

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 *

StateName.csv

Name: length

Year: range (span of years)

Gender: length

State:length

Count: maximum value (most common name), minimum value (least common name), mode (highest count would be top name), median (the name that is common -not most popular or least common), average (the name that is common -not most popular or least common), frequency (would be related between count and name)

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. *

Categorical: state (nominal), name (ordinal- highest count name would have a hierarchy), gender (nominal)

Quantitative: year(continuous), count(discrete)

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. Could possibly be skewed when getting to the most common names.

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	Is the data skewed?
Question 2	Which name is the top name in a given year?
Question 3	Which name occurs the most often throughout the range of years?
Assumptions	I assume that count is the number of babies born in that state that had the name in reference for the specified year.