THE RELATIONSHIP BETWEEN WEATHER AND CAR ACCIDENT

JUNE 19, 2020

ZILONG ZHENG
29499496

Table of Contents

1. Introduction:

In today society, the number of people using car is gradually increasing. When people drive on the road. There are a lot of factors that affect the drive, some of which are weather factors. When it rains, it will increase the number of car accidents. The car accident is a very terrible thing, he will take the drive's life. This paper mainly counts the total number of car accidents in each state of Australia each year, and the factors are wind speed, maximum temperature, humidity, air pressure, and rainfall. Analyse the weather factors in each state year by year. And analyse the relationship between various weather factors and car accidents, and then analyse the general trend of car accident and weather.

This paper is focus on three question. The three questions are 1) what is the total annual number of car accident in each state in Australia? 2) what is the annual average of various weather factors in each Australian state, and what are the trend of these weathers in recent years? 3) what is the relationship between each weather factor and car accident?

This report summarizes the weather factors in Australia, the number of car accidents and their relationship, People can observe the trend of Australian weather. People can understand the relationship between weather factors and the number of cars. Through the relationship, drives can understand the impact of weather factors on car accidents. Drivers can change driving strategies through weather factors to achieve the purpose of reducing the probability of car accidents.

2. Design

The design of this paper uses 5 design sheets. Through this method, all data visualization methods are analysed, and the best method is selected to form the final data visualization results.

2.1 First sheet

5 The first page of the design sheet is brainstorming. On this page, we draw all the ideas and choose the best one among these ideas.

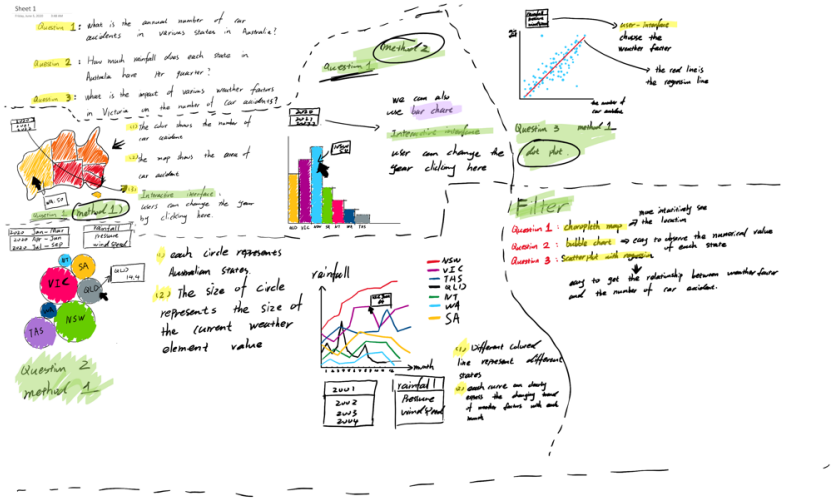


Figure 1: 1 sheet brainstorming

For these problems, these problems can be solved by choropleth, histogram, line chart, bubble chart and dot chart with regression line. choropleth and bar chart can be used to

display the total number of car accidents in each state of Australia each year. Line charts and bubble charts can be used to display the average annual values of weather factors in each state of Australia. Dot plots with regression lines can be used to show the relationship between weather factors and car accidents.

By comparing the advantages and disadvantages of each picture, three pictures are finally selected, and each picture can solve a problem. For the first question, I chose to use choropleth, which is more convenient for users to observe the location and data of each state, and it is more beautiful than the histogram. For the second question, I chose a line chart, which can show the trend more conveniently. Dot plot with regression line to solve the third problem.

