Introduction to Data Science
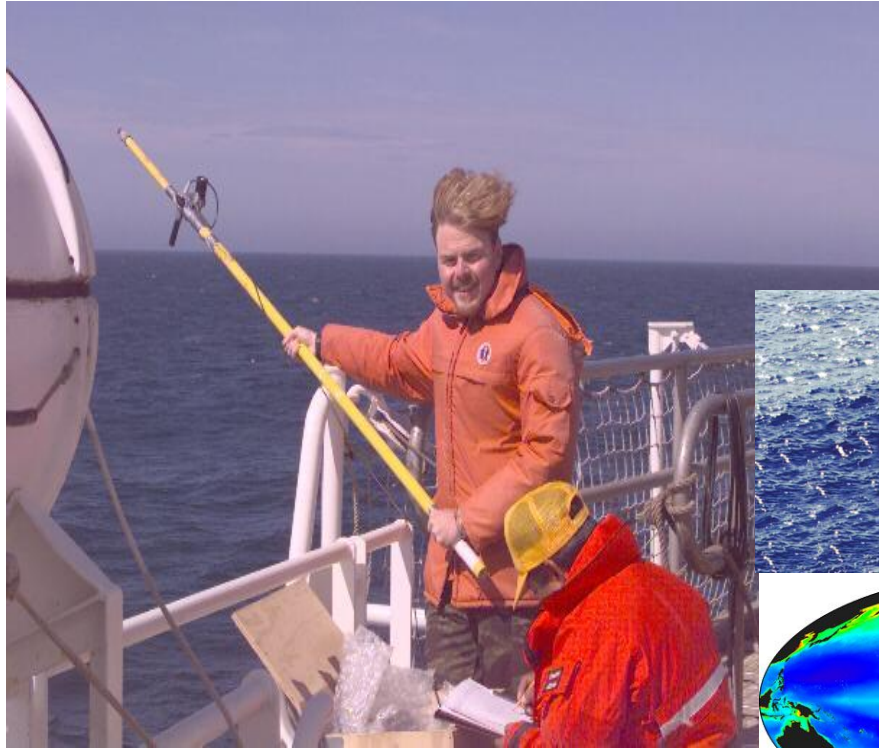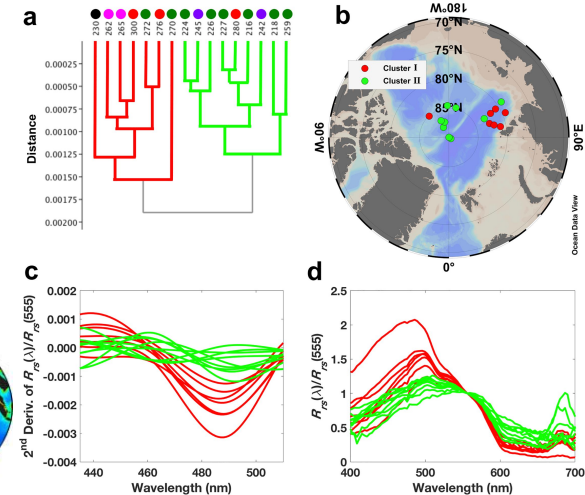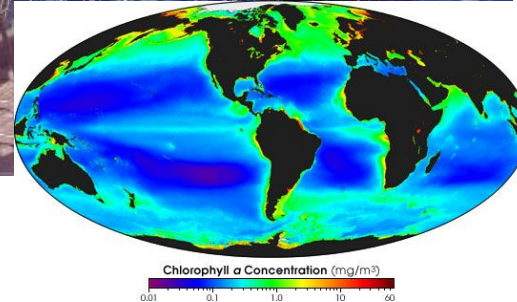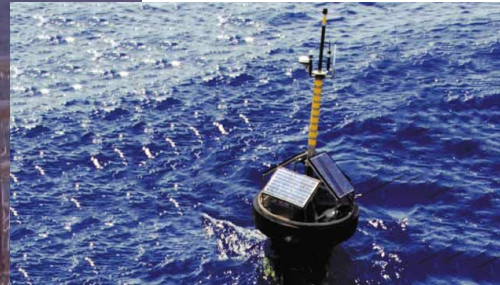
Servet Ahmet Çizmeli

ahmet@pranageo.com

PRANAGE

About me

My name is Servet Ahmet Çizmeli. I am researcher in **environmental sciences**. Now I decided to go in the direction of **tech entrepreneurship**.

