Math 260 Project Data Description

Section: 1:20PM

Sandy, Pete & Tanush

11/5/18

[Reminder: You do **not** need to do any inference for this part of the project. That is, I do not expect to see confidence intervals or hypothesis tests yet, just descriptions of your data as we did in Chpaters 1 and 2.]

1. Write a brief background paragraph describing the data and relationships you plan to study. (No R commands needed here.)

The videogame industry is a multi-billion dollar industry and a significant part of the world economy. We plan to investigate certain trends and patterns within this industry to better understand it. The specific relationships we are interested in are maturity rating and global sales, genre and Japanese sales, global sales and critic score, console and genre. From this we will be able to learn which types of video games are the most successful in both global and Japanese economies, how the critic score relates to how successful the game will be, and to see if specific consoles are more prone to specific genres of games.

2. Use R to import and describe your data including cases and variables. Using the `head()` command will show the basic structure of the data.

```
Games2 = read.csv(file="C:\\Users\\tanus\\Documents\\College\\Homework\\Year 4\\Fall\\Applied St
ats\\Project\\video-game-sales-with-ratings\\VGS.csv")
head(Games2, 5)
```

```
##                            Name Platform Year_of_Release  Genre
## 1           Grand Theft Auto V      PS3            2013 Action
## 2 Grand Theft Auto: San Andreas      PS2            2004 Action
## 3           Grand Theft Auto V     X360            2013 Action
## 4    Grand Theft Auto: Vice City      PS2            2002 Action
## 5         Grand Theft Auto III      PS2            2001 Action
##              Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
## 1 Take-Two Interactive     7.02     9.09     0.98        3.96        21.04
## 2 Take-Two Interactive     9.43     0.40     0.41       10.57        20.81
## 3 Take-Two Interactive     9.66     5.14     0.06        1.41        16.27
## 4 Take-Two Interactive     8.41     5.49     0.47        1.78        16.15
## 5 Take-Two Interactive     6.99     4.51     0.30        1.30        13.10
##   Critic_Score Critic_Count User_Score User_Count      Developer Rating
## 1           97           50        8.2       3994 Rockstar North      M
## 2           95           80          9       1588 Rockstar North      M
## 3           97           58        8.1       3711 Rockstar North      M
## 4           95           62        8.7        730 Rockstar North      M
## 5           97           56        8.5        664     DMA Design      M
```

3. Provide graphs and numerical summary statistics for your variables individually (e.g., frequency tables, bar charts, histograms, or dotplots as appropriate). Briefly annotate or caption your graphs.
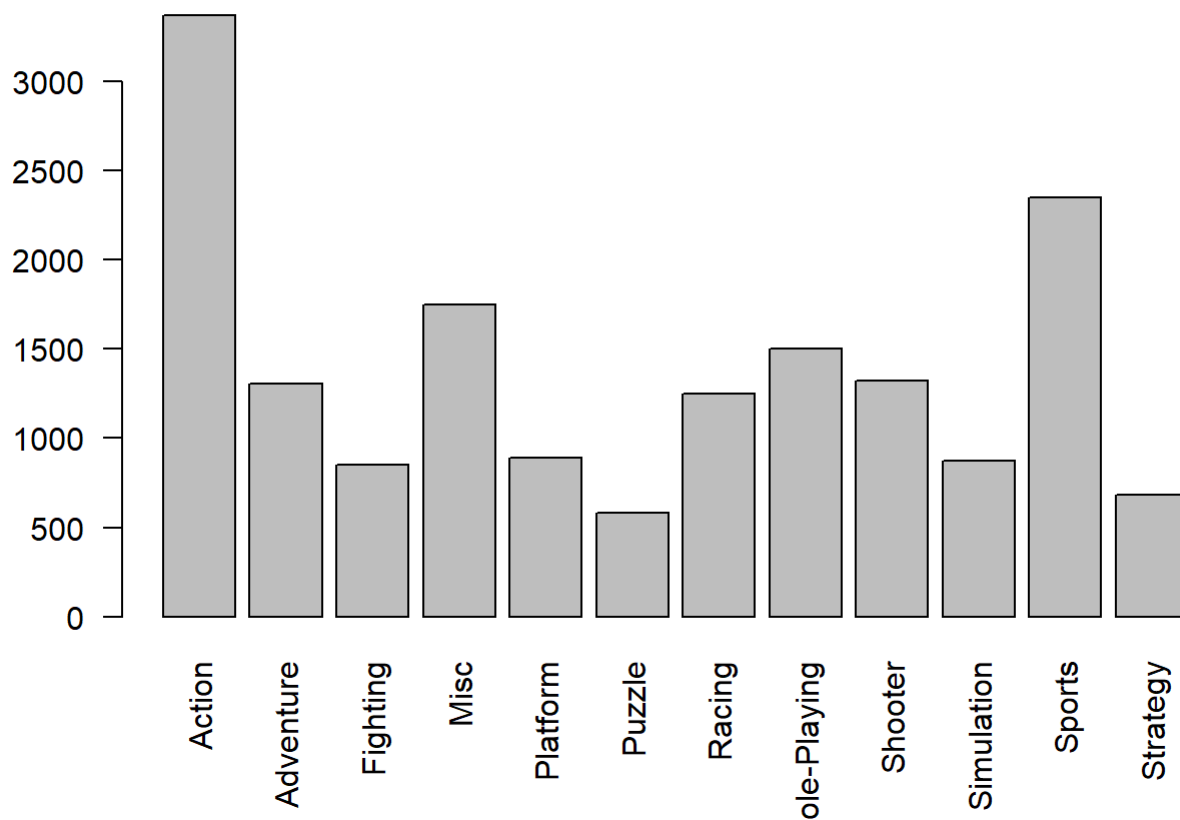
```
boxplot(Games2$Critic_Score, ylab = "Critic Score")
```
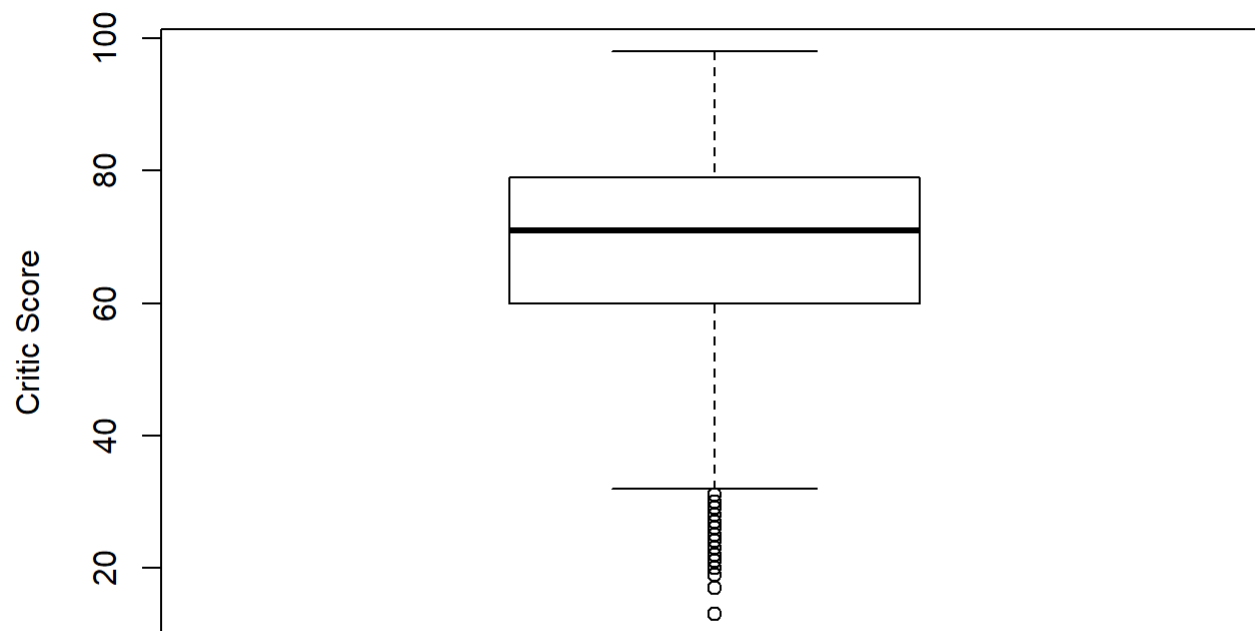

