

# DIGITAL DATA VIZ

## FINAL THESIS REPORT

---

I'm teaching myself how to build data visualisations online and I want to share my learning through live examples, code tutorials and useful resources.

Duncan Struthers  
ID 000748329-5

MA Web Design  
& Content Planning

[www.digitaldataviz.com](http://www.digitaldataviz.com)

# CONTENTS

## **P.3 #1 CONCEPT**

- 1.1 Elevator pitch
- 1.2 The problem
- 1.3 The value proposition
- 1.4 Unique selling proposition (USP)

## **P.7 #2 BUSINESS**

- 2.1 Business environment
- 2.2 SWOT analysis
- 2.3 Costs and value

## **P.11 #3 RESEARCH**

- 3.1 Cognate websites
- 3.2 Non-cognate websites
- 3.3 Direct competitor analysis
- 3.4 Research conclusions

## **P.18 #4 COMMODITY**

- 4.1 User centred vs. self-centred
- 4.2 Identifying a target audience
- 4.3 Target audience survey
- 4.4 User profiles
- 4.5 Content
- 4.6 User experience

## **P.26 #5 CULTURAL CONTEXT**

- 5.1 Background
- 5.2 The present
- 5.3 Privacy and big data

## **P.30 #6 FIRMNESS**

- 6.1 Requirements
- 6.2 Proposal

## **P.33 #7 DELIGHT**

- 7.1 Design research
- 7.2 My design work
- 7.3 Data visualisation principles
- 7.4 Brand values
- 7.5 Typography
- 7.6 Colour palette
- 7.7 Logo
- 7.8 Layout
- 7.9 Design conclusions

## **P.47 #8 IMPLEMENTATION**

- 8.1 First prototype
- 8.2 Schedule of works
- 8.3 Second prototype
- 8.4 Search engine optimisation (SEO)
- 8.5 Social media
- 8.6 Future marketing plans

## **P.53 #9 ANALYSIS**

- 9.1 Site performance
- 9.2 Accessibility
- 9.3 Browser testing
- 9.4 Feedback
- 9.5 Analytics

## **P.59 #10 CONCLUSION**

## **P.63 #11 REFERENCES**

# #1 CONCEPT

## 1.1 Elevator pitch

Data is an increasingly important factor in shaping our societies and the modern world. Data visualisation is a method of quickly analysing a dataset through visual representation. It is currently a popular tool for scientists, the media, business, government and activists, among many others and until recently it was primarily presented using static visuals. With the interactive potential of the modern web more and more tools are now being created online using web-based data.

I'm a data visualisation designer and web developer and I want to bridge the gap between my two disciplines by learning how to build online data visualisations. Because the subject area, technologies and communities surrounding them are still relatively in their infancy I have found many resources to build basic visualisations online, but few at an intermediate level. *Digital Data Viz* will be a way of documenting and sharing my learning and reducing the learning curve that exists beyond the basics of online data visualisation.

**90%**  
**OF ALL**  
**DATA IN**  
**THE WORLD**  
**HAS BEEN**  
**GENERATED**  
**OVER THE LAST**  
**TWO YEARS<sup>1</sup>**

---

SINTEF, May 2013

## 1.2 The problem

Data visualisation is a method of representing complex information with the aim of clearly and efficiently communicating the narrative behind the data.

Data visualisation has become increasingly popular in society and business since the beginning of the digital age in the late 1970s. This trend has accelerated dramatically with the rise of the internet in the 1990s and its increasing influence on daily life.

The popularity of static charting, graphs and infographics in print mediums and desktop applications is now being overtaken by the potential of data driven web visualisations. Data visualisation is, by its very nature, a perfect medium for the interactivity the web can offer. Data itself is digital and the primary source of data creation in modern society is the web.

There are now a growing number of web-based frameworks to create data visualisations online and JavaScript (JS) is the primary language in which these libraries are based. There are primarily two categories of these frameworks. The first consists of proprietary frameworks with the intention of making data visualisations online easier to create through pre-built templates and functionalities. An example of this first category would be Google Charts API, which makes a large amount of design and interaction decisions for the users but prevents bespoke customisations beyond the options it provides.

The second category consists of open source libraries that give you the tools to bind your data to Document Object Model (DOM) elements, in essence the functional building blocks. This allows you to create a bespoke visualisation and gives you a great deal of autonomy and flexibility in both design and development. Many of the proprietary frameworks in the first category are built using these libraries.

The leading example of a library within the second category is d3.js (D3). The problem with using a tool like D3 lies in the lack of resources available online to learn its intermediate level techniques that create the CSS3 transitions and Scalable Vector Graphic (SVG) interactions that put data driven web visualisations at the forefront of information design. Because online data visualisation and D3 itself is still relatively in its infancy there are freely available tutorials for a user to learn the basics of D3 online but a steep learning curve is widely acknowledged beyond this point.

## 1.3 The value proposition

The value of this website will be in providing freely available tutorials, resources and live examples in web-based data visualisation that goes beyond the basics.

My initial focus will be on D3, as I believe it is the leading open source visualisation library with the most future potential and it has a strong emphasis on web standards. I also plan to provide insights into how to use proprietary frameworks, focusing in on customisations that go beyond the basic offering.

## 1.4 Unique selling proposition (USP)

Through my initial experiences of learning how to build online data visualisations in D3, I found there were few resources available to learn past the basics and that the growth in proprietary frameworks which offer template options was largely due to the steep learning curve associated with libraries such as D3. With the present popularity of bespoke infographics in a modern world increasingly shaped and influenced by data analysis, I believe that the trend towards online data visualisations will increase and that the ability to control the design and functionality to fit the subject will rise in importance.

I have therefore chosen to focus my website and learning primarily on flexible open source data visualisation libraries, beginning with D3. I believe that by sharing my learning through collecting, curating and recommending useful resources and examples I will assist those wishing to create online data visualisations and reduce the steep learning curve in understanding analysis tools.

I also believe that in providing a similar service to users of proprietary tools such as Google Charts API will help to allow users with less time or knowledge of web development to learn customisations that go beyond the basic offering.

## #2 BUSINESS

### 2.1 Business environment

Edward Tufte is an American statistician and professor of political science, statistics, and computer science at Yale University. He is noted for his writings on information design and as a pioneer in the field of data visualisation. His first book on the subject, *Envisioning Information*,<sup>2</sup> was published in 1990 and since then data visualisation's popularity began a slow but steady rise towards mainstream visual culture. It wasn't until 2010 with the huge success of David McCandless's book, *Information is Beautiful*,<sup>3</sup> that the popularity of data visualisation reached a peak. This coincided with the rise to prominence of Twitter, and infographics soon became a perfect medium to visualise the narratives in topical data.

At the same time that static infographics were rising in prominence, data visualisation on the web was also progressing. In 2007 Google Charts was officially launched. Originally it was an internal tool using PNG files to support rapid embedding of charts within Google's own applications, like Google Finance.

Later in 2011 Mike Bostock, now Graphics Editor of The New York Times, then studying a PhD at Stanford University published *D3: Data-Driven Documents*.<sup>4</sup> D3 is a JS library for manipulating documents and primarily SVG elements based on data. D3 was such an important development because it allowed for interactivity, had an emphasis on web standards and had a data-driven approach to DOM manipulation. The release of D3 accelerated interactive web based data visualisation and in April 2012 Google followed suit and updated Google Charts API to use SVG. Much as static infographics were propelled into the mainstream by David McCandless, interactive online data visualisation is now coming to the fore and this was reflected by McCandless updating his website,

[www.informationisbeautiful.net](http://www.informationisbeautiful.net),<sup>5</sup> in January 2013 by beginning to introduce interactive infographics.

There are now many different JS libraries, proprietary frameworks and software that can produce interactive online data visualisations. However, this is an emerging field and there is a great variation in quality and ease of use between them.

## 2.2 SWOT analysis

### Threats

The main threat to my site is my competitors within the market. These fall into three categories. Firstly, popular gallery sites that began based on static infographics and have recently added interactive sections such as [www.visual.ly](http://www.visual.ly).<sup>6</sup> However, these sites focus on the visual output rather than the resource that built them. Because they are galleries that simply present attractive visuals rather than offer information on how the infographics are built, they are not direct competition for users who are looking to build themselves.

Secondly, there are sites that are in closer competition with my proposal. Andy Kirk's visualising data site, [www.visualisingdata.com](http://www.visualisingdata.com),<sup>7</sup> makes recommendations on which tools to use and these are lent weight by his reputation as a data visualisation expert. The same is true for Interactive Things,<sup>8</sup> a design and technology studio based in Zürich, who specialise in producing interactive data visualisation and publicise their favourite tools at [www.selection.datavisualization.ch](http://www.selection.datavisualization.ch).<sup>9</sup>

Thirdly, there are sites that are covering the same topics as my own. Scott Murray's [www.alignedleft.com](http://www.alignedleft.com)<sup>10</sup> has introductory tutorials in D3 and links to his ebook on the subject, *Interactive Data Visualization for the Web*,<sup>11</sup> which is available to read online. Sebastian Gutierrez's [www.dashingd3js.com](http://www.dashingd3js.com)<sup>12</sup> covers a basic introduction to D3 and offers paid for tutorials in intermediate



and advanced D3. Malcolm Maclean's [www.d3noob.org](http://www.d3noob.org)<sup>13</sup> is again a site for those new to D3, as is his ebook on the subject, *D3 Tips and Tricks, Interactive Data Visualization in a Web Browser*.<sup>14</sup>

## **Strengths**

The strengths of my proposed site become clear when considering the threats and competition already available on the web. As discussed above there are already websites that provide expert information about producing interactive data visualisation on the web. However these are still not sites dedicated to providing tutorials and resources for the tools they use.

There are however websites providing tutorials in the basics of D3. These are excellent resources and actually provide good places to direct users who need to hone their understanding of the core concepts before looking at my more intermediate tutorials. Dashing D3 does provide intermediate and advanced tutorials, but as these are paid for services my website should be a useful resource for those who cannot afford the fees.

## **Weakness**

There are two main weaknesses to my proposal. Firstly, I am not a JS or programming expert and many of the tools I will be using rely on this heavily. However, these are skills I am keen to learn as this is a field I am very interested in, so I'm looking forward to the challenge. The second problem is time. I first want to learn advanced JS, then D3 as well as Google Charts API, and I'm keen to design my site around my content. My response to this was to consider the scope of the project and be realistic about what I could achieve within the thesis timescale and what would be goals for the future. After initially planning to look at other libraries beyond D3 and Google Charts API, including Raphael.js and Processing.js I decided to focus my learning, and only look into other libraries if time permitted.

