

Introduction to R and Data Science

Bhavesh Shah, Scientist (Environment), V&A

And student on MSc Data Science Course at UCL



b.shah@vam.ac.uk

#IconMMN



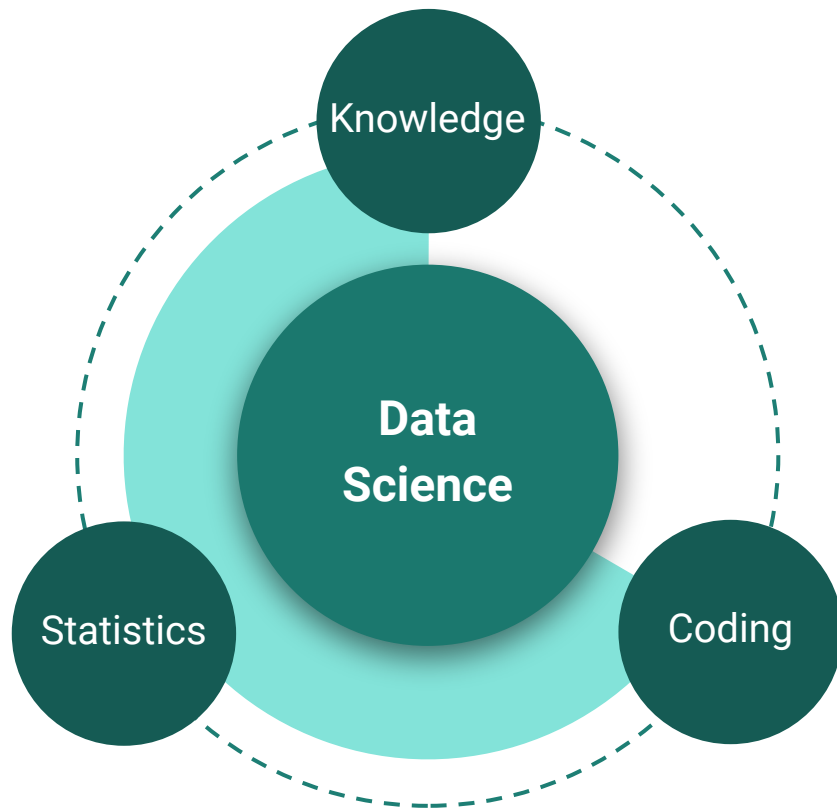
Data science

Breakthroughs in technology, computing and algorithms

And lots and lots of data...

New kid on the block?

Or just research using a computer?



Why code?

Let's try to answer a different question:

If you have (managed to find) flour and eggs,
would you rather have...



