Visualization Dashboard – Accidents in London, UK

IAT 814 Final Project Report

Submitted to: Dr. Lyn Bartram
Simon Fraser University

Submitted by :
Jamshed Khan (Student ID 301369381)
Jaideep Misra(Student ID

Overview

London, the capital and largest city of the United Kingdom is considered one of the worl'd most important global cities[1]. London has taken substantial steps to ensure the prosperity of its roads and transportation sector by undertaking transportation planning, traffic managements, road safety projects etc.[2] Traffic collisions are however, a common and unfortunate part of day to day life and last year, London made news headlines for a dramatic increase in collision related deaths and injuries[3] which were reported to be increasing significantly year after year.

The purpose of our course project for IAT 814 is to create a data visualization dashboard that help relevant stakeholders to analyse the raw data pertaining to accidents in London, spot trends and associated factors, and draw conclusions accordingly. The city of london has made available a wide variety of information-rich, flawless datasets with a great many dimensions. Using these as a starting point along with data collected from various other sources such as weather stations, automobile manufacturers etc. we aim to generate suitable data visualizations that help users understand the relations, causes and effects that accidents have within the city of London. *The main goal of our data visualization dashboard is to examine various factors that may possibly have a relationship with accidents and to see the extent of their impact.* A few examples of such factors include Date/Time, weather, road conditions and so on. The audience for this dashboard can be anyone who has an interest in analysing relationships of accidents in London with underlying factors. A few suitable stakeholders may include vehicle insurance companies, police personnel, researchers, government and non-governmental organizations etc.

Introduction

The city of London has approximately 10,000 miles (16000 km) of roads. Each year, thousands of traffic collisions occur on London roads. Such collisions occur when one or more vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstructions, such as a tree, pole or building. Traffic collisions often result in injury, disability, death, and property damage as well as financial costs to both society and the individuals involved. Risk of collision is usually attributed to factors like vehicle design, speed, road design, road environment, driving skills, impairment due to alcohol or drugs and so on. The value of societal harm from motor vehicle crashes includes both economic impacts and valuation for lost quality-of-life. Estimated at around $870.8 billion in 2010 worldwide[4], this value represents lost quality-of-life, as well as economic impacts.

A visual analysis of the road accident data will provide a better understanding if these assumptions are correct and to what degree. It can also go a long way in revealing hidden trends or patterns which may not be obvious by simply looking at staistics but once brought in the appropriate visual format using the data visualization principles learnt throughout this course, such data might speak volumes to the relevant user and highlight potential factors related to traffic accidents within London allowing the stakeholders to make decisions to address them.

To communicate the elemental information clearly and efficiently, our data visualization dashboards makes use of a variety of statistical graphs, plots, maps, filters and other tools. We believe effective data visualization will help users analyze and reason about data by making complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or finding causes and effects. We hope that the design principles we have followed while preparing the dashboard makes such tasks easier for the users.

Description of Underlying Data

The data used to power the visualization dashboard in this project has been gathered from various sources online such as open data archives and other services. A majority of the raw data was obtained from the open data portal by Transport for London, particularly the road safety which includes extracts for road collision data. Transport for London is a local government body responsible for the transport system in Greater London, England. The data they have collected and made available provides information for road traffic collisions that involve personal injury occurring on the public highway reported to the police. Damage only collision are not included and all data belongs to the Greater London area which comprises 32 London boroughs and the City of London. Data is collected by police at the scene of an accident or in some cases reported by a member of the public at a police station, then processed and passed by the police to Transport for London for checking and analysis.

Our approach to data engineering for the purposes of this project was to collect datasets from various sources, each pertaining to one relevant aspect that we felt could pose as an underlying factor that contributes to accidents in London, for example: weather, vehicle details etc. Some of these individual datasets were joined to one another on some common dimension such as date, while others were used independently throughout the process. As such, here is a brief description of the major datasets that were used in this project:

