

# R Programming For Natural Resource Professionals

Lecture 2: Data Structures, Data Types,  
Finding Help, Coding Etiquette

# Paper discussions

- Group assignments

# Paper discussions

- Open this link now: <https://bit.ly/3GVPwuV>
- Talk with your group to identify one or two thoughts, questions, and epiphanies that resonate then record them in the Google Doc.
- Read through the list as it updates.

# Homework reminders

- Be sure to cite sources, including your colleagues.
- Remember to comment your code.
  - In many cases, line-by-line commenting is appropriate
- Follow instructions exactly.
- Homework will be returned via email. Grades will be on Canvas.
  - Read through the R markdown that I return. It includes notes from me (JH: ...)

# Topics for today

- Data types
- Data structures
- Finding help
- Coding etiquette
- Workflows

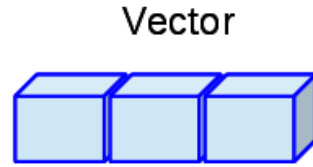
# Data types

Data type	Example	Verify
Logical	TRUE, FALSE	<pre>a &lt;- TRUE class(a)</pre>
Numeric	42.1, 4, 2215	<pre>b &lt;- 23.4 class(b)</pre>
Integer	2L, 24L, 0L	<pre>c &lt;- 2L class(c)</pre>
Character	"a", "good", "perch"	<pre>d &lt;- "perch" class(d)</pre>
Complex	1+4i	<pre>e &lt;- "1+4i" class(e)</pre>

# Topics for today

- Data types
- Data structures
- Finding help
- Coding etiquette
- Workflows

# Data structures



**Vector**: A sequence of items of the same type.

Most basic data structure in R

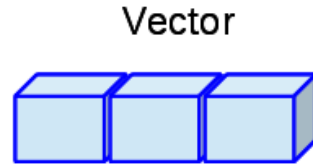
Items in a vector can be accessed using `[]`

Length of vector displayed using `length()`

```
> apple <- c("red", "green", "yellow")  
> apple  
> class(apple)  
> length(apple)  
> apple[2]
```



# Data structures

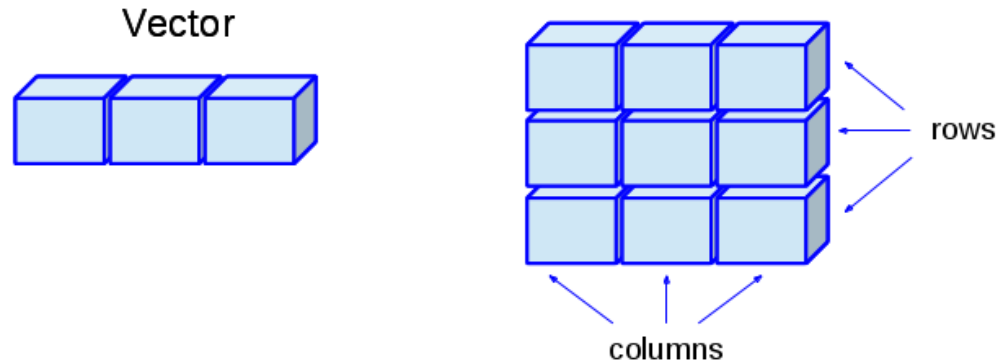


**Factor**: A common type of vector used in plotting and modeling.

Forces values of the vector into categories

```
> apple <- c("red", "green", "yellow")  
> factor_apple <- as.factor(apple)  
> factor_apple  
> str(factor_apple)
```

# Data structures

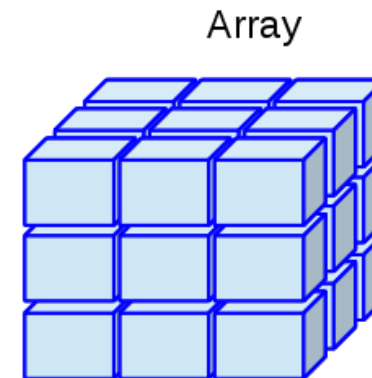
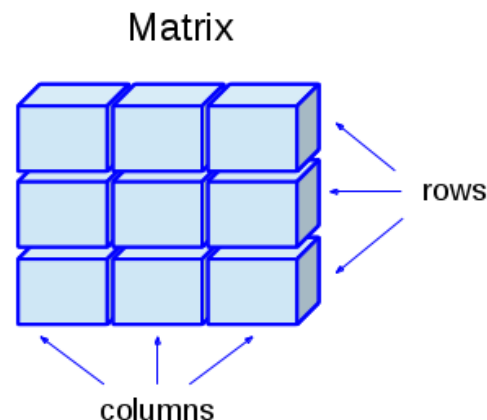
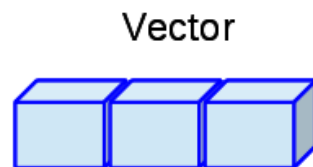


