

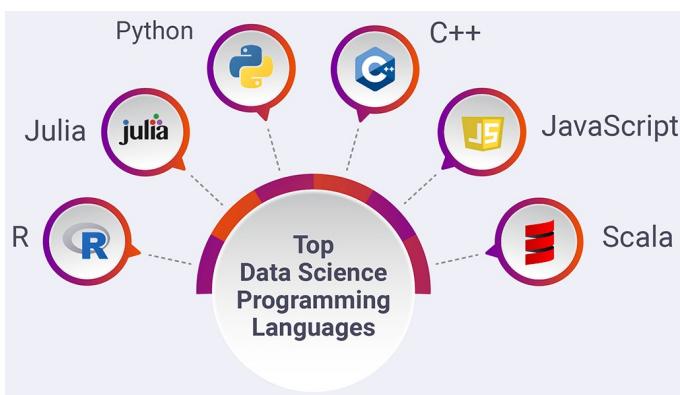
Introduction to & R-Markdown

 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 programming language for statistical computing



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

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

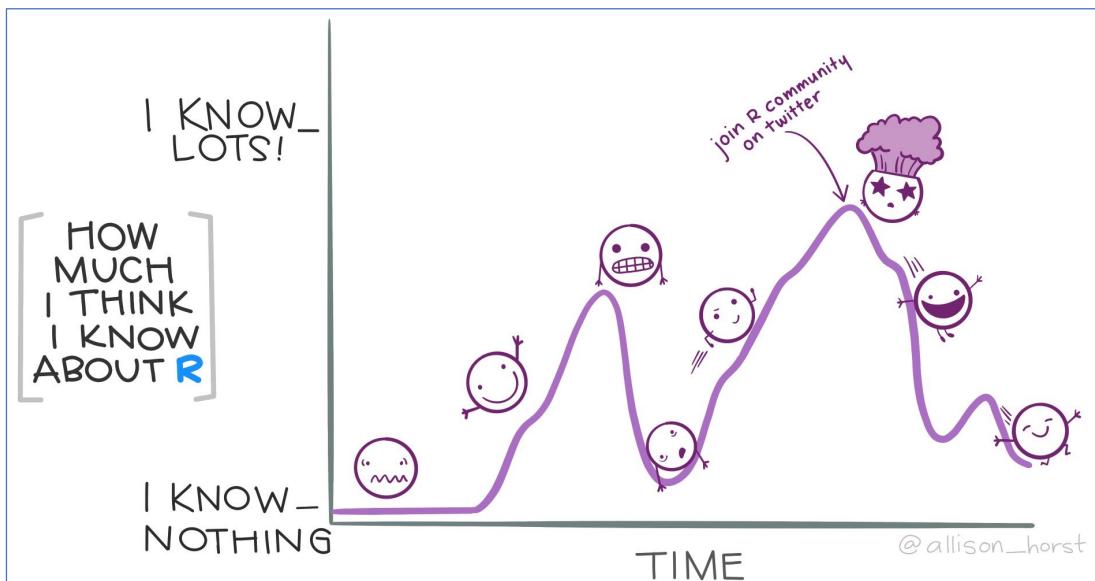


Link: <http://www.rstudio.com/download>

3

How to learn R?

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



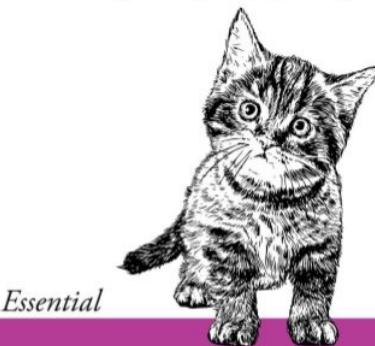
That how I learn R and C

4

How to learn R?

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

How to actually learn any new programming concept



Essential

Changing Stuff and
Seeing What Happens

O RLY?

@ThePracticalDev

The internet will make those bad words go away



Essential

Googling the
Error Message

O RLY?

