FIT5147 Data Exploration and Visualisation S1-2022

DATA VISUALISATION REPORT ON:

Analysis of the frequency of Road Crash in the ACT.

Prachi Jaiswal, 32192673

Tutorial 01
Tutors: Mohit Gupta and Abishek Sriramulu

TABLE OF CONTENTS

1. Aim

The purpose of the visualisation is to examine the regularity of road crashes and underlying patterns over time, with specific attention to the suburbs of Canberra, Australia.

2. Motivation

Driving is an essential part of life in a country as vast as Australia. The necessity of driving here cannot be overstated. The government have enforced several strict and comprehensive road safety regulations in order to ensure the safety of Australians in road accidents. Consequently, the number of road accidents has noticeably declined in the past few years in part because of such directives. I, therefore, sought to learn more by investigating how crashes have decreased over time as a result of the following facts. In order to understand what causes these collisions, I conducted a study on the frequency of road crashes.

3. Proposed Questions

3.1. The trend and pattern of road accidents in the ACT (Canberra) in the past decade.

3.2. Identifying the most accident-prone suburbs/areas, as well as the factors/conditions that contribute to the highest number of collisions.

3.3. An examination of how frequently traffic accidents occur in different situations/elements time of the day.

4. Design

This section discusses the five design sheets in detail, as well as the ideas employed in each of the sheets. Along with that, I will discuss the final design sheet utilised in the visualisation project. This section also discusses the rationale for the addition of other aspects of interactions for graphs.

4.1. Design Sheet 1

The first design sheet depicts the potential graphs' ideas, filters, combinations, and refining for the final visualisation project. A total of 12 concepts were produced which are included in the appendix (image 1). A few concepts were chosen from a slew of others. The density graph is one of them, while the histogram is another. They both represent the distribution

of a numeric value. Furthermore, a few graph configurations were designed to address the questions while also making the visualisation presentable. The sheets that follow (Sheets 2, 3, and 4) illustrate combinations as well as their positive and negative aspects for each.

4.2. Design Sheet 2

Sheet 2 is a combination of three visualisations: A line graph for yearly accidents, a Heat-map for hourly accidents and a sankey diagram that shows crash severity for different types of conditions that cause road crashes. The goal is to visualise the frequency of road accidents for specific years on the line graph and the hourly analysis of collision severity on the heat map. The Sankey diagram depicts the flow of a process, encompassing all conceivable scenarios for accidents, as well as the severity of the crash. The thicker the line or arrow, the greater the quantity of energy involved.

The sheet contains advantages such as a dropdown menu for analysing the pattern for each area, however, the user can only choose one suburb for the analysis. For a better understanding of the collision severity, the sheet includes additional operating tools such as a slider input for changing the year and radio buttons for a better grasp of knowledge about the accidents. The sheet answers almost all questions, however, it fails to display the geographical locations for a better understanding.

4.3. Design Sheet 3

Sheet 3 has a set of three visualisations: A bar graph for hourly analysis of accidents, a Box plot for accident's yearly pattern and a pie chart that shows crash severity for different types of conditions that cause road accidents. The visualisation is not suitable for a proper representation to answer the proposed questions, as the piechart fails to explain key assumptions, causes, effects, or patterns about the data. The boxplot explanation can be overwhelming as such detailed information is not needed for the analysis.

The sheet does not contain a dropdown menu so the user won't filter out the suburbs and the representation will be clustered. For a better understanding of the collision severity, the sheet includes additional operating tools such as a slider input for changing the year and radio buttons for a better grasp of knowledge about the accidents.

4.4. Design Sheet 4

Sheet 4 includes again a combination of three visualisations: A Lollipop chart for hourly analysis of accidents, a Violin plot for accident's yearly pattern and an alternative to a pie chart known as a doughnut chart. The doughnut chart shows crash severity for different types of conditions that cause road accidents. The visualisation is still better for visualization but not as good as a Sankey diagram to explain key assumptions, causes, effects, or patterns about the data. The boxplot explanation can be overwhelming as such detailed information is not needed for the analysis.