2.2 Second sheet

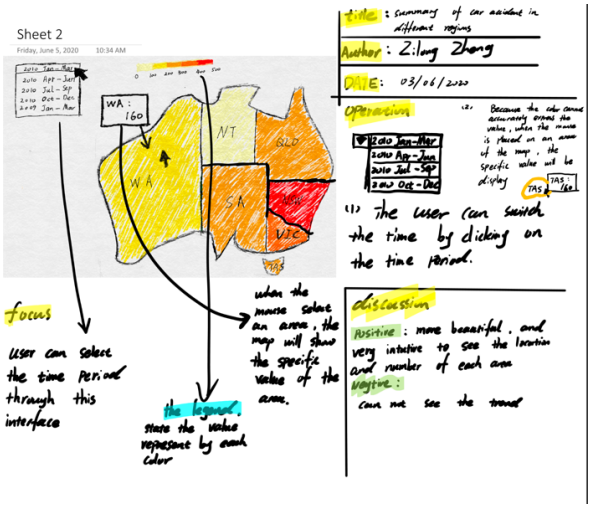


Figure 2: sheet 2

The second picture discusses the design of choropleth. Choropleth shows the number of car accidents in each area by colour. When the number of car accidents is red, it means that the number of car accidents in this area is high. When the colour is lighter, it indicates the car accidents in the area. Not many. However, the user can only compare the difference between the two regions by colour, and it cannot see the specific difference. Therefore, a label is added to the map. When the user moves the mouse to a state, the state's label will be displayed, and the specific number will be displayed on the label. There is also a selection bar on the left of the map for users to choose. The user can select the year to be displayed according to the selection bar. The advantage of choropleth is that it is beautiful, and users can easily determine the location of each state through the map. The comparison between the states is also very obvious. The disadvantage is that the trend of the number of car accidents with the year cannot be seen.

If choropleth is used alone, the annual total number of car accidents and the annual average of weather can be displayed, but the trend of weather changes and the relationship between car accidents and weather cannot be seen through this graph.

2.3 Third sheet

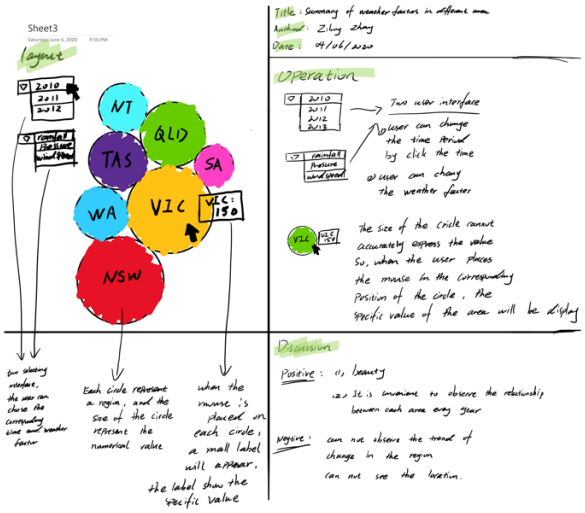


Figure 3: third sheet

The third sheet is some details about the visualization of the second problem. In the original design, the bubble chart was used. The bubble chart is very beautiful and can show the difference between the two regions, but the use of the bubble chart does not reflect every A change trend of weather element with time, so in the final design, I changed this part to a line chart. The line chart can not only observe the difference between each area in each year, but also can observe the trend of weather factors in all areas over time. There is a choice of state and weather factors in the upper left corner of the line chart. Users can choose according to their needs. Select the state and weather factors to be displayed

If only the design of sheet2 is used, the factors of car accidents and weather cannot be displayed, but the change of sheet2 is that shee2 can show the trend of weather changes with time.

2.4 Fourth sheet

Figure 4: the fourth sheet

The fourth picture mainly solves the third problem. In this part of the picture, a dot plot with a regression line is used. The regression line in the dot plot shows the relationship between the abscissa and the ordinate. The abscissa in the dot plot is the user. The value of the selected weather, the abscissa is the number of car accidents. There are two selection bars on the right for users to choose state and weather factors.

The third sheet can show the relationship between weather factors and car accidents, but the separate third sheet does not show the total number of car accidents in each state and the annual average of weather.

2.5 The final sheet

Through the analysis of the previous three sheets, the final design is to stitch together the three solutions above, and each sheet can solve a problem.

