16 February 2019

Jurgen Tan 3B01

Self-reflection

# Challenges faced

One of the challenges that I faced was choosing of the dataset, as the dataset had many values and different categories. This made it hard to drill down as to which category to choose from. The dataset originally had data from ITE all the way up to Universities. Hence, I decided to use only the Universities data as it contained the main chunk of the dataset and since most of the audience of the presentation are going to be enrolling into University soon.

Subsequent challenge that I faced was different date ranges for each dataset. The datasets that were found on the data.gov website as inconsistent in the data ranges. Some of the datasets had data from the 1980s while come only had data for a single year. This made it really hard to draw a connection and inference based on data that are from different years as there are many other factors that could affect in different years.

Another challenge which I faced was the lack of examples and documentation from Seaborn. This could be a personal peeve, as I like to read proper documentations and methods of usage before trying it out. Hence, trying to find information about the library was quite challenging. However, in the end I had come to realized that it is very similar to Matplotlib and many codes that were used in Matplotlib could be used in Seaborn.

# Acheivements

I feel that I had managed to achieve the skill of being able to code out proper graphs using different libraries and doing basic cleansing of unclean and unstructured data using the Pandas library. While doing this assignment, I had also managed to stumble upon the Plotly library, which allowed the graphs that were produced to be interactive and show the other statistics or information about the graph. Which proved to be very informative for me when I was trying to draw an inference and conclusion.