Police Dashboard: Exploratory Data Analysis (EDA)
DATA301: Group Project

Maria DaRocha (300399718), George Pope (300416638), Usanthini Sasikumar (300451063)

8/30/2020

GitHub Repository:

https://github.com/Marianette/Police-Dashboard

The datasets that were used for this report are Offender Proceedings (offender demographics), Offender Proceedings (police stations), Police Dataset (Victimisation Time and Place), Victimization (Police Stations) and Unique Victims (Police Stations).

Background and Description of Datasets Offender Proceedings (Offender Demographics) and Offender Proceedings (Police Stations)

These datasets are of particular interest as they show each crime committed, the type of crime, age, gender, the ethnicity of the offender, and where it was committed. It could be used to identify which types of crime are most common in certain areas of New Zealand, and which demographic is most likely to commit these of crimes.

The types of data present in these datasets are both numerical and categorical. The types of offenses, location, and other variables are stored as characters, as well as time-series data, and integers for certain variables. There are some values missing in the dataset, where they were not stated, or where people's information had been suppressed.

Background and Description of Datasets Police Dataset (Victimisation Time and Place)

This dataset has 107089 observations of 14 variables. This dataset is of particular interest as it shows each crime committed, the type of crime, when it was committed, and where it was committed. It could be used to see and identify if there are any links or patterns between the type of crime committed and the area of the crime committed (Territorial authority) around Wellington.

The types of data present in this dataset are both numerical and categorical. The types of crimes, location, and other variables are stored as characters, as well as time-series data and Integers for certain variables such as the number of records, victimizations, and other variables. There are some values missing in the dataset, where they were not listed clearly, or where the information had been reduced.

Background and Description of Datasets Victimisations & Unique Victims (Police Stations)

Among the datasets we decided to work on, two of them were: Victimisations (Police Stations), and Unique Victims (Police Stations). These datasets were of particular interest because oftentimes when we discuss

crime, the majority of our conversations revolve around the offender: their background, their demographic, etcetera. While these details may help us better understand an individual's propensity towards crime, focusing solely on the offender can mean overlooking the other involved party: their victims. Thoughtfully done, analyzing information about victims can lead to a greater understanding of commonalities between victims of similar crimes. This layer of insight could help us better understand what might make someone more vulnerable to crime, or susceptible to being targeted by a specific type of crime.

The datasets themselves consisted of largely *categorical data*. The data is best suited to time series models, as it is intended to offer insights about victimisations and unique victims within a specific time period (2015-2020).

While there were no missing values in the Victimisations (Police Stations) dataset, there were some corrupted entries in which a categorical variable (Ages 10-19) had become stored as a date (Oct-19). This issue could be related to the files being exported from policedata.nz as 'Microsoft Excel csv files,' instead of regular csv files. Data integration took place on an as-needed basis. For example, when comparing ethnicity and crime division features in each dataset, the four features were merged into a single dataframe. The Unique Victims dataset was missing just over 68,000 values.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

General EDA, Victimisations (Police Stations):

Missing Values, Data Types, Etc.

```
##   ï..Table.1         Anzsoc.Division    Anzsoc.Group       Anzsoc.Subgroup
##  Length:631458      Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##   Year.Month         Age.Group          Anzsoc.Subdivision Ethnic.Group
##  Length:631458      Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
```

```
##
##
##    Ethnicity             Location           Location.Type
##  Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Location.Type.Description Location.Type.Division Person.Organisation
##  Length:631458             Length:631458          Length:631458
##  Class :character          Class :character       Class :character
##  Mode  :character          Mode  :character        Mode  :character
##
##
##
##  Police...Boundary  Police.Area        Police.Area.TA     Police.District
##  Length:631458      Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Police.District.TA Police.Station     Police.Station.TA  Territorial.Authority
##  Length:631458      Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    Weapon            Weapons.Used       X..Variance        X.Variance
##  Length:631458      Length:631458      Length:631458      Length:631458
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Number.of.Records Previous.Period    Selected.Period      Variance
##  Min.   :1         Min.   :  0.0000   Min.   :  0.000   Min.   :-122.00000
##  1st Qu.:1         1st Qu.:  0.0000   1st Qu.:  0.000   1st Qu.:   0.00000
##  Median :1         Median :  0.0000   Median :  0.000   Median :   0.00000
##  Mean   :1         Mean   :  0.4139   Mean   :  0.442   Mean   :   0.02805
##  3rd Qu.:1         3rd Qu.:  0.0000   3rd Qu.:  0.000   3rd Qu.:   0.00000
##  Max.   :1         Max.   :122.0000   Max.   :199.000   Max.   : 199.00000
##  Victimisations
##  Min.   :  1.000
##  1st Qu.:  1.000
##  Median :  1.000
##  Mean   :  2.113
##  3rd Qu.:  2.000
##  Max.   :199.000


## [1] 68114
```

3

```
## [1] 0
```
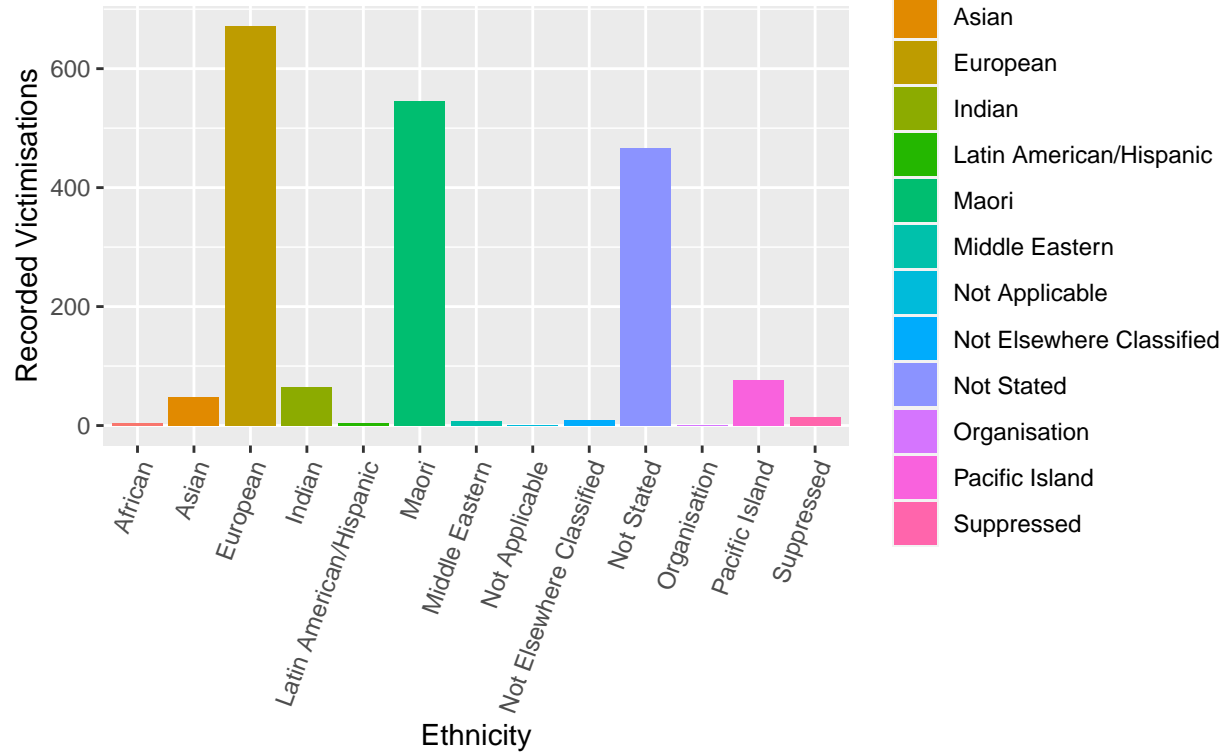
Victimisations (Police Stations):
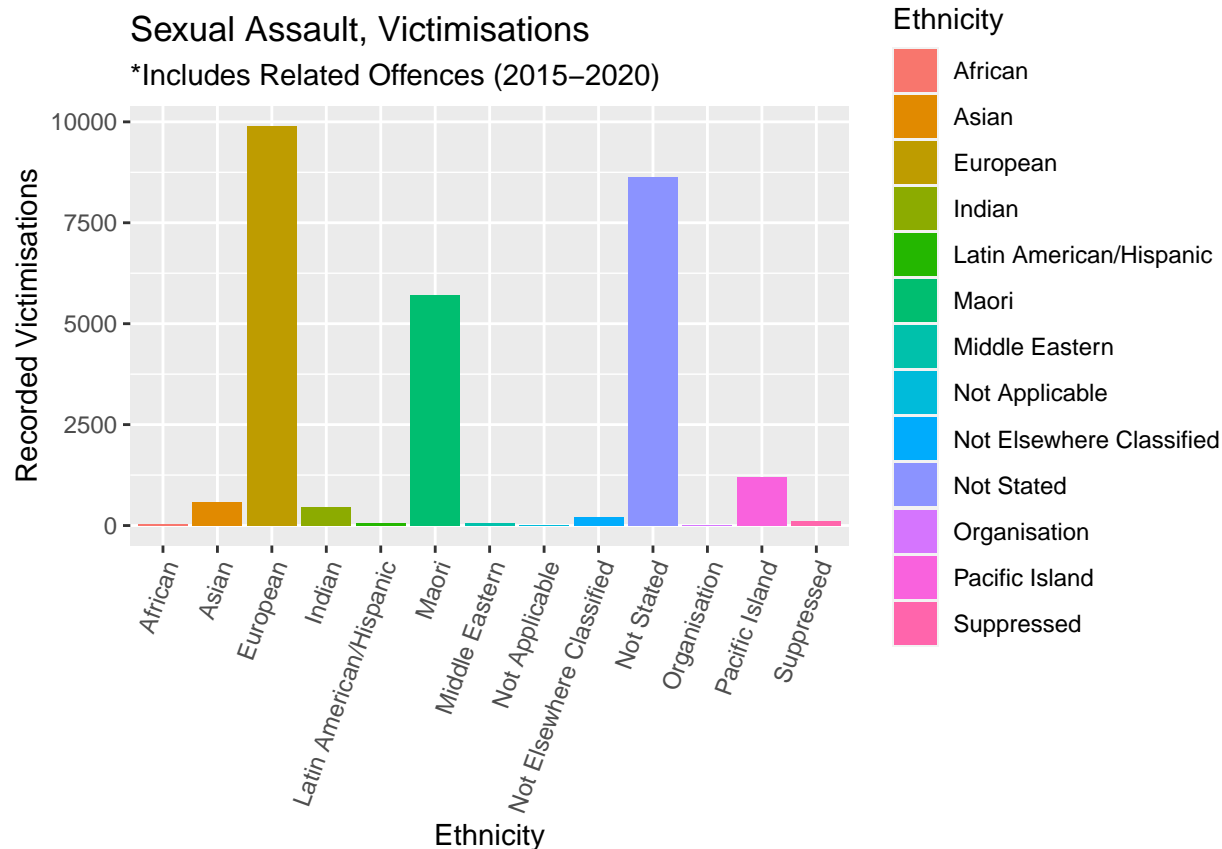
Crime Division, Ethnicities of Victims

