Response Summary:

1. Student Information *

First Name	Caroline
Last Name	Dixon
Major	Game Development
Course (e.g. CGT 270-001)	CGT 270 - 001
Term (e.g. F2019)	S2022

2. Email Address *
(University Email Address is required.)

dixon128@purdue.edu

3. Visualization Assignment *
- Lab Assignment

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *

Show ID (int): identifier for each Netflix title, 6-8 digits each

Title (str): the legal name of each title on Netflix

Description (str): contents of the user interface description for each Netflix title

Rating: maturity rating for each title on Netflix

Duration Minutes(int): movies only, duration of the full length of the movie

Duration Seasons(int): TV shows only, duration of the show by groups of episodes

Rating(str): maturity rating for each title on Netflix (eg. Rated NC-17, R, etc)

Date Added(Date): exact date each title was first uploaded to netflix, not counting updates or additional seasons.

Release Year(int): year each title was released on television/cinema, not related to Netflix at all

Director (str): Director of each title on netflix, only one/two per title but many are listed multiple times

Cast(str): Cast of each title on netflix, each title has always more than one cast member and some actors are listed in this set multiple times as they star in different titles

Listed In(str): lists genre categories of each title on netflix

Country(str): lists which countries each title is available in

Type(str): lists whether each entry is a movie or tv show

5. Categorize: consider what is similar and what is different? Categorize the data. Are the variables categorical (normal, ordinal, or rank). Are they quantitative (discrete or continuous)? Show categories. Explain. *

Identifiers: Show ID (int) is the one common variable across all the different workbooks in this dataset. It helps connect all of the different pieces of media to eachother in a succinct way.

Description: (str) This also gets its own category considering how this data set is completely for user interface and not measurable data whatsoever. There are too many inconsistencies across the descriptions for different titles for there to be much to measure or display.

Names: Director, Cast (str). These are all real people that participated in the creation of each piece of media on Netflix.

Categories: Listed In, Country, Type, Rating (str). These are qualitative data points that give more information about what the content of each title is, as well as who its audience is.

Quantitative Info.: Duration Minutes, Duration Seasons, Date Added, Release Year (int). Same as above but all of these are number variables, and completely objective statistics within the metadata of Netflix.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *

The data is not streaming data as it is not continuously updated every second by robots, but rather is manually mined data that is updated about once a month when Netflix adds more media to their website. There isn't really a need to have streaming data when the data itself isn't continuously updated either.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

This data set is very large considering how many individual titles Netflix has in their catelogue and to that extent how many actors in each title listed in the metadata, and how many categories each title fits into. The data is relatively evenly spread considering each title has to have at least one entry in most variables, but the casting can fluctuate wildly.

Evaluate

8. Questions and Assumptions: list at least 3 questions you plan to answer with the data or list the questions if they were provided. Must be complete sentences and end in a question mark. What assumptions are you making? *

Question 1	How many titles have each director participated in the creation of on Netflix?
Question 2	What is the average length of a movie on Netflix for each movie category?
Question 3	Do the countries that each title is available in affect the rating of that title?
Assumptions	My assumption is that one cannot infer much quantitative data from this dataset and thus must mainly rely on asking questions about the qualitative data instead.