A Quick Introduction to iNEXT via Examples

T. C. Hsieh, K. H. Ma, and Anne Chao

Latest version 3.0.0 (July, 2022)

Latest Updates in July 2022: (1) We have modified (in the main function `iNEXT`) the bootstrap method used to obtain confidence intervals for the coverage-based rarefaction and extrapolation curves. We have expanded the `iNEXT` output (`$iNextEst`) to include two lists (`$size_based` and `$coverage_based`). (2) In the function `estimateD`, for a given coverage value, we have refined our algorithm to find the corresponding sample size (not necessarily restricted to integers) to obtain more accurate diversity estimates. (3) We have changed some column names in the output in order to conform to our forthcoming iNEXT series (iNEXT.3D, iNEXT.4steps, iNEXT.link). Please download the latest version of `iNEXT` available from CRAN or from Anne Chao's iNEXT_github, or use the latest version of iNEXT Online available from Shiny iNEXT-Online.

`iNEXT` (iNterpolation and EXTrapolation) is an R package modified from the original version which was supplied in the Supplement of Chao et al. (2014). In the latest updated version, we have added more user-friendly features, improved some algorithms, and refined the graphic displays. In this document, we provide a quick introduction demonstrating how to run `iNEXT`. Detailed information about `iNEXT` functions is provided in the iNEXT Manual, also available in CRAN. See Chao & Jost (2012), Colwell et al. (2012) and Chao et al. (2014) for methodologies. A short review of the theoretical background and a brief description of methods are included in an application paper by Hsieh, Ma & Chao (2016). An online version of iNEXT-online is also available for users without an R background.

`iNEXT` focuses on three measures of Hill numbers of order q: species richness (`q = 0`), Shannon diversity (`q = 1`, the exponential of Shannon entropy) and Simpson diversity (`q = 2`, the inverse of Simpson concentration). For each diversity measure, `iNEXT` uses the observed sample of abundance or incidence data (called the "reference sample") to compute diversity estimates and the associated 95% confidence intervals for the following two types of rarefaction and extrapolation (R/E):

1. Sample-size-based (or size-based) R/E sampling curves: `iNEXT` computes diversity estimates for rarefied and extrapolated samples up to an appropriate size. This type of sampling curve plots the diversity estimates with respect to sample size.
2. Coverage-based R/E sampling curves: `iNEXT` computes diversity estimates for rarefied and extrapolated samples based on a standardized level of sample completeness (as measured by sample coverage) up to an appropriate coverage value. This type of sampling curve plots the diversity estimates with respect to sample coverage.

`iNEXT` also plots the above two types of sampling curves and a sample completeness curve (which depicts how sample coverage varies with sample size). The sample completeness curve provides a bridge between the size- and coverage-based R/E sampling curves.

HOW TO CITE iNEXT

If you publish your work based on the results from the `iNEXT` package, you should make references to the following methodology paper (Chao et al. 2014) and the application paper (Hsieh, Ma & Chao, 2016):

- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs, 84, 45–67.

- Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: An R package for interpolation and extrapolation of species diversity (Hill numbers). Methods in Ecology and Evolution, 7, 1451-1456.

SOFTWARE NEEDED TO RUN iNEXT IN R

- Required: R
- Suggested: RStudio IDE

HOW TO RUN iNEXT:

The `iNEXT` package is available from CRAN and can be downloaded with a standard R installation procedure or can be downloaded from Anne Chao's iNEXT_github using the following commands. For a first-time installation, an additional visualization extension package (`ggplot2`) must be installed and loaded.

```
## install iNEXT package from CRAN
install.packages("iNEXT")
## install the latest version from github
install.packages('devtools')
library(devtools)
install_github('AnneChao/iNEXT')
## import packages
library(iNEXT)
library(ggplot2)
```

MAIN FUNCTION: iNEXT()

We first describe the main function `iNEXT()` with default arguments:

```
iNEXT(x, q=0, datatype="abundance", size=NULL, endpoint=NULL, knots=40, se=TRUE, conf=0.95,
         nboot=50)
```

The arguments of this function are briefly described below, and will be explained in more details by illustrative examples in later text. This main function computes diversity estimates of order q, the sample coverage estimates and related statistics for K (if `knots = K`) evenly-spaced knots (sample sizes) between size 1 and the `endpoint`, where the endpoint is described below. Each knot represents a particular sample size for which diversity estimates will be calculated. By default, endpoint = double the reference sample size (total sample size for abundance data; total sampling units for incidence data). For an example, if `endpoint = 10`, `knots = 4`, diversity estimates will be computed for a sequence of samples with sizes (1, 4, 7, 10). In a later real-data example, we have `endpoint = 336`, `knots = 40`; diversity estimates will be computed for a sequence of samples with sizes (1, 10,19, 28, …, 318, 327, 336).

Argument	Description
`x`	a `matrix`, `data.frame`, `lists` of species abundances, or `lists` of incidence frequencies (see data format/information below).
`q`	a number or vector specifying the diversity order(s) of Hill numbers.
`datatype`	type of input data, "`abundance`", "`incidence_raw`" or "`incidence_freq`".

Argument	Description
`size`	an integer vector of sample sizes for which diversity estimates will be computed. If `NULL`, then diversity estimates will be calculated for those sample sizes determined by the specified/default endpoint and knots.
`endpoint`	an integer specifying the sample size that is the endpoint for R/E calculation; If `NULL`, then endpoint=double the reference sample size.
`knots`	an integer specifying the number of equally-spaced knots between size 1 and the endpoint; default is 40.
`se`	a logical variable to calculate the bootstrap standard error and `conf` confidence interval.
`conf`	a positive number < 1 specifying the level of confidence interval; default is 0.95.
`nboot`	an integer specifying the number of bootstrap replications; default is 50.

This function returns an `"iNEXT"` object which can be further used to make plots using the function `ggiNEXT()` to be described below.

DATA FORMAT/INFORMATION

Three types of data are supported:

1. Individual-based abundance data (`datatype="abundance"`): Input data for each assemblage/site include species abundances in an empirical sample of n individuals (“reference sample”). When there are N assemblages, input data consist of an S by N abundance matrix, or N lists of species abundances.
2. Sampling-unit-based incidence data: There are two kinds of input data.

a. Incidence-raw data (`datatype="incidence_raw"`): for each assemblage, input data for a reference sample consisting of a species-by-sampling-unit matrix; each element in the raw matrix is 1 for a detection, and 0 otherwise. When there are N assemblages, input data consist of N lists of raw matrices, and each matrix is a species-by-sampling-unit matrix.
b. Incidence-frequency data (`datatype="incidence_freq"`): input data for each assemblage consist of species sample incidence frequencies (i.e., row sums of the corresponding incidence raw matrix). When there are N assemblages, input data consist of an (S+1) by N matrix, or N lists of species incidence frequencies. The first entry of each column/list must be the total number of sampling units, followed by the species incidence frequencies.

RAREFACTION/EXTRAPOLATION VIA EXAMPLES

Four data sets are included in the `iNEXT` package for illustration. There are two abundance data sets: `spider` (list of two vectors) and `bird` (in data.frame format), and two incidence data sets: `ant` (list of 5 vectors) and `ciliates` (list of 3 matrices). The input datatypes are the same for the two abundance data sets (`datatype="abundance"`), but the input datatypes are different for the `ant` data (`datatype="incidence_freq"`) and the `ciliates` data (`datatype="incidence_raw"`). We first use the `spider` data for illustration; see Chao et al. (2014) for analysis details and data interpretations. The spider data consist of abundance data from two canopy manipulation treatments (“Girdled” and “Logged”) of hemlock trees (Ellison et al. 2010). For these data, the following commands run the `iNEXT()` function for `q = 0`.

```
data(spider)
str(spider)
```

```
iNEXT(spider, q=0, datatype="abundance")
```

The `iNEXT()` function returns the `"iNEXT"` object including three output lists: `$DataInfo` for summarizing data information; `$iNextEst` for showing size- and coverage-based diversity estimates along with related statistics for a series of rarefied and extrapolated samples; and `$AsyEst` for showing asymptotic diversity estimates along with related statistics. `$DataInfo`, as shown below, returns basic data information including the reference sample size (`n`), observed species richness (`S.obs`), sample coverage estimate for the reference sample (`SC`), and the first ten frequency counts (`f1-f10`). This part of output can also be computed by the function `DataInfo()`

```
$DataInfo: basic data information
  Assemblage   n S.obs     SC f1 f2 f3 f4 f5 f6 f7 f8 f9 f10
1    Girdled 168    26 0.9289 12  4  0  1  0  2  0  1  1   0
2     Logged 252    37 0.9446 14  4  4  3  1  0  3  2  0   1
```

For incidence data, the list `$DataInfo` includes the reference sample size (`T`), observed species richness (`S.obs`), total number of incidences (`U`), sample coverage estimate for the reference sample (`SC`), and the first ten incidence frequency counts (`Q1-Q10`).

In the Girdled treatment assemblage, by default, 40 equally spaced knots (samples sizes) between 1 and 336 (= 2 x 168, double the reference sample size, Chao et al. 2014) are selected. Diversity estimates and related statistics are computed for these 40 knots (corresponding to sample sizes m = 1, 10, 19, …, 327, 336), which locates the reference sample at the mid-point of the selected knots. If the argument `se=TRUE`, then the bootstrap method is applied to obtain the 95% confidence intervals for each diversity and sample coverage estimate.

The list `$iNextEst` output includes two data frames: `$size_based` and `$coverage_based`. (Note the output in the list `$iNextEst` is different from that obtained from earlier `iNEXT` versions < 3.0.0, due to a modification in the bootstrap method.) For the sample size corresponding to each knot, the first data frame (as shown under `$size_based`) includes the name of `Assemblage`, the sample size (`m`, i.e., each of the 40 knots), the method (`Rarefaction`, `Observed`, or `Extrapolation`, depending on whether the size `m` is less than, equal to, or greater than the reference sample size), the diversity order (`order.q`), the diversity estimate of order q (`qD`), the 95% lower and upper confidence limits of diversity (`qD.LCL`, `qD.UCL`), and the sample coverage estimate (`SC`) along with the 95% lower and upper confidence limits of sample coverage (`SC.LCL`, `SC.UCL`). These sample coverage estimates with confidence intervals are used for plotting the sample completeness curve.

```
$iNextEst: diversity estimates with rarefied and extrapolated samples.
$size_based (LCL and UCL are obtained for fixed size.)

   Assemblage   m        Method Order.q     qD qD.LCL qD.UCL    SC SC.LCL SC.UCL
1  Girdled      1   Rarefaction       0  1.000  1.000  1.000 0.122  0.089  0.156
10 Girdled     84   Rarefaction       0 18.912 15.902 21.923 0.900  0.872  0.927
20 Girdled    168      Observed       0 26.000 21.492 30.508 0.929  0.904  0.954
30 Girdled    248 Extrapolation       0 30.883 25.149 36.618 0.948  0.918  0.979
40 Girdled    336 Extrapolation       0 34.731 27.187 42.275 0.964  0.931  0.996
41  Logged      1   Rarefaction       0  1.000  1.000  1.000 0.145  0.109  0.180
50  Logged    126   Rarefaction       0 28.268 24.935 31.600 0.908  0.886  0.930
60  Logged    252      Observed       0 37.000 31.789 42.211 0.945  0.925  0.964
70  Logged    371 Extrapolation       0 42.786 35.844 49.727 0.958  0.935  0.980
80  Logged    504 Extrapolation       0 47.644 38.485 56.804 0.969  0.946  0.991

NOTE: The above output only shows five estimates for each assemblage; call
iNEXT.object$iNextEst$size_based to view complete output.
```

The second data frame (as shown under `$coverage_based`) includes the name of `Assemblage`, the standardized sample coverage (`SC`), the corresponding sample size for the standardized coverage (`m`, i.e., each of the 40 knots), the method (`Rarefaction`, `Observed`, or `Extrapolation`, depending on whether the sample coverage `SC` is less than, equal to, or greater than the reference sample coverage), the diversity order (`order.q`), the diversity estimate of order q (`qD`), and the 95% lower and upper confidence limits of diversity (`qD.LCL`, `qD.UCL`). These diversity estimates and confidence intervals are used for plotting the coverage-based R/E curves.

```
$coverage_based (LCL and UCL are obtained for fixed coverage; interval length is wider due to
varying size in bootstraps.)

   Assemblage    SC   m        Method order.q     qD qD.LCL qD.UCL
1  Girdled    0.122   1   Rarefaction       0  1.000  0.857  1.143
10 Girdled    0.900  84   Rarefaction       0 18.912 10.761 27.064
20 Girdled    0.929 168      Observed       0 26.000 13.239 38.761
30 Girdled    0.948 248 Extrapolation       0 30.883 12.129 49.638
40 Girdled    0.964 336 Extrapolation       0 34.731  9.788 59.673
41  Logged    0.145   1   Rarefaction       0  1.000  0.796  1.204
50  Logged    0.908 126   Rarefaction       0 28.268 20.192 36.343
60  Logged    0.945 252      Observed       0 37.000 20.209 53.791
70  Logged    0.958 371 Extrapolation       0 42.786 21.977 63.594
80  Logged    0.969 504 Extrapolation       0 47.644 23.357 71.932

NOTE: The above output only shows five estimates for each assemblage; call
iNEXT.object$iNextEst$coverage_based to view complete output.
```

In the above output (`$size_based` and `$coverage_based`), the confidence intervals of any standardized diversity are obtained by a bootstrap method. In the size-based standardization, the sample size is fixed in each regenerated bootstrap sample. In the coverage-based standardization, for a given standardized coverage value, the corresponding size needed to attain the same level of coverage may vary with regenerated bootstrap samples. Thus, the sampling uncertainty is greater in the coverage-based standardization and the resulting confidence interval is wider than that in the corresponding size-based standardization. For example, if the size for a future survey will be fixed at a sample size of 84, we can obtain a 95% CI of (15.9, 21.9) for the expected diversity (q = 0) based on the first data frame (`$size_based` output). However, if the coverage of a survey is fixed at the level of 0.9, the size needed for the current data is 84, but the size needed for a regenerated bootstrap sample may be different from 84; the second data frame (`$coverage_based` output) shows a CI of (10.8, 27.1), which is wider than the former one based on a size of 84. Because we use a random bootstrapping/regeneration process, with 50 replications (default), to obtain each CI, the output for `qD.LCL` and `qD.UCL` may vary slightly each time you enter the same data.

