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.

   Acquire: gathering the data or retrieving a dataset and then reviewing it. User should make assumptions, see how/if the data is categorized, and identify the audience.
   Parse: looking at what types of data are in set and if there is anything interesting about the data.
   Mine: making calculations to find out certain aspects of the data.
   Filter: Filtering the data to fit your purposes.
   Represent: visually representing the data you have chosen.
   Critique: critiquing your own or getting help from others to see what you need to improve in your visualization.
   Refine: using the feedback from the critique stage, improve the visualization.

2. Explain the interactions between the stages of the data visualization process.

   Parse and Mine are related because you can mine calculations between the same types of data such as integers. Filter and Represent are closely related because you must think

of what you want to represent before you filter your data to exclude unnecessary information. Critique sets up the process for refinement.

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.**

        Temperatures seem to be about the same regardless of shadow result. There seems to be more "no shadow" results in recent years.

        ii. What's Punxsutawney Phil's track record for seeing his shadow? (Hint: Full shadow counts, Partial shadow counts, No Shadow counts). **Explain.**

        No Shadow: 16. Partial Shadow: 1. Full Shadow: 102.

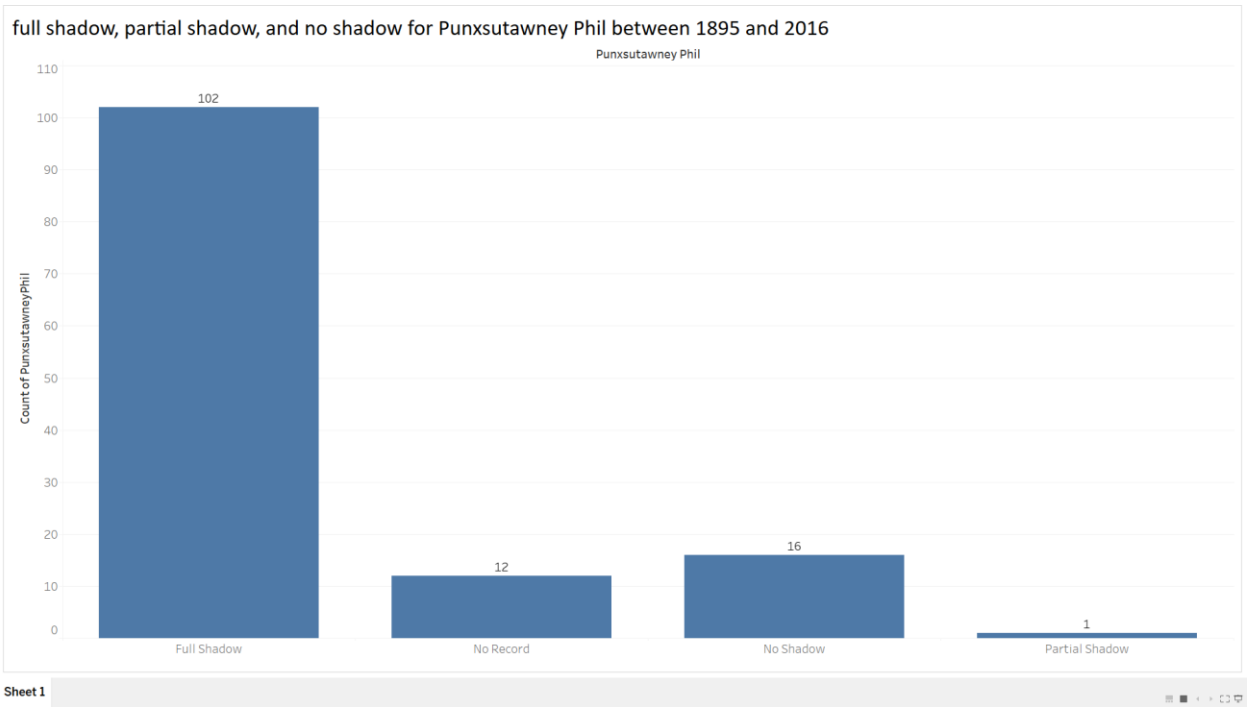
        iii. What's Punxsutawney Phil's track record for No Record recorded? Anything interesting about this data? **Explain.**

        No record: 12. It is interesting that for 12 years there was nothing recorded because this is almost the same amount of times there was no shadow at all. 12 is a substantial number when calculating the difference between how many full shadows, partial shadows, and no shadows there is and could easily change the data.

    d. Filter & Represent Worksheet (save your work)
        i. How many records are there in the dataset? _____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).