In the victimisations (by police stations) dataset, we chose to explore ethnicity's connection to each crime division. If patterns are found in later, deeper analysis - then we will be able to conclude some correlation between ethnicity and a particular type of crime. In a real-world application, this would help identify a population that may be more vulnerable to certain crimes. Initially, we explored the data using the classifications for ethnicity provided. Below, we show the distributions of Ethnicity classes for crimes involving abduction and sexual assault.

Ethnicity	Abduction, Harassment and Other Related Offences Against a Person	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
African	4	665	63	47	967	0
Asian	48	4563	1053	575	11849	0
European	672	62113	4287	9901	85993	0
Indian	65	6395	2331	461	12351	0
Latin American/Hispanic	5	354	33	69	1079	0
Maori	546	50251	1580	5705	35300	0

Abduction and Harassment, Victimisations
*Includes Related Offences (2015–2020)

Sexual Assault, Victimisations
### *Includes Related Offences (2015–2020)

Victimisations (Police Stations):

Crime Division, (Aggregated) Ethnicities of Victims

Upon seeing the data heavily reflect European, Maori, and unspecified (or suppressed) ethnicities, it seemed appropriate to aggregate non-Maori and non-Eurpoean ethnic groups into a "Minority (Other)" category. This aggregation provides a visual model which improves the comparability of minority groups that are less prevalent New Zealand. In doing this, we ensure that individuals who separately fall into the (now aggregated) "Minority (other)" category are fairly represented - and that visually, the model no longer reflects any one minority category appearing less targeted than Maori or European persons.

Ethnicity	Abduction	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
Minority (Other)	206	24628	3983	2411	38771	0
European	672	62113	4287	9901	85993	0
Maori	546	50251	1580	5705	35300	0
Not Applicable	0	0	0	0	0	83748
Not Elsewhere Classified	10	1508	95	216	1792	0

Ethnicity	Abduction	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
Not Stated	467	41133	4526	8644	102914	0

Abduction and Harassment, Victimisations
*Includes Related Offences (2015–2020)
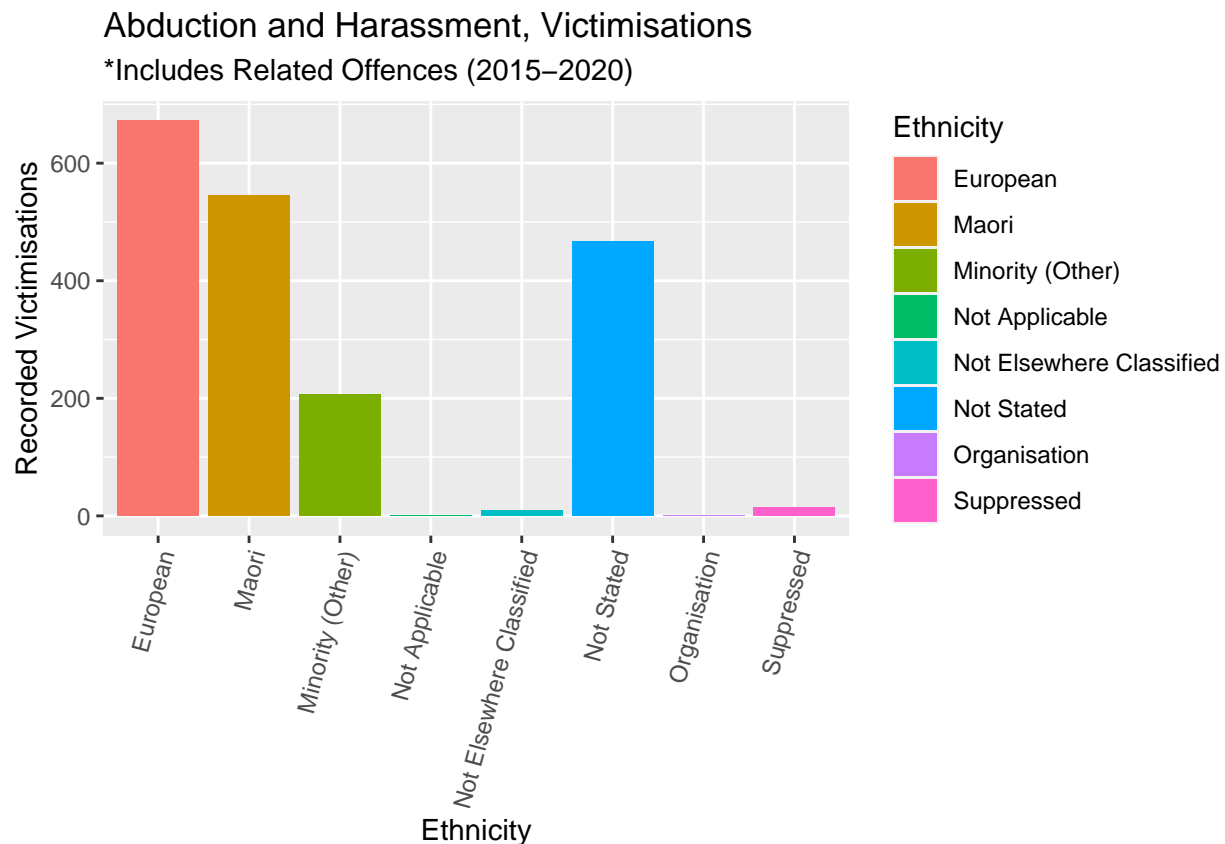
Victimisation (Police Stations):

Crime Division to Year of Report

The next interesting feature of the data is "Year.Month" - i.e. the basis for our time series. In future analysis, we will investigate this feature in greater detail. For now, we can aggregate these values to just the year on record and observe basic trends.
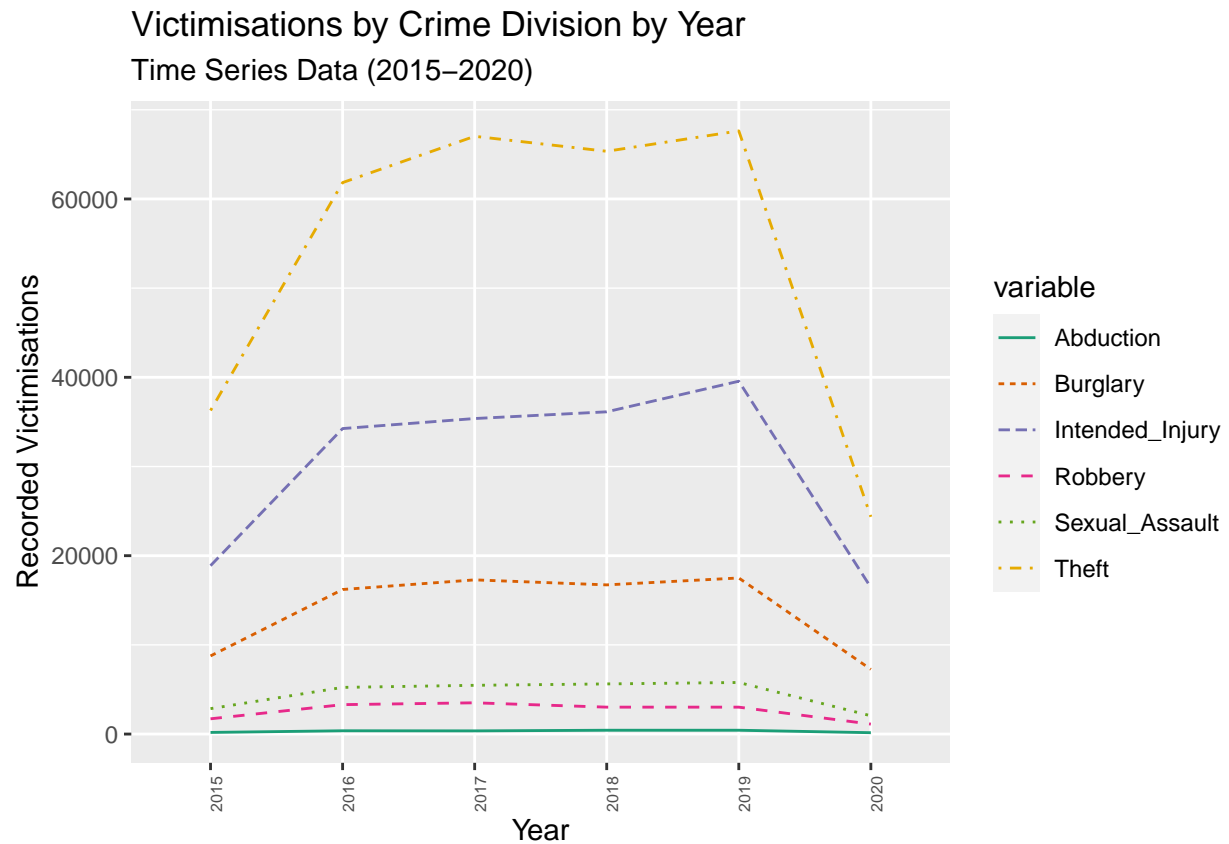
Year	Abduction	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
2015	182	18879	1712	2849	36279	8760
2016	369	34255	3300	5234	61834	16211
2017	356	35380	3506	5465	67026	17290
2018	428	36132	3017	5621	65344	16728
2019	428	39562	3019	5783	67628	17502

Year	Abduction	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
2020	152	16444	1117	2048	24361	7257

Time Series, Victimisations (Police Stations):

Change in Recorded Victimisations by Crime Division (Annually)
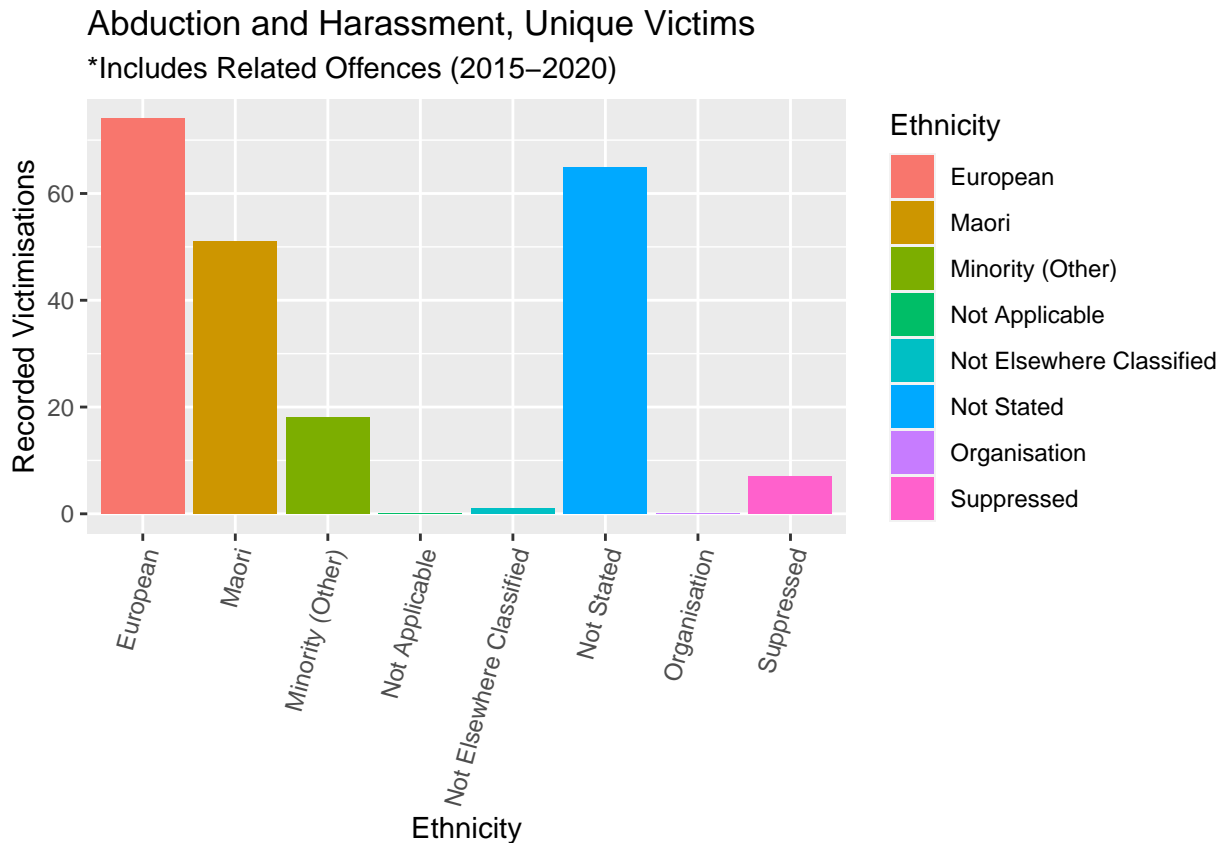
Below, we model this abstracted data.

Unique Victims (Police Stations):

Crime Division, (Aggregated) Ethnicities of Victims

We conduct a near-identical analysis of Unique Victims (Police Stations) below.

Ethnicity	Abduction	Acts Intended to Cause Injury	Robbery, Extortion and Related Offences	Sexual Assault and Related Offences	Theft and Related Offences	Unlawful Entry With Intent/Burglary, Break and Enter
Minority (Other)	18	3426	544	417	3927	0
European	74	7614	691	1386	7489	0
Maori	51	6229	255	897	3445	0
Not Applicable	0	0	0	0	0	6268
Not Elsewhere Classified	1	260	17	42	225	0
Not Stated	65	5788	742	1460	9445	0

Abduction and Harassment, Unique Victims
*Includes Related Offences (2015–2020)

Unique Victims & Victimisations (Police Stations):

Distributional Analysis with Standard Deviation

For the final part of the exploratory analysis on this data, it seemed like a reasonable inquiry to test the comparablity of the victimisations and unique victims datasets (in regards to variability). Ethnicity was used as a dummy variable to track the consistency of a single feature's variation across the two datasets. We

