

# STAT7123/STAT8123

## Statistical Graphics Assignment 2

Umut Demirhan

Due 11:55 pm, Friday September 29th, 2023

### Question 1

- a) The total number of offences by local government area (LGA).

```
## # A tibble: 6 x 2
##   Offence_LGA      total_offences
##   <chr>                  <int>
## 1 Bayside                   14
## 2 Bega Valley                 2
## 3 Blacktown                  107
## 4 Blue Mountains                19
## 5 Burwood                    127
## 6 Byron                      6
```

Given the summary provided, one challenging feature regarding plotting this data is the significant disparity in the total number of offences between different Local Government Areas (LGA). Some LGAs have over a hundred offences, while many others have only one, leading to a high variance in the data which can make visualization unclear or misleading if not handled properly.

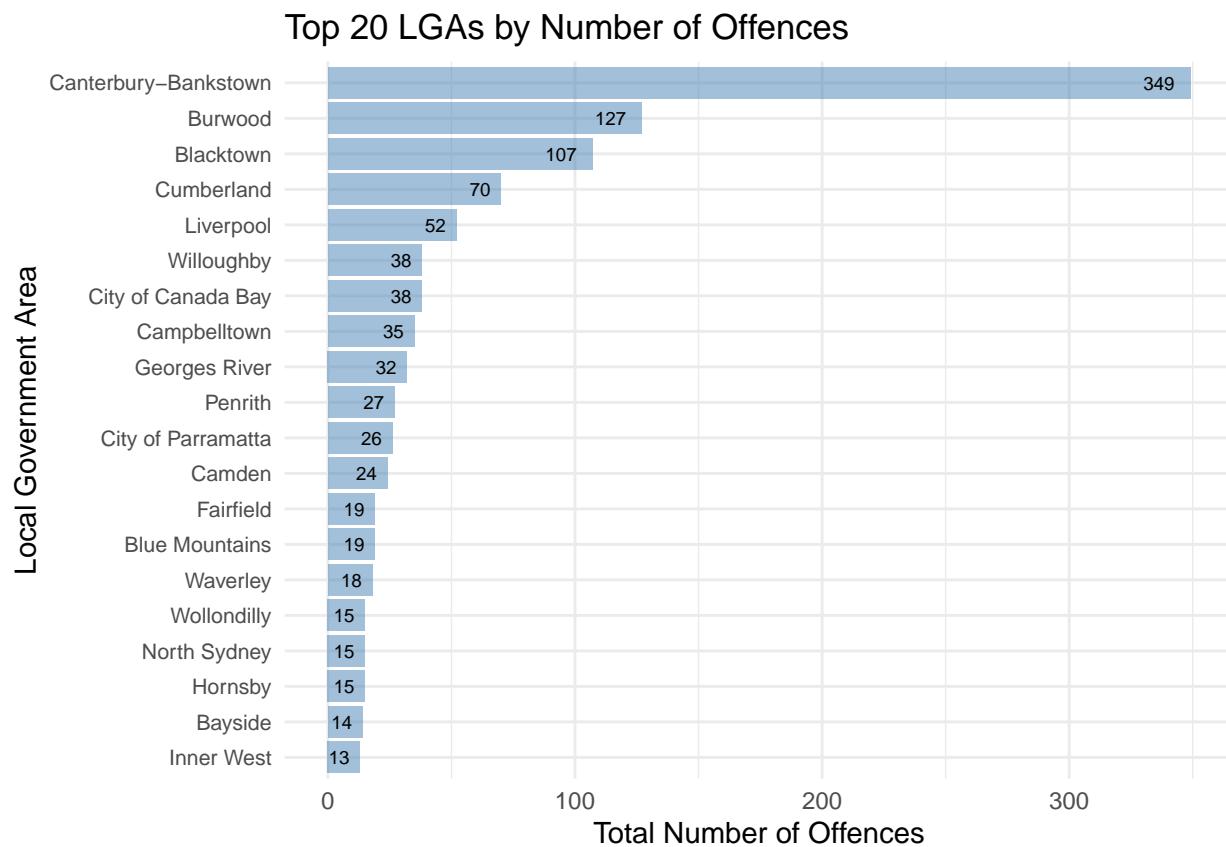
Secondly, there are a considerable number of LGAs listed, which could result in a cluttered and unreadable plot, especially if every LGA is labeled, hence requiring careful consideration of labeling, positioning, and choice of plot type.

Lastly, since this is count data, choosing the right type of plot is crucial. If a choropleth map is not the choice, bar plot might be the most straightforward choice, but with the disparity in the number of offences, using a logarithmic scale might be necessary to visualize low and high count LGAs together effectively.

- b) Which LGA has the most offences?