## Opportunities

As I work as a Designer and Front-end Developer I want to tie my learning into my professional work to benefit both my studies and my job.

I have discussed my thesis project with my employers, Ipsos Mori, who are a market research company. They are currently looking to expand their online interactive data visualisation offering to clients and are very interested in the growth of this area. There may be an opportunity for funding and allocated time at work for projects that benefit the business and I believe my project would fulfil this criteria.

There is also the opportunity to use some of the research data that my company produces in the data visualisation examples on my site. Before pursuing this I will have to consider how much control of the project I am willing to surrender and whether our objectives would align.

## 2.3 Costs and value

The project will not be aimed at revenue generation.

The alternative value of the project will be as a portfolio site to promote my work both inside and outside of my company and to raise my profile within the data visualisation sector.

The only cost of the project will be hosting and my own time. If I can get funding, then I will consider using paid for tutorials to increase the speed of my learning.

## #3 RESEARCH

### 3.1 Cognate websites

<http://visual.ly/>

<b>Global traffic rank</b>	2,500
<b>Google page rank</b>	7/10
<b>Page load</b>	Slow (3.018s) 78% of sites are faster

www.visual.ly<sup>6</sup> is the most popular infographic site on the web. They promote themselves as 'the world's marketplace for visual content' and offer companies the equivalent of a design agency service drawing from their large infographic community.

They display a gallery of their latest uploads on the homepage under the categories: infographic, video and interactive. Visually have huge investment behind them and in January 2014 announced the end of their latest financing round, totalling \$8.1 million.

Although they are the largest cognate website, I do not see them as direct competitors for traffic. While my target audience may look at their galleries for inspiration and ideas and note the listed technologies used to build a particular example, my site primarily offers knowledge and advice on how to create rather than simply showing the final visual examples.

<http://www.visualizing.org/>

<b>Global traffic rank</b>	66,000
<b>Google page rank</b>	7/10
<b>Page load</b>	Very Slow (3.78s) 86% of sites are faster

[www.visualizing.org](http://www.visualizing.org/)<sup>15</sup> is an infographic community website where users can share their work, find data sources, enter challenges and visit events. Visualizing was created by GE and Seed Media Group to help make data visualisation more accessible to the general public; to promote information literacy through the creation, sharing, and discussion of data visualisations; and to provide a unique resource to help simplify complex issues through design. The site's content is a mixture of interactive sites and static infographics.

I believe that Visualizing differs from my concept in that it is very much a site focused on the community aspect of infographics. Their organised challenges and data marathon meetups help to create a social exchange focus rather than a skills based focus. They are also a source of data and data channels for creating infographics as well as a gallery or inspiration site so again they are not direct competition but instead provide a great resource to direct my users to.

## 3.2 Non-cognate websites

<http://www.moneysupermarket.com>

<b>Global traffic rank</b>	6,800
<b>Google page rank:</b>	6/10
<b>Page load</b>	Average (1.51s) 56% of sites are slower

Money Supermarket<sup>16</sup> is a good example of a non-cognate website because it is based on communicating and visualising data quickly to users. Money Supermarket is a commercial site based on generating income through referrals when users register for an account or policy through moneysupermarket.com. Starting with the homepage, immediately the navigation is broken down into the separate areas they compare. Once you drill down to a comparison page it is clear they are not keen on visualising the comparisons and instead they simply list the features and rely on a filtering system at the top for users to interact with the data presented. This use of a filtering system allows users to select the parameters that are important to them first, before seeing the comparison.

### 3.3 Direct competitor analysis

<http://www.visualisingdata.com/>

<b>Global traffic rank</b>	414,700
<b>Google page rank</b>	6/10
<b>Page load</b>	Very Slow (7.01s) 80% of sites are faster

*(Please note at the time of publication the site is in the process of being redesigned.)*

Andy Kirk is a UK-based freelance data visualisation design consultant, trainer, author, speaker and editor of [www.visualisingdata.com](http://www.visualisingdata.com)<sup>7</sup>. He is the founder of Visualising Data Ltd undertaking design consultancy and delivering training. Visualising Data is his personal site that has details about his books, training courses, lectures and, importantly for my site, a data visualisation resources section.

In terms of popularity Andy Kirk is a prominent figure in data visualisation through his lecturing, publishing and training courses and this lends weight to his recommendations when it comes to interactive data visualisation tools.

The business model for the site is to promote himself and his business by creating a list of recommended resources for data visualisation. He intends to increase traffic to his website and thereby promote his books, courses and lectures.

I can learn a great deal from Andy's approach as I also intend to use the website to raise my profile within the infographic community. I feel it will be important to have a profile section on the site for myself and to ensure the site is well connected to my social networks and other online profiles.

Andy himself states in the resources section that his list is a grand overview of all the technologies available and that “some are great, some are less great... people can try them out for themselves and judge to see what works for them.”<sup>7</sup>

Therefore, one way I can differentiate my website is through my focus on tools that I recommend and use myself. I can also improve the visual aspect of the presentation by implementing a clearer, cleaner design and practicing what I preach by presenting my recommendations using data visualisation principles.

<http://selection.datavisualization.ch/>

<b>Global traffic rank</b>	210,350
<b>Google page rank</b>	5/10
<b>Page load</b>	Fast (2.12s)
	68% of sites are slower

Interactive Things<sup>8</sup> are a user experience and data visualisation studio based in Zürich, Switzerland.

The [www.selection.datavisualization.ch](http://www.selection.datavisualization.ch)<sup>9</sup> website is a set of preferred libraries, applications and services that they use regularly in their interactive data visualisation work. It includes libraries for plotting data on maps, frameworks for creating charts, graphs and diagrams, and tools to simplify the handling of data. These tools sit within their site [www.datavisualization.ch](http://www.datavisualization.ch)<sup>9</sup> which is a news and knowledge resource for data visualisation and infographics.

Interactive Things are highly successful within the sphere of data visualisation as they fulfil a rare niche of being specialised in producing digital interactive data visualisations. This again lends weight to their recommendation of tools based on their reputation within the field.

The business model behind creating their selected tools and the site in which it sits is again to promote their studio and increase traffic to their site. There are many blog posts that list the top 20 data visualisation tools. However, here they are producing a curated collection that is far more valuable to the user and therefore a greater asset to a business promoting the quality of their expertise.

I can learn a great deal from this site because it is very close to my original concept of reviewing and comparing the different available tools. However, I believe my time is better spent focused on a small range of flexible and powerful tools beginning with D3, so that I can provide an informed and valuable opinion to the user.

<http://alignedleft.com/tutorials/d3/>

<b>Global traffic rank</b>	242,500
<b>Google page rank</b>	5/10
<b>Page load</b>	Very Fast (1.21s) 85% of sites are slower

Scott Murray<sup>10</sup> is a code artist who writes software that generates visual and interactive experiences. He is an assistant professor of design at the University of San Francisco, and the author of Interactive Data Visualization for the Web.<sup>11</sup>

In the process of learning how to use D3, Murray put together tutorials on how to use it, which he hosts on his personal website. He is well known within the professional world of data visualisation and has created a profile for himself through regular conference talks and through publishing his book on D3. The purpose of publishing tutorials on how to use D3 is likely to be to attract traffic to his site from those interested in learning about D3 and therefore interested in buying his book on the subject. It also increases his profile in the field as an expert in the technology.



The first thing I have learnt about Murray's tutorials is how much content needs to be produced just to teach users how to produce a simple bar chart and scatter plot. This reaffirms my intent to create intermediate tutorials and Murray's site can be an excellent resource to suggest to beginners in my resource section. I also think it would be best if I make the starting point of my tutorials an existing basic example like the ones created by Mike Bostock so users can see how it was created from scratch rather than just the intermediate adaptations.

The second important point I took from Murray's tutorials was the importance of tone of voice. Murray's writing style is similar to other popular web authors in that it is light hearted and jovial. This makes sense because code as an entity can be quite a dry subject and a playful tone adds interest for user and engagement with the website. It also injects personality into the website and as this will be a site to showcase my work I want the text content to reflect my own character.

### 3.4 Research conclusions

It was from analysing Murray's and Interactive Thing's websites that I realised my initial concept to learn, compare and advise users on a large range of tools would leave me little time to learn them all and would result in a low quality service to the user. I therefore plan to reduce the scope of this initial phase and focus my learning.

Through my competitor analysis I have come to appreciate the clarity, focus and simplicity in the production and concept of other websites and this is something I feel is important to consider moving forward.

## #4 COMMODITY

### 4.1 User centred vs. self-centred

The main reason for choosing digital data visualisation as the topic for my thesis is because it is an area of interest for me as well as becoming an area in which I want to specialise professionally.

My aim is to increase my knowledge of the area, learn the best tools and how to use them. It is my intention to use this knowledge to produce a site that documents this process and helps people learn and build upon what I have learned myself. By creating this resource I plan to increase my reputation and profile within the data visualisation community and in my profession.

My site's content strategy will therefore not be solely based on a user centred approach but a balance between that and my own learning objectives.

I have developed some parameters based on what I want to learn and this will help me in identifying and targeting a specific audience. Firstly, as stated previously, I am keen to learn D3 and I want to focus my learning and therefore my content on intermediate tutorials, so my primary users need experience of basic HTML, CSS and JS.

## 4.2 Identifying a target audience

Research into prospective target audiences is difficult as online data visualisation is an area that has only appeared relatively recently. One research method is to find out about my competitors' users, and this can be done through Twitter and [www.alexacom.com](http://www.alexacom.com),<sup>17</sup> a web based analytics tool.

As discussed in section 3.3, Andy Kirk is a data visualisation specialist who runs [www.visualisingdata.com](http://www.visualisingdata.com).<sup>7</sup>

### Who visits [visualisingdata.com](http://visualisingdata.com)?

According to [alexacom.com](http://alexacom.com) in comparison to the general internet population there is:

- A slightly higher amount of male visitors
- A greatly over represented amount of graduate school visitors
- A greatly over represented amount of work based visitors

### Twitter followers: 12,500

In a random sample of 50 followers 26% were Data artists, Developers, Data Visualisers and Digital Designers.

