Untitled

Library

```
library(tidyr)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Load Data

```
ratings <- read.delim(gzfile("Data_files/title.ratings.tsv.gz"), sep="\t")
basics <- read.delim(gzfile("Data_files/title.basics.tsv.gz"), sep="\t")
```

Combine and deal with data.

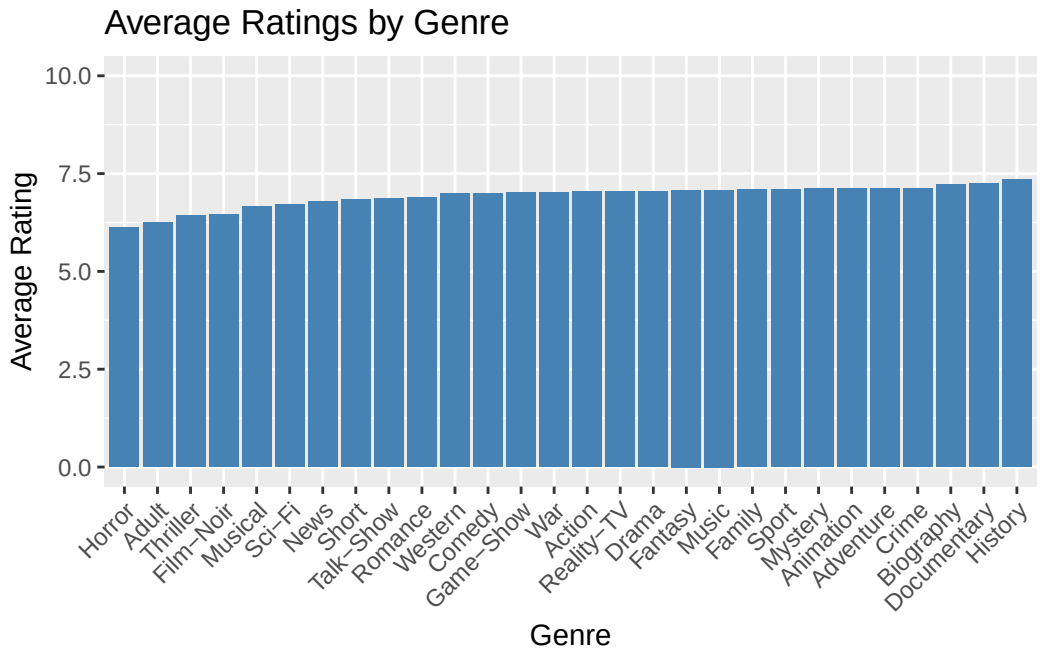
```r
# Performing an inner join on the 'tconst' column
combined_data <- merge(basics, ratings, by="tconst")

genre_filter_data <- combined_data %>%
  separate_rows(genres, sep = ",") %>%
  filter(genres != "" & !is.na(genres) & genres != "\\N")
```

Average Ratings By Genre

```r
average_ratings_by_genre <- genre_filter_data %>%
  group_by(genres) %>%
  summarise(averageRating = mean(averageRating, na.rm = TRUE), .groups = 'drop')
```

Visualization:

```r
ggplot(average_ratings_by_genre, aes(x = reorder(genres, averageRating), y = averageRating
  geom_bar(stat = "identity", fill = "steelblue") +
  ylim(0, 10) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(x = "Genre", y = "Average Rating", title = "Average Ratings by Genre")
```



Rating Trends Over Time

```
startYear_filter_data <- combined_data %>%
  filter(startYear != "" & !is.na(startYear) & startYear != "\\N")

startYear_filter_data$startYear <- as.numeric(startYear_filter_data$startYear)
```

```
average_ratings_by_year <- startYear_filter_data %>%
  group_by(startYear) %>%
  summarise(averageRating = mean(averageRating, na.rm = TRUE), .groups = 'drop')
```

```
ggplot(average_ratings_by_year, aes(x = startYear, y = averageRating)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = seq(min(average_ratings_by_year$startYear, na.rm = TRUE), ma
  labs(x = "Start Year", y = "Average Rating", title = "Rating Trends Over Time") + theme_
```