1.1 Introduction

Applied Data Analysis (ADA) laboratory
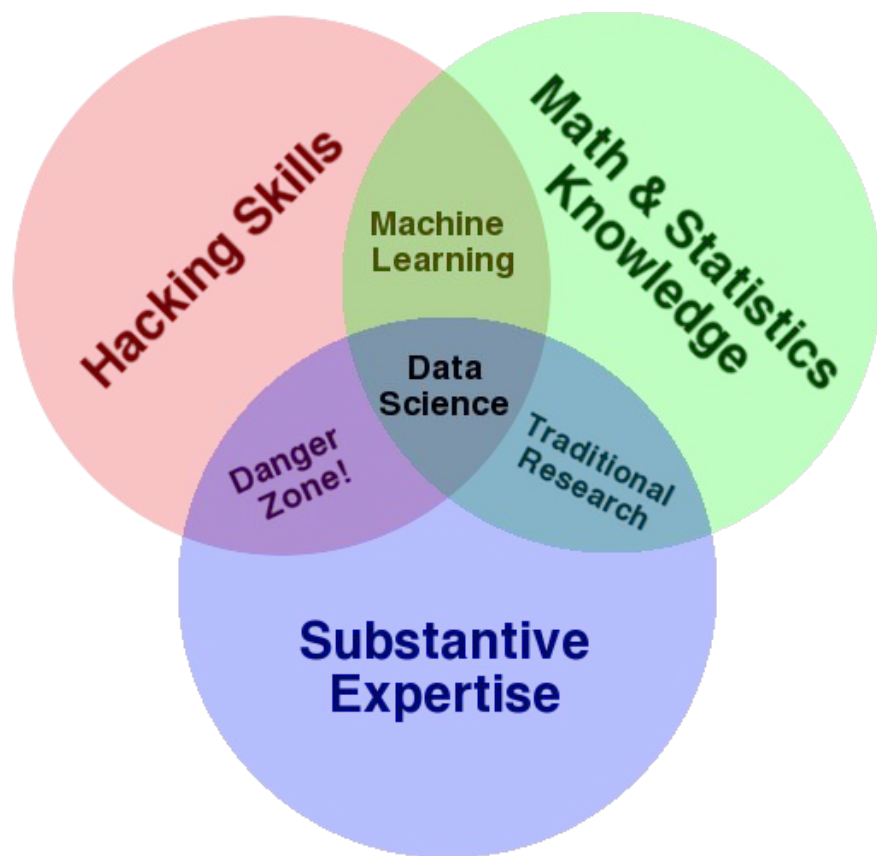
DHDK UniBo - a.a. 2023/2024

INTERNET ARCHIVE

Google Books

WIKIDATA

JSTOR
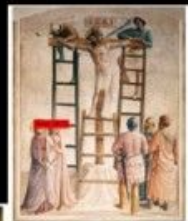
What's in a title

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.

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

How we think about ADA
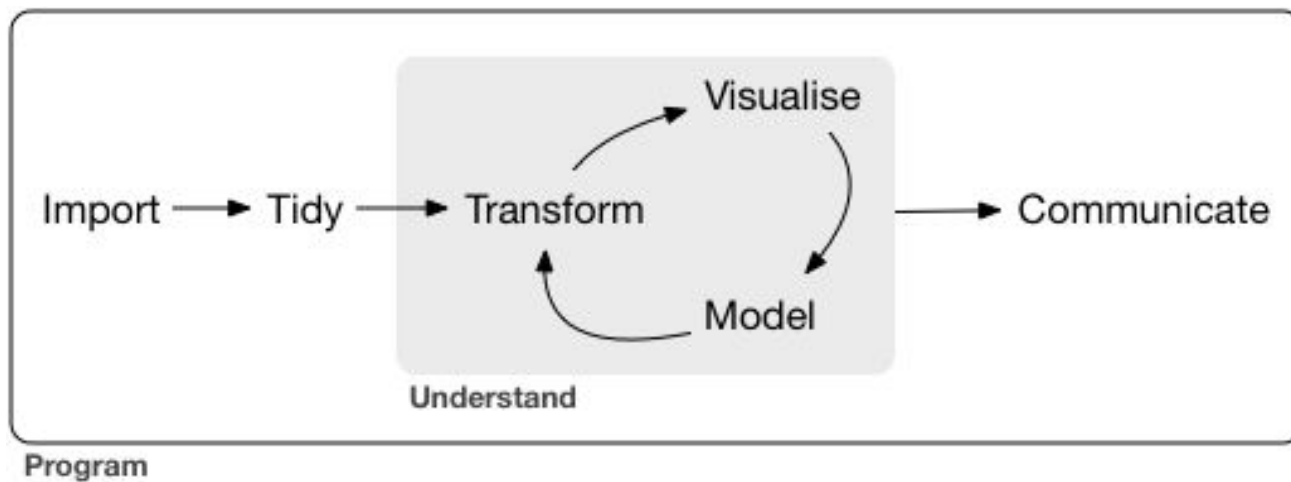
How we think about ADA

* Observational rather than causal evidence.

