2024

New Zealand Crime visualization

MXB362 CASE STUDY

COLIN HAN MING TAN

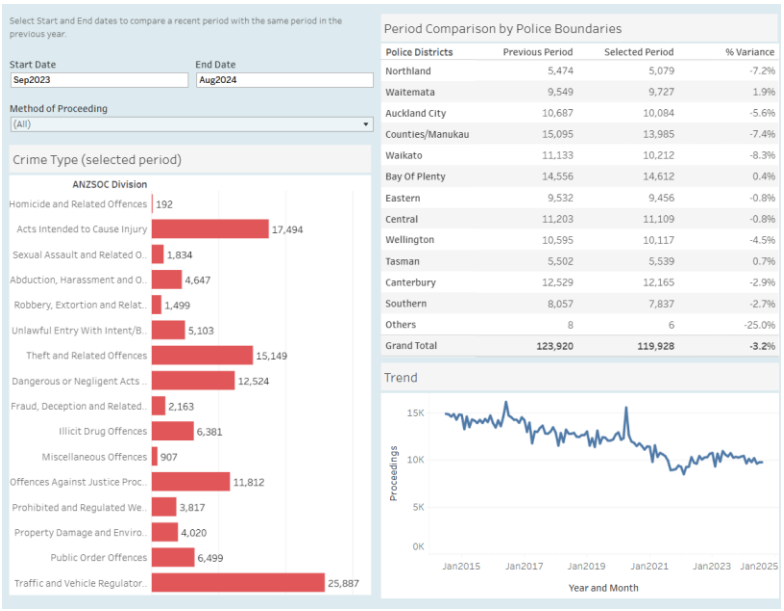
Executive summary

Crime is prevalent in every country, and one way to reduce the crime rate is to gather information from analysing data with the goal of implementing effective laws or actionable plans. This report aims to provide better insight into the vast amount of crime data in New Zealand. The data are taken from the New Zealand Police website[1] and the New Zealand government Justice website[2]. These data are used to create an interactive map of New Zealand, showing crime rates for different crime types over a 10-year period, along with an animated visualization of total crimes per region. Viewers should be able to get a general idea of crime numbers and trends in New Zealand by looking at these visualizations, rather than sifting through the large volumes of raw data available on the aforementioned websites.

