

**Submitted by :**

**Yash Aggarwal (2K22/MAE/21)**

**Chetan Singhal (2K22/MAE/20)**

**Dhruv Thapliyal (2K22/MAE/17)**

**Vishesh kumar pal (2K22/MAE/53)**

**Aryan bhati (2K22/MAE/)**

**INTRODUCTION**

* The purpose of our project was to leverage the capabilities of ggplot2 and explore its functionalities for data visualization.
* We aimed to showcase various aesthetics, layering techniques, statistical transformations, and customizations available in ggplot2 to create informative and visually appealing visualizations.
* For our project, we worked with a dataset (P2.Movie.Ratings) consisting of movie information, including the film name, genre, critic rating, audience rating, budget in millions, and year.

Summary of the presentation

ANALYSIS

**Scatter plot**

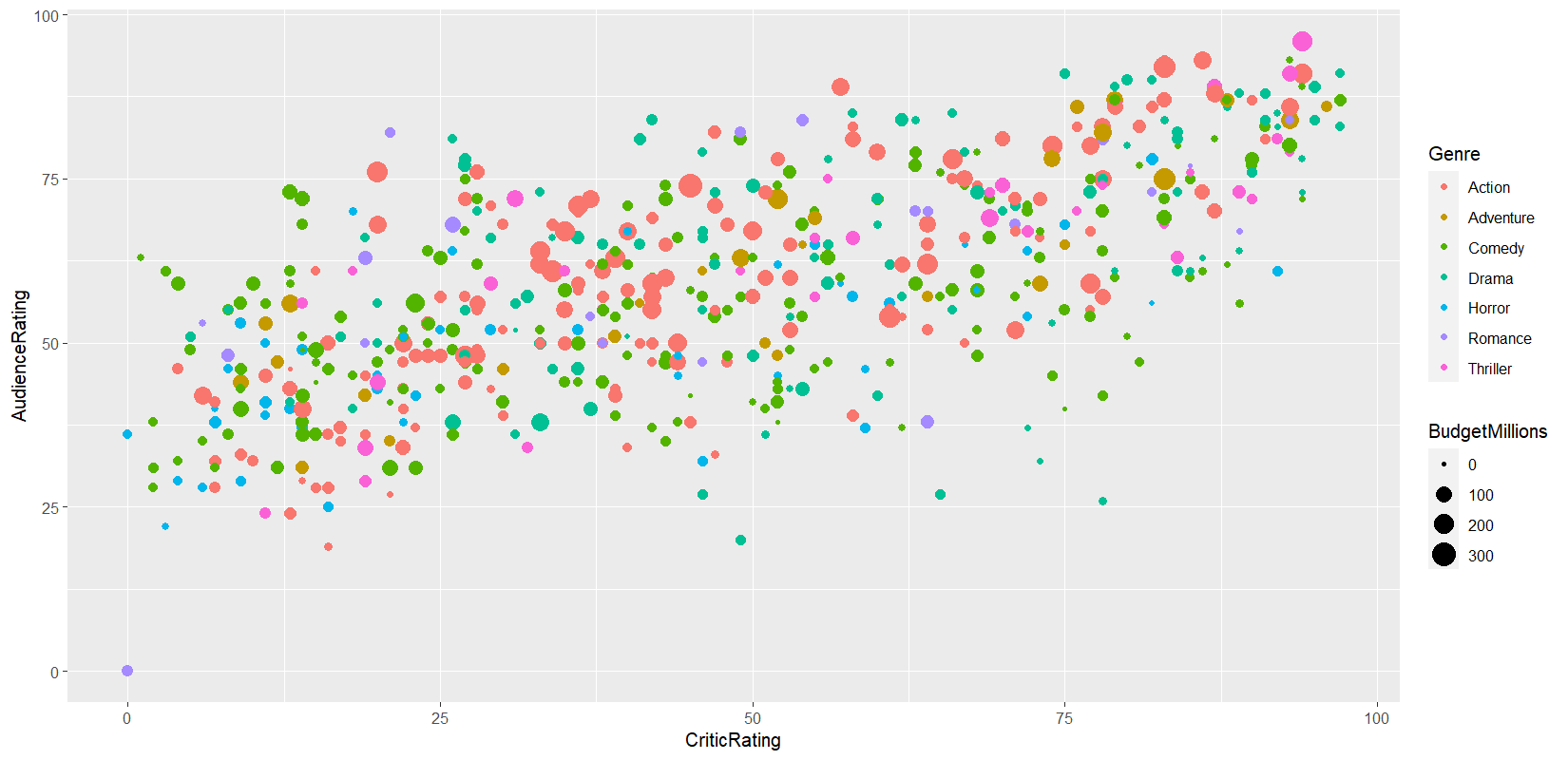
Scatter plots are particularly useful for identifying patterns or trends in the data, such as whether there is a positive or negative correlation between the two variables. They can also be used to identify outliers or unusual data points.

