FIT 5147 Data Exploration & Visualisation

Data Visualisation Project

Correlation between Rainfall and Aircraft Movements in 8 major Australian Airports

Tan Kah Wang

2944 2826

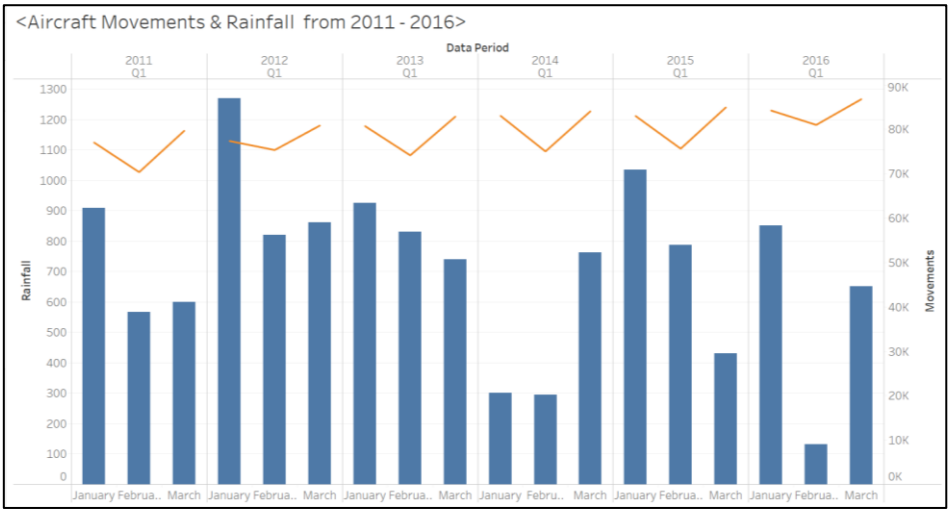
Submission Date – 31/05/2018

Table of Contents

1. Introduction

The purpose of this data visualisation project was to identify findings from the previous data exploration project that I hoped to communicate out to my chosen target audience. A narrative visualisation in the form of a web-based presentation using R or D3, should be implemented in an attempt to aid the communication of key messages to the intended audience.

The topic of my project was studying the possible correlation between rainfall and aircraft movements in 8 major Australian airports, namely Adelaide, Brisbane, Canberra, Gold Coast, Melbourne, Perth Sydney and Tasmania from the data period of January 2011 to December 2016. The key message that I hope to convey to the target audience is there was a noticeable pattern *(refer to Figure A)* between rainfall and aircraft movements in the month of February consistently across the years from 2011 to 2016. A drop in rainfall coincidentally saw a drop in aircraft movements in the month of February.



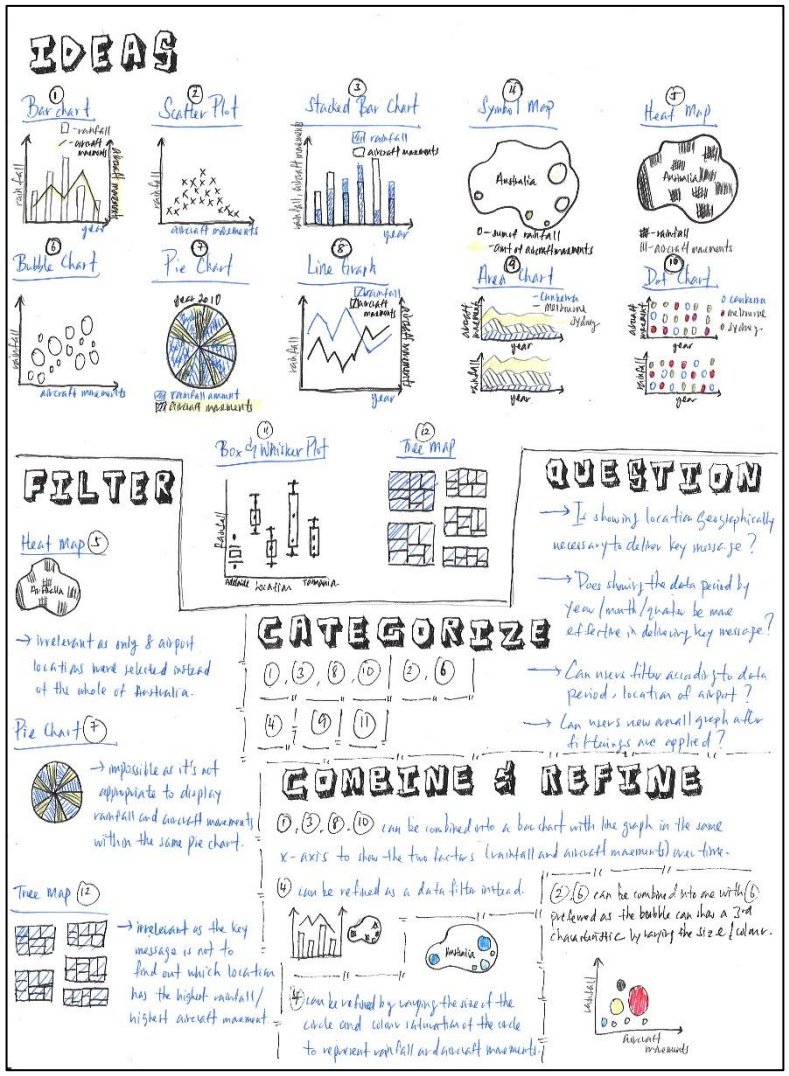
(Figure A – Tableau Bar & Line Chart of Aircrafts Movement & Rainfall from 2011 to 2016)

The intended target audience for communicating this key message will be Australian Government Department of Infrastructure, Regional Development and Cities. Main roles of this government body includes supporting economic growth through transport and increasing transport access[1]. By communicating the key message to them, it may potentially lead to more flights and routes planned out in the month of February when it was evident that there was clearly a drop in aircraft movements in the month of February across 2011 to 2016, with a coincident drop in rainfall as well. This is in line with the responsibilites of the government data as it will will to greater flight access to the general public and in turn, boost the economic growth with increased revenues. Furthermore, the dataset[2] containing the amount of aircraft movements originated from the same government body and by presenting analysis of their own dataset, it may lead to greater credibility of the data resulting in stronger communication of the key message.

2. Design

i. 5 Design Sheets Methodology – Sheet 1

A total of 12 different graphs/sketches were generated in the “IDEAS” portion and with the key message in mind, irrelevant graphs/sketches were removed in the “FILTER” section. Heat Map was deemed irrelevant as the dataset for this project did not cover the whole of Australia and only 8 airport locations were selected. Pie chart was also crossed out as it was inappropriate to represent two data factors (Rainfall and Aircraft Movements) in one pie chart. Tree Map was removed as well as the key message was not to compare the highest/lowest rainfall/aircraft movements. The remaining ideas were then categorized into five main categories as shown in the “CATEGORIZE” section and they were then combined into three possible designs namely a combined bar and line chart, a symbol map and a bubble chart that will be discussed in the next 3 sheets.



(Figure B – Five Design Sheet: Sheet 1)

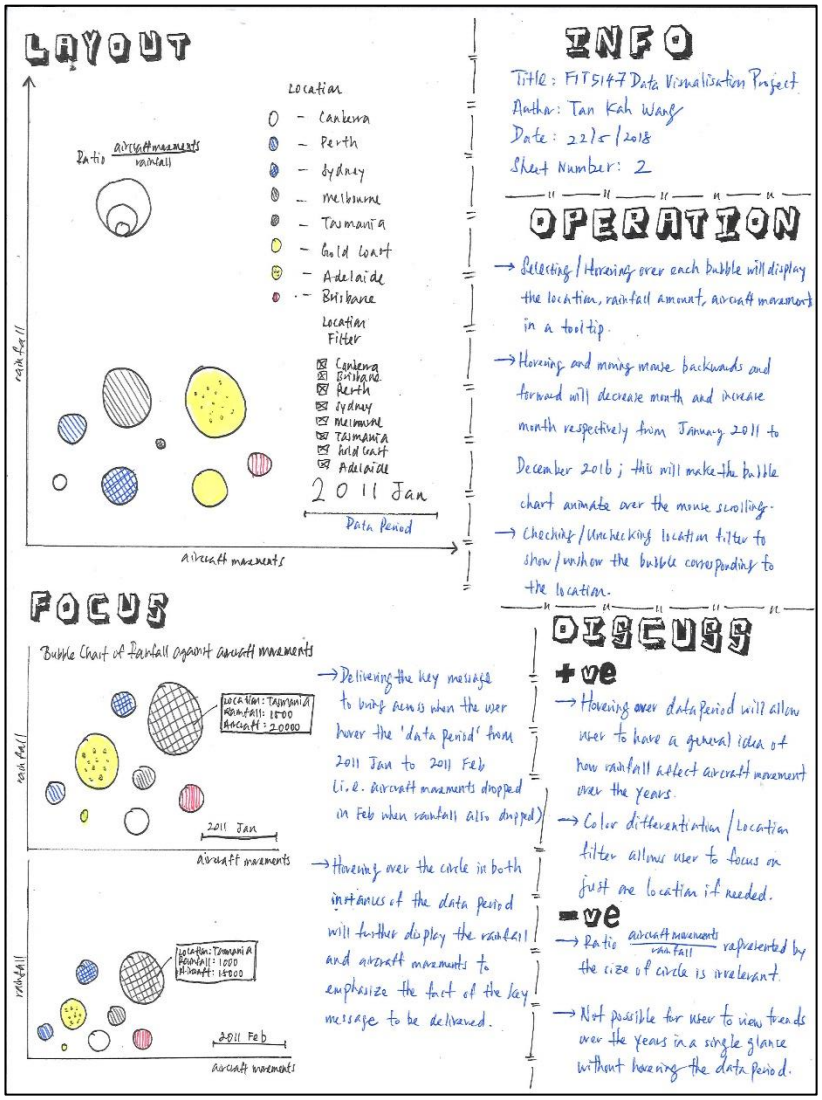
While coming out with possible ideas for the narrative visualisation to be able to communicate the key message to the target audience, some questions were asked and noted down:

- Is showing location geographically necessary to deliver the key message?
- Does showing the data period by year/month/quarter be more effective in delivering the key message?
- Can users filter according to data period, location of airport?
- Can users view overall graph after filtering have been applied?

ii. 5 Design Sheets Methodology – Sheet 2

The first design being considered was a bubble chart. Using this visualisation, the key message will be communicated when the user drags the data period filter horizontally which will make the bubble chart change over time. In this way, when the user drags from January to February, the position of the circles will be lower as there were both a lower rainfall and aircraft movements in February compared to January. This will allow the user to have a general idea of how rainfall affect aircraft movements and with the different colours representation of the bubbles based on the location, users can easily identify possible trends in different locations with the help of the location filter as well. Finally, a tooltip will be displayed when users hover over the bubbles showing the location, rainfall amount and total aircraft movements.

This design was not chosen as the final design as since there were only two value-based data factors namely aircraft movements and rainfall, there was no pre-existing third data factor that can be shown together with the width of the circle. Thus, a ratio between these two factors needed to be computed to represent the width of the bubbles which I thought was irrelevant and provide little context in communicating the key message. Furthermore, axes of a bubble chart are best represented by two value-based data and since our dataset and key message include a time data factor, it can only be included in this visualisation as a filter which makes it hard for users to identify the key message over the years in a single glance without dragging horizontally on the data period filter.

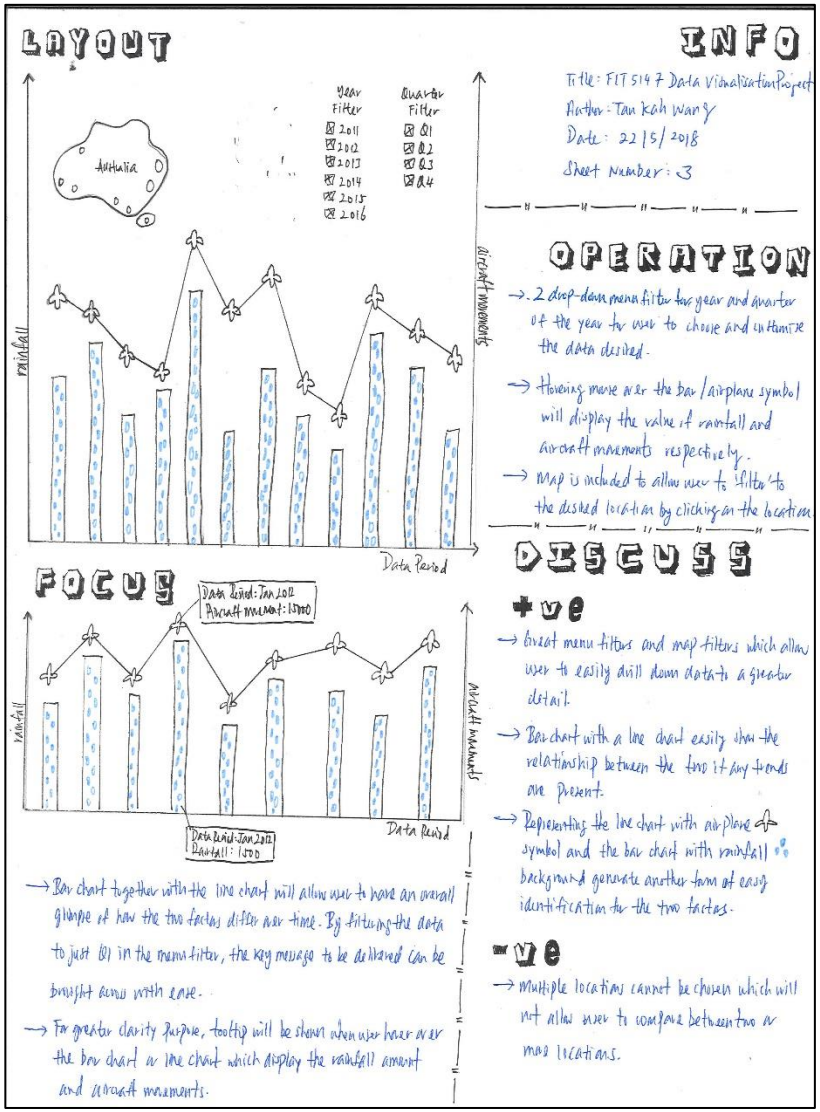


(Figure C – Five Design Sheet: Sheet 2)

iii. 5 Design Sheets Methodology – Sheet 3

The second design being considered was a combination of bar chart and a line chart. Through this visualisation, it will allow users to have an overall glimpse of how the two data factors (Rainfall and Aircraft Movements) differed over time. And with the additional users' interaction of filtering the data to the desired year and quarter, it will be easier for users to focus on just quarter one of the data periods which will thus lead to more effective communication of the key message. Furthermore, tooltip will be displayed when users hover over the bar charts or the airplane symbol on the line chart. A third users' interaction filter in the form of a map was included for users to filter between the 8 different locations, while at the same time, identify the key message geographically. Representing the bar charts with rainfall as background and the line chart with an aircraft symbol will generate another form of easy identification between the two data factors.