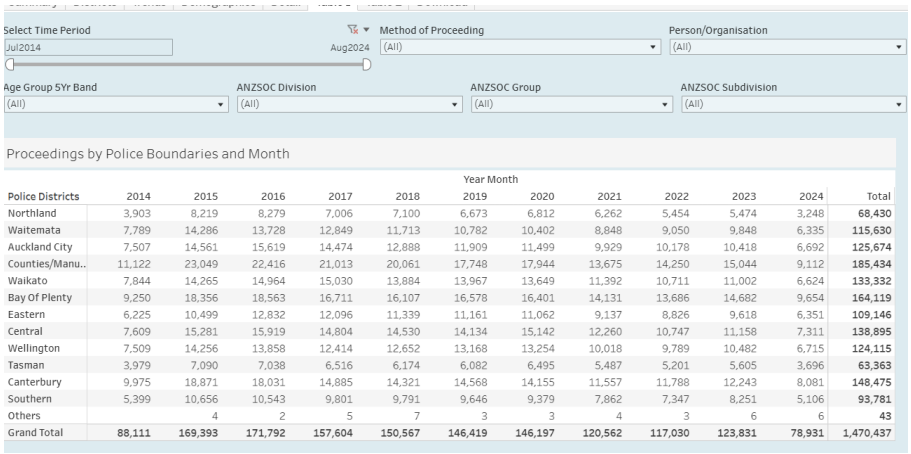
Description of project

Data sources

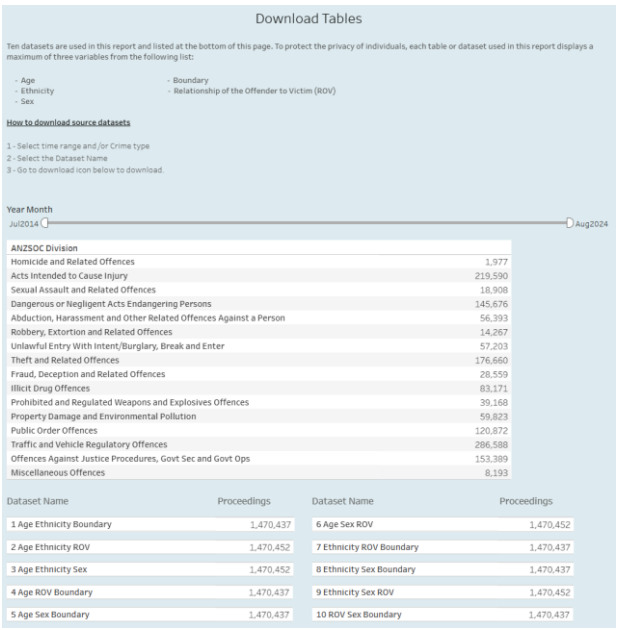
The data collected are from the New Zealand Police and New Zealand Justice websites. As both are part of government bodies, it is safe to assume that the data published are accurate and up to date. With so much data available, the first task is to decide which pieces of information are needed or valuable. The New Zealand Police website publishes reports on a monthly basis, but some of their reports are focused on crime prevention rather than crime numbers. The biggest issue with their data is that they are summarized or categorized when presented, meaning the closest way to access raw data is through 'filtering' their on-site visualization tool to display everything. Below are some screenshots to illustrate this.



Data is messily represented.



Huge volumes of data and elements that provide minimal information.



Data is only shown and the dataset is not downloadable.

This is where the New Zealand Justice website came in. Their data was represented by charges and convictions, based on the region of the court where the hearing took place. By cross-referencing the data from these two sources, a resourceful and effective dataset was created that can be used to analyse and create visualizations from. This dataset contained the number of crimes(charges) based on the year, crime type and where it occurred(region).

Visualization techniques

The main purpose of using a visualization is to send a message to the viewer. One concept of data visualization is deciding if the reader or the author decides the sequence of data being show and whether the data is shown all at once or parts at a time.

Visualization environments and tools

Rstudio is a popular and powerful development environment software that is made for processing and analysing data. To create visualizations, imported libraries such as ggplot2 and plotly are used. For this case study, libraries such as leaflet, shiny and gganimate are used to create the interactive and animated visualizations that are created on a separate user interface for easy viewing.

Results and outputs

Background research

According to *The New Zealand Herald[3],* The Crime rate in New Zealand has been at a certain constant for the past few years which leaves citizens still feeling unsafe[4], especially with an increase in fraud cases. In order to crack down on crime, one would need to analyse all the available information in order to create policies or initiatives to discourage criminal activity.

Data massaging and manipulations

As mentioned, the data sources provided huge amounts of data that had to be cleaned and filtered [show examples of original csv files]. The data from the New Zealand Justice Department are in the form of multiple excel sheets, each labeled by the type of crime, whether the crime was just a charge or a conviction and the percentage. Each data sheet is further filtered by region, sub-region and the total counts or convictions for each year from 2014 to 2023. The goal is to show a general idea of the total crime across New Zealand, so the data were collated from these excel sheets, cross referenced by the New Zealand Police website to create a consolidated comma separated values file that contains the following information: region where the crime took place, category of the crime committed and number of charges for each year. These variables are the main necessities to relay the relevant information.

In Rstudios, the data is reformatted so that it can be analyzed and presented using the R functions. Originally, the years were the part of the row(x-axis) titles and the name of some regions and crime categories had inconsistent spelling or syntax thus the data frame was reformatted to have the years as entries in columns(y-axis) and any categorical names were replaced with the correct corresponding syntax to main consistency and avoid data anomalies. Along with this data, to pinpoint the regions on the interactive map, a data frame with the latitude and longitude coordinates of each region was created to be merged with the rest of the data to create one consolidated data frame when creating the interactive map.

```r
# coordinates of each region
nz_regions <- data.frame(
  Region = c("Taitokerau", "Waitematā", "Auckland", "South Auckland",
             "Waikato", "Bay of Plenty", "Waiariki", "East Coast",
             "Taranaki/Whanganui", "Manawatū/Wairarapa", "Northern Wellington",
             "Wellington", "Nelson/Marlborough/West Coast", "Canterbury",
             "Otago", "Southland"),
  lon = c(174.29, 174.76, 174.76, 174.89,
          175.28, 176.17, 176.25, 178.35,
          174.37, 175.61, 175.02, 174.78,
          173.28, 172.63, 170.50, 168.35),
  lat = c(-35.43, -36.85, -36.85, -37.00,
          -37.78, -37.68, -38.14, -38.31,
          -39.30, -40.35, -41.13, -41.29,
          -41.27, -43.53, -45.87, -46.41)
)
```

```r
## create different data frames for the different categories ##
## data frame for year and crime by type
long_data_total <- data %>%
  pivot_longer(cols = starts_with("X"),  # Replace if column names are not prefixed with "X"
               names_to = "year",
               values_to = "Crime_Count") %>%
  mutate(year = as.integer(gsub("X", "", year)),
         Crime_Count = as.numeric(Crime_Count))

# Clean Region names in long_data_total by removing leading/trailing spaces and fixing capitalization
long_data_total <- long_data_total %>%
  mutate(Region = trimws(Region),  # Remove leading/trailing spaces
         Region = case_when(
           Region == "Waitemata\u0081" ~ "Waitematā",  # Fix special character issue
           Region == "South Auckland " ~ "South Auckland",  # Fix trailing space
           Region == "Bay of plenty " ~ "Bay of Plenty",  # Fix trailing space and capitalization
           Region == "Bay of plenty" ~ "Bay of Plenty",
           Region == "East Coast " ~ "East Coast",  # Fix trailing space
           Region == "Manawatu/Wairarapa" ~ "Manawatū/Wairarapa",  # Fix missing macron or spelling
           TRUE ~ Region  # Keep other regions unchanged
         ))

# Clean Region names in nz_regions in a similar way
nz_regions <- nz_regions %>%
  mutate(Region = trimws(Region))
```

