Critical Findings and Data Analysis on Movies of Different Genres

Implemented by: Sadman Sakib Mridul

Date: July 27th, 2023

Dataset Information:

The dataset that was utilized for this assignment includes information on a wide range of movies from different genres, as well as information on their budget, release year, and critic and audience ratings. Using the aid of the following visualizations: **Multiple Histograms, Scatter plot, Boxplot,** and **Polynomial Curve**, the goal of this project is to derive some intriguing and insightful conclusions from this dataset.

```
# Importing the dataset

movies = read.csv("P2-Movie-Ratings.csv")


head(movies)

##                      Film      Genre Rotten.Tomatoes.Ratings.. Audience.Ratings..
## 1 (500) Days of Summer     Comedy                        87                 81
## 2         10,000 B.C. Adventure                         9                 44
## 3            12 Rounds     Action                        30                 52
## 4            127 Hours Adventure                        93                 84
## 5            17 Again     Comedy                        55                 70
## 6                 2012     Action                        39                 63
##    Budget..million... Year.of.release
## 1                  8             2009
## 2                105             2008
## 3                 20             2009
## 4                 18             2010
## 5                 20             2009
## 6                200             2009


# changing some of the column names
colnames(movies) = c("Film", "Genre", "CriticsRatings", "AudienceRatings",
"BudgetMillions", "YearReleased")



head(movies)

##                      Film      Genre CriticsRatings AudienceRatings BudgetMillions
## 1 (500) Days of Summer     Comedy             87              81              8
## 2         10,000 B.C. Adventure              9              44            105
## 3            12 Rounds     Action             30              52             20
## 4            127 Hours Adventure             93              84             18
## 5            17 Again     Comedy             55              70             20
```

```
## 6                        2012      Action                      39              63                    200
##    YearReleased
## 1          2009
## 2          2008
## 3          2009
## 4          2010
## 5          2009
## 6          2009
```

```r
str(movies)
```

```
## 'data.frame':    562 obs. of  6 variables:
##  $ Film          : chr  "(500) Days of Summer " "10,000 B.C." "12 Rounds " "127
Hours" ...
##  $ Genre         : chr  "Comedy" "Adventure" "Action" "Adventure" ...
##  $ CriticsRatings : int  87 9 30 93 55 39 40 50 43 93 ...
##  $ AudienceRatings: int  81 44 52 84 70 63 71 57 48 93 ...
##  $ BudgetMillions : int  8 105 20 18 20 200 30 32 28 8 ...
##  $ YearReleased   : int  2009 2008 2009 2010 2009 2009 2008 2007 2011 2011 ...
```

```r
library("ggplot2")

# Plotting individual histograms for budgets of the movies, grouped by their Genres
p = ggplot(data = movies, aes(BudgetMillions))
q = p + ggtitle("Budgets of movies grouped by different genres") +
geom_histogram(binwidth = 10, aes(fill = Genre), colour = "Black") + facet_grid(Genre~.,
scales = "free")

print(q)
```

