Assignment Two Tyler Jensen

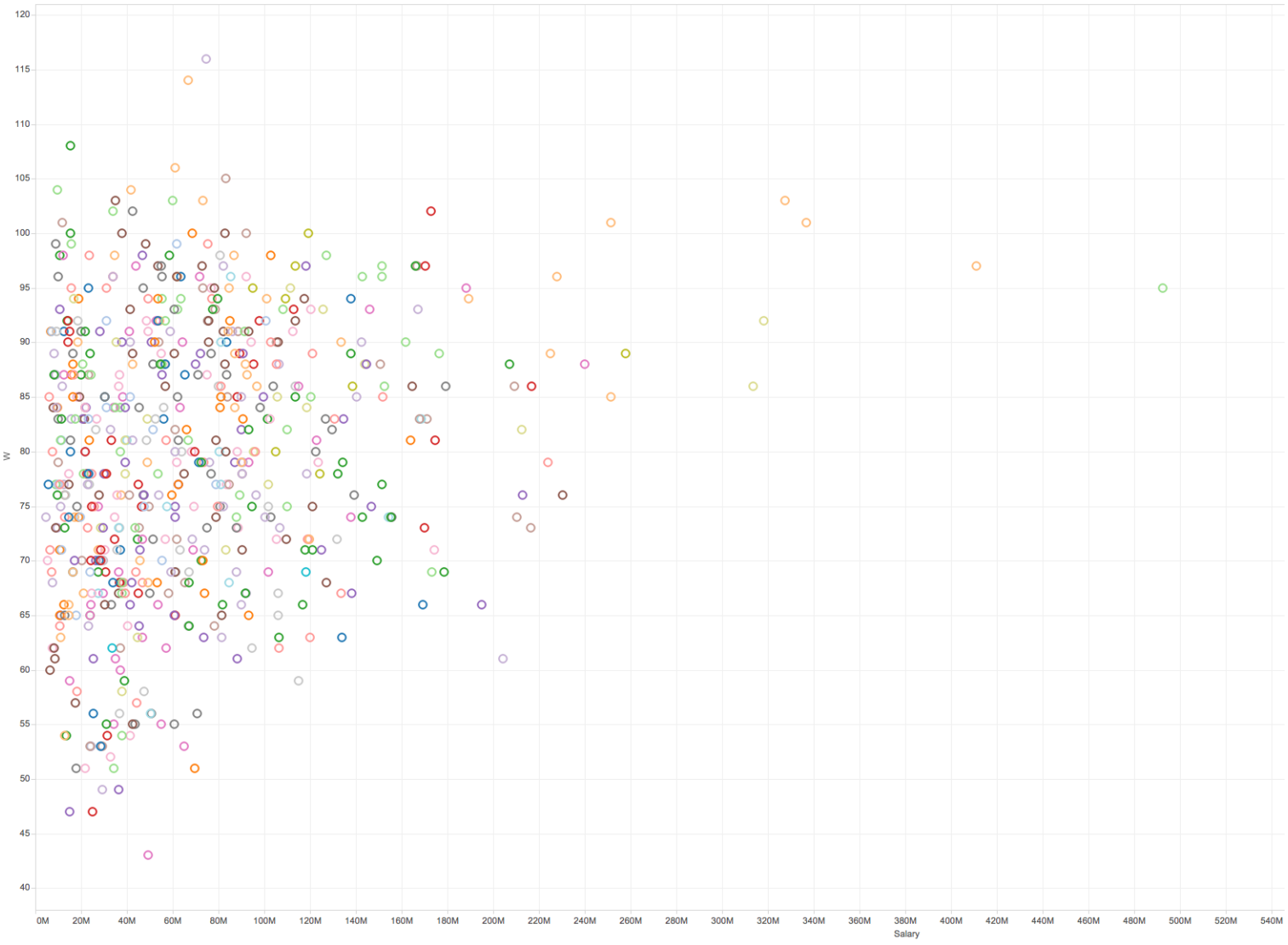
The question I set out on answering was "how much effect does a baseball teams salary have on it's wins?"