The Practical Developer
@ThePracticalDev

5

Introduction to R

R in 2 hours

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



6

Introduction to R



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



7

RStudio Screen

The screenshot shows the RStudio interface with four main panels labeled 1. SOURCE, 2. CONSOLE, 3. Environment / History, and 4. Files / Plots / Packages / Help.

- 1. SOURCE:** This panel contains a code editor where you can write R code. A callout box points to the "Run" button with the text: "Click 'Run' to send your code to the console".

This is where you write your code!
Your code will not be evaluated until you "Run" them to the console.
- 2. CONSOLE:** This panel shows the R command-line interface. A callout box points to the text: "This is where your code from the Source is evaluated by R."

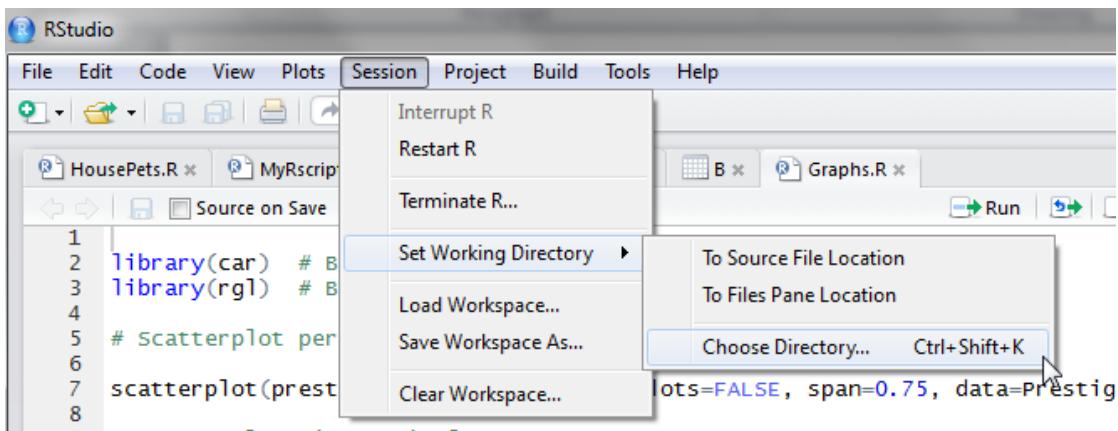
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-pc-linux-gnu (32-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but WITHOUT ANY GUARANTEES.
R is a collaborative effort with many contributors.
Type 'contributors()' for more information.
Type 'citation()' for writing papers using R.
Type 'demo()' for some basic examples.
Type 'help.start()' to view help browser.
Type 'q()' to exit R.
[Workspace locked]
> |

This is where your code from the Source is evaluated by R.

You can also use the console to perform quick calculations that you don't need to save
- 3. Environment / History:** This panel shows the R environment browser and history. A callout box points to the text: "Here you can see what objects are in your working space (Environment) or view your command history (History)".
- 4. Files / Plots / Packages / Help:** This panel shows the R file browser, plots, packages, and help. A callout box points to the text: "Here you can see file directories, view plots, see your packages, and access R Help".

