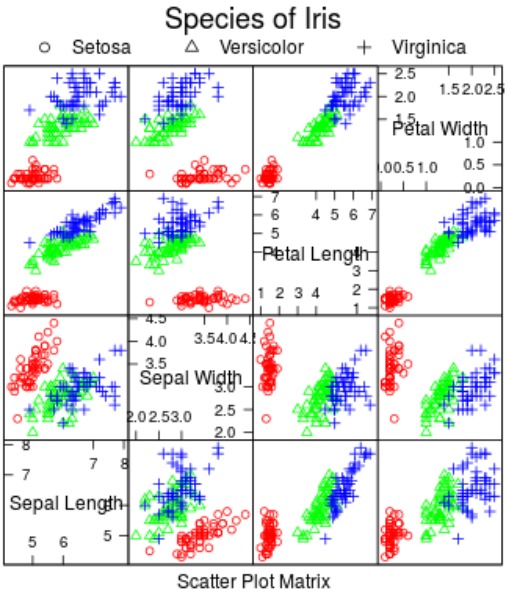
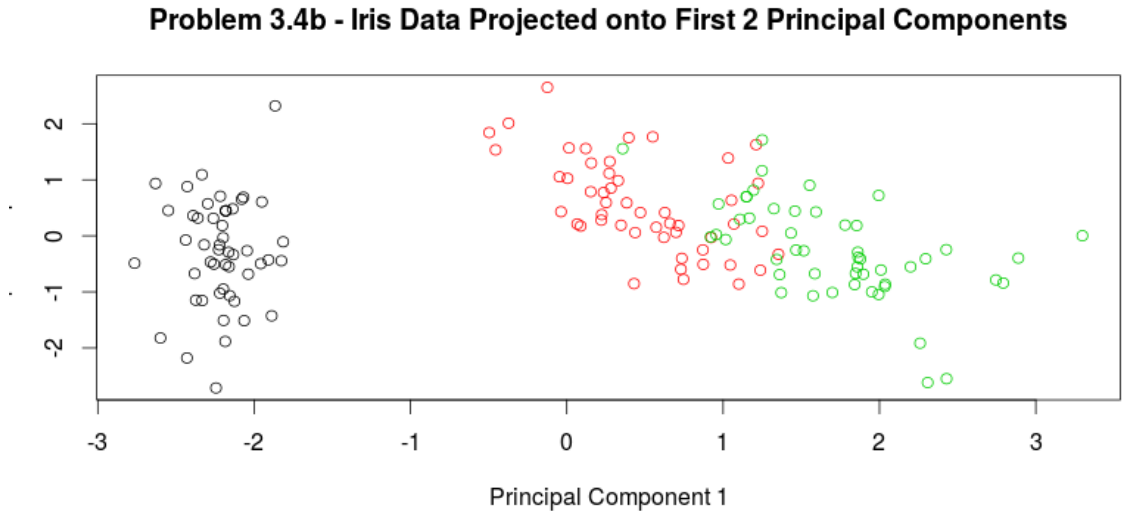
Neeraj Asthana
nasthan2
2/22/2016
CS 498 Homework 3 Report

All principal components and projections were retrieved using the prcomp function included in R.
All PLR1 direction components were retrieved using the plsreg2 function inside the plsdepot library.
All 3dplots were producing using the scatterplot3d library.

Problem 3.4a:



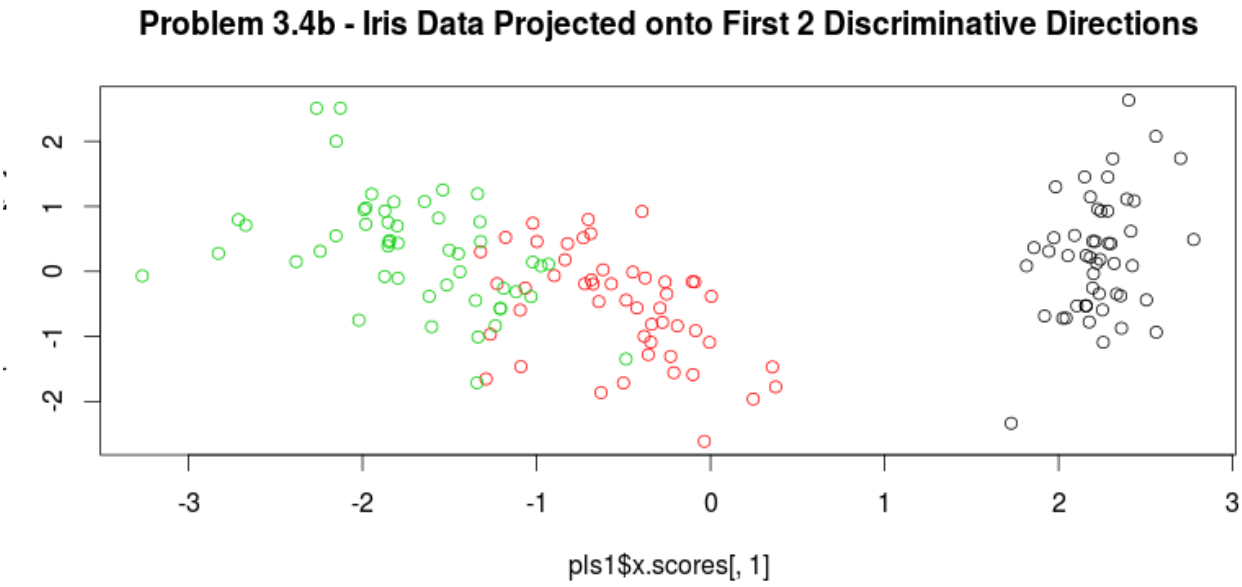
Problem 3.4b:



The plot of the first 2 Principal Components has not introduced much distortions to the iris

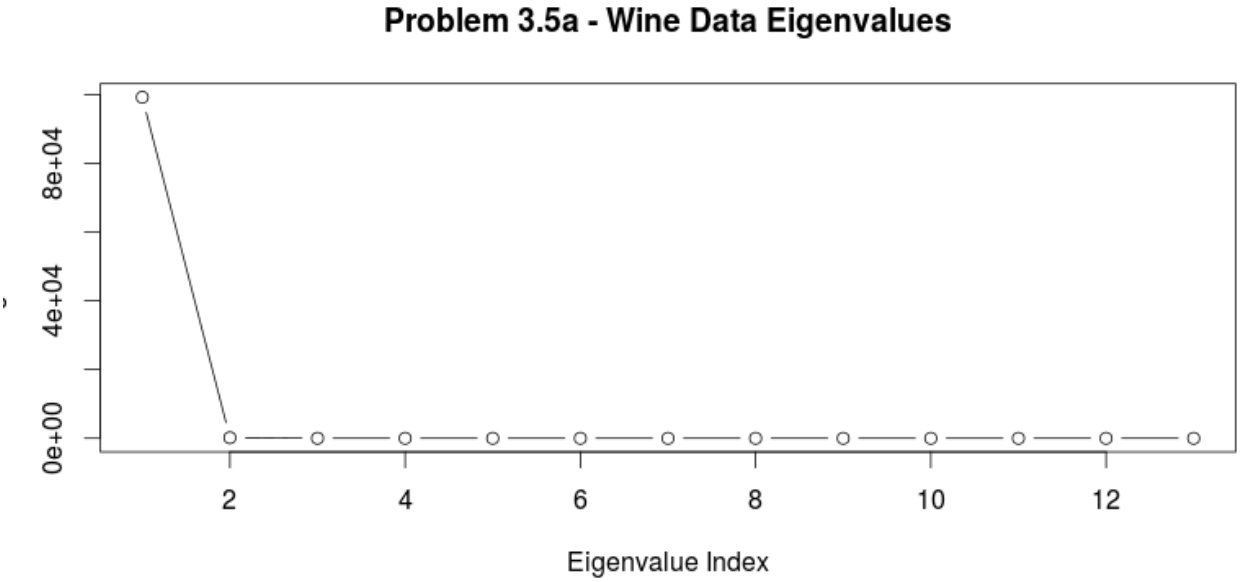
data set. The Setosa dataset is still completely separated from the other two types of irises and the Versicolor and Virginica data still slightly overlap but are somewhat separated.

