Data Visualization Coursework:

1. Research Question(s):

There are a number of research questions that motivated and shaped my visualization. The data blog provided a comprehensive dataset about both divorce and marriage. Extensive data was provided from years 1974 -2012 so only data from these years was used in visualizations as to not skew visualizations for years where there is no information.
This data allowed for vast exploration into divorce trends and prompted research questions like;

- How have divorce trends changed in the last forty years relative to marriage trends?
- What is the most common reason for divorce in each year?
- How have patterns in reasoning for filing divorce changed over the past forty years?
- Who are divorces granted to the most, wives or husbands?
- At what age are people getting divorced? Has this changed over time?
- How long do marriages that end in divorce last on average? Are marriages longer or shorter over time?

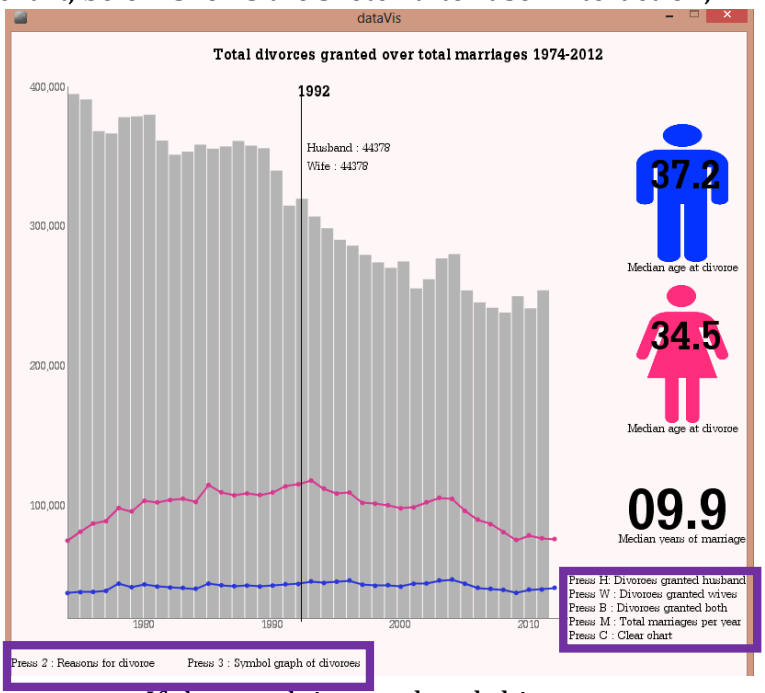
These research questions have helped to shape visualizations with regards to both aesthetic and content. Establishing questions to be answered through visualizations allow strong objectives to be set, allowing for effective designing an implementation.

2. Data Source(s):

Divorce rates data, 1858 to now: how has it changed?
http://www.theguardian.com/news/datablog/2010/jan/28/divorce-rates-marriage-ons

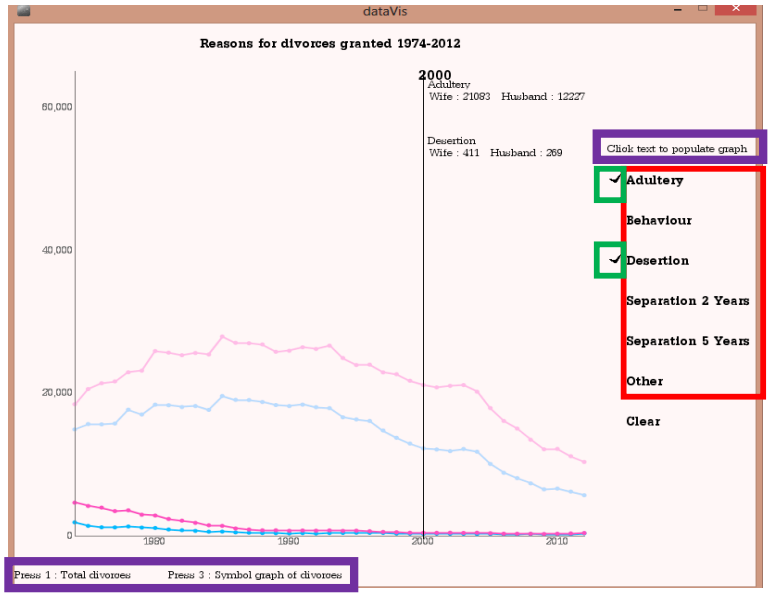
3. User Instructions:

The visualization consists of three different views; total divorces over marriages chart (sketch1), reasons for divorce(sketch2) and total divorce symbol chart(sketch3).
On compilation of the code the sketch displayed is the blank total divorces over marriages chart, below shows the sketch after user interaction;