The sheet contains a dropdown menu but the user can only filter out the one suburb at a time. For a better understanding of the yearly trend, the sheet does not include operating tools like a slider input. The sheet also uses radio buttons for a better grasp of knowledge about the accidents.

4.5. Design Sheet 5 (Final Design)

The final design is a refinement of sheet 2 which consists of a Sankey Chart, Heat Map and Line Graph which are chosen as final designs along with a Choloropeth map to aid in understanding the relationship between the conditions affecting the road crashes at different hours of the day. Because both variables (hour and year) cannot be presented on the same axis in a graph, they were separated into two visualisations with some user-operated tools which resulted in consistency in design and interaction. This will be done with the help of HTML, CSS and Javascript D3 library.

5. Implementation

The model was developed using D3, JavaScript, HTML, and CSS. A little preprocessing was done to make the implementation easier. The first visualisation, a map, makes use of an act.json file to draw precise coordinates. In addition, I retrieved the year from the dataset's date column in order to construct a slider input for user interactivity. In addition, for the suburb selection, I filtered out unique locations from the data. In addition, I included radio button features that allow the user to pick different situations that cause traffic collisions in the ACT. These methods made the data set easier when designing my final sheet in the D3 library.

5.1. Technology Stack

Stack: HTML, CSS and Javascript

Library: D3

Font: Roboto

The Roboto font from the google API was also used for the graphs. Also, the Sankey package from the D3 library was used to create the Sankey chart. Some of the implementations done on the D3 library were used as a reference to build the charts.

5.2. UI Interface

The user interface includes four interactive visualisations answering almost every question proposed during the proposal.

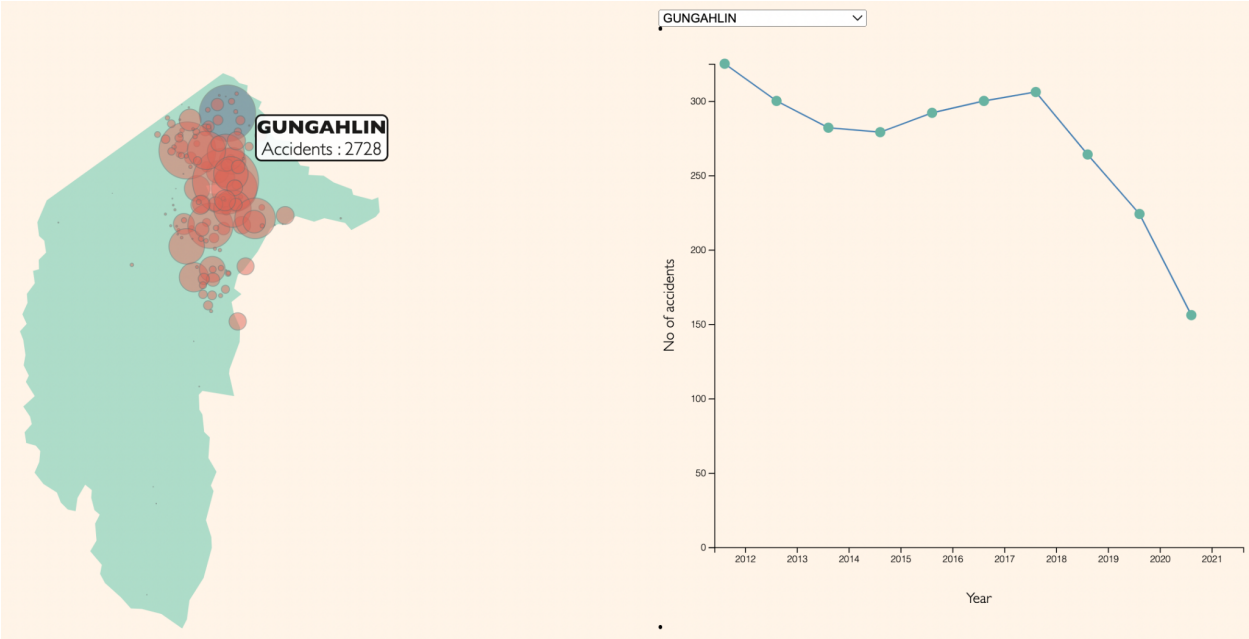


Figure 1: The figure shows the interaction between the Canberra map and the line graph.

