UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Impact of Data Visualisation on Users in CRM Systems

Kristijan Korać
M.Sc. Thesis
August 2016

Supervisors:
dr. Arjan Van Hessen
dr. Mariët Theune
Michael Zirngibl

Human Media Interaction
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede

Contents

Abstract

Data visualisation offers lot of advantages like revealing the information hidden in the form of numbers or structured in tables etc. It makes a comprehension of large amounts of data almost instantaneous. Because of its advantages, it has also been applied to the field of business and sales, more specifically to the field of customer relationship management or CRM. People working in this domain, especially sales representatives and managers are using CRM systems on a daily basis and they are using data visualisation in order to perceive, interpret and comprehend given data effectively and efficiently. In this thesis I investigate the effects of data visualisation on users of a real-world CRM system called Leadscore. I have created new visualisations for this system and tested them in comparison to the old visualisations that existed. Testes were conducted with actual domain experts and results showed that new visualisations were perceived to be more efficient and effective when conveying specific data.

1. Introduction

Nowadays having a graphical representation of data, a visualisation, offers a lot of advantages, revealing the information hidden in the form of numbers or structured in tables etc. It provides the ability to comprehend large amounts of data instantaneously, without the need of analysing each data separately [1]. It also allows us to track down the changes over time, make comparisons, scale the data and also exploits the patterns and trends which cannot be easily seen.

Because of all the advantages, a Berlin based company called interact.io was interested in adding data visualisation to their CRM system and this motivated the topic for this thesis - Data visualisation in CRM systems.

Before moving further, it is essential to set a proper definition of this term of data visualisation. One of the greatest minds in the data visualisation field was Edward Tufte. His work is often considered the gold standard as he provided sets of visual principles that should be followed in construction of diagrams with sequencing and multi-variant data. In one of his works he defines graphical excellence as the one that consists of complex ideas communicated with clarity, precision and efficiency [2]. To build an effective, well designed graphic, Tufte indicates some principles. In essence, a graphic should [3]:

- show the data
- avoid distorting what the data has to say
- present many numbers in a small space
- make large data sets easy to understand
- encourage inferential processes, such as comparing different pieces of data
- give different perspectives on the data - from broad overview to the fine structure

The following figure is given in order to get an insight into a high-level visualisation that follows Tufte's principles and which according to him may well be the *"best statistical graphic ever drawn"* [2].

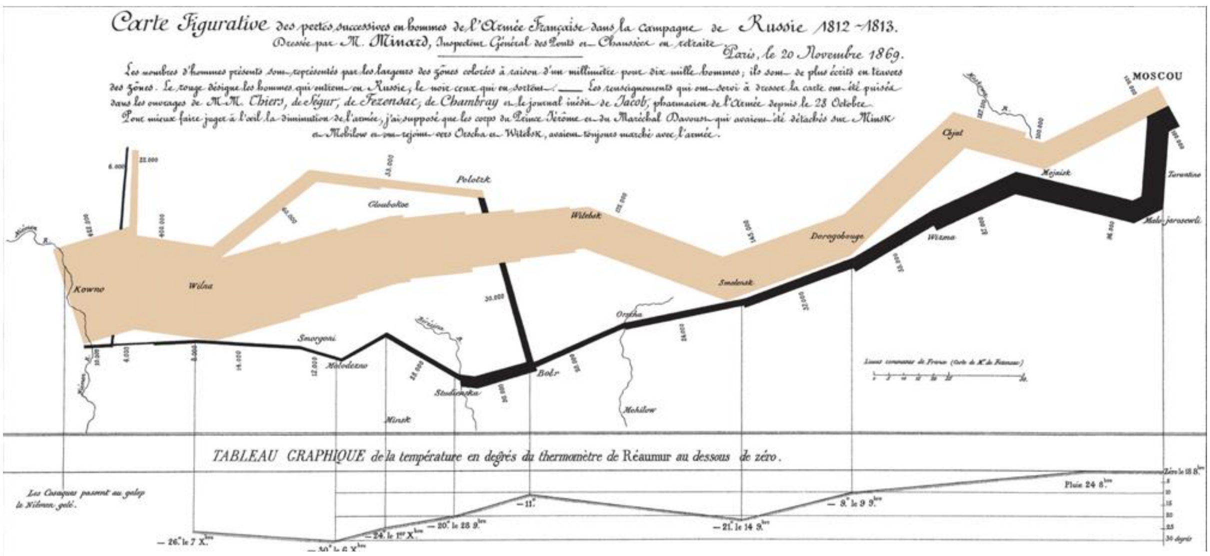

Figure 1 - Visualisation of Napoleon's march on Russia in 1812-13; visualisation shows physical progression and retreat of French army as well as its losses of men (source: Wikipedia)

Being created in 1869 by C. J. Minard, the visualisation shown in Figure 1 depicts Napoleon's march on Russia in years of 1812 and 1813. Starting from the Russian border, the French army (which counted around 400 000 soldiers) is presented as a thick line which is getting thinner over time in order to depict Napoleon's military losses. The position of the army (latitude and longitude) is also encoded in this visualisation which easily reveals the exact path that was undertaken during the march. Below the main visualisation it is possible to notice a second chart which shows the changes in the temperature over time that are directly related to Napoleon's men losses during his retreat.

The visualisation implements the basic idea of the composition technique which is an orthogonal placement of axes that encode the same information to create a 2D metric space of multi-dimensional data [3]. Variables army size, army latitude and longitude and temperature share the same horizontal axis, which is the time making the visualisation an example of single-axis composition technique. It is necessary to point out that these types of data visualisations are not a subject of the work of this thesis due to the nature of the data and target audience. The data being visualised for this thesis is strongly related to business metrics and statistics while the target audience are people working in the field of business (more specifically in the sales domain).

Having seen one example of state-of-the-art visualisation and basic principles for creating an effective and well-designed graphic, it is possible to bring forth the definition of the term data visualisation. Even though it is a relatively new discipline, different definitions can be found describing it. However, all of them are unified in the fact that data visualisation is concerned with the creation of visual artefacts in order to amplify cognition. So there goes the definition:

"Data visualisation is the use of computer-supported interactive visual representations of abstract data to amplify cognition. It is a form of external cognition, using resources in the world outside the mind to amplify what the mind can do." [3, p. 6]

This definition relies heavily on an argument that visualisation is a cognitive activity that is facilitated by external visual and graphical representations from which humans construct internal mental representation of the world. It is possible to facilitate this visualisation process with modern visualisation tools, such as modern software and libraries dedicated to visualizing data. However, it is important to notice that above definition takes a distinction from computers, pointing out that visualisation is an activity that only resides within the mind [3].

The following six ways[1] are the ones visualisation uses to amplify human cognition [3]:

- Increases the memory and processes resources available to the users
- Reduces the search for information
- Enhances the detection of patterns in data by using visual representations
- Enables perceptual inference operations
- Uses perceptual perception mechanisms for monitoring
- Encodes information in a manipulable medium

[1] Based by [3] on the work by Card et al. (1999): Card, S.K.; Mackinlay, J.D.; Shneiderman, B, Readings in information visualization: using vision to think, Morgan Kaufmann Publishers Inc., 1999, San Francisco

Since data visualisation has the ability to boost our cognition, it is becoming an important discipline in modern world that offers a vast number of diverse tools that can be used for designing and developing complex visualisations and mapping the same ones with corresponding data. As a famous saying states that *“a picture is worth a thousand words”*, a similar comparison can be drawn with the visualisation of the real-world data. Deciding upon which visualisation to use is not as trivial as it may seem considering the fact that it largely depends on what kind of message one actually wants to convey, the audience one is targeting and the type of data itself.

However, before the data is actually visualised and shown to the user, it needs to go through several basic stages. There are four basic visualisation stages [1]:

- Stage of data collection and storage
- Data pre-processing stage where data is modified and converted into a form that is easier to manipulate
- Stage of mapping the data to its corresponding visualisations - consists of computer algorithms responsible for creating the visualisations
- Stage of human perception, visual and cognitive processing

All of the four stages are also combined with corresponding feedback loops, making the process iterative to strive for the most optimal solution. All of them carry a certain weight for the visualisation process. However, the most important stage here is the final stage because it is directly responsible for communicating the data to the user. In this stage viewers (target audience) perceive a visualisation and put their cognitive effort in to understand the data.

When talking about the audience, data visualisation also found its place in the field of business. What is interesting for this thesis is an application of data visualisation in a field of customer relationship management, specifically in CRM systems since the company interact.io showed interest in implementing data visualisation in their own CRM system.

In the fast developing world, managing customers and establishing relationships with them has become one of the key activities companies need to perform in order to accomplish sustainable and healthy business growth. Since a company’s operations and profits are greatly dependent on their customers, these activities need to be carefully designed and executed [4]. However, when the business starts to grow and get more complex, and when the number of its customers begins to grow, it becomes rather difficult to manage all the customers, to retain the old ones as well as to acquire new ones. In addition, in a world where many great products exist, (personal) service and relationship are becoming increasingly important. To help nourish these activities and to deal with the pains that occur within them, the customer relationship management had to evolve into a serious process and become one of the key activities a modern company needs to perform.

2. Research question

As said before, an important thing to notice first is the audience the data visualisation is used by and consequently is influencing. The target audience in this paper are people within the field of sales, more specifically sales representatives and managers. The motivation for this thesis is to get an insight and to find out why certain visualisation performs better with sales people than others. To unravel the potential of data visualisation on the target audience, a cloud-based CRM system of a Berlin based company called interact.io will be used and its data visualised and visualisation evaluated. Different data is collected by the company's cloud platform through different micro-apps on different devices that waits to be exploited and visually utilized to bring the information to the user more effectively and efficiently. Examples of collected data are numbers of interactions of the users over time, over different type and direction, number of leads in the system (over different sales stages), leads/contacts information, etc. The current system offers completely basic and rudimentary visualisations, offering no more than a bare minimum of its data potential that is often found in form of data tables.

Considering the motivation for the thesis, the research question is the following:

How does data visualisation help salespeople (as the users of a particular CRM system) in terms of their work performance? Does the visualisation of CRM data offer users easier and quicker ways to grasp the information? Do they find this new visualisation more appealing, attractive, motivating and easier to use than the old one?

To support this research question, tests for evaluating the CRM system's data visualisation are needed. Since the thesis is focused on the users of the CRM system and influence on data visualisation on them, tests and evaluation need to involve the users familiar with such systems to give their own opinion. How users interpret, perceive and why do they prefer one visualisation over another is what this thesis is focused on.

From an academic perspective, data visualisation in CRM is still infancy compared to the commercial perspective. The reason for this may lie in the fact that data visualisation is difficult to evaluate. Its evaluation is complex and it is characterized by diverse tasks, data sets and participants [5]. This may also be a fundamental reason for having only few user studies about the topic of data visualisation evaluation out there. Since evaluation of data visualisation is still in its infancy, there are even less studies about this topic in the field of CRM. Therefore, this thesis presents interest in a scientific way because it dives into an unexplored field of data visualisation and its evaluation in CRM. Particularly, it discusses a user evaluation of data visualisation by using actual domain experts in testing.

To understand the evaluation of data visualisation in this field of CRM, one must first get acquainted with the topic of data visualisation. Chapter 3 discusses this topic, starting with the understanding the data that need to be visualised. The same chapter discusses certain topics such as human perception, Gestalt principles and visual encoding. Also, to understand the evaluation of data visualisation in CRM, one must get acquainted with the topic of customer relationship management which is discussed in the chapter 4. That chapter discusses the CRM process and CRM system, giving an overview of its constituent parts. Chapter 5 then discusses data visualisation in CRM systems, the data that can be found in CRM systems and how to choose a proper visualisation when visualising certain CRM data. It also discusses what kind of impact data visualisation can make on the users of a CRM system. Some parts of this thesis are based on earlier work. Since this thesis came from its topic proposal [6], chapters 3, 4 and 5 reference this proposal and expand its content. Chapter 6 discusses gives an overview of building the visualisations and their testing. Subsection 6.1.2 and section 6.2 also give an overview of building and testing the funnel chart visualisation which was also described in my minor thesis [7]. However, since the minor thesis discusses the influence of data visualisation on CRM's value proposition, not all the questions given in the chapter 6 correspond to the ones in the minor thesis, since they evaluate different things. Chapter 7 is reserved for results discussion and gives the limitations of this research. Chapter 8 gives the conclusion, possible improvements and future work.

3. Behind Data Visualisation

3.1. Understanding the Data

In order for us to understand the data visualisation and to utilize its power, it is essential to understand the data itself and its purpose at the first place. One must know what he wants from the available data and what use can he get from it.

Data is any (abstract) information one can collect. It can be presented in different forms like numbers, measurements, words, object properties, etc. The adjective “abstract” has been added since it describes things that are not physical.

Data is an invaluable source of knowledge and information whether it comes to exploring something new, still unfamiliar or even when trying to confirm already existing research. In this section I will describe different types of data, their characteristics and categories. It is necessary to get familiar with the term data in order to be able to dive into data visualisation.

3.1.1 Quantitative and qualitative data

There are two fundamental types of data: quantitative and qualitative.

Quantitative data is basically any data that can be expressed as a number or quantified, displayed in numerical rather than narrative form [8] [9]. Quantitative data is crucial in natural and social sciences to support quantitative research as the systematic empirical investigation of observable phenomena through different research techniques such as mathematical, statistical or computational [9]. Any information that can be expressed in a numerical way represents quantitative data. Height of the tallest mountain and depth of the deepest ocean, average car speed on the highway, number of planets in our Solar system and carbon’s atomic mass can all be measured and quantified, expressed in a form of number. By going further and expressing these numbers with statistics and percentages, one is analysing the data to extract first meaningful information valued for quantitative research which is used to create mathematical models, theories and to prove or disprove hypotheses related to certain phenomena. Outside of the academic community, quantitative data is also known as numerical data.

Qualitative data, on the other hand, cannot be expressed as a number but rather in different narrative forms and nominal scales [8]. This type of data can be observed, but it cannot be measured or quantified [9]. Person’s gender, hair and eye colour, their thoughts and opinion about certain products and their behaviour is considered as qualitative data. This type of data creates one of the essentials for dealing with qualitative research. Being heavily influenced by various sub-disciplines, qualitative research meant different things throughout the history and people have observed different situations, phenomena or events and have gathered different qualitative data in various disciplines and areas. With different theoretical learnings and methodological preferences, a goal was to identify, analyse and understand different social processes as well as patterned behaviours [8]. Outside of the academic community, this type of data is also known by the name categorical data.

The next example will be used to depict the previously mentioned data types, extracted from the same source of information. Figure 2 shows a collection of randomly positioned and scaled circles filled with different colours, giving excellent example of an information source that can contain both qualitative and quantitative data.

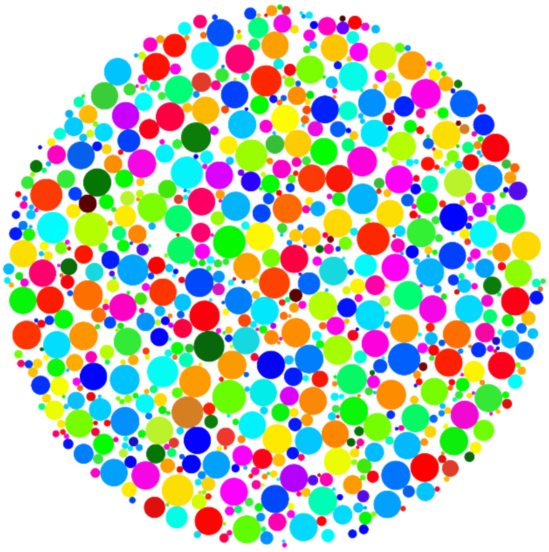

Figure 2 - A messy collection of colourful random-sized circles that presents a rich source of different types of information (source: openclipart)

As mentioned, this information source contains both types of data, waiting to be extracted.

Regarding the quantitative data, there are different features which can be expressed numerically. The first and most obvious one is the count of the circles, how many of them are present. That number is expressed as a positive integer, whole value. The second feature that can be quantified is the area size of each and every circle in the set, and it can be expressed as any positive real number.

Considering qualitative data, the first feature to be qualitatively observed is the colour of the circles. It is possible to create different categories of the circles regarding the basic colour (blue, green, red, etc.) and their variations (dark blue, light blue, etc.). Other features could be representing the size of the circles, perhaps in terms of being big, medium, small or extra small.

Just by quickly taking a look at Figure 1, it was possible to extract different types of data and their features. However, by taking a closer look on their features, it is possible to notice a difference between them.

In quantitative data on the one hand, the number of circles in the set can only take particular values, integer ones (1, 2, 294, 67, 11, etc.). Thus, this type of quantitative data is called a discrete one, since its values can only be taken from an already defined set. On the other hand, the area size of the circles can take any positive value, as a real number (2.324, 1.2, 5.324.., 2). This type of quantitative data is also known as continuous since values can take on any value from a given interval [10].

The first feature in the qualitative data was the colour of the circles. While it is possible to count how many circles contain a specific colour, it is not possible to order and rank that data [10]. This is an example of nominal data. The second feature was the size of the circles which contained several categories regarding the size of the circles. While one can count how many circles are there that belong to distinct categories, it is also possible to put these categories in order - to rank them [11]. This type is referred to as ordinal data.

3.1.2 Cross-sectional and longitudinal studies

Data is generally dissected in two ways, either through a cross-sectional or a longitudinal approach. Which one to undertake depends greatly on the nature of the research question. It is important to know what kind of information the study aims to collect in order to make a first step in determining how the study should be carried out [12]. An example of (visualised) cross-sectional and longitudinal data can be seen in Figure 3.

Both of the approaches are observational, and therefore the information is collected and recorded without manipulating and influencing the study environment.

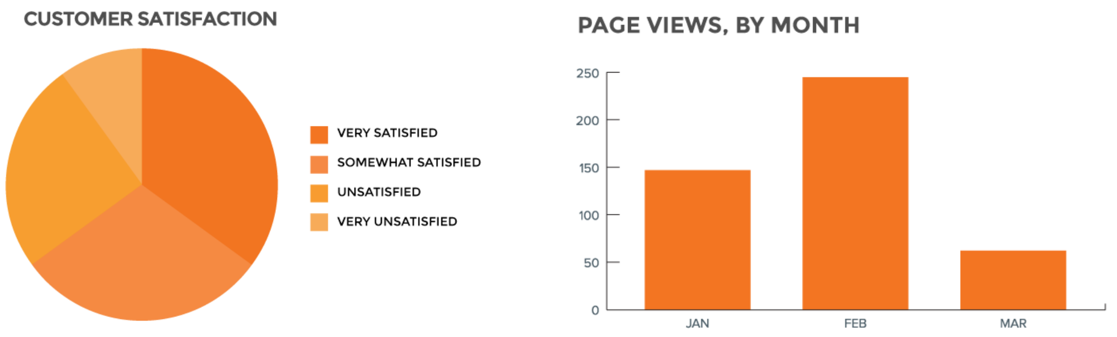

Figure 3 - A sample visualisation of cross-sectional data (left) and longitudinal (right) [9]

A cross-sectional study measures different variables only once at a single point in time. It comes in handy in practice when it is needed to compare many different variables at the same time. One can think of it as taking a snapshot of variables at a certain point of time [9]. Drawback of this approach is that it does not provide any information about the relationships between causes and the effects. Since these studies take a snapshot in a single point of time, it is not possible to discover what happens before and after the snapshot is taken [12].

A longitudinal study differs from the cross-sectional one in a way that observations are conducted repeatedly over a certain period of time, rather than just once. The great benefit of this study is being able to discover changes and development of certain subjects progressively over time. The study extends beyond a single point in time, thus enables the creation of sequences and patterns [12].

3.1.3 Data sets and relationships

A data set is a collection of related sets of information that may be accessed individually, in combination or managed as a whole entity. Simply put, it is a collection of data. A data set is comprised out of individual data points, the items that are measured or counted, commonly referred to as variables. They can take up different values and can be examined on their own or can be compared to the other variables [9].

When examining a whole set of quantitative variables of the same type, different operations and relations can be established between them, such as:

- The sum of all variables divided by their number - the mean
- The difference between the highest and the lowest ones in the set - range
- The variables' distance in relation to the mean - standard deviation
- The data distribution around the central value - distribution
- Variables with abnormal distance from the rest of the data set - outliers

These are just few out of many relations between the variables of the same data set.

Regarding the comparison of the data, one can observe different relationships between the data variables [9]. Quantitative messages are the ones revealing these relationships, telling stories that deserve one's attention [13]. These messages can be described as one or as a combination of the following:

- Nominal comparison - a simple comparison of quantitative values of categorical subdivision
- Time series - tracks change in value taken at equidistant points in time
- Ranking - shows how two or more values compare to each other ordered by size
- Part-to-Whole - measures individual subset of data compared to the larger whole. Used to show proportion or percentages
- Deviation - comparison of categorical subdivision of each variable and reference value, expressed as their difference
- Frequency distribution - counts per each categorical subdivision of quantitative interval
- Correlation - comparison of two or more variables to show whether they correlate to each other (follow approximately same pattern) or not

Entering the area of data visualisation, both qualitative and quantitative messages are the ones that inform the user, giving clear reasoning about the data presented. Thus, it is important to select the appropriate visualisation in order to express the right message. While some of the charts, like bar and line charts are useful to express different things, others, like pie charts, are limited in the messages they convey. Seeing what messages they are good conveying at will be shown in the upcoming sections, but before that we will dive into the origins and development of the data visualisation throughout the history as well as into the human perception to see how we actually perceive visual inputs.

3.2 Data visualisation throughout the ages

Since very beginning of the human race, people have expressed themselves visually. Dating back to prehistoric ages, people from every corner of the world drew paintings on the caves' walls and ceilings. They were communicating visually, whether it was drawing hunters hunting their prey, everyday life scenes, wild and sacred animals, or even by simply leaving their handprints [14]. While today it has been considered as an art, often referred as "Palaeolithic art" [15], it is not difficult to realize that it presented something more to prehistoric people, and that is a way to express themselves and to communicate, to convey a certain message.

Regarding the true origins of the data visualisation as we know it today, very little is actually known. However, according to Funkhouser, the first known example of data visualisation dates all the way back to the tenth or eleventh century. Discovered by Sigmund Gunter in 1877 as a part of the manuscript called "*De cursu per zodiacum"* owned by Bayerische Staatsbibliothek in Munich, this example contained the graphical representation of the orbits' inclination of our Solar system planets over time [16].

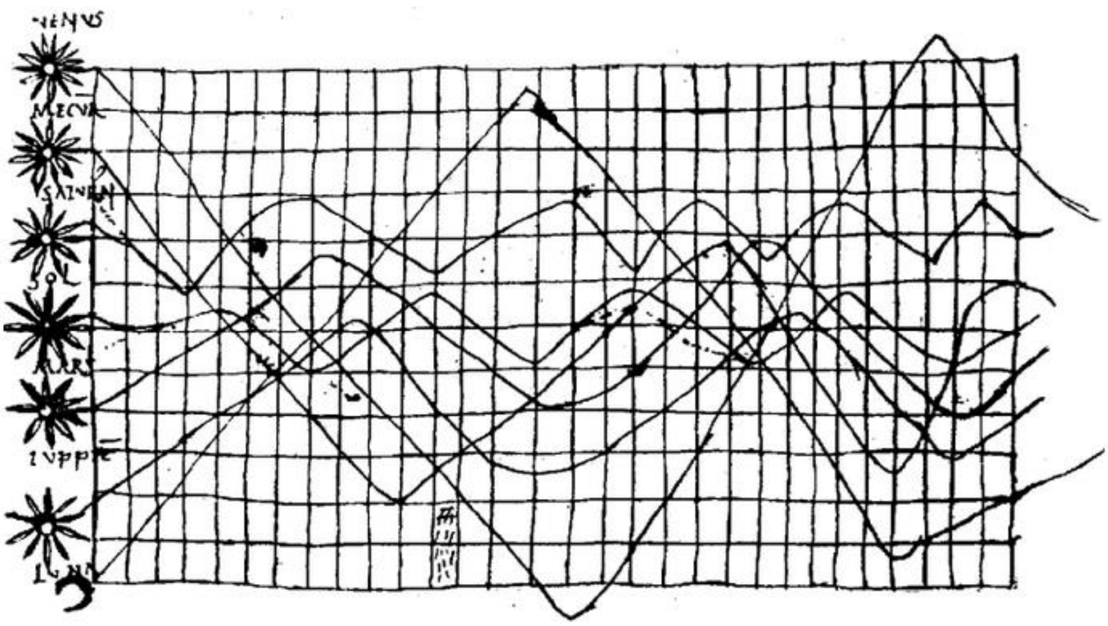

Figure 4 - First known example of data visualisation, part of manuscript De cursu per zodiacum [16]

In Figure 4 it is possible to see the celestial latitude over time for the planets like Venus, Mercury, Saturn, the Sun, Mars, Jupiter and the Moon. While it is still a mystery who created this chart, an interesting fact is that the chart was created centuries before Copernicus and Galileo introduced their heliocentric theory.

Even though people have been arranging data in a tabular presentation, the idea of graphically presented information did not occur before the 17th century. The French mathematician and philosopher Rene Descartes introduced his two-dimensional coordinate system for displaying values. It was comprised out of two axes, one horizontal and one vertical each for different variable [17]. Credited as the father of analytical geometry, he bridged algebra and geometry setting the foundations for future generations to exploit their potentials, like data visualisation among them.

Later on in the 18th century people have begun to exploit the potential of graphics for communication of quantitative data. A Scottish engineer named William Playfair was one of the key figures at that time. He pioneered most of the charts we are using today, including line, area and line chart as well as the pie chart which is used to depict part-to-whole relations [18]. In his line charts, he was the first person to show the changes in altitude of the line from left to right to demonstrate variable changes and progression over time.

He was also directly inspired in creation of the bar chart by Joseph Priestley who had invented timeline charts that used individual bars to visualise the life span of a person to compare the life spans of multiple persons. His bar chart invention was, in fact driven by the lack of the data (Scotland's trades with different countries per month) to create certain line charts. Having collected data about import and export from different countries, he used line charts to visualise that data. However, due to a lack of data related to Scotland's trades, he used a bar chart to graph its trade data for a single year, taking one bar for each of Scotland's trading partner countries [19].

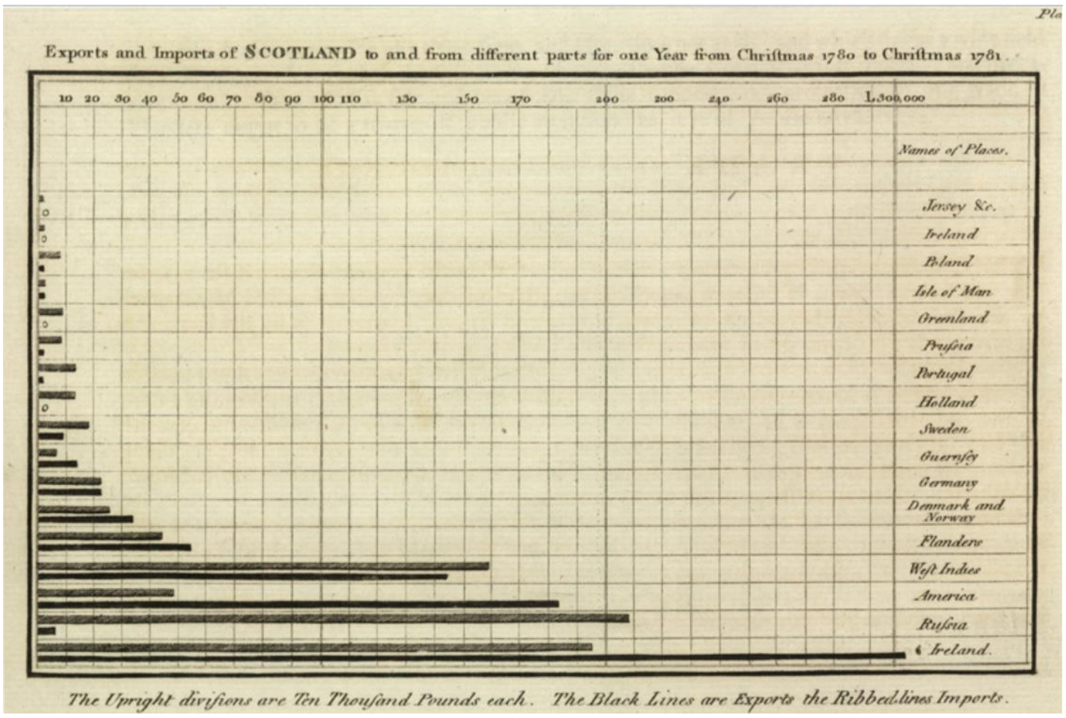

Figure 5 - Playfair's visualisation of Scotland's imports and exports over one year [19]

Playfair is also considered to be the inventor of the first pie chart, which can be found in his *Statistical Breviary*.

Until the second half of the 20th century, usability and effectiveness of graphs did not improve much. In his book, *Semiologie graphique* published in 1967, Jacques Bertin set the foundations for the progress made by the upcoming generations, discovering the rules by which our visual perception operates and mapping them to data visualisation to deliver the message hidden in the data efficiently, accurately and clearly [17]. These rules were the basis for the study by Cleveland and McGill [20](see Visual encoding 3.3.3).

Later in the century, John Tukey and Edward Tufte made a significant impact in the field of data visualisation. Tukey brought the power of data visualisation up to light by exploring and making sense of quantitative data [18]. In his book named *The Visual Display of Quantitative Information*, Tufte pointed out that the majority of people are using the data visualisation in a wrong way [17].

3.3 Human perception and importance of data visualisation

In order to understand the data visualisation and how to design the best and optimal solutions for end users, it is essential to understand how humans actually perceive the outside world, how important our visual system is and how our brain functions when our eyes are stimulated by visual inputs.

Thus, this section is focused on examining human perception, exploring its power when it comes to visual stimulation. It also focuses on examining two distinct visual representations, sensory and arbitrary. Attention is also given to examining the process of human perception and its workflow.

3.3.1 Human perception and visualisation

Visual arousal is the most sensitive and dominant arousal in human beings. Through our vision, we acquire more information than with all of other senses together. Our visual perception is handled by a visual cortex which is located in the rear part of the brain and is extremely efficient and fast, enabling us to see things immediately, without much effort [17]. That visual cortex contains around 20 billion neurons working on analysing this visual information [1]. Our cognitive ability is largely influenced by our ability to visually perceive the world, and it can be improved by optimizing the search for important information and patterns. This way we would have to use less effort in order to understand what we are observing and therefore improve our decision-making process enabling us to find and execute right decisions effectively and efficiently.

As seen from the Introduction, there are four basic visualisation stages (see Figure 6). Even though the first stage is the longest one, where one must seek and gather all of the relevant data, the second stage also carries great importance where one must analyse the data, distinguish the relevant and useful data from the rest and transform it into something that can be easily compiled or processed by computer algorithm [1]. The third stage of mapping the data to a corresponding visualisation is the one that is making direct impact and influences the final stage of human perception and cognition.

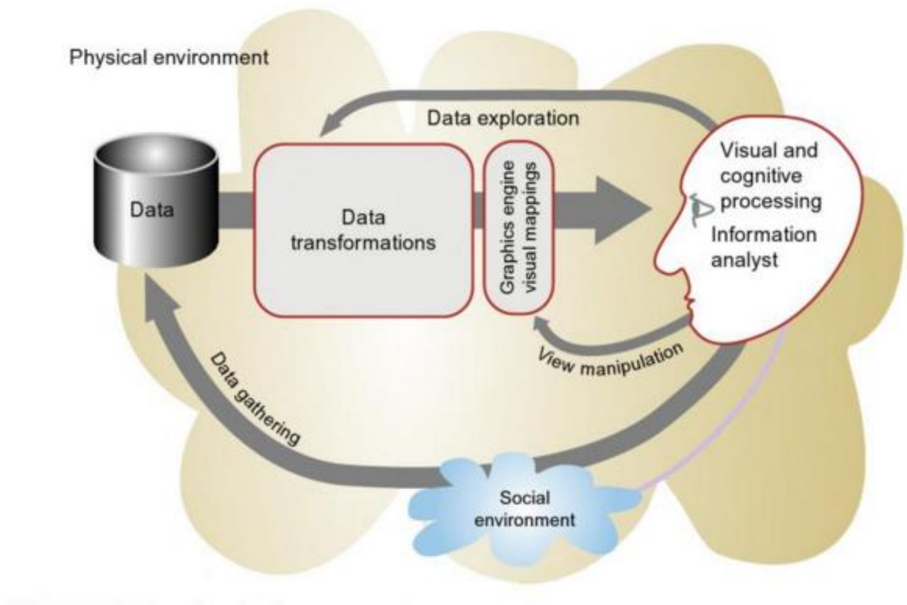

Figure 6 - Four stages of the data visualisation process [1]

The third step in the visualisation process is especially interesting for this thesis since it is the one that contains the desired visualisation and therefore influences user's perception and cognitive thinking.

Some scholars have argued that visualisation cannot be understood as a science at all but rather as a learned language [1]. Relating the visualisation to diagrams and diagrams to symbols which are based on social conventions and interactions, they have put visualisation in the same group a language which needs to be learned in order to derive sense out of it and in the end, perceive it accurately.

Believing that all representations have a certain value, famous Swiss linguist and semiotician Ferdinand de Saussure laid the foundation for thinking of data visualisation as something that has to be taught. According to him, truth is relative to its social context and it is only meaningful to those who understand it. However, Ware reports in his book [1] that various studies and experiments like the ones conducted by Deregowski (1968)[2] and Hochberg and Brooks (1962)[3] proved this thinking wrong. They have proved that basic understanding and interpretation of pictures is not a learned skill and can be done without training [1].

[2] Deregowski, J.B. (1968). Pictorial recognition in subjects from a relatively pictureless environment. *African Social Research,* 5, 356-364.
[3] Hochberg, J. & Brooks, V. (1962). Pictorial recognition as an unlearned ability: a study of one child's performance. *American Journal of Psychology,* 75, 624-628.

Considering human perception and cognition, there are two distinct aspects of visualisation, sensory and arbitrary one. Sensory aspects of visualisation derive their expressive power from being well designed to stimulate the visual sensory system and use the brain's perceptual power without training. Arbitrary visualisation refers to the aspects that derive their power from how well they are learned. In order for them to represent a useful information, they need to have a perceptual basis [1]. Examples of each can be seen in Figure 7 with the left figure being more sensory since it perceives the relation instantly while the right one requires the viewer to "learn" about the relation and then fully perceive it.

Sensory representations and symbols can be understood without training and are resistant to instructional bias. Our brain processes them very quickly and parallel and they are valid for almost every human being and stable across individuals, cultures and time. However, since it the brain so fast, poor data mapping can also lead to misunderstanding and wrong interpretation.

Arbitrary representation and symbols are formally powerful, deriving their power from how well they are learned. Due to the influences of the outside world, they may already be learned (as for example having a same word with the same meaning in two different languages). Drawback of these presentations is that it takes a lot of time to learn them. They vary across different cultures and applications and can be difficult to learn while at the same time they can be easily forgotten.

Figure 7 - Two different arbitrary representations with left one being more "sensory" since it enables the viewer to perceive the relationships instantly [1]

Even though these representations differ radically, they are still used in combination to each other in data visualisation.

In order to understand the process of the human visual perception, a simplified model of perceptual processing can be introduced. This model contains three stages of perceptual processing. The first stage is responsible for extracting basic features from the environment and processing that information in parallel. Second stage is the one segmenting the visual scene into specific regions and making a distinction between colour, texture and patterns in motion. The third stage, also the highest level of perception, uses active mechanisms of attention to reduce information only to the one contained in the visual working memory [1]. Through these stages, the basis of visual thinking is formed. These stages can be seen in a simplified model in Figure 8.

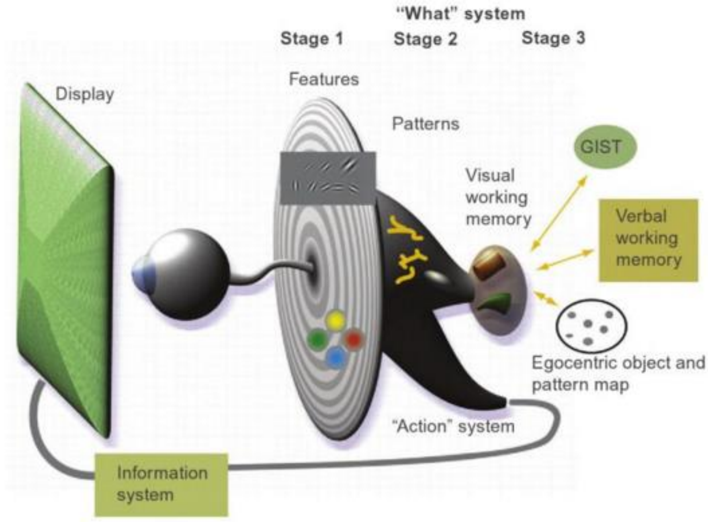

Figure 8 - Simplified three-stage model of processing visual information [1]

3.3.2 Gestalt ideas and principles

The Gestalt School of Psychology, founded at the University of Berlin in the early 20th century, was the most noticeable one in the area of cognitive psychology making the earliest contributions to the science of perception. Studying our abilities of acquiring and maintaining meaningful perception, Gestalt psychology is focused on understanding psychological phenomena by observing them as a whole entity rather than the sum of their parts [21].

“The whole is other than the sum of its part”, is a famous quote by Gestalt psychologist Kurt Koffka, pointing out that a whole entity has an independent existence in perceptual system compared with the mere sum of its constituent parts [22].

Gestalt psychologists tried to reveal how we actually perceive organisation, form and pattern in our observations [17]. With their research, they have noticed that we as humans organize the information we perceive in a certain ways in order to make sense of it. This observation yielded a number of Gestalt ideas and principles that are still today considered as accurate descriptions of our visual behaviour.

The key ideas of Gestalt School are emergence, reification, multistability and invariance. All of them can be easier to understand with an appropriate picture that supports corresponding description.

The idea of *emergence* is one of those that support Gestalt theory of observing the whole entity. It is a process of forming complex patterns from simple rules [23]. Emergence is instant, it appears out of nowhere, almost like magic. The most famous example of this principle is the picture of a Dalmatian dog sniffing the ground in the shade of the trees. The dog cannot be perceived by just observing his individual body parts, but rather is perceived as whole at once. This example can be seen in Figure 9.

Figure 9 - Dalmatian dog and emergence – the dog cannot be seen while observing its individual parts but is rather perceived immediately when observed a as a whole entity (source: ChangingMinds)

Idea of *reification* is extremely powerful since it uses the constructive aspect of our perception in order to convert an abstract object into something concrete, more meaningful. This principle is the proof of our perception being constructive for it constructs a whole mental representation from less explicit sensory information. Basically, our mind fills the gaps [23]. Figure 10 presents simple example of building a concrete image out of something that “does not exist”.

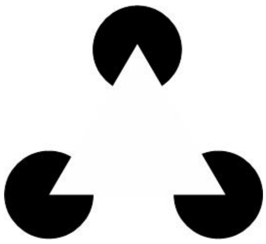

Figure 10 - Reification - our mind constructs the image of triangle even though it does not technically exist (source: Wikipedia)

Multistability enables us to interpret different images out of the same object while switching our focus on the other parts of the object, while the principle of in variance is describing our perceptual ability to recognize the same objects independent of their scaling, translation, rotation, lightning, etc. [23]. Their examples can be seen in the Figure 11.

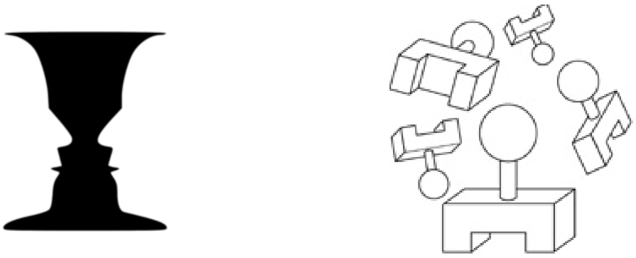

Figure 11 - Multistability (left) and invariance (right); Multistability describes how we actually perceive different images from the same object when shifting our focus on different parts; Invariance describes our ability to recognize the same object independent of its scale, translation, rotation, lightning, etc. (source: Wikipedia)

These fundamental ideas brought a new point of view in cognitive and perceptual psychology. Even though they present the foundation of Gestalt psychology, their application cannot be directly addressed in the topic of data visualisation. However, using these ideas as a background, Gestalt psychologists created their principles which can find direct application in data visualisation.

One of the fundamental principles is the Law of Prägnanz (also known as Good Figure or Law of Simplicity). It suggests that "people will perceive and interpret ambiguous or complex images as the simplest form(s) possible". When confronted with a complex form or shape, we tend to reorganize them into simplest forms that are easier to perceive and take less time to process.

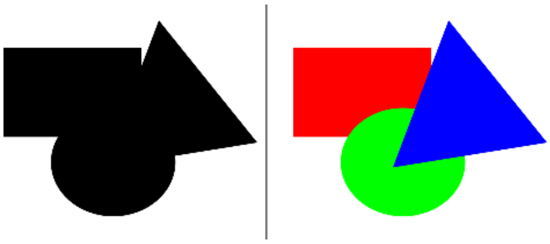

Figure 12 - Law of Prägnanz, Good Figure suggests that we perceive complex image by breaking it down into simplest form(s) possible [23]

Figure 12 shows an example where it is easier to perceive distinct objects rather than the whole object itself. However, that is not always the case. The next principle shows us when it is easier to perceive a complex form rather than the simplified, distinct objects.

Closure is the principle opposite of Good Figure, in which we are combining parts to form a simpler whole [23]. In that way, we are perceiving the objects such as shapes, letters and pictures as being whole even when they are not. When parts of an image are missing, our brain fills in the visual gaps to perceive the whole entity. In Figure 13, even though geometrical shapes are not completed, our brain fills in the gaps enabling us to perceive them as whole objects.

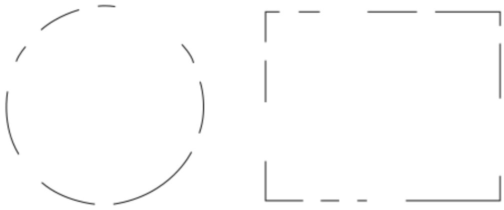

Figure 13 - Principle of closure – Even though a circle and a rectangle are not drawn on the picture (just groups of organized lines), our mind aids us and constructs these images (source: [Wikipedia](#))

While it is a human tendency to seek for meaningful patterns, closure can be thought of as the glue that is holding elements together [23]. For closure to occur, enough information needs to be provided to the eye so our brain can fill the rest. Not having enough information makes us interpret parts as separate instead of as a whole entity.

Another interesting principle is the one about *proximity*. This principle suggests that objects that are in near vicinity to each other tend to be perceived as a group whether they are in a relationship or not [24]. Figure 14 shows an example of the proximity principle. Circular shapes that are positioned close to the others are perceived as a group rather than individual objects.

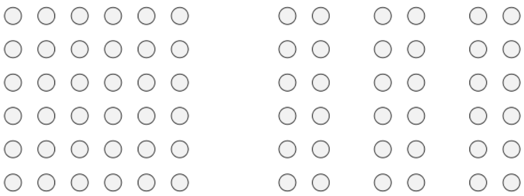

Figure 14 – The principle of proximity describes how people tend to perceive objects in near vicinity as a group (source: [Wikipedia](#))

When it comes to grouping, other principles are also important. Among them are principles of similarity, continuity and symmetry. They all tend to group individual objects into meaningful groups, while easily making a distinction among the groups themselves.

The principle of *similarity* suggests that elements sharing similar characteristics, such as shape, colour and size, tend to be perceived as more related than elements that do not share those characteristics [23]. These elements are perceived as a group regardless of whether this relationship exists or not.

The principle of *continuity* however, states that elements that are connected by lines are seen in way that follows the smoothest path [24]. This visually occurs when we observe and follow a specific path, we choose the one that progresses smoothly. It is in our instinct to follow a certain path or a line, just as much as we follow up a river path.

Symmetry is the principle that gives us a feeling of order, in perceptual pleasure. We tend to seek order over chaos. Therefore, this principle allows us to perceive symmetrical elements as the ones grouped around a centre point. When observing two symmetrical shapes that are not connected, our mind perceptually binds them into the same group [23].

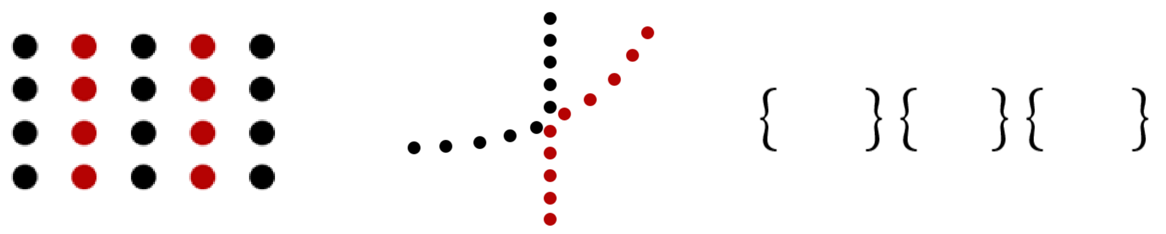

Figure 15 - From left to right, principles of similarity, continuity and symmetry [23]

Figure 15 gives some examples that make it easier to understand these principles. Thanks to similarity, the left image people usually perceive as consisting of five different columns, grouped by colour rather than perceiving it in rows. The middle image is mostly perceived as being comprised out of two intersecting lines, the vertical one and the curved one. Even though there is also grouping by similarity in this image, this example demonstrates the dominance of the continuity principle over the similarity principle. The image on the right is usually perceived as consisting of three pairs of curly brackets rather than consisting of each bracket separately.

Summarizing, Gestalt psychology introduced new principles that shed the light on distinction between our perception and our interpretation. While our eyes are responsible for seeing the visual cue and perceiving it, its real interpretation occurs inside the brain while it builds up the missing parts, creates distinctions and groups the given elements, all in the favour of reducing brain's cognitive load. Reducing the cognitive load while interpreting certain information and especially data is one of the key assets of data visualisation. However, one may ask how to actually make a relation between these visual cues, or elements, and the data that needs to be interpreted. Next section introduces the mapping between the display elements, which we perceive in the visualisation and interpret, and the actual data which we want to visualise and exploit.

3.3.3 Visual encoding

When creating a new memory, encoding is the critical first step. It enables the perceived item to be stored within the brain's memory by converting it into a construct, which can be later recalled from short or long-term memory [25]. Encoding is a biological event or a process that begins with our perception through our senses. After perceiving an external stimulus, the process of memorizing it begins with an attention since it causes our neurons to fire more frequently. That increases the likelihood of encoding that event as a memory since it makes the experience more intense. Emotion is the experience that tends to enhance the attention and this emotional part of the event is processed through an unconscious pathway to a part of the brain called amygdala. This part of the brain, located in within the medial temporal lobe, has a primary role in processing emotional reactions and plays an important role in visual encoding.

There are three to four types of memory encoding. However, due to the scope of this thesis, only the visual one is of interest. Visual encoding is the process of encoding visual sensory information [25]. Before being encoded into long-term memory, this information is stored within the iconic memory. The amygdala's part in this encoding is accepting the visual input in addition to other inputs from other senses and encoding the values as positive or negative ones of conditioned stimuli.

Simply put, visual encoding is the process by which we remember visual images. This definition however, is given from a biological and cognitive perspective since it maps visually perceived information with constructs that are stored in the memory. Mapping the information can be also be observed from a design point of view in which visual encoding actually refers to the way how data is mapped into visual structures which are used to build the visualisations [26]. In order to be able to craft meaningful data visualisations, this type of visual encoding needs to be understood.

Two types of variables are supported in visual encoding: planar and retinal. Planar variables are those which locate points in space, whereas retinal variables are used for some other properties (like size, colour, shape, etc.). To understand how these variables are plotted in corresponding visualisation, one must first recall what types of data are there (see 3.1.1). To sum it here, there are two basic types of data: quantitative and qualitative data. Quantitative data can be measured, expressed in numbers. Qualitative data cannot be measured but is expressed in narrative forms and nominal scales. Two basic subcategories of qualitative data are nominal and ordinal. Nominal variables have two or more categories, but do not have any order that sorts them. Ordinal variables however, have also two or more categories, but these categories can be compared and ordered.

Planar and retinal variables

Now that we have recalled which types of data are there, to visually encode them we need to use planar and retinal variables. The position of a dot in coordinate system is considered a planar variable since it locates a point in space. It is probably one of the most prominent visual encodings that occur in data visualisation and also a display element that can be perceived with great accuracy. Therefore, planar variables thrive in presenting any quantitative data. Figure 16 gives an example of position along x and y axis as being a planar variable.

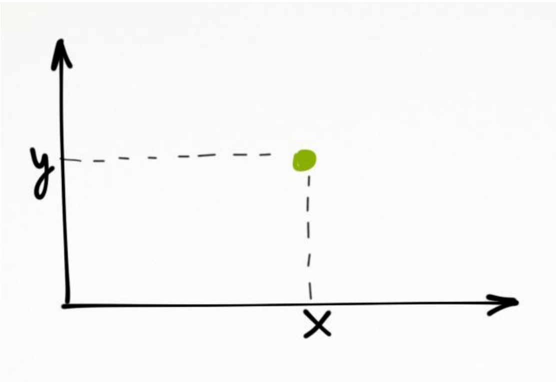

Figure 16 - Position as a planar variable - locates a point in space and allows a perception with great accuracy [26]

When it comes to visualising higher dimensions of data, plotting more than two variables in space, for example three variables, requires a usage of a third dimension of the space. However, 3D models are generally poorly perceived by our eyes since they make quantitative comparison between the points way more difficult. That is the case where retinal variables come to the rescue since they enable encoding additional variables for the data set. First retinal variable that comes in handy is the size (of a point for example). It enables the user to spot the difference right away and is a good visualizer for quantitative data [26]. Size as a retinal variable is also particularly effective for ordered data since it allows different quantitative values to be classified into distinct categories which can be ordered by their size. Other retinal variables which are proven effective for order data are orientation and colour saturation which can be used if there is a need for encoding more data variables. However, their usage may be tricky. While there is no problem in distinguishing vertical and horizontal lines for orientation, it is hard to perceive quantitative differences between different orientations [26]. For colour saturation, even though it is helpful to visualise the ordered data, perceived value difference between vicinal colour categories (i.e. light blue and slightly lighter blue) is not so obvious. Figure 17 shows these three retinal variables.

Figure 17 - Size, orientation and colour saturation as retinal variables - particularly effective for displaying ordered data [26]

On the other hand, retinal variables that are particularly useful for encoding nominal data are colour hue, shape and texture. Colour is great for displaying separate categories and often a number one choice. It is followed by a shape which also enables users to easily distinguish categories. Texture is however, less common in practice since it is usually less catchy than previous two variables. Figure 18 gives an example of visual encoding for these types of retinal variables.

Figure 18 - Colour hue, shape and texture as retinal variables - particularly effective for displaying nominal data [26]

So far, different types of visual encoding variables have been introduced to match and encode graphically different data types. However, it is obvious that differences between their effectiveness exist because of their nature and perception of our eyes. How to know which encoding variables to use when presenting, for example quantitative data? In their 1985 paper, William Cleveland and Robert McGill compared the effectiveness of different display elements as visual encoding variables in terms of users being able to interpret exact number values from the graphs (quantitative information). The ranking of these encodings is as follows [20]:

1. Position along a common scale
2. Position on an identical but nonaligned scale
3. Length
4. Angle, slope
5. Area
6. Volume, density, colour saturation
7. Colour hue

Position along a common scale has been proven to be the most accurate of the encodings while colour hue, saturation, density and volume were the least accurate ones. When it comes to creating a visualisation and mapping data variables to visual encodings, it is crucial that one understands the rankings of visual encodings and thus chooses appropriate encodings for the design. It is essential that most important variables in given data set are mapped to the appropriate encodings that are ranked higher since they are more accurate in conveying the right information.

To sum up this sub-section, when creating a data visualisation, there are two types of encoding variables that need to be considered: planar and retinal variables. Planar variables are used to locate a point in space (such as a position), while retinal are used for different, descriptive type of encoding (such as size or colour). For displaying data with different types of data variables (quantitative, qualitative), encoding variables have a different weight in terms of successfully conveying (visualising) the message that data is telling. In order to design an effective visualisation, encoding variables and their strengths need to be taken into account.

4. Introduction to Customer Relationship Management

Customer relationship management is the process of managing a company's interactions with its customers in order to get a greater insight into their needs [27]. By collecting and analysing the data about the interactions a customer has with a company, businesses are able to draw certain conclusions about the customer, make forecasts and ultimately make a response by making a change in their offering of a product or a service.

However, this process is not always fully understood and can sometimes lead to customer dissatisfaction and even loss. Results of a research released in May 2003 by an American research and advisory company called Gartner show that 41.9 percent of total bought software licenses were not deployed and about 70 percent of customer relationship management projects resulted in loss [28] [29]**.** It is not enough to install certain software just to track the customer related information. Businesses need to be aware that understanding the process thoroughly is crucial for effective management of the relationships with their customers.

4.1 Understanding the CRM Process

Understanding how to effectively manage relationships with customers has become a very important topic in modern businesses and the academic world. Businesses have started to move away from a product-based approach to a customer-based one, communicating different offerings to different groups of customers and adapting their strategies accordingly since they have realized that different customers represent different economic values for the company [28].

While the previous definition of customer relationship management, as a process of company's interactions with its customers, is not wrong, it does not actually contain more insightful information regarding the process itself. The definition of CRM greatly depends on the level it is practiced within the company or the organization. The most interesting level, for the companies, is the customer-facing level which emphasizes the importance of coordinating customer information over time in order to successfully manage customer relationship. That means building a single-view of a customer across time and all contact channels.

When it comes to conceptualizing the CRM process, there are four distinct factors that need to be considered [28]:

- The essence of marketing concept is delivered by building and managing ongoing customer relationships
- Customer relationships evolve with different phases (customer acquisition and initiation, maintenance and retention and finally termination)
- At each stage, companies are managing relationships and interacting with their customers
- The value of the relationship is not distributed homogeneously

While the first factor contains the essence of customer relationships management and its main tasks, it also stresses that building the right types of relationships is much more important and better than just building more and more relationships [28].

The second factor suggests that relationships are not comprised out of isolated and independent transactions but they rather evolve throughout different phases [28]. These transactions are interdependent and thus are creating unique relationship dynamic, which defines a CRM process as longitudinal one.

The third factor points out that it is crucial to recognize the relationships' evolution over phases since it makes direct impact on the company's success rate. Managing relationships and interacting with customers should be different at each stage, which is one of the goals of customer relationship management [28]. In these stages, relationships should be managed proactively and systematically.

The final factor suggests that the distribution of relationship value to the company is not homogenous. Nowadays, companies have adopted the CRM approach and are able to make profitability statements along customer relationship process. It is not so rare that the best customers receive way less attention than they should, while marginal customers are given way more attention than they deserve [28]. This factor indicates that companies should allocate their resources in a better manner and to define different resource allocations for different groups of customers.

These factors present great importance in customer relationship management. At every stage of the process (customer acquisition and initiation, maintenance and retention and finally termination), CRM activities should be continuously balanced in order to nourish the relationships with a customer and consequently maximize the value coming from them. CRM process entails proactive management of relationships from their beginning till the end, with executing different activities throughout different stages [28]. In addition, these activities result with feedback information about the customer, which is further analysed in order to create better customer image.

Following previous sections and definitions, the CRM process can be defined as a systematic process to manage customer relationship initiation, retention and termination across all customer contact points in order to maximize the value of the relationship portfolio [28].

A company's task therefore becomes also to lead the customers through and manage these three customer lifecycle stages and by applying CRM process on the way, it brings up the company's performance in two of three stages, retention and initiation [28].

Technology becomes important when performing these tasks, since it could help company's employees in performing their activities on a daily basis in order to achieve their goals. Since businesses started to apply customer relationship management more and more out of economic reasons, there was a need for creating software that can aid and provide support to their users in order to manage their customer relationship effectively and efficiently, throughout the whole customer lifecycle. Therefore, CRM systems were introduced.

4.2 CRM Systems

A customer Relationship Management (CRM) System is a software system that is used for performing the tasks of managing a company's relationships with its customers. In other words, a CRM System is a software that gathers, stores, analyses and consolidates customer information in order for business users to access and manage it easily. That type of system enables tracking of every interaction with current and future users that was established through different channels, including telephone, email, company's website, social media, etc. That information is stored in the system database so it can be easily retrieved and used. Using a CRM approach and System, companies analyse gathered information in order to improve their business relationships with customers by performing necessary activities and offering certain products or service at specified points of time in the customer lifecycle.

There are different types of CRM Systems out there, each one offering some additional specific feature or service. However, there are some common features most CRM Systems include like marketing automation, sales force automation, contact centre automation and location-based services [30].

With marketing automation, a sales representative as a user of a system can have it automatically send necessary marketing materials to sales prospects that have entered the system. Sales force automation enables automation of certain business tasks such as sales processing, tracking of customer interactions as well as sales forecast analysis [31]. Automation of the contact centre provides a simplified customer service process with pre-recorded audio that assists customers in solving their problems, while location-based services are used to create specified geographic customer campaigns based on customers' physical locations [30].

Since they have become one of the key companies' resources and assets, sales representatives have started using CRM Systems on a daily basis. Therefore, the system started to play significant role in a business since it makes a direct impact on the salespeople, users who are using it constantly for their work. In order to get a better understanding of a CRM System and how it works, the next section will introduce some of its key parts and their roles, making a reference to the system which the writer of this paper has been working on, called Leadscore.

4.2.1 Constituent parts of a CRM System

Every CRM system has some specific and unique parts and features. However, the most important and key ones are what probably all CRM systems have in common. All of these parts are responsible for different tasks in order to drive the sales and revenue growth. Acting individually, they are built specific for their task whether it is showing the user's current activity, predicting the potential future outcome or notifying the user of the hottest leads. It is crucial that necessary information and data is shared among them to provide real time and synchronized insights as well as for keeping the consistency throughout the entire system.

Before diving into the constituent parts of the system, its physical parts, it is important to notice the key tasks a CRM system should perform. By providing insightful analytics, response tracking and flexible segmentation tools to marketing professionals, the CRM system should improve marketing effectiveness. Consequently, this should result in an increased pipeline filled with qualified leads that can be prioritized and forwarded to the right sales teams and representatives. Managing and tracking marketing campaigns from lead to close is useful since it gives an insight in which marketing campaign results in the highest number of sales [32].

A CRM system should provide the ability to record and build a complete customer history. This way it enables sales representatives to easily retrieve necessary customer information and history. However, it should also provide the ability of forecasting and planning customer activities such as contract renewals and selling upgrades. This information should be available to representatives and managers to gain visibility across deals and prioritize their selling time, giving them the ability to close right deals faster and increase their deal size [32].

A CRM system's task should not be finished when the deal is closed, but rather provide helpful tools and necessary information in order to assist teams in customer service and technical support. Arming a service with customer history and knowledge bases necessary to effectively and efficiently resolve customer issues will result in customer thrill and delight [32].

The next sub-section discusses some of the basic physical components of a CRM system, in terms of software implementation. As stated before, each of them is given a specific task, while their combination creates a complete solution known as a CRM system. It is important to mention that some components may contain different information regarding the user of the system, whether it is a sales representative or sales manager.

Email and Call integration

Automating emails and integrating calls were one of the first features of a CRM system. Since its task is to track down and store all the information about the interactions with the customers, most CRM systems today have integrated email functionality, mimicking existing desktop email services and applications in offering full email support. Therefore, it is easy for the system to track the email interactions and store them into the customer database. Becoming arguably the most important way of business interaction nowadays, it is crucial that this functionality is implemented in today's systems. Another advantage of email integration is emails automation. This presents great value for sales representatives in terms of establishing first contacts with their leads in order to assess their potential as a buyer. That assessment is done by gathering necessary information about the lead, and when he is considered as a buyer, he is referred to as a qualified lead. Hence, email automation can be considered as a first step in leads qualification. An example of email automation can be found in a case when new leads are leaving their personal information on company's website such as in the case of form fill-in. Afterwards, CRM system automatically sends emails to those leads with corresponding messages and offers based on leads' preferences and provided information from the form.

As stated before, call integration is also essential in CRM systems. Calling the leads is also one of the main ways of business interaction today, and having a system that supports this activity presents significant value to the sales representatives. They can easily establish a call with their leads directly from the system itself, while in the background it tracks the information such as a length of a call, current date, etc. Figure 19 shows an example of call integration.

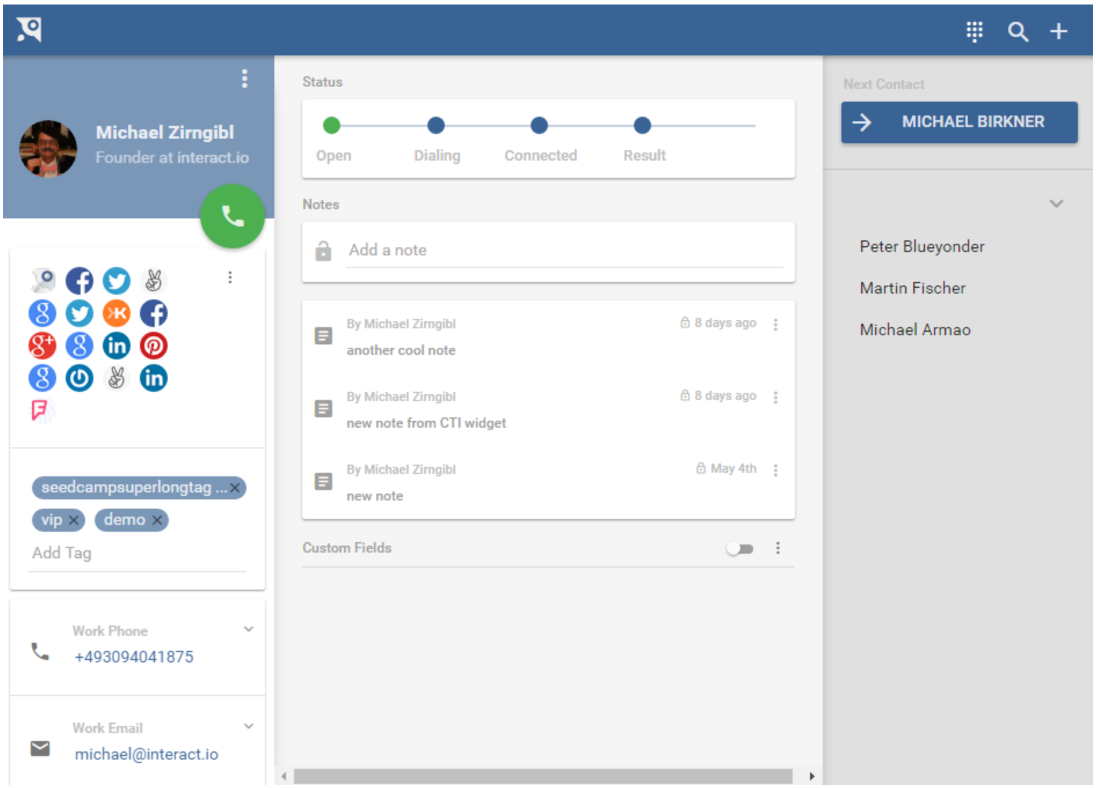

Figure 19 - Example of a CRM's call integrated feature - a call is initiated by pressing the green button below the contact information, the status of the call (i.e. dialling, rejected, etc.) is tracked in the steps and right panel shows the next lead to contact along with the others (source: Leadscore CRM)

The calling activity becomes even easier today with the development of cloud-based CRM systems and their deployment to the modern mobile platforms. Having an app installed on a smart-phone, the user can easily make a call no matter of his physical location at any time while the system again tracks all the necessary information. After a call has finished, a good CRM system should rank other leads that need to be contacted and bring forth the one with the highest rank as a suggestion for making the next call activity.

Contacts page

Every CRM system needs to contain the information and relative data about the potential leads and customer contacts. That stored information needs to be easily retrieved for the users of the system while they try to perform their daily tasks of interacting with customers and managing customer relationships.

Having the contacts page, the user can easily access any contact relative information in order to get an insight and complete view of their customer. This information is useful in terms of decision making for future interactions, for example which stage the customer is at, when and what was the last interaction as well as when to make a new one with the highest chance of converting the customer to the next stage of the business pipeline (see Sales Pipeline).

The contacts page is usually presented as a table or a list of customer contacts. This list can be constantly updated depending on whether new customer contacts are entering the system. One important feature this page needs to carry is the possibility of filtering the list of contacts according to their entry date, stage in a pipeline and date of the last interaction. Ranking the contacts according to their names, entry date and date of the last interaction also carries a value in terms of easier contact information retrieval. Figure 20 presents one example of a CRM system's contacts page.

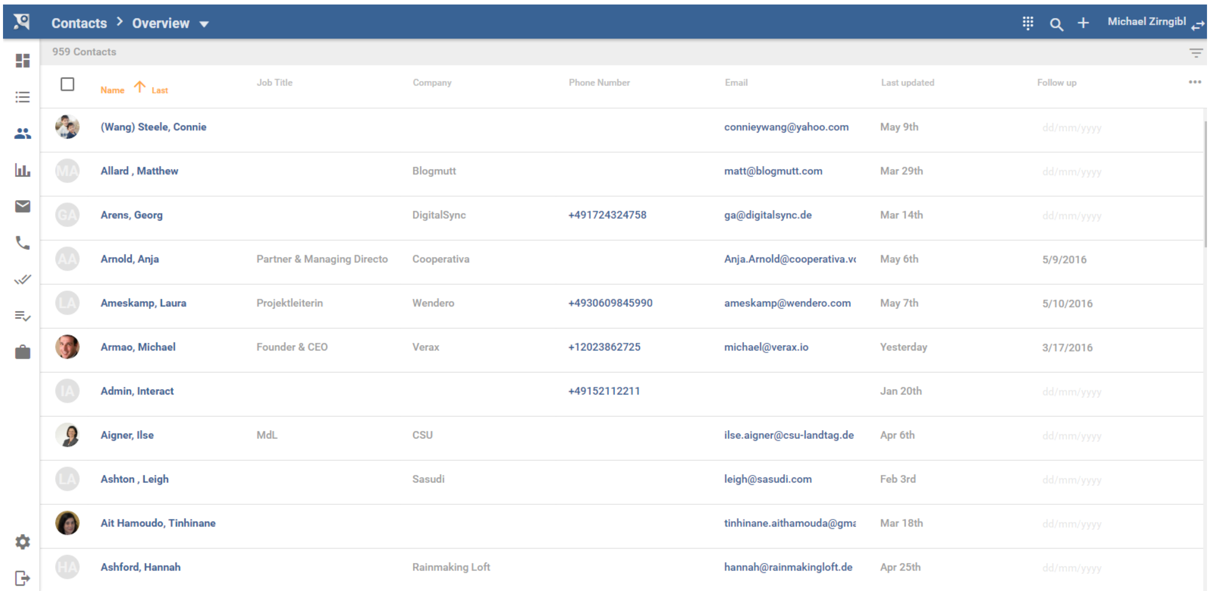

Figure 20 - Example of a contacts page with additional features such as easy search, contact type filter, different ranking options and possibility of adding a new contact (source: Leadscore)

Dashboards

When mentioning a CRM system, a lot of people with basic knowledge about it will think of the dashboards. Providing a high level, composite overview of sales and activities, dashboards are one of the key components of a system, since they drive the user's activities and show a quick overview of data related to a particular job. Showing current tasks and work that sales representatives need to do, dashboards are focused on the future, on the things that need to be done in order to satisfy the currently set goal. They are telling a story about the data in the system, such as the number of current leads, new leads, new incoming messages, etc. That story goes beyond the mere numbers and expands to the suggestion of which leads should be contacted first ("hottest leads"), when is the best time of a day to make a contact with the leads of a specific stage in the business pipeline, etc. Depending on a type of a business, the dashboard also can present the information of the overall growth as well as the revenue increase or decrease.

