---

title: Scatterplot

author: Maria Williams

date: '2017-10-05'

slug: scatterplot

categories: []

tags:

- data science

- scatterplot

---

Today we will learn how to visualize our data using scatter plots. The first thing we need to do is tell R studio where we are getting our data from. We need to bring in the Lahman Baseball database, SQLDF, and ggplot2 with library.

After these databases are brought in, we are going to look for some baseball player's career homeruns and strikeouts. The one caught is that they need to have more than 400 homeruns.The coding is as follows:

```{r}

query<-"SELECT playerID,sum(HR),sum(SO)

FROM Batting

GROUP BY playerID

HAVING sum(HR)>400

"

sqldf(query)

```

Yay! We have gathered all of the data needed to make this scatter plot!To make or lives easier, let's clean up the information and save it as result.

```{r}

query<-"SELECT playerID,sum(HR) AS CareerHR,sum(SO) AS CareerSO

FROM Batting

GROUP BY playerID

HAVING sum(HR)>400"

result<-sqldf(query)