The LGA with the most offences is stored in the variable top\_LGA. According to the sorted data, the LGA with the most offences is Canterbury-Bankstown, having 349 total offences.

c) A bar chart to display a subset of the LGAs

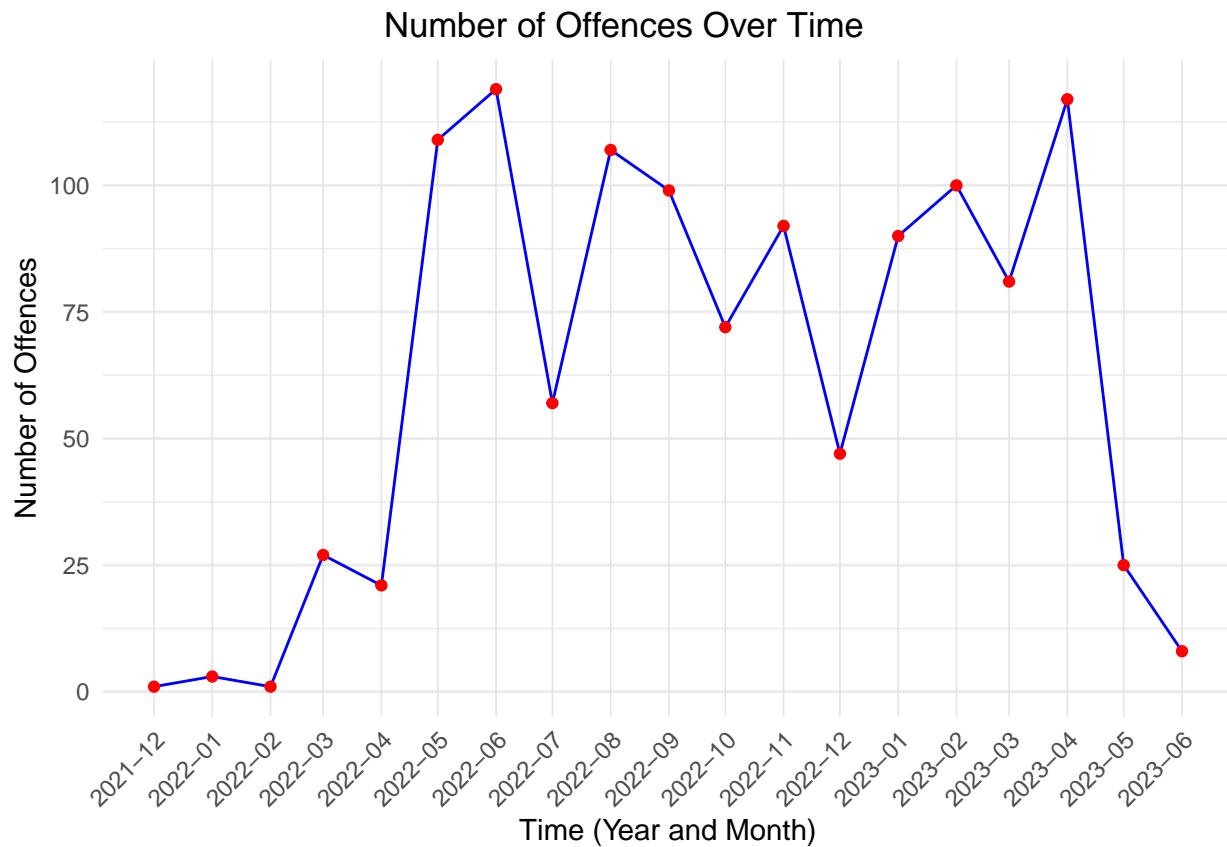


The LGA with the most offences is Canterbury-Bankstown, with a total of 349 offences. This is significantly higher than the second and third-ranking LGAs, Burwood and Blacktown, which have 127 and 107 offences respectively. The chart also shows that the number of offences varies considerably among these top 20 LGAs, ranging from a high of 349 in Canterbury-Bankstown to a low of 13 in Inner West.

The horizontal bars are color-coded in steel blue and are ordered in descending fashion, making it easy to identify LGAs with more frequent offences. Text labels indicate the exact number of offences for each LGA, providing a quick numerical reference for the viewer. Overall, the chart provides a succinct yet comprehensive overview of food safety offences in the top 20 LGAs, highlighting areas that may require more stringent regulatory attention.