8

Change Working Directory



Or alternatively you can type the command

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

9

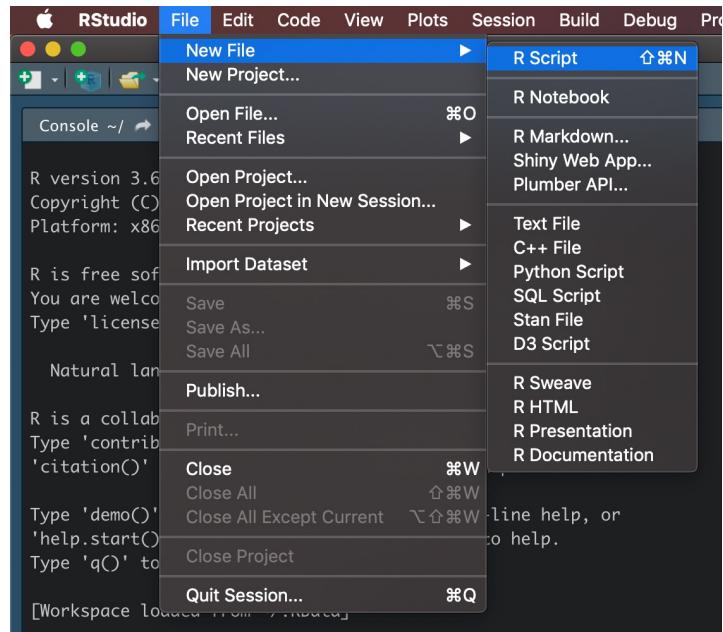
Set Default Working Directory

A screenshot of the RStudio interface. The menu bar at the top includes File, Edit, Code, View, Plots, Session, Project, Build, Tools, and Help. The 'Tools' menu is currently active, showing options like Import Dataset, Install Packages..., Check for Package Updates..., Version Control, Shell..., and Options... (which is highlighted with a blue box). Below the menu, a console window shows R version information and a message about redistributing the software. On the right side, an 'Options' dialog box is open, specifically the 'General' tab. This tab contains settings for the default working directory (set to 'H:/MyData/RFiles'), restore behavior, save workspace options, and text encoding. A callout box with a blue border and white text is overlaid on the left side of the dialog, stating: '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

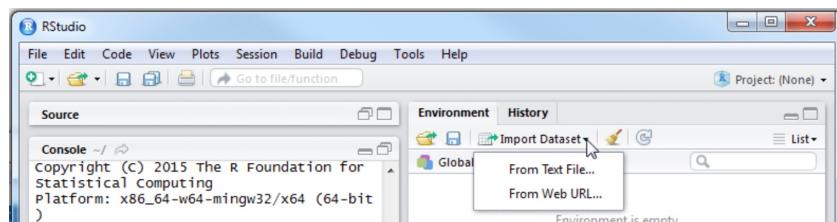


11

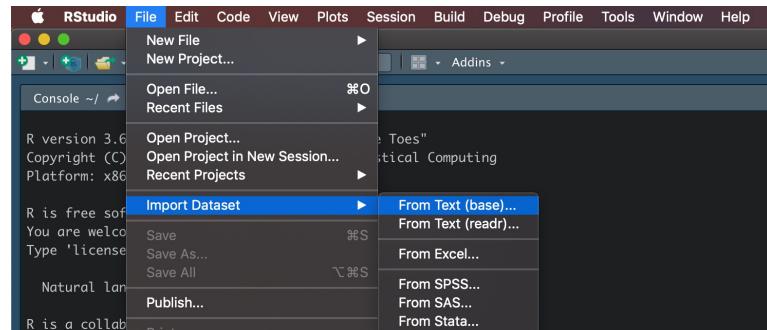
Load Data vis R Studio Menu Items

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



12

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" )
```

13

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" )
```

14

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)
```

15

Basic Commands

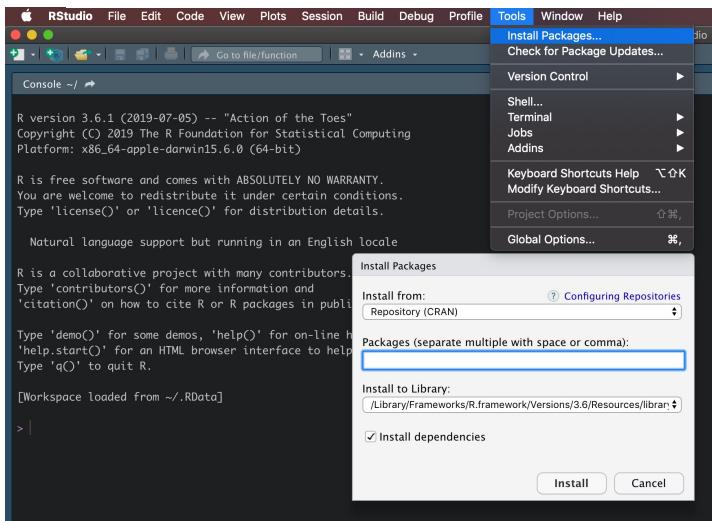
R uses functions to perform operations

```
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
```

16

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
```

17

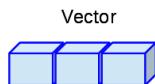
What are the basic datatypes?

```
# Declare variables of different types  
  
my_numeric = 42  
  
my_character = "universe"  
  
my_logical = FALSE  
  
# Note that R is case sensitive!
```

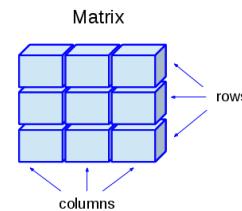
18

Data types in R

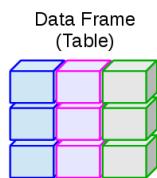
- Vector



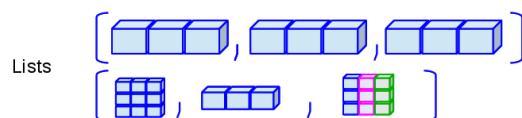
- Matrix



- Data frame



- List



19

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" )
```

20

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
```

21

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 )
```

22

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)
```

23

More info about R

- Free Online Course at DataCamp:

<https://www.datacamp.com/courses/free-introduction-to-r>

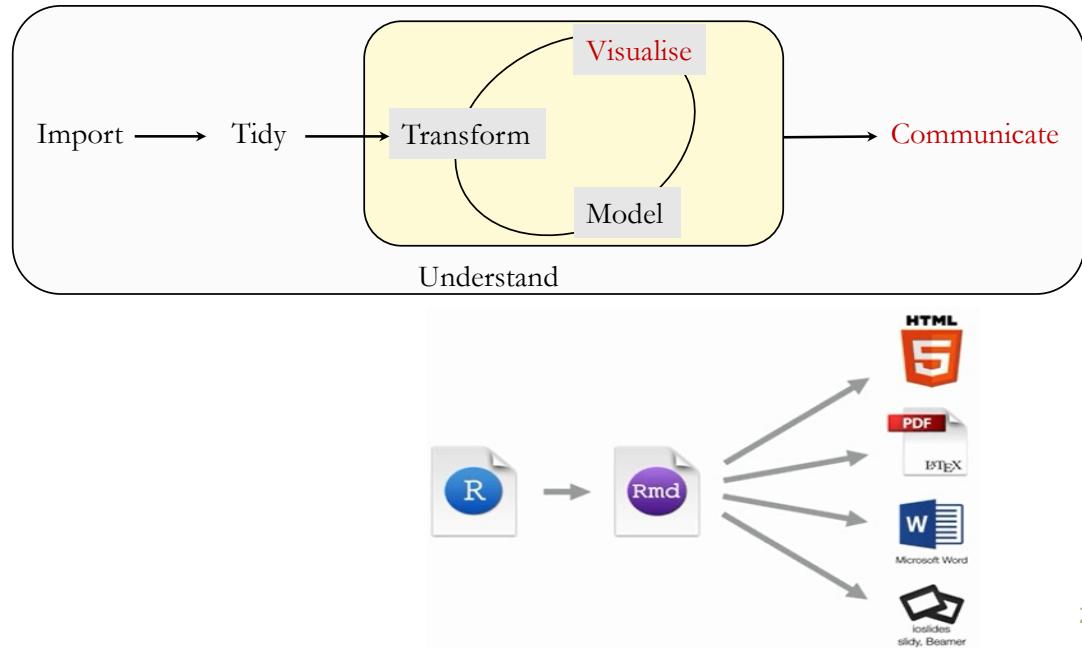
- YouTube - R Programming Tutorial:

https://youtu.be/_V8eKsto3Ug

24

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.



