

Visualising Food Safety Breaches in New South Wales

Question 1

Use tidyverse commands to calculate the total number of offences by local government area (LGA)

a) Using 3-5 sentences, describe what features of this summary make the data challenging to plot

```
## # A tibble: 54 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
## 7 Camden                     24
## 8 Campbelltown                 35
## 9 Canterbury-Bankstown        349
## 10 Central Coast                  3
## # i 44 more rows
```

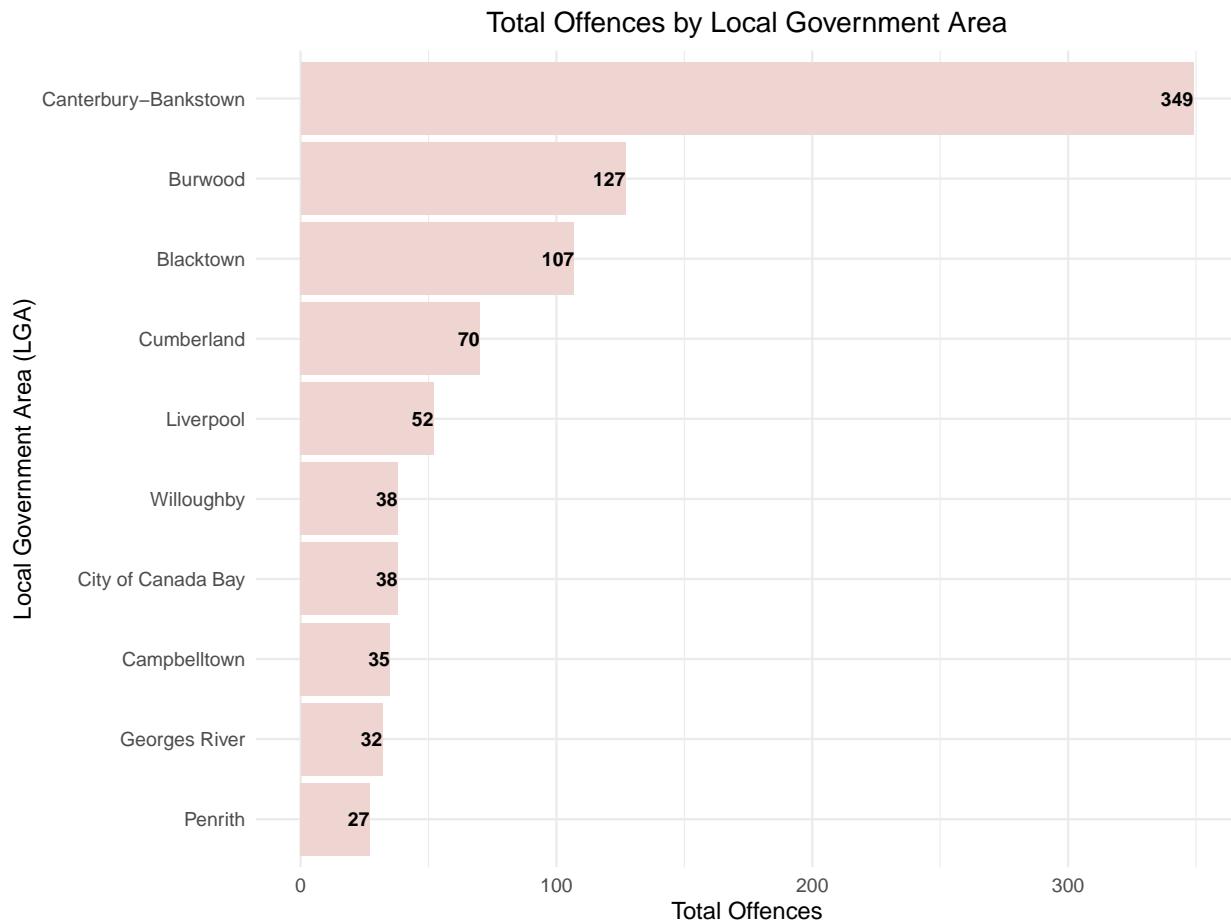
Visualizing data containing the number of offences across 54 Local Government Areas (LGAs) poses a number of challenges. Scale issues can also be caused by a wide range of violation counts ranging from a minimum of 1 to a maximum of 349 as LGAs with lesser violation counts are overshadowed in the plots. As for LGAs with low offence amounts, overplotting might obfuscate the data. Moreover, it is difficult to create geographic or spatial representations when there is no geographical data available.

b) Which LGA has the most offences?

```
## # A tibble: 1 x 2
##   Offence_LGA      Total_Offences
##   <chr>                <int>
## 1 Canterbury-Bankstown        349
```

Canterbury-Banktown suburb has the highest number of offences with 349 offences.

- c) Use an appropriate plot to display a subset of the LGAs (do not use a map here) that could be used to answer the research question: “Which LGAs have the highest number of offences?”

