R Notebook

```
# load libraries
library(phyloseq)
library(ggplot2)
# set plotting theme
theme_set(theme_bw())
# make sure you are in 'metag' folder
getwd()
```

```
## [1] "/Users/husenzhang/Documents/GitHub/FAES_metagenomics"
```

```
# if not in 'metag' folder
# setwd("~/Desktop/metag")
# getting otu_table and tree files into a
# phyloseq object named ps
ps <- import_biom('out/json', treefilename = 'out/rep_set.tre')
# take a look at 'ps'
ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 76 taxa and 6 samples ]
## sample_data() Sample Data:       [ 6 samples by 2 sample variables ]
## tax_table()   Taxonomy Table:    [ 76 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 76 tips and 74 internal nodes ]
```

```
# aha 76 otus in 6 samples - we knew this since Tuesday
# just to double-check the dimention of our otu_table
dim(otu_table(ps))
```

```
## [1] 76  6
```

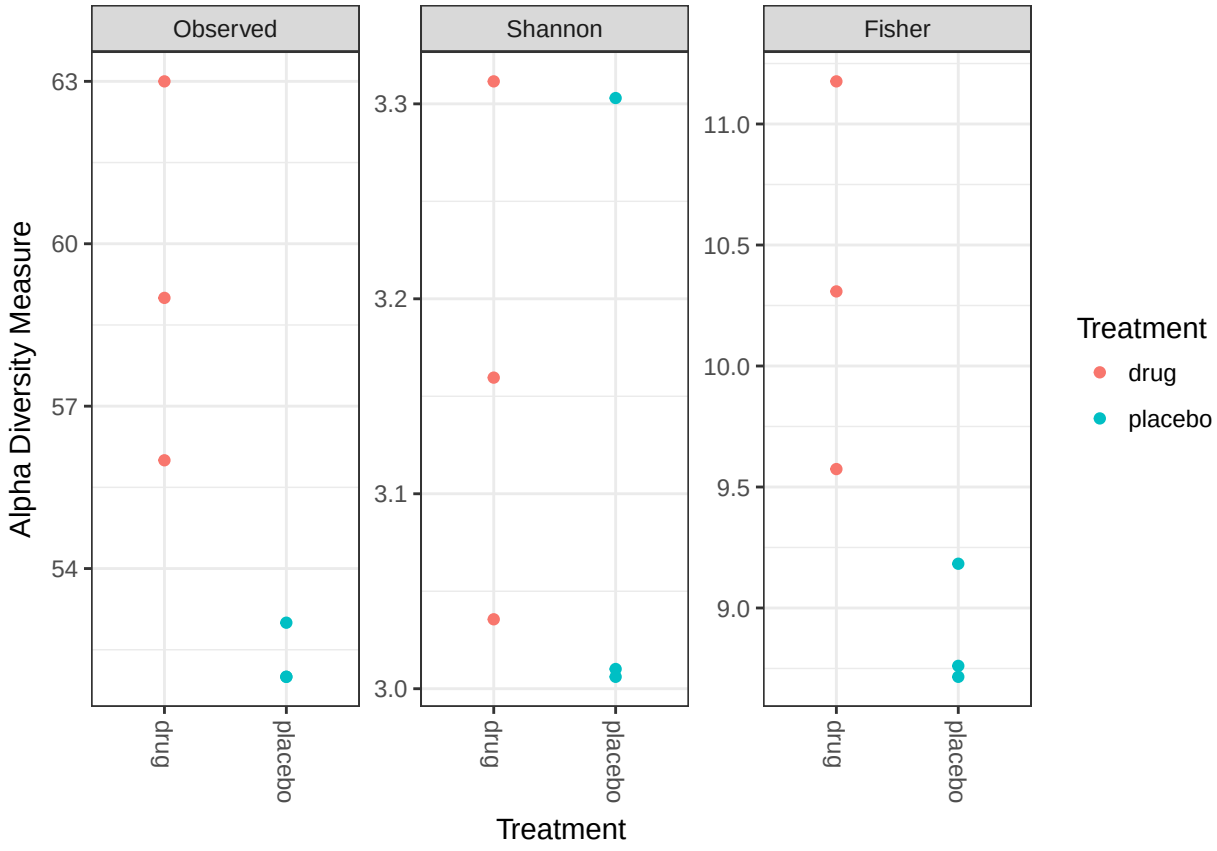
```