This is by no means a robust test, however it has provided me with a clearer picture of my typical target audience who are interested in learning about online data visualisation and fit my learning parameters.

## 4.3 Target audience survey

To find out more about this group's demographics, experience, needs and suggestions I designed an online survey on Survey Monkey in February 2014.

<https://www.surveymonkey.com/s/XLNSKBW>

I publicised this survey within the data visualisation community through email, Facebook and Twitter and managed to reach over 24,000 potential users in particular via retweets from Andy Kirk and Alberto Cairo. So far 32 people have taken the survey and below are the results.

- **75%** of users were Male
- **70%** were between 21 and 39

### **35% had professions within my hypothesised target audience**

Web developer / Infographics editor / Interaction designer /  
Web developer / Data visualiser / Designer / Mobile developer /  
Front-end engineer

### **The rest fell into three groups**

1. Journalists and Scientists
2. Professors, Academics and students
3. Miscellaneous

- **59%** had experience of designing static infographics
- **66%** had quite a bit or a lot of experience of **HTML**
- **55%** had quite a bit or a lot of experience of **CSS**
- **38%** had quite a bit or a lot of experience of **JS**
- **93%** had interest in creating online interactive data visualisations

## Recommendations for content

- **24%** thought there should be live examples
- **24%** thought there should be tutorials accompanying the examples
- **14%** suggested there should be a news or blog section featuring tips and trends
- **14%** wanted data viz tools that could be exported and embedded in sites, excel, pdfs, etc
- **Only 7%** wanted examples that removed the need for coding expertise

## Other individual useful recommendations included

- Links to further information
- Intuitive to use and colourful
- Forum for questions
- Open data sources
- Alternative visualisation methods to traditional ones
- Background information on statistical theory

Of the online data visualisation tools they have experience in the most used were **56% D3** and **48% Google charts API**.

## Quotes from participants

"There are a few free online data viz tools for non-tech people like myself to use (tab.ly, muraly etc) but they are all a bit clunky and look rubbish. I want something simple to use with custom colour ways and fonts and logos I can import, nothing that's too rubbish looking, challenging or time consuming!"

"Introductions to new tools are interesting, as are showcases with examples on specific techniques, but theory is just as important. Tools are easier to find, theory (e.g. Tufte) is not so much."

## Survey conclusions

My target audience are mostly male, between 21 and 39, working in either the web, design or data visualisation. They may have experience in static infographics, have some experience in HTML and CSS and they are not experts in JS. The most popular libraries are D3 and Google Charts API and they want tutorials and coded examples.

## 4.4 User profiles

Because the website will be documenting tools and techniques I believe are necessary to learn about data visualisation, I am the first example of the target market for this kind of site.

### PERSONA 1

---

<b>User profile</b>	Duncan Struthers
<b>Profession</b>	Digital Designer, specialising in data visualisation
<b>Age</b>	30
<b>Coding level</b>	Front-end beginner
<b>DDV experience</b>	Beginner

### Requirements

- To create the best available visualisation for my web projects
- Learn how to create interactive examples
- Find useful resources and examples

## PERSONA 2

---

<b>User persona</b>	Matthew Gibbons
<b>Profession</b>	Web Developer for digital agency
<b>Age</b>	26
<b>Coding level</b>	Advanced
<b>DDV Experience</b>	Intermediate, has used Google Charts API and D3 before

### Requirements

- Has an upcoming work web project that needs to include data visualisation
- Needs tutorials for creating charts in D3
- Looking for existing coded examples to show clients in pitches
- Would like links to further quality resources

## PERSONA 3

---

<b>User persona</b>	Dave Beasant
<b>Profession</b>	Interaction Designer for the Guardian
<b>Age</b>	35
<b>Coding level</b>	Intermediate
<b>DDV Experience</b>	Beginner to intermediate, was a Flash expert and now specialises in Adobe Muse and Edge

### Requirements

- Wants more knowledge of how to create charts from hand coding rather than Adobe packages
- Needs to see visuals to compare the tools and test the interactive behaviours
- Wants detailed step by step tutorials to follow
- Would like links to further quality resources

### Unifying persona requirements

Prospective users already have their data set and the narrative they want to communicate. What they are after are examples that can be adapted and either instructions or resources to create their work.

## 4.5 Content

Because my learning will form the basis of my content, I need to pick a technique and tutorial topic that is widely needed, but also one that currently has insufficient intermediate level tutorials already online.

Based on my user research my content should fall into two sections.

**Section 1.** I should display live coded examples, both my own and other examples I find online to increase my content and the user experience.

**Section 2.** I should include my source code, tutorials and links to further information.

### **Source of data**

I will use topical data to improve search engine optimisation (SEO) and exposure from open data sources such as the Guardian Datablog.



## 4.6 User experience

### Practice what you preach

I plan to create a user experience adhering to the principles of data visualisation: clear, intuitive to use, organised and communicates information effectively. It will be an interactive site reflecting the interactive content.

### Responsive design

Although some of the tools and charts on the site may be responsive, this is an area which data visualisation has yet to conquer on the whole, so the majority of charts are likely to be fixed width at present. However I believe this can be a growth area and a particular niche for myself with my experience in creating responsive websites. So in an effort to future proof the project I feel it is important that the website is built responsively. From my research the majority of my users will be accessing the site from work or in a professional capacity and therefore are less likely to be viewing the site on their phone, so whilst I will follow a mobile first development approach, my design approach will be desktop first.

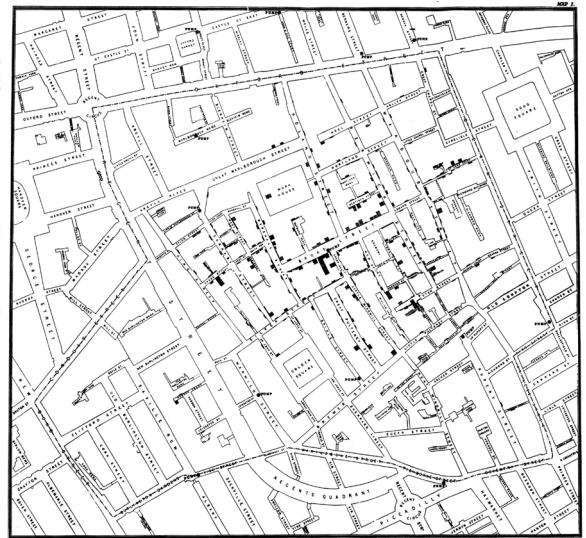
# #5 CULTURAL CONTEXT

## 5.1 Background

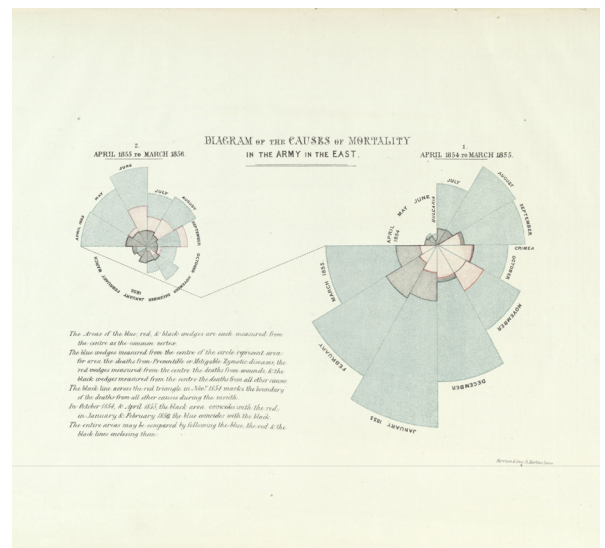
In 1854 a severe outbreak of cholera in London's Soho killed 500 people. At the time of the outbreak the disease was believed to be spread by miasma in the air. However, Dr John Snow decided to map the locations of known cases to find the source.<sup>18</sup>

On the map each death is marked as a bar and it became apparent that the cases were clustered around the pump in Broad Street, which on further inspection was found to be polluted by sewage from a nearby cesspit.<sup>18</sup>

In the same year as Snow's cholera outbreak, war broke out in Crimea. Florence Nightingale was a nurse to the wounded and in 1859 famously created polar area diagrams depicting causes of death among the soldiers.<sup>19</sup> The diagrams showed that preventable diseases dwarfed battle wounds in mortality, and according to John Maindonald in the Journal of Statistics Education, Nightingale "made strong, consistent and carefully argued cases for enlightened and data-based public decision-making."<sup>19</sup>



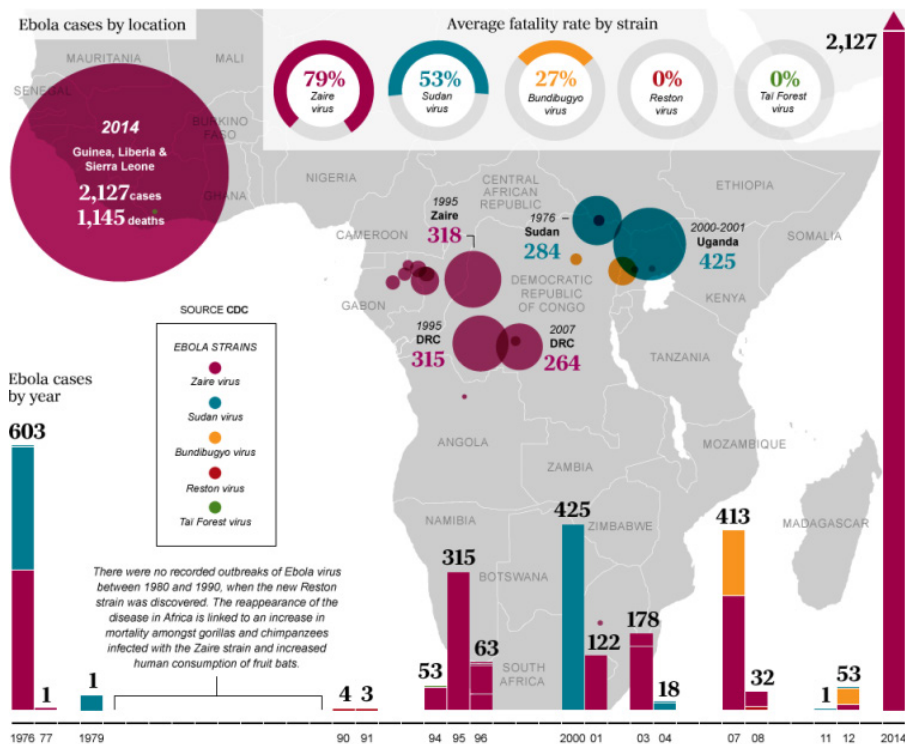
Original map made by John Snow in 1854. Cholera cases are highlighted in black.  
SOURCE: <http://commons.wikimedia.org/wiki/File:Snow-cholera-map-1.jpg>



Florence Nightingale's contribution to the sanitary history of the British army during the late war with Russia (London, 1859).  
SOURCE: <http://nyamcenterforhistory.org/2013/10/23/a-passionate-statistician-florence-nightingale-and-the-numbers-game/>

## 5.2 The present

Today, public health bodies rely on visualisations like these to analyse disease. The World Health Organization and other organisations use the data they collect themselves as well as making it open and available to everyone online. Through these open data sets journalists and the general public create their own visualisations, which help to publicise the problem. The infographic below, for instance, featured in the Daily Telegraph newspaper and was created by using open source historical data to map the history of Ebola in West Africa.