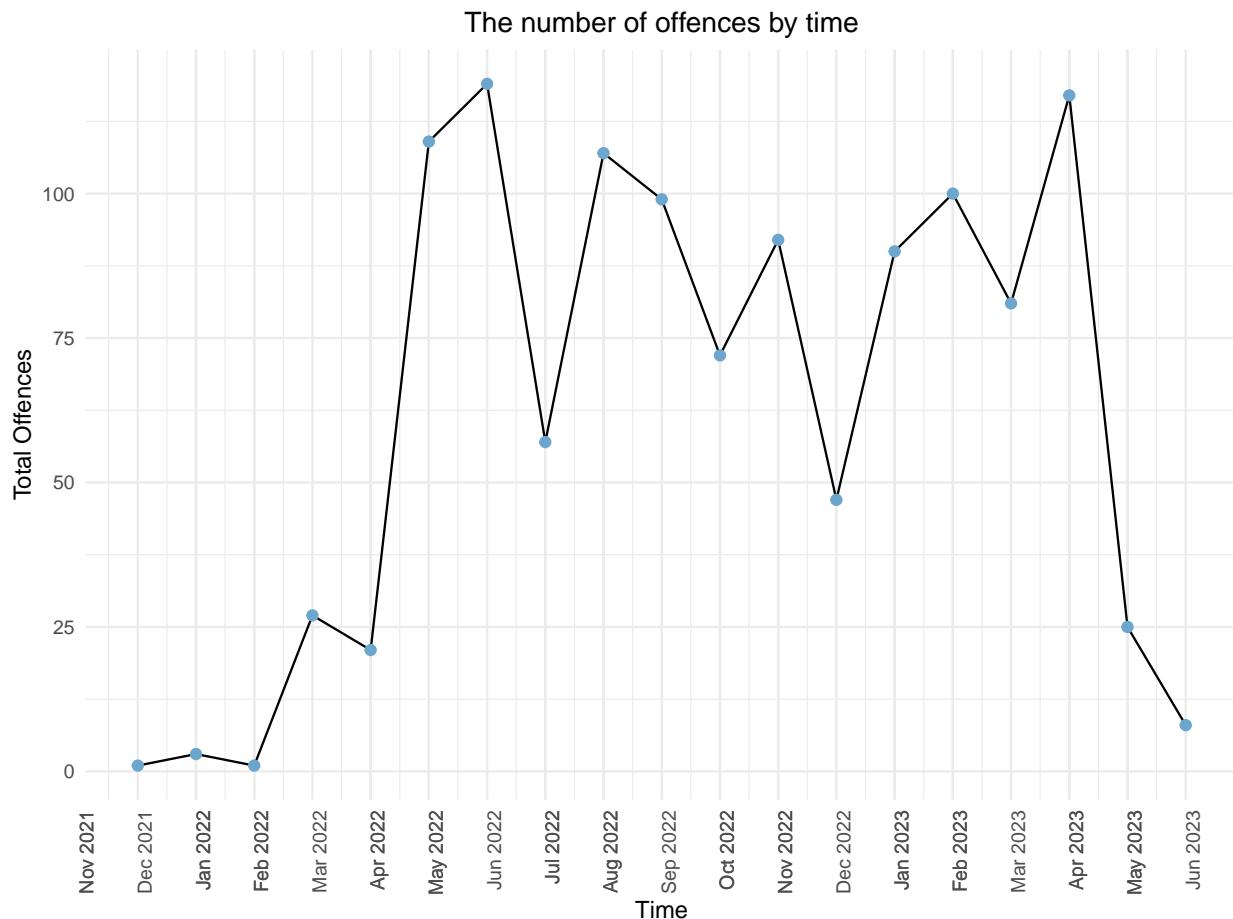


The bar chart shows the top 10 Local Government Areas have the highest number of total offences. Among the LGAs, *Canterbury-Bankstown* has the highest number of offences with 349 incidents, nearly triple more than that of Burwood which is in second rank. While Blacktown comes in third with 107 offences. Cumberland, Liverpool, Willoughby and City of Canada Bay follow suit with decreasing numbers of offences.

Question 2

Use tidyverse commands to calculate the number of offences and average amounts payable by month per year

