**Analysis of Lithium Ion Silicate Data Set**

**Na’il Mitchell**

**data set description:**Data set of chemical and physical properties of lithium ion silicate cathodes.

**data set url:** <https://www.kaggle.com/divyansh22/crystal-system-properties-for-liion-batteries>

**data set dimensions:** 339 rows, 11 columns

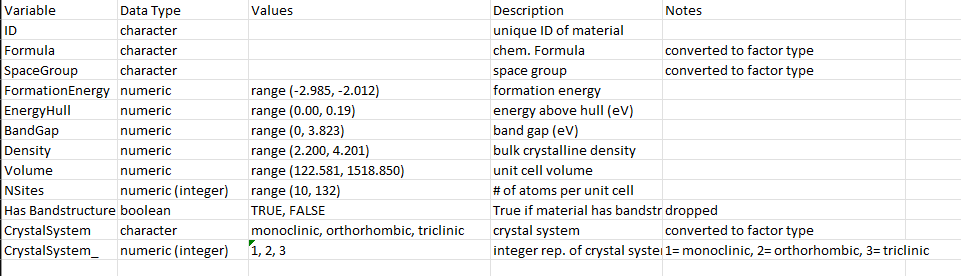


Table 1. *Data dictionary*

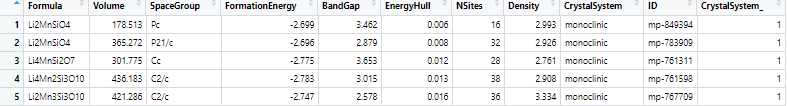


Table 2. *First five rows in data set.*

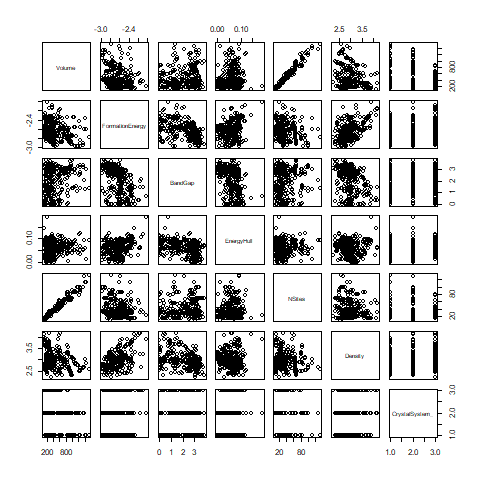


Figure 1. *Scatter plot matrix of all numeric variables in the data set, and Crystal System. High linear correlation is evident between Volume and NSites. Density and Formation energy also have noticeable correlation. Most correlation in the data appears to be weak to little.*

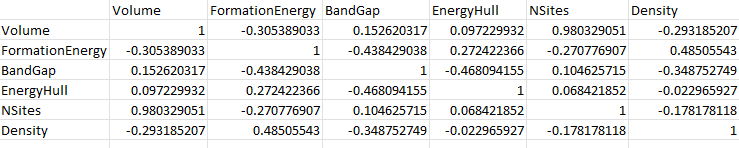


Table 3. *Correlation matrix of the numeric variables in the data set showing correlation coefficients. Values are between -1 and 1. -1 represents perfect negative linear correlation and 1 represents perfect positive linear correlation.*

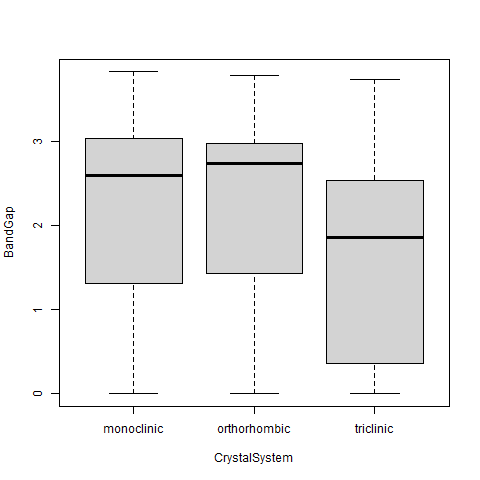


Figure 2*. Bar plots showing differences in median band gap values for the different crystal system groups.*

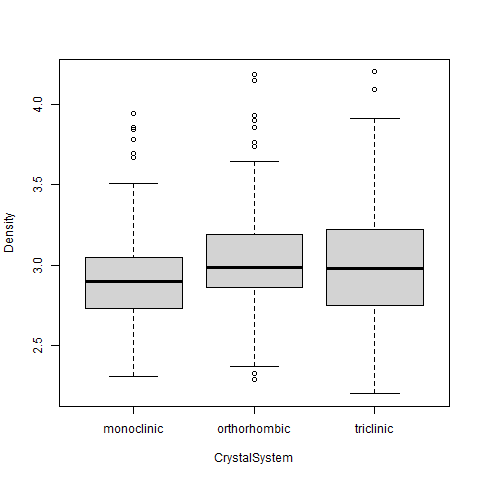


Figure 3. *Bar plots showing difference of monoclinic crystal system density from other crystal systems.*

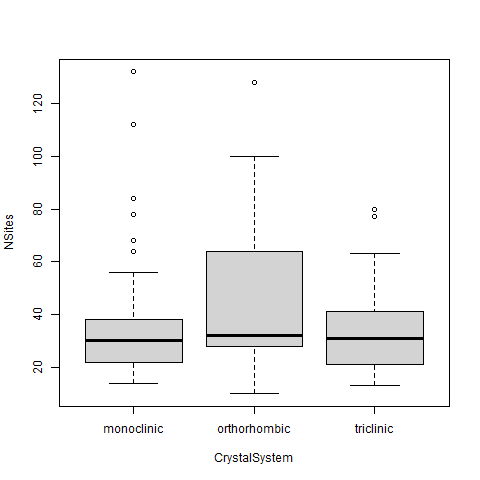


Figure 4. *No evident difference in median of number of atoms per unit cell for the crystal systems.*

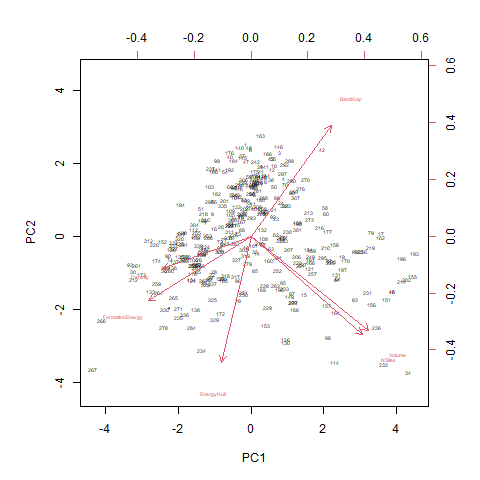


Figure 5. *Biplot of first two principal components of the numeric variables in the data set. Principal components are uncorrelated, linear combinations of variables in a data set. Principal component analysis is a dimension reduction technique that can be used to visualize high dimensional data. Right and top axes are principal component loadings, bottom and left axes are scaled loadings. Data points are plotted as numbers; 1 being the first data point. Red axes represent the numeric variables in the data, represented as vectors. Nearly 70% of the variation in the data is explained by the first two principal components.*