

ICA10Turnin

2023-10-10

```
setwd("~/Desktop/STT810/Data")
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.3      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr    1.5.0
```

```
## v ggplot2    3.4.3      v tibble     3.2.1
```

```
## v lubridate  1.9.3      v tidyr      1.3.0
```

```
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# I return a row x which is index, looking into dropping from command
```

```
movies <- read.csv('movies.csv')
```

```
batting = read.csv('batting.csv')
```

```
# 6.6 what is the movie with the highest mean rating that has been rated at least 30 times?
```

```
movies %>%
```

```
  group_by(title) %>%
```

```
  summarize(mn_rat = mean(rating), n_rat = n()) %>%
```

```
  filter(n_rat >= 30) %>%
```

```
  arrange(mn_rat)
```

```
## # A tibble: 882 x 3
```

```
##   title                                mn_rat n_rat
```

```
##   <chr>                                <dbl> <int>
```

```
## 1 Godzilla (1998)                      1.95    33
```

```
## 2 I Know What You Did Last Summer (1997) 2.11    32
```

```
## 3 Wild Wild West (1999)                 2.21    53
```

```
## 4 Batman & Robin (1997)                 2.21    42
```

```
## 5 Hollow Man (2000)                     2.29    39
```

```
## 6 Flintstones, The (1994)               2.36    42
```

```
## 7 Free Willy (1993)                     2.39    37
```

```
## 8 Striptease (1996)                     2.41    41
```

```
## 9 Coneheads (1993)                     2.42    63
```

```
## 10 Showgirls (1995)                     2.47    35
```

```
## # i 872 more rows
```

6.11 which player has been hit-by-pitch the most number of times?

```
batting %>%  
  group_by(playerID) %>%  
  summarize(hit = sum(HBP)) %>%  
  arrange(desc(hit))
```

```
## # A tibble: 20,469 x 2  
##   playerID   hit  
##   <chr>     <int>  
## 1 jennihu01   287  
## 2 biggicr01   285  
## 3 tucketo01   272  
## 4 baylodo01   267  
## 5 kendaja01   254  
## 6 huntro01    243  
## 7 mcganda01   230  
## 8 utleych01   204  
## 9 rizzoan01   201  
## 10 robinfr02  198  
## # i 20,459 more rows
```

6.12 how many doubles were hit in 1871

```
batting %>%  
  group_by(yearID) %>%  
  summarize(doubles = sum(X2B)) %>%  
  filter(yearID == 1871)
```

```
## # A tibble: 1 x 2  
##   yearID doubles  
##   <int>   <int>  
## 1   1871     434
```

6.13 which team has the omst total number of home runs?

```
batting %>%  
  group_by(teamID) %>%  
  summarize(homeruns = sum(HR)) %>%  
  arrange(desc(homeruns))
```

```
## # A tibble: 149 x 2  
##   teamID homeruns  
##   <chr>     <int>  
## 1 NYA      16785  
## 2 CHN      14818  
## 3 DET      14193  
## 4 BOS      14167  
## 5 PHI      13724  
## 6 CIN      13633  
## 7 CLE      13558  
## 8 SLN      12441  
## 9 CHA      11945  
## 10 PIT     11846  
## # i 139 more rows
```

6.14 whihc player who has played the most in at least 500 games has socres the most numver of runs pe

```
batting %>%
  group_by(playerID)%>%
  summarize(runs = sum(R),nc = n())%>%
  arrange(desc(runs))
```

```
## # A tibble: 20,469 x 3
##   playerID  runs    nc
##   <chr>    <int> <int>
## 1 henderi01 2295    29
## 2 cobbt01  2247    24
## 3 bondsba01 2227    22
## 4 aaronha01 2174    23
## 5 ruthba01  2174    22
## 6 rosepe01  2165    25
## 7 mayswi01  2062    23
## 8 rodrial01 2021    22
## 9 ansonca01 1999    27
## 10 musiast01 1949    22
## # i 20,459 more rows
```

6.15 a. which player has the ost lifetime at bats without ever hitting a home run?

```
batting %>%
  group_by(playerID)%>%
  summarize(homeruns = sum(HR), atbats = sum(AB))%>%
  filter(homeruns == 0)%>%
  arrange(desc(atbats))
```

```
## # A tibble: 12,149 x 3
##   playerID homeruns atbats
##   <chr>      <int> <int>
## 1 eggleda01      0  2544
## 2 mcgeami01      0  2481
## 3 holbebi01      0  2335
## 4 bondto01       0  1975
## 5 oliveto01      0  1931
## 6 hallir01       0  1904
## 7 holdsji01      0  1484
## 8 deaslpa01      0  1466
## 9 lathaju01      0  1431
## 10 waltero01     0  1426
## # i 12,139 more rows
```

b. whihc active player has the most lifetime at bats wihout ever hittings a homerun?