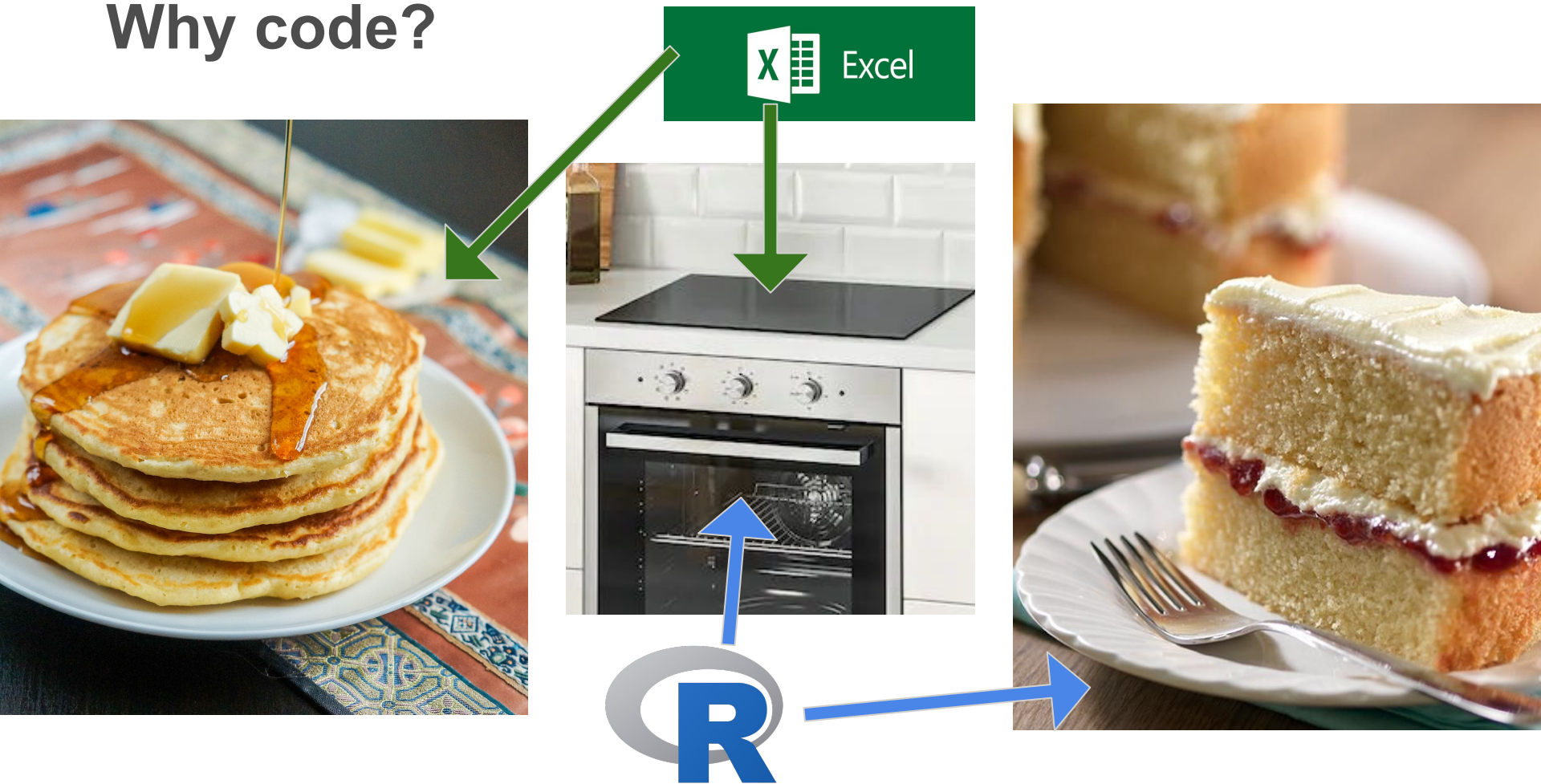
Why code?



Pancake
or
cake?



Why code?



Coding

- 1) You write the recipes (code) and prepare the ingredients (data)
- 2) R does the baking - you cross your fingers
- 3) Add the finishing touches - and serve!



The data ingredients

Understanding data is the most important concept:

- Data types
- Data structures
- Data grammar
- How data interacts
- Volume, Veracity, Variety



Understanding the ingredients

Alphabet - the building blocks

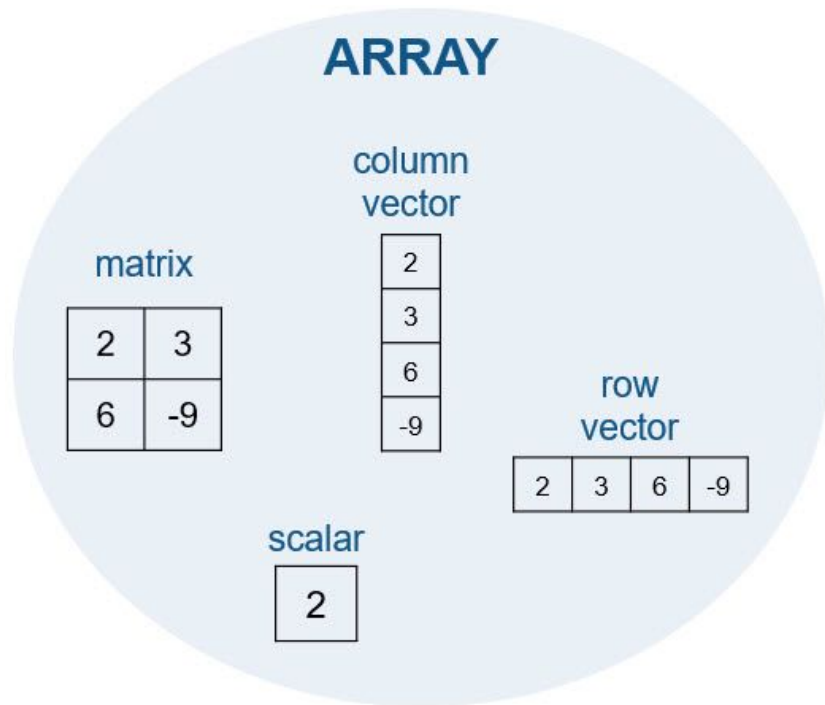
Nouns - also known as 'objects'

