

CGT 27000 Midterm Part II

Visualization Challenge



Spring 2022

Student: Keegan Palonis

Professor: Dr. Vetria Byrd

TA: Kushal Nerella

Term: Spring 2022

Contents

[Punxsutawney Phil – 6 More Weeks of Winter 2](#_Toc96923259)

[Data Description 2](#_Toc96923260)

[Lets’ get started 3](#_Toc96923261)

[Parse the data 3](#_Toc96923262)

[Let’s explore the data 4](#_Toc96923263)

[Filter the data 6](#_Toc96923264)

[Mine the Filtered Data 9](#_Toc96923265)

[The Assignment 10](#_Toc96923266)

[Represent 12](#_Toc96923267)

[Critique 14](#_Toc96923268)

[Refine 16](#_Toc96923269)

[What’s the story? 18](#_Toc96923270)

# 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).

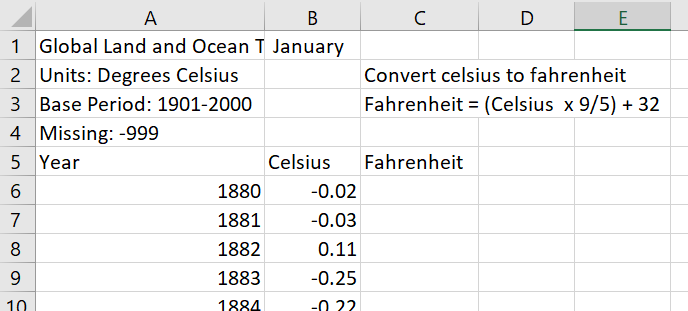
Additional data has been “acquired” from NOAA’s website, and is saved on an additional tab in the Punxsutawney Phil workbook. The data records **anomalies** that have occurred in years when Phil’s shadow sightings (or not) were recorded. This data is saved in a separate tab in the Punxsutawney Phil data. 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.



* 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.

Year: Integer and includes a range of 1886-2016.

Punxsutawney Phil: String with Values of No Record, No Shadow, Partial Shadow, and Full Shadow.

February Average Temperature: Float, with a range of 25.23-41.41, but possible range infinite

February Average Temperature (Northeast): Float with a range of 10.4-31.6, but possible range infinite

February Average Temperature (Midwest): Float with a range of 20.3-41.4, but possible range infinite

February Average Temperature (Pennsylvania): Float with a range of 15.2-35.8, but possible range infinite

March Average Temperature: Float with a range of 35.44-50.41, but possible range infinite

March Average Temperature (Northeast): Float with a range of 24.2-43.4, but possible range infinite

March Average Temperature (Midwest): Float with a range of 28.5-56.3, but possible range infinite

March Average Temperature (Pennsylvania): Float with a range of 24.5-47.7, but possible range infinite

In the space below, parse the anomalies data, 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 with a range of 1880-2022

Celsius (formerly Value): Float with a range of -0.74-1.14, but possible range infinite

Fahrenheit: Float with a range of 30.668-34.052, but possible range infinite

# 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?
  + 131 rows of data
* What’s the range of dates (in years) included in the Punxsutawney Phil data?
  + 1886-2016
* How many rows of **data** are there in the Anomaly data?
  + 142
* What’s the range of dates (in years) included in the Anomaly data?
  + 1880-2022
* 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.

Save your work!

