

# Making Sense of Multivariate Data: An Iterative Design of the Visualization and Interactions with Parallel Coordinates

Niklas Bruce and Nicolás Salas

DEPARTMENT OF DESIGN SCIENCES  
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY  
2020

MASTER THESIS

Qlik



---

# **Making Sense of Multivariate Data:**

An Iterative Design of the Visualization and Interactions  
with Parallel Coordinates

Nicolás Salas and Niklas Bruce



**LUND**  
UNIVERSITY

---

# Making Sense of Multivariate Data: An Iterative Design of the Visualization and Interactions with Parallel Coordinates

Copyright © 2020 Nicolás Salas and Niklas Bruce

*Published by*

Department of Design Sciences  
Faculty of Engineering LTH, Lund University  
P.O. Box 118, SE-221 00 Lund, Sweden

*Publicerad av*

Institutionen för designvetenskaper  
Lunds Tekniska Högskola, Lunds universitet  
Box 118, 221 00 Lund

Subject: Interaction Design (MAMM01)

Supervisor: Christofer Rydenfält

Co-supervisor: Quan Ho, Qlik

Examiner: Kirsten Rasmus-Gröhn

Ämne: Interaktionsdesign (MAMM01)

Huvudhandledare: Christofer Rydenfält

Bitr. handledare: Quan Ho, Qlik

Examinator: Kirsten Rasmus-Gröhn

---

# Abstract

There exists today vast accumulations of data that we don't always know what to do with. Qlik is a company that specializes in the visualization of data and specifically in being able to offer as many different types of charts as possible to their customers. Parallel coordinates is a chart that efficiently can visualize multivariate data and this master thesis is about developing such a chart for Qlik, using techniques from interaction design and with a focus on usability.

Using techniques like user interviews, focus groups and usability tests the needs of the user were elicited and a parallel coordinates diagram could be developed. The results showed that a parallel coordinates plot could be very effective in visualizing multivariate data but also that it might not always be the most optimal chart for all data sets. The resulting product had basic functionality with some additional tools for further analysis.

In this project it was established that there are many factors to think about when deciding how the data should first be presented for the user, which tools that should be implemented and how the user should interact with said tools. Some of those factors may be, who the users are, their background, level of competence and type of data they wish to visualize.

**Keywords:** Parallel Coordinates, Interaction, Chart, Visualization.

---

# Sammanfattning

Idag existerar det stora samlingar av data som man inte riktigt vet vad man ska göra med. Qlik är ett företag som specialiserar sig inom visualisering av data och erbjuder så många olika typer av diagram som möjligt till sina kunder. Parallella koordinater är ett diagram som på ett effektivt sätt visualiserar multivariat data och detta examensarbetet handlar om att utveckla ett sådant diagram åt Qlik och med tekniker från interaktionsdesign lägga fokus på användarvänlighet.

Genom att använda tekniker som användarintervjuer, fokusgrupper och användartester så eliciterades användarnas behov och ett parallella koordinater diagram kunde utvecklas. Resultaten visade att parallella koordinater kan vara väldigt effektiva när man vill visualisera multivariat data, men också att parallella koordinater inte alltid är det bästa alternativet för alla dataset. Den slutgiltiga produkten hade grundläggande funktionalitet med tillägg för vidare analys.

I detta projekt så fastställdes det att det finns många faktorer att tänka på när man ska bestämma hur data ska presenteras för användaren, vilka verktyg man ska använda och hur användaren ska interagera med verktygen. Några av dessa faktorer är, vem användarna är, deras bakgrund, kompetensnivå och vilken typ av data de vill visualisera.

**Nyckelord:** Parallella Koordinater, Interaktion, Graf, Visualisering.

---

# Acknowledgments

Many thanks to Christofer Rydenfält, our supervisor at LTH for his patience and continued support throughout all our work.

Many thanks as well to Quan Ho and Anders Fridell who gave us a warm welcome into Qlik and helped us with whatever we needed during our time there.

We would also like to thank all the other Qlickies who helped us. It could have been with participating in tests, pondering over bugs or just sharing their expertise, we could not have done it without all of you.

And lastly our friends and families who were always there when we needed to frustratingly talk about what did not work and they just listened and made us feel better.

Lund, May 25, 2020

Nicolás Salas and Niklas Bruce

# Contents

---

	<b>Page</b>
<b>1 Introduction</b>	<b>10</b>
1.1 Background . . . . .	10
1.2 Qlik . . . . .	13
1.2.1 Qlik Sense . . . . .	13
1.3 Goals of the Thesis . . . . .	14
1.4 Work Distribution . . . . .	14
<b>2 Theory</b>	<b>15</b>
2.1 Interaction Design . . . . .	15
2.1.1 Norman´s Seven Principles of Design . . . . .	15
2.1.2 Usability Goals . . . . .	16
2.1.3 Iterative Design Process . . . . .	16
2.1.4 User interview . . . . .	17
2.1.5 Photo-Elicitation . . . . .	17
2.1.6 Focus Group . . . . .	18
2.1.7 Pugh Matrix . . . . .	18
2.1.8 Usability Test . . . . .	18
2.2 Prototyping . . . . .	21
2.2.1 Lo-fi Prototype . . . . .	21
2.3 Visualization . . . . .	22
2.3.1 Visualizing Multivariate Data . . . . .	22
<b>3 Literature Study</b>	<b>23</b>
3.1 Method . . . . .	23
3.2 Results . . . . .	26
3.2.1 Parallel Coordinates . . . . .	27
<b>4 User Study</b>	<b>33</b>
4.1 Method . . . . .	33
4.1.1 Preparation . . . . .	33

---

4.1.2	The Interview . . . . .	34
4.1.3	Compiling the Results . . . . .	37
4.2	Results . . . . .	38
4.3	Analysis & Discussion . . . . .	44
<b>5</b>	<b>Concept Generation &amp; Focus Group</b>	<b>45</b>
5.1	Concept Generation . . . . .	45
5.1.1	Method . . . . .	45
5.1.2	Results . . . . .	46
5.1.3	Analysis & Discussion . . . . .	48
5.2	Focus Group . . . . .	49
5.2.1	Method . . . . .	49
5.2.2	Results . . . . .	50
5.2.3	Analysis & Discussion . . . . .	52
5.3	Final Concept . . . . .	53
<b>6</b>	<b>Application Development</b>	<b>54</b>
6.1	Theory . . . . .	54
6.1.1	Picasso.js . . . . .	54
6.1.2	Nebula.js . . . . .	54
6.2	Method . . . . .	55
6.3	Resulting Prototype . . . . .	57
6.4	Analysis & Discussion . . . . .	61
<b>7</b>	<b>Focus Group &amp; Usability Test</b>	<b>62</b>
7.1	Focus Group . . . . .	62
7.1.1	Method . . . . .	62
7.1.2	Results . . . . .	64
7.2	Usability Test . . . . .	67
7.2.1	Method . . . . .	67
7.2.2	Results . . . . .	70
7.3	Analysis & Discussion . . . . .	73
<b>8</b>	<b>Final Analysis &amp; Discussion</b>	<b>75</b>
8.1	Methodologies Used . . . . .	75
8.1.1	User Interview . . . . .	75
8.1.2	Focus Group . . . . .	76
8.1.3	Usability Test . . . . .	76
8.2	Sources of Error . . . . .	77
8.3	Research Questions . . . . .	79
8.4	Future Work . . . . .	80
<b>9</b>	<b>Conclusion</b>	<b>81</b>
	<b>References</b>	<b>82</b>
	<b>Appendix A Manuscript of the User Study</b>	<b>86</b>

---



<b>Appendix B Concepts One, Two &amp; Three</b>	<b>93</b>
<b>Appendix C Requirements Specification</b>	<b>96</b>
<b>Appendix D Manuscript of the Usability Test</b>	<b>102</b>

# Chapter 1

## Introduction

---

This chapter will introduce the reader into the project. We give a background into multivariate data and parallel coordinates as well as to why this work was needed, all of this with an emphasis on interaction design. Also introduced are the company that facilitated the project as well as their product that served as a basis for the work. Lastly the goals and the scope for the thesis will be presented.

### 1.1 Background

Data is becoming a bigger part of everyone's lives as the day progresses. Big companies are in international scandals about buying or selling data [1] and data even plays a part in central plot points in modern TV shows and movies[2]. Data is all around us, even if we don't notice it. But data in itself can have very little value if one does not know how to read it, interpret it, analyze it and understand it. For this one would need to be fluent in what is called data literacy[3], the language of understanding data and knowing what to do with it. At the same time, having the perspective of interaction design and trying to find usable ways to visualize and interact with data that at the same time provide a good user experience, is also key. There are many ways of visualizing and conveying data since there are many different types of data and as data becomes more complex the comparison of data with large amounts of variables becomes more common and vital.

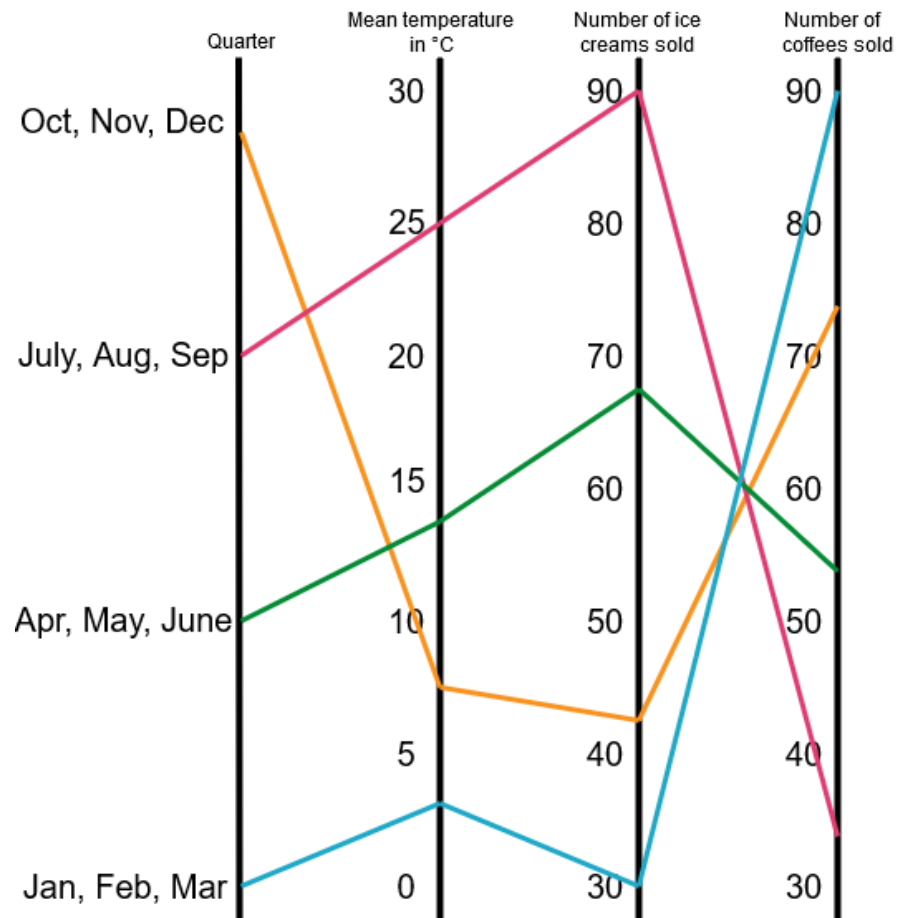
Multivariate data is data where the analysis is based on more than two variables. Bivariate data analysis, data where the analysis is based on two variables, can get you far but may not show the whole truth. An example: imagine a scatter plot showing that the sales of ice cream increase the higher the temperature is on any given day at a lakeside store. The correlation may be true but it maybe doesn't paint the whole picture. For example, there could be a correlation to big groups of families going to the lake during

the summer months because people tend to take their vacations on those months. It just happens to be that they take their vacations during warm months. More people coming to the lake means more sales for the store but they are not selling more because it is warm, it just happens to be so. In general more variables can paint a clearer picture, but it is not always easy to compare and understand multivariate data. One diagram developed for the visualisation of multivariate data is parallel coordinates[4].

Parallel coordinates are a common way to visualize and analyze multivariate data. An example of an over simplified parallel coordinates plot can be seen in figure 1.1 for better understanding of the concept. There are two or more axes that are parallel to each other and there is one axis per variable. Each axis's scale can have their own range or unit depending on the variable. This makes parallel coordinates handy to compare very different things in one chart. Values for each variable are shown as connected *polylines* between the axes. Furthermore the order in which the axes are arranged can greatly impact ones understanding of the chart. This is because the relationships between adjacent variables are easier to perceive than for non-adjacent variables.

The example shown in figure 1.1 shows the correlation between how much ice cream and coffee are sold at an imaginary store during the four quarters of a year. Also one can see the mean temperature during each of the quarters so that one can see a correlation between the mean temperature and numbers of either ice creams or coffees sold.

Parallel coordinates are also handy in that you can both find outliers in the polylines or see entire patterns from many polylines and draw conclusions from them. Parallel coordinates plots can be very powerful tools but when handling big amounts of data the charts can get cluttered very fast. There have been many features developed that either try to counter the clutter or help facilitate the understanding of the data in a parallel coordinates plot. Some of these techniques and tools are discussed in more depth in later chapters.



Data			
Quarter	Mean temperature in °C	Number of ice creams sold	Number of coffees sold
Oct, Nov, Dec	7	42	72
July, Aug, Sep	25	90	33
Apr, May, June	14	67	53
Jan, Feb, Mar	3	30	89

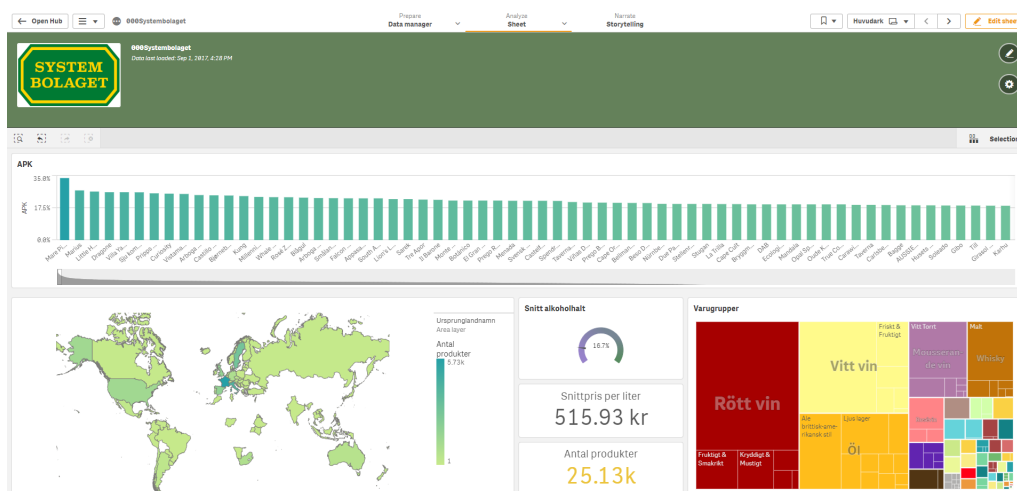
**Figure 1.1:** An example of a simple parallel coordinates plot.

## 1.2 Qlik

Qlik[5] is a software vendor company founded in 1993 in Lund, Sweden. Qlik specialises in data visualization, data integration and user-driven business intelligence products.

### 1.2.1 Qlik Sense

While Qlik as a company has many products this thesis will only concentrate on the one that is interesting for the project, Qlik Sense. Qlik Sense is a self service analytics platform. Qlik provides both a downloadable desktop version and a web based version. The software has an analytics engine that helps the user sort their data and choose what could be of interest for them to analyse. Qlik Sense then has over fifteen different charts to help the user visualise their data. Some of the charts that they have are: bar charts, pie charts, treemaps, scatterplots and many more. The user can with simple drag and drop tools build a visualisation page that matches what the user needs. After the desired charts are in place the user can interact with the charts and see in real time reaction from all the charts.



**Figure 1.2:** An example of how Qlik Sense can look by showcasing a sheet with test data from Systembolaget.

## 1.3 Goals of the Thesis

The main goal of the thesis is to make the basis of a new chart for Qlik Sense in the form of a parallel coordinates plot. This should be developed by using iterative interaction design techniques so as to develop something that the potential user of the chart would be happy with.

Another aim of this master thesis is to analyze and review how we can make parallel coordinates as perspicuous as possible. Also to make an assessment as to which of the required features, and other features we might find along the way, are the optimal tools to be used in conjunction with parallel coordinates to be able to facilitate the understanding of the data visualized. We hope to reach these goals by answering these research questions:

- **RQ-1:** What is important to think about when choosing which type of visualization tool to use when visualizing multivariate data?
- **RQ-2:** What is important to think about when visualizing parallel coordinates?
- **RQ-3:** How does the choice of coloring influence the users' understanding of the data shown in a parallel coordinates chart?
- **RQ-4:** Which feature is the most important to ease the users' understanding of a cluttered parallel coordinates chart?

## 1.4 Work Distribution

Both of the authors to this master thesis had during the entirety of the project an even distribution of the work in all its stages. The only time when their work differed was during the user study, the focus groups and the usability tests. In those instances Nicolás served as moderator and Niklas took notes.

# Chapter 2

## Theory

---

This chapter will present the background of some of the main theories and techniques, within interaction design and visualization, that were used during the project.

### 2.1 Interaction Design

The aim of interaction design is to create user experiences that augment and enhance the way people work, interact and communicate. Interaction design is not just one way of doing things, there are a lot of different frameworks, methods and techniques one can use in interaction design and which processes are used depend on the context [6].

#### 2.1.1 Norman´s Seven Principles of Design

Norman´s seven design principles are commonly used in interaction design. The aim of the principles is to create good user experience. [7]

1. **Discoverability:** the ability to discover what actions can be performed on a system by the user.
2. **Feedback:** after a performed action the system should send information about the new state of the system to communicate a change to the user.
3. **Conceptual Model:** a conceptual model is an explanation of how something works and it is usually highly simplified. As long as the explanation is useful it does not have to be complete or accurate.
4. **Affordances:** An affordance refers to a potential action in a system, for example turning a knob.

5. **Signifiers:** a signifier are used to signify where an action takes place. Using signifiers in an effective way, ensures discoverability and makes sure that feedback is communicated to the user in a way that is intelligible.
6. **Mappings:** the relationship between the elements in two sets of things is called mapping. Take for example set switches that pull down the blinds for windows, the mapping would specify which switch pulls down which of the blinds.
7. **Constraints:** A constraint is a clue that limits the set of possible actions the user can perform, a constraint can be:
  - Physical, a battery needs to be properly oriented.
  - Cultural, each culture has a set of actions that are allowed for social situations.
  - Semantic, relies on the meaning of the situation, for example on a motorcycle there is only one meaningful place for the rider to sit.
  - Logical, relies on the logic in the system, for example the left light switch would control the left lamp and the right light switch would control the right lamp.

## 2.1.2 Usability Goals

Usability has the aim to make sure that an interactive product is enjoyable, easy and effective to use. To achieve this there are a set of goals set by Preece et al [6]:

1. **Effectiveness:** is the measure of how well a product performs the task it is supposed to perform.
2. **Efficiency:** is the measure of how well a product supports the user in performing the tasks.
3. **Safety:** refers to protecting the user from conditions that could be dangerous and unwanted.
4. **Utility:** is to which extent the product gives the user functionality to do what they need or want to do.
5. **Learnability:** is the measure of how easy it is to learn how to use a new system.
6. **Memorability:** is the measure of how easy it is for the user to remember how to use the system.

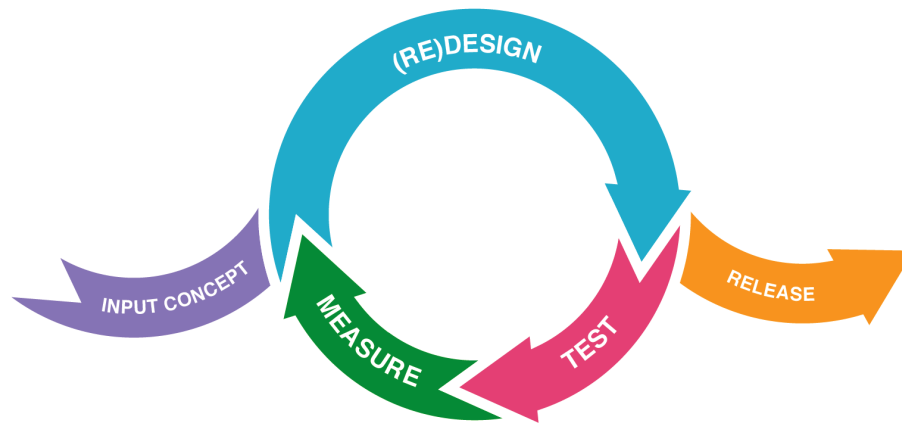
## 2.1.3 Iterative Design Process

Iterative design means that the development of the product proceeds in iterations. The iteration design cycle with its different phases can be seen in figure 2.1.

