

Loading Microbiome Data

Install biom package

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

Load biom package

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

Load global gut data using biom package

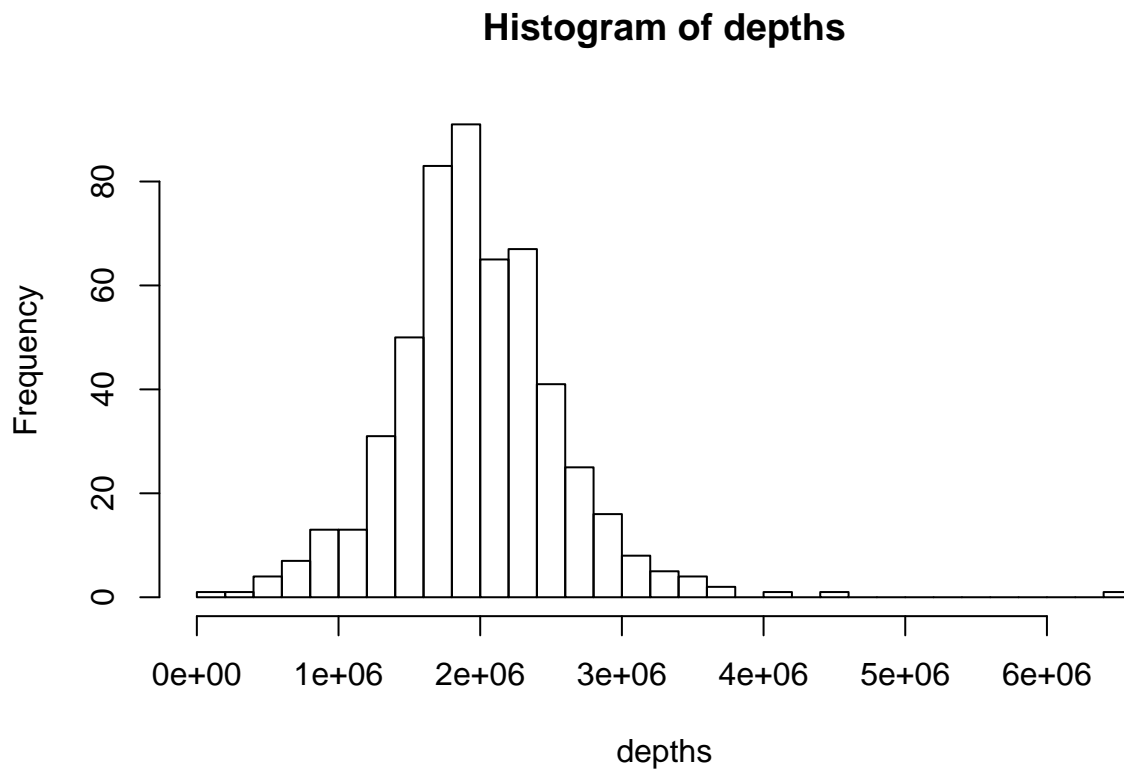
```
gg.otus.biom <- read_biom('../data/globalgut/otutable.biom')
```