	Region	Category	year	Crime_Count
1	Taitokerau	violent offence	2014	616
2	Taitokerau	violent offence	2015	711
3	Taitokerau	violent offence	2016	697
4	Taitokerau	violent offence	2017	618
5	Taitokerau	violent offence	2018	546
6	Taitokerau	violent offence	2019	636
7	Taitokerau	violent offence	2020	670
8	Taitokerau	violent offence	2021	543
9	Taitokerau	violent offence	2022	519
10	Taitokerau	violent offence	2023	564
11	Taitokerau	harmful digital communications	2014	0
12	Taitokerau	harmful digital communications	2015	0
13	Taitokerau	harmful digital communications	2016	0
14	Taitokerau	harmful digital communications	2017	1
15	Taitokerau	harmful digital communications	2018	2
16	Taitokerau	harmful digital communications	2019	6
17	Taitokerau	harmful digital communications	2020	0
18	Taitokerau	harmful digital communications	2021	1
19	Taitokerau	harmful digital communications	2022	1
20	Taitokerau	harmful digital communications	2023	2
21	Taitokerau	drug offences	2014	280
22	Taitokerau	drug offences	2015	292
23	Taitokerau	drug offences	2016	281
24	Taitokerau	drug offences	2017	253

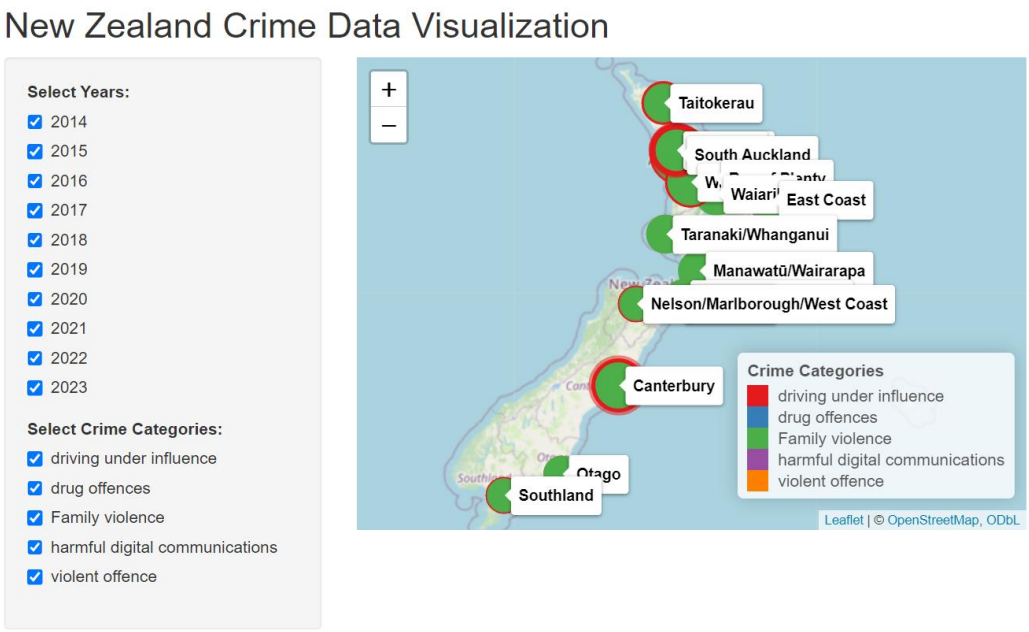
Reformatted data frame with years as entries instead of column names

