Answers 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)
```

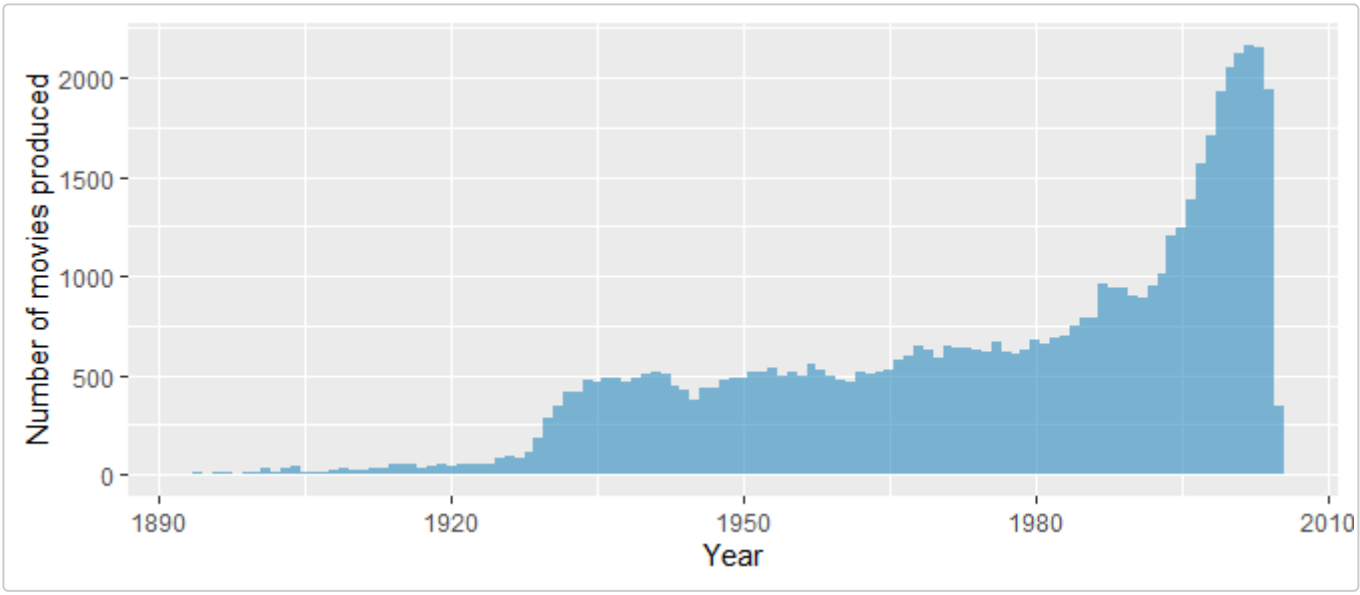
Plot some information

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?

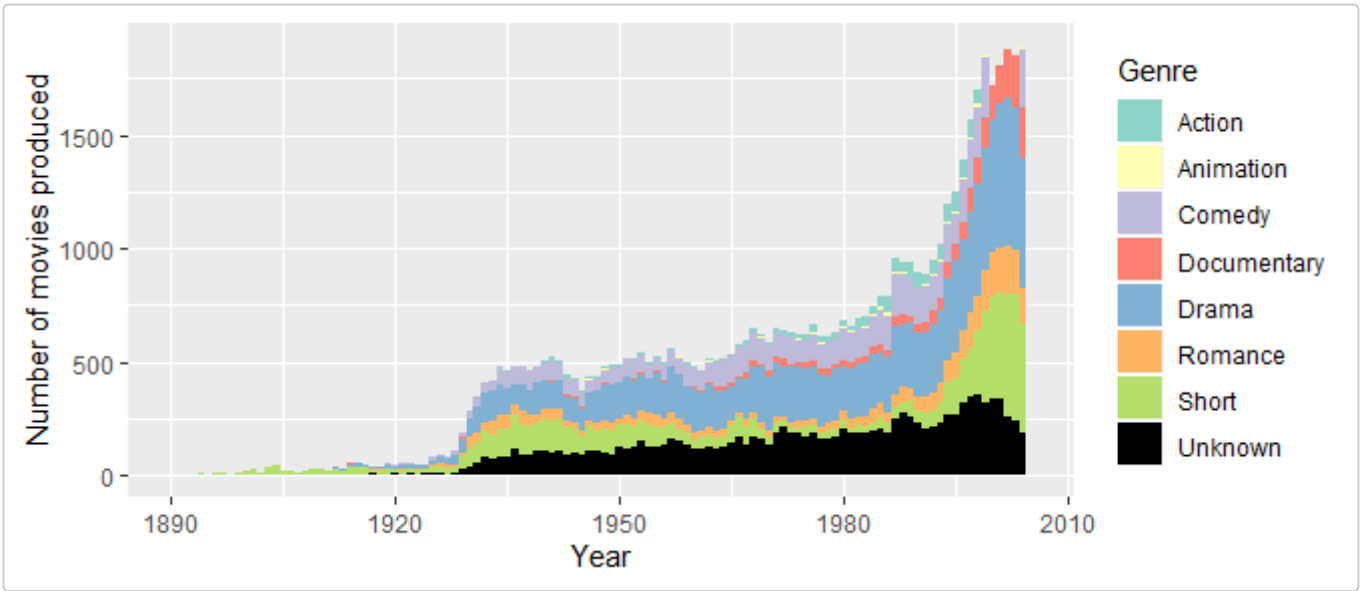
```
g = ggplot(data=movies, aes(x=year))
g1 = g + geom_histogram(binwidth = 1, fill="#2b8cbe", alpha=0.6) +
  xlab("Year") + ylab("Number of movies produced")
```

```
g1
```



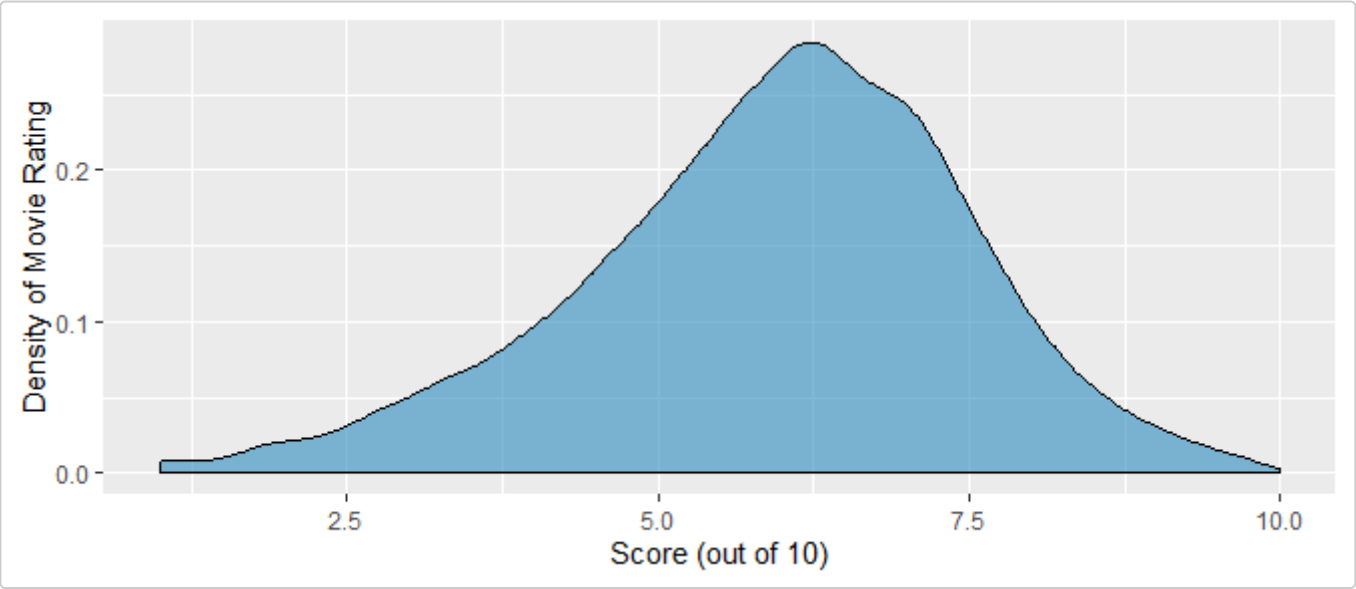
2. What is the number of movies produced per year per genre (action, animation ect)?

```
# 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


# define a vector for colors to be used in the plot
clr <- c('#8dd3c7','#ffffb3','#bebada','#fb8072','#80b1d3','#fdb462','#b3de69', '#000000')

ggplot(movies, aes(x=year, fill=Genre)) + geom_histogram(binwidth=1) + scale_fill_manual(values =
clr) +
  xlab("Year") + ylab("Number of movies produced") + ylim(0,1900) + xlim(1890,2005)
```



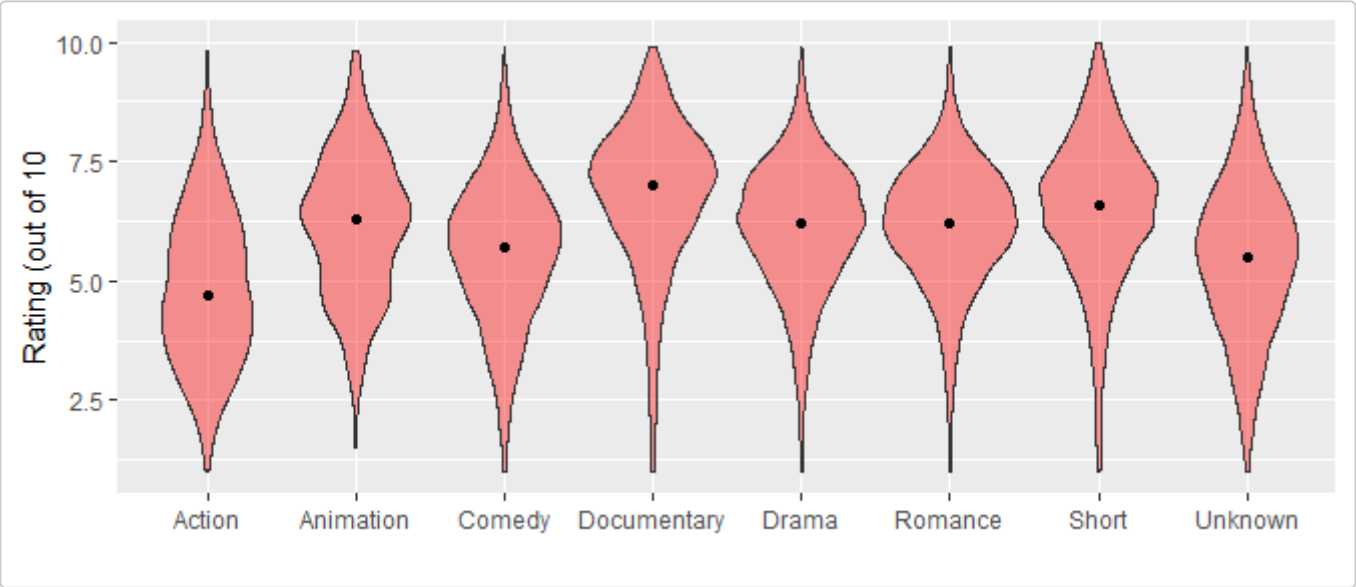
3. Create a graph to present information on the rating of movies.

```
ggplot(movies, aes(x=rating)) + geom_density(fill="#2b8cbe", alpha=0.6) +
  ylab("Density of Movie Rating") + xlab("Score (out of 10)")
```



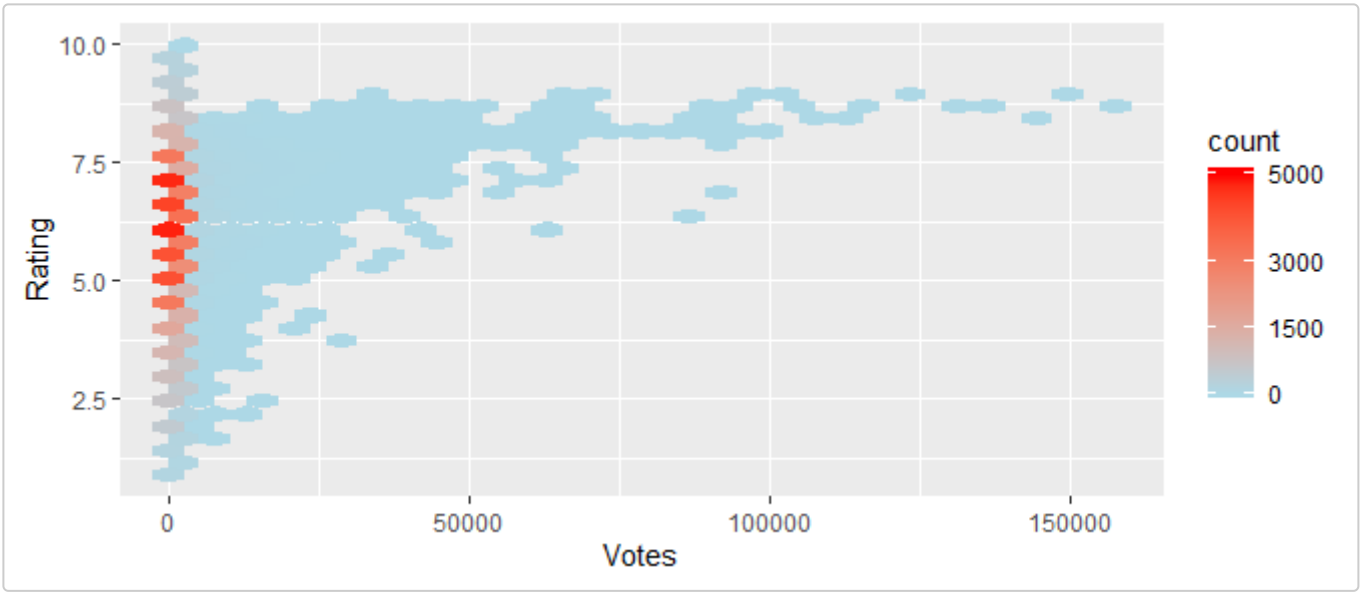
4. Is there a difference on rating depending on genre?

```
ggplot(movies, aes(x=factor(Genre), y=rating)) + xlab("") + ylab("Rating (out of 10") +
  geom_violin(fill="red", alpha=0.4) +
  stat_summary(fun.y = median, geom='point')
```



5. Is the rating influenced by the number of votes?

```
ggplot(movies, aes(x=votes, y=rating)) + xlab("Votes") + ylab("Rating") +
  stat_binhex() +
  scale_fill_gradient(low="lightblue", high="red", breaks=c(0, 1500, 3000, 5000),
                      limits=c(0, 5000))
```


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