DataVis Helper:
A Tool for Exploring the Design Space of Data Visualization

University of Twente

A Creative Technology Graduation Project

Supervised by Yuri Engelhardt and Barend Köbben

Written by Roseidys Primera

9th of August 2017

Abstract

This graduation project is about developing a tool that will encourage and support users in exploring the design space of data visualisation. Existing systems that help users in choosing the right type of visualisation have been analysed. The main problems encountered were that they are barely data-centred, the focus is on how the looks of the result and not what the data can visualise. Moreover, users are given a variety of options on how to visualise their data, but these tools do not guide users through the process. The idea is to develop a web tool that applies the principles of visual encoding. The design process for creative technology was used. Through several iterations, brain storm sessions and feedback sessions all the requirements and analyses for this research project were determined. A visualisation analysis has been performed. This analysis consisted of determining which principles of visual encoding fit each visualisation type. A final design of the web tool was created and this was tested with two groups, with experts such as data visualisation designers and universal participants such as students with little to no experience with data visualisation. The outcome was that such a tool help users to be aware of the abundance of visualisation types that exist since most participants were surprised by the number of suggestions of visualisations given. It has also been proven that by sketching users might understand their data better and that they can create new visualisation types that did not exist before. Therefore, the tests conclude that the tool did support and encourage users in exploring the design space of data visualisation. Further work is needed to improve the universal aspect of the tool. The tool is currently not applicable for complete novice users, they will be able to use it but will take a bit longer to grasp everything. Users with experience with data visualisation will have little to no problem with figuring out the ins and outs the web tool.

Acknowledgements

Firstly, I would like to thank my advisors; my supervisor, Yuri Engelhardt and my critical observer, Barend Köbben. I thank you, both for asking insightful questions, for all the advice on how to perfect my research project and for always being very enthusiastic and supporting when giving advice.

Secondly, I would like to thank my loving and supporting mother, Marlene Maduro for supporting me through the entire process.

Thirdly, I would like to thank the tool saviour and helper, Sebastiaan den Boer. I thank you for helping me out so much on the development of the tool; your expertise is impeccable.

Lastly, I would like to thank my supporting friends for always being there, Gabi, Birgit, Gelieke, Justin, Miriam, Hinke, Laura, Thérèse, Justin, Urney, Sade, Anja, Daan, Christiaan and everyone else that contributed in some way or another in supporting and helping me out.

Table of Contents

List of Figures

List of Tables

1. Introduction

The use of technology is rising day by day. We are now able to collect a lot of more data than we were used to years ago. 'Visualization is also one of the most important components of research presentation and communication due to its ability to synthesize large amounts of data into effective graphics'. [1] The possibility to learn from the data that is being gathered is even bigger due to the collection of data that is available. The act of knowing how to visualise data is becoming more important day by day, hence the relevance of this research project. Furthermore, there are plenty of tools such as catalogues consisting of many types of visualisations that can be used to represent data, but the downfall of most of these catalogues is that they are rarely data centred. The focus is on the functionality of the visualisation, for example; to show comparisons, relationships or hierarchy. Unfortunately, the focus is not on firstly analysing the data and then finding out what this data can visualise. Moreover, most 'data visualizers' such as experts, students, journalists in the field of data visualisation and everyday people who would like to visualise data tend to use the most convenient types of graphics, which introduces the goal of the research paper. The goal is to make these 'data visualizers' aware that there are many more options in the design space to how they can visualise and represent their data.

1.1 Problem Statement

The first problem as briefly mentioned previously, is that some catalogues and tools used for visualising data are limited when it comes to understanding which visualisation is used for which combination of data types. A possible solution is to analyse the design space of data visualisation by finding out which data types are suitable for which visualisations based on the fundamental principles of visual encoding. For that matter, the design space of data visualisation must be analysed and defined. This design space of data visualisation will be explored in this research project.

1.2 Goal

The primary aim of this research is to develop a web tool that will encourage, support and motivate users to use different visualisation types than just the most convenient ones. Users are mostly confused and overwhelmed by the number of visualisation types, and they do not know where to start since existing tools give users a bag of options full of visualisation types with little to no guidance. The aim is to also focus more on the data that the user wants to visualise. The tool will be data-centred. This goal will be achieved by using the principles of visual encoding, a forthcoming book by Clive Richards and Yuri Engelhardt about the principles of visual encoding.

1.3 Main Questions

The central research question discussed in this paper is 'How can we support and encourage end users in exploring the diversity of the data visualisation design space, in ways that are based on the fundamental principles of visual encoding?'. For answering the above question, research has to be made in five distinct areas; therefore the main research question has been broken down in five sub-questions, (1) 'what has been written and developed regarding the design space of possible types of data visualisations?', (2) 'what has been written and developed regarding making choices in the design process from data to the visualisation?', (3) 'how can the different types of data visualisations be analysed regarding the basic principles of visual encoding?', (4) 'how can all of the above be applied in a tool that can support and encourage users in exploring the design space of data visualisation?' and (5) 'what would be benefits and problems that users experience when using such a tool?'. By exploring these questions, the answers will be applied in the making and evaluating of a tool for users to explore the diversity of the design space of data visualisation and therefore answering the central research question.

2. Background

In this chapter, the background research is discussed. This chapter focusses on finding answers to the first two sub-questions, (1) 'what has been written and developed regarding the design space of possible types of data visualisations?' and (2) 'what has been written and developed regarding making choices in the design process from data to the visualisation?'. Firstly, this chapter will clarify what is meant by 'The Design Space of Data Visualization.' Secondly, the relevance and the application of 'The Basic Principles of Visual Encoding' are explained. Thirdly, a summary of the state of the art research concerning existing systems which help users explore the design space is discussed. Lastly, a possible solution to the problem is proposed.

2.1 The Design Space of Data Visualization

The central research question of this research paper revolves around 'The Design Space of Data Visualization.' However, what is this design space? In this section, the design space of data visualisation is defined.

The design space of data visualisation is defined differently by every data visualisation expert or any data visualisation user, novice or expert. Looking at the bigger picture, data visualisation experts agree that the design space is vast and undefined. 'The design space of possible vis idioms[1] is huge and includes the considerations of both how to create and how to interact with visual representation's as defined by Munzner. The design space of data visualisation is the space where design choices are made. What data the user sees, why the user intends to use a vis tool, and how the visual encoding and interaction idioms are constructed in terms of design choices as defined by Munzner.' [2] For this research, this same definition of the design space of data visualisation will be used. The design space of data visualisation is the process of finding out how to turn the data into a visualisation.

This definition is broad and can be broken down in three sections. Firstly, users can find out how to turn the data into a visualization by using data visualization helpers, such as catalogues that consist of a variety of visualization types. Secondly, users can sketch, to find out how they want to visually represent their data. Lastly, users can directly use data visualization tools, such as Tableau and Excel to visualise their data.

[1] Vis idioms are visualization types, such as the different types of charts available to visualise data

2.2 The Basic Principles of Visual Encoding

As mentioned in the introduction this research project will be applying the basic principles of visual encoding a forthcoming book by Clive Richards and Yuri Engelhardt. The basic principles of visual encoding are used to help users visualise their data. In this section, the basic principles of visual encoding will be explained.

The basic principles of visual encoding are principles which users can use to represent their data visually. These principles are grouped into five types, depicting, linking, grouping, ordering and scaling. There are 13 principles in total; picturing, mapping, connecting, nesting, grouping by enclosure (enclosing), grouping by proximity or alignment (spatial grouping), grouping by colour or shape, ordering by intensity, ranking based spatial ordering (attribute based spatial sequence), chronological spatial ordering (time-based spatial sequence), positioning on a time axis, positioning on an attribute axis and scaling by size.

‘Depicting encompasses both ‘Picturing’ and ‘Mapping.' Depicting is used to represent the visual appearance and/or spatial location of entities in the physical world (existing or imagined). These principles answer questions such as “What?” and “Where?”. Picturing is used to represent the visual appearance of objects or scenes. Picturing can be done by utilising representational methods such as perspective projection. Picturing is the key visual encoding principle used in technical illustrations and other pictorial representations. Mapping is used to represent the two-dimensional layout of physical configurations. Mapping can be done by utilising representational methods such as cartographic projection. Mapping is the key visual encoding principle in world maps, street maps and floor plans.’ [3]

‘Linking encompasses both ‘Connecting’ and ‘Nesting.' Linking is used to represent relationships between entities, such as connections, pathways, chronological order or hierarchies. These principles answers questions such as "Does a given relationship hold (between two entities)?". Connecting can be done by utilising lines, arrows, other joining devices between visual components, or by spatial alignment of visual components. Connecting is the key visual encoding principle in flow charts, family trees and network graphs. Nesting can be done by placing one visual component within another visual element. Nesting is the key visual encoding principle in 'tree map' and 'circle packing' visualisations.’ [3]

‘Grouping encompasses 'Grouping by enclosure,' 'Grouping by proximity or alignment' and 'Grouping by colour or shape'. Grouping is used to represent nominal attributes of entities, i.e. category membership. These principles answers questions such as "Which group?", "Which category?". Grouping by enclosure is the key visual encoding principle in Venn diagrams. Grouping by proximity or alignment is a visual encoding principle in tabular representations and in all diagrams that use spatial clustering. Grouping by colour or shape is a visual encoding principle in all diagrams that use colour coding or symbol coding.’ [3]

‘Ordering encompasses 'Ordering by intensity,' 'Ranking-based spatial ordering' and 'Chronological spatial ordering.' Ordering is used to represent ordinal attributes of entities, such as level, rank or chronological order. These principles answers questions such as "Which level?", "Which ranking?", "Which chronological order?". Ordering by intensity can be done by utilising a grey scale or brightness scale. Ordering by intensity is the key visual encoding principle in grey scale maps and maps with brightness gradients. Ranking-based spatial ordering is the key visual encoding principle in ordered lists and 'bump charts.' Chronological spatial ordering is a key visual encoding principle in comic strips and pictorial (graphic) instructions (in addition to 'Picturing’).’ [3]

‘Scaling encompasses 'Positioning on a time axis,' 'Positioning on an attribute axis' and 'Scaling by size.' Scaling is used to represent quantitative attributes of entities, such as amounts/numerical values or points in time. These principles answers questions such as "How many?", "How much?", "When?". Positioning on a time axis is the key visual encoding principle in timelines and clock faces. Positioning on an attribute axis is the key visual encoding principle in line charts and scatter plots. Scaling by size is the key visual encoding principle in bar charts and pie charts.’ [3]

The above-defined principles are applied in the project of this research paper. In chapter 5 the explanation will be elaborated on how these principles are implemented in the project.

2.3 State of the Art

In this chapter, the focus is on three main components. Firstly, on exploring existing systems such as, online tools, articles, blogs and research papers which help users in exploring the design space of data visualisation. Secondly, on describing the problem analysis by determining the limitations, disadvantages and improvements of the existing systems. Lastly, the focus of this research paper is elaborated.

2.3.1 Existing Systems

Background research has been done by finding out which online tools exist and that can help users in exploring the design space. This exploration has been done to obtain a better understanding of what has already been developed and is available when it comes to the design space of data visualisation. Furthermore, a description of the existing systems is given by stating their purpose, functionalities and characteristics. In total 15 existing systems have been examined: The Graphic Continuum, The Periodic Table of Visualization Methods, The Data Visualization Catalogue, The TimeViz Browser, A Classification of Chart Types, The Qualitative Chart Chooser, Chart Suggestion – A Thought-Starter, Chart Chooser,

Chart Chooser Cards, Infographic Taxonomy, The Visual Vocabulary, A Tour through the Visualization Zoo, Anychart-Chartopedia, Tableau and Microsoft Excel.

These 15 existing systems were chosen, because some were recommended by the supervisor of this research project, some were recommended during 'Data Visualization' courses given to University of Twente students studying Creative Technology and others were simply found by searching for 'how to choose the right type of chart' and 'data visualization catalogues'.

1. The Graphic Continuum

'The Graphic Continuum' is a platform which provides users multiple ways to help them in exploring the design space of visualisation, such as via the use of flash cards and posters which is based on the visualizations of the Data Visualization Catalogue. 'The Graphic Continuum' consist of a more comprehensive view of graphic types and how they are classified into different categories. The purpose of these tools is to help people decide how to choose the best graph for their data or to expose people to less common graphic types. The flashcards can be seen in figure 1. In figure 2 the poster version can be seen.

The Graphic Continuum flashcards are based on 'The Graphic Continuum' poster. The cards are categorised into six categories, namely (1) distribution, (2) time, (3) comparing categories, (4) geospatial, (5) part-to-whole and (6) relationships. At one side of the card a visualisation type is shown, and on the other hand, the category which the visualisation type belongs to is displayed.

The Graphic Continuum poster identifies some presentation methods, and it illustrates some of the connections that can bind different representations together. The use of the cards together with the poster will help users gain new insights since the poster also shows some relationships between categories and types of visualisations.

Figure 1: The Graphic Continuum - Flash Cards[2]

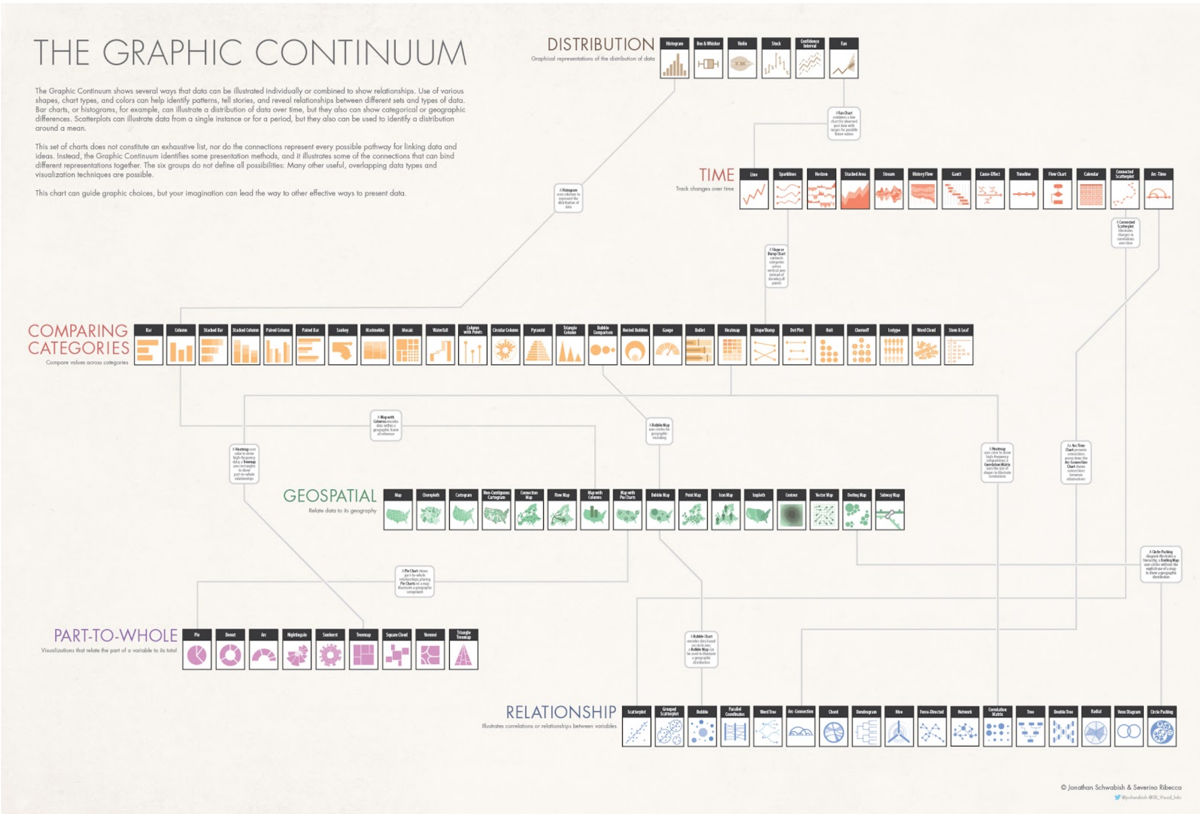


Figure 2: The Graphic Continuum - Poster[3]

[2] https://visual.ly/blog/graphic-continuum/
[3] https://policyviz.com/product/graphic-continuum-cards/

2. A Periodic Table of Visualization Methods

The periodic table of visualisation methods is an interactive overview of many visualisation types. The visualisations are categorised by different types of visualisation, such as (1) data visualisation, (2) information visualisation, (3) concept visualisation, (4) strategy visualisation, (5) metaphor visualisation and (6) compound visualisation which can be seen in figure 3. By hovering the mouse over 'an element' an example of the visualisation type will be shown. The purpose of this tool is to give a clear overview of many existing visualisation types. This tool does consist of many flaws. Firstly, the different forms of visualisation are not universal and do not give a proper description to when to use the visualisation types. Secondly, an image example of the visualisation type is provided, but the functionality of the visualisation types is not given. Thirdly, there is no useful organisation as mentioned by Few. 'What is the point of doing an entire research project to force a list of visualization methods into a paradigm that does not fit it? Apart from breaking the visualization methods into general categories (information visualization, concept visualization, etc.), this table exhibits no useful organization.' [4] The structure of the periodic table does not have an added value.

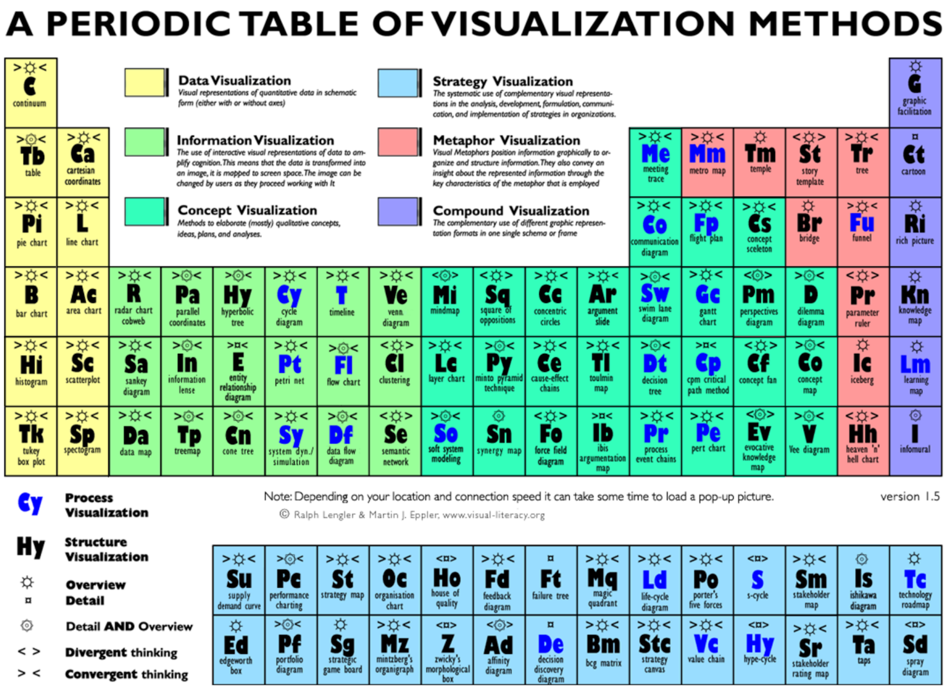


Figure 3: A Periodic Table of Visualization Methods[4]

[4] http://www.visual-literacy.org/periodic_table/periodic_table.html

3. The Data Visualization Catalogue

The Data Visualization Catalogue' is used for exploring the design space of possible data visualisation types. 'The Data Visualization Catalogue' is one of the most descriptive tools; It can be viewed in two ways, firstly by viewing all the types which can be seen in figure 4 and secondly by considering them by what the user wants to show which can be seen in figure 5. The first way lets the user click on one of the visualisation types, once clicked, a short description of the type is given, the anatomy of the visualisation, the functions, variations, examples and possible tools which can generate the visualisation. The second viewing way shows a list of functions, consisting of the functions; comparisons, proportions, relationships, hierarchy, concepts, location, part-to-a-whole, distribution, how things work, processes & methods, movement or flow, patterns, range, data over time, analysing text and reference tool.

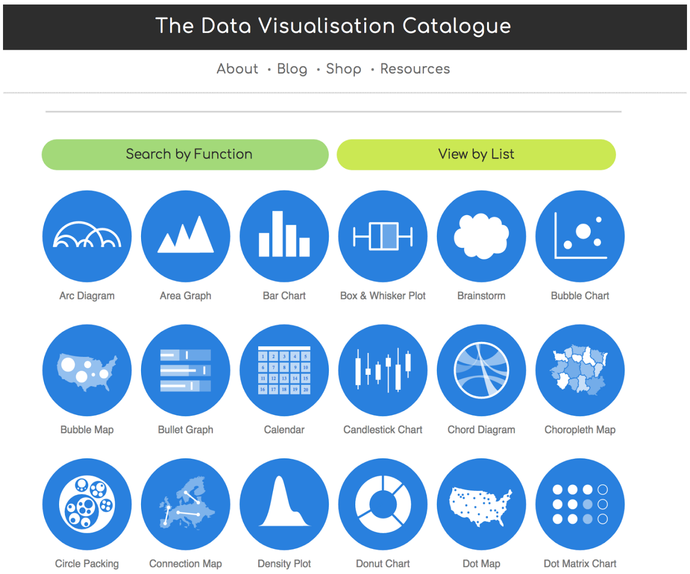


Figure 4: The Data Visualization Catalogue – Type View[5]

[5] http://www.datavizcatalogue.com/index.html

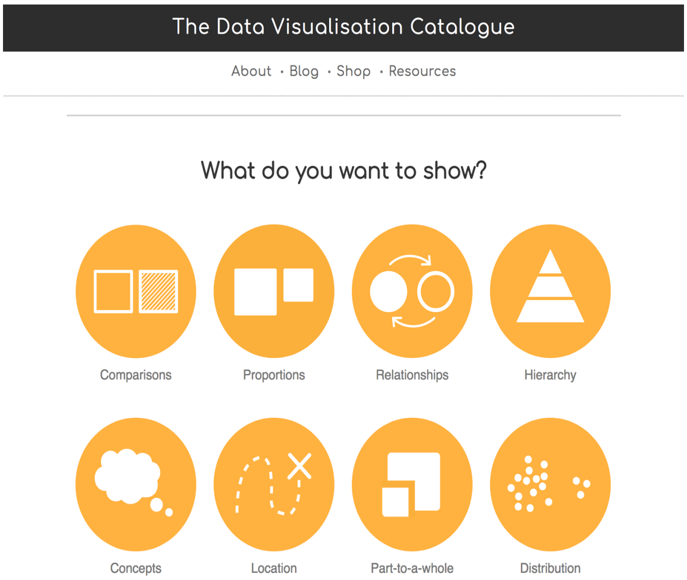


Figure 5: The Data Visualization Catalogue – Function View[6]

4. The TimeViz Browser

The TimeViz Browser' is a visual survey of visualisation techniques for time-oriented data which is illustrated in figure 6. It consists of 115 visualisation techniques, and it functions as a browser or a generator. The user selects what the attributes of his data are, such as (1) if the data is abstract or spatial, (2) if the data is univariate or multivariate, (3) if the time arrangement is linear or cyclic, (4) if the time primitives are instant or an interval, (5) if the visualization mapping is static or dynamic and (6) if the dimensionality of the visualization is 2D or 3D. The only two attributes which are purely data focused are the selections (1) if the data is abstract or spatial and (2) if it is univariate or multivariate. The other attributes focus on how the data will be visualised. Also, as stated in the name of this tool it is only focused on time-oriented data and this type of data is only a small representation of all the data types that exist.

