Crime Trends in New Zealand: An Analysis of Demographics and Offense Types

Anju Sambasivan

Introduction

This project analyzes crime trends in New Zealand to understand how demographic factors such as age, gender, and ethnicity influence criminal activities. Using data from the New Zealand Police's Recorded Crime Offenders Statistics (RCOS) from July 2022 to July 2024, the study identifies the most common types of crimes and the demographic groups most involved in criminal proceedings. The goal is to provide insights that can help develop effective crime prevention strategies and targeted interventions.

Methods

The data for this analysis was sourced from the New Zealand Police Recorded Crime Offenders Statistics (RCOS), covering the period from July 2022 to July 2024. The dataset includes key variables such as offense type (ANZSOC Division), offender demographics (age group, gender, and ethnicity), and the number of proceedings for each case.

Data analysis was conducted using R, with the `ggplot2` package employed for creating visualizations such as bar charts and boxplots to identify trends. The `dplyr` package was used for data cleaning, grouping, and summarizing. Categorical variables like ANZSOC Division, Ethnicity, Age Group, and Gender were converted to factors to ensure accurate analysis. Exploratory data analysis was performed to uncover patterns in criminal proceedings across demographic groups.

Results

1. Libraries Loaded

```r
library(ggplot2)
library(knitr)
library(dplyr)
```

- **ggplot2**: Used for creating bar charts, boxplots, and other visualizations.
- **knitr**: Used to format tables for a report.
- **dplyr**: Provides tools for data manipulation, including filtering, summarizing, and grouping.

2. Data Preparation

```r
data <- read.csv("data.csv")
head(data)
```

```
    Date                ANZSOC_Division Ethnicity Gender Age_Group Proceedings
1 Jul-22 Acts Intended to Cause Injury  European Female     20-24           1
2 Jul-22 Acts Intended to Cause Injury  European Female     20-24           2
3 Jul-22 Acts Intended to Cause Injury  European Female     20-24           3
4 Jul-22 Acts Intended to Cause Injury  European Female     25-29           2
5 Jul-22 Acts Intended to Cause Injury  European Female     25-29           3
6 Jul-22 Acts Intended to Cause Injury  European Female     25-29           4
```

```r
str(data)
```

```
'data.frame':   76656 obs. of  6 variables:
 $ Date           : chr  "Jul-22" "Jul-22" "Jul-22" "Jul-22" ...
 $ ANZSOC_Division: chr  "Acts Intended to Cause Injury" "Acts Intended to Cause Injury" "Act
 $ Ethnicity      : chr  "European" "European" "European" "European" ...
 $ Gender         : chr  "Female" "Female" "Female" "Female" ...
 $ Age_Group      : chr  "20-24" "20-24" "20-24" "25-29" ...
 $ Proceedings    : int  1 2 3 2 3 4 4 1 2 2 ...
```

- **read.csv("data.csv")**: Reads the dataset into R.
- **head(data)**: Displays the first six rows to preview the dataset.
- **str(data)**: Displays the structure of the dataset, including column names and data types.

```r
data$ANZSOC_Division <- as.factor(data$ANZSOC_Division)
data$Ethnicity <- as.factor(data$Ethnicity)
data$Age_Group <- as.factor(data$Age_Group)
data$Gender <- as.factor(data$Gender)
```