Verbs - the 'recipes' and the bits you code

Alphabet - numbers, characters, dates,
missing, logical,

Sentences - vectors, matrices, arrays, lists,
data frames

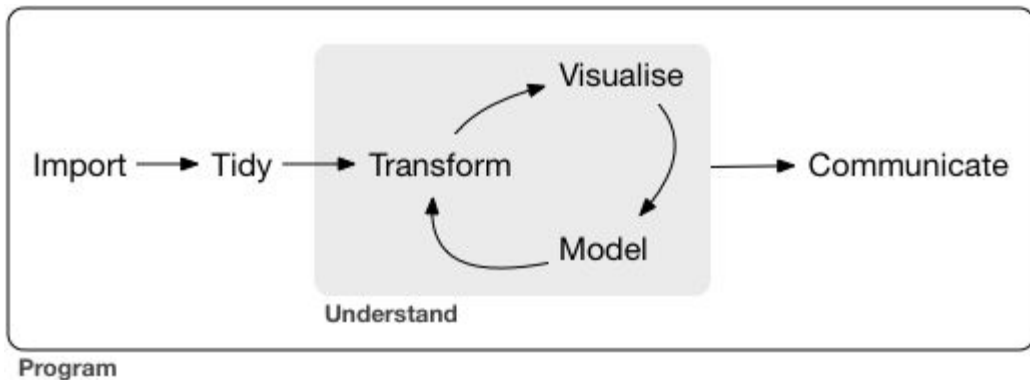
Spelling and grammar - are important



Writing recipes

Decide what you want to do, who is your audience (your client) and work backwards

Be flexible and adaptive - as you'll likely end up somewhere else



Icing on the cake

Final pun

Modelling

Machine learning

Graphing

Reporting

Interactives

Online



Icing on the cake

Final pun

Modelling

Machine learning

Graphing

Reporting

Interactives

Online

Statistics

Known data
Explanatory power

Machine learning

Unseen data
Predictive power

Report



THE R GRAPH
GALLERY

V&A Environmental Report

April 2019 – Winter conditions

South Kensington

Temperature (15-25°C, winter)

Relative Humidity (35-65%rh, annual)

Red: out of specification >30% of the time

Green: in spec ≥70% of the time

Grey: N/A (no objects or FuturePlan)

-* is <30%rh

+* is >70%rh

+** is >70%rh sustained for over 3 days, potentially leading to a risk of mould growth



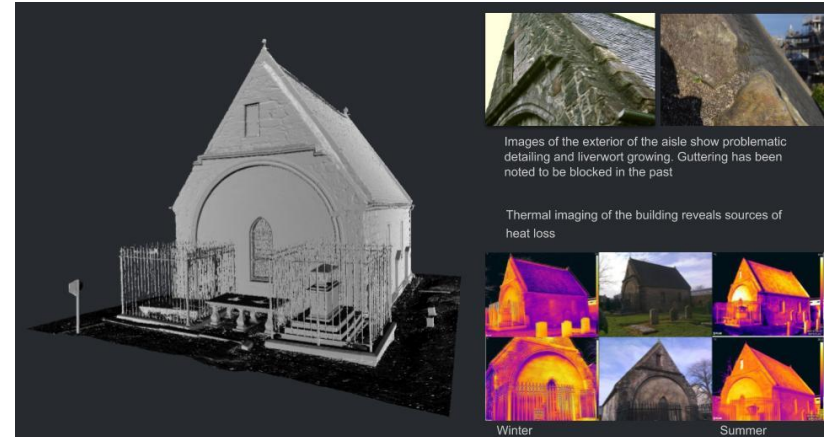
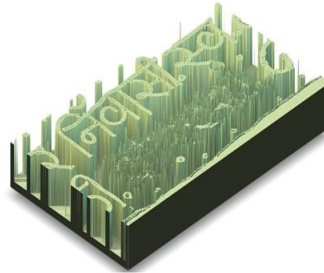
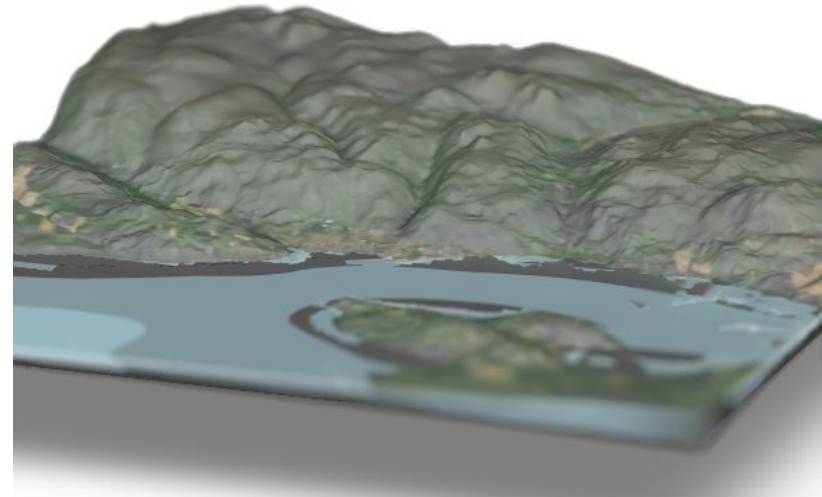
Icing on the cake

