Introduction: Analysis of past data is one of the important sections of society-economy studies that enable researchers, politicians, economists, and societies to find a better understanding of the effect of poverty on the rate of theft. To do this analysis, we have three objectives: 1. What is the probability of crimes over the months of a year? How about the days of a week and days of a month? 2. Does weather have a significant effect on crime? 3. Does weather affect crime in geographies differently? Through our study, we used statistical methods like: 'Pearson-correlations', 'rank-biserial correlation' 'scatterplots', 'averages', 'Median', 'standard deviations', and 'Mann-Whitney U test' to address our objectives.

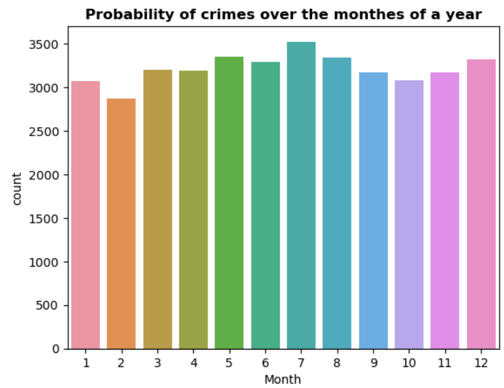
Our results were surprising; they are available on GitHub and Doc Slides. Based on our results for the study Crime Distribution Over Time, crime is relatively consistent throughout the year, with slight peaks in July and December. Mondays have the highest crime rates, while weekends (Saturday and Sunday) have the lowest. Crime rates remain fairly stable throughout the month, with a slight spike at the start and a drop on the 31st day. To evaluate the Poverty and Crime Correlation, Violent crimes such as aggravated assault and robbery are more prevalent in poverty-stricken areas compared to non-poverty areas. When normalized for population size, poverty areas still exhibit higher crime rates for most violent offenses. Also, for evaluate the Weather's Impact on Crime, there is a weak positive correlation between high temperatures and crime rates (Pearson correlation of 0.1637). High temperatures (≥90°F) are associated with a statistically significant increase in violent crime rates. Lastly, to evaluate the Temperature Effect by Area Type, Poverty areas show a stronger correlation between high temperatures and increased crime rates (Pearson correlation of 0.2942) compared to non-poverty areas. Non-poverty areas do not exhibit a significant difference in crime rates between high and low temperatures.

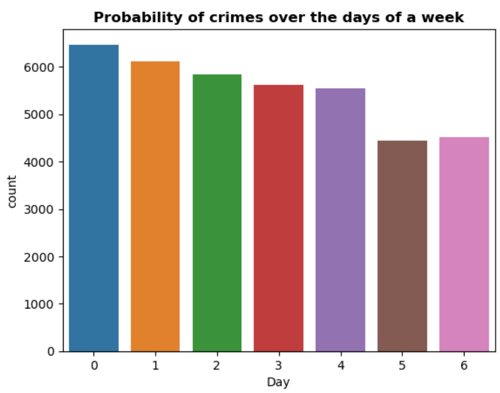
Datasets: The dataset we were given for this project is information compiled about crime statistics in Austin between 2015-01-01 and 2015-12-31 range. We used two important datasets: "crime-housing-austin-2015" and "AustinZipCodes". Also, to find the relationship between temperature and crime rate and answer the objective 2, we downloaded the "austin_weather" dataset. In these datasets, there are many columns that we selected some important columns to answer our questions. We used the 'Highest_Offense_Desc', 'Report_Date', 'Populationbelowpovertylevel', 'Zip_Code_Crime', 'Zip Code', and 'Population' columns. Further, in the 'austin_weather' dataset, there is information about 'Temp,' 'Dew,' 'Humidity,' 'SeaLevelPressure,' 'Wind,' and 'VisibilityHighMiles,' and we worked with 'TempHigh' and 'TempLow' columns.

Analysis Technique: To start our analysis, we first checked the datasets and dropped the rows with the 'NaN' values to ensure the dataset's integrity; then, using the group-by technique, we compared the count of the crimes over months of a year, days of a week, and days of a month. To answer the second question, we defined "poverty" and "non-poverty" zip codes based on these two criteria: **Poverty Level:** Zip codes where the population below the poverty level is greater than 20% ("Populationbelowpovertylevel" > 0.2); and **Median Household Income:** Zip codes where the median household income is less than $50,000 ("Medianhouseholdincome" < 50000). Any zip code that meets both these conditions is considered a "poverty" zip code. Those that do not meet these conditions are classified as "non-poverty" zip codes. By defining these two categories, we can compare Violent Crime by Offense Type and Poverty Status. To better understand, we normalized this data and calculated the adjusted population.