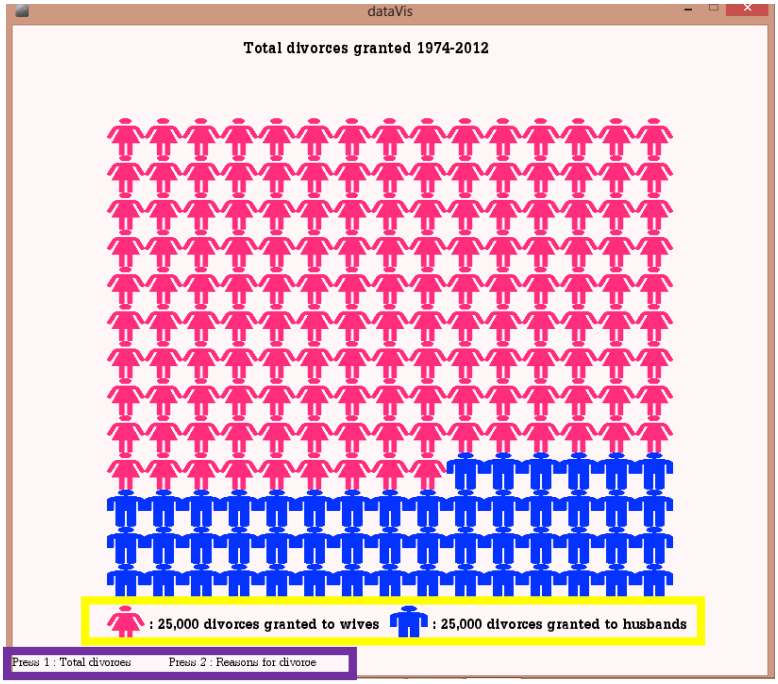


This sketch can also be accessed from sketch2 and sketch3. Key press instructions are highlighted in purple. The chart is populated using different key press events (H, W, B, M or C) as instructed in the sketch. Keys can be pressed in any order and the chart will be updated reflecting the key press event. The user can also switch views in the application using number keys (2 or 3) as instructed in the sketch. To reveal more specific information about the graphs, move the mouse horizontally across the graph area, the year label updates to the year that mouse and line position represent. If the graph is populated this mouse movement will also dynamically update the corresponding years median age at divorce and median length of marriage display (if only husband chart is shown only husband median age is shown and so on).

Press ‘2’ to display Reasons for divorce (sketch2), sketch2 can be accessed from sketch1 and sketch3. When this view has been selected via the key press event ‘2’ an empty chart will be displayed. To populate this chart, click the reasons highlighted in red, more than one reason can be selected; on selection, the corresponding chart will be plotted and a tick will be displayed confirming requested selection. Similarly to sketch1, to reveal more information about the displayed chart, move the mouse horizontally over the chart area. Labels will update showing the reason and the corresponding number of divorces granted to either the husband or wife for the selected year. To remove a chart, click the tick next to the reason, highlighted in green. Selection and removal of reasons charts can be done on multiple occasions. To remove all charts, click the ‘Clear’ option.



The user can also switch views in same way as before using number keys (1 or 3) as detailed in the sketch. All instructions for this sketch are highlighted in purple.



Press ‘3’ to display total divorce symbol chart. This sketch can be accessed from sketch1 and sketch2. This sketch has limited interaction; the user can switch to sketch1 or sketch2 as per the instructions highlighted in purple. The key is found below the chart and is highlighted in yellow.

4. Design Justification:

The design of this visualization has been carefully considered to provide maximum impact to the user. All aspects of design were chosen using various design concepts and ideologies to reflect the data honestly and present the user with the opportunity for exploration.

4.1 Colour:

The data used to populate visualizations was predominantly focused on statistics for the husband and the wife separately with only small amounts of data detailing a combined

total. The choice of colours to display data was dictated by the gender difference described in the datasets. Since 2005, JeongMee Yoon has been carrying out a project "The Pink and Blue Projects" inspired by her daughters favorite colour; pink. This project found girls preference towards pink and boys towards blue since the second world war was "universal and widespread" attributed to persuasive media campaigns (Popova, 2009). With this wide spread association with colour and gender, data about the wife was to be displayed in a shade of the colour pink and data about the husband to be displayed in a shade of the colour blue. Whilst using two different colours to demonstrate the data should allow the user to make an instantaneous assumption that the colours mean two different things. The use of the gender associated colours attempts to encourage the user to categorize the data correctly for enhanced understanding. This initial introduction to the colour differences occurs in sketch1, when the user selects either husband or wife data to display, this early association with the colours is carried out through the application and should allow the user to feel informed.

