24 June 2020

Atelier Python | Python workshop

- Pandas, tracer et l'analyse des données | Pandas, plotting and data analysis

Présenté par : Jarno van der Kolk
Helpers : Purvasha Patnaik, Masoumeh Soflaei, and Will Kochtitzky

Getting Started

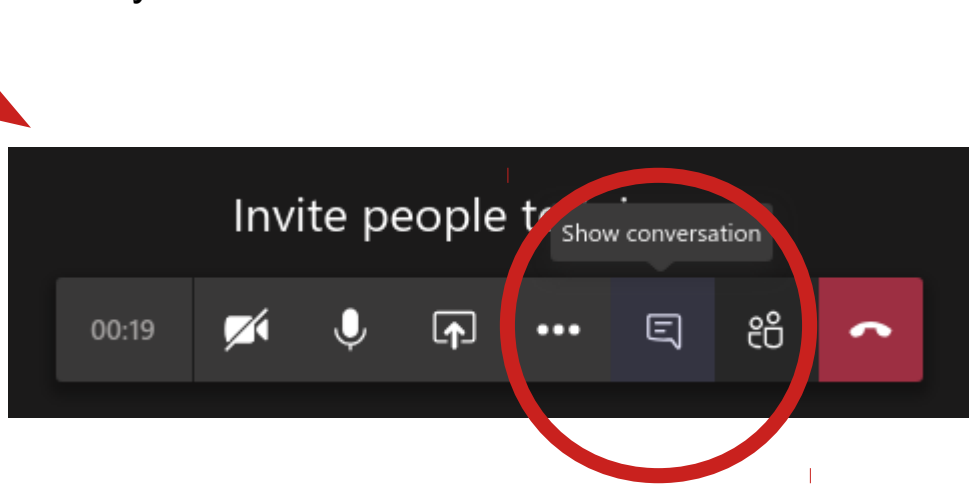
- Download the slides: jarno.ca/python.pdf
- Go to syzygy.ca
  - Click on LAUNCH in top-right corner
  - Choose UOTTAWA
  - Click on the red house with “Log in” in small print
  - Use your uoAccess credentials to log in
- If you don’t have uoAccess credentials, talk to us over Teams

Program

9:00	Intro
9:30	Reading data
10:30	Break
10:45	Plotting
12:00	Lunch break
13:00	Data manipulation
14:30	Break
14:45	Basic scripting with if-then and for loops
16:00	END

Say Hello!

- We will use the chat function in Teams for interactivity
- Say hello and introduce yourself!
  - Why are you here?



Helpers

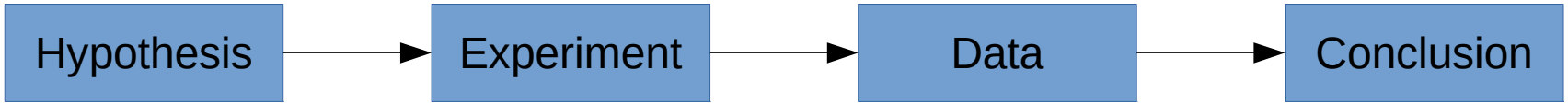
- We have three helpers with us today!
  - Will Kochtitzky
  - Purvasha Patnaik
  - Fabrizio Donzelli
- They will help if you get in trouble!

*All pictures shown are for illustration purpose only. Actual appearance may differ from images shown.

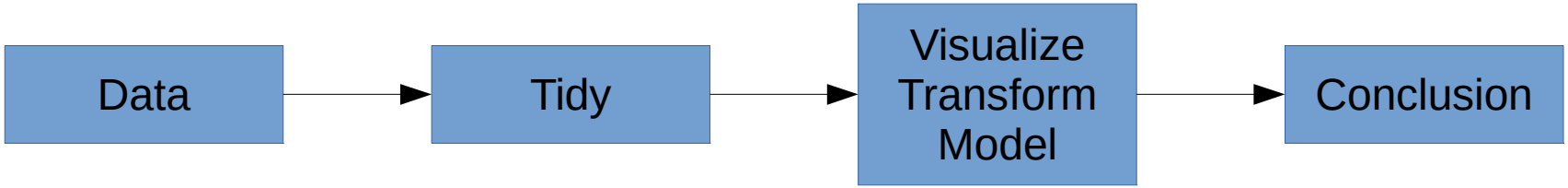
Data Science

- What do you think Data Science is?
- Type your answers in the chat!
- Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge.