**Matrix**: A two-dimensional array

```
> matrix1 <- matrix(c("a", "a", "b", "c", "c", "a"),  
  nrow = 2, ncol = 3, byrow = TRUE)
```

```
> matrix1
```

# Data structures



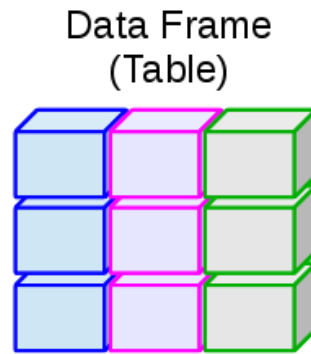
**Array**: A multidimensional matrix

Any number of dimensions.

The array function's dim attribute is used to specify dimensionality.

```
> array1 <- array(c("green", "yellow"), dim = c(3,3,2))  
> array1
```

# Data structures



**Data frame**: A table, similar to a matrix, but each variable (column) can be a different data type

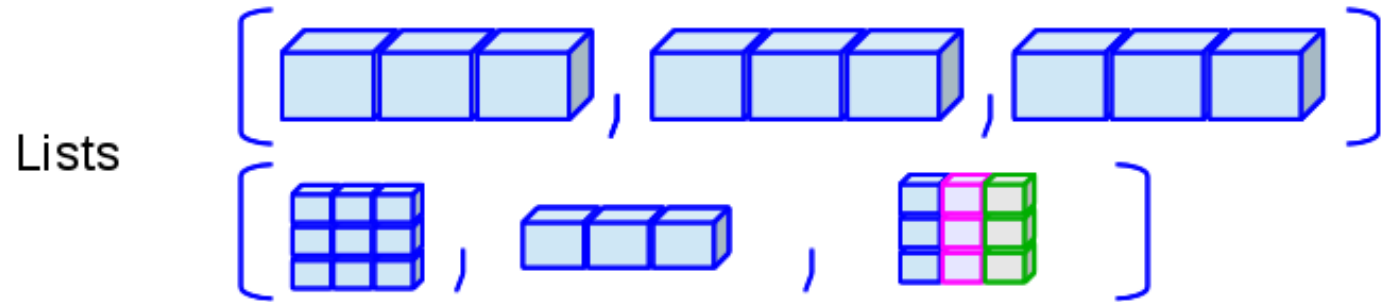
```
> df1 <- data.frame(gender = c("Male", "Female",  
"Female"), height = c(152, 171.5, 165), weight = c(81,  
93, 83), age = c(42, 38, 26))
```

```
> df1
```

```
> str(df1)
```

```
> summary(df1)
```

# Data structures



**Lists**: Can contain many different types of elements, such as vectors, arrays, data frames, or even other lists.

```
> list1 <- list(c(2,5,3), 21.3, "tree")  
> str(list1)  
> list1[3]  
> list1[[3]]  
> list1[[1]][[2]]
```