[6] http://www.datavizcatalogue.com/search.html

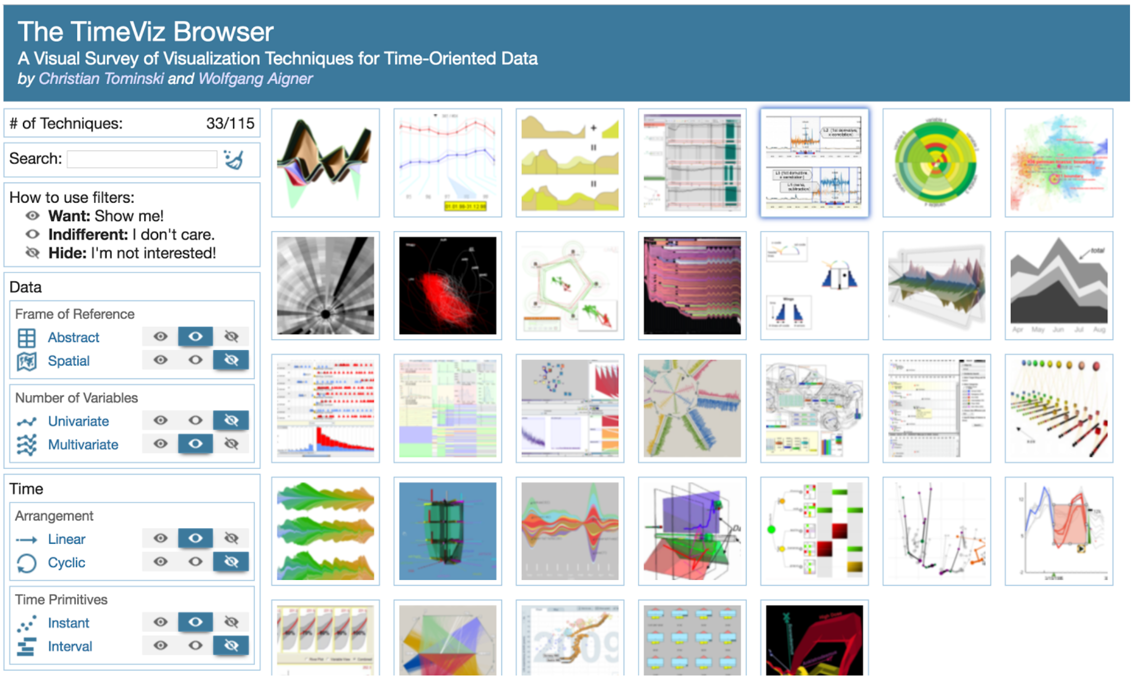


Figure 6: The TimeViz Browser[7]

5. A Classification of Chart Types

'A Classification of Chart Types' classifies the charts in two sections, (1) data comparison charts and (2) data reduction charts which can be seen in figure 7. The first section is then divided into three subsections, namely (1) comparison, (2) composition and (3) distribution. The second section is also divided into three subsections, specifically (1) evolution, (2) relationship and (3) profiling. This tool divides visualisation types once again into categories, which is only helpful to a certain extent.

[7] http://browser.timeviz.net/

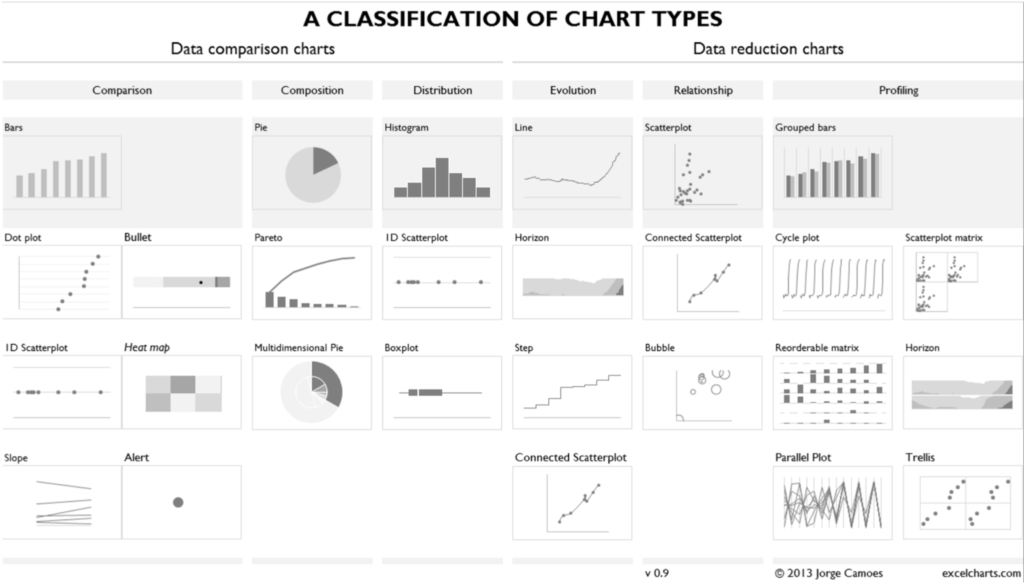


Figure 7: A Classification of Chart Types[8]

6. Qualitative Chart Chooser

Two versions of the 'Qualitative Chart Chooser' have been made. The 'Qualitative Chart Chooser 1.0' focuses on how to present the data, and the 'Qualitative Chart Chooser 2.0' concentrates on the story the user is trying to tell. 'Qualitative Chart Chooser 1.0' which can be seen in figure 8 consist of six categories; (1) how to highlight a person's comment, (2) how to align a goal or outcome, (3) how to show a process, (4) how to present themes, (5) how to show change over time and (6) how to display parts of a whole. 'Qualitative Chart Chooser 2.0' which can be seen in figure 9 consist of four categories; (1) diagrams, (2) pictures, (3) quotes & words and (4) tables. The use of these four categories is shown in two sections; the first section shows an overview of the possible visualisations and the second section answers the question 'What story are you trying to tell?' By giving possible answers corresponding with the visualisations. The answers provided are categories, namely (1) highlight a person's comment, (2) align with a goal or outcome, (3) show a process, (4) present themes, (5) show change over time, (6) display parts of a whole, (7) quantification, (8) exploratory and (9) explanatory. The first version did not show the relationship between the visualisations, where this has been added in the second version, where users can see the possibilities to combine visualisation types.

[8] https://excelcharts.com/classification-chart-types/

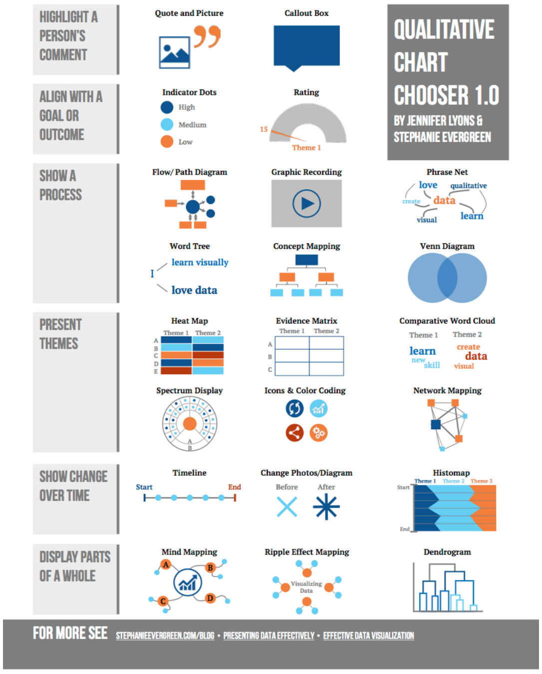


Figure 8: Qualitative Chart Chooser 1.0[9]

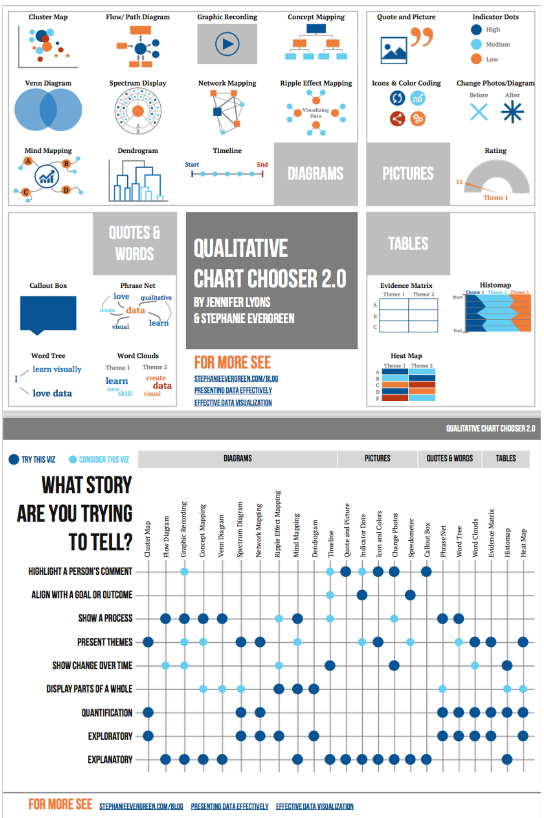


Figure 9: Qualitative Chart Chooser 2.0[10]

[9] http://stephanieevergreen.com/qualitative-chart-chooser/
[10] http://stephanieevergreen.com/qualitative-chart-chooser/

7. Chart Suggestion – A Thought-Starter

In comparison with other chart choosers, the 'Chart Suggestion' which is shown in figure 10 gives more detailed guidance for the user. It starts with the question 'What would you like to show?', the answer is then divided into four categories; (1) comparison, (2) distributions, (3) composition and (4) relationship. Both the category distribution and relationship are then divided into subsections depending on the number of variables the user wants to showcase. The category comparison is divided into two subsections; over time and among items. Lastly, the category composition is split into two subsections; changing over time and static. This chart chooser is a more detailed approach of knowing how to show data and is very descriptive and breaks down to the data, but its disadvantage is that it only gives the user the possibility to explore 20 visualisation types.

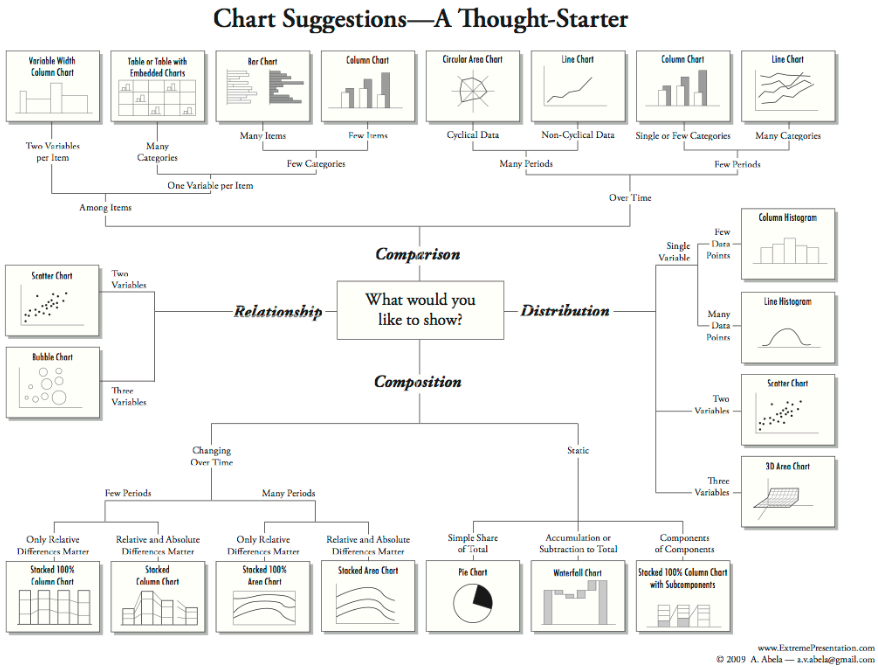


Figure 10: Chart Suggestion – A Thought-Starter[11]

8. Chart Chooser

The 'Chart Chooser' which is shown in figure 11 aims to help users find the right chart type for their needs; the visualisations can then be downloaded as an Excel or PowerPoint template. The central question asked is 'What type of visualization do you need for your data?', the user can then choose several options, such as (1) all, which shows all the visualizations and when the mouse hovers over a visualization, the function(s) of the visualization is shown, (2) comparison, (3) distribution, (4) composition, (5) trend, (6) relationship and (7) table, are the six categories which the visualizations are placed in. The advantage of this tool is that the user can choose multiple categories at once, so this results in selecting, for example, the category

[11] http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf

comparison and distribution which results that the tool gives visualisation types which can show both categories. This tool provides, therefore, the relationship between categories.

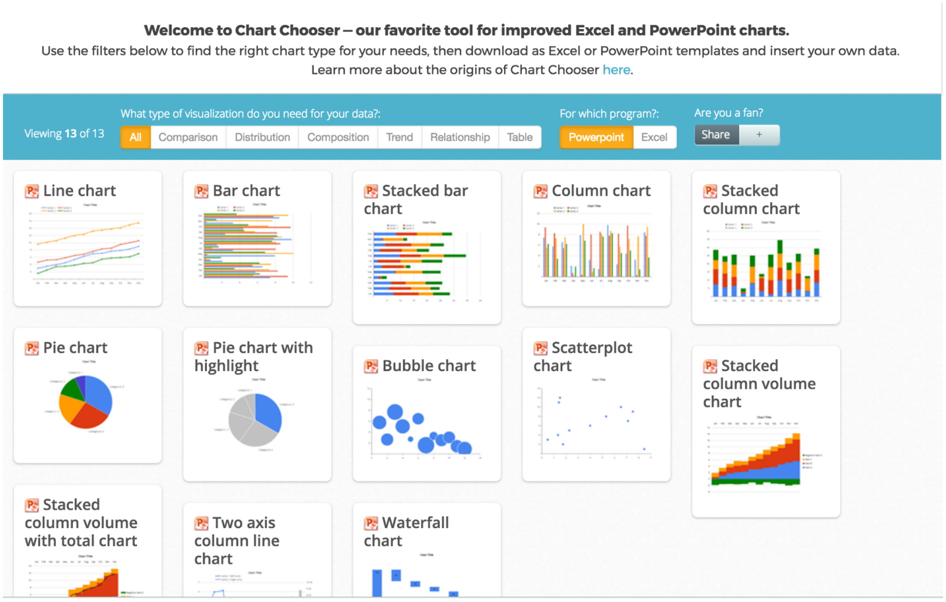


Figure 11: Chart Chooser[12]

9. Chart Chooser Cards

'Chart Chooser Cards' which can be seen in figure 12, are meant to be simple. They help the user choose the best type of chart. The aim of 'Chart Chooser Cards' is to cut through the data noise to help users decide which graph type is best and which chart sells their story or message. As shown in figure 13, on each card the common name of the chart type is shown, along with a description, a visual example, when it is used and what type of data set it is best for. As can be seen in Figure 14, the chart types are organised into six categories; (1) small numbers, percentages, frequencies, (2) time, (3) survey responses, (4) comparisons, (5) place and (6) it is complicated. This tool is very descriptive and falls parallel in the same category as 'The Data Visualization Catalogue.' There are both descriptive when it comes to the visualisation types. The disadvantage here again is that the combinations of categories are not given.

[12] http://labs.juiceanalytics.com/chartchooser/index.html

Figure 12: Chart Chooser Cards[13]

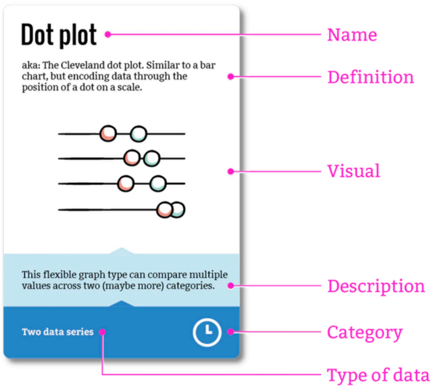


Figure 13: Chart Chooser Cards – Description[14]

Figure 14: Chart Chooser Cards – Categories[15]

[13] https://www.kickstarter.com/projects/99144298/data-visualization-chart-chooser-cards
[14] https://www.kickstarter.com/projects/99144298/data-visualization-chart-chooser-cards
[15] https://www.kickstarter.com/projects/99144298/data-visualization-chart-chooser-cards

10. Infographic Taxonomy

'Infographic Taxonomy' which can be seen in figure 15, focuses on the importance of the function of infographics. Stating that infographics are not just about picking a random visualisation; the result varies depending on the structure of the data and the (combinations of) questions. 'Infographic Taxonomy' consist of 6 main questions which are also paired with each other. The questions are, (1) who/which is involved, (2) where is it, (3) when did it happen, (4) what is it about, (5) how/why does it work and (6) how much is it. The focus is primarily on the interrogative words such as, who/which, where, when, what, how/why, how much. This 'Infographic Taxonomy' focuses therefore on the relationship and combination between the questions, which in this case are the categories. Similarly, to the 'Chart Suggestions - A Thought-Starter,' the Infographic Taxonomy is very data focused, but its disadvantage is that only 20 visualisations are represented.

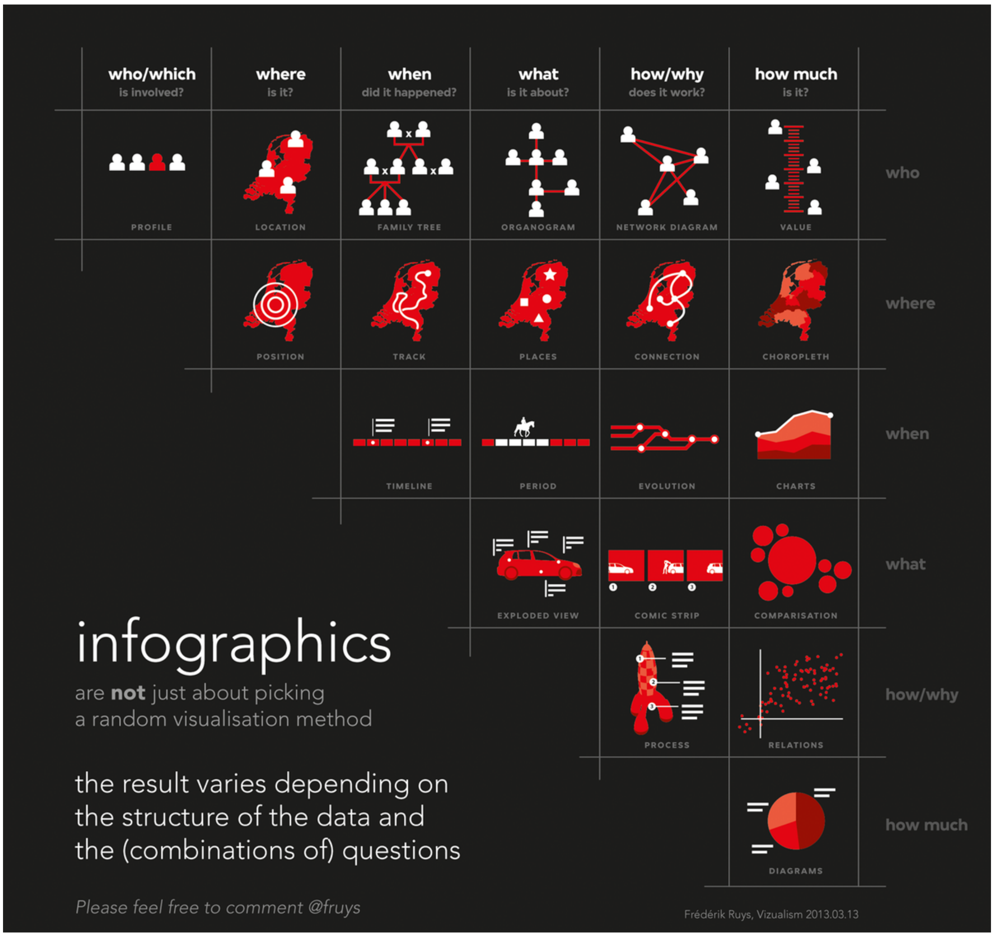


Figure 15: Infographic Taxonomy[16]

---

[16] https://www.vizualism.nl/infographic-taxonomy/

11. Visual Vocabulary

The 'Visual Vocabulary' which is shown in figure 16, aims at users making an informative and meaningful data visualisation by deciding which data relationship is most important in the user's story, then look at the different types of the chart within the category to form some initial ideas about what might work best. The 'Visual Vocabulary' has an interactive online version and a static offline version in the shape of a poster which can be seen in figure 16. It consists of nine categories, namely (1) deviation, (2) correlation, (3) ranking, (4) distribution, (5) change over time, (6) magnitude, (7) part-to-whole, (8) spatial and (9) flow. The 'Visual Vocabulary' also gives a short description of each category and examples of when the category is used. Users get a broad overview of all the visualisation types available to use, but the combinations of the categories and relationships between the visualisations are not given.

Figure 16: Visual Vocabulary[17]

[17] http://ft-interactive.github.io/visual-vocabulary/

12. A Tour through the Visualization Zoo

'A Tour through the Visualization Zoo' is an extensive article that provides a brief tour through the "visualisation zoo," showcasing techniques for visualising and interacting with diverse data sets. Here the focus is on a few of the more sophisticated and unusual techniques that deal with complex data sets. Analogously, some of the more exotic forms of visual data representations are covered, starting with one of the most common, (1) time-series data; continuing to (2) statistical data and (3) maps, and then completing the tour with (4) hierarchies and (5) networks. Most of the visualisations are also accompanied by interactive examples. 'The article concludes that the DNA underlying all visualisations remains the same: the principled mapping of data variables to visual features such as position, size, shape, and colour.' [5] In figure 17, a sneak peek of 'A Tour through the Visualization Zoo' is shown, each section is structured in the same way. First, the visualisation is described, and then an interactive example is provided, where the reader can experience the description in a 'live' environment within a context.

The article is focused on very unusual visualisation types, and it does an excellent job at describing each type and its functionalities, but as a disadvantage only describes a small scope of the entire zoo of visualisations.

Figure 17: A Tour through the Visualization Zoo – Sneak Peak

13. AnyChart- Chartopedia

Chartopedia calls himself 'your guide for choosing the right chart types', this tool can be seen in figure 18. Chartopedia has two view function one where the user can see all the 64 different chart types and one where the chart types are grouped by usage with the categories: (1) comparison, (2) data over time, (3) range, (4) distribution, (5) proportion, (6) location, (7) part of the whole and (8) finance. All types mentioned in the catalogue are supported by JavaScript charting libraries which means all these types can be made using JavaScript. Chartopedia focuses on grouping the types into different types of categories of usage. This catalogue also does not focus on the data, but on how you will want to use the chart type.

Figure 18: AnyChart - Chartopedia[18]

14. Tableau

A quick overview of Tableau can be seen in figure 19 and 20. Tableau is a data visualisation tool where users can directly import their data and start visualising by dragging and dropping their data types. Tableau can be used to get an idea of how to represent data visually quickly. Tableau has two main disadvantages it does not guide a user through the process, unless the user searches for an online tutorial. Secondly, only 24 visualisation types are advised to the "Show Me" table that the user can use to visualise his data.

[18] http://www.anychart.com/chartopedia/

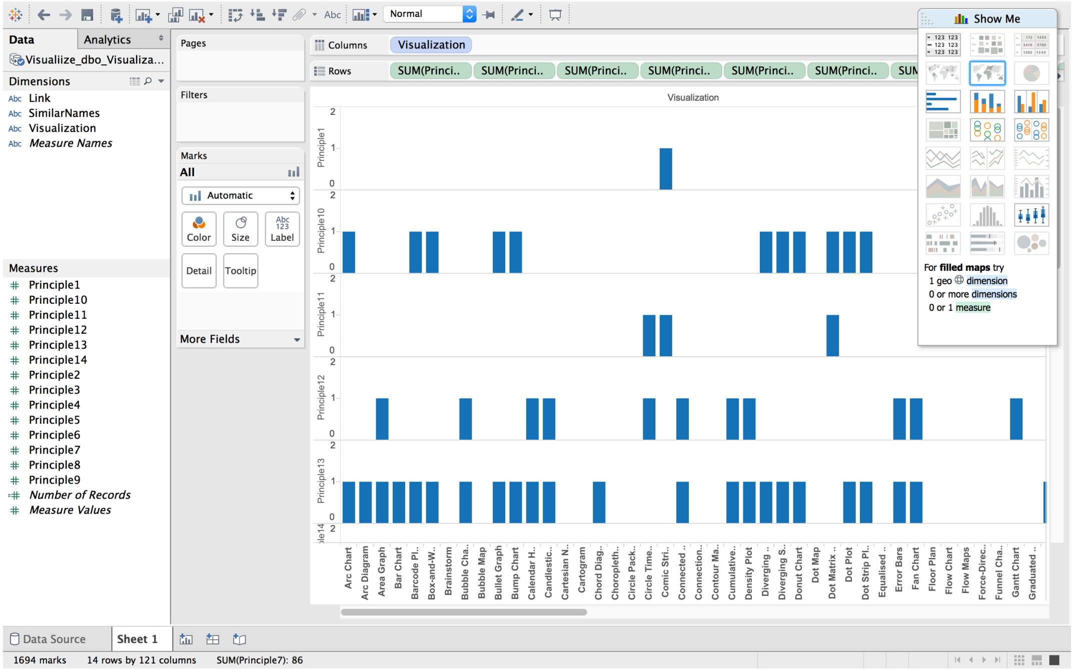


Figure 19: Tableau Interface[19]

Figure 20: Tableau Visualization Possibilities[20]

[19] https://www.tableau.com/
[20] https://www.tableau.com/

15. Microsoft Excel

Microsoft has very similar features as Tableau. The data can be directly visualised by choosing one of the options provided by Excel. The Microsoft Excel interface and visualisation examples can be seen in figure 21 and 22. Microsoft Excel is not as intuitive as Tableau. Tableau uses the 'drag and drop' functionality, by using excel the user must select the data he would like to visualise, but this is not informed in advance, the user must either search for it or figure it out by himself. Also, Excel provides the option to represent data in 35 different visualisation types visually.