This iteration design cycle lets the developers redesign the product in both early and late stages depending on what the potential user wants and needs. At the same time, with this process, big mistakes should be noticed early while the cost of making mistakes is low. Later in a developing process when mistakes cost more time and resources



the perceived flaws by the potential users should be smaller and easier to rectify. Using iterative design allows designers to redefine the design in regards to the response received from tests, interviews and other forms of feedback and data gathering [6].



**Figure 2.1:** The iteration design cycle.

### 2.1.4 User interview

A user interview is a method in user experience (UX) research where the researcher asks a potential user questions about certain topics to learn more about those topics. The topics can for example be, use of a system, habits and behaviours. User interviews can be useful for many reasons. They help the researcher understand what the users think about different things, for example a product or a process. Depending on where in the development process you perform the interview you get different kinds of data. If the interview is performed in the beginning of the project before you have a design, then the goal is to gather data that will help you make good design choices. Whereas if the interview is performed at the end of a usability test, then the goal is to collect more information about how the user perceived and experienced the product and why they interacted with the product in the ways that they did [8].

### 2.1.5 Photo-Elicitation

Images invoke deeper elements of the human consciousness than words do. This is because the part of the brain that processes images is, from an evolutionary perspective, older than the part that processes verbal information [9]. Photo elicitation is the idea of inserting an image in to a research interview to elicit comments from the participants. Photo elicitation can add reliability and validity to a word based survey. However, only using images won't automatically produce good results. To create good results the images need to make the person interviewed reflect [9].

## 2.1.6 Focus Group

A focus group is like an interview with more people. Normally a focus group consists of three to ten people with varied demographics but in a design process one would prefer for the people to be future users of the product. In a focus group the discussion is led by a facilitator [6]. One of the benefits of a focus group is that it lets diverse issues be raised that one might otherwise miss, since the aim of the focus group is to provide a supportive environment for the participants to discuss and in that way enable them to put forward their ideas and opinions. However, a focus group can have disadvantages or aspects that one needs to take into consideration. When deciding on which participants to have as a part of the group one has to give thought to the the social skills of the participants. If they lack the social skills needed then they might not bring to light opinions they perhaps would have shared if it was a one-on-one interview. Also, having one or several participants that are too dominant could be a disadvantage since they might not let the other participants talk.

## 2.1.7 Pugh Matrix

A Pugh matrix, is a matrix diagram that lets the designers query a user, by letting the user compare designs and choose which fits the set criteria the most [10]. A Pugh matrix is created and carried out in the manner described below:

- Defining the criteria that will be used to compare the different designs.
- Decide upon one design to use as the baseline for the comparison.
- Compare each design against the baseline design, one criteria at a time.
- Score the criteria with a + if its better than the baseline, an S if it is as good as the baseline or a – if it is worse than the baseline.
- See which design had the best score by counting the + and – signs.
- Discuss potential hybrids between the concepts.
- Decide on which design or design hybrid to choose and develop.

## 2.1.8 Usability Test

Usability testing is a process where participants that are either taking on the role of a user or are real future users of a product evaluate a product to see if the project meets the requirements [11]. The goal of a usability test is to gather data from the testers and using this data, inform the designers of usability problems in the product. According to Rubin and Chisnell, usability tests are a great way to make sure that a product has the following metrics:

- The product is useful to and valued by the target audience.
- The products learning curve is not too steep.
- The product helps people to be effective and efficient at what they want to accomplish.

- The product is satisfying to use

## Test Plan

A test plan is the foundation of a usability test. It describes how the developer will perform the test, with whom, where and why [11]. Depending on the type of test and how formal the test will be, the test plan can vary, but there are some standards that test plans usually contain. These are:

- The goals, purpose and objectives.
- Participants characteristics.
- Task list.
- Test moderator role.
- Report contents and presentation.
- Research questions.
- Method.
- Test environment.
- Data to be collected.

The task list contains the tasks that the tester will perform during the test. When testing in late stages of the development the tasks will provide realistic details and context so the the testers can understand and perform the tasks with little intervention from the test moderators.

## Post test Questionnaire

After the test is completed the tester is often requested to fill out a post test questionnaire. The goal of the post test questionnaire is to collect information from the testers to get a clearer knowledge about the product's weaknesses and strengths [11]. An important part of the questionnaire is to make sure that the questions are asked in the same way to all testers. When designing the questionnaire the content and the format are the two biggest considerations. The content is about the subject you want to gather data about and the format has to do with the wording and arrangement of the questions. By keeping the questions concise and to the point one keeps the level of abstraction low [11].

## Likert Scale

Post test questionnaires can have both qualitative questions, in the form of open text responses, and quantitative questions in the form of scales, amongst others. Such a scale is the Likert Scale where the tester is asked a question and the predetermined answers the tester can choose from varies in a n-point scale from disagree to agree or vice versa. The number n can vary from three to large numbers but five is a standard number for a Likert scale to have.

A standard way of representing a five point Likert scale is:

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

The answers of the users can then be quantified on a 1-5 scale [11]. This makes comparison of testers preferences easier.

## 2.2 Prototyping

A prototype is an early version of a product [12]. It represents a design idea and by making a prototype you can find problems early on in the developing stage and correct them. They are most commonly divided into either Lo-fi (low fidelity) or Hi-fi (high fidelity) prototypes. A lo-fi prototype is a prototype with low level of detail [12] while a Hi-fi prototype has a much higher level of detail and thus should look and feel closer to the proposed final product.

### 2.2.1 Lo-fi Prototype

The definition of a Lo-fi prototype is "an artefact that has been created independently from the likely form of a finished product" [13]. This means that instead of focusing on how and with which tools and methods the prototype will be implemented, the designer can instead focus on the design ideas. Lo-fi prototyping lets the designer make quick and easy changes to the design and Lo-fi prototypes also lets the designer address features or ideas that current technology might not be able to implement. One way of developing a lo-fi prototype is using "Paper Prototyping" which utilizes paper, post it notes and index cards to create a representation of the system.

## 2.3 Visualization

Visualization, is a cognitive activity where people build an internal mental representation of the world, facilitated by external visual representations. Data on its own has very little meaning. The data must first be processed, organized and then presented in meaningful way to be utilized to its maximum value. With the right process the once useless data, becomes useful information [4]. There are some criteria that are good to keep in mind when one hopes to create good visual representations:

- The visualization needs to be effective. It should communicate as much information as possible with the least amount of effort [4].
- The visualization needs to have integrity, meaning that the visualization should not distort the data or in other ways make it possible to misinterpret the meaning.
- The visualization should not contain unnecessary elements, all the elements should be relevant to conveying the message of the visualization and extra artistic elements like backgrounds or 3D effects can distract the person trying to read the data.
- The visualization should also have "elegance", meaning that it should have great level of detail, avoid unnecessary decorations and also have the proper format and design.

### 2.3.1 Visualizing Multivariate Data

As mentioned in the introduction, multivariate data is data with more than two variables. This kind of data is very common since most real problems depend on a large number of different variables that needs to be analyzed [4]. Parallel coordinates are not the only way of visualising multivariate data but are one of the few that can visualise virtually unlimited amounts of variables. If we compare it to the scatter plot we can see where other types of graphs falter. If one wishes to visualize more than two dimensions with a scatter plot one can either give the graph another dimension and have a 3D scatter plot or one could stack entire scatter plots to get a more complex view by comparing entire graphs side to side. This gets very convoluted and if one adds more variables and have a defined screen size the entire graph has to shrink to give place to the new variable. In comparison, when using parallel coordinates one would only need to keep adding one more axis for every added variable.

# Chapter 3

## Literature Study

---

This chapter will present the literature study where the goal was to to gain knowledge within the areas of parallel coordinates and visualization of multivariate data to increase the effectiveness of the work to come and increase the quality of the data gathered.

### 3.1 Method

The literature study was the first part of the master thesis. This phase began with setting up a number of questions that needed to be answered to get an understanding of parallel coordinates as a science as well as how it is viewed from a user's perspective. The goal of this phase was to understand what is available in the field of parallel coordinates and also which tools and features that can make parallel coordinate plots easier to interpret and analyse. To help us do this step we brainstormed [6] forth what was considered to be valuable questions to be answered.

The questions to be answered were the following:

- Where is the science right now on parallel coordinates?
- Are there any unanswered questions still in the field of parallel coordinates?
- How can parallel coordinates be visualized?
- What are the benefits and drawbacks of different visualization methods in terms of usability?
- Is there a big difference in how you visualize big data compared to normal data in parallel coordinates?
- What types of features are used right now in conjunction with parallel coordinates to make them easier to understand?

To begin answering the questions the first thing that was done was to go through an article from 2016 by Johansson et al. [14] about parallel coordinates and their usability purely from a user's perspective. This article was a review of many studies that had been done about parallel coordinates that were all based on user testing of varied sizes, ranging from 9 to 1687 participants. The article painted a picture of some of the features that are available to facilitate understanding of the data to the user. Many of the features discussed were of great interest to the project and it was decided to dive deeper into some of the research papers that the article referenced. We started by reading the first pages of some of the articles and it was quickly apparent that the older the text was the less relevant it was since the newer ones also included the older ideas, which the newer ones were built upon. The articles' sources were listed after their relevancy to the master thesis, and because of what was just discussed, that was guided mostly by year of publication. Using *Google Scholar* and similar sites to it we tried to find free versions of as many of the relevant articles as possible.

Another method used to look for new articles was to utilize the search tool *Google Scholar*. When using the search engine we needed to use the correct search strings so that we could elicit the correct type of articles. Below is a list of the search strings we thought would find articles that could answer the questions we had posed:

- "Parallel Coordinates".
- "Parallel Coordinates" AND "Visualization".
- "Parallel Coordinates" AND "Big Data".
- "Parallel Coordinates" AND ("Tools" OR "Features").

Using this search strings we found four more relevant articles. There were many more articles found but they were either very outdated, not relevant or they were already on the list since they also were a source in Johansson Et al. [14].

In total, by using the snowball method of searching we got the seven most relevant references from the original text from 2016 by Johansson Et al. [14]. Those seven articles together with the original article in itself and the four articles that were found through *Google Scholar* built up the foundation for the literature study. Twelve articles may seem like a small number of articles based on the searching we did and there were many articles that were not chosen. The main reasons for that were that they were either too similar and did not provide new information or that they were outdated and the information they gave was built upon in the articles that were chosen or the articles were only about the mathematics behind parallel coordinates, something that had little relevance to our project.

When the list was set all the articles were read by both students. During and after reading each article we wrote down as much as we found relevant on the articles that could help us answer the main questions that we had posed. This stage was done on purpose without regard to what types of features Qlik wanted to have, we did this to get an unbiased view of the subject and of the different features that are used in conjunction with parallel coordinates. During the study everything that could be seen as relevant to



making parallel coordinate plots easier to understand was made a note of to be taken into consideration for the user study. The compilation of our findings and everything that was on these notes can be seen in the *Results* in the coming pages.

## 3.2 Results

This part will showcase the articles that were found and explain different concepts and features of parallel coordinates that were found during the literature study. Below in table 3.1 one can see all of the articles found, their titles, the name of all their authors and their main contribution to our work. If more information is desired they are cited and the rest of the information can be found in the references list.

Title	Authors	Area of Contribution
Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research [14]	J. Johansson, C. Forsell	Evaluation of different features
Uncovering Clusters in Crowded Parallel Coordinates Visualizations [15]	A. O. Artero, M. Oliveira, H. Levkowitz	Clustering, opacity modification
Visual Clustering in Parallel Coordinates [16]	H. Zhou, X. Yuan, H. Qu, W. Cui, B. Chen	Clustering, edge bundling
Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships [17]	J. Johansson, C. Forsell, M. Lind, M Cooper	3D parallel coordinates
The Effect of Aesthetic on the Usability of Data Visualization [18]	N Cawthon, A. Vande Moere	User study
An Edge-Bundling Layout for Interactive Parallel Coordinates [19]	G. Palmas, M. Bachynskyi, A Oulasvirta, H. Seidel, T Weinkauff	Edge bundling with Bézier curves
Interactive Local Clustering Operations for High Dimensional Data in Parallel Coordinates [20]	P. Guo, H Xiao, Z Wang, X Yuan	Clustering, edge bundling
Evaluation of a Bundling Technique for Parallel Coordinates [21]	J Heinrich, Y Luo, A. E. Kirkpatrick, H. Zhang, D. Weiskopf	Edge bundling
Slope-Dependent Rendering of Parallel Coordinates to Reduce Density Distortion and Ghost Clusters [22]	D. Pomerence, F. L. Dennig, D.A. Keim, J. Fuchs, M. Blumenschein	Clustering, opacity modification
Revealing Structure within Clustered Parallel Coordinates Displays [23]	J. Johansson, P. Ljung, M. Jern, M. Cooper	Clustering, opacity modification
Revealing structure in visualizations of dense 2D and 3D parallel coordinates [24]	J. Johansson, P. Ljung, M. Jern, M. Cooper	Clustering, opacity modification, 3D parallel coordinates
Angular Histograms: Frequency-Based Visualizations for Large, High Dimensional Data[25]	Z. Geng, Z. Peng, R. S.Laramee, R. Walker, J.C. Roberts	Histograms in parallel coordinates

**Table 3.1:** Table showing basic information for all of the articles gathered during the literature study.

### 3.2.1 Parallel Coordinates

During the literature study many different features for parallel coordinates were found, the most prominent are listed below and discussed more in depth in the coming pages. They contribute mostly to the facilitation of understanding the data through the tidying up of poly-lines and making patterns that should stick out more prominent. They are:

- Brushing.
- Edge Bundling.
- Coloring.
- Modifying Opacity.
- Clustering.
- 3D Parallel Coordinates.
- Histograms.

#### Brushing

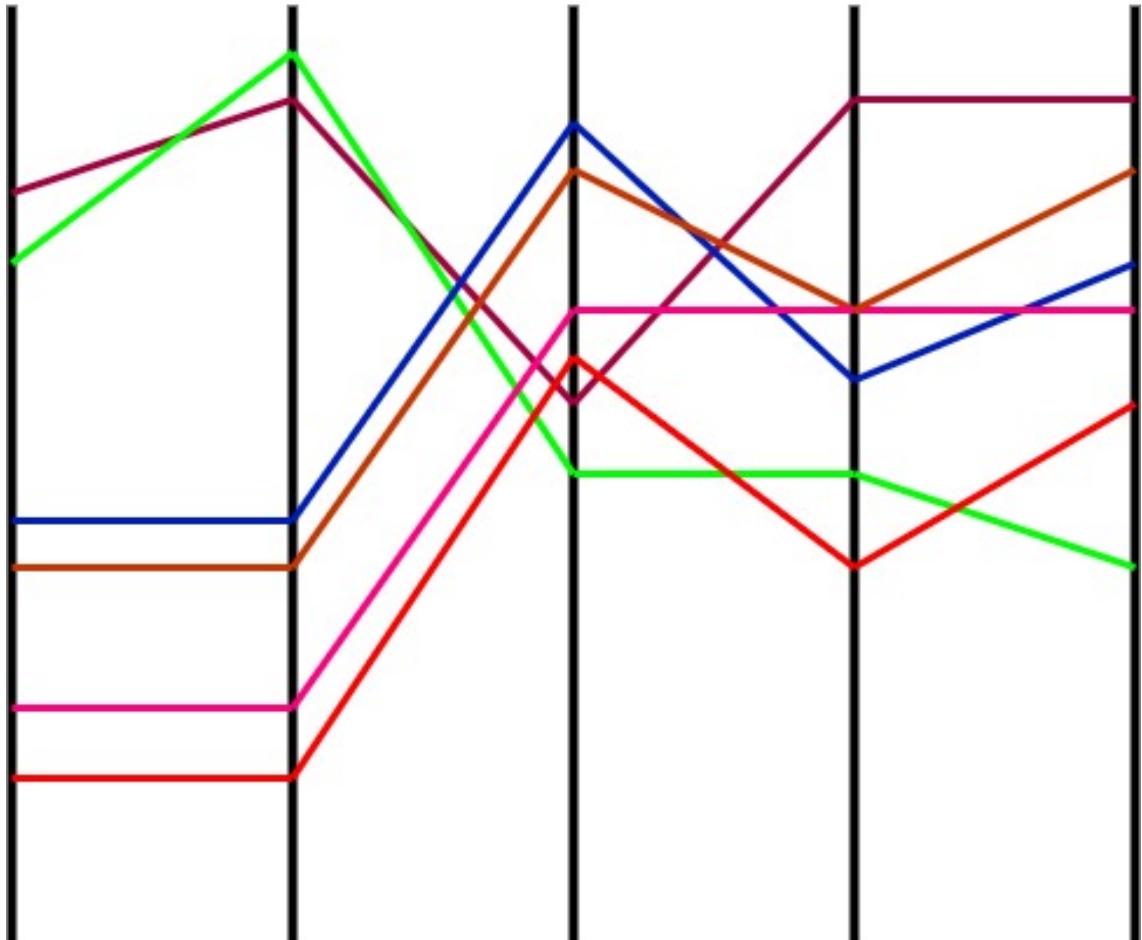
One of the most common ways to facilitate the understanding of parallel coordinates is via a method called brushing. One could also call it filtering since it is the same principle. In brushing one chooses one or several polylines to be highlighted resulting in the rest of the polylines to be either scaled down in visibility or made completely non visible. It is also possible for the plot to have a function where one instead chooses a scale in each of the parallel axes where one delimits which polylines that pass through each axis that can be visible. By brushing, the amount of data visible decreases, making it easier for the user to understand the information. [14]

#### 3D Parallel Coordinates

This feature is more like another type of graph(s) in itself. Even though the implementation of the project was specified to be in two dimensions it was decided that it would be good to read about the areas around it to get a better understanding of the subjects in the vicinity. As the name implies parallel coordinate plots in three dimensions are types of graphs that have one more dimension to be able to compare one more variable at the same time. An example could be that instead of axes we have planes and so on. This topic will not be discussed more in this report [24].

## Coloring

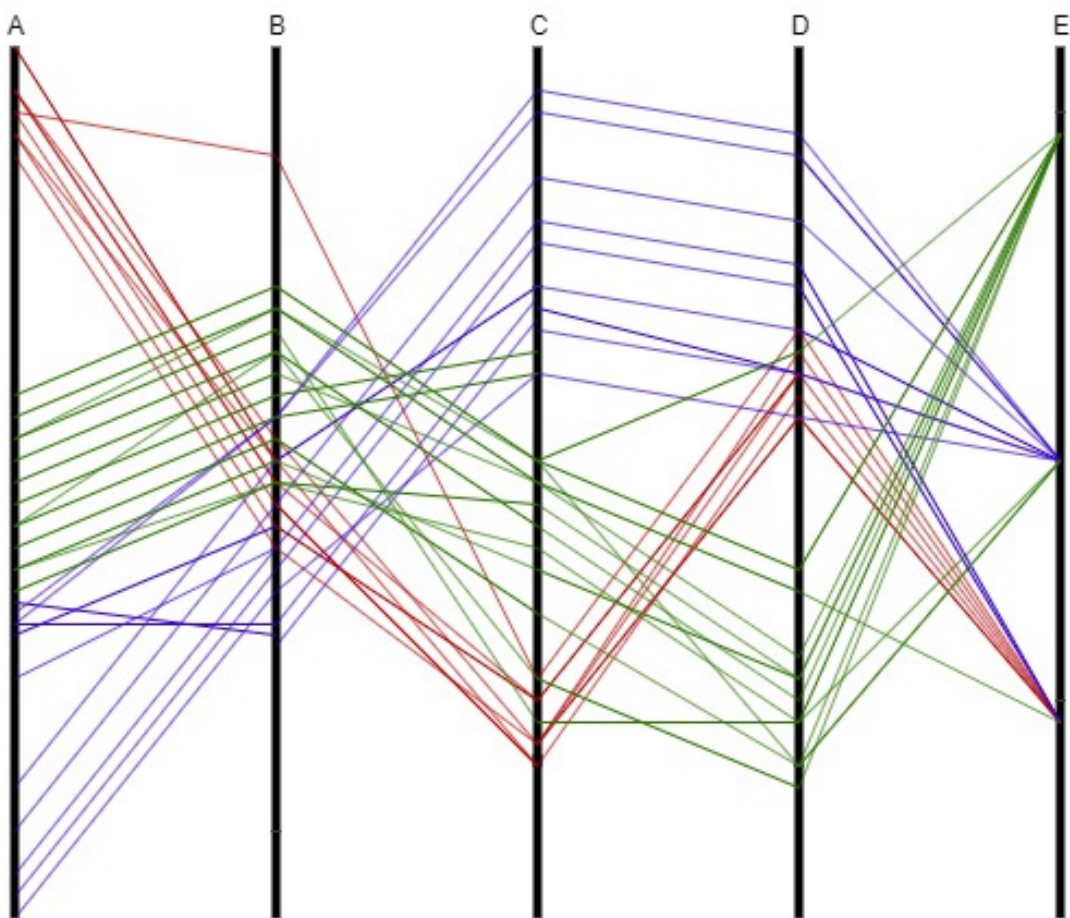
Another of the common features of parallel coordinates plots is the coloring of the polylines. By coloring the polylines based on a metric one can differentiate between groups or simply have a gradient of colors based on one of the variables so one can follow along in other axes [16]. An example of this feature can be seen in figure 3.1.