The first issue with this visualisation was users will not be allowed to choose multiple locations at the same time. The location filter will only allow one location to be chosen at the same time and this will prevent users from comparing between two locations. Furthermore, as the key message focus primarily on the data period of February, a single bar chart with 12 individual bars representing the 12 months of a year might not be appropriate to focus on month of February only. Representing the whole data period at once by months will lead to a total of 72 bars plotted in a single bar chart which might be hard for users to perceive the information and the key message.

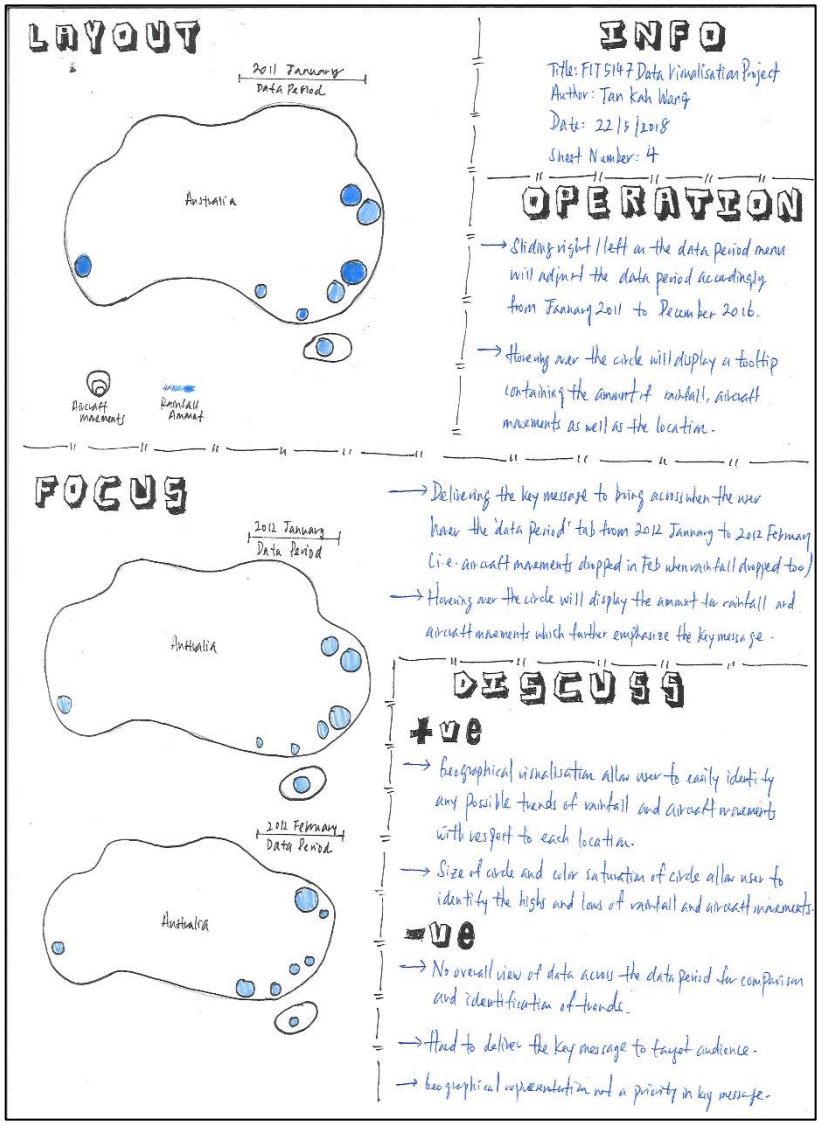


(Figure D – Five Design Sheet: Sheet 3)

iv. 5 Design Sheets Methodology – Sheet 4

The third design being considered was a symbol map. Aircraft movements will be represented by size of the circles on the map and rainfall will be represented by the colour saturation of the circles. A filter is included whereby users can drag horizontally to filter between data period. The key message can be make prominent when users access the data period filter from January to February which will see a decrease in size of circle and lighter saturation of the circle as there were lower aircraft movements and rainfall in February compared to January. Hovering over each circle will also display the amount for rainfall and aircraft movements which allow users to grasp the real data values. This geographical visualisation allows users to easily identify any possible trends with respect to each location and the size and colour saturation of the circles allow users to further identify the highs and lows of the two data factors.

However, as the key message focus primarily on just the time factor (i.e. February), geographical representation should not be the focus as it might lead users away from the key message. Furthermore, it will be hard for users to view the whole data period plotted at once in the same map which might make the key message less prominent. Lastly, though the highs and lows of rainfall and aircraft movements of different locations can be easily identified with the colour saturation and size of the circles, this is not a component of the key message.



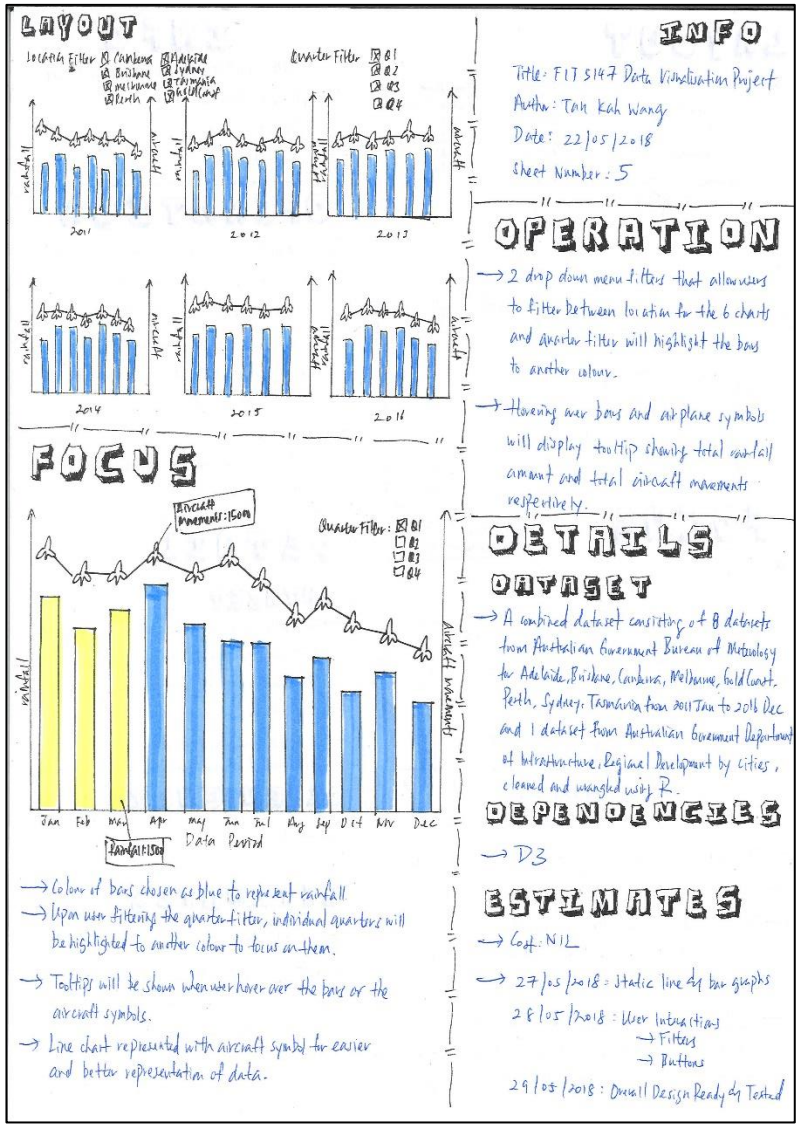
(Figure E – Five Design Sheet: Sheet 4)

v. 5 Design Sheets Methodology – Sheet 5

The final chosen design was an improvement from Sheet 3 – combination of line and bar charts. Instead of plotting the whole dataset in a single chart, the dataset will be plotted on 6 different charts with each chart representing a year (i.e. 2011 to 2016).
A quarter filter will aid the communication of the key message as the bars corresponding to the users' selection of quarter will be highlighted across all the 6 charts which allow users to focus on just those bars. A location filter will be included to allow users to filter between locations. Finally, tooltip will be displayed when users hover over the bars represented by blue which corresponds to colour of the rainfall, and the airplane symbol representing the data factor of aircraft movements.

Main reason for choosing this design was it will be easier for users to view the whole dataset in a single glance for easy identification of the key message instead of filtering the data period year by year or having a single chart with 72 bars. The second reason was I felt this is the simplest and most effective representation of two data factors over the time factor. Bearing in mind the target audience is a government body, a visualisation that can deliver messages promptly should be implemented.

Users will first have an overall view of the whole dataset regardless of the locations and highlighting the bars to quarter 1 using the quarter filter will allow users to see a noticeable pattern of drop in aircraft movements when there was a drop-in rainfall in months of February across 2011 to 2016 shown in the 6 different charts. This will be effective for the communication of the key message which was to highlight the month of February. And if the users are interested to identify if this pattern is consistent across all 8 locations or more prominent at any location, the location filter can be used.



