Response Summary:

1. Student Information *

First Name	Grace
Last Name	Combs
Major	Web Development
Course ***(e.g. CGT 270-001)***	CGT 270-003
Term ***(e.g. F2019)***	SP2022

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

3. Visualization Assignment *
- Lab Assignment

Generate

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

Data source 1	PunxsutawneyPhil.csv The data was given through an assignment, most likely publicly available given the context.
Data source 2	N/A
Data source 3	N/A

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 is a mix of text and numbers. This data appears to be secondary data to me because it is over a wide span of years and was probably recompiled by using outside sources for temperature across the country.

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.)? *
There is temperatures, dates, and descriptions of shadows. They are stored in columns and rows in an excel sheet in .csv format.

Evaluate

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

Year: independent. 1886-2016
Punxsutawney Phil: independent. No shadow, No record, Partial shadow, Full shadow
February Average Temperature: dependent on year
February Average Temperature (Northeast): dependent on year
February Average Temperature (Midwest): dependent on year
February Average Temperature (Pennsylvania): dependent on year
March Average Temperature: dependent on year
March Average Temperature (Northeast): dependent on year
March Average Temperature (Midwest): dependent on year
March Average Temperature (Pennsylvania): dependent on year

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

I think the audience is the general public who likes to know what the result is on groundhogs day. I assume there is no relationship between temperature and the result of the shadow.

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. *

The data does show a timeline or history of events dating back to the 1800s.

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 data is incomplete because it is added onto each year. The quality is marginal because a lot of the results are "no record" or there is no temperature listed. It is likely the data will be available in the future.

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

The emphasized information is the result of if there is a shadow or not in a given year, and then what the temperatures were. I think the result of the shadow is the most centrally focused data.

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 and may be summarized or raw. It is unknown what is exactly recorded on the days that the groundhog shadow is recorded, but if the temperature and year are the only things recorded then it is raw.

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. *

There is a time range and season range. Only the topic of groundhog's day and the shadow can be covered really, but one could use the temperature aspect of the data for other purposes.