FIT5147 – Data Visualisation Project

VICTORIA CRIMINAL INCIDENTS ANALYSIS

APP LINK

Hoang Viet Do

29780977

Applied Session 7 – Michael Niemann

Table of Contents

Introduction

Criminal incidents are a pressing concern in any community, profoundly impacting the safety and well-being of its residents. In this visualization project, we embark on a journey to dissect the landscape of criminal incidents in Victoria, Australia. Our main objective is to provide an app that facilitates comprehensive exploration and research of the criminal landscape in Victoria across all regions.

At the forefront of our discoveries are the three LGAs—Melbourne, Yarra, and Latrobe—with the highest criminal rates per 100,000 population. Moreover, while statistical models hinted at a potential link between socio-economic factors and crime rates, a visual examination of the data presented a different story. Our visualization showed limited to no discernible patterns in the data. This discrepancy underscores the importance of acknowledging the limitations of our dataset, including biases, the absence of multi-year data, and data obsolescence. This further emphasizes the need for ongoing research and data refinement to arrive at conclusive insights.

Our narrative visualization caters to a diverse and engaged audience. Foremost among our audience are the citizens of Victoria, who are deeply invested in understanding and enhancing the safety of their communities. Moreover, recognizing the unique vulnerability of international students in our region, we strive to equip them with essential insights for their well-being.

Design

The design style is closely aligned with the 5-Sheet Design methodology and is implemented using the *'shinydashboard'* package in R. This design layout features a dashboard header at the top left corner, prominently displaying the application or project title. Below the header, there is a left-hand sidebar, and adjacent to it is the main panel. The initial configuration of the left sidebar includes five distinct tab items: "Total Incident Rate," "Sub-category Analysis," "Map," "Socio-economic Factor," and "About." These tabs serve as entry points, offering a concise overview of the project's analytical focus. The main panel on the right houses a variety of visualizations, which facilitate a comprehensive analysis of criminal incidents in Victoria from various perspectives.

Initially, various design approaches were considered for visualizing the project's findings. Contemplation included the use of diverse plot types, such as bar charts, scatter plots, heat maps, choropleths, and others. The challenge lay in achieving a delicate balance between complex visualization methods while

ensuring clarity and comprehensibility for users. Consequently, the concept of a dashboard emerged as a practical solution. The dashboard accommodates multiple types of plots, input elements for users, result descriptions, and a user guide. For the year-on-year comparison of Local Government Areas (LGAs), a combination of bar plots and additional line plots depicting the total Victoria criminal rate was deemed sufficient for providing accessible insights to users. In the case of sub-category analysis, the objective was to present the top 5 location types, suburbs, and principal offenses rated by criminal incidents. These visualizations needed to be responsive to user inputs related to LGAs and years. Initially, a horizontal bar plot appeared to be a suitable choice. However, as the concept of year-on-year comparison was integrated, a radar chart emerged as an effective solution that met all visualization requirements. Furthermore, the radar chart proved to be more visually engaging compared to the repetitive use of bar charts, potentially capturing, and retaining user attention. In addition to this, a choropleth was included to support spatial analysis. Finally, a parallel coordinate plot was employed to illustrate the relationship between socio-economic factors and the rate of criminal incidents.

In the initial design phase, the dashboard for the 5-Sheet Design featured several components and elements. These included a bar plot for year-on-year comparisons of the criminal rate per 100,000 population, combined with the total rate for all of Victoria. Additionally, there were visualizations illustrating the total criminal incidents over two consecutive years, with color-coding indicating improvements or deteriorations. A proportional horizontal bar in the middle of the dashboard presented information about the criminal case status and served as a visual divider for the dashboard content. Furthermore, the lower part of the dashboard displayed three horizontal bar plots for sub-category analysis, focusing on location types, suburbs, and principal offenses. Each element of the dashboard was controlled by user inputs, specifically the Local Government Area and the year. On hover, these plots and figures offered additional information. However, during the implementation phase, it became evident that the initial design was overly complex, lacking a user guide or concise result explanations. This complexity could distract general users and hinder understanding, particularly for those without technical backgrounds.

As a result, modifications were made to streamline the dashboard and achieve the primary goal of providing detailed yet easily comprehensible insights. Year-on-year crime rate per 100,000 population and sub-category analyses were separated into two different tabs. The first tab featured a bar plot with a trendline, and additional information appeared on hover, providing a user guide and guidance on interpreting the results. Hovering on bars revealed details about the chosen Local Government Area, including the year, crime rate, and the percentage compared to the total Victoria rate. Hovering over the trendline

displayed the year and the crime rate for all of Victoria. The tab was structured with user input on the tab panel, the main plot in the main panel, and user instructions in a text box below.

The second tab for sub-category analysis introduced a significant change from the initial design. While it retained elements such as user input, a user guide text box, and the main plot, it departed from horizontal bar plots for each sub-category and adopted radar charts instead. The radar chart offered a broader range of visualization options, enhancing user engagement. It allowed for multiple chart elements for each year, while still focusing on comparing the top 5 location types, suburbs, and principal offenses in terms of the total number of criminal incidents. Users could select the year (ranging from 2014 to 2023) as an additional input. Tooltips provided more information, and sub-category labels could be rotated to ensure legibility.

