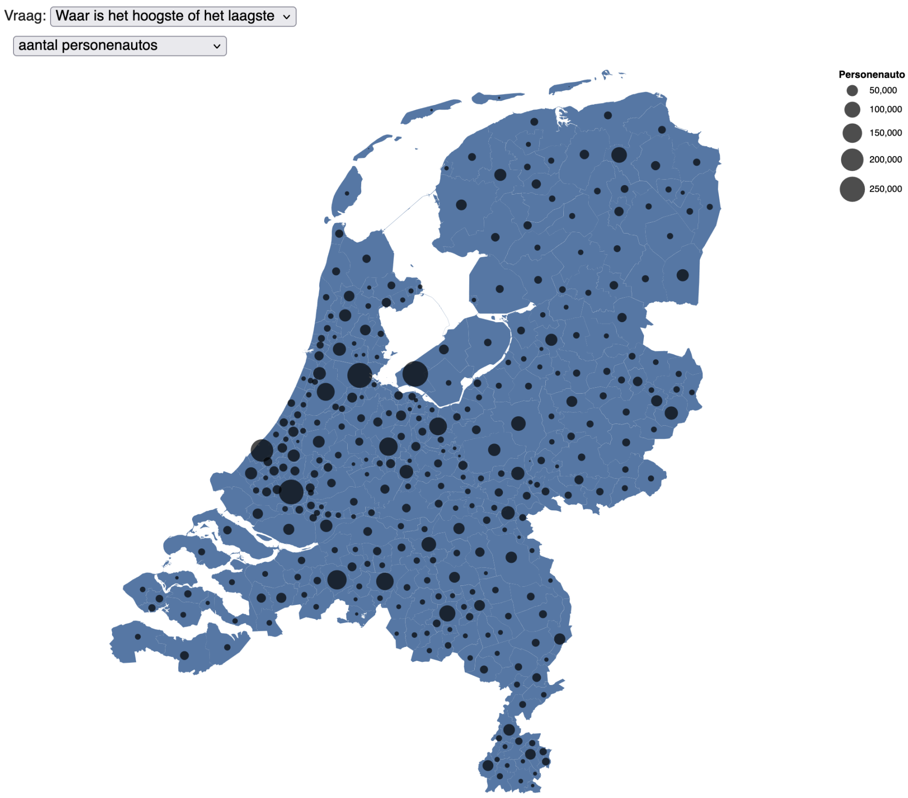


FIGURE D.6: Result: Where is the highest or lowest amount of cars?

