

University of Trento --- M.Sc. Data Science

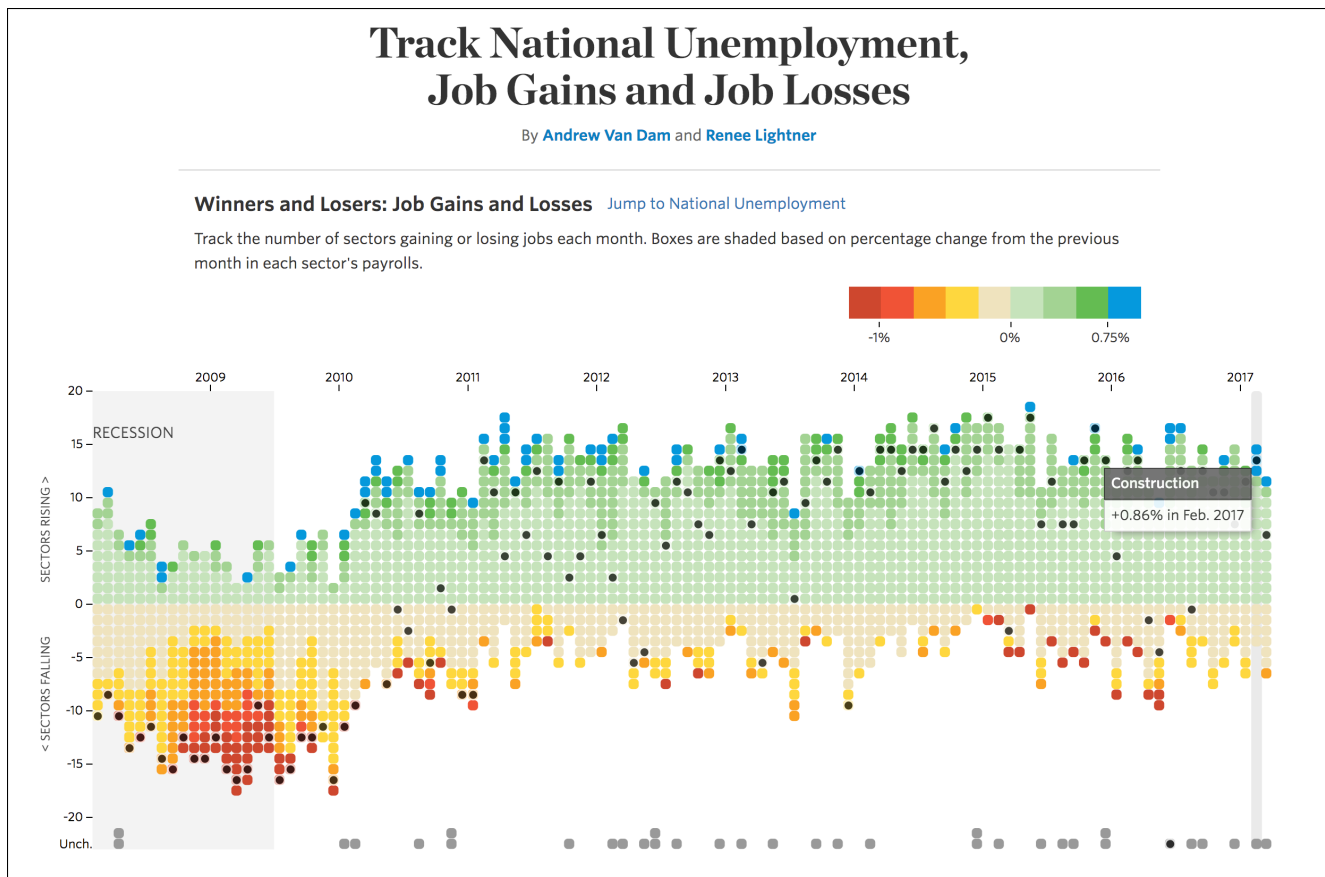
Data Visualization Lab Exam

10 July 2020

Create a Jupyter and/or R Notebook, named
name_surname.{ipynb,Rmd}

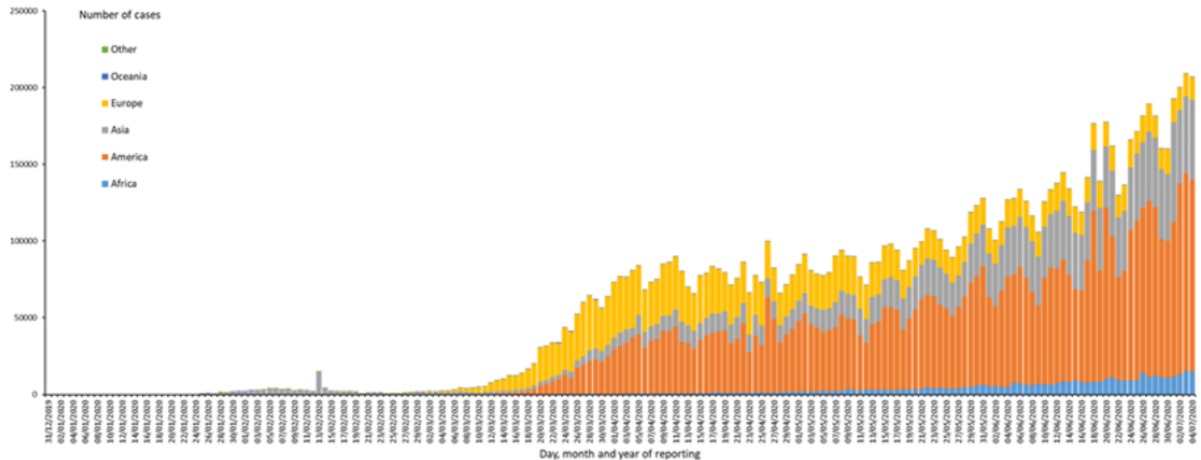
Answer the questions (in a Markdown cell/ as plain text) and solve the exercises listed hereafter:

1. [0-5 points] Describe in detail the elements of visual encoding (marks and annotations), their meaning and their use. Explain and comment the meaning of the visual encoding elements in the following picture.



2. [0-5 points] Discuss in detail the issue of *trustworthiness* in