# take a look at data
head(otu_table(ps))
```

```
## OTU Table:          [6 taxa and 6 samples]
##                      taxa are rows
##        S1 S2 S3 S4  S5 S6
## OTU_37  1  0 15  9   0  0
## OTU_67  2  3  1  1   0  0
## OTU_25 17 27 23 15 111 89
## OTU_48  0  0  3  0  31 25
## OTU_62  3  1  1  4   4  1
## OTU_42 26  3  0  0   0  0
```

```
# 6 samples: S1 to S6
# 2 sample variables: Description and Treatment
sample_data(ps)
```

```
##    Description Treatment
## S1           1      drug
## S2           2   placebo
## S3           3      drug
## S4           4      drug
## S5           5   placebo
## S6           6   placebo
```

```
# plot alpha diversity
plot_richness(ps, x='Treatment' , measures = c('Observed','Fisher','Shannon'), color = 'Treatment')
```




```
# other valid measures are "Chao1", "ACE", "InvSimpson".
# these are ecological richness measures - google their meanings.
```

```
# now take a look at our rep_set.tre
phy_tree(ps)
```

```
##
## Phylogenetic tree with 76 tips and 74 internal nodes.
##
## Tip labels:
##  OTU_37, OTU_67, OTU_25, OTU_48, OTU_62, OTU_42, ...
## Node labels:
##  , 0.981, 0.926, 0.875, 0.768, 0.000, ...
##
## Unrooted; includes branch lengths.
```

```
# which OTUs are most abundant?
taxa_sums(ps)
```

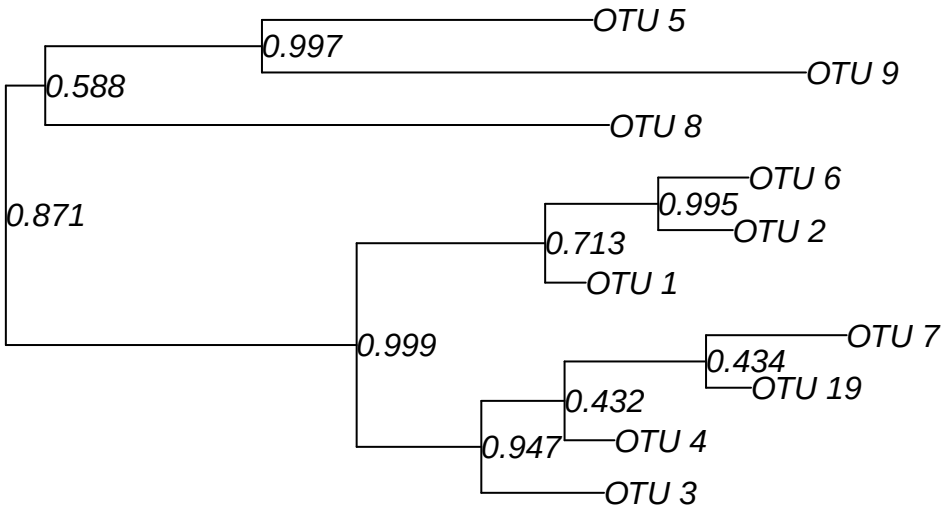
```
## OTU_37 OTU_67 OTU_25 OTU_48 OTU_62 OTU_42 OTU_15 OTU_31 OTU_53 OTU_57
##     25      7    282     59     14     29    279    271    130     16
## OTU_14 OTU_26 OTU_16 OTU_74 OTU_69 OTU_65 OTU_71 OTU_76 OTU_54 OTU_68
##    253    156    525     19     23     42      6     17    111     54
## OTU_66 OTU_55 OTU_45 OTU_17 OTU_47 OTU_60  OTU_3  OTU_4 OTU_20 OTU_19
##     13     69     14    308     48     75    784   1948    358    779
```

```