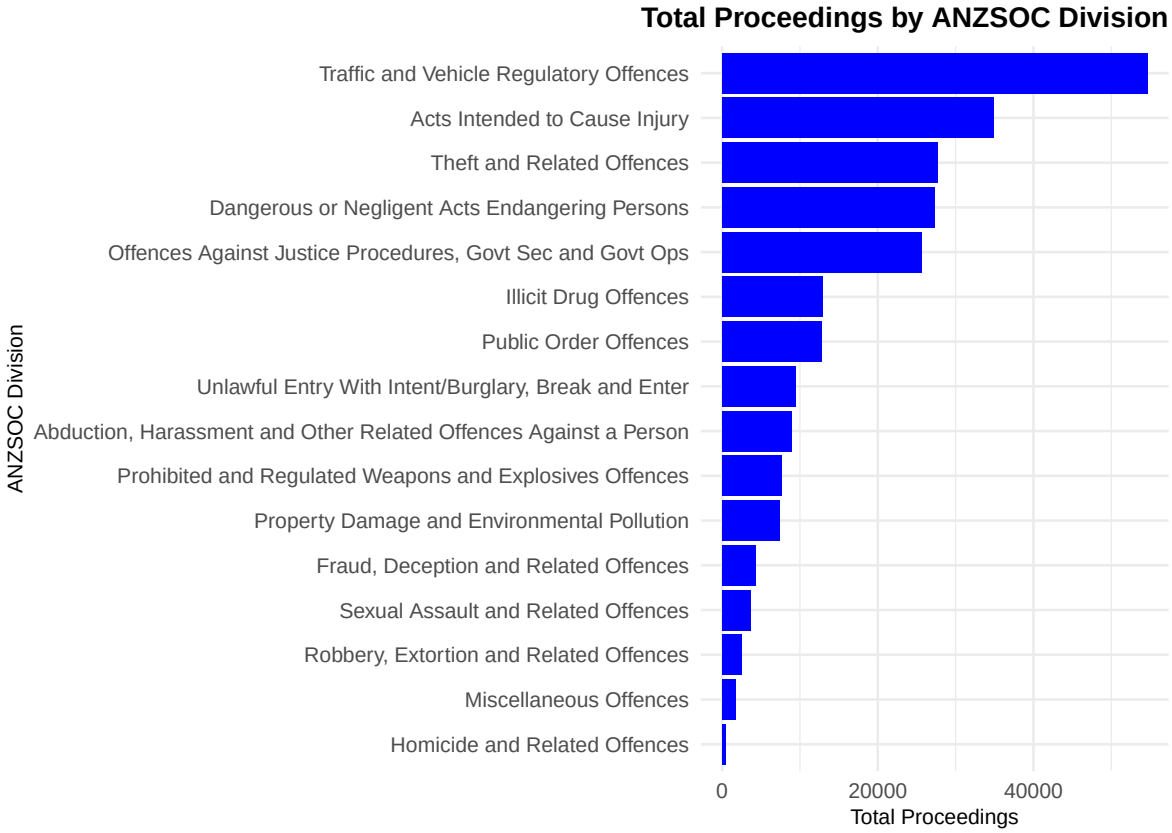
Converts specific columns into **factor variables**, ensuring they are treated as categorical data for analysis and visualization.

3. Total Proceedings by Offense Type

```r
anzsoc_division_summary <- data %>%
  group_by(ANZSOC_Division) %>%
  summarise(Number_of_Proceedings = sum(Proceedings)) %>%
  arrange(desc(Number_of_Proceedings))
```

- **`group_by(ANZSOC_Division)`**: Groups the dataset by offense type.
- **`summarise(Number_of_Proceedings = sum(Proceedings))`**: Calculates the total number of proceedings for each offense type.
- **`arrange(desc(Number_of_Proceedings))`**: Sorts the offense types in descending order by total proceedings.

```r
ggplot(anzsoc_division_summary,
       aes(x = reorder(ANZSOC_Division, Number_of_Proceedings),
           y = Number_of_Proceedings)) +
  geom_bar(stat = "identity", fill = "blue") +
  coord_flip() +
  theme_minimal() +
  labs(
       title = "Total Proceedings by ANZSOC Division",
       x = "ANZSOC Division", y = "Total Proceedings"
       ) +
  theme(
        plot.title = element_text(size = 12, face = "bold", hjust = 1),
        axis.text.y = element_text(size = 9),
        axis.text.x = element_text(size = 9),
        axis.title.x = element_text(size = 9),
        axis.title.y = element_text(size = 9)
        )
```



- **`ggplot(anzsoc_division_summary, aes(...)):`** Tells R to create a plot using the dataset `anzsoc_division_summary`. `x = reorder(ANZSOC_Division, Number_of_Proceedings)`: Orders the offense categories (`ANZSOC_Division`) based on the total number of proceedings, so the bars appear in order of size. `y = Number_of_Proceedings`: The length of the bars corresponds to the total number of proceedings.
- **`geom_bar(stat = "identity", fill = "blue"):`** Creates the bars for the chart. `stat = "identity"`: Uses the actual values of `Number_of_Proceedings` to draw the bars. `fill = "blue"`: Fills the bars with a blue color.
- **`coord_flip():`** **F**lips the x and y axes, making the bars horizontal for easier readability.
- **`theme_minimal():`** Applies a clean and simple style to the chart, removing unnecessary gridlines and clutter.
- **`labs(...):`** Adds labels to the chart. `title`: Sets the title to "Total Proceedings by ANZSOC Division". `x`: Labels the x-axis as "ANZSOC Division". `y`: Labels the y-axis as "Total Proceedings".

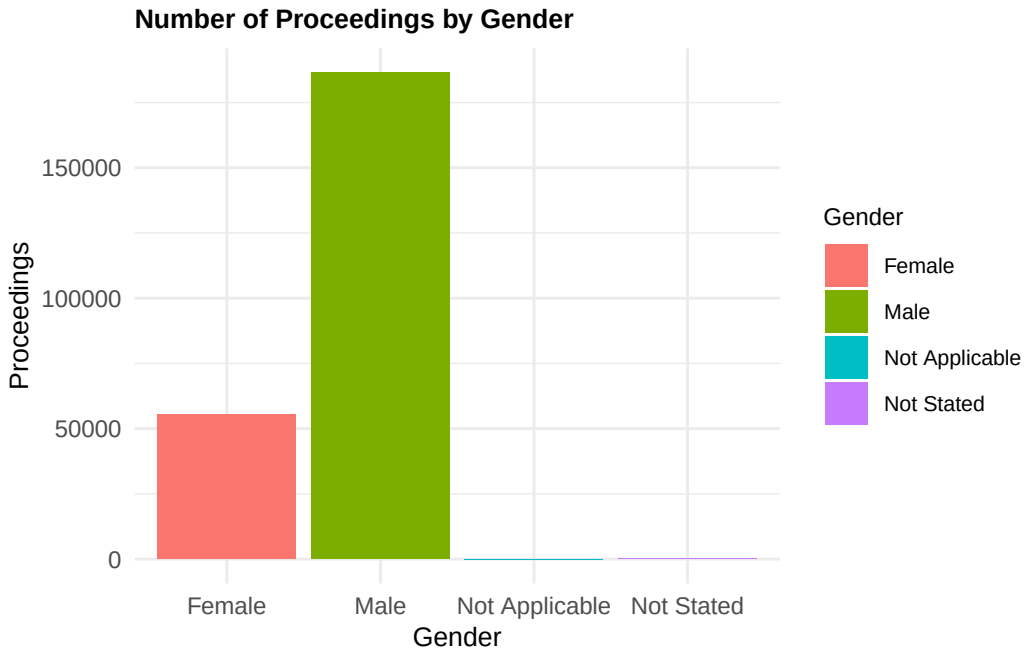
- **`theme(...)`:** Adjusts the appearance of text in the chart. `axis.text.y`: Sets the size of the text for y-axis labels (offense categories) to 9. `axis.text.x`: Sets the size of the text for x-axis labels (number values) to 9. `plot.title`: Sets the title text size to 12, makes it bold, and aligns it slightly to the right (`hjust = 1`). `axis.title.x` and `axis.title.y`: Sets the size of the x-axis and y-axis titles to 9.

4. Proceedings by Gender

```
gender_crime_summary <- aggregate(Proceedings ~ Gender, data = data, sum)
```

- **`aggregate()`**: Groups the data by gender and calculates the sum of proceedings for each group.

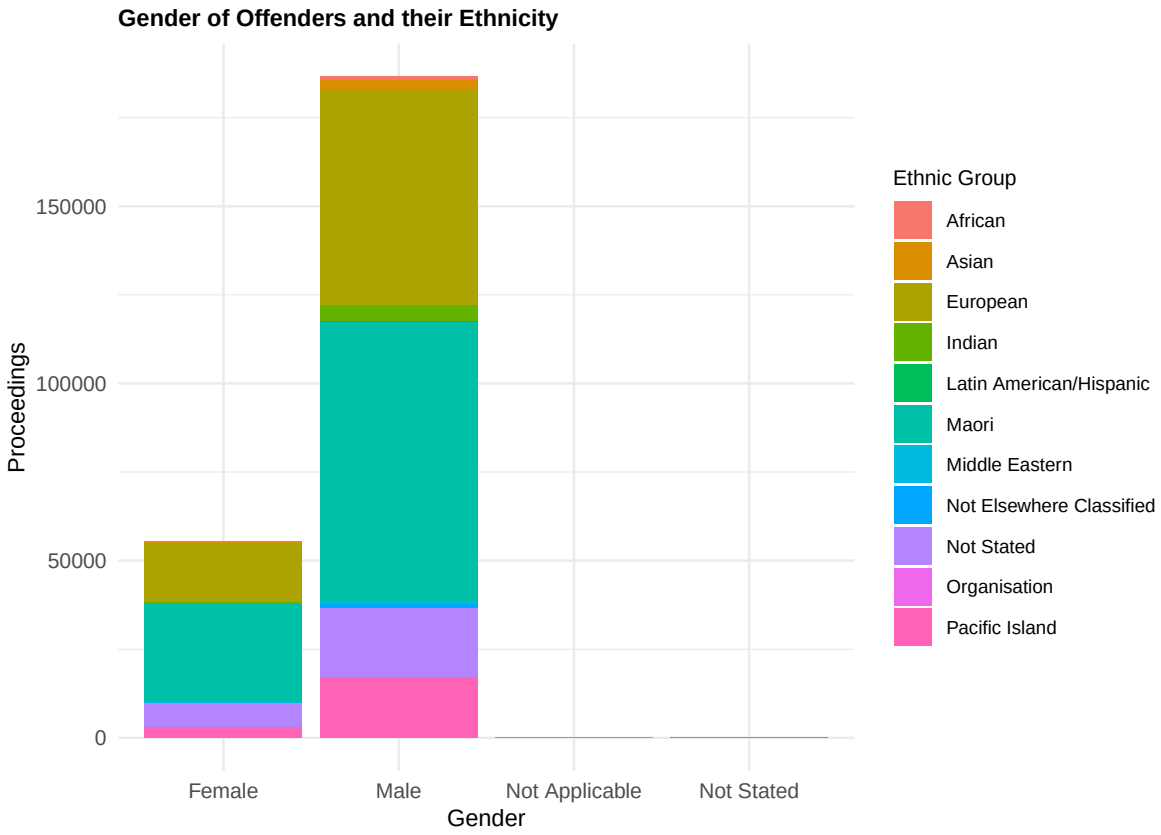
```
ggplot(gender_crime_summary, aes(x = Gender, y = Proceedings, fill = Gender)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(
    title = "Number of Proceedings by Gender",
    x = "Gender",
    y = "Proceedings"
  ) +
  theme(
    plot.title = element_text(size = 10, face = "bold"),
    axis.text.y = element_text(size = 9),
    axis.text.x = element_text(size = 9),
    axis.title.x = element_text(size = 10),
    axis.title.y = element_text(size = 10),
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 9)
  )
```



