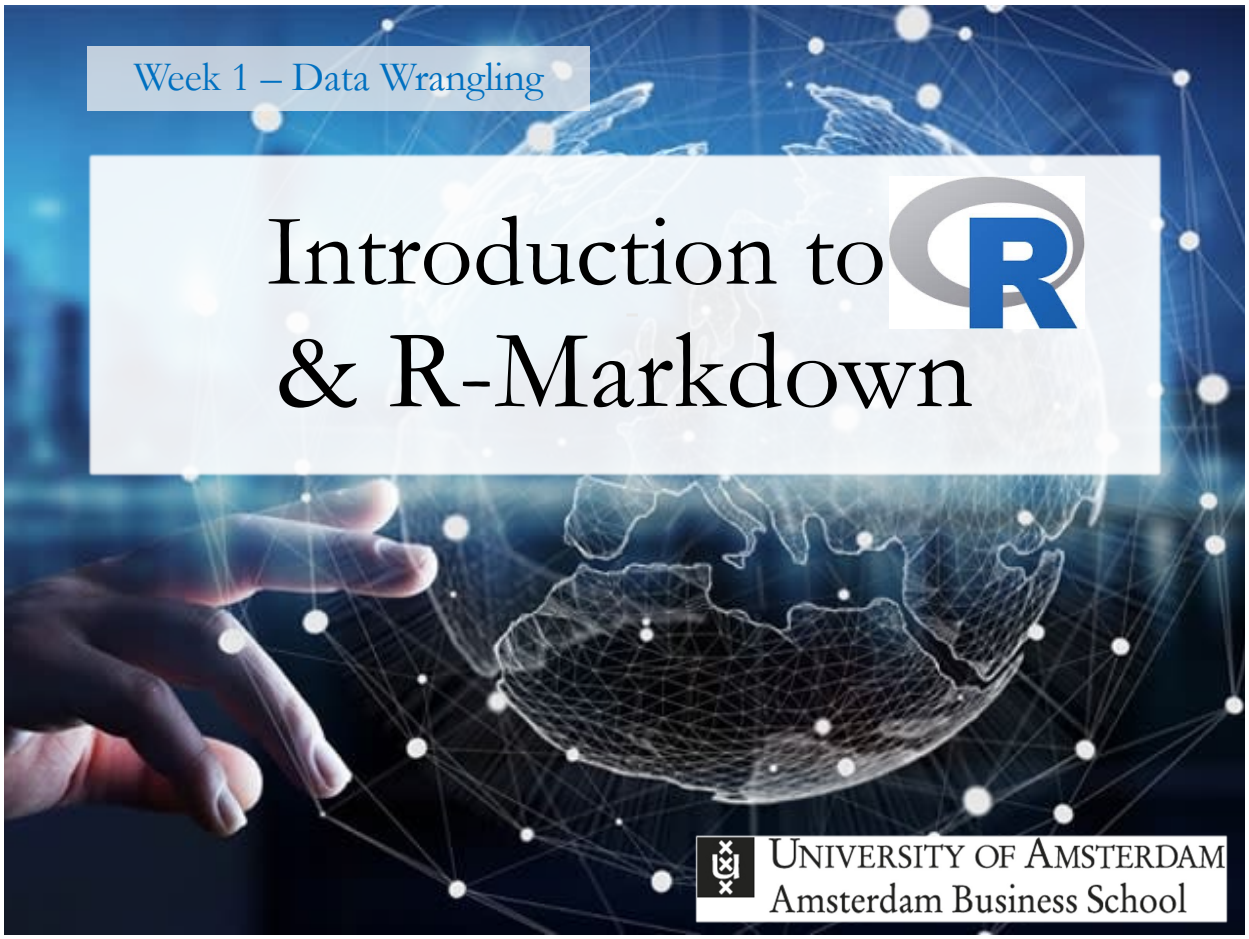
Week 1 – Data Wrangling

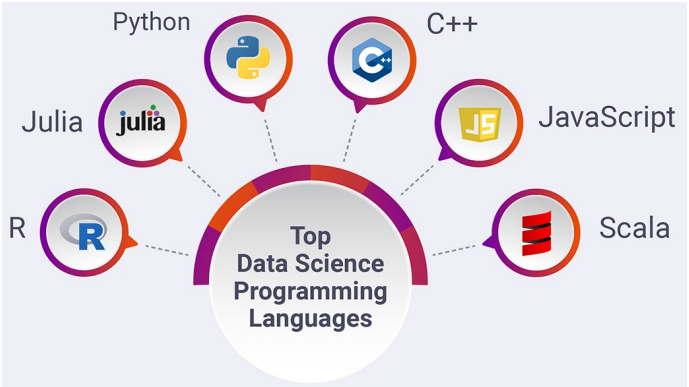
Introduction to R & R-Markdown



UNIVERSITY OF AMSTERDAM
Amsterdam Business School

Why R Software?

- R is free and Open-source Tool.
- R is one of the leading tools for Data Science, Statistics, and Machine Learning.
- For data-driven businesses, lack of Data Scientists is a huge concern. Companies are using R as their core platform and are recruiting trained R programmers.
- R contains actual machine and statistical techniques; new techniques are made available in R very quickly.



