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	Shreya
Last Name	Vasant
Course *(e.g. CGT 270-001)*	CGT 27000-LC4
Term *(e.g. F2019)*	F2021

2. Email Address *
svasant@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 *
TopBabyNamebyState.csv
For the string variables such as state, gender, and top name you can find the length of the strings. For the integer variables year and occurrence you can find the range and the median. For the variable occurrence specifically you can find the maximum value (the highest number of occurrence which shows which name was repeating the most), the minimum values ((the lowest number of occurrence which shows which name was repeating the most), the range (represents the diversity in names), the mode (would be the most repeating name-same as maximum- because that would be the greatest occurring name). With the frequency found that would be the most commonly occurring top name (could be a range of names since they could all have the same occurrence). The median would be the name that was most used in the range of years in the middle state of the dataset. You can find an average with occurrence but it does little to nothing to tell you about the data.

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. *

The variables for state, gender, and top name are categorical variables that are nominal for state and gender and ordinal for top name. The variables year and occurrence would be quantitative and would be continuous for year and discrete for occurrence.

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 collected over several years.

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

The data is spread out but dense at certain points.

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 name is most likely to be top name?
Question 2	Which year had the most occurring name?
Question 3	What state had the highest frequency of a top name?
Assumptions	I assumed that occurrence represents the number of newborns in the referenced year that were born to the state.