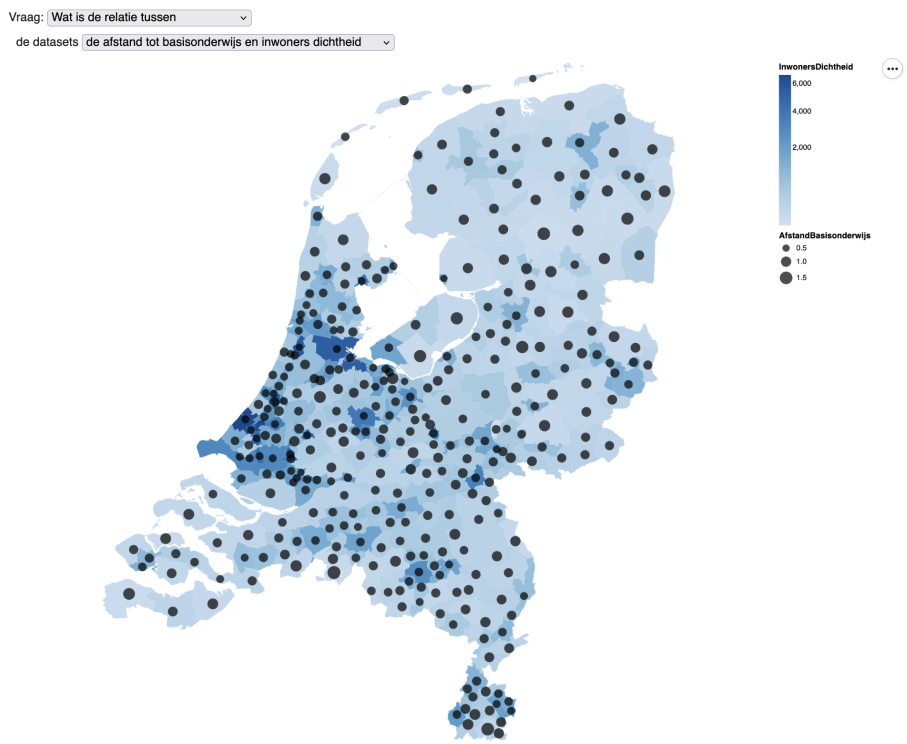


FIGURE D.7: Result: What is the relationship between the distance between primary schools and the population density?

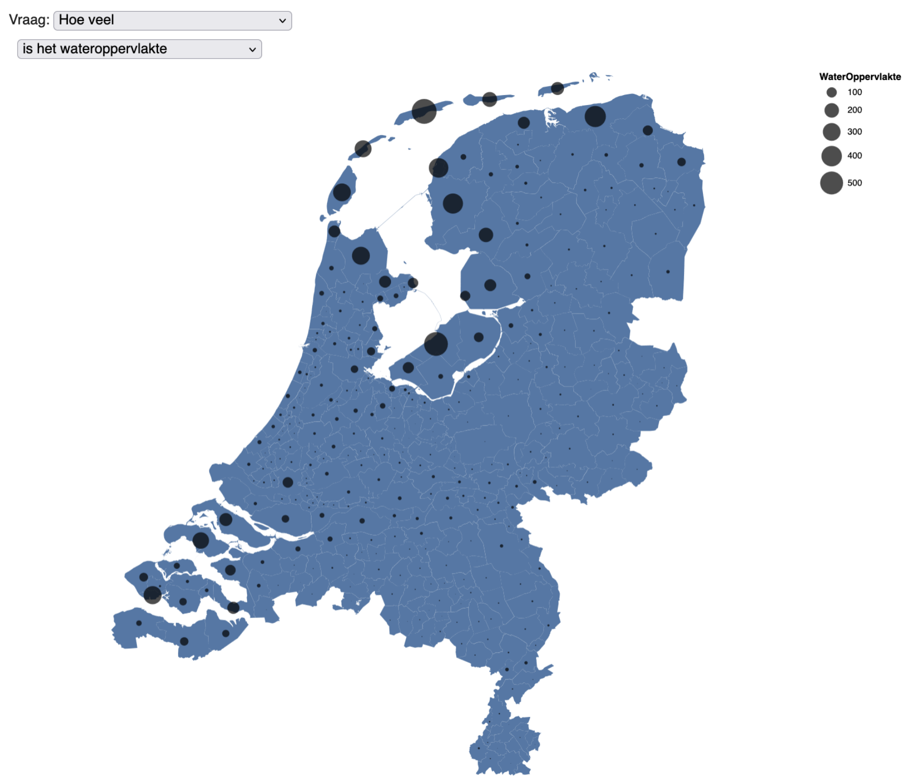


FIGURE D.8: Result: How much is the water surface in $km^2$?

