

**Lab Assignment: Midterm Review**

**This week you will use the Byrd Data-Vis Tool and your self-assessments to review and prepare for the midterm.**

**There are 5 Learning Objectives**

1. Remember and define the stages of visualizing data
2. Explain the interactions between the stages of the data visualization process.
3. Demonstrate what happens in each stage of the data visualization process.
4. Generate/produce data visualization that provide insight.
5. Self and per-critique the data visualization process and outcomes.

**Lab Assignment**

*Exploratory Analysis of Punxsutawney Phil Data.* This exercise is designed to demonstrate your competency in applying the data visualization process. Complete the following worksheets using the Punxsutawney Phil Dataset:

- a. Acquire Activity Worksheet.
- b. Parse Activity Worksheet
- c. Mine Activity Worksheet
- d. Filter & Represent Worksheet
- e. Rate your visualization (rate the visualization you created)
- f. Refine Worksheet (refine the visualization(s) you created).

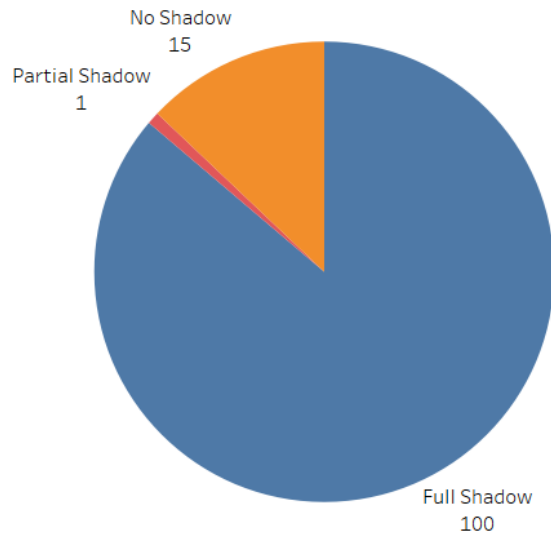
Data Visualization Activity worksheets can be accessed using The Byrd Data-vis Tool. The tool is available for download on the course and lab web pages in Brightspace (See Resources Module). Once in the application look for Worksheets on the left navigation panel. The application is also installed on all computers KNOY 314 and KNOY 306.

**Understanding the Data Visualization Process**

1. In your own words, list AND define, the stages of visualizing data discussed in class and lab.
  - a. Acquire: Gather data, either from your own experiment or study, or use data made publicly available.
  - b. Parse: Make sure the data is organized and comprehensible.
  - c. Filter: Separate the data that you're interested in, and make sure all the data that you need is present. Also make sure that the data you're choosing is complete.
  - d. Mine: Analyze the data you have filtered, look for patterns and relationships, tie the data to the questions or information you are looking for.
  - e. Represent: Make a simple graph appropriate for the questions you've asked.
  - f. Critique: Share the data visualization with others in order to receive feedback on where things could improve, and make sure that you haven't missed anything.
  - g. Refine: Make changes to your visualization, and improve it to be even better, more engaging.
  - h. Interact: Add ways to adjust the data represented, and what data is displayed.
2. Explain the interactions between the stages of the data visualization process.