Problem 3.4c:



The plot of the first 2 Discriminative Directions of the Iris dataset does not look much different from the plot of the first 2 Principal Components (just flipped). The Setosa dataset is still completely separated from the other two types of irises and the Versicolor and Virginica data still slightly overlap but are somewhat separated.

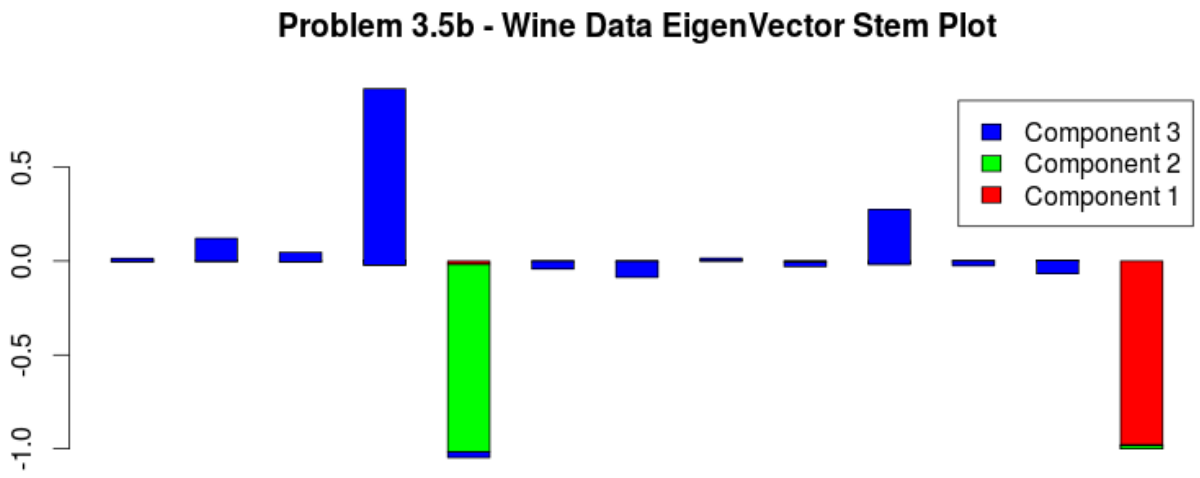
Problem 3.5a:



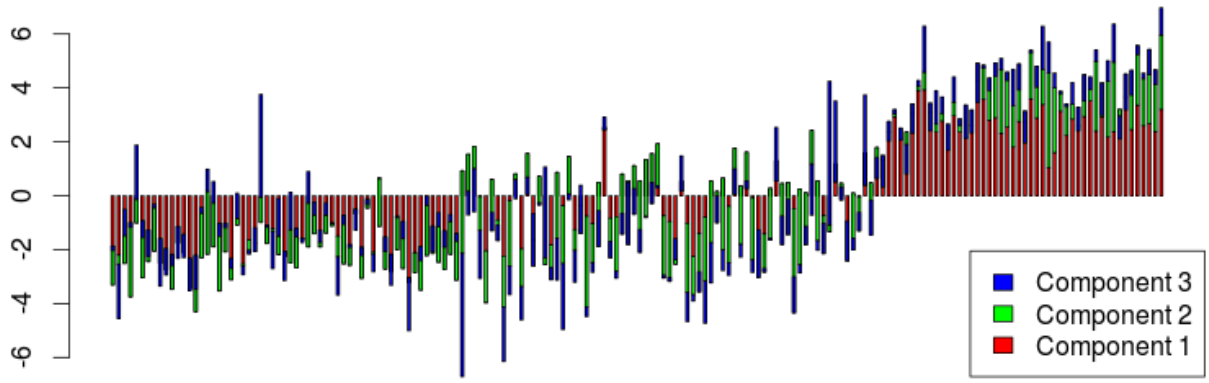
Eigenvalues: `9.920179e+04 1.725353e+02 9.438114e+00 4.991179e+00 1.228845e+00 8.410639e-01 2.789735e-01 1.513813e-01 1.120968e-01 7.170260e-02 3.757598e-02 2.107237e-02 8.203703e-03`

It appears as if the first eigenvalue is very large compared to the other eigenvalues. I would use the first 3 principal components to represent this dataset as the corresponding eigenvalues for these components are very close to or greater than 10 (much larger than the other values).

