Crime Statistics in Victoria

Abhishek Sinha

31322743

Tutorial No: FIT5147 Lab 03-P1

Tutor Name: Bruno Luis Mendivez Vasquez

Introduction:

In today’s world, crimes have become an inseparable part of our lives. Every day we read about them, hear about them and some of us might be even involved in at least one of them during our lifetime. Victoria is no different.

Being cautious and improving safety are not just simple instruction anymore. The community needs to be aware of the crime scenario in their area so as to take precautionary measures whereas on the other side the authorities need to be aware of the crime rates in their jurisdiction so as to assess their effectiveness.

This report is motivated by an article published in the ABC News, which said that the Victoria’s total crime rate measured as incidents per 100,000 people rose 1.6 per cent over the 12 months to September, halting a three-year downward slide as police recorded an extra 14,367 incidents compared to the same period the previous year [1].

The main goal of this report is to present the situation of crime in Victoria in the past decade to answer the primary question, ‘Is Crime in Victoria increasing?’. And if it is, which LGA (Local Government Area) of Victoria witness the highest rate of crimes and which crime is the most often reported. Additionally, the report also looks at the locations that are more prone to crime incidents and what is the investigation status of the recorded crime incidents. To answer these questions, we need an authentic data source that can provide us with a rich data set on crime for the past decade. For this report the data has been taken from Crime Statistics Agency (CSA) website [3] which maintains a database sourced from Victorian Police’s Law Enforcement Assistance Program (LEAP) database.

This report targets a broader audience. Crime statistics is one measure the community uses to judge our relative safety and security. They show the rate of recorded crime across local government areas or regions and can highlight trends such as increases or decreases in types of crimes and how and where they are committed. The government can use crime statistics as an indicator of the effectiveness of criminal justice policies, and the Victoria Police can use this report to help streamline its resources [2].

Design:

This report is prepared using the Five Design-Sheet (FDS) Methodology. The Five Design-Sheet (FDS) Methodology was introduced as a way of formalising sketching and lo-fidelity prototyping to allow visualisation designers explore their ideas without worrying on the technical aspects. As suggested, there are 5 sheets in this methodology and this report will touch on what each sheet is about and how were they being used to develop the visualisation report for Victorian Crime Statistics.

Sheet 1: *The ideas sheet*

To begin the design process to visualise and present my findings, the first and initial step is to relate the research questions with the possible plots that can be visually attractive and informative to present the analysis in a simple and interactive way. The ***primary research*** question that this report is based on is to analyse the situation of crime rate trend in Victoria in the last decade. Looking at this analysis the idea is to identify the LGAs with high crime rate in the previous decade and thus help the target audience by providing them the status of crime rate in their respective suburbs and LGAs. ***Secondly***, I also want to look at the crime rates across offence division i.e., the crime rates across different type of offence recorded by the Victorian Police in the last decade. ***Additionally***, I also want to present to the readers of this report, my findings about identifying locations prone to higher crime rate in a particular LGA and the current investigation status of the recorded crime incidents.

To present these ideas and my analysis, I need to first decide on the maps and an efficient design for my interactive tool that will not only help the user to interact with the data in a seamless way but also make the analysis genuine and informative.

I start the process with thinking about the possible plots that can help me present my analysis in a visually appealing way and at the same time be a correct plot for the analysis.

My initial **ideas** about the possible plots were:

1. Choropleth Leaflet Map, Choropleth Static Map
2. Histogram, Bar Plot
3. Line Plot
4. Bubble Graph
5. Tree Map
6. Sunburst Graph
7. Coxcomb Plot

Although all these maps are visualising appealing and correct in different situations, I needed to **filter** them out but also being in line with the research questions.

To answer the primary question of identifying the trend in Crime Rate across LGAs in Victoria in the last decade, I choose Choropleth leaflet map to identify the LGAs or suburbs that saw the worst of crime rate and LGAs that did a good job to curb the crime rate. This visually appealing and interactive plot along with the legend can tell the user, about the situation of crime rate across LGAs.

