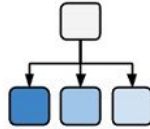


R Workshop I

Carlos Utrilla Guerrero
Data Science - Researcher



Who I am?



Carlos Utrilla Guerrero
Data Science - Researcher
Institute of Data Science
Research focus: Planetary Health & Open Science
c.utrillaguerrero@maastrichtuniversity.nl

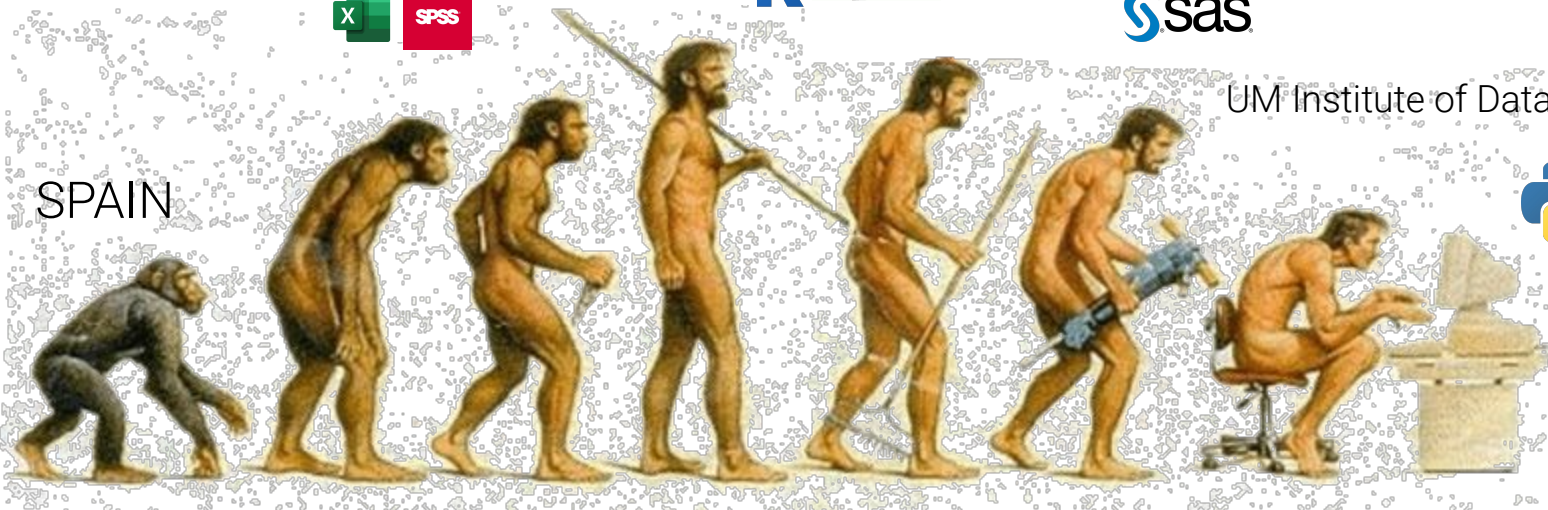
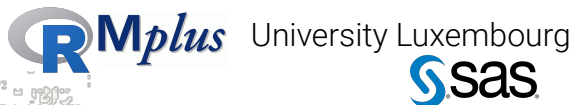
A data scientist at IDS,
teaching the fundamentals
of Data Science,
implementing Responsible
Data Science by design for
research and enabling
cross-discipline
collaborations



Journey to R and Open Science

BSc Economics

MSc Social Research Methods & Stats



SPAIN

UM Institute of Data Science

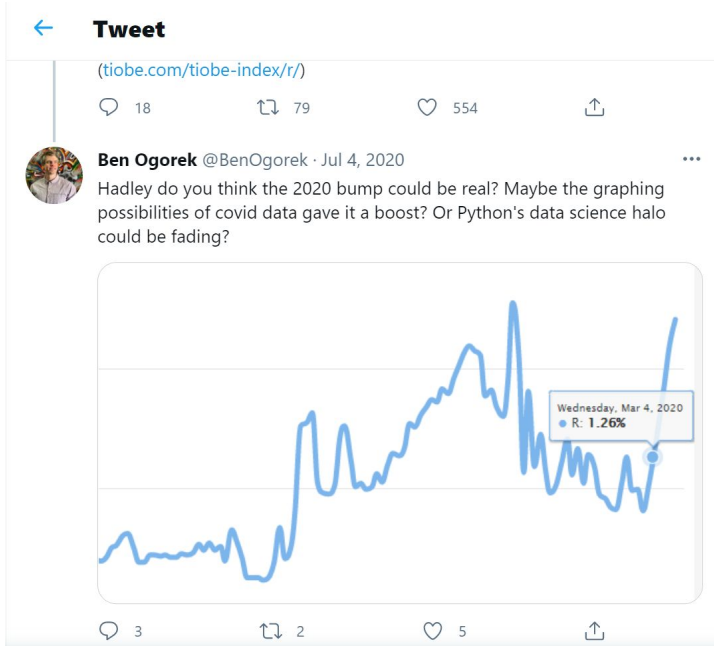


R and Open Science Journey

1986



Present



Paul Jansen, CEO of Tiobe Software, now reckons R and Python have benefited from demand in universities and from global efforts to find a vaccine for the COVID-19 virus.

"The days of commercial statistical languages and packages such as SAS, Stata and SPSS are over," he writes in the July update.

"Universities and research institutes embrace Python and R for their statistical analyses," he continues. "Lots of statistics and data mining needs to be done to find a vaccine for the COVID-19 virus.

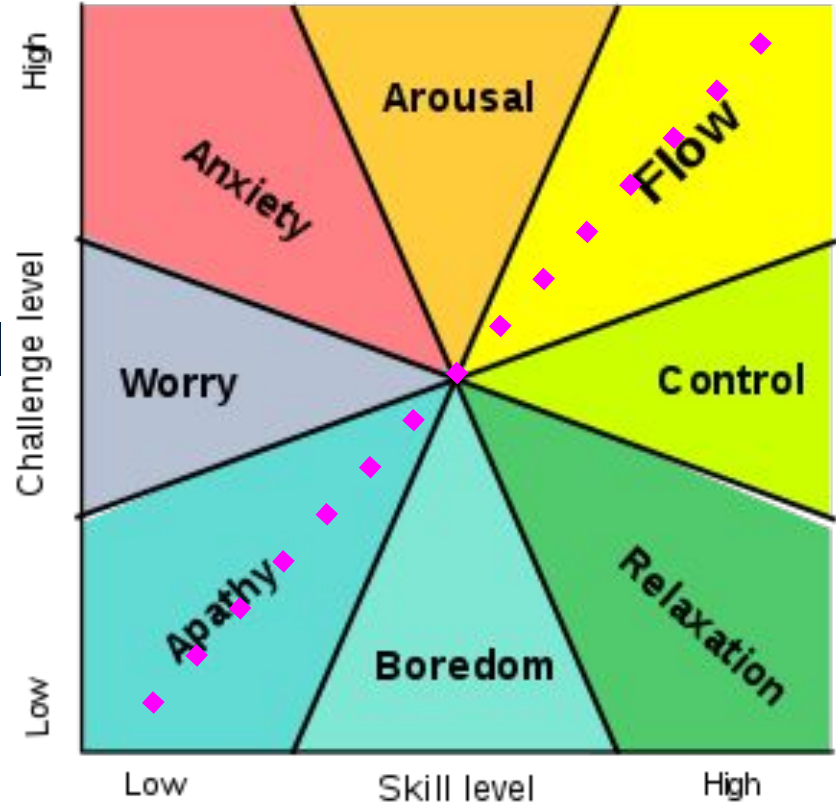
As a consequence, statistical programming languages that are easy to learn and use, gain popularity now."

<https://www.zdnet.com/article/programming-language-rankings-r-makes-a-comeback-but-theres-debate-about-its-rise/>

Achieve 'flow' state *

- Learn and enjoy with R
- Continuous feedback through our workshops
- Balance act challenge-skill

*Mihaly Csikszentmihalyi ([1990](#))



What is R?

- R is an open-source system for the manipulation, statistical and numerical analysis, and graphical display of data



Why use R for your research analysis?

- Connect with an amazing community
- Programming is a superpower that everyone has access to
- Clean, analyze, plot, and communicate with your data all in one place
- Reproducibility
- Automation
- IT'S FREE

What we are covering today

- Basic of R data types
- R's packages system
- Simple statistics
- Visualisation
- Functions





What we aren't doing today

- Covering the data science pipeline
- Deep dive into statistical programming
- Expecting you to 100% get it the first time!



Ultimate Goal

Download and install R and R studio

- Download and install the latest version of R:
 - <https://www.r-project.org/>
- Download and install RStudio:
 - <https://www.rstudio.com/>
- Install R in Windows:
 - <https://www.youtube.com/watch?v=q0PjTAylwoU>
- Install R in Mac:
 - <https://www.youtube.com/watch?v=5-ly3kyxwEg>



R and R studio

R: Engine



RStudio: Dashboard



Rstats - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

~/projects/Rstats/

```
> library(ggplot2)
> ggplot(mtcars, aes(x = disp, y = mpg, color = as.factor(cyl))) + geom_point(size = 3)
> |
```

Environment History Connections Git Tutorial

Global Environment

Environment is empty

Files Plots Packages Help Viewer

as.factor(cyl)

