

Loading Microbiome Data

Back to [Table of Contents](#)

After you have [installed QIIME](#) and [generated an OTU table](#), you can load the OTU table into R.

Install biom package

```
install.packages('biom',repo='http://cran.wustl.edu')
```

Load biom package

```
library('biom')
```

Convert BIOM file to JSON format

If you have data in a “new” BIOM format (HDF5), you first need to convert to JSON format first.

```
# (on the command-line, in the course repo)  
# make a JSON-formatted OTU table for loading into R  
cd data/globalgut-66-adults  
biom convert -i otu_closed_ref/otu_table.biom -o otu_table_json.biom --to-json
```

Load global gut data using biom package

```
gg.otus.biom <- read_biom('../data/globalgut-66-adults/otu_table_json.biom')
```

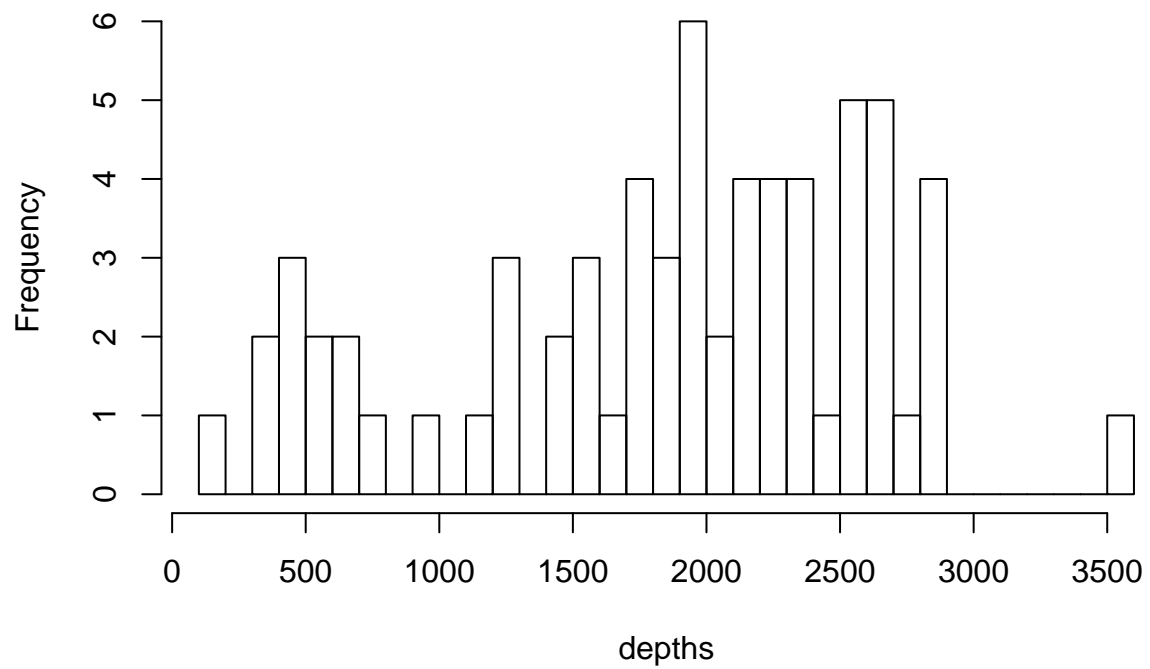
Extract data matrix (OTU counts) from biom table

```
gg.otus <- as.matrix(biom_data(gg.otus.biom))  
  
# transpose so that rows are samples and columns are OTUs  
gg.otus <- t(gg.otus)
```

Plot histogram of sample depths

```
depths <- rowSums(gg.otus)  
hist(depths,breaks=30)
```