`$AsyEst` lists the name of `Assemblage`, the `Diversity` (species richness for q = 0, Shannon diversity for q = 1, and Simpson diversity for q = 2), the observed diversity (`Observed`), the asymptotic diversity estimate (`Estimator`), the s.e. of the asymptotic estimator (`s.e.`) and the associated 95% lower and upper confidence limits (`LCL`, `UCL`). The estimated asymptotes are calculated via the functions `ChaoRichness()` for q = 0, `ChaoShannon()` for q = 1 and `ChaoSimpson()` for q = 2; see Chao et al. (2014) for the formulas of all asymptotic estimators. The output for the spider data is shown below.

```
$AsyEst: asymptotic diversity estimates along with related statistics.
  Assemblage         Diversity  Observed Estimator    s.e.     LCL     UCL
1    Girdled  Species richness    26.000    43.893  17.219  26.000  77.642
2    Girdled Shannon diversity    12.060    13.826   1.339  11.201  16.451
3    Girdled Simpson diversity     7.840     8.175   0.934   6.344  10.006
4     Logged  Species richness    37.000    61.403  19.692  37.000  99.998
5     Logged Shannon diversity    14.421    16.337   1.864  12.684  19.990
6     Logged Simpson diversity     6.761     6.920   0.926   5.106   8.734
```

The user may specify an integer sample size for the argument `endpoint` to designate the maximum sample size of R/E calculation. For species richness, the extrapolation method is reliable up to double the reference sample size; beyond that, the prediction bias may be large. However, for measures of q = 1 and 2, the extrapolation can usually be safely extended to the asymptote if data are not sparse; thus there is no limit for the value of the `endpoint` for these two measures.

The user may also specify the number of `knots` in the sample size range between 1 and the endpoint. If you choose a large number of knots, then it may take a long time to obtain the output due to the time-consuming nature of the bootstrap method. Alternatively, the user may specify a series of sample sizes for R/E computation, as in the following example:

```
# set a series of sample sizes (m) for R/E computation
m <- c(1, 5, 20, 50, 100, 200, 400)
iNEXT(spider, q=0, datatype="abundance", size=m)
```

Further, `iNEXT` can simultaneously run R/E computation for Hill numbers with q = 0, 1, and 2 by specifying a vector for the argument `q` as follows:

```
out <- iNEXT(spider, q=c(0,1,2), datatype="abundance", size=m)
```

A data.frame input format for abundance-based analysis is also supported:

```
data(bird)
str(bird) # 41 obs. of 2 variables
iNEXT(bird, q=0, datatype="abundance")
```

GRAPHIC DISPLAYS: FUNCTION ggiNEXT()

The function `ggiNEXT()`, which extends `ggplot2` to the `"iNEXT"` object with default arguments, is described as follows:

```
ggiNEXT(x, type=1, se=TRUE, facet.var="None", color.var="Assemblage", grey=FALSE)
```

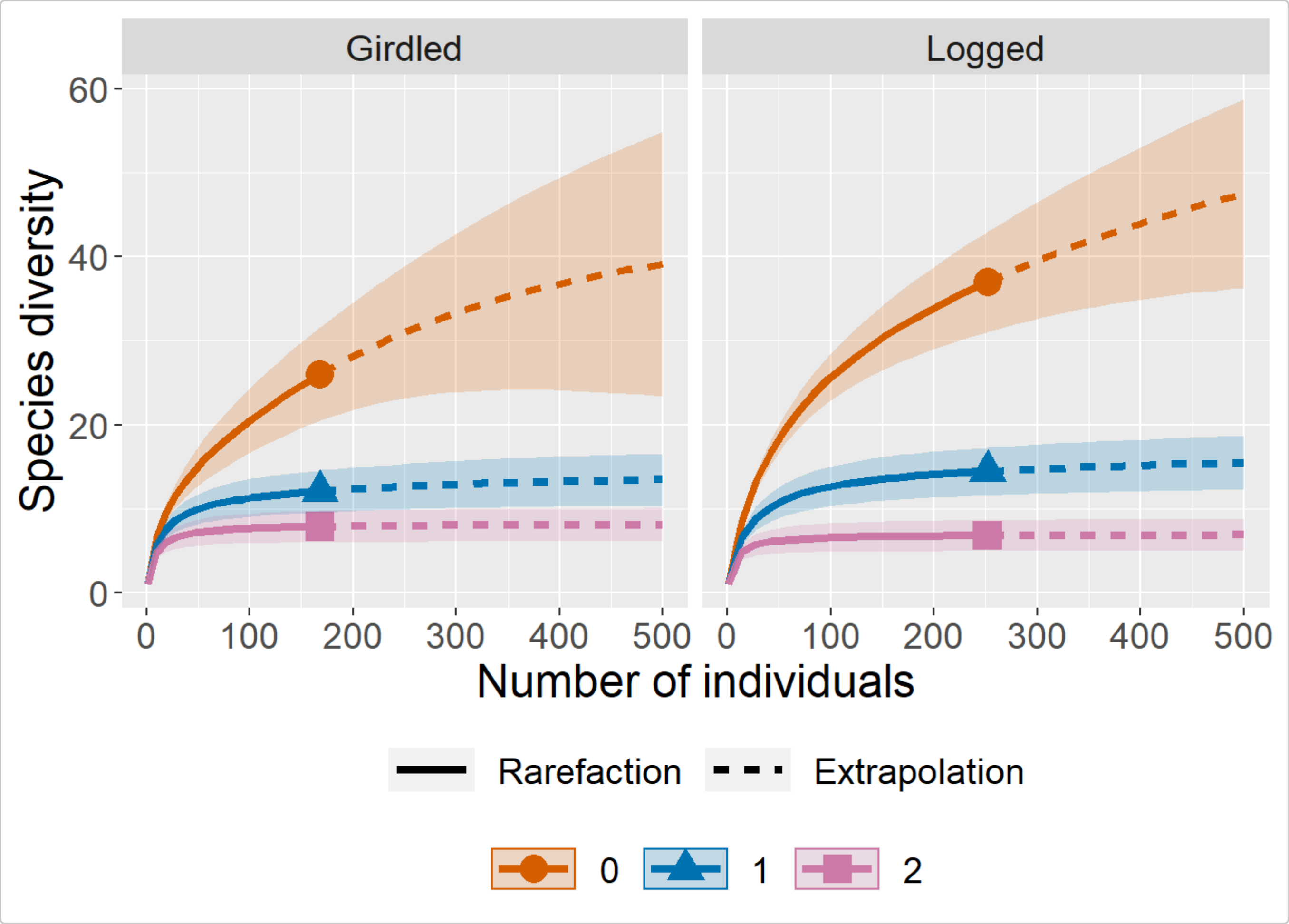
Here `x` is an `"iNEXT"` object. Three types of curves are allowed:

1. Sample-size-based R/E curve (`type=1`): see Figs. 1a and 2a in Hsieh et al. (2016). This curve plots diversity estimates with confidence intervals (if `se=TRUE`) as a function of sample size up to double the reference sample size, by default, or a user-specified `endpoint`.
2. Sample completeness curve (`type=2`) with confidence intervals (if `se=TRUE`): see Figs. 1b and 2b in Hsieh et al. (2016). This curve plots the sample coverage with respect to sample size for the same range described in (1).
3. Coverage-based R/E curve (`type=3`): see Figs. 1c and 2c in Hsieh et al. (2016). This curve plots the diversity estimates with confidence intervals (if `se=TRUE`) as a function of sample coverage up to the maximum coverage obtained from the maximum size described in (1).

The `ggiNEXT()` function is a wrapper around the `ggplot2` package to create a R/E curve using a single line of code. The resulting object is of class `"ggplot"`, so it can be manipulated using the `ggplot2` tools. The argument `facet.var=("None", "Order.q", "Assemblage" or "Both")` can be used to create a separate plot for each value of the specified variable. See the following examples.

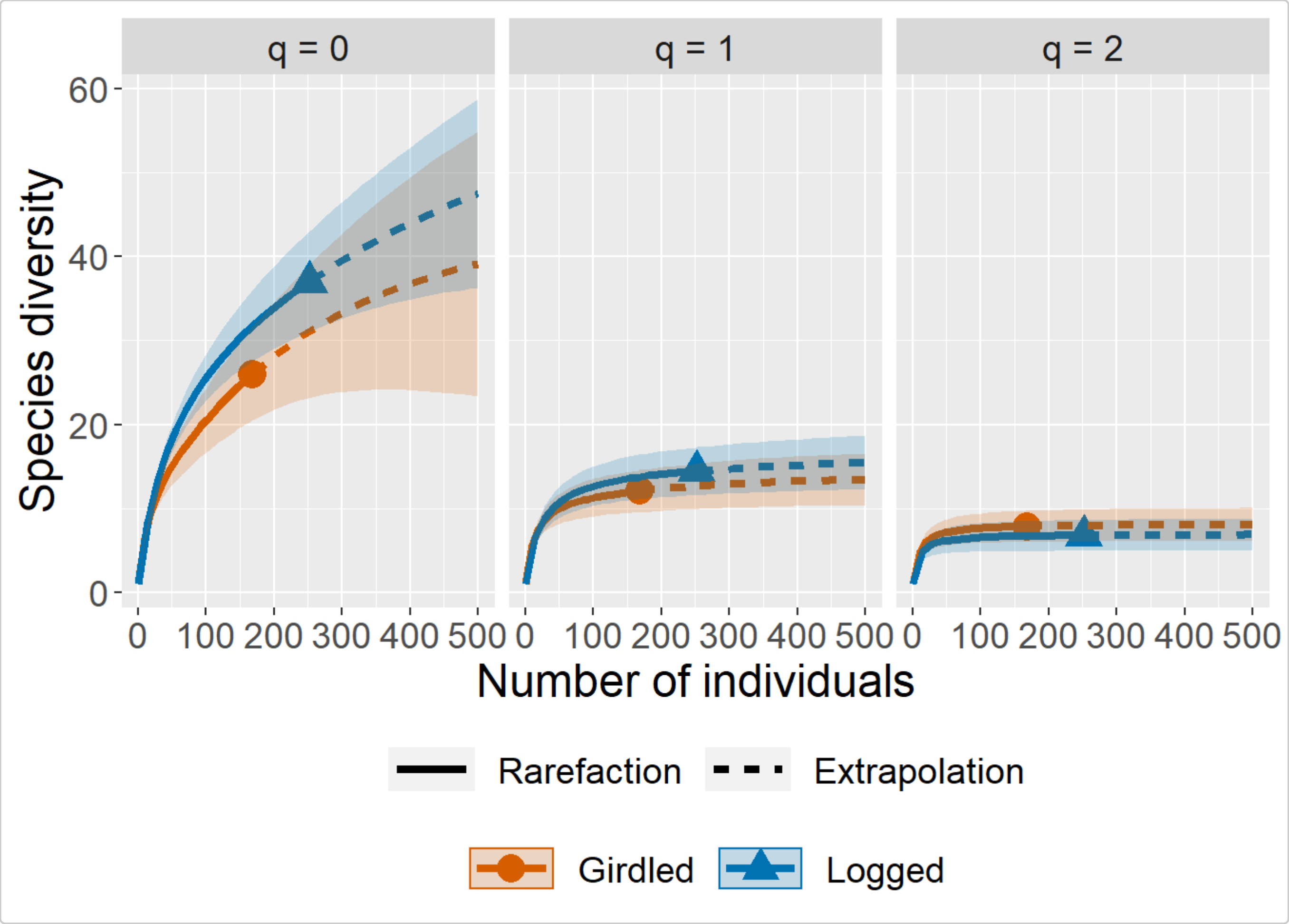
The argument `facet.var="Assemblage"` in the `ggiNEXT` function creates a separate plot for each assemblage as shown below:

```
# Sample-size-based R/E curves, separating by "Assemblage""
out <- iNEXT(spider, q=c(0, 1, 2), datatype="abundance", endpoint=500)
ggiNEXT(out, type=1, facet.var="Assemblage")
```



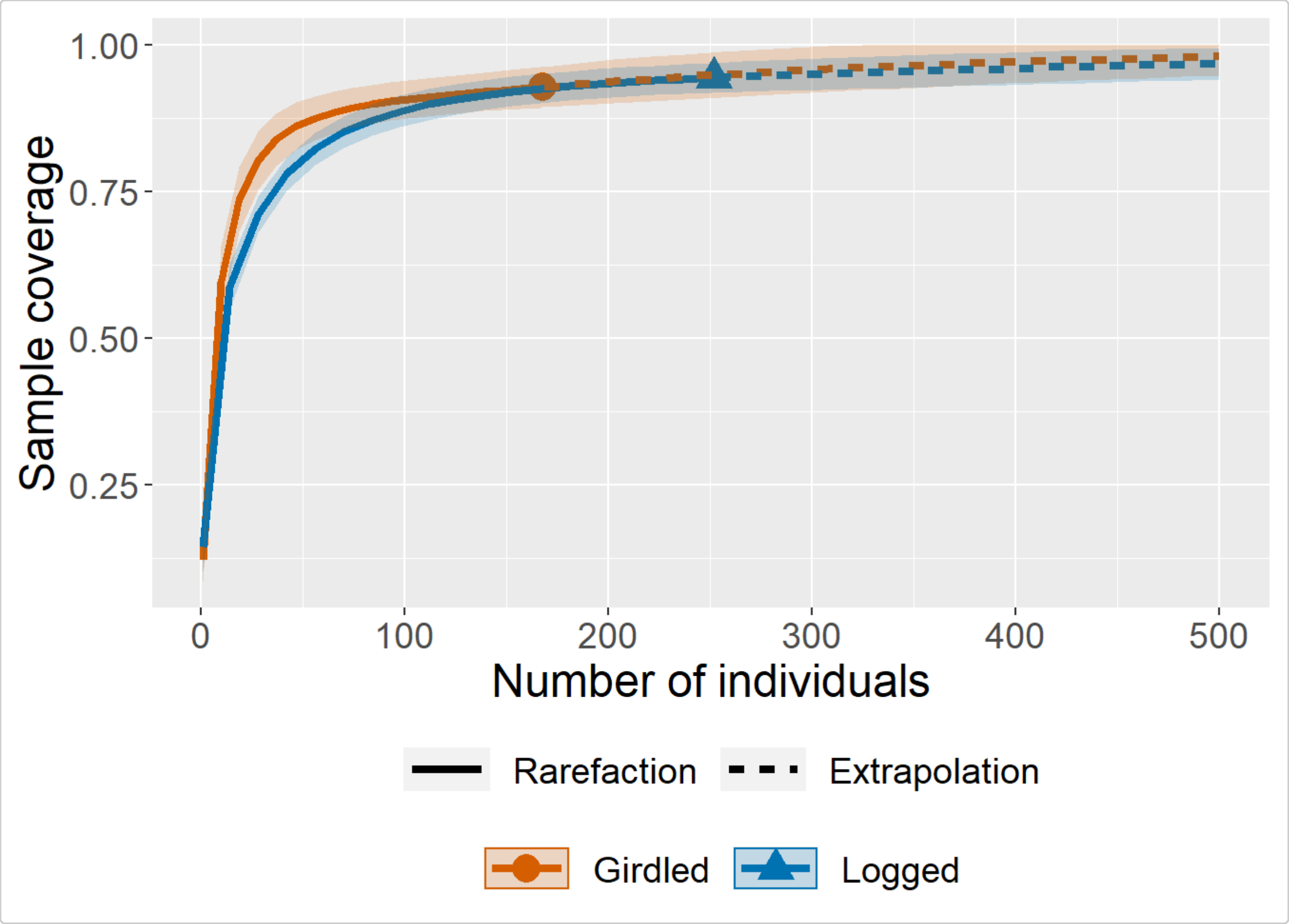
The argument `facet.var="Order.q"` and `color.var="Assemblage"` creates a separate plot for each diversity order assemblage, and within each plot, different colors are used for the two assemblages.

```
ggiNEXT(out, type=1, facet.var="Order.q", color.var="Assemblage")
```



