

# 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

Click here

The screenshot shows a Windows File Explorer window titled 'OBS Video'. The address bar indicates the current location is 'This PC > Data disk (D:) > OBS Video'. The left sidebar shows 'This PC' selected under 'Quick access'. The main area displays a list of 71 items, which are video files. The files are organized into columns: Name, Date modified, Type, and Size. The files are sorted by date modified, with the most recent files at the top. The file names follow a pattern of 'YYYY-MM-DD HH-MM-SS.mp4' or 'YYYY-MM-DD HH-MM-SS.mkv'. The file types are either 'MP4 Video File (VLC...)' or 'MKV Video File (V...)'. The sizes range from 1,331 KB to 135,155 KB.

Name	Date modified	Type	Size
2020-04-03 15-47-18.mp4	06/04/2020 13:33	MP4 Video File (VLC...)	135,155 KB
2020-04-03 15-46-50.mp4	06/04/2020 13:33	MP4 Video File (VLC...)	2,544 KB
2020-04-02 17-16-17.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	2,574 KB
2020-04-02 17-18-03.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	12,206 KB
2020-04-02 17-20-47.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	2,549 KB
2020-04-02 17-22-34.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	15,688 KB
2020-04-02 17-31-25.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	24,818 KB
2020-04-02 17-53-55.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	22,011 KB
2020-04-02 18-02-21.mp4	06/04/2020 11:22	MP4 Video File (VLC...)	20,411 KB
2020-04-03 16-11-31.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	25,824 KB
2020-04-03 16-13-59.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	43,336 KB
2020-04-03 16-18-14.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	8,727 KB
2020-04-03 16-19-05.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	1,331 KB
2020-04-03 16-19-15.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	2,744 KB
2020-04-03 16-19-30.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	93,017 KB
2020-04-03 16-33-29.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	18,830 KB
2020-04-03 16-38-41.mp4	03/04/2020 16:40	MP4 Video File (VLC...)	6,815 KB
2020-04-03 16-38-41.mkv	03/04/2020 16:39	MKV Video File (V...)	6,747 KB
2020-04-03 15-25-14.mp4	03/04/2020 16:39	MP4 Video File (VLC...)	34,736 KB
2020-04-03 15-38-41.mp4	03/04/2020 16:39	MP4 Video File (VLC...)	24,411 KB
2020-04-03 15-42-40.mp4	03/04/2020 16:39	MP4 Video File (VLC...)	2,335 KB
2020-04-03 16-33-29.mkv	03/04/2020 16:38	MKV Video File (V...)	18,512 KB
2020-04-03 16-19-30.mkv	03/04/2020 16:26	MKV Video File (V...)	92,613 KB
2020-04-03 16-19-15.mkv	03/04/2020 16:19	MKV Video File (V...)	2,729 KB
2020-04-03 16-19-05.mkv	03/04/2020 16:19	MKV Video File (V...)	1,324 KB
2020-04-03 16-18-14.mkv	03/04/2020 16:18	MKV Video File (V...)	8,684 KB
2020-04-03 16-13-59.mkv	03/04/2020 16:17	MKV Video File (V...)	43,137 KB
2020-04-03 16-11-31.mkv	03/04/2020 16:13	MKV Video File (V...)	25,729 KB
2020-04-03 15-47-18.mkv	03/04/2020 15:57	MKV Video File (V...)	134,547 KB
2020-04-03 15-46-50.mkv	03/04/2020 15:47	MKV Video File (V...)	2,531 KB
2020-04-03 15-42-40.mkv	03/04/2020 15:43	MKV Video File (V...)	2,280 KB
2020-04-03 15-38-41.mkv	03/04/2020 15:41	MKV Video File (V...)	24,264 KB
2020-04-03 15-25-14.mkv	03/04/2020 15:36	MKV Video File (V...)	34,038 KB

71 items

This is the path, it will start with a letter like C: or D: and have \ as separator

