

Response Summary:

1. Student Information *

First Name	Ema
Last Name	Westerfeld
Major	Animation
Course (e.g. CGT 270-001)	CGT 270-02
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

ewesterf@purdue.edu

3. Visualization Assignment *

- Lab Assignment

Generate

4. Identify appropriate data sources: is the data publicly available? What search methods were used? *

Data source 1	Tableau Sample Data, Secondary Source
Data source 2	Informationisbeautiful.net, Secondary Source
Data source 3	Publicly available statistics for each of the films, Primary Source

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *

The data has been put into a spreadsheet format, structured by the titled columns. There is a combination of text and numbers depending on the category. Multiple sources were used to find the data, making the dataset a secondary source.

6. Data types: what types of data are in the data? How are they stored? What is the access to the data (API, JSON, txt, csv, etc.)? What structure holds the data (data base, spreadsheet, etc.)? *

The data set includes the title of the film, its genre, the lead studio, audience score, profitability, its Rotten Tomatoes score, its worldwide gross, and the year it was released. The file format of the spreadsheet is csv.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

The data's variables are genre, studio, audience score, profitability, Rotten Tomatoes score, gross, and year. The parameters would be the profitability for movies released in 2007-2012. The independent variables are genre, studio, and year. The dependent variables are audience score, profitability, Rotten Tomatoes score, and gross.

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

I assume that this data is accurate for the time it was compiled, as some pieces, such as Rotten Tomatoes or audience scores can change over time. I assume that the genres are generalized, as many films do not fit cleanly into one genre. My audience would be people who are interested in what films were the most successful during that time period, such as film executives and producers.

Generate

9. What real life behavior does the data reflect? Does it show patterns of activity, regularity of events, a timeline, population data, etc? Explain. *

The data shows how the different variables, such as which studio is producing a film, affect its ability to make money.

11. What are the weaknesses of the data source? Is it likely that the source will be available in the future? Is the data complete? What is the quality of the data? Is it specific to your needs for the current project? Is the data in the format you need? Are there missing data? Explain. *

There are holes in the data for certain categories, which could make it difficult to get a full picture for those films. It does not always specify units, which can be confusing. Because some of the data has changed since this set was collected, I cannot check the accuracy of the data on my own.

12. What information is emphasized? What is the central focus of the data? Explain. *

The focus of the data is profitability and gross, as this data would likely be used to determine what films make the most amount of money.

13. At what level of granularity is the data provided? Is the data summarized, or do you have access to the raw data? Is the data categorized or is the data in a format that allows you to create your own categories, etc. Explain. *

The data is summarized and does not give each individual source. It has already been categorized into columns, sorted alphabetically by the title of each film.

14. What is the scope of the data? What topics can be covered using the data? Is there a time range/frame? Is the data for a specific area/discipline/demographic etc.? Explain. *

This data could be used to look at the quality of the movies releasing at this time, because both the Rotten Tomatoes and audience scores are available. The data is specific to 2007 - 2012 and would be most helpful to the film industry.