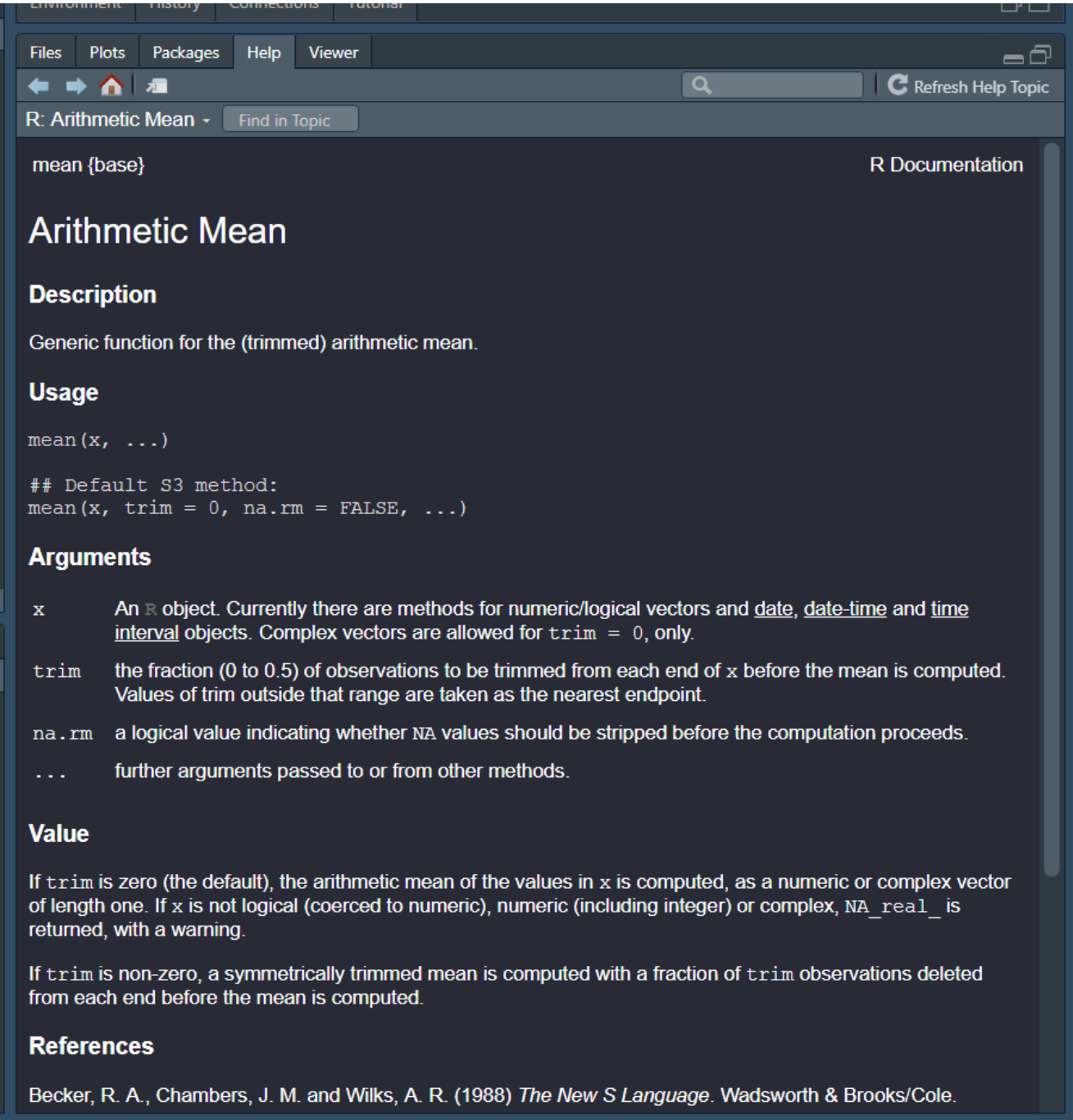
# Topics for today

- Data types
- Data structures
- Finding help
- Coding etiquette
- Workflows

# Finding help

R's built-in help pages

> ?mean



The screenshot shows the R help interface for the 'mean' function. The top navigation bar includes tabs for 'Files', 'Plots', 'Packages', 'Help', and 'Viewer'. Below this is a search bar and a 'Refresh Help Topic' button. The main content area is titled 'R: Arithmetic Mean' and includes a 'Find in Topic' button. The page content is organized into sections: 'mean {base}', 'Arithmetic Mean', 'Description', 'Usage', 'Arguments', 'Value', and 'References'. The 'Description' section states it is a generic function for the (trimmed) arithmetic mean. The 'Usage' section shows the function signature 'mean(x, ...)' and a comment about the default S3 method. The 'Arguments' section lists 'x' (an R object), 'trim' (fraction of observations to trim), 'na.rm' (logical value for NA handling), and '...' (further arguments). The 'Value' section explains the output based on the 'trim' argument. The 'References' section cites the book 'The New S Language' by Becker, Chambers, and Wilks (1988).

mean {base} R Documentation

## Arithmetic Mean

### Description

Generic function for the (trimmed) arithmetic mean.

### Usage

```
mean(x, ...)
```

## Default S3 method:  
mean(x, trim = 0, na.rm = FALSE, ...)

### Arguments

<code>x</code>	An <code>R</code> object. Currently there are methods for numeric/logical vectors and <a href="#">date</a> , <a href="#">date-time</a> and <a href="#">time interval</a> objects. Complex vectors are allowed for <code>trim = 0</code> , only.
<code>trim</code>	the fraction (0 to 0.5) of observations to be trimmed from each end of <code>x</code> before the mean is computed. Values of <code>trim</code> outside that range are taken as the nearest endpoint.
<code>na.rm</code>	a logical value indicating whether NA values should be stripped before the computation proceeds.
<code>...</code>	further arguments passed to or from other methods.

### Value

If `trim` is zero (the default), the arithmetic mean of the values in `x` is computed, as a numeric or complex vector of length one. If `x` is not logical (coerced to numeric), numeric (including integer) or complex, `NA_real_` is returned, with a warning.

If `trim` is non-zero, a symmetrically trimmed mean is computed with a fraction of `trim` observations deleted from each end before the mean is computed.

