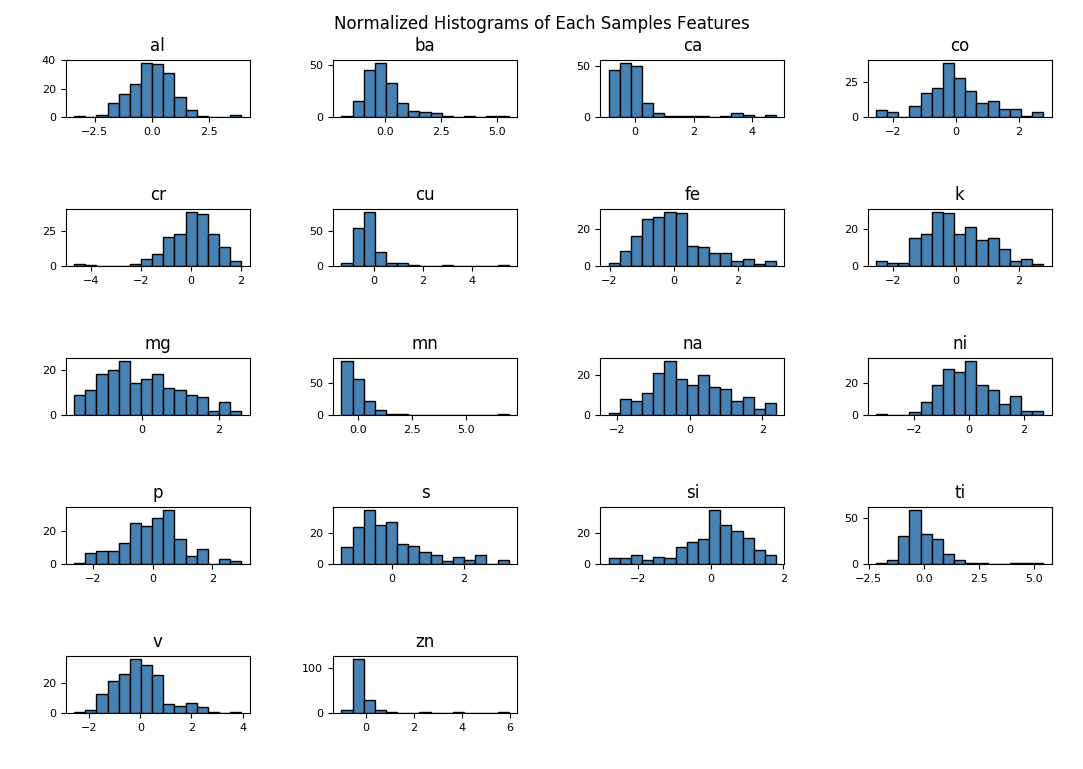
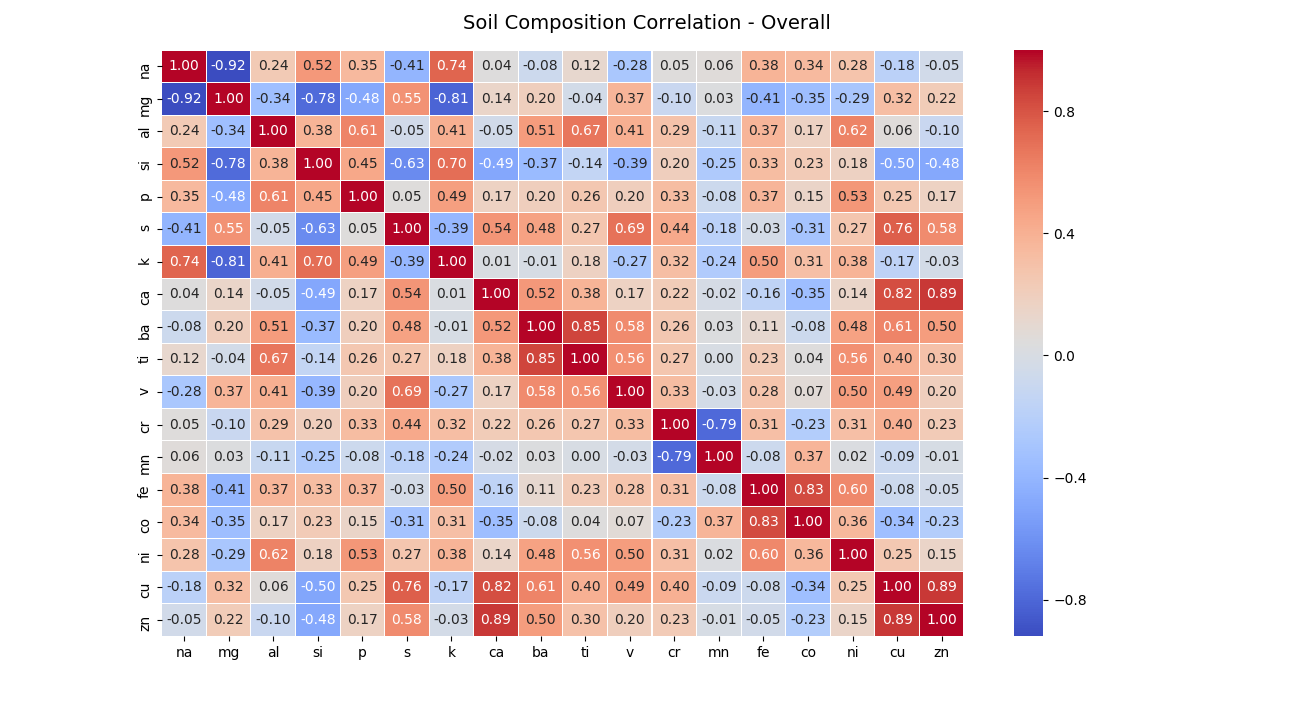
A preliminary analysis of soil samples collected in VCP. These samples were analyzed as a whole population and classified amongst each samples assigned region of collection.

