Data visualization using ecocomDP

Preparing and loading data

```
library(tidyverse)
library(neonUtilities)
devtools::load_all("C:/Users/savan/Documents/GitHub/savannahrayegonzales/ecocomDP")
#library(ecocomDP)
```

Link to NEON token tutorial: https://www.neonscience.org/resources/learning-hub/tutorials/neon-api-tokens-tutorial

```
NEON_TOKEN = Sys.getenv('NEON_TOKEN')
```

Load data using ecocomDP package and (optional) NEON token

```
inv <- ecocomDP::read_data(
  id = "neon.ecocomdp.20120.001.001",
  site = c('ARIK','MAYF'),
  startdate = "2017-06",
  enddate = "2020-03",
  token = NEON_TOKEN, #optional argument
  check.size = FALSE)
```

```
## Finding available files
##   |                                                                      |
##
## Downloading files totaling approximately 0.959824 MB
## Downloading 19 files
##   |                                                                      |
##
## Unpacking zip files using 1 cores.
## Stacking operation across a single core.
## Stacking table inv_fieldData
## Stacking table inv_persample
## Stacking table inv_pervial
## Stacking table inv_taxonomyProcessed
## Stacking table inv_taxonomyRaw
## Copied the most recent publication of validation file to /stackedFiles
## Copied the most recent publication of categoricalCodes file to /stackedFiles
## Copied the most recent publication of variable definition file to /stackedFiles
## Finished: Stacked 5 data tables and 3 metadata tables!
## Stacking took 0.8730092 secs
```

Plotting

Graph frequencies of each taxon rank. Produces a bar graph displaying which rank is most common.

```
ecocomDP::plot_taxa_rank(inv[[1]]$tables$observation,
               inv[[1]]$tables$taxon,
               inv[[1]]$tables$observation$location_id)
```

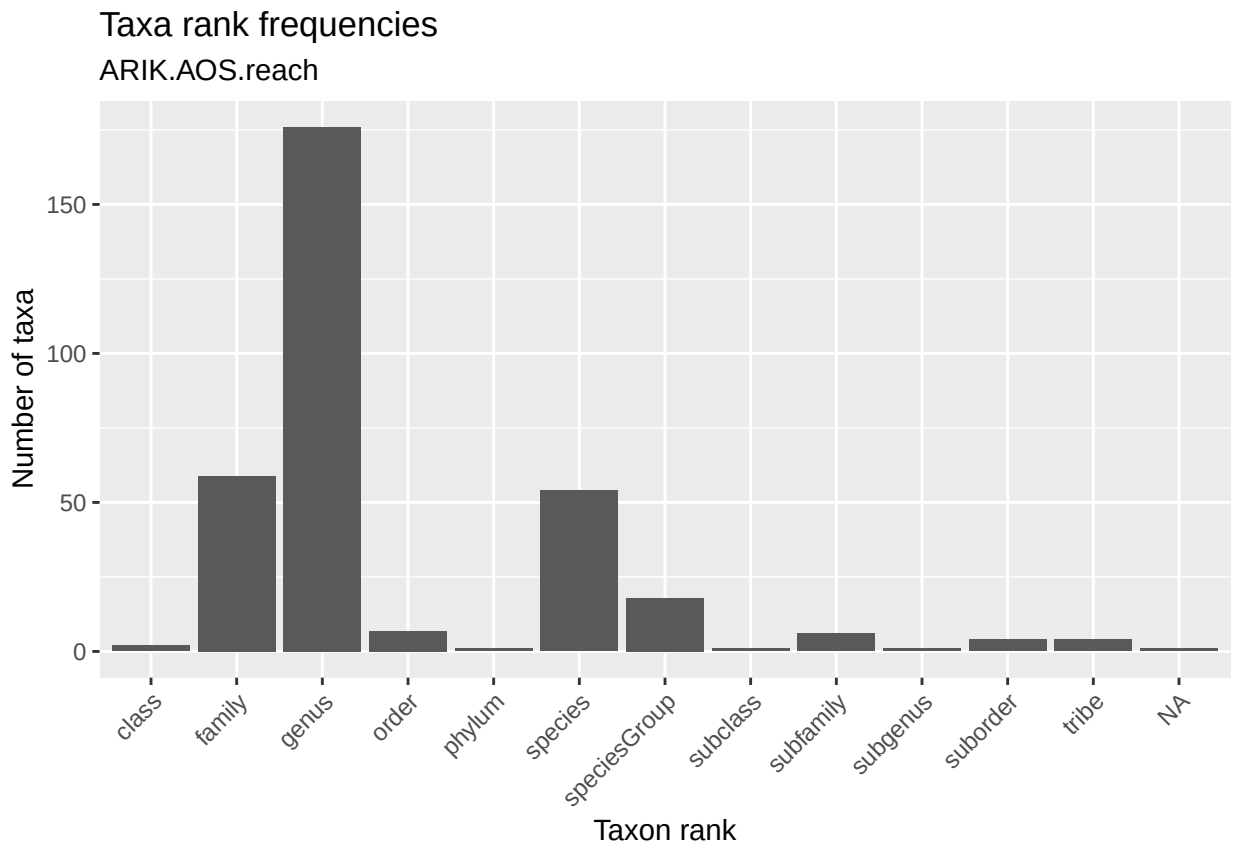


Figure 1: Over both sites, the taxon rank genus is most commonly recorded.

