Comparisons\_new

library(dplyr)

library(tibble)

library(ggplot2)

library(ggpubr)

metadata <- read.csv("C:/Users/jguitart/OneDrive/DOCTORAT-IRTA/AA\_PROJECTE/METAGENOMICS/Diversity\_Abundance\_analysis/microeco\_new/metadata\_table.csv", sep = ";", header = T, check.names = FALSE, row.names = 1)  
index <- read.csv("C:/Users/jguitart/OneDrive/DOCTORAT-IRTA/AA\_PROJECTE/METAGENOMICS/Diversity\_Abundance\_analysis/microeco\_new/alpha\_diversity\_new.csv", header = TRUE, row.names = 1, dec = ",")  
index <- index %>% rownames\_to\_column(var = "id")  
head(index)

## id Observed Chao1 se.chao1 ACE  
## 1 3052\_GG\_V1 19350 20246.3784431138 66.9734302958306 19957.4218044669  
## 2 3052\_GG\_V2 19752 20401.6229007634 55.8907434464514 20211.0093318185  
## 3 3052\_GG\_V3 14793 16901.9034090909 98.8912010317053 16918.3061940011  
## 4 3052\_GG\_V4 11460 13966.2577377603 120.245882785372 14118.7873674996  
## 5 3053\_GG\_V1 13524 15993.5316455696 121.107013456299 15847.5213436566  
## 6 3053\_GG\_V2 19905 20675.4577656676 61.7396178019291 20414.625332725  
## se.ACE Shannon Simpson InvSimpson  
## 1 64.0871428552733 3.7733962188044 0.936988933091034 15.8702280258916  
## 2 59.2104179591465 4.18286266280225 0.944713560680049 18.087618090448  
## 3 61.5767267724221 3.74036962962848 0.938182881295513 16.1767487866984  
## 4 59.4255867365707 2.90314076757783 0.90304029516951 10.3135627500956  
## 5 62.5077759153045 2.91195766975776 0.903776606280377 10.3924831721672  
## 6 62.5870090160915 3.65496230425797 0.928144845515527 13.9168860908383  
## Fisher Coverage  
## 1 1664.42621324602 0.999993429859469  
## 2 1920.98039016946 0.999983549870857  
## 3 1725.32465026907 0.99965967419637  
## 4 1172.69819932645 0.999854885255073  
## 5 1161.46950006685 0.999978375363819  
## 6 1747.16775667075 0.999993132245525

diversity <- data.frame(metadata, index)  
sapply(diversity, class)

## ID\_number Group Visit ID\_mother   
## "character" "character" "integer" "integer"   
## Parity Wgt\_i Wgt\_f Wgt\_gain.day   
## "integer" "character" "character" "character"   
## Clinic Enfermería Average\_variable id   
## "character" "integer" "character" "character"   
## Observed Chao1 se.chao1 ACE   
## "integer" "character" "character" "character"   
## se.ACE Shannon Simpson InvSimpson   
## "character" "character" "character" "character"   
## Fisher Coverage   
## "character" "character"

cols.num <- c("Parity", "Wgt\_i", "Wgt\_f", "Wgt\_gain.day", "Observed", "Chao1", "se.chao1", "ACE", "se.ACE", "Shannon", "Simpson", "InvSimpson", "Fisher", "Coverage")  
cols.fact <- c("ID\_mother", "Visit", "Group", "Clinic","Enfermería")  
diversity[cols.num] <- sapply(diversity[cols.num], as.numeric)

diversity[cols.fact] <- sapply(diversity[cols.fact], as.factor)  
G1 <- subset(diversity, Group==1)  
G2 <- subset(diversity, Group==2)  
G3 <- subset(diversity, Group==3)  
G4 <- subset(diversity, Group==4)  
G5 <- subset(diversity, Group==5)  
G6 <- subset(diversity, Group==6)  
GG <- subset(diversity, Group=='G')  
V1 <- subset(diversity, Visit=='1')  
V2 <- subset(diversity, Visit=='2')  
V3 <- subset(diversity, Visit=='3')  
V4 <- subset(diversity, Visit=='4')

# All data

##Wilcoxon pairwise t-test (non-parametric)

pairwise.wilcox.test(x = diversity$Shannon, g = diversity$Group, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: diversity$Shannon and diversity$Group   
##   
## 1 2 3 4 5 6   
## 2 0.558 - - - - -   
## 3 0.876 0.558 - - - -   
## 4 0.391 0.876 0.391 - - -   
## 5 0.876 0.558 0.876 0.391 - -   
## 6 0.072 0.421 0.086 0.558 0.072 -   
## G 0.723 1.000 0.558 0.876 0.704 0.558  
##   
## P value adjustment method: BH

pairwise.wilcox.test(x = diversity$Shannon, g = diversity$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum test with continuity correction   
## data: diversity$Shannon and diversity$Visit   
##   
## 1 2 3   
## 2 0.00073 - -   
## 3 0.00289 0.45693 -   
## 4 0.13981 0.07398 0.15075  
##   
## P value adjustment method: BH

##One-way ANOVA (non-parametric)

kruskal.test(G3$Shannon ~ G3$Visit)

## Kruskal-Wallis rank sum test  
##   
## data: G3$Shannon by G3$Visit  
## Kruskal-Wallis chi-squared = 9.6688, df = 3, p-value = 0.0216

kruskal.test(GG$Shannon ~ GG$Visit)

## Kruskal-Wallis rank sum test  
##   
## data: GG$Shannon by GG$Visit  
## Kruskal-Wallis chi-squared = 19.226, df = 3, p-value = 0.0002455

kruskal.test(V2$Shannon ~ V2$Group)

## Kruskal-Wallis rank sum test  
##   
## data: V2$Shannon by V2$Group  
## Kruskal-Wallis chi-squared = 17.785, df = 6, p-value = 0.006792

kruskal.test(V4$Shannon ~ V4$Group)

## Kruskal-Wallis rank sum test  
##   
## data: V4$Shannon by V4$Group  
## Kruskal-Wallis chi-squared = 21.262, df = 6, p-value = 0.001646

# 

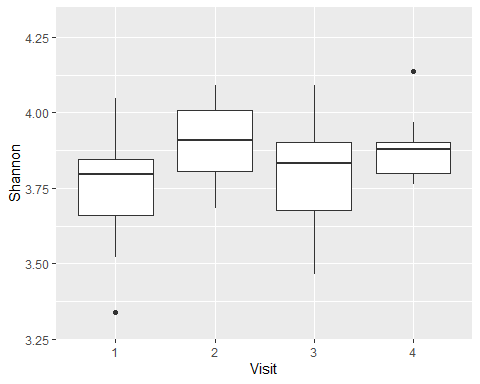
# Comparison between Visits within Groups

## Group 1

pairwise.wilcox.test(x = G1$Shannon, g= G1$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G1$Shannon and G1$Visit   
##   
## 1 2 3   
## 2 0.57 - -   
## 3 0.74 0.59 -   
## 4 0.59 0.74 0.59  
##   
## P value adjustment method: BH

ggplot(G1, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

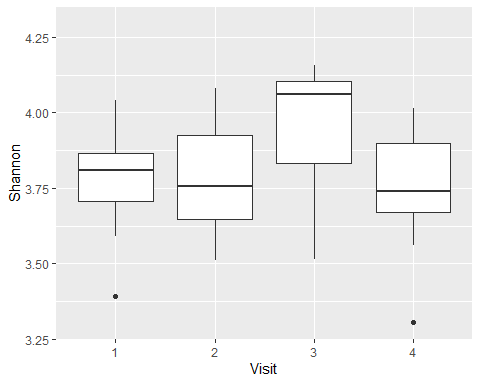


## Group 2

pairwise.wilcox.test(x = G2$Shannon, g= G2$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G2$Shannon and G2$Visit   
##   
## 1 2 3   
## 2 0.80 - -   
## 3 0.33 0.33 -   
## 4 0.80 0.80 0.33  
##   
## P value adjustment method: BH

ggplot(G2, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)



## 

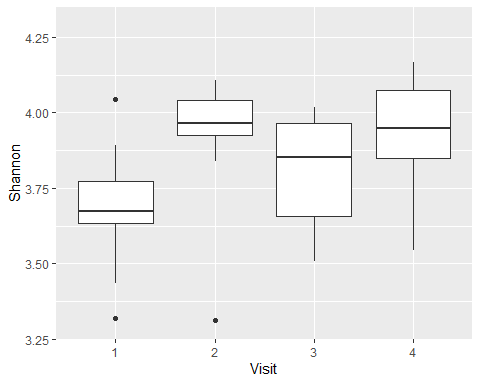
## Group 3

pairwise.wilcox.test(x = G3$Shannon, g= G3$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G3$Shannon and G3$Visit   
##   
## 1 2 3   
## 2 0.054 - -   
## 3 0.199 0.185 -   
## 4 0.070 0.971 0.185  
##   
## P value adjustment method: BH

ggplot(G3, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

## Warning: Removed 1 rows containing non-finite values (`stat\_boxplot()`).

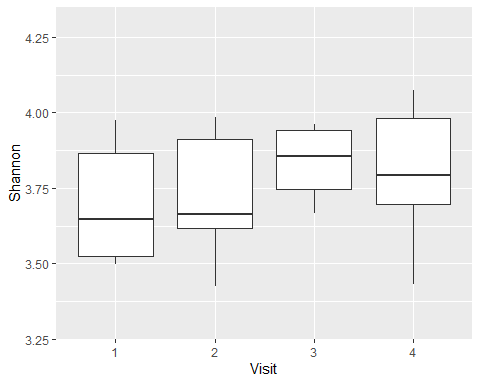


## Group 4

pairwise.wilcox.test(x = G4$Shannon, g= G4$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G4$Shannon and G4$Visit   
##   
## 1 2 3   
## 2 1.00 - -   
## 3 0.37 0.37 -   
## 4 0.37 0.37 1.00  
##   
## P value adjustment method: BH

ggplot(G4, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

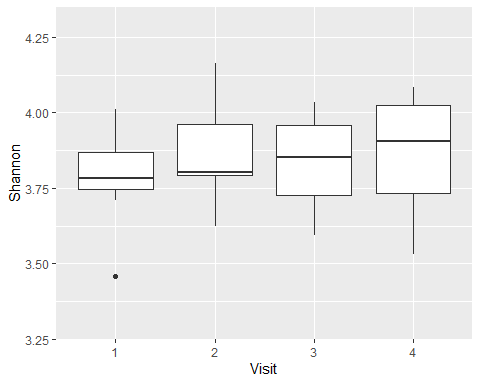


## Group 5

pairwise.wilcox.test(x = G5$Shannon, g= G5$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G5$Shannon and G5$Visit   
##   
## 1 2 3   
## 2 0.57 - -   
## 3 0.71 0.97 -   
## 4 0.57 0.96 0.87  
##   
## P value adjustment method: BH

ggplot(G5, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

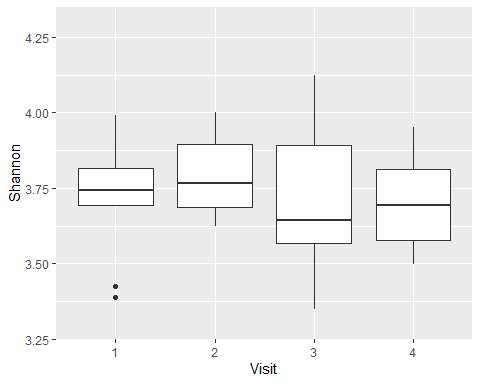


## Group 6

pairwise.wilcox.test(x = G6$Shannon, g= G6$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: G6$Shannon and G6$Visit   
##   
## 1 2 3   
## 2 0.63 - -   
## 3 0.97 0.63 -   
## 4 0.97 0.63 0.97  
##   
## P value adjustment method: BH

ggplot(G6, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

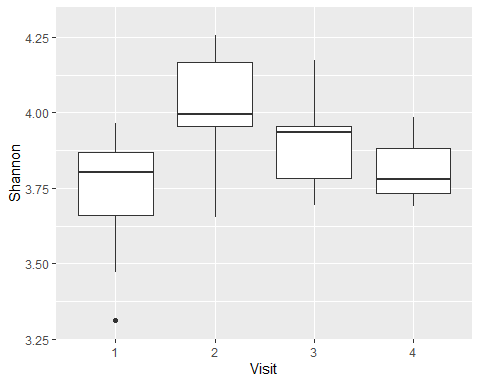


## Group G

pairwise.wilcox.test(x = GG$Shannon, g= GG$Visit, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: GG$Shannon and GG$Visit   
##   
## 1 2 3   
## 2 0.0032 - -   
## 3 0.0532 0.1261 -   
## 4 0.1903 0.0019 0.0042  
##   
## P value adjustment method: BH

ggplot(GG, aes(x = Visit, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)



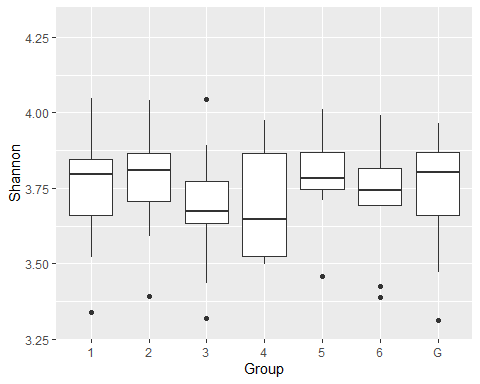
# Comparison between Groups within Visits

## Visit 1

pairwise.wilcox.test(x = V1$Shannon, g= V1$Group, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: V1$Shannon and V1$Group   
##   
## 1 2 3 4 5 6   
## 2 0.96 - - - - -   
## 3 0.91 0.91 - - - -   
## 4 0.91 0.91 0.91 - - -   
## 5 0.91 0.91 0.91 0.91 - -   
## 6 0.91 0.91 0.91 0.96 0.91 -   
## G 0.91 0.91 0.96 0.91 0.91 0.97  
##   
## P value adjustment method: BH

ggplot(V1, aes(x = Group, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

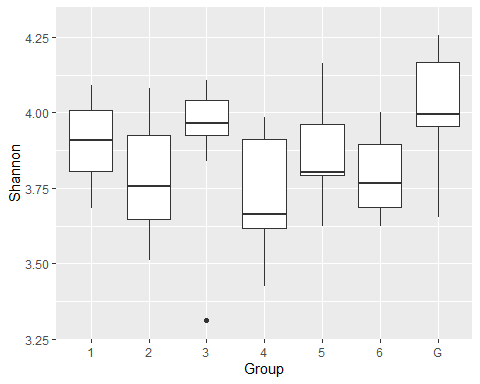


## Visit 2

pairwise.wilcox.test(x = V2$Shannon, g= V2$Group, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: V2$Shannon and V2$Group   
##   
## 1 2 3 4 5 6   
## 2 0.208 - - - - -   
## 3 0.486 0.138 - - - -   
## 4 0.124 0.971 0.124 - - -   
## 5 0.608 0.267 0.267 0.267 - -   
## 6 0.221 0.608 0.124 0.346 0.509 -   
## G 0.285 0.055 0.413 0.055 0.138 0.103  
##   
## P value adjustment method: BH

ggplot(V2, aes(x = Group, y = Shannon)) + geom\_boxplot()+ ylim(3.30,4.30)

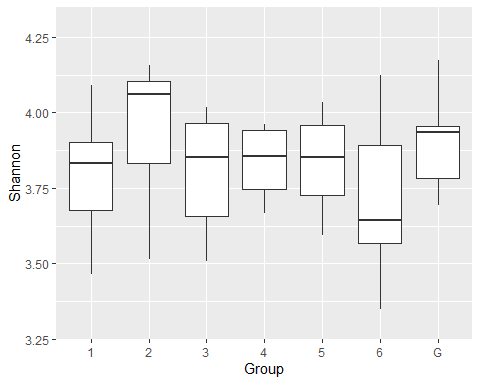


## Visit 3

pairwise.wilcox.test(x = V3$Shannon, g= V3$Group, p.adjust.method = "BH")

##   
## Pairwise comparisons using Wilcoxon rank sum exact test   
##   
## data: V3$Shannon and V3$Group   
##   
## 1 2 3 4 5 6   
## 2 0.65 - - - - -   
## 3 1.00 0.65 - - - -   
## 4 0.88 0.65 1.00 - - -   
## 5 0.98 0.65 0.98 1.00 - -   
## 6 0.70 0.65 0.70 0.65 0.65 -   
## G 0.65 1.00 0.70 0.70 0.72 0.65  
##   
## P value adjustment method: BH

ggplot(V3, aes(x = Group, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)

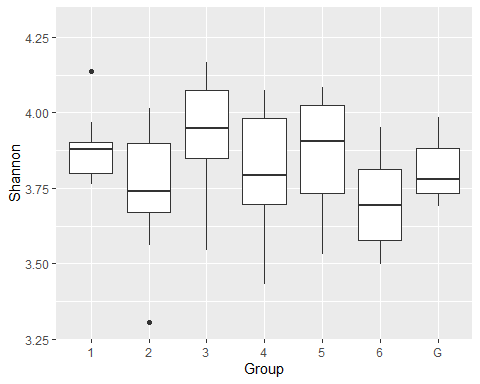


## Visit 4

pairwise.wilcox.test(x = V4$Shannon, g= V4$Group, p.adjust.method = "BH")

## Pairwise comparisons using Wilcoxon rank sum exact test   
## data: V4$Shannon and V4$Group   
##   
## 1 2 3 4 5 6   
## 2 0.199 - - - - -   
## 3 0.516 0.158 - - - -   
## 4 0.640 0.494 0.326 - - -   
## 5 0.684 0.199 0.640 0.640 - -   
## 6 0.065 0.662 0.062 0.199 0.114 -   
## G 0.020 0.114 0.020 0.020 0.020 0.158  
##   
## P value adjustment method: BH

ggplot(V4, aes(x = Group, y = Shannon)) + geom\_boxplot() + ylim(3.30,4.30)



# ANOSIM test

library(vegan)

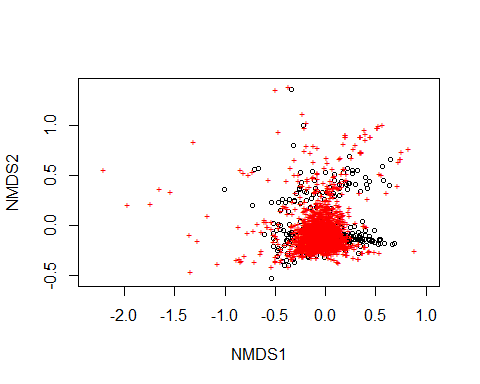
nmds\_input <- read.csv("C:/Users/jguitart/OneDrive/DOCTORAT-IRTA/AA\_PROJECTE/METAGENOMICS/Diversity\_Abundance\_analysis/microeco\_new/taxa\_abund\_bacteria\_new/nmds\_input.csv", sep=";", header = TRUE)  
com = nmds\_input[,7:ncol(nmds\_input)]  
m\_com = as.matrix(com)  
nmds = metaMDS(m\_com, distance = "bray")

## Run 0 stress 0.1250546   
## Run 1 stress 0.2095999   
## Run 2 stress 0.1412279   
## Run 3 stress 0.1631198   
## Run 4 stress 0.1470478   
## Run 5 stress 0.1392201   
## Run 6 stress 0.1320628   
## Run 7 stress 0.1475268   
## Run 8 stress 0.1419066   
## Run 9 stress 0.1408826   
## Run 10 stress 0.187665   
## Run 11 stress 0.1552786   
## Run 12 stress 0.1534262   
## Run 13 stress 0.1465906   
## Run 14 stress 0.2160852   
## Run 15 stress 0.1251549   
## ... Procrustes: rmse 0.00221234 max resid 0.03254529   
## Run 16 stress 0.1444962   
## Run 17 stress 0.1538149   
## Run 18 stress 0.1517287   
## Run 19 stress 0.2056574   
## Run 20 stress 0.1579979   
## \*\*\* No convergence -- monoMDS stopping criteria:  
## 1: no. of iterations >= maxit  
## 14: stress ratio > sratmax  
## 5: scale factor of the gradient < sfgrmin

nmds

##   
## Call:  
## metaMDS(comm = m\_com, distance = "bray")   
##   
## global Multidimensional Scaling using monoMDS  
##   
## Data: m\_com   
## Distance: bray   
##   
## Dimensions: 2   
## Stress: 0.1250546   
## Stress type 1, weak ties  
## No convergent solutions - best solution after 20 tries  
## Scaling: centring, PC rotation, halfchange scaling   
## Species: expanded scores based on 'm\_com'

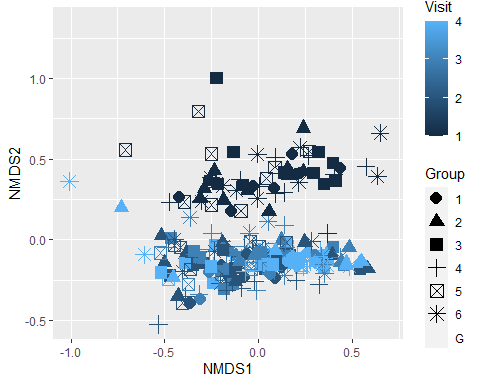
plot(nmds)



data.scores = as.data.frame(scores(nmds))  
data.scores$Number = nmds\_input$Number  
data.scores$Group = nmds\_input$Group  
data.scores$Visit = nmds\_input$Visit  
data.scores$Mother = nmds\_input$Mother  
data.scores$Parity = nmds\_input$Parity  
head(data.scores)

## NMDS1 NMDS2 Number Group Visit Mother Parity  
## 1 -0.3599444 0.255563236 3052 G 1 5599 2  
## 2 -0.1442166 -0.009824572 3052 G 2 5599 2  
## 3 0.2456311 -0.187187792 3052 G 3 5599 2  
## 4 0.6875227 -0.174714978 3052 G 4 5599 2  
## 5 -0.6708990 0.564880154 3053 G 1 5680 2  
## 6 -0.5207760 -0.132525448 3053 G 2 5680 2

library(ggplot2)  
xxx = ggplot(data.scores, aes(x = NMDS1, y = NMDS2)) +   
 geom\_point(size = 4, aes( shape = Group, colour = Visit))  
xxx



ano = anosim(m\_com, nmds\_input$Group, permutations = 999, distance = "bray")  
ano

##   
## Call:  
## anosim(x = m\_com, grouping = nmds\_input$Group, permutations = 999, distance = "bray")   
## Dissimilarity: bray   
##   
## ANOSIM statistic R: 0.01137   
## Significance: 0.041   
##   
## Permutation: free  
## Number of permutations: 999

There is statistical significance that microbial communities are not very dissimilar based on “Group”.

ano\_v = anosim(m\_com, nmds\_input$Visit, permutations = 999, distance = "bray")  
ano\_v

##   
## Call:  
## anosim(x = m\_com, grouping = nmds\_input$Visit, permutations = 999, distance = "bray")   
## Dissimilarity: bray   
##   
## ANOSIM statistic R: 0.3425   
## Significance: 0.001   
##   
## Permutation: free  
## Number of permutations: 999

There is statistical significance that microbial communities are somewhat dissimilar based on “Visit”.

ano\_n = anosim(m\_com, nmds\_input$Number, permutations = 999, distance = "bray")  
ano\_n

##   
## Call:  
## anosim(x = m\_com, grouping = nmds\_input$Number, permutations = 999, distance = "bray")   
## Dissimilarity: bray   
##   
## ANOSIM statistic R: -0.02784   
## Significance: 0.914   
##   
## Permutation: free  
## Number of permutations: 999

There is no statistical significant difference based on ID number.

ano\_m = anosim(m\_com, nmds\_input$Mother, permutations = 999, distance = "bray")  
ano\_m

##   
## Call:  
## anosim(x = m\_com, grouping = nmds\_input$Mother, permutations = 999, distance = "bray")   
## Dissimilarity: bray   
##   
## ANOSIM statistic R: -0.01709   
## Significance: 0.829   
##   
## Permutation: free  
## Number of permutations: 999

There is no statistical significant difference based on the ID of the mother.

ano\_p = anosim(m\_com, nmds\_input$Parity, permutations = 999, distance = "bray")  
ano\_p

##   
## Call:  
## anosim(x = m\_com, grouping = nmds\_input$Parity, permutations = 999, distance = "bray")   
## Dissimilarity: bray   
##   
## ANOSIM statistic R: -0.0145   
## Significance: 0.783   
##   
## Permutation: free  
## Number of permutations: 999

There is no statistical significant difference based on the parity number.

# Beta-diversity PCA

library(vegan)  
library(phyloseq)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1  
## v purrr 0.3.4   
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(patchwork)

library(agricolae)

## Registered S3 methods overwritten by 'klaR':  
## method from   
## predict.rda vegan  
## print.rda vegan  
## plot.rda vegan

library(FSA)

## Registered S3 methods overwritten by 'FSA':  
## method from  
## confint.boot car   
## hist.boot car   
## ## FSA v0.9.3. See citation('FSA') if used in publication.  
## ## Run fishR() for related website and fishR('IFAR') for related book.

library(rcompanion)

## Warning: package 'rcompanion' was built under R version 4.1.3

pca\_input <- read.csv("C:/Users/jguitart/Desktop/curs\_estadistica/beta\_diversity\_bacteria\_new/bray.csv", row.names=1, check.names = FALSE)

OTU = otu\_table(as.matrix(m\_com), taxa\_are\_rows = FALSE)  
data\_phylo <- phyloseq(OTU, metadata)  
pcoa <- ordinate(data\_phylo, "PCoA", "bray")  
plot\_ordination(data\_phylo, pcoa, color = "Visit") + geom\_point(size=3)

## Warning in plot\_ordination(data\_phylo, pcoa, color = "Visit"): Full  
## functionality requires `physeq` be phyloseq-class with multiple components.

## Warning in plot\_ordination(data\_phylo, pcoa, color = "Visit"): Color variable  
## was not found in the available data you provided.No color mapped.

## No available covariate data to map on the points for this plot `type`

