STA141B Homework 3

Zeng Fung Liew

11/7/2020

Question 1

What years does the data cover? Are there data for each of these years?

To approach this problem, I used the same query statement (shown below) for each table. The reason I decided not to join all the tables and get a single row in my resulting data frame is because I want to make sure I know the years covered in each table. In other words, if I join all tables that contain `yearID` and find that the earliest year available is X and the latest year available is Y, I cannot be confident that that is the case for every table, and that every table contains data for each of the years between X and Y.

In order to preserve space, the query below is only shown once in this report since the other queries are identical. The full code can be found in the R script attached to the report.

- Code to obtain query:

```
# get years data from batting table
numyears_batting = dbGetQuery(db,
    "SELECT
      'Batting' AS table_name,
      MIN(yearID) AS min_year,
      MAX(yearID) AS max_year,
      (MAX(yearID) - MIN(yearID) + 1) AS years_between,
      COUNT(DISTINCT(yearID)) AS distinct_years
    FROM
      Batting;")
```

```
##             table_name min_year max_year years_between distinct_years
## 1              Batting     1871     2013           143            143
## 2             Pitching     1871     2013           143            143
## 3             Fielding     1871     2013           143            143
## 4          AllstarFull     1933     2013            81             81
## 5           HallofFame     1936     2014            79             76
## 6             Managers     1871     2013           143            143
## 7                Teams     1871     2013           143            143
## 8          BattingPost     1884     2013           130            117
## 9         PitchingPost     1884     2013           130            117
## 10          FieldingOF     1871     1955            85             85
## 11        FieldingPost     1903     2013           111            109
## 12        ManagersHalf     1892     1981            90              2
## 13           TeamsHalf     1981     1981             1              1
## 14            Salaries     1985     2013            29             29
## 15          SeriesPost     1884     2013           130            117
```

```
## 16      AwardsManagers     1936     2013           78            78
## 17       AwardsPlayers     1877     2013          137           115
## 18 AwardsShareManagers     1983     2013           31            31
## 19  AwardsSharePlayers     1911     2013          103            95
## 20         Appearances     1871     2013          143           143
```

As we can see in the above data frame, most of the tables contain data from each of the years between 1871 to 2013, with the following exceptions:

- `AllstarFull`, `Salaries`, `AwardsManagers`, and `AwardsShareManagers` tables started collecting data in 1933, 1985, 19636 and 1983 respectively.
- `HallofFame` table started collecting data in 1936, and contains the data up until the year 2014. However, the table is missing of 3 years of data.
- `BattingPost`, `FieldingPost`, and `SeriesPost` all started collecting data in 1884, and all three tables are missing of 13 years of data.
- `FieldingOF` table started stop collecting data in the year 1955.
- `FieldingPost` started collecting data in 1903. However, the table is missing of 2 years of data.
- The `ManagersHalf` and `TeamsHalf` tables are quite interesting. The former only contains data from the years 1892 and 1981, while the latter only contains data from 1981.
- `AwardsPlayers` table started collecting data in 1877, but is missing of 22 years of data in between.
- `AwardsSharePlayers` table started collecting data in 1911, but is missing of 8 years of data in between.

Question 2

How many (unique) people are included in the database? How many are players, managers, etc?

To count the number of unique people, I used the `COUNT(DISTINCT(playerID))` in my query. However, getting the number of unique people in the `Master` table isn't enough. Therefore, to get more information about the number of players in each position (Hitter, Pitcher, Fielder, and Manager), I performed similar queries on the other tables. On top of that, I also added `JOIN` statements in order to find out the number of players who have multiple roles in the team. I also used the `UNION` keyword in order to join the rows I obtain from each query and avoid having to consistently declare new variables.

```
dbGetQuery(db,
    "SELECT
       COUNT(DISTINCT(playerID)) ppl_count
     FROM
       Master;")
```

```
##   ppl_count
## 1     18354
```

From the above query, we find that there are 18,354 unique people in the database. We now want to dive deeper to find out if each unique person has a specific role, or that each person can carry multiple roles.

```
dbGetQuery(db,
    "SELECT
       1 AS hitter,
       0 AS pitcher,
       0 AS fielder,
       COUNT(DISTINCT(playerID)) AS ppl_count
     FROM
       Batting

     UNION

     SELECT
       0 AS hitter,
       1 AS pitcher,
       0 AS fielder,
       COUNT(DISTINCT(playerID)) AS ppl_count
     FROM
       Pitching

     UNION

     SELECT
       0 AS hitter,
       0 AS pitcher,
       1 AS fielder,
       COUNT(DISTINCT(playerID)) AS ppl_count
     FROM
       Fielding

     UNION

     SELECT
       1 AS hitter,
       0 AS pitcher,
       1 AS fielder,
       COUNT(DISTINCT(Batting.playerID)) AS ppl_count
     FROM
       Batting
     INNER JOIN
       Fielding
     ON
       Batting.playerID = Fielding.playerID

     UNION

     SELECT
       0 AS hitter,
       1 AS pitcher,
       1 AS fielder,
       COUNT(DISTINCT(Pitching.playerID)) AS ppl_count
     FROM
       Pitching
     INNER JOIN
```

```
    Fielding
ON
    Pitching.playerID = Fielding.playerID

UNION

SELECT
    1 AS hitter,
    1 AS pitcher,
    0 AS fielder,
    COUNT(DISTINCT(Batting.playerID)) AS ppl_count
FROM
    Batting
INNER JOIN
    Pitching
ON
    Batting.playerID = Pitching.playerID

UNION

SELECT
    1 AS hitter,
    1 AS pitcher,
    1 AS fielder,
    COUNT(DISTINCT(Batting.playerID)) AS ppl_count
FROM
    Batting
INNER JOIN
    Pitching
ON
    Batting.playerID = Pitching.playerID
INNER JOIN
    Fielding
ON
    Batting.playerID = Fielding.playerID
ORDER BY
    ppl_count DESC;")
```

```
##   hitter pitcher fielder ppl_count
## 1      1       0       0     18107
## 2      0       0       1     17981
## 3      1       0       1     17918
## 4      0       1       0      8810
## 5      0       1       1      8810
## 6      1       1       0      8747
## 7      1       1       1      8747
```

From this query, we get the following findings (by constructing a Venn Diagram):

- There are 17918 players who play as a Hitter and Fielder, among which 8747 of those players also took up the role as a pitcher.
- Only 63 players play exclusively as a Fielder and a Pitcher but not a Hitter.

- There are 189 Hitters who do not have any other roles. However, there are no players who have sole roles of either as a Pitcher or as a Fielder.

```
dbGetQuery(db,
    "SELECT
        'Manager' AS role,
        COUNT(DISTINCT(playerID)) AS ppl_count
     FROM
        ManagerS;")
```

```
##      role ppl_count
## 1 Manager       682
```

On the other hand, there are 682 managers in the database.

Question 3

How many players became managers?

To approach this problem, I made sure to set the result of the query to be `GROUP BY plyrMgr`.

```
dbGetQuery(db,
    "SELECT
        (CASE
          WHEN plyrMgr = 'Y' THEN 'Player-Manager'
          ELSE 'Manager'
        END) AS role,
        COUNT(DISTINCT(Managers.playerID)) AS ppl_count
     FROM
        Managers
     INNER JOIN
        Master
     ON
        Managers.playerID = Master.playerID
     GROUP BY
        plyrMgr;")
```

```
##             role ppl_count
## 1        Manager       512
## 2 Player-Manager       247
```

Here, we can see that about one-third of the Managers are actually Player-Managers.

Problem 4

How many players are there in each year, from 2000 to 2013? Do all teams have the same number of players?

Once again, the `GROUP BY` clause is very important in solving this problem.

Based on the query below, we find that there are around 1250 players on average each year.

```
dbGetQuery(db,
  "SELECT
     yearID AS year,
     COUNT(DISTINCT(playerID)) AS player_count
   FROM
     Appearances
   WHERE
     yearID >= 2000
   GROUP BY
     yearID;")
```

```
##    year player_count
## 1  2000         1230
## 2  2001         1220
## 3  2002         1218
## 4  2003         1230
## 5  2004         1247
## 6  2005         1237
## 7  2006         1242
## 8  2007         1278
## 9  2008         1291
## 10 2009         1266
## 11 2010         1249
## 12 2011         1295
## 13 2012         1284
## 14 2013         1304
```

However, we can see that not all teams have the same number of players.

```
# Number of players per team each year
dbGetQuery(db,
  "SELECT
     Appearances.yearID AS year,
     Teams.name AS team,
     COUNT(DISTINCT(Appearances.playerID)) AS player_count
   FROM
     Appearances
   JOIN
     Teams
   ON
     Teams.teamID = Appearances.teamID AND
     Teams.yearID = Appearances.yearID
   WHERE
     Appearances.yearID >= 2000
   GROUP BY
     Appearances.yearID,
     Appearances.teamID
   LIMIT
     15;")
```

```
##    year                 team player_count
## 1  2000       Anaheim Angels           45
```

```
## 2  2000 Arizona Diamondbacks          41
## 3  2000       Atlanta Braves          47
## 4  2000    Baltimore Orioles          50
## 5  2000       Boston Red Sox          52
## 6  2000    Chicago White Sox          42
## 7  2000         Chicago Cubs          51
## 8  2000      Cincinnati Reds          45
## 9  2000    Cleveland Indians          55
## 10 2000     Colorado Rockies          49
## 11 2000       Detroit Tigers          45
## 12 2000      Florida Marlins          40
## 13 2000       Houston Astros          46
## 14 2000   Kansas City Royals          44
## 15 2000  Los Angeles Dodgers          46
```

```
# Number of players per team overall
dbGetQuery(db,
  "SELECT
     Teams.name AS team,
     COUNT(DISTINCT(Appearances.playerID)) AS player_count
   FROM
     Appearances
   JOIN
     Teams
   ON
     Teams.yearID = Appearances.yearID AND
     Teams.teamID = Appearances.teamID
   WHERE
     Appearances.yearID >= 2000
   GROUP BY
     Appearances.teamID
   LIMIT
     15;")
```

```
##                            team player_count
## 1                Anaheim Angels          103
## 2           Arizona Diamondbacks          321
## 3                Atlanta Braves          299
## 4             Baltimore Orioles          337
## 5                Boston Red Sox          366
## 6             Chicago White Sox          255
## 7                  Chicago Cubs          325
## 8               Cincinnati Reds          313
## 9             Cleveland Indians          322
## 10             Colorado Rockies          353
## 11               Detroit Tigers          287
## 12              Florida Marlins          268
## 13               Houston Astros          293
## 14           Kansas City Royals          326
## 15 Los Angeles Angels of Anaheim          164
```

Question 5

What team won the World Series in 2010? Include the name of the team, the league and division.

The problem faced when searching for the answer to this question is that the data needed is split into two different table, `SeriesPost` and `Teams`. Therefore, a `JOIN` clause is needed.

To ensure that we are getting the league that the World Series winner is in correctly, we join the two tables on two conditions, ie. `SeriesPost.teamIDwinner = Teams.teamID` and `SeriesPost.yearID = Teams.yearID`.

As for the answer to this question, San Francisco Giants won the World Series in 2010, as shown in the table below.

```
dbGetQuery(db,
    "SELECT
        SP.yearID AS year,
        SP.round,
        SP.teamIDwinner AS team,
        T.name AS team_name,
        SP.lgIDwinner AS league,
        T.divID AS division
     FROM
        SeriesPost AS SP
     LEFT JOIN
        Teams AS T
     ON
        SP.teamIDwinner = T.teamID AND
        SP.yearID = T.yearID
     WHERE
        SP.yearID = 2010 AND
        SP.round = 'WS';")
```

```
##   year round team            team_name league division
## 1 2010    WS  SFN San Francisco Giants     NL        W
```

Question 6

What team lost the World Series each year? Again, include the name of the team, league and division.

The query used in this problem is very similar to the previous one (Question 5). The only difference is that we are looking for the losers instead of the winners, and that we want the data for every single year instead of just a single year.