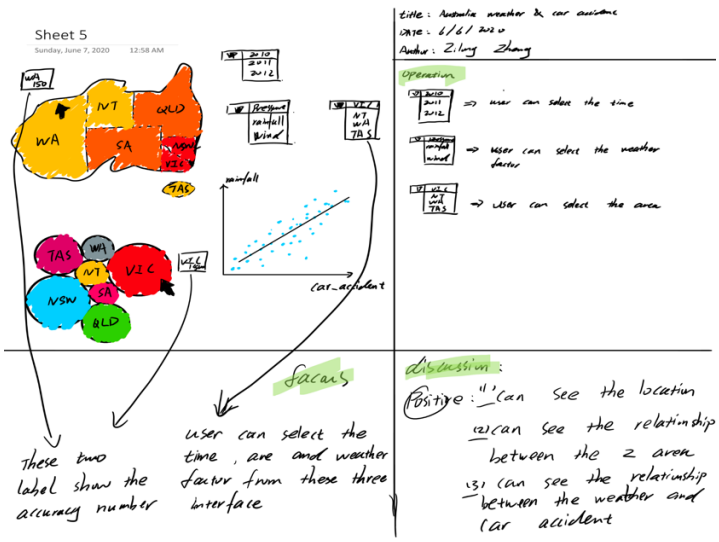


Figure 5: the final sheet

In the final design, the map can answer a question, and users can use the map to understand the total number of car accidents in each state each year. The bubble chart in the final design was replaced by a line chart. Although the bubble is very beautiful and the difference between the two states can be seen more clearly, the trend of weather factors over the years cannot be observed, so in the final product Replaced the bubble chart with a line chart that better shows the changing trend. The user can answer the second question according to the line chart. The third question can be answered based on a dot plot with a regression line. The regression line in the dot plot shows the relationship between weather factors and car accidents.

In terms of user interaction, because the map determines the colour depth by the size of the value, the user cannot get the exact number by the colour depth, so when the user puts the mouse on a state in the map, a label is displayed. The label will show the specific value to go to this state. There is also a selection bar and check box in the final design to facilitate users to filter data and choose to visualize the data they need.

Because each picture corresponds to a problem, the final design puts each of the three pictures into a sub-interface, so that the structure is clearer, and users can switch the interface according to their own choice.

3. Implement

This report uses r language to realize data visualization. These include some libraries of r language, which are shiny, shinythemes, leaflet, ggplot2, dplyr, htmltools and geojasonio.

3.1Shiny and shinythemes.

Shiny is an open source R package, and Shiny is based on the development framework of the r language web front. Users can create web pages that users can interact with through this library. The reason for using this library is that the software can realize data visualization and allow users to interact with these diagrams.

Shiny themes is a library that can change the style of shiny interface. In the final design result, the three pictures are not on one page, and pagination needs to be made. Shiny themes can change the style of this UI to make it more beautiful.

3.2 Leaflet

Leaflet is a modern, open source library developed for building mobile device friendly interactive maps. The Leaflet design adheres to the ideas of simplicity, high performance and good usability. It can operate efficiently on all major desktop and mobile platforms. It will take advantage of HTML5 and CSS3 in modern browsers, and also supports old browser access. Supports plug-in extensions, with a friendly and easy-to-use API document and a simple, readable source code. Leaflet's powerful open source library plug-in involves all aspects of map applications including map services, data provision, data format, geocoding, route and route search, map control and interaction. There are more than 140 types of plug-ins. These controls enrich the functions of Leaflet, and at the same time, it is very convenient to implement custom controls with good scalability.

For question 1, users need a map of Australia to map the number of car accidents in each state. Leaflet can help us achieve this function, and it can realize the interaction between the map and the user.

3.3 ggplot2

ggplot2 is a powerful drawing library in r language. ggplot2 can draw all graphs, such as line charts, histograms, etc. Questions 2 and 3 require a line graph and a dot plot with a regression line. These two graphs can be displayed by using the drawing library ggplot2.

3.4 dplyr

The R package dplyr can be used to process structured data inside or outside R. Compared to the plyr package, dplyr focuses on accepting dataframe objects, which greatly increases the speed and provides a more robust database interface. At the same time, the dplyr package can be used to operate Spark's dataframe. This article is just the basic dplyr package study notes, so it will not discuss some advanced applications or compare the performance with the data.table package.

In the process of drawing, some data needs to be processed to filter out the required data. This library needs to be used to help us achieve this function.

3.5 Htmltools

The function of the htmltools library is mainly to generate and output html format results. In the map interaction function, the colour does not display the exact number of car accidents in each region, so label is required to display accurate numbers, and it needs to be generated in the label format. And output html format results, so you need to use this library to help generate and output html format results.

3.6 Geojasonio

For question one, the output needs to be a map of Australia, so we cannot use the world map in leaflet to draw the result of question one, so we need to draw the outline of Australia, which requires the use of the Australian geojson file. This file draws the outline of Australia and the outline of each state through coordinates. The function of geojsonio is to read this data, and the file can be used by leaflet.

4. User guide

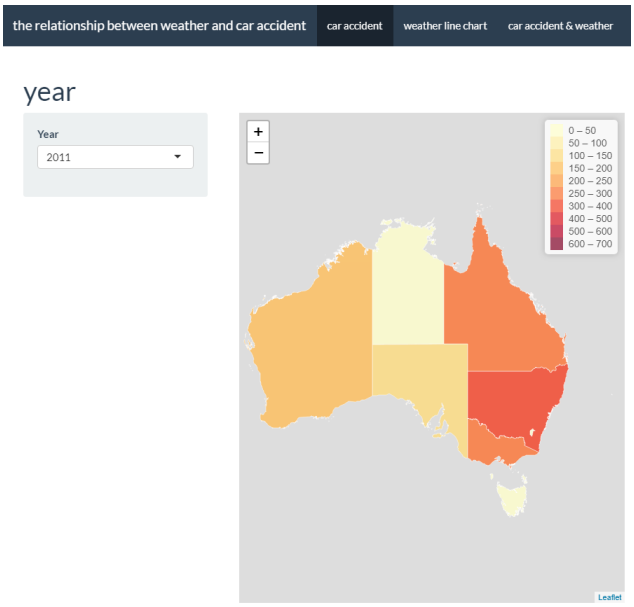


Figure 6: First page

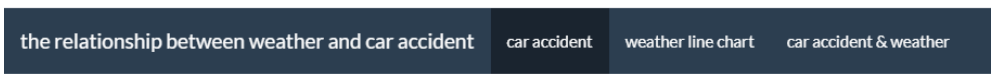


Figure 7: Navigation bar

This picture is the final result, and the top is a navigation bar. Users can switch the displayed content by clicking car accident, weather line chart, car accident & weather. The content of the car accident is shown in the picture. This page mainly shows the number of annual car accidents in each state of Australia.

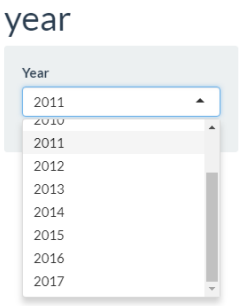


Figure 8: Year selection bar

There is a selection bar on the right side of the figure, the user can switch the displayed year according to their needs.

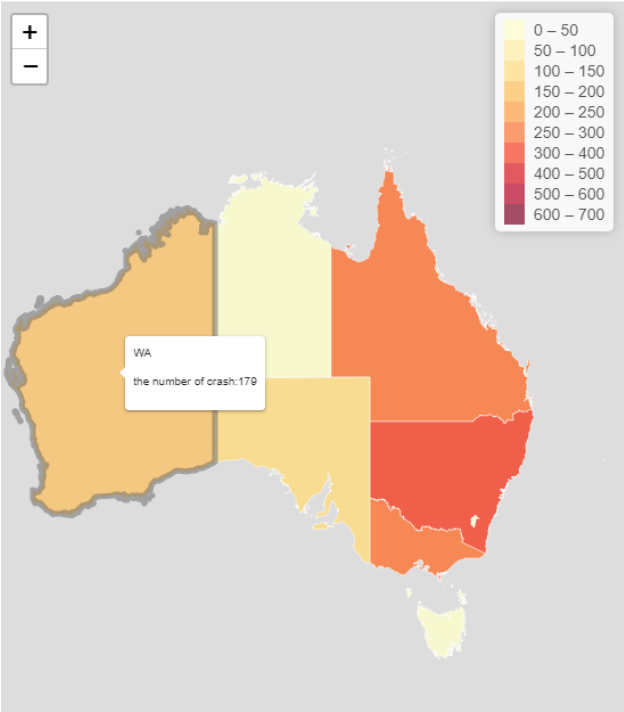


Figure 9: Choropleth

According to the colour depth of the map, it is judged which area has a large number of car accidents, which is not convenient for users to observe accurate values. When the user drags the mouse to a state in the map, a label will be displayed, and the name and specific value of the state will be displayed on the label, and the border of the state will become grey and thick.