The following graph shows the distribution of samples in accordance with measured soil composition. Each samples measurement scale has been normalized and therefore presents all data on the same scale. The y axis shows the n number of samples in each histogram bucket for the 180 samples collected.



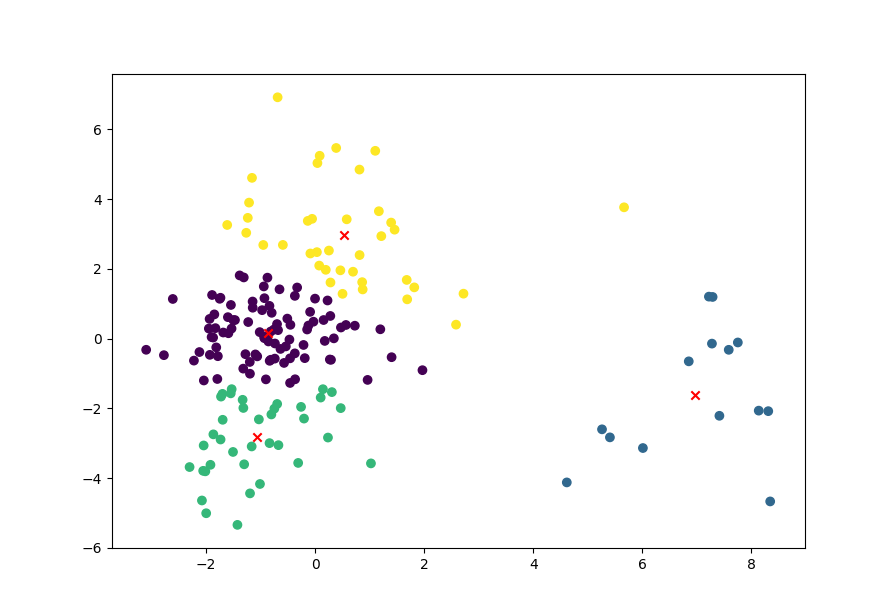
This Figure shows a high level summary of the features collected during the course of this soil study.

The relationship between each feature in the population can be shown using a correlation matrix presented below,



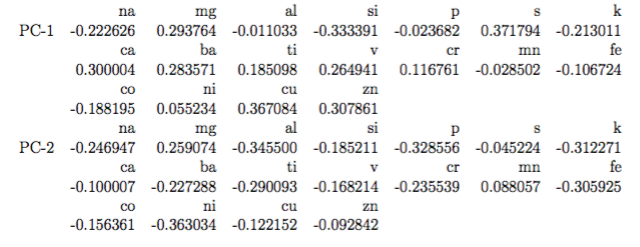
As expected all elements along the diagonal of the correlation matrix are 100% correlated. This is expected because this represents the relationship between the element and itself. Inversely related elements are therefore represented as values which are close to a value of -1.0 in the matrix, or dark blue on the heat map.

An initial look at KMeans clustering among these 4 different regions (tb, vh, nwf, cw) initially show that clustering is possible, although the accuracy of these results needs further analysis. Visually it is presented as



where the x-axis represents the first principal component, and the y axis represents the 2nd principal component. Each centroid is denoted by a red X. These samples may be clustering for a specific reason although it is not clear if this reason is geographically based, which seems initially to not be the case.

The PCA analysis breaks down for each component by analyzing the Eigen values for each component versus each feature. The number of components was supplied at 2 to allow for the ability to visualize the results. These represents the influence or weights of each feature as it influences the components makeup.



A Support Vector Machine analysis, specifically Support Vector Classifier produced results of classification score of close to 0.5972 or 59.72%. This was only an initially test as no cross validation or tuning has been performed as of yet.