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 *

- Training Data

Understand

4. Parse Data: List each field and its data type. Refer to Fry (page 8-9, 2007) for examples of description of different data types (string, float, character, integer), you can also create user defined types (some combination that uniquely identifies data like the Index type in the Fry 2007 page 9 example) *

Date Added (date), Description (string), Duration Minutes (int), Duration Seasons (int), Rating (int), Release Year (int), Show ID (int), Title (string), Type (string), Count of Netflix Titles (int).
Number of variables/columns: 9
Number of records/rows (i.e. the total number of Netflix titles catalogued in this data set): 6,236

5. Assumptions: List any assumptions you are making about the data and/or the visualization challenge (aka the project) *

The assumption I'm making is that we're going to use the data to construct a narrative out of it, probably in regards to what kinds of media on Netflix are more popular as they get added to the database more frequently. I'm anticipating that analyzing the data further will reveal more details about what shows Netflix chooses to add to their platform, at the very least.

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

Understand

4. Parse Data: List each field and its data type. Refer to Fry (page 8-9, 2007) for examples of description of different data types (string, float, character, integer), you can also create user defined types (some combination that uniquely identifies data like the Index type in the Fry 2007 page 9 example) *
Cast (string), Show ID (int)
Number of columns/variables: 2
Number of rows/records (ie the number of references to an actor in the cast of a title in Netflix data): 44,311

5. Assumptions: List any assumptions you are making about the data and/or the visualization challenge (aka the project) *
For this portion of the data, the assumption I'm making is that this dataset is meant to be combined with the netflix_titles dataset, since the Show ID lines up - in other words, my assumption is that this particular set of data in this file is not useful on its own. A mixed list of names with numbers attached doesn't mean much on its own, as opposed to the previous dataset that can be analyzed on its own given the wealth of variables it has.

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

Understand

4. Parse Data: List each field and its data type. Refer to Fry (page 8-9, 2007) for examples of description of different data types (string, float, character, integer), you can also create user defined types (some combination that uniquely identifies data like the Index type in the Fry 2007 page 9 example) *

Listed In (string), Show ID (int)
Number of columns/variables: 2
Number of rows/records(ie number of entries in each Netflix category): 13,670

5. Assumptions: List any assumptions you are making about the data and/or the visualization challenge (aka the project) *

The assumptions I'm making about this data is that the reason there's more records than Netflix titles in this dataset is that each title can be listed under multiple categories, so there are going to be significantly more entries in each genre category of Netflix than there are titles. I'm also making the assumption that this dataset cannot be analyzed on its own either, since there aren't that many variables and the Show ID seems to be a connecting thread between this dataset and the other two.