Same as above, but separate into bar graphs for each site.

```
ecocomDP::plot_taxa_rank_by_site(inv[[1]]$tables,
                      inv[[1]]$tables$observation$location_id)
```

Stacked taxa by site, with argument rank="order" The frequencies for each observation when taxon_rank=order is plotted. This doesn't include more specific ranks (ex: doesn't include the order when the observation's taxon_rank=species, despite that species belonging to the same order).

```
ecocomDP::plot_stacked_taxa_by_site(inv[[1]]$tables,
                         inv[[1]]$tables$observation$location_id,
                         "order")
```

One way to improve this plot is by ordering the bars' sum from greatest to least, rather than only ordering the red bars from the first site.

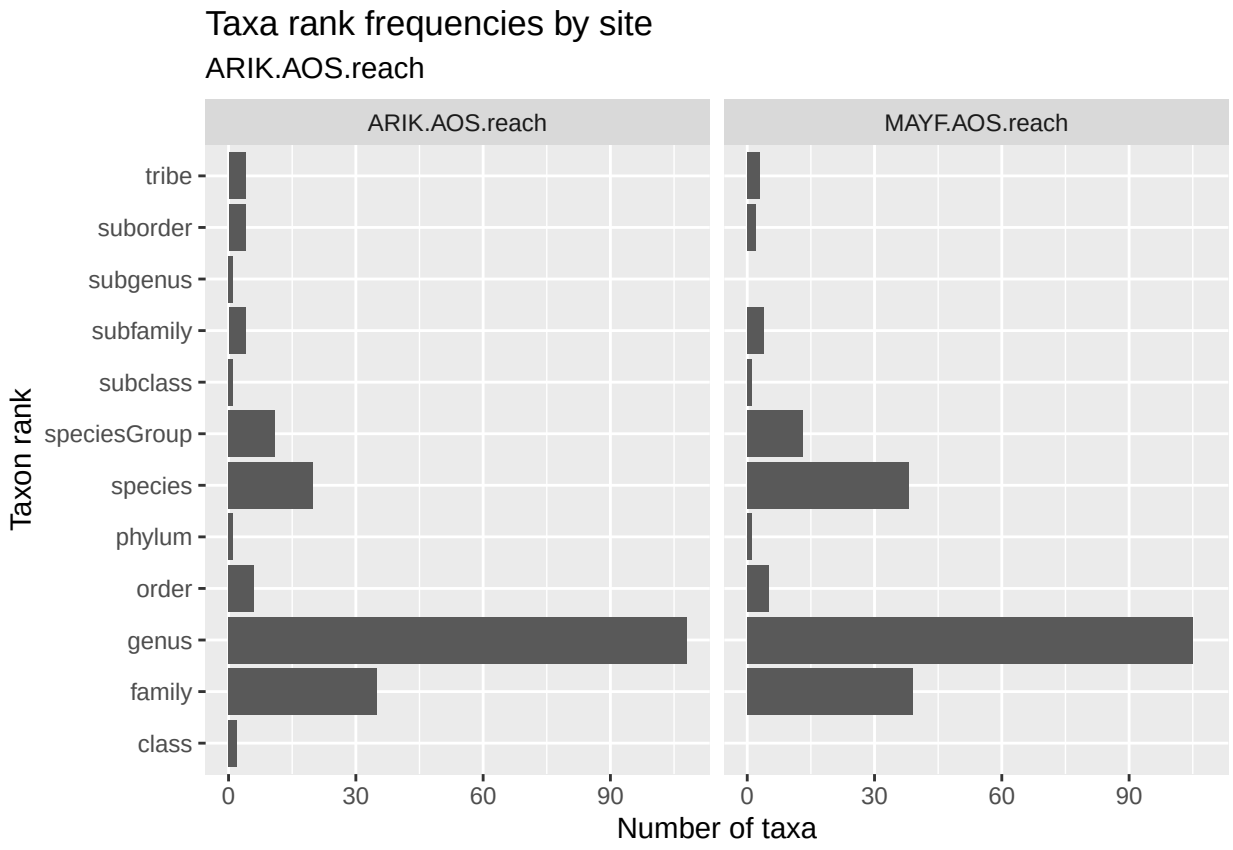


Figure 2: The taxon rank genus is most commonly recorded at both sites individually.