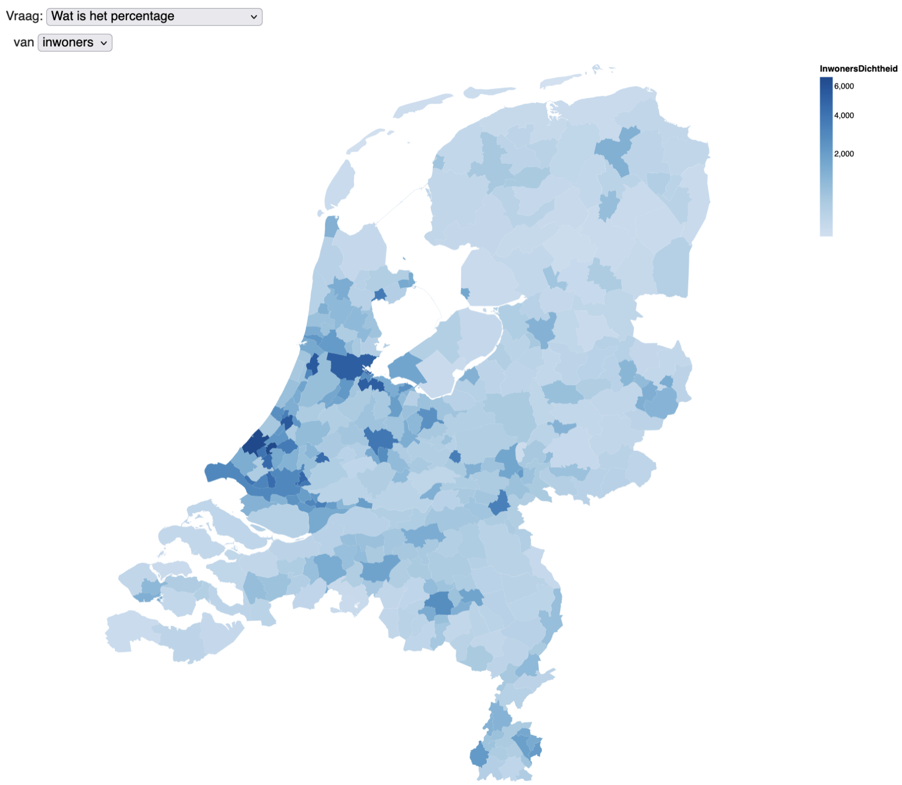


FIGURE D.9: Result: What is the percentage of inhabitants?

# Appendix E

# Counter-tool: Cijfers op de Kaart

Cijfers op de Kaart [8] is a website made by CBS, that visualises socio-economic data on maps. It is also the tool used as a counter-product during the usability tests, so that users could compare the tools. The figures below are screenshots of Cijfers op de Kaart. Figure E.1 is what a user sees first when loading in the website. If one would click "Theme selecteren" ("Select theme"), and select one of the theme options available, this results in, for example, an interface similar to figure E.2. If a user does not choose a location, it automatically zooms to middle of the Netherlands, but can be zoomed out again. A user can also search for a place, that result is showed in figure E.3. Text pops up under "Select theme", pointing towards the button.