* Complementary and enriching rather than exclusive.

Course overview

Week 1	Week 2	Week 3	Week 4	Week 5
Introduction to ADA	Tidy data	Exploratory DA	Data visualization	Publishing code/datasets & communicating results
- course intro - intro to 🐼 - data carpentry	- basics of TD - data wrangling - operations on tidy data frames	- descr. statistics - distributions - plotting with Pandas, matplotlib and seaborn	- explor. data viz. - data viz best practices - geo mapping - network analysis	- dataset construction - communicating analysis results - best practices for publishing data/code

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)

 Sample of historical British newspapers (19th century, metadata and text)

Want more? Have a look at *[Humanities Datasets in Context](#)* by Humanities Computing at Princeton.

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 highly encouraged.

There is a **Slack channel** to facilitate communication both with me and among you (e.g. for group project work).

Please come to me with comments and feedback at any time: things can be improved as we go.

Assessment methods

1. **Group project** (50% of final grade)
   - 2-3 people max.
   - final presentation of all group projects
   - projects can cover several (or all) steps of an ADA pipeline (data cleaning, exploratory data analysis, modelling, static/interactive visualizations)
   - dataset: existing, new (ad-hoc), existing but unpublished

2. **Oral exam** (50% of final grade)
   - will address both course content and group project

Round of introductions

Examples of data analysis applications

Example of ADA: Mining Classics Citations from JSTOR

Mining of canonical references from Classics articles in JSTOR, to study scholarly reception of classical texts via citations as a proxy for scholar's attention.

Pre-print: https://doi.org/10.5281/zenodo.3736455

# Data processing: mining canonical citations

Example of canonical references

first 'block' of arrival scenes in the Telemachy (*Odyssey* 1–4): the arrival of Athena-Mentes at Ithaca (Hom. Od. 1.103–324):⁹ the goddess appears on the threshold of the palace (1.103–4) where she finds the suitors engaged in their respective activities (1.106–12). She is then seen by Telemachus (1.113, 1.118), who rises to accommodate the disguised goddess (1.119–20), takes her by the hand (1.121), and makes her enter (1.125); he welcomes her (1.122–4), takes her spear (1.121 and 1.127–9), and invites her to sit down (1.130–2); the dinner is prepared (1.136–43), consumed (1.149), and concluded (1.150); the visitor's identity is finally revealed (1.169–93) and information is exchanged (1.194–305) before Athena escapes Telemachus' attempt to retain her (1.309–19). This 'classic' hospitality scene highlights Telemachus'


https://citedloci.org/
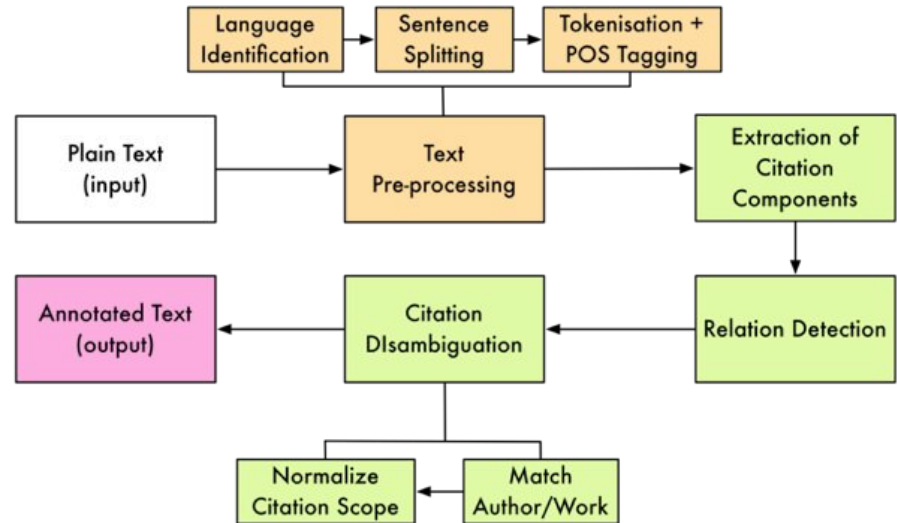
Data overview: Classics articles in JSTOR
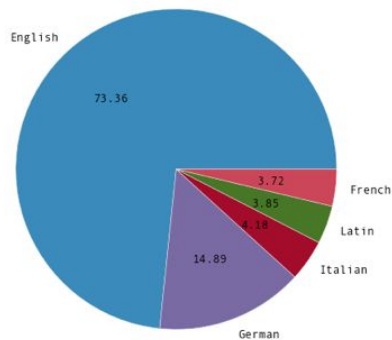
	Number
Total articles	138,821
Successfully processed articles	119,723
Sentences	34,853,399
Tokens	865,075,857
Extracted canonical references	1,649,868
Extracted author mentions	1,448,163
Extracted work mentions	4,665