The second design sheet, dedicated to map and spatial analysis, closely resembled the initial design with minimal differences. It utilized a choropleth map for analysis, focusing on the Victoria region. Each Local Government Area was distinguishable by border colour, and the choropleth provided a comparison of criminal rates per 100,000 population among LGAs for the same year. The saturation of one colour (in this case, a green palette) indicated the criminal rate, with darker colours signifying higher rates. This spatial analysis aimed to enable users to compare rates among different LGAs within a single year, in contrast to the first tab's objective of comparing rates across different years for a specific LGA. To facilitate this comparative analysis, an additional input was introduced: Police Region, allowing users to examine larger-scaled regions beyond individual LGAs. Like the first two tabs, the choropleth map was interactive, featuring tooltips and a user guide text box.

The third design incorporated a parallel coordinate plot to analyse all three socio-economic factors and crime rates concurrently. It featured four y-axes, with each axis representing a socio-economic factor and one for the crime rate per 100,000 population. Due to data limitations, the plot remained static. Each line represented an LGA, classified into two groups based on crime rates above or below the average, each marked with distinct hues. Additionally, the three LGAs with the highest crime rates per 100,000 population—Melbourne, Yarra, and Latrobe—were assigned unique hues, allowing users to assess how socio-economic factors influenced crime rates in these regions. As this plot might be less common for non-technical users, a user guide and a concise discussion of the plot results were included to enhance the user experience.

Implementation

To translate the design concept into a functional application, R Shiny was employed as the implementation platform. While D3 offers enhanced capabilities for user interactions and the creation of visually appealing visualizations, the decision to use R Shiny was primarily based on familiarity with the tool. Implementing D3 without sufficient knowledge would have risked suboptimal results.

Several packages were essential for developing the final app, aligning with the project's objectives. These packages included *tidyverse, shiny, dplyr, shinydashboardPlus, shinydashboard, plotly, leaflet, sf, colorspace, htmltools, reshape2,* and *janitor*. The tidyverse package was employed for data preparation prior to analysis. It addressed the need to format Local Government Area (LGA) values in proper case, as data was imported with uppercase values. While this formatting didn't impact the app's functionality, it significantly improved the user experience. The *mutate()* and *str_to_title()* functions were used for this purpose. Additionally, the leaflet package facilitated the creation of a map for spatial analysis. To differentiate LGAs as distinct polygons on the map, the sf package and sf data format were utilized. Data processing for the map data mirrored that of the crime data. Lastly, the shiny package was central to building the application for data visualization. Various shiny themes, including *bootswatch*, were available for app decoration, but the adoption of *dashboardPage, dashboardHeader, dashboardSidebar,* and *dashboardBody* components from the *shinydashboardPlus* and *shinydashboard* packages allowed for more flexible app design with an external CSS file. This approach was favored over the conventional use of *fluidPage*.

The app's introduction and objectives were presented in a text box within a separate tab item, accompanied by information about the author in a text box below. Upon initial browsing or app reload, a welcome page featuring the app title and an image would appear, as demonstrated below. The tab panel, containing five different tab items, was displayed on the left side, allowing users to switch between them. The user interface (UI) and server setup were configured with some reference to programming exercise #2 and primarily based on Shiny Dashboard Rdocumentation. A combination of *fluidRow()*, *column()*, *mainPanel*(), and *sidebarPanel()* components maintained a consistent design. Two to three *selectizeInput* elements were used to facilitate user selections for Year, Local Government Area, and Police Region values.

Within the first tab item's bar plot, a reactive function in the server filtered data to reflect the user's input for LGAs. The bar plot elements were equipped with tooltips that provided additional information upon hovering, enhancing user

understanding. The colour of the plot elements, including bars and lines, was determined automatically. Given the substantial variance between crime rates of individual LGAs and Victoria, two different y-axes were employed for LGA and Victoria values. Furthermore, the app featured a dynamic plot title that reacted to user input of LGAs, achieved through the use of the *textOutput* interactive element positioned above the plot, replacing the default title of the plot.

To facilitate sub-category analysis, three distinct datasets representing location types, suburbs, and principal offenses were filtered within the reactive functions to respond to user input regarding LGAs and years. The implementation of radar charts, realized through the use of *plot_ly(),* presented a noteworthy challenge. To address this challenge, reference was made to Plotly documentation for radar charts (Radar Charts in R, n.d.).

Spatial analysis implementation proved to be the most intricate among all visualizations due to the data structure and the utilization of the *leaflet* package. Initially, the *sf* polygon data had to be transformed into a format suitable for Google Maps plotting using *st_transform()* and *st_as_sf*(), which is compatible with the leaflet package. The filtering process to ensure data responsiveness to user input, encompassing *Police Region, LGAs, and Year*, was achieved through the application of a reactive function. The decision to use a green colour palette for the choropleth map from personal preference, although it is essential to note that the visualization may not be colourblind-friendly. Given that the differentiation of crime rates relies on colour saturation, inherent to choropleth maps, this limitation could not be entirely mitigated. Consequently, insights about three regions with the highest crime rates per 100,000 population were provided in the user guide text box.

