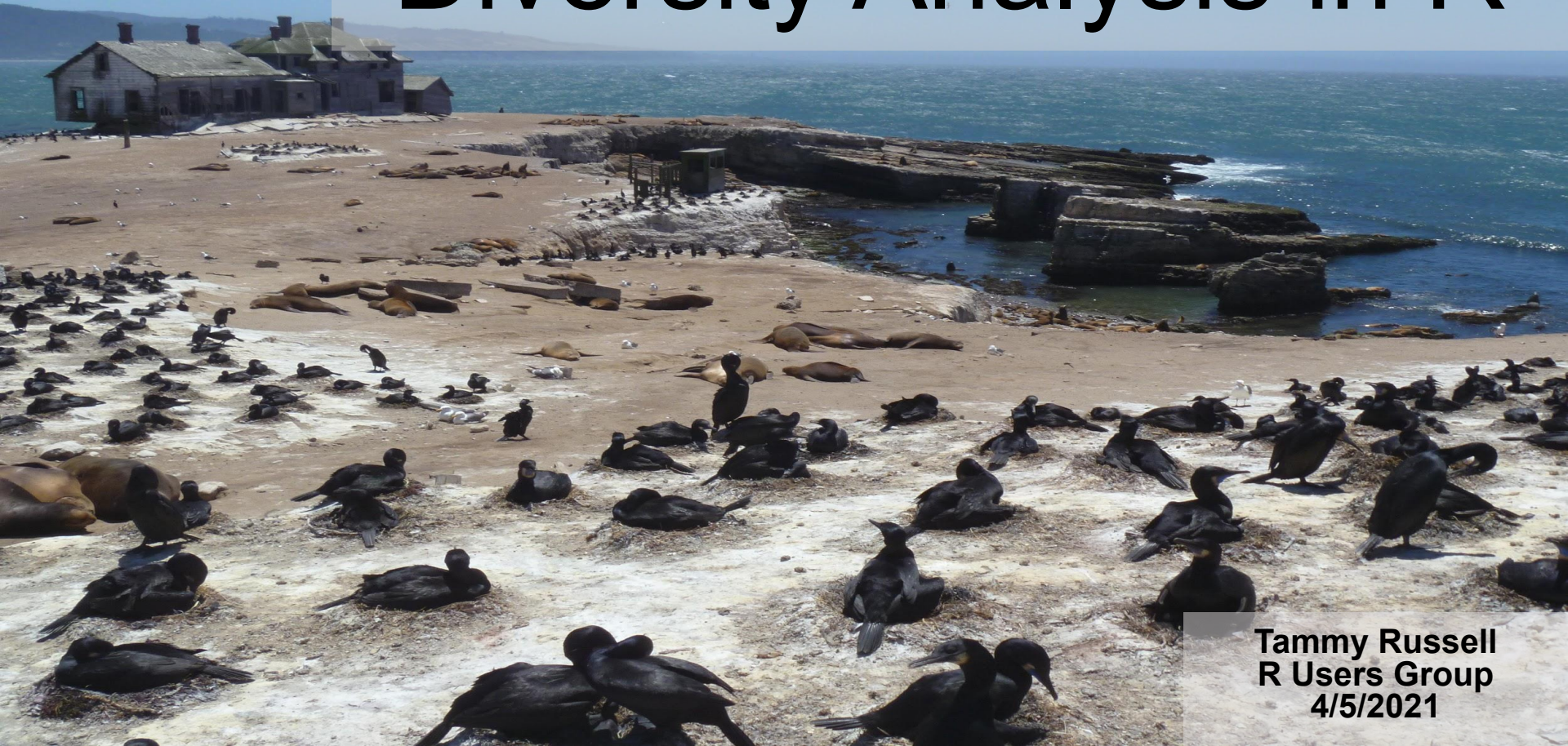


# Diversity Analysis in R



Tammy Russell  
R Users Group  
4/5/2021

# Measurements of Diversity

Species Richness: Number of species

Species Evenness: Relative abundances of species

Diversity Indices: Incorporates both species richness and evenness

- Shannon Index ( $H'$ )
  - Quantifies the uncertainty in predicting the species identity of an individual that is taken at random
  - The higher the index = more diverse (higher uncertainty)
- Simpson's Index ( $D$ )
  - An estimate of the probability that in the entire community you sampled, two randomly drawn individuals would be the same species
  - Heavily skewed by the dominant species
  - 0 is high, 1 is low, so usually the inverse or  $1-D$  is used

# Measurements of Diversity

But what if you want to compare the diversity between sites?