Budgets of movies grouped by different genres

*(handwritten annotations: "high variation", "low variation", "20 to 70 mill")*
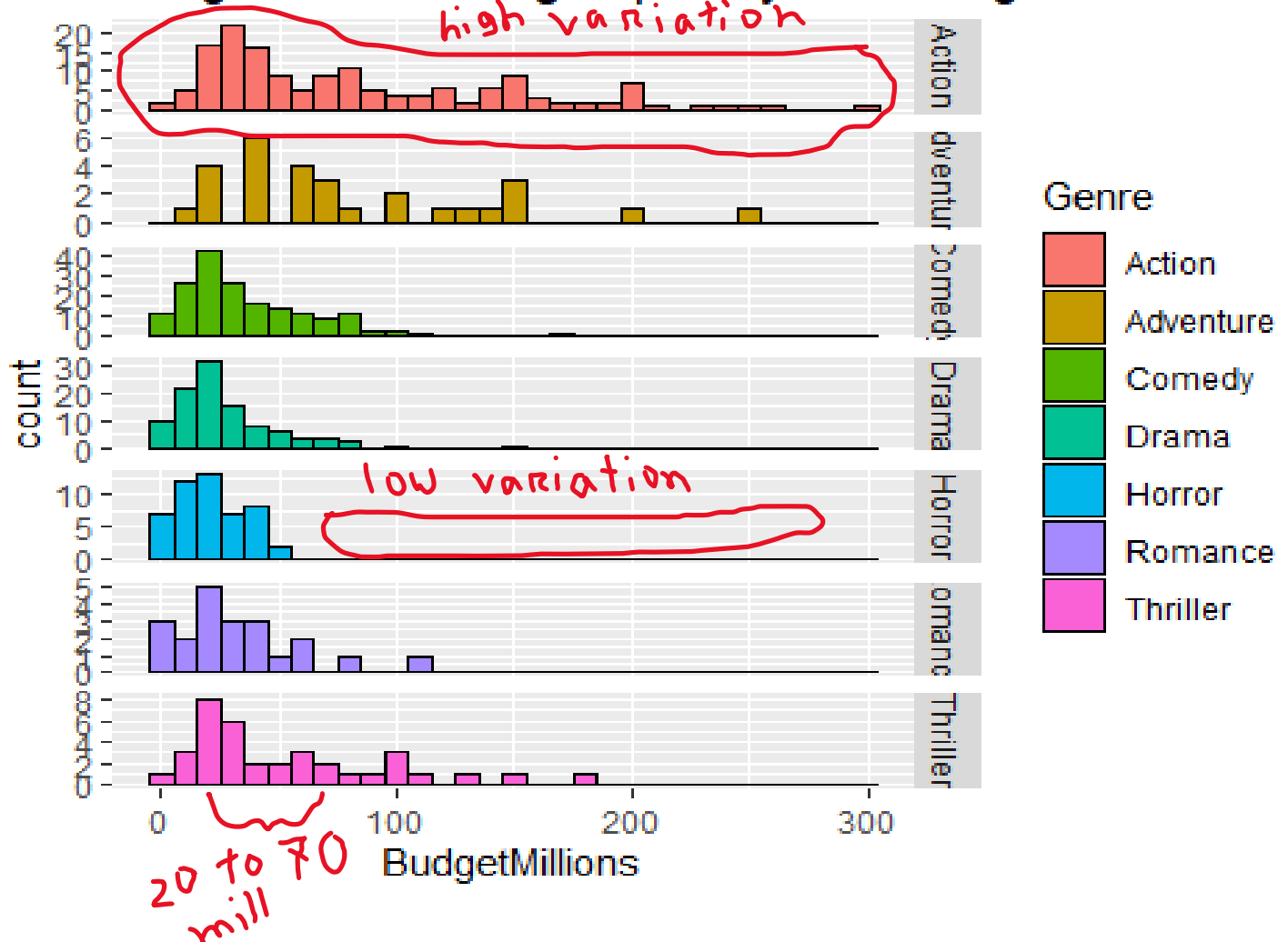
The above chart shows the budgets of movies grouped by different genres with the help of their individual histograms. There are numerous insightful conclusions that may be drawn from the figure. First, **action** movies have the greatest budget variance between lowest and highest. Many movies, across all genres, have budgets between **20** and **70** million. On the other hand, compared to films from other genres, **horror** films appear to have a very low overall budget. None of the horror films have a budget more than **$50** million. Only **action** and **adventure** films have budgets that are **$200** million or more. Generally speaking, the budgets for **adventure** and **thriller** films seem to be pretty similar.

Scatter plot:

```
s = ggplot(data = movies, aes(x = BudgetMillions, y = AudienceRatings, colour = Genre))

s + ggtitle("Relationship between Audience Ratings
and Budget") + geom_point(size = 1)
```

## Relationship between Audience Ratings and Budget
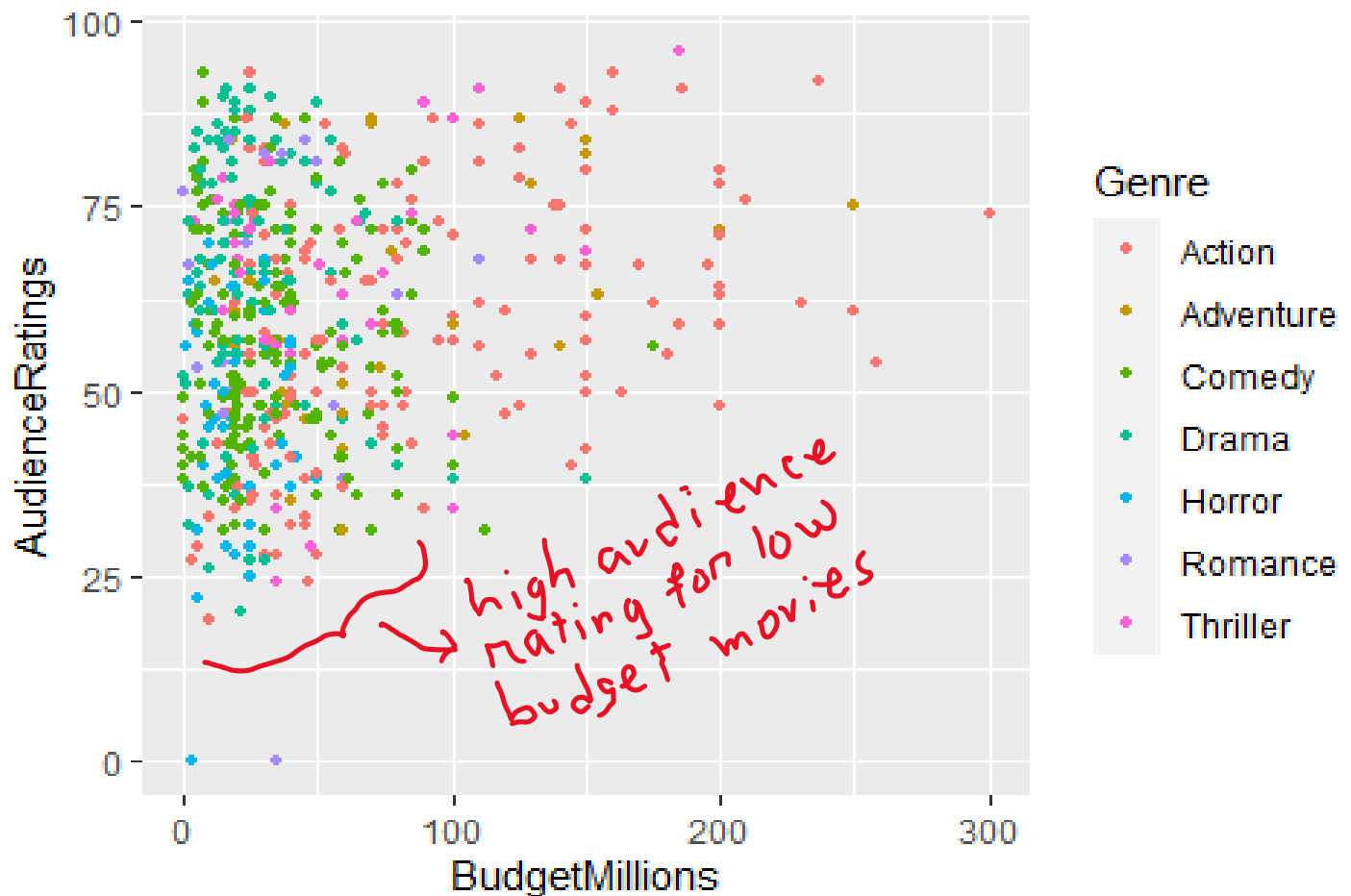
The above graph shows the relationship between the **audience rating** and their corresponding **budgets** for each movie. **Action** films typically have a substantially larger budget than those in other genres. Budgets for many action films typically ranged from $100 million to $250 million. Also, it is clear from the graph that just because a movie has a bigger budget does not guarantee that audiences would think favorably of it. In this instance, the audience is not being biased; rather, they are sharing their genuine opinions about the reviews.

Boxplot:

```
r = ggplot(data = movies, aes(x = Genre, y = AudienceRatings, colour = Genre))

t = r + ggtitle("Relationship between Audience Ratings
and Genre") + geom_jitter(size = 0.5) + geom_boxplot(size = 0.4, alpha = 0.4)
print(t)
```

Relationship between Audience Ratings and Genre
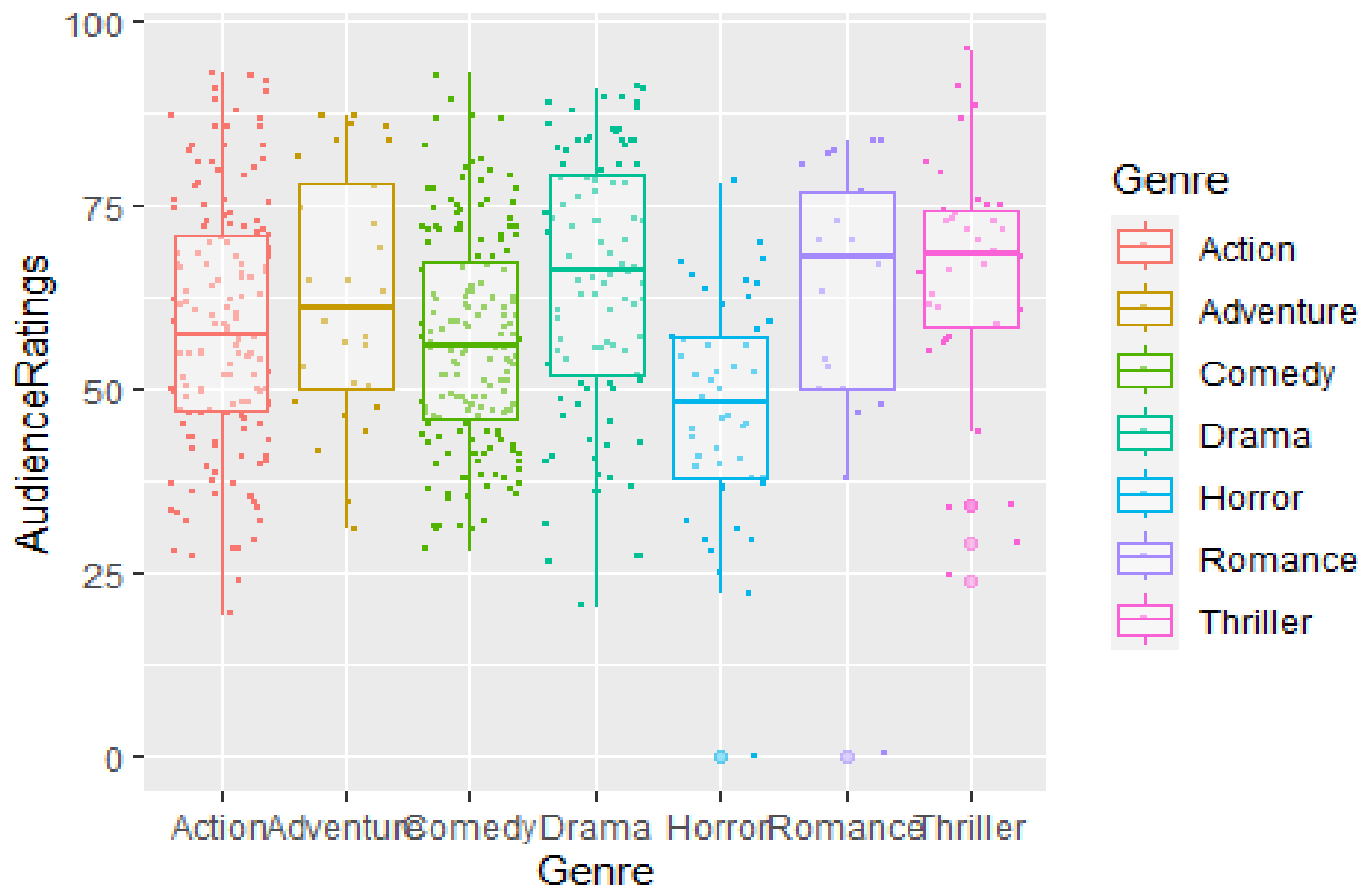
If we look at the boxplot above, we can see that audiences tend to give **thriller** movies the highest ratings because the mean values are the greatest and the variation is the lowest because the box for these films is the narrowest. Strangely, **horror** films appear to be the least popular of all the genres, suggesting that the horror film industry may prove to be very risky and less lucrative.
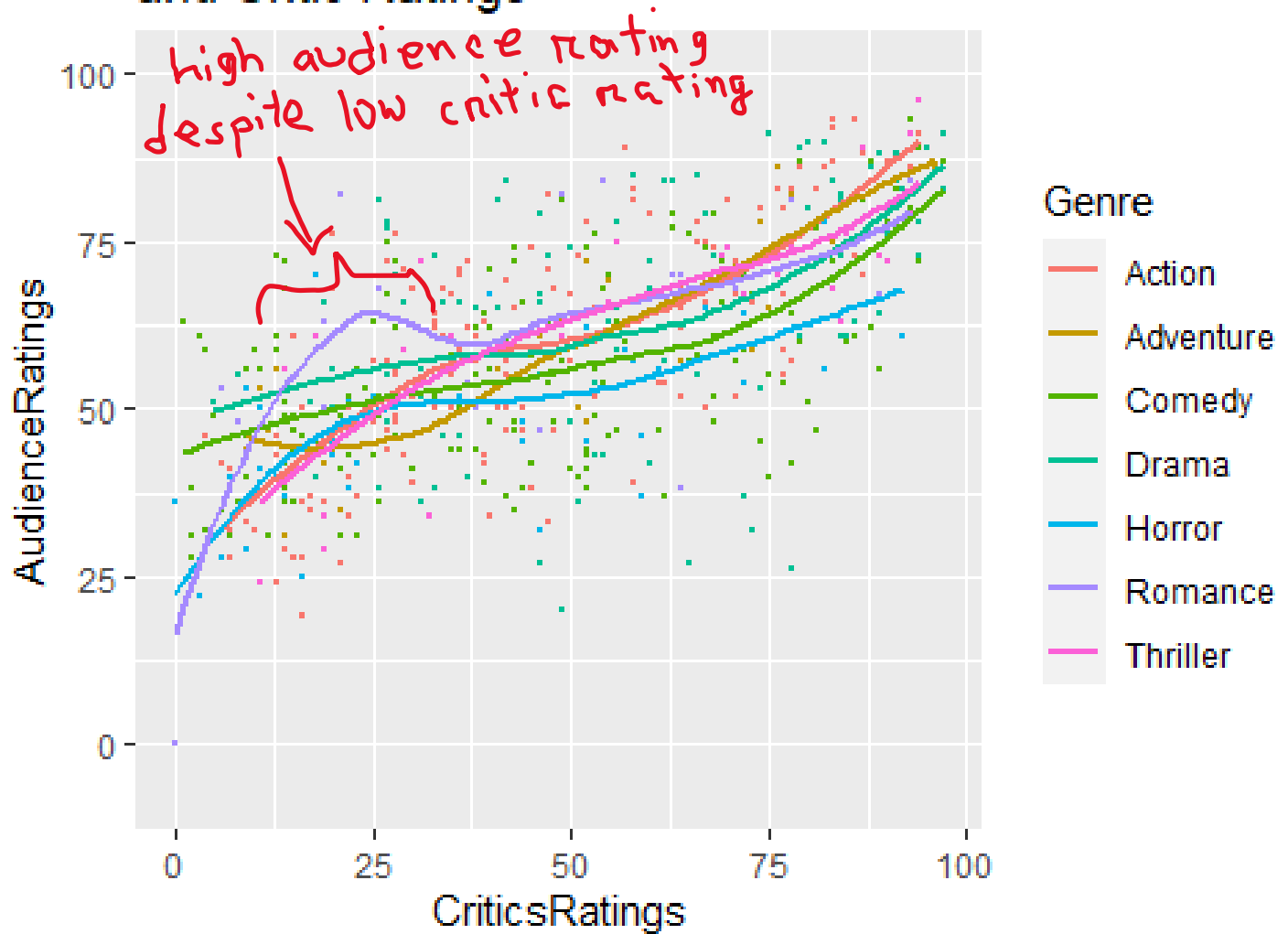
Polynomial Curve:

```
m = ggplot(data = movies, aes(x = CriticsRatings, y = AudienceRatings, colour = Genre))

n = m + ggtitle("Relationship between Audience Ratings
and Critic Ratings") + geom_point(size = 0.5) + geom_smooth(linewidth = 0.8, fill = NA)

print(n)
```

Relationship between Audience Ratings and Critic Ratings

From to the above graph, there is quite an interesting insight that can be derived. For **romance** movies (marked with purple colored line), a critic rating of around **25** out of **100** is very likely to have a high audience rating of around **63** out of **100**. This shows that the opinions of critics and the audience about romance films are very different. In this instance, the audience may appear to have a strong bias in favour of romantic movies. Another thing interesting point to note is that **horror** films (blue coloured line) appear to be continuously scored poorly by audiences when compared to critic experts.

Conclusion:

Just to sum up the results, it is clear that action movies tend to have the biggest variance and can range in budget from low to high. In most cases, they also have the largest budget, as opposed to horror movies, which typically have minimal budgets. From the audience's point of view, it appears that movies are not concerned with their budgets prior to giving reviews. Instead, they evaluate a film just on how much they like it. Also, it appears that audiences will give high ratings to romantic and thriller films but lower ratings to horror films.

Reference:

https://www.superdatascience.com/pages/rcourse

https://www.superdatascience.com/pages/rcourse

https://www.udemy.com/course/r-programming/