

Demo 1

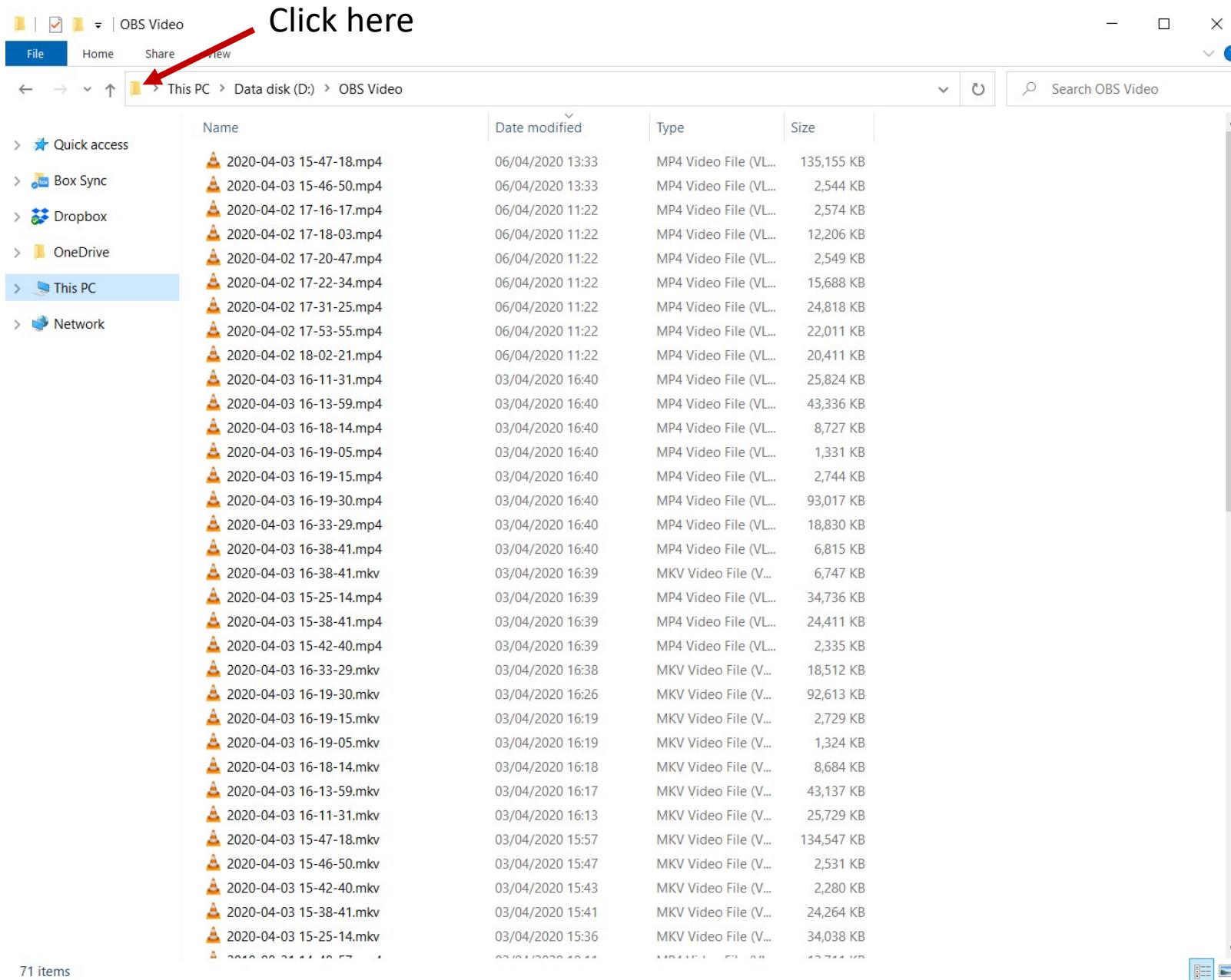
R for statistical analysis

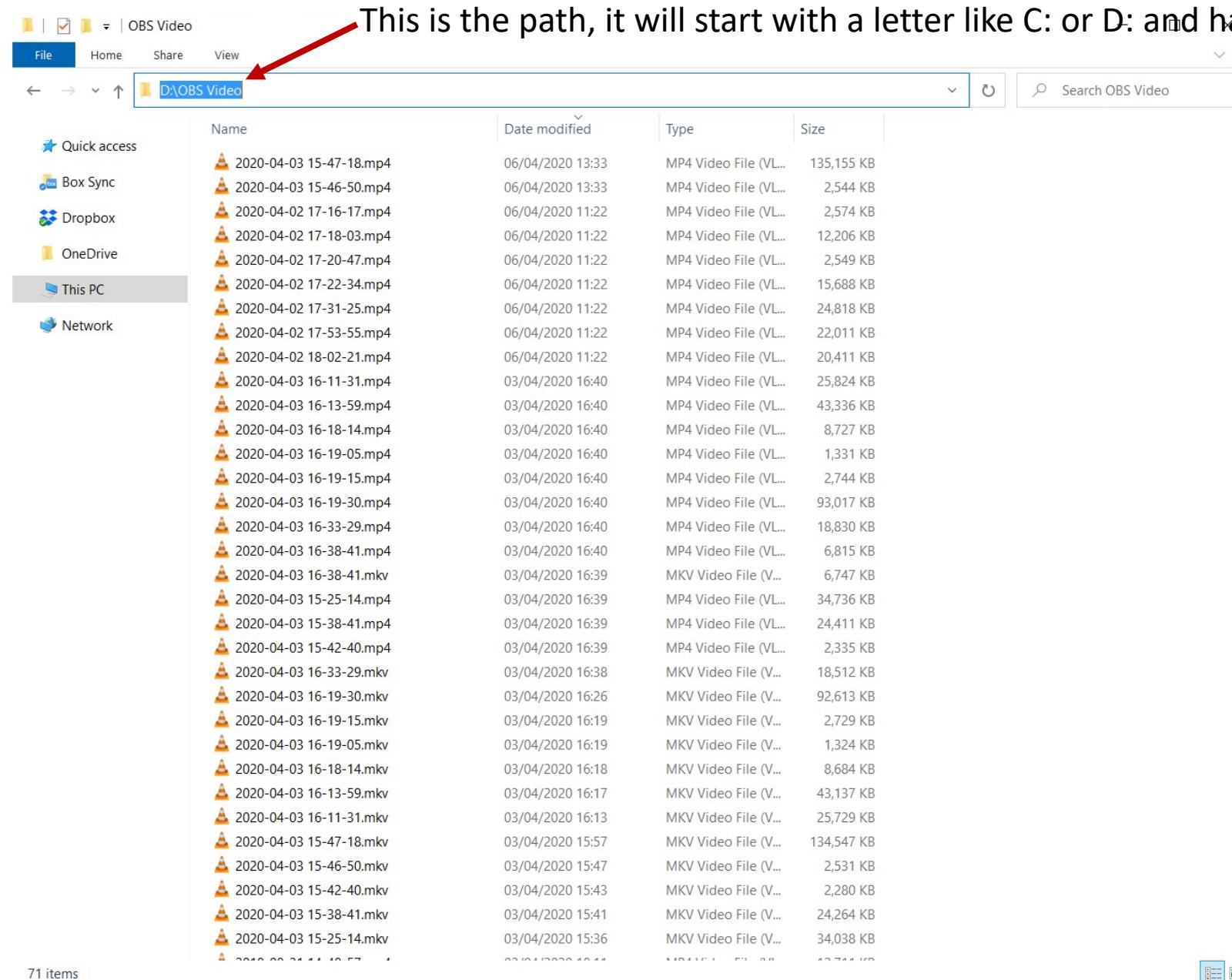
Juulia T. Suvilehto D.Sc.(tech)

