Exploring Criminal Offences and Crime Patterns in Proximity to University Campuses in Victoria, Australia

Salman Tahir

Contents

1 Introduction **1**

2 Data Sources **2**

3 Data Wrangling **3**

3.1 Import Libraries 3

3.2 Load Tabular Datasets 4

3.2.1 University Locations 4

3.2.2 Recorded Offences by Offence Type 4

3.3 Load Spatial Datasets 5

3.3.1 Victorian Local Government Area (LGA) Boundaries 5

3.3.2 Victorian Suburbs Boundaries 7

4 Data Analysis **9**

4.1 Prevalent Crime Trends near University Campuses 10

4.2 Correlating Offences in University Suburbs with their Neighbouring suburbs (LGAs) 18

4.3 Exploring Patterns in Criminal Offences 26

5 Conclusion **30**

6 Bibliography **31**

1 Introduction

The safety and well-being of university students is of paramount importance to educational institutions, law enforcement agencies and society at large. While universities are generally considered secure environments for intellectual and personal growth, they are not immune to criminal activities. Incidents of theft, assault and other forms of misconduct can have long-lasting psychological and emotional impacts on victims, as well as the wider university community.

The motivation for this capstone project in Data Visualisation and Analytics stemmed from a personal experience at Monash University, Clayton. An unfortunate incident involving the theft of my personal belongings, specifically Apple AirPods, from a temporarily vacated study space led to concerns about the security of personal items on university campuses. This incident, albeit minor, not only left a feeling of vulnerability but also raised broader questions about the prevalence of such incidents around educational institutions in Victoria, Australia. To better understand the scope and nature of the problem, this project aims to answer the following research questions:

1. What are the prevalent crime trends in areas surrounding university campuses across Victoria, Australia?
2. Are there specific types of criminal offences that occur more frequently near university campuses compared to their neighbouring suburbs?
3. What patterns in recorded criminal offences can be identified between suburbs in proximity to university campuses and the broader local government areas (LGAs)?

These research questions aim to delve into the nuances of criminal activities in proximity to university campuses. By investigating various types of crimes and comparing them with broader geographic locations, this project also aims to uncover any discernible patterns or trends. Furthermore, the significance of this study extends beyond academic curiosity, as the findings have potential to influence institutional policies and law enforcement strategies. By identifying criminal hotspots and trends, universities (Safer Community Unit 2023) and law enforcement agencies can allocate their resources more efficiently. Additionally, this study also serves as a data-driven foundation for community engagement and awareness, empowering individuals to take precautionary measures.

2 Data Sources

- **AURIN - National Education Facilities - Universities 2018**
  - Provides geographic coordinates, addresses and location attributes of university locations (educational institutes and campuses) in Australia.
  - Structure: Tabular data in CSV (Comma Separated Values) format, with 268 rows and 15 columns. The data includes text and spatial attributes such as (longitude, latitude and multipoint geometry).
  - Link to dataset: https://data.aurin.org.au/dataset/aurin-aurin-national-education-dataset-universities-2018-na
- **Crime Statistics Agency - LGA Recorded Offences Year Ending March 2023**
  - Contains counts of recorded offences by offence type, in LGAs and suburbs, detailing offence division, subdivision and subgroup.
  - Structure: Tabular data in XLSX (Microsoft Excel) format with 359133 rows and 9 columns. The data includes date, text and numerical attributes.
  - Link to dataset: https://files.crimestatistics.vic.gov.au/2023-06/Data_Tables_LGA_Recorded_Offences_Year_Ending_March_2023.xlsx
- **Victoria Local Government Areas (LGAs) Boundaries**
  - Geospatial dataset containing geometry points for the LGA boundaries in Victoria.
  - Structure: Geospatial shapefiles with polygon geometry and coordinate reference system (CRS): GDA2020/EPSG:7844
  - Link to dataset: https://data.gov.au/dataset/ds-dga-bdf92691-c6fe-42b9-a0e2-a4cd716fa811/details
- **Victoria Suburb Boundaries**
  - Geospatial dataset containing geometry points for suburb boundaries in Victoria.
  - Structure: Geospatial shapefiles with polygon geometry and coordinate reference system (CRS): GDA94/EPSG:4938

- Link to dataset: https://data.gov.au/data/dataset/vic-suburb-locality-boundaries-geoscape-administrative-boundaries

3 Data Wrangling

In the context of our project, data wrangling served a major role for subsequent downstream analyses. It allowed us to filter out irrelevant information, handle missing values and merge datasets to correlate information. Our project contains multiple datasets from various sources, each with its own set of characteristics, missing values and potential outliers, making data wrangling an essential step of the overall analytics process.

The raw data from the above sources was cleansed and transformed to create structured datasets. This process was done in Jupyter Notebook using R and the notebook is made available in the repository. The following is a summary of the steps taken to clean and transform the data:

1. Loading and Inspecting Data The datasets were loaded as dataframes and inspected to understand the structure of the data. The initial inspection of the datasets gives us an overall view of the data, the attributes and their types.
2. Filtering for VIC universities The university locations dataset was filtered to only contain universities belonging to the state of Victoria, as our project primarily focuses on crime patterns in Victoria. Similarly, when focusing on a specific suburb, LGA or offence type, data filtering was performed.
3. Column Reduction Unnecessary columns were dropped from both datasets, and specific columns were renamed for easier manipulation and better readability. Further to this, the case sensitivity was also inspected and handled to ensure smooth interoperability of data.
4. Handling NULL values The datasets were checked for any NULL values, no null values were found.
5. Handling duplicate values Some instances of data pertaining to university locations had duplicates, these were identified and removed to ensure data redundancies do not occur.
6. Exporting clean data for downstream analysis The cleaned datasets were then exported for subsequent downstream analyses.

The cleaned datasets were then imported into R/RStudio for further analysis and visualisation.

3.1 Import Libraries

```r
library(tidyverse)
library(leaflet)
library(sf)
library(scales)
library(ggthemes)
library(treemap)
library(treemapify)
library(fmsb)
library(viridis)
library(htmlwidgets)
```

3.2 Load Tabular Datasets

3.2.1 University Locations

```r
# Read university locations data
df_uni <- read_csv("../data/input/VIC_Universities.csv")
```

```
## Rows: 80 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): institution, campus_name, address, suburb, wkb_geometry
## dbl (3): postcode, latitude, longitude
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
dim(df_uni)
```

```
## [1] 80  8
```

```r
head(df_uni)
```

```
## # A tibble: 6 x 8
##   institution              campus_name address suburb postcode latitude longitude
##   <chr>                    <chr>       <chr>   <chr>     <dbl>    <dbl>     <dbl>
## 1 Australian Catholic Un~ Melbourne ~ 115 Vi~ FITZR~     3065    -37.8      145.
## 2 Australian Catholic Un~ Ballarat (~ 1200 M~ BALLA~     3350    -37.6      144.
## 3 Australian Catholic Un~ Ararat Hos~ Girdle~ ARARAT     3377    -37.3      143.
## 4 Charles Darwin Univers~ CDU Melbou~ 271 Co~ MELBO~     3000    -37.8      145.
## 5 Charles Sturt Universi~ CSU Study ~ 30 Chu~ MELBO~     3000    -37.8      145.
## 6 CQ University           Melbourne   120 Sp~ MELBO~     3000    -37.8      145.
## # i 1 more variable: wkb_geometry <chr>
```

3.2.2 Recorded Offences by Offence Type

```r
# Read recorded offences data
df_offtype <- read.csv("../data/input/VIC_RecordedOffences_OffenceType.csv")

dim(df_offtype)
```

```
## [1] 359132      8
```

```r
head(df_offtype)
```

```
##   Year Local.Government.Area Postcode Suburb.Town.Name
## 1 2023                Alpine     3691         Dederang
## 2 2023                Alpine     3691         Dederang
## 3 2023                Alpine     3691         Dederang
```

```
## 4 2023                  Alpine     3691         Dederang
## 5 2023                  Alpine     3691         Dederang
## 6 2023                  Alpine     3691       Glen Creek
##                         Offence.Division                   Offence.Subdivision
## 1           A Crimes against the person    A20 Assault and related offences
## 2           A Crimes against the person     Other crimes against the person
## 3     B Property and deception offences                          B40 Theft
## 4     B Property and deception offences                          B40 Theft
## 5 D Public order and security offences D10 Weapons and explosives offences
## 6     B Property and deception offences                      B50 Deception
##                    Offence.Subgroup Offence.Count
## 1      A212 Non-FV Serious assault             1
## 2 Other crimes against the person             1
## 3  B42 Steal from a motor vehicle             2
## 4                   B49 Other theft             1
## 5             D11 Firearms offences             1
## 6 B53 Obtain benefit by deception             1
```

3.3 Load Spatial Datasets

We load the datasets as simple feature (sf) dataframes and plot to visualise the correctness of the data.