Figure 1 illustrates the initial portion of the website, in which the map (left) shows the geographical location of ACT (Canberra), and the size of the bubbles denotes the number of accidents in each suburb. Because it needed a geojson file, the map was one of the more challenging things to implement in the project. The line graph, on the other hand, was quite simple to create, but the interaction between both visualisations took some time for me to develop. The figure also depicts a dropdown menu that allows the user to filter the suburb and obtain a clear analysis for that specific region for this section of the page.

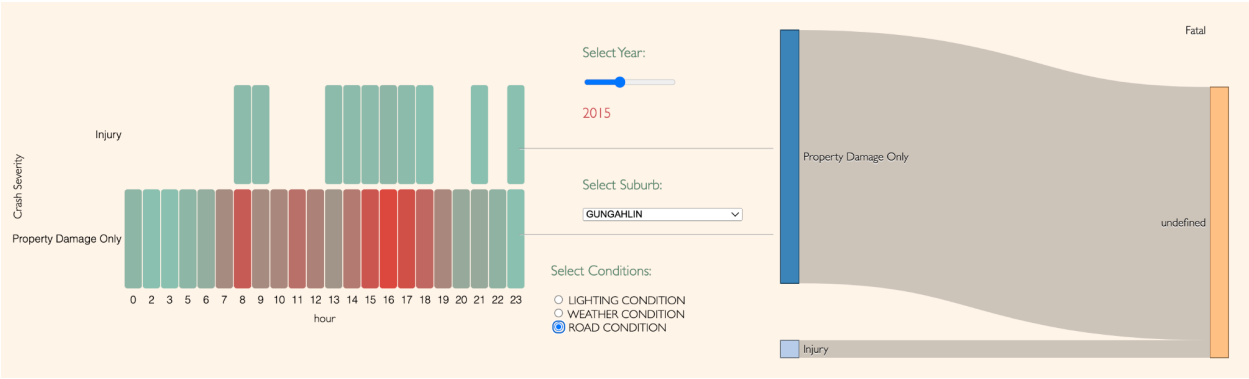


Figure 2: The figure shows the interaction between the heat map and the Sankey diagram

Figure 2 explains the hourly analysis and crash severity with the help of a heat map and a sankey diagram explains the crash severity against three other conditions displayed on the radio button. This section also has a drop-down menu similar to what is provided in Figure

1, as well as a slider input so that the user can analyse the difference in the pattern of car crashes from 2012 to 2021.

Overall, the project was pretty challenging to complete due to a few problems such as the structuring of the homepage and aligning every graph and text so that the page looks appropriate. However, I am unable to adjust the zoom ratio for the screen to create an acceptable representation. Another issue I encountered was the Sankey diagram, which was not as appealing as it could have been. Finally, the Canberra map took me a while to grasp and apply.

(Holtz (2022) assisted me in filtering the graphs because it was extremely challenging for me to distinguish between various ideas. (Using D3.js to Plot an Interactive Map," 2022), was a really helpful blog for helping me out with map interactivity. Furthermore, I was able to offer adequate beautifying and alignment of the project by using other sources such as ("D3.js", 2022) and a few StackOverflow responses.

6. User Guide

Follow these steps to get the most out of the visualisations:

- Navigate to the folder containing all of the source code.
- Launch index.html directly or load it on the live server (After installing the appropriate VS Code extension)
- It is advised that you open the html file with the Google Chrome web browser.
- Since the website is scrollable, scroll down to see the second half of the page.
- **Visualisation 1:**
  - Hover over the map to view the geographical location of each Canberra suburb.
  - Hover over the Line Graph to view the trend in the number of accidents in each Canberra suburb.
  - The map and line graphs are interactive and change according to user inputs.
  - Both visualisations change according to the selection of the suburb from the drop-down menu.
  - The line graph also changes according to the selection of the location on the map.
- **Visualisation 2:**
  - Hover over the Heat Map to view an hourly breakdown of the frequency of Canberra road accidents.
  - Hover over the Sankey Chart to see how catastrophic the accidents were in the suburbs under different circumstances.

