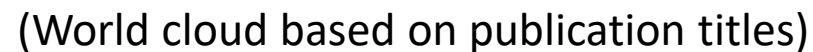


# Analysis and Visualisation of Complex Agro-Environmental Data - 2025

## Lesson 01

- Welcome and self-introductions
- Contents, teaching approach and assessment criteria
- Introduction to data analysis and visualization

[psegurado@isa.ulisboa.pt](mailto:psegurado@isa.ulisboa.pt)



Write up to **3 keywords** that best describe your interests



<https://www.menti.com/aloozbhn2cg>

# Contents: what will I learn?

## Classes

## Contents

1


Welcome, contents, teaching approach and assessment criteria  
Data analysis and visualization: introduction to concepts with examples  
Data visualisation: good design practices  
Introduction to data analysis and statistics  
Data and table types  
Descriptive statistics  
Data distribution analysis  
Hypothesis testing  
Correlation and regression analysis  
Cluster analysis  
Ordination analysis  
Interactive data visualization  
Geospatial Data visualization

14

6 ECTS:

- 35 contact hours
- 133 hours - Autonomous work and assessments

# Teaching approach: how will I learn?

- Two Learning components: **Numerical analysis**  **Visualization**
- **“Learning by doing”** approach: practical exercises with real agro-environmental case studies

# Teaching approach: how will I learn?

## A typical class

1. Solve the exercises of the previous week: 30 min
2. Introduction to the topic: 45 min
3. Working example (jupiter notebook): 30 min
4. Exercise: 30 min

# Assessment criteria: How I will be assessed?

3 criteria:

- Two **assignments** (or one final exam) – focused on theoretical-practical knowledge: **40%**
- One **final project** – a **group work** focused on data storytelling: **40%**
- **Participation** – deliver of weekly exercises: **10%** - with an extra bonus of up until **10%** depending on how well the exercises were solved.



talia.pt

~~~~~  
perfection is being  
courageously imperfect  
~~~~~



# Final project

**Goal:** tell a coherent and appealing story from a complex agro-environmental dataset

## **Steps:**

1. Problem definition
2. Database queries
3. Summary statistics
4. Exploratory data analysis
5. Inferential statistics
6. Final visualisation product and storytelling

## **Assessment:**

1. A **live presentation** of your story, a **poster**, or an **interactive dashboard** and a short **written report**, including the code as an Appendix.

# Teaching approach: how will I learn?

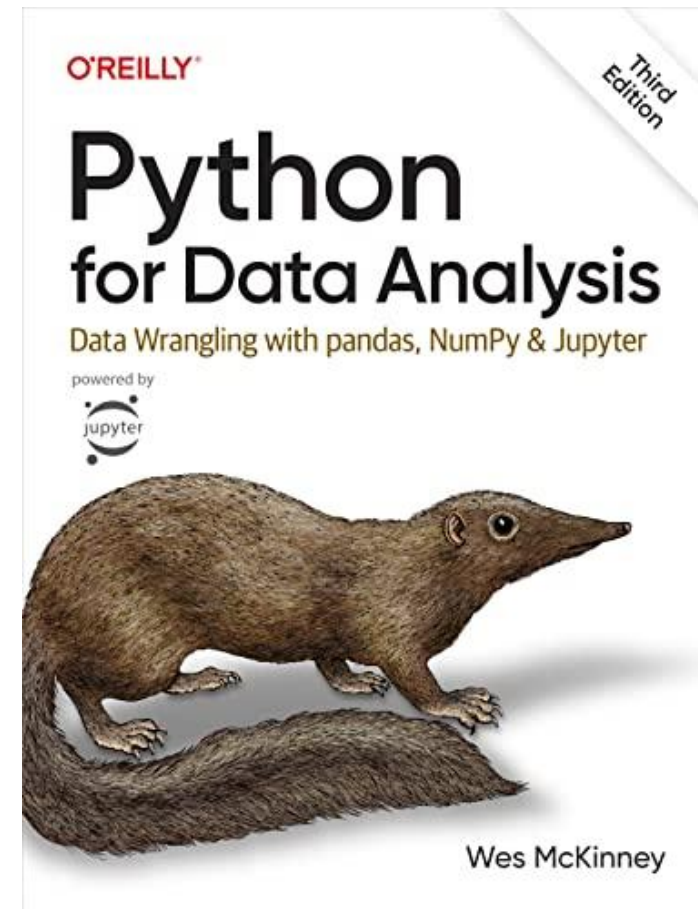
Programming language



# Materials (autonomous work)

- Main book:  
Wes McKinney. *Python for Data Analysis*. <https://wesmckinney.com/book/>

- Shared folder with some books



# Materials (autonomous work)

## Pandas

Python Tutorials by Corey Schafer – Pandas tutorials (parts 1 to 11 – vídeos #120 to #138)

<https://www.youtube.com/playlist?list=PL-osiE80TeTt2d9bfVyTiXJA-UTHn6WwU>

## Statistics

- Statistics and probability – Khan Academy (in particular units 10, 11, 12, 14 and 16):

<https://www.khanacademy.org/math/statistics-probability>

- Statquest videos:

<https://statquest.org/video-index/>

## Statistics with python

Videos by Shashank Kalanithi based on the book: *Practical Statistics for Data Scientists*

- Chapter 1 - Exploratory Data Analysis: <https://www.youtube.com/watch?v=wwsizzg6UjU>
- Chapter 2 - Data and Sampling Distributions: <https://www.youtube.com/watch?v=7mi6cJSTj6Y>
- Chapter 3 - Statistical Experiments Significance Testing: <https://www.youtube.com/watch?v=mzjooX4OyFs>

Original book (also available in the google drive): [https://www.researchgate.net/profile/Janine-](https://www.researchgate.net/profile/Janine-Zitianellis/post/Can_anyone_please_suggest_a_books_on_machine_learning_using_R_Programming/attachment/613a5b83647f3906fc975a71/AS%3A1066204907204608%401631214467436/download/Practical+Statistics+for+Data+Scientists+50%2B+Essential+Concepts+Using+R+and+Python+by+Peter+Bruce%2C+Andrew+Bruce%2C+Peter+Ge)

[Zitianellis/post/Can\\_anyone\\_please\\_suggest\\_a\\_books\\_on\\_machine\\_learning\\_using\\_R\\_Programming/attachment/613a5b83647f3906fc975a71/AS%3A1066204907204608%401631214467436/download/Practical+Statistics+for+Data+Scientists+50%2B+Essential+Concepts+Using+R+and+Python+by+Peter+Bruce%2C+Andrew+Bruce%2C+Peter+Ge](https://www.researchgate.net/profile/Janine-Zitianellis/post/Can_anyone_please_suggest_a_books_on_machine_learning_using_R_Programming/attachment/613a5b83647f3906fc975a71/AS%3A1066204907204608%401631214467436/download/Practical+Statistics+for+Data+Scientists+50%2B+Essential+Concepts+Using+R+and+Python+by+Peter+Bruce%2C+Andrew+Bruce%2C+Peter+Ge)  
deck.pdf

Original github repository of the book - Jupiter notebooks: <https://github.com/gedeck/practical-statistics-for-data-scientists/tree/master/python/notebooks>

## **Short questionnaire - previous skills on data visualization and statistics**

## **Short questionnaire - previous skills on data visualization and statistics**

# Why Python (and not R)?











vs.



# Why Python (and not R)?

## 1. Popularity

January 2023

Jan 2023	Jan 2022	Change	Programming Language		Ratings	Change
1	1			Python	16.36%	+2.78%
2	2			C	16.26%	+3.82%
3	4	▲		C++	12.91%	+4.62%
4	3	▼		Java	12.21%	+1.55%
5	5			C#	5.73%	+0.05%
6	6			Visual Basic	4.64%	-0.10%
7	7			JavaScript	2.87%	+0.78%
8	9	▲		SQL	2.50%	+0.70%
9	8	▼		Assembly language	1.60%	-0.25%
10	11	▲		PHP	1.39%	-0.00%
11	10	▼		Swift	1.20%	-0.21%
12	13	▲		Go	1.14%	+0.10%
13	12	▼		R	1.04%	-0.21%



In 2024, R dropped to the 21st position...