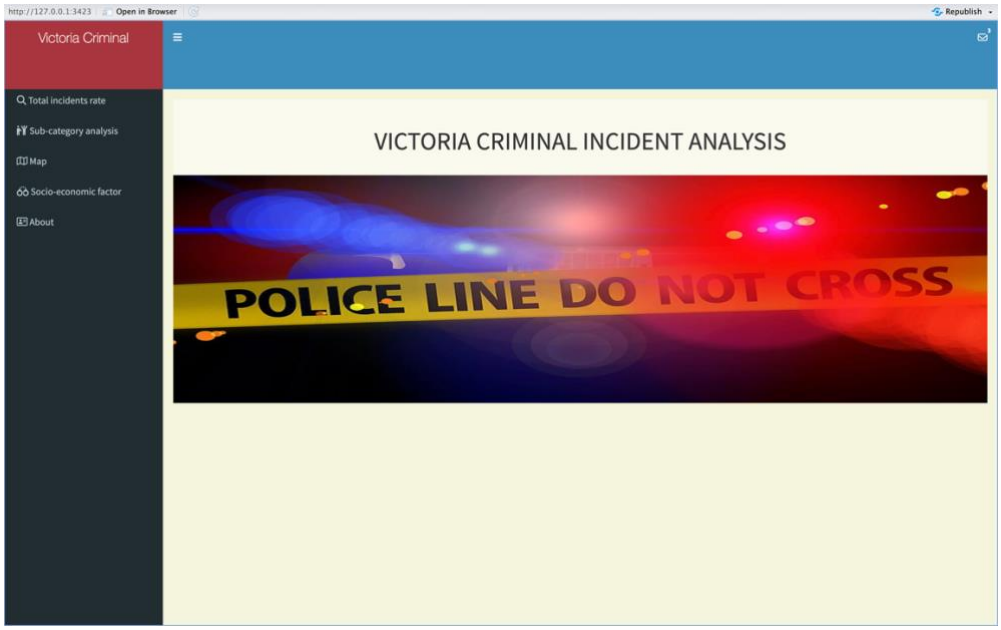
In the case of the parallel coordinate plot, the merging of three distinct datasets pertaining to socio-economic factors, including the satisfaction of living standard index, socio-economic disadvantage score, and education and occupation disadvantage index score, was executed. Subsequently, LGAs were classified based on the mean crime rate per 100,000 population, aligning with the inherent nature of the parallel coordinate plot. Additionally, three regions with the highest crime rates were distinguished by a unique colour, such as yellow in this instance, to enable users to identify potential connections between socio-economic factors and crime rates. Since the plot is static, attributed to the availability of socio-economic data for only one year (2016 for satisfaction of living standard and 2021 for socio-economic disadvantage and education and occupation disadvantage index), the implementation process was relatively straightforward.

Regarding the data itself, numerous datasets with multiple sheets were employed for this project. Consequently, meticulous data processing and wrangling were conducted, as elaborated in detail in the Data Exploration Project (DEP) section. While the utilization of tabular data might be perceived as simplistic, it aligns with the purpose of developing a useful app capable of delivering comprehensive insights to the intended audience.

In terms of user interaction, the principal interactivity revolves around filters based on selection inputs. While these interactions may not be characterized as highly sophisticated, they effectively assist users in seeking the answers they seek concerning crime rates in Victoria.

