Phage Alpha and Beta Diversity Using GPD

2024-07-31

Load and prepare data for phyloseq

Create counts matrix

```
# Load counts
gpd_counts <- read.csv("features_reads_raw_count.tsv", header = TRUE,
                        skipNul = TRUE, sep = "\t", as.is = TRUE)
#gpd_counts
#str(gpd_counts)
#row.names(gpd_counts)
#class(gpd_counts)
colnames(gpd_counts)
```

```
##    [1] "phage"    "C102"    "C103"    "C104"    "C105"    "C1"      "C107"
##    [8] "C111"     "C114"    "C116"    "C118"    "C119"    "C123"    "C124"
##   [15] "C134"     "C135"    "C136"    "C137"    "C140"    "C142"    "C146"
##   [22] "C147"     "C148"    "C15"     "C152IIP" "C18"     "C19"     "C20"
##   [29] "C21"      "C23"     "C24"     "C26"     "C28"     "C30"     "C32"
##   [36] "C33"      "C34old"  "C35"     "C40"     "C44"     "C46"     "C47"
##   [43] "C48"      "C49"     "C5"      "C51"     "C54"     "C59"     "C65"
##   [50] "C68"      "C69"     "C70"     "C7"      "C72"     "C74"     "C75"
##   [57] "C76II"    "C80"     "C82"     "C85"     "C86"     "C87"     "C88"
##   [64] "C89"      "C90"     "C9"      "C95"     "C96"     "C98"     "P100"
##   [71] "P103"     "P104IIP" "P105"    "P107"    "P10old"  "P114"    "P115"
##   [78] "P116"     "P118"    "P119"    "P11old"  "P120"    "P12"     "P14"
##   [85] "P15"      "P16"     "P17"     "P18"     "P19"     "P20"     "P24"
##   [92] "P26"      "P28"     "P31"     "P34"     "P37"     "P38"     "P4"
##   [99] "P42"      "P43"     "P45"     "P46"     "P47"     "P48"     "P50"
##  [106] "P5"       "P51"     "P52"     "P53"     "P56"     "P57"     "P58"
##  [113] "P59"      "P60"     "P61"     "P62P"    "P63"     "P66"     "P67"
##  [120] "P68"      "P69"     "P70"     "P71"     "P72"     "P73"     "P74"
##  [127] "P77"      "P79"     "P8"      "P83"     "P85"     "P87"     "P88"
##  [134] "P9"       "P94"     "P95II"   "P99"
```

```
#head(colnames(gpd_counts))
#dim(gpd_counts)
#gpd_counts$phage

## Put this in `matrix` class for `phyloseq`'s `otu table`, and create it:
counts <- as.matrix(sapply(gpd_counts, as.numeric))
#head(counts)
#head(colnames(counts))
rownames(counts) <- gpd_counts$phage
#head(counts)
#head(colnames(counts))
```

```r
# delete/remove the newly created "phage" column:
counts <- counts[ , -grep("phage", colnames(counts))]
#head(counts)
#head(colnames(counts))
#dim(counts)

otu.table <- otu_table(counts, taxa_are_rows = TRUE)
head(otu.table)
```

```
## OTU Table:          [6 taxa and 136 samples]
##                 taxa are rows
##         C102 C103 C104 C105 C1 C107 C111 C114 C116 C118 C119 C123 C124 C134 C135
## ivig_1     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
## ivig_2     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
## ivig_3     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
## ivig_6     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
## ivig_7     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
## ivig_8     0    0    0    0  0    0    0    0    0    0    0    0    0    0    0
##         C136 C137 C140 C142 C146 C147 C148 C15 C152IIP C18 C19 C20 C21 C23 C24
## ivig_1     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
## ivig_2     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
## ivig_3     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
## ivig_6     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
## ivig_7     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
## ivig_8     0    0    0    0    0    0    0   0       0   0   0   0   0   0   0
##         C26 C28 C30 C32 C33 C34old C35 C40 C44 C46 C47 C48 C49 C5 C51 C54 C59
## ivig_1    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
## ivig_2    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
## ivig_3    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
## ivig_6    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
## ivig_7    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
## ivig_8    0   0   0   0   0      0   0   0   0   0   0   0   0  0   0   0   0
##         C65 C68 C69 C70 C7 C72 C74 C75 C76II C80 C82 C85 C86 C87 C88 C89 C90 C9
## ivig_1    0   0   0   0  0   0   0   0     0   0   0   0   0   0   0   0   0  0
## ivig_2    0   0   0   0  0   0   0   0     0   0   0   0   0   0   0   0   0  0
## ivig_3    0   0   0   0  0   0   0   0     0   0   0   0   0   0   0   0   0  0
## ivig_6    0   0   0   0  0   0   0   0     0   0   0   0   1   0   0   0   0  0
## ivig_7    0   0   0   0  0   0   0   0     0   0   0   0   0   0   0   0   0  0
## ivig_8    0   0   0   0  0   0   0   0     0   0   0   0   0   0   0   0   0  0
##         C95 C96 C98 P100 P103 P104IIP P105 P107 P10old P114 P115 P116 P118 P119
## ivig_1    0   0   0    0    0       0    0    0      0    0    0    0    0    0
## ivig_2    0   0   0    0    0       0    0    0      0    0    0    0    0    0
## ivig_3    0   0   0    0    0       0    0    0      0    0    0    0    0    0
## ivig_6    0   0   0    0    0       0    0    0      0    0    0    0    0    0
## ivig_7    0   0   0    0    0       0    0    0      0    0    0    0    0    0
## ivig_8    0   0   0    0    0       0    0    0      0    0    0    0    0    0
##         P11old P120 P12 P14 P15 P16 P17 P18 P19 P20 P24 P26 P28 P31 P34 P37 P38
## ivig_1       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_2       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_3       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_6       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_7       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_8       0    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##           P4 P42 P43 P45 P46 P47 P48 P50 P5 P51 P52 P53 P56 P57 P58 P59 P60 P61
## ivig_1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_2   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_6   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_7   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ivig_8   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##         P62P P63 P66 P67 P68 P69 P70 P71 P72 P73 P74 P77 P79 P8 P83 P85 P87 P88
## ivig_1    0   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
## ivig_2    0   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
## ivig_3    0   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
## ivig_6    0   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
## ivig_7    0   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
## ivig_8   24   0   0   0   0   0   0   0   0   0   0   0   0  0   0   0   0   0
##         P9 P94 P95II P99
## ivig_1  0   0     0   0
## ivig_2  0   0     0   0
## ivig_3  0   0     0   0
## ivig_6  0   0     0   0
## ivig_7  0   0     0   0
## ivig_8  0   0     0   0
```

```r
head(colnames(otu.table))
```

```
## [1] "C102" "C103" "C104" "C105" "C1"   "C107"
```

```r
dim(otu.table)
```

```
## [1] 142809    136
```

```r
#otu.table
```

Create dummy tax table for TAX

```r
taxmat = matrix(sample(letters, 142809, replace = TRUE),
                nrow = nrow(otu.table), ncol = 7)
rownames(taxmat) <- rownames(otu.table)
colnames(taxmat) <- c("Domain", "Phylum", "Class", "Order",
                      "Family", "Genus", "Species")
#taxmat
TAX = tax_table(taxmat)
```