A sales representative is a company's employee that works in sales (calls potential customers, arranges a meetings, sales a service or a product). A sales manager is an employee with a higher rank within the company who manages and supervises a team of sales representatives. While a sales representative as a user of the system will get the information about his performance, leads and contacts, interactions and tasks, a sales manager will get the information about his/her overall team performance. Typically managers also have access into the performance of every individual team member. This gives them the ability to discover the best ones in their teams which could be awarded, and also to detect who has the poorest performance in order to help them catch up with the rest of the team. Figure 21 shows an example of a dashboard (this is not an example of a real CRM but rather a template that can be used for building a CRM dashboard).

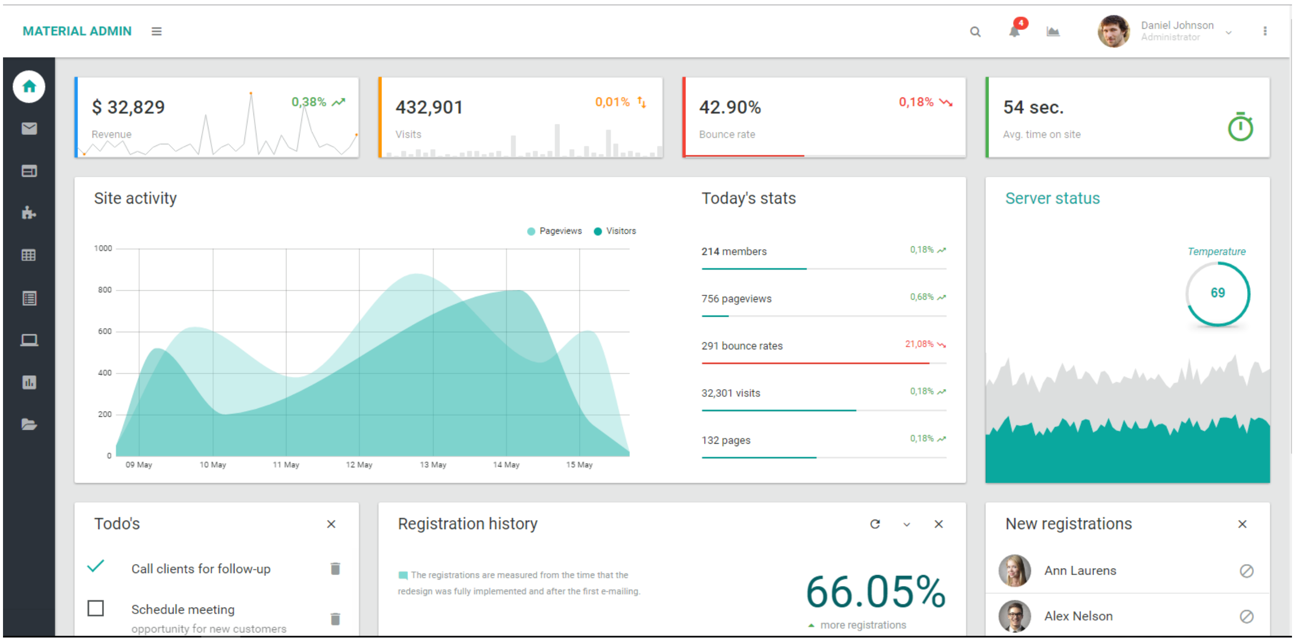

Figure 21 - Example of a dashboard showing current revenue growth (upper left), activities over time (left chart), current tasks (bottom left) and new contact registrations (bottom right) (source: Material Admin)

Reporting module

While the dashboard is also comprised of charts that present information dependent on a time variable, that information is mostly focused on the future because it motivates and drives the user forward. The reporting module is however, focused on the past events and interactions, presenting the data at different time intervals in the past. It shows how well the sales representative performed, how many interactions has he/she carried out and so on.

This information presents an enormous value for both types of the system users, sales representatives and managers. To sales representatives, reports present a value in terms of building a complete view of their customer(s). By looking at the reports, sales representatives can get an insight in their sales history, customer's behaviour and their purchasing habits. Another important value the reporting section gives them is the ability to track their activities over time, make comparison with past interactions and get an insight in their performance improvement. That value is especially important to managers, in order to track down their representatives' performances, their effectiveness and efficiency. Managers are able to make different aggregations, filtering and ranking of the performance data of their representatives to get information like what is the average team performance, who is performing the best in the team and who is lagging behind the average.

Therefore, the reporting module is usually filled with different data visualisations, such as line charts, bar charts and pie charts. This module is responsible for quickly delivering the right information to the user as it aggregates the available data, pre-processes it and generates the visualisation. As mentioned before, reports tend to make comparison between certain time intervals (weeks, months, quartiles), which carries the core of their value proposition. Having different charts, like line and bar charts, presenting the same data variables over different time spans can reveal trends in users' performance and progress. The role of data visualisation is in fact, to deliver this information as smoothly, easily and efficiently as possible. Figure 22 gives an example of a reporting module, which can be found in a CRM system. The example contains early version of the visualisation that was created and discussed in Chapter 6.

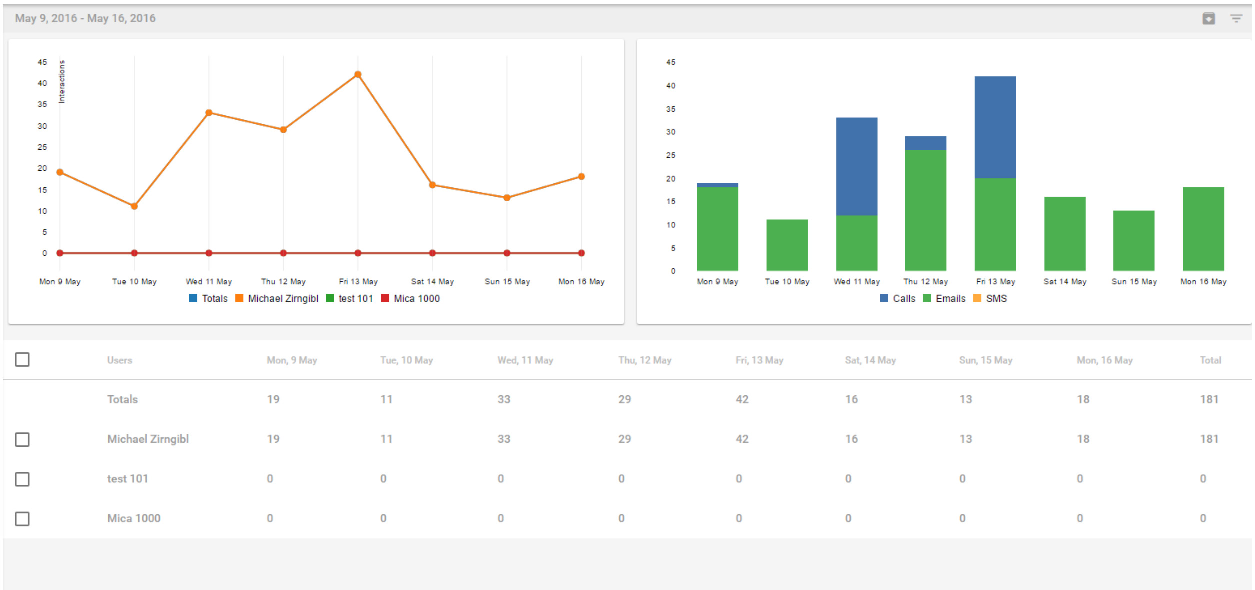

Figure 22 - Example of CRM system's reporting module - data visualisation in form of charts gives quick insight into trends and progress, while bottom table offers detailed information (source: Leadscore CRM)

Sales Pipeline

When being visually presented, the sales pipeline can also considered as a different type of the reporting module and could have also been introduced in the previous section. However, due to its special function and information that it reveals, it is discussed separately.

The sales pipeline describes an approach to selling by representing a sales process as a one consisting of all sales steps needed in order to sell a product or a service to the customer. Each of these sales steps represents a single or group of actions that need to be undertaken in order to complete the current step and move to the next one within the sales pipeline [33]. The first step is usually initial contact with a potential customer that progresses to the next step of converting a customer and qualifying him/her as a lead. To make it clear, lead is a potential consumer of a product or service, created when he or she (or it - business) shows interest and provides contact information. Later on that lead is validated into a sales opportunity which is then, followed by other steps progressed to the closing stage. The sales pipeline is what actually presents all of these step together and actions that need to be taken in order to move the sales opportunity from the initial stage to the final stage.

Information that reveals how many customers (or sales opportunities) actually convert to the next stage is called Closing Probability. It is measured in percentages that show the likelihood of winning the opportunity for the current step and moving a prospect to the next one. This information is especially important since it determines with how many customers a user needs to set up initial contact to reach a desired (or set) goal. It also helps a user determine the “weight” of one’s pipeline. The weighted value describes the likelihood of converting a particular customer as a paying one [33]. It is equal to the sum of the total opportunity values in each sales step multiplied by the closing probability for that particular step.

The sales pipeline can be presented in two possible ways. One possible pipeline view is presenting a list of all pipeline stages and corresponding leads or opportunities that are currently in that stage. This view is useful for getting a more detailed view of which leads are actually in which stage, how long have they been there, etc.

Another way of presenting a pipeline is with the funnel chart which, using its data visualisation, gives a quick overview of the statistics of the sales pipeline. A funnel chart is especially useful when tracking the leads’ progress over the stages and their conversion. Due to its visualisation progression (from top to bottom or from left to right), it conveys the information of the progress of the number of leads or opportunities through the stages. That is particularly valuable since it quickly and easily reveals which stage transition is the weakest, indicating the existence of a certain problem. This way it is possible to identify where business is losing its opportunities. Backed up with conversion rates, this pipeline representation is usually complemented with a detailed breakdown of each stage, in the form of a table which shows certain statistics, such as the number and the percentage of the converted opportunities and unconverted ones, etc.

Combining two funnel charts into one opens up a possibility of data comparison for different timespans. This integration, yet again enables quick insight in which compared stages are more dominant and which compared transition is more successful. Figure 23 gives an example of funnel chart pipeline representation (early version of the visualisation created in Chapter 6).

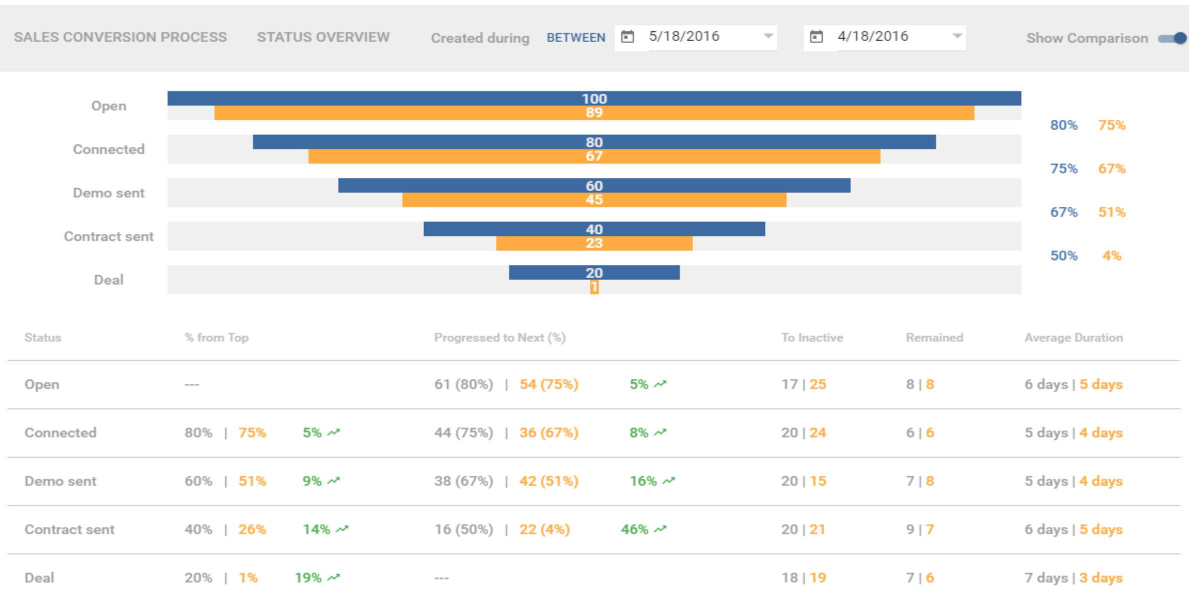

Figure 23 - Sales pipeline funnel chart example - comparing two different time spans backed up with breakdown details table (source: Leadscore CRM)

4.2.2 CRM and customer satisfaction

Customer satisfaction represents one of the key goals companies need to strive for. Since happy customers can also be thought of as free advertising for the company, the significance of customer satisfaction cannot be neglected. Customers need to be put in the centre of business according to business strategies, events and processes [34]. Actually, finding new customers is way more difficult and less profitable than just selling the products or services to the present ones. Gone are the days when companies used CRM systems just to manage their leads and customers. Today, successful businesses are deploying strategies to ensure customer retention and are using their CRM systems to manage the relationship with their customers. Thus they are changing their employees to be more customer-oriented. Through marketing, prospecting, sales process and service, CRM should help its users to create an ideal customer experience resulting in long-lasting positive relationships with their customers [35]. Companies are not only trying to satisfy their customers but they are aiming to do this in a more effective and efficient way than their market rivals to attain their goals.

However, some also argue that having satisfied customers is not enough [36] because customer satisfaction needs to be efficient in order to produce the impact on customer loyalty. Loyalty of a customer is what represents a higher level of satisfaction, it retains him/her and therefore builds a lifetime customer. Research have shown that CRM has a positive relationship with customer satisfaction and loyalty, two dependent variables [34]. While behaviour of the employees and relationship development are mostly contributing to customer satisfaction, it is employee's knowledge about the product and his communication skills that are creating the relationship with the customers by utilizing their satisfaction and finally creating customer loyalty. Therefore, it is essential for employees to have a reliable and efficient CRM system that supports their knowledge (about the product and customer) in order to communicate the right message to the customer. Other studies [37] have also shown that CRM has a significant effect on customer satisfaction. With improving the customer relationship management and making it strong and reliable, customers are getting more satisfied and there is a higher chance of their retention within the company. It is the most efficient way in nurturing relationships with the customers since it is not pure business, but it rather supports personal bonding with them. It is easy to identify the right needs of customers and offer them a proper solution once a personal bond has been built [34]. With more sophisticated implementation of a CRM, there is a higher chance of making a business more fruitful. Thus, companies should consider in allocating their effort and resources in implementing CRM systems in their workplaces and training their employees to provide better quality service in order to have a higher chance of customer retention that results in the end with lower costs and higher profit.

4.3 Cloud-based CRM systems

The previous section mentioned one interesting fact when it comes to making a certain activity, such as calling or sending an email. Users are able to contact their leads or perform certain tasks regardless of their physical location at any point of time. This property however, cannot be claimed by all of the systems. Traditional CRM systems, usually oriented around big-sized companies, are specifically installed within the companies that own them, together with corresponding hardware and are backed up and maintained by the company's personal IT support staff. Due to their high prices, complexity and technical requirements, traditional systems were not affordable to small and mid-sized companies eager for their functionality and service. These pains were relieved when systems started to move to the cloud and when the new generation was introduced - Cloud-based CRM systems.

Cloud computing is largely characterized by obtaining resources (data) according to user's needs and necessities. When one expresses a need to operate a larger computational services, it is not necessary to purchase additional hardware and software components, but rather just connect to the cloud [38]. This brings forth a lot of benefits, especially for small and mid-sized companies, allowing them more agile and flexible development. The days when a CRM system was imagined as bulky, complex and expensive software that needed specialized IT staff as a backup for its installation and maintenance are gone.

Shifting a CRM solution to a remote servers on the internet, or to a cloud, brought a revolution in the business world when it comes to managing customer relationships. The most significant benefits are described in the following paragraph.

The first benefit of using the cloud approach is eliminating the fear that a CRM comes with a complex installation process [39]. A cloud-based solution is, so to say, “IT-worry-free” and users can easily get up and running in no time. There is no need to pay for hardware, software and server installation, upgrades and maintenance - all of these things are done remotely. The only requirement is having a stable internet connection.

Probably the best benefit of a cloud CRM system is consistent accessibility. In a case when a user needs to make an urgent call or send an email while being out of the office, cloud CRM enables it all giving the user access from anywhere at any time and on any device whether it is a desktop, a laptop, a smartphone or a tablet [39]. The internet connection is the only thing required and logging in the system gives the user access to all necessary and required information since the system is always connected to the centralized database.

Ease of use and affordability are other great advantages of a cloud-based approach. The user subscribes to a monthly package at a fixed price and having a device to work on and strong internet connection, he/she only needs to login the system with provided credentials and that is it [39]. System maintenance is carried out by the cloud CRM system provider and hence user has nothing to worry about it.

5. The role of data visualisation in CRM

5.1 Data and visualisation in CRM

When it comes to visualising specific data, such as the data in the CRM systems, the first step is to understand the data itself, available data types as well as the variables. This section is focused on the data that can be usually found in modern CRMs with a slight expansion on data that is found in a specific CRM system – Leadscore CRM.

5.1.1 Data in CRM

Modern CRM systems are a huge source of lots of kinds of data and data types. The number of documents that are stored through the system and retrieved back is measured on a big scale. These documents need to be organized in a proper manner in order to enable quick access to necessary information. That is accomplished by having proper algorithms that are in charge of data manipulation and aggregation, storage, traversal and retrieval. Even though this is out of the scope of this thesis, the next chapter will briefly introduce a high level concept of company's software architecture to acquaint the reader with the data flow present in the company's service offer.

Interactions data

As seen in the fourth chapter of this thesis (see 4.2.1), modern CRM systems are comprised of different parts that are responsible for carrying out different tasks and presenting different information. Since the systems are primarily about tracking the interactions users have with their leads, this is the first type of data that can be examined. In order to communicate with their leads and potential customers, sales representatives mostly use phone calls and emails as a primary ways of interaction, with some minor ones such as SMS, social media, etc. Depending on which type of interaction is used, different information can be tracked and recorded in order to acquire as much information about the lead as possible with the aim of using that insight to build a strong and complete image of a customer and finally sell him/her a service or a product.

When talking about the interactions (phone calls, emails, SMS, etc.), the first information that is tracked here is simply the absolute number of interactions made. This information is first of the many to which usually a timestamp is added in order to track the progress in numbers of interactions over time. Having an absolute number of interactions is valuable information to the user (sales representative) since it can show user's progress over time (in the reporting module, see 4.2.1) as well as indicate how much work is there to be done in order to accomplish a certain goal whether it was a daily goal, weekly or monthly perhaps. However, the absolute number of interactions is also of crucial importance to the sales manager, who is responsible for monitoring his representatives and analysing their performance. This information is of great benefit when it comes to determining the overall and average team performance and setting up the number of certain interaction a team needs to accomplish. It also enables the discovery of the best performers in the team so they can be rewarded and the identification of the ones that are performing the poorest in the team in order to give them additional help and coaching so they can keep up with the rest.

When talking about different types of interactions, it is possible to track different information for each type. That especially is seen with phone calls where the call duration can be tracked which can bring up information like how much effort a user needs to spend on a particular lead. Regarding emails, information about the email status can be tracked so the user can see if a certain email has been sent and if the lead has received the email, opened it and responded.

Interactions data is mostly present in systems' dashboards and reporting modules. As stated before (see 4.2.1), dashboards are used to give a quick overview of users' current performance states and to drive them forward with an incentive of how much work remains for them before completing a certain goal. Therefore, dashboards can be observed as the modules that are more oriented on the future and what still has to come. As opposed to the dashboards, there are reporting modules which are oriented on the past. They deliver the information about past events and therefore are useful in reporting the one's performance and the work that has already been accomplished. Types of this data that can be found in these systems are mostly quantitative, represented in numbers with different statistics. Qualitative information that is found here is mostly to distinguish different types of interactions. This is important to address when thinking of mapping the data to visual elements.

Contacts data

This type of data is usually the first one that enters the system. It is the data that contains different information about the contacts, leads and potential customers. It can be acquired in various ways. One way of getting this data is allowing potential leads to leave their information on company's website by registering for a specific service or newsfeed and simply expressing their interest in company's offer. Another way of acquiring this data can be by scraping the data that is open and available on the internet. Social networks are huge data-open platforms that can be observed as a potential source of valuable information. By scraping, filtering and managing the open data from social networks, companies are able to acquire more precise data about their potential customers. Other methods of collecting contacts data are more traditional ones. One can acquire contact information by having meetings with people or maybe collecting a large number of business cards and contacts on business gatherings, events and meetups.

Contact data is what actually makes up a company's lead. This data normally contains general information about the contact like name, address, profession, phone number and email address. One should be able to perform different operations with this information such as making a phone call interaction with a contact by using his/her phone number and dialling it though CRM's integrated telephony service. A similar activity can be accomplished by using contact's email address and sending him/her an email through the system's integrated email service. However, there is one kind of important information that is assigned to each contact on entering the system. By entering the system, contacts also enter a special procedure which is known in customer relationship management as sales pipeline (see 4.2.1) which contains several steps or statuses. Upon entering the sales pipeline, contacts are assigned a beginning status and while they are progressing through the pipeline, their status is changed by moving to each stage. Tracking this progress over time can bring a lot of benefits such as identifying a conversion rate by comparing the number of total contacts that entered the pipeline with the number of those (out of total ones) that managed to get to the final stage. By observing the conversion between nearby stages it is also possible easily to identify which stages have a problem in leads conversion. Other important analysis of contact status and pipeline data is tracking the numbers of all statuses over time. This information quickly indicates whether the company is growing or not.

This type of data is comprised of both qualitative data which reveals different lead information as well as statistics drawn to these same leads. This needs to be considered when doing a visual encoding of the data to the visualisations.

