

Visualization of UK Road Traffic Accidents

Team Collision: Chua Feng Ru, Dina Heng, Lim Ting Zhi

Abstract – Britain has one of the best records in the world for road casualties. However, hundreds still die on the roads every year. This has led to a cause for concern. Using official data recorded by police in Great Britain (data.gov.uk) between 1999 and 2010, we have plotted every road collision in which someone died. To help the common users understand and gain insights from all the UK road traffic accidents that have happened before, we plan to visualize demographics of drivers and casualties and examine the underlying factors which contribute to accidents. Our application makes use of bar charts which are dynamically linked to an interactive Radial Bar Chart diagram for users to explore and visualize UK traffic road accidents in an easy way.

Index Terms – UK Road Traffic Accident, Visual Analytics, Radial Bar Chart, Bar Chart, Visualization

1 INTRODUCTION

For an everyday activity, travelling by road is probably the riskiest thing many of us do on a regular basis. On average, some seven people are killed every day on the roads in Great Britain. Hundreds more are injured, many of them seriously, often with life changing consequences. In 2008 alone, 2,538 people died and nearly a quarter of a million were injured. In the past 10 years, the death toll has amounted to 32,298.

We can only expect the number of UK road traffic accidents to continue rising in the near future. In order to facilitate users to better understand the details of the UK road traffic accidents, we provide a visualization application using the data of all the attacks from 2000 to 2012. This application provides them a platform to discover and explore the road traffic accidents on simple interactive charts to help them examine the underlying factors which contributes to accidents, understand demographics of drivers and casualties.

In this project, our team will be focusing on 2015 road safety data. The data is obtained from [data.gov.uk](https://data.gov.uk/dataset/road-accidents-safety-data) (<https://data.gov.uk/dataset/road-accidents-safety-data>). It contains only personal injury accidents on public roads that are reported to the police and recorded using the UK STATS19 accident reporting form. It consists a total of 3 datasets that provide information about the accidents, the types of vehicles involved and the demographics of the casualties.

Most of the data attributes are coded and re-coding would be done with the lookup tables provided by data.gov.uk.

In this paper, we are interested in observing and analyzing traffic data, reviewing existing work in presenting UK traffic accident data from 2000 to 2012 and introducing the research and methods of creating our web-enabled application. It consists of 7 sections: section 1 provides a general introduction of the paper, followed by the problem and motivation of this research in Section 2; section 3 will explain the process of data gathering and preparation before we develop the visualizations; Section 4 details the approach in developing the application; Section 5 entails the system architecture consideration; Section 6 consists of an explanation of our final visualization and section 7 will conclude our findings and highlighting possible future additions to the research.

2 MOTIVATION AND OBJECTIVES

In United Kingdom (UK), road traffic accidents have resulted in 1,732 deaths in the year 2015, which is a 2% dip as compared to the year 2014. [1] Despite the drop in the number of deaths, the casualties across all severities remained at an alarming figure of 186,209. As there is an increasing demand for the use of public roads, there is a strong need for us to prevent road traffic accidents and make the public roads as safe as possible. In order to prevent such

accidents, it is therefore crucial to understand what are the different factors that contribute to road traffic accidents, and these understandings may then be used to prevent road traffic accidents from occurring. The main objectives of our research include:

- To understand the demographics of drivers
- To understand the demographics of casualties (Distribution of age of casualties, severity of casualties and type of casualties)
- To examine the underlying factors which contributes to accidents. The following are some factors, but not limited to:
 1. **Temporal patterns:** Accident records based on time.
 2. **Weather conditions:** Which type of weather conditions would cause more accidents?
 3. **Road conditions:** Which type of road conditions would cause more accidents?
- Develop appropriate interactive visualization to allow discovery of insights from multiple dimensions from the dataset.

3 DATA GATHERING AND PREPARATION

In this project, our team will be focusing on 2015 road safety data. The data is obtained from data.gov.uk (<https://data.gov.uk/dataset/road-accidents-safety-data>). It contains only personal injury accidents on public roads that are reported to the police and recorded using the UK STATS19 accident reporting form. It consists a total of 3 datasets that provide information about the accidents, the types of vehicles involved and the demographics of the casualties. Most of the data attributes are of ordinal data types and re-coding would be done with the lookup tables provided by data.gov.uk. As there are many columns associated to each row of all the road safety data, we had to download the data dictionary to identify the columns that are of interest in our visualizations. The following data attributes are used in this project:

