Movie dataset

1. Getting to know the data

a. Import the data
b. View the data
c. Look at column names
d. Look at dimension of data (rows and columns)

2. Scatterplots

a. do scatter plot of Tickets Sold and Gross (Is the trend expected?)
b. What is the correlation between tickets sold and sales? Is this expected?

3. Other plots

a. do boxplot
b. do histogram for type of films
c. do histogram of ticket sales. Try different bin numbers.
d. Add frequency count to top of bars.
e. do barplot of genre

4. Create a set of functions that compute specific metrics by genre:

- average_number_of_tickets_sold
- average_gross_sales

5. Use the functions created in 4 to compute these metrics by distributor and genre.

6. Make a shiny app that allow to explore the datasets using the above points and add any additional feature you wish.

Winter Olympic

Create a `RMarkdown` (.Rmd) document that answers that addresses the following requirements.

1. Getting to know the data:

a. Import the data
b. View the data
c. Look at column names
d. Look at dimension of data (rows and columns)

2. Data is currently sorted by Rank. Sort data by total medals and country. Assign sorted data to a new data frame. Call it sort_total.

3. Use some function to look at data.

4. Look at some statistics

a. What is median of number of gold, silver, bronze and total medals?
b. Also look at the mean and total number of G, S, B and T medals

6. More statistics - subset

a. Redo above statistics, this time group by Region
b. Which region won the highest mean total medals?
c. How many countries are in this Geographic Region?
d. How many countries are in the EUROPE group?
e. What is the max number of medals won? What country won the max?

Summer-winter Olympic

1. Dealing with Data

a. Look at the column names and change names to more meaningful names.
b. The data represent, in order:
    1. country
    2. number of summer games played, gold, silver, bronze, total,
    3. number of winter games played, gold, silver, bronze and total, total
    4. total (Winter + Summer) games, gold, silver, bronze, total
c. use table() to find frequency of total summer games played
d. explore the data with other variables

4. Graphs

a. do histogram of summer games (total)
b. do histogram of winter games (total)
c. put above two histograms on one page
d. do two histograms on one page: total summer, total winter medals won
e. is there a correlation between number of medals given out in winter and summer? (do plot)
f. how about number of games each country competes in. Is there correlation between winter and summer?
g. look at distribution of each of the types of medals, by season (6 histograms on one page)
h. redo g with different number of bins (10 instead of 20)
i. explore data on your own