- Same sample size or effort? Use indices!
- Different sample size/effort
  - Create rarefaction curves for abundance or sample size
  - Compare estimated number of species based on the same sample size (min sample or extrapolation)
- Effective Species Number
  - Indices aren't linear, so not directly comparable (2.4 vs 1.2 is not twice as diverse)
  - Effective species number (Hill numbers)
    - The number of equally-common species required to give a particular value of an index

# Measurements of Diversity

Outline for today:

- Vegan package
  - Rarefaction
  - Shannon and Simpson indices
  - Effective species number
- iNEXT
  - Hill numbers (effective species)

Other Packages that may be useful:

-preseqR	-betapart
-abdiv	-devtools
-simboot	-prabclus
-BiodiversityR	

# Diversity Analysis

Data Format for `vegan`:

- species abundance (columns) for each sample or region (rows)

Example data: `dune` or `BCI`

	Abarema.macradenia	Vachellia.melanoceras	Acalypha.diversifolia	Acalypha.macrostachya
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0

# Diversity Analysis

## Species Richness

```
library(vegan)
```

```
# Load Sample data below
```

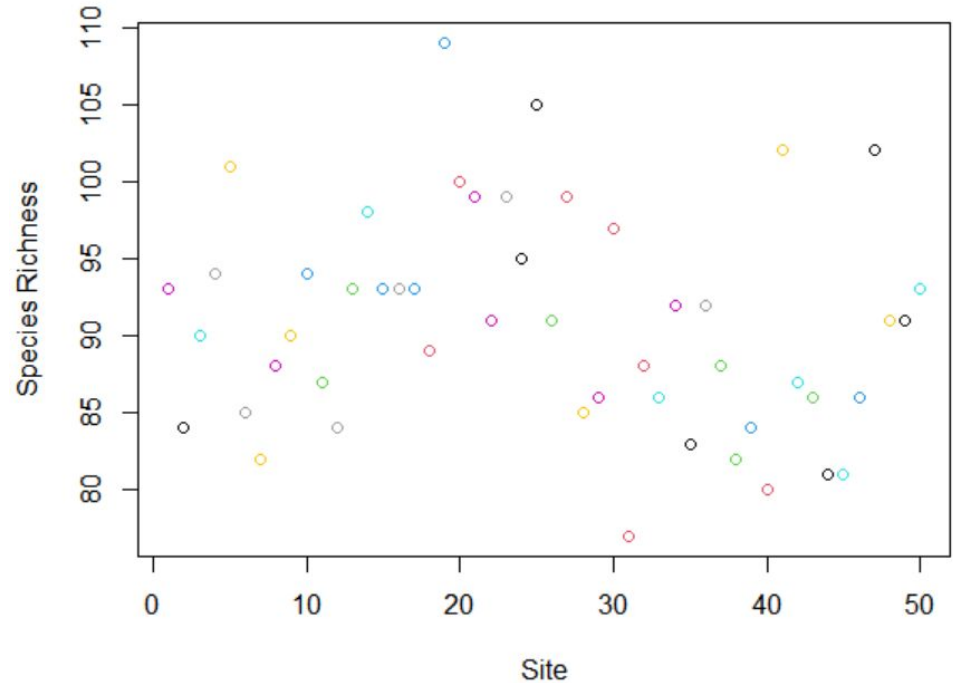
```
data(BCI)
```

```
# calculate species richness per site
```

```
test_species_richness <- specnumber(BCI)
```

```
> head(test_species_richness)
```

1	2	3	4	5	6
93	84	90	94	101	85



# Diversity Analysis

## Did we do a good job sampling?

- Use rarefaction to see if you've gotten close to the asymptote. Can also use to look at expected number of species per number sampled to compare areas of different sampling sizes

Rarefaction: Use when accumulating individuals instead of sites

```
# Number of INDIVIDUALS per site
```

```
raremax <- min(rowSums(BCI))
```