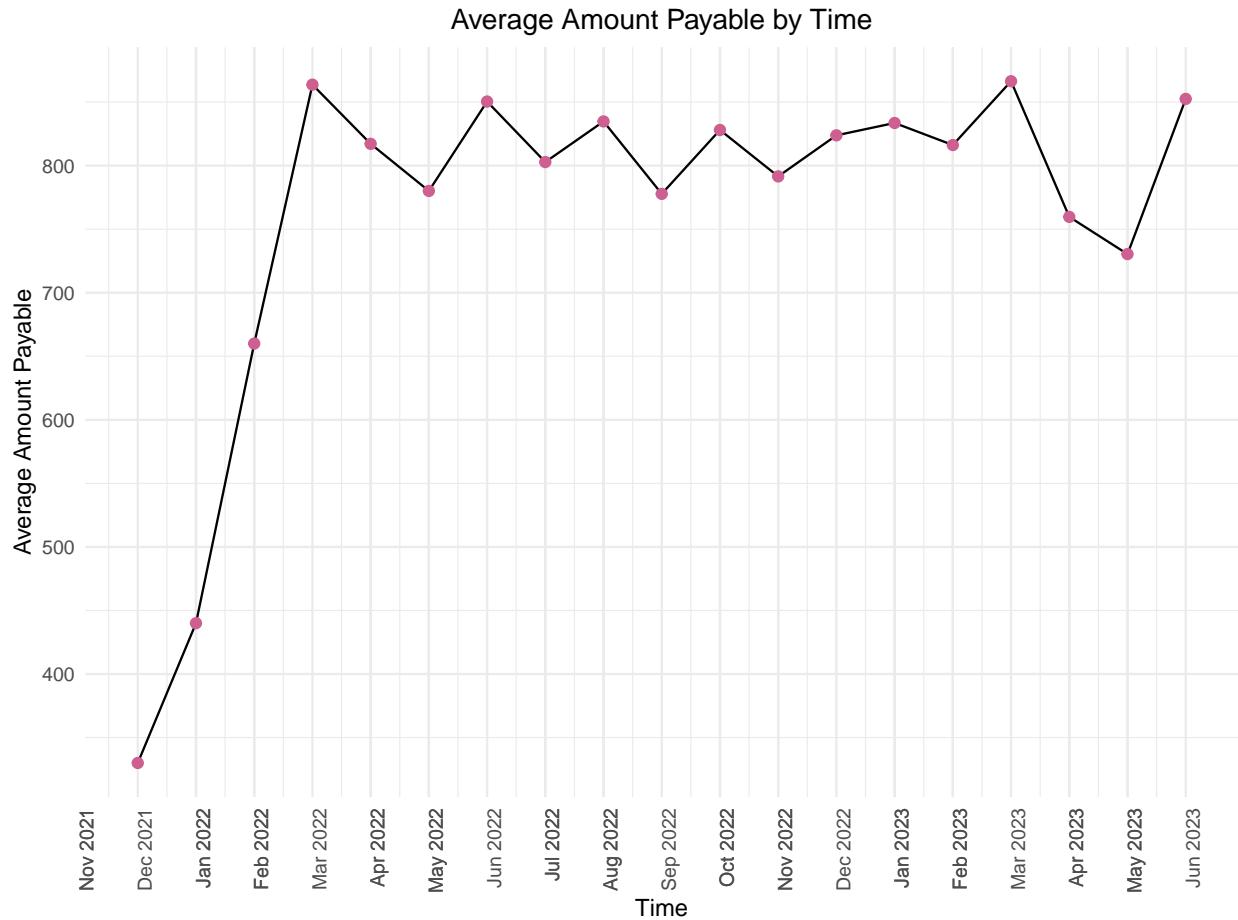
a) Plot the number of offences by time



b) In 3-5 sentences, interpret your plot to answer the research question “Are there any trends, patterns or seasonality in the number of events over time?”

The graph illustrates the number of reported offences from December 2021 to June 2023. Notably, a clear seasonality is evident, with consistent peaks in June 2022 and April 2023, indicating a recurring pattern of increased criminal activity during these months. Furthermore, there is an overall upward trend in reported offences, particularly from December 2021 to June 2022, suggesting a sustained increase in criminal incidents during this timeframe. Furthermore, periodic fluctuations are apparent, with periods of lower offence counts followed by spikes, hinting at potential cyclic patterns or external factors influencing the reported number of offences.

c) Plot the average amount payable by time



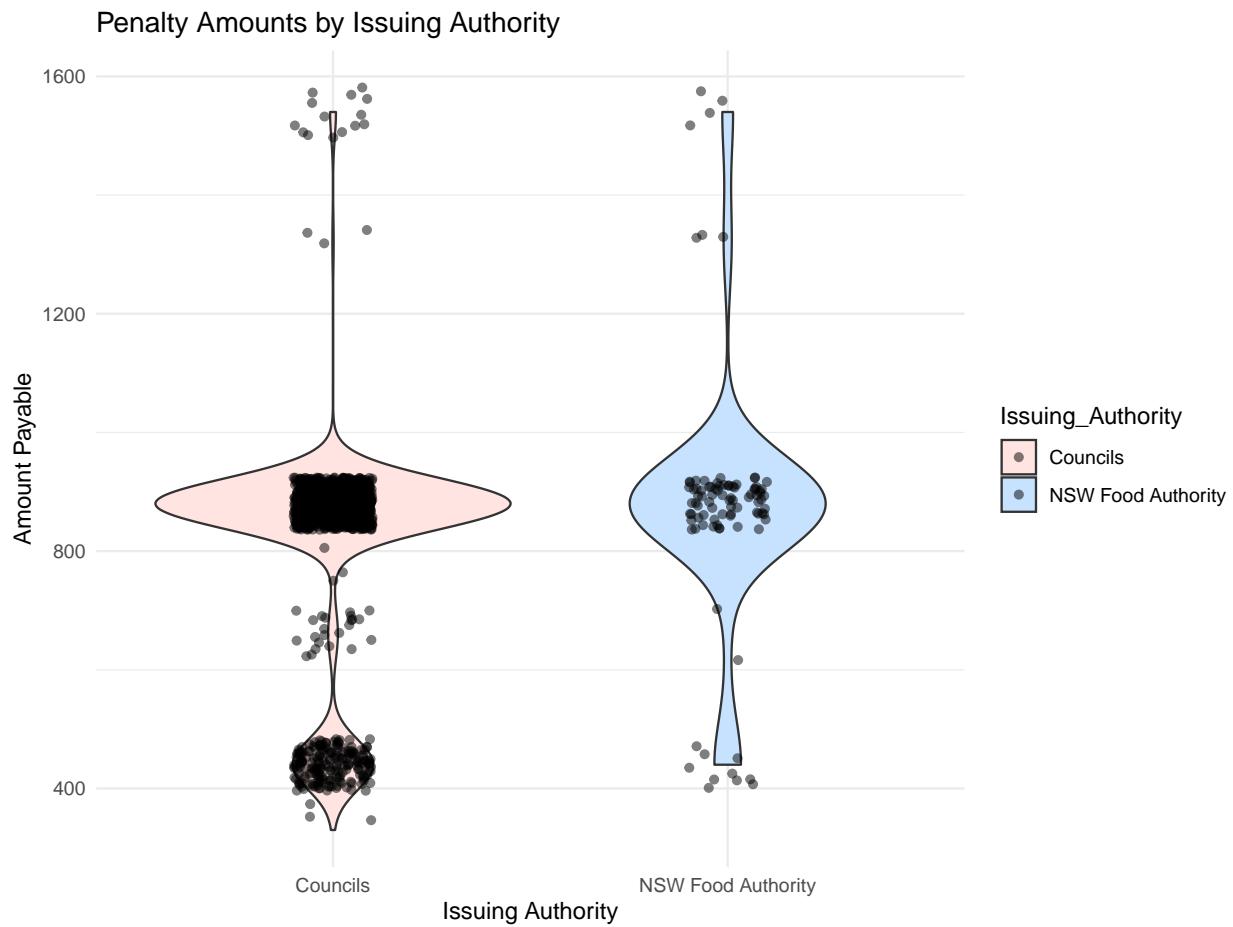
d) In 3-5 sentences, interpret your plot to answer the research question “Are there any trends, patterns or seasonality in average amounts payable over time?”