Import the metadata:

```r
sampledata <- as.data.frame(read.csv
                    (file = "pd_meta_with_ffq_and_scfa_only_oursamples_3variables.csv",
                     header = TRUE, sep = ",", row.names = 1))
#rownames(sampledata)
#colnames(otu.table)
identical(rownames(sampledata), colnames(otu.table))
```

```
## [1] TRUE
```

```r
sampledata = sample_data(sampledata)
```

Create physeq

```
physeqfinal <- phyloseq(otu.table,TAX, sampledata)
#physeqfinal
#sample_data(physeqfinal)
#summary(sample_data(physeqfinal))

physeqfinal.2 <- subset_taxa(physeqfinal, taxa_sums(physeqfinal) >0)
#head(sort(taxa_sums(physeqfinal.2), decreasing = FALSE))

physeqfinal.2
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:        [ 119827 taxa and 136 samples ]
## sample_data() Sample Data:      [ 136 samples by 4 sample variables ]
## tax_table()   Taxonomy Table:   [ 119827 taxa by 7 taxonomic ranks ]
```

```
physeqfinal
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:        [ 142809 taxa and 136 samples ]
## sample_data() Sample Data:      [ 136 samples by 4 sample variables ]
## tax_table()   Taxonomy Table:   [ 142809 taxa by 7 taxonomic ranks ]
```

```
summary(sample_data(physeqfinal.2))
```

```
##     Group              gender          age_at_stool_collection      BMI
##  Length:136         Length:136         Min.   :51.00           Min.   :17.51
##  Class :character   Class :character   1st Qu.:61.00           1st Qu.:24.13
##  Mode  :character   Mode  :character   Median :65.00           Median :26.31
##                                        Mean   :64.98           Mean   :26.69
##                                        3rd Qu.:69.00           3rd Qu.:28.62
##                                        Max.   :78.00           Max.   :37.87
##                                                                NA's   :13
```

Alpha Diversity for Group (C vs P)

```
richness.table <- estimate_richness(physeqfinal.2, split = TRUE,
                                    measures = c("Observed", "Chao1",
                                    "ACE", "Shannon", "InvSimpson"))

head(richness.table)
```
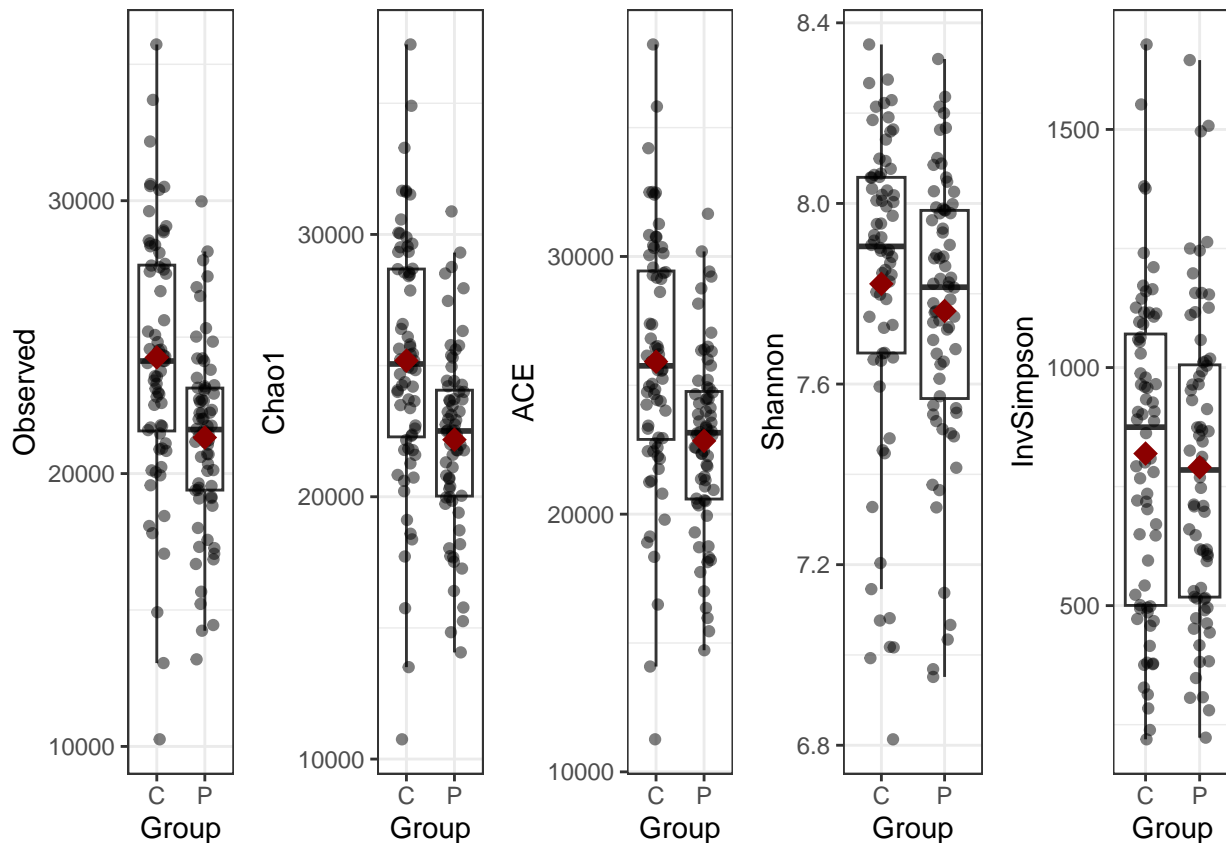
```
##      Observed   Chao1 se.chao1      ACE   se.ACE  Shannon InvSimpson
## C102    20050 20832.24 43.38079 21756.94 72.80999 7.076252   498.5324
## C103    17815 18374.95 36.74295 18902.23 67.94507 7.730928   378.2800
## C104    28092 29338.91 60.68695 30122.85 85.01034 8.228803  1115.2664
## C105    19571 20211.91 39.99259 20795.47 71.12560 7.925058   887.3270
## C1      28441 29547.82 55.87518 30334.29 85.17621 8.019084   964.4550
## C107    24564 25644.02 55.67177 26408.50 80.03626 7.954481   890.9713
```

```
richness.table$Group <- sample_data(physeqfinal.2)$Group
```

Plot Alpha Diversity

```
theme_set(theme_bw())

grid.arrange(
  ggplot(richness.table, aes(x = Group, y = Observed)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = Group, y = Chao1)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = Group, y = ACE)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = Group, y = Shannon)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = Group, y = InvSimpson)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  nrow = 1)
```

Observed Richness Wilcoxon rank sum test (Group)

```r
wilcox.test(richness.table$Observed ~ sample_data(physeqfinal.2)$Group,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Observed by sample_data(physeqfinal.2)$Group
## W = 3289, p-value = 2.137e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  1534 4355
## sample estimates:
## difference in location
##               2854.328
```

CHAO1 Wilcoxon rank sum test (Group)

```r
wilcox.test(richness.table$Chao1 ~ sample_data(physeqfinal.2)$Group,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Chao1 by sample_data(physeqfinal.2)$Group
## W = 3268, p-value = 3.201e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  1602.108 4481.770
## sample estimates:
## difference in location
##               2953.591
```