What is R? RStudio?

- R – open-source programing language for statistical computing

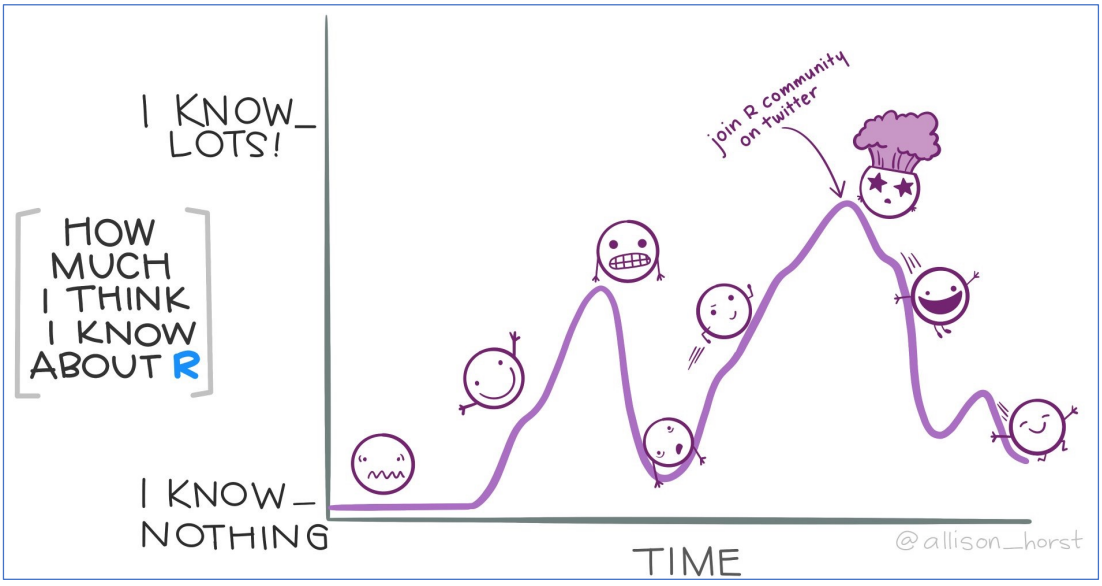
Link: https://cloud.r-project.org

- RStudio – popular editor to write R scripts and interact with R

Link: http://www.rstudio.com/download

How to learn R?

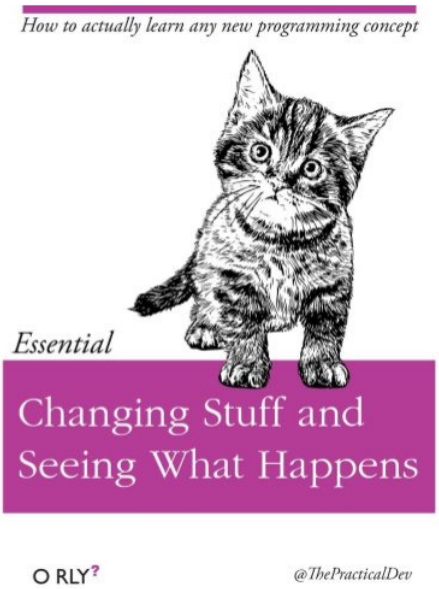
For those of you who are interested to learning a new programming language

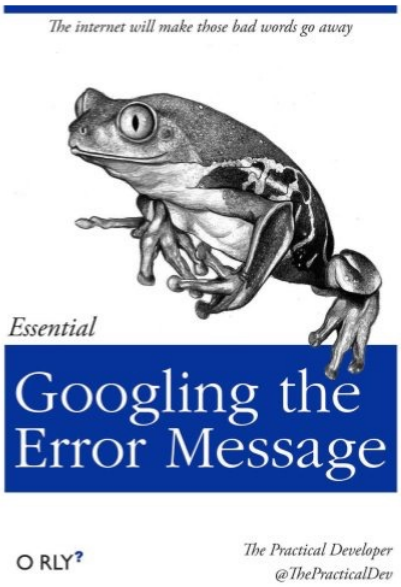


That how I learn R and C

How to learn R?

Here are some important things to keep in mind as you learn (these are joke book covers):