        I created new tabs on the excel worksheet titled as above.

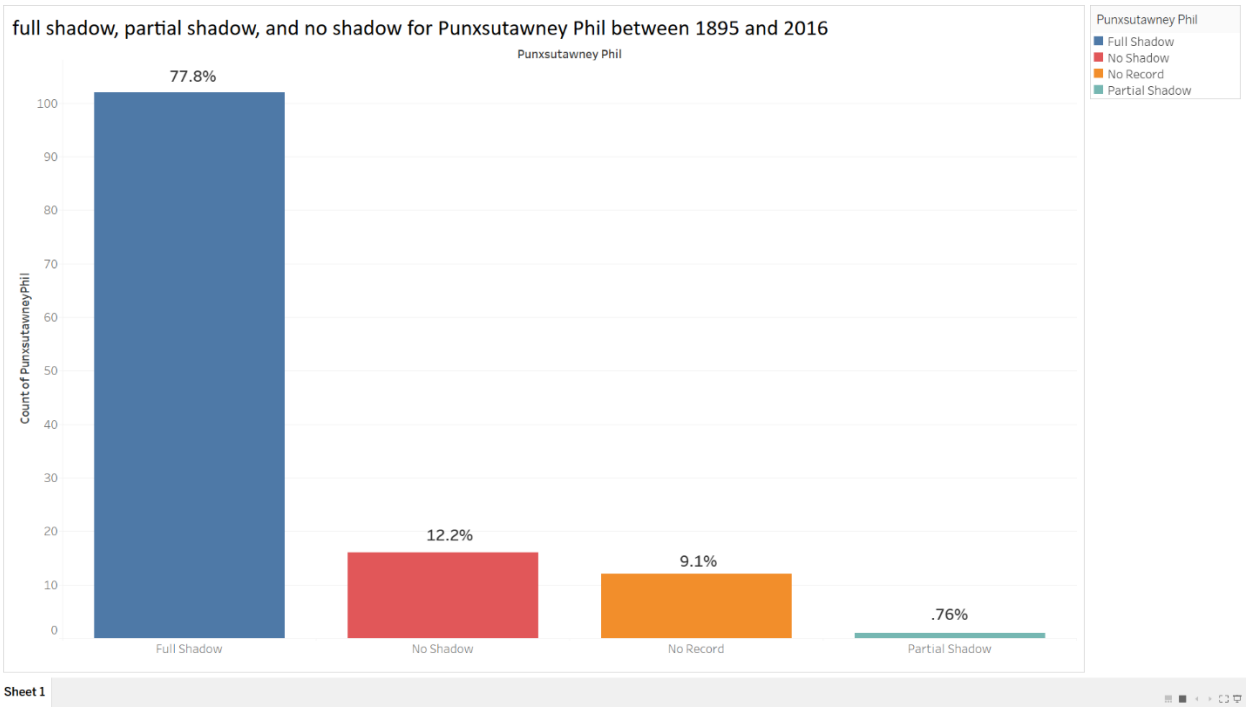
        iii. Create a visualization comparing the number of times the data show, on average there was full shadow, partial shadow, and no shadow for Punxsutawney Phil between 1895 and 2016; show the data values.



