Understanding the Data Visualization Process

1. In your own words, list AND define, the stages of visualizing data discussed in class and lab.
   1. **Acquire** – View the data, gather sources, analyze, and make assumptions
   2. **Parse** – Separate out all variables and identify data types
   3. **Mine** – Analyze to determine basic descriptors, range, distribution, etc. and evaluate by making questions about the data
   4. **Filter & Represent** – Remove unnecessary data and create a visual representation of the data
   5. **Refine** – Based on critiques, refine and adjust the visualization.
2. Explain the interactions between the stages of the data visualization process.
   1. Acquire and Parse act as introductions to the data, giving basic information and setting up the ability to take a deeper dive.
   2. Mine is very descriptive and uses the information gathered in the previous steps to help determine its portions. It also provides questions to build the later visualizations upon. This is, arguably, the most essential step in the visualization process.
   3. Filter & Represent lead very nicely from Mine as it provided the sections and details that are now able to be filtered upon. With the newly filtered data, a visualization can be made
   4. Refine allows for iteration and improvements upon the previous visualization made and is the culmination of all the previous steps taken
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:
   1. Acquire Activity Worksheet (save your work)
   2. Parse Activity Worksheet (save your work)
   3. Mine Activity Worksheet (save your work)
      1. What patterns can/do you detect in the data? **Explain.**
         1. In most of the years, the record shows that there is a full shadow. There are never back-to-back years of No Shadows, and there is only 1 partial shadow.
      2. What’s Punxsutawney Phil’s track record for seeing his shadow? (Hint: Full shadow counts, Partial shadow counts, No Shadow counts). **Explain.**
         1. 100 Full Shadows
         2. 15 No Shadows
         3. 1 Partial Shadow
         4. This indicates that Phil has a bias towards seeing his shadow.
      3. What’s Punxsutawney Phil’s track record for No Record recorded? Anything interesting about this data?  **Explain.**
         1. There are 12 instances of No records reported. The main interesting part of this is that there are also no temperatures recorded for the majority of the “No Record” entries.
   4. Filter & Represent Worksheet (save your work)
      1. How many records are there in the dataset?
      2. 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).
      3. 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.

A picture containing rectangle

Description automatically generated

1. 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/>
   1. 20/24 – Does not represent how they relate to each other. Put into pie chart?
   2. 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.

Chart, pie chart

Description automatically generated

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.

1. 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
   1. How may records show “No Record” for Punxsutawney Phil?
      1. 12 Records
   2. How many records show “No Record” for Punxsutawney Phil but actually show data points (temperatures) for the associated year?
      1. 5 Records
   3. **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.

Chart, line chart

Description automatically generated

1. 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.
   1. How many records indicate Punxsutawney Phil saw his Full Shadow?
      1. 100
   2. What is the highest and lowest Average temperature recorded for February?
      1. High: 41.41
      2. Low: 25.23
   3. Complete the following table:

|  |  |  |
| --- | --- | --- |
| **Punxsutawney Phil Full Shadow** | **High** | **Low** |
| February Average Temperature (Northeast) | 31.6 | 10.4 |
| February Average Temperature (Midwest) | 41.4 | 20.3 |
| February Average Temperature (Pennsylvania) | 35.8 | 15.2 |

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

Application

Description automatically generated with low confidenceGraphical user interface, application

Description automatically generated

1. Revisit the data to compare High and Low temperatures for February and March across the regions
   1. 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 | 31.60 | 43.40 |
| Low | 12.10 | 24.20 |

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

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

OR Create a line chart.

1. What assumptions did you make about the data?
   1. I made the assumption that the data is not fully accurate due to the innovations over the time periods, resulting in progressively more and more accurate results.

**What to turn in:**

* This document (saved as PDF) showing your understanding of the data visualization process

1. Listing and defining each stage of visualizing data, in your own words.
2. Explain the interactions between the stages of the data visualization process, in your own words.
3. List each stage and the interaction, in your own words.

* Combine all files:
  1. this document, with screenshots of visualizations included.
  2. the data file showing your filtered data (.xlsx)
  3. one set of Data Visualization Activity worksheets (acquire, parse, mine, filter& represent, refine) – with visualizations included
  4. 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**