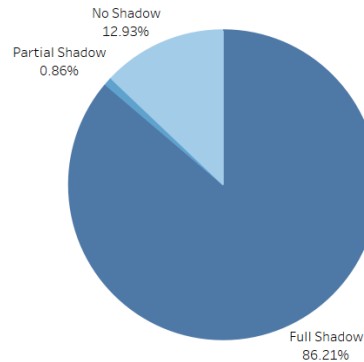
- a. Most interactions are sequential. Acquire -> Parse -> Filter -> Mine -> Represent -> Critique -> Refine -> Interact.
  - b. Sometimes the Critique stage will require you to go back to Acquire stage to gather more data, Filter stage to adjust what data you are including in your visualization, Mine stage to look for more relationships or patterns, or the Represent stage to adjust what type of visualization you have chosen.
  - c. The refine stage can take you back to the critique stage in order to get more feedback on the adjustments you had made to your visualization.
  - d. Finally, the interact stage could send you back to either the critique or refine stage as you finalize the visualization in order to make adjustments.
3. Locate and download the Punxsutawney Phil dataset in Week 7's Lab assignment. You will use this dataset to demonstrate your understanding and competency in each stage of the data visualization process. **Complete the following worksheets** using the Punxsutawney Phil dataset:
- a. Acquire Activity Worksheet (save your work)
  - b. Parse Activity Worksheet (save your work)
  - c. Mine Activity Worksheet (save your work)
    - i. What patterns can/do you detect in the data? **Explain.**
      1. Over both February and March, the overall average temperature has increased in all regions
      2. March is almost always warmer on average than February
      3. The Average Temperature for No Shadow is higher in every category than Full Shadow, and same is true for March.
    - ii. What's Punxsutawney Phil's track record for seeing his shadow? (Hint: Full shadow counts, Partial shadow counts, No Shadow counts). **Explain.**
      1. Punxsutawney Phil sees his Full shadow 102 times, Partial Shadow 1 time, and No shadow 16 times. This means that Punxsutawney Phil sees his full shadow 85.71% of the time.
    - iii. What's Punxsutawney Phil's track record for No Record recorded? Anything interesting about this data? **Explain.**
      1. Most of the times there is No Record is in the beginning of the data set, the last No Record occurs in 1943, and the one before that is in 1902.
  - d. Filter & Represent Worksheet (save your work)
    - i. How many records are there in the dataset?
      1. 131
    - ii. Show your filtered data: No record, Full shadow, Partial Shadow, No entry, and No Shadow. Hint: it might be helpful to save each filtered dataset to a separate tab in the same workbook (If using Microsoft Excel, make sure you name the tabs accordingly).
    - iii. Create a visualization comparing the number of times the data show, on average there was full shadow, partial shadow, and no shadow for Punxsutawney Phil between 1895 and 2016; show the data values.

## Punxsutawney Phil Shadow Record from 1895 - 2016



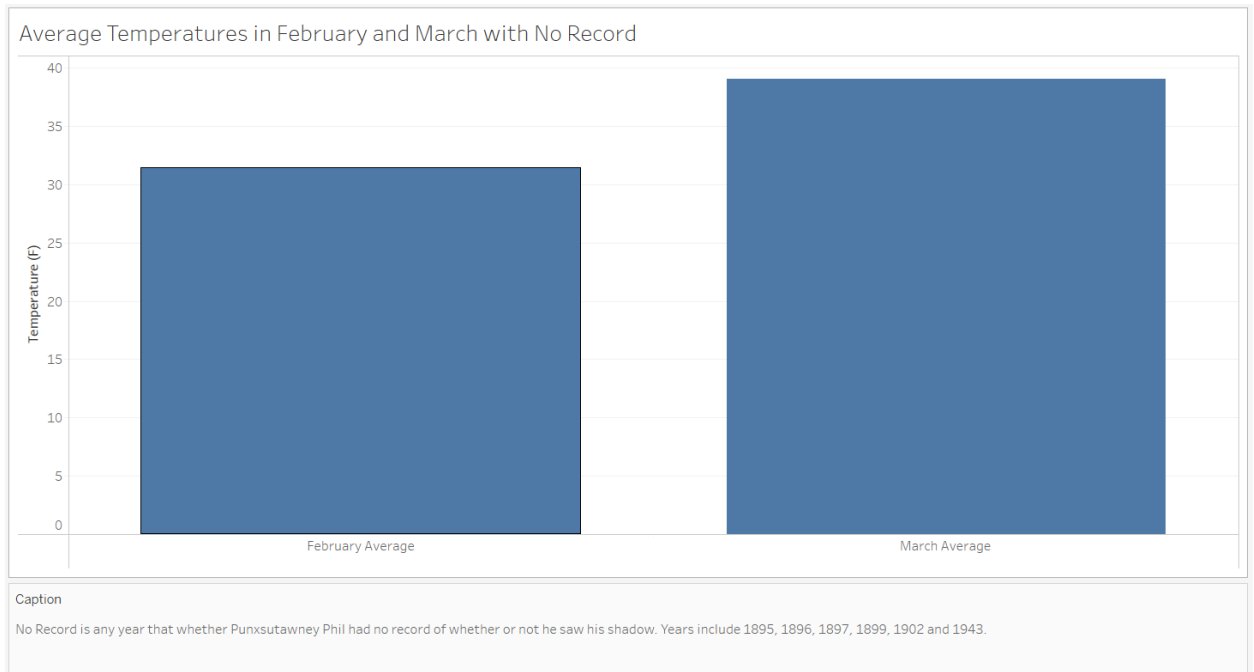
4. Rate your visualization rate the visualization you created step 3d. Use the data visualization checklist to examine your output from Step 4. <https://stephanieevergreen.com/data-visualization-checklist/>
  - a. Refine Worksheet (refine the visualization(s) you created in step 3a and rated in step 3e. Edit the visualization created in 3d (iii) and replace the data values with percentages.

Punxsutawney Phil Shadow Record from 1895 - 2016



You may recall, data visualization is an iterative process. Now that you are familiar with Punxsutawney Phil's track record, examine the temperatures recorded. This will require you to "revisit" several stages in the data visualization process.

5. In Step 3c (iii), you were asked if there was anything interesting about Punxsutawney Phil's track record for "No Record" recorded. Go back and Mine the data again to answer the following questions
  - a. How many records show "No Record" for Punxsutawney Phil? 12
  - b. How many records show "No Record" for Punxsutawney Phil but actually show data points (temperatures) for the associated year? 6
  - c. **Create a visualization** comparing the February Average Temperature to the March Average Temperature for the "No Record" data.
    - Only include records that actually have data.
    - Make sure you use data visualization best practices, and refer to the data visualization checklist. For starters: your graphs should have descriptive title, include the range of dates from which the data is visualized and your axis should be properly labeled.



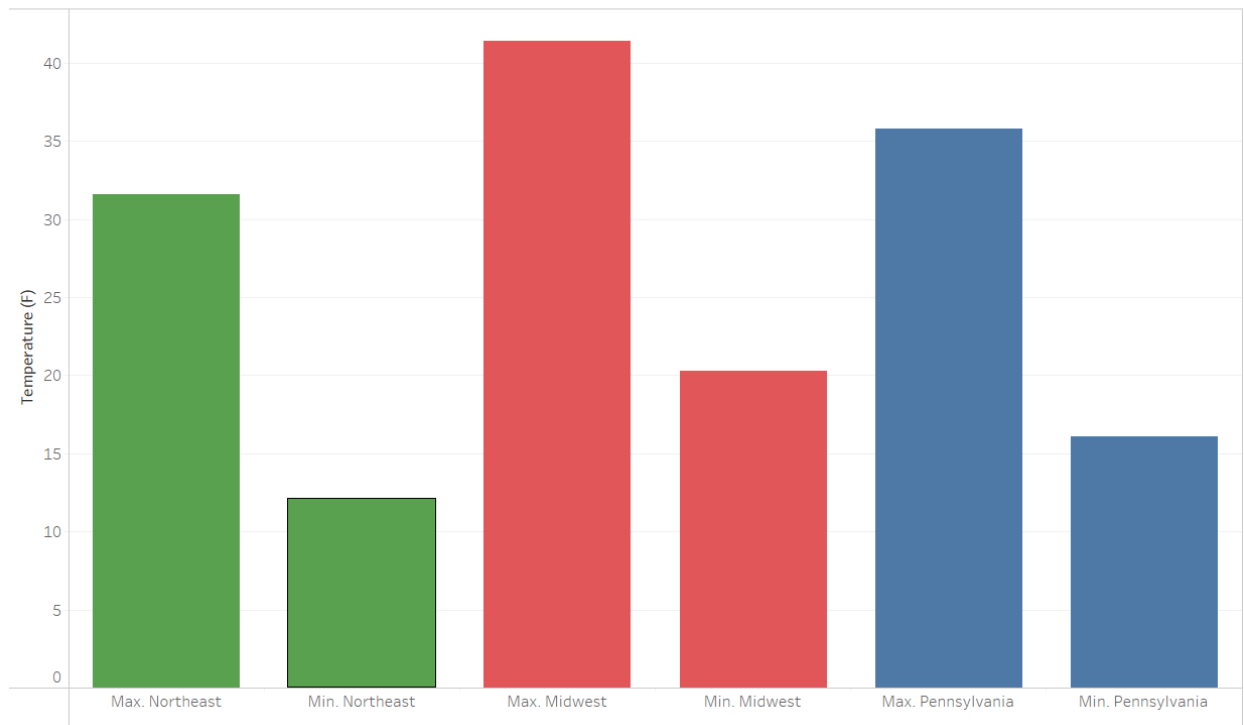
6. Average temperatures are provided for Northeast, Midwest and Pennsylvania. Filter and Mine the data again to answer the following questions for reports of Phil seeing his Full Shadow.

