Response Summary:

1. Student Information *

First Name	David
Last Name	Luo
Major	Data Science
Course *(e.g. CGT 270-001)*	CGT 270-003
Term *(e.g. F2019)*	S2022

2. Email Address *
(University Email Address is required.)
luo354@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: range, 1895-2016
Punxsutawney Phil: categories: No Record, Full Shadow, No Shadow, Partial Shadow
February Average Temperature: Mean=33.8, std=3.28
February Average Temperature (Northeast): mean=22.69, std=4.28
February Average Temperature (Midwest): mean=32.69, std=4.86
February Average Temperature (Pennsylvania): mean=26.52, std=4.53
March Average Temperature: mean=41.70, std=2.95
March Average Temperature (Northeast): mean=32.36, std=3.88
March Average Temperature (Midwest): mean=42.56, std=4.51
March Average Temperature (Pennsylvania):mean=35.9, std= 4.20

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. *
year: Discrete
Punxsutawney Phil: Ordinal
all else: continuous

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 not streaming data. The values range from 1895 to 2016. So far no updates have been made.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *
The spread of data for temperature is fairly evenly spread.
The spread for Punxsutawney Phil is uneven, with 100 records of Full Shadow, 15 No Shadow, 6 missing, and 1 Partial Shadow.

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	Does a higher February temperature correlate with a higher March Temperature?
Question 2	How does shadow correlate with change in temperature?
Question 3	How accurate has Phil been in his predictions?
Assumptions	No further assumptions made other than the accuracy of the data.