CENSUS IRELAND

DATA VISUALISATION OF CENSUS & REFERENDUM STATISTICS USING TABLEAU

Jonathan Dempsey – N00112462
Web Engineering/Multimedia Systems – DL824
April 4, 2016

Declaration of Plagiarism

The incorporation of material without formal and proper acknowledgement (even with no deliberate intent to cheat) can constitute plagiarism.

If you have received significant help with a solution from one or more colleagues, you should document this in your submitted work and if you have any doubt as to what level of discussion/collaboration is acceptable, you should consult your lecturer or the Course Director.

WARNING: Take care when discarding program listings lest they be copied by someone else, which may well bring you under suspicion. Do not to leave copies of your own files on a hard disk where they can be accessed by other. Be aware that removable media, used to transfer work, may also be removed and/or copied by others if left unattended.

Plagiarism is considered to be an act of fraudulence and an offence against Institute discipline.

Alleged plagiarism will be investigated and dealt with appropriately by the Institute. Please refer to the Institute Handbook for further details of penalties.

The following is an extract from the B.Sc. in Computing (Hons) course handbook. Please read carefully and sign the declaration below

Collusion may be defined as more than one person working on an individual assessment. This would include jointly developed solutions as well as one individual giving a solution to another who then makes some changes and hands it up as their own work.

DECLARATION:

I am aware of the Institute's policy on plagiarism and certify that this thesis is my own work.

Student : Jonathan Dempsey

Signed_____

Failure to complete and submit this form may lead to an investigation into your work.

Acknowledgements

I would firstly like to express my gratitude to my project supervisor Cyril Connolly for his continuous support and guidance throughout the duration of this project. His help and advice is greatly appreciated.

I would also like to extend my gratitude to my fellow classmates for the encouragement and wisdom they have offered over the course of the project.

Finally, I would like to thank my family for their unwavering love and support throughout my years of study in IADT. None of this would be possible without them.

Abstract

This report details the research and development of data visualisations in a web application which displays visualized census and referendum data. The report will document the approach taken in determining the feasibilities and technologies required for the visualisations, as well as steps taken towards implementing the end product.

The visualisations are created in Tableau, which extracts data from Excel worksheets for each dataset. The census data was primarily acquired through the CSO (Central Statistics Office) website while the referendum data was acquired from Referendum Ireland. The visualisations are uploaded and stored on Tableau Public's server and are delivered to the front end web application in HTML using the embed code that Tableau Public provides for each visualisation. The site is designed using CSS while the dashboards are designed using Tableau.

Testing the system was done through unit and integration testing. Each unit was developed on its own and then implemented one by one to ensure they all worked in the final iteration of the project. Many of the potential pitfalls and bugs were found early on but through extensive user testing additional bugs were found, that some would later be fixed. The user tests were done after the main application was completed using questionnaires and test tasks.

Table of Contents

Declaration of Plagiarism 1

Acknowledgements 2

Abstract 3

1 Introduction 7

2 Research & Analysis 10

2.1 Introduction 10

2.2 The Components of Data Visualisation 10

2.2.1 Brief History 10

2.2.2 Modern Process to Creating Data Visualisations 11

2.3 Qualitative & Quantitative Data 15

2.3.1 Mixed Methods 16

2.4 Usability 17

2.4.1 Gestalt Principles 17

2.5 Tableau 20

2.6 Similar Product Analysis 21

2.6.1 AIRO Content 21

2.6.2 Tableau Public Server Content 22

2.7 User Requirements 23

2.8 Task Analysis 25

2.9 Functional Requirements 26

2.10 Non-Functional Requirements 27

2.11 Technical Requirements 27

2.12 Implementation Platforms 27

2.13 Data Sources 27

2.14 Potential Issues 28

2.14.1 Server Side Lag 28

2.14.2 Time 28

2.14.3 Testing 28

2.14.4 Front End 28

2.14.5 Backup Platforms 29

2.15 Summary 29

3 Design 31
3.1 Introduction 31
3.2 Description of Proposed System 31
3.2.1 System Model *31*
3.2.2 Server Side *31*
3.2.3 Client Side *32*
3.3 System Architecture Design 32
3.4 Tableau 32
3.4.1 Dashboards *33*
3.4.2 Data *33*
3.4.3 Backup *34*
3.5 Sequence Diagrams 34
3.6 User Interface Design 35
3.6.1 Example User Interfaces *36*
3.6.2 Prototype Design *38*
3.7 Final User Interface Design 40
3.7.1 Census Ireland Banner *40*
3.7.2 Navigation *41*
3.7.3 Tableau Dashboard Layout *41*
3.7.4 Colour Scheme *43*
3.8 Summary 43
4 Implementation 44
4.1 Introduction 44
4.2Tableau Polygon Map 44
4.3 Intergrating a Data Source and Data Mapping 46
4.4 Tooltip Data 47
4.5 Additional Graphs 49
4.6 Filtering 52
4.7 Creating Dashboards 55
4.8 Uploading to Tableau Public 55
4.9 Embedding a Dashboard to HTML 56
4.10 Issues 58

4.11 Summary 59
5 Testing 60
5.1 Introduction 60
5.2 Unit Testing 60
5.3 Quantitative Usability Goals 61
5.3.1 Goals Overview *61*
5.4 Test Materials 63
5.4.1 Pre-test Questionnaire *63*
5.4.2 Test Tasks *63*
5.4.3 Post-test Questionnaire *64*
5.5 Data Collection 64
5.6 Summary of Gathered Data 64
5.6.1 Analysis & Interpretation *64*
5.6.2 Statistical Analysis *66*
5.7 Summary 66
6 Conclusion 68
7 References 70
8 Appendices 74
8.1 Appendices A 74
8.2 Appendices B 75
8.3 Appendices C 76
8.4 Appendices D 78
8.5 Appendices E 81
8.6 Appendices F 82
8.7 Appendices G 83
8.8 Appendices H 85
8.9 Appendices I 86

1 Introduction

The primary goal is to create a website based application that enables users to view and filter data visualisations that display data from the 2011 census and the 2015 referendums. The website will focus on user interactivity and dynamic filtering to create an engaging experience for the user and display easy to understand data visualisations to determine trends in the data that would not be clear without the use of a visualisation.

The aim is to make it simple for users to learn more about Ireland on a county level. For example, users can view the marriage equality referendum dashboard to see which counties voted in favour or against together with other variables including; the electorate, total poll, invalid and valid ballot papers, total votes in favour and against, percentage turnout, in favour and against the referendum.

After researching various technologies that could achieve the goals and aims of the project, it was decided that Tableau would be used to create the visualisations. Tableau is a software platform that allows users to see and understand their data. It is an intuitive platform allowing dashboards and visual analytics to be displayed. As Tableau is becoming more widely used, and with easy access to their public servers which highlights the best visualisations created by its users, it will provide a vast array of visualisations to pull inspiration from and provide insight to Tableau's capabilities.

The project will focus on using Tableau's ability to draw polygon maps, which will be used to draw out the counties of Ireland and allow census and referendum data to be appended to them. Tableau will read in the data through Excel sheets downloaded from the CSO (Central Statistics Office) and Referendum Ireland websites.

The projects development will follow a schedule of specified intervals. Within these intervals, additional tasks will be carried out. However these additional tasks are much more flexible than the dates shown in *figure 1* and can be completed at any time within the given dates. These tasks can include things such as acquiring data, creating a graph in Tableau and binding data to a map. In *figure 1*, the project plan is displayed in a Gantt chart to assist with the time management of the project:

Requirements & Feasibility Document:	14th Sep – 30th Oct
Research and Outline Design:	30th Oct – 4th Dec
Midterm Presentation Preparation:	4th Dec – 6th Dec
Midterm Presentation:	7th Dec
Detailed Design Document:	8th Dec – 22nd Jan
Tested Source Code Due:	23rd Jan – 26th Feb

Testing and First Draft of Project Report: 27th Feb – 11th Mar

Final Project Report Due: 12th Mar – 4th Apr

Final Presentation: 9th or 10th May

Figure 1: Project Gantt Chart

Research and analysis:

The requirements and analysis' objective is to better understand the projects functional and non-functional requirements, gain background research and to learn what technologies will have to be implemented during development. The functional and non-functional requirements will determine what the system should do and how it should do it. Background research carried out will focus around data visualisation as well as researching publicly accessible data that the project will require. Finally, the technological implementation will identify the appropriate software platforms and frameworks that will be required to carry out the visualisation of the data that was gathered for the project.

Outline design:

The outline designs objective is to better understand the projects application logic design, the outline database and the user interface layout. This section will determine the database, interface and application logic that will be used in the system and attempt to gain a better grasp on the system as a whole.

Detailed design:

The detail designs objective is to specify the projects application logic design, the outline database and the user interface layout. After outlining the general design of the system, this section will delve into much greater depths of how the system will function.

Implementation:

The implementation's objective is to carry out the implementation of system components and the integration of system components. This will include joining all the separate units

of the system together to work as one, which will likely lead to bugs and errors that will require identifying and fixing.

Testing:

The Testing's objective is to carry out unit testing, system testing, the integration testing and the test plan. This will include error checking to ensure the system runs the way it should, testing each section alone and then as a whole and user testing the improve the projects usability.

Each of the intervals in the development cycle mentioned above will be detailed in this report. However, although they are described in the order they were completed, in the final iteration of this report their order may vary from chapter to chapter.

2 Research & Analysis

2.1 INTRODUCTION

Data visualisation is a way to help people better understand relationships and patterns by using visual context to display data which may have been missed if left in text format. The primary aim of data visualisation is to not only broadcast information data in a manner that is clearly understood but to also provide a stimulating experience that helps tell a story with the data through a graphical interface.

Through the use of internet, people have access to a vast amount of data based on any specific topic. However, the majority of this data isn't visualised in a practical way making it difficult to understand or potentially useless. In parallel to this, people also have access to many visualisation tools but lack the knowledge or skill to present it efficiently. With this lack of expertise *"designers often tend to discard the balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose – to communicate information" [1].* Not only do people tend to create inefficient graphs, but they also apply the incorrect sets of data to graphs that are not designed to display that dataset. In this chapter, exploration will be done on these visualisations and the data they use and ways in which these visualisations are created practically and effectively.

2.2 THE COMPONENTS OF DATA VISUALISATION

2.2.1 Brief History

To create functioning data visualisations, one must first understand the core principles and stages to produce data clearly and effectively. Although to fully understand this, it is important to know where the standards for data visualisation derives from. A prime example is the French civil engineer, Charles Joseph Minard, who is highly regarded for his work using geographical maps. He is most well-known for his depiction of Napoleon's Russian campaign in 1812, which is considered to be the *"greatest statistical graph ever drawn" [2]* by many information designers such as David McCandless, Edward Tufte and Howard Wainer.

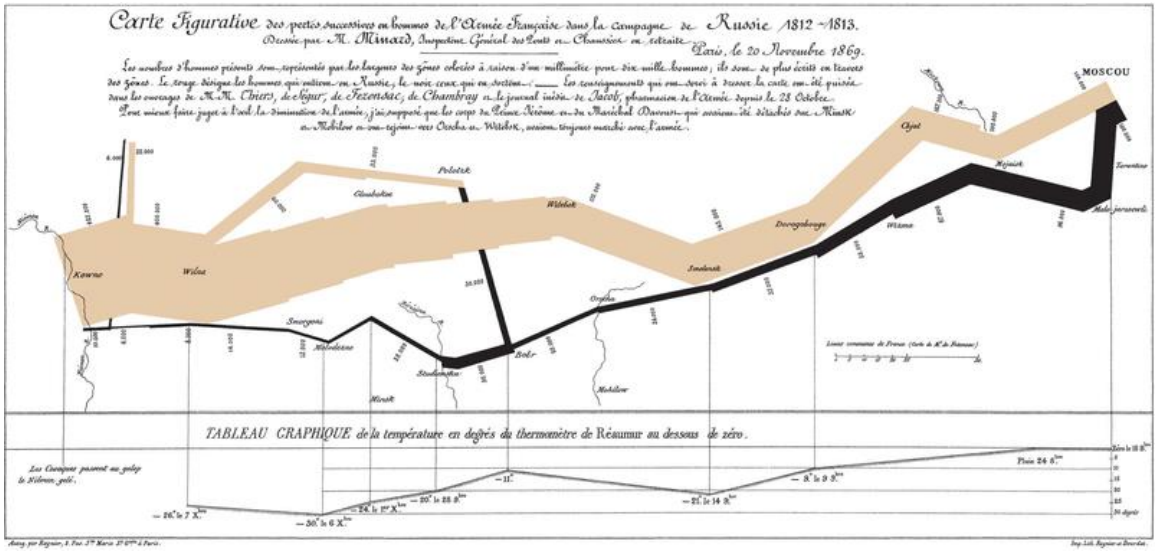


Figure 2: Napoleon's Russian Campaign 1812 [3]

The graph presents a large amount of data on a two dimensional plane. The tan line represents the French armies movement towards Russian while the black line shows those who have retreated. In addition to movement, the graph depicts six other types of data such as latitude, longitude, location relative to dates, temperature, movement direction, troop numbers and travel distance. The reason Minard's graph is praised is that it not only relays the data to the reader in a clear way, it manages to tell the reader a story and achieves its goal of showing how Napoleon's campaign was a failure. Graphically it *"shows the catastrophic loss of life in Napoleon's Grand Army. The diminishing size of the army, initially 422,000 strong (including conscripts from his empire), is shown by the width of a steadily diminishing line, overlaid on the map of Russia, ending with 10,000 returning at the end of the campaign" [3].*

2.2.2 Modern Process to Creating Data Visualisations

David McCandless, a British data journalist and information designer, supports the key components used by Minard in his graphs by stating *"these elements form the backbone of my process" [4]* in creating visualisations. McCandless discusses how four key components, when combined, create functioning data visualisations including: Information (data) integrity, an interesting story, a useful goal and a beautiful visual form. In *figure 3*, it shows the result depending on what way they're combined. For example; a visualisation with only information and story only gives you a script or research article. Interestingly, even combing three of four does not produce a functioning visualisation. For example, combing information, story and visual form without a goal is deemed 'useless'. *"All four elements in [his] graph seem essential. With just two, you get something prototypical and sketchy – which is okay if you're at an early stage. With three, the end result is strangely lacking" [4].*

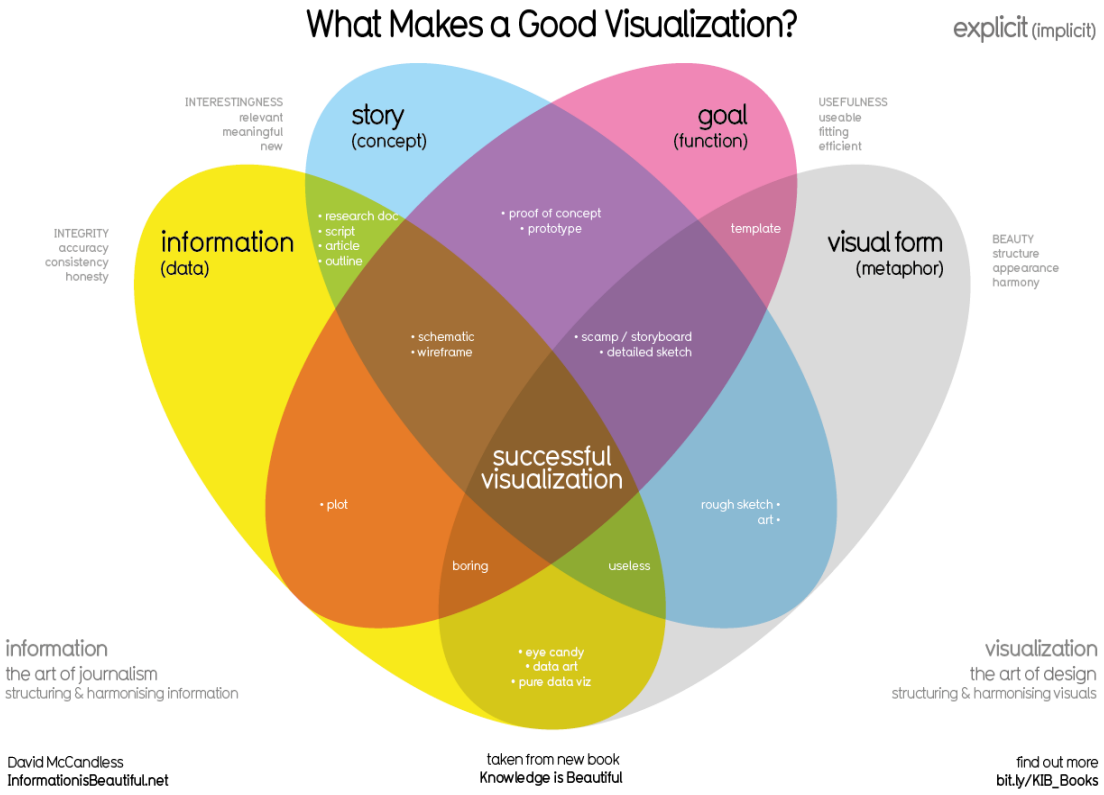


Figure 3: What Makes a Good Visualisation? [4]

Edward Tufte also uses the same general components but goes into more detail about them in the first chapter of *Visual Display of Qualitative Information [2].* Tufte's components include:

- With the data you have, choose the proper format and design. If the incorrect format is chosen for the acquired data, it may become difficult to communicate that data's purpose making it useless.

- The data acquired should achieve a sense of relevant scale and be balanced and proportionate. *Figure 4* below is an example of what not to do, the chart feels cluttered and does not clearly convey information. *"...the only worse design than a pie chart is several of them..." [2][5].*

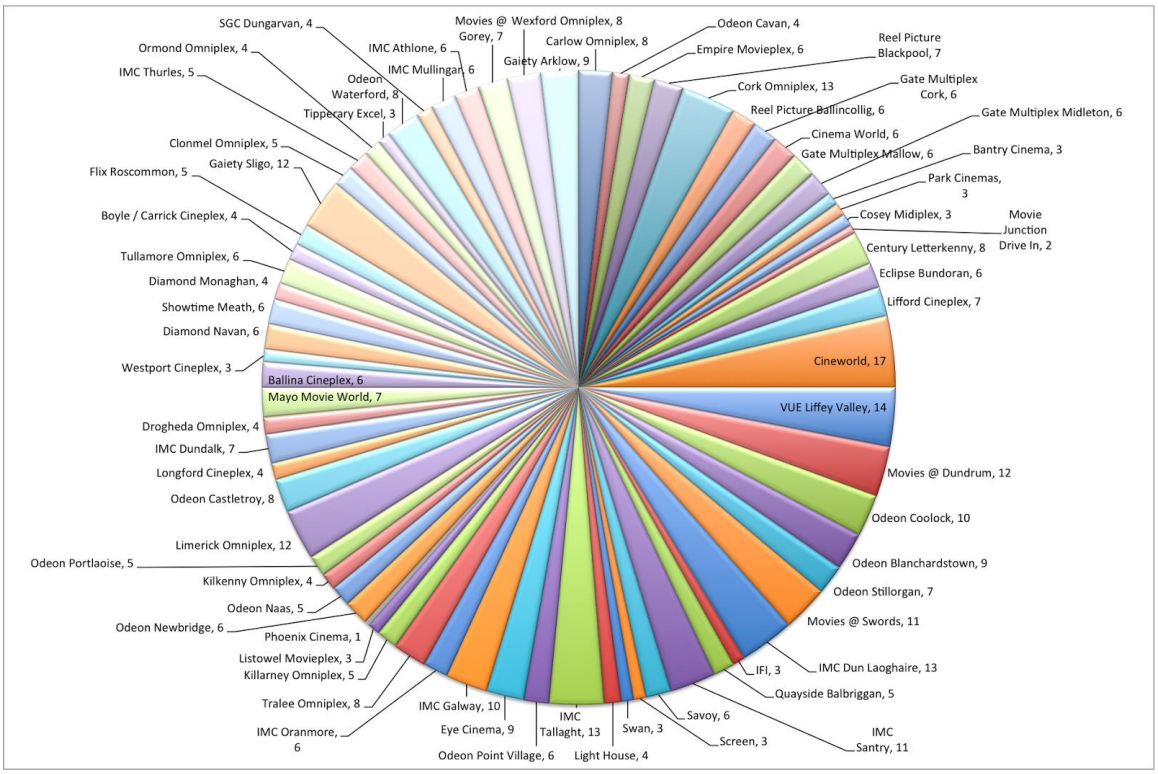


Figure 4: Complicated Pie Chart [5]

- Tell a story about the data providing it a narrative. Readers are more invested and helps them better understand what the data is trying to convey if they can get a feeling of context.

- The use of drawings, words and numbers together presented in a professional manner. A visualisation can work well and produce great results with the combination of these three if they are done correctly. A viewer of a rushed, shoddy visualisation will have gained no knowledge of the information you were trying to display, which is a situation to be avoided at all costs.

In a more recently study in *Creating Business Value from Your Data* by Jean-Paul Isson and Jesse Harriott, they also follow the same general concepts using an acronym *C.O.N.V.I.N.C.E. [6]* to ensure effective visualisations.

Convey Meaning

For a visualisation to convey its meaning, it has to present its data in a fashion that gives it an obvious purpose. This can be achieved by simply using labels and titles that describe what's being displayed.

Objectivity

It is very easy for certain companies and businesses to skew statistics that negatively portray them and so they either exaggerate results or make them entirely obscure by adjusting time period or scale. *Figure 5* shows a real life example of how data can be skewed. Fox News wants to show that if the Bush tax cuts expire it would be a tragedy, but looking closer at the y-axis the starts at 34% instead of 0%, making the 4.6% difference look exaggerated.

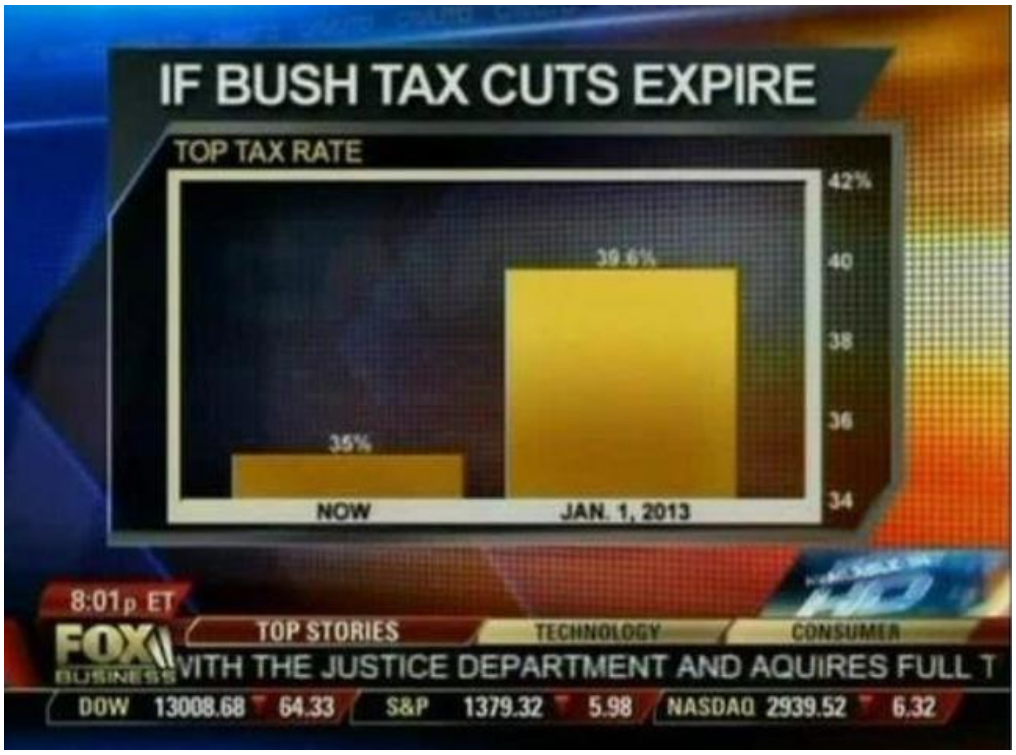


Figure 5: Lie Factor Bar Chart [7]

Necessity

Sometimes adding more details to a visualisation can make it much more complicated and difficult for readers to understand, the data can get lost if the reader has too much to look at. Removing background details and avoiding clutter, or chart junk, can help the reader focus more on the data being presented, or as Harriott and Isson summed up, *"Don't boil the ocean" [6].*

Visual Honesty

Visual Honesty and Objectivity share many of the same properties, which is to keep the visualisation impartial and genuine by following some general guidelines. Continuing on from *figure 5*, start plot points from zero to avoid making the visualisation misleading. In

addition, using correct scales, lines, axis and diligent use of 3D charts to evade visual dishonesty.