Table 2: Basic statistics about the JSTOR data.
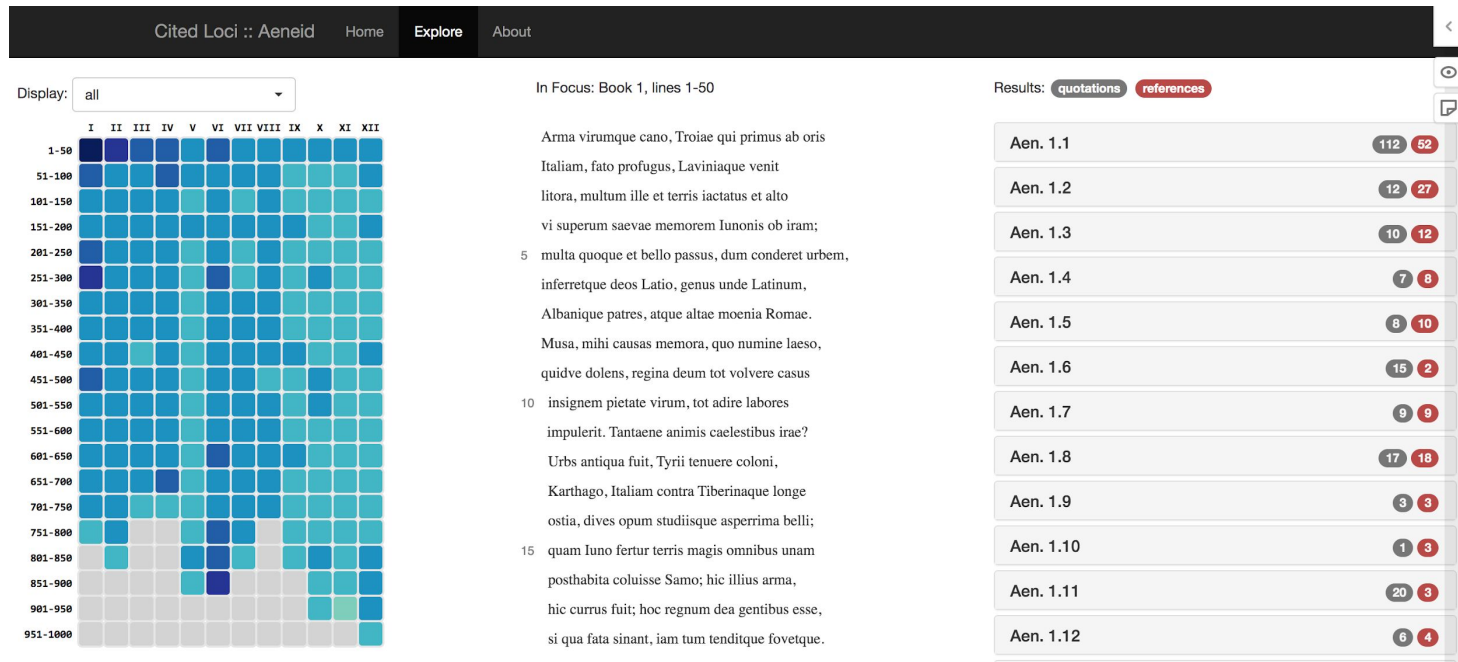

