Information Visualization

Course intro

Lesson 1

Marilena Daquino
Assistant Professor

Department of Classical Philology and Italian Studies

marilena.daquino2@unibo.it

Table of contents

01

Introduce yourself

What are you interested in (topic-wise) and what you want to take home form this course (data analysis, visualization, web tech, communication skills…

01

What it is about

Data analysis

Methods to query, process, analyse data

Data visualization

Concepts and techniques to plot data

Web communication

Presentation techniques for explanatory projects

01

What you'll learn

Data analysis

Manipulate Linked Open Data with Python

Data visualization

Select charts and plot data with Python and Javascript

Web communication

Present your results with appropriate narratives

02

Background

What I wish you knew and what you will (hopefully) get from this course

Background

02

What I wish you already knew

Comp Think

Python (intermediate)

Install libraries, Jupyter notebook, read/write CSV and JSON data

Web tech: UI / UX

HTML, CSS, JS (good)

JQuery for UI, modify DOM, interactivity

Introductory methods

GitHub (good)

A short introduction
Github guides
sourcetree GUI

Knowledge mgmt.

RDF, SPARQL, OWL (good)

Read RDF in several syntaxes, read/write SPARQL queries, understand basics of ontologies

Background

02

What I will show you

Comp Think

Python and Jupyter

Libraries for data exploration.
Jupyter to document your work

Web tech: UI / UX

JS (good)

Libraries for data visualization,
Digital storytelling strategies

Introductory methods

Github, Colab and Binder

Publish your work on github (data, software and website)

Knowledge mgmt.

RDF, SPARQL, OWL (good)

Python APIs for RDF/SPARQL

Classes overview

09/11

Introduction to the course

10/11

Preliminaries on data viz

November						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

Classes overview

16/11

Introduction to RDFlib

17/11

Data access and SPARQL query

November						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

Classes overview

23/11

Data sense making

24/11

Data analysis

November						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

Classes overview

30/11 

Data visualization

1/12 

Digital storytelling

November / December

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	1		

Questionnaire

Classes overview

14/12

Additional methods

15/12

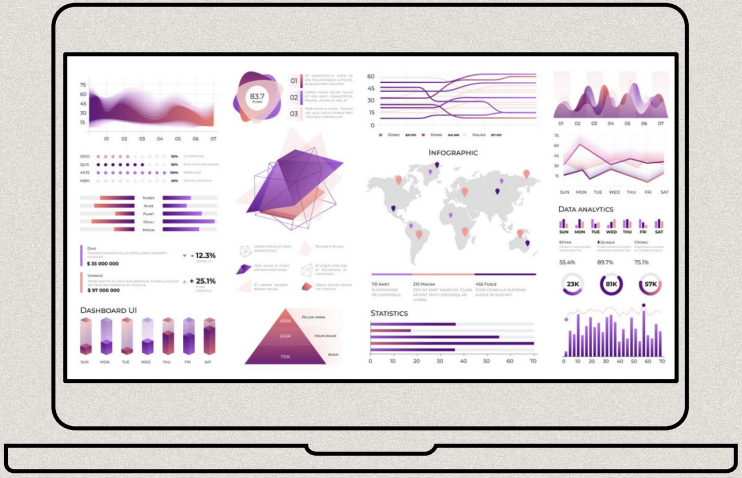
Publication,
review, and wrap up

December

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15		

03

What you'll do



The project

Choose a **topic**, find data-driven **questions**, **analyse** and **visualize** data, produce a **notebook** with your code, and a **website** for presenting results.

03

What you'll do

MUST

Topic

Pick a domain, and quantitative questions

Data

Find or create data that support your analysis

Group

3-people group to share the pain

03

Topic selection

The course will make use of examples from the History of Art. The topic is free as long as you are able to find good questions, data, and a group by your own.

Topics

03

Some suggestions

Art history

Artworks provenance
E.g. What are the artworks that travelled most in Europe?

