Data Science and Interactive Visualization 2

Analysis and Visualization of Big Data

Franziska Peter and Josep Perelló

UNIVERSITAT DE BARCELONA

Contents - Creating and editing graphics

Session II

Scientific Plots in Practice. Static plots for scientific publication. Python plotting libraries.

Sessions VI

Data Science and Interactive Visualizations. Introduction Pandas.

Sessions VIII, (IX), XI, XII:

Selecting Databases. Developing visualizations. What is a dashboard? Streamlit. | Bokeh, Plotly, Altair. Displaying Geographical Data. Dashboards.

UNIVERSITAT DE BARCELONA

Evaluation

Gradual and incremental set of tasks (in class and through Campus Virtual)

Task 1: Data Management Plan Forensics, in group (Tues 9, JPerelló): 10%

Task 2: Sharing code in Github, individual (Wed 10, FPeter): 10%

Task 3: Write an abstract (Mon 15, JPerelló): 10%

Task 4: Create a dashboard (Thu 18, FPeter): 30%

Task 5: Oral presentation, in group (Fri 19, JPerelló + FPeter): 40%

To set a group between 2 and 4. You will work together during the course.

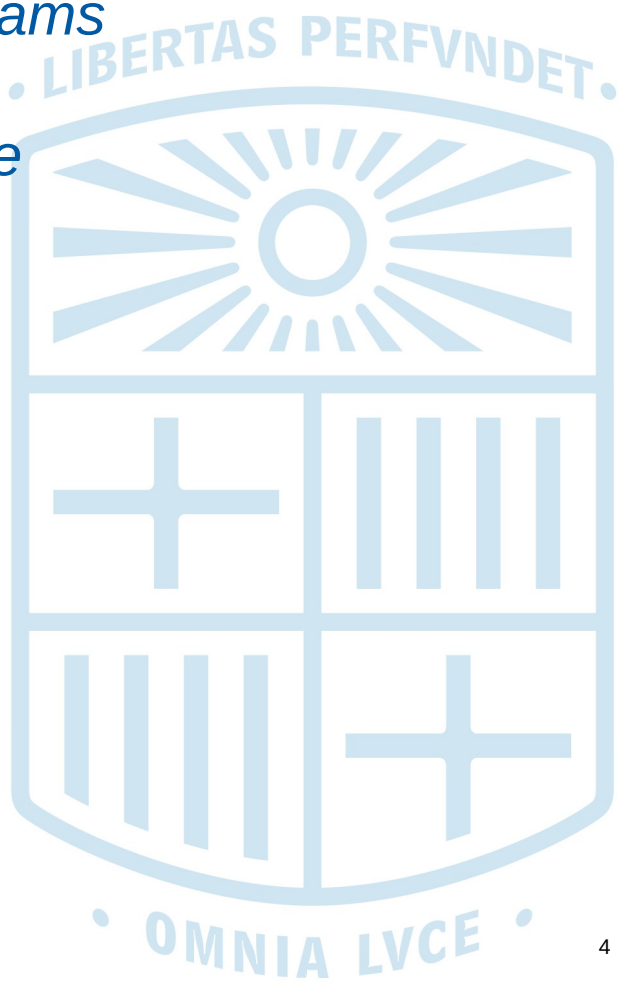
UNIVERSITAT DE BARCELONA

Outline

1. Sift your favourite data bases + Form Teams

2. Develop visualizations from the database

3. Dashboards and Streamlit



UNIVERSITAT DE BARCELONA

Some references

1. references from last session (dashboards)

2. Streamlit:
   - https://docs.streamlit.io/
   - https://pythonwife.com/introduction-to-streamlit/

3. Inspiration: https://www.informationisbeautifulawards.com/ and Josep's slides

UNIVERSITAT DE BARCELONA

Dashboard

Collection of interactive figures and diagrams on some specific topic

Dashboards are often required to:

- update themselves automatically
- be self-explanatory (depending on user community)
- be accessible online

UNIVERSITAT DE BARCELONA

Dashboard

Collection of interactive figures and diagrams on some specific topic

My favourite Dashboard is unfortunately in German..
https://www.zeit.de/wissen/corona-karte-deutschland-aktuelle-zahlen-landkreise

They also explain how the dashboard evolved and which programs they use. Unfortunately in German as well…
https://blog.zeit.de/fragen/2021/02/26/was-hinter-dem-corona-dashboard-auf-zeit-online-steckt/