The time series plot depicting the “Average Amount Payable by Time” exhibits several noteworthy observations. Firstly, there is a discernible overall upward trajectory in the average amount payable by offenders as time progresses, indicating a general increase in the financial penalties imposed. Secondly, a recurring pattern with peaks and troughs emerges, suggesting the presence of seasonality. Notably, there are spikes in average amounts payable in February and March each year, potentially linked to specific periods of heightened offence activity or legal alterations. Moreover, isolated anomalies, such as the peaks observed in March in both years, imply transient periods of unusually high average payments, which may be attributed to external factors.

Question 3

Using all of the data:

- a) Create a violin plot that includes the individual data points to determine if there is a difference in the penalties (amounts payable) issued by the NSW Food Authority versus Councils



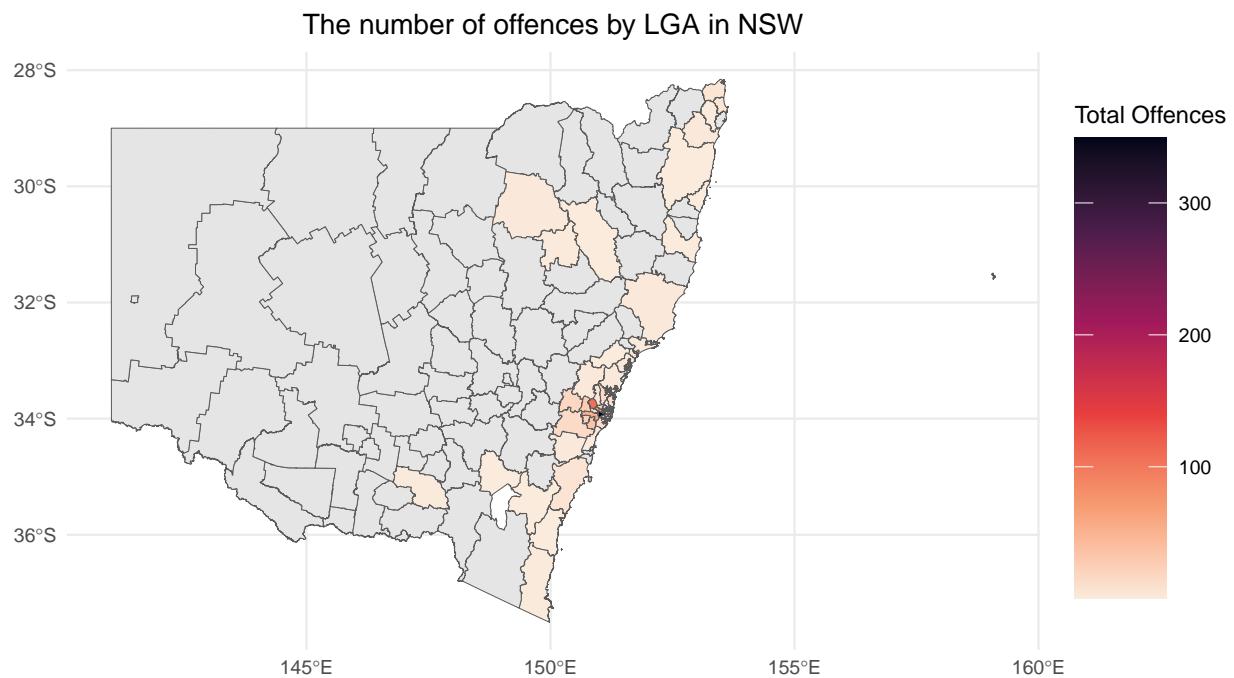
- b) Interpret the plot to answer the research question: “Is there a difference in the amounts payable issued by the NSW Food Authority compared to Councils?”

