



**Trinity College Dublin**  
Coláiste na Tríonóide, Baile Átha Cliath  
The University of Dublin

# Interactive Data Visualisation

Applying an interactive data visualisation to Irish road accident data  
to study usability and memorability.

Alessa Gross

A research Paper submitted to the University of Dublin,  
in partial fulfilment of the requirements for the degree of  
Master of Science Interactive Digital Media

Supervisor: David Kaneswaran

May, 2015

# Declaration

I declare that the work described in this research Paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: \_\_\_\_\_

Alessa Gross  
14<sup>th</sup> May 2015

## Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this research Paper upon request.

Signed: \_\_\_\_\_

Alessa Gross  
14<sup>th</sup> May 2015

# Acknowledgements

I would like to thank my supervisor David Kaneswaran who proved to be the pillar of strong support and regular advice that I needed to write this research paper.

I also wish to thank my classmates in this year's IDM course who turned out to be invaluable through their continuous support, motivating words and the unforgettable moments together with them.

I would also like to thank my parents who supported me in taking up this course in the first place.

# Abstract

The digital revolution in recent years has led to various data gathering devices and services which have become common in everyday life. It also has led to the big data discussion that deals with the questions of what to do and where to go with these information in numbers and text? One approach is to make sense of the incomprehensible amount of data in a human-comprehensible way through informative visualisations of the data. This has recently contributed to various different studies on the data visualisations and especially interactive data visualisations. Research in interactive data visualisation touches, amongst others, the fields of computing, design and human computer interaction. However, it is still a relatively young field of research and hence the formula to quickly find the best visual representation for every data set has yet to be found. Therefore the full range of features that make a good visualisation, and enhance comprehension and memorability of the presented information cannot be clearly defined.

For this study, an interactive data visualisation has been created that illustrates road accidents in Ireland in 2010. The design decisions for it were based on the findings of current research. The visualisation focuses on a compact representation and easy access of the presented information. It was implemented with modern web- and visualisation tools.

In an experiment the visualisation was compared to another interactive data visualisation and a static visualisation that are both based on the same data set. The experiment investigated the response time and response accuracy of participants that were asked to answer specific questions with the help of one of the three visualisations. It also tested the participant's memorability on information they got from the visualisations. Furthermore, the participants were asked for their personal perception of the visualisations. The results of this experiment showed a relationship between a quick response time and a low memorability as well as a high usability and a good user perception. The results of the experiment of this work add to the clarification of the above named questions on the features of interactive data visualisations.

# Table of Contents

<b>Declaration</b> .....	<b>II</b>
<b>Permission to lend and/or copy</b> .....	<b>III</b>
<b>Acknowledgements</b> .....	<b>IV</b>
<b>Abstract</b> .....	<b>V</b>
<b>Table of Contents</b> .....	<b>VI</b>
<b>List of Figures</b> .....	<b>VII</b>
<b>List of Tables</b> .....	<b>VII</b>
<b>List of Code Snippets</b> .....	<b>VII</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>2 State-of-the-art</b> .....	<b>3</b>
2.1 The definition of data visualisation .....	3
2.2 Data Visualisation Frameworks .....	4
2.3 Interaction .....	5
2.4 Storytelling .....	7
2.5 A good data visualisation .....	10
2.6 Memorability .....	11
<b>3 Interactive Data Visualisation Design &amp; Implementation</b> .....	<b>13</b>
3.1 Design Decisions.....	14
3.2 Implementation and Tools .....	18
3.2.1 JSON .....	18
3.2.2 Google Maps API and the MarkerClusterer .....	19
3.2.3 Google Fusion Tables .....	21
3.2.4 D3.js.....	23
<b>4 Usability and Memorability Tests</b> .....	<b>25</b>
4.1 Experiment Set-up.....	29
4.2 Questions .....	30
4.2.1 Introductory questions.....	30
4.2.2 Main questions.....	30
4.3 Participants.....	32
4.4 Results and Evaluation .....	32
<b>5 Conclusion</b> .....	<b>40</b>
5.1 Future Work .....	40
<b>Abbreviations</b> .....	<b>43</b>
<b>References</b> .....	<b>44</b>
<b>Appendix A</b> .....	<b>47</b>
<b>Appendix B</b> .....	<b>53</b>

## List of Figures

Figure 1: “PlaneTruth” Interactive data visualisation .....	6
Figure 2: “But Will It Make You Happy?” Interactive data visualisation.....	7
Figure 3: Napoleons March to Moscow, early data visualisation .....	8
Figure 4: Comparison of single-frame (a) and multi-frame (b) interactivity.....	9
Figure 5: “Monstrous Costs” Data visualisation.....	12
Figure 6: “Road accidents: Ireland, 2010.” Interactive data visualisation .....	13
Figure 7: Visualisation of a parent node and two child nodes in a sunburst chart.....	15
Figure 8: The province of Munster highlighted in the map .....	17
Figure 10: Grid-based clustering .....	20
Figure 11: Google Maps Polygon .....	21
Figure 12: Google Maps Circles.....	22
Figure 13: Accuracy of the polygons from the Natural Earth Fusion Table .....	23
Figure 14: All road collisions in 2010.....	25
Figure 15: Serious road collisions in 2010.....	26
Figure 16: The first 25 entries of the spreadsheet visualisation .....	27
Figure 17: The filter and sort option of the spreadsheet visualisation.....	28
Figure 18: Label informing about the number of filtered results in Microsoft Excel 2010.....	33
Figure 19: Response accuracy results .....	34
Figure 20: Response time results .....	35
Figure 21: Memorability results.....	37
Figure 22: User perception results.....	38
Figure 23: Cluster marker icon suggestion.....	41

## List of Tables

Table 1: Comparison of information coverage in the visualisations of the experiment.....	29
Table 2: Questions of the usability and memorability experiment .....	31

## List of Code Snippets

Code 1: Accidents 2010 JSON object example .....	18
Code 2: Details of County Dublin JSON object example .....	19

# 1 Introduction

Data visualisation, and specifically interactive data visualisation as a research field, is a relatively young discipline and can be seen as a result of the big data discussion of recent years. It has evolved hand in hand with the revolution of data gathering services and devices that have become common in everyday life. A thorough analysis of the causes, the means and the risks of big data go beyond the scope of this work. However, it should be stated that the need for making use and sense of the large amount of digital information in text and numbers forms the basis of the whole discussion. Yet, it is not straightforward to make it accessible in a human comprehensible way; “big data is a resource and a tool. It is meant to inform, rather than explain; it points toward understanding, but it can still lead to misunderstanding, depending on how well it is wielded.” (Cukier & Mayer-Schoenberger 2013).

Hence, the urge to make use of the data available, has led to the development of appropriate tools that ease the process of understanding this information. Digital data visualisations that can be updated to the most recent datasets in milliseconds are the most common examples, for example.

Different definitions of the term ‘data visualisation’ by different authors are explained in Chapter 2.1. Based on these slightly varying explanations, the following definition shall be kept in mind, while reading the present paper: Generally, the term data visualisation can be summarised as the process of putting data into a visual context in order to improve its comprehensibility. This is often reached by finding different representative visual objects for the data objects, such as rectangles in bar charts or polygons in pie charts, and putting them in relation to each other in a two-dimensional or three-dimensional space.

This is especially true when these data visualisations are opened up for interaction with the user. Then they prove to be of essential help in making sense of large data sets. Through interaction data visualisations can be manipulated to suit the particular needs of a user, and relationships between the illustrated information can become even clearer to the user than in static visualisations.

Current discussions in this area of research include methods to measure the usability of a data visualisation as well as the impact of adding interactivity to it. Along with this paper, an interactive data visualisation on road accidents in Ireland has been created based on the results of research on visual design, composition, user experience and storytelling techniques. The response accuracy and response time of users engaging with the visualisation as well as their ability to remember facts from the visualisation and their subjective opinion were tested in comparison to comparable established visualisations. These tests were carried out in order to research, which features of a visualisation improve the response accuracy, response time or memorability. The results of this paper will



contribute to the clarification of the advantages and disadvantages of the use of interactive data visualisations on road accidents in Ireland.

This introduction has discussed the motivations behind this study and has given a brief overview of the experimental method that follows. Chapter 2 will examine related work and fundamental knowledge on the topic of data visualisations. Chapter 3 discusses the development, design decisions and implementation of the interactive data visualisation named “Road accidents: Ireland, 2010” created specifically for this study. Following this, Chapter 4 describes the experiment undertaken using this purpose built visualisation and discusses the results. The paper then concludes in Chapter 5 by discussing to the results of the experiment giving a deeper insight into the field of interactive data visualisation. It also gives an overview of possible future work in this field.

As it is difficult to fully capture the dimensions of an interactive visualisation in a static representation such as screenshots and text, the reader may want to view the “Road accidents: Ireland, 2010” visualisation, before continue reading. The website is available here, the credentials for a Trinity College School of Computer Science and Statistics (SCSS) login will be required:

[www.scss.tcd.ie/~grossa/ResearchProject/DataVis/](http://www.scss.tcd.ie/~grossa/ResearchProject/DataVis/)

For public viewing a short screen capture illustrating a common use case of the visualisation can be seen here:

<https://www.youtube.com/watch?v=MRi-Pyv5hAs>

## 2 State-of-the-art

As stated data visualisation is a research topic only since recent years. Hence, the formula to find the best way to present data quickly and most comprehensible has yet to be found. This chapter presents the current research in data visualisation, as well as considering its related parts within the big data discussion and the research field of user experience. This includes the clarification of the definition of the term in the first subchapter. Afterwards a brief insight into early data visualisations is given that discusses their purpose and their use. This is followed by a summary of frameworks that have been proposed to help in the process of creating helpful data visualisations. Furthermore, the research on interactivity in data visualisations that feeds into the field on human computer interaction is being analysed. After this a relative new idea of introducing storytelling techniques to data visualisation is presented. Finally, this chapter sums up with an introduction on the psychology of visualisations, how to improve memorability of information and factors that influence their perception.

### 2.1 The definition of data visualisation

What is the definition of data visualisation? This question has been attempted to be answered a number of times and needs to be reviewed on a regular basis as tools, methodology and theories around data visualisation are constantly evolving. In Chapter 1 a general definition is given that forms the basis of this paper and that was based on the following research.

Manovich (2002) described visualisation as the process, “when quantified data which by itself is not visual [...] is transformed into a visual representation”. However his definition excludes data that is already of a visual nature in its raw form. A detailed definition needs to consider that this data can be (re)arranged into another visual representation as well in order to make its content more accessible for the viewer. Later Manovich observed this gap in his definition and clarified the aspects of the various meanings and applications of visualisation (2011). He introduces the approach of “direct visualisation” for data that is visual in its raw form<sup>1</sup>. Different to most visualisation techniques, this approach uses thumbnails of the raw data as a visual representation and does not replace it by another visual object, such as a rectangle. But not only shapes can represent a data record, colour and especially the hue of a colour can be used for visualisations (Driscoll 2010). Roberts *et al.* take both visual and non-visual raw data into account when stating that generally the most successful visualisations “[map] data to an appropriate visual form” (2014). This study will later use a CSV document containing raw Irish road collision data. Even though the raw data used for the visualisations

---

<sup>1</sup> In Manovich’s example the thumbnails (small digital representations of images), of the front covers of the Times magazine are arranged next to each in a way that illustrates the development of their design over the years.

of this work (see Chapter 3), is not visual material itself, but numbers and text, the considerations named above grant a certain thought process in finding the appropriate visual form and therefore need to be taken into account when designing any data visualisation.

Every attempt to find an appropriate definition leads inescapably to the question; what is the aim of data visualisation? “The purpose of visualization is insight, not pictures. [...] The term insights makes clear that we are discussing a human experience.” (Shneiderman 2008, p.3) is one direct answer to this question. While revising his definition, Manovich broadens his definition by declaring the aim is to “make visible patterns and relations in the data” (2011). They give meaning to big sets of data and make them “comprehensible and accessible to a broader range of people” (CACM Staff 2014). In the words of Langseth data visualisations “[provide] beautiful, easy-to-use interfaces that lie on top of massive amounts of number crunching from a growing array of sources” (as cited in CACM Staff 2014).

A well-established and often cited definition by Card *et al.* summarises both the method and the aim of data visualisation. In their words data visualisation refers to the “use of computer-supported, interactive, visual representations of data to amplify cognition, or the acquisition and use of knowledge” (1999).

## 2.2 Data Visualisation Frameworks

The process of obtaining data visualisations can be generalised to provide a framework that can be worked through when creating new visualisations. In short the data visualisation designer explores the given data, “selects the subset(s) [and] selects the presentation format(s) [that are] appropriate for [the] users”(Janvrin *et al.* 2014).

These steps were analysed in detail by Fry along with the development of the programming language “Processing” (2004). He proposed a framework that approaches the visualisation of large data sets in a holistic manner by borrowing from the toolsets of computer science, statistics, design and human computer interaction. His scheme is made of the seven stages of acquiring, parsing, filtering, mining, representing, refining the data and making it interactively accessible (2008). These stages might not seem novel itself, but rather like an established good-practice approach. However, Fry gives the different parts involved in the process of producing useful data visualisation a name and he describes it in a methodical way.

Similar to this McCandless approaches data sets for his visualisations that he publishes both on his website (2015a) and his books (2009; 2014). He even includes a data visualisations influences on a viewer after its read in a six steps process from unstructured data to “inter-connected knowledge”

(2014, p.14). First data is “gathered, measured [and] mined” (2014, p.13), which is then “examined, recognised [and] classified” (2014, p.13) in order to get structured data sets. The next step is to “filter, interpret [and] arrange” (2014, p.13) the unstructured data to obtain the information in the data. The fourth step includes to “connect, sequence [and] condense” (2014, p.14) the information to get “linked information” (2014, p.14). This step is the last step that can be controlled by the designer, everything that after this step is depending on the viewer and on the degree of quality that went into the four steps before. In the fifth step the viewer needs to “evaluate, understand [and be able to] explain” (2014, p.14) the linked information, which can be then called knowledge. In the last step the viewer needs to be able to “integrate, extrapolate [and] generate” the new knowledge and prior knowledge, then the data / information of the visualisation is part of an inter-connected knowledge” (2014, p.14).

These frameworks were used in order to approach the data set that forms the basis of this work. As stated by Janvrin *et al.* every step in any framework must be undertaken while considering the principles (2014). Firstly the visualisations should not overwhelm the viewer with too much information, while including enough data in order to inform the user, tell an engaging story and help with decision making. Secondly, the user’s knowledge of both the contents domain and the domain of data visualisation, needs to influence every design decision.

## 2.3 Interaction

To comprehend the full dimensions of the techniques used in this study here, the impact of applying interactivity to data visualisation must be clarified. Traditionally static visualisations illustrate preselected parts of the full amount of available data in preselected visual forms. The viewer is not able to change the format or filter the visualised data. In contrast, for interactive data visualisations the designer selects a larger range of data from the full data set and provides more than one alternative visual form of the data for the viewer (Janvrin et al. 2014). The viewer takes on an active role and becomes a user.

As an approach to the design of interactive data visualisations Shneiderman coined the “Visual Information Seeking Mantra: Overview first, zoom and filter, then details-on-demand” (1996, p.337). Meaning that a visualisation should show a well extracted summary of the data at first sight, while offering the option to zoom or filter the data. Every view should offer an option to show the details of the data, if requested (details-on-demand). Shneiderman further proposes that these design methods such as zooming, filtering or even aggregating data in markers that represent a number of records will become more and more important as data sets grow and screens shrink (2008).

An example of an interactive data visualisation titled “Plane Truth” (McCandless 2015b) is illustrated by a screenshot in Figure 1. As a data journalist, McCandless regularly incorporates the most recent research findings in the field of interactive data visualisations in his designs. The visualisation follows strictly Shneiderman’s approach as it displays an overview over the plane accidents, while offering filtering options and details-on-demand, when hovering over a square.

