

CTFS R FORMATTED DATA TABLES

R. CONDIT

FULL TREE TABLES

These tables are stored in files *Plotname.full#.rdata*, where # refers to the census number. The R object within is a dataframe of the same name. There is one file and one dataframe for each census.

All the full dataframes from one plot have exactly the same number of rows, with one row for every tree ever found in any census, in exactly the same order. Trees thus appear before they recruit, while they are alive, and after they die. Because rows match, it is convenient to work with multiple censuses, as if all the data were in one table. Additionally, because tables from every census have precisely the same format, identical analyses can easily be repeated on more than one census, and demographic rates can be calculated between any pair of censuses.

0.1. Census 5 from BCI data. Attach the Rdata file for a single census. (The folder defined by 'mydatapath' is where it is stored on my computer.)

```
> mydatapath='/home/fullplotdata/full/'
> if(!exists('bci.full15'))
+   attach(paste(mydatapath,'bci.full15.rdata',sep=''))
> str(bci.full15)

'data.frame':      368122 obs. of  20 variables:
 $ treeID   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ stemID   : int  NA 1 ...
 $ tag      : chr  "-05599" "-22851" "-24362" "-26589" ...
 $ StemTag  : chr  NA NA NA NA ...
 $ sp       : chr  "swars1" "hybapr" "aegipa" "beilpe" ...
 $ quadrat  : chr  "4007" "0718" "0417" "0007" ...
 $ gx       : num  800.2 151.5 95.2 11.7 7.7 ...
 $ gy       : num  152.2 378.8 357.5 151.1 96.2 ...
 $ MeasureID: int  NA 1 ...
 $ CensusID : int  NA 5 ...
 $ dbh      : num  NA ...
 $ pom      : chr  NA NA NA NA ...
 $ hom      : num  NA ...
 $ ExactDate: chr  NA NA NA NA ...
 $ DFstatus : chr  "dead" "dead" "dead" "dead" ...
 $ codes    : chr  NA NA NA NA ...
 $ nostems  : num  NA 1 ...
 $ status   : chr  "D" "D" "D" "D" ...
 $ date     : num  14801 14656 14666 14636 14629 ...
 $ agb      : num  [1:368122(1d)] 0 0 0 0 0 0 0 0 0 0 ...
 ..- attr(*, "dimnames")=List of 1
 .. ..$ : chr  "1" "2" "3" "4" ...
```

Date: September 20, 2010.

Description of columns in the *full* R format tables. Each row is data from a single tree. If a tree has multiple stems, the dbh of only one is given. Usually, the same stem of a tree appears in this table in successive censuses, unless that stem was lost or damaged. The stemID is used to determine whether the stem changed. Usually, the largest stem is the one whose dbh is given, but not always (see the *stem* for all dbhs). The biomass in this table, though, is for the entire tree, or the sum of the biomass of each stem.

Column name	Description
treeID	The unique tree identifier in CTFS database. Useful to be certain in matching trees.
stemID	The unique stem identifier in CTFS database. Useful to be certain in matching stems.
tag	Tag number used in the field.
StemTag	Tag number on the individual stem, if present.
sp	The species mnemonic. To get full species names, the taxonomy table must be downloaded from the CTFS database: http://ctfs.arnarb.harvard.edu/CTFSReports
quadrat	Quadrat designation.
gx	The x coordinate within the plot, relative to one edge of the plot.
gy	The y plot coordinate.
MeasureID	The unique identifier of a single measurement in the CTFS database.
CensusID	The numeric identifier of the census.
dbh	Diameter of one stem on the tree, the stem whose stemID is given.
pom	The point-of-measure, where the diameter was taken, identical to hom, but a character variable with only 2 decimal places.
hom	The height-of-measure, identical to pom but a numeric variable with full precision.
ExactDate	The date on which a tree was measured.
DFstatus	The status taken from the DFtemp table of the CTFS database: alive, dead, lost_stem, missing, or prior ¹
codes	The codes for the measurement as recorded in the field.
nostems	The number of living stems on the date of measurement.
status	An abbreviated version of status, for compatibility with earlier versions of functions in the CTSF R package: A, D, M, P
date	The julian date, for date arithmetic.
agb	Above-ground-biomass of all stems on the tree, in Mg (= metric tons or 10 ⁶ g). Note that agb=0 for dead trees.

¹ Status. Alive (A) and dead (D) refer to the entire tree, so if any stem is alive, the tree is alive, and a tree is only dead when every stem is dead. Status = 'lost_stem' indicates that the stem had the associated code; it usually means the was broken in the given census, while the tree had no other stem. Status = 'missing' (M) are cases where dbh and codes for a tree were not given, so it is not certain whether the tree was alive or dead. Status = 'prior' (P) indicates a tree had not yet recruited at this census. The lost_stem status is now deprecated, since it should always be safer to check stemID to determine whether a tree's measurement changed stems between censuses.

Sample queries of tree tables from two censuses, using BCI census 5 and 6 to illustrate.

Attach a dataset from another census:

```
> if(!exists('bci.full6'))
+ attach(paste(mydatapath, 'bci.full6.rdata', sep=' '))
```

Because rows match, it is possible to check the status in two censuses of all trees. This shows that of all trees alive in census 5 at BCI, 198778 were still alive in census 6, while 28432 had died.

```
> table(bci.full5$status, bci.full6$status, exclude=NULL)
```

	A	D	M	<NA>
A	199037	28432	0	0
D	276	117034	0	0
M	0	1	52	0
P	23290	0	0	0
<NA>	0	0	0	0

The number of diameters of sizes 10-19 mm:

```
> head(table(subset(bci.full5, status=='A')$dbh), 10)
 10  11  12  13  14  15  16  17  18  19
4704 8607 9671 9162 8835 8705 7748 7369 6941 6133
```

Again, with matching rows, growth is easy to calculate (multiplying by the number of days in a year, since the date field is in days. (I rounded down to 2 digits for clearly display.) This shows the 10 fastest growth rates, of which the first 8 at least must be errors, then checks that tree which grew over 30 cm per year (it's a giant strangler fig for which a decimal place must have been missed).

```
> growth=round(365.25*(bci.full6$dbh-bci.full5$dbh)/(bci.full6$date-bci.full5$date), 2)
```

```
> head(sort(growth, decreasing=TRUE), 10)
```

```
[1] 319.35 220.18 171.06 98.03 96.59 90.78 88.93 79.79 58.51 54.50
```

```
> subset(bci.full5, growth>300,
```

```
+   select=c('tag', 'stemID', 'sp', 'dbh', 'agb', 'pom', 'ExactDate', 'status'))
```

	tag	stemID	sp	dbh	agb	pom	ExactDate	status
929	000922	1	ficutr	203	0.1917631	3.94	2000-09-20	A

```
> subset(bci.full6, growth>300,
```

```
+   select=c('tag', 'stemID', 'sp', 'dbh', 'agb', 'pom', 'ExactDate', 'status'))
```

	tag	stemID	sp	dbh	agb	pom	ExactDate	status
929	000922	1	ficutr	1817	36.33648	6.70	2005-10-10	A