

Visualization

Haochen Wang

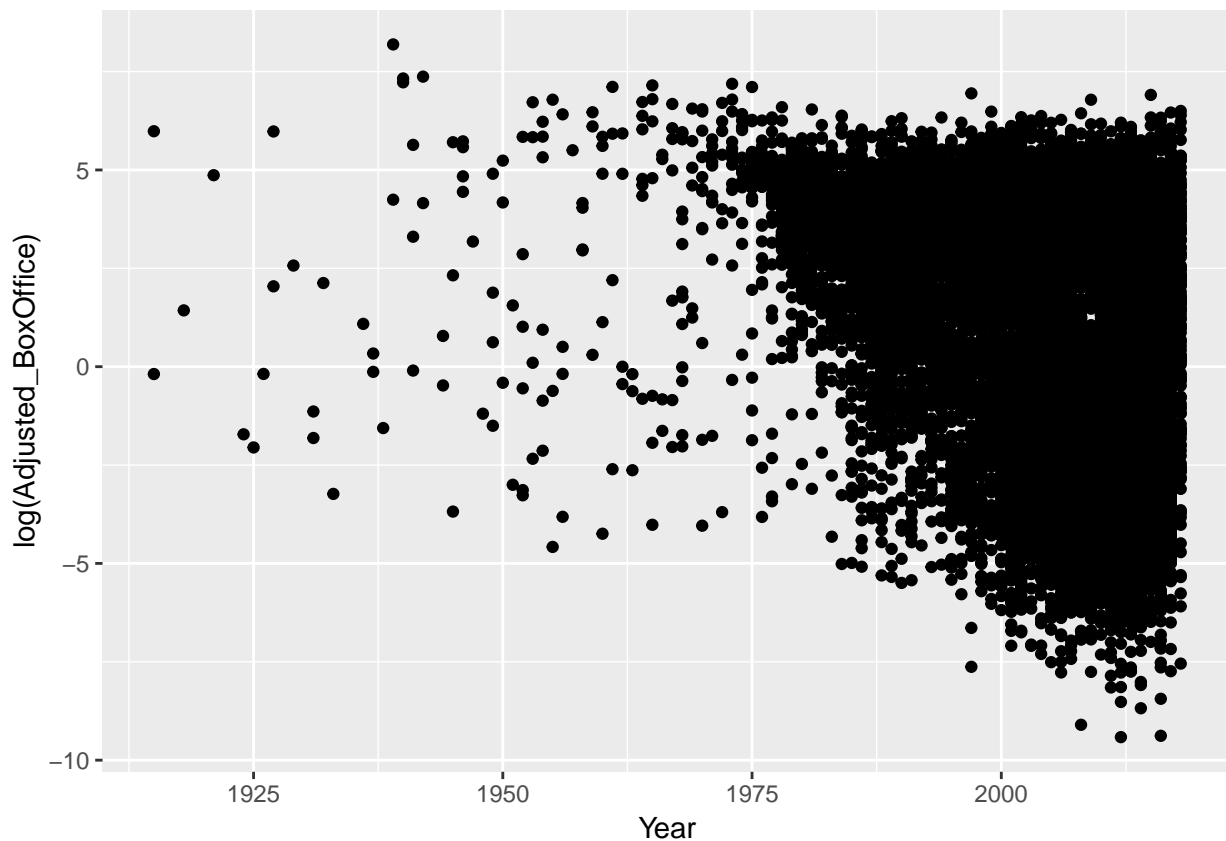
12/5/2018

```
library(ggplot2)
```

```
data <- read.csv("data_combined_budget.csv")
```

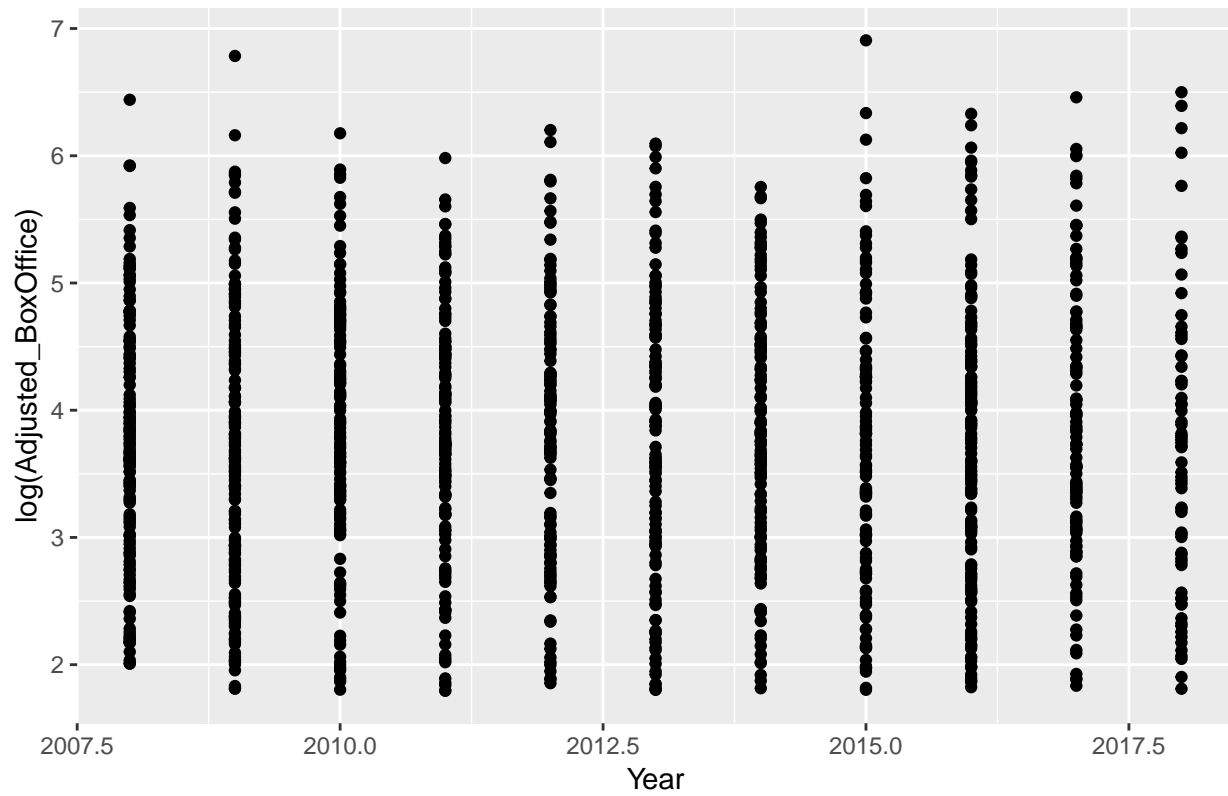
Scatter Plot

```
ggplot(data = data) +  
  geom_point(mapping = aes(y = log(Adjusted_BoxOffice), x = Year))
```

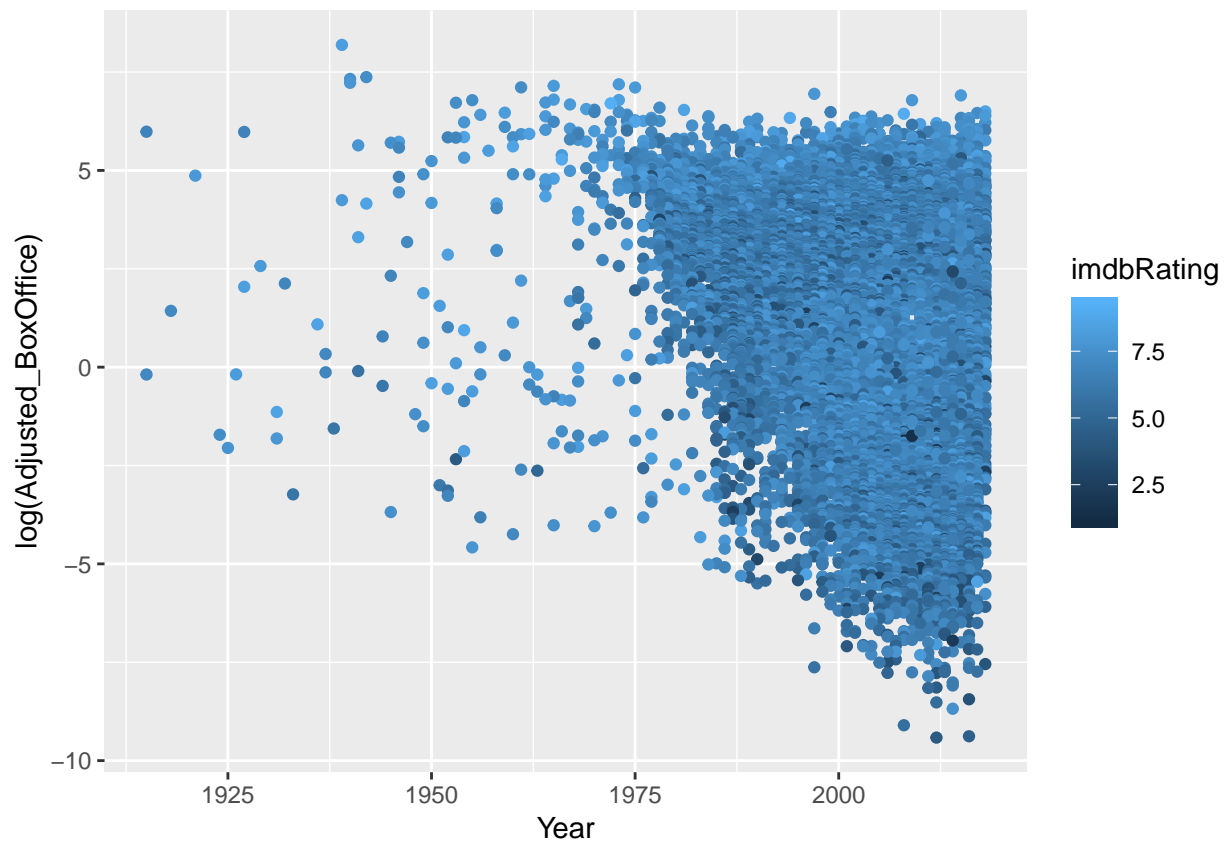


```
ggplot(data = data[data$Year>=2008 & data$Adjusted_BoxOffice>=6 ,]) +  
  geom_point(mapping = aes(y = log(Adjusted_BoxOffice), x = Year)) +  
  ggtitle("Movies In Recent 10 Years has more than 10e6 Box office")
```

