1.1 Introduction

Applied Data Analysis (ADA)

Oxford DH Summer School - 2019

Reminder!

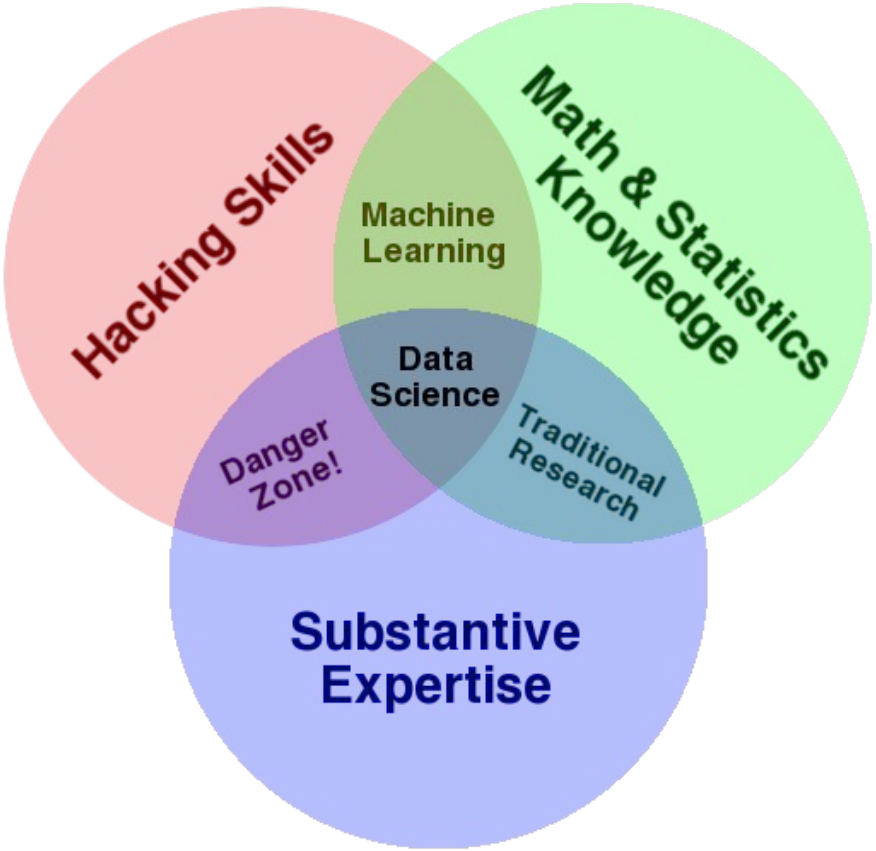
Pre-school survey:
https://forms.gle/RQjt5VssHabQKh2U7


© DHLAB B.Seguin-I.di Lenardo


Google
Books


INTERNET ARCHIVE


FAKE
NEWS
LinkedIn
foursquare
LIVEJOURNAL
vimeo
YouTube
facebook
Blogger
twitter
flickr
Like


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

What's in a title

Applied: we foreground techniques and methods in real-world scenarios, over implementations (but not theory!). I.e., we focus on what is done by a technique or method, rather than how [1].

Data: we use datasets which are too large to process (i.e., read) manually. We also consider datasets which are too large to be perused manually in their entirety without high risk of bias.

Analysis: we strive for insight. Data and tools serve little purpose without a motivation, question or information need, which should in turn help in creating new insight or knowledge.

