

# Project

Anju Sambasivan

2024-09-07

```
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr)
library(dplyr)
library(knitr)

# Read the data
data <- read_csv("data.csv", show_col_types = FALSE)

# Extract structure information
data_structure <- data.frame(
  Column_Names = names(data),
  # DataType = sapply(data, class),
  stringsAsFactors = FALSE
)

# Display the structure in a table format
kable(data_structure)
```

Column_Names
Date
Victimisations
AgeGroup
AnzsocDivision
LocationType
PoliceArea
ROVDivision
TerritorialAuthority
PersonOrOrganization

```
summary(data)
```

```
##      Date      Victimisations      AgeGroup      AnzsocDivision
## Length:1048575 Min.   : 1.000   Length:1048575 Length:1048575
## Class :character 1st Qu.: 1.000   Class :character Class :character
## Mode  :character Median : 1.000   Mode  :character Mode  :character
##                      Mean  : 2.248
##                      3rd Qu.: 2.000
##                      Max.   :605.000
## LocationType      PoliceArea      ROVDivision      TerritorialAuthority
## Length:1048575    Length:1048575    Length:1048575    Length:1048575
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## PersonOrOrganization
## Length:1048575
## Class :character
## Mode  :character
##
##
##
```

```
# Change into factor Categorical
```

```
data$AnzsocDivision <- as.factor(data$AnzsocDivision)
data$LocationType <- as.factor(data$LocationType)
data$TerritorialAuthority <- as.factor(data$TerritorialAuthority)
data$PoliceArea <- as.factor(data$PoliceArea)
data$Date <- as.Date(data$Date, format = "%y/%m/%d")
data$PersonOrOrganization <- as.factor(data$PersonOrOrganization)
data$AgeGroup <- as.factor(data$AgeGroup)
data$ROVDivision <- as.factor(data$ROVDivision)
```

```
str(data)
```

```
## spc_tbl_ [1,048,575 x 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Date      : Date[1:1048575], format: NA NA ...
## $ Victimisations : num [1:1048575] 1 1 1 1 1 1 1 1 1 ...
## $ AgeGroup      : Factor w/ 5 levels "10-19","0-9",...: 4 4 4 4 4 4 4 4 4 ...
## $ AnzsocDivision : Factor w/ 6 levels "Abduction, Harassment and Other Related Offences Againsts",...: 32 17 16 32 32 32 32 32 32 ...