3.3.1 Victorian Local Government Area (LGA) Boundaries

```r
# Read Victoria LGA boundaries
vic_lga <- st_read("../data/input/VIC_lga_boundary/vic_lga.shp")
```

```
## Reading layer 'vic_lga' from data source
##   'C:\Data\Repos\Online\VIC-university-crime-analysis\data\input\VIC_lga_boundary\vic_lga.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 92 features and 6 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 140.9619 ymin: -39.13657 xmax: 149.9763 ymax: -33.98127
## Geodetic CRS:  GDA2020
```

```r
dim(vic_lga)
```

```
## [1] 92  7
```

```r
head(vic_lga)
```

```
## Simple feature collection with 6 features and 6 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 142.3535 ymin: -38.56752 xmax: 147.3923 ymax: -36.39268
## Geodetic CRS:  GDA2020
##         LG_PLY_PID         LGA_PID  DT_CREATE          LGA_NAME   ABB_NAME STATE
## 1 lgpKRtJNYEope2S lga136d886cbd2c 2023-08-11      Alpine Shire     Alpine   VIC
```

```
## 2 lgp74f72df1ae34 lga48c76dfa462e 2021-08-10 Ararat Rural City     Ararat   VIC
## 3 lgpbXCewirplDR0 lgab70a9914e5bc 2023-08-11     Ballarat City   Ballarat   VIC
## 4 lgpMOUMGpTebStd lga5591321694d6 2023-08-11      Banyule City    Banyule   VIC
## 5 lgp3b89ab2a2346 lgaac2e88625ea2 2022-08-11  Bass Coast Shire Bass Coast   VIC
## 6 lgpJaZIDU4dadFA lgaac2e88625ea2 2023-08-11  Bass Coast Shire Bass Coast   VIC
##                         geometry
## 1 POLYGON ((147.3209 -37.0231...
## 2 POLYGON ((142.4191 -37.4734...
## 3 POLYGON ((143.8432 -37.6702...
## 4 POLYGON ((145.028 -37.7641,...
## 5 POLYGON ((145.3461 -38.5085...
## 6 POLYGON ((145.3587 -38.5175...
```

```r
# Figure 1(a) Victoria LGA Boundaries
ggplot(data = vic_lga) +
  geom_sf() +
  ggtitle("Victoria LGA Boundaries") +
  xlab("Longitude") +
  ylab("Latitude") +
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5,
                              margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    # Adjust x-axis and y-axis spacing
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
  )
```

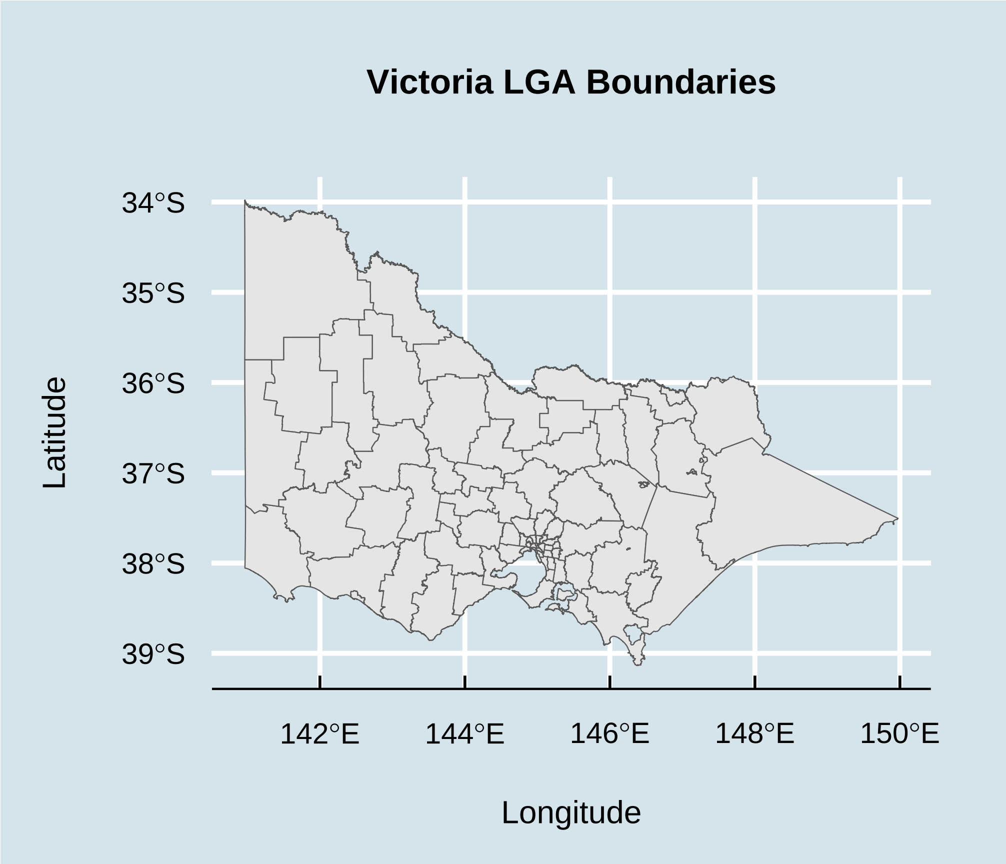

Figure 1(a). Geospatial data of Victoria Local Government Area (LGA) Boundaries

3.3.2 Victorian Suburbs Boundaries

```
vic_sub <- st_read(
  "../data/input/VIC_suburb_boundary/VIC_LOCALITY_POLYGON_shp.shp"
)
```

```
## Reading layer ‘VIC_LOCALITY_POLYGON_shp’ from data source
##   ‘C:\Data\Repos\Online\VIC-university-crime-analysis\data\input\VIC_suburb_boundary\VIC_LOCALITY_POI
##   using driver ‘ESRI Shapefile’
## Simple feature collection with 2973 features and 12 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 140.9619 ymin: -39.13658 xmax: 149.9763 ymax: -33.98128
## Geodetic CRS:  GDA94
```

```
dim(vic_sub)
```

```
## [1] 2973   13
```

```
head(vic_sub)
```

```
## Simple feature collection with 6 features and 12 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 141.0672 ymin: -37.97886 xmax: 148.6839 ymax: -34.99838
## Geodetic CRS:  GDA94
##   LC_PLY_PID  DT_CREATE DT_RETIRE LOC_PID VIC_LOCALI VIC_LOCA_1 VIC_LOCA_2
## 1       6670 2011-08-31      <NA> VIC2615 2012-04-27       <NA>  UNDERBOOL
## 2       6671 2011-08-31      <NA> VIC1986 2012-04-27       <NA>     NURRAN
## 3       6672 2011-08-31      <NA> VIC2862 2012-04-27       <NA>   WOORNDOO
## 4       6673 2011-08-31      <NA>  VIC734 2017-08-09       <NA>   DEPTFORD
## 5       6674 2011-08-31      <NA> VIC2900 2012-04-27       <NA>      YANAC
## 6       6405 2011-08-31      <NA> VIC1688 2012-04-27       <NA>    MINIMAY
##   VIC_LOCA_3 VIC_LOCA_4 VIC_LOCA_5 VIC_LOCA_6 VIC_LOCA_7
## 1       <NA>       <NA>          G       <NA>          2
## 2       <NA>       <NA>          G       <NA>          2
## 3       <NA>       <NA>          G       <NA>          2
## 4       <NA>       <NA>          G       <NA>          2
## 5       <NA>       <NA>          G       <NA>          2
## 6       <NA>       <NA>          G       <NA>          2
##                         geometry
## 1 POLYGON ((141.7455 -35.0722...
## 2 POLYGON ((148.6688 -37.3957...
## 3 POLYGON ((142.9229 -37.9788...
## 4 POLYGON ((147.8234 -37.6600...
## 5 POLYGON ((141.2798 -35.9985...
## 6 POLYGON ((141.3307 -36.6411...
```

```r
# Figure 1(b) Victoria Suburb Boundaries
ggplot(data = vic_sub) +
  geom_sf() +
  ggtitle("Victoria Suburb Boundaries") +
  xlab("Longitude") +
  ylab("Latitude") +
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5,
                              margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    # Adjust x-axis and y-axis spacing
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
  )
```

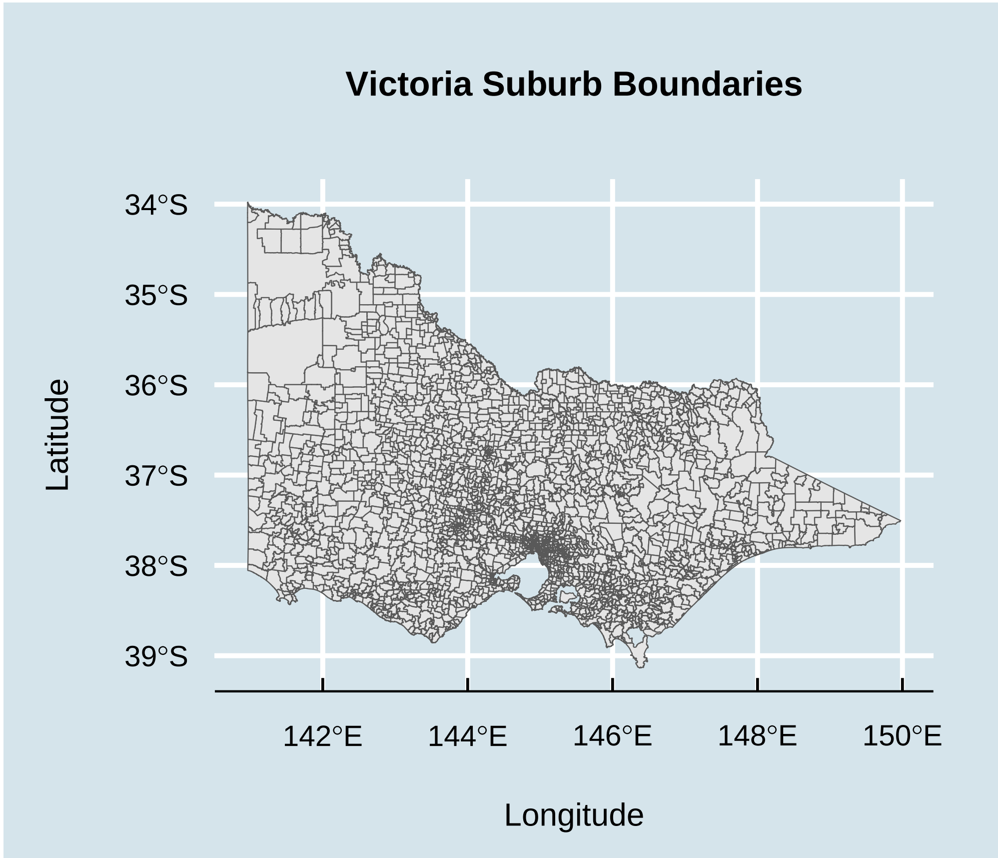

Figure 1(b). Geospatial data of Victoria Suburb Boundaries.

For subsequent analyses, the university locations dataset was also transformed into a spatial object, therefore the coordinate reference system (CRS) for all spatial data was transformed to GDA2020 or EPSG:7844 to ensure that the data aligns correctly when displayed on a map/plot.