would expect, if the datasets are comparable, for the standard deviations to exhibit the same (or relatively similar trends). This proved to be the case - and going forward, it will be possible to use this test holding other features of interest constant, and potentially use the two datasets interchangeably. **This is invaluable knowledge for avoiding temporally and computationally expensive (unnecessarily redundant) investigations.**
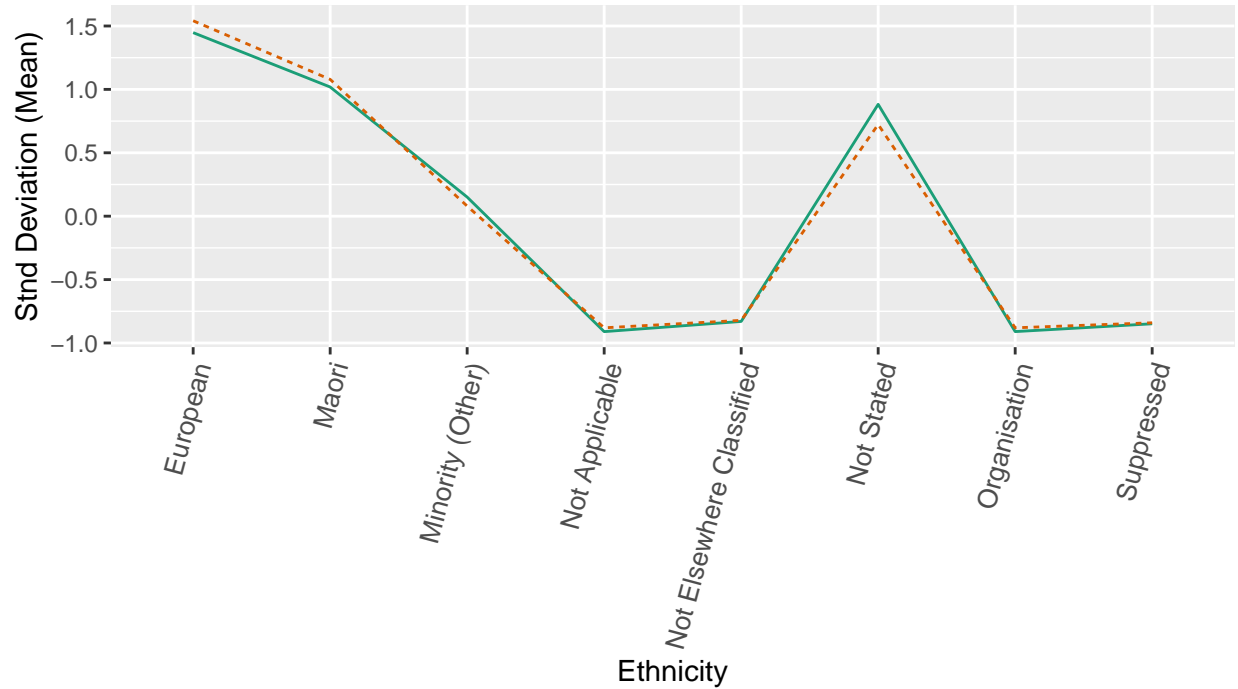
Abduction

Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015–2020)
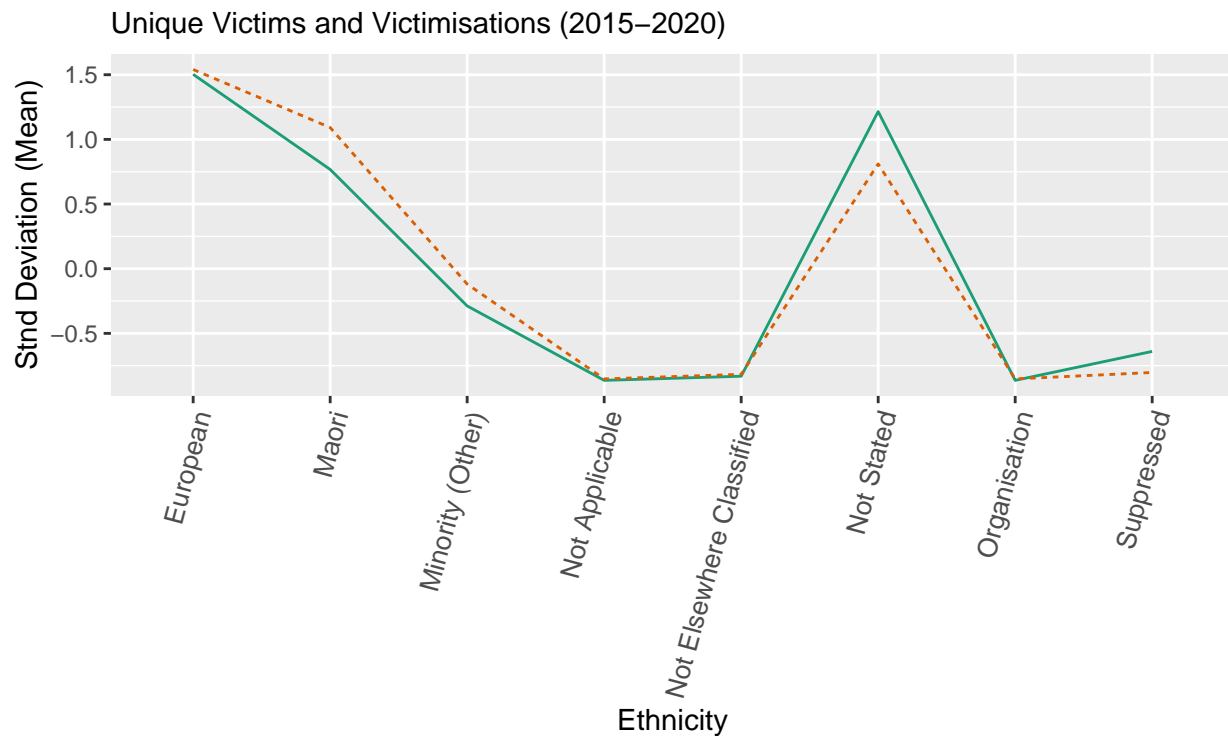
Acts Intended to Cause Injury

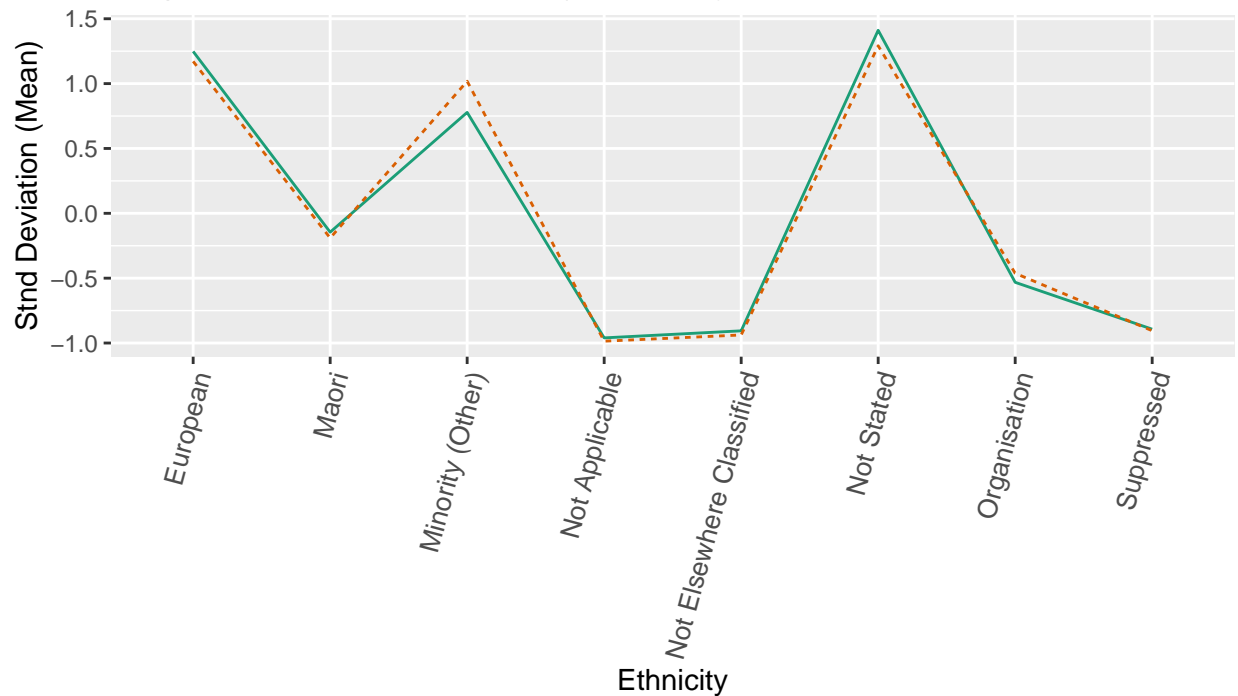Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015–2020)