Introduction to R

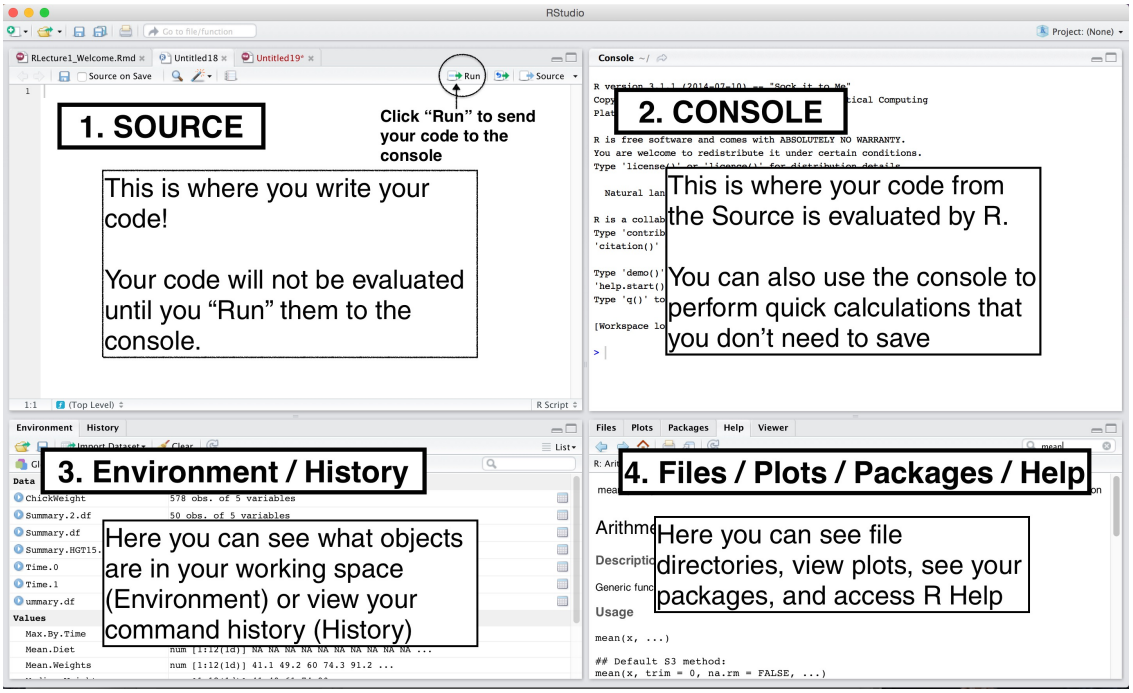
R in 2 hours

YouTube link: https://youtu.be/_V8eKsto3Ug

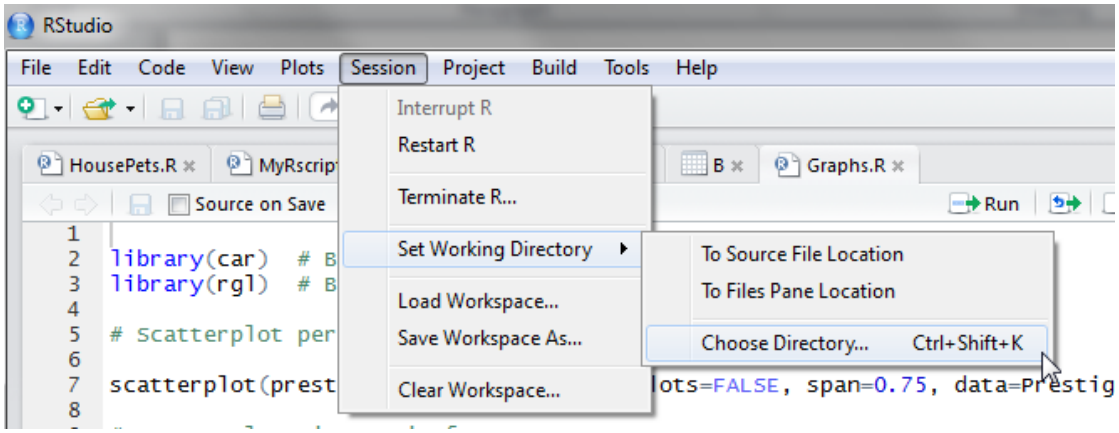
Introduction to R

DataCamp: https://app.datacamp.com/groups/data-wrangling-2022

RStudio Screen



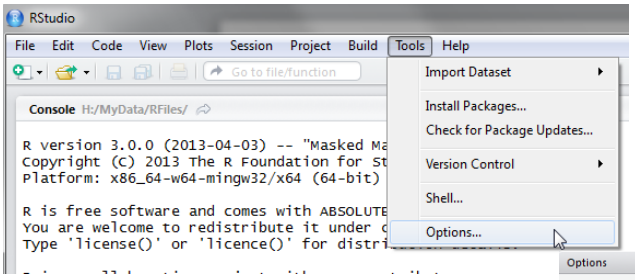
Change Working Directory

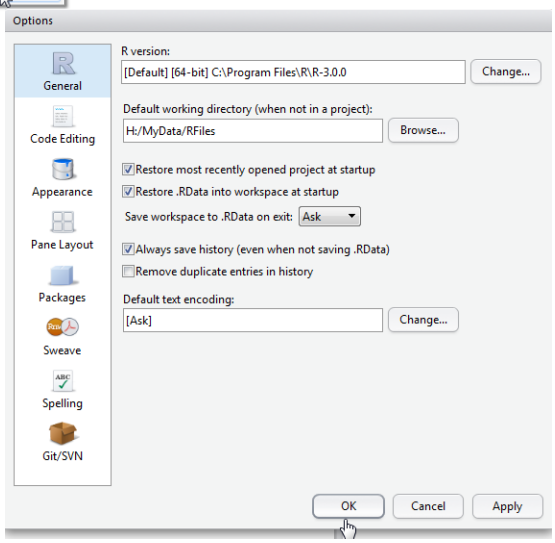


Or alternatively you can type the command

```
setwd( "c:/docs/mydir" ) # set working directory
getwd()    # shows the working directory
```


Set Default Working Directory



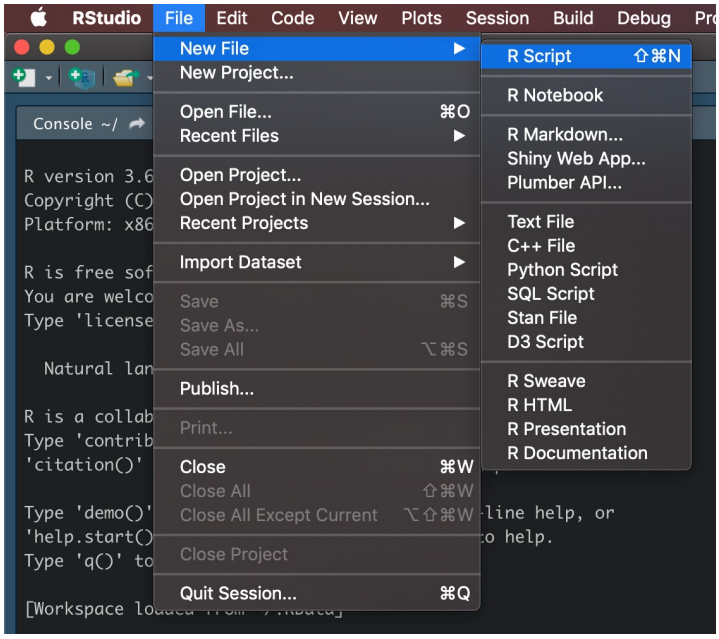


Each time you open R you go to a default working directory.

How to Create R Script?

Create an R Script file from Rstudio and save it with a name e.g. "week1-Rcode.R".

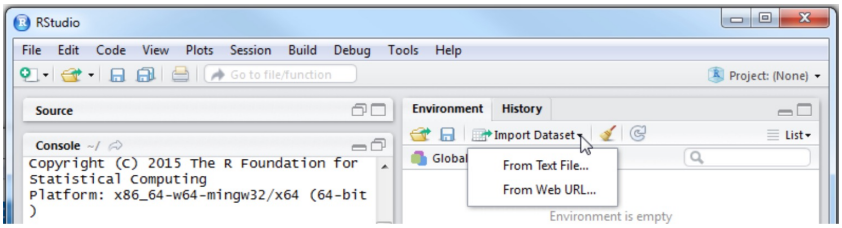
Then ready to go ….



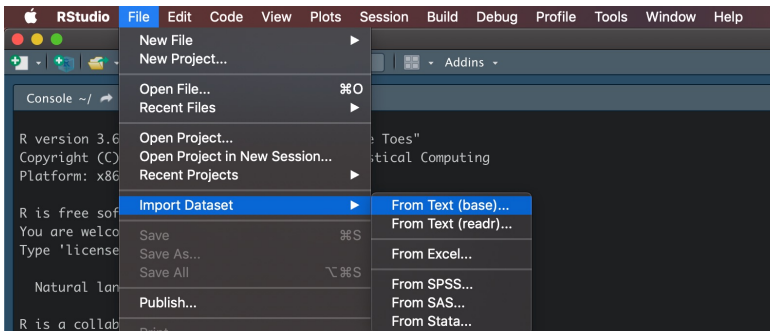
Load Data vis R Studio Menu Items

Two different places to load data in Rstudio:

1. In the toolbar of the upper right section of R Studio:



2. The screenshot shows where the "Import Dataset" menu item is located in R Studio's top menu



Load Data into R

We could load data into R by using "read.csv()" function.

Example: Loading the corona dataset:

```
data = read.csv( "corona.CSV" )
```

Note: In the above, only the file name itself is shown. Then R expects to find the file in the same directory R is running from. If you want to specify the full path to the file, you can do so too. Here is an example of how that looks on Mac:

```
data = read.csv( "~/Dropbox/Datasets/corona.CSV " )
```

The same file path on a Window machine could look like:

```
data = read.csv( "c:\\Dropbox\\Datasets\\corona.CSV " )
```


Load Data into R from online source

By using "read.csv()" function, you could load the data from online sources.

