

DS 203 : Programming for Data Science
Tutorial and Assignment Sheet – 5
Data Visualization

Submission guidelines:

- Prepare an ipython notebook and name it <roll no.>.ipynb and submit it on Moodle before 11:59pm on September 23, 2020.
1. For the data source at <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>, visualize the following data with the appropriate type of graph, and use the right options to make the graph look good (such as using legends and axes titles with the legible font size, and exploring color palettes):
 - a. Pick the top six countries by average yearly suicides, and display their suicide for each year separately.
 - b. For the same six countries compare the mix of age groups. What does the plot tell you about the differences or similarities by country?
 - c. Plot an appropriate set of graphs or charts that highlight the consistency of difference between males and females when it comes to suicide rates.
 - d. Using an appropriate graph, show the worst year for each generation in the US.
 - e. Plot a bihistogram for a few specific countries (for a year, say 2000) for male and female populations by age ranges to highlight some differences in sex ratios between countries. Check out: <https://www.itl.nist.gov/div898/handbook/eda/section3/bihistog.htm> and <https://stackoverflow.com/questions/62678411/how-to-plot-a-paired-histogram-using-seaborn> for ideas
 - f. Show a bubble plot to show the relation between suicide rates, human development index (HDI), and population. Due to the large spread in population, you might have to use a transform. Is there any interesting observation?
 2. Repeat the exercise for data at URL <https://www.kaggle.com/jmmvutu/summer-products-and-sales-in-ecommerce-wish>. Here, explore six different relations of your choice, and try to plot a different type of plot of your choice (but appropriate for the insight being sought) for each part.