MAST90138 Week 4 Lab

Problems:
The iris data contain various measurements (sepal length, sepal width, petal length and petal width) of 50 flowers from each of 3 species of iris flowers. Type `help(iris)` to learn about the formal of these data.

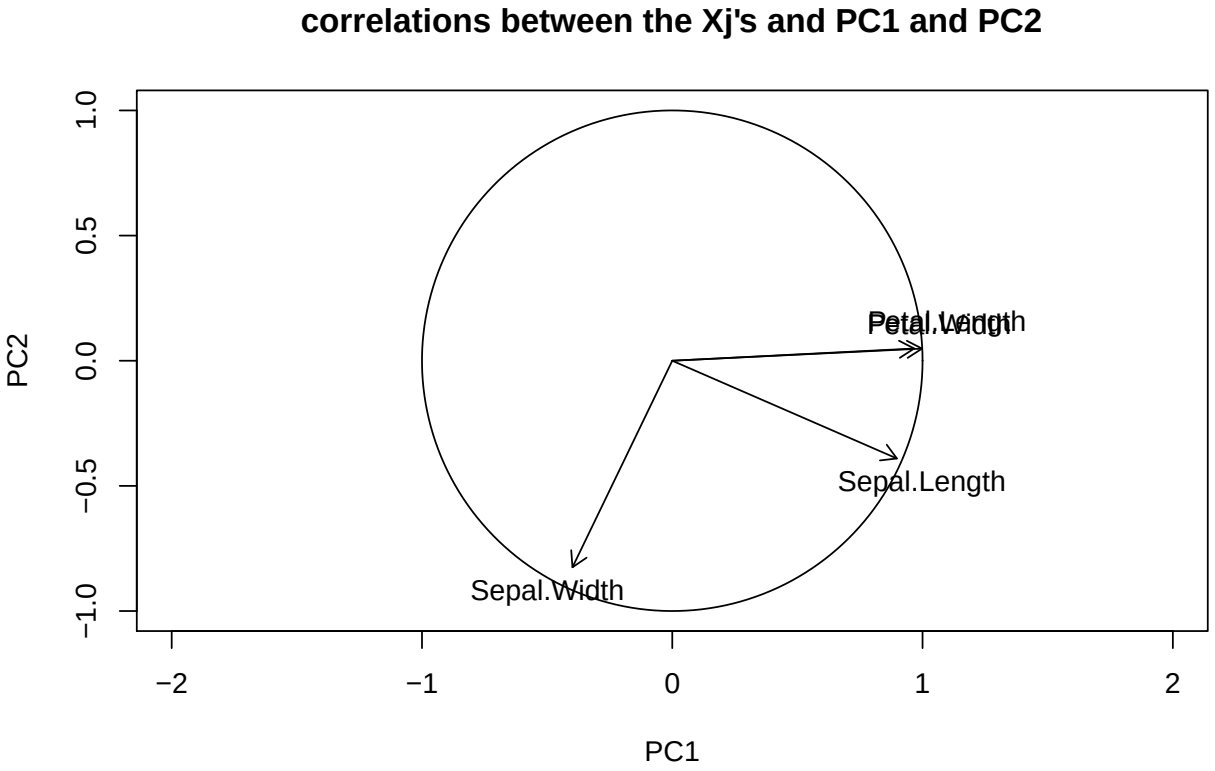
1. Load the `iris` data in R (they are already in R).

2. Do a PC analysis of these data using only the numerical variables, this time using the prcomp command. Using the output of this function, store the eigenvectors of the covariance matrix S in a matrix and the eigenvalues in a vector. Also store the Y_{ik}'s in a matrix Y, again using the output of prcomp.

3. Draw a screeplot for these data and recall the ψ_j's (cumulative proportion of variance explained by each component) from last week. How many component does this suggest you should keep?

4. What is the weight of each original variable in the linear combination use to create PC1 and PC2? Which variables are the most correlated with each PC (describe PC by PC and support your answer by some calculations)?

5. The correlation graph showing the correlation between each of the original variable and two PCs is given below, where we also plot PC1 versus PC2. We also provide a table with the values of the correlations between each original variable and each PC. Use these graphs and this table to provide more insight into the results of the analysis.



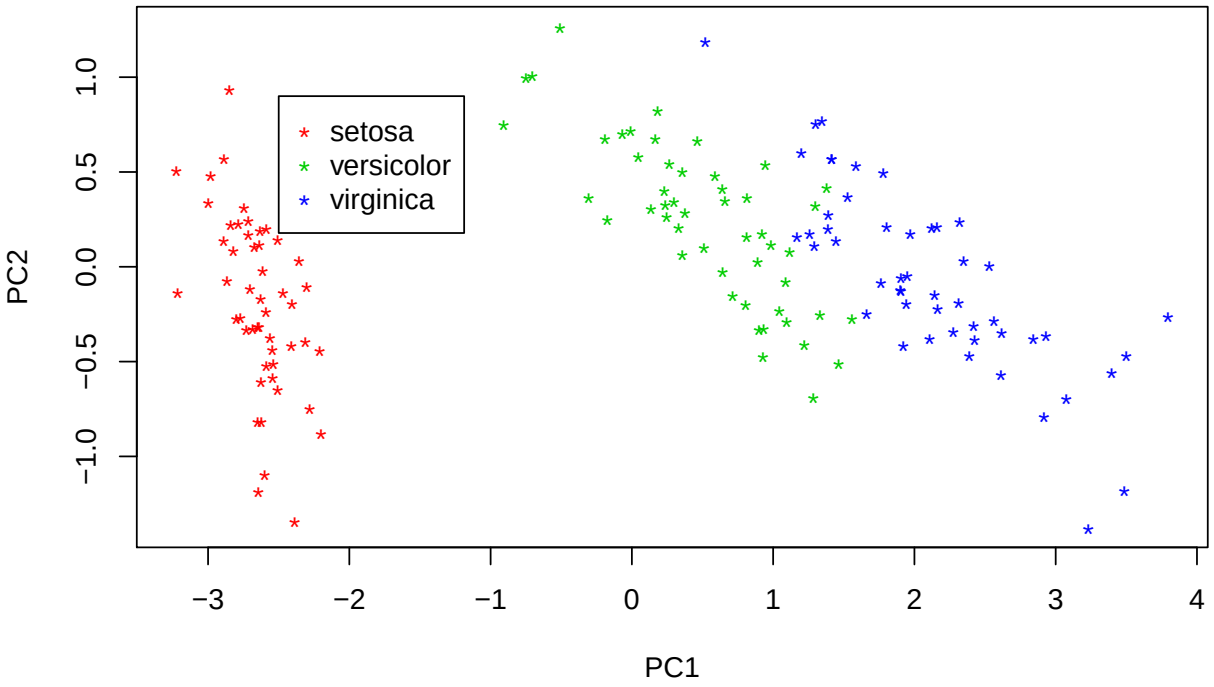


Table 1: Correlations between original variables and the principal components:

	PC1	PC2	PC3	PC4
Sepal length	0.8974018	-0.3906044	0.19656672	0.05882002
Sepal width	-0.3987485	-0.8252287	-0.38363030	-0.11324764
Petal length	0.9978739	0.0483806	-0.01207737	-0.04196487
Petal width	0.9665475	0.0487816	-0.20026170	0.15264831

6. Now compute the variances of the original variables. What does it suggest for the PC analysis done above and what can we consider doing instead (do it)?