ACE Wilcoxon rank sum test (Group)

```r
wilcox.test(richness.table$ACE ~ sample_data(physeqfinal.2)$Group,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$ACE by sample_data(physeqfinal.2)$Group
## W = 3269, p-value = 3.141e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  1633.451 4551.437
## sample estimates:
## difference in location
##               3038.283
```

InvSimpson Wilcoxon rank sum test (Group)

```r
wilcox.test(richness.table$InvSimpson ~ sample_data(physeqfinal.2)$Group,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$InvSimpson by sample_data(physeqfinal.2)$Group
## W = 2439, p-value = 0.5819
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -77.22045 150.79223
## sample estimates:
## difference in location
##                32.37002
```

Shannon Wilcoxon rank sum test (Group)

```r
wilcox.test(richness.table$Shannon ~ sample_data(physeqfinal.2)$Group,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Shannon by sample_data(physeqfinal.2)$Group
## W = 2725, p-value = 0.0726
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.01002372  0.18103460
## sample estimates:
## difference in location
##              0.08479553
```
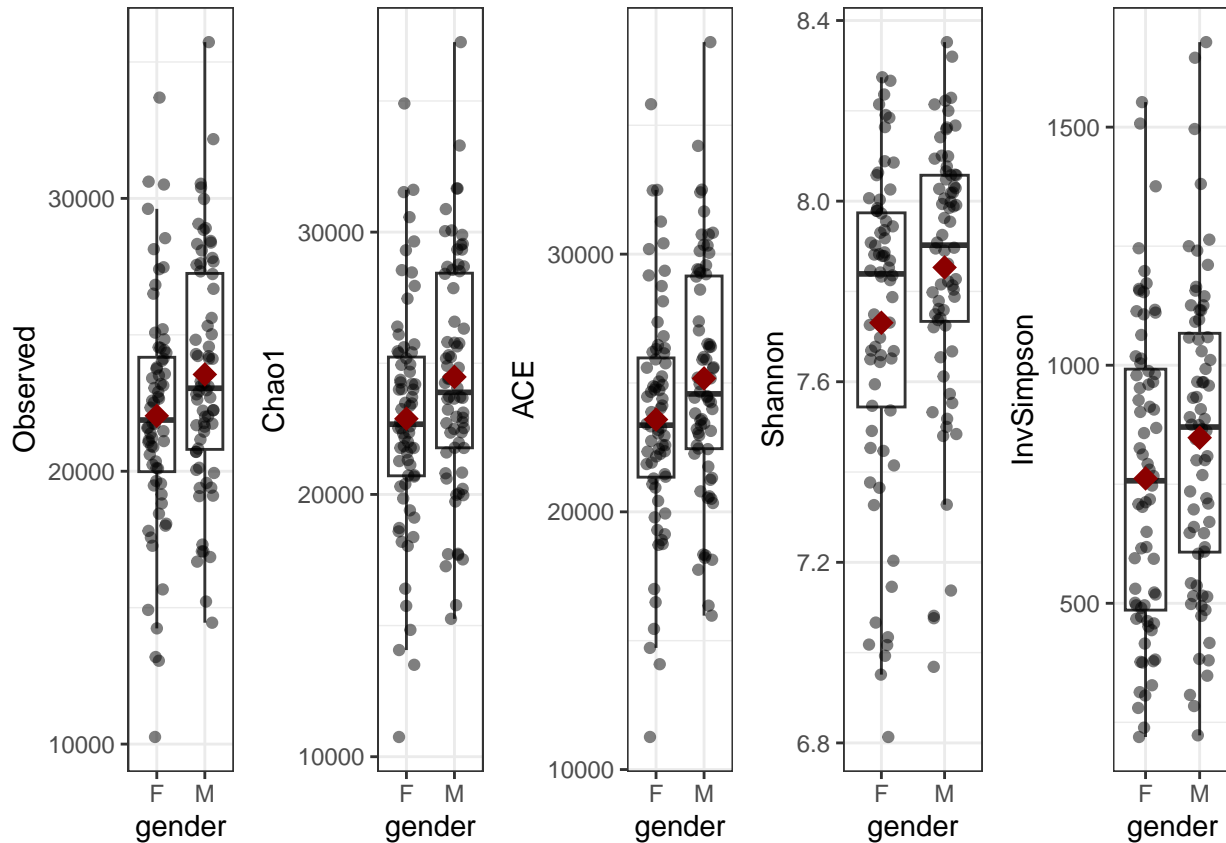
Alpha Diversity for Gender (F vs M)

```r
richness.table$gender <- sample_data(physeqfinal.2)$gender
```

Plot Alpha Diversity for Gender (F vs M)

```r
theme_set(theme_bw())

grid.arrange(
  ggplot(richness.table, aes(x = gender, y = Observed)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = gender, y = Chao1)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
  ggplot(richness.table, aes(x = gender, y = ACE)) +
    geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
    stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
```

```
ggplot(richness.table, aes(x = gender, y = Shannon)) +
  geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
  stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
ggplot(richness.table, aes(x = gender, y = InvSimpson)) +
  geom_boxplot(outlier.shape = NA) + geom_jitter(width = 0.2, alpha =.5) +
  stat_summary(fun = mean, color = "darkred", geom = "point", shape = 18, size = 4),
nrow = 1)
```


Observed Richness Wilcoxon rank sum test (gender)

```
wilcox.test(richness.table$Observed ~ sample_data(physeqfinal.2)$gender,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Observed by sample_data(physeqfinal.2)$gender
## W = 1879, p-value = 0.05978
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -2780.00003    63.00006
## sample estimates:
## difference in location
##                   -1310
```

CHAO1 Wilcoxon rank sum test (gender)

```
wilcox.test(richness.table$Chao1 ~ sample_data(physeqfinal.2)$gender,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Chao1 by sample_data(physeqfinal.2)$gender
## W = 1861, p-value = 0.04991
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -2971.0483727    -0.7484694
## sample estimates:
## difference in location
##              -1391.321
```

ACE Wilcoxon rank sum test (gender)

```
wilcox.test(richness.table$ACE ~ sample_data(physeqfinal.2)$gender,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$ACE by sample_data(physeqfinal.2)$gender
## W = 1870, p-value = 0.05466
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -3004.15480     37.69457
## sample estimates:
## difference in location
##              -1401.348
```

InvSimpson Wilcoxon rank sum test (gender)

```
wilcox.test(richness.table$InvSimpson ~ sample_data(physeqfinal.2)$gender,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$InvSimpson by sample_data(physeqfinal.2)$gender
## W = 1958, p-value = 0.1239
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -196.80937    23.55243
## sample estimates:
## difference in location
##               -86.4198
```

Shannon Wilcoxon rank sum test (gender)

```
wilcox.test(richness.table$Shannon ~ sample_data(physeqfinal.2)$gender,
            conf.level = 0.95, conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  richness.table$Shannon by sample_data(physeqfinal.2)$gender
## W = 1829, p-value = 0.03573
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.194331848 -0.008515514
## sample estimates:
## difference in location
##              -0.1048387
```

Alpha Diversity for BMI (Continuous variable)

```
richness.table$BMI <- sample_data(physeqfinal.2)$BMI
```

Plot Alpha Diversity for BMI (Continuous variable)

```
common_theme <- theme(
  axis.text.x = element_text(angle = 45, hjust = 1),
)

grid.arrange(
  ggplot(richness.table, aes(x = BMI, y = Observed)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "BMI", y = "Observed") +
    scale_x_continuous(breaks = seq(20, 35, by = 5)) +  # Adjust range as needed
    common_theme,

  ggplot(richness.table, aes(x = BMI, y = Chao1)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "BMI", y = "Chao1") +
    scale_x_continuous(breaks = seq(20, 35, by = 5)) +
    common_theme,

  ggplot(richness.table, aes(x = BMI, y = ACE)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "BMI", y = "ACE") +
    scale_x_continuous(breaks = seq(20, 35, by = 5)) +
    common_theme,

  ggplot(richness.table, aes(x = BMI, y = Shannon)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
```

```
    labs(x = "BMI", y = "Shannon") +
    scale_x_continuous(breaks = seq(20, 35, by = 5)) +
    common_theme,

  ggplot(richness.table, aes(x = BMI, y = InvSimpson)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "BMI", y = "InvSimpson") +
    scale_x_continuous(breaks = seq(20, 35, by = 5)) +
    common_theme,

  nrow = 1
)
```
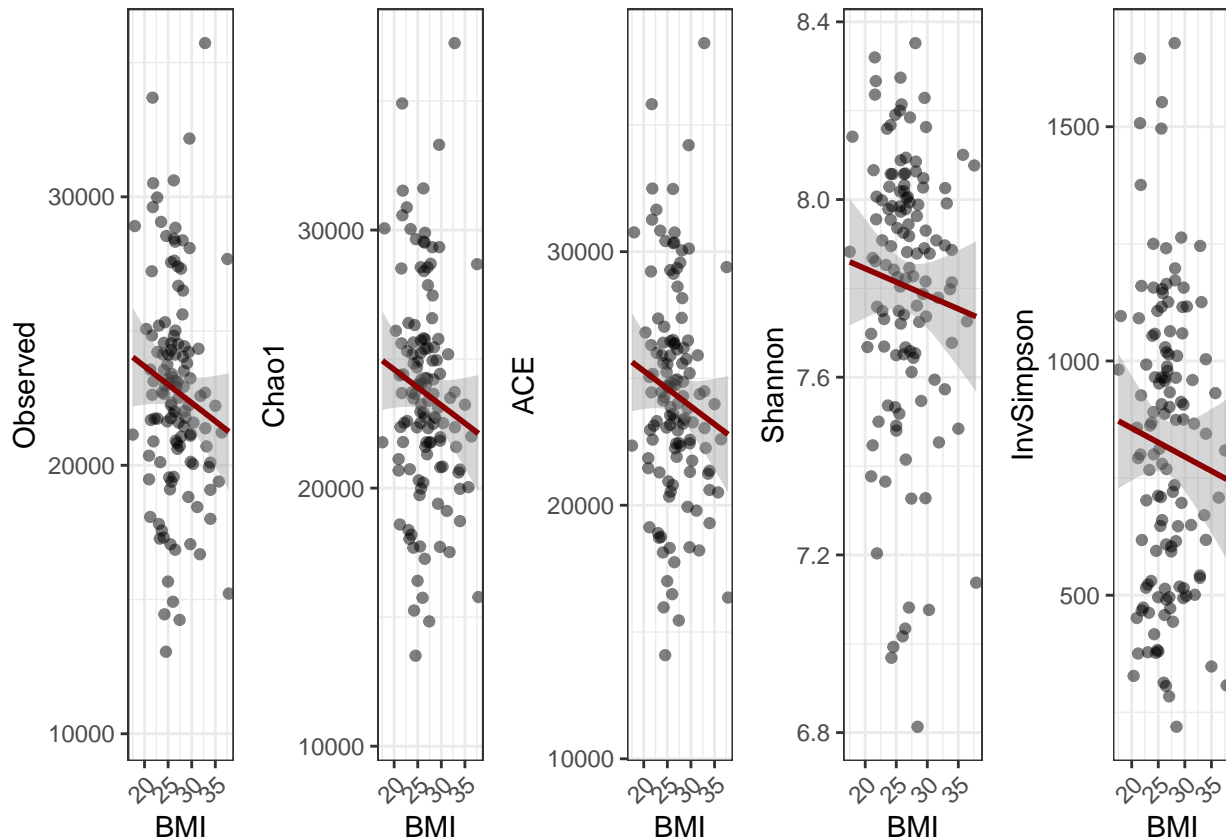
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```


Observed Richness Spearman rank sum test (BMI)

```
cor.test(richness.table$Observed, sample_data(physeqfinal.2)$BMI, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
```

11

```
## data:  richness.table$Observed and sample_data(physeqfinal.2)$BMI
## S = 353345, p-value = 0.1242
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1393682
```

CHAO1 Richness Spearman rank sum test (BMI)

```r
cor.test(richness.table$Chao1, sample_data(physeqfinal.2)$BMI, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$Chao1 and sample_data(physeqfinal.2)$BMI
## S = 350757, p-value = 0.1486
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1310231
```

ACE Richness Spearman rank sum test (BMI)

```r
cor.test(richness.table$ACE, sample_data(physeqfinal.2)$BMI, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$ACE and sample_data(physeqfinal.2)$BMI
## S = 350039, p-value = 0.156
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1287078
```

InvSimpson Richness Spearman rank sum test (BMI)

```r
cor.test(richness.table$InvSimpson, sample_data(physeqfinal.2)$BMI, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$InvSimpson and sample_data(physeqfinal.2)$BMI
## S = 313483, p-value = 0.9054
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.01083126
```

Shannon Richness Spearman rank sum test (BMI)

```r
cor.test(richness.table$Shannon, sample_data(physeqfinal.2)$BMI, method = "spearman")
```

```
## 
##  Spearman's rank correlation rho
## 
## data:  richness.table$Shannon and sample_data(physeqfinal.2)$BMI
## S = 330886, p-value = 0.4619
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho 
## -0.06694806
```

Alpha Diversity for Age (Continuous variable)

```
richness.table$age <- sample_data(physeqfinal.2)$age_at_stool_collection
```