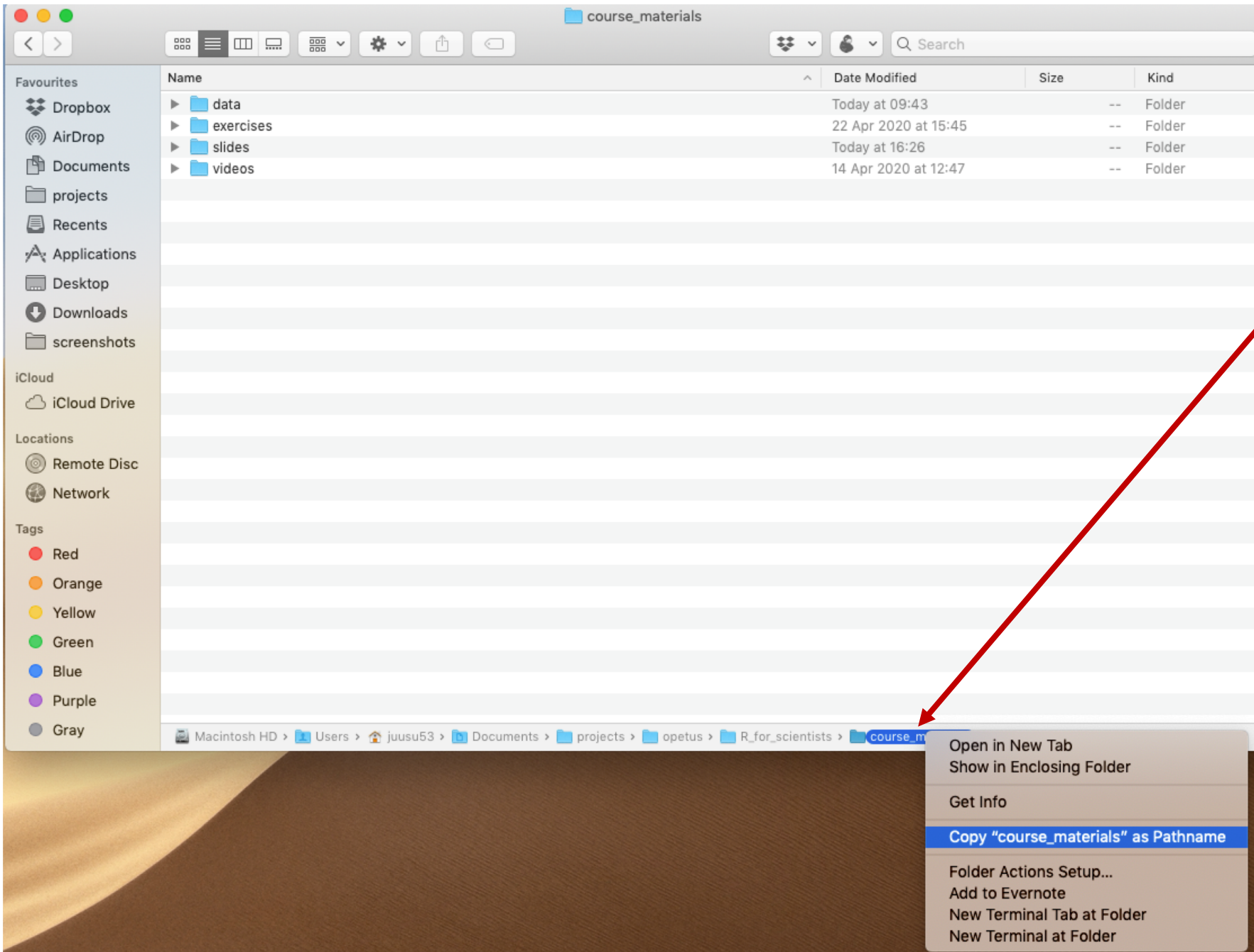
The screenshot shows a Windows File Explorer window with the address bar containing the path `D:\OBS Video`. A red arrow points to this path with the text "This is the path, it will start with a letter like C: or D: and have \ as separator". The left sidebar shows "This PC" selected. The main area displays a list of 71 items, all video files, with columns for Name, Date modified, Type, and Size.