Number of articles per year in JSTOR (1900-2009)
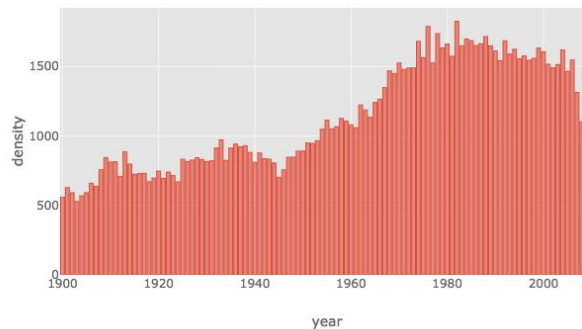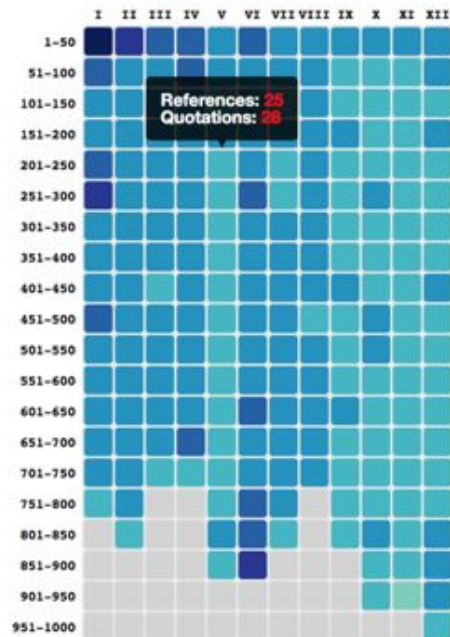

Number of canonical references per year
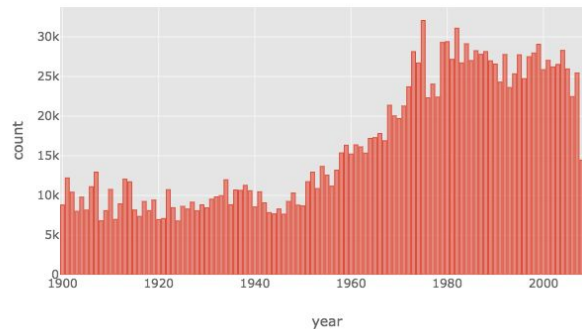
Interactive data visualization: *Aeneid* in JSTOR

https://mromanello.github.io/Aeneid_in_JSTOR/

http://labs.jstor.org/blog/cited-loci-of-the-aeneid/

Interactive data visualization: *Aeneid* in JSTOR

Distant reading

In Focus: Book 6, lines 251-300

    adventante dea. "Procul O procul este, profani,"

    conclamat vates, "totoque absistite luco;

260 tuque invade viam, vaginaque eripe ferrum:

    nunc animis opus, Aenea, nunc pectore firmo."

    Tantum effata, furens antro se immisit aperto;

    ille ducem haud timidis vadentem passibus aequat.

☞   Di, quibus imperium est animarum, umbraeque silentes,

Close reading
(primary literature)

Results: quotations references

Aen. 6.258	10 7
Aen. 6.260	2 1
Aen. 6.263	0 1
Aen. 6.264	8 13

Proserpina's Tapestry in Claudian's "De raptu": Tradition and Design
MICHAEL VON ALBRECHT
Illinois Classical Studies 1989
DOI: 10.2307/23064383

Besides the role of the Sibyl, Claudian also takes the role of Virgil, invoking the gods of the underworld, as his great predecessor had done (1. 20 ; Aen. 6. 264).

— (reference: 571a493ab634c1c699eed3fa)