## $ LocationType  : Factor w/ 36 levels "Administrative/Professional",...: 32 17 16 32 32 32 32 32 32 ...
## $ PoliceArea     : Factor w/ 39 levels "Auckland Central Area",...: 31 27 10 7 7 7 5 35 4 1 ...
## $ ROVDivision    : Factor w/ 5 levels "Known To Victim",...: 2 2 2 2 2 2 2 2 2 ...
## $ TerritorialAuthority: Factor w/ 68 levels "Area Outside Territorial Authority.",...: 56 13 3 3 3 3 3 3 3 ...
## $ PersonOrOrganization: Factor w/ 2 levels "Organisation",...: 1 1 1 1 1 1 1 1 1 ...
## - attr(*, "spec")=
## .. cols(
## ..   Date = col_character(),
## ..   Victimisations = col_double(),
## ..   AgeGroup = col_character(),
## ..   AnzsocDivision = col_character(),
## ..   LocationType = col_character(),
## ..   PoliceArea = col_character(),
## ..   ROVDivision = col_character(),
```

```
## .. TerritorialAuthority = col_character(),
## .. PersonOrOrganization = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
summary(data)
```

```
##      Date      Victimisations      AgeGroup
## Min.   :NA      Min.    : 1.000  '10-19      :107557
## 1st Qu.:NA      1st Qu.: 1.000  0-9         : 25285
## Median :NA      Median : 1.000  20+         :558161
## Mean   :NaN      Mean   : 2.248  Not Applicable:259975
## 3rd Qu.:NA      3rd Qu.: 2.000  Not Specified : 97597
## Max.   :NA      Max.    :605.000
## NA's    :1048575
##
##                                     AnzsocDivision
## Abduction, Harassment and Other Related Offences Against a Person: 3072
## Acts Intended to Cause Injury                                     :353043
## Robbery, Extortion and Related Offences                          : 25732
## Sexual Assault and Related Offences                             : 46085
## Theft and Related Offences                                       :479496
## Unlawful Entry With Intent/Burglary, Break and Enter            :141147
##
##      LocationType      PoliceArea
## Dwelling - Private :204429  Canterbury Metro Area : 76406
## Street/Footpath    :130614  Manawatu Area          : 49279
## Retail N.E.C.      : 94942  Hawkes Bay Area        : 47066
## Dwelling N.F.D.    : 83201  Western Bay Of Plenty Area: 43811
## Unspecified Location: 78066  Hamilton City Area      : 41591
## Dwelling Land      : 76783  Waitemata West Area     : 40318
## (Other)            :380540  (Other)                 :750104
##
##      ROVDivision      TerritorialAuthority
## Known To Victim      :136587  Auckland.              :302441
## No Offender Identified:632988  Christchurch City.     : 76914
## Not Applicable       :173523  Hamilton City.         : 38776
## Not Stated           : 33979  Wellington City.       : 33114
## Stranger             : 71498  Tauranga City.         : 33032
##                      :         Far North District.: 25504
##                      :         (Other)             :538794
##
##      PersonOrOrganization
## Organisation:259975
## Person      :788600
##
##
##
##
```

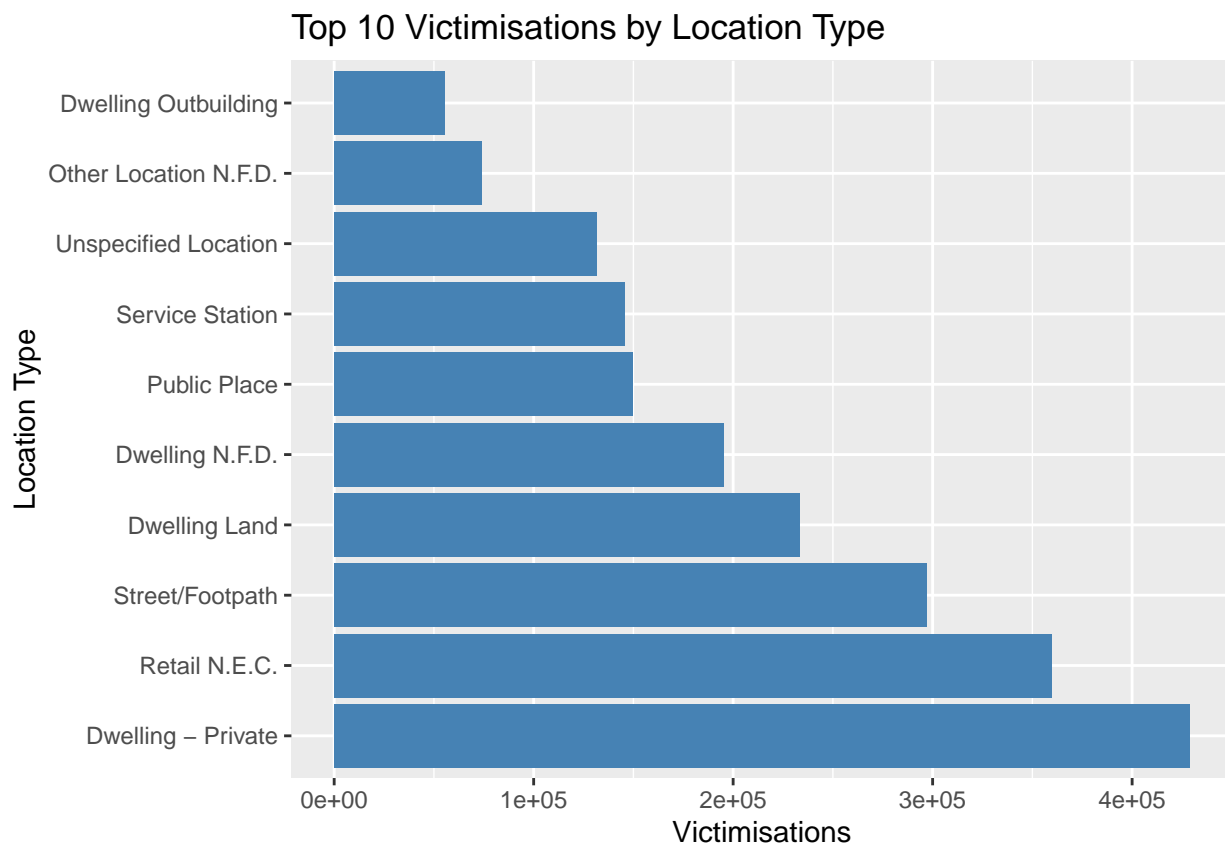
```
location_summary <- data %>%
  group_by(LocationType) %>%
  summarise(Victimisations = sum(Victimisations, na.rm = TRUE)) %>%
  arrange(desc(Victimisations))
```

```
# Sort the data by Victimisations in descending order
```

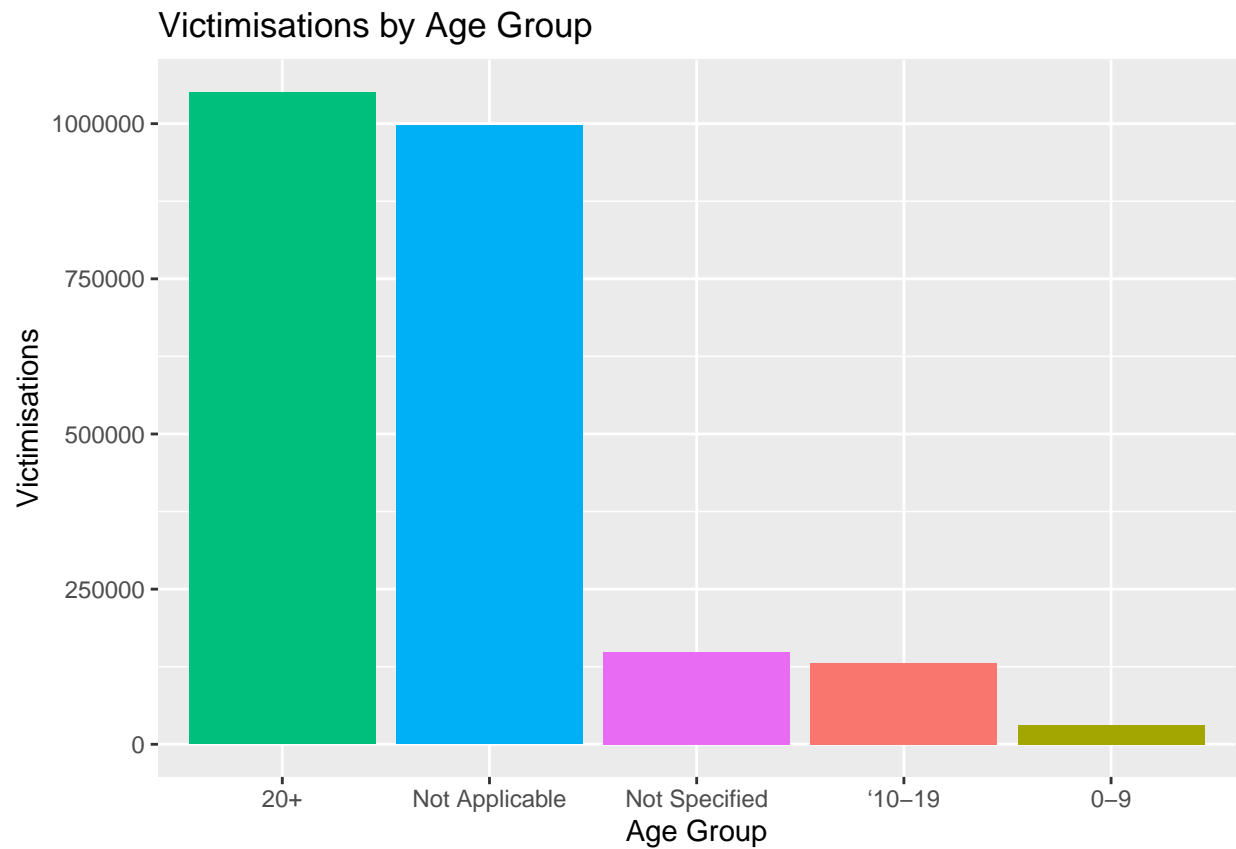
```
location_summary <- location_summary[order(-location_summary$Victimisations), ]
```

```
# Select the top 10 locations
top_10_locations <- head(location_summary, 10)

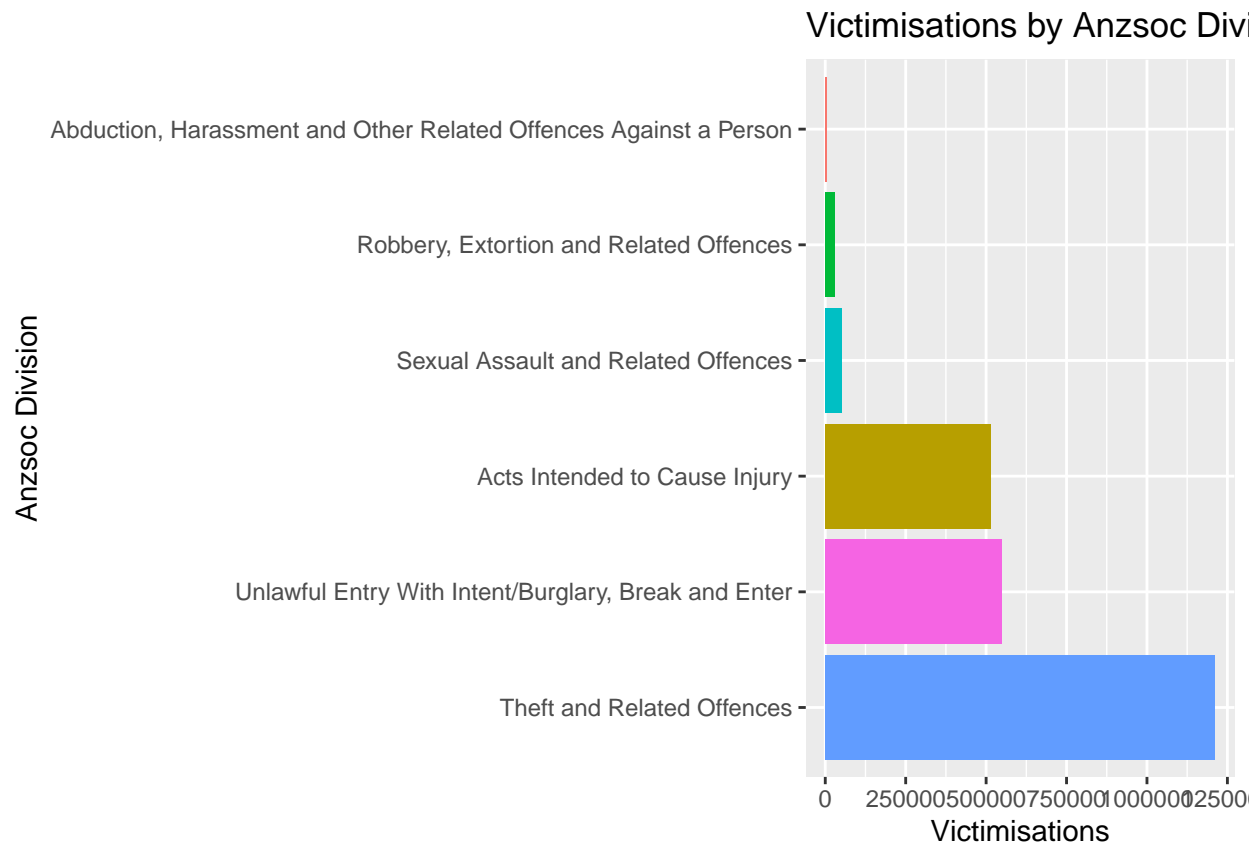
ggplot(top_10_locations, aes(x=reorder(LocationType, -Victimisations), y=Victimisations)) +
  geom_bar(stat="identity", fill="steelblue") +