**Figure 3.1:** How a parallel coordinates chart with polylines of different colors may look like.

## Clustering

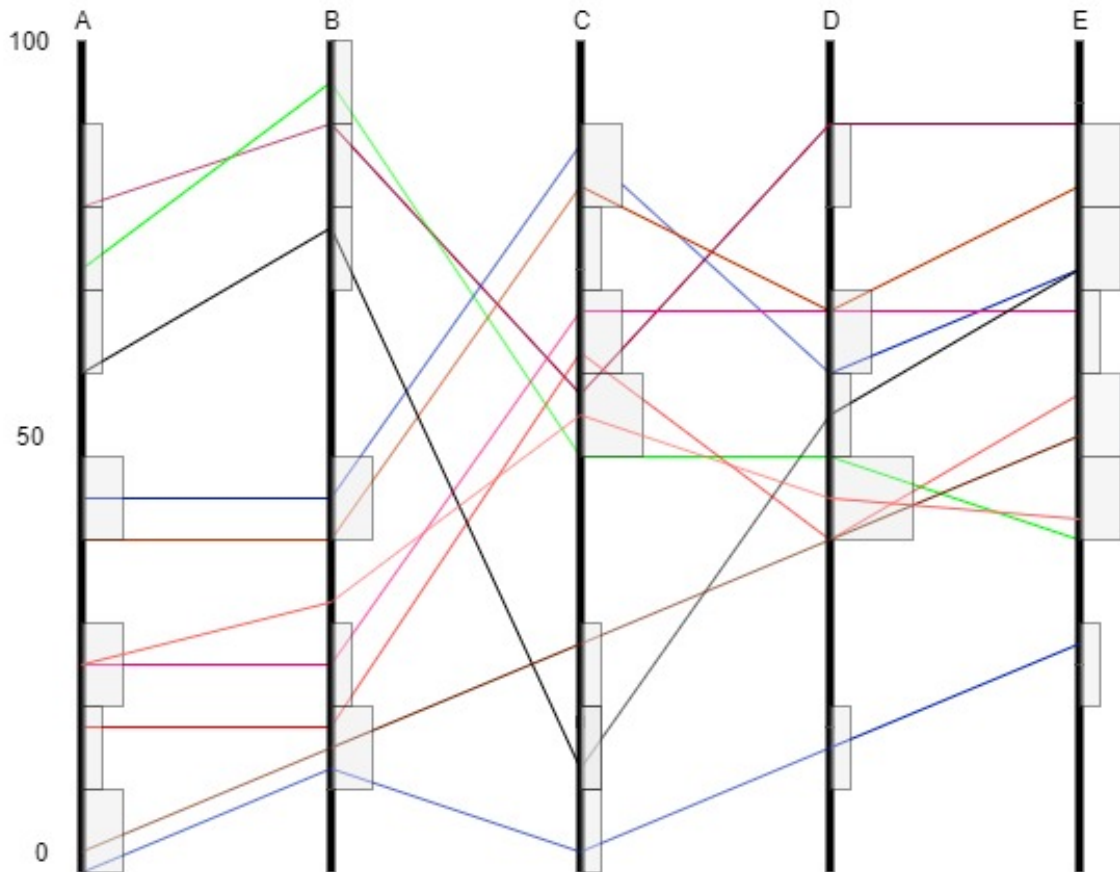
This feature is most commonly used in conjunction with some sort of coloring. The polylines are grouped together in one to several groups based on which of the polylines that follow a pattern throughout the whole graph [16]. The clusters are then usually colored with distinct colors to more easily see the difference between the clusters [16]. In the resulting graph its usually easier to see some sort of pattern since clusters of polylines with a pattern is what one is usually looking for in the first place. There are algorithms that try to make the process more effectively if one doesn't know where to begin looking, i.e. which of the axis to choose and which parameters to try [16]. An example of this feature can be seen in figure 3.2.



**Figure 3.2:** How a parallel coordinates chart with the clustering feature may look like.

## Histograms

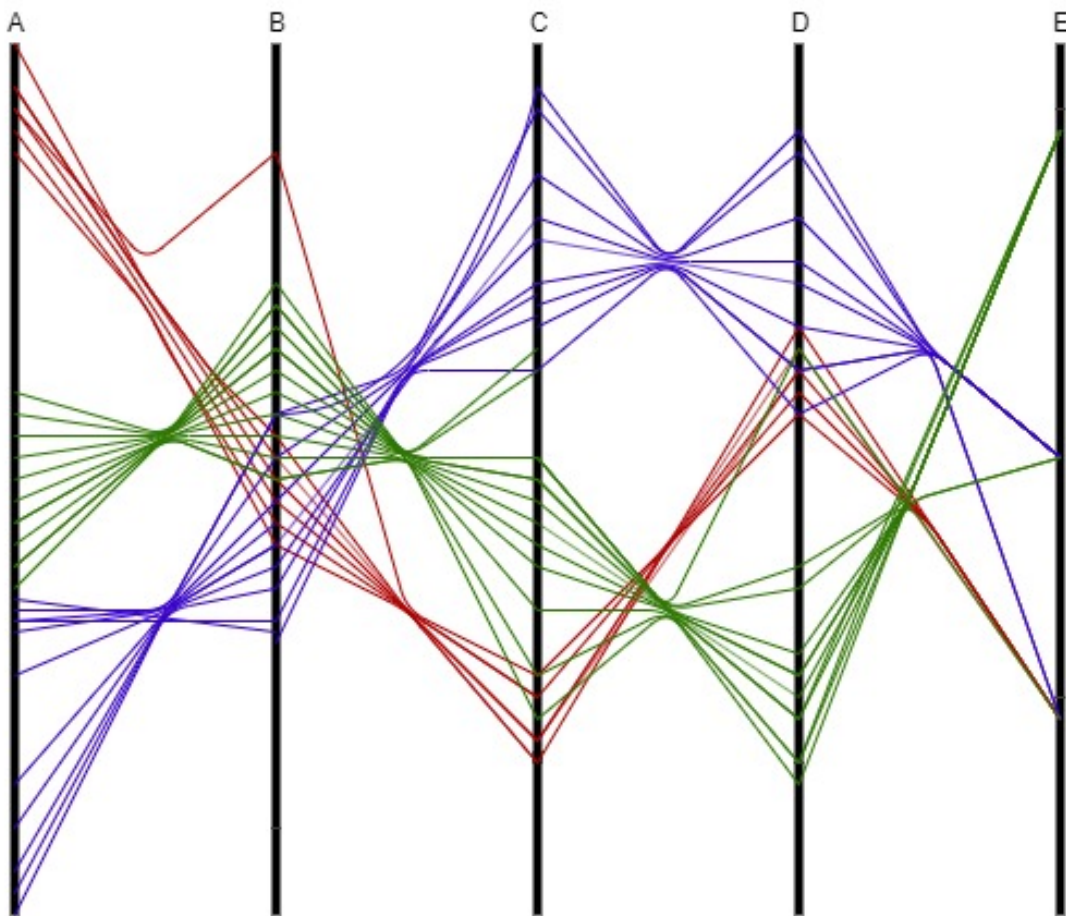
Another common method used both by itself and in conjunction with other features shown here is histograms. They are usually shown horizontally along one or both sides of the axes to show where in the axes there is a higher or lower concentration of poly-lines. This way one can see patterns in concentration even if the plot looks like there is an similar amount of poly-lines in different places [25]. An example of this feature can be seen in figure 3.3.



**Figure 3.3:** How a parallel coordinates chart with histograms enabled may look like.

## Edge Bundling

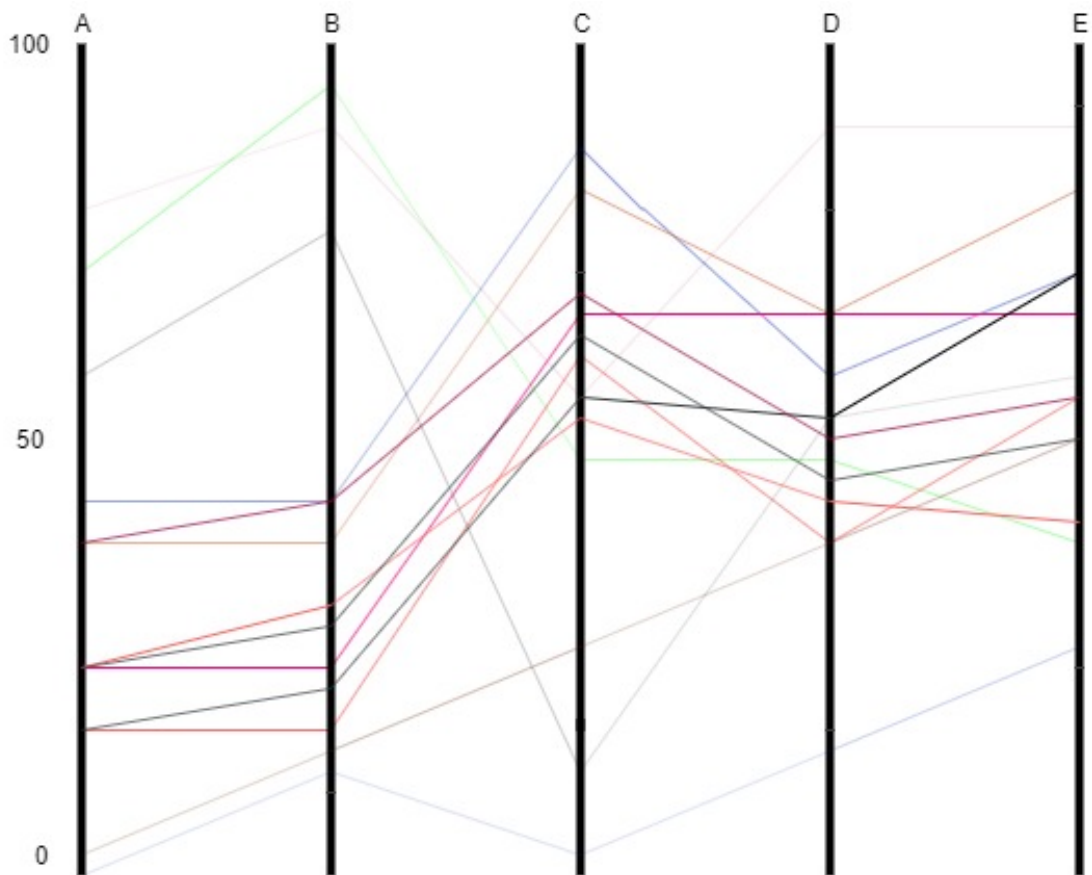
This feature can be seen as a type of clustering. After the polylines are clustered and divided (and sometimes colored) one can change the appearance of the polylines themselves to try and make the patterns more easily understood or just to declutter the graph even more. There are several types of edge bundling but they all try and concentrate and group the polylines. Some types bend the lines into Bézier curves so as to make the clusters even more uniform. Another more extreme example takes all the lines in a cluster and transforms it instead into a polygonal strip with constant width [21]. An example of this feature can be seen in figure 3.4.



**Figure 3.4:** How a parallel coordinates chart with an edge bundling feature may look like.

## Modifying Opacity

Another feature that came up during the literature study was the modifying of opacity of the polylines based on concentration in different spots of the graph with the help of a transfer function. In this method the polylines are drawn with different opacity based on how many there are grouped together. More lines usually equals that they are drawn more opaquely. With the help of a transfer function one could also choose if one would prefer for the less crowded areas to be the more opaque ones. This helps the user to see where there is a concentration of lines or lack thereof faster than usual [22]. An example of this feature can be seen in figure 3.54.



**Figure 3.5:** How a parallel coordinates chart with an opacity feature may look like.



# Chapter 4

## User Study

---

This chapter will present the user study. The study was performed to gather the opinions and knowledge from people with expertise within the area of parallel coordinates and data visualization. The goal was to investigate which methods and techniques that were gathered during the literature study, would prove useful throughout the project.

### 4.1 Method

#### 4.1.1 Preparation

The point of the user study was to elicit the needs of the potential users of the product [8]. Before being able to do this several meetings with Qlik employees of varying backgrounds were planned. The purpose of the first set of meetings was to gain a more specific understanding of how Qlik Sense works, what it produces and who the target user is. These meetings were actually done from the start of the project, even during the literature study. At the end of the literature study there was a meeting with Quan, the thesis supervisor at the company. The goal was to understand the specific constraints of the platform that the software was going to be developed for. For one, Picasso.js the framework which was to be used to build the software has limitations on what types of diagrams can be developed on it. Also the engine side of Qlik Sense has limitations as to what kind and amount of calculations can be made for the program to work fluently. These constraints narrowed down the amount of features for parallel coordinates, presented in the previous chapter, that could be implemented, by quite a great deal. The final meeting was with two Qlik employees working with user studies and surveying. The goal was to see how a big international company does surveying and if there were any assets that we could use for our user study. Another thing revealed during these meetings was the importance and relevance of having a qualitative study instead of a

quantitative study. According to the UX department at Qlik, after interviewing five persons the quality of the data discovered by the accumulative results does not increase that much [26]. Our goal was then set to try and find around five people to interview. We wanted to interview people who were the target users for the product that was to be developed. That meant finding people who develop apps in Qlik Sense and not necessarily people that develop Qlik Sense in itself. We got some tips of people that work in house exclusively in developing apps for Qlik Sense and from them we got other recommendations of other people also working developing apps for Qlik Sense. In total seven different people were asked if they would like to be a part of the user study and six of them replied that they would be a part of it. Five of them worked in the Qlik Lund offices and one of them in Boston.

The first draft of the interview was created and pilot tested on a colleague at the company. This gave a good understanding of what worked and what didn't, what direction the interview would take, and approximately how much time the interview could end up taking. After the pilot test, the questions were revised and changed so that they either asked a clearer question or removed if they felt redundant. The next step was to make sure that the questions gave enough depth to actually be able to create the concepts for the product. The questions were reviewed by the supervisor at the university and changed accordingly. A few questions were added to get even more depth and knowledge and some questions were opened up to be less constricted so that more information could be gained. The last thing before starting the interviews was to make a second pilot test with a different colleague to make sure that all was correct. This time the mock interview had a better flow and the person felt they understood more of how the interview was set up and what was expected of them.

### 4.1.2 The Interview

To make sure that all the participants were given the exact same information an orientation script containing background to the interview was developed [8]. The interviews were performed in a conference room, either in person or over video chat using Webex. Nicolás was the moderator during the interviews and Niklas took notes of what the person being interviewed said. Low-fidelity visualization sketches had been made digitally and were presented as a set of slides in conjunction with the questions [12] [13]. The main aim of using the images was to use photo elicitation[9] to convey a clearer image and elicit comments and opinions from the participants to promote reflection. The slideshow was controlled by Nicolás and was presented on a laptop in front of the interviewee. See *Appendix A* for the entire manuscript of the user study.

#### Before the Interview

The interview started with a presentation of the project and the purpose of the user study. After the introduction to the interview basic information on the interviewee and their rudimentary knowledge on visualisation was collected via pre test question-

naire. Here they were asked questions in line with how much multivariate data they have worked with and how proficient they felt using Qlik Sense. At this time the interviewee was asked if it was alright for the audio of the interview to be recorded. All of them answered yes. This gave us a safety net where we could listen to the answers more times after the interview if we felt some of the answers were not clear during the interview. The recording was realized with *Audacity* on the same computer where the slides where, the computer that the interviewee had in front of them.

At the beginning of the coloring section and all of the subsequent segments of the interview the user was asked to describe their thoughts on the matter before getting any at all information from us, except for the title of the subject to be discussed. We wanted to as often as possible encourage the interviewee to come forth with own ideas not influenced by what we showed or described to them. This was done to decrease interviewer bias as much as possible [27].

## **General Visualization**

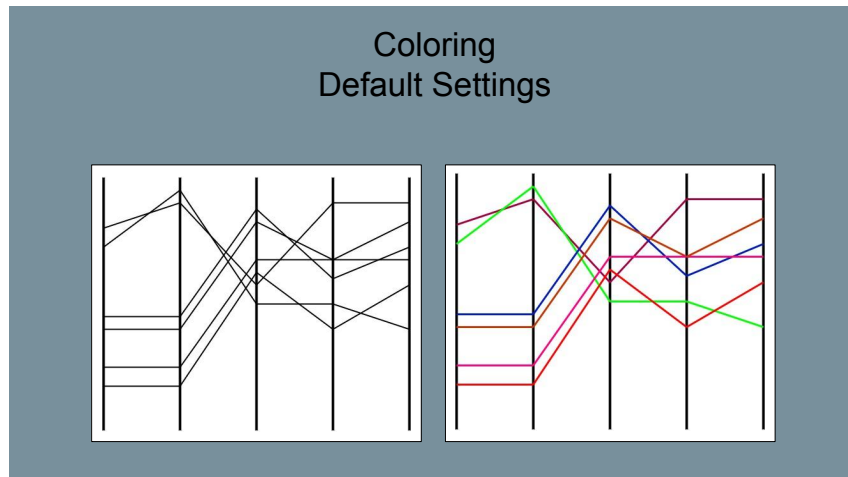
Although the project is about parallel coordinates there is also value in asking questions about visualization in general as well since they can grant new perspectives or new ideas. This was done at the beginning of the interview to try and reduce the angle towards parallel coordinates in these questions. The general visualization questions asked were mainly about multivariate data, how the interviewee uses visualisation tools differently depending on the amount of data, and their thoughts on how color can help a user understand multivariate data compared to one or two-dimensional data.

## **Coloring**

In this section the questions were centered around coloring features and techniques that have directly to do with parallel coordinates. Questions varied from which color should the chart show in its default state, as seen in figure 4.1, to advanced coloring features that a future user would want.

## **Brushing**

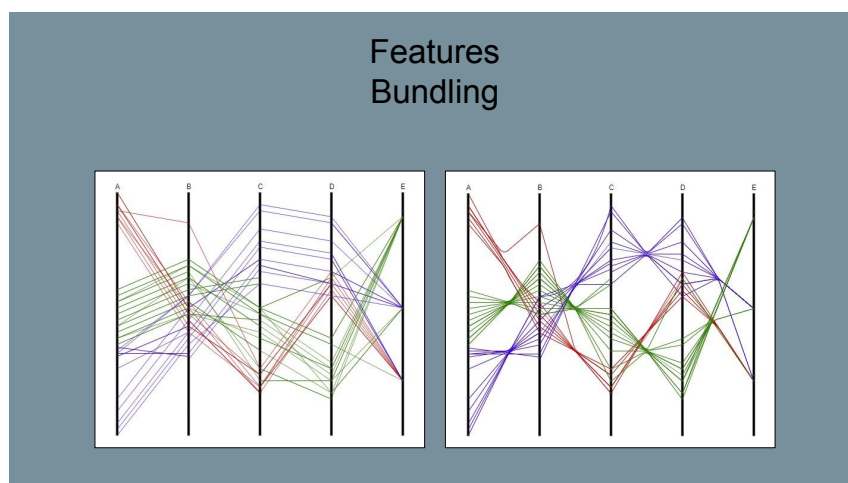
Here the first question was if the interviewee knew what brushing in the context of parallel coordinates was and if they didn't then we had prepared a short explanation. During this subsection of questions the interviewee amongst other things got to see four different types of brushing usually used with parallel coordinates and they got the chance to compare them with each other and made choices about which combinations of them would make for better brushing for which types of data [14].



**Figure 4.1:** One of the slides of the slideshow. This one presenting the interviewee with the question of what they would prefer as the default coloring when the chart is first plotted out.

### Handling for Large Amounts of Data

In parallel coordinates there is a big risk that the data sets have a lot of data. This can become very cluttered in the chart and therefore it's important to find out viable ways that make large sets of data more comprehensible. Questions were in the same structure as the last subsection where the interviewee got to see three different types of way of arranging the data in the chart and asked which of them that they thought conveyed the data in a better manner. In figure 4.2 one can see one of the slides comparing two different ways one can represent the data.



**Figure 4.2:** One of the slides of the slideshow. This one presenting the interviewee with the comparison of clustering and bundling.

## Histograms

Histograms can be used as a part of a parallel coordinates plot to better convey where on the different axes there is a higher concentration of polylines. Here several ways of using histograms were shown. Questions varied from if the histograms should be on one or both sides of the axes to if the histograms should be opaque or transparent and much more. We had this subsection as a separate segment since it was one of the more unique features that our supervisor at the company had shown lot of interest in us having in the finished product [25].

## Finishing Questions

After all of the questions regarding parallel coordinates and visualisation were asked we talked and asked with the interviewee for a while to see what they had thought of not only the questions but of the whole interview process, what they had thought was well done and if there was something they felt was lacking or missing. The entire manuscript with all the text spoken and questions asked can be found in the appendices at the end of the paper.

### 4.1.3 Compiling the Results

After each individual interview the written down answers were analyzed and compared to the recording to see if there was anything that had been missed. All the answers from all the interviews were gathered in a big matrix document for easy assessment and overview. After all the interviews were concluded the next step was to draw conclusions from the results so as to be able to generate several concepts. This was done by weighting the participants backgrounds and relevancy of their skill set. The persons who had a better understanding and expertise in the subject of parallel coordinates and visualization were weighted higher. This was not an easy choice to make and something that is discussed in more depth in the discussion section of this chapter. Discussion ensued so that the answers mirrored what the interviewed as a group sought after in this type of chart and followed what was expected of the project by the guidelines set by the company. Then the participants' answers were compiled into a single answer per question. This stage then made it possible to compile a set of concepts before beginning with the implementation of the product. More on this in the next chapter.

## 4.2 Results

Below in tables 4.1 through 4.6 are the results from the user study. The results are divided into the categories that the questions were divided into. Here one can see the complete list of the compiled answers for all of the questions asked during the interviews. In figures 4.3 through 4.5 one can see more examples of the slides used during the user study.

### Open Ended Questions

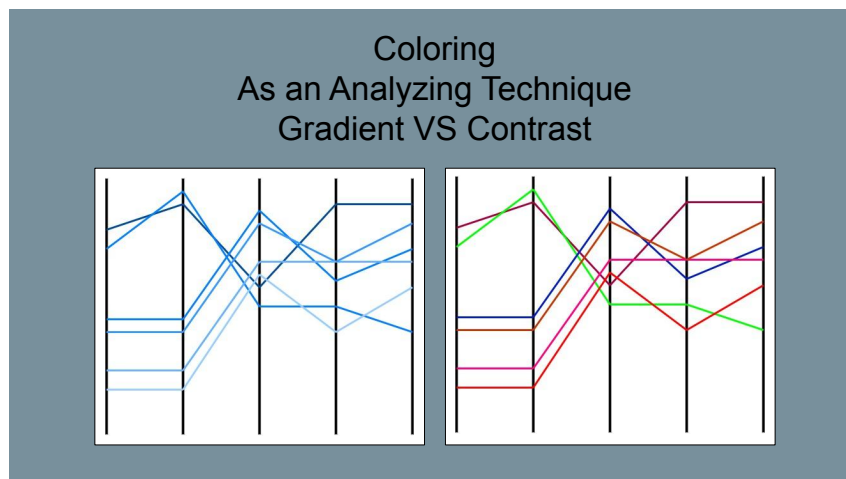
Question	Compiled Answer
How much multivariate data do you work with or have worked with? Not at all? Some? A lot?	It was quite divided, half had worked with a lot of it and the other half had either limited amounts of experience or none at all.
What would you say are the most important aspects of data visualization in general?	Visualization should be easy to understand for the user. It should be effective in communicating the "answer to the question asked".
What would you say is the difference in how you visualize and present small quantities of data and large quantities of data?	If there is too much data it can be hard to see patterns or anything in a clear way. One might need to condense data to show what's important. A lot of data also takes a lot of computing power if all is shown at the same time.
What would you need from a graph visualizing multivariate data to better understand the data? (In comparison to one or two dimensional data)	Highlight paths or trends and make sure that other interesting information is easy to perceive. Highlight what the "extra" variables are actually doing for you in comparison to lower dimensional data.
Do you have any thoughts regarding how color can be used when visualizing multivariate data? What purpose, do you think, can color serve when visualizing multivariate data?	Color can be very important, it can for example be the carrier of information or be used to highlight things, but at the same time it can't always be relied upon since a big part of the population has some form of color blindness and then how the user interprets the information is not consistent.

**Table 4.1:** Table showing the compiled results of the *open ended* questions obtained during the user study.

## Coloring

Question	Compiled Answer
Do you have any thoughts or ideas on how you would like to use color to help the visualization of parallel coordinates plots?	Depends on what you want to show whether it is distinguishing and highlighting path/outliers/groups or something else. Should be easy to understand what the color means in the diagram.
When first plotting out a parallel coordinates graph how would you like the default coloring to be, before editing anything?	Most of the participants agreed that the polylines should be single colored (but not the same color as the axes) as a default. There was no consensus on a specific color but one should not choose colors that have a widely accepted meaning, like green can mean positive or red mean negative for example.
When you add colors to a parallel coordinates plot as an analysing technique, how would you like to be able to display the colors? Would you like for the colors to be decreasing or increasing in a gradient for polylines that are beside each other?	Gradients was something that everyone seemed to like, it is also appropriate for color blind people. Having lines beside each with have high contrast colors was also generally liked. User choice was very important, they should have the option to choose how they want the colors to be formatted.

**Table 4.2:** Table showing the compiled results of the *coloring* questions obtained during the user study.

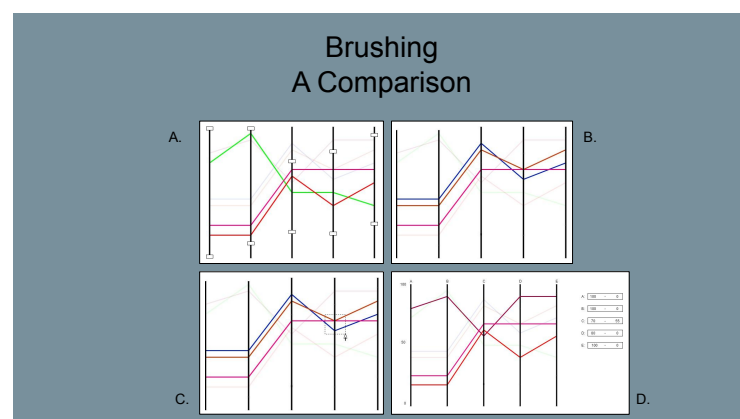


**Figure 4.3:** One of the slides of the slideshow. This one presenting the interviewee with the difference in polylines beside each others having colors in a gradient or with contrast.

## Brushing

Question	Compiled Answer
Before we show you anything we would like to know how you would like to be able to brush or filter parallel coordinates? Any ideas of your own?	before anything was shown some of the participants suggested being able to pick/click one line at a time. No other suggestions.
Now that you have seen some common ways of brushing do you have a suggestion on what other methods you would like to see implemented?	ABC would cover most user cases, D was either liked or really disliked, some participants raised concerns that data would be overlooked if using D since it's the same as A but the inputs are not as visual as the sliders and not as interactive. B and C were mostly seen as a must have since the user expects it of a Qlik product
Of the different brushing techniques that we showed you: Which one or which ones did you like the most and why? And which one or which ones did you like the least and why?	A together with B and C are comparable, many said that D is a necessary option but several really disliked D. B was generally in the top for every participant. B is a must for how Qlik works, C is also seen as a must since a lot of customers will expect for that function to be there.
Could you order them from best to worst?	Hard to order them when they all had different orders but mostly: B-C-A-D.
Is there a certain combination of them that you think is the most optimal?	B works great both A and C.
When brushing, the polylines that will be filtered out, would you prefer for them to disappear entirely or for them to be translucent and or lighter in color?	Ultimately leave it to the user, but translucent is the more preferred one because then one can still see what has been removed which can give some information. But one should be able to remove them if they just become clutter or noise.

**Table 4.3:** Table showing the compiled results of the *brushing* questions obtained during the user study.



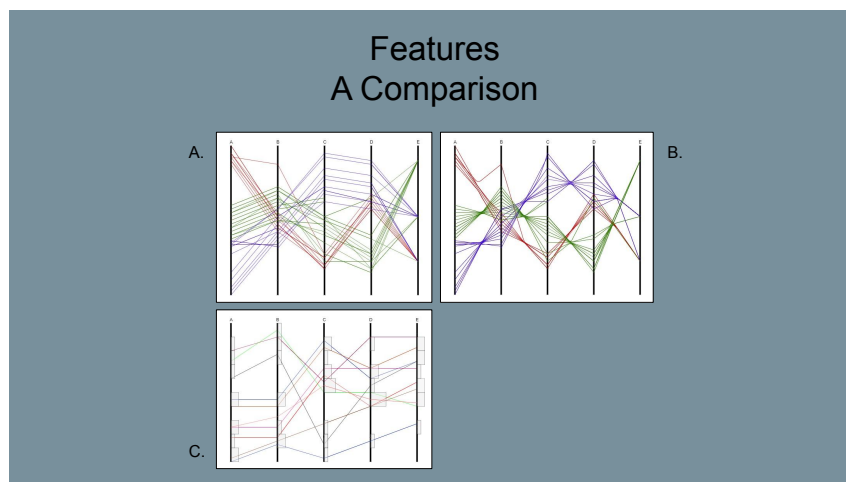
**Figure 4.4:** One of the slides of the slideshow. This one presenting the interviewee with the different types of brushing we were showcasing.



## Handling for Large Amounts of Data

Question	Compiled Answer
Do you have any ideas on how best to visualize large amounts of data on a parallel coordinates plot?	Most had no ideas of their own. One said to show where there is a higher concentration.
So now that you have heard of these three different features please rank them from best to worst and comment on why you think one feature is better or worse than the other ones?	Clustering (A) was generally well perceived. Bundling (B) was powerful but some participants didn't really see a use for it and since it alters the lines some users could have trouble interpreting it. A is great for trends but is a heavier cognitive load than B. Histograms (C) were also the subject of both approval and disapproval. But most agreed that histograms should be an optional tool.
Now after having seen these three features do you have any other ideas on how best to organize large amounts of data on a parallel coordinates plot?	Adding the possibility to have dotted or dashed lines (in conjunction with colors) to be able to perceive some lines differently. Others thought more along the path of changing the thickness of some of the lines.

**Table 4.4:** Table showing the compiled results of the *handling for large amounts of data* questions obtained during the user study.



**Figure 4.5:** One of the slides of the slideshow. This one presenting the interviewee with the different types of features for handling large amounts of data we were showcasing.

## Histograms

Question	Compiled Answer
Based on you already seeing histograms compared to other features. Are histograms something that you would like to have in a parallel coordinates plot or not? Would you like them to be there as a default or as something you add to the plot when you want the feature?	One participant saw this as something that should not be implemented at all, but all the others saw it as a nice thing to have but should be something the user adds themselves not as a default.
Would you like to have the histograms on just one side of the axis or on both sides?	The participants were divided equally on this one. It is more common to have it one side for histograms to make sense. One could have in the edit field if one wants it on the right, left or both sides. The participants were both divided and with very strong opinions on this one.
Should they be semi transparent or opaque?	semitransparent but with the option to make them opaque as a very last resort.
Would you like the histogram to show the value of the amount of polylines that are crossing through it? How? Where?	People like the hover option, where hovering the mouse on top of the histogram showed the line amount. Most saw the option where the number is always showing as more clutter than something helpful. Location wise, most agreed that inside the histogram would be best if at all.
Is there any other shape that you think would be better than a rectangle as a histogram? That you think is either better at conveying information or more visually appealing?	None of the participants wanted anything else than rectangles.
Do you have any ideas about how you would like the coloring of the histograms to be? Would you like the histograms to all be the same color? Different colors depending on the concentration of polylines going through it?	Coloring should be optional. Maybe to highlight outliers. A lot of them also thought that it would be counter productive for histograms to change colors or have different colors seeing how they work.
Histograms can also be used in brushing. One could, for example, click on a specific histogram and filter out all the other lines. Compared to the other brushing techniques that we showed you, where would you put brushing with histograms in?	Most of them thought that histograms as a brushing option would be very confusing but one of them really liked it. Still most of them agreed that if the histograms are implemented one should be able to brush with them.

**Table 4.5:** Table showing the compiled results of the *histograms* questions obtained during the user study.

## Questions at the End of the Interview

Question	Compiled Answer
Are there any ideas or features that you would like to see implemented that we haven't mentioned?	To try and keep it quite simple for the user in the beginning. The first window when it is plotted out shouldn't have too much information that could confuse a user.
Of the things we have mentioned, which ones would you say are a must have and are there any features that we have talked about that you think that we should not implement?	Herein lies the conflict of histograms, some said it could be a nice feature to have, one said it's a must have, and one said it "should not be implemented and is redundant". Many lifted up coloring as very important. Two of the participants also had the idea to have a horizontal line across of all of the axes giving a level or bar. For example: If one is above the bar (the bar could describe how much money a company made in a quarter) then the polylines would be green colored and if one is below the bar then the lines would be red colored.
Lastly, is there anything else you would like to add? Any feedback on either parallel coordinates or the interview itself?	People seemed to appreciate the interview. Some hinted for future interviews to give more explanation about what we were doing. Also some of them would have liked real data, so as to say that our low fidelity slide examples were to abstract for them sometimes.

**Table 4.6:** Table showing the compiled results of the *questions at the end of the interview* obtained during the user study.

## 4.3 Analysis & Discussion

As commented before we did not take lightly the fact that we weighted the background of our interviewees and gave more weight to the answers of people that had more experience. We are aware that by making that choice we gave more weight to the ideas that are already standardised and maybe those people have a harder time thinking about new concepts in their field since they already work with something that works for them. The problem is that the experience that we are speaking of that some had and some others lacked was not only within parallel coordinates but also in general visualization as well. Some of the people we interviewed were not as accustomed to visualization as we had been made to understand by the ones that recommended them to our interview. Even though we didn't want to color the results we had to step in at this stage to weigh the answers so as to not get concepts with features that could clearly be seen as an error of not knowing more about a field. At first it can be seen as a bias that we had but we hope that the reader understands that the different interviews, test and so on that we have done during this thesis were not meant for persons without a background. We need experts to test our software because experts are the ones that are going to use it. We could not always find experts so we did what we could with what we had.

We resume by asking, what did we get out of the results of the user study that we can use? We got an understanding on which features the potential users would like to have implemented. Up until now we have not been taking in consideration the limitations that Qlik Sense and Picasso.js pose. Picasso.js is the language used for writing the code, better explained in the chapter about implementation. The limitations must be taken into account now in deciding which features that we would like to develop before building up potential concepts and presenting them to the focus group. We are still content in not taking them into account this far so that Qlik gets a feel for which features their software should support if they wish to accommodate all types of users. The features we finally chose to go with were all part of the requisites set up by Qlik at the beginning of the project. These were, the ability to: brush in some way, perform selections, have histograms for both brushing and highlighting and the ability to change basic cosmetics things in the chart. Features we dropped at this stage were for example all of the features presented for handling large amounts of data. When learning more about Qlik Sense we understood that those features either had not the required framework in Picasso.js in place for us to be able to build it up right now or there were restrictions from Qlik Senses side in how fast the data could be called from the engine side.

One comment from a participant that stood out was in reference to which colors we should choose. He did not have give specific requests but brought up the fact that for many people different colors possess different meanings. For example red often signifies something negative and green often signifies something positive. This comes from the cultural and historical meaning of different colors, where red is the color of warfare, fire and violence, whereas green is the color of renewal, abundance, and is considered to signify stability [28]. This is just to bring up the two colors named during the interview but made us think about color in visualization for the steps to come.

# Chapter 5

## Concept Generation & Focus Group

---

After the user study the next step was to compile the results with the theory and create concepts for prototypes and then discuss the advantages and disadvantages of each concept with potential users at Qlik with a focus group. After that a final concept could be constructed from all the feedback.

### 5.1 Concept Generation

#### 5.1.1 Method

The aim of the concept generation was to create three concepts that differed slightly in how a user would interact with them so that we could evaluate them in a focus group and in that way get more feedback on how users should interact with the chosen features before going into the development phase. The data gathered from the focus group would then be compiled into a fourth concept that would be the one to be implemented. The concepts were created with the same features but with different approaches to the interaction since the product still needed to have the same functions to comply with what Qlik had sought. More about why just these features were chosen can be read in the discussion part of this section. The features are:

- The ability to brush on the main dimension or any measures.
- The ability to perform selections.
- The ability to use histograms for brushing or highlighting.
- The ability to change the basic appearances of the chart.

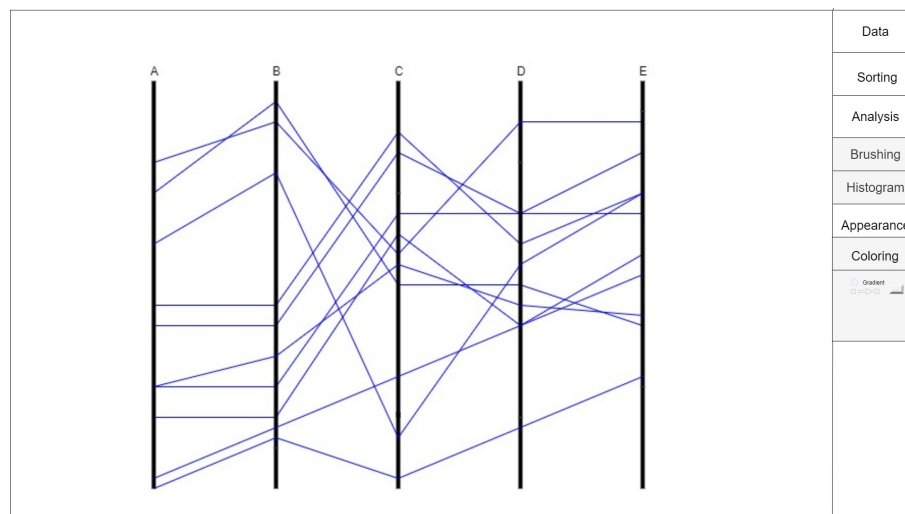
Important to note that brushing and selection can both at least be done by clicking and using a lasso function and the user can switch between the functions. Qlik's lasso function is extensively described in section 6.3. Outside of the constraints that we had we

tried to appease the suggestions of the volunteers from the user study as much as possible. The different ways to brush that were chosen were based on what was preferred, brushing types *A*, *B* and *C*. The standard coloring state of the polylines is monochromatic, also what the interviewees preferred during the user study.

## 5.1.2 Results

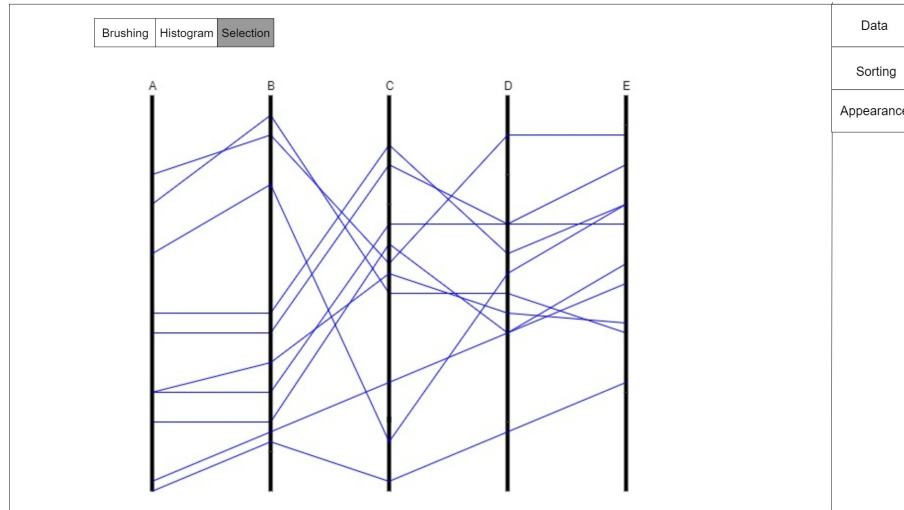
Along with the descriptions of the concepts, visualizations were created in the form of lo-fi sketches of the GUI where the focus laid on the process of using the chart and the interactions one would need to do instead of the design choices [13]. The in-depth written descriptions of concepts one through three can be found in *Appendix B*. Below follows a short description of how the interaction varies in concepts one through three.

**The first concept** was made into a very bare bone structure that builds directly on the standard style that Qlik's chart usually have. With all the functionality in one place in a side bar that you reach by pressing the edit button that would exist in the app but outside of the chart itself. In figure 5.1 one can see one of the low-fi sketches created for this concept.



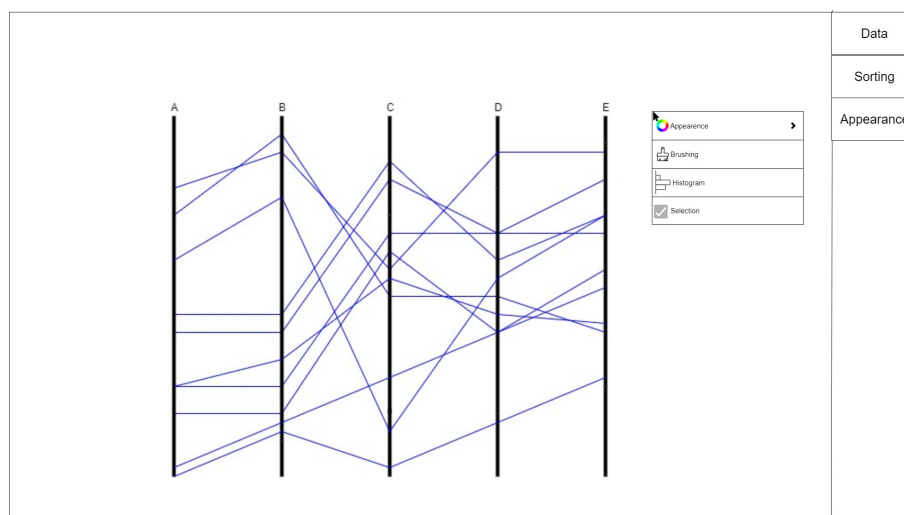
**Figure 5.1:** One of the low-fi sketches created to better explain concept 1.