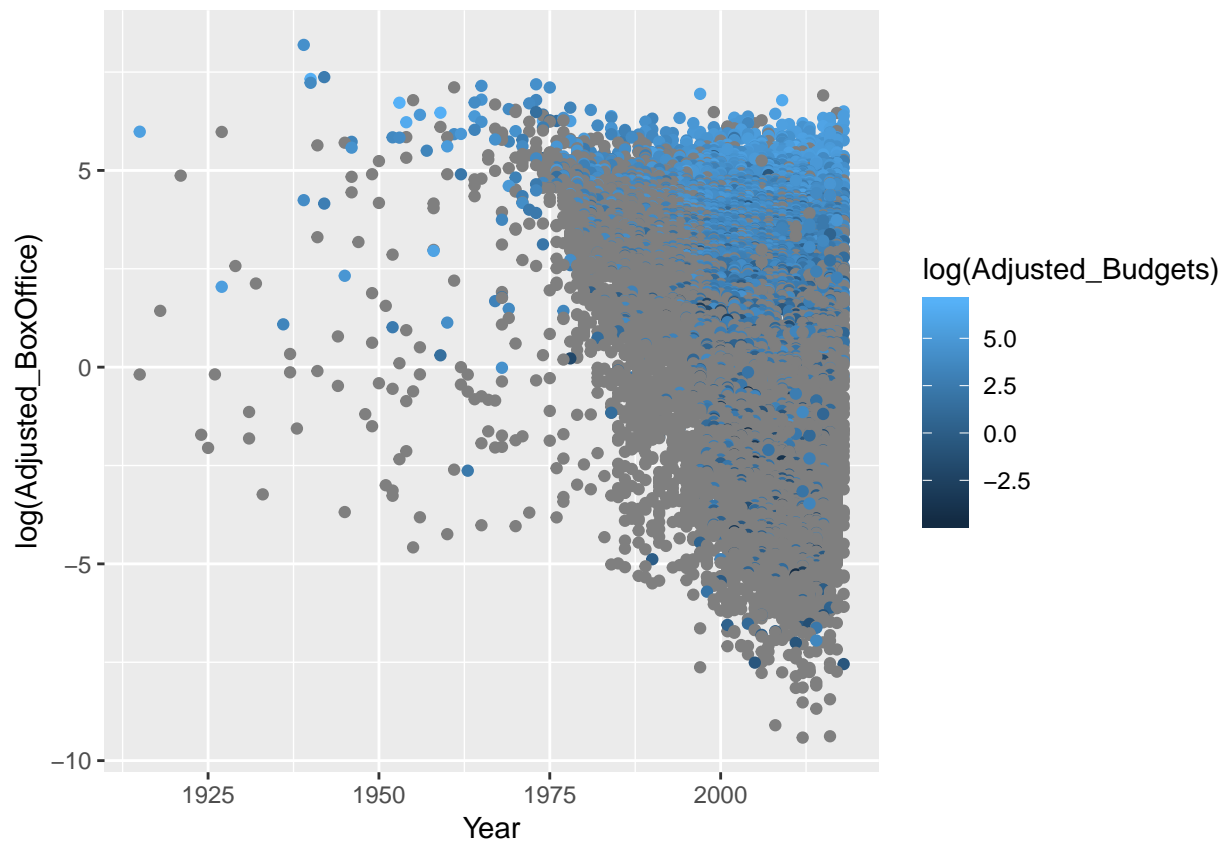
Movies In Recent 10 Years has more than 10e6 Box office



```
ggplot(data = data) +  
  geom_point(mapping = aes(x = log(Adjusted_BoxOffice), y = Year,  
    color = imdbRating)) + coord_flip()
```

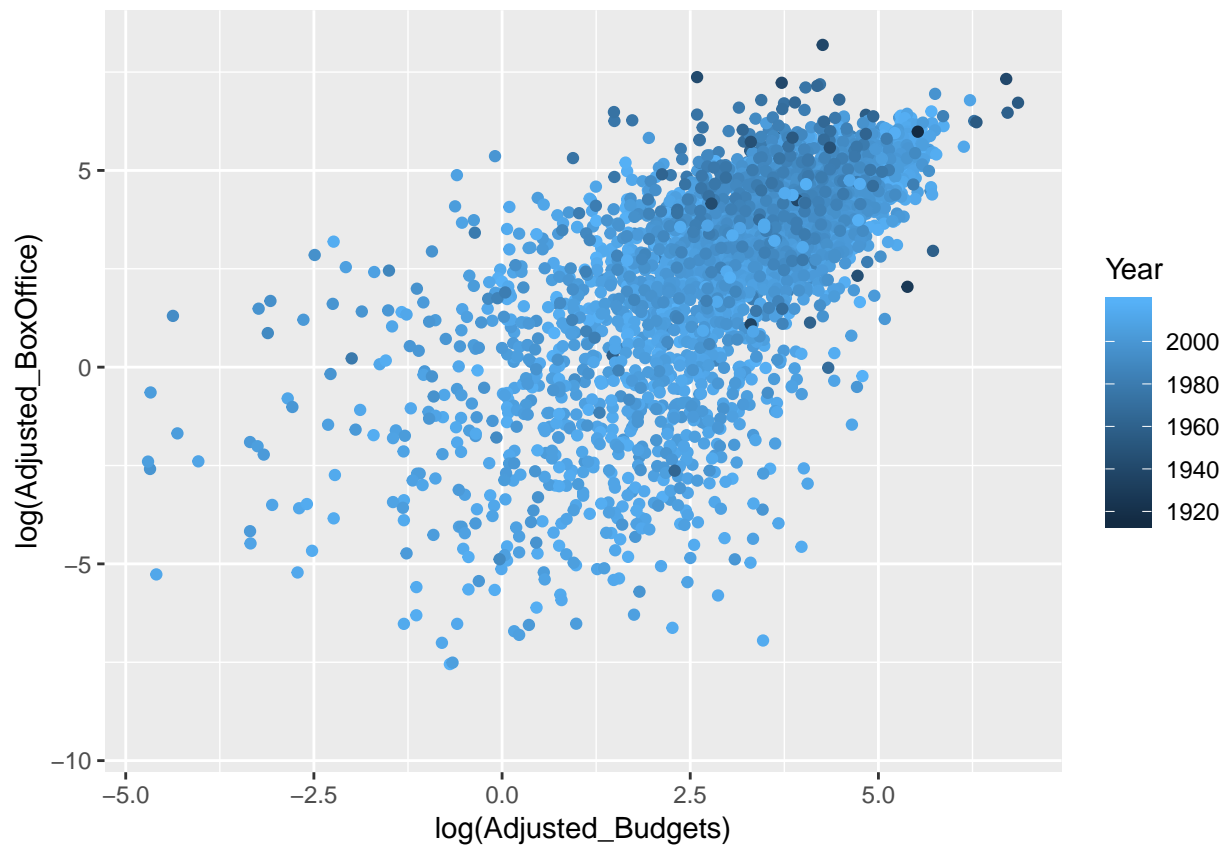


```
ggplot(data = data) +  
  geom_point(mapping = aes(x = log(Adjusted_BoxOffice), y = Year,  
    color = log(Adjusted_Budgets))) + coord_flip()
```



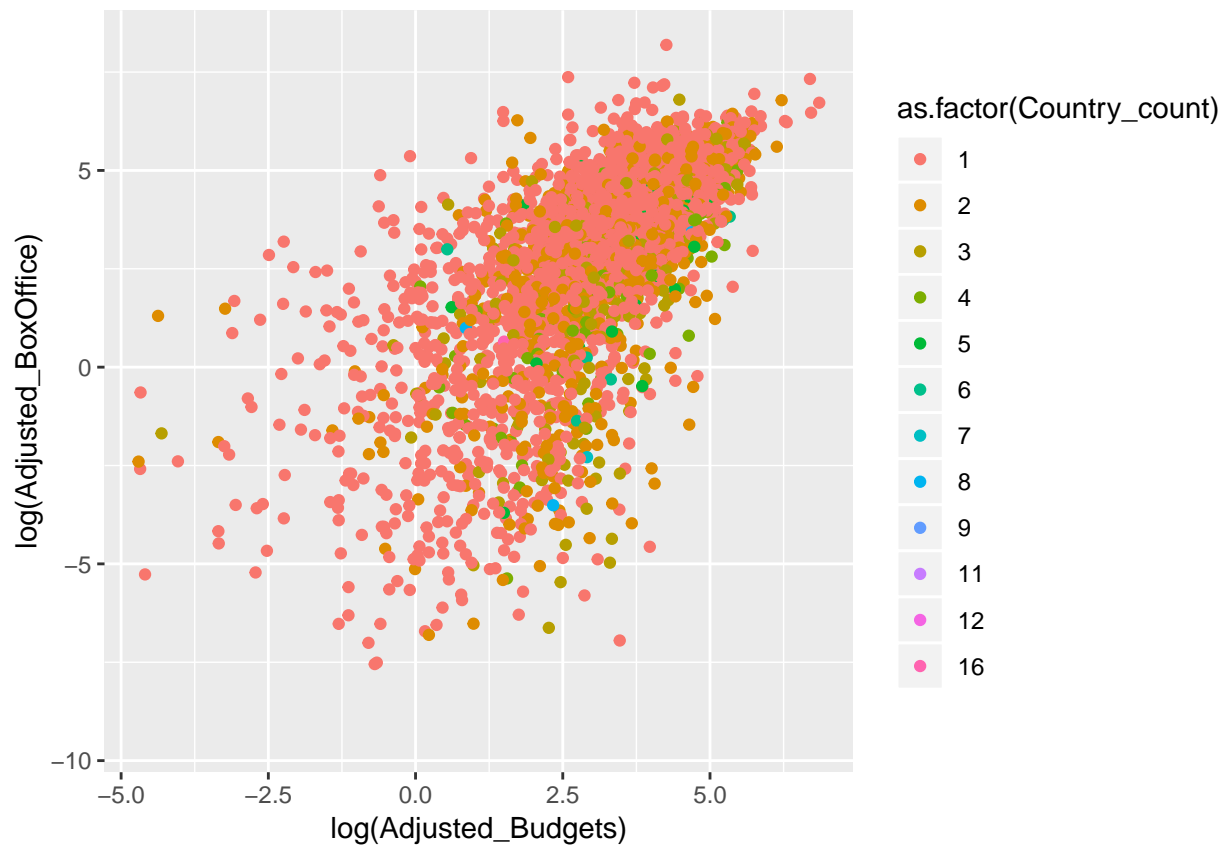
```
ggplot(data = data) +  
  geom_point(mapping = aes(x = log(Adjusted_Budgets),  
                           y = log(Adjusted_BoxOffice), color = Year))
```

Warning: Removed 7357 rows containing missing values (geom_point).



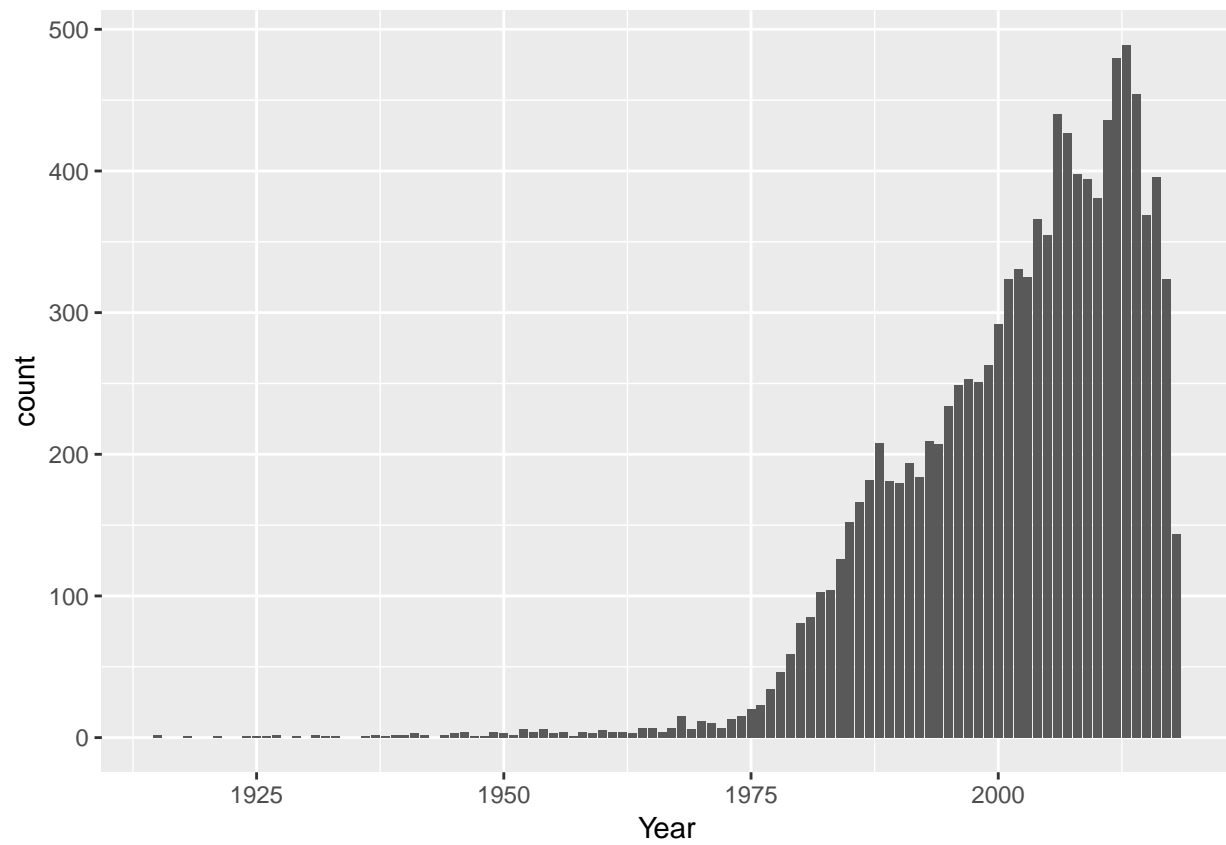
```
ggplot(data = data) +  
  geom_point(mapping = aes(x = log(Adjusted_Budgets),  
    y = log(Adjusted_BoxOffice),  
    color = as.factor(Country_count)))
```

```
## Warning: Removed 7357 rows containing missing values (geom_point).
```

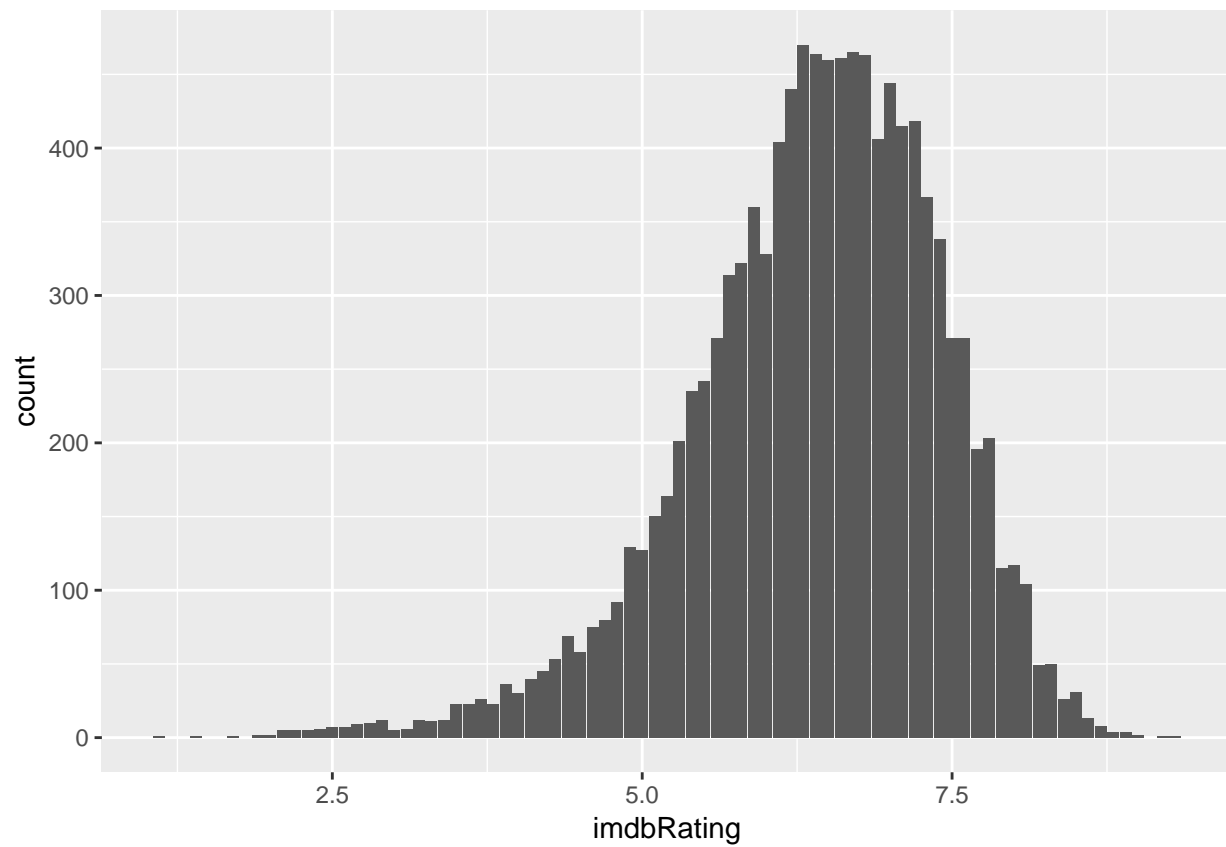


Bar Plot

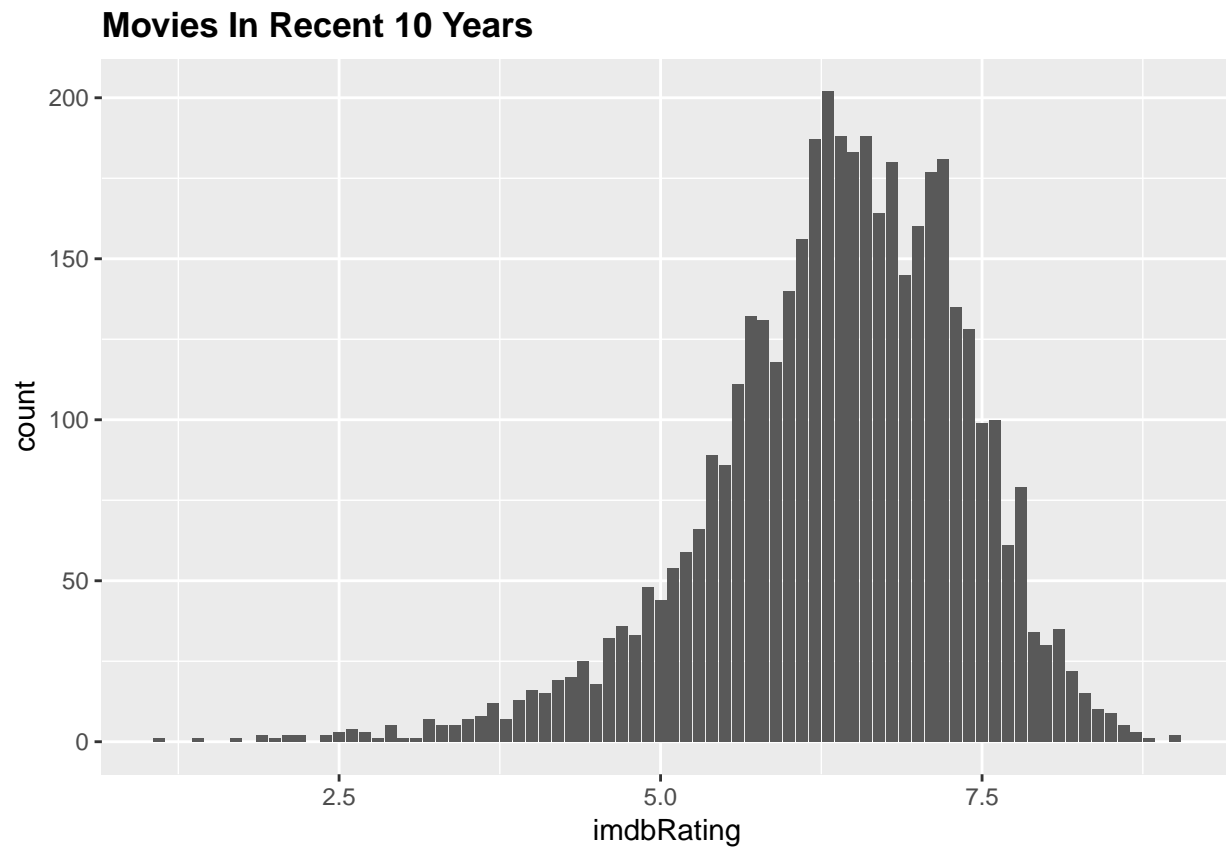
```
ggplot(data = data) +  
  geom_bar(mapping = aes(x = Year))
```



```
ggplot(data = data) +  
  geom_bar(mapping = aes(x = imdbRating))
```

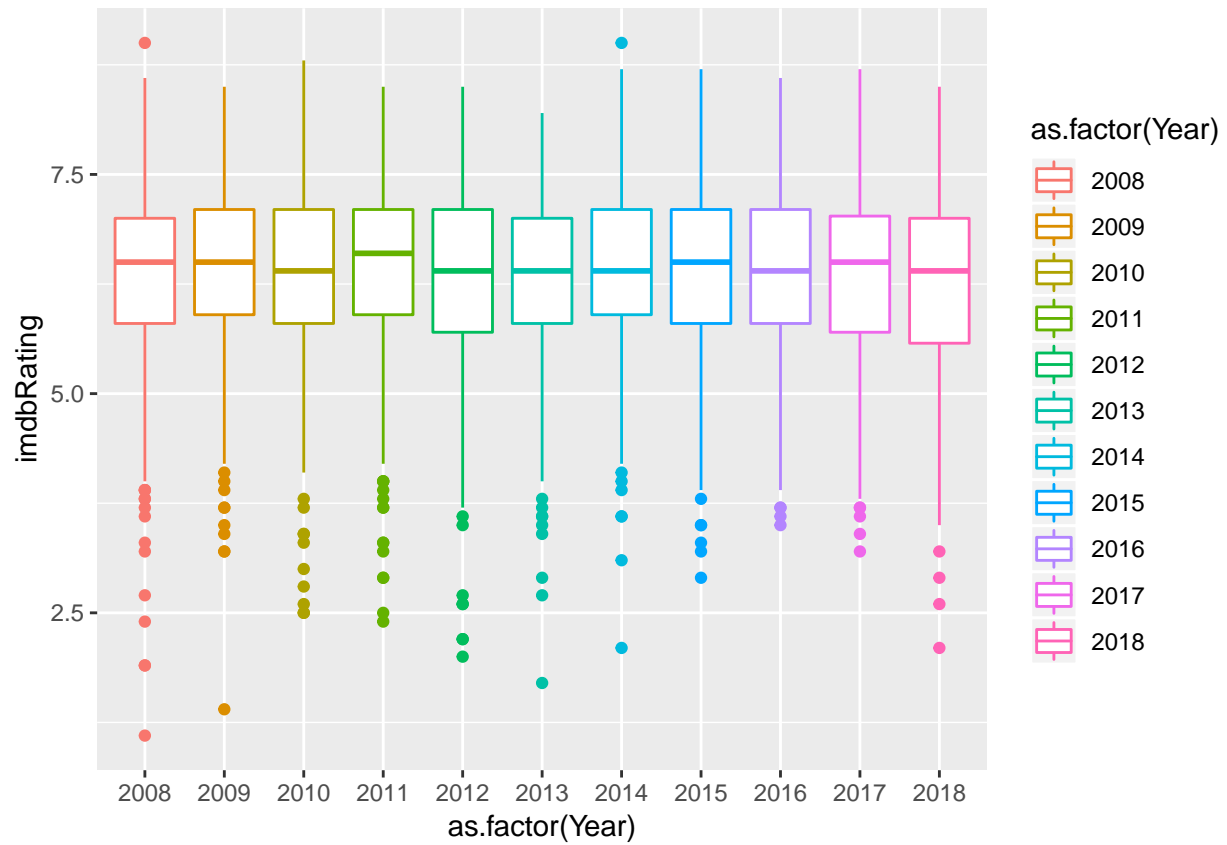


```
ggplot(data = data[data$Year>=2008,]) +  
  geom_bar(mapping = aes(x = imdbRating)) +  
  ggtitle("Movies In Recent 10 Years") +  
  theme(plot.title = element_text(lineheight=.8, face="bold"))
```

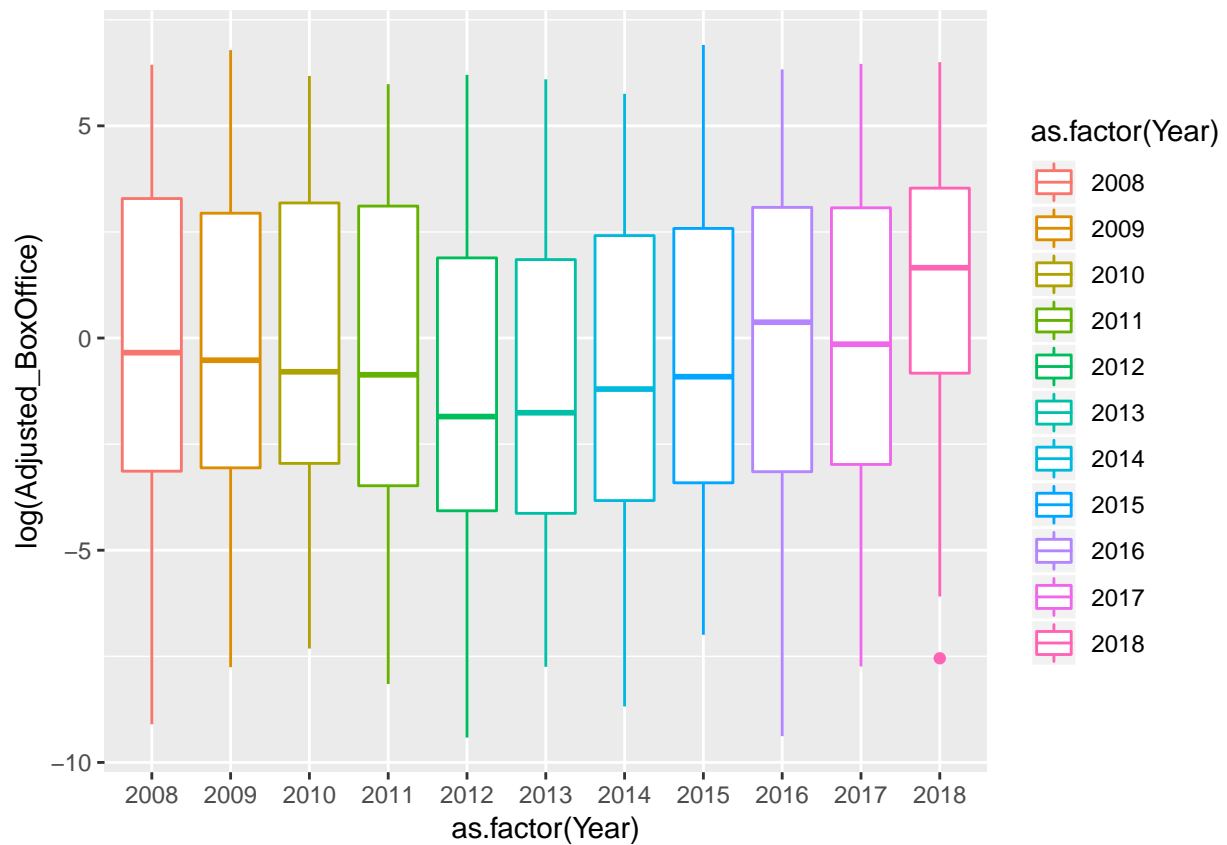



Boxplot

```
ggplot(data = data[data$Year>=2008,],  
  mapping = aes(x = as.factor(Year), y = imdbRating,  
    colour = as.factor(Year))) + geom_boxplot()
```



