

1. For the data source at

<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>,

perform the following steps in python using pandas, matplotlib and/or seaborn. Use code cells to perform functions with a comment for each line explaining what it is doing (and using intuitive variable names), and mark-down cells to note down any significant observations after each code cell (e.g., "Variable X appears to be normal distributed"):

a. Preliminaries:

- i. Print the datatype of each variable.
- ii. For each variable, print the number of unique values.
- iii. For each variable, display the number of missing entries.
- iv. Find the number of records with no missing entries.

b. Discrete variables:

- i. For each variable, plot the frequency of each unique value (histogram).
- ii. For each variable, identify the mode value.

c. Continuous variables:

- i. For each variable, print mean, variance, skew, min, max, median, 25th percentile, 75th percentile, and interquartile range.
- ii. For each variable, use QQ-plot to see the extent to which the variable deviates from normal distribution, and how (left-skew, right-skew, or more like uniform distribution)
- lii. Plot the heatmap and observe the correlation matrix between various variables. (Highly correlated variables do not give us much new information, hence you may drop them from your analysis)

Bonus - Try to perform such (and even more elaborate analysis) on datasets of your choice!!