**The second concept** was generated in a way that the design lets the user access the features that are specific for the parallel coordinates plot directly from the chart with the help of a toolbar with toggle buttons on it. The user would then have easier access to the specific features, but it would also make the toolbar a part of the chart itself. In figure 5.2 one can see one of the low-fi sketches created for this concept.



**Figure 5.2:** One of the low-fi sketches created to better explain concept 2.

**The third concept** would build on the idea of the second concept, to have the functionality of the features within the chart itself but this time instead of an always visible toolbar, the same options would be reached by right clicking with the mouse instead. In the context menu one would be able to reach the same functionality as the ones available in concept two but the actions to some of the functions would vary. In the context menu one would also find customization options which in concept two you would still need to go outside of the chart area to reach. In figure 5.3 one can see one of the low-fi sketches created for this concept.



**Figure 5.3:** One of the low-fi sketches created to better explain concept 3.

### 5.1.3 Analysis & Discussion

Even though the user study resulted in some features to be more desirable than others there were a few baseline features requested from Qlik at the beginning of the thesis that we can not overlook. This resulted in histograms being part of the concepts even though some of the interviewees greatly advised against it since it was a prerequisite from Qlik. The user study gave some new ideas that could be interesting to develop, like clustering and bundling, but were not realistic for the time frame we had and our developing skills. This resulted in all three concepts having the same features and instead concentrated on having different interactions. So because we had a framework of instructions, time and skill to follow there were some good ideas that we could not bring into the concepts and some less desirable features that we had to have in the concepts.



## 5.2 Focus Group

### 5.2.1 Method

#### Before the Focus Group

The focal point of the focus group was to encourage discussion surrounding the differences in interaction with the product that the three concepts illustrated. A manuscript loosely resembling the one used in the user study was created to guide the group in the right direction so we could get a varied view of opinions on all of the concepts. Slides were also created containing the aforementioned visualizations created for each concept to better show the differences between the concepts.

#### During the Focus Group

The focus group took place in a Qlik office with a projector assisting us in showing the slides. It consisted of four people with various backgrounds. Two of them were users of the product and two of them were developers. During the focus group Nicolás was the moderator and in charge of what was being shown from the projector while Niklas co-moderated and concentrated on writing what the participants said. The focus group [6], was divided into four parts.

**The first part** consisted of a presentation of the three concepts, the ideas behind them and the reason behind some of the changes made to the standard Qlik graphic user interface (GUI).

**The second part** consisted of a discussion about the different concepts. The aim of the discussion was to gather the participants' opinions regarding the differences between the concepts as well as advantages and disadvantages on the differences in the way the user interacts with the product in the different concepts.

**The third part** consisted of more specific questions that needed answering, the questions were not necessarily connected to one specific concept but still gave very important data and metrics.

During the second and third parts of the focus group an inadvertent discussion began. Some of the features we had brought up seemed to be more thought provoking than we thought. During this part we helped the participants by searching for things on the web so that they could better elaborate and convey their ideas. The biggest outcome of this spontaneous dialogue was that some of the features that Qlik had requested, like histograms and the switching between selection and brushing, were things that not all of the participants were positive towards.

**The fourth part** consisted of the participants filling in a Pugh Matrix to evaluate the concepts with the *Qlik way* as the baseline. The Pugh matrix had 3 goals for every function and if the participants thought it was an improvement compared to the baseline

they filled in a + sign, if they thought it was inferior they put a – sign and if they thought it was neither better nor worse compared to the baseline they put a zero in the cell in the matrix [10]. The Pugh matrices were filled individually by all four participants and the four Pugh matrices were summed together into one Pugh matrix.

## 5.2.2 Results

The focus group gave both a wide array of varied thoughts and opinions on the different concepts as well as a more concentrated and measurable result in the Pugh matrix. None of the concepts were badly received but one concept's way of interacting with the features presented was a clear winner, concept two. Below one can see the thoughts of the participants on each of the concepts. There are also remarks made towards the conclusive sums that all of the concepts got in the Pugh matrix, which can be seen in its entirety in the next page.

**Response for the first concept** was quite mild. All of the participants felt that they both understood and recognized the proposed workflow that concept would have. They only strictly positive feedback was that they thought that chart would look good in Qlik Sense.

**Response for the second concept** was much more varied. All of the participants liked the added accessibility that the toolbar added for the user to be able to access certain things without needing to go to the edit menu. But one of the participants later added a comment that all the other participants later agreed with. Toolbars take up valuable *real estate*, or space, in the chart which is highly unwanted since simplicity and aesthetics are very important for the workflow of Qlik Sense. Although this was a widespread view it seems the participants didn't value the simplicity as much as the functionality that the toolbar could provide as can be seen in the result of the Pugh Matrix where the second concept was by far the most liked one of the three.

**Response for the third concept** was again mixed but contrary to the second concept response here was at first good and then as the discussion ensued the approval of it just got lower and lower. They all liked the fact that they could access the features so easily, while at the same time not sacrificing any *real estate*, one of the avid users of Qlik Sense even wished he could have that ease of accessibility right now on other charts. But they all agreed that what was gained in accessibility was lost in visibility. Right clicking is not a common part of the workflow of Qlik Sense and they feared this could make those features feel hidden for some of the users. This shared view is clear in the result of this concepts sum in the Pugh matrix where the resulting score was even worse than the baseline set by concept one, the default Qlik Sense experience.

Functions	Goals	Concept 1 - Classic (Baseline)				Concept 2 - Toolbar				Concept 3 - Right Click Menu			
		User #1	User #2	User #3	User #4	User #1	User #2	User #3	User #4	User #1	User #2	User #3	User #4
Coloring	Ease of Use	0	0	0	0	+	0	0	0	+	+	-	+
	Ease of Understanding	0	0	0	0	0	0	0	0	0	0	-	+
	Aesthetics	0	0	0	0	+	+	+	+	+	-	-	-
Histograms	Ease of Use	0	0	0	0	+	0	0	0	+	0	0	+
	Ease of Understanding	0	0	0	0	-	0	0	0	-	0	0	+
	Aesthetics	0	0	0	0	0	0	0	+	0	0	+	-
Brushing	Ease of Use	0	0	0	0	+	+	+	0	+	-	0	0
	Ease of Understanding	0	0	0	0	+	+	+	0	+	-	0	-
	Aesthetics	0	0	0	0	0	0	+	0	+	0	0	-
Selection	Ease of Use	0	0	0	0	+	0	+	-	+	0	0	-
	Ease of Understanding	0	0	0	0	+	-	+	0	+	-	0	-
	Aesthetics	0	0	0	0	0	0	+	-	+	0	0	-
	SUM	0	0	0	0	6	2	7	0	8	-6	5	-9
	Total SUM	0				15				-2			

### 5.2.3 Analysis & Discussion

Even though the participants at first liked the toolbar in concept two thanks to its ease of accessibility and then disliked it because of the high cost on screen real state they seem to have settled down on approval of the feature, at least from what can be read from the Pugh matrix. This may be because a toolbar can be very useful for the user since it conveys features that the user otherwise might not have known existed. Although the toolbar takes precious *real state* the participants nonetheless rated high the ability for the user to be aware that the features exists.

The spontaneous discussion between the participants in part two and three took us by surprise but we let it go on within a certain arbitrary margin we both felt was acceptable during the discussion so that the conversation did not become an argument. It later resulted that a lot of the things discussed were of great interest for us so we could better understand how widespread the acceptance of certain features that Qlik had requested of us were.

## 5.3 Final Concept

After the data gathered was compiled into cohesive, measurable answers we could compile what the participants of the focus group thought were the best parts of each concept and create a final concept.

None of the concepts conflicted against each other so we could have just implemented all of the features with all of the different ways of interacting with them, that could have satisfied a larger audience of users but we had a time frame to be mindful about. The clear choice to make from the options we had was to create a final concept based on the structure of concept two. In contrast to other iterative interaction design processes we decided at the start of the thesis that due to time constraints we were not going to do a high fidelity prototype and instead go directly into implementing something that Qlik had direct benefit of. This meant that when the final concept was being worked on we had to think of it as software to be implemented, a time consuming process if done without proper preparation. Therefore, a comprehensive requirements specification was created to hopefully cover most of the unexpected turns that could come during the developing stage. The requirements specification contained:

- General information about the basic framework the program will have and the most rudimentary requirements for the software to have.
- Descriptions of what the toolbar would contain and basics functions it would perform.
- Information on what the appearance of the chart should be in the default stage and what could be changed by the user.
- In-depth description on how the brushing would work, how the brushed poly-lines would behave and where the function could be accessed from.
- In-depth descriptions on how the highlighting and selection features would work and where the functions would be accessed from.
- A very comprehensive view of how the Qlik Sense standard drop down edit menu would look like and what functionalities should be accessible from here.

The requirements specification in its entirety can be viewed in *Appendix C*.

# Chapter 6

## Application Development

---

With the prototype conceptualized it was time to implement the prototype for the parallel coordinates chart. This chapter presents the ups and downs of the development phase and why the final product differs from the presented final concept in the former chapter. The language used in the coding process was Picasso.js and the platform which it was developed on was Nebula.js.

### 6.1 Theory

#### 6.1.1 Picasso.js

Picasso is a JavaScript based, chart building framework created by Qlik. The elements that the chart is made up of are written in Picasso.js and JavaScript. For that reason this is also the framework that we used to implement this project.

#### 6.1.2 Nebula.js

Nebula.js is a scaled down platform that works very similar to Qlik Sense but with only the core functionality. When a new chart is developed it is first tested in Nebula.js to make sure that the chart works as intended. When the chart is fully developed it is then created as an extension or native element in Qlik Sense.

## 6.2 Method

This is a master thesis in interaction design and not software development and thus this method section of the implementation stage is purposely short and straight to the point and not overly detailed and complex.

The final concept was structured into a list of features to implement, listed in decreasing priority order. This was not so much deciding ourselves which features were important and not taking into account the opinions of the participants of the user study and focus group. It was mostly that we wanted to be clear that we needed to create the basic functions of the chart, such as axes and polylines, before concentrating on which brushing functions were to be implemented. The next step was to establish a minimum viable product with the help of the aforementioned list. This would give us the product with the lowest possible amount of features that would still make for a viable parallel coordinates chart for the users. The most important factor to take into consideration when establishing the minimum viable product was "what is the least amount of functionality needed to just display parallel coordinates"? It was established that to have parallel coordinates there needed to be a user chosen number of parallel axes that have lines between them where every line is representing the values mapped to a specific dimension. Another factor was that the product needed to have the basic Qlik functionality *selection*, otherwise it could not function as a Qlik Sense extension. Other features like coloring and filtering were deemed important but non essential for the minimum viable product.

When the minimum value product was established the implementation of the product began. During the few following weeks, with help from both the existing Picasso documentation and the assistance of the developers in the department, a functioning parallel coordinates plot was created. That first prototype had only one function and that was to display the parallel coordinates.

The next features to be implemented were the selection and highlighting functions. They would let the user interact with the chart to help them understand and analyse the contents of the chart better. Again, this stage was realised with the help of both preexisting Picasso.js documentation and aid from colleagues. The ability to highlight a polyline by using both a mouse click and a lasso function were implemented into the product. The lasso function is native to the Picasso.js framework and therefore our lasso function looks and works identical to how it would look in another chart of Qlik Sense. The connection to the selection button, which is native to the Nebula.js framework since it is a Qlik Sense staple, was also done. Now the chart could both show a parallel coordinates plot and be interacted with in the most basic way.

With highlighting and selection implemented the next feature to develop was filtering/brushing. After much trial and error a filtering mode in the toolbar was created that let the user remove lines from the chart by using the lasso function. Many features and functions remained to be implemented for this to be a complete product but due to constraints better argued about in the discussion section of this chapter this is as far

as the software was developed.

## Data Set

While reading the articles in the literature study we encountered, several times, a data set containing cars and metrics of cars that was often used in conjunction with parallel coordinates. We always saw it referred to as "mtcars data set", when it was referred to at all. A quick search on the internet for "mtcars data set" reveals that this is a staple of multivariate data and parallel coordinates. The data seems to be extracted from a 1974 *Motor Trend* magazine and describes fuel consumption and ten other aspects of cars performance for 32 automobiles models from 1973 and 1974. We could not find a reliable source for that information as everyone using the data and citing its origins seems to use the same quote. As a homage to the seemingly famous data set we used it in the development stage and forth in our work.



## 6.3 Resulting Prototype

The final product that was developed was a parallel coordinates chart with highlighting and selecting features and restricted brushing abilities. The colors were chosen by a predetermined set of colors that Picasso.js offers.

In figure 6.1 the overview of the software can be seen. Here and in all of the coming examples a revised version of the "mtcars data set" was used. We choose to use all of the cars but not all of the aspects that they were compared with so as to not clutter the screen and make the images clearer to follow. The software is presented in Nebula.js that consists of three main parts.

- **At the bottom left:** The chart in itself.
- **At the bottom right:** The fields for feeding in information about which measures and dimensions one wants.
- **At the top:** Other controls that all of the Qlik Sense charts have.

The space for the chart is where the data can be visualized in the form of parallel coordinates. The chart is made up of one dimension axis to the far left and, in this case, five dimension axes to the right of it.

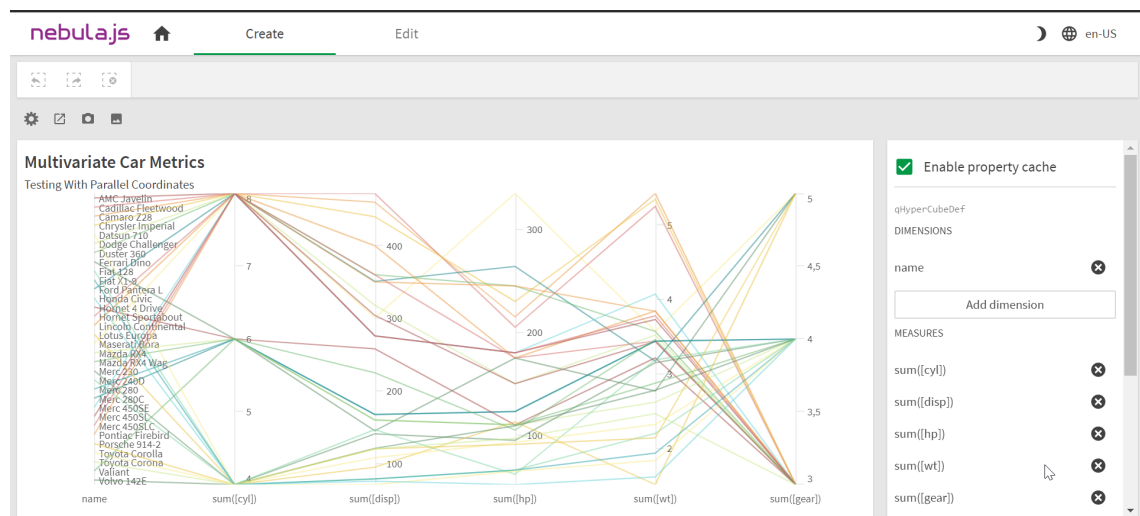
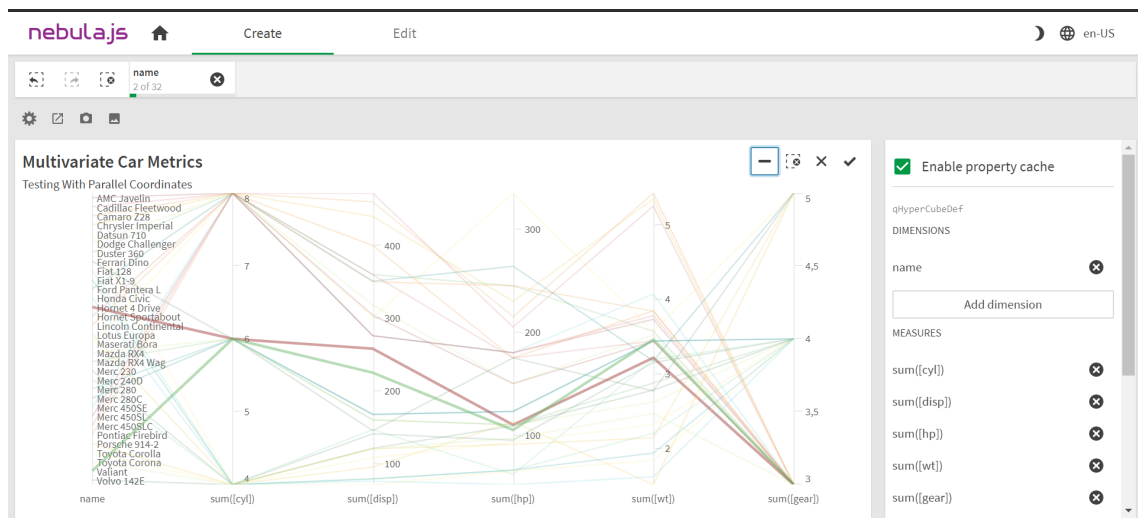


Figure 6.1: An overview of the chart.

## Highlighting & Selection

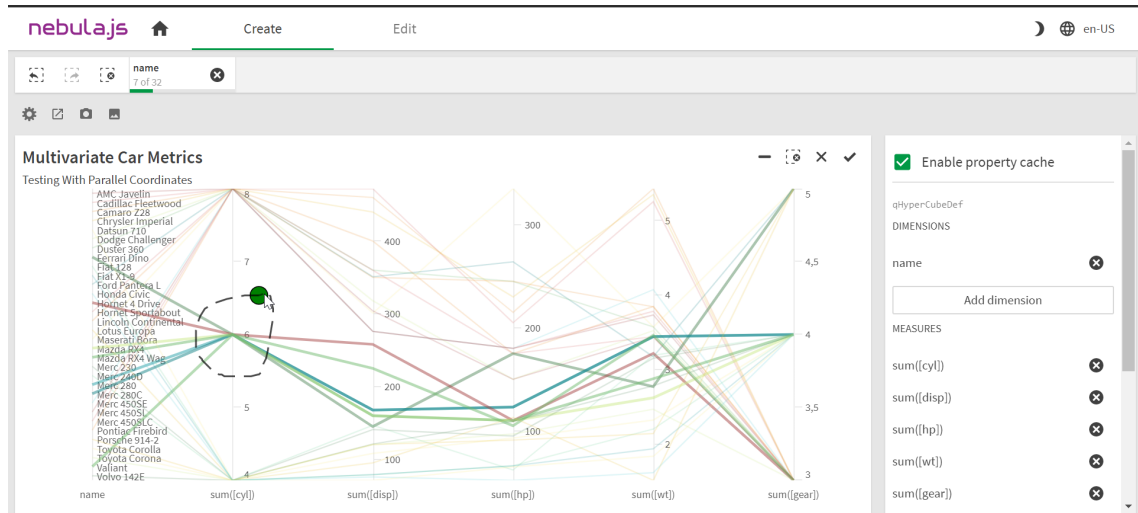
The chart has a selection feature where the user can highlight one or several polylines and then remove all other polylines from the chart and just look at the selected poly-lines by pressing the *confirm selection* button in the toolbar, the one with a *check mark* symbol on. To highlight polylines the user can simply either click with the mouse cursor one line at a time or use the lasso function to highlight all polylines within the lasso area. If the user wants to un-highlight a highlighted polyline, they can do so by clicking with the mouse cursor on a highlighted polyline.

In figure 6.2 one can see what the chart looks like when one polyline is highlighted, note how the opacity of the highlighted polyline is greater than the opacity of the other polylines.



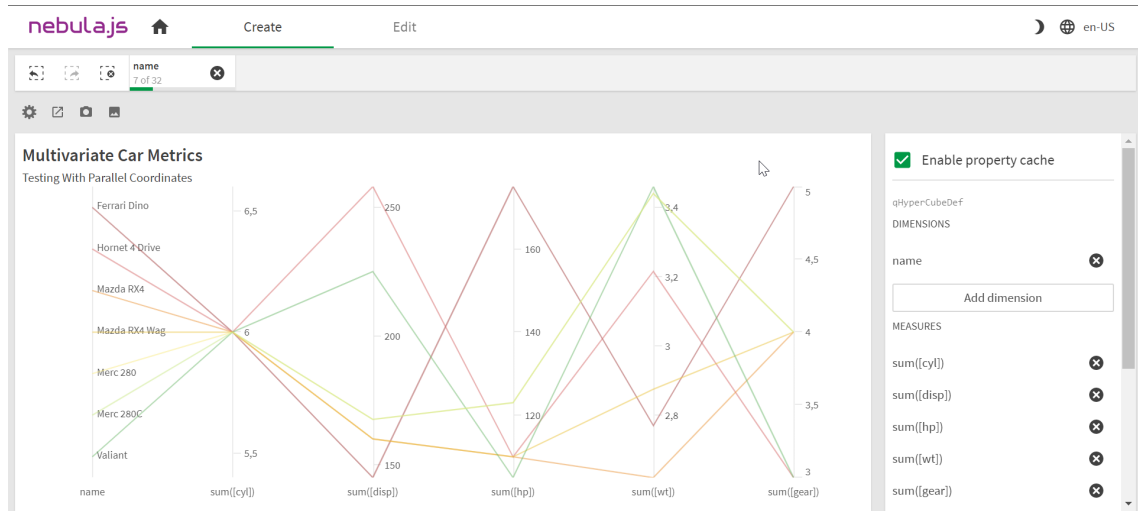
**Figure 6.2:** A view of the chart with one highlighted polyline.