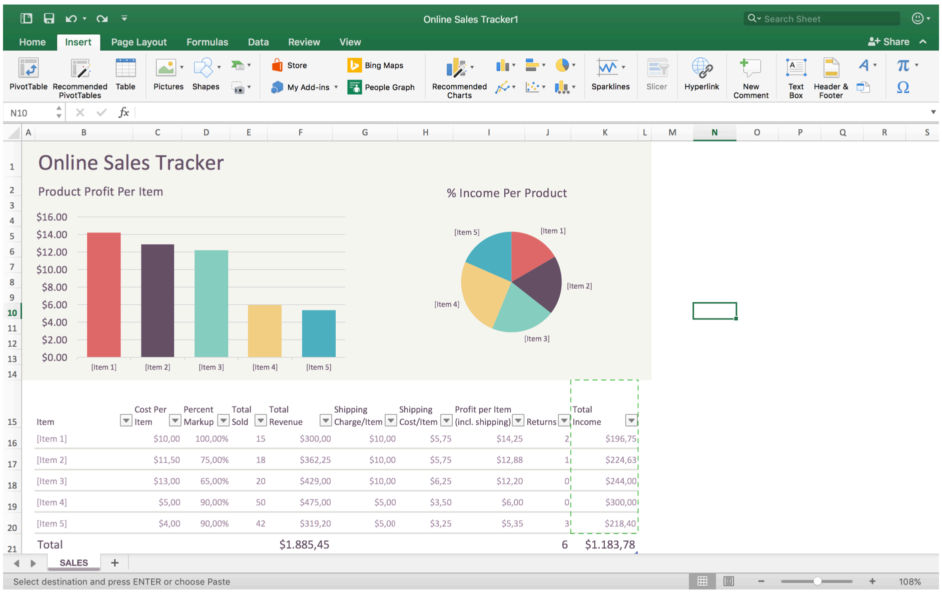


Figure 21: Microsoft Excel Interface

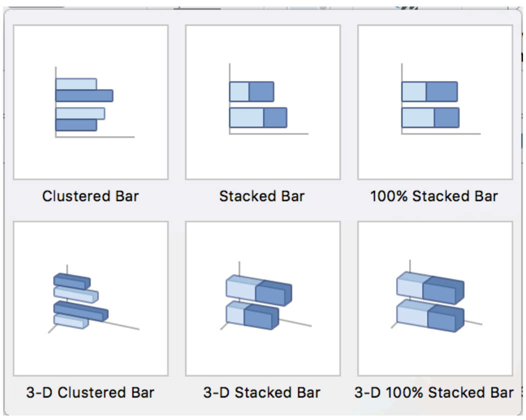


Figure 22: Example of Visualization Types of Microsoft Excel

2.3.2 Problem Analysis

There is now a clear overview of what is available to use when it comes to exploring the design space of data visualisation. In this section, the limitations, disadvantages for the user and improvements of the existing tools will be discussed.

2.3.3 Limitations and Disadvantages

Limitations and disadvantages are considered factors which limit the user from exploring the complete design space of data visualisation and factors which restrict the user to understand how to represent his data visually fully. While going through the functionalities, purpose and description of all the tools mentioned above, some problems and limitations kept emerging. This list was gathered by picking all the functionalities of each existing system and putting them in one list, since some systems offer other functionalities than other, if you both all the functionalities together that would make a tool that will encourage the user to explore the complete design space.

Some limitations of these systems are:

1. The relationship between visualisation types is not always given; this also includes showing different variations of visualisation types with similar functionalities.
2. A clear description of the categories in which the visualisations are categorised is not always present.
3. Description of the visualisation types is not always present.
4. Mentioning in which tool the visualisation can be made is not always present.
5. The relationship between categories is not always given.
6. The definition of categories is not always given.
7. The possibility to combine categories is not always present.
8. Some visualisation support 20 visualisations, other 100 or more visualisations.
9. The focus is mostly on what the user wants to show, instead of what the data can show. Therefore, these tools are not always data-centred, they are visualization-centred.
10. The tool does not guide the user through the process on how to get from data to visualisation.
11. The tool does not encourage creativity or the freedom to explore new types of visualisation types or least known visualisation types.

A matrix has been made to get an overview of which tools have which limitations, that can be seen in Figure 23. The red boxes indicate that the tool completely consists of the limitation, orange shows that it partly has the limitation, green shows that it does not have the restriction.

			Limitations										
			1	2	3	4	5	6	7	8	9	10	11
Existing Tools	1	The Graphic Continuum								87			
	2	A Periodic Table of Vis. Methods								100			
	3	The Data Visualization Catalogue								60			
	4	The TimeViz Browser								115			
	5	A Classification of Chart Types								27			
	6	Qualitative Chart Chooser								23			
	7	Chart Suggestion								21			
	8	Chart Chooser								30			
	9	Chart Chooser Cards								46			
	10	Infographic Taxonomy								21			
	11	The Visual Vocabulary								70			
	12	A Tour through the Visualization Zoo								20			
	13	Anychart-Chartopedia								64			
	14	Tableau								24			
	15	Excel								35			

Figure 23: Limitation Matrix

2.3.4 Improvements

As can be seen in figure 23, the tools with the least limitations are: 1) The Data Visualization Catalogue, 2)) The Visual Vocabulary, 6) A Tour through The Visualization Zoo and 7) Anychart-Chartopedia.

The main improvements for the above-explored tools are; (1) to show the combinations and relationship between visualization types, (2) to describe the categories, if the visualization types are categorized, (3) defining the visualization type, by explaining its function and also giving a visual example, (4) mentioning in which tool such a visualization can be made, (5) showing variations of the same type of visualization, (6) to focus on what the data consist of and what can be demonstrated which this data, (7) to guide the user through the entire process and (8) to encourage creativity and freedom of exploration, to come up with new visualization types and come across least known types.

These main improvements were chosen by looking at Figure 23, the 'Limitation Matrix'. The boxes that where the most red, which means that they have the limitation. These limitations or disadvantages can be improved by applying the opposite. For example, point 11 consist of the limitation: the tool does not encourage creativity or the freedom to explore new types of visualisation types or least known visualisation types. So, the improvement would be to make sure the tool encourages creativity and freedom of exploration of the design space of data visualization.

2.3.5 Focus

The focus and the main problem which can be observed in all the existing systems is that the focus is not on the data, but on the result, the visualisation. The step to analyse the data to get to a visualisation is missing. Therefore, the focus of this research project is to make an interactive tool for exploring the design space of data visualisation by applying the principles of visual encoding to take that step from data to visualisation. This tool will focus on the data which will be visualised; what are the attributes of the data, that is the focal point and how to visually represent these data attributes.

2.4 A Solution

As mentioned in the previous section the focus of this research project is to make an interactive tool for exploring the design space of data visualisation by applying the principles of visual encoding from data to visualisation. In this section, the concept, application area and goal of this idea will be elaborated.

2.4.1 The Concept

The tool will be a suggestion web tool. It will help the user explore the design space by starting with his data and potentially ending up with ideas on how to visualise his data. In chapter five and six the concept will evolve into a prototype, therefore in those chapters, the tool will be explained more elaborately. Here a brief idea of the tool will be given. The user will be guided throughout the entire process. Therefore, the process will be divided into steps: Step 1) Analyse the dataset, 2) Define the data types, 3) Connecting the data types to the principles of visual encoding, 4) Sketching, here the user will get the freedom to creatively apply the principles in any way he or she wants to and 5) The user will get suggestions on which visualization types to use that fit the data types and visual encodings. The idea is that the tool is clean, straightforward and easy to use.

2.4.2 Application Area and Goal

The developer of the tool which is the author of this research project realised that there could be misunderstandings on what the difference is between the intended tool and existing data visualisation tools. There are a few data visualisation designers, who use data visualisation tools for the exploration of the design space. In chapter 4, the difference between data visualisation tools and this tool will be explored in more detail.

The goal of this tool is to encourage and motivate users to explore the design space by sketching and by finding out about least known visualisation types. Furthermore, to make users aware of the abundance of visualisation types that exist and that can be used.

The application of this tool is universal; this tool will have a similar function as a calculator, the tool helps the user with the exploration of visualisation types. The user decides when, how and why to use it. Anyone that has access to the internet can make use of the tool. The tool will not only be used by data visualisation experts, but it is highly focused also to be used by novices, such as students who would like to represent a data set visually and who need inspiration, since they might not be aware of all the possibilities.

3. Methods and Techniques

In this chapter, the methods and techniques used to realise this research project are elaborated. The creative technology design process used as the red line throughout this research project. 'The process of starting with the ideation, then the specification, the realisation and finally the evaluation.' [6] In this section, the stakeholders, the tool requirement analysis and the evaluation procedure will be discussed.

3.1 Stakeholders

In this section, everyone who is affected by the web tool, who have power over it or who have an interest in its successful or unsuccessful development will be discussed. Thus, the stakeholders will be discussed. The stakeholders will be divided into four groups: (1) Influencers, (2) Advisors, (3) Users and (4) End Users.

The Influencers are the people who have a direct influence or power over the web tool and research project. A direct influence in the sense of that they can potentially change the look, the feel, and functionality of the tool. The developer of the tool and the supervisor of the research project are known to be 'the main influencers.' These two stakeholders have the power to determine the functionalities of the tool and the look-&-feel of it as well. The developer makes all the necessary and final decisions in collaboration with the supervisor.

The Advisors are people who have an interest in its successful or unsuccessful development, but also might influence the functionality and look-&-feel of the tool. Moreover, they also might be affected by the tool. Advisors are the critical observer of this research project and experts in the field of data visualisation, which can be teachers of data visualisation and data visualisation designers. The Advisors do not have a direct influence on the development of the tool, they are as the name suggests, advisors. They have the freedom and choice to give their opinion on the web tool, and the developer decides whether he takes their advice into account or not. They are data visualisation experts and designers. Therefore they are also affected by the tool, so they will have the freedom to creatively use the tool whenever they need to explore how to visualise their data set.

The Users can be anyone between the ages of 14 and up to as old as time. The users can have any occupation and any background. They must be able to have access to the internet, to know how to work with an online web tool and have knowledge about datasets. These Users are affected by the tool since they are the ones who will use it.

The End-Users are the people who will see the end visualisation which the Users make. Their impact on the tool itself is not significant. However, it can happen that the Users advice the End-Users to use the tool, and then they also become Users.

Thus, one of the main stakeholders of the research projects is 'The Users' that will use the tool. These users can be both novices to the field of data visualisation and data visualisation enthusiasts.

3.2 Tool Requirements Analysis

The tool must fulfil some requirements before development starts. The requirements are divided into three groups: (1) Main Requirements, (2) Desired Requirements and (3) Additional Requirements.

The Main Requirements are that the tool must be available online and easily accessible to anyone with access to the internet. Therefore, the performance of the web tool must be reasonably good. It will be a web tool; therefore, it is expected to work on any mobile or computer device.

The Desired Requirements are based on the software that will be applied to develop the tool. All software used is desired to be free and simple to use, to showcase that students with any background can be able to develop a web tool. Therefore, 'Visual Studio Code' is used for the frontend and backend development, due to its working collaboration with 'Azure,' a free cloud computing service for students to develop their web applications. 'Data Grip' is used for the management of the database and 'Bootstrap Studio' to make the interface of the web tool.

The Additional Requirements are needed to make the entire process of the development of the tool smoother and easier. Therefore, 'GitHub' is used for the source code management.

A more detailed approach on how to the tool is made will be given in chapter 5, the chapter on the specification of this research project.

3.3 Evaluation Procedure

For the evaluation of the research project, two principal evaluations have been used. Mainly, the so-called expert evaluation and a universal testing.

Expert evaluations, expert feedback and brainstorm sessions have been performed throughout the entire process of this research project. Each week there was a feedback session with the supervisor since each week there was progress on the development of the tool. After each session, the feedback given was applied to the tool, to improve it before the universal testings began. In these cases, the expert was the supervisor of this research project, a data visualisation expert and enthusiast, also the co-author of the forthcoming work 'The Fundamental Principles of Visual Encoding', Yuri Engelhardt. The feedback from two known Dutch Data Visualization Designers was also considered during the testing of the prototype, namely Nadieh Bremer and Jan Willem Tulp.

Universal testings are tests which were performed with potential users of the web tool. The participants had a diverse background with a bit of experience in visualising data. The testings were focused on universal users to test whether this tool can be used by anyone with any background and any age between 14 and infinity. Both types of evaluations will be further discussed in chapter 6, the chapter on realisation.

4. Ideation

In this chapter, the analyses for the development of the envisioned product will be discussed. The idea of the envisioned web tool will be elaborated throughout the following chapters. Firstly, the application of the principles of visual encoding will be explained. Secondly, the visualisation analysis by using the principles will be elaborated, which explores the answer to the third sub question 'How can the different types of data visualisations be analysed regarding the basic principles of visual encoding?'. Thirdly, the idea of the envisioned tool will be compared with existing visualisation tools. Lastly, the final idea for the envisioned tool will be discussed.

4.1 Application of Principles of Visual Encoding

In chapter 2, section 2.2, the principles of visual encoding were explained. In this section, the application of the principles of visual encoding will be explained. In other words, how the principles will be used and applied in the tool will be elaborated.

The principles of visual encoding will be applied twice in the process of the tool. Firstly, they will be applied by providing the user with the principles which are fitting to his data type, an illustration of this process can be seen in Table 1. This will be further explained in detail in section 4.4. Secondly, with the help of the visualisation analysis, the principles of visual encoding will be the connecting piece to the data types and the existing visualisation types. This will be explained in detail in section 4.2.

Data Type	**Principles**
Visual Appearance and/or Spatial Location	Picturing
	Mapping
Relationship	Connecting
	Nesting
Nominal Attribute	Grouping by enclosure
	Grouping by proximity
	Grouping by colour or shape
Ordinal Attribute	Ordering by intensity

	Ranking-based spatial ordering
	Chronological spatial ordering
Quantitative Attribute	Positioning on a time axis
	Positioning on an attribute axis
	Scaling by size

Table 1: Data Type to Principle

4.2 Visualization Analysis

The principles of visual encoding play a significant role in this research project. The goal is to advise users on which visualisation types to use that fit their data, to achieve this the basic principles are used. Each data type can be represented by several principles as mentioned in section 2.2. Each visualisation type consists of a combination of the principles of visual encoding. This is considered the visualisation analysis.

The visualisation analysis consists of analysing all the existing visualisation types. In Chapter 2, existing tools and catalogues have been analysed, and the ones who had the least limitations were (1) The Data Visualization Catalogue, (2) The Visual Vocabulary, (3) A Tour through The Visualization Zoo and (4) Anychart–Chartopedia.

The analysis consisted of three steps. First; all the existing visualisations of the catalogues were gathered, which were in total 220 visualisations. Second; all the similar visualisations were grouped, and duplicates were taken out of the analysis. This concluded with the analysis containing a total of 125 unique existing visualisations. Third; a matrix was made to analyse each visualisation with the basic principles of visual encoding. In other words, examining which principles are used in each visualisation. For example, a line chart consists of applying the principles 'Positioning on an attribute axis' and 'Positioning on a time-axis'. In Appendix A and Appendix B, the procedure of this analysis can be seen. This analysis will be used as part of the database of the tool, to retrieve the data that is needed to give the user suggestions on which visualisation types to use.

4.3 The Idea of the Tool vs. Existing Data Visualization Tools

In this section, the focus is on clearly explaining the difference between the functionalities and purpose of the envisioned tool and existing data visualisation tools. This analysis compares the data visualisation tool 'Tableau' with the idea of the envisioned tool.

Tableau is used by some data visualisation designers as an exploration tool; to quickly place the data and get an overview of what can be visualised. Other data visualisation designers also use 'Tableau' for visualising their data for clients and making their end visualisations in 'Tableau'. In Table 2 the functionalities of Tableau and the envisioned tool (DataVis Helper) are compared.

Tableau	**DataVis Helper**
Data set can quickly be imported	Data set cannot be imported, only gives advice on how the data can be visually represented
Does give the possibility to create visualizations with the imported data	Does not give the possibility to create an end visualization with the user's data set. The user can explore possibilities on which visualizations to use based on the advice given by the DataVis Helper
Exploration of 24 visualization types is available	Exploration of 125 visualization types is available
Does not encourage users to sketch	Encourages users to sketch and use their creativity
Does not give users suggestions on what is most fitting for their data	Will give users suggestions based on what is most fitting for their data

Table 2: Tableau vs. DataViz Helper Functionalities

As summarised in Table 2, 'Tableau' and other data visualisation tools have a different goal and functionality as the DataVis Helper. The 'DataVis Helper' will suggest up to 125 different existing visualisation types, and it encourages users to sketch and use their creativity by applying the principles of visual encoding. 'Tableau' and other data visualisation tools have the advantage that the user can directly import their dataset and immediately start visualising their imported data and end up with a visualisation.

4.4 Final Idea

In this section, the process from the initial idea to the final idea of the tool will be discussed. Initially, the idea of the tool was to have as little steps as possible, to make the process straightforward and easily understandable for the user. After the visualisation analysis has been done, a few new insights were found which will be explained by going through the first idea of the process of the tool.

The first process of the tool consisted of the following steps; (1) analyse the dataset, (2) define the data types, (3) find fitting principles of visual encoding for the data types, (4) sketch with the principles given, here the user will get the freedom to creatively apply the principles in any way he or she wants to and (5) the user will get suggestions on which visualization types to use that fit the data types and visual encodings.

One critical insight that was found is that some variables in the dataset might be more important than others. This depends on the importance of the variable itself, or it depends on the message or story that the user wants to tell. An extra step has been decided to be added where the user must rank his chosen variables, from most important to least important. A second insight gained was that this tool needs to have a maximum number of variables the user can fill in. It has been chosen that the user can pick a maximum of six variables to fill in.

The concluded steps which are illustrated in Table 3 are: (1) fill in the names of the variables, (2) rank the variables, (3) assign data types to the variables, (4) principles will be given based on the data types, (5) the user can now sketch with the given principles and (6) the user will get existing visualization suggestions that he can use.

Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Name the variables	Rank the variables	Assign the data types	Principles will be given	User must sketch	Visualisation types suggestions are given

Table 3: Flowchart of the Process of the DataViz Helper

5. Specification

The final idea has been specified in the previous chapter. In this chapter, the idea will be further specified with the help of performing human prototyping, feedback sessions with the major stakeholders, the making of the user requirements and possible user scenarios.

5.1 Prototyping

In this part of the ideation phase, a human-paper version of the tool has been made and tested. Firstly, by letting students of a Data Visualization course at the University of Twente analyse visualisation types according to the principles of visual encoding, the insights of this occurrence will be further discussed in section 5.2.1. Secondly, a human-paper prototype has been tested to find out whether the process and flow of the tool will work, this will be further discussed in section 5.2.2.

5.1.1 Student Prototyping

A guest lecture on the topic of 'The Design Space of Data Visualization' was given by Yuri Engelhardt, Yuhang Yu and Roseidys Primera to a group of University College students of the University of Twente. The students received an explanation of the design space of data visualisation and on the principles of visual encoding. After these explanations, the students received an exercise where they had to find the principles of visual encoding which were present in the visualisation types in the article 'A Tour through the Visualization Zoo'. Students were asked to form pairs so that they could discuss the answers together.

The exercise which the students had to perform is the similar process that was mentioned in section 4.2, the so-called 'Visualization Analysis' where four data visualisation catalogues were analysed, and the principles present in the visualisation types of these catalogues were found.

Table 4 was given to the students so they can analyse the visualisations of 'A Tour through the Visualization Zoo'. The article 'A Tour through the Visualization Zoo' is divided into five sections consisting each of a different type of visualisation, each pair got assigned a section to analyse.

Data type	**Question**	**Principle**
Visual Appearance	What?	Picturing
Spatial Location	Where?	Mapping
Relationships between entities	Does a given relationship hold between two entities?	Connecting
		Nesting
Nominal Attributes	Which category?	Enclosing
		Spatial Grouping
		Grouping by colour
		Grouping by shape
Ordinal Attributes	Which sequence, order or level?	Chronological Spatial Ordering
		Spatial Ordering
		Ordering by intensity
Quantitative Attributes	When?	Time Axis
	How many? How much?	Axis
		Scaling by size

Table 4: Data Type-Question-Principle Process

The insights gained by testing out the procedure of finding out which principles of visual encoding are present in different types of visualisations were the following: (1) adjusting the terminology of some principles and (2) finding out whether everyday people such as students can grasp the purpose and process of the tool.

The terminology of the following principles was adjusted. Firstly, the difference between 'nesting' and 'enclosing' was not clear to the students; therefore 'nesting' was kept as 'nesting', and 'enclosure' has been changed to 'grouping by enclosure'. Secondly, the students did not understand what was meant by 'spatial grouping', this term has been changed to 'grouping by proximity or distance'. Thirdly, the difference between 'Chronological Spatial Ordering' and 'Spatial Ordering' was not clear, and these principles have been changed into 'Time-based spatial ordering' and 'Attribute-based spatial ordering'. Lastly, the students did not know the difference between the principle 'time axis' and 'axis'. Therefore, the principles consisted of axes has been adjusted into 'Positioning on a time axis' and 'Positioning on an attribute axis'.

The rest of the procedure of assigning principles to each visualisation type was perfectly understood by the students. The purpose of the tool was also more or less clear to the students; most students started questioning the difference between this tool and data visualisation tools such as 'Tableau' which has already been explained in section 4.3.

5.1.2 Human- Paper Prototype

A human-paper prototype has been made and tested to gain insights on how the tool will work. The functionality, the flow and the content of the tool were tested. The human part of the prototype is in this case, the developer of the tool. The developer has all the knowledge that the tool will also have. 'I' is in this section the developer, also known as the human part of the human-paper prototype.

The procedure of testing the human-paper prototype with a potential user consisted of following the six steps mentioned in section 4.4 as the final idea.

Step 1: The participant was asked to search for a dataset and give the variables he wanted to visualise a name. The names for the variables were the same names stated in his data set as the column names. The insight that was gathered here is that participants might tend to use variables that are known as 'unique variables'. Unique variables or name tags are variables with consist of a name. For example, the number of each bus line starting from 1 till 30 or the name or number of students, such as Maria and s123456.

Step 2: The participant was then told to analyse the dataset, this consisted of finding out which variables he wanted to visualise based on the analysis he wanted to make or the message he wants to showcase. Based on his analysis on the importance of his variables, he was asked to rank these variables from one being the most significant and three the least important.

Step 3: The participant had to define his variables with the suitable data type. The process of defining the variables into data types was a bit challenging at first since I only gave the names of the data types to the participant such as 'Visual Appearance or Spatial Location of Entities' and 'Relationships between Entities'. These terms are not universally used or known. The participant needed more help figuring out which data types fit with his variables, so I started asking the participant questions, such as, which of these questions does your variable answer. For example, 'What is x?' or 'Where is x?', 'x' represents the variable. The participant got a clearer idea of what kind of data type his variables were by me asking these questions.

Step 4: The data types were now known, which made the rest of the process run very smoothly. Each data type has certain principles belonging to them, and each combination of principles have accompanying visualisations. At this step, the participant was given a list of the principles fitting to his data types. The participant was given the time and freedom to sketch his first visualisation with the given principles of visual encoding before I gave the participant the possible suggestions on which visualisations fit with his variables.

Step 5: After some sketching, the suggestions were given. While the participant performed the sketching part, a question or insight arose; that I must make sure that users sketch with the given principles and do not skip this step entirely. The freedom to create your non-existing unique visualisation types is an important part of the goal of the tool. The use of creativity is part of the exploration of the design space and cannot be skipped.

The testing concluded with the user receiving three possibilities of visualisation types on how to visualise his variables. The participant decided on which visualisation type he will apply to visualise his variables. In figure 24 the procedure of the human prototyping is illustrated with information on the chosen variables, suitable data type, fitting principles, sketches and the end suggestions.

Step 1: Name of Variables

FuelUsed

VehicleBrand

DestinationToDestinationAmountOfKM

Step 2: Ranking of Variables

VehicleBrand

FuelUsed

DestinationToDestinationAmountOfKM

Step 3: Variables Datatypes

VehicleBrand - Nominal Attribute (Which category?)

FuelUsed – Quantitative Attribute (How much?)

DestinationToDestinationAmountOfKM – Quantitative Attribute (How much?)

Step 4: Principles of Visual Encoding

Vehicle Brand – Grouping by enclosure, proximity, colour or shape

FuelUsed – Axis or scaling by size

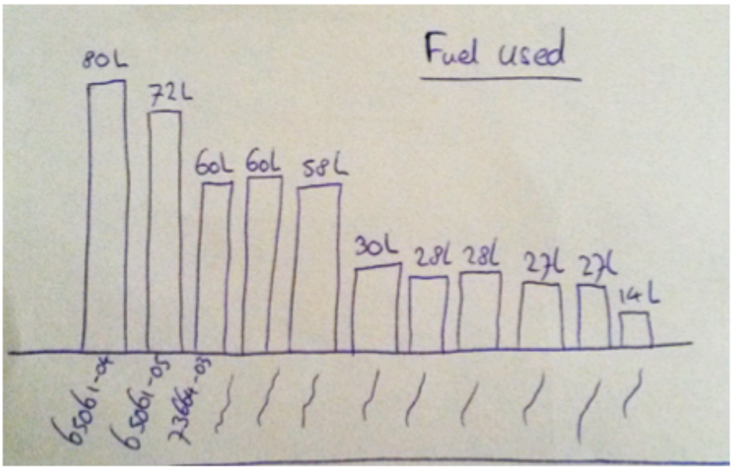
Step 6: Suggestions

Sankey Diagram

Parallel Sets

Chord Diagram

Step 5: Sketching



Chosen Suggestion

Sankey Diagram;

VehicleBrand will use grouping by colour

FuelUsed will use scaling by size

DestinationToDestinationAmountofKM will use attribute axis.