Source: <https://www.tiobe.com/tiobe-index/>



# Why Python (and not R)?

## 1. Popularity

January 2025

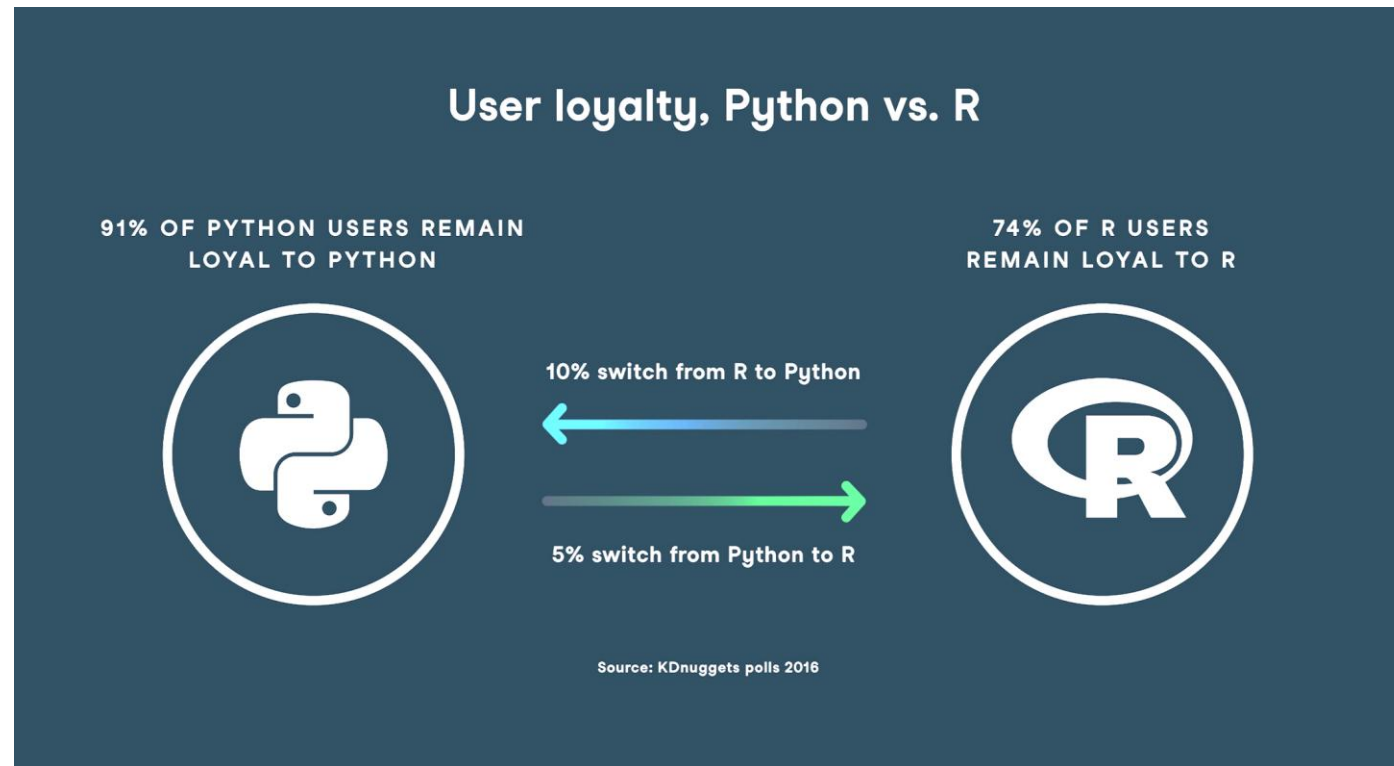
Feb 2025	Feb 2024	Change	Programming Language		Ratings	Change
1	1			Python	23.88%	+8.72%
2	3	^		C++	11.37%	+0.84%
3	4	^		Java	10.66%	+1.79%
4	2	v		C	9.84%	-1.14%
5	5			C#	4.12%	-3.41%
6	6			JavaScript	3.78%	+0.61%
7	7			SQL	2.87%	+1.04%
8	8			Go	2.26%	+0.53%
9	12	^		Delphi/Object Pascal	2.18%	+0.78%
10	9	v		Visual Basic	2.04%	+0.52%
11	11			Fortran	1.75%	+0.35%
12	15	^		Scratch	1.54%	+0.36%
13	18	^^		Rust	1.47%	+0.42%
14	10	v		PHP	1.14%	-0.37%
15	21	^^		R	1.06%	+0.07%