The violin plot illustrates the distribution of penalty amounts payable, differentiating between the issuing authorities, namely the NSW Food Authority and Councils. There is a difference in the amounts payable issued by the NSW Food Authority compared to Councils. As for Councils, the distribution appears to be bimodal, with two prominent peaks in the density around 440 and 880. This suggests that Councils often issue penalties at these specific amounts, possibly reflecting standardized penalty structures. The distribution for Councils is relatively symmetric, indicating that they have a relatively balanced distribution of penalty amounts on either side of the median. In contrast, the NSW Food Authority's distribution is notably different. This suggests that the NSW Food Authority predominantly issues penalties at or around 880, with fewer deviations from this amount. The distribution also appears slightly skewed to the right, indicating a tendency towards higher penalty amounts than Councils. Therefore, while Councils have a more evenly distributed range of penalty amounts, the NSW Food Authority tends to issue penalties more consistently.

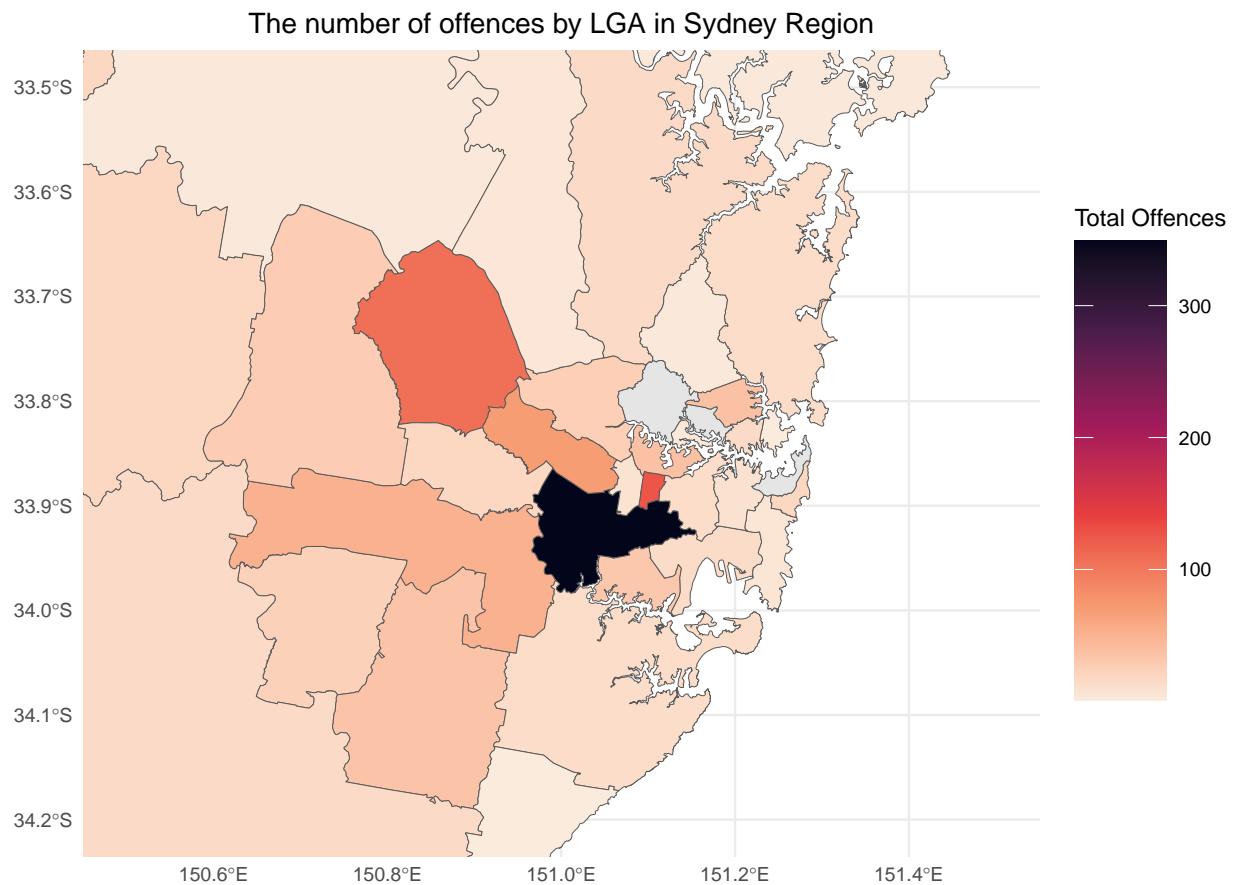
Question 4

Using all of the data:

- a) Create and present a map that shows the number of penalties by LGA in NSW (this could be a choropleth map or a dot density map)



b) Focus the map on the Sydney region



c) In one sentence, explain why we may want to focus on the Sydney region.

Based on the distribution of the darkest colors in the choropleth map for the number of penalties by LGA in NSW, Sydney region has the highest number of penalties by LGA.

d) In about 3 sentences, summarise the main results of the plot presented in part (b)

The choropleth map for Sydney region shows clearly with a color scheme from darkest to lightest showing areas with the most to least food offence, respectively. Moreover, with the legend and range being similar to the NSW's choropleth map color scheme may suggest that Sydney represents efficiently for the allocation of offending areas to NSW. Western and Southern Sydney areas have a much higher offence rate than northern Sydney.