Figure 24: Human-Prototyping Procedure & Results

5.2 Expert Requirements

In this section, the expert's requirements will be elaborated. The main expert of this research project as discussed previously is the supervisor and co-author of the forthcoming book on the basic principles of visual encoding. The expert and the developer had several feedback sessions and brainstorming sessions together to further work out the process of the tool regarding the principles. In this section, the requirements of the expert based on the steps of the tool will be discussed.

Step 1

It must be clear to the user that the use of 'unique identifiers' or 'nametags' should not be chosen as a variable to visualise. These variables can be added as a label to a value, but cannot be chosen as variables to be visually represented in one way or another. A limit must be set for the user; the user can not fill in endless amounts of visualisations. It has been decided that the user can fill in a maximum of six variables.

Step 2

The user's ranking of his variables: it has been decided that this step will be optional. Not every user knows what kind of analysis he wants to make and not every user knows the message that they want to showcase. If the user chooses to rank his variables, he must rank them from most important to least important. For the following steps during the process of the tool, his preference will be taken into consideration. The top variables will be assigned to principles of visual encoding first.

Step 3

Users must assign data types to their variables. The users will not have to assign the data type directly, but indirectly by assigning variables to a fitting question that the variables answer. The data types and the suitable questions can be seen in Table 5.

Data type	**Question Answered**
Visual Appearance	What?
Spatial Location	Where?
Relationships between entities	Does a given relationship hold between two entities?
Nominal Attribute	Which category?
Ordinal Attribute	Which sequence, order or level?
Time Quantitative Attribute	When?
Quantitative Attribute	How many? How much?

Table 5: Data Type - Question Answered

Step 4

In the background, the tool will also have a ranking system to determine which principle will be assigned to a variable. The principles of visual encoding that are based on a spatial dimension always will be suggested first.

Nominal Attributes have the principles; enclosing (grouping by enclosure), spatial grouping (grouping by proximity), grouping by colour and grouping by shape. The suggested principle, in this case, is 'spatial grouping (grouping by proximity). The other principles are on the same level of importance.

Ordinal Attributes have the principles; time-based spatial ordering, attribute-based spatial ordering and ordering by intensity. Firstly, time-based spatial ordering will be suggested, then attribute-based spatial ordering and then ordering by intensity.

Quantitative Attribute has the principles: attribute axis and scaling by size. Attribute axis will be suggested first and then scaling by size.

Step 5

The users get the principles of visual encoding that are fitting to their variables. They will have the opportunity to sketch. At this step examples will be given on when the principles are used, to give the user an idea what is meant by each principle.

Step 6

Suggestions of existing visualisation types consisting of combinations of the principles will be given to the user. First, the visualisation types with the most combinations of principles will be shown and then the visualisation types that combine at least two principles.

5.3 Developer Requirements

In this section, the developer's requirements will be discussed. The software requirements, the structure of the database, the structure of the tool and the structure of the interface will be discussed.

5.3.1 Software Requirements

The developer had the goal to create this tool entirely based on free web application software. The use of these programs was mostly based on the fact that the tool could have been made completely free by using the Azure Microsoft Web Application services, where users can create a free web application connected on the Microsoft Servers and connect all these services with one another.

The final decision on which software that has been chosen:

1) For the management of the database: Data Grip.
2) For the making of the interface: Bootstrap Studio.
3) For the management of the web development: Visual Studio Code.
4) For the tracking of the documentation: GitHub.

In Chapter 6, the software will be explained in further detailed.

5.3.2 Structure of the Database

The tool consists of two main engines, one being where the tool will be developed, and one is the database of the tool where all the information will be retrieved. For the most efficient information retrieval, the following structure has been chosen for the database. The structure consists of three databases, one for the datatypes, one for the principles and one for the visualisations. This structure and the process of retrieval of the data will be detailed explained in chapter 6.

5.3.3 Structure of the Tool

The structure of the backend and frontend communication of the tool is illustrated in figure 25. In this illustration, the different steps of that tool and the user must perform are illustrated. The main tasks of the tool are to store and retrieve information, and the main tasks of the user are to give and use information. This process will be explained in more detail in chapter 6.

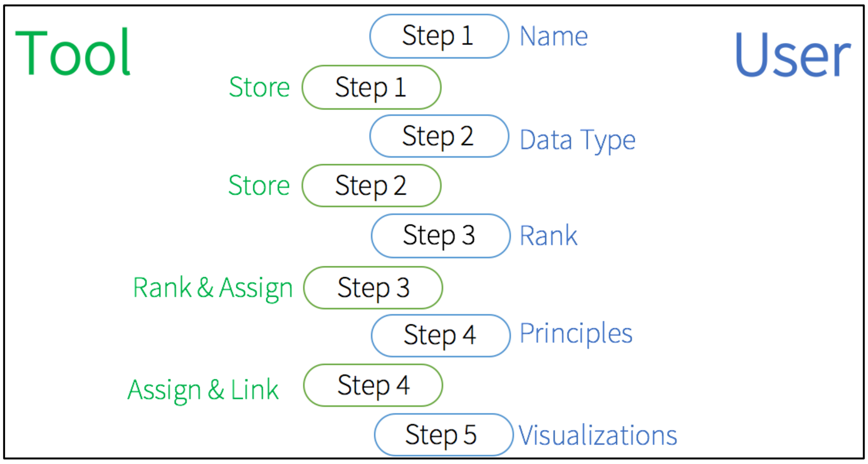


Figure 25: Tool (Back-End) & User (Front-End) Communication

5.3.4. Structure of the Interface

The main requirements for the structure of the user interface have been decided to be as follows; the steps must be easily understandable, and the navigation of the process must run smoothly.

5.4 User Requirements

In this section, the user requirements will be summarised.

The tool will be a universal tool which can be used by anyone. Therefore, it must be easy to understand for all types of users. This suggestion tool, also known as a 'helper', is a new concept in the field of data visualisation design space, which means the users do not know what to expect, so the tool must be explained for the user, each step, definitions and examples given must be clear.

The tool is so universal that there are no clearly defined requirements for the user. The goal is to find out if the user has requirements by performing the universal testings with the first web prototype. These findings will be explored in chapter 6 and 7.

6. Realization

In this chapter, the focus is on the making and testing of the tool which has been developed. This chapter focuses on answering the fourth sub question ‘how can all be applied in a tool that can support and encourage users in exploring the design space of data visualisation?’. This chapter starts with explaining all the components of the tool and then ending with explaining the procedure of the usability testings which have been performed.

6.1 Tool

In this section, the development of the tool will be fully explained which consists of explaining the components, process and different versions of the tool.

6.1.1 The Database

The database of the web tool is made of three main tables which manage the information needed to provide the users with the visualisation suggestions. This information consists of the data types, the principles of visual encoding and the visualisation suggestions. The structure of the database can be seen in the following illustrations. The first table connects the data type with an encoding type, which is illustrated in Table 6.

Data Type	**Encoding**
Visual Appearance	Depicting
Spatial Location	Depicting
Relationships	Linking
Nominal Attributes	Grouping
Time Ordinal Attribute	Time Ordering
Ordinal Attributes	Ordering
Time Quantitative Attribute	Time Scaling
Quantitative Attributes	Scaling

Table 6: Data Type to Encoding

The second table consists of the information of the principles, illustrated in Table 7. This information includes the type of encoding of the principle, the name of the principles and examples of when the principles are applied.

Encoding Type	Principle	Example of use
Depicting	Picturing	Pictorial/Technical Illustration
Depicting	Mapping	Street map, Floor plan
Linking	Connecting	Flow chart, family tree, network graph
Linking	Nesting	Tree map, circle packing
Grouping	Grouping by enclosure (Enclosing)	Venn diagram
Grouping	Grouping by proximity (Spatial grouping)	Diagrams that use spatial clustering
Grouping	Grouping by colour	Diagrams that use colour coding
Grouping	Grouping by shape	Diagrams that use shape coding
Time Ordering	Chronological Spatial Ordering (Time based spatial sequence)	Comic strip, Pictorial instructions
Ordering	Rank-based Spatial Ordering (Attribute based spatial sequence)	Ordered list, bump chart
Ordering	Ordering by intensity	Greyscale/ Intensity map
Time Scaling	Positioning on a time axis (Time Axis)	Time line, clock face
Scaling	Positioning on an attribute axis (Attribute Axis)	Scatter plot
Scaling	Scaling by size	Bar chart, pie chart

Table 7: Principles

The terms in the brackets are the terms which were used in the final version of the prototype of the tool. The terms which are not in the brackets are the most recent developed terms by the forth coming work 'The Fundamentals of Visual Encoding'. The purpose of the first and second table is to get from a datatype to a principle. The element 'Encoding' in the first table and the element 'Encoding Type' in the second table were created to connect one table with the other. This structure was chosen since it gives a clear overview and the information is easily accessible; this makes the workings of the database and the web development software more efficient.

The third table consists of the visualisation analysis which was elaborated in section 4.2, and this table can be seen in Appendix C. This table consists of all the names of the existing visualisation types with their similar other used names and a link to an explanation on how to use the visualisation type and how the visualisation type looks like. Furthermore, in this table, the principles present in each visualisation type are marked. The markings consist of two values, true (1) or false (0), if a principle is present, the value is 'true' which is a '1', and if the principle is not present, the value is 'false' which is a '0'.

6.1.2 The Process

In this chapter, the process of the web tool is explained. The process consists of the steps and content of the tool when it comes to the front-end development or so called interface of the tool. This section discusses each step one by one with illustration on how the tool developed from start till end with two different versions of the web tool. The latest version of the web tool can be found at www.datavishelper.com

Home Page

This step consists of the home page of the tool. The homepage of the tool is simple and has one goal; for the user to get started with the entire process of finding out more about the design space of data visualisation.

The very first version consisted of a big 'Let's Get Started!' button for you to start the process and a menu with an 'About', a 'FAQ' and a 'Contact' page. The about page explains in short what the tool is, its goal and history, mentioning that this tool is part of a research project. The FAQ (Frequently Asked Questions) page has been added so that users can send their feedback, suggestions, complaints or questions. The feedback received will be evaluated and may be implemented to make the tool better. The 'Contact' page focuses on everyone who has contributed to the development of the tool. In figure 26, the first version of the tool is shown. The final version of the pages 'About', 'FAQ' and 'Contact' can be seen in Figures 27, 28 and 29.

The second version of the homepage is the final version used for the testing which can be seen in figure 27. The main addition was adding a short introduction of the tool; 'The DataVis Helper; a tool that helps you explore the design space of data visualisation.' The user can then click the button to get started with the process.

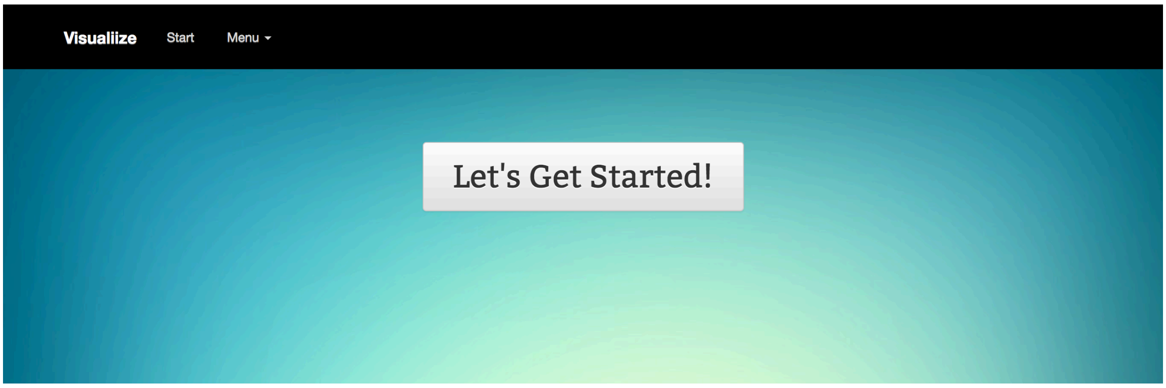


Figure 26: Version 1.0 Interface Design 'Home Page'

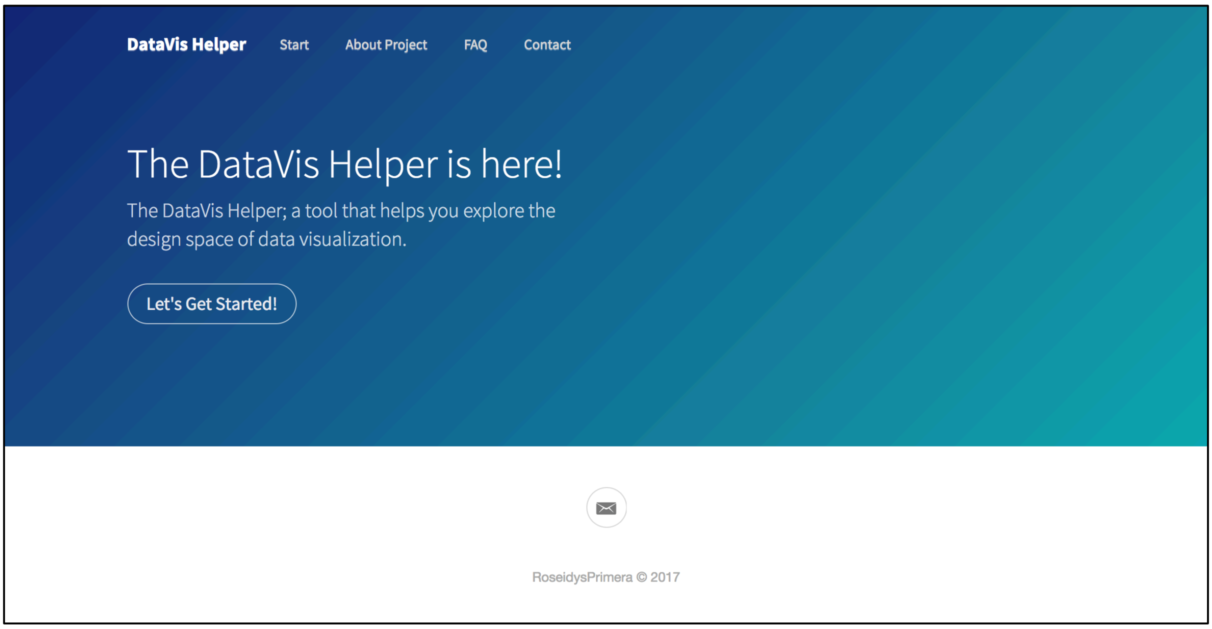


Figure 27: Version 2.0 Interface Design 'Home Page'

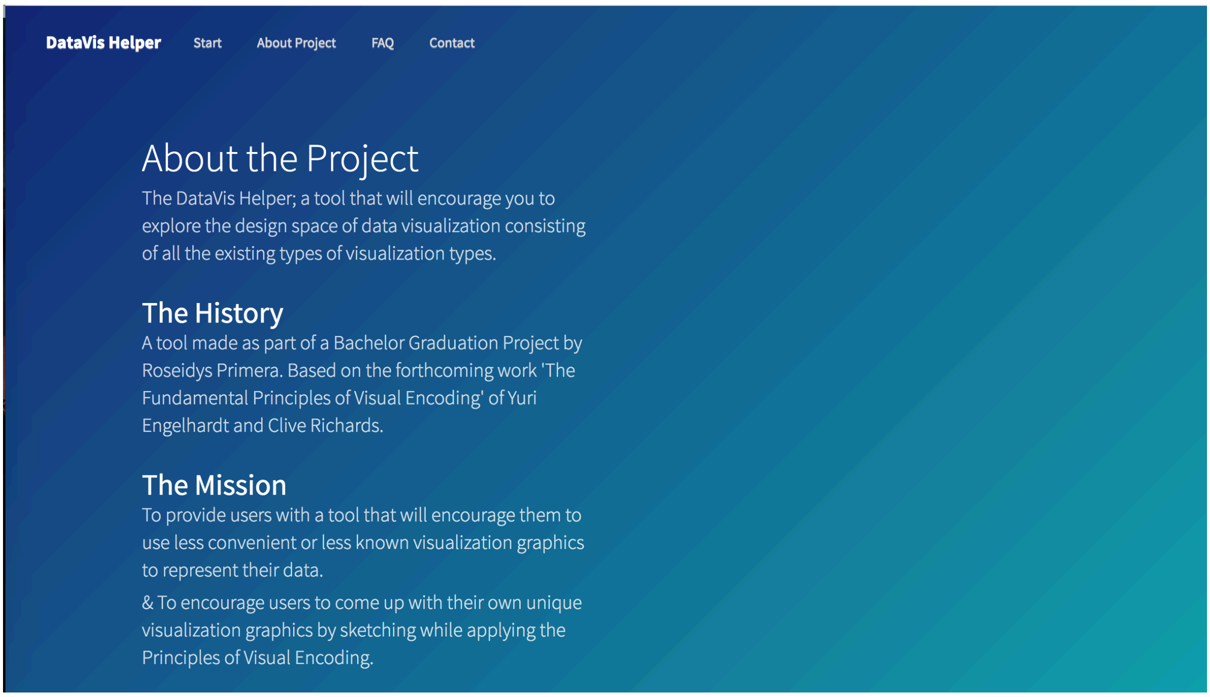


Figure 28: Version 2.0 Interface Design 'About'

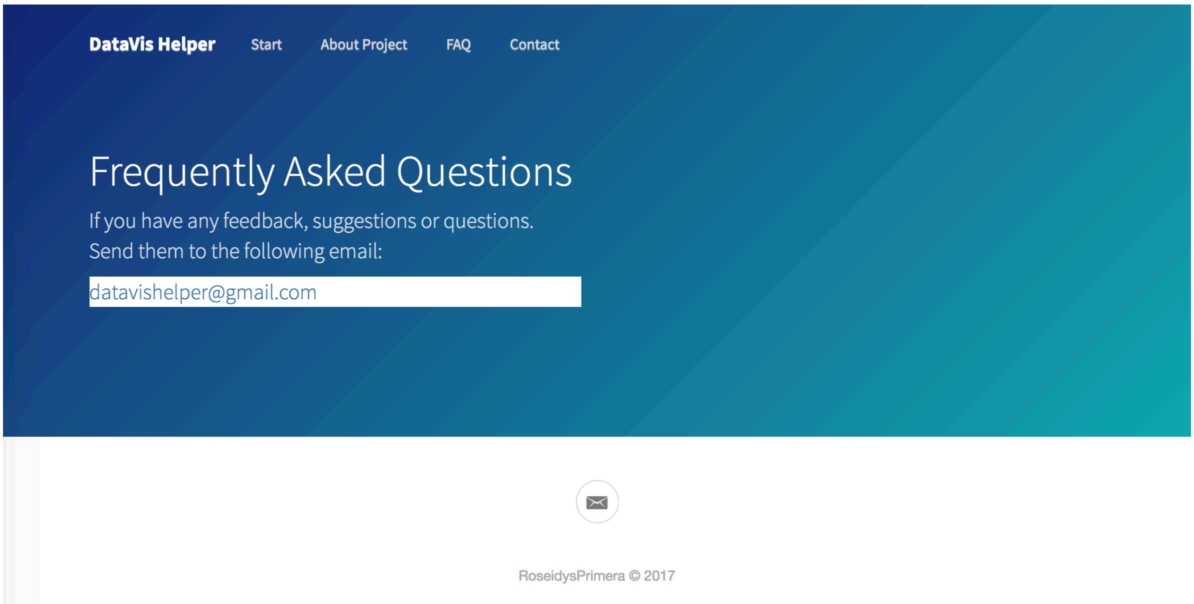


Figure 29: Version 2.0 Interface Design 'FAQ'

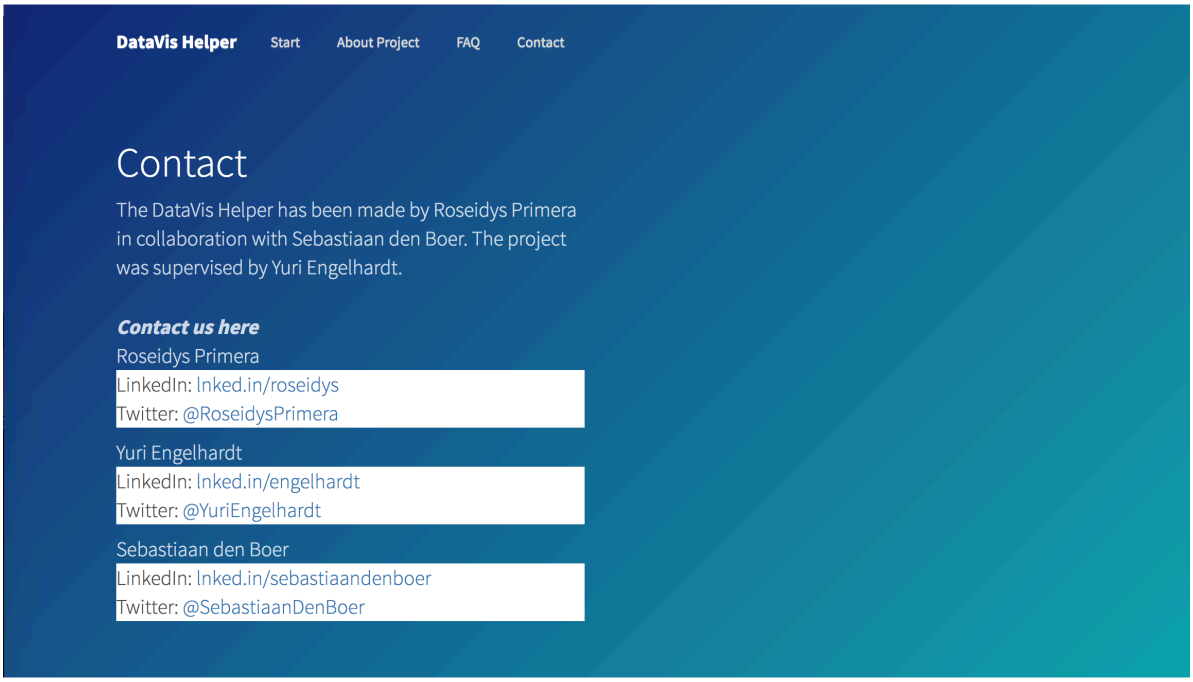


Figure 30: Version 2.0 Interface Design 'Contact'

Step 1: Name your variables

The first step of the process is where the user must choose a dataset and add the names of his variables. The very first version consisted of the user adding his variables by filling in six individual boxes which can be seen in figure 31. This layout was used for quick testing by the developer to see if the back end of the tool worked properly. In the final version, the entire interface was changed and user centred. In the final version step 1 consists of getting a short description on what has to be done and on the right side of the page so-called 'tips' on how the decision on naming your variables have to be made. The user must fill in one variable at a time and by clicking 'add variable' a variable will be added. This can be seen in figure 32.