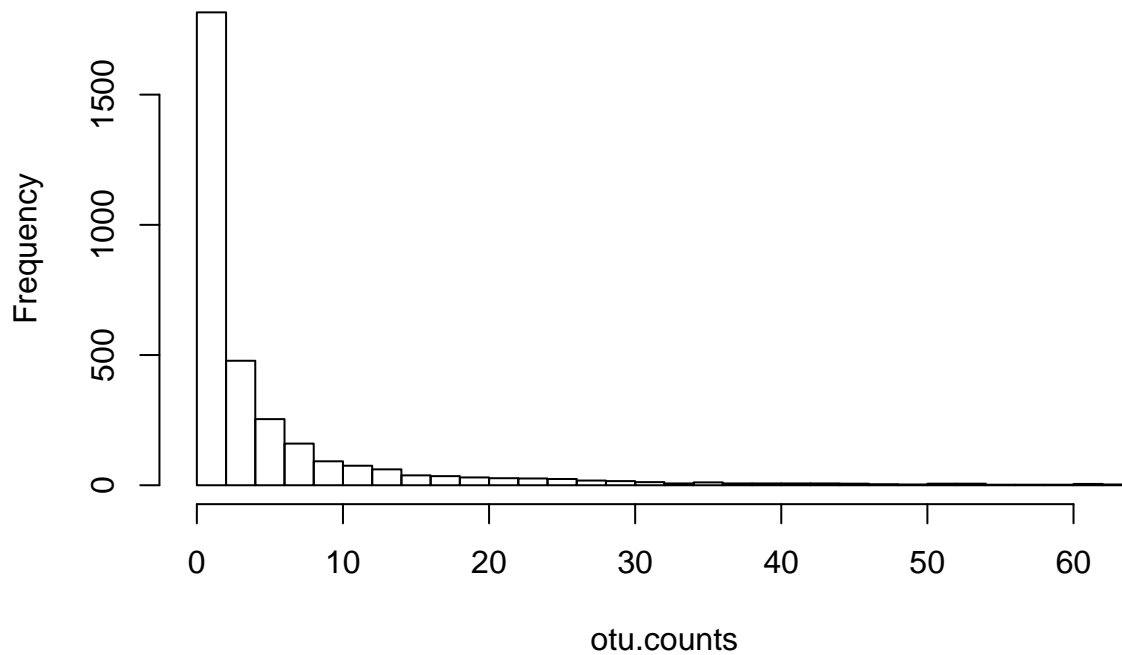
Histogram of depths



Plot histogram of OTU frequencies

```
otu.counts <- colSums(gg.otus > 0)
hist(otu.counts,breaks=30)
```

Histogram of otu.counts



Remove OTUs present in < 10% of samples

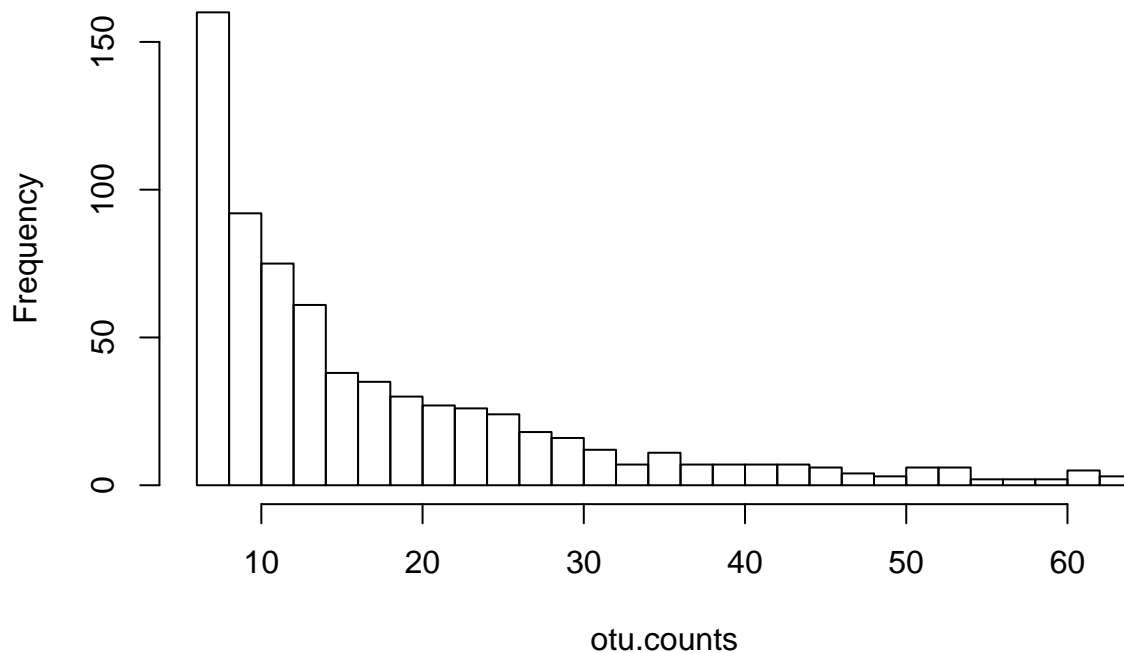
```
gg.otus <- gg.otus[,colMeans(gg.otus > 0) >= .1]
depths <- rowSums(gg.otus)
dim(gg.otus)
```

```
## [1] 66 699
```

Re-plot histogram of OTU frequencies now that we removed singletons

```
otu.counts <- colSums(gg.otus > 0)
hist(otu.counts,breaks=30)
```

Histogram of otu.counts



Remove any samples with very low depth

```
sort(depths)[1:10]
```

```
## Amz33eld.418866 Amz24chld.418412 Amz31adlt.418759 Amz7adltF.418395
##          125          336          361          398
## Amz6teen.418569 Amz23chld.418722 Amz6eldM.418356 Amz25chld.418466
##          408          460          480          563
## Amz5chldF1.418757 Amz6eldF.418597
##          587          638
```

```
gg.otus <- gg.otus[depths >= 1000,]
dim(gg.otus)
```

```
## [1] 53 699
```

Load mapping file

```
gg.map <- read.table('../data/globalgut-66-adults/map.txt', sep='\t', head=T, row=1, check=F, comment='')
```

Ensure that mapping file and OTU table contain the sample samples in the same order

```
sample.ids <- intersect(rownames(gg.otus), rownames(gg.map))

# might as well put the samples in alphabetical order
sample.ids <- sort(sample.ids)

# in R you can subset using sample IDs or numerical indices. Most languages only use indices.
gg.otus <- gg.otus[sample.ids,]
gg.map <- gg.map[sample.ids,]
dim(gg.otus)
```

```
## [1] 53 699
```

```
dim(gg.map)
```

```
## [1] 53 22
```