

A primer on R for the Intermediate Distance Sampling workshop



CREEM

Centre for Research into Ecological
and Environmental Modelling



University
of
St Andrews

Learning goals

- Obtain a working knowledge of the **R language** and environment, to be able to implement analysis for the Intermediate Distance Sampling workshop
- Learn how to access R via the **RStudio** software

The learning session will be almost entirely hands-on.

Thanks to Tiago Marques and Danielle Harris for tutorial material.

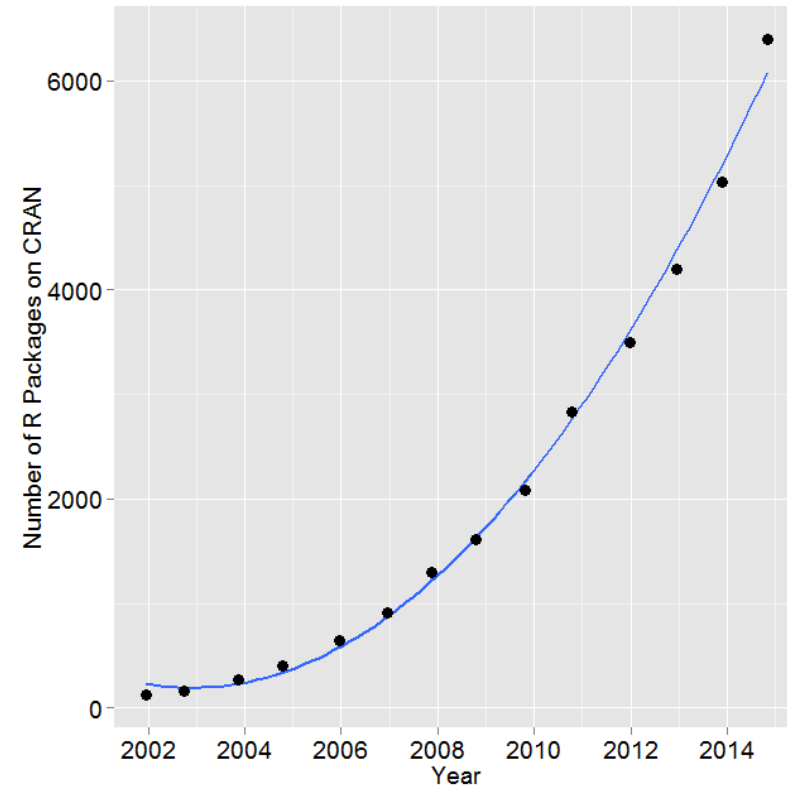
About R

- R is both a **programming language** and an **environment for statistical computation**
- R is free **open source** software
- Created in 1995 by Ross Ihaka and Robert Gentleman, (University of Auckland, Auckland)
- Based on previously established language, S
- Now managed by the “R Core Development Team”
- Software and resources available at
<http://www.r-project.org>

Latest version was **R version 3.4.1**

The R base + packages

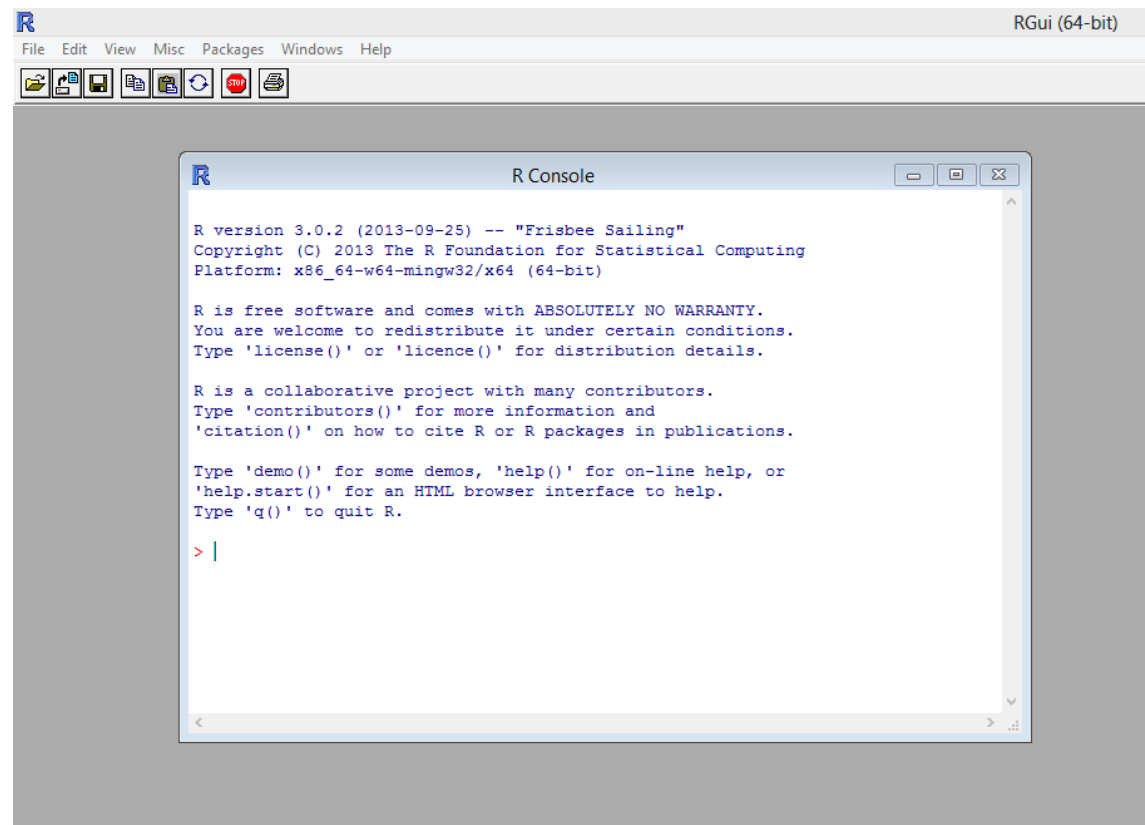
- **R base installation comes with some packages** that provide access to a large number of commonly used features (e.g. fancy plotting, survival analysis, nonlinear modelling, etc)
- **Many more packages are available online**, contributed by users



Source: <http://r4stats.com/articles/popularity/>

The R environment

- R comes with a simple interface, allowing you to
 - type/paste commands
 - view plots



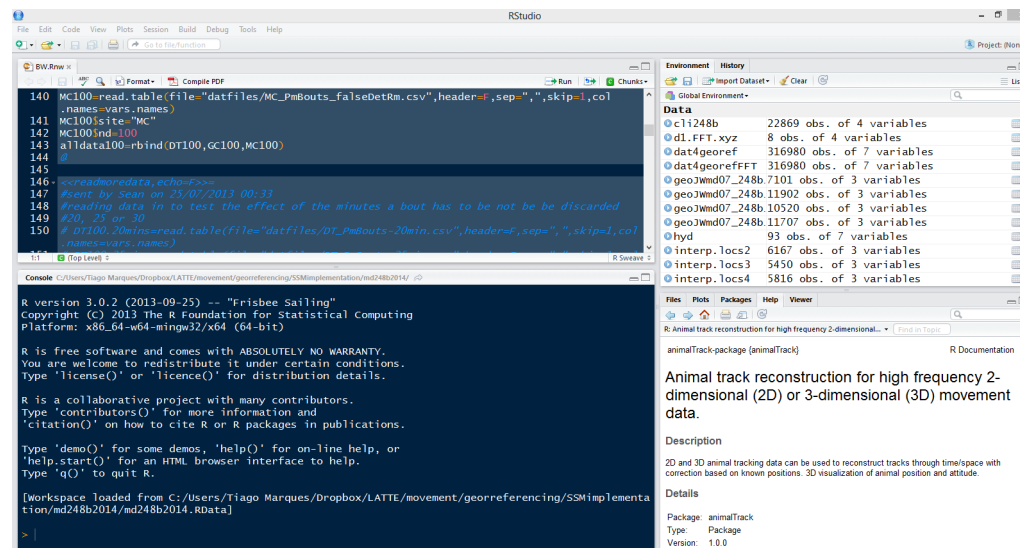
The screenshot shows the RGui (64-bit) window. The title bar reads "RGui (64-bit)". The menu bar includes "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations and execution. The main area contains an "R Console" window with the following text:

```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"  
Copyright (C) 2013 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> |
```

- ... but RStudio is much easier to use

RStudio

- An feature-rich “integrated development environment” for data analysis
- Simplifies the user interaction with R
- In a single application you have access to a number of windows containing your code files, an execution console, your plots, help, a file explorer, etc.
- (as R) Free and open source



4 windows; 4 ways to interact with R

code file (“script”)

The screenshot displays the RStudio interface with four main windows:

- Code Editor (Script):** Contains the following R code:

```
1 temp=rnorm(50)
2 plot(temp)
3
```
- Environment:** Shows the current environment with the following data:

Data	Values
dat4fhsmm	num [1:6292, 1:3] 26.1 39.4 52.6 65.6 79.4 ...
temp	num [1:50] -0.518 0.0561 -0.7382 -0.2479 0.620...
- Console:** Shows the R startup message and the execution of the code:

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from C:/Users/Tiago Marques/Dropbox/LATTE/movement/sim3Dtracks/DTAGissues/dat4fhsmm.Rdata]

> plot(temp)
Error in plot(temp) : object 'temp' not found
> temp=rnorm(50)
> plot(temp)
>
```
- Plots:** Shows a scatter plot of the variable 'temp' against the 'Index' (1 to 50). The y-axis ranges from -1.5 to 1.5.

active console

“environment” –
i.e., list of variables

plots

The code and plots tabs

Sends and executes current line or selected text to console

Move back and forth in a figure log

Export figures as pdf or jpeg

The screenshot shows the RStudio interface with the following components:

- Code Editor:** Contains R code:


```
1 temp=rnorm(50)
2 plot(temp)
3
4
5 temp2=rnorm(50)
6 plot(temp2)
```

 A red arrow points from the 'Run' button to the code editor.
- Environment:** Shows the current workspace with variables:

Object	Class	Attributes	Values
dat4fhsmm	num	[1:6292, 1:3]	26.1 39.4 52.6 65.6 79.4 ...
temp	num	[1:50]	-0.518 0.0561 -0.7382 -0.2479 0.620...
temp2	num	[1:50]	-1.562 -0.356 -1.489 1.309 -0.792 ...
- Console:** Shows the execution history:


```
S.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in pub
lications.

Type 'demo()' for some demos, 'help()' for on-line
help, or
'help.start()' for an HTML browser interface to he
lp.
Type 'q()' to quit R.

[workspace loaded from C:/Users/Tiago Marques/Drop
box/LATTE/movement/sim3Dtracks/DTAGissues/dat4fhs
m.Rdata]

> plot(temp)
Error in plot(temp) : object 'temp' not found
> temp=rnorm(50)
> plot(temp)
> temp2=rnorm(50)
> plot(temp2)
>
```
- Plots:** Displays a scatter plot of 'temp2' vs 'Index'. The x-axis is labeled 'Index' (0 to 50) and the y-axis is labeled 'temp2' (-2 to 1). A red arrow points from the 'Export' button to the plot.

The environment tab

The screenshot displays the RStudio interface with the following components:

- Environment Tab:** Shows the Global Environment with objects:
 - `dat4fhsmm`: num [1:6292, 1:3] 26.1 39.4 52.6 65.6 79.4 ...
 - `temp`: num [1:50] -0.518 0.0561 -0.7382 -0.2479 0.620...
 - `temp2`: num [1:50] -1.562 -0.356 -1.489 1.309 -0.792 ...
- Data Table:** A table with 12 rows and 3 columns (V1, V2, V3). The first row is highlighted in pink.

	V1	V2	V3
1	2.610553e+01	448.00	1
2	3.944256e+01	449.00	1
3	5.259969e+01	450.00	1
4	6.556488e+01	451.00	1
5	7.936468e+01	452.00	1
6	9.393605e+01	453.00	1
7	1.093265e+02	454.00	1
8	1.255684e+02	455.00	1
9	1.424350e+02	456.00	1
10	1.598421e+02	457.00	1
11	1.776251e+02	458.00	1
12	1.955942e+02	459.00	1
- Console:** Shows R code and output:


```

      Type 'citation()' for more information and
      'citation()' on how to cite R or R packages in pub
      lications.

      Type 'demo()' for some demos, 'help()' for on-line
      help, or
      'help.start()' for an HTML browser interface to he
      lp.
      Type 'q()' to quit R.

      [Workspace loaded from C:/Users/Tiago Marques/Drop
      box/LATTE/movement/sim3Dtracks/DTAGissues/dat4fhsm
      m.Rdata]

      > plot(temp)
      Error in plot(temp) : object 'temp' not found
      > temp=rnorm(50)
      > plot(temp)
      > temp2=rnorm(50)
      > plot(temp2)
      > View(dat4fhsmm)
      > View(dat4fhsmm)
      >
      
```
- Plots Panel:** Displays a scatter plot of `temp2` vs `Index`. The x-axis is labeled "Index" (0 to 50) and the y-axis is labeled "temp2" (-2 to 1).

Shortcut for data
import

Objects
available in
workspace (+
details)

Pressing
most
objects
names
gives you
a preview
of what
they are

The packages tab

The screenshot displays the RStudio interface with the following components:

- Environment Pane:** Shows the current workspace with variables:

Variable	Class	Attributes	Values
dat4fhsmm	num	[1:6292, 1:3]	26.1 39.4 52.6 65.6 79.4 ...
temp	num	[1:50]	-0.518 0.0561 -0.7382 -0.2479 0.6203 ...
temp2	num	[1:50]	-1.562 -0.356 -1.489 1.309 -0.792 ...
- Packages Tab:** Lists installed and available packages.

Package Name	Description	Version	Status
<input type="checkbox"/> animalTrack	Animal track reconstruction for high frequency 2-dimensional (2D) or 3-dimensional (3D) movement data.	1.0.0	Not Installed
<input type="checkbox"/> boot	Bootstrap Functions (originally by Angelo Canty for S)	1.3-9	Not Installed
<input type="checkbox"/> CircStats	Circular Statistics, from "Topics in circular Statistics" (2001)	0.2-4	Not Installed
<input type="checkbox"/> circular	Circular Statistics	0.4-7	Not Installed
<input type="checkbox"/> class	Functions for Classification	7.3-9	Not Installed
<input type="checkbox"/> cluster	Cluster Analysis Extended Rousseeuw et al.	1.14.4	Not Installed
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-8	Not Installed
<input type="checkbox"/> compiler	The R Compiler Package	3.0.2	Not Installed
<input checked="" type="checkbox"/> datasets	The R Datasets Package	3.0.2	Installed
<input type="checkbox"/> expm	Matrix exponential	0.99-1.1	Not Installed
<input type="checkbox"/> fields	Tools for spatial data	6.9.1	Not Installed
<input type="checkbox"/> foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase, ...	0.8-55	Not Installed
<input checked="" type="checkbox"/> graphics	The R Graphics Package	3.0.2	Installed
<input checked="" type="checkbox"/> grDevices	The R Graphics Devices and Support for Colours and Fonts	3.0.2	Installed
<input type="checkbox"/> grid	The Grid Graphics Package	3.0.2	Not Installed
<input type="checkbox"/> KernSmooth	Functions for kernel smoothing for Wand B. Jones (1995)	2.23-10	Not Installed
<input type="checkbox"/> lattice	Lattice Graphics	0.20-23	Not Installed
- Console:** Shows R help text and error messages:


```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,
or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from C:/Users/Tiago Marques/Dropbox/LA
TTE/movement/sim3Dtracks/DTAGissues/dat4fhsmm.Rdata]

> plot(temp)
Error in plot(temp) : object 'temp' not found
> temp=rnorm(50)
> plot(temp)
> temp2=rnorm(50)
> plot(temp2)
> View(dat4fhsmm)
> View(dat4fhsmm)
>
```

Annotations on the right side of the image:

- install a package (points to the 'Install Packages' button)
- an installed package (points to the 'dat4fhsmm' variable in the Environment pane)
- an installed and loaded package (points to the 'datasets' package in the Packages tab)
- click the name for package description (points to the 'KernSmooth' package name in the Packages tab)

The history and files tabs

The screenshot displays the RStudio interface with four main panels:

- Environment:** Shows a data frame named 'dat4fhsmm' with 6292 observations and 3 variables (V1, V2, V3). The data is displayed in a table with columns V1, V2, and V3, and rows 1 through 12.
- History:** Shows a list of R commands executed in the session, including:


```
Sweave("DTAGissues.Rnw")
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
plot(-dat4fhsmm[,1],ylim=c(-1600,0),pch=".",col=smd248bnmd,xlab="Time",y
plot(-obs[1:6300,1],ylim=c(-1600,0),pch=".",col=sJABES[1:6300],xlab="Ti
rrow(dat4fhsmm)
Sweave("DTAGissues.Rnw")
Sweave("DTAGissues.Rnw")
plot(temp)
temp=rnorm(50)
plot(temp)
temp2=rnorm(50)
plot(temp2)
View(dat4fhsmm)
View(dat4fhsmm)
```
- Console:** Shows the R prompt and the output of the commands. The output includes:


```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publicati
ons.

Type 'demo()' for some demos, 'help()' for on-line help,
or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from C:/Users/Tiago Marques/Dropbox/LA
TTE/movement/sim3Dtracks/DTAGissues/dat4fhsmm.Rdata]
> plot(temp)
Error in plot(temp) : object 'temp' not found
> temp=rnorm(50)
> plot(temp)
> temp2=rnorm(50)
> plot(temp2)
> View(dat4fhsmm)
> View(dat4fhsmm)
>
```
- Files:** Shows a directory listing of files in the working directory. The files include:

Name	Size	Modified
..		
.RData	18.5 MB	Feb 28, 2014, 2:55 PM
.Rhistory	25.2 KB	Mar 3, 2014, 1:32 PM
dat4fhsmm.Rdata	62 KB	Feb 6, 2014, 5:04 PM
DTAGissues-concordance.tex	204 B	Mar 3, 2014, 1:14 PM
DTAGissues.aux	3.9 KB	Mar 6, 2014, 12:57 PM
DTAGissues.bbl	1 KB	Mar 3, 2014, 12:27 PM
DTAGissues.blg	145 B	Mar 3, 2014, 12:27 PM
DTAGissues.log	36.1 KB	Mar 6, 2014, 12:57 PM
DTAGissues.out	286 B	Mar 6, 2014, 12:57 PM
DTAGissues.pdf	1.1 MB	Mar 6, 2014, 12:57 PM
DTAGissues.Rnw	37.8 KB	Mar 3, 2014, 1:17 PM
DTAGissues.synctex.gz	32.9 KB	Mar 6, 2014, 12:57 PM
DTAGissues.tex	10.8 KB	Mar 3, 2014, 1:18 PM
DTAGissues.toc	367 B	Mar 6, 2014, 12:57 PM
figs		

The command history (with shortcuts to send commands to source or console)

To see the files in your working directory (select a file to load or open it)

The help tab

The screenshot shows the RStudio interface with the following components:

- Environment/History:** Shows the execution of `temp=rnorm(50)`, `plot(temp)`, `temp2=rnorm(50)`, `plot(temp2)`, `View(dat4fhsmm)`, `View(dat4fhsmm)`, and `?mean`.
- Files/Plots/Packages/Help/Viewer:** The Help tab is active, showing the documentation for the `mean` function.
- Console:** Shows the execution of `plot(temp)` (error: object 'temp' not found), `temp=rnorm(50)`, `plot(temp)`, `temp2=rnorm(50)`, `plot(temp2)`, `View(dat4fhsmm)`, `View(dat4fhsmm)`, and `?mean`.
- Help Pane:** Displays the documentation for "Arithmetic Mean", including a description, usage, arguments, and value.

Annotations in the image highlight the following features:

- If you call for help...:** Points to the `?mean` command in the History pane.
- This is where it will show up:** Points to the Help tab in the top navigation bar.
- With hyperlinks for easy navigation:** Points to the `date`, `date-time`, and `time interval` links in the Arguments section of the help pane.

Now... open the hands on tutorial

A hands on introduction to R tutorial

prepared by Tiago A. Marques, Danielle Harris & Len Thomas

Introduction

This tutorial was created as a gentle introduction to the R environment. It does not assume any basic knowledge about R, but some basic programming notions would be desirable.

There is an extensive community revolving around R, and abundant courses, tutorials, books, blogs, list servers, etc, freely available online. We provide here a small list of some of these:

- [R webpage](#) - the main R webpage, including links to downloading R, manuals, tutorials, dedicated search engines, etc.
- [R video tutorials](#) - video how to's in R
- [Online tutorial](#) - a course with interactive exercises
- [Online course](#) - DataCamp commercial site
- [Reference card](#) - A very handy list of useful R functions
- [Short reference card](#) - A longer reference card with most commonly used R functions

To facilitate the interaction with R we leverage on RStudio, a piece of software which allows users to have at a click's distance many useful features in R. In the following sections of the tutorial you will be guided through a first session of R via RStudio.

The tutorial is intended to follow a brief presentation about R and RStudio, their interaction and capabilities. It assumes that R and RStudio have been previously installed in the computer you are using. The latest version of both software packages is recommended. Both are free and open source.

Introduction to RStudio

Working with Rmarkdown files

- Having downloaded and opened *01-Rtutorial.Rmd*
 - You will see a mixture of readable text along with R code
 - The bits of R code are called “chunks” and those chunks usually have labels
 - We are likely to refer to these chunk names when describing computation steps
 - Lines 62-66 is a chunk named ‘calc’
 - Green box indicates icons that can be clicked to send code within chunk to console for processing
 - Orange ellipse indicates icon that can be used to render entire document

The screenshot shows the RStudio interface with the file `01-Rtutorial.Rmd` open. The toolbar includes a 'Knit' button (circled in orange) and a 'Run' button. A code chunk is visible with the following content:

```

59 We can print an object to the screen by simply typing its name
60 and press enter (despite the fact that currently you can
61 actually see the values on these objects `Environment` tab -
62 but that is because they are simple objects and the workspace
63 is almost empty. )
64
65 59
66 R is a very powerful calculator! Try some simple maths, say
67 for example (you need to press enter after each line so that
68 the line is evaluated)
69
70 62 ~~~{r calc}
71 63 4+3
72 64 log(8)
73 65 sin(pi)
74 66 ~~~
75
76 68 Tip: There is actually a simpler way to do sourcing from the
77 console. For example, in RStudio, Ctrl-Enter is a keyboard shortcut for
  
```

The 'Run' and 'Source' icons at the end of the code chunk are highlighted with a green box.