Ecological dioversity of human microbiome

Bahar Sayoldin, bahar.sayoldin@gmail.com

This script investigates ecological diversity in human microbiome Publicly available data of Human Microbiome Project are used in this exercise

Reference: Lloyd-Price, J., Mahurkar, A., Rahnavard, G. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature 550, 61–66 (2017). https://doi.org/10.1038/nature23889

Load libraries

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-7
```

```
library(ggplot2)
library(ape)
library(rmarkdown)
library(knitr)
```

Set working directory

Read data

Load sample_info, a tab-delimited file

Load microbial species abundances, , a tab-delimited file

```
sample_info <- read.delim('data/metadata.txt', sep = "\t", row.names = 1)
Microbial_Species <- read.delim('data/Microbial_Species.txt', sep = "\t", row.names = 1)
dim(sample_info)
```

```
## [1] 2355    3
```

```
dim(Microbial_Species)
```

```
## [1] 2355 1091
```

Calculate similarity between samples based on microbial species abundance

```
veg_dist <- as.matrix(vegdist(Microbial_Species, method="bray"))
```

Write the distance matrix in you computer as a tab-delimited file

```
write.table( veg_dist, 'data/bray_disatnce.txt', sep = "\t", eol = "\n",
             na = "", col.names = NA, quote= F, row.names = T)
```

Ecological diversity

Create a data fram to hold statistics

```
diversity_data <-
  setNames(data.frame(matrix(
    nrow = dim(Microbial_Species)[1],  ncol = 4,
  )), c("Sample", "Group", "Alpha_diversity", "Beta_diversity"))
rownames(diversity_data) <- rownames(Microbial_Species)
```

Alpha diversity(within sample diversity)

```
shannoon <- diversity(Microbial_Species, "shannon")

diversity_data$Alpha_diversity <- shannoon
diversity_data$Body_area <- sample_info[rownames(Microbial_Species), "Body_area"]
```

Kruska Wallis test

**if microbial alpha diversity is different among human body areas**

```
res_kruskal.test <- kruskal.test(diversity_data$Alpha_diversity, sample_info$Body_area)
pvalue <- res_kruskal.test$p.value
```

Visulization

ggplot2 for manuscript quality figures

```
alpha_plot <- ggplot2::ggplot(
  data = diversity_data, ggplot2::aes(factor(Body_area), y=Alpha_diversity)) +
  ggplot2::geom_boxplot(ggplot2::aes(fill = Body_area), alpha=0.25) +
  ggplot2::geom_point(ggplot2::aes(fill = Body_area), size=2, alpha=0.5,
                      shape = 21, colour = "black",
                      position = ggplot2::position_jitterdodge(),
                      show.legend = F) +
  ggplot2::annotate(
    geom = "text",
    x = Inf,
    y = 0,
    hjust = 1,
    vjust = 1,
    label = sprintf(
      "P-value: %1.3e",
      pvalue
    ) ,
    color = "black",
    size = 4,
    fontface = "italic"
  )+
  xlab("Body area") + ylab("Shannon index")+
  theme_classic()
```

```
alpha_plot
```
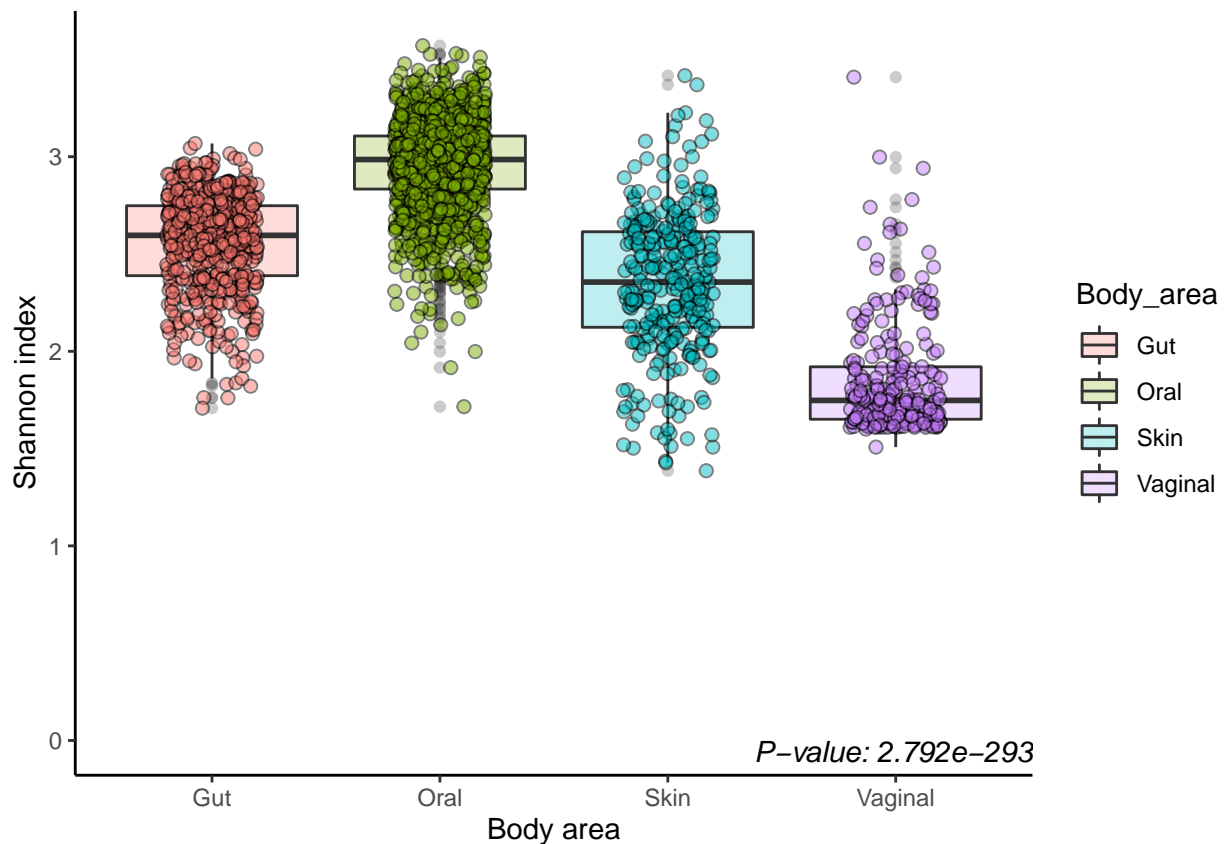
Save the plot with specifics

```
ggsave("analysis/alpha_diversity.pdf",alpha_plot, width = 5.6, height = 3.6,
       units = "in", dpi = 350  )
```