Question 5

The data contains a column showing a detailed description of the offence (`Nature_of_Offence_Full`). The description is quite long, and so it cannot be used directly in plots. However, it is of interest to summarise. It may be necessary to simplify the information contained in this column without losing important information about the nature of the offence.

- a) Describe in detail, using a series of bullet points, how you could handle (simplify) the data in this column so that it can be graphically presented.

To simplify the data of the offence (`Nature_of_Offence_Full`) for graphic presentation, the data should be followed by:

Step 1: Text Preprocessing (to ensure uniformity and cleanliness of the data):

- Convert all text to lowercase to make every value is in the same format.
- Remove any leading or trailing whitespaces.

Step 2: Categorization (categorize the offences based on common keywords or themes in descriptions):

- Create a new column, “`New_Nature_of_Offence`,” for this purpose.
- Use regular expressions to search for specific keywords or phrases that indicate the nature of the offence, e.g., “animal/pest control,” “equipment condition,” “contamination,” etc.
- Assign a simplified label or category to each offence based on the identified keywords.
- Group similar offences together under broader categories to reduce complexity.

Step 3: Labeling for ‘Warning’ (identify offences associated with warnings):

- Create another column, “`Warning_Classification`,” to indicate whether a warning is associated with the offence description.
- Use regular expressions to identify phrases related to warnings and label them accordingly.

Step 5: Data Aggregation (aggregate the data based on the simplified offence categories):

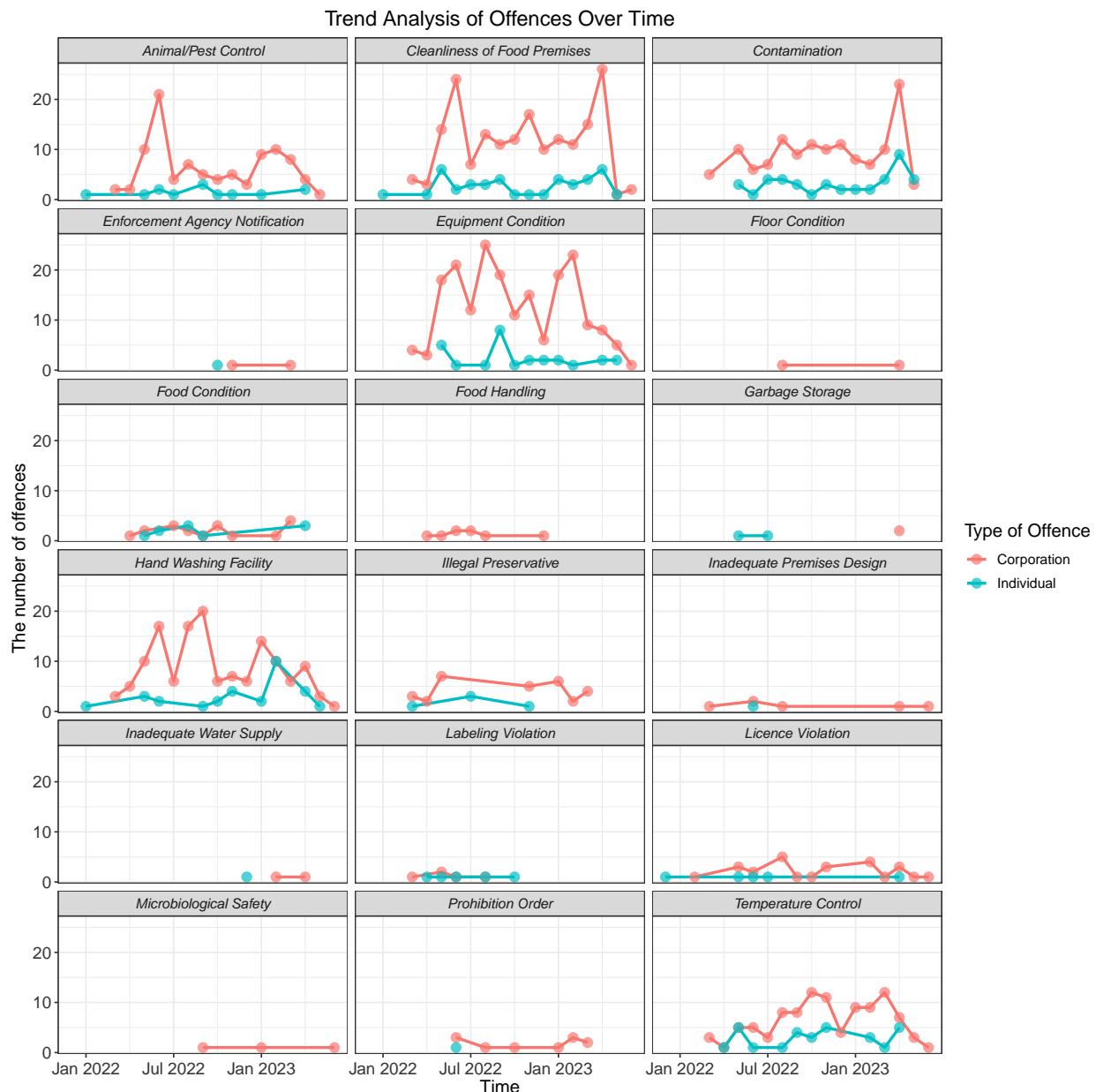
- Group the data by the simplified offence categories.
- Perform aggregations like counting the occurrences of each category or summing associated penalty amounts.

