Lab Part I: Baseball Names:

One source of data for this lab is the public `Lahman` database which contains a number of data sets with different units of observation. Below are the first few rows and some of the columns for two of these data sets: `Teams` and `Batting`. They contain data going back to 1871. Use these excerpts to help you answer the following questions.

yearID	teamID	franchID	G	W	L	R	RA	name
2014	NYA	NYY	162	84	78	633	664	New York Yankees
2013	WAS	WSN	162	86	76	656	626	Washington Nationals
1997	NYA	NYY	162	96	66	891	688	New York Yankees
1981	PIT	PIT	103	46	56	407	425	Pittsburgh Pirates
1931	NYA	NYY	155	94	59	1067	760	New York Yankees
1928	BSN	ATL	153	50	103	631	878	Boston Braves

playerID	yearID	teamID	G	AB	R	H	BB	SO
hammeja01	2016	CHN	35	65	6	16	1	24
mcewijo01	2002	NYN	105	196	22	39	9	50
jeffcmi01	1985	CLE	9	0	0	0	0	0
richmbe01	1933	CHN	5	1	0	0	0	0
cooneji02	1925	SLN	54	187	27	51	4	5
mcinnst01	1919	BOS	120	440	32	134	23	11

Question 1
What is the unit of observation for the Teams data set? What about for the Batting data set?

Question 2
Write out two questions about baseball that could answered purely through *summaries* of these data sets (numerical summaries or plots).

Question 3
Write out *predictive* questions (two classification and two regression) that you could answer about baseball using the data sets above. Identify a response variable for each.

Question 4
What is a question that we would need more granular (measured on a finer/more specific part of the game) data than the `Teams` and `Batting` data sets provide to answer?

Question 5
Roughly since 1962 MLB Teams have played 162 games in a season. What do you think the distribution of wins (`W`) looks like? Sketch a plot of what you think the *entire* wins column looks like, adding axis tick marks with plausible values, and describe the shape in words.

Question 6
What do you think the relationship between wins (`W`) and runs (`R`) looks like? Sketch a plot, adding axis tick marks with plausible values, and describe the shape in words.

Question 7
Some people believe analytics is ruining baseball because Teams are more cautious which makes the games less entertaining. Do you agree or disagree? Why? Answer in two or more sentences.