Secondly, to identify the number of recorded incidents across different offence type, bar plot and coxcomb plot were my possible choices. Coxcomb Plot is a visualising appealing plot to present the relationship and also reduce the space taken by bar plot thus reducing clutter. But on further reading about this plot, I realised that this kind of circular plots suffers from a very serious issue interpretability. It is very hard to interpret the plot correctly and people often end up making subjective analysis using this plot. As a wedge gets a larger radius, the area taken up by the wedge increases disproportionately to the radius increase. Since area also factors into how people judge the size of a chart component, wedges with larger radii are much larger than they are. Thus, I choose bar plot as it not only shows the count of the cases overtime across different offence type but also let the user identify a possible trend.

To identify the locations in a suburb or a LGA that see a high number of recorded offences and answer my third question I choose to go ahead with the tree map. With the inputs from my tutor, I realised that the number of levels that I want to present on bubble plot or tree map is huge and this will result into a possible clutter which will compromise the user experience and will also lead to false and broken findings. With this in mind I choose Tree map as the better solution as square brackets will minimize the possible clutter and at the same time keep the information.

Additionally, to answer my final question about the investigation status of the recorded incidents, I had two options for the plots. One was using sunbursts plot to present the status in a visually unique way differentiating it from other plots and the other one was bar plot which would have been a similar plot as the one above and caused possible visual mistakes. So, I choose Sunbursts plot.

Thus, of the filtered plots that I will proceed my analysis with, I can **categorize** them as Choropleth Map being used for my spatial data, Bar plots for my tabular data and tree map for data with hierarchy.

While deciding the possible ideas for the narrative visualisation to communicate the key findings to the target audience, some questions still remained:

1. Is showing a leaflet map necessary to deliver the message? Is presenting suburbs on leaflet plot makes it more informative or highly cluttered?
2. Is sunburst the best plot to complete my narrative and present a complete story?

Sheet 2: *Alternative Design sheet*

Now that plots are finalized, the next step is to start with the first design sheet towards the final realisation tool. This initial design consisted of using leaflet plot, a bar plot to visualise recorded incidents trend across offence type, a tree map to plot counts across multiple locations and a sunburst plot with bar plot to present investigation status of the recorded incidents as sunburst plot alone cannot tell the trend in data.

The user can hover over the map to look at the normalised rate of recorded incidents per 100,000 population across LGAs. The user can also select a LGA from the drop-down menu and a type of offence to view the trend in the recorded cases overtime. Further the user can select a year to look at the cases by location in a LGA in a particular year. Further selecting a year and a LGA the user can look the status of cases.

This design sheet is highly interactive and easy to follow. With detailed and informative plots this sheet can present a complete data story to answer possible research questions. But this particular design sheet was not chosen as the final design sheet as though being detailed it is a bit cluttered with multiple plots. Secondly, a map plot alone offers a limited answer to the primary question of trend in recorded incidents. As choropleth map is clustered to include 79 LGAs it becomes difficult to identify small LGAs and the scenario there. Also, by looking at the legend I can tell which LGA saw the highest crime incidents, but I will not be able to say anything about trend.

Sheet 3: *Alternative Design sheet*

This design sheet builds on the first design sheet and tries to overcome previous negative and present a new solution. This design sheet consisted of using leaflet plot along with a bar plot to answer the primary question, a bar plot to visualise recorded incidents trend across offence type, a tree map to plot counts across multiple locations and a combination of three individual bar plots to present investigation status of the recorded incidents leaving out the sunburst plot. The bar plot can handle both the comparison and pointing out the trend in data. Also, as a possible solution to the congestion issue in Design 1, I decided to try multiple tabs instead of one. All the individual research questions can be answered in individual tabs.

The user can now easily navigate between tabs to analyse the data. They can still hover over the map to look at the normalised rate of recorded incidents per 100,000 population across LGAs. The user can also select a LGA from the drop-down menu and a type of offence to view the trend in the recorded cases overtime. Further the user can select a year to look at the cases by location in a LGA in a particular year. Further selecting a year and a LGA the user can look the status of cases.

This design sheet is highly interactive and user friendly. It’s easy to navigate as well. With detailed and informative plots this sheet can present a complete data story to answer possible research questions. But this particular design sheet was not chosen as the final design sheet. This is because using multiple tabs can help to declutter and analyse the present question more easily but makes comparison really difficult. Secondly, for research question four to identify the conviction status using three separate maps together makes the analysis messy.

Sheet 4: *Alternative Design sheet*

The final design sheet looks at whether to use a leaflet plot characterise by LGAs or by suburbs. There are many issues with going down to suburbs as first it will make the plot highly cluttered. It will be really hard to identify any difference between suburbs given the limitations of the plot size as well. Additionally, finding a correct shapefile is challenging as there are a lot of missing values in either the shapefile or my data. In the last design sheet, there was a problem of multiple plots for individual investigation status which was not visually appealing. To overcome this issue this design sheet now let the user decide using a drop-down menu the investigation status of choice. Thus, this design sheet was discarded as well.

Sheet 5: *Realization*

The final design sheet is an improvement to all three design sheets which incorporates the positives of all the design sheets while maintain a trade off with the negatives. The final design will use a single tab where the user can scroll down to view multiple plots. Using a single tab allows user to compare and get a complete data story.

To answer the primary question a combination of map and bar plot will guide the user to identify a LGA’s performance and also look at its trend in the last decade. Plot 3 will help user to answer research question 2 of identifying trend across different offence type and identify the most common offence in an LGA. Plot 4, a tree map will let user study the number of recorded cases across multiple location types and identify vulnerable spots in an LGA. Finally, plot 5 will help user to know about the investigation status of the recorded incidents in different LGAs overtime. User will have a lot of interactivity with the visualisation tool with multiple drop-down menus to filter data accordingly and also hover cover plots to get underline figures.

Implementation:

In order to realise my design for my visualization report and provide the user with an interactive tool to interact with the data and make their own analysis, I need to first select an interactive tool. For this report I choose R shiny dashboard. R shiny is an interactive tool like Tableau, that provides the user with the option to interact with the data and create their own analysis or repeat one.

To create my interactive tool I used the following set of libraries to help me with data reading, wrangling, visualising and rendering on web:

1. library(leaflet)
2. library(ggplot2)
3. library(dplyr)
4. library(rgeos)
5. library(maptools)
6. library(ggmap)
7. library(broom)
8. library(shiny)
9. library(shinydashboard)
10. library(plotly)
11. library(tidyverse)
12. library(readxl)

The counting methodology used in this report is identified as Crime Incident. A recorded criminal incident is a criminal event that may include multiple offences, alleged offenders and/or victims that is recorded on the LEAP database on a single date and at one location.

The data for this analysis has been taken from the Crime Statistics Agency website. The ‘Crime Incidents year ending March 2020’ data table was taken which is a very detailed data table and includes incidents data filtered down to postcodes and suburb along with offence type for a more comprehensive picture.

In Figure 1 we can look at the structure of the downloaded dataset in excel. The data is divided across multiple sheets which needs to be read in carefully and joined further for analysis.