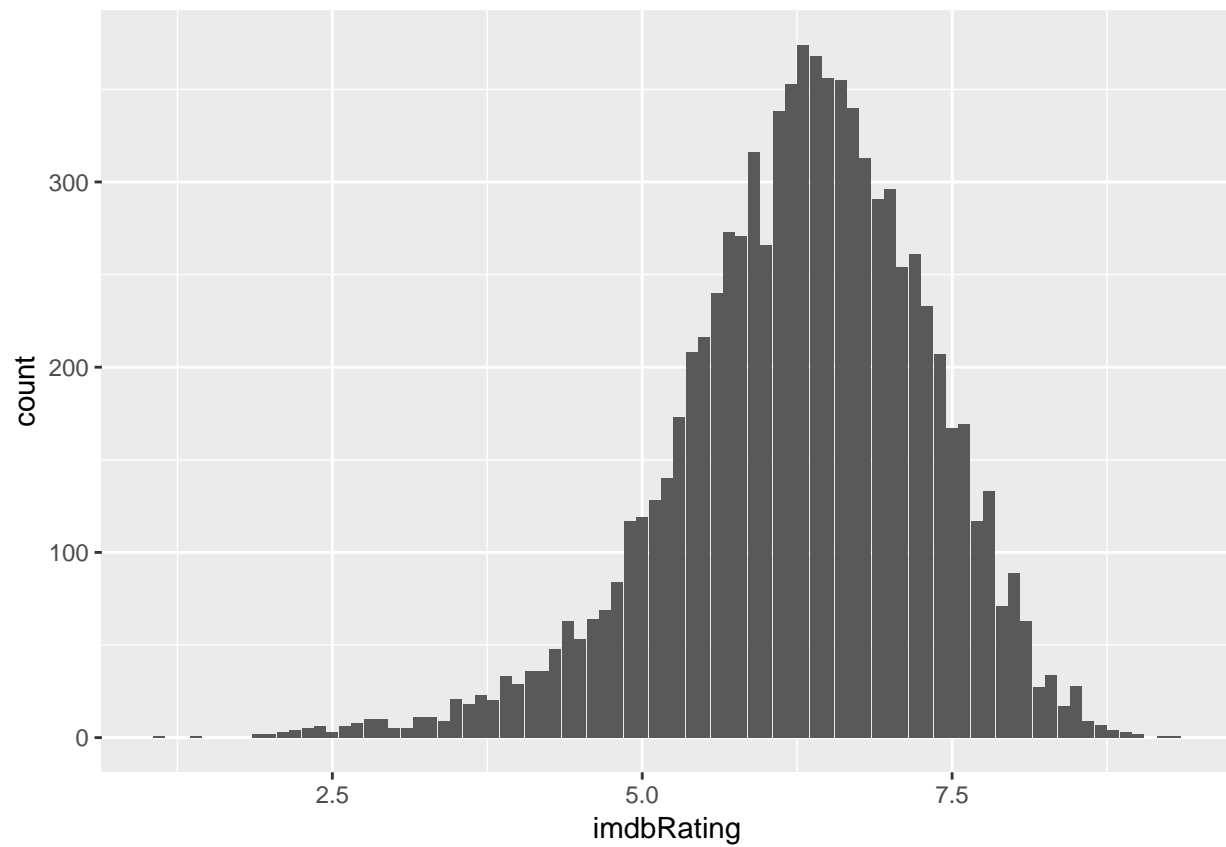
```
ggplot(data = data[data$Year>=2008,],
  mapping = aes(x = as.factor(Year), y = log(Adjusted_BoxOffice),
    colour = as.factor(Year))) + geom_boxplot()
```



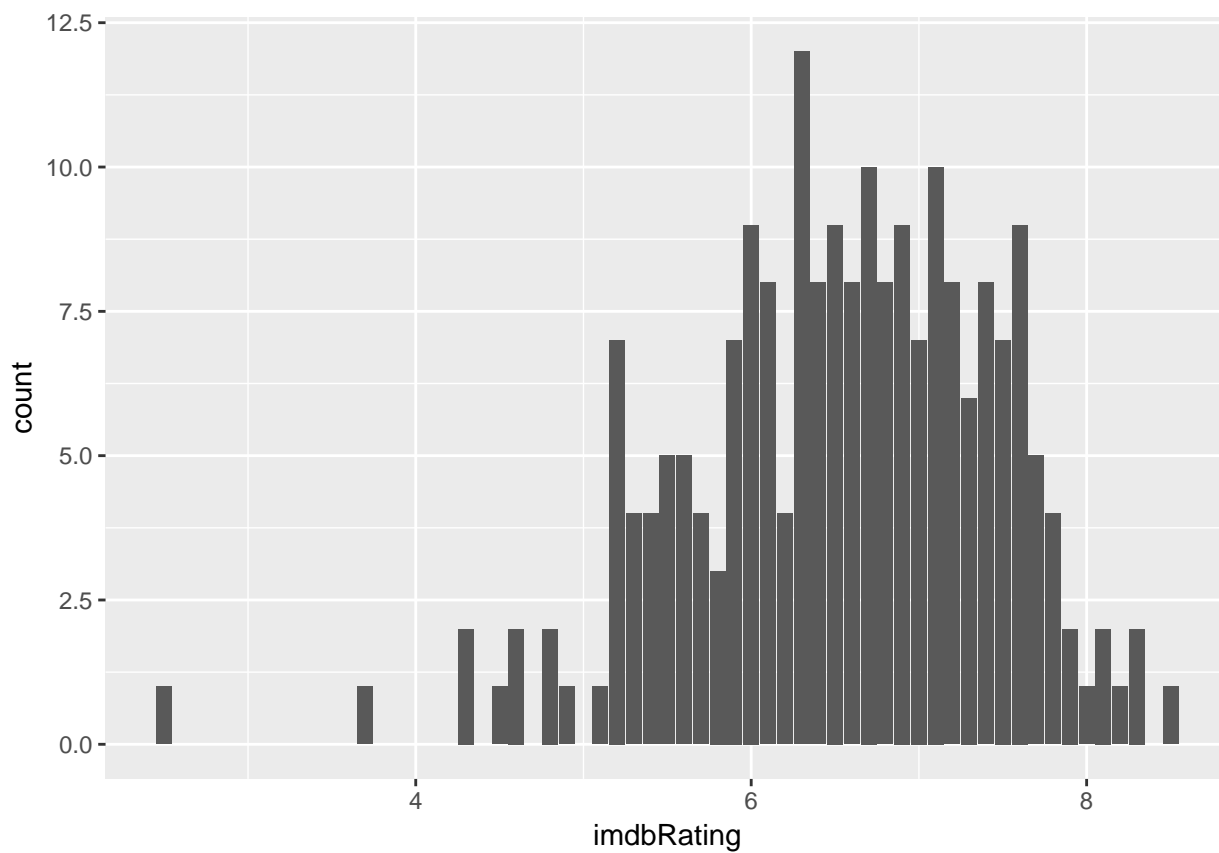
USA vs CHINA

```
USA = data[grep("USA", data$Country), ]
CHINA = data[grep("China", data$Country), ]
```

```
ggplot(data = USA) +
  geom_bar(mapping = aes(x = imdbRating))
```



```
ggplot(data = CHINA) +  
  geom_bar(mapping = aes(x = imdbRating))
```



```
CHINA[, "CRY"] = "CHINA"  
USA[, "CRY"] = "USA"  
USA_CHINA = rbind(CHINA, USA)
```

```
ggplot(data = USA_CHINA) +  
  geom_point(mapping = aes(y = imdbRating, x = as.factor(CRY)))
```

