With this data of NFL arrests, we were hoping to see any possible correlations between the amount of arrests and the other rows to measure correlation, statistics and the like.

We have numerical values in season (year), week number, game time, home score, away score and arrests. Weekday, home team and away team are categorical, and division game is boolean.

The first variable we explored is day of the week, where we compared both the average and total sum of each arrest per game day. There were five days in the data set: Sunday, Monday, Wednesday, Thursday and Saturday. Tuesday and Friday had no entries. For this, Sunday had by far the largest amount of arrests, 5217, nearly ten times more than Monday, who had a second-highest total of 527. The means were different--Wednesday had only one entry in the dataset with 39 arrests, so while it had the lowest sum (barring Tuesday and Friday), it had the highest average of 39.

The next section explored was score range. For this, a row was added to the dataset that subtracted the away score from the home score and gave us the absolute value, so as to not have potentially negative integers. The range from 0-10 in difference provided the most amount of sums (3,666), nearly 2,000 more than the 10-20 range (1,679).

For each game, the average number at home games was measured vs. the start time of game. We found no correlation.

For team record vs. average number of arrests, we merged NFL arrests per game and NFL team records. We then measured team record vs. average arrests and the moving team record (average team record up to that year, to see if it plays a part in the data). The graphs both have spikes at the 0.5 mark for most arrest occurrences and showcase. This similar behavior could be to our dataset, and a set with a larger number of years may be a better indicator of the graphing results.

For week in the season vs. average number of arrests, we analyzed the seasons to see if there were weekly patterns. Then, we checked the five seasons to see if there were outliers. Although we did not find an intended correlation, we found that weeks 4, 8 and 11 have much lower average arrests. But, becauses the medians of this data are below the means, there must be major outliers below the mean.

The average number of arrests home game showed San Diego at the highest, with Tampa Bay in last.

With accounted outliers, we found that 40% of outlier games are divisional games. But, through our own research, we also found that that only 6 out of a total of 16 games are divisional, in which case they are not an apt barometer for measurement.

Our last measurement was number of arrests vs. city population. We used a public dataset and arrest records per city (home game) and then found out that there isn't much a correlation between the two.

The limitations for our survey are the limited number of years--we don't have the statistics from 2016 to 2017 (or, the current weeks of 2018, per NFL year) in the dataset, so we cannot hypothesize how much this data has changed. Beyond that, the reason for arrests was not provided, so we can't do a further analysis for types of arrest. We also dropped any rows that did not have full datasets.