

DSCI 401 - Homework 1

Isabel Heard

09/08/2023

Contents

1.Set up a github account and create a repository called DSCI401	1
2. Create an R Markdown file and run the following code:	1
A) How many total games have the Cubs won and lost between 1876 and 2022?	1
B) What year did the Cubs score the most runs?	1
B2) What year did the Cubs score the fewest runs?	2
B3) Do you have any thoughts about the year that the Cubs scored the fewest runs?	2
C) In how many seasons was the Cubs total attendance (i.e. the variable Attendance) over 3 million?	2
Python code link	2

1.Set up a github account and create a repository called DSCI401

<https://github.com/iisabel/DSCI-401.git>

2. Create an R Markdown file and run the following code:

```
df <- read.csv("/Users/isabelheard/Downloads/DSCI 401/DSCI401_git/data/cubs_all_time.csv")
#head(df)
```

A) How many total games have the Cubs won and lost between 1876 and 2022?

```
sum(df$W)
```

```
## [1] 11141
```

```
sum(df$L)
```

```
## [1] 10592
```

11141 games won between 1876 and 2022.

10592 games lost between 1876 and 2022.

B) What year did the Cubs score the most runs?

```
#max(df$R)
which.max(df$R)
```

```
## [1] 129
```

```
df[129, ]
```

```
##      Year      Tm Lg   G  W  L Ties  W.L.  pythW.L.   Finish  GB Playoffs
## 129 1894 Chicago Colts NL 137 57 75    5 0.432    0.49 8th of 12 34.0
##      R    RA Attendance BatAge PAge X.Bat X.P      Top.Player      Managers
## 129 1056 1080    239000    27.3 26.6    22   9 C.Griffith (5.9) C.Anson (57-75)
```

The cubs scored the most runs (1056) in 1894.

B2) What year did the Cubs score the fewest runs?

```
#min(df$R)
which.min(df$R)
```

```
## [1] 3
```

```
df[3, ]
```

```
##      Year      Tm      Lg   G  W  L Ties  W.L.  pythW.L.   Finish GB
##   3 2020 Chicago Cubs NL Central 60 34 26    0 0.567    0.545 1st of 5 --
##      Playoffs   R    RA Attendance BatAge PAge X.Bat X.P      Top.Player
##   3 Lost NLWC (2-0) 265 240          NA    27.9 30.3    47  26 Y.Darvish (2.8)
##      Managers
##   3 D.Ross (34-26)
```

The cubs scored the least runs (265) in 2020.

B3) Do you have any thoughts about the year that the Cubs scored the fewest runs?

COVID-19 meant that there were fewer games.

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

```
#df$Attendance >= 3000000
length(which(df$Attendance >= 3000000))
```

```
## [1] 12
```

12 seasons the total club attendance was over 3 million.

Python code link

https://colab.research.google.com/drive/1Pg0RwYnDq1VCvyB0_iPm5WyZk1lQ98Cy?usp=sharing