Data X Berkeley

HW5: Data Visualization III: Tableau

(Total 60 points)

Data Visualization using Tableau for IMDB Movie analysis. Following the instructions and requirements to generate your Tableau worksheets and dashboard. Submit a **report** as well as your **.twb** file. This homework will be graded manually.

Instructions:

1. Tableau Installation
   To download Tableau free education license, visit https://software.berkeley.edu/tableau. Choose "Students Request Form" and fill in all the information. You will then receive an email containing your Tableau Desktop Product Key. Download Tableau Desktop and activate it with your product key.

2. Dataset Download
   Download the dataset **movies.csv**. This dataset contains 6820 movies from 1986 to 2016. The details of the dataset can be found in https://www.kaggle.com/danielgrijalvas/movies.

Tableau Questions:

1. Load Dataset and Generate Bar Plot (9 points)
   Load the dataset in Tableau, generate a horizontal bar plot showing the distribution of the scores. The x-axis will be scores. Set the bin size to 0.5. The y-axis will be the number of movies.
   Additional requirements:
   - Modify the x-axis to the actual score ranges by editing the alias (i.e. 3.5-3.9, 4.0-4.4, etc.).
   - Rename the y-axis to "Number of movies"
   - Set the worksheet name to "Movie scores bar plot"
   - Add a caption describing what is on this sheet.
   - Export the sheet to an image and paste it into your report.

2. Generate Pie Chart (9 points)
   Generate a pie chart based on movie genres. Only plot the 4 movie genres: Action, Comedy, Drama, and Horror. The color shows details about the genre. Size shows the count of movies.

Additional requirements:
- Use different colors for each genre.
- Show the names of each genre by adding the labels.
- For each genre, show the percent of total under the genre names. (for example, 28% under Drama)
- Set the worksheet name to "Movie genre pie chart"
- Add a caption describing what is on this sheet.
- Export the sheet to an image and paste it into your report.

3. Generate Highlight Table (9 points)

Generate a highlight table showing the number of movies that belong to each country.

Additional requirements:
- Modify the color settings so that it only contains 2 colors based on the number of movies: (if more than 100 movies, the color will be red; if less than 100 movies, the color will be blue)
- Set the worksheet name to "Movie numbers of each country"
- Add a caption describing what is on this sheet.
- Export the sheet to an image and paste it into your report.

4. Generate Treemap (9 points)

Generate a treemap containing the number of movies from each country, and the **average** score from each country. The size of each rectangle indicates the number of movies from each country, and the color indicates the average scores of the country. Only show the countries with **more than 50** movies.

Additional requirements:
- Set the color to be "red-blue diverging", with blue represents lower scores, and red represents higher scores.
- Show the country names, and the average scores of each country by adding the labels.
- Set the worksheet name to "Movie scores and counts of each country"
- Add a caption describing what is on this sheet.
- Export the sheet to an image and paste it into your report.

5. Generate Packed Bubble (9 points)

Generate a packed bubble plot showing the genre of all the movies. Use different colors for different genres, and the bubble size indicates the number of movies.

Additional requirements:
- Add the name of genres into the label.
- Set the worksheet name to "Movie genre bubble plot"
- Add a caption describing what is on this sheet.
- Export the sheet to an image and paste it into your report.

6. Generate Symbol Map (9 points)

    Generate a symbol map showing the number of movies from each country, and the average scores. The **size** of the circles depends on the number of movies in this dataset, and the **color** of the circles indicates the **average** movie scores in the countries. Only mark the countries with more than 20 movies.

    Additional requirements:

    ● Set the color to "red-blue diverging". Red indicates higher scores, and blue represents lower scores.
    ● Add the country names as the label.
    ● Set the worksheet name to "Movie number and score symbol map"
    ● Add a caption describing what is on this sheet.
    ● Export the sheet to an image and paste it into your report.

7. Generate Dashboard (6 points)

    Generate a Dashboard and add the above 6 worksheets to it. Format this dashboard as you like, name it as "HW5_Yourname", export it to an image, and paste the image into your report.

Submit:

A **report** (.pdf) with the 7 questions above, and your **Tableau file** (.twb). You don't need to write too many words for the report, just include the figures.