Name	Date modified	Type	Size
2020-04-03 15-47-18.mp4	06/04/2020 13:33	MP4 Video File (VL...	135,155 KB
2020-04-03 15-46-50.mp4	06/04/2020 13:33	MP4 Video File (VL...	2,544 KB
2020-04-02 17-16-17.mp4	06/04/2020 11:22	MP4 Video File (VL...	2,574 KB
2020-04-02 17-18-03.mp4	06/04/2020 11:22	MP4 Video File (VL...	12,206 KB
2020-04-02 17-20-47.mp4	06/04/2020 11:22	MP4 Video File (VL...	2,549 KB
2020-04-02 17-22-34.mp4	06/04/2020 11:22	MP4 Video File (VL...	15,688 KB
2020-04-02 17-31-25.mp4	06/04/2020 11:22	MP4 Video File (VL...	24,818 KB
2020-04-02 17-53-55.mp4	06/04/2020 11:22	MP4 Video File (VL...	22,011 KB
2020-04-02 18-02-21.mp4	06/04/2020 11:22	MP4 Video File (VL...	20,411 KB
2020-04-03 16-11-31.mp4	03/04/2020 16:40	MP4 Video File (VL...	25,824 KB
2020-04-03 16-13-59.mp4	03/04/2020 16:40	MP4 Video File (VL...	43,336 KB
2020-04-03 16-18-14.mp4	03/04/2020 16:40	MP4 Video File (VL...	8,727 KB
2020-04-03 16-19-05.mp4	03/04/2020 16:40	MP4 Video File (VL...	1,331 KB
2020-04-03 16-19-15.mp4	03/04/2020 16:40	MP4 Video File (VL...	2,744 KB
2020-04-03 16-19-30.mp4	03/04/2020 16:40	MP4 Video File (VL...	93,017 KB
2020-04-03 16-33-29.mp4	03/04/2020 16:40	MP4 Video File (VL...	18,830 KB
2020-04-03 16-38-41.mp4	03/04/2020 16:40	MP4 Video File (VL...	6,815 KB
2020-04-03 16-38-41.mkv	03/04/2020 16:39	MKV Video File (V...	6,747 KB
2020-04-03 15-25-14.mp4	03/04/2020 16:39	MP4 Video File (VL...	34,736 KB
2020-04-03 15-38-41.mp4	03/04/2020 16:39	MP4 Video File (VL...	24,411 KB
2020-04-03 15-42-40.mp4	03/04/2020 16:39	MP4 Video File (VL...	2,335 KB
2020-04-03 16-33-29.mkv	03/04/2020 16:38	MKV Video File (V...	18,512 KB
2020-04-03 16-19-30.mkv	03/04/2020 16:26	MKV Video File (V...	92,613 KB
2020-04-03 16-19-15.mkv	03/04/2020 16:19	MKV Video File (V...	2,729 KB
2020-04-03 16-19-05.mkv	03/04/2020 16:19	MKV Video File (V...	1,324 KB
2020-04-03 16-18-14.mkv	03/04/2020 16:18	MKV Video File (V...	8,684 KB
2020-04-03 16-13-59.mkv	03/04/2020 16:17	MKV Video File (V...	43,137 KB
2020-04-03 16-11-31.mkv	03/04/2020 16:13	MKV Video File (V...	25,729 KB
2020-04-03 15-47-18.mkv	03/04/2020 15:57	MKV Video File (V...	134,547 KB
2020-04-03 15-46-50.mkv	03/04/2020 15:47	MKV Video File (V...	2,531 KB
2020-04-03 15-42-40.mkv	03/04/2020 15:43	MKV Video File (V...	2,280 KB
2020-04-03 15-38-41.mkv	03/04/2020 15:41	MKV Video File (V...	24,264 KB
2020-04-03 15-25-14.mkv	03/04/2020 15:36	MKV Video File (V...	34,038 KB

