Ksooklall_Homework2

Kenan Sooklall

2/11/2021

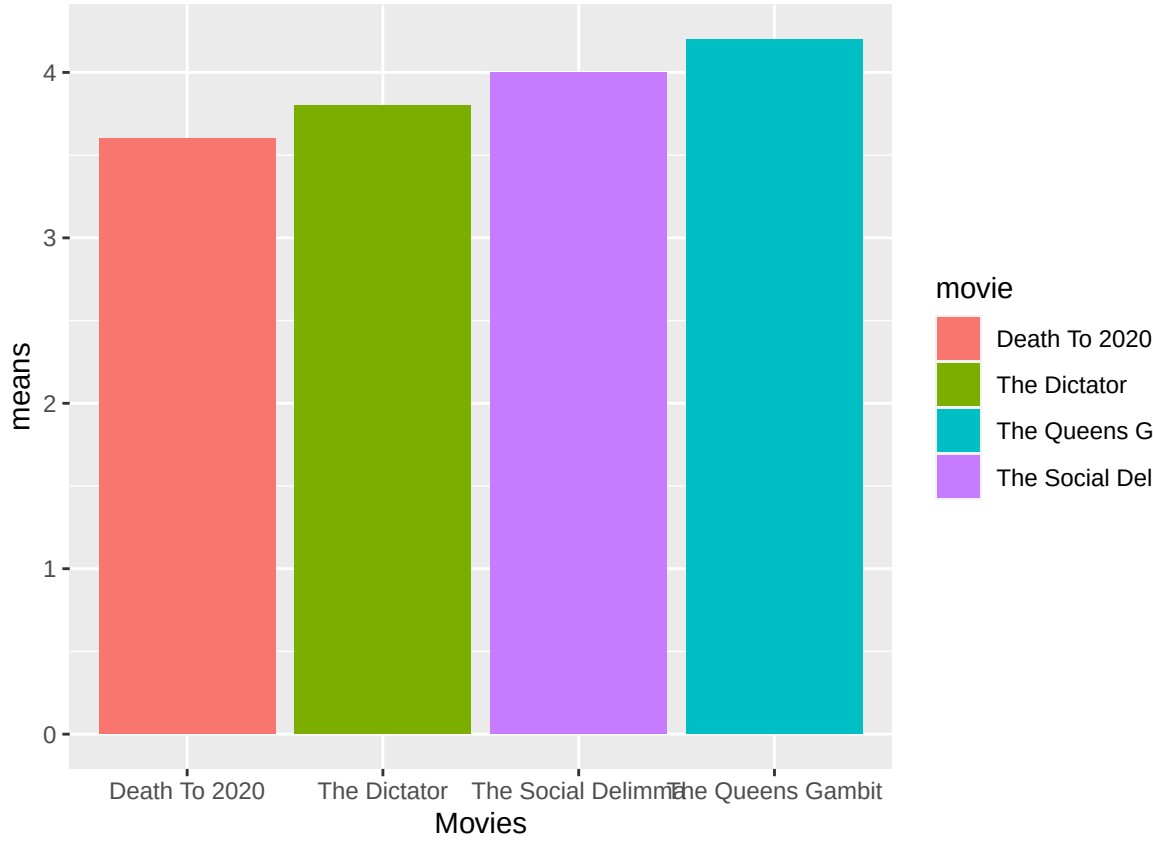
R and PostgreSQL

Connect and pull a table into a data frame from postgres

```
conn <- dbConnect(RPostgres::Postgres(), user='kenan', password='password', dbname='postgres', host='lo
sql <- 'SELECT * from movie_ratings'
df <- dbGetQuery(conn, sql)
disconnected <- dbDisconnect(conn)
df
```

```
##    id              movie reviewer ratings
## 1   1 The Social Delimma    Kenan       4
## 2   2 The Social Delimma   Barbie       4
## 3   3 The Social Delimma   Prince       4
## 4   4 The Social Delimma      Fay       4
## 5   5 The Social Delimma    Kelly       4
## 6   6      Death To 2020    Kenan       4
## 7   7      Death To 2020   Barbie       4
## 8   8      Death To 2020   Prince       4
## 9   9      Death To 2020      Fay       3
## 10 10      Death To 2020    Kelly       3
## 11 11  The Queens Gambit    Kenan       5
## 12 12  The Queens Gambit   Barbie       4
## 13 13  The Queens Gambit   Prince       4
## 14 14  The Queens Gambit      Fay       4
## 15 15  The Queens Gambit    Kelly       4
## 16 16       The Dictator    Kenan       3
## 17 17       The Dictator   Barbie       4
## 18 18       The Dictator   Prince       5
## 19 19       The Dictator      Fay       3
## 20 20       The Dictator    Kelly       4
```

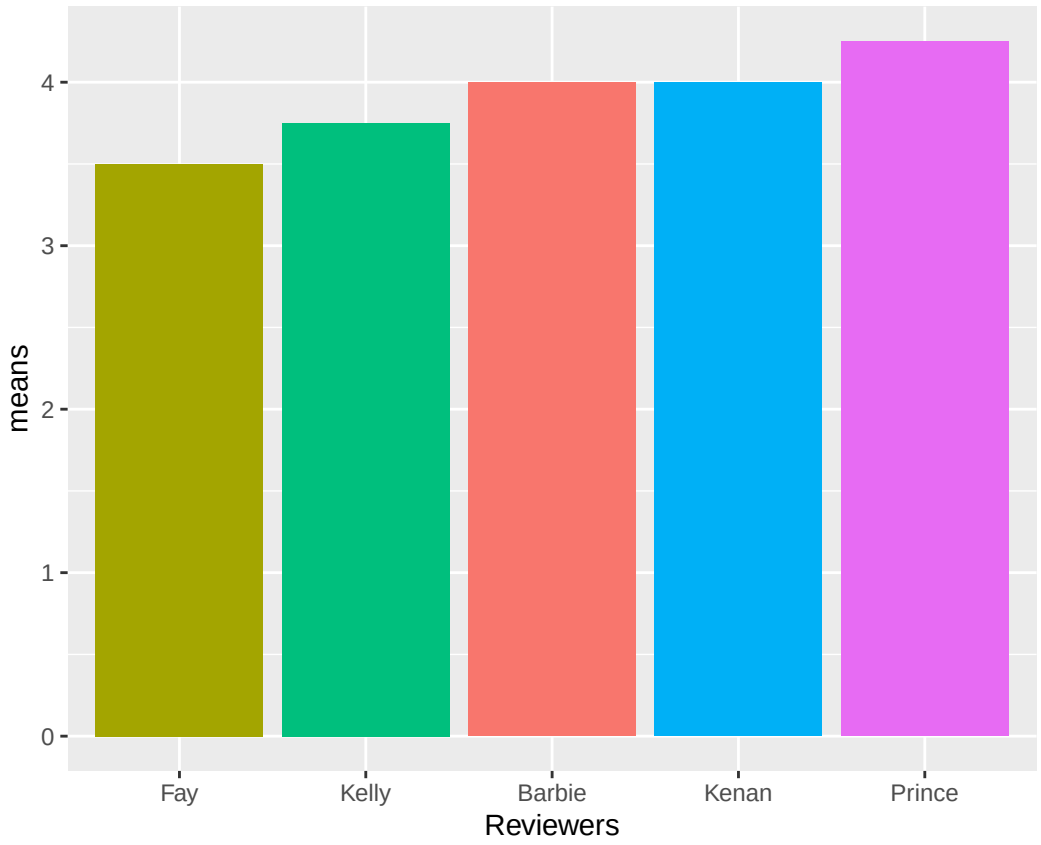
```
df %>% group_by(movie) %>% summarise(means = mean(ratings), .groups='drop') %>%
ggplot(aes(x=reorder(movie, means), y=means, fill=movie)) +
  geom_col() + labs(x= 'Movies')
```



Rating distribution

The most loved movie from my group is the queens gambit

```
df %>% group_by(reviewer) %>% summarise(means = mean(ratings), .groups='drop') %>%
ggplot(aes(x=reorder(reviewer, means), y=means, fill=reviewer)) +
  geom_col()  + labs(x='Reviewers')
```



Reviewer distribution

The easiest person to please here is Prince, he likes everything.

```
mdf <- df %>%
  group_by(movie) %>%
  summarise(min=min(ratings),
            mean=mean(ratings),
            median=median(ratings),
            max=max(ratings),
            iqr=IQR(ratings), .groups='drop')
mdf
```

Tabular statistics

```
## # A tibble: 4 x 6
##   movie                min  mean median   max   iqr
##   <chr>              <int> <dbl>  <int> <int> <dbl>
## 1 Death To 2020          3   3.6      4     4     1
## 2 The Dictator           3   3.8      4     5     1
## 3 The Queens Gambit      4   4.2      4     5     0
## 4 The Social Delimma     4   4        4     4     0
```