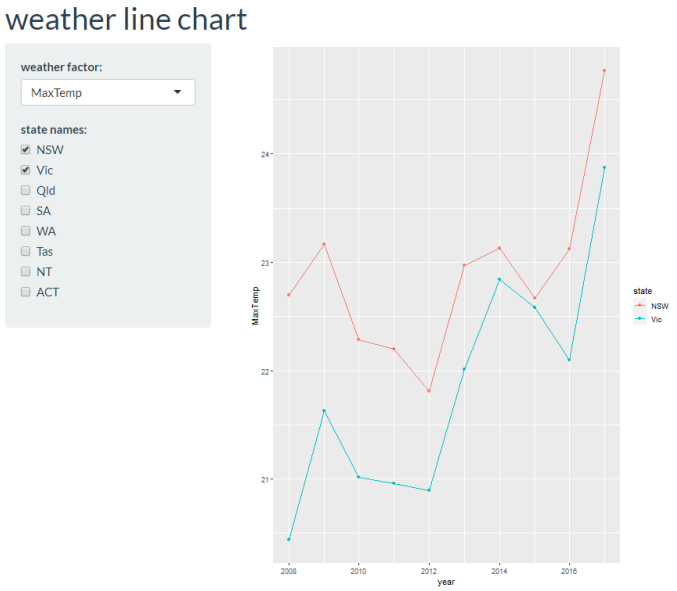


Figure 10: line chart

Click the weather liner chart on the navigation bar to enter the second page. This page shows the change of each weather factor over time. Each line in the line chart represents a state.

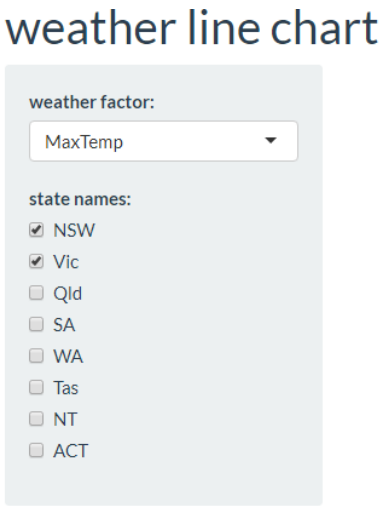


Figure 11: weather select bar

The user can select the weather factor whose curve needs to be displayed through the selection bar on the left. There is a check box under the selection bar. Users can check the state name according to their needs. Only the state with the checked name will be displayed in the line chart on the right.

Click car accident & weather on the navigation bar to enter the third page. The third page shows the relationship between weather factors and car accidents.

Figure 12: Third page

There are two selection bars on the left side of this page. The first selection bar can select states. The user can switch between states according to his needs. The second selection bar is weather factors, because there are many kinds of weather factors in the data , And the relationship between each weather factor and car accident is different, so the user can use this selection bar to switch the weather factors that want to observe the relationship with the car accident.

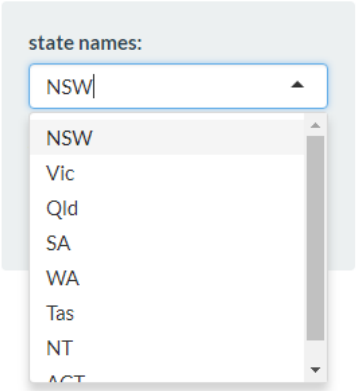


Figure 13: State selection bar

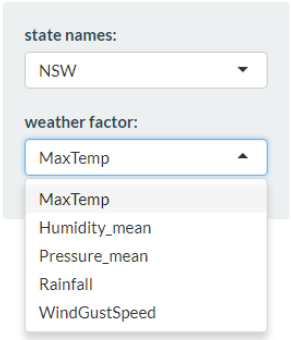


Figure 14: weather select bar

The left side of the figure is a dot plot, the abscissa is the data of weather factors, and the ordinate is the number of car accidents. The line in the figure is the regression line. According to these points, the regression line will show the approximate relationship between weather factors and the number of car accidents. Users can understand their relationship through this line.

5. Summary

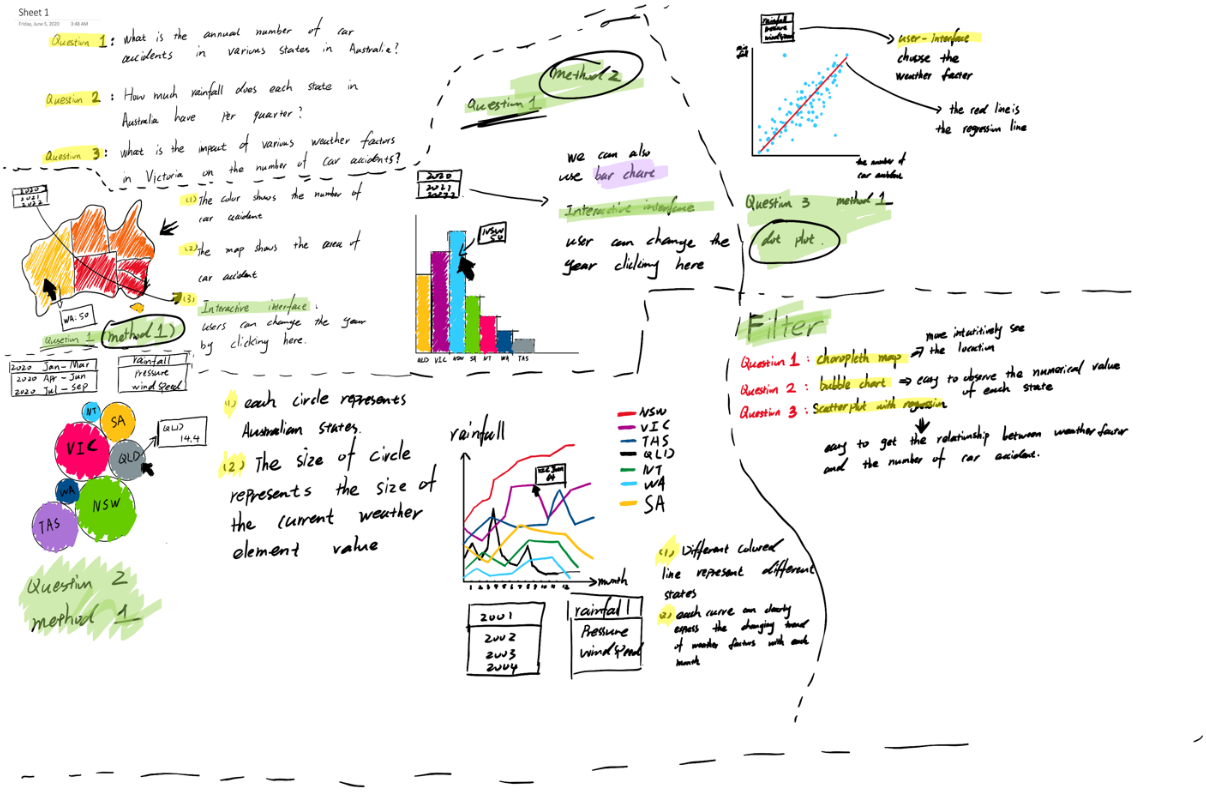
Through the final design, users can know the number of car accidents in each state each year, and which state has the most car accidents and which state has the least car accidents each year. The user can know the change trend of each weather in every state over time, and user can also compare the weather difference of different states every year. Users can understand the relationship between each weather factor and the number of car accidents in each state of Australia according to the third page.

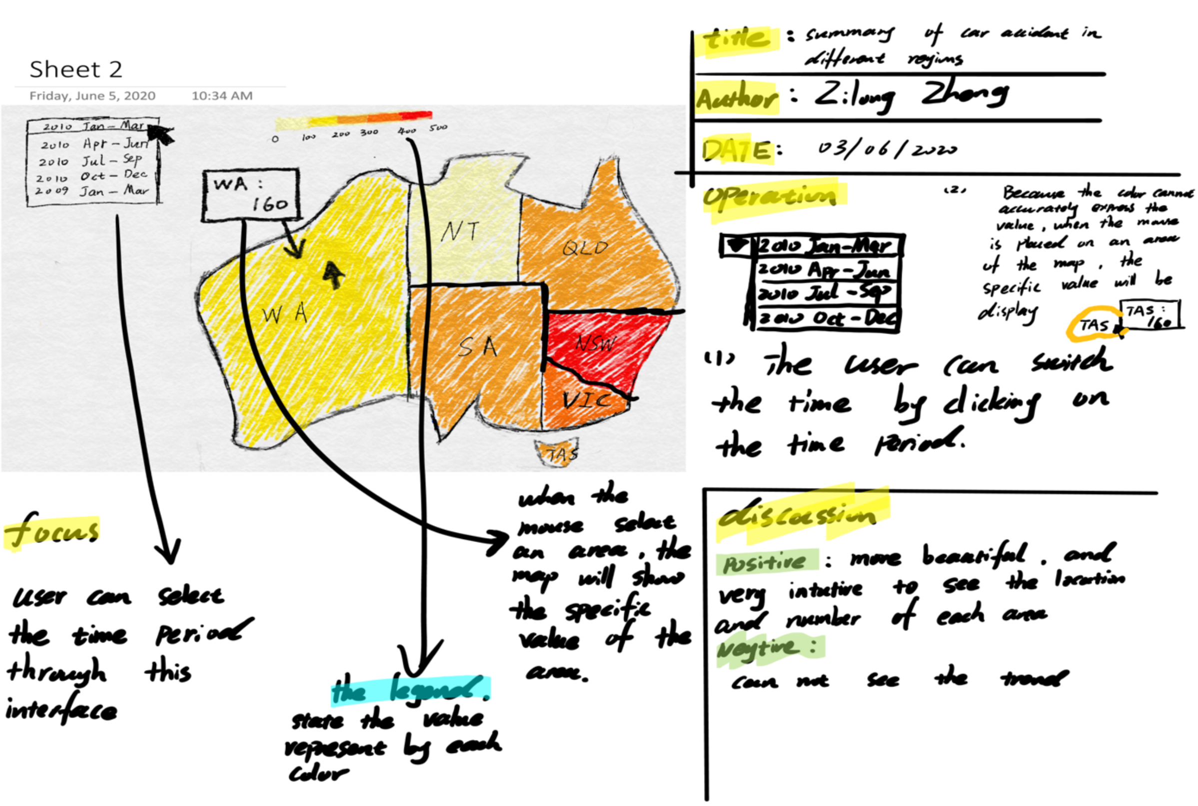
I learned the design method of 5 sheets. Using this method will make my design more perfect. For data visualization, I learned how to use shiny. By using the leaflet library, I learned how to draw a map, and how to implement map interaction.