To analyze the relationship between temperature and violent crime rates, daily crime data was merged with population data using zip codes as the common identifier. Population data was converted to numeric format, and daily crime rates per 1,000 people were calculated for each zip code. Then, we performed the KDE histogram, which allowed us to visually inspect the data's distribution to see if it resembled a normal distribution (bell curve). The Q-Q plot enabled us to compare the quantiles of the data against the quantiles of a standard normal distribution. For further understanding, Temperature data was categorized into high-temperature days (≥90°F) and low-temperature days (<90°F). The crime rate distributions were examined through a histogram and a Q-Q plot to assess normality. Lastly, to assess the effect of temperature on crime rates in both poverty and non-poverty areas, the datasets were split into four groups: high-temperature poverty areas, low-temperature poverty areas, high-temperature non-poverty areas, and low-temperature non-poverty areas. Pearson correlation analysis was used to explore the relationship between temperature and crime rates in each group. Non-parametric Mann-Whitney U tests were conducted to test the statistical significance of differences in crime rates between high and low-temperature days. Median crime rates were also calculated for comparison, and the rank-biserial correlation was used to determine effect sizes.

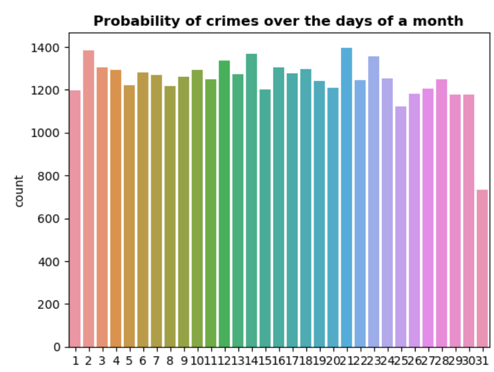
Results: The Figs. 1~3 shows the probability of the crimes over months of a year, days of a week, and days of a month. The first bar chart shows the distribution of crimes across the months of the year, indicating that criminal activities are relatively consistent throughout the year, with only minor fluctuations. The months of July (7) and December (12) exhibit a slightly higher occurrence of crimes, reaching around 3500 incidents, whereas February (2) shows a slightly lower count, close to 3000 incidents, indicating that while crimes can happen in any month, certain periods, particularly mid-summer and the end of the year, might experience a marginal increase in criminal activities.



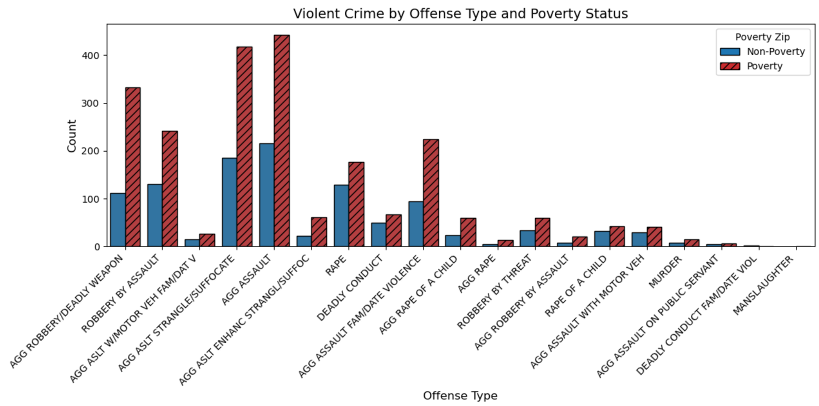


The second bar chart displays the crime rates over the different days of the week are decreasing, with the highest happening on Mondays, with the count surpassing 6000 incidents, and the lowest happening on Saturdays and Sundays, both slightly below 5000 incidents. This pattern indicates a potential decrease in criminal activities towards the end of the week.

