An Introduction to Data Analysis using R

Dr. Calum Webb
Sheffield Methods Institute, the University of Sheffield
c.j.webb@sheffield.ac.uk

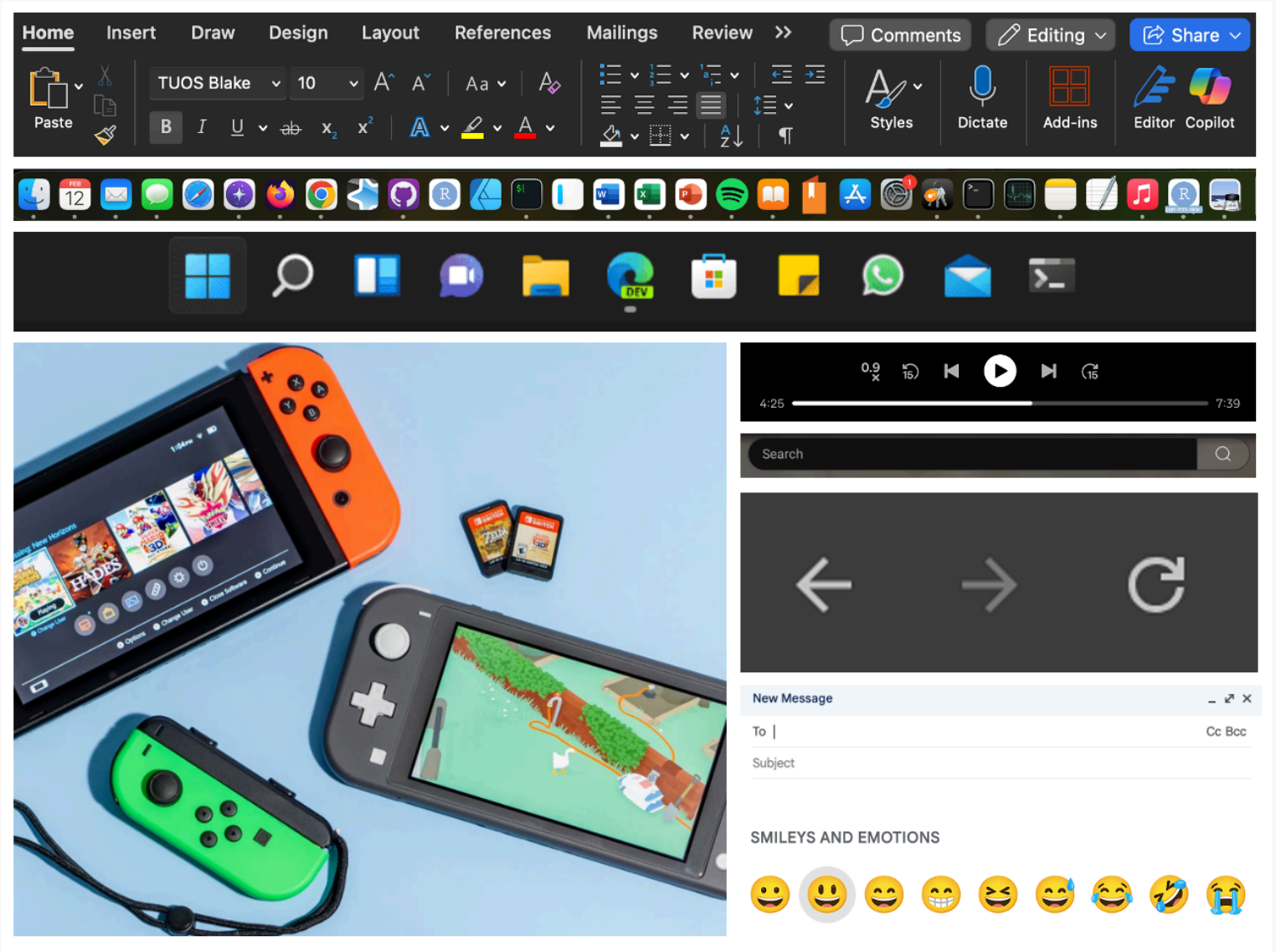
This is going to be a hands-on training course: all of the information that you need is included in the materials, rather than being split across the slides and the exercises.

R compared to software for Data Analysis

- **R** is a programming language, rather than just a piece of software.
- **R** has a central repository that holds packages that can expand what you are capable of doing with it.
- **R** is free.
- **R** has a very supportive and active community around it.
- **R** can do pretty much everything that any commercial software package can do, with a few very niche exceptions..

Artwork by Allison Horst

"But I'm not a programmer, I never studied computer science/mathematics/intro to programming before"

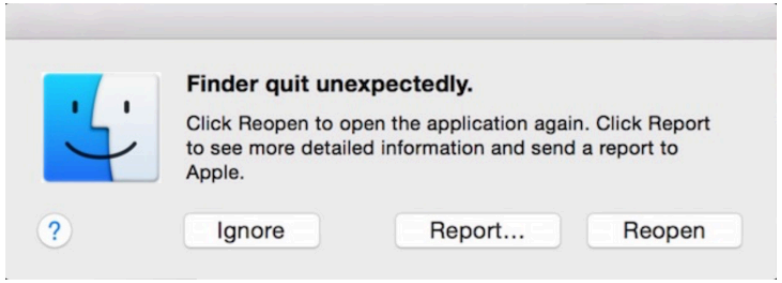

Home
Insert
Draw
Design
Layout
References
Mailings
Review
Comments
Editing
Share
Paste
TUOS Blake
10
Styles
Dictate
Add-ins
Editor
Copilot
4:25
7:39
Search
New Message
To
Cc Bcc
Subject
SMILEYS AND EMOTIONS

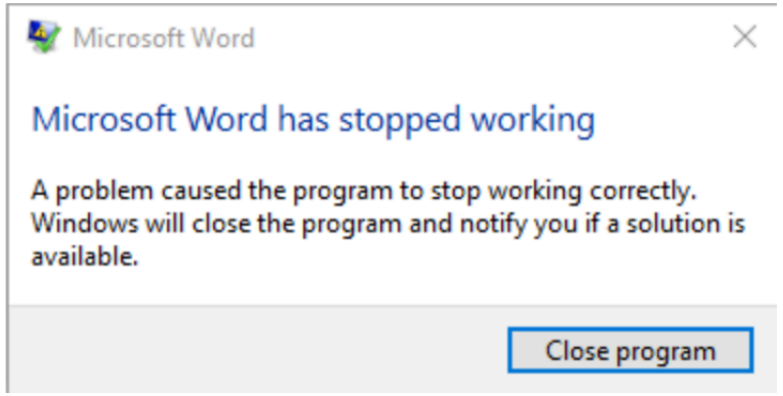
If you can use a computer, you can program.

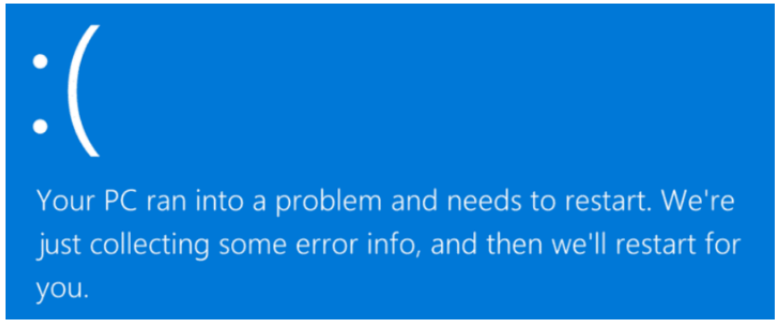
- It's not difficult, it's different. Getting used to something different takes time.
- Instead of clicking virtual or real buttons that apply functions (i.e. make things happen), we give commands by writing text.
- They are called **languages** for a reason — they have their own rules, grammar, and syntax — you have to practice them to get good at them.
- **But**, that doesn't mean you can't use phrasebooks, especially if you only use the language occasionally.

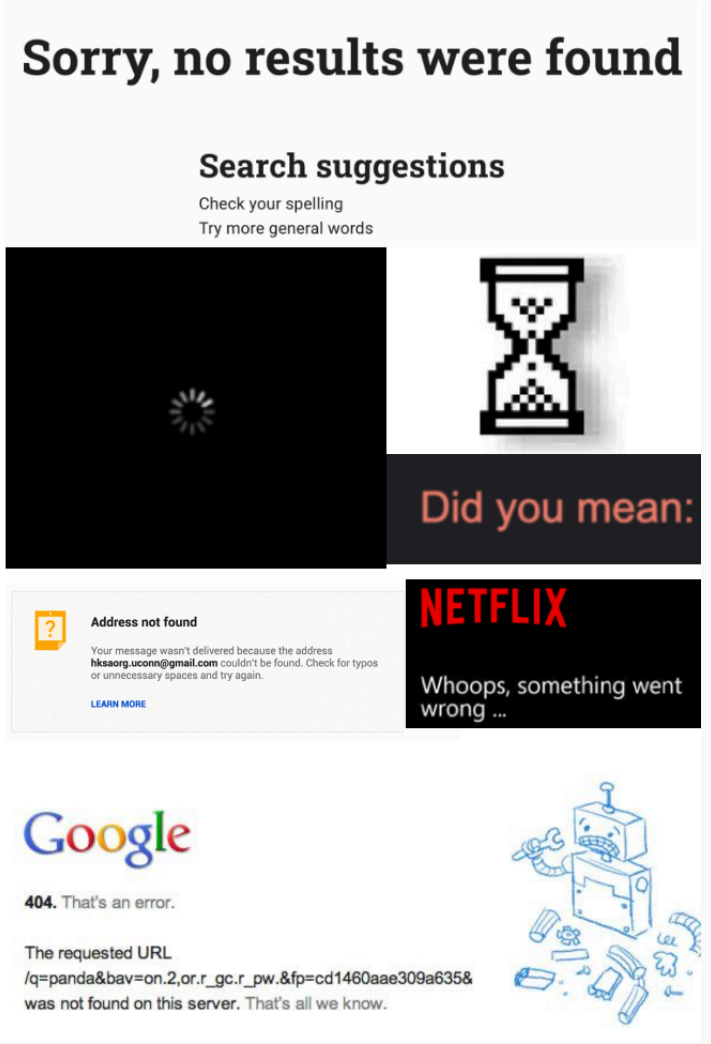
```
open.spotify()

search.playlists(search = "instrumental")

play.playlist(playlist = "Instrumental Studying
Mix")

skip.track()

rewind.track(seconds = 20)

add.track.to.favourites()

stop()

quit.spotify()
```

```
> read.csv("testdat.csv")
Error in file(file, "rt") : cannot open the connection
In addition: Warning messages:
1: In prose_index(x) : Code fences are not balanced
2: In prose_index(x) : Code fences are not balanced
3: In file(file, "rt") :
  cannot open file 'testdat.csv': No such file or directory
>
```