Task data

This data tells the user what kind of tasks he/she has been assigned to do. This data usually is comprised of the information about a certain task (task name and a description) and other data like contacts data (contact name, phone number, and email address). One kind of crucial information that needs to present here is the data for the task deadline since it is important to execute certain tasks in a certain amount of time. Another important data is assigning a priority status to each of the tasks in order to know which tasks are more important and carry more value. Priority status is subject to change due to different factors, such as the previously mentioned deadline factor. Priority status also depends on the value a certain task carries in terms of the amount of profit related to the task (the bigger the amount, the higher priority the task has).

Task data is also comprised out of qualitative and quantitative data. Qualitative data can give deeper information about particular tasks, related descriptions and leads while quantitative data can contain different statistics of completed or pending tasks.

Even though CRM systems contain even more different data and aggregate with the existing ones in different ways, the previously mentioned data types are the ones that are most commonly found in these systems. The question that arises is how to actually convey this data into a meaningful information that can be easily understood and perceived. The following section explains which visualisation to select when trying to visually express certain data and data types that are found in modern CRM systems.

5.1.2 Choosing the proper visualisation

Visualising data in customer relationship management is crucial for users in order to understand the data, but having also a certain visual interest in the visualisation could be appealing to the sales representatives if it would complement the data that is presented in order to drive them forward and boost their motivation. However, when visualising the data, there are different factors that play a role. One must first realize which type of data is manipulated and then decide what message he/she wants to convey to the reader. In order to realize what he wants to show, one should ask himself following questions [40]:

- Does one want to compare values?
- Does one want to show the composition of something?
- Does one want to understand the distribution of the data?
- Is there any interest in analysing the trends in the data set?
- Does one want to better understand the relationship between value sets?

Answering the previous questions brings us closer to choosing the proper visualisation. Figure 16 offers a conceptual overview in suggesting which chart to pick for the data that we want to present. When it comes to simple charts (e.g. line, bar and pie charts) as visualisation (which is often the case in CRM systems), one must first decide what does he/she want to present with his/her data. Four main things that can be presented are as follows:

1. Data comparison
   - comparing two or more variables of given data. The easiest way to think of this is having a bar chart with two bars presenting two different variables. Comparison is shown by observing the difference (in size) between the bars – the longer bar presents a higher value. (see Figure 24 - top)

2. Distribution of data
   - displaying a frequency of different outcomes from a data sample. An example would be a histogram that shows how many values of the same variable fall into one of the predefined categories. (see Figure 24 - right)

3. Composition of the data
   - describing out of which parts is one entity is composed. A pie chart is a good example here since it shows how variables, in relation to each other, are used to create a whole entity. (see Figure 24 - bottom)

4. Relationship of data
   - describing the relationship of a bigger number of data points over two (or more variables). Examples here are scatterplot and bubble chart. The first one shows a relationship of more data points over two variables (horizontal and vertical position) while a bubble chart also introduces a third variable in form of size of a bubble. (see Figure 24 - left)

Choosing between any of the previously listed options offers several visualisation possibilities, with some of them offering the same chart types, however used differently to convey specific information. The choice of the graph depends greatly on the nature of its variables (time-based, comparison, relational, etc.) which can create subcategories for determining and picking up the right choice.

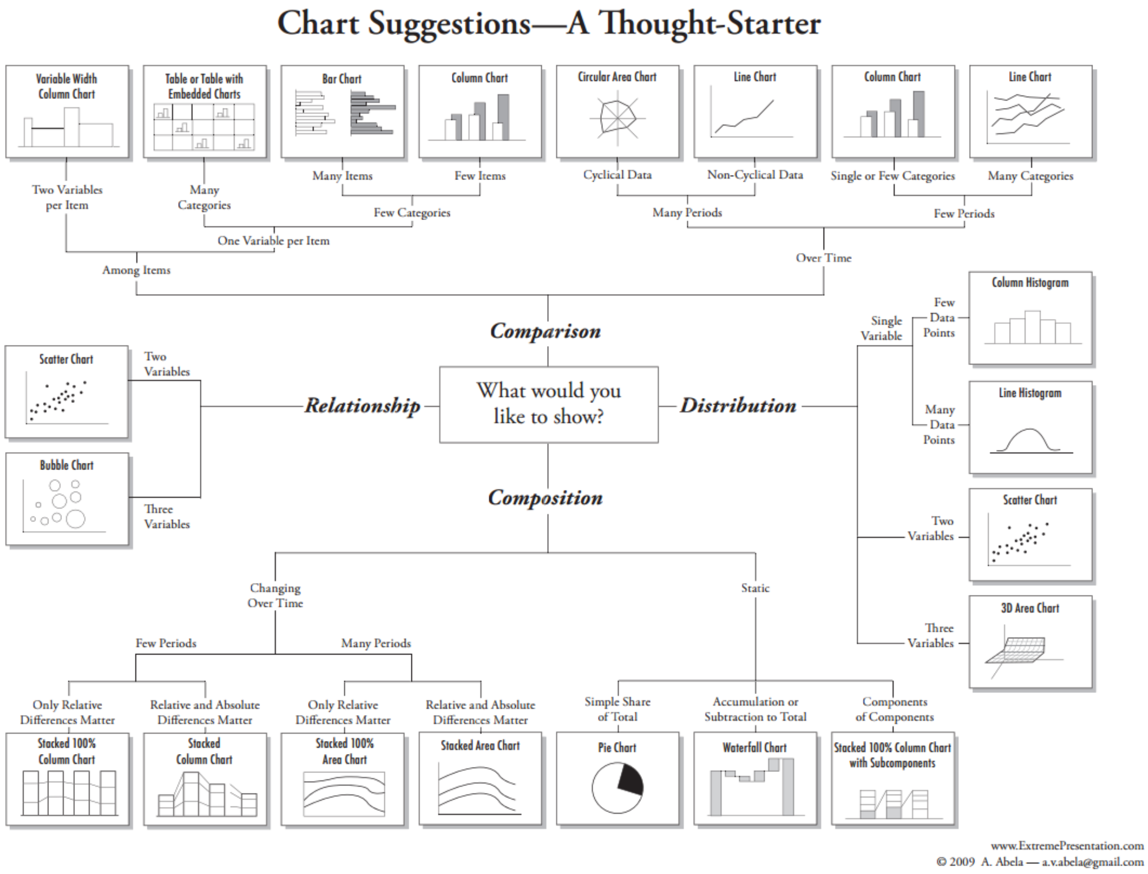

Figure 24 - Different categories and subcategories for determining the proper choice of chart [13]. In order to choose a proper chart, one must first decide what he/she wants to show with particular dataset

After deciding what needs to be shown with a given dataset, one must find the most appropriate chart for conveying that particular information. Some of the basic charts and their overviews are shown as follows.

Column chart

This chart is mostly used for comparing the data among different items or over time for analysing the trends in the dataset [41] [40]. With this chart, human perception easily grasps the information about low and high values by comparing columns' heights. This chart also performs well when one wants to visualize the data distribution (histogram).

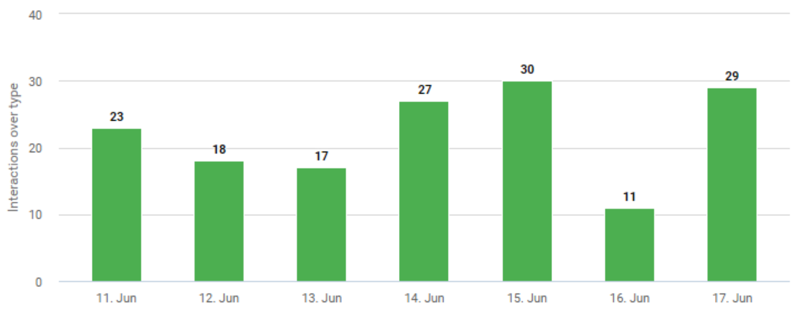

Figure 25 – A column chart is great for comparing the data among different items or over time (source: screenshot - Leadscore CRM)

Line chart

This type of chart (Figure 26) is the most common one when it comes to visualizing data that is dependent on a time variable [41]. It is used to reveal the trends and progress over time and it can also be used to show the data comparison or distribution [40].

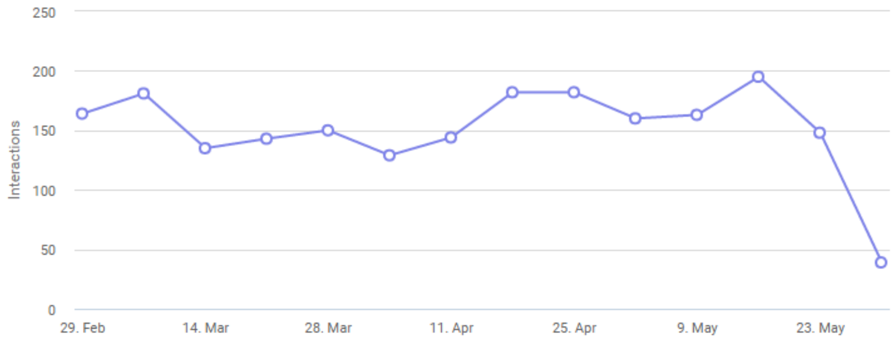

Figure 26 – A line chart is the most efficient one for conveying the data dependent on a time variable (source: screenshot - Leadscore CRM)

Stacked column chart

This type of chart (Figure 27) is used to compare different items while showing the composition of each item that is being compared [40]. Each bar incorporates separate data series so that it is possible to see how the components contribute to the whole [41].

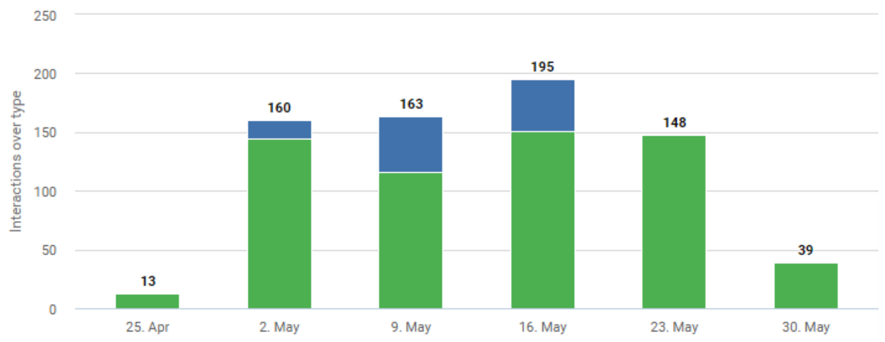

Figure 27 – A stacked column chart enables comparison while showing composition (source: screenshot - Leadscore CRM)

Pie chart

A pie chart (Figure 28) is used to show a static number (the one that does not change over time) of the composition of something. It is divided into slices to illustrate the numerical proportion of different categories and how they represent a part of a whole relation [41]. Numbers presented by a pie chart are not absolute but in percentages which always sum up to 100% [40].

However, this type of chart is rarely used in scientific literature. Having too many pieces makes a pie chart less effective and readable [42]. It also has problems when comparing the data slices since the human eye is not used to deciphering angles.

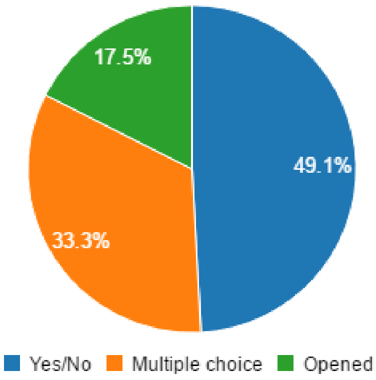

Figure 28 – A pie chart is showing a part of a whole relationship in percentages (source: screenshot - Leadscore CRM)

Funnel chart

A funnel chart (Figure 29) shows a series of steps and the conversion rate for each step [22]. In customer relationship management, it is can be used for tracking the sales process or customer conversion over the sales pipeline.

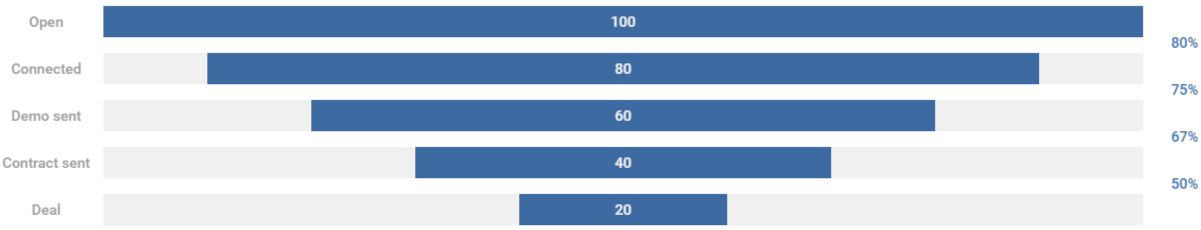

Figure 29 – A funnel chart is particularly useful for tracking a conversion of leads in the sales pipeline (source: made by me - Leadscore CRM)

In the world of CRM systems the previously listed charts can be responsible for tracking and displaying previously mentioned data such as interactions with leads over time and type, contact information statistics, etc. It is obvious by now that visualised data yields more insightful information by revealing hidden patterns than a tabular presentation of mere numbers. Section 5.2 gives an overview of how data visualisation influences the user's everyday work, in particular their decision-making.

5.2 Impact on users (sales people)

An important thing to notice first is the audience the data visualisation is used by and consequently is influencing. As said in the Introduction, the target audience in this paper are people within the field of sales, more specifically sales representatives and managers. Their work ranges from selling the products, goods and services, working with customers to find what they want to creating the solutions and ensuring a smooth sales process [43]. As mentioned in the fourth chapter, a CRM system is there to make these activities easier for them and to ensure they are always up-to-date with their deadlines, meeting their customers' expectations, making interactions with them etc. Data visualisation, in CRM systems, needs to be a tool that helps its users comprehend the exposed data more efficiently and meaningfully. Data presented to the user is usually found in tabular format, simple rows and columns with “raw” numbers in their cells. That format proves to be useful when a user wants to dive deep into the data by examining each data point individually and seeing it in more detailed. However, a tabular format does not appear to be useful when it comes to grasping the data more efficiently and easily, discovering new trends and patterns and interacting with data.

Data visualization has a potential to make a different impact on its audience and to create a different “value” for its users depending on the purpose of its application. When it comes to the domain of business and sales, data visualization can be significant since it directly influences the sales representatives, the ones in charge of making the company’s profit. One of the potential influences that data visualization imposes on those kinds of users is its impact on users’ decision making. Informed decision making is the foundation for building a successful business. With growing business, the amount of collected information grows too. Thus, as a business decision maker, one needs access to highly visual tools that provide help to make the right decisions quickly [44]. This can also be critical in strategic decision making, in which data interfaces need to be designed in order to support the most critical decisions an organization can make [45]. However, due to the scope of this paper, operational decision making (which is focused on everyday operations) represents more interest for data visualization discussed in this paper. Operational decisions are related to everyday operations within the company. A lot of interactions that take place on a daily basis represent the result of operational decisions. Since they are repeated, they have usually have a short-term. In these everyday actions, data visualization has a potential to aid the decision maker (whether he/she is a manager or a sales representative) to make rational and well informed decisions.

A question that arises is that of how data visualization can actually aid in decision making. To answer this question, we must a take step back and find out how our brain makes complex decisions. Business people tend to combine statistical calculations, past experiences and intuition in order to make impactful decisions. Statistical calculations are referred here as binary processing while past experiences are referred to as pattern recognition (recognition of something already seen/experienced). Binary processing is something where machines excel especially on a more complex level while humans are capable of processing merely simple binary calculations (e.g. knowing the revenue last quarter). Pattern recognition however, on a basic level manifests itself as instinct - we will perform some action based on our instinct (e.g. when catching a ball, we need only the visual cue of the ball in the air and our past experiences to instinctually know where to go) [46]. Because our brain is especially sensitive to visual cues, bridging the gap between binary processing and pattern recognition can often be achieved with certain visualisation forms. Simple charts and graphs help us see certain progress such as declining revenue performance, and complex visualizations can aid us in finding different patterns and trends that our brains are not otherwise able to identify from strings of data. At the highest levels of decision-making, we must make challenging leaps to understand different uncertainties. We are relying on data and analysis in tandem with our intuition to guide our process of cognition. Making analysis more digestible and readable for the users through data visualization is an important element in complementing our decision-making.

6 CRM data visualisation research and evaluation

This chapter describes designing, developing and evaluating interactive visualisations supplied by real-time data from interact.io platform. First it lists the technology used to build the visualisations (section 4.1). That section is followed by one that describes the data visualisation pipeline process (characteristic for all the four visualisations that were built) and design and visual encoding of each visualisation (section 4.2). The last section in this chapter describes the evaluation of the designed visualisations and their testing in order to validate and support the given hypothesis (section 4.3).

6.1 Technology and implementation

The CRM system that was used for getting the relevant data and building the visualisations upon that very data within the system itself, is a cloud-based CRM system called Leadscore. That means that the system is built for the web as a web application and offers its services via the user's browser and internet connection. On the other end is a platform responsible for collecting and tracking the user's actions and reporting them to the client (CRM system) when requested.

Since the system was created as a web application, technologies used to build the corresponding visualisations were also from the same domain of the technologies and tools - the web based ones. For this paper, they can be divided into two groups:

1. Ones used for getting and pre-processing the data, wrapping up the visualisation and determining its position within the system

2. Ones that are used for visualisation in terms of creating visual graphics, shapes and artefacts and mapping the data to corresponding visual encodings

6.1.1 Technology stack (tools)

The first group of technologies and tools include the following ones: HTML, CSS, JavaScript, AngularJS and Moment.js. Among others, these technologies and libraries played the most important role in receiving the data and its manipulation. Each one of them is given a short description in following sections.

HTML (Hypertext Mark-up Language) is a mark-up language for describing web pages. It is used for creating and visually representing a web page and it presents its basic building component. It is in charge of determining the web page's content, but not its functionality [47]. This mark-up language describes the structure and semantic content of a web page. Being a mark-up language means the language itself is comprised out of mark-up tags, in this example HTML tags. Each of the HTML tags describes different content of a web page.

CSS (Cascading Style Sheets) is a stylesheet language used to describe how HTML (or XML) elements should be rendered and displayed on screen, paper or in other media. It is in charge of web page's presentation in terms of its style. CSS binds custom style specifications with mark-up languages through the use of their elements' names, classes and id's in order to perform certain formatting and appearance of a corresponding tag or element.

JavaScript is an interpreted programming language built for the web. It is usually interpreted in a web browser's scripting engine. Initially used only as a scripting language for web pages, development and expansion of JavaScript started to grow in order to develop greater and more complex web applications and eventually enabled the language to expand also too many non-browser based environments [48]. JavaScript is an event driven language meaning it is possible to declare a certain piece of code to be executed when a specified event occurs. When a certain event occurs, a piece of code called callback function is executed. This "event loop" is non-blocking which enables other pieces of code to be executed while still awaiting an answer from some other function. In example of the cloud-based CRM system, JavaScript will enable the execution of other pieces of code while waiting for a platform to return an answer to a certain query.

AngularJS is a complete open-source web application framework based on the programming language JavaScript for creating rich client side web applications. It uses HTML as a template language and enables extension of HTML syntax in order to make it more expressive. Using Angular's data binding and dependency injection the amount of code is reduced, code structure quality is increased and testing is made easier. When extending existing HTML, Angular teaches the browser a new syntax through constructs called directives [49]. Leadscore CRM system is based on Angular as main client-side framework. In terms of building the visualisations for the project, Angular is used to make and send queries for retrieving the desired data. It is also in charge of data manipulation and pre-processing and data communication to the other parts of the application. It is also responsible for tracking and processing user interaction.

Moment.js is a JavaScript library used for parsing, validating, manipulating and displaying dates. In Leadscore CRM it was used to transform date-related data into the appropriate formats needed for visualisation libraries to parse. It was also used in defining and applying different date filters that needed a consistent date format when communicating with system's backend and corresponding platform.

The second groups of technologies and tools are the ones used for creating the graphics and visualisations: SVG, D3.js, C3.js and Highcharts.js

SVG (Scalable Vector Graphics) is an XML-based mark-up language for describing two-dimensional vector graphics [50]. It can be thought of being to graphics what HTML is to text. SVG specification is an open standard and its tags, elements and their behaviour are defined in XML text files. Three types of graphic objects are supported in SVG which are vector graphics (vector defined graphics), raster graphics (pixel defined graphics) and text. By using different mark-up elements (tags), it is possible to define simple as well as complex shapes which are rendered and drawn upon browser interpretation of a text file. SVG can be thought of as a foundation upon which different visualisations are built.

D3.js (Data-Driven Documents) is a JavaScript visualisation library for manipulating documents based on data. It helps bring data to life using standard web technologies such as HTML, SVG and CSS [51]. It allows users to bind arbitrary data to Document Object Model (DOM) and then apply different data-driven transformations and shape manipulations to the document [52]. In this project, it was used to create two visualisations which can be seen in sub-section 4.1.2.

C3.js is a JavaScript charting library built upon D3.js. It provides an easy, out-of-the-box application programming interface (API) for manipulating SVG elements in order to create a handful of basic charts much faster than using standalone D3.

Highchart.js is a JavaScript charting library that offers an easy way of adding interactive charts to a web page. Highcharts supports only basic charts, however with great flexibility, compatibility and interactivity. Its rich API offers high customization of available charts.