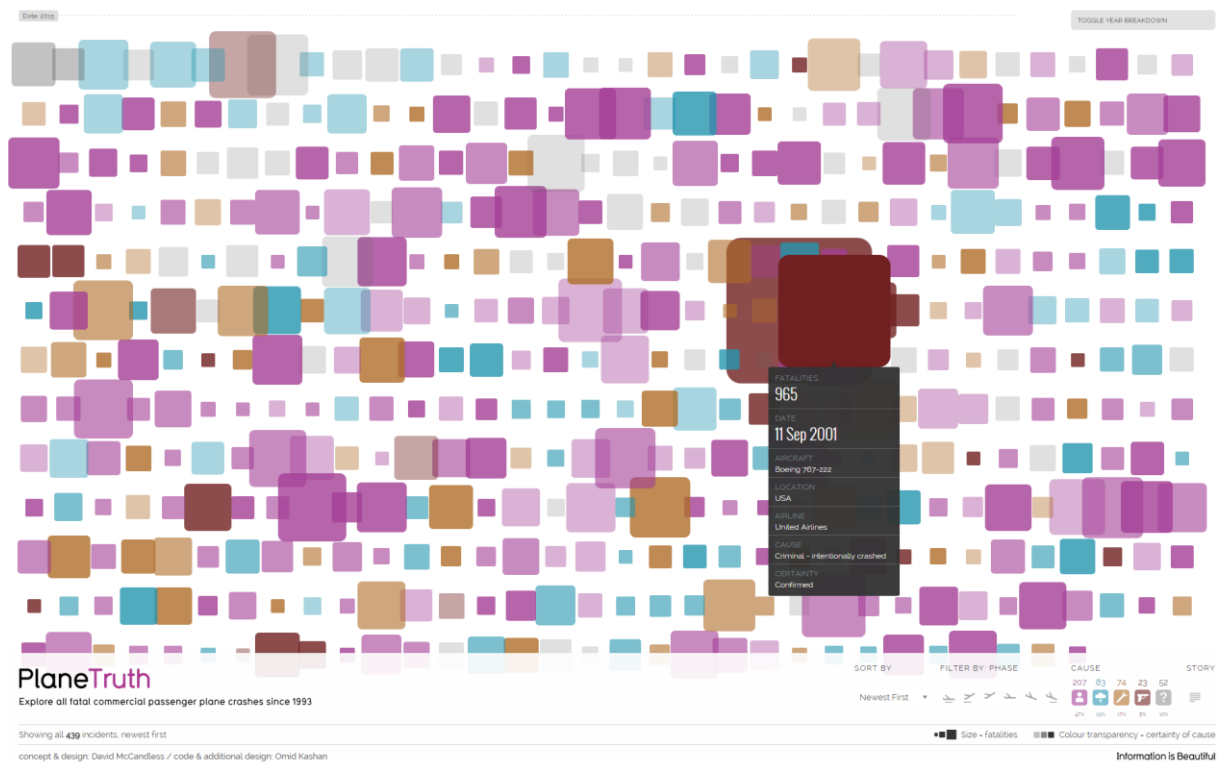


Figure 1: “PlaneTruth” interactive data visualisation (McCandless 2015b) – Every square represents a plane crash. Hovering over one of the squares opens up the shown information box with details on the represented flight. Clicking on one of the icons below or the options above the squares, triggers the filtering of data by year, cause and/or phase. A short animation is run then, in which rejected data disappears and filtered data rearranges to fill the available space appropriately.

Figure 2 illustrates another example of an interactive data visualisation. It is created with the visualising tool “Cascade” (Thorp 2011; 2012). The tool was designed to create interactive visualisations about a tweet on Twitter on demand. The tool visualises the impact of the tweet in retweets and conversations in a cascade styled tree. The cascade is visualised in a 3D space and hence it allows panning and tilting the view, zooming, and folding as an alternative to filtering.

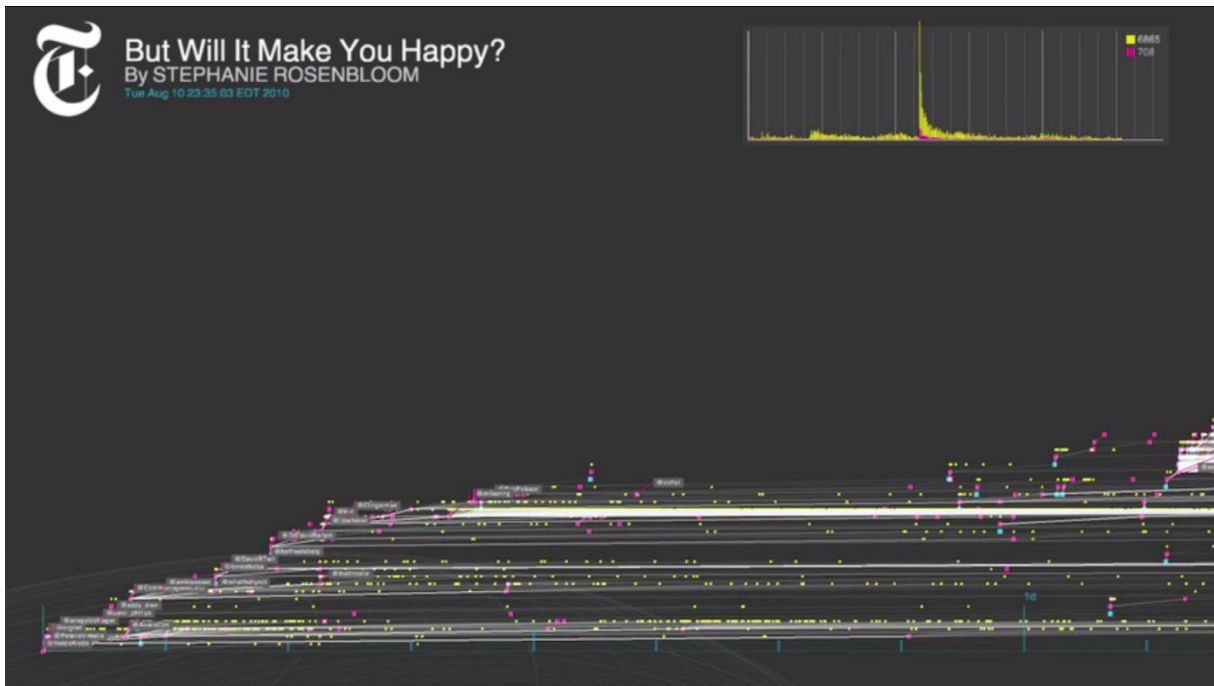


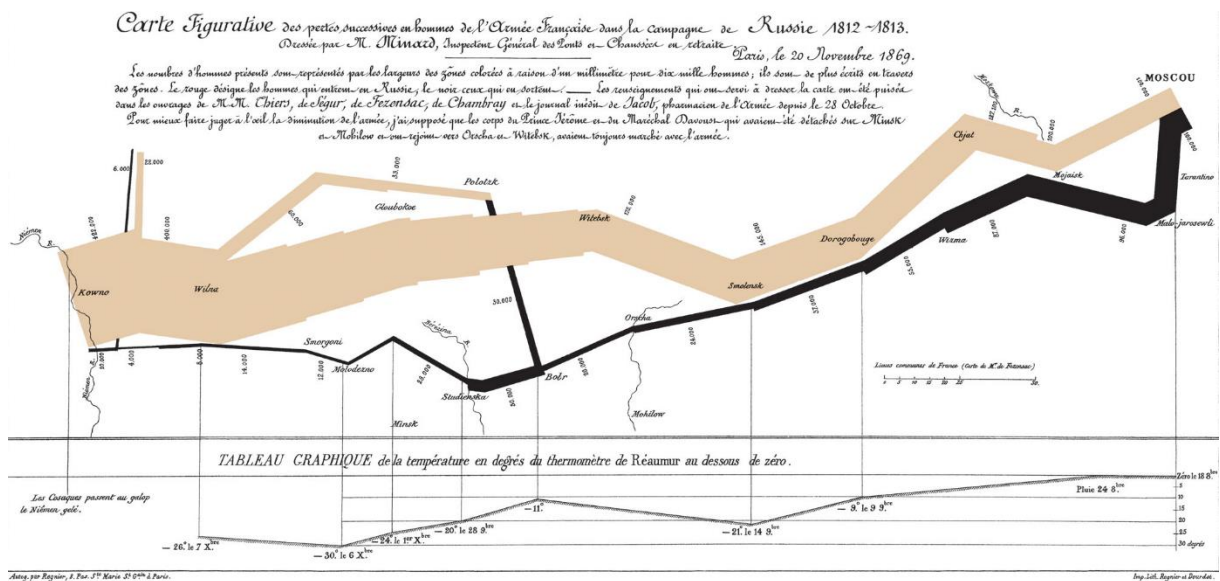
Figure 2: “But Will It Make You Happy?” (Rosenbloom 2011) – An interactive data visualisation based on a New York Times article by Stephanie Rosenbloom (2010), created with the help of the tool Cascade (Thorp 2011). The visualisation shows the first tweet on Twitter about the article in the bottom left hand-side corner and its impact on conversations and retweets in the next 13 hours. Every little dot represents a retweet or a reference in a conversation.

## 2.4 Storytelling

Storytelling can provide data visualisation designers with established techniques that draw connections between single facts, and add characteristics to the visualisations that make them memorable. Therefore, “narrative visualisations”, a term first coined by Segel and Heer (Segel & Heer 2010), is one of the fastest growing research fields with regard to data visualisation. Kosara and Mackinlay propose, now that a solid basis of research is done on exploring and analysing big amounts of data, “storytelling is the next logical step”, which focuses on the appropriate way of presenting data (2013). Telling a story with data can improve “engagement and interest, [the] ability to remember key points [and] to make more informed decisions”. These factors are traditionally not a part of the evaluation of information visualisations, which used to only focus on the speed of the user to answer a question by use of the visualisation and the accuracy of the answer.

An early data visualisation by Minard is a good example of a static data visualisation which tells a well-known story. Figure 3 shows his visualisation about the march of Napoleon’s army to Moscow and back, between the years 1812 and 1813. This space-filling, “figurative map” (Minard 1869) tells the story of a month long march in temperatures up to  $-38^{\circ}\text{C}$ . Minard used a map representation as well as the width of the bands to give information on place, time and size of the army. Furthermore he used a line chart to illustrate the temperature on the march back home. He also used only two different

colours, and it can be fit on one piece of paper (or the viewport of a computer screen in the digital age) in order to make all information of the story available to the viewer.



condensing additional knowledge”. They recommend that “recording the kinds of questions asked” in these presentations can be of use for later revisions.

The data visualisation produced for this work will most often be presented in situations similar to the first scenario, described above. However during the evaluation phase of the visualisation, the designer will be in personal contact with the test users as described in the third scenario and even though this does not require additional flexibility of the visualisation, the recording of questions raised by the user is being carried out as part of the best practice of user evaluation.

When it comes to interactive data visualisations such as the example discussed later in this work, applying storytelling techniques can be challenging as the linearity of facts and events, traditionally required for storytelling, cannot be assumed anymore by the designer. However, in the field of computer games “stories that respond to and change based on interaction” (Kosara & Mackinlay 2013) have been explored already and the approaches can be easily transferred to interactive data visualisation.

Segel and Heer analysed 58 visualisations with regard to their storytelling design features (2010). It is not an exhaustive collection of features that need to be considered when applying storytelling to a data visualisation, but a summary of good practise. A few shall be named that are relevant to the work in Chapter 1. How they are applied to the data visualisation in this work will be explained at a later context.



Figure 4: Comparison of single-frame (a) and multi-frame (b) interactivity – Interaction between the user and a visualisation can either trigger a change of the data within the frame (a), such as filtering, rearrangement, colour change or it can trigger the visualisation to move to a completely new frame (b).

Regarding the interactive navigation of a visualisation, Segel and Heer defined the term “single-frame interactivity” while referring to interactivity that “manipulates items within a single frame without taking the user to new visual scenes”. A comparison between single-frame and multi-frame interactivity are is illustrated in Figure 4. For example, filtering the data or selecting different data (sub) sets within a single data visualisation can be considered as single-frame interactivity. Furthermore,



they found that a “tacit tutorial” such as a initially shortly shaking button help to explain the interactivity to the user. Additionally, to avoid an overloaded layout or to introduce another layer of detail, “details-on-demand” can be introduced that are short hidden facts in the form of text for example, which show up when the user hovers over or clicks on an element. Finally, Segel and Heer claim that “transition guidance” such as gradually animated transitions, when navigating between visual representations, help make the presented data clearer to the viewer.

These storytelling features might raise the question, if they are more ‘chartjunk’<sup>2</sup> than helpful. However, the question of what makes a good data visualisation and if chartjunk can be a helpful feature will be discussed in the following.

## 2.5 A good data visualisation

The research areas above only briefly mentioned, if at all, the question of what factors make a good data visualisation. Most imply that a good data visualisation fulfils its aim of making a big, incomprehensible amount of data accessible and comprehensible. How much a visualisation reaches this aim can be determined by asking test users to clarify a question with the help of the visualisation and measuring the time and accuracy of their response (Kosara & Mackinlay 2013). However, the quality of data visualisation can be classified by many more factors that consider how much users can engage with it, which features improve memorability of the presented information and in the long term how much it influences the user in making decisions based on the information learned from the visualisation (Kosara & Mackinlay 2013).

Research that tries to clarify these questions mainly discuss the “chartjunk debate” (Few 2011). This debate is about whether a visualisation benefits from embellishments or not. Catering all opinions and points in this debate is beyond the scope of this work and hence only the main parts shall be mentioned: One of the first advocates for a minimalist approach is Tufte. His strong opinion claims that “it is wrong to distort the data measures in order to make an editorial comment or fit a decorative scheme” (Tufte 1983, p.59). To put the degree of minimalism in visualisations into numbers, Tufte introduced the data-ink ratio, where the ink used to illustrate the information is divided by the total amount of information. He claims the closer this ratio is to 1 the higher the objectivity and the more accurate the viewer’s interpretation. Later he goes a step further by saying “credibility vanishes in clouds of chartjunk” (Tufte 1990, p.34). Keeping it minimalistic in order to maximise the objectivity of a visualisation received wide support. Kress and van Leeuwen state also “visual structures point to

---

<sup>2</sup> Chartjunk are embellishments such as colour or pictures in data visualisations that are not necessary to comprehend the data presented. (Kosara & Mackinlay 2013)

particular interpretations of experience” (2006). In order to avoid pointing to a particular interpretation, data visualisations should not include more visual structures than needed to get the information across. Tufte’s strong statements were recently attenuated by Few, who sees himself as a supporter of the minimalist side of the debate (Few 2011). However, he hinted at the broadness of the term ‘chartjunk’ and stated that the actual question of the debate is whether embellishments are of use for visualisations. In his opinion they have a right to exist, if they “[engage] the interest of the reader”, “[draw] the reader’s attention to particular items that merit emphasis” or if they “[make] the message more memorable” (Few 2011).

This leads to the other side of the debate. A lot of research has been done to clarify whether embellishments are improving the memorability of the content of data visualisations (e.g. Bateman et al. 2010; Borkin et al. 2013). This topic shall be analysed in the following paragraph.

## 2.6 Memorability

It is not always the main objective of a data visualisation, however, enhancing the ability of the user to remember the information in the visualisation can be another feature that improves user perception. After all the information in the data visualisation is supposed to inform the user and hence, support his/her decision making process.

The features that make a visualisation memorable have been identified as the ones that also differentiate them from other visualisations (Kosara & Mackinlay 2013). Features that were identified multiple times throughout the research are colour, size, embellishments, orientation and human recognisable objects (Segel & Heer 2010; Borkin et al. 2013; Bateman et al. 2010; Gershon & Page 2001).

The data-ink ratio, a term coined by Tufte (1983), is mentioned regularly in relationship to memorability. Tufte proposed a high data-ink ratio, which means that if the visualisation was printed, most of the used ink would actually convey data and only a minimum of ink is used for embellishments. However in terms of memorability, it was found that a lower data-ink ratio increases memorability (Bateman et al. 2010; Borkin et al. 2013).

An advocate of embellishments in charts is Holmes. One of his charts is illustrated in Figure 5. It incorporates a lot of embellishments that point at the topic of the visualisation. Hence it features only a low data-ink ratio. However a study of Bateman et al. investigated the memorability different visualisations of Holmes in comparison to plain versions of the visualisations featuring a high data-ink

ratio. They proved that viewers are able to remember the content and details of the embellished visualisation better than of the visualisations without ‘chart junk’ (Bateman et al. 2010).

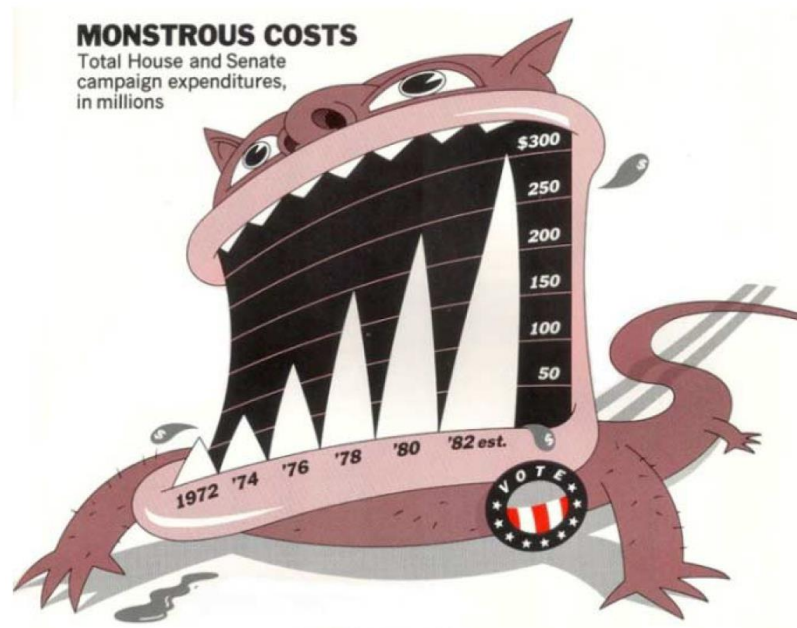


Figure 5: “Monstrous Costs” (Holmes 2010) – This static data visualisation informs about the increasing costs for political campaigns in the USA.

