Title : Data Visulisation Project

Student name - Dishi Jain

Student ID - 30759307

Tutorial Number - 24

Tutor Name - Pratik

Table Of Contents

INTRODUCTION

This data visualisation project focuses on the information about the crashes on the roads of Victoria for the last couple of years. The data is spread across different categories giving details about the different aspects of the accidents, examples where the accident took place, how many young and old drivers were involved,etc. This project hence is useful in depicting and talking about certain questions like the number of accidents on the basis of the location(using latitude and longitude). This visualisation will help the audience to understand and evaluate where the accidents are taking place when seen on the map. This visualisation also gives the user an option to select from a drop down the value of the year. Based on this value the cluster map of accidents is refreshed.
Another topic of discussion in this project is to identify which regions suffer the most number of accidents. Here we wish to educate the audience and show them a pattern which would depict which area/region sees the most number of accidents and why. Based on the region selection we can also see how the particular region has accidents based on the light conditions of the environment.
This project contains another topic of interest which focuses on the count of accidents based on the light conditions and the age of the driver i.e. a young driver or an old driver. This visualization will help the audience to know the importance of the street lights/lightning condition of the environment and also find a relationship between this and the age of the driver.
Hence all these visualisations will focus on the number of accidents in Victoria divided into different categories like, location , region name , light condition and driver type(old or young).

The intended audience for this project is the public as well as the victorian road safety department. For the general public these visualizations will be helpful to educate them how light conditions affect the number of accidents and whether the age of the driver affects the count of accidents. With this information they can themselves take better care while driving. It would also help them to know which regions see the most number of accidents and they can be extra careful while driving in such regions.
For the victorian road safety department this project is useful in many ways. They can compare how yearwise the accidents are spread over different locations. They can also focus on which regions in particular see the most number of accidents and can apply stricter rules in those regions. It would also be helpful for them to know how light conditions affect the count of accidents and hence they would take measures to ensure that the light conditions are so that the accidents decrease.

Hence, this project will be useful for the general public as well as the victorian road safety department.

DESIGN

The visualisations that are selected for this project has been selected after carrying out immense processing and analysing of the different options. The different topics covered in this project are using different kinds of visualisations for depiction of the data in the most suitable manner such that it reaches the intended audience in the best way possible. Different visualisations like cluster map, area chart, doughnut chart and radar chart are finally chosen as the final visualisations that will be used for this project. However this selection was made after attempting and trying out different options in R. Different charts like bar charts, circular barplots, heatmaps and a lollipop chart were used and applied to answer the above questions. However they were not depicting the information in the best possible manner and hence they were not chosen. The best visualization that was clearly showing the required information is chosen and used for this project. The five design sheet methodology is shown below(in Apendix).

This five design sheet methodology contains 5 sheets each with a different motive. The sheet 1 is where the brainstorming is carried out. This contains all the ideas and their visulisations that I saw when simply looking at the data. This sheet contains many different visualisations and the messages that it tenders to convey to the audience. It is basically a brainstorming session including any possible ideas that may arise when looking at the data.
The sheet 2 contains the visualizations that were first picked while answering the questions. This includes the alternative designs that were picked which are heat map, bar chart or a circular barplot. These visulisations were not very helpful in depicting the answers to the audience and hence they were not picked as the final visulisations. The sheet2 contains some other alternative designs for answering the questions which are lollipop chart and line chart. These visualizations again however did not convey the message to the most extent. The sheet 3 contains the final visulisations that were picked in answering the questions. This includes the cluster map, the area chart and the radar chart. These visulisations were finally selected as the most efficient visulisations and hence were picked. They convey the message in the most easy and interactive way. These visulisations are also user interactive and pleasing to the eyes which will connect to the audience and hence have a better impact on them.
The final design was chosen after implementation of the alternatives. The final design for the implementation of the accidents on the map was chosen as a cluster map. This visualization was chosen as the number of accidents were so high that they could not be depicted as individual points in the first go. Hence the cluster map was chosen which forms clusters of the nearby points. Then with the user interaction, the user can zoom in and zoom out to expand a cluster and see the exact points where the accidents took place. Hence a cluster map was the best choice to show such a huge number of accidents count. This visualization actually has two cluster maps where one shows the accidents overall in Victoria and one shows the yearwise accidents in Victoria based on the selection of Year made by the user. Hence a cluster map was chosen to depict this information.
The final design for the implementation of the region wise count of accidents was chosen as the area chart. This map clearly depicts how the accident count is changing over the different regions seen in Victoria, for example North, South, etc. The area chart was chosen as it is user

interactive in the sense that by hovering over the chart we can see the region where the most accidents took place. The area chart also depicts clearly which regions suffer the most accidents and displays each region name by different colours hence making it more appealing to the user.
The final design for the depiction of how light conditions and type of driver affects the accidents, we choose a radar chart. This chart clearly depicts how different light conditions are affecting the number of accidents. It also shows two groups inside the radar chart that shows how young and old drivers' counts differ from each other. The user interaction is taken into consideration by asking the user which region's information he/she wishes to see.
Hence the final designs are taken into consideration and plotted in R using R shiny to convey the respective message to the intended audience.