- How many records indicate Punxsutawney Phil saw his Full Shadow? 102
- What is the highest and lowest Average temperature recorded for February?  
High: 41.41; Low: 25.23
- Complete the following table:

Punxsutawney Phil Full Shadow	High	Low
February Average Temperature (Northeast)	31.6	12.1
February Average Temperature (Midwest)	41.4	20.3
February Average Temperature (Pennsylvania)	35.8	16.1

- Create line or bar chart to support your answer to Step 6c. Use data visualization best practices!

Minimum and Maximum Average Temperatures with Full Shadow in February

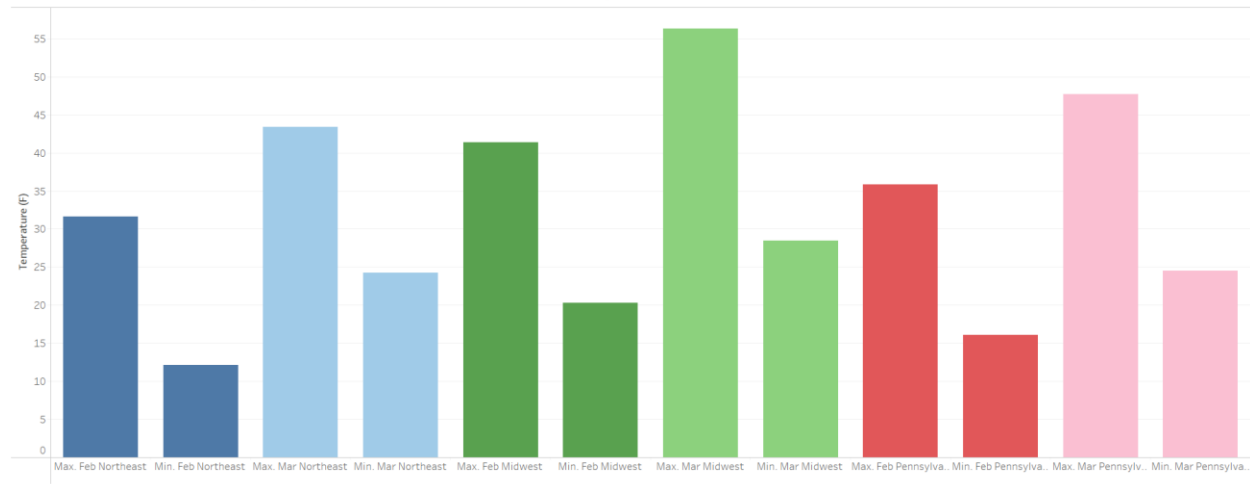


7. Revisit the data to compare High and Low temperatures for February and March across the regions
- Complete the following table showing the average high and low temperatures when Punxsutawney Phil saw his Full Shadow in the Northeast (1898 – 2016)

	February Average Temperature (Northeast)	March Average Temperature (Northeast)
High	31.6	43.4
Low	12.1	24.2

- Create and include visualization(s) that compare the high and low temperatures for February and March across the regions.

Minimum and Maximum Average Temperatures with Full Shadow in February and March



OR Create a line chart.

8. What assumptions did you make about the data?
  - a. Temperature in Fahrenheit, data is complete

#### What to turn in:

- ✓ This document (saved as PDF) showing your understanding of the data visualization process
  - a. Listing and defining each stage of visualizing data, in your own words.
  - b. Explain the interactions between the stages of the data visualization process, in your own words.
  - c. List each stage and the interaction, in your own words.
- ✓ Combine all files:
  - a. this document, with screenshots of visualizations included.
  - b. the data file showing your filtered data (.xlsx)
  - c. one set of Data Visualization Activity worksheets (acquire, parse, mine, filter& represent, refine) – with visualizations included
  - d. Data visualization process self-assessment (yes, complete this again in lab)

into one (1) zip file. Save the zip file as **LastNameFirstInitial\_CGT270\_Lab7.pdf**