Plot Alpha Diversity for Age (Continuous variable)

```
common_theme <- theme(
  axis.text.x = element_text(angle = 45, hjust = 1),
)

grid.arrange(
  ggplot(richness.table, aes(x = age, y = Observed)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "age", y = "Observed") +
    scale_x_continuous(breaks = seq(50, 80, by = 10)) +  # Adjust range as needed
    common_theme,

  ggplot(richness.table, aes(x = age, y = Chao1)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "age", y = "Chao1") +
    scale_x_continuous(breaks = seq(50, 80, by = 10)) +
    common_theme,

  ggplot(richness.table, aes(x = age, y = ACE)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "age", y = "ACE") +
    scale_x_continuous(breaks = seq(50, 80, by = 10)) +
    common_theme,

  ggplot(richness.table, aes(x = age, y = Shannon)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
    labs(x = "age", y = "Shannon") +
    scale_x_continuous(breaks = seq(50, 80, by = 10)) +
    common_theme,

  ggplot(richness.table, aes(x = age, y = InvSimpson)) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", color = "darkred") +
```

```
    labs(x = "age", y = "InvSimpson") +
    scale_x_continuous(breaks = seq(50, 80, by = 10)) +
    common_theme,

  nrow = 1
)
```
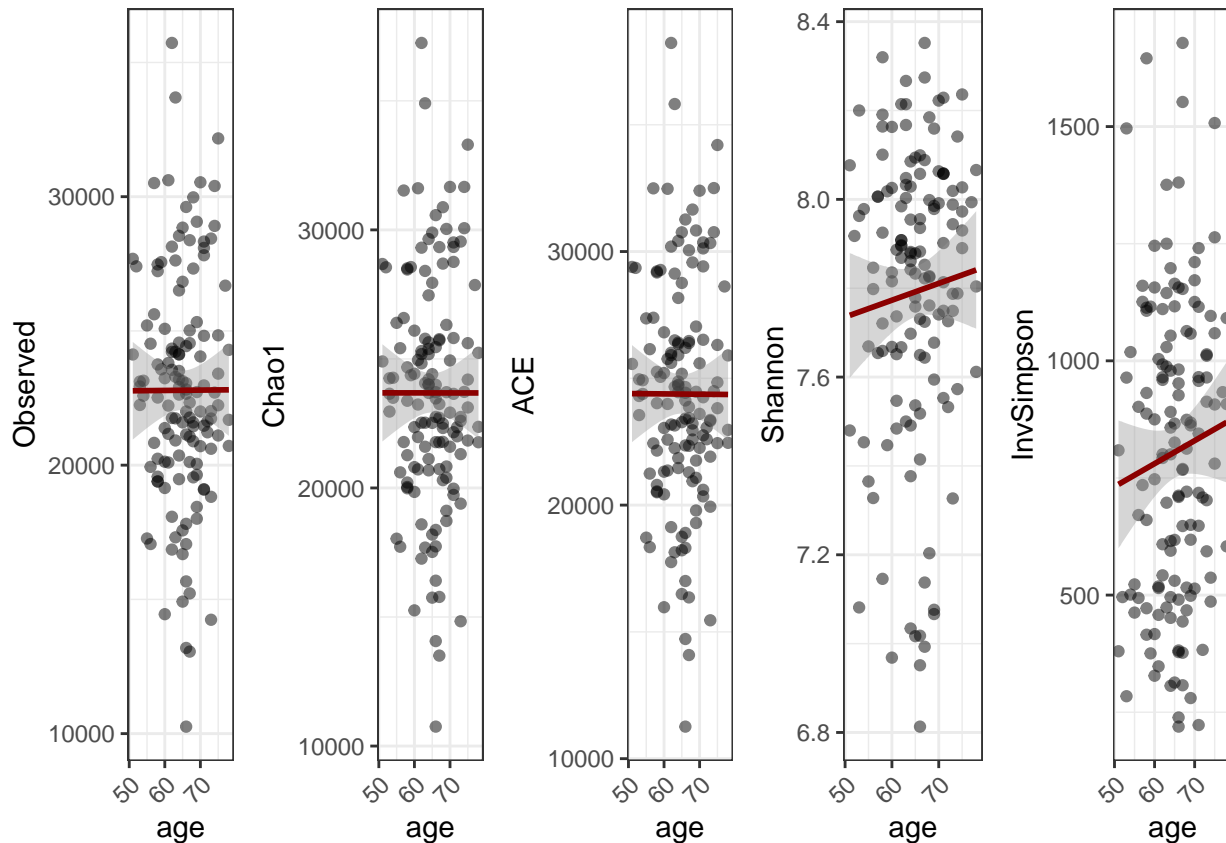
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```


Observed Richness Spearman rank sum test (age)

```
cor.test(richness.table$Observed,
         sample_data(physeqfinal.2)$age_at_stool_collection, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$Observed and sample_data(physeqfinal.2)$age_at_stool_collection
## S = 425498, p-value = 0.8626
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.01497588
```

CHAO1 Richness Spearman rank sum test (age)

```
cor.test(richness.table$Chao1,
         sample_data(physeqfinal.2)$age_at_stool_collection, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$Chao1 and sample_data(physeqfinal.2)$age_at_stool_collection
## S = 425243, p-value = 0.8682
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.01436671
```

ACE Richness Spearman rank sum test (age)

```
cor.test(richness.table$ACE,
         sample_data(physeqfinal.2)$age_at_stool_collection, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$ACE and sample_data(physeqfinal.2)$age_at_stool_collection
## S = 426074, p-value = 0.8502
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.01634948
```

InvSimpson Richness Spearman rank sum test (age)

```
cor.test(richness.table$InvSimpson,
         sample_data(physeqfinal.2)$age_at_stool_collection, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$InvSimpson and sample_data(physeqfinal.2)$age_at_stool_collection
## S = 382799, p-value = 0.3145
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.08687872
```

Shannon Richness Spearman rank sum test (age)

```
cor.test(richness.table$Shannon,
         sample_data(physeqfinal.2)$age_at_stool_collection, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  richness.table$Shannon and sample_data(physeqfinal.2)$age_at_stool_collection
```

```
## S = 395535, p-value = 0.5136
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.05649697
```

Distance-based multivariate analysis / beta diversity

CLR transformation followed by Euclidian Distance (a.k.a. Aitchinson Distance)

```r
physeqfinal.2.clr <- microbiome::transform(physeqfinal.2, transform = "clr", target = "OTU",
                                           shift = 0, scale = 1)
physeqfinal.2.clr.eucl_dist <- vegdist(t(as.data.frame(as.matrix(otu_table(physeqfinal.2.clr)))),
                                       method = "euclidean", binary = FALSE)

sample_data(physeqfinal.2.clr)$Group <- factor(sample_data(physeqfinal.2.clr)$Group,
                                                levels=c("C" , "P"),
                                                ordered = FALSE)
sample_data(physeqfinal.2.clr)$gender <- factor(sample_data(physeqfinal.2.clr)$gender,
                                                 levels=c("F" , "M"),
                                                 ordered = FALSE)
```

