

DSCI401_HW1_LexBrunett

Lex Miguel Brunett Chavez

2023-09-07

R Markdown Homework 1

Homework 1 should be submitted as an R Markdown file with links to Google colab notes where necessary. Homework should be turned in on Sakai, all the code and other copies are in my github repository of the class.

```
cubs <- read.csv("https://raw.githubusercontent.com/gjm112/DSCI401/main/data/cubs_all_time.csv")
```

1. How many total games have the Cubs won and lost between 1876 and 2022?

In the homework we have three questions to answer showing some code and writing to explain:

We have to sum the values from 1876 to 2022 (no filter in the years) using the columns W and L for Wins and Losses respectively.

```
Cubs_Wins = sum(cubs$W)
Cubs_Loses = sum(cubs$L)
```

Between 1876 and 2022, Chicago Cubs obtain **11141** wins and **10592** losses.

2. What year did the Cubs score the most runs? What year did the Cubs score the fewest runs? Do you have any thoughts about the year that the Cubs scored the fewest runs?

For the next question we need to know how was the best (maximum value) and the worst (minimum value) of the runs during the years of the Chicago cubs:

After we obtain the values we can make some observations of the seasons.

```
Cubs_Best_Runs = cubs[which.max(cubs$R), "Year"]
Cubs_Worst_Runs = cubs[which.min(cubs$R), "Year"]
```

The most runs season of the Chicago Cubs history occurred in **1894**, in the other hand the fewest runs season of the Chicago Cubs history occurred in **2020**.

It seems that the Chicago cubs suffer the fewest season in 2020, looking the year, the fewest runs season was during pandemic where people suffered some breathing problems and it could be a reason of the decrease of the quantity of the runs in that year.

3. In how many seasons was the Cubs total attendance (i.e. the variable Attendance) over 3 million?

For the last questions we need to know how many seasons was the Cubs total attendance exceed 3 million, we can obtain the value with a iteration and conditioning each value to count the total quantity.

```
Season_num = 0

for (x in cubs$Attendance) {
  if (!is.na(x) & (x >= 3000000)){
    Season_num <- Season_num + 1
  }
}
```

In the Chicago Cubs history, the total amount of seasons which the Cubs recieved over 3 million in attendance is **12** seasons.

Please review the same homework in google colab (Python) following this link: **Google Colab notebook**

And the same file has been uploaded in my Github repository of the class please follow this link: **My GitHub Repository**