4 Data Analysis

Rigorous data checking was an integral part of this project during Analysis. The following steps were performed to ensure further cleanliness and completeness of the data.

1. Column Renaming Columns like 'Suburb/Town Name', where the data may introduce ambiguity, were renamed to 'Suburb' to maintain consistency across the datasets.

2. Data Filtering Data filtering was performed extensively to explore the data and later make visualisations according to our specific research questions. For example, the recorded offences' dataset was filtered to only include offence related information for suburbs that are in proximity to university campuses.

3. Case Handling The case of suburb names was standardised to title case to ensure that string matching operations would not fail or display incorrect statistics due to case sensitivity.

4. Data Grouping and Summarizations The data was grouped and summarised multiple times, based on specific requirements to answer our research questions. For example, the data was grouped by offence types and summarised by the total of offence counts for each type.

5. Data Transformations The structure of the data was changed to facilitate the creation of data visualisations such as grouped bar charts and heatmaps. Transformations were made to the data to convert

it into long format, as to accommodate for offence groupings when creating the subsequent visualisations. Such transformations were also done in the data checking process to ensure one dataset joins successfully with another dataset.

Furthermore, the data attributes were made consistent across our datasets. No missing values or duplicates were found during the data checking phase, and all identified errors were corrected.

The data exploration part of this project dives deep into the patterns, trends and any findings uncovered from exploring and analysing our datasets. Using various plotting libraries such as ggplot2, leaflet and treemapify, we address and visualise each of our research questions in a focused manner. The data exploration part of our project is segmented into three main categories (see the headings below) where each segment aims to uncover nuanced facets of criminal activity in and around suburbs with university campuses in Victoria, Australia

4.1 Prevalent Crime Trends near University Campuses

Focusing on our first research question, we aim to analyse and explore the prevalent crime trends in and around areas surrounding university campuses across Victoria. The data pertaining to recorded criminal offences in Victoria ranged from the year 2014 to 2023. At the initial stage of our exploration, we account for offence counts in all years. Additionally, the data was filtered to include offence information for only suburbs that have universities. This helps us focus and explore the prevalent crime trends in university suburbs only. Data grouping and summarizations were performed to visualise the total count of offences by offence division, subdivision and subgroups. Our approach incorporated the use of visual analytics, generating horizontal bar charts to visualise the total offence counts by offence division and subdivision, as seen in Figure 2. and Figure 3. below. And, a tree map was used to visualise and understand the distribution and frequency of different subgroups of criminal offences, as seen in Figure 4 below.

```r
# Capitalise the first letter of each suburb in df_uni
df_uni$suburb <- str_to_title(df_uni$suburb)

head(df_uni, 3)
```

```
## # A tibble: 3 x 8
##   institution             campus_name address suburb postcode latitude longitude
##   <chr>                   <chr>       <chr>   <chr>     <dbl>    <dbl>     <dbl>
## 1 Australian Catholic Un~ Melbourne ~ 115 Vi~ Fitzr~     3065    -37.8      145.
## 2 Australian Catholic Un~ Ballarat (~ 1200 M~ Balla~     3350    -37.6      144.
## 3 Australian Catholic Un~ Ararat Hos~ Girdle~ Ararat     3377    -37.3      143.
## # i 1 more variable: wkb_geometry <chr>
```

```r
# Rename 'Suburb/Town Name' to Suburb
df_offtype <- df_offtype %>%
  rename(Suburb = `Suburb.Town.Name`)

head(df_offtype, 3)
```

```
##   Year Local.Government.Area Postcode   Suburb
## 1 2023                Alpine     3691 Dederang
## 2 2023                Alpine     3691 Dederang
## 3 2023                Alpine     3691 Dederang
##                   Offence.Division             Offence.Subdivision
## 1       A Crimes against the person A20 Assault and related offences
```

```
## 2       A Crimes against the person  Other crimes against the person
## 3 B Property and deception offences                        B40 Theft
##                   Offence.Subgroup Offence.Count
## 1     A212 Non-FV Serious assault             1
## 2 Other crimes against the person             1
## 3  B42 Steal from a motor vehicle             2
```

```r
# Unique suburb names in df_uni
length(unique(df_uni$suburb))
```

```
## [1] 41
```

```r
# Unique suburb names in df_offtype
length(unique(df_offtype$Suburb))
```

```
## [1] 2850
```

```r
# Filter the offence data to only include university suburbs
filtered_offtype <- df_offtype %>%
  filter(Suburb %in% df_uni$suburb)
length(unique(filtered_offtype$Suburb))
```

```
## [1] 39
```

```r
# Group and summarize
crime_trends_uni <- filtered_offtype %>%
  group_by(Offence.Division) %>%
  summarise(Total_Count = sum(Offence.Count))
```

```r
crime_trends_uni
```

```
## # A tibble: 6 x 2
##   Offence.Division                     Total_Count
##   <chr>                                      <int>
## 1 A Crimes against the person               145182
## 2 B Property and deception offences         529841
## 3 C Drug offences                            65416
## 4 D Public order and security offences       90442
## 5 E Justice procedures offences             155584
## 6 F Other offences                           14820
```

```r
# Rename Offence Divisions for better readability
crime_trends_uni <- crime_trends_uni %>%
  mutate(Offence.Division = case_when(
    Offence.Division == "A Crimes against the person" ~ "Crimes Against Person",
    Offence.Division == "B Property and deception offences" ~ "Property & Deception",
    Offence.Division == "C Drug offences" ~ "Drug Offences",
    Offence.Division == "D Public order and security offences" ~ "Public Order & Security",
    Offence.Division == "E Justice procedures offences" ~ "Justice Procedures",
    Offence.Division == "F Other offences" ~ "Other Offences",
```

```r
    TRUE ~ Offence.Division
  )) %>%
  arrange(desc(Total_Count))

crime_trends_uni
```

```
## # A tibble: 6 x 2
##   Offence.Division        Total_Count
##   <chr>                         <int>
## 1 Property & Deception         529841
## 2 Justice Procedures           155584
## 3 Crimes Against Person        145182
## 4 Public Order & Security       90442
## 5 Drug Offences                 65416
## 6 Other Offences                14820
```

```r
# Figure 2. Bar Chart showing the prevalent types of crimes near university campuses
ggplot(
  data = crime_trends_uni,
  aes(x = reorder(Offence.Division, Total_Count), y = Total_Count)
) +

  # Bar Columns
  geom_col(aes(fill = Total_Count), width = 0.8) +

  # Text Labels for Bar Columns
  geom_text(aes(label = scales::comma(Total_Count)),
    hjust = -0.1,
    position = position_dodge(width = 0.9)
  ) +

  # Title and Axis Labels
  ggtitle("Prevalent Types of Crimes in Suburbs with University Campuses") +
  xlab("Offence Division") +
  ylab("Offence Count") +

  # Theme and Formatting
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5,
                              margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    # Space between x-axis label and plot
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    # Space between y-axis label and plot
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    legend.position = "right",
    legend.key = element_blank(),
    legend.title = element_text(size = rel(1)),
    legend.text = element_text(size = rel(1))
  ) +
```

```
# Y-axis Configuration
scale_y_continuous(
  labels = scales::comma,
  expand = expansion(c(0, 0.15))
) +

# Colour Gradient and Legend Configuration
scale_fill_gradient(
  name = "Offence Count",
  low = "lightblue",
  high = "darkblue",
  trans = "identity",
  labels = scales::comma
) +

# Flip coordinates
coord_flip(clip = "off", ylim = c(0, max(crime_trends_uni$Total_Count) * 1.2))
```

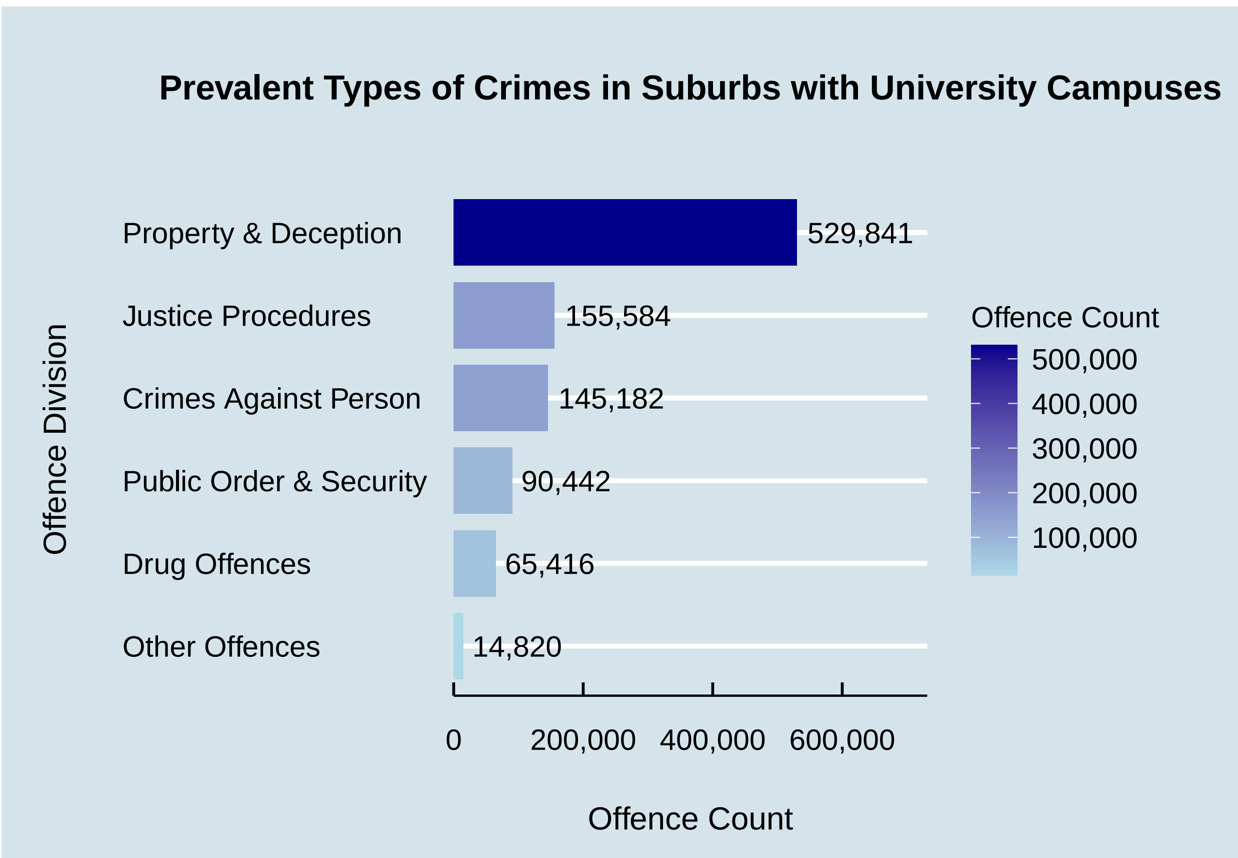

Figure 2. A horizontal bar chart representing the prevalent types of crimes (grouped by offence division) in suburbs with university campuses. The sequential colour palette and legend show the count of offences for each offence division.