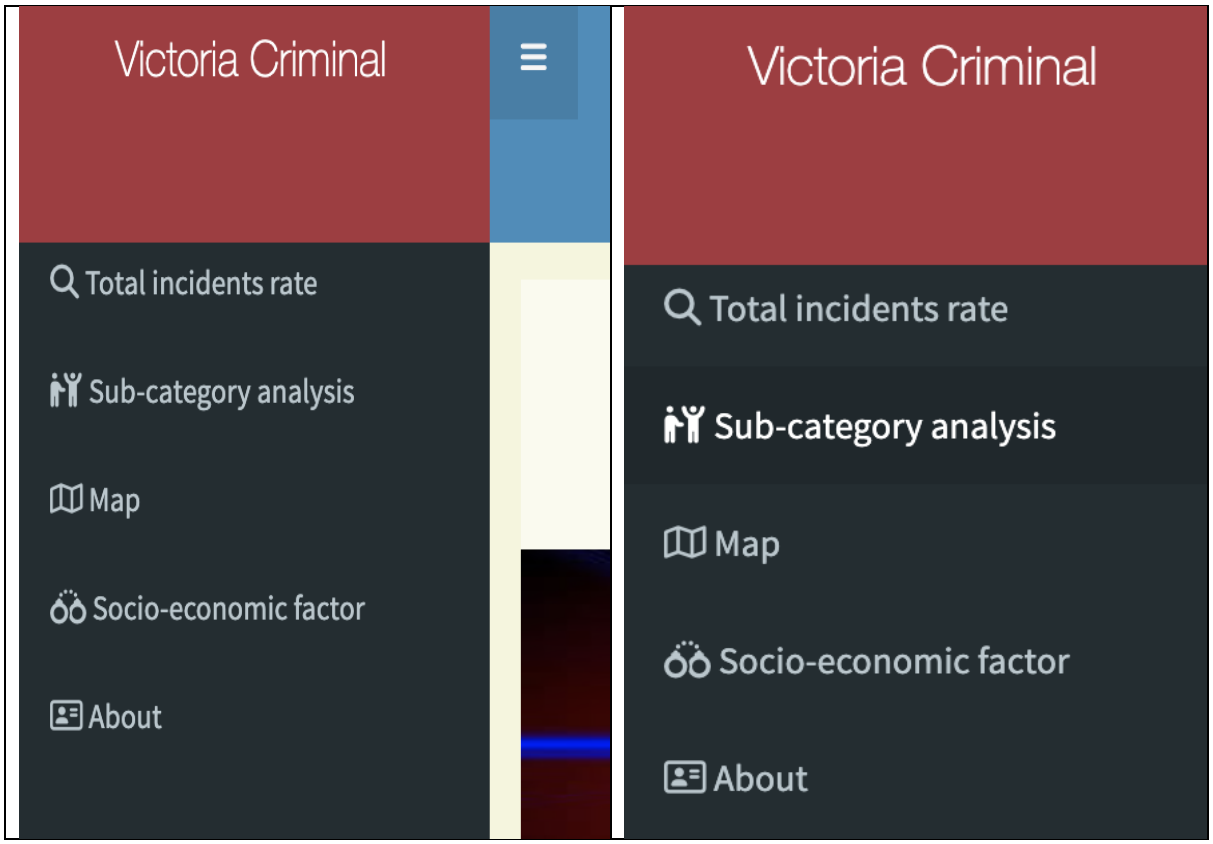
User Guide

For users with access to local files, launching the application is straightforward. They can initiate the application by simply clicking on the 'Run App' tab. This action triggers the automatic loading of the required libraries and data, subsequently displaying the application. Alternatively, users can access the app via the provided app link. Upon running the application, users are welcomed to the main page, featuring the project title accompanied by an image.

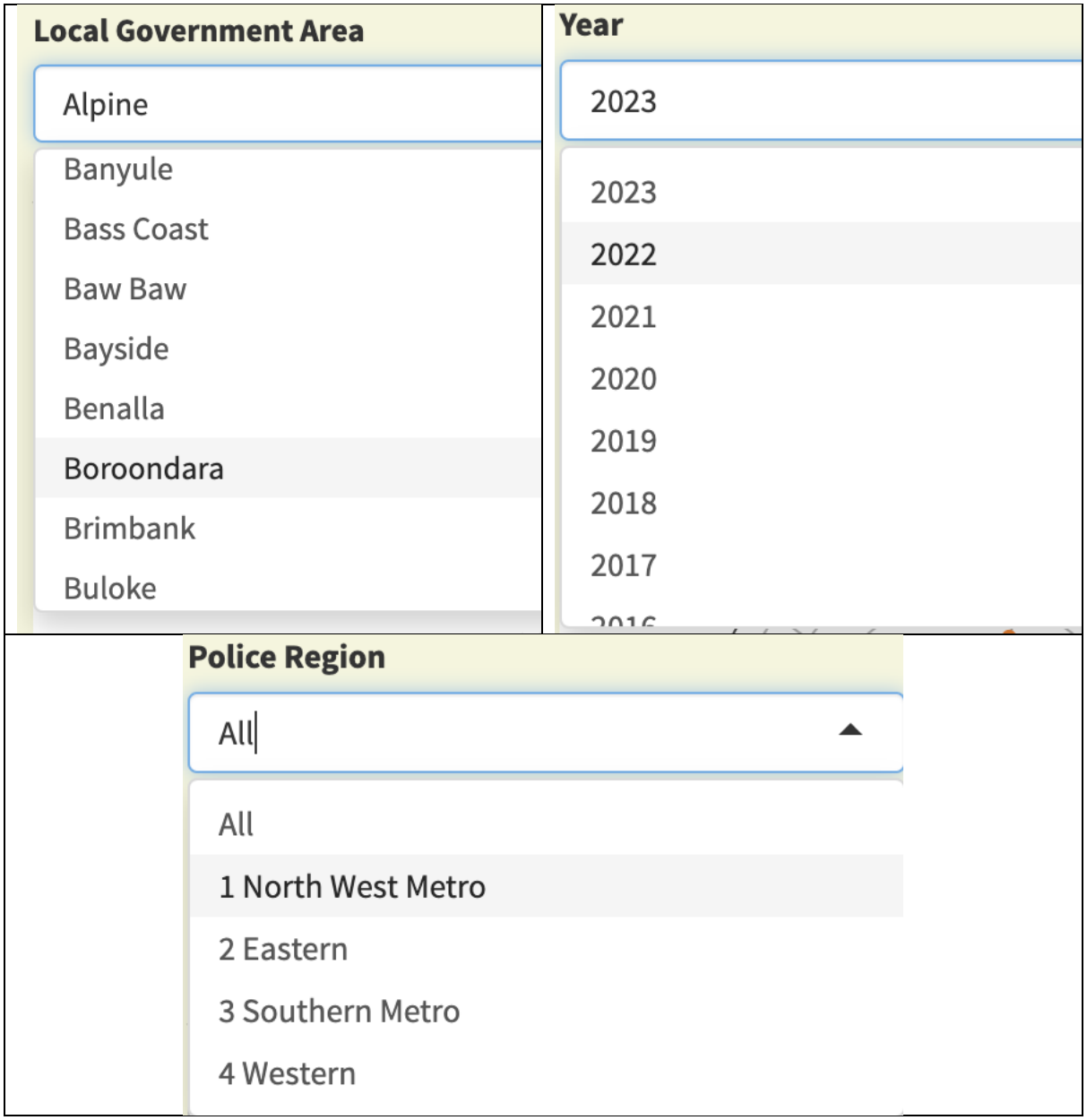


The tab panel is conveniently situated on the left side of the app. To enhance user experience, it can be minimized by clicking on the three-line button, thus expanding the main interface. This functionality is particularly useful if certain