- Visualizes total proceedings by gender using a bar chart.

5. Gender and Ethnicity Breakdown

```
ggplot(data, aes(x = Gender, y = Proceedings, fill = Ethnicity)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(
    title = "Gender of Offenders and their Ethnicity",
    x = "Gender",
    y = "Proceedings",
    fill = "Ethnic Group"
  ) +
  theme(
    plot.title = element_text(size = 10, face = "bold"),
    axis.title.x = element_text(size = 10),
    axis.title.y = element_text(size = 10),
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 9),
    axis.text.y = element_text(size = 9),
    axis.text.x = element_text(size = 9)
  )
```

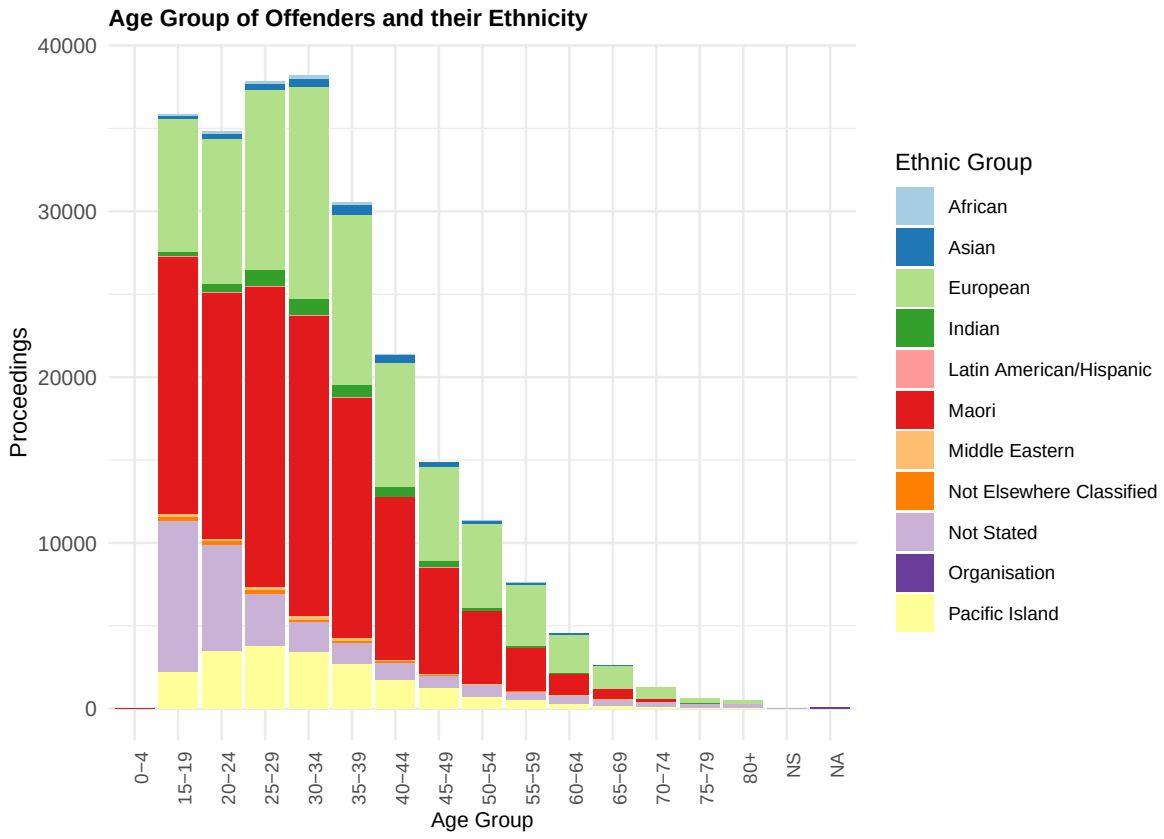


- Creates a **stacked bar chart**, where proceedings for each gender are further divided by ethnic group.

6. Age Group and Ethnicity

```
ggplot(data, aes(x = Age_Group, y = Proceedings, fill = Ethnicity)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(
       title = "Age Group of Offenders and their Ethnicity",
       x = "Age Group",
       y = "Proceedings",
       fill = "Ethnic Group"
       ) +
  theme(
        plot.title = element_text(size = 10, face = "bold"),
        axis.text.x = element_text(angle = 90, hjust = 1, size = 8),
```

```
        axis.text.y = element_text(size = 9),
        axis.title.x = element_text(size = 9),
        axis.title.y = element_text(size = 10),
        legend.text = element_text(size = 8),
        legend.title = element_text(size = 10),
        ) +
  scale_fill_brewer(palette = "Paired")
```

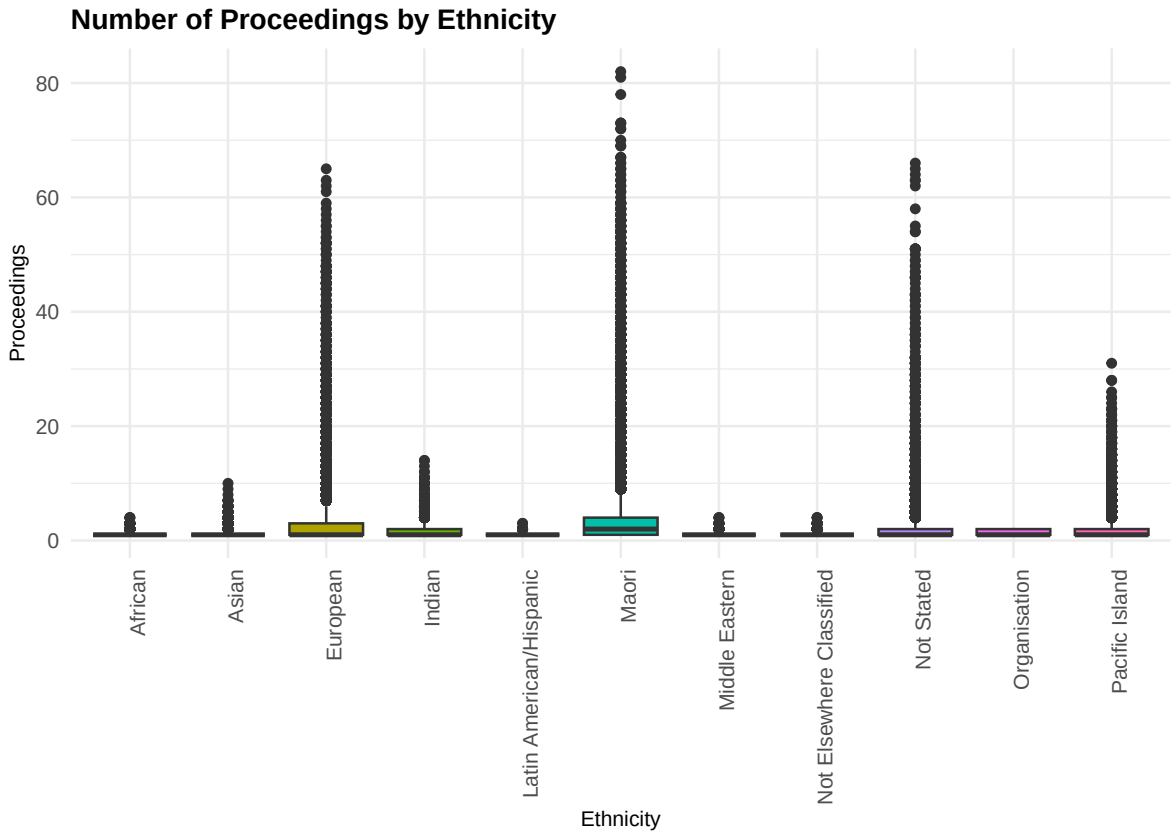


- Groups data by age and ethnicity.
- **`scale_fill_brewer(palette = "Paired")`**: Adds a predefined color palette for better visualization.