6.1.2 Building the visualisations

For certain parts, the design of the Leadscore CRM system lacked visualisations of the data. Those parts were the dashboard and reporting modules. Also, the already provided visualisation in the email campaign was not satisfactory for single email campaign as well (for the email campaign, user sends email to more than one lead while for a single email campaign, the user sends single email to a single lead). Using the previously listed technologies and tools, four visualisations were built for this project. The idea behind them was to offer a visualisation of the data to give support to the future users of the system in understanding the data efficiently. Three built visualisations were using real, gathered data supplied from the backend and system's platform. The data used for the fourth visualisation simulated the real data that still had to be provided from the system's backend (the visualisation was designed in advance – without real-time data since the company's backend development team was focused on more important work). Figure 30 shows the data visualisation pipeline for this particular project. Each of the visualisations took the same approach. Available data was retrieved and collected from the system's backend via AngularJS. Following retrieval, data

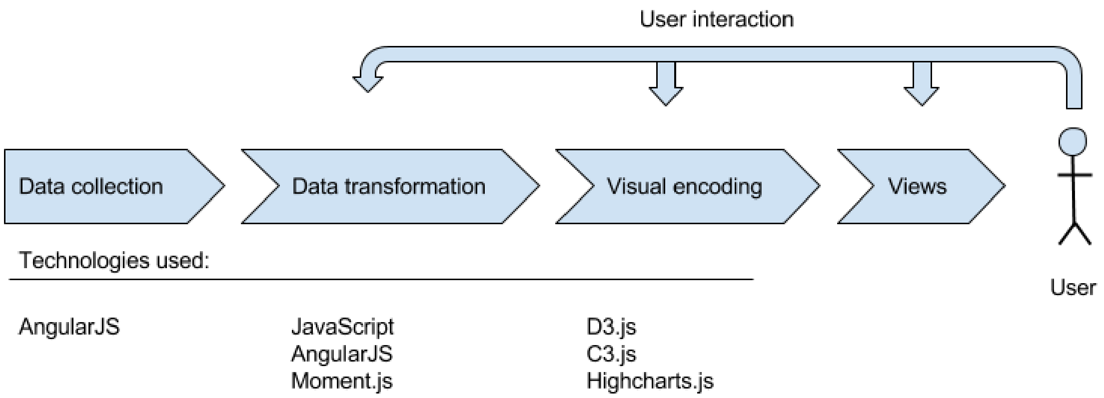

Figure 30 - Data visualisation pipeline for this research (source: created by me)

was manipulated with JavaScript and AngularJS in order to accommodate it to visualisation and charting libraries. Moment.js was used to manipulate retrieved dates and to bring consistency for visualisation libraries. Furthermore, one of the following libraries was used: D3, C3 or Highcharts in order to encode the data visually and generate the views for the user. By interaction with visualisation (applying different filters, hovering) the user can transform the given data, which is encoded again in real time and views are generated and displayed to the user.

Funnel Chart

The funnel chart was used to visualise the data related to lead conversion (see Figure 31). It used dummy data that simulated the real data of which retrieval (during the writing of this thesis) still needed to be implemented. Due to work overload, the backend developers hadn't been able to finish this data retrieval. However, its content had been discussed and the data had been defined. Simulated data that was visualised contained three variables: time, type of sales pipeline status, and number of leads that had been in a certain status (current ones + ones that have passed it). Since one of the most effective and therefore important visual encodings is length (see 3.3.3), it was used to encode the key data for this visualisation: the number of leads that passed a certain status. Bar length was used to visually convey this quantitative information. The type of sales pipeline status was encoded with vertical position and status name. Each status has its own bar and name and the statuses are aligned vertically, starting from the first one on top till the last status at the bottom of the visualisation. An additional, aggregated variable in this visualisation is conversion rate. It has been added as a number in percentages to the blank space between each of the status bars. However, by changing the visualisation filter from absolute to relative numbers, the length of the bar becomes a visual encoding for the conversion rate rather than for the number of leads passing the status. This is to provide useful insights when triggering a comparison which generates additional funnel chart for a different time period to see the difference between the conversion

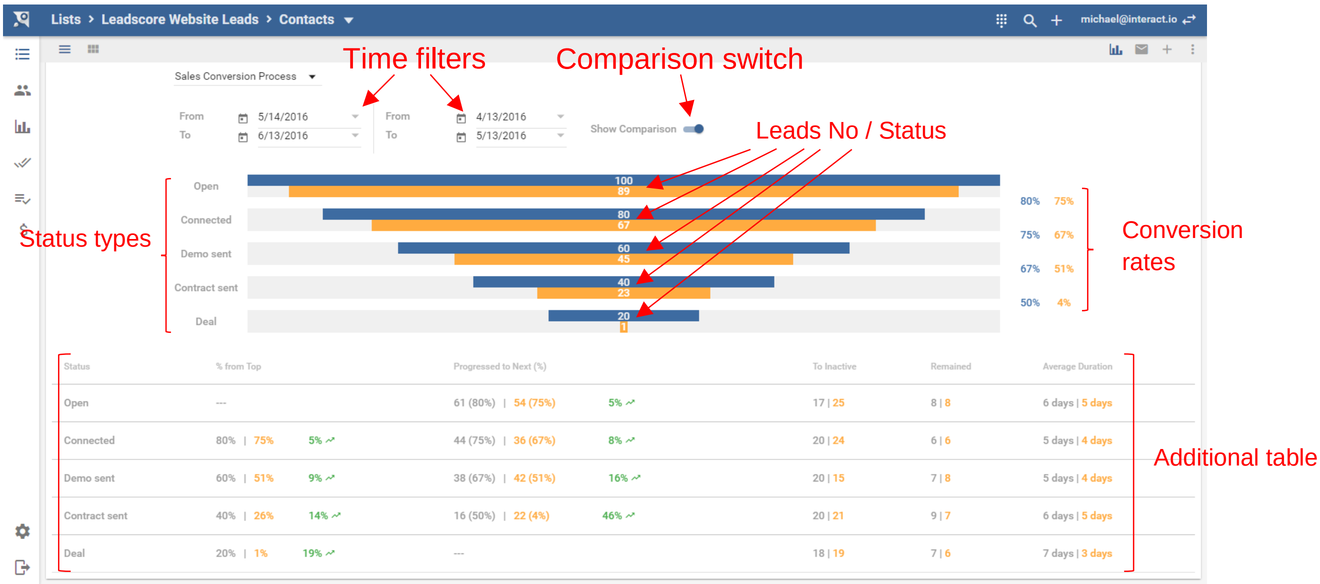

Figure 31 – Created visualisation - Interactive funnel (blue and orange bars) chart – this chart is interactive chart since it enables user interaction through setting the date filter and triggering the comparison switch to see compared chart (yellow bars)(source: screenshot – Leadscore CRM)

rates of two periods. Two different funnel charts (qualitative information) are encoded with colour hue to provide a clear distinction between them. For this visualisation, time was not directly visually encoded but rather exposed in the visualisation itself as the whole visualisation was put into a time span that was able to change by changing the time periods. As additional visualisation, a table was provided below the funnel chart to display summary information. This visualisation is also described in my minor thesis [7] since it was tested to assess the impact of data visualisation on the system's value proposition. For this thesis however, different questions were asked to evaluate and assess the visualisation itself.

Line chart & stacked column chart

The charts seen in Figure 32 were used for the reporting module for the managers in the CRM system. The data provided for the visualisation contained real-time statistics and metadata of users' interactions (phone calls, emails and SMSs) received from the backend through the system's platform.

For the line chart, the data contained three variables: a time variable, the type of a user and number of user interactions that is dependent on the time variable and the type of the user. Time (as quantitative variable) was encoded by a planar variable of position along the horizontal axis while the number of interactions (also quantitative variable) was encoded by position along the vertical axis. As stated, the variable of number of interactions was dependent on the time variable and therefore was encoded with a planar variable. Connecting the dots that are positioned with planar variables, trends are emerging and a line chart is created. The third variable, the type of a user is a qualitative data type. Furthermore, it is a nominal type of qualitative data since it cannot be meaningfully ordered, but rather represents distinct, equivalent categories. Since colour hue performs particularly well for encoding

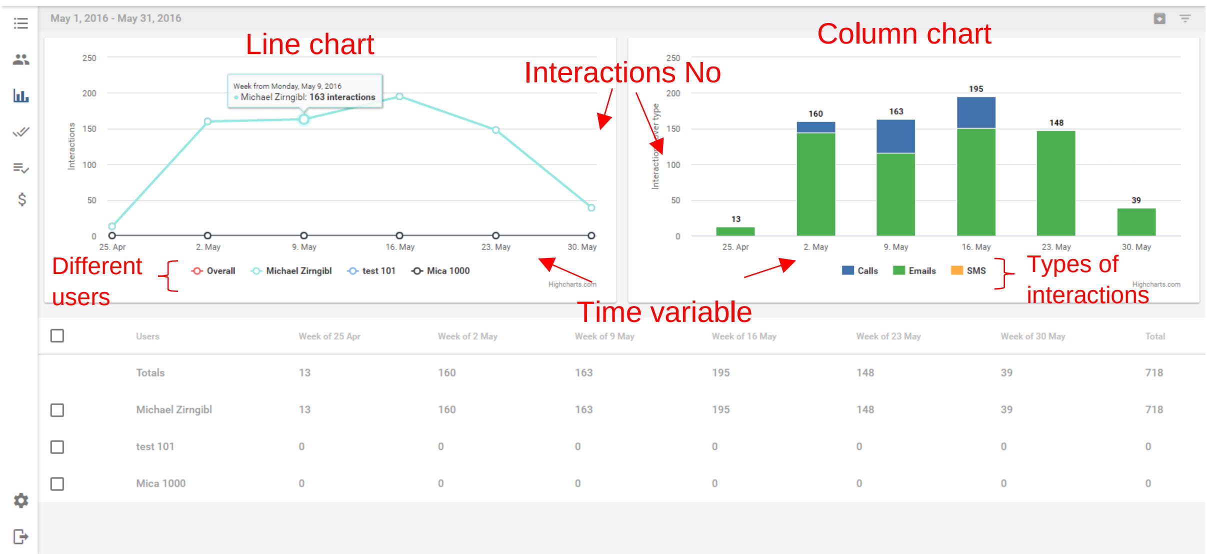

Figure 32 - Created visualisation - interactive line and stacked column chart (date, type, direction filters; tooltips on hover) (source: screenshot – Leadscore CRM)

nominal data (see 3.3.3), different users were encoded with distinct colours, meaning that lines representing their interactions in time contained different colour hues.

On the other hand, the stacked column chart data contained also three variables. Those are: a time variable, the number of interactions per type and the type of the interaction. Similar as in the line chart, the time variable was encoded on the horizontal axis while the number of interactions per type was encoded on the vertical axis. However, the number of interactions was encoded also with a length of a column thus forming a column chart. To distinguish different types of interactions colour hue was used for each type as a retinal variable for encoding the type. Furthermore, to find out the contribution of each type to the total number of interactions, columns for each type were stacked on a single column per one time value and thus creating a stacked column chart.

Pie charts

A pie chart type of visualisation was used to visually express real-time data that was related to the datapoints (see Figure 33). Datapoint is a term specific to the tested CRM system, and it represents additional information for a particular lead. That information is presented in the form of labels with different colour coding and textual abbreviations. Data interesting for visualisation in this example consists of aggregated datapoints and their statistics over a certain amount of contacts to get an insight in different information (e.g. seeing how many contacts possess an active datapoint or how many of them have certain datapoints). In these examples we are talking about part-to-whole relationships which gives a certain direction for choosing the appropriate chart. Two visualisations are given, both of them with two variables. The first visualisation contains the type of a datapoint according to its activeness (open or not) as a nominal variable and number of leads per datapoint activeness as a quantitative variable. The second visualisation contains the type of an open datapoint as a nominal variable and number of leads per different open datapoint.

A pie chart was used here to visualise the part-to-whole relationship between the datapoints. Shapes and (pie) slices were used to encode quantitative variables related to the total amount. Colour hue was used to encode nominal variables to make a distinction between the datapoint types.

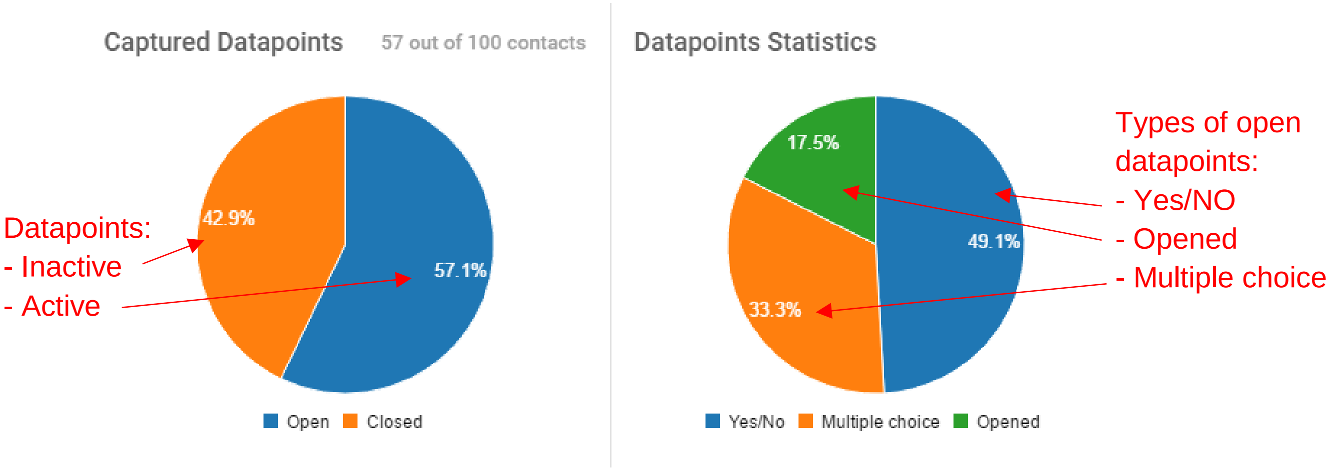

Figure 33 – Created visualisation - interactive pie charts (tooltips on hover) (source: screenshot – [Leadscore CRM])

Email campaign timeline graph

The timeline visualisation seen in Figure 34 is concerned with an activity a user of a CRM system performs - an email campaign. The user can select multiple leads and send them an email simultaneously. The information whether they have received the email, opened it, responded, etc. is tracked down, recorded and stored by the platform. However, this visualisation focuses on sending an email to a single lead to track down his activities with an email he has received. The data presented here is real-time data received from system's platform and contains multiple variables. For the purpose of visualisation, three of them were used: time variable, the type of an email activity status and state of an email activity status (which can be fulfilled or not). First, in this visualisation type of an email activity status was encoded with a circular shape as a planar variable in horizontal position. The state of an email activity status was encoded with a colour hue – a fulfilled status had inner circle filled with colour, while an unfulfilled status had white colour as a circle fill-in. Time was encoded with text and numbers and was dependent on the activity status type variable.

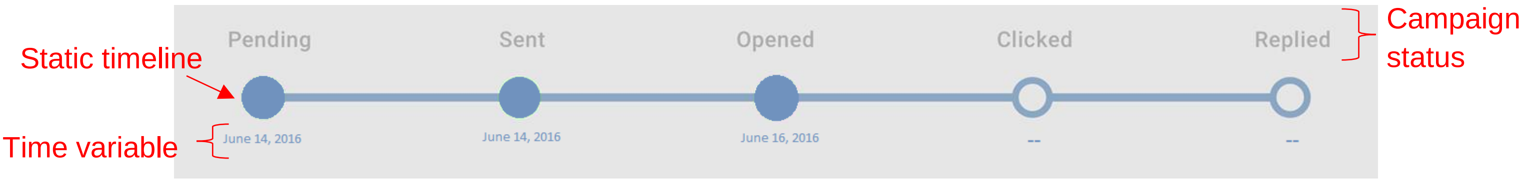

Figure 34 – Created visualisation - static timeline with completed steps and time context

6.2 Testing the visualisations

After developing these four data visualisations for the CRM system, it was time to test and evaluate them. Given the fact that the new visualisations were evaluated in relation to the old visualisations currently present in the system, A/B testing had been selected as evaluation method for this research. A/B testing, or split testing, is a process where the tester is running a simultaneous experiment between two system features (or systems themselves) to see which one performs better [52]. It is a method that validates any new addition or change in a system that has a potential of improving its usability. An A/B test consists of creating two alternative features for a specific part of a system and showing each of them to a predetermined percentage of the target audience. The existing feature design is commonly used as version A (the control version) while version B is represented by a new feature design (the test version). In this research, the control version was represented by the old system's data visualisations and the test version is represented by the newly designed data visualisations. General practice in A/B testing is to make an even split for testing in control and test groups (50/50). Alternatively when having a clear champion, a variable that one is confident in its success, a practice is to apply uneven split (90/10) to the test [53]. The latter one, alternative approach is used in this research project, as the test version of system's feature (new data visualisation) is regarded as a clear champion between two tested versions. It is important to mention that these were not the laboratory experiments, but tests that were done with actual users that performed given tasks with the two versions of the system.

While A/B testing is generally conducted using online surveys to test two versions and collect the data of users' opinions, in this research the interview was the main usability method in conducting the testing. More specifically, one-on-one interviews were used. Interviews are a valuable method for exploratory user research and are useful when there is a need to explore user's general attitudes, beliefs and thoughts surrounding a certain problem, application or its feature and specific tasks that application or its feature supports [54] [55]. With the provided information one is responsible for iterating and redesigning certain features in order to address the problem as well as for having it tested. In this research, interviews were conducted for both of the tested versions, the control one and the test one. Interviews were conducted over four tasks that correspond to four visualisation features that were to be tested. In the control version interviews were conducted over the old visualisations with a focus on finding out whether a user can find specific information hidden in the data. In the test version interviews were conducted over the newly designed interactive visualisations described in the previous section in order to find out if and what kind of information a user can discover in them. Both user groups were asked nearly the same interview questions. Key questions that have been asked were open ended ones to avoid constraining the users from expressing themselves freely. Concrete questions being asked for all four tasks in the interviews can be found in the Appendix. I have asked the participants specific questions in order to find out whether they can perceive and interpret the data from the given visualisation. I was interested in whether they perceived this information quickly and efficiently and whether that influenced on their decision-making. Data from the interviews was collected in qualitative form in written and mostly audio format. Data that was gathered in the written format was metadata about the participants of the study, their personal information - age, gender, education and nationality (see Appendix). Data gathered in the audio format was the one related to the subject of the interview.

Regarding the target audience, people from a business domain - sales people with an experience of using CRM systems were most eligible and were preferred for the test. However, due to the lack of sales representative audience and accessibility limitations, not all of the participants (but most of them) were from the most desirable target audience. The number of users participated in the tests was nine (9). Six (6) participants were from the most desirable target audience while others were represented by two (2) developers and one (1) product manager. The reason for allowing them to participate in testing was because of their close relation to the field of sales since all three of them were working tightly in cooperation with sales representatives. Therefore, it is correct to think of them as qualified as eligible representatives. All of the participants possessed minimum a university bachelor degree. Their age group was from 21 to 30 years old.

Regarding the split among control and test group, since there was a clear expected champion between them, an uneven split was applied to the target groups. Control version was given 2 users (from most desirable audience) to test while the test group had 7 users participating. The uneven ratio applied in this A/B testing was ~20/80. None of the participants had been a part of interact.io company and none of them had already worked with the old system. As stated before, all four tasks and the questions that have been asked to the user groups can be found in this paper's Appendix. The first out of four visualisations is also described in the minor thesis [7]. However, evaluation in the minor thesis is focused on the influence of data visualisation on the CRM's value proposition and therefore different things have been tested and the participants have been asked different questions (but it was part of the same evaluation session).

Before each test, the participants were briefly introduced to the corresponding system's part and necessary information about the system (they were given help to get familiar with the parts of the Leadscore system). Then, for each test the participants were shown the corresponding visualisation (old visualisation for control version; new visualisation for test version). Depending on whether the given visualisation contained any interactivity, they were asked to interact with the visualisation. That included changing the time filters (funnel chart, line and column charts), toggling comparison (funnel chart), hovering for a tooltip (bar and line charts, pie charts). During the testing procedure, interview was conducted with the participants to collect the research data. The interview contained the same questions for both of the tested groups. With all four tests combined, interviews lasted from 15 minutes (shortest test of the control version) till 27 minutes (longest test of the control version). These timings refer only to the questions that were asked for the evaluation this thesis discusses (time that was taken to ask questions and acquire participants' answers in the first test that was for the minor thesis [7] was excluded).

6.3 Results

The first results that are important to mention are those related to time that took participants to recognize the data. The data about the time was extracted from the audio that was recorded during each interview. For the A version, the time taken for participants to recognize the data ranged from ~10 seconds (when the number of leads was low – in the first two tests) to the point they could not recognize the data. For the B version, the time it took for the participants to recognize the data ranged from ~4 seconds (second test) to ~16-20 seconds to recognize the visualisation they haven't perceived at first. These first results indicate that B version visualisations performed well in conveying the data faster and more efficient.

The interview results are listed in the form of a table containing two columns out of which one represents the A version (control version) and the other the B version (test version). Since A/B testing is about comparing two different features, it is important to enable the comparison of the results as efficiently as possible thus the table contains two columns of different features to get a quick glimpse in which version is the winner. Since there were four different A/B tests corresponding to four different visualisations, it would not be convenient to present the questions and their results in a strictly textual format. Rather the results (answers to questions) could be reduced by grouping them into different ordinal categories corresponding to their weight of positivity or negativity. Even though ordinal variables are not usually encoded with colour hue, these five ordinal categories were created and encoded with colour hue due to the meaning these colours convey in western society (red colour for something negative like a warning or a threat; green colour for something positive and successful):

- Participants' answers indicated serious deficiencies in the design
- Participants' answers indicated some deficiencies in the design
- Participants' answers indicated that design was not either good nor bad
- Participants' answers indicated positive features in the design with the potential for improvement
- Participants had positive opinion about the feature and answered the question with understanding and ease

The results are comprised of visually encoded (colour hue) participants' answers. These answers are then aggregated and their average colour (considering red equals to value 1 and blue equals to value 5) is taken as a result. The final result can be altered (lower or higher) than the most frequently occurring answer if an extreme case in found among the answers (e.g. if all the answers are green and blue except for the one or two that are red – then the answer is lower).

Tables with results of the answers and the assigned colour codes can be found in the Appendix 1. Averaged results are shown in the following tables.

Questions asked (reduced)	**A feature (control) - old visualisations**	**B feature (test) - new visualisations**
Test 1 - Funnel chart	A raw table	Interactive funnel chart vis.
1. What data do you see?	🟡	🟢
2. How many leads were in a specific status?	🟠	🔵
3. Leads progress in last month?	🔴	🟢
4. Conversion rates over the stages?	🔴	🔵
5. Conversion rates comparison of two time spans?	🔴	🔵
6. Opinion about the visualisation	🟠	🟢

Table 1 - Comparison of visualised A/B results of Test 1

Questions asked (reduced)	**A feature (control) - old visualisations**	**B feature (test) - new visualisations**
Test 2 - Line & Stacked bar chart	A raw table	Interactive line and column charts (straight lines and labelled columns)
1. What data do you see?	🟠	🟢
2. Could you easly track the progress of your data?	🔴	🔵
	Additional curved line and column chart (without labels)	
3. What was happening with the data on certain dates? (curved line present)	🟠	🔵
4. Counts of interactions per one (individually and together)?	🟢	🔵
5. Opinion about the visualisation	🟡	🔵

Table 2 - Comparison of visualised A/B results of Test 2

Questions asked (reduced)	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 3 - Pie charts	A raw table	Interactive pie charts
1. What data do you see?	● orange	● green
2. Detection of datapoints (in numbers and active ones)	● red	● green
3. Active datapoints per type	● red	● yellow
4. Opinion about the visualisation	● orange	● yellow

Table 3 - Comparison of visualised A/B results of Test 3

Asked questions	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 4 - Email (single) campaign timeline	Visualisation in form of unconnected rings (Figure 41)	Static timeline with dates
1. What data do you see?	● yellow	● cyan
2. Number of emails in the campaign?	● yellow	● green
3. Current campaign email status	● orange	● cyan
4. Opinion about the visualisation	● yellow	● green

Table 4 - Comparison of visualised A/B results of Test 4

7. Discussion

From the tables in the previous chapter, it was easy to get a clear distinction between the two visualisations and their level of design friendliness for the users. The interview backed up A/B testing and the corresponding visualisation of the results enabled an easy comparison between two tested features' designs.