(Figure F – Five Design Sheet: Sheet 5)

3. Implementation

i. Description of implementation

The visualisation was built using D3 and the dataset consisting of 8 different spreadsheets of Rainfall data[3]-[10] from the website of Australian Government, Bureau of Meteorology and 1 spreadsheet of Aircraft Movements data[2] obtained from Australia Department of Infrastructure and Regional Development, were combined and cleaned using R in a single dataset to be read into D3. No external libraries were used in D3.

ii. Reasons for implementation

The main reason to choose a combination of bar and line chart was because it's the simplest, yet most effective way of visualising two data factors over a time period. Users can easily identify any possible trends or patterns by comparing the bars and the lines. Choosing to implement 6 individual charts instead of one was due to the ease of interpretation of the data in the former. If the dataset was to be plotted on a single chart, it would consist of 72 bars which makes it hard for users to interpret. And because I will want the users to have an overall glimpse of the whole dataset in a single glance, it would thus be feasible to plot them in 6 different charts. With the quarter filter implemented as well, the key message which was to highlight the drop in aircraft movements when there was a drop in rainfall in the month of February across 2011 to 2016 can be further made known when the user chooses quarter 1 in the filter, which will make the bars corresponding to January, February and March be highlighted. As mentioned in one of the questions during the design process, geographical representation of the data was not a priority as it's not a significant component in the key message. Initial plan to include a map as a filter in the bar and line chart was changed to just a simple drop-down menu. Another reason for not implementing the map filter was due to the difficulty in plotting maps in D3. Another minor change made was to make the bars blue instead of having raindrop as background due to the difficulty for users to read the bars when plotted using the raindrop background in the bars.

iii. Difficulty in implementation

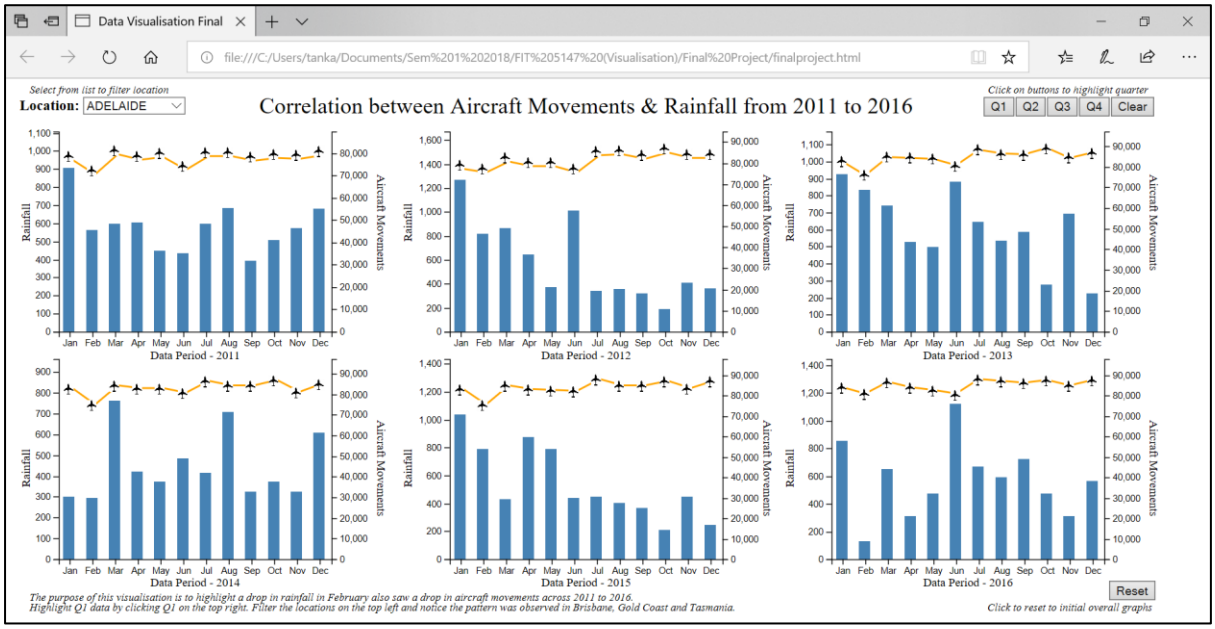
Using D3 for building the design was indeed tough with no prior knowledge other than the third assignment in this course and thus extensive research was needed to be carried out on the web on the following components:

- Nesting and summing the data to plot the individual charts
- Setting the domains and axis of the charts based on different datasets
- Appending airplane symbols on top of the line charts
- Rotating the axis labels based on different positions in the visualisation
- Creating drop down filters and changing the plots accordingly
- Involving transition when the user filters the data
- Creating buttons for users to change between quarters
- Creating a reset button to return graph to original state

4. User Guide

i. Overview of the visualisation

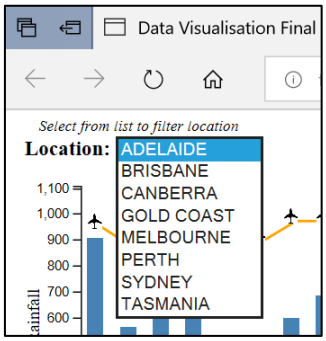
When the user enters into the HTML file, 6 individual charts *(refer to Figure G)* will be shown which show the individual plots for each year from 2011 to 2016 of the overall rainfall (represented by the bars) and aircraft movements (represented by the lines and airplane symbols). One drop down menu filter for location can be found on the top left and another quarter filter in the form of buttons can be found on the top right. Note that both filters can work together at once. A reset button can be found on the bottom right if users decide to return the initial overall view after having filtered on the location. A brief description of this visualisation and intended key message can be found on the bottom left.



