

<https://intro2r.com/> Chapter 1

CSCI 297b, Spring 2023

April 19, 2023

Advantages of R

- R is open source and freely available.
- R is available for Windows, Mac and Linux operating systems.
- R has an extensive and coherent set of tools for statistical analysis.
- R has an extensive and highly flexible graphical facility capable of producing publication quality figures.
- R has an expanding set of freely available 'packages' to extend R's capabilities.
- R has an extensive support network with numerous online and freely available documents.

RStudio panels

The screenshot displays the RStudio Desktop interface with the following panels and content:

- Source Editor:** Contains a script named `squid1.txt` with the following R code:

```
1  
2 squid <- read.delim("data/squid1.txt")  
3 str(squid)  
4 dotchart(squid$nid.length)  
5
```
- Console:** Shows the output of the executed code:

```
R 4.1.2 ~ ~ ~  
Hit <Return> to see next plot:  
  
pairs() ## put (absolute) correlations on the upper panels,  
## with size proportional to the correlations.  
pairs() panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)  
pairs+ {  
pairs+   usr <- par("usr"); on.exit(par(usr))  
pairs+   par(usr = c(0, 1, 0, 1))  
pairs+   r <- abs(cor(x, y))  
pairs+   txt <- format(c(r, 0.123456789), digits = digits)[1]  
pairs+   txt <- paste0(prefix, txt)  
pairs+   if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)  
pairs+   text(0.5, 0.5, txt, cex = cex.cor * r)  
pairs+ }  
  
pairs() pairs(USJudgeRatings, lower.panel = panel.smoother, upper.panel = panel.cor,  
pairs+   gap=0, rowlratop=FALSE)  
Hit <Return> to see next plot:
```
- Environment:** Displays the current environment with the following data objects:

Object	Class	Attributes
given.depth	num	[1:4, 1:2] 39.5 96.5 238.5 534.5 107.5 ...
squid	data.frame	519 obs. of 14 variables
whale	data.frame	100 obs. of 8 variables
whaledata	data.frame	100 obs. of 8 variables
- Files:** Shows the file explorer with the following files:

Name	Description	Version
palmerpenguins	Palmer Archipelago (Antarctic) Penguin Data	0.1.1
abind	Combine Multidimensional Arrays	1.4-5
ade4	Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences	1.7-19
AER	Applied Econometrics with R	1.2-10
afex	Analysis of Factorial Experiments	1.1-1
AICcmodavg	Model Selection and Multimodel Inference Based on AICc	2.3-1
airports	Data on Airports	0.1-0
annotate	Annotation for microarrays	1.72-0
AnnotationDbi	Manipulation of SQLite-based annotations in Bioconductor	1.56-2
antword	Extract Text from Microsoft Word Documents	1.3.1
asword	Analysis of Overdispersed Data	1.3-2

RStudio panels left side

- Source window
 - Enter and edit commands for the R interpreter.
 - Not sourced into the R interpreter until you hit Source button.
 - Can be sourced with control+enter (windows) or command+enter (mac)
- Console/Terminal/Launcher window
 - Type commands directly into the R interpreter.
 - Use as a calculator

RStudio panels right side

- Environment/History/Connections/Tutorial window
 - Environment tells you what names you have defined.
 - History gives list of commands you've entered in console.
- Files/Plots/Packages/Help/Viewer window
 - File browser
 - Plot viewer
 - Help viewer
 - Web viewer

RStudio console and source

- Mostly we use editor and console
- Can enter commands in console for quick answers
- Almost always better to enter commands into an R script
- Save your R scripts with a .R extension
- R scripts are plain text files that can be edited with any text editor

R packages

- Large collection already installed on server.
- CRAN packages:

```
1 install.packages('remotes', dependencies=  
  TRUE)
```

- Using packages:

```
1 library(remotes)
```

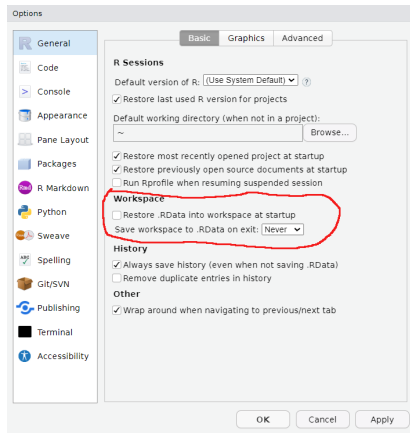
- GitHub packages

```
1 library(remotes)  
2 install_github('tidyverse/dplyr')
```

- You can install additional packages on your own computer.
- You can install additional packages on the server, if they are just for your own use.

Global options

- Tools → Global Options ...
- Turn off .RData save and restore
- This prevents new R sessions from being influenced by previous R sessions.



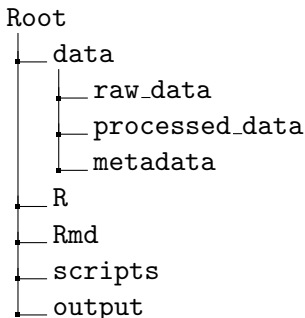
R projects

- Remembers settings, history, *etc.*
- Organizes all files, scripts, data, *etc.*
- File → New Project
- Usually use a new directory
- Create Project

Working directory

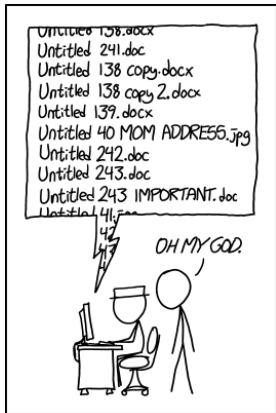
- Current working directory seen at top of console window
- Default folder R looks in when loading scripts, data, etc.
- Opening a project file automatically sets working directory
- Can use `getwd()` to identify working directory
- Can use `setwd()` to change working directory

Directory structure



- Root holds the .Rproj file
- raw_data holds data from the field that should not be edited
- processed_data holds data that has been cleaned up
- metadata holds text files describing how the data was collected, etc.
- R holds R functions of general utility
- Rmd holds Rmarkdown files
- scripts holds current work
- output holds plots, data summaries, etc.

File names



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

- Short
- Informative
- No spaces
- No special characters
!@#\$\$%^&*()
- Pad with zeros
file001 file002 file003
- Use ISO 8601 date format
YYYY-MM-DD or YYYYMMDD

R script meta information

```
1 # Title: Time series analysis of snouters
2
3 # Purpose : This script performs a time series analyses on
4 #           snouter count data.
5 #           Data consists of counts of snouter species
6 #           collected from 18 islands in the Hy-yi-yi
7 #           archipelago between 1950 and 1957.
8 #           For details of snouter biology see:
9 #           https://en.wikipedia.org/wiki/Rhinogradentia
10
11 # Project number: #007
12
13 # DataFile: 'data/snouter_pop.txt'
14
15 # Author: A. Nother
16 # Contact details: a.nother@uir.ac.uk
17
18 # Date script created: Mon Dec 2 16:06:44 2019 _____
19 # Date script last modified: Thu Dec 12 16:07:12 2019 _____
20
21 # package dependencies
22 library(PopSnouter)
23 library(ggplot2)
24
25 print('put your lovely R code here')
26
27 # good practice to include session information
28
29 xfun::session_info()
```

Style guide

`https://google.github.io/styleguide/Rguide.html`

Citing R

```
1 > citation()
2
3 To cite R in publications use:
4
5 R Core Team (2021). R: A language and environment for
6 statistical computing. R Foundation for Statistical
7 Computing, Vienna, Austria. URL https://www.R-project.org/.
8
9 A BibTeX entry for LaTeX users is
10
11 @Manual{,
12   title = {R: A Language and Environment for Statistical Computing},
13   author = {{R Core Team}},
14   organization = {R Foundation for Statistical Computing},
15   address = {Vienna, Austria},
16   year = {2021},
17   url = {https://www.R-project.org/},
18 }
19
20 We have invested a lot of time and effort in creating R,
21 please cite it when using it for data analysis. See also
22 'citation("pkgname")' for citing R packages.
```

Exercise 1