We Create a scatter plot with CriticRating on the x-axis, AudienceRating on the y-axis, color mapped to Genre, and size mapped to BudgetMillions

code

ggplot(data = movies, aes(x = CriticRating, y = AudienceRating)) +

geom\_point(aes(colour = Genre, size = BudgetMillions))



Brief explain

In this section, we created a scatter plot using ggplot2. We mapped the critic rating to the x-axis and the audience rating to the y-axis to examine the relationship between these two variables.

- To enhance the scatter plot, we added color aesthetics based on the movie genre and size aesthetics based on the budget in millions. This allowed us to incorporate additional dimensions into the visualization.

Boxplot

In a box plot, the data is divided into quartiles, and the box represents the middle 50% of the data, with the median shown as a line inside the box.

Box plots are particularly useful for comparing the distribution of data across different groups or categories, as multiple box plots can be drawn side-by-side to show differences in the data.

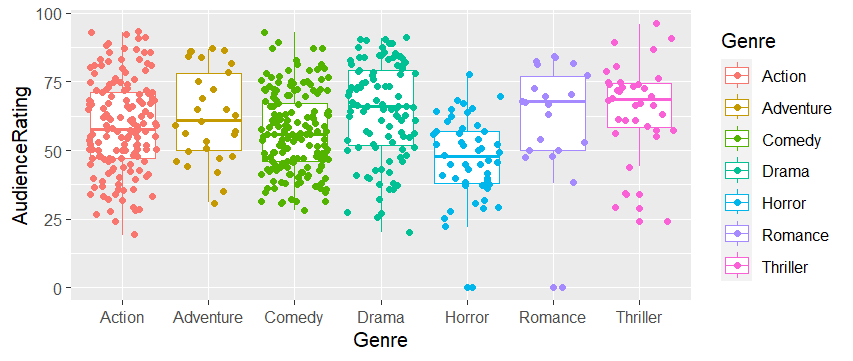
We Create a boxplot of AudienceRating based on Genre, with color differentiated by Genre, and add jitter points for better visualization

Codes –

ggplot(data = movies, aes(x = Genre, y = AudienceRating, colour = Genre)) +

geom\_boxplot() +

geom\_jitter()



Brief explanation

We employed boxplots to analyze the distribution of audience ratings across different movie genres.

- To better understand the distribution and identify potential outliers, we added jitter points to the boxplots. These jitter points allow us to visualize individual data points more clearly.

Histogram

In a histogram, the data is divided into a set of intervals, or "bins," and the number of data points that fall into each bin is shown as a bar. The height of each bar represents the frequency of data points in that bin.

Histograms are particularly useful for identifying patterns or trends in the data, such as whether the data is normally distributed or skewed.

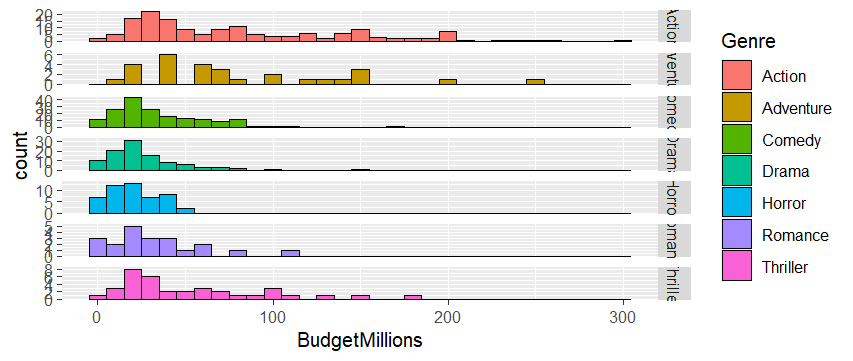
# Create histograms of BudgetMillions, with bins of width 10, filled by Genre,

# and organized by Genre using facet\_grid

ggplot(data = movies, aes(x = BudgetMillions)) +

geom\_histogram(binwidth = 10, aes(fill = Genre), colour = "black") +

facet\_grid(Genre ~ ., scale = "free")



Brief explanation

Density chart

In a density plot, the data is represented as a smooth curve that shows the estimated probability density function of the variable.

Density plots are particularly useful for identifying patterns or trends in the data, such as whether the data is normally distributed or skewed.

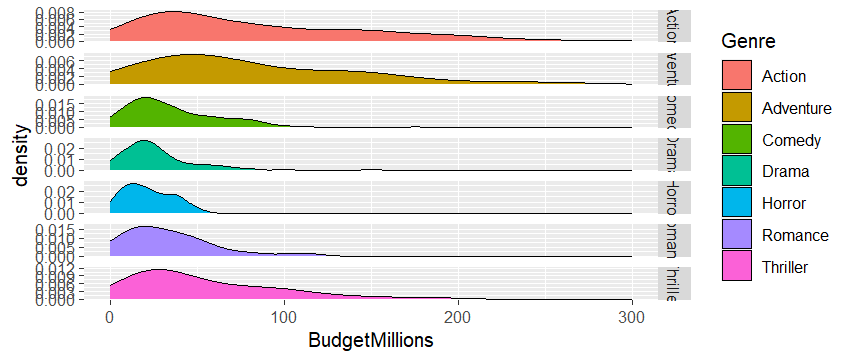
We Create density charts of BudgetMillions, filled by Genre, and organized by Genre using facet\_grid

code

ggplot(data = movies, aes(x = BudgetMillions)) +

geom\_density(aes(fill = Genre)) +

facet\_grid(Genre ~ ., scale = "free")



Explain

Scatter plot

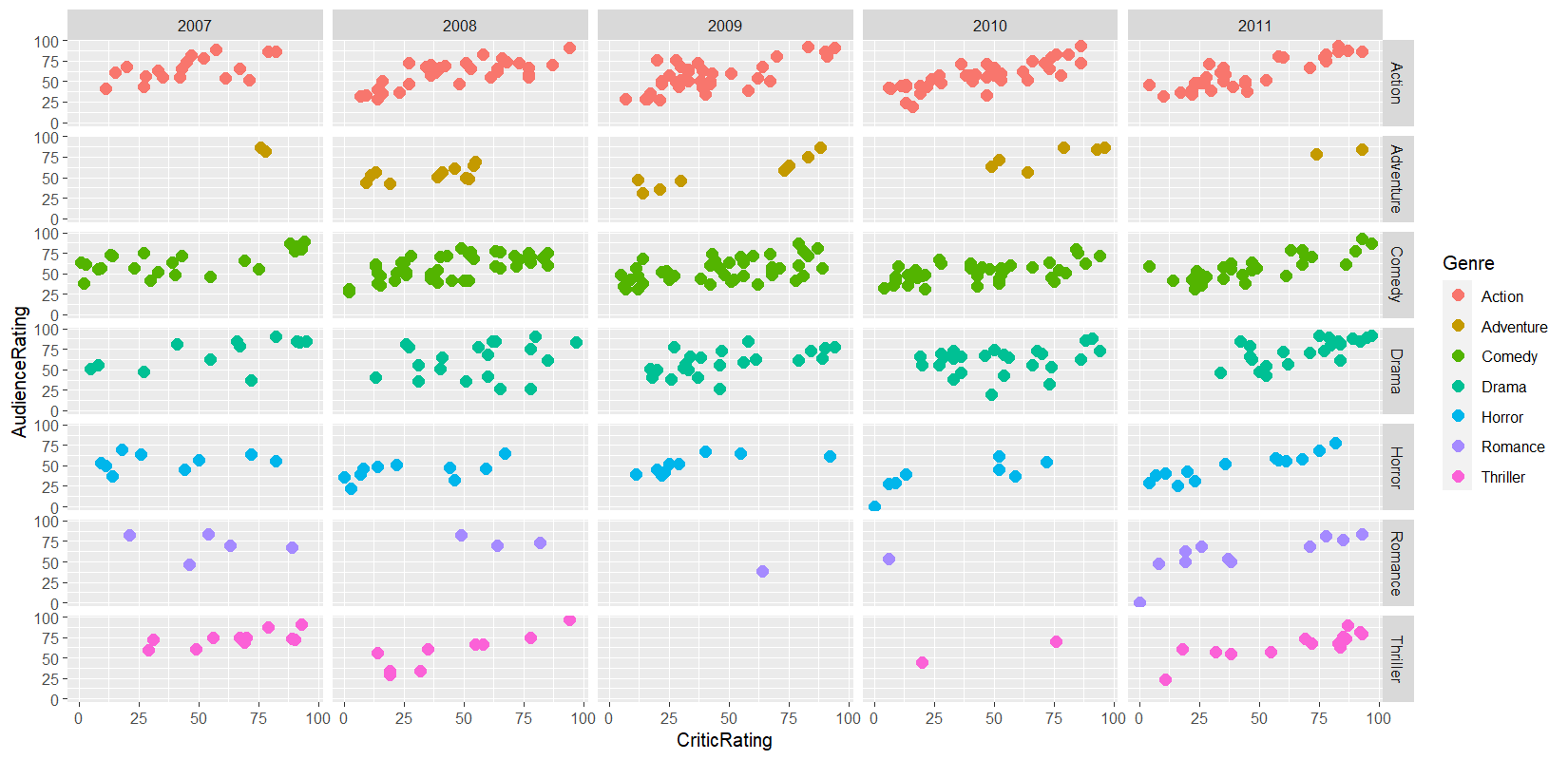
# Create scatter plots of CriticRating vs. AudienceRating, with color mapped to Genre,

# and facets organized by Genre and Year

ggplot(data = movies, aes(x = CriticRating, y = AudienceRating, colour = Genre)) +

geom\_point(size = 3) +

facet\_grid(Genre ~ Year)



# Create scatter plots of CriticRating vs. AudienceRating, with color mapped to Genre,

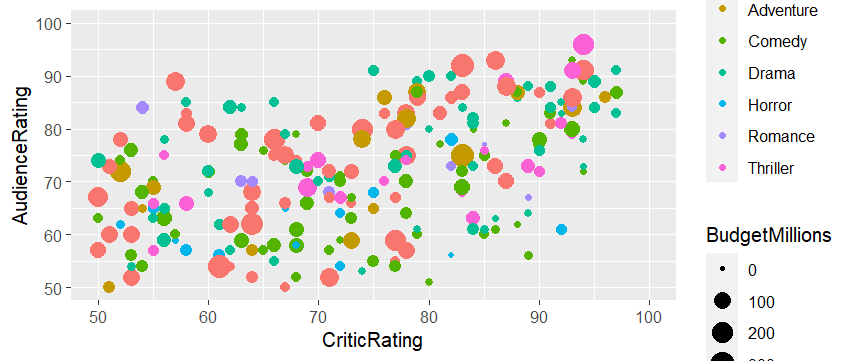
# size mapped to BudgetMillions, and control the coordinate limits

ggplot(data = movies, aes(x = CriticRating, y = AudienceRating, size = BudgetMillions, colour = Genre)) +

geom\_point() +

xlim(50, 100) +

ylim(50, 100)



