Lab 5 Data Visualization (Project 1)

The Movie Data Set

Data for 5000 movies has been provided in your lab folder in the spreadsheet 5000movies.xlsx. The data set include the following information about movies:

budget, genre, id, original_language, original_title, popularity, main_production_company, production_country, release_date, revenue, runtime, spoken_languages, title, vote_average, vote_count

Preparing the data

• Import the data set into R.
• Document the steps for the import process and any preprocessing that needs to be done prior to or after the import (e.g. creating new variables, etc.)

Analyzing the data

Do the analysis as in the lecture for categorical and numerical data.

- Perform a qualitative univariate analysis: report the appropriate descriptive statistics, showing the appropriate plots for your data
- Perform a quantitative univariate analysis: report the appropriate descriptive statistics, showing the appropriate plots for your data
- Perform a qualitative bivariate analysis and show the appropriate plots for your data
- Perform a quantitative bivariate analysis and show the appropriate plots for your data
- Perform a qualitative and quantitative bivariate analysis and show the appropriate plots for your data

Requirements

- You must perform all the five analyses mentioned above.
- You must provide at least one plot using base graphics system and one ggplot for each analysis.
- Make sure you have covered different types of plots and discover new ones. (Doing this will give you more points.)
- Don't forget to provide descriptive statistics.
- Briefly discuss what you learned from the data and analysis. Examples:

- Movies made in the U.S. generate higher revenue.
- Action movies have higher average vote.

Submitting Your Work

Write a report which documents every step for this lab. Your report should include R code and results. Submit the report and R files.