

 shirleywang.rbind.io

 shirleywang@g.harvard.edu

 @shirleybwang

# Introduction to R

Day 1: Data Cleaning & Wrangling

Shirley Wang

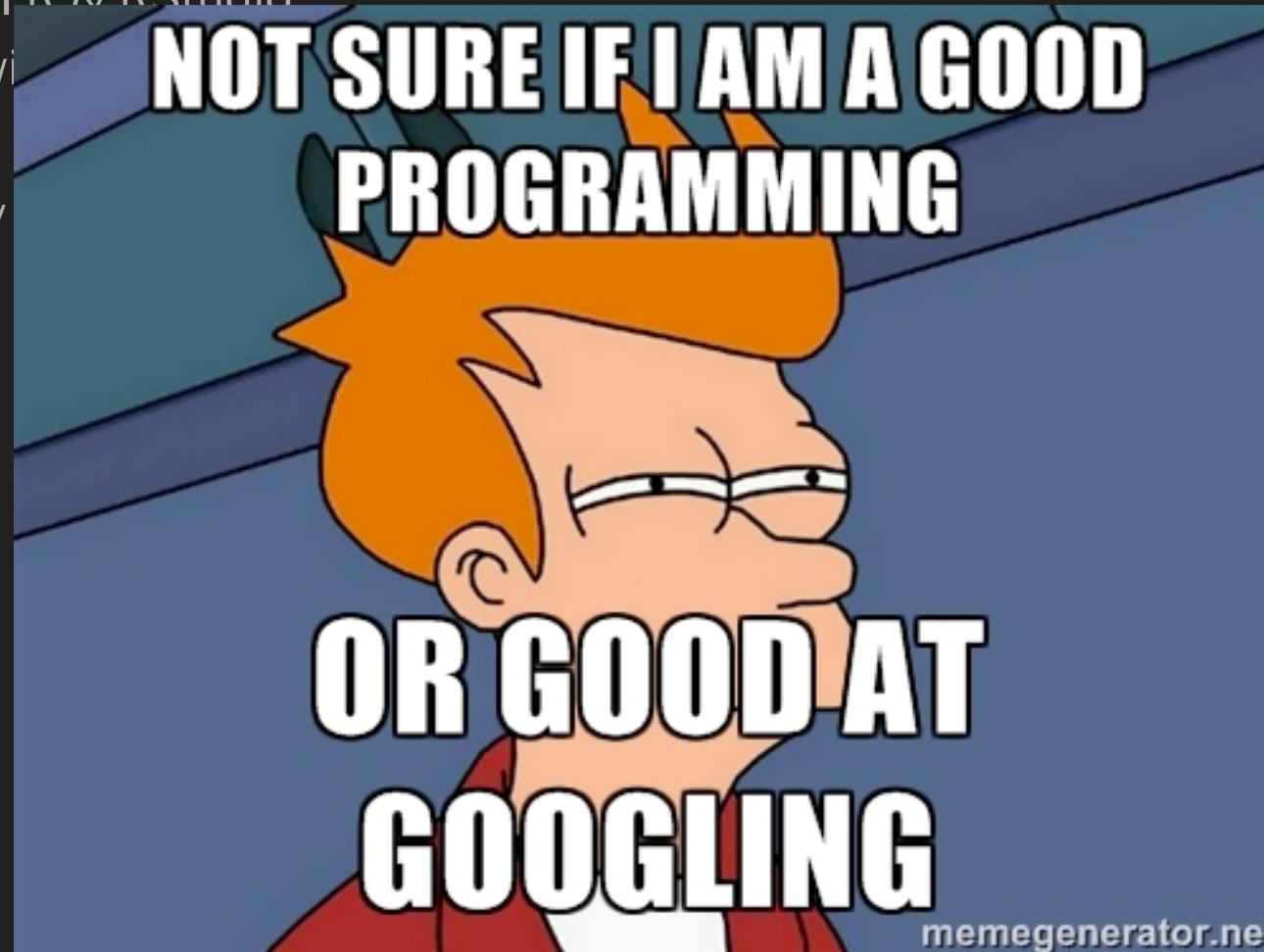
March 18, 2021

# Day 1 Goals

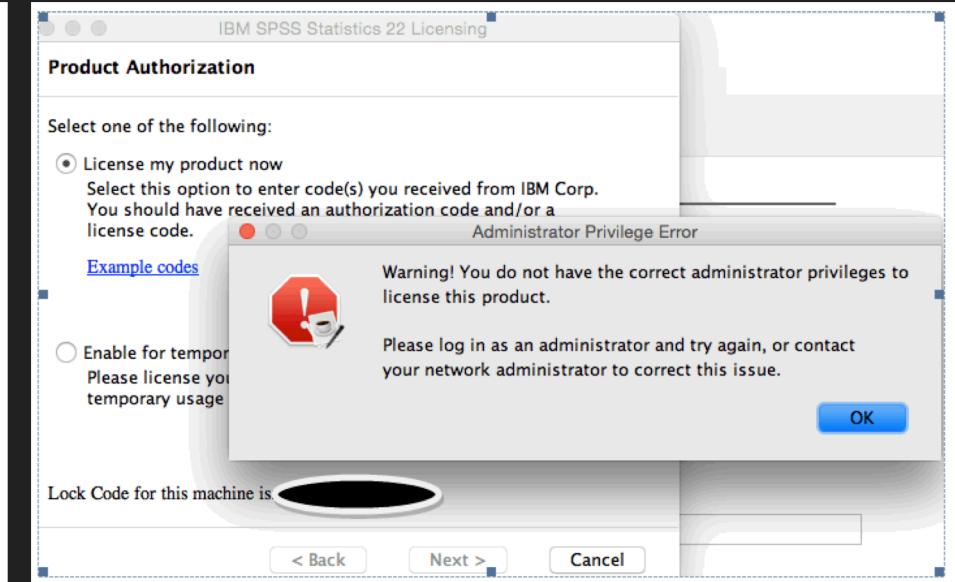
