NFL Statistics and Arrests

Names:
- Josh Janda
- Jack Massey
- Ashley Arroyo
- Isaias Lopez

Our tentative title for the project is ***NFL Statistics and Arrests – Do better teams have more player arrests?***

The two data files we are using are:
- https://data.world/chrishayles/nfl-data/workspace/file?filename=receiving.csv
- https://www.kaggle.com/patrickmurphy/nfl-arrests

The first dataset contains 14 variables and 67,761 observations regarding receiving player statistics from the 2000-2016 regular season. Some variables we deem important from this dataset are:
- Name (Used to merge datasets together)
- Team (Used to merge/compare datasets)
- `Rec` - Total number of receptions
- `yds` - Total number of yards received
- `td` - Total number of touchdowns

These variables are all that look to be important, note that this is not an exhaustive list of all variables in the dataset.

The second dataset contains 8 variables and 850 observations regarding player arrests from 2000-2017. Some variables we deem important from this dataset are:
- Team (Used to merge/compare datasets)
- `Date` - Date of incident
- Name (Used to merge datasets together)
- `Position` - Position of player
- `Case` - Incident type
- `Category` - Incident Crime Categories
- `Description` - Description of crime
- `Outcome` - Outcome of arrest

Some background information of these datasets is that they contain NFL statistics, for both arrests by player and team as well as receiving statistics by player and team between 2000-2017 and 2000-2016 respectively. The sources for these datasets are given above.

Our interest in these datasets are to utilize both to ultimately answer the questions:
- Do better teams have a greater number of arrests?
- What makes a `better` team?

- What teams have more arrests? Where are they located?
- What defines a `good` player?
- What do the distribution of player statistics look like?

Overall, our interest is to gain insight from both datasets alone, as well as together to see how player arrests affect teams and their statistics. Through answering this question, we will gain insight to all following questions above.

A quick description of data preparation is that we plan to load in the data utilizing a comma-separated delimiter as both of these datasets are CSV files. We plan to merge them together on different variables, such as player team or name depending on what questions we are looking to answer. We also plan on subsetting the data to look deeper at certain teams, variables, and so forth.