4.2 Layout:

The choice to allow the user to switch between three different sketches was to emphasize the difference in data. There was enough data about both; the total number of divorces and marriages, and the reasons for divorce to warrant standalone visualizations. Different sketches were used to display these visualizations as to not clutter the screen and confuse the user. Edward Tufte describes the importance of graphical integrity for visualizations in "*The Visual Display of Quantitative Information*" He explains the principle of this is "show data variation, not design variation" (Tufte, 2001, p53-77). With this principle in mind it was important to let the data speak for itself and provide simple and well informed charts to explain large proportions of data.

To "enable user familiarity and adoption" (Goodman, 2007) emphasizes the need to strive for consistency with design. These visualizations incorporated standardized colours, layout and navigation to strive for a high level of consistency and satisfy the users needs.

4.3 Symbolisation:

To denote information about husbands and wives separately male and female symbols were used as below;

Male (husband)

Female (wife)

The symbols used were chosen because they are widely synonymous with males and females respectively. These symbols are instantly recognizable to the user and are used to produce a symbolic graph showing total numbers of divorces over the past 38 years. Each symbol is representative of 25,000 divorces granted and who it was granted to.

This symbolism was also used to represent husbands and wives median age at divorce in the "Total divorces granted over total marriages 1974-2012" chart. Consistent symbolization was used to ensure that the user was able to establish a connection between the symbols and the corresponding data sets allowing quick comparisons to be made.

4.4 Interaction:

Interaction with this visualization comes in numerous forms; key presses, mouse clicks and mouse overs. The application layout and interaction design was loosely based upon the Visual Information Seeking Mantra "Overview first, zoom and filter, then details-on-

demand" (Shniederman, 2006). The application seeks to provide an overview of data, filtering of data through interaction and details on demand through interaction.
Key press interaction enables the user to interact with the "Total divorces over marriages 1974- 2012" chart, the user is clearly instructed about which keys to press in order to populate the chart, satisfying the "filter" aspect of the mentioned mantra. Key presses are also used to switch between sketches to display different aspects of the data set. Although key presses are intuitive (press 'h' to see husband data, 'w' for wife data) there are clear instructions provided to the user so they are able to knowingly switch between sketches.
Another way of adhering to "filtering" aspects was to integrate mouse presses in interaction with sketch2 "Reasons for divorces granted 1974-2012". The user is directed to select reasons via a click interaction to populate the page, the user is also able to remove the reason's data from the chart by another simple mouse click. This is a familiar form of interaction with an application, using this standard method yields a very effective type of filtering.
Overview and details on demand are fulfilled by the mouse movement interaction with the charts. A simple mouse interaction using mouse movement is not explicitly described in the sketches themselves, it relies on users intuition to 'play around' with the application. Specific details are provided when the user moves their mouse over the chart area in sketch1 and sketch2, a guideline will appear with a year label that corresponds to the year represented at the mouse location. In sketch1 it was decided not to have too many labels on this guideline, rather to populate other areas of the screen as to not clutter the already busy chart. The simpler chart in sketch2 displays labels detailing exact total figures for each reason that is currently plotted on the chart. These additional details allow the user to draw conclusions based on real facts and figures.

5. Further Work:

Further enhancements to these sketches could involve a further interaction on sketch1. It could be possible to derive the year people got married from the year they divorced combined with the median length of marriage statistics. This may enable further scope for investigation into marriage trends and could be possible to make predictions about the future.
Details about facts and figures could be encased in a smooth tooltip rather than using a text label on the guideline, this may help toward clarity.
The final feature which could be enhanced is the graphs, adding a sort function to find out which year has the most total divorces and the most reason for divorce in each year. The user could still currently do this using sketch1 and sketch2 although an interaction which does the sorting for the user may enable quicker comparison and prove more effective.

6. References:

Goodman, R. (2007). *Design Principle 5. Consistency.* Available: http://ryangoodman.net/blog/2007/08/20/design-principle-5-consistency/. Last accessed 29th April 2014.

Popova, M. (2009). *The Pink and Blue Projects: Exploring the Genderization of Color.* Available: http://www.brainpickings.org/index.php/2009/12/11/pink-and-blue-project/. Last accessed 25th April 2014.

Tufte, E. (2001). Graphical Integrity. In: Tufte, E *The Visual Display of Quantitative Information*. 2nd ed. Connecticut: Graphics Press. p53-77.

Shneiderman, B. (1996) The eyes have it: A task by data type taxonomy for information visualization, *Proceedings of the IEEE Symposium on Visual Languages*, pp.336-343.