- **Attendant dataset** – This is a general dataset having 27 dimensions or measures and each row represents a separate accident. Every accident is uniquely identified by an accident reference number. This number is the primary key which can then be used to join other datasets from Transport For London. Some of the dimensions are quantitative such as date while others are of a qualitative nature such as condition of road surface(dry, wet) or accident severity(slight, serious). We collected individual datasets for each year and merged them vertically so that the data spans 8 years worth of accidents from 1st Jan 2010 to 31st Dec 2017.
- **Casualty Dataset** – This is a separate dataset which we also merged for the same 8 year period. It details the casualties involved in a particular accident. Given a reference number for a particular accident, this dataset contains one or more rows for casualties (i.e. people) associated with a particular accident. A "casualty" is any human that has suffered an injury stemming from the accident. The injuries can be fatal(death occurs due to injury within 30 days), serious injuries (requiring hospitalization) or slight injuries(not requiring medical treatment eg. cuts, bruises).
- **Vehicle dataset** – This dataset describes details about the vehicles involved within a specific crash and certain related phenomena such as the manoevre that caused the accident to occur or if there was skidding involved, point of impact and so on. As in the previous case, this dataset is much bigger as there may be multiple vehicles associated with a single accident.
- **Weather Dataset** – This dataset was obtained from the Met Office UK from historic data of the London weather station. The relevant dimensions such as temperature, precipitation, snowfall amount etc. was joined to the original dataset on the common date and time, thus allowing us to factor in the weather for a given accident and see if it had any part to play.

The first three datasets can be found here, the Road Safety archive of Transport for London while the weather data can be accessed over here by specifying a given date range (2010-2017 in our case). After extracting and joining the raw datasets, various cleaning, aggregations and transformations had to be performed so as to apply it for our particular use case. Some examples include generation of latitude, longitude from address names so as to plot markers on a map, standardization of datetime and other dimensions, correcting misspelled or missing data, cleaning incorrect data and so on. All of the data used for this project is real and has been a basis for many of our design decisions for the dashboard. In other words, we picked plots and charts that best complement the underlying data, using visualization design principles that we learnt throughout this course.

Visualization Design

We decide to inculcate a dashboard design for our project because we feel that it works best to provide the user with at-a-glance views of the most important measures and key performance indicators that allow them to capture all the most relevant information they might need. It also gives the user a variety of control over the interface and empowers them to perform analyses the way they might choose to. While it is true that the viability of a dashboard is often up to the design decisions behind it, we hope that by inculcating a wide range of controls for the user might make their tasks more comprehensible and easier.

Main page of our visualization dashboard

Our dashboard design follows a "tab" like structure meaning that visualizations are divided into a series of Tabs. Each tab represents a major factor which we feel plays a role in accidents. The four tabs we have on our dashboard are "Date and Time", "Weather", "Road Conditions" and "Vehicle Details". The webpage visualizes the first tab by default which contains 6 plots and a bunch of KPIs. Here are some of the visualizations that we have used throughout our dashboard:

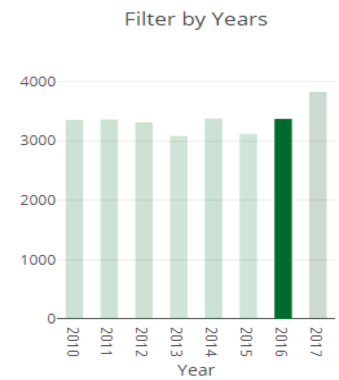
- **KPI Containers** – While it is true that these are not data visualizations but rather just numbers being shown to the user, we feel that these KPIs may be important to people who are looking for a quick glance at information from the dashboard. These indicators are redundant due to users being able to see the same information by using the visualizations themselves but we felt that these containers are familiar to people who use software such as Tableau or Spotfire and they also help if there is no documentation on how to use the visualizations. For example: a user may not know that hovering on a chart would show them accident counts.
- **Bar Charts** – We use bar charts to depict comparisons between discrete categories. The orientation is horizontal or vertical depending on the type of visual effect we want to achieve. For example: when one bar dominates so that the rest are barely visible, we use the horizontal orientation so as to utilize more space and allow greater visibility. We have also made use of clustered bar charts when necessary such as showing distributions based on severity.
- **Line Charts** – We have used line charts having one or more lines to depict categorical changes between variables. For example: a plot showcasing accidents in different months where lines of different colors represent accident severity.
- **Heatmap** – We made use of heatmap as visualization technique that shows magnitude of a phenomenon as color in two dimensions. The variation in hue (intensity), gives visual cues to the a user about how the dimensions varies over space. For example: we have a heatmap depicting casualty counts across the dimensions of weekdays and hour of the day by intensity of color.
- **Sunburst Chart** – We use sunbursts which can be though of as a multiple-level pie chart to visualize hierarchical data structures. It gives a good understanding in portions of a category while drilling down across dimensions. While it appears to be visually confusing on the outer rims of the chart, the interactivity of the chart helps in solving this as at each level, clicking on a category causes previous parents to abate and shows the individual percentage proportions of its children.
- **Map Visualization** – A map visualization as in our dashboard helps users identify the exact geographical of each accident on a map of London. Accident sites are depicted by markers and the markers are colored differently according to their severity(slight,serious,fatal). The map theme is light so that the natural map colors don't get in the way and confuse the user. However, roads and streets are clearly highlighted so that users can follow the major routes/intersections/junctions where accidents have occurred.
- **Scatter and Bubble Plots** – Several scatter plots are used throughout our dashboard where relationships between two categories need to be shown by markers. In one case, where another dimension was necessary to be depicted, the size of the marker was used to do so resulting in a bubble plot.