From the bar chart visualisation in Figure 2, we identified that the most prevalent type (offence division) of criminal offences, in Suburbs with university campuses, are related to property and deception. Once we identified the high-frequency offence divisions, we can now break down the analysis to take a closer look at the subdivisions within the 'Property & Deception' offence division. To do so, we visualised the total count of offences by offence subdivision for this particular offence division in a horizontal bar chart, as can be seen in Figure 3. below.

```
# Dimensions of our offence data
dim(filtered_offtype)
```

```
## [1] 23284     8
```

```
# Filter the data to only include "B Property and deception offences" Offences
property_deception_filtered <- filtered_offtype %>%
  filter(Offence.Division == "B Property and deception offences")

# Dimensions of our filtered offence data
dim(property_deception_filtered)
```

```
## [1] 7371    8
```

```
# Group by Offence Subdivision and summarize
property_deception_subdivision <- property_deception_filtered %>%
  group_by(Offence.Subdivision) %>%
  summarise(Subdivision_Total_Count = sum(Offence.Count))

property_deception_subdivision
```

```
## # A tibble: 6 x 2
##   Offence.Subdivision          Subdivision_Total_Count
##   <chr>                                          <int>
## 1 B10 Arson                                       3651
## 2 B20 Property damage                            73409
## 3 B30 Burglary/Break and enter                   63938
## 4 B40 Theft                                     301938
## 5 B50 Deception                                  86886
## 6 B60 Bribery                                       19
```

```
# Rename Offence subdivisions for better readability
property_deception_subdivision <- property_deception_subdivision %>%
  mutate(Offence.Subdivision = case_when(
    Offence.Subdivision == "B10 Arson" ~ "Arson",
    Offence.Subdivision == "B20 Property damage" ~ "Property Damage",
    Offence.Subdivision == "B30 Burglary/Break and enter" ~ "Burglary",
    Offence.Subdivision == "B40 Theft" ~ "Theft",
    Offence.Subdivision == "B50 Deception" ~ "Deception",
    Offence.Subdivision == "B60 Bribery" ~ "Bribery",
    TRUE ~ Offence.Subdivision # keep original value if above conds. are not met
  )) %>%
  arrange(desc(Subdivision_Total_Count))

property_deception_subdivision
```

```
## # A tibble: 6 x 2
##   Offence.Subdivision Subdivision_Total_Count
```

```
##   <chr>                                     <int>
## 1 Theft                                    301938
## 2 Deception                                 86886
## 3 Property Damage                           73409
## 4 Burglary                                  63938
## 5 Arson                                      3651
## 6 Bribery                                      19
```

```
# Figure 3. Bar Chart breaking down "Property & Deception" Offences
ggplot(
  data = property_deception_subdivision,
  aes(
    x = reorder(
      Offence.Subdivision,
      Subdivision_Total_Count
    ),
    y = Subdivision_Total_Count
  )
) +

  # Bar Columns
  geom_col(aes(fill = Subdivision_Total_Count), width = 0.8) +

  # Text Labels for Bar Columns
  geom_text(aes(label = scales::comma(Subdivision_Total_Count)),
    hjust = -0.1,
    position = position_dodge(width = 0.9)
  ) +

  # Title and Axis Labels
  ggtitle("'Property & Deception' Offences in Suburbs with University Campuses") +
  xlab("Offence Subdivision") +
  ylab("Offence Count") +

  # Theme and Formatting
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5,
                              margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    # Space between x-axis label and plot
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    # Space between y-axis label and plot
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    legend.position = "right",
    legend.key = element_blank(),
    legend.title = element_text(size = rel(1)),
    legend.text = element_text(size = rel(1))
  ) +

  # Y-axis Configuration
  scale_y_continuous(
    labels = scales::comma,
```

```r
  expand = expansion(c(0, 0.15))
) +

# Colour Gradient and Legend Configuration
scale_fill_gradient(
  name = "Offence Count",
  low = "lightblue",
  high = "darkblue",
  trans = "identity",
  labels = scales::comma
) +

# Flip coordinates with additional space
coord_flip(clip = "off",
           ylim = c(0,max(property_deception_subdivision$Subdivision_Total_Count) * 1.2))
```

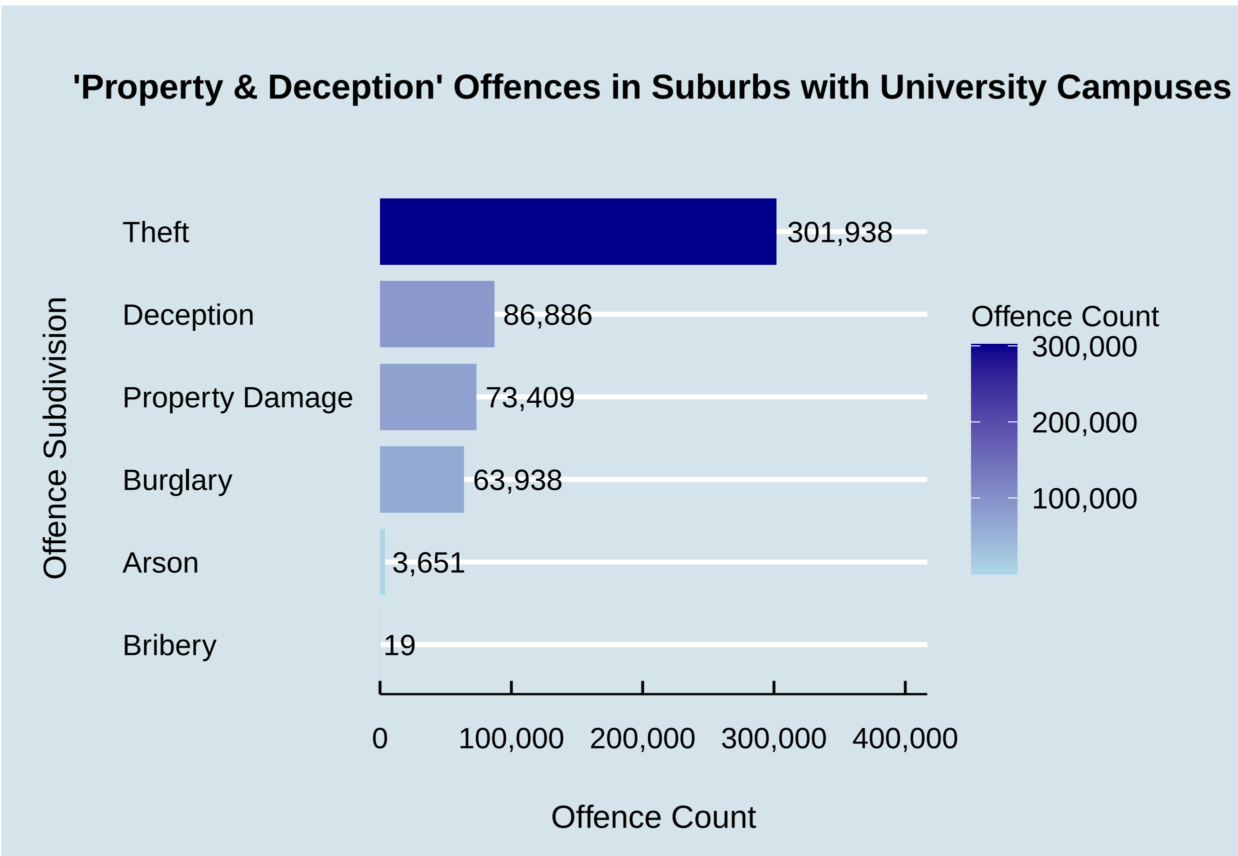

Figure 3. A horizontal bar chart, breaking down the most prevalent criminal offence 'Property & Deception' into offence subdivision. The sequential colour palette and legend show the frequency of offence counts for each offence subdivision.