Example: Loading the corona dataset from the online source as follow.

```
# Read the Corona Dataset sheet into "R":

data=read.csv("https://opendata.ecdc.europa.eu/covid19/case
              distribution/csv", na.strings = "",
              fileEncoding = "UTF-8-BOM" )
```

Basic Commands

R uses functions to perform operations

```
x <- c( 1,3,2,5 ) # use function c() to make a vector
                  containing the numbers 1,3, 2 and 5
                  and save them as vector named x

mean(x)  # calculate mean of vector x

? mean() # get help on function mean()

mean(x, na.rm = TRUE) # na.rm = TRUE removes missing
                        values before the mean is
                        calculated (the default value for
                        na.rm is FALSE)
```


Basic Commands

R uses functions to perform operations

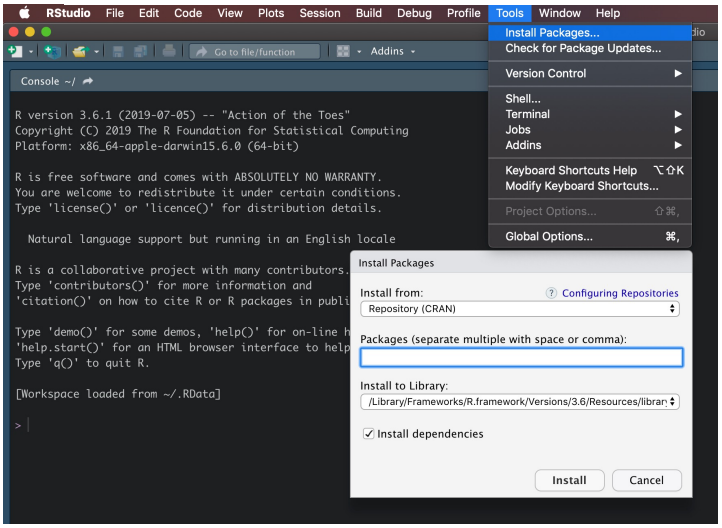
```
ls() # get a list of all objects such as data and vectors
       that have been used so far

rm( x ) # remove vector x

rm( list = ls() ) # remove all objects
```

Install/load Packages

Sometimes we need a function/dataset which is not in R. So, we need to install and load the package that includes this function. (We need to install the package once; the next time we open R we only need to load the package.)



Code in R:

```
install.packages( "ggplot2" )

library( "ggplot2" )   4
```


What are the basic datatypes?

```
# Declare variables of different types

my_numeric = 42

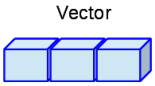
my_character = “universe”

my_logical = FALSE

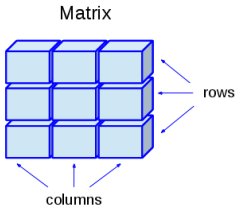
# Note that R is case sensitive!
```

Data types in R

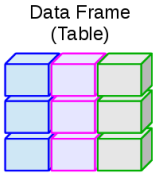
- Vector



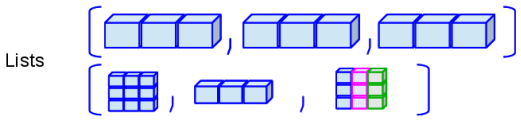
- Matrix



- Data frame



- List



How to create a vector?

A one-dimensional array that can hold numeric data, character data, or logical data

Example:

```
numeric_vector   = c( 1, 2, 3 )
character_vector = c( “a”, “b”, “c” )
```

How to create a Matrix?

A two-dimensional array with a fixed number of rows and columns that can hold numeric data, character data, or logical data

Example:

```
matrix( 1 : 9, byrow = TRUE, nrow = 3 )

     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9
```


How to create a data frame?

A data frame has variables of different data types of a data set as columns and the observations as rows

Example:

```
# Definition of vectors
name     = c( “Mercury”, “Venus”, “Earth”, “Mars” )
diameter = c( 0.382, 0.949, 1, 0.532 )
rotation = c( 58.64, -243.02, 1, 1.03 )
rings    = c( FALSE, FALSE, FALSE, FALSE )

# Create a data frame from the vectors
planets_df = data.frame( name, diameter, rotation, rings )
```

How to create a list?

A list allows us to gather a variety of objects under one name (that is, the name of the list) in an ordered way. These objects can be matrices, vectors, data frames, even other lists, etc.

Example:

```
# Vector with number from 1 up to 10
my_vector = 1 : 10
# Matrix with numerics from 1 up to 9
my_matrix = matrix(1 : 9, ncol = 3)
# First 10 elements of the built-in data frame mtcars
my_df = mtcars[1 : 10, ]
# Construct list with these different elements:
my_list = list(my_vector, my_matrix, my_df)
```