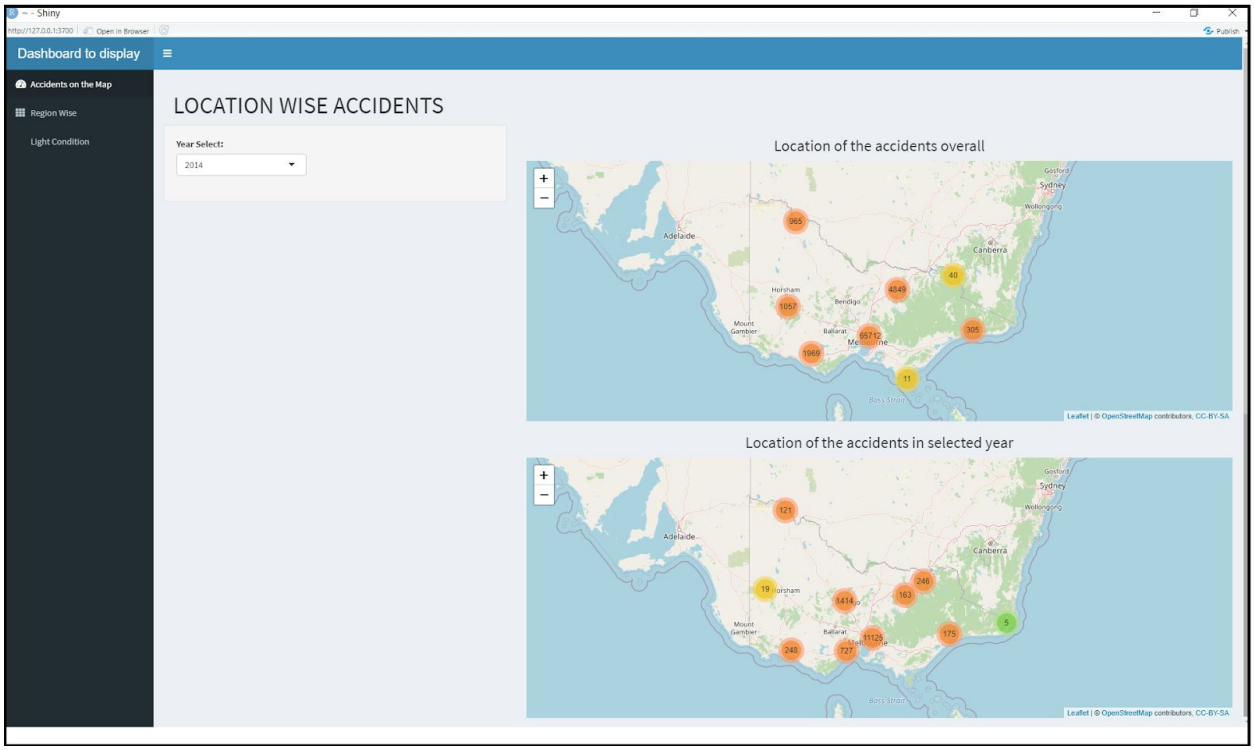
IMPLEMENTATION

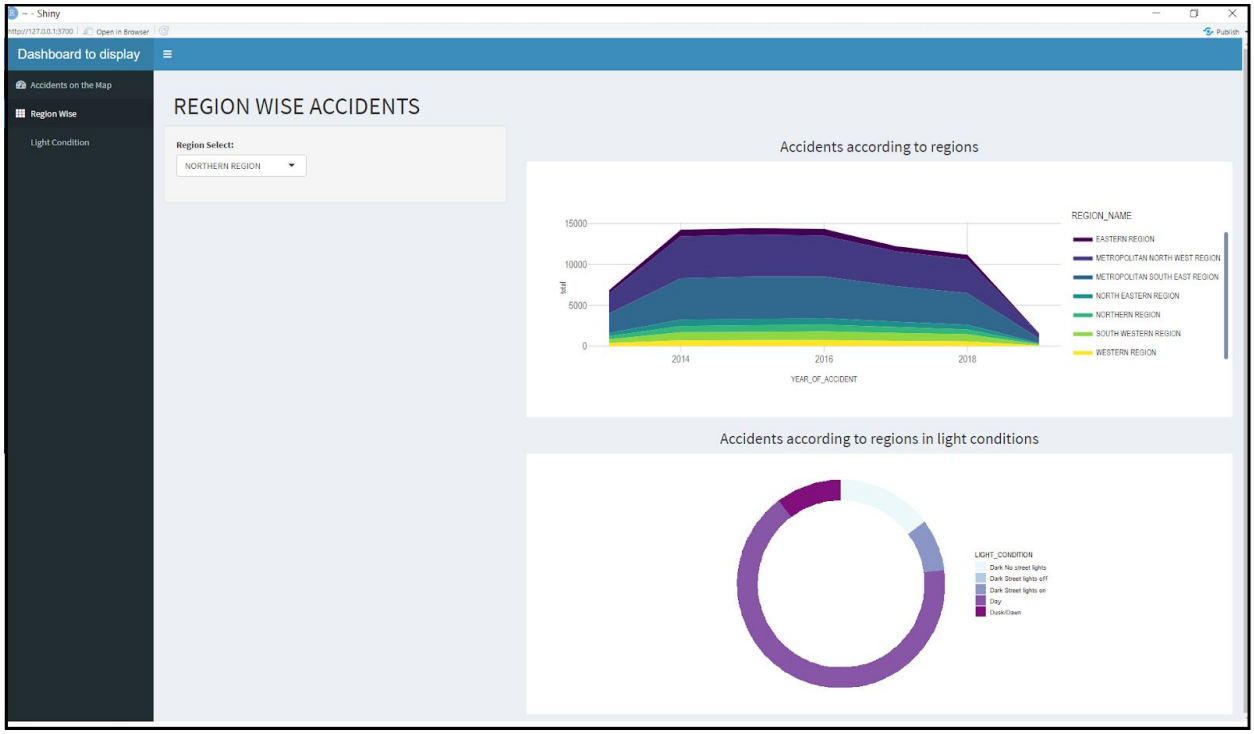
The implementation of this project was carried out after finalising the designs in the five design sheet methodology. The data was then explored and used to make the final visulisations. For the implementation of the cluster map the data was taken after exploration and using the leaflet function in R a cluster map was constructed. This takes the inputs as the data itself and the column like latitude and longitude over which the map is generated. Then by using the cluster option as markerClusterOptions, a cluster map is generated. This visualization actually contains two cluster maps. The first cluster map is the overall accidents in Victoria shown on the map as clusters. The user can then zoom in and zoom out on each cluster to get the exact location of the accidents. The second cluster map contains the cluster map of the accidents based on some year that is taken as an input value from the user. For creating this cluster map,the data is chosen according to the input of the user and if input is let's say 2015 then only those rows are taken from the data that have the year of accident as 2015. Using this new data we plot the cluster map and show it below the first cluster map. This is done so that the user can easily find any relationship that may exist between the overall accidents and yearwise accidents.
For the implementation of the area chart, a new dataframe is constructed that contains the groupwise count of accidents based on the year of accident and region name. Then ggplot is used to create the final visualization of the area chart which on hovering displays the region name. A legend is also displayed besides the graph to provide the exact name of the region based on the colors to the user.
The doughnut chart depicts how the light conditions were during the time of the accident. When a particular region is selected the doughnut chart changes as per the regions. This chart is constructed grouping the data based on light condition and region name and then counting the number of accidents. Then based on the input of the user, a region is selected and then the final doughnut chart is created.
For the implementation of the radar chart, I have grouped by original data as the number of accidents based on the light condition and region names to get the sum of cases where old drivers were involved and where young drivers were involved in the accident. This has been done using the summarise function in R and taking the sum of the column old driver and young