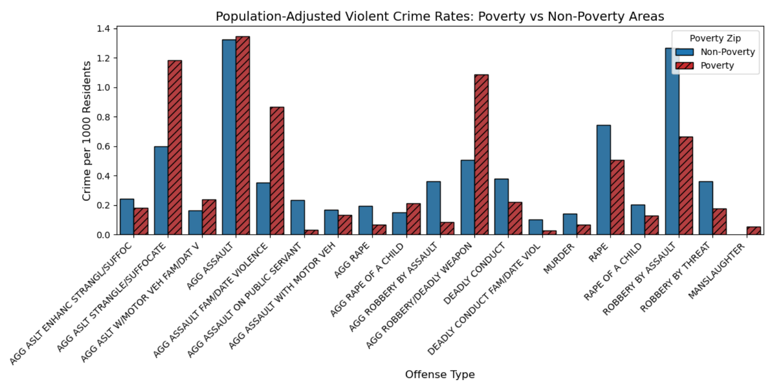
The third bar chart shows the distribution of crimes over the days of a month. This indicates that the number of incidents remains fairly stable throughout the course of time. The crime rates appear slightly elevated on the 1st and 2nd days of the month, each reaching above 1400 incidents, suggesting a potential spike in criminal activity at the start of the month. There is a notable drop on the 31st day, which may be due to months with fewer than 31 days affecting the count. Overall, the chart indicates a relatively uniform distribution of crimes across most days of the month, highlighting only minor variations except at the month's beginning and end. This consistency suggests that no specific day within a month significantly influences crime occurrence.

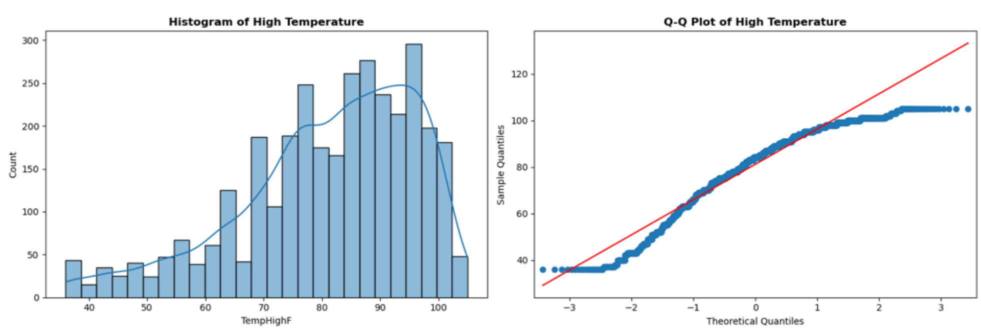


The other result is comparing the crime rate in 'poverty' and 'non-poverty' zip codes. This chart illustrates the counts of violent crimes by offense type in areas categorized by poverty status. It shows that crimes like aggravated assault and robbery are significantly more prevalent in poverty-stricken areas compared to non-poverty areas. The counts for offenses like "AGG ASSAULT" and "AGG ROBBERY/DEADLY WEAPON" are notably higher in poverty areas, which indicates a correlation between poverty and certain types of violent crimes.



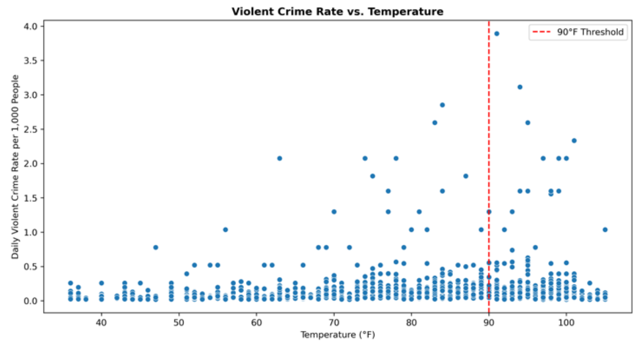
The second chart shows these crime rates for population size, presenting the crime rate per 1,000 residents in both poverty and non-poverty areas. When normalized for population, poverty areas continue to exhibit higher crime rates for most violent offenses. For example, offenses like "AGG ASSAULT" and "AGG ASLT STRANGLE/SUFFOCATE" show significantly higher rates in poverty areas, reinforcing the association between higher crime rates and impoverished conditions.