Run the statistics for group variable

```r
set.seed(1337)
adonis.Res.clr <- adonis2(physeqfinal.2.clr.eucl_dist ~
                          sample_data(physeqfinal.2.clr)$Group,
                          perm = 10000,
                          na.action = na.exclude,
                          parallel = 10)

adonis.Res.clr
```
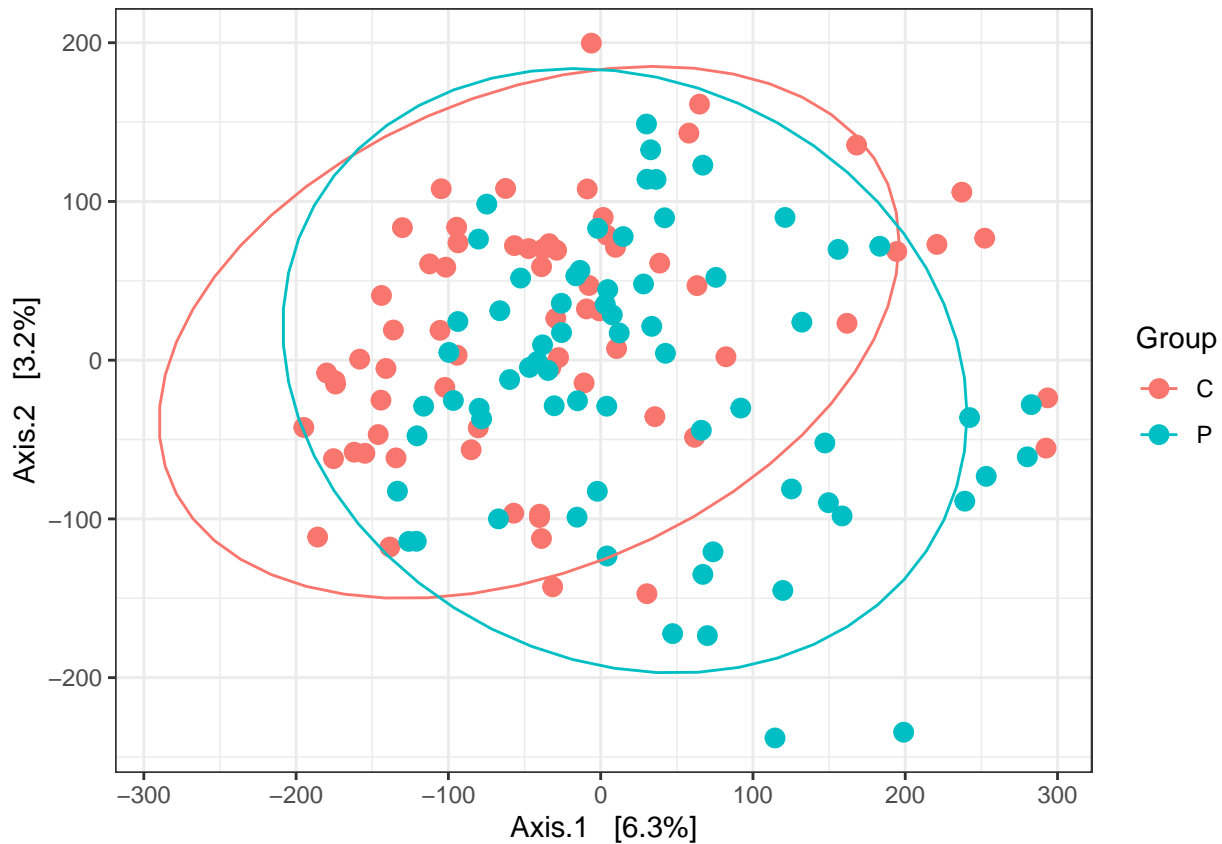
```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = physeqfinal.2.clr.eucl_dist ~ sample_data(physeqfinal.2.clr)$Group, permutations =
##                                     Df SumOfSqs     R2      F Pr(>F)
## sample_data(physeqfinal.2.clr)$Group    1   367500 0.01284 1.743   2e-04 ***
## Residual                            134 28252410 0.98716
## Total                               135 28619910 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Principal Coordinates Analysis (PCoA) for Group

```r
ord_clr <- ordinate(physeqfinal.2.clr, method = "PCoA",
                    distance = "euclidean", binary = FALSE, autotransform = FALSE,
                    trymax = 1000, maxit = 10000, sratmax = 0.999999,
                    previous.best, parallel = 10)
```

```
plot_ordination(physeqfinal.2.clr, ord_clr,
                type = "Samples", color = "Group") + geom_point(size = 3) +
  stat_ellipse(level = 0.95)
```


Run the statistics for gender variable

```
set.seed(1337)
adonis.Res.clr <- adonis2(physeqfinal.2.clr.eucl_dist ~
                          sample_data(physeqfinal.2.clr)$gender,
                          perm = 10000,
                          na.action = na.exclude,
                          parallel = 10)
adonis.Res.clr
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = physeqfinal.2.clr.eucl_dist ~ sample_data(physeqfinal.2.clr)$gender, permutations =
##                                          Df SumOfSqs     R2      F Pr(>F)
## sample_data(physeqfinal.2.clr)$gender   1   248936 0.0087 1.1758 0.05329 .
## Residual                               134 28370974 0.9913
## Total                                  135 28619910 1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
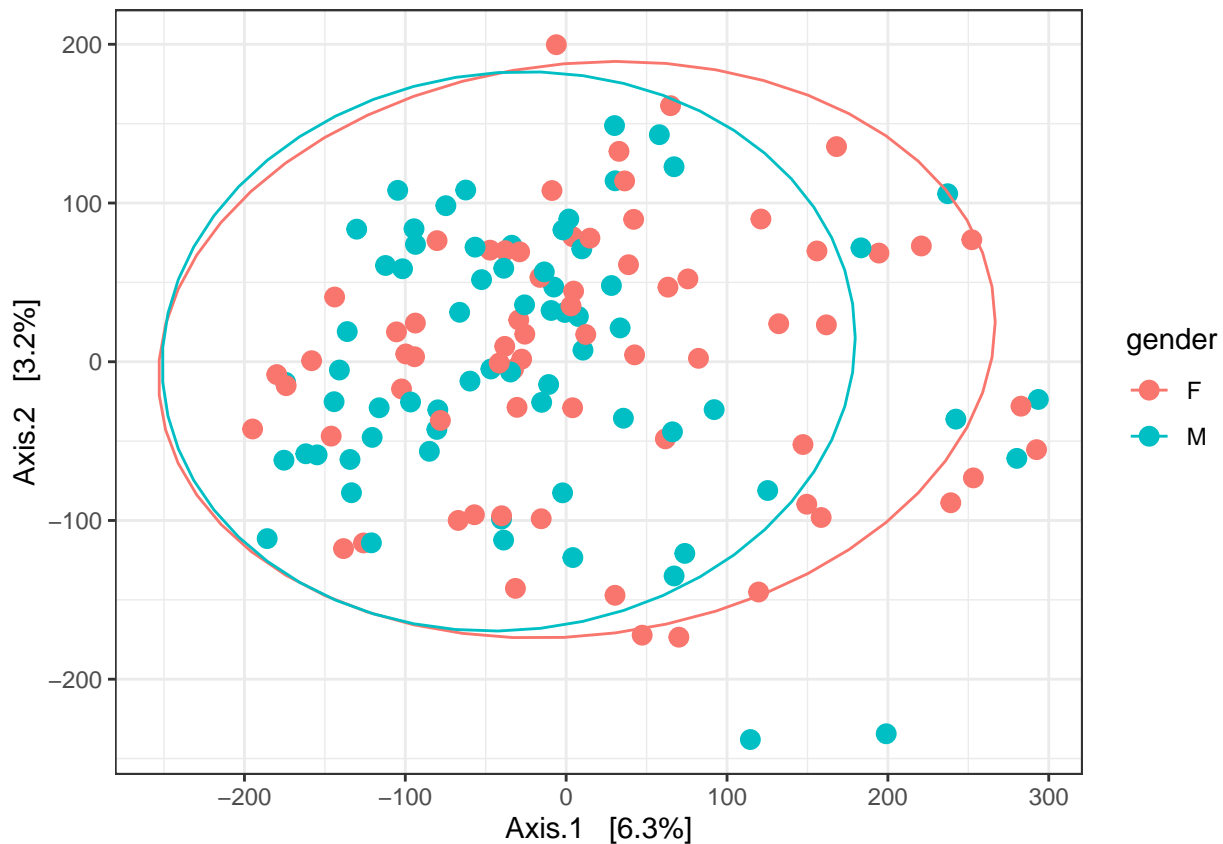
Principal Coordinates Analysis (PCoA) for Gender

```r
ord_clr <- ordinate(physeqfinal.2.clr, method = "PCoA",
                    distance = "euclidean", binary = FALSE, autotransform = FALSE,
                    trymax = 1000, maxit = 10000, sratmax = 0.999999,
                    previous.best, parallel = 10)
plot_ordination(physeqfinal.2.clr, ord_clr,
                type = "Samples", color = "gender") + geom_point(size = 3) +
  stat_ellipse(level = 0.95)
```