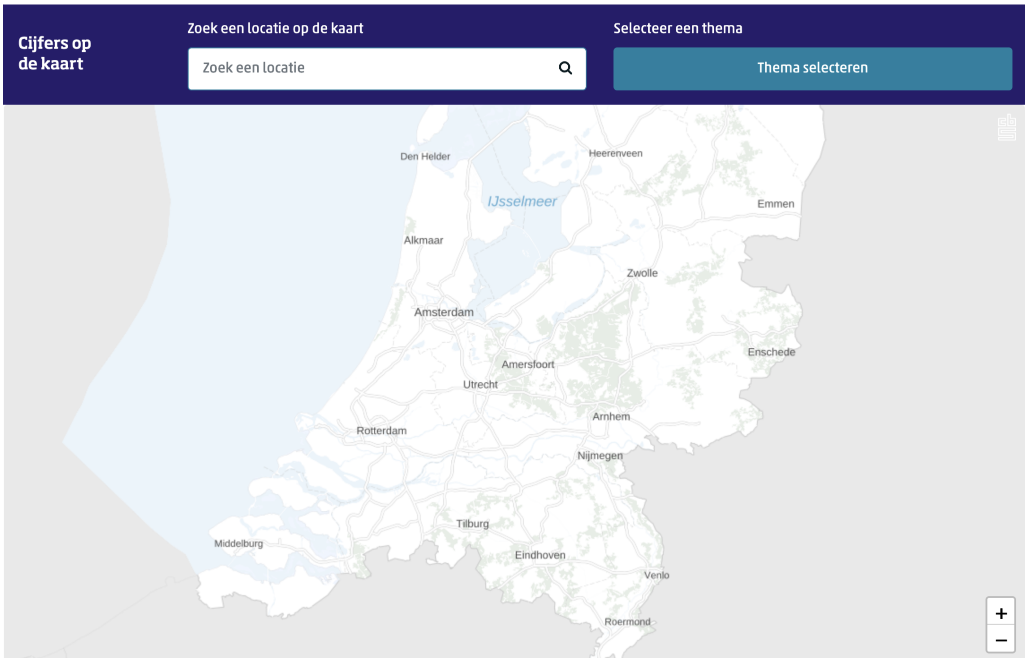


FIGURE E.1: Cijfers op de Kaart: First interface [8]

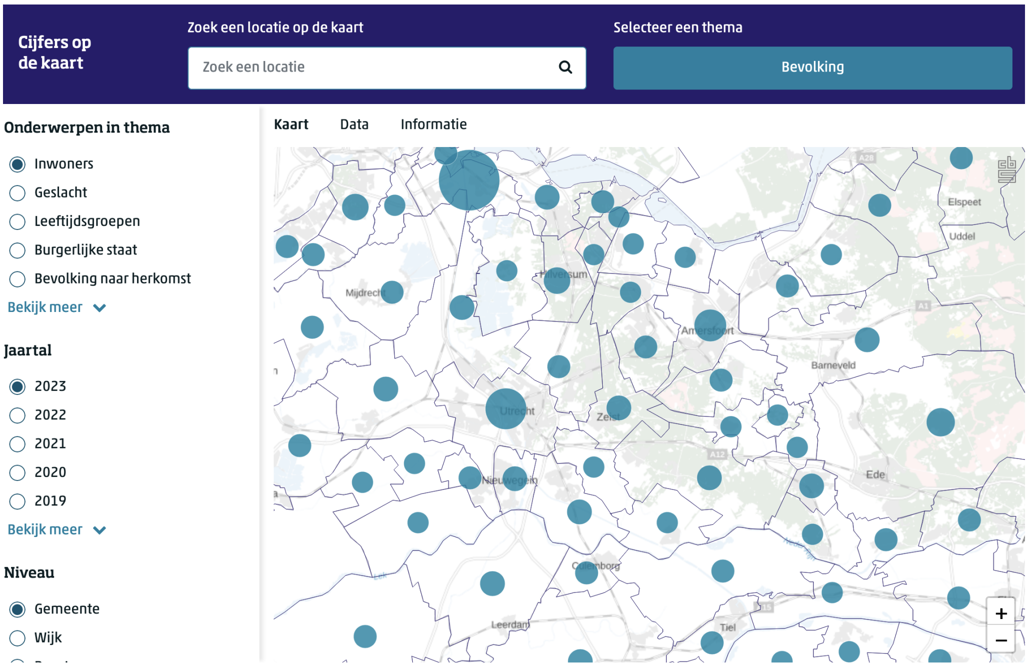


FIGURE E.2: Cijfers op de Kaart: When dataset is selected (Number of inhabitants) [8]