### References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

# Finding help



## Online resources

Stack Overflow

Package vignettes on CRAN

- <https://cran.r-project.org/web/packages/vegan/index.html>

R-bloggers

Package's GitHub sites

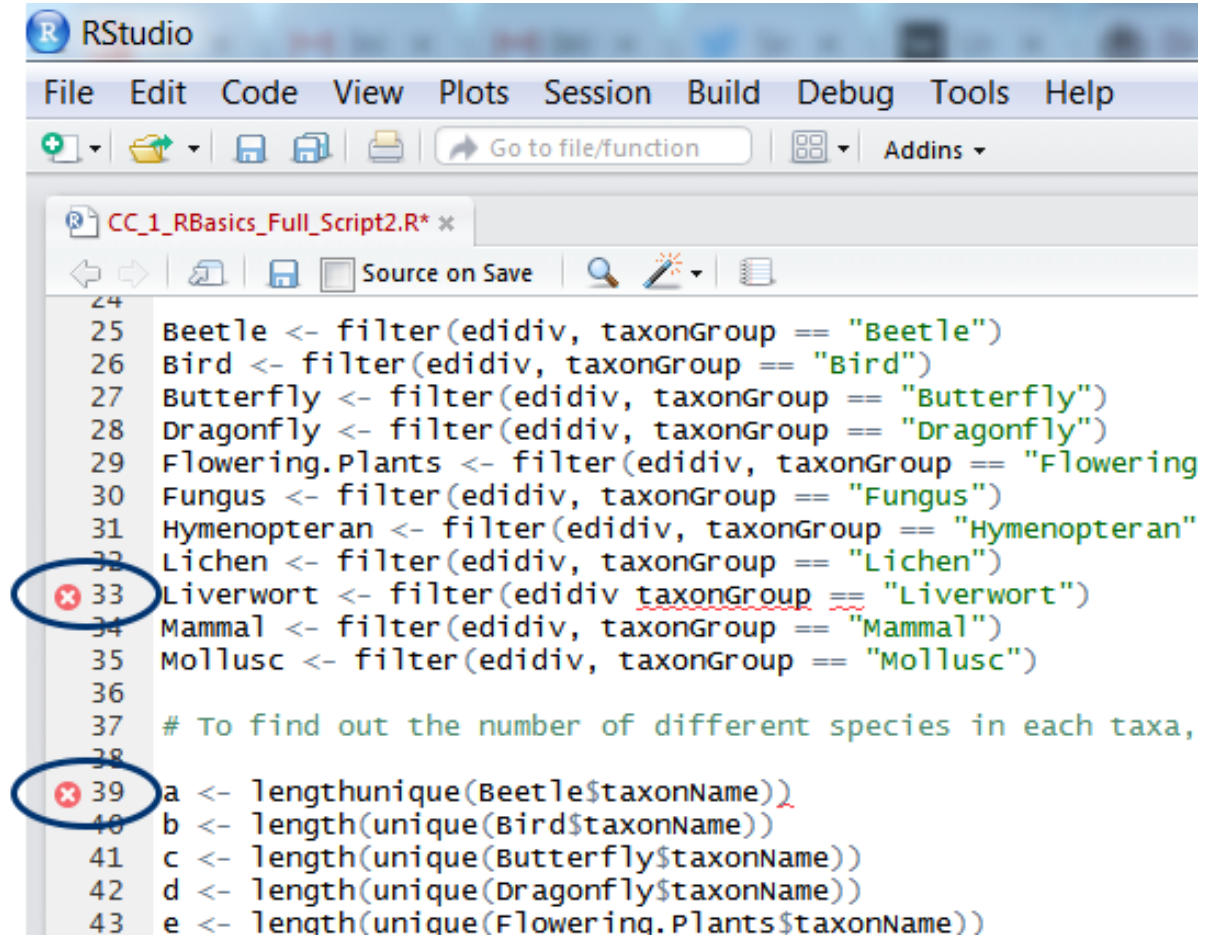


# Finding help

## Realtime Error Checking in Rstudio

Syntax errors are common and easy to make.

- Open bracket or parentheses
- Missing commas
- Extra character or other typo



The screenshot shows the RStudio interface with a script file named 'CC\_1\_RBasics\_Full\_Script2.R'. The script contains several lines of R code. Two lines are circled in red, each with a red 'x' icon indicating a syntax error. The first error is on line 33, where the closing parenthesis for the 'filter' function is missing. The second error is on line 39, where there is an extra closing parenthesis at the end of the line.

```
24  
25 Beetle <- filter(edivid, taxonGroup == "Beetle")  
26 Bird <- filter(edivid, taxonGroup == "Bird")  
27 Butterfly <- filter(edivid, taxonGroup == "Butterfly")  
28 Dragonfly <- filter(edivid, taxonGroup == "Dragonfly")  
29 Flowering.Plants <- filter(edivid, taxonGroup == "Flowering")  
30 Fungus <- filter(edivid, taxonGroup == "Fungus")  
31 Hymenopteran <- filter(edivid, taxonGroup == "Hymenopteran")  
32 Lichen <- filter(edivid, taxonGroup == "Lichen")  
33 Liverwort <- filter(edivid, taxonGroup == "Liverwort")  
34 Mammal <- filter(edivid, taxonGroup == "Mammal")  
35 Mollusc <- filter(edivid, taxonGroup == "Mollusc")  
36  
37 # To find out the number of different species in each taxa,  
38  
39 a <- length(unique(Beetle$taxonName))  
40 b <- length(unique(Bird$taxonName))  
41 c <- length(unique(Butterfly$taxonName))  
42 d <- length(unique(Dragonfly$taxonName))  
43 e <- length(unique(Flowering.Plants$taxonName))
```

# Common errors

Error: Could not find function 'functionName'

Likely solution: Package containing function not loaded. `library(packageName)`

Error: There is no package called 'packageName'

Likely solution: Package isn't installed

Error: object 'objectName' not found

Likely solution: Check your environmental panel that object is loaded. Check for typos.

Error: unexpected symbol in 'lineOfCode'

Likely solution: A forgotten or extra comma, bracket, etc.

# Topics for today

- Data types
- Data structures
- Finding help
- Coding etiquette
- Workflows

# Coding Etiquette

File names should be meaningful

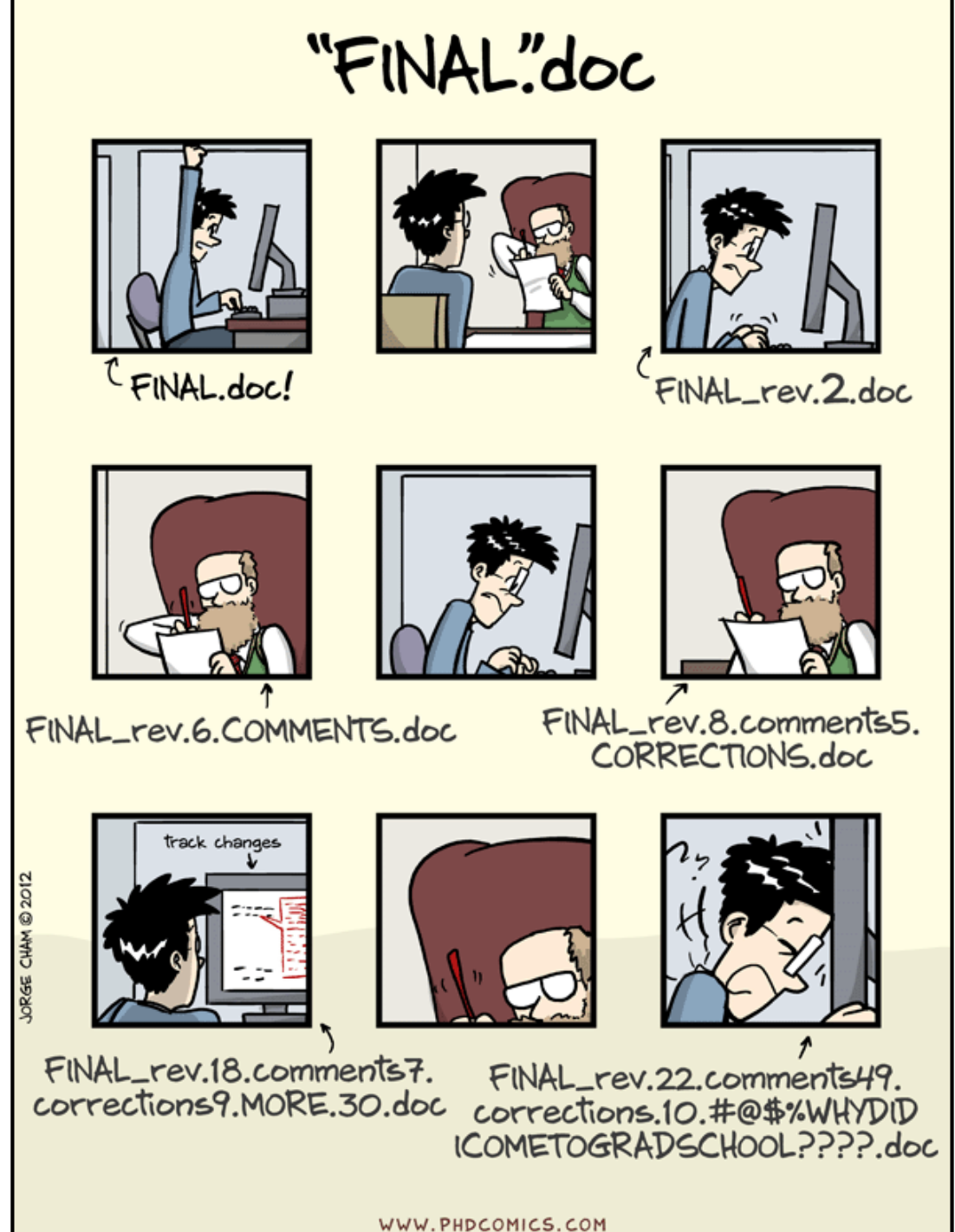
- dataWrangling.R
- repeatedMeasuresModel.R
- modelSelection.R

# Coding Etiquette

If files must be run in sequence,  
give them numerical prefixes:

- 1\_dataWrangling.R
- 2\_modelSelection.R
- 3\_plotting.R

Word from the wise- no matter how  
much you think it is appropriate, never  
name anything "final."



# Coding Etiquette

Function names should be **verbs**

Object names should be **nouns**

## Good function names

parse\_data

generate\_boxplots

## Good object names

green\_bay\_env

mccurdy\_woodlot\_dbh

# Coding Etiquette

Do not assign names to existing functions

For example:

- C
- T
- F
- data

# Coding Etiquette

Put **space** around all operators and after commas.  
Just like English. Make your code readable.

Good: `average <- mean(feet / 12 + inches, na.remove = TRUE)`

Bad: `average<-mean(feet/12+inches,na.remove=TRUE)`



# Coding Etiquette

Put **space** around all operators and after commas.  
Just like English. Make your code readable.

Exception to the rule!

```
base::mean
```

# Coding Etiquette

**Opening curly brackets** do not go on new line but are followed by a new line.  
**Closing curly brackets** always go on a new line.

```
if (y < 0){  
    message("Y is negative")  
}
```

# Coding Etiquette

**Assignment** should always be done using “<-”, never “=”

Good: `x <- 10`

Bad: `x = 10`

# Coding Etiquette

**Case:** lots of case options. Just be consistent.

- lower\_snake
- UPPER\_SNAKE
- lowerCamelCase
- UpperCamelCase
- kebab-case

# Coding Etiquette

Use commenting frequently

Titling scripts:

```
#####  
#### Length frequency plots ####  
#####
```

In R Studio: ##### generates a collapsible code block

# Coding Etiquette Exercise

```
x=330/12
```

```
c <- 1972
```

```
if (y < 0){message("Y is negative")}
```

# Topics for today

- Data types
- Data structures
- Finding help
- Coding etiquette
- **Workflows**

# Workflows

Goal: Portable code

- R scripts and data files recreate the environment

R Studio facilitates a project-oriented workflow to help with this



# Workflows

## **Projects** include:

- Input data
- R scripts
- A default home directory
  - Enables writing of relative paths (e.g., data/species\_counts.csv)

*One folder = one project*

# Workflows

**Projects** directory structures can vary, but a good starting point is:

1. data folder: all input data and metadata
2. doc folder: manuscript and other documents
3. figs folder: all output figures
4. output folder: intermediate outputs (e.g., wrangled data)
5. R folder: R scripts
6. reports folder: R Markdown files that generate reports