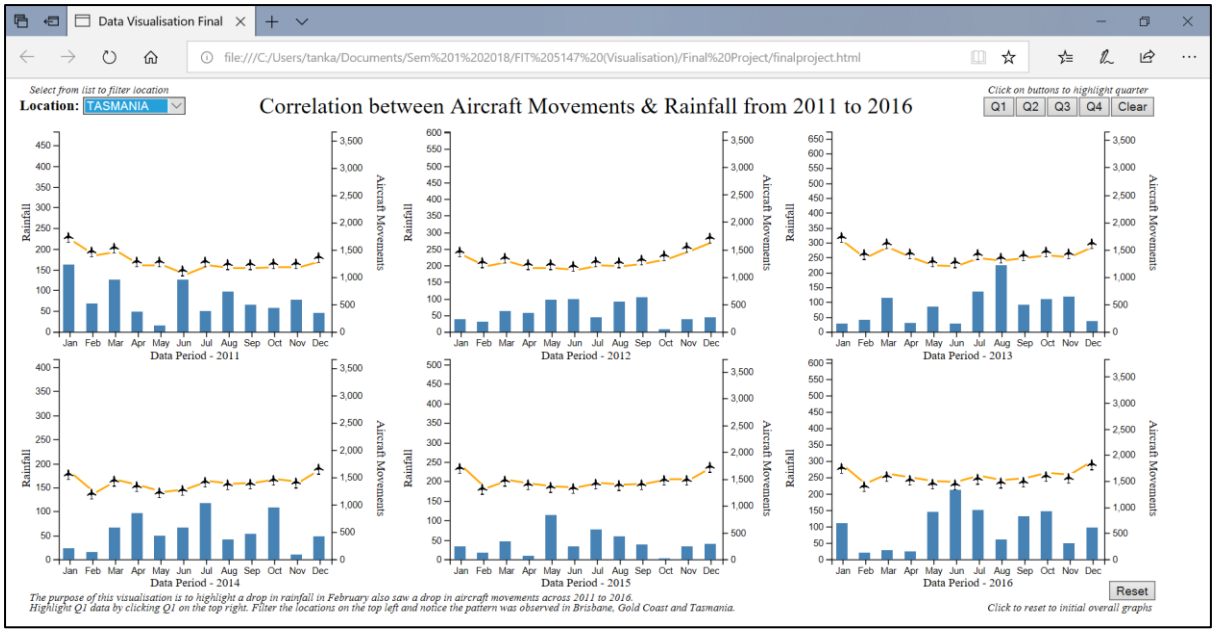
(Figure G – Initial view of the visualisation)

ii. User Interactions – Location Filter

The first user interaction is the location filter found at the top left corner. Users can filter between the locations by selecting any one from the listed eight airport locations *(refer to Figure H)*. Once selected, the 6 charts will change accordingly with new axes corresponding to their values plotted as well *(refer to Figure I)*.



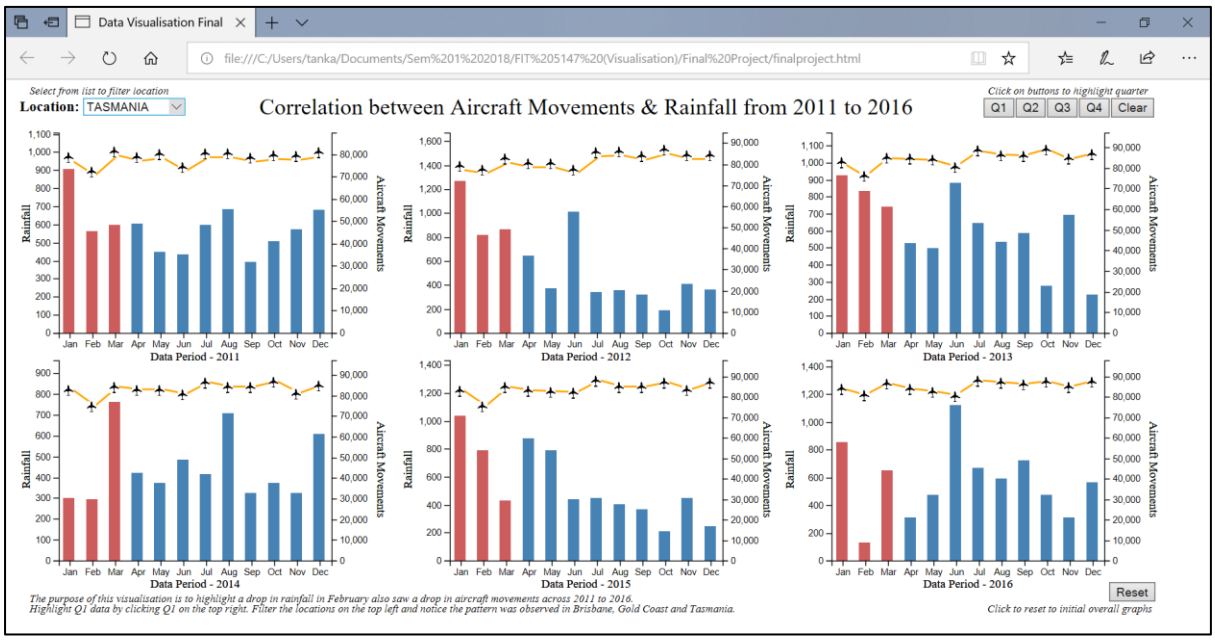
(Figure H – View of the location filter)