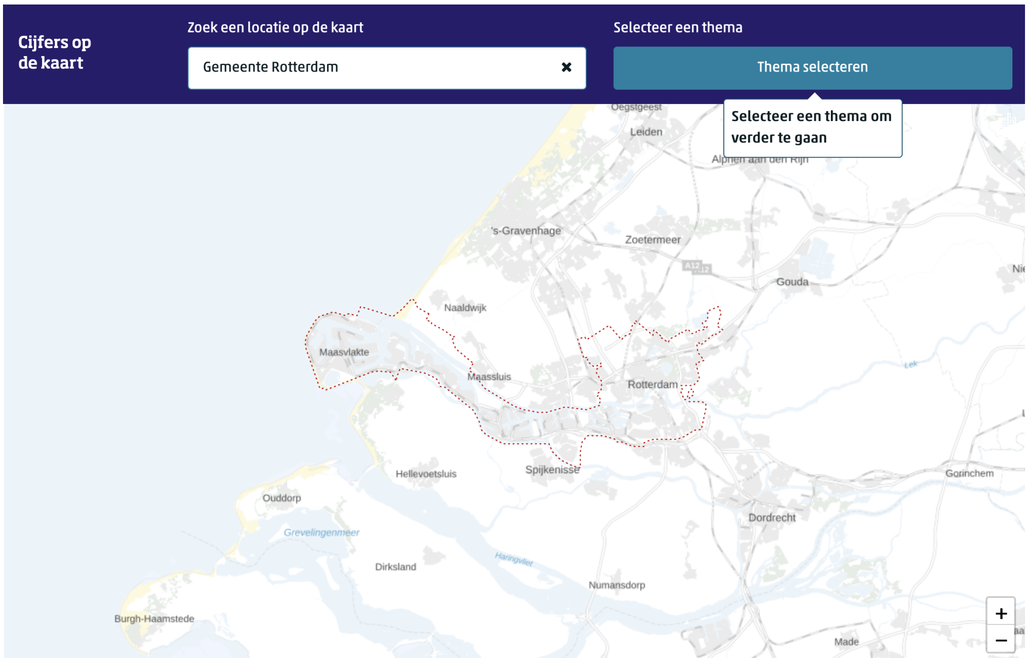


FIGURE E.3: Cijfers op de Kaart: When the location "Rotterdam" is searched [8]

# Appendix F

# Future work: Sorted tool

As presented in future work, the inclusion of a sorted list with data can be added to the visualisation. Vega-Lite does have the options for this to be developed, as can bee seen in figure F.1. A user can click on a municipality and the corresponding municipality lights up with the same colour in the bar chart. This also works the other way around, so that municipalities on the map can be found more easily. The tooltip functionality still works on this map as well. This page is just an example of a visualisation that shows the potential of this functionality and does not include the grammar based on questions yet.

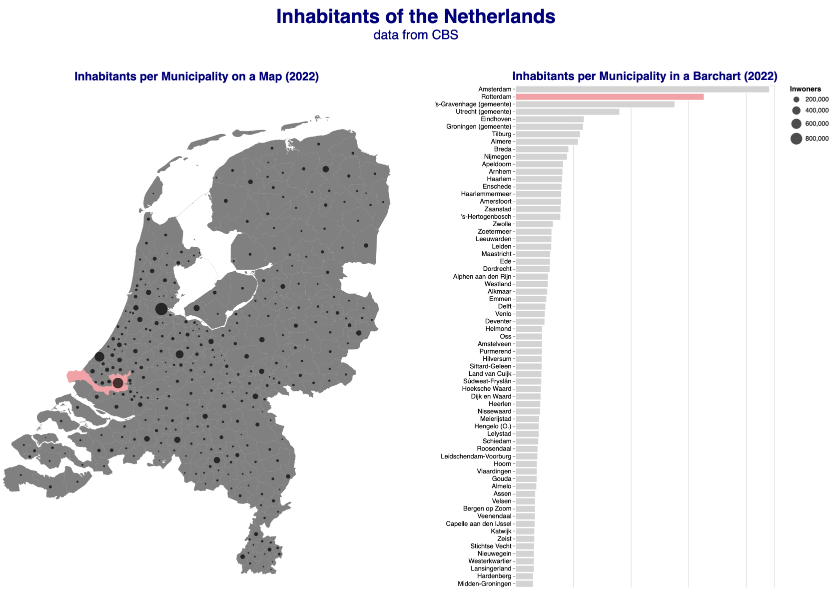


FIGURE F.1: Screenshot of the potential future tool with sorted data