Explain

----

# Customize the theme by modifying axis labels, tick mark sizes, legend appearance, and plot title

ggplot(data = movies, aes(x = BudgetMillions)) +

geom\_histogram(binwidth = 10, aes(fill = Genre), colour = "black") +

xlab("money axis") +

ylab("number of movies") +

theme(axis.title.x = element\_text(colour = "brown", size = 30),

axis.title.y = element\_text(colour = "Red", size = 30),

axis.text.x = element\_text(size = 20),

axis.text.y = element\_text(size = 20),

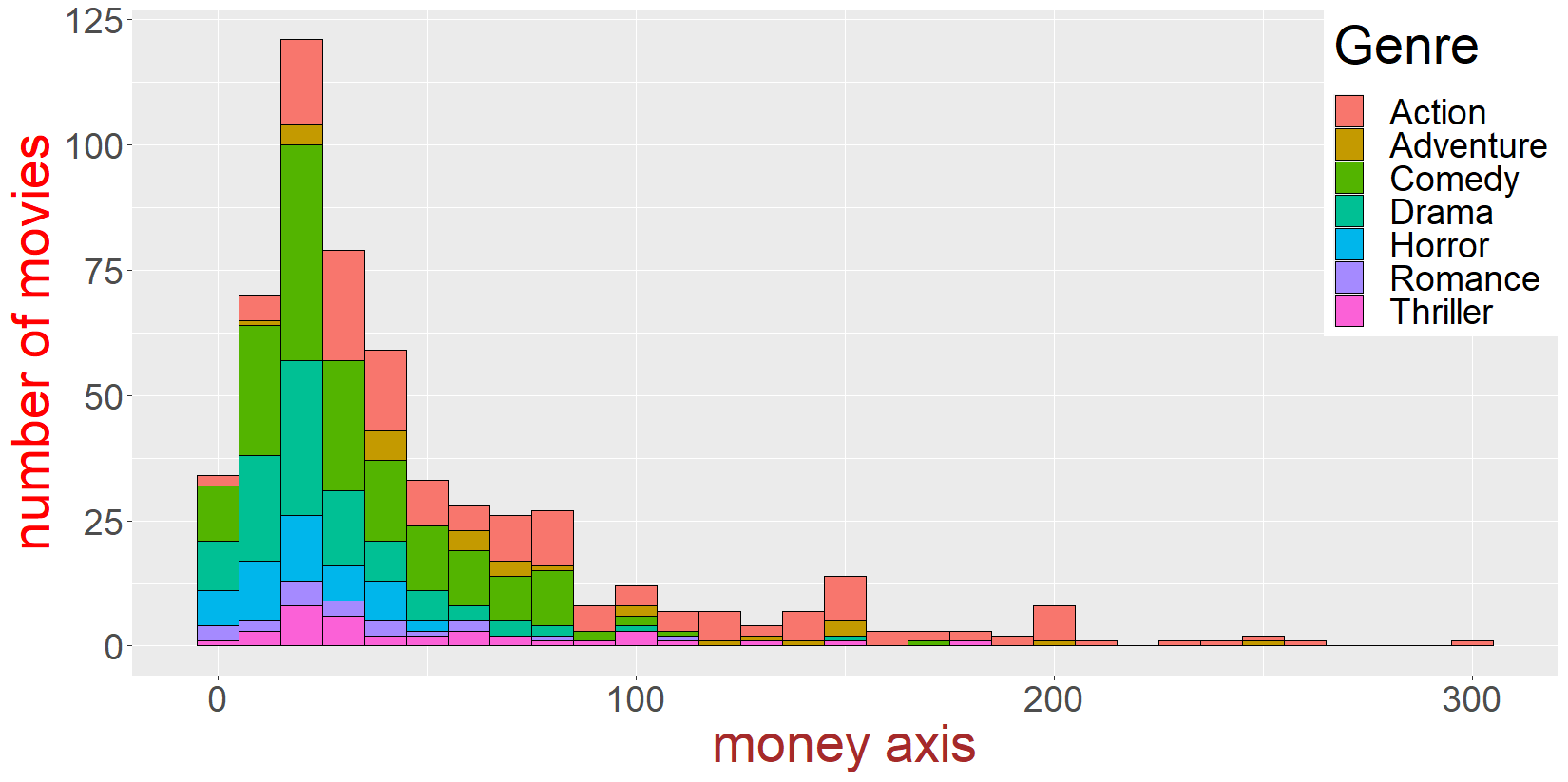
legend.title = element\_text(size = 30),

legend.text = element\_text(size = 20),

legend.position = c(1, 1),

legend.justification = c(1, 1),

plot.title = element\_text(colour = "DarkBlue", size = 40, family = "Courier"))



Explain

- With ggplot2, we can control the coordinate limits to focus on specific regions of interest within the plot.

- We demonstrated how to customize the axes labels, adjust tick mark sizes, modify legend appearance, and add a plot title for improved clarity and aesthetics.

Conclusion

- In conclusion, ggplot2 is a powerful data visualization library in R that offers a wide range of capabilities for exploratory data analysis.

- Throughout this presentation, we have showcased various techniques, including scatter plots, boxplots, histograms, density charts, facets, and coordinate control, that can be used to create informative and visually appealing visualizations.

- We hope this demonstration has highlighted the potential of ggplot2 for data visualization and inspires you to explore further and apply these techniques to your own data analysis projects.