

Data Visualization Project of Analysing Crashes on Victorian roads.

Student name: Xitong Jiang

Student ID: 28402987

1.Introduction

My project is based on the data of amount of accidents in victoria state between 2014-2018. It includes the trend of number of crashes between 2014-2018, the amount of crashes in a selected region and the top region have most crashes was given. Besides, trends of different light condition, accident type and so on between years and hours were included. My target audience are those who interested in the data visualization of crashes, which might not have any background for this area.

2. Design

2.1 design process

2.1.1 *ideas sheet*

For the first sheet, ideas sheet, I put all ideas given by myself or potential users. Here is some of them:

- Map: show the amount of crashes in a region?(Region: Kingston)
- Is the amount of crashes increasing or decreasing?
- Any relationess between severity and time or location?
- Higher speed cause more damage?
- Is light condition influencing the rate of crashes?
-
-

After the process of filter and category, some ideas such as 'Higher speed cause more damage' and 'Any relationess between severity and time or location' was discarded due to the original dataset cannot answer these question.

Navbar structure is decided to used in my website.

2.1.2 Alternative Designs

Choropleth map, motion charts, table charts were included in my alternative design.

Choropleth map can directly show the comparison between different regions. However, my original dataset lack geographic details of fixed boundaries. So the proper choice of the map should be re-designed. In addition, table chart was chosen to show the ranking list.

However, it cannot show comparison effectively. Hence how to show the rank should be reconsidered.

2.1.3 Realisation

The final design includes proportional data maps, motion charts, bar charts, and line charts. Proportional data maps compare the number of crashes in different regions in the 'Summary' page. With the `groupby()` and `cluster` function, the number of crashes in a region can be calculated. Since the vector of longitude and latitude was provided, proportional data maps are more suitable than the choropleth map mentioned in the first design.

Bar chart presents data with rectangular bars with the size proportional to the values that they represent, which helps to show the rank of a region in terms of the number of crashes. The bar chart is more clear and direct than the table chart in the first design.

2.2 Human perceptual system and human communication assumptions

For the design of my project, Nielsen and Molich's Ten User Interface Design Guidelines was followed. First, by giving a navbar, users can go to view different aspects of analysis by clicking tabs. The navbar also contributes to help users knowing what state the system is in which follows the guideline of user Control and Freedom. Secondly, for the selectInput box and radio buttons, users can only select the choices provided by the system rather than input something unexpected which influence the flexible operation of the website, which helps for error Prevention. Interview with representative of my target user was done for me to get suggestions from a user's perspective.

3. Implementation

3.1 libraries used in the project

- Shiny library: shiny provides framework to build interactive web apps straight from R.
- GoogleVis library: Requiring an Internet connection. googleVis package helps with providing dynamic visualization, especially the data between years.
- Leaflet library: Leaflet helps producing mobile-friendly interactive maps. For the original dataset in my project have attribute of longitude and latitude, leaflet was suitable for my design.

- Shinythemes library: Get access to the theme provided by shiny.
- Plotly library: In my project, the change in 24 hours was included to show in a day which hour have most crashes. Previous googleVis library help only show the change between years or months. plotly library makes interactive, publication-quality graphs online.
- Shinyjs library: Support js function can be used in shiny

In terms of difficulty, a js function was used to support that user can localize himself/herself on the map. Different layer of a map was also applied.

3.2 Reasons for the implementation decisions for narrative visualisation

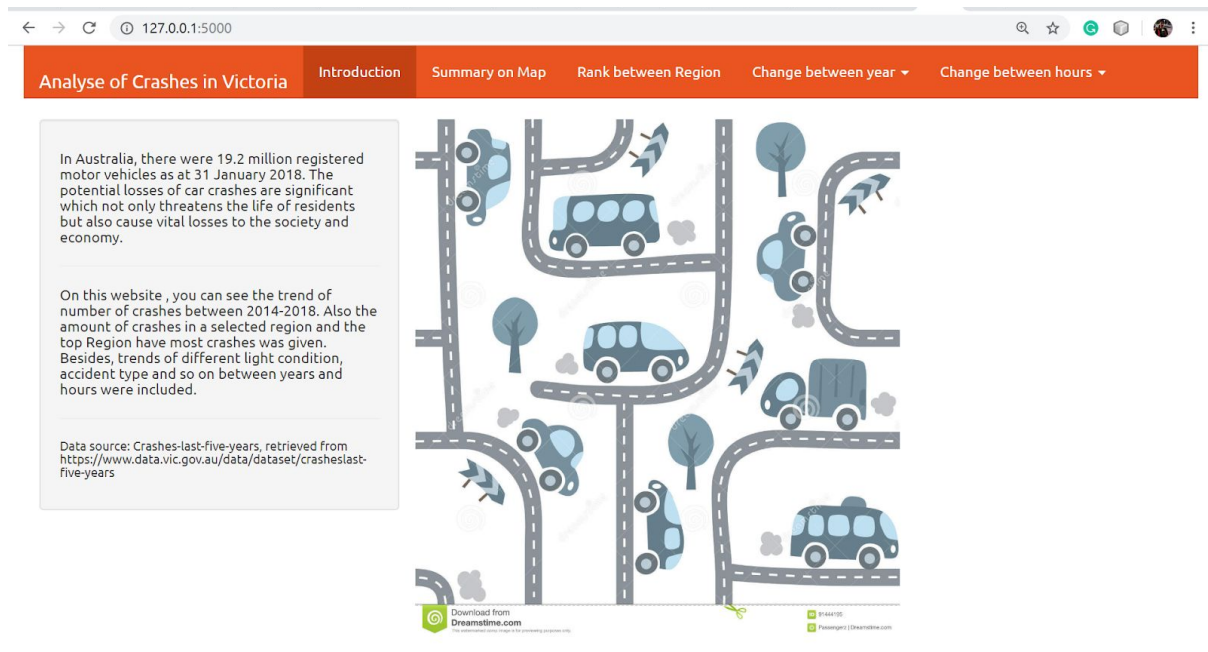
In my first design, choropleth map, motion charts, table charts were included. Compare to my first design, proportional data maps, motion charts, bar charts, and line charts were presented in final design.