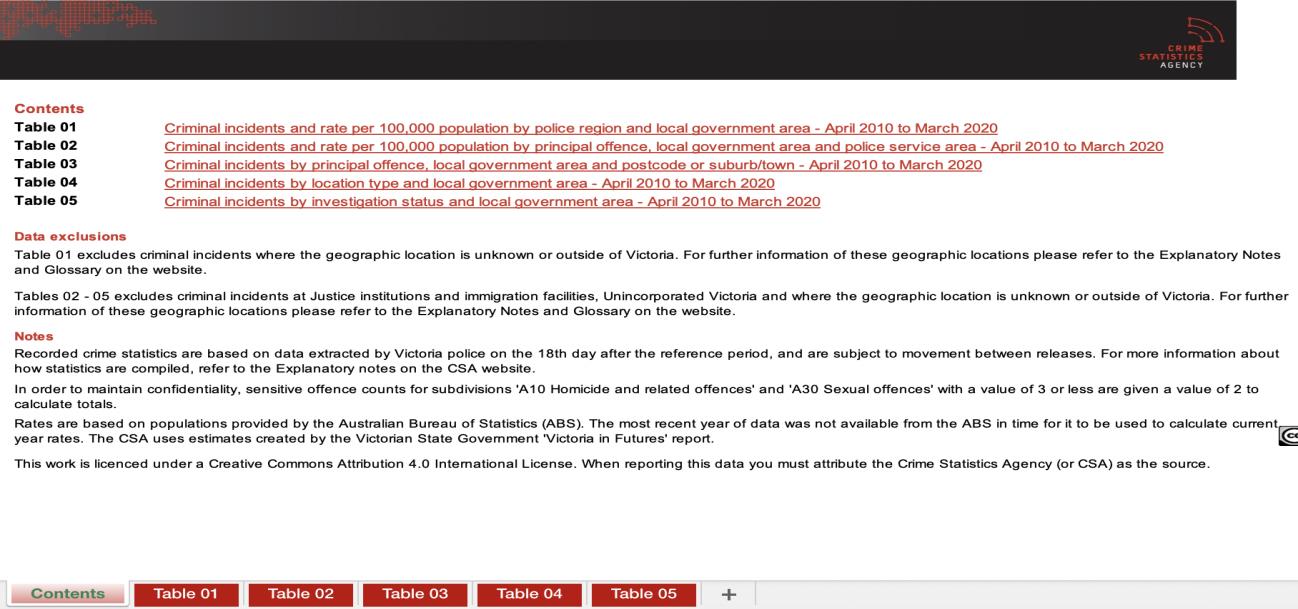


Figure 1: Crime Incidents Data Table structure

For the purpose of this analysis, I used data from Table 2, Table 3, Table 4 and Table 5. Table 2, details down Table 1 to Criminal Incidents and rate per 100,000 population by Police Service Areas and Principal Offence. Table 3 provides a different context to the crime incidents reporting where it records the Criminal incidents by principal offence and LGA and postcode allowing for a more detailed analysis. The data from April’2010 to March’2020 which represents a good 10 years’ worth of data. Here, A Local Government Area (LGA) is a geographical area under the responsibility of an incorporated local government council [5] and Principal offence is the offence type determined by the CSA offence index as the most serious offence type committed within an incident [5]. Table 4 further enhances the analysis as it provides crime incidents by location type. Table 5 allows us to analyse the investigation status of the crime incidents recorded by LGA.

Secondly, to create a leaflet map, I needed a shape file data with the information for 79 LGAs present in my dataset. After an initial struggle to find a correct spatial datafile which was either not very big or had missing or incorrect information I finally, got the shape file from the Victorian Government Data Directory. The shape file I had contained LGAs which contained coordinates for locations like islands and mountains. Removing them was my initial approach but it led to white spots on the plot making it look inaccurate. Since recorded incidents are more related to whole LGA and my dataset does not have any information for these places, I realised there was not information loss or incorrect result. I customised the plot to present a colour palette where the difference can be seen easily. I also added legend and title.

Secondly, I had to change the lower case LGAs in my dataset to upper case and rename shape file column to join both files correctly.

I used a shiny dashboard app instead of shiny app as it provides more flexibility in customising the end product. I followed the approach of keeping all drop down tabs on left and the plots on right. Further I used functions like renderPlotly, renderPlot to render plots on web app.