Although memorability is not the primary objective of the visualisation that is developed along with this study, these features will have to be considered, when making the design decisions for it.

This chapter gave an overview over the current state of the art. The chapter also outlined features that help a visualisation to convey its story, factors that make it pleasing to the eye and other examples that make it a good visualisation. How this research fed into the design and development of the “Road accidents: Ireland, 2010” visualisation will be explained in the following chapter.

## 3 Interactive Data Visualisation Design & Implementation

This chapter firstly gives an explanation on the “Road accidents: Ireland, 2010” interactive data visualisation illustrated in Figure 6 that was developed during this work. Afterwards the design decisions that take the findings of Chapter 2 into account are analysed. This chapter finishes with an explanation of the technology and visualisation tools that were used to implement the visualisation, while pointing out the challenges that arose during the implementation process.

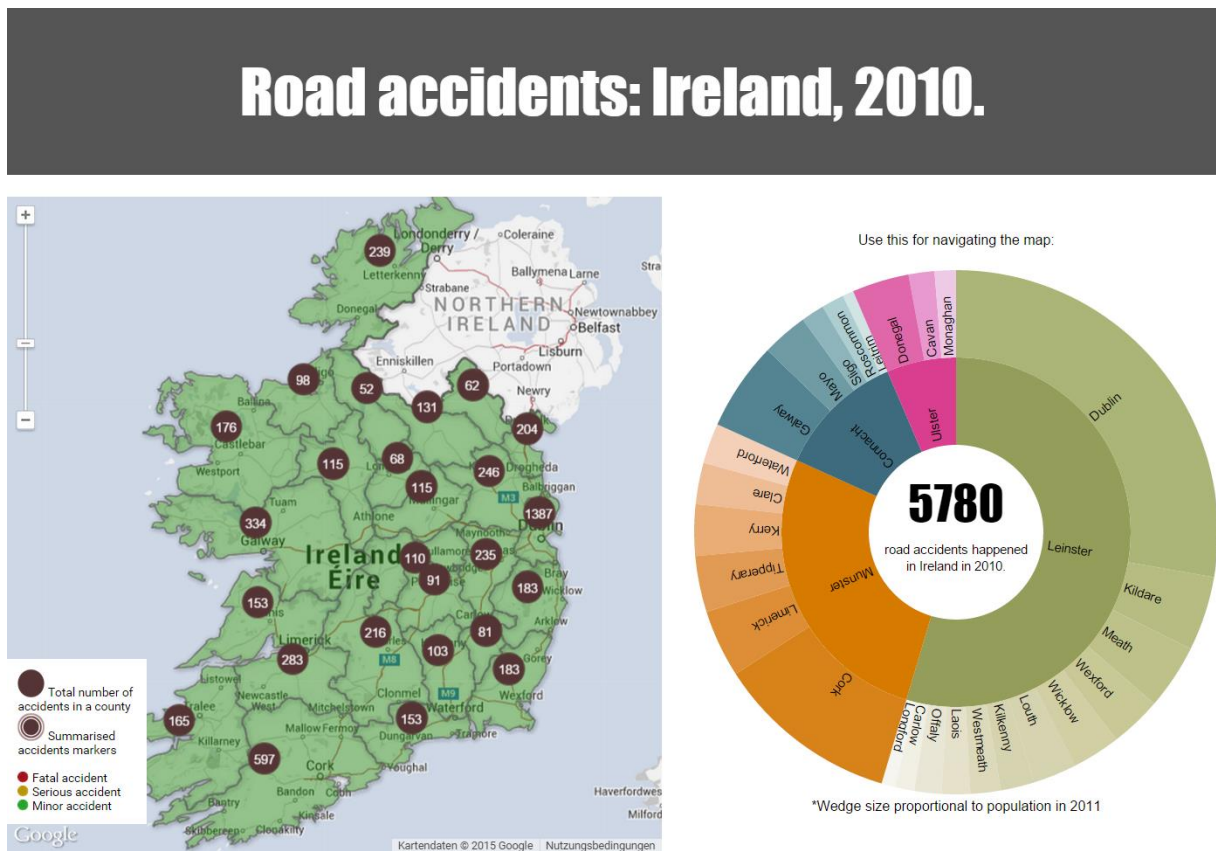


Figure 6: “Road accidents: Ireland, 2010.” Interactive data visualisation – The screenshot shows the initial state of the data visualisation. By clicking the wedges of the sunburst chart or zooming in and out on the map, the user can interact with the data and customise the view.

All views of the data visualisation display a map of Ireland and a sunburst chart. This layout is not changed by the interaction of the user. Furthermore the map includes a small legend that explains the meaning of the five different types of markers that are used on the map. A footnote in the bottom left-hand corner serves as the legend for the sunburst chart. It explains that the size of the wedges correspond to the size of population in each county and province. Throughout all the views of the

interactive data visualisation, only one instruction is given to the user; placed above the sunburst chart a sentence hints the user at using the chart to navigate the map.

Different sources contributed towards the data acquisition. Firstly, the information on road collisions in 2010 was provided by the Irish Road Safety Authority (Road Safety Authority Ireland 2015). Secondly, the population of the counties and provinces in 2011 was accessible through the Irish Central Statistics Office (Central Statistics Office 2011). Furthermore, the general coordinates for the country and the counties were obtained from Natural Earth and OpenStreetMap (Natural Earth 2011; OpenStreetMaps 2015). Additionally, the map itself is part of the Google Developer Map API (Google Developers 2013), which will be further explained later. Finally the sunburst chart is inspired by Rodden's Zoomable Sunburst (2013).

This chapter first explains the design decisions. After this, the implementation and the applied tools and libraries will be explained.

## 3.1 Design Decisions

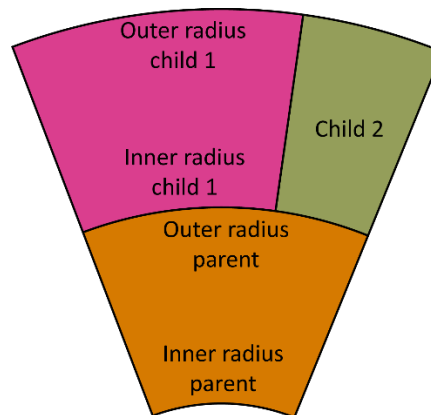
The visualisation's design is based on the findings in Chapter 2. To approach the design, the static screenshot of the initial state in Figure 6 will be referred to first. The parts of the design that become visual throughout the interaction with the user will be discussed afterwards.

In the initial view, only one type of the markers is displayed: 26 circles with labels inform the user about the total number of accidents in each county. The visibility of the other types of markers is triggered by the interaction of the user, which will be explained later. Furthermore, a label in the white space in the centre of the sunburst chart adds all the numbers displayed in the map together and displays the total sum of accidents in Ireland in 2010. Therefore, the initial view delivers an overview of the accident data as a summary, while hiding the specific details. This way the user gets to see the most general information at first sight and can go further into the details of the data by using the interactive navigation. The main navigation step is to click a county or province wedge in the sunburst chart.

The sunburst chart was initially created by Stasko et al. (2000) as an alternative for a Treemap visualisation as suggested by Shneiderman (1992) and Johnson and Shneiderman (1991). The sunburst chart as well as the Treemap are space-filling charts for visualising hierarchical data (Stasko et al. 2000). However the sunburst chart was designed to "better [convey] a hierarchy's structure, without sacrificing the use of area to present attributes of the hierarchy's node" (Stasko & Zhang 2000, p.1). In a user study that compared response time and accuracy as well as user perception, Stasko et al. (2000)

found that the sunburst chart mostly performed better in all three categories. They claimed that the sunburst chart is generally preferred by the user over a Treemap visualisation.

As explained by Stasko & Zhang (2000), in a circular sunburst chart nodes that are deeper in the hierarchy are displayed further away from the centre of the circle. This can be seen in Figure 6, where counties are ranked at a lower level than provinces and hence, they are further away from the centre of the sunburst chart. Every node is visualised by a wedge with a certain inner radius and certain outer radius. Furthermore, the inner radii of all child node wedges are adding up to exactly the outer radius of the parent node wedge. Therefore they can be drawn as an extension of the parent node. The relationship between parent and child nodes are illustrated in Figure 7. Lastly the wedge sizes, the colour of the wedges, as well as the labels of each node visualise a certain attribute of the nodes within the hierarchy.



*Figure 7: Visualisation of a parent node and two child nodes in a sunburst chart – Deeper levels of the hierarchy are further away from the centre of the circuit than higher levels, while the sum of the inner radii of all child nodes equals the outer radius of the parent node.*

Stasko & Zhang also suggested various improvements of the original sunburst chart including a zoomable version, where a clicked wedge “expands [...] radially in place [...] while compressing all other items” (2000, p.9). In their opinion, the transition from the general view to a detailed view after a click should be animated and gentle in order to make it easier for the user to track the changes in the visualised data as claimed by Segel & Heer in Chapter 2.4 (see Segel & Heer 2010).

An implementation of the sunburst chart with the above mentioned D3.js JavaScript library became popular by Rodden, who applied it to the hierarchical ordered data set of user navigation sequences on the YouTube subpages (2014). She also made available the D3.js implementation of a zoomable version as proposed by Stasko & Zhang (2000) called “Zoomable Sunburst” (Rodden 2013).

The Zoomable Sunburst chart in Figure 6 visualises the spread of the population over the different counties in Ireland. The hierarchy of this data set is very flat and contains only two levels: the provinces and the counties. Due to its nature, the sunburst chart preforms well for visualising the hierarchy of regions in Ireland in a space-filling way, while the interactive zoom feature facilitates the use as a navigation element very well.

A click on one of the county or province wedges triggers the zoom in to the next level of the hierarchy. The addressed province wedge expands to the full size of the inner ring and the counties of the provinces group around it. This transition happens so smoothly that the user is able to follow the details of the change. The click also triggers the map to zoom in and pan to the addressed county or province and highlight its boundaries, which is illustrated in Figure 8.

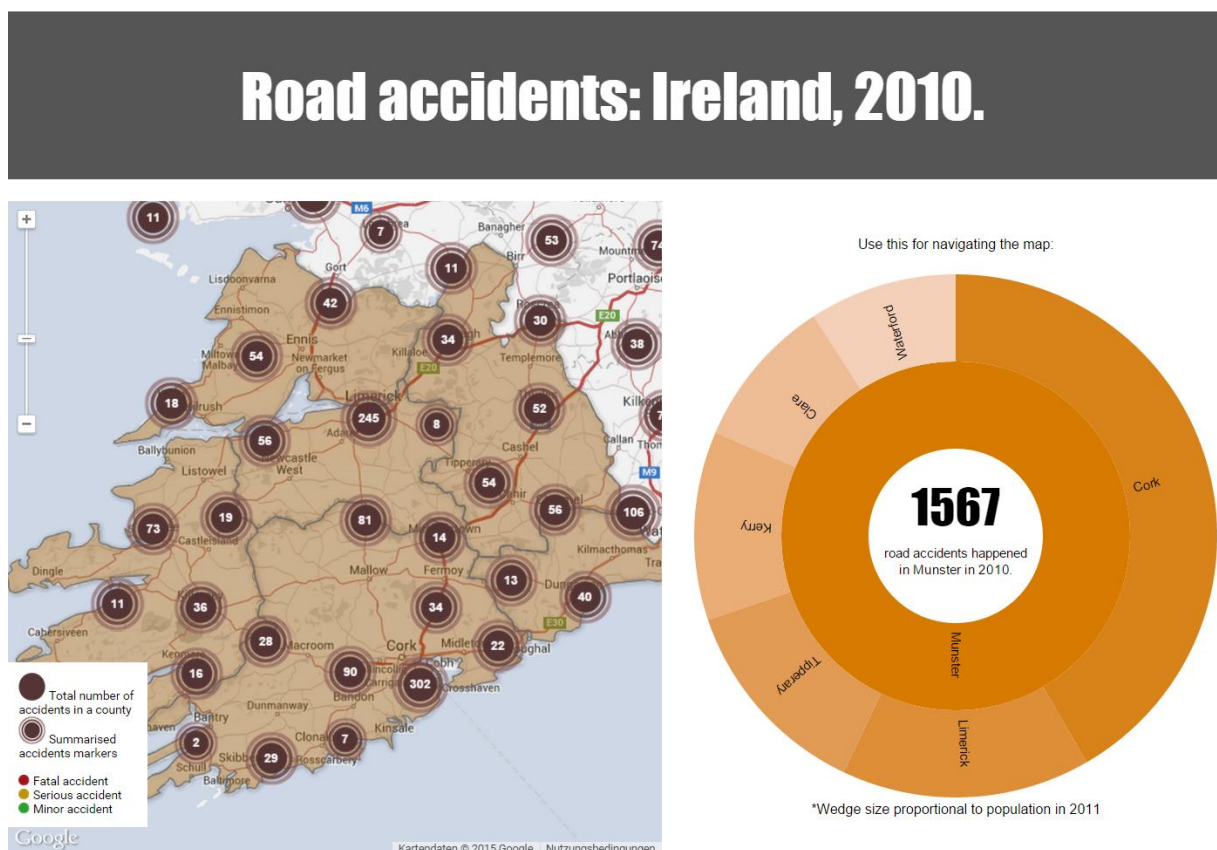


Figure 8: The province of Munster highlighted in the map – A click on the wedge labelled “Munster” in the sunburst chart triggers a zoom into the province on the map and the highlighting of its outline.

The colour of the wedge of each province reflect a main colour of the province’s flag, while the counties exhibit a brighter shade of their province’s colour as suggested by Driscoll (2010).

The click on a wedge of the sunburst chart also triggers a change in the markers on the map. While the overview of the whole country shows the total amount of accidents in each county, when zoomed into



a county or a province, every single accident is present on the map, but summarised in so-called 'clusters'. This preserves a tidy and not overloaded view, and therefore a better user experience (Mahe & Broadfoot 2010). The exact procedure of the clustering algorithm that decides which markers are summarised together will be explained later in Chapter 3.2.2.

A cluster's centroid is visualised by a circle surrounded by rings with gradually increasing opacity and a label indicating the number of summarised accidents as can be seen in Figure 8. Hence the marker icon for a cluster is very similar to the marker indicating the overall number of accidents in a county. However it is different enough to make sure the user recognises which marker he / she is looking at. The clustering is only applied up until zoom level ten. In a higher zoom level, the markers are displayed separately as shown in Figure 9. This ensures that the user is able to investigate the specific details of the accidents in a certain area.

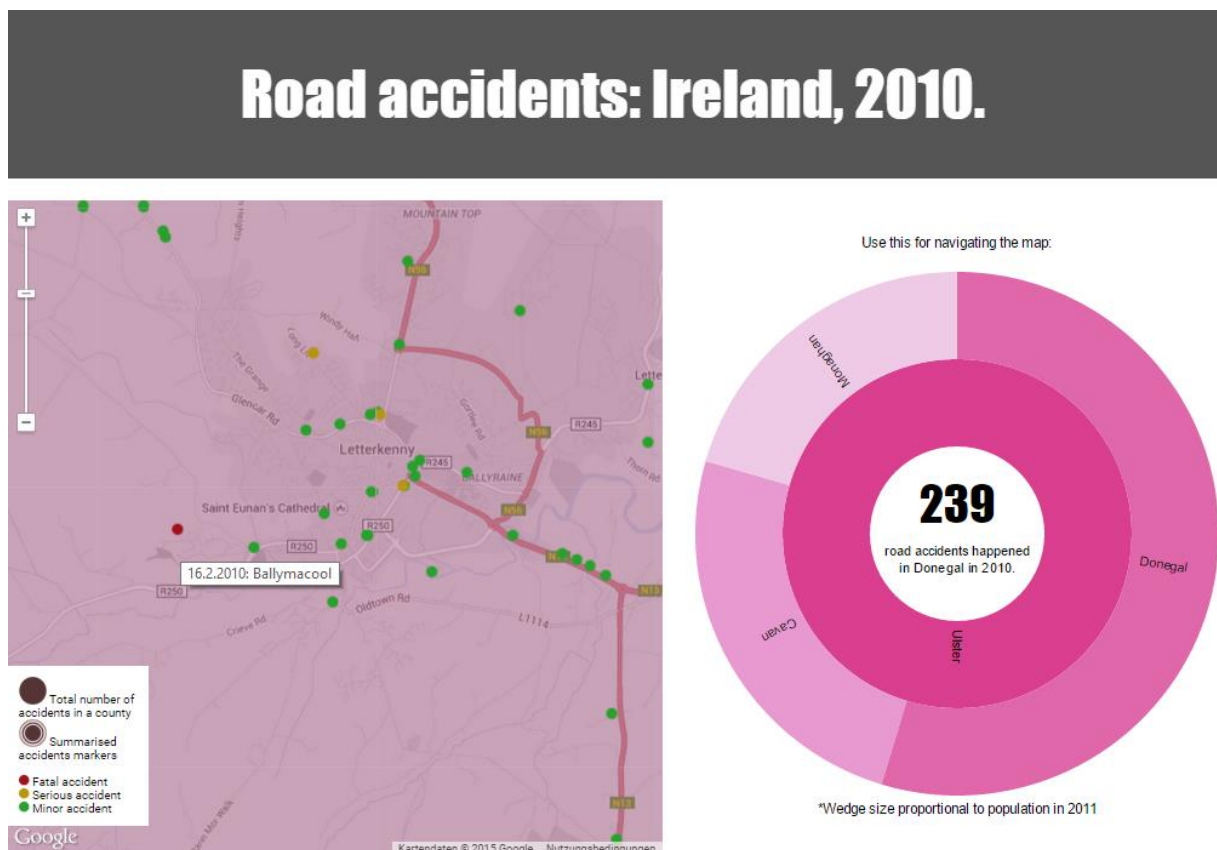


Figure 9: Markers for every single accident in Letterkenny in Donegal – A zoom higher than ten disables the cluster algorithm so that every single accident is represented by a marker on the map. Hovering over an accident marker reveals an information box with details on the accident.