How to use R

Before starting

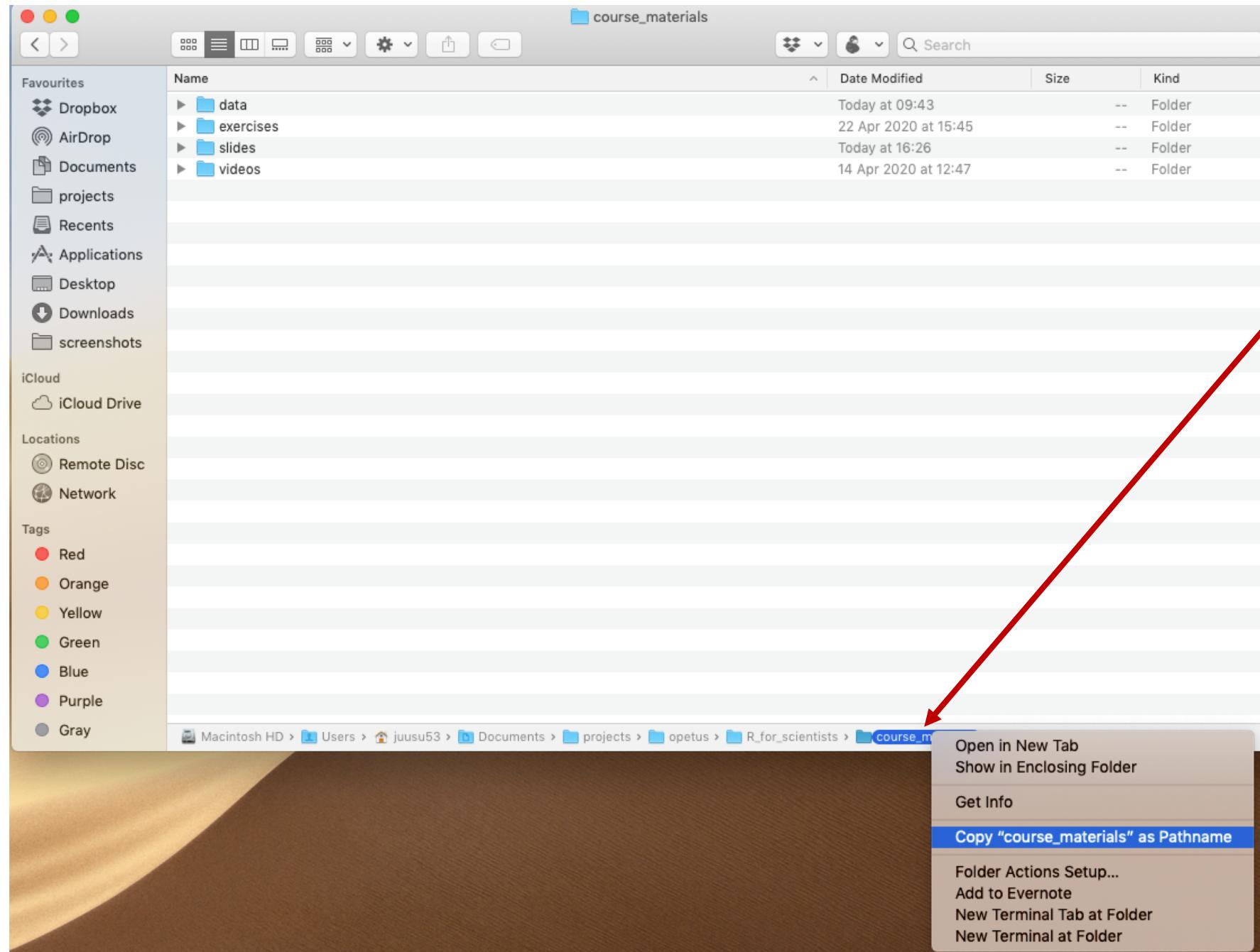
- Make sure you have both R (the programming language) and RStudio (an integrated development environment for R) installed
- Make a folder where you will save all code, data etc. related to this course
- Make a note of the location (absolute path) of this folder
 - In windows





Before starting

- Make sure you have both R (the programming language) and RStudio (an integrated development environment for R) installed
- Make a folder where you will save all code, data etc. related to this course
- Make a note of the location (absolute path) of this folder
 - In windows
 - In mac



Right click on the folder name here, then select copy as pathname

Path will look something like
`/Users/juusu53/Documents/projects/opetus/R_for_scientists/course_materials`

Path will start with /, and the separator will be / (different than in windows!)