... but in 2025, R rose again to the 15th position.

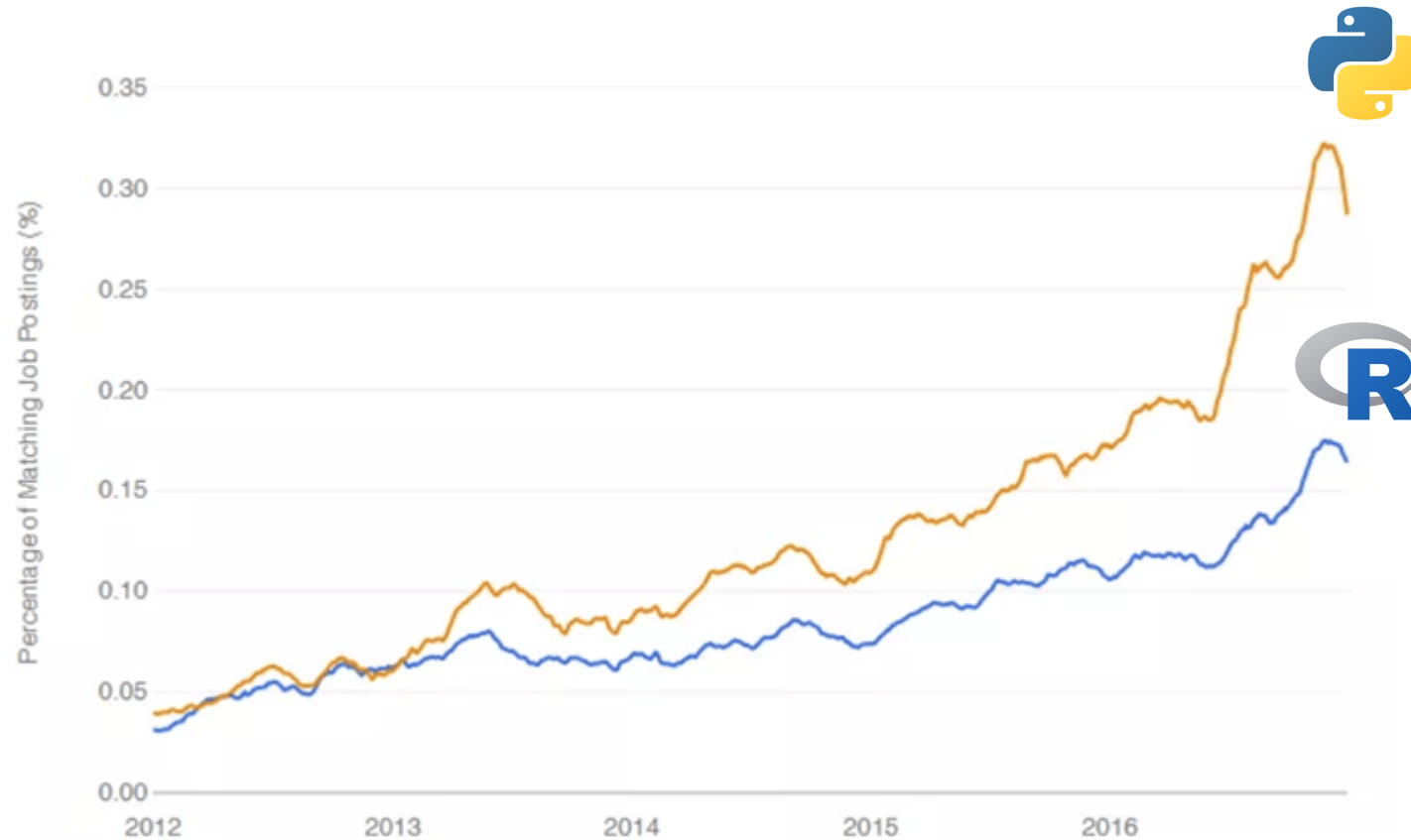
# Why Python (and not R)?