February Average Temperature Max: 41.41

February Average Temperature Min: 25.23

February Average Temperature (Northeast) Max: 31.6

February Average Temperature (Northeast) Min: 10.4

February Average Temperature (Midwest) Max: 41.4

February Average Temperature (Midwest) Min: 20.3

February Average Temperature (Pennsylvania Max: 35.8

February Average Temperature (Pennsylvania) Min: 15.2

March Average Temperature Max: 50.41

March Average Temperature Min: 35.44

March Average Temperature (Northeast) Max: 43.4

March Average Temperature (Northeast) Min: 24.2

March Average Temperature (Midwest) Max: 56.3

March Average Temperature (Midwest) Min: 28.5

March Average Temperature (Pennsylvania) Max: 47.7

March Average Temperature (Pennsylvania) Min: 24.5

Celsius Max: 1.14

Celsius Min: -0.74

Fahrenheit Max: 34.052

Fahrenheit Min: 30.668

# 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).

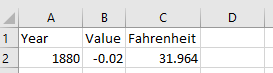


Rename Sheet2 to Filtered: right click Sheet2 🡪 rename 🡪 type 1895-2016

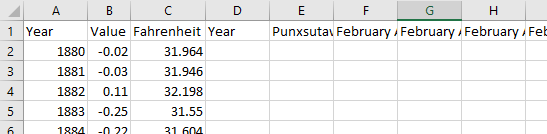
* 

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

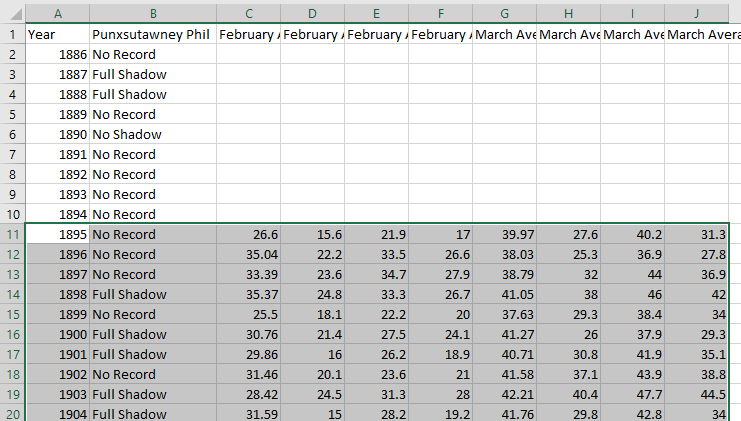
1. Copy the three columns of data from the Anomalies tab (Year, Value and Fahrenheit) to the 1895-2016 tab, starting in cell A1.



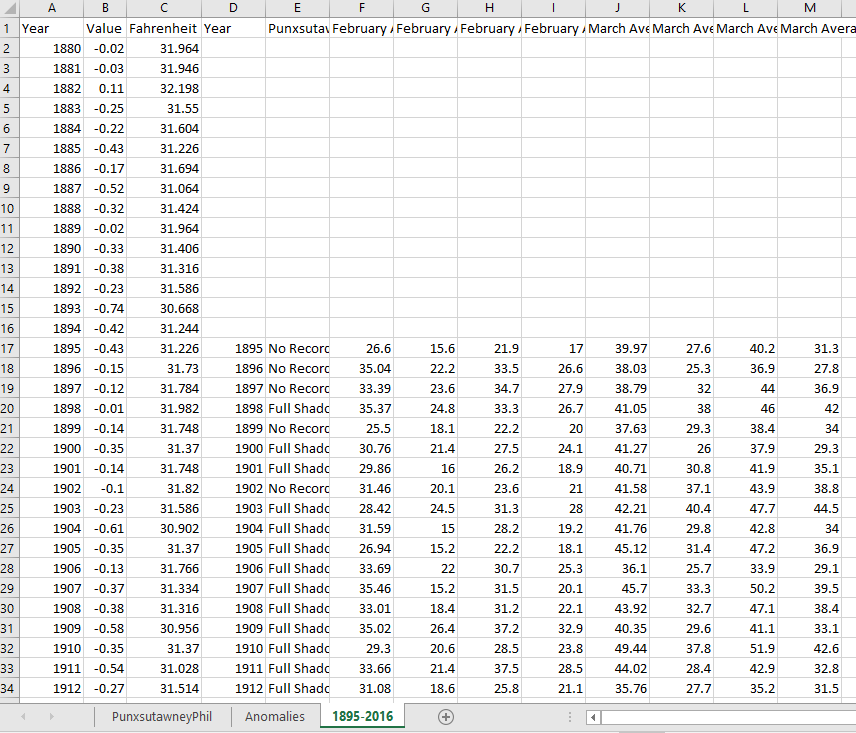
1. Copy the header row (row 1) from Punxsutawney Phil tab and paste it into the 1895 – 2016 tab, starting in cell D1



1. 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.