Background : Unlike some other leagues in the US, baseball doesn't have a salary cap, which means that teams are allowed to pay players as much as they like. This causes teams like The New York Yankees to win consistently every year, while teams with smaller payrolls (like the Mariners) are at a severe disadvantage

Data processing : I took two different CSV files and opened them in excel. One had stats on teams since the 1800s including win and loss records, the other had payroll information for teams since 1985. Something that was really difficult was the payroll CSV only had individual players, so I had to do some excel trickery to sum up the players salaries for any given team.

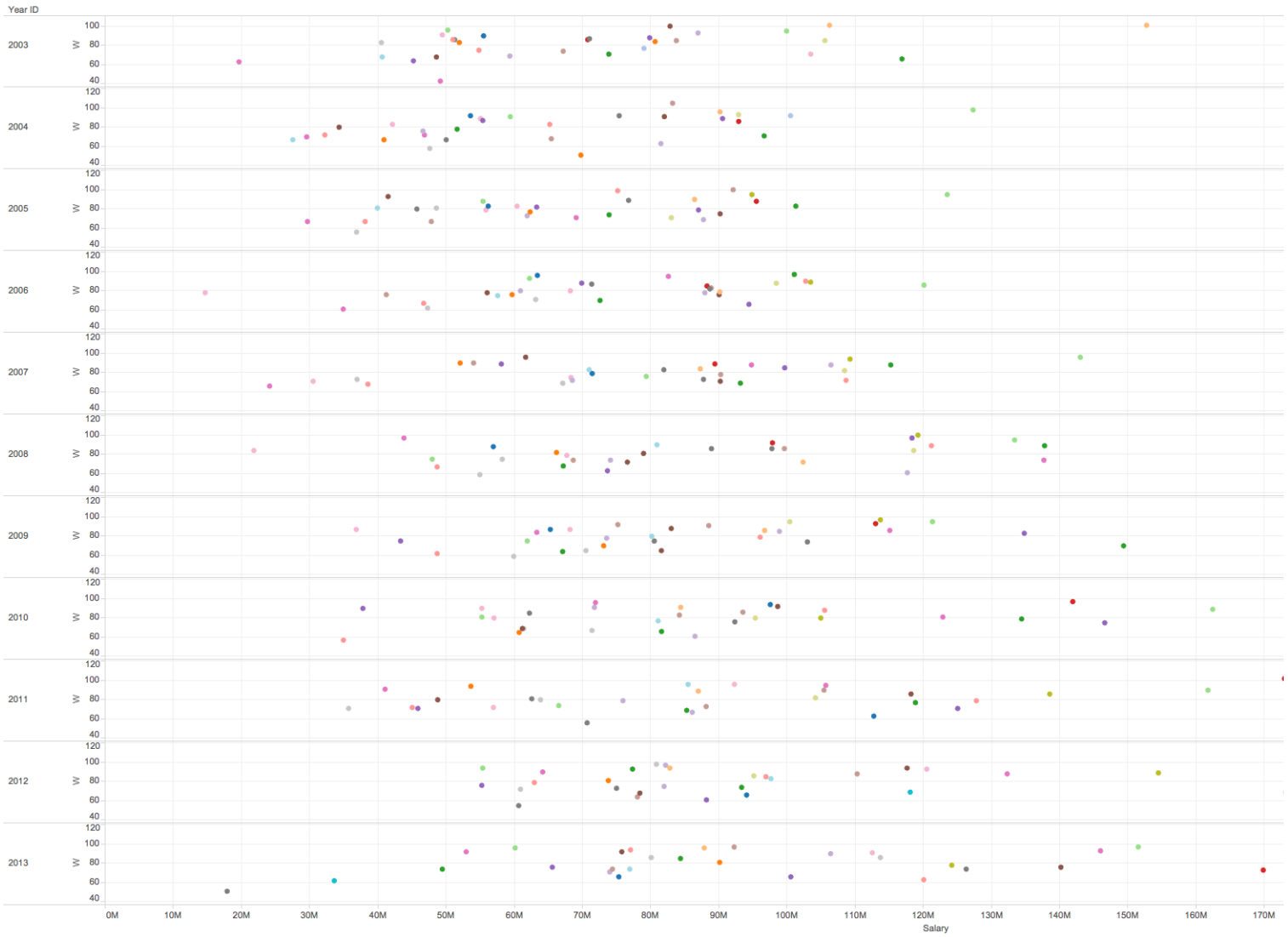
Process

Salary VS Wins in baseball since 1985



The first visualization I figured I could do was plot every team since 1985 according to their salary and wins that year. There were a few problems with this. First : Since salaries have increased drastically since 1985 it causes basically a big cluster of data, and I realized I have to compare teams by the year they were in. Second : Some years (like 1994) there were players strikes so less games ended up getting played. This could also skew the data.

