Homework 5

Zabih Yousuf

Problem 1

Part (A)

```
library(ggplot2)
total_ratings <- read.csv("Homework 5 Data 1.csv")
ggplot(data = total_ratings, aes(y=rating.avg, x=Release.period, group =
Release.period)) + geom_boxplot(show.legend = FALSE, color = "red") +
geom_dotplot(binaxis = "y", stackdir = "center", fill = "white")

## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```

Part (B)

The movies made in the earlier period were not that many which left little room for bad reviews. This was also the time where movies were on the come up so every movie was deemed original and good.

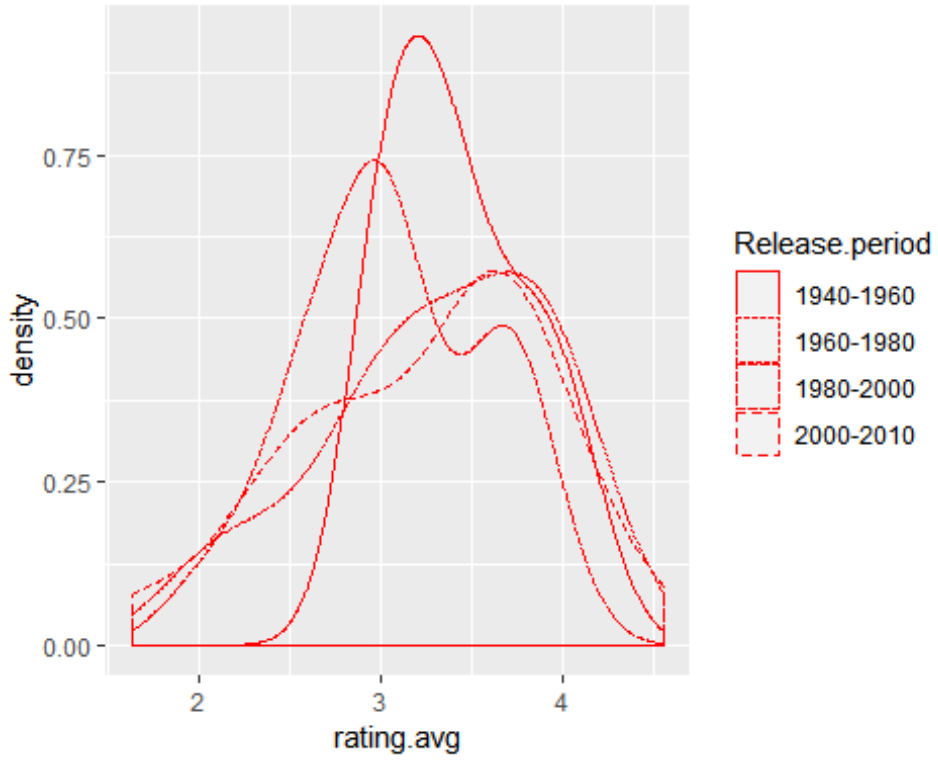
Part (C)

The boxplot is able to list out the summary values such as the Q1 and Q3 and median etc, but the dotplot is able to show all the values. The combination of both allows for maximum amount of information.

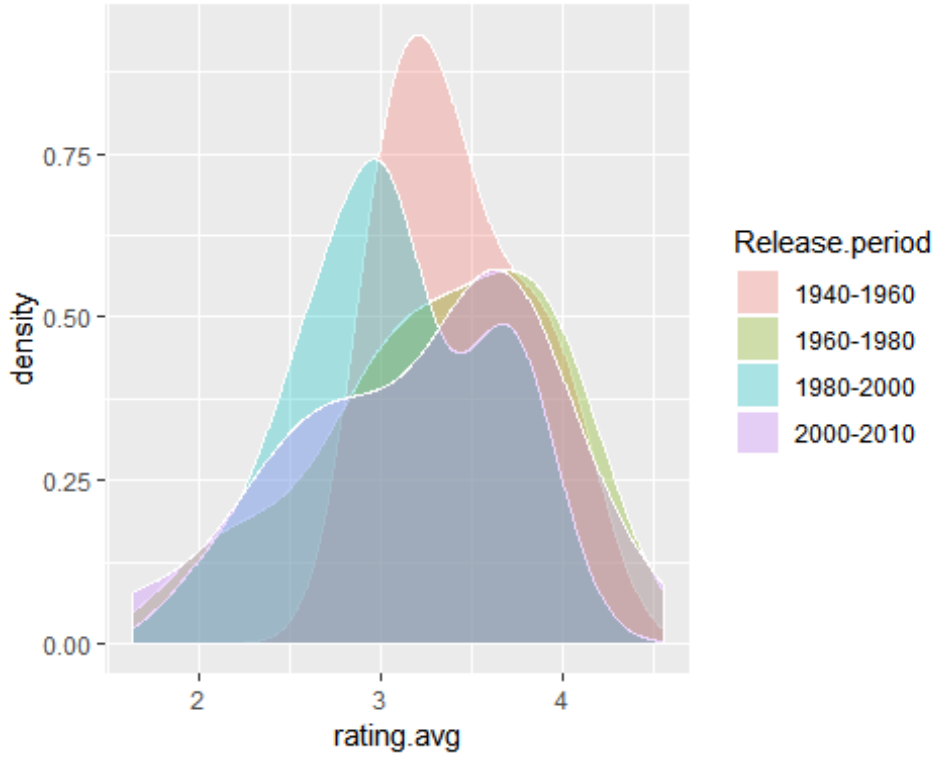
Part (D)

The easiest to read in my opinion is probably the last graph simply because I am able to visualize each period with clearity.

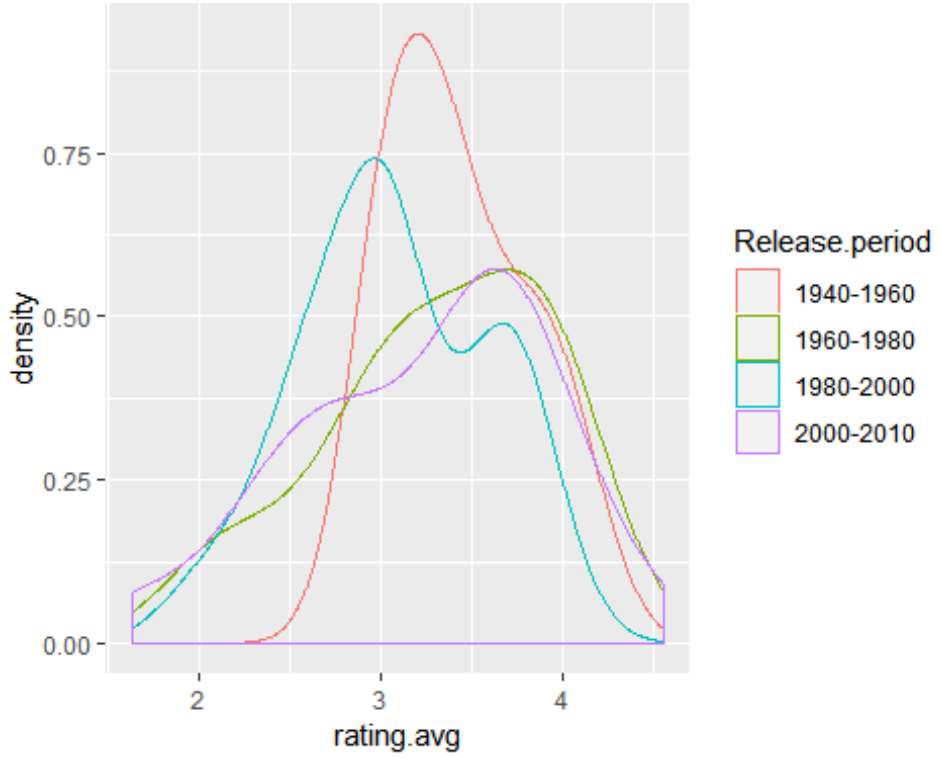