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)
```

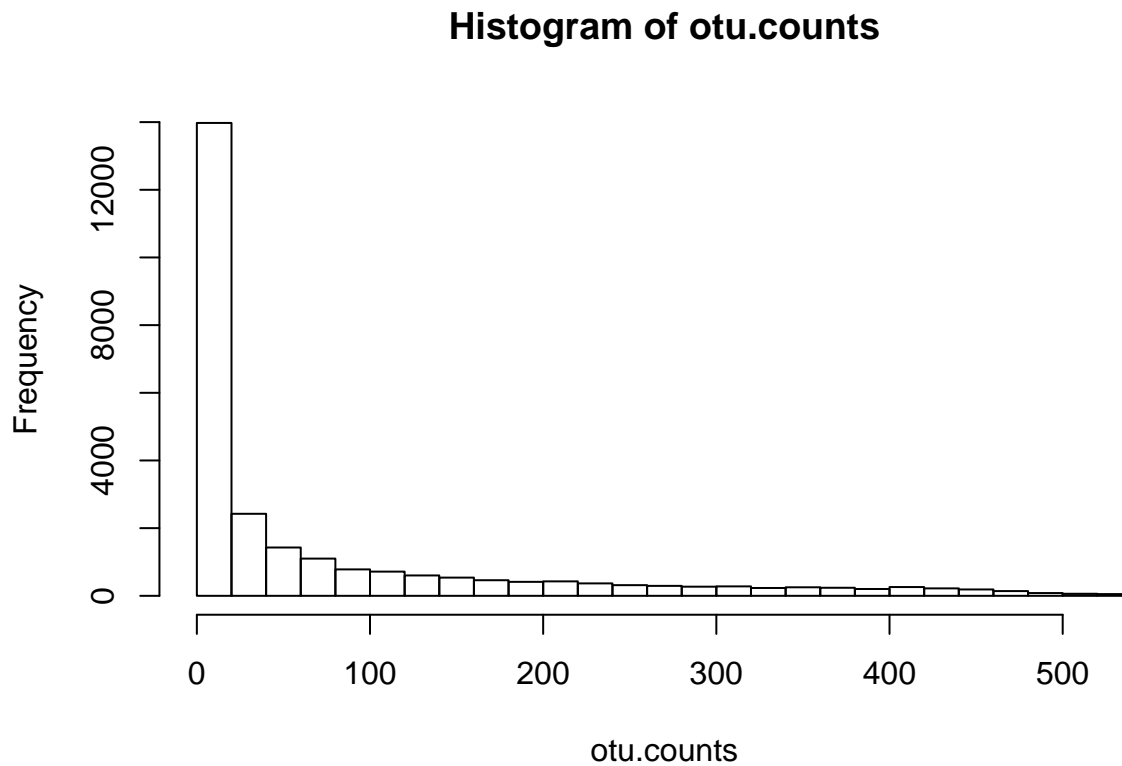
Plot histogram of sample depths

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



Plot histogram of OTU frequencies

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



Remove OTUs present in < 10% of samples

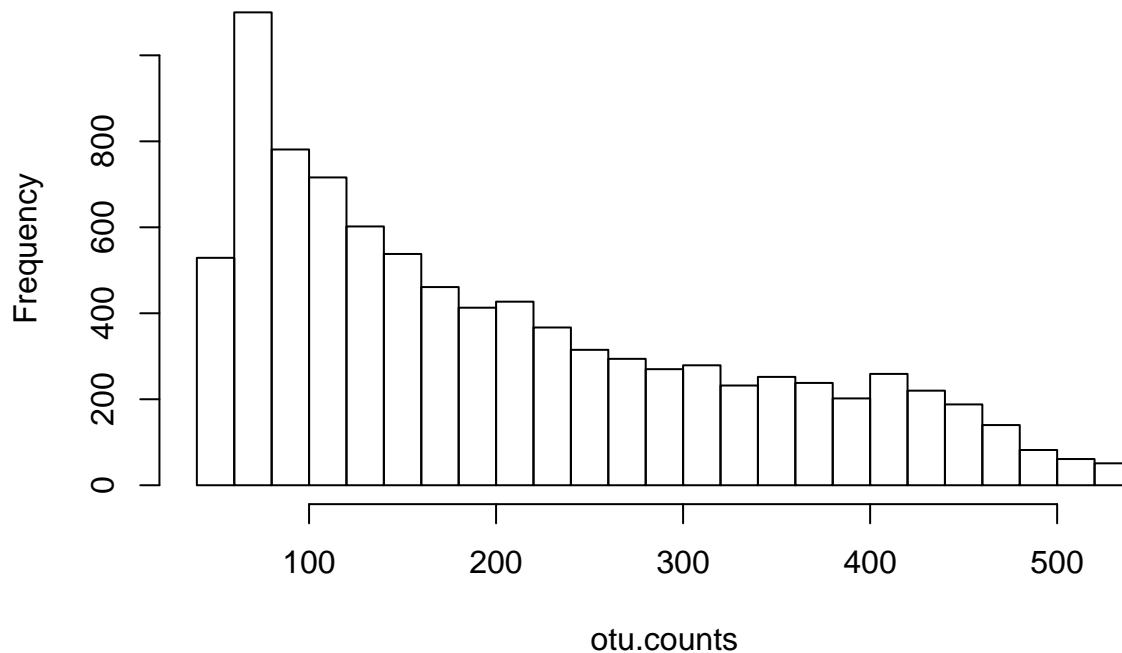
```
gg.otus <- gg.otus[,colMeans(gg.otus > 0) >= .1]
depths <- rowSums(gg.otus)
dim(gg.otus)
```

```
## [1] 530 9017
```

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]
```

```
## USygt34.T2.418554 USygt25.M.418835 AmzC3adltM.418849
## 1 309608 405006
## USygt3.M.418496 Amz33eld.418866 USinfTw1.2.418491
## 420719 504968 576975
## AmzC30adltF.418737 AmzC9adltM.418480 AmzC22chldM.418704
## 595770 631957 646988
## USygt12.F.418502
## 679153
```

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

```
## [1] 529 9017
```

Load mapping file

```
gg.map <- read.table('../data/globalgut/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] 529 9017
```

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

```
## [1] 529 63
```