Robbery, Extortion and Related Offences
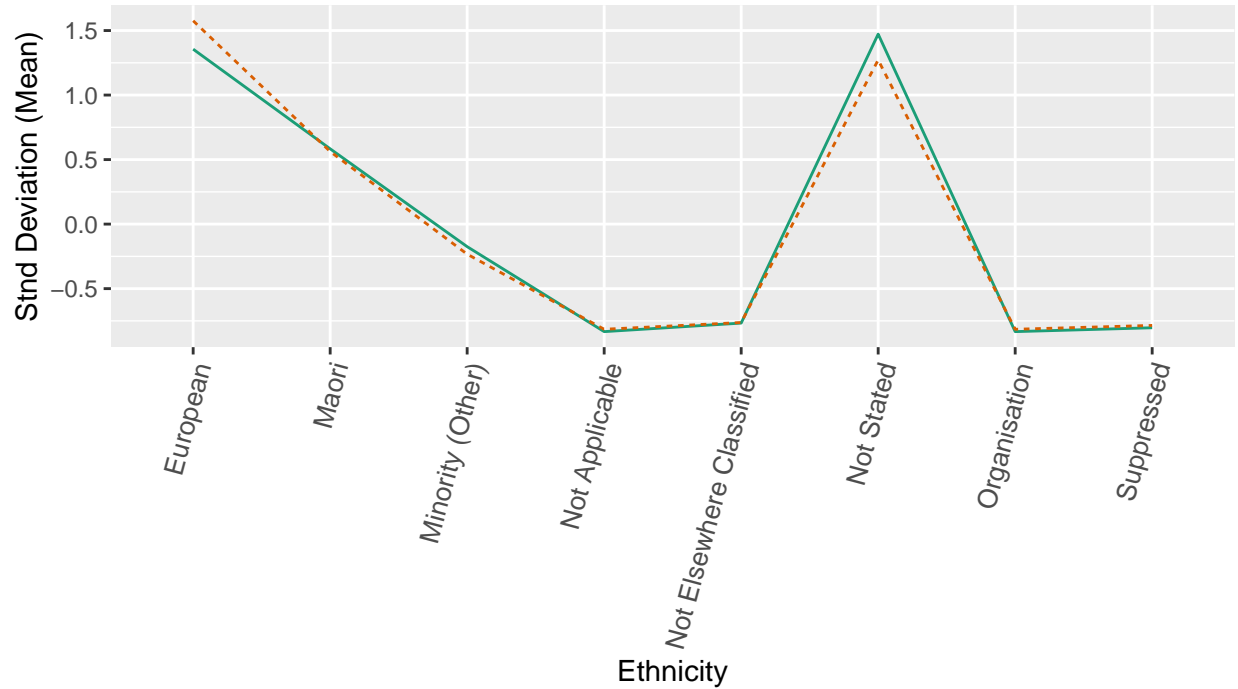
Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015–2020)

Sexual Assault and Related Offences
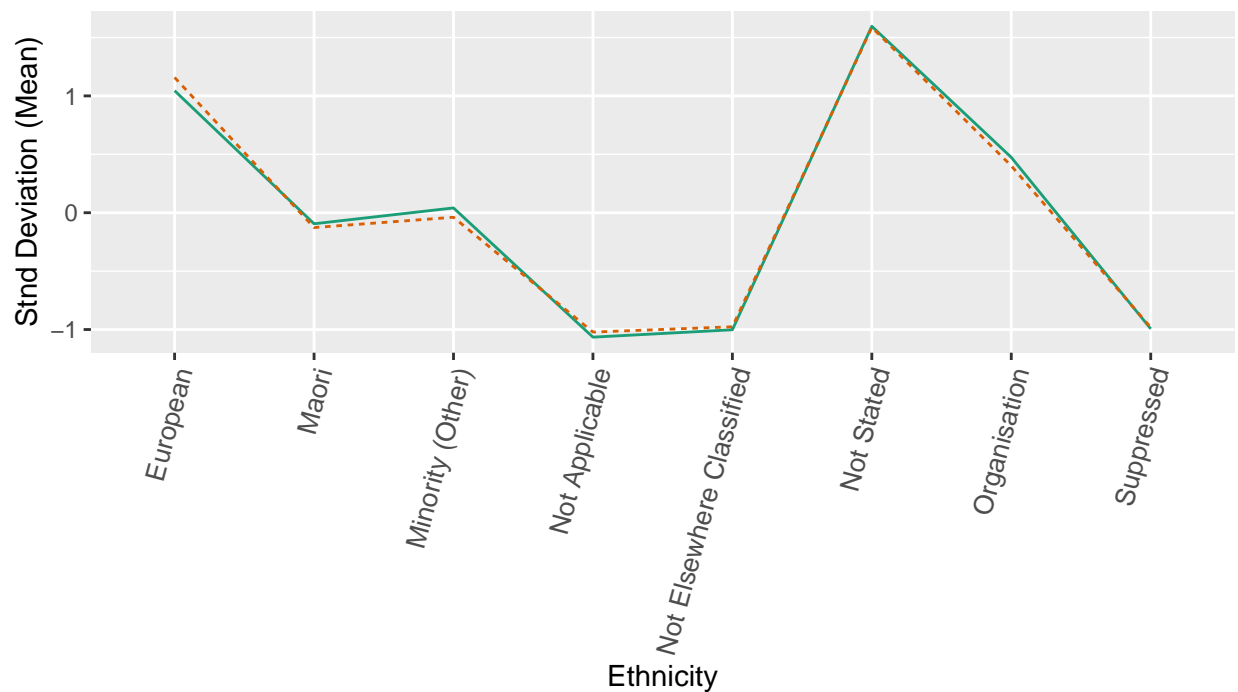Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015−2020)
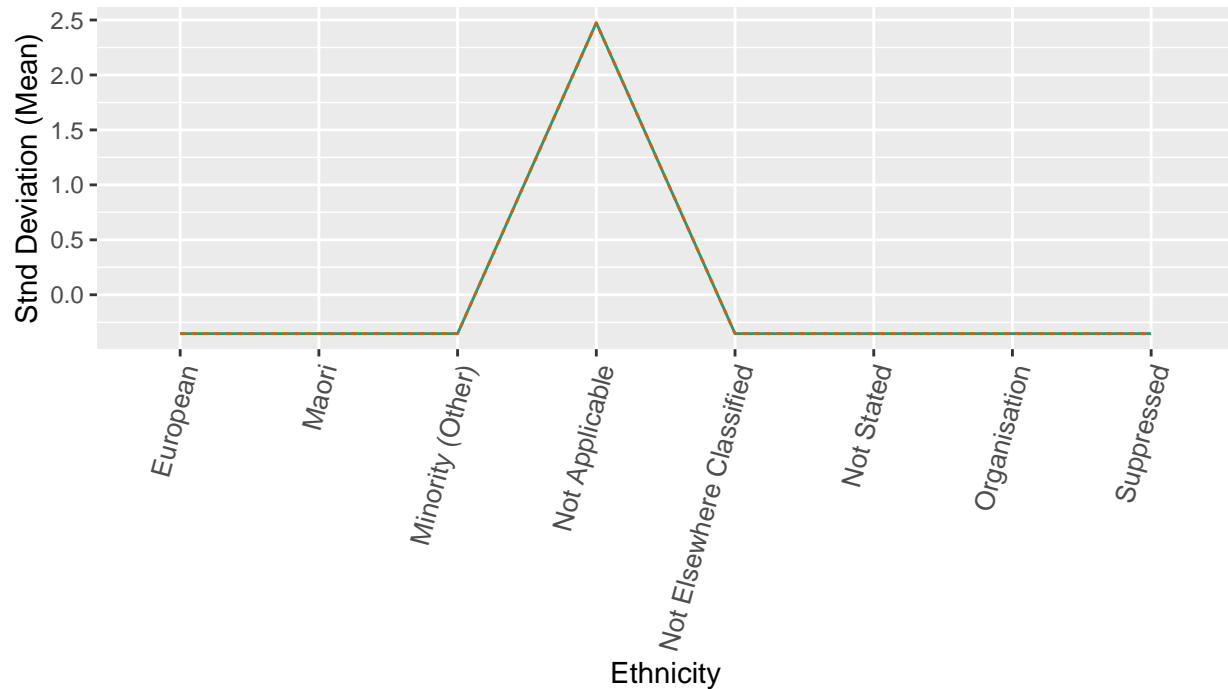

Theft and Related Offences
Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015−2020)

Unlawful Entry With Intent/Burglary, Break and Enter
Fluxuation in Distribution of Ethnicities:

Unique Victims and Victimisations (2015–2020)

Proceedings (Police Stations) & Proceedings (Offender Demographics):

Crimes committed by Europeans and Maori:

```
##                      European Maori
## Abduction/Harassment    15693 11931
## Assault                 40321 48592
## Endagerment             21072 13460
## Fraud                    7249  5777
## Homicide                  413   343
## Drugs                   22347 16629
## Miscellaneous            2160  1068
## Govt/Justice            19392 23092
## Weapons                  8758  8736
## Property/Environment    12556 15015
## Public Order            22511 22822
## Robbery/Extortion        1234  4294
## Sexual Assault           4894  2921
## Theft                   28541 42019
## Traffic/Vehicle         34339 35018
## Break and Enter          8968 15395
```