Imagine the Audience

When creating data visualisations it is important to consider what type of user will be viewing the data and for what purpose. For example, if the visualisations purpose is to quickly show the data in a short amount of time (i.e. an advertisement), the visualisation needs to be concise and direct with its data as the audience may not have the time or ability to view a more complicated graph. However if the target audience is, for example, a statistician, the visualisation can more sophisticated and use more complicated data visualisation techniques.

Nimble

Nimble and Necessity also share many of the same properties, which is to keep the selected audience attentive by concentrating on a single topic at one time. Creating visualisations that respond to the audiences' interaction can hold users who want to know more.

Context

Context helps users gain a better understanding of the data presented by adding more information within the visualisation.

Encourage Interaction

An effective way to gain an audience's attention is to make data visualisations interactive, encouraging them to explore the data.

2.3 QUALITATIVE & QUANTITATIVE DATA

Qualitative data is data that deals with descriptive qualities. It is data that is not measured but observed. The qualities general include descriptions likes appearance, textures, smells, etc. Quantitative data is data that tends to be measureable. Examples of quantitative data includes height, width, volume, costs, ages, etc. Typically, quantitative data can be analysed statistically which allows for more detailed assessments of the data. In contrast, qualitative data is more exploratory, the findings are not as inclusive as quantitative data and generally is used to provide a broad base of insight. For example, both types of data can be found in almost anything, such as painting. The qualitative data would be that it's a painting of a river, texture shows brush strokes, and is painted shades of blue and white. The quantitative values is that the painting costs 500euro, its surface area is 120" and that's its 10" by 12" *[8].*

Typically, quantitative data is considered much more valuable than qualitative data because with quantitative data, results are much clearer to analyse and is a bigger part of people's lives. For example, when buying a pack of crisps from a supermarket, the food packets are required to display their nutritional values. Everything from calories, fat, sodium levels and vitamins are shown as quantitative data and effect how people perceive what they are buying, compared to if it was described as qualitatively.

2.3.1 Mixed Methods

In data visualisation, many situations would suggest not to combine both qualitative and quantitative data together *"because of the misguided assumptions that more is better, that it is the fashionable thing to do, or, most importantly, that qualitative research is incomplete without quantitative research" [9].* Nonetheless, through mixed-method techniques it is possible to create functional data visualisations using both data types. Mixed-methods involves gathering and researching qualitative and quantitative data. Generally, it is not enough to just gather and analyse the data, they need to be mixed together in a way that they form a more complete story than they do alone. There are a number of ways in which mixed method can be performed *(see figure 6):*

In the first example the data is connected, in examples two and three the data is embedded and in example four the data is merged.

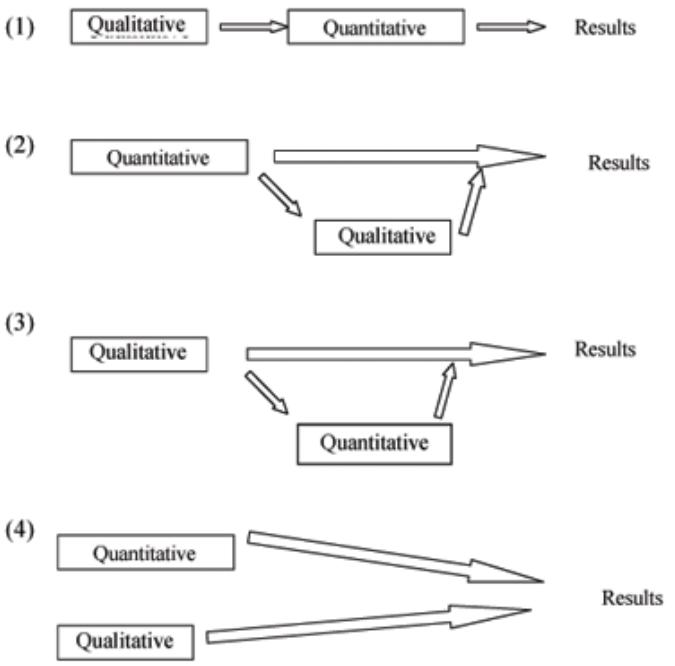


Figure: Possible ways qualitative and quantitative can be integrated.

Figure 6: Mixed Methods [10]

Using these methods produces advantages to better understand data. Mixed methods are useful in comprehending contradictions between the findings of qualitative and the results of quantitative. It also helps in gathering rich and comprehensive data by mirroring the way in which people naturally collect information. For example, in sports articles the qualitative data (images of highlights and descriptions) with qualitative data (scores) assist in creating a complete story than either data set would have alone. It can also be disadvantageous as method-mixing requires more resource finding than it would take to only research one and also increases the complexity as combing the two data sets requires more planning. Integrating mixed method into creating a functional data visualisation has potential to make a powerful enriching visualisations if planned and implemented carefully.

2.4 USABILITY

2.4.1 Gestalt Principles

Usability is an important part of creating functional data visualisations. It is how the user will actually use the data visualisation, which expands on the components of functional data visualisation. The Gestalt Principles, of which there are many, is a psychological study that tries to understand how people perceive the whole on its own, separate from its parts, or as psychologist Kurt Koffka phrased it, *"The whole is other than the sum of the parts" [11].* A prime example of this is how the lights of a movie theatre marquee turn on and off. In reality, it is a series of lights flashing on and off but to someone viewing it appears as though it is one light travelling from one bulb to another.

The most common gestalt principles seen in data visualisation are as follows:

- **Proximity** is one of the main principles that user's perception focuses on first and is one of the simplest to employ. Objects that are closer together are seen to be more related than objects that are far apart.

- **Similarity** is when objects that have similar characteristics, such as shape, size and colour, are perceived to belong to the same group and helps in reinforcing already established relationships.

In the example in *figure 7*, when the balls are scattered they are seen to be related based on their colour. However, when placed next to one another, they are perceived related based on their location or proximity of each other.

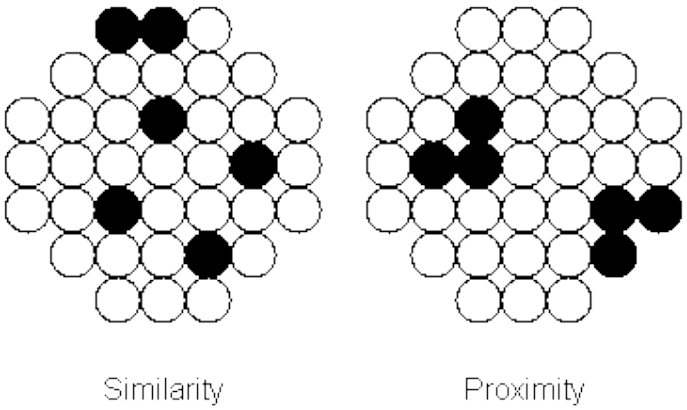


Figure 7: Similarity & Proximity [12]

- **Continuity or Common Fate** is when objects that are moving in unison or in the same direction are seen as related. Alternatively, objects moving in different directions are seen to be less related. This is useful in establishing or negating relationships between objects. In *figure 8*, all the pink bars are the same colour and shape but are seen to be different groups of bars based on the direction they are pointing.

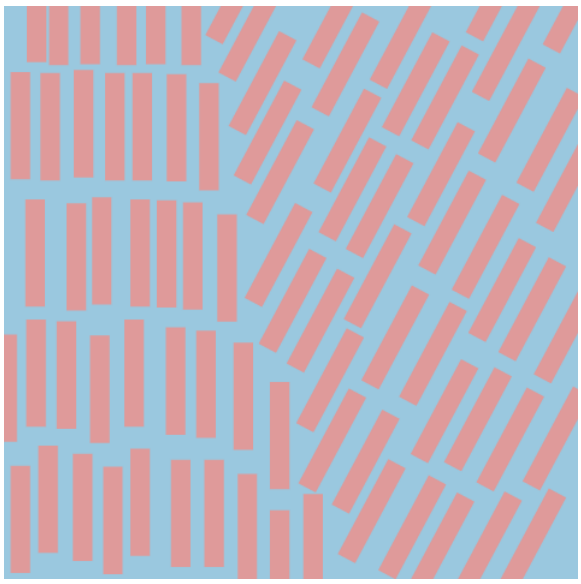


Figure 8: Continuity or Common Fate [13]

- **Closure** is about drawing conclusions and occurs when an object is incomplete or not completely enclosed. If key points of the shape are shown, people will perceive the whole shape themselves. For example, in *figure 9*, the image consists of a

number of spikes of a variety of sizes but the way in which they are displayed gives the impression that it's an image of a spiky sphere.

Figure 9: Closure Example [14]

- **Common Region** is when objects that are grouped together in the same enclosed region appear to be related, even is colour, shape, and size is different. This can be carried out by simply providing relating objects a border from unrelated objects. In the example in *figure 10*, the sign is displaying the lovely message "Stop War. Peace Now". Although when the common region of the sign is changed from being horizontal to vertical, the signs meaning is the complete opposite message as before, reading as "Stop Peace. War Now".

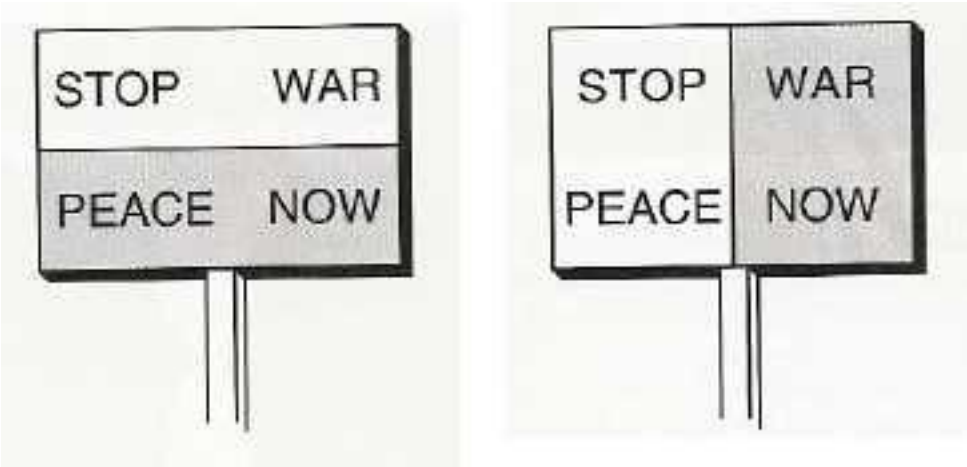


Figure 10: Common Region [15]

- **Past Experience** is possibly the weakest gestalt principle as every individual has had unique past experiences and typically should not be relied heavily upon as it could be perceived in any number of ways depending on the user, but can be effective if used correctly. The most common use of past experience being used in practice is the computer desktop. To make the transition from physical to digital

easy to understand, components of the software were named after the real life objects they would be simulating. For example, the desktop has folders, and within those folders you keep your files

2.5 TABLEAU

Tableau is a business analytic software platform that allows users to easily create data visualisations. The company was founded by Chris Stolte, Christian Chabot and Pat Hanrahan in January 2003 in Stanford University. Between 1999 and 2002, Pat Hanrahan and Chris Stolte were researching the use of table-based displays to view multidimensional relational databases with the aim to commercialise the research. To accomplish this, they created a database visualisation language known as VizQL (Visual Query Language), "*a formal language for describing tables, charts, graphs, maps, time series and tables of visualizations. All these visual representations are unified into one framework, lowering the overhead of switching from one visual representation to another" [16].* In October 2003, Tableau was sent out of Stanford and the company moved to Seattle, Washington, where it still remains.

In 2016, for the fourth year in a row, Tableau is the leader in the Magic Quadrant for Business Intelligence and Analytics Platforms *[17].* The Magic Quadrant is a yearly report that details all major analytic software companies and details their strengths and cautions and maps out the results on a quadrant chart.

Figure 11: Gartner Magic Quadrant [17]

Tableaus strengths are that it exceeds best at its "land-and-expand" sales model, which is when an initial sale is made (landing) and over time the product becomes more widely used within the company, resulting in more sales (expanding). Tableau focusing on its customers' needs has helped the company grow exponentially which is evident from Tableau's user conferences, "*attendance at Tableau's user conference topped 10,000 attendees in 2015, nearly double the 2014 attendance and an increase of more than 50 times the 187 attendees at its inaugural user conference in 2008" [17].* Tableaus ability to allow it's users to connect to a wide array of data sources, which is constantly expanding, provides rapid access to visualisations and interactions.

Tableau's cautions are that it's cost. In a survey, 44% of people *"cited the cost of software as a barrier" [17],* which is much higher than many of Tableau's competitors. Tableau faces a problem many rapidly growing companies, its rapid growth means that it struggles to scale to meet support for its clients and demand for more complex deployments. "*The reference survey also suggests that buyers of Tableau have encountered some software limitations as they attempt to scale their deployments (to meet the demands of more users trying to solve more complex problems)" [17].*

2.6 SIMILAR PRODUCT ANALYSIS

There were not many websites that had similar aims or provided similar functionality to what Census Ireland aimed to achieve, Airo.maynoothuniversity.ie *[18]* was the most similar example found. Exploring through Tableau's public servers *[19]* resulted in the discovery of multiple visualisations that achieve the projects interaction and dynamic goals, many of which were referred to when deciding on a suitable layout for the project and confirming or rejecting possible functionality decisions made during the projects production.

2.6.1 AIRO Content

The AIRO (All-Island Research Observatory) site has a 'Featured Mapping Source' section on their home page. The map shows *"hundreds of socio-economic variables at the Small Area (SA) level" [18].* Unfortunately at the time of the writing of this report, the visualisation was in beta so much of the functionality was not available. However, the primary functionality was there which included an interactive map of Ireland split into segments which had a tooltip with data appear when a segment was clicked on. The data shown could also be filtered based on census data such as population and religion *(See figure 12).*

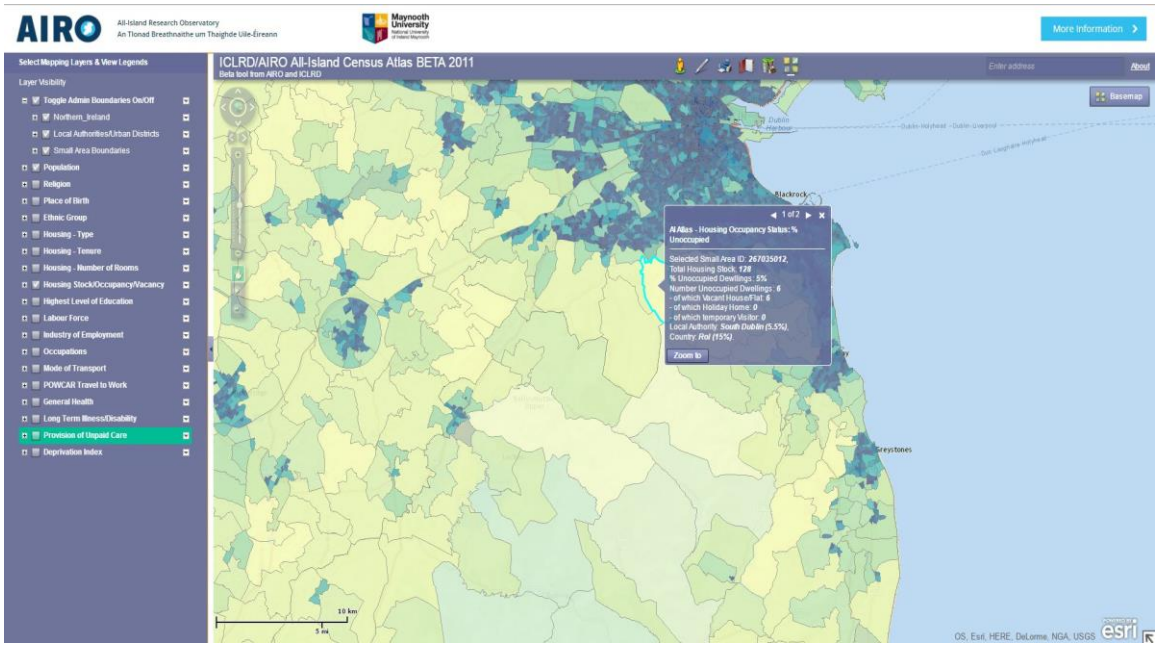


Figure 12: All-Island Atlas (BETA) [20]

2.6.2 Tableau Public Server Content

On the Tableau Public home page, navigating to 'Gallery' will display a list of visualisations called 'Viz of the Day', *[21]* that shows the best visualisations made by Tableau users. Although there are no maps specifically about Irish census data on a county level, there are many about other countries and world maps that show similar functionality goals of the project. For example, *figure 13* shows a map of the world displaying data about the prevalence of diabetes. The user can filter the data by region which results in all the graphs dynamically changing to only display the dataset selected. The user can also mouse over countries to display a tooltip showing addition data specific to that country. *[22]*

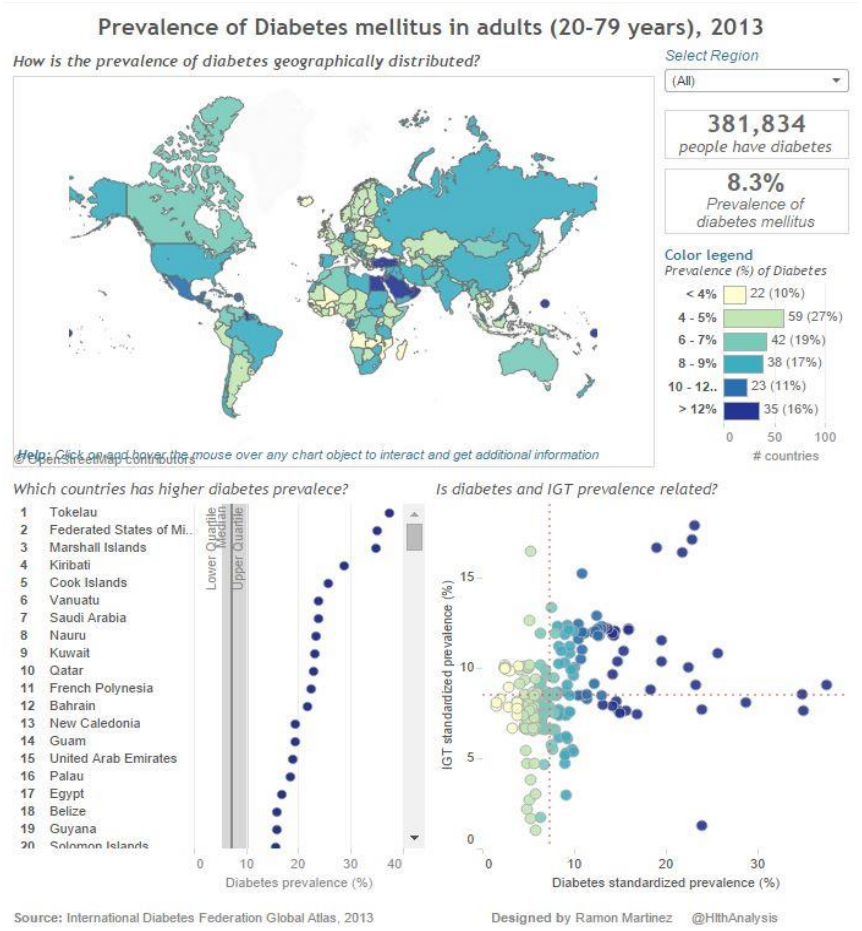


Figure 13: Prevalence of Diabetes Map [22]

2.7 USER REQUIREMENTS

The user requirements are what the user expects the system to do. Users should be able to request and receive data as well as be able to view the visualised data, and in some cases be able to alter which data appears in place of others. An example would be if a map was showing population statistics, the user could check a tick box to change the data to birth rate statistics.

Figure 14 shows a use case diagram of the user requirements. As shown, the user can request and receive data. When the user is viewing data the use case extends to allow the user to change data being displayed.

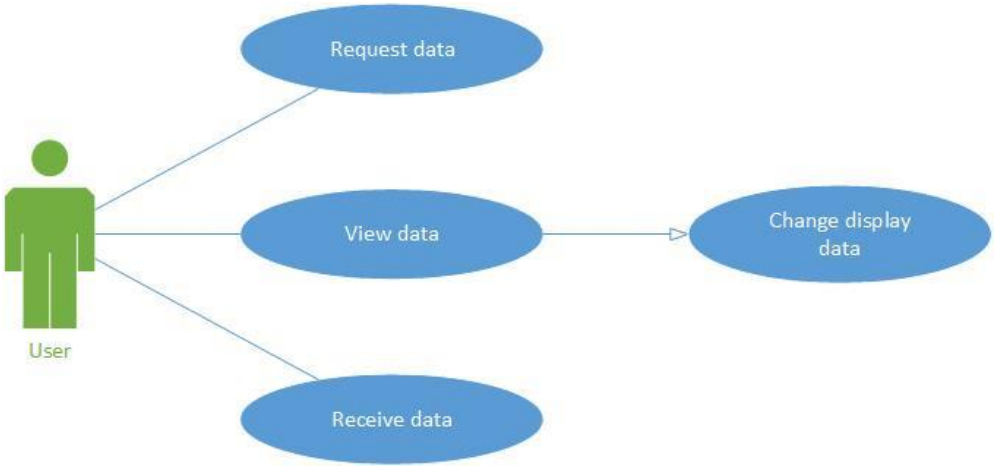


Figure 14: Use Case Diagram

To help better understand the user requirements for Census Ireland, personas were created. Personas are a way to summarize the way in which an identified type of user would use the system. They depict a specific person but not a real individual. The persona is a representation of a large portion of people in the real world that help focus on a memorable cast of people, rather than hundreds of individuals.

In the case of Census Ireland, personas were made to represent the two main users that would likely be using the web application. The first *(See figure 15)* is a user with a vast amount of experience in data visualisation.

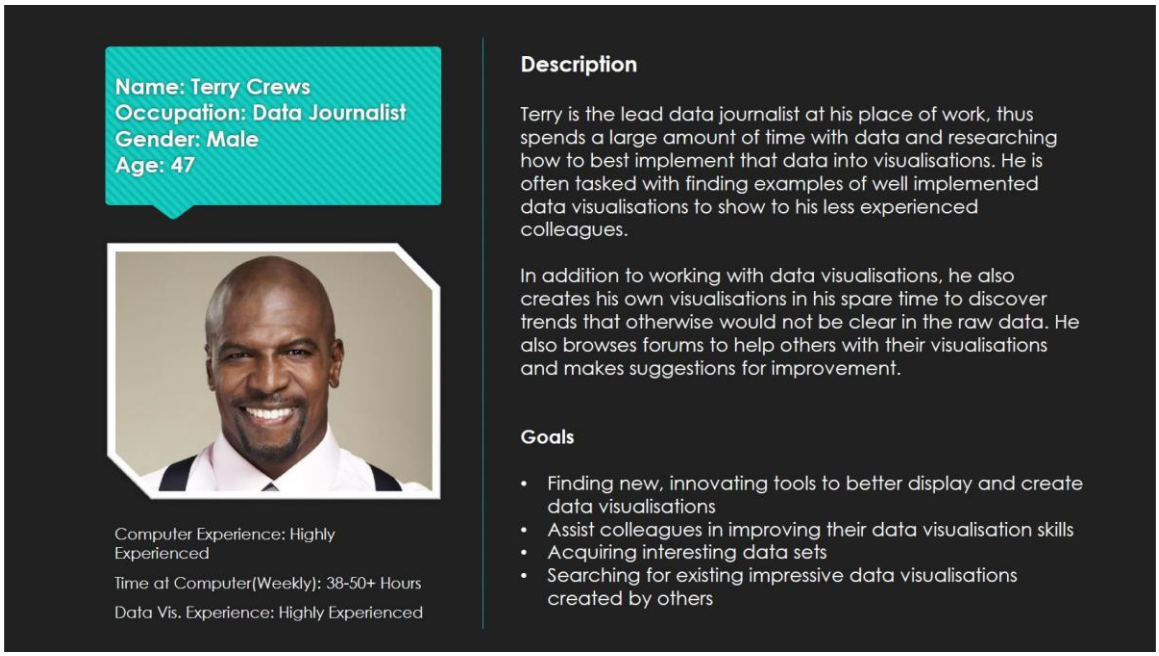


Figure 15: Persona, Experienced User

The persona describes a user named Terry who works as a data journalist. He is described as spending much of his professional and personal time working with data visualisations. Census Ireland should be able to cater to Terry's goals and meet his needs as an experienced user.