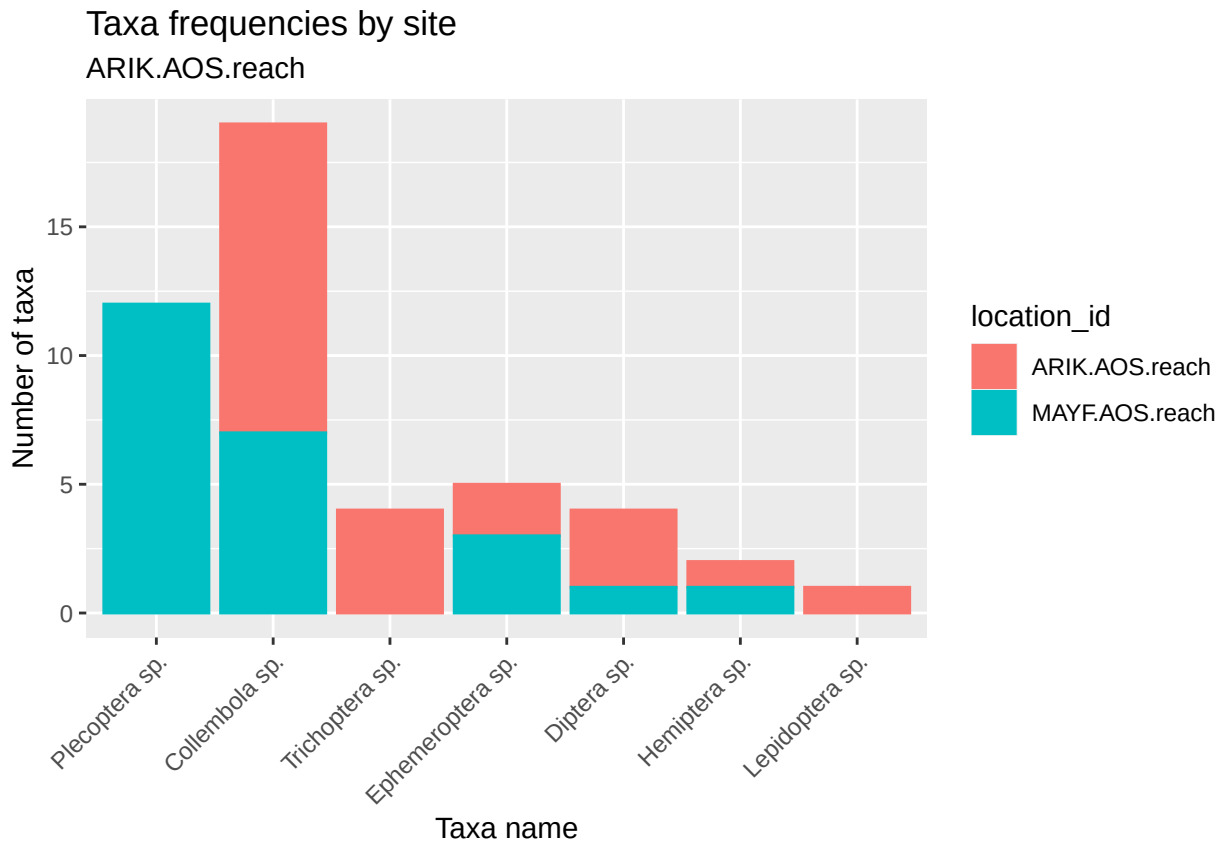


Figure 3: Collembola sp. is the most commonly recorded order at both at the first site (ARIK) and overall. Plecoptera is the most commonly recorded order at the second site (MAYF).

Similar to figure above, but given the argument rank="family"

```
ecocomDP::plot_stacked_taxa_by_site(inv[[1]]$tables,
                           inv[[1]]$tables$observation$location_id,
                           "family")
```

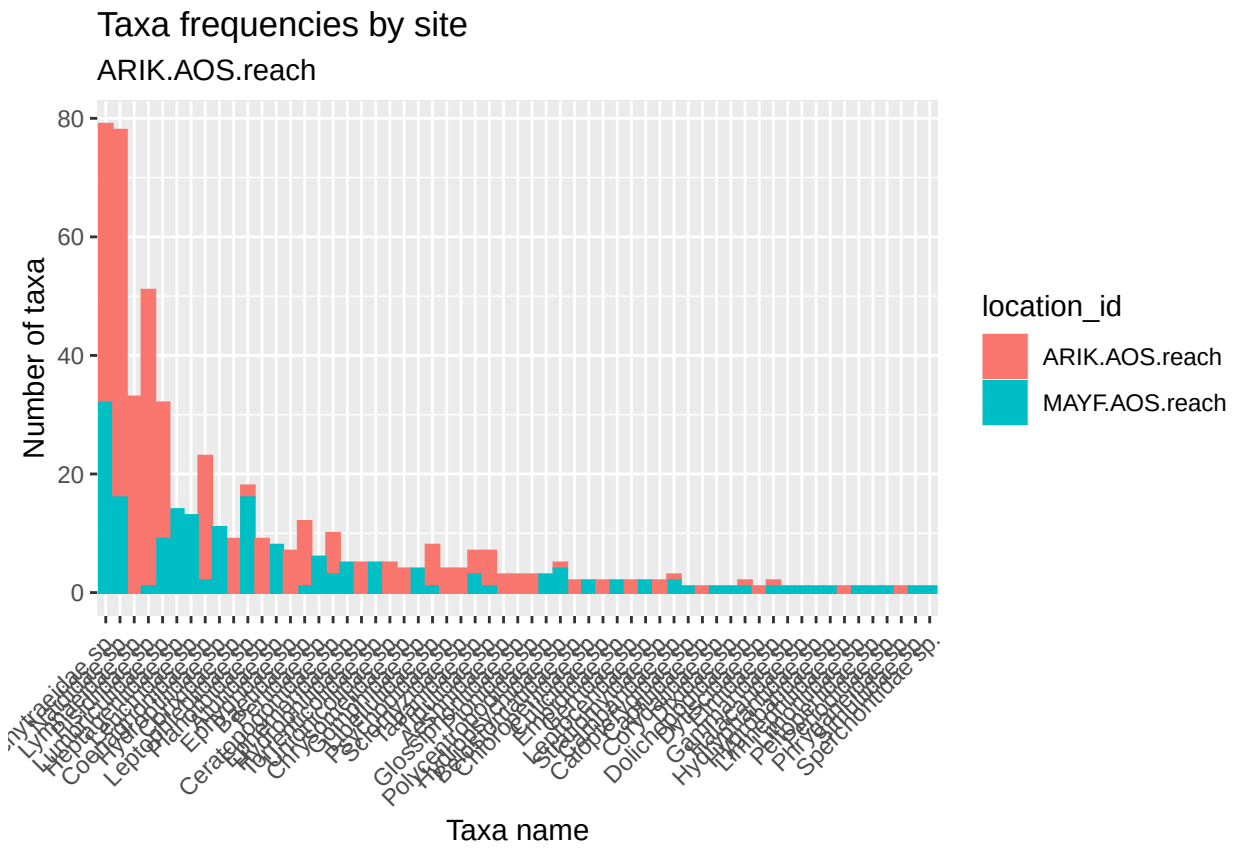
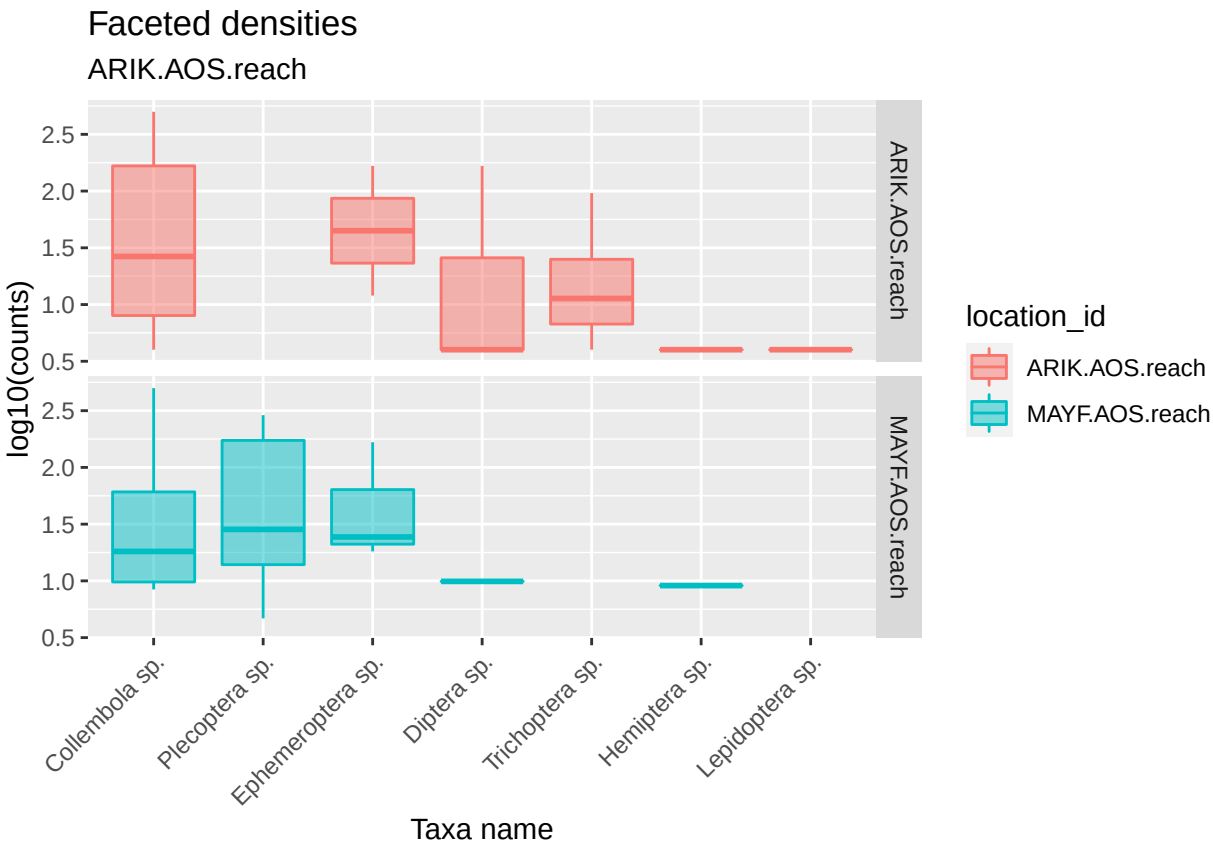