Run the statistics for BMI variable

```r
set.seed(1337)
adonis.Res.clr <- adonis2(physeqfinal.2.clr.eucl_dist ~
                          sample_data(physeqfinal.2.clr)$BMI,
                          perm = 10000,
                          na.action = na.exclude,
                          parallel = 10)

adonis.Res.clr
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = physeqfinal.2.clr.eucl_dist ~ sample_data(physeqfinal.2.clr)$BMI, permutations = 10
```

```
##                                        Df SumOfSqs     R2     F Pr(>F)
## sample_data(physeqfinal.2.clr)$BMI    1    230448 0.00898 1.097 0.1308
## Residual                            121  25419005 0.99102
## Total                               122  25649454 1.00000
```

Principal Coordinates Analysis (PCoA) for BMI

```
ord_clr <- ordinate(physeqfinal.2.clr, method = "PCoA",
                    distance = "euclidean", binary = FALSE, autotransform = FALSE,
                    trymax = 1000, maxit = 10000, sratmax = 0.999999,
                    previous.best, parallel = 10)
plot_ordination(physeqfinal.2.clr, ord_clr,
                type = "Samples", color = "BMI") + geom_point(size = 3) +
  stat_ellipse(level = 0.95)
```
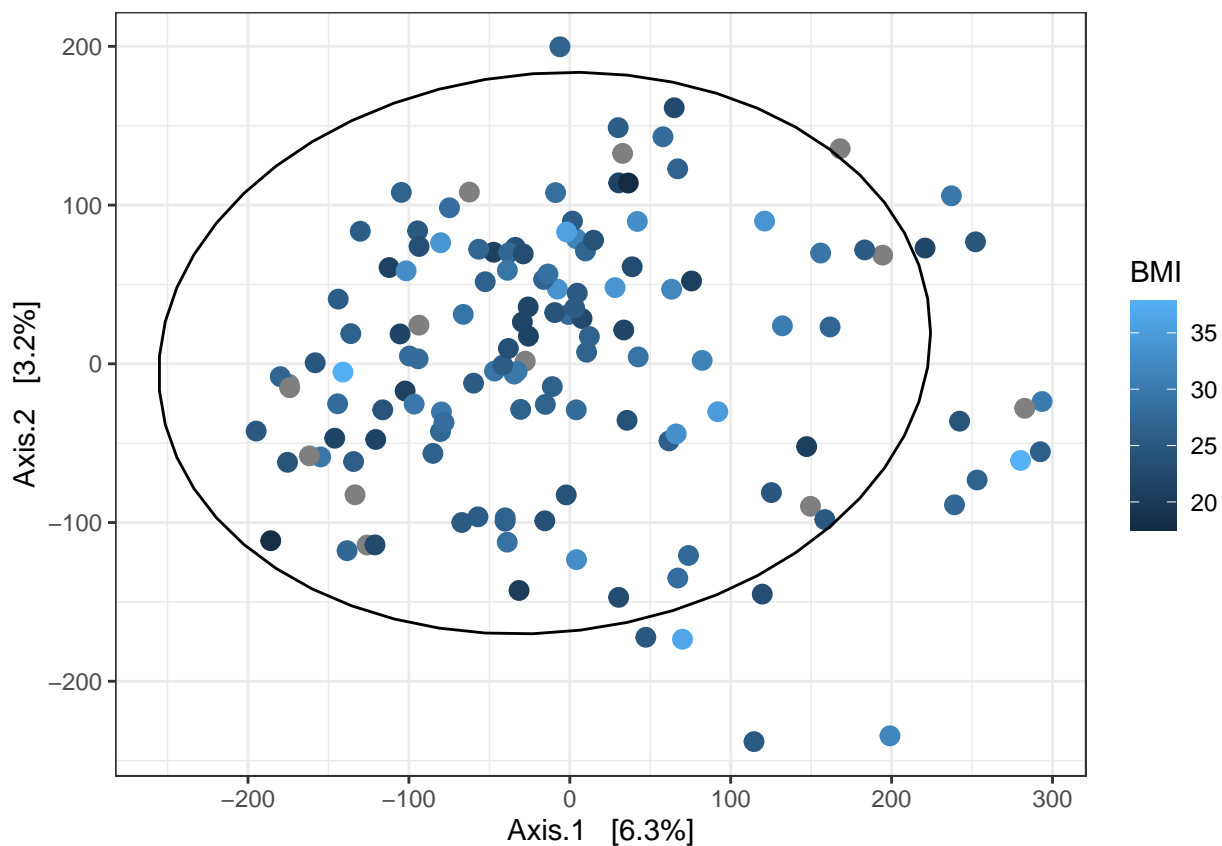

Run the statistics for Age variable

```
set.seed(1337)
adonis.Res.clr <- adonis2(physeqfinal.2.clr.eucl_dist ~
                          sample_data(physeqfinal.2.clr)$age_at_stool_collection,
                          perm = 10000,
                          na.action = na.exclude,
                          parallel = 10)

adonis.Res.clr
```

```
## Permutation test for adonis under reduced model
```

```
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = physeqfinal.2.clr.eucl_dist ~ sample_data(physeqfinal.2.clr)$age_at_stool_collectio
##                                                    Df SumOfSqs      R2
## sample_data(physeqfinal.2.clr)$age_at_stool_collection  1   210452 0.00735
## Residual                                          134 28409458 0.99265
## Total                                             135 28619910 1.00000
##                                                        F Pr(>F)
## sample_data(physeqfinal.2.clr)$age_at_stool_collection 0.9926  0.437
## Residual
## Total
```