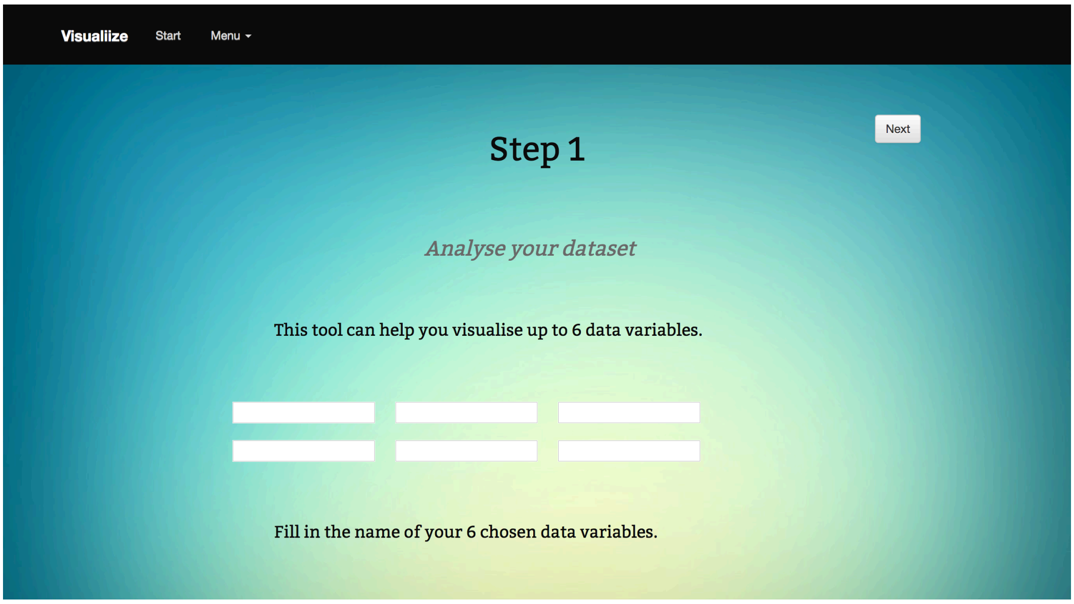


Figure 31: Version 1.0 Interface Design 'Step 1"

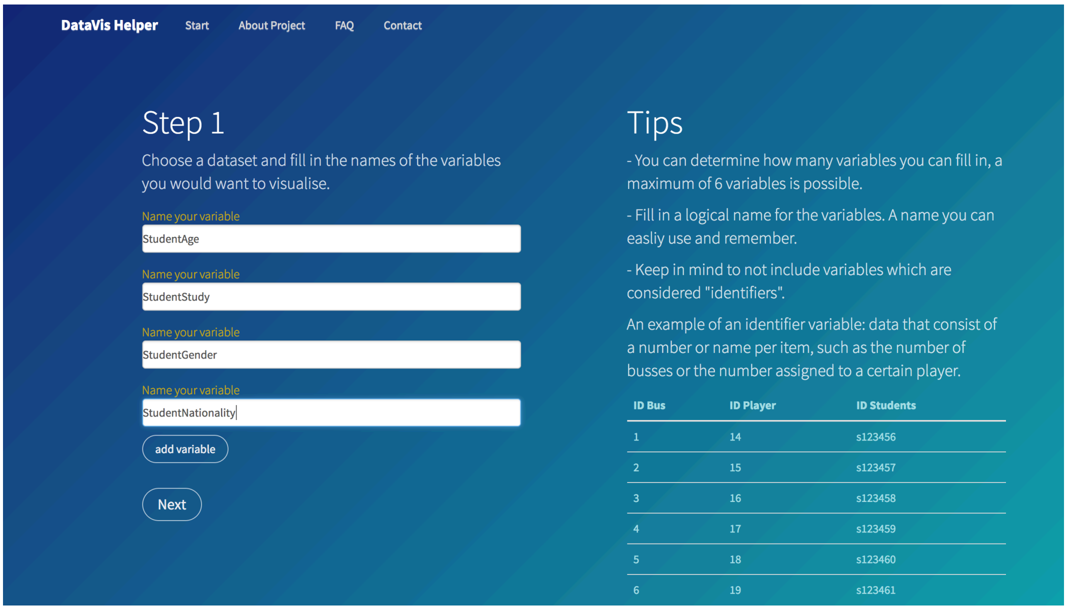


Figure 32: Version 2.0 Interface Design 'Step 1'

Step 2: Assigning and Defining Data Types

Step 2 is defining the data types of the variables. This step can be seen in figure 33. The user gets the choice between different data type questions such as; What? and/or Where?', 'Does a given relationship hold?', 'Which category?', 'Which order or level?', 'Which chronological order?', 'When?', 'How much? Which numerical value?'. These questions help the user understand what kind of data their variables consist of. On the right of the page, tips are being provided on how the user should make his selection. The table gives examples on how the questions should be answered, for example, 'Which category or group does it belong to?' which makes this a Nominal Attribute.

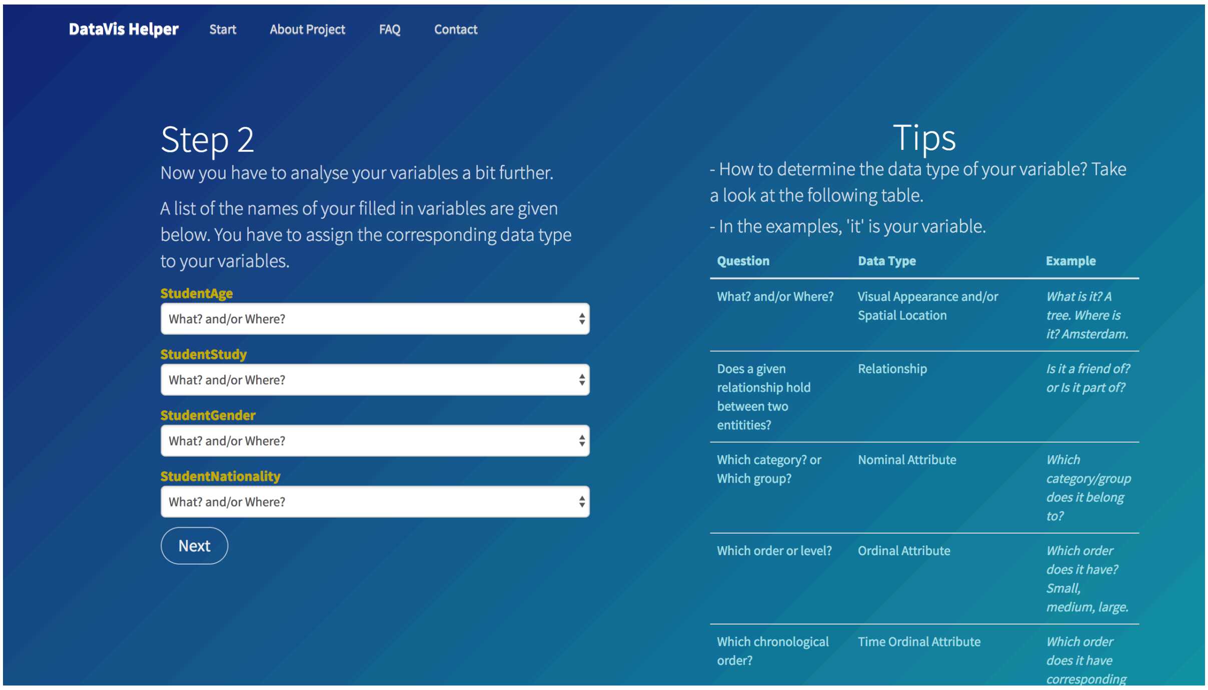


Figure 33: Version 2.0 Interface Design 'Step 2'

Step 3: Ranking

The user must rank his variable by dragging and dropping the variables in the correct order. This can be seen in figure 34. The user's ranking depends on the importance of his analysis or the message he wants to tell with his visualisation. The top variables are the most important and the low variables are the least important. Also, just as by step 1 and 2, on the right side of the page 'tips' are being given on how the user should rank his variables. This step is also optional for the user. The users choose to rank their variables if they know the importance of the variables according to his analysis or if they know the message they want to tell, if the users do not know the answer to either of these two, they can choose to skip this step.

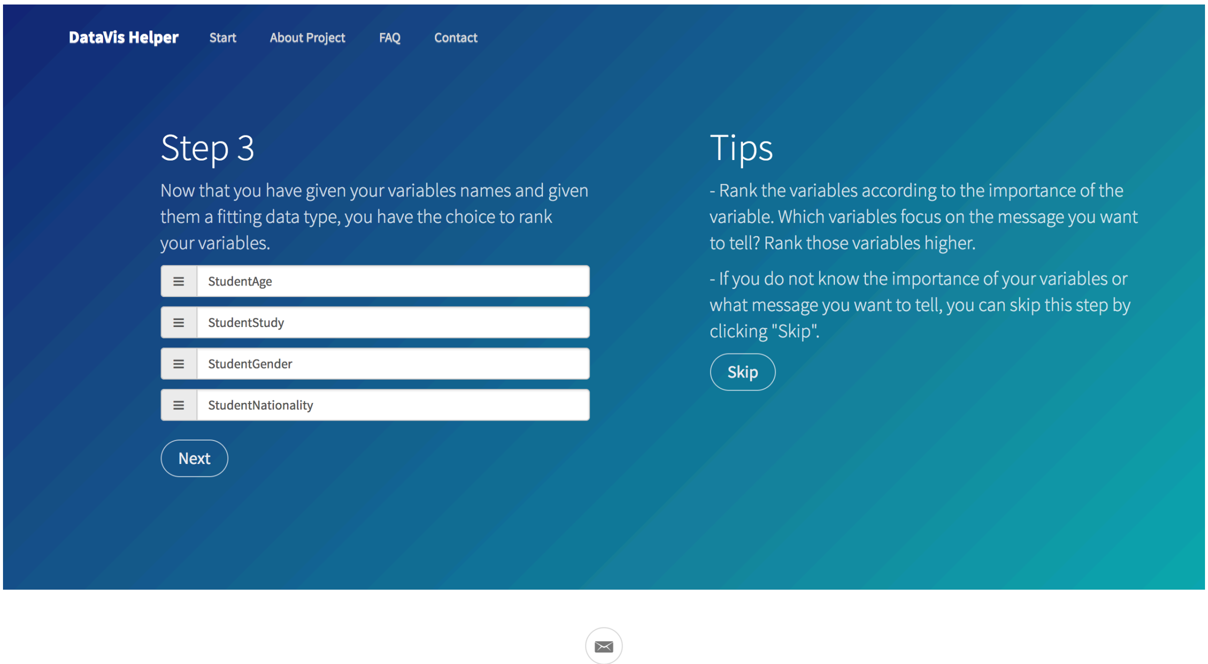


Figure 34: Version 2.0 Interface Design 'Step 3'

Step 4: Principles of Visual Encoding and Sketching

Step 4 consists of giving the user suggestions on which principles of visual encoding to use to represent their variables visually. This step can be seen in figure 35. The users are encouraged to sketch different types of visualisations with the principles given before moving on to the next step.

The back-end consists of a ranking system when it comes to assigning which principle gets selected by the user. The ranking is based on first assigning the spatial dimensions which are mentioned in section 5.2.

Step 4 also consists of a very specific backend working, aside from the ranking system of first assigning the spatial dimensions, the tool also has to make sure that the principles are blocked, once selected or suggested. In Table 8 the principles of visual encoding can be seen, most of the variables can only be assigned to a variable only once, such as Picturing, Mapping, Connecting, Nesting, Enclosing, Spatial Grouping, Grouping by colour, Grouping by shape, Ordering by intensity and Scaling by size. Four principles can be used twice; these are Time based spatial sequence, Attribute based spatial sequence, Time axis and Attribute Axis, if they are suggested twice, they can no longer be suggested again. There is also the principle Mapping that blocks the possibility to suggest the two axes.

Principle
Picturing
Mapping
Connecting
Nesting
Grouping by enclosure (Enclosing)
Grouping by proximity (Spatial grouping)
Grouping by colour
Grouping by shape
Chronological Spatial Ordering (Time based spatial sequence)
Rank-based Spatial Ordering (Attribute based spatial sequence)
Ordering by intensity
Positioning on a time axis (Time Axis)
Positioning on an attribute axis (Attribute Axis)
Scaling by size

Table 8: Principles of Visual Encoding

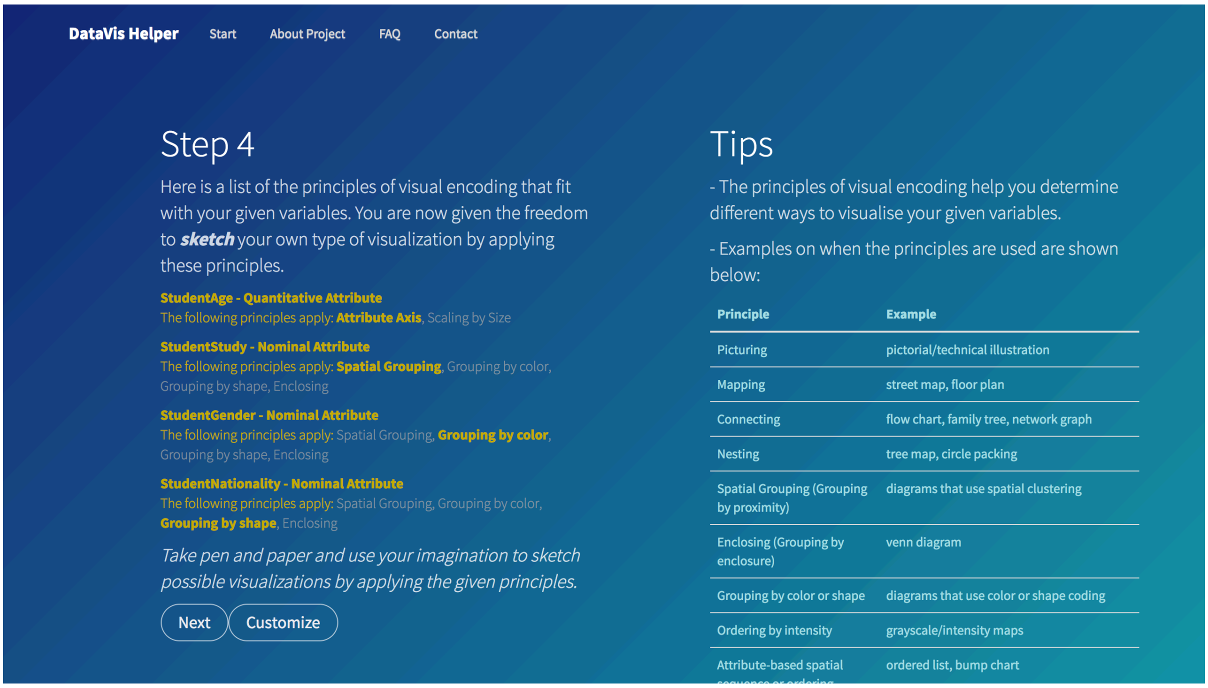


Figure 35: Version 2.0 Interface Design 'Step 4'

Step 5

The suggestions on which existing visualisation types to use are given to the user. This step can be seen in figure 36. A suggestion will be given if the combination of principles has a suitable visualisation type. The tool gives suggestions of at least a combination of two principles unless the user wants to only visualise one variable then suggestions on how to visualise this one variable will be given. As soon as there are 3 or more variables, the suggestions for the combination of at least two variables will be shown.

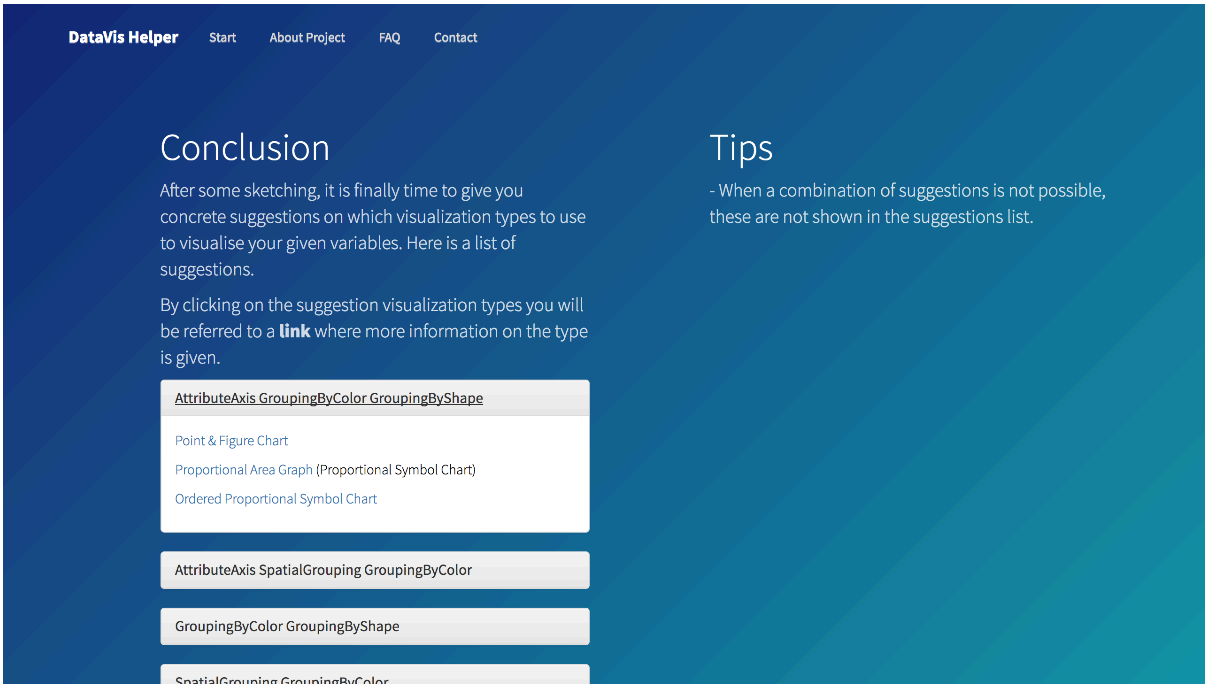


Figure 36: Version 2.0 Interface Design 'Step 5'

6.2 Usability Testing

In this section, the procedure of the usability testing will be explained. As discussed in section 3.4, two types of testings will take place (1) Expert Testings and (2) Universal Testings. Here both usability testings will be elaborated. All testings were recorded and documented and the usability testing consent form that each universal participant had to fill in can be seen in Appendix F.

6.2.1 Expert Testing

This usability testing is focused on performing the test on the supervisor and potentially other data visualisation designers which are known as the 'experts'. The experts that will be tested are two known Dutch Data Visualisation Designers; Nadieh Bremer and Jan Willem Tulp. The procedure of the testings went as follows; (1) they will have the freedom to use their dataset, (2) they must follow the steps of the web tool and (3) they must give critical, constructive and their honest advice on the content and process of the web tool.

The supervisor was tested by a tool walk-through where each step was discussed in detail. The data visualisation designers, on the other hand tested the tool by mail, they were asked to follow the steps of the web tool and provide feedback on their experience with the tool.

6.2.2 Universal Testing

Universal usability tests must be performed to find out if the tool is useful and fulfils the goal and the vision. The technique that is planned to be used is by testing control groups and doing individual sessions. Five types of universal testing groups have been tested: 1) University Students with average experience with data visualization, 2) University Students with little experience with data visualization, 3) University Students with no experience with data visualization, 4) Adults (around the age of 40) with no experience with data visualization and 5) Teenagers (around the age of 15) with no experience with data visualization.

All participants can choose their own data set, and they will be asked to follow the steps of the web tool. At the end the usability test, the users will receive a survey where they will have to describe their experience and improvements for the tool if they have any. The survey used is the 'The System Usability Scale' (SUS). 'The SUS is perhaps the most popular standardised usability questionnaire, accounting for approximately 43% of unpublished usability studies. It is a 10-item questionnaire designed to measure users' perceived usability of a product or system.' [7]

This survey consists of the ten following questions mentioned in figure 37.

	The System Usability Scale Standard Version		Strongly disagree 1	2	3	4	Strongly agree 5
1	I think that I would like to use this system.		○	○	○	○	○
2	I found the system unnecessarily complex.		○	○	○	○	○
3	I thought the system was easy to use.		○	○	○	○	○
4	I think that I would need the support of a technical person to be able to use this system.		○	○	○	○	○
5	I found the various functions in the system were well integrated.		○	○	○	○	○
6	I thought there was too much inconsistency in this system.		○	○	○	○	○
7	I would imagine that most people would learn to use this system very quickly.		○	○	○	○	○
8	I found the system very cumbersome to use.		○	○	○	○	○
9	I felt very confident using the system.		○	○	○	○	○
10	I needed to learn a lot of things before I could get going with this system.		○	○	○	○	○

Figure 37: The System Usability Scale Survey

This usability testing focuses on testing whether the tool is applicable for everyday people, starting from individuals that have zero to a lot of experience with visually representing a dataset. These users also vary in age, have different interests and the reason of why they will use the envisioned tool will differ. Each person had the freedom to determine their own data set during the testings, therefore also their own variables.

7. Evaluation

In this section, the results of the testings will be discussed. This section will also explore and potentially answer the fifth sub question 'what would be benefits and problems that users experience when using such a tool?'.

7.1 Results

As mentioned in section 6.2, there were two types of testings, the testing with experts and a universal testing. In the following sections, the results and feedback given by the participants will be discussed.

7.1.1 Expert Results

In this section, a summary of the most important points is mentioned. The detailed testing notes can be found in Appendix D.

1. The Supervisor

The supervisor chose to think of his own variables.

Step 1: 'The number of the bus lines' can be seen as a quantity, but it is meant to be a nametag. So, use the key word; unique identifier or unique name.

Step 2: The question 'How much?', does not contribute to all the quantitative variables that exist, so the question should be changed to 'How much? Or What is the numerical value?'.

The question 'What? And/or Where?' can be separated since the principles are different, but there are a few exceptions, it was first thought of putting these questions together to form one data type, but there are two different types. 'What' is a broad universal question that can be used for any variable. Therefore, it has been decided to change it into 'What does it look like?', since this data type is 'Visual Appearance' and 'Where?' is a 'Spatial Location'.

Step 3: The formulation on how to guide the user to rank his variables should be changed since knowing which variables are "central" can be tricky. So, users should rank the variables according to the message he or she wants to tell or the analysis of the variables he or she wants to make. The supervisor found the drag and drop functionality very intuitive.

Step 4: The highlighted options mean they are the ones selected and the preferred principle that the user should use.

Step 5: It is not clear which variable is using which principle when suggestion the existing visualisations.

The developer suggested the following solution to the problem at step 5. The variable name can be put in brackets after the principle and add an extra explanation so that the user knows that the text in brackets is the variable name.

The sketches drawn by the supervisor as results of using this web tool can be seen in appendix D.

2. The Data Visualization Designers

As mentioned previously two Data Visualization Designers have been asked for their feedback and testing the web tool; Nadieh Bremer and Jan Willem Tulp.

Feedback received from Nadieh

Step 2: Suggests using the data type names instead of the questions.

Step 4: Users might not know what all the examples in the table mean or look like which means they would not know how to use them. This also applies to the principles; some users might not understand what 'spatial grouping' is.

Step 5: Instead of adding a link to the visualisation, add a picture of each visualisation.

Feedback received from Jan Willem Tulp

Step 1: Novices wouldn't know where to start, give the users an example video to work with or give sample datasets.

Step 5: Suggested to add picture of each type instead of textual information.

He thinks that there might be a gap between what the tool does and the knowledge the user needs to understand all the terminology. Experts will understand the terms, but novices will have a harder time figuring it out. He also suggests explaining to the user on how the suggestions were given, now the suggestions are just being given, but how were they determined.

7.1.2 Universal Results

As mentioned previously all participant got the freedom to choose their own dataset. The testings were documented, and five different groups were tested, 1) University Students with average experience with data visualization, 2) University Students with little experience with data visualization, 3) University Students with no experience with data visualization, 4) Adults (around the age of 40) with no experience with data visualization and 5) Teenagers (around the age of 15) with no experience with data visualization

The following main points were gathered from the universal testings:

- Users thought that the program would give them immediate advice on how to visualise their data with known/existing visualisation types without the need of sketching.
- Users will have a preference on which visualisation to sketch.
- Photos should be added for visual feedback.
- Add a delete button so that you can remove variables.
- The term "choose a dataset" can be confusing, do the users get the option to choose a dataset out of a set of examples or do they have to choose this by themselves. This was not clear for some users.
- The tool should give feedback when assigning the data types, users might make a mistake, but they do not know this. Therefore, feedback should be provided.
- Clarify the point of the tool, or what is going to happen and what is expected. Most users expected the tool to tell them immediately which visualisations to use which meant that the sketching part was a surprise to most of them.
- The principles are a bit too difficult to understand. A definition of their names should be provided.
- Some additional description is needed to explain the difference between a time axis and an attribute axis.
- People tend always to use grouping by colour, even though another principle was suggested. Subconsciously everyone used grouping by colour, sometimes even when it was not mentioned at all or when they did not know the meaning of the other principles.
- The drag and drop function was liked by all users, they all found it very intuitive.
- People get the gist of the principles that are highlighted and not highlighted.
- "Spatial grouping" is not an intuitive term for most the participants.
- Give examples of how to use the principles in combination, people do not know that they can use two attribute axes, or two time-axis, or two axes in general.
- Users suggested to show examples on how to use the tool, a video or tiny clips at the start of each step, mostly helpful for novice users, but also experts.
- Add a search bar, so that users can search for terms on the web page instead of opening a new window and searching somewhere else.
- Create a version of the web tool where the user gets asked what his experience level is. Depending on the experience level, more or less explanation and guidance will be given to the user.
- Participants are surprised by the number of visualisations possible to visualise their data.
- Users understood their data better.

7.2 Results of SUS

Results from the System Usability Scale Questionnaire. The questionnaire was filled in by every universal participant, and the results are the following, '1' stands for 'strongly disagree' and '5' stands for 'strongly agree'.

The result by question 1 is clear, most participants chose either '4' or '5', so most participants would like to use the system again. This can be seen in figure 38.

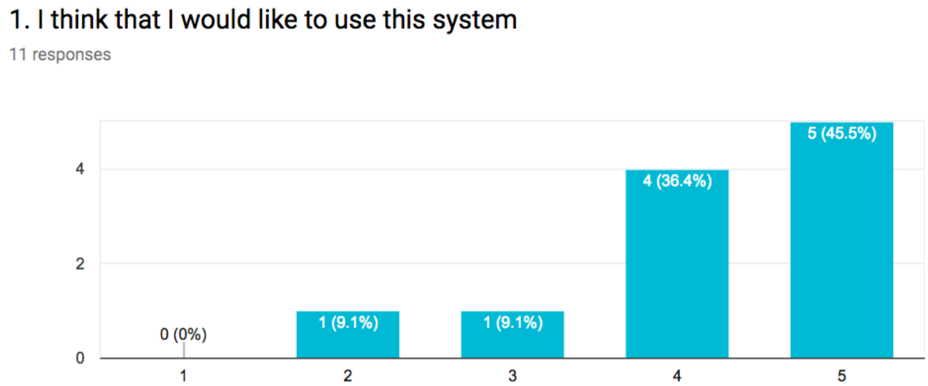


Figure 38: Question 1

The result by question 2 is also clear, most participants chose '2', so most participants did not find the system unnecessarily complex. This can be seen in figure 39.