All of the visualizations are interactive. Brushing and linking is supported throughout our dashboard and on each particular tab, every visualization chart is linked to every other chart. As such, making a change such as a click or a selection by dragging will result in a visual change in that as well as every other chart on the tab as well as the KPI containers according to what is to be accompished. Besides, the charts support other features such as hovering to show details or zooming. The interactions are further elaborated on in the next section.

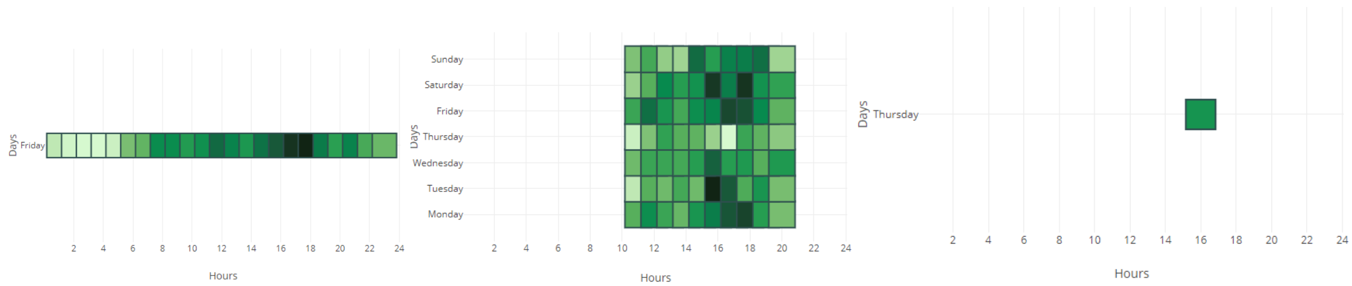
Usage of Interactions and Controls

Date and Time Tab (Main Page)

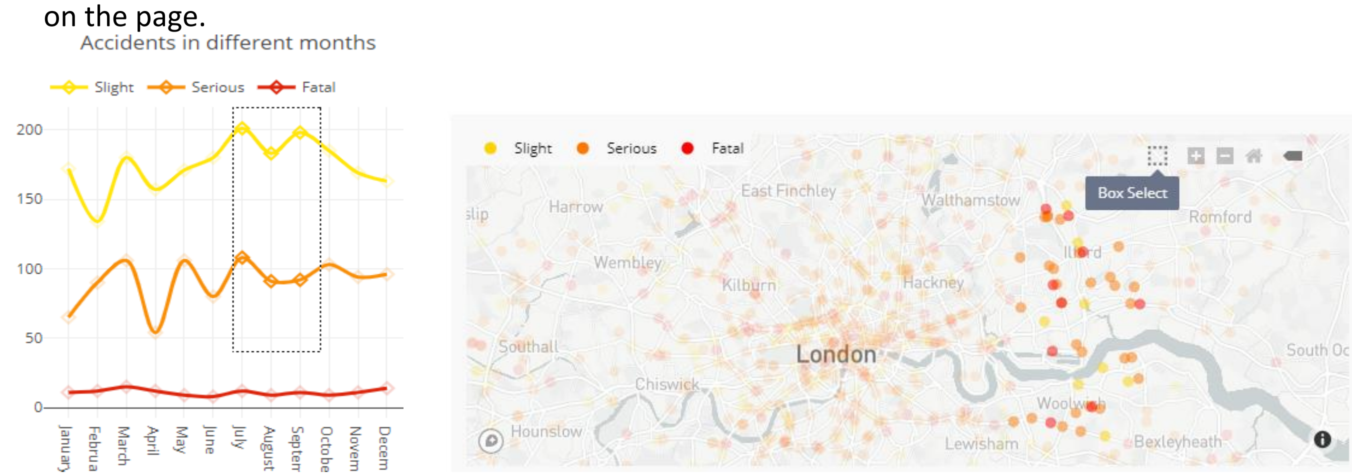
1. <u>Filter by Severity </u>: These are three checkboxes that are all checked by default: fatal, slight and Serious. Unchecking or checking them removes or adds accident incidents of those severity.