Before starting

- Make sure you have both R (the programming language) and RStudio (an integrated development environment for R) installed
- Make a folder where you will save all code, data etc. related to this course
- Make a note of the location (absolute path) of this folder
 - In windows
 - In mac
- Open RStudio

Getting to know RStudio

RStudio

Untitled1 x

Source on Save | Run | Source | List |

1

1:1 (Top Level) R Script

Console Terminal

~/

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
> |
```

Environment History Connections

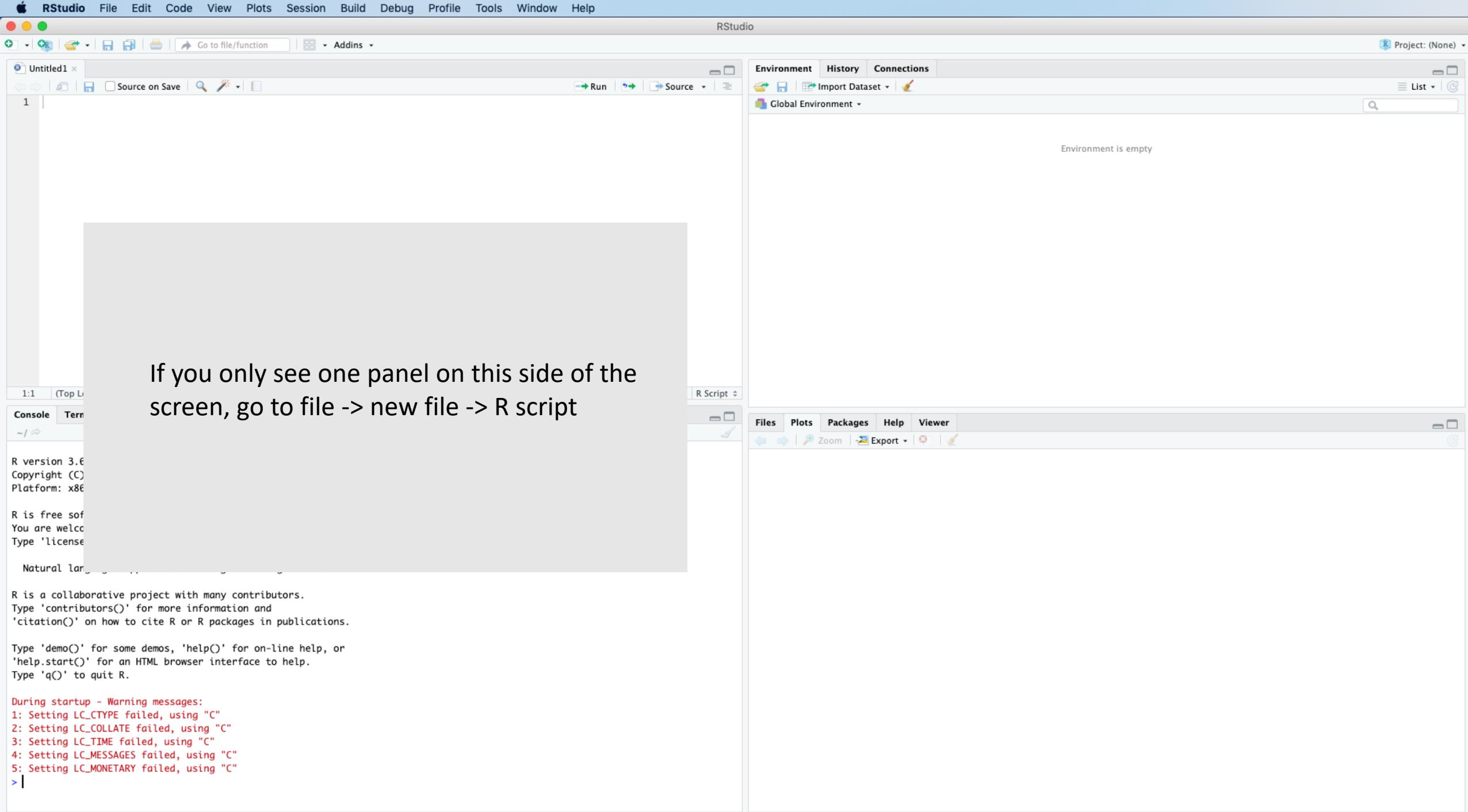
Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

Zoom Export



RStudio

Untitled1 x

Source on Save | Run | Source | List |

1

1:1 (Top Level) R Script

Console Terminal

~/

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
> |
```

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

Zoom Export

RStudio

Untitled1 x

Source on Save | Run | Source | List |

1

1:1 (Top Level) R Script

Console Terminal

~/

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
> |
```

Console

Environment History Connections

Import Dataset Global Environment

Environment is empty

Files Plots Packages Help Viewer

Zoom Export

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help RStudio Addins Untitled1 x Source on Save Run Source Environment History Connections Import Dataset Global Environment List Project: (None) Environment is empty

Script window

You create a new script file with
File > New File > R script

1:1 (Top Level) R Script

Console Terminal ~/

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
> |
```

Console

The image shows the RStudio interface with several windows open:

- Script window:** A new script file named "Untitled1" is open, showing the code "1".
- Console:** Displays the R startup message and various system information.
- Plots & Help:** A tabbed interface showing help and plots.
- Environment:** Shows the Global Environment tab with the message "Environment is empty".

Script window

You create a new script file with
File > New File > R script

1:1 (Top Level) R Script

Console Terminal

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
```

Console

Plots & Help

Will show help for functions and plots when you make them (in different tabs)

The image shows the RStudio interface with several sections:

- Script window:** A new script file named "Untitled1" is open, showing a single line of code "1".
- Environment:** The environment is empty.
- Console:** Displays the R startup message and license information.
- Plots & Help:** Shows text about displaying help and plots.

Script window

You create a new script file with
File > New File > R script

Environment

Will list your variables etc

Console

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

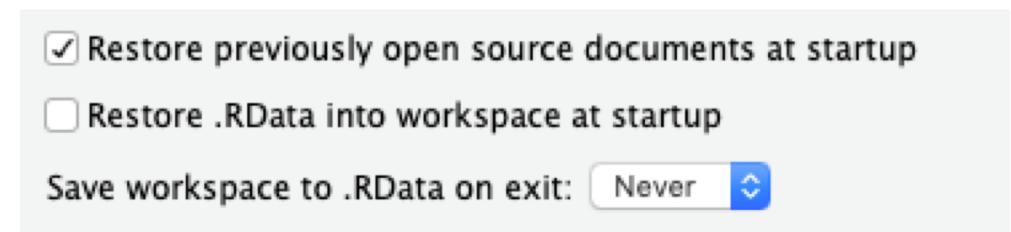
During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
> |
```

Plots & Help

Will show help for functions and
plots when you make them (in
different tabs)

RStudio settings

- You can change the look and some functionality of RStudio by going to tools -> options (windows) or RStudio -> preferences (mac)
- I warmly recommend you change your RStudio general settings to **not** restore .Rdata and to **never** save workspace to .RData on exit
- You can also set your default working directory (also part of general settings) to point to any folder you want, for example the folder you made for this course



How to use R

Calculations

- R knows the following simple mathematical operators

Addition	Substraction	Multiplication	Division	Power	Modulus
+	-	*	/	^	%%

- In order to "run" code (i.e. make the code execute), you need to either
 - Type it in the console and then click enter
 - Type it in the script and with the cursor on that row, press ctrl+enter (windows) or cmd+enter (mac)

Variables

- Variables are like the x you always had to solve for in math class
- Variable does not just have to hold a single numerical value, it can hold all kinds of things (more on this soon!)
- The variable name is a handle, which helps keep your code more readable and makes it easier for you to work with your data
- Naming your variables
 - Future You will be very happy if you name your variables in a descriptive way. For example, 'avg_women' is a better variable name than 'x3'
 - Variable names have to start with a letter
 - Variable names can also include numbers, underscores, and dots. For example, this_is_a_valid_variable_name, ThisOnels2, this.would.also.work
 - Variable names cannot include spaces or other special characters

Functions

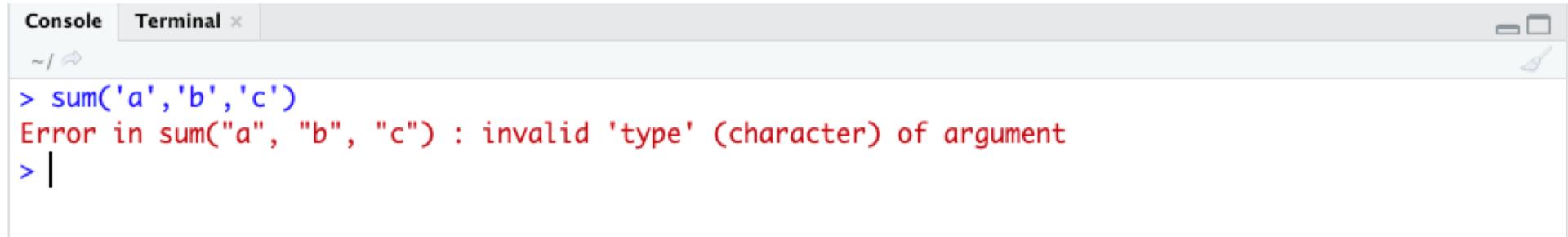
- Functions are operations (simple or complicated) which you can do to data
- A function call looks like this

function(arguments)