25

R Markdown

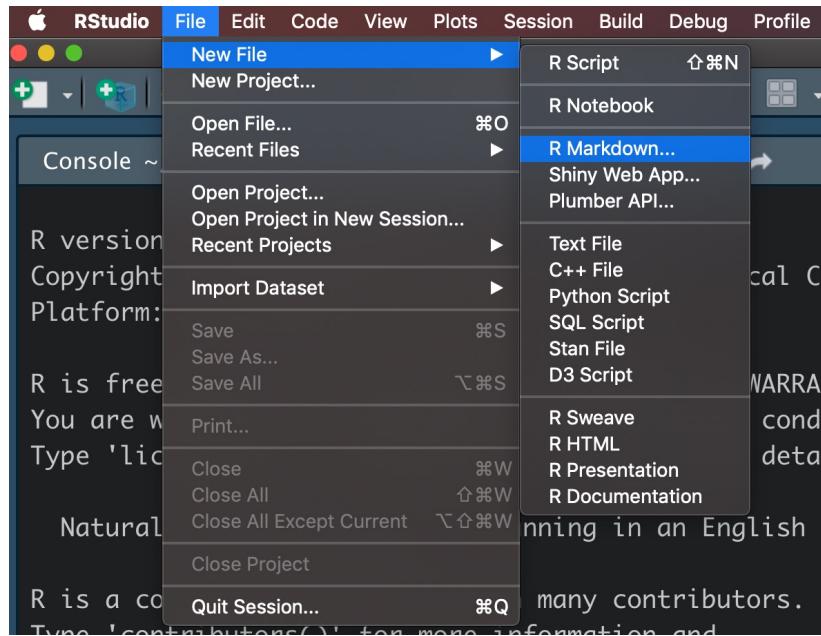
YouTube link: <https://youtu.be/DNS7i2m4sB0>



26

How to create R Markdown?

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



27

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

```
1---  
2title: "Viridis Demo"  
3output: html_document  
4---
```

R code chunks surrounded by ```s

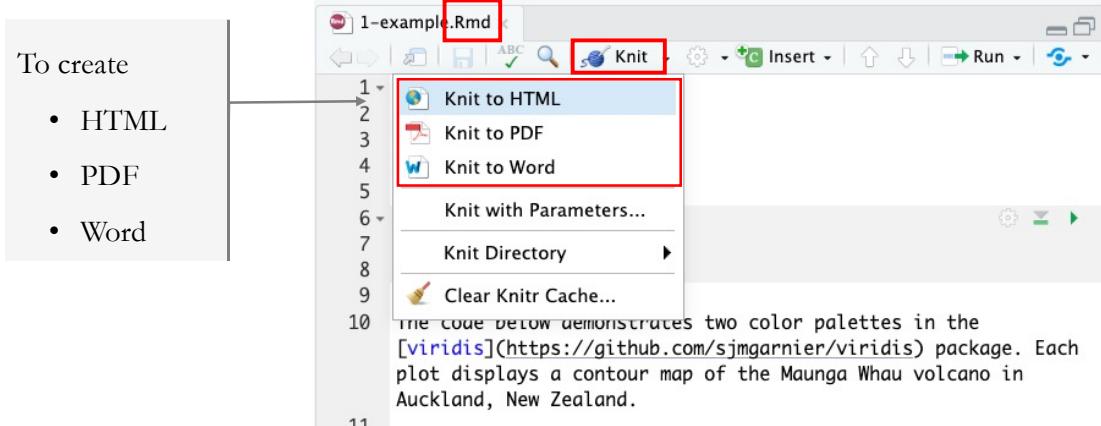
```
6```{r include = FALSE}  
7library(viridis)  
8```
```

Text mixed with simple text formatting.

```
10The code below demonstrates two color palettes in the  
[viridis](https://github.com/sjmgarner/viridis) package.  
Each plot displays a contour map of the Maunga Whau  
volcano in Auckland, New Zealand.  
11## Viridis colors  
12```{r}  
13image(volcano, col = viridis(200))  
14```  
15## Magma colors  
16```{r}  
17image(volcano, col = viridis(200, option = "A"))  
18```
```

28

How R Markdown works?