Iconography
E.g. What subjects are mostly represented in artworks of 16th century?

Connoisseurship
E.g. What are the most reliable criteria to justify an artwork attribution?

History of Art history

Art historians' relations
E.g. Which countries are historians from and where did they work?

The debate on research topics (artists, periods, movements)
E.g. In which periods the research focused on Mannerism?

Resources
E.g. which types of resources in archival collections represent research topics?

Topics

03

Some suggestions

Photography of Art

Photographers' relations
E.g. Which photographers worked in Italy in 20th century?

The most photographed artworks
E.g. What art genres are mostly represented in professional photography?

Archives and museums
E.g. Which museums commissioned photographs of artworks?

Gender in Arts

Representativeness of female or non-binary gender in history
E.g. How influential were female photographers in the 20th century?
E.g. Are female photographers under-represented in photo archives?

Photographers' occupations
E.g. What other occupations had female art historians?

Women and market
E.g. Who buys women's art?

Create social value

You create social value by considering the economic, environmental, and social aspects that can impact people’s life, increasing their well-being and development.

For instance, in Arts social value is given by **the effects of artistic activities** not taken into account by the market: e.g. increased self-belief, self-empowerment

Ask for advice

Once you defined your research questions, drop an email to me **marilena.daquino2@unibo.it** for feedback and suggestions.

This will help your work not to be trashed right after the exam and may be useful to somebody in the near future.

03

The data

The course will make use of data from a few Linked Open Data for Art History and popular sources. Consider integrating multiple sources to answer your questions. You must use at least one Linked Open Dataset.

Data

03

Some suggestions

Artchives

http://artchives.fondazionezeri.unibo.it

ARTchives includes data created by cataloguers of art historical photo archives and reuses data harvested from Wikidata. Data includes information on art historians, archival collections, debated art genres, and keeping institutions.

Zeri Photo archive

http://data.fondazionezeri.unibo.it/

Zeri & LODE includes data created by the Zeri Foundation and reuses data harvested from Wikidata, DBpedia, ICONCLASS, AAT Getty, VIAF. Data includes information on artworks and photographs of artworks collected by Federico Zeri, one of the most notable connoisseurs of last century. The dataset is limited to artworks of Modern Art.

Data

03

Some suggestions

Wikidata

https://query.wikidata.org/

Wikidata is a general purpose Linked Open Dataset, originally born to represent structured data of Wikipedia (the right-side boxes) in RDF. It includes a variety of information, such as people biographical data, scholarly publications, historical events, and so on.

And more...

You can use ARTchives/ZERI data only or (**recommended**) integrate these with other data, Linked Data or not (e.g. Wikidata, DBpedia,artistorians.info).

Find other sources that best suit your goals, e.g. PHAROS https://vision.artresearch.net/spargl

You can work directly on other sources (e.g. Wikidata)

03

The Group

3 people max. Justify your contribution to the project. Grades are individual (you are judged for your contribution).

You can work alone, but you need to achieve the same results. No discounts :(

04

Evaluation

What to prepare, what happens the day of the exam, how I grade your work.

04

The D-Day

WHAT

Presentation

You present your project and show the **website** (15')

Dig into the code

I look into the **jupyter** notebook

Q&A

Discussion of methods, **results** and questions

04

The day before the D-Day

YOU HAVE

Presentation

15’ **presentation** (optional slides) according to a given template

Jupyter notebook

A **jupyter** notebook including the data processing and visualizations

Website

An **online webpage** presenting results of your work to a broader audience

Preparation checklist

Jupyter notebook

Install

Jupyter [1]

Notebook

1 python notebook

Abstract

Introduce the scope of the notebook

Data

Manipulate data via python and save results ad CSV/JSON

Markdown

Document functions and operations with markdown [2]

Clean up

Keep it short, group functions and imports, use titles and Table of contents

Preparation checklist

Website

HTML / CSS /JS

Create a static webpage, you can use templates

Data

Access your data from JS (CSV/JSON or APIs)

Visualize

Create charts with your data

Present

Add titles, sections, descriptions of charts, and discuss **results**

Preparation checklist

04

Github

Repo

Create a repo for the project

Upload

Upload the notebook, the website and the data

README

Add a README file with project title, people and resp., licence, badge Binder

Binder

Connect the notebook to Binder

Issue

Open or comment an issue called “Exam DD/MM/YYYY” with: Project title, Website URL, Repository URL, People [here]

Responsibilities and licenses

Pay attention!