- What the function *returns* is called an *output*
- If you run a function without assigning the value to a variable, it will print out the output on the console
- If you assign the output to a variable, you will then need to call that variable to see the result

Errors

- If you try to do something R is not ready for, you will get an *error*



A screenshot of the RStudio interface showing the Console tab. The console window displays the following text:

```
Console Terminal ~ / 
> sum('a','b','c')
Error in sum("a", "b", "c") : invalid 'type' (character) of argument
> |
```

- Large part of learning how to program is learning how to figure out what went wrong & how to fix it
- Often reading the *error message* is helpful. You can also try to google your error message if it seems cryptic

Saving

- You can save your script file from file -> save, or using typical keyboard shortcuts
- Saving like this will only save the script file (top left), not the console or environment (your variables)
 - To make sure you are able to re-run the analysis from the script, you should make sure that you are not executing anything on the command line that will be needed in a later step

"99% of the time you are using R, you will do the following: 1) Define objects. 2) Apply functions to those objects. 3) Repeat!. Seriously, that's about it. ...

...However, as you'll soon learn, the hard part is knowing how to define objects they way you want them, and knowing which function(s) will accomplish the task you want for your objects."

From "YaRrr! The Pirate's Guide to R"
<https://bookdown.org/ndphillips/YaRrr/>

Simple data types

What are data types

- Data types are pre-defined data storage formats which can hold specific types of values
- It is important to understand the basics since different data types behave differently
- Numeric
 - Fairly intuitive
 - Two subtypes: double (default) and integer

Character (i.e. text)

- To differentiate character strings from variable names, they should be surrounded by double "" or single " quotes
- If the string itself needs to have a quote, you can use backslash to escape the character, like so 'Don\'t let anyone tell you you can\'t use apostrophes in a character string!' or you can use the other quote type to enclose the string
- Tip: R (like many programming languages) can be picky about the quotes – if you copy paste a string with quotes and running it in R gives you an error, try replacing the quotes in the text by ones you type yourself

Logical

- Two possible values, TRUE (or T or 1) and FALSE (or F or 0)
- Can be explicitly determined (`x <- TRUE`) or by logical expressions (`5 < 6`)
- Logical expressions can contain any number of variables and are very handy for indexing your data
 - For example, if you want to do an analysis on people who are women and over 50 years of age, the easiest way to find all of the participants who match that condition is a logical expression

Logical operators

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x y	x OR y
x & y	x AND y

Vectors (a data structure)

- Objects of the same type (e.g. numeric, character) can be grouped together into a 'vector' (1-d matrix)
- If you group together data types which can be coerced to the same data type, R will do that without asking you.
 - For example, `c(1,'a')` will create a vector with `[1, 'a']`, i.e. the number has been converted into a character
- Please note: A **list** is similar to vector in shape, but can mix data types. (More about lists later!)

Vector indexing

- We sometimes only want to access a single **element** of the vector
- Most common way to do this is to run something like this
 - vector[5]

which would return the 5th element of the variable called 'vector'

- The square brackets indicate that you want to get only part of the vector and the value(s) inside the square brackets indicate which part(s) you want to retrieve
- You can also use logical expressions to select elements, which fulfill specific criteria

Loops

What are loops?

- Recipes for repeated actions
- Same effort to do the same thing twice or 10000 times
- Generic format:

```
for (element in vector) {  
    do something to element  
}
```
- Note the indentation! R is not as picky with it as other programming language, but those extra spaces make your code much more readable
- Sidenote: other loops than for-loops exist, but these will be most useful for data analysis

Packages

Packages

- What
 - Groups of functions developed by someone else than the people who develop base R
 - Range from new approaches to data manipulation (like the tidyverse, which we'll discuss next week) to highly specialised packages for single analysis type (like linear mixed effect analyses)
- Why
 - Saves you a lot of effort developing your own functions
 - Usually faster & more reliable than anything you might put together
 - Good documentation & help forums available

How do I access these packages?

1. **Find** out which package you want (e.g. googling or being told by someone)
2. **Install** that package (only done once)
3. **Load** the package (at the beginning of every script)