## Lahman Baseball Database Exercise

- this data has been made available [online](http://www.seanlahman.com/baseball-archive/statistics/) by Sean Lahman

- you can find a data dictionary [here](http://www.seanlahman.com/files/database/readme2016.txt)

### Use SQL queries to find answers to the \*Initial Questions\*. If time permits, choose one (or more) of the \*Open-Ended Questions\*. Toward the end of the bootcamp, we will revisit this data if time allows to combine SQL, Excel Power Pivot, and/or Python to answer more of the \*Open-Ended Questions\*.

\*\*Initial Questions\*\*

1. What range of years for baseball games played does the provided database cover?

2. Find the name and height of the shortest player in the database. How many games did he play in? What is the name of the team for which he played?

3. Find all players in the database who played at Vanderbilt University. Create a list showing each player’s first and last names as well as the total salary they earned in the major leagues. Sort this list in descending order by the total salary earned. Which Vanderbilt player earned the most money in the majors?

4. Using the fielding table, group players into three groups based on their position: label players with position OF as "Outfield", those with position "SS", "1B", "2B", and "3B" as "Infield", and those with position "P" or "C" as "Battery". Determine the number of putouts made by each of these three groups in 2016.

5. Find the average number of strikeouts per game by decade since 1920. Round the numbers you report to 2 decimal places. Do the same for home runs per game. Do you see any trends?

6. Find the player who had the most success stealing bases in 2016, where \_\_success\_\_ is measured as the percentage of stolen base attempts which are successful. (A stolen base attempt results either in a stolen base or being caught stealing.) Consider only players who attempted \_at least\_ 20 stolen bases.

7. From 1970 – 2016, what is the largest number of wins for a team that did not win the world series? What is the smallest number of wins for a team that did win the world series? Doing this will probably result in an unusually small number of wins for a world series champion – determine why this is the case. Then redo your query, excluding the problem year. How often from 1970 – 2016 was it the case that a team with the most wins also won the world series? What percentage of the time?

8. Using the attendance figures from the homegames table, find the teams and parks which had the top 5 average attendance per game in 2016 (where average attendance is defined as total attendance divided by number of games). Only consider parks where there were at least 10 games played. Report the park name, team name, and average attendance. Repeat for the lowest 5 average attendance.

9. Which managers have won the TSN Manager of the Year award in both the National League (NL) and the American League (AL)? Give their full name and the teams that they were managing when they won the award.

10. Find all players who hit their career highest number of home runs in 2016. Consider only players who have played in the league for at least 10 years, and who hit at least one home run in 2016. Report the players' first and last names and the number of home runs they hit in 2016.

\*\*Open-ended questions\*\*

11. 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.

12. In this question, you will explore the connection between number of wins and attendance.

<ol type="a">

<li>Does there appear to be any correlation between attendance at home games and number of wins? </li>

<li>Do teams that win the world series see a boost in attendance the following year? What about teams that made the playoffs? Making the playoffs means either being a division winner or a wild card winner.</li>

</ol>

13. It is thought that since left-handed pitchers are more rare, causing batters to face them less often, that they are more effective. Investigate this claim and present evidence to either support or dispute this claim. First, determine just how rare left-handed pitchers are compared with right-handed pitchers. Are left-handed pitchers more likely to win the Cy Young Award? Are they more likely to make it into the hall of fame?