3.4

a)



Iris Data scatter splot matrix

Scatter Plot Matrix
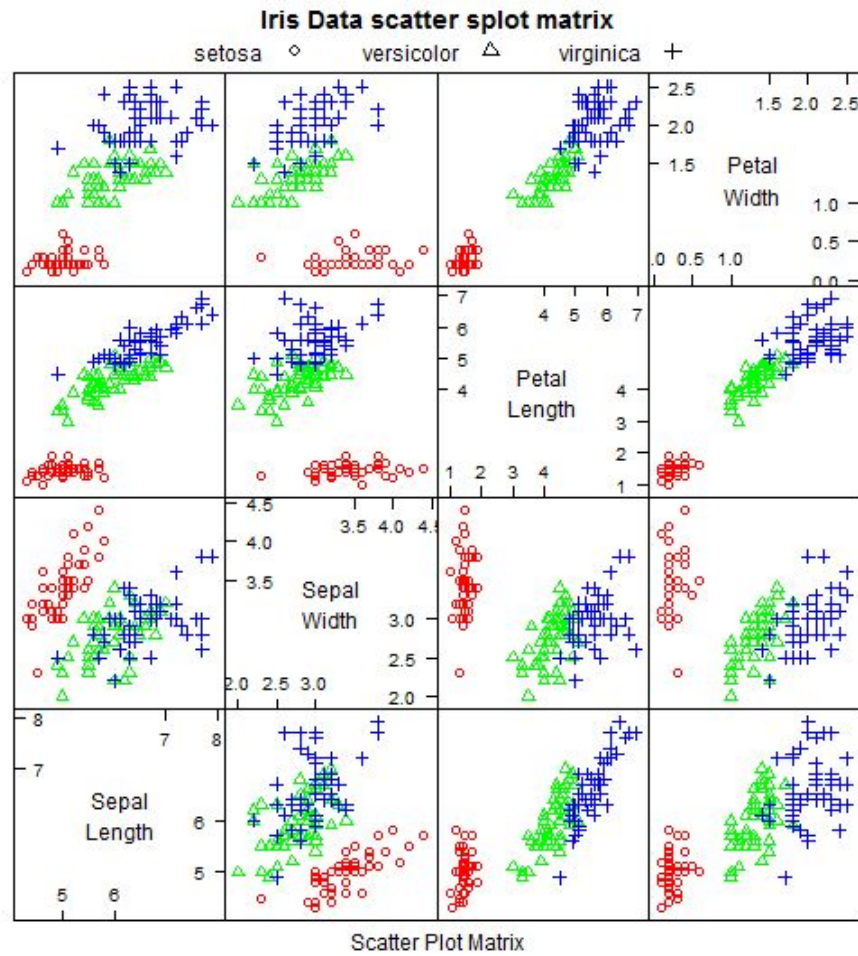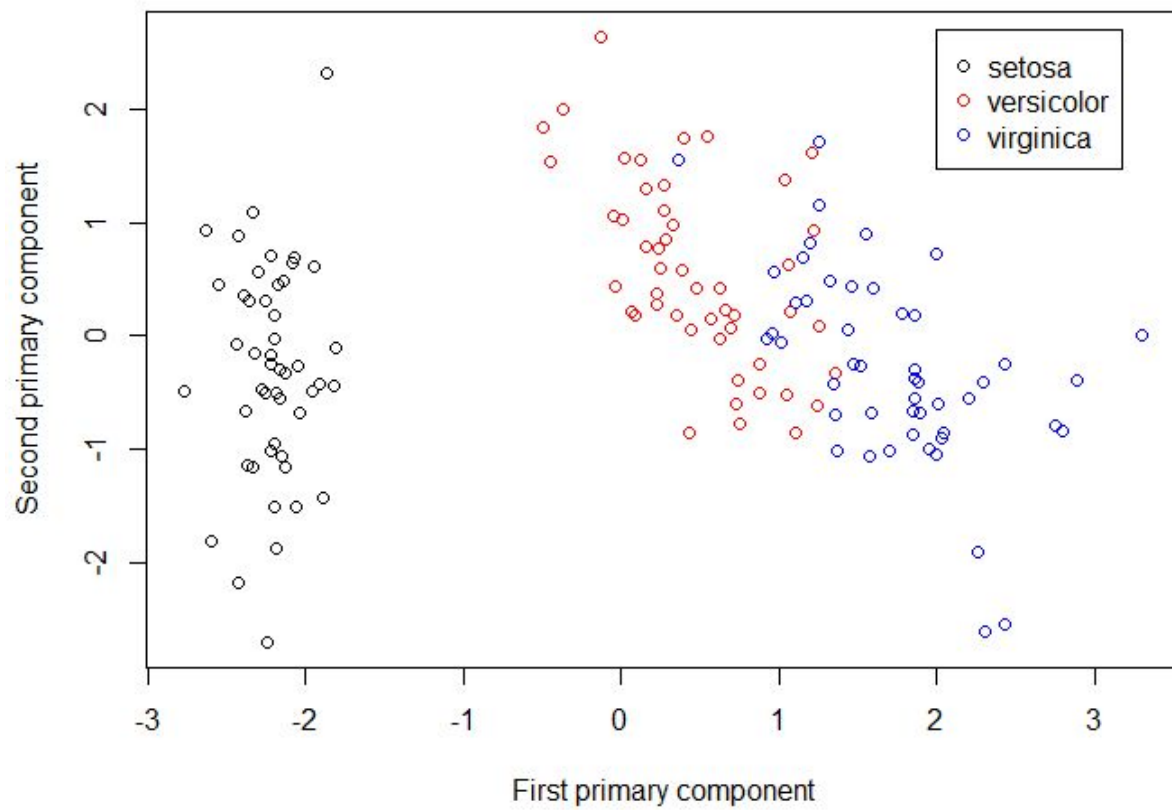