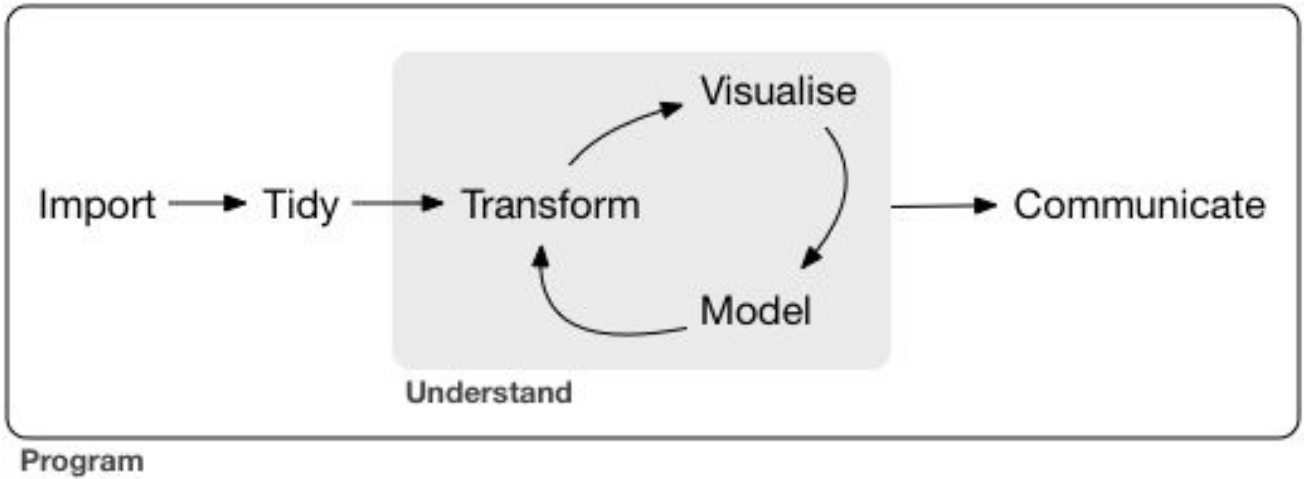
What's also there but not in the title

Humanities: we focus on data of interest to humanities scholars, professionals, practitioners.

Advanced: we assume some previous coding grounding on your side.

[1] http://dhdebates.gc.cuny.edu/debates/text/99.

How we think about ADA



See https://r4ds.had.co.nz/introduction.html

How we think about ADA

* Observational rather than causal evidence.

* Complementary and enriching rather than exclusive.

Topics

1. Python ADA skills
2. Import data
3. Tidy data
4. Wrangling
5. Visualizing
6. Modelling
7. Communicating

Schedule

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Introduction	Skills	Wrangling	Exploratory DA	Modelling
Data import	Tidy data	Skills + Viz	Visualizing	Adv. application
Afternoon session	*Afternoon session*	*Afternoon session*	*Afternoon session*	Communicating and wrap-up

Afternoon sessions

Options (attendees chose from the below):

* **Catching-up**: assistance is provided to clarify any issue from the previous classes or in setting-up your Python environment.

* **Exercises/project**: exercises or mini-projects will be provided for practice. Alternatively, attendees can bring their own mini-project to the class and work on it, individually or with others.

* **Lectures at Text to Tech**: attend the invited lectures given as part of the Text to Tech strand https://www.dhoxss.net/from-text-to-tech.

Datasets

* Tweets from Elon Musk (21st century, text as a time series)

* 19th-century books from the British LIbrary (metadata and text)

* Contracts of apprenticeship from Venice (16-17th centuries, numerical data)

** Early African-American film database (20th century, metadata)*

** Network of crypto art transactions (21st century, network data)*

for afternoon sessions

Teaching methods

Most classes are using **Jupyter notebooks**: interactive snippets of code and comments. Several short or not-so-short **assignments** are there for you to engage with. You should try to **play with code and data** as we go along, don't just execute our code.

Some classes use slideshows or the board.

Questions and comments are encouraged.

One of us will always move around: use **post-its** to signal if you have an issue (orange) or not (green, or something).

Afternoon sessions are for you to decide what to do, with us moving around to help.

Please come to us with comments and feedback at any time during the week: we can always improve as we go.

Introductions

Results of the pre-school survey

Tools and materials

Code and data are on GitHub: https://github.com/mromanello/ADA-DHOxSS2019

Slides are on Google Drive: https://bit.ly/2NV8gDJ

Collaborative question area: shared GDoc

Pre-school survey: https://forms.gle/RQjt5VssHabQKh2U7

Let's try these out! https://mybinder.org/v2/gh/mromanello/ADA-DHOxSS2019/master

Twitter: #DHOxSS2019 and #ADA

Set-up and warm-up

- Launch Binder from the repo (it might take a little while)
- Go to /notebooks
- Open the *HelloWorld* notebook
- Play along and see if it’s all sound and clear, use post-its to signal if there is any issue

If you want to work locally, you are welcome to fork the repo and use your own copy.

We can help setting you up during lunch time and during the last afternoon session.