71 items

# 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`

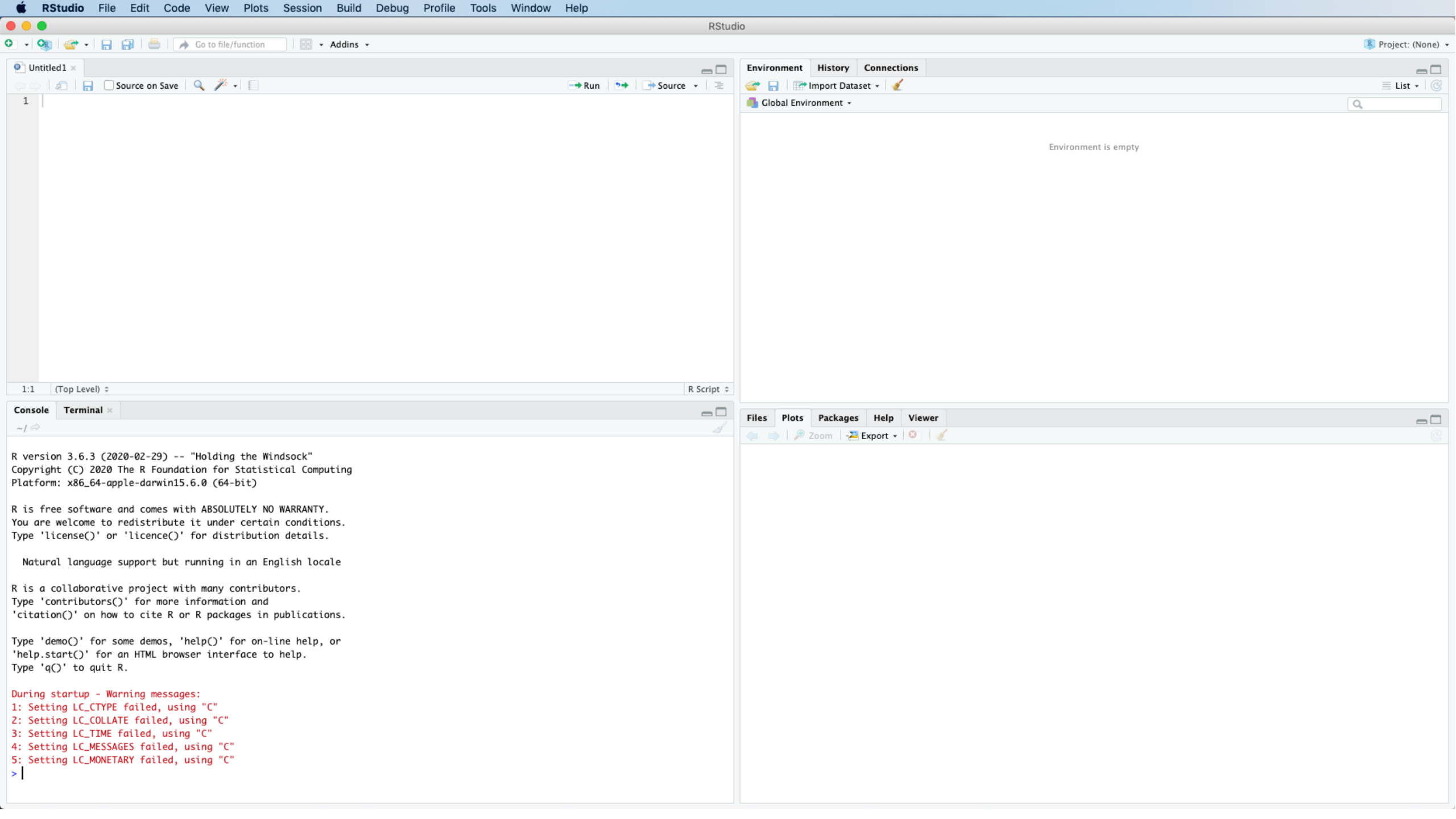
Path will start with /, and the separator will be / (different than in windows!)