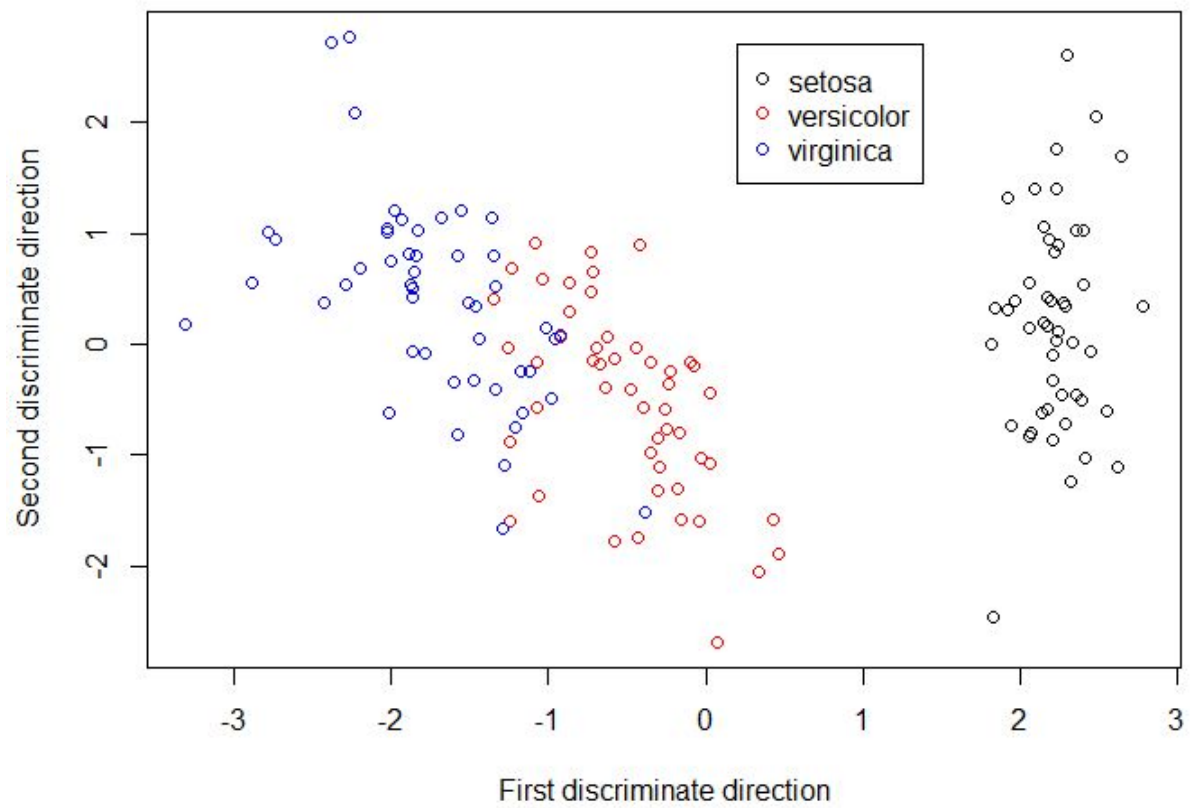
b) The plot has been rotated and the two clusters squeezed to become level with each other. It is very easy to separate the