One can also highlight lines by using the lasso function. As soon as one clicks down with the mouse (and does not click on a line) and starts to move the cursor, a green ball appears as the starting point of the lasso that is being performed. When one moves the cursor now a dashed line appears tracing the mouse cursor of the user. The shape does not have to be a circle or square just a free form figure. If the cursor is near enough the green ball when one releases the mouse button a highlight is performed on all the lines that are partially inside the area that the lasso has now formed. If one releases the mouse button when the cursor is not near enough the green ball no highlighting is performed. This can also be used as a way of cancelling a lasso in the process of it being drawn if one is not happy with it. As stated before the lasso function is native to Picasso.js so we could not control neither the size and color of the ball nor the type of dashed line used. The lasso function can be seen in figure 6.3. The feature is more intuitive when one uses it than can be described in words.



**Figure 6.3:** Highlighting polylines using the lasso function.

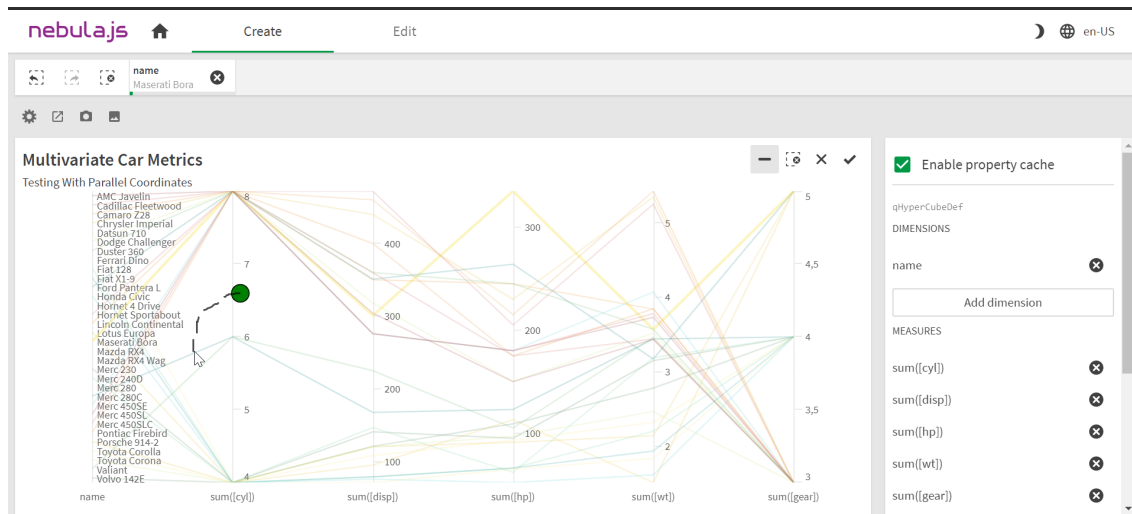
When the user has highlighted the polylines they want to look closer at they can press the *confirm selection* button in the toolbar. The chart is then redrawn and only the previously highlighted lines are left as the only polylines in the new chart, these polylines are now not highlighted. This can be seen in figure 6.4.



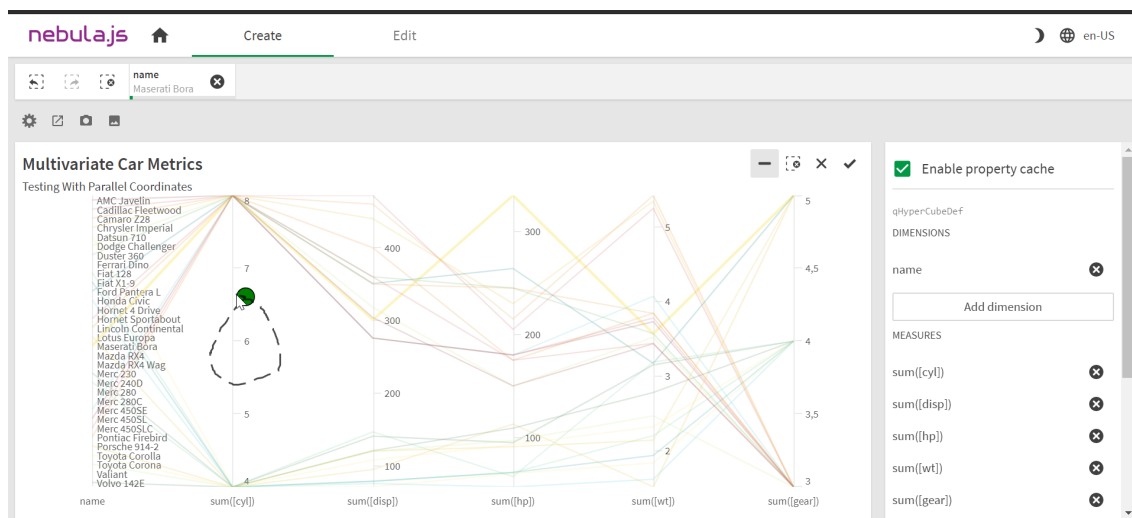
**Figure 6.4:** An overview of the chart where some of the polylines have been selected.

## Brushing

The filtering function is a little less straightforward. One must begin by highlighting any polyline so that the toolbar becomes visible. One can compare figure 6.1 with the other pictures in this section to see that in that figure there is no toolbar because no highlight has yet been made. When the toolbar is visible one can press the toggleable *filtering* button that has a *minus sign* symbol on to toggle the filtering function on and off. If the filtering function is **on** the lasso function will only work as a way of filtering and not highlighting. This is the only implemented way to brush or filter lines in the chart. If one wishes to be able to highlight lines again one must remember to turn off the filtering function. In figures 6.5 and 6.6 one can see the lasso filtering function in action.



**Figure 6.5:** Brushing lines with the filtering-lasso function (part 1).



**Figure 6.6:** Brushing lines with the filtering-lasso function (part 2).

## 6.4 Analysis & Discussion

Due to time constraints the scope of the project had to be narrowed down to just one development stage instead of the planned two. This was mainly because of the added complexity of developing from home due to the spread of the Covid-19 virus. The altered working environment due to working from home, the lack of documentation for Picasso.js and Nebula.js plus the added inconvenience of not being able to talk in real life with our colleagues at the Qlik office when we encountered problems were the biggest reasons for the decreased efficiency in the development.

There are others factors worth mentioning to why the product did not meet the standards we had hoped for. We set a scope for the project that was too large for the time frame we had, regardless of the world pandemic. At the same time when we developed the final concept we wanted to be true to the feedback we had gotten.

Our knowledge in Picasso.js did not increase over time as fast as we had hoped for and Picasso.js had less documentation for some of the problems we encountered than what we could have expected. We knew going in that Picasso.js was a new language for us and that it had limited documentation. We could not however have foreseen when planning the final concept that what we wanted to develop would be a particular hard task to achieve with the aid of Picasso.js. Some of the features so complicated that even the people that coded and built Picasso.js from the ground up gave pause to before giving a complicated answer on how to solve what could have been a less convoluted solution if Picasso.js had support for some of the features that we wanted to develop. For example: a basic attribute of parallel coordinates are the titular parallel axes in the chart. With Picasso.js one can easily create axes at the top, bottom, left or right sides of a chart. But if one wants to add several axes in the middle portion of the chart, something fundamental to parallel coordinate, it somehow is a convoluted affair. It is doable as can be seen in our final product, but the way there took longer than expected. This type of frustrating slow advancement happened several times when trying to develop features for the chart which slowed down the process and delayed our time frame significantly.

# Chapter 7

## Focus Group & Usability Test

---

This chapter will present the final stage of the thesis, the various testings of the product with users and how we circumvented the obstacles caused by Covid-19 to be able to have in-person testing.

Because of the changed circumstances in development, the reduction of the scope and the ongoing situation with the world pandemic we revised the evaluation process. Instead of having qualitative usability tests with experts we performed two online focus groups with both experts and non-experts and five quantitative usability tests with non-experts.

### 7.1 Focus Group

#### 7.1.1 Method

##### Preparation

When the situation changed and it was decided that we were going to do focus groups it was important that we could get the right people for it since this was going to in part substitute having in-person usability testing with experts. We got six volunteers to help us, which were all employees of Qlik. Three of them had been part of the user study and three of them knew of the project we were doing but had not participated in it yet. One group consisted of two visualisation experts and a developer. The other group consisted of one visualisation expert and two developers. We choose to mix the groups based on the volunteers backgrounds so as to get a more varied discussion in each of the focus groups.

For the focus group a manuscript and set of questions were made. This allowed us to present the product in the same way to both of the groups. Since we would not be in the same room as the volunteers and most importantly they would not be getting hands-on with the product we needed to be sure the questions were formed so that they could elicit as much information as possible. The focus group [6], consisted of five parts:

1. Introduction.
2. Basic showcasing of the software.
3. Showcasing the highlighting and selection features.
4. Showcasing the brushing feature.
5. General questions about everything we had shown.

### During the Focus Group

Throughout both of the sessions everyone participating was at their own selected location and participated exclusively via Webex. During the sessions Nicolás moderated the conversations and Niklas shared the screen on his computer where he demonstrated the software. We asked both groups for consent to record the sessions so as to be able to review the answers calmly after the focus groups were done but we both wrote down some of their answers during the discussions as well. Everyone consented to being recorded.

In part **one** we introduced ourselves for those who had not participated in one of our studies before and the goal of the thesis. In part **two** we presented the rudimentary view of the chart without showcasing any features. In part **three** we presented the highlighting and selecting features, showing both how one can highlight lines with the mouse cursor and with the lasso function. In part **four** we presented the brushing function as described in the chapter before. In part **five** we asked general questions about everything they had seen during the presentation. During parts **two**, **three** and **four** we tried to elicit as many opinions from the participants as possible. For each of those sections they were asked:

- Which actions they thought that they could perform by just looking at the screen.
- Their thoughts on the design of the feature.
- What was good about the feature.
- What could be improved with the feature.
- If anything was especially unintuitive.
- What worked well and what did not work at conveying the information.
- If there was anything they thought was completely missing that they thought was crucial.

During part **five** the participants were asked:

- What they would choose to implement next if they could only choose one thing.
- Which of the implemented features was the best and why.
- Which of the implemented features felt worse and why.
- Any last comments that they would have liked to add.

## 7.1.2 Results

Here follows a compilation of all the participants answers divided into which part of the focus group there was an emphasis on. The response in general was very positive towards what had been produced. There was no response to **part one** as it was just an introduction.

### Part Two - Rudimentary Attributes

The first thing they all felt that they could do was both click on the polylines with the mouse cursor and use the lasso function. In general they were all impressed with what had been achieved and how good it looked.

On the more negative but constructive side of things they all thought it looked a little messy on the left side of the chart with all the names of the cars beneath the lines. There was discussion about just moving them to the left of the axis but then one would lose real state on the chart. Another proposition was to take out the names all together, they could show if one hovered over a specific polyline but in general the names of the attributes of the dimension need not be there since one does not usually look at specifics in parallel coordinates, one looks at patterns over the whole chart.

Some did not like that the entire name of the measure is the title of the axes in the bottom, it could be better if the user could choose the title for the name of each axis. There were also other comments about the axes. There could be a buffer on top and below each axis so that the smallest and largest values are not at the edge of the chart. Also to put the value of the smallest and largest polyline crossing each axis, not only the arbitrary scale that Picasso.js chooses by itself.

Another thought was that if lines are on top of each other right now one could interpret it as data not being shown. If one instead only had one color for all polylines and worked with opacity of lines in conjunction with concentration of lines this problem would be solved. If there are lines on top of each other it would just show as a darker line.



## Part Three - Highlighting & Selecting Features

In general they all liked the look and interactions that could be made. They liked how the selected line or lines got thicker and all the other lines got dimmer. We got a lot of comments saying a variation of: "that is exactly how it's supposed to look and work". Participants were also happy that the lasso function did not have a static round or square shape. They liked that one could draw an arbitrary shape to get a highlight over an exact area.

Other feedback we got was around how it could be useful to dim out the names of the cars not only the lines. Also that there were so many values for the dimension it could be useful to be able to highlight by clicking on the name and not just the polyline. It could also be useful with a list beside the chart with the names of all the cars of the lines that have been highlighted.

Another suggestion that we got was that it could be practical to have a function that when one presses the name of an axis the value of the axis gets inverted. This would result in some x-shaped patterns in the chart being *untangled*. This function could give other insights into the data that one could not inherently get right now.

## Part Four - Brushing Feature

This part was where we got the most constructive criticism. But it was not all bad. Some of the experts really liked the way one used lasso to filter out lines. They expressed interest in having such a function in other line diagrams. Right now they would need to highlight and then invert the selection. They thought this way of doing it was faster and more simple. Other positive comments were that they liked the functionality in what could be achieved by it, but they all thought that it was a bit unintuitive.

The first thing they mentioned was that they did not think it was obvious that the symbol we had chosen for our toggle button would signify that one could perform brushing [7]. Two of them suggested maybe using the icon of an eraser but most of them did not know what could be a good substitute.

The main complaint was that they thought it was complicated for the user to first need to highlight an arbitrary line to then be able to have access to the toolbar and the features that were there. This is something inherent in Qlik Sense but there are no charts we know of that have such an elementary function, as brushing is to parallel coordinates, *hidden* in the toolbar. The testers thought the feature both required too many steps to activate it and it was not a clear set of steps to get there, for such a common feature used in the analyzing of parallel coordinates.

There was also a discussion about if the polylines that are being filtered away should disappear entirely, the way that it is right now, or if they should just become very dimmed out. But others then argued that if they are just dimmed it would just be a reversed highlight function. If the lines disappear completely one participant felt there

should be a type of *undo* function that could take the user back one step. Since if the line is not visible one can not press it to un-brush it.

Another comment was that it could have been practical with sliders on the top and bottom of each of the axes that would filter out lines when moved along the axis. This was a function that we really wanted to have implemented and the one that was half developed when time ran out.

## Part Five - General Questions

In this part the participants all expressed that they were impressed with how much we had done with the time that we have had and considering everything that changed around us as we were implementing the project. In general they thought that this was a good first iteration that has a good base for its continued development into something Qlik may be able to release later. They also expressed that at this stage of development it's quite common to get a lot of feedback since the base of a product can not be accustomed to every type of user. In the coming stages of development one can put more time and energy into specific solutions catered to specific users.

When asked about what function or feature they thought should be implemented next if they could choose these were some of their individual responses:

- Somehow fix the clutter of the names and the polylines that are around the dimension axis.
- Be able to hover over a line to see its exact value in any given position.
- Brush sliders on every axis.
- For the default stage to only have one color and work with opacity of that one color to show information.
- Have the titles of the axes on top of them, not beneath them.
- Have the value for the polyline containing the maximum respectively minimum value on each axis.

When asked what they thought was the best thing about the whole presentation and everything we had shown them these were some of their individual responses:

- The interaction of the graph when one highlighted lines.
- That we succeeded building so much of the chart when Picasso.js is not really suited for this type of chart.
- The lasso function when used in conjunction with the highlighting feature.
- The lasso function when used in conjunction with the brushing feature, something he had not seen before and positively surprised him.
- That we had come so far and this could become something that Qlik customers could use.

## 7.2 Usability Test

### 7.2.1 Method

#### Preparation

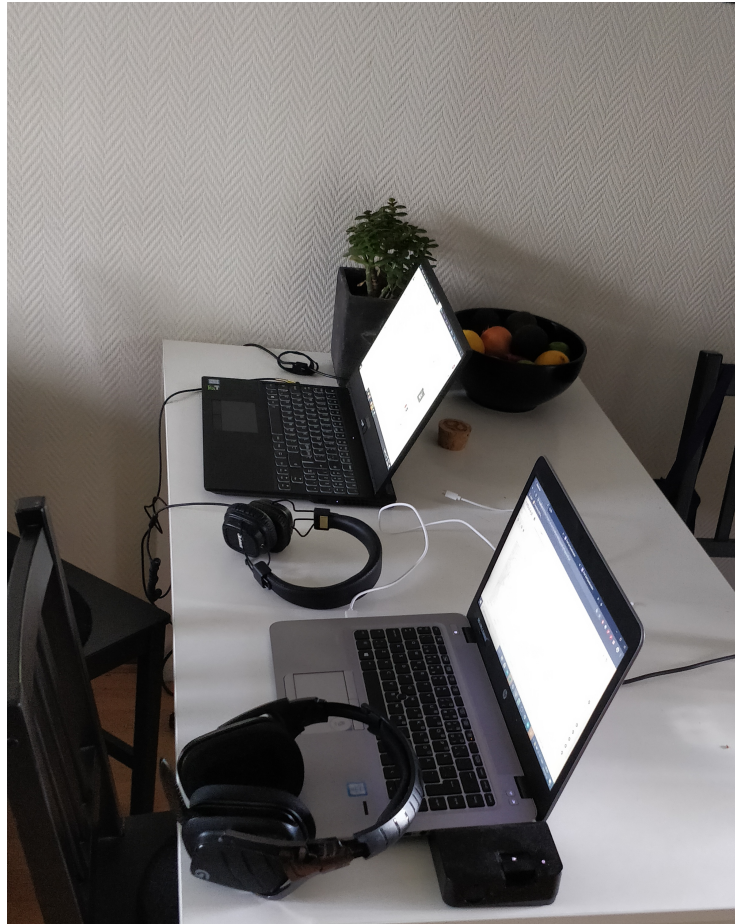
Since we could not get the experts at Qlik to do one on one sessions with us so that they could use the software themselves we needed to find someone who we could get in contact with at this difficult times that at the same time could give us some valuable data. We got five volunteers. All of them computer engineering students. All of them living in close proximity to Niklas's home.

When developing the test we also had to customize it to the type of people we were going to test, namely people without any knowledge of neither Qlik Sense nor parallel coordinates. This meant that in a certain part of the test we gave them a brief introduction to both of those subjects to try and simulate how the product would be tested by an expert. Not a perfect solution but something we thought would give us better measurements. Another difficult part when creating the usability test was coming up with the right tasks for the volunteers to do. We had to have such tasks that they would not feel over redundant but still tested all of the features that we wanted tested in the software so that we could gather feedback from all aspects of the developed product [11]. The usability test, whose manuscript can be seen in its entirety in *Appendix D*, consisted of seven parts:

1. Introduction.
2. Exploratory Testing.
3. Qlik Sense & Parallel Coordinates.
4. General Usability.
5. Highlighting Selection.
6. Brushing and Switching Between Features.
7. Post-Test Questionnaire.

#### During The Usability Test

The usability test was performed in two locations. Nicolás was at his home. Niklas and the participant were in Niklas home where Niklas had prepared a specific setup for the usability test to be able to work when Nicolás was in another location. The setup that can be seen in figure 7.1 consisted of two computers and their respective headset. One for the volunteer to be able to interact with the software and hear Nicolás instructions as he was the moderator. The other for Niklas to be able to hear Nicolás speak at the same time that he could write down not only the opinions of the volunteer but their actions as well as he was sitting in such a position that he could see how the volunteer interacted with the software. Niklas also assisted the volunteer if any questions arose where there needed to be an explanation done on-site. The communication between the computers was done via Webex.



**Figure 7.1:** An overview of the setup used in Niklas's home for the usability test. The fruit in the background was not part of the test.

In part **one** we introduced our thesis and explained the layout of the usability test. Here we also were thorough in explaining to the volunteer that we wanted them to *think aloud* [29] throughout the whole usability test so that we could know what they were thinking. In part **two**, exploratory testing [30], we gave them free reins with the software. This was a way for us to let them get comfortable with the software up to a level that maybe someone who is used to working with Qlik Sense would be. We also wanted to get a first impression on which affordances [6] the software had. In part **three** after the volunteers had gotten a bit more comfortable around the software we gave a short introduction to parallel coordinates, something none of them had heard of. We also had a short explanation of what dimensions and measures mean. This is part of everyday language at Qlik but something none of them had any experience with.

In the next three parts, **four**, **five** and **six**, we wanted the volunteer to test different features. So that we would not give anything away and they could discover as much as possible on their own we gave them tasks that they had to perform. The goal was that as they were performing the task they would simultaneously test the feature we had intended to be tested for that task. This way we could see which parts of the software were more intuitive than others. In part **four** the volunteer was presented with a blank slate

of a chart without any data and their first task to perform was to add the dimension and measures that we were going to need to use in the other tasks, that information was not conveyed to them. In part **five** the task the volunteer got was about how the highlighting and selection features worked. They got to try out both clicking with the mouse cursor and the lasso function as well as confirming a selection and resetting the chart to no selections made. In part **six** the task revolved about the brushing feature and the switching between brushing and highlighting. They got to brush out unwanted polylines, highlight interesting metrics in wanted lines and lastly confirm the selection.