The Telegraph newspaper on Ebola's history  
SOURCE: <http://simonrogers.net/2014/09/23/ebola-in-charts-data-journalism-and-the-outbreak/>

Disease mapping is now a growth area for digital innovation, with doctors able to mark patients on outbreak maps with their smartphones,<sup>20</sup> and social networks being data-mined to show how disease spreads.<sup>21</sup>

With the viral impact of social media, organisations have the power to connect with billions around the world without expensive advertising. 2014 saw an awareness campaign for amyotrophic lateral sclerosis (ALS) go viral through the use of the ice bucket challenge, primarily on Facebook. The scale of the campaign's success is evident with hits to the English Wikipedia article on ALS growing from an average of 163,300 views per month to 2.89 million views in August 2014.<sup>22</sup> Within weeks of the challenge going viral, The New York Times reported that the ALS Association had received \$41.8 million in donations from July 29 until August 21 2014.<sup>22</sup>

For charitable organisations to connect with potential supporters, the use of narrative advertising has been popular since television commercials in the 1980s. Data visualisation's growth from print to digital covers multiple mediums and the use of narrative when visualising data is strongest when presented through animation. The present culture of video sharing through social media websites such as Youtube and Vimeo means charities can educate and motivate a wide audience through visual storytelling, making a potentially more powerful and lasting impact. The animated video "Water Changes Everything,"<sup>23</sup> created by the organisation Charity: Water, is a good example of this trend and its impact, with over a million views on Youtube.

## 5.3 Privacy and big data

Data privacy is now one of the foremost topical social issues around the world. Questions over the rights of personal data have been pushed into the public eye through the acts of Edward Snowden, who provided the Guardian newspaper with top-secret NSA documents leading to revelations about US surveillance on phone and internet communications, as well as through the work of the website WikiLeaks.<sup>24</sup>

Alongside the issue of data privacy is the topic of big data. In May 2013, SINTEF, the largest independent research organisation in Scandinavia released findings to show that 90% of all the data in the world had been generated over the last two years.<sup>1</sup> With the current trend for increasing the amount and complexity of online data, there is a growing need for digital data visualisation techniques to make sense of it and help to shape our societies for the better.

# #6 FIRMNESS

## 6.1 Requirements

### **Prioritising JavaScript**

I will be investing a large amount of my time in learning JS. The priority for me is to learn advanced JS before tackling how D3 as a JS library works so I can understand and explain the processes behind D3 enough to write tutorials about them. Therefore I do not want to create an unmanageable workload for myself by choosing numerous other technologies or frameworks that require significant additional time investment to learn and execute in producing the website itself.

### **Consistent structure**

The structure of my content will need to be consistent throughout the project, from the homepage to the primary and secondary content pages as I want the site's design to be based around the data visualisation principles of clarity and visual efficiency.

### **Sole contributor**

I am the sole web designer for the project. I will be making any updates and additions. There are no other contributors to edit and administrate or user accounts to generate content.

### **Responsiveness**

As previously stated in the Commodity section, I believe that although many of the current examples of online data visualisation are not natively responsive this is an area of growth that should be expected. I also have experience of responsive web design (RWD) that I'm keen to develop further in a data visualisation project in the future.

### **Practical and applicable**

I want my learning on the thesis to be practically applicable to my current work as a Digital Designer in a market research company. I am responsible for creating data visualisation micro-sites that present data visually from a variety of research reports. This means not just creating the charts and visuals but making concise succinct interactive web apps.

Below is an example of online data visualisation my company has produced in the past.

[http://www.ipsos-mori.com/\\_assets/topcity/](http://www.ipsos-mori.com/_assets/topcity/)

Below is an example of the more interactive web apps that you can use to interrogate the data that I aim to produce in the future.

<http://www.pop.pt/en/>

## **6.2 Proposal**

### **Markup**

I have chosen to use HTML5 as the markup language for this project as this will benefit my future progression in web development. It also means greater semantics in the website's code through the new structural elements and I intend to combine this with the use of WAI-ARIA roles to create a site built for optimal accessibility.

### **Presentation**

CSS3 will be used for the styling of the website with the option to include animations and transitions to enhance the interaction with the user.

### **Chart libraries**

After conducting my user research survey in my Commodity section I found the most popular libraries were D3 and Google

Charts API. I plan to focus my content primarily on D3 to begin with and move on to Google Charts API once I'm happy with my progress.

### **Data and code sharing**

As part of my tutorials I will need my data and code files to be available as an open source resource. I also want to share my work and receive comments, feedback and suggestions from others in a similar position. I have chosen to use Github for this purpose and specifically Gist, because Mike Bostock's [www.bloomberg.com](http://www.bloomberg.com)<sup>25</sup> website collates and collects D3 examples hosted there to share with the community.

### **CMS**

I have chosen to use Perch as my CMS because of its ease to implement and simplicity to integrate. This is an important consideration as I want to dedicate as much time to learning JS and D3 as possible. Because Perch can be implemented after the site is designed and developed in HTML, CSS and JS I can leave this step till last in my schedule and know that it will not affect my final site if I run over and am not able to implement it. I also believe that Perch's excellent support and documentation will be invaluable if time becomes tight over the summer.

### **Hosting**

I'm planning on using Guru for my hosting account. The cloud based hosting they offer means it's essentially pay as you use rather than committing to a longer contract, which I have found an issue in the past with my current host Bluehost. This also means the website is less likely to come offline and will be able to manage traffic spikes if it is successful. They are based in the UK and support ecological hosting through the use of a renewable-energy datacentre, which are both factors of importance to me.



# #7 DELIGHT

## 7.1 Design research

In designing my website it is important to consider my users and what will appeal to them in terms of layout, typography and visuals. From my target audience survey in Commodity section, I have estimated that there is 70:30 split of male and female. The majority are between 21 and 39yrs old and their professions are based around development, design and data visualisation. Over half of those surveyed had experience designing infographics.

It is also important to consider the design trends and conventions in the existing online community in which my website will sit. From my research into these websites I have noted two different types of layout categories, a primarily single column blog style and a grid based gallery format.

### **CATEGORY 1:** **Blog style layouts**

#### **Trend analysis**

- Minimalist, white clean
- Colour coming from chart/work
- Vertical simple layout
- Imagery is snapshot

#### **Case Study** [thewhyaxis.info/](http://thewhyaxis.info/)

I have chosen [www.whyaxis.info](http://www.whyaxis.info)<sup>26</sup> as an example to analyse in depth because I feel it has the strongest design amongst the websites in the blog category.

The simple iconic branding has a subtle 3D treatment that I feel brings some depth to the design without becoming skeuomorphic. The font used throughout is Proxima Nova, which is clear and geometric like the visualisations the site is posting. There is also a coherent brand running from the yellow logo to the yellow colour treatments on images. This helps to unify the design and makes the visuals consistent despite the different content sources.

## **CATEGORY 2:**

### **Gallery style layouts**

#### **Trend analysis**

- Minimalist, white clean
- Colour coming from chart/work
- Organised grid used
- Varied colour palette

#### **Case Study**

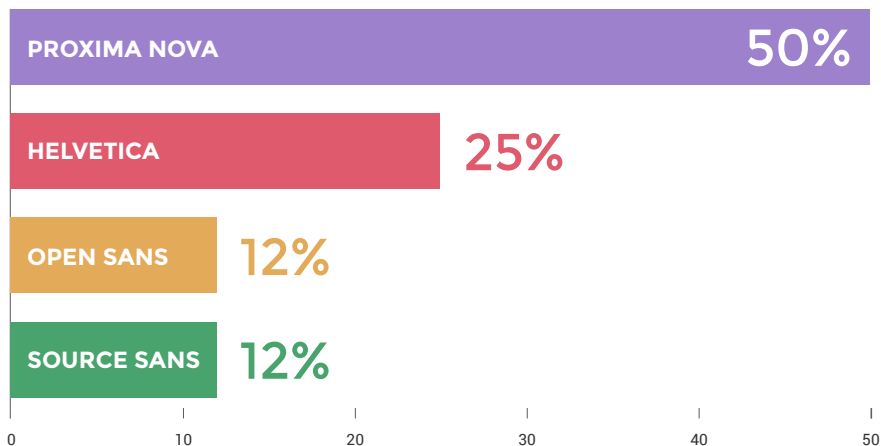
[bl.ocks.org/mbostock](http://bl.ocks.org/mbostock)

I am using Mike Bostock's portfolio page, [bl.ocks.org/mbostock](http://bl.ocks.org/mbostock),<sup>25</sup> as an example design to analyse. It's clear the use of a grid based layout makes sense, not just because of the amount of overview based gallery style content, but also because visual organisation is a key factor in presenting clearly in information design and data visualisation.

The background is white, like the majority of data visualisation sites, however the colour palette is extremely mixed as the colours come from the great variety of thumbnail chart examples. There is also no branding to speak of, probably due to the website being intrinsically linked to the open source library of D3 and to emphasise the sense of ownership for the D3 community. The simple functional design and layout also differentiates it from its competitors offering branded proprietary frameworks.

## Typography

Whilst analysing the designs of the websites within the online data visualisation community I surveyed which fonts each site was using.



There are close similarities between all these fonts.

Firstly, they are all sans-serif. Secondly, they are geometric in nature with open bowls and counters in the letterforms.

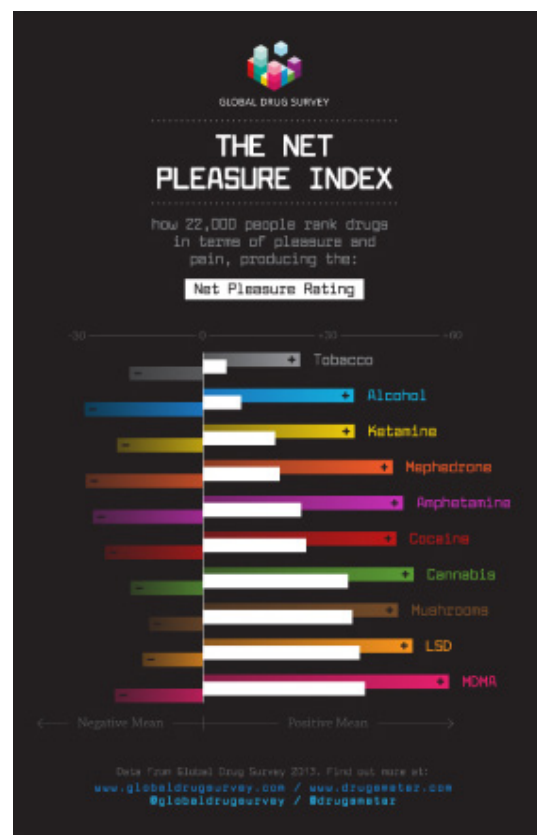
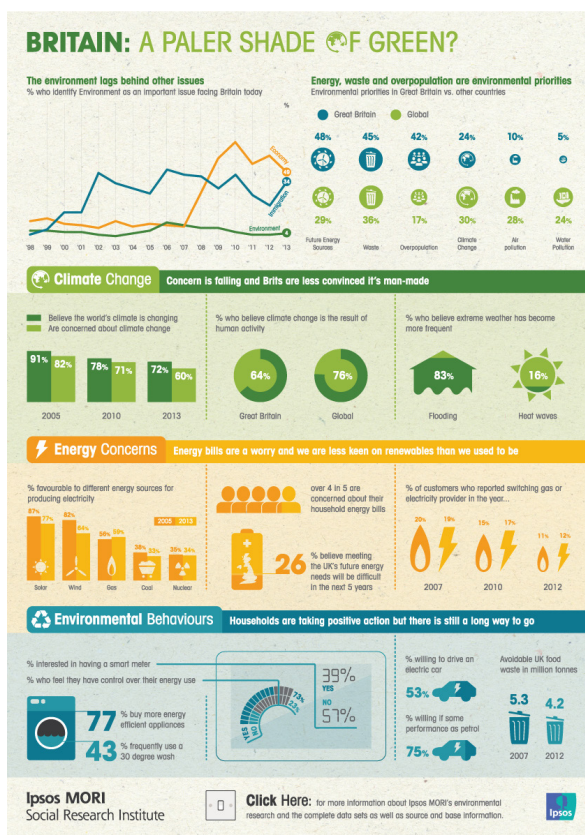
These fonts are highly legible and thus communicate clearly, as well as being utilitarian in style and thereby not distracting from the visualisations.

## Design research summary

The overriding trends in the existing online data visualisation community are for a clean white background, a blog or gallery grid style layout, and for the majority of colours being generated by examples.

## 7.2 My design work

The aims for my *Digital Data Viz* website are not only to help users learn and work with data visualisation tools but also to promote my own design and digital data visualisation work. In considering the design of the site I therefore not only have to consider the users, but also my own style and preferences as a designer.



Above are some examples of my work. I created the infographic on the left based on research data by Ipsos MORI. The design on the right was for the Global Drug Survey to publicise their 2013 survey data regarding drug use. On the whole my work is typified by the use of flat design, bright colours and simplicity. I always design using a grid and place a great deal of importance on typography and use of space.

## 7.3 Data visualisation principles

With the website's subject matter I feel it is very important that I also consider my design in regard to common principles of data visualisation.

### **Simplify**

Good data visualisation represents data efficiently without oversimplifying. A visualisation is complete not when there is nothing left to add, but when there is nothing left to take away.

### **Comparison**

Data visualisation is a method of representing data for the viewer to make conclusions. Therefore the ability to visually contrast differing information is important.

### **Importance of context**

A visualisation without clear explanation and labelling is meaningless. To communicate clearly data visualisation need to find a balance between text annotation and intuitive visuals.

### **Documentation**

Citing the source of your data is critical to providing openness and trust in the information you present.

### **Content is most crucial**

Good quality content is the most important part of a visualisation. Uninteresting visualisations are normally a data problem, not a design problem.

## **Applying these principles to web based visualisations**

During my initial research phase I looked at the website created by Swiss data visualisation agency Interactive Things<sup>8</sup> to display the collection of online tools they use in their work. From following their work since that time I believe they are one of the best examples of practitioners applying and adapting data visualisation principles successfully to the web.

The first thing that strikes me about their portfolio of work is their use of colour in considered, unified colour palettes for each different visualisation project. The second is the use of space within their work and the uncluttered visuals they produce. Finally, it is the balance of textual information and annotation that, through legible and considered typography, complements the data rather than overshadowing it.

Through researching the online data visualisation community it is clear that the quality of both design and development work present in Interactive Things, as well as a handful of other leading data visualisation studios, is the exception rather than the norm. The ability of agencies like these to bring together specialists in both data visualisation design and development and combine their skills through art direction is an expensive and lengthy process. As individual practitioners there is normally a weakness in one of these sides and for the most part my personal experience is that design, rather than development is weaker within the community. I do feel however, as an individual with a background in data visualisation design and front-end development, that I am well suited to combining the two areas.

## 7.4 Brand values

### Clean / Simple / Elegant / Organised

Through combining my design research, my own design style and the principles of data visualisation I chose these four values to build the *Digital Data Viz* website around. These don't simply apply to design and brand, but have influence over the content, structure and functionality.

## 7.5 Typography

### Requirements

- Legibility to communicate clearly and efficiently
- Work at small sizes on screen
- Range of weights to allow for flexibility
- Typefaces for headline, body copy and code examples that work together
- Open source webfonts

### Headline and title font choice

**Montserrat** - Similar to Gotham, the font used in Barack Obama's first US Presidential election campaign. It is robust, clear, geometric and trustworthy.

### Body copy font choice

**Roboto** - Created specifically for screen user interface (UI). It works well at small sizes and has a large range of weights.

**Consolas** - Designed for on-screen use, specifically for programming applications. It is free to use through the JS plugin prism.

Using the Typecast webfont service I also ensured all fonts worked well together by choosing those with a similar x-height. They are all available as open source fonts online which is important as my project primarily focuses on open source technology.

## 7.6 Colour palette

During my design research I found that a large amount of websites within the data visualisation community had a varied colour palette due the differing sources of visualisations. I believe that this was to the detriment of the overall designs, creating a busy and chaotic look and feel. Websites like The Why Axis<sup>26</sup> stood out due to their use of a unified colour palette.

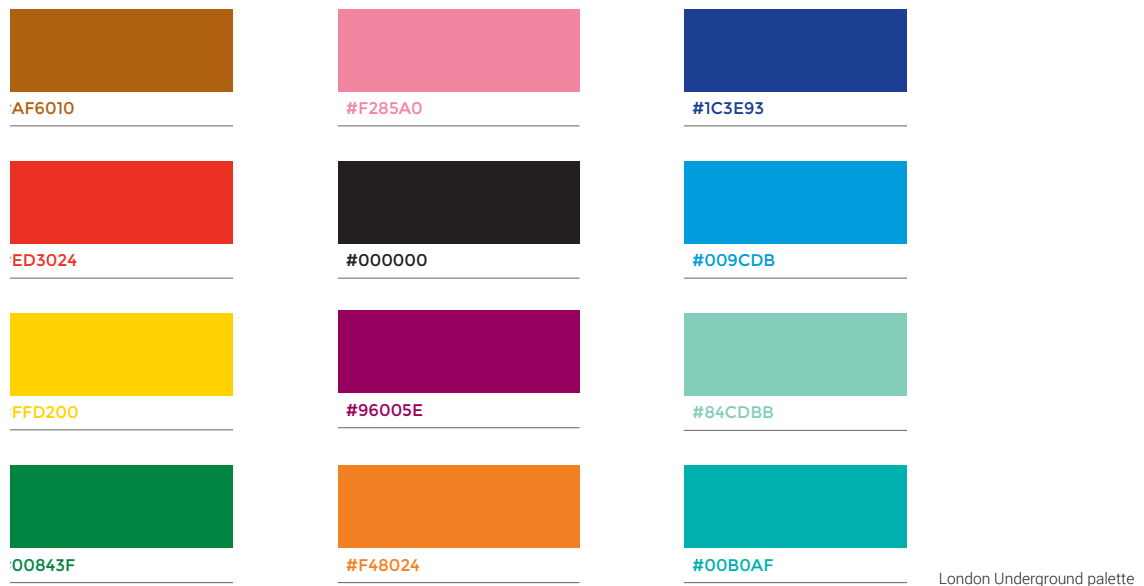
I have chosen to use a unified colour palette. This is firstly because I am keen to promote the design side of data visualisation as well as the development side. Secondly, I want a standardisation of colour to allow for clarity and lack of distraction so users can focus on function and form.

I also need a palette that works as a spectrum as data sets and variables may require specific colour meanings. They may also include a large amount of differing variables.