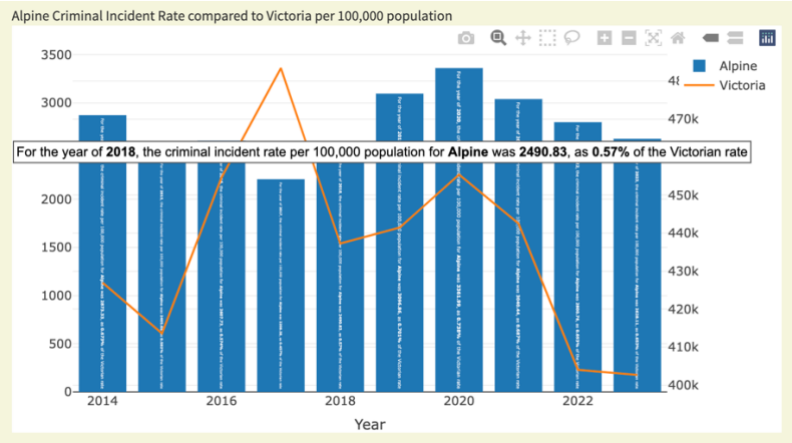
plots are partially obscured due to their size. Users can effortlessly navigate between tabs by clicking on the tab corresponding to their area of interest.

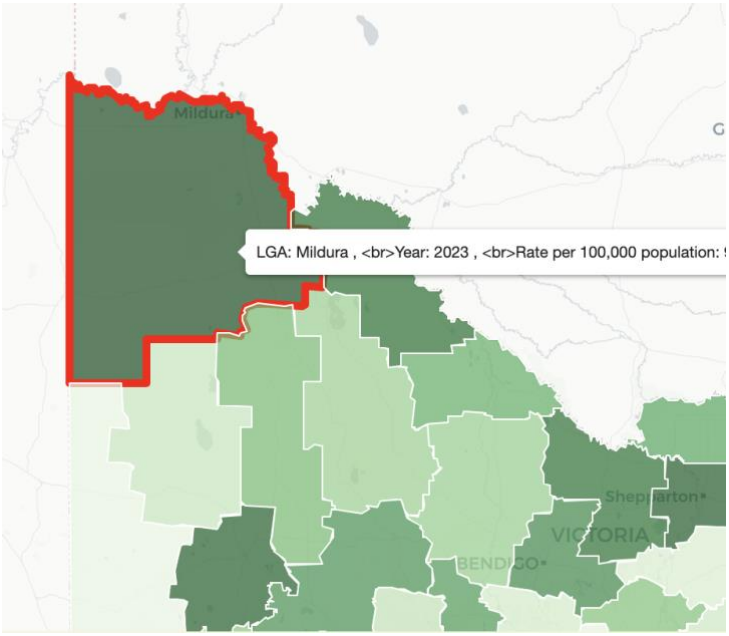


Each tab incorporates a drop-down box, affording users the option to select their focus of investigation. Their selection dictates the presentation of insights tied to the chosen Year, Police Region, LGAs, or the entirety of data.

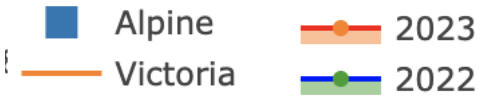

Local Government Area
Alpine
Banyule
Bass Coast
Baw Baw
Bayside
Benalla
Boroondara
Brimbank
Buloke
Year
2023
2023
2022
2021
2020
2019
2018
2017
Police Region
All
All
1 North West Metro
2 Eastern
3 Southern Metro
4 Western

For an enriched understanding of the visualizations, users can hover over tooltips on plot elements, revealing additional information related to the specific plot.





Furthermore, in the case of bar plots and radar charts, users have the option to de-select components directly on the legend by clicking on the specific element they wish to remove.



In the event that users encounter challenges in comprehending and interpreting the plot results, a valuable resource is at their disposal—the "Instruction for users" text box. Alternatively, users can seek further clarification by reaching out to the author via email, with contact details available in the "About" tab.

Instruction for users

- These interactive radar charts enable users to assess **Year on Year** total incidents across the top 5 items in *3 distinct categories* .
- These three radar charts display the *top 5 location types* , *top 5 suburbs* , *and top 5 principal offenses*, respectively, for the selected Local Government Area and year.
- Feel free to hover over data points for additional details. User can also rotate the axis by hovering over the titles on the left side of the chart.