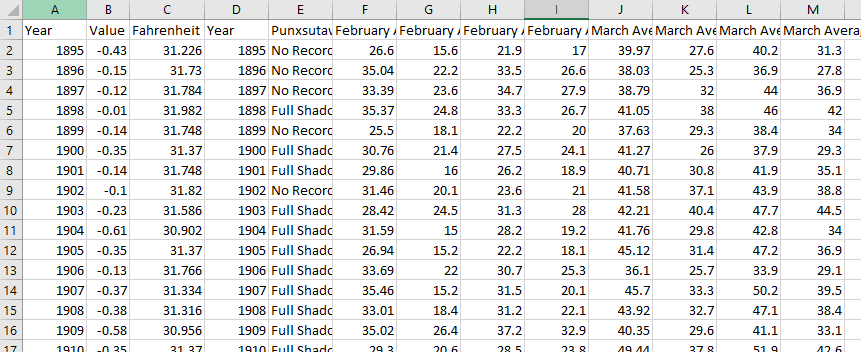
1. Paste the data in cell D17. Assuming you have not altered the order of the data, your 1895-2016 tab should look like the figure below.



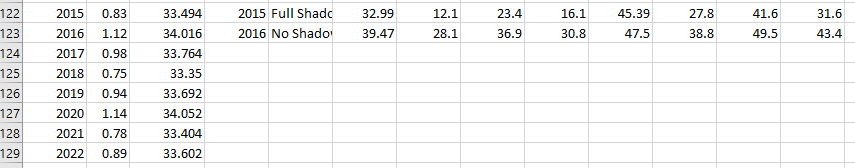
A little more work is needed to clean up the data. First, we only want data ranging from 1895 – 2016, so we need to remove (delete) the first 15 rows of data. Select and delete rows 2 – 16.

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.

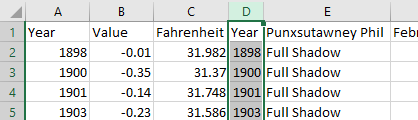


Scroll to the bottom of the file and remove any rows of data that have a year greater than 2016.



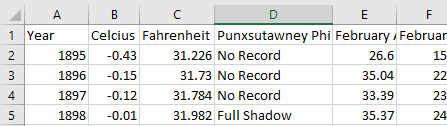
Remove rows 124 – 129. Let’s remove the duplicate YEAR column. Delete column D.

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



Rename the Value column to Celsius.

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



# Mine the Filtered Data

1. How many rows of data are there in the 1895-2016 filtered data?
   1. 122
2. How many times did Phil see is Full shadow between 1895 and 2016?
   1. 100
3. How many times did Phil see a Partial shadow 1895 and 2016?
   1. 1
4. How many times did Phil did not see his shadow 1895 and 2016?
   1. 15
5. How many times were there no record of Phil’s shadow 1895 and 2016?
   1. 6
6. What’s the Max and Min March Average Temperature?
   1. Min: 35.44
   2. Max: 50.41
7. What’s the Max and Min March Average Temperature (Northeast)?
   1. Min: 24.2
   2. Max: 43.4
8. What’s the Max and Min March Average Temperature (Midwest)?
   1. Min: 28.5
   2. Max: 56.3
9. What’s the Max and Min March Average Temperature (Pennsylvania)?
   1. Min: 24.5
   2. Max: 47.7
10. What’s the Max and Min Celsius temperatures converted to Fahrenheit?
    1. Min: C: -0.61 F: 30.902
    2. Max: C: 1.12 F: 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 March 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

