

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 *

- Training Data

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *

Gender (String): mining procedure: mode(?)

Year (Integer): mining procedure: median

Top Name (String): mining procedure: length, mode

Occurrences (Integer): mining procedure: maximum(8,265), minimum(214), average(4,248).

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: Top Name

Nominal: Gender

Interval: Year (discrete)

Ratio: Occurences (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 temporal because it stores data from 1910 to 2012 by year.

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

The data set is large but not extremely large as it has 205 rows of data. The data seems to be evenly distributed since the average is 4,248 and the range is 214-8,265.

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 names appear most often as the top name throughout the years?
Question 2	How have occurrences increased with the years?
Question 3	What name has the longest continuous period of being the top name?
Assumptions	I assume some names hold the top spot for a couple of years (top names last for a few years).