

SQL Final Project

Group 1

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Question

Is there any correlation between number of wins and team salary? Use data from 2000 and later to answer this question. As you do this analysis, keep in mind that salaries across the whole league tend to increase together, so you may want to look on a year-by-year basis.

SQL Code

- ```
SELECT s.yearid, s.teamid, t.name, team_salary, t.w as wins
FROM (SELECT salaries.yearid, salaries.teamid, SUM(salary) as team_salary
FROM salaries
WHERE salaries.yearid >= 2000
GROUP BY salaries.yearid, salaries.teamid) s
LEFT JOIN teams t on s.teamid = t.teamid AND s.yearid = t.yearid
ORDER BY name;
```

# SQL Results – ordered by team name

|   | yearid | teamid | name                 | team_salary | wins |
|---|--------|--------|----------------------|-------------|------|
| ▶ | 2000   | ANA    | Anaheim Angels       | 51464167    | 82   |
|   | 2002   | ANA    | Anaheim Angels       | 61721667    | 99   |
|   | 2004   | ANA    | Anaheim Angels       | 100534667   | 92   |
|   | 2001   | ANA    | Anaheim Angels       | 47535167    | 75   |
|   | 2003   | ANA    | Anaheim Angels       | 79031667    | 77   |
|   | 2000   | ARI    | Arizona Diamondbacks | 81027833    | 85   |
|   | 2015   | ARI    | Arizona Diamondbacks | 61834000    | 79   |
|   | 2011   | ARI    | Arizona Diamondbacks | 53639833    | 94   |
|   | 2006   | ARI    | Arizona Diamondbacks | 59684226    | 76   |
|   | 2013   | ARI    | Arizona Diamondbacks | 90132000    | 81   |
|   | 2004   | ARI    | Arizona Diamondbacks | 69780750    | 51   |
|   | 2002   | ARI    | Arizona Diamondbacks | 102819999   | 98   |
|   | 2010   | ARI    | Arizona Diamondbacks | 60718166    | 65   |
|   | 2007   | ARI    | Arizona Diamondbacks | 52067546    | 90   |
|   | 2008   | ARI    | Arizona Diamondbacks | 66202712    | 82   |

# SQL Results – ordered by team salary (descending)

| yearid | teamid | name                | team_salary | wins |
|--------|--------|---------------------|-------------|------|
| 2013   | NYA    | New York Yankees    | 231978886   | 85   |
| 2013   | LAN    | Los Angeles Dodgers | 223362196   | 92   |
| 2016   | NYA    | New York Yankees    | 222997792   | 84   |
| 2016   | LAN    | Los Angeles Dodgers | 221288380   | 91   |
| 2014   | LAN    | Los Angeles Dodgers | 217014600   | 94   |
| 2015   | LAN    | Los Angeles Dodgers | 215792000   | 92   |
| 2015   | NYA    | New York Yankees    | 212751957   | 87   |
| 2005   | NYA    | New York Yankees    | 208306817   | 95   |
| 2008   | NYA    | New York Yankees    | 207896789   | 89   |
| 2010   | NYA    | New York Yankees    | 206333389   | 95   |
| 2011   | NYA    | New York Yankees    | 202275028   | 97   |
| 2009   | NYA    | New York Yankees    | 201449189   | 103  |
| 2014   | NYA    | New York Yankees    | 197543907   | 84   |
| 2012   | NYA    | New York Yankees    | 196522289   | 95   |
| 2016   | DET    | Detroit Tigers      | 194876481   | 86   |

# SQL Results – ordered by wins (descending)

|   | yearid | teamid | name                  | team_salary | wins |
|---|--------|--------|-----------------------|-------------|------|
| ► | 2001   | SEA    | Seattle Mariners      | 74720834    | 116  |
|   | 2004   | SLN    | St. Louis Cardinals   | 83228333    | 105  |
|   | 2002   | NYA    | New York Yankees      | 125928583   | 103  |
|   | 2002   | OAK    | Oakland Athletics     | 40004167    | 103  |
|   | 2009   | NYA    | New York Yankees      | 201449189   | 103  |
|   | 2016   | CHN    | Chicago Cubs          | 154067668   | 103  |
|   | 2001   | OAK    | Oakland Athletics     | 33810750    | 102  |
|   | 2011   | PHI    | Philadelphia Phillies | 172976379   | 102  |
|   | 2002   | ATL    | Atlanta Braves        | 92870367    | 101  |
|   | 2003   | ATL    | Atlanta Braves        | 106243667   | 101  |
|   | 2003   | NYA    | New York Yankees      | 152749814   | 101  |
|   | 2004   | NYA    | New York Yankees      | 184193950   | 101  |
|   | 2003   | SFN    | San Francisco Gia...  | 82852167    | 100  |
|   | 2005   | SLN    | St. Louis Cardinals   | 92106833    | 100  |
|   | 2008   | LAA    | Los Angeles Angel...  | 119216333   | 100  |