The breakdown of property and deception offences, as displayed in Figure 3. reveals that 'Theft' is by far the highest criminal offence in university suburbs, with a total count of 301,938. Now, we can further analyse the data to identify which subgroups of the 'Theft' offence subdivision are most prevalent. We effectively visualise this information using a tree map to represent the distribution of theft related offences, as can be seen in Figure 4. below.

```r
# Further filter the data to include only Theft subgroup
theft_subdivision <- property_deception_filtered %>%
  filter(Offence.Subdivision == "B40 Theft")

# Group by Offence subgroup and summarise
theft_subgroup_summary <- theft_subdivision %>%
  group_by(Offence.Subgroup) %>%
  summarise(Subgroup_Total_Count = sum(Offence.Count)) %>%
  arrange(desc(Subgroup_Total_Count))
```

```r
# Rename Offence subgroup for better readability
theft_subgroup_summary <- theft_subgroup_summary %>%
  mutate(Offence.Subgroup = case_when(
    Offence.Subgroup == "B49 Other theft" ~ "Other Theft",
    Offence.Subgroup == "B42 Steal from a motor vehicle" ~ "Steal from Motor Vehicle",
    Offence.Subgroup == "B43 Steal from a retail store" ~ "Steal from Retail Store",
    Offence.Subgroup == "B45 Receiving or handling stolen goods" ~ "Handling Stolen Goods",
    Offence.Subgroup == "B41 Motor vehicle theft" ~ "Motor Vehicle Theft",
    Offence.Subgroup == "B44 Theft of a bicycle" ~ "Bicycle Theft",
    Offence.Subgroup == "B46 Fare evasion" ~ "Fare Evasion",
    TRUE ~ Offence.Subgroup
  ))

theft_subgroup_summary
```

```
## # A tibble: 7 x 2
##   Offence.Subgroup         Subgroup_Total_Count
##   <chr>                                   <int>
## 1 Other Theft                             91111
## 2 Steal from Motor Vehicle                79031
## 3 Steal from Retail Store                 55833
## 4 Handling Stolen Goods                   31965
## 5 Motor Vehicle Theft                     25583
## 6 Bicycle Theft                           18155
## 7 Fare Evasion                              260
```

```r
# Figure 4. Treemap of Theft Offences by Subgroup
gg <- ggplot(theft_subgroup_summary,
             aes(area = Subgroup_Total_Count,
                 fill = Subgroup_Total_Count,
                 label = Offence.Subgroup,
                 subgroup = Offence.Subgroup)) +

  # Add treemap
  geom_treemap() +

  # Add text
  geom_treemap_text(reflow = T,
                    aes(label = paste0(Offence.Subgroup)),
                    color = "white",
                    place = "centre",
                    size = 11) +
```

```r
  # Add gradient color scale and rename legend title
  scale_fill_gradient(name = "Offence Count",
                      low = "lightblue",
                      high = "darkblue") +

  # Add title
  ggtitle("Distribution of Theft Offences by Subgroup") +

  # Customise theme
  theme(
    plot.title = element_text(hjust = 0.5, margin = margin(t = 10, b = 20)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    plot.background = element_rect(fill = "#d6e3eb", color = NA),
    legend.background = element_rect(fill = "#d6e3eb", color = NA),
    panel.background = element_rect(fill = "#d6e3eb", color = NA)
  )
```

Figure 4. Treemap distribution of offences inside the 'Theft' offence subdivision. The size and colour palette represent the count of offences, where the bigger size and darker colour refers to a higher frequency of recorded offences. The legend represents offence counts for offence subgroups related to Theft.

Therefore, by narrowing our focus to suburbs with university campuses, we were able to shed light on the specific types of criminal activities that are most frequent in these areas. The visual analytics, Figures 2 and 3, served as compelling evidence for effectively communicating these trends. They revealed that 'Property & Deception' offences are predominant, with 'Theft' being the most frequent offence subdivision within the category. Furthermore, the tree map visualisation, Figure 4, offered a nuanced understanding of the various subgroups within the 'Theft' category, where we identify that 'Other Theft' and 'Steal from Motor Vehicle' are most prevalent. The Crime Statistics Agency of Victoria, states that 'Other Theft' offence subgroup relates to all other theft that cannot be classified elsewhere in the 'Theft' subdivision. (Crime Statistics Agency, 2023). Therefore, our analysis demonstrates that theft (except motor vehicles) and stealing from motor vehicles are the most prevalent criminal offences in Victorian suburbs with university campuses

4.2 Correlating Offences in University Suburbs with their Neighbouring suburbs (LGAs)

The second research question of our study seeks to explore the specific types of criminal offences that occur more frequently in suburbs with university campuses compared to their neighbouring suburbs. By correlating criminal activities in university suburbs with their neighbouring areas, we can identify any distinct patterns or types of crimes that are more prevalent near educational institutions. To address this question, we identified the local government areas (LGAs) associated with each university campus using spatial joins. Since we are looking for a correlation in offences between university suburbs and their surrounding areas, instead of comparing the offence counts in university suburbs with the broader LGA, we instead opt for a more granular approach. A suburb shares more local characteristics with its immediate neighbouring suburbs than the broader LGAs that may introduce more noise into the data such as diverse types of localities (rural, suburban or urban). Therefore, we identified the immediate neighbouring suburbs that touch the boundary of the university suburbs and calculate the total offence counts for each area type (university suburbs and neighbouring suburbs). To allow for a more meaningful comparison of criminal offences in university suburbs and their neighbouring suburbs, instead of comparing the total offence counts, we instead normalised the data to make it comparable as a fraction of the total for each area. This was done because the number of university suburbs '39' and the number of neighbouring suburbs '215' is different, therefore the offence counts may be more biased towards the area type with the higher quantity of suburbs. We instead calculate the concentration of offences for each area type as:

$$Concentration = \frac{Total\,Offence\,Count}{Total\,Offence\,Count\;for\;Area\,Type}$$

This provides a more meaningful comparison of the types of offences prevalent in university suburbs and their neighbouring suburbs, as can be seen in Figure 5. below.

```r
# Make a spatial object for uni data
df_uni_spatial <- st_as_sf(
  df_uni,
  coords = c("longitude", "latitude"), crs = 4326
)

# Transform CRS of df_uni_spatial to GDA2020 explicitly
df_uni_spatial <- st_transform(x = df_uni_spatial, crs = st_crs(7844))

# Spatial join uni data with LGA boundaries to get LGA for each university
uni_with_lga <- st_join(df_uni_spatial, vic_lga, join = st_within)
```

```r
# Select and rename data
uni_with_lga <- uni_with_lga %>%
  select(-"LG_PLY_PID", -"LGA_PID", -"DT_CREATE", -"STATE", -"LGA_NAME") %>%
  rename(lga_name = ABB_NAME)

uni_with_lga
```

```
## Simple feature collection with 80 features and 7 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: 142.0584 ymin: -38.39142 xmax: 146.8488 ymax: -34.20562
## Geodetic CRS:  GDA2020
## # A tibble: 80 x 8
##    institution                  campus_name address suburb postcode wkb_geometry
##    <chr>                        <chr>       <chr>   <chr>     <dbl> <chr>
##  1 Australian Catholic Univers~ Melbourne ~ 115 Vi~ Fitzr~     3065 MULTIPOINT ~
##  2 Australian Catholic Univers~ Ballarat (~ 1200 M~ Balla~     3350 MULTIPOINT ~
##  3 Australian Catholic Univers~ Ararat Hos~ Girdle~ Ararat     3377 MULTIPOINT ~
##  4 Charles Darwin University    CDU Melbou~ 271 Co~ Melbo~     3000 MULTIPOINT ~
##  5 Charles Sturt University     CSU Study ~ 30 Chu~ Melbo~     3000 MULTIPOINT ~
##  6 CQ University                Melbourne   120 Sp~ Melbo~     3000 MULTIPOINT ~
##  7 Deakin University            Burwood     221 Bu~ Burwo~     3125 MULTIPOINT ~
##  8 Deakin University            Waurn Ponds Pigdon~ Geelo~     3216 MULTIPOINT ~
##  9 Deakin University            Waterfront  1 Gher~ Geelo~     3220 MULTIPOINT ~
## 10 Deakin University            Warrnambool Prince~ Warrn~     3280 MULTIPOINT ~
## # i 70 more rows
## # i 2 more variables: geometry <POINT [°]>, lga_name <chr>
```

```r
# Capitalise suburb names in uni_with_lga
uni_with_lga$suburb <- str_to_upper(uni_with_lga$suburb)
```

```r
# Filter vic_sub to only contain university suburbs
vic_sub_uni <- vic_sub[vic_sub$VIC_LOCA_2 %in% uni_with_lga$suburb, ]
```

```r
# Identify neighbouring suburbs
neighboring_suburbs <- st_touches(vic_sub_uni, vic_sub)
```

```r
# Convert list of indices to list of names
uni_neighbor_names <- lapply(neighboring_suburbs, function(indices) {
  return(vic_sub$VIC_LOCA_2[indices])
})

# Flatten the list and make it unique
uni_neighbor_names <- unique(unlist(uni_neighbor_names))
```

```r
# Capitalise suburb names in uni_with_lga
uni_with_lga$suburb <- str_to_title(uni_with_lga$suburb)

# Summarising for University Suburbs
uni_suburb_summary <- df_offtype %>%
  filter(Suburb %in% uni_with_lga$suburb) %>%
  group_by(`Suburb`,`Offence.Division`) %>%
  summarise(Uni_Suburb_Total_Count = sum(`Offence.Count`))
```

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

```r
# TitleCase suburb names in df_offtype
df_offtype$Suburb <- str_to_upper(df_offtype$Suburb)

# Filter the df_offtype data to include only offenses from neighboring suburbs
df_offtype_neighbors <- df_offtype %>%
  filter(`Suburb` %in% uni_neighbor_names)
```

```r
# Summarize the offense data for neighboring suburbs
neighbor_suburb_summary <- df_offtype_neighbors %>%
  group_by(Suburb, `Offence.Division`) %>%
  summarise(Neighbor_Suburb_Total_Count = sum(`Offence.Count`))
```