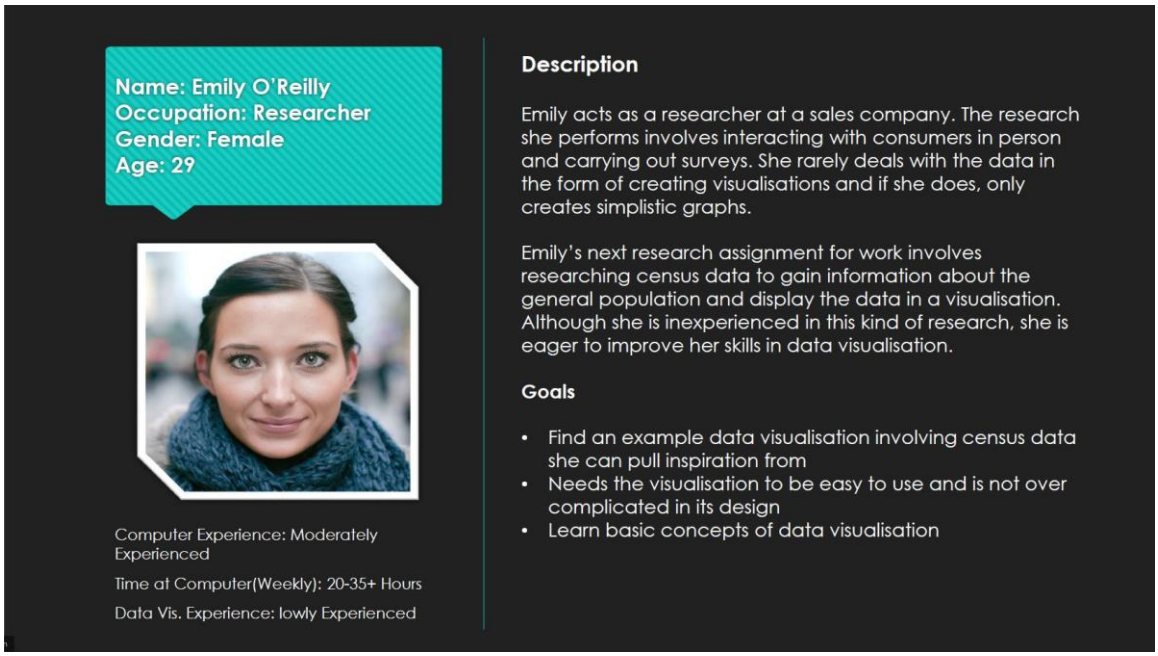


Figure 16: Personas, Inexperienced User

The second persona *(see figure 16)* describes a user named Emily who works as a researcher. Unlikely Terry, she spends very little of her time working with visualisations and is unfamiliar with how they are implemented. Census Ireland should also be able to cater to Emily's goals and needs as an inexperienced user.

2.8 TASK ANALYSIS

The task analysis grid outlines the goals users wish to achieve when using Census Ireland. The grid then describes a scenario in which he user would need to achieve the goal, the type of users the specific goal applies to, any influences that may affect how the user reaches the goal or might prevent them from achieving it, and then delves into the tasks and sub tasks, describing the steps that would need to be taken to reach the goal.

Task Analysis Grid			
Goal:	**VIEW CENSUS DATA**	**CHANGE CENSUS DATASET**	**FILTER CENSUS DATA**
Scenario:	The user wants to view a visualisation of Irish census data	The user wants to change the dataset from the set they are currently viewing to a new set	The users wants to filter out data from the current dataset they are viewing
Applies to:	All users	All users	All users
Influencers:	Tableau's public servers. An internet connection is required	Tableau's public servers. An internet connection is required	Tableau's public servers. An internet connection is required
Tasks & Sub Tasks:	**(1)** The user makes sure they have an active internet connection. **(2)** The user needs to load the web application. **(3)** The user waits for the web application to load the data visualisation off of Tablea's public server. **(4)** The user can now interact with the data visualisation	**(1)** The user makes sure they have an active internet connection. **(2)** The user needs to click into a new tab that appears at the top of the visualisation in order to switch to a new dataset. **(3)** The user waits for the web application to load the new data visualisation off of Tableau's public server.	**(1)** The user makes sure they have an active internet connection. **(2)** The user can filter variables that are present in the current dataset by using dropdown menu/slider/tick box filters as well as clicking on specific elements on the data visualisation

Figure 17: Task Analysis Grid

Working through the first goal to view census data shown in *figure 17*, it shows the scenario in which the user would need to achieve the goal is that the user wants to view a visualisation of Irish census data. It applies to all users who would be accessing Census Ireland. There are two influencers that affect the end goal and there are that the Tableau public servers need to be online and that the user needs an active internet connection. The tasks and sub tasks to achieve the goal to view census data are as follows:

1. The user makes sure they have an active internet connection
2. The user needs to load the web application
3. The user waits for the web application to load the data visualisation off of Tableau's public servers
4. The user can now interact and view the data visualisation

2.9 FUNCTIONAL REQUIREMENTS

The functional requirements are the systems functions and components, basically what the system should do. The functional requirements of the project should display data in clear, visible and easy to understand fashion for users, allow the user to filter the data being viewed by either using a dropdown menu or using the interactive map of Ireland, display tooltips when the user hovers over elements in all graphs and maps, and let the user easily navigated between the many different visualisations on the web application.

2.10 NON-FUNCTIONAL REQUIREMENTS

Unlike functional requirements, non-functional requirements describe how the system should behave and work. Performance is important, as the project grows and has more data added, load time may increase. The system should be built to be scalable, as the main goal is to get detailed statistics of each county across multiple datasets. Usability should also be a main focus. The data acquired could be worthless if it is not easily understood by the user.

2.11 TECHNICAL REQUIREMENTS

The data for the data visualisations will be formatted on Excel worksheets before being read into Tableau. Tableau then extracts the data from the worksheets before being uploaded to Tableau's public servers.

The user interface will be designed to be easy to navigate, easy to read and easy to understand what the visualisations are trying to convey. Although many of the users are likely to be experienced with data visualisations, it is important to cater to those users who may be less familiar with the visualisations. Except for the navigation section and logo, the rest of the web application will consist of varying types of graphs which the user can interact with and filter.

The way in which the system will process these interaction will be through a client – server – client system. For example, when the user requests a visualisation of a dataset, the request is sent to the Tableau public server which will fetch the visualisation from Tableau Public's database and is sent back to the user in the web application.

2.12 IMPLEMENTATION PLATFORMS

The technical requirements outlined above will be implemented primarily using Tableau 9.1 to create the data visualisations and draw the polygonal map of Ireland, while Microsoft Excel will be used to format and temporarily store the census and referendum data. The project developer has experience in using Excel but has no prior experience in using Tableau, or creating data visualisations. Research will need to be carried out to properly implement the projects main functionality, along with learning how to append data into Tableau's visualisations from an Excel worksheet.

2.13 DATA SOURCES

The data for the project will be acquired primarily from the CSO website (Central Statistics Office) *[23]* and the Referendum Ireland *[24]* website. The data acquired from

the CSO is formatted as a CSV (Comma Separated Value) file while the data from Referendum Ireland had to be manually added into an Excel Worksheet. Additional data sources, such as AIRO (All-Island Research Observatory) and data.gov.ie *[25]*, were located as a precaution if the CSO could not provide suitable datasets.

2.14 POTENTIAL ISSUES

2.14.1 Server Side Lag

The data visualisations are being stored on Tableau Public servers. There is the possibility that if the visualisations contain a large amount of data, the time to pull them from the database into the web application could be an issue.

2.14.2 Time

This is going to play a major part in this project. It is understood the full task at hand and the length of work that needs to be completed in order to complete the project. The longest part of this project to implement will be the creating and uploading of the data visualisations.

2.14.3 Testing

Testing will be conducted on each visualisation in Tableau before being uploaded to make sure each one is fully functional and ready to be implemented into the web application. After each visualisation gets created, it will then be tested to ensure it is functional. However, even after it is uploaded, as long as the original Tableau Workbook of the visualisation is available it can be edited and overwriting on the server at any time.

After the entire application is complete, it will be made available to testers to test and report any bugs. Ideally, the developer would want to solve most of the bugs before beta testing and to allow user testing as a final phase of testing to fix small minor issues that the ordinary user may have such as bad navigation or unclear processes that the user may encounter, or difficult to understand visualisations.

2.14.4 Front End

To help in improving the best user experience possible, a HTML and CSS minifier will be used so the files do not cause too much of a slowdown on the web application. When tested on a sample code of CSS, the minifier cut the file size down by half. A minifier takes the code that has been written and lays it out in a way that reduces the file size, primarily

getting rid of white space by placing all the code in a long single line and getting rid of any unnecessary spaces.

2.14.5 Backup Platforms

In the event that Tableau is no longer a viable option as the software platform. Quantum Geographic Information System (QGIS) *[26]* can be used along with OpenStreetMaps *[27]* to ensure the project continues, though QGIS has been found to have a much higher learning curve and focus less on visualisations than Tableau.

QGIS is a free and open source geographic information system that allows users to create, edit, visualise, analyse and publish geospatial information on Windows, Mac, Linux and BSD *[26].* OpenStreetMap is built by a community of mappers that contribute and maintain data about roads, trails, cafes, railway stations and much more, all over the world. *[27].* It is also open data, meaning it is free to use for any purpose as long as OpenStreetMap and its contributors are credited.

2.15 SUMMARY

Creating data visualisations require much consideration before being deemed functional. Back as far as 1812, Minard shows with his work that qualitative data and quantitative data can be combined using mixed-methods that follow key components to create functional data visualisations. Minard's work inspired many great analytical statisticians and designers to take his work and allowed them to create their own thoughts on key components of data visualisation. Tufte, McCandless and Isson and Harriott all have varying ideas of these key components but all of them seem to follow the general concept that functional data visualisations have:

- Solid data and information to be presented in a professional manner.
- A story or narrative that helps the user understand the presented data by conveying context and meaning.
- A goal or objective that is visually honest and does not skew the data, but is balanced and non-biased.
- The data presented is shown in a visually pleasing way to gain and keep the selected audience's attention.

The data type selected, be it qualitative, quantitative or both are taken into consideration in creating a functional data visualisation.

Through researching the attributes of functional data visualisations, it makes the process of identifying and analysing the requirements needed for accomplishing the projects goals. Determining the types of users and the tasks they will need to carry out assist in defining how the project will need to be build and laid out in order to allow them to achieve those tasks with ease. Analysing the functional and non-functional requirements help clarify the projects aims while the technical requirements identify what software will be needed to complete those aims. Finally, recognising potential issues that may arise during the projects implementation will allow planning to prevent these issues from derailing the project.

3 Design

3.1 INTRODUCTION

The design chapter will detail the projects system architecture. It will include a system model and sequence diagram to help better understand how different parts of the system will interact with one another. The user interface design will be discussed and examples of similar UI's will be shown which were used as points of reference of how to layout Tableau dashboards in an intuitive way. The projects own prototype will be detailed before moving on to the projects final UI design, showing the progress the project made over the course of the design phase and how it changed over time.

3.2 DESCRIPTION OF PROPOSED SYSTEM

3.2.1 System Model

The system will follow basic architecture. A user will be able to access the system via laptop, desktop and ideally on mobile devices. The user's device will call the dashboard/system to access the data and its display. For the user to change the data type, they will have to call the system and the system will return the data change.

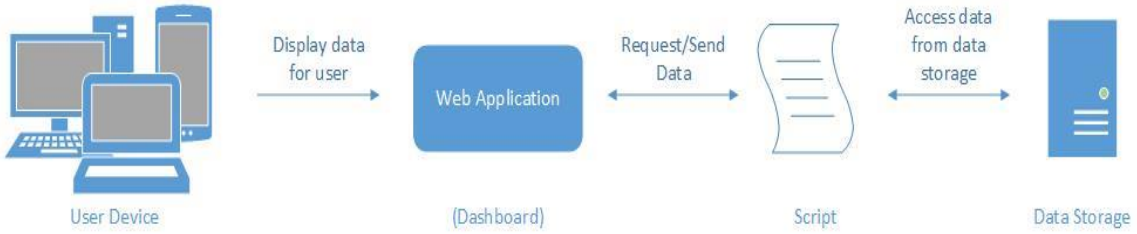


Figure 18: System Model

Figure 18 shown above shows a predicted system model the project will implement. The web application where the data is displayed can be accessed by a variety of user devices such as desktops, laptops, tablets and mobile devices. Whether or not the application will work responsively on the smaller devices (tablets and mobile devices) will depend on how much time is still available when the main functions are working. The web application will then call script that will be able to pull data from data storage and have it become available for display on the web application.

3.2.2 Server Side

The web application will be pulling the stored data visualisations and data off of Tableau's public servers for the users to view and interact with. The data will be acquired from the Central Statistics Office (CSO) and Referendum Ireland. The data off CSO is available to

download as a CSV but requires reformatting to be compatible with Tableau. Referendum Ireland has no download so the data needs to be entered into Excel manually and recalculated to fit constituency data to work as county data.

3.2.3 Client Side

The web application itself will use a small amount of HTML5 and CSS for basic web page structuring. Tableau will be implemented into the HTML by first uploading the visualisations to Tableau public and then embedding code acquired from Tableau Public that will allow the web application to call each visualisation for display.

3.3 SYSTEM ARCHITECTURE DESIGN

The functionality involves displaying census data on a dynamic map and graphs that users can interact with. Techniques that will be used to design the user interface have been researched which will assist in making the application more user friendly and visually appealing. The final design has yet to be fully realized and will continue to evolve over time as more skills in UI design are acquired. The hardware for the project is standard and does not require any additional resources.

Many datasets to be used in the final application have been downloaded and correctly formatted. There are additional datasets on the CSO and other data gathering sites that have been identified but still require reformatting for use in Tableau. Research has been done with Tableau on how to best implement the data, along with multiple tutorials to learn how to use all the features needed to fully implement all the applications functionality.

3.4 TABLEAU

Tableau is a platform that allows users to see and understand their data and use them in visualisations. It is an intuitive platform and once one is familiar with the layout they can easily start creating dashboards and visual analytics. However, there is a slight learning curve, especially when working with polygon maps and multiple sets of data. The primary reason Tableau is being used to produce the application is its ability to display polygon maps. A polygon map is created in tableau using a number of variables:

- **Polygon ID and subPolygon ID**: Draws out the separate polygons.
- **Point ID**: Tells Tableau in what order to draw the outline of the polygons.

- **Longitude and latitude:** Marks the Id locations on a geographic map, which is supplied by OpenStreetMaps.

Once these variables have been placed on the map, data in the form of CSV's can be linked to them, typically using the county name as the link.

3.4.1 Dashboards

Once a number of graphs have been created on Sheets, they can be brought together on one main page called a Dashboard. Dashboards are visual tools that help people view and interpret data. Noting relevant and interesting data will be required as a Dashboard filled with unimportant data is boring for the user and using irrelevant data does not create a story. As the main purpose is displaying the data on polygon maps, the creation of charts is to aid in the analysis of the data. The user needs to be able to determine if there are any trends, are certain values larger than others and how can they compare those values. It needs to be easy for the user to compare the data and have the data be memorable.

3.4.2 Data

The data acquired from the CSO is formatted as a CSV (Comma Separated Value) file. Unfortunately when downloaded, the CSV's are laid out in a way that Tableau cannot read so each data set used in the application has to be layout properly be readable in Tableau. *Figure 19* shows an example of how the data is initially downloaded. In the current layout, Tableau would not detect that the years 2006 and 2011 are meant to be different columns. In addition, each numeric value needs to have the name of the county, the sex and the year in the same row for Tableau to read it.

				2006	2011
State					
	Both sexes				
		Leinster			
			Population (Number)	2295123	2504814
			Actual Change Since Previous Census (Number)	189544	209691
			Population Change Since Previous Census (%)	9	9.1
		Carlow			
			Population (Number)	50349	54612
			Actual Change Since Previous Census (Number)	4335	4263
			Population Change Since Previous Census (%)	9.4	8.5
		Dublin			
			Population (Number)	1187176	1273069
			Actual Change Since Previous Census (Number)	64355	85893
			Population Change Since Previous Census (%)	5.7	7.2

Figure 19: Incorrect Excel Layout

Figure 20 shows how the layout should be formatted for Tableau to read it. The numeric values for each county are clearly displayed in its associated row, repeating multiple times

based on the year and the sex columns. With the data layout formatted this way, Tableau can easily sort the data to be appended onto the polygon maps and charts.

COUNTYNAME	Year	Sex	Population	Actual Change	Population Change Since Previous Census (%)
Carlow County	2006	Both sexes	50349	4335	9.4
Kilkenny County	2006	Both sexes	87558	7219	9
Cavan County	2006	Both sexes	64003	7457	13.2
Monaghan County	2006	Both sexes	55997	3404	6.5
Clare County	2006	Both sexes	110950	7673	7.4
Cork County	2006	Both sexes	361877	37110	11.4

Figure 20: Correct Excel Layout

3.4.3 Backup

Once certain milestones of the application are finished, it gets uploaded to Tableau's public server. This means if the Tableau workbook get corrupt or lost at the source files, the uploaded application is still fit for use.

3.5 SEQUENCE DIAGRAMS

Figure 21 shows how the user will receive data once it's requested. The user requests the data once the web application is opened. Embedded HTML script initializes and makes a requests data from the data storage (Tableau public server). The data storage then acquires the correct dataset off the server and is sent back to be displayed in the web application for the user to view.

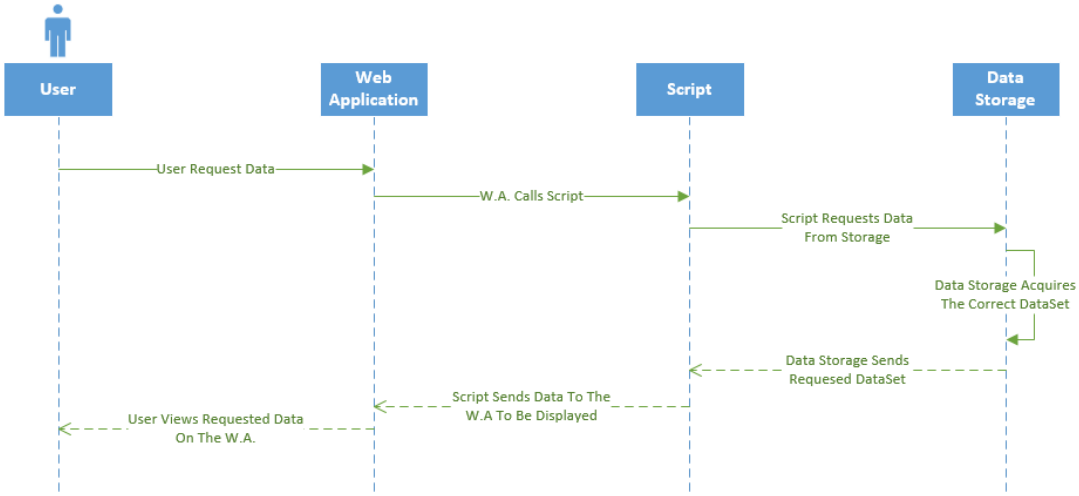


Figure 21: Sequence Diagram, Display Data

Figure 22 shows how the user can change the data being displayed in the web application. The steps taken in *figure 22* follow the same architecture as *figure 21* in the way the user interacts with the web application and how it runs code to pull data from the data storage to be displayed back in the web application.

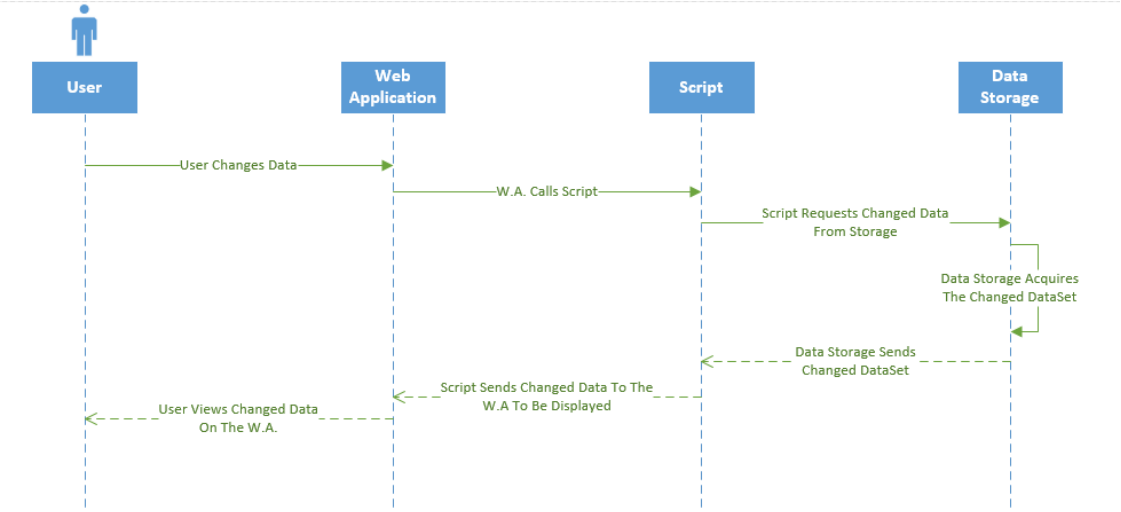


Figure 22: Sequence Diagram, Change Data

3.6 USER INTERFACE DESIGN

The user interface (UI) focuses on what the user needs to do and ensures that the interface has the required elements to carry out these tasks. The UI for the application will need to incorporate good interaction, information and visual design. When designing the user interface there are core fundamentals that should be followed:

- **Page Layout:** The web applications elements should be organized in a way which allows users to tell which elements are more important than others. The screen layout should be consistent to allow tasks to be performed quickly and efficiently.

- **Simple Interface:** Keeping the interface simple is important and can be accomplished by avoiding unnecessary elements that could clutter the UI and using language that the user can fully understand, instead of using terminology a regular user would not be able to comprehend.

- **Visual Elements:** The use of colour and typography is a useful way to direct or redirect a user's attention to particular elements. Colour is simple to implement and is extremely effective. Warmer colours like red and yellow attract more attention from users than colder colours such as green and blue. Contrast is used to catch the user's eye as making elements lighter or darker against the background to make more vital information stand out from less important parts.

Following these three fundamentals will assist in deciding the final design of the user interface.