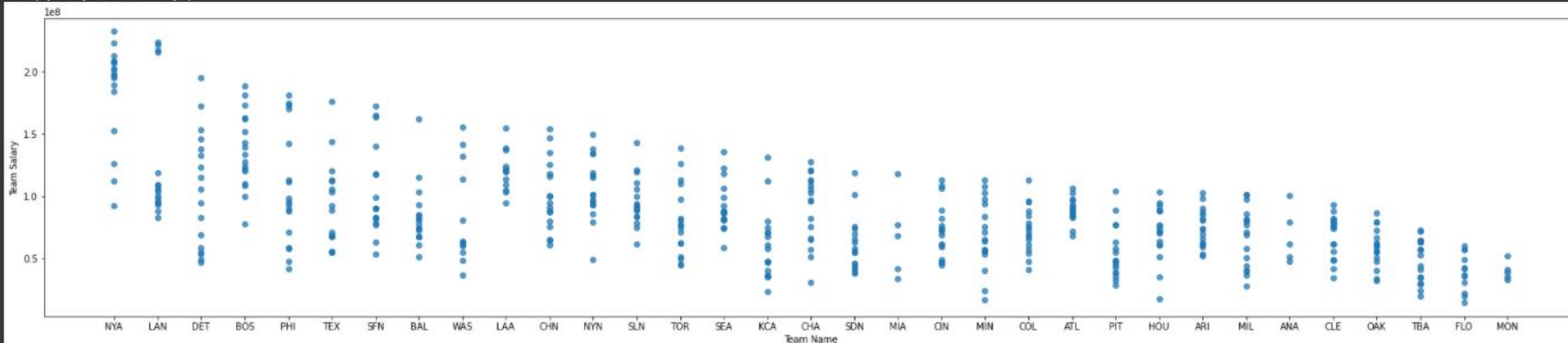
# Selected Visualizations

```
fig = plot.subplots(figsize = (30,6))
```

```
plot.scatter(x = df['teamid'],
 y = df['team_salary'],
 alpha = 0.75,)
```

```
plot.xlabel('Team Name')
plot.ylabel('Team Salary')
```

```
Text(0, 0.5, 'Team Salary')
```



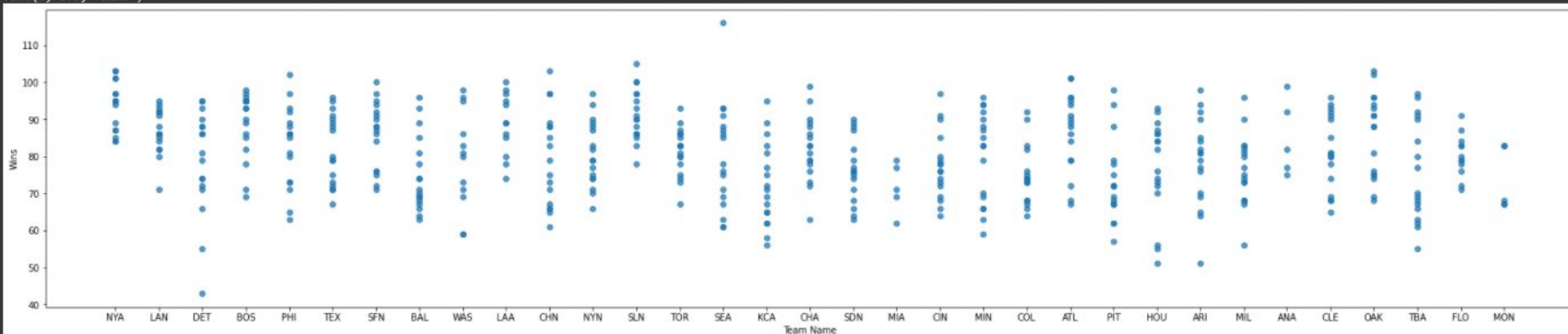
# Selected Visualizations

```
fig = plot.subplots(figsize = (30,6))
```

```
plot.scatter(x = df['teamid'],
 y = df['wins'],
 alpha = 0.75,)
```

```
plot.xlabel('Team Name')
plot.ylabel('Wins')
```

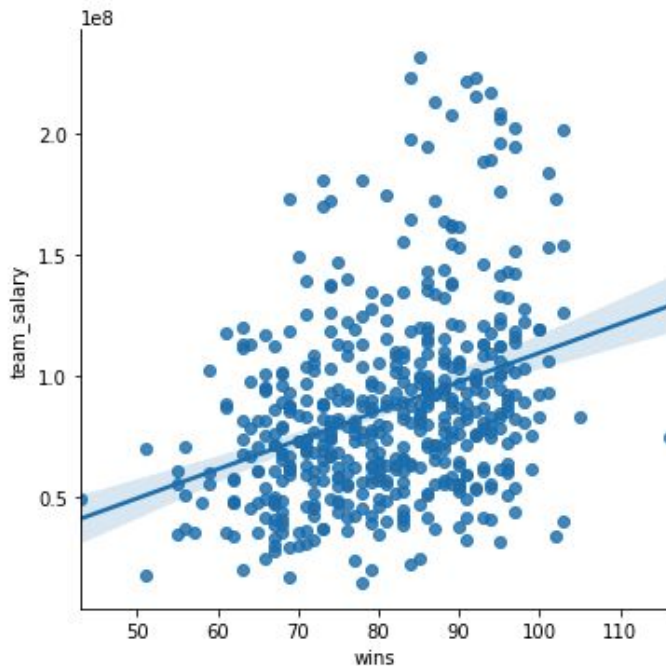
```
Text(0, 0.5, 'wins')
```





# Python Visualization

```
1 sns.lmplot(x="wins", y="team_salary", data=team_salary)
2 plt.show()
```



# Conclusion

- No significant correlation between number of team wins and salary
- Team salary based on other factors... may be location, attendance, population density, etc.
- New York Yankees in 2013 and the Los Angeles Dodgers in 2013 were the top highest salaried teams
- Seattle Mariners in 2001 and the St. Louis Cardinals in 2004 had the most wins