Finder quit unexpectedly.
Click Reopen to open the application again. Click Report to see more detailed information and send a report to Apple.
Ignore
Report...
Reopen


Microsoft Word
Microsoft Word has stopped working
A problem caused the program to stop working correctly. Windows will close the program and notify you if a solution is available.
Close program


:(
Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you.


Sorry, no results were found
Search suggestions
Check your spelling
Try more general words
Did you mean:
Address not found
Your message wasn't delivered because the address hksaorg.uconn@gmail.com couldn't be found. Check for typos or unnecessary spaces and try again.
LEARN MORE
NETFLIX
Whoops, something went wrong ...
Google
404. That's an error.
The requested URL /q=panda&bav=on.2,or.r_gc.r_pw.&fp=cd1460aae309a635& was not found on this server. That's all we know.

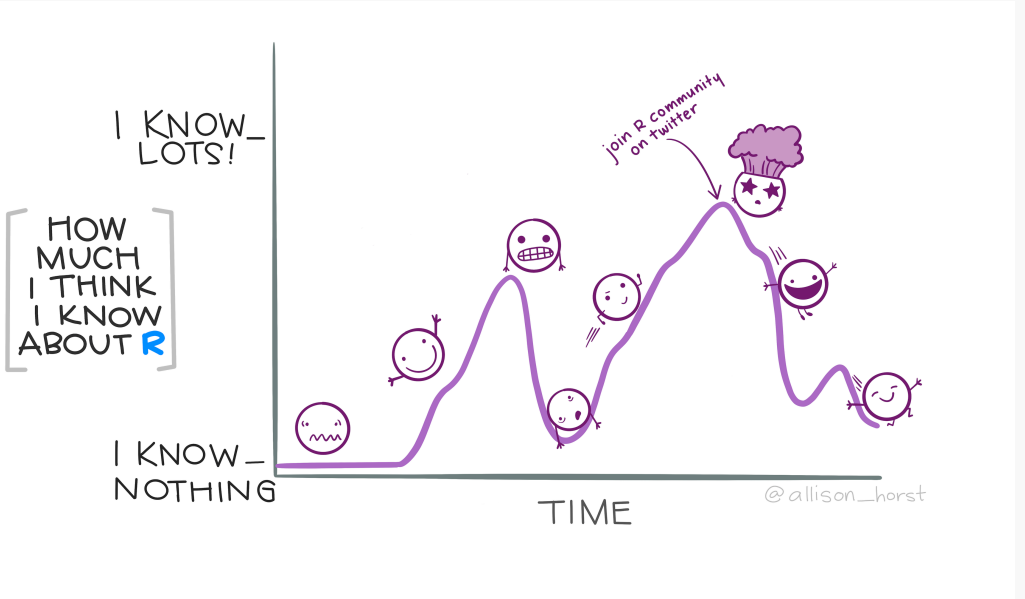
We get errors all the time when using computers. The only difference between the errors you just saw and R errors is that the R errors are at least trying to tell you what went wrong.

If you've ever used the "back" ‹ or "undo" ↺ button, change something, and then do it again to get the result you want — congratulations, you've corrected an error.

What we're going to cover today:

- A general rule for using code in R: things you have to change and things that stay the same.
- Basic univariate and bivariate descriptive statistics using Base R
- Basic data visualisations using base R
- Common bivariate inferential statistics using base R
- Installing additional packages
- Reading datasets into R, including SPSS datasets
- Descriptive statistics using `tidyverse` and `janitor` tools
- Data visualisations using `ggplot2`

This is your phrasebook to help you on your R learning journey.



Artwork by Allison Horst