  coord_flip() +
  labs(title="Top 10 Victimisations by Location Type", x="Location Type", y="Victimisations") +
  theme(legend.position = "none") # Remove the legend if not needed
```



```
age_group_summary <- aggregate(Victimisations ~ AgeGroup, data=data, sum)
ggplot(age_group_summary, aes(x=reorder(AgeGroup, -Victimisations), y=Victimisations, fill=AgeGroup)) +
  geom_bar(stat="identity") +
  labs(title="Victimisations by Age Group", x="Age Group", y="Victimisations") +
  theme(legend.position = "none")
```



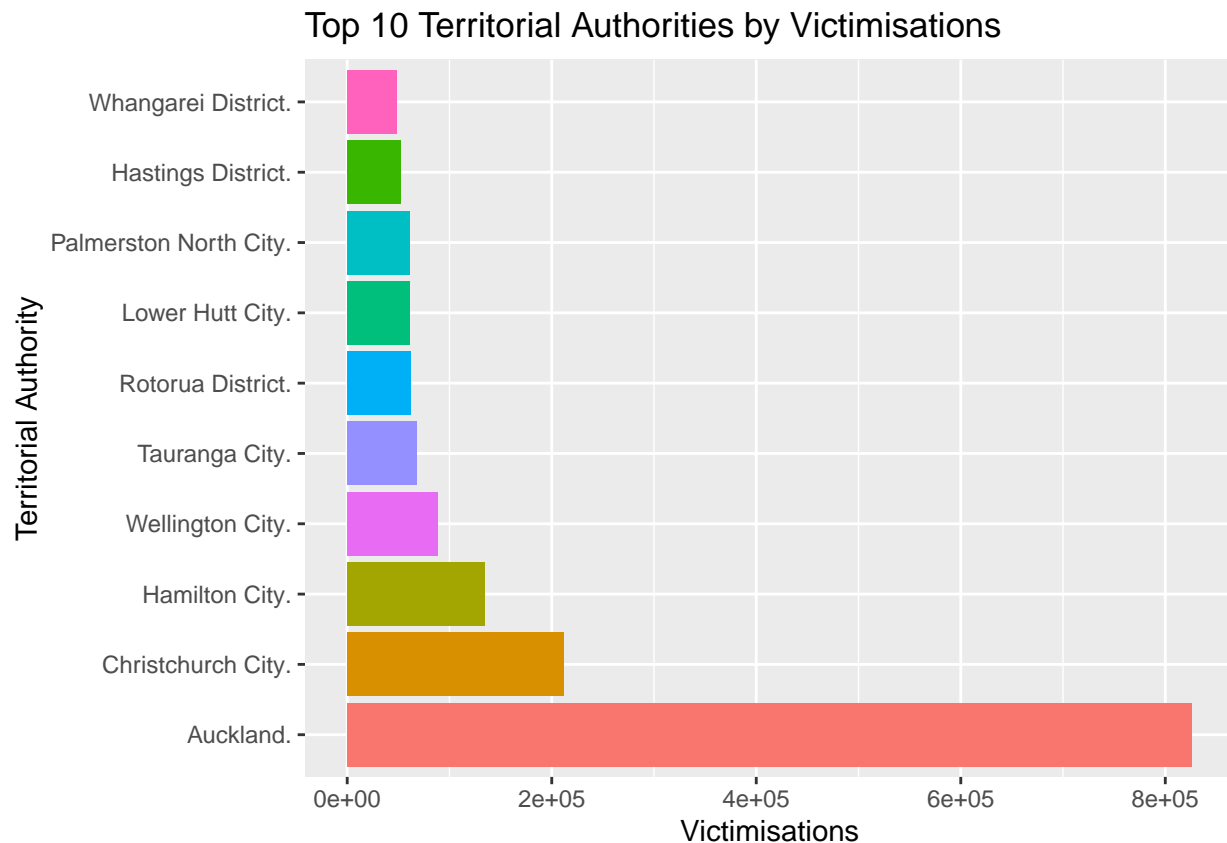
```
anzsoc_summary <- aggregate(Victimisations ~ AnzsocDivision, data=data, sum)
ggplot(anzsoc_summary, aes(x=reorder(AnzsocDivision, -Victimisations), y=Victimisations, fill=AnzsocDiv
  geom_bar(stat="identity") +
  coord_flip() +
  labs(title="Victimisations by Anzsoc Division", x="Anzsoc Division", y="Victimisations") +
  theme(legend.position = "none")
```



```
# Summarizing data by TerritorialAuthority
territorial_summary <- data %>%
  group_by(TerritorialAuthority) %>%
  summarise(Victimisations = sum(Victimisations, na.rm = TRUE)) %>%
  arrange(desc(Victimisations))