6. Bibliography

1. Shiny. (2020). Retrieved 19 June 2020, from https://shiny.rstudio.com/
2. Leaflet for R - Introduction. (2020). Retrieved 19 June 2020, from https://rstudio.github.io/leaflet/
3. v3.3.1, g. (2020). ggplot2 package | R Documentation. Retrieved 19 June 2020, from https://www.rdocumentation.org/packages/ggplot2/versions/3.3.1
4. v0.7.8, d. (2020). dplyr package | R Documentation. Retrieved 19 June 2020, from https://www.rdocumentation.org/packages/dplyr/versions/0.7.8
5. Story Map Journal. (2020). Retrieved 19 June 2020, from https://www.arcgis.com/apps/MapJournal/index.html?appid=75eff041036d40cf8e70df99641004ca
6. Sheet 1 – Brain storm | Five Design Sheet. (2020). Retrieved 19 June 2020, from http://fds.design/index.php/2015/06/25/sheet-1-brain-storm/

7. Appendix


Sheet 1
Question 1: what is the annual number of car accidents in various states in Australia?
Question 2: How much rainfall does each state in Australia have per quarter?
Question 3: what is the impact of various weather factors on the number of car accidents? in Victoria
(1) the color shows the number of car accident
(2) the map shows the area of car accident
(3) Interactive interface: users can change the year by clicking here.
Question 1 method 1
method 2
Question 1
we can also use bar chart
Interactive interface
user can change the year clicking here
user-interface chose the weather factor
the red line is the regression line
Question 3 method 1
dot plot
Filter
Question 1: choropleth map → more intuitively see the location
Question 2: bubble chart ⇒ easy to observe the numerical value of each state
Question 3: Scatter plot with regression
easy to get the relationship between weather factor and the number of car accident.
(1) each circle represents Australian states.
(2) The size of circle represents the size of the current weather element value
rainfall
NSW
VIC
TAS
QLD
NT
WA
SA
month
(1) Different coloured line represent different states
(2) each curve can clearly express the changing trend of weather factors with each month
Question 2 method 1


Sheet 2
Friday, June 5, 2020 10:34 AM
2010 Jan – Mar
2010 Apr – Jun
2010 Jul – Sep
2010 Oct – Dec
2009 Jan – Mar
WA : 160
NT
QLD
WA
SA
NSW
VIC
TAS
title: summary of car accident in different regions
Author: Zilong Zheng
DATE: 03/06/2020
Operation
(1) The user can switch the time by clicking on the time period.
(2) Because the color cannot accurately express the value, when the mouse is placed on an area of the map, the specific value will be display
TAS : 160
focus
user can select the time period through this interface
the legend. state the value represent by each color
when the mouse select an area, the map will show the specific value of the area.
discussion
Positive: more beautiful, and very intuitive to see the location and number of each area
Negtive: can not see the trend