```
ggplot(total_ratings, aes(x=rating.avg, group = Release.period, linetype =
Release.period)) + geom_density(color = "red")
```



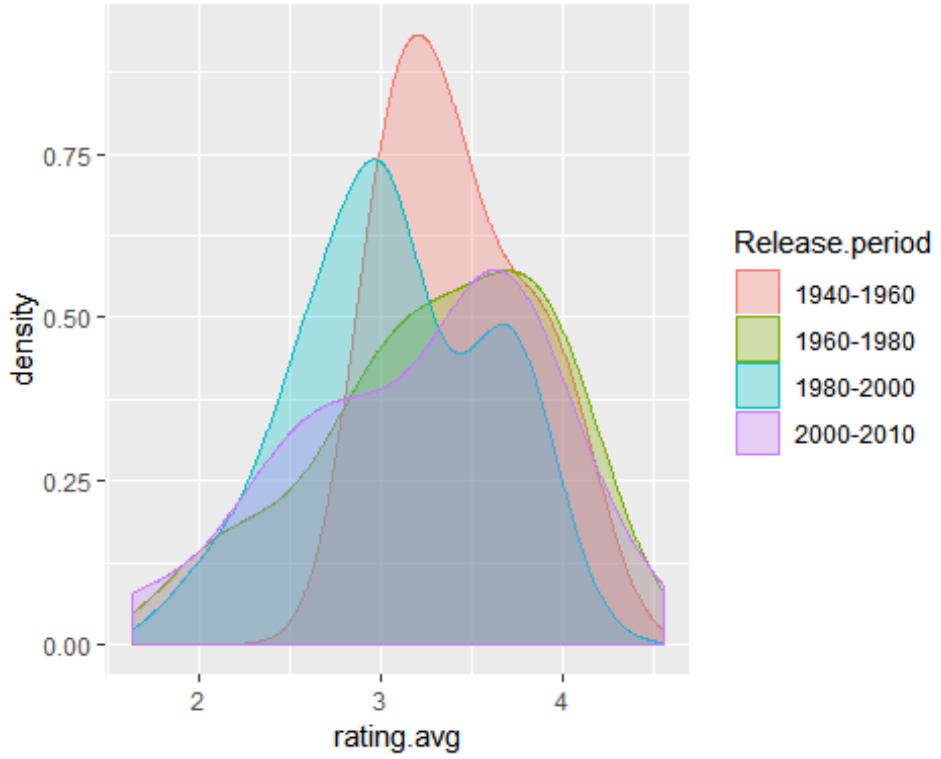

```
ggplot(total_ratings, aes(x=rating.avg, group = Release.period, fill =
Release.period)) + geom_density(alpha = .3, color = "white")
```




```
ggplot(total_ratings, aes(x=rating.avg, group = Release.period, color =
Release.period)) + geom_density()
```



```
ggplot(total_ratings, aes(x=rating.avg, group = Release.period, color =
Release.period, fill= Release.period)) + geom_density(alpha = .3)
```



Part (E)

The graphs suggest that time has affected the spread of the average rating. Besides from the 1940-1960 period, the other periods show a steady increase towards the higher ratings.

Part (F)

I prefer the dotplot combined with the box and whisker plot. The visual for that combination allows the user to get a sense of all the information needed. The 5 number summary is available with the box and whisker plot and the dotplot shows all data points.