The following table is the result from the query. Notice that the `Teams` table did not contain data of some of the losing teams prior to 1890. Additionally, there aren't any established divisions prior to the year 1969.

```
dbGetQuery(db,
    "SELECT
        SP.yearID AS year,
        SP.round,
        SP.teamIDloser AS team,
        T.name AS team_name,
```

```
    SP.lgIDloser AS league,
    T.divID AS division
  FROM
    SeriesPost AS SP
  LEFT JOIN
    Teams AS T
  ON
    SP.teamIDloser = T.teamID AND
    SP.yearID = T.yearID
  WHERE
    SP.round = 'WS';")
```

```
##     year round team              team_name league division
## 1   1884    WS  NYP                   <NA>     AA     <NA>
## 2   1885    WS  STL                   <NA>     AA     <NA>
## 3   1886    WS  CHC                   <NA>     NL     <NA>
## 4   1887    WS  STL                   <NA>     AA     <NA>
## 5   1888    WS  STL                   <NA>     AA     <NA>
## 6   1889    WS  BRO                   <NA>     AA     <NA>
## 7   1890    WS  LOU                   <NA>     AA     <NA>
## 8   1903    WS  PIT     Pittsburgh Pirates     NL     <NA>
## 9   1905    WS  PHA Philadelphia Athletics     AL     <NA>
## 10  1906    WS  CHN           Chicago Cubs     NL     <NA>
## 11  1907    WS  DET         Detroit Tigers     AL     <NA>
## 12  1908    WS  DET         Detroit Tigers     AL     <NA>
## 13  1909    WS  DET         Detroit Tigers     AL     <NA>
## 14  1910    WS  CHN           Chicago Cubs     NL     <NA>
## 15  1911    WS  NY1        New York Giants     NL     <NA>
## 16  1912    WS  NY1        New York Giants     NL     <NA>
## 17  1913    WS  NY1        New York Giants     NL     <NA>
## 18  1914    WS  PHA Philadelphia Athletics     AL     <NA>
## 19  1915    WS  PHI  Philadelphia Phillies     NL     <NA>
## 20  1916    WS  BRO        Brooklyn Robins     NL     <NA>
## 21  1917    WS  NY1        New York Giants     NL     <NA>
## 22  1918    WS  CHN           Chicago Cubs     NL     <NA>
## 23  1919    WS  CHA      Chicago White Sox     AL     <NA>
## 24  1920    WS  BRO        Brooklyn Robins     NL     <NA>
## 25  1921    WS  NYA       New York Yankees     AL     <NA>
## 26  1922    WS  NYA       New York Yankees     AL     <NA>
## 27  1923    WS  NY1        New York Giants     NL     <NA>
## 28  1924    WS  NY1        New York Giants     NL     <NA>
## 29  1925    WS  WS1    Washington Senators     AL     <NA>
## 30  1926    WS  NYA       New York Yankees     AL     <NA>
## 31  1927    WS  PIT     Pittsburgh Pirates     NL     <NA>
## 32  1928    WS  SLN    St. Louis Cardinals     NL     <NA>
## 33  1929    WS  CHN           Chicago Cubs     NL     <NA>
## 34  1930    WS  SLN    St. Louis Cardinals     NL     <NA>
## 35  1931    WS  PHA Philadelphia Athletics     AL     <NA>
## 36  1932    WS  CHN           Chicago Cubs     NL     <NA>
## 37  1933    WS  WS1    Washington Senators     AL     <NA>
## 38  1934    WS  DET         Detroit Tigers     AL     <NA>
## 39  1935    WS  CHN           Chicago Cubs     NL     <NA>
## 40  1936    WS  NY1        New York Giants     NL     <NA>
```

```
## 41  1937    WS  NY1        New York Giants     NL     <NA>
## 42  1938    WS  CHN           Chicago Cubs     NL     <NA>
## 43  1939    WS  CIN        Cincinnati Reds     NL     <NA>
## 44  1940    WS  DET         Detroit Tigers     AL     <NA>
## 45  1941    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 46  1942    WS  NYA       New York Yankees     AL     <NA>
## 47  1943    WS  SLN    St. Louis Cardinals     NL     <NA>
## 48  1944    WS  SLA       St. Louis Browns     AL     <NA>
## 49  1945    WS  CHN           Chicago Cubs     NL     <NA>
## 50  1946    WS  BOS         Boston Red Sox     AL     <NA>
## 51  1947    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 52  1948    WS  BSN          Boston Braves     NL     <NA>
## 53  1949    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 54  1950    WS  PHI  Philadelphia Phillies     NL     <NA>
## 55  1951    WS  NY1        New York Giants     NL     <NA>
## 56  1952    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 57  1953    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 58  1954    WS  CLE      Cleveland Indians     AL     <NA>
## 59  1955    WS  NYA       New York Yankees     AL     <NA>
## 60  1956    WS  BRO       Brooklyn Dodgers     NL     <NA>
## 61  1957    WS  NYA       New York Yankees     AL     <NA>
## 62  1958    WS  ML1       Milwaukee Braves     NL     <NA>
## 63  1959    WS  CHA      Chicago White Sox     AL     <NA>
## 64  1960    WS  NYA       New York Yankees     AL     <NA>
## 65  1961    WS  CIN        Cincinnati Reds     NL     <NA>
## 66  1962    WS  SFN   San Francisco Giants     NL     <NA>
## 67  1963    WS  NYA       New York Yankees     AL     <NA>
## 68  1964    WS  NYA       New York Yankees     AL     <NA>
## 69  1965    WS  MIN        Minnesota Twins     AL     <NA>
## 70  1966    WS  LAN    Los Angeles Dodgers     NL     <NA>
## 71  1967    WS  BOS         Boston Red Sox     AL     <NA>
## 72  1968    WS  SLN    St. Louis Cardinals     NL     <NA>
## 73  1969    WS  BAL      Baltimore Orioles     AL        E
## 74  1970    WS  CIN        Cincinnati Reds     NL        W
## 75  1971    WS  BAL      Baltimore Orioles     AL        E
## 76  1972    WS  CIN        Cincinnati Reds     NL        W
## 77  1973    WS  NYN          New York Mets     NL        E
## 78  1974    WS  LAN    Los Angeles Dodgers     NL        W
## 79  1975    WS  BOS         Boston Red Sox     AL        E
## 80  1976    WS  NYA       New York Yankees     AL        E
## 81  1977    WS  LAN    Los Angeles Dodgers     NL        W
## 82  1978    WS  LAN    Los Angeles Dodgers     NL        W
## 83  1979    WS  BAL      Baltimore Orioles     AL        E
## 84  1980    WS  KCA     Kansas City Royals     AL        W
## 85  1981    WS  NYA       New York Yankees     AL        E
## 86  1982    WS  ML4      Milwaukee Brewers     AL        E
## 87  1983    WS  PHI  Philadelphia Phillies     NL        E
## 88  1984    WS  SDN       San Diego Padres     NL        W
## 89  1985    WS  SLN    St. Louis Cardinals     NL        E
## 90  1986    WS  BOS         Boston Red Sox     AL        E
## 91  1987    WS  SLN    St. Louis Cardinals     NL        E
## 92  1988    WS  OAK      Oakland Athletics     AL        W
## 93  1989    WS  SFN   San Francisco Giants     NL        W
## 94  1990    WS  OAK      Oakland Athletics     AL        W
```

```
## 95  1991    WS  ATL         Atlanta Braves     NL        W
## 96  1992    WS  ATL         Atlanta Braves     NL        W
## 97  1993    WS  PHI  Philadelphia Phillies     NL        E
## 98  1995    WS  CLE      Cleveland Indians     AL        C
## 99  1996    WS  ATL         Atlanta Braves     NL        E
## 100 1997    WS  CLE      Cleveland Indians     AL        C
## 101 1998    WS  SDN       San Diego Padres     NL        W
## 102 1999    WS  ATL         Atlanta Braves     NL        E
## 103 2000    WS  NYN          New York Mets     NL        E
## 104 2001    WS  NYA       New York Yankees     AL        E
## 105 2002    WS  SFN   San Francisco Giants     NL        W
## 106 2003    WS  NYA       New York Yankees     AL        E
## 107 2004    WS  SLN    St. Louis Cardinals     NL        C
## 108 2005    WS  HOU         Houston Astros     NL        C
## 109 2006    WS  DET         Detroit Tigers     AL        C
## 110 2007    WS  COL       Colorado Rockies     NL        W
## 111 2008    WS  TBA         Tampa Bay Rays     AL        E
## 112 2009    WS  PHI  Philadelphia Phillies     NL        E
## 113 2010    WS  TEX          Texas Rangers     AL        W
## 114 2011    WS  TEX          Texas Rangers     AL        W
## 115 2012    WS  DET         Detroit Tigers     AL        C
## 116 2013    WS  SLN    St. Louis Cardinals     NL        C
```

Question 7

Compute the table of World Series winners for all years, again with the name of the team, league and division.

This query is once again very similar to the previous one, where I join the `SeriesPost` and `Teams` tables on `teamID` and `yearID`. Once again, we do not have the data for some of the winners prior to the year 1890 and there is not established division prior to the year 1969. The results of the query is shown below:

```
dbGetQuery(db,
    "SELECT
       SP.yearID AS year,
       SP.teamIDwinner AS team,
       T.name AS team_name,
       SP.lgIDwinner AS league,
       T.divID AS division
     FROM
       SeriesPost AS SP
     LEFT JOIN
       Teams AS T
     ON
       SP.teamIDwinner = T.teamID AND
       SP.yearID = T.yearID
     WHERE
       SP.round = 'WS';")
```

```
##     year team              team_name league division
## 1   1884  PRO       Providence Grays     NL     <NA>
## 2   1885  CHC                   <NA>     NL     <NA>
## 3   1886  STL                   <NA>     AA     <NA>
```

```
## 4   1887  DTN     Detroit Wolverines     NL     <NA>
## 5   1888  NYG                    <NA>     NL     <NA>
## 6   1889  NYG                    <NA>     NL     <NA>
## 7   1890  BRO   Brooklyn Bridegrooms     NL     <NA>
## 8   1903  BOS       Boston Americans     AL     <NA>
## 9   1905  NY1        New York Giants     NL     <NA>
## 10  1906  CHA      Chicago White Sox     AL     <NA>
## 11  1907  CHN           Chicago Cubs     NL     <NA>
## 12  1908  CHN           Chicago Cubs     NL     <NA>
## 13  1909  PIT     Pittsburgh Pirates     NL     <NA>
## 14  1910  PHA Philadelphia Athletics     AL     <NA>
## 15  1911  PHA Philadelphia Athletics     AL     <NA>
## 16  1912  BOS         Boston Red Sox     AL     <NA>
## 17  1913  PHA Philadelphia Athletics     AL     <NA>
## 18  1914  BSN          Boston Braves     NL     <NA>
## 19  1915  BOS         Boston Red Sox     AL     <NA>
## 20  1916  BOS         Boston Red Sox     AL     <NA>
## 21  1917  CHA      Chicago White Sox     AL     <NA>
## 22  1918  BOS         Boston Red Sox     AL     <NA>
## 23  1919  CIN        Cincinnati Reds     NL     <NA>
## 24  1920  CLE      Cleveland Indians     AL     <NA>
## 25  1921  NY1        New York Giants     NL     <NA>
## 26  1922  NY1        New York Giants     NL     <NA>
## 27  1923  NYA       New York Yankees     AL     <NA>
## 28  1924  WS1    Washington Senators     AL     <NA>
## 29  1925  PIT     Pittsburgh Pirates     NL     <NA>
## 30  1926  SLN    St. Louis Cardinals     NL     <NA>
## 31  1927  NYA       New York Yankees     AL     <NA>
## 32  1928  NYA       New York Yankees     AL     <NA>
## 33  1929  PHA Philadelphia Athletics     AL     <NA>
## 34  1930  PHA Philadelphia Athletics     AL     <NA>
## 35  1931  SLN    St. Louis Cardinals     NL     <NA>
## 36  1932  NYA       New York Yankees     AL     <NA>
## 37  1933  NY1        New York Giants     NL     <NA>
## 38  1934  SLN    St. Louis Cardinals     NL     <NA>
## 39  1935  DET         Detroit Tigers     AL     <NA>
## 40  1936  NYA       New York Yankees     AL     <NA>
## 41  1937  NYA       New York Yankees     AL     <NA>
## 42  1938  NYA       New York Yankees     AL     <NA>
## 43  1939  NYA       New York Yankees     AL     <NA>
## 44  1940  CIN        Cincinnati Reds     NL     <NA>
## 45  1941  NYA       New York Yankees     AL     <NA>
## 46  1942  SLN    St. Louis Cardinals     NL     <NA>
## 47  1943  NYA       New York Yankees     AL     <NA>
## 48  1944  SLN    St. Louis Cardinals     NL     <NA>
## 49  1945  DET         Detroit Tigers     AL     <NA>
## 50  1946  SLN    St. Louis Cardinals     NL     <NA>
## 51  1947  NYA       New York Yankees     AL     <NA>
## 52  1948  CLE      Cleveland Indians     AL     <NA>
## 53  1949  NYA       New York Yankees     AL     <NA>
## 54  1950  NYA       New York Yankees     AL     <NA>
## 55  1951  NYA       New York Yankees     AL     <NA>
## 56  1952  NYA       New York Yankees     AL     <NA>
## 57  1953  NYA       New York Yankees     AL     <NA>
```

```
## 58  1954  NY1        New York Giants     NL     <NA>
## 59  1955  BRO       Brooklyn Dodgers     NL     <NA>
## 60  1956  NYA       New York Yankees     AL     <NA>
## 61  1957  ML1       Milwaukee Braves     NL     <NA>
## 62  1958  NYA       New York Yankees     AL     <NA>
## 63  1959  LAN    Los Angeles Dodgers     NL     <NA>
## 64  1960  PIT     Pittsburgh Pirates     NL     <NA>
## 65  1961  NYA       New York Yankees     AL     <NA>
## 66  1962  NYA       New York Yankees     AL     <NA>
## 67  1963  LAN    Los Angeles Dodgers     NL     <NA>
## 68  1964  SLN    St. Louis Cardinals     NL     <NA>
## 69  1965  LAN    Los Angeles Dodgers     NL     <NA>
## 70  1966  BAL      Baltimore Orioles     AL     <NA>
## 71  1967  SLN    St. Louis Cardinals     NL     <NA>
## 72  1968  DET         Detroit Tigers     AL     <NA>
## 73  1969  NYN          New York Mets     NL        E
## 74  1970  BAL      Baltimore Orioles     AL        E
## 75  1971  PIT     Pittsburgh Pirates     NL        E
## 76  1972  OAK      Oakland Athletics     AL        W
## 77  1973  OAK      Oakland Athletics     AL        W
## 78  1974  OAK      Oakland Athletics     AL        W
## 79  1975  CIN        Cincinnati Reds     NL        W
## 80  1976  CIN        Cincinnati Reds     NL        W
## 81  1977  NYA       New York Yankees     AL        E
## 82  1978  NYA       New York Yankees     AL        E
## 83  1979  PIT     Pittsburgh Pirates     NL        E
## 84  1980  PHI  Philadelphia Phillies     NL        E
## 85  1981  LAN    Los Angeles Dodgers     NL        W
## 86  1982  SLN    St. Louis Cardinals     NL        E
## 87  1983  BAL      Baltimore Orioles     AL        E
## 88  1984  DET         Detroit Tigers     AL        E
## 89  1985  KCA     Kansas City Royals     AL        W
## 90  1986  NYN          New York Mets     NL        E
## 91  1987  MIN        Minnesota Twins     AL        W
## 92  1988  LAN    Los Angeles Dodgers     NL        W
## 93  1989  OAK      Oakland Athletics     AL        W
## 94  1990  CIN        Cincinnati Reds     NL        W
## 95  1991  MIN        Minnesota Twins     AL        W
## 96  1992  TOR      Toronto Blue Jays     AL        E
## 97  1993  TOR      Toronto Blue Jays     AL        E
## 98  1995  ATL         Atlanta Braves     NL        E
## 99  1996  NYA       New York Yankees     AL        E
## 100 1997  FLO        Florida Marlins     NL        E
## 101 1998  NYA       New York Yankees     AL        E
## 102 1999  NYA       New York Yankees     AL        E
## 103 2000  NYA       New York Yankees     AL        E
## 104 2001  ARI   Arizona Diamondbacks     NL        W
## 105 2002  ANA         Anaheim Angels     AL        W
## 106 2003  FLO        Florida Marlins     NL        E
## 107 2004  BOS         Boston Red Sox     AL        E
## 108 2005  CHA      Chicago White Sox     AL        C
## 109 2006  SLN    St. Louis Cardinals     NL        C
## 110 2007  BOS         Boston Red Sox     AL        E
## 111 2008  PHI  Philadelphia Phillies     NL        E
```

```
## 112 2009  NYA       New York Yankees     AL        E
## 113 2010  SFN   San Francisco Giants     NL        W
## 114 2011  SLN    St. Louis Cardinals     NL        C
## 115 2012  SFN   San Francisco Giants     NL        W
## 116 2013  BOS         Boston Red Sox     AL        E
```

Question 8

Compute the table that has both the winner and runner-up for the World Series in each tuple/row for all years, again with the name of the team, league and division, and also the number of games the losing team won in the series.