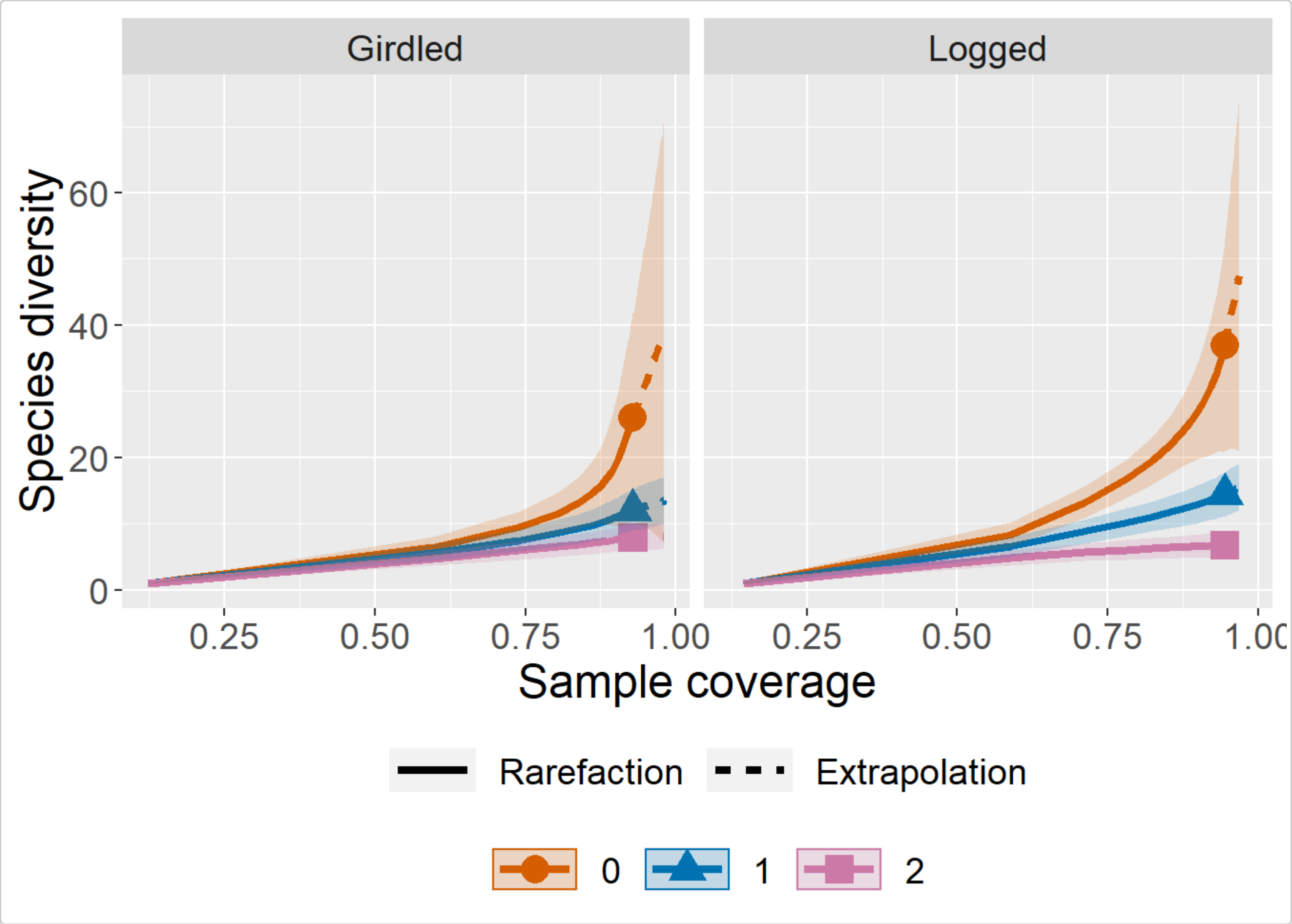
The following commands return the sample completeness curve in which different colors are used for the two assemblages:

```
ggiNEXT(out, type=2, facet.var="None", color.var="Assemblage")
```

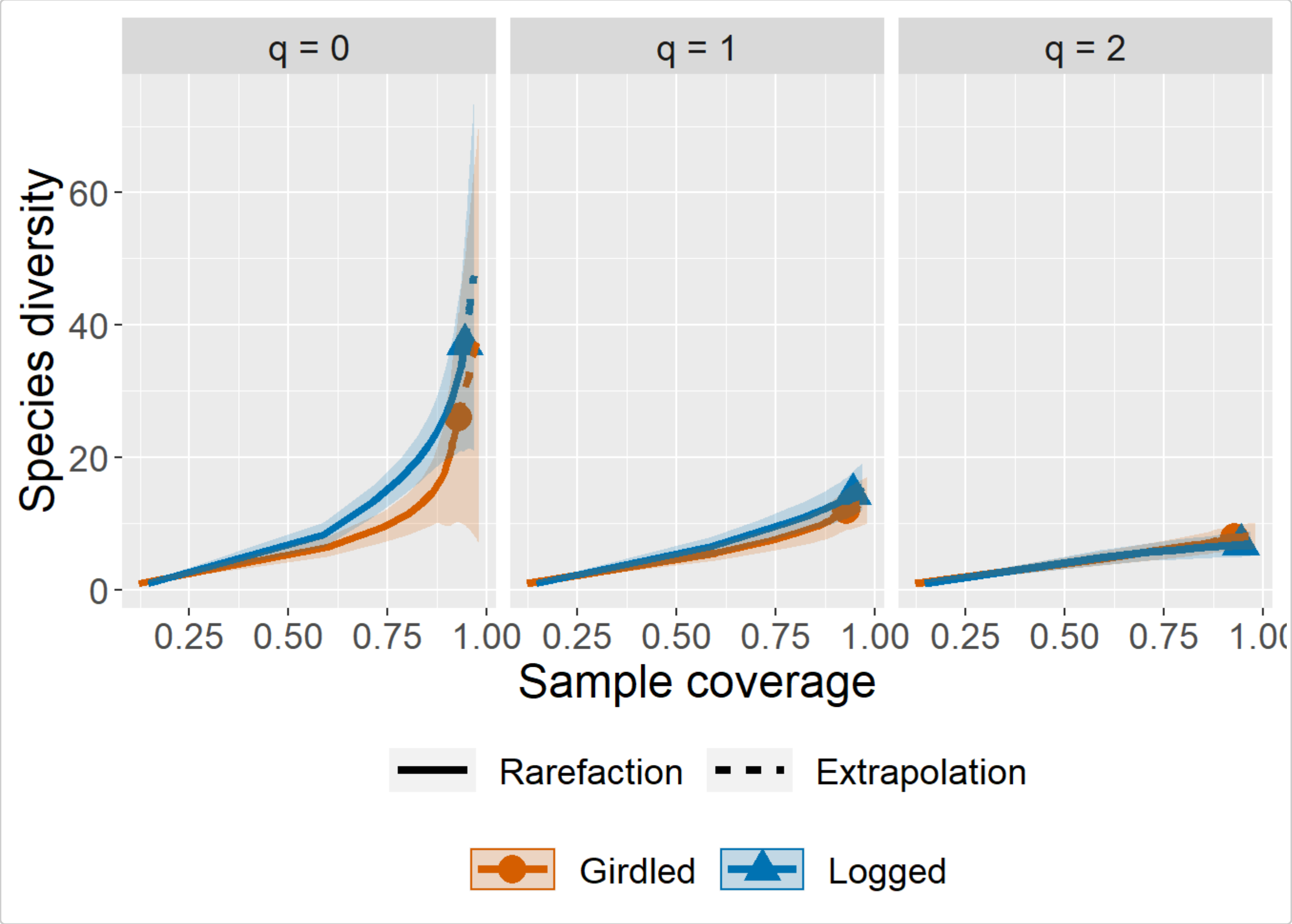


The following commands return the coverage-based R/E sampling curves in which different colors are used for the two assemblages (`facet.var="Assemblage"`) and for three orders (`facet.var="Order.q"`)

```
ggiNEXT(out, type=3, facet.var="Assemblage")
```



```
ggiNEXT(out, type=3, facet.var="Order.q", color.var="Assemblage")
```



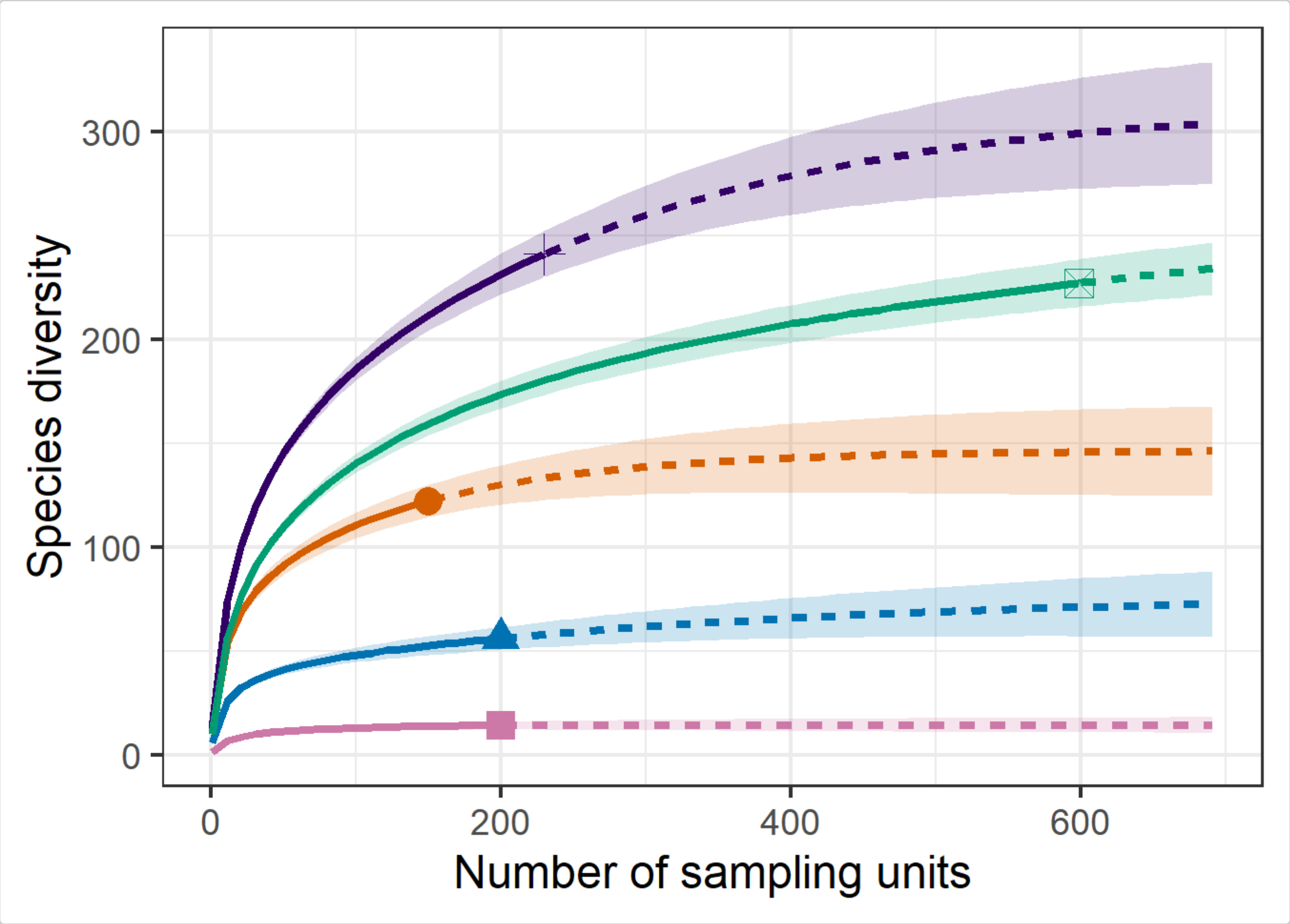
INCIDENCE DATA with datatype="incidence_freq"

For illustration, we use the tropical ant data (in the dataset `ant` included in the package) at five elevations (50m, 500m, 1070m, 1500m, and 2000m) collected by Longino & Colwell (2011) from Costa Rica. The 5 lists of incidence frequencies are shown below. The first entry of each list must be the total number of sampling units, followed by the species incidence frequencies.