- **Accident Dataset** [Accidents_2015.csv]
 1. Accident_Index [Accident No.]
 2. Day_of_Week
 3. Date

4. Time
 5. Weather_Conditions
 6. Road_Surface_Conditions
- **Vehicles Dataset** [Vehicles_2015.csv]
 1. Accident_Index [Accident No.]
 2. Vehicle_Reference [Vehicle No.]
 3. Age_of_Driver
 - **Casualties Dataset** [Casualties_2015.csv]
 1. Accident_Index [Accident No.]
 2. Vehicle_Reference [Vehicle No.]
 3. Casualty_Reference [Casualty No.]
 4. Casualty_Class [Driver/Rider, Passenger or Pedestrian]
 5. Age_of_Casualty
 6. Casualty_Severity

We saved this data into a Comma-Separated Values (CSV) file format for better readability through our JavaScript codes.

4 VISUALIZATION APPROACH

We began to research on the visualizations on UK traffic accidents that are available for public viewing over the web and also developed several charts to assess the effectiveness of visualizing the data with the various charts. Through our research, we also discovered that there is no existing visualization work, which allow users to analyze factors of casualty-related accident. An interesting we found was from the Guardian Data Blog, as shown in Fig. 1 below.

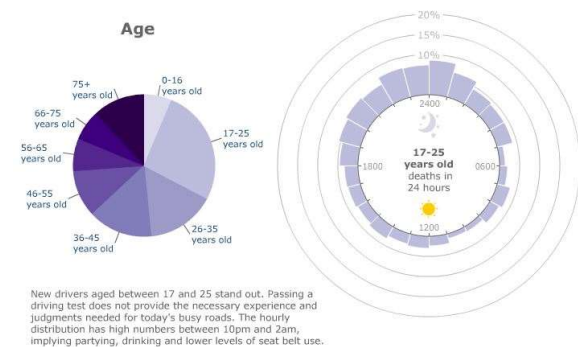


Fig. 1: UK road accidents 2000-2011 Radial Bar Chart

Interactive Visualization to Track Accidents Involving Casualties

We felt that the Visualization provided in Fig. 1 was very informative as it allows the user to

select the desired category of data to display. The radial bar chart will then allow you to look at the distinctive pattern of each age group over a time-period. This chart will be beneficial for high number of bins, where we will be able to look at all the bars or columns from one view without scrolling back and forth.

Following our research on the appropriate visualizations for the data, we proceeded with building charts and integrating them into an interactive application. The following subsections below details our process of selecting and building of the charts.

All of our data visualizations charts are developed using Data-Driven Document (D3.js), a JavaScript library for manipulating documents based on data library [2] written by Michael Bostock [3]. The D3 library allows us to manipulate elements of a web page in the context of our dataset. These elements can be HTML, SVG or Canvas elements and can be manipulated according to the contents and structure of dataset [4].

4.1 Initial Visualization Strategy

The main goal of the visualization design is to be easy to use, fast-processing, simple to understand, and also to avoid chart junks. This is so as we believe that it will aid user in their visual analytical process.

Our initial visualization strategy is to use a drill-down approach, where our visualization is being segmented into 3 major portions. With reference to Fig. 2, at the top, the first visualization is a map of UK, where it allows users to focus on different regions of UK. The second visualization serves as an intermediate navigation-step, where users can then focus on more in-depth details such as the different casualty types. Lastly, at the bottom section shows the underlying factors. Through the first 2 visualizations, users will be able to study how the casualties are associated with the underlying factors.

With reference to Fig. 3, our intention to use the sunburst diagram would require us to define the hierarchy within the casualty-related data as there is no existing hierarchy within the casualty-

related dataset. As from our strategy and exploration of the data, we have decided the hierarchy to be in the following format:

1. Age group with each bin of 10 years old.
2. Casualty class, which can be either driver, rider, passenger or pedestrian.
3. Casualty severity, which can be minor, serious or fatal.