3.6.1 Example User Interfaces

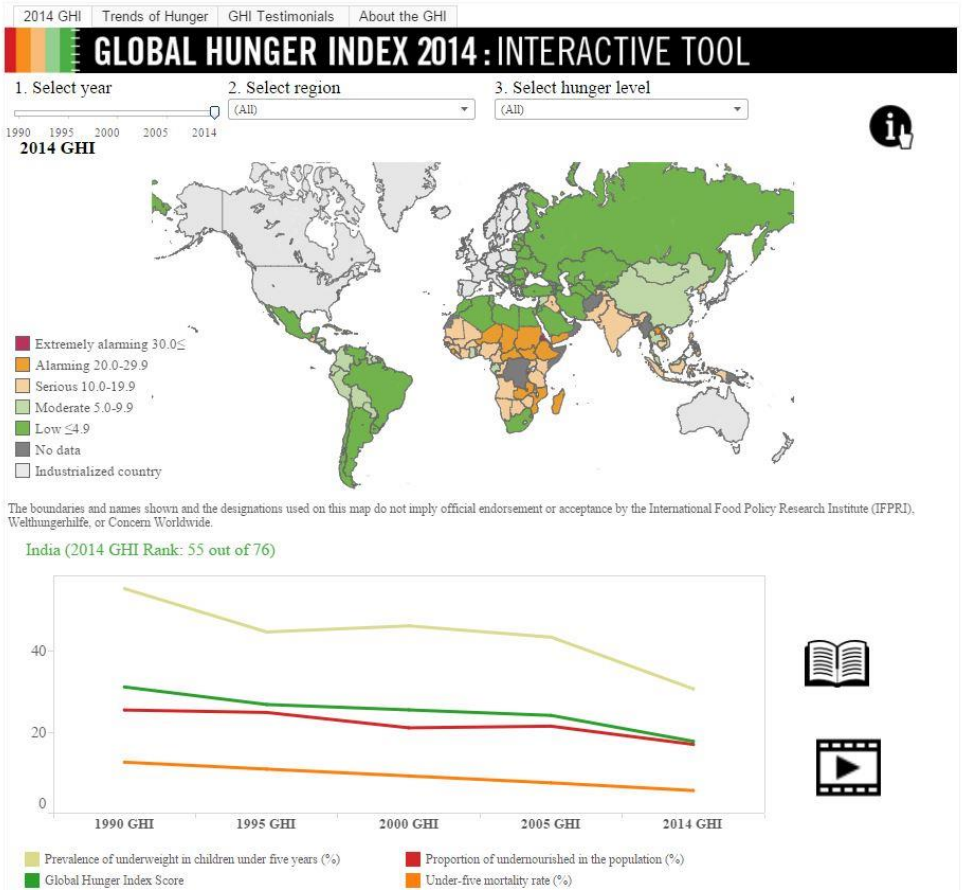


Figure 23: Global Hunger Index [28]

Figure 23 shows an example of the type of detail Tableau is capable of when dealing with geographic data. On the top of the page are tabs that open additional graphs and information regarding global hunger, which is an effective way to separate up the data so one page is not cluttered with graphs. In the visualisations main body, a map of the world is shown with colours depicting the varying levels of hunger along with a clearly labelled legend so the user knows what each of the colours mean.

However, even if the legend was not shown, the user would likely associate green with being good and red being the worst. This is a common gestalt principle used in data visualisations. Above the map are three filters that allow the user to change the data based on year, region and hunger level. The filtering function is taken even further as when the user clicks on a country, the line chart below the map updates and shows data relating to the selected country.

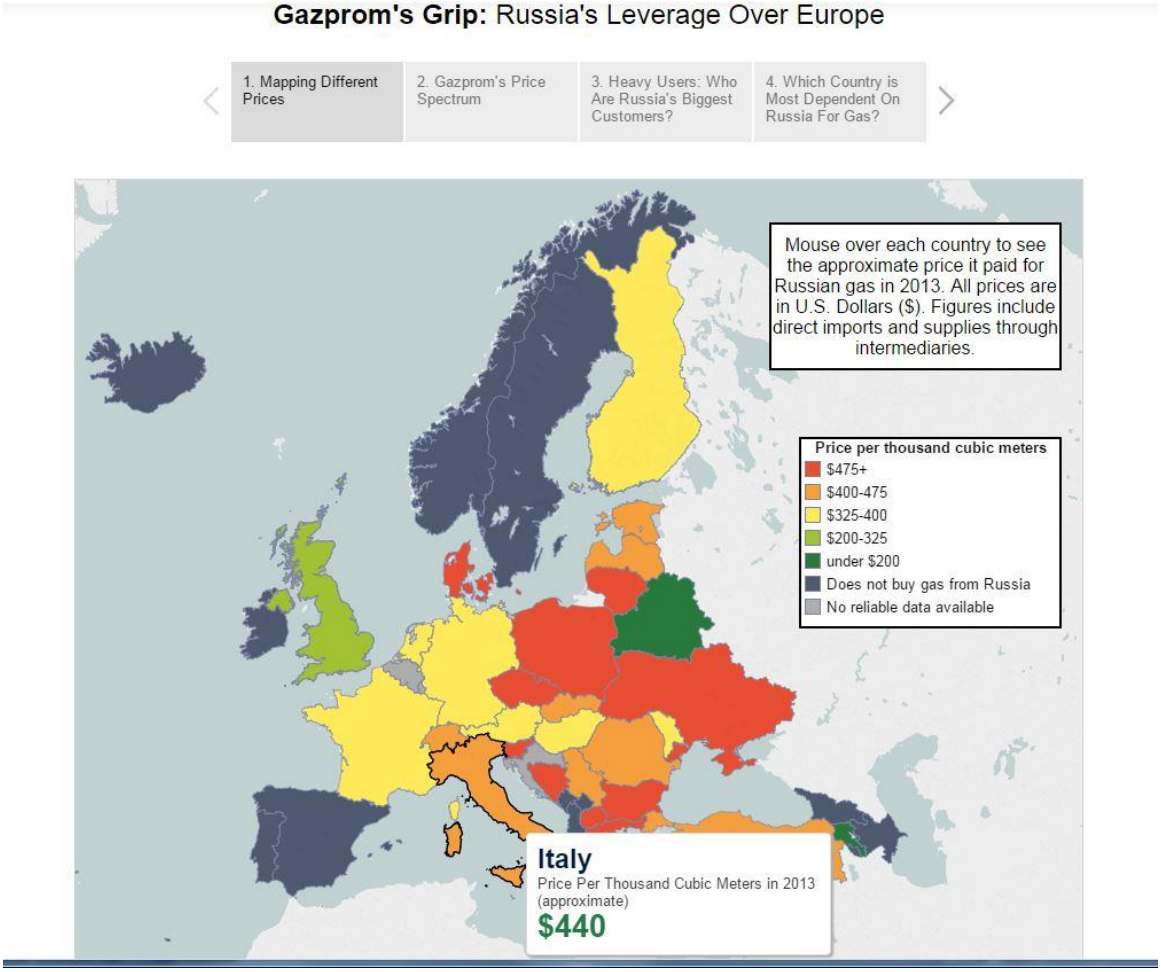


Figure 24: Gazprom's Grip [29]

Figure 24 is another example of data visualisation being implemented with Tableau using geographic data. It displays the amount each country pays for Russian gas in 2013 in US dollars. Similar to *figure 23*, the map is coloured going from green being the least significant pricing to red being the most significant pricing. The visualisation also uses the hover function that allows data to appear on a tooltip relating that countries statistics. Rather than having multiple graphs on one page, the author decided to keep the graphs separated on different pages.

Figure 25 shows a graph based off Gazprom's Grip. The graph is coloured using blue and orange which are complimentary colours to each other, this helps make clear distinctions between the data being presented.

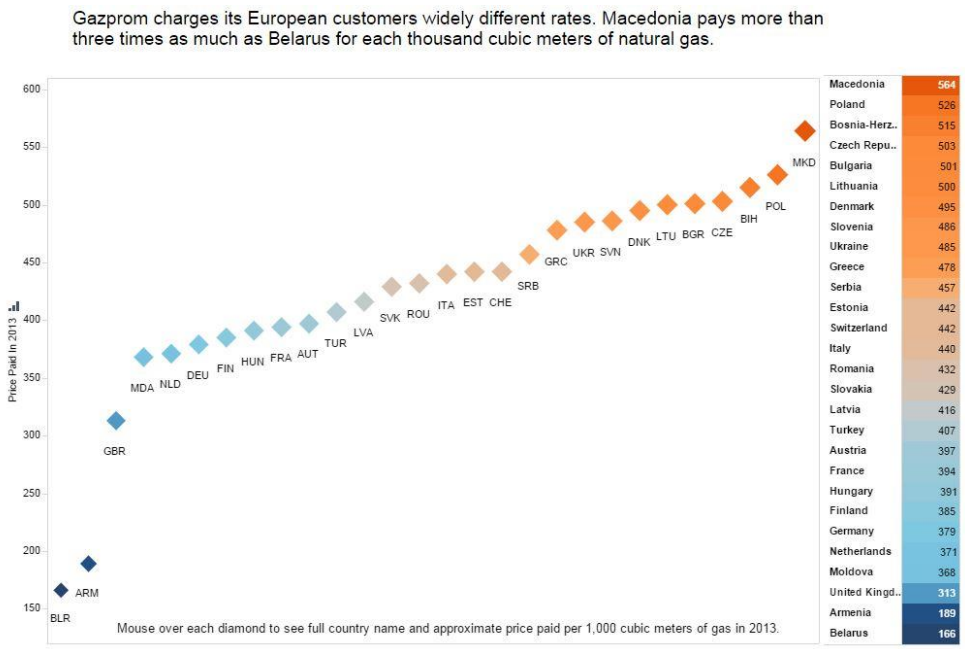


Figure 25: Gazprom's Grip, Graph [29]

Identifying the colour usage, filtering capabilities and layout of the two examples shown in *figure 23, 24* and *25* will help greatly in creating the final design for the application. Keeping the graphs simple to understand with the use of legends is important, as well as having the graphs clearly labelled. Using complimentary colours can make the differences in the data stand out more and using colour combinations people already understand (green = good, red = bad) makes the visualisation easier to understand.

3.6.2 Prototype Design

The prototype shown contains data from the marriage equality referendum 2015 displayed in Tableau using a variety of visualisations. *Figure 26* shows how the data needs to be formatted in excel so it can be read correctly by Tableau. The referendum is counted by constituency, not by county, so a number of counties will display the same data in Tableau. For example, Carlow and Kilkenny are in the same constituency and thus, share the same data.

COUNTYNAME	Constituency	Electorate	Total Poll	% Turnout	Invalid Ballot Papers	Valid Poll	Votes in Favour	Votes Against	% in Favour	% Against
Carlow County	Carlow - Kilkenny	104,735	68,531	65.43	668	67,863	38,166	29,697	56.24	43.76
Kilkenny County	Carlow - Kilkenny	104,735	68,531	65.43	668	67,863	38,166	29,697	56.24	43.76
Cavan County	Cavan - Monaghan	99,265	56,774	57.19	517	56,257	28,494	27,763	50.65	49.35
Monaghan County	Cavan - Monaghan	99,265	56,774	57.19	517	56,257	28,494	27,763	50.65	49.35
Clare County	Clare	81,809	48,627	59.44	336	48,291	28,137	20,154	58.27	41.73

Figure 26: Prototype Excel

A polygon map of the counties and some major cities are created in Tableau. The excel data is brought into Tableau and the data is linked by a common field, in the case by 'COUNTYNAME'. The data is then refined and only relevant data is displayed on the visualisation. For the prototype, the data tells the user which counties were in favour of the referendum *(See figure 27)*.

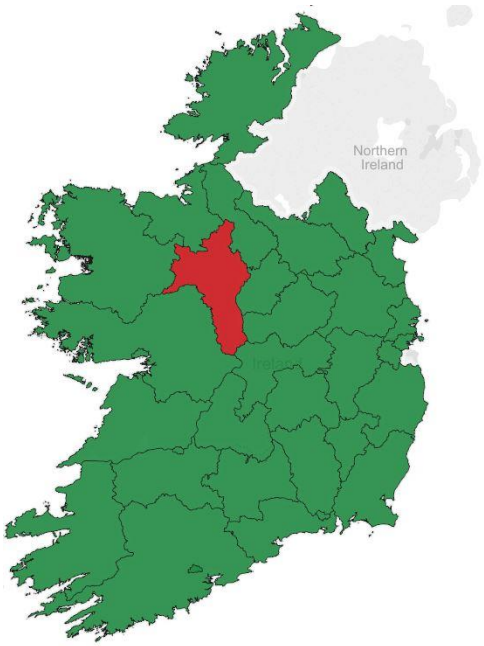


Figure 27: Counties in Favour of Referendum

Graphs in separate worksheets are created to show data that cannot be effectively displayed on the polygon map. In *figure 28*, a scatterplot graph is created that shows that counties that had a higher turnout also had higher voters in favour of the referendum.

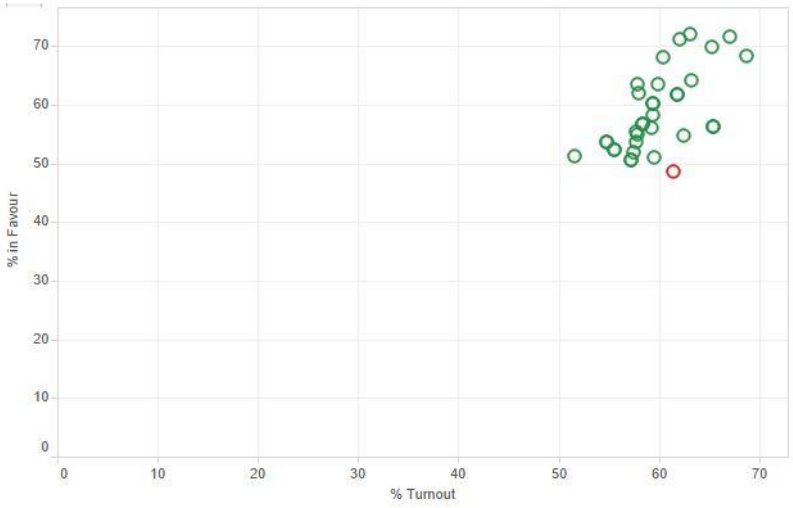


Figure 28: Marriage Equality Scatterplot

Figure 29 shows a simple bar chart that shows a better visual of the ratio of voters in favour and against the referendum. The chart also provides dynamic interaction which allows users to change the order of the chart based on county with the highest % in favour, county with the highest % against and county by alphabetical order.

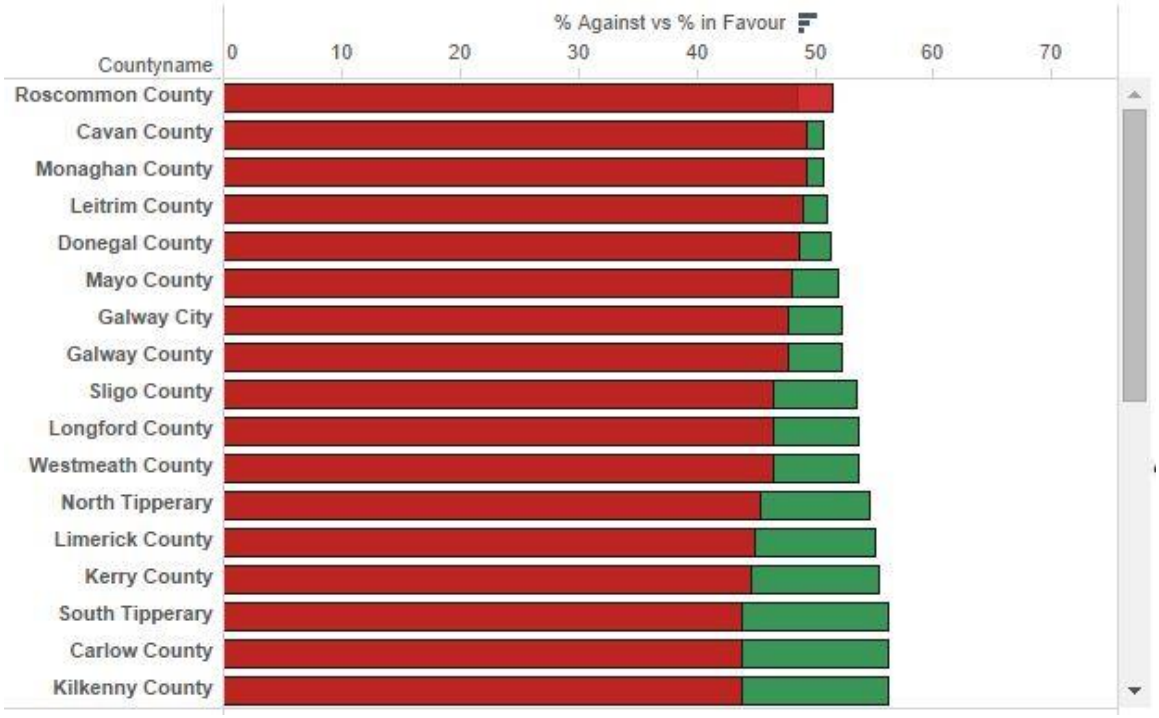


Figure 29: Marriage Equality Bar Chart

Once completed, these visualisations are merged together into a dashboard. A simple filter is added that allows the user to filter the data by county which links all the charts simultaneously. A prototype shows a visualisation with the polygon map occupying the majority of the space on the dashboard as it's what users should be drawn to initially. Beside the map the two graphs are displayed in a smaller size and above the graph is the county filter (*Appendices A).*

3.7 FINAL USER INTERFACE DESIGN

3.7.1 Census Ireland Banner

The Census Ireland logo is designed using a colour palette that starts from a deep blue up to light blue before transitioning through a white heptagon, then moving through a light orange to a dark orange. The concept came to mind after the Tableau dashboard layout was completed. When viewing some of the polygon maps, the colour moves from a dark blue to dark orange, representing values on the map. The logo aims to mimic the polygon maps by following the same patterns *(See figure 30).*

Figure 30: Census Ireland Banner

3.7.2 Navigation

The menu was originally designed to appear within the dashboard. However, Tableau only allows a limited number of elements per dashboard. After applying the core maps, graphs and filter elements, Tableau will prevent the application of any additional elements. (**See prototype).** The menu had to be implemented within the HTML code to provide the application with navigation. Each navigation brings the user to its associated dashboard. To help provide more insight to what the navigation is about, small icons are placed next to them which somewhat represent what the dashboard is about. For example, the births navigation shows an icon of a baby *(See figure 31).*

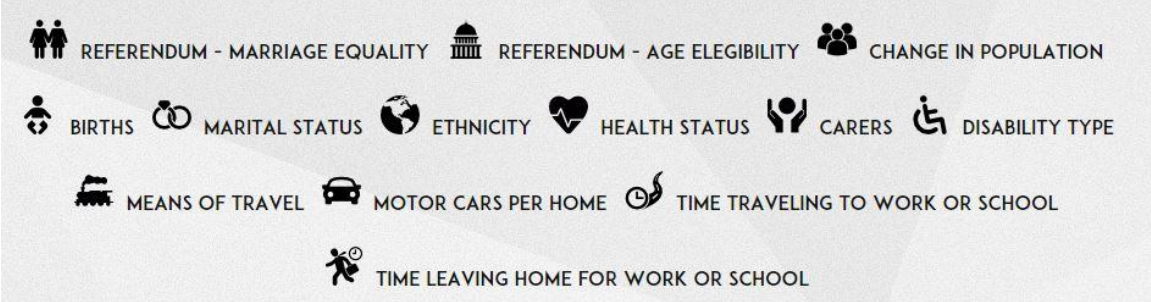


Figure 31: Census Ireland Navigation Menu

3.7.3 Tableau Dashboard Layout

The dashboards created in Tableau follow the same general layout. The Census Ireland logo is placed on the top of the dashboard. It should be noted that originally the navigation was meant to appear below the logo on the dashboard but due to Tableau's limit on dashboard elements, the navigation now appears above the visualisations in the HTML. Below the logo is the page title along with its accompanying icon *(See figure 32).*

Figure 32: Census Ireland Dashboard Header

On the right side of the dashboard the filters, legend and filter instructions for the map are placed along the top with the interactive map of Ireland placed below. On the left side of the dashboard, two graphs are placed. These two graphs vary by type based on the dataset the dashboard is currently displaying *(See figure 33).*

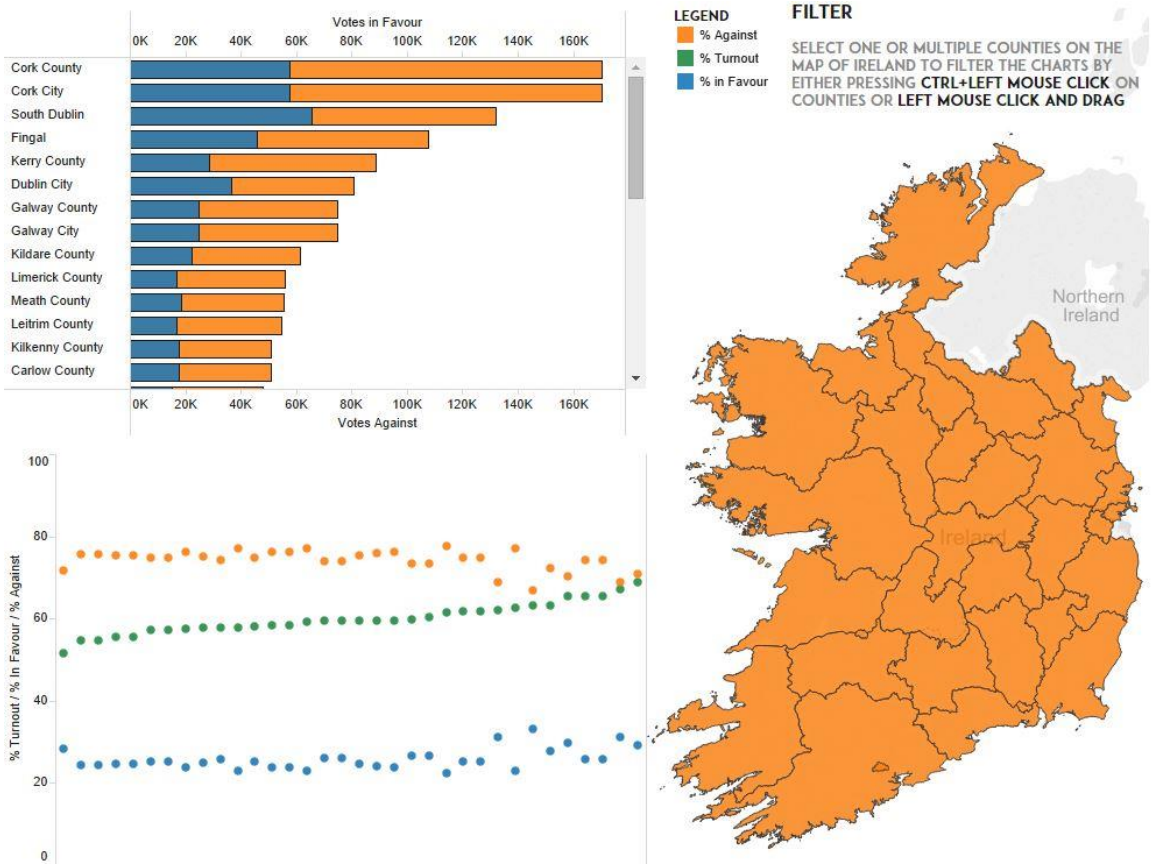


Figure 33: Census Ireland Visualisation Layout

3.7.4 Colour Scheme

The projects colour scheme was originally planned to be a combination of reds and greens to represent the data in the visualisations, as seen in the projects prototype. However after research it was determined that the colour scheme should be changed since statistics from Colour Blind Awareness show that worldwide about 8% of men and 0.5% of women suffer from colour blindness *[30]*. Despite an even smaller percentage of those being colour blind specifically to reds and greens, its best to ensure the data displayed in the visualisations are clear and easy to understand. Switching to an orange and blue colour scheme means that the project can maintain colours that clash. In addition to orange/blue and red/green, *figure 34* shows what other possible combinations the project could have adopted.

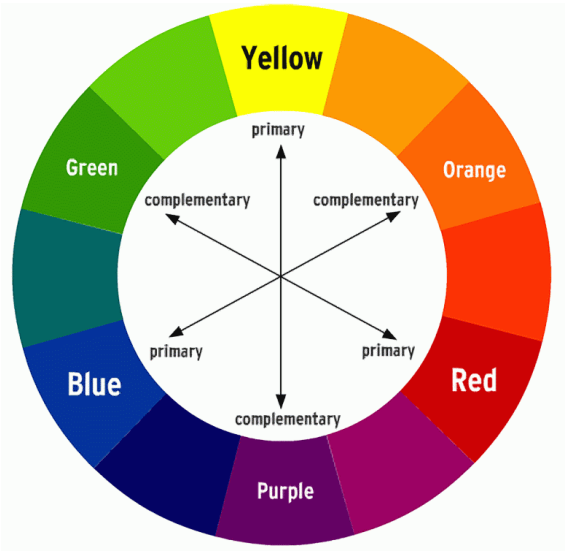


Figure 34: Colour Wheel

3.8 SUMMARY

The system will interact with the user through a web application. The web application allows the user to connect with the server side aspect of the project to deliver the data visualisations back to the web application for the user to view. These visualisations will be created using Tableau and stored on Tableau's public server. Through prototyping, the general layout of the project was decided. Though overtime, elements of the interface had to be removed or placed somewhere else on the web application. The colour scheme was also altered to cater to users who may suffer from colour blindness. With the system architecture and user interface of the project decided, it will make the implementation primarily focused on creating the visualisations with dynamic and interactive interactions.