## Post-Test Questionnaire

In part **seven** we asked all of the volunteers to fill an online post-test questionnaire [11]. They could do so in the computer that they were already sitting in front of. The questionnaire consisted of three parts. A part where the test subject could write open text answers about certain things we wondered especially about. A part where the test subject, using a Likert scale, gave feedback on how intuitive certain aspects that they were tested on were [11]. And lastly another open text question where they could add any last comments they had, either about the chart or the usability test in itself, if they wished. The open text questions were:

- How did you experience the interface in itself? What was hard and what was easy? What was clear and what was not clear?
- Did all the functions do what you expected them to do? If not please elaborate.
- In your opinion what is the main use for the filtering function?

The questions asked, were the volunteers responded using a Likert scale from one to five were one was not intuitive at all and five was very intuitive, were:

- How intuitive was it to load the dimension and measures?
- How intuitive was the selection function?
- How intuitive was the filtering function?
- How intuitive was it to find the filtering function?
- How intuitive was the lasso function?
- How intuitive was it to understand the information from the graph?

## 7.2.2 Results

### Exploratory Testing

In part **two**, exploratory testing, we can divide the participants interactions into two groups, interaction with the chart and interaction with Nebula.js. All of them except for one started by interacting with the chart by clicking on lines to highlight them. The one person that started by interacting with Nebula.js chose for some reason to never interact with the chart. All of the others eventually found themselves to the *confirm selection* button in the toolbar. One of them said aloud that they were looking for a function to be able to zoom with, something that was not implemented. Four of them lastly interacted with the ways of entering data and took away some of the measures before they themselves felt the need to put them back.

### The Tasks

Everything went smoothly in part **four**. Since they all had explored their way around this section of Nebula.js in the previous part they all knew what to do when asked to choose certain dimensions and measures. The only noteworthy thing is that in Nebula.js when you press to put in a new dimension or measure a drop down menu comes forth. Three of the volunteers exclusively scrolled down to the wanted values while two of them exclusively used the search function that can be found at the top of the drop down menu and wrote what they were searching for.

Part **five** is where we encountered for the first time that some of the volunteers did not know what to do. Two of them had no problems whatsoever and cruised through the task without complications. One thought that, when the task was to select a car, they could only press on the car name, that gives no response. After some trial and error they understood that they only got a response if they clicked on a line. None of the participants found it hard to proceed when given the task to *confirm the selection of the highlighted lines* meaning that the *confirm selection* button was well understood. The biggest problem in this part was when three of the volunteers could not find their way back to clear the selections they had made. There are two ways to do so. One of them is the same as one would find in Qlik Sense and would not have been a problem for an expert user and the other is an overlay found in Nebula.js. Nonetheless it took them quite some trial and error before understanding how that part of the system worked.

In part **six** there was sadly a part of a task that none of the participants could proceed through without help from us. None of them highlighted a line to then be able to press the toggleable button for brushing. They thought that it was unintuitive one would need to highlight an arbitrary line to then be able to brush. They also thought it was unintuitive that all of a sudden the lasso function would now clear away lines instead of highlighting. They did not understand either that the brushing function does not reset if one resets the chart. Only if one resets the whole window on the browser or presses the toggleable button again. The rest of the task went smoothly as it was based on things they had done in the previous task. Throughout all the tasks if they had al-

ready done a thing, next time it came up they thought it was very intuitive.

## Post-Test Questionnaire

### Open-Text Questions

There was not much added to the context of the data collected by the open-text questions. The participants felt that they had, through the talk aloud method, gotten out most of what they wanted said already in the previous parts of the usability test. Below is a compilation of the **new data** that the answers gave divided by question.

#### **How did you experience the interface in itself? What was hard and what was easy? What was clear and what was not clear?**

- Clean chart with good visualisation.
- The interface was complicated at first when I did not understand parallel coordinates but got easier along the way.
- It was unintuitive to exclude data by brushing it away.

#### **Did all the functions do what you expected them to do? If not please elaborate.**

- The toggleable brushing button was the only one that did not have information when you hovered on it, that was unclear.
- The brushing feature should be accessible directly at the beginning without the need to highlight a line.
- It was very unexpected that that the brushing button was toggleable.

#### **In your opinion what is the main use for the filtering function?**

- To remove information from the chart.
- To make the chart clearer to view by taking away unwanted information.

#### **Do you have anything that you want to add?**

- To be able to hover on a line and know which car it corresponds to.
- Tool tip for the brushing button.
- The ability to choose an axis and then type in values and see which polylines that are between those values would be good.

**Likert Scale Questions**

In table 7.1 one can see the mean value of the answers to the Likert scale questions. Just a reminder that the scale was from one to five with one being not intuitive at all and five being very intuitive.

<b>Questions</b>	<b>Mean Value</b>
How intuitive was it to load the dimension and measures?	<b>4.4</b>
How intuitive was the selection function?	<b>4.4</b>
How intuitive was the filtering function?	<b>3</b>
How intuitive was it to find the filtering function?	<b>2.6</b>
How intuitive was the lasso function?	<b>4</b>
How intuitive was it to understand the information from the graph?	<b>4</b>

**Table 7.1:** Table showing the mean value results for the Likert scale questions



## 7.3 Analysis & Discussion

This is purposely a joint discussion about both the focus group and the usability test since the participants on both those activities reviewed the same product. We did not alter the software at all between the focus groups and the usability test. Also worth noting both of the activities were done entirely in Swedish. Both in the previous focus group and in the user study we had non Swedish speaking participants and since we wanted all the experiences to be as similar as possible we developed everything in English. This time we knew all of the participants were Swedish speaking before developing too much of the material so we chose to hold both the focus groups and the usability test in Swedish. We felt this also gave freedom to all of the participants to be able to express themselves better when they could use their native language. Before the results are discussed a few observations about how everything went will be made.

Having a focus group where everyone was sitting remotely in their respective homes worked surprisingly well. In comparison to the first focus group where all the participants were in the same room, being on Webex kept the dialog flowing with no one interrupting one another. Worth mentioning is also that no one in the group really had differentiating opinions and that could also be a reason for the good flow and lack of interruptions. Something else that happened in the focus groups was that as soon as a question was asked the participants got to talking and not only answered that question but perhaps also parts of other questions we had not yet asked. We did not want to halt conversation so we did as best we could to keep a tally on what had been answered and which things had not yet been discussed about the topic at hand.

The usability test went very smoothly and there was a good rapport between everyone involved as we all know each other since before. We think this made the participants calm and let them feel that they could admit when they did not know how to proceed. We wanted to have people with more varied background for this part to see if the answers we would get would vary more but considering everything we are very happy with the participants we got. As we are ourselves computer technology students it is only elementary that most of our social circle, people that would help in a situation like this, would be in the same demographic.

During the usability test another thing that separated the participants was the way that they chose to search a drop down menu. Some of them scrolled down to the measure we wanted them to choose while others used the search function at the top of the menu. We think that this is strictly a preference based on what one is used to work with. If the list is long maybe the search function would be more effective but if the user is not used to it it could take longer to begin the process. Nonetheless the way this function works is inherent in Nebula.js and is neither something we designed nor something that can be changed.

Most of the participants thought that the chart in general was well done and well designed. They really liked the highlighting and selection features and thought it felt intuitive. In part five of the usability test some of the participants had trouble under-

standing how to highlight a car. There were simply not enough signifiers for everyone to immediately understand that you could click on a polyline to highlight a car [6]. Instead, many of them tried first to click on the name of a car to highlight it.

The main negative thing that was remarked upon was also something that all of the participants brought up. That was that the way to activate the brushing function was too complicated and felt unintuitive. This indicated that the product both lacked discoverability and learnability since the participants did not understand how they should approach the solving of the task given [7] [6]. When the participants were given the explanation on how to solve the task it was revealed that they thought the toggle button for the filtering function did not give enough feedback that the button was toggleable [7]. Several of the participants instead thought that you used the button to filter out lines that were already highlighted. The affordances for filtering out data were there and were received but the signifiers did not convey the type of interaction we wanted the participants to take [7]. In general, they thought that feature was very unintuitive. Unintuitive, is a word we have used a lot but we feel describes exactly what the participants felt. A way to solve this could be to not have it in the toolbar and instead have a button that is always part of the chart. The negative side to that would be the real state it would occupy and that it would make the graph even more complicated at a first viewing. Some of the participants in the usability test, people not accustomed to Qlik Sense gave the idea to just have the toolbar always be visible. Even if this is something we agree on it's something that is not a Qlik Sense standard and we have a hard time believing that the company would change a fundamental way of how their software works for a chart that is not released yet.

Looking at the argument from the other side, Qlik Sense is known for having a learning curve, something that will not change anytime soon. Maybe not all functionalities have to be obvious and self evident from the first time one sees them. As stated, as soon as we had shown to the participants how a function was accessed they had no trouble in using it the second time they were asked. And most of the criticism was the accessibility of the brushing function. The brushing function in itself got some mixed reviews but most people liked it after getting comfortable with it and understood its purpose.

# Chapter 8

## Final Analysis & Discussion

---

In this chapter we will discuss how the use of the iterative design process worked for us, why we did not get all the results we expected, the answers to the research questions and how the product may continue to be developed in the future.

### 8.1 Methodologies Used

In this project, several types of methods have been used during the iterative design process [6]. Using interviews, focus groups and usability tests have been a great way to elicit the needs of the users. The project has been user centered and the aim of the methods used has always been to elicit the need of the users. In our opinion all of the methods have worked well and produced relevant and helpful data. The only negative thing associated to the methods used during the project is not due to a fault of them or how we utilized them. A recurring problem in the project in general was finding suitable test participants. Either due to full schedules or a shortage of people to ask with the right background. This posed a problem since without users to test with it can be difficult to understand what the users need and want.

#### 8.1.1 User Interview

The user interviews took a long time to prepare and execute, mostly due to the troublesome process of finding test participants and then finding the time in the schedules to actually have the interview. The interviews did give a lot of valuable metrics and new insights that were not previously taken into consideration [8]. As this was the first of the methods used in conjunction with participants during the project it gave a disproportionately large amount of feedback that could be used in later stages of the project.

### 8.1.2 Focus Group

There were three focus groups performed in the project, the first one was to discuss the elicited needs that were gathered in the user interviews and the second and third ones aimed at discussing the results of the implementation and possible ways to develop the application further [6]. The first focus group provided some much needed background to both the Qlik Sense functionality and also some context about why some functionality like filtering can be problematic to implement, either because of limitations or because of principles. The results from this focus group provided a clearer image about how the features would be implemented and in what priority order. The second and third focus groups also provided valuable data. The second group consisted of more seasoned Qlik Sense apps developers which could provide experienced and technical feedback while the third consisted of more novice or design oriented developers which provided more user based feedback which focused more on the GUI. The data gathered from the second and third focus groups, serves as the background to what should be developed in the future to make sure that the chart has high usability and is a powerful analysing tool.

### 8.1.3 Usability Test

The usability tests was performed late in the project. The original plan was to develop and test and the repeat the cycle at least once [11]. But due to the narrowing of the scope, time constraints and other factors only one development cycle and one test was performed. The data from the test however provided some insights that had not surfaced before, that of users who were new to both parallel coordinates and visualization in general. These insight were more focused on the intuitive nature of the graphical user interface and the signifiers of said interface [6].

## 8.2 Sources of Error

Here we will discuss some of the factors that made have been responsible for either diminishing the scope or influence why the original goals of the thesis could not be met.

Due to corporate restrictions and processes we did not have access to interviewing and testing people from the target audience, customers of Qlik who use Qlik Sense daily. This was also the case with both the focus groups and the usability tests, the consequences of this is that we might have lost some qualitative insight and data that we could only have accessed through testing a target user.

Continuing the discussion on the participants. We are content with the experts at Qlik who gave their valuable time helping us but it was not always easy to coordinate meetings and tests with them, neither when we were testing them individually nor in groups. Many of their schedules are full and since we did not have a large pool of participants to choose from we needed to wait sometimes for many days before being able to advance to the next stage. This was most prominent during the user study, which took almost two weeks to complete the interview phase.

Another hindrance was the restricted amount of features Picasso.js supported since it's a fairly new library and things tend to get implemented when the need for new features arises. A lot of what we wanted to develop were new concepts that had not been previously developed with Picasso.js. Had it not been for the other developers in the department that could provide assistance, answer questions and make modifications to the source code to accommodate us the development might have gone even slower. More about this can be read in in section 5.4.

Due to the Covid-19 situation the scope had to be dramatically reduced in several aspects that influenced each other. When everyone at Qlik got ordered to work from home we lost the easy access we had to our colleagues at the office. This meant that less features were developed and the minimal features that were developed took longer than expected. That in turn resulted in a time constraint that reduced the scope further. When the developing stage took so long it was decided to only do one round of testing instead of the initially planned two rounds. Since we could only get access to the Qlik experts via Webex we could not do the usability test with them because the software could not be exported to any system in the stage that it was in. This meant dividing the testing phase into focus groups with experts and the usability test with whomever we could get to meet us in person. We could only get other engineering students for the usability test and they sadly lacked the expertise required to properly understand how Qlik Sense works and what parallel coordinates are to give as much meaningful data as possible. At the same time this gave us the opportunity to test the software on fresh participants not accustomed to this type of program. They could then react to the affordances in our program and in Nebula.js with a fresh set of eyes and give more feedback into some things that the experts would have not noticed, since they would already be accustomed to them [7].

Another aspect that might have had an impact on the usability test is the fact that the usability test was performed in Nebula.js and not in Qlik Sense. The reason that this might have affected the outcome is that Nebula.js is a scaled down version of Qlik Sense and they have some differences in their interface. For example the fields where one chooses dimensions, measures and other metrics are not part of the initial screen in Qlik Sense.

## 8.3 Research Questions

### **RQ-1: What is important to think about when choosing which type of visualization tool to use when visualizing multivariate data?**

Just because a chart has the possibility to visualize multivariate data it doesn't mean it is the right tool for the job. One has to know which tool is the appropriate one depending both on the type of data and what one wishes to get out of it. One can for example use a modified scatter plot to visualize multivariate data but it might not enhance the right aspects of what wants to understand with said data. If one wants to see the patterns in data with lots of variables maybe parallel coordinates is the right tool for you. If you instead want to find and showcase outliers in the data it might be better to show another visualization tool.

### **RQ-2: What is important to think about when visualizing parallel coordinates?**

From our own experiences developing the chart and foremost from the feedback we have gotten from the volunteers in all the tests, focus groups and interviews it seems that people want a very clean and uncluttered design around the chart. A parallel coordinates plot can already be challenging for some people to gain insight into and for them to have an uncluttered view around the chart is important. At the same time all the common features should be available to the user if they wish to interact with the chart.

### **RQ-3: How does the choice of coloring influence the users understanding of the data shown in a parallel coordinates chart?**

We narrowed the way that a chart can be colored into two options, depending on the data visualized. Either one wants all the polylines to be different colors so as to make it easier for the user to see specific values outlined in the plot. Or one would want all the polylines to be the same color. Then, lines that are on top of each other increase in opacity so as to visualize the increase in data intensity in an area with the increased opacity of the polylines.

### **RQ-4: Which feature is the most important to help the users understanding of a cluttered parallel coordinates chart?**

Feedback from the tests performed indicated that both highlighting and brushing in conjunction with selecting were all very appreciated when trying to declutter a parallel coordinates chart or simply to parse through the data in a chart. Something that many participants, both expert users and developers, were disappointed that we had not had the time to develop were ranged sliders on the top and bottom of every axis. This is something that can be seen in many prototypes of parallel coordinates but not all. Those who have had interactions with it swear by its function to be very intuitive, easy to interact with and give a better understanding of the data being shown.

## 8.4 Future Work

As commented before we had to reduce the scope of the project but we had plans as to what we would have liked to continue doing if we have had more time. We would have foremost liked to do another usability test after having revised the version we did the first usability test on. Next, we would have liked to do a test with the same persons we did the first test with as well as do the same test with some experts to see how the feedback would change. We would also have liked to implement some features that we think would have made the interacting with the chart more fluent. First we would have liked to implement sliders on all of the axes for them to work as a way to brush the polylines, that was the last thing we were working on when we realised we had no more time for that phase. The other thing we would have liked to do is export our software as an extension so that everyone with access to Qlik Sense could download it and test it for themselves.



# Chapter 9

## Conclusion

---

In this chapter we close the report with some finishing thoughts.

From the feedback we got from the expert users and developers as well as the participants in the usability test it seems that what we have achieved has potential to continue to be developed. Our goal was never to produce a final chart that Qlik could put into their next patch. We wanted to make a working prototype of a chart and export it as an extension and we came a few steps short of that. We believe that Qlik wants to develop a fully functional parallel coordinates chart sometime soon and we hope that they can use the work we have produced to help them carry out that task. Either that be the technical side of it with the code that has been written or it be insights in user interaction and user preferences from the different tests and interviews we have performed.

# References

---

- [1] Various. Facebook–cambridge analytica data scandal. [https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge\\_Analytica\\_data\\_scandal](https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal), accessed:2020-05-20.
- [2] Various. Westworld (tv-series). [https://en.wikipedia.org/wiki/Westworld\\_\(TV\\_series\)](https://en.wikipedia.org/wiki/Westworld_(TV_series)), accessed:2020-05-20.
- [3] Qlik. What is data literacy. <https://www.qlik.com/us/bi/data-literacy>, accessed:2020-05-21.
- [4] R. Mazza. *Introduction to Information Visualization*. Springer-Verlag London Limited, 2009.
- [5] Qlik. Qlik homepage. <https://www.qlik.com/us/> accessed:2020-05-21.
- [6] J. Preece, H Sharp, and Y Rogers. *Interaction Design*. John Wiley & Sons Ltd, 2015.
- [7] Don Norman. *The Design of Everyday Things*. Basic Books, 2013. New York.
- [8] Kara Pernice. User interviews: How, when, and why to conduct them. <https://www.nngroup.com/articles/user-interviews/>, accessed:2020-05-22.
- [9] Douglas Harper. Talking about pictures: a case for photo elicitation. *Visual Studies*, 17(1), 2002.
- [10] S. Burge. The systems engineering tool box. 2009.
- [11] J. Rubin and D. Chisnell. *Handbook of Usability Testing*. Wiley Publishing, Inc., 2008.
- [12] M. Arvola. *Interaktionsdesign och UX*, volume 1. Studentlitteratur AB, 2008.
- [13] C. Magnusson, K. Rassmus-Gröhn, K. Tollmar, and E. Deaner. User study guidelines. 2009. [http://www.certec.lth.se/fileadmin/certec/publikationer/HaptiMap\\_d12.pdf](http://www.certec.lth.se/fileadmin/certec/publikationer/HaptiMap_d12.pdf).

- 
- [14] J. Johansson and C. Forsell. Evaluation of parallel coordinates: Overview, categorization and guidelines for future research. *IEEE Transactions on Visualization and Computer Graphics*, 22:579–588.
- [15] A.O. Artero, M. Oliveira, and H. Levkowitz. Uncovering clusters in crowded parallel coordinates visualizations. *IEEE Symposium on Information Visualization*, pages "81–88", 2004.
- [16] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. *IEEE-VGTC Symposium on Visualization*, 27:1047–1054, 2008.
- [17] J. Johansson, C. Forsell, M. Lind, and M. Cooper. Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization*, 7:152–162.
- [18] N. Cawthon and A. Vande Moere. The effect of aesthetic on the usability of data visualization. *2007 11th International Conference Information Visualization (IV '07)*, page 637–648, 2007.
- [19] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. Seidel, and T. Weinkauff. An edge-bundling layout for interactive parallel coordinates. *2014 IEEE Pacific Visualization Symposium*, page 57–64, 2014.
- [20] P. Guo, H. Xiao, Z. Wang, and X. Yuan. Interactive local clustering operations for high dimensional data in parallel coordinates. *2010 IEEE Pacific Visualization Symposium (PacificVis)*, 2010.
- [21] J. Heinrich, Y. Luo, A. E. Kirkpatrick, H. Zhang, and D. Weiskopf. Evaluation of a bundling technique for parallel coordinates. *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*, 1:594–602, 2011.
- [22] D. Pomerence, F. L. Dennig, D. A. Keim, J. Fuchs, and M. Blumenschein. Slope-dependent rendering of parallel coordinates to reduce density distortion and ghost clusters. *2019 IEEE Visualization Conference (VIS)*, 2019.
- [23] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005.
- [24] J. Johansson, P. Ljung, M. Jern, and M Cooper. Revealing structure in visualizations of dense 2d and 3d parallel coordinates. *Information Visualization*, 5(2):125–136, 2006.
- [25] Z. Geng, Z. Peng, R.S. Laramée, R. Walker, and J.C Roberts. Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12).
- [26] Qlik UX department, Interview, November 2019.