Proportional data maps compare the number of crashes in different regions in the 'Summary' page. For my first design, Choropleth map was preferred to directly show the comparison between regions. However, my original dataset lack geographic details of fixed boundaries, which makes Choropleth map is applicable. With the groupby() and cluster function, the number of crashes in a region can be calculated. Since the vector of longitude and latitude was provided, proportional data maps are more suitable.

Bar chart was used to show the rank of Region in terms of the number of crashes. However, at my beginning stage, the table chart was chosen to show the ranking list. But during my development, I find it not effective to show comparison only with the numeric value in the table. Hence I re-pick bar charts to show differences.

4. User guide

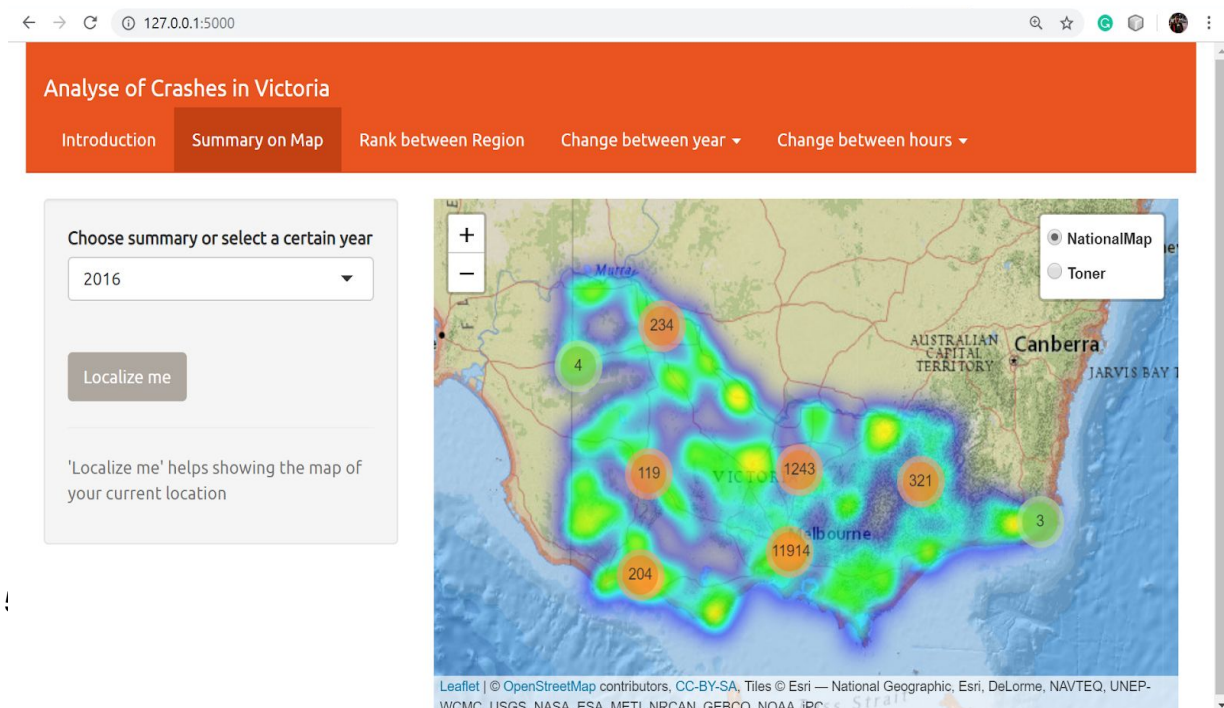
Javascript and flash need to enable in the web browser. Also need Internet connection. Once successfully run the app in shiny, the project shows in the web browser.