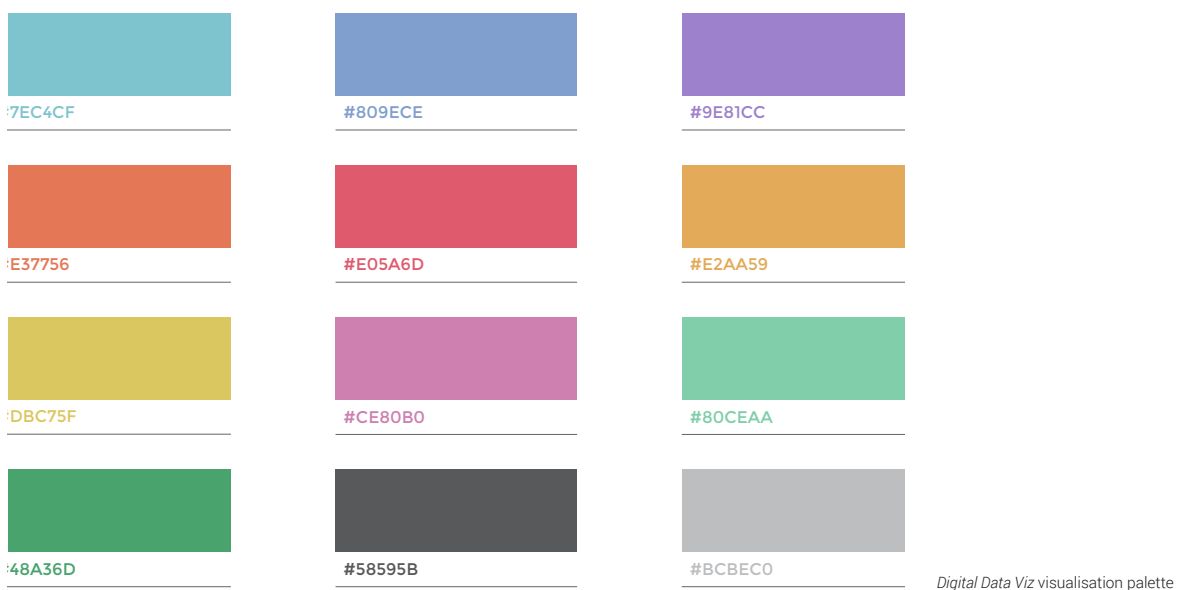


## Palette development

As a starting point for finding a unified spectrum based colour palette I began by looking at the colour scheme for one of the most famous pieces of information design, the London Underground map.



The second stage was to tone down the stark contrast and vibrancy of the colours so that they will work more harmoniously together.



The second stage then became the basis for colour spectrums used in visualisations like the multi-series interactive. Through different stages of prototyping and design/development iteration the primary website palette below, evolved through practical use to suit all the needs of the main site.



## 7.7 Logo

### Development

I began working with the font Montserrat for my logo design as I planned on using it in headlines throughout the website. It's a very simple, clean typeface, with an organised and consistent geometric grid like structure that suits the subject well. The rounded bowls in the uppercase D and G contrast well with the angular V and Z and create almost diagrammatical visuals.

Because data visualisation is the combination of graphical and text elements, I wanted the logo embody the subject matter rather than visually reference it through the use of text and an icon. Through iteration I found that I could tie the text elements together using a line graph through the A(s) and V.



Although I also tried to include horizontal bars to represent bar charts, during my prototype presentation and later testing with others, I found that they were competing too much visually and that neither was immediately evident to the viewer.

I therefore chose the stronger visual of the two, the line graph element and iterated the design to accentuate it. This was achieved by grouping the rest of the text in a neutral light grey and highlighting the zig-zag line graph with an accent colour.

Through my final live prototype I decided that the logo needed more legibility where the line graph overlaid the A(s) and V.

I also felt it needed more depth and texture to bring it out of the screen and make it appear almost touchable and interactive like

its subject matter. I therefore decided to add a light shadow to each A and V which increased the emphasis on these areas and through that their legibility. It had the added effect of introducing a sense of movement to the logo, which was unexpected but quite apt considering the use of animated SVG transitions within D3 and other visualisation tools.



*Digital Data Viz final logo set*

## 7.8 Layout

I based my layout around my chosen body text size and baseline measurement of 16px / 1em. Horizontally I used major gridlines at 128px / 8em. Those were divided in half, quarter and eighths for the minor gridlines at 64px / 4em, 32px / 2em and 16px / 1em. These measurements were used in margins and padding as well as in the vertical grid which used the baseline without major gridlines.

To translate my designs and grid layout from Adobe Illustrator into the browser I used `hashgrid.js`<sup>27</sup> that was built by Jon Tan and Jon Gibbins. The grid can still be seen on the live site by pressing and holding the G key and kept in place by pressing the H key once the grid is showing. Through iteration, the alignment

of the design to the vertical grid had to be adapted optically from an initial strict use. However the horizontal grid remains integral to the layout.

The use of colour categorisation is the most efficient method of differentiation in data visualisation and therefore it is integral to the structure and layout of *Digital Data Viz*. Tutorials, resources and examples are each assigned a colour and this runs through the navigation, link colours, the logo styling for each section, the background shades and the section summary or internal adverts on the homepage. The use of colour categorised backgrounds deviates from the trend within the online data visualisation community for white backgrounds. However I feel my use of this method adheres closely to the principles of data visualisation by quickly and effectively communicating orientation within the website and does not affect example visualisations because they are hosted on Github and bl.ocks.org.

## 7.9 Design conclusions

Overall I think my design challenges the convention of the trends within online data visualisation. *Digital Data Viz* provides a strong contrast to the trend within the community for white backgrounds and colour palettes generated through differing visualisations. I think this is largely because *Digital Data Viz* showcases my web design abilities and style at the same time as delivering useful content to the user. It is one of the few data visualisation websites that is fully responsive. The text is large and designed for easy legibility on screen. The colours are bold and used within a unified palette which is only used sparingly until you visit an example like the Multi-series Line Chart, where a spectrum is required to colour each variable. There is more personality in the design of *Digital Data Viz* and as I have outlined this is important because I want the website to reflect my character and particular design style.

I believe that the example visualisations I have created are still very much demonstrations of functionality rather than examples of visualisations designed to be optimised for use on the web. I plan to further develop the Multi-series Line Chart interactive example professionally as a tool for Ipsos MORI and through this process I will adapt its design to become more intuitive and user friendly. Once completed this will be part of my next tutorial series.

## #8 IMPLEMENTATION

### 8.1 First prototype

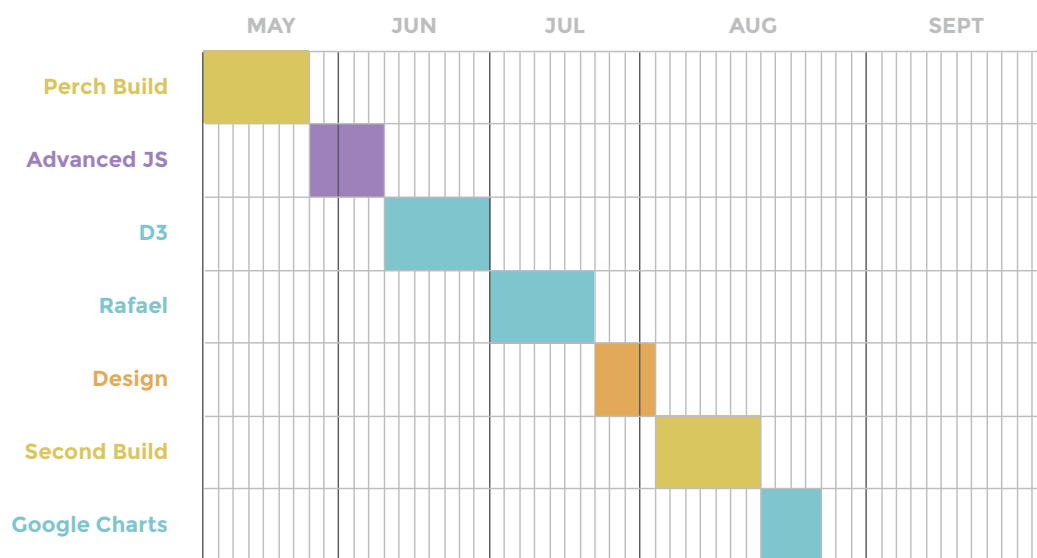
<http://grestrudun.users36.interdns.co.uk/prototype/>

For my initial prototype I was advised to write, design and build a sample article on web visualisation using a live example and code tutorials. As this would be my primary content page it would be the starting place for my layout and design considerations. The biggest challenge I found was the time involved in producing the content itself and manually creating and styling code indentations and colours.

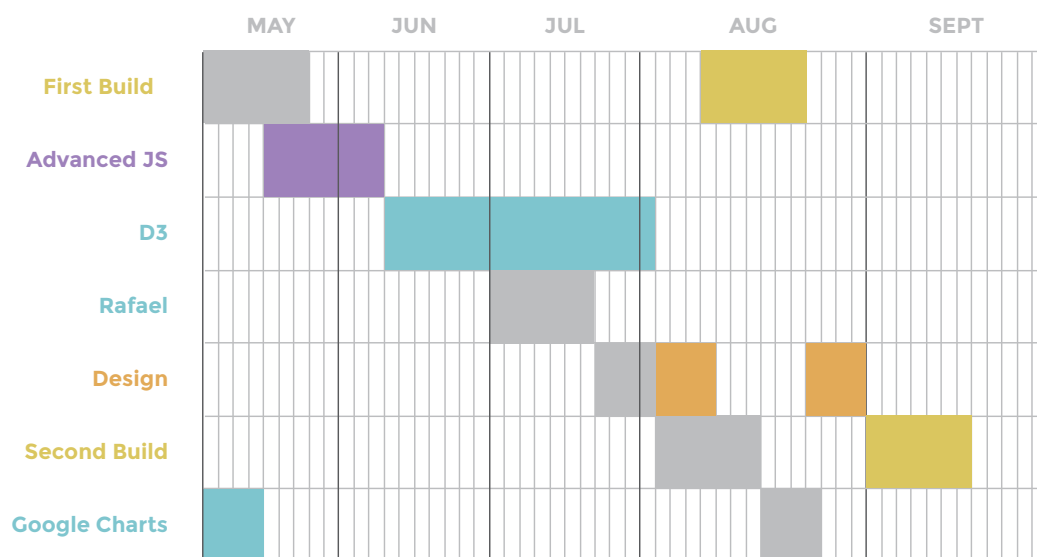
The critical feedback I received for my design was that using a single column design made the content difficult to read due to the large line lengths. It also meant that supporting information and links to the tutorial were lost in the page header and it was recommended I introduce a sidebar to combat both these concerns. Others mentioned that the coloured code text could be difficult to read and that the use of black boxes and the need to switch from reading black text on white to white and coloured text on black was too great a contrast and not comfortable for the reader.

## 8.2 Schedule of works

At the same time as creating my prototype I also created a gant chart with my proposed schedule for the summer to ensure I could estimate and plan my time well. The chart was not based on the amount of days in the summer, but how many days I knew I had available to work on the thesis.