Figure 4: Most commonly recorded families at both sites

Faceted densities plot, shows averages of each taxa from a specified rank. Only includes when taxon_rank=order, not any more specific ranks.

```
ecocomDP::plot_faceted_densities(inv[[1]]$tables,
                         inv[[1]]$tables$observation$location_id,
                         "order")
```

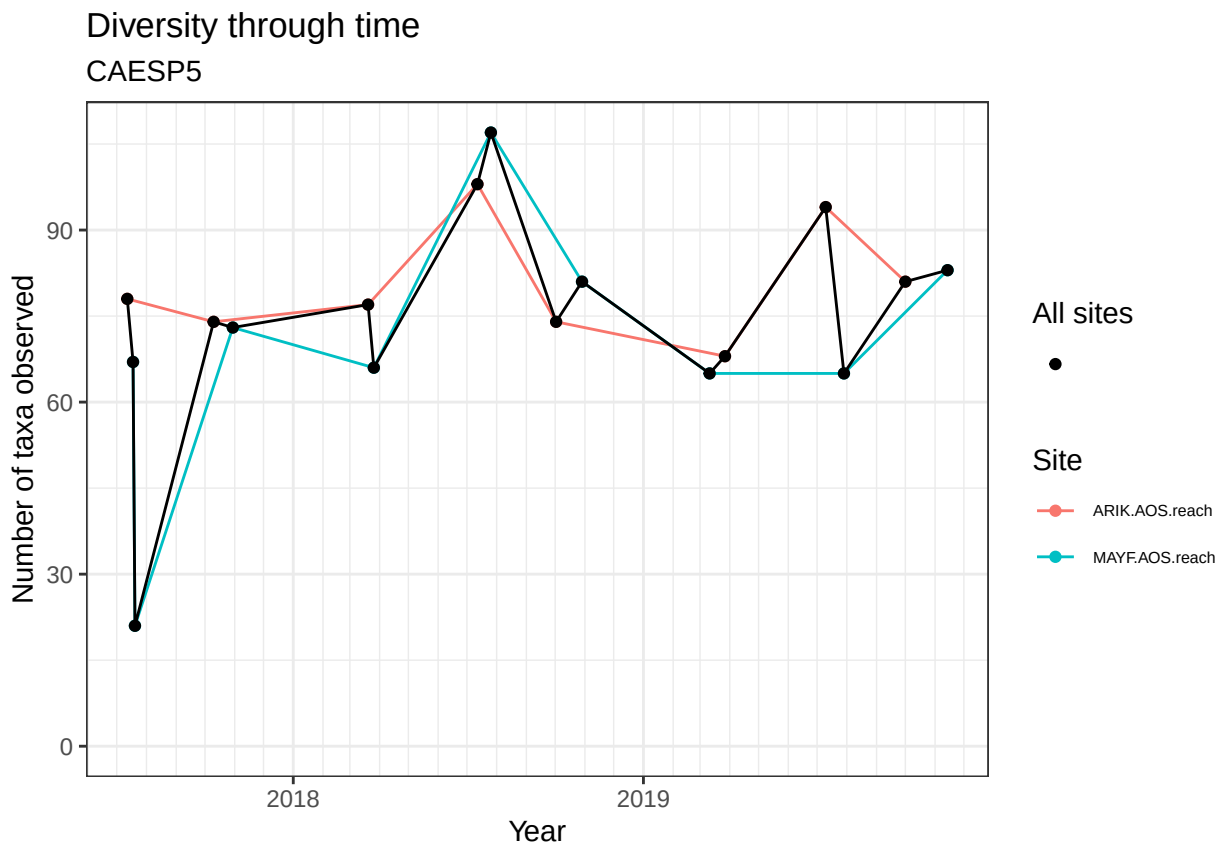


Need to include what the specified rank is in the final plot - either on x-axis label or in title.

