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

SELECT MIN(debut)

FROM people; **1871-05-04**

SELECT MAX(debut)

FROM people; **2017-04-03**

1. 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?

SELECT

namegiven,

height

FROM people

order by height asc; **Edward Carl w/ playerid gaedeed01- (there are 2 Edward Carl’s!!)** **He was 43 inches tall! 3.5 feet**

SELECT \*

FROM people

WHERE namegiven = 'Edward Carl';

He Played one game on 08-19-1951.

SELECT \*

FROM APPEARANCES

WHERE playerid ='gaedeed01';

SELECT \*

FROM teams

WHERE teamid = 'SLA'; **Which is the St. Louis Browns**

1. 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? **David Taylor**

SELECT

c.schoolid,

s.salary,

p.namegiven

FROM collegeplaying as c

LEFT JOIN salaries as s

ON c.playerid = s.playerid

JOIN people as p

ON p.playerid = c.playerid

WHERE c.schoolid = 'vandy'

AND s.salary IS NOT NULL

ORDER BY salary DESC;

1. 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.
2. 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?

WITH sohr\_game as

(SELECT

FLOOR(yearid/10) \* 10 AS decade,

ROUND(AVG(soa+so)/2,2) AS avg\_strikeouts,

ROUND(AVG(hr),2) AS avg\_homeruns

FROM teams

GROUP BY decade)

SELECT \*

FROM sohr\_game

WHERE decade >= 1920

ORDER BY decade ASC;

1. 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.

SELECT

b.playerid,

p.namegiven,

b.sb,

b.cs,

ROUND((b.cs/b.sb)\*100,2) AS success\_rate

FROM batting AS b

JOIN people AS p

ON b.playerid = p.playerid

WHERE yearid = 2016 AND sb > 20

1. 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?
2. 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.