three classes with just the two principal component dimensions.
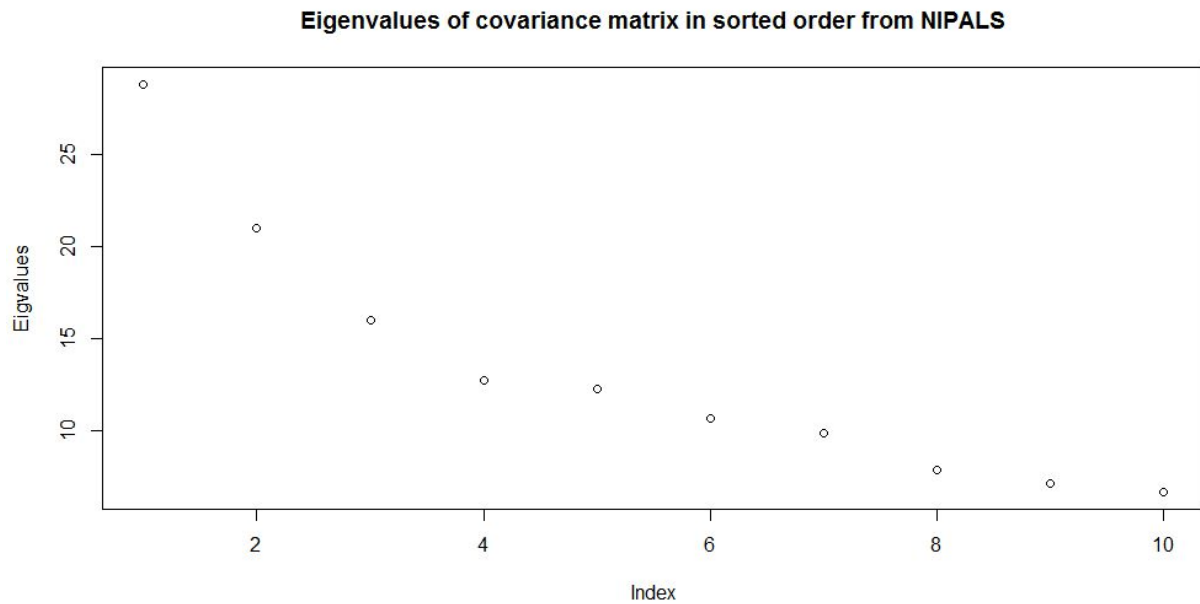
**Iris data plotted on two primary components**



c) The PSL result is similar to the PCA allowing the three clusters to be easily differentiated. While the setosa cluster is very separate and distinct, the versicolor and virginica are more difficult to speciate.

**Iris data plotted on Two discriminate directions**

setosa
versicolor
virginica

Second discriminate direction

First discriminate direction

3.5

 a)

**Eigenvalues of covariance matrix in sorted order from NIPALS**



Judging from comparing "eigenvalues.png" to Figure 3.34 in the notes, 4 components should be used to represent the dataset. The difference between the ith and jth eigenvalues with i from 1 to 3, and j from 2 to 4, respectively, corresponding to their respective principal components, is noticeably greater than the difference between the 4th and 5th eigenvalues.
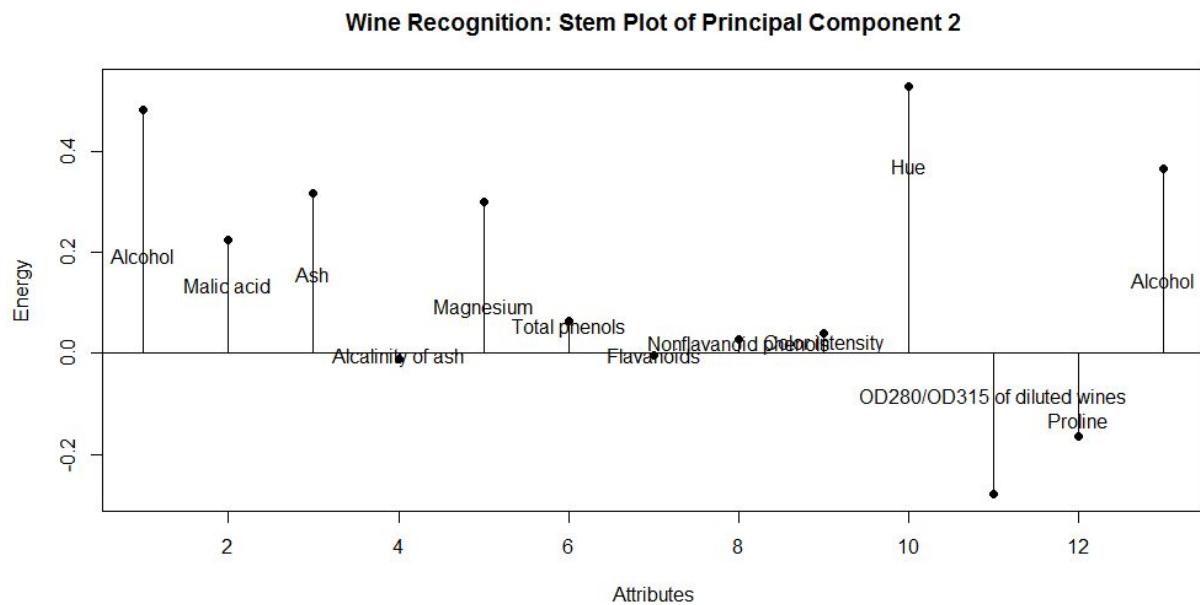
b)

    i)

**Wine Recognition: Stem Plot of Principal Component 1**
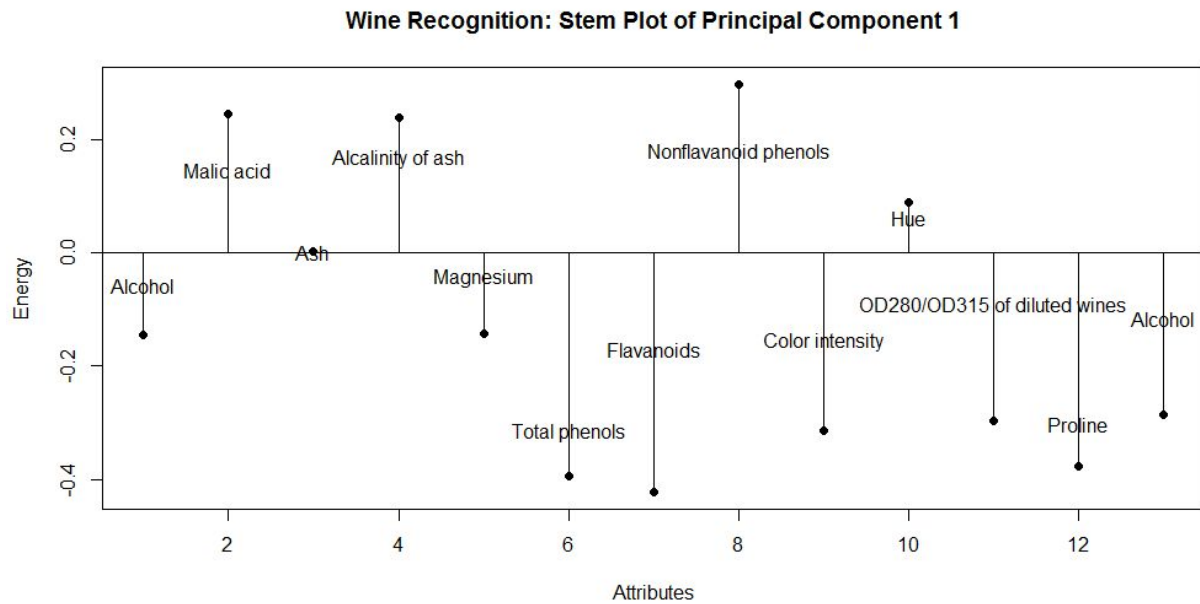


In "principal_component_1.png", ash has nearly no leverage, alcohol, magnesium, and hue have some leverage, and the remaining features have significant leverage.

    ii)

**Wine Recognition: Stem Plot of Principal Component 2**



In "principal_component_2.png", alcalinity of ash, flavanoids, nonflavanois phenols, and color intensity have little leverage, proline has some leverage, and the remaining features have significant leverage.

iii)

