Practical 3 - Advanced graphics

Osama Mahmoud

Instructions

Please note: *The graphs shown at each question is only a suggested plot for the solution. You may want to reproduce it or create a diiferent plot to answer the corresponding question.*

Getting started

To get you familiar with the underlying `ggplot2` concepts, we'll recreate some standard graphics. Some of these plots aren't particularly useful, we are just using them for illustration purposes.

To begin with, load the `ggplot2`

```
library("ggplot2")
```

Next we load the `movies` data set

```
# Details of the movies dataset can be displayed by:
library(ggplot2movies)
data(movies, package="ggplot2movies")
?movies
```

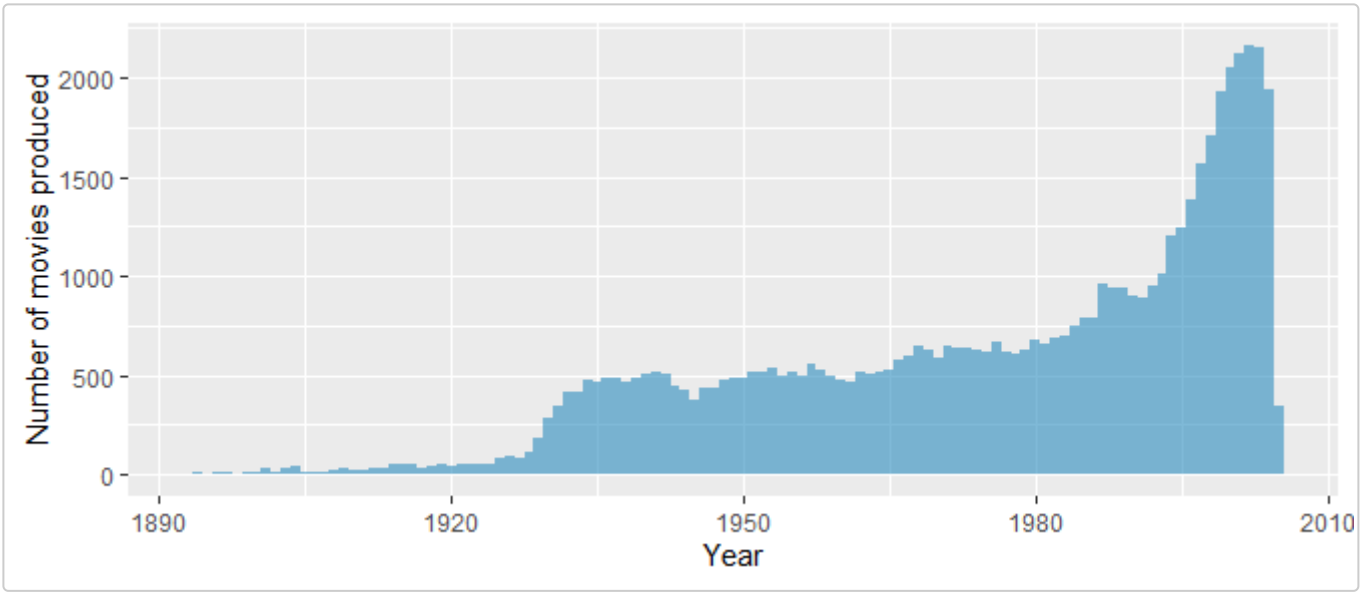
When loading in data, it's a good idea to check some basic characteristics:

```
str(movies)
dim(movies)
names(movies)
head(movies)
```

Plot some information

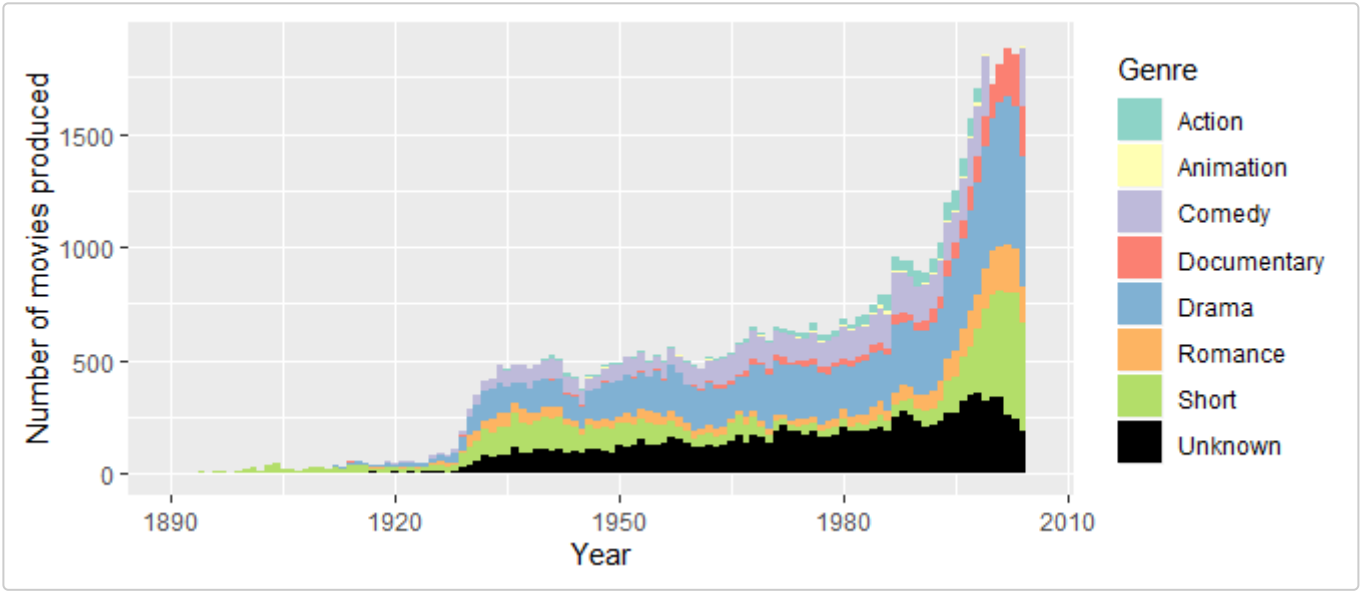
Feel free to experiment with your own ideas. I present some graphs as a reference that you may try to reproduce if you wish.