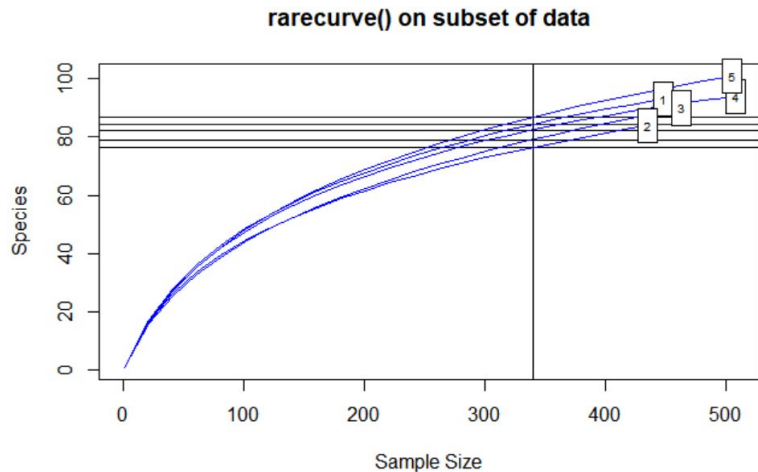
```
# rarefy, w/ raremax as input (?)
```

```
Srare <- rarefy(BCI, raremax)
```

```
# Plot a subset of that data
```

```
rarecurve(BCI[1:5,], step = 20, sample = raremax, col =  
"blue", cex = 0.6,
```

```
main = "rarecurve() on subset of data")
```



# Diversity Analysis

## Species Accumulation Curves

-Use when plotting sites (the overall sample)

### # Species accumulation curves

```
test_BCI_accum <- specaccum(BCI)
```

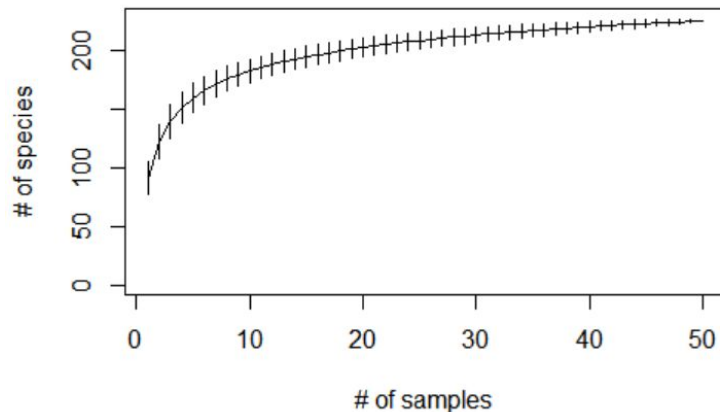
```
plot(test_BCI_accum, xlab = "# of samples", ylab = "# of species")
```

```
# Find the expected sp rich for site 1
```

```
test_BCI_accum[[4]][1]
```

```
# Find the sd for site 1
```

```
test_BCI_accum[[5]][1]
```



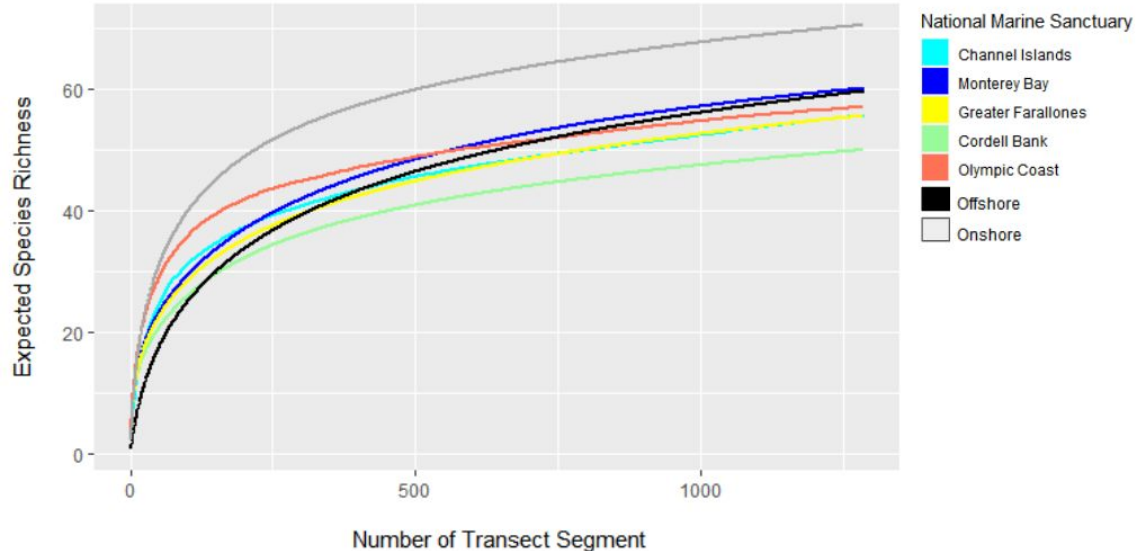


# Diversity Analysis

## Species Accumulation Curves

- Can be used to compare expected species richness between areas with different sampling effort.

