

Kairavi Chahal  
36-315 B Exam 1  
Professor Thomas  
February 27, 2013

### Question 2

```
barplot(table(mpaa), main="Bar Plot: MPAA Ratings", ylim=c(0, 400),  
ylab="Number of Movies", xlab="MPAA Rating", col=c("green", "yellow",  
"red", "grey"))
```

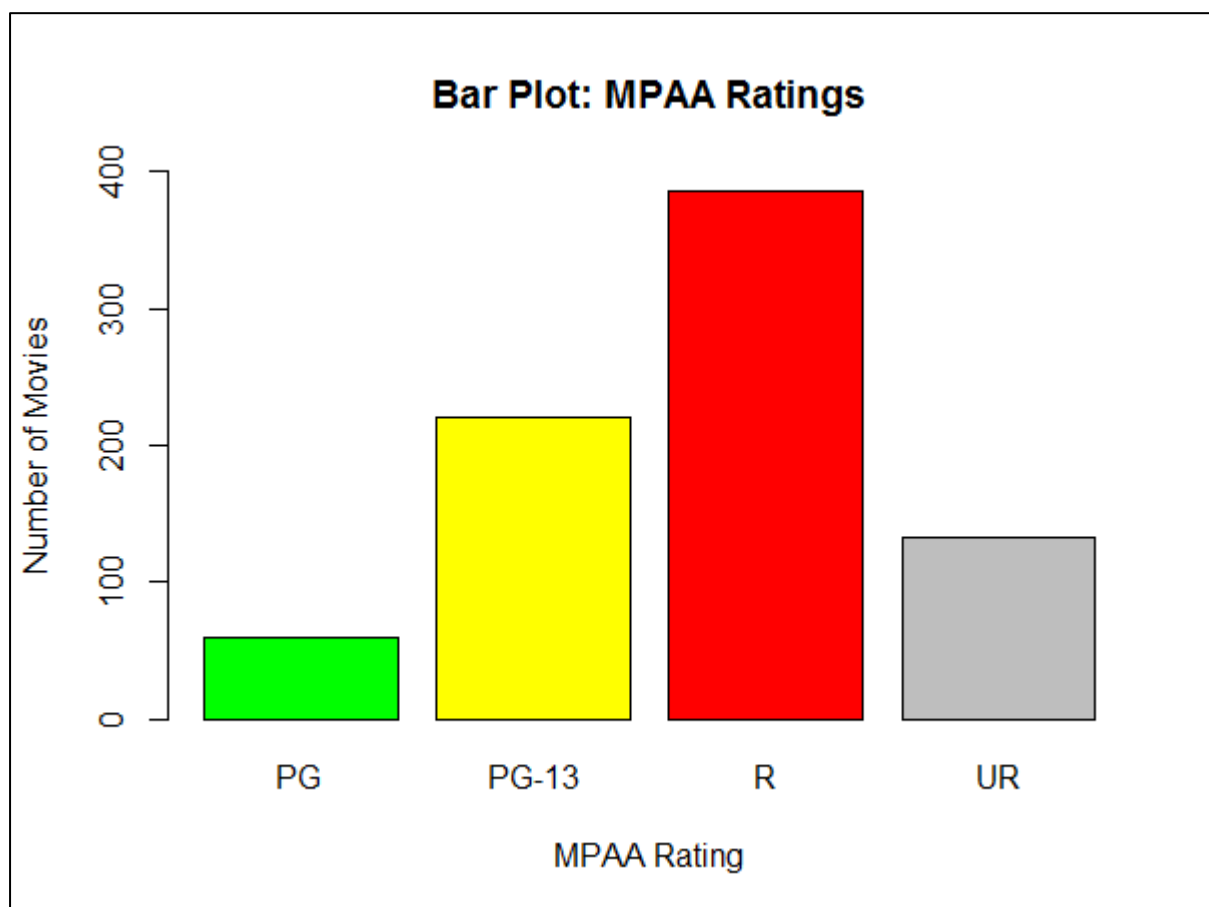
```
table(mpaa)
```

```
mpaa  
   PG PG-13    R    UR  
   60   220   386   132
```

```
sum(table(mpaa))  
[1] 798
```

```
386/798  
[1] 0.4837093
```

```
60/798  
[1] 0.07518797
```



The movies are not evenly distributed by rating. Some ratings seem to be more popular than others. This could either imply that the MPAA is very harsh while awarding ratings or filmmakers prefer to make movies of a certain kind.

The most popular rating is "R."

The least popular rating is "PG."

132 movies were not rated by the MPAA.

Less than half (but almost half) of all films are rated "R."

7.52 percent of all movies were rated "PG."

I chose a bar plot because we are comparing one categorical variable, and it is easiest to see the difference between each category in a bar plot. Other graphs that would have worked for this data are a pie chart or a spine chart. A spine chart would have been most helpful in describing what proportion of movies had a certain rating, which is hard to do in a bar plot, since the bars are split up. You cannot determine the exact percentage from a graph, a table and basic arithmetic would suffice for this.