1. What is the number of movies produced per year?

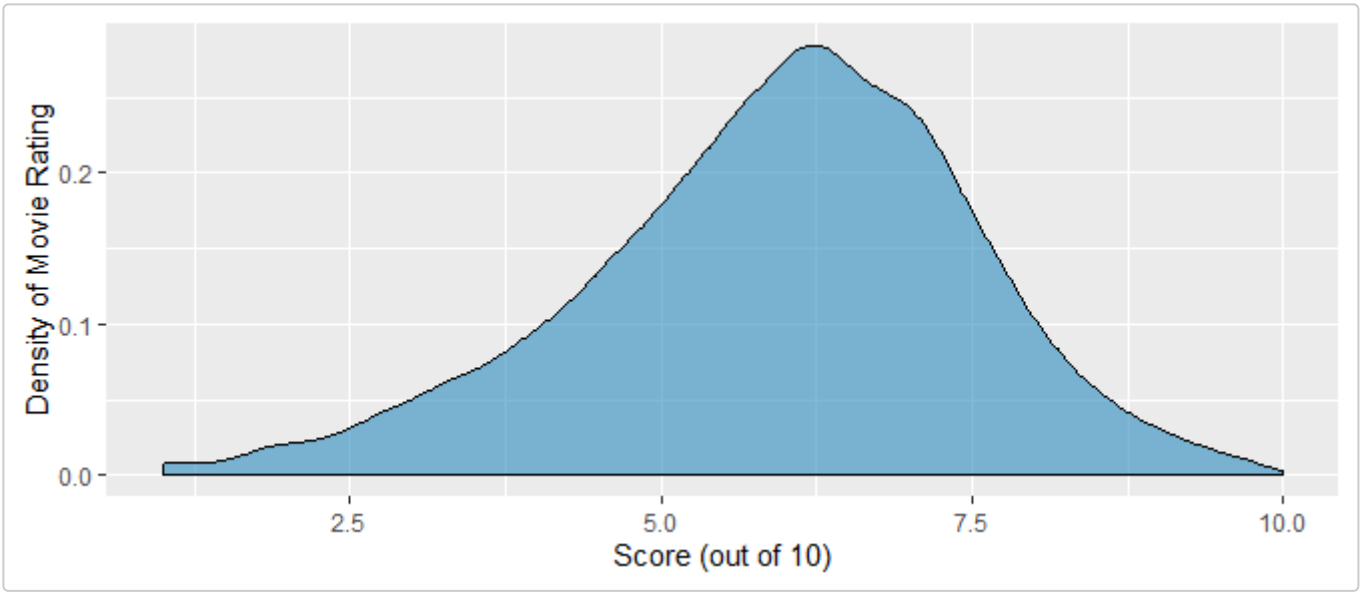


2. What is the number of movies produced per year per genre (action, animation ect)?

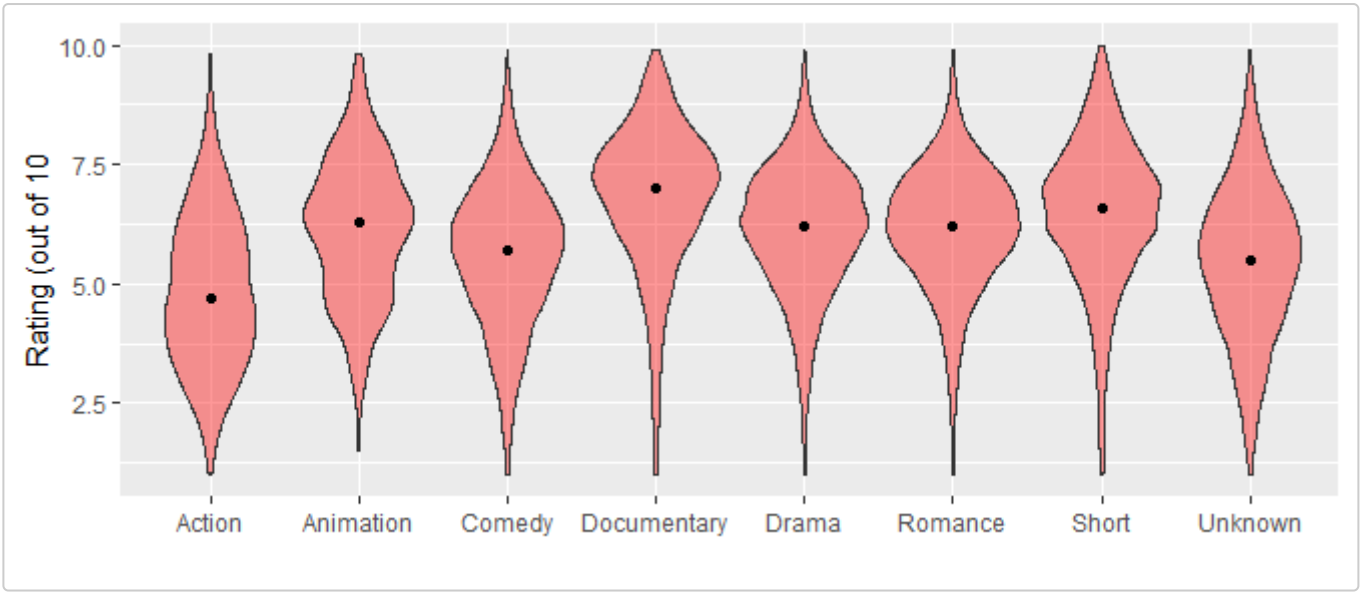
```
# TIP: You need first to create a genre variable:
genre <- rep(0, nrow(movies))
for(i in 18:24)
{
  genre[movies[,i]==1] <- names(movies)[i]
}; genre[genre==0] <- "Unknown"