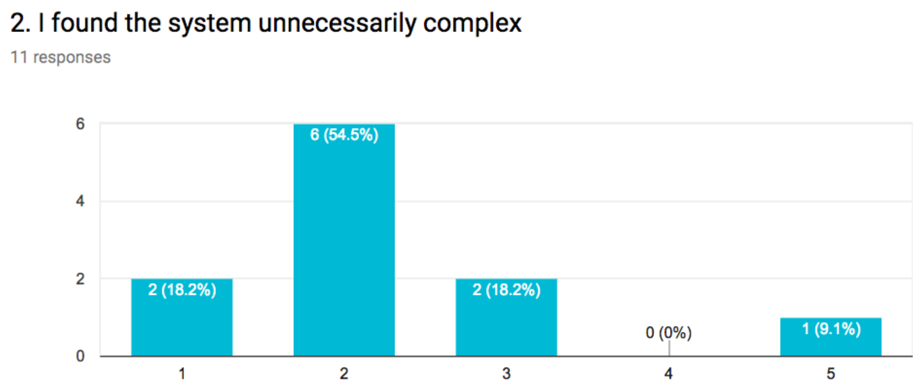


Figure 39: Question 2

The result by question 3 is also clear. Most participants chose '4' which means they agreed with the statement that they thought the system was easy to use. This can be seen in figure 40.

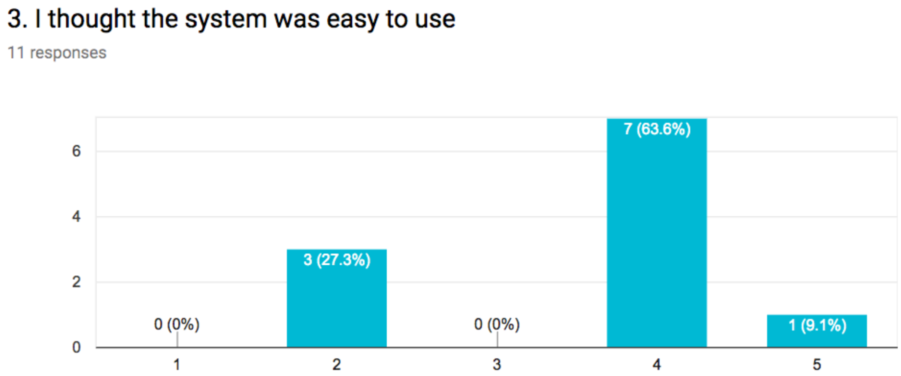


Figure 40: Question 3

The result for question 4 is also clear. Most participants will not need additional support to use the system. This can be seen in figure 41.

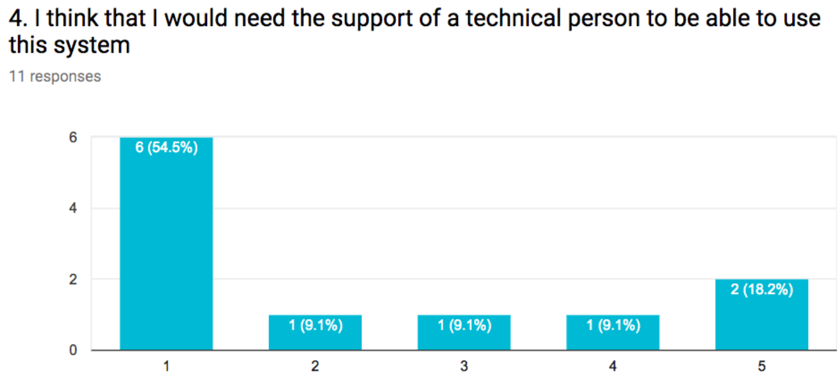


Figure 41: Question 4

The result for question 5 is also clear. Most participants found the functions well integrated. This can be seen in figure 42.

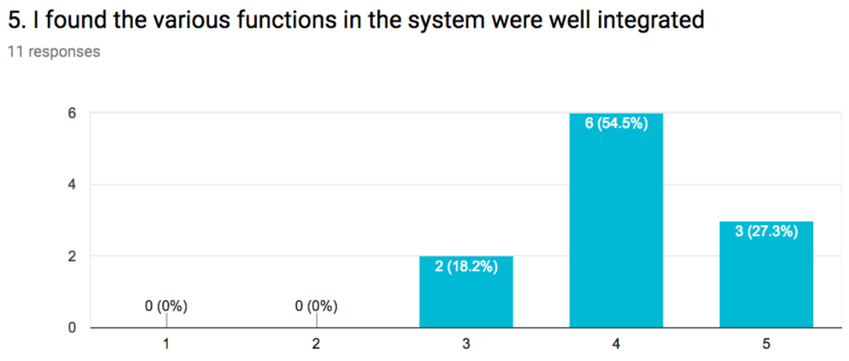


Figure 42: Question 5

The result of question 6 is also clear. Most participants thought that the system did not have inconsistency. This can be seen in figure 43.

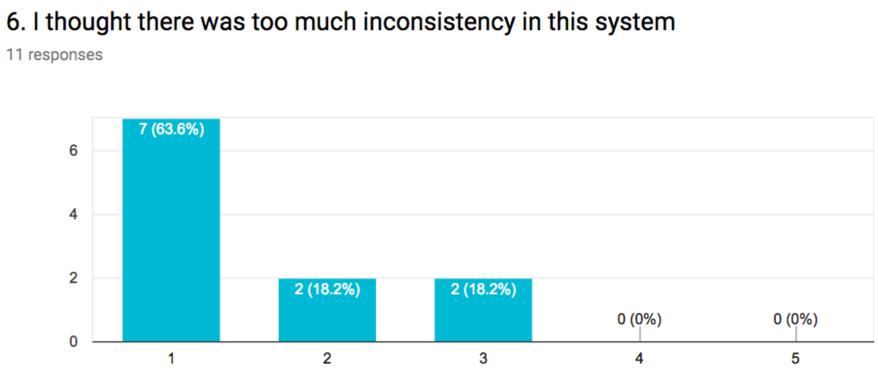


Figure 43: Question 6

The result for question 7 is also clear; most participants agree that they can imagine people learning to use this system very quickly. This can be seen in figure 44.

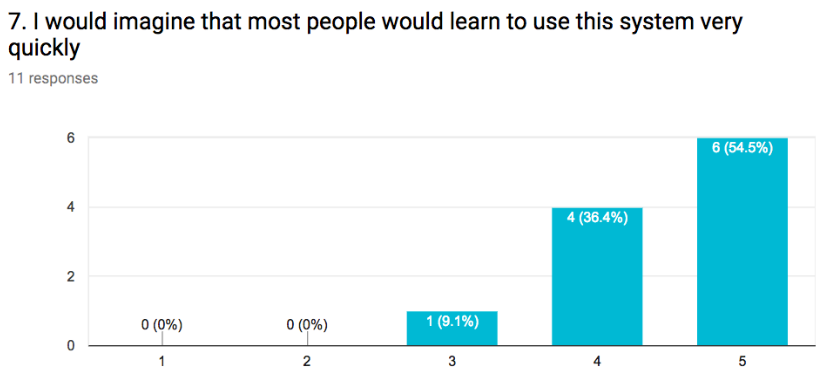


Figure 44: Question 7

The result for question 8 is also clear; most participants did not find the system cumbersome or difficult. This can be seen in figure 45.

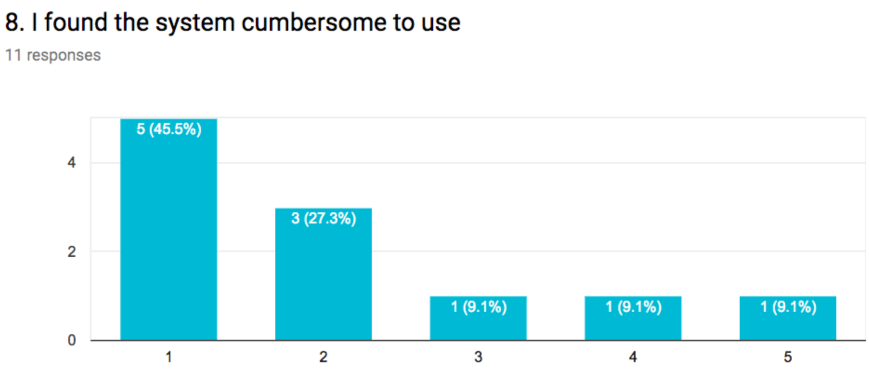


Figure 45: Question 8

The result for question 9 is also clear, the users were not too confident while using the system, most participants were still doubtful. This can be seen in figure 46.

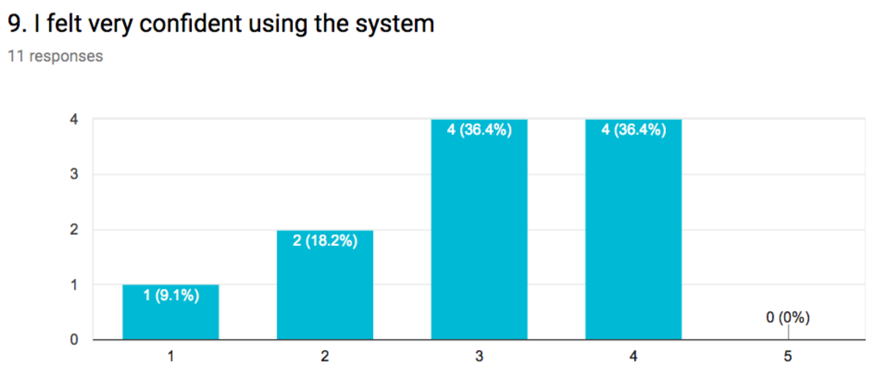


Figure 46: Question 9

The result for question 10 is also clear, most participants, agreed that they did not needed to learn a lot of things before they got going with the system. Some needed more time than others. This can be seen in figure 47.

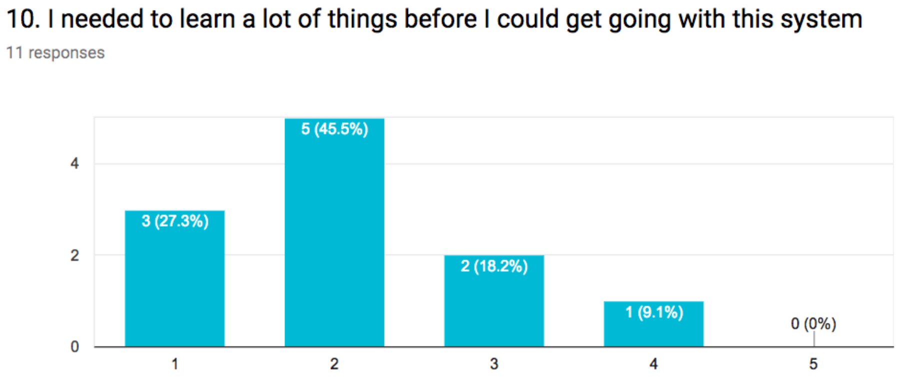


Figure 47: Question 10

8. Conclusion

In this section, the research problem of this research project will be concluded. In this section, it will be concluded if the tool helped with answering the research question 'How can we support and encourage users in exploring the diversity of the data visualisation design space, in ways that are based on the fundamental principles of visual encoding?', it is time to see if the goal was achieved.

The focus of this research project was to explore the design space of data visualisation by creating a web tool that would eventually support and encourage users in exploring the diversity of the data visualisation design space, in ways that are based on the fundamental principles of visual encoding. A web tool was developed which is data centred, creativity centred and consist of 125 unique existing visualisation types. Different versions of prototypes were made, expert and universal usability testings have been performed. According to the results obtained by the SUS questionnaire, users will use this system. According to the results obtained by participants using the web tool to visualise their dataset, the number one argument that every participant agreed on is that they were all surprised by the number of visualisations they can use to visualise their variables. This means that the users were encouraged in exploring the diversity of the design space of data visualisation. The tool still needs to be improved when it comes to supporting users in exploring the design space since every user is different and their level of experience and knowledge differs.

The tool also encouraged users creatively to explore the design space of data visualisation based on the principles. This can be seen by observing all the sketches that the participants made, and the insight gained that most participants understood their data better while sketching their variables visually on paper.

So, did this research project answer the research question? Yes, it did. It has been proven that such a tool help users to be aware of the abundance of visualisation types that exist. It has also been proven that by sketching users might understand their data better and create new visualisation types that did not exist before. The tool encourages and supports the exploration of the design space of data visualisation, but still needs a few improvements to be a universally useful tool for any users.

9. Future Work

This chapter focuses on what will make the tool encourage and support users to explore the design space of data visualisation, including any other recommendations not related to the goal of this research project. This chapter will, therefore, be divided into two sections, the major recommendations and the minor recommendations.

9.2 Major Recommendations

- The tool currently consists of 125 existing visualisation types, if more catalogues were explored this amount might be larger, so for a future task is to explore every data visualisation catalogue or chart chooser that exists.
- To create a web tool where users can select their level of experience. This will make it more appealing for any users, novice and experts.
- Add a "helper-bot". Instead of adding a video-clip on each web page explaining how the step should be performed, a "helper-bot" can be added in the web tool. Users can ask the bot all types of questions.
- Adding pictures to the web tool. The users need visual feedback.
- Give the user feedback when the wrong data type is assigned to a variable or letting the tool assign a variable directly. This consist of potentially building an 'AI' (artificial intelligence) in the web tool, where the user should fill in the names or import the data set, and the tool does everything else, except the sketching part.
- The user should be informed about what is to be expected. The user should know why he should use the tool, how to use the tool and what the tool does.

9.3 Minor Recommendations

- Making the code of the web tool open source, since the community of data visualisation designers is constantly working together and working on other's work. It might be that if the code of the tool is put online, software developer or data visualisation experts can add functionalities that no one would not have thought of.

Appendix A – Visualization Analysis A

The Visualization Analysis of the separate tools.

https://docs.google.com/spreadsheets/d/1iDrSn98UtI1BVHpxeR_fO9VMp0j0iH1Lyh8SiKRi-iA/edit#gid=1326833207

	Version 2		1	2	3	4	5	6	7	8	9	10	11	12	13
			Depicting		Linking		Grouping			Ordering			Scaling		
			Picturing	Mapping	Connectin	Containin	Grouping by	Grouping by color	Grouping by shape	Ordering	Chronolo gical	Ordering by	Axis	Time Axis	Scaling by size
	Visualization Zoo														
1	Index Chart														
2		Line Chart											x	x	
3		Stacked Line Chart											x	x	
4	Area Chart												x	x	
5	Stacked Area Chart												x	x	
6	Small Multiples Area Chart									x			x	x	
7	Horizan Graphs										x		x	x	
8	Histogram										x		x	x	
9	Box-and-Whisker Plot										x		2x		
10	Stem-and-Leaf Plot						x			x			2x		
11	Q-Q Plot									x			x		

12	Scatter Plot Matrix							x	x	x			x	x	
13	Parallel Coordinates				x								x		
14	Maps			x											
15		Flow Maps		x	x										x
16		Choropleth Maps		x								x			
17		Graduated Symbol Maps		x											x
18	Cartogram			x											x
19	Node-Link Diagram				x										x
20	Adjacency Diagram	(linear cordinates)			x		x			x					
21		Sunburst (polar coordinates)			x		x			x					
22	Enclosure Diagram					x	x					x			x
23		Treemap				x	x					x			x
24	Nested Cricles					x	x					x			x
25	Force-directed Layout				x			x							x
26	Arc-Diagram				x					x			x		x (optional)
27	Matrix Views				x		x					x	2x		
	DataViz Catalogue														
1	Bar Chart												x		
2		Histogram									x		x	x	
3		Stacked Bar Graph (instead of multisets,											x		

		bars are stacked upon each other)													
4		Multi-set Bar Chart											x		
5		Population Pyramid (Back-to-back Histogram)									x		x	x	
6		Radial Bar Chart											x		
7		Radial Column Chart											x		
8	Pie Chart	(center of the donut cut out)								x			x		x
9	Donut Chart									x			x		x
10	Sunburst Diagram				x		x			x					
11	Spiral Plot							x (optional)	x (optional)		x		x	x	
12	Nightingale Rose Chart												x	x	
13	Gantt Chart	(useful for project management)						x (optional)			x		x	x	x
14	Span Chart												2x		
15	Candlestick Chart												x	x	
16	Box & Whisker Plot										x		2x		
17	Bullet Graph										x		2x		
18	Parallel Sets				x		x	x					x		x
19	Sankey Diagram				x			x (optional)					x		x
20	Chord Diagram				x										x
21	Maps			x											

22		Bubble Map		x				x (optional)	x (optional)						x
23		Choropleth Map		x								x			
24		Connection Map		x	x			x (optional)							x (optional)
25		Dot Map		x		x				x					
26		Flow Map		x	x			x (optional)							x
27	Scatter Plot							x (optional)	x (optional)				2x		
28	Bubble Chart	Simplified version is the Proportional Area Chart						x (optional)	x (optional)				2x		x
29	Venn Diagram					x		x							
30	Circle Packing	similar to the treemap qua function				x		x (optional)			x				
31	Word Cloud		x					x (optional)							x
32	Brainstorm	(Mind-Map)		x?	x			x (optional)							x (optional)
33	Heatmap											x	2x		
34	Marimekko Chart	(Mosaic Plot) = Two Way 100% Stacked Bar Graph						x (optional)					2x		
35	Treemap	Alternative way to visualise a Tree Diagram				x	x	x (optional)				x			x
36	Pictogram Chart		x					x (optional)	x				x		
37	Point & Figure Chart		x				x	x	x				x	x	

38	Dot Matrix Chart							x	x (optional)		x				x (optional)
39	Line Graph												2x	x (optional)	
40	Kagi Chart	comparible with the Candlestick Chart						x			x		x	x	
41	Timeline		x (optional)	x (optional)	x			x (optional)			x			x	x (optional)
42	Arc Diagram				x					x			x		x (optional)
43	Open-high-low-close Chart							x					x	x	
44	Network Diagram				x			x (optional)							x (optional)
45	Error Bars	Addition to a scatterplot, dotplot, bar chart and line graph											2x	x (optional)	
46	Non-ribbon Chord Diagram				x										
47	Flow Chart		x		x			x (optional)	x						
48	Parallel Coordinates Plot				x			x (optional)		x			multiple vertical axes		
49	Illustration Diagram		x	x											
50	Tree Diagram		x (optional)		x			x (optional)	x (optional)						
51	Tally Chart						x						x		
52	Area Graph												2x	x (optional)	

53	Density Plot	comparible with an histogram						x					x	x	
54	Stacked Area Graph							x					2x	x (optional)	
55	Stream Graph							x					x	x	x
56	Proportional Area Chart							x (optional)	x				x		x
57	Violin Plot	Combination of a density plot and box plot						x (optional)					x	x	
58	Radar Chart				x			x			x		x		
	Visual Vocabulary	See above!													
1	Bar												x		
2		Ordered Bar									x		x	x (optional)	
3		Paired Bar (multiset bar)									x (optional)		x	x (optional)	
4		Spine Chart								x			x		
5		Diverging Bar								x			x		
6		Population Pyramid (back to back histogram)									x		x	x	
7		Proportional Stacked Bar											x		x
8		Bullet Chart									x		2x		
9		Boxplot									x		2x		
10		Diverging Stacked Bar								x			x		
11		Priestley Timeline									x		x	x	x
12		Sankey				x			x (optional)					x		x

13	Column												x		
14		Ordered Column									x		x	x (optional)	
15		Histogram									x		x	x	
16		Paired Column									x (optional)		x	x (optional)	
17		Line + Column									x		x	x	
18		Stacked Column											x		
19		Waterfall			x			x (optional)	x (optional)	x			x	x (optional)	x
20	Pie									x			x		x
21	Donut									x			x		x
22	Arc				x					x			x		x (optional)
23	Isotype (Pictogram)		x					x (optional)	x				x		
24	Grouped Symbol	(Dot Matrix Chart)						x	x (optional)		x				x (optional)
25	Ordered Proportional Symbol							x (optional)	x		x		x		x
26	Proportional Symbol	(Proportional Area Chart)						x (optional)	x				x		x
27	Gridplot							x					2x		
28	XY Heatmap							x (optional)				x	2x		x (optional)
29	Calendar Heatmap							x (optional)				x		2x	x (optional)
30	Dot strip plot						x	x (optional)		x	x (optional)		x		
31	Scatterplot							x (optional)	x (optional)				2x		

32	Bubble							x (optional)	x (optional)				2x		x
33	Maps			x											
34		Basic Choropleth (rate/ratio)		x								x			
35		Proportional Symbol (count/magnitude)		x				x (optional)	x (optional)						x
36		Flowmap		x	x			x (optional)							x
37		Contour map		x								x			
38		Equalised Cartogram		x				x	x (optional)			x (optional)			
39		Scaled Cartogram (value)		x				x	x (optional)			x (optional)			x
40		Dot Density		x			x	x (optional)		x	x (optional)		x		
41		Heat Map		x								x			
42	Slope				x		x (optional)	x (optional)			x				
43	Line												2x	x (optional)	
44		Cumulative Curve											2x		
45		Dot plot											x		
46		Lollipop Chart											x		
47		Circle Timeline						x (optional)			x	x (optional)		x	x
48		Connected Scatterplot			x			x (optional)	x (optional)				2x		
49		Stock Price (Open-high-low-close Chart)						x					x	x	
50		Parallel Coordinates			x			x (optional)		x			multiple verti		

													cal axes		
51		Vertical Timeline	x (optional)	x (optional)	x			x (optional)			x			x	x (optional)
52		Barcode Plot (comparible with the dot strip plot)					x	x (optional)		x	x (optional)		x		
53		Seismogram (comparible with the circle timeline)						x (optional)			x	x (optional)		x	x
54		Chord			x										x
55		Network			x			x (optional)							x (optional)
56	Area Chart												2x	x (optional)	
57		Surplus/Deficit Filled Line ?													
58		Fan Chart (projections)											2x	x (optional)	
59		Radar Chart			x			x			x		x		
60		Violin Plot						x (optional)					x	x	
61		Venn				x		x							
62		Voronoi				x	x					x			x
63		Treemap				x	x					x			x
	Anychart (Chartopedia)		See above												
1	Map			x											
2		Flow Map		x	x			x (optional)							x

3		Connector Map		x	x			x (optional)							x (optional)
4		Dot Map		x			x	x (optional)		x	x (optional)		x		
5		Choropleth Map		x								x			
6		Bubble Map		x				x (optional)	x (optional)						x
7	Bar Chart												x		
8		Bullet Chart									x		2x		
9		Range Bar Chart											x		
10		Stacked Bar Chart											x		
11		Percent Stacked Bar Chart											x		
12		Gantt Chart						x (optional)			x		x	x	x
13		Resource Chart (example of an gannt chart)						x (optional)			x		x	x	x
14		Linear Gauge						x					x		
15	Pie Chart									x			x		x
16	Donut Chart									x			x		x
17	Circular Gauge							x					x		
18	Funnel Chart							x					x		
19	Pyramid Chart							x			x		x	x	
20	Column Chart												x		
21		Range Column Chart											2x	x (optional)	
22		Pareto Chart											x		
23		Error Chart											2x	x (optional)	

24		Japanese Candlestick Chart											x	x	
25		Box Chart									x		2x		
26		Precent Stacked Column Chart											x		
27		Stacked Column Chart											x		
28	Area Chart												2x	x (optional)	
29		Range Area Chart											2x	x (optional)	
30		Range SplineArea Chart											2x	x (optional)	
31		Range StepLineArea Chart											2x	x (optional)	
32		StepLine Area Chart											2x	x (optional)	
33		Spline Area Chart											2x	x (optional)	
34		Percent Stacked Area Chart											x	x	
35		Percent Stacked SplineArea Chart											x	x	
36		Percent Stacked StepLineArea Chart											x	x	
37		Stacked SplineArea Chart											x	x	
38		Stacked Area Chart											x	x	
39		Stacked StepLineArea Chart											x	x	
40		Polar Chart			x			x			x		x		
41		Radar Chart			x			x			x		x		

42	Line Chart												2x	x (optional)	
43		Stock Chart											2x	x (optional)	
44		Spline Chart											2x	x (optional)	
45		OHLC Chart						x					x	x	
46		StepLine Chart											2x	x (optional)	
47		Stick Chart		x			x	x (optional)		x	x (optional)		x		
48		Sparkline Chart		x			x	x (optional)		x	x (optional)		x		
49		Jump Line Chart											2x	x (optional)	
50		PERT Chart			x			x (optional)							
51	Dot Chart			x			x	x (optional)		x	x (optional)		x		
52	Bubble Chart							x (optional)	x (optional)				2x		x
53	Heatmap							x (optional)				x	2x		x (optional)
54	Treemap					x	x					x			x
55	Seatmap		x	x											
56	More?														

