

Attached is a dataset titled 'Films' which has data of Hollywood Movies from 1975 to 2018. The dataset was taken from Kaggle (source: <https://www.kaggle.com/bidyutchanda/top-10-highest-grossing-films-19752018>) . Please look at the column specifications to identify the variables and the related information. It is worth noting that the column 'worldwide_gross' is the revenue produced by each movie in a particular year and its value is in US \$. Before you read the questions understand the dataset very carefully after importing it in R-Studio.

1. With reference to the given dataset, you are invited to answer the following questions.

Please use R-Studio for analysis. Make use of the concepts already taught and demonstrated in the previous lectures and the workshops.

Step 1: Load the library

Step 2: Import the dataset in R-Studio

Step 3: Which studio has earned the highest worldwide gross over these years and what is this value? Please give the codes and screenshot in your answers.

Step 4: Visualise the sales of studio "Warner Bros" over years. Which year has been the best year? Provide codes, screenshot of the visualisation.

Note: Copy the codes used (OR Screenshots crop only to show codes), visualisations etc. on a word doc. Your answer sheet has to be a pdf file. Please write the question numbers clearly before answering.

2. Objective type questions. Choose only one answer out of the given four.

Q 2.1. Which R-Package is used for cleaning, reshaping, sorting and filtering data?

- a. Ggplot
- b. Dplyr
- c. Stringr
- d. readxl

Q 2.2. 'mutate' function is used to

- a. visualize a particular column
- b. break the existing dataframe

- c. create/add a new row or column to a dataframe which may contain some new values but related to the existing values.
- d. Join two dataframes

Q 2.3. To draw a scatter plot, we may require using:

- a. `geom_point`
- b. `geom_bar`
- c. `geom_line`
- d. `geom_blank`

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs