Introduction to R for distance sampling

Workshop, 19-20 August 2019

Centre for Research into Ecological and Environmental Modelling

Brief introduction to R

This document is meant as a quick reference guide to some of the terms and special characters used in `R`. It is not a guide to the functions used throughout the workshops.

R (R Core Team, 2018) is a software environment for statistical computing and graphics. It is freely available and can be downloaded from https://cran.r-project.org/. There are versions for Windows and Mac users. RStudio is an interface designed to make R easier to use: it can also be downloaded for free from https://www.rstudio.com/.

Users of R have developed packages to expand the basic features and capabilities of R. The main packages that will be used throughout the workshops are listed below.

Terminology

arguments/default values - parameters that need to be specified by the user in order to execute a function. Some arguments may have pre-specified, or default, values and so don't need to be changed unless different options are required.

function - collection of `R` commands to perform a particular task. Can be in-built (e.g. `mean`) or written by the user.

package - collection of functions. A package must be installed (i.e. listed on the Packages tab) before accessing.

Rstudio - graphical user interface to `R`.

command line - `>` in the console panel. If R thinks that a command is not complete when return has been pressed (e.g. a bracket is missing), the `>` will change to `+` until the command is finished.

script files - created by 'File > New File > R Script'. Text file used to write commands (and comments) which can then be executed (Run) from the script window using 'Run', or copied and pasted to the command line. These files have the subscript xxxx.R where xxxx is the name provided by the user.

working directory - directory specified by user and subsequently used, for example, to read files from or to save files to.

workspace - file containing data objects created during R session. Has filename suffix '.RData' by default, but this can be changed by user.

Useful commands/functions

`citation()` - provides the citation for a package (e.g. `citation(package="Distance")`)

`getwd()` - returns the working directory

`help` function - provides details about the specified function and what arguments it requires (e.g. `help(mean)`)

`library` - loads specified package (e.g. `library(Distance)`). Package must be installed.

`load(".RData")` - loads workspace

`ls()` - lists objects in the R workspace

`q()` - quit R session. You will be asked if you want to save your workspace. If 'yes' this will save the objects created in an R workspace. Script files can also be saved.

`require` - similar to `library` (e.g. `require(tidyverse)`)

`rm()` - deletes objects from workspace(e.g. `rm(xval)`). The command `rm(list=ls())` deletes everything - use with caution!

`save.image()` - saves everything in current workspace to `.RData` file, `save.image(file="xxx.RData")` will save to `xxx.RData`

`setwd` function - sets the working directory (e.g. `setwd("C:/workshop")`)

Useful menu items in RStudio

Some of the basic commands in R used to navigate your directories can be performed using the menu items in RStudio. A few are listed below.

File > Open file - opens file (this could be an '.rmd' file or a text file containing R commands)

File > Save - saves the active file.

Session > Set working directory > Choose directory - sets the working directory

Session > Load workspace - loads the chosen `.RData` file

Distance sampling packages

The main packages that will be used for the Introductory and advanced distance sampling workshops are:

'Distance' - package for fitting detection functions and estimating abundance using distance sampling methodology (Miller 2017)

'dsm' - implements spatial density models for distance sampling data (Miller et al.)

'fancy' - ?

mrds - mark-recapture distance sample (Laake et al 2018)

Other packages that may be of interest are:

'dsdata' - data from Distance for Windows projects used for workshops have been exported and packaged together,

To obtain more information about these packages use, for example:

```
help(package="Distance")
```

Special characters

`+`,`-`,`*`,`/` arithmetic operators

`:` as in `a:b` produces a vector from `a` to `b`, inclusive (e.g. `5:10` produces a vector 5 6 7 8 9 10)

`<-` assignment symbol (e.g. `x <- 5:10`)

`[]` used for selection (e.g. `x[1]` selects first element of a vector `x`) {} - used for grouping lines of code together, for example in a function

`[rows, columns]` - used for selection of a dataframe where `rows` is the criteria for selecting rows and `columns` is the criteria for selecting the columns

`&` intersection, 'and'

`|` or

`!` not

`<` less than

`<=` less than or equal to

`>` greater than

`>=` greater than or equal to

`==` logical equals, used in selection to compare two values

`!=` not equal to

`=` used to specify values for function arguments (e.g. `ds(data=dsdat, key="hn")`)

`~` notation for specifing a model as in `y ~ x`

`^` to the power of (e.g. `x^2` is x^2)

`#` comment; any code/text after `#` on the same line is ignored

`$` used to define elements (columns/object) within an R object (e.g. data$X)

`NA` represents missing values

`%` used for matrix multiplication (e.g. `x %*% y`)

`%>%` pipe used in `tidyverse` to mean 'take the object before the pipe and feed into the next step

`pi` is π i.e. 3.141593

Useful hints

uparrow key - recalls previous command in Console panel

R is case specific and so `A` and `a` are treated as separate objects

If a package is not listed on the 'Packages' tab, it can be installed from the CRAN Repository

'Esc' - the escape key will halt execution of a command as well as using the red 'Stop' symbol on the Console tab (only displayed when a command is being executed).

References

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Miller DL (2017) Distance: Distance Sampling Detection Function and Abundance Estimation. R package version 0.9.7. https://CRAN.R-project.org/package=Distance

Miller DL, Rexstad E, Burt L, Bravington MV and Hedley S (NA) dsm: Density Surface Modelling of Distance Sampling Data. R package version 2.2.16. http://github.com/DistanceDevelopment/dsm

Laake J, Borchers D, Thomas L, Miller D and Bishop J (2018) mrds: Mark-Recapture Distance Sampling. R package version 2.2.0. https://CRAN.R-project.org/package=mrds