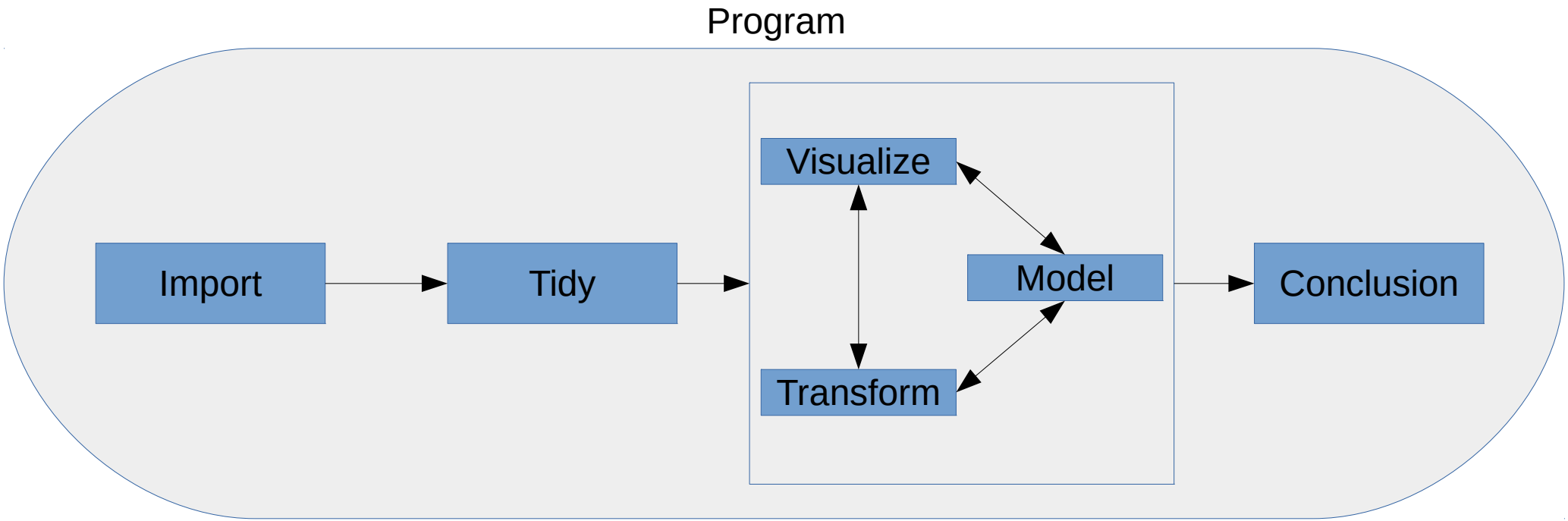
Normal Science



Data Science

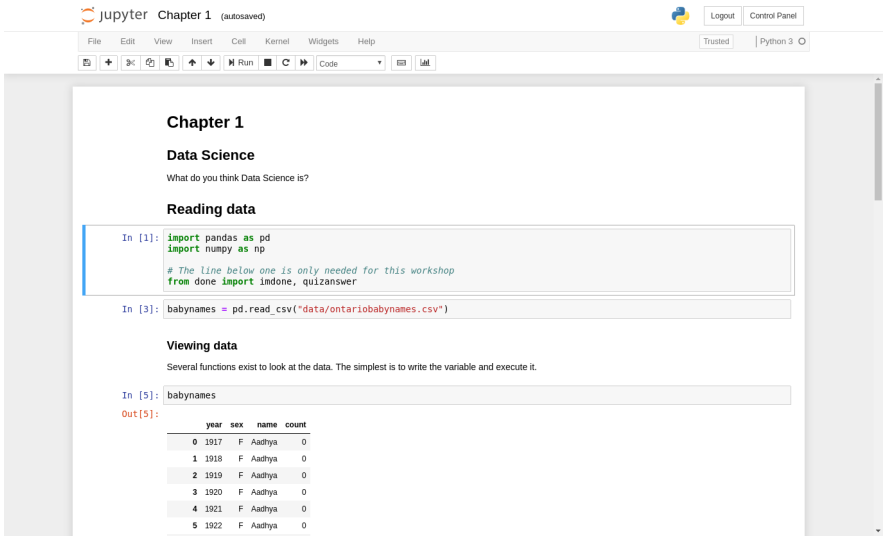


Data Science



Reproducible Science

- Programs are reproducible
- Using Jupyter Notebook
- Mix between program and report
- Takes data and reproduces conclusions



Python

- High-level
- General-purpose
- Interpreted
- Huge ecosystem

Python – The Language

- Values: -1 1.3 "Ottawa"
- Object: n = -1 x = 1.3 city_name = "Ottawa"
- Functions: print(n) len(city_name)

- Special values:
  - Lists: [1,2,3,4] ["Hello", "world"]
  - Dictionaries: {"num": 12, "name": "something"}

Warm Up

- Which of these are numbers?

1 "1" "one" one

Warm Up

- Which one of these will work? (assume one=1)

log(1) log("1") log("one") log(one)

Speed test

- Numerical integration of $f(x,y) = x^2 + x\,y + y^2$ from -10 to 10 with steps of 0.001.
- Python

```
Integral of f(x,y) = x^2 + xy + y^2 is 66.67333333332641

real    1m44.903s
user    1m44.866s
sys     0m0.004s
```

- C

```
Integral of f(x,y) = x^2 + xy + y^2 is 66.673333

real    0m1.097s
user    0m1.097s
sys     0m0.000s
```

Factor of 96 difference!!

Well, that sucks... but!

- The power of Python is in its ecosystem.
- Thousands of packages are available. Many written in C but usable from Python.
- Python is the glue to connect them all

Python Packages

- Python Package Index (PyPI)
  - https://pypi.org
- 233,536 projects
- Package installation depends on how Python was installed, most common is conda or pip.
- Our environment already has the most common packages installed

- Using packages with import

```
import math
print("sin(3) = ", math.sin(3))
```

Jupyter Notebooks

- Open your browser and go to
  - https://uottawa.syzygy.ca
  - Open Chapter 1.ipynb

Further Assistance

- For help with
  - Compute Canada
  - Programming
  - Servers
  - Any research-related computing needs
- Contact Jarno van der Kolk through TopDesk
  - https://topdesk.uottawa.ca
- Or by e-mail, jvanderk@uottawa.ca