Step 6: Visualization

- Using plots to visualize to represent the simplified offence categories and their relationships with other relevant variables.

b) Complete your data handling (you can ask for help from your conveners with this), and create and present three statistical graphics. The graphics should display the nature of the offence, along with other relevant variables that you think provide insight into the data. For each plot presented, explain in detail (about 100 words) what the plot shows so that a non-expert can understand the pattern/trend.

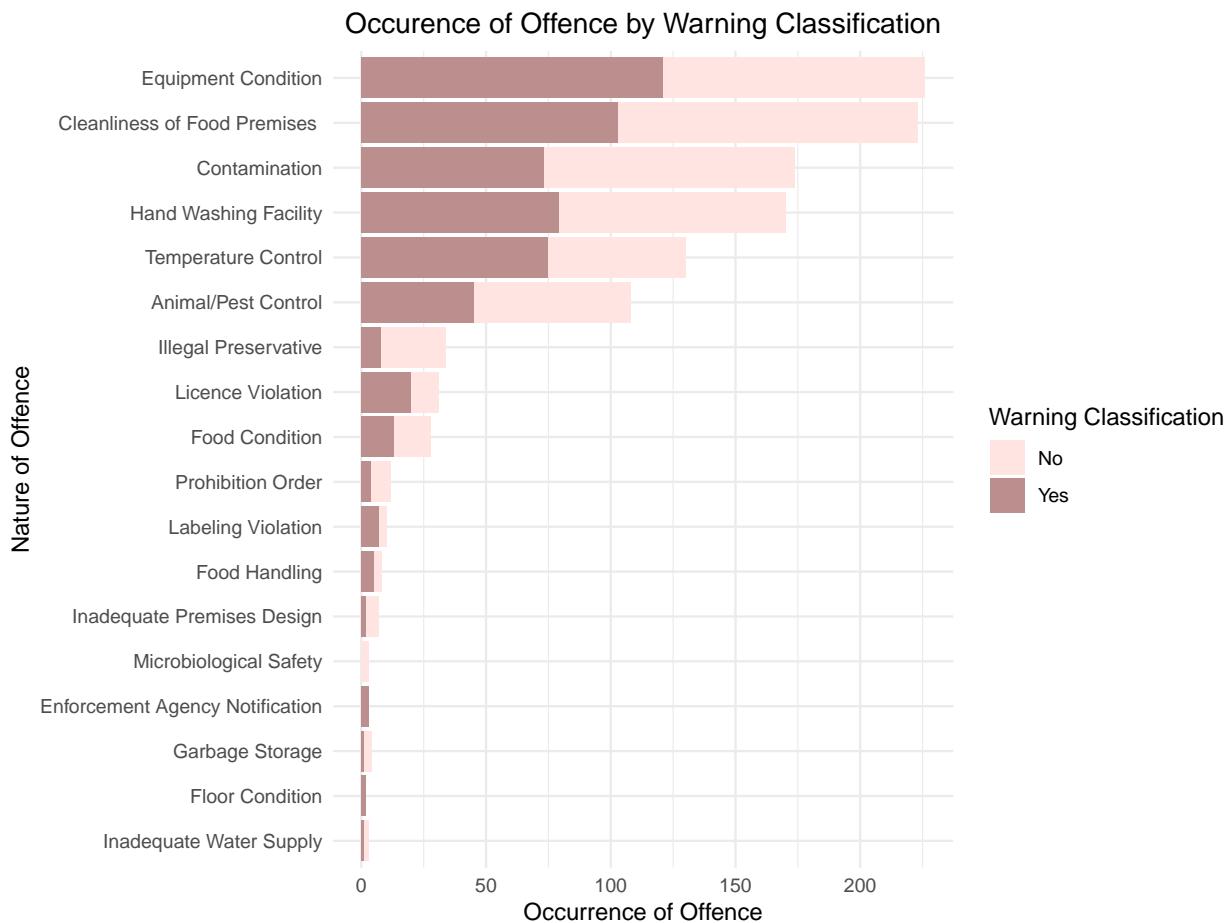
Graphic 1: Trend Analysis of Offences Over Time



The plot illustrates the impact of time on each food safety offence in the years 2022 and 2023, categorized by type of offence and whether they were committed by corporations or individuals. It highlights a consistent trend of reduction in food safety violations throughout the observed period. Notably, corporate entities consistently outpace individuals in the commission of such offences, underscoring the necessity for stringent oversight and regulatory measures within the corporate sector. The top three most prevalent offences