Conclusion

The interactive visualizations provided an effective means to delve into the dataset and address the project's set of fundamental questions, which were as follows:

1. Are certain areas in Victoria more prone to criminal activity than others?

2. Do specific socio-economic indicators correlate with higher crime rates in certain areas?

The application readily facilitates the response to the initial question. It reveals that Melbourne, Yarra, and Latrobe stand out as the three most high-risk areas in Victoria, which may contrast with common perceptions often influenced by media portrayals. Additionally, the exploration of principal offenses offers valuable insights into preventive measures, such as safeguarding motor vehicles in secure locations, particularly in regions with high incidents of vehicle theft. The findings from this analysis can partly influence individuals' decisions regarding rental or property purchase locations for enhanced safety

However, the response to the second question may not fully meet expectations. Several factors contribute to this, as elucidated in the report, including data quality, limitations, and the chosen analytical approach. Therefore, future projects should emphasize a more meticulous selection of data sources. This would entail exploring different data sources with broader temporal coverage and reduced bias.

This project has served as an extensive exploration of R Shiny. While there is room for improvement, the results achieved are satisfying, considering the constraints posed by the dataset, as discussed in the Data Exploration Project. Nonetheless, this project effectively addresses the essential questions it aimed to answer, and its thematic approach is captivating, capturing the audience's attention. The underlying concept of the project was to distil and present only

the critical aspects of the data, which explains the use of four distinct visualization options. Consequently, the project contributes to a refined set of design sheets, incorporating feedback from the tutor, enhancing overall quality.

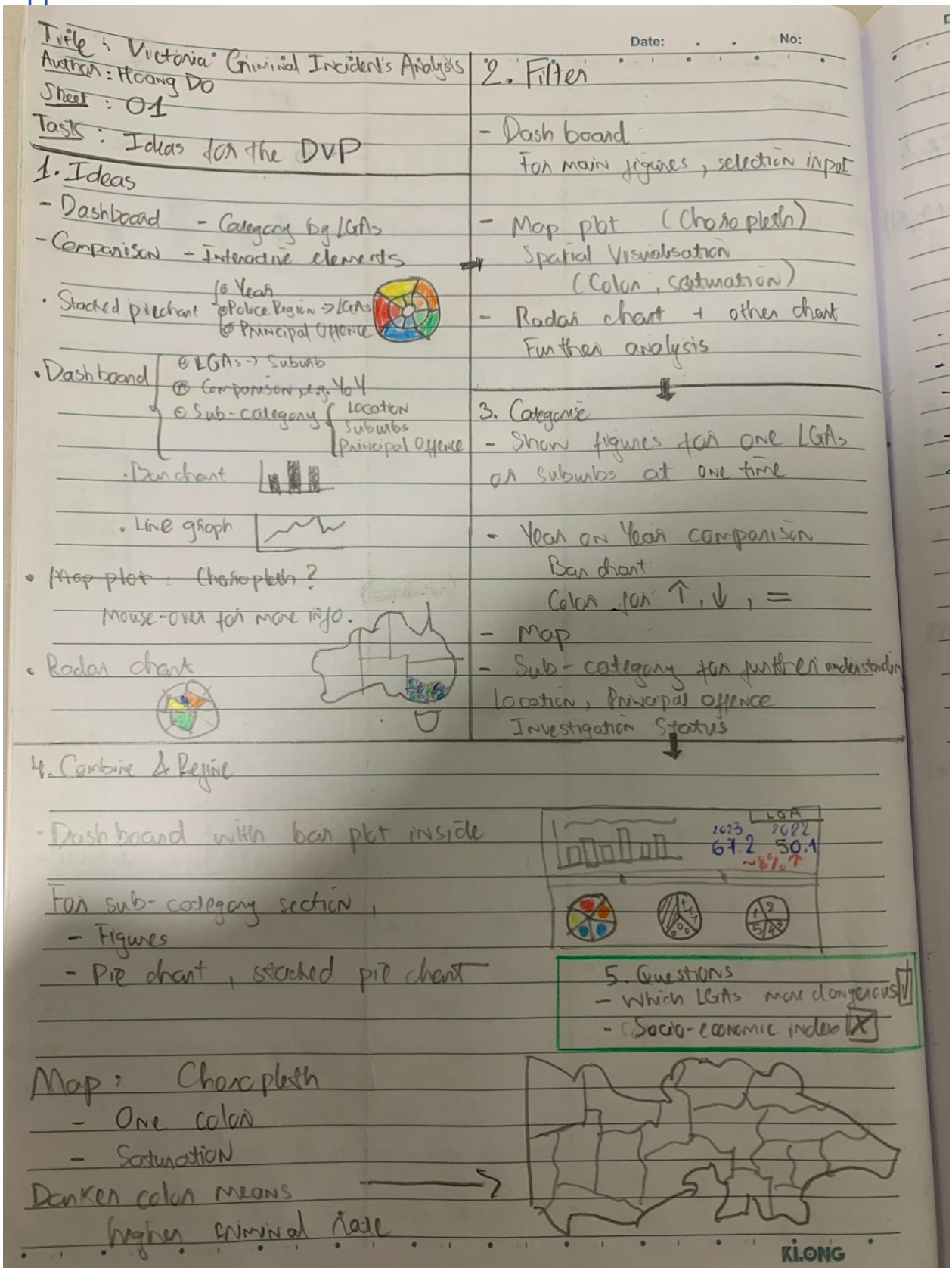
Bibliography

Chang, W. (2021). Shiny Dashboard. Retrieved from RDocumentation: https://rdocumentation.org/packages/shinydashboard/versions/0.7.2

Plotly. (n.d.). Retrieved from https://plotly.com/r/radar-chart/

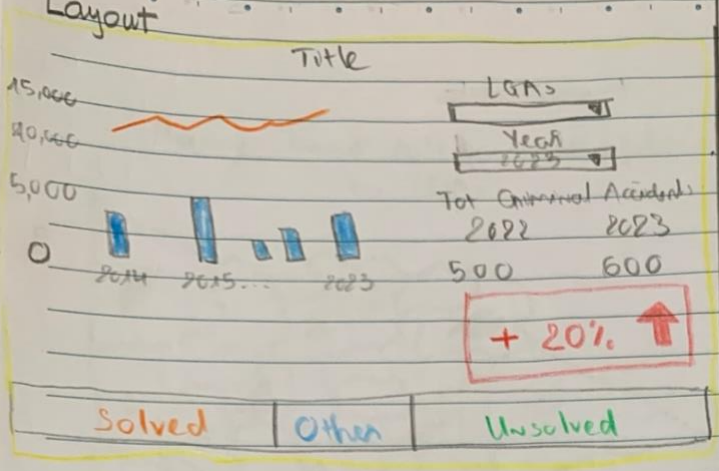
Wickham, H. (2020). Mastering Shiny. Retrieved from Mastering Shiny: https://mastering-shiny.org/index.html

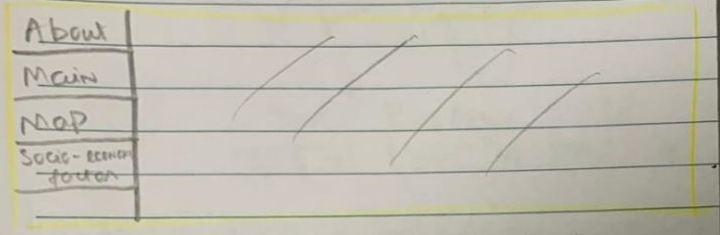
Appendix


Title : Victoria Criminal Incident's Analysis
Author : Hoang Do
Sheet : 01
Tasks : Ideas for the DVP
Date:
No:
1. Ideas
- Dashboard - Category by LGAs
- Comparison - Interactive elements
• Stacked piechart
Year
Police Region → LGAs
Principal Offence
• Dashboard
LGAs → Suburb
Comparison e.g. Y-o-Y
Sub-category
Location
Suburbs
Principal Offence
• Bar chart
• Line graph
• Map plot : Choropleth ?
Mouse-over for more info.
• Radar chart
2. Filter
- Dashboard
For main figures, selection input
- Map plot (Choropleth)
Spatial Visualisation
(Color, saturation)
- Radar chart + other chart
For further analysis
3. Categorie
- Show figures for one LGAs or Suburbs at one time
- Year on Year comparison
Bar chart
Color for ↑, ↓, =
- Map
- Sub-category for further understanding
location, Principal offence
Investigation Status
4. Combine & Refine
• Dashboard with bar plot inside
For sub-category section,
- Figures
- Pie chart, stacked pie chart
LGA
2023
67.2
2022
50.1
~8% ↑
5. Questions
- Which LGAs more dangerous ✓
- Socio-economic index ✗
Map : Choropleth
- One color
- Saturation
Darker color means
higher criminal rate
KLONG

Date: . . No:

Layout





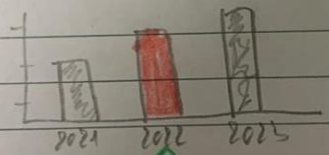
Title: Victoria Criminal Incidents
Author: Hoong DO
Sheet: 02
Task: Dashboard Design

Operations

- Criminal incident rate from 2014 to 2023, compared to total of VIC
- Allow user to choose LGA & year
  e, Year can also be chosen on bar chart
- Main box with colon for Y/Y comparison
- Hover the mouse on elements for further information

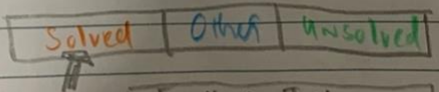
Focus

① Intuitive



Click on the bar to choose
All other bars blurred out

② Mouse Hovering



→ Hover the mouse on elements

③ Tab switching : Click on tab to switch

Discussion

+ve

① Allow users give interactive attention by selective input & color switching
② Also provide handful information & summarize

-ve

① Might be complex to fully understand all functions
② Too much info.
⇒ May cause distraction

KLONG

Date: . . No:

Layout

Title: Victoria Criminal Incidents
Author: Hoong Do
Sheet: 03
Task: Spatial Analysis

Average Incident rate per 100,000 population by LGAs

Year | Police Region | LGA
2023 | All | All

1,631.7 — 52,631.2

Operations
- Each LGA or Police region will have different saturation to indicate incidents rate
  - Darker → ↑
  - Lighter → ↓
- Selection of LGA or Police Region will zoom in to that block on map
- User can get more info. by hovering mouse to a region

Focus

(1) Saturation — From light to dark to indicate rate of incidents

(2) Interactive — Box to choose input

(3) Mouse Hover — Hover mouse on LGA for further info.

