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	Thomas
Last Name	Cluff
Course *(e.g. CGT 270-001)*	CGT 270-009
Term *(e.g. F2019)*	F2021

2. Email Address *
tcluff@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 *
ID is an integer that acts as a primary key to the data.
Pokemon.Name is a string that gives the name of the Pokémon
Gen.2.Steps, Gen.3.Steps, Gen.4.Steps, Gen.5.6.Steps, Gen.7.Steps are all integers that denote how many steps need to be taken to hatch that Pokémon depending on the generation of game.

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. *
ID is technically a nominal continuous integer because it serves only as a primary key. mathematic functions do not help visualize the data. This could be replaced this a random string of characters, it just has to be unique.
Pokemon.Name is a nominal string whose length can be calculated.
Gen.2.Steps, Gen.3.Steps, Gen.4.Steps, Gen.5.6.Steps, Gen.7.Steps are all continuous ratio integers. They have a meaningful zero at no steps needed to hatch the egg. As the number increases, it takes longer to hatch the egg. It is very comparable to age in this case.

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

Gen.2.Steps, Gen.3.Steps, Gen.4.Steps, Gen.5.6.Steps, and Gen.7.Steps are all technically temporal data as the generation is defines when the game what released. So generation 2 was released before 3, 3 before 4, etc. While this doesn't mean anything when comparing the direct numbers (as release dates are not consistent) it can show how the game designer's design process and balancing technique changed as they made more games. So the time between games is not important but the order is (as if the columns are ordinal data).

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

Obviously the name and ID of each Pokémon is uniformly distributed. For the egg hatching steps, data started sparse and becomes dense as you move into late generations. This is because the newest Pokémon did not have data for the earlier generations because they did not exist.

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	Which generation has the smallest distribution of egg hatching steps?
Question 2	If you were to scale a newer Pokémon's egg steps relative to its generational average, how many steps would it take in an older generation?
Question 3	Is there a strong correlation between egg steps and the Pokémon elemental type from the other datasets?
Assumptions	I am assuming that the Pokémon ID's in this dataset align with the ID's from my other two datasets when comparing data between the two.