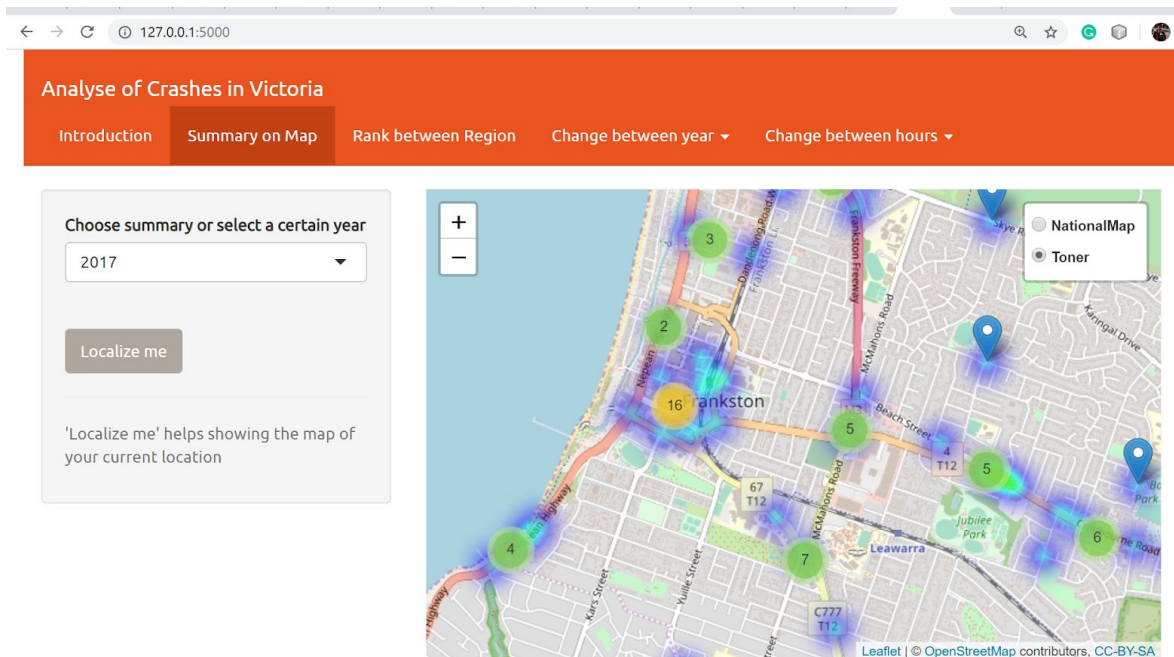


4.1 Summary on the map

When select "Summary on the map" tab, the heat map about crashes in Victoria state was given. Besides, user can view the data in summary (whole data from 2014 to 2018) or in a certain year.

On one hand, when 'Localize me' button was clicked, the map automatically zooms to the user's position, which is more user-friendly and efficient than user zooms to find his or her location on the map. On the other hand, two types of the map were given to be selected, they are nation map and Toner map. User can select the style of the map they like.





