Exploratory 16S data analysis

2019-01-11

Follow analyses presented in microbiomeSEQ package

http://userweb.eng.gla.ac.uk/umer.ijaz/projects/microbiomeSeq_Tutorial.html

Load data

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 7665 taxa and 9107 samples ]
sample_data() Sample Data:       [ 9107 samples by 9 sample variables ]
tax_table()   Taxonomy Table:    [ 7665 taxa by 7 taxonomic ranks ]
```

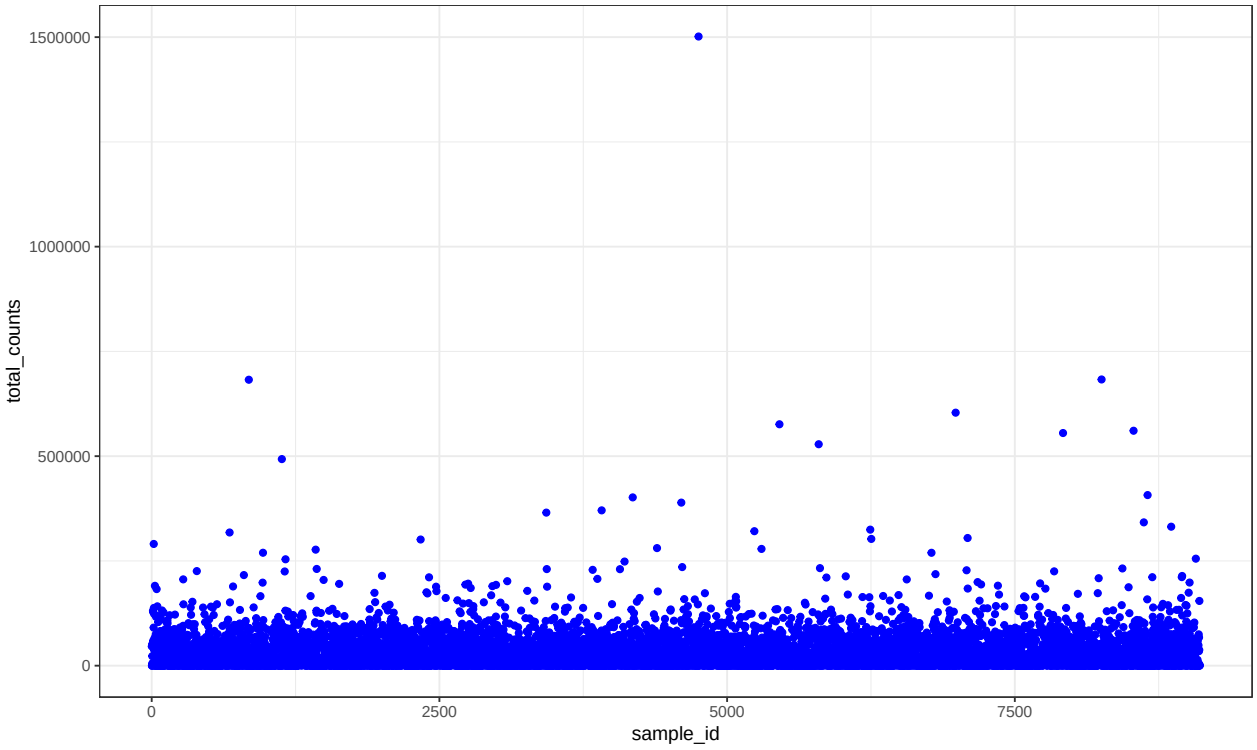
```
names(sample_data(mtx_phyloseq_16S))
```

```
[1] "sample_id"        "subject_id"       "sample_body_site"
[4] "visit_number"     "subject_gender"   "subject_race"
[7] "study_full_name"  "project_name"     "file"
```

Merge taxa based on phylogenetic information