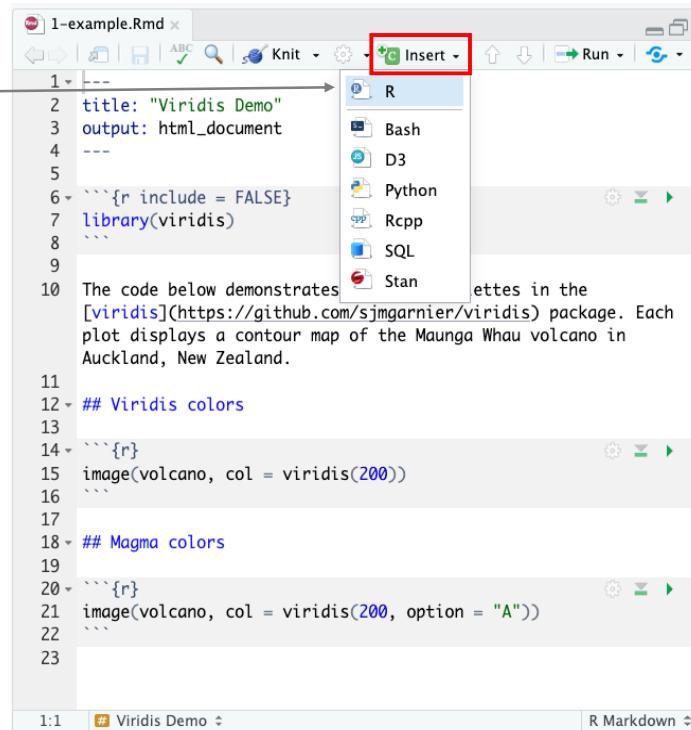


To generate a report from the .Rmd file, use the “Knit” button in the RStudio to render the file and preview the output with a single click or keyboard shortcut ($\text{⌘} \text{⌘} \text{K}$).

29

How R Markdown works?

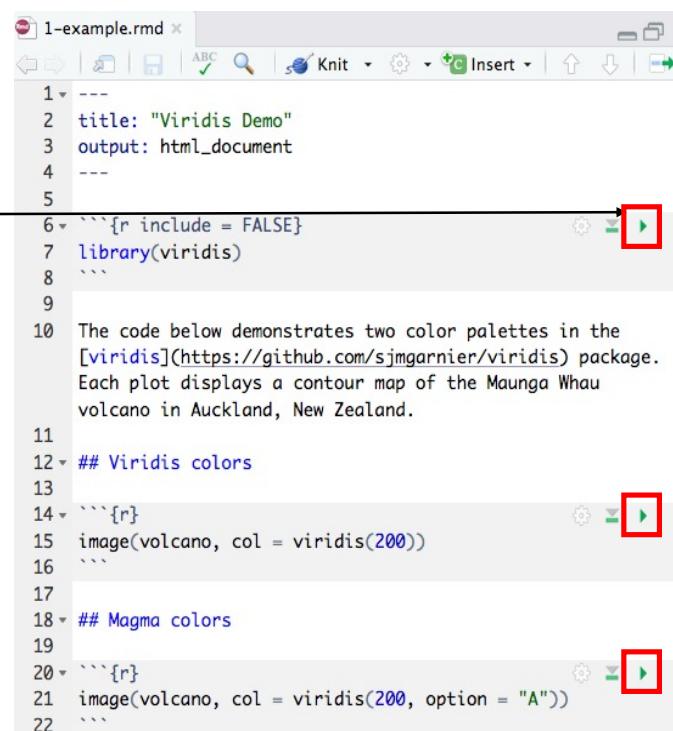
To insert code in the report. E.g. R code.



30

How R Markdown works?