4 Implementation

4.1 INTRODUCTION

In this section of the document, the implementation stage of the project will be explained. The process of creating the interactive visualisations will be detailed as well as how the excel data was integrated into the system. To acquire a working version of Tableau, the Tableau team had to be send an email with proof that the developer is a student. A free version of Tableau was quickly made available to the developer to begin work on the project. Though much of the implementation was self-taught through experimentation with the platform, two main sources helped in better understanding the software. The first was an article about creating custom polygons on a background image. Although not related to creating the polygon map, it explained how the system operates when working with polygons *[31].* The second was a tutorial on Lynda.com that goes into how to create intuitive dashboards in Tableau through the layout design and interactivity *[32].*

4.2TABLEAU POLYGON MAP

A polygon map was downloaded which was created by Adam Riley in the form of a Tableau Workbook *[33].* The workbook will draw out the all the counties excluding most of Ulster. It splits Dublin into four sections; Dublin City, Dublin South, Dun Laoghaire-Rathdown and Fingal. It also includes major cities such as Galway City, Limerick City, Cork City and Waterford City. With the Workbook opened, double clicking the '*longitude'* and '*latitude'* pills from the '*measurements*' section will place them in the column and row slots, respectively.

Pills are pill shaped elements in Tableau, they represent measurements and dimensions that allow users to create visualisations. The user can drag the pills into various tabs and sections to get functionality regarding that pill *(See figure 35).*

From the '*measurements'* tab, the *PointId, PolygonId* and *SubPolygonId* pills need to be dragged into the '*dimensions*' tab. Doing this will allow Tableau to map these points onto the map with the corresponding longitude and latitude.

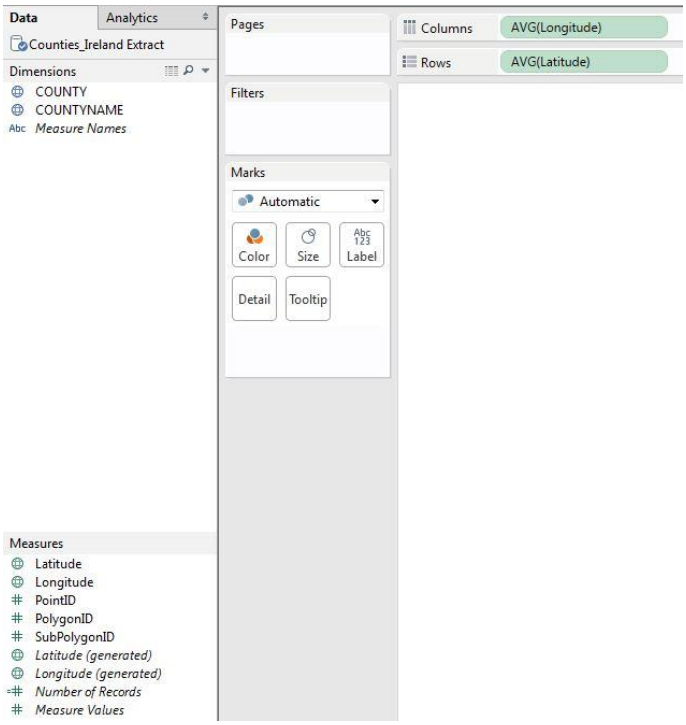


Figure 35: Starting Layout of Pills

In the '*marks*' tab, polygon needs to be selected which will open the path and detail. Once available, the *Polygon* and *SubPolygon* pills are dragged into *'detail'* while *PointId* is dragged into '*path'*. Doing this will draw the outline of each county in Ireland. The counties are drawn but are not visible with the current design. In *'colour'*, a border is added and the opacity of the background colour can be set. The background colour can be set to relate the any dimension or measurement, for example in this case the map has been set to have different colours for each county/city provided at random *(fig 36)*.

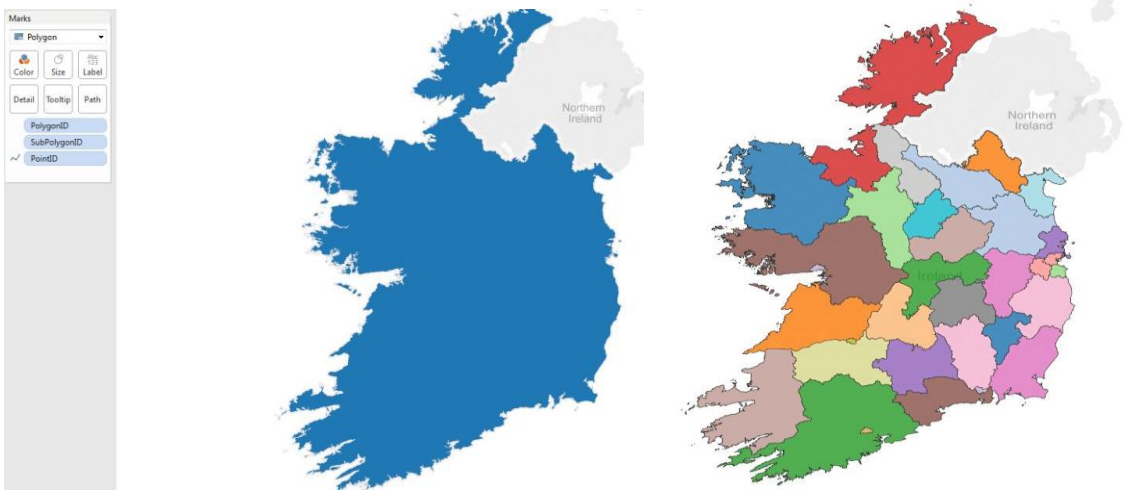


Figure 36: Ireland Map: No colour/No border - colour/border

4.3 INTERGRATING A DATA SOURCE AND DATA MAPPING

To add a new data source, the '*add new data source*' option is selected, in this the data file to connect to will be an Excel worksheet *(fig 37)*.

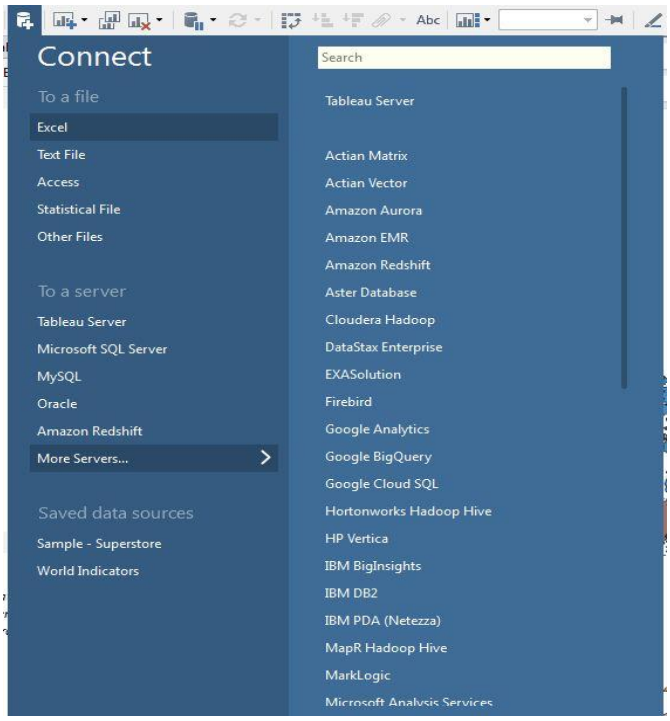


Figure 37: Connect to Data Source

Once the new data file is added, it needs to be linked to the polygon map by creating a new relationship between the two. In the data tab the Counties extract has to be highlighted. The 'Edit Relationship…' option is opened, the primary data source should be the polygon map with the secondary data source should be the selected excel worksheet. By default the automatic tick box is selected, tick the custom tick box instead. Click the 'Add…' button and set the relationship of the two data sources, in this case its set between the county names of each data source *(fig 38)*.

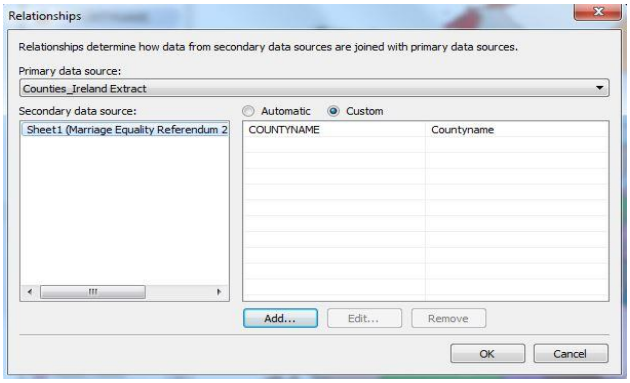


Figure 38: Creating Relationships

In the dimensions of the secondary data source, if the relationship is valid an orange link should appear beside the linking relationship *(fig 39)*.

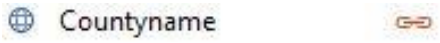


Figure 39: Relationship Link

The Measurements from the secondary data source can now be implemented into the polygon map. For example, the % in favour pill is dragged into the colour section. The colour is set to Orange-Blue Diverging given four steps of colour with the centre value being set to 50, a visualisation is created that displays the counties/cites that voted against the marriage equality referendum in orange, and those who voted in favour in blue, with dark blue being the most in favour constituencies *(fig 40)*.

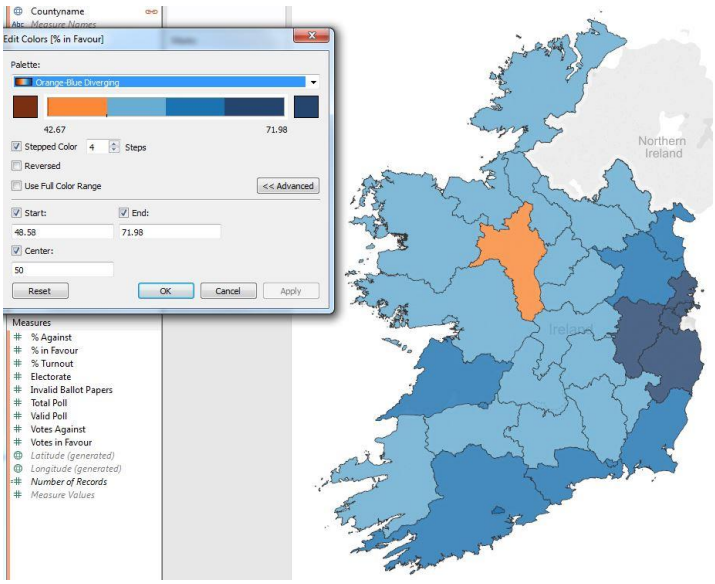


Figure 40: Coloured Marriage Equality Map

4.4 TOOLTIP DATA

The visualisation created up to this point is interactive but its tooltip currently displays values that most users do not need to see and since it is not designed yet, only displays default fonts that are black. These tooltips appear and show data based on what county is being hovered over, in this case Roscommon County is being hovered over *(fig 41)*.

PointID: **1522**
PolygonID: **30**
SubPolygonID: **1**
Countyname: **Roscommon County**
Latitude: **53.914**
Longitude: **-8.451**
% in Favour: **48.58**

Figure 41: Default Tooltip

In the *'marks'* tab, clicking on tooltip will open the edit tooltip window. There are several variables currently displayed that can be removed such as the *PointId, PolygonId, SubPolygonId, longitude* and *latitude*. The linking relationship is County name and since the tooltip will display data relating to that county, it's important to have it clearly visible on the top of the tooltip. The font, font weight and font colour can all be altered as well as the ability to add and remove data from any linking data source *(fig 42)*.

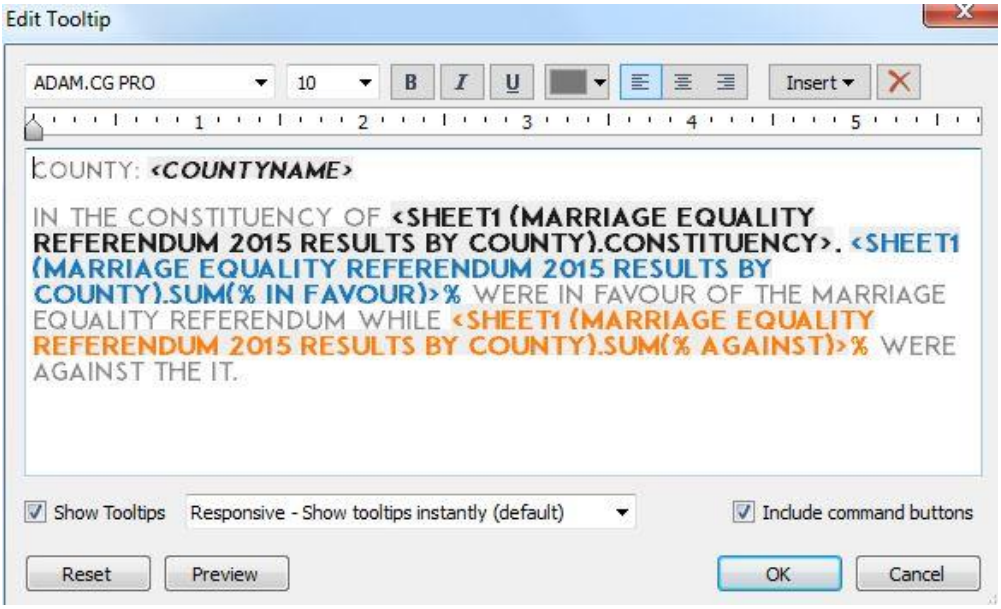


Figure 42: Edit Tooltip Dialog Box

In the text area, values are encased between < *value* >. As county name is being taken from the primary data source, it does not need to be specified and can be left as *'<COUNTYNAME>'*. The next value is pulling the constituency value off the marriage equality data source and as it is not the primary data source, it's linking sheet needs to be referenced. *SHEET1 (MARRIAGE EQUALITY REFERENDUM 2015 RESULTS BY COUNTY)* is the name of the table the data is being stored on. That is followed by *'.CONSTITUENCY'* which pulls the constituency value from the marriage equality data source and as county name is the linking relationship, the value of constituency will dynamically change based on what county is being hovered over on the main map. The same principles apply to the % in favour and % against values. As the counties in favour of the referendum appear in blue on the map, to keep consistency, the value is highlighted in blue on the tooltip with % against changed to orange *(fig 43)*.

Figure 43: Designed Tooltip

4.5 ADDITIONAL GRAPHS

The map and its tooltip provides a lot of information but it's not ideal for certain types of data. To combat this issue, additional graphs are added to the dashboard and change depending on the data that needs to be displayed. Following the creation of the marriage equality dashboard, there are measurements that show the total votes, votes in favour and votes against per county. This information can be displayed on a map but in this case sticking with a basic bar chart would convey the information much more clearly.

In Tableau, a new worksheet is created. To allow for global filtering which will be explained later on, the primary data sources linking relation needs to be present in each graph in some way. In this case from the primary data source *COUNTYNAME* is dragged into rows with the sum of votes in favour and votes against dragged into columns. This creates two separate bar charts *(fig 44)*.

Figure 44: Separated Bar Chart

To combine them into one overlapping chart, the second pill in columns (votes against) is right clicked and dual axis needs to be ticked.

The charts are now combined but their axis are not aligned, votes against top value is only 58k while votes in favour has 145k meaning the graph currently shows skewed values. To synchronize the axis, the pill that had dual axis ticked needs to be right clicked again, and have 'Synchronize Axis' ticked. The graph now displays correct comparison between the two values *(fig 45)*.

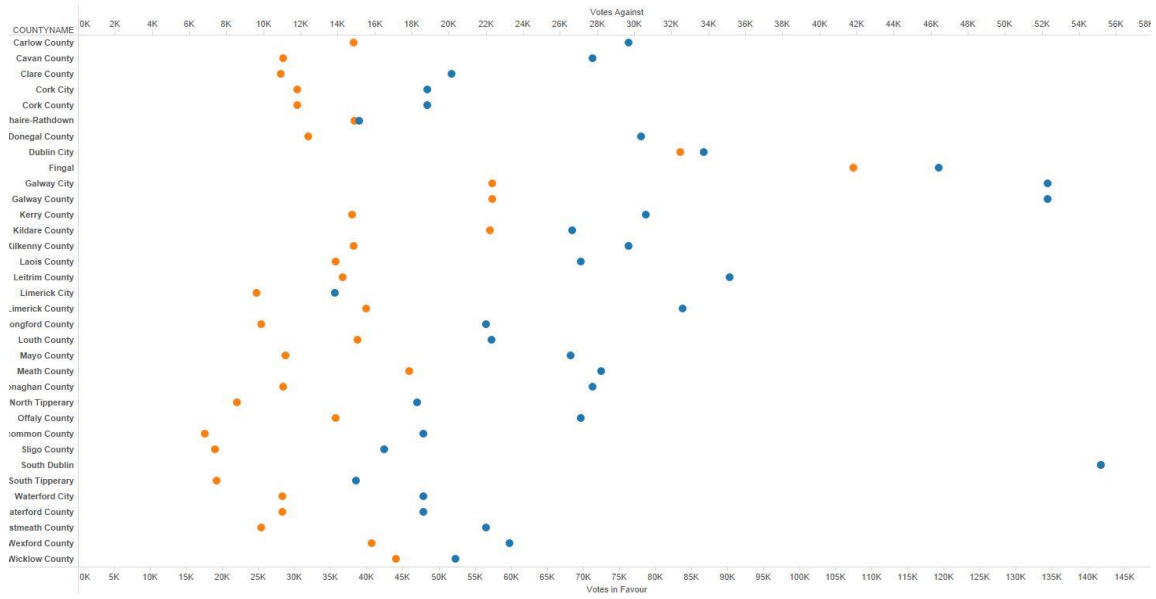


Figure 45: Bar Chart Combined

To change the visualisation from circles to a bar chart, change the dropdown menu selection from automatic to bar in the marks tab *(fig 46)*.

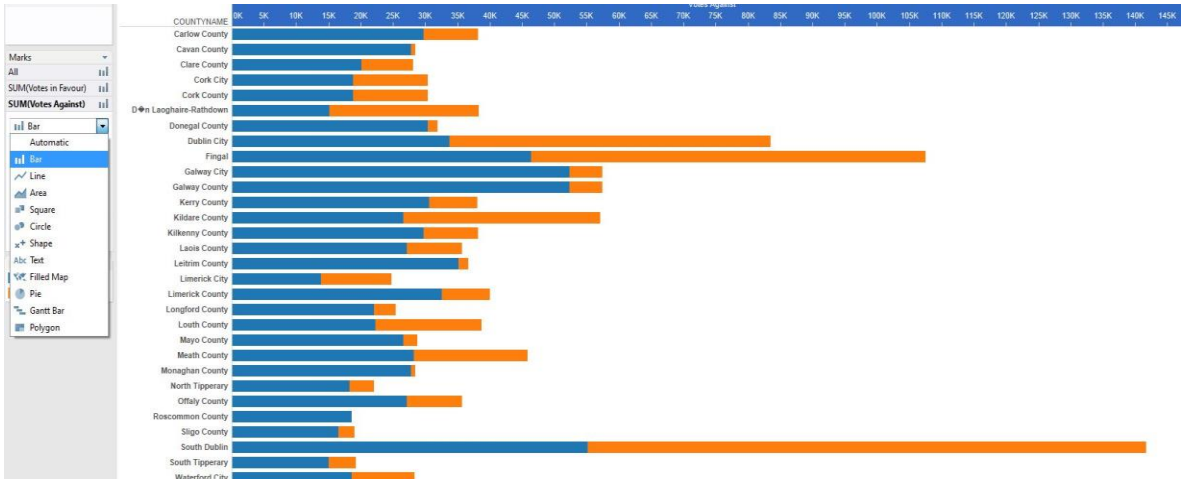


Figure 46: Bar Chart Designed

With the main structure of the graph complete, it's just a matter of altering the design of the graph and the graphs tooltip. Adding a custom filter by clicking on '*votes against'*, the graph can be changed to display the counties/cities in order of the amount of votes against in ascending or descending order. The tooltip for the bar chart will alter depending on if the blue or orange bars are hovered over. The finished graph can be seen in *figure 47*.

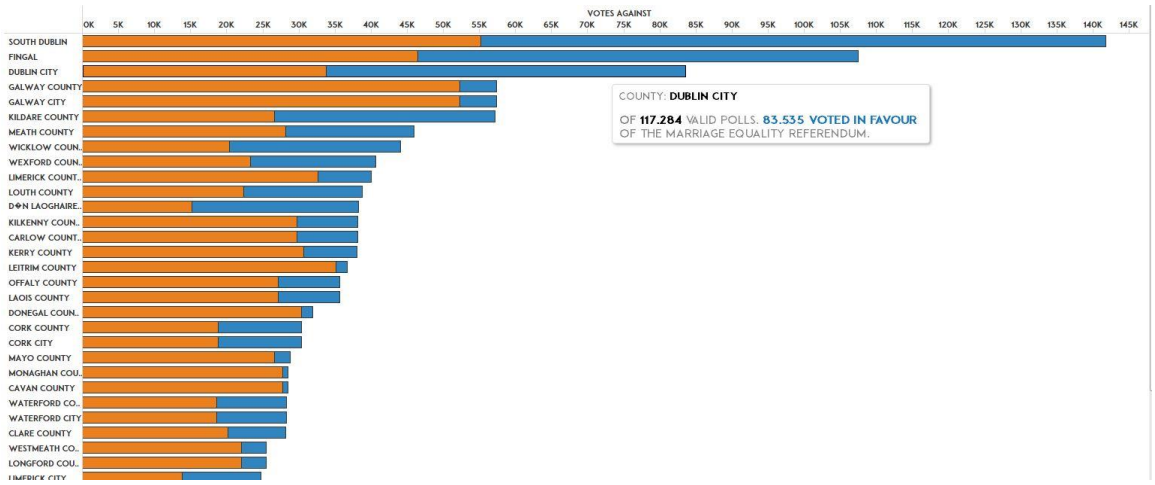


Figure 47: Finished Bar Chart

Adding three variables to a chart is generally not recommended in Tableau as adding more than two takes away the ability to create detailed tooltips. Regardless of this, the visualisation itself without a tooltip can show interesting data. In this example the % turnout (Green) per county is compared versus the % in favour (Blue) and % against (Orange). The visualisation shows that counties with higher turnouts were more in favour of the referendum *(fig 48)*.

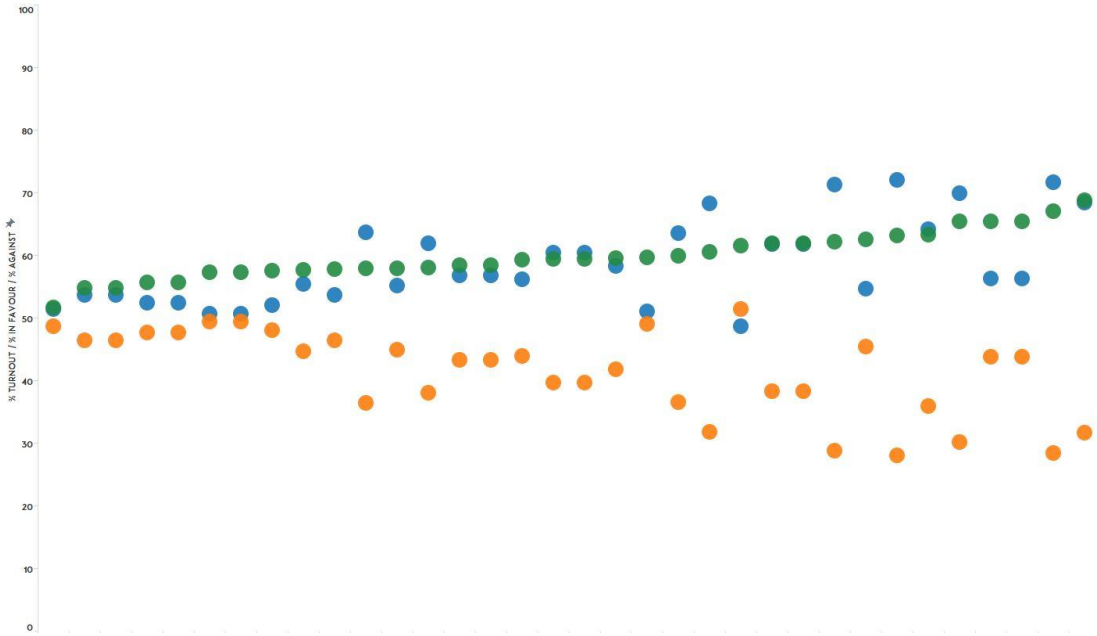


Figure 48: Scatterplot Graph

To implement three variables to one chart, the primary value (% turnout) needs to be placed in the column section. The two additional values need to be selected and dragged onto the same columns simultaneously. If done correctly, two green squares should appear on the cursor when hovering over the % turnout column *(fig 49)*.

