Center for Tropical Forest Science R Package Manual
Pamela Hall, Suzanne Lao, Ellen Connell and Marie Massa
Version 1.00 March 29, 2006

4.0 File organization, Data, Functions and Results

CTFS colleagues working with R generate many large datasets, graphs and tables. Although there are a number of ways to effectively organize CTFS R files, we recommend a system which assures the integrity of the CTFS data files, R functions and helps to keep you from spending lots of time searching through files to figure out how they differ or how and why they were created.

4.1 Basic File Organization of R

The first step to organizing personal R files is to develop an understanding of where general R packages, functions, manuals and documentation are located.

The R folder is the folder into which all of the R files are placed upon installation, i.e. $R_HOME, which contains folders with various files. The actual R program file and Rgui is located in $R_HOME. The file *.rData* which is created and modified when the user saves their workspace is also located in $R_HOME. In addition to the files needed to run R, $R_HOME also contains help documentation as well as various other files.

$R_HOME's folder "library", contains all of the R packages installed on a given computer. Each installed package is contained within its own folder. As explained in Chapter 1, a package is a bundle of functions, documentation, and datasets. Upon installation, R includes more than 30 packages each with its own folder in the library folder. Each package folder contains a sub-folder called "R." This folder contains all of functions specific to a particular package. Oftentimes, all of a package's functions are contained within a single file. Most package folders also contain a sub-folder called "html" in which all of the package's html help pages are stored. These are the html help pages that the user can access using the R help facility described in chapter 3. The user can also directly open these files by selecting the files themselves. This will open a Windows Explorer window outside of R.

In the MacOS the file structure is substantially the same, depending on whether R was installed for a single or multiple users. If in stalled for multiple users (at the system level) than the R files will be in the Library folder in an R.framework folder. A locally installed R appears in ~user/Library/R/library. ("~user" refers to the home folder of a given user). Packages can be installed at the system or user level.

In general, it is not necessary to know where R is actually installed. In fact, DO NOT PLACE ANY OF YOUR PERSONAL CTFS R DATASETS OR FUNCTIONS IN THE $R_HOME OR THE /LIBRARY/R FOLDER. Leave these folders alone, but do explore them and see where the R functions are kept, the html pages, manual chapters and the source files for help pages and functions. The help pages and manual chapters can be directly accessed from here or used during an R session. Do not access directly.

4.2 CTFS Data Files

I strongly suggest you use the features of your operating system that allows you to keep a separate folder structure for user specific files. In Windows this is by using the shortcut "MyDocuments". Don't worry about where exactly this is located on the C drive, just use the shortcut and keep ALL of the CTFS datafiles, results and functions YOU write, here. In MacOS the equivalent is the Documents folder in your user folder. In Unix, each user has their own folder location also. In the rest of this document I will refer to this location as the user's home folder or USERHOME. DO NOT intermingle CRAN R and CTFS R Package functions and data with those you create yourself.

You need to retain the integrity of your CTFS master datasets. Create a folder within USERHOME for these master datasets. As you create datasets from analyses that are derivative of the master datasets and are sure these new datasets are of lasting value, you may wish to keep safe copies of them also in this folder location. I recommend subfolders to differentiate between datasets that contain census and site data vs those that are derived from this data.

For example:

USERHOME/CTFS datasets
    /Census
    /Plot
    /Neighborhood

The /Census folder contains all the census datasets, the /Plot folder contains topographic data, and /Neighborhood contains datasets derived from the Census data that describe the neighborhoods for each tree and/or species they have been created for.

4.3 CTFS R Functions

Place functions that you write to do an analysis in a separate folder structure. If you are working on a specific analysis, you may want to place the related functions into separate folders. You can also put related functions into the same file, which is how the CTFS R package functions are structured. You may have gotten functions from colleagues and you can mix these in with yours as is appropriate or keep them separate. But DO NOT put these functions in with the R or CTFS R package functions.

4.4 CTFS results

While using R, you will generate lots of results. These only exist in R environment when a given session is running. As soon as you quite R without saving files or workspace, all of these files will be lost. If you want to save files you can save the workspace, but remember ALL variables you have created in the workspace are saved, so this can get very cumbersome and confusing if you do it too much. If there are results you want to save to disk, do this explicitly with the save function or by saving a graphic window (either in a function or via the Print/PDF facility on the MacOS). If you are in the middle of analysis and not sure of the value of a given file, you can save it with "tst" or "tmp" in its name to indicate it temporary and disposable nature. Don't forget to CLEAN UP once in a while!

We highly recommend that users name their datasets, functions and results files clearly and with consistency. Functions should have names that relate to what they actually do. Dataset names should be sufficient for you to figure out their contents and origin with a glance.
Use a short abbreviations you can easily remember to indicate the type of information in the file. "res" for results from a function, eg "mort.res" for results from a mortality analysis. More specifically something like "bci.mort.dbh.9095.res" is the results of running the mortality function using dbh categories for the 1990 to 1995 census on the BCI dataset. If results are a graph, try something like "mort.dbh.9095.grf" so that you know it's a graph and not the direct results of the mortality function.

4.5 CTFS related help pages

A CTFS R package help page that is relevant to this chapter is *CTFS.datafiles*. It describes the structure of the CTFS datasets.