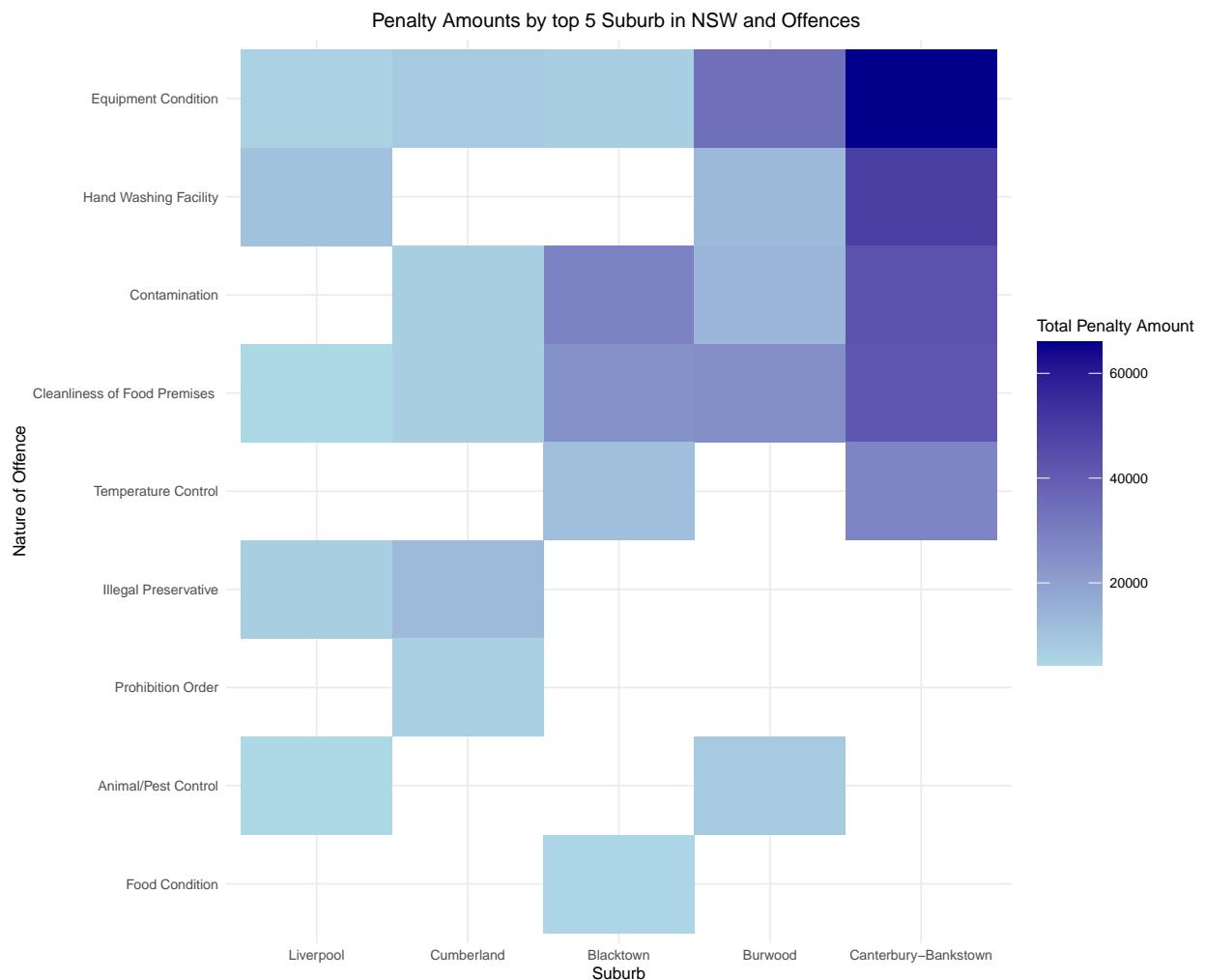
encompass issues related to equipment condition, cleanliness of food premises and hand washing facility, respectively. The number of offences from the middle of 2022 to early 2023 remained at a higher rate in the whole period. However, in general, until mid-2023, the number of crimes tends to decrease, which is a positive sign in reducing food violations.

Graphic 2: Occurrence of Offence by Warning Classification



The stacked bar chart provides a comprehensive view of the relationship between the occurrence of offences and the presence of prior warnings. Each bar represents a distinct type of offence, with two distinct categories: "Yes" indicating that a warning was issued before the offence, and "No" suggesting that there was no prior warning. The plot reveals that offences preceded by warnings ("Yes") seemly exhibit a notably lower occurrence rate compared to those without prior warnings ("No"). This suggests that warnings play a pivotal role in deterring specific offence types. Notably, offences related to "Equipment Condition," "Licence Violation," and "Temperature Control" emerge as particularly responsive to pre-warning measures, underscoring the effectiveness of such interventions in reducing violations within these categories. Conversely, offences such as "Contamination" and "Hand Washing Facility" appear less likely to be forewarned, highlighting areas where preventative actions may need to be reinforced for more comprehensive compliance with regulations.

Graphic 3: Penalty Amounts by top 5 Suburb in NSW and Offences



The heatmap illustrates penalty amounts for the 5 prominent suburbs in New South Wales (NSW) concerning the top 5 food safety offences in each area. The color intensity within each cell serves as a visual indicator of the cumulative penalty amount attributed to that suburb-offence combination. “Equipment Condition” and “Cleanliness of Food Premises” are frequently in the top 5 causes of food offence in all 5 regions, followed by “Contamination” in 4 out of 5 regions. Notably, suburbs such as Canterbury-Bankstown and Blacktown exhibit significantly elevated penalty amounts for offences related to “Equipment Condition” and “Hand Washing Facility”. Meanwhile, in Burwood and Cumberland, offences pertaining to “Equipment Condition” and “Contamination” also garner substantial penalties. Conversely, “Cleanliness of Food Premises” and “Temperature Control” offences consistently result in comparatively lower penalties across all examined suburbs.