Existing plotting functions in the ecocomDP package

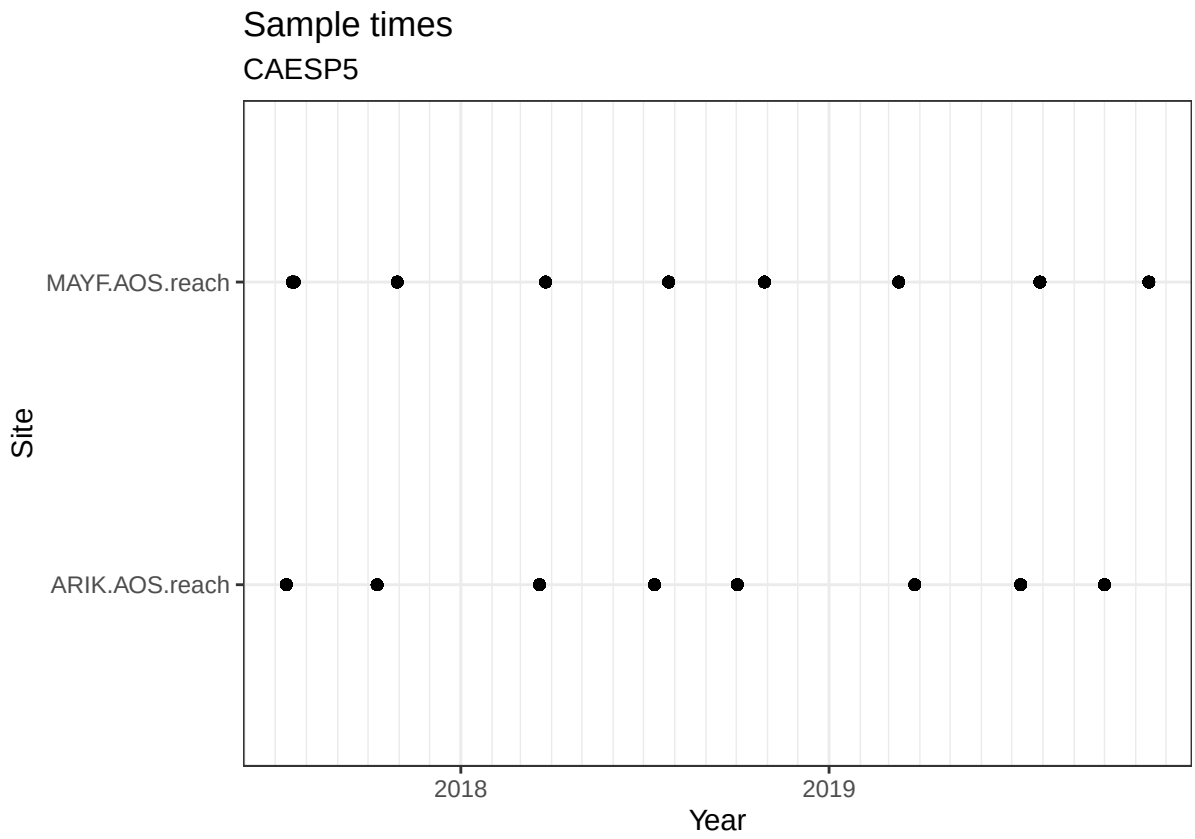
Explain arguments for each of these 4 functions.

```
ecocomDP::plot_taxa_diversity(inv[[1]]$tables$observation, inv[[1]]$tables$observation$taxon_id)
```



Different sites and times in the data set.