## Types of Crimes Committed by Maori and Europeans
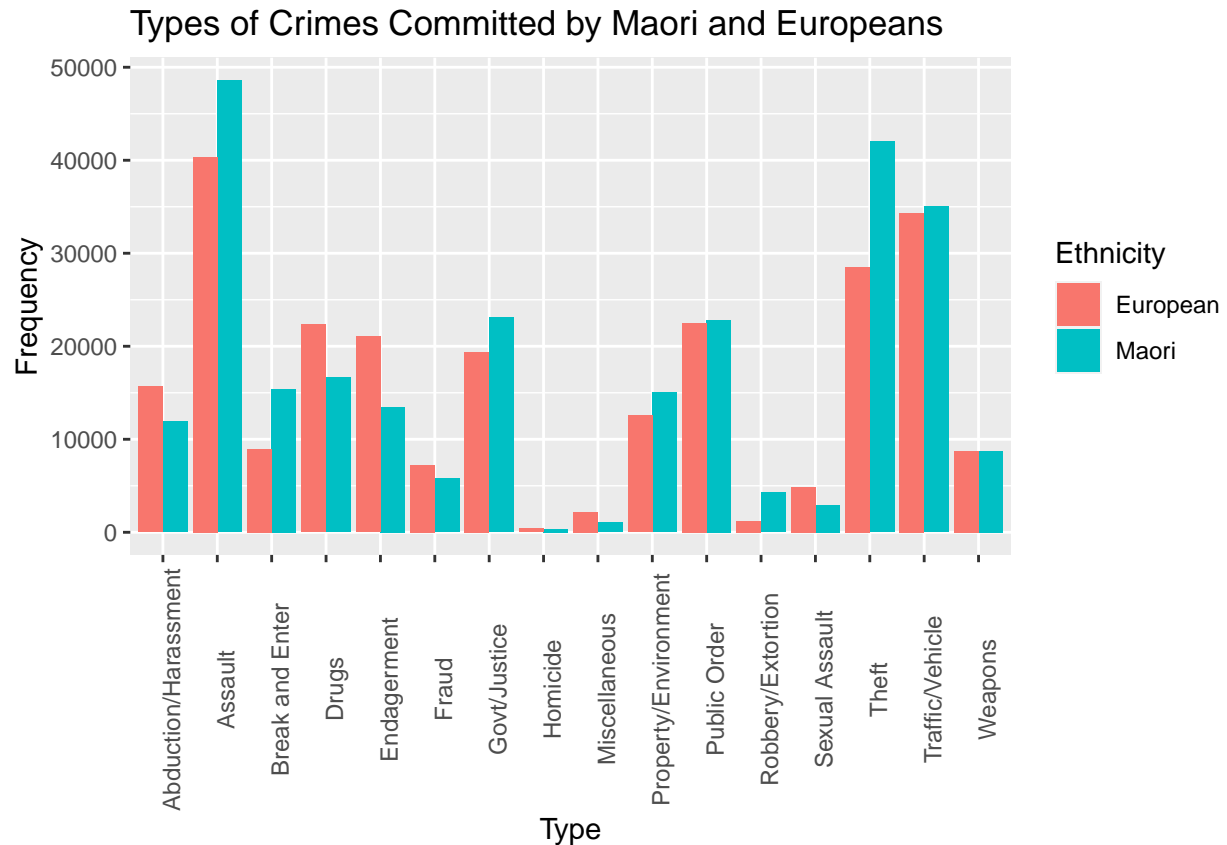
The variables along the x-axis refer to the types of crimes committed by particular ethnicities. The naming of types have been simplified to allow easy fit onto the graph.The y-axis refers to the number of times these types of crimes occurred.

It is interesting that despite the Maori population being far smaller than the European population in New Zealand, they have similar crime rates, even exceeding Europeans in certain crime types, such as theft and assault.

Types of crimes committed in certain regions:

```
##                     Auckland Wellington
## Abduction/Harassment     2469       2783
## Assault                  8623      10017
## Endagerment              5067       5056
## Fraud                    1848       1483
## Homicide                   50         59
## Drugs                    3861       4085
## Miscellaneous             312        372
## Govt/Justice             5600       4635
## Weapons                  1609       1942
## Property/Environment     2426       3041
## Public Order             5691       4769
## Robbery/Extortion         983        487
## Sexual Assault            906        940
## Theft                    8379       7297
## Traffic/Vehicle          8346       9058
## Break and Enter          2311       2355
```

Types of Crimes Committed in the Auckland City and Wellington Districts
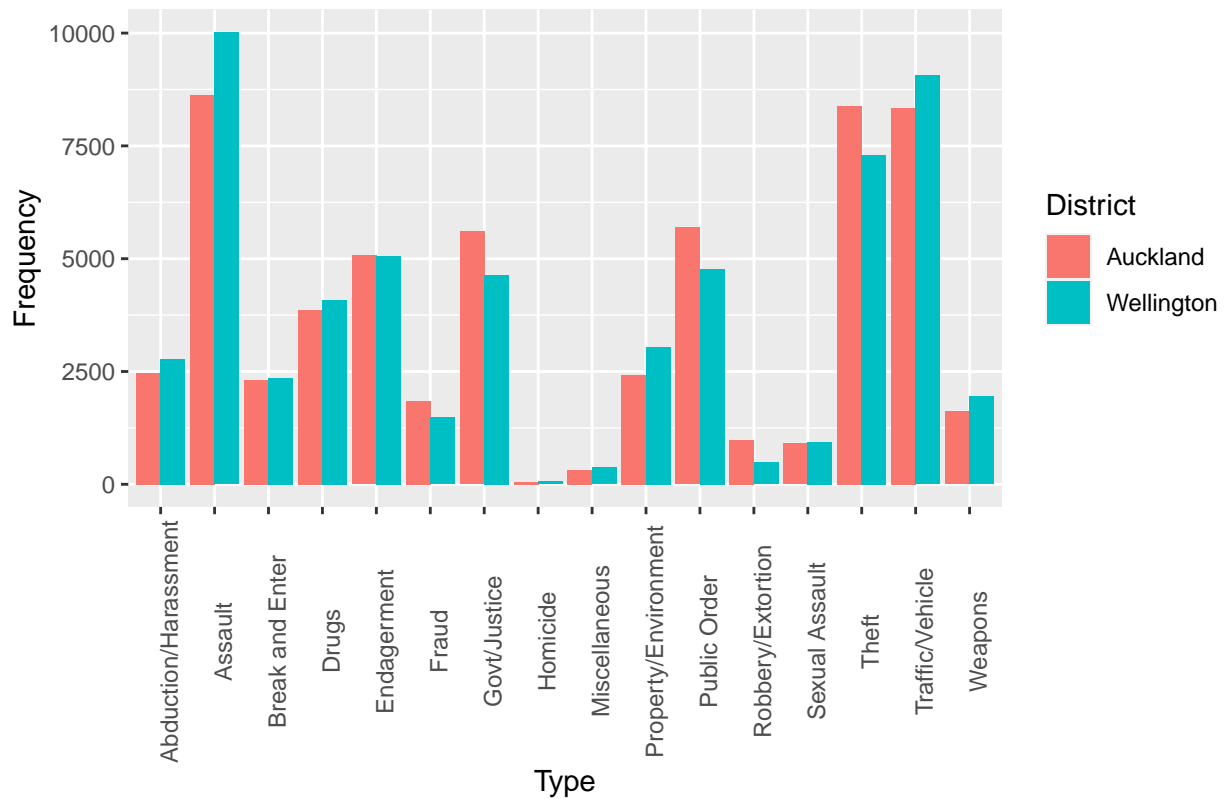
It is important to note that Auckland here refers to the Auckland City district, not the entirety of Auckland. Wellington in this instance refers to the entirety of the Wellington district, which includes areas like Upper Hutt.

Each of the variables along the x-axis are types of crimes that are committed. The naming of types have been simplified to allow easy fit onto the graph. The y-axis is the number of times specific crimes have been committed in the district. It is interesting to note that although the Auckland City District has the greatest population out of all possible districts, it has a very similar number of crimes compared to the Wellington District.