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/

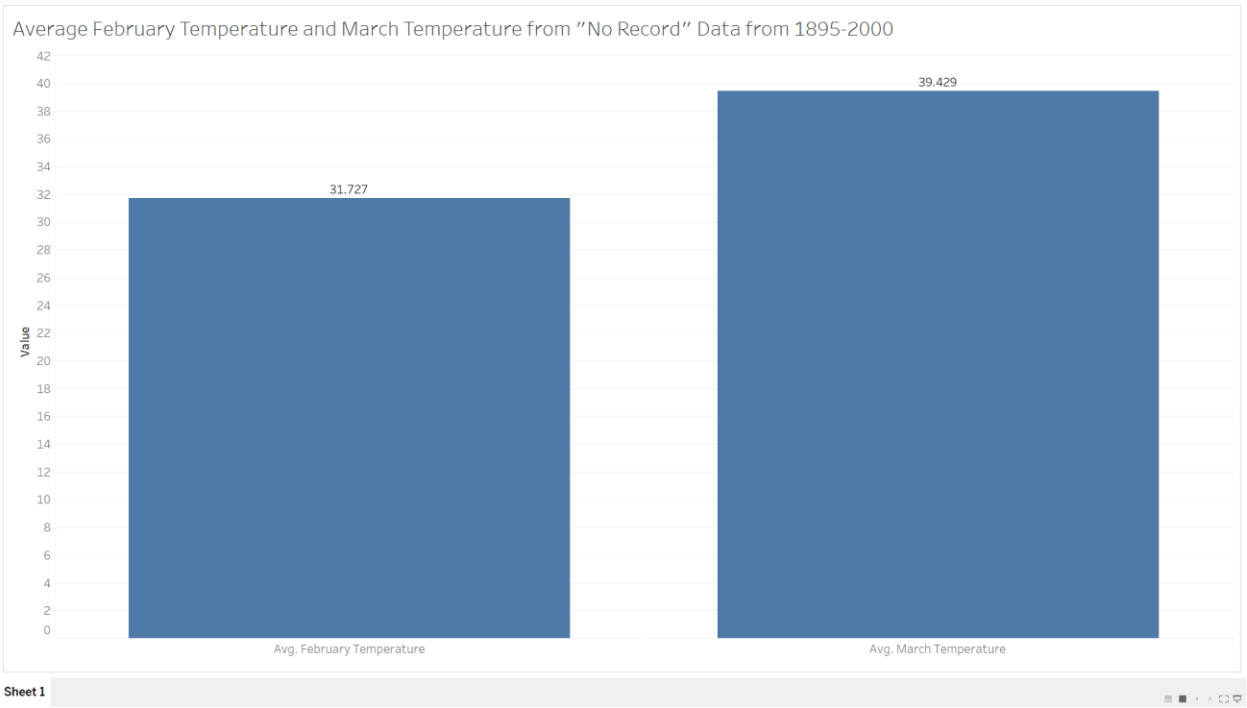
   I received an 80%.

   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.



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.

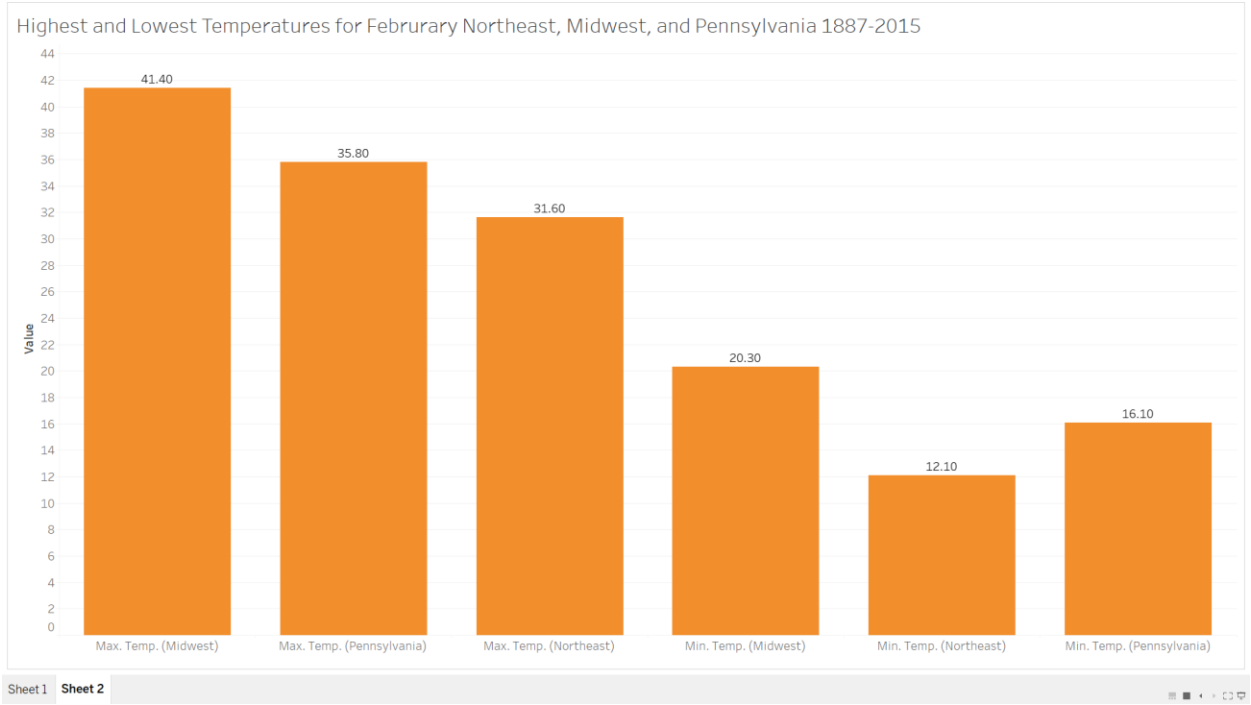
5. In Step 3c (iii), you were asked if there was anything interesting about Punxsutawney Phil's track record for "No Record" recorded. Go back and Mine the data again to answer the following questions
    a. How may records show "No Record" for Punxsutawney Phil? _____13_____
    b. How many records show "No Record" for Punxsutawney Phil but actually show data points (temperatures) for the associated year? _____7____
    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.
   a. How many records indicate Punxsutawney Phil saw his Full Shadow? ____102_____
   b. What is the highest and lowest Average temperature recorded for February?
      High:____41.41____; Low: __25.23_____
   c. 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