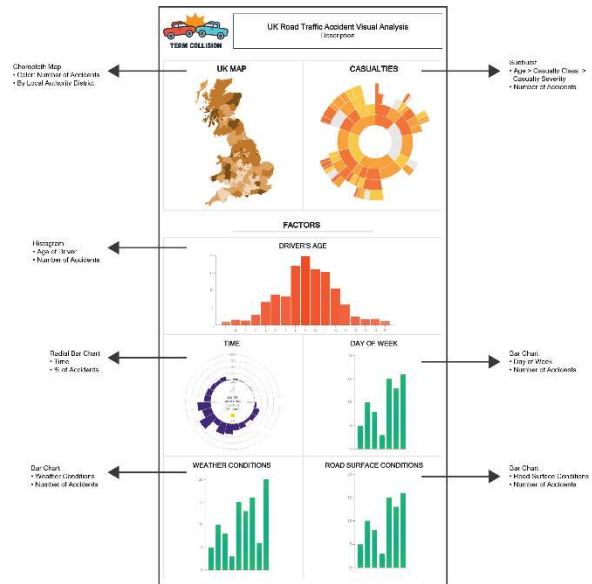


Fig. 2: Initial Visualization Strategy

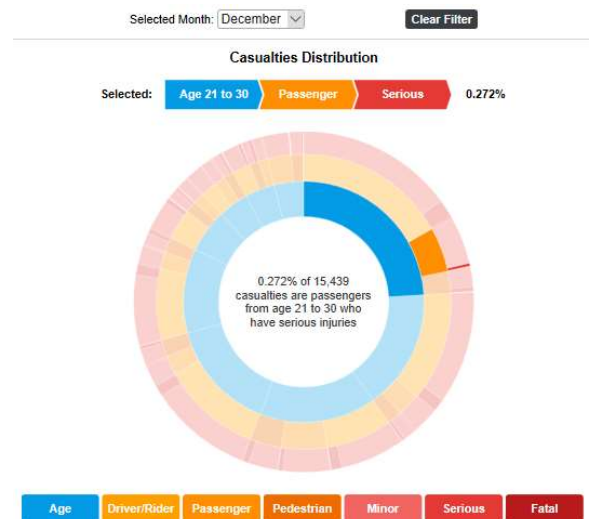


Fig. 3: Casualties Distribution

4.2 Revised and Finalized Visualization Strategy

After our preliminary assessment of the visualization strategy in the previous subsection, our team realized that the initial visualization strategy would result in a visual clutter and lack of space, which may hinder the visual analytical process of the user. Thus, we decided to make a decision between choosing a Sunburst diagram and Choropleth map. Originally, we selected the Choropleth map to visualize the number of accidents per land authority district as the users might want to visualize which district has the most and least accidents. Furthermore, we felt that using all three charts would be a problem when we integrate our charts together as it might not display the information effectively if only a small space is allocated for the Choropleth map.

On the other hand, a sunburst diagram would show the distribution of casualties more effectively. We implemented a Sunburst diagram for users to drill down according to the age, driver/ rider, passenger, pedestrian, minor/ serious/ fatal injuries. The darker the shade represents the more serious the injury is, as shown in Fig. 4.



Fig. 4: Sunburst implementation

Furthermore, our team also included a “Month” filter, and this is because of 2 main reasons:

1. The entire dataset is too large, and it is infeasible to load a huge amount of data

to the Document Object Model (DOM) within D3.js.

2. Certain weather or road conditions, such as snow conditions, only appear in the winter seasons from December to February. Thus, it will be logical to allow the user to visualize the data by month.

Modification of Visualization after Revised Strategy: After above assessment on the initial visualization strategy, we modified the initial visualization ensure that the graphical furniture can fit within 1 single page. With reference to Fig. 5, we chose to maximize the screen estate by using a landscape view, and segment the visualization into 2 major portions (e.g., left with general details, and right shows the detailed factors).

From the left, the first visualization (Left) serves as the main navigation, where users can focus on the types of casualties to analyze on.

On the right, the graphical furniture shows the underlying factors. This will allow the users to study how the various factors in the data contribute to different casualties for different age-groups.

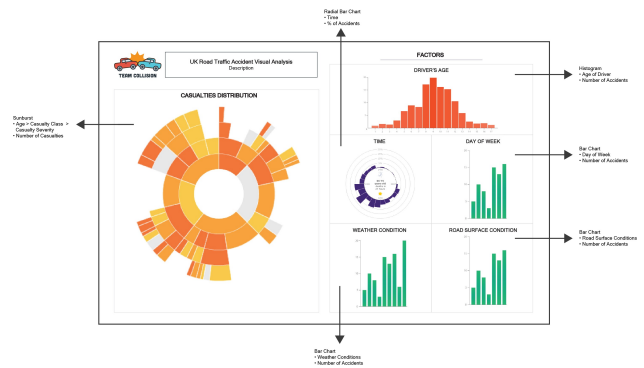


Fig. 5: Revised and finalized visualization strategy

4.3 Visualizing the No. of Accidents According to Driver's Age

To visualize the distribution of accidents across driver's age, we implemented a histogram to display the age distribution of the driver's age, as shown in Fig. 6. This allows user to analyze how driver's age plays its part in casualty-related

accidents, whether are the drivers involved old or young.