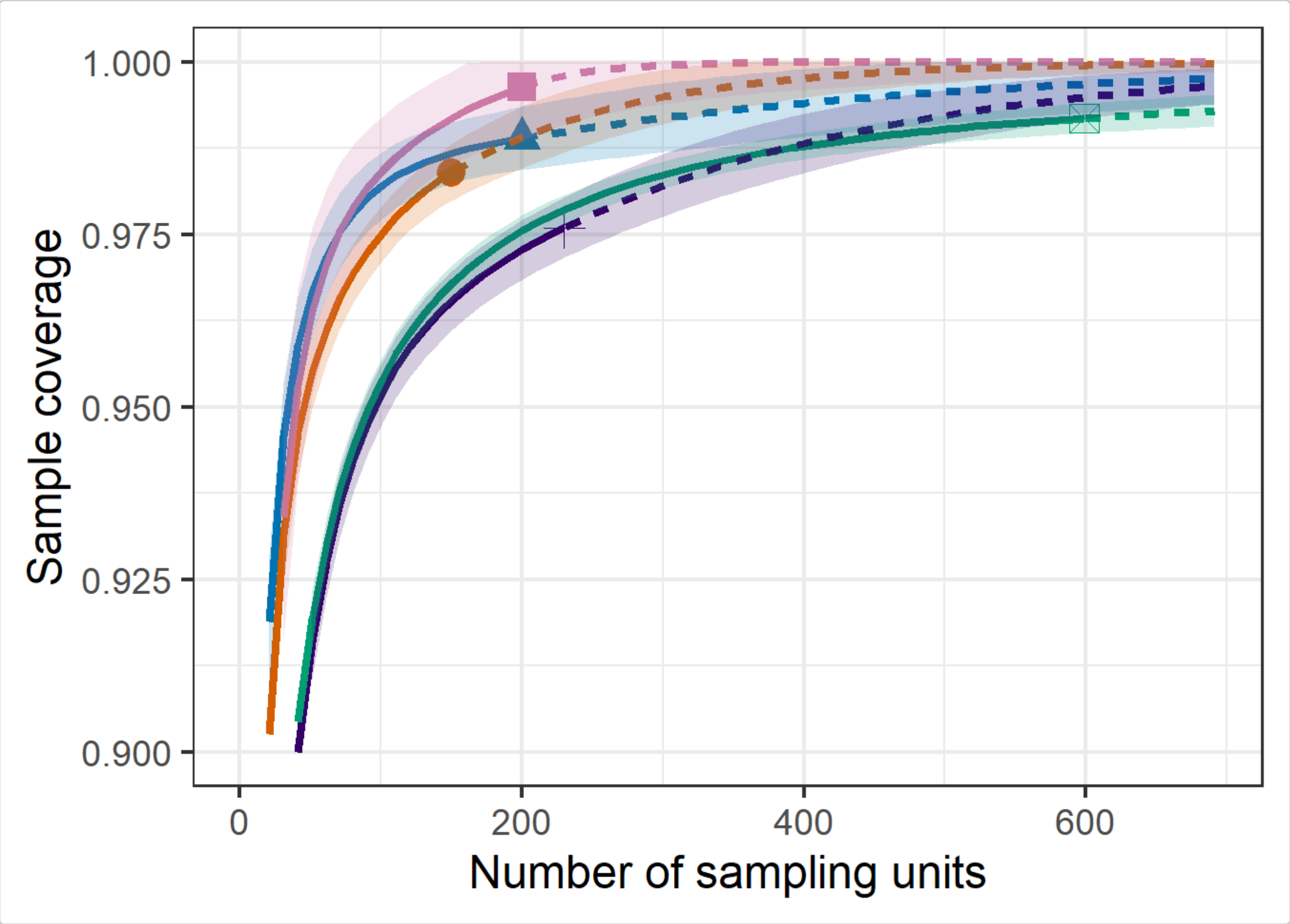
```
data(ant)
str(ant)
List of 5
 $ h50m  : num [1:228] 599 330 263 236 222 195 186 183 182 129 ...
 $ h500m : num [1:242] 230 133 131 123 78 73 65 60 60 56 ...
 $ h1070m: num [1:123] 150 99 96 80 74 68 60 54 46 45 ...
 $ h1500m: num [1:57] 200 144 113 79 76 74 73 53 50 43 ...
 $ h2000m: num [1:15] 200 80 59 34 23 19 15 13 8 8 ...
```

The argument `color.var = ("None", "Order.q", "Assemblage" or "Both")` is used to display curves in different colors for values of the specified variable. For example, the following code using the argument `color.var="Assemblage"` displays the sampling curves in different colors for the five assemblages. Note that `theme_bw()` is a ggplot2 function to modify the display setting from a grey to a white background with black gridlines. The following commands return three types of R/E sampling curves for the `ant` data.

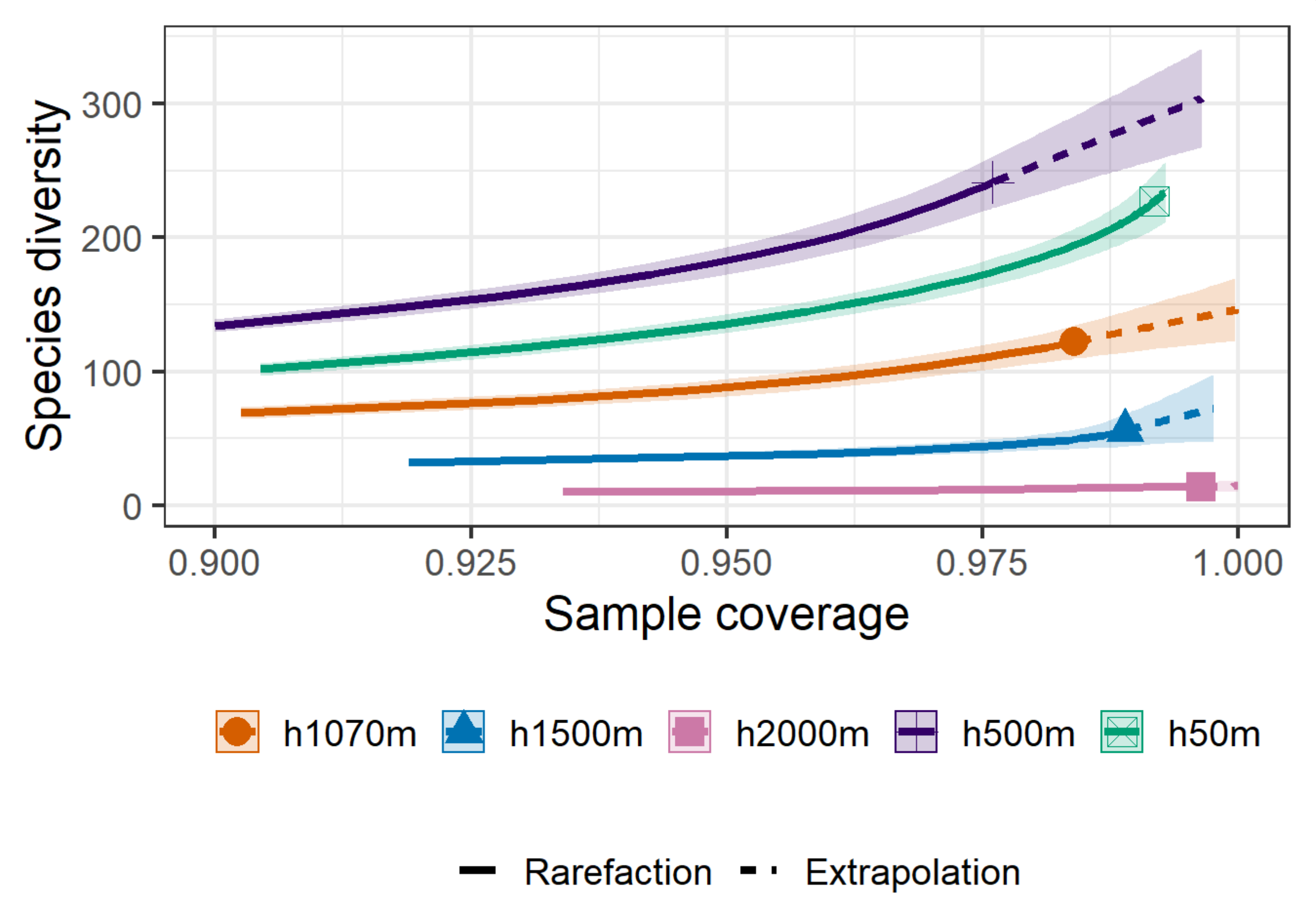
```
t <- seq(1, 700, by=10)
out.inc <- iNEXT(ant, q=0, datatype="incidence_freq", size=t)

# Sample-size-based R/E curves
ggiNEXT(out.inc, type=1, color.var="Assemblage") +
  theme_bw(base_size = 18) +
  theme(legend.position="None")
```



```
# Sample completeness curves
ggiNEXT(out.inc, type=2, color.var="Assemblage") +
  ylim(c(0.9,1)) +
  theme_bw(base_size = 18) +
  theme(legend.position="None")
```



```
# Coverage-based R/E curves
ggiNEXT(out.inc, type=3, color.var ="Assemblage") +
  xlim(c(0.9,1)) +
  theme_bw(base_size = 18) +
  theme(legend.position="bottom",
        legend.title=element_blank(),
        text=element_text(size=18),
        legend.box = "vertical")
```



INCIDENCE DATA with datatype="incidence_raw"

We use the `ciliates` data collected from three coastal dune habitats to demostrate the use of the input `datatype="incidence_raw"`. The data set (`ciliates`) included in the package is a list of three species-by-plots matrices. Run the following commands to get the output as shown below.

```
data(ciliates)
str(ciliates)
List of 3
 $ EtoshaPan          : int [1:365, 1:19] 0 0 0 0 0 0 0 0 0 0 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:365] "Acaryophrya.collaris" "Actinobolina.multinucleata.n..sp."
         "Afroamphisiella.multinucleata.n..sp." "Afrothrix.multinucleata.n..sp." ...
  .. ..$ : chr [1:19] "x53" "x54" "x55" "x56" ...
 $ CentralNamibDesert : int [1:365, 1:17] 0 0 0 0 0 1 0 0 0 0 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:365] "Acaryophrya.collaris" "Actinobolina.multinucleata.n..sp."
         "Afroamphisiella.multinucleata.n..sp." "Afrothrix.multinucleata.n..sp." ...
  .. ..$ : chr [1:17] "x31" "x32" "x34" "x35" ...
 $ SouthernNamibDesert: int [1:365, 1:15] 0 0 0 0 0 0 0 0 0 0 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:365] "Acaryophrya.collaris" "Actinobolina.multinucleata.n..sp."
         "Afroamphisiella.multinucleata.n..sp." "Afrothrix.multinucleata.n..sp." ...
  .. ..$ : chr [1:15] "x9" "x17" "x19" "x20" ...
```

```
out.raw <- iNEXT(ciliates, q = 0, datatype="incidence_raw", endpoint=150)
out.raw
```

```
# ggiNEXT(out.raw)
```

```
Compare 3 assemblages with Hill number order q = 0.
$class: iNEXT

$DataInfo: basic data information
           Assemblage  T   U S.obs     SC  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10
1           EtoshaPan 19 516   216 0.8017 107 44 26 14  6  5  4  3  2   2
2  CentralNamibDesert 17 379   130 0.8425  63 28 13  4  3  7  1  2  1   0
3 SouthernNamibDesert 15 358   150 0.7816  82 28 14  8  6  1  1  2  2   1

$iNextEst: diversity estimates with rarefied and extrapolated samples.
$size_based (LCL and UCL are obtained for fixed size.)

             Assemblage   t        Method Order.q      qD  qD.LCL  qD.UCL    SC SC.LCL SC.UCL
1             EtoshaPan   1   Rarefaction       0  27.158  24.893  29.423 0.190  0.164  0.216
10            EtoshaPan  10   Rarefaction       0 153.260 141.984 164.536 0.680  0.644  0.716
20            EtoshaPan  20 Extrapolation       0 221.386 203.508 239.263 0.810  0.778  0.843
30            EtoshaPan  88 Extrapolation       0 333.606 291.244 375.967 0.991  0.985  0.997
39            EtoshaPan 150 Extrapolation       0 338.901 293.346 384.456 0.999  0.999  1.000
40   CentralNamibDesert   1   Rarefaction       0  22.294  20.579  24.009 0.374  0.340  0.408
49   CentralNamibDesert  10   Rarefaction       0  98.993  90.442 107.545 0.764  0.736  0.791
58   CentralNamibDesert  24 Extrapolation       0 151.018 135.480 166.557 0.892  0.856  0.929
67   CentralNamibDesert  87 Extrapolation       0 195.191 156.456 233.925 0.996  0.990  1.000
76   CentralNamibDesert 150 Extrapolation       0 196.656 154.935 238.377 1.000  0.999  1.000
77  SouthernNamibDesert   1   Rarefaction       0  23.867  21.566  26.167 0.282  0.245  0.319
85  SouthernNamibDesert   9   Rarefaction       0 112.485 103.115 121.854 0.699  0.657  0.741
94  SouthernNamibDesert  30 Extrapolation       0 207.213 185.149 229.276 0.893  0.846  0.940
103 SouthernNamibDesert  93 Extrapolation       0 259.337 209.899 308.776 0.995  0.988  1.000
111 SouthernNamibDesert 150 Extrapolation       0 261.886 209.174 314.598 1.000  0.999  1.000

NOTE: The above output only shows five estimates for each assemblage; call
iNEXT.object$iNextEst$size_based to view complete output.

$coverage_based (LCL and UCL are obtained for fixed coverage; interval length is wider due to
varying size in bootstraps.)

             Assemblage        SC   t        Method Order.q      qD  qD.LCL  qD.UCL
1             EtoshaPan 0.1901402   1   Rarefaction       0  27.158  24.894  29.422
10            EtoshaPan 0.6799226  10   Rarefaction       0 153.260 135.451 171.069
20            EtoshaPan 0.8103610  20 Extrapolation       0 221.386 194.992 247.779
30            EtoshaPan 0.9909111  88 Extrapolation       0 333.606 288.842 378.369
39            EtoshaPan 0.9994305 150 Extrapolation       0 338.901 293.087 384.715
40   CentralNamibDesert 0.3743404   1   Rarefaction       0  22.294  20.572  24.016
49   CentralNamibDesert 0.7635124  10   Rarefaction       0  98.993  86.617 111.369
58   CentralNamibDesert 0.8921419  24 Extrapolation       0 151.018 125.356 176.681
67   CentralNamibDesert 0.9964228  87 Extrapolation       0 195.191 153.805 236.576
76   CentralNamibDesert 0.9998814 150 Extrapolation       0 196.656 154.657 238.655
77  SouthernNamibDesert 0.2821229   1   Rarefaction       0  23.867  21.581  26.152
85  SouthernNamibDesert 0.6993084   9   Rarefaction       0 112.484  97.518 127.451
94  SouthernNamibDesert 0.8931001  30 Extrapolation       0 207.213 171.199 243.227
103 SouthernNamibDesert 0.9946808  93 Extrapolation       0 259.337 207.038 311.636
111 SouthernNamibDesert 0.9996478 150 Extrapolation       0 261.886 208.751 315.021
```

```
NOTE: The above output only shows five estimates for each assemblage; call
iNEXT.object$iNextEst$coverage_based to view complete output.

$AsyEst: asymptotic diversity estimates along with related statistics.
          Assemblage         Diversity Observed Estimator   s.e.     LCL     UCL
1  CentralNamibDesert  Species richness  130.000   196.706 19.523 158.441 234.971
2  CentralNamibDesert Shannon diversity   81.812   106.480  5.291  96.110 116.850
3  CentralNamibDesert Simpson diversity   54.225    59.556  3.175  53.333  65.778
4           EtoshaPan  Species richness  216.000   339.255 23.121 293.938 384.571
5           EtoshaPan Shannon diversity  158.367   222.936 11.125 201.130 244.741
6           EtoshaPan Simpson diversity  116.677   142.833  8.700 125.780 159.885
7 SouthernNamibDesert  Species richness  150.000   262.067 30.638 202.018 322.115
8 SouthernNamibDesert Shannon diversity  103.705   149.910  9.301 131.681 168.139
9 SouthernNamibDesert Simpson diversity   72.327    84.597  5.276  74.255  94.938
```