About us

PranaGEO.com is a young **Research** and **Development** company in Bodrum, Turkey. We provide data analysis & software solutions for **Data Science** projects.

We would like to create software to make the dream of **Reproducible Research** come true.

Reproducible Research
Problem

"an article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result."

(Donoho, 2010)

we need a RR-aware platform. If possible, **ONE** platform that binds all the required services/stages

Reproducible Research
Why?

Computational research today needs to be 100% **open and reproducible**. Why?

- Research advances when others construct **on top of** what we build, and vice versa.

- Scientific results are subjective and multi-dimensional. Full access to **datasets**/**figures**/**code** is needed to prove/**refute**/**compare**/**improve** existing research

- On the contrary to common, old fashioned belief, **we get richer when we share!**
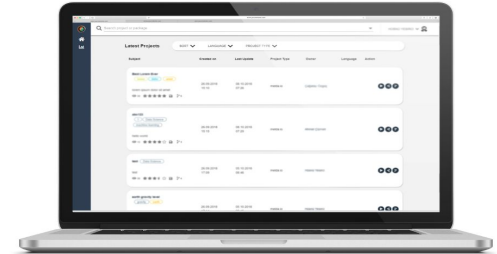
Our product : melda.io

Modern, Elegant **D**ata **A**nalysis

melda.io is a new, innovative, web-based cloud-native data science platform. You can create data analysis projects with R, python, publish your work, interact and co-create with others.

Start using melda.io for free

it was designed to promote reproducibility, online **collaboration**, to make it **easy** to conduct environmental/data scientific studies, hence speed up **innovation**

save time

Data science

Innovation + business+ data + coding

Innovation

Definition of Innovation

🔊 innovation

/ɪnəˈveɪʃ(ə)n/

noun

the action or process of innovating.
"innovation is crucial to the continuing success of any organization"

Similar: (change) (alteration) (revolution) (upheaval) (transformation)

- a new method, idea, product, etc.

Example

Definition of Healthcare Innovation

Medical Innovation leads to the discovery and development of products and therapies that **save**, **improve** and **extend** lives — the primary goal of any healthcare program.

And by doing that, innovation helps keep healthcare **costs down** and drives **economic growth**.

Data science
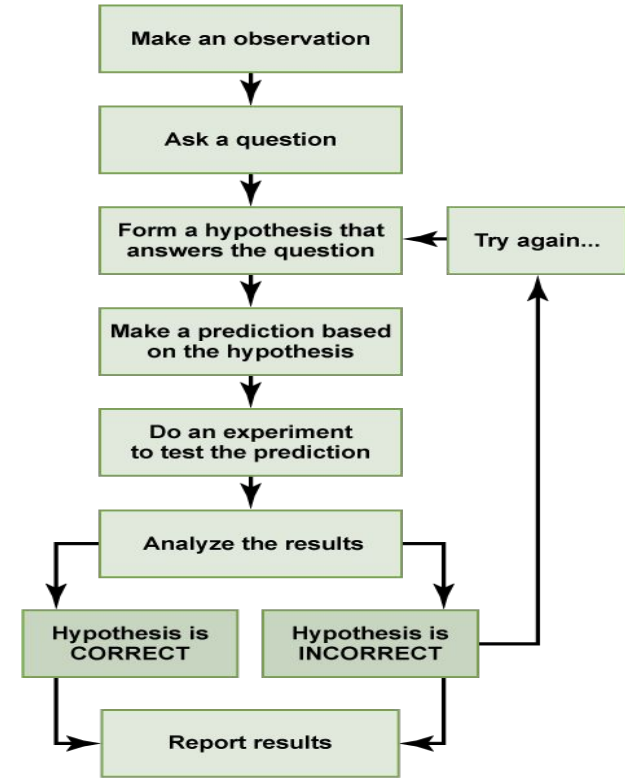
It's in the business world!