This graph displays the spread in the critic score for all games.

```
counts <- table(Games2$Genre)
barplot(counts, main="Frequency by Genre", las=2)
```
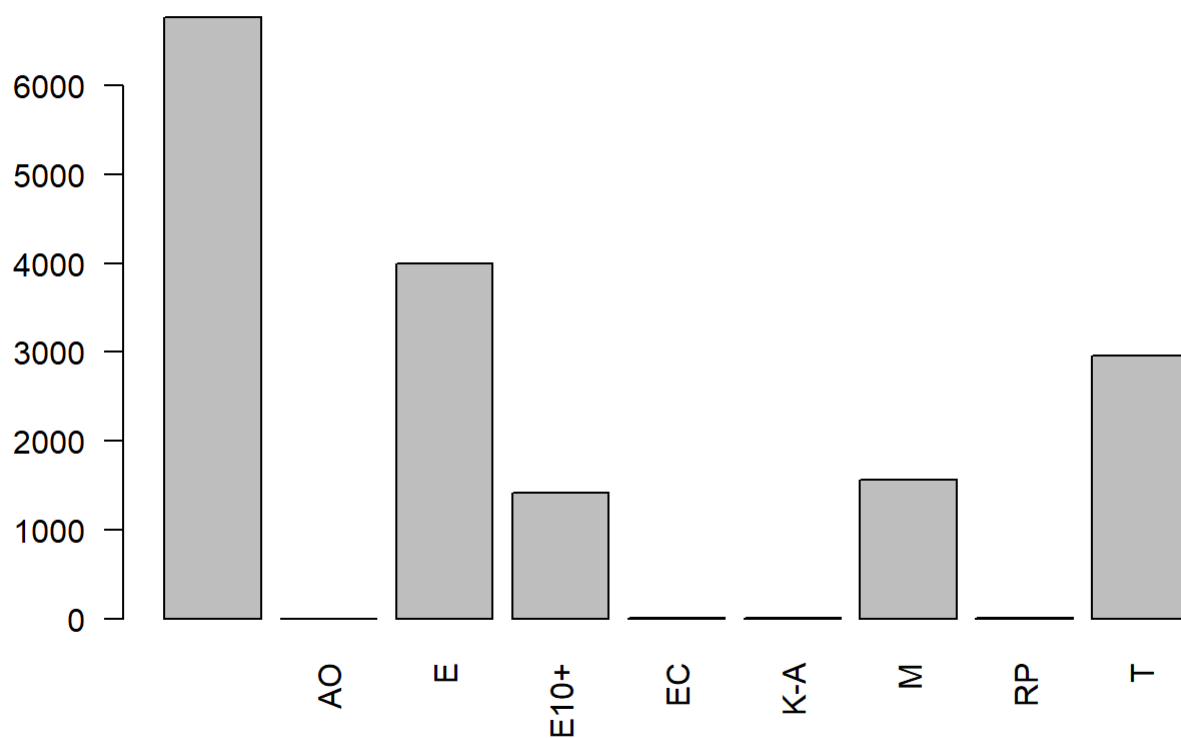
Frequency by Genre