movies$Genre <- genre
```



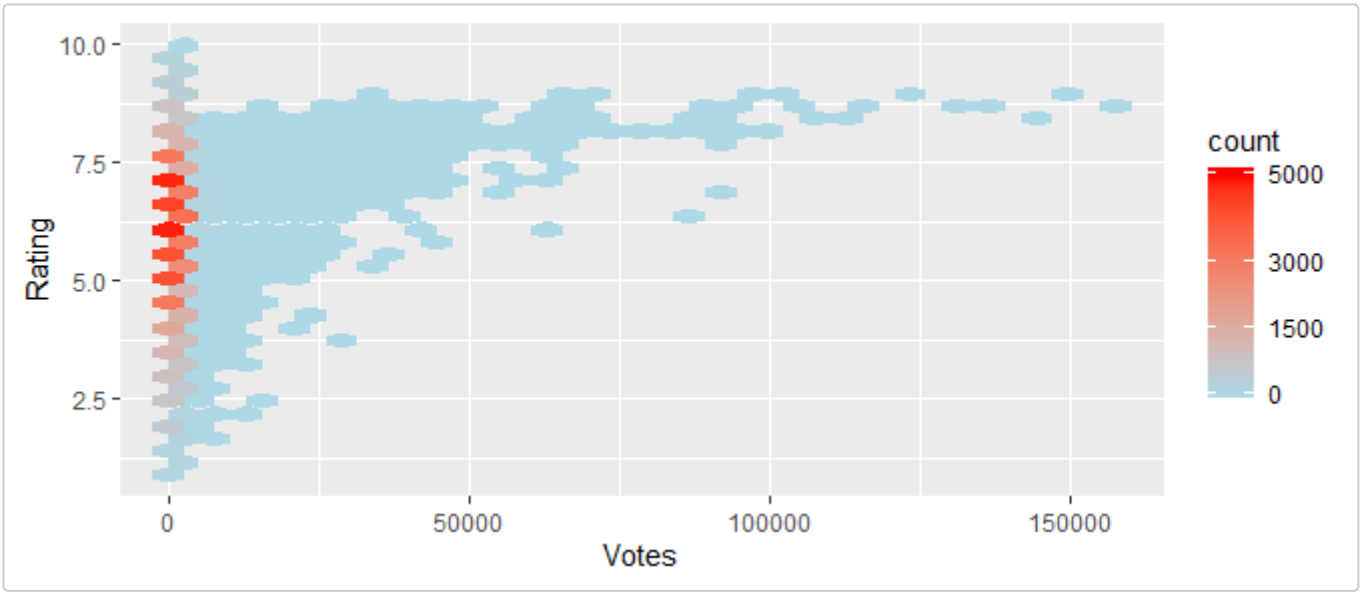
3. Create a graph to present information on the rating of movies.



4. Is there a difference on rating depending on genre? (Hint: you may use `geom_violin`)



5. Is the rating influenced by the number of votes? (Hint: you may use `stat_binhex`)


10.0
7.5
5.0
2.5
Rating
0
50000
100000
150000
Votes
count
5000
3000
1500
0