## Question 2

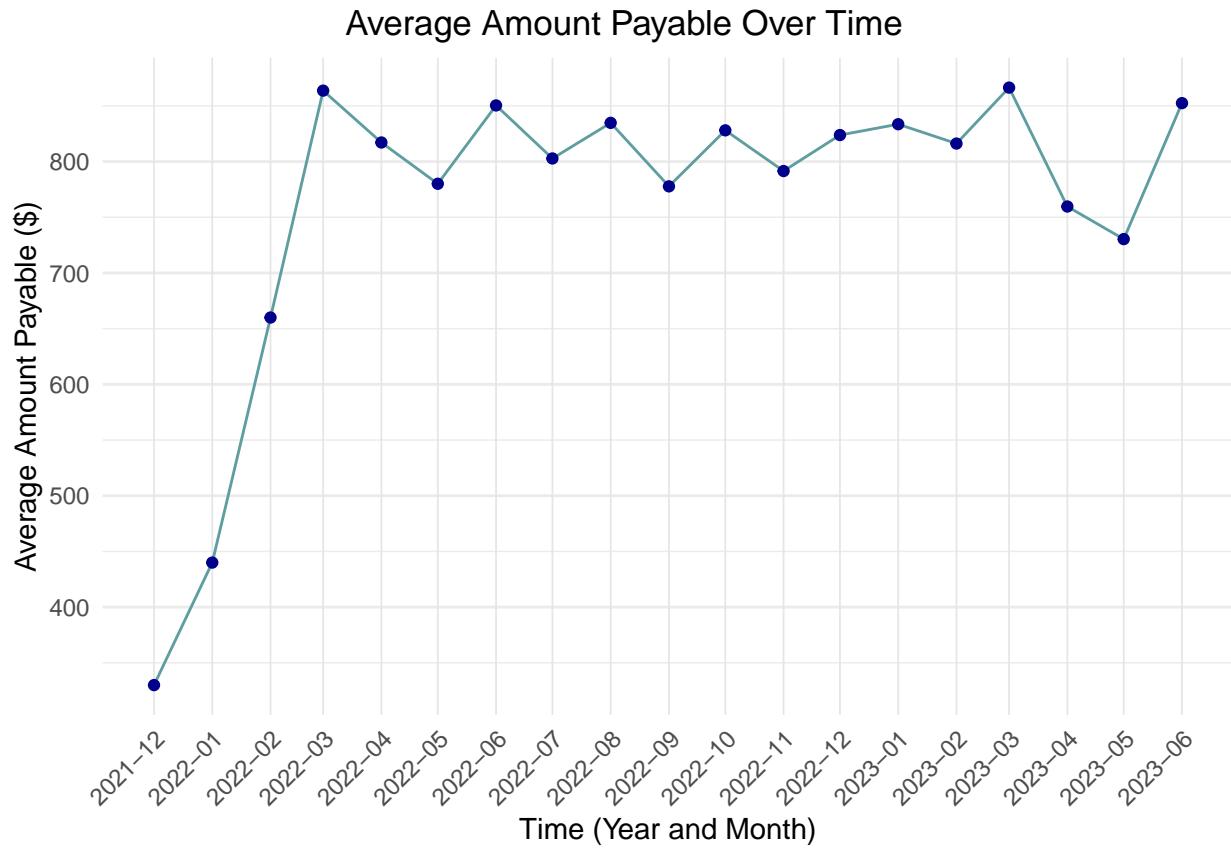
- a) A plot for number of offences by time



- b) Interpretation of the plot

The line chart reveals a dramatic fluctuation in the number of offences from December 2021 to June 2023. Notably, there are peaks observed around May and June of 2022 and troughs in early 2022 and June 2023. While there is no pattern for seasonality, the drastic increase in May 2022 and decline in offences in recent months suggests that other factors may also be at play such as lifted quarantine in February 2022 in Sydney. Overall, the chart illustrates that the number of offences has not remained consistent over time and warrants further investigation to understand these variations.

