

# Process BIOM Data

Team YOLO

## R Markdown for processing BIOM data

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Loaded BIOM data

## [1] 1 0 0 0 0 0

##           77.TS127 77.TS31.2 77.TS49 77.TS31 77.TS30 77.TS25 77.TS14 77.TS55
## 188753           1           2           1           1           4           4           2           1
## 181342           0           0           0           0           0           0           0           0
## 3589405          0           0           0           1           0           0           0           0
## 4467993          0           0           0           0           0           0           0           0
## 355102           0           2          66           6           9           0           2           1
## 4465907          0           6           3          51           7           3          12           1
##           77.TS27 77.TS21
## 188753           1           1
## 181342           0           0
## 3589405          0           0
## 4467993          0           0
## 355102           0          11
## 4465907          0           2

## Transposed data

##           188753 181342 3589405 4467993 355102
## 77.TS127           1           0           0           0           0
## 77.TS31.2          2           0           0           0           2
## 77.TS49            1           0           0           0          66
## 77.TS31            1           0           1           0           6
## 77.TS30            4           0           0           0           9
## 77.TS25            4           0           0           0           0
```

## Save data

```
write.csv(b_data_m_t, "/Users/hussainwani/Codeathon/Microbiome Data/OTUdata.csv")
```

## Get observation data

```
#' get observations
observation <- observation_metadata(biom_tbl)
head(observation)
```

```
##           taxonomy1      taxonomy2      taxonomy3      taxonomy4
## 188753 k__Bacteria    p__Firmicutes c__Clostridia o__Clostridiales
## 181342 k__Bacteria    p__Firmicutes c__Clostridia o__Clostridiales
## 3589405 k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales
## 4467993 k__Bacteria p__Actinobacteria c__Actinobacteria o__Actinomycetales
## 355102 k__Bacteria    p__Firmicutes c__Clostridia o__Clostridiales
## 4465907 k__Bacteria    p__Firmicutes c__Clostridia o__Clostridiales
##           taxonomy5      taxonomy6 taxonomy7
## 188753 f__Lachnospiraceae g__Blautia s__
## 181342 f__Clostridiaceae g__02d06 s__
## 3589405 f__Ruminococcaceae g__ s__
## 4467993 f__Actinomycetaceae g__Actinomyces s__
## 355102 f__Lachnospiraceae g__ s__
## 4465907 f__Lachnospiraceae g__Blautia s__
```

## Read Metadata

```
#' read metadata
metad <- read.csv("/Users/hussainwani/Codeathon/Microbiome Data/77_20180101-113930.txt",
                  header = TRUE, sep = "\t")

# If you want to save it
write.csv(metad, "/Users/hussainwani/Codeathon/Microbiome Data/MetaData.csv", row.names = FALSE)
```

## Combine metadata and OTUs

```
##      Row.names 188753 181342 3589405 4467993
## 1      77.TS1      0      0      0      0
## 2      77.TS1.2    0      0      0      0
## 3      77.TS10     0      0      0      0
## 4      77.TS10.2   0      0      0      0
## 5      77.TS100    1      0      0      0
## 6      77.TS100.2  0      0      0      0
```

## If you need only a few important columns

```
#' now we don't need all the phenotype information  
#' We can drop what we don't need  
cols <- c("sample_name", "age", "ancestry", "family", "obesitycat", "sex",  
          "twin_mother", "zygosity")  
comb_df_sub <- merge(b_data_m_t, metad[, cols], by.x = "row.names", by.y = "sample_name")  
head(comb_df_sub[, 1:5])
```

```
##      Row.names 188753 181342 3589405 4467993  
## 1      77.TS1      0      0      0      0  
## 2      77.TS1.2    0      0      0      0  
## 3      77.TS10     0      0      0      0  
## 4      77.TS10.2   0      0      0      0  
## 5      77.TS100    1      0      0      0  
## 6      77.TS100.2  0      0      0      0
```

```
write.csv(comb_df_sub, "/Users/hussainwani/Codeathon/Microbiome Data/Obese_Lean_Microbiome_Data_sub.csv")
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.