The parameter choices in this plot are very basic, with the colours representing how kid-friendly the movie is. The y-axis limits were adjusted to range from 0 to the max value of the data.

#### Question 4

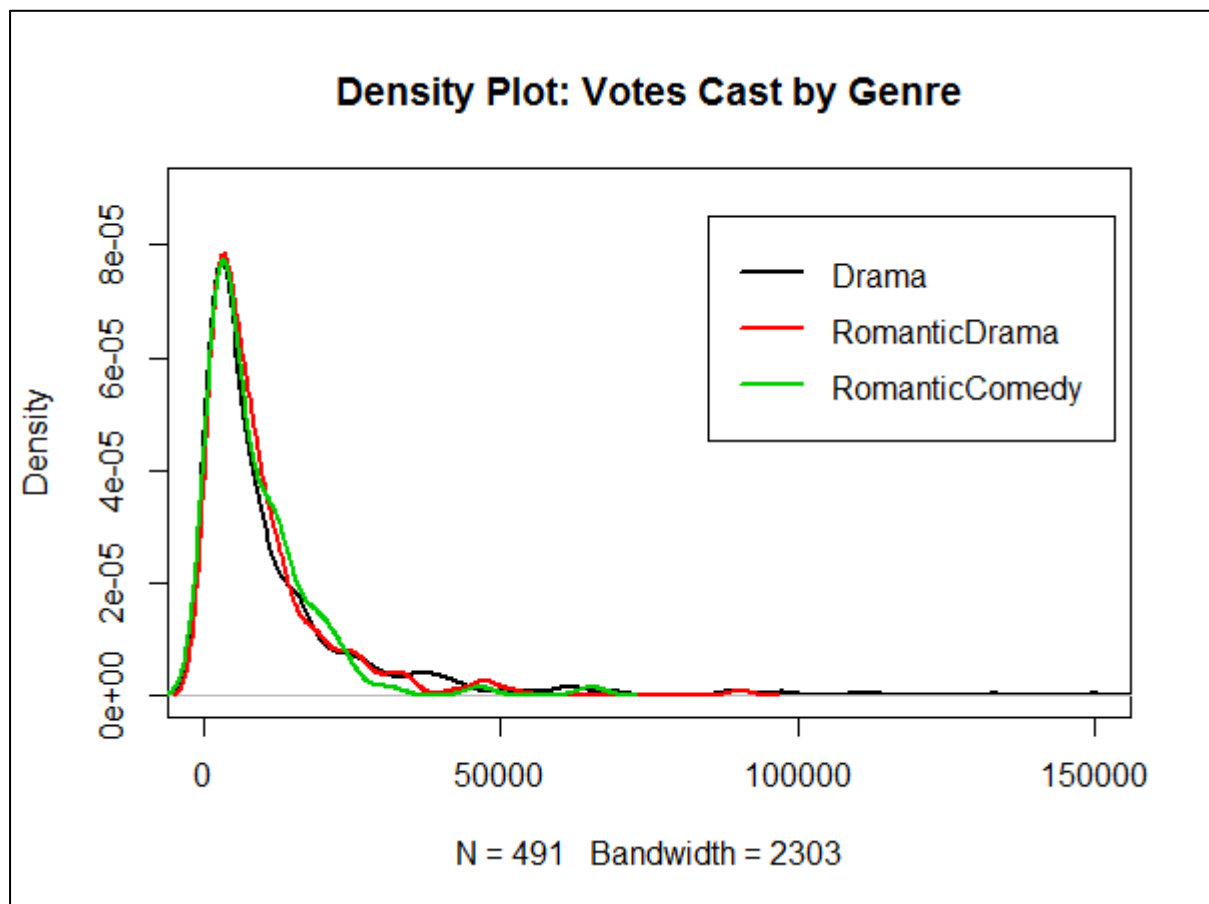
```
v.d = votes[Genre=="Drama"]
v.rd = votes[Genre=="RomanticDrama"]
v.rc = votes[Genre=="RomanticComedy"]

plot(density(v.d), xlim=c(0,150000), ylim=c(0,0.00009), col=1, lwd=2,
main="Density Plot: Votes Cast by Genre")

lines(density(v.rd), col=2,lwd=2)

lines(density(v.rc), col=3,lwd=2)

legend(85000, 0.000085, lwd=2, col=c(1, 2, 3), c("Drama", "RomanticDrama",
"RomanticComedy"))
```



The distribution for all genres is right-skewed, with a high density concentrated near (approximately) 3000 votes. There isn't much of a difference between the distributions. However, the sample sizes for each genre vary drastically.

I chose to use the density plot as it is easy to put them on one graph and compare them. This data could have also been represented with three separate strip charts, where each chart represented a genre and each point was the votes cast. That would have given a sense of the distribution, but it would have been hard to determine the area of highest density.