Discussion

⊕ VE
- Interactive elements to engage users
- Use of saturation on map allows for easier comparison of rates

⊖ VE
- Cannot do year on year comparison

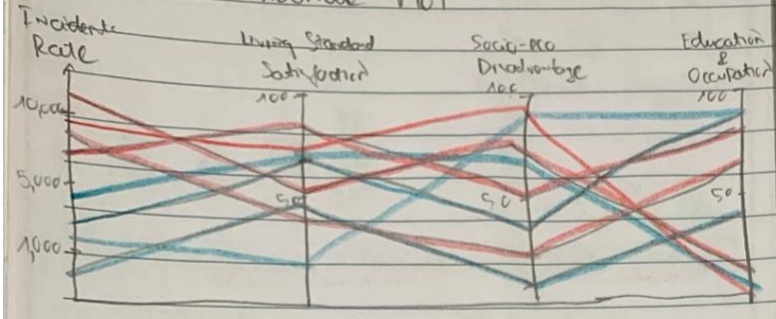
KI.ONG

Date: . . No:

Title: Victorian Criminal Incidents
Author: Hoang Do
Sheet: 01
Task: ~~Sub-category Analysis~~

Layout

① Socio-economic factors
② Parallel Coordinate Plot



② Sub-category

Top 5 location type

House	104
Street	90
Driveway	100
Other	50
Market	35

Top 5 suburbs

Melb	16,052
Southbank	5,122
Carlton	3,227
Docklands	2,789
North Melb	1,521

Operations

- Interactive with other elements of the dashboard
- Using color for easier V&Y comparison & classification
- User can get more info. by hovering the mouse to a bar or line

Focus

① Color
- Parallel Coordinate Plot
  - : Red -> higher than average
  - : Blue -> lower than average
- Sub-category bar plot
  - : Higher than last year
  - : Lower than last year

② Mouse Hover

House 104

Hover mouse on the bar for more info.

Discussion

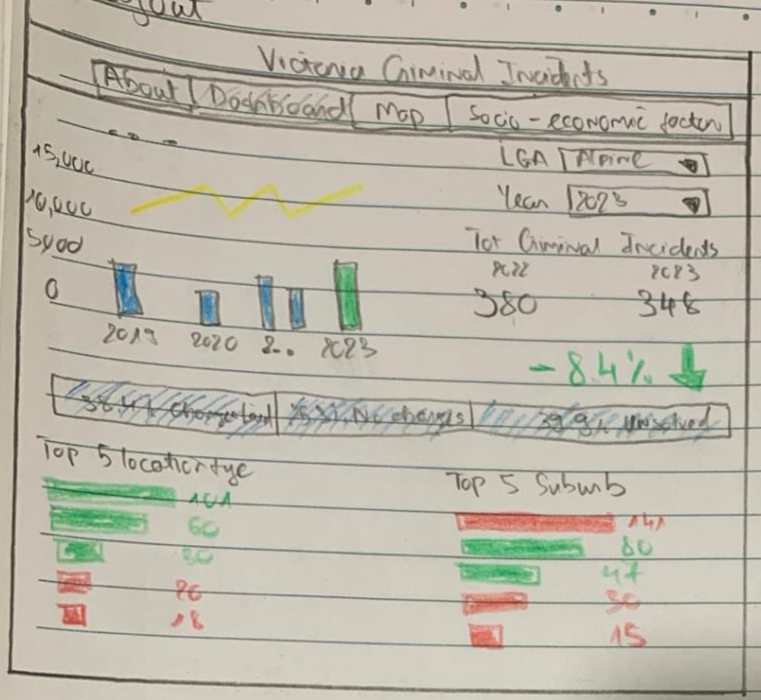
+ve
- Simple but effective visualisation
- Able to incorporate a lot of information

-ve:
- Parallel Coordinate Plot may be hard to understand w/o proper guidance

KLONG

Date: . . No:

Layout

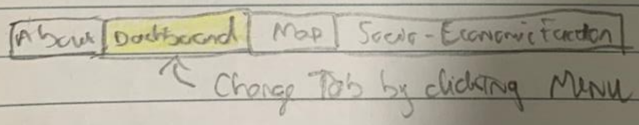


Title: Victoria Criminal Incidents
Author: Hoang Do
Sheet: 05
Task: Build final interface

Operation

- User select MENU tab to move to other page of the dashboard
- Each page provides different information based on the dataset
- Category: Provides top 5 info

Focus

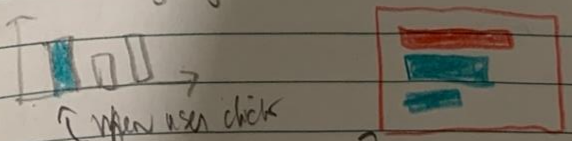


Change Tab by clicking MENU

- MENU -> Page

When user select the menu, page will be changed

- Sub-category



When user click the bar

This plot section will change too

Discussions

- Visualisation graph might be changed based on implementation
- Aim to provide info to audience
- The board will be designed by R Shiny
- Data wrangling & clean by R studio

KLONG