# 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



```
Untitled1 x
Source on Save
1
Run Source
1:1 (Top Level) R Script
```

Environment History Connections

Global Environment

Environment is empty

```
Console Terminal
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windssock"
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

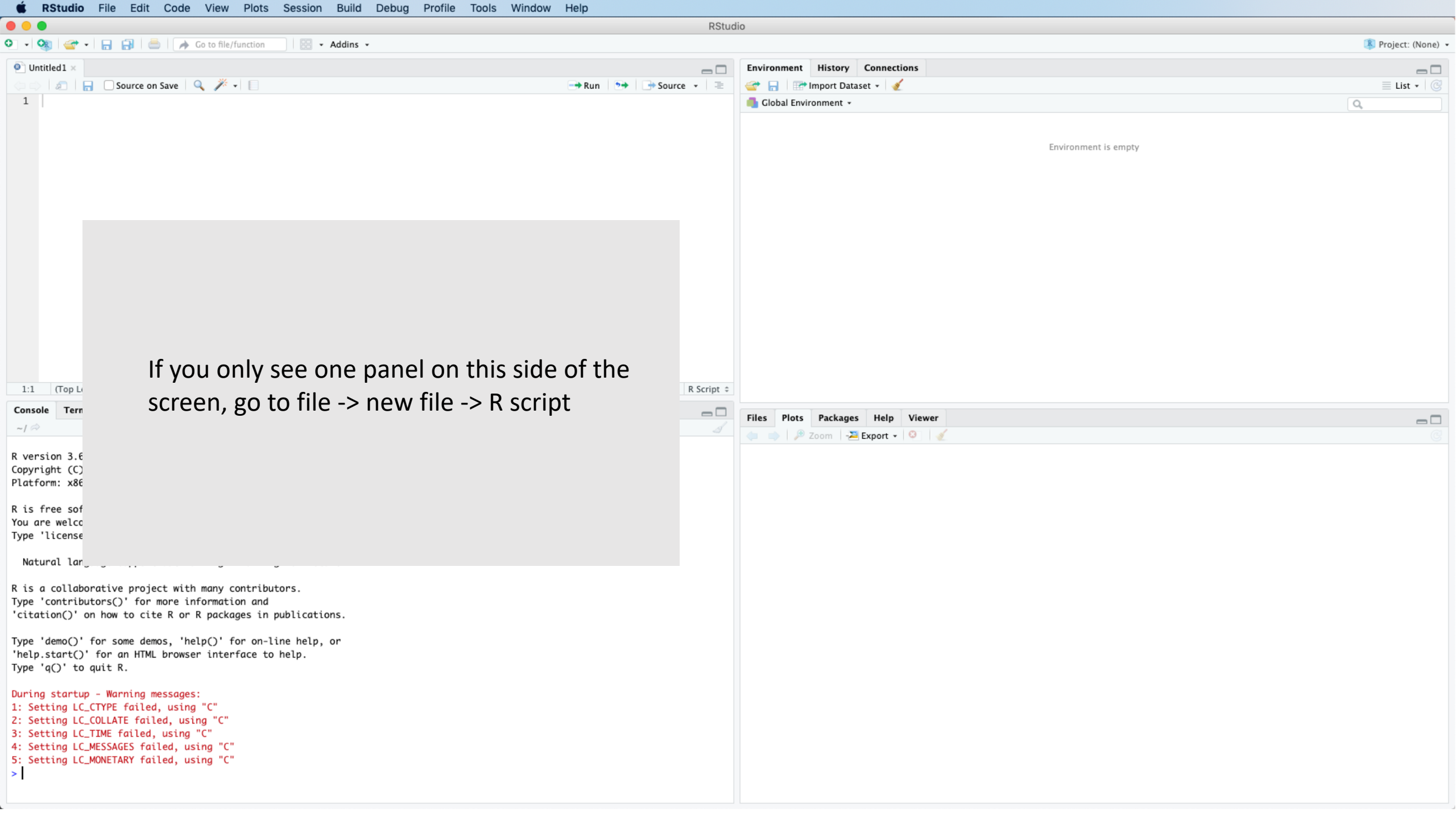
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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"
> |
```