- The heat map and Sankey diagram are interactive and change according to user inputs.
- The user can choose different years ranging from 2012 to 2020 from the slider input and both the graphs change accordingly.
- The user can choose only one suburb location from the drop-down menu and both the graphs change accordingly.
- The last user input only works for the sankey diagram which shows the crash severity and different conditions chosen from the radio buttons.

7. Conclusion

All of the questionnaires provided in the introduction were answered by the visualisation project. We can observe how the number of accidents in each year and each suburb has decreased over time unit the pandemic hit in the year 2020. Furthermore, the interactive chloropeth map outlines all of the geographical locations of road crashes and accident hotspots. It also suggests that none of the extreme weather conditions played a significant effect in Canberra road crashes since the roads are dry and the weather is pleasant all through the day.

Contrary to the belief that night driving leads to road accidents, it was observed that most accidents happen during the day and on good surface roads as none of the extreme weather conditions plays a significant influence on Canberra road crashes even when the weather is fine during the day. The heat-map shows that the highest number of accidents occur during the office hustle, when people rush to go or depart from work, particularly in the city, where the most businesses are centred. However, accidents do occur mostly during the rainy season which is in May.

The project introduced me to the javascript library d3 and the ideas of five design sheets. It allowed me to experiment with various visualisation platforms besides R Shiny and helped me produce intriguing and visually appealing outcomes. The data collection I picked made it difficult to put the spots on the map in d3 in a good representation prohibiting the data from being loaded into HTML through the server.

Additionally, the main restriction of this visualisation is that it takes a bit longer to load because the dataset contains around 71k traffic accidents over the previous decade. The visualisation presently displays the crash sites as points, however, I could have investigated the direction of the crashes while owing to a large number of suburbs and locations in the dataset, this would have been difficult to make presentable. Nevertheless,

the entire process of developing and implementing a visualisation has been quite interesting and has boosted my understanding of this area.

Finally, to conclude, it's worth emphasising that the government is doing its job effectively by enforcing strict road safety laws and regulations, which has led to a substantial decrease in the general trend of accidents from 2012 to 2021 keeping the people in Canberra safe.

8. Bibliography

I. Road Safety. Retrieved 1 JUNE 2022, https://www.infrastructure.gov.au/roads/safety/

II. Connelly, & Supangan. (2016, November 6). Science Science Direct. Retrieved June 2, 2022, https://www.sciencedirect.com/science/article/abs/pii/S0001457506000649

III. ACT Road Crash Data. (2021). Retrieved 1 JUNE 2021, from https://www.data.act.gov.au/Transport/ACT-Road-Crash-Data/6jn4-m8rx

IV. Holtz, Y. (2022). The D3 Graph Gallery – Simple charts made in d3.js. Retrieved 29 May 2022, from https://d3-graph-gallery.com/

V. Using D3.js to Plot an Interactive Map. (2022). Retrieved 1 May 2022, from https://medium.com/@ivan.ha/using-d3-js-to-plot-an-interactive-map-34fbea76bd78

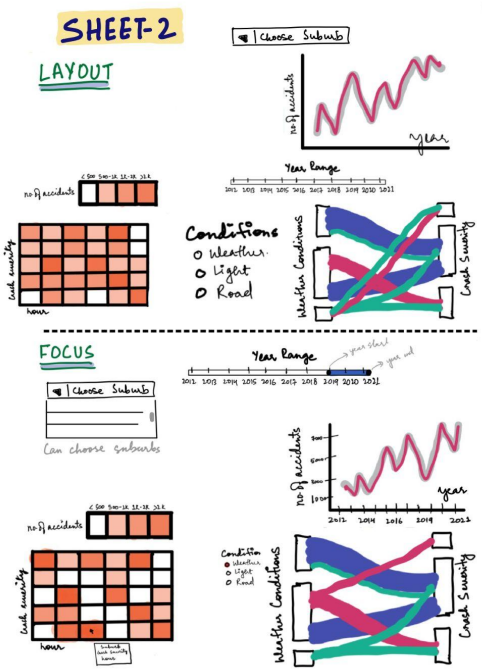
VI. D3.js. (2022). Retrieved 3 June 2022, from https://www.w3schools.com/ai/ai_d3js.asp

VII. D3.js - SVG Transformation. (2022). Retrieved 1 June 2022, from https://www.tutorialspoint.com/d3js/d3js_svg_transformation.html

VIII. Car Accident Statistics 2020 | Car Research & Statistics — Budget Direct™. (2020). Retrieved 2 JUNE 2021, from https://www.budgetdirect.com.au/car-insurance/research/car-accidentstatistics.html

9. Appendix

Design Sheets:



Title Frequency of Road Accidents in ACT
Submitted by Prachi Jaiswal
Date 16-05-2022
Sheet 2
Task Analyse trends of road accidents in ACT

Operations

→ The drop down menu shows the list of accident prone suburbs in ACT. The operation also lets the user select upto ten suburbs.

→ The slider input shows the year of accidents given in the data. The input lets the user select a range of years for the analysis.

→ The tooltip will show when the user hover on any location on the chloropeth map. The information it provides is the location name and the no. of accidents.

Discussion

Positive

→ Have sufficient operation tools for a proper user understanding.

→ Hovering will let the user know the no. of accidents for the given location.

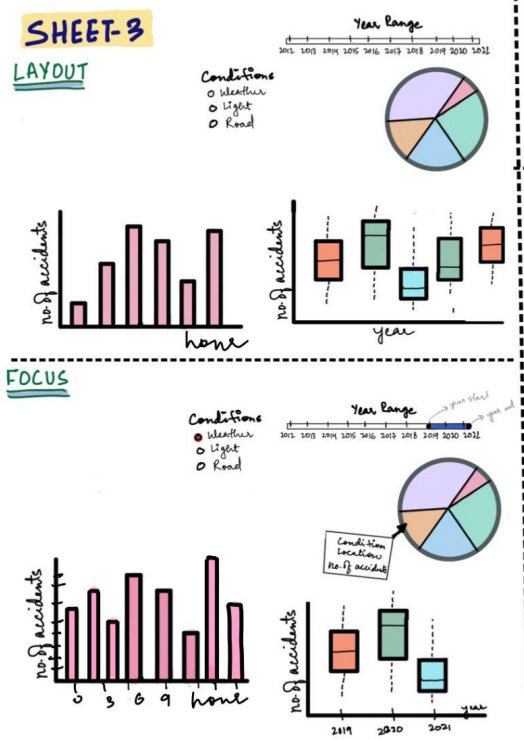
→ The visualizations won't be cluttered

→ All graphs are interactive

Negative

→ The user won't see any geographical representation of suburbs.

→ The user can choose only one suburb



Title Frequency of Road Accidents in ACT
Submitted by Prachi Jaiswal
Date 16-05-2022
Sheet 3
Task Analyse trends of road accidents in ACT

Operations

→ The radiobuttons display the conditions that may cause road accidents in ACT. The user can select only one condition.

→ The slider input lets the user select a range of years when the road accidents occured in ACT.

→ The tooltip shows the user information of the location, number of accidents and the condition

Discussion

Positive

→ The user can select any of the three conditions and directly analyse it with the accident severity.

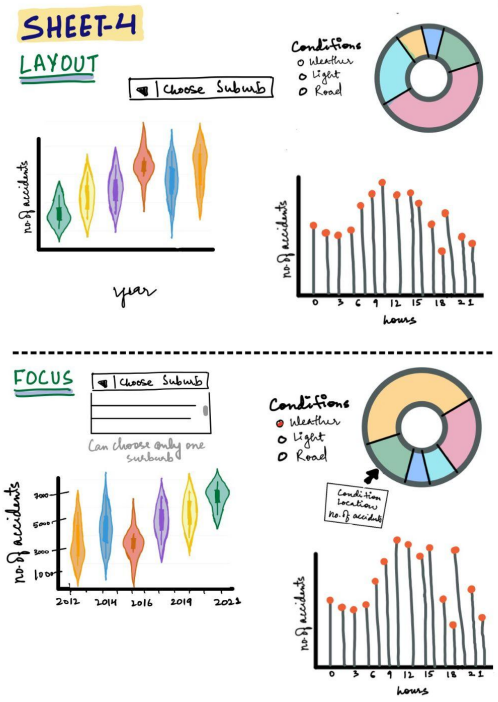
→ The user can filter out the range of year for a better and specific analysis.

Negative

→ The user won't see any geographical representation of the suburbs.

→ The user don't have any option to choose suburbs.

→ The representation will be overwhelming.



Title Frqueney of Road Accidents in ACT
Submitted by Prachi Jaiswal
Date 16-05-2022
Sheet 4
Task Analyse trends of road accidents in ACT

Operations

→ The dropdown list will provide all accident prone suburbs where the user can only select one suburb for analysing hourly crash severity.

→ The tooltip operation will display the name of the location along with time of the day and the crash severity on the heat map.

Discussion

Positive

→ The tooltip will help user to view the cells when honered on.

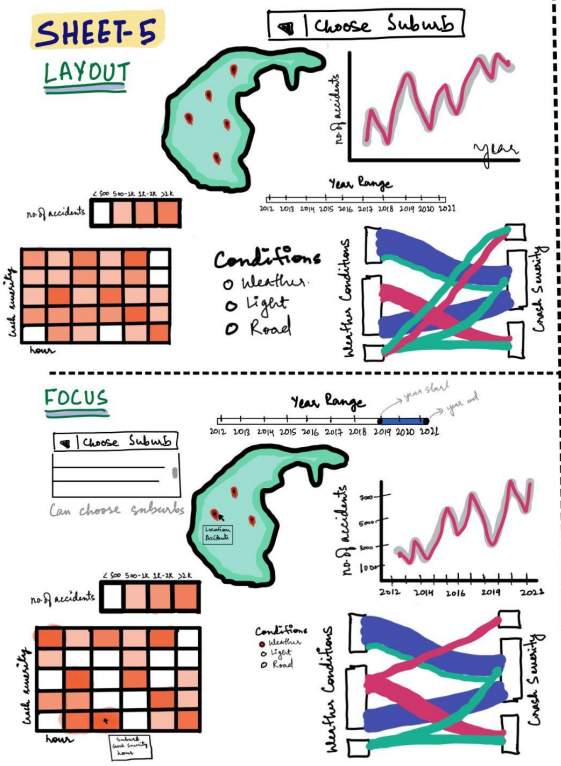
→ The violin plot shows in depth accidents information for the year variable.

Negative

→ The visulisation does not show geographical locations of suburb.

→ The user can only choose one suburb at a time for analysis

→ The user won't understad the violin plot if they are not statistically sound.



Title Frqueney of Road Accidents in ACT
Submitted by Prachi Jaiswal
Date 16-05-2022
Sheet 5
Task Analyse trends of road accidents in ACT

Operations

→ The dropdown menu shows the list of accident prone suburbs. It lets the user select suburbs.

→ The slider input shows the year of accidents given in the data. The input lets the user select a range of years

→ The radio buttons displays the conditions that may cause road accidents in ACT. The user can choose only one condition at a time.

→ Tooltip operation will display the location name along with the no. of accidents when the user honers on the chloropeth map.

Detail

Dependencies → HTML, CSS, D3

Dataset

The dataset is taken from ACT government open data portal which provides the geographical data (latitude and longitude points) of road accidents in the various suburbs of ACT. The data is provided for the years 2012 to 2021. The data also provides accident's condition and crash severity.

Estimate

Cost : NIL

Basic map of suburbs → 17-05-2022
Add filters and interactions → 18-05-2022
Design Testing → 19-05-2022