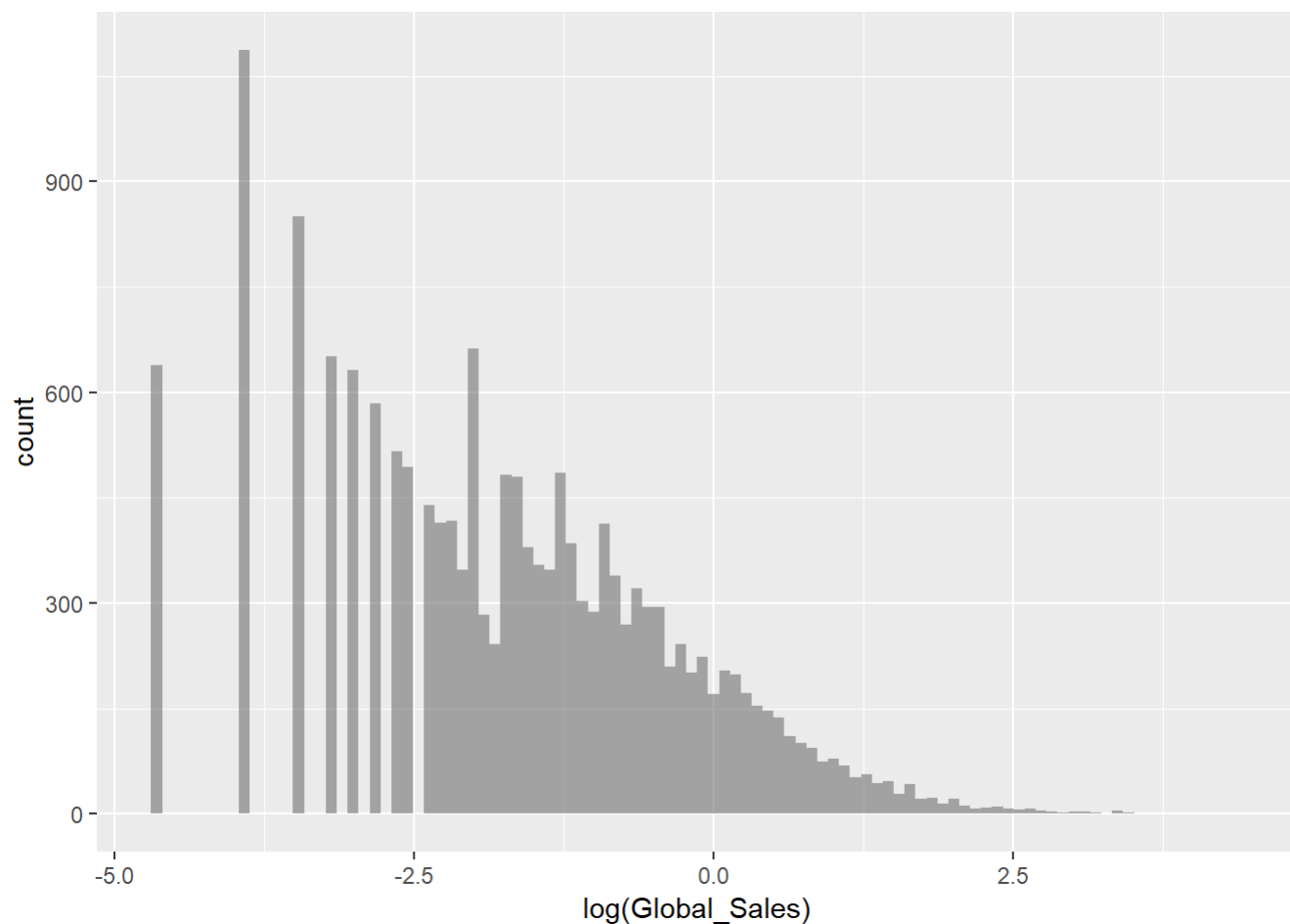This graph displays the number of games for each ganre.

```
count3 <- table(Games2$Rating)
barplot(count3, main="Frequency by Rating", las=2)
```

Frequency by Rating

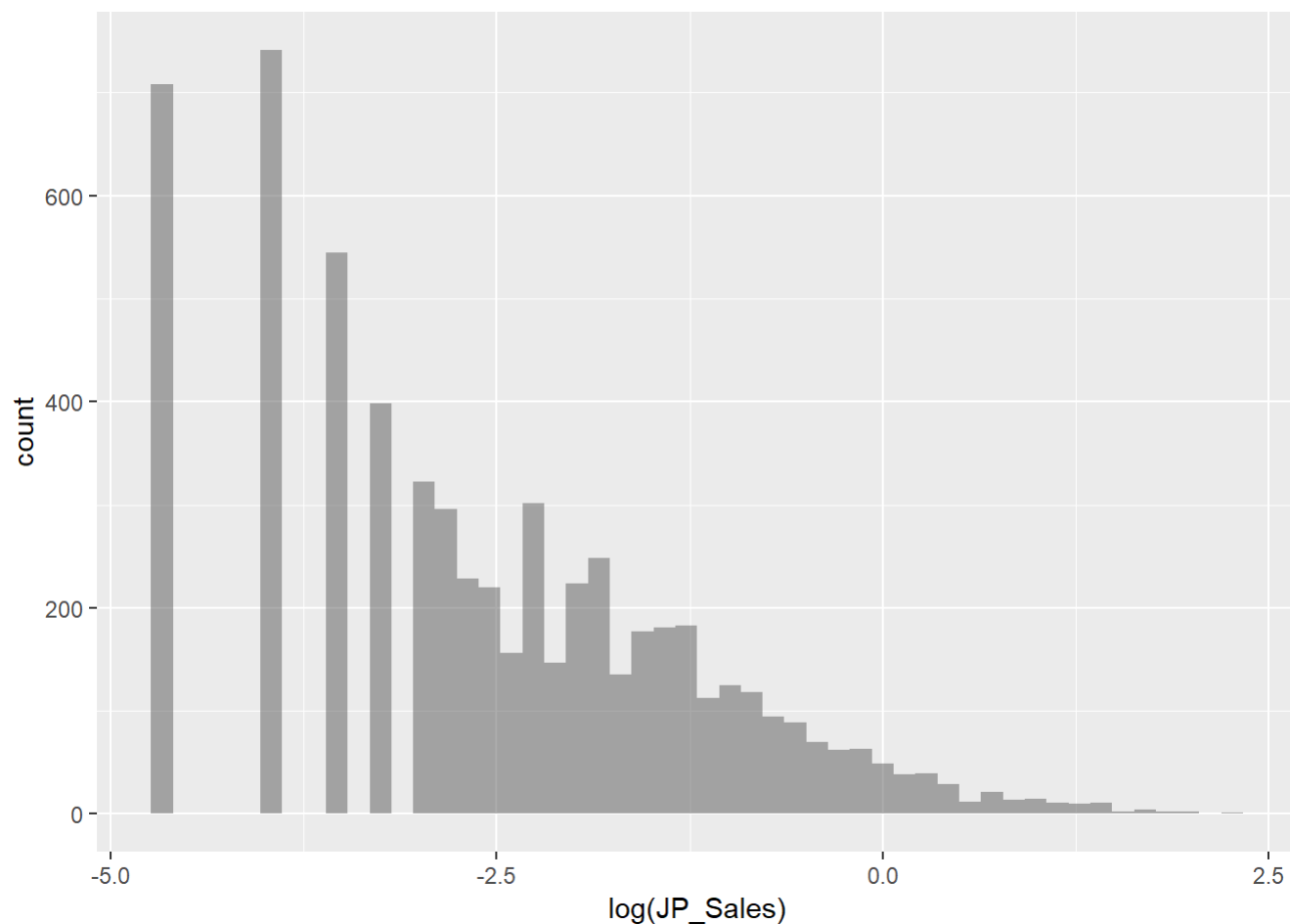This graph displays the number of games in each rating category. Unlabeled bar is for no rating.

```
gf_histogram(~log(Global_Sales), data = Games2, bins = 100)
```

This graph displays the log of the distribution of all sales. With global sales being in the millions.

```
gf_histogram(~log(JP_Sales), data = Games2, bins = 50)
```
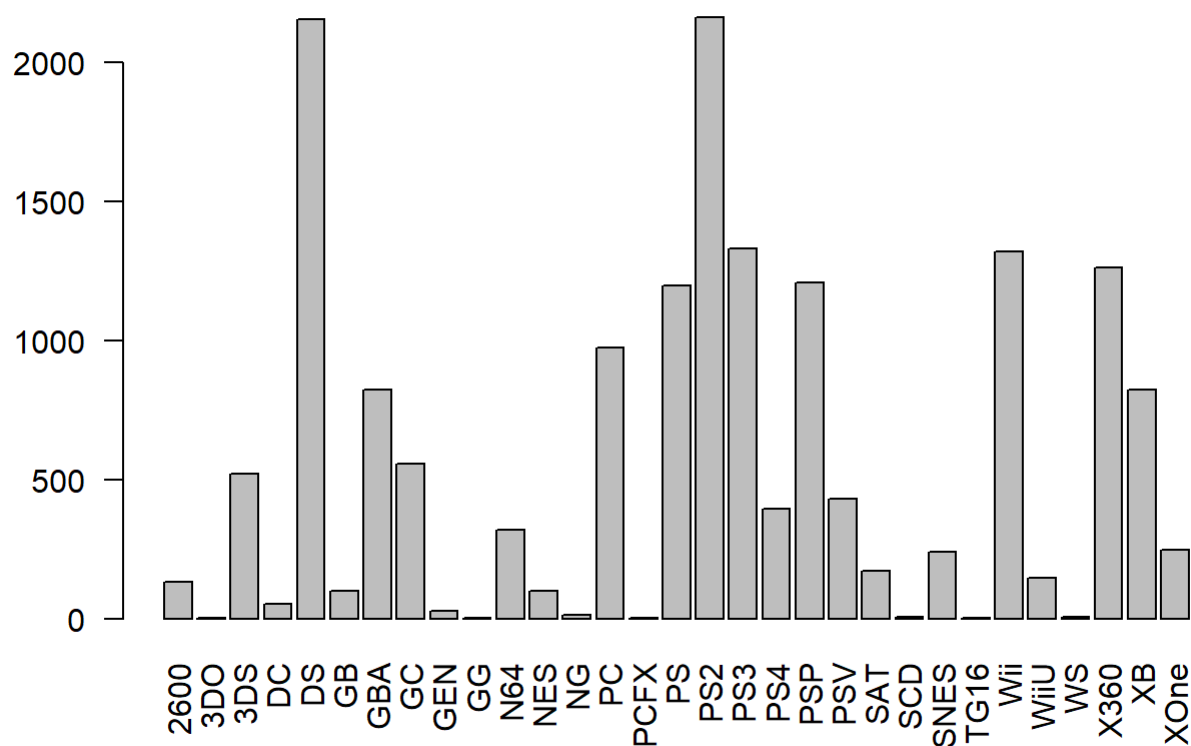
```
## Warning: Removed 10514 rows containing non-finite values (stat_bin).
```

This graph displays the log of the distribution of Japanese sales. With JP_Sales being in the millions.

```
count4 <- table(Games2$Platform)
barplot(count4, main="Frequency by Platform", las=2)
```

Frequency by Platform

This graph displays the number of games per platform.