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

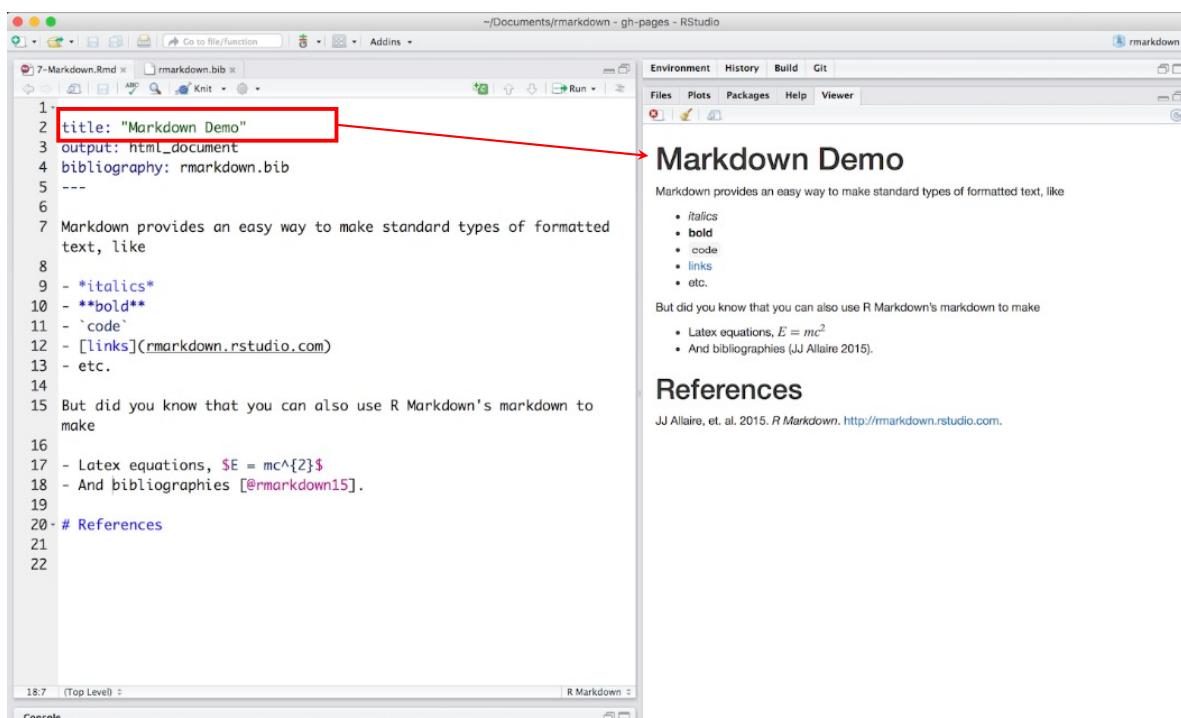


The screenshot shows an RStudio interface with an R Markdown file named "1-example.rmd". The code includes a YAML header and several R code chunks. Each code chunk has a green execution icon with a white triangle pointing right, which is highlighted with a red box. The code itself is as follows:

```
1 ---  
2 title: "Viridis Demo"  
3 output: html_document  
4 ---  
5  
6 ```{r include = FALSE}  
7 library(viridis)  
8 ```  
9  
10 The code below demonstrates two color palettes in the  
[viridis](https://github.com/sjmgarnier/viridis) package.  
Each plot displays a contour map of the Maunga Whau  
volcano in Auckland, New Zealand.  
11  
12 ## Viridis colors  
13  
14 ```{r}  
15 image(volcano, col = viridis(200))  
16 ```  
17  
18 ## Magma colors  
19  
20 ```{r}  
21 image(volcano, col = viridis(200, option = "A"))  
22 ``
```

31

Rmd syntax



The screenshot shows an RStudio interface with an R Markdown file named "7-Markdown.Rmd". A red box highlights the "title: 'Markdown Demo'" line in the YAML header. A red arrow points from this line to the rendered output on the right. The rendered output shows a section titled "Markdown Demo" with a list of features and a "References" section. The code in the file is as follows:

```
1  
2 title: "Markdown Demo"  
3 output: html_document  
4 bibliography: rmarkdown.bib  
5 ---  
6  
7 Markdown provides an easy way to make standard types of formatted  
text, like  
8  
9 - *italics*  
10 - **bold**  
11 - `code`  
12 - [links](rmarkdown.rstudio.com)  
13 - etc.  
14  
15 But did you know that you can also use R Markdown's markdown to make  
16  
17 - Latex equations, $E = mc^2$  
18 - And bibliographies [@rmarkdown15].  
19  
20 # References  
21  
22
```

32

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
```

33

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.

34

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.

35

## 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
...```

```

36

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.

37

Table of chunk options

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

38

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.

39

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>

40