Salary VS Wins in baseball since 1985

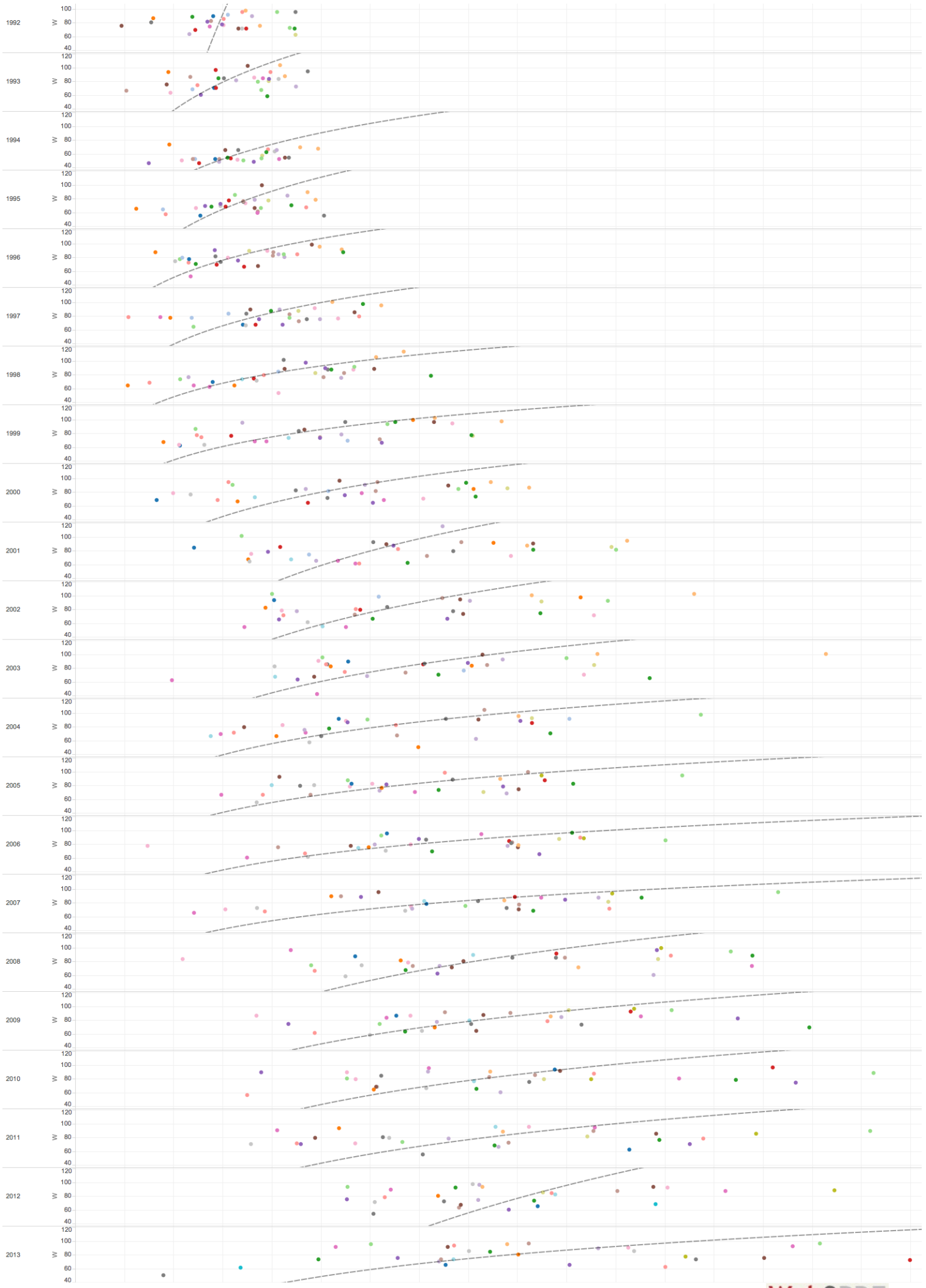


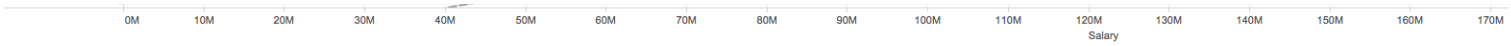
Now I just made a chart with each year being a seperate graphic. Because the x axis is so long, it's hard to see if salary is correlated with wins. I figured I had two options here : make the x-axis shorter, or