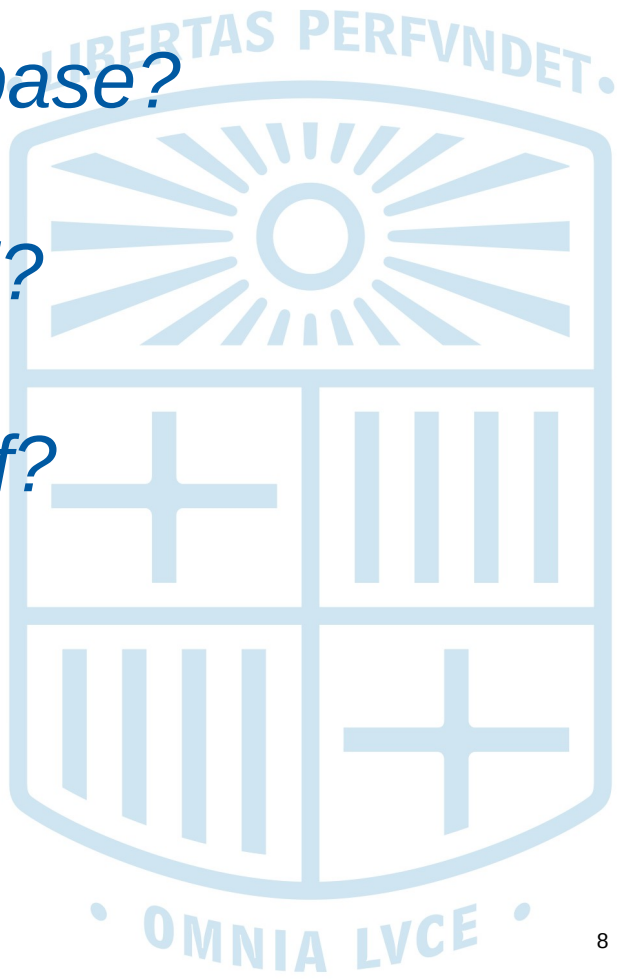
UNIVERSITAT DE BARCELONA

Your favourite Databases

? What is the topic of the data base?

? Which columns are contained?

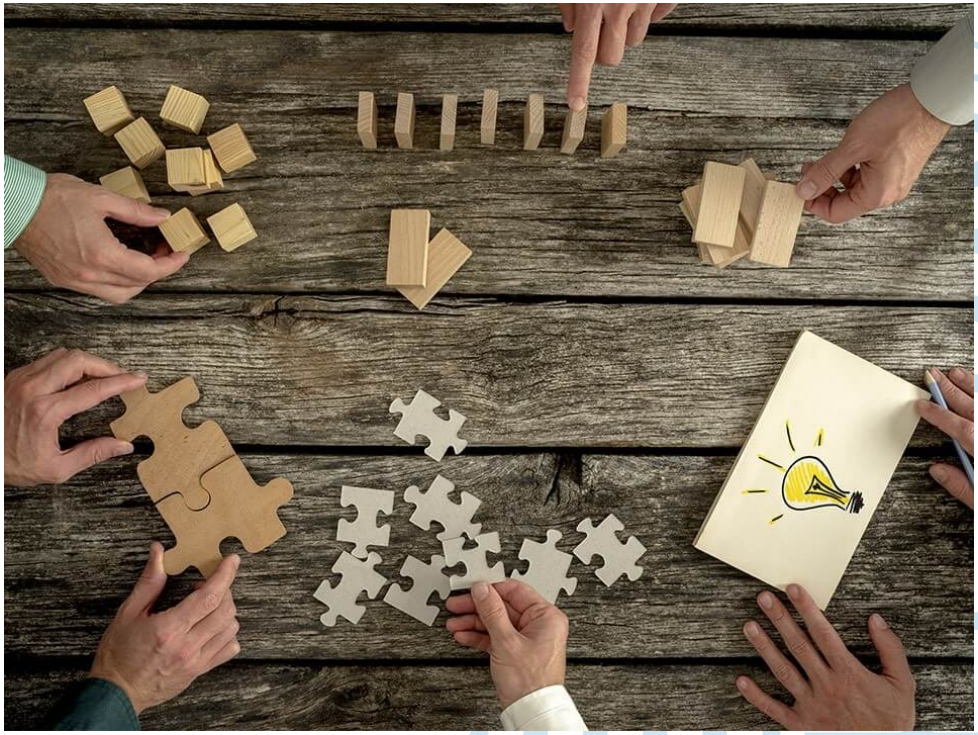
? Which plots could we think of?*



UNIVERSITAT DE BARCELONA

We need two teams, and in each team we need:

* an editorial board (2):
  - coordinates what will end up on the page
* task coordinators (2):
  - coordinates who does what when
* a representative (1):
  - communicates issues and progress
* a documenter (1):
  - writes down work progress
* git responsible (2):
  - make sure all code is bundled in one repo