4. Provide graphs that show the relationships you intend to study (e.g, boxplots, side-by-side histograms or dotplots, scatterplots as appropriate). Again, briefly annotate or caption your graphs.
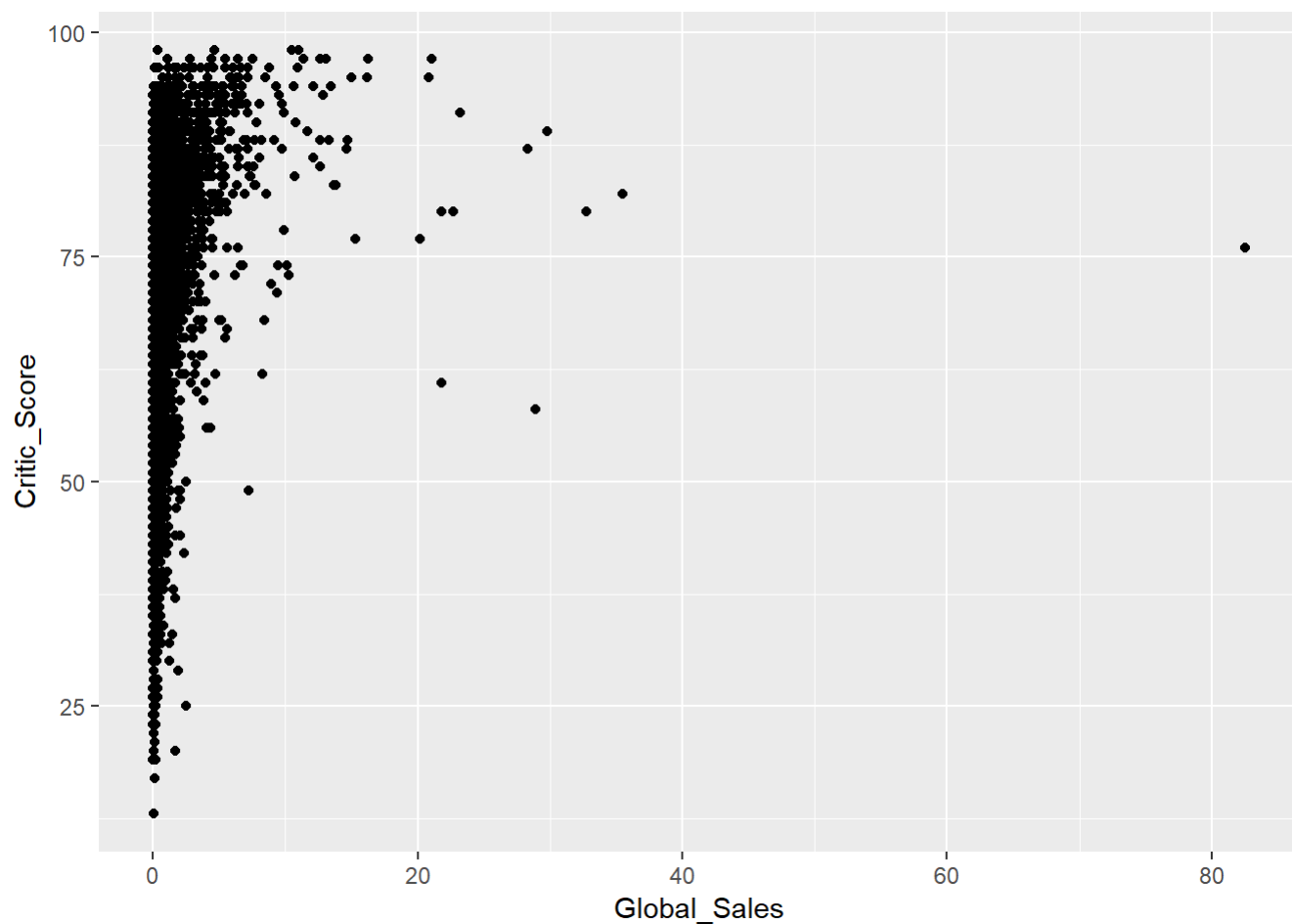
```
gf_point(Critic_Score~Global_Sales, data=Games2[-c(13687, 7273),])
```

```
## Warning: Removed 8579 rows containing missing values (geom_point).
```
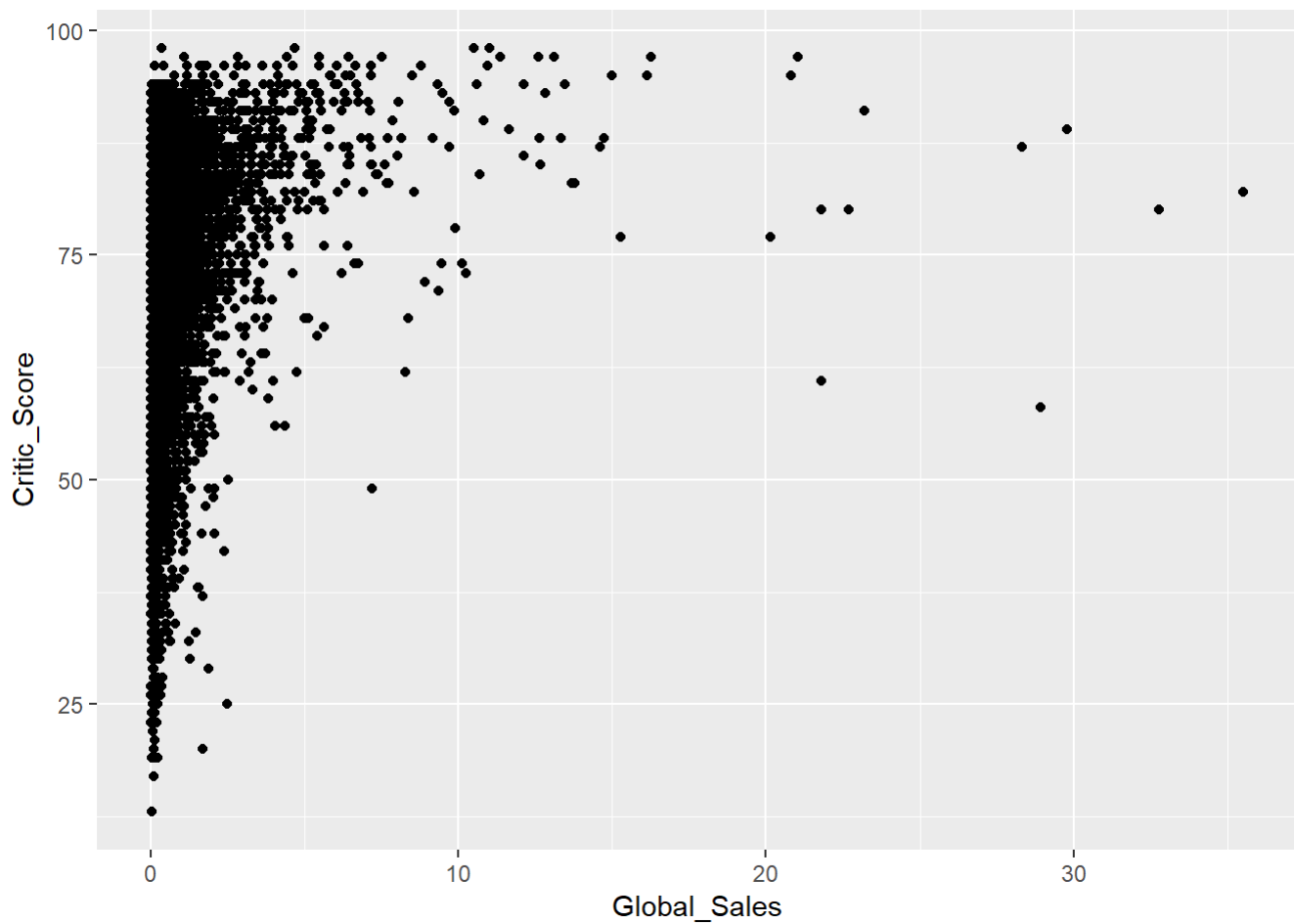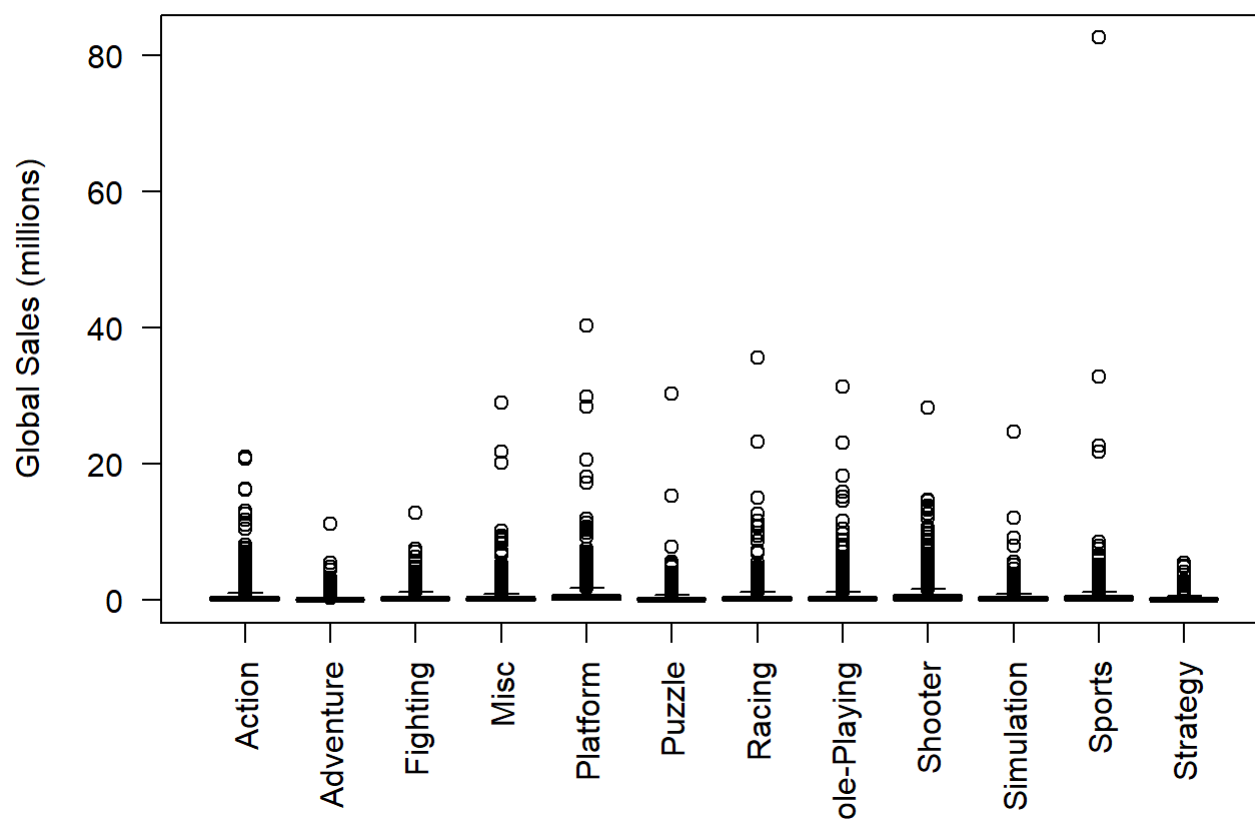
```
gf_point(Critic_Score~Global_Sales, data=Games2)
```

```
## Warning: Removed 8580 rows containing missing values (geom_point).
```
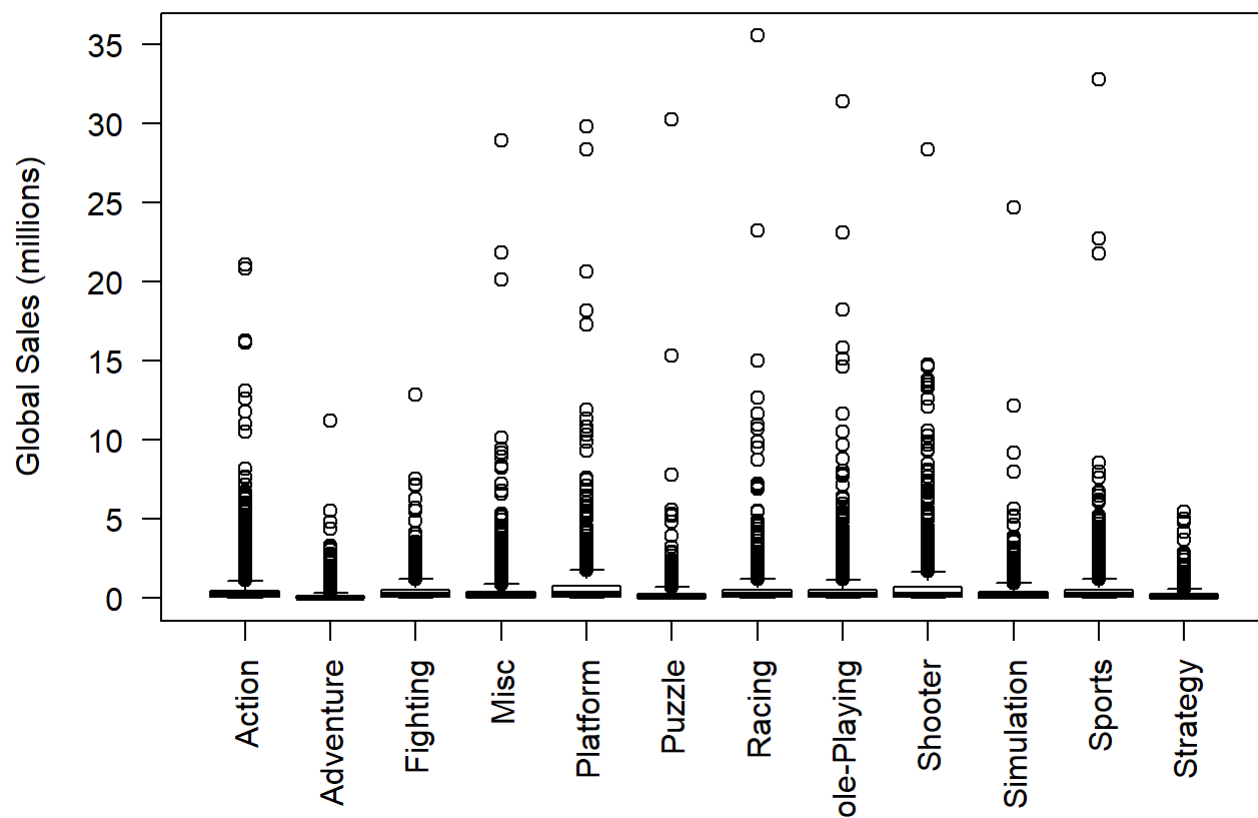
This graph displays the relationship between the critic scores the games recieved along with the gobal sales outcome.

```
boxplot(Global_Sales ~ Genre, data = Games2, ylab = "Global Sales (millions)", las = 2)
```
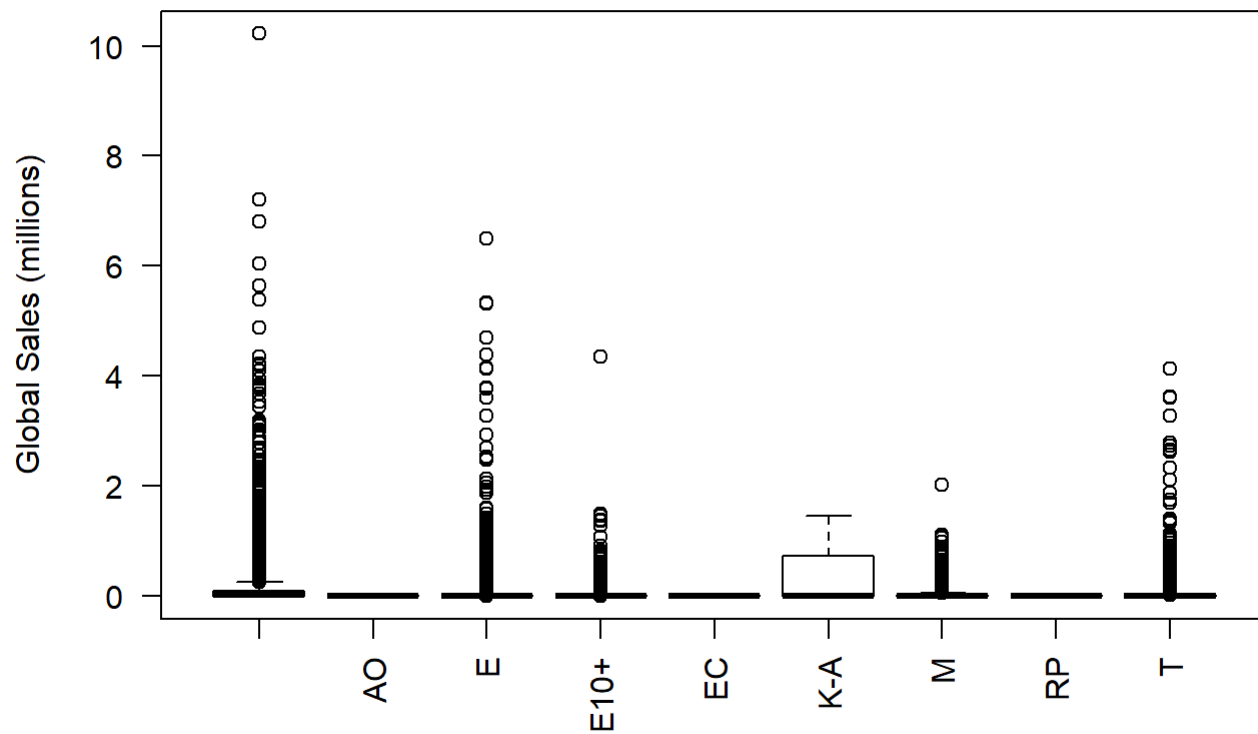
```
boxplot(Global_Sales ~ Genre, data = Games2[-c(13687, 7273),], ylab = "Global Sales (millions)",
 las = 2)
```
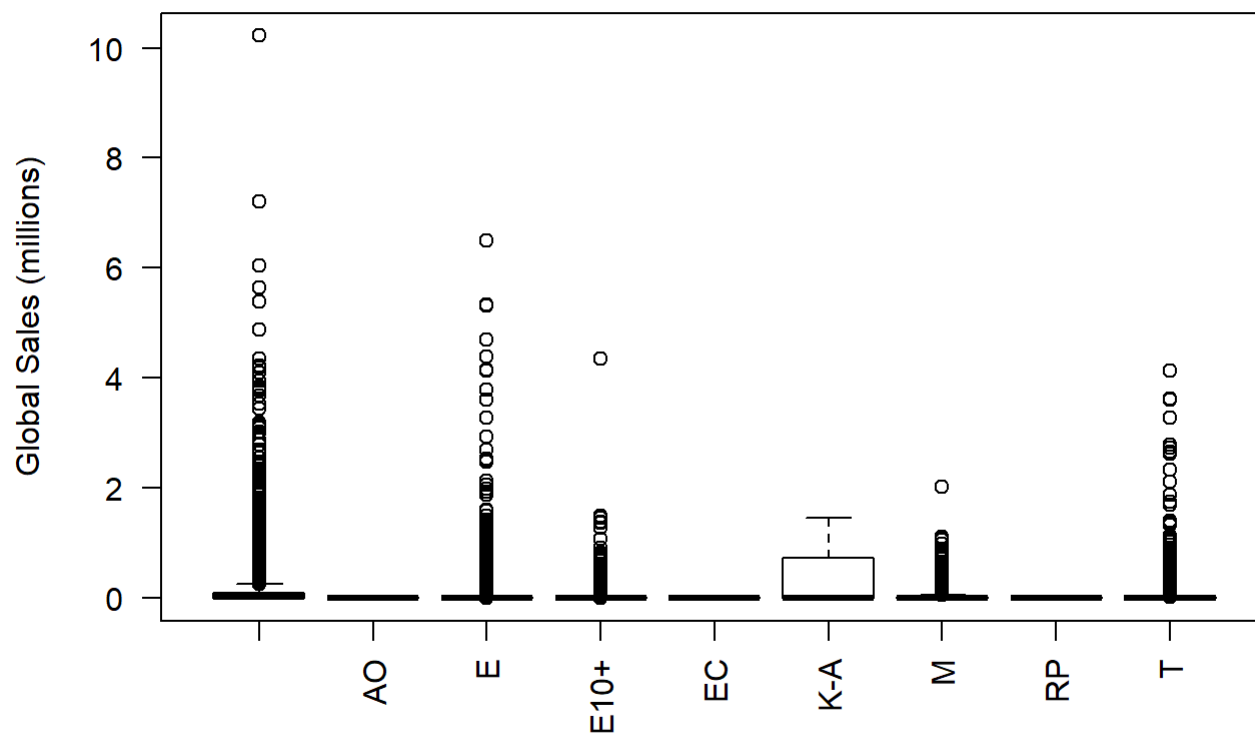
This graph displays the distribution of game sales based on their genre.

```