Choose summary or select a certain year

2016

Summary

2014

2015

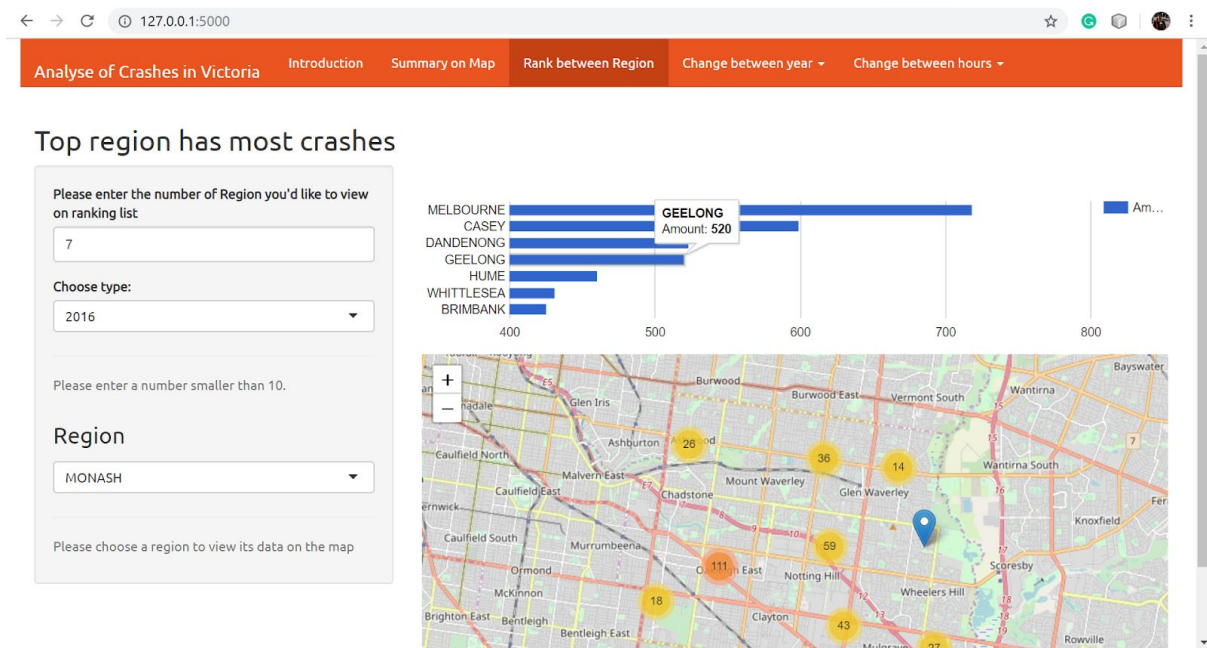
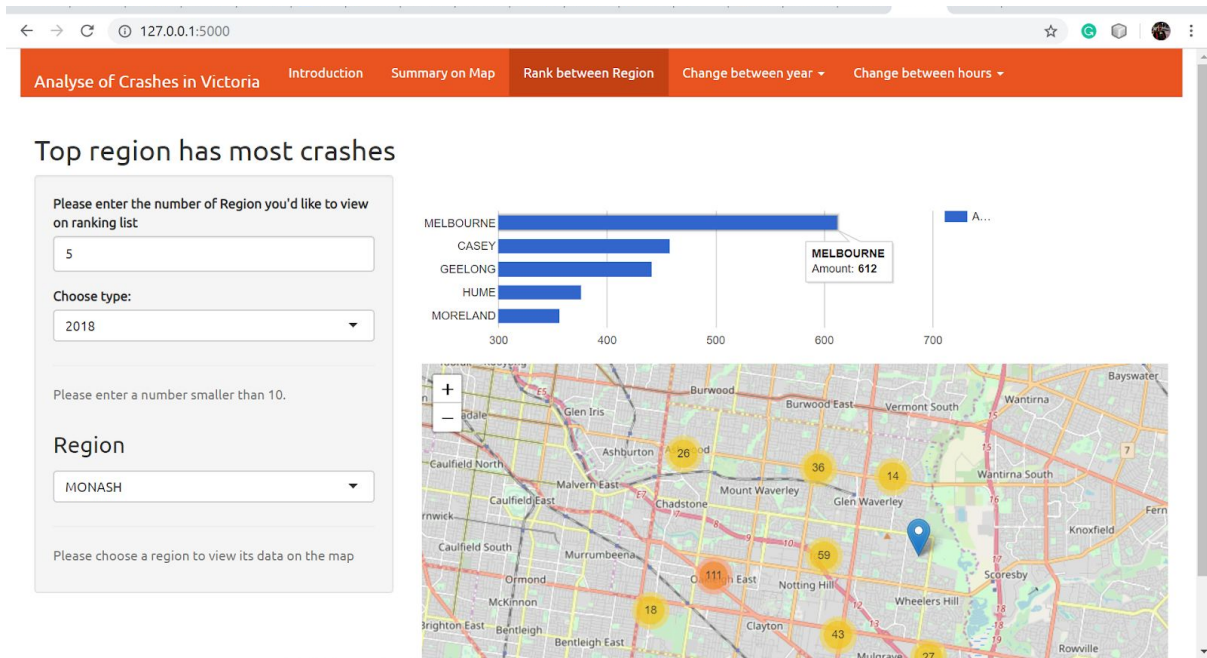
2016