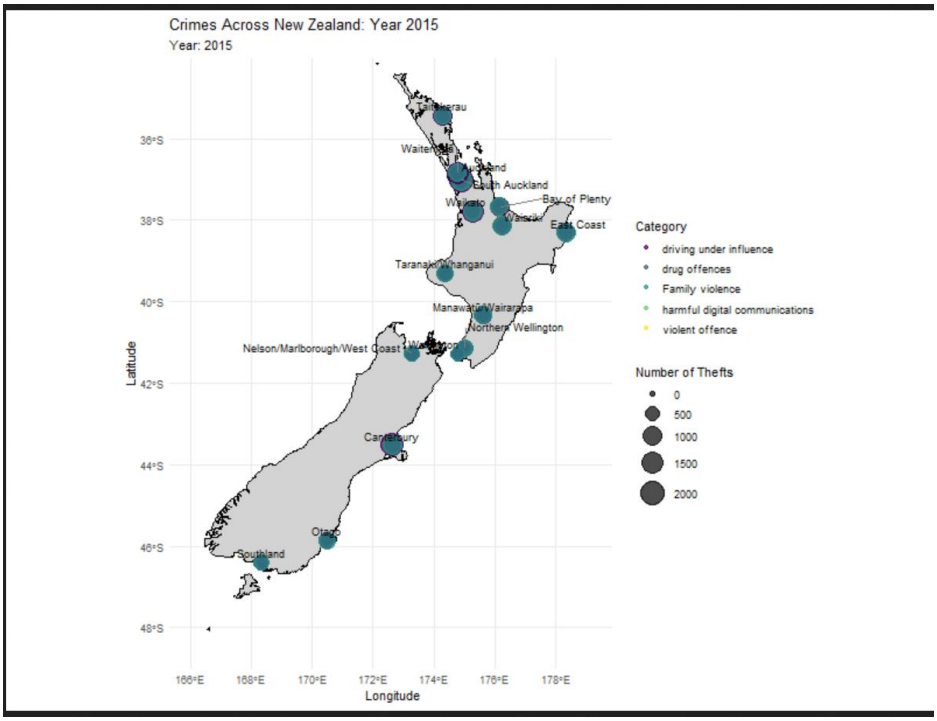
Code explanation

Now that the data was prepared, it can be properly filtered according to the different years and categories. As mentioned above, Rstudio packages like fastmap, bslib were used to create the interactive map of New Zealand with crime data by creating different layers and applying them together. The first layer is the map of New Zealand itself using the leaflet package and entering the latitude and longitude coordinates of New Zealand. Then a layer for each year and crime type was created using a for loop to ensure that every combination is filtered out into a data frame which will be called upon during the creating of the interactive map itself using another for loop. A group of colours are assigned to the different crime types and each region is also given a label to make

identification easier. The last point is to include a legend and compile it all together to create the interactive map. Due to the fact that there are a lot of elements in this visualization making unappealing to view though the Rstudios viewer, an interactive user interface window was created using the shiny package which allows readers to choose which filters they want and the data will reflect accordingly.

Visualization outputs





Above are the still images of the interactive and animated map. As mentioned, viewers can choose which year, crime type or any combination of the two they would like shown on the map. Each crime type is colour coded and the size of the circle implies the number of crime committed in that region. Users can also click on the specific region and the corresponding numbers will be shown. The animated visualization shows the total crime committed across the years therefore there is no difference in colour for the circles again but the size of the circle still represents the crime tally.

Effective visualization issues

The visualizations produced in this case study performs a better job of relaying information about the crime rate of New Zealand, with the chief reason being that the information is much more quicker and easier to absorb, especially compared to what is shown on the source website. The main difference is that instead of having to search through huge amounts of data and numbers, and reader can take a quick look at the visualizations, and if they want to inquire more, choose which conditions and filters they want and be able to take in the new information just as easily.