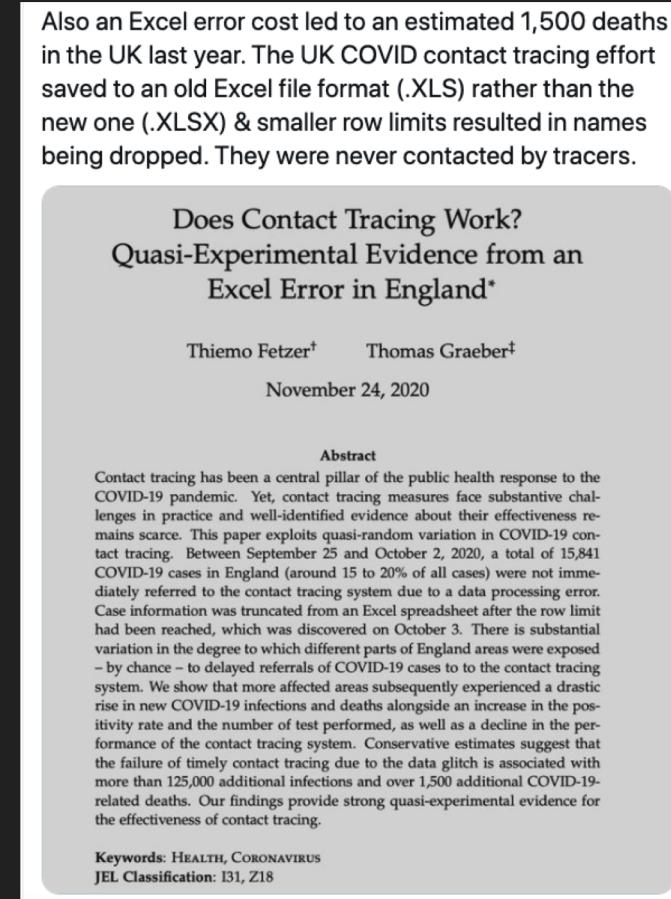
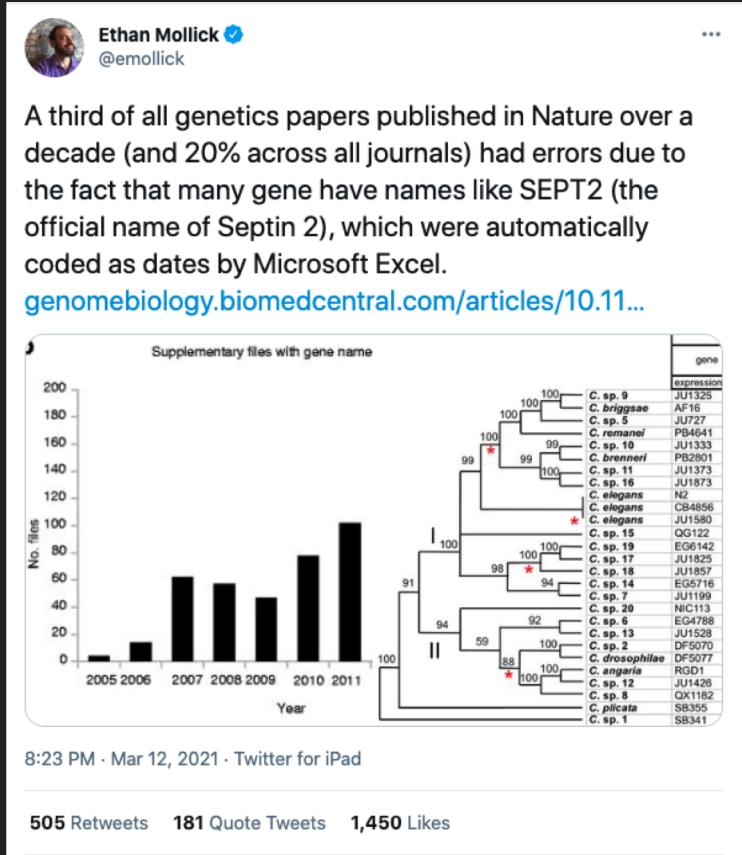
- Download & install R & RStudio
- Become familiar with the RStudio environment
- Operations
- Objects (what they are & how to create them)
- Data types
- Functions