- [27] BusinessDictionary. Interviewer bias. <http://www.businessdictionary.com/definition/interviewer-bias.html>, accessed:2020-05-21.
- [28] Cameron Chapman. Color theory for designers, part 1: The meaning of color. <https://www.smashingmagazine.com/2010/01/color-theory-for-designers-part-1-the-meaning-of-color/>, accessed:2020-06-4.
- [29] R. Jääskeläinen. Think-aloud protocol. *Handbook of Translation Studies*, 2010.
- [30] J. Itkonen and Rautiainen. K. Exploratory testing: A multiple case study. *2005 International Symposium on Empirical Software Engineering*, 2005.

# Appendices

# **Appendix A**

## **Manuscript of the User Study**

---

In the written manuscript anything written within parentheses is a note to the interviewers and was never said aloud during an interview, they were only clarifications and instructions.

# Opening Text

--SLIDE--

(Each time the --SLIDE-- sign is shown it is a cue to change the slide to the next one)

The goal of this interview is to get to know what you as a user of Qlik sense would like in a visualization graph. There are no wrong answers and we encourage you to speak your mind as much as possible so we can get as much information as possible. If there is a question that you don't understand or if you think I am not being clear in my either my reading or explaining of the questions don't hesitate to ask for me to repeat the question or for me to explain the question further.

The interview will be divided into seven different topics that will begin with an explanatory text of each section.

The topics of the questions will get more complicated as the interview advances. This doesn't mean that the questions themselves will get harder but the concepts will be more complex. This means that some of the questions in the beginning might seem rudimentary.

We also ask of you if you could say when you are done answering a question so that we know when to proceed with the interview.

Do you have any questions before we begin?

## Open Ended Questions - General Visualization

--SLIDE--

### **Opening text:**

We are now going to ask you some general questions about how you would like to visualize data.

### **Question # 1:**

How much multivariate data do you work with or have worked with?

Not at all? Some? A lot?

### **Question # 2:**

What would you say are the most important aspects of data visualization in general?

### **Question # 3:**

What would you say is the differences in how you visualize and present small quantities of data and large quantities of data?

### **Question # 4:**

What would you need from a graph visualizing multivariate data to better understand the data?

(In comparison to 2 dimensional data)

### **Question # 5:**

Do you have any thoughts regarding how color can be used when visualizing multivariate data?

What purpose, do you think, can color serve when visualizing multivariate data?

# Parallel Coordinates Explanation

**--SLIDE--**

**If the volunteer said that they didn't know what a parallel coordinates plot is:**

Explanation of Parallel coordinates:

Parallel coordinates are a common way to visualize and analyze multivariate data.

An example of an over simplified parallel coordinates plot can be seen in the figure shown for you now for better understanding of the concept.

There are two or more axes that are parallel to each other and there is one axis per variable.

Each axis scale can have their own range or unit depending on the variable.

Values for each variable are shown as connected polylines between the axes.

Furthermore, the order in which the axes are arranged can greatly impact one's understanding of the graph. This is because the relationships between adjacent variables are easier to perceive than for non-adjacent variables.

Most of the questions are just about your perception and how we could develop it into a type of graph that you would like to use so do not feel any pressure and as stated before there are no wrong answers.

## Colouring

**Opening text:**

**--SLIDE--** We are now going to ask you some questions about color and coloring in regards to parallel coordinates plots.

**Question # 6:**

Do you have any thoughts or ideas on how you would like to use color to help the visualization of parallel coordinates plots?

**Question # 7:**

**--SLIDE--** When first plotting out a parallel coordinates graph how would you like the default coloring to be, before editing anything?

**--SLIDE--** Should all the polylines have the same color to begin with or **--SLIDE--** should they all be of different colors?

**Question # 8:**

**--SLIDE--** When you add colors to a parallel coordinates plot as an analysing technique, how would you like to be able to display the colors?

(let them answer before continuing with follow up questions)

**--SLIDE--** Would you like for the colors to be decreasing or increasing in a gradient for polylines that are beside each other?.

Or should polylines that are beside each other have colors with high contrast to each other?

Which of these two do you think gives a better understanding of the data?

Or would you like to have the option to give color to each and every single line separately?.



# Brushing

## Opening text: --SLIDE--

The next set of questions are about a technique called brushing, in the context of parallel coordinates, do you know what brushing refers to?

(If they answer no: Brushing is basically filtering, is the means to limit how much of the information one sees at any one time. This is because when one plots a lot of data, parallel coordinates tend to get very cluttered and one needs tools to declutter, so to say, the information.)

(If they answer yes: Great!)

## Question # 9:

Before we show you anything we would like to know how you would like to be able to brush or filter parallel coordinates? Any ideas of your own?

## Explanatory text about the next questions:

We will now show you an array of different ways that one can brush parallel coordinates.

We will go through each and every one and we would like to hear both what you like and what you do not like about each one.

What brushing options would you like to have in a parallel coordinates plot?

- --SLIDE-- Manual sliders along every axis that diminish the amount of lines that can be seen. All polylines outside of these manual sliders would be filtered out ?
- --SLIDE-- By clicking on one or several polylines and having the rest of them be filtered out? In this example we have three polylines that we have clicked.
- --SLIDE-- Marking an area in the plot with the mouse cursor, in effect filtering out all lines that are outside?
- --SLIDE-- A field where you can type the exact boundaries for each axis, this field would most likely be in the “edit fields” in Qlik sense. All polylines outside of these boundaries would be filtered out?.

## Question # 10:

--SLIDE-- Now that you have seen some common ways of brushing do you have a suggestion of what other methods you would like to see implemented?

## Question # 11:

Of the different brushing techniques that we showed you:

Which one or which ones did you like the **most** and why?

And which one or which ones did you like the **least** and why?

## Question # 12:

Could you order them from best to worst?

## Question # 13:

Is there a certain combination of them that you think is the most optimal?

## Question # 14:

--SLIDE-- When brushing, the polylines that will be filtered out, would you prefer for them to disappear entirely or for them to be translucent and or lighter in color?

# Features to Analyze Large Amounts of Data

**--SLIDE--** As we said before, parallel coordinates plots tend to get very cluttered very fast when you have large amounts of data. Apart from coloring and/or brushing there are other features that have been developed to help one analyze large amounts of data better.

**Question # 15:** Do you have any ideas on how best to visualize large amounts of data on a parallel coordinates plot?

**Explanatory text about the next questions:**

We will now show you three different features and describe them shortly. After that we want you to order them from best to worst. Do this in regards to both visual appeal and how good they are at conveying information. With the first one being the most visually appealing and/or best at conveying information.

**--SLIDE--** First we have **Clustering**. Clustering is grouping together polylines based on a pattern that follows through the whole parallel coordinates plot. Clustering is often used in conjunction with coloring to denote the different groups, with different patterns, that appear in the graph.

**--SLIDE--** Then we have **Bundling (and Edge Bundling)**. This can be seen as a type of clustering. Apart from dividing the polylines into patterns you also manipulate the form of the polylines themselves. There are many different types of this. The most common is to bend the polylines into Bézier curves or to group all lines in a cluster together in the middle of two axes.

**--SLIDE--** And lastly we have **Histograms**. They are usually shown horizontally along one or both sides of the axes to show where in the axes there is a higher or lower concentration of polylines. This is used to show the different concentrations of polylines at different areas of every axis.

**Question # 16:**

**--SLIDE--** So now that you have heard of these three (four) different features please rank them from best to worst and comment on why you think one feature is better or worse than the other ones.

**Question # 17:** Now after having seen these three (four) features do you have any other ideas on how best to organize large amounts of data on a parallel coordinates plot?

# Histograms

## Opening text:

**--SLIDE--** The next set of questions are about one of the features that we just showed you, histograms. They are a powerful tool that can be customised in various ways.

## Question # 18:

Based on you already seeing histograms compared to other features. Are histograms something that you would like to have in a parallel coordinates plot or not?

(let them answer before continuing with follow up questions)

Would you like them to be there as a default or as something you add to the plot when you want the feature?

## Question # 19:

**--SLIDE--** Would you like to have the histograms on just one side of the axis or on both sides?

## Question # 20:

**--SLIDE--** Should they be semi transparent or opaque?

## Question # 21:

**--SLIDE--** Would you like the histogram to show the value of the amount of polylines that are crossing through it? How? Where?

**--SLIDE--** As a permanent number or only when hovered over with the mouse cursor?

## Question # 22:

**--SLIDE--** Is there any other shape that you think would be better than a rectangle as a histogram? That you think is either better at conveying information or more visually appealing? For example, circles, ovals, triangles, star shaped?

(When they choose a shape or stay with rectangles)

Why do you think this shape is better than rectangles?

Or

Why do you think that rectangles are the best shape for this?

## Question # 23:

**--SLIDE--** Do you have any ideas about how you would like the coloring of the histograms to be?

## Question # 24:

Would you like the histograms to all be the same color? Different colors depending on the concentration of polylines going through it?

Or maybe for them to change in a gradient that is relative to the size it has i.e. that changes in color the further the histograms edge gets from the axis?

**Question # 25:**

**--SLIDE--** Histograms can also be used in brushing. One could, for example, click on a specific histogram and filter out all the other lines. Compared to the other brushing techniques that we showed you, where would you put brushing with histograms in? **--SLIDE--**  
(here we will show a slide with all the brushing techniques again).

## Final Questions

**Opening text:**

**--SLIDE--** Now we just have some last questions to sum up the whole interview.

**Question # 26:**

Are there any ideas or features that you would like to see implemented that we haven't mentioned?

**Question # 27:**

Of the things we have mentioned, which ones would you say are a must have and are there any features that we have talked about that you think that we should not implement? **--SLIDE--**

**Question # 28:**

**--SLIDE--** Lastly, is there anything else you would like to add? Any feedback on either parallel coordinates or the interview itself?

## Closing Text

Thank you very much for agreeing to participate in our interview.

Later during our developing stage we will do an open interview with a focus group to see which of the concepts derived from these interviews will be the one that we are to implement.

Later after the implementation phase we will also need participants for the testing phase. Is it ok if we contact you when these are relevant, to see if you would like to participate in those tests?

Thank you very much for your time, have a nice day.

# Appendix B

## Concepts One, Two & Three

---

### Concept One

Very standard and similar to how Qlik charts usually work and how the user usually interacts with them while at the same time allowing for the requests that Qlik had for our project, specifically that there would be a clear distinction between filtering and selection.

All of the different changes to the chart are done in the edit tab. With the major things one can change being:

- Turning brushing on and off
- Turning histograms on and off
- Changing the appearances of the chart

If one clicks or lassos something in the chart it will automatically be highlighted and a button to approve selection will come up. If that button is pressed one will go into selection.

The only brushing that can be done in this concept is by using sliders on each of the axes. These are not on the chart by default meaning that there is no brushing in the default state of the chart. In the edit tab one can go into analysis, then brushing then turn on the sliders from there. In this concept brushing is completely separated from highlighting and selection so that neither of them can influence or react with each other.

---

## Concept Two

Data, sorting and appearances are as usual in the edit tab but all of the other interactive tools are now found in the toolbar located above the chart.

The toolbar has buttons for easy access to:

- Brushing
- Histograms
- Selection

The standard state of the PCP is that none of the buttons are on the on mode and that if you click or lasso something it will become highlighted.

Interaction between features:

### Brushing

- When the brushing buttons is pressed both the clicking and lasso functions become ways to brush as well as sliders on each of the axes appearing also as a way to brush. To turn off press the button again
- If histograms are on one can also brush with them
- While the Brushing button is pressed the selection button is in the off mode
- When you turn the brushing off the selection button will go to whichever state it was before pressing the brushing on.

### Histograms

- When one presses down the histogram button histograms appear along each of the axes on the right of the axes. Several changes to this can be done in the appearance field in the edit tab
- Histograms are independent of the other two buttons
- If you hover over a histogram the number of polylines that are not brushed will appear inside the histogram
- If histograms is on and brushing is on one can use the histograms as a way to brush. If you press on a histogram is as if one had lassoed that section
- If one turns off brushing after making a histogram brush and histograms are still on the lines brushed appear translucent/brushed but one can not interact with the histogram any more in a brushing way
- If one has done a histogram brush and one turns off histograms, regardless of whether brushing is on or off the histogram and the brushing done with the histogram will disappear, other brushing will be left as it was

### Selection

- When the selection button is in the on mode all the clicking and lassoing one does will not only highlight what one has chosen but selected them as well. And that will reflect in what shows in the other charts in the app.

---

## Concept Three

Qlik's standard features such as data and sorting are as usual in the edit tab but all of the other interactive tools are found only in a context menu that appears when you right click the computer mouse.

The context menu gives you easy access to:

- Appearance
- Brushing
- Histograms
- Selection

The interaction between brushing and highlighting/selection is the same as the interactions described in concept two.

# **Appendix C**

## **Requirements Specification**

---



# Requirements Specification

## General

There will be n number of axes. ( $n \geq 1$ ).

There can only be one dimension per axis and a total of 10000 measures in total.

The main window will look in its outline like the other Qlik charts.

The toolbar on top of the chart will look the same as in the other charts.

There will be one polyline connecting two connected measures in adjacent axes.

There is the possibility to change the order in which the axes are arranged.

The standard order of the axes is based on the information the user feeds in.

The axes have numbers on their edges based on the min and max of the measure.

## Toolbar

In the toolbar to the left of the always existing buttons there will be:

- A button for turning brushing on and off.
- A button for selection (the one that is already there) (Confirming the selection).
- A button for “clear selection”.
  - This button is only available for pressing if at least one polyline is highlighted.
  - To remove all highlighting that is done in the chart.
- A button for “cancel selection” (big red X).
  - To remove all highlighting that is done in the chart and go back to the standard mode at the beginning.
  - This button is only available for pressing if at least one polyline has been highlighted.

## Appearances

One can not change the appearance of the axes. The color is predetermined.

One can change the interval of the numbers describing the measures in the axes.

The chart will follow Qlik Sense themes.

# Brushing

## Basic functions

- If brushing is on all the interactions with the chart will not be highlighting/selection.
- If brushing is on interactions with the chart will not change or interact with the other charts in the app.
  
- If brushing is turned on brushing slides will appear on the top and bottom of each axis.
- If you move the sliders the lines that are outside of the parameters of one of the sliders will be in brushed mode.
  
- One can also brush lines by selecting the lines that one does not want brushed either by clicking on them or by using the lasso function when Brushing mode is on. All the other lines that are not selected are in brushed mode. One can select or lasso multiple times.

## Brushed Lines

- A brushed line has other properties than a normal or highlighted line.
- A brushed line is more transparent.
- One can change the transparency of the line in the edit menu with a slider.
- The lines can also be made to disappear entirely when they are brushed with a button in the edit menu.
- If the lines have been taken out entirely one can not click or lasso them anymore.
- One can change if one wishes for the brushed line to keep its color or be grey.
- The line will be grey with a 50% transparency as a default.

## Brushing mode interactions

- If one has moved the sliders and then turn off brushing the slides will:
  - Not disappear entirely, they will be translucent, all of them. Not just the moved ones.
  - Stay in their exact place, all of them. Not just the moved ones.
- One can not interact with the sliders when brushing mode is turned off
- The sliders will disappear entirely when they are all reset to their starting positions and brushing mode is turned off.

# Highlighting / Selection

## Basic

- If brushing is turned off the standard mode will be highlighting/selection.
- One can click or lasso polylines to highlight them.
- One can click lines to unhighlight them
- Highlighted lines will be a little thicker and change in color depending on a color scheme that is predetermined (all lines a different color) until a certain amount, after that all the highlighted lines will be of the same color, also predetermined.
- When polyline(s) are highlighted the other charts in the app will react accordingly (in contrast to when brushing is on)
- One can remove all the selections with the “clear selection” button.
  - All the lines will go back to the default state.
  - This button will work the same as the “cancel selection” button and it's mostly here to have continuity with other Qlik charts. For example the “map” chart also has this button and its use is no different than the “cancel selection” button.
- One can reset all of the highlighting done by pressing the “cancel selection” button the big red X button beside the confirm selection button at the right of the toolbar.
  - This will at the same time reset any change that has been made in the other charts because of a highlight in the PCP chart

## Selection

- When polylines are highlighted one can press the “confirm selection” button in the toolbar to confirm the selection.
- When highlighted lines are selected first then will the PCP chart change itself (although the other charts will already have changed).
- When highlighted lines are selected all other lines (that are not selected) will disappear entirely and one will get, in essence, a whole new chart.
- If one wishes to go back to the original chart one can press the “remove selection” button that is in the big toolbar locked on top of the whole app when a selection has been made.
  - There needs to be a separate button for this when one is trying this as an extension and not in Qlik sense.
  - The button will say “Remove selection” and will be on the top left corner of the chart.

# Edit Menu

- **Data**
  - Dimensions (add)
    - Field
    - Label
    - Include null values
    - Limitations
    - Master Items (add) (delete)
  - Alternative dimension (name) (add)
    - Field
    - Label
    - Include null values
    - Limitations
    - Master Items (add) (delete)
  - Measure (add)
    - Modifier
    - Number formatting
    - Formatting type
    - Trendlines (add) (delete)
  - Alternative measure (add)
    - Add or add as an expression (field for expression)
- **Sorting**
  - Sorting of the the axes against each other and their order
  - Axis one
    - Sorting (auto) (custom)
    - If custom:
      - Sort numerically (ascending descending)
      - Sort alphabetically (ascending descending)
  - Axis two
    - Sorting (auto) (custom)
    - If custom:
      - Sort by expression (ascending descending) (field for expression)
      - Sort numerically (ascending descending)
      - Sort alphabetically (ascending descending)
- **Add-ons**
  - Data handling
    - Include zero values
    - Calculation condition (field for expression)
  - Reference lines
    - Add reference line

- **Appearance**
  - General
    - Show titles (toggle on off)
      - Title (field) (field for expression)
      - Subtitle (field) (field for expression)
      - Footnote (field) (field for expression)
    - Show details (toggle on off)
  - Presentation
    - Styling
      - Polyline thickness
      - Axis thickness
    - Horizontal or Vertical
    - Scrollbar (mini chart, bar none)
    - Scroll alignment (start , end)
    - Spacing between the axes ( toggle auto custom)
      - Spaces ( narrow, normal, far, very far)
    - Value labels (toggle on off)
  - Axes (for naming and presentation of them)
    - Menu for each and every axis
      - Label and title (labels only, title only, none, both)
      - Label orientation ( auto, horizontal, vertical)
      - Position of label (position around the axes)
      - Scale (narrow, medium large) Space between values shown
  - Colors
    - Color of all the polylines at the same time (color chart)
    - Gradient (on or off) (of all the polylines)
      - Two color charts for choosing the two colors
      - Inverse (inverse)
  - Brushing
    - Brushed lines
      - Transparency (slider)
      - Color (same color as before or grey) (toggle)
      - Completely off when brushed (toggle)

# **Appendix D**

## **Manuscript of the Usability Test**

---

In the written manuscript anything written within parentheses is a note to the testers and was never said aloud during a usability test, they were only clarifications and instructions.

## Part 1 - Introduction

During the whole usability test we would like to encourage you to use the *think aloud* method. We want you to all the time say what you are thinking so we can better gauge your spontaneous thoughts on each of the different functions that we have developed.

It is understandable that you will feel unsure at some of the questions we will ask you. This software was developed with Qlik sense users in mind. To try and narrow the gap of a first time user and an expert we have prepared two parts to both help you feel more comfortable with the test at the same time that we continue to urge you to use Think Aloud so we know what you are thinking the whole time.

## Part 2 - Exploratory Testing

We will begin with a preset of the software we want to show you and do the usability test on. The point of this part is to gain your first reactions to the software without first giving you any input. You will have as little or up to three minutes time for you to explore around the software.

## Part 3 - What are Parallel Coordinates & Qlik Language

(The second part of the “teaching” phase is teaching very briefly the basics of parallel coordinates and some of the Qlik standard language.)

(Read prepared text about parallel coordinates.)

(Read prepared text about what *Dimensions* and *Measures* are.)

## Part 4 - General Usability

(The participant is presented with a blank screen of Nebula.js with no data loaded in.)

Start by selecting a dimension, choose “names” as your dimension.

Now follow that up by choosing the first measure. Choose the sum of cylinders or CYL as your first measure.

To be able to do the rest of the test we will need some more measures.

Please choose the sum of DISP, HP, WT and GEAR as your other four measures.

## Part 5 - Highlighting & Selection

Now that we have all the data we need to carry out the test we will begin by trying out the selection function

Choose the car "Volvo 142E".

Choose another car at random, any of your choice. You may pick more than one if you want to.

Confirm your selection and that these highlighted cars are the ones you want to look closer at.

Reset the window/your selections.

(To test the lasso function.)

Highlight all the cars that have "six cylinders".

(If they pick each and everyone by hand and not use the lasso function.)

There may be a function to do this task more easily.

Unhighlight the car "Hornet 4 drive".

Confirm your selection and that these highlighted cars are the ones you want to look closer at.

## Part 6 - Filtering/Brushing & Switching Between Modes

We begin by resetting the window/your selections.

Lets now try and instead of highlighting, brush off polylines.

Brush/filer all the cars that have "four gears".

Now highlight the four cars that have the highest "DISP" value.

Confirm your selection and that these highlighted cars are the ones you want to look closer at.