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

2. Email Address *
myers436@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 *
Row number: Integer - Mode, Median
Animal Name: String - Mode
Animal Gender: Boolean - Mode
Animal birth Month: Integer - Mode, Median
Breed Name: String - Mode
Borough: Boolean - Mode
Zip Code: Integer - Mode, Median Average
Community: Integer - Mode, Median Average
CensusTract2010: Integer - Mode, Median
NTA: Alphanumeric - Mode, Median
City Consul District: Integer - Mode
Congressional District: Integer - Mode
State Senatorial District: Integer - Mode
License Issue Data: Integer - Mode, Median Average
License Expired Date: Integer - Mode, Median Average

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. *

Nominal: Animal Name, Animal Gender, Breed Name, Borough, City Consul District, Congressional district, State Senatorial District.
Ordinal: Animal Birth Month
Interval: Row Number, CensusTract2010, NTA,
Ratio: Zipcode, Community, License Issue Date, License Expired Date

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

Yes, this data was only updated in 2015-2016. The license issue dates are years old and are already expired.

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

Very large data set with 121,951 rows! The data is dense showing many different categories for dog likeness's.

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	When do most of the license's expire?
Question 2	Are most licenses for boy or girl dogs?
Question 3	Which breed is most popular in NYC?
Assumptions	It appears that most of the license's are already expired. The data has not been updated since it was first made and that was years ago.