CGT 27000 MIDTERM PART II

Visualization Challenge

Student: James Connor Colbert

Professor: Dr. Vetria Byrd

TA: Kushal Nerella

Term: Spring 2022

SPRING 2022

Contents

Punxsutawney Phil – 6 More Weeks of Winter

In last week's lab (Week 7) you performed exploratory analysis of Punxsutawney Phil data. In this challenge you will take a deeper look at the data and compare regions of data. You should create two (2) visualizations, this can be a collection of charts or a dashboard, whatever is necessary to tell the story or support the insight gained through your visualizations. Make sure you READ and FOLLOW ALL Instructions. The goal is to demonstrate your understanding of the data visualization process.

Data Description

For this assignment you will work with the Punxsutawney Phil dataset from Week 7's lab AND you will utilize an additional dataset from the National Centers for Environmental Information National Oceanic and Atmospheric Administration's (NOAA) web site. Both datasets are provided in Brightspace.

The first dataset is the Punxsutawney Phil data. This data includes shows years where Phil either saw his full shadow, partial shadow, and when there was no shadow recorded. The data also includes average temperatures for February and March, average temperatures for Northeast, Midwest and Pennsylvania for the months of February and March over a period of years. This dataset is available for you to download from Brightspace. You should download the Midterm version of the Punxsutawney Phil data (DO NOT use the file from Lab 7).

Additional data has been "acquired" from NOAA's website, and is saved on an additional tab in the Punxsutawney Phil workbook. The data records anomalies that have occurred in years when Phil's shadow sightings (or not) were recorded. This data is saved in a separate tab in the Punxsutawney Phil data. Unlike the Punxsutawney data, the amomolies data requires some pre-processing before it can be visualized.

The anomalies data: contains two columns of data: Year and Value. The units for the Value column are Degrees Celsius. The temperatures recorded in the Punxsutawney Phil data is in Fahrenheit. In order to perform your exploratory analysis, you will need to make sure the temperature variables are in the same units.

Lets' get started

After downloading the Punxsutawney Phil data open the Excel file (in Excel, recommended) and prepare the data for processing.

- Go to the Anomalies Tab and add a new column and name it Fahrenheit. Suggestion use cell C5

- You need to convert the Celsius data to Fahrenheit in this new column.

	A	B	C	D	E
1	Global Land and Ocean T	January			
2	Units: Degrees Celsius		Convert celsius to fahrenheit		
3	Base Period: 1901-2000		Fahrenheit = (Celsius x 9/5) + 32		
4	Missing: -999				
5	Year	Celsius	Fahrenheit		
6	1880	-0.02			
7	1881	-0.03			
8	1882	0.11			
9	1883	-0.25			
10	1884	-0.22			

- The function to convert Celsius to Fahrenheit is (Celsius degree x 9/5) + 32. Type this the following function into cell C6 =(B6*(9/5))+32

  Use the auto fill function in Excel to duplicate the formula in the remaining rows of data. An Excel Quick Reference (aka Cheat Sheet) has been provided with instructions for Basic Excel Skills. Refer to this document, located in Week 8 → Midterm Part II module in Brightspace, for basic Excel skills needed to complete this visualization challenge.

  Save your work before you proceed!

Parse the data

In the space below, parse the Punxsutawney Phil data, listing each variable and each data type that best describes the data. Also list the range of values for each variable.

In the space below, parse the <u>anomalies data</u>, listing each variable and each data type that best describes the data. Include the new Fahrenheit column in your parsing. Also list the range of values for each variable.

Year: integer *(1880 -> 2022)*

Value: float/double *(-0.74 -> 1.14)*

Fahrenheit: float/double *(30.668 -> 34.052)*

Let's explore the data

Mining involves math, statistics and data mining. The data in this case will receive only simple treatment. Let's do a few sanity checks:

- How many rows of **data** are there in the Punxsutawney Phil data?

  o 10

- What's the range of dates (in years) included in the Punxsutawney Phil data?

  o 1886 -> 2016 (130 years)

- How many rows of **data** are there in the Anomaly data?

  o 3

- What's the range of dates (in years) included in the Anomaly data?

  o 1880 -> 2022 (142 years)

- Calculate the min and max values for all quantitative fields for both the Punxsutawney Phil data and the anomalies data. (Hint: Use Excel Min and Max functions, and auto fill to duplicate the formulas). List the variables, and their max and min values for both datasets.

Punxsutawney Phil		
Variable	Min	Max
Year	1886	2016
February Average Temperature	25.23	41.41
February Average Temperature (Northeast)	10.4	31.6
February Average Temperature (Midwest)	20.3	41.4
February Average Temperature (Pennsylvania)	15.2	35.8
March Average Temperature	35.44	50.51
March Average Temperature (Northeast)	24.2	43.4
March Average Temperature (Midwest)	28.5	56.3
March Average Temperature (Pennsylvania)	24.5	47.7

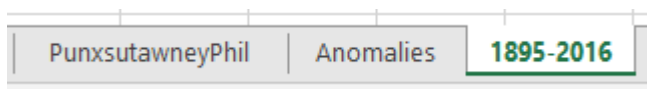| Anomalies		
Variable	Min	Max
Year	1880	2022
Value	-0.74	1.14
Fahrenheit	30.668	34.052

Filter the data

It will be easier to work with one tab instead of two, so let's add a tab that includes only the data of interest.

