Problem Set 2:

1. This one is a bit of a doozy and has several parts. You're going to need MatLab, or something, and a data set available with this assignment on canvas. You'll need the file "raman_spectrum_alcohols.mat". In this file you'll find three variables:

   alcohol_raman, this is the raman spectrum of several alcohols

   RamanShift, this is a vector for the X-axis

   Sample_labels, labels corresponding to each sample

   You're going to need to plot and manipulate this data. It turns out that different alcohols have different Raman spectra. This is because they often have different additives or various aromatic compounds either added to them or dissolved in them through the aging process. All the data and these questions come from the paper "Through-bottle whisky sensing and classification using Raman spectroscopy in an axicon-based backscattering configuration" written by Holly Fleming and their colleagues.

   a. Plot the Raman spectrum of "Absolut Vanilla" and "Talisker 10 Year". Which peaks on the plot are the same in each beverage? Which is alcohol?
   b. Now we're going to need to correct for the fluorescence background of the sample. We have the whole internet at our disposal to find this information. Find and use an appropriate algorithm or package to remove the fluorescence background and correct for the baseline. Plot it
   c. Now that we know how to do that plot the whole data set.
   d. Are you able to make any useful observations using the plot at hand? Is this display conducive to classifying alcohols?
   e. Now to move to a different space. Plot the first two principal components of each spectra with respect to each other. There's a few good subroutines out there for this.
   f. Is there a better combination of principal components to distinguish between the two than the first and second?
   g. Is it easier to distinguish different types of alcohol by their Raman spectra or the principal components of it?
   h. Last thing, plot all the spectra without baseline correction. Why do the different types of alcohol have different baselines? For example, why is the whole spectrum for Talisker shifted higher than absolut?

For e through f, you should have a scatter plot like so: