<http://afodor.github.io/classes/stats2015/prePostPhylum.txt>

1. Download the dataset. Perform PCA ordination.

(For example:

rm(list=ls())

setwd("C:\\Users\\afodor\\git\\afodor.github.io\\classes\\stats2015\\")

inFileName <- paste("prePostPhylum.txt", sep ="")

myT <-read.table(inFileName,header=TRUE,sep="\t")

numCols <- ncol(myT)

myColClasses <- c(rep("character",4), rep("numeric", numCols-4))

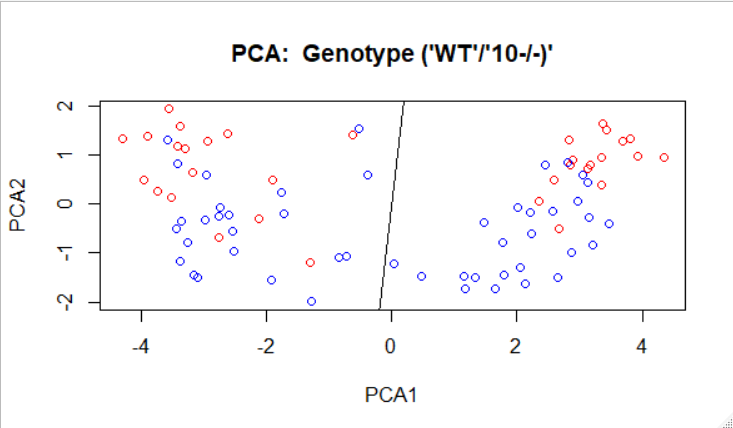
myT <-read.table(inFileName,header=TRUE,sep="\t",colClasses=myColClasses)

myTData<-myT[,5:10]

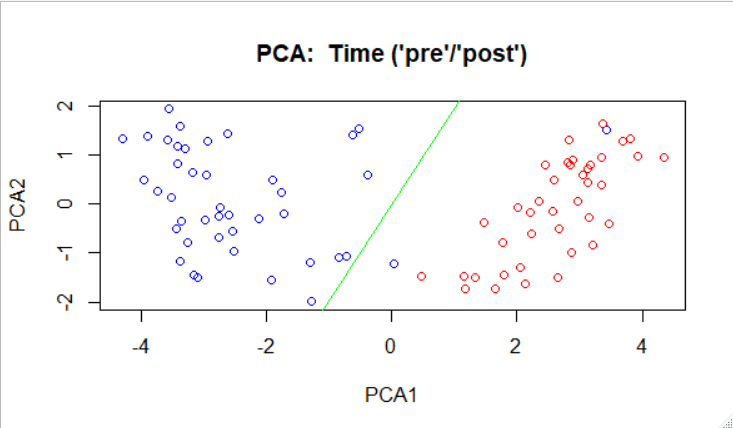
myPCOA <- princomp(myTData)

1. Graph PCA1 vs. PCA2. Make three versions of the graph. One colored by genotype,

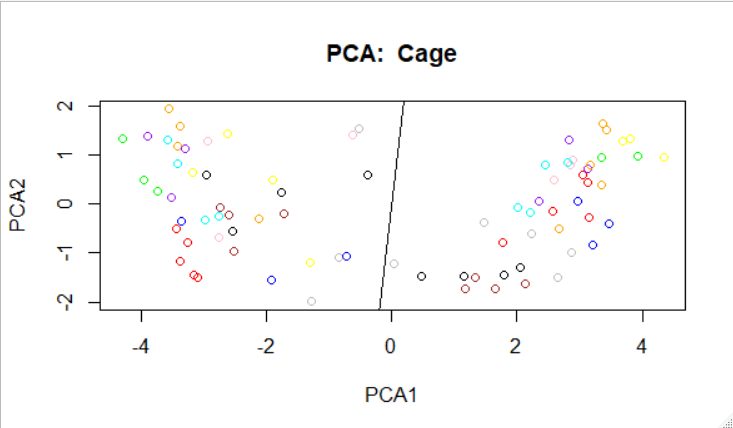
one colored by cage and one colored by timepoint (pre-vs-post)



**Figure 1: Genotype Plot**



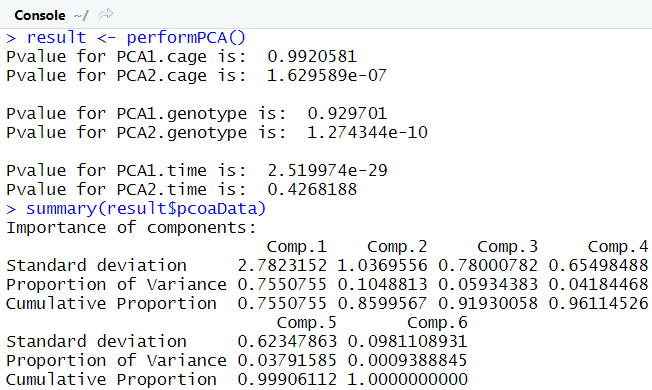
**Figure 2: Time Plot**



**Figure 3: Cage Plot**

1. Fill in the following table for p-values testing the null hypothesis for PCA 1 and 2. For cage, use a way one-ANOVA. For genotype and timepoint (“pre” vs “post”) use a t-test

|  |  |  |
| --- | --- | --- |
|  | PCA1 | PCA2 |
| Cage | 0.9920581 | 1.629589e-07 |
| Genotype | 0.929701 | 1.274344e-10 |
| Time (pre vs. post) | 2.519974e-29 | 0.4268188 |



Which variable seems to be most associated with the first PCA axis? Which variable is most associated with the second PCA axis? Does cage seem to be having an effect on these data?