2. <u>Filter by Year</u> : This is a simple bar graph with height of bar corresponding to counts of accidents and each bar represents a year from 2010 to 2017. By clicking on a particular bar, the data is filtered out to reflect only the corresponding year and the changes will be reflected across all charts and KPI containers. The selected bar will get highlighted.
3. <u>Filter by Day and Time</u> : This interaction can be performed by dragging the cursor to make selections on the heatmap visualization. Selecting one or more weekday-hour tiles will cause data to be filtered out and reflected in the other charts, map and KPIs. The visualization supports both click and selection. Thus clicking on a single tile will select only that value while selecting multiple tiles by the mouse dragging action will select more. Some examples of when selections are made are shown as follows:



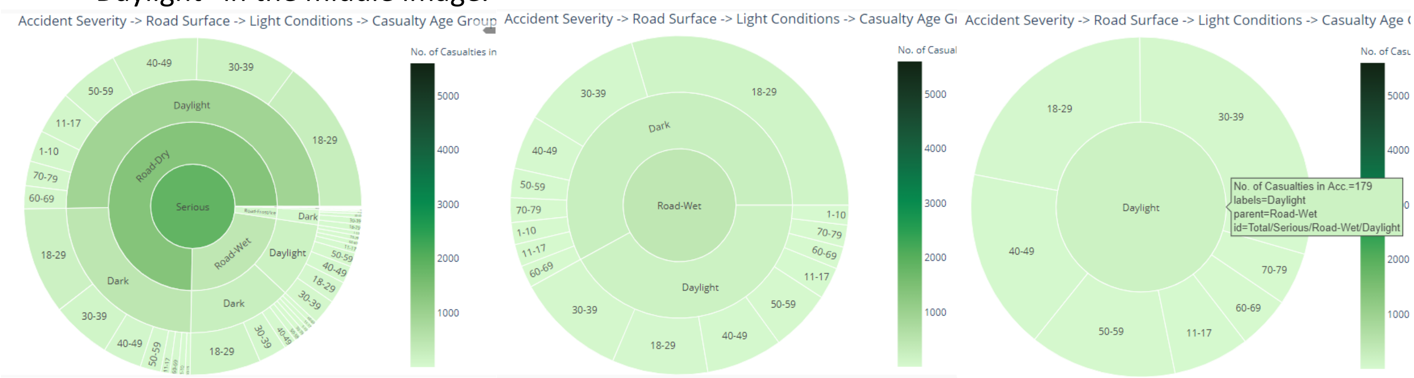
4. <u>Filter by Month</u> : This interaction can be performed by the line chart which shows accident rates by month. We can use the mouse to drag a box selection around one or more months to select data for only those months. The selection may be made for points across one, two or all three lines. Upon selection, the line chart will be transformed to reflect only the selected months as will all other charts on the page.



