APPLICATIONS

OF DATA SCIENCE

Introduction

Applications of Data Science - Class 0

Giora Simchoni

`gsimchoni@gmail.com and add #dsapps in subject`

Stat. and OR Department, TAU

2023-03-12

You are here.

- You have a basic background in Statistics, Data Science, Machine Learning (Stat. Learning is a prerequisite)

- You're OK with R, Python or both (Some CS introduction is a prerequisite)

- You want to know more

- You need some experience in real data projects

- You're thinking about actually getting a job

- You need a portfolio

I must insist.

- You have a terminal, a.k.a shell (Mac/Windows 10 - See installation instructions [here](). Windows 7 and less - 😰)
- You start becoming friendly with your terminal

- You have git on your system
- You have a personal Github account
- You are familiar with Google. And later, [Stack Overflow]().

And either:

- You have Docker on your system
- And then, you won't need to install anything and we'll work on the exact same environment. Almost.

Or:

- You have the latest R (4.2.X) and latest RStudio IDE (2022.X) and you are able to install packages with `install.packages("package_name")`
- You have Python (at least 3.6), Jupyter, and you are able to install packages with `pip install`. Consider working with Anaconda/~~PyCharm~~ VSCode.

# Speak absolutely no R? Absolutely no Python?

- Well you better start (R):
  - [DataCamp: Introduction to R](#)
  - [Swirl](#)
  - [edX: Data Science: R Basics](#)
  - [RStudio: R for Data Science](#)
- Well you better start (Python):
  - [DataCamp: Introduction to Python](#)
  - [W3Schools: Python Introduction](#)
  - [edX](#), [Coursera](#) (really, just pick one)
  - [Andreas Ernst: Python-Lectures](#)

We learn by doing, we strive for reproducibility

R notebooks for R
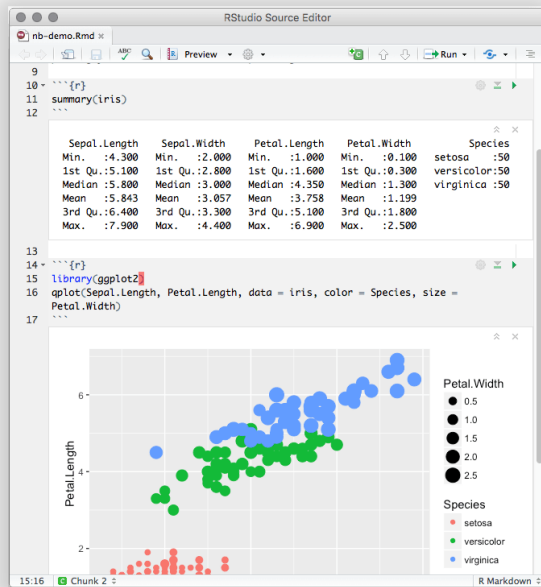
Jupyter notebooks for Python
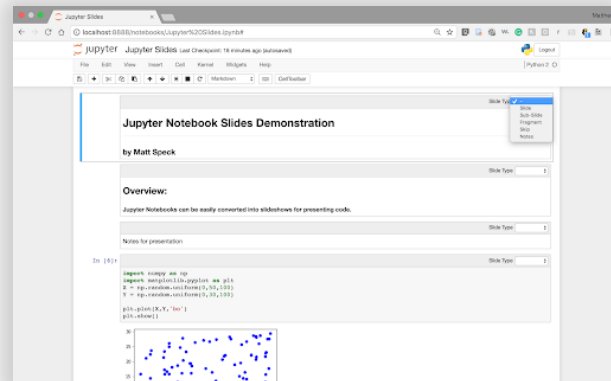
Grade

- 7 home assignments (50%)

  - Miss one or have your lowest grade dismissed

  - HW6 is cool and mandatory 👠

  - Submitted via Github, more on that later

- One final project (50%)

  - Something you can show other humans

  - More on that later

- Though I strongly encourage you to not worry about the grade, start submerging yourselves into the material or if you have to worry about something, worry about what you're going to show future employers

Finally, what we'll learn

- Part I: Data Wrangling + Viz in the Tidyverse (R) and in pandas (Python)

- Part II: Intro to Network Analysis (Python)

- Part III: More Predictive Modeling (R)

- Part IV: Deep Learning – beyond the basics. Or maybe not, up to you. (Python)

- ~~Part V: Building a DS Project (R and a bit of Python)~~ (I doubt it)

APPLICATIONS
OF DATA SCIENCE

Don't Panic.

Posit Cloud, Binder, Colab Demos

Git Demo

Docker Demo

Are you ready?

Go to: bit.ly/dsapps2023_tidyverse