Homework: Exploratory data analysis and distribution

Use the same data set we used in the class for questions 1-4.

1.   Please plot histograms of any four features in one plot. Use different colors and transparency to make them visible.
2.   Please plot these features in a plot and use different axes to include them in one plot.
a.   SWU, T, tc, and ice.
3.   Please plot these features in a plot and use the standardization to include them in one plot.
a.   LWU, u, cbh, and ice.
4.   Make two box diagrams (any two features of your choice) in one plot and plot them in different colors so that they are easily distinguishable.

5.   Now Load the wine dataset and plot a scatterplot matrix and a heat map.
df_wine = pd.read_csv('https://archive.ics.uci.edu/'
            'ml/machine-learning-databases/wine/wine.data',
            header=None)
6.   Plot the histograms for the proline concentrations in wine class levels 1 and 3. Then plot the Gaussian distribution function for proline on top of the histograms. You might see multiple lines when you plot the distribution since the Proline values are not in order. That's why lines begin from one value and jump to the other and many lines appear. Can you fix that problem and plot just one line?
7.   What is the probability of having a concentration of
a.   600 and 1400 in class 1 wine
b.   400 and 1200 in class 3 wine.
c.   If someone dislikes the taste of Proline and likes to drink wine, which wine should he choose?
   *** Hint: you can use np. where function to find the index values related to each wine.

For any questions, contact me via Slack.

Please submit your code and the plots on blackboard.