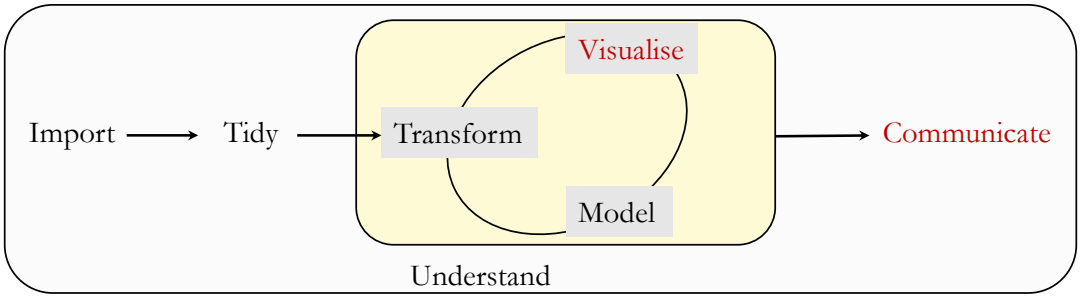

More info about R

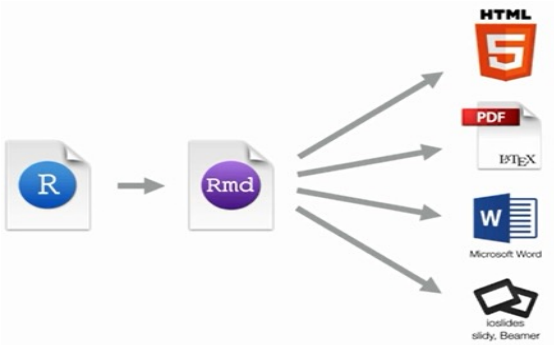
- Free Online Course at DataCamp:
  https://www.datacamp.com/courses/free-introduction-to-r

- YouTube - R Programming Tutorial:
  https://youtu.be/_V8eKsto3Ug

Reporting with R Markdown

It doesn't matter how great your analysis is unless you can explain it to others: you need to communicate your results.



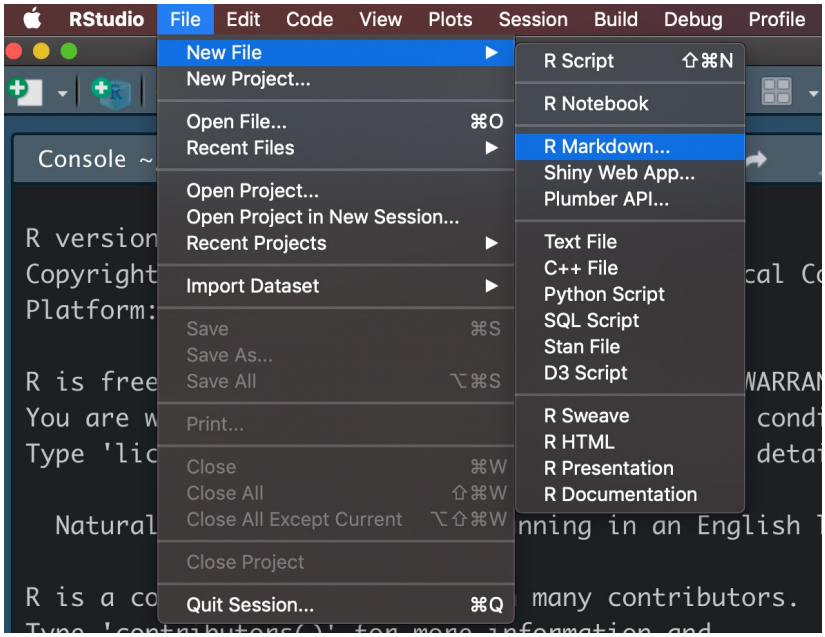


R Markdown

YouTube link: https://youtu.be/DNS7i2m4sB0

How to create R Markdown?

We can create R Markdown in Rstudio by going to "File > New File > R Markdown":



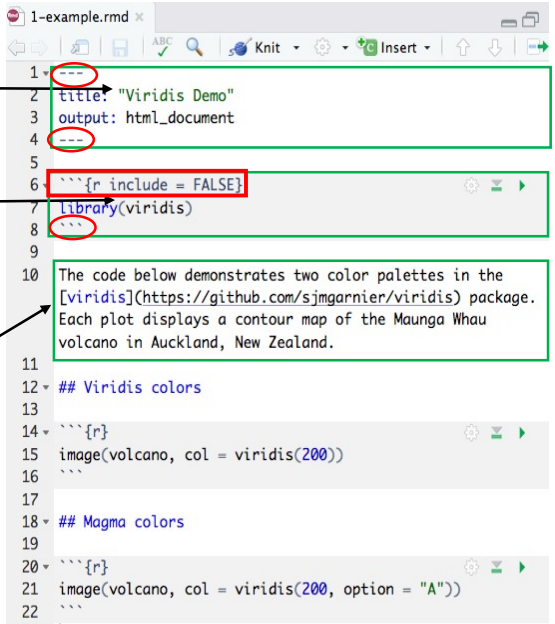
How R Markdown works?

Open the file "RMarkdown-Example-1.Rmd" which is at Canvas.
This file contains three types of content:

An (optional) YAML header surrounded by ---s

R code chunks surrounded by ```s

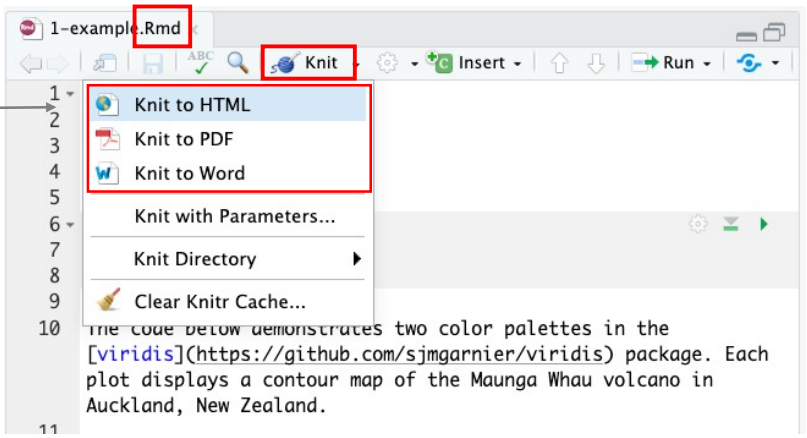
Text mixed with simple text formatting.