Files Plots Packages Help Viewer

Zoom Export



If you only see one panel on this side of the screen, go to file -> new file -> R script

```
R version 3.6
Copyright (C)
Platform: x86

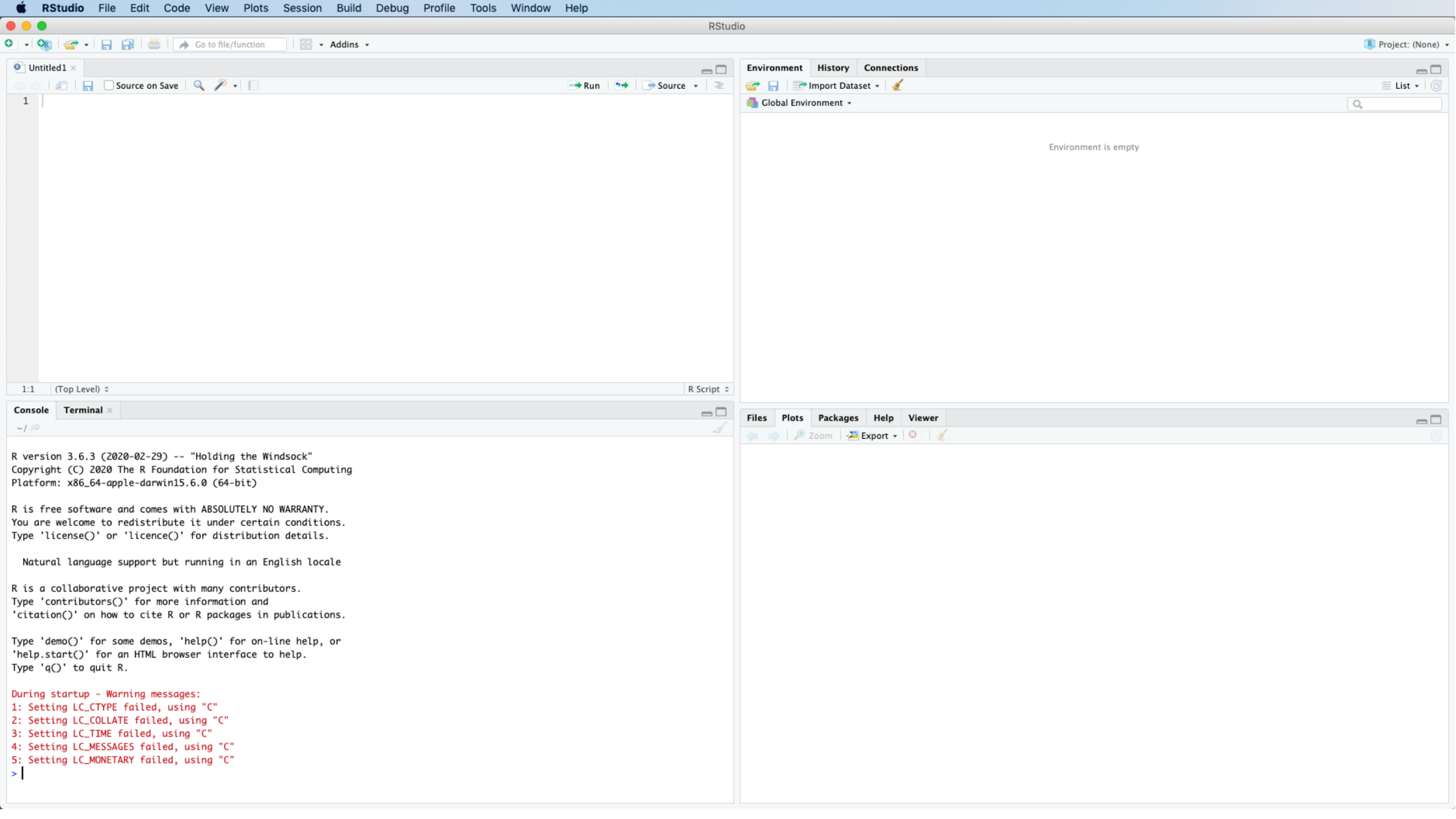
R is free sof
You are welcc
Type 'license

Natural lan

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"
> |
```



```
Untitled1 x
Source on Save
1
```

