Jim Chatterton Project #4 – Interactive Visualization Using D3 or R Shiny

For each visualization, I used Plot.ly for Python.

**Visualization 1**

The first data set I used was baseball team attendance and payroll data from each year from 2016 to 2019 from Baseball Reference. I specifically pulled the Miscellaneous Team Info table from the following page for the 2016, 2017, 2018, and 2019 seasons: <https://www.baseball-reference.com/leagues/MLB/2019-misc.shtml>

The data set references each team’s attendance, attendance per game, All Stars from that year, estimated payroll, average time for each game, their manager’s, the season the stats are from, as well as some other data.

The visualization I used is a scatterplot with the estimated payroll as the x-axis and their attendance as the y-axis. Additionally, each point was colored based on the number of all stars the team had that season. Based on the scale, the lighter the color of green, the fewer players on that team were all stars. The plot shows that as payroll increases, attendance increases. There are also more points that are a darker green the further up and to the right on the graph you go.

![A close up of text on a white background

Description automatically generated]()

From the sample code for the scatter plot, I updated the color to pull from a specific scale while also editing the size and the format of the points of data so that they stood out on the graph more. I also changed the label of the #A-S column to be All Stars so that it was reader friendly.

**Visualization 2**

The next data set is also from Baseball Reference which details the attendance data for every season of the San Francisco Giants’ history including their years as other team names back in the 1800s. I retrieved the data from the Franchise Attendance, Stadiums and More table on the following page: <https://www.baseball-reference.com/teams/SFG/attend.shtml>

The data shows the year, team name, league, wins, losses, place finished, playoff results, attendance, attendance per game, their attendance rank for their league, estimated payroll, park factors for batters and pitchers, and the stadium name.

I used a line graph to visualize the attendance on a year to year basis for the Giants while also showing which stadium they were playing in at the time. The data had many individual stadium names listed but I grouped them into four groups that represented the building they played in during that time. Each stadium is represented as a different color on the graph.

![A screenshot of text

Description automatically generated]()

I changed up the line graph from the example quite a bit as I wanted to split the graph into different colors while still having a continuous line instead of four different ones. I did this by using the add\_scattergl function to specify a line color based on the stadium name. I also changed the x-axis ticks to be every five years.

**Visualization 3**

The final data set is also from Baseball Reference. It is from the table Team Game-by-Game schedule for the New York Mets 2019 Schedule and Results page here: <https://www.baseball-reference.com/teams/NYM/2019-schedule-scores.shtml>

The data shows every single game the Mets played, the date, their opponent, if they won or lost, runs scored, runs allowed, their record after the game, rank in the division after the game, winning pitcher, losing pitcher, the pitcher that recorded the save, the time of the game, if the game was in day or night, the attendance, and the win/loss streak after the game.

I used a polar scatterplot to visualize this data. Each team the Mets played against is represented as the points around the chart (theta), and the radius of the chart is the run differential of each game (runs scored – runs allowed). Each win is represented as blue and each loss as orange.

![A screenshot of a cell phone

Description automatically generated]()

I changed up the example plot by editing the theme and the color scale to use as I felt it important to use the Mets team colors. Additionally, like the first example, I edited the dots used in the plot to stand out more.