**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*** | Cristina |
| ***Last Name*** | Pascua |
| ***Course*** *(e.g. CGT 270-001)* | CGT 270-009 |
| ***Term*** *(e.g. F2019)* | F2021 |

**2. Email Address** *
cpascua@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** *
Year (Integer): mining procedure: median
Gender (String): mining procedure:
Rank (Integer): mining procedure: range maybe
Name (String): mining procedure: length, mode
Count (Integer): mining procedure: average(2,361), minimum(762), maximum(9,444)

**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.** *
Textual: Name
Nominal: Gender
Interval: Year (discrete)
Ratio: Count(quantitative, discrete), Rank? (categorical)

**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 temporal because it only contains data within a set time range (1960-2019). It is stored over several years by each year.

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

The dataset contains 3003 rows of data which is large but not as big as the previous dataset. The average is for count 2,361 and the range is 762-9,444. This shows that the data is not very evenly distributed. There are more names with low counts than names with high counts.

# 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*** | The data records the top 25 names in each year (for each gender). Is there a big difference in count between the rank 1 name and the rank 25 name (within a year)? |
| ***Question 2*** | Are there any similarities between the top 25 names within a year (i.e. length, occurrence of a character, first initial)? |
| ***Question 3*** | Throughout the years, do some names show up continuously but move up or down the rankings? |
| ***Assumptions*** | I assume within a time period, names appear regularly but go down or up gradually as time goes on, |