## 2. Loyalty



# Why Python (and not R)?

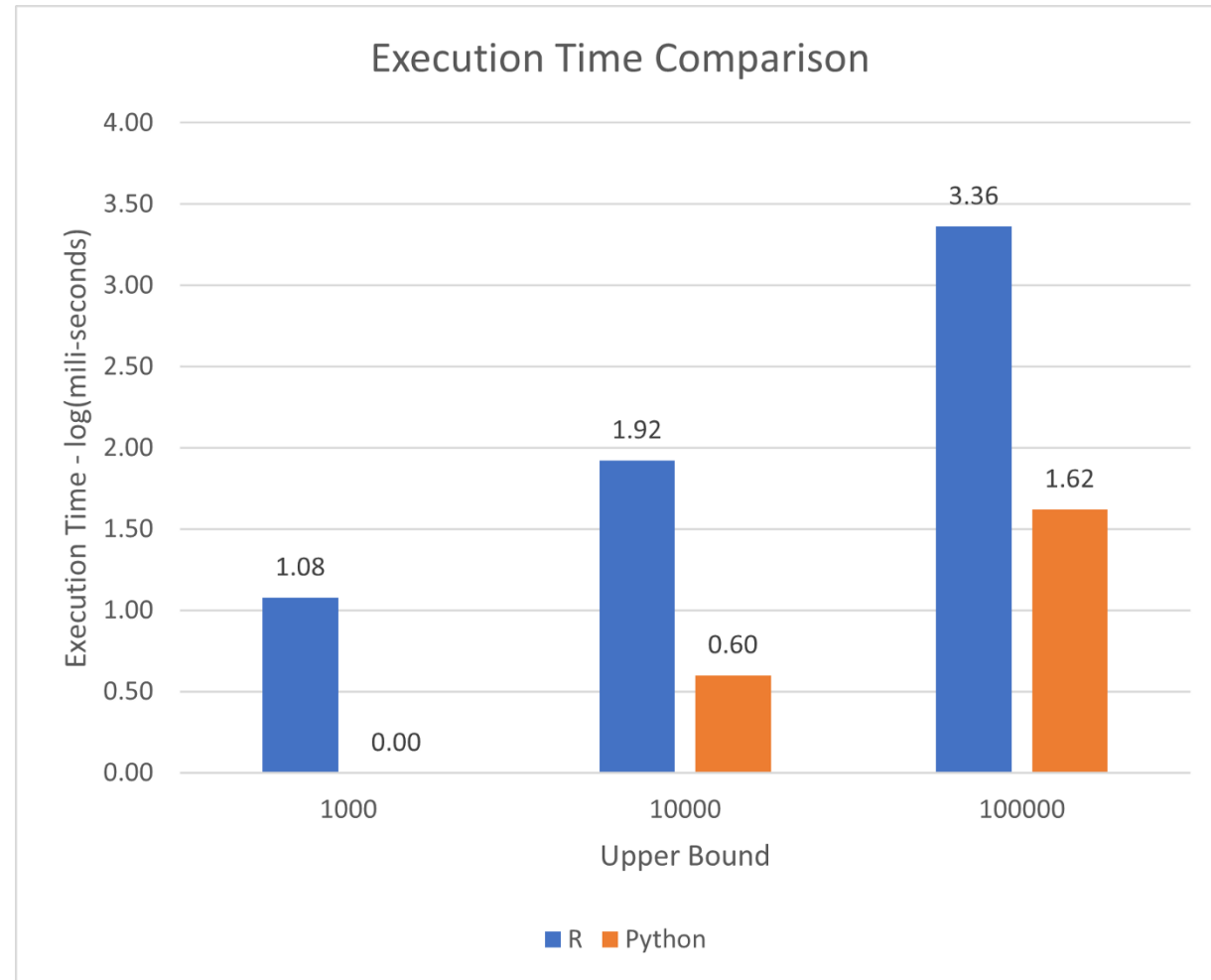
## 3. Job opportunities



# Why Python (and not R)?

## 4. Execution time

Example: a code block written in Python and R to obtain prime numbers up to a pre-defined upper bound



# Why Python (and not R)?

## Summary



Criteria	Python	R
Popularity	Variety of industries.	Amongst Data Scientists/ Statisticians
Community	Large	Statisticians and Academia
Complexity	Depends on your skills	Depends on your skills
Performance	Excellent	Needs improvement
Data Analysis	Good	Excellent
Visualization	Excellent	Excellent
Data Modeling/Machine Learning	Excellent	Good
Deployment and Production	Excellent	Needs improvement

# Python libraries used in the AVDAC course



## Data manipulation and analysis

- pandas (data manipulation)
- numpy (mathematical functions)
- scipy (mathematical functions)
- Statsmodels (statistical and modelling functions)



## Visualization

- matplotlib (basic plots)
- seaborn (more advance statistical plots)
- bokeh (interactive visualization)
- plotly (interactive Geospatial Data visualization)
- GeoPandas (Geospatial Data visualization)
- Folium (Geospatial Data visualization)



GeoPandas



Folium

# Software/IDEs



**Bash/PowerShell** (Terminal Command Line Interfaces) – Manipulate directories and files; install software and packages.



**Git** – to exchange files between local and remote machines (github)



**VSCode** – running python, git, markdown, editing jupyter notebooks



**Jupyter notebook** – editing jupyter notebooks

# Repositories and platforms



**Fenix ISA** – classes; documentation



**Google drive** – other supporting materials



**Github** – data; exercises; examples; outputs



**Discord** – interaction among students and the teacher





What is data analysis and visualization

# What is data analysis and visualization

**03:00**



<https://www.menti.com/aln2ry6zo3p1>

## Data analysis is ...

the process of **inspecting, cleaning, transforming**, and **modeling** data with the goal of **discovering** useful information, **informing** conclusions, and **supporting** decision-making.

(Brown, 2014)

# What is data analysis and visualization?

**03:00**



<https://www.menti.com/alaxvjbzdgqt>

# Data visualization is ...

the process of graphically **representing information** and **data** aiming at placing complex data in a visual format to facilitate real-time interpretation.

or putting it more simply:

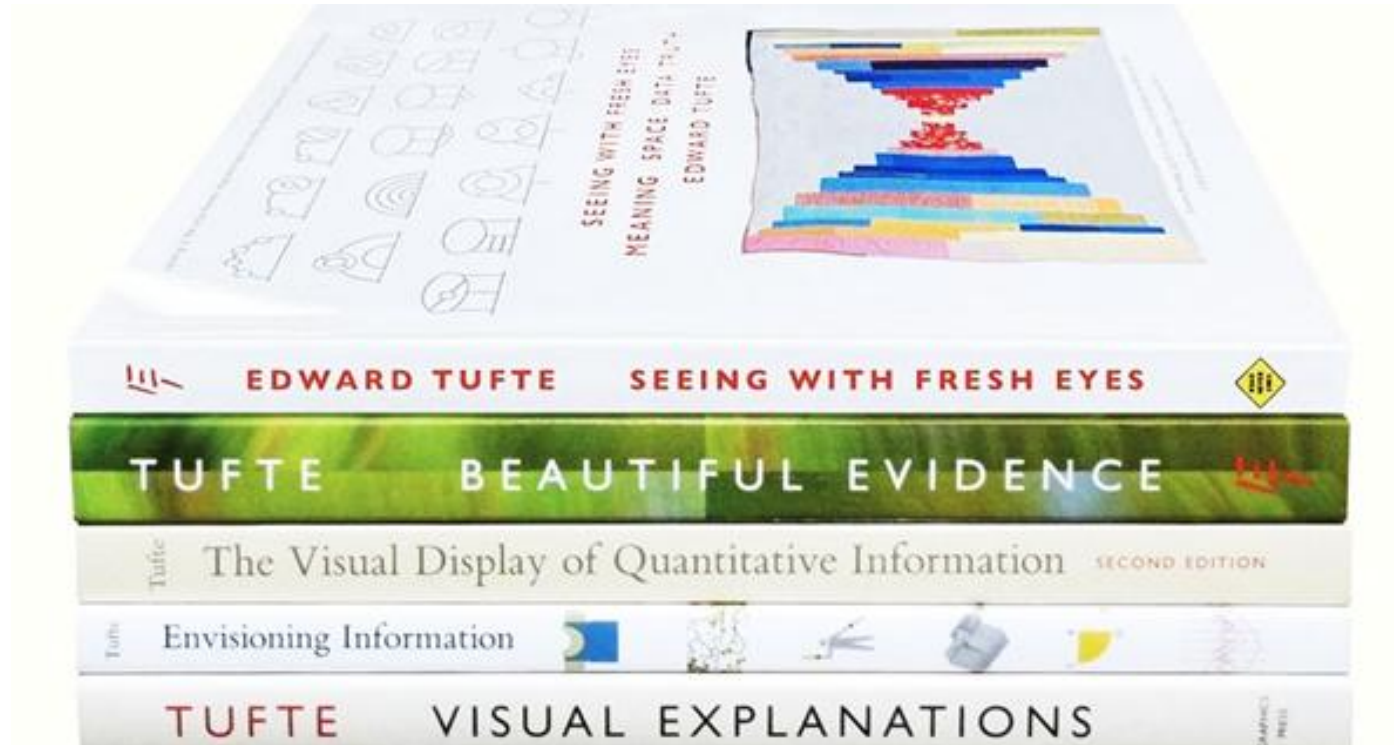
“The visual representation and presentation of data to facilitate understanding” (Kirk, 2019).

## The Art of Data Visualization



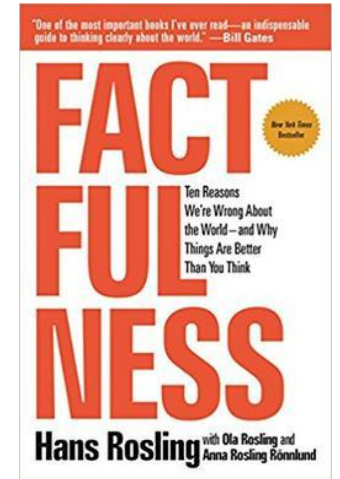
# Big names of data visualization and storytelling

Edward Tufte (EUA, 1942-) - <https://www.edwardtufte.com/tufte/>



Hans Rosling (Sweden, 1948-2017) – Author of the best seller Factfulness

200 Countries, 200 Years, 4 Minutes - The Joy of Stats (Trendalyzer software)



<https://www.youtube.com/watch?v=jbkSRLYSojo>

Watch also: DON'T PANIC — Hans Rosling showing the facts about population <https://www.youtube.com/watch?v=FACK2knC08E>



# Exercise 1

<https://github.com/isa-ulisboa/greends-avcad-2025/blob/main/exercises/avcd-exerc1-wordcloud.ipynb>

Create a personal word cloud with python, using:

- the text from your CV, or ...
- the text of a motivation letter, or ...
- A text with keywords that best describe your background and interests

You may also use a shape of something related with your interests or hobbies.