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

```r
length(unique(uni_suburb_summary$Suburb))
```

```
## [1] 39
```

```r
length(unique(neighbor_suburb_summary$Suburb))
```

```
## [1] 215
```

```r
# Calculate crime type concentration
aa_uni_suburb_summary <- uni_suburb_summary %>%
  group_by(Offence.Division) %>%
```

```
  summarise(Total_Count = sum(Uni_Suburb_Total_Count)) %>%
  mutate(Concentration = Total_Count / sum(Total_Count))

aa_neighbor_suburb_summary <- neighbor_suburb_summary %>%
  group_by(Offence.Division) %>%
  summarise(Total_Count = sum(Neighbor_Suburb_Total_Count)) %>%
  mutate(Concentration = Total_Count / sum(Total_Count))
```

```
# Combining data into one data frame for plotting
aa_combined_data <- bind_rows(
  aa_uni_suburb_summary %>% mutate(Area_Type = 'University Suburbs'),
  aa_neighbor_suburb_summary %>% mutate(Area_Type = 'Neighbouring Suburbs')
)
```

```
# Rename Offence.Division values for better readability
aa_combined_data <- aa_combined_data %>%
  mutate(Offence.Division = case_when(
    Offence.Division == "A Crimes against the person" ~ "Crimes Against Person",
    Offence.Division == "B Property and deception offences" ~ "Property & Deception",
    Offence.Division == "C Drug offences" ~ "Drug Offences",
    Offence.Division == "D Public order and security offences" ~ "Public Order & Security",
    Offence.Division == "E Justice procedures offences" ~ "Justice Procedures",
    Offence.Division == "F Other offences" ~ "Other Offences",
    TRUE ~ Offence.Division # keeps other values unchanged
  ))

# Display the modified data frame
aa_combined_data
```

```
## # A tibble: 12 x 4
##    Offence.Division        Total_Count Concentration Area_Type
##    <chr>                         <int>         <dbl> <chr>
##  1 Crimes Against Person        145182        0.145  University Suburbs
##  2 Property & Deception         529841        0.529  University Suburbs
##  3 Drug Offences                 65416        0.0653 University Suburbs
##  4 Public Order & Security       90442        0.0903 University Suburbs
##  5 Justice Procedures           155584        0.155  University Suburbs
##  6 Other Offences                14820        0.0148 University Suburbs
##  7 Crimes Against Person        211363        0.134  Neighbouring Suburbs
##  8 Property & Deception         957935        0.609  Neighbouring Suburbs
##  9 Drug Offences                 97267        0.0618 Neighbouring Suburbs
## 10 Public Order & Security      110519        0.0703 Neighbouring Suburbs
## 11 Justice Procedures           173624        0.110  Neighbouring Suburbs
## 12 Other Offences                22299        0.0142 Neighbouring Suburbs
```

```
# Re-level the Area_Type variable to make 'Neighbour Suburbs' appear first
aa_combined_data$Area_Type <- factor(aa_combined_data$Area_Type,
                                     levels = c("University Suburbs",
                                                "Neighbouring Suburbs"))

# Figure 5. Grouped Bar Chart showing the concentration of criminal offences
# in University Suburbs and Neighbouring Suburbs
```

```r
ggplot(aa_combined_data, aes(x = Offence.Division,
                             y = Concentration, fill = Area_Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Concentration of Offence Divisions grouped by Area Type ",
       x = "Offence Division",
       y = "Concentration *") +
  scale_fill_manual(
    values = c("Neighbouring Suburbs" = "#ef7f6f",
               "University Suburbs" = "#3fbbc3"),
    breaks = c("Neighbouring Suburbs", "University Suburbs"),
    name = "Area Type") +
  scale_y_continuous(labels = scales::percent_format(scale = 100)) +
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5, margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    # Space between x-axis label and plot
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    # Space between y-axis label and plot
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    legend.position = "right",
    legend.key = element_blank(),
    legend.title = element_text(size = rel(1)),
    legend.text = element_text(size = rel(1))
  ) +
  coord_flip(clip = "off")
```

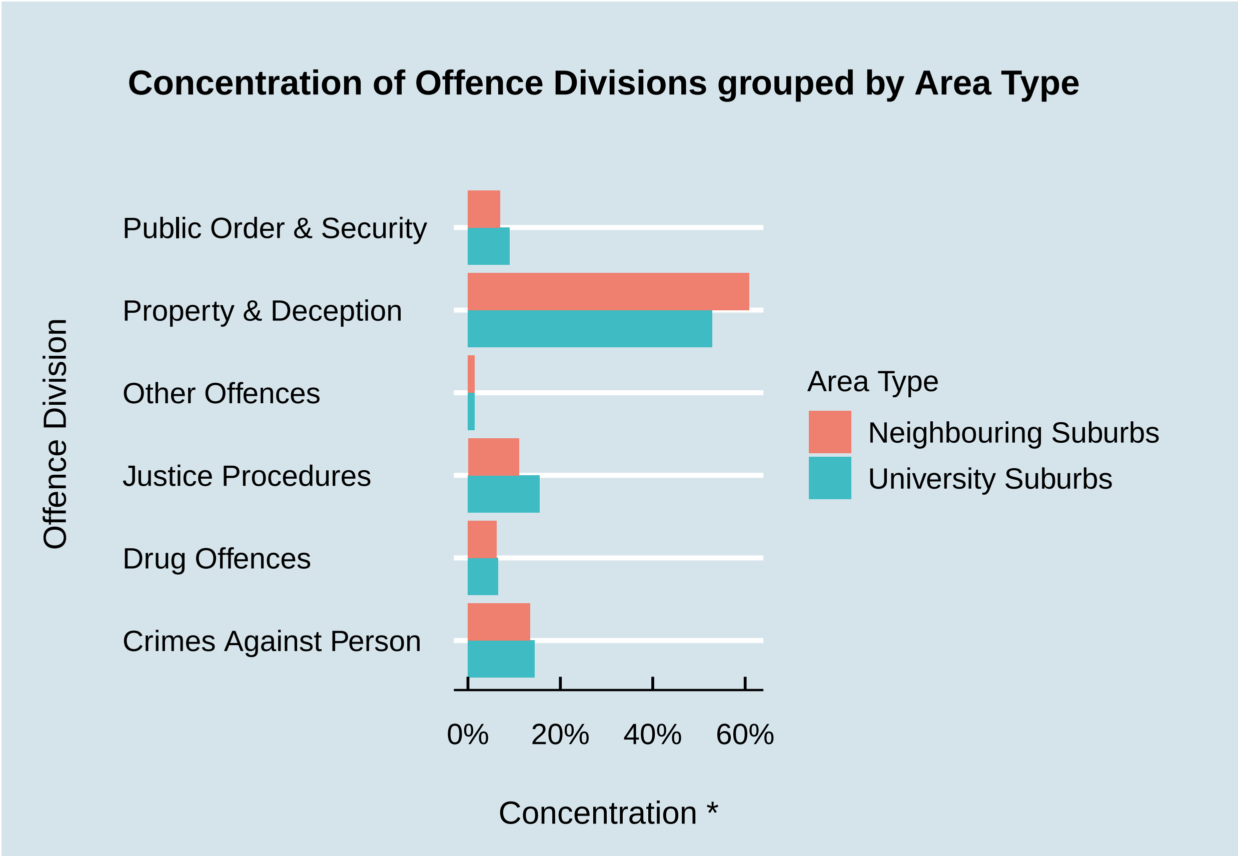

Figure 5. A horizontal grouped bar chart representing the concentration of criminal offences for each offence division (offence type) in university suburbs and their neighbouring suburbs. The asterisk on x-axis Concentration refers to the formula used for calculation, as seen above.

As seen in Figure 5. criminal offences related to crimes against person, drug offences, justice procedures and public order/safety are more prevalent in university suburbs with 'Crimes Against Person' being the most prominent variable. Furthermore, 'Property & Deception' crimes form a larger part of total crimes in 'Neighbouring Suburbs' (60.9%) and in University Suburbs (52.9%). Although there are various other factors, such as population sizes and rates per capita, that may affect our analysis. The key takeaway from our visualisation is that almost all offence types, with the exclusion of Property & Deception, are more prevalent in neighbouring suburbs than suburbs with university campuses. To delve deeper into the distribution and frequency of offences, we utilise heatmap analysis, specifically focusing on property-related crimes (Property & Deception). We selected a subset of our data to focus on one university suburb, Clayton, and its neighbouring suburbs, as seen in heatmap visualisation, Figure 6. below.

```
# Identify the row corresponding to Clayton in your spatial object
clayton_row <- which(vic_sub$VIC_LOCA_2 == "CLAYTON")

# Use st_touches to find the indices of polygons that touch Clayton
clayton_neighbors_idx <- st_touches(vic_sub[clayton_row, ], vic_sub)

# Convert list of indices to actual suburb names
clayton_neighbor_names <- vic_sub$VIC_LOCA_2[unlist(clayton_neighbors_idx)]

# Display the neighboring suburbs
print(clayton_neighbor_names)
```

```
## [1] "CLARINDA"       "NOTTING HILL"   "OAKLEIGH EAST"  "MOUNT WAVERLEY"
## [5] "OAKLEIGH SOUTH" "CLAYTON SOUTH"  "SPRINGVALE"     "MULGRAVE"
```

```r
# Filter df_offtype to only include Clayton and its neighboring suburbs
df_offtype_clayton_neighbors <- df_offtype %>%
  filter(Suburb == "CLAYTON" | Suburb %in% clayton_neighbor_names)
```

```r
# Summarize the offense data
clayton_summary <- df_offtype_clayton_neighbors %>%
  group_by(Suburb, `Offence.Subdivision`) %>%
  summarise(Total_Count = sum(`Offence.Count`))
```

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

```r
# Filter data to only include 'B Property and Deception Offences'
df_offtype_clayton_B <- df_offtype_clayton_neighbors %>%
  filter(`Offence.Division` == "B Property and deception offences")
```

```r
# Summarize the offense data
clayton_B_summary <- df_offtype_clayton_B %>%
  group_by(Suburb, `Offence.Subdivision`) %>%
  summarise(Total_Count = sum(`Offence.Count`))
```

