Response Summary:

1. Student Information *

First Name	Liam
Last Name	McGuire
Major	CGT 270
Course *(e.g. CGT 270-001)*	CGT 270
Term *(e.g. F2019)*	SP2022

2. Email Address *
(University Email Address is required.)
mcguir53@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 of Punxsutawney Phil is publicly available. You can just search the shadow history and more than a few results with quite a bit of data will show up, so you can just google it. In this case, I was provided an excel sheet of the data.
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 was given in an XLXS format(Excel format). It is structured due to being related to multiple reports and databases, so it is from multiple sources over when the shadow did and didn't show, telling us it is secondary data since it was carried over to other researchers over the last century.

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 access to the data is, again, given in an XLXS(Excel) format. A spreadsheet holds all this data, storing it all with a mix of nominal and ordinal data via the times the shadow appeared during the years and the temperature of the month.

Evaluate

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

Year, Punxsutawney Phil, Average temperature. The parameters are the years themselves and when a shadow appeared during those respective years. I'd say the independent variables is very much average temperature, as doesn't impact the shadow appearing. The years and shadows are dependent with one another since without either of them, the tradition can't happen and the most important part of the data can't be entered in.

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

The audience are the likely the people who don't live in Punxsutawney as they might want to know how often the groundhog's shadow has appeared over the years. The assumptions I have is that these were measured during the day of the tradition.

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. *

It shows a common pattern of full shadows being shown the most often and consistently. Other than that, it doesn't reflect too much.

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 source will of course be available in the future since it's tradition, meaning it will keep happening as a part of Punxsutawney culture. It is complete and lists the relevant data needed. There isn't really any missing data, so I don't see any weaknesses with the source.

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

The times when full shadows have appeared since it helps predict how much longer winter will last and when spring will start, so it will of course be the most important piece of 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 summarized quite well quantitatively as it have me the raw data, allowing me to create my own categories.

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. *

How often the shadows have appeared and how often they have not, listing a time frame of when these events took place. The only specific area of data is the average monthly temperature in certain regions.