|  |
| --- |
| Data Analysis Hollywood Movie Project |
| Assignment 3 |

|  |
| --- |
| Lozarta V  5-31-2024 |

**R Studio**

I loaded the Hollywood Movie dataset into R studio as CSV, cleaned it and ran summary statistics, exporting as CSV. Below is a screenshot from the script editor in R studio.

A screenshot of a computer

Description automatically generated

**Explanation of steps:**

1. Loading data: reads file in CSV format, creating a data frame from it – assigning data frame to the variable df

df<- read.csv("<https://public.tableau.com/app/sample-data/HollywoodsMostProfitableStories.csv>")

1. Viewing data: allows user to view the data table

View(df)

1. Loading library: tidyverse packages loaded for data analysis, ‘tidy data’.

install.packages("tidyverse")

1. Importing library: tidyverse imported

library(tidyverse)

1. Checking data types: the data types used are Numeric (integers & float), Integer(whole number), Char(string of text).

str(df)

1. Checking for missing values: important to identify for data analysis

colSums(is.na(df))

1. Dropping missing values: in order to handle missing values, in this case the amount of missing data is small

df <- na.omit(df) or df <- df %>% drop\_na()

1. Checking to make sure the rows have been removed

colSums(is.na(df))

1. Summary statistics: exploratory data analysis

summary(df)

1. Scatterplot: exploratory data analysis

ggplot(df, aes(x=Lead.Studio, y=Rotten.Tomatoes..)) + geom\_point()+ scale\_y\_continuous(labels = scales::comma)+coord\_cartesian(ylim = c(0, 110))+theme(axis.text.x = element\_text(angle = 90))

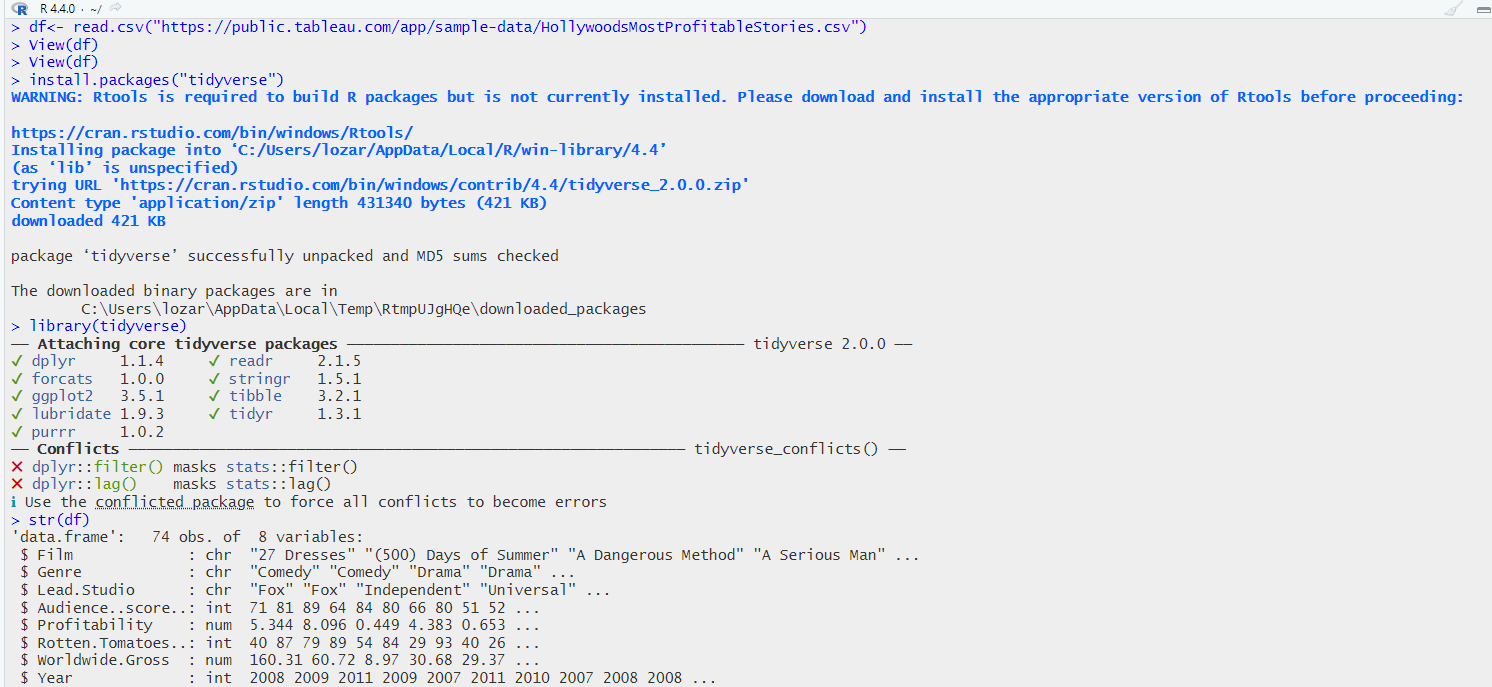
1. Bar Chart: exploratory data analysis

ggplot(df, aes(x=Year)) + geom\_bar()

1. Exporting clean data: CSV file created as clean\_df.csv

write.csv(df, "clean\_df.csv")

Below is a screenshot showing steps from loading the data to checking the data types.



This screenshot shows steps from checking for missing values to exporting clean data as CSV.

A close-up of a computer screen

Description automatically generated

Exploratory Analysis

**Exploratory Analysis**

The scatter plot below shows each lead studio by the number and spread of Rotten Tomatoes ratings.

A graph with black dots

Description automatically generated

The bar chart below shows the count of films for each year.

A graph of a number of bars

Description automatically generated with medium confidence

In addition, I produced a bar chart for count of each genre below.A graph of a bar chart

Description automatically generated

A graph with a number of squares

Description automatically generated with medium confidenceI also produced a boxplot of lead studios and Rotten Tomatoes ratings.

**Power BI Dashboard**

Below is a screenshot of my Power BI dashboard, after importing the CSV and transforming the data.

In the table view, I changed the column names for better readability, as well as reducing the number of decimal points for the Profitability and Worldwide Gross. I also set the ‘count’ of films by year, and calculated the average Rotten Tomatoes Rating by Genre using a DAX calculation.

A screenshot of a graph

Description automatically generated