POINT ESTIMATION FUNCTION: estimateD()

We also supply the following function

```
estimateD(x, datatype="abundance", base="size", level=NULL)
```

to compute diversity estimates with q = 0, 1, 2 for any particular level of sample size (`base="size"`) or any specified level of sample coverage (`base="coverage"`) for abundance data (`datatype="abundance"`) or incidence data (`datatype="incidence_freq" or "incidence_raw"`). If `base="size"` and `level=NULL`, then this function computes the diversity estimates for the minimum among all doubled reference sample sizes. If `base="coverage"` and `level=NULL`, then this function computes the diversity estimates for the minimum among the coverage values for samples extrapolated to double the size of the reference sample.

The following command returns the species diversity with a specified level of sample coverage of 98.5% for the ant data. For some assemblages, this coverage value corresponds to rarefaction (i.e., less than the coverage of the reference sample), while for the others it corresponds to extrapolation (i.e., greater than the coverage of the reference sample), as indicated under the `method` column of the output.

```
estimateD(ant, datatype="incidence_freq",
          base="coverage", level=0.985, conf=0.95)
```

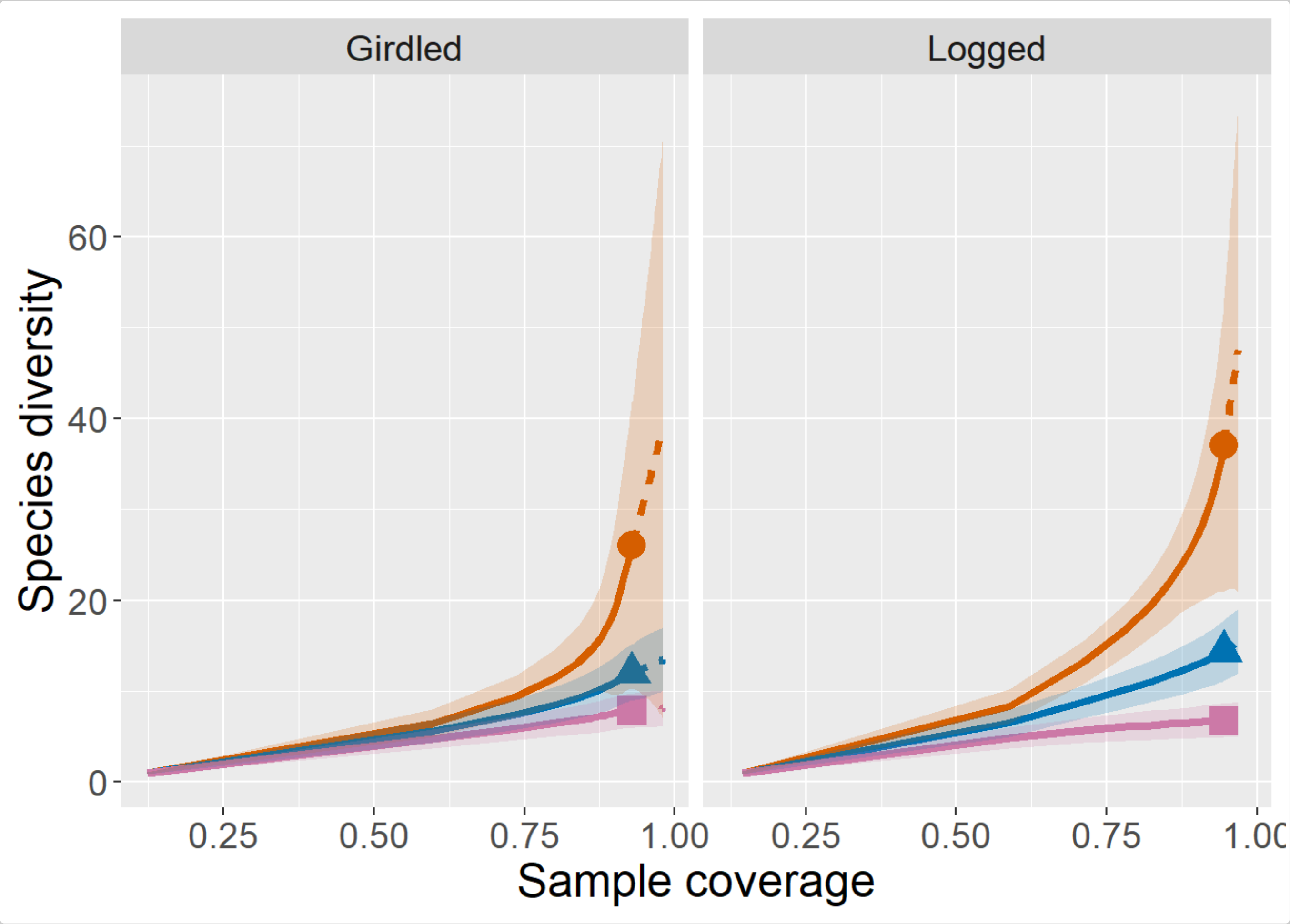
```
   Assemblage       t        Method Order.q    SC      qD  qD.LCL  qD.UCL
1        h50m 327.165   Rarefaction       0 0.985 197.488 186.058 208.918
2        h50m 327.165   Rarefaction       1 0.985  78.053  75.389  80.717
3        h50m 327.165   Rarefaction       2 0.985  50.461  48.640  52.282
4       h500m 342.859 Extrapolation       0 0.985 268.726 242.802 294.650
5       h500m 342.859 Extrapolation       1 0.985 103.847 100.256 107.438
6       h500m 342.859 Extrapolation       2 0.985  64.758  61.983  67.534
7      h1070m 158.951 Extrapolation       0 0.985 123.609 113.000 134.218
8      h1070m 158.951 Extrapolation       1 0.985  59.592  56.903  62.280
9      h1070m 158.951 Extrapolation       2 0.985  41.775  39.465  44.085
10     h1500m 125.959   Rarefaction       0 0.985  50.479  41.666  59.291
11     h1500m 125.959   Rarefaction       1 0.985  26.249  24.575  27.923
12     h1500m 125.959   Rarefaction       2 0.985  18.649  17.446  19.852
13     h2000m 104.631   Rarefaction       0 0.985  12.910  11.002  14.817
14     h2000m 104.631   Rarefaction       1 0.985   7.711   6.915   8.506
15     h2000m 104.631   Rarefaction       2 0.985   5.795   5.079   6.510
```

Hacking ggiNEXT()

The `ggiNEXT()` function is a wrapper around the `ggplot2` package to create a R/E curve using a single line of code. The resulting object is of class `"ggplot"`, so it can be manipulated using the `ggplot2` tools. The following are some useful examples for customizing graphs.

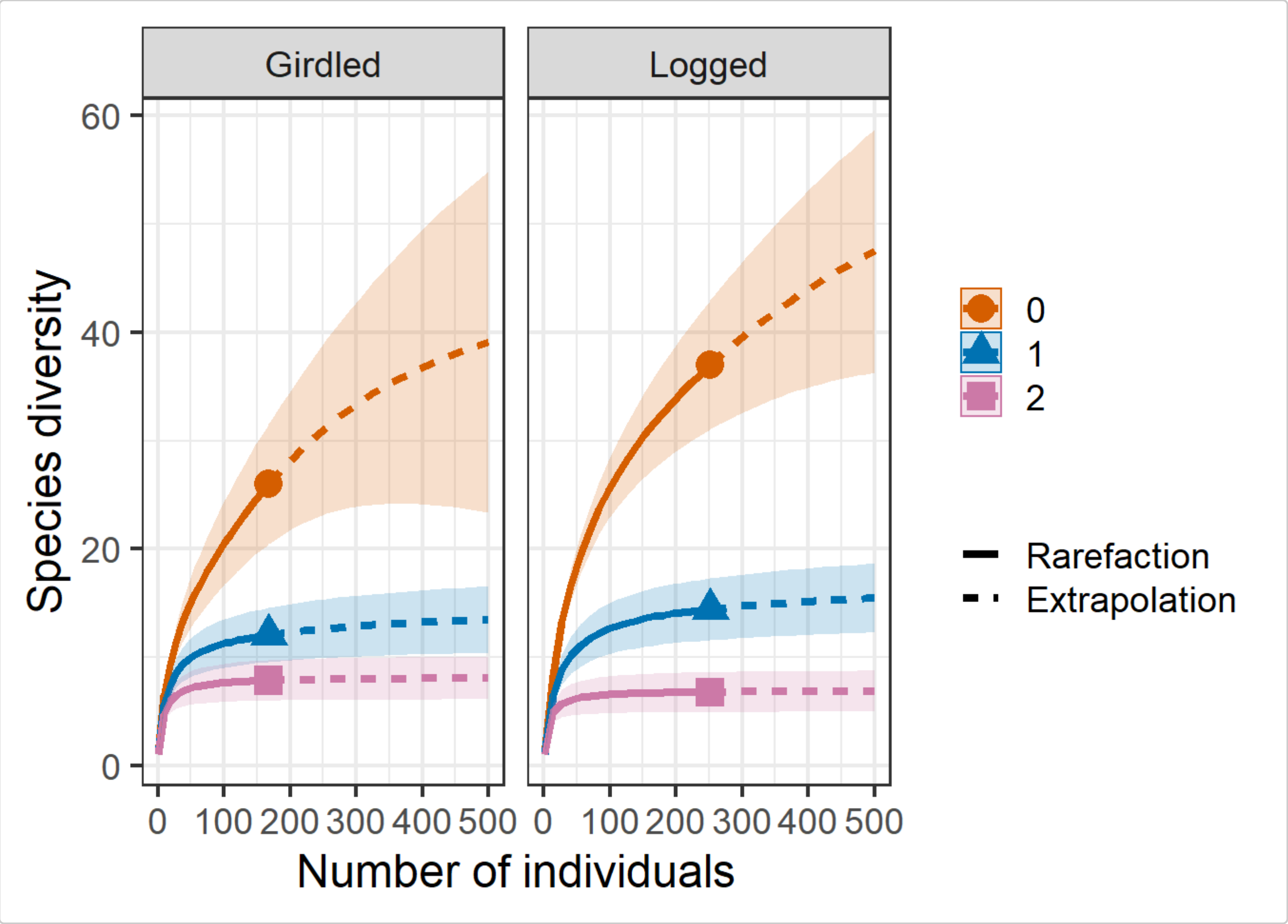
Remove legend

```
ggiNEXT(out, type=3, facet.var="Assemblage") +
  theme(legend.position="None")
```



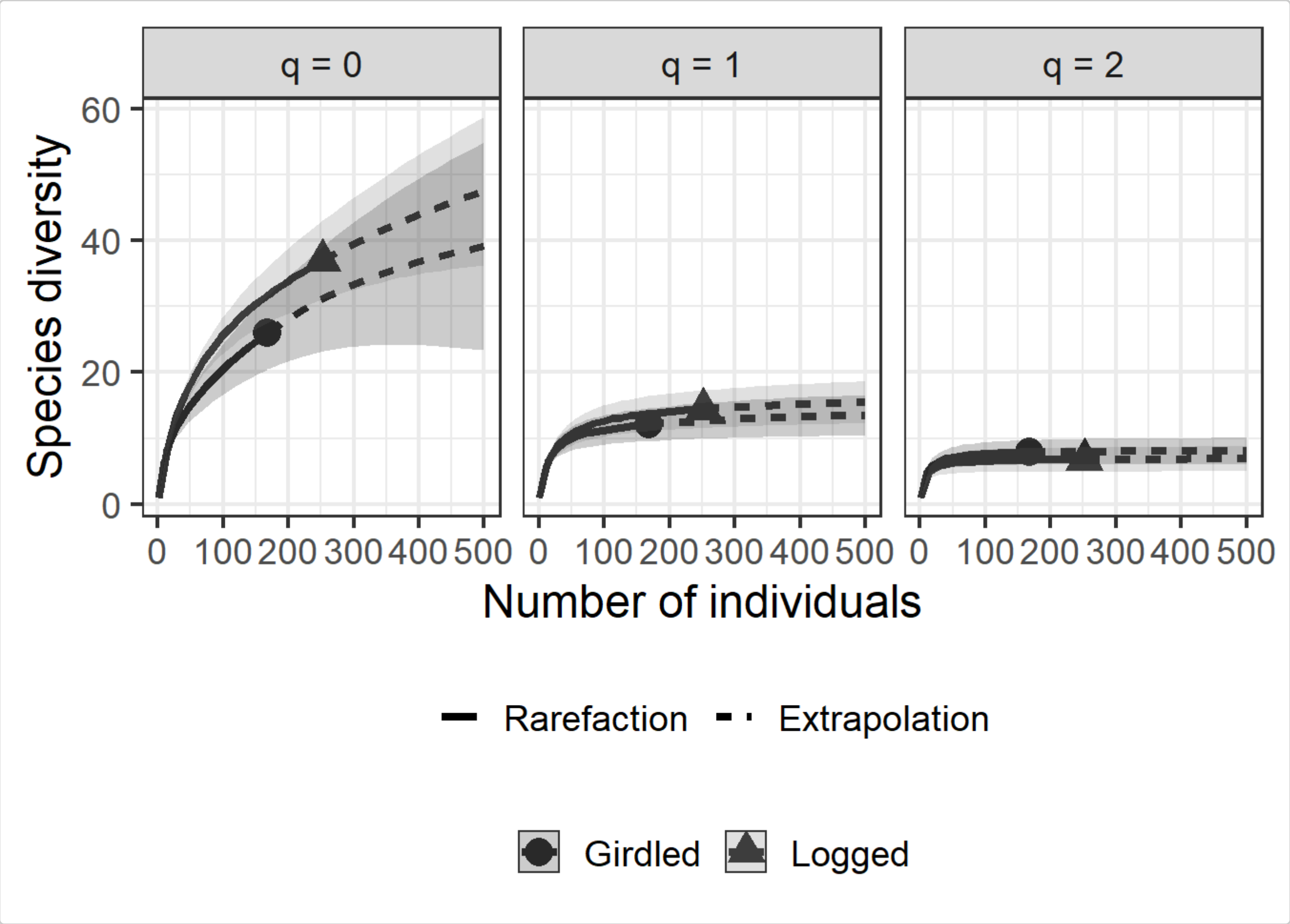
Change the theme and legend.position

```
ggiNEXT(out, type=1, facet.var="Assemblage") +
  theme_bw(base_size = 18) +
  theme(legend.position="right")
```



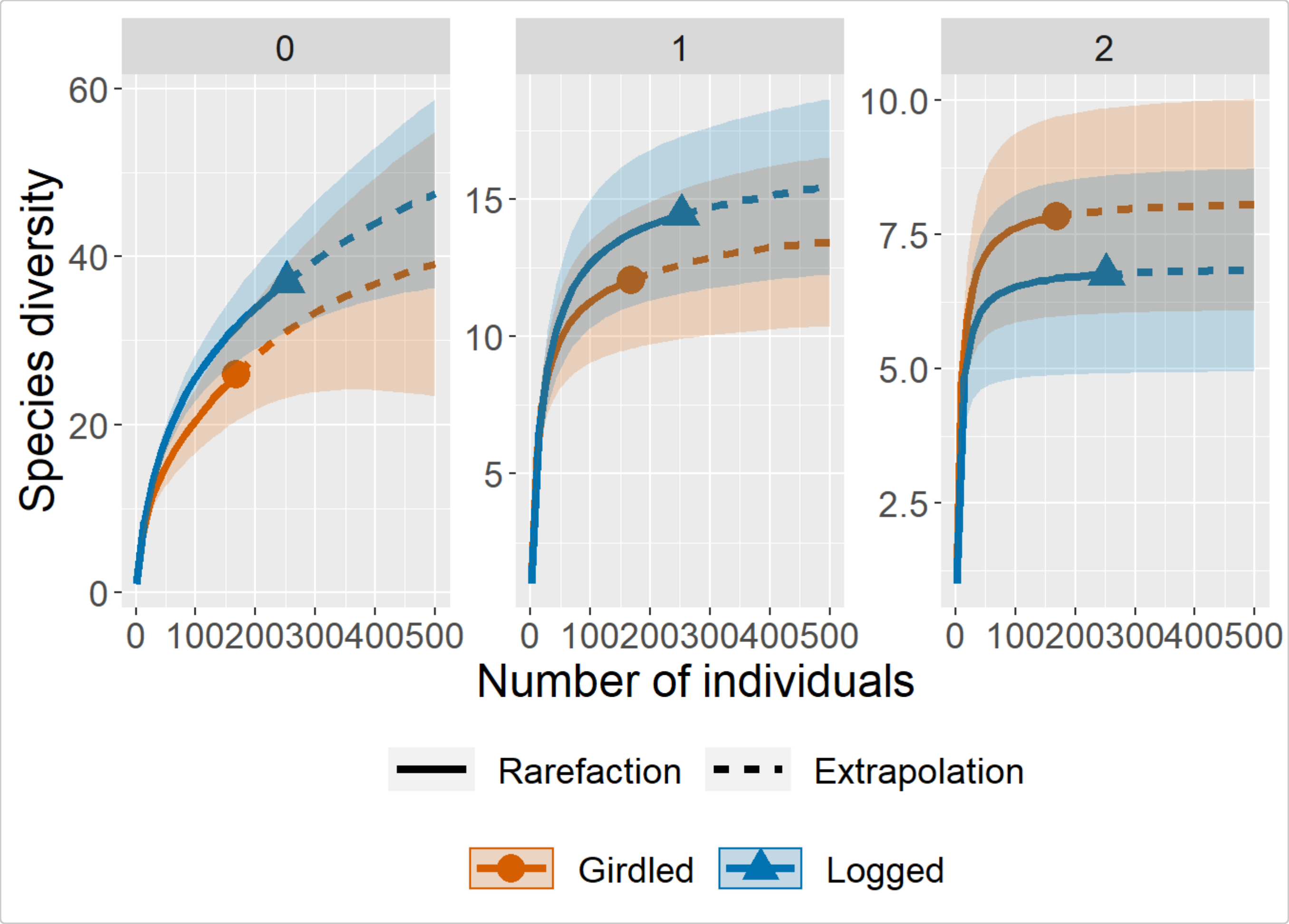
Display black-white figures

```
ggiNEXT(out, type=1, facet.var="Order.q", grey=TRUE)
```



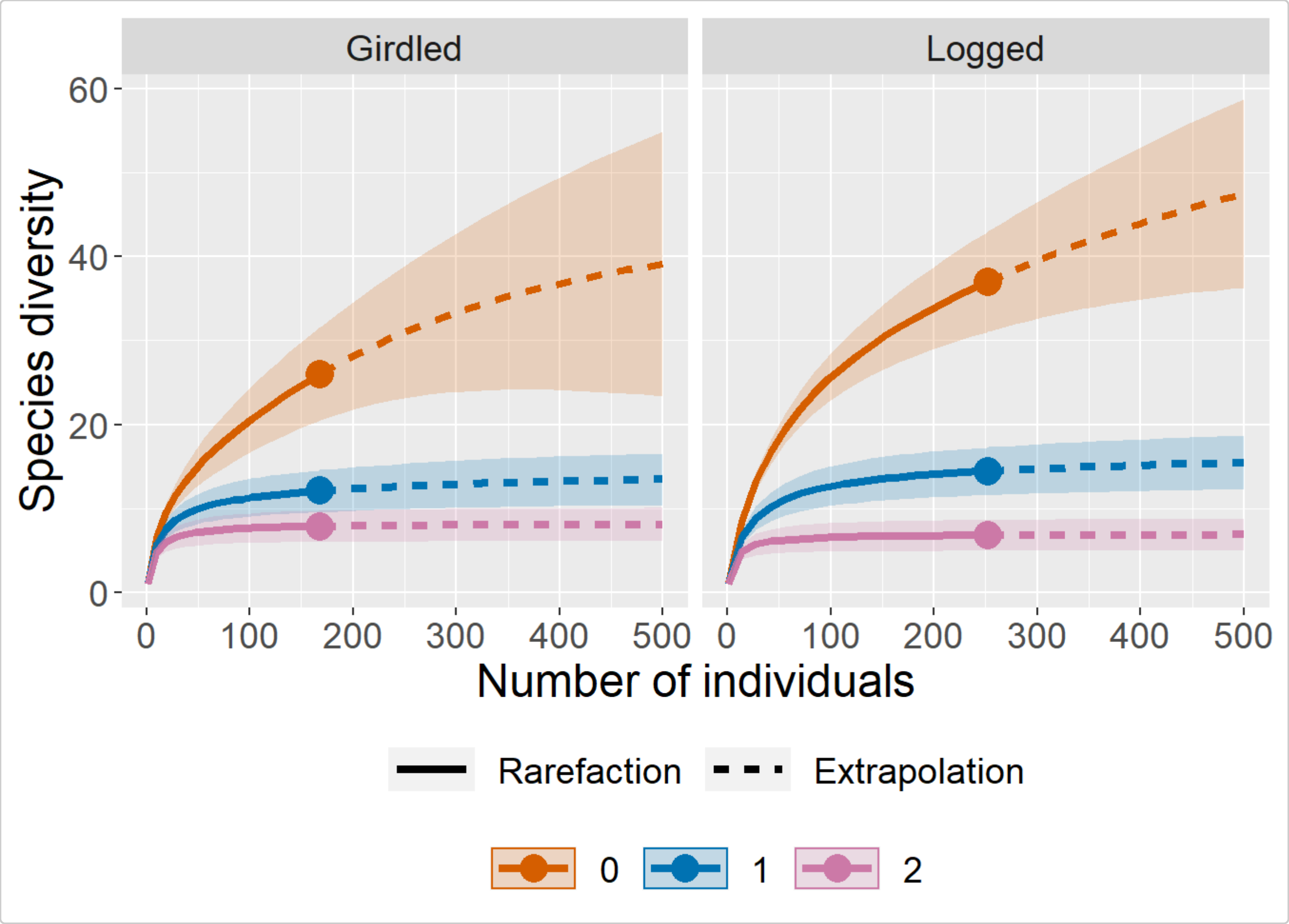
Free the scales of the axis

```
ggiNEXT(out, type=1, facet.var="Order.q") +
  facet_wrap(~Order.q, scales="free")
```



change the shape of the reference sample point

```
ggiNEXT(out, type=1, facet.var="Assemblage") +
  scale_shape_manual(values=c(19,19,19))
```



General customization

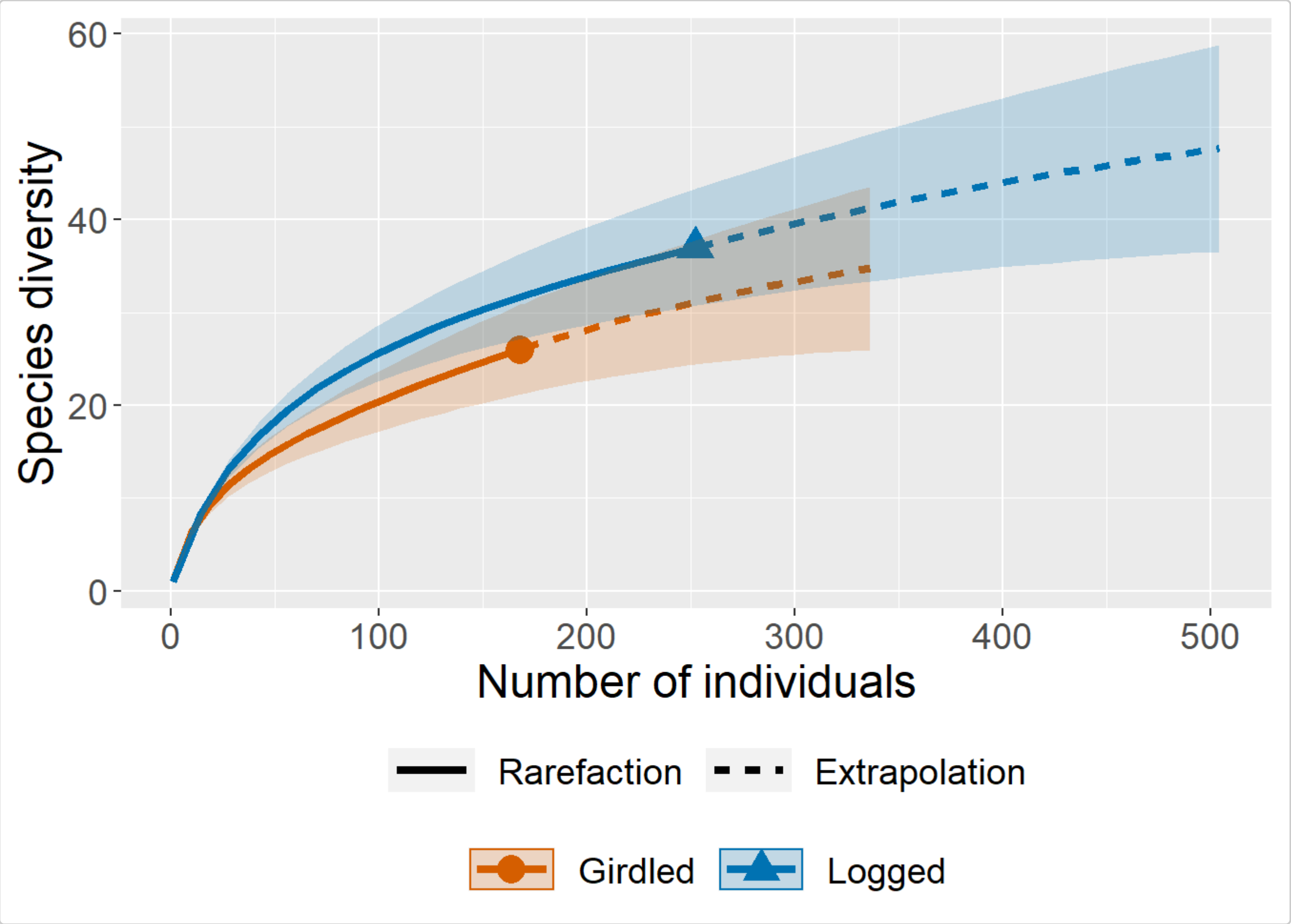
The data visualization package `ggplot2` provides the `scale_` function to customize data which is mapped into an aesthetic property of a `geom_`. The following functions would help user to customize `ggiNEXT` output.

- change point shape: `scale_shape_manual`
- change line type : `scale_linetype_manual`
- change line color: `scale_colour_manual`
- change band color: `scale_fill_manual`
  see quick reference for style setting.