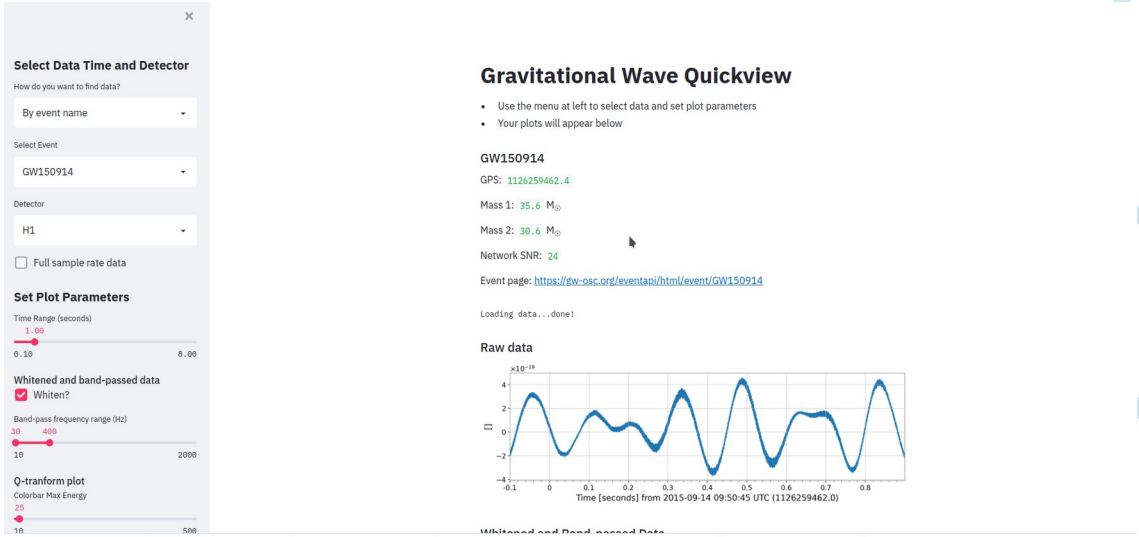
… and of course, all of you are skillful programmers and everyone should contribute some visualization(s).

UNIVERSITAT DE BARCELONA

Streamlit

= one and maybe the simplest way to create a dashboard with python

Example:
https://share.streamlit.io/jkanner/streamlit-dataview/master/app.py/+/



Streamlit

= *one and maybe the simplest way to create a dashboard with python*

other python dashboards:

- plotly dash
- holoviz panel

UNIVERSITAT DE BARCELONA

Streamlit

Installation:

```
pip install streamlit  or  conda install streamlit
```

Execution:

```
streamlit run some_app.py
```

Quick tutorial:

```
streamlit hello
```

Gallery:

https://streamlit.io/gallery

Documentation:

https://docs.streamlit.io/library/api-reference

Cheat Sheet:

https://github.com/daniellewisDL/streamlit-cheat-sheet

UNIVERSITAT DE BARCELONA

Streamlit / basic commands

```
import streamlit as st

# Text
st.write("Hello World!") # understands markdown

# Interactive widgets
st.selectbox("Choose!",("Option1", "Option2"))
st.number_input("How old are you?", value=18, step=1)

# Sidebar
st.sidebar.selectbox("Choose!",("Option1", "Option2"))

# Plots
from bokeh.plotting import figure, show
p = figure()
p.line([1, 2, 3], [4, 5, 6])
st.bokeh_chart(p)
```

UNIVERSITAT DE BARCELONA

Your streamlit projects: modular structure

Folder structure:

```
starting_with_streamlit/
....covid_vaccines_app.py
....plots/
........geo_plot_vaccines.html
........geo_plot_vaccines.py
........__init__.py
```

To compose the dashboard of several plots made by different student we **modularize** the code. Each module should run in the app and on its own.

streamlit App

File contents:

covid_vaccines_app.py

```
from plots.geo_plot_vaccines import plot_abundance_for_list_of_postal_codes
some_input = df["some filter"]
plot_abundance_for_list_of_postal_codes(some_input)
```

geo_plot_vaccines.py

```
def plot_abundance_for_list_of_postal_codes(some_input):
    """
    documentation!
    """
    return p

if__name__ == "__main__":
    some_input = [1, 3, 5, 7, 3, 5, 6]
    plot_abundance_for_list_of_postal_codes(some_input)
```

one of the plots

UNIVERSITAT DE BARCELONA

Until Tuesday:

* **create plots** from the database of your group
- under the premises presented by Josep
S1_Storytelling_Plots_v1.pdf slides 15, 17, 20, 28-34, 39: Graphical Excellence, Visual Complexity

Until Wednesday:

* go through the bokeh Turorial: glyphs, tools
Master_Visualizations_2021/Topic_3_DataScience_and_Interactive_Visualizations/interactive_plots/Dive into Bokeh.ipynb
* improve your plots according to the results of the Tuesday session and in communication with your group
* make your plots runnable as modules
* git responsible: share your code with Chaotique