Proposed schedule of works



Actual working pattern



The second gant chart demonstrates how my actual working pattern differed from my initial proposal. After the feedback on my initial prototype I still felt I was some way off knowing the final design and structure of my website so I decided not to move straight into integrating my prototype design with Perch CMS. This was the correct decision because learning JS and D3 took longer than I had anticipated and it was important for me to design around my content rather than fit it into a pre-existing template.

As you can see from the second gant chart JS and D3 in particular took longer than I had expected to learn and meant I could not fit in learning Raphael.js and integrating Perch CMS into my website. These are now both part of my future plans for the website.

The major reason for D3 taking so much longer than anticipated to learn was the difficulty in overcoming the steep learning curve. The example visualisation I chose to create was an interactive multi-series line chart. I took the opportunity described in the Business section of using Ipsos MORI's survey data and chose to visualise their long-term study into UK public opinions of the top issue facing Britain.<sup>28</sup> The data had been collected every month since 1974 and required plotting forty different issues. I chose this visualisation and dataset because it was an example of a popular requirement that had few D3 examples and no comprehensive online tutorials. This meant I needed to do wide research to bring together the different elements required from disparate examples and to deconstruct code that had not been annotated.

## 8.3 Second prototype

<http://grestrudun.users36.interdns.co.uk/tutorials/>

After successfully creating the multi-series interactive I decided to design and build a prototype based around my primary content. I designed the page in Adobe Illustrator for desktop first, before moving into responsive views, although I developed the prototype mobile first in the browser.

To overcome the challenge of manually styling the code boxes I researched and chose Lea Verou's JS plugin prism.js.<sup>29</sup> I also chose to use a tab based navigation design that helped orient the user experience as well as fitting with the brand values of simple and organised.

The greatest problem I encountered in building the second prototype was how to integrate my logo within the structure of my website. This led me back to my final design phase and final build.

## 8.4 Search engine optimisation (SEO)

My primary approach to SEO is in providing high quality, relevant and unique content. I decided to use a URL that references the

Keyword	Avg. monthly searches	Competition
d3	49,500	Low
d3.js	18,000	Low
d3 examples	2,400	Low
learn d3	140	Low
d3 tutorials	260	Low
data visualisation tools	2,900	High
interactive data visualisation	40	High
data visualisation	1,300	High

website's content subject and represented some of my keywords without creating too long an address. I feel that *Digital Data Viz* is succinct, memorable and avoids the need to use the long form of the word visualisation, which is not simple to type.

## Keywords

I used Google Adwords Keyword tool to pick the keywords. They appear in search engine friendly locations such as page titles, meta tags, headings, the first few paragraphs on the homepage as well as file and folder names. I also ensured there is a high keyword density throughout my content which occurs naturally rather than being contrived.

I have registered the site with Google webmaster tools and to avoid duplication of content and marking down by search engines I have a 301 redirect in my .htaccess file to change requests for digitaldataviz.com to www.digitaldataviz.com.

## 8.5 Social media

Through the promotion of my research survey I generated a number of contacts through Twitter and email, most notably Albert Cairo, Scott Murray and Andy Kirk. I have updated these contacts on the website's progress and asked that they again retweet a link to the site with the potential of reaching a possible 29,000 people interested in data visualisation.

I also plan to continue tweeting my latest work and tutorials to generate interest. By targeting both my personal contacts within data visualisation and well known practitioners I expect to generate a larger following and professional network in which to promote my work. I will also be creating a link from my personal portfolio website to *Digital Data Viz* to help promote the site and my experience in online data visualisation.

## 8.6 Future marketing plans

In addition to promoting the website on social media I also have the following plans to market the website online.

Once I have completed all the sections of the multi-series interactive tutorial I plan to post the link on the D3 tutorials wiki on Github. My current example is shown on Mike Bostock's site [bl.ocks.org](http://bl.ocks.org) and I plan to develop my portfolio within that site. Firstly, I will develop the design aspect of the current multi-series chart and post further examples that I create in work and in my spare time. Once I have a full multi-series chart tutorial I plan to go back to the D3 Google Group and Stackoverflow posts which I found requesting help and assistance in creating an interactive chart like mine and post my link. This will hopefully solve other users' problems as well as increasing traffic to the site and work as an advertising tool for my services.

Once I have built up a body of knowledge and a number of examples I will further promote *Digital Data Viz* and my own profile within the community by entering competitions in data visualisation. Whether these are successful or not they will provide useful experience and networking opportunities with others in a similar position to my own.

## #9 ANALYSIS

### 9.1 Site performance

Using Google PageSpeed Insights *Digital Data Viz* scored **74/100** for rendering speed on desktop. The site was confirmed to have avoided landing page redirects, minified HTML, optimised images, prioritised visible content and had a reduced server response time.

The site scored **64/100** for mobile browsing speed. The site was also given **95/100** for user experience on mobile. *Digital Data Viz* was confirmed to have avoided plugins, as well as rendering properly on all devices and had used legible font sizes.

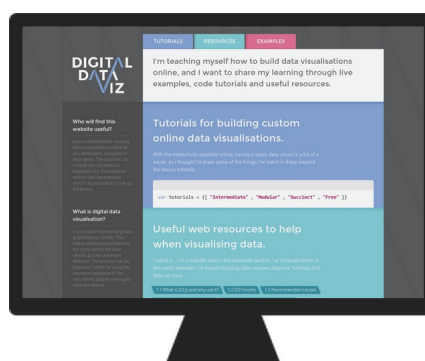
To increase the speed of the site Google recommends moving the render-blocking JS files for hashgrid and JQuery from the <head> of the document. This is something I have considered but the current setup is required for hashgrid to work and I'm keen for my site to be assessed with this in place as it was part of my building process. Once assessment is completed I can then remove hashgrid from the site and move JQuery to the bottom of the markup.

I also used the website testing tool Nibbler<sup>30</sup> with *Digital Data Viz* scoring **7/10** as an overall score. This was based on a number of different factors outlined below:

#### **Amount of content 8.4/10**

The website was found to have an average of 433 words per page. The amount of content on a website has been shown to correlate with its search engine ranking. However, I am confident that the more text heavy tutorial content will increase this average with the addition of more of these pages in the future.

- 7.0 Overall**  
The overall score for this website.
- 9.9 Accessibility**  
How accessible the website is to mobile and disabled users.
- 7.2 Experience**  
How satisfying the website is likely to be for users.
- 2.0 Marketing**  
How well marketed and popular the website is.
- 9.4 Technology**  
How well designed and built the website is.



Analysis results from Nibbler

## Internal links 9.4/10

Nibbler found the site had excellent use of linking to other pages, both internally and externally. Search engines can index content better if it is well linked to, and it can help those using non visual browsers like screen readers. This was achieved through the use of descriptive and succinct link text.

## Mobile optimisation 10/10

The site scored highly here because of my use of CSS media queries to create a responsive design. The major breakpoints were at 480, 640, 768 and 1024px or 30, 40, 48 and 64em. The fact that the breakpoints were at standard device sizes is not due to a rigid and inflexible development approach. I am instead a proponent of device agnostic responsive development, placing break points based on the design, rather than waiting for pre-set device dimensions. However, the breaks came naturally to the design due to my use of a horizontal grid system that was based on my baseline typographic measurement of 16px / 1em and through designing responsive views in Adobe Illustrator before developing in the browser.

## Page titles, Headings, Meta tags and URL format 10/10

All pages were found to use page titles appropriately which is important as it aids SEO. The headings were clearly defined, summarising the content of webpages quickly and increasing accessibility. Nibbler noted a good use of an instructive meta description on all pages, and included the visuals of how the site would look on Google search results. The use of appropriate web addresses and folder structure throughout the site means greater usability, SEO and marketability in future.

## Code quality 10/10

Nibbler classed the site as World Wide Web Consortium (W3C) compliant and this supports my validation testing during browser testing, which showed I had passed.

Nibbler found that *Digital Data Viz* scored low for marketing. This will be my first area to address in the next phase of the project, using the social media and SEO plan described in my Implementation section.

## 9.2 Accessibility

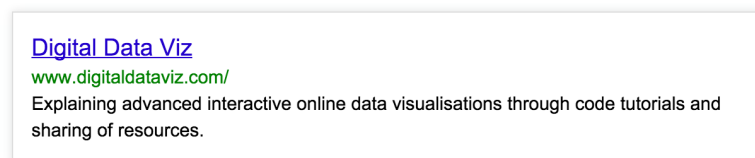
After analysing the website, Nibbler gave *Digital Data Viz* a 10/10 for accessibility. This was because of the use of the new semantic HTML5 structural elements and the use of ARIA roles, which identify and describe structural areas within the page aiding users with disabilities through optimising the website for screen-readers and keyboard navigation. There was also shown to be a good use of descriptive alt tags within HTML <img> tags.

However, through further analysis on the accessibility testing services WAVE<sup>31</sup> and checkmycolours,<sup>32</sup> I found that the contrast between some of the site's text and the colours used behind the text could be a concern for users with vision

impairment. This was echoed in feedback during my final presentation so it is an area I plan on addressing in the short term through the badly affected areas of sidebar and footer.



Wordcloud generated from  
*Digital Data Viz* headers



How *Digital Data Viz* appears  
in google search results

Moving forward I also plan to create a high contrast style option for visually impaired users.

## 9.3 Browser testing

Due to the fact that D3 is designed for use with the full capabilities of modern browsers, this defined the scope of browser support I needed the site to work with. This meant excluding testing for IE8 and below. *Digital Data Viz* is tested against Firefox, Chrome, Safari, Opera, IE9+, Android and iOS. The parts of *Digital Data Viz* that use D3, such as the tutorial and example pages may work in older browsers as the core D3



library has the minimal requirements of JS and the W3C DOM API. However modern browsers are required to use D3's SVG and CSS3 Transitions.

## 9.4 Feedback

### Tutor: David Watson

"Good explanation of the context. D3 JS library - open source but complex. The look and feel is generally good - particularly the layout. I find the desaturated colours a little too muted. Is it a good idea to make the tutorial copy UK specific (references from UK TV)? Maybe you should be hoping for an international audience. I think you've found a really nice niche and the project site looks well thought through and rounded."

**Response:** I feel I may have overstated the idea of UK specific references in the site during my final presentation. I mentioned this in relation to introducing a light-hearted tone of voice to my content. For example, in my multi-series interactive tutorial the second section is titled "Are you local?," and the first sentence in section states "A local chart for local data!" This is a reference to a skit from the BBC comedy show The League of Gentlemen. However, as it is relevant and adapted to suit the context it will not alienate users who are not aware of the reference.