## OTU_33 OTU_43 OTU_36  OTU_7 OTU_59 OTU_75 OTU_10 OTU_23  OTU_1  OTU_2
##    252    231    146   1411     25     24    346    107    884   2272
## OTU_12 OTU_30 OTU_29  OTU_6 OTU_50 OTU_61 OTU_63  OTU_8 OTU_24 OTU_49
##    316    361    127    977     58      8     16    575    200      7
## OTU_72 OTU_41 OTU_46 OTU_64 OTU_44 OTU_73 OTU_38 OTU_35 OTU_56 OTU_39
##      3     36      6     23      5      9     38     89     71    100
##  OTU_9 OTU_52 OTU_18 OTU_11 OTU_22  OTU_5 OTU_13 OTU_21 OTU_28 OTU_51
##    809     83    500    271    254    609    375    191     51     76
## OTU_58 OTU_34 OTU_32 OTU_27 OTU_70 OTU_40
##     26     50    150    278     19     37
```

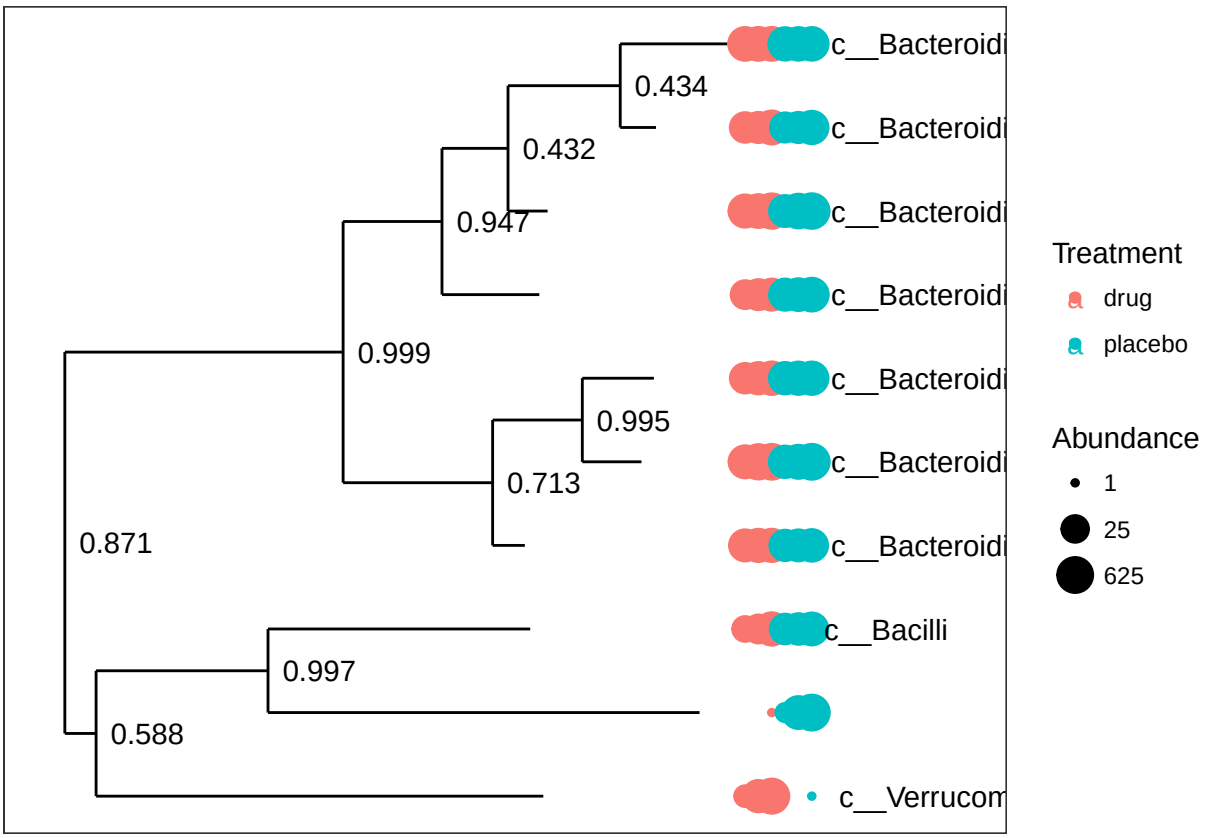
```
# not obvious! let's sort them
sort(taxa_sums(ps), decreasing = TRUE)[1:10]
```

```
##  OTU_2  OTU_4  OTU_7  OTU_6  OTU_1  OTU_9  OTU_3 OTU_19  OTU_5  OTU_8
##   2272   1948   1411    977    884    809    784    779    609    575
```

```
# let's plot a subset tree containing 10 most abundant OTUs
# step1 : myTaxa is a list of OTU names
myTaxa <- names(sort(taxa_sums(ps), decreasing = TRUE)[1:10])