**Wine Recognition: Stem Plot of Principal Component 3**
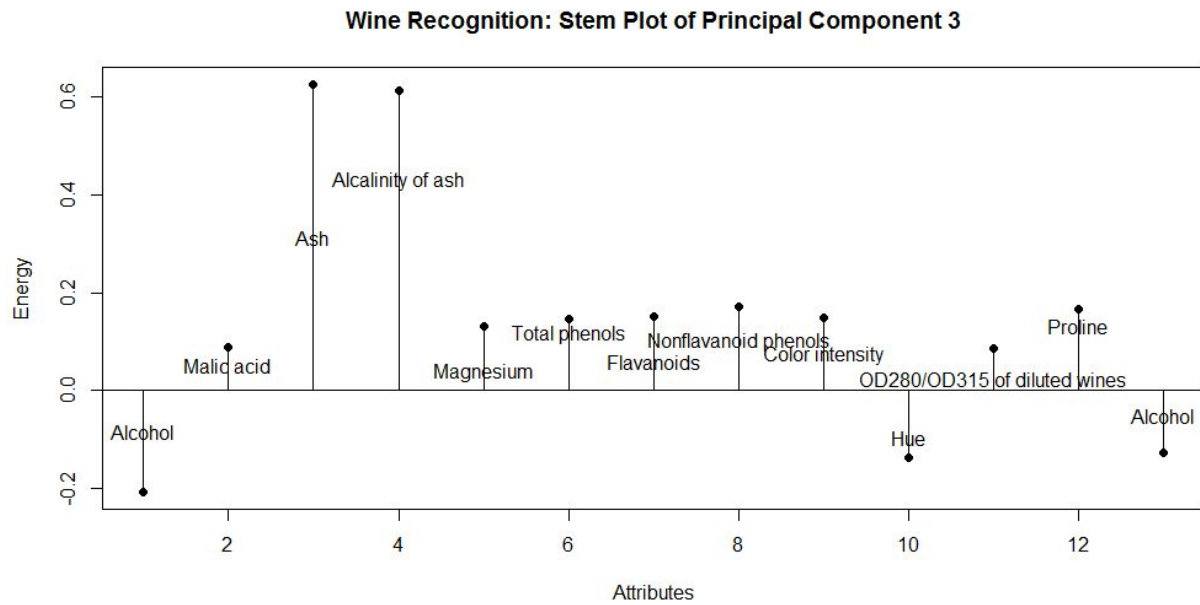


In "principal_component_3.png", ash and alcalinity of ash have significant leverage, while the remaining features have some leverage.
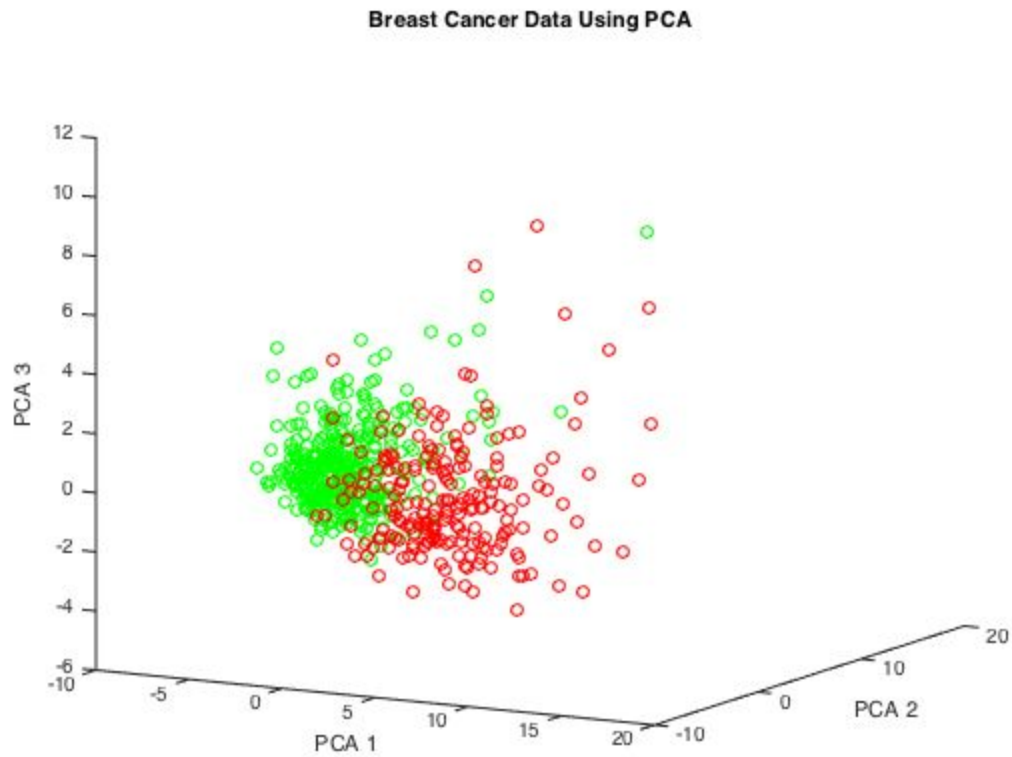
c)