Problem 3.5b:

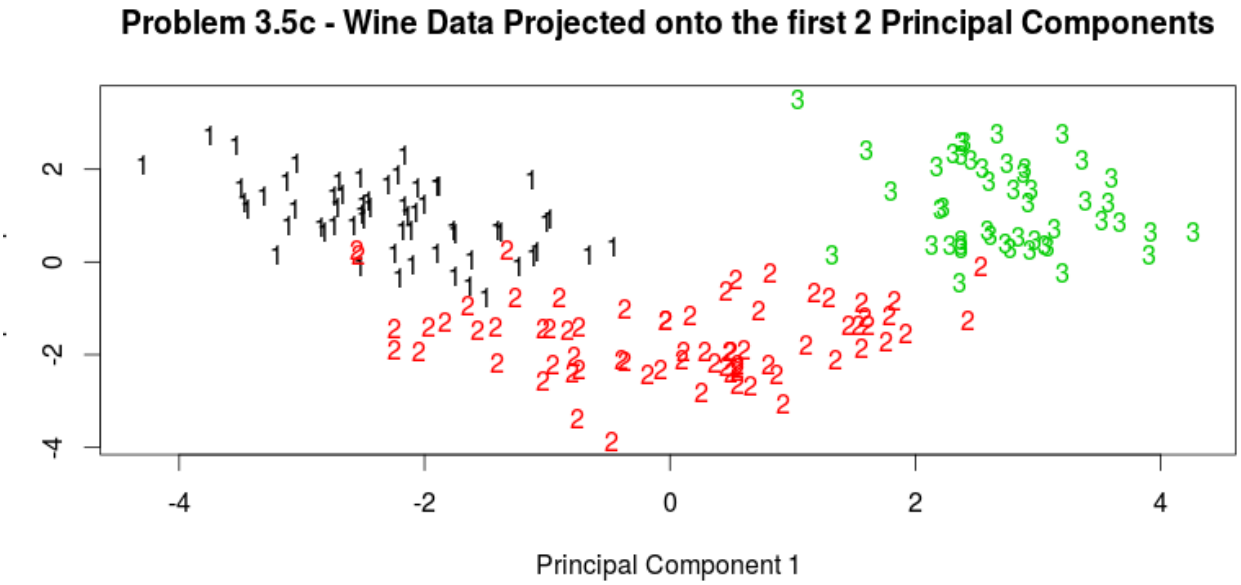


This stemplot (above) shows the first 3 principal components and there contributions to the 13 different features.