Another feature has an effect in this part of the visualisation: Hovering over a single accident marker gives the available details about the accident in a small information box also shown in Figure 9.



## 3.2 Implementation and Tools

The “Road Accidents: Ireland, 2010” visualisation is a collaboration of different web technologies that are commonly used for visualisations today. Generally, the visualisation is written in HTML5, CSS3 and JavaScript. However, some tools and libraries are helping with visualising the data set in an appropriate way. These will be explained in the following.

### 3.2.1 JSON

The JavaScript Object Notation (JSON) is a standardised notation model for any kind of data originating from the use in the programming language JavaScript. Alternative models exist, but for web technologies, JSON is the standard. Many JavaScript libraries provide interfaces that require the attributes to be a JSON object.

The Google Maps API, the MarkerClusterer, the Google Fusion Tables and D3.js, which are the technologies used for the road accident visualisation, all work with JSON. Mostly they require option definitions or customisations to be passed on in a JSON object.

```
var accidents2010 = {
  "year": 2010,
  "total": 5780,
  "entries": [
    {
      "city": "Ballinacourty",
      "lat": 52.08361111111111,
      "lon": -7.562222222222222,
      "severity": 3,
      "date": "1.1.2010"
    },
    {
      "city": "Carrigtwohill",
      "lat": 51.91305555555555,
      "lon": -8.23,
      "severity": 3,
      "date": "11.8.2010"
    }
  ]
}
```

*Code 1: Accidents 2010 JSON object example – The accident data that feeds into the “Road Accidents: Ireland, 2010” visualisation is structured in a JSON object. The “entries” array in the object is shortened for reasons of simplicity. In the live system it contains 5780 JSON objects that specify an accident each.*

JavaScript naturally implements easy ways to create and access a JSON object. Hence, the JavaScript Object Notation was the natural choice to structure the accident data and the details about the counties and provinces used in the road accident visualisation. The object for the accident data is

illustrated in Code 1. It stores information about the year, the total amount of accidents, and details such as the place name, the coordinates, the severity and the date of each accident.

```
var dublin = {  
  
  "id": 11,  
  "naturalEarthName1": "Dublin",  
  "latMax": 53.6342156,  
  "latMin": 53.1781813,  
  "lonMax": -5.9942565,  
  "lonMin": -6.5468919,  
  "centreLat": 53.40619845,  
  "centreLon": -6.2705742,  
  "zoom": 9,  
  "color": "#acb578",  
  "yearlyData": {  
    "2010": {  
      "roadAccidents": {  
        "total": 1387  
      }  
    },  
    "2011": {  
      "inhabitants": 1273069  
    }  
  }  
}
```

*Code 2: Details of County Dublin JSON object example – Every detail about a county, a province or the Republic of Ireland that is needed in the “Road Accidents: Ireland, 2010” visualisation is contained in the corresponding JSON object.*

On the other hand, Code 2 shows the JSON object with details for County Dublin as an example on how the information for every county, all provinces and the country in general are stored. In the visualisation, details such as the bounding box (defined by the smallest and largest longitude and latitude values), the centre coordinates or the name within the Natural Earth Fusion Table (see Chapter 3.2.3) are regularly required. Through the structured JSON objects the data can be easily accessed from the code, maintained, extended or replaced.

### 3.2.2 Google Maps API and the MarkerClusterer

The Google Maps API allows the integration of an interactive map into many applications. It offers many different interfaces and customisation options to reuse the usual Google Maps behaviour in a different way or to extend it. For the map that is part of the visualisation developed for this work in Figure 6 and Figure 8, the JavaScript (Google Developers 2015) library in particular came to use.

Through the Google Maps API the markers for each accident, the information appearing when hovering over a marker, and the zoom and pan effects, when clicking a wedge of the sunburst chart, are

visualised. Furthermore, the clustering of the markers and the colouring of the clicked provinces and counties are implemented with the help of libraries and tools that integrate the Google Maps API. They are explained in the following.

The MarkerClusterer (Google Code 2015) is an additional JavaScript library that extends the Google Maps API by a simple grid-based clustering algorithm (Mahe & Broadfoot 2010). As illustrated in Figure 10 the algorithm iterates over the markers it is given and assigns each marker to the nearest cluster. To find the nearest cluster, the algorithm calculates, if the coordinates of a marker lie within the square bounds of an existing cluster. If two or more clusters come into consideration, the closest cluster is calculated. If the marker is not within any boundaries of the existing clusters, a new cluster with the marker's coordinates as the centroid is being created.

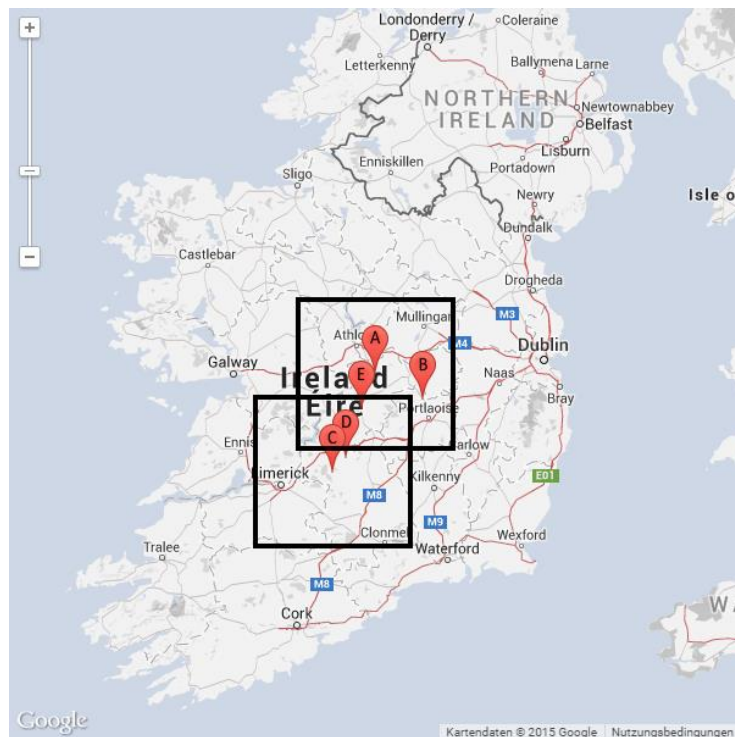


Figure 10: Grid-based clustering – The algorithm iterates over all markers. It starts with A and hence the square with the coordinates of A as the centroid becomes the first cluster. After this the algorithm checks B. B lies within the boundaries of the first cluster, so B becomes part of the first cluster. Now the algorithm finds out that the coordinates of C lie outside of the first cluster, so the square box with C as the centroid becomes a second cluster. D lies only within the second cluster and hence, it becomes part of it. Finally the algorithm checks D. D lies within the boundaries of both existing clusters, but becomes part of the cluster around A, as it is closer to its centroid than to the centroid of the second cluster.

The MarkerClusterer is used for the visualisation presented in this work to summarise the road accident markers until a zoom level of 10. For a higher zoom level than 10 the MarkerClusterer is not active anymore, so that the user is able to see the accidents separately.

### 3.2.3 Google Fusion Tables

Commonly used data management systems are designed and built with regard to a proprietary use within a single company and/or for the use by people with professional know-how. Acknowledging the contemporary needs for a more flexible data management outside of the business world, Google decided to design an equivalent cloud-based system for people with less technical skills that would implement “functionality for today’s connected world” (Gonzalez et al. 2010).

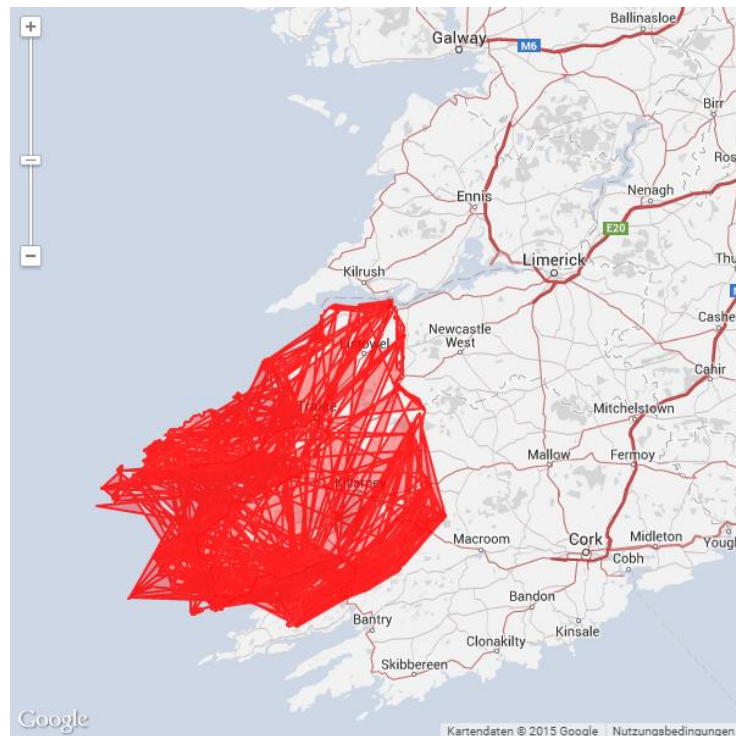


Figure 11: Google Maps Polygon based on the coordinates for the boundaries of County Kerry from OpenStreetMap – The result cannot be used for the visualisation, however it was rendered in under a second.

In 2009 they launched the Google Fusion Tables, a data management system that also supports a community, where data can be shared with other users of the system or the public and where multiple users can contribute to the data, while not being limited to one company or team. Google furthermore aimed to design the Fusion Tables to be easily accessible to people with less technical skills as well as easily accessible from many web technologies (Gonzalez et al. 2010). This way Google claimed to have taken the burden of data management design and scalability off the user, so that the users can spend more time on maintaining the data and using it for storytelling (Madhavan et al. 2012).

In the data visualisation developed for this work the interface for Fusion Tables within the Google Maps API was used to integrate a public database by Natural Earth (2011) that stores the geographical

position and shape of every county in Ireland. The shapes are requested through a simple SQL statement and are loaded into the FusionTablesLayer that lies as a separate layer on top of the map.

The Fusion Tables helped solving a performance problem that the initial prototype of the visualisation had. In a first approach to visualise the boundaries of each county, the coordinates for the boundaries were scraped from OpenStreetMap (2015) and converted into a Google Maps polygon. However, the coordinates for OpenStreetMap polygons are not saved in the same order that a Google Maps polygon reads it and the resulting visualised polygons are not useful as shown in Figure 11.

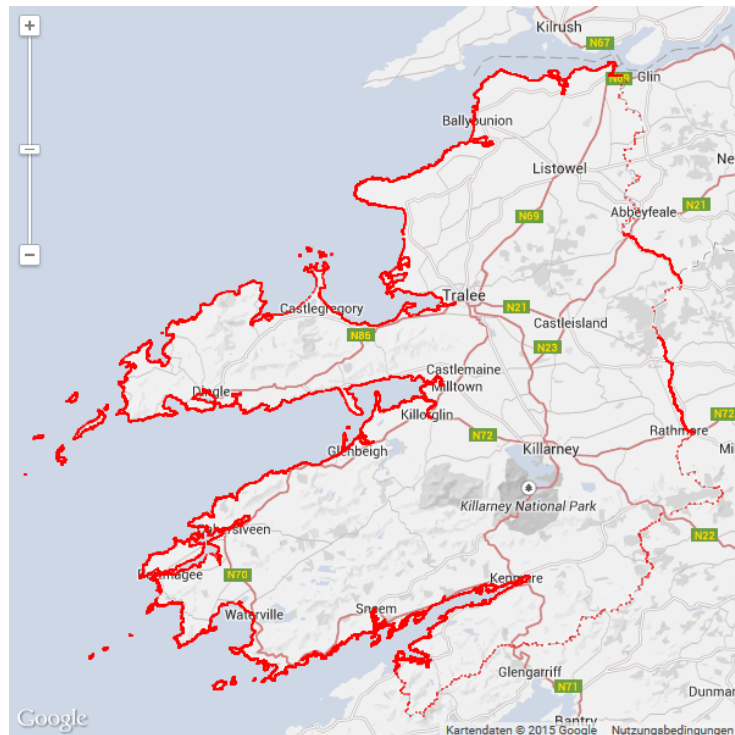


Figure 12: Google Maps Circles for every coordinate of the boundaries of County Kerry from OpenStreetMap – The result is a very accurate outline of County Kerry, however the visualisation takes several minutes to render the circles.

In a second approach the scraped coordinates were interpreted and loaded into the map as a small single Google Maps circle or a Google Maps marker each, which is illustrated in Figure 12. The result had the desired effect and showed in very high detail, where the boundaries of a county are. However, the rendering of many thousand circles or markers slowed down the client enormously so that the boundaries would not be visible before a rendering time of several minutes.

Obtaining the desired result, but in an unbearable rendering time, drew the attention to an approach that uses Google servers for rendering instead of the client (Mahe & Broadfoot 2010) by applying a FusionTablesLayer. The polygons drawn through this approach are not as accurate as the results of the second approach (see Figure 13), however every time the user selects a new county, the requested

polygons are visible in under a second. For a zoom rate that is low enough so that the users can overview the whole county or province the inaccuracy is not recognisable as illustrated in Figure 8.



Figure 13: Accuracy of the polygons from the Natural Earth Fusion Table – The boundaries of county Kerry and Clare are drawn not very accurate at high zoom rates, but the polygons are loaded into the map in under one second.

The Google Fusion Tables can deliver spatial and geological related data sets that usually involve highly detailed data, very quickly as their design focuses on sampling data sets in order to deliver only a small, but representative part of the data set (Madhavan et al. 2012). Furthermore, the Google Fusion Tables keeps possibly requested columns in the memory of the server, so that they can be quickly accessed (Madhavan et al. 2012).

### 3.2.4 D3.js

The JavaScript library D3.js provides functionality for designing interactive data visualisations based on dynamically changing data sets. Developed by Bostock et al., D3.js provides an interface to directly bind data attributes to elements of the Document Object Model<sup>3</sup> and manipulate them (2011).

<sup>3</sup> Document Object Model (DOM): The model treats elements of documents written in HTML, XML or SVG as a tree representation. It provides a well-established and standardised way of accessing these elements from programming languages such as JavaScript (Pfeiffer et al. 2015).

It allows to select elements from the DOM based on the W3C Selectors API. Functions for binding data, styling, transforming or manipulating DOM elements in other ways can be applied to the selection. All functions return the updated selection. Hence, D3.js allows “method chaining” (Bostock et al. 2011), where long chains of function calls are concatenated to each other to shorten the code.

For the visualisation developed for this work, D3.js was the main tool to create the sunburst chart, which serves as the main interface to navigate the map. D3.js also supports the chart’s interactive zoom function. Additionally, the transition that is triggered by a click event on the chart is defined with the help of D3.js.

The data that defines the wedge sizes, colours and labels in the sunburst chart are read from the same JSON that feeds the map with data. The data is rearranged to another JSON that is accepted by the data binding method of D3.js. This design of the code allows an easy extension of the sunburst chart by data of other years or a quick change of the colour coding of both the sunburst wedges and the corresponding county highlighting in the map.

This chapter introduced the “Road accidents: Ireland, 2010” visualisation and described the process and challenges of designing and implementing it. It explained the Zoomable Sunburst chart and the utilised clustering technique. Furthermore it described the purpose of the technologies and tools and their use in the visualisation. After its successful implementation the “Road accidents: Ireland, 2010” visualisation was compared to other visualisations in an experiment that will be outlined in the next chapter.



## 4 Usability and Memorability Tests

An experiment was set up in order to investigate the usability of the visualisation introduced in Chapter 3 and its effect on the memorability of the information presented. This chapter firstly explains the purpose and the content of the experiment. In addition, details on the set up are given and the procedure is explained. Thirdly, this chapter explains the questions that are subject to the experiment and discusses what part of the investigation each question contributes to. The chapter finishes by presenting and analysing the results of the experiment.

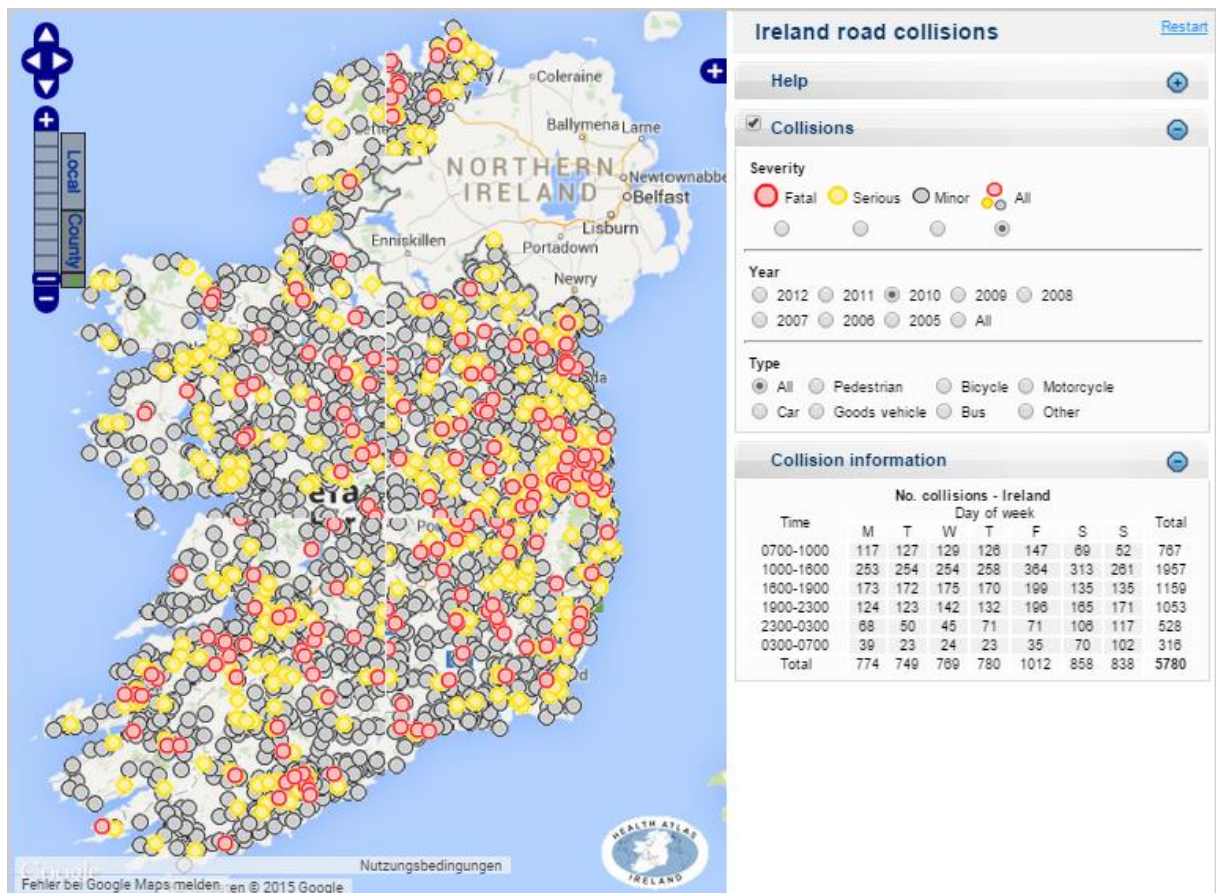


Figure 14: All road collisions in 2010 displayed by the RSA visualisation (Road Safety Authority Ireland 2015).

The experiment compares the visualisation generated by this study (referred to as “VIS-1”) with two other established visualisations. The two other established visualisations are based on the same data set. In the experiment a participant uses one specific visualisation at a time to answer several questions regarding road accidents in Ireland in 2010.

Visualisation two is the “Ireland Road Collision web interface” by the RSA of Ireland (Road Safety Authority Ireland 2015), which is illustrated in Figure 14. It will be referred to as “VIS-2”. This



visualisation is part of the website of the RSA and seemingly serves as an informative or educational tool for the public. Therefore it is delivered through a web browser just like VIS-1. This interactive visualisation is comparable to VIS-1 as it presents a map of Ireland on the left hand-side and the navigation on the right hand-side. It furthermore offers the user the ability to zoom in to the map through a vertical slider in the top left-hand corner of the map in the same manner as VIS-1. Each accident is visualised by a circle that is colour coded by the different colours red, yellow and green to differentiate between fatal, serious and minor accidents respectively. Finally VIS-2 facilitates the highlighting of a specific county as well as VIS-1. In Figure 15 the outline of County Galway is highlighted to illustrate this. VIS-2 is interactive by letting users zoom in to the map and navigate it through the commonly known Google Maps interfaces. It also gives the user the opportunity to filter the illustrated data by year, severity of the accidents and the type of collision. The only animation in the visualisation is the zoom animation of Google Maps.

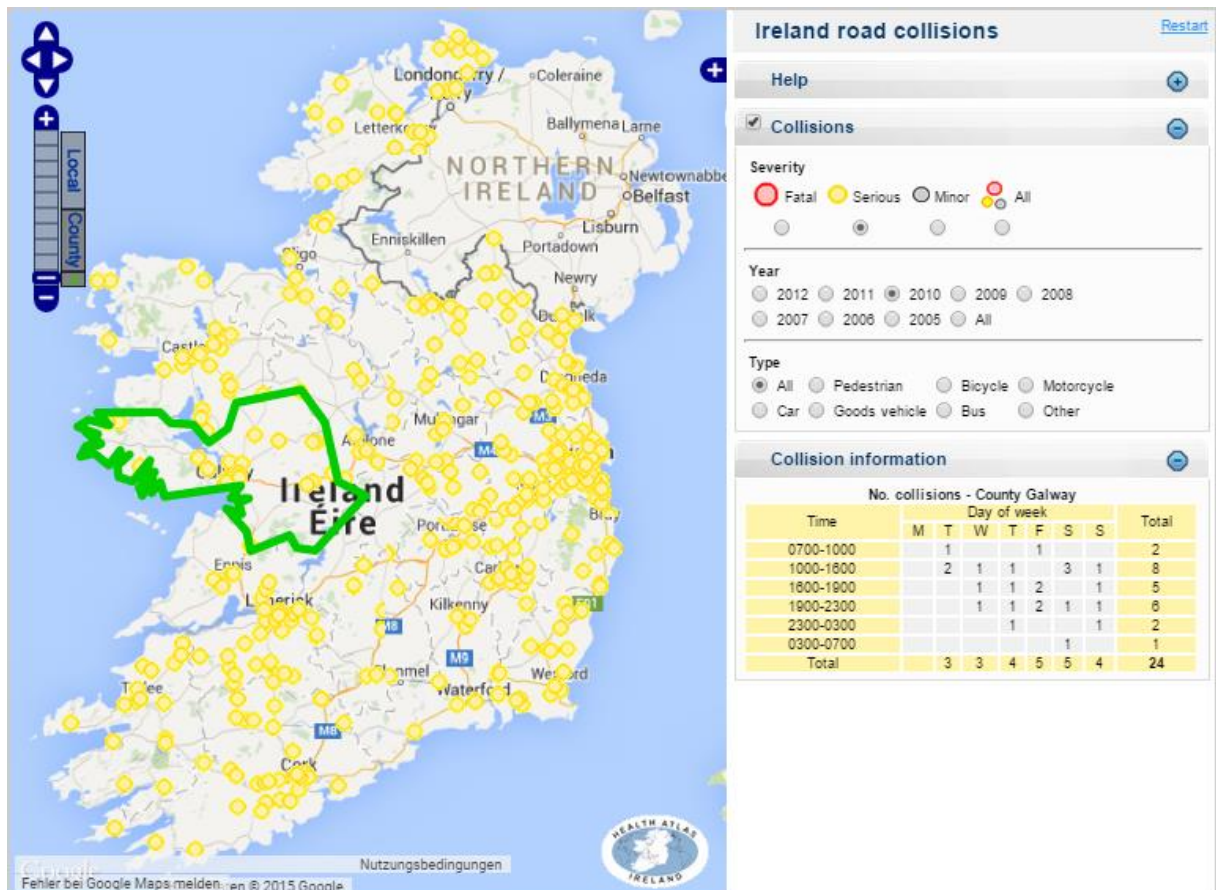


Figure 15: Serious road collisions in 2010 and County Galway highlighted in the RSA visualisation (Road Safety Authority Ireland 2015).

The second visualisation that is compared to VIS-1 in the experiment is a spreadsheet. It covers all reported road accidents in Ireland in 2010 by their geological location in latitude and longitude, name

of the place, county, severity and date of the year. A screenshot shows the first 25 entries of the table in Figure 16. The spreadsheet is referred to as VIS-3 in this experiment. All columns in this spreadsheet are unmodified columns from the original data set (Road Safety Authority Ireland 2015) that serves as the base for VIS-1 and VIS-2. However columns that were not relevant to the other two visualisations were extracted in the spreadsheet for clarity, while the adopted ones were rearranged.

	A	B	C	D	E	F	G	H	
1	County	City / Townland	Latitude	Longitude	Accident Type	Day	Month	Year	
2	23	Ballinacourty	52° 5' 1.439"N	7° 33' 44.821"W	Minor		1	1	10
3	1	Tullow	52° 49' 15.071"N	6° 48' 8.410"W	Minor		1	1	10
4	15	Dunleer	53° 50' 18.276"N	6° 24' 54.803"W	Minor		1	1	10
5	6	Dublin	53° 21' 39.740"N	6° 16' 57.582"W	Minor		1	1	10
6	5	Ballyderowen	55° 3' 7.584"N	7° 24' 4.119"W	Minor		1	1	10
7	15	Drogheda	53° 43' 21.743"N	6° 22' 15.775"W	Minor		1	1	10
8	13	Finnoe	52° 33' 48.825"N	9° 10' 39.574"W	Fatal		1	1	10
9	2	Cavan	53° 59' 0.809"N	7° 22' 56.869"W	Minor		1	1	10
10	24	Rochfortbridge	53° 24' 8.531"N	7° 17' 35.328"W	Minor		1	1	10
11	4	Aherla	51° 51' 5.566"N	8° 44' 49.246"W	Minor		1	1	10
12	23	Bunmahon	52° 8' 21.044"N	7° 20' 23.686"W	Minor		1	1	10
13	16	Balla	53° 48' 3.962"N	9° 6' 32.080"W	Minor		1	1	10
14	6	Dublin	53° 21' 33.183"N	6° 14' 44.765"W	Minor		1	1	10
15	23	Kilmacthomas	52° 12' 8.931"N	7° 25' 44.403"W	Minor		1	1	10
16	6	Westmanstown	53° 22' 7.740"N	6° 26' 0.764"W	Minor		1	1	10
17	23	Dungarvan	52° 5' 30.235"N	7° 37' 11.075"W	Minor		1	1	10
18	19	Kilcormac	53° 10' 33.150"N	7° 43' 51.498"W	Minor		2	1	10
19	13	Raheen	52° 37' 54.816"N	8° 39' 17.096"W	Serious		2	1	10
20	5	Bridgend	55° 2' 35.219"N	7° 22' 55.082"W	Minor		2	1	10
21	5	Muff	55° 4' 3.135"N	7° 16' 8.131"W	Serious		1	1	10
22	5	Drumfries	55° 11' 46.400"N	7° 23' 45.055"W	Minor		2	1	10
23	11	Mountrath	53° 0' 17.948"N	7° 28' 45.060"W	Minor		1	1	10
24	6	Meekstown	53° 24' 19.439"N	6° 17' 53.963"W	Minor		2	1	10
25	5	Raphoe	54° 52' 51.439"N	7° 32' 54.840"W	Minor		2	1	10
26	2	Belturbet	54° 5' 55.208"N	7° 25' 51.921"W	Serious		2	1	10

Figure 16: The first 25 entries of the spreadsheet visualisation – Every entry represents one accident and is specified by the county, the name of the place, the longitude and latitude, the severity and the date.

The spreadsheet visualisation is considered to be a static alternative to VIS-1. It resembles a digital table. However, because it is opened in a spreadsheet processing application, it offers a few basic interactive elements to the user. Firstly, the user must scroll down, as the full length of the table is too long for the viewport of the screen used in the experiment. Secondly the user can sort or filter the entries in any column through the option next to the table headings (see Figure 17). Because of these options the spreadsheet visualisation cannot entirely be seen as a static visualisation. However due to the size of the table, the visualisation would not be able to compete with the response times of the other visualisations. Furthermore the visualisation does not include any animation.

	A	B	C	D	E	F	G	H	I
1	County	City / Townland	Latitude	Longitude	Accident Type	Day	Month	Year	
2	23	Ballinacourty	52° 5' 1.4			1	1	10	
3	1	Tullow	52° 49' 15			1	1	10	
4	15	Dunleer	53° 50' 18			1	1	10	
5	6	Dublin	53° 21' 39			1	1	10	
6	5	Ballyderowen	55° 3' 7.5			1	1	10	
7	15	Drogheda	53° 43' 21			1	1	10	
8	13	Finnoe	52° 33' 48			1	1	10	
9	2	Cavan	53° 59' 0.			1	1	10	
10	24	Rochfortbridge	53° 24' 8.			1	1	10	
11	4	Aherla	51° 51' 5.			1	1	10	
12	23	Bunmahon	52° 8' 21			1	1	10	
13	16	Balla	53° 48' 3.			1	1	10	
14	6	Dublin	53° 21' 39			1	1	10	
15	23	Kilmacthomas	52° 12' 8.			1	1	10	
16	6	Westmanstown	53° 22' 7.			1	1	10	
17	23	Dungarvan	52° 5' 30			1	1	10	
18	19	Kilcormac	53° 10' 33			2	1	10	
19	13	Raheen	52° 37' 54			2	1	10	
20	5	Bridgend	55° 2' 35			2	1	10	
21	5	Muff	55° 4' 3.1			1	1	10	
22	5	Drumfries	55° 11' 46			2	1	10	
23	11	Mountrath	53° 0' 17			1	1	10	
24	6	Meekstown	53° 24' 19			2	1	10	
25	5	Raphoe	54° 52' 51.439"N	7° 32' 54.840"W	Minor	2	1	10	
26	2	Belturbet	54° 5' 55.208"N	7° 25' 51.921"W	Serious	2	1	10	

Figure 17: The filter and sort option of the spreadsheet visualisation.

The screenshots of all three visualisations above hint at a difference in the delivered aspects of the road accident data set. The questions asked throughout this experiment aim to use data that is delivered by at least two visualisations. Questions regarding data that is not delivered by one of the three visualisations, only need to be answered with the help of one of the other two visualisations. Hence, every question in the experiment can be answered correctly using the corresponding visualisation in one way or another. Table 1 puts the different aspects of the data delivered by the three visualisations into comparison.

In this experiment the questions exclusively ask for information from the year 2010, as VIS-1 and VIS-3 only cover this time range. Furthermore, questions regarding the county boundaries will only be asked referring to VIS-1 and VIS-2. Other questions refer to the total number of accidents in a certain region, such as counties or specific numbers on fatalities in these regions, as this information is provided in all three visualisations.

	<b>VIS-1</b>	<b>VIS-2</b>	<b>VIS-3</b>
<b>Time range</b>	2010	2005-2012	2010
<b>County borders</b>	x	x	
<b>Province borders</b>	x		
<b>Total number of accidents in regions (county, province, national)</b>	x	x (indirect through separate "Collision information" tables of the 3 severity filters)	x
<b>Relative number of population in counties and provinces</b>	x		
<b>Clustering of accidents that are close to each other</b>	x		
<b>Main type involved in road collision (pedestrian, bicycle, car, etc.)</b>		x	
<b>(Presumably average) spread of collisions throughout a day per county per severity</b>		x	
<b>Data on single accident</b>			
Latitude & longitude			x
Geological location on the map	x	x	
Place name	x		x
Date	x		x
Severity	x	x	x
County	x		x

Table 1: Comparison of information coverage in the visualisations of the experiment.

## 4.1 Experiment Set-up

Each participant gets to see and use all three visualisations. The visualisations are delivered digitally on a Windows 7 computer with a 24-inch screen and a standard computer mouse. Furthermore, the two interactive visualisations are opened in a separate Google Chrome browser window, whereas the spreadsheet is opened in Microsoft Office Excel 2010.

In the beginning every participant needs to answer an introductory question on each visualisation in order to get familiar with the navigation. These questions are the same for every participant. After this, each participant answers ten primary questions. Before the next question is asked, the participant is asked to refresh the window of the required visualisation (or to reset the filters in the spreadsheet visualisation). This way every question starts with the initial view of the visualisation. Different questions apply to different participant groups as explained in the next section. There are four main participant groups in total.

If a participant's response is wrong and the correct answer is needed for a subsequent follow up question such as for question #02, the instructor will help the participant to correct the mistake.

Furthermore, the instructor is allowed to help if a participant struggles with the geological location of a place or a region that is subject to a question.

