Response Summary:

1. Student Information *

First Name	Rae
Last Name	Fu
Major	Animation and Visual Effects
Course ***(e.g. CGT 270-001)***	CGT 270-004
Term ***(e.g. F2019)***	S2022

2. Email Address *
(University Email Address is required.)
fu345@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	The data is publicly available. It was found by searching online. http://www.stormfax.com/ghogday.htm
Data source 2	https://www.foxnews.com/lifestyle/punxsutawney-phil-predictions-5-decades
Data source 3	https://web.archive.org/web/20120513160448/http://www.maa.org/pubs/cmj_1_01.pdf

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. There is a combination of text, integers, and decimals. It is secondary data, because it compiles the data from each year.

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 categorical qualitative information about whether a shadow is seen or not with years as discrete measures and the temperature as continuous data. They are stored as text and numeric. The access to the data is csv. The structure of the data is a spreadsheet.

Evaluate

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

The data variables are year, Punxsutawney Phil, 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), and March Average Temperature (Pennsylvania). Dependent variables are the temperatures and Punxsutawney Phil. The year is the independent variable.

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

Assumptions about the data is that there might not be a relationship between having a shadow and the temperature. The audience is people who want to know whether or not the predictions due to the groundhog are accurate.

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 reflects regularity of events because it shows the weather and whether or not there is a shadow.

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. *

It is likely that the source will be available in the future since it is a tradition, so it would be unlikely to be hidden. The data is not complete, because it is ongoing and some years have no records. The quality of the data is relatively good, since it is mostly complete. It is specific to finding our whether there is a relationship between a shadow and the weather. The data is in the format needed, they are text and numbers. There is missing data for some of the years because it says no record.

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

The information emphasized is whether or not there was a shadow and the temperature for that year, since those are the dependent variables.

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 fine, there are many temperatures. The data is raw, and in a format that allows for creation of categories because the temperatures are separate for the same month.

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. *

Topics on the shadow on groundhog day, the temperature according to year can be covered. The time range is 1895 to 2016. The data is for a specific discipline of groundhog day.