Total Crimes Committed by Age Group
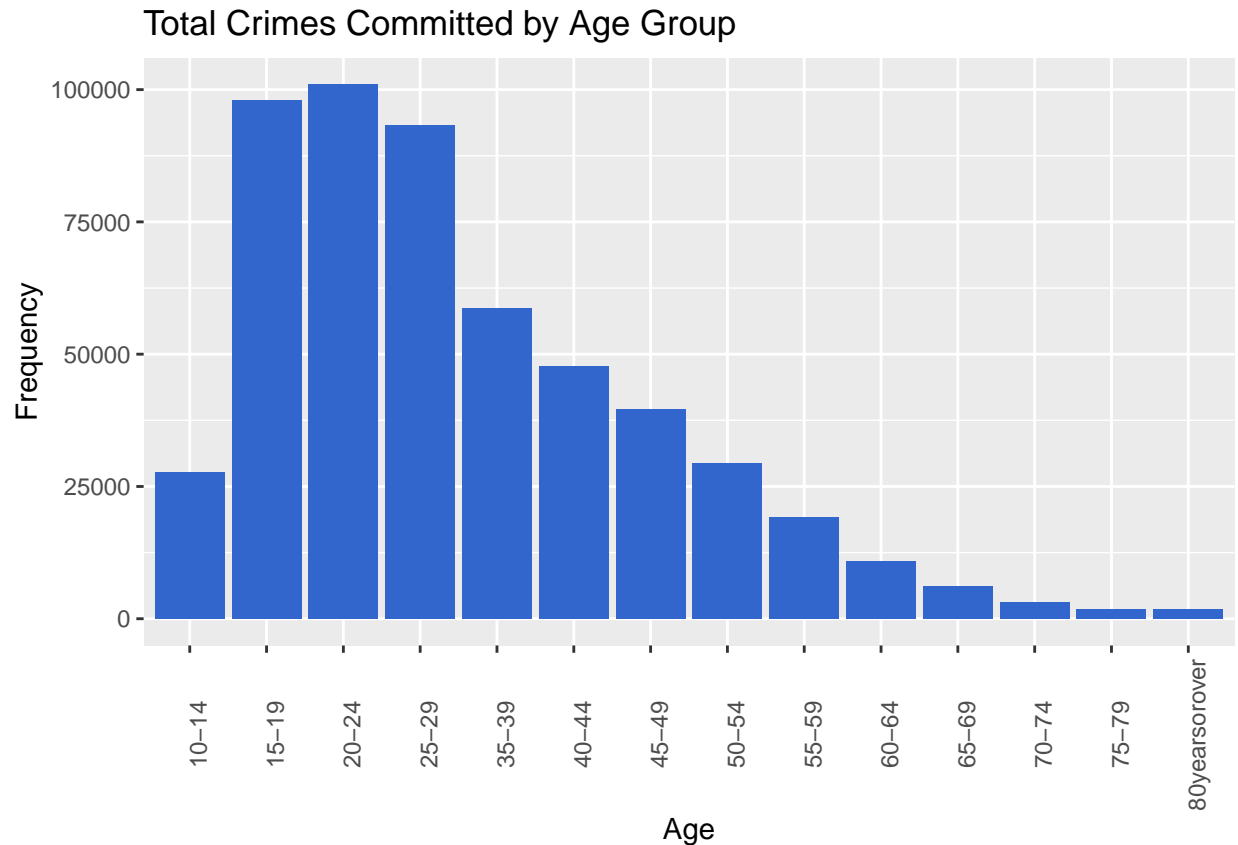
Note ages, 0-4, 5-9, NotApplicable and NotSpecified were removed as they were so small compared to the other age groups.

The x-axis refers to the age groups of offenders and the y-axis is the total number of crimes recorded for that specific age group.
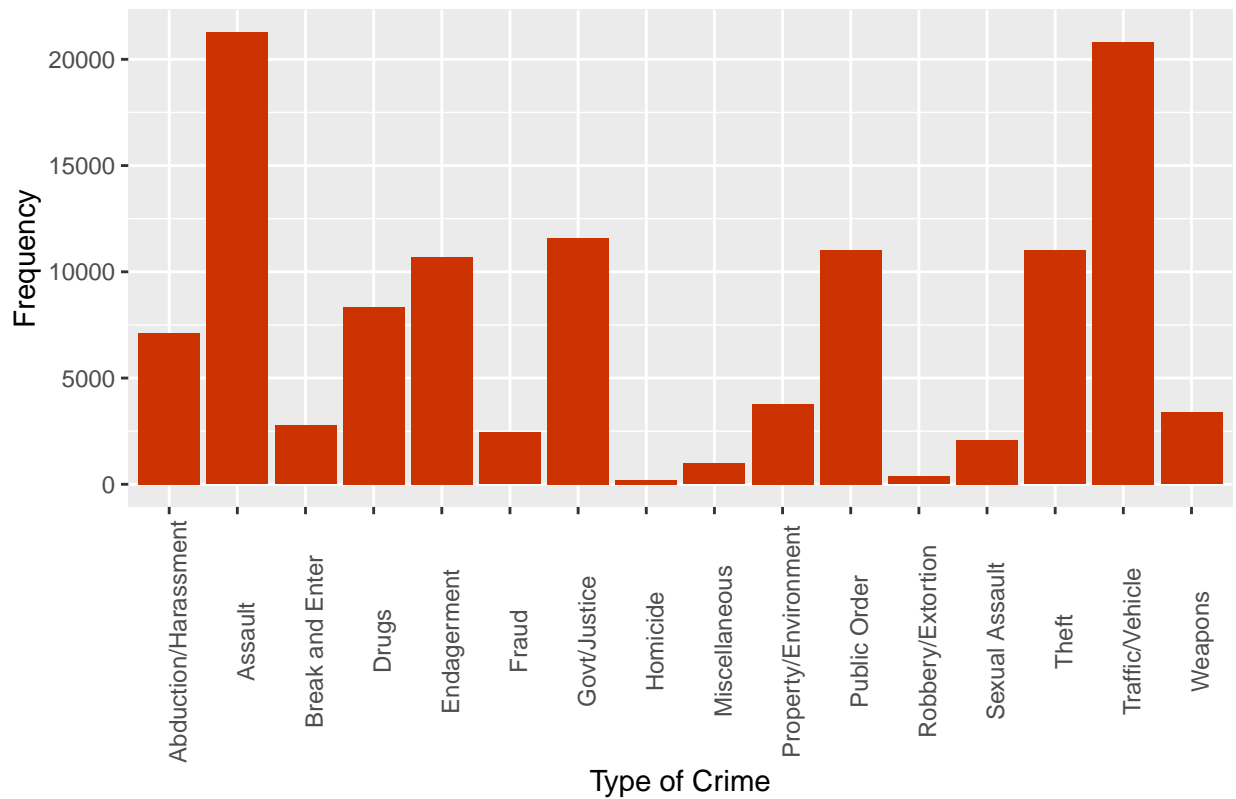
A point of interest is that crimes occur the most commonly at ages 15 through to 29, and then begins to taper off as you get older. There is a steady decline in crimes commited as you reach age 30, at a decreasing rate.

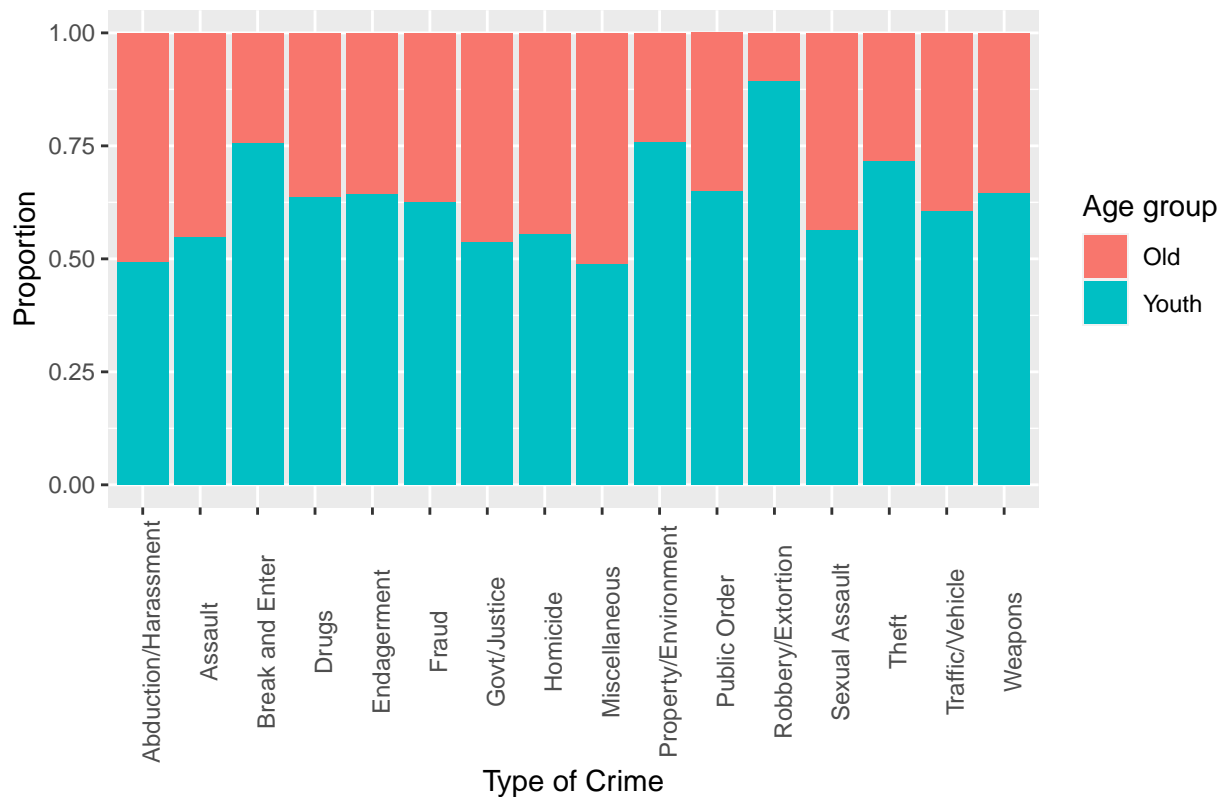Crimes committed by youths (15-24) and people aged 40-54