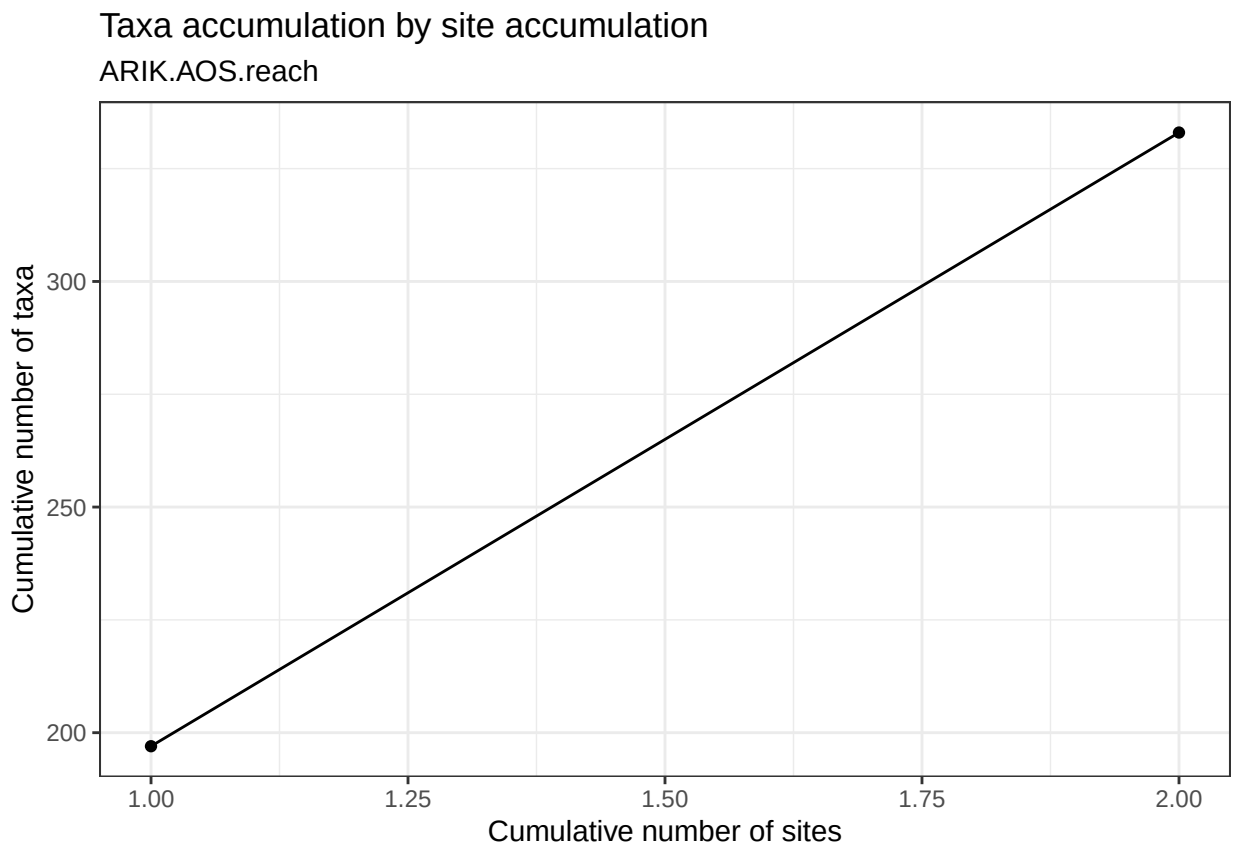
```
ecocomDP::plot_taxa_sample_time(inv[[1]]$tables$observation, inv[[1]]$tables$observation$taxon_id)
```



One way to potentially improve this plot is by making each dot size bigger when more observations are present at a particular site and time.

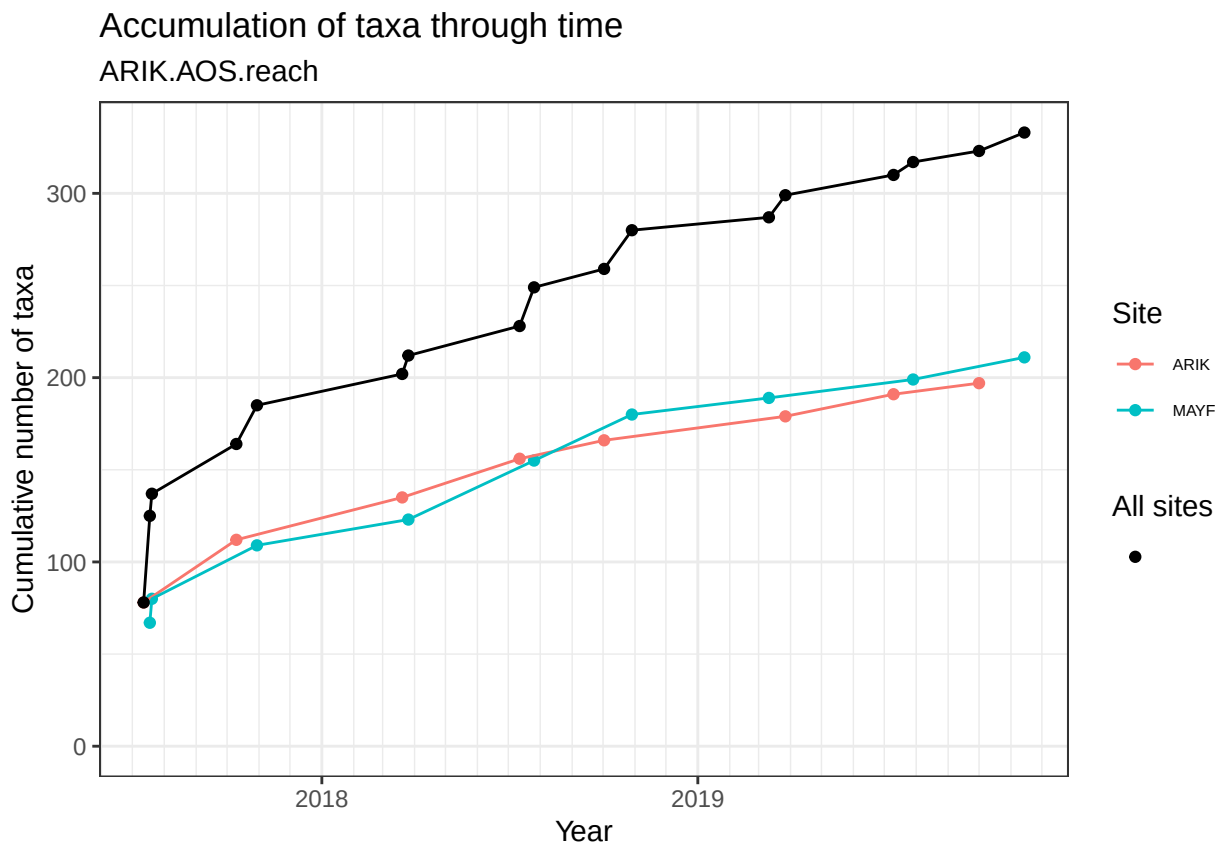
Number of taxa counted as the number of sites increases. This example only includes 2 sites, so looks correct, only 2 data points.

```
ecocomDP::plot_taxa_accum_sites(inv[[1]]$tables$observation, inv[[1]]$tables$observation$location_id)
```