Scientific Method
Defined

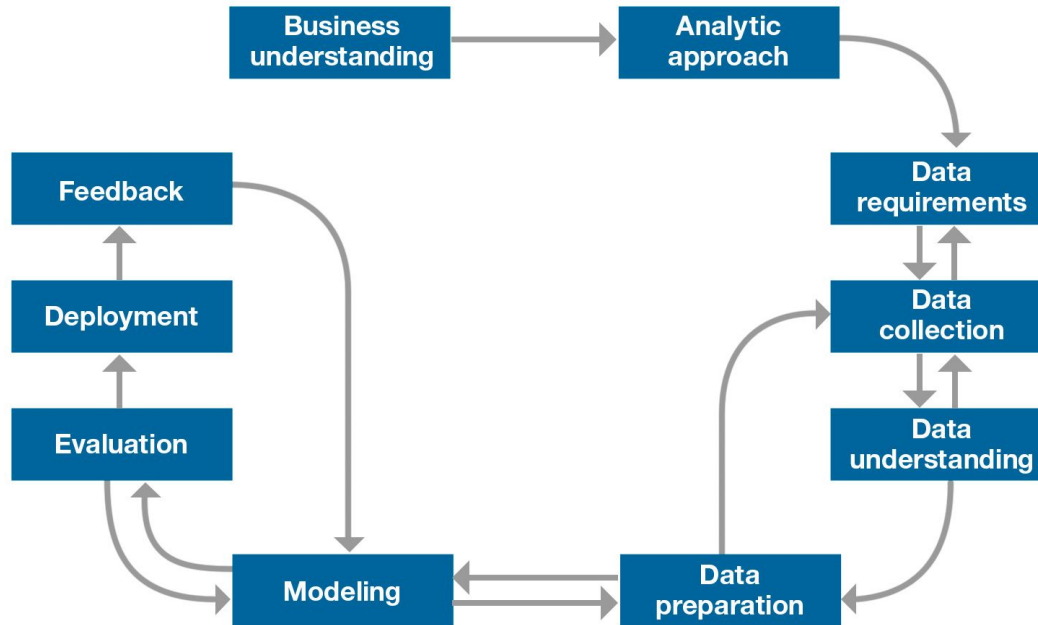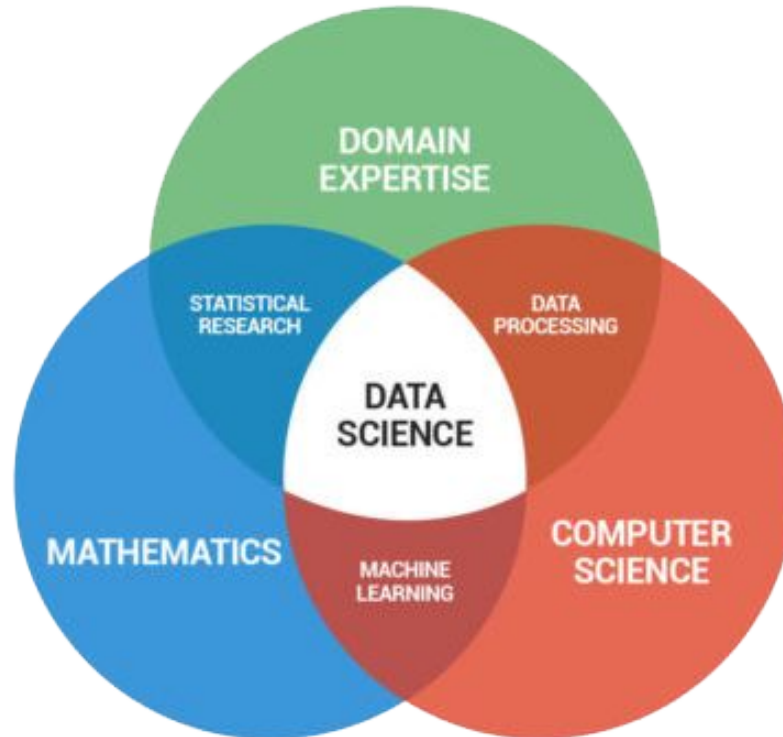PRANAGEO

https://biologydictionary.net/scientific-method

https://courses.lumenlearning.com/boundless-psychology/chapter/the-scientific-method/

Scientific Method
Implemented in data science

Data science
components

Data science
Who is a data scientist?