show trend lines. I decided to show trend lines as I also wanted to highlight how much salaries have grown in the past 30 years.

Main Visualization

Salary VS Wins in baseball since 1985

Sheet 1





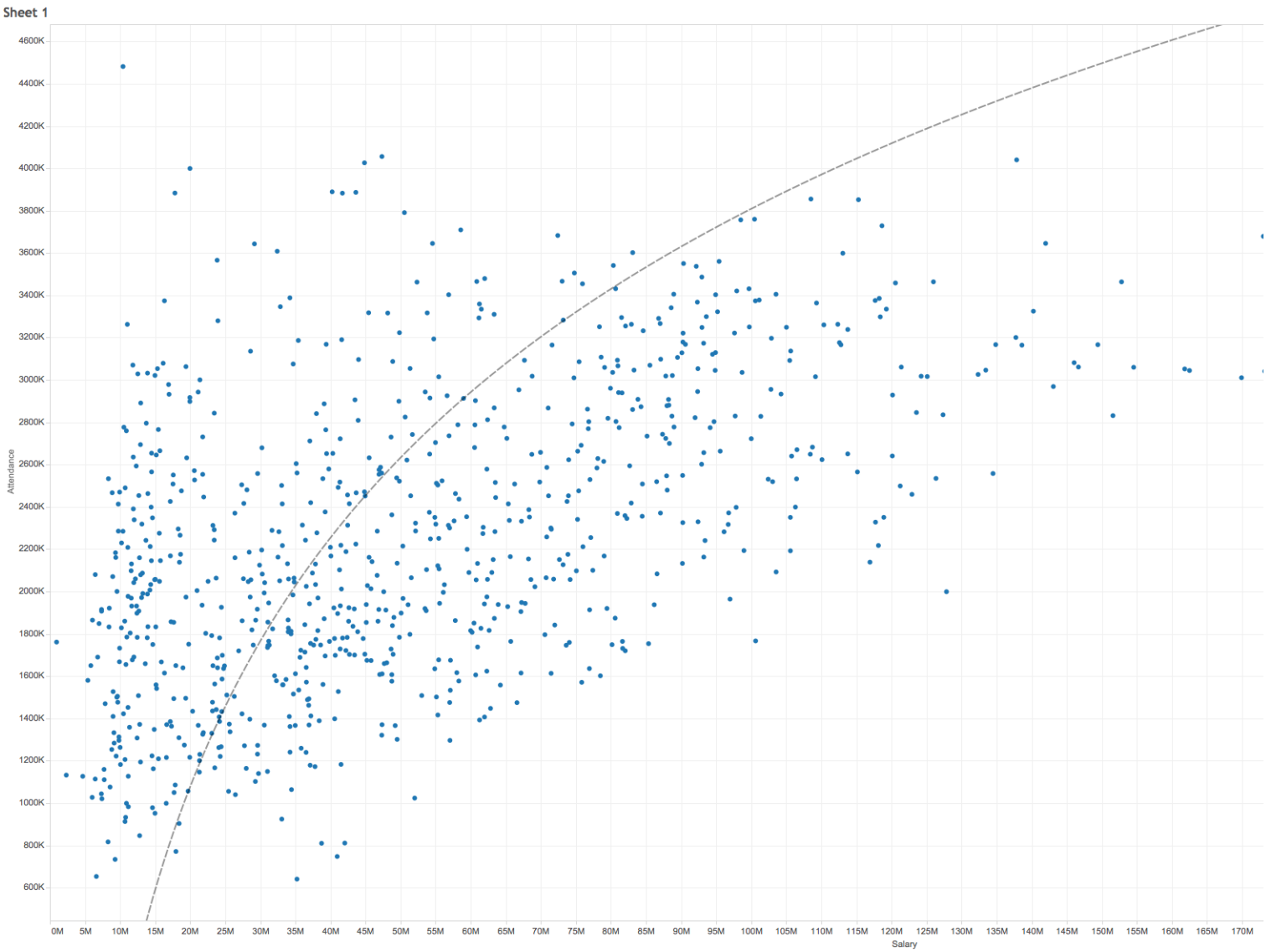
The trend of sum of Salary for W broken down by Year ID. Color shows details about Team ID.

Adding trendlines showed that salary infact does have a huge effect on wins, post 1992-1994. This is really interesting as in 1994 there was a players strike which destroyed any effective baseball salary cap and causes wages to skyrocket, as baseball players felt like they weren't being payed enough. Another interesting thing to note here is if you've even seen the movie Moneyball, the light green data point in 2000-2001 is about that Oakland A's team. This team recorded a huge amount of wins with a very low salary by using statistics (and Jonah Hill) to their advantage

More visualizations

Now I wanted to answer other questions. How much does a baseball teams salary correlate with their attendance?

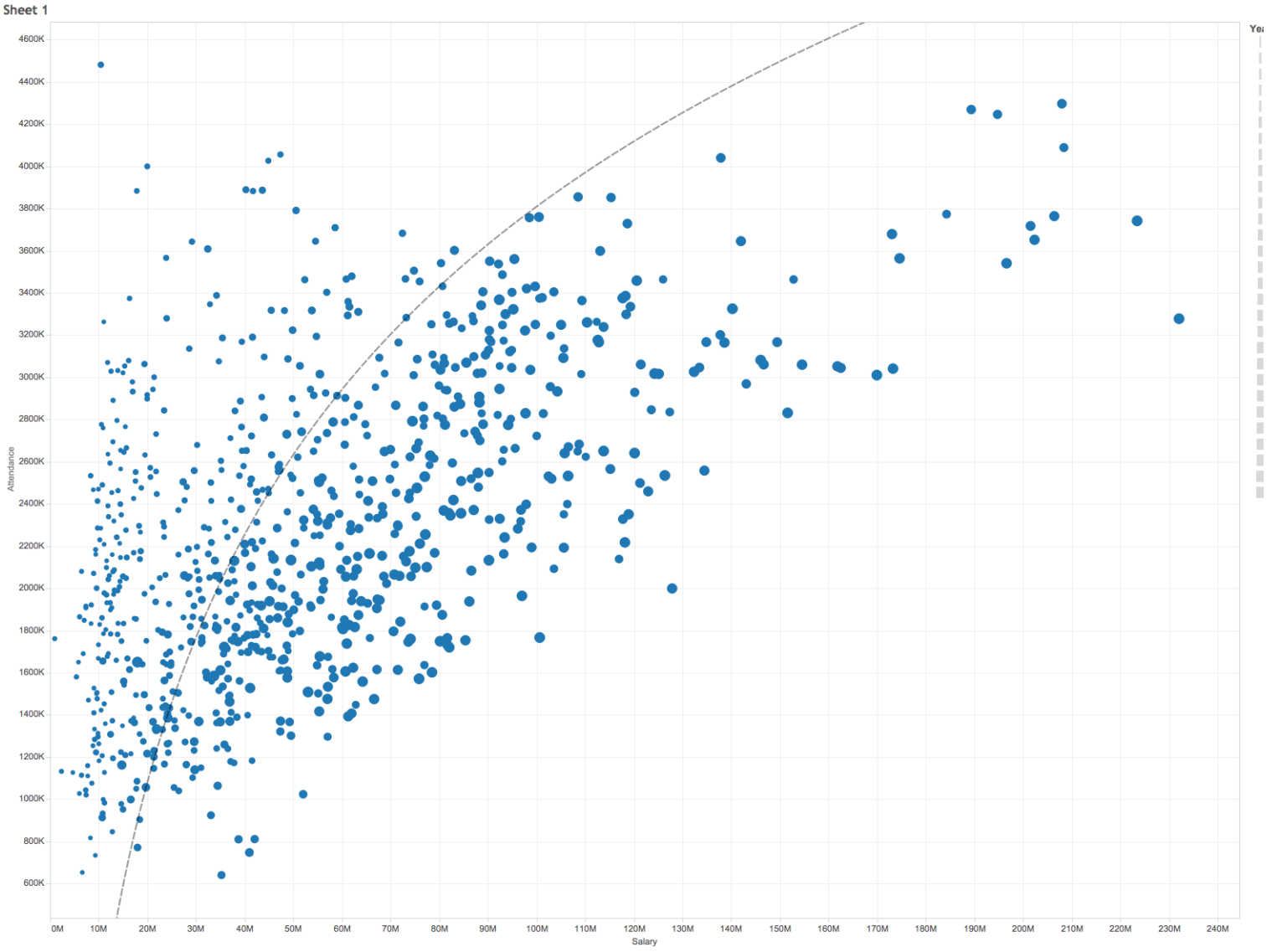
Salary VS Attendance by team in baseball since 1985



The trend of sum of Salary for Attendance. Details are shown for Year ID and Team ID.

Same issue here as before. Plotting every game on the same graph is strange because salaries have drastically increased.