# Day 1 Goals

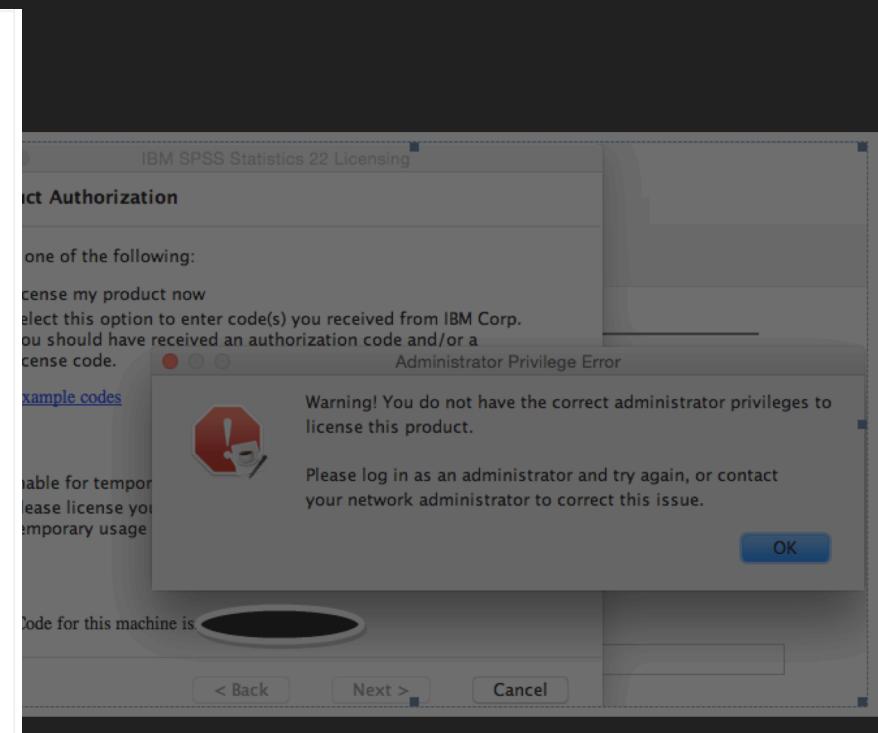
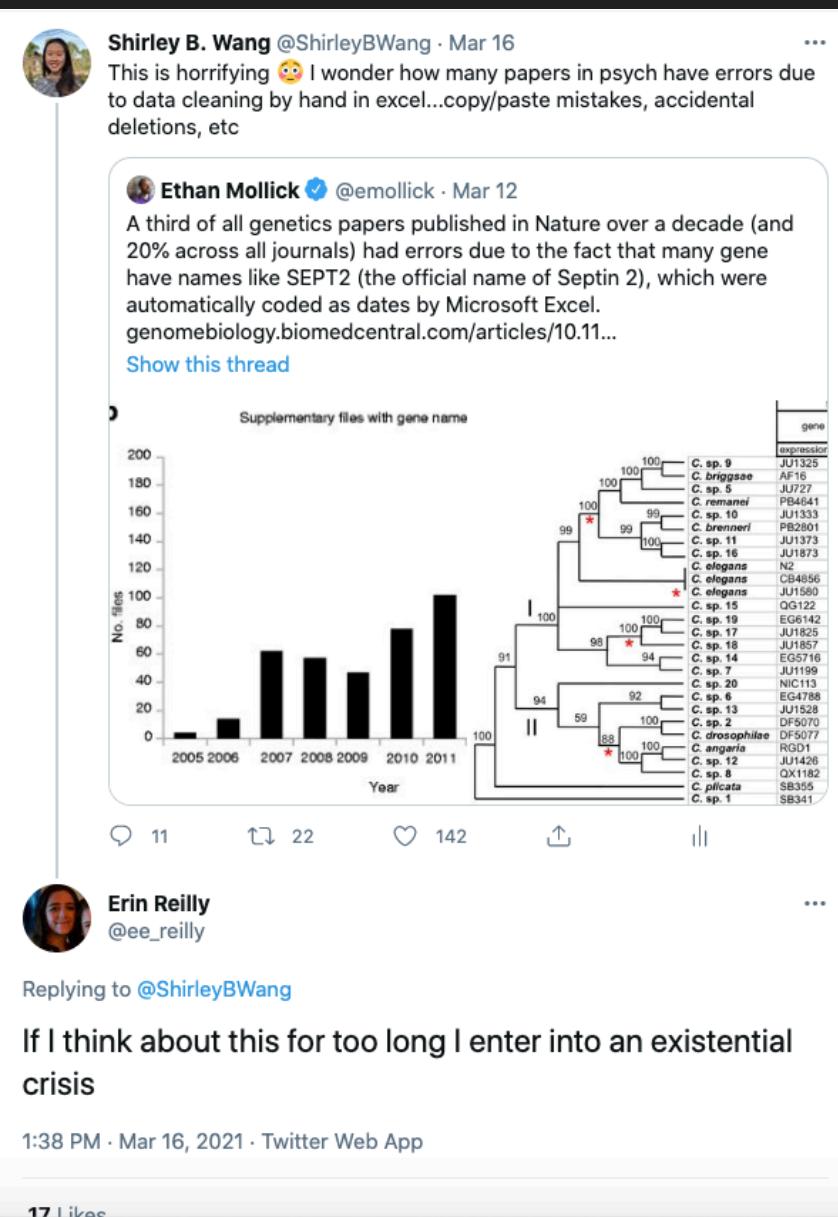
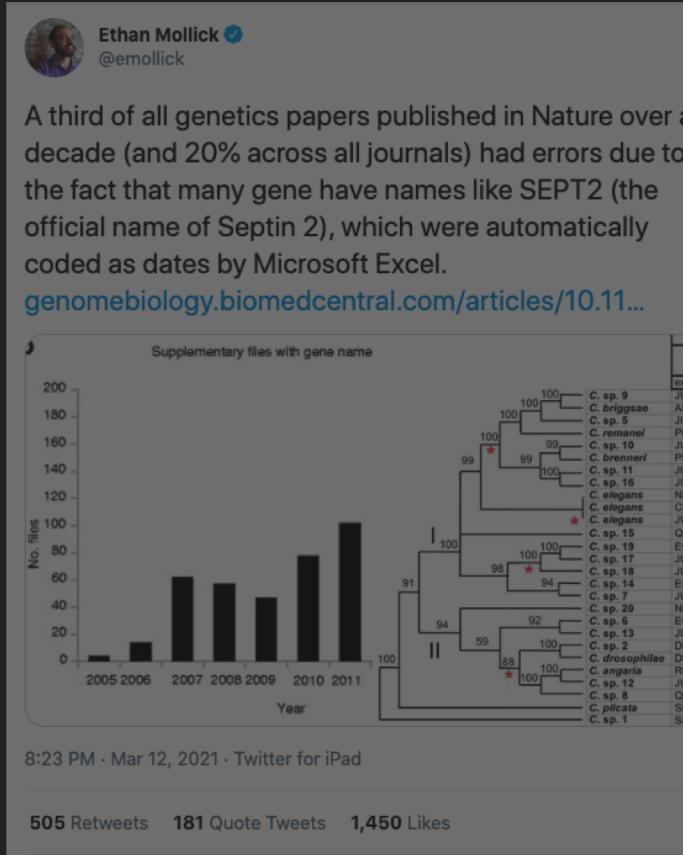
- Download & install R & RStudio
- Become familiar with:
- Operations
- Objects (what they are)
- Data types
- Functions