R	53%
python	53%
SQL	40%
MATLAB	19%
Java	18%
C/C++	18%

Male
70% of Data Scientists in our research were male

2 Languages
Data scientists speak at least 1 foreign language on average

2 years
This is a new profession. The median experience as data scientists of professionals in our research was 2 years

4.5 years
People who work as data scientists currently have a median work experience of 4.5 years (including previous positions)
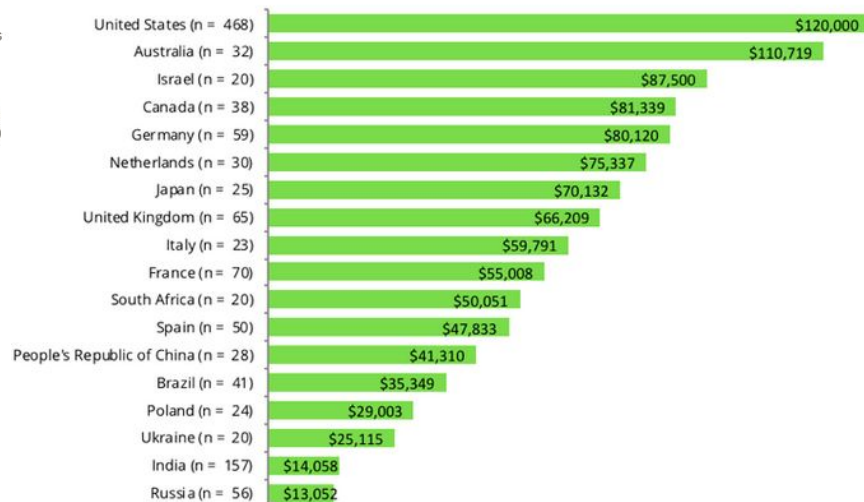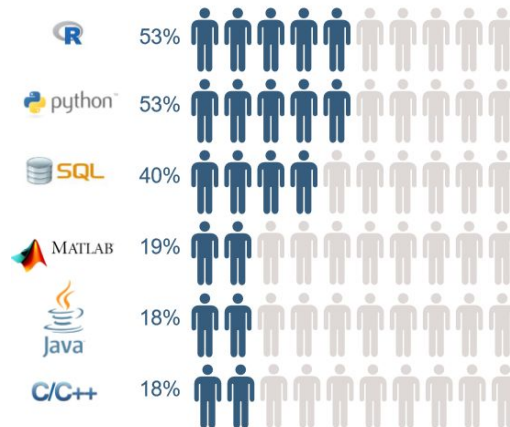
R and/or Python
More than 50% of the data scientists in our research work in R and/or Python

Master or PhD
75% of data scientists have a PhD (27%) or a Master (48%) degree

Country	Salary
United States (n = 468)	$120,000
Australia (n = 32)	$110,719
Israel (n = 20)	$87,500
Canada (n = 38)	$81,339
Germany (n = 59)	$80,120
Netherlands (n = 30)	$75,337
Japan (n = 25)	$70,132
United Kingdom (n = 65)	$66,209
Italy (n = 23)	$59,791
France (n= 70)	$55,008
South Africa (n = 20)	$50,051
Spain (n = 50)	$47,833
People's Republic of China (n = 28)	$41,310
Brazil (n = 41)	$35,349
Poland (n = 24)	$29,003
Ukraine (n = 20)	$25,115
India (n = 157)	$14,058
Russia (n = 56)	$13,052

Data Science
Preliminary questions to answer

Do you already have data?

If not, do you know where to get it?

What kind of data analysis software/platform do you currently use?

What kind of analysis do you plan on using?

Does that involve use of hardware?
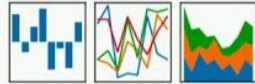
Data Science
Things to know

Data Science
Open source tools we prefer

https://melda.io

Contact us:

melda.io

ahmet@pranageo.com

PRANAGE