## 4.2 Questions

### 4.2.1 Introductory questions.

In the beginning of the experiment, each participant is asked one question for each visualisation. The responses are not recorded and there is no time restriction so that the participant can get familiar and comfortable with the navigation. The questions aim to make the user familiar with all parts of the visualisation that are relevant for the main questions. The introductory questions are the same for every participant group. They are as follows:

VIS-1:

How many serious accidents happened in the city of Waterford in 2010?

The correct answer is 2.

VIS-2:

How many minor road accidents happened in County Galway in 2010?

The correct answer is 5186.

VIS-3:

How many serious accidents happened in Waterford town in 2010?

The correct answer is 2.

### 4.2.2 Main questions

The main questions are investigating three different aspects of the data visualisations: response time and accuracy, user perception and memorability. Following this structure, the experiment is divided into three different parts. The questions of the experiment are listed in Table 2. The order in Table 2 is almost the same order in which they were presented to the participants. The one only exception will be explained later.

First the participants answer five questions, while their response time is being timed by the instructor. Three of the five questions are asked twice, which means that they need to be answered with the help of VIS-1, and another time with the help of one of the other visualisations. These three questions aim to directly compare the response time and accuracy of each user on the different visualisations. The other two questions of part one are only asked once. Even though the response time for these two questions is timed as well, they investigate the memorability of the information at a later stage of the experiment.

ID	Test purpose	Question	Visualisation
#01	Response time & accuracy / Memorability part 1	How many road accidents happened in County Dublin in 2010? The correct answer is 1387.	Group 1: VIS-1 and VIS-3 Group 2: VIS-1 and VIS-3 Group 3: VIS-1 Group 4: VIS-3
#02	Response time & accuracy / Memorability part 1	How many fatal road accidents happened in the city of Limerick in 2010? The correct answer is 1.	Group 1: VIS-1 Group 2: VIS-3 Group 3: VIS-1 and VIS-3 Group 4: VIS-1 and VIS-3
#03	Response time & accuracy / Memorability part 1	How many road accidents happened in County Mayo in 2010? The correct answer is 176.	Group 1: VIS-2 Group 2: VIS-1 Group 3: VIS-1 and VIS-2 Group 4: VIS-1 and VIS-2
#04	Response time & accuracy	Can you highlight the County Cork?	VIS-1 and VIS-2
#05	Response time & accuracy / Memorability part 1	How many fatal road accidents happened in Ballina in the North of County Mayo in 2010? The correct answer is 2.	Group 1: VIS-1 and VIS-2 Group 2: VIS-1 and VIS-2 Group 3: VIS-2 Group 4: VIS-1
#06	User perception	In your opinion, which visualisation conveys the total number of accidents in a specific county best?	VIS-1 and VIS-3
#07	User perception	In your opinion, which visualisation conveys the spread of accidents in County Cork best?	VIS-1 and VIS-2
#08	User perception	If somebody would ask you a question on road collisions in Ireland, which would visualisation would you most likely open for reference?	All visualisations.
#09	Memorability part 2	Can you remember the total amount of road accidents in Dublin / Mayo in 2010?	No visualisation for help allowed. Group 1: Mayo Group 2: Mayo Group 3: Dublin Group 4: Dublin
#10	Memorability part 2	Can you remember how many fatal road accidents happened in the city of Limerick / Ballina in 2010?	No visualisation for help allowed. Group 1: the city of Limerick Group 2: the city of Limerick Group 3: Ballina Group 4: Ballina

Table 2: Questions of the usability and memorability experiment – The questions investigate different aspects of the visualisations such as response time and accuracy, user perception and memorability. The participants were split into four different groups to investigate memorability.

The participant is not told in advance that the experiment has a third part, in which these two questions will be repeated and answered without the help of the visualisations. The timing is done in order to not raise suspicion. The participant shall be as unprepared and impartial for the memorability part as possible so that their memorability is exclusively influenced by the features of the visualisations. The

memorability questions are a subset of two different questions out of the five for each participant group.

The questions #01, #02, #03 and #05 in Table 2 are the questions that investigate either aspect response time and accuracy or memorability for the four different user groups. Question #04 is the only question that investigates the participant's response time and accuracy only.

In part two of the experiment, three questions compares the user's perception of the visualisations. The questions aim to ask the participant directly for their opinion on certain features of the visualisations. Questions #06, #07 and #08 in Table 2 cover this aspect of the experiment. However they are not presented to the participants in exactly this order: Question #07 refers to data that is required for question #04 as well and hence it is asked right after question #04.

The full protocol form that was used during the experiment can be found in Appendix A.

### 4.3 Participants

Eight volunteers, involved in the academic course for which this work was carried out, participated in the experiment. For this reason, it was possible to put two participants into each participant group. The classification was conducted in the order of the participation: The first participant was part of group 1, the second participant was part of group 2, and the fifth participant was part of group 1 again.

The participants were all familiar with the computer equipment and the tools provided in the experiment. Specifically, they were asked, if they have used a computer, Microsoft Windows 7, Microsoft Office Excel 2010, Google Chrome and Google Maps before.

Furthermore, 75% of the participants were Irish, so that a basic knowledge of the Irish geography could be assumed. However, the help of the instructor in relation to geological location of certain places was not restricted to non-Irish participants. The other 25% of the participants have lived for an average of three years in Ireland, before they participated in the experiment. Finally the participants were between aged 23 and 35 years old, while the average age was 26 years of age.

### 4.4 Results and Evaluation

In order to be able to analyse the hand-written protocols, the responses were summarised digitally in an Excel spreadsheet. From this, percentages and average quotes could be calculated for each question. The results are illustrated in the diagrams in Figure 19 to Figure 21 and will now be discussed.

Firstly looking at Figure 19, which compares the response accuracy of the visualisations, it can be seen that all three performed well. The highest accuracy of 72.2% was achieved by VIS-2, while VIS-1 achieved 70% and VIS-3 achieved 66.7%.

The overall results for VIS-1 and VIS-2 are very close in that they both got the same results in question #03 to #05. This shows that the different features of both visualisations did not affect the user's accuracy. In fact, because they visualise a single road accident in a very similar way, it could be observed that wrong answers were given for the same reasons. For example, question #05 was only answered correctly by a third of the participants, because it is not clear in both visualisations which of the two fatal accident indicators in the area of Ballina actually count towards Ballina. VIS-1, however, would be able to clarify this question as it shows the name of the place of an accident when the user hovers over the marker of the accident on the map. Nevertheless, none of the participants made use of this feature. It could be stated that this feature is not clear enough to the user within the visualisation. Otherwise the visualisation could have reached a higher accuracy for both question #05 and question #02.

Furthermore, the reason for answering incorrectly to question #03 that asked for the total amount of accidents in County Mayo in VIS-2, was mainly due to the complicated process that is required in order to find the correct answer. The obvious solution is to filter the collisions by the correct year and click "All" for the severity and then highlight the county in the map. Most participants expected to see the desired details in the "Collision information" section (see Figure 14). However the visualisation does not provide information here, if "All" is selected for severity. The correct process involves repeating the steps named above for all three severity types "Fatal", "Serious" and "Minor" and to add the three numbers together. Hence there are too many steps involved and this resulted in many of the participants feeling doubtful about the visualisation. There is no obvious reason why the visualisation does not provide the information as most participants expected.

210	6 Mountjoy	53°
211	6 Finglas West	53°
214	6 Baldoyle	53°
217	6 Finglas	53°
232	6 Bridewell	53°

Single collisions

Ready 1387 of 1048575 records found

Figure 18: Label informing about the number of filtered results in Microsoft Excel 2010 – The label is positioned in the bottom left hand-side corner of the window and was not found by many participants.



In contrast, VIS-1 provides the information for question #03 in one click: The user needs to click the correct wedge in the sunburst chart and the information box in the centre of the chart updates to present the required information. However, the participants that did not get this answer right with the help of VIS-1 tried to acquire the information from the map and added up the labels of the clusters within County Mayo. Nevertheless, the MarkerClusterer that was explained in Chapter 3.2 does not consider county boundaries. It puts accidents that are located in different counties into the same cluster, if they are close enough to each other. For this reason the wrong answers for question #03 that were given with the help of VIS-1 were mostly higher than the correct answer.

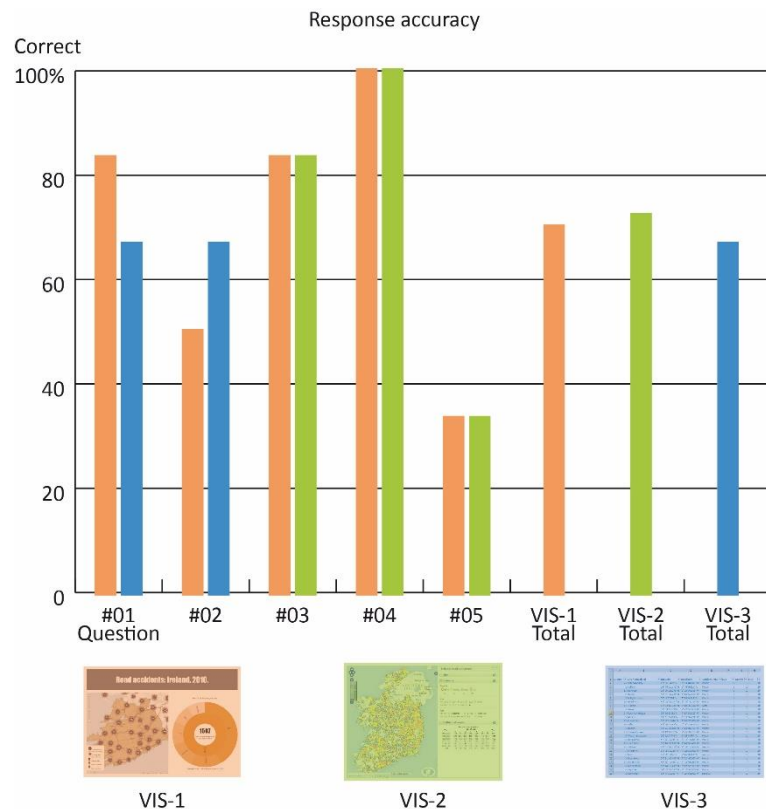


Figure 19: Response accuracy results – The diagram shows how much percentage of the participants answered the questions #01 to #05 correctly. The different colours indicate, which visualisation was used by the participants, while answering a question. The diagram also summarizes how accurate the response where for each visualisation in total.

VIS-3 achieved a constant result of 83.3% accuracy. In question #01 the mistakes were made even though the filters for the relevant column were used extensively and correct. The reason was mostly due to the participant not being able to find the small label that is illustrated in Figure 18 in the bottom left-hand side corner of the Excel window that informs about the total number of results of the applied filters. Without the knowledge of this label, the participant needed to highlight all results after the filters were applied and was then able to find a similar label to the first one in the bottom right-hand

side corner of the window. These steps were highly error prone for the high number of accidents in County Dublin.

Another problem with VIS-3 that caused incorrect answers for question #02 was that it does not strictly differentiate between differently spelled versions of the same place. Hence most wrong answers for question #02 that were acquired with the help of VIS-3 were too low as the participants only selected accidents for the city name “Limerick City” and they ignored the fact that there were also entries for “Limerick” that also counted towards the correct answer.

The three different visualisations lost percentages for accuracy for various different reasons. However, their different features that either helped the participants to answer correctly or that confused them and made them answer incorrect, balanced each other out. Hence the three visualisations achieved similar overall results and can be seen as equally good when it comes to response accuracy.

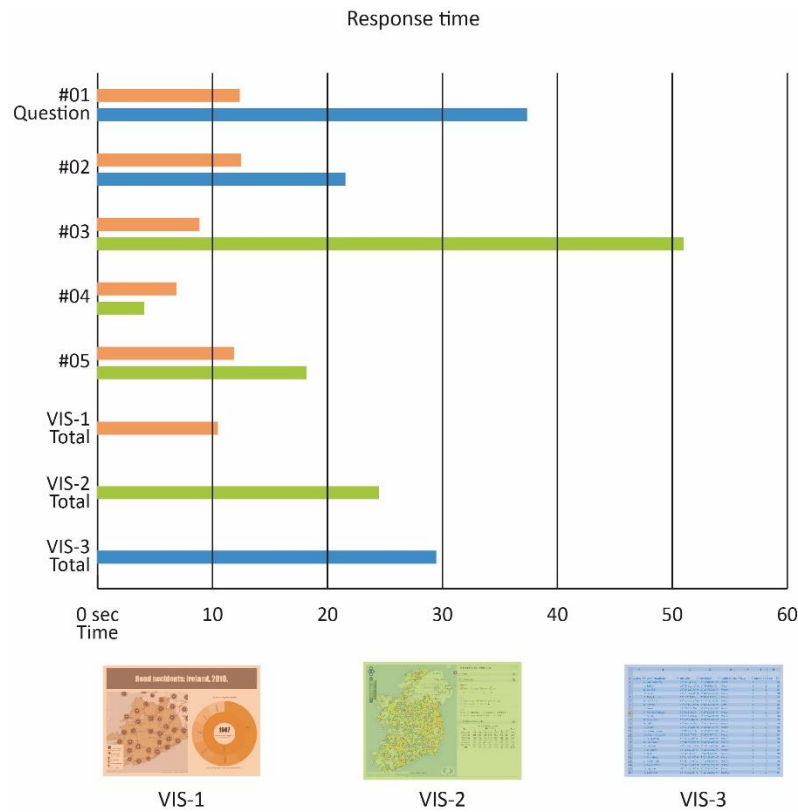


Figure 20: Response time results – The diagram shows the average response times of the participants for the questions #01 - #05. The colours indicate, which visualisation was used by the participants, while answering a question. The diagram also summarises how quickly the average response where for each visualisation.

A clearer difference can be found in the results of the response times that are illustrated in Figure 20. As the idea for VIS-1 was initially found because of the long and confusing steps of VIS-2, VIS-1 was

among other aims designed<sup>4</sup> to make the user get to important information quicker than in VIS-2. This aim was totally achieved as questions answered with the help of VIS-1 achieved an average response time of ten seconds. Questions for VIS-2 were answered in an average time of 24 seconds, while answers found in VIS-3 took an average of 29 seconds.

VIS-1 helped the participants to find the information quicker than the compared visualisation in every test question bar one. In question #04 the participants were asked to highlight County Cork. Every participant clicked immediately into the area of the map, which they assumed was the county. This was the correct step for VIS-2, however in VIS-1, this is not a feature and the user needs to click the corresponding wedge in the sunburst chart. VIS-1 does not fulfil the participants' expectations here. However it is not clear, if the participants' expectations were influenced by the behaviour of VIS-2. This is because following the complicated steps of VIS-2 involved in question #03, the participants had to highlight a county three times in VIS-2, right before they were asked to do the same in VIS-1.

The complicated steps in question #04 cost the participants a lot of time and hence the average response time is the highest in the whole test. VIS-1 offers the required information after one click. However participants that did not use the label in the centre of the sunburst diagram, and instead tried to add up the cluster labels within the county, which took more time and led to the wrong answer as explained above.

Answering question #05 with VIS-2 cost more time than with VIS-1 for two reasons: Firstly, the participant had to set the year first, and secondly the zoom option of VIS-2 is not clearly visible. The visualisation lets the user do a zoom by a double click in to the map only twice. After this the zoom slider in the top left-hand side corner of the map needs to be used for zooming. This confused most of the participants as they were used to the double click zoom from previous Google Maps applications.

The problems named above also affected the response times of questions #01 and #02 as they asked for similar information.

Clear differences as well became evident in the memorability results, which are illustrated in Figure 21. It first needs to be stated that the accuracy of the answers in the first part of the test did not influence the memorability results. It was simply tested if the first answer of the participant agreed with the answer they gave, when they were asked to repeat it without the help of a visualisation.

The quick workflows in VIS-1 made participants not engage with the numbers they answered long enough. This mirrors the memorability result of the visualisation: Only 37.5% of the participants could

---

<sup>4</sup> The design decisions are explained in Chapter 3.1.

remember their answer of a question that they answered with the help of VIS-1. In fact, they were only able to remember the answer of question #10, which asked for a small number such as 1 or 2. None of the participants were able to remember the amount of accidents in a specific county, which was subject to question #09.