Getting the results for this question means that I need to get the team names for both the winner and runner-up. The issue faced here is that both of those data come from the same table, `Teams`. It is impossible to do a join by writing `JOIN teamIDwinner = teamID = teamIDloser` since it wouldn't give us the desired result. To solve this problem, I did `JOIN`s to the `Teams` table twice and used a different alias for each of them. That way, we can obtain the accurate result as shown below.

```
dbGetQuery(db,
    "SELECT
       SP.yearID AS year,
       SP.teamIDwinner AS 'team(w)',
       TW.name AS 'name(w)',
       SP.lgIDwinner AS 'league(w)',
       TW.divID AS 'division(w)',
       SP.wins AS 'wins(w)',
       SP.teamIDloser AS 'team(l)',
       TL.name AS 'name(l)',
       SP.lgIDloser AS 'league(l)',
       TL.divID AS 'division(l)',
       SP.losses AS 'wins(l)'
     FROM
       SeriesPost AS SP
     LEFT JOIN
       Teams AS TW, Teams AS TL
     ON
       SP.teamIDwinner = TW.teamID AND
       SP.yearID = TW.yearID AND
       SP.teamIDloser = TL.teamID AND
       SP.yearID = TL.yearID
     WHERE
       SP.round = 'WS'
     ORDER BY
       SP.yearID ASC;")
```

```
##     year team(w)               name(w) league(w) division(w) wins(w) team(l)
## 1   1903     BOS       Boston Americans        AL        <NA>       5     PIT
## 2   1905     NY1        New York Giants        NL        <NA>       4     PHA
## 3   1906     CHA      Chicago White Sox        AL        <NA>       4     CHN
## 4   1907     CHN           Chicago Cubs        NL        <NA>       4     DET
## 5   1908     CHN           Chicago Cubs        NL        <NA>       4     DET
## 6   1909     PIT     Pittsburgh Pirates        NL        <NA>       4     DET
## 7   1910     PHA Philadelphia Athletics        AL        <NA>       4     CHN
```

```
## 8   1911     PHA Philadelphia Athletics       AL       <NA>       4     NY1
## 9   1912     BOS         Boston Red Sox       AL       <NA>       4     NY1
## 10  1913     PHA Philadelphia Athletics       AL       <NA>       4     NY1
## 11  1914     BSN          Boston Braves       NL       <NA>       4     PHA
## 12  1915     BOS         Boston Red Sox       AL       <NA>       4     PHI
## 13  1916     BOS         Boston Red Sox       AL       <NA>       4     BRO
## 14  1917     CHA      Chicago White Sox       AL       <NA>       4     NY1
## 15  1918     BOS         Boston Red Sox       AL       <NA>       4     CHN
## 16  1919     CIN        Cincinnati Reds       NL       <NA>       5     CHA
## 17  1920     CLE      Cleveland Indians       AL       <NA>       5     BRO
## 18  1921     NY1        New York Giants       NL       <NA>       5     NYA
## 19  1922     NY1        New York Giants       NL       <NA>       4     NYA
## 20  1923     NYA       New York Yankees       AL       <NA>       4     NY1
## 21  1924     WS1    Washington Senators       AL       <NA>       4     NY1
## 22  1925     PIT     Pittsburgh Pirates       NL       <NA>       4     WS1
## 23  1926     SLN    St. Louis Cardinals       NL       <NA>       4     NYA
## 24  1927     NYA       New York Yankees       AL       <NA>       4     PIT
## 25  1928     NYA       New York Yankees       AL       <NA>       4     SLN
## 26  1929     PHA Philadelphia Athletics       AL       <NA>       4     CHN
## 27  1930     PHA Philadelphia Athletics       AL       <NA>       4     SLN
## 28  1931     SLN    St. Louis Cardinals       NL       <NA>       4     PHA
## 29  1932     NYA       New York Yankees       AL       <NA>       4     CHN
## 30  1933     NY1        New York Giants       NL       <NA>       4     WS1
## 31  1934     SLN    St. Louis Cardinals       NL       <NA>       4     DET
## 32  1935     DET         Detroit Tigers       AL       <NA>       4     CHN
## 33  1936     NYA       New York Yankees       AL       <NA>       4     NY1
## 34  1937     NYA       New York Yankees       AL       <NA>       4     NY1
## 35  1938     NYA       New York Yankees       AL       <NA>       4     CHN
## 36  1939     NYA       New York Yankees       AL       <NA>       4     CIN
## 37  1940     CIN        Cincinnati Reds       NL       <NA>       4     DET
## 38  1941     NYA       New York Yankees       AL       <NA>       4     BRO
## 39  1942     SLN    St. Louis Cardinals       NL       <NA>       4     NYA
## 40  1943     NYA       New York Yankees       AL       <NA>       4     SLN
## 41  1944     SLN    St. Louis Cardinals       NL       <NA>       4     SLA
## 42  1945     DET         Detroit Tigers       AL       <NA>       4     CHN
## 43  1946     SLN    St. Louis Cardinals       NL       <NA>       4     BOS
## 44  1947     NYA       New York Yankees       AL       <NA>       4     BRO
## 45  1948     CLE      Cleveland Indians       AL       <NA>       4     BSN
## 46  1949     NYA       New York Yankees       AL       <NA>       4     BRO
## 47  1950     NYA       New York Yankees       AL       <NA>       4     PHI
## 48  1951     NYA       New York Yankees       AL       <NA>       4     NY1
## 49  1952     NYA       New York Yankees       AL       <NA>       4     BRO
## 50  1953     NYA       New York Yankees       AL       <NA>       4     BRO
## 51  1954     NY1        New York Giants       NL       <NA>       4     CLE
## 52  1955     BRO       Brooklyn Dodgers       NL       <NA>       4     NYA
## 53  1956     NYA       New York Yankees       AL       <NA>       4     BRO
## 54  1957     ML1       Milwaukee Braves       NL       <NA>       4     NYA
## 55  1958     NYA       New York Yankees       AL       <NA>       4     ML1
## 56  1959     LAN    Los Angeles Dodgers       NL       <NA>       4     CHA
## 57  1960     PIT     Pittsburgh Pirates       NL       <NA>       4     NYA
## 58  1961     NYA       New York Yankees       AL       <NA>       4     CIN
## 59  1962     NYA       New York Yankees       AL       <NA>       4     SFN
## 60  1963     LAN    Los Angeles Dodgers       NL       <NA>       4     NYA
## 61  1964     SLN    St. Louis Cardinals       NL       <NA>       4     NYA
```

```
## 62  1965     LAN    Los Angeles Dodgers        NL        <NA>       4     MIN
## 63  1966     BAL      Baltimore Orioles        AL        <NA>       4     LAN
## 64  1967     SLN    St. Louis Cardinals        NL        <NA>       4     BOS
## 65  1968     DET         Detroit Tigers        AL        <NA>       4     SLN
## 66  1969     NYN          New York Mets        NL           E       4     BAL
## 67  1970     BAL      Baltimore Orioles        AL           E       4     CIN
## 68  1971     PIT     Pittsburgh Pirates        NL           E       4     BAL
## 69  1972     OAK      Oakland Athletics        AL           W       4     CIN
## 70  1973     OAK      Oakland Athletics        AL           W       4     NYN
## 71  1974     OAK      Oakland Athletics        AL           W       4     LAN
## 72  1975     CIN        Cincinnati Reds        NL           W       4     BOS
## 73  1976     CIN        Cincinnati Reds        NL           W       4     NYA
## 74  1977     NYA       New York Yankees        AL           E       4     LAN
## 75  1978     NYA       New York Yankees        AL           E       4     LAN
## 76  1979     PIT     Pittsburgh Pirates        NL           E       4     BAL
## 77  1980     PHI  Philadelphia Phillies        NL           E       4     KCA
## 78  1981     LAN    Los Angeles Dodgers        NL           W       4     NYA
## 79  1982     SLN    St. Louis Cardinals        NL           E       4     ML4
## 80  1983     BAL      Baltimore Orioles        AL           E       4     PHI
## 81  1984     DET         Detroit Tigers        AL           E       4     SDN
## 82  1985     KCA     Kansas City Royals        AL           W       4     SLN
## 83  1986     NYN          New York Mets        NL           E       4     BOS
## 84  1987     MIN        Minnesota Twins        AL           W       4     SLN
## 85  1988     LAN    Los Angeles Dodgers        NL           W       4     OAK
## 86  1989     OAK      Oakland Athletics        AL           W       4     SFN
## 87  1990     CIN        Cincinnati Reds        NL           W       4     OAK
## 88  1991     MIN        Minnesota Twins        AL           W       4     ATL
## 89  1992     TOR      Toronto Blue Jays        AL           E       4     ATL
## 90  1993     TOR      Toronto Blue Jays        AL           E       4     PHI
## 91  1995     ATL         Atlanta Braves        NL           E       4     CLE
## 92  1996     NYA       New York Yankees        AL           E       4     ATL
## 93  1997     FLO        Florida Marlins        NL           E       4     CLE
## 94  1998     NYA       New York Yankees        AL           E       4     SDN
## 95  1999     NYA       New York Yankees        AL           E       4     ATL
## 96  2000     NYA       New York Yankees        AL           E       4     NYN
## 97  2001     ARI   Arizona Diamondbacks        NL           W       4     NYA
## 98  2002     ANA         Anaheim Angels        AL           W       4     SFN
## 99  2003     FLO        Florida Marlins        NL           E       4     NYA
## 100 2004     BOS         Boston Red Sox        AL           E       4     SLN
## 101 2005     CHA      Chicago White Sox        AL           C       4     HOU
## 102 2006     SLN    St. Louis Cardinals        NL           C       4     DET
## 103 2007     BOS         Boston Red Sox        AL           E       4     COL
## 104 2008     PHI  Philadelphia Phillies        NL           E       4     TBA
## 105 2009     NYA       New York Yankees        AL           E       4     PHI
## 106 2010     SFN   San Francisco Giants        NL           W       4     TEX
## 107 2011     SLN    St. Louis Cardinals        NL           C       4     TEX
## 108 2012     SFN   San Francisco Giants        NL           W       4     DET
## 109 2013     BOS         Boston Red Sox        AL           E       4     SLN
##                    name(l) league(l) division(l) wins(l)
## 1       Pittsburgh Pirates        NL        <NA>       3
## 2   Philadelphia Athletics        AL        <NA>       1
## 3             Chicago Cubs        NL        <NA>       2
## 4           Detroit Tigers        AL        <NA>       0
## 5           Detroit Tigers        AL        <NA>       1
```

```
## 6           Detroit Tigers        AL        <NA>       3
## 7             Chicago Cubs        NL        <NA>       1
## 8          New York Giants        NL        <NA>       2
## 9          New York Giants        NL        <NA>       3
## 10         New York Giants        NL        <NA>       1
## 11  Philadelphia Athletics        AL        <NA>       0
## 12   Philadelphia Phillies        NL        <NA>       1
## 13         Brooklyn Robins        NL        <NA>       1
## 14         New York Giants        NL        <NA>       2
## 15            Chicago Cubs        NL        <NA>       2
## 16       Chicago White Sox        AL        <NA>       3
## 17         Brooklyn Robins        NL        <NA>       2
## 18        New York Yankees        AL        <NA>       3
## 19        New York Yankees        AL        <NA>       0
## 20         New York Giants        NL        <NA>       2
## 21         New York Giants        NL        <NA>       3
## 22    Washington Senators        AL        <NA>       3
## 23        New York Yankees        AL        <NA>       3
## 24     Pittsburgh Pirates        NL        <NA>       0
## 25    St. Louis Cardinals        NL        <NA>       0
## 26            Chicago Cubs        NL        <NA>       1
## 27    St. Louis Cardinals        NL        <NA>       2
## 28  Philadelphia Athletics        AL        <NA>       3
## 29            Chicago Cubs        NL        <NA>       0
## 30    Washington Senators        AL        <NA>       1
## 31          Detroit Tigers        AL        <NA>       3
## 32            Chicago Cubs        NL        <NA>       2
## 33         New York Giants        NL        <NA>       2
## 34         New York Giants        NL        <NA>       1
## 35            Chicago Cubs        NL        <NA>       0
## 36         Cincinnati Reds        NL        <NA>       0
## 37          Detroit Tigers        AL        <NA>       3
## 38        Brooklyn Dodgers        NL        <NA>       1
## 39        New York Yankees        AL        <NA>       1
## 40    St. Louis Cardinals        NL        <NA>       1
## 41       St. Louis Browns        AL        <NA>       2
## 42            Chicago Cubs        NL        <NA>       3
## 43          Boston Red Sox        AL        <NA>       3
## 44        Brooklyn Dodgers        NL        <NA>       3
## 45           Boston Braves        NL        <NA>       2
## 46        Brooklyn Dodgers        NL        <NA>       1
## 47   Philadelphia Phillies        NL        <NA>       0
## 48         New York Giants        NL        <NA>       2
## 49        Brooklyn Dodgers        NL        <NA>       3
## 50        Brooklyn Dodgers        NL        <NA>       2
## 51       Cleveland Indians        AL        <NA>       0
## 52        New York Yankees        AL        <NA>       3
## 53        Brooklyn Dodgers        NL        <NA>       3
## 54        New York Yankees        AL        <NA>       3
## 55        Milwaukee Braves        NL        <NA>       3
## 56       Chicago White Sox        AL        <NA>       2
## 57        New York Yankees        AL        <NA>       3
## 58         Cincinnati Reds        NL        <NA>       1
## 59    San Francisco Giants        NL        <NA>       3
```

```
## 60        New York Yankees        AL        <NA>       0
## 61        New York Yankees        AL        <NA>       3
## 62         Minnesota Twins        AL        <NA>       3
## 63     Los Angeles Dodgers        NL        <NA>       0
## 64          Boston Red Sox        AL        <NA>       3
## 65     St. Louis Cardinals        NL        <NA>       3
## 66       Baltimore Orioles        AL           E       1
## 67         Cincinnati Reds        NL           W       1
## 68       Baltimore Orioles        AL           E       3
## 69         Cincinnati Reds        NL           W       3
## 70           New York Mets        NL           E       3
## 71     Los Angeles Dodgers        NL           W       1
## 72          Boston Red Sox        AL           E       3
## 73        New York Yankees        AL           E       0
## 74     Los Angeles Dodgers        NL           W       2
## 75     Los Angeles Dodgers        NL           W       2
## 76       Baltimore Orioles        AL           E       3
## 77      Kansas City Royals        AL           W       2
## 78        New York Yankees        AL           E       2
## 79       Milwaukee Brewers        AL           E       3
## 80   Philadelphia Phillies        NL           E       1
## 81        San Diego Padres        NL           W       1
## 82     St. Louis Cardinals        NL           E       3
## 83          Boston Red Sox        AL           E       3
## 84     St. Louis Cardinals        NL           E       3
## 85       Oakland Athletics        AL           W       1
## 86    San Francisco Giants        NL           W       0
## 87       Oakland Athletics        AL           W       0
## 88          Atlanta Braves        NL           W       3
## 89          Atlanta Braves        NL           W       2
## 90   Philadelphia Phillies        NL           E       2
## 91       Cleveland Indians        AL           C       2
## 92          Atlanta Braves        NL           E       2
## 93       Cleveland Indians        AL           C       3
## 94        San Diego Padres        NL           W       0
## 95          Atlanta Braves        NL           E       0
## 96           New York Mets        NL           E       1
## 97        New York Yankees        AL           E       3
## 98    San Francisco Giants        NL           W       3
## 99        New York Yankees        AL           E       2
## 100    St. Louis Cardinals        NL           C       0
## 101         Houston Astros        NL           C       0
## 102         Detroit Tigers        AL           C       1
## 103       Colorado Rockies        NL           W       0
## 104         Tampa Bay Rays        AL           E       1
## 105  Philadelphia Phillies        NL           E       2
## 106          Texas Rangers        AL           W       1
## 107          Texas Rangers        AL           W       3
## 108         Detroit Tigers        AL           C       0
## 109    St. Louis Cardinals        NL           C       3
```