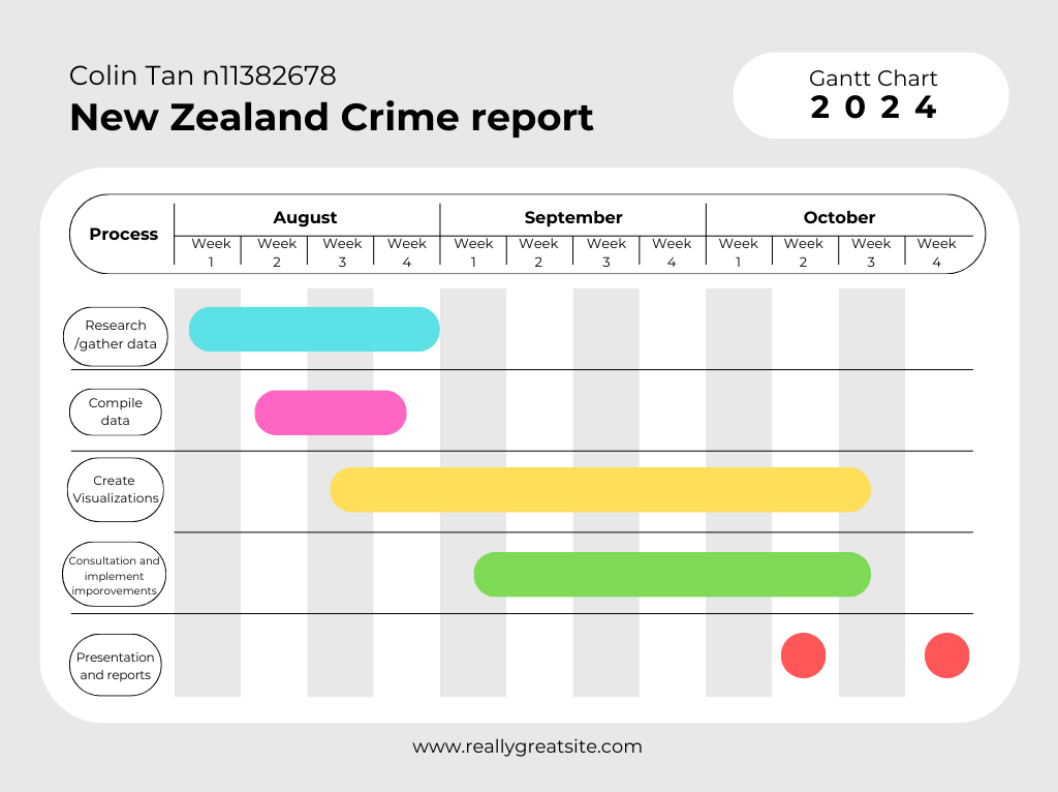
Insights and knowledge gained from case study

From the visualizations, readers can gleam that certain crimes such as family violence and driving under the influence have high number while crimes such as harmful digital communications are low. Viewers can also see that the crime rate in New Zealand has not seen a drastic increase or decrease in the past 10 years, which is not bad news but any crime prevention plans should be implemented sooner rather than later.

Exegesis of the effectiveness of visualization output

The goal of the interactive and animation maps is to visualize the crime rates across the New Zealand over the course of 10 years. Normally, using area or size to represent numerical data is not highly accurate but because the main goal is to show the trends and differences, and with the huge amounts of data, consolidating all the data and representing it as such is adequate. Each type of crime is represented by a colour, providing a quick and intuitive way for readers to differentiate between types of crime. This applies Keller & Keller's principle of using colour to classify and reveal categories within data[5]. Having an interactive map gives the reader control of the intake of information. Choosing to look at a specific region, filtering by year and crime type and clicking the circles to get certain information shift the responsibility of data flow from the author to the reader, fostering interactive data exploration rather than passive consumption. Another key point is reducing the data-to-ink ratio[6] to ensure that the focus remains on the data. Non-essential visual elements like gird lines and highlighted borders are minimized or removed and Iin this case, having an external user interface window helps. Both the interactive and animated map uses a simple and clean design layout and the only outstanding elements are highlighted regions.

Update Tasks and project Timeline



August:

- Dataset is found and read through. Found guides on visualization techniques that potentially improve the intended visualization.
- Read through available data from sources to discern if data was useful
- Did research and reading into crime rates in New Zealand

September

- First few iterations of animated visualization is made however still working on creating the proper animation (there is animation but not as intended)
- Debugging and making sure code produced intended visualizations
- Consulted Lecturer and tutor on thought process

October

- Started putting together presentation and report
- Still working to debug and implement features in to visualization