- c) A plot for average amount payable by time

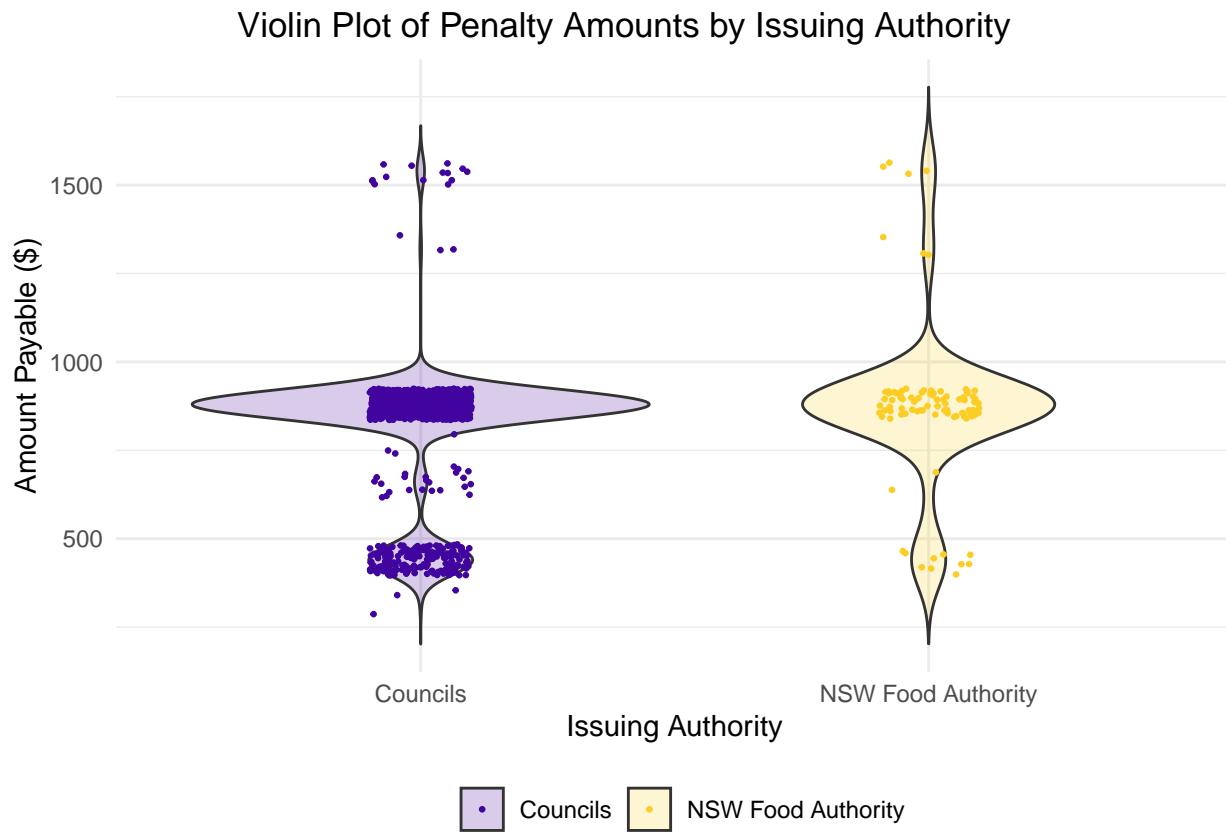


d) Interpretation of the plot

The plot of average amounts payable over time predominantly exhibits no discernible seasonal or linear trends, with the exception of a short period from December 2021 to March 2022, which may reflect increased monitoring coinciding with a spike in offences. The data points fluctuate up and down, but these movements are somewhat sporadic and do not follow a consistent pattern across the observed period. While there are some high and low points, they don't occur with the regularity that would suggest seasonality. Therefore, the plot reveals variability but does not provide evidence for a distinct trend or seasonality in the average amounts payable for offences.