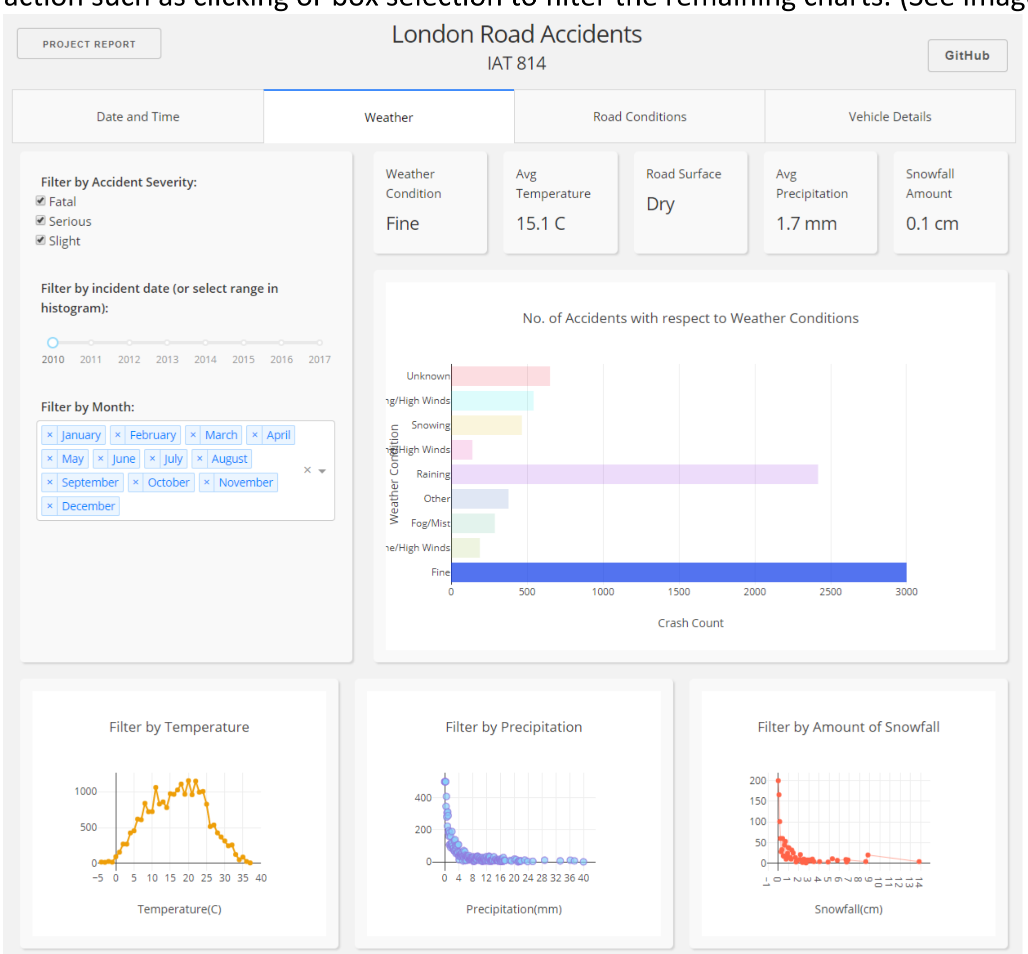
Filtering by month – box selection made on months to filter(left). Location filtering – selected locations highlighted upon selection on map(Right)

5. <u>Filter by Location</u> : We can also filter the data based on location. Simply choose the "box select" option at the top right of the map visualization(shown in the image above) and then use the mouse to drag a box across desired markers on the map. The selected markers will then be highlighted while others will fade and the remaining charts and KPIs will update accordingly.