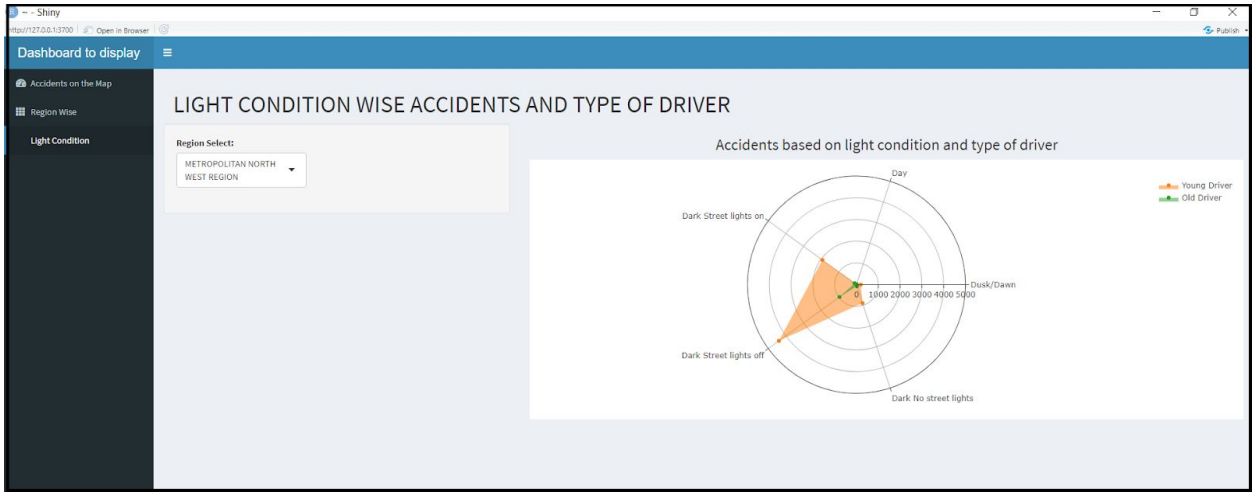
driver individually. Then based on the input from the user i.e. the region name, the particular visulisation is shown to the user as a radar chart.
The libraries used in the above implementation include

- library(shinydashboard)
- library(ggplot2)
- library(shiny)
- library(dplyr)
- library(htmltools)
- library(leaflet)
- library(scales)
- library(datasets)
- library(ggplot2)
- library(viridis)
- library(plotly)
- library(quantmod)
- library(gganimate)
- library(babynames)
- library(hrbrthemes)

After the implementation the final designs look like as shown below -







USER GUIDE

The user can run the shiny app and he will be displayed with a R shiny dashboard. This dashboard contains 3 tabs that displays the 3 visualizations used for this project. These tabs appear on the left side of the dashboard. On clicking the first tab the user can see two cluster maps and a drop down for user input. The first cluster map remains unchanged and it is displaying overall accidents. The second cluster map keeps on changing as per the input taken from the user. The user can click on the drop down and select the Year he is interested in.

Based on this selection the second cluster map is shown. The user can also zoom in and zoom out on the cluster maps to get the exact location of the accidents.
The user can now toggle to the second tab in the dashboard where the area chart will be displayed. Here the user can hover onto the area chart and according to this the region name will be displayed. The user can hover along the area chart across the different years.
The user can also select the region from the drop down available and then based on the region the doughnut chart changes its values to depict how the light conditions affect the accident counts.
The user can now toggle to the third tab in the dashboard where the radar chart is displayed. Here the user can select an input from the drop down which displays the different region names. On the basis of the user selection the radar chart shows information about the accidents in different light conditions and based on the different types of drivers. The user can also zoom in the radar chart using the zoom option displayed on the top right corner of this visualization. The user can also zoom in by holding the cursor(right mouse button) and sliding in and out into the radar chart. The user can also double click on the radar chart to get the original radar chart without any zoom in zoom outs.

CONCLUSION

After the visualization of this project, I have learned many things and I have also had some interesting observations. From the visualizations it can be clearly seen which year sees the most number of accidents by using the cluster map. It is seen that 2019 has the most number of accidents. The number of accidents can also be compared by the regions and the years using the area chart. Hence in which year and in which region the accidents were seen the most can be seen using the area chart. Hence we see that Metropolitan South East and North West show the most number of accidents. Hence extra care should be taken for thm. Also the relationship between the light conditions and type of driver can be concluded from the radar chart.

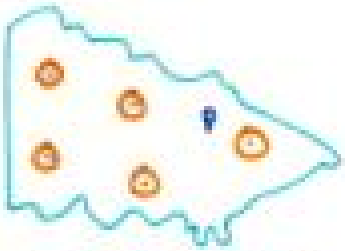
Talking about regions we see a big difference in the values. For the regions Metropolitan North West and South East a huge number of accidents have occured. Hence special consideration and care should be focused on these regions to minimize the number of accidents. The cause should also be identified that are leading to such a high number. The visulisation process hence answered and depicts all the questions that were initially put up.