### Tutor: Prisca Schmarsow

"~ wow - great achievement to get your head around JS!!  
a few initial comments:

- great intro to data viz, lovely slides as always :)
- beautiful transition from learning and goals to final site, nice work
- very well presented homepage with its sections, very strong and considered visuals
- great resource site and journey ~ amazed to tackled the heavy JS libraries

- your logo does work very well, colour scheme works well though the text on the sidebar might be a little hard to read
- beautiful use of space"

## 9.5 Analytics

Due to the scale of the work I undertook in learning JS and D3 before producing my own authored content, it has not been possible to launch the site in time to collect meaningful traffic analysis.

The domain was registered on 18th April 2014. However the site was launched in late September. The site is registered with Google webmaster tools and has Google analytics tracking embedded. Once I have put my previously explained marketing plans in place I expect to generate a steadily increasing flow of traffic to the site which I can maintain through posting of examples and tutorials of my latest online data visualisations created in my professional role.

## #10 CONCLUSION

To analyse the success of the *Digital Data Viz* project I have looked at the website against my initial brief and the aims developed through my research.

The value proposition of the website was in providing freely available tutorials, resources and live examples in web-based data visualisation that went beyond the basics, with the aim of assisting those wishing to create online data visualisations and reducing the intermediate level learning curve. The initial focus was to be on D3, with insights also provided on how to use proprietary frameworks, focusing in on customisations that go beyond the basic offering.

I feel the site as it currently stands achieves that proposition and has laid the structure to develop *Digital Data Viz's* value even further in the future.

Before developing the website, I identified two main potential weaknesses to my proposal. The first was my lack of JS and programming experience. The second was the ability to achieve my ambitions within the thesis timescale of learning D3 from scratch and overcoming the steep learning curve which was to form the content niche for my website before designing and developing the website around my content.

I feel I was correct to identify these two areas as the main potential dangers to the project as they were the biggest challenges to overcome. Looking back they are my proudest achievements from the project and though they proved the most difficult hurdles, they also gave me the greatest satisfaction and enjoyment.

From a cost perspective the project was not aimed directly at revenue generation. The alternative value of the project was to act as a portfolio site to promote my work both inside and

outside of my company and to raise my profile within the data visualisation sector.

Although marketing and social media promotion form the basis for the next phase of the project, I have had a distinct increase in enquires about undertaking professional data visualisation work in the last month. This is likely to be due to my increasing Twitter following from data visualisation practitioners and a general increase in my online profile. However it gives me confidence that I can build traffic and interest in *Digital Data Viz* and that there is a current demand for professionals that demonstrate experience in online data visualisation.

From a user perspective, my use of personas earlier in the project led me to conclude that prospective site visitors already had their data set and narrative and that they would be after examples that can be adapted or either instructions or resources to create their own work. I feel that my current tutorials and linked examples are an excellent start to providing instructional content to these users. Whilst I am populating the website with more instructional content I have ensured my Resources section provides a large variety of sources for examples, both at basic and advanced levels, as well as personally recommending a range of tutorials.

I also considered the user in relation to the tone of voice within the website's content. This was motivated through personal experience of some of my competitors' websites that delivered the code tutorials in a very dry monotone format and lacked enthusiasm for the subject, making the learning experience a struggle. I chose a light-hearted approach for my writing style, which reflects a trend within web technology blogging. I believe this was delivered without distracting from the content and was integrated through a light touch that would not alienate users without a similar sense of humour or background.

Before beginning the design process I outlined the following brand values for *Digital Data Viz*: clean, simple, elegant

and organised. I believe the site embodies these values and utilises the principles of data visualisation.

In analysing the project as a whole, I must also consider what I would have done differently in light of what I have learned from the experience.

I think the main point I have learnt from my experience is to integrate a website's social media campaign closely with the building process. I began promoting the project on Twitter very early on through the use of my target audience survey. However looking back I believe it would have been beneficial to follow this up by tweeting some initial design concepts, my live prototype pages and branding ideas.

I was not comfortable at the time of exposing unfinished work into the public domain. However, I now realise that I could have benefitted greatly from the feedback as a form of remote user testing during the design and development process. Also tweeting D3 experts with my problems whilst trying to overcome the learning curve would have had the beneficial effect of networking and site promotion. I am not naturally comfortable with social media use in a personal context, but I feel I have grown to understand its huge application to professional promotion and networking and this is something I will take into my freelance work and future projects.

I believe that this website is not a finished product but an excellent starting point for me learning digital data visualisation. My future plans involve three initial areas of immediate development. Firstly, to put my social media and online marketing plans into place to promote the website and increase my profile within the community. Secondly, I want to integrate Perch CMS into the existing structure to enable me to update it quickly and regularly in future while maintaining it as a repository and resource for my learning and professional development. Thirdly, I want to complete the multi-series interactive tutorial and produce a follow up series based on adapting a functional

prototype example into a data visualisation optimised for online and onscreen use.

I am immensely proud of what I have achieved in the process of creating *Digital Data Viz*. I feel I could not have put more time, energy and soul into the project. However the success of the project was not only down to myself.

## **Acknowledgements**

I would like to thank my wife Julie Struthers, my friend and coursemate Kate Lawton and my tutors Cole Henley, David Watson and Prisca Schmarsow for all their advice, support and encouragement in the last twelve months.

# #11 REFERENCES

1. SINTEF. (2013). *Big Data, for Better or Worse: 90% of World's Data Generated Over Last Two Years*. Available: <http://careers.theguardian.com/government-cuts-career-services>  
Last accessed 4th Oct 2014
2. Edward Tufte (1990). *Envisioning Information*. USA: Graphic Press
3. David McCandless (2010). *Information is Beautiful*. London: Collins
4. Mike Bostock. (2011). *D3: Data-Driven Documents*. Available: <http://vis.stanford.edu/files/2011-D3-InfoVis.pdf>  
Last accessed 5th Jun 2014
5. David McCandless. (2013). *Information is Beautiful.net update*. Available: <http://www.informationisbeautiful.net/>  
Last accessed 18th Feb 2014
6. Visually. (2014). *Visual.ly*. Available: <http://visual.ly/>  
Last accessed 18th July 2014
7. Andy Kirk. (2014). *Visualising Data*. Available: <http://www.visualisingdata.com/>  
Last accessed 15th September 2014
8. Interactive Things. (2014). *User experience & data visualization studio*. Available: <http://www.interactivethings.com/>  
Last accessed 28th July 2014
9. Interactive Things. (2011). *Selected tools*. Available: <http://www.selection.datavisualization.ch/>  
Last accessed 10th March 2014
10. Scott Murray. (2011). *Aligned Left*. Available: <http://www.alignedleft.com/>  
Last accessed 8th October 2014
11. Scott Murray (2013). *Interactive Data Visualization for the Web*. USA: O'Reilly
12. Sebastian Gutierrez. (2012). *Learn how to make Data Visualizations with D3.js*. Available: <https://www.dashingd3js.com/>  
Last accessed 29th September 2014
13. Malcolm Maclean. (2014). *D3 Noob*. Available: <http://www.d3noob.org/>  
Last accessed 18th September 2014

14. Malcolm Maclean (2014). *D3 Tips and Tricks, Interactive Data Visualization in a Web Browser*.  
USA: Leanpub
15. GE and Seed Media Group. (2014). *Visualizing*  
Available: <http://www.visualizing.org/>  
Last accessed 10th August 2014
16. Money Supermarket. (2014).  
Available: <http://www.moneysupermarket.com/>  
Last accessed 24th Feb 2014
17. Alex. (2014).  
Available: <http://www.alex.com/>  
Last accessed 4th October 2014
18. Simon Rogers, Guardian. (2013). *John Snow's data journalism: the cholera map that changed the world*  
Available: <http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map>  
Last accessed 3rd October 2014
19. Public Health Perspectives. (2013). *Lessons from the ghost of John Snow*  
Available: <http://blogs.plos.org/publichealth/2013/03/13/lessons-from-the-ghost-of-john-snow/>  
Last accessed 3rd October 2014
20. Ian Sample, Guardian. (2012). *ClickClinica, the app that maps disease outbreaks*  
Available:  
<http://www.theguardian.com/science/blog/2012/nov/26/clickclinica-app-map-disease-outbreaks>  
Last accessed 3rd October 2014
21. Mashable. (2013). *Twitter Can Track Disease – Can It Predict Outbreaks?*  
Available: <http://mashable.com/2012/06/08/social-media-disease-tracking/>  
Last accessed 3rd October 2014
22. Wikipedia. (2014). *Amyotrophic lateral sclerosis*  
Available: [http://en.wikipedia.org/wiki/Amyotrophic\\_lateral\\_sclerosis](http://en.wikipedia.org/wiki/Amyotrophic_lateral_sclerosis)  
Last accessed 3rd October 2014
23. Charity: Water. (2011). *Water changes everything*  
Available: <https://www.youtube.com/watch?v=BCHhwvxvQxg>  
Last accessed 4th October 2014
24. The Guardian. (2013). *The computer analyst whistleblower with top-secret NSA documents*  
Available: <http://www.theguardian.com/us-news/edward-snowden>  
Last accessed 4th October 2014
25. Mike Bostock. (2011). *mbostock's blocks*  
Available: <http://bl.ocks.org/mbostock>  
Last accessed 5th October 2014



26. The Why Axis. (2014).  
Available: <http://thewhyaxis.info/>  
Last accessed 14th September 2014
27. Jon Tan & Jon Gibbins. (2012). *hasgrid.js*  
Available: <http://hashgrid.com/>  
Last accessed 1st September 2014
28. Ipsos MORI. (1974-2014). *The Issues Index*  
Available: <https://www.ipsos-mori.com/researchpublications/researcharchive/2420/Issues-Index-Archive.aspx>  
Last accessed 30th September 2014
29. Lea Verou. (2012). *Prism.js*  
Available: <http://prismjs.com/>  
Last accessed 30th September 2014
30. Silktide. (2012). *Nibbler*  
Available: <http://nibbler.silktide.com/>  
Last accessed 6th October 2014
31. WebAIM. (2011). *Wave*  
Available: <http://wave.webaim.org/>  
Last accessed 6th October 2014
32. Giovanni Scala. (2011). *Checkmycolours*  
Available: <http://www.checkmycolours.com/>  
Last accessed 6th October 2014