Tucson Rain Vs Crimes Data Analysis

By:

Raine Ikagawa, Josie Mazzone, & Cora Ricoy

For our final project, we are investigating the potential correlation between rain amounts and crime type in Tucson, AZ. We are using the supplied data set from Dr. Haiquan's class and the data set from the Tucson Police Data & Analysis web page ([link](https://policeanalysis.tucsonaz.gov/pages/reported-crimes)).

Hypothesis:

We hypothesize that an inverse relationship exists between the quantity of rain and the type of crimes in Tucson, AZ, for each month from 2018 to 2021.

Workflow:

To test this, we first imported both datasets and assigned Tucson rain as 'rain' and Tucson crimes as 'crimes'. Once labeled, we could see a difference in date formatting. To help the merging process, we used the lubridate() function to easily match date formats.

We formatted the 'readingDate' from the 'rain; file and the 'Date.Occurred' from the 'crime' file to match the format of year, month, and day.

After changing the date format, we selected only the essential columns for both data sets

For Tucson crimes the following columns are selected and concatenated into a separate data frame.

colnames: Incident.ID, Date.Occurred, Incident.Type, UCR.Number

For Tucson rain, we chose the following and did the same procedure.

colnames: readingId, readingDate, quality, rainAmount

After the two data frames were established, Josie and Raine each made a function that can produce multiple plots using ggplots.

Function 1:

Josie's function, 'data\_selection', performs the data manipulation and cleaning needed to merge both starting data frames. After the function runs, it will create a data frame called 'data'; 'data' contains the years and months for 2018-2021. The total "good" rain amounts, crimes, and crime types are given for each year and month.

Josie's primary function generates a summary comparison plot. The graph uses bars for the crime types and a line plot for the rain amounts. The x-axis gives the rain amount, and the y-axis is the number of crimes. For this graph, we withheld the crime type 'Larceny' because the crime totals for larceny were considered an outlier to the rest, which created a visual disruption for interpreting the data.

[figure 1. – Summary Plots]Diagram

Description automatically generated

Results 1:

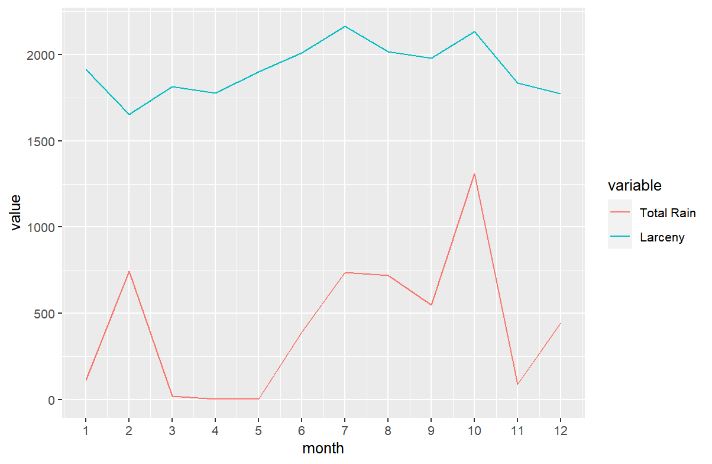
From the graphs, we decerned the overall increase/decrease of rain does not show a specific relationship either inverse or direct with the crime types. However, the infrequent number of events for each crime type impacts the overall trends of each plot. This large variability for each crime type makes it difficult to view small trends in each graph, even with the outliers removed. To better visualize our comparisons, we created a graph that lets the y-axis adjust with the number of events for each crime type for one year.

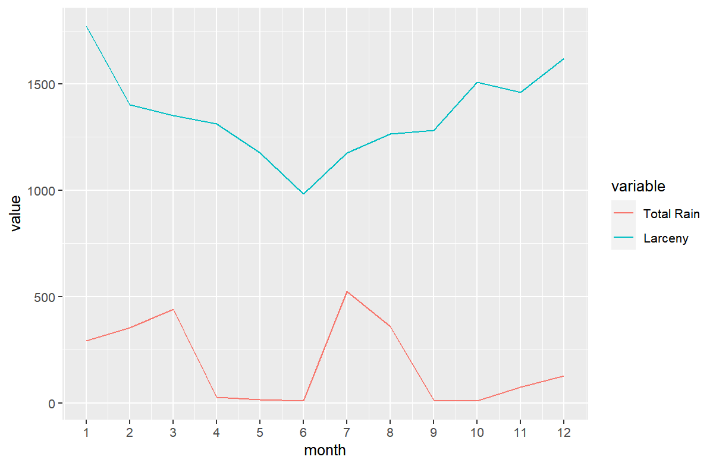
Function 2:

For further inspection between rain amounts per month and crime types, we use Raine's function 'crime\_compare'. This function takes the mentioned data frames and further manipulates them to create a line plot with a flexible y-axis, labeled as "values". The values for each line plot are pulled from both crime type events and total monthly rain counts. This is made possible by using the melt() function, which melds both summaries of rain and crimes, 'rain\_summary' and 'crime\_summary'; which then stores the column names according to the given values which ggplot() can discern. With crime\_compare we can populate line plots for the summarized values making it easy to decern any trends within each comparison. We chose to examine 'Larceny' because it has the highest number of events, thus, the most reliable for showing trends.

[Larceny: figure 2.a - Year 2018] [Larceny figure 2.b - Year 2019]

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated[ Larceny figure 2.c - Year 2020] [Larceny figure 2.d - Year 2021]

Results 2:

Analyzing the last four years shows a slight trend between rain amount and larceny events. This trend is most robust between 2019 and 2020, where a sharp decrease in crime correlates with the expected decline in rainfall for July of both years. We wanted to see if the absence or presence of rain for other crime types for 2019-2020 had any similar characteristics.

We focused on both burglary and aggravated assault for 2019 and 2020.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generatedBurglary: [ figure 3.a 2019] [figure 3.b 2020]

Aggravated Assault: Chart, line chart

Description automatically generatedChart, line chart

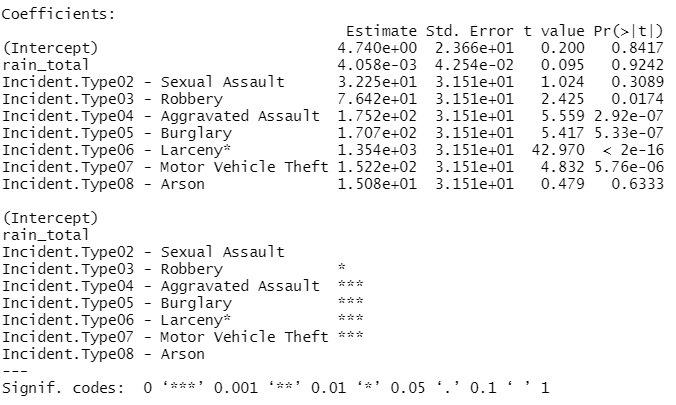
Description automatically generated[figure 4.a 2019] [figure 4.b 2020]

Results 3:

Looking at the plots for Burglary we saw an inverse relationship where a decrease in rain per month has a peak of increased burglary events, clearly shown in figures 3.a. For Aggravated Assault, a direct relationship was observed with the rain amounts per month, clearly displayed in 2020.

Regression Analysis:

In addition to the plots, we created a function that would run a regression analysis for a given year. This analysis allowed us to see if there were significant relationships between types of crime and rain for each year. An example of the results for 2020 is shown below. The asterisks indicate there is a significant relationship between several types of crime (aggravated assault, burglary, larceny & motor vehicle theft) and the total rain throughout the year.



It is important to note this function did not test for normal distribution in the data. Additionally, a relationship between rain and a type of crime does not indicate causation as there are many other factors that could contribute to whether or not a crime is committed. However, this does provide an example of a function that could be used to look at statistical analyses of a specific facet of a dataset.

Conclusion:

We looked over the plots and saw a noticeable trend where the total rain amounts can have an inverse or direct relationship for each type of crime. This helps confirm that our hypothesis is partially true; that the rain amounts have an inverse relationship with certain crime types i.e., larceny and burglary. With this data, the inverse relationship is possible because one type of crime might be easier to commit in a month with less rain than another. Unfortunately, the results are inconclusive until more data can be collected for all crime types. A bigger city might be able to present concrete relationships for each crime type because Tucson, AZ, does not have a high number of homicide, arson, and sexual assault events that are reported. Overall, we can say that a connection between rain amounts, and crime types is present; to know how much, we will need more data to clarify each correlation.

Contributions:

**Josie Mazzome**: Created Function 1

33.3333333%

**Raine Ikagawa**: Created Function 2

33.3333333%

**Cora Ricoy**: Wrote the Report

33.3333333%

All source documents are submitted in D2L and are located on GitHub.

GitHub Project Location: <https://github.com/jmazzone11/BE502.git>

Tucson Crime Report Webpage: <https://policeanalysis.tucsonaz.gov/pages/reported-crimes>