```
# takes long time to run mtx_phyloseq_16S.genus = tax_glom(mtx_phyloseq_16S,
# 'Genus') plot_tree(mtx_phyloseq_16S.genus, color='SampleType',
# shape='Class', size='abundance')
```

Total number of counts for this sample data



Alpha Diversity Analysis

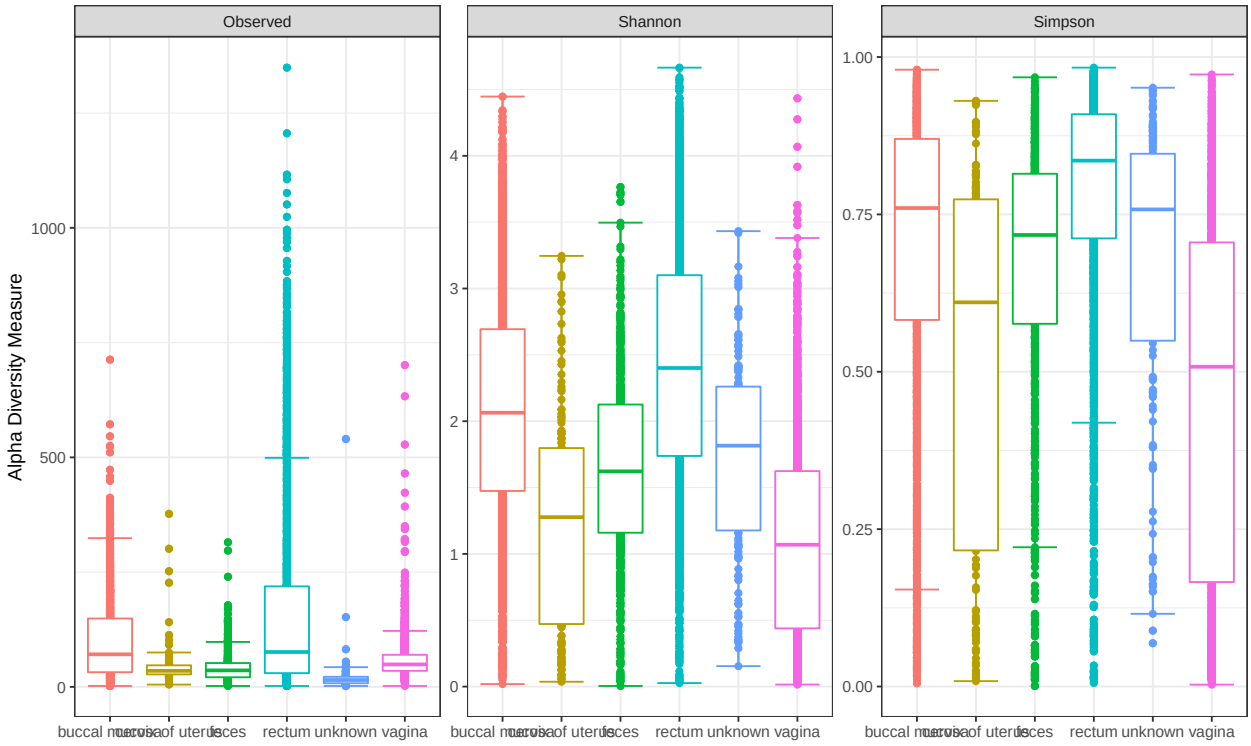
Alpha diversity measures the taxonomic variation within a sample and *phyloseq* provides a method, `plot_richness`, to plot various alpha diversity measures.

First a vector of richness (i.e. alpha diversity) measures is created to be passed to the `plot_richness` method.

We start by comparing alpha diversity metrics for each body site. Alpha diversity is estimated using raw microbiome counts.

```
richness_measures <- c("Observed", "Shannon", "Simpson")
```

```
alpha_div <- estimate_richness(mtx_phyloseq_16S, measures = richness_measures)
mtx_phyloseq_16S %>% plot_richness(x = "sample_body_site", color = "sample_body_site",
    measures = richness_measures) + stat_boxplot(geom = "errorbar") + geom_boxplot() +
    theme_bw() + theme(axis.title.x = element_blank(), legend.position = "none")
```



Principle Coordinates Analysis

Bray-Curtis distance multidimensional scaling plots based on proportions data. This one takes a lot of time to run, use a sample of 50 subjects instead.

```
mtx_phyloseq_16S.prop <- mtx_phyloseq_16S
row.sums <- apply(otu_table(mtx_phyloseq_16S.prop), 2, sum)
otu_table(mtx_phyloseq_16S.prop) <- t(t(otu_table(mtx_phyloseq_16S.prop))/row.sums)

ids <- sample(sample_data(mtx_phyloseq_16S.prop)$subject_id, 50)

mtx_phyloseq_16S.prop.samp <- subset_samples(mtx_phyloseq_16S.prop, subject_id %in%
    ids)
```

Remove unobserved taxa

```
mtx_phyloseq_16S.prop.samp %<>% taxa_sums() %>% is_greater_than(0) %>% prune_taxa(mtx_phyloseq_16S.prop

mtx_phyloseq_16S.prop.samp
```

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 4754 taxa and 915 samples ]
sample_data() Sample Data:       [ 915 samples by 9 sample variables ]
tax_table()   Taxonomy Table:    [ 4754 taxa by 7 taxonomic ranks ]
```

```
mtx_ordination <- ordinate(mtx_phyloseq_16S.prop.samp, method = "PCoA", distance = "bray")

mtx_phyloseq_16S.prop.samp %>% plot_ordination(mtx_ordination, color = "sample_body_site",
    shape = "sample_body_site") + theme_bw() + theme(legend.position = "bottom")
```