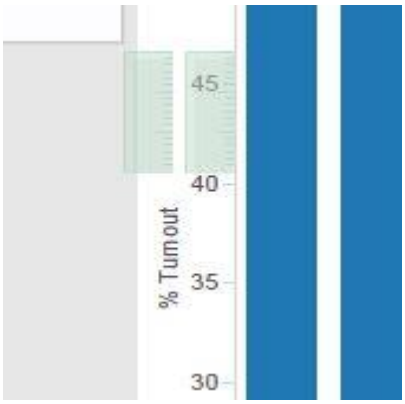


Figure 49: Combining the Variables

After combining the three variables, go to the marks tab and change the dropdown from automatic to circles which will clearly display all values *(fig 50)*.

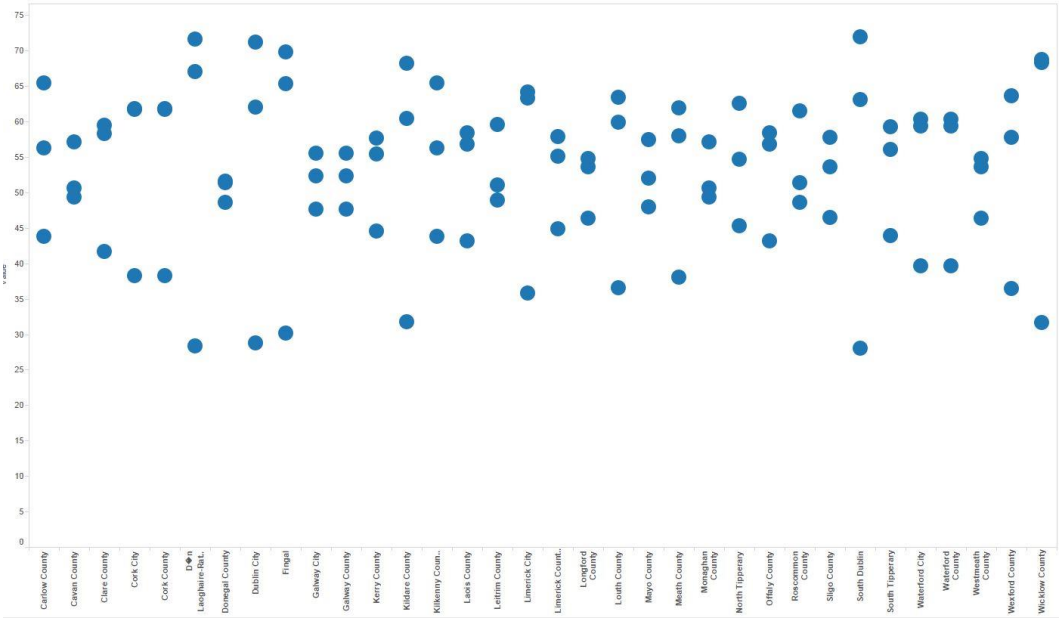


Figure 50: Scatterplot Uncoloured

Same as the bar chart, the scatterplot chart can be filtered to show % turnout in an ascending or descending pattern. Adding colour representing each value will give you the final result as shown in *figure 48*.

4.6 FILTERING

In Tableau, the filtering is a very powerful tool to help make visualisations more dynamic and interactive. For all dashboards the map of Ireland acts as a visualisation and as an interactive filter that alters the two accompanying graphs based on what counties are

selected. However in some dashboards, such dynamic linking is not possible because the dataset being displayed cannot be filtered by county/city. For graphs to be filtered as one, they all need to share a common value within them. As the map is based off the COUNTYNAME value, all other graphs that are to be filtered using the map need to incorporate the COUNTYNAME value somewhere within the graph.

Upon creating a new dashboard and after adding all the graphs that are to be filtered, mouse over the map element. A grey border should appear around it with an 'x' and an arrow in the top right. Click the arrow and tick the 'Use as Filter' option, the dashboard will then take a few seconds the re calculate the newly added filter function. Once completed and if the user has correctly incorporated 'COUNTYNAME' into the graphs, the map will now filter those graphs when a county is clicked on *(fig 51)*.

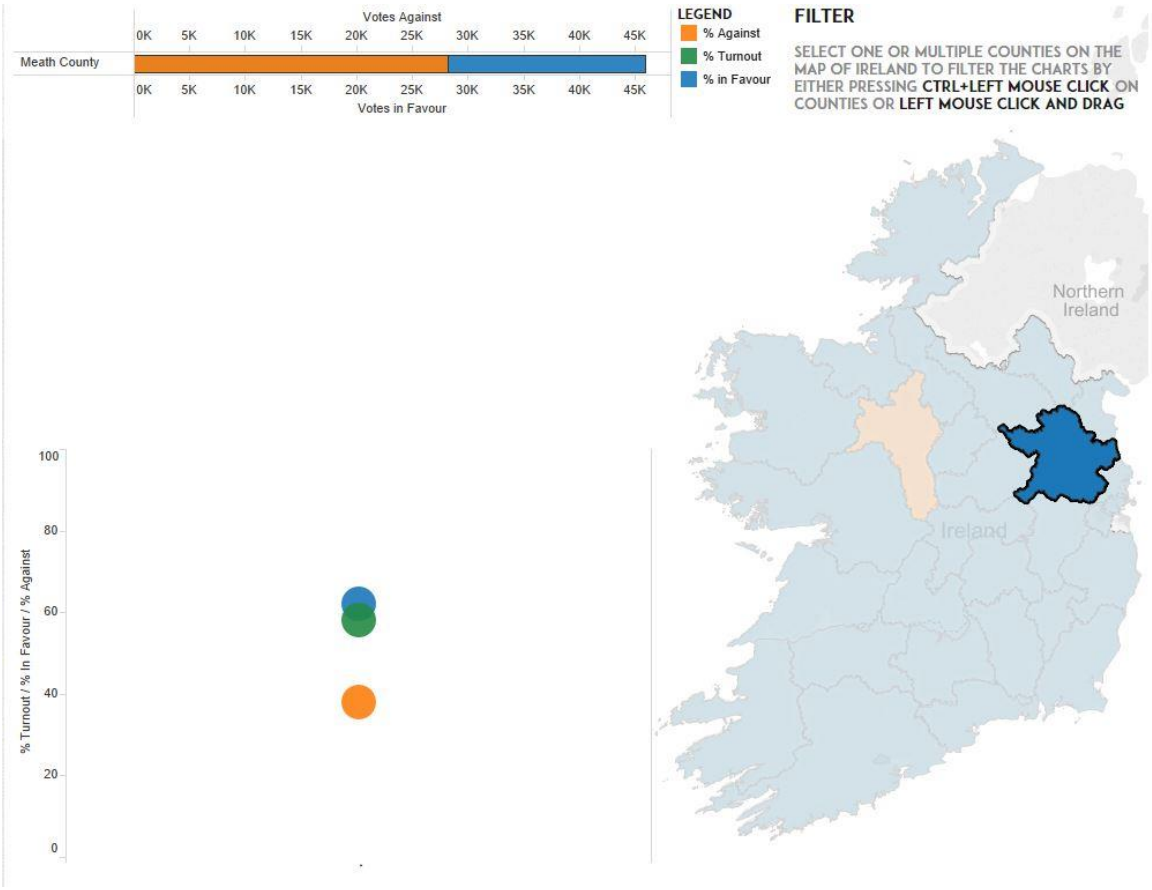


Figure 51: Using Map as a Filter

The user can *CTRL+click* on additional counties to keep adding more countries/cities to the graphs to efficiently compare specific counties against each other. To add the drag ability to the map to allow the user to select multiple constituencies at once, the map option on the top of the page needs to be clicked and *'Allow pan and zoom'* needs to be unchecked. This is useful because the data relates to Ireland only so users should not need to pan the map, adding extra functionality *(fig 52)*.

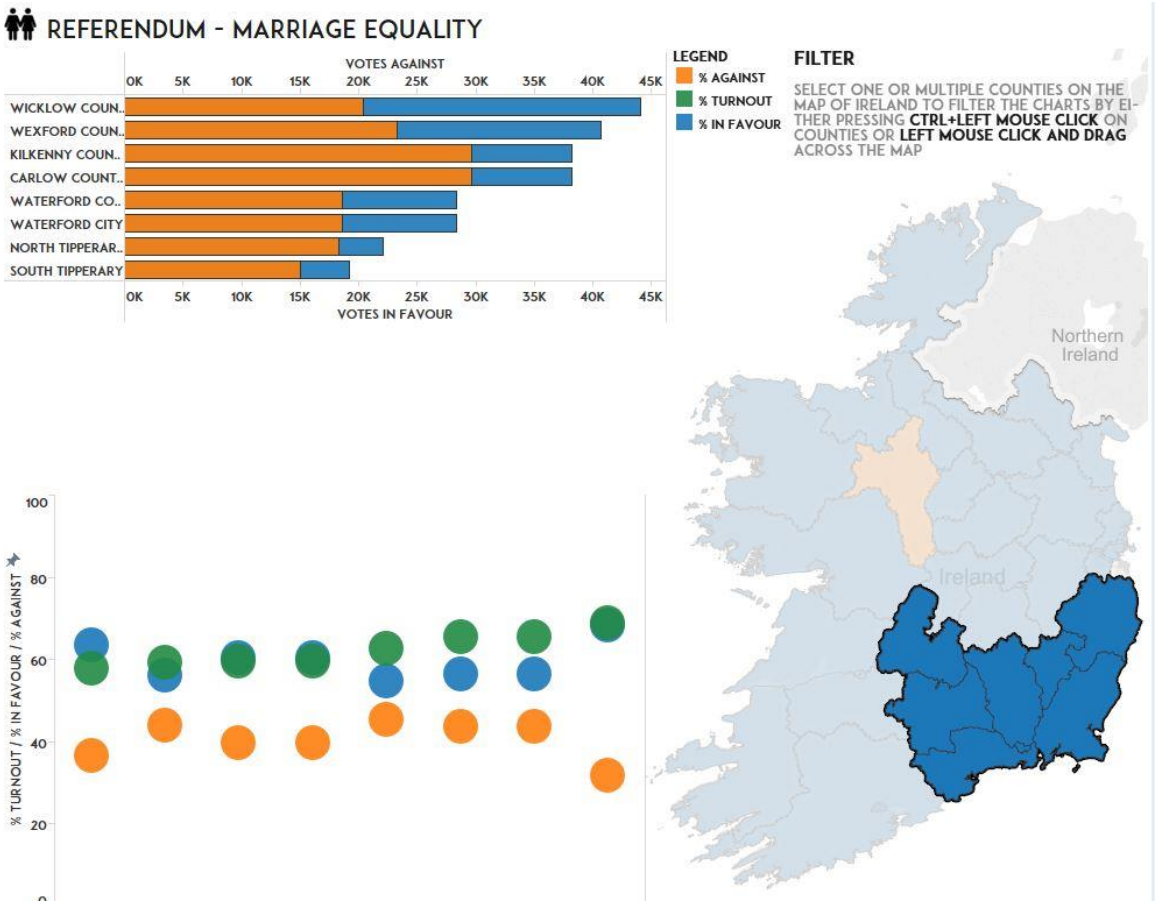


Figure 52: Map Filter Group Select

In cases where there are variables other than county to be filtered that cannot use the map, a dropdown menu filter is implemented. For example, in the 'Births' dashboard, only the top graph can be filtered but it can be filtered by county/city using the map or filtered by year by using the dropdown menu filter.

The dropdown filter follows the same principles as the map filter. The filter will automatically work with all graphs in the dashboard but it can be edited to only filter specific graphs by hovering over the filter which will make the grey border appear around it. Clicking on the arrow in the top right of the filter and selecting *'Applying to worksheet'* you can select *'only this worksheet'* , *'selected worksheets'* or *'All using the data source'*.

The filter also has various options that allow you to select single or multiple values to be filtered. For most of the dashboards, *'single menu (dropdown)'* is selected as allowing users to filter multiple values would take away some of the clarity from the data being displayed.

4.7 CREATING DASHBOARDS

Once the worksheets are completed, a new dashboard worksheet can be created. The dashboard allows you to drag these completed worksheets onto the dashboard off of the '*dashboards*' tab. Worksheets already on the dashboard have a blue tick in their icon *(fig 53)*.

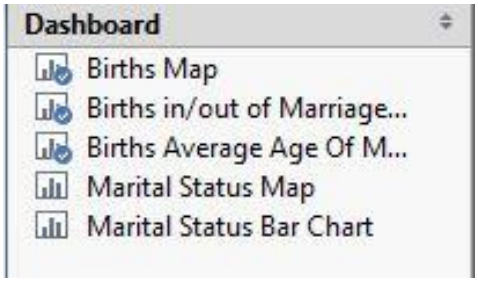


Figure 53: Dashboard Tab

Once an element has been placed on the dashboard it will automatically float, meaning it and all other elements of the graph will be resized to allow it to fit on screen. However the way these dashboards are laid out, the floating function has been switched off which allows the user to manually alter the x and y coordinates of the element, as well as its width and height *(fig 54)*.

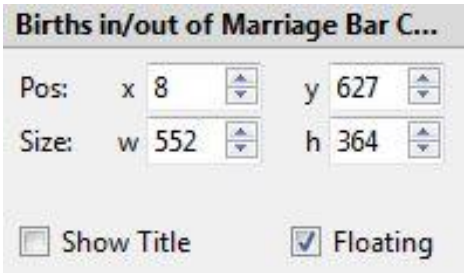


Figure 54: Dashboard Coordinates

When all the elements are laid out, the dashboard is ready to be uploaded to Tableau's public server.

4.8 UPLOADING TO TABLEAU PUBLIC

To upload to Tableau Public, a Tableau public account needs to be created. Clicking on server at the top of the screen will present the option to open Tableau Public, save to Tableau Public or manage account. Once signed in, selecting '*save to tableau public'* will open up a dialog box which will request that the data is extracted. Extracting the data means that only the data actually used in the dashboard is uploaded to the server instead of uploading everything. This helps save space which is helpful as Tableau Public only provides 1GB of free memory to store on the server.

When the dashboard is uploaded it is placed in the users Workbook on their personal profile which allows the title and description of the dashboard to be edited. It also provides the option to allow the dashboard to be visible to other users viewing the list of Workbooks, allows the dashboard to be downloaded by others or not and allows the user to delete the dashboard off of the server. This was useful in the project as many dashboards were uploaded for testing and prototyping and would not be suitable for others to view *(fig 55)*.

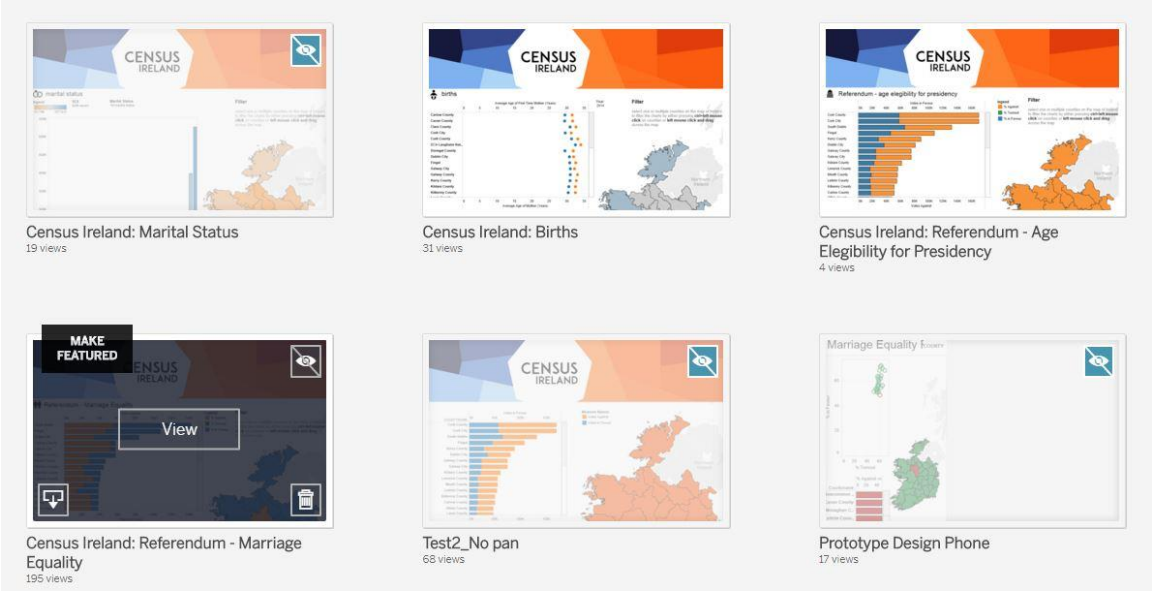


Figure 55: Tableau Public Dashboards

4.9 EMBEDDING A DASHBOARD TO HTML

Tableau Public provides embed code that allows the dashboard to be implemented within a web browser. Each dashboard has its own unique embed code. On the bottom right hand side of the dashboard is a share button. Clicking on the share button opens a dialog box which provides the ability to share the dashboard to Facebook, Twitter or email it. It also provides a link to the dashboard and an embed code.

Figure 56: Share Dashboard Dialog Box

The embed code is placed in a html file but appears in one long line of code and needs to be manually cleaned up to make it easy to read and understand. It starts by linking to a JavaScript file that links to Tableau Publics API. A placeholder is then established which sets the dimensions the dashboard will have within the web browser, it also calls the dashboard itself. A list of parameters are set up that will enable or disable various functions that are as follows:

- Show toolbar
- Animate transitions
- Display static image
- Display spinner
- Display overlay
- Display count
- Show tabs

All these parameters can be given a value of 'yes' or 'no' which will enable or disable their functionality as shown in the script below *(fig 57)*.

```
<script type='text/javascript' src='https://public.tableau.com/javascripts/api/viz_v1.js'></script>

<div class='tableauPlaceholder' style='width: 1004px; height: 1069px;'>
    <noscript><a href='#'><img alt='Marriage Equality Dashboard ' src='https:&#47;&#47;public.tableau.com&#47

    <object class='tableauViz' width='1004' height='1069' style='display:none;'>
        <param name='host_url' value='https%3A%2F%2Fpublic.tableau.com%2F' />
        <param name='path' value='shared&#47;9CD9J922Z' />
        <param name='toolbar' value='no' />
        <param name='static_image' value='https:&#47;&#47;public.tableau.com&#47;static&#47;images&#47;9C&#47
        <param name='animate_transition' value='yes' />
        <param name='display_static_image' value='yes' />
        <param name='display_spinner' value='yes' />
        <param name='display_overlay' value='yes' />
        <param name='display_count' value='yes' />
        <param name='showVizHome' value='no' />
        <param name='showTabs' value='y' />
        <param name='bootstrapWhenNotified' value='true' />
    </object>
</div>
```

Figure 57: HTML Embed Script

The navigation menu is also set up in the HTML using a series of *ul* tags to create each new line of navigation links. Each *ul* tag has a row of three *href's* which link to the various dashboards. Each navigation also has its icon placed next to it and is also clickable to bring the user to its associated dashboard *(fig 58)*.

```
<div class="container">
    <ul class="navbar">
        <li class="nav"><a href="index.html"><img class="image" src="./Logo/Lesbian Couple_Black.png">Referendum - Marriage Equality</a></li>
        <li class="nav"><a href="AgeEquality.html"><img class="image" src="./Logo/Capitol Building_Black.png">Referendum - Age Elegibility</a></li>
        <li class="nav"><a href="ChangeInPop.html"><img class="image" src="./Logo/Multiple Users Silhouette_Black.png">Change in Population</a></li>
    </ul>
    <ul class="navbar">
        <li class="nav"><a href="Births.html"><img class="image" src="./Logo/Baby With Diaper_Black.png">Births</a></li>
        <li class="nav"><a href="MaritalStatus.html"><img class="image" src="./Logo/Interlocking Rings_Black.png">Marital Status</a></li>
        <li class="nav"><a href="Ethnicity.html"><img class="image" src="./Logo/World_Black.png">Ethnicity</a></li>

        <li class="nav"><a href="HealthStatus.html"><img class="image" src="./Logo/Heart Beats_Black.png">Health Status</a></li>
        <li class="nav"><a href="Carers.html"><img class="image" src="./Logo/Circle Between Hands_Black.png">Carers</a></li>
        <li class="nav"><a href="DisabilityType.html"><img class="image" src="./Logo/Wheelchair Accessible Sign_Black.png">Disability Type</a></li>
    </ul>
    <ul class="navbar">
        <li class="nav"><a href="MeansOfTravel.html"><img class="image" src="./Logo/Old Train_Black.png">Means Of Travel</a></li>
        <li class="nav"><a href="MotorCarsPerHome.html"><img class="image" src="./Logo/Sports Car_Black.png">Motor Cars Per Home</a></li>
        <li class="nav"><a href="TimeTravel.html"><img class="image" src="./Logo/Clock on Road_Black.png">Time Traveling to Work or School</a></li>
    </ul>
    <ul class="navbar">
        <li class="nav"><a href="TimeLeavingHome.html"><img class="image" src="./Logo/Going to Work_Black.png">Time Leaving Home for Work or School</a></li>
    </ul>
```

Figure 58: HTML Navigation

4.10 ISSUES

When the excel sheets they were unable to be implemented with Tableau. The Excel worksheets had to be reformatted for Tableau to read them. The process in which the formatting took place is discussed in the Design chapter.

A major issue with performance is arises due to the way the dashboards are created. The map of all the counties is based on a large data source, every time a dashboard is loaded into the web application Tableau has to pull all the points off of the server and draw out the map along with the graphs. Then every time a filter involving the map is initiated, Tableau has to redraw the map again but also rearrange the data into the map based on what type of filter was activated. This means when a filter is activated, Tableau can take up to 20 seconds connecting to the server and redrawing out the data. This is a huge issue for the user as the user may think the application is broken, or they will be too impatient to wait for it to load fully.

A minor issue with the dashboard is the font currently being used. A font called ADAM.CG PRO *[34],* as seen in the logo, is used for much of the titles and text in the dashboard. However, if the dashboard is opened on a computer without the font installed, it reverts back to times new roman. A list of fonts is available on Tableau's website and an article of aesthetically appealing font combinations *[35]* is available to draw inspiration from.

When data is appended to the Dun Laoghaire-Rathdown section on the map is disappears from view and cannot be interacted with, but still appears in additional graphs. The cause of this is unknown but is speculated to be because the u in Dun appears as an unknown character in the data table.

When the user filters the dashboard when it's not scrolled down fully because the mouse location being displayed on the monitor actually interacts with the dashboard a few pixels above it. A fix to this would either try change the navigation appear beside the dashboard instead of above so the web application does not need to scroll.

An issue with the navigation occurred when trying to implement within the dashboard below the logo. However Tableau has a limit on the amount of elements that can be placed in a dashboard. This was solved by implementing the navigation within the HTML above the dashboard.

4.11 SUMMARY

The polygon map of Ireland with the counties and major cities borders were drawn out in Tableau which will act as the main visualisation for the dashboards. Data acquired from the CSO and Referendum Ireland websites will be formatted in Excel Worksheets which will allow Tableau to append the data into the map visualisation. The tooltip that appears when the user hovers their cursor over a county was designed to better display additional data about the county and to help the user feel more engaged through interaction with the visualisation.

