Response Summary:

Mine Worksheet

Goal: to identify patterns, extreme and subtle features about the data
Objectives: Students will identify basic descriptors for the data, and categorize the data according to the specifications from the Parse Worksheet
Outcomes: Three (3) specific questions to be answered using the data

1. Student Information *

First Name	Nathan
Last Name	Rusk
Course *(e.g. CGT 270-001)*	CGT27000-09
Term *(e.g. F2019)*	F2021

2. Email Address *
nrusk@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 *
disney_plus_shows.csv
imdb_id - String length, title - String length, plot - String length, type - String length, rated - String length, year - Average, max, min, released_at - String length, added_at - String length, runtime - String length, genre - String length, director - String length, writer - String length, actors - String length, language - String length, country - String length, awards - String length, metascore - Average, max, min, imdb_rating - Average, max, min, imdb_votes - Average, max, min.

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. *

disney_plus_shows.csv
imdb_id - Textual, title - Textual, plot - nominal, type - nominal, rated - interval, year - nominal, released_at - interval, added_at - interval, runtime - interval, genre - nominal, director - Textual, writer - Textual, actors - Textual, language - nominal, country - nominal, awards - nominal, metascore - interval, imdb_rating - interval, imdb_votes - interval.

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 like a snapshot of what is currently on the streaming platform at the time of creating the data set. This particular set of data does not get updated, added, or removed from. It is stored all at one time, the time it is compiled into the excel sheet.

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

This is not a small data set, as it has over 50 rows. The data in this set is pretty sparse it some areas, as it has quite a few missing or N/A. It is fairly dense as it has quite a few variables per TV show or movie. The ranges, such as the runtime of the movies and TV shows, vary quite a bit since this data set has runtime for TV shows and movies rather than just for the movies.

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	What is the average length of the movies/shows?
Question 2	What is the most common rating?
Question 3	What is the most common release year?
Assumptions	-The average release year will not be the current year -The average runtime will not be very low -The most common rating will likely not be that low