Drama median: 5997

Drama max: 149494

Romantic Drama median: 6498

Romantic Drama max: 90195

Romantic Comedy median: 5968.5

Romantic Comedy max: 65146

### Question 7

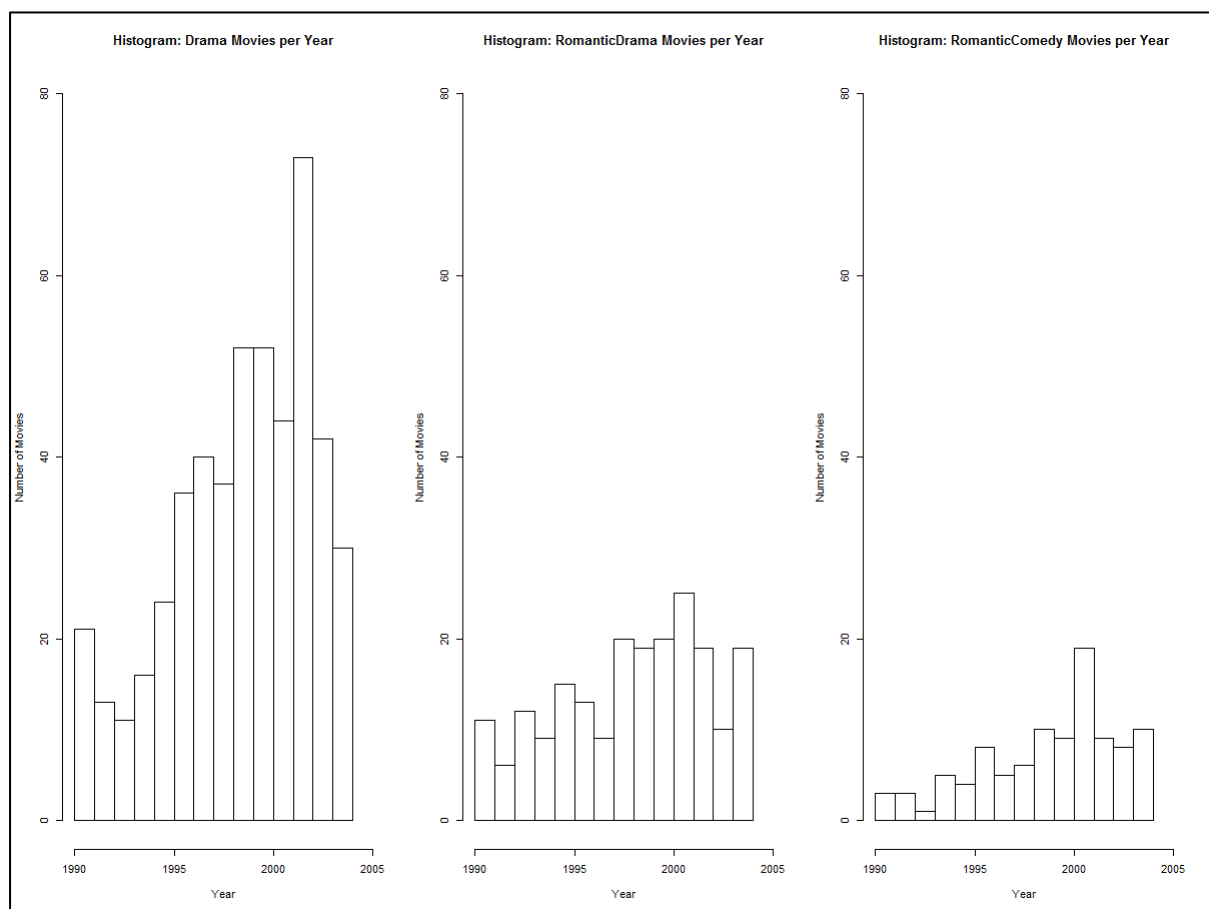
```
y.d = year[Genre=="Drama"]  
y.rd = year[Genre=="RomanticDrama"]  
y.rc = year[Genre=="RomanticComedy"]
```

```
par(mfrow=c(1,3))
```

```
hist(y.d, ylim=c(0,80), xlim=c(1990,2005), breaks=15, main="Histogram:  
Drama Movies per Year", xlab="Year", ylab="Number of Movies")
```

```
hist(y.rd, ylim=c(0,80), xlim=c(1990,2005), breaks=15, main="Histogram:  
RomanticDrama Movies per Year", xlab="Year", ylab="Number of Movies")
```

```
hist(y.rc, ylim=c(0,80), xlim=c(1990,2005), breaks=15, main="Histogram:  
RomanticComedy Movies per Year", xlab="Year", ylab="Number of Movies")
```



Yes there have been substantial changes in the number of movies per genre over time. All genres seem to have seen a rise in the number of movies, however, Drama movies seem to be most popular, with a steeper rise and a greater absolute number. The second most popular genre is RomanticDrama, followed by RomanticComedy.

RomanticDrama movies were maximum in the year 2002.

The largest increase was in 1996 to 1997.

I chose this plot because it is easiest to see the maximums and the changes from year to year. We could have also used three separate strip charts.