Response Summary:

1. Student Information *

First Name	Quenton
Last Name	Hostetter
Major	Web Programming and Design
Course ***(e.g. CGT 270-001)***	CGT 270-001
Term ***(e.g. F2019)***	S2022

2. Email Address *
(University Email Address is required.)
qhostett@purdue.edu

3. Visualization Assignment *

- Training Data

Generate

4. Identify appropriate data sources: is the data publicly available? What search methods were used? *

Data source 1	Cat vs Dog Popularity in the US Tableau Data Set
Data source 2	Government survey of pet data recorded for the purpose of legislating on pet ownership. Public survey
Data source 3	Local Humane society data on pet owners based on adoption records. Probably private records

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *
The data is structured, and numerical in nature for the most part. The labels are strings, but everything else is float values. The data was recorded via survey and is still raw and unfiltered, therefore it is still primary data.

6. Data types: what types of data are in the data? How are they stored? What is the access to the data (API, JSON, txt, csv, etc.)? What structure holds the data (data base, spreadsheet, etc.)? *
The data is contained in a spreadsheet of file type .xlss and is organized as floats and strings.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

Location: string -> independent var
Number of Households (in 1000): int -> dependent var
Percentage of households with pets: float -> dependent var
Number of Pet Households (in 1000): int -> dependent var
Percentage of Dog Owners: float -> dependent var
Dog Owning Households (1000s): int -> dependent var
Mean Number of Dogs per household: float -> dependent var
Dog Population (in 1000): int -> dependent var
Percentage of Cat Owners: float -> dependent var
Cat Owning Households: int -> dependent var
Mean Number of Cats: float -> dependent var
Cat Population: int -> dependent var

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

The audience is people looking into US pet population data, and I am assuming that the data was recorded impartially and using accurate survey methods.

Generate

9. What real life behavior does the data reflect? Does it show patterns of activity, regularity of events, a timeline, population data, etc? Explain. *

This data reflects population data of dogs and cats in the US.

11. What are the weaknesses of the data source? Is it likely that the source will be available in the future? Is the data complete? What is the quality of the data? Is it specific to your needs for. the current project? Is the data in the format you need? Are there missing data? Explain. *

The dataset is weakened by the fact that it missed two states of the US.

12. What information is emphasized? What is the central focus of the data? Explain. *

The central focus is the dicotomy of pet households that contain dogs or cats.

13. At what level of granularity is the data provided? Is the data summarized, or do you have access to the raw data? Is the data categorized or is the data in a format that allows you to create your own categories, etc. Explain. *

The data is categorized, but it does seem to summarize slightly as it rounds its numbers to the 1,000s and as such doesn't give the raw numbers on households and population.

14. What is the scope of the data? What topics can be covered using the data? Is there a time range/frame? Is the data for a specific area/discipline/demographic etc.? Explain. *

The data contained is very region specific, namely it is specific to the united states excluding Alaska and Hawaii.