Example: `spider` data

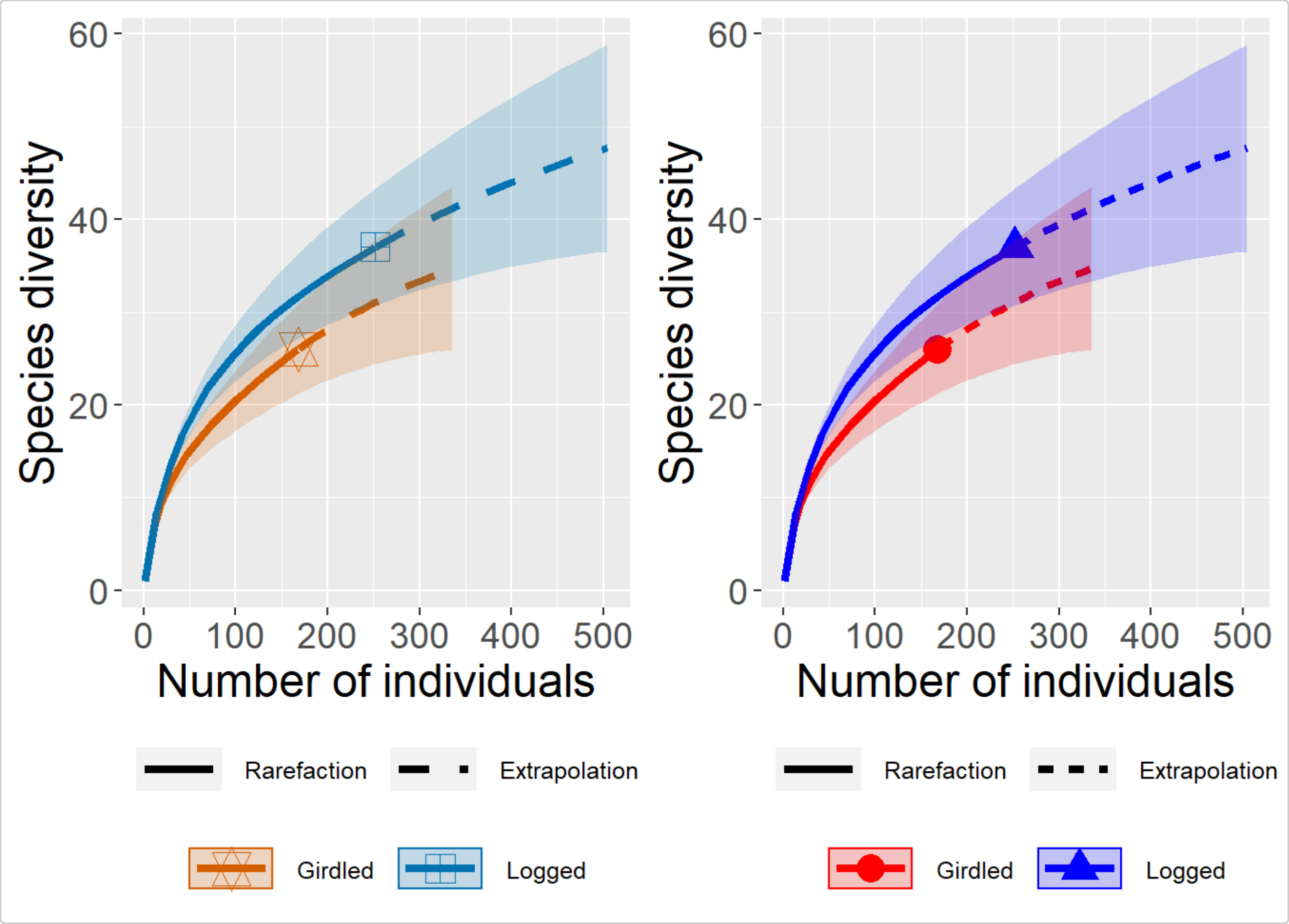
To show how to custmize `ggiNEXT` output, we use abundance-based data `spider` as an example.

```
library(iNEXT)
library(ggplot2)
library(gridExtra)
library(grid)
data("spider")
out <- iNEXT(spider, q=0, datatype="abundance")
g <- ggiNEXT(out, type=1, color.var = "Assemblage")
g
```



Change shapes, line types and colors

```
g1 <- g + scale_shape_manual(values=c(11, 12)) +
          scale_linetype_manual(values=c(1,2)) +
          theme(legend.text = element_text(size = 9.5))
g2 <- g + scale_colour_manual(values=c("red", "blue")) +
          scale_fill_manual(values=c("red", "blue")) +
          theme(legend.text = element_text(size = 9.5))
# Draw multiple graphical object on a page
# library(gridExtra)
grid.arrange(g1, g2, ncol=2)
```



Customize point/line size by hacking

In order to change the size of the reference sample point or rarefaction/extrapolation curve, the user need to modify the `ggplot` object.

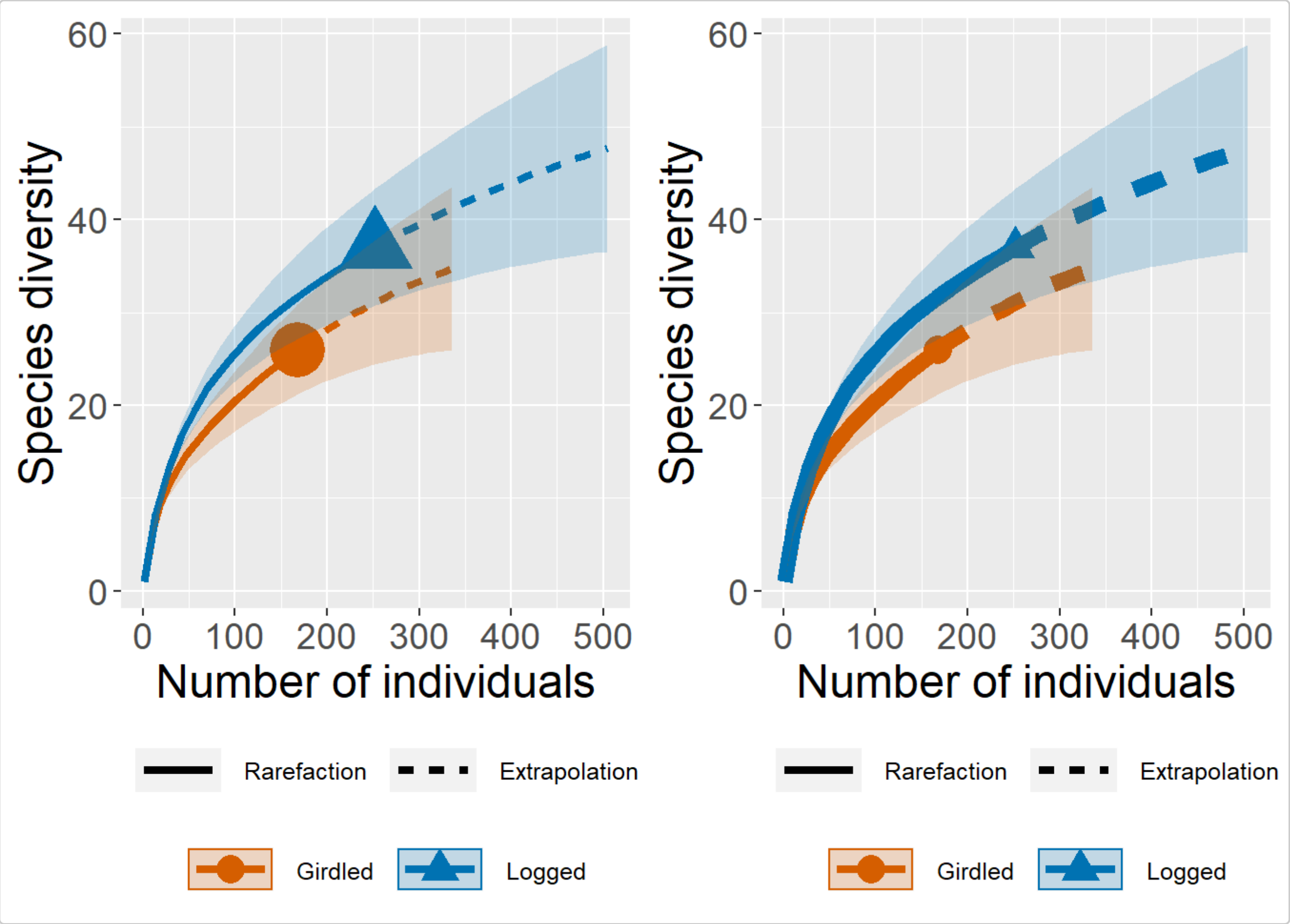
- change point size:
  the reference sample size point is drawn on the first layer by `ggiNEXT`. Hack the point size by the following

```
# point is drawn on the 1st layer, default size is 5
gb3 <- ggplot_build(g + theme(legend.text = element_text(size = 9.5)))
gb3$data[[1]]$size <- 10
gt3 <- ggplot_gtable(gb3)
# use grid.draw to draw the graphical object
# library(grid)
# grid.draw(gt3)
```

- change line width (size):
  the reference sample size point is drawn on the second layer by `ggiNEXT`. Hack the point size by the following

```
# line is drawn on the 2nd layer, default size is 1.5
gb4 <- ggplot_build(g + theme(legend.text = element_text(size = 9.5)))
gb4$data[[2]]$size <- 3
gt4 <- ggplot_gtable(gb4)
# grid.draw(gt4)
```