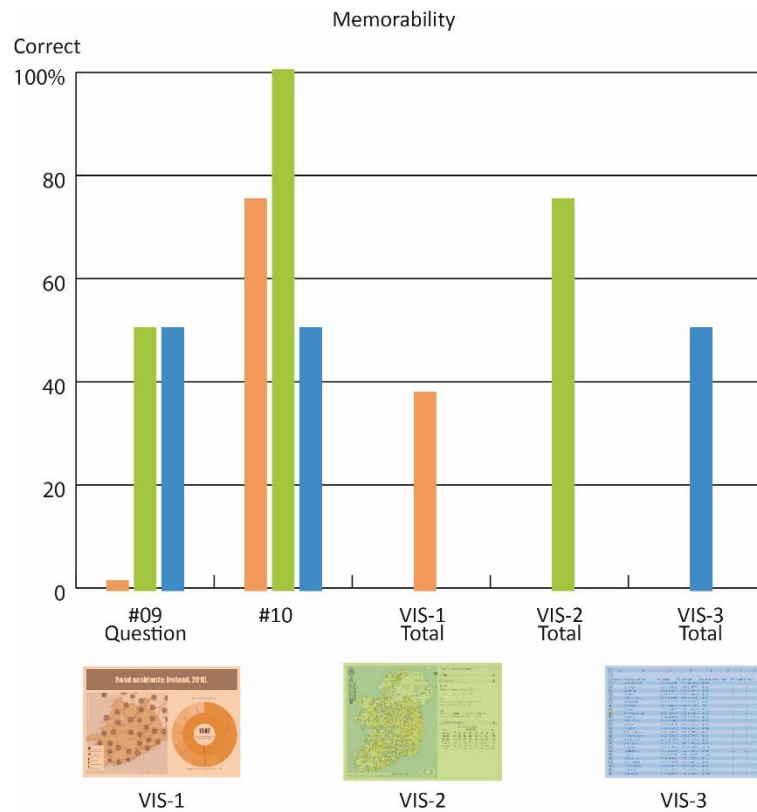


Figure 21: Memorability results – The diagram shows how much percentage of the participants could remember the answers of questions #09 and #10, which they gave in the first part of the experiment. The colours indicate, which visualisation was used by the participant to answer the question in the first part of the experiment. It needs to be mentioned that VIS-1 in question #09 scored 0%, but is displayed by a higher value in the diagram. This was done in order to make it visible enough, so that it is obvious that VIS-1 was also subject of question #09.

VIS-2, the other hand, supported the participants' memorability best. 75% of the participants overall were able to remember their answers from before. Every participant was able to remember the small amount of accidents that were subject to question #10, while some were incorrect with the high numbers requested in question #09. The fact that VIS-2 achieved the highest results can be explained by the combination of visual representation and the longer steps that make the user engage with the answers for a longer period of time.

Finally, VIS-3 achieved a continuous result of 50% for both questions. It gives a certain visual representation by showing accident in a separate line, while its workflows are neither very short as in VIS-1 nor very long as in VIS-3. This combination was enough to help half of the participants to

remember, while it was not enough engagement to help the remaining other half of participants. This supports that tables or spreadsheets are helpful to remember information for certain types of people.

So overall the quick work flows in combination with the easy to read markers on the map and the colour-coded sunburst chart made VIS-1 achieve very high results in the user perception test. Participants commented on the markers in VIS-2 as too cluttered and the spreadsheet in VIS-3 as not appropriate for the large amount of accidents.

However, some participants voted for VIS-2 in direct comparison of VIS-1 and VIS-2: Question #07 asked, which visualisation conveys the spread of accidents in County Cork best. 25% of the participants answering this question said that although they found the map too full, being able to see all accident markers for the whole county makes it more obvious where the dangerous roads are.

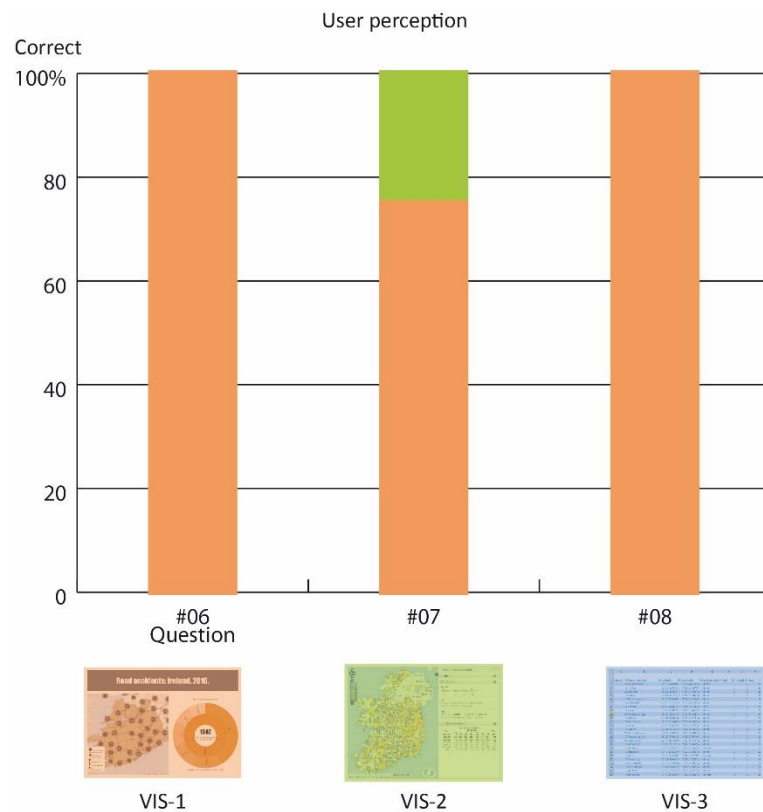


Figure 22: User perception results – The diagram shows how much percentage of the participants voted for which visualisation in the user perception questions #06 - #08. The colours indicate the visualisations. It should be stated that not every visualisation was subject in every question. In question #06 the participant chose between VIS-1 and VIS-3. In question #07 they decided between VIS-1 and VIS-2. Finally, only in question #08 the decision needed to be made between all three visualisations.

100% of the participants would consult VIS-1 again if they were asked a question on road accidents in Ireland in 2010 in the future.

This chapter presented an experiment that investigated the response accuracy, response time, memorability and use perception of three different visualisations. The first visualisation was created to particularly address problems in the other two visualisations regarding response time and user perception. The chapter explained the setup, the procedure, the content and the results of this experiment. It was found that a quick response time of a visualisation results in low memorability as users are not engaging long enough with the values to remember them. It was also found that an easy usability, assembled by a quick response time and a high response accuracy, results in a good user perception and that complicated steps lead to the opposite.

## 5 Conclusion

This work first introduced the current state of research in interactive data visualisation. The features that improve - according to the current research - the story of a visualisation, its interactivity, its memorability and its user experience were named and analysed in this part of the work.

These findings served as the basis for the next part of this work, where a new interactive data visualisation was designed and developed. Different web technologies and techniques were tested in order to create an appropriate visualisation that tells an engaging story with the data set of road accidents in Ireland in 2010. A visualisation that makes the data easily accessible, readable and memorable. The development process and the finalised tools and techniques were presented in this paper.

Afterwards an experiment on the data visualisation was presented that investigated response accuracy, response time, memorability and use perception of the visualisation in comparison to two other visualisations that were based on the same data set. The results of this experiment showed the relationship between a quick response time and a low memorability as well as a high usability and a good user perception.

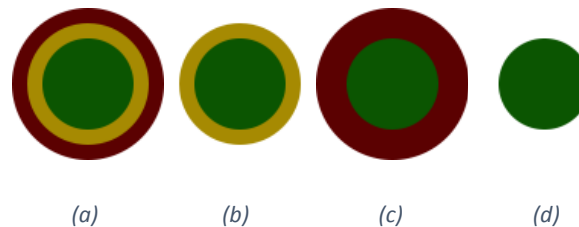
This paper contributed to the clarification of the question, what makes a good interactive data visualisation. It also helped to raise new research questions in the field of data visualisations, which will be explained in the following.

### 5.1 Future Work

In case of the “Road accidents: Ireland, 2010”, visualisation future work can be conducted by extending the data set by more years, especially recent years. The visualisation by the RSA, embraces the years 2005-2012. This means that data, which is younger than 2012, is not available in visualised form yet. The integration of more data sets is made easy by the decision to structure the data for the visualisation in JSON objects. However, the main task, when including more years, is to find an appropriate visual representation that allows the user to switch between years or even compare different years.

Furthermore a marker management as explained in by Mahe & Broadfoot (2010) can be implemented. This includes not rendering markers that are not in the viewport of the map and to cluster markers in a more intelligent way than grid-based clustering allows, i.e. clustering by type or in case of the visualisation of this work by severity of the accident. A marker management aims at improving performance when handling many markers and saving bandwidth. The performance has not been an

issue in the visualisation anymore since the FusionTablesLayer was introduced. However, it could become an issue again, when including data of more recent years. Implementing the marker management would be good-practice and would improve scalability of the visualisation. Clustering not just by distance but also by severity could be used to customise the cluster icon depending on how many fatal, serious or minor accidents are contained in a cluster. This is visualised in Figure 23.



*Figure 23: Cluster marker icon suggestion – Icon (a) visualises a cluster marker that contains all three severity types: minor (green), serious (yellow) and fatal (red). The cluster icon (b) visualises that no fatal accidents are contained in the cluster, while the cluster with icon (c) does not contain a serious accident. Finally the icon (d) applies to clusters that contain minor accidents only. Illustrated are the most common*

The cluster algorithm could furthermore be improved by only clustering accidents within the boundaries of a county. This would better convey what users expect of the cluster markers as the experiment in Chapter 4 has shown.

Additionally future work regarding the design of the visualisation should rethink the colour scheme of the markers. Participants noticed difficulties to differentiate between the colours of markers, especially when the county was highlighted in its colour by the FusionTablesLayer. This also applies to the cluster icon suggestions above in Figure 23 as colour blindness could impair certain users from reading the information conveyed by the markers correctly.

Future work on the experiment on the other hand, should be performed with a higher number of participants. A wider spread of age, computer knowledge, nationality, cultural and professional background will help the results to reflect a real world scenario better.

Further investigation should also include the rearrangement of the questions and tasks in the experiment, so that the learning effect of the previous question does not affect the results of the next question as seen in question #04 (see Chapter 4.4).

Another extension to the experiment could be to let questions not only start with the initial view. This means that the participants will not be asked to refresh the visualisation before every question. Certain questions could build on the interaction that happened in the previous question. This will simulate a more realistic workflow, where more than one question in a row is answered with the visualisation.



This approach investigates if a visualisation is still user friendly after the user has triggered certain changes through the interactive features. This is a research question for interactive data visualisations that has not been investigated extensive enough before. As an example, considering the participants did not reset the filters in the spreadsheet visualisation, the experiment as carried out above leads to an obvious problem: For the next question the participant will have to open every filter dialog in order to set it to the new requirements. This workflow, depending on the nature of the questions, could include many more clicks than filtering from the initial view.

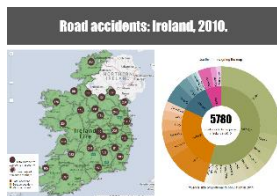
Furthermore, regarding the influences on the memorability of the visualisations in a future experiment another step should be included that investigates long term memorability. This means that participants would be asked to repeat the answer of a question after a period of a few hours or a few weeks. Long term memorability was not the primary aim when designing the “Road accidents: Ireland, 2010” visualisation, and taking it into consideration for the experiment was beyond the scope of this work. However, the effect of certain features of interactive data visualisation on long term memorability has yet to be completely clarified at the current state of research and requires further investigation.

Finally regarding the part of the experiment that tested the user perception, in a future experiment the question should be answered as a rating of the visualisation instead of a single-choice question. This needs to be considered as participants commented that they do not find it easy to commit exclusively to one visualisation for each question, while answering the user perception questions.

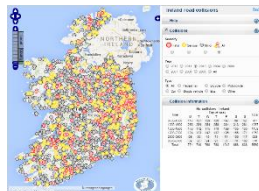
All in all the work for this paper added to the clarification on the features of interactive data visualisations that make a good visualisation and enhance comprehension and memorability of the presented information. The connection between a short response time and a low memorability was shown as well as the influence of usability on user perception was established. However, there is still a need for a substantial amount of research and investigation on interactive data visualisation. The almost unpredictable paths that users take through an interactive visualisation in comparison to a static visualisation require new ways of conducting experiment results and an appropriate set up has yet to be found as shown above.

# Abbreviations

- API: Application Programming Interface
- CSV: Comma Separated Values
- D3: Data Driven Documents
- DOM: Document Object Model
- HTML: HyperText Markup Language
- JSON: JavaScript Object Model
- SQL: Structured Query Language
- VIS-1: Visualisation one refers to the “Road accident: Ireland 2010” interactive data visualisation.



- VIS-2: Visualisation two refers to the interactive data visualisation titled “Ireland Road Collision web interface” by the Irish Road Safety Authority.



- VIS-3: Visualisation three refers to the spreadsheet visualisation for road accidents in Ireland in 2010.

	A	B	C	D	E	F	G	H	I
1	County / District	Latitude	Longitude	Accident Type	Day	Month	Year		
2	Dublin City	53.34987	-6.26031	Minor	1	1	10		
3	Dublin	53.34987	-6.26031	Minor	1	1	10		
4	Dublin	53.34987	-6.26031	Minor	1	1	10		
5	Dublin	53.34987	-6.26031	Minor	1	1	10		
6	Dublin	53.34987	-6.26031	Minor	1	1	10		
7	Dublin	53.34987	-6.26031	Minor	1	1	10		
8	Dublin	53.34987	-6.26031	Minor	1	1	10		
9	Dublin	53.34987	-6.26031	Minor	1	1	10		
10	Dublin	53.34987	-6.26031	Minor	1	1	10		
11	Dublin	53.34987	-6.26031	Minor	1	1	10		
12	Dublin	53.34987	-6.26031	Minor	1	1	10		
13	Dublin	53.34987	-6.26031	Minor	1	1	10		
14	Dublin	53.34987	-6.26031	Minor	1	1	10		
15	Dublin	53.34987	-6.26031	Minor	1	1	10		
16	Dublin	53.34987	-6.26031	Minor	1	1	10		
17	Dublin	53.34987	-6.26031	Minor	1	1	10		
18	Dublin	53.34987	-6.26031	Minor	1	1	10		
19	Dublin	53.34987	-6.26031	Minor	1	1	10		
20	Dublin	53.34987	-6.26031	Minor	1	1	10		
21	Dublin	53.34987	-6.26031	Minor	1	1	10		
22	Dublin	53.34987	-6.26031	Minor	1	1	10		
23	Dublin	53.34987	-6.26031	Minor	1	1	10		
24	Dublin	53.34987	-6.26031	Minor	1	1	10		
25	Dublin	53.34987	-6.26031	Minor	1	1	10		
26	Dublin	53.34987	-6.26031	Minor	1	1	10		
27	Dublin	53.34987	-6.26031	Minor	1	1	10		
28	Dublin	53.34987	-6.26031	Minor	1	1	10		
29	Dublin	53.34987	-6.26031	Minor	1	1	10		
30	Dublin	53.34987	-6.26031	Minor	1	1	10		

---

## References

- Bateman, S. et al., 2010. Useful junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. New York, USA: ACM Press, pp. 2573–2582.
- Borkin, M.A. et al., 2013. What makes a visualization memorable. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), pp.2306–2315.
- Bostock, M., Ogievetsky, V. & Heer, J., 2011. D<sup>3</sup>: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), pp.2301–2309.
- CACM Staff, 2014. Visualizations make big data meaningful. *Communications of the ACM*, 57(6), pp.19–21.
- Card, S.K., Mackinlay, J.D. & Shneiderman, B., 1999. *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann.
- Central Statistics Office, 2011. Population of each Province, County and City, 2011. Available at: <http://www.cso.ie/en/statistics/population/populationofeachprovincecountyandcity2011/> [Accessed April 29, 2015].
- Cukier, K.N. & Mayer-Schoenberger, V., 2013. The Rise of Big Data. *Foreign Affairs*. Available at: <http://www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data> [Accessed March 3, 2015].
- Driscoll, M., 2010. Color: The Cinderella of Data Visualization. In J. Steele & N. Iliinsky, eds. *Beautiful Visualization*. O'Reilly Media, Inc., p. 416.
- Few, S., 2011. The Chartjunk Debate - A Close Examination of Recent Findings. *Visual Business Intelligence Newsletter*, (April, May and June).
- Fry, B., 2004. *Computational Information Design*. Massachusetts Institute of Technology.
- Fry, B., 2008. *Visualizing Data* 1st ed. A. Oram, ed., USA: O'Reilly.
- Gershon, N. & Page, W., 2001. What storytelling can do for information visualization. *Communications of the ACM*, 44(8), pp.31–37.
- Gonzalez, H. et al., 2010. Google Fusion Tables: Web-Centered Data Management and Collaboration. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, pp. 1061–1066.
- Google Code, 2015. MarkerClusterer. *Google Maps Utility Library*. Available at: <http://google-maps-utility-library-v3.googlecode.com/svn/trunk/markerclusterer/> [Accessed May 5, 2015].
- Google Developers, 2013. Google Maps API Version 3. Available at: <https://developers.google.com/maps/> [Accessed April 29, 2015].
- Google Developers, 2015. Google Maps JavaScript API v3. Available at: <https://developers.google.com/maps/documentation/javascript/> [Accessed May 5, 2015].