# step 2: cut these 10 OTUs out from the whole dataset
ex1 = prune_taxa(myTaxa, ps)
# step 3: plot just these 10 OTUs
plot(phy_tree(ex1), show.node.label = TRUE)
```

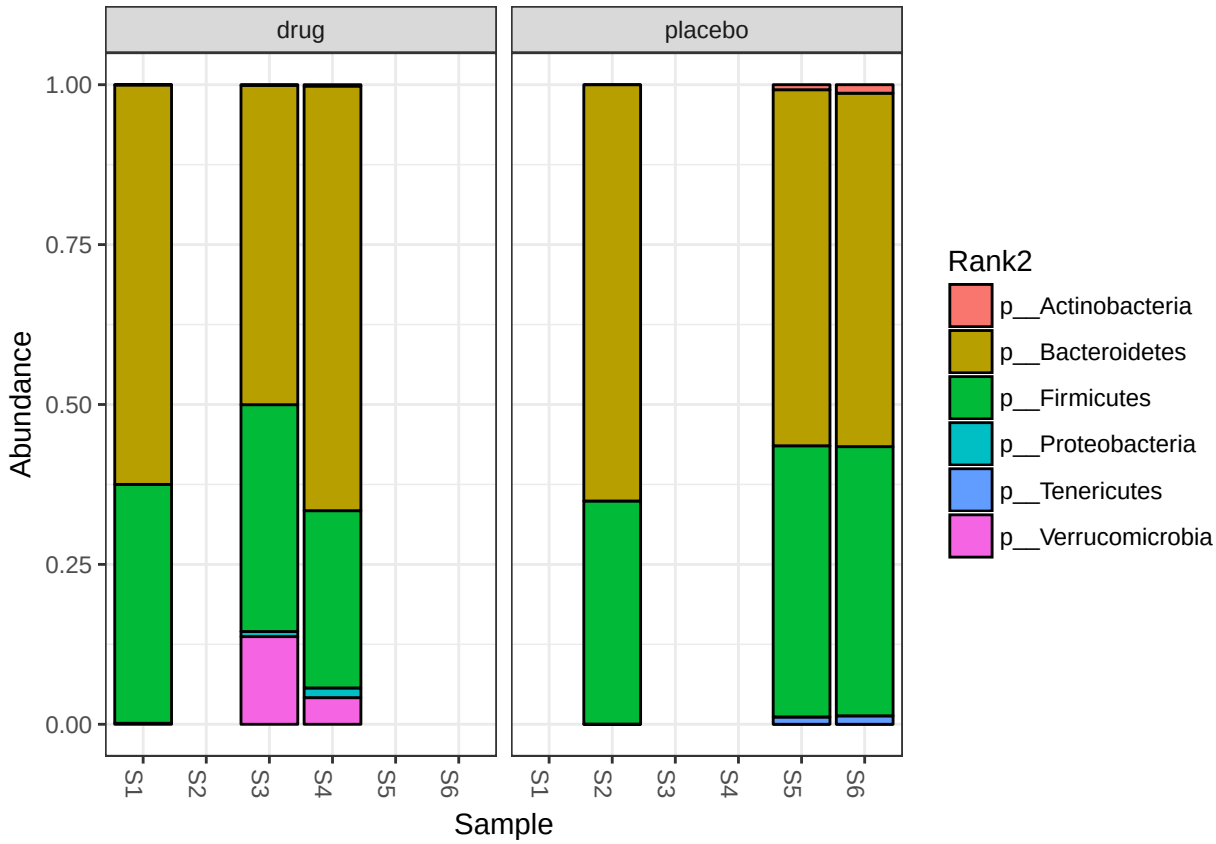



```
# now we plot the same tree with drug/placebo information
# Rank3 means at the bacterial Class level (Rank 2 would be phylum)
plot_tree(ex1, color = "Treatment", label.tips = "Rank3", ladderize = "left", justify = "left" , size =
```



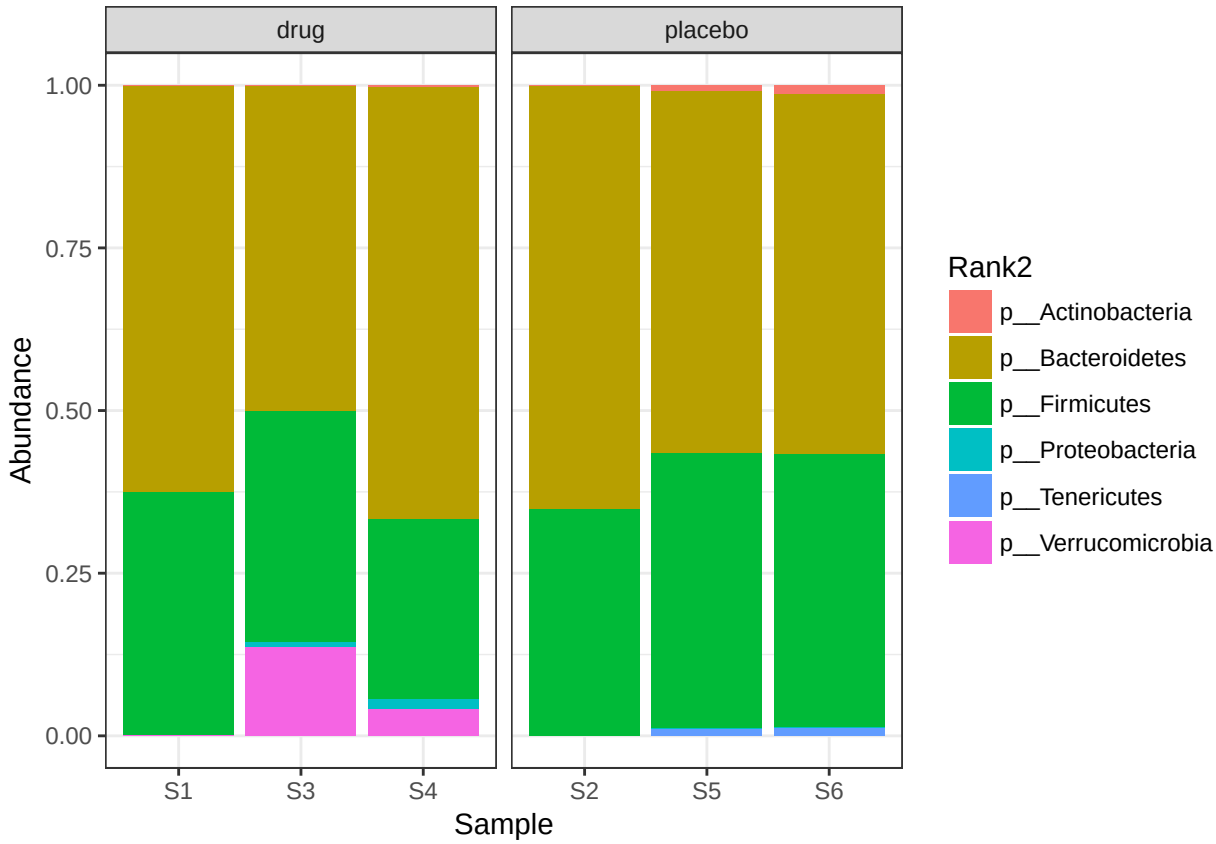

```
# what do we conclude? Verrucomicrobia appears to be more in drug group
```

```
# convert counts to relative abundance
psr  = transform_sample_counts(ps, function(x) x / sum(x) )
# filter out low abundance OTUs
fr = filter_taxa(psr, function(x) mean(x) > 1e-5, TRUE)
# plot bar
plot_bar(tax_glom(psr, "Rank2"), fill = "Rank2", facet_grid = ~Treatment)
```