### Question 3

- a) A violin plot for penalty amounts by Issuing Authority

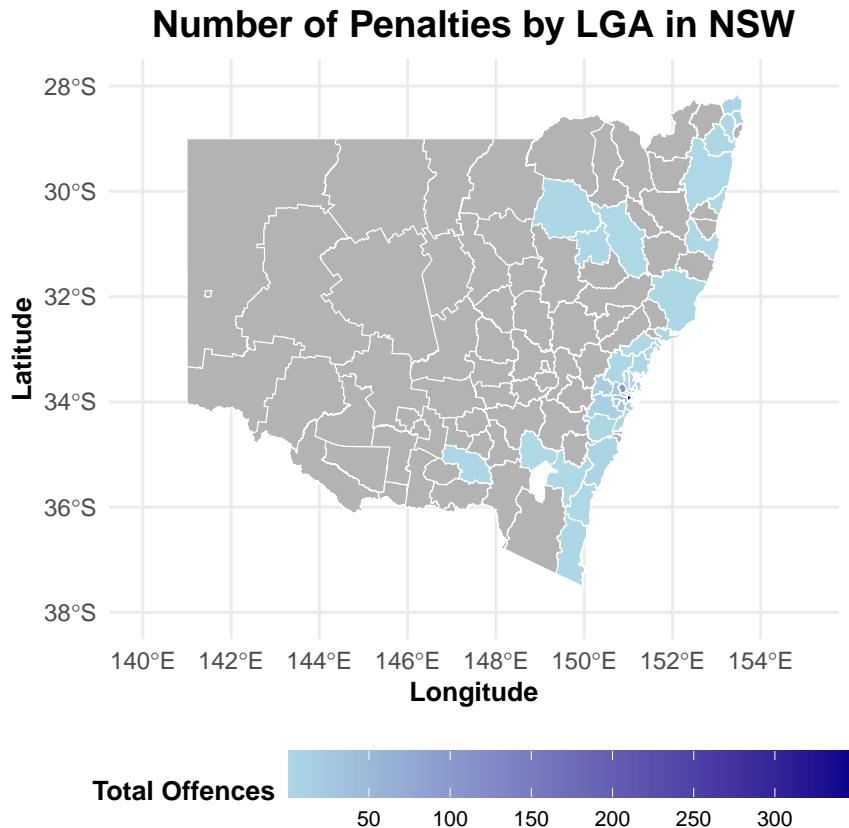


- b) Interpretation of the plot

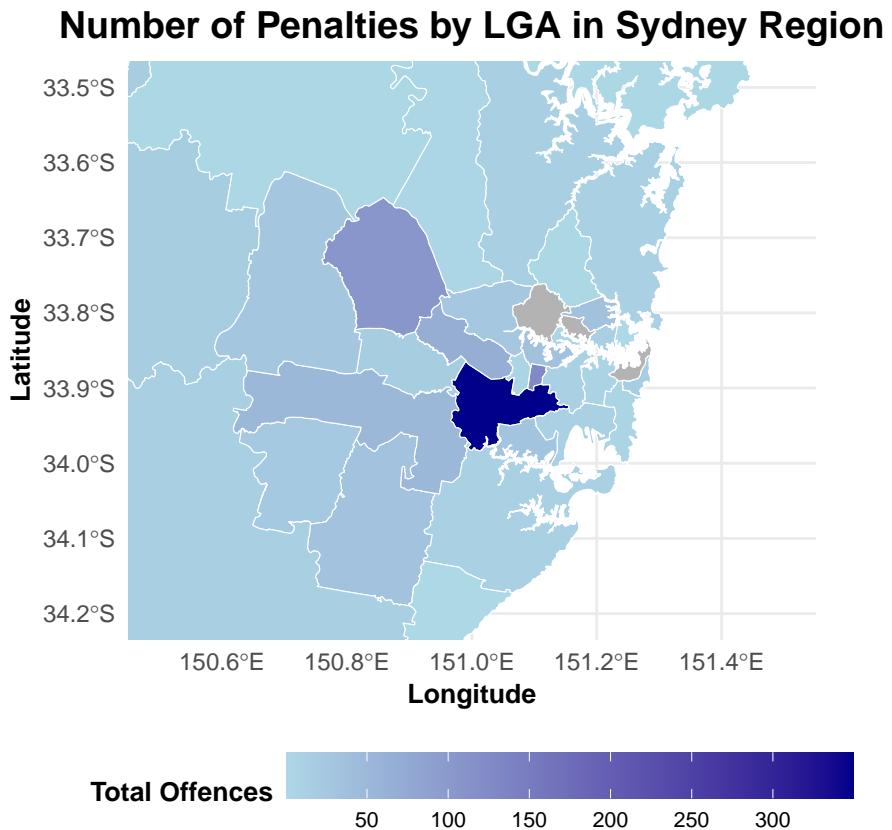
- There are seven distinct values for amount payable, and the violin plot reveals those penalty amounts for both the councils and the NSW Food Authority.
- Councils frequently issue penalties below \$500 (\$440 and \$330) and mostly around \$800 (\$880, \$770, and \$660), whereas the NSW Food Authority focuses on the \$800 range while also being more diverse in penalties.
- The NSW Food Authority's plot is more balanced, indicating a greater diversity in penalty amounts, including those over \$1000.
- The general pattern for both indicates that penalties over \$1000 (\$1540 and \$1320) and \$660 are sparse, penalties below \$500 (\$440 and \$330) are pretty popular, and penalties around \$800 (\$880, \$770) are the most frequent.

## Question 4

- a) A choropleth map that shows the number of penalties by LGA in NSW



b) A choropleth map focusing on Sydney region



c) Why focus on the Sydney region for number of penalty per LGA?

Focusing on the Sydney region allows us to highlight densely populated areas and, therefore, more meaningful variations in penalty numbers, as it contains the areas with the highest number of penalties, whereas other regions in the expansive NSW territory often have so few penalties that they only appear in the lightest colours on a statewide map.

d) Summary of the main results of the choropleth plot presented.

Within the Sydney-centric map, Canterbury-Bankstown emerges prominently due to its darker hue, signifying the highest number of offences among the Local Government Areas (LGAs). Although some other regions like Burwood and Blacktown register elevated offence levels, they are markedly lower than those in Canterbury-Bankstown. Intriguingly, despite hosting the highest number of restaurants, the City of Sydney doesn't reflect a proportional penalty rate, emphasizing the urgent need for further investigations of the reason behind it and/or targeted interventions in areas like Canterbury-Bankstown.