- Add a new sheet to your workbook (click the + to the right of the Anomalies tab).

| PunxsutawneyPhil | **Anomalies** | ⊕ |
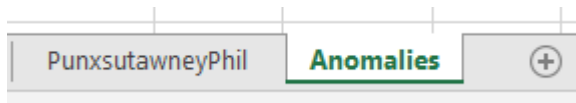
Rename Sheet2 to Filtered: right click Sheet2 → rename → type 1895-2016

| PunxsutawneyPhil | Anomalies | **1895-2016** |

-

To further assess the data, you need to filter out the data of interest and work from one spreadsheet.

1. Copy the three columns of data from the Anomalies tab (Year, Value and Fahrenheit) to the 1895-2016 tab, starting in cell A1.

	A	B	C	D
1	Year	Value	Fahrenheit	
2	1880	-0.02	31.964	

2. Copy the header row (row 1) from Punxsutawney Phil tab and paste it into the 1895 – 2016 tab, starting in cell D1

	A	B	C	D	E	F	G	H
1	Year	Value	Fahrenheit	Year	Punxsuta\	February /	February /	February /	Feb
2	1880	-0.02	31.964					
3	1881	-0.03	31.946					
4	1882	0.11	32.198					
5	1883	-0.25	31.55					
6	1884	-0.22	31.604					

3. Notice there are two columns with the heading Year. We are interested in data from 1895 to 2016 from both data sets. Locate and copy data ranging from 1895 to 2016 from Punxsutawney Phil and tab and copy and paste it into the 1895-2016 tab. Make sure the years align with the same years in the 1895-2016 data.

	A	B	C	D	E	F	G	H	I	J
1	Year	Punxsutawney Phil	February	February	February	February	March Ave	March Ave	March Ave	March Avera
2	1886	No Record								
3	1887	Full Shadow								
4	1888	Full Shadow								
5	1889	No Record								
6	1890	No Shadow								
7	1891	No Record								
8	1892	No Record								
9	1893	No Record								
10	1894	No Record								
11	1895	No Record	26.6	15.6	21.9	17	39.97	27.6	40.2	31.3
12	1896	No Record	35.04	22.2	33.5	26.6	38.03	25.3	36.9	27.8
13	1897	No Record	33.39	23.6	34.7	27.9	38.79	32	44	36.9
14	1898	Full Shadow	35.37	24.8	33.3	26.7	41.05	38	46	42
15	1899	No Record	25.5	18.1	22.2	20	37.63	29.3	38.4	34
16	1900	Full Shadow	30.76	21.4	27.5	24.1	41.27	26	37.9	29.3
17	1901	Full Shadow	29.86	16	26.2	18.9	40.71	30.8	41.9	35.1
18	1902	No Record	31.46	20.1	23.6	21	41.58	37.1	43.9	38.8
19	1903	Full Shadow	28.42	24.5	31.3	28	42.21	40.4	47.7	44.5
20	1904	Full Shadow	31.59	15	28.2	19.2	41.76	29.8	42.8	34

4. Paste the data in cell D17. Assuming you have not altered the order of the data, your 1895-2016 tab should look like the figure below.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Value	Fahrenheit	Year		Punxsuta	February	February	February	February	March Ave	March Ave	March Ave	March Avera
2	1880	-0.02	31.964										
3	1881	-0.03	31.946										
4	1882	0.11	32.198										
5	1883	-0.25	31.55										
6	1884	-0.22	31.604										
7	1885	-0.43	31.226										
8	1886	-0.17	31.694										
9	1887	-0.52	31.064										
10	1888	-0.32	31.424										
11	1889	-0.02	31.964										
12	1890	-0.33	31.406										
13	1891	-0.38	31.316										
14	1892	-0.23	31.586										
15	1893	-0.74	30.668										
16	1894	-0.42	31.244										
17	1895	-0.43	31.226	1895	No Recorc	26.6	15.6	21.9	17	39.97	27.6	40.2	31.3
18	1896	-0.15	31.73	1896	No Recorc	35.04	22.2	33.5	26.6	38.03	25.3	36.9	27.8
19	1897	-0.12	31.784	1897	No Recorc	33.39	23.6	34.7	27.9	38.79	32	44	36.9
20	1898	-0.01	31.982	1898	Full Shadc	35.37	24.8	33.3	26.7	41.05	38	46	42
21	1899	-0.14	31.748	1899	No Recorc	25.5	18.1	22.2	20	37.63	29.3	38.4	34
22	1900	-0.35	31.37	1900	Full Shadc	30.76	21.4	27.5	24.1	41.27	26	37.9	29.3
23	1901	-0.14	31.748	1901	Full Shadc	29.86	16	26.2	18.9	40.71	30.8	41.9	35.1
24	1902	-0.1	31.82	1902	No Recorc	31.46	20.1	23.6	21	41.58	37.1	43.9	38.8
25	1903	-0.23	31.586	1903	Full Shadc	28.42	24.5	31.3	28	42.21	40.4	47.7	44.5
26	1904	-0.61	30.902	1904	Full Shadc	31.59	15	28.2	19.2	41.76	29.8	42.8	34
27	1905	-0.35	31.37	1905	Full Shadc	26.94	15.2	22.2	18.1	45.12	31.4	47.2	36.9
28	1906	-0.13	31.766	1906	Full Shadc	33.69	22	30.7	25.3	36.1	25.7	33.9	29.1
29	1907	-0.37	31.334	1907	Full Shadc	35.46	15.2	31.5	20.1	45.7	33.3	50.2	39.5
30	1908	-0.38	31.316	1908	Full Shadc	33.01	18.4	31.2	22.1	43.92	32.7	47.1	38.4
31	1909	-0.58	30.956	1909	Full Shadc	35.02	26.4	37.2	32.9	40.35	29.6	41.1	33.1
32	1910	-0.35	31.37	1910	Full Shadc	29.3	20.6	28.5	23.8	49.44	37.8	51.9	42.6
33	1911	-0.54	31.028	1911	Full Shadc	33.66	21.4	37.5	28.5	44.02	28.4	42.9	32.8
34	1912	-0.27	31.514	1912	Full Shadc	31.08	18.6	25.8	21.1	35.76	27.7	35.2	31.5