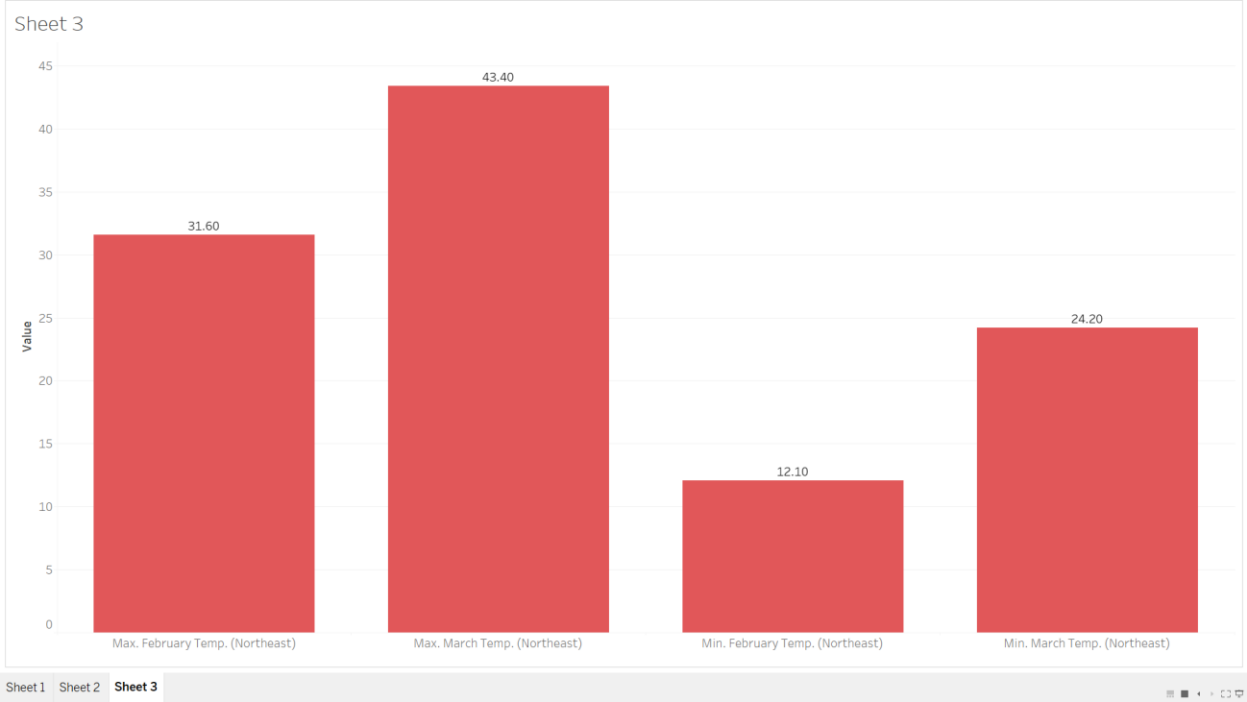
   d. Create line or bar chart to support your answer to Step 6c. Use data visualization best practices!



7. Revisit the data to compare High and Low temperatures for February and March across the regions
    a. 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

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



OR Create a line chart.

8. What assumptions did you make about the data?
    a. I assumed the lowest and highest temperatures for March would be higher than those of February.

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**