boxplot(JP_Sales ~ Rating, data = Games2, ylab = "Global Sales (millions)", las = 2)
```
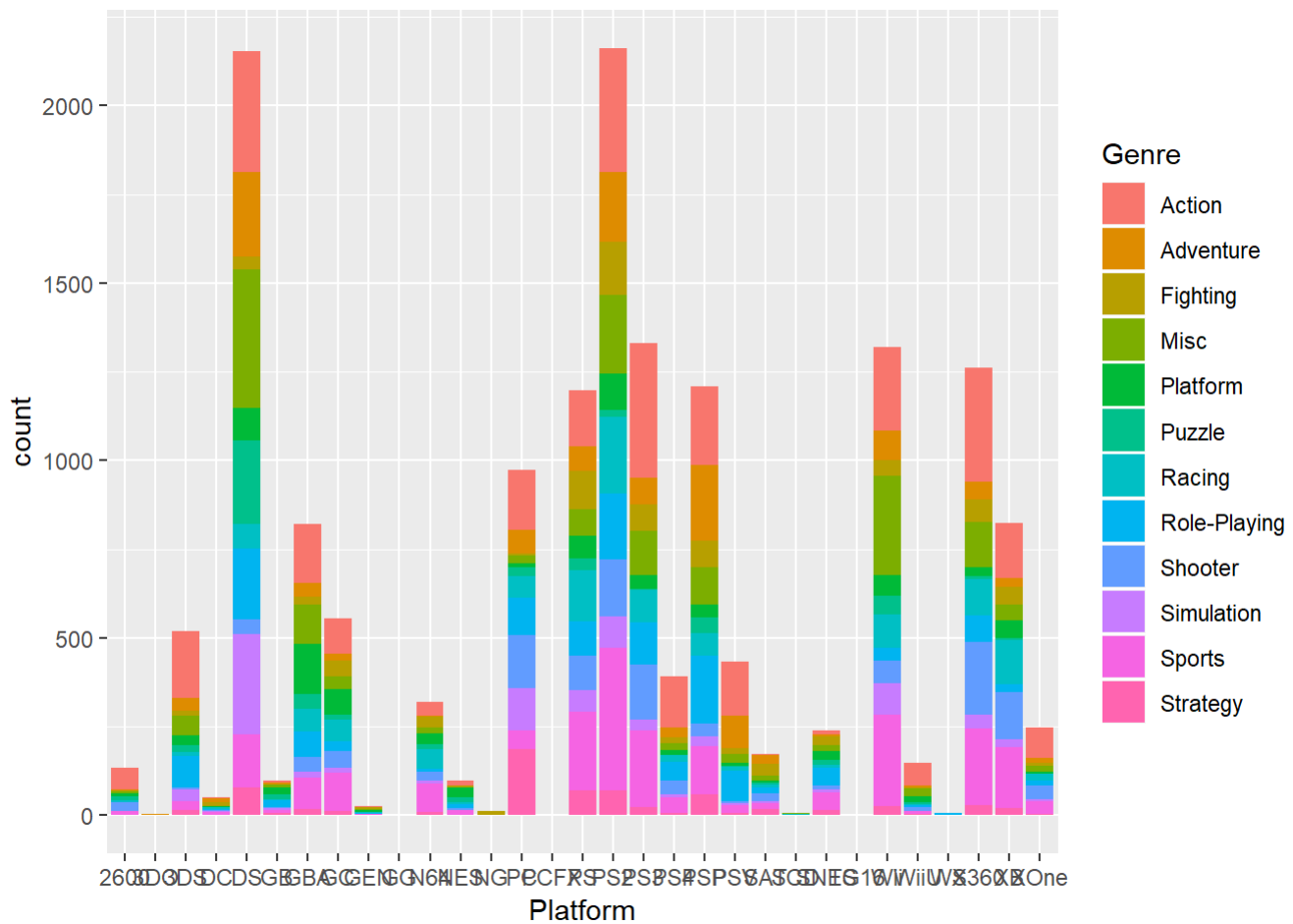
```
boxplot(JP_Sales ~ Rating, data = Games2[-c(13687, 7273),], ylab = "Global Sales (millions)", la
s = 2)
```

This graph displays the distribution of game sales based on their age rating.

```
gf_bar(~Platform, fill= ~Genre, data=Games2)
```

This graph displays the distribution of game genres for each platform.

5. Write three or four preliminary observations, likely in bullet point form, based on your initial investigations.

- Games rated E seemed to mostly sell more than games with a higher age rating. This is probably because the user base is significantly larger since more people (children) can play the game but it also indicates that there is a substantial market for casual gaming.
- Nintendo has the 15 top selling games globally based on our datatset.
- Action and Sports games are the most popular genres, this is also reflected in their respective sales.
- We see a slight relationship between critic score and sales indicating that reviews do indeed have an impact on a game's sales.