# Sort the territorial_summary dataframe by Victimisations in descending order and select top 10
top_10_territorial <- territorial_summary[order(-territorial_summary$Victimisations), ][1:10, ]

# Plotting the top 10 Territorial Authorities
ggplot(top_10_territorial, aes(x=reorder(TerritorialAuthority, -Victimisations), y=Victimisations, fill=TerritorialAuthority)) +
  geom_bar(stat="identity") +
  coord_flip() +
  labs(title="Top 10 Territorial Authorities by Victimisations", x="Territorial Authority", y="Victimisations") +
  theme(legend.position = "none")
```



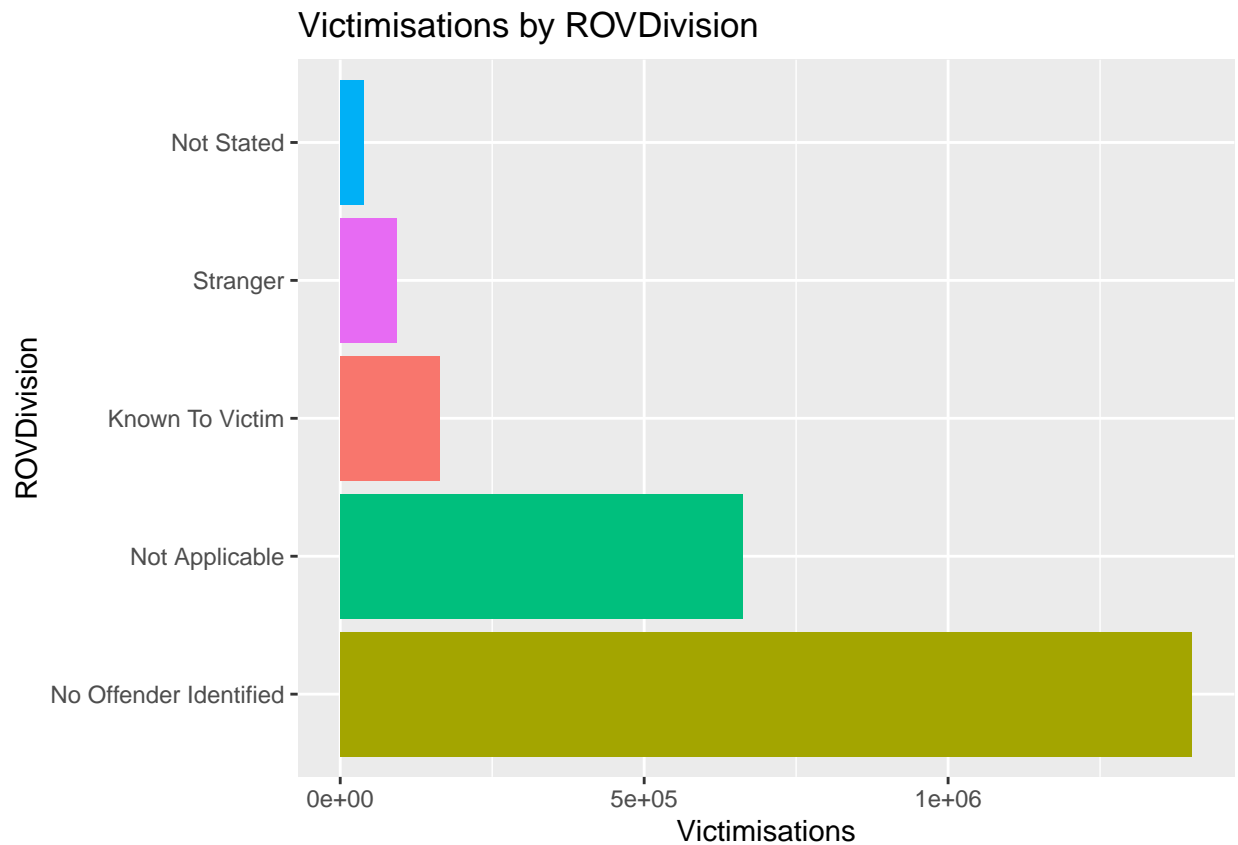
```
# Summarize victimisations by ROVDivision
rovddivision_summary <- aggregate(Victimisations ~ ROVDivision, data=data, sum)

# Sort the summary to get the top ROVDivisions
rovddivision_summary <- rovddivision_summary[order(-rovddivision_summary$Victimisations), ]

# Display the ROVDivision with the most victimisations
head(rovddivision_summary, 10)
```

```
##          ROVDivision Victimisations
## 2 No Offender Identified      1400950
## 3      Not Applicable         661941
## 1      Known To Victim       163070
## 5              Stranger        92259
## 4      Not Stated           39329
```

```
# Plot the victimisations by ROVDivision
ggplot(rovddivision_summary, aes(x=reorder(ROVDivision, -Victimisations), y=Victimisations, fill=ROVDivision)) +
  geom_bar(stat="identity") +
  coord_flip() +
  labs(title="Victimisations by ROVDivision", x="ROVDivision", y="Victimisations") +
  theme(legend.position = "none")
```



```
# Summarize victimisations by PoliceArea
policearea_summary <- aggregate(Victimisations ~ PoliceArea, data=data, sum)

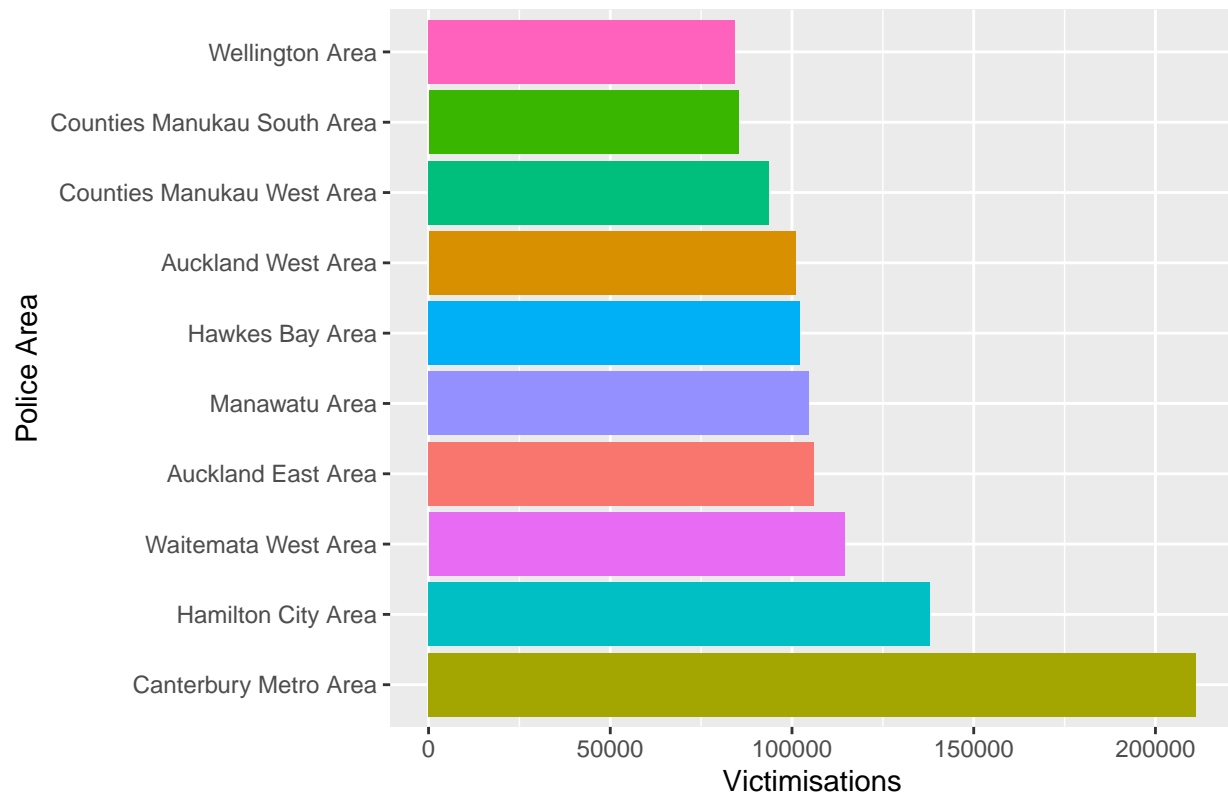
# Sort the summary to get the top PoliceAreas
policearea_summary <- policearea_summary[order(-policearea_summary$Victimisations), ]

# Subset the data to the top 10 PoliceAreas
top_10_policeareas <- head(policearea_summary, 10)

# Plot the victimisations for the top 10 PoliceAreas
ggplot(top_10_policeareas, aes(x=reorder(PoliceArea, -Victimisations), y=Victimisations, fill=PoliceArea)) +
  geom_bar(stat="identity") +
  coord_flip() +
  labs(title="Top 10 Police Areas by Victimisations", x="Police Area", y="Victimisations") +
  theme(legend.position = "none")
```



### Top 10 Police Areas by Victimisations

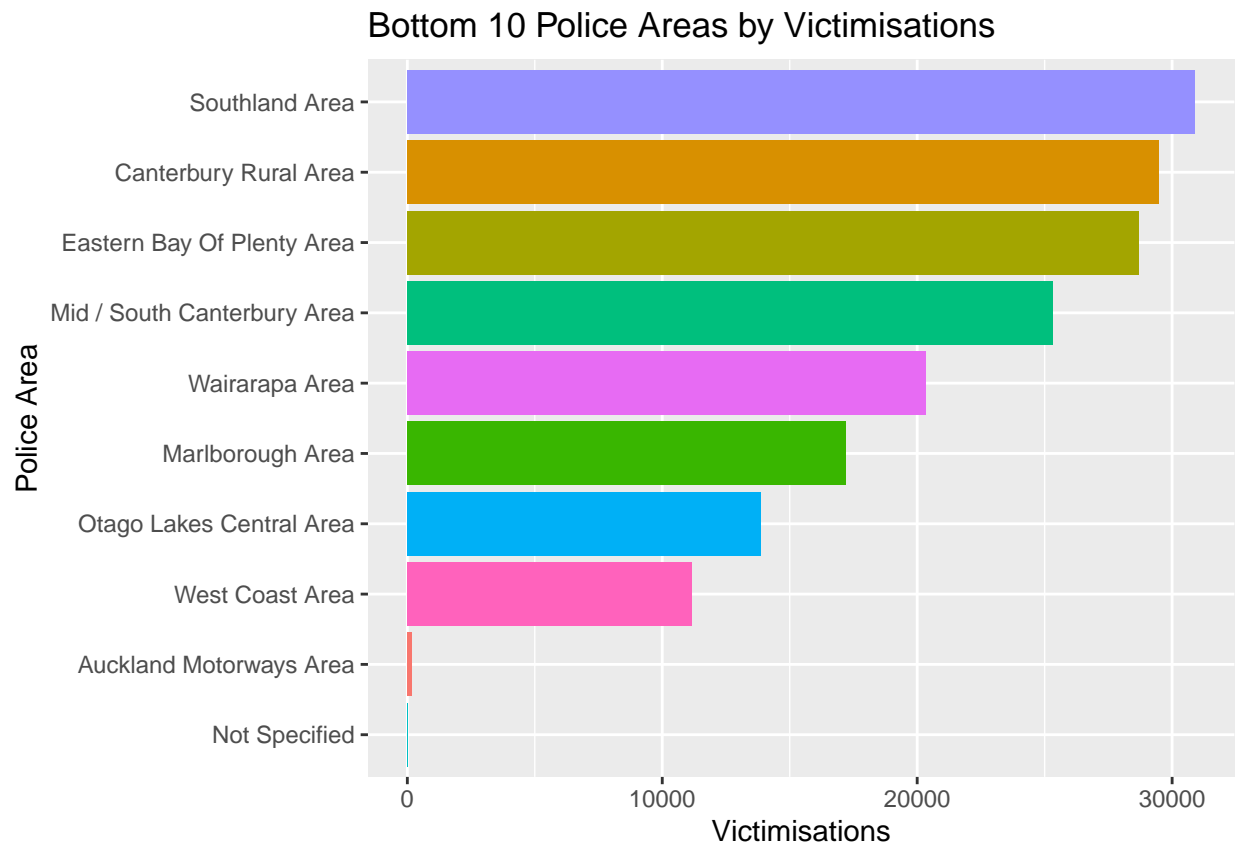


```