2017

2018

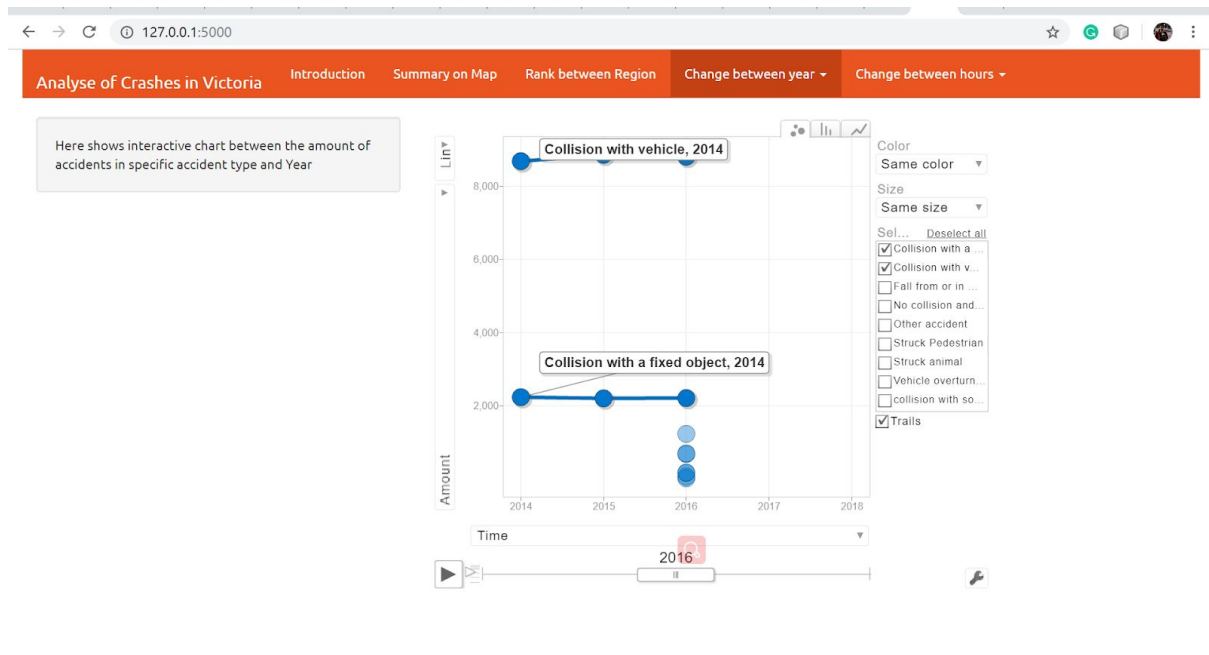
4.2 Rank between Region

In this page, a bar chart of the top Region has most car accidents was given. Something special of the page is that the number of Regions shows in the ranking list depends on the number user put in the first input box. In addition, only data with the selected year was given.



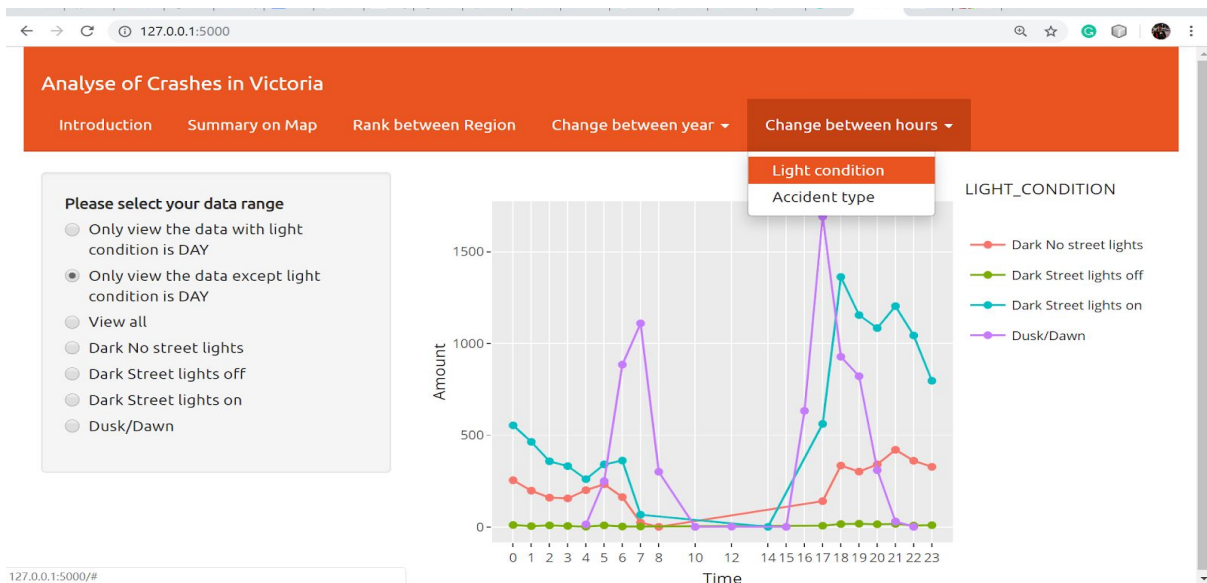
4.3 Change between year

In here, user can view the tendency of crashes in various severity, light condition and accident type from motion chart, which is a dynamic bubble chart with interactive exploration.