(Figure I – View of the visualisation after filtering to TASMANIA in the top left location filter)

iii. User Interactions – Quarters Filter

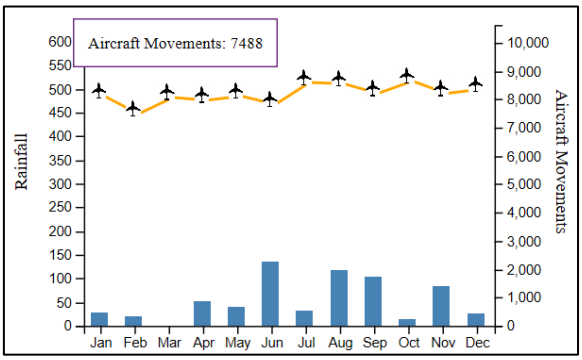
The second user interaction is the quarter filter in the form of buttons found at the top right corner. Users can click one or more buttons and the bars corresponding to the selected quarters will be highlighted red *(refer to Figure J)*. A “Clear” button has also been included for users to clear the highlights of the bars.



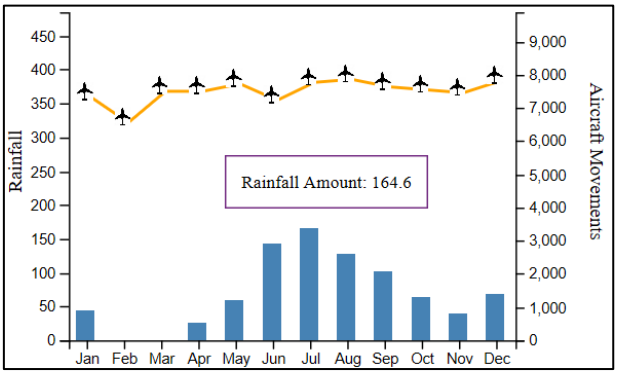
(Figure J – View of the visualisation after clicking on the Q1 button on the quarter filter at the top right corner)

iv. User Interactions – Tooltip on Hovering

The third user interaction is the tooltip which will be displayed once the user hovers over the airplane symbol showing the total aircraft movements *(refer to Figure K)* or when the user hovers over the individual bar showing the total rainfall amount *(refer to Figure L)*.



(Figure K – Cropped view of the visualisation after hovering over the aircraft symbol showing a tooltip)



(Figure L – Cropped view of the visualisation after hovering over the bar showing a tooltip)

5. Conclusion

It was certainly not easy to implement an interactive data visualisation and in particularly, in D3, which is a tool I've always yearned to learn before taking up this course but refrained from due to minimal exposure in writing and understanding Javascript language. Prior knowledge and exposure to R Programming back in undergraduate days further pushed me to try using D3 instead of R so as to learn something new and develop a new skill through experience. Though the final interactive visualisation was incomparable to the sophiscated D3 examples found on the web, it was definitely satisfying to be able to design an interaction that can allow users some sort of basic interactions rather than the usual static visualisations.

Through this project, I've also got to know designing an user interactive design definitely is a long process with plenty of details to consider before implementation. Identifying the target audience and key message at the start of the implementation is definitely the key to implementing any effective visualisation, be it static or interactive. A visualisation designed for a particular group of target audience might not deliver the same message to another group of audience and not as effective as different audience groups have different knowledges and perspectives. The design itself is also another important aspect in designing an effective user interactive visualisation. From choosing the colours, shapes, types and the way the user interactions work, these are definitely important aspects in developing an effective visualisation as at the end of the day, what we hoped to archieve through effective visualisation is for users to grasp the intended key messages in the easiest manner and shortest time possible.

The first thing that could have been done better in this project was to include more datasets that could have affected aircraft movements. As the intended target audience is a government body, it would have been better if the key message I'm trying to deliver can be supported by other datasets which could have been factors that affect aircraft movements other than just one focus in this project which was the rainfall factor. With only a single factor, it might be hard to convey the key message nor convince the governmnet body to look into planning more routes and flights in the month of February. The second thing that could have been improved was implementing the map filter in the visualisation. Though it was mentioned earlier that geographical representation of the data was not a priority in delivering the key message, it could have been useful for users to visualise the location on the map itself when exploring the visualisation. With better skills in D3 and Javascript language, this could have been implemented.

6. References

[1]

https://infrastructure.gov.au/department/about/index.aspx

[2]

https://bitre.gov.au/publications/ongoing/files/WebMonthlyAirportJanuary2018.xls

[3]

http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=023034

[4]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=40913

[5]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=70351

[6]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=40196

[7]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=86282

[8]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=9021

[9]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=66037

[10]
http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=96081

THE END