APPENDIX

SHEET 1 - BRAINSTORM

ROAD ACCIDENTS IN VICTORIA

Overall Accidents



Accidents in 2015

FILTER

on the basis of year, find road accidents in Victoria

Select Year

2014
2015
2016

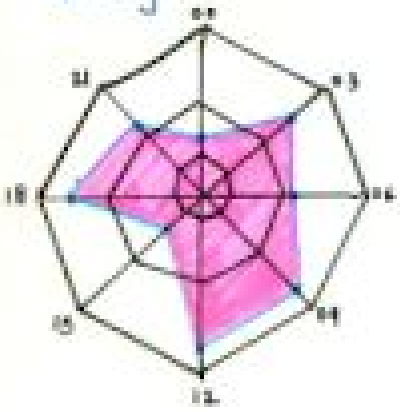
Dropdown

Light's effect on accidents

Select Region

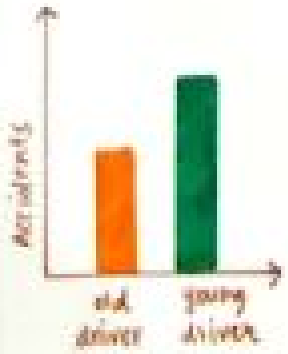
North
West
South

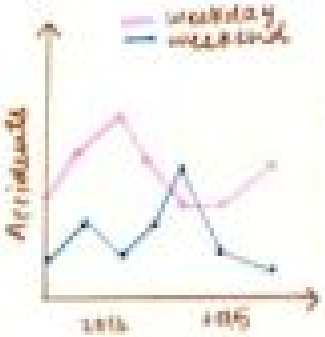
Dangerous Hour

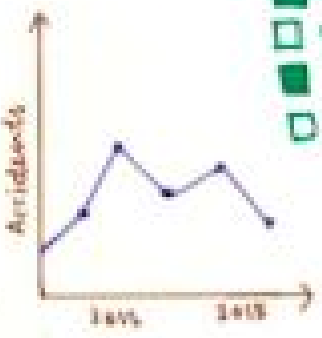


→ Accidents on the basis of light condition of environment

→ The most dangerous hour that sees the maximum no. of accidents

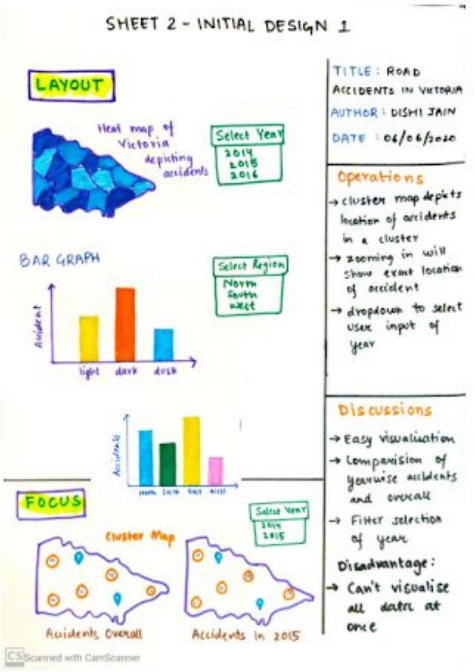




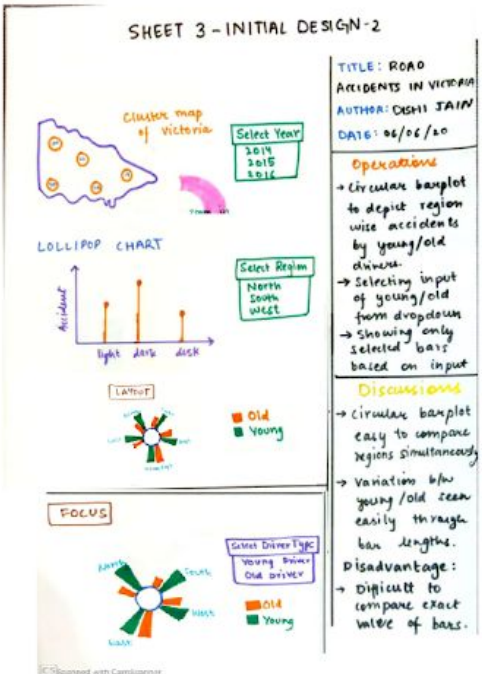


old driver
young driver
weekend
weekday

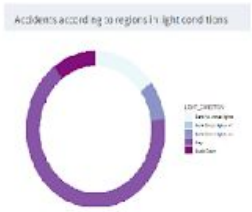
FIVE DESIGN SHEET - SHEET 2



FIVE DESIGN SHEET - SHEET 3



FIVE DESIGN SHEET - SHEET 4



FIVE DESIGN SHEET
SHEET 5