Region	Species Richness	Estimated Richness
CINMS	57	55.14 +/- 1.26
MBNMS	80	60.02 +/- 2.43
CBNMS	51	49.93 +/- 0.856
GFNMS	66	55.67 +/- 2.279
OCNMS	57	57 +/- 0
CA Current	114	64.62 +/- 3.21



# Diversity Analysis

## Calculating Shannon Index

```
### Calculate Shannon Index
```

```
BCI_shannon <-diversity(BCI, index="shannon")
```

```
# can also omit the index since Shannon is the default
```

## Calculating Simpson Index

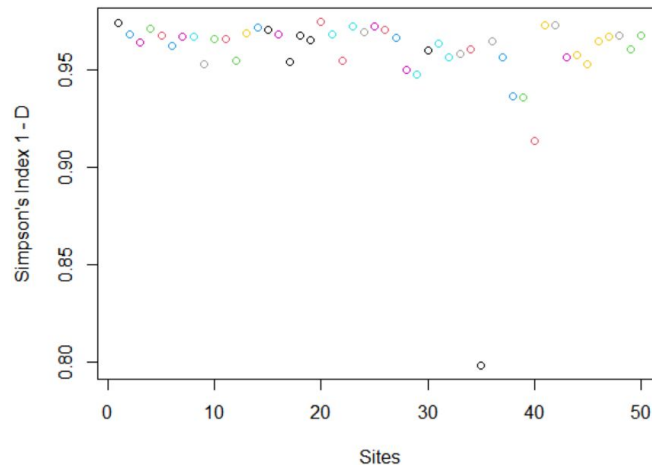
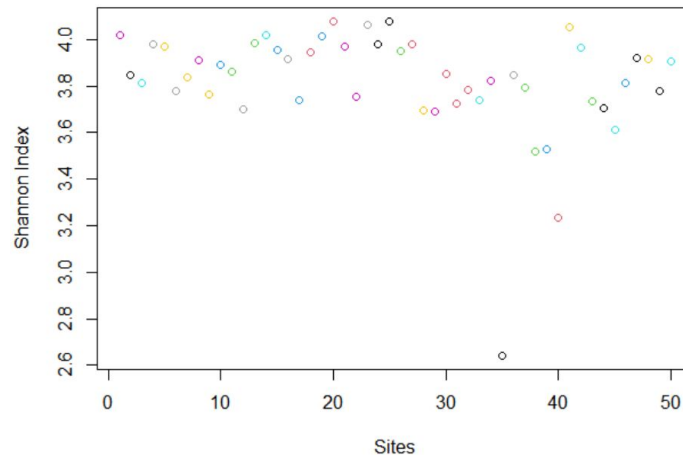
```
### Calculate Simpson Index
```

```
BCI_simpson <-diversity(BCI, index="simpson")
```

```
# 1-D so the closer to zero, the less diverse
```

```
BCI_simpson <-diversity(BCI, index="invsimpson")
```

```
# This returns 1/D (also closer to zero, less diverse)
```



# Diversity Analysis

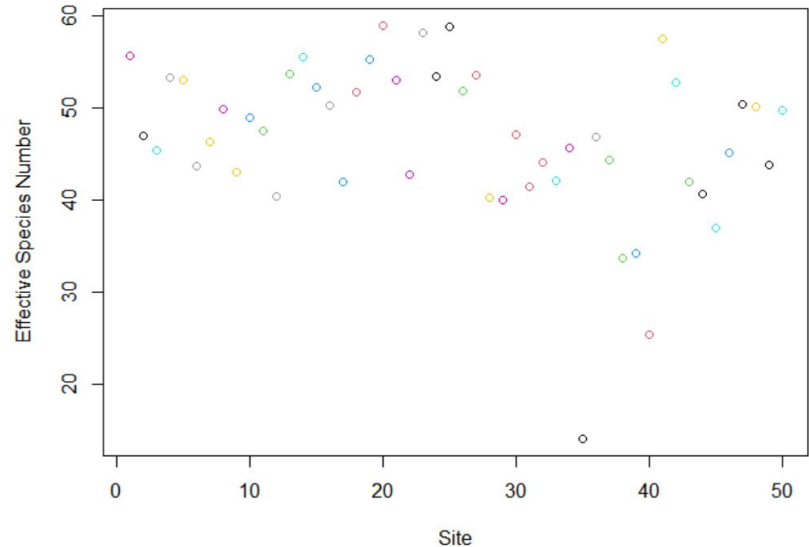
## Calculating Effective Species Number

- The exponential of the Shannon Index

```
### Calculate Effective Species number
```

```
BCI_effective_1 <- exp(BCI_shannon)
```

```
plot(BCI_effective_1, xlab="Site", ylab="Effective Species Number  
(q=1)", col=factor(BCI_effective_1))
```



# Diversity Analysis: iNEXT

Use either Individual-based abundance data (sample data: `spider`) or Sampling-unit-based incidence data (sample data: `ants`)

\*Uses lists, not dataframe

- Calculates rarefaction curves and extrapolates them to a given endpoint.

Focuses on the 3 Hill numbers:

- Species richness (when  $q = 0$ ), the exponential of the Shannon index (when  $q = 1$ ), and Simpson's index (when  $q = 2$ )

```
library(iNEXT)
```

```
library(ggplot2)
```

# Diversity Analysis: iNEXT

## Example 1: Individual-based abundance data

```
# Use the BCI data, but need to remove zeros first

#RUN iNext for species richness (q=0), exponential of
Shannon (1) or Simpson (2)

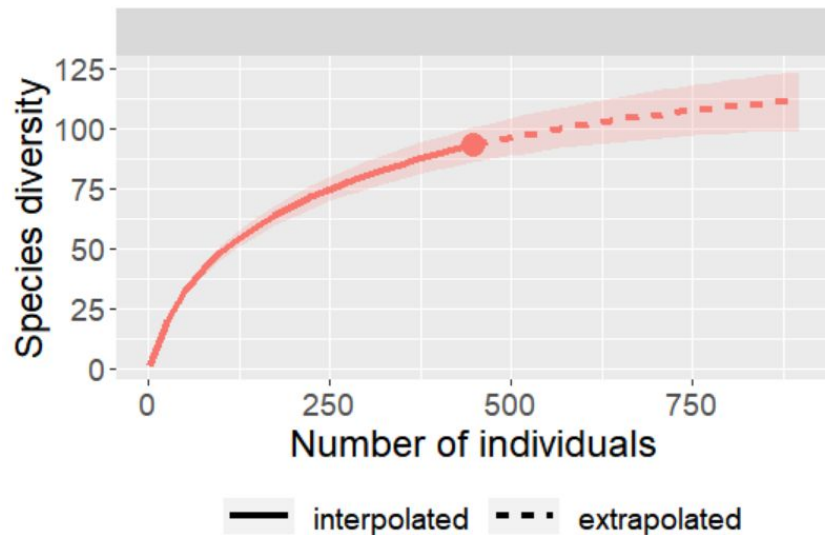
BCI_inext <-iNEXT(as.vector(BCI_out), q=0,
datatype="abundance")

ggiNEXT(BCI_inext, type=1, facet.var="site")

plot_1 <-ggiNEXT(test_ant , type=1, color.var="site") +

  theme_bw(base_size = 18) +

  theme(legend.position="none")
```



# Diversity Analysis: iNEXT

Using bird data to compare all 3 Hill numbers between two sites

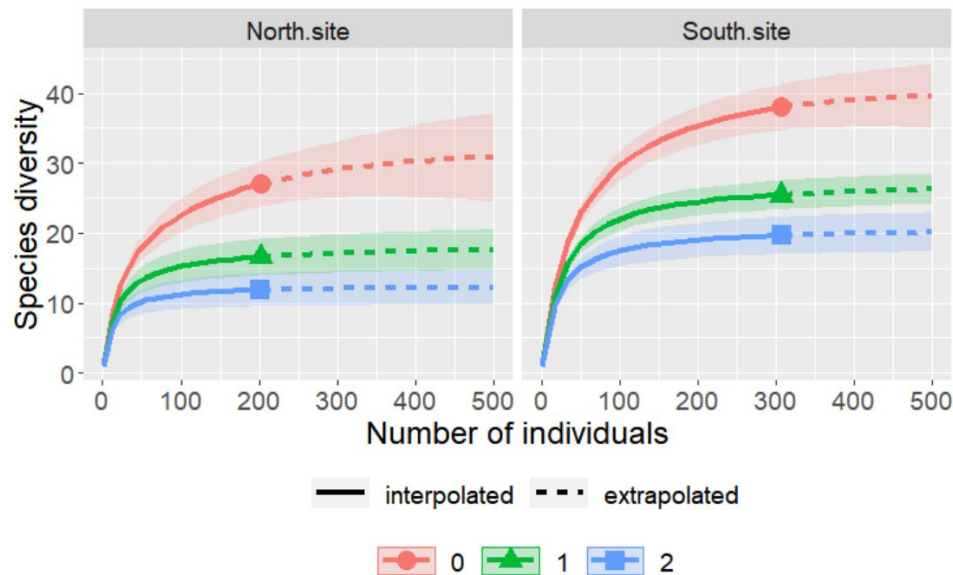
```
data(bird)
```

```
bird_out <- iNEXT(bird, q=c(0, 1, 2),
```

```
  datatype="abundance", endpoint=500)
```

```
# Sample-size-based R/E curves, separating plots by "site"
```

```
ggiNEXT(bird_out, type=1, facet.var="site")
```