Specify your tasks (e.g. data access and cleaning, data analysis, data visualisation, web development, graphics, communication strategy and so on) during the presentation

Specify your **names and tasks**:

- in the README file of the repository
- on the webpage (e.g. in the footer, in a dedicated page called Credits)

In the README file of the repository specify:
The **license** of data you reused (look at their websites)
The license of your derivative data (please, use either CC0 or CC-BY licenses to allow future reuse)

No panic

Pay attention!

- You can reuse existing **templates** for the website (e.g. CMS, HTML templates)
- You can reuse any **py/JS libraries**
- **Github** is a free-of-charge solution to publish a project website and to handle the code in the same environment. Feel free to use other solutions for publishing the website! However, it is mandatory for publishing the notebook and the data.
- We will set up all the pieces of your project (install Jupyter, use libraries, create an account on github, create and publish a webpage) in dedicated hands-on classes

Presentation template

Pay attention!

If you present with a slide presentation, please make it 10 slides max (No need to share it before the exam day).

Use the following template to prepare the **talk**.

1. Title
2. Background (the domain, the problem)
3. Goals
4. Research questions
5. Data preparation and data analysis
6. Data visualisations selected and reasons
7. Data communication strategies
8. Summary of results

An example

Pay attention!

Title: Trends in the study of artistic periods

Background. Artistic periods are differently studied by art historians over time. An artistic period may be overlooked at a certain time, possibly due to a decreased interest (market, research discoveries, exhibitions) towards some artist or genre.

Goals. Discover trends of interest towards artistic periods by looking at the dates of activity of art historians (and their archival collections) that studied a certain period.

Research questions.

1. **When** there is **more interest** towards a certain artistic period?
2. **How** this interest **evolves** over time?
3. **Which** artistic periods show a significant trend over time?

An example

Pay attention!

Data preparation and data analysis. We studied ARTchives and Wikidata data models. Data about archival collections, art historians' activity dates, and related artistic periods are collected from ARTchives. Descriptions and dates of artistic periods are collected from Wikidata. We query ARTchives and Wikidata SPARQL endpoints, we reconcile the data, we prune duplicates (e.g. "Baroque" and "Baroque art").
We perform some preliminary analysis to understand the distribution of periods over the archival collections.
We analyse the trend of artistic periods as subject of art historians' collections over time.

Data visualisations selected and reasons. We show trends in a line chart having on the x axis the dates of art historians activities (corresponding to the time range of their collection) and on the y axis the artistic periods. So doing a user can see all trends at the same time and figure if there are correlations in trends.

An example

Pay attention!

Data communication strategies. We first show preliminary exploratory visualisations about the distribution of periods as subjects addressed by art historians so as to demonstrate the validity and representativity of results. We provide brief descriptions of artistic periods for whom may not be acquainted with art history. We show results of our investigation as an interactive line chart where users can select/remove the period to be shown in the graph.

Summary of results. Most significant results show that artistic period XX received lots of attention in the 19XXs, while period YY shows a significant loss of attention in the same period. Notably, period XX and ZZ have similar trend over time.