Complex - collaborative - multi-disciplinary projects

Insights

Storytelling

State-of-art tools



Pros and cons

Free

Community support

10,000 packages

Development

Transferable and high demand skill (and it pays)

The future

Hard

Need to keep learning

Coding 'rage'

Tips

Give yourself plenty of time

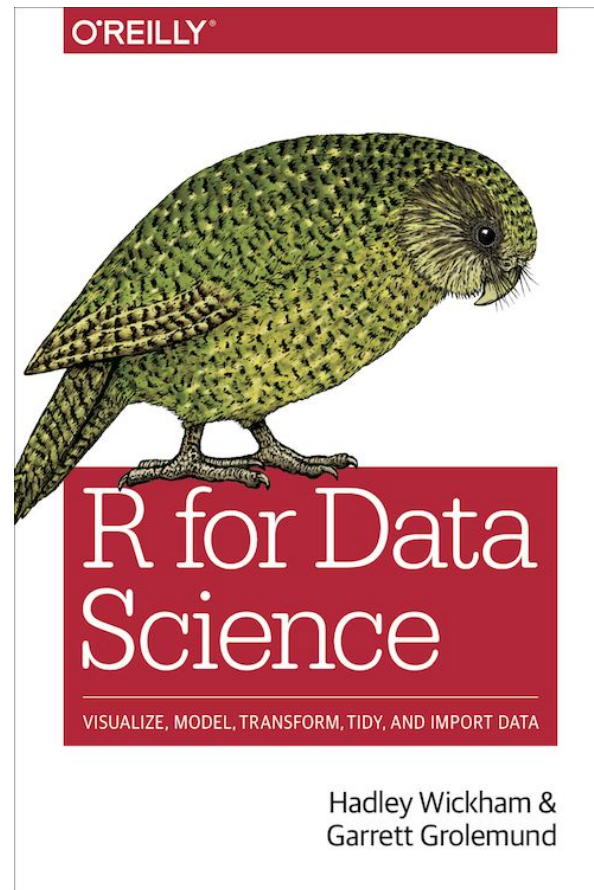
Think of your future self - Comment as you go along

Seek help before code rage!

StackOverflow, Blogs, Twitter, Meetups, User Groups
(LondonR), Cheat Sheets, YouTube, Kids coding books
-these are your friends

Good place to start: <https://r4ds.had.co.nz/>

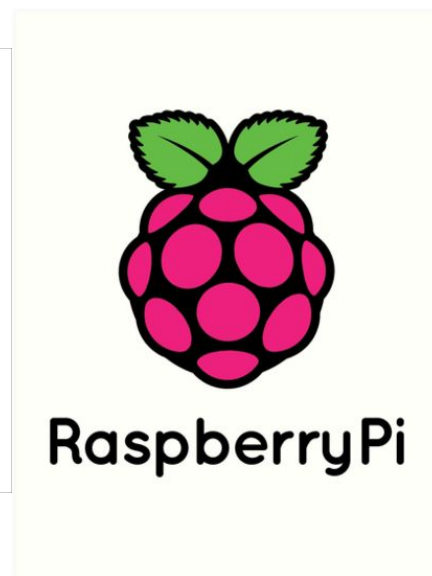
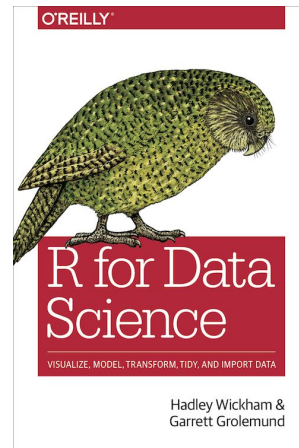
Google: “**R for data science Hadley**”



Follow instructions at RStudio for
how to download and get started:



<https://www.vam.ac.uk/blog/author/bhaves-shah>



File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Training

R.R x Training.Rmd x
Source on Save Run Source

1

Don't panic!

1:1 (Top Level)

R Script

Console Terminal x Jobs x

~/ALL SENSORS/Training/

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.

> |

Environment History Connections Git

Import Dataset

List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > ALL SENSORS > Training

	Name	Size	Modified
↑	..		
■	.gitignore	44 B	May 27, 2020, 4:29 PM
■	.Rhistory	0 B	May 28, 2020, 11:56 AM
■	R.R	0 B	May 27, 2020, 11:24 PM
■	Training.html	628.9 KB	May 27, 2020, 10:54 PM
■	Training.Rmd	880 B	May 27, 2020, 10:54 PM
■	Training.Rproj	218 B	May 28, 2020, 10:13 PM



**Ingredients
(Data)**



Outputs (and Help)

PM
6 AM
4 PM
4 PM
4 PM
3 PM



```
R is free
You are we
Type 'lic
Natural
R is a co
Type 'cont
'citation
Type 'demo
'help.start
Type 'q()
> |
```

~ / ALL SENSORS / Training - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Run Source

Environment History Connections Git

Global Environment

Environment is empty

<https://oceanonline.shinyapps.io/MuseumClimateApp/>

&

1:1 (Top Level) R Script

Console Terminal Jobs

~/ALL SENSORS/Training/

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 trademark of its proprietor. Use and distribution of R is governed by the R Project for Statistical Computing.
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.

> |

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > ALL SENSORS > Training

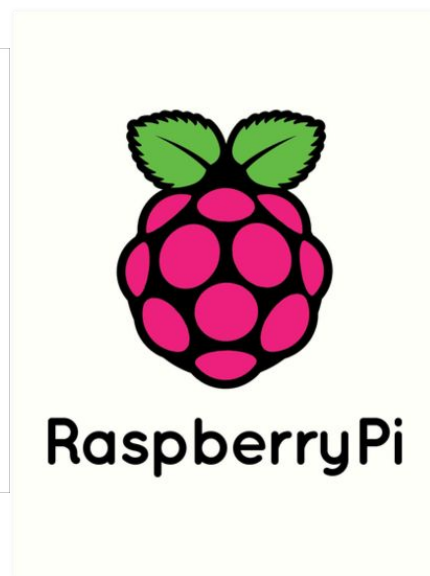
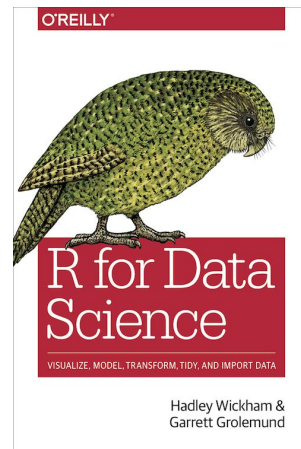
	Name	Size	Modified
	..		
	.gitignore	44 B	May 27, 2020, 4:29 PM
	.Rhistory	0 B	May 28, 2020, 11:56 AM
	R.R	0 B	May 27, 2020, 11:24 PM
	Training.html	628.9 KB	May 27, 2020, 10:54 PM
	Training.Rmd	880 B	May 27, 2020, 10:54 PM
	Training.Rproj	218 B	May 28, 2020, 10:13 PM

https://oceanonline.shinyapps.io/Demo_RmarkdownShiny/

Follow instructions at RStudio for
how to download and get started:



<https://www.vam.ac.uk/blog/author/bhaves-shah>



Thankyou

<https://github.com/BhavShah01/Training>

b.shah@vam.ac.uk

[@bhav_shah](#)

V&A