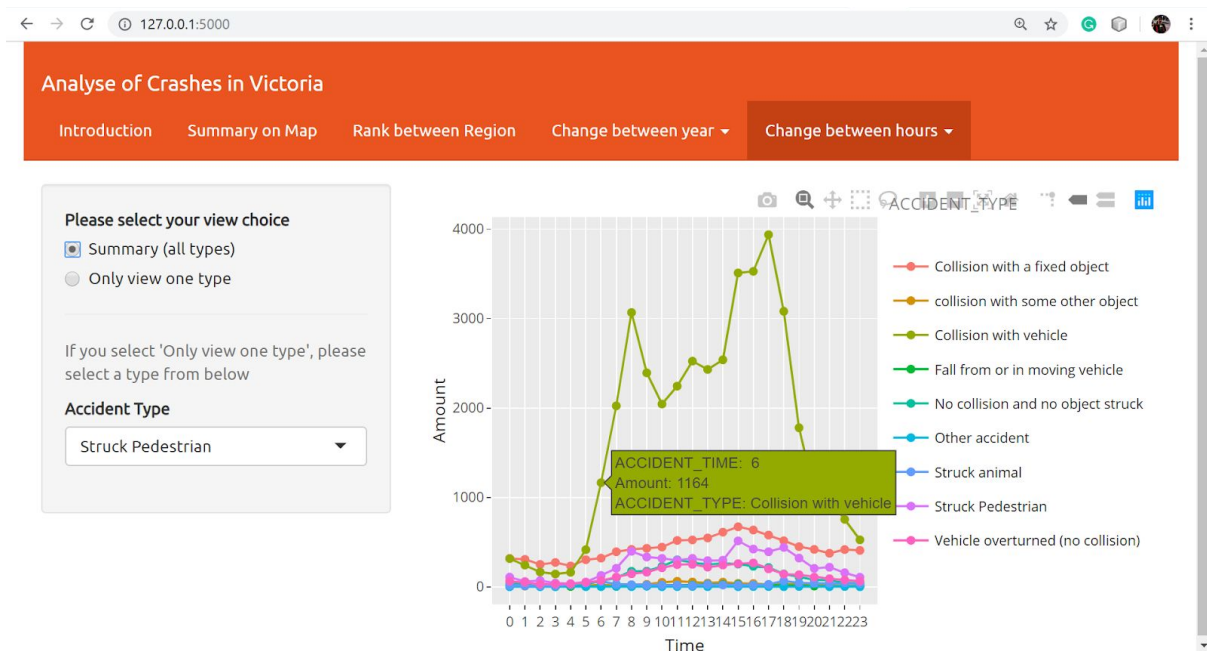


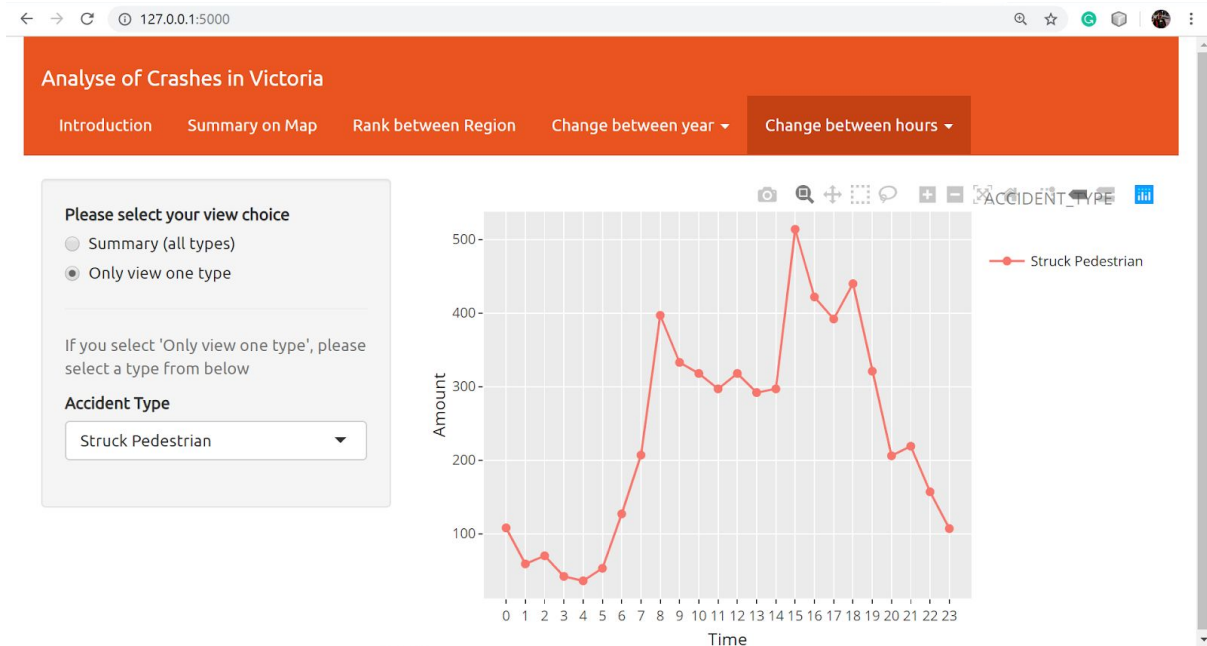
4.4 Change between hours

On this page, users can flexibly select his/her data range. Multiple choice for users here since users come from different background and have different preference on data range. On one hand, some users prefer to view the whole data. On the other hand, some might only interested in some specific conditions, such as the data with light condition under day, or data excepted light condition under day, or data with a specific light condition such as dusk/dawn.



For 'Accident type' tab, various options also given to guarantee the demand from most users can be realized.





5. Conclusion

Data visualization transforms invisible data into visible graphical symbols. It can connect and correlate complex data that cannot be explained or correlated to gain insights and values. With appropriate charts to express data clearly and intuitively, data visualization realize data self-interpretation and data speaking. The right side of the brain can remember images a million times faster than the left brain can remember abstract words. Therefore, data visualization can strengthen the audience's understanding and memory of data.

During my development process, the type of charts showing the trend in a day(24 hours) had been doubted. First motion chart was decided to use, however, motion chart does not support to show the hours properly, which makes the message transformed to user so confused. Consequently, I decided to choose another type of chart which can express correct and clear information to users, ploty package was used in this circumstance. The design of data visualization is an iterative process and it is unavoidable to change the design again.