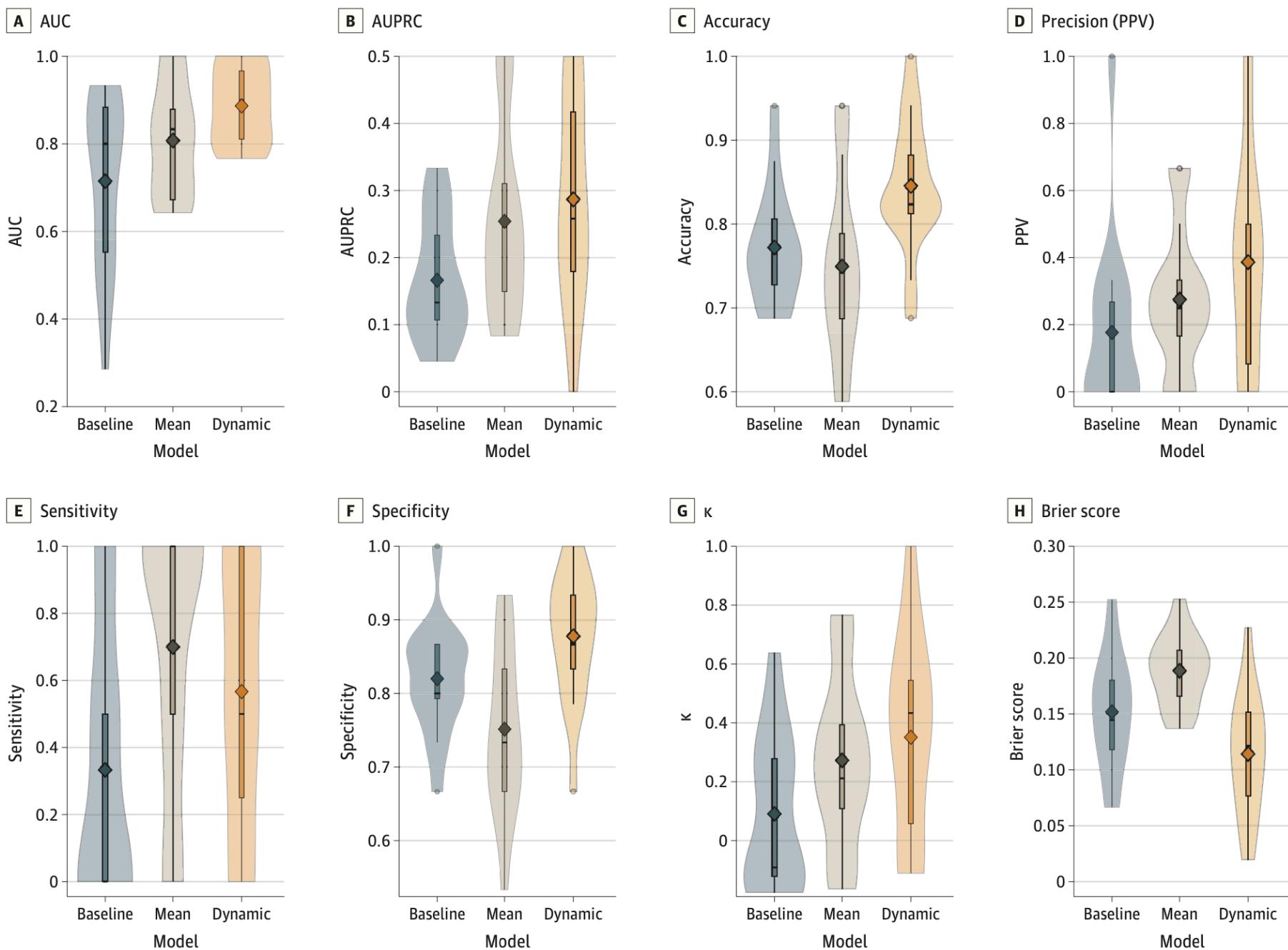
# Why use R?



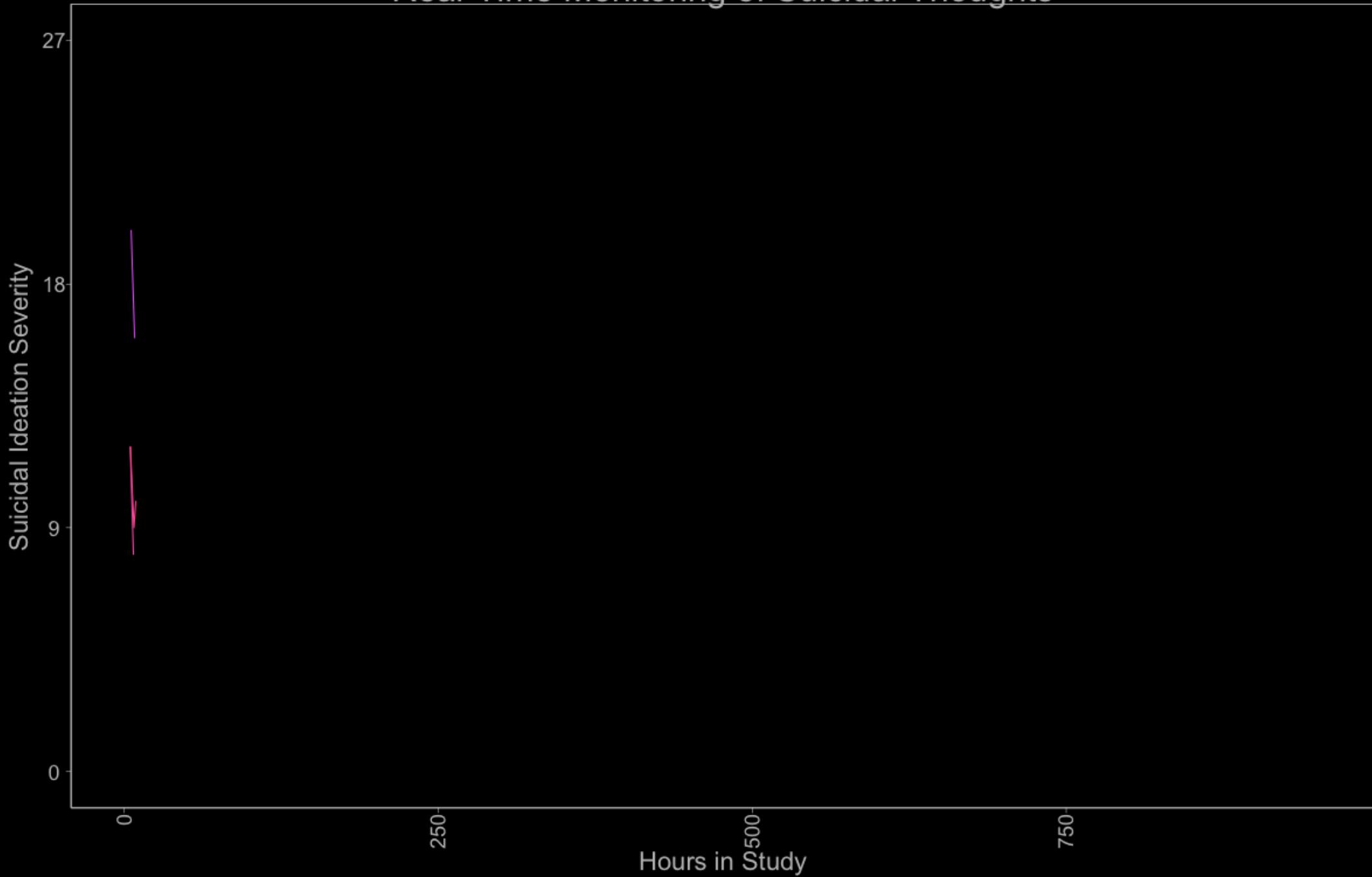
# Why use R?



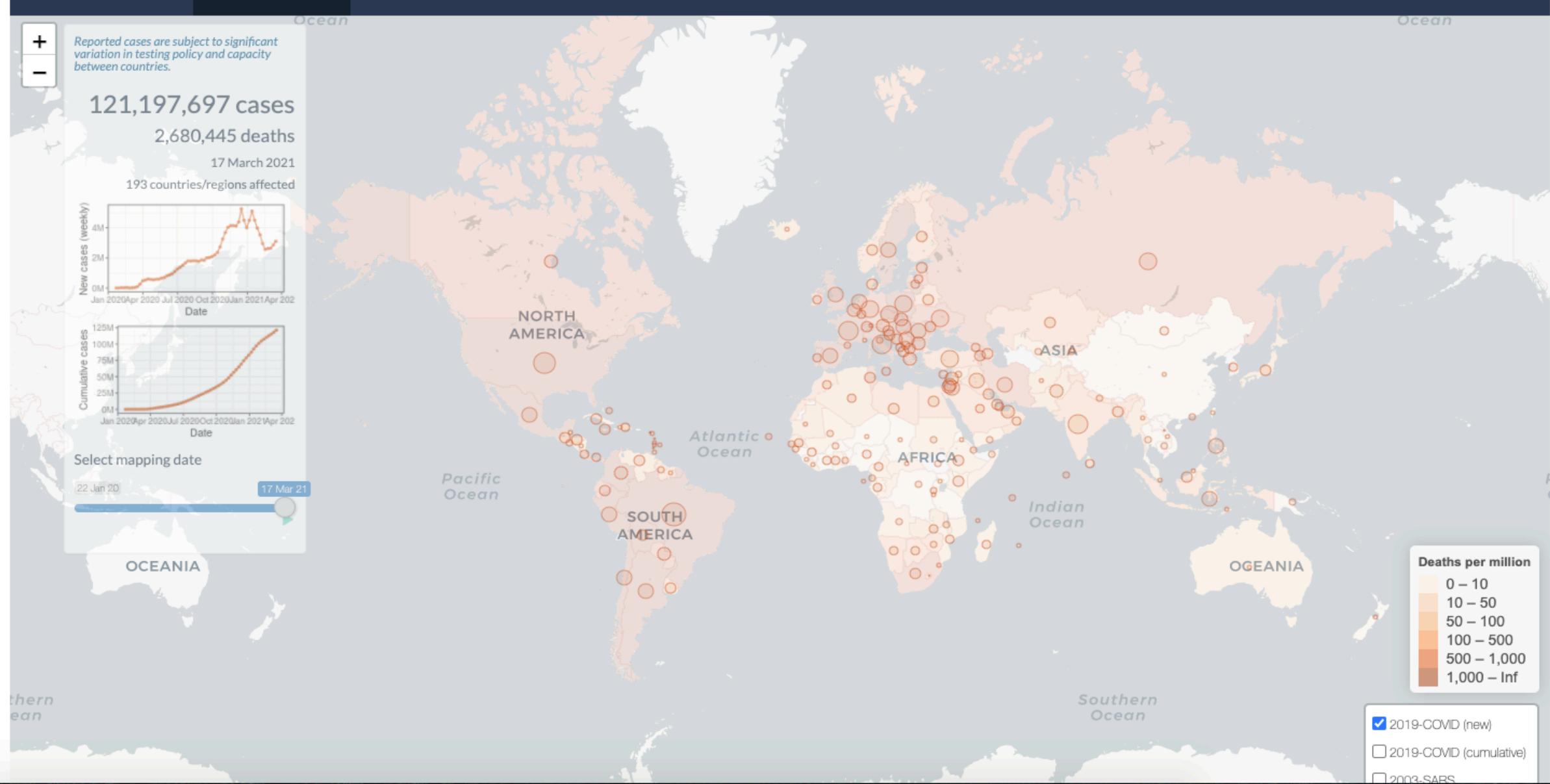
**Figure 1. Suicide Attempt Prediction Model Metrics**



# Real-Time Monitoring of Suicidal Thoughts



# covid19, epidemiology

[COVID-19 tracker](#)[COVID-19 mapper](#)[Region plots](#)[SARS mapper](#)[Outbreak comparisons](#)[Data](#)[About this site](#)

[Introduction](#)[How It Works](#)[Code Chunks](#)[Inline Code](#)[Code Languages](#)[Parameters](#)[Tables](#)[Markdown Basics](#)[Output Formats](#)[Notebooks](#)[Slide Presentations](#)[Dashboards](#)[Websites](#)[Interactive Documents](#)[Cheatsheets](#)

# Slide Presentations

R Markdown renders to four presentation formats:

- `beamer_presentation` - PDF presentations with beamer
- `ioslides_presentation` - HTML presentations with ioslides
- `slidy_presentation` - HTML presentations with slidy
- `powerpoint_presentation` - PowerPoint presentation
- `revealjs::revealjs_presentation` - HTML presentations with reveal.js

Each format will intuitively divide your content into slides, with a new slide beginning at each first or second level header.

Insert a horizontal rule ( `---` ) into your document to create a manual slide break. Create incremental bullets with `>-`, as in the .Rmd file below, which is available [here](#) on RStudio Cloud.

The screenshot shows the RStudio interface. On the left is the RStudio sidebar with various options like Introduction, How It Works, etc., and the "Slide Presentations" option is highlighted with a blue background. The main workspace shows two panes: the left pane displays the R Markdown (.Rmd) source code, and the right pane displays the resulting HTML presentation. The source code includes R code chunks and a manual slide break. The generated HTML page has a dark background with a central plot titled "PLASMA COLORS".

```
1: ---
2: title: "Viridis Presentation"
3: output:
4:   revealjs::revealjs_presentation:
5:     theme: league
6: ---
7:
8: ```{r include = FALSE}
9: knitr::opts_chunk$set(echo = FALSE)
10: library(viridis)
11: ```
12:
13: The [viridis](https://github.com/sjmgarnier/viridis) package contains four color palettes, revealed in the plots that follow.
14:
15: >- Viridis
16: >- Magma
17: >- Inferno
18: >- Plasma
19:
20: Each plot displays a contour map of the Maunga Whau volcano in Auckland, New Zealand.
21:
22: ## Viridis colors
23:
24: ```{r}
```

PLASMA COLORS



## Shirley Wang

Ph.D. candidate  
Harvard University



## About

I am a PhD candidate in clinical psychology with a secondary in computational science and engineering at Harvard University. I work with [Matt Nock](#) and am funded by the National Science Foundation Graduate Research Fellowship Program. My research examines why people engage in behaviors that are harmful to themselves, including eating disorder behaviors, nonsuicidal self-injury, and suicide. I use a wide range of methods to study these problems, including laboratory-based behavioral experiments, real-time monitoring, and large-scale longitudinal studies. I am particularly interested in using mathematical and computational modeling to formalize theories in psychopathology.

I value diversity, inclusion, and belonging, and am committed to promoting these values in research, academia, clinical practice, and beyond. I believe that increasing representation of people from marginalized and historically disadvantaged backgrounds is critical for conducting ethical, comprehensive, and innovative clinical science research.

### Interests

- eating disorders
- suicide
- computational modeling
- real-time monitoring

### Education

-  Ph.D. in Clinical Science, expected 2023  
Harvard University
-  A.M. in Clinical Science, 2019  
Harvard University

## Preface

Why read this book

Structure of the book

Software information and convention...