7. Boxplot by Ethnicity

```
ggplot(data, aes(x = Ethnicity, y = Proceedings, fill = Ethnicity)) +
  geom_boxplot() +
```

```
theme_minimal() +
labs(
     title = "Number of Proceedings by Ethnicity",
     x = "Ethnicity",
     y = "Proceedings"
     ) +
theme(
       plot.title = element_text(size = 12, face = "bold"),
       axis.text.x = element_text(angle = 90, hjust = 1),
       axis.text.y = element_text(size = 9),
       axis.title.x = element_text(size = 9),
       axis.title.y = element_text(size = 9),
       legend.position = "none"
      )
```



- **`geom_boxplot()`**: Creates a boxplot showing the spread of proceedings for each ethnic group.

- Highlights outliers, medians, and distributions of proceedings.

8. Descriptive Statistics

```
summary_stats <- summary(data$Proceedings)
std_dev <- sd(data$Proceedings)
variance <- var(data$Proceedings)
```

- **`summary()`**: Generates summary statistics (min, max, median, quartiles, etc.).
- **`sd()`**: Calculates the standard deviation.
- **`var()`**: Calculates the variance.

```
summary_table <- data.frame(
  Statistic = c("Min", "1st Qu.", "Median",
                "Mean", "3rd Qu.", "Max",
                "Standard Deviation", "Variance"),
  Value = c(summary_stats["Min."],
            summary_stats["1st Qu."],
            summary_stats["Median"],
            summary_stats["Mean"],
            summary_stats["3rd Qu."],
            summary_stats["Max."],
            std_dev, variance)
)
```

- Combines all summary statistics into a single table for reporting.

```
kable(summary_table, col.names = c("Statistic", "Value"),
      caption = "Descriptive Statistics for Criminal Proceedings")
```

Table 1: Descriptive Statistics for Criminal Proceedings

Statistic	Value
Min	1.000000
1st Qu.	1.000000
Median	1.000000
Mean	3.157470
3rd Qu.	3.000000
Max	82.000000
Standard Deviation	5.515392

Statistic	Value
Variance	30.419545

9. Most Common Crimes by Ethnic Group

```
crime_by_ethnicity <- data %>%
  group_by(Ethnicity, ANZSOC_Division) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count))
```

```
`summarise()` has grouped output by 'Ethnicity'. You can override using the
`.groups` argument.
```

```
# To remove all grouping and produce a flat (ungrouped) data frame:
# summarise(Count = n(), .groups = "drop")

# To keep the data grouped at the current level:
# summarise(Count = n(), .groups = "keep")
```

If we don't need the grouping structure after summarization, always use .groups = "drop". This ensures the result is ungrouped and avoids any unintended issues in downstream operations. If we do want the grouped structure for further operations, let the default behavior apply or explicitly set .groups = "keep".

- **`group_by(Ethnicity, ANZSOC_Division)`**: Groups data by ethnicity and offense type.
- **`summarise(Count = n())`**: Counts the number of occurrences for each ethnicity-offense pair.

```
most_common_crime_by_ethnicity <- crime_by_ethnicity %>%
  group_by(Ethnicity) %>%
  slice_max(Count, n = 1) %>%
  arrange(desc(Count)) %>%
  select(Ethnicity, ANZSOC_Division, Count)
```

- Finds the most common offense for each ethnic group by selecting the row with the maximum count.
- **`slice_max(Count, n = 1)`**: For each ethnic group (created by `group_by`), selects the row with the highest value in the `Count` column. `n = 1` ensures only the row with the maximum value is retained.

```
kable(most_common_crime_by_ethnicity, caption = "Most Common Crimes by Ethnic Group")
```