6. Reference

- Crashes-last-five-years, retrieved from
<https://www.data.vic.gov.au/data/dataset/crasheslast-five-years>
- 4 tricks for working with r leaflet and shiny,
<https://www.r-bloggers.com/4-tricks-for-working-with-r-leaflet-and-shiny/>
- <http://gradientdescending.com/queensland-road-accidents-mapped-with-shiny-and-leaflet-in-r/>
- <https://stamen.com/>
- <https://rstudio.github.io/leaflet/basemaps.html>
- <https://deanattali.com/blog/advanced-shiny-tips/>

7. Appendix

Ideas Sheet

5 Design Sheet Cici

IDEAS

- User can see visualization from different aspects, Alcohol, speed zone, time.
- Show the amount of crashes alcohol-related.
- Show the increasing / decreasing trend of crashes.
- What location in Victorian state have most crashes?
- Light condition VS Time.
- Speed zone VS Time.
- can select certain year / month / date?
- Alcohol-related have relationship with severity?
- Map = can select a small region, like cantfield?
- Accident Type VS Region = How many crash with vehicle in Kingston?

Average in Year? or total?

Filter

- Map (show location, leafout, longitude, latitude).
- Severity distribution from 2014-2018
- Speed distribution (Under Alcohol)
- Light condition distribution under 24 hours
- Region = which region is a high-risk area (Top 10)

Total 5 years?
Or select average.

Category

Map:

1. Summary.
2. Alcohol-related
3. Top 10 (leafout) - (bar chart).

Analyse in a Day (24 hour)

1. Light condition.
2. Speed zone. (Motion chart)

Analyse in a Year

1. Severity distribution.

Combine & Refine

Map

Analyse in a Day

Analyse in a Year

Amount

4394

Serious, 2014

Year

2014

2018

Select

Amount

512

(Motion chart)

00:00

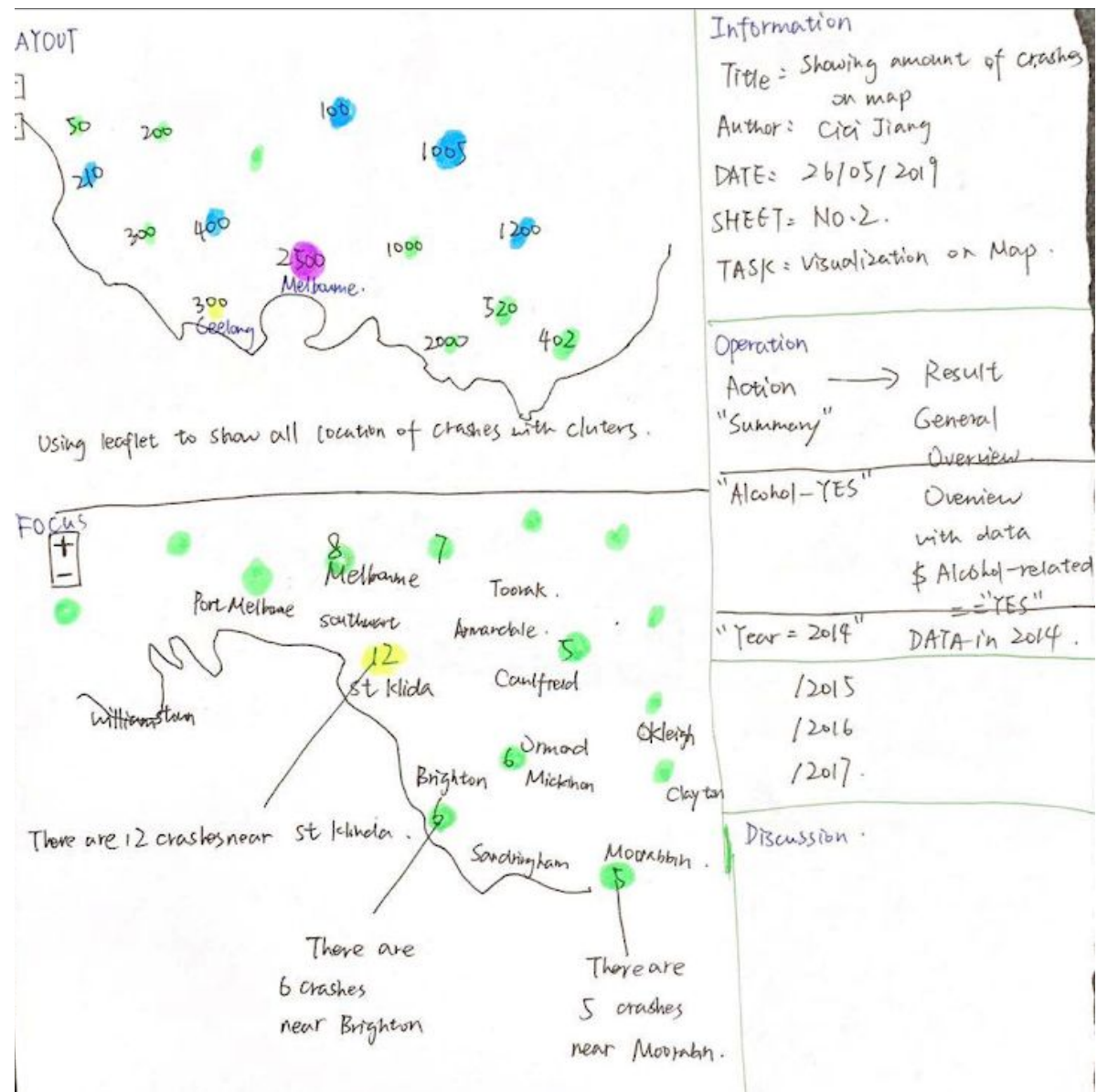
24:00

Time

Select

About Data:
Data source.