Principal Coordinates Analysis (PCoA) for Age

```
ord_clr <- ordinate(physeqfinal.2.clr, method = "PCoA",
                    distance = "euclidean", binary = FALSE, autotransform = FALSE,
                    trymax = 1000, maxit = 10000, sratmax = 0.999999,
                    previous.best, parallel = 10)
plot_ordination(physeqfinal.2.clr, ord_clr,
                type = "Samples", color = "age_at_stool_collection") + geom_point(size = 3) +
  stat_ellipse(level = 0.95)
```
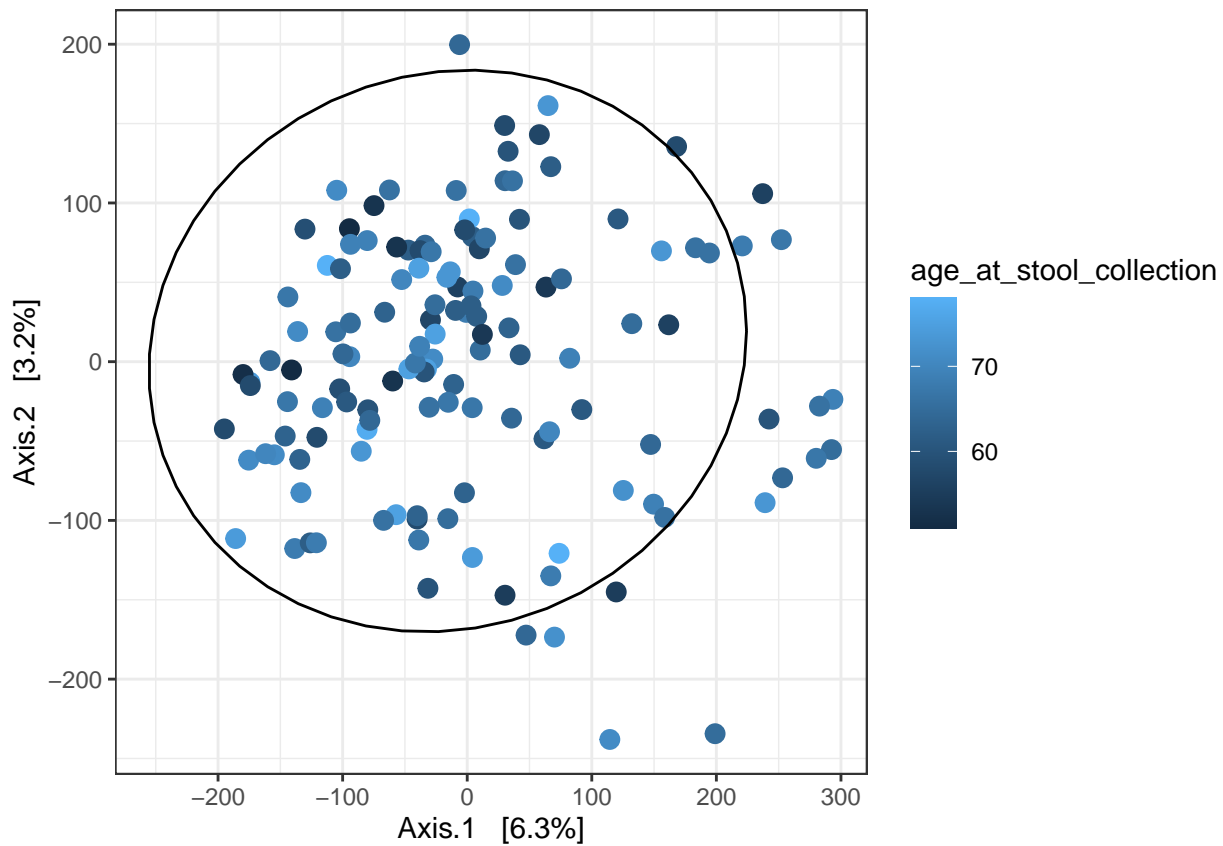
Session Info for reproducibility

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0
##
## locale:
##  [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
##  [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8
##  [7] LC_PAPER=en_GB.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Helsinki
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] microbiome_1.26.0         fido_1.1.1
##  [3] DESeq2_1.44.0             SummarizedExperiment_1.34.0
##  [5] Biobase_2.64.0            MatrixGenerics_1.16.0
##  [7] matrixStats_1.3.0         GenomicRanges_1.56.1
##  [9] GenomeInfoDb_1.40.1       IRanges_2.38.1
## [11] S4Vectors_0.42.1          BiocGenerics_0.50.0
## [13] vegan_2.6-6.1             lattice_0.22-5
## [15] permute_0.9-7             phyloseq_1.48.0
## [17] dabestr_2023.9.12         coin_1.4-3
## [19] survival_3.7-0            ggridges_0.5.6
## [21] qqplotr_0.0.6             MatrixCorrelation_0.10.0
## [23] energy_1.7-11             corrr_0.4.4
## [25] GGally_2.2.1              patchwork_1.2.0
## [27] cowplot_1.1.3             gridExtra_2.3
## [29] kableExtra_1.4.0          magrittr_2.0.3
## [31] purrr_1.0.2               reshape2_1.4.4
## [33] tidylog_1.1.0             tidyr_1.3.1
## [35] dplyr_1.1.4               RColorBrewer_1.1-3
## [37] ggplot2_3.5.1             BiocParallel_1.38.0
## [39] knitr_1.48
##
## loaded via a namespace (and not attached):
##   [1] libcoin_1.0-10           tensorA_0.36.2.1         rstudioapi_0.16.0
##   [4] jsonlite_1.8.8           TH.data_1.1-2            modeltools_0.2-23
##   [7] farver_2.1.2             rmarkdown_2.27           zlibbioc_1.50.0
##  [10] vctrs_0.6.5              multtest_2.60.0          tinytex_0.52
```

```
##  [13] htmltools_0.5.8.1       S4Arrays_1.4.1        progress_1.2.3
##  [16] distributional_0.4.0    plotrix_3.8-4         tidybayes_3.0.6
##  [19] Rhdf5lib_1.26.0         SparseArray_1.4.8     rhdf5_2.48.0
##  [22] pracma_2.4.4            plyr_1.8.9            sandwich_3.1-0
##  [25] zoo_1.8-12              igraph_2.0.3          lifecycle_1.0.4
##  [28] iterators_1.0.14        pkgconfig_2.0.3       Matrix_1.6-5
##  [31] R6_2.5.1                fastmap_1.2.0         GenomeInfoDbData_1.2.12
##  [34] digest_0.6.36           colorspace_2.1-1      RSpectra_0.16-2
##  [37] labeling_0.4.3          fansi_1.0.6           httr_1.4.7
##  [40] abind_1.4-5             mgcv_1.9-1            compiler_4.4.1
##  [43] withr_3.0.1             doParallel_1.0.17     gsl_2.1-8
##  [46] backports_1.5.0         ggstats_0.6.0         highr_0.11
##  [49] MASS_7.3-61             DelayedArray_0.30.1   biomformat_1.32.0
##  [52] caTools_1.18.2          tools_4.4.1          ape_5.8
##  [55] qqconf_1.3.2            glue_1.7.0           nlme_3.1-165
##  [58] rhdf5filters_1.16.0     grid_4.4.1           Rtsne_0.17
##  [61] checkmate_2.3.2         cluster_2.1.6        ade4_1.7-22
##  [64] generics_0.1.3          gtable_0.3.5         data.table_1.15.4
##  [67] hms_1.1.3               xml2_1.3.6           utf8_1.2.4
##  [70] XVector_0.44.0          ggdist_3.3.2         foreach_1.5.2
##  [73] pillar_1.9.0            stringr_1.5.1        posterior_1.6.0
##  [76] robustbase_0.99-3       splines_4.4.1        tidyselect_1.2.1
##  [79] locfit_1.5-9.10         Biostrings_2.72.1     arrayhelpers_1.1-0
##  [82] svglite_2.1.3           xfun_0.46            DEoptimR_1.1-3
##  [85] stringi_1.8.4           UCSC.utils_1.0.0      yaml_2.3.10
##  [88] boot_1.3-30             evaluate_0.24.0       codetools_0.2-19
##  [91] twosamples_2.0.1        tibble_3.2.1         cli_3.6.3
##  [94] pbmcapply_1.5.1         systemfonts_1.1.0     munsell_0.5.1
##  [97] Rcpp_1.0.13             coda_0.19-4.1         svUnit_1.0.6
## [100] parallel_4.4.1          prettyunits_1.2.0     opdisDownsampling_1.0.1
## [103] bitops_1.0-8            viridisLite_0.4.2     mvtnorm_1.2-5
## [106] scales_1.3.0            crayon_1.5.3          clisymbols_1.2.0
## [109] rlang_1.1.4             multcomp_1.4-26
```