- 4
- 6
- 8

mpg

disp

Console

Output

R: New
Phone



R Packages:
Apps you can
download



Available on the
App Store



GET IT ON
Google Play

Live demo

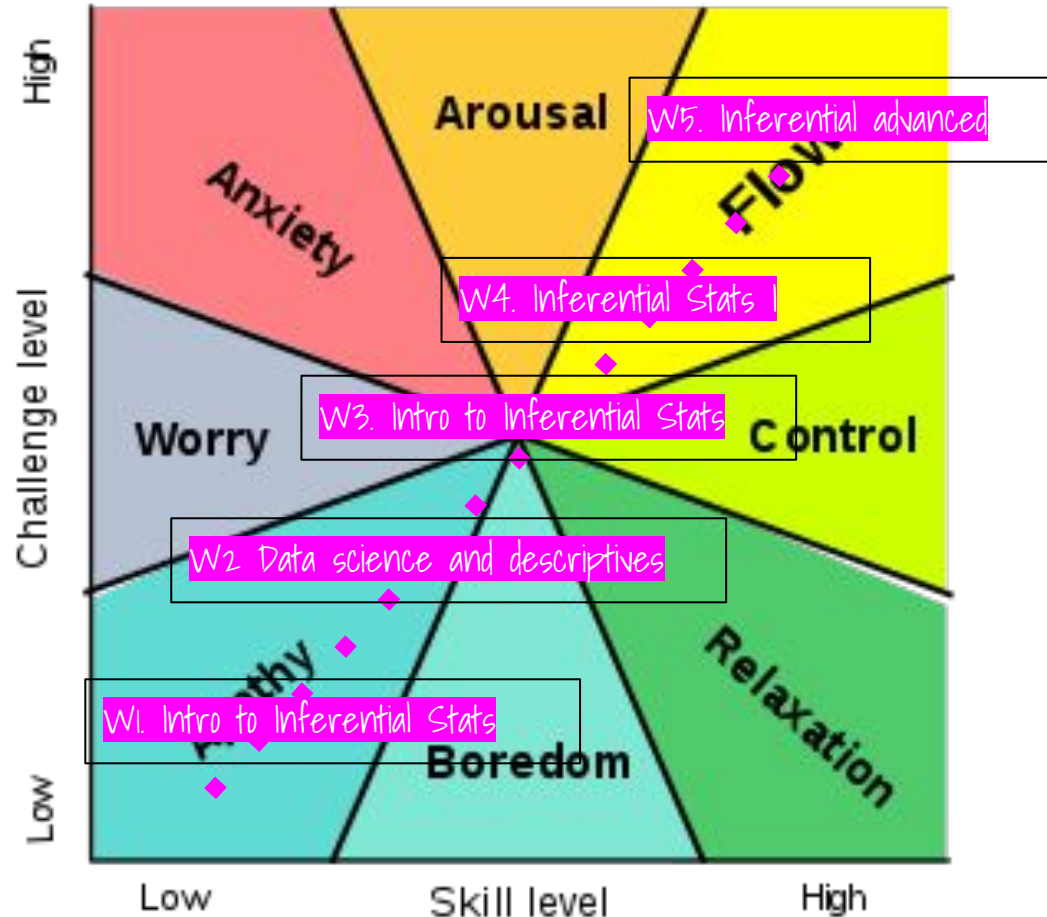


Next steps!

Practice

- After completion assignment 1, you'll be able to:
 - Understand R's basic data types
 - Understand R's basic data structures including vectors and matrix
 - Understand R's package system
 - Perform simple statistics
 - Visualise real-world dataset using R's package
 - Write your own functions

We have a plan for our learning curve



Recommended R training online material

<https://rstudio.cloud/learn/primers>

- Interactive
- Ready to go

Recommended R training online material

<https://datacarpentry.org/r-socialsci/01-intro-to-r/index.html>

- Open Science community
- Step-by-step lessons
- Exercises

1. Before we Start	How to find your way around RStudio? How to interact with R? How to manage your environment? How to install packages?
2. Introduction to R	What data types are available in R? What is an object? How can values be initially assigned to variables of different data types? What arithmetic and logical operators can be used? How can subsets be extracted from vectors? How does R treat missing values? How can we deal with missing values in R?
3. Starting with Data	What is a data.frame? How can I read a complete csv file into R? How can I get basic summary information about my dataset? How can I change the way R treats strings in my dataset? Why would I want strings to be treated differently? How are dates represented in R and how can I change the format?
4. Data Wrangling with dplyr and tidyr	How can I select specific rows and/or columns from a dataframe? How can I combine multiple commands into a single command? How can create new columns or remove existing columns from a dataframe? How can I reformat a dataframe to meet my needs?



Book guides

- Learning statistics with R
 - <https://learningstatisticswithr.com/book/>
- R for Data Science
 - <https://r4ds.had.co.nz/>
- Data Visualization
 - <https://socviz.co/>
- Intro research methods
 - <https://bookdown.org/ejvanholm/Textbook/>