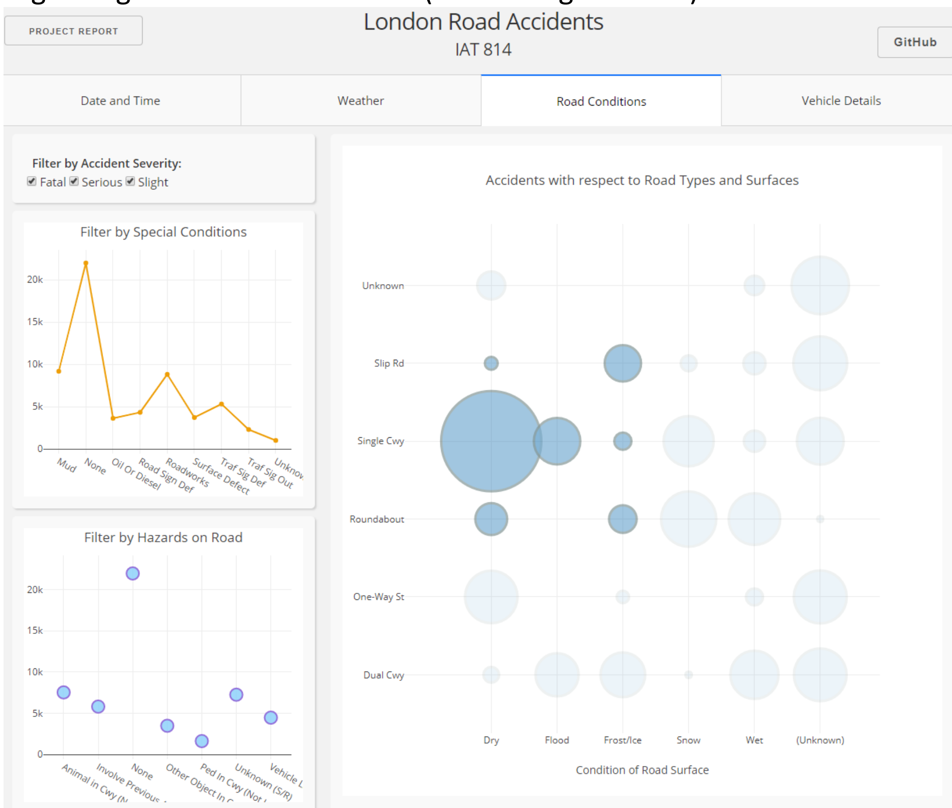
6. Drill down for Sunburst Chart – This is supported only in the Sunburst plot. The drilling down feature requires clicking on a sub-level of the sunburst chart. This can be demonstrated through an example of the images shown below. The chart as it appears originally is pictured on the main dashboard image above. The image in the left is the result when the user clicks on “Serious” in the 2nd innermost circle (for reference, innermost circle says “Total”). The image in the middle is the result when the user clicks on “Road wet” in the first image. while the image on the right is what happens when the user clicks on “Daylight” in the middle image.