**Wine Recognition plotted on two principal components**
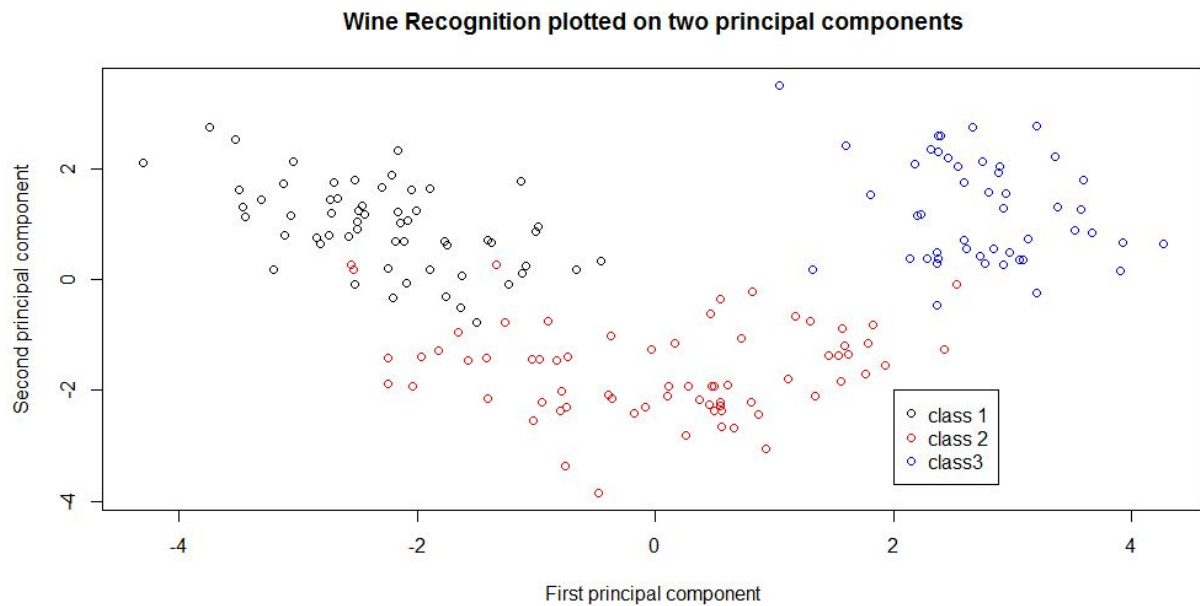


See "two_principal_components.png"

3.7

a)



Breast Cancer Data Using PCA

Plotting the first principal component against the second principal component yields the most defined clustering of benign and malignant data classes. There is a distinct separation between the two classes. Plotting the first against the third principal component yields a somewhat defined clustering of benign and malignant data classes. Plotting the second against the third principal component yields the least defined results as the benign data overlaps the malignant data.

b)

**Breast Cancer Data Using PLS1**
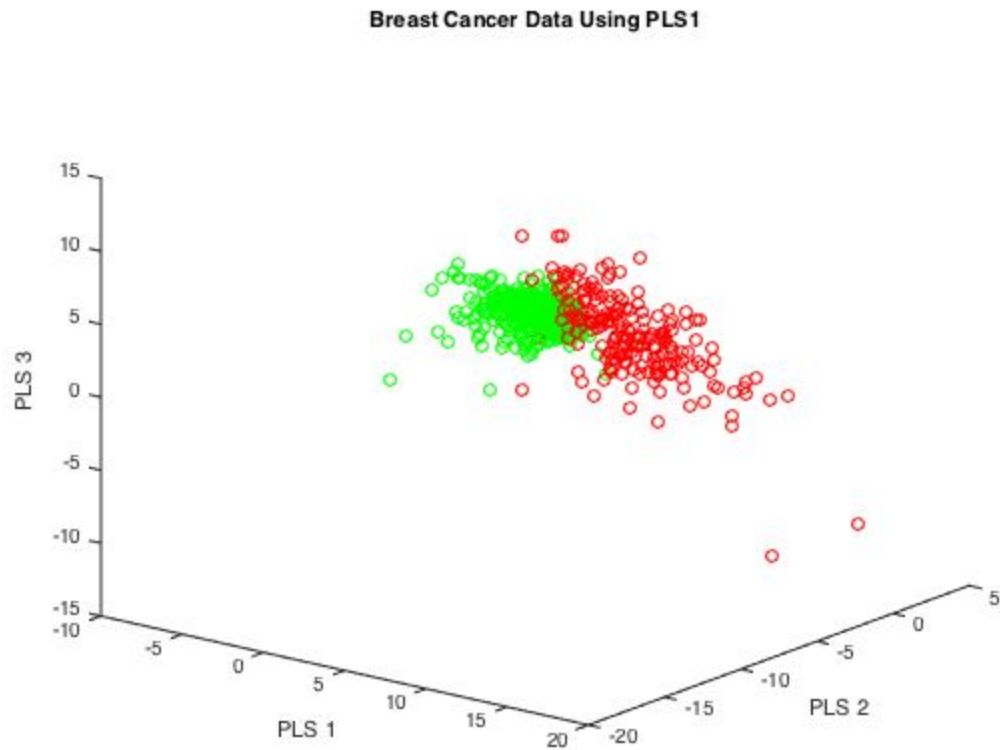


Using PLS1 yields a plot with a denser clustering of benign and malignant data classes compared to using PCA. Thus, PLS1 does a better job of performing dimension reduction.