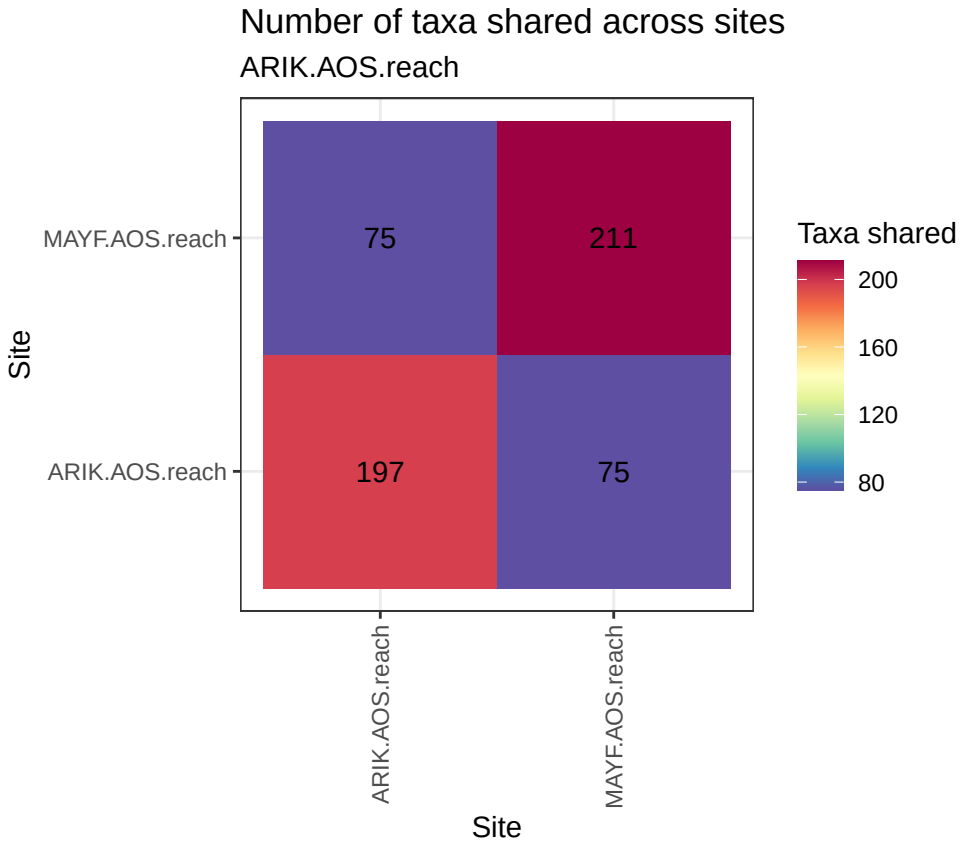
Number of taxa counts at each site over time. The curves for the two sites appear correct, but the curve representing all sites does not look like what is expected; the black line should show the summation of both sites.

```
ecocomDP::plot_taxa_accum_time(inv[[1]]$tables$observation, inv[[1]]$tables$observation$location_id)
```



Matrix-like plot representing the taxa shared between all sites.

```
ecocomDP::plot_taxa_shared_sites(inv[[1]]$tables$observation, inv[[1]]$tables$observation$location_id)
```


Number of taxa shared across sites
ARIK.AOS.reach
MAYF.AOS.reach
ARIK.AOS.reach
75
211
197
75
Site
ARIK.AOS.reach
MAYF.AOS.reach
Site
Taxa shared
200
160
120
80