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

SELECT MIN(year) AS starting_year,
      MAX(year) AS ending_year
FROM homegames;

/* 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? */

select namefirst || ' ' || namelast as full_name,
      height,
      a.teamid,
      t.name,
      a.G_all,
      p.playerid
from people as p
join appearances as a using(playerid)
join teams as t using(teamid)
group by full_name, height, a.teamid, t.name, a.G_all, p.playerid
order by height
limit 1;

/* 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? */

select distinct namefirst || ' ' || namelast as full_name,
            cp.playerid,
            schoolid,
            sum(salary) as total_salary
from collegeplaying as cp
join people as p using(playerid)
join salaries as s using(playerid)
where schoolid = 'vandy'
group by cp, full_name, playerid, schoolid
-- you can apparently group by (collapse) whole tables
-- who knew?
order by total_salary desc;

/* 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. */

select sum(po) as total_putouts,
	case
		when pos = 'OF' then 'Outfield'
		when pos = '1B' or pos = '2B' or pos = 'SS' or pos = '3B' then 'Infield'
		else 'Battery'
	end as position_group
from fielding
where yearid = '2016'
group by position_group;

/* 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? */

WITH decades AS (
		select yearid,
			g,
			so,
			hr,
			case
				when yearid between 1920 and 1929 then '1920s'
				when yearid between 1930 and 1939 then '1930s'
				when yearid between 1940 and 1949 then '1940s'
				when yearid between 1950 and 1959 then '1950s'
				when yearid between 1960 and 1969 then '1960s'
				when yearid between 1970 and 1979 then '1970s'
				when yearid between 1980 and 1989 then '1980s'
				when yearid between 1990 and 1999 then '1990s'
				when yearid between 2000 and 2009 then '2000s'
				when yearid between 2010 and 2019 then '2010s'
			end as decade
		from teams
		where yearid >= 1920
	)
select decade,
	sum(so) as total_k,
	sum(g) as total_g,
	sum(hr) as total_hr,
	round(sum(so::numeric)/sum(g), 2) as k_per_g,
	round(sum(hr::numeric)/sum(g), 2) as hr_per_g
from decades
```

```
group by decade
order by decade;

/* 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. */

with sb_attempts as (
                select playerid,
                      yearid,
                      sb,
                      cs,
                      round((sb::numeric/(sb+cs))*100, 2) || '%' as sb_percentage
                from batting
                where sb+cs >= 20
                and yearid = 2016
               )
select namefirst||' '||namelast as full_name,
      yearid,
      sb,
      cs,
      sb_percentage
from sb_attempts join people using(playerid)
order by sb_percentage desc;

/* 7.  From 1970 – 2016, what is the largest number of wins for a team that did not win the world
series? --SEA, 116 wins
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? */

--d
with max_w_data as (
                select yearid,
                      max(w) as max_w
                from teams
                where yearid between 1970 and 2016
                group by yearid
                order by yearid
```

```
                )
select sum(case when wswin = 'Y' then 1 else 0 end) as maxw_wsw,
--       count(wswin) as total_years,
      concat(round(100*avg(case when wswin = 'Y' then 1 else 0 end), 2), '%') as
maxw_wsw_pct
--select max_w, wswin, teams.yearid
from max_w_data
join teams
    on max_w_data.yearid = teams.yearid
    and max_w_data.max_w = teams.w
where wswin is not null;


/* 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. */

WITH avg_att AS (SELECT (SUM(attendance)/games) AS avg_attendance, park, team, year
                                            FROM homegames
                                             WHERE games >= 10
                                             GROUP BY games, park, team, year)

SELECT parks.park_name, team, avg_attendance
FROM avg_att INNER JOIN parks
ON parks.park = avg_att.park
WHERE year = 2016
ORDER BY avg_attendance DESC --add/remove DESC to see highest and lowest avg
LIMIT 5;

/* 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. */

select namefirst || ' ' || namelast as full_name,
       name as team_name,
       awardsmanagers.lgid as league,
       awardsmanagers.yearid as year,
       awardid as award
from awardsmanagers
join people
    using(playerid)
```

```
join managers
    on managers.yearid = awardsmanagers.yearid
    and managers.playerid = awardsmanagers.playerid
join teams
    on teams.teamid = managers.teamid
    and teams.yearid = managers.yearid
where awardid ilike 'tsn%'
and awardsmanagers.playerid in
                        (
                         select playerid
                         from awardsmanagers
                         where awardid ilike 'tsn%'
                         and lgid = 'AL'
                         intersect
                         select playerid
                         from awardsmanagers
                         where awardid ilike 'tsn%'
                         and lgid = 'NL'
                        )
order by full_name, year desc;

/* 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. */

WITH player_max AS(
        SELECT
                playerid,
                MAX(hr) AS max_hr
        FROM batting
        GROUP BY playerid
        )

SELECT
        namefirst||' '||namelast AS full_name,
        yearid AS year,
        SUM(hr) AS homeruns
FROM people
        INNER JOIN batting USING (playerid)
        INNER JOIN player_max USING (playerid)
WHERE 2016 - EXTRACT(year FROM debut::date) >= 10
        AND hr > 0
        AND yearid = 2016
```

```
        AND max_hr = hr
GROUP BY full_name, yearid, max_hr
ORDER BY max_hr DESC;
```

Bonus

/* 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. */

```
--a
with
team_data as (
                select yearid as year,
                      teamid as team,
                      w as wins,
                      sum(salary::numeric::money) as team_salary
                from salaries join teams using(yearid, teamid)
                where yearid >= 2000
                group by year,
                        team,
                        wins
                order by year,
                        team
            ),
rank_data as (
            select rank() over(partition by year order by team_salary desc) as salary_rank,
                  rank() over(partition by year order by wins desc) as wins_rank
            from team_data
            )
select salary_rank,
      round(avg(wins_rank), 2) as avg_wins_rank
from rank_data
group by salary_rank
order by salary_rank;

--b
with
home_att_data as (
                select yearid as year,
                      attendance/ghome as avg_home_att,
                      w as wins
                from teams
```

```
                where attendance/ghome is not null
                and yearid >= 1998 -- 29th and 30th team added to the MLB
               ),
rank_data as (
            select rank() over(partition by year order by avg_home_att desc) as aha_rank,
                   rank() over(partition by year order by wins desc) as wins_rank
            from home_att_data
           )
select aha_rank,
      round(avg(wins_rank), 2) as avg_wins_rank
from rank_data
group by aha_rank
order by aha_rank;

/* 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? */
```