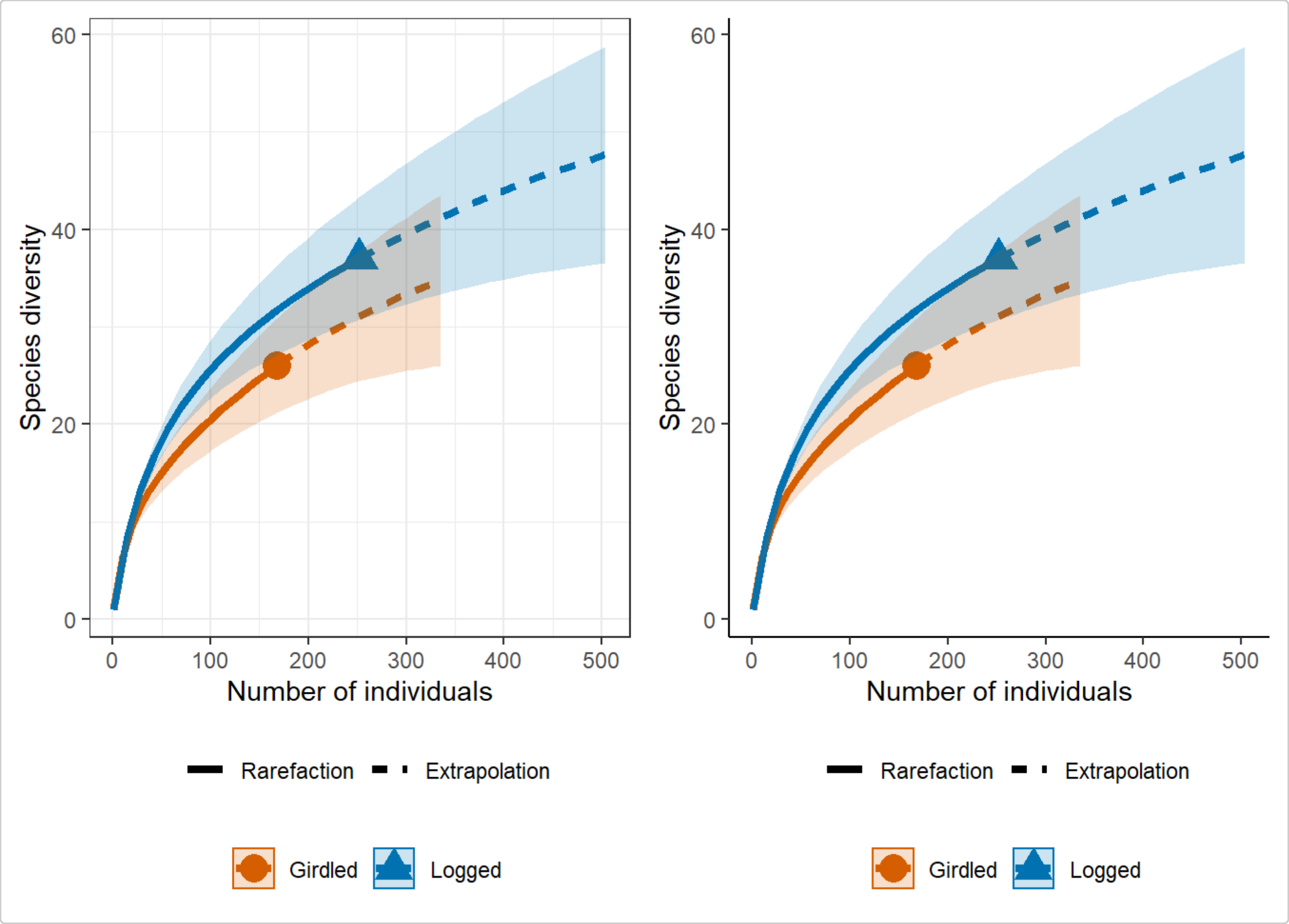
```
grid.arrange(gt3, gt4, ncol=2)
```



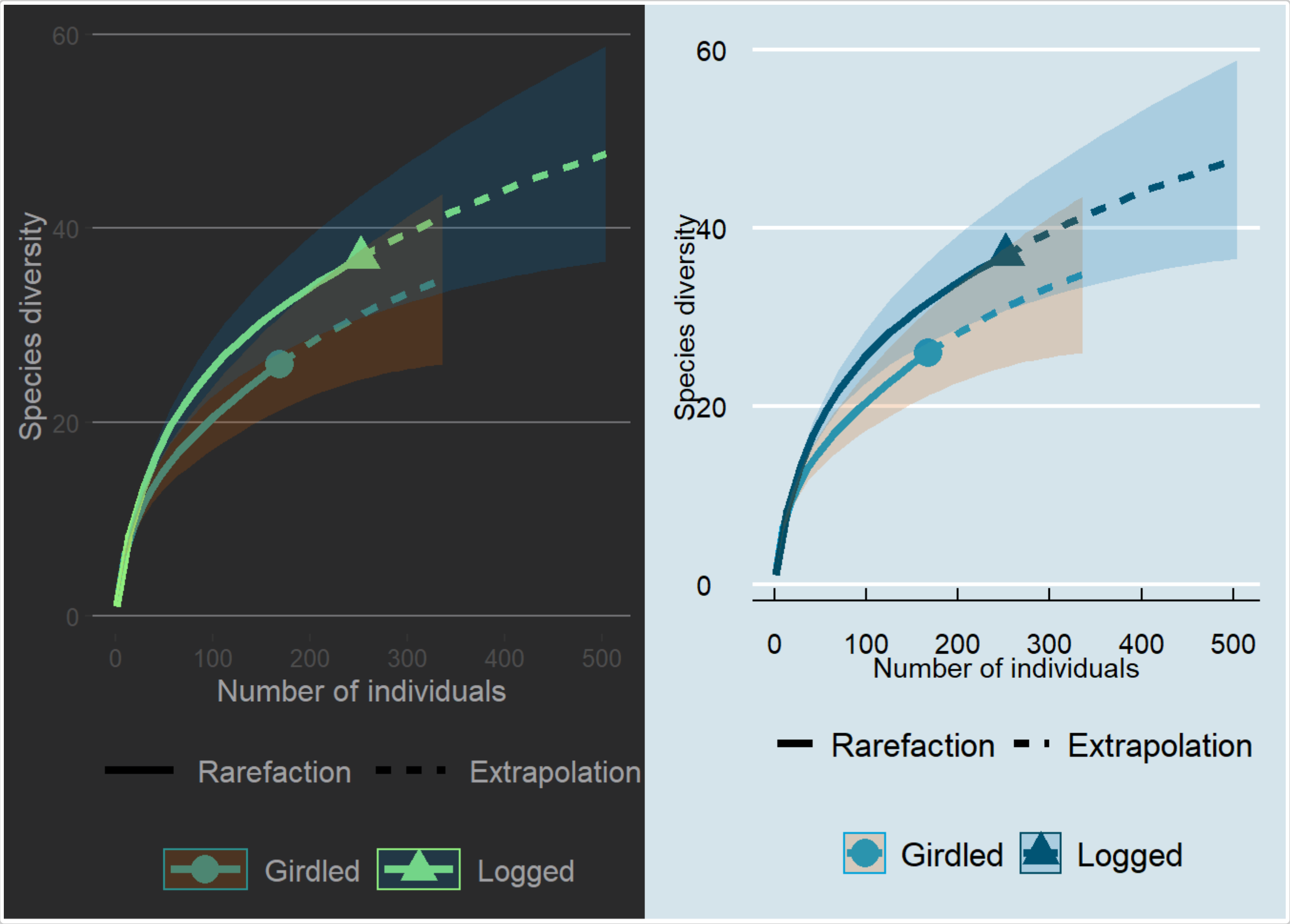
Customize theme

A `ggplot` object can be themed by adding a theme. The User could run `help(theme_grey)` to show the default themes in `ggplot2`. Further, some extra themes are provided by the `ggthemes` package. Examples are shown in the following:

```
g5 <- g + theme_bw() + theme(legend.position = "bottom", legend.box = "vertical")
g6 <- g + theme_classic() + theme(legend.position = "bottom", legend.box = "vertical")
grid.arrange(g5, g6, ncol=2)
```



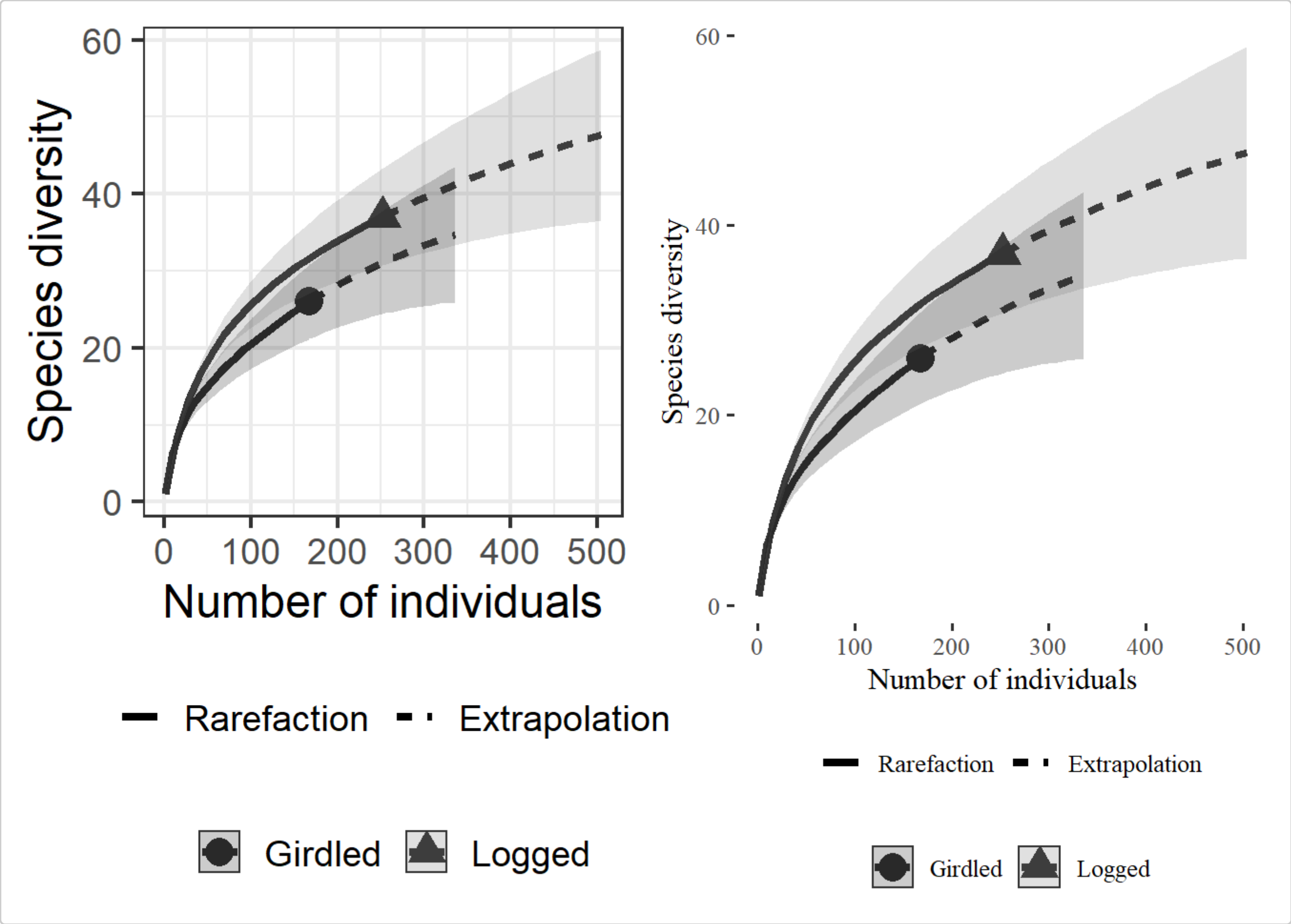
```
library(ggthemes)
g7 <- g + theme_hc(bgcolor = "darkunica") +
  theme(legend.box = "vertical",
        legend.text = element_text(size = 12)) +
  scale_colour_hc("darkunica")
g8 <- g + theme_economist() +
  theme(legend.position="bottom",
        legend.box = "vertical",
        legend.text = element_text(size = 13)) +
  scale_colour_economist()
grid.arrange(g7, g8, ncol=2)
```



Black-White figures

The following are custmized themes for black-white figures. To modifiy the legend, see Cookbook for R for more details.

```
g9 <- g + theme_bw(base_size = 18) +
      scale_fill_grey(start = 0, end = .4) +
      scale_colour_grey(start = .2, end = .2) +
      theme(legend.position="bottom",
            legend.title=element_blank(),
            legend.box = "vertical")
g10 <- g + theme_tufte(base_size = 12) +
    scale_fill_grey(start = 0, end = .4) +
    scale_colour_grey(start = .2, end = .2) +
    theme(legend.position="bottom",
          legend.title=element_blank(),
          legend.box = "vertical")
grid.arrange(g9, g10, ncol=2)
```

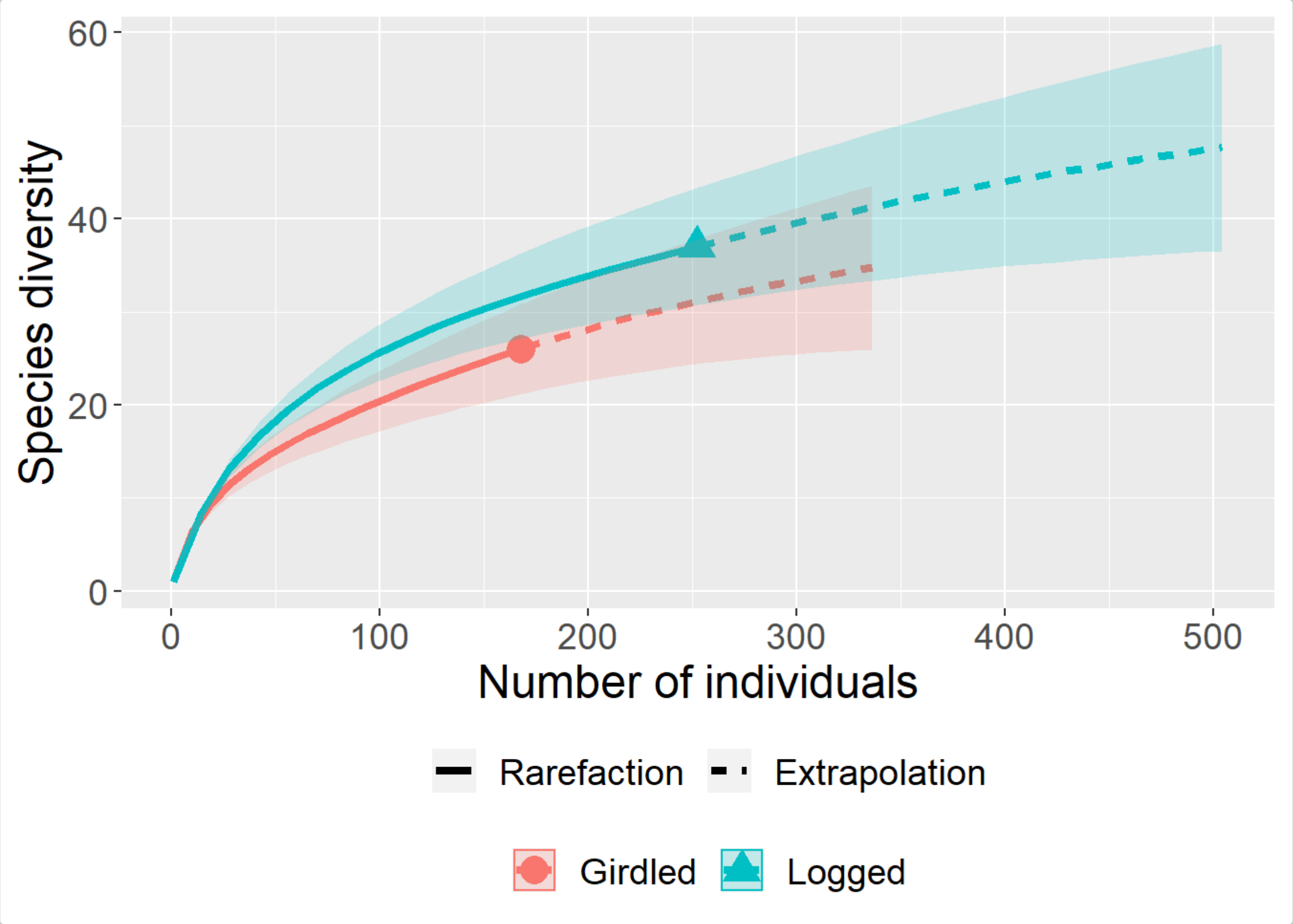


Draw R/E curves by yourself

In iNEXT, we provide a S3 `ggplot2::fortify` method for class `iNEXT`. The function `fortify` offers a single plotting interface for rarefaction/extrapolation curves. Set argument `type = 1, 2, 3` to plot the corresponding rarefaction/extrapolation curves.

```
df <- fortify(out, type=1)
head(df)
   datatype plottype Assemblage      Method Order.q  x         y     y.lwr    y.upr
1 abundance        1    Girdled Rarefaction       0  1  1.000000  1.000000  1.00000
2 abundance        1    Girdled Rarefaction       0 10  6.478617  5.983295  6.97394
3 abundance        1    Girdled Rarefaction       0 19  9.450323  8.530705 10.36994
4 abundance        1    Girdled Rarefaction       0 28 11.514220 10.253775 12.77466
5 abundance        1    Girdled Rarefaction       0 37 13.126817 11.575192 14.67844
6 abundance        1    Girdled Rarefaction       0 47 14.622424 12.778255 16.46659
df.point <- df[which(df$Method=="Observed"),]
df.line <- df[which(df$Method!="Observed"),]
df.line$Method <- factor(df.line$Method,
                        c("Rarefaction", "Extrapolation"),
                        c("Rarefaction", "Extrapolation"))

ggplot(df, aes(x=x, y=y, colour=Assemblage)) +
  geom_point(aes(shape=Assemblage), size=5, data=df.point) +
  geom_line(aes(linetype=Method), lwd=1.5, data=df.line) +
  geom_ribbon(aes(ymin=y.lwr, ymax=y.upr,
                  fill=Assemblage, colour=NULL), alpha=0.2) +
  labs(x="Number of individuals", y="Species diversity") +
```

```
  theme(legend.position = "bottom",
        legend.title=element_blank(),
        text=element_text(size=18),
        legend.box = "vertical")
```



License

The iNEXT package is licensed under the GPLv3. To help refine `iNEXT`, your comments or feedback would be welcome (please send them to Anne Chao or report an issue on the iNEXT github iNEXT_github.

References

- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs, 84, 45–67.
- Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology, 93, 2533–2547.
- Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.-Y., Mao, C.X., Chazdon, R.L. & Longino, J.T. (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology, 5, 3–21.
- Ellison, A.M., Barker-Plotkin, A.A., Foster, D.R. & Orwig, D.A. (2010) Experimentally testing the role of foundation species in forests: the Harvard Forest Hemlock Removal Experiment. Methods in Ecology and Evolution, 1, 168–179.
- Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: An R package for interpolation and extrapolation of species diversity (Hill numbers). Methods in Ecology and Evolution, 7, 1451-1456.

- Longino, J.T. & Colwell, R.K. (2011) Density compensation, species composition, and richness of ants on a neotropical elevational gradient. Ecosphere, 2:art29.