Question 9

Do you see a relationship between the number of games won in a season and winning the World Series?

To solve this problem, all we have to do is join the `SeriesPost` table with the `Teams` table on `yearID` and `teamID = teamIDwinner`, and then get all the necessary fields. the query and results are shown below:

```
winner_stat = dbGetQuery(db,
    "SELECT
       SP.yearID AS year,
       SP.teamIDwinner AS team,
       T.name AS name,
       T.W AS wins,
       T.L AS losses
     FROM
       SeriesPost AS SP
     LEFT JOIN
       Teams AS T
     ON
       SP.yearID = T.yearID AND
       SP.teamIDwinner = T.teamID
     WHERE
       SP.round = 'WS' AND
       (T.W IS NOT NULL OR
       T.L IS NOT NULL);")
winner_stat
```

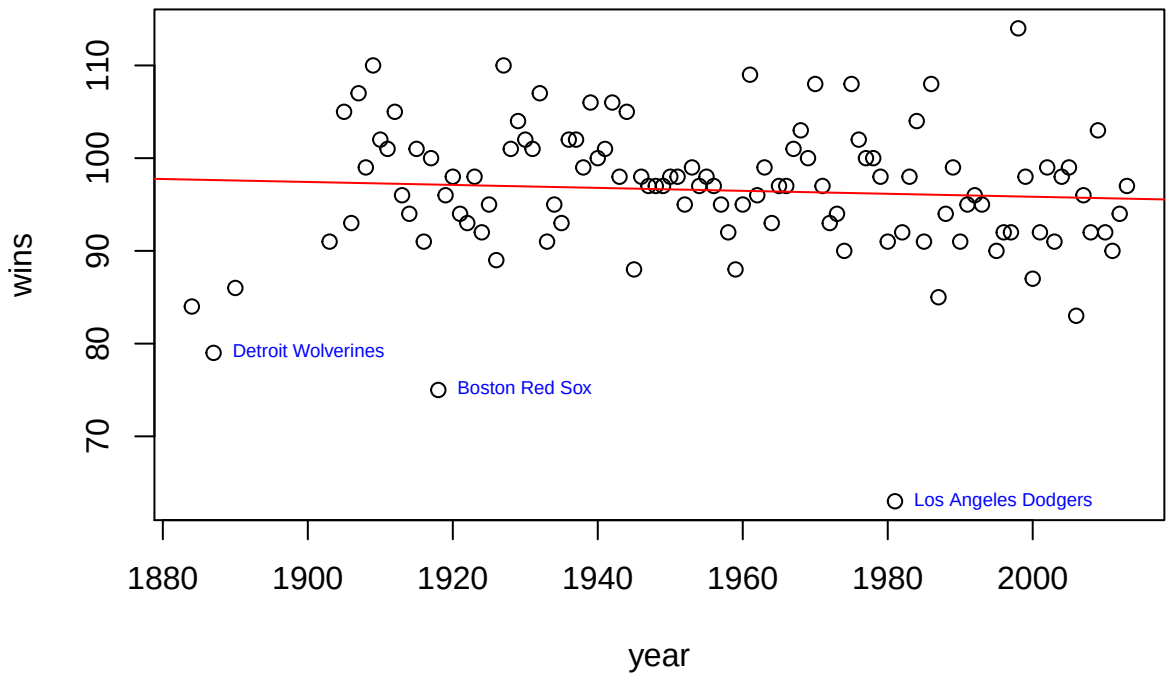
```
##     year team                   name wins losses
## 1   1884  PRO       Providence Grays   84     28
## 2   1887  DTN     Detroit Wolverines   79     45
## 3   1890  BRO   Brooklyn Bridegrooms   86     43
## 4   1903  BOS       Boston Americans   91     47
## 5   1905  NY1        New York Giants  105     48
## 6   1906  CHA      Chicago White Sox   93     58
## 7   1907  CHN           Chicago Cubs  107     45
## 8   1908  CHN           Chicago Cubs   99     55
## 9   1909  PIT     Pittsburgh Pirates  110     42
## 10  1910  PHA Philadelphia Athletics  102     48
## 11  1911  PHA Philadelphia Athletics  101     50
## 12  1912  BOS         Boston Red Sox  105     47
## 13  1913  PHA Philadelphia Athletics   96     57
## 14  1914  BSN          Boston Braves   94     59
## 15  1915  BOS         Boston Red Sox  101     50
## 16  1916  BOS         Boston Red Sox   91     63
## 17  1917  CHA      Chicago White Sox  100     54
## 18  1918  BOS         Boston Red Sox   75     51
## 19  1919  CIN        Cincinnati Reds   96     44
## 20  1920  CLE      Cleveland Indians   98     56
## 21  1921  NY1        New York Giants   94     59
## 22  1922  NY1        New York Giants   93     61
## 23  1923  NYA       New York Yankees   98     54
## 24  1924  WS1    Washington Senators   92     62
## 25  1925  PIT     Pittsburgh Pirates   95     58
```

```
## 26  1926  SLN    St. Louis Cardinals   89     65
## 27  1927  NYA       New York Yankees  110     44
## 28  1928  NYA       New York Yankees  101     53
## 29  1929  PHA Philadelphia Athletics  104     46
## 30  1930  PHA Philadelphia Athletics  102     52
## 31  1931  SLN    St. Louis Cardinals  101     53
## 32  1932  NYA       New York Yankees  107     47
## 33  1933  NY1        New York Giants   91     61
## 34  1934  SLN    St. Louis Cardinals   95     58
## 35  1935  DET         Detroit Tigers   93     58
## 36  1936  NYA       New York Yankees  102     51
## 37  1937  NYA       New York Yankees  102     52
## 38  1938  NYA       New York Yankees   99     53
## 39  1939  NYA       New York Yankees  106     45
## 40  1940  CIN        Cincinnati Reds  100     53
## 41  1941  NYA       New York Yankees  101     53
## 42  1942  SLN    St. Louis Cardinals  106     48
## 43  1943  NYA       New York Yankees   98     56
## 44  1944  SLN    St. Louis Cardinals  105     49
## 45  1945  DET         Detroit Tigers   88     65
## 46  1946  SLN    St. Louis Cardinals   98     58
## 47  1947  NYA       New York Yankees   97     57
## 48  1948  CLE      Cleveland Indians   97     58
## 49  1949  NYA       New York Yankees   97     57
## 50  1950  NYA       New York Yankees   98     56
## 51  1951  NYA       New York Yankees   98     56
## 52  1952  NYA       New York Yankees   95     59
## 53  1953  NYA       New York Yankees   99     52
## 54  1954  NY1        New York Giants   97     57
## 55  1955  BRO       Brooklyn Dodgers   98     55
## 56  1956  NYA       New York Yankees   97     57
## 57  1957  ML1       Milwaukee Braves   95     59
## 58  1958  NYA       New York Yankees   92     62
## 59  1959  LAN    Los Angeles Dodgers   88     68
## 60  1960  PIT     Pittsburgh Pirates   95     59
## 61  1961  NYA       New York Yankees  109     53
## 62  1962  NYA       New York Yankees   96     66
## 63  1963  LAN    Los Angeles Dodgers   99     63
## 64  1964  SLN    St. Louis Cardinals   93     69
## 65  1965  LAN    Los Angeles Dodgers   97     65
## 66  1966  BAL      Baltimore Orioles   97     63
## 67  1967  SLN    St. Louis Cardinals  101     60
## 68  1968  DET         Detroit Tigers  103     59
## 69  1969  NYN          New York Mets  100     62
## 70  1970  BAL      Baltimore Orioles  108     54
## 71  1971  PIT     Pittsburgh Pirates   97     65
## 72  1972  OAK      Oakland Athletics   93     62
## 73  1973  OAK      Oakland Athletics   94     68
## 74  1974  OAK      Oakland Athletics   90     72
## 75  1975  CIN        Cincinnati Reds  108     54
## 76  1976  CIN        Cincinnati Reds  102     60
## 77  1977  NYA       New York Yankees  100     62
## 78  1978  NYA       New York Yankees  100     63
## 79  1979  PIT     Pittsburgh Pirates   98     64
```

```
## 80  1980  PHI  Philadelphia Phillies   91     71
## 81  1981  LAN    Los Angeles Dodgers   63     47
## 82  1982  SLN    St. Louis Cardinals   92     70
## 83  1983  BAL      Baltimore Orioles   98     64
## 84  1984  DET         Detroit Tigers  104     58
## 85  1985  KCA     Kansas City Royals   91     71
## 86  1986  NYN          New York Mets  108     54
## 87  1987  MIN        Minnesota Twins   85     77
## 88  1988  LAN    Los Angeles Dodgers   94     67
## 89  1989  OAK      Oakland Athletics   99     63
## 90  1990  CIN        Cincinnati Reds   91     71
## 91  1991  MIN        Minnesota Twins   95     67
## 92  1992  TOR      Toronto Blue Jays   96     66
## 93  1993  TOR      Toronto Blue Jays   95     67
## 94  1995  ATL         Atlanta Braves   90     54
## 95  1996  NYA       New York Yankees   92     70
## 96  1997  FLO        Florida Marlins   92     70
## 97  1998  NYA       New York Yankees  114     48
## 98  1999  NYA       New York Yankees   98     64
## 99  2000  NYA       New York Yankees   87     74
## 100 2001  ARI   Arizona Diamondbacks   92     70
## 101 2002  ANA         Anaheim Angels   99     63
## 102 2003  FLO        Florida Marlins   91     71
## 103 2004  BOS         Boston Red Sox   98     64
## 104 2005  CHA      Chicago White Sox   99     63
## 105 2006  SLN    St. Louis Cardinals   83     78
## 106 2007  BOS         Boston Red Sox   96     66
## 107 2008  PHI  Philadelphia Phillies   92     70
## 108 2009  NYA       New York Yankees  103     59
## 109 2010  SFN   San Francisco Giants   92     70
## 110 2011  SLN    St. Louis Cardinals   90     72
## 111 2012  SFN   San Francisco Giants   94     68
## 112 2013  BOS         Boston Red Sox   97     65
```

From the plot below, we can see that the World Series winners each year gets more than 90 wins the entire season most of the time. There are a few exceptions though, LA Dodgers, Boston Red Sox, and Detroit Wolverines are some of the World Series winners with the least season wins, with each of them winning less than 80 games in 1981, 1918, and 1887 respectively.

```
lm1 = lm(wins ~ year, data = winner_stat)
plot(wins ~ year, data = winner_stat)
abline(lm1$coefficients, col = "red")
leastwins = subset(winner_stat, winner_stat$wins < 80)
text(leastwins$year, leastwins$wins, leastwins$name, cex=0.6, pos=4, col="blue")
```



Question 10

In 2003, what were the three highest salaries? (We refer here to unique salaries, ie. there may be several players getting the exact same amount.) Find the players who got any of these 3 salaries with all of their details.