Furthermore, corresponding results visualisation enabled even easier discussion of the results and picking up and confirming a right winner. Having encoded users' answers with colour hue and putting the values in a form of a table, this data presentation gives us the opportunity to extract the variables to analyse the results. Those variables are the already mentioned ordinal variable representing the positivity of the answer (encoded with colour hue in tabular visualisation) and a quantitative variable that corresponds to the number of occurrences of a particular answer (category of an ordinal variable) within a single test for a single feature. These two variables are used to visualise the results data which enables easier interpretation and analysis. The first variable (the ordinal one already colour encoded) is encoded with vertical position, by ordering the categories from the most positive one in descending order. Second variable is simply encoded with horizontal position by adding the number of values that correspond to each of the categories. Single visualisation for each test gives an easier comparison between the two versions. Each circle represents an answer to a question. Since both tested groups received the same questions, they have the same number of the answers. A winner is determined by the number of more positive answers.

7.1 Determining a winner

Test 1

The first test clearly shows that version A has some serious issues that had to be addressed (more than half of the responses). Two (aggregated) responses indicated some deficiencies while only one (aggregated) response was neutral. Participants managed to recognize the data and how many leads were in a specific status only if the number of leads was low. When it came to recognizing and perceiving the data about the conversion and conversion rates, they were not able to answer any of the given questions. This visualisation was particularly poor in conveying that information and no positive responses were gathered (see Appendix – Figure 39).

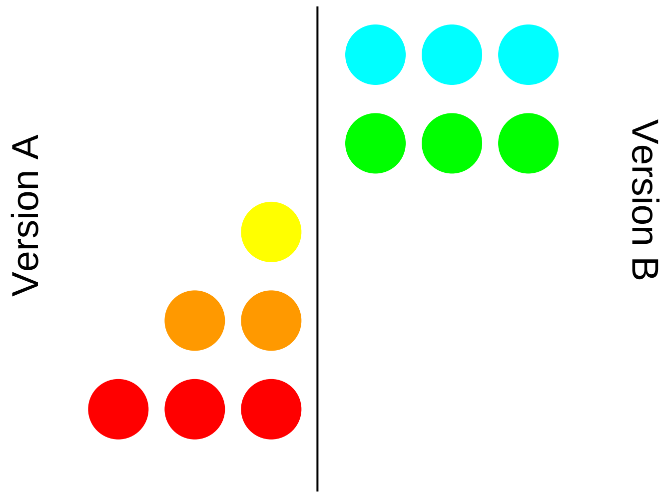

Figure 35 - Visualisation of Test 1 results for easier comparison

Version B on the other hand clearly indicates that the proposed solution was a proper one with half of the responses being the most positive (cyan) and other half with some improvement potential (green). Visualisation performed particularly well in conveying the information about lead conversion and conversion rates (funnel bar length progression – from longest on the top to the shortest at the bottom). It also performed well in visualising the number of leads per specific status (funnel bar length). Figure 35 shows the visualised results of the Test 1. **Winner:** Version B

Test 2

Even though only one response with a critical issue was reported, version A in test 2 still had flaws with its users which is easily seen with two orange dots representing features with some deficiencies. However when complemented with additional curved line and column chart without labels, this visualisation also had positive responses, in terms of positive users' interpretation of the data and its usefulness. However, its performance was not precise since its curved lines on the line chart gave a wrong impression of data (fluctuations of data even though there weren't any).

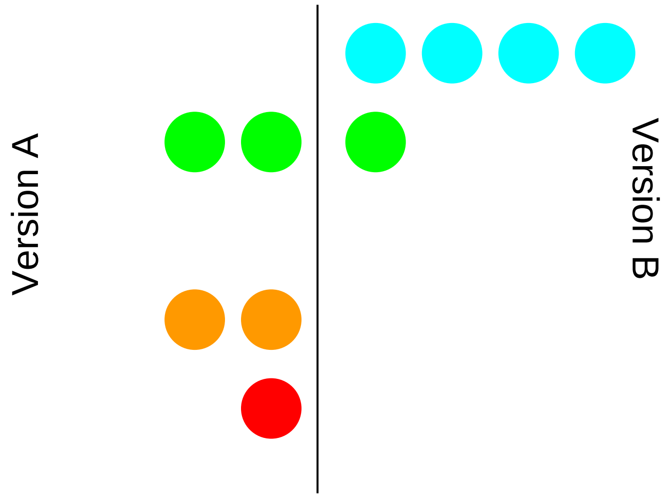

Figure 36 - Visualisation of Test 2 results for easier comparison

Version B however, clearly takes the lead here having almost all of the responses as the most positive. It performs extremely well when visualising the data dependent on time variable (both charts), precise (for line chart with straight lines compared to curved lines) and efficient in its accuracy (for column chart with labels). Version B makes it a suitable visualisation for given data. Visualised results of the test can be seen in Figure 36.

Winner: Version B

Test 3

The first visualisation in the third test also indicates serious deficiencies in data presentation. Both halves of the responses were distributed on the lowest scores. This was mainly due to inability for the users to track all of the datapoints and their distribution and status over a large number of leads - aggregated data was missing (see Appendix - Figure 39). Version B, even though clearly better, was also not the perfect solution for the users. While having half of the responses were positive with improvement potential, the other half was neutral. This was mainly due to the fact that the pie chart was having issues in presenting a larger number of different

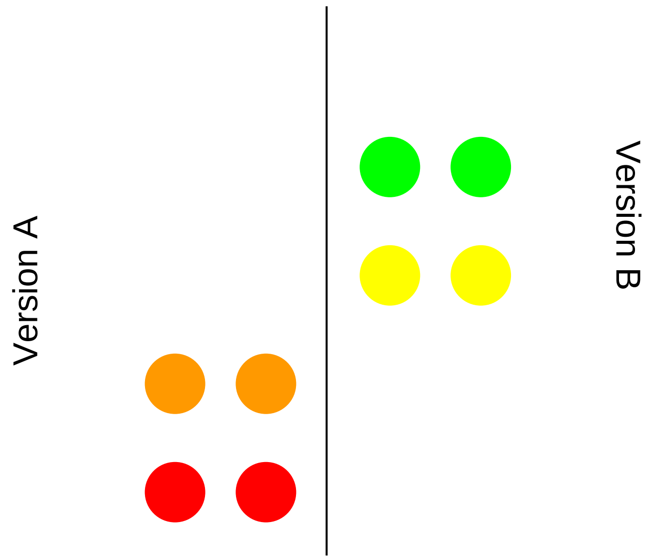

Figure 37 - Visualisation of Test 3 results for easier comparison

types of datapoints. Visualised results of the test can be seen in Figure 37.

Winner: Version B (with a large potential for improvement)

Task 4

The results for version A in task 4 indicated that this version did not have any serious critical issues. It rather had potentially confusing but also appealing visualisation (see Appendix – Figure 41). The thing that was bad in this visualisation were the numbers contained within. For multiple email campaigns those numbers were performing well in the visualisation, while they would make a confusion for a single campaign (there would be multiple number 1s – one for each fulfilled step, which would be confusing since it would make users think that more campaigns are included).

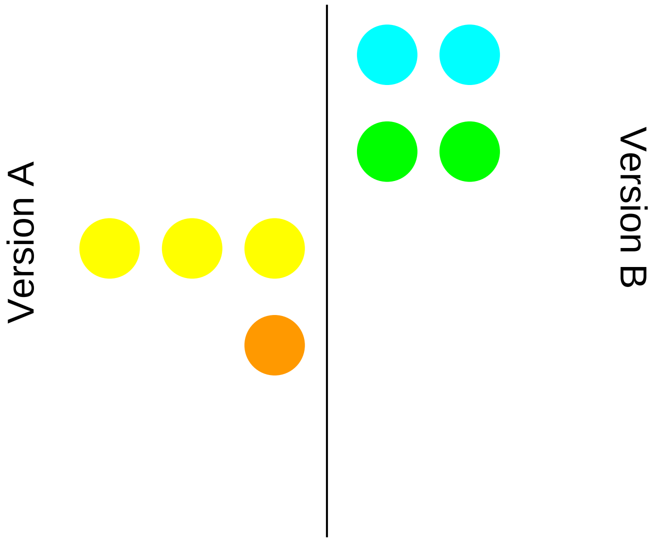

Figure 38 - Visualisation of Test 4 results for easier comparison

Results for the version B are showing that the suggested visualisation was correctly picked with all responses being either most positive or positive with some potential for improvement. This was mainly due to proper addressing of given data in a given context, separating visualisation for a single campaign and giving it a new one in a form of timeline. Visualisation performed well in encoding the email campaign as a whole entity, a process with several stages that need to be fulfilled. Participants recognized this design pattern and thus correctly interpreted the single email campaign. Visualised results of the test can be seen in Figure 38.

Winner: Version B

7.2 Impact of data visualisation

The previous section clearly indicated that version B was the dominating one in all of the four tests. While version B in test 3 showed some deficiencies and bigger potential for improving the visualisation in form of the pie charts, the other three tests proved their test versions were a suitable choice for visualising the given data. However, as said, the last section only gave a winner between two versions of A/B testing and which version should have won was not that hard to assume. The true question that underlies this choice is why the version B was the dominating one. Except for the fact that it introduced something new to the system, the test version used that new feature to convey the right information to the right audience in a rightful manner.

Three factors are significant here: the importance of the data being manipulated, how is it conveyed and the effectiveness, efficiency and satisfaction of the provided visualisation design.

The data present in both visualisations was found useful, because it was giving the metrics important for users of the system to perform their work (number of interactions, conversion rates, different statistics, etc.). That data was mostly in the quantitative form that was referring to the actions already done by the users and the actions that had to be done in order to satisfy certain goals. The data was already part of the system's database and pulled off the system's platform and therefore its quality was not questionable. The same data was used for making both versions of the system's feature which tells us that there was no difference from the data point of view and since it was the same, data was not the factor that determined the overall advantage and domination of the B version.

We have already seen how this data was visualised. Key variables were encoded visually to express the story of the data in a fruitful way. It was already seen in the chapter 3 that visualisation is very powerful since it speaks the visual language, which is the most dominating one in humans' interpretation of the environment.

In the Appendix 2 (see Appendix - Opinions about the visualisation) it is possible to see the respondents' answers and their overall opinions about both tested versions. From the respondents' answers about the A versions, it is possible to notice the strong presence of dissatisfaction in users' opinions. Two key problems can be inferred from the answers and those are time consumption and lack of data exploitation and utilization. It was clear that getting some information from version A would require a lot of time and effort to be put, from counting to calculating statistics. It was also clear that participants found the data useful, but did not have a proper way to explain that or to use that same data.

The winning visualisation, version B revealed why it was the dominating one. From the answers in the Appendix, it can be easily seen that mostly positive feedback was provided, with few exceptions. Two main advantages can be inferred from the answers. The first one is reducing user's time to perceive and rightfully interpret the data. The second one is enabling a user to make right decisions and perform necessary activities related to the issues or opportunities gained from the insights of the visualised data. However, this advantage is inferred only from the participants' answers (see Appendix – Opinions about the visualisation) and it was not measured separately. Therefore presents a research limitation. In short, two main benefits of these data visualisations were time-saving and expected better, real-time data-driven decision-making. These two benefits were a perfect fit for the issues found in the A version visualisations since they addressed and affected those issues completely. Therefore, it can be rightfully said users preferred the version B visualisations because they addressed the stated problems by reducing the time for understanding what the data had to say and by enabling the users to make data-driven decisions based on real insights and real-time data.

It had been inferred that data visualisation can significantly reduce user's time by quickly giving a needed overview in a visual format that a user can interpret more easily. Data visualisation also influences the user in a way it enables him to make his decisions faster and more reliable. Having a lot of these decisions (or micro-decisions) driven by data determines the course of the business itself in the end and sets a path to a more healthy and sustainable business.

7.3 Research limitations

I am aware that this research has some limitations which have to be mentioned. The first one is limitation in the number of the participants that took a part in testing the visualisations. A higher number of the participants would have given me a better overview of participants' opinions (due to a higher number of their answers) and therefore a more precise insight into the effects of the visualisations. This limitation was caused by the fact that it was challenging to find people working as sales managers, representatives (or someone closely related) without interrupting their schedule, both working and private. Because of this, research was conducted only on a small number of target audience. Second, another limitation that is also noticeable is the choice for tested visualisations – the old one against the new one. It would have been less predictable to have two new but different visualisations tested against each other. Third, a limitation that can be mentioned is the possibility that participants' answers were not completely honest due to various reasons. There is a possibility that they were just polite so their opinion about the visualisation was positive. What also needs to be mentioned is the fact that the interviews were conducted during participants' working hours. Thus it is also possible that their answers were positive just to finish the test as soon as possible since they had to go back to work (or since they wanted to save some time for their break). Final limitation is the one about second advantage for the users – better decision making. This advantage was only inferred from their answers (they said they could make certain decision) and it was not tested with the users to check which decisions are they making and how useful they are.

8. Conclusion

Looking back at the research question, we can ask did this research actually answer its question. Inferred from the results of the testing and their discussion, it can be stated that users have found the new visualisation more appealing, attractive, motivating and easier to use and interpret than the older one. Results clearly pointed out the winner of the A/B testing, having version B always winning over the A version with having a larger amount of positive points. When testing the new visualisation users managed to understand and answer most of the given questions. It can be said that the new visualisation satisfied users' needs and provided user satisfaction in almost all of the tests (Test 3, version B still has more space for improvement).

It has been seen that data visualisation influenced the users mainly in two ways: it reduced their time to understand the data and they thought it would affect their decision-making by enabling them to undertake certain actions based on real-time data. The new visualisation offered users an easier and quicker way to grasp the given data since it gave it a visual aspect, one that is easily understood by human beings. Influence of the data visualisation was also such as it made users' decision-making faster. Reducing the time spent on understanding the data and enabling more effective and efficient decision making greatly influences the user's performance and setting a path for healthy company growth and sustainability.

It can be stated that data visualisation definitely represents an important aspect in CRM users' work. It influences users' performance that essentially represents one of the company's key assets, working strength. Data visualisation relieves the company employees of their pains and boosts their gains and furthermore promises to deliver more productivity and overall growth in the end by providing efficient and effective information interpretation and user satisfaction.

8.1 Improvement and future work

The thing that could be improved in the tested visualisations is the design of some of them. While the second visualisation performed particularly well in its tasks and was attractive to the users, other visualisations showed some space for improvement. The first visualisation could be redesigned due to the answer provided by one participant, mentioning that having bars that stretch on both sides of the funnel chart is hard to compare – they should be only stretching on one side. The third visualisation is something that definitely needs to be improved and redesigned (one participant stated that she "does not necessarily like these charts"). That fact opens up a space for having a new visualisation designed instead of the pie charts. Also the fourth visualisation is up for a redesign since one participant mentioned that the "clicked" stage does not have to be fulfilled in order to progress to the next one.

In an ideal situation with enough manpower, time and money resources better testing could be conducted. More users would be tested and ratio of the A/B test user groups would be 50/50. Also, due to more time and money resources for the research, different visualisations would be created and the evaluation would be conducted on the two or more different visualisation versions instead of on the old visualisation versus the new one. During the testing phase, one person would be responsible for conducting the test (briefly explaining the users about the CRM system and leading them through the tasks) while another person would be responsible for writing down what the users are saying. That person would also be responsible for taking the notes about users' facial expressions and body language.

References

[1] C. Ware, Information Visualization: Perception for Design, Morgan Kaufman, 2000.

[2] D. R. Voss, “The Development of a Model for Nonverbar Factors Impacting The Design of Visual Information,” PhD thesis, Wayne State University, Detroit, 2008.

[3] R. Mazza, “Introduction to Information Visualisation,” Faculty of Communication Sciences, University of Lugano, Lugano, 2004.

[4] K. Storbacka, T. Strandvik and C. Grönroos, “Managing Customer Relationships for Profit: The Dynamics of Relationship Quality,” *International Journal of Service Industry Management,* vol. 5, no. 5, pp. 21-38, 1994.

[5] G. Ellis and A. Dix, “An explorative analysis of user evaluation studies in information visualisation,” *ACM,* pp. 1-7, 2006.

[6] K. Korać, “Impact of Data Visualisation on Users in CRM Systems,” Topic Proposal for the Master Thesis, University of Twente, June, 2016.

[7] K. Korać, “Data visualisation influence on a CRM system’s value proposition,” I&E Minor Thesis, University of Twente, EIT Digital Master School June 24, 2016.

[8] L. M. Given, The SAGE Encyclopedia of Qualitative Research Methods, Thousand Oaks, CA: SAGE Publications, Inc., 2008.

[9] K. French, “The Data Basics You Need to Know,” 3 December 2014. [Online]. Available: http://www.columnfivemedia.com/data-basics-need-know.

[10] “Numerical Data: What’s The Difference Between Discrete and Continuous?,” Australian Bureau of Statistics, [Online]. Available: http://www.abs.gov.au/websitedbs/CaSHome.nsf/Home/CaSQ+3B+NUMERICAL+DATA:+WHAT%27S+THE+DIFFERENCE+BETWEEN+DISCRETE+AND+CONTINUOUS.

[11] “Types of Variable,” Laerd Statistics, [Online]. Available: https://statistics.laerd.com/statistical-guides/types-of-variable.php.

[12] “What researchers mean by cross-sectional vs. longitudinal studies,” Institute for Work & Health, 2009.

[13] S. Few, “Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message,” 18 September 2014. [Online]. Available: https://www.perceptualedge.com/articles/ie/the_right_graph.pdf.

[14] P. Ghosh, “Cave paintings change ideas about the origin of art,” BBC News, 8 October 2014. [Online]. Available: http://www.bbc.com/news/science-environment-29415716.

[15] R. D. Guthrie, “Preface: Reassembling the Bones,” in *The Nature of Paleolithic Art*, Chicago, University of Chicago Press, 2006 .

[16] E. J. Beh and R. Lombardo, Correspondence Analysis - Theory, Practice and New Strategies, John Wiley & Sons, Ltd, 2014.

[17] S. Few, “Data Visualization for Human Perception,” Interaction Design Foundation, [Online]. Available: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception.

[18] M. Friendly, “Milestones in the history of thematic cartography, statistical graphics, and data visualization,” 24 August 2009. [Online]. Available: http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf.

[19] J. R. Beniger and D. L. Robyn, “Quantitative graphics in statistics: A brief history,” *The American Statistician,* p. pp. 1–11, 1978.

[20] W. S. Cleveland and R. McGill, “Graphical Perception and Graphical Methods for Analyzing Scientific Data,” *Science,* vol. 229, pp. 828-833, 1985.

[21] M. Soegaard, “Gestalt principles of form perception,” Interaction Design Foundation, [Online]. Available: https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/gestalt-principles-of-form-perception.

[22] R. Dewey, “"The Whole is Other than the Sum of the Parts",” [Online]. Available: http://www.intropsych.com/ch04_senses/whole_is_other_than_the_sum_of_the_parts.html.

[23] S. Bradley, “Design Principles: Visual Perception And The Principles Of Gestalt,” 28 March 2014. [Online]. Available: https://www.smashingmagazine.com/2014/03/design-principles-visual-perception-and-the-principles-of-gestalt/.

[24] H. Stevenson, “Emergence: The Gestalt Approach to Change,” Cleveland Consulting Group, Inc, [Online]. Available: http://www.clevelandconsultinggroup.com/articles/emergence-gestalt-approach-to-change.php.

[25] L. Mastin, “The Human Memory - Memory Encoding,” 2010. [Online]. Available: http://www.human-memory.net/processes_encoding.html.

[26] M. Dubakov, “Targetprocess - Visual Encoding,” September 2012. [Online]. Available: https://www.targetprocess.com/articles/visual-encoding/.

[27] B. &. Company, “Customer Relationship Management, Guide - Bain & Company,” 10 June 2015. [Online]. Available: http://www.bain.com/publications/articles/management-tools-customer-relationship-management.aspx.

[28] W. Reinartz, M. Krafft and W. D. Hoye, “The CRM Process: Its Measurement and Impact on Performance,” *Journal of Marketing Research,* 2004.

[29] B. L. Eisenfeld, E. Kolsky and T. Topolinski, “Gartner Survey: 42 Percent of CRM Software Goes Unused,” Gartner, 28 February 2003. [Online]. Available: https://www.gartner.com/doc/387369/gartner-survey--percent-crm.

[30] M. Rouse, “customer relationship management (CRM),” Searchcrm - Techtarget, November 2014. [Online]. Available: http://searchcrm.techtarget.com/definition/CRM.

[31] M. Rouse, “Sales force automation (SFA),” Whatis - Techtarget, January 2011. [Online]. Available: http://whatis.techtarget.com/definition/sales-force-automation-SFA.

[32] W. Seliga, “The Three Main Components of a CRM System,” SWC - Technology partners, 9 July 2012. [Online]. Available: https://www.swc.com/blog/customer-relationship-management/the-three-main-components-crm-system.

[33] C. Czarnecki, “The Sales Pipeline Revealed,” 22 June 2014. [Online]. Available: http://blog.pipelinersales.com/sales-pipeline-management/what-is-sales-pipeline/.

[34] C. S. Long, R. Khalafinezhad, W. K. W. Ismail and S. Z. A. Rasid, "Impact of CRM Factors on Customer Satisfaction and Loyalty," *Procedia Economics and Finance,* vol. 23, pp. 563-567, 2015.

[35] H. Swift, "How to Grow Customer Satisfaction with Your CRM," RingLead, 17 June 2014. [Online]. Available: https://www.ringlead.com/blog/crm-customer-satisfaction/.

[36] J. T. Bowen and S. Chen, "The relationship between customer loyalty and customer satisfaction," *International Journal of Contemporary Hospitality Management,* vol. 13, no. 5, pp. 213 - 217, 2001.

[37] R. S. Hassan, A. Nawaz, M. N. Lashari and D. F. Zafar, "Effect of Customer Relationship Management on Customer," in *2nd Global Conference on Business, Economics, Management and Tourism*, Prague, 2014.

[38] X. Liang, "CRM business cloud computing," in *International Conference on Innovative Computing and Cloud Computing, 2011*, Beijing, 2013.

[39] Z. Plaksij, "Why Cloud-based CRM Suits the Needs of Small Businesses," SuperOffice, 27 July 2016. [Online]. Available: http://www.superoffice.com/blog/cloud-based-crm-for-small-business/.

[40] J. Oetting, "A Simple Guide to Choosing the Right Chart for Your Data," HubSpot, 15 May 2015. [Online]. Available: http://blog.hubspot.com/marketing/data-visualization-choosing-chart.

[41] S. Coles and J. Rowley, "Creating effective graphs and charts," *Industrial Management & Data Systems,* vol. 97, no. 3, pp. 114-118, 1997.

[42] C. Finch, "Advantages & Disadvantages of a Pie Chart," [Online]. Available: http://www.ehow.com/list_6715678_advantages-disadvantages-pie-chart.html.

[43] "Sales Representative Job Description," Snagajob, [Online]. Available: http://www.snagajob.com/job-descriptions/sales-representative/.

[44] R. Blewett, "The Importance of Data Visualization to Business Decision Making," 12 June 2011. [Online]. Available: http://multipeers.itpeers.com/fotos/gca/the_importance_of_data_visualization_to_business_decision_making_83713439550bf20f8eaddf.pdf.

[45] A. Shen-Hsieh and M. Schindler, "Data Visualization for Strategic Decision Making," *Case Studies of CHI2002*, ACM, 2002 .

[46] J. Haight, "How Data Visualization Empowers Decision Making and Who Is Getting Us There," 13 October 2014. [Online]. Available: http://bluehillresearch.com/how-data-visualization-empowers-decision-making-and-who-is-getting-us-there/..

[47] "HTML," Mozilla Development Network, 1 July 2016. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTML.

[48] "JavaScript," Mozilla Development Network, 6 June 2016. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/JavaScript.

[49] "What Is Angular?," AngularJS, [Online]. Available: https://docs.angularjs.org/guide/introduction.

[50] "SVG," Mozilla Development Network, 9 February 2016. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/SVG.

[51] M. Bostock, "Data-Driven Documents," Data-Driven Documents, 2015. [Online]. Available: https://d3js.org/.

[52] M. Bostock, V. Ogievetsky and J. Heer, “D3 Data-Driven Documents,” *IEEE Transactions on Visualization and Computer Graphics,* vol. 17, no. 12, pp. 2301-2309 , 2011 .

[53] J. M. Sonberg, “How to Build a Strong A/B Testing Plan That Gets Results,” 2014. [Online]. Available: http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results/.

[54] U. Bar-Joseph, “8 Rules of A/B Testing – The Art in Marketing Science,” Search Engine Watch, 6 August 2014. [Online]. Available: https://searchenginewatch.com/sew/how-to/2196754/8-rules-of-a-b-testing-the-art-in-marketing-science.

[55] J. Nielsen, “Interviewing Users - Nielsen Norman Group,” 26 July 2010. [Online]. Available: https://www.nngroup.com/articles/interviewing-users/.

[56] J. Ludeman, “Edward Tufte: The DaVinci of Data,” 22 February 2016. [Online]. Available: https://sites.psu.edu/julieludeman/2016/02/22/edward-tufte-the-da-vinci-of-data/.

[57] S. Card, J. Mackinlay and B. Shneiderman, Readings in information visualization: using vision to think, San Francisco: Morgan Kaufmann Publishers Inc., 1999.

Appendix – Summary of users' answers (encoded)

Asked questions	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 1 - Funnel chart	A raw table	Interactive funnel chart vis.
1. What data can you tell us from this presentation?	All the respondents recognized only information about basic lead information and current status. No conversion rate was recognized here.	All the respondents recognized sales conversion rates and their pipeline progress in the visualisation, among the other information contained within the table like basic lead information, current status, etc.
2. Could you or how fast could you tell us how many leads were in a specific status?	If the number of contacts was relatively low, they could count them; otherwise it is impossible task; ~ *"I would clearly be inconvenient to count all of them"* - participant nr. 2	Respondents easily answered this question by reading the data from the chart. Three of them needed brief assistance in understanding the flow of the leads; in the end managed to perceive the right information.
3. Could you tell us what was the progress of the leads over the stages in last month?	None of the respondents managed to answer this question.	All the respondents managed to read and comprehend progress of the number of leads over their statuses in pipeline.
4. Could you tell us the conversion rates over the stages?	As before, since they weren't able to comprehend the leads' progress over time, respondents were constrained from interpreting conversion rates over stages.	All respondents successfully perceived and interpreted this information. Some of them had to be provided a bit of extra guidance, but in the end comprehended the information completely.
5. Could you make a comparison of conversions and conversion rates of currently presented time span with some other one?	No respondents was able to answer this question.	All of the respondents managed to answer the question by triggering the comparison switch and comparing current and past time span.
6. What is your opinion about the data and the visualisation? What do you think of it in terms of being helpful for your work?	Respondents found this information useful, but would prefer some manipulations with the data for it to be useful.	Data was also found useful for all respondents. Visualisation is appealing and useful in terms of saving time and making better decisions; ~ *"This would save me a lot of time"* - participant nr. 3

Table 5 – Summary of respondents' answers for Test 1

Asked questions	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 2 - Line & Stacked bar chart	A raw table	Interactive line and column charts (straight lines and labeled columns)
1. What data can you tell us from this presentation?	Respondents recognized the number of interactions over time; saw only simple numbers. 🟠	Respondents discovered certain patterns and behaviour of number of interactions over time. 🟢
2. Could you track the progress of your data easly?	The answer was not positive since the respondents were presented by mere numbers - it was difficult for them to track the progress quickly. 🔴	All the respondents answered positively; it was easy for them to track the progress of the data. 🔵
	Additional curved line chart and empty column chart	
3. Could you tell us what is happening with the interactions in Week 2 of May till the 9th in the line chart (under curved line) and from Sat 5 Dec till Sun 6 Dec?	Respondents detected slight rise and then fall (and vice-versa) of the numbers of interactions even though they are consistent. 🟠	Respondents perceived the progression of the data accurately. 🔵
4. Could you tell us how many interactions are there per one column (individually and together)?	Respondents managed to answer but had slight delay to scale the columns. 🟢	With labeled columns, all of the respondents answered the question in a glimpse. 🔵
5. What is your opinion about the data and the visualisation? What do you think of it in terms of being helpful for your work?	Table presentation was useful, but certainly not enough; other visualisation was useful complement but needs some redesign in order to completely convey right information ~ *"The data I see here is useful for me, but it's hard to understand the big picture from the table"* - participant nr. 1 🟡	Complementing the table, this types of charts were very useful (according to the participants), appealing and time saving and enabled more efficient decision making. Because of the visualisation, data was found more useful by most of the users. ~ *"Seeing this data visually I am actually discovering its potential"* - participant nr. 7 🔵

Table 6 – Summary of respondents' answers for Test 2

Asked questions	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 3 - Pie charts	A raw table	Interactive pie charts
1. What data (about the datapoints) could you tell us from this visualisation?	By reading the datapoint column, participants were able to perceive how many datapoints are available for one contact but couldn't find out which datapoints (since abbreviations were used). 🟠	Participants recognized the number of available datapoints for a single contacts and their names. 🟢
2. How many datapoints are there? How many of contacts have active datapoints?	Having a large number of contacts, it was impossible for them to count the distribution of datapoints; the task became time consuming and inconvenient. 🔴	Participants easily answered this question by reading the visualisation - it was possible to see the total number of datapoints and how many of them were active. 🟢
3. Could you tell us, out of all active ones how many, what is the presence of all datapoints? How many active datapoints are present (per type)?	Even with a low number of contacts, it was inconvenient for the participants to answer this question since it required data aggregation and then additional operations (mean). 🔴	All the respondents managed to read and comprehend numbers and percentage of active datapoints and datapoints per type. 🟡
4. What is your opinion about the data and the visualisation? What do you think of it in terms of being helpful for your work?	Seems useful, but participants could have extracted more inside information and gain more insight. Since it was giving additional info about the leads, the data itself was found useful but respondents would have liked to see it in real context. 🟠	Visualisation seems useful, providing easy access to aggregated information (time saving and decision making oriented). Due to this reason, data was also found useful for the participants - however, one participant was not satisfied with the visualisation (wanted different visualisation for this useful data) ~ *"I think the data is useful, but I believe some other visualisation would be better"* - participant nr. 6 🟡

Table 7 – Summary of respondents' answers for Test 3

Asked questions	A feature (control) - old visualisations	B feature (test) - new visualisations
Test 4 - Email (single) campaign timeline	Visualisation in form of unconnected rings	Static timeline with dates
1. What data could you tell us from this visualisation?	Participants were able to perceive how many emails are in which stage (single campaign - only 1). 🟡	Participants recognized the progress of the email single campaign and in which state is it now and which ones were completed and when. 🔵
2. How many emails are we talking about here?	One email, but potentially more. 🟡	Respondents perceived easily that visualisation is related only to the single campaign. 🟢
3. What is the current status of the campaign?	Participants answered that one sent, one was received (which created confusion in their perception about single campaign). 🟠	All participants easily perceived current campaign status and put it within a time context. 🔵
4. What is your opinion about the data and the visualisation? What do you think of it in terms of being helpful for your work?	Participants found this visualisation useful for having multiple campaigns, but can be confusing for a single campaign. Therefore, there was a confusion in what kind of data is the visualisation conveying. ~ *"Why is there one sent and one open mail if it's only a single campaign? Oh wait.. "* - participant nr. 1 🟡	Participants found the visualisation attractive and useful for it saves their time and enables them to meet their leads more by observing time between different steps in email campaign (i.e. how long does it take for lead to answer the email). 🟢

Table 8 – Summary of respondents' answers for Test 4

Opinions about the visualisation

To determine how the visualisation actually helps the users of the system, their opinion was asked as the final question of the interview. Key snippets were taken from their answers and listed below, to provide help in determining the "influence" of data visualisation on users' performance.

Opinions about the visualisations are following:

Test 1

Version A: 🟠

- Participant nr. 1 - *"Tabular presentation looks good, but I would like to see something done with it."* – encoded: 🟡
- Participant nr. 2 - *"If I want to extract more information from this data, like when you asked me to count down the statuses, this presentation becomes totally poor and furthermore time consuming. It seems to me this data potential is not fully used with this presentation."* – encoded: 🟠

Version B: 🟢

- Participant nr. 3 - *"I like this presentation. I find it readable and easy. This would save me a lot of time."* – encoded: 🔵
- Participant nr. 4 - *"It's good because it lets you see the progress you have and which is the breaking point in the process (sales pipeline). In that way you can tell your reps to work more on this point."* – encoded: 🟢
- Participant nr. 5 - *"It would definitely be more helpful to get more detail into this. I like the chart. I don't necessarily like that is goes to the both ends (funnel chart), makes it difficult to compare, but it's definitely the best option to represent these numbers."* – encoded: 🟢
- Participant nr. 6 - *"This (visualisation) definitely provides a useful insight. I can see which conversion rate is the lowest and undertake certain actions to raise the numbers."* – encoded: 🔵
- Participant nr. 7 - *"Seeing this I don't need to count all the statuses and I have the progress of my leads over time easily presented. This is time saving and I can focus more on what do I need to do next."* – encoded: 🔵
- Participant nr. 8 - *"I like the chart, it's nice and appealing. I didn't see the conversion rates (on the right) immediately, maybe that could be highlighted somehow, but seeing them now I can see my progress and focus on some parts that are underperforming."* – encoded: 🟢
- Participant nr. 9 - *"Definitely a good visualisation, gives me an insight about rates and corresponding problems while saving time. Why should I count the statuses myself? System should do it for me and I should focus on what comes after I see that data."* – encoded: 🟢

Test 2

Version A:

- Participant nr. 1 - *"The data I see here is useful for me, but it's hard to understand the big picture from the table."* – encoded:
- Participant nr. 2 - *"Table is good for me to see what happened on which day by which user, but of course there is more potential here. This new presentation (curved line chart and no-label bar chart) is way to go for me, but would need some redesign not to confuse me (curved line)."* – encoded:

Version B:

- Participant nr. 3 - *"This is needed here, tracking the interactions over time is needed and these charts are doing that well."* – encoded:
- Participant nr. 4 - *"Nice charts, I like the design and labels in bar charts. It can give me the overview of the data really quickly."* – encoded:
- Participant nr. 5 - *"The table is good for the details, but if I only need an overview, this (the charts) is what I would prefer."* – encoded:
- Participant nr. 6 - *"With this one (line chart), I can easily compare between different users and their performances. That can tell me do I need to stimulate someone to work more efficiently or do I need to award someone for their effort."* – encoded:
- Participant nr. 7 - *"Seeing this data visually I am actually discovering its potential. Data in the table is useful when I want to see the exact number of a particular users for a particular period. However with the visualisation, time for me to understand the data is vastly reduced."* – encoded:
- Participant nr. 8 - *"I like these chart because they can show me comparison between my users while table cannot do that - it's good only for details."* – encoded:
- Participant nr. 9 - *"I can see how my users are making progress and easily draw a comparison between them. In that way I can see if someone is falling behind so I can motivate him."* – encoded:

Test 3

Version A:

- Participant nr. 1 - *"Like I said before (in Test 1), it's good to see things stored in form of spreadsheet, but this could be exploited."* – encoded:
- Participant nr. 2 - *"Ok, this is similar to that test before (Test 1). When I want aggregated data, it becomes hard.. No, impossible for me to count down all the datapoints and calculate the statistics if I want to see them. And think I would like to see some."* – encoded:

Version B:

- Participant nr. 3 - *"As in the other examples (tests) this is obviously time saving since I don't have to count datapoints by myself. However, I think this visualisation can still be worked on."* – encoded:

- Participant nr. 4 - *"Pie charts can show me the breakdown of datapoints by their activeness and by their type. If I see that lot of my leads share certain datapoints, depending on a type of datapoint I could adjust my effort to that group of leads."* – encoded:
- Participant nr. 5 - *"The data this chart is presenting is obviously useful, but I don't necessarily like these charts."* – encoded:
- Participant nr. 6 - *"I think the data is useful, but I believe some other visualisation would be better."* – encoded:
- Participant nr. 7 - *"This visualisation would save me time."*
- Participant nr. 8 - *"It's good for me to see these charts, especially the one with active/inactive datapoints. That tells me which leads we still haven't reached yet and therefore we must contact."* – encoded:
- Participant nr. 9 - *"The percentage of inactive datapoints can indicate that I haven't contacted these leads and sets my task. This chart can also show me if there are a lot of those leads whose datapoints' statuses are not so attractive so I can focus my effort somewhere else."* – encoded:

Test 4

Version A:

- Participant nr. 1 - *"Not quite sure about this. Why is there one sent and one open mail if it's only a single campaign? Oh wait.. "* – encoded:
- Participant nr. 2 - *"This presentation is good if I would like to see multiple email campaigns, but for a single campaign these numbers (ones) are bit confusing."* – encoded:

Version B:

- Participant nr. 3 - *"In these single email campaigns, it obvious that those steps always include the previous ones. With this visualisation you can easily see that."* – encoded:
- Participant nr. 4 - *"I think this visualisation is good because it gives your email campaign time context, for example when was your email opened by the customer, and it tells you what the current status of your campaign is."* – encoded:
- Participant nr. 5 - *"I like the fact that I can see when something happened with the email. That information is pretty good conveyed in this timeline."* – encoded:
- Participant nr. 6 - *"This visualisation is nice but beware of this stage ('clicked') since the customer does not have to click on a link in the email in order reply (go to the other stage)."* – encoded:
- Participant nr. 7 - *"Gives me quick look what has happened when with the email and what is happening now."* – encoded:
- Participant nr. 8 - *"Not sure whether 'clicked' stage should be there, but the timeline is really good presentation of a single email campaign."* – encoded:
- Participant nr. 9 - *"This timeline is a good representation of this campaign because the campaign is set in time and this visualisation is giving us a nice overview."* – encoded:

Appendix 3 - Participants' metadata:

- User 1:
  - Age: 25
  - Gender: Female
  - Education: BSc of Marketing
  - Profession: sales representative
  - Nationality: Argentinian

- User 2:
  - Age: 28
  - Gender: Female
  - Education: MSc of Advertising
  - Profession: sales representative
  - Nationality: Romanian

- User 3:
  - Age: 21
  - Gender: Female
  - Education: BSc of Science
  - Profession: sales representative
  - Nationality: German

- User 4:
  - Age: 23
  - Gender: Male
  - Education: MSc of Computer Science
  - Profession: developer
  - Nationality: French

- User 5:
  - Age: 30
  - Gender: Female
  - Education: BSc of Management
  - Profession: sales manager
  - Nationality: Brazilian

- User 6:
  - Age: 25
  - Gender: Male
  - Education: BSc of Economics
  - Profession: sales representative
  - Nationality: German

- User 7:
  - Age: 26
  - Gender: Male
  - Education: MSc of Computer Science
  - Profession: developer
  - Nationality: German

- User 8:
  - Age: 22
  - Gender: Female
  - Education: BSc of Management
  - Profession: product manager
  - Nationality: French

- User 9:
  - Age: 25
  - Gender: Female
  - Education: MSc of Marketing
  - Profession: sales representative
  - Nationality: Polish

Appendix 4 - Leadscore CRM - Old Visualisations:

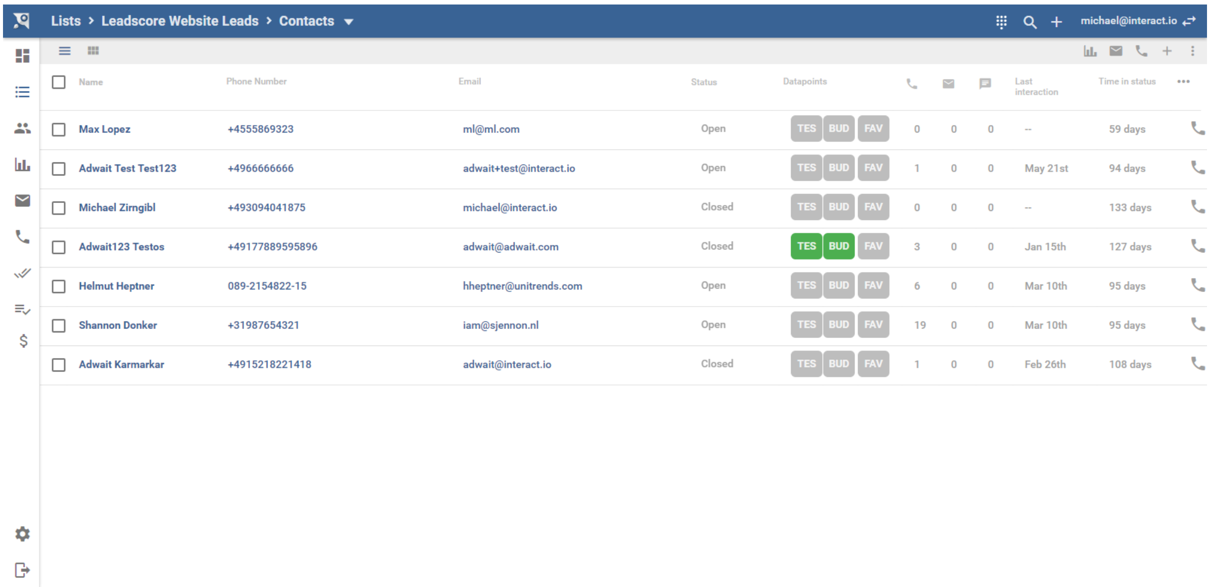

Figure 39 - The old visualisation for both Test 1 and Test 3 (screenshot: Leadscore)

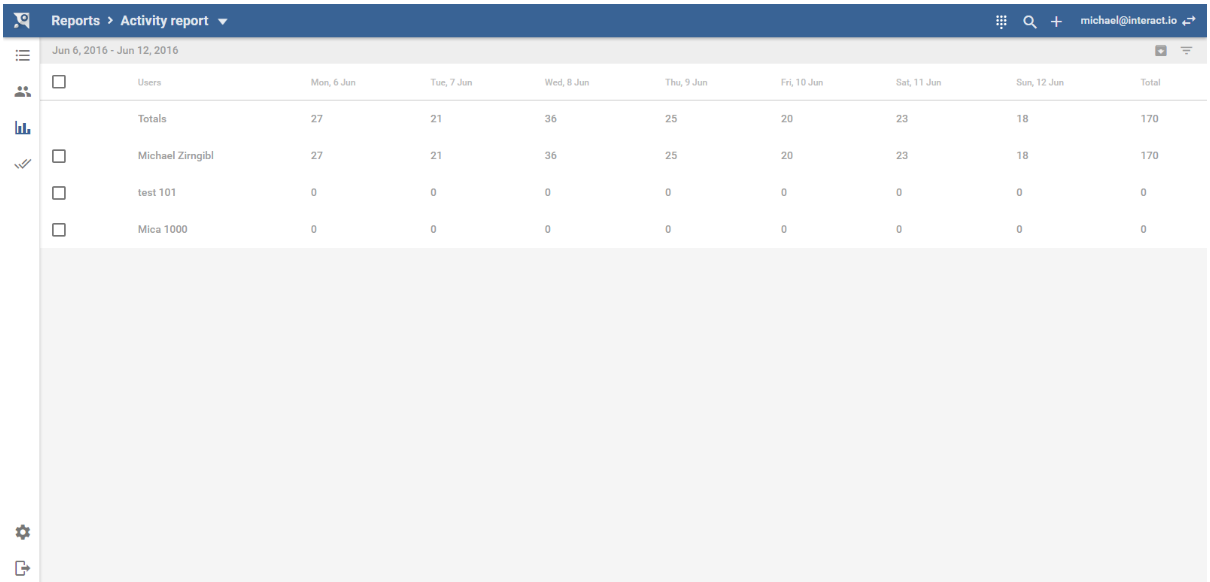

Figure 40 - The old visualisation for Test 2 (screenshot: Leadscore)

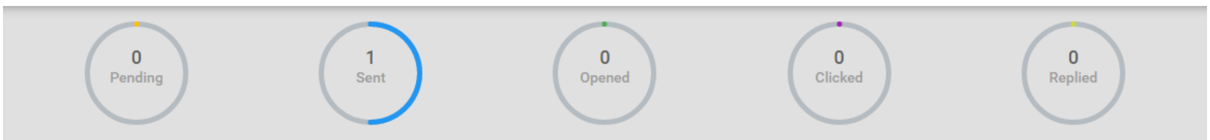

Figure 4139 - Email campaign (old) visualisation in form of unconnected rings (screenshot: Leadscore)