Fig. 6: No. of Accidents on a bar chart according to driver's age

4.4 Visualizing the Time of Accidents

To further present the data in a more interesting manner, we implemented a radial bar chart to allow users to visualize the percentage of accidents that occurred in each hour of the day in the month. This can be seen in Fig. 7. below. This allows users to find out the time at which most casualty-related accidents happen.

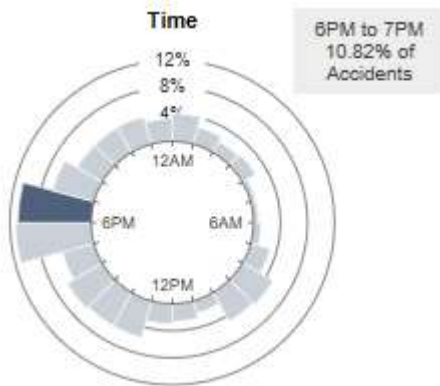


Fig. 7: Sunburst chart displaying the percentage of accidents in each hour per day

When a user mouse-over a block, a tool-tip function will display the time and the percentage of accidents that happened in that hour.

4.5 Visualizing the No. of Accidents According to the Day of Week, Weather Condition and Road Surface Condition

To visualize the number of accidents held in each day of the week, we implemented three bar charts displaying the number of accidents according to the day of the week, weather condition and road surface condition respectively.

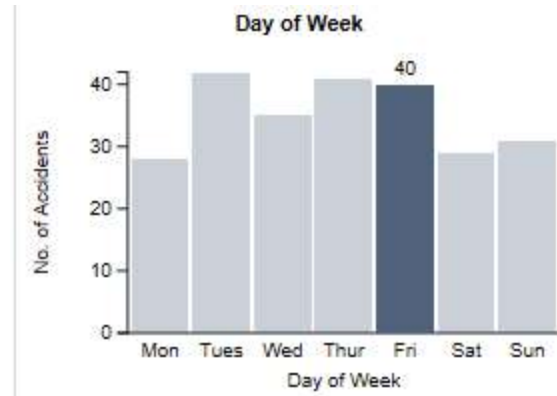


Fig. 8: No. of Accidents on a bar chart according to day of week

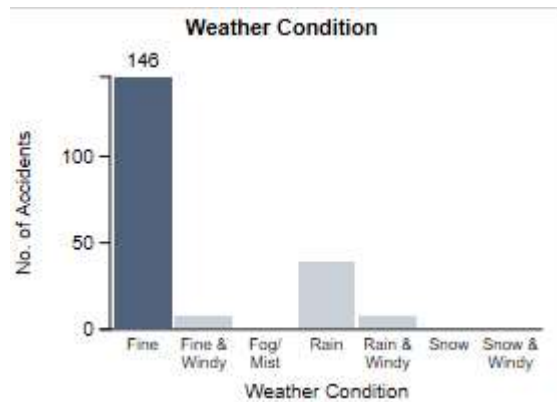


Fig. 9: No. of Accidents on a bar chart according to weather condition

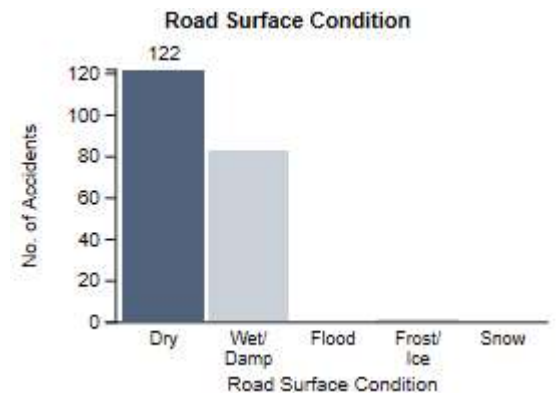


Fig. 10: No. of Accidents on a bar chart according road surface condition

Thus, the user would be able to identify what kind of road surface or weather condition causes the most casualties. Also, if day of the week plays

a part, as there may be more traffic in the weekdays and less traffic in the weekends.

