

Final project

2024-04-12

load the library

call the data set

```
AnnualTicketSales <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/AnnualTicketSales.csv")
HighestGrossers <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/HighestGrossers.csv")
PopularCreativeTypes <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/PopularCreativeTypes.csv")
TopDistributors <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/TopDistributors.csv")
TopGenres <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/TopGenres.csv")
TopGrossingRatings <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/TopGrossingRatings.csv")
TopGrossingSources <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/TopGrossingSources.csv")
TopProductionMethods <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/TopProductionMethods.csv")
WideReleasesCount <- read.csv(file = "/Users/wangyi/Desktop/STAT 184/Final-Project/final project/WideReleasesCount.csv")
```

Some EDA of Q1

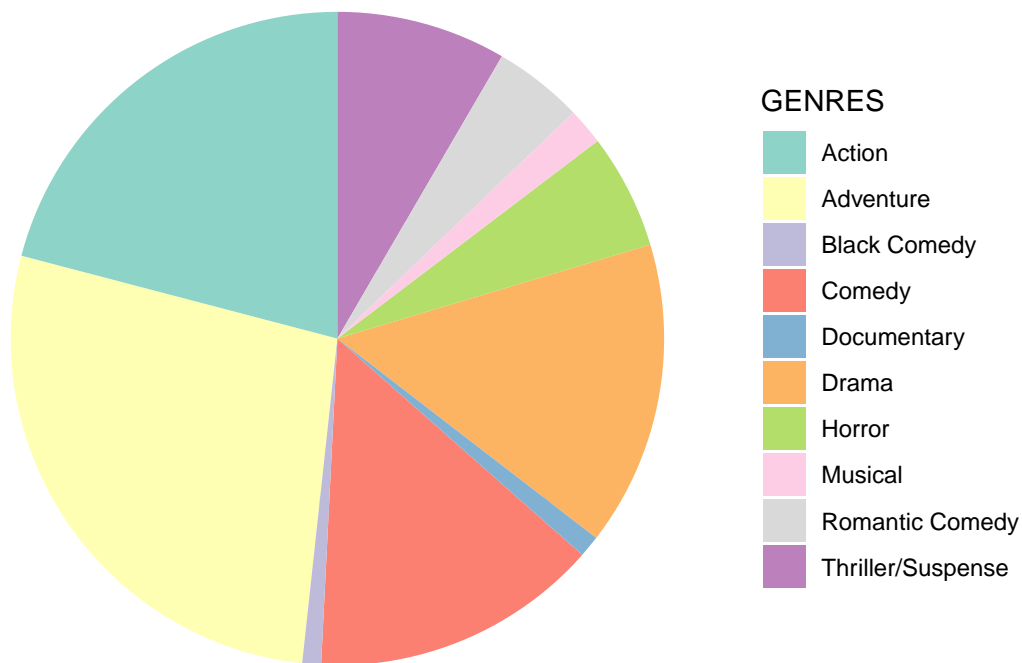
##Most popular Geners basde on Earning ##arrange from the highest to the lowest

```
# Data preparation
TopGenres$`TOTAL.GROSS` <- as.numeric(gsub("\\$|,", "", TopGenres$`TOTAL.GROSS`))

# Reorder GENRES based on TOTAL.GROSS
TopGenres <- TopGenres[order(TopGenres$`TOTAL.GROSS`, decreasing = TRUE), ]

# Create a pie chart using ggplot2
ggplot(TopGenres, aes(x = "", y = `TOTAL.GROSS`, fill = GENRES)) +
  geom_bar(width = 1, stat = "identity") + # Use geom_bar with identity stat to create pie chart effect
  coord_polar("y", start = 0) + # Convert the bar chart into a pie chart
  theme_void() + # Remove axes and gridlines
  scale_fill_brewer(palette = "Set3") + # Set color palette for genres
  labs(title = "Total Gross by Genre") + # Chart title
  theme(plot.title = element_text(hjust = 0.5)) # Center the title
```

Total Gross by Genre



```
# using ggplot to explore the specific detail of total gross of genres
ggplot(TopGenres, aes(x = GENRES, y = TOTAL.GROSS)) +
  geom_col(width = 0.5, fill = "skyblue", color = "black") + # Customize bar appearance
  geom_segment(aes(x = GENRES, xend = GENRES, y = 0, yend = TOTAL.GROSS + 5), # Add pin segments
    color = "black", size = 1) +
  geom_point(aes(x = GENRES, y = TOTAL.GROSS + 5), shape = 18, size = 2.5, color = "black") + # Add pin
  labs(x = "Genres Of Movies", y = "Total Gross Price in billion of dollars", title = "Bar Graph Of Total Gross By Genre") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate x-axis labels
  scale_y_continuous(limits = c(2000000000, 70000000000)) +
  scale_y_continuous(labels = function(x) format(x / 1e7, scientific = FALSE, big.mark = ",", decimal.mark = ".")) +
  theme_minimal() # Apply a minimal theme
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```

Bar Graph Of Total Gross Price In Billion Of Dollars by Genres Of Movies

