Lab_10_Metagenomics_MBI3100A_2022_Assignment

2022-11-12

R Markdown

```
# clear the R environment
rm(list = ls())
```

Install required libraries

Installing these libraries may take some time. Try to update all other dependencies when prompted (type "a" and enter).

```
if (!require("BiocManager")) install.packages("BiocManager")
```

```
## Loading required package: BiocManager
```

```
if (!require("phyloseq")) BiocManager::install("phyloseq")
```

```
## Loading required package: phyloseq
```

```
if (!require("tidyverse")) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
if (!require("DT")) install.packages("DT")
```

```
## Loading required package: DT
```

```
if (!require("dendextend")) install.packages("dendextend")
```

```
## Loading required package: dendextend
## Registered S3 method overwritten by 'dendextend':
##   method     from
##   rev.hclust vegan
##
## --------------------
## Welcome to dendextend version 1.16.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## You may ask questions at stackoverflow, use the r and dendextend tags:
##   https://stackoverflow.com/questions/tagged/dendextend
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## --------------------
##
##
## Attaching package: 'dendextend'
##
## The following object is masked from 'package:stats':
##
##     cutree
```

Load Libraries

```
library(phyloseq)
library(ggplot2)
library(dplyr)
library(DT)
library(dendextend)
```

Data import

Questin 1: Import the three files named as 'GP_sp_assignment_otu_table_df.csv', 'GP_sp_assignment_sample_data_df.csv', and 'GP_sp_assignment_tax_table_df.csv' and make a phyloseq object named 'asgmt_physeq'? (2 mark)

Please provide the correct file/folder path

```
GP_sp_assignment_otu_table = read.table("./assignment_files/GP_sp_assignment_otu_table_df.csv",
           sep = "\t",
           header =  T,
           row.names = "otus")
my_OTU_table = otu_table(GP_sp_assignment_otu_table, taxa_are_rows = TRUE)
GP_sp_assignment_sample_data = read.table("./assignment_files/GP_sp_assignment_sample_data_df.csv",
           sep = "\t", header =  T,
```

```
           row.names = "sampleid")
my_Sample_data = sample_data(GP_sp_assignment_sample_data)
GP_sp_assignment_tax_table = read.table("./assignment_files/GP_sp_assignment_tax_table_df.csv",
           sep = "\t",
           header =  T,
           row.names = "otus")
my_tax_table = tax_table(as.matrix(GP_sp_assignment_tax_table))

asgmt_physeq = phyloseq(my_OTU_table, my_Sample_data, my_tax_table)
```

Print asgmt_physeq

```
print(asgmt_physeq)
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 1413 taxa and 9 samples ]
## sample_data() Sample Data:       [ 9 samples by 8 sample variables ]
## tax_table()   Taxonomy Table:    [ 1413 taxa by 7 taxonomic ranks ]
```

Question 2: How many taxa, samples, and sample vairables are there in asgmt_physeq? (1mark)

```

```

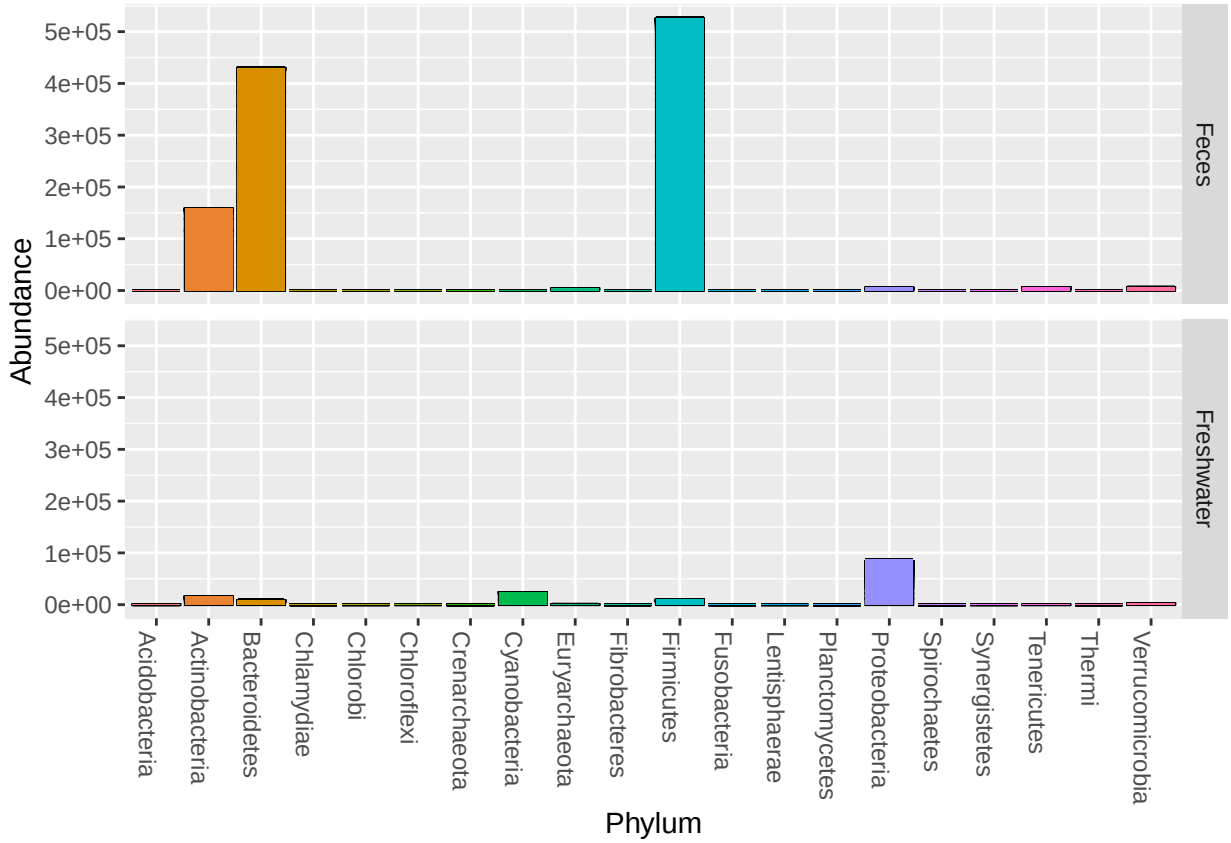
Question 3: List the catagories present in the sample variable "SampleOrigin". (1 mark)

```
sample_data(asgmt_physeq)$SampleOrigin %>% as.factor %>% levels()
```

```
## [1] "Feces"      "Freshwater"
```

Question 4: Generate a bar plot for sample vs abudance and facet it for the catagories in SampleOrigin (1 mark)

```
p = plot_bar(asgmt_physeq, x= "Phylum", fill = "Phylum", facet_grid=SampleOrigin~.)
p + theme(legend.position="none") + geom_bar(stat = "identity")
```



Question 5: Based on the plot generated in question 4, name all the phylum which big difference in abundance between "Feces" and "Freshwater" samples? (1 mark)

Transform the absolute abundance into relative abundance and filter the taxa which have mean relative abundance less than 0.0001

```
# To convert to relative abundance
asgmt_physeq_r  = transform_sample_counts(asgmt_physeq, function(x) x / sum(x) )
# Keep the taxa which have a mean values at least 0.0001
asgmt_physeq_rf = filter_taxa(asgmt_physeq_r, function(x) mean(x) > 0.0001, TRUE)
asgmt_physeq_rf
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 238 taxa and 9 samples ]
## sample_data() Sample Data:       [ 9 samples by 8 sample variables ]
## tax_table()   Taxonomy Table:    [ 238 taxa by 7 taxonomic ranks ]
```

Question 6: How many taxa are left after the above filtering?

#Extract OTU table

```
ps_rel_otu = data.frame(phyloseq::otu_table(asgmt_physeq))

ps_rel_otu = t(ps_rel_otu) # transpose the table (required by vegdist )
```

```
?hclust
```

Options availabe for hclust

```
## starting httpd help server ... done
```

Question 7: Generate a Hierarchical clustering plot