VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

Lab exercise-4

Code/Course	**:**	**CSE3020 – Data Visualisation**	**Date**	**:**	**09/02/2022**
Lab Experiments		**PCA and LDA Analysis**	**Slot**	**:**	**L15+L16**

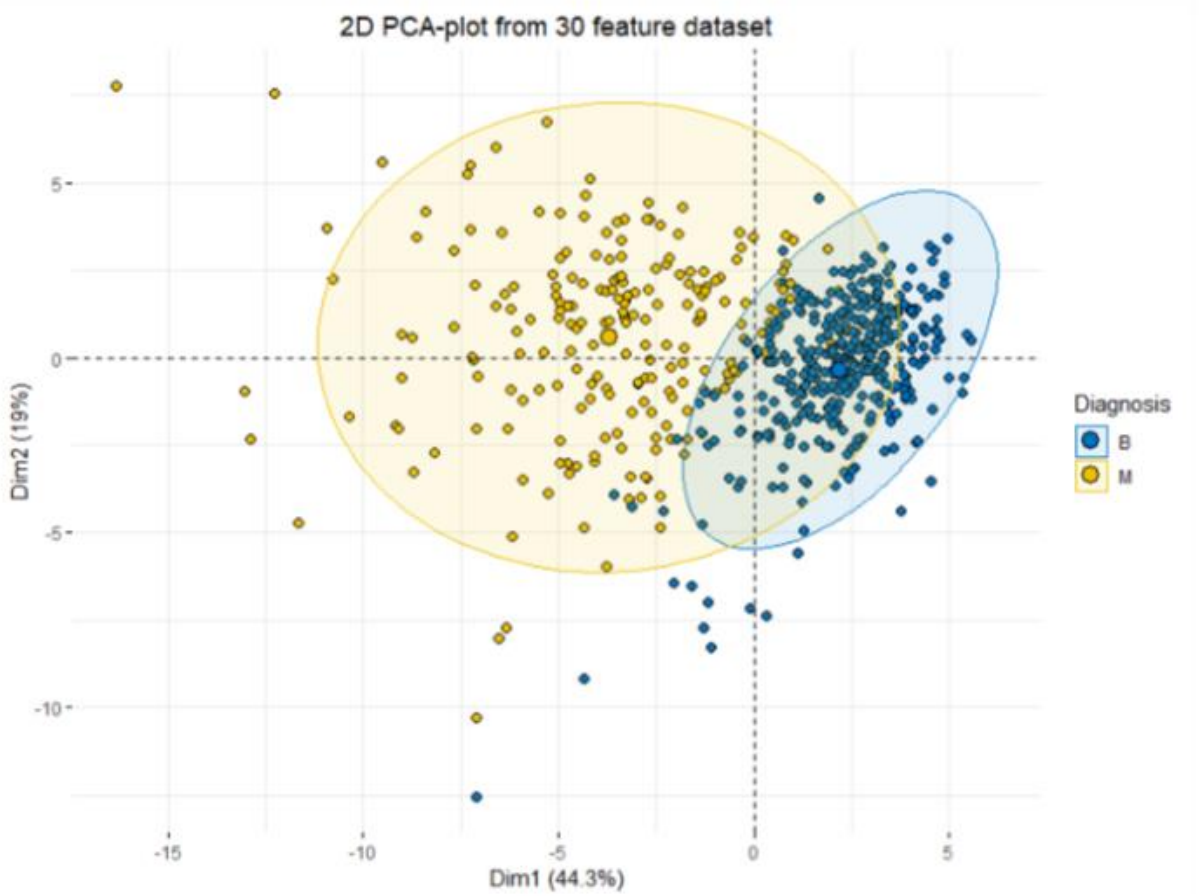
Pre-requisite: Moderately familiar with basic concepts in R, including variables and functions, and with Studio, the integrated development environment for programming in R.

1. Use Breast Cancer Wisconsin data set from the *UCI Machine learning repo* to plot the PCA analysis. Use the '*prcomp*' function runs PCA on the data.

i. You want to explain difference between **malignant** and **benign** tumors using Visualisation and add the **response variable** (*diagnosis*) to the plot
ii. Construct some kind of model using the first 6 principal components to predict whether a tumor is benign or malignant and then compare it to a model using the original 30 variables.

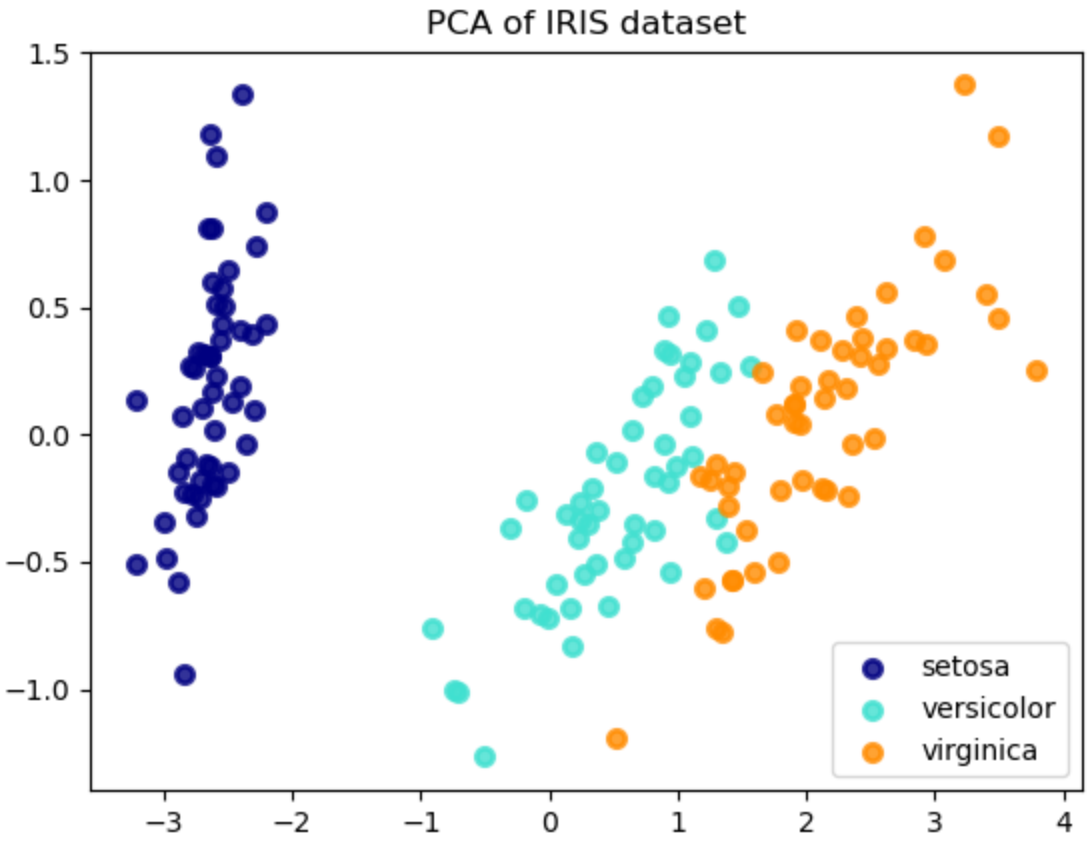
Use the below given packages to improve your results

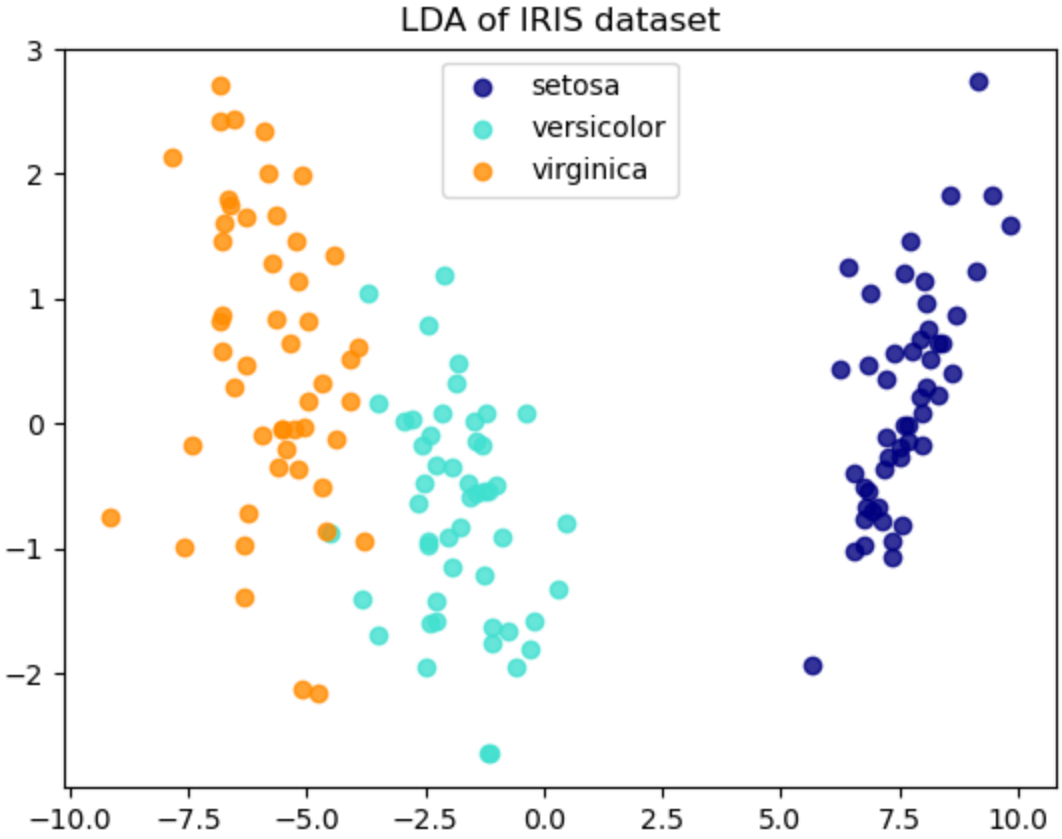
```
library(devtools)
install_github("vqv/ggbiplot")
```



2. Use the built-in **iris** dataset in R to plot the LDA analysis. Use the *lda* function of the *MASS* package in R

Project the LDA visual output and Compare the LDA and PCA 2D Projection of Iris dataset


PCA of IRIS dataset
1.5
1.0
0.5
0.0
−0.5
−1.0
−3
−2
−1
0
1
2
3
4
setosa
versicolor
virginica


LDA of IRIS dataset
setosa
versicolor
virginica
3
2
1
0
−1
−2
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0