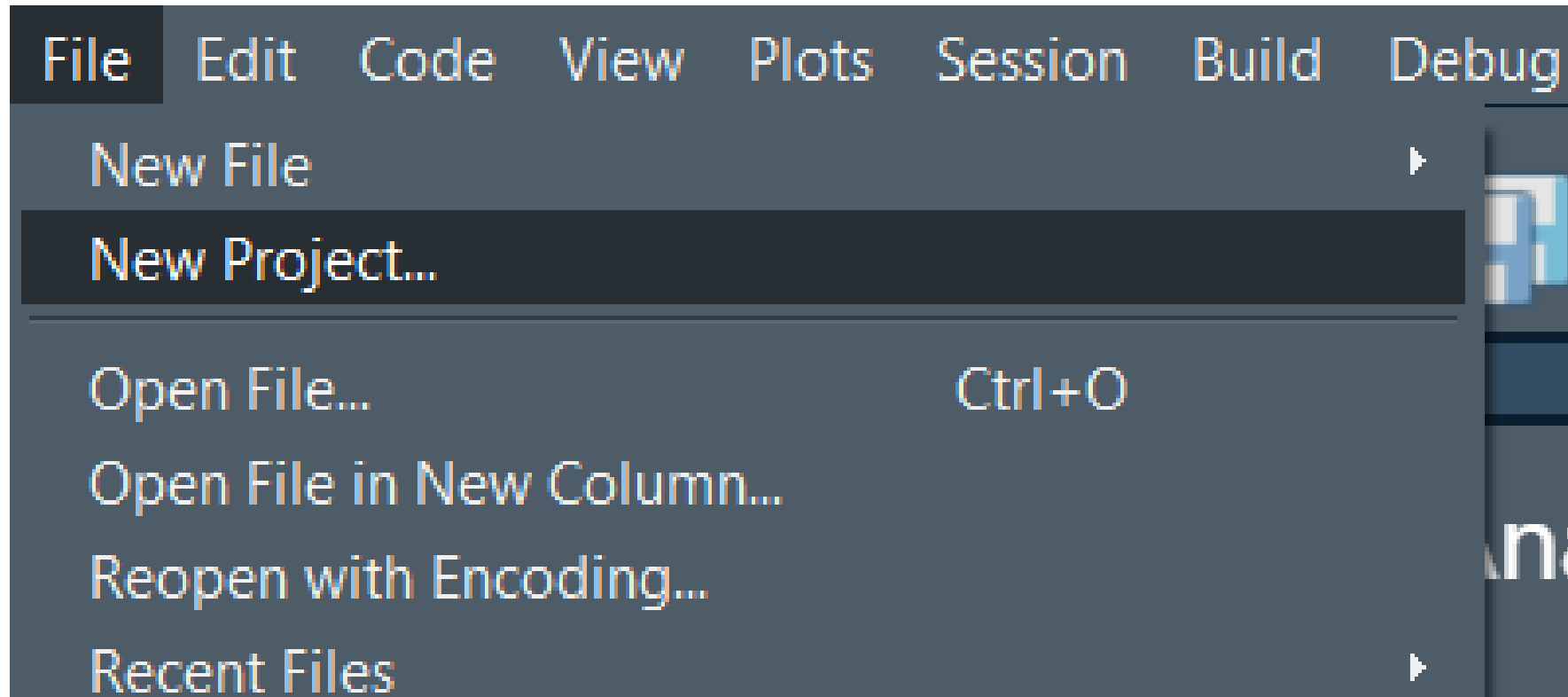
# Workflows

**Projects** should contain scripts organized for a logical workflow:

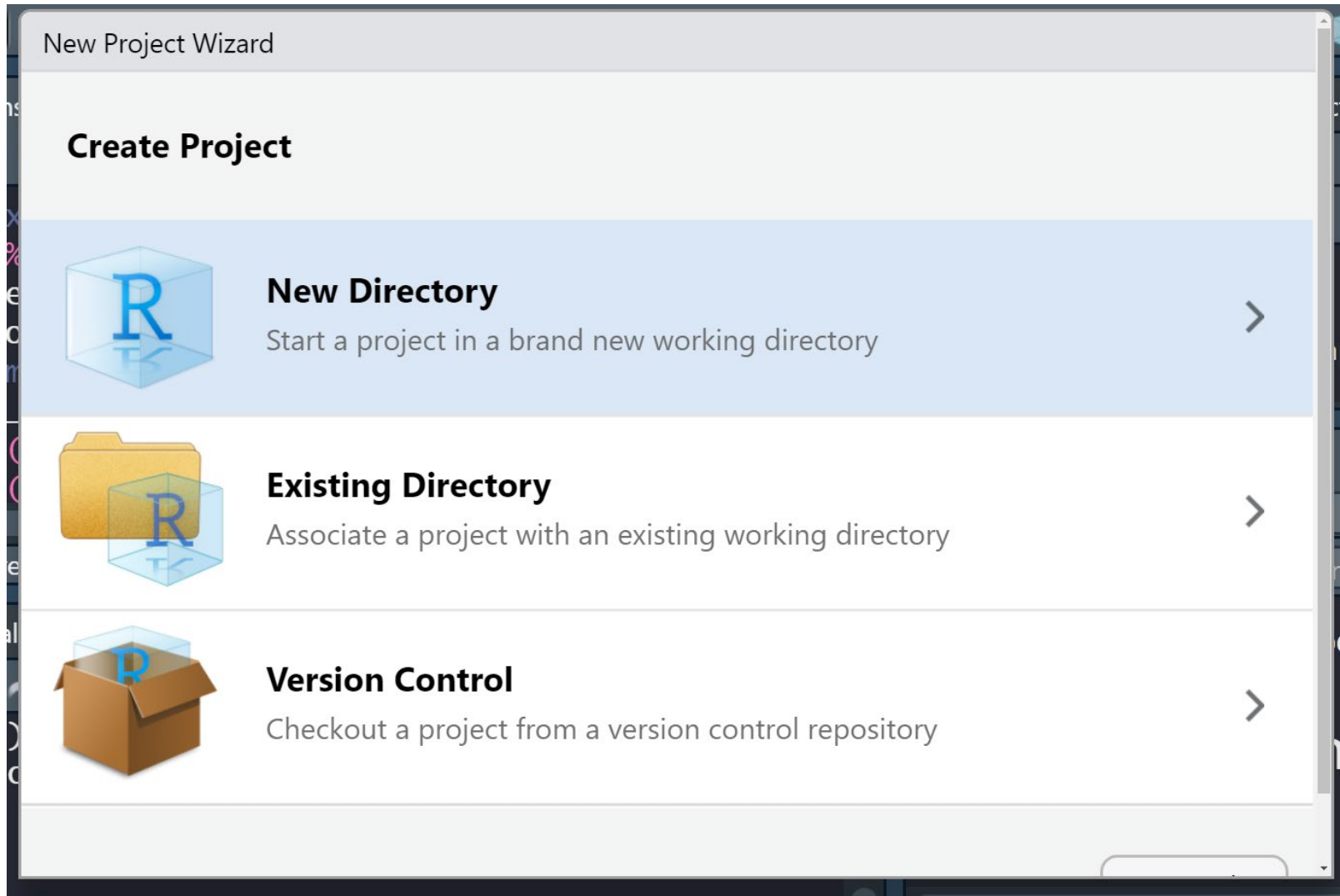
1. Load and merge data (Always work from your raw data!)
2. Data wrangling
3. Data analyses
4. Generate outputs such as tables and figures

