Erika Maldonado-Rosado

20 April 2020

Lab 11

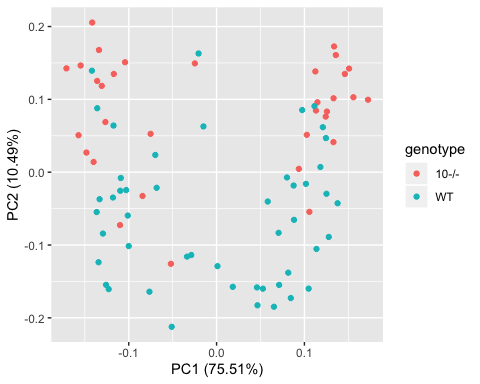
#1  
setwd("~/Desktop")  
inFileName <- paste("prePostPhylum.txt", sep ="")  
  
myT <-read.table(inFileName,header=TRUE,sep="\t")  
numCols <- ncol(myT)  
myColClasses <- c(rep("character",4), rep("numeric", numCols-4))  
myT <-read.table(inFileName,header=TRUE,sep="\t",colClasses=myColClasses)  
  
myTData<-myT[,5:10]  
  
myPCOA <- prcomp(myTData)  
  
summary(myPCOA)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 2.7997 1.0434 0.78487 0.65907 0.62736 0.09872  
## Proportion of Variance 0.7551 0.1049 0.05934 0.04184 0.03792 0.00094  
## Cumulative Proportion 0.7551 0.8600 0.91930 0.96115 0.99906 1.00000

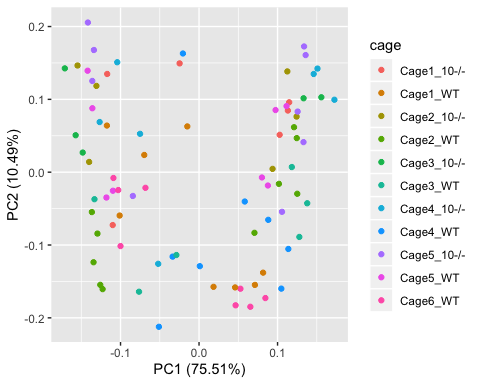
#2  
library(ggfortify)

## Loading required package: ggplot2

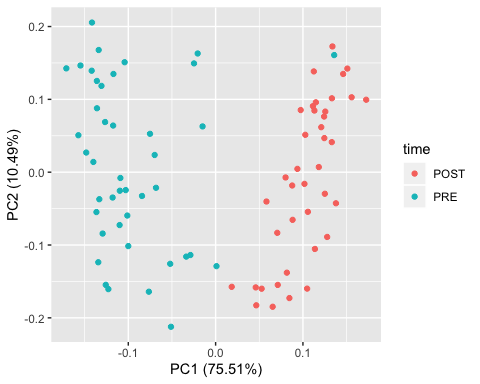
autoplot(myPCOA, data = myT, colour = 'genotype')



autoplot(myPCOA, data = myT, colour = 'cage')



autoplot(myPCOA, data = myT, colour = 'time')



PC\_1 <- myPCOA$x[,1]  
PC\_2 <- myPCOA$x[,2]  
#If I use scores it is only when I use princomp. I prefer prcomp because it shows PC1 instead of comp. 1  
#PC\_1 <- myPCOA$scores[,1]  
#PC\_2 <- myPCOA$scores[,2]

#3  
#cage  
summary(aov(PC\_1 ~ myT$cage, data = myT))

## Df Sum Sq Mean Sq F value Pr(>F)  
## myT$cage 10 20.2 2.015 0.232 0.992  
## Residuals 70 606.9 8.670

summary(aov(PC\_2 ~ myT$cage, data = myT))

## Df Sum Sq Mean Sq F value Pr(>F)   
## myT$cage 10 43.59 4.359 7.013 1.63e-07 \*\*\*  
## Residuals 70 43.51 0.622   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#genotype  
t.test(PC\_1 ~ myT$genotype)

##   
## Welch Two Sample t-test  
##   
## data: PC\_1 by myT$genotype  
## t = 0.088614, df = 56.616, p-value = 0.9297  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.281180 1.399803  
## sample estimates:  
## mean in group 10-/- mean in group WT   
## 0.03514763 -0.02416400

t.test(PC\_2 ~ myT$genotype)

##   
## Welch Two Sample t-test  
##   
## data: PC\_2 by myT$genotype  
## t = 7.4362, df = 76.441, p-value = 1.274e-10  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.9701189 1.6797915  
## sample estimates:  
## mean in group 10-/- mean in group WT   
## 0.7851586 -0.5397966

#time  
t.test(PC\_1 ~ myT$time)

##   
## Welch Two Sample t-test  
##   
## data: PC\_1 by myT$time  
## t = 19.454, df = 67.256, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 4.530785 5.566732  
## sample estimates:  
## mean in group POST mean in group PRE   
## 2.617875 -2.430884

t.test(PC\_2 ~ myT$time)

##   
## Welch Two Sample t-test  
##   
## data: PC\_2 by myT$time  
## t = -0.79877, df = 78.935, p-value = 0.4268  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.6475421 0.2766614  
## sample estimates:  
## mean in group POST mean in group PRE   
## -0.09615426 0.08928610

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 | .992 | 1.63e-07 |
| Genotype | .9297 | 1.274e-10 |
| Time (pre vs. post) | 2.2e-16 | 0.4268 |

**Which variable seems to be most associated with the first PCA axis?**

I believe time is most associated with PCA1

**Which variable is most associated with the second PCA axis?**

I believe that genotype is most associated with PCA2

**Does cage seem to be having an effect on these data?**

I believe that cage does not seem to have an effect on the data.