## Question 5

- a) Data Simplification Steps for Nature\_of\_Offence\_Full for Graphical Presentation

### Objective:

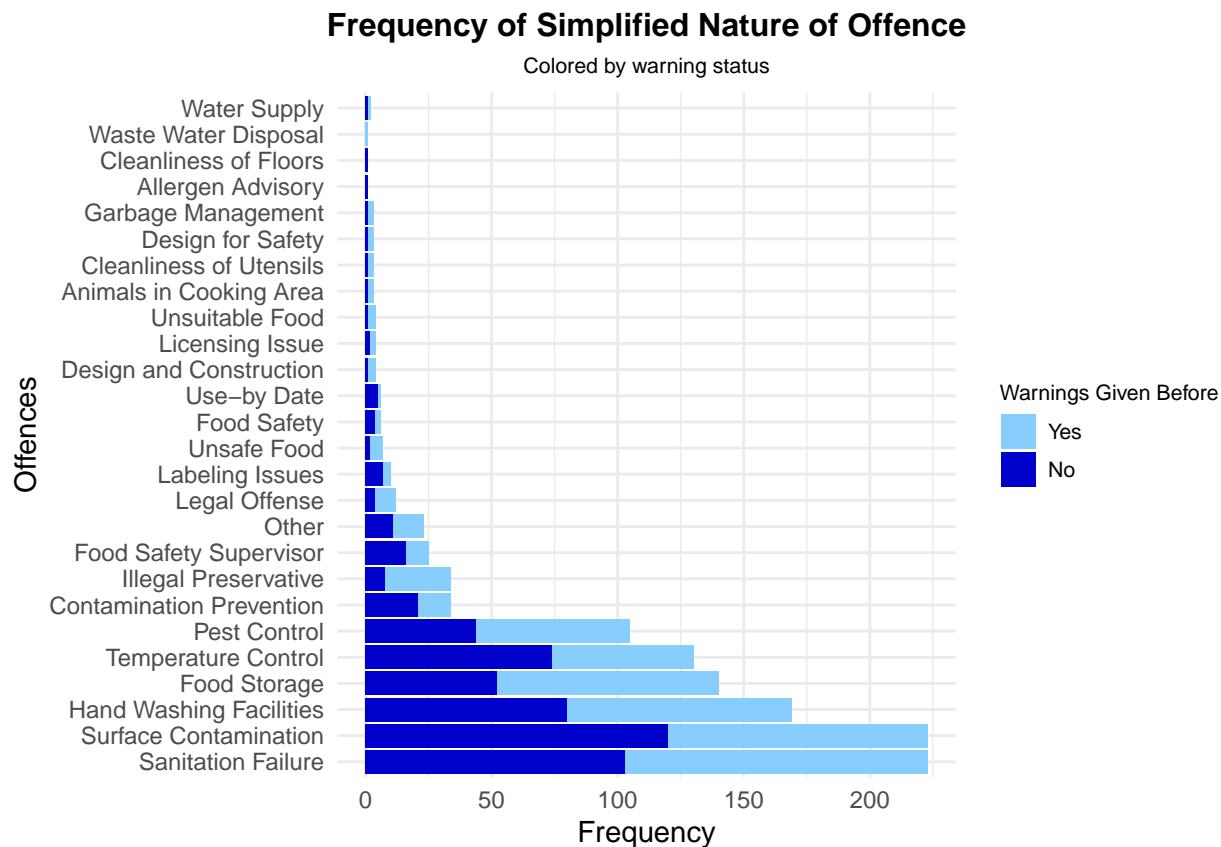
To refine the lengthy descriptions in `Nature_of_Offence_Full` to a simplified, categorized format suitable for graphical representation, without losing the crucial information.

### Proposed Steps:

- **Standardization:**
  - Ensure uniformity in text to prevent redundancy and discrepancies during categorization.
  - Convert the text in `Nature_of_Offence_Full` to lowercase for uniformity.
  - Standardize variants of similar phrases to a common representation, such as different forms of ‘previous warning given’.
- **Categorization and Simplification:**
  - Generalize and simplify the nature of offences to manageable categories for effective graphical representation.
  - Identify key terms or phrases from the detailed descriptions to create simplified categories, e.g., “Illegal Preservative”, “Cleanliness of Utensils”.
  - Assign each entry a simplified category based on the identified key terms or phrases. Classify unmatched entries as “Other”.
- **Creating Indicator Variables:**
  - Extract additional meaningful information from the detailed descriptions.
  - Create an indicator variable to denote whether previous warnings were mentioned, providing insights into the compliance history of the entity involved.
- **Data Aggregation:**
  - Facilitate the creation of graphical representations by summarizing the data.
  - Aggregate the data based on the new simplified categories and indicator variables to find the frequency of each type of offence.