Environment History Connections

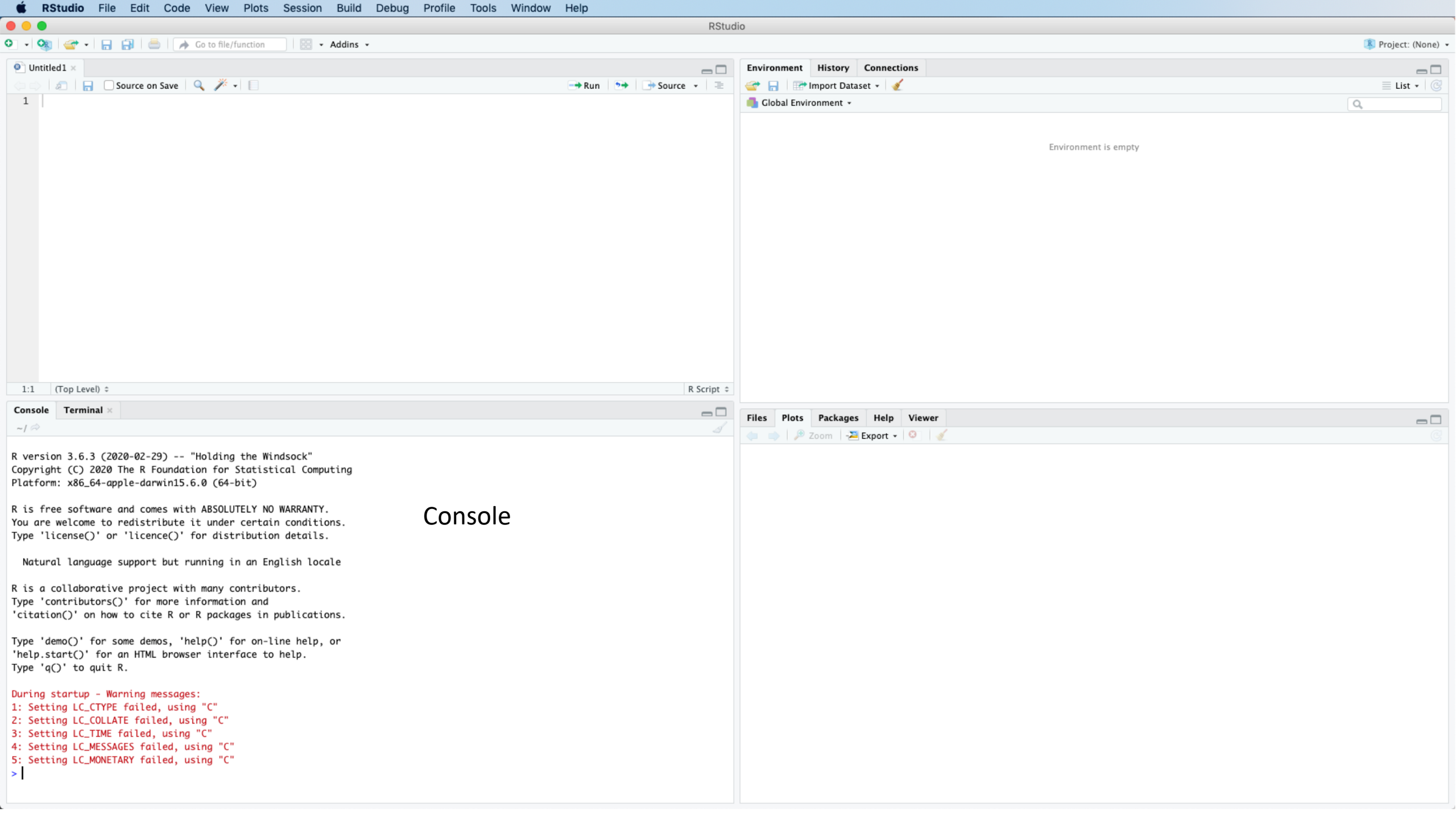
Global Environment

Environment is empty

```
Console Terminal x
~/
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"
>
```

Files Plots Packages Help Viewer

Zoom Export



## 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"  
>|
```

Untitled1

Source on Save

Run Source

```
1
```

1:1 (Top Level) R Script

## Script window

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

Environment History Connections

Global Environment

Environment is empty

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

Files Plots Packages Help Viewer

Zoom Export

Go to file/function Addins

Untitled1

Source on Save Run Source

1

## Script window

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

1:1 (Top Level) R Script

Environment History Connections

Import Dataset

Global Environment

Environment is empty

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

Files Plots Packages Help Viewer

Zoom Export

## Plots & Help

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

Go to file/function Addins

Untitled1