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

```r
# Rename Offence subdivisions for better readability
clayton_B_summary <- clayton_B_summary %>%
  mutate(Offence.Subdivision = case_when(
    Offence.Subdivision == "B10 Arson" ~ "Arson",
    Offence.Subdivision == "B20 Property damage" ~ "Property Damage",
    Offence.Subdivision == "B30 Burglary/Break and enter" ~ "Burglary",
    Offence.Subdivision == "B40 Theft" ~ "Theft",
    Offence.Subdivision == "B50 Deception" ~ "Deception",
    Offence.Subdivision == "B60 Bribery" ~ "Bribery",
    TRUE ~ Offence.Subdivision # keep original value if above conds. are not met
  ))
```

```r
# Remove rows where Offence.Subdivision is "Bribery" (Mistake in Data)
clayton_B_summary <- clayton_B_summary %>%
 filter(Offence.Subdivision != "Bribery")
```

```r
# Calculate the sum of Total_Count for normalization
total_sum <- sum(clayton_B_summary$Total_Count)

# Add a new column for percentages
clayton_B_summary <- clayton_B_summary %>%
  mutate(Percentage = (Total_Count / total_sum) * 100)
```

```r
# Figure 6. Heatmap of Property-Related Offences in Clayton and Neighbouring Suburbs
ggplot(clayton_B_summary, aes(x = Suburb,
                              y = `Offence.Subdivision`, fill = Total_Count)) +

  # Tile plot
  geom_tile() +

  # Text annotations for percentages, centered and white
  geom_text(aes(label = sprintf("%.2f%%", Percentage)),
            vjust = 0.5, hjust = 0.5, col = "white") +

  # Title and Axis Labels
  ggtitle("Heatmap of Property Related Offences in Clayton and Neighbouring Suburbs") +
  xlab("Suburb") +
  ylab("Offence Subdivision") +

  # Colour Gradient Configuration for Total Count
  scale_fill_gradient(
    low = "#58a6ff",
    high = "#f85149",
    name = "Total Count of\nOffences"
  ) +

  # Theme and Formatting
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.6, margin = margin(t = 10, b = 30),
                              size = rel(1.1)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    # Space between x-axis label and plot
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1.1)),
    # Space between y-axis label and plot
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1.1)),
    # Rotate x-axis labels and adjust alignment
    axis.text.x = element_text(angle = 35, hjust = 1, vjust = 1, size = rel(1)),
    legend.position = "right",
    legend.key = element_blank(),
    legend.title = element_text(size = rel(1)),
    legend.text = element_text(size = rel(1))
  ) +

  # Flip coordinates if necessary
  coord_flip(clip = "off")
```

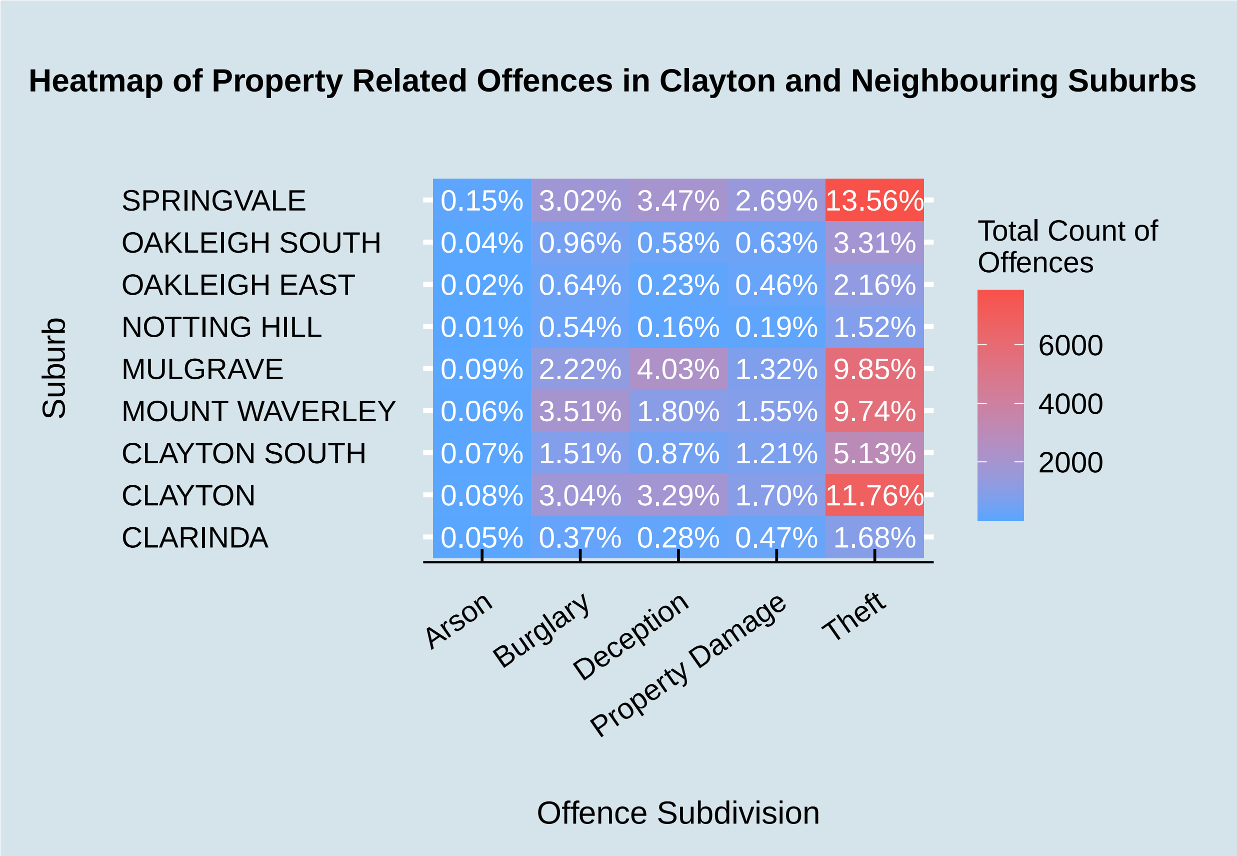

Figure 6. A heatmap visualisation of property related offences in Clayton and its neighbouring suburbs. The colour palette shows the frequency of offences, and the percentage inside each box represents the proportion of a specific offence subdivision within a particular suburb.

The heatmap analysis revealed that the suburb of Springvale recorded the highest count of offences for theft, followed closely by Clayton. This granular analysis allows us to identify hotspots for specific types of criminal activities in suburbs surrounding universities. While the bar chart in Figure 5. gives us a broad understanding of the types of crimes that are more prevalent in each area type, the heatmap goes a step further by pinpointing the exact suburbs where these crimes are most concentrated (in a subset of data).

4.3 Exploring Patterns in Criminal Offences

As we have previously identified, that property related offences (Property & Deception) are most prevalent in and around areas with university campuses, we now aim to identify any patterns in criminal offences between university suburbs, neighbouring suburbs and the broader LGAs. We visualise the suburb boundaries and university locations on an interactive map of Victoria, using leaflet, as seen in Figure 7. below. This geographical representation allows for an immediate visual interpretation of crime prevalence in different regions of Victoria.

```
property_offences <- df_offtype %>%
  filter(`Offence.Division` == "B Property and deception offences") %>%
  group_by(Suburb) %>%
  summarise(Total_Count = sum(`Offence.Count`))

# Removing Melbourne
property_offences <- subset(property_offences, Suburb != "MELBOURNE")
```

```r
vic_sub_data <- left_join(vic_sub, property_offences,
                          by = c("VIC_LOCA_2" = "Suburb"))
```

```r
# Standardise datasets to the same CRS
vic_sub_data <- st_transform(vic_sub_data, crs = 4326)
df_uni_spatial <- st_transform(df_uni_spatial, crs = 4326)
```

```r
# Create a vector of colors from blue to red
color_vector <- colorRampPalette(RColorBrewer::brewer.pal(11, "RdGy"))

# Create a reversed vector of colors
reversed_colors <- rev(color_vector(11))

# Create a reversed vector of colors
final_map <- leaflet(vic_sub_data) %>%
  addTiles() %>%

  # Add polygons with color gradient
  addPolygons(
    fillColor = ~colorQuantile(reversed_colors, Total_Count)(Total_Count),
    color = "grey",
    weight = 1,
    fillOpacity = 0.7
  ) %>%

  # Add markers for universities
  addMarkers(
    data = df_uni_spatial,
    label = ~institution
  ) %>%

  # Add legend
  addLegend(
    pal = colorQuantile(reversed_colors, NULL, n = 5),
    values = ~Total_Count,
    opacity = 0.7,
    title = "Frequency of Property-Related Offences",
    position = "bottomright"
  )

# Save interactive map as HTML file
# saveWidget(final_map, file = "../images/fig7.html", selfcontained = FALSE)
```