# Summarize victimisations by PoliceArea
policearea_summary <- aggregate(Victimisations ~ PoliceArea, data=data, sum)

# Sort the summary to get the bottom PoliceAreas
policearea_summary <- policearea_summary[order(policearea_summary$Victimisations), ]

# Subset the data to the bottom 10 PoliceAreas
bottom_10_policeareas <- head(policearea_summary, 10)

# Plot the victimisations for the bottom 10 PoliceAreas
ggplot(bottom_10_policeareas, aes(x=reorder(PoliceArea, Victimisations), y=Victimisations, fill=PoliceArea)) +
  geom_bar(stat="identity") +
  coord_flip() +
  labs(title="Bottom 10 Police Areas by Victimisations", x="Police Area", y="Victimisations") +
  theme(legend.position = "none")
```



```
# Summarize victimisations by PersonOrOrganization
personorg_summary <- aggregate(Victimisations ~ PersonOrOrganization, data=data, sum)

# Plot the victimisations by PersonOrOrganization
ggplot(personorg_summary, aes(x=reorder(PersonOrOrganization, -Victimisations), y=Victimisations, fill=Victimisations)) +
  geom_bar(stat="identity") +
  labs(title="Victimisations by Person or Organization", x="Person or Organization", y="Victimisations") +
  theme_minimal() +
  theme(legend.position = "none")
```