# Create project for RforNatRes

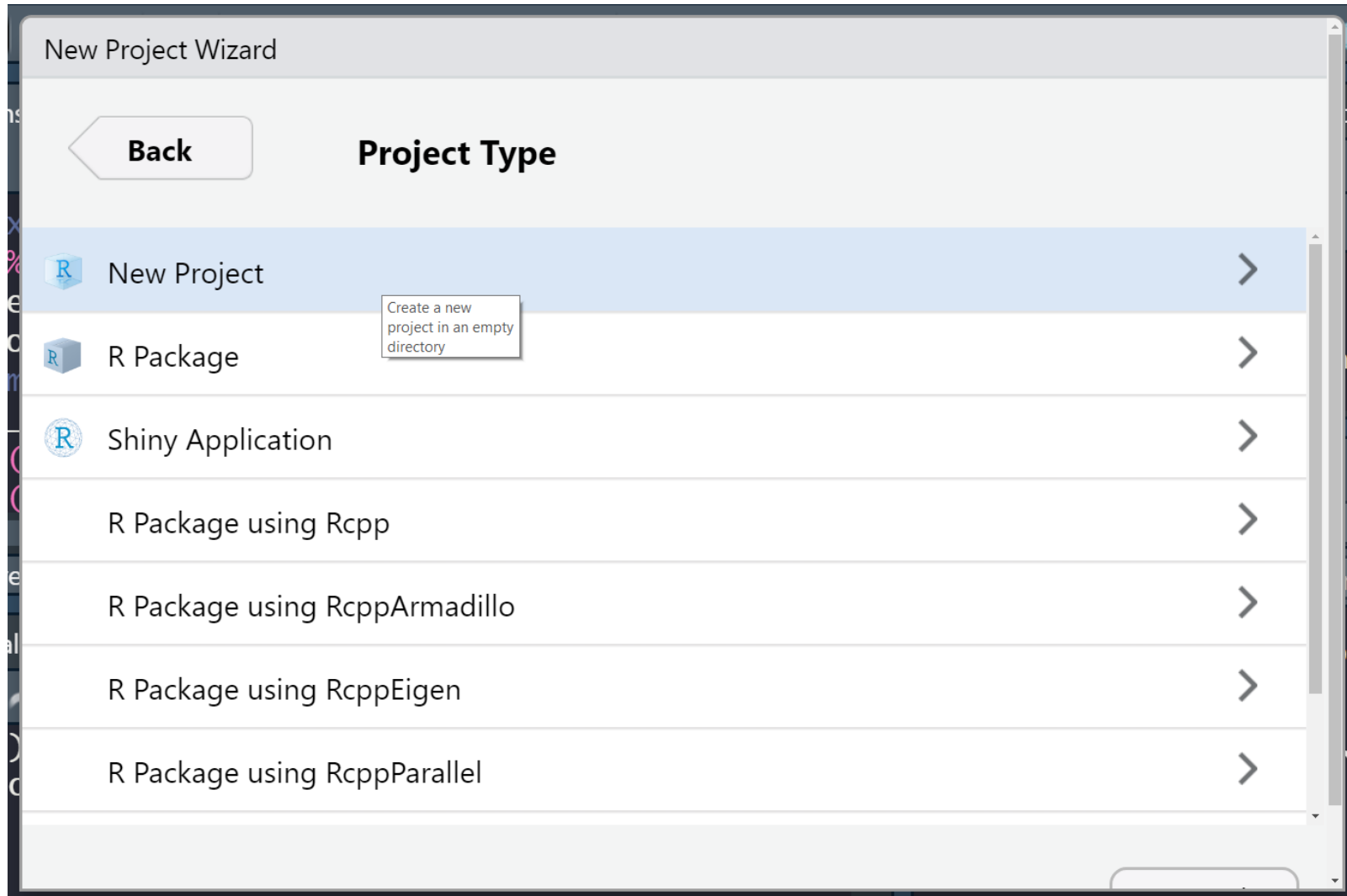
 RStudio



# Create project for RforNatRes




# Create project for RforNatRes



# Create project for RforNatRes

New Project Wizard

[Back](#) **Create New Project**



Directory name:

Create project as subdirectory of:  
 [Browse...](#)

☐ Create a git repository

☐ Use renv with this project

☐ Open in new session