5 System Architecture

Even though browser technologies and capabilities have improved much from the past, it would be infeasible to load a big chunk of the data into the Document Object Model, which the D3.js library uses, as browsers are not designed to load such huge amount of data.

Thus, in order to visualize a huge volume of data and to fulfil the team’s goal of ensuring a positive visual analytical process for the users, we have derived a few technical considerations:

- Segment large datasets to be by month
As the dataset contains around 180,000 rows of data, it would cause the loading of data to be slow. In order to resolve such issues, our team has decided to segment the data into months.

6 FINAL APPLICATION

Our final application went through a painstaking integration and layout designing process. The main page of our application consists of the sunburst chart and the different bar charts (Fig. 11).

From the left, the first visualization (Left) serves as the main navigation, where users can focus on the types of casualties to analyze on.

On the right, the graphical furniture shows the underlying factors. This will allow the users to study how the various factors in the data contribute to different casualty profiles for different age-groups.

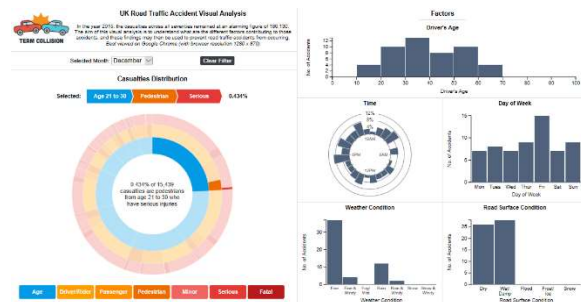


Fig. 11: Application Main Page

The sunburst chart displays the casualties’ distribution and acts as a month filter for the rest of the charts on the right. A user can just click on a bar on the sunburst chart and the rest of the charts on the right will only display the casualties occurred during the month range selected. This is done through the implementation of Crossfilter [5].

The rest of the charts on the right allows the drilldown of the details on the casualties as mentioned in the previous subsection. Once a user selects the sequence within the sunburst chart, the other charts on the right segment will filter and display the desired data accordingly.

While the integration of the various charts proved to be challenging, due to no prior D3.js experience within the team, the utmost challenge lies in developing a visualization which is easy to use and understand. Such challenges include ensuring a scroll-free visualization, color choice within graphical furniture, positioning of the graphical furniture and ensuring that the overall experience is simple and fulfilling. The team feels that this is of utmost importance, as failure to overcome these challenges can potentially affect the visual analytical experience or process of the users.

7 CONCLUSION

Through our application, users are able to explore the specific casualties through sunburst chart, drill down on the casualties’ distribution by filtering the months and understand the underlying factors which contributes to accidents through the bar charts. The interactivity of our application provides users with a more refreshing and effective way of visualizing casualty-related road accidents in the UK. From our application, users are able to examine the underlying factors which contributes to accidents. These information is useful as it helps people to make decision on evaluating the risk of travelling.

Our application provides a general overview of distribution of casualties in the UK occurring each month in 2015. For future improvement, we feel that the scope of our research can be expanded to include a choropleth map to analyze and visualize the location of accidents in UK. This can be displayed through a choropleth map integrated with our current application. Users will be allowed to select a city and view the number of accidents occurred on that city itself. This would be useful in detecting any geo-spatial accident patterns, especially for a big country like UK.

Our team wishes that our research and application will help educate more users and aid them in visualizing the UK road accidents data in a refreshing manner. We would also like to see more application and tools that help to analyze and visualize casualties out on the web as it is crucial to understand what are the different factors that contribute to road traffic accidents, and these understandings may then be used to prevent road traffic accidents from occurring.

Our team wishes to thank Dr. Kam Tin Seong for his kind support and guidance throughout the course of this research.

REFERENCES

- [1] Reported Road Casualties in Great Britain, 2015: <https://www.gov.uk/government/statistics/reported-road-casualties-in-great-britain-main-results-2015>
- [2] Bostock, M. (2012) Data-Driven Document. Available at d3js.org. Accessed 31 Oct. 2012.
- [3] Bostock, M. Ogievetsky, V. and Heer, J. (2011) "D3: Data-Driven Documents" IEEE Trans Vis and Comp Graphics, 17(12), pp. 2301- 2309 Dec 2011.
- [4] Dewar, M (2012) Getting Started with D3. O'Reilly Media Inc. USA.
- [5] Crossfilter: Fast Multidimensional Filtering for Coordinated Views. (2012). Retrieved October 20, 2014, from <http://square.github.io/crossfilter/>

ACKNOWLEDGMENT