User Guide:

In order to proceed with the analysis please ensure all required libraries are installed on your system. The information about these libraries can be traced back to Implementation.

When you run the app.R file in Visualisation project folder, it will run and present in front of you the final web app.

You can notice that on left you have a tab named ‘Crime Rate’ which when clicked on reveals the further drop-down menus. On the right are the plots which can be scrolled down and can be identified by plot label what they represent.

First plot is the leaflet map plot which plot the log values of the sum of normalised count of recorded incidents across LGAs for the last decade.

The Second plot is to look at the trend in the recorded incidents overtime for an LGA. The user can select the LGA from the second drop down menu.

The third plot is about analysing the trend in the incidents recorded overtime for an offence type in an LGA. This plot helps the user to identify the most common offence type and its history for an LGA. The user can select offence type from drop down menu one and LGA from drop down menu 2.

The fourth plot is a tree map that presents the number of recorded incidents across different location types identified by Victorian police. The user can select a LGA and a year from the drop down menu to personalize analysis.

The fifth plot is about the investigation status of the overall recorded incidents across LGAs which the user can modify by selecting a LGA and the type of status they want to see the trend about.

Conclusion:

From the analysis performed above it can be certainly said that the crime has increased in the last decade in Victoria. The data taken from the Crime Statistics website is a very detailed dataset allowing for different analysis to be performed and interpretation of the crime statistics.

From the figures plotted as part of data visualisation the user can get a complete picture of the crime statistics in Victoria witnessed in the past decade.

From the figures plotted as part of data exploration we can see how in the last decade the recorded crime incidents have increased along with the average rate of incidence per 100,000. Local Government Areas like Melbourne and Greater Dandenong witnessed the highest number of cases. The most frequent type of crime recorded by the Victorian police is Property and deception offences with the highest number of crime incidents recorded at residential private properties. This dataset also allows us to look at the investigation status if the recorded incidents which has shown that the number of unsolved cases is increasing as compared to cases with charges laid which is not a positive sign. Based on the analysis it can be said that the crime situation in Victoria is not good with rising number of cases and rising number of unsolved cases. This analysis can be used as a starting step by both community and the authorities to understand the situation and create measures to control rising crime in Victoria.

This project helped me to incorporate my learnings from the FIT5147 Laboratory and workshop into handling such huge dataset and appropriately handling data wrangling and interpreting data variables. I learned data scraping and learned multiple ways to create shape file including geojson data. I learned to create an end-to-end web app to communicate my analysis and findings to the broader audience. I learned in deep about R Shiny and various functions to plot and render different type of plots. I learned about more about tree plot while analysing this dataset and how it can be used to present complex hierarchical data. I learned about different types of plots and their advantages and disadvantages. I learned about the five design sheet methodology and how it can be used to design effective data visualisations.

Bibliography:

[1] **Victorian crime statistics reveal the worst Melbourne suburbs for thefts from cars.** Available at: <https://www.abc.net.au/news/2020-02-01/victorias-crime-statistics-show-rising-crime-rate/11825694>

**[**2] Crime Data – Victorian Auditor-General’s office. Available at: <https://www.audit.vic.gov.au/report/crime-data?section>

**[**3] Download Data – Crime Statistics Agency. Available at: <https://www.crimestatistics.vic.gov.au/crime-statistics/latest-crime-data/download-data>

**[**4] Data Explanatory Notes – Crime Statistics Agency. Available at: <https://www.crimestatistics.vic.gov.au/about-the-data/explanatory-notes>

**[**5] Data Glossary – Crime Statistics Agency. Available at: <https://www.crimestatistics.vic.gov.au/about-the-data/glossary-and-data-dictionary>

**[**6] Victoria LGAs shapefile. Available at: <https://discover.data.vic.gov.au>

Appendix:

A close up of text on a whiteboard

Description automatically generated

Text, letter

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

Text, letter

Description automatically generated