Alternative Design



LAYOUT

Crashes in Victoria.

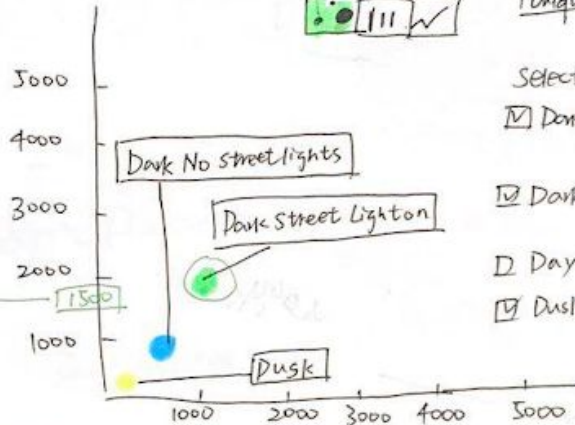
Map Motion Chart on Day Motion Chart on Year Alcohol selected

Group Data by

- Light condition (selected)
- Speed zone
- Severity

can select the data with alcohol related

- All (selected)
- No Alcohol
- Alcohol - related



Color

Unique Color

Select

☒ Dark No street light

☒ Dark street light on

☐ Day

☒ Dusk / Dawn

Amount



22:00

Means = 5723 crashes happen in 19:00

Means mouse click on represents 1500 crashes under dark street light on when 22:00 pm

5723
19:00



Color

Unique Color

Select Deselect All

Discussion:

- Average / Sum.

Information:

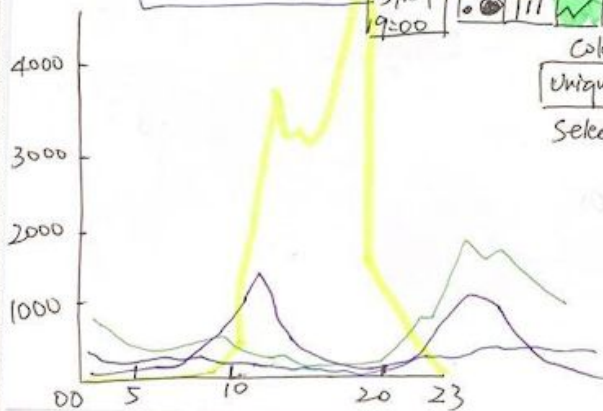
Title: Show number of crashes under various light condition

Author: Cici Jiang

DATE: 27/05/2019.

Sheet: No. 3.

TASK: Motion chart on Light



LAYOUT

Crashes in Victoria.

Map
Summary		
Top Risky		
More Details		

Number of Rank to view

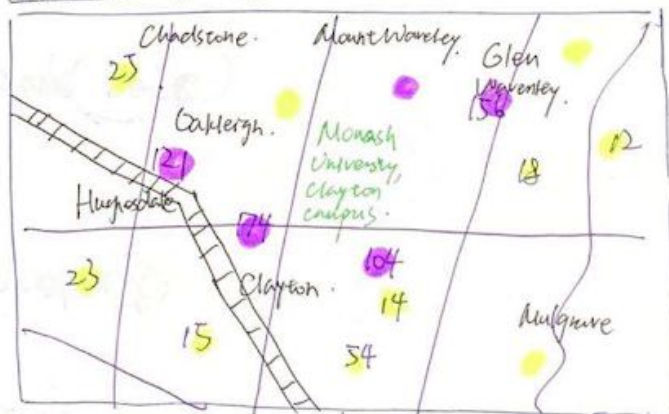
10

Region	Amount
Melbourne	4068
CASEY	2865
GEE LONG	2667
DANDENONG	2470
HUME	2267
BRIMBANK	2236
MORELAND	2137
YARRA RANGES	2111
MONASH	2081
WHITLESBA	1761
YARRA	1761

Select Region to see details of crashes in this Region

Monash 2017

Select Year



FOCUS:

- The number of crashes near Monash University, Clayton campus.
- Can give users something special based on their input, more personalized customization.
- Ranking can start from 1 to 20.

Information:

Title: More Detail about Maps.

Author: Cici Jiang.

DATE: 27/05/2019.

Sheet: No.4.

Task: Deeper analyze on regions.

Discussion:

Operation:

Action	Result
"Year"	select data with certain year
"Region"	select data with certain Region

"Number"	Show data with top "Number" rank
----------	----------------------------------

Realisation