The map visualisation is accompanied by two additional graphs in each dashboard. These graphs will help better display data that users might not be able to understand if it was shown in the map visualisation. Most of these graphs will dynamically filter the data they are displaying when the user uses the map visualisation as a global filter. Doing this will allow users to get a better understanding of the data and displays only the data they wish to see. Both the graphs and the map visualisation are typically affected by the dropdown filters that appear on some of the dashboards. For example, this means instead of the dashboards filtering by county, they will filter by gender.

Once the dashboards are completed, they are uploaded to the developers Tableau Public profile. The developer can then decide whether to make the dashboards publicly visible or hidden and edit details, such as the dashboards title. With the dashboards on the server, an embed code is used that is unique to each visualisation is placed into a HTML file. Each HTML has its own dashboard and a navigation section which allows users to access other dashboards.

Issues that have arisen during the implementation have effected user experience. The primary issue being that dashboards being initially pulled into the web application is a slow process and can take up to 20 seconds, while the map filtering can take up to 30 seconds. This is a major issue for users as the majority of users would likely lose interest before the dashboard is finished loading or think it is broken.

5 Testing

5.1 INTRODUCTION

To test the system functioned correctly, unit tests were performed. Dashboards were create one function at a time and were slowly implemented together over time to ensure they worked in unison. In the usability testing for Census Ireland, three quantitative usability and system goals are identified and discussed along with the type of tests that will be performed to examine whether the application has achieved those goals. Ten participants will be gathered and ask to perform a pre-test questionnaire which will identify the demographic of the participants. They will then be asked to perform three tasks which will test the usability and system goals. They will then be asked to fill out a post-test questionnaire which will determine the user satisfaction of the web application. The data gathered will then be analysed and interpreted to help determine the applications success in achieving its goals. Finally, an evaluation of the results gathered will be carried out to determine what may have gone wrong, what would be done differently if the project could be restarted, and formulate recommendations for future changes of the application.

5.2 UNIT TESTING

As Tableau was a software the developer was unfamiliar with, each function had to be tested on its own before being implemented into the dashboards. The first unit test performed was embedding data from an excel sheet into the map of Ireland to work interactively. Instead of testing the map with all the counties, a map layer with only the four provinces was used, as well as a small dataset in the excel sheet.

Once the interactivity was achieved, the dropdown filter function was added which filtered the provinces. Along with the dropdown filtering, the tooltips were also set up so they changed what data they displayed depending on the filter selected. Separate graphs were created and dragged into the dashboard with the map data to ensure they all filtered with the dropdown filter. The advanced filtering was then added and tested extensively to get it functioning correctly with all the other functions.

Once the province prototype had all its functions working, the map was changed from provinces to counties, the excel data was increased and the whole process was started again. Starting with the provinces ensured the developer had a full understanding of how Tableau worked and allowed for an easier transition into working with the counties/cities map.

5.3 QUANTITATIVE USABILITY GOALS

- **Time comparison between a user finding data on the map visualisation versus user finding data on formatted table should be in favour of the map visualisation.**

Census Irelands primary focus is to help users easily view census and referendum data on visualisations. The purpose of this test is to ensure that the user can find the data more efficiently on the visualisations than they can on the formatted table.

- **The time taken for a user to identify the counties/cities in Ireland where the female population has declined and the percentage of that decline from the previous census should be less than or equal to two minutes.**

As Census Ireland is a relatively small web application, the amount of time to complete any task within the system should be minimal. However, the fact that the dashboards can take nearly up to 20 seconds to load should be taken into consideration. Novice users may get impatient with the long wait time and try to refresh or go back a page to reload the dashboard, only increasing the time to complete the task.

- **The time taken for a user to learn how to use advanced filtering should be less than or equal to two minutes and thirty seconds.**

Advanced filtering is the functionality of the dashboard that allows users to interact with the map of Ireland as a filter, which dynamically changes the displayed data in the accompanying graphs to show data relative to the selected counties. Testing how long the participant takes to learn how to use this functionality will help determine if it is manageable to new users or too difficult.

5.3.1 Goals Overview

The tests that are to be performed should determine if Census Ireland is easy to use, easy to learn and performs better in comparison versus data found on a formatted Excel Worksheet. Participants will be given a pre-test questionnaire before being given specific tasks to perform, followed by a post-test questionnaire. The data will be documented in an Excel worksheet and be displayed in visualisations.

Ease of Use

Ease-of-use focuses on how easy it is for a user to use a product. In this case, users will be asked to perform tasks on Census Ireland to determine if the product is user friendly. For an ease-of-use goal to be considered a success, the results after testing must meet the characteristics of the 5 E's *[36]:*

- **Effectiveness**: Determined by looking at whether the user met their given tasks and whether they performed them correctly. This can be achieved through well-presented information on an interface and the use of terminology the user can understand.
- **Efficiency**: Determined by how fast and accurately the user completed the tasks. This can be achieved through navigational keyboard short-cuts and having less path ways to achieve certain tasks.
- **Engaging**: Determining if the product is satisfying to use, typically associated with the visual design and aesthetics of the product. This can be achieved by providing a theme for the product with a visually appealing colour scheme and having interactivity for the user.
- **Error Tolerance**: How the product prevents errors from occurring and the ability to help users recover if an error is encountered. This can be achieved by making it difficult for users to take actions that are incorrect, invalid or irreversible.
- **Easy to Learn**: How the product presents new functionality in a way that the user can quickly learn without much effort. This can be achieved by keeping the product consistent and providing instructions for more complex tasks.

Ease of Learning

Ease-of-learning focuses on how easy it is for infrequent users and first time users to learn to use the product, and how easy it is to remember. In this case, users who have no experience using Census Ireland will be asked to perform tasks to determine if the product is intuitive, or if it has a forgiving learning curve. The system is small and does not focus on catering to "*Power Users" [37],* users who will use the product extensively, so the testing will primarily target casual users. This will result in assuring the user interface is easy to use and understand and that the more complex functions are clearly specified to the user.

Performance

Performance focuses on the actual user performance in a system. Performance links well with ease-of-use and ease-of-learning because if the user can easily use and learn the functionality of the product, their performance will generally be better. To calculate performance metrics, the time and errors made is recorded when the user is given a task. This helps in determining if a specific function takes the user too long to figure out or if the user interface is too confusing that it misleads the user.

5.4 TEST MATERIALS

5.4.1 Pre-test Questionnaire

To obtain demographics, a pre-test questionnaire was issued to participants to complete before proceeding to the test tasks. The pre-test questionnaire will determine the participant's age, sex and how experienced they are with data visualisations. Rather than ask outright if they consider themselves experienced with data visualisations, the participants are asked if they prefer visualisations or raw data, if they find data visualisations typically easy to understand and if they create data visualisations themselves. Depending on how they answer these questions, it can be determined if they are a novice or expert. The pre-test questionnaire can be found in *Appendices B*.

5.4.2 Test Tasks

Once the pre-test questionnaire is complete, the participants will be asked to perform three tasks the will test Census Ireland's ease of use, ease of learning and performance quantitative usability goals. The tasks carried out will measure the time taken by participants to complete the tasks and the amount of errors the system or participant encountered.

The first task given will ask participants to identify the amount of counties/cities that voted in favour of the age equality referendum displayed on an Excel Worksheet and then will be asked to identify the amount of counties/cities that voted against the marriage equality referendum shown on the map visualisation.

In the second task given to participants, they will be asked to identify the counties/cities in Ireland where the female population has declined and the percentage of that decline from the previous census within two minutes. To complete this task the user will have to take advantage of the majority of the applications functionality excluding the map filtering (though it is optional). The time taken and the amount of errors the system/participant encountered will be documented.

For the final task, participants will be asked to navigate to the "Births" dashboard and display data from Donegal, Galway County, Mayo and Wicklow in 1992 in the "Average Age of First Time Mothers" graph. Performing this task will require the user to use the dropdown filter to switch the data to 1992 and then use the map filter to display data from the previously mentioned counties. The time taken to complete the task and the errors the system/participants encountered will be documented.

5.4.3 Post-test Questionnaire

The post-test questionnaire will be given to participants after completed the test tasks. It will be used to determine user satisfaction of the web application. The questions are taken from the System Usability Scale (SUS). The post-test questionnaire can be seen in *Appendices C.*

5.5 DATA COLLECTION

Firstly, participants will be required to answer the questions on the pre-test questionnaire. The questions will be presented to them by the interviewer on an A4 page and will be answered on paper and later added into Excel. The data collected for the pre-test questionnaire can be seen in *Appendices D.*

The participants will then be asked to complete the tasks detailed in the "Test Tasks" section. For the first task the user will be shown the Excel worksheet of raw data and then shown the map visualisation. As they are different sets of data, the order in which they are shown is irrelevant. A stopwatch will be used to keep track of the time taken by starting when the data/visualisation is displayed on screen and stopped when the user says the correct answer out loud.

For the second and third task, the user will be placed at the applications landing page. Time will start when the user is told to go and time will be stopped when the user completes the task. If the system or participant encounters any errors or bugs, they will be documented. As each of the tasks should be completed in two minutes and thirty seconds, if the participant surpasses this time by a significant amount they will be told to stop, and it will be noted that they did not complete the task. In addition, if the user is unable to complete the task and opts to give up before the two minute mark, it will be documented as an incomplete task. The data collected for all three tasks can be found in *Appendices E.*

Once the test tasks have been carried out, the participant will be provided a post-test questionnaire. Same as the pre-test questionnaire, the post-test will be presented to them by the interviewer on an A4 page and will be answered on paper and later added into Excel. The data collected for the post-test questionnaire can be found in *Appendices F.*

5.6 SUMMARY OF GATHERED DATA

5.6.1 Analysis & Interpretation

From the ten participants used in the usability testing of Census Ireland:

- four participants were female and six were male

- four participants were aged 18-24, four were aged 25-34 and two were aged 55+
- three participants agreed to preferring visualised data over raw data while seven strongly agreed
- Six participants strongly disagreed to not understanding visualisations, two disagreed and two agreed to not understanding visualisations
- Five participants strongly disagreed to creating their own visualisations, four disagreed and one agreed.

With the data gathered from the pre-test, it can be determined that regardless of gender and age, the participants taking part in the testing of Census Ireland are intermediate users who prefer visualised data over raw data, generally understand visualisation but do not create their own.

When participants were given the first task to discover information on a visualised map versus a raw data table, all the users passed the test and were much quicker in finding the data on a visualised map. As the participants did not need much interaction with either, there was nothing significant to note that may have affected the results.

In task two, only one participant went over the expected time limit because of the applications long load times and were asked to stop the task after three minutes. The other nine participants passed the task under the two minute mark. There were a few key notes:

- Participant #2 navigated to the incorrect page before finding the correct page
- Participants #3 and #5 encountered the cursor error which misplaces the cursor down a few pixels from where the actual interaction on the visualisation should occur
- Participant #9 failed the test because of the long load times, while participant #6 also did the same thing but managed to pass the test eleven seconds before the two minute mark

In the third task, all the participants managed to complete the task under the two and a half minute mark, though six of the participants encountered errors:

- Participants #1, #4, #7 and #8 all encountered the cursor error
- Participants #2 and #10 tried advance filtering the counties first, but found when the used the dropdown filter to access the 1992 results that it reset the map filter, so they had to perform the advanced filter twice
- Participant #9 needed to be told how the advanced filter operated

The post-test questionnaire contains the System Usability Scale (SUS) which includes ten items using a five point response scale (strongly agree – strongly disagree). With the exception of participant #9, the other participants responded well to Census Ireland in terms of usability, ease-of-use and ease-of-learning. However, a number of participants encountered errors that hindered and negatively affected the appeal of the application.

5.6.2 Statistical Analysis

For task one, a two-tailed paired t-test was carried out to determine if the time difference between the visualised and raw data were statistically significant. A decision risk of 5% is used. The decision risk (α) allots half to testing the statistical significance in one direction and half to testing the other direction, giving 2.5% to each tail in the distribution of the test statistic. Using the T-table with 9 degrees of freedom and the cumulative probabilities of 97.5%, the critical values are calculated at ± 2.262. As the calculated t is -3 which appears within the critical region, it can be determined that the difference between the visualisation and raw data table are statistically significant in favour of the map visualisation. To see the worked out statistical analysis and the visualised results, see *Appendices G.*

For task two, a statistical dashboard was created. The first chart shows the time it took each participant to complete the task in seconds. Participants who passed have their bar in green while participants who failed have a red bar. Out of ten participants, one did not pass the task. The blue chart shows which participants encountered an error in the application. Ideally all should be zero but two participants encountered the cursor error. The dashboard can be seen in *Appendices H.*

For task three, another statistical dashboard was created. The charts show the same data sets as the previous task, the time completing the task in seconds with the bar colour showing green if the passed and red if they failed. In this case however, all participants passed. With the more advanced functionality, more errors arose. This time four participants encountered the cursor error and two encountered a filtering error. The statistical dashboard can be seen in *Appendices I.*

5.7 SUMMARY

From analysing the gathered results from the user testing, it can be determined that Census Ireland achieved its ease-of-use and ease-of-learning quantitative usability goals. However, although the application generally performed well with users, the cursor error was encountered too many times and will have to be fixed to ensure the system meets its performance goal. Another issue with performance was the long load times of dashboards. Though rare, some participants were confused by the long load times and thought the application was broken or needed a page refresh.

As mentioned in the research and analysis chapter, the server side lag did become a problem with achieving the performance usability goal. Unfortunately, as the visualisation dashboards are stored on Tableau's public server, nothing can be done to improve the long load times.

The usability testing revealed some unforeseen bugs and errors. The cursor error was encountered enough by participants to be perceived as a threat to achieving the usability goals. Though it did not prevent participants from completing tasks, it definitely hindered their progress, as participants who encountered the bug had higher completion times on tasks. A minor error encountered through testing found that computers that did not have the font used in the dashboards installed, had their font set to the default of times new roman. This issue can be fixed easily by simply changing the font used to a universal font. The filter error was also discovered through user testing. Though the cause of the error is still unknown, additional work on the application will have to be done to remove the bug.

To conclude, through the testing carried out on Census Ireland, it was determined that the application achieved its ease-of-use and ease-of-learning usability goals, but additional work will have to be done on the application to achieve its performance goal.

6 Conclusion

The project set out with the aim to create interactive and dynamic data visualisations that represent census and referendum data to users that were easy to use and understand. Through multiple iterations and tests, Census Ireland has definitely achieved its goal. The web applications primary focus, the map of Ireland, became fully interactive and even became a filter that dynamically changed the displayed data in the accompanying graphs based on the counties and cities the user selected. With the amount of data acquired, thirteen dashboards were created from the 2011 census and 2015 referendum results that users can access.

There would be a few things that would be done differently if the project were to be started again. Primarily, finding a way to improve the applications performance would be vital. Finding an alternate way to store the visualisations off a private server with better performance than Tableau's public server may have been a better direction to go. A couple of the dashboards created show data that is predictable, such as Health Status and Carers. The map visualisations show what would be expected, which is that the higher population areas of Ireland have the higher amounts of people with health issues and carers. If starting again, these dashboards would be removed entirely and more focus would have gone into dashboards with more interesting data. The project also initially was meant to implement D3.js charts along with the Tableau dashboards, but were removed when it was discovered Tableau could achieve the projects goals on its own. If starting again, finding a way to implement D3 with Tableau in a way that was still as functional and visually appealing would be an interesting undertaking.

At the time of this report, the project has a huge opportunity to be expanded. The 2016 census is taking place on April 26th. This means that once the census results are released, the application would have a huge amount of new data to implement. Many new dashboards could be created that would analyse the new data, and new dashboards could be created that would compare the 2011 census results to the new 2016 results.

There is a lot to be taken from the experience gained from the development of this project. Over the course of the project, a lot was learned about data visualisation, about how the key to creating a well-functioning visualisation four things are required; Information, Story, Goal and Visual Form. These four along with the information learned about Gestalt principles, C.O.N.V.I.N.C.E. and colour schemes, came together to help make what would become Census Ireland.

Working with Tableau proved to be a valuable learning experience. Working with new technologies can often be difficult, the same can be said about Tableau. Though there are tutorials and articles available, the ones found did not help when it came to binding data to a custom polygon map. Only after finding a seemingly irrelevant tutorial about a man mapping data about beer to a supermarket shelf was the core function of the project

found. With help from Cyril, and some self-learning did I understand how to implement the full functionality of the project using Tableau.

I also learned the value of user testing. User testing carried out on Census Ireland helped ensure that it was easy to use and that the visualisations were easy to understand. It also shed light on some problem areas that helped improve the overall quality of the project.

With the primary aims and goals of Census Ireland achieved, and with the knowledge I gained throughout the development of the project, I am pleased with how the final iteration of the project turned out and to have had this unique experience creating it.

7 References

[1]. Friedman, V (2008). "Data Visualization and Infographics", Smashing Magazine. January 14th, 2008. Retrieved November 8th, 2015. Source: http://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/

[2]. Tufte, E (2001). The Visual Display of Quantitative Information. Graphic Press. 2001.

[3]. Friendly, M (2002). "Visions and Re-Visions of Charles Joseph Minard", Journal of Education and Behavioural Statistics, Vol. 27. Spring 2002.

[4]. McCandless, D (2009). Information is Beautiful. HarperCollins. UK. March 31st, 2009.

[5]. Sleeper, R (2015). Evolving Analytics. Retrieved: November 9th, 2015. Source: http://www.evolytics.com/data-driven-storytelling-tips-iron-viz-champion/8-dont-use-pie-charts/

[6]. Isson, J-P., Harriott, J. (2012). Win with Advanced Business Analytics: Creating Business Value from Your Data. John Wiley Sons Inc. October, 2012.

[7]. Fox News Exaggerated Bar Chart Image. Retrieved: November 10th, 2015. Source: https://twitter.com/DanaDanger/status/230851016344600576/photo/1/large

[8] Examine the differences between qualitative and quantitative data. Retrieved: November 10th, 2015. Source: http://regentsprep.org/regents/math/algebra/ad1/qualquant.htm

[9]. Sandelowski, M (2000). Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Method Studies. University of North Carolina.

[10]. Mixed Method Techniques Image. Retrieved: November 11th, 2015. Source: http://caribbean.scielo.org/img/revistas/wimj/v57n5/a06fig01.gif

[11]. Koffka, K (2005). Principles of Gestalt Psychology, The International Library of Psychology. C.K. Ogden. Originally published in 1935.

[12] Similarity & Proximity Image. Retrieved: December 4th, 2015. Source: http://music.miami.edu/programs/mue/Research/etrama/web/sepa_files/image037.gif

[13]. Continuity or Common Fate Image. Retrieved: December 4th, 2015. Source: http://www.teaching.louisabufardeci.net/111/files/weblog/designTips/worksheets/gestalt_common-fate.gif

[14]. Closure Image. Retrieved: December 4th, 2015. Source: https://s-media-cache-ak0.pinimg.com/originals/3e/95/32/3e9532cbd50d7272ed05759d5054f426.gif

[15]. Common Region Image. Retrieved: December 4th, 2015. Source: http://www.teaching.louisabufardeci.net/111/files/weblog/designTips/worksheets/gestalt_common-fate.gif

[16]. Stolte C., Mackinlay J. (2006). VizQL: A language for query, analysis and visualisation. Stanford EE Computer Systems Colloquium. May 3rd, 2006.

[17]. Parenteau J., Sallam R., Howson C., Tapadinhas J., Schlegel K., Oestreich T. Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner. Retrieved: March 23rd, 2016. Source: https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204&st=sb

[18]. All-Island Research Observatory. Maynooth University. Retrieved: March 11th, 2016. Source: http://airo.maynoothuniversity.ie/

[19]. Tableau Software. Retrieved: March. 11th, 2016. Source: https://public.tableau.com/s/

[20]. All-Island Research Observatory. Maynooth University. Retrieved: March 11th, 2016. Source: http://airo.maynoothuniversity.ie/external-content/all-island-atlas-0

[21]. Tableau Software. Retrieved: March 11th, 2016. Source: https://public.tableau.com/s/gallery

[22]. Tableau Software. Prevalence of Diabetes Mellitus in Adults (20-79 years), 2013. Retrieved: March 11th, 2016. Source: https://public.tableau.com/s/gallery/prevalence-diabetes-world

[23]. Central Statistics Office (CSO). Retrieved: March 11th, 2016. Source: http://www.cso.ie/en/index.html

[24]. Referendum Ireland. Retrieved: March 11th, 2016. Source: http://www.referendum.ie/index.php

[25]. Data.gov.ie. Ireland's Open Data Portal. Retrieved: March 11th, 2016. Source: https://data.gov.ie/data

[26]. QGIS (Quantum Geographic Information System). Retrieved: March 11th, 2016. Source: http://www.qgis.org/en/site/

[27]. OpenStreetMap. Retrieved: March 11th, 2016. Source: https://www.openstreetmap.org/#map=15/53.2533/-6.1177&layers=QN

[28]. International Food Policy Research Institute (IFPRI) (2014). Retrieved: January 19, 2016. Source: https://public.tableau.com/profile/ifpri.td7290#!/vizhome/2014GHI/2014GHI

[29]. Li Lou (2015). Gazprom's Grip: Final. Retrieved: January 19, 2016. Source: https://public.tableau.com/s/gallery/gazproms-grip-russias-leverage-over-europe

[30]. Colour Blind Awareness. Types of Colour Blindness. Retrieved: March 23rd, 2016. Source: http://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/

[31]. Tableau and Behold! Creating Custom Polygons on a Background Image. Retrieved: January 11th, 2015. Source: http://tableauandbehold.com/2015/04/13/creating-custom-polygons-on-a-background-image/

[32]. Makdad N. Creating Interactive Dashboards in Tableau. Retrieved: January 11th, 2015. Source: https://www.lynda.com/Tableau-tutorials/Why-Tableau/417094/442249-4.html

[33]. Adam Riley. Ireland Counties. Alteryx. Retrieved: March 14th 2016. Source: http://tableaumapping.bi/2013/11/25/ireland-counties/

[34]. Shrenik Ganatra. ADAM.CG PRO Font. Behance. Retrieved: March 23rd, 2016. Source: https://www.behance.net/gallery/13756975/ADAMCG-PRO-Free-Typeface

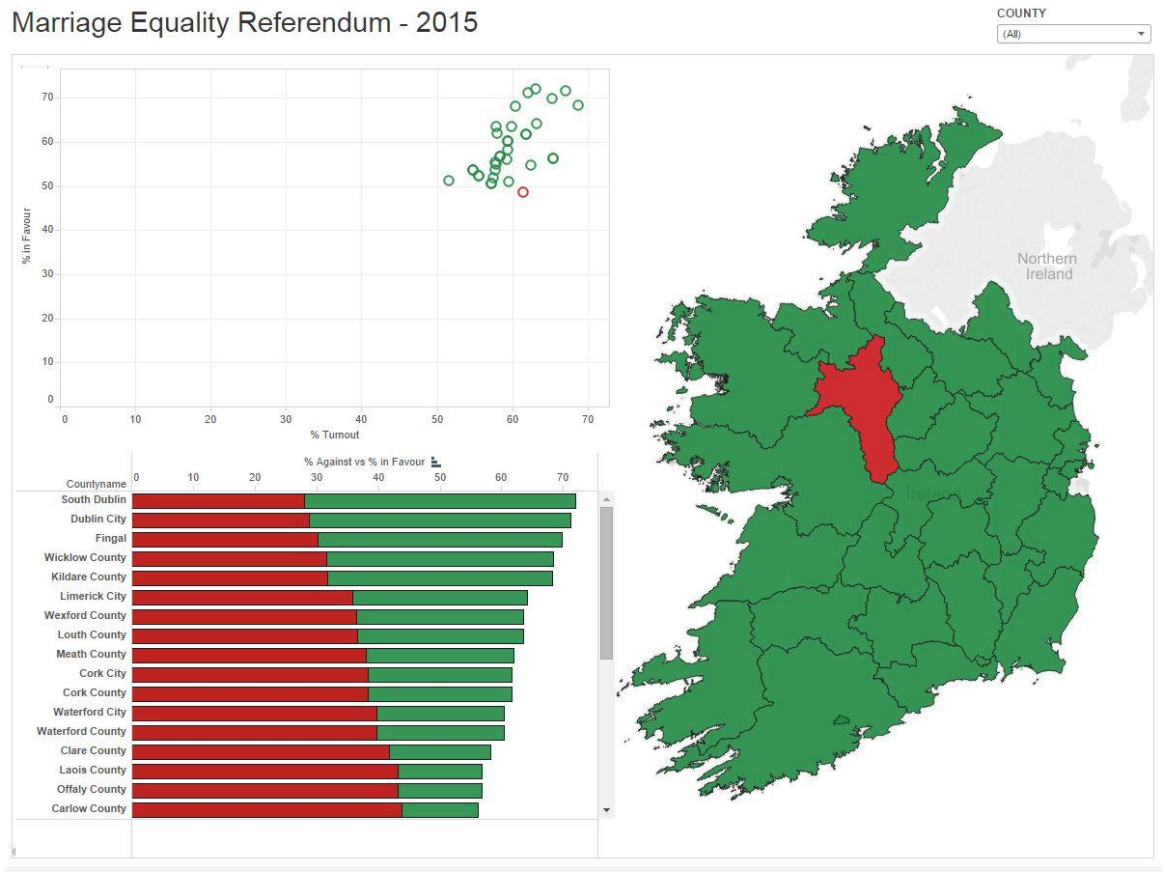
[35]. VizCandy. Finishing Touches: Fonts. Retrieved: March 23rd, 2016. Source: http://vizcandy.blogspot.ie/2013/12/finishing-touches-fonts.html

[36]. Quesenbery W., (2001). Society for Technical Communication. What Does Usability Mean: Looking Beyond 'Ease of Use'. Retrieved: March 24th, 2016. Source: http://www.wqusability.com/articles/more-than-ease-of-use.html

[37]. Mayhew D., (2013). Besticaux Community. Ease of Learning vs. Ease of Use – They Aren't the Same Thing. Retrieved: March 24th, 2016. Source: http://community.bestica.com/profiles/blogs/ease-of-learning-vs-ease-of-use-they-aren-t-the-same-thing

8 Appendices

8.1 APPENDICES A



8.2 APPENDICES B

What is your gender?