Salary VS Attendance by team in baseball since 1985

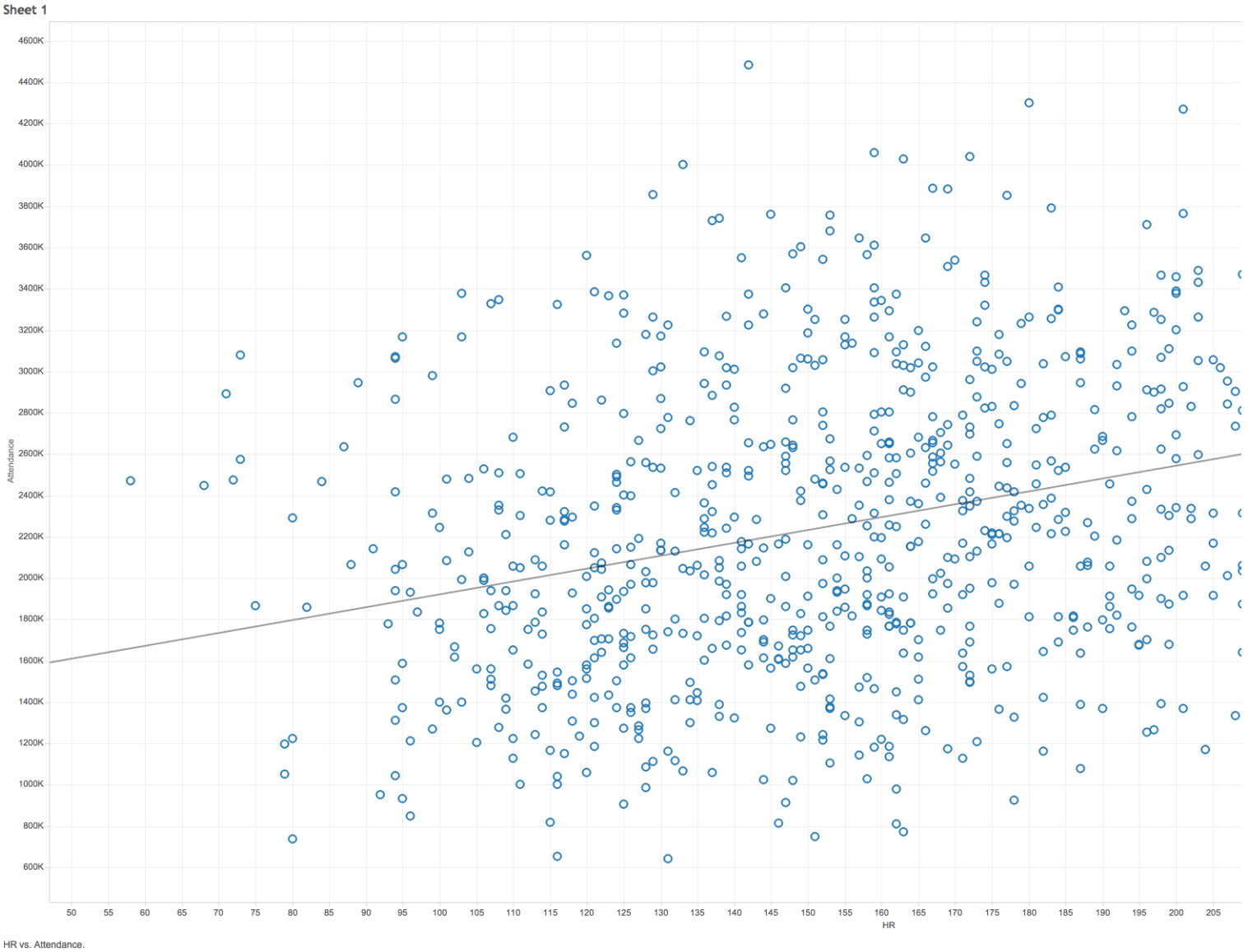


The trend of sum of Salary for Attendance. Size shows details about Year ID. Details are shown for Team ID.

Instead of creating a bunch of different charts, I thought about what other variables I could change to visually encode the year. I ended up choosing size/area based on the fact year was a quantitative variable in this case, position was taken, and length angle and slope didn't seem applicable to a scatterplot

It seems pretty clear that attendance and salary are correlated, but the real reason probably has to do more with something like wins. Maybe it has to do with the most exciting part of baseball, homeruns.