How R Markdown works?

To create

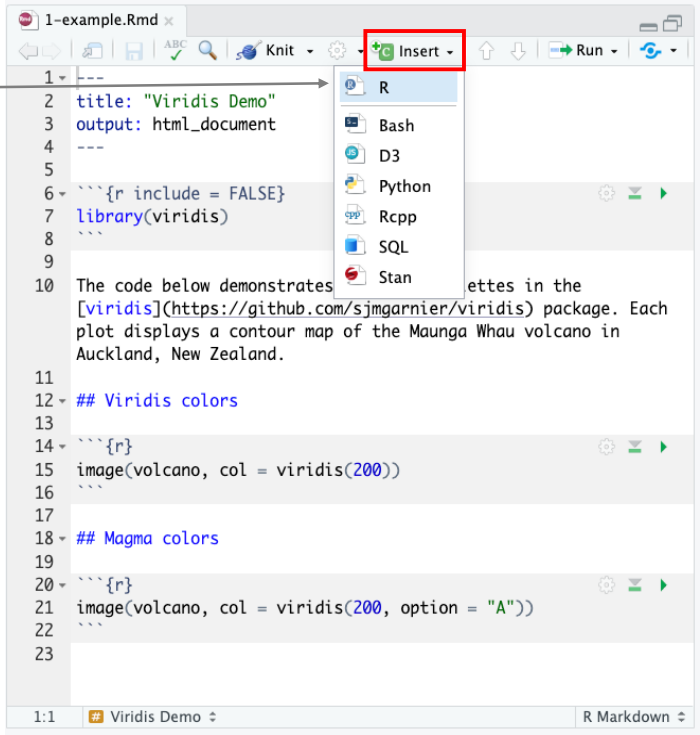
- HTML
- PDF
- Word



To generate a report from the .Rmd file, use the "Knit" button in the RStudio to render the file and preview the output with a single click or keyboard shortcut (⇧⌘K).

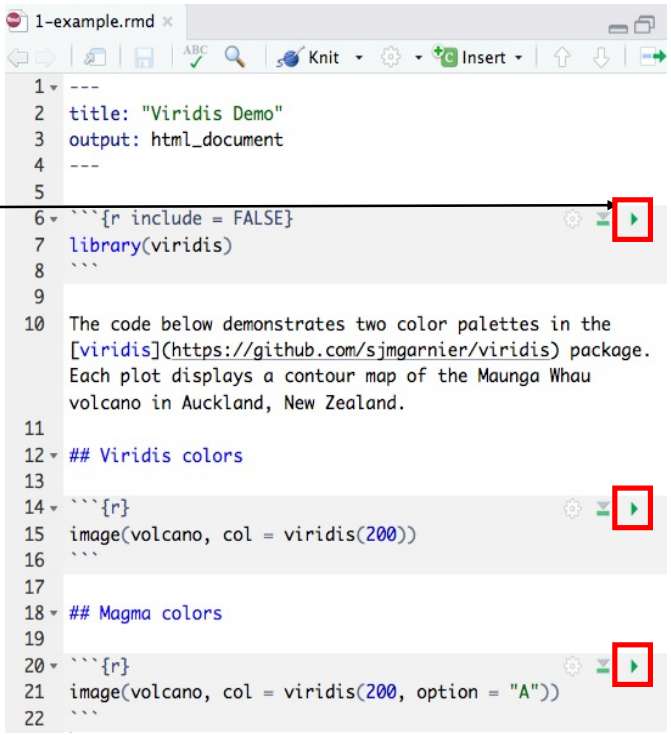
How R Markdown works?

To insert code in the report. E.g. R code.