PunxsutawneyPhil | Anomalies | **1895-2016** | ⊕

A little more work is needed to clean up the data. First, we only want data ranging from 1895 – 2016, so we need to remove (delete) the first 15 rows of data. Select and delete rows 2 – 16.

See Excel Cheat Sheet for instructions on how to delete a row in Excel (Course web page → Week 8 → Midterm Part II

The first 16 rows of your data should look like the snapshot below.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Value	Fahrenheit	Year	Punxsutav	February	February	February	February	March Ave	March Ave	March Ave	March Avera
2	1895	-0.43	31.226	1895	No Record	26.6	15.6	21.9	17	39.97	27.6	40.2	31.3
3	1896	-0.15	31.73	1896	No Record	35.04	22.2	33.5	26.6	38.03	25.3	36.9	27.8
4	1897	-0.12	31.784	1897	No Record	33.39	23.6	34.7	27.9	38.79	32	44	36.9
5	1898	-0.01	31.982	1898	Full Shadc	35.37	24.8	33.3	26.7	41.05	38	46	42
6	1899	-0.14	31.748	1899	No Record	25.5	18.1	22.2	20	37.63	29.3	38.4	34
7	1900	-0.35	31.37	1900	Full Shadc	30.76	21.4	27.5	24.1	41.27	26	37.9	29.3
8	1901	-0.14	31.748	1901	Full Shadc	29.86	16	26.2	18.9	40.71	30.8	41.9	35.1
9	1902	-0.1	31.82	1902	No Record	31.46	20.1	23.6	21	41.58	37.1	43.9	38.8
10	1903	-0.23	31.586	1903	Full Shadc	28.42	24.5	31.3	28	42.21	40.4	47.7	44.5
11	1904	-0.61	30.902	1904	Full Shadc	31.59	15	28.2	19.2	41.76	29.8	42.8	34
12	1905	-0.35	31.37	1905	Full Shadc	26.94	15.2	22.2	18.1	45.12	31.4	47.2	36.9
13	1906	-0.13	31.766	1906	Full Shadc	33.69	22	30.7	25.3	36.1	25.7	33.9	29.1
14	1907	-0.37	31.334	1907	Full Shadc	35.46	15.2	31.5	20.1	45.7	33.3	50.2	39.5
15	1908	-0.38	31.316	1908	Full Shadc	33.01	18.4	31.2	22.1	43.92	32.7	47.1	38.4
16	1909	-0.58	30.956	1909	Full Shadc	35.02	26.4	37.2	32.9	40.35	29.6	41.1	33.1
17	1910	-0.35	31.37	1910	Full Shadc	29.3	20.6	28.5	23.8	49.44	37.8	51.9	42.6

Scroll to the bottom of the file and remove any rows of data that have a year greater than 2016.

	A	B	C	D	E	F	G	H	I	J	K	L	M
122	2015	0.83	33.494	2015	Full Shadc	32.99	12.1	23.4	16.1	45.39	27.8	41.6	31.6
123	2016	1.12	34.016	2016	No Shadov	39.47	28.1	36.9	30.8	47.5	38.8	49.5	43.4
124	2017	0.98	33.764										
125	2018	0.75	33.35										
126	2019	0.94	33.692										
127	2020	1.14	34.052										
128	2021	0.78	33.404										
129	2022	0.89	33.602										

Remove rows 124 – 129. Let's remove the duplicate YEAR column. Delete column D.

See Excel Cheat Sheet for instructions on how to delete a row in Excel (Course web page → Week 8 → Midterm Part II

	A	B	C	D	E
1	Year	Value	Fahrenheit	Year	Punxsutawney Phil	Febr
2	1898	-0.01	31.982	1898	Full Shadow	
3	1900	-0.35	31.37	1900	Full Shadow	
4	1901	-0.14	31.748	1901	Full Shadow	
5	1903	-0.23	31.586	1903	Full Shadow	

Rename the Value column to Celsius.

The first 5 rows of your 1896-2016 tab should look similar to this

	A	B	C	D	E	F
1	Year	Celcius	Fahrenheit	Punxsutawney Phi	February	Februar
2	1895	-0.43	31.226	No Record	26.6	15
3	1896	-0.15	31.73	No Record	35.04	22
4	1897	-0.12	31.784	No Record	33.39	23
5	1898	-0.01	31.982	Full Shadow	35.37	24

Mine the Filtered Data

1. How many rows of data are there in the 1895-2016 filtered data?

    a. 122

2. How many times did Phil see is Full shadow between 1895 and 2016?

    a. 100 times

3. How many times did Phil see a Partial shadow 1895 and 2016?

    a. 1 time

4. How many times did Phil did not see his shadow 1895 and 2016?

    a. 15 times

5. How many times were there no record of Phil's shadow 1895 and 2016?

    a. 6 times

6. What's the Max and Min March Average Temperature?

    a. Min: 35.44

    b. Max: 50.41

7. What's the Max and Min March Average Temperature (Northeast)?

    a. Min: 24.2

    b. Max: 43.4

8. What's the Max and Min March Average Temperature (Midwest)?

   a. Min: 28.5

   b. Max: 56.3

9. What's the Max and Min March Average Temperature (Pennsylvania)?

   a. Min: 24.5

   b. Max: 47.7

10. What's the Max and Min Celsius temperatures converted to Fahrenheit?

    a. Min: 30.902

    b. Max: 34.016

The Assignment

There are multiple parts to this assignment. Make sure you read the entire assignment before starting.

Now that you have cleaned the data, determine a story or goal to support two visualizations you will create using the 1895-2016 data you just curated. Your two (2) visualizations MUST be different chart types. **This means DO NOT create two bar charts or two-line charts or two of the same chart types!** Challenge yourself. This is your time to show what you know. Hint: some more filtering and sorting of data might be helpful.

Required:

1. One of your visualizations (Figure 1) must include the anomalies data you converted to Fahrenheit.

2. Recreate the figure below using <u>March</u> Average Temperatures for Northeast, Midwest and Pennsylvania. Update the dates to reflect the 1895 -2016 date range of your filtered data. Name this figure: Figure 2
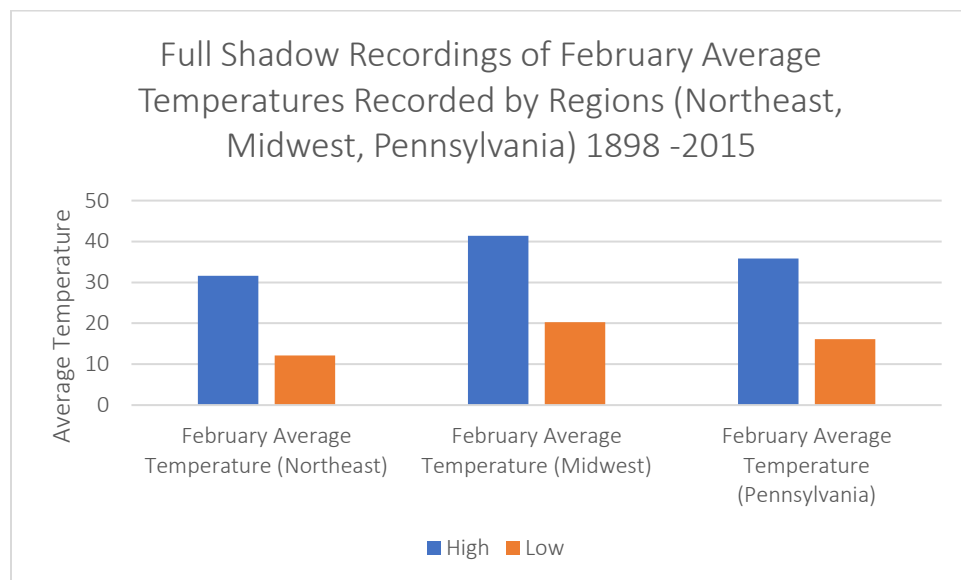
Use data visualization best practices!

Not sure what to do? Here are some suggested examples of what you could do:

- Dashboard summarizing Punxsutawney Phil's track record
- Forecast future shadow sightings
- Explore variation of the temperature when there was a full shadow/partial shadow/ no shadow

[Be creative and think of other things you could do](#)