Crimes Committed by Youths (15–24)
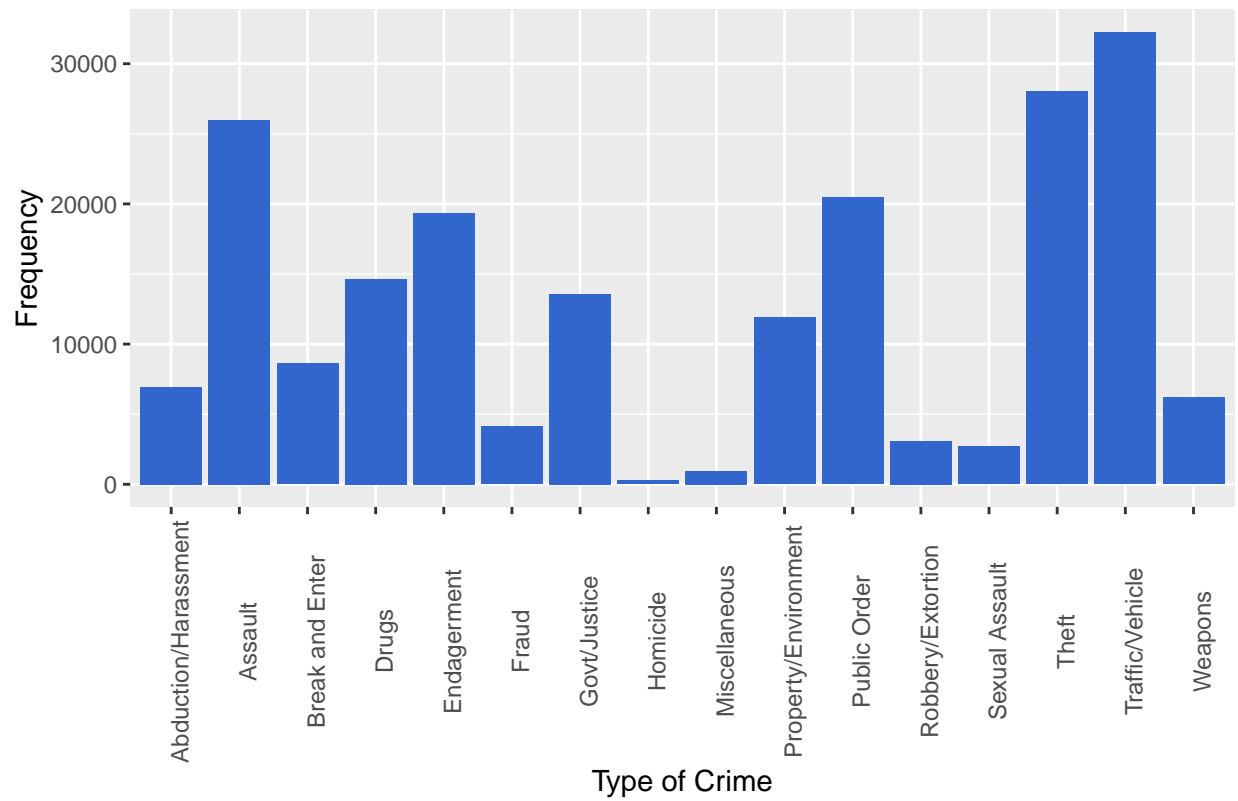
Crimes Committed by people aged 40–54

Proportion of Crimes Committed by Youths (15–24) and Old (40–59)

Each of the variables along the x-axis are types of crimes that are committed. The naming of types have been simplified to allow easy fit onto the graph. The y-axis is the number of times specific crimes have been committed by particular age brackets.

The most common types of crime for youths to commit are assault, theft, and traffic/vehicle offences, which are similar to the most common crimes committed by people in the age bracket 40-54, with the exception of Offences Against Justice Procedures, Govt Sec and Govt Ops (Govt/Justice). Theft is also much more common for youths when compared to the older age group.

Assault is the most common crime for the older age bracket, whereas traffic/vehicle offences are the most common for youths.

Police Data (Victimisations Time and Place):

```
##   ANZSOC.Division      Month.Year         ANZSOC.Group        ANZSOC.Subdivision
##  Length:107089      Length:107089      Length:107089       Length:107089
##  Class :character   Class :character   Class :character    Class :character
##  Mode  :character   Mode  :character   Mode  :character    Mode  :character
##
##
##
##    Area.Unit          Locn.Type.Division    Meshblock          Occurrence.Day.Of.Week
##  Length:107089      Length:107089      Min.   :    -289   Length:107089
##  Class :character   Class :character   1st Qu.:1966800    Class :character
##  Mode  :character   Mode  :character   Median :2052500    Mode  :character
##                                        Mean   :2038403
##                                        3rd Qu.:2136000
##                                        Max.   :2277300
##  Occurrence.Hour.Of.Day   Table.1           Territorial.Authority
##  Min.   : 0             Length:107089      Length:107089
##  1st Qu.:13             Class :character   Class :character
##  Median :21             Mode  :character   Mode  :character
##  Mean   :52
##  3rd Qu.:99
##  Max.   :99
##     Weapon           Number.of.Records Victimisations
##  Length:107089      Min.   :1          Min.   : 1.000
##  Class :character   1st Qu.:1          1st Qu.: 1.000
##  Mode  :character   Median :1          Median : 1.000
##                     Mean   :1          Mean   : 1.062
##                     3rd Qu.:1          3rd Qu.: 1.000
##                     Max.   :1          Max.   :11.000
```

```
## [1] 0
```

Ethical Considerations

1.) Many ethical remarks can be raised about the morality of undertaking feature analysis on perpetrators of different divisions of crime.

- One of the most pronounced issues would be applying learnings from patterns in data to individuals in the real world. Putting firm labels on features that offenders appear to have in common is inherently

discriminatory. Just as equally, we may be able to identify trends (feature patterns, or offender commonalities) that might shift focus more onto pre-crime interventions (for subgroups discovered to be "more likely" to tend towards crime), and less onto post-crime rehabilitation.

- An additional ethical consideration is the moral quandary of choosing whether or not to use offender data to predict future trends in crime. A positive reason for doing so might be to reduce victimisations for certain subgroups of people. However, a significant consequence of this would be the influence of data-driven bias leading to selffulfilling scenarios. The example here would be: An upwards trend in crime is identified for (insert police district). More police are tasked with community surveillance. More surveillance leads to more detected crimes for said district, and fewer for another (because police are a finite resource). This district becomes more marginalised, causing the members of its community to fall into a lower-income decile.

Privacy Considerations

2.) Privacy issues that may arise when using these datasets:

- Ensure that suppressed information cannot be related back to people it is pertaining to.

- Data collection and use must meet the standards listed in the NZ Privacy Act 1995, and the Statistics Act of 1975 – amended 2012

Security Considerations

3.) Steps that could be taken to keep the project data and results secure:

- Using complex passwords provide essential protection from fraud and hackers from accessing our data as guessing passwords is one of the most common method hackers use to breaks into computers.

- Data encryption will further secure the data by encoding the data so that it remains hidden from unauthorized users. Encryption protects private and sensitive information and enhances the security of communication between clients and servers.

- Allowing permitted individuals only to access those computer areas, e.g. by restricting entry by ID cards or magnetic swipe cards to these areas

Individual Contributions

1.) Maria DaRocha:
+ Background Description Victimisations (Police Stations) & Unique Victims (Police Stations)
+ EDA Victimisations & Unique Victims (Police Stations)
+ Ethical considerations
+ Code integration
+ Report consolidation
+ Project Organisation (Trello)
+ Version control (GitHub)

2.) George Pope:
+ Background Description Offender Proceedings (Offender Demographics) and Offender Proceedings (Police Stations)

+ EDA Offender Proceedings (Offender Demographics) and Offender Proceedings (Police Stations)
+ Privacy considerations
+ References
+ Version control (GitHub)

3.) Usanthini Sasikumar:
+ Background Description Police Data (Victimisation Time and Place)
+ EDA Police Data (Victimisation Time and Place)
+ Security Considerations

References

- NZ Police. (2017). Victimisations Time and Place (V2:00) [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/victimisation-time-and-place
- NZ Police. (2017). Offender Proceedings (Police Stations). [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/proceedings-police-stations
- NZ Police. (2017). Offender Proceedings (offender demographics). [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/proceedings-offender-demographics
- NZ Police. (2017). Unique Victims (demographics) (02.00) [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/unique-victims-demographics
- NZ Police. (2017). Unique Offenders (demographics) [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/unique-offenders-demographics
- NZ Police. (2017). Victimisations (Police Stations) [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/victimisations-police-stations
- NZ Police. (2017). Unique Victims (Police Stations) (01.00). [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/unique-victims-police-stations
- NZ Police. (2020). Covid-19 Response Data (01.01) [Datafile]. Available from NZ Police Website: https://www.police.govt.nz/about-us/statistics-and-publications/data-and-statistics/covid-19-response