Problems

Some issues that arose while doing this case study were in the form of data collection as well as the coding of the interactive visualization. The New Zealand Police website was the first source of data found and was the only source mentioned in the project proposal. But upon deeper reading, the data presented was not the most useful and the website itself was challenging to navigate. The data shown on the website is already “filtered” and “visualized” in charts and tables and claimed that the data was ablet to be downloaded how ever, what is downloaded is simply the data represented as the chart or table on the website. Fortunately the New Zealand Justice Department also has their own website and has huge amounts of data about criminal cases as well. The trouble now was navigating through the huge amounts of data that has been compiled based on the law structure and classifications of New Zealand, requiring to read up on the laws themselves to understand how the data is categorized. Another key issue that arose was coding the interactive map to produce the intended outcome. The interactive map has many layers and elements so the data has to be accurately cleaned and prepared else it will affect many processes down the line. This also makes it very difficult to incorporate features outside of the R packages that help to create the interactive map. Originally, the plan was to have custom icons represent each of the crime type, but because the icon size represents the numerical data(crime count), it required that the custom icons had to have different size to represent the difference because the icons are not dynamic, and on top of that because of the nature of the function, it is better if whatever represents the crime tally is a dynamic element.

Project journal

9 August: Found that New Zealand police publish reports about their arrests monthly since 2014. Can use this data to supplement visualization.

10 August: To ensure visualization is complex enough, I have decided to include New Zealand Police crime data.

18 August: The official website data is unorganized, scattered and categorized in a confusing matter, will take slightly longer than anticipated to ensure I get enough of the correct information.

10 September: Noticed that a lot of advanced techniques introduced in the unit involves 3-dimensional data structures or visualizations, need to consult if my project meet the requirements.

15 September: Got feedback regarding my project, will carry on with intended.

19: Started working on the animating visualization in Rstudio, slightly confusing but thankfully many resources online about the libraries that I am using.

21 September: looked through online resources for documentation and examples of using leaflet and shiny

24 September: seems the code for creating an interactive visualization is simple, its just a matter of creating layers to be put together. Now need to make sure I think thorough on what message I want to send

5 October: I have many ideas I want to implement for the interactive visualization like custom icon but it seem that a lot of work like editing the size and pixel of the image is needed and the leaflet package only accepts certain image formats.

11 October: take what I said about coding in R to be simple. Even in the tutorials I am getting stuck on some functions. Same situation with this case study. Getting an error when compiling all the custom layers and it seems to be caused by the data in the custom data frames created. Lots of tedious work to go through just to debug. Hoping to get the proper visualization for the presentation.

18 October: the interactive and animated visualizations are mostly done, gone through some testing and debugs but happy with the results. Slightly disappointed that I could

not do or implement everything I intended but maybe can look into it in the future when I have more time.

Conclusions

This case study summarizes and presents the crime rate of New Zealand for the past 10 years, with the intention of providing insight and information to policy makers and law enforcement officials to think of ways to prevent more crime from occurring. Having easy access to crime data on the official websites was convenient and while there is no argument about the analytical capabilities that MATLAB has, I personally prefer Rstudios because it seems easier to use, and with so many packages and libraries to help you with your task. As a data science major, doing this case study gave me a glimpse of what I might potentially do in the future and I enjoy it. However, I do understand why it takes a lot of time and energy to process data, analyse it and present it in a way that makes it easy to understand. Working with huge amounts of data and deciding if that information is useful requires a certain level of attention to detail. If given more time, I would have like to delve deeper into the vast data available and look to see if there are certain algorithms that can be formed from it.

References

*1. **Data and statistics**. (n.d.). New Zealand Police. https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics*

*2. **Data tables | New Zealand Ministry of Justice**.* (n.d.). https://www.justice.govt.nz/justice-sector-policy/research-data/justice-statistics/data-tables/

3. *Cheng, D. (2024, June 27). **Crime: Latest survey shows overall crime steady, rise in feeling unsafe, biggest increases in fraud, vehicle theft.** NZ Herald. https://www.nzherald.co.nz/nz/politics/crime-latest-survey-shows-overall-crime-steady-rise-in-feeling-unsafe-biggest-increases-in-fraud-vehicle-theft/UALI6KIV2NFANKP745GI6CZRXU/*

*4. Daniels, C. (2024, July 9**). Crime crackdown: What the data says about rising fear, but steady rates - The Front Page**. NZ Herald. https://www.nzherald.co.nz/nz/crime-crackdown-what-the-data-says-about-rising-fear-but-steady-rates-the-front-page/SUP7KGJD4BGUBHSLNBEBTNEF7U/*

*5. **Keller, P. R., & Keller, M. M.** (1993). Visual Cues: Practical Data Visualization. Los Alamitos, CA: IEEE Computer Society Press.*

*6. **Tufte, E. R.** (1983). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.*