[ ] Male [ ] Female

What is your age?

[ ] 18 – 24 [ ] 25 – 34 [ ] 35 – 44 [ ] 45 – 54 [ ] 55+

(Answer the questions below based on how much you agree with the statements)

When viewing information, I would rather see the information as a visualisation rather written as numbers.

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

When viewing data visualisations, I often don't understand the data being presented.

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

If I find information that interests me, I would create a data visualisation to help myself better understand it.

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

8.3 APPENDICES C

I think that I would like to use this website frequently

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I found the website unnecessarily complex

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I thought the website was easy to use

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I think that I would need the support of a technical person to be able to use this website

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I found the various functions in this website well integrated

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I thought there was too much inconsistency in the website

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I would imagine that most people would learn to use this website very quickly

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I found the website very cumbersome to use

Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I felt very confident using this website

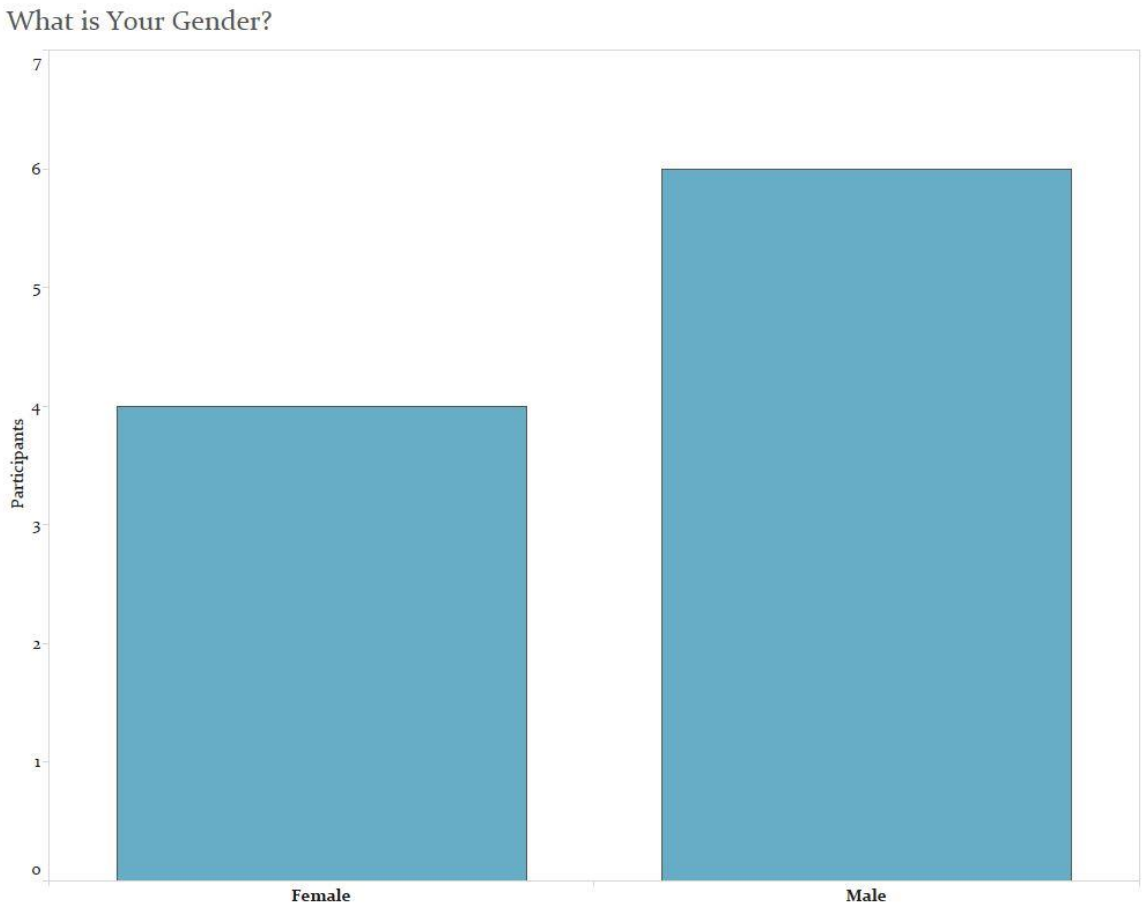
Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

I needed to learn a lot of things before I could get going with this website

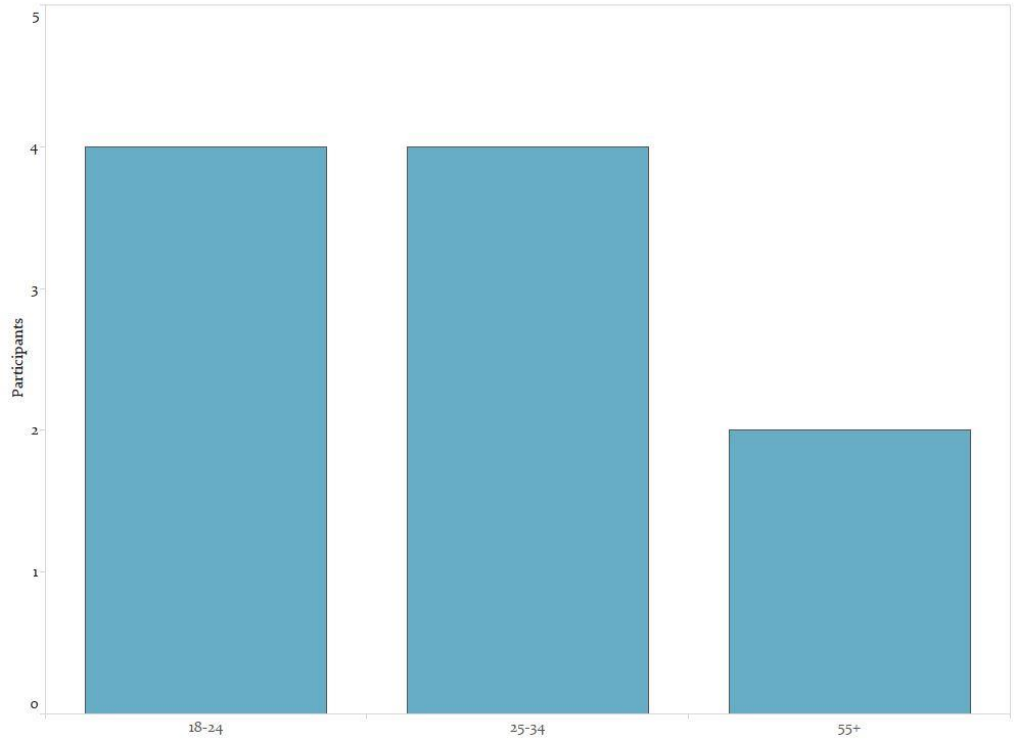
Strongly Agree	Agree	Neither	Disagree	Strongly Disagree

8.4 APPENDICES D

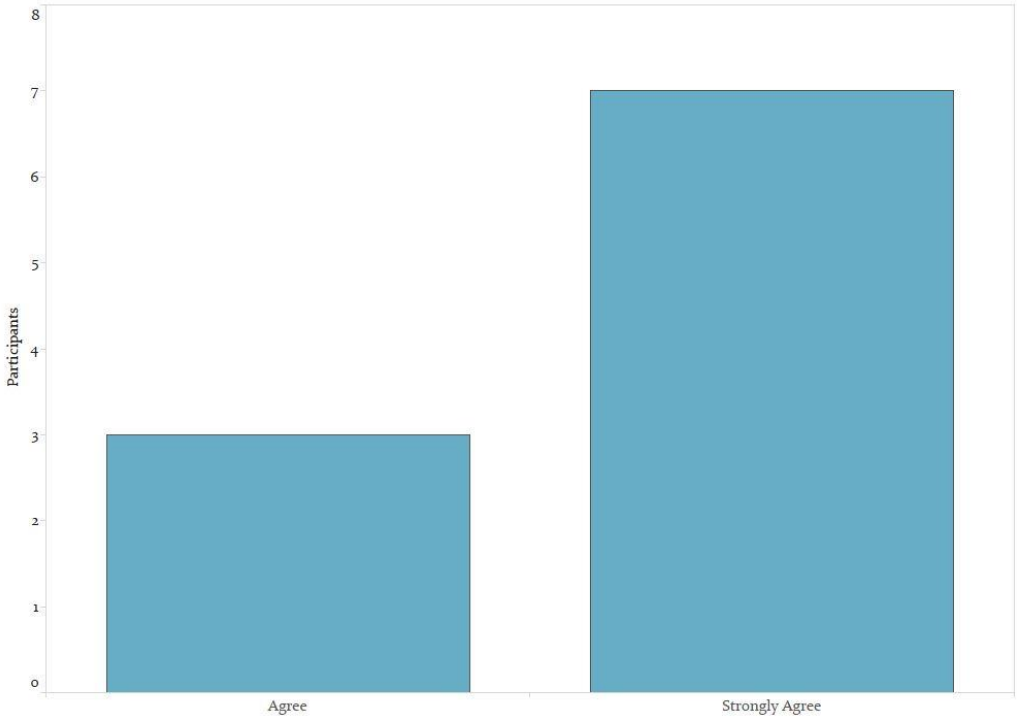
Participants	Gender	Age	Visualised vs Raw	Understand Visualisations?	Create Visualisations?
Participant #1	Male	18-24	Strongly Agree	Strongly Disagree	Disagree
Participant #2	Male	18-24	Strongly Agree	Strongly Disagree	Disagree
Participant #3	Female	18-24	Strongly Agree	Strongly Disagree	Disagree
Participant #4	Male	25-34	Strongly Agree	Strongly Disagree	Agree
Participant #5	Female	18-24	Agree	Agree	Disagree
Participant #6	Male	25-34	Strongly Agree	Strongly Disagree	Strongly Disagree
Participant #7	Female	25-34	Strongly Agree	Strongly Disagree	Strongly Disagree
Participant #8	Male	25-34	Strongly Agree	Agree	Strongly Disagree
Participant #9	Male	55+	Agree	Disagree	Strongly Disagree
Participant #10	Female	55+	Agree	Disagree	Strongly Disagree



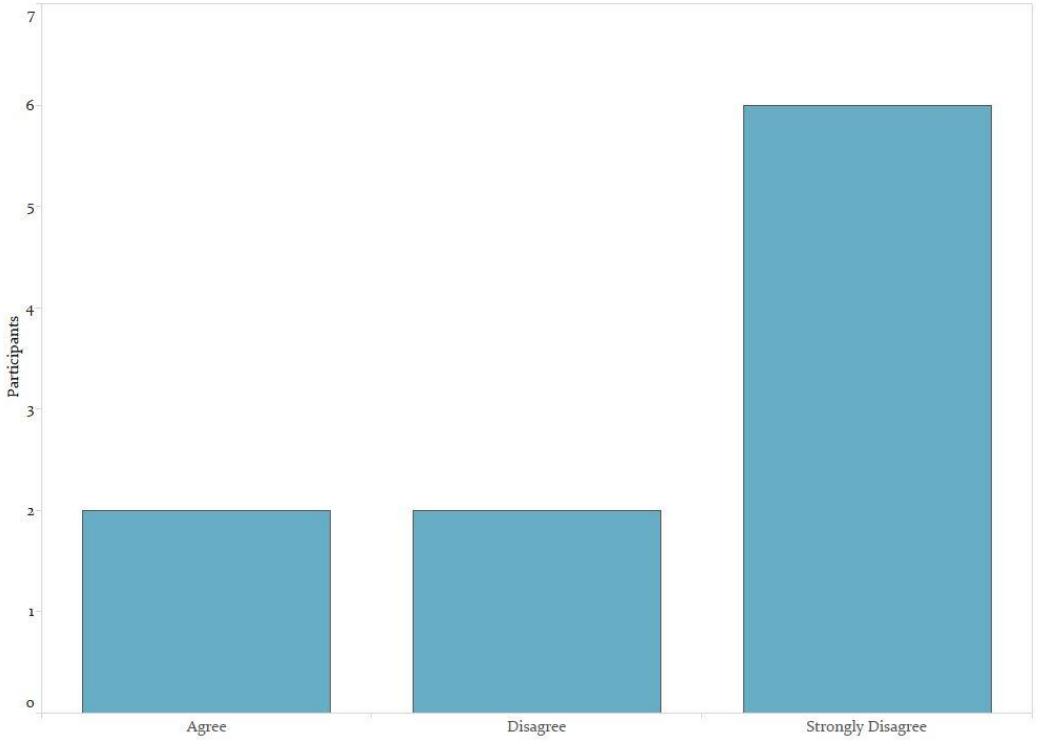
What is Your Age?



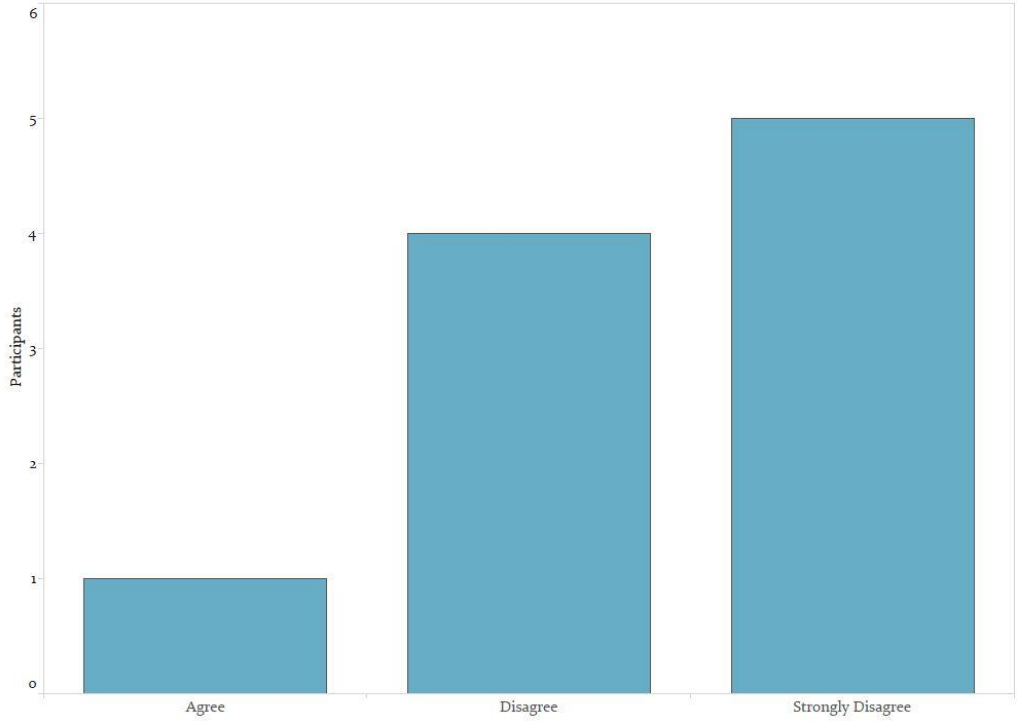
When viewing information, I would rather see the information as a visualisation rather written as numbers.



When viewing data visualisations, I often don't understand the data being presented.



If I find information that interests me, I would create a data visualisation to help myself better understand it.



8.5 APPENDICES E

Participants	Task 01 Time Visualised	Task 01 Time Raw	Results	Notes
Participant #1	5	12	PASS	...
Participant #2	4	11	PASS	...
Participant #3	5	14	PASS	...
Participant #4	6	15	PASS	...
Participant #5	4	13	PASS	...
Participant #6	3	13	PASS	...
Participant #7	5	15	PASS	...
Participant #8	3	14	PASS	...
Participant #9	8	17	PASS	...
Participant #10	6	15	PASS	...

Participants	Task 02 Time	Task 02 Errors	Results	Notes
Participant #1	01:13	0	PASS	...
Participant #2	01:22	0	PASS	Navigated to the wrong dashboard before finding the correct one
Participant #3	01:38	1	PASS	Encountered cursor error
Participant #4	01:26	0	PASS	...
Participant #5	01:36	1	PASS	Encountered cursor error
Participant #6	01:49	0	PASS	Refreshed the page from long load time
Participant #7	01:25	0	PASS	...
Participant #8	01:41	0	PASS	...
Participant #9	03:00		FAILED	Kept refreshing the page because of long load times. Was asked to stop after 3 mins
Participant #10	01:51	0	PASS	...

Participants	Task 03 Time	Task 03 Errors	Results	Notes
Participant #1	01:46	1	PASS	Encountered cursor error
Participant #2	02:07	1	PASS	Filtered the counties first, resets if dropdown is filtered last
Participant #3	01:33	0	PASS	...
Participant #4	01:42	1	PASS	Encountered cursor error
Participant #5	01:45	0	PASS	...
Participant #6	02:13	0	PASS	...
Participant #7	01:55	1	PASS	Encountered cursor error
Participant #8	02:15	1	PASS	Encountered cursor error
Participant #9	02:47	0	PASS	Needed to be told how CTRL+Click worked
Participant #10	02:21	1	PASS	Filtered the counties first, resets if dropdown is filtered last

Participants	I think that I would like	I found the website unnec	I thought the website was easy	I think that I would need the su	I found the various functions
Participant #1	Agree	Disagree	Agree	Strongly Disagree	Agree
Participant #2	Agree	Strongly Disagree	Strongly Agree	Strongly Disagree	Strongly Agree
Participant #3	Agree	Disagree	Strongly Agree	Disagree	Strongly Agree
Participant #4	Disagree	Disagree	Agree	Strongly Disagree	Agree
Participant #5	Disagree	Strongly Disagree	Strongly Agree	Disagree	Agree
Participant #6	Agree	Disagree	Agree	Disagree	Agree
Participant #7	Agree	Disagree	Strongly Agree	Strongly Disagree	Strongly Agree
Participant #8	Agree	Disagree	Agree	Disagree	Strongly Agree
Participant #9	Disagree	Agree	Disagree	Agree	Neither
Participant #10	Disagree	Disagree	Agree	Disagree	Agree

I thought there was too much inc	I would imagine that most	I found the website very cumb	I felt very confident usi	I needed to learn a lot of thi
Disagree	Agree	Disagree	Agree	Disagree
Strongly Disagree	Strongly Agree	Strongly Disagree	Strongly Agree	Strongly Disagree
Disagree	Agree	Disagree	Agree	Disagree
Strongly Disagree	Agree	Strongly Disagree	Agree	Disagree
Strongly Disagree	Agree	Strongly Disagree	Agree	Disagree
Disagree	Agree	Disagree	Agree	Disagree
Strongly Disagree	Strongly Agree	Disagree	Strongly Agree	Strongly Disagree
Disagree	Agree	Disagree	Agree	Disagree
Disagree	Agree	Agree	Disagree	Agree
Disagree	Strongly Agree	Disagree	Agree	Disagree

8.7 APPENDICES G

Participants	Task 01 Time Visualised	Task 01 Time Raw	Difference	Square of Difference
Participant #1	5	12	-7	49
Participant #2	4	11	-7	49
Participant #3	5	14	-9	81
Participant #4	6	15	-9	81
Participant #5	4	13	-9	81
Participant #6	3	13	-10	100
Participant #7	5	15	-10	100
Participant #8	3	14	-11	121
Participant #9	8	17	-9	81
Participant #10	6	15	-9	81
SUM			**-90**	**824**

- Sample size

N = 10

- Degrees of Freedom

Df = 9

- Sum of Difference

Σdi = -90

- Sum of Difference Squared

Σdi^2 = 824

- Mean of Difference

d = -90 / 10

d **= -9**

- Standard Deviation of Differences

S^2d = 824 – 10 (-9)^2 / 10 – 1

S^2d = 822.5 / 9

S^2d = 91.4

∴ sd = 9.6

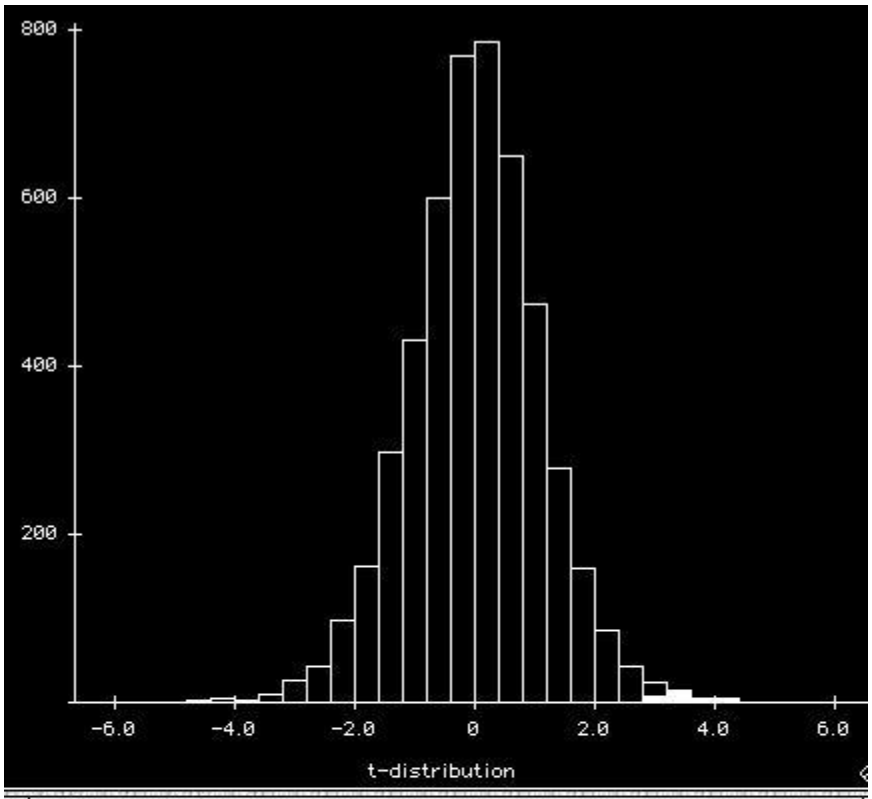
- Test Statistic

t = -9 / 9.6 / √10

t = -9 / 9.6 / 3.2

t = -9 / 3

t = -3



% Results ≥ than t

Frequency breakdown of **Test Result**
No Selector

Total Cases	5000
Number of Categories	2

Group	**%**
Percent of distribution greater than calculated t	0.800
Percent of distribution less than calculated t	99.200

Calculated t	Number of Simulat...	SampleSize
3	5000	10

8.8 APPENDICES H

8.9 APPENDICES I