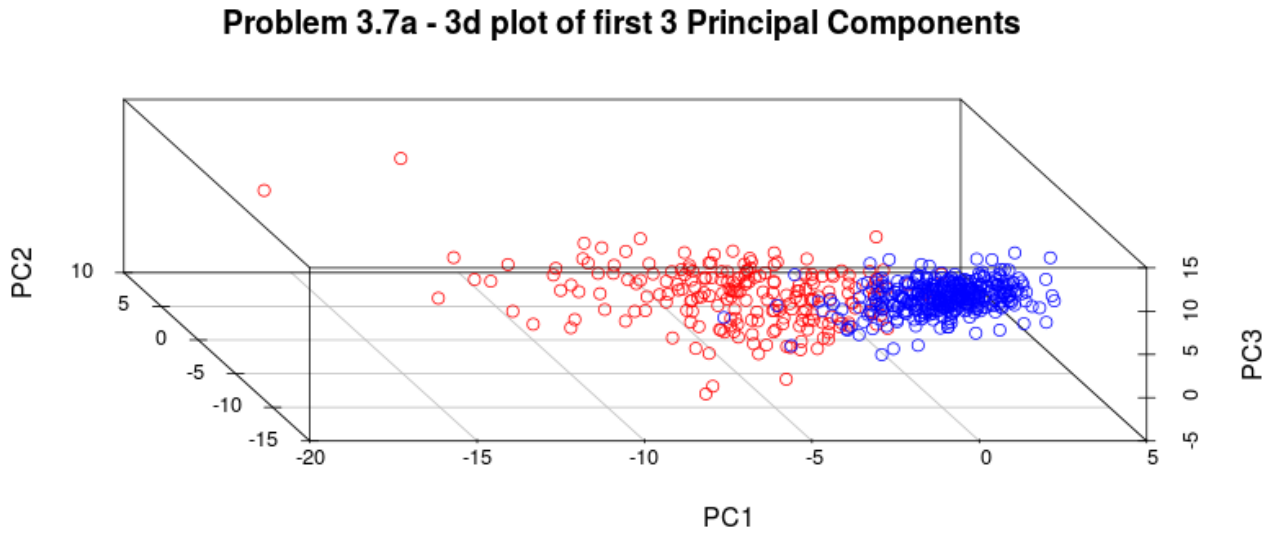


Although the question did not specifically ask for this stem plot, I created a stem plot which demonstrates the contribution of each principal component on each of the data items as it seems more informative than the previous stem plot. From this stem plot it appears that component one contributes to most of the build up of a single observation in the data, followed by the second component and the third component respectively.

Problem 3.5c:

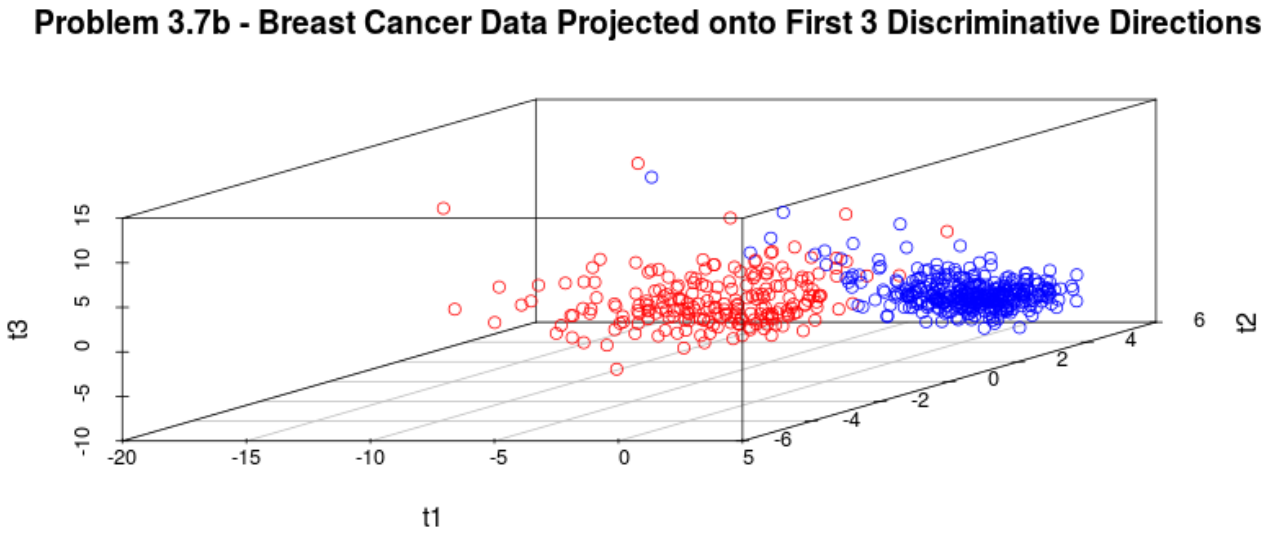


Problem 3.7a:



Legend: Red – Malignant, Blue – Benign
After plotting the first 3 principal components of the breast cancer data set, it appears as though the malignant and the benign cases are pretty separable, however there are a few example that overlap significantly.

Problem 3.7b:



Legend: Red – Malignant, Blue – Benign
After plotting the first 3 discriminative directions of the breast cancer data set, it appears as though the malignant and the benign cases are pretty separable, however there are a few example that overlap. The discriminative directions appear to separate a little better than the principal components plot.