Represent
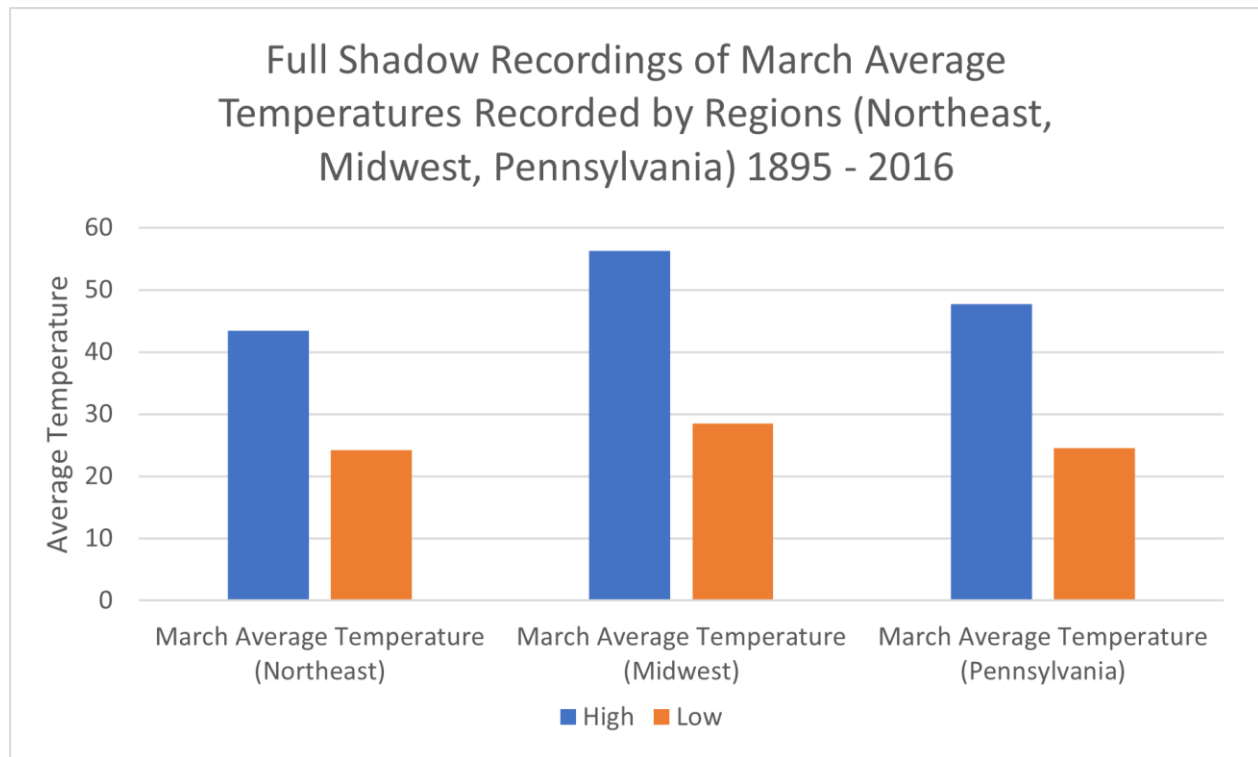
Figure 1. The highest and lowest recorded temperatures between 1895 and 2016 in March by region (Northeast, Midwest, Pennsylvania).

Average Temperature in 1895 - 2016 By Punxsutawney Phil's Shadow Compared to Total Average
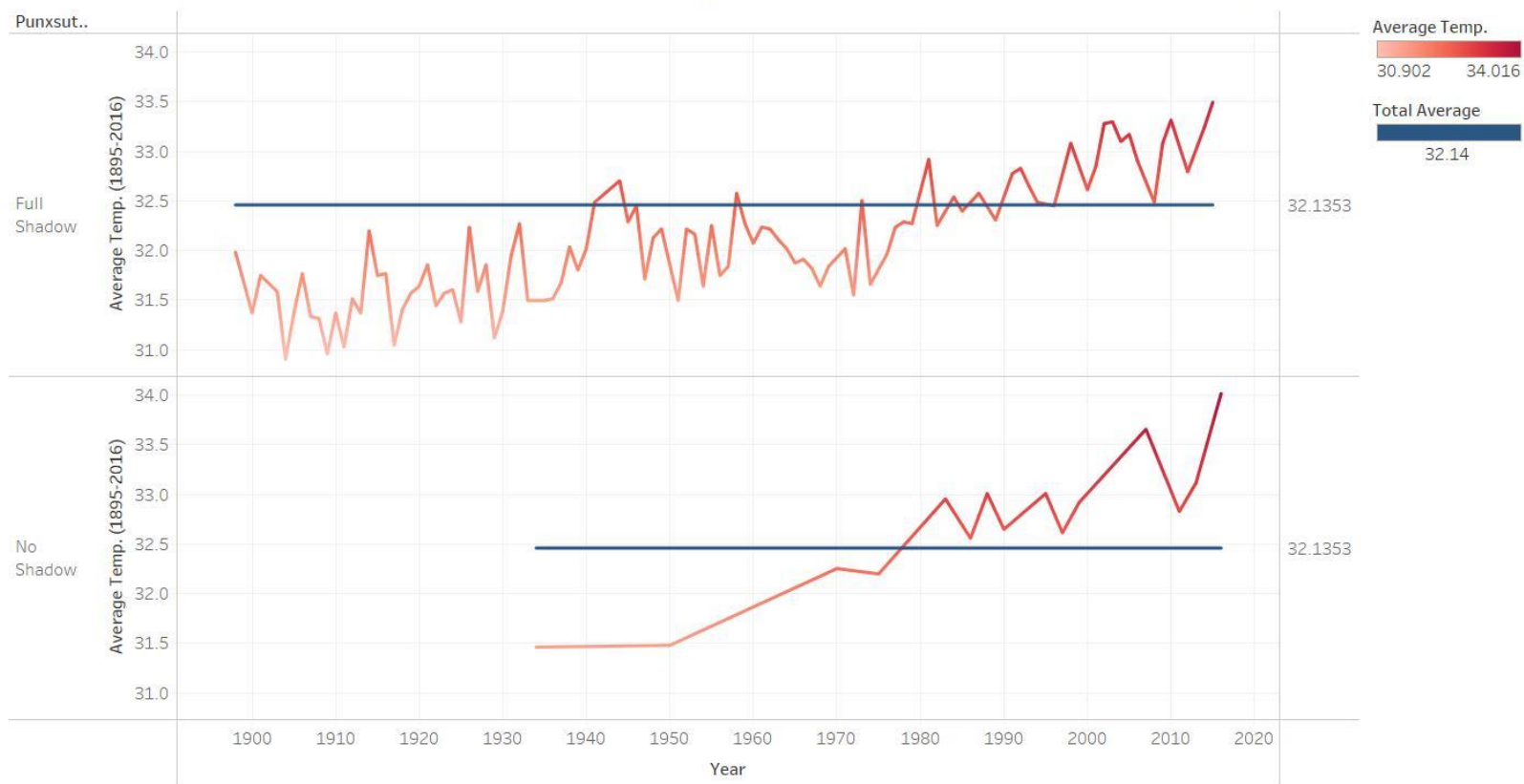
Figure 2. The average temperature between 1895-2016 compared to total average temperature by Punxsutawney Phil's shadow. Only Full Shadow and No Shadow are used due to minimal data from other entries.

Critique

Rate your visualizations (Figure 1 and Figure 2). Refer to the Data Visualization Checklist and rate your visualizations. This document is saved in the Midterm Part II Module on the course web page. DO NOT search for it on the internet.

Figure 1 Rating

Rate your visualization (using the Data Visualization Checklist)

Criteria	Maximum Points/Criteria	Fill in Your self-rating	
Text	12	10	
Arrangement	10	8	Calculate your % Score for Figure 1
Color	10	8	
Lines	8	10	_41_ / _48_ = _85_%
Overall	8	7	
Score	48	41	

Figure 2 Rating

Rate your visualization (using the Data Visualization Checklist)

Criteria	Maximum Points/Criteria	Fill in Your self-rating	
Text	12	8	
Arrangement	10	8	Calculate your % Score for Figure 2
Color	10	8	
Lines	8	5	___37___ / _48___ = __77___%
Overall	8	8	
Score	48		

Refine

In this part of the visualization challenge, you should identify one or more characteristics of the visualizations you created (Figure 1 and Figure 2) and update the figures. Include an updated version of each Figure below. In the figure caption, state what changes were made.
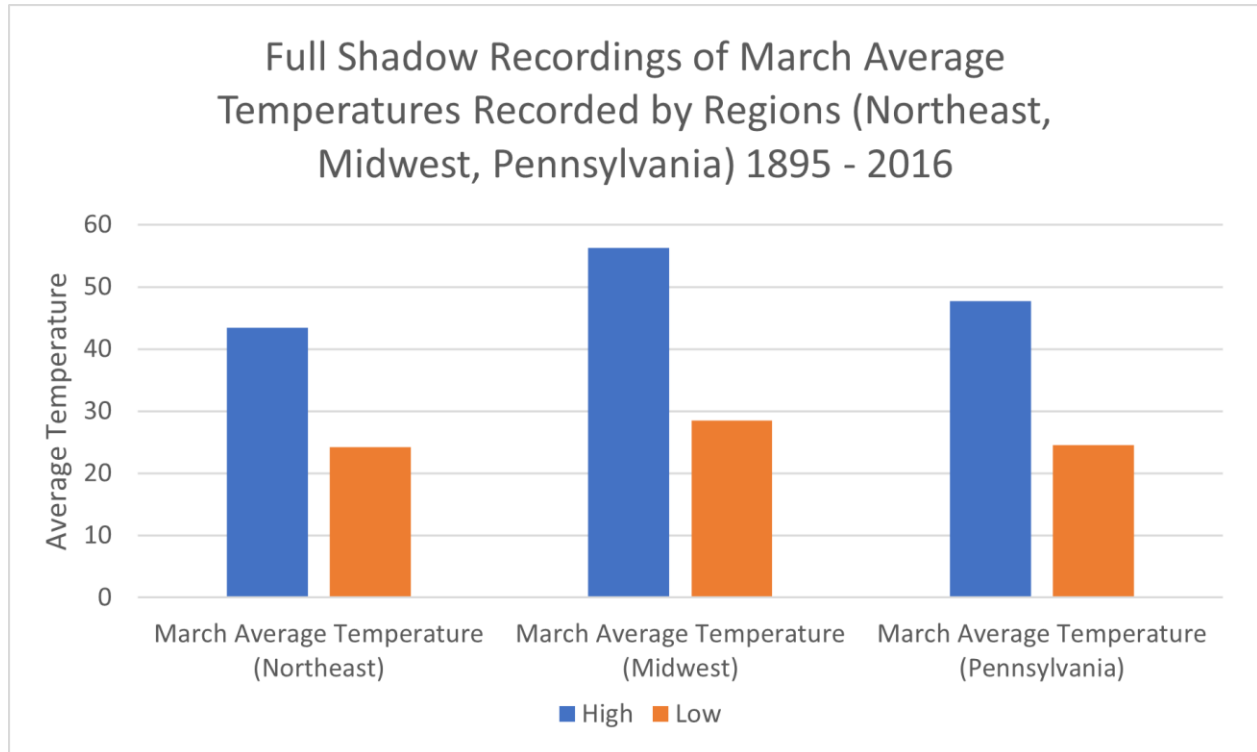
Figure 1 Refined.
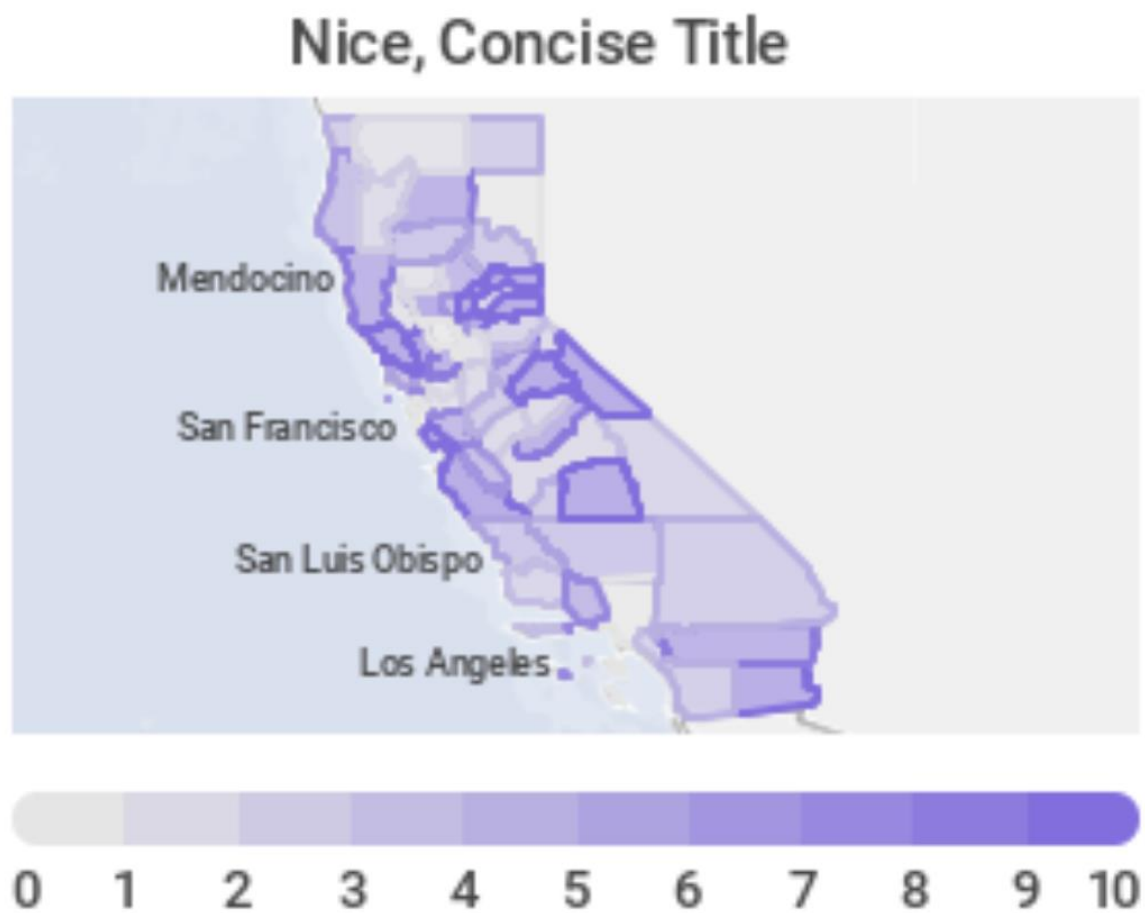
Nice, Concise Title

Figure 2 Refined. Added a nice, concise title, a progress bar to replace the cluttered legend.

