Week 4 note taking. Hunju , Harsh

* Visualization techniques and priorities.
We can visualize the characteristics of data by using various characteristics of data. For example, position in the x and y axis, different color, size, shape, or showing relationship connection are most frequently used techniques in data visualization.

However, it is important to be selective in creating visualization as a data scientist. The goal of visualization is to convey the clear message of data to the audiences in a effective way. Therefore, if you are trying to add so many messages, your customer won't catch the point and you won't attain the goal.

* What is Continuous Data?
There are two different type of data, discrete and continuous data. It will be easy to think of 'MRI data' as an example of continuous data. If you are taken MRI, millions of data will be stored in just one slice of picture and these data are called 'continuous data'. This type of data are continuously connected without specific regulation.

* How should we represent the data points?
If there are many data points in the x and y axis, it is a good way to show the trends of data by making some connection of each points. Also, to gain more correct answer from the data, we also have to specify the bin on the x-axis. This work is more crucial when the trend is abruptly changed(such as weather and precipitation change). However, if we had not gained any data from specific bin size, we can combine the bin with adjacent bin to compensate for 'no data' situation. You also have to be reminded that more than one data should be required to show the trend. If not, this one data could distort the overall trends.

* Deciding bin size
If bin sizes are not uniform, you have to decide the bin size by drawing the rationale or using formulas.

* Arithmetic Average vs Weighted Average
Arithmetic Average is 'mean'. To get this number, you can sum all the values and divide it by number of values. If you want to get Weighted Average, you can weight the value which tends to occur more.