Acknowledgments

About the Author

## 1 Introduction

1.1 Motivation

1.2 Get started

1.3 Usage

1.4 Two rendering approaches

1.5 Some tips

## 2 Components

2.1 Markdown syntax

2.1.1 Inline formatting

2.1.2 Block-level elements

2.1.3 Math expressions

2.2 Markdown extensions by bookdown

2.2.1 Number and reference examples

2.2.2 Theorems and proofs

2.2.3 Special headers

2.2.4 Text references

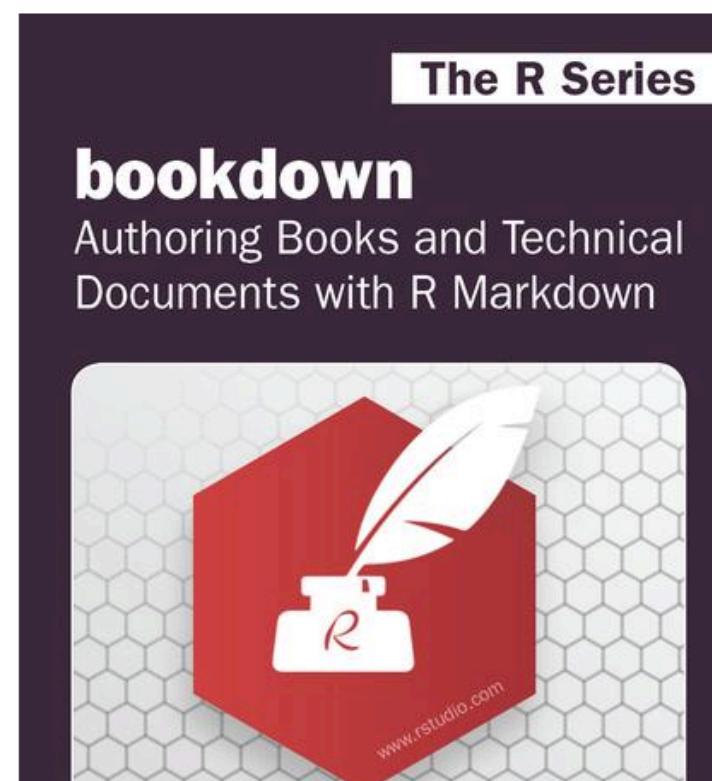
# bookdown: Authoring Books and Technical Documents with R Markdown

Yihui Xie

2021-03-15

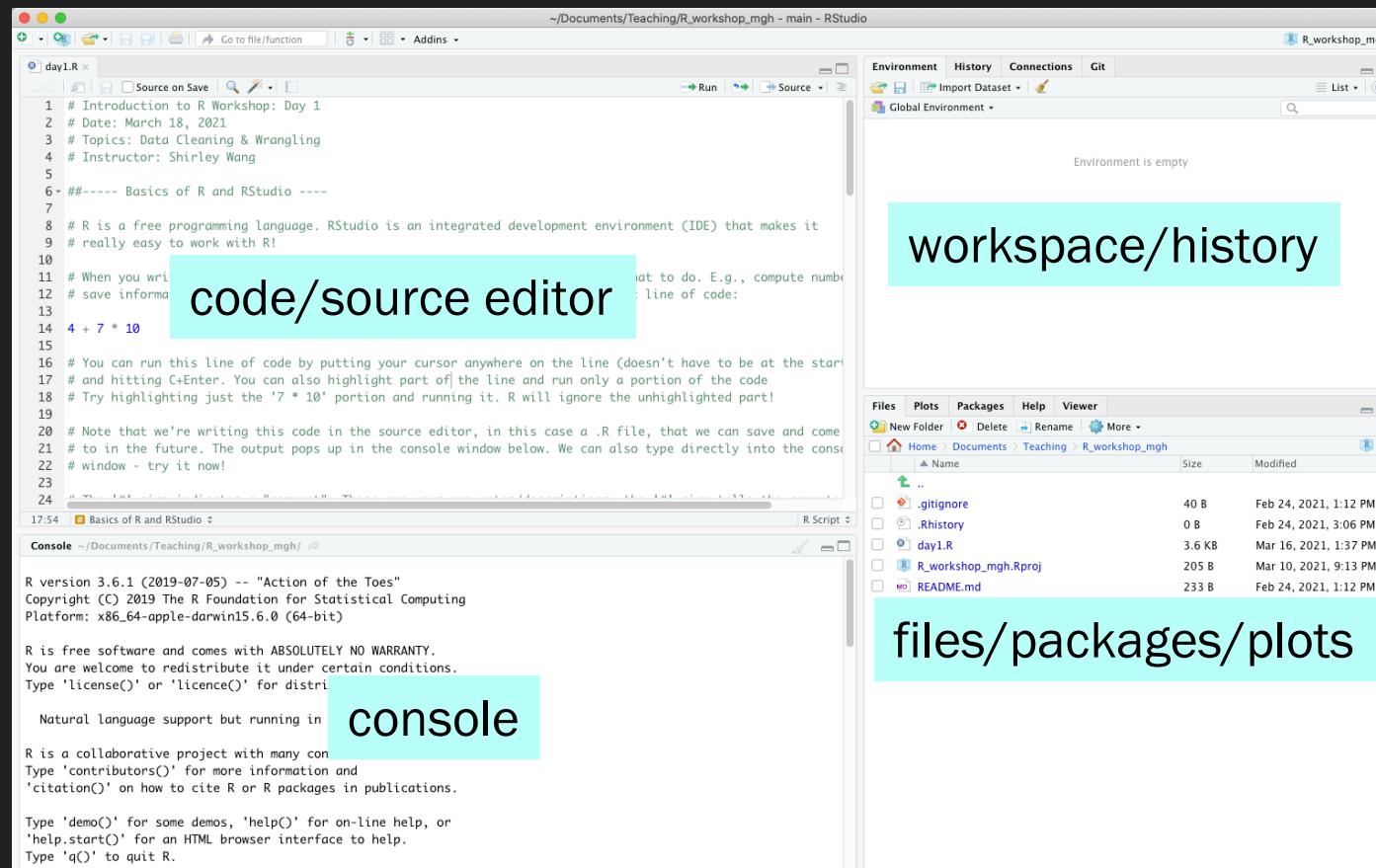
## Preface

---



# R & RStudio Orientation

- R is a free programming language
- RStudio is an IDE that makes it easier write, run, and save R code!