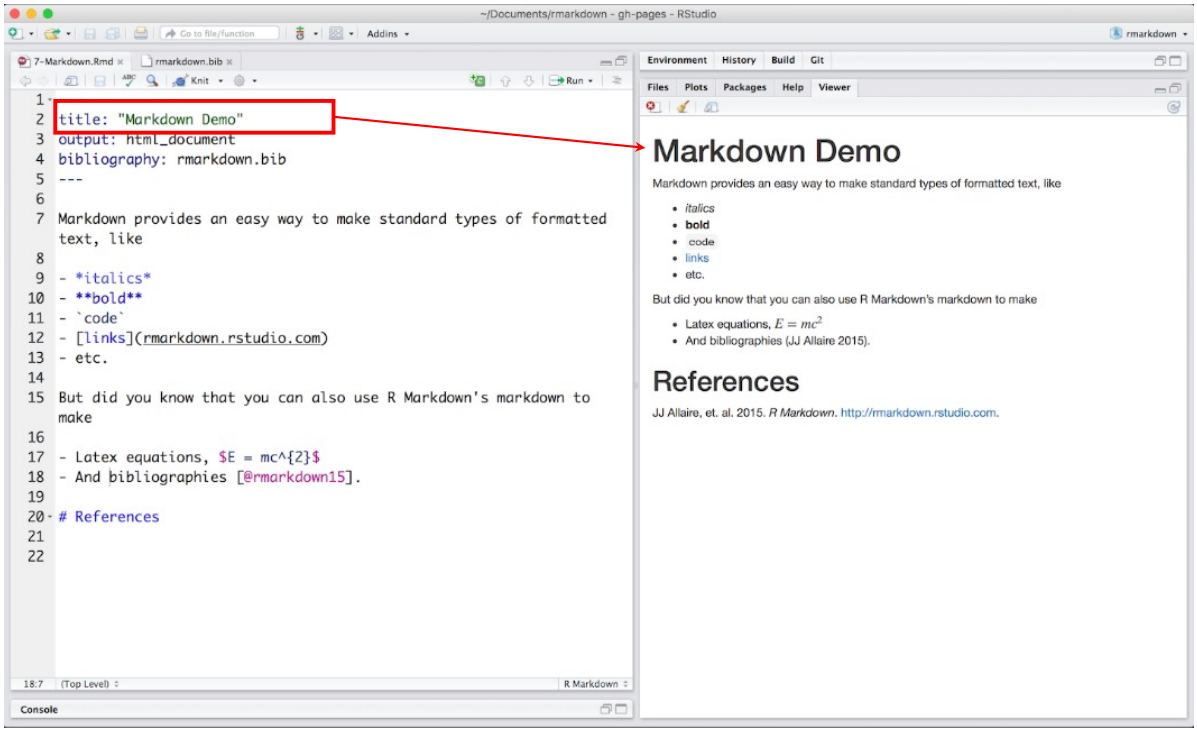


How R Markdown works?

You can run each code chunk of R code by clicking the ▶ icon.



Rmd syntax



Styling narrative sections

In Markdown, we can embed formatting instructions into your text. e.g.:

- Italicized by surrounding it in asterisks: *italics*.
- Bold by surrounding it in two asterisks: **bold**.
- monospaced (like code) by surrounding it in backticks: `code`.

You can turn a word into a link like:

```
[RStudio](www.rstudio.com)
```

To create title and headers, use leading hashtags. The number of hashtags determines the header's level:

```
# First level header
## Second level header
### Third level header
```


List

To make a bulleted list in Markdown, place each item on a new line after an asterisk and a space, like:

```
* item 1
* item 2
* item 3
```

Similarly, you can make an ordered list like:

```
1. item 1
2. item 2
3. item 3
```

In each case, you need to place a blank a blank line between the list and any paragraphs that come before it.

R code chunks

You can embed R code into your R Markdown report with the knitr syntax, by surrounding R code with two lines:

```
```{r}
  # some R code
```
```

When you render the report, R will execute the code. If the code returns any results, R will add them to your report.

Customize R code chunks

You can customize each R code chunk by providing optional arguments after the r in ```{r}.

R functions sometimes return messages, warnings, and error messages. By default, R Markdown will include these messages in your report. You can use the message, warning and error options to prevent R Markdown from displaying these.

Example: R Markdown would ignore any errors or warnings generated by the chunk below.

```
```{r  warning = FALSE,  error = FALSE}
  1 + 2
```
```

Popular chunk options

Three most popular chunk options are echo, eval and results:

- If echo = FALSE, R Markdown will not display the code in the final document, but still run the code and display its results.
- If eval = FALSE, R Markdown will not run the code or include its results, but still display the code.
- If results = `hide`, R Markdown will not display the results of the code, but still run the code and display the code itself.

Table of chunk options

Option	Run code	Show code	Output	Plots	Messages	Warnings
eval = FALSE	X	✓	X	X	X	X
include = FALSE	✓	X	X	X	X	X
echo = FALSE	✓	X	✓	✓	✓	✓
results = FALSE	✓	✓	X	✓	✓	✓
fig.show = FALSE	✓	✓	✓	X	✓	✓
message = FALSE	✓	✓	✓	✓	X	✓
warning = FALSE	✓	✓	✓	✓	✓	X

Inline R code

R code can be embedded into the text of document with the `r ` syntax; Be sure to include the lower case r. R Markdown will run the code and replace it with its result.

For example, the line below uses embedded R code to create a complete sentence:

```
The factorial of four is `r factorial(4)`.
```

When you render the document, the result will appear as:

```
The factorial of four is 24.
```

Inline code provides a useful way to make your reports completely automatable.

More Information

- See an example on Canvas > Week 1.
- Online exercise on DataCamp:

  https://learn.datacamp.com/courses/reporting-with-rmarkdown

- Other useful links:

  https://rmarkdown.rstudio.com/lesson-1.html

  http://r4ds.had.co.nz/r-markdown.html