Temperature in 1895-2016 Compared to Average Temperature by Punxsutawney Phil's Shadow
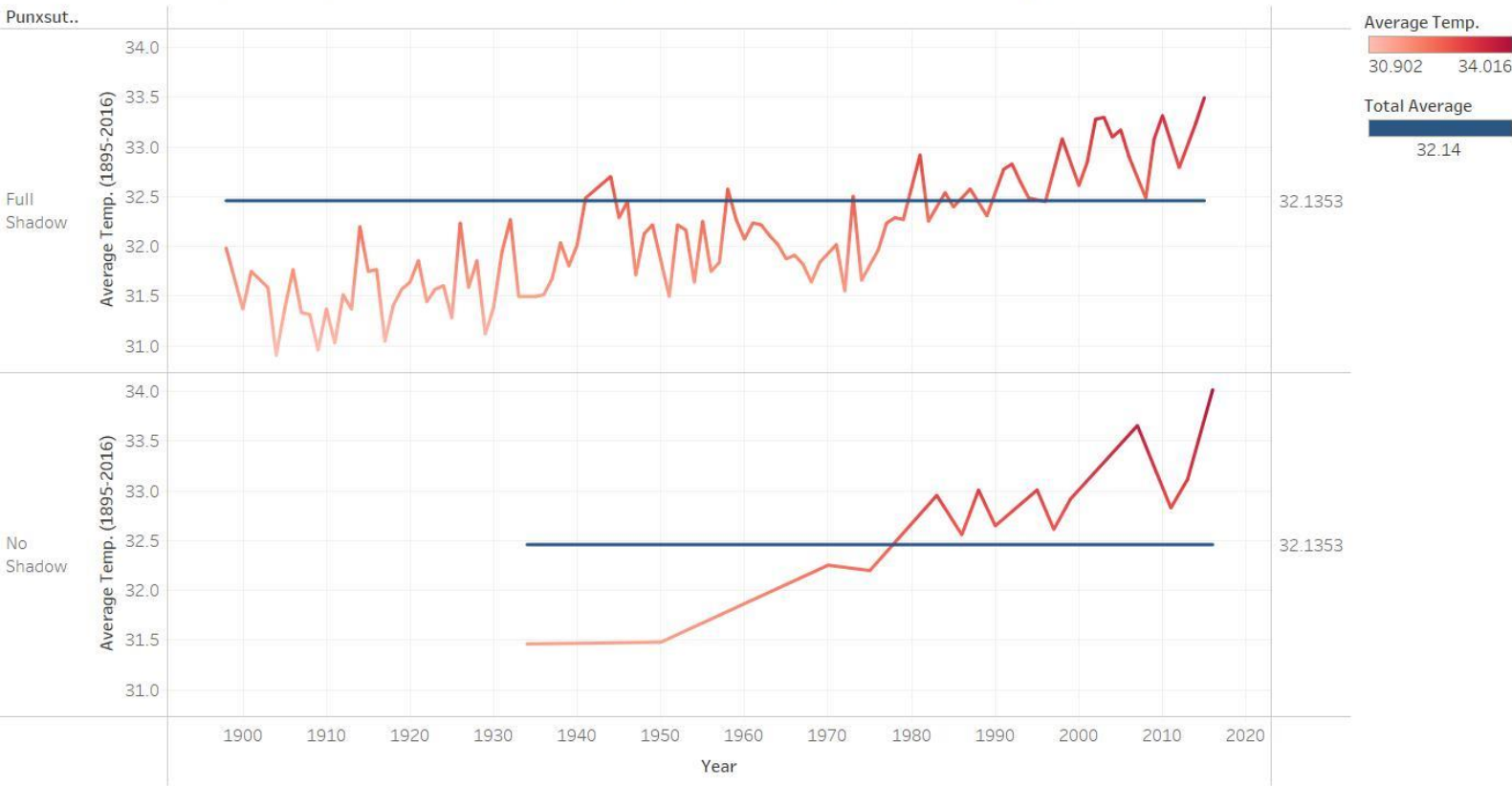
Figure 2 Refined. The average temperature between 1895-2016 compared to total average temperature by Punxsutawney Phil's shadow. Only Full Shadow and No Shadow are used due to minimal data from other entries.

What's the story?

<u>Using complete sentences</u>, answer the following questions:

1. The story I told with the data was that, when Phil sees a full shadow, the temperature typically dips below the average. In comparison, when Phil sees no shadow, the temperature does not dip as much. In Figure 1 this can be seen.

2. My audience consists of the citizens of the United States.

3. List 2 insights gained.

    a. Phil is not 100% accurate with his predictions.

    b.

Points will be taken off for incomplete sentences.

Bonus points for REALLY GOOD stories!

<u>Checklist of what to submit (2 files):</u>

- Save this file as LastnameFirstInitial_CGT270Spring2022_MidtermPartII.pdf

- Make sure your visualizations are included in the pdf file.

- Upload your data file showing your filtered data and basic calculations

Failure to follow these instructions will result in your work NOT being graded.

<u>General Deductions (others made accordingly)</u>

- No name on the first page of the document: -5 pts

- Did not use File naming convention: -2 pts

- Did not remove filler text/figures: -5 pts

- Altered template: -10 pts

- No figures included: -15 pts for each missing figure

- No figure captions: -5 pts for each missing figure caption

- Late submissions: Will NOT be graded (-80 pts)

- Provided a link to visualizations instead of including the visualization in line with the text: this will be treated as no figure, no figure caption (-25 pts)

- Failure to follow data visualization best practices (data visualization checklist): deductions made appropriately.

- General Rubric:

    o Mine Data   20

    o Represent   40

    o Refine   10

    o Storytelling (Your short story)   <u>10</u>

    Total   80 pts

You must submit your work in Brightspace by the end of lab.

Byrd Data Visualization Lab