Appendix B – Visualization Analysis B

The Visualization Analysis of the merged existing visualization types.

https://docs.google.com/spreadsheets/d/1iDrSn98UtI1BVHpxeR_fO9VMp0j0iH1Lyh8SiKRi-iA/edit#gid=1326833207

	Merging		Similar	Link to	Descriptio	1	2	3	4	5	6	7	8	9	10	11	12	13	14
							Depicting		Linking				Grouping			Ordering			Scaling
						Picturing	Mapping	Connectin	Nesting	Enclosing	Grouping	Grouping	Grouping	Ordering	Chronolo	Ordering	Axis	Time Axis	Scaling
1	Line		Line	http://ww													x	x	
2		Stacked		http://ww	Same, but												x	x	
3		Small		http://ww	Not												x	x	
4		Cumulativ		https://git	Can be												2x	x	
5		Dot plot	(Similar to	https://git	Can also						x			x			2x		

6		Stem-		http://ww							x			x			2x		
7		Q-Q Plot		http://que	Can be									x			x		
8		Scatter	SPLOM	http://que	Can also							x	x	x			x	x	
9		Lollipop		https://git	Can be												x		
10		Circle		https://git								x			x	x		x	x
11		Connecte		https://git				x				x	x				2x		
12		Vertical		https://git		x	x	x				x			x			x	x
13		Barcode		https://git							x	x		x	x		x		
14		Seismogr		https://git								x			x	x		x	x
15		Stock		http://ww													2x	x	
16		Spline		http://ww													2x	x	

17		StepLine		http://ww													2x	x	
18		Stick		http://ww			x				x	x		x	x		x		
19		Sparkline		http://ww			x				x	x		x	x		x		
20		Jump		http://ww													2x	x	
21		PERT	a type of	http://ww				x				x							
22	Area		Area	http://ww													x	x	
23		Stacked		http://ww													x	x	
24		Small		http://que										x			x	x	
25		Horizan		http://que											x		x	x	
26		Surplus/D	SAME AS	https://git															
27		Fan Chart		https://git													2x	x	

28		Voronoi							x		x					x			x
29		Range															2x	x	
30		Range															2x	x	
31		Range															2x	x	
32		StepLine															2x	x	
33		Spline															2x	x	
34		Percent															x	x	
35		Percent															x	x	
36		Percent															x	x	
37		Stacked															x	x	
38		Stacked															x	x	

39		Polar						x				x			x		x		
40	Parallel		Parallel					x									x		
41	Maps						x												
42		Flow					x	x											x
43		Choroplet					x									x			
44		Basic					x									x			
45		Graduate	Proportio				x												x
46		Bubble					x					x	x						x
47		Connectio	Connecto				x	x				x							x
48		Dot Map	Dot				x		x					x					
49		Contour					x									x			

50		Equalised					x					x	x			x			
51		Scaled					x					x	x			x			x
52		Heat Map					x									x			
53	Cartogra						x												x
54	Node-							x											x
55		Tree				x		x				x	x						
56	Adjacenc	(linear						x			x			x					
57		Sunburst						x			x			x					
58	Enclosure								x		x					x			x
59		Treemap							x		x					x			x
60		Nested	Circle						x		x					x			x

61	Force-							x				x							x
62	Arc-							x						x			x		x
63	Matrix							x			x					x	2x		
64	Bar Chart																x		
65		Histogra													x		x	x	
66		Stacked															x		
67		Multi-set															x		
68		Populatio													x		x	x	
69		Radial															x		
70		Radial															x		
71		Bullet													x		2x		

72		Box-and-	Box Chart												x		2x		
73		Ordered													x		x	x	
74		Paired													x		x	x	
75		Spine												x			x		
76		Diverging												x			x		
77		Proportio															x		x
78		Diverging												x			x		
79		Priestley													x		x	x	x
80		Line +	Pareto												x		x	x	
81		Waterfall						x				x	x	x			x	x	x
82		Range															2x	x	

83		Precent															x		
84	Pie Chart	(center of												x			x		x
85	Donut													x			x		x
86	Spiral											x	x		x		x	x	
87	Nightingal																x	x	
88	Gantt	(useful for										x			x		x	x	x
89	Span																2x		
90	Candlesti		Japanese														x	x	
91	Parallel							x			x	x					x		x
92	Sankey							x				x					x		x
93	Chord		Chord					x											x

104	103	102	101	100	99	98	97	96	95	94
Point &	Pictogram	Marimekk			Heatmap	Brainstor	Word	Venn	Bubble	Scatter
		(Mosaic	Calendar	XY		(Mind-			Simplified	
	Isotype								Bubble	
x	x						x			
						x?				
						x				
								x		
x										
x	x	x	x	x		x	x	x	x	x
x	x								x	x
			x	x	x					
x	x	2x		2x	2x				2x	2x
x			2x							
			x	x		x	x		x	

105	Dot		Grouped									x	x		x				x
106	Kagi	comparibl										x			x		x	x	
107	Timeline					x	x	x				x			x			x	x
108	Open-		"Stockpric									x					x	x	
109	Network		Network					x				x							x
110	Error	Addition	Error														2x	x	
111	Non-							x											
112	Flow					x		x				x	x						
113	Illustratio					x	x												
114	Tally										x						x		
115	Density	comparibl										x					x	x	

116	Stream											x					x	x	x
117	Proportio		Proportio									x	x				x		x
118	Ordered											x	x		x		x		x
119	Violin Plot	Combinati										x					x	x	
120	Radar							x				x			x		x		
121	Arc	(not the	Arc Bar					x						x			x		x
122	Gridplot											x					2x		
123	Dot strip										x	x		x	x		x		
124	Slope							x			x	x			x				
125	Bump																2x	x	

Appendix C – Visualization Analysis C

The Visualization Analysis Database Version.

https://docs.google.com/spreadsheets/d/1WYn67blLR_gG9qcLhFmRIj3PPGHqs8egli2AfhMvUDo/edit#gid=1473409783

VisualizationNam	SimilarNames	Link	Picturing	Mapping	Connecti	Nesting	Enclosin	SpatialGr	Grouping	Grouping	Ordering	Attribute	TimeBas	TimeAxis	Attribute	ScalingR
'Line Graph'	'Line Chart'	'http://www.datavizcatalogue.com/methods/line_graph'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stacked Line Graph'	'Stacked Line Chart'	'http://www.datavizcatalogue.com/methods/line_graph'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Dot Plot'	'Stem-and-Leaf Plot'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Cumulative Curve'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stem-and-Leaf Plot'	null	'http://www.datavizcatalogue.com/methods/stem_and_leaf_plot.html'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Q-Q Plot'	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	0	0	0	1	0	0	0	0	0	0	1	0
'Scatter Plot Matrix'	'SPLOM'	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	0	0	0	1	1	0	0	0	0	1	1	0
'Lollipop Chart'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	0	0	0	0	0	0	1	1
'Circle Timeline'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	0	1	1	0	1
'Connected	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	1	0	0	0	1	0	0	0	0	1	1	0

Scatter Plot'																
'Vertical Timeline'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	0	1	1	0	0
'Barcode Plot'	'Dot Strip Plot'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Seismogram '	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	0	0	1	0	1	1	0	1
'Stock Chart'	null	'http://www.anychart.com/chartopedia/chart-types/stock-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Spline Chart '	'A version of a line chart'	'http://www.anychart.com/chartopedia/chart-types/spline-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stepline Chart'	null	'http://www.anychart.com/chartopedia/chart-types/stepline-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stick Chart'	'A bar chart with the use of lines instead of bars'	'http://www.anychart.com/chartopedia/chart-types/stick-chart/'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Sparkline Chart'	'Combination of a line and column chart'	'http://www.anychart.com/chartopedia/chart-types/sparkline-chart/'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Jumpline Chart '	null	'http://www.anychart.com/chartopedia/chart-types/jump-line-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'PERT Chart'	'A type of workflow model'	'http://www.anychart.com/chartopedia/chart-types/pert-chart/'	0	0	1	0	0	0	1	0	0	0	0	0	1	0
'Area Graph '	'Area Chart'	'http://www.datavizcatalogue.com/methods/area#_graph.html'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stacked Area	'Stacked Area Graph'	'http://www.datavizcatalogue.com/methods/stacked_area_graph.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0

Graph '																
'Horizon Graphs'	'Surplus/ Deficit Filled Line'	'http://queue.acm.org/detail.cfm?id=1805128 and https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf'	0	0	0	0	0	0	0	0	0	0	1	1	1	0
'Fan Chart'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Cartesian Node-Link Diagram '	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	0	0	0	0	0	0	0	0	0	0	0	0
'Funnel Chart '	null	'http://www.anychart.com/chartopedia/chart-types/funnel-chart/'	0	0	0	0	0	0	0	0	0	0	0	0	0	0
'Voronoi'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	1	0	1	0	0	1	0	0	0	0	0
'Tree map'	null	'http://www.datavizcatalogue.com/methods/treemap.html'	0	0	0	1	0	1	1	0	0	0	0	0	0	0
'Circle Packing'	'Nested Circles '	'http://www.datavizcatalogue.com/methods/circle_packing.html'	0	0	0	1	0	1	1	0	0	0	0	0	0	0
'Range Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/range-area-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Range Spline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/range-splinearea-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Range Stepline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/range-steplinearea-chart/'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Stepline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/stepline-area-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Spline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/spline-area-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0

'Percent Stacked Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/percent-stacked-area-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Percent Stacked Spline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/percent-stacked-splinearea-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Percent Stacked Stepline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/percent-stacked-steplinearea-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Stacked Spline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/stacked-splinearea-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Stacked Stepline Area Chart'	null	'http://www.anychart.com/chartopedia/chart-types/stacked-steplinearea-chart/'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Polar Chart'	null	'http://www.anychart.com/chartopedia/chart-types/polar-chart/'	0	0	1	0	0	0	1	0	0	1	0	0	1	0
'Parallel Coordinates'	null	'http://www.datavizcatalogue.com/methods/parallel_coordinates.html'	0	0	1	0	0	0	1	0	0	0	0	0	1	0
'Flow Maps'	null	'http://www.datavizcatalogue.com/methods/flow_map.html'	0	1	0	0	0	0	1	0	0	0	0	0	0	1
'Choropleth Maps'	null	'http://www.datavizcatalogue.com/methods/choropleth.html'	0	1	0	0	0	0	0	0	1	0	0	0	0	0
'Graduated Symbol Maps'	'Proportional Symbal Maps (count/m	'http://queue.acm.org/detail.cfm?id=1805128'	0	1	0	0	0	0	1	1	0	0	0	0	0	1

	agnitude)'															
'Bubble Map'	null	'http://www.datavizcatalogue.com/methods/bubble_map.html'	0	1	0	0	0	0	1	0	0	0	0	0	0	1
'Connection Map'	'Connector Map'	'http://www.datavizcatalogue.com/methods/connection_map.html'	0	1	1	0	0	0	1	0	0	0	0	0	0	0
'Dot Map'	'Dot Density Map'	'http://www.datavizcatalogue.com/methods/dot_map.html'	0	1	0	1	0	0	1	0	0	0	0	0	0	0
'Contour Map'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	1	0	0	0	0	0	0	1	0	0	0	0	1
'Equalised Cartogram'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	1	0	1	1	0	1	0	1	0	0	0	0	0
'Scaled Cartogram'	''	'http://ft-interactive.github.io/visual-vocabulary/'	0	1	0	0	0	0	1	0	1	0	0	0	0	1
'Heat Map'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	1	0	0	0	0	1	0	1	0	0	0	0	0
'Cartogram'	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	1	0	0	0	0	1	0	0	0	0	0	0	1
'Radial Node-Link Diagram'	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	1	0	0	0	1	0	0	1	0	0	0	0
'Tree Diagram'	null	'http://www.datavizcatalogue.com/methods/tree_diagram.html'	0	0	1	0	0	0	1	0	0	1	0	0	0	0
'Icicle Tree Layout'	''	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	1	0	0	1	1	0	0	1	0	0	0	0
'Sunburst Diagram'	null	'http://www.datavizcatalogue.com/methods/sunburst_diagram.html'	0	0	1	0	0	1	1	0	0	1	0	0	0	0
'Force-Direct	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	1	0	0	1	1	0	0	0	0	0	0	1

ed Layout'																
'Arc Diagram'	null	'http://www.datavizcatalogue.com/methods/arc_diagram.html'	0	0	1	0	0	0	1	0	0	0	0	0	1	1
'Matrix View'	null	'http://queue.acm.org/detail.cfm?id=1805128'	0	0	1	0	1	0	1	0	0	0	0	1	1	0
'Bar Chart'	null	'http://www.datavizcatalogue.com/methods/bar_chart.html'	0	0	0	0	0	0	1	0	0	0	0	0	1	0
'Histogram'	null	'http://www.datavizcatalogue.com/methods/histogram.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Stacked Bar Graph'	null	'http://www.datavizcatalogue.com/methods/stacked_bar_graph.html'	0	0	0	0	0	0	1	0	0	0	0	0	1	0
'Population Pyramid'	'Back-to-Back Histogram'	'http://www.datavizcatalogue.com/methods/population_pyramid.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Radial Bar Chart'	null	'http://www.datavizcatalogue.com/methods/radial_bar_chart.html'	0	0	0	0	0	0	1	0	0	0	0	0	1	0
'Radial Column Chart'	null	'http://www.datavizcatalogue.com/methods/radial_column_chart.html'	0	0	0	0	0	0	1	0	0	0	0	0	1	0
'Bullet Graph'	null	'http://www.datavizcatalogue.com/methods/bullet_graph.html'	0	0	0	0	0	0	0	0	1	1	0	0	1	0
'Box-and-Whisker Plot'	'Box Plot or Box Chart'	'http://www.datavizcatalogue.com/methods/box_plot.html'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Ordered Bar/Column'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Paired Bar/C	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0

olumn'																
'Spine Chart'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Diverging Bar'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Proportional Stacked Bar'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	0	0	0	1	1
'Diverging Stacked Bar'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Priestley Timeline'	'Bar Graph and Timeline'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	1	1	0
'Line + Column'	'Pareto Chart'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	0	1	1	1	0
'Waterfall'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	1	0	0	0	1	0	0	1	0	0	1	0
'Range Column Chart'	null	'http://www.anychart.com/chartopedia/chart-types/range-column-chart/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Percent Stacked Column Chart'	null	'http://www.anychart.com/chartopedia/chart-types/percent-stacked-column-chart/'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Pie Chart'	null	'http://www.datavizcatalogue.com/methods/pie_chart.html'	0	0	0	0	0	0	1	0	0	1	0	0	1	1
'Donut Chart'	null	'http://www.datavizcatalogue.com/methods/donut_chart.html'	0	0	0	0	0	0	1	0	0	1	0	0	1	1
'Spiral Plot'	null	'http://www.datavizcatalogue.com/methods/spiral_plot.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Nightingale	null	'http://www.datavizcatalogue.com/methods/nightingale_rose_chart.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	1

Rose Chart'																
'Gantt Chart'	null	'http://www.datavizcatalogue.com/methods/gantt_chart.html'	0	0	0	0	0	0	1	0	0	1	0	1	0	0
'Span Chart'	null	'http://www.datavizcatalogue.com/methods/span_chart.html'	0	0	0	0	0	0	1	0	0	1	0	0	1	0
'Candlestick Chart'	'Japanese Candlestick Chart'	'http://www.datavizcatalogue.com/methods/candlestick_chart.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Parallel Sets'	null	'http://www.datavizcatalogue.com/methods/parallel_sets.html'	0	0	1	0	0	0	1	0	0	1	0	0	1	1
'Sankey Diagram'	null	'http://www.datavizcatalogue.com/methods/sankey_diagram.html'	0	0	1	0	0	0	1	0	0	1	0	0	1	1
'Chord Diagram'	null	'http://www.datavizcatalogue.com/methods/chord_diagram.html'	0	0	1	0	0	0	1	0	0	0	0	0	1	1
'Scatter Plot'	null	'http://www.datavizcatalogue.com/methods/scatterplot.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Bubble Chart'	''	'http://www.datavizcatalogue.com/methods/bubble_chart.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	1
'Venn Diagram '	null	'http://www.datavizcatalogue.com/methods/venn_diagram.html'	0	0	0	0	1	0	1	0	0	0	0	0	0	0
'Word Cloud'	null	'http://www.datavizcatalogue.com/methods/wordcloud.html'	1	0	0	0	0	0	1	0	0	0	0	0	0	1
'Brainstorm '	'Mind-Map'	'http://www.datavizcatalogue.com/methods/brainstorm.html'	0	1	1	0	0	0	1	1	0	0	0	0	0	0
'XY Heatmap'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	1	0	0	0	1	0
'Calendar Heatmap'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	1	0	0	1	1	0
'Marimekko Chart '	'Mosaic Plot '	'http://www.datavizcatalogue.com/methods/marimekko_chart.html'	0	0	0	1	0	0	1	0	0	0	0	0	1	0

'Pictogram Chart'	'Isotype Chart'	'http://www.datavizcatalogue.com/methods/pictogram.html'	1	0	0	0	0	0	1	1	0	1	0	0	0	0
'Point & Figure Chart'	null	'http://www.datavizcatalogue.com/methods/point_and_figure_chart.html'	0	0	0	0	0	0	1	1	0	0	0	1	1	0
'Dot Matrix Chart'	'Grouped Symbol'	'http://www.datavizcatalogue.com/methods/dot_matrix_chart.html'	0	0	0	0	0	0	1	0	0	1	1	0	0	0
'Kagi Chart'	null	'http://www.datavizcatalogue.com/methods/kagi_chart.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Timeline '	null	'http://www.datavizcatalogue.com/methods/timeline.html'	0	0	0	0	0	0	0	0	0	0	1	0	0	0
'Open-high-low-close Chart'	'"Stockprice" and OHLC Chart'	'http://www.datavizcatalogue.com/methods/OHLC_chart.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Network Diagram'	'Node-link Diagram'	'http://www.datavizcatalogue.com/methods/network#_diagram.html'	0	0	1	0	0	1	1	0	0	0	0	0	0	1
'Error Bars'	'Error Chart'	'http://www.datavizcatalogue.com/methods/error_bars.html'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Non-Ribbon Chord Diagram'	null	'http://www.datavizcatalogue.com/methods/non_ribbon_chord_diagram.html'	0	0	1	0	0	0	1	0	0	1	0	0	0	1
'Flow Chart'	null	'http://www.datavizcatalogue.com/methods/flow_chart.html'	0	0	1	0	0	0	1	1	0	1	0	0	0	0
'Illustration Diagram'	null	'http://www.datavizcatalogue.com/methods/illustration_diagram.html'	1	1	0	0	0	0	0	0	0	1	0	0	0	0
'Tally Chart '	null	'http://www.datavizcatalogue.com/methods/tally_chart.html'	0	0	0	0	0	1	0	0	0	1	0	0	0	0
'Density Plot'	null	'http://www.datavizcatalogue.com/methods/density_plot.html'	0	0	0	0	0	0	1	0	0	0	0	1	1	0
'Stream Graph '	null	'http://www.datavizcatalogue.com/methods/stream_graph.html'	0	0	0	0	0	0	1	0	1	0	0	1	1	1

'Proportional Area Graph'	'Proportional Symbol Chart'	'http://www.datavizcatalogue.com/methods/area_chart.html'	0	0	0	0	0	0	1	1	0	0	0	0	1	1
'Ordered Proportional Symbol Chart'	null	'https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf'	0	0	0	0	0	0	1	1	0	0	1	0	1	1
'Violin Plot'	null	'http://www.datavizcatalogue.com/methods/violin_plot.html'	0	0	0	0	0	0	0	0	0	0	0	1	1	0
'Radar Chart'	null	'http://www.datavizcatalogue.com/methods/radar_chart.html'	0	0	1	0	0	1	0	0	0	1	0	0	1	0
'Arc Chart'	'Arc Bar Graph'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	0	1	0	0	1	1
'Grid Plot'	'Dot Matrix Plot'	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	1	1	0	0	0	0	0	1	0
'Dot Strip Plot'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	1	0	0	0	1	0	0	1	0
'Slope'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	1	0	0	0	1	0	0	1	1	0	0	0
'Bump Chart'	null	'http://ft-interactive.github.io/visual-vocabulary/'	0	0	0	0	0	0	1	0	1	1	0	0	1	1
'Comic Strips'	'Story Board'	null	1	0	1	0	0	0	0	0	0	0	1	0	0	0
'Floor Plan'	null	null	0	1	0	0	1	0	1	1	0	0	0	0	0	0

Appendix D – Experts Testing Results

The testing results of the experts; the supervisor, Yuri Engelhardt and two data visualization designers, Nadieh Bremer and Jan Willem Tulp.

1. The Supervisor, Yuri Engelhardt.
The supervisor chose to think of his own variables. The supervisor suggested to change minor mistakes in the text which was used in the tool.

Step 1: 'The number of the bus lines' can be seen as a quantity, but it is meant to be a nametag. So, use the key word; unique identifier or unique name.

Step 2: The question 'How much?', does not contribute to all the quantitative variables that exist, so the question should be changed to 'How much? Or What is the numerical value?'.

The question 'What? And/or Where?' can be separated since the principles are clearly different, since there are a few exceptions, it was first thought of putting these questions together to form one data type, but there are two different types. 'What' is a broad universal question that can be used for any variable, therefore it has been decided to change it into 'What does it look like?', since this data type is 'Visual Appearance' and 'Where?' is a 'Spatial Location'.

Step 3: The formulation on how to guide the user to rank his variables should be changed, since knowing which variables are "central" can be tricky. So, users should rank the variables according to the message he or she wants to tell or the analysis of the variables he or she wants to make. The supervisor found the drag and drop functionality very intuitive.

Step 4: The highlighted options mean they are the ones selected and the preferred principle that the user should use. Considering the layout, there should be a space between the buttons 'next' and 'customize'.

Step 5: Make sure whenever the user opens a visualization that it opens a new tab instead of a new window. Concerning the test, take away the 'Finally', this word makes the process seem like it is exhausting to sketch. Just mention that there are existing types of visualizations that you can also use to visualise. It is not clear which variable is using which principle when suggestion the existing visualizations.

It has been discovered that the principle 'Connecting' can also be used for answering the question 'When?'.

Sketches of the supervisor:

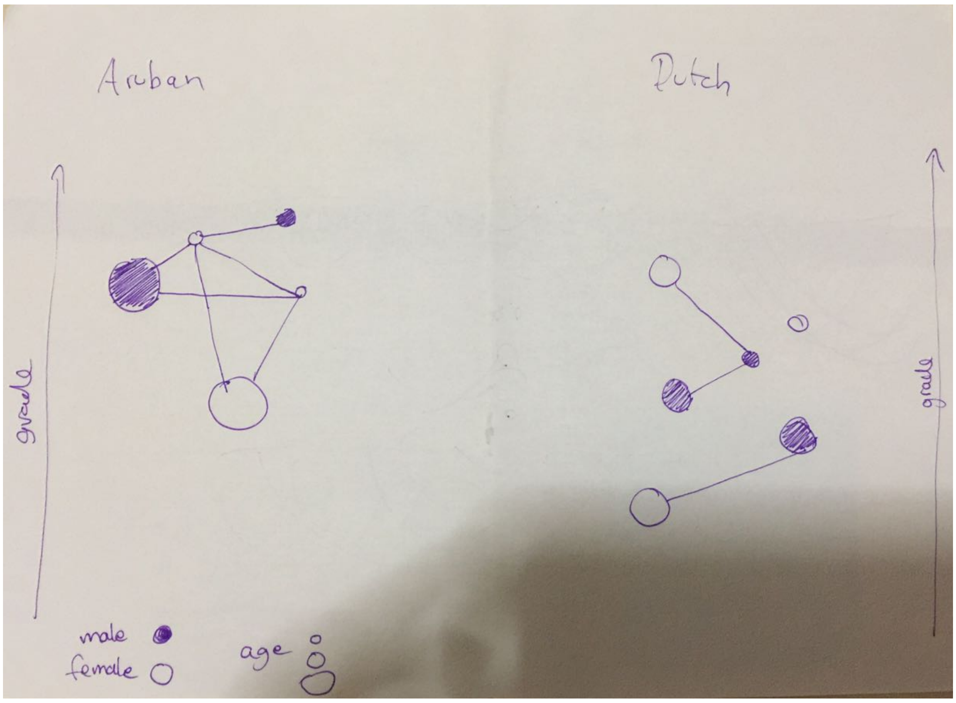


Figure 48: A Connected Spatial Bubble Chart

Feedback received from Nadieh

In step 2, you fill the drop down boxes with the questions, such as "does a relationship hold?" or "which category?" - I was actually quite confused at the start, I somehow felt like I had to answer those questions, but there was no way to answer them. Personally, I think I wouldn't have had any confusion if you had filled the drop down box with the actual data type options, such as "relationship","nominal", and then I would've looked to the right section of the page to get a better understanding of what these data types mean. Although I also feel the example questions can be improved on, because several of them don't help in my understanding of the data type. Such as for nominal. I don't quite know what to do with the sentence "Which category", instead I would've expected something more along the lines of "can the variable be split up into different groups?"

On step 4 you supply examples in the tips section, but I don't think that everybody knows what all of those charts mean. Perhaps you can create a hover over that shows a mini example of the chart when you hover over the chart's name, or that it is a link to the corresponding page on the dataviz catalogue. But the same also applies to some of the suggestions made on the left for the sketching. Although "grouping by color" is obvious, I don't really know what to do with "Attribute-Based Spatial Sequence", perhaps here you could also have a hover with a mini example

And that's also my suggestion for the final page in which you show links to websites for each suggested chart type. But to make it easier for people to assess, I think a hover with a mini chart would really help, because then people would not have to click each link to see how it looks, they can quickly hover over your suggestions and see approximately what it might look like before they decide to investigate further.

Feedback received from Jan Willem Tulp

On the first step, when clicking enter, I would expect the 'add variable' would be activated so that I can add another one, but it moved to the next step

For step 3, skipping or moving to the next step without making any modifications is basically the same, so you could get rid of the skip button

For the conclusion, I would display the links not as if they are in a textbox, but differently, so that they clearly look like links

I think sketching based on just a list of variables, but without a sample dataset is a bit difficult to do. Maybe you could generate a small sample dataset based on the choices people have made

I think that conceptually there is a bit of a gap between what the tool does (helping you decide on what visualuzation to use) and the knowledge you need to have to understand what the terminology is about. If you already understand what it means to nest data, what ordinal is, to use treemaps, etc. then you probably would not need this tool. So, my suggestion would be to be more explicit about what you're talking about, for instance by introducing images to illustrate concepts like nesting or treemaps

In my experience, one of the most challenging parts people have is to think creatively on how to map data to visual attributes. And there are many options, and many combinations possible. I personally would have emphasized that a bit more. So, that if you have a quantitative value, how can you represent that? What are different options? Etc.

Finally, what I would love to see a bit more is how your choices relate to the end conclusion and examples. So, for instance, I've what if I would have changed the order of the variables, what difference would that have made? If I would have chosen different roles of my variables, how would that differ? In the examples in the conclusion your provide, how is the variable type, role, order etc. used in these examples?

Appendix E – Universal Testing Results

All the notes taken during the universal testing results and the sketches of each participant.

Participant 1

- Thought the program will give him advise on how to visualise his data with known/existing visualization types
- Maybe change the order of steps
- Sketching at least 5-10 minutes, takes time for people to come up with something
- Preference for pie charts
- Giving feedback when users pick the wrong type of variable
- The tool was compared with “Excel’s” graphic wizard
- Add photos for visual feedback

Participant 2,3,4,5 and 6

- People were confused by the term ‘dataset’, some didn’t know what it was
- Also, some suggested to add a ‘delete’ button
- “Choose a dataset” is confusing, do they get options of datasets from the tool or do they have to choose this by themselves
- Words used should be more clear and simple
- The tool should also give feedback. When assigning the data types, users might make a mistake, but they do not know this, therefore feedback should be provided.
- Testing two people at once, is difficult, but interesting results are obtained.
- Amount of text is good
- Overview simple, but boring
- Clarify the point of the tool, or what is going to happen, what is expected. Most users expected the tool to tell them immediately which visualizations to use, the sketching part was a surprise.
- The principles are a bit too difficult to understand. A definition of their names has to be provided.
- Provide the option to also add more variables when assigning the datatypes
- Extra description is needed to explain the difference between a time axis and an attribute axis.
- People tend to always use grouping by colour, even though another principle was suggested. Subconsciously everyone used grouping by colour, sometimes even when it wasn’t mentioned at all. Or when they didn’t know the meaning of the other principles.
- The tool is helpful, not invasive
- Intuitive drag and drop function
- Confusing example of what not to use
- Telling people that there is still a big is completely fine
- People get the gist of the priniciples highlighted and not highlighted
- “Spatial grouping” is not an intuitive term
- Give examples on how to use the principles in combination, people don’t know that they can use two attribute axes, or two time axis, or two axis in general
- Even during the testings, I had to give feedback, since some participants were really doubting
- Some participants didn’t know the meaning of a variable, or what was meant by the term variable
- The goal of the tool was also not clear

Participant 7

- At the beginning the user didn't grasp what they must fill in as a variable
- The user also didn't know how he should rank his variables
- User suggested to show examples on how to use the tool, a video or tiny clips at the start of each step
- The goal is unclear: Different types of way to visualise data, to get different connections between datasets, since people tend to always use the same type?
- Add a search bar, to search for the terms

Participant 8

- Online test (older participant)
- Got the jist of the tool, only visualised one variable, which was doable.
- Some terms were unclear, and google had to be used in some occasions

Participant 9

- Online test
- It was difficult to understand, since all the online tests will completely performed alone
- User didn't know that the process was done/finished
- Also, user didn't know if everything she did was good, no feedback was given
- The main page doesn't provide enough information for a complete novice user

Participant 10

- Online test, younger novice user
- She suggests asking at the start the experience level of the user to give a more catered
- explanation or guidance depending on the experience level.
- print screen of process as example in the tips section

Participant 11

- A bit confusing at first as to what to write for variables. But the tips sections helped. Maybe more examples would be helpful
- A lot of technical terms which might be confusing to people who have never used these types of tools before.
- Step 3 is a bit confusing
- It would be better to give visual examples of this.
- I was very surprised by the amount of visualization possible for my type of data.
- The ones I drew are very simple and typical. Your tool gave me a surprisingly big amount of options.
- When I would be doing an actual data visualization I would use it to check what are my options for my type of data.
- For a future version of the tool it would be nice to have a small example of each type of proposition given.
- I would use the tool again since sometimes it is difficult to imagine all of the actual options out there for my specific type of data. Most of the time I just rely on what the data visualisation tool I am using can do but I rarely think about other options.
- Personally I did not find the system unnecessarily complex. What is a little trickier are the concepts. I understood them since I have experience with data visualization but I could

- imagine someone less experienced having troubles with it.
- The system felt very smooth. I did redo it but that was because I wanted to change the variables I choose.
- I did not find any inconsistency in the system.
- Someone with experience with data visualization tools will be able to pick this up in no time but I am not so sure about people with 0 experience with it.
- I felt confident using the system but I had to google some terms and it would have been easy with more visual tips and examples instead or in addition to the text.

Sketches of Participants:

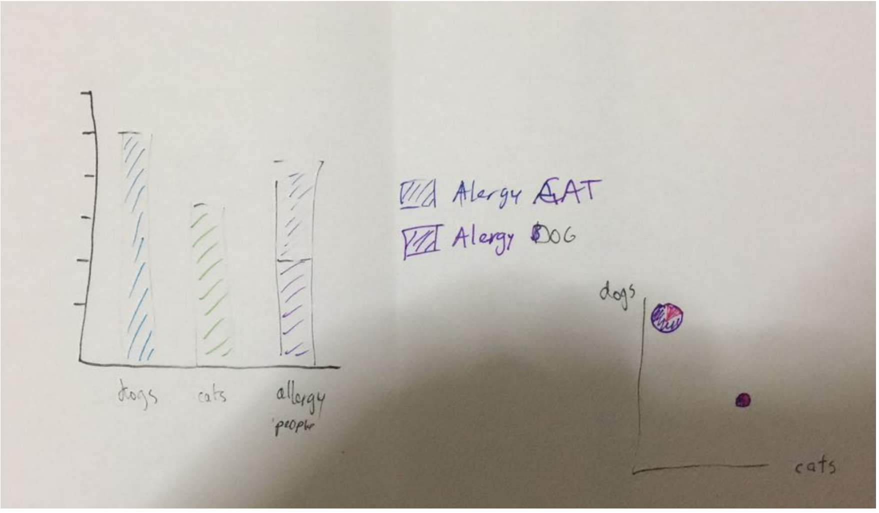


Figure 49: Animal Allergy Bar Chart and Bubble Chart

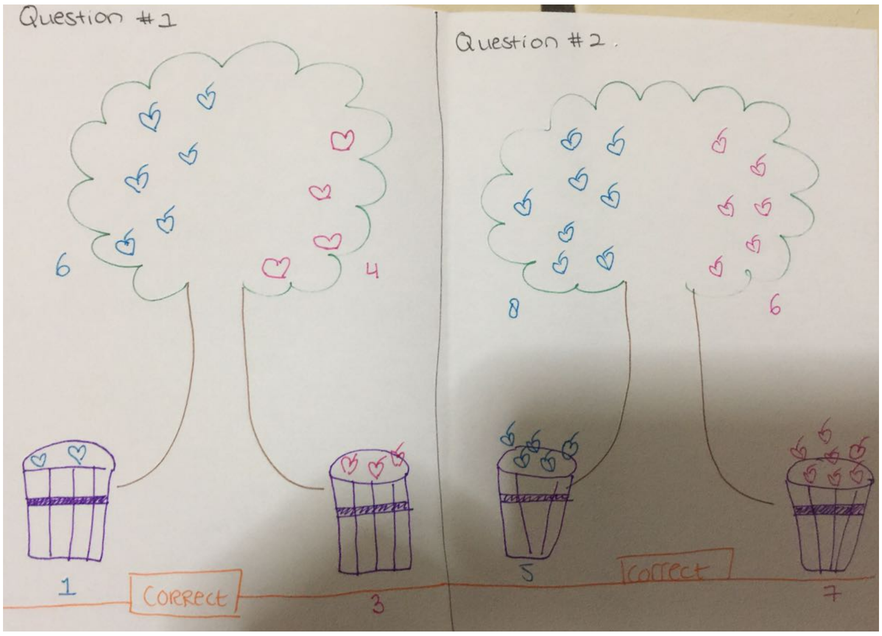


Figure 50: Tree Illustration representing female and male points achieved in a test

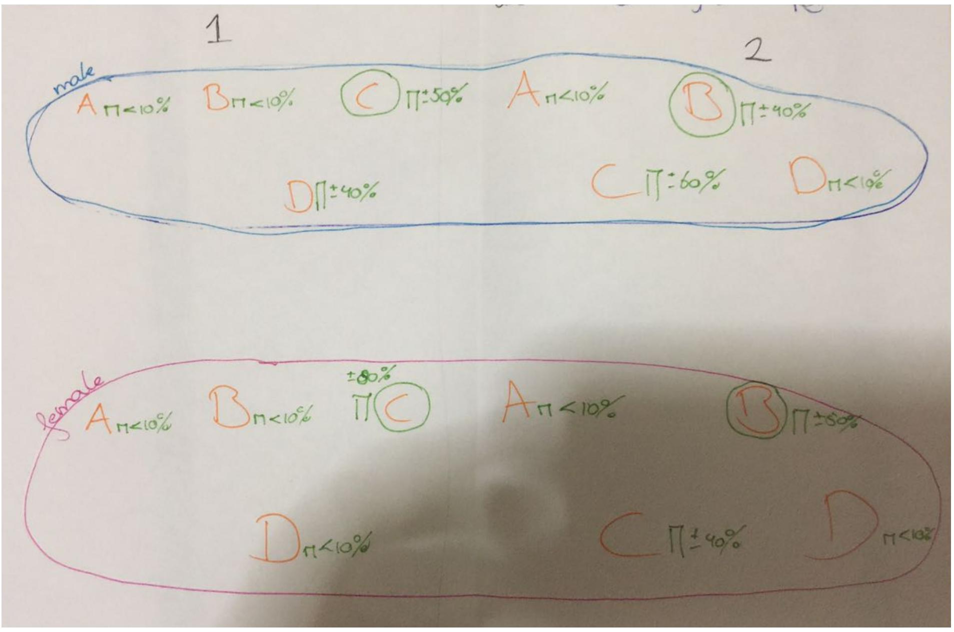


Figure 51: Female and Male Test Percentages Spatial Illustration

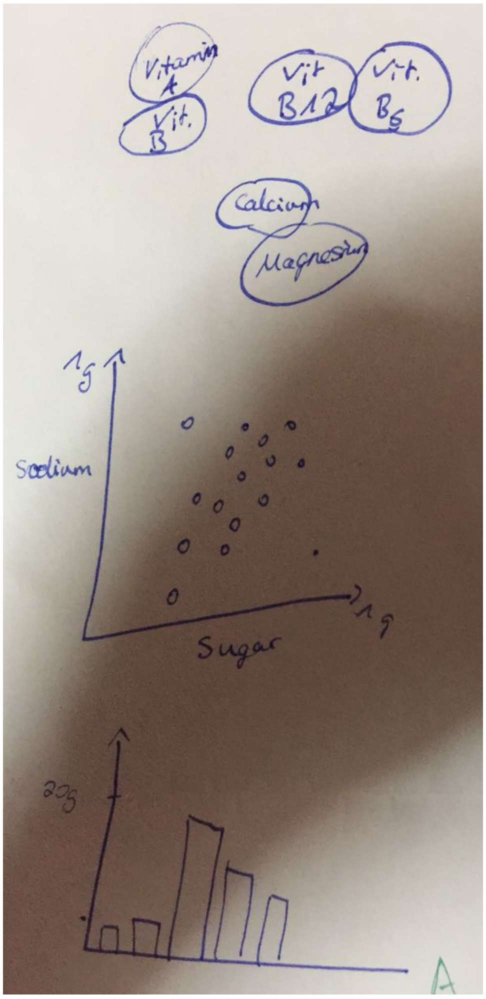


Figure 52: Vitamin Spatial Chart, Scatter Plot and Bar Chart

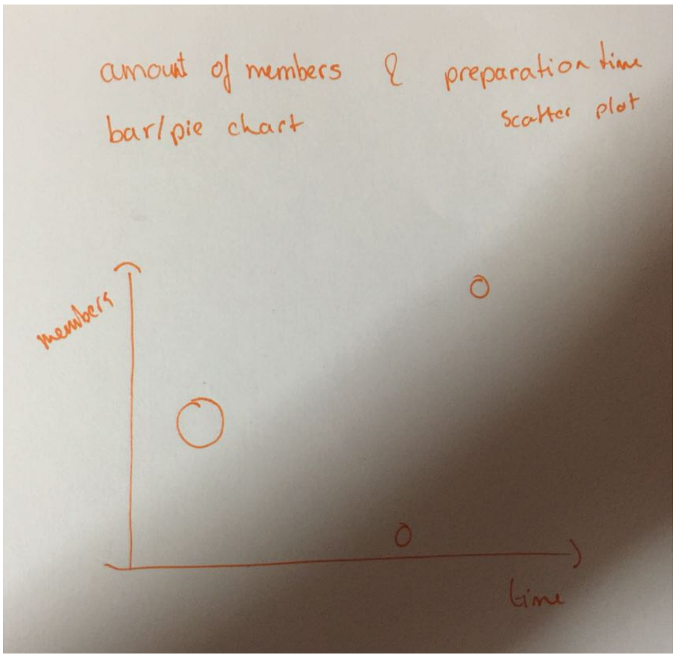


Figure 53: Bubble Chart representing time members take over an activity

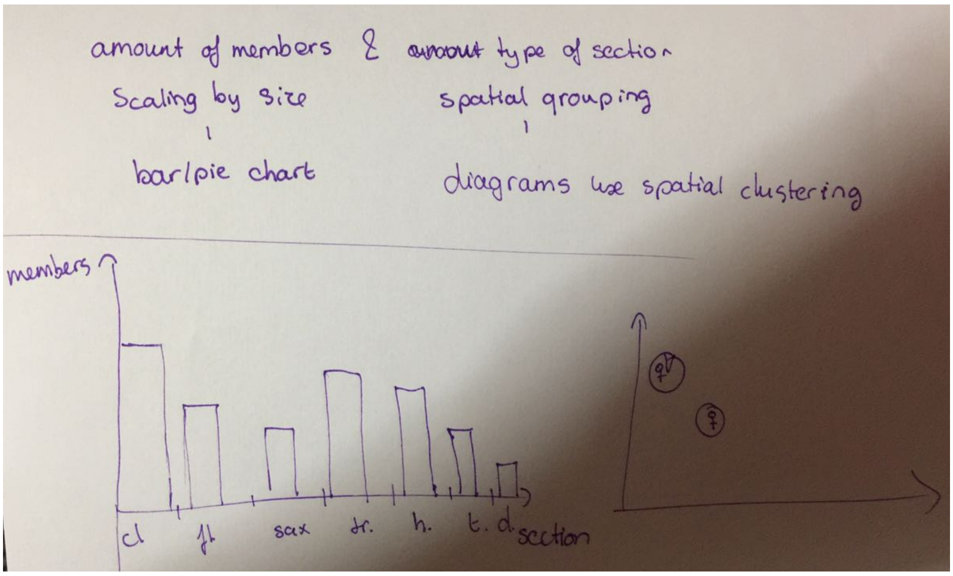


Figure 54: Bar Chart representing amount of people in each section of an orchestra

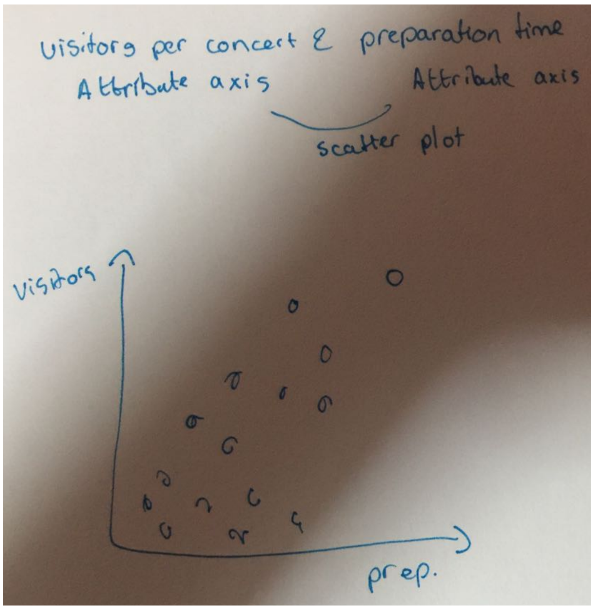


Figure 55: Scatter Plot representing the number of visitors for an event and the preparation time needed

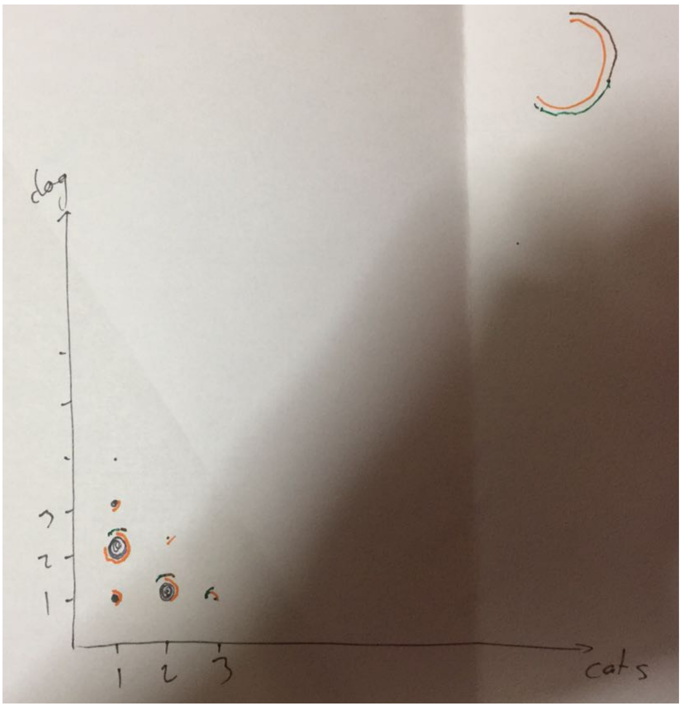


Figure 56: Radial Bubble Chart

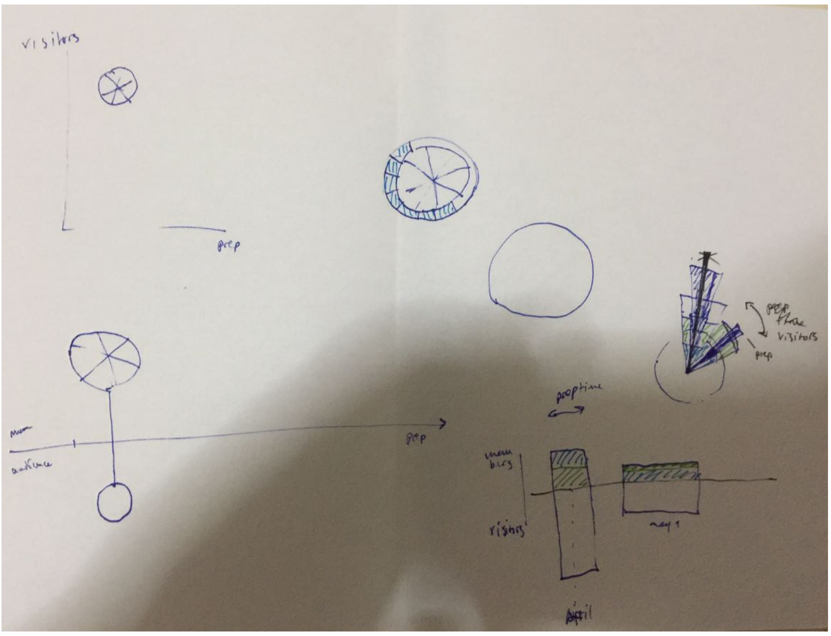


Figure 57: Several Sketches of different types of bar chart types

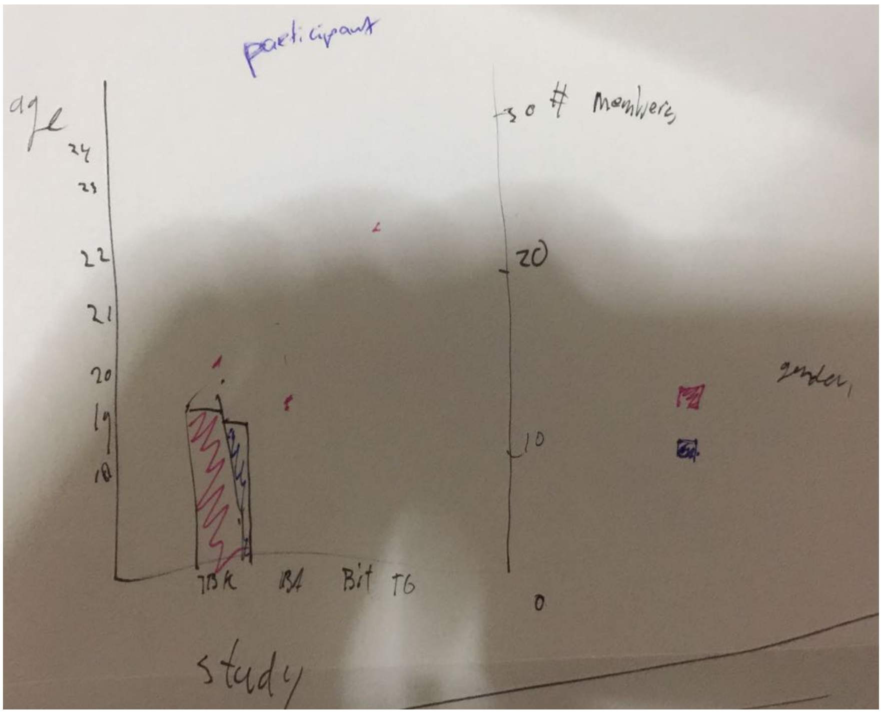


Figure 58: Simple Colorcoded Bar Chart

Appendix F – Usability Testing Consent Form

Usability Test Consent Form

Please read and sign this form.

In this usability test:
● You will be asked to perform certain tasks on a computer.
● You will be asked to fill in a questionnaire after performing the tasks.
● We will also conduct an interview with you regarding the tasks you performed.

Participation in this usability study is voluntary. All information will remain strictly confidential. The descriptions and findings may be used to help improve the DataVis Helper application. However, at no time will your name or any other identification be used. You can withdraw your consent to the experiment and stop participation at any time.

If you have any questions after today, please contact Roseidys Primera at r.c.m.primera@student.utwente.nl

I have read and understood the information on this form and had all of my questions answered

____________________________ ________________
Subject's Signature Date

Bibliography

[1] Ware.C., "Information Visualization: Perception for Design," Morgan Kaufmann Publishers, CA, 2000.

[2] T. Munzner, Visualization Analysis and Design, FL: CRC Press, 2014.

[3] C. R. a. Y. Engelhardt, "The Principles of Visual Encoding," in *The Principles of Visual Encoding*, TBA.

[4] S. Few, "Visual Business Intelligence," 8 January 2007. [Online]. Available: https://www.perceptualedge.com/blog/?p=81. [Accessed 1 August 2017].

[5] M. B. a. V. O. Jeffrey Heer, "A Tour through the Visualization Zoo," 2010.

[6] W. E. Angelika Mader, "A Design Process for Creative Technology," in *Proceedings of the 16th International conference on Engineering and Product Design*, Bristol, UK, 2014.

[7] A. Garcia, "UX Research | Standardized Usability Questionnaire," ChaiOne, 27 November 2013. [Online]. Available: https://chaione.com/blog/ux-research-standardizing-usability-questionnaires/. [Accessed 1 August 2017].