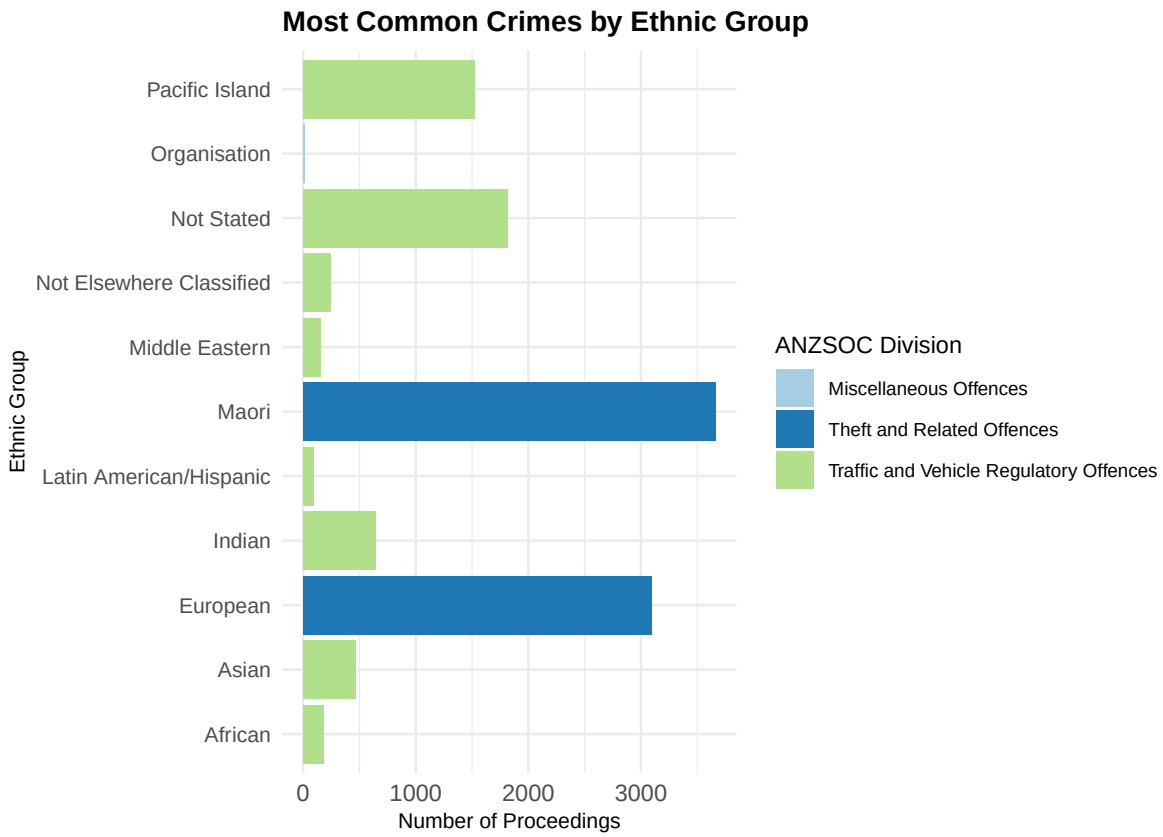
Table 2: Most Common Crimes by Ethnic Group

Ethnicity	ANZSOC_Division	Count
Maori	Theft and Related Offences	3664
European	Theft and Related Offences	3095
Not Stated	Traffic and Vehicle Regulatory Offences	1815
Pacific Island	Traffic and Vehicle Regulatory Offences	1519
Indian	Traffic and Vehicle Regulatory Offences	647
Asian	Traffic and Vehicle Regulatory Offences	468
Not Elsewhere Classified	Traffic and Vehicle Regulatory Offences	243
African	Traffic and Vehicle Regulatory Offences	178
Middle Eastern	Traffic and Vehicle Regulatory Offences	158
Latin American/Hispanic	Traffic and Vehicle Regulatory Offences	97
Organisation	Miscellaneous Offences	14

- Visualizes the most common crimes for each ethnicity using a horizontal bar chart.

```
ggplot(most_common_crime_by_ethnicity,
       aes(x = Ethnicity, y = Count, fill = ANZSOC_Division)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(
    title = "Most Common Crimes by Ethnic Group",
    x = "Ethnic Group",
    y = "Number of Proceedings",
    fill = "ANZSOC Division"
  ) +
  theme(
    plot.title = element_text(size = 12, face = "bold"),
    axis.title.x = element_text(size = 9),
    axis.title.y = element_text(size = 9),
    axis.text.x = element_text(size = 10),
```

```
    axis.text.y = element_text(size = 9),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 8)
  ) +
  scale_fill_brewer(palette = "Paired")
```



10. Appendix

Summarize total proceedings by ANZSOC_Division and arrange in descending order

```
# Summarize total proceedings by ANZSOC_Division and arrange in descending order
anzsoc_division_counts <- data %>%
  group_by(ANZSOC_Division) %>%
  summarise(Total_Proceedings = sum(Proceedings, na.rm = TRUE)) %>%
  arrange(desc(Total_Proceedings))
```

```
# View the summarized data in descending order
kable(anzsoc_division_counts, caption = "Total Proceedings by ANZSOC Division")
```

Table 3: Total Proceedings by ANZSOC Division

ANZSOC_Division	Total_Proceedings
Traffic and Vehicle Regulatory Offences	54627
Acts Intended to Cause Injury	34935
Theft and Related Offences	27728
Dangerous or Negligent Acts Endangering Persons	27272
Offences Against Justice Procedures, Govt Sec and Govt Ops	25654
Illicit Drug Offences	12903
Public Order Offences	12785
Unlawful Entry With Intent/Burglary, Break and Enter	9459
Abduction, Harassment and Other Related Offences Against a Person	8867
Prohibited and Regulated Weapons and Explosives Offences	7667
Property Damage and Environmental Pollution	7446
Fraud, Deception and Related Offences	4343
Sexual Assault and Related Offences	3672
Robbery, Extortion and Related Offences	2518
Miscellaneous Offences	1739
Homicide and Related Offences	424