Sheet3

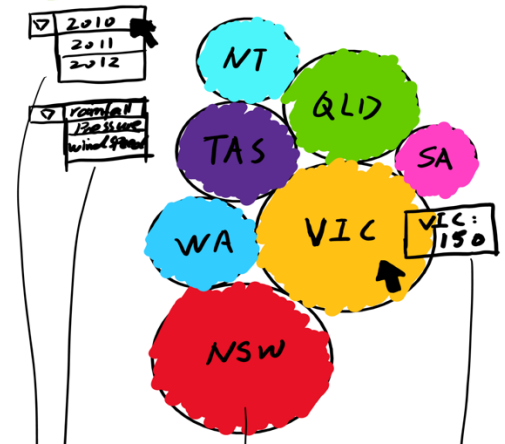
Saturday, June 6, 2020 9:56 PM

Title : Summary of weather factors in different area

Author : Zilong Zhang

Date : 04/06/2020

layout



two selecting interface, the user can chose the corresponding time and weather factor

Each circle represent a region, and the size of the circle represent the numerical value

when the mouse is placed on each circle, a small label will appear, the label show the specific value

Operation

Two user interface

① user can change the time period by click the time

② user can chang the weather factor

The size of the cricle cannot accurately express the value So, when the user places the mouse in the corresponding position of the circle, the specific value of the area will be display

Discussion

Positive : (1) beauty

(2) It is convenient to observe the relationship between each area every year

Negtive : can not observe the trend of change in the region

can not see the location.

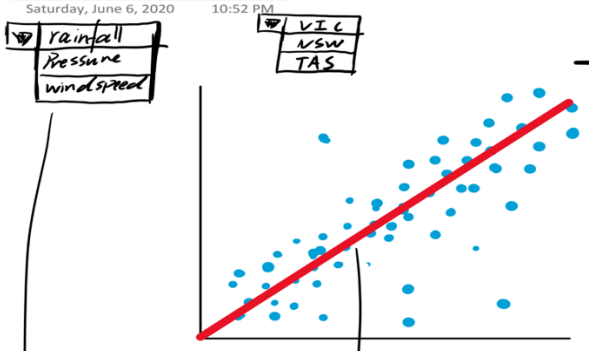
Sheet4

Saturday, June 6, 2020 10:52 PM

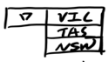
Title : the relationship between the number of accident and weather factor

Author : Zilong Zhang

Date : 5/6/2020



operations



↳ the user can change the area

the user can change the y-axis variable through this interface

focus

the selecting interface. user can select the weather factor.

Line regression. this line can help users more intuitively observe the relationship between weather factors and the number of car accidents

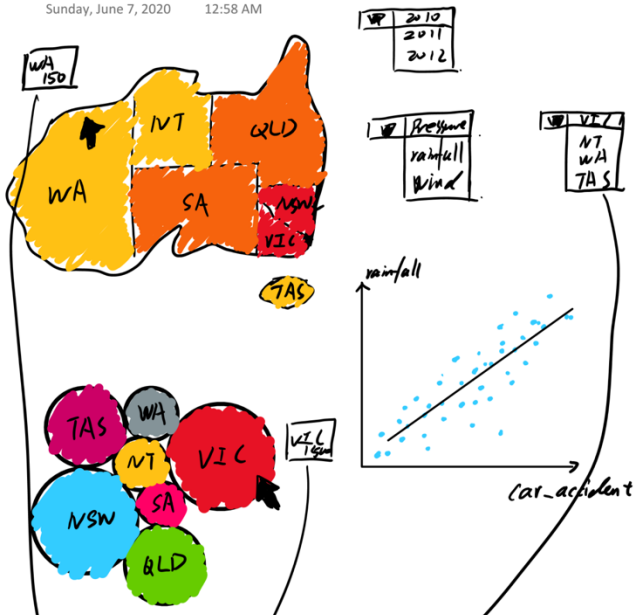
discussion

Positive : easy to see the relationship

negtive : just can see one area

Sheet 5

Sunday, June 7, 2020 12:58 AM



title : Australia weather & car accidents

DATE : 6/6/2020

Author : Zilong Zhang

Operation

=> user can select the time

=> user can select the weather factor

=> user can select the area

facaus

These two label show the accuracy number

user can select the time, area and weather factor from these three interface

discussion :

Positive : (1) can see the location

(2) can see the relationship between the 2 area

(3) can see the relationship between the weather and car accident