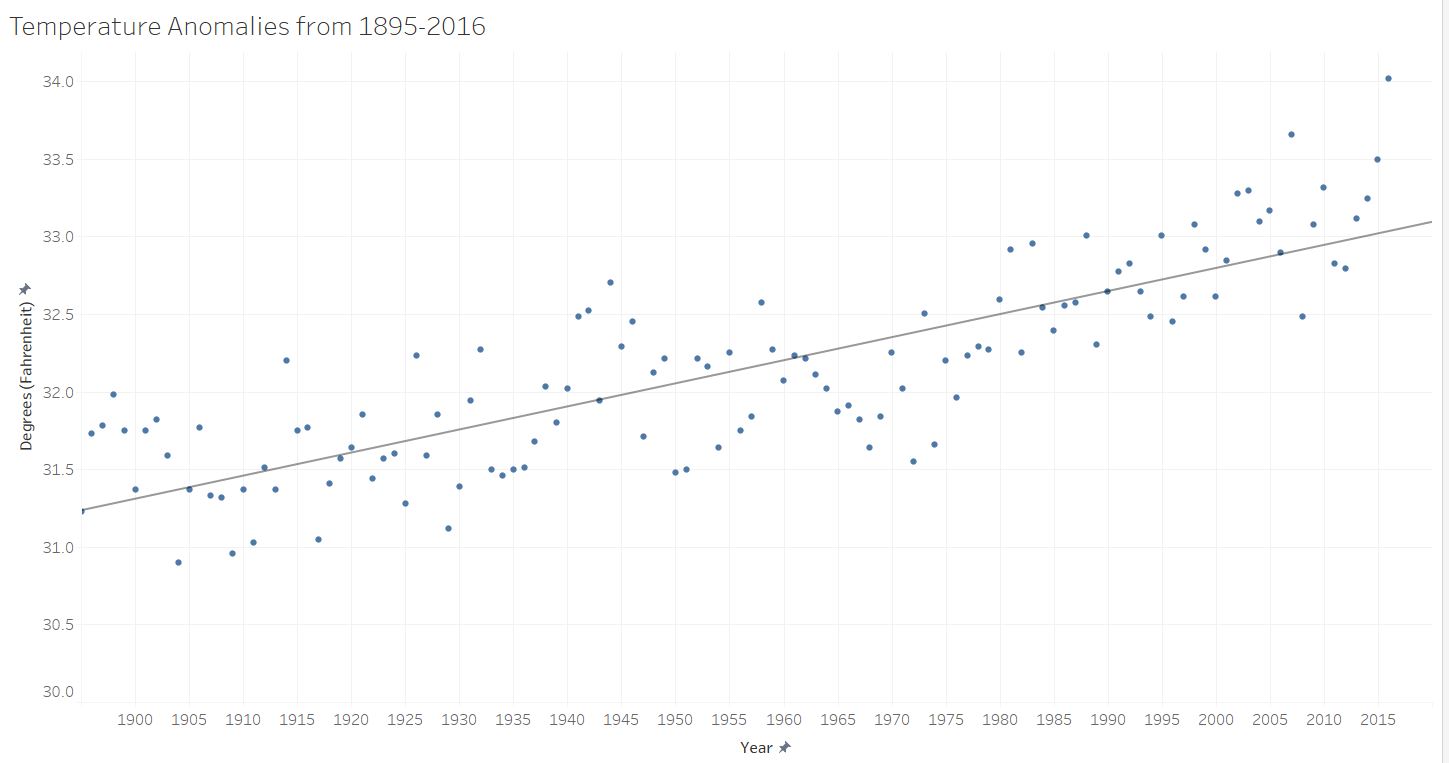
**Use data visualization best practices!**

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

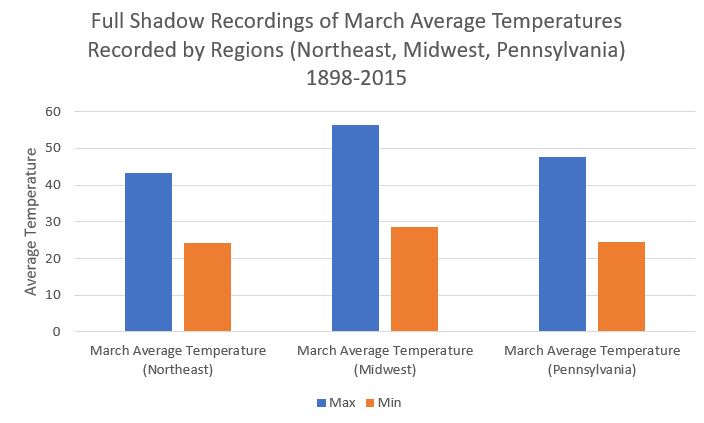
**Be creative and think of other things you could do**

# Represent

How to replace this figure: Right-click on the figure below, select Change Picture 🡪 From a File. Locate your figure.

Temperature Anomalies are the difference between the average temperature and a temperature that was experienced. For example if the Average Temperature was 50 degrees but have an anomaly of 20 degrees, the temperature could have ranged from 30-70 degrees.

How to replace this figure: Right-click on the figure below, select Change Picture 🡪 From a File. Locate your figure.



This bar graph shows the March Maximum and Minimum Average Temperatures when Punxsutawney Phil sees his full shadow. Data is selected from 1898-2015.

**Helpful Tip: Utilize the space that you have. Do NOT create a tiny visualization that is unreadable. Remember, the purpose of visualization is insight, but all insight is lost if it cannot be seen.**

# 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**

<add screenshot of rating for Figure 1 here>

Rate your visualization (using the Data Visualization Checklist)

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Maximum Points/Criteria | Fill in Your self-rating | Calculate your % Score for Figure 1  \_\_\_45\_\_\_ / \_48\_\_ = \_94\_% |
| Text | 12 | 11 |
| Arrangement | 10 | 10 |
| Color | 10 | 8 |
| Lines | 8 | 8 |
| Overall | 8 | 8 |
| Score | 48 | 45 |

**Figure 2 Rating**

<add screenshot of rating for Figure 2 here>

Rate your visualization (using the Data Visualization Checklist)

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Maximum Points/Criteria | Fill in Your self-rating | Calculate your % Score for Figure 2  45 / \_48\_\_\_ = 94% |
| Text | 12 | 10 |
| Arrangement | 10 | 10 |
| Color | 10 | 9 |
| Lines | 8 | 8 |
| Overall | 8 | 8 |
| Score | 48 | 45 |

# 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.

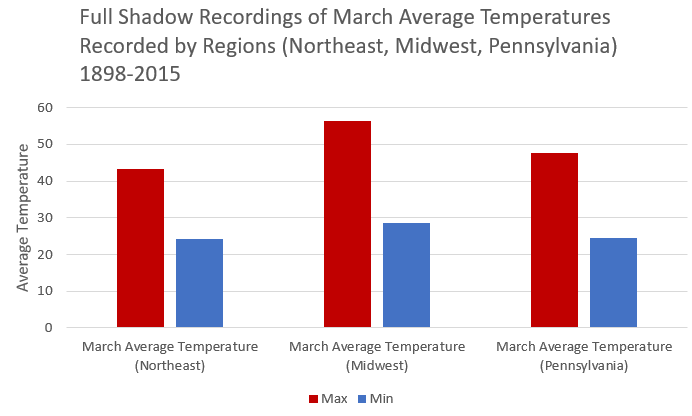


Figure 1 Refined. Left Adjusted Title and Selected Colors

Replace the picture with your visualization, remove this text before submission.

