Visualization Programming course

Master on Visual Tools to Empower Citizens

Universitat de Girona
2020-2021

Anton Bardera

Long Term practical work
Elaboration of a complete data science process to extract knowledge from a real data set

Working asset and delivery
You will work in groups of 3 or 4 persons and apply the contents of the course to the reference dataset. Each group works independently.
A report will be delivered by Friday, January 8th, 2021 at 12:00 M.
A public presentation of each work will be done on Friday, January 8th, 2021 at 17:00 PM.
Common discussion of the different approaches followed by the working teams on the same dataset will be held as the final part of your training program in this course.
In the following, details on the different aspects of this practical work are detailed.

Reference DATASET
For this practical work, you will work with COVID-19 data maintained by *Our World in Data*. Data consists of a different time series for COVID-19 tests, new cases, deaths, and other related information. A codebook for available information can be found as a CSV file at GitHub repository available at
https://github.com/owid/covid-19-data/tree/master/public/data
A second dataset is available on the statistics course with information of different (not all) countries (country-info.xlsx). The column location can be used as a primary key to join with the COVID-19 dataset.

The general workplan
This project is part of the global final project for this first trimester. As you know, in this final project you will have to create a descriptive analysis of this data and create 4 data mining methods (statistical part) and create a data system to maintain the data updated (data engineering part).

The visualization programming workplan
This visualization programming part consists of the implementation of different Observable notebooks that will show the statistical findings and will use the data provided by your database

system. First, you will have to create an exploratory tool to understand the multidimensional raw data that you have to use. Then, for each one of the statistical models that you have created, you have to produce a notebook explaining the model and implementing visualizations that support the explanation. Each notebook may contain one or more visualizations.

Deliverable

An Observable Notebook Collection with the 5 notebooks (1 for data exploration and 4 for model explanation). Each notebook will be composed of the following parts:

1. Introduction
2. Data preparation
3. Visualization and description
4. Conclusions
5. Implementation details
6. Appendix

The implementation details must describe the which have been the main implementation decisions, references that have been used, difficulties found during the implementation, and future work to improve the notebook.