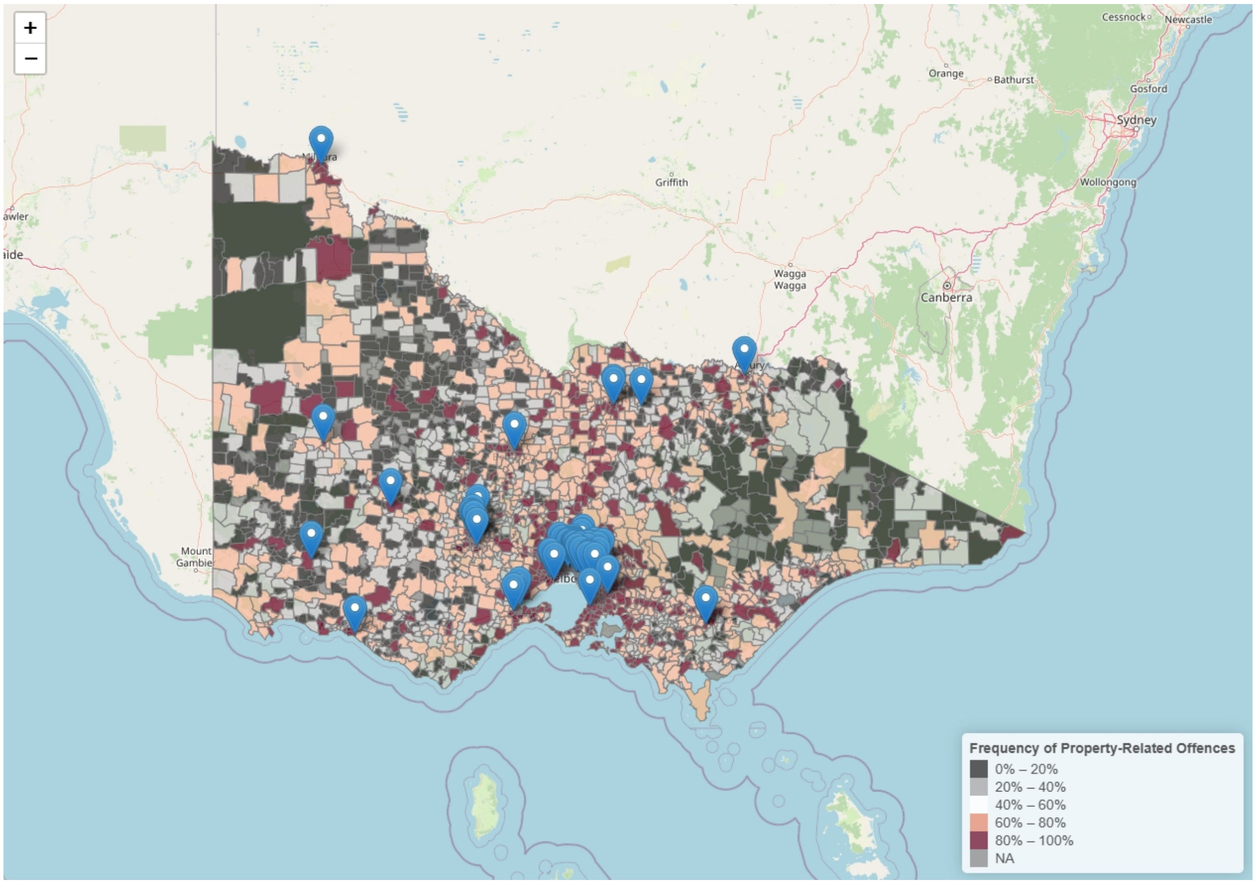

Figure 7. Representation of an interactive leaflet map of property-related offences in Victoria. The blue markers represent the locations of universities, and the colour gradient indicates the frequency of offences in each suburb. The colour intensity increases with frequency of offences.

The leaflet map, Figure 7, shows that suburbs containing university campuses have a significantly higher frequency of property-related offences, ranging between 80-90%. In contrast, the neighbouring suburbs show a frequency ranging between 60-80%. This clear pattern underscores the need for targeted interventions in areas with educational institutions.

```
# Group and summarize the data, including Offence.Division
time_trend_data <- df_offtype_clayton_neighbors %>%
  group_by(Year, Suburb, `Offence.Division`) %>%
  summarise(Total_Count = sum(`Offence.Count`))
```

```
## ‘summarise()‘ has grouped output by ’Year’, ’Suburb’. You can override using
## the ‘.groups‘ argument.
```

```
# Generate the facet grid of line charts
ggplot(time_trend_data, aes(x = Year, y = Total_Count, color = Suburb)) +
  geom_line(size = 0.3) +
  geom_point(size = 1) +

  # Title and Axis Labels
  ggtitle("Offence Trends Over Time: Clayton vs Neighboring Suburbs") +
  xlab("Year") +
  ylab("Total Offence Count") +
```

```r
  # Facet by Offence Division
  facet_wrap(~`Offence.Division`) +

  # Theme and Formatting
  theme_economist() +
  theme(
    plot.title = element_text(hjust = 0.5,
                              margin = margin(t = 10, b = 30),
                              size = rel(1.2)),
    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm"),
    # Space between x-axis label and plot
    axis.title.x = element_text(margin = margin(t = 20), size = rel(1)),
    # Space between y-axis label and plot
    axis.title.y = element_text(margin = margin(r = 20), size = rel(1)),
    legend.position = "right",
    legend.key = element_blank(),
    legend.title = element_text(size = rel(1)),
    legend.text = element_text(size = rel(0.7)),
    # Facet label and panel spacing
    strip.text = element_text(size = rel(0.5)),
    # x and y axis text size
    axis.text = element_text(size = rel(0.5)),
    panel.spacing.y = unit(2, "lines"),
    panel.spacing.x = unit(1, "lines")
  )
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

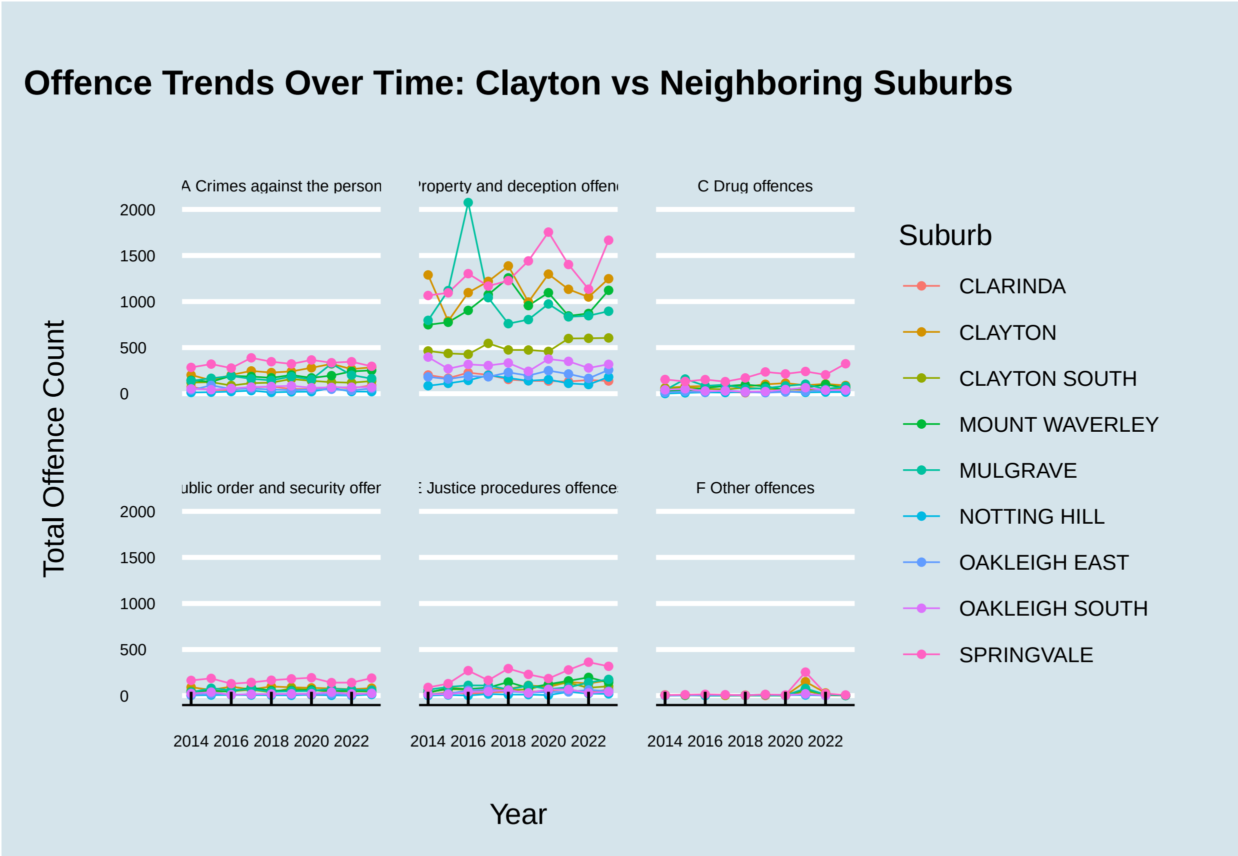

Figure 8. A faceted time-series analysis of criminal offence trends in Clayton and neighbouring suburbs. The different coloured, line styles represent each suburb's offence trends over time.

The time series analysis in Figure 8. above provides insights into the evolution of criminal behaviour in Clayton and its neighbouring suburbs. It revealed that although Springvale consistently has the highest overall criminal activity in all offence divisions, a group of suburbs including Springvale, Clayton, Mount Waverley, and Mulgrave shows notably high offence counts specifically for property and deception offences. This trend over time emphasises the persistent nature of certain crimes in these areas. Additionally, we also identified that in this subset of data, the suburb Mulgrave had the highest count of offences related to property and deception in the year 2016. After further research, it became evident that in 2016 more than 20,000 vehicles were stolen and railway stations in Mulgrave were a common spot. (Stephen Johnson, 2018) Hence, our time series analysis also offers a historical context, showcasing the trends in criminal activities and how they fluctuated over time.

5 Conclusion

The primary goal of this study was to explore the criminal landscape surrounding university campuses in Victoria, Australia. By focusing on three main research questions, our analyses revealed that property-related offences, particularly 'Theft', are the most common types of crimes in university suburbs. This aligns with the personal experience that motivated this study, suggesting that theft of personal items on university campuses is a concern worth addressing. Furthermore, the data showed a higher concentration of crimes against persons, drug offences, and public order/safety crimes in university suburbs compared to neighbouring ones. This finding provides valuable insights for law enforcement agencies and university security services to focus their preventative measures on specific types of crimes near educational institutions. Our spatial analyses highlighted that property related crimes are particularly prevalent in university suburbs, with frequencies between 80-90%. Neighbouring suburbs displayed lower frequencies, between 60-80% of offences. The time-series analyses indicated that university suburbs have persistent high counts of property related criminal offences, adding a temporal dimension to our spatial findings. While our study offers valuable insights, it also has limitations. For instance, our data does not account for population sizes or rates per capita, which could offer additional insights into crime rates. Future research could incorporate these variables and possibly extend the study to other states and educational institutions for a more comprehensive view. Overall, this study serves as a data-driven foundation for policymaking and community engagement. Law enforcement agencies and university security services can use these findings to better allocate resources and implement targeted interventions.

6 Bibliography

Crime Statistics Agency. (2023, July 21). Offence classification. Crime Statistics Agency Victoria.
https://www.crimestatistics.vic.gov.au/about-the-data/classifications-and-victorian-map-boundaries/offence-classification

Safer Community Unit. Current students. (2023, September 12).
https://www.monash.edu/students/support/safety-security/safer-community-unit

Stephen Johnson. (2018, January 7). More than 20,000 vehicles stolen in Victoria last year. Daily Mail Online.
https://www.dailymail.co.uk/news/article-5243417/More-20-000-vehicles-stolen-Victoria-year.html