# R Basics

- Case sensitive
- Run lines of code with cmd+Enter
- Get help with ? or `help()`
- Comment your code with #

# R Basics

- Case sensitive
- Run lines of code with cmd+Enter
- Get help with ? or
- Comment your code

When you trying to look at  
the code you wrote a month ago

IT'S SOME KIND OF ELVISH

I CAN'T READ IT

# Operations in R

## Arithmetic Operators

Operator	Description
+	addition
-	subtraction
*	multiplication
/	division
<sup>^ or </sup> **	exponentiation
x %% y	modulus (x mod y) 5%%2 is 1
x %/% y	integer division 5%/%2 is 2

## 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
isTRUE(x)	test if X is TRUE

# Objects in R

- Objects store information (can be an entire dataset, a single value, a string of text, an image, etc)
- Assign values using ‘`<-`’
- Value on the right-hand side gets assigned to object name on the left-hand side (e.g., `myname <- 'shirley'`; `myage <- 26`)
- Rules for object names:
  - Must start with a letter
  - Can only contain letters, numbers, periods, underscores
  - Case-sensitive

# Data Types

- Numeric
- Integer
- Character
- Factor
- Logical
- Check data types in R with `class()`

# Data Structures

- Vector
- Matrix
- Array
- Data frame
- List

# Data Structures

- Vector
  - Matrix
  - Array
  - Data frame
  - List

# Data Frames

- Access columns with ‘\$’ (e.g., `df$age` accesses the column named ‘age’ in the data frame ‘df’)
- `View(df)` to view data frame in excel-like format
- `dim(df)` to get dimensions (# rows, # columns)
  - `nrow(df)`
  - `ncol(df)`
- `str(df)` to view structure of data frame
- `colnames(df)` to view all column (variable) names

# Functions

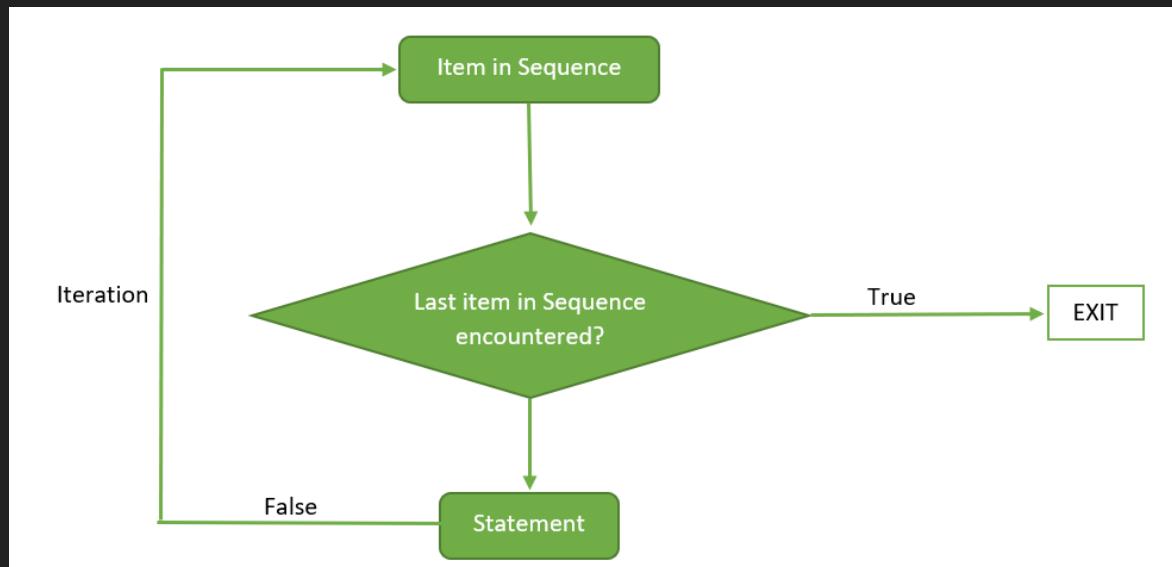
- Sets of instructions/statements to complete a specific task; takes input (e.g., numbers) and returns an output (e.g., mean of those numbers)
- R comes with many built-in functions (e.g., `mean()`, `min()`, `max()`, `print()`)
- You can also write your own functions to perform specific tasks!
- `function(arguments)`
- `myFunction <- function(input) {# do something; return output}`

# Packages

- R packages are bundles of functions, documentation, and data to perform tasks (e.g., machine learning, natural language processing, data visualization)
- Anyone can build an R package to share with other R users!
- Hosted on the central R repository CRAN: <https://cran.r-project.org/>
- Currently over 17,000 packages available
- Packages need to be installed on your computer, and loaded at the start of each new R session
- `install.packages('ggplot2')`
- `library(ggplot2)`

# Loops

- Loops are useful to repeat a block of code multiple times
- Rather than copy/pasting code several times, you can iterate through a loop! This can shorten hundreds of lines of code to just a few lines.



 shirleywang.rbind.io

 shirleywang@g.harvard.edu

 @shirleybwang

# Thank you!

## Helpful resources:

- <https://psyr.djnavarro.net/>
- <https://swirlstats.com/>
- <https://r4ds.had.co.nz/>
- <https://www.r-bloggers.com/>
- <https://stackoverflow.com/>