*[TO EXPLAIN **WHY** THIS HAPPENS IS NOT UP TO YOU - if you are not an art historian]*

04

My grades

WHAT

Data

Correct and efficient SPARQL queries

Correct final CSV/JSON data

Data integration with multiple sources

Questions

Soundness and usefulness of research questions

Use of adequate visualizations

Graphic skills

30

Presentation

Clarity of the presentation during the exam

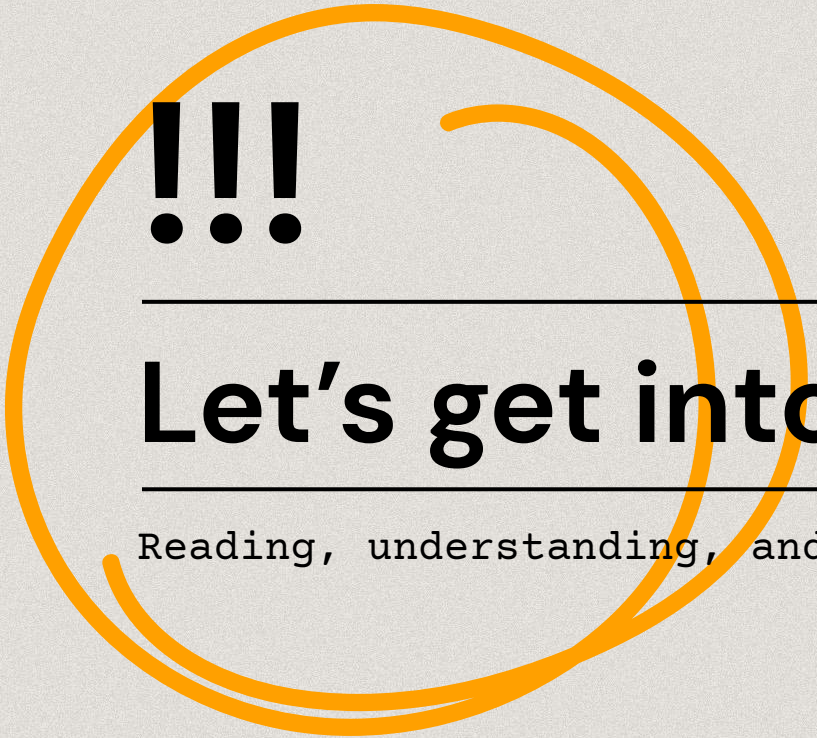
24

Ability to summarise complex issues

28

Surprise me!

30L



Let’s get into the spirit

Reading, understanding, and answering questions

A gentle introduction to EDA

!!!

EDA

Read (20′)

Read this article (now or at home)

Explore (30′)

Visit ARTchives and try to figure an answer for the questions

Answer

Fill in the form with your answers

Take away message (If you run out of time)
The preliminary questions when exploring a new dataset:

1. What question(s) are you trying to solve (or prove wrong)?
2. What kind of data do you have and how do you treat different types?
3. What's missing from the data and how do you deal with it?
4. Where are the outliers and why should you care about them?
5. How can you add, change or remove features to get more out of your data?

Assignments

Pay attention!

A few **questionnaires/exercises** will be given to you over time. These are primarily meant to give the teacher an estimate of your general understanding.

They are **not mandatory**, you should not be afraid to answer wrong (this won't be used against you).

We will **review** results at the end of the course, so please fill in all the forms you want **by November 30.**

Assignments

Pay attention!

However...

Remember the final presentation lasts 15 minutes, and it's the first moment I get to know you closely (which can be either good or bad).

If I knew you had **good** results in the questionnaires and you gave a **good** presentation this may encourage me to give you the maximum grade.	If I knew you had **good** results in the questionnaires but you gave a **bad** presentation, I'd take into account your constant effort and you'd not be disadvantaged.
If I knew you had **bad** results in the questionnaires but you gave a **good** presentation this would highly impress me and I'd tend to be more generous.	If I knew you had **bad** results in the questionnaires and you gave a **bad** presentation, well...at least you tried!

Thanks!

Do you have any questions?

marilena.daquino2@unibo.it

https://github.com/marilenadaquino/information_visualization