- b) Data Simplification process for Nature\_of\_Offence\_Full for Graphical Presentation and Three Statistical Graphics

Graphic 1: Bar Plot of Top 5 Most Common Offences per Year

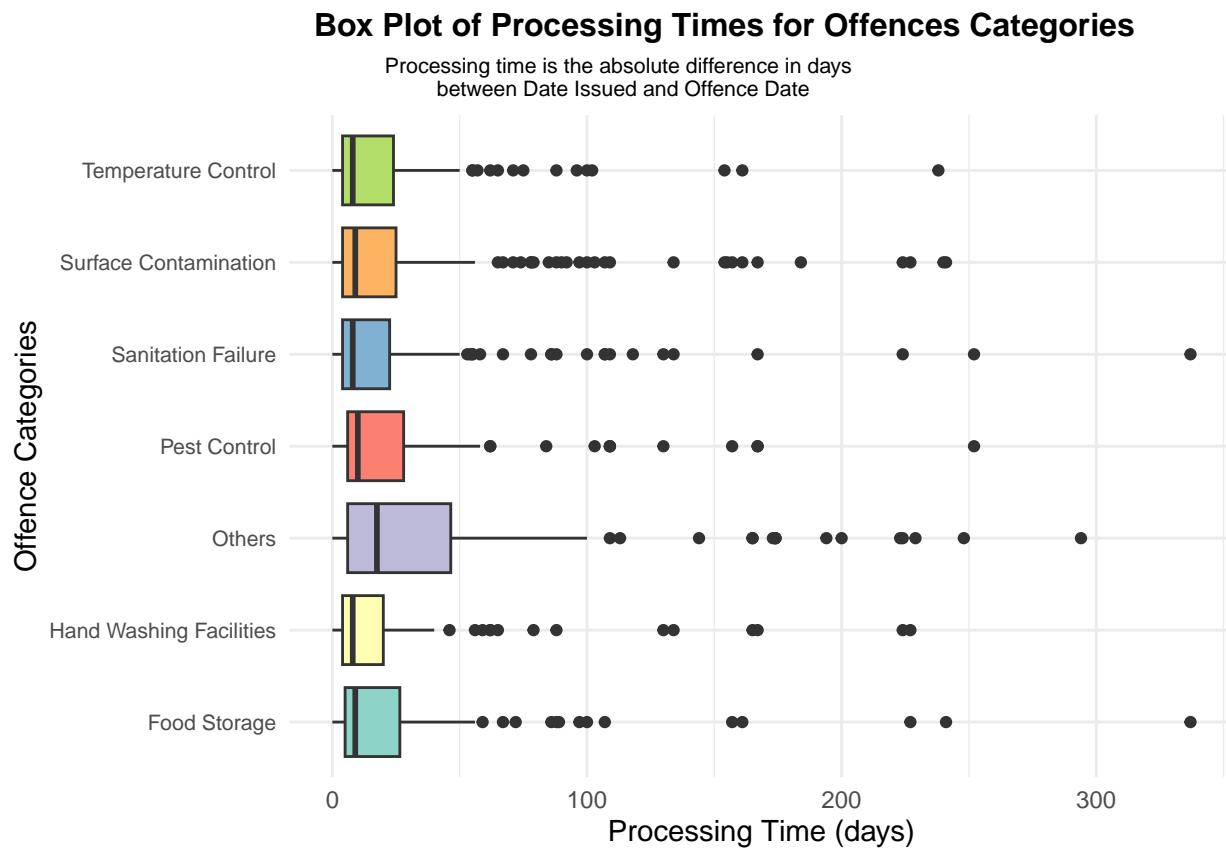


#### Interpretation of the Frequency of Simplified Offences Bar Chart

The bar chart provides a clear visualization of the frequency of various offenses, distinguished by whether warnings were given before for the same issue.

- “Surface Contamination(Food Contact)” and “Sanitation Failure (Food Premises)” are predominant, with an almost equal distribution of warnings and non-warnings.
- “Hand Washing Facilities,” “Food Storage,”, “Pest Control” and “Temperature Control” occur significantly, with variances in warning distributions.
- Issues with “Illegal Preservative” seems to reoccur as many penalties are given more than once.
- The bar chart indicates that “Labelling Issues” and “Use-by Date” cases tend to not repeat.

