Response Summary:

1. Student Information *

First Name	Connor
Last Name	Colbert
Major	Game Development and Design
Course ***(e.g. CGT 270-001)***	CGT 270
Term ***(e.g. F2019)***	SP2022

2. Email Address *
(University Email Address is required.)
colberj@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	National Oceanic and Atmospheric Administration (NOAA) - Found monthly and regional temperature data going back to 1985.
Data source 2	York Daily Record - Has record of all of Punxsutawney Phil's data on if he saw his shadow since 1886.
Data source 3	United States of America: War Department (Monthly Weather Review) - Data on temperature and weather kept via the War Department in the late 1800s.

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 in column and row format. It is a combination of text and numerical values from multiple sources. This is secondary 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 types of data present are Strings, Integers, Enums (No Shaodow, Full Shadow, etc.), and doubles. The data is in a .csv file format and accessible in a spreadsheet.

Evaluate

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

All data variables in this dataset are independent. They include:
Year,
Punxsutawney Phil (what Phil saw),
February Average Temperature,
February Average Temperature (Northeast),
February Average Temperature (Midwest),
February Average Temperature (Pennsylvania),
March Average Temperature ,
March Average Temperature (Northeast),
March Average Temperature (Midwest),
March Average Temperature (Pennsylvania).

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

For this data, I can make the assumptions that the data going back to the late 1800s to early 1900s are not as accurate as the data from the 21st century. The audience for this data consists of the citizens of the United States, especially those who live in Pennsylvania.

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 shows the trends in weather over the years and the correlation between it and Punxsutawney Phil's predictions.

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 largest weakness is that there are the many data entries without records. The quality of the data is acceptable; however, the phrasing for many of the variables leads to some confusion. The source will be available in the future and the data is not misrepresentative.

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

The central focus of the dataset is the predictions made by Punxsutawney Phil and the weather of the subsequent months.

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 somewhat summarized as the temperature is averaged across regions as opposed to individual states (except for Pennsylvania)-- the numerical data itself is raw, however.

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 scope for this data is quite large as it goes back to 1895, more than a century ago. The data is largely intended for states in the Eastern regions of the United States (i.e., states that have significant, measurable winters every year).