To find the answer to this solution, we have to rank the players based on their salaries. There are three types of ranks in SQL:

- `ROW_NUMBER()`: ranks the players distinctly from 1 to N regardless if we have players with the same salary.
- `RANK()`: ranks the players from 1 to N. Players with the same salary will have the same rank, but the following ranking will be skipped, eg: 1,2,3,3,3,6,...
- `DENSE_RANK()`: ranks the players from 1 to N. Players with the same salary will have the same rank, and the following ranking will not be skipped, eg: 1,2,2,3,3,3,4,...

Clearly, we will need to use `DENSE_RANK()` in this case based on the question that we are asked. With that, we have the following query and result:

```
dbGetQuery(db,
    "SELECT
        *
     FROM (
```

```
        SELECT
            T.name AS team,
            S.lgID,
            M.nameFirst || ' ' || M.nameLast AS name,
            S.salary,
            DENSE_RANK() OVER (ORDER BY S.salary DESC) AS rank
         FROM
            Salaries AS S
         INNER JOIN
            Teams AS T
         ON
            S.teamID = T.teamID AND
            S.yearID = T.yearID
         INNER JOIN
            Master as M
         ON
            S.playerID = M.playerID
         WHERE
            S.yearID = 2003
         ORDER BY
            rank
     ) AS ranked_salary
     WHERE
        rank <= 3;")
```

```
##                team lgID           name   salary rank
## 1     Texas Rangers   AL Alex Rodriguez 22000000    1
## 2    Boston Red Sox   AL  Manny Ramirez 20000000    2
## 3 Toronto Blue Jays   AL Carlos Delgado 18700000    3
```

Question 11

For 2010, compute the total payroll of each of the different teams. Next compute the team payrolls for all years in the database for which we have salary information. Display these in a plot.

The first part of the problem is relatively easy to compute, we only need to sum up all the players' salaries across each team using the `GROUP BY` clause for the year 2010. The results are shown below:

```
salary2010_info = dbGetQuery(db,
    "SELECT
        T.name AS team,
        SUM(S.salary) AS total_payroll
     FROM
        Salaries AS S
     JOIN
        Teams AS T
     ON
        S.teamID = T.teamID AND
        S.yearID = T.yearID
     WHERE
        S.yearID = 2010
     GROUP BY
```

```
        S.teamID
    ORDER BY
        total_payroll DESC;")
salary2010_info
```

```
##                            team total_payroll
## 1                New York Yankees     206333389
## 2                  Boston Red Sox     162447333
## 3                    Chicago Cubs     146609000
## 4           Philadelphia Phillies     141928379
## 5                   New York Mets     134422942
## 6                  Detroit Tigers     122864928
## 7               Chicago White Sox     105530000
## 8  Los Angeles Angels of Anaheim     104963866
## 9            San Francisco Giants      98641333
## 10                Minnesota Twins      97559166
## 11             Los Angeles Dodgers      95358016
## 12             St. Louis Cardinals      93540751
## 13                  Houston Astros      92355500
## 14                Seattle Mariners      86510000
## 15                  Atlanta Braves      84423666
## 16                Colorado Rockies      84227000
## 17               Baltimore Orioles      81612500
## 18               Milwaukee Brewers      81108278
## 19                  Tampa Bay Rays      71923471
## 20                Cincinnati Reds      71761542
## 21              Kansas City Royals      71405210
## 22               Toronto Blue Jays      62234000
## 23            Washington Nationals      61400000
## 24               Cleveland Indians      61203966
## 25            Arizona Diamondbacks      60718166
## 26                  Florida Marlins      57029719
## 27               Oakland Athletics      55254900
## 28                   Texas Rangers      55250544
## 29                San Diego Padres      37799300
## 30              Pittsburgh Pirates      34943000
```

This next section is slightly trickier. It isn't all that different from the previous query, except that this time we use the `GROUP BY` clause to group years and teams. Note that I have added certain fields in this query to answer Question 12.

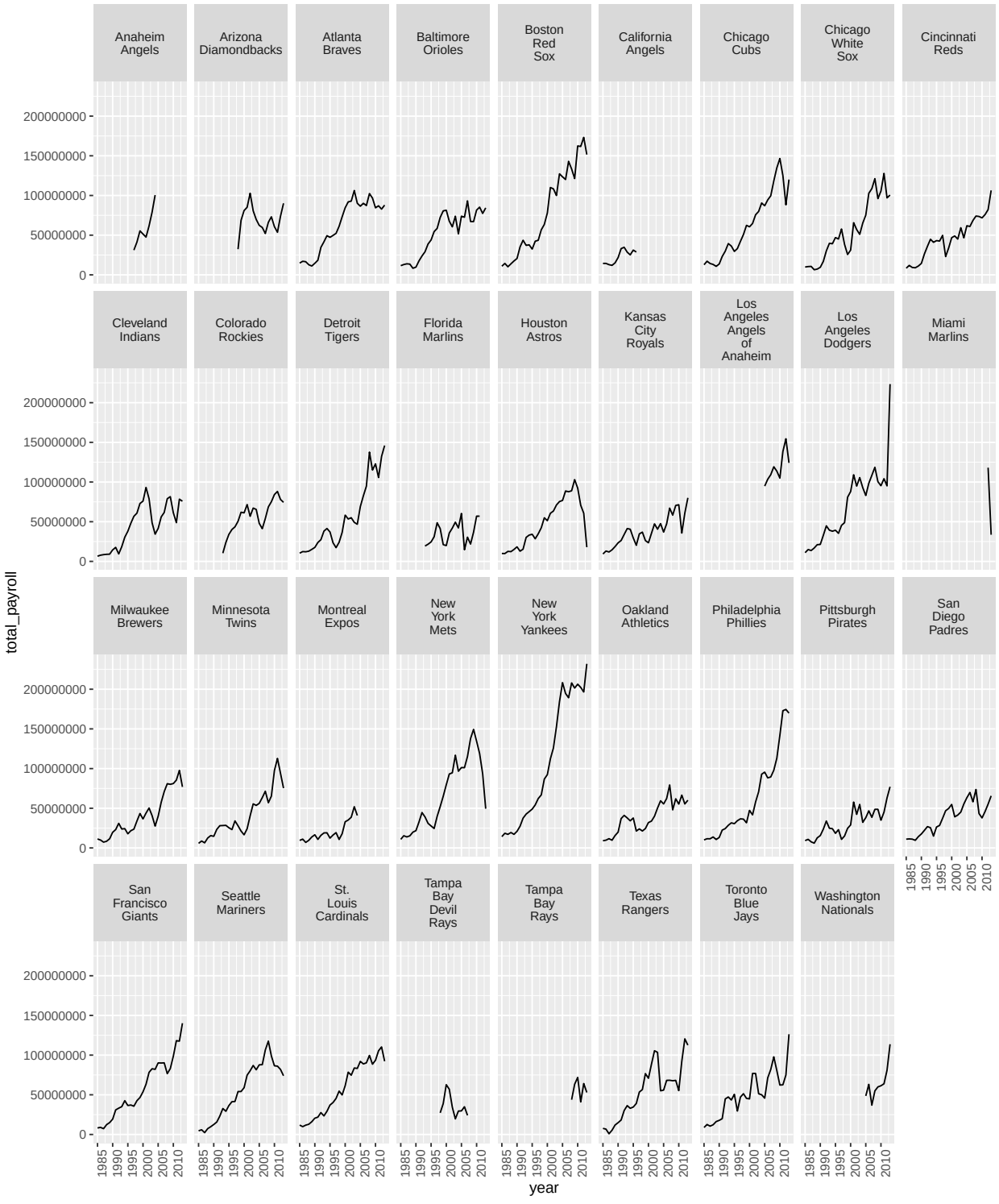
The result of this query is a relatively large data frame.

```
salary_info = dbGetQuery(db,
    "SELECT
        S.yearID AS year,
        S.teamID AS team_id,
        T.name AS team,
        T.lgID AS league,
        T.divID AS division,
        IFNULL(T.WSWin, 'N') AS WS_winner,
        IFNULL(T.LgWin, 'N') AS Lg_winner,
        SUM(S.salary) AS total_payroll
```

```
    FROM
       Salaries AS S
    JOIN
       Teams AS T
    ON
       S.teamID = T.teamID AND
       S.yearID = T.yearID
   GROUP BY
       S.teamID,
       S.yearID
    ORDER BY
       year ASC,
       total_payroll DESC;")
```

The following set of plots show the payroll trend of each team throughout the years. It is interesting to notice that teams with small total payrolls that do not have huge payroll increases each year tend to not have much data. This could mean that those teams have probably gone bankrupt and do not have enough financial ability to pay their players. Some examples are the California Angels, the Tampa Bay Devil Rays, and the Tampa Bay Rays. Otherwise, teams that have been around for a long time tend to exhibit rapid increases in total payroll (almost exponential for some!).

```
options(scipen = 10)
ggplot(salary_info, aes(x = year, y = total_payroll)) +
  geom_line() +
  facet_wrap(vars(team), ncol = 9, labeller = label_wrap_gen(1)) +
  theme(axis.text.x = element_text(angle = 90))
```



Question 12

Explore the change in salary over time. Use a plot. Identify the teams that won the world series or league on the plot. How does salary relate to winning the league and/or world series?

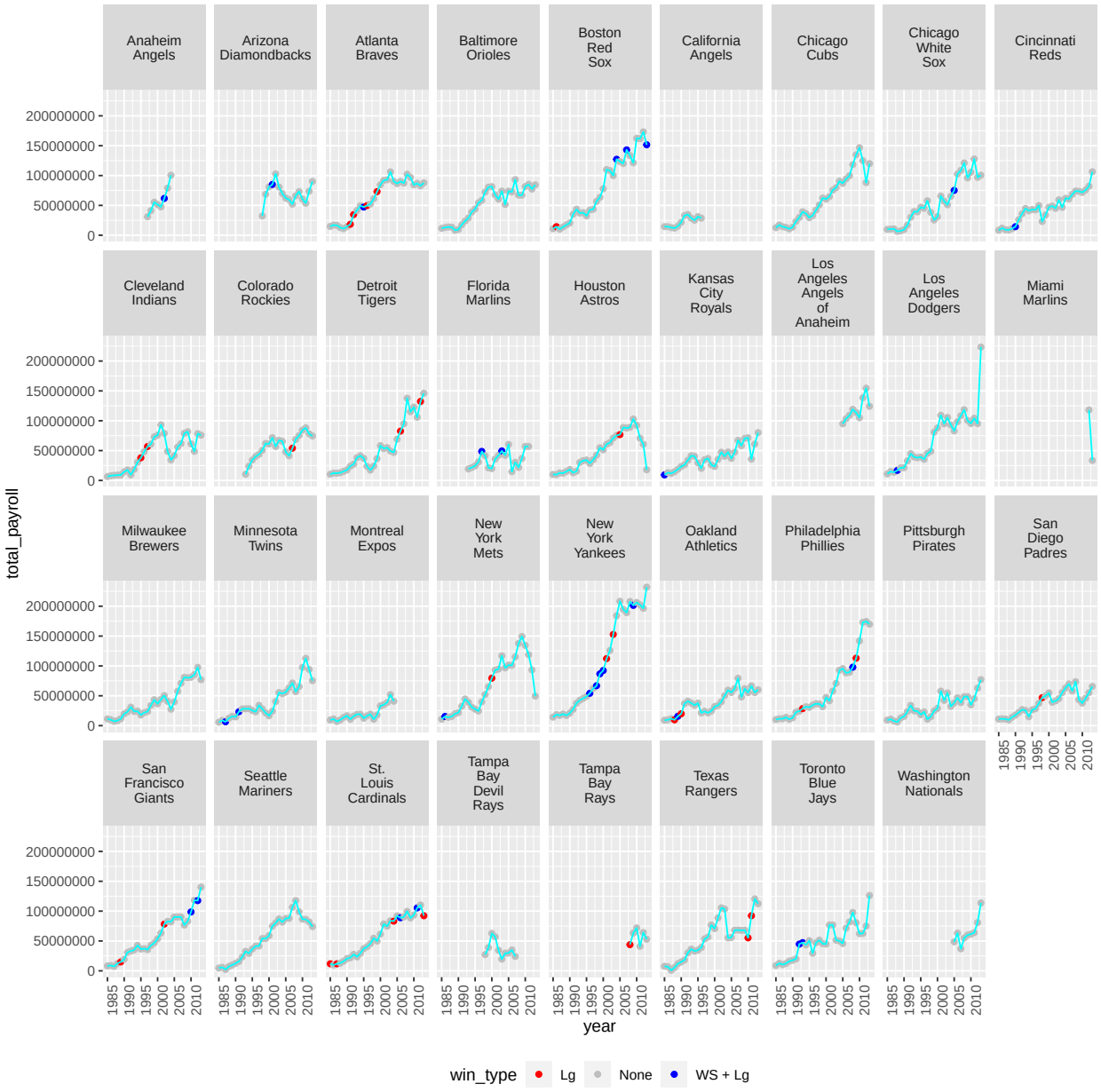
The data frame to answer this question has been obtained in the query in Question 11. In the plot below,

we can see that wins in the league and the World Series is often accompanied with a huge increase in total payroll. Teams that do not incorporate this model however, often see a long period of failure to win the league and World Series. Some examples here are the Arizoa Diamondbacks, the Florida Marlins, and the Houston Astros. On the other hand, the New York Yankees' exponential increase in total payroll is accompanied by various league and World Series succeesses.

```
options(scipen = 10)

salary_info$win_type = sapply(paste0(salary_info$WS_winner, salary_info$Lg_winner),
                              switch,
                              "NN" = "None",
                              "NY" = "Lg",
                              "YN" = "WS",
                              "YY" = "WS + Lg")

ggplot(salary_info, aes(x = year, y = total_payroll)) +
  geom_point(aes(col = win_type)) +
  geom_line(col = "cyan") +
  facet_wrap(vars(team), ncol = 9, labeller = label_wrap_gen(1)) +
  scale_color_manual(values = c("red", "gray", "blue")) +
  theme(axis.text.x = element_text(angle = 90),
        legend.position = "bottom")
```

