## Response Summary:

**1. Student Information** *

| | |
|---|---|
| ***First Name*** | Keegan |
| ***Last Name*** | Palonis |
| ***Major*** | Data Visualization |
| ***Course*** ***(e.g. CGT 270-001)*** | CGT 270-003 |
| ***Term*** ***(e.g. F2019)*** | S2022 |

**2. Email Address** *
**(University Email Address is required.)**
kpalonis@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*** | Punxsutawney Phil, Given information as well as average temperatures in February and March, in different regions |
| ***Data source 2*** | https://www.ncdc.noaa.gov/cag/statewide/time-series/36/tavg/1/1/1895-2022?base_prd=true&begbaseyear=1901&endbaseyear=2000<br>Includes state by state temperature information, and is public information, search was "average temperature by year" |
| ***Data source 3*** | https://www.usclimatedata.com/<br>Has data by state and city from 2007 to 2019, data is publicly available search was "u.s. average temperature by year" |

**5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data?** *
Data is Structured, and is all text, with Data Source 2 having additional graphs. The data seems to be primary.

**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 are both string and float data points, stored in tables. Some are csv and some are text. Spreadsheets are how the data is structured.

# Evaluate

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

Punxsutawney Phil (Either No shadow, Full shadow, partial shadow or no record). Then Average temperature for February and March in the US, in the Northeast, Midwest and Pennsylvania.

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

Audience is the general public, and it is assumed that all available data is included, and that the data is complete. Also, that the temperature is in Fahrenheit.

# 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.** *

Shows a timeline of temperatures in March and February in different regions and locations, and whether the groundhog Punxsutawney Phil saw his shadow or not.

**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.** *

Not all data is available, there are a few years with no records of Punxsutawney Phil, and some years in the 1890s that don't have any temperature information.

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

Focus of the data seems to be whether Punxsutawney Phil saw his shadow and what the average temperature was.

**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.** *

Data is categorized, though we do have access to raw data. We have access to the monthly average and whether Punxsutawney Phil saw his shadow or not.

**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 topics covered are temperature in March and February for the Northeast, Midwest, overall, and Pennsylvania