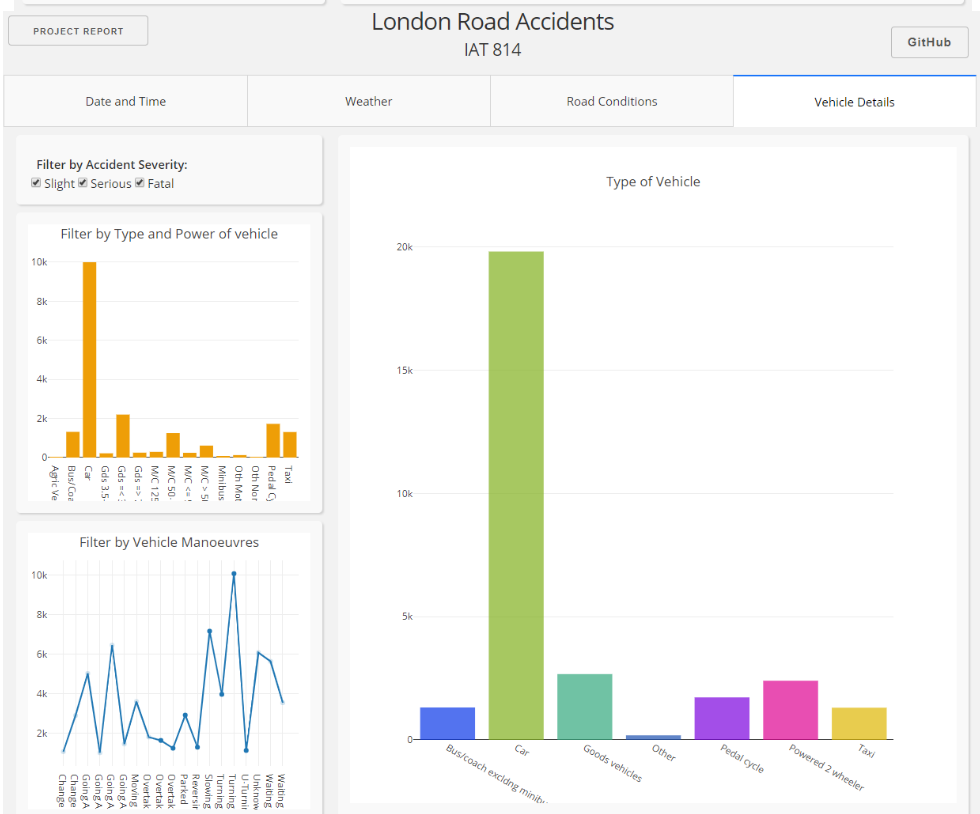


7. Controls for Weather – The weather tab consists of 4 charts, KPI containers and a fiter controls panel. The KPI containers depict information that can be changed as per the controls in the panel. These consists a slider that changes the year and a dropbox containing names of months as tiles. The KPIs contain information such as Average temperature, precipitation etc. For example, if one keeps the month of July in the month filter(dropbox), the averages would reflect that. The 4 charts are not connected to these controls and are instead interconnected to one another, so that the user can perform action such as clicking or box selection to filter the remaining charts. (See image below)



8. <u>Controls for remaining tabs</u> – The last two tabs attempt to depict causal relations between road conditions and Vehicle details respectively. Each representation consists of three graphs each. In both cases the plots on the page can be filtered by accident severity using a checklist provided on that page. The three plots also support linking and are interconnected amongst themselves so that making selections by clicking or dragging a box around items will highlight those values and cause corresponding changes in the other charts. (Refer images below)





Further Work

Our original design was somewhat different to the dashboard currently hosted online. For one, we had a separate control panel on the left hand side of our dashboard on the main page. This control panel was similar to the one present on the weather tab page. It consisted a slider for controlling the year, a dropbox menu that contained tiles for each month, a range slider for choosing hours and several other controls. By changing these controls, we could change the appearances of our plots and these were meant to be a way to filter the data. However, after a project feedback session with the course instructor and TA, we realized that a simpler and much more efficient way of filtering our data was by brushing and linking the plots themselves so that making selections in plots for year, month, day-time and so on we could achieve the desired results and also save valuable dashboard space that was being occupied by the filter control panel.

Although we managed to inculcate the necessary changes, there are still some alterations and additions that we are looking to make to our current dashboard design. A few of these are elaborated as follows:

- Adding interactivity through charts to the KPI containers in the Weather Tab – While the charts on the weather tab are linked to each other so that they can be filtered by brushing each other, the original design only supported filtering through controls on the filter panel. While we made updates to the chart, we did not have time to make the same updates to the KPI containers at the top and they can only be controlled by the filter panel. We intend to change these containers to display relevant information such as accident counts which can then be filtered by plot such as temperature.
- Filtering by Speed limits – There is a plot on our main page that shows accidents that happened based on speed restrictions. Currently it is the only plot which cannot be brushed to filter the other graphs on that page. We shall add this functionality to it also.
- Multiple map configurations – Currently our map visualization is a geospatial configuration where accident sites are visualized by markers on a base map layer. We wish to also add more configurations which the user can then choose. We are working on a chloropleth map since our dataset has a dimension for borough(like a municipality) within the city. The city of london can be divided into these boroughs then and the color intensity could be used to show accident rates by borough. We are also working on a density heatmap so as to use color intensity to visualize accident counts geospatially. The user can switch between these configurations using an option to do so on the map visualization.

Conclusion

We have utilized the principles of good visualization design that we learned throughout the course of IAT 814 to prepare a dashboard that can be help the relevant audience to analyse accidents in London and ask important questions such as :

“What are the factors involved in the origin of accidents that happen in London?”

“Are there any causal relationships between Road conditions, weather or date/time and accidents in London?”

“What are the effects of accidents in London on certain age groups ?”

Our dashboard will help individuals to gain insights about these and many more questions. It will allow them to access relevant and important information at a glance and perform desired analytical tasks. Hidden patterns, trends and relationships that are not inherently observable have a better chance of being highlighted through our visualization dashboard.

References

[1]"Global Power City Index 2017". Institute for Urban Strategies – The Mori Memorial Foundation. Retrieved 23 November 2018.
[2] https://www.london.ca/residents/Roads-Transportation/Pages/default.aspx
[3] https://www.standard.co.uk/news/london/1220-killed-or-seriously-injured-in-london-road-accidents-over-three-months-a4069436.html
[4] https://en.wikipedia.org/wiki/Traffic_collision#Economic_costs

We would also like to cite the Traffic Accidents in UK project by Richard Muir which served as a brainstorm to our own dashboard and also the Dash App Gallery from where we obtained some of the styles that we used in our dashboard design.