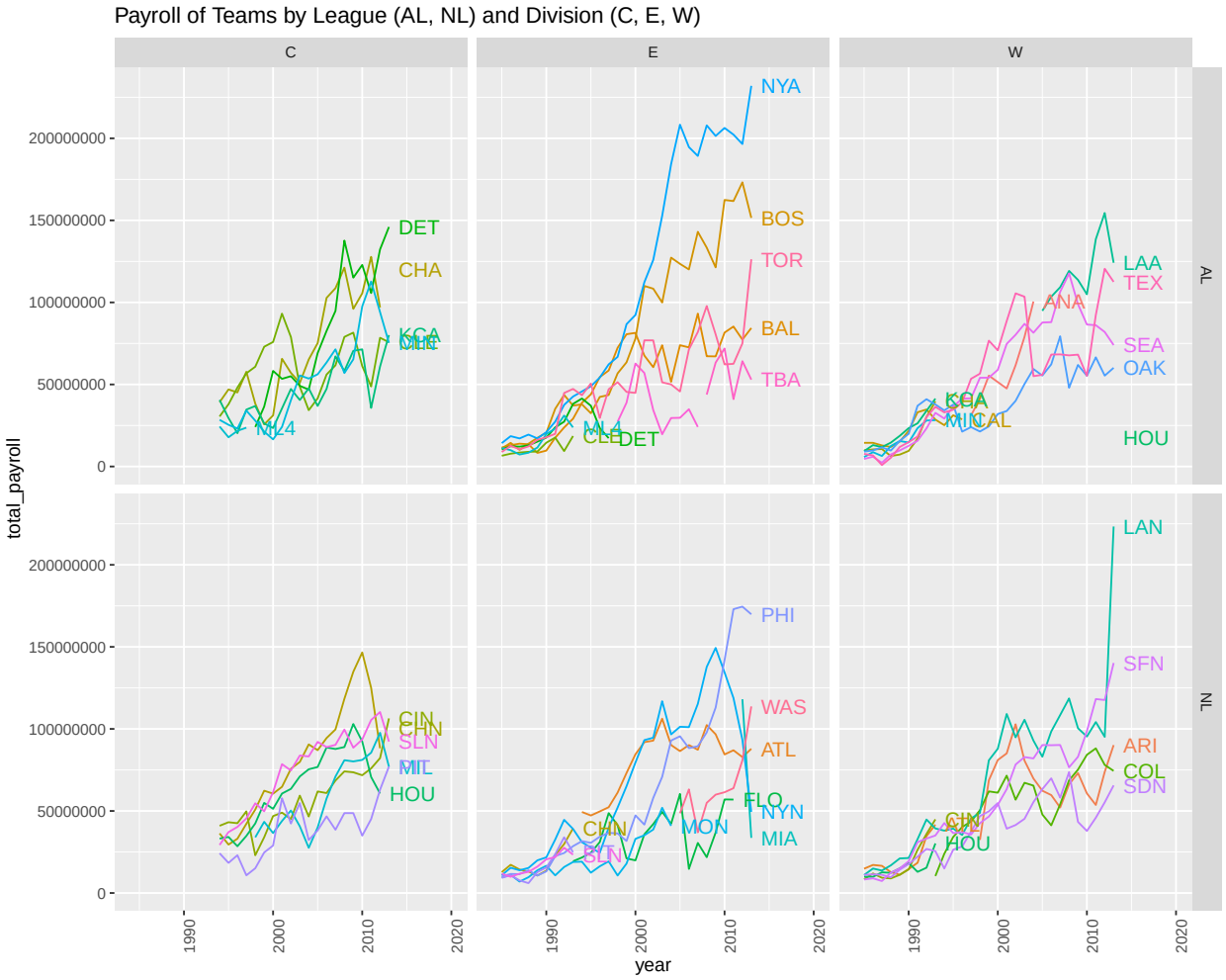


Question 13

Compare the payrolls for the teams that are in the same leagues, and then in the same divisions. Are there any interesting characteristics? Have certain teams always had top payrolls over the years? Is there a connection between payroll and performance?

Based on the plot below, it seems that there isn't much of a difference in payroll between teams in the same league or division prior to the year 2000. After the year 2000, the total payroll within each league and division has been dominated by some specific teams, such as the New York Yankees (NYA) and the Boston Red Sox (BOS) in the AL league of the Eastern division, and Atlanta Braves (ATL), New York Mets(NYN), and Philadelphia Phillies (PHI) in the NL league of the Eastern division.

```
options(scipen = 10)
ggplot(salary_info, aes(x = year, y = total_payroll)) +
  geom_line(aes(col = team)) + xlim(1984,2020) +
  facet_grid(vars(league), vars(division), labeller = label_wrap_gen(1)) +
  ggtitle("Payroll of Teams by League (AL, NL) and Division (C, E, W)") +
  theme(axis.text.x = element_text(angle = 90),
        legend.position = "none") +
  directlabels::geom_dl(aes(label = team_id, col = team),
                        method = list(directlabels::dl.trans(x = x + .2),
                                      "last.points"))
```



Question 14

Which player has hit the most home runs? Show the number per year.

To solve this problem, I first had to create a temporary table in SQLite using `CREATE VIEW` to sum up the number of home runs of each player from the Batting table and then compute a `DENSE_RANK()` over them.

```
dbExecute(db,
    "CREATE VIEW IF NOT EXISTS
```

```
        total_hr
    AS
        SELECT
            playerID,
            SUM(IFNULL(HR,0)) AS home_runs,
            DENSE_RANK() OVER (ORDER BY SUM(HR) DESC) AS rank
         FROM
            Batting
         GROUP BY
            playerID;")
```

With that total home run temporary table computed, I then write a query to get the home runs from each year for the top ranked player, ie `rank = 1`. The query is shown below.

```
most_hr = dbGetQuery(db,
    "SELECT
        B.yearID AS year,
        B.playerID AS id,
        M.nameFirst || ' ' || M.nameLast AS name,
        B.HR AS home_runs
     FROM
        Batting AS B
     LEFT JOIN
        Master AS M
     ON
        B.playerID = M.playerID
     WHERE
        B.playerID = (
            SELECT
                playerID
            FROM
                total_hr
            WHERE
                rank = 1);")
```

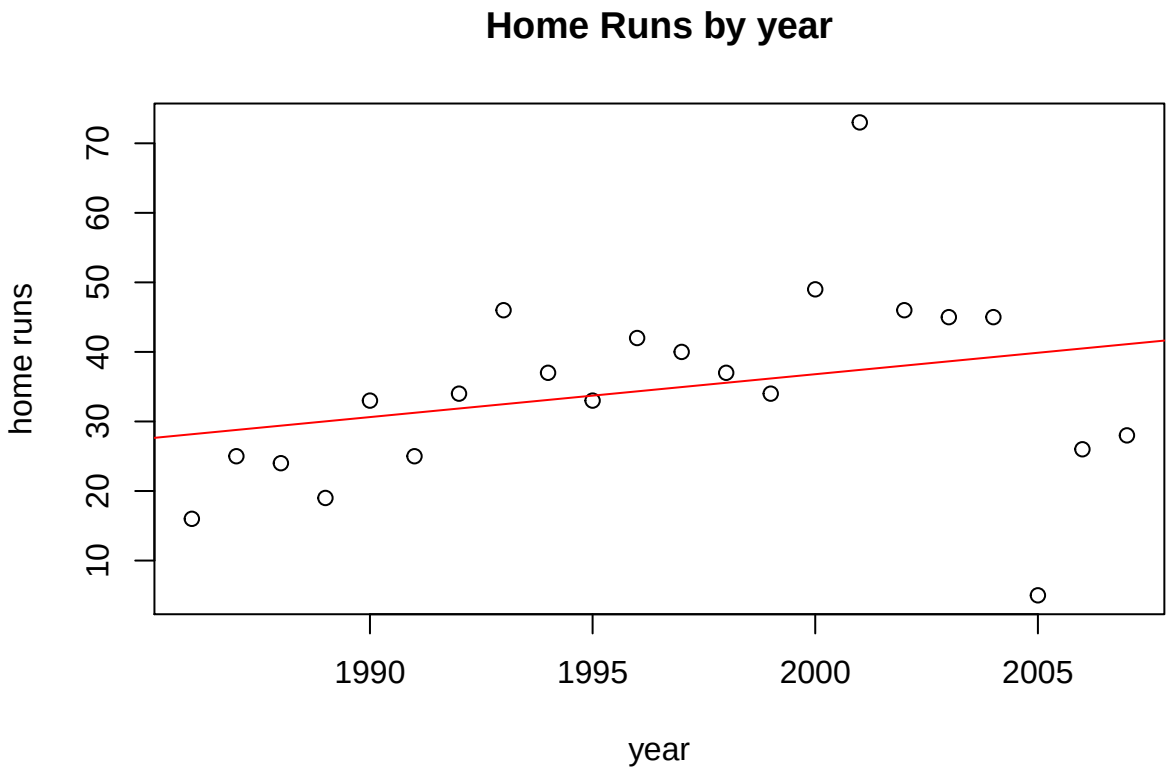
The following table shows the top 5 seasons in terms of number of home runs by the top ranked player, Barry Bonds. We can see that he has had an insane season in 2001, where he scored 73 home runs, at least 24 more than any other season.

```
most_hr[order(most_hr$home_runs, decreasing = T)[1:5],]
```

```
##    year        id        name home_runs
## 16 2001 bondsba01 Barry Bonds        73
## 15 2000 bondsba01 Barry Bonds        49
## 8  1993 bondsba01 Barry Bonds        46
## 17 2002 bondsba01 Barry Bonds        46
## 18 2003 bondsba01 Barry Bonds        45
```

The following plot is a visualization of Barry Bonds' home run hits over the years. We can see that his home run numbers has been consistently increasing over the years.

```
lm2 = lm(home_runs ~ year, most_hr)
plot(most_hr$year, most_hr$home_runs,
     main = "Home Runs by year",
     xlab = "year", ylab = "home runs")
abline(lm2$coefficients, col = "red")
```




```
dbExecute(db, "DROP VIEW total_hr")
```

Question 15

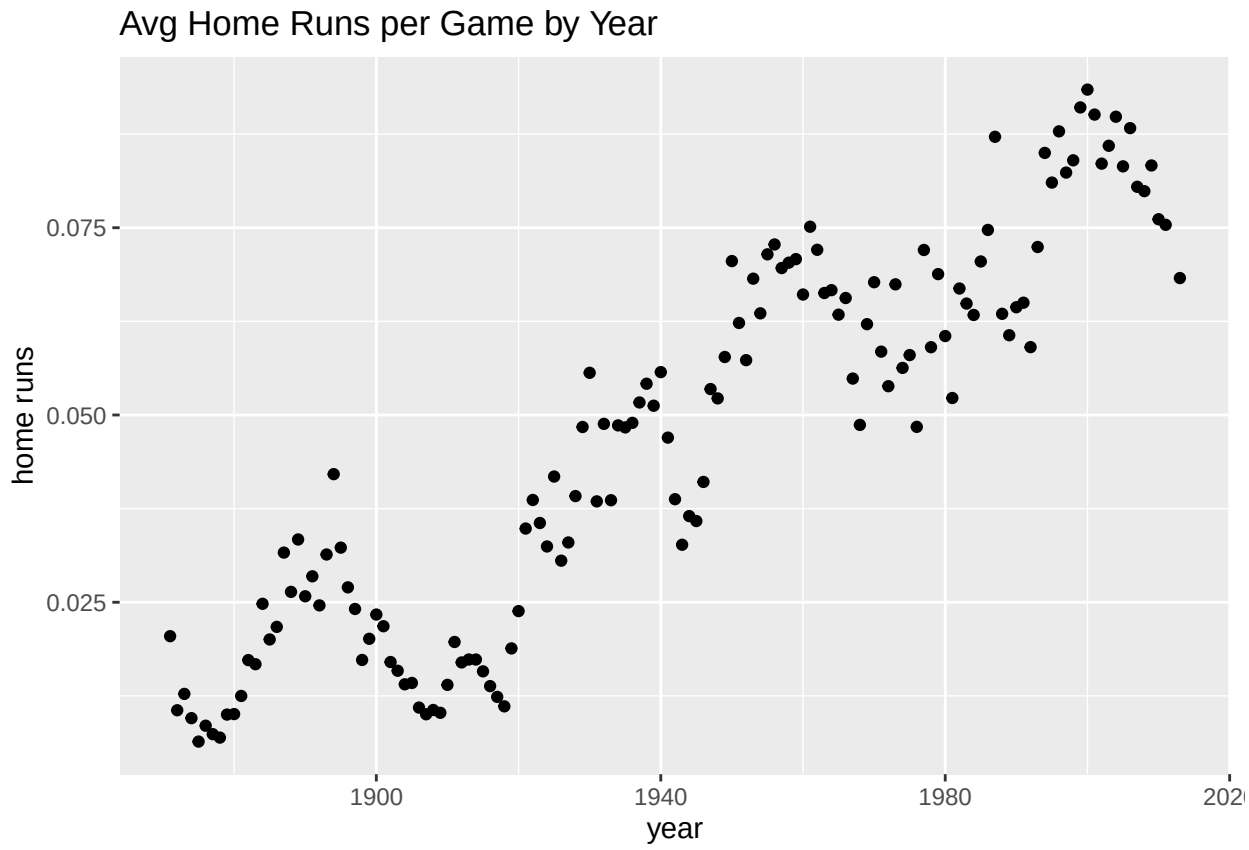
Has the distribution of home runs for players increased over the years?