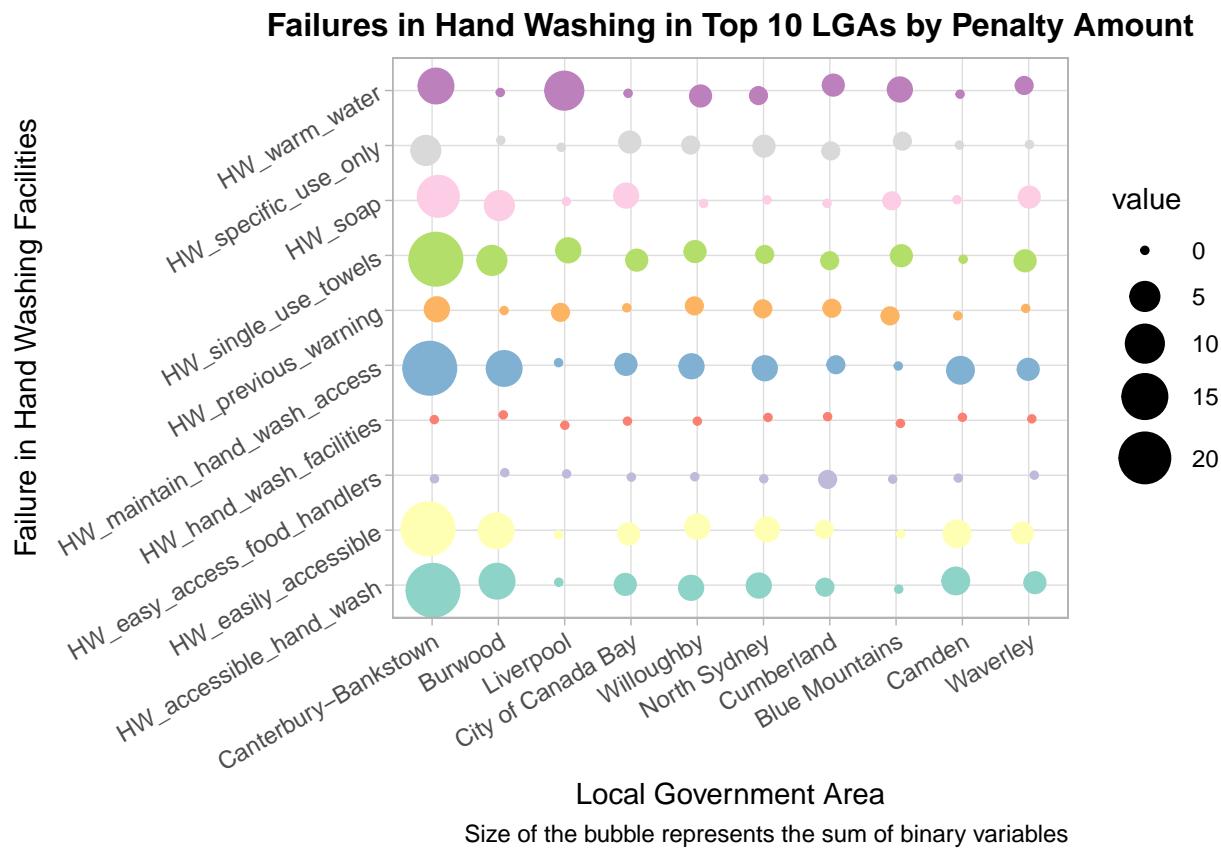
Graphic 2: Stacked Area Plot of Processing Time for Different Types of Offences Over Time



#### Interpretation of Processing Times across Offence Categories

- Processing time is the number of days between the date of offence and the date penalty notice issued.
- Most of the time, for all types of offences, processing times are shorter rather than longer.
- However, in every offence category, there are several cases that take much longer time to process, these are the “outliers.”
- Thus, the all the processing time are right-skewed.
- Other combined categories has a wide range of processing times and a lot of variation. This could mean some specific cases take longer to process than others on the plot.

Graphic 3: Failures in Hand Washing in Top 10 LGAs by Penalty Amount



#### Interpretation:

- **Canterbury-Bankstown:** This LGA likely stands out with several bubbles of various sizes, depicting numerous instances of non-compliance across different aspects of hand washing facilities.
- **Burwood and Liverpool:** these LGAs would also showcase several bubbles. Restaurants in Liverpool may not have hot water in their hand washing facilities.
- **Camden, Blue Mountains and Waverley:** These LGAs would probably have the smallest and least amount of bubbles, reflecting the lowest total penalties due to fewer non-compliance occurrences in terms of hand washing.
- **Variety in Non-Compliance:** Each LGA might have varying levels and types of non-compliance, with some being more prone to specific hand washing facility failures than others like Liverpool stands out with hot water.
- **Severity and Frequency of Non-Compliance:** LGAs like Canterbury-Bankstown, having more and larger bubbles, depict more frequent and severe non-compliance with hand washing facilities.

#### Important Note:

The code chunks are hidden as advised. The codes can be found in the R Markdown file attached.

#### AI Use Acknowledgement

I have used ChatGPT to revise my writing and debugging codes. It helped make the report more professional and concise.