Summarize the number of crimes by Gender (Sex)

```
# Summarize the number of crimes by Gender (Sex)
gender_crime_summary <- aggregate(Proceedings ~ Gender, data = data, sum)

# Sort the table in descending order by Total Proceedings
gender_crime_summary <- gender_crime_summary[order(-gender_crime_summary$Proceedings), ]

# Display the summarized table using kable
kable(gender_crime_summary, col.names = c("Gender", "Total Proceedings"),
      caption = "Total Criminal Proceedings by Gender")
```

Table 4: Total Criminal Proceedings by Gender

	Gender	Total Proceedings
2	Male	186485
1	Female	55396
4	Not Stated	112
3	Not Applicable	46

Summarize the total proceedings by Age Group

```
# Summarize the total proceedings by Age Group
age_group_summary <- data %>%
  group_by(Age_Group) %>%
  summarise(Total_Proceedings = sum(Proceedings, na.rm = TRUE)) %>%
  arrange(desc(Total_Proceedings))

# Display the summarized table using kable
kable(age_group_summary, col.names = c("Age Group", "Total Proceedings"),
      caption = "Total Criminal Proceedings by Age Group")
```

Table 5: Total Criminal Proceedings by Age Group

Age Group	Total Proceedings
30-34	38172
25-29	37845
15-19	35845
20-24	34836
35-39	30543
40-44	21368
45-49	14868
50-54	11323
55-59	7593
60-64	4525
65-69	2622
70-74	1302
75-79	629
80+	496
NA	46
NS	25

Age Group	Total Proceedings
0-4	1

Calculate the proportions for each Ethnic Group and arrange in descending order

```
# Create ethnic group counts
ethnic_group_counts <- data %>%
  group_by(Ethnicity) %>%
  summarise(Count = n())

# Calculate the proportions for each Ethnic Group and arrange in descending order
ethnic_group_proportions <- ethnic_group_counts %>%
  mutate(Proportion = Count / sum(Count)) %>%
  arrange(desc(Proportion))

# View the sorted proportions
kable(ethnic_group_proportions,
      caption = "Proportions of Criminal Proceedings by Ethnic Group")
```

Table 6: Proportions of Criminal Proceedings by Ethnic Group

Ethnicity	Count	Proportion
Maori	24771	0.3231450
European	24126	0.3147307
Not Stated	9782	0.1276091
Pacific Island	8896	0.1160509
Indian	3132	0.0408579
Asian	2384	0.0311000
Not Elsewhere Classified	1129	0.0147281
African	1010	0.0131757
Middle Eastern	1007	0.0131366
Latin American/Hispanic	383	0.0049963
Organisation	36	0.0004696

Finding the most common crime for each ethnic group and showing their count

```r
# Finding the most common crime for each ethnic group and showing their count
most_common_crime_by_ethnicity <- crime_by_ethnicity %>%
  group_by(Ethnicity) %>%
  slice_max(Count, n = 1) %>%
  arrange(desc(Count)) %>%
  select(Ethnicity, ANZSOC_Division, Count)

# View the result
kable(most_common_crime_by_ethnicity,
      caption = "Most Common Crimes by Ethnic Group")
```

Table 7: Most Common Crimes by Ethnic Group

Ethnicity	ANZSOC_Division	Count
Maori	Theft and Related Offences	3664
European	Theft and Related Offences	3095
Not Stated	Traffic and Vehicle Regulatory Offences	1815
Pacific Island	Traffic and Vehicle Regulatory Offences	1519
Indian	Traffic and Vehicle Regulatory Offences	647
Asian	Traffic and Vehicle Regulatory Offences	468
Not Elsewhere Classified	Traffic and Vehicle Regulatory Offences	243
African	Traffic and Vehicle Regulatory Offences	178
Middle Eastern	Traffic and Vehicle Regulatory Offences	158
Latin American/Hispanic	Traffic and Vehicle Regulatory Offences	97
Organisation	Miscellaneous Offences	14

Conclusion

This project analyzed crime trends in New Zealand, focusing on how gender, ethnicity, and age influence criminal activities. The findings show that men are responsible for most criminal proceedings, suggesting a need for programs that address their involvement in crimes.

The Maori ethnic group is overrepresented in criminal proceedings, followed by Europeans and Pacific Islanders, while other ethnicities, such as Asians and Indians, are less involved. This highlights the need for culturally tailored solutions to reduce crime rates among these groups.

Young adults aged 15–39 are the most involved in criminal activities, with fewer crimes committed by older age groups. Early intervention programs for youth could help prevent re-offending and address the causes of their involvement in crimes.

The most common offenses are traffic violations, acts intended to cause injury, and theft. Efforts to reduce these crimes, such as stricter enforcement and public awareness, could have a big impact.

These findings provide useful insights for developing crime prevention strategies that create a more safer and fairer society in New Zealand.