In this query, we compute the average number of home runs per game over each year. The result is shown in the following plot. Clearly, there is a significant increasing trend in number of home runs from the 1870s to the 2010s. Note that information from 2012 is not included since there is no data on the number of games as batter for each player.

```
homerun_yr = dbGetQuery(db,
    "SELECT
       yearID AS year,
       SUM(G_batting) AS games,
       SUM(HR) AS home_runs
     FROM
       Batting
```

```
    GROUP BY
       yearID
    HAVING
       G_batting IS NOT NULL;")

ggplot(homerun_yr, aes(x = year, y = home_runs/games)) +
  geom_point() +
  xlab("year") + ylab("home runs") +
  ggtitle("Avg Home Runs per Game by Year")
```



Question 16

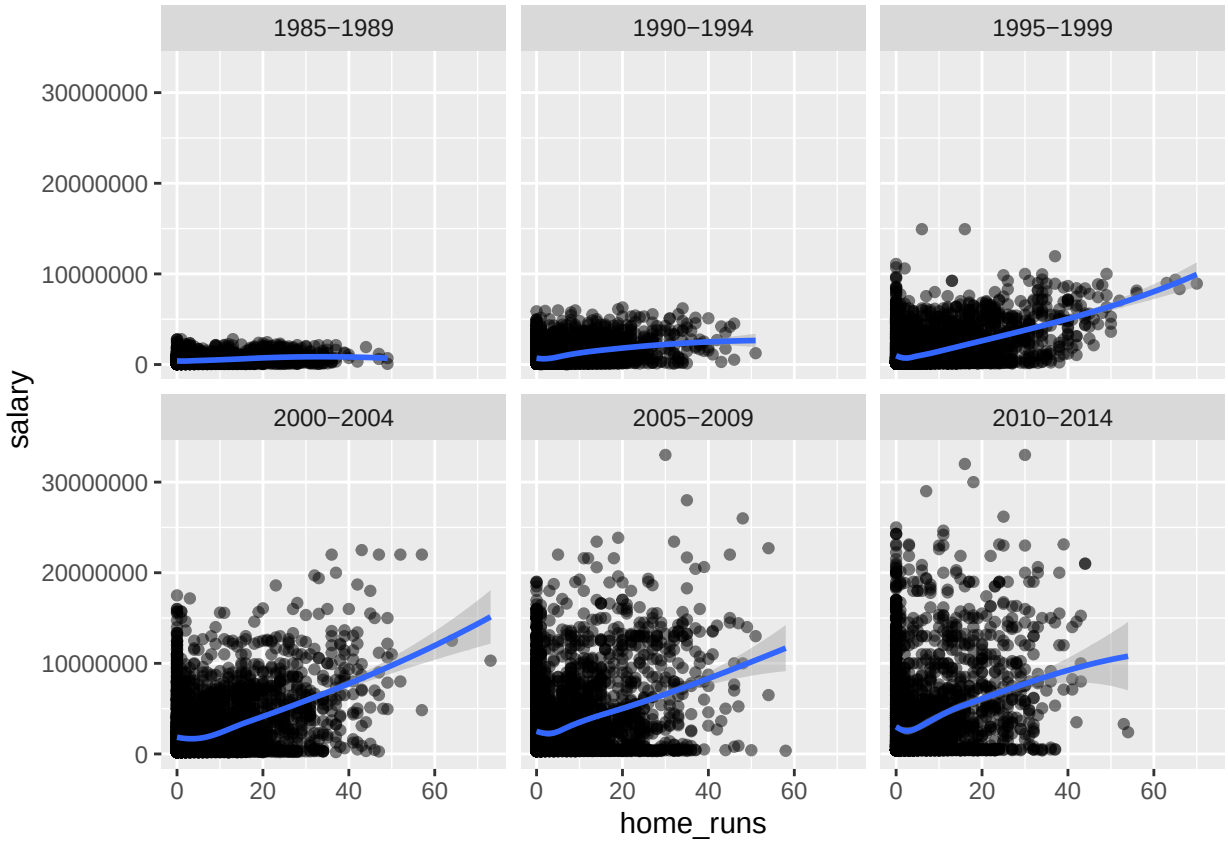
Do players who hit more home runs receive higher salaries?

In this query, I collected the number of home runs and the salary of each player by years. Since there is a significantly huge number of years of data being recorded and also the unfairness due to salary inflation, I decided to conduct this analysis based on 5 year periods.

Due to the `Salaries` data only being available starting from the 1980s, all data points prior to that decade has been ignored. As we can see in the plot below, it is arguably not the case that players who hit more home runs receive higher salaries prior to the year 1994. After that, it seems that in general players who hit more home runs receive higher salaries, especially when looking at recent years. Of course, there are plenty of players with high salaries without hitting much home runs, but this could be due to their other contributions to their teams.

```
homerun_salary = dbGetQuery(db,
    "SELECT
        B.yearID AS year,
        B.HR AS home_runs,
        S.salary AS salary
     FROM
        Batting AS B
     INNER JOIN
        Salaries AS S
     ON
        B.playerID = S.playerID AND
        B.yearID = S.yearID;")

homerun_salary$period = ifelse(homerun_salary$year %% 10 < 5,
                               paste0(homerun_salary$year %/% 10, "0-",
                                      homerun_salary$year %/% 10, "4"),
                               paste0(homerun_salary$year %/% 10, "5-",
                                      homerun_salary$year %/% 10, "9"))
options(scipen = 7)
ggplot(data = homerun_salary,
       aes(x = home_runs, y = salary, na.rm = TRUE)) +
  geom_point(alpha = 0.5, na.rm = TRUE) +
  geom_smooth(method = "loess", formula = y ~ x, na.rm = TRUE) +
  facet_wrap(vars(period))
```



Question 17

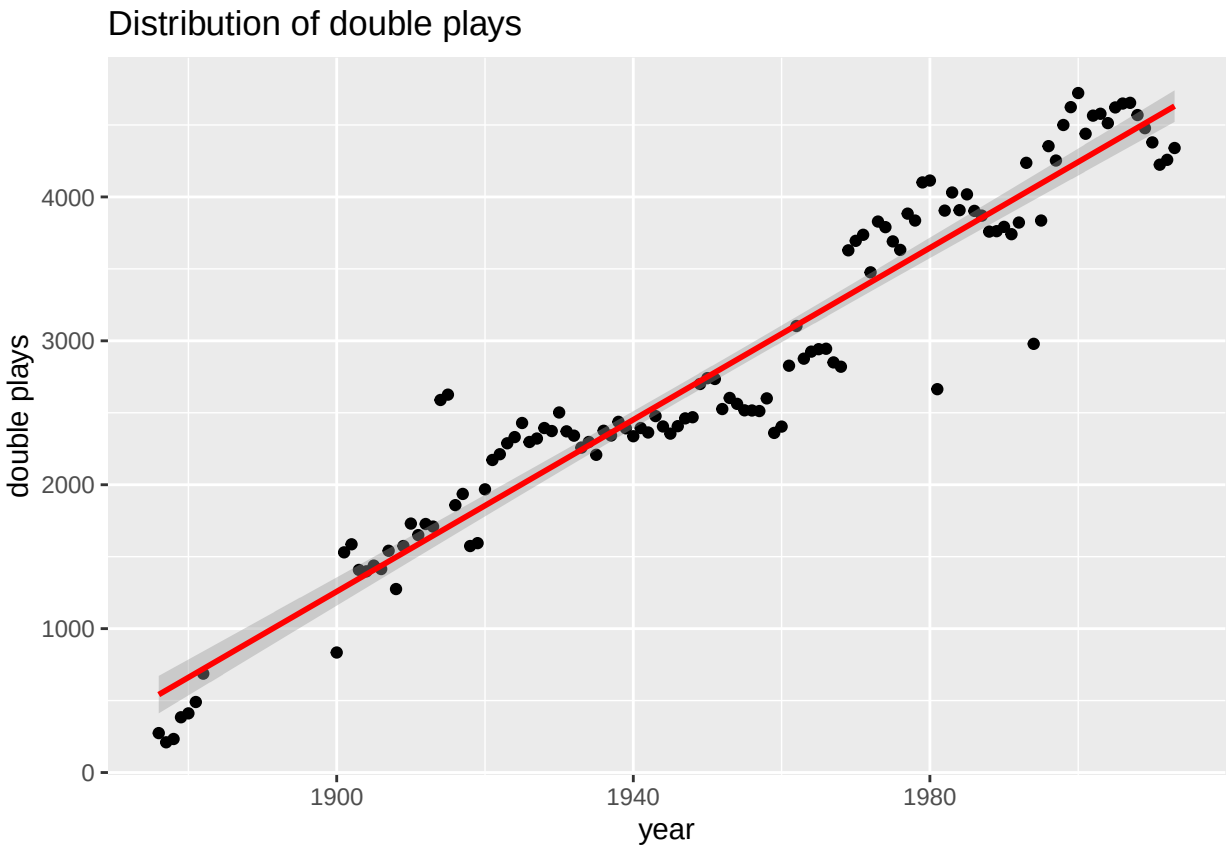
What is the distribution of double plays? Triple plays?

The approach to this problem is relatively simple, as all the data that we needed are in the `Teams` table. All we have to do is to sum up the total amount of double plays by year. However, it is important to note that there are some earlier years that that do not contain any data on double plays. Those years are subsequently removed.

The plot of total double plays vs year is shown below. We can see that the number of double plays has been steadily increasing in a linear fashion over the years.

```
dbl_plays = dbGetQuery(db,
    "SELECT
       yearID AS year,
       SUM(DP) as total_double_plays
     FROM
       Teams
     WHERE
       DP IS NOT NULL
     GROUP BY
       year
     ORDER BY
       year;")

ggplot(dbl_plays, aes(x = year, y = total_double_plays)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x, col = "red") +
  xlab("year") + ylab("double plays") +
  ggtitle("Distribution of double plays")
```



Question 18

What pitchers have a large number of double or triple plays? Again, give their details (names, team, year, . . .).

We get this data from the `Pitching` table. We need to join 3 tables together to answer this question since the names and teams of the players are from separate tables.

The top 10 pitchers with the largest number of double plays are shown below. It is interesting to see that all of the data come from the year 2013, as the years before did not provide any double plays data.

```
dbGetQuery(db,
    "SELECT
        M.nameFirst || ' ' || M.nameLast AS name,
        T.name AS team,
        P.yearID AS year,
        GIDP AS double_plays
     FROM
        Pitching AS P
     JOIN
        Master AS M
     ON
        P.playerID = M.playerID
     JOIN
        Teams AS T
     ON
```

```
        P.teamID = T.teamID AND
        P.yearID = T.yearID
     ORDER BY
        GIDP DESC
     LIMIT 10;")
```

```
##               name                 team year double_plays
## 1  Adam Wainwright  St. Louis Cardinals 2013           32
## 2   Jhoulys Chacin     Colorado Rockies 2013           30
## 3    Kevin Correia      Minnesota Twins 2013           26
## 4      Doug Fister       Detroit Tigers 2013           26
## 5     Hyun-jin Ryu  Los Angeles Dodgers 2013           26
## 6     Joe Saunders     Seattle Mariners 2013           26
## 7  Felix Hernandez     Seattle Mariners 2013           24
## 8    Lucas Harrell       Houston Astros 2013           23
## 9       Mike Leake      Cincinnati Reds 2013           23
## 10      Wade Miley Arizona Diamondbacks 2013           23
```

Question 19

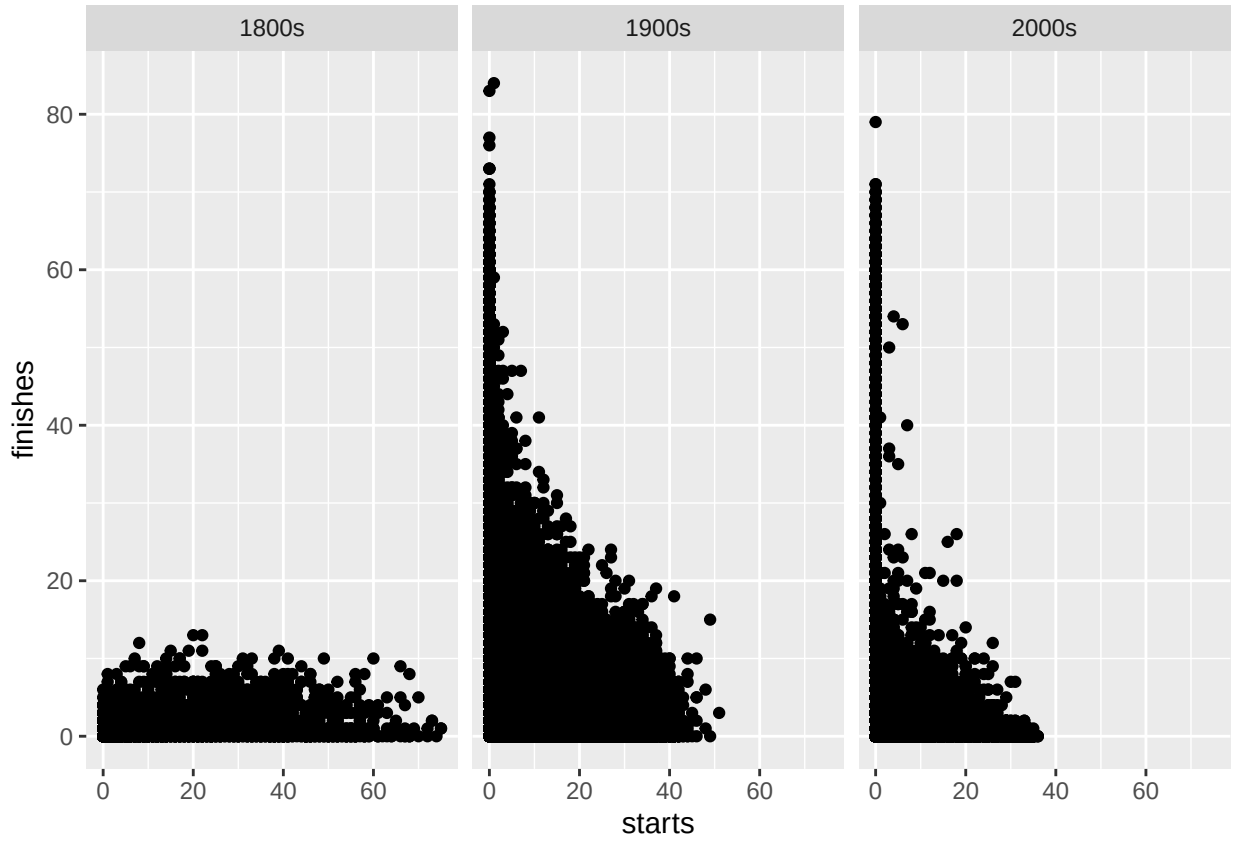
How many games do pitchers start in a season? Plot this against games finished in a season.