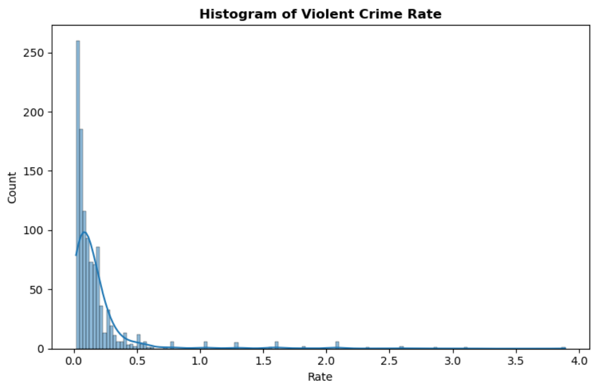


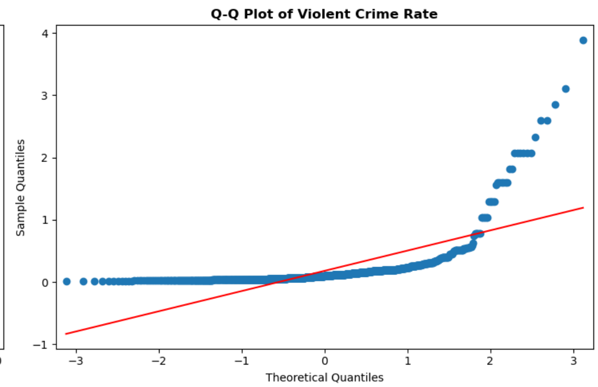


This figure also contains two subplots: the crime rate and temperature relationship. The histogram of high temperatures (TempHighF) appears to be right-skewed, with most of the temperatures concentrated between 70°F and 100°F. This indicates that in warm conditions, the crime rate increases. The right panel displays a Q-Q plot to assess the normality of the temperature distribution. The deviation from the red line, particularly at the tails, indicates that the high-temperature data does not follow a normal distribution.

This set of results shows the crime rate and its relationship with Temperature. As we can see in the scatterplot, the crime rate is distributed in different temperatures, while in the temperature higher than the threshold the 'Daily Violent Crime Rate per 1.000 people' increases which is the same as the previous results that in high temperatures the crime rate increases. Also, the histogram and Q-Q plot proves the dataset's non-normality.





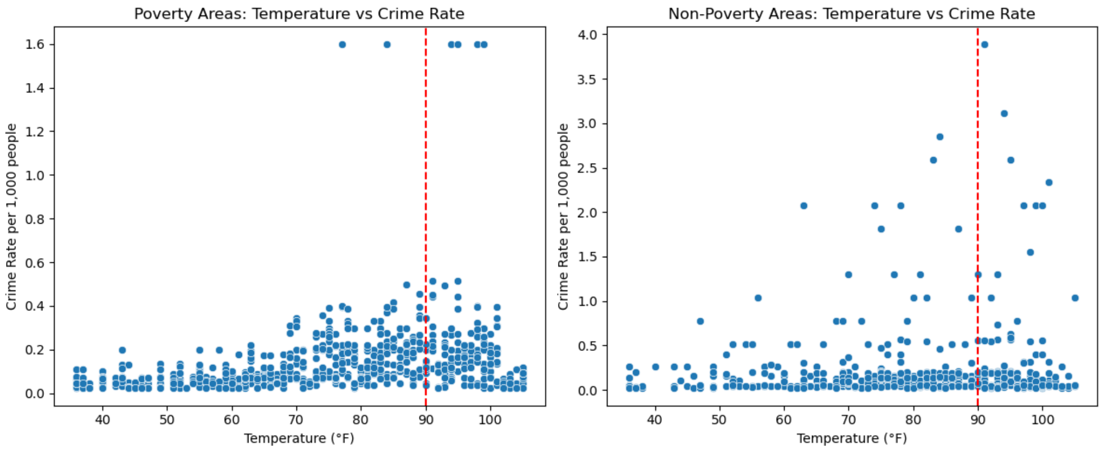


This table shows the statistical analysis of temperature and violent crime rates. The Pearson correlation coefficient between temperature and violent crime rate is 0.1637 (p-value = 0.0000), indicating a weak positive correlation. The median crime rate for high-temperature days (≥ 90°F) is 0.123 per 1,000 people, with a standard deviation of ± 0.438. The Mann-Whitney U test results in a U statistic of 144286.0 and a p-value of 4.407e-05, suggesting a statistically significant difference in violent crime rates between high and low-temperature days. This implies that violent crime rates are generally higher on days with temperatures reaching or exceeding 90°F.