```
# above graph okay but have ugly gaps!
# fix them now by using ggplot2
df = plot_bar(tax_glom(psr, "Rank2"), fill = "Rank2", facet_grid = ~Treatment)$data

ggplot(df, aes(Sample, Abundance, fill = Rank2)) +
  geom_bar(stat = 'identity') +
  facet_wrap(~Treatment, scales = "free_x")
```

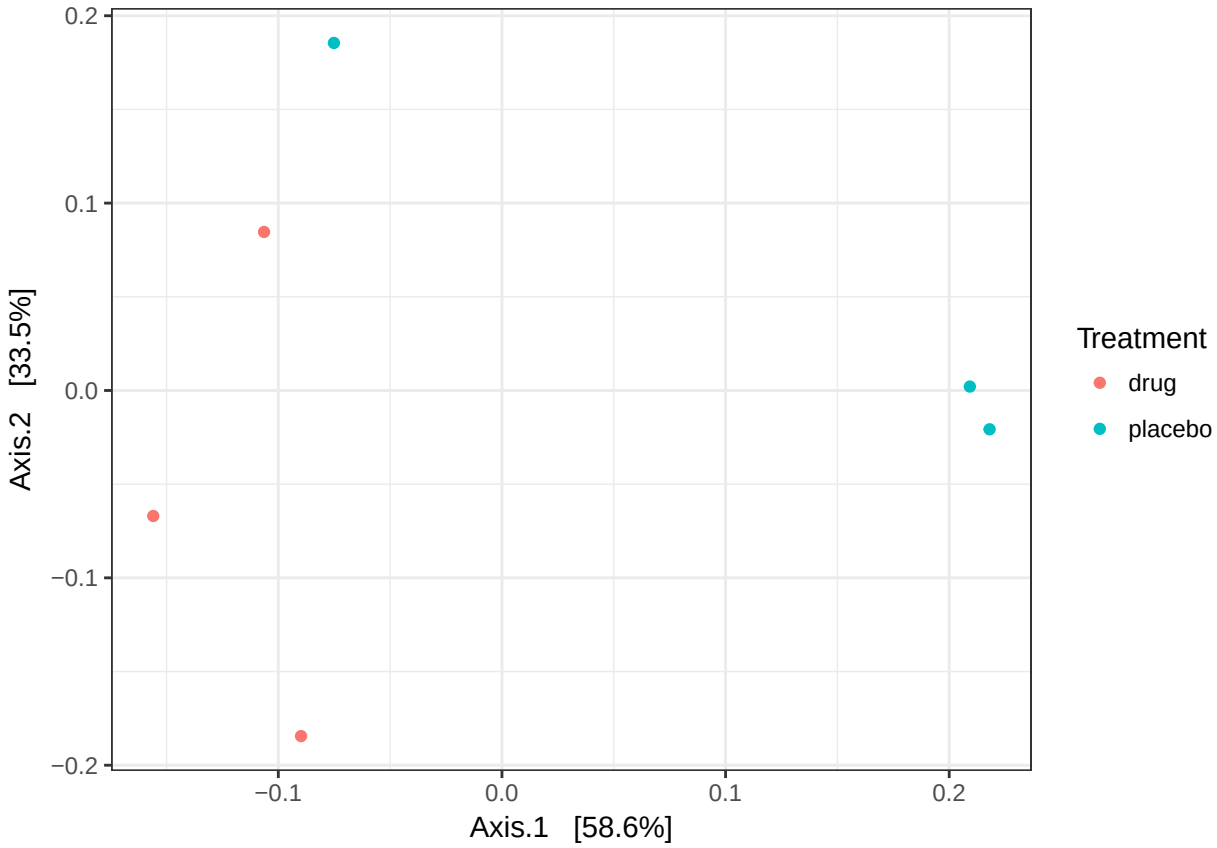



```
# this is publication-ready graphics!
```

```
# pca plots
pso <- ordinate(ps, "PCoA", "unifrac")
```

```
## Warning in UniFrac(physeq, ...): Randomly assigning root as -- OTU_58 -- in
## the phylogenetic tree in the data you provided.
```

```
plot_ordination(ps, pso, type="samples", color = "Treatment")
```

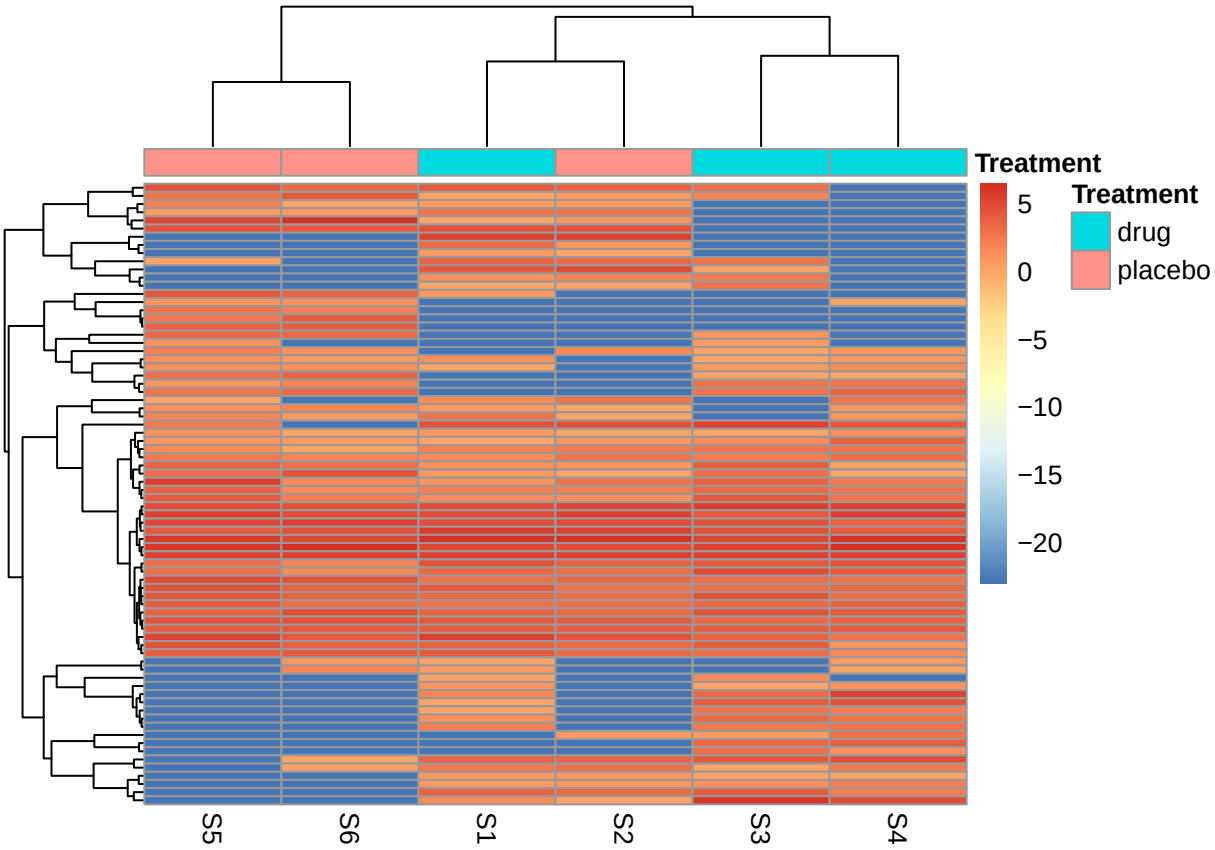



```
# plot a heatmap
# may need install.packages('pheatmap')
if (!require(pheatmap)){
  install.packages("pheatmap")
  library(pheatmap)
}
```

```
## Loading required package: pheatmap
```

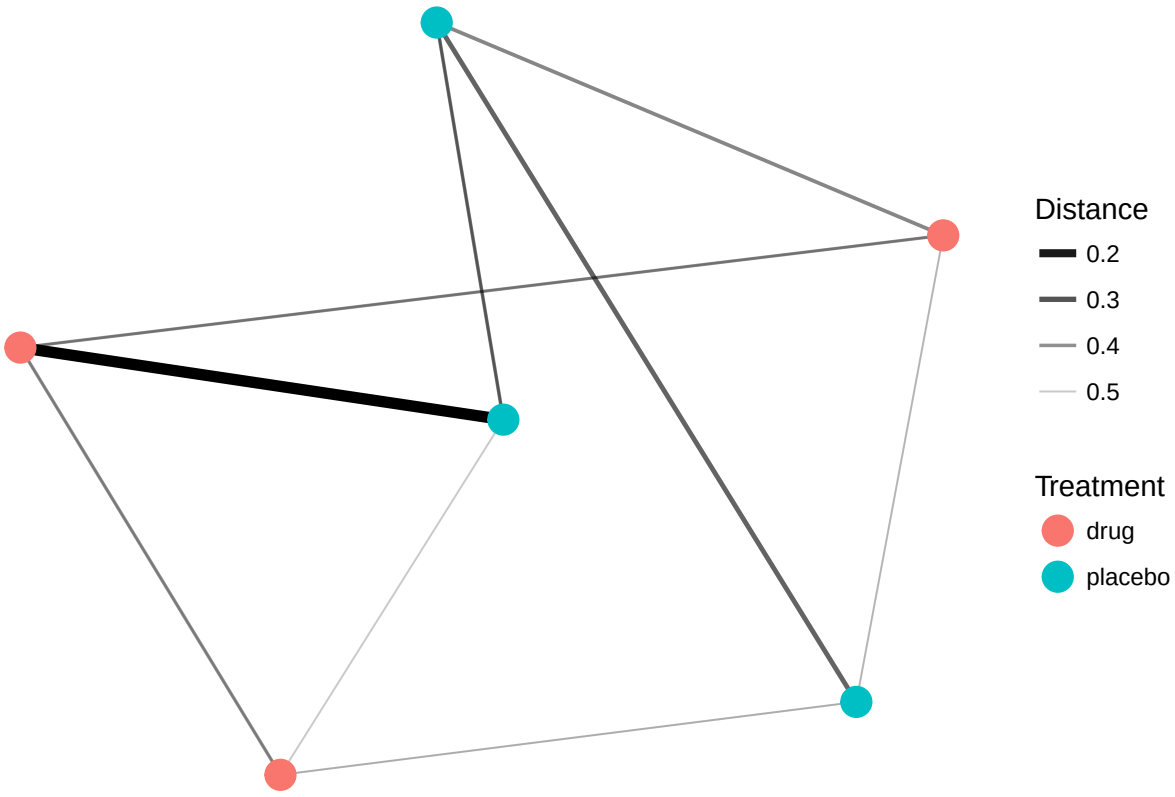
```
EPS = 1E-10 # a tiny number
mat <- data.frame(log(otu_table(ps) + EPS, 2),
                  check.names = FALSE)

pheatmap(mat,
         show_rownames=FALSE,
         show_colnames=TRUE,
         #treeheight_row=0,
         annotation_col=data.frame(sample_data(ps)[, "Treatment"])
         )
```




```
# we see three clusters in the heatmap
# it would be nice if we see two (drug, placebo)
```

```
# plot network
plot_net(ps, color = "Treatment")
```


Distance
0.2
0.3
0.4
0.5
Treatment
drug
placebo