The following query will be used to answer this question and Questio 21. This query is just a simple query of obtaining the necessary fields from the `Pitching` table.

From the plot below, we see that in the 1800s most players have a high number of starts when compared to the games they finished. Note that there are certain years before 1900 where the data on the game finishes of players are missing. However, the trend seemed to have changed in the 1900s as there were a lot of players with close to zero starts but have a huge amount of finishes.

```
pitchers = dbGetQuery(db,
    "SELECT
        yearID AS year,
        G AS games,
        W AS wins,
        GS AS starts,
        GF AS finishes
     FROM
        Pitching;")

pitchers$century = paste0(pitchers$year %/% 100, "00s")
ggplot(pitchers, aes(x = starts, y = finishes, na.rm = TRUE)) +
  geom_point(na.rm = TRUE) + facet_wrap(vars(century))
```

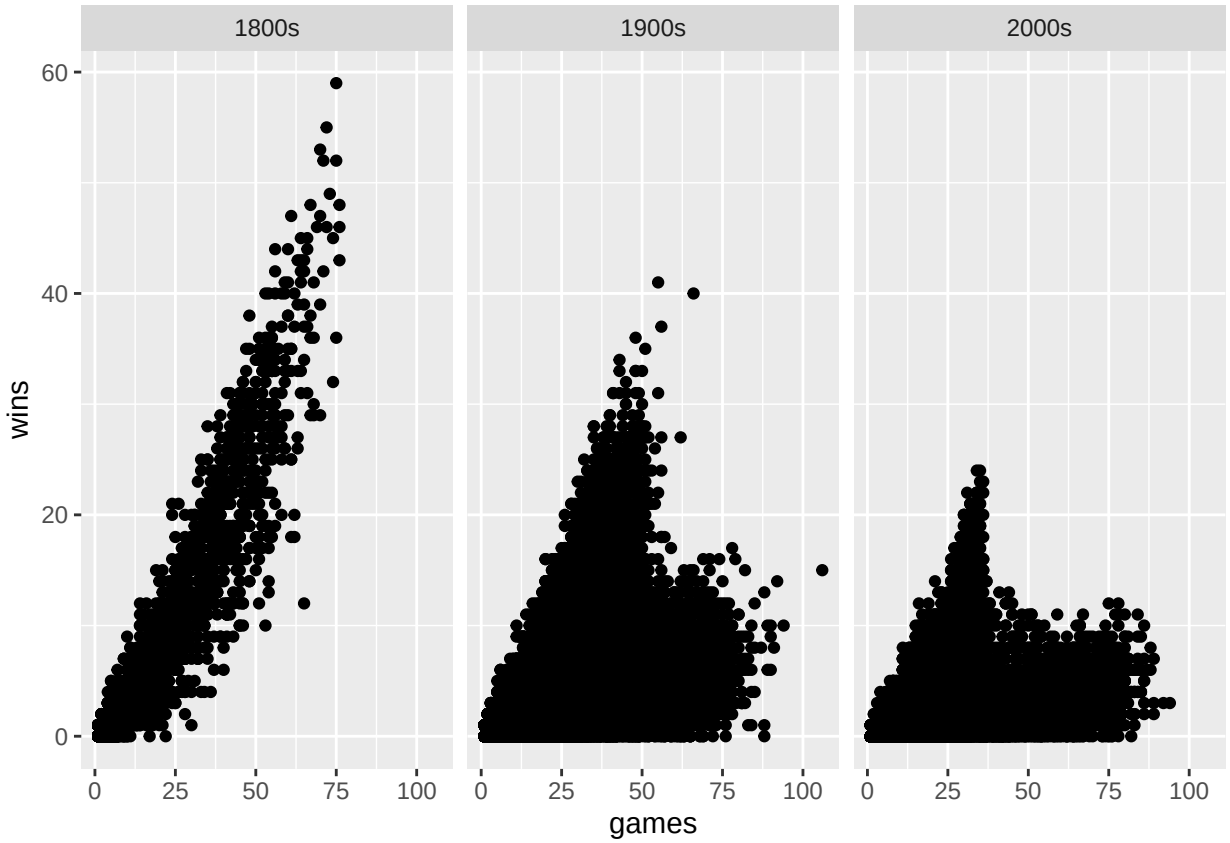


Question 20

How many games do pitchers win in a season?

Based on the following plot, it seems that the roles of pitchers are less effective as they were in the past. In the 18th century, the distibution of a pitcher's wins in a season seems to be linear to the number of games played. However, these became less of a case as the years went by.

```
ggplot(pitchers, aes(x = games, y = wins)) +
  geom_point() +
  facet_wrap(vars(century))
```



Question 21

How are wins related to hits, strikeouts, walks, homeruns and earned runs?

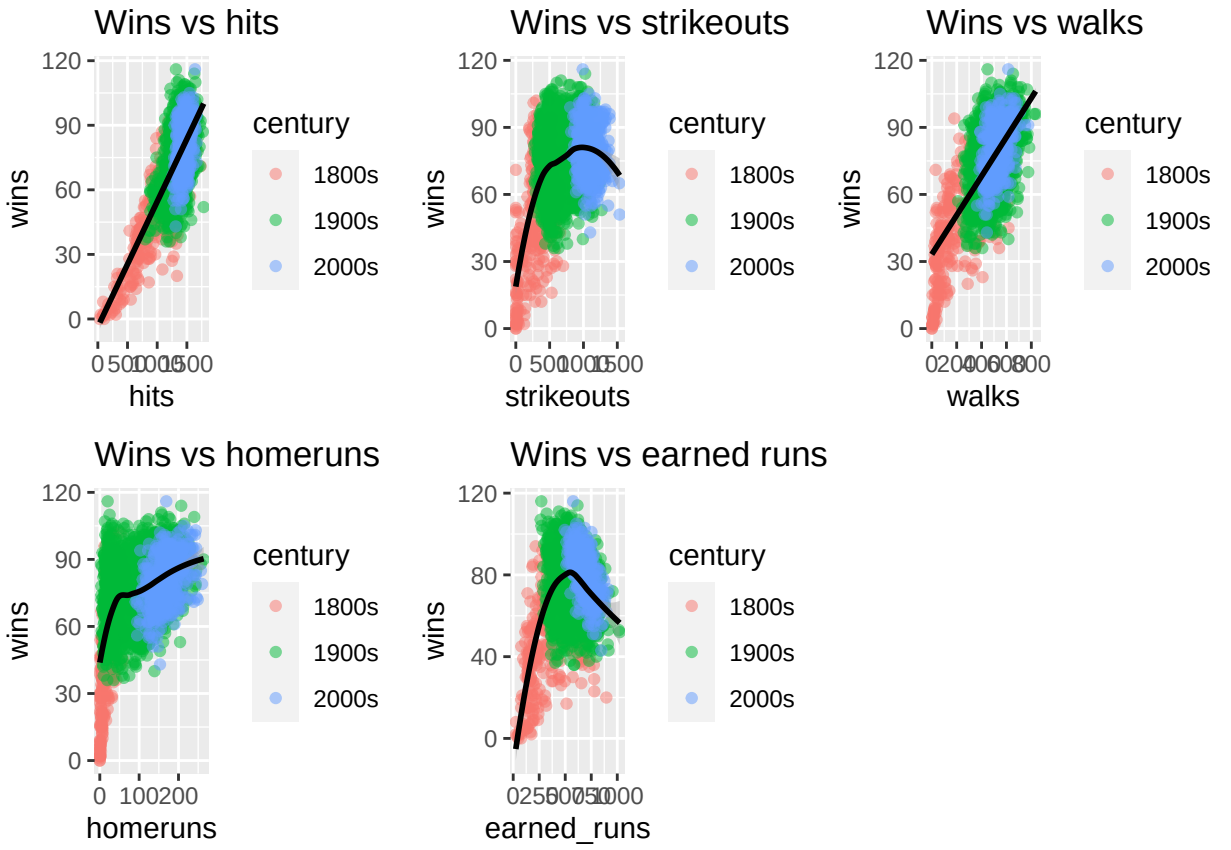
This problem is once again relatively easy to solve since all the required data is in the `Teams` table. Note that in this query, there are a number of null values in the strikeout columns for the years between 1903 and 1912. After obtaining the data, we plot wins against each predictor variable and observe their relationship.

Based on the plot below, it seems that the number of hits, walks and homeruns are the best predictors for the amount of wins as there is a more significant relationship between these variables and the number of wins.

Another interesting thing that we find is that the number of wins in the increases by century. This can be explained by the increasing number of games played by each team throughout the years.

```
win_stats = dbGetQuery(db,
    "SELECT
        yearID AS year,
        W AS wins,
        H AS hits,
        SO AS strikeouts,
        BB AS walks,
        HR AS homeruns,
        ER AS earned_runs
     FROM
        Teams;")
```

```
win_stats$century = paste0(win_stats$year %/% 100, "00s")
plt1 = ggplot(win_stats, aes(x = hits, y = wins, col = century)) +
  geom_point(alpha = 0.5) + ggtitle("Wins vs hits") +
  geom_smooth(method = "lm", col = "black")
plt2 = ggplot(win_stats, aes(x = strikeouts, y = wins, col = century, na.rm = TRUE)) +
  geom_point(alpha = 0.5, na.rm = TRUE) + ggtitle("Wins vs strikeouts") +
  geom_smooth(method = "loess", col = "black", na.rm = TRUE)
plt3 = ggplot(win_stats, aes(x = walks, y = wins, col = century)) +
  geom_point(alpha = 0.5) + ggtitle("Wins vs walks") +
  geom_smooth(method = "lm", col = "black")
plt4 = ggplot(win_stats, aes(x = homeruns, y = wins, col = century)) +
  geom_point(alpha = 0.5) + ggtitle("Wins vs homeruns") +
  geom_smooth(method = "loess", col = "black")
plt5 = ggplot(win_stats, aes(x = earned_runs, y = wins, col = century)) +
  geom_point(alpha = 0.5) + ggtitle("Wins vs earned runs") +
  geom_smooth(method = "loess", col = "black")
ggpubr::ggarrange(plt1, plt2, plt3, plt4, plt5, nrow = 2, ncol = 3)
```



Question 22

What are the top ten collegiate producers of major league baseball players? How many colleges are represented in the database?

To find the top ten collegiate producers of major league baseball players, we `COUNT` the players `GROUP BY` their schools, and then `ORDER BY` the number of players from each school. The results are shown below:

```
dbGetQuery(db,
    "SELECT
       S.schoolName,
       COUNT(SP.playerID) AS count
     FROM
       SchoolsPlayers AS SP
     JOIN
       Schools AS S
     ON
       SP.schoolID = S.schoolID
     GROUP BY
       SP.schoolID
     ORDER BY
       count DESC
     LIMIT
       10;")
```

```
##                                  schoolName count
## 1           University of Southern California   102
## 2               University of Texas at Austin   100
## 3                    Arizona State University    98
## 4                         Stanford University    82
## 5                       University of Michigan    77
## 6                    College of the Holy Cross    75
## 7                     University of Notre Dame    70
## 8  University of Illinois at Urbana-Champaign    68
## 9       University of California, Los Angeles    66
## 10                       University of Arizona    66
```

To find the total number of colleges represented in the database, we do a `DISTINCT` count on the `schoolID` field and obtain the result below, ie. there are 713 colleges in this database.

```
dbGetQuery(db,
    "SELECT
       COUNT(DISTINCT schoolID) AS number_of_colleges
     FROM
       SchoolsPlayers;")
```

```
##   number_of_colleges
## 1                713
```

Question 23

What players have pitched in the post season and also hit a home run in their career?

For this question, we are post season related tables, ie. `PitchingPost` and `BattingPost`. I joined both tables on `playerID` and obtained the number of games they have played as pitchers and the number of homeruns they have. Since the tables provide data for mulitple years instead of the total games played and the total number of homeruns, I needed to group the tables by `playerID` and obtain the `SUM` of games played as pitchers and `SUM` of homeruns. Lastly, I also joined the `Master` table to obtain the names of those players. The results are shown below:

```
dbGetQuery(db,
    "SELECT
       M.nameFirst || ' ' || M.nameLast AS name,
       SUM(P.G) AS 'games_as_pitcher',
       SUM(B.HR) AS homeruns
     FROM
       PitchingPost AS P
     JOIN
       BattingPost AS B
     ON
       P.playerID = B.playerID
     JOIN
       Master AS M
     ON
       M.playerID = P.playerID
     GROUP BY
       P.playerID
     HAVING
       homeruns > 0
     ORDER BY
       homeruns DESC;")
```

```
##              name games_as_pitcher homeruns
## 1       Babe Ruth               30       30
## 2    Dave McNally               98       18
## 3     Don Gullett              180       11
## 4   Steve Carlton              144       10
## 5     Joe Blanton               80        9
## 6    Mike Cuellar               72        8
## 7     Jeff Suppan               80        8
## 8    Ken Holtzman               52        7
## 9      Kerry Wood               75        7
## 10     Bob Gibson               27        6
## 11     Dave Foutz               50        4
## 12   Jesse Haines               24        4
## 13  John Clarkson               24        3
## 14      Rosy Ryan               18        3
## 15    Rick Ankiel               12        2
## 16   Jack Bentley               10        2
## 17   Lew Burdette               12        2
## 18       Ed Crane               14        2
## 19   Mudcat Grant                8        2
## 20  Mickey Lolich               10        2
## 21 Rick Sutcliffe                6        2
## 22  Bucky Walters                8        2
## 23      Jim Bagby                2        1
## 24    Bill George                1        1
## 25  Jose Santiago                3        1
```