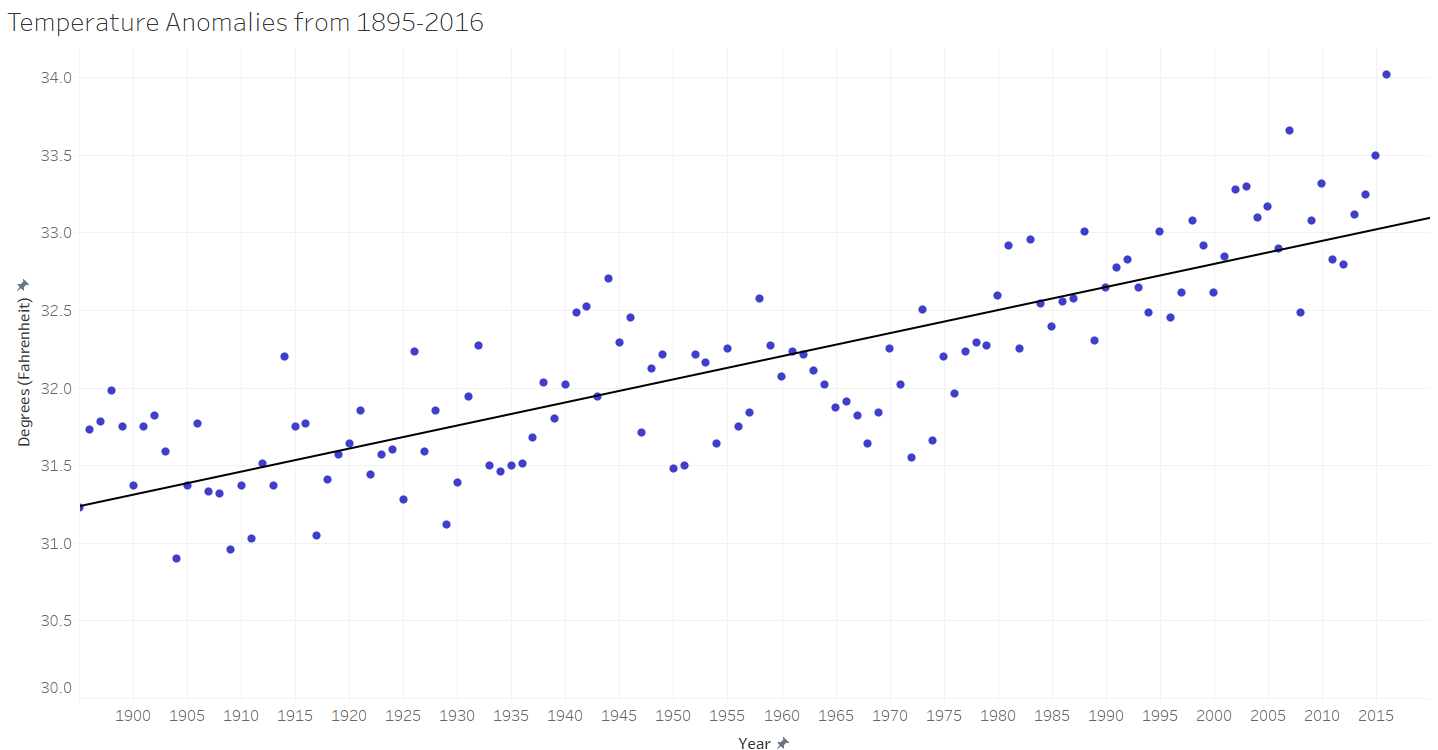


Figure 2 Refined. Changed colors and made Trend Line Bolder

# What’s the story?

**Replace the text on this page with your story**. The story should be no more than one-page **If you go over the page limit, your story will NOT be read (-5 pts)**. Single space Calibri Light (Body) font, max font size 13.

Using complete sentences, answer the following questions:

1. What story did you tell using the data? (5 pts) *(Note: your story must be supported by your visualizations: Figure 1, Figure 1 Refined, Figure 2, Figure 2 Refined). (***500 words max, with complete sentences). Make sure you mention the figures (by name, for example, as seen Figure 1, Figure 1 Refined, etc.)**
2. Who is your audience? (Use complete sentences) (1 pts)
3. List 2 insights gained. (4 pts)
   1. <insight# 1: complete sentences, 150 words max>
   2. <insight# 2: complete sentences, 150 words max >

Points will be taken off for incomplete sentences.

Bonus points for REALLY GOOD stories!

**Checklist of what to submit (2 files):**

* **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.**

**General Deductions (others made accordingly)**

* **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:**
  + **Mine Data 20**
  + **Represent 40**
  + **Refine 10**
  + **Storytelling (Your short story) 10**

**Total 80 pts**

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



Byrd Data Visualization Lab