Source on Save Run Source

1

## 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

Environment History Connections

Import Dataset

Global Environment

Environment is empty

## Environment

Will list your variables etc

Files Plots Packages Help Viewer

Zoom Export

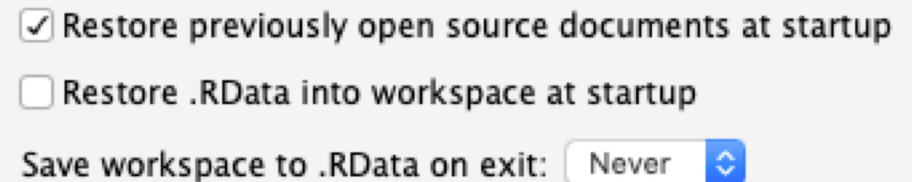
## Plots & Help

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`, `ThisOnes2`, `this.would.also.work`
  - Variable names cannot include spaces or other special characters

# Functions

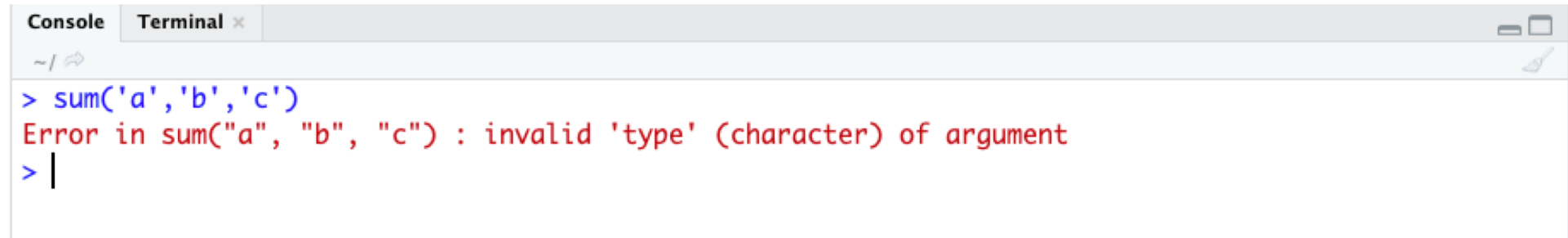
- Functions are operations (simple or complicated) which you can do to data
- A function call looks like this

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 an R console window. The window has a title bar with "Console" and "Terminal x" tabs. The main area shows the R prompt ">" followed by the command "sum('a','b','c')". Below the command, a red error message is displayed: "Error in sum("a", "b", "c") : invalid 'type' (character) of argument". The prompt ">" is followed by a vertical bar cursor "|".

```
Console Terminal x
~/
> 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)