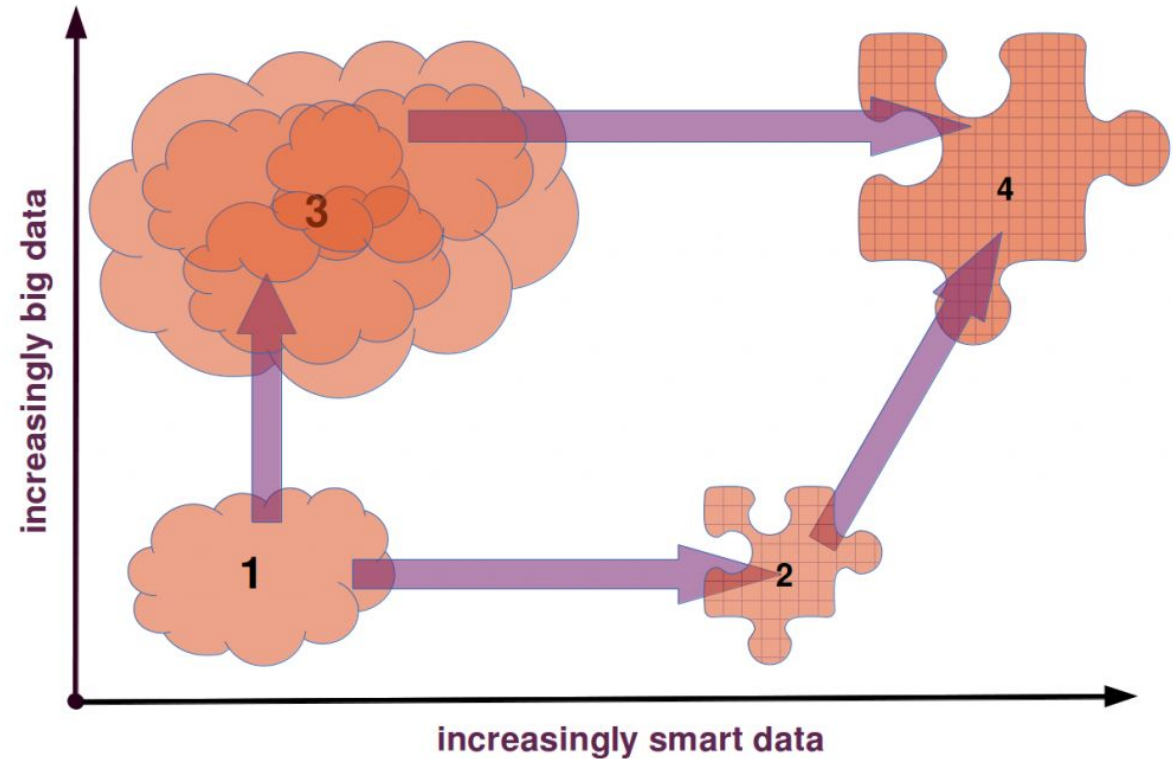**Close reading**
(secondary literature)

Longitudinal analysis: Waves of reception in Ovid and Vergil
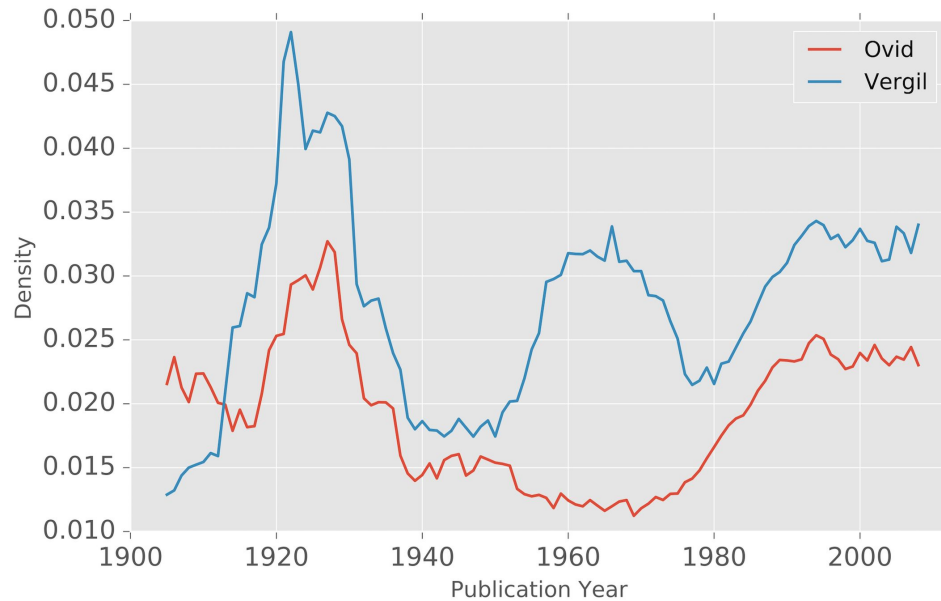
T. Ziolkowski 2009, *Ovid in the Twentieth Century.*

Conclusion

Links

Course description:
https://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2023/498219

All teaching materials are on GitHub: https://github.com/mromanello/ADA-UniBo

Slack channel: invitation link https://bit.ly/ada_unibo_slack_invite

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

Setup and warm-up

- Launch Binder from the repo (it might take a little while)
- Go to /notebooks
- Open the *1.1 Skills Python* notebook
- Play along and see if it's all sound and clear

If you want to **work locally**
- you are welcome to **fork** the repo and use your own copy
- make sure to pull changes from upstream at least once a week (best on Wed)