Homeruns VS Attendance by team in baseball since 1985

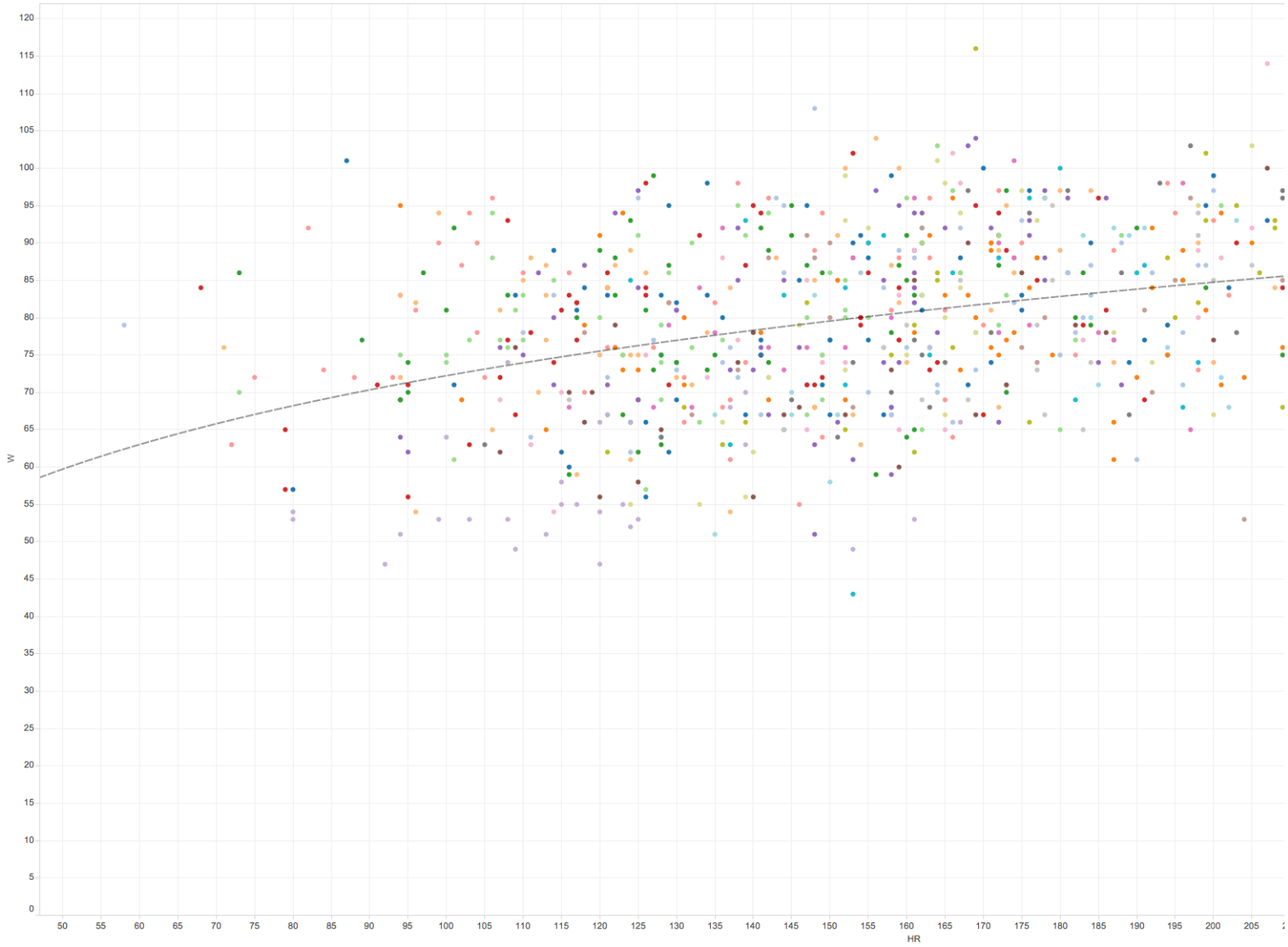


HR vs. Attendance.

There seems to be a very slight correlation between homeruns and attendance, but not as much as I expected. In retrospect, I wish I had encoded the year as the size of each point so I could highlight the steroid era of baseball when there were way more home runs.

Homeruns VS Wins by team in baseball since 1985

Sheet 1



The trend of sum of W for HR. Color shows details about Year ID. Details are shown for Team ID.

Finally, I wanted to see how much homeruns correlated with wins. From what I can tell, it seems like once you hit a very basic number of Homeruns (like 125ish) hitting more stops mattering as much