# Diversity Analysis: iNEXT

## Sample size rarefaction & extrapolation using ant data

- The first entry of each list must be the total number of sampling units.

```
data(ant)
```

```
t <- seq(1, 700, by=10)
```

```
ant_out <- iNEXT(ant, q=0,
```

```
  datatype="incidence_freq", size=t)
```

```
> str(ant)
```

```
List of 5
```

```
$ h50m : num [1:228] 599 330 263 236 22
```

```
...
```

```
$ h500m : num [1:242] 230 133 131 123 78
```

```
$ h1070m: num [1:123] 150 99 96 80 74 68
```

```
$ h1500m: num [1:57] 200 144 113 79 76 7
```

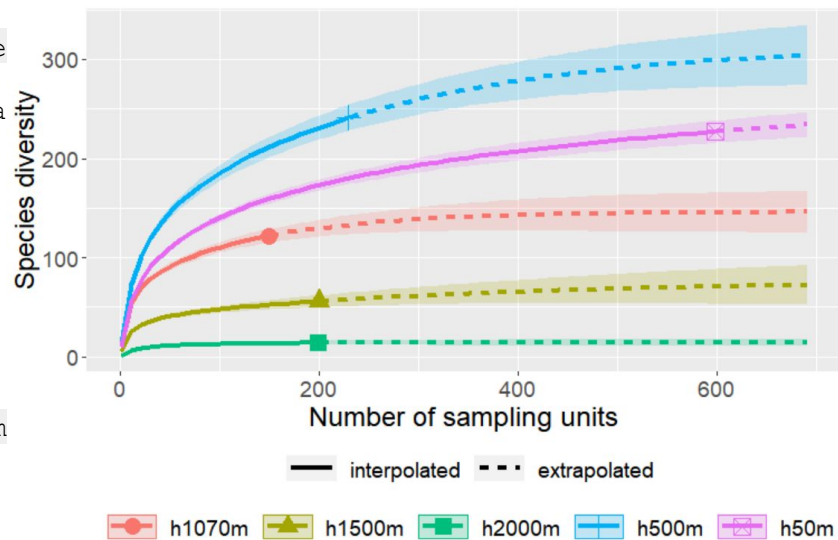
```
$ h2000m: num [1:15] 200 80 59 34 23 19
```

# Diversity Analysis: iNEXT

## Sample size rarefaction & extrapolation using ant data

- The first entry of each list must be the total number of sampling units.

```
# Row 1: Names of regions,  
# Row 2: Sampling effort, Then rows are sums of each spe  
# Sample size rarefaction & extrapolation using ant data  
data(ant)  
  
t <- seq(1, 700, by=10)  
  
ant_out <- iNEXT(ant, q=0,  
  datatype="incidence_freq", size=t)  
  
ggiNEXT(ant_out, type=1) + theme(legend.position="bottom"  
  theme(legend.position="bottom",  
    legend.title=element_blank())
```





# Community Composition

What else in vegan?

- Beta Diversity using `vegdist`
- NDMS using `metaMDS`
- Species Ranked Abundance using `radfit`
  - Examples of these in R doc

Thank you!

What has worked for your data and what  
advice do you have??



# Resource Links

Vegan

<https://cran.r-project.org/web/packages/vegan/vignettes/diversity-vegan.pdf>

iNEXT

[https://rstudio-pubs-static.s3.amazonaws.com/210845\\_860b29c5e0a643f987a80179b61b616.html](https://rstudio-pubs-static.s3.amazonaws.com/210845_860b29c5e0a643f987a80179b61b616.html)

<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12613>

<https://besjournals.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2F2041-210X.12613&file=mee312613-sup-0001-AppendixS1.pdf>

<https://github.com/JohnsonHsieh/iNEXT>