**Exploratory Questions:**

* How does elevation affect the weather (rain, snow, temperature)? (line)
* Which month does it rain / snow the most? What month is the hottest? (line)
* How does wind speed affect rain / snowfall / temperature?
* Which state has the most snowfall / rainfall? What is the hottest / coldest state? (bin)
* How does the wind speed compare for the hottest and least hot state? (bar code plot and pull out states)
* How does snowfall compare to snow depth in different states for different months? (dual line)

**Visualizations to create:**

* Elevation vs. avg temp (line)
* Month vs. precipitation (line)
* Avg temp vs. state (bar) + highlight hottest & coldest state
* Count of states (bar)
* Wind speed (bar code) + highlight hottest and coldest state wind speed
* Wind speed vs. state + colored by avg temp (bar) x2
* Snowfall vs. snow depth vs. month (dual line)

**Analysis Process & Data Transformations:**

The data set that I decided to use is the “Daily Weather in the U.S., 2017” CSV file. First look at this data, it contains:

* Station
* State
* Latitude
* Longitude
* Elevation
* Date
* PRCP = Precipitation (inch)
* SNOW = Snowfall (inch)
* SNWD = Snow depth (inch)
* TMAX = Maximum temperature (F)
* TMIN = Minimum temperature (F)
* TAVG = Average temperature (F)
* AWND = Average daily wind speed (miles / hour)
* WSF5 = Fastest 5-second wind speed (miles / hour)
* WDF5 = Direction of fastest 5-second wind (degrees)

To conduct initial exploratory analysis, I used what we did in lecture to look at the distribution of max temperature as a barcode plot. I also used Python to look at the summary statistics of the data like the averages, medians, etc. Then, I utilized Tableau to conduct further exploratory analysis since I was more familiar with it and could create more visualizations quicker. The questions that I initially thought of just by looking at the data were:

* How does elevation affect the weather (rain, snow, temperature)?
* Which month does it rain / snow the most? What month is the hottest?
* How does wind speed affect rain / snowfall / temperature?
* Which state has the most snowfall / rainfall? What is the hottest / coldest state?

With these questions, I created visualizations in Tableau to begin my investigation. As I was creating the visualizations and looking at the data, two other questions came to mind:

* How does the wind speed compare for the hottest and least hot state?
* How does snowfall compare to snow depth in different states for different months?

After creating my visualizations in Tableau, I began thinking of which visualizations I wanted to create in JavaScript and how to translate them from Tableau to JavaScript. After looking through my visualizations, I decided to create a line graph for elevation vs. average temperature, a line graph for month vs. precipitation, a bar graph of average temperature by state to find out what the hottest and coldest states are, a bar graph of the count of the states to make sure there was an understanding of where most of the weather data was coming from, a barcode plot of average wind speed with the hottest and coldest state’s wind speed highlighted, a bar graph of wind speed vs. state that has different shades by average temperature, and a dual line graph of snowfall vs. snow depth vs. month. I chose to create these particular visualizations from my questions since I thought they were the most interesting and appeared to be somewhat related to each other.

Going through the process for each of my visualizations:

* Line Graph: Elevation vs. Average Temperature
  + Creating this visualization, I had to look up how to create a line graph using d3, and found that there is a d3 graph gallery that provides starter code on how to create a basic line graph. I used this code as a template and was able to create a line graph using the weather data. However, when I first created the line graph the lines seemed to be connected to many different data points, which made the graph completely unreadable and dense. To fix this, I had to figure out how to order the data … had to cut down the amount of data…
* Line Graph: Month vs. Precipitation
  + Since I already created a line graph for the elevation and average temperature, I was able to use that code and just replace the data with the month and precipitation. The main difference was the date scale since I had to pull out the month from the dates using d3’s timeParse and then used scaleTime to put the months on the x scale.
* Bar Graph: State vs. Average Temperature w/ Highlight on Hottest & Coldest State
  + Creating this visualization, I utilized the binning that we did in lecture and also information from the Internet. To show the highlight, I made the hottest and coldest states a darker shade to make it stand out.
* Bar Graph: Count of States
  + I used the same method as the state and average temperature bar graph to create this one, except I had to create a new column of data for the count of states.
* Barcode Plot: Average Wind Speed w/ Highlight on Hottest & Coldest State’s Average Wind Speed
  + To create this visualization, I used the bar code plot code we did in class as a template to create this visualization. To highlight the hottest and coldest state’s average wind speed, I had to pull out that data and create separate barcode plots for them. I then put them next to the average wind speed plot to highlight and compare them.
* Bar Graph: State vs. Average Wind Speed & Colored by Average Temperature
  + I used my bar graph code as a template to create the graph, and then had to look up how to color the graphs by average temperature.
* Dual Line Graph: Month vs. Snowfall & Snow Depth
  + Creating this graph, I had to look up how to do this and found starter code from d3’s API.

Insights (fig 1, fig 2…):

* Line Graph: Elevation vs. Average Temperature
  + It appears that when the elevation is around 0-200 m, the data there is very dense. Additionally, around 2400 m, appears to be the biggest spikes and where the average temperature is the highest. Between 200~1400, the data seems very sparse and the average temperature is low.
* Line Graph: Month vs. Precipitation
  + There is a clear decline in precipitation inches as the year progresses. This makes sense since January is the start of the year and still winter. As the year progresses, it begins to become warmer which may be why precipitation decreases. It is interesting though, that there appears to be a spike in August for some reason.
* Bar Graph: State vs. Average Temperature w/ Highlight on Hottest & Coldest State
  + The hottest state is Texas and the coldest state is NT.
* Bar Graph: Count of States
  + It appears that Utah has the highest count of data and MP has the lowest.
* Barcode Plot: Average Wind Speed w/ Highlight on Hottest & Coldest State’s Average Wind Speed
* Bar Graph: State vs. Average Wind Speed & Colored by Average Temperature
  + It appears that the highest average wind speed is in Texas and the second is California.
  + The highest average temperature is Utah, but the average wind speed is pretty low.
* Dual Line Graph: Month vs. Snowfall & Snow Depth
  + It appears that both the snowfall and depth decline as the months go by, which makes sense since the season is moving towards summer. It’s interesting how the greatest snowfall is in January but the snow depth in January is not the largest. As the snowfall declines, the snow depth increases until it hits March, where it then begins to decrease. This might be due to the snow not sticking or maybe just building up.

**Main Lessons Learned:**

One of the main lessons I learned from analyzing this data, is that it’s important to understand the tools that you are using to create the visualizations. With an understanding, you are able to better manipulate the data and create polished graphics. Through this assignment, I was able to learn and gain experience with using d3 and vega in JavaScript. I’ve never created visualizations in JavaScript before, so this was a bit of a learning curve for me. It was initially difficult for me to even create a basic line graph, but by reading through the APIs and getting help from the Internet, I was able to create a graph using the weather data. Another lesson I learned was that it’s important to have clean data and transform your data before attempting to create any visualization. This makes the process much easier and smoother. Additionally, one more lesson I learned was how important initial investigation is. It’s important because through the process, you encounter interesting statistics and comparisons, which leads to more complex questions.

**Feedback Given:**

* Use different shades of the same color
* Make the visualization bigger
* Change the shape of points to indicate a different variable on the same graph
* Add labels to x and y + a title