- Holmes, N., 2010. Monstrous Costs, R. Kosara, ed. *eagereyes.org*. Available at: <https://eagereyes.org/criticism/chart-junk-considered-useful-after-all> [Accessed May 13, 2015].
- Janvrin, D.J., Raschke, R.L. & Dilla, W.N., 2014. Making sense of complex data using interactive data visualization. *Journal of Accounting Education*, 32(4), pp.31–48.
- Johnson, B. & Shneiderman, B., 1991. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91*.
- Kosara, R. & Mackinlay, J., 2013. Storytelling: The Next Step for Visualization. *Computer*, 46(5), pp.44–50.
- Kress, G.R. & van Leeuwen, T., 2006. *Reading Images: The Grammar of Visual Design*, Routledge.
- Madhavan, J. et al., 2012. Big Data Storytelling Through Interactive Maps. *IEEE Data Eng. Bull.*, 35, pp.46–54.
- Mahe, L. & Broadfoot, C., 2010. Too Many Markers! *Google Developers*. Available at: <https://developers.google.com/maps/articles/toomanymarkers> [Accessed May 4, 2015].
- Manovich, L., 2002. Data Visualisation as New Abstraction and Anti-Sublime.
- Manovich, L., 2011. What is visualisation? *Visual Studies*, 26(1), pp.36–49.
- McCandless, D., 2015a. Information Is Beautiful. Available at: <http://www.informationisbeautiful.net/> [Accessed May 9, 2015].
- McCandless, D., 2009. *Information is Beautiful*, Collins.
- McCandless, D., 2014. *Knowledge is Beautiful*, HarperCollins Publishers.
- McCandless, D., 2015b. Plane Truth - Every single commercial plane crash of the last 20 years. *Information Is Beautiful*. Available at: <http://www.informationisbeautiful.net/visualizations/plane-truth-every-single-commercial-plane-crash-visualized/> [Accessed May 12, 2015].
- Minard, C.J., 1869. Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813. *Regnier et Dourdet, Paris*. Available at: <https://datavizblog.files.wordpress.com/2013/05/map-full-size1.png> [Accessed May 13, 2015].
- Natural Earth, 2011. States and Provinces. *Google Fusion Tables*. Available at: <https://www.google.com/fusiontables/DataSource?dsrclid=420419#rows:id=1> [Accessed April 29, 2015].
- OpenStreetMaps, 2015. Ireland/Boundaries. *OpenStreetMaps Wiki*. Available at: <http://wiki.openstreetmap.org/wiki/Ireland/Boundaries> [Accessed April 29, 2015].
- Pfeiffer, A. et al., 2015. Document Object Model (DOM). *Mozilla Developer Network*. Available at: [https://developer.mozilla.org/en-US/docs/Web/API/Document\\_Object\\_Model](https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model) [Accessed May 5, 2015].

- Road Safety Authority Ireland, 2015. Ireland Road Collision web interface. *rsa.ie*. Available at: <http://rsa.ie/en/RSA/Road-Safety/Our-Research/Ireland-Road-Collisions/> [Accessed April 29, 2015].
- Roberts, J.C. et al., 2014. Visualization beyond the Desktop - the Next Big Thing. *IEEE Computer Graphics and Applications*, 34(6), pp.26–34.
- Rodden, K., 2014. Applying a sunburst visualization to summarize user navigation sequences. *IEEE Computer Graphics and Applications*, 34(5), pp.36–40.
- Rodden, K., 2013. Zoomable sunburst with updating data. Available at: <http://bl.ocks.org/kerryrodde/477c1bfb081b783f80ad> [Accessed January 8, 2015].
- Rosenbloom, S., 2010. But Will It Make You Happy? *NYTimes.com*. Available at: [http://www.nytimes.com/2010/08/08/business/08consume.html?pagewanted=all&\\_r=0](http://www.nytimes.com/2010/08/08/business/08consume.html?pagewanted=all&_r=0) [Accessed May 13, 2015].
- Rosenbloom, S., 2011. But Will It Make You Happy? Data visualisation with Cascade. *NYTLabs.com*. Available at: <http://nytlabs.com/projects/video5.php?file=http://cdn.nytlabs.com/movies/zappos.m4v&w=852&h=480> [Accessed May 13, 2015].
- Segel, E. & Heer, J., 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp.1139–48.
- Shneiderman, B., 2008. Extreme visualization: squeezing a billion records into a million pixels. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. ACM, pp. 3–12.
- Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*.
- Shneiderman, B., 1992. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1), pp.92–99.
- Stasko, J. et al., 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53(5), pp.663–694.
- Stasko, J. & Zhang, E., 2000. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*. IEEE Comput. Soc, pp. 57–65.
- Thorp, J., 2011. Cascade. *NYTLabs.com*. Available at: <http://nytlabs.com/projects/cascade.html> [Accessed May 13, 2015].
- Thorp, J., 2012. The Weight of Data. *TEDx Talks*. Available at: [http://www.ted.com/talks/jonathan\\_harris\\_tells\\_the\\_web\\_s\\_secret\\_stories?language=en](http://www.ted.com/talks/jonathan_harris_tells_the_web_s_secret_stories?language=en) [Accessed April 13, 2015].
- Tufte, E.R., 1990. *Envisioning information*, Graphics Press.
- Tufte, E.R., 1983. *The Visual Display of Quantitative Information*, Graphics Press.

## Appendix A

# Usability Testing

Test outline and protocol

## Participant details

<b>Number</b>	
<b>Age</b>	
<b>Nationality (Irish or other)</b> <b>If other, lived in Ireland for how long?</b>	
<b>Familiar with presented equipment / tools</b>	<input type="checkbox"/> Computer (Screen, mouse) <input type="checkbox"/> Windows 7 <input type="checkbox"/> Microsoft Excel 2010 <input type="checkbox"/> Google Chrome <input type="checkbox"/> Google Maps
<b>Group</b>	

## Introductory questions and tasks (not recorded)

Group 1	Group 2	Group 3	Group 4
Always VIS01			

**1. How many serious accidents happened in Waterford town in 2010?**

Group 1	Group 2	Group 3	Group 4
Always VIS02			

**2. How many minor road accidents happened in county Galway in 2010?**

Group 1	Group 2	Group 3	Group 4
Always VIS03			

**3. How many serious accidents happened in the city of Waterford in 2010?**

## Response Time & Accuracy (VIS01 vs. VIS03)

Group 1	Group 2	Group 3	Group 4
VIS01 + VIS03	VIS01 + VIS03	VIS01	VIS03

### #01 How many road accidents happened in county Dublin in 2010?

	Answer	Response time
VIS01		
VIS03		

Comments:

Group 1	Group 2	Group 3	Group 4
VIS01	VIS03	VIS01 + VIS03	VIS01 + VIS03

### #02 How many fatal road accidents happened in the city of Limerick in 2010?

	Answer	Response time
VIS01		
VIS03		

Comments:

## Response Time & Accuracy (VIS01 vs. VIS02)

Group 1	Group 2	Group 3	Group 4
VIS02	VIS01	VIS01 + VIS02	VIS01 + VIS02

### #03 How many road accidents happened in county Mayo in 2010?

	Answer	Response time
VIS01		
VIS02		

Comments:

Group 1	Group 2	Group 3	Group 4
Always VIS01 + VIS02			

### #04 Can you highlight the county Cork in both visualisations?

	Answer	Response time
VIS01		
VIS02		

Comments:

Group 1	Group 2	Group 3	Group 4
VIS01 + VIS02	VIS01 + VIS02	VIS02	VIS01

### #05 How many fatal road accidents happened in Ballina in the North of county Mayo in 2010?

	Answer	Response time
VIS01		
VIS02		

Comments:



## Memorability

Group 1	Group 2	Group 3	Group 4
Mayo	Mayo	Dublin	Dublin

**#09 Can you remember the total amount of road accidents in XXXXXXXX in 2010?**

Answer: \_\_\_\_\_

Comments:

Group 1	Group 2	Group 3	Group 4
the city of Limerick	the city of Limerick	Ballina	Ballina

**#10 Can you remember how many fatal road accidents happened in XXXXXXXX in 2010?**

Answer: \_\_\_\_\_

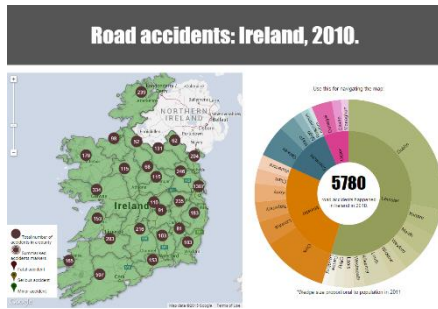
Comments:

# User Perception

**#06 In your opinion, which visualisation conveys the total number of accidents in a specific county best?**

VIS01

VIS03

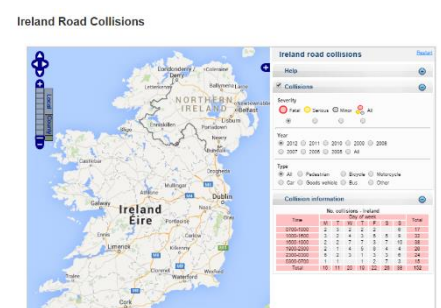
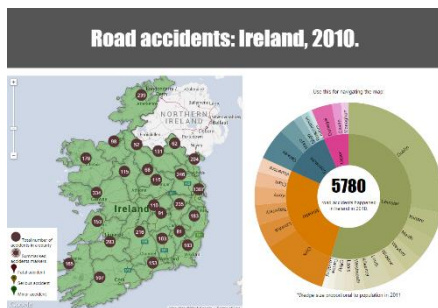


County	City/Townland	Latitude	Longitude	Accident Type	Day	Month	Year
23	Ballinacorney	52° 5' 1.439"N	7° 33' 44.821"W	Minor	1	1	10
1	Tullow	52° 49' 15.072"N	0° 48' 8.410"W	Minor	1	1	10
15	Dunleely	53° 50' 36.278"N	0° 28' 56.803"W	Minor	1	1	10
5	Dublin	53° 21' 35.740"N	6° 16' 57.542"W	Minor	1	1	10
3	Ballydowen	55° 3' 7.588"N	7° 24' 4.119"W	Minor	1	1	10
7	Drogheda	53° 42' 21.743"N	0° 22' 15.770"W	Minor	1	1	10
8	Finnep	52° 33' 48.823"N	5° 10' 39.574"W	Fatal	1	1	10
2	Cavan	53° 59' 0.809"N	7° 22' 56.489"W	Minor	1	1	10
24	Rochehorbridge	54° 34' 6.531"N	7° 17' 35.328"W	Minor	1	1	10
4	Aherla	51° 51' 5.566"N	6° 44' 49.240"W	Minor	1	1	10
23	Burmahom	52° 8' 21.044"N	7° 20' 23.688"W	Minor	1	1	10
16	Balla	53° 42' 16.502"N	0° 8' 32.687"W	Minor	1	1	10
6	Dublin	53° 21' 33.183"N	6° 14' 44.760"W	Minor	1	1	10
23	Kilmachomas	52° 12' 8.931"N	7° 29' 44.403"W	Minor	1	1	10
8	Westmalinstown	53° 22' 7.867"N	0° 28' 0.746"W	Minor	1	1	10
23	Dungarvan	52° 5' 30.257"N	7° 37' 11.070"W	Minor	1	1	10
19	Kilcomac	53° 50' 33.150"N	7° 43' 51.498"W	Minor	2	1	10
15	Kobreen	52° 17' 54.818"N	0° 29' 17.096"W	Serious	2	1	10
5	Brinsford	55° 2' 25.257"N	7° 22' 59.847"W	Minor	2	1	10
19	Kilcomac	53° 50' 33.150"N	7° 43' 51.498"W	Minor	2	1	10
5	Muff	55° 4' 3.153"N	7° 16' 8.111"W	Serious	1	1	10
5	Drumcliffe	55° 11' 46.407"N	7° 29' 45.030"W	Minor	2	1	10
11	Mountbath	53° 07' 17.945"N	7° 28' 45.067"W	Minor	1	1	10
6	Meekstown	53° 28' 16.439"N	6° 17' 53.963"W	Minor	2	1	10
5	Raphoe	54° 52' 51.459"N	7° 32' 54.440"W	Minor	2	1	10
2	Ballinabert	54° 5' 55.368"N	7° 20' 31.017"W	Serious	2	1	10

**#07 In your opinion, which visualisation conveys the spread of accidents in county Cork best?**

VIS01

VIS02

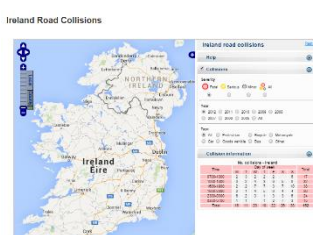
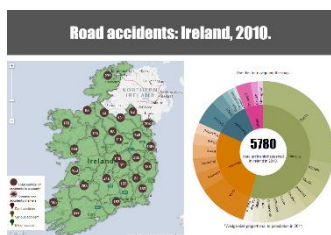


**#08 If somebody would ask you a question on road collisions in Ireland, which would visualisation would you most likely open for reference?**

VIS01

VIS02

VIS03



County	City/Townland	Latitude	Longitude	Accident Type	Day	Month	Year
23	Ballinacorney	52° 5' 1.439"N	7° 33' 44.821"W	Minor	1	1	10
1	Tullow	52° 49' 15.072"N	0° 48' 8.410"W	Minor	1	1	10
15	Dunleely	53° 50' 36.278"N	0° 28' 56.803"W	Minor	1	1	10
5	Dublin	53° 21' 35.740"N	6° 16' 57.542"W	Minor	1	1	10
3	Ballydowen	55° 3' 7.588"N	7° 24' 4.119"W	Minor	1	1	10
7	Drogheda	53° 42' 21.743"N	0° 22' 15.770"W	Minor	1	1	10
8	Finnep	52° 33' 48.823"N	5° 10' 39.574"W	Fatal	1	1	10
2	Cavan	53° 59' 0.809"N	7° 22' 56.489"W	Minor	1	1	10
24	Rochehorbridge	54° 34' 6.531"N	7° 17' 35.328"W	Minor	1	1	10
4	Aherla	51° 51' 5.566"N	6° 44' 49.240"W	Minor	1	1	10
23	Burmahom	52° 8' 21.044"N	7° 20' 23.688"W	Minor	1	1	10
16	Balla	53° 42' 16.502"N	0° 8' 32.687"W	Minor	1	1	10
6	Dublin	53° 21' 33.183"N	6° 14' 44.760"W	Minor	1	1	10
23	Kilmachomas	52° 12' 8.931"N	7° 29' 44.403"W	Minor	1	1	10
8	Westmalinstown	53° 22' 7.867"N	0° 28' 0.746"W	Minor	1	1	10
23	Dungarvan	52° 5' 30.257"N	7° 37' 11.070"W	Minor	1	1	10
19	Kilcomac	53° 50' 33.150"N	7° 43' 51.498"W	Minor	2	1	10
15	Kobreen	52° 17' 54.818"N	0° 29' 17.096"W	Serious	2	1	10
5	Brinsford	55° 2' 25.257"N	7° 22' 59.847"W	Minor	2	1	10
19	Kilcomac	53° 50' 33.150"N	7° 43' 51.498"W	Minor	2	1	10
5	Muff	55° 4' 3.153"N	7° 16' 8.111"W	Serious	1	1	10
5	Drumcliffe	55° 11' 46.407"N	7° 29' 45.030"W	Minor	2	1	10
11	Mountbath	53° 07' 17.945"N	7° 28' 45.067"W	Minor	1	1	10
6	Meekstown	53° 28' 16.439"N	6° 17' 53.963"W	Minor	2	1	10
5	Raphoe	54° 52' 51.459"N	7° 32' 54.440"W	Minor	2	1	10
2	Ballinabert	54° 5' 55.368"N	7° 20' 31.017"W	Serious	2	1	10

	Connacht
7	Galway
12	Leitrim
16	Mayo
20	Roscommon
21	Sligo
	Leinster
1	Carlow
6	Dublin
9	Kildare
10	Kilkenny
11	Laois
14	Longford
15	Louth
17	Meath
19	Offaly
24	Westmeath
25	Wexford
26	Wicklow
	Munster
3	Clare
4	Cork
8	Kerry
13	Limerick
2	Tipperary
23	Waterford
	Ulster
5	Donegal
2	Cavan
18	Monaghan

---

## Appendix B

Trinity College Dublin  
School of Computer Science and Statistics

# Consent from

Usability test of “Road accidents: Ireland, 2010.” (Interactive Data Visualisation)

The usability test investigates the accuracy and the response time of tasks carried out on a newly developed interactive data visualisation in comparison to established visualisations. It furthermore investigates the user’s perception, while using it. During the test data will be exclusively gathered in form of written notes by the instructor. This research may benefit the field of data visualisation by clarifying user behaviour and expectations. The usability test is part of a research project carried out by Alessa Gross and supervised by David Kaneswaran. Alessa Gross is also the instructor in this test.

I am participating voluntarily in this usability test and I know that I can withdraw my participation at any time. I know that any information or data, which is obtained from me during the test, is treated anonymously and confidentially.

If I have any questions about this research, I know I can ask Alessa Gross ([grossa@tcd.ie](mailto:grossa@tcd.ie)) or David Kaneswaran ([kaneswad@tcd.ie](mailto:kaneswad@tcd.ie)).

I have been given a copy of this consent form. I understand what this test is about and I agree to participate.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of instructor