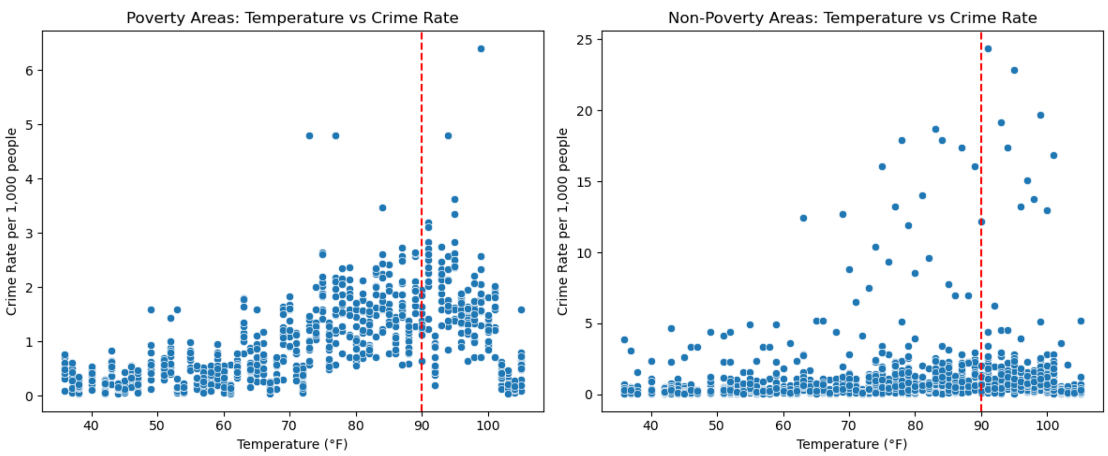
Pearson Correlation	Median Crime Rate (High Temp)	Mann-Whitney U statistic	p-value
0.1637 (p-value: 0.0001)	0.123 ∓ 0.089	144286	4.41E-05

The last section is about finding the relationship between High/low temperature and poverty/poverty areas. The analysis indicates a relationship between temperature and crime rates in both poverty and non-poverty areas, with a more pronounced impact observed in poverty areas. The first scenario is for violent crime rates, and the second one considers all types of crimes. For violent crime rates, the Pearson correlation in poverty areas was 0.2942 (p-value = 0.0000), suggesting a weak positive correlation, and the Mann-Whitney U test revealed a statistically significant difference in crime rates between high-temperature (≥90°F) and low-temperature days (p-value = 1.29e-06). In contrast, non-poverty areas showed a negligible correlation (0.1047, p-value = 0.0206) and no significant difference in crime rates between high and low temperatures (p-value = 0.1697). Similarly, for overall crime rates, poverty areas displayed a moderate positive correlation (0.5364, p-value = 0.0000), with crime rates significantly higher on hot days (median = 1.383 per 1,000 people) than on cooler days (median = 0.597 per 1,000 people). Non-poverty areas also showed a significant increase in crime rates on high-temperature days, though the effect was less pronounced (effect size = -0.2465).

In summary, the findings suggest that high temperatures have a greater influence on crime rates in poverty areas compared to non-poverty areas. Specifically, violent crime rates are significantly higher on hot days in poverty-stricken areas, while non-poverty areas do not exhibit the same pattern. When considering all crime types, both poverty and non-poverty areas experience an increase in crime rates on hot days, with poverty areas showing a stronger and more significant effect. This implies that interventions aimed at reducing temperature-related crime surges might be particularly beneficial in poverty-stricken regions.



Violent crime



All types

Crime	Zip code	Pearson Correlation	Median Crime Rate (High Temp)	Mann-Whitney U statistic	p-value
Violent crime	Poverty Areas	0.2942 (p-value: 0.0000)	0.143 ∓ 0.247	45621	1.29E-06
	Non-Poverty Areas	0.1047 (p-value: 0.0206)	0.123 ∓ 0.438	28245	0.17
All types	Poverty Areas	0.5364 (p-value: 0.0000)	1.38 ∓ 0.884	82987.5	7.28E-16
	Non-Poverty Areas	0.1722 (p-value: 0.0206)	0.701 ∓ 3.010	223274	4.81E-12

Technical: For the data preparation process for this project, we used three datasets, “crime-housing-austin-2015”, “AustinZipCodes”, and “austin_weather”. The last dataset we downloaded (link) was used to see the effect of the weather on the crime rate. All datasets were clean, and we just dropped the “NaN” rows to ensure our analysis was consistent. We used the bar chart to see what is the trend of the crime rate on different days of the week, different months of the year, and different days of the year. Then, we defined two categories, "Poverty" and "non-poverty," based on two conditions explained in detail in the Analysis Technique section to see the effect of the Zip code on the crime rate and trends. We also created the histogram and Q-Q plot to determine the normality of our results.

In the last section, we used scatterplots for several analyses because they seemed to display the data best and made it very clear if our correlations were positive or negative, such as the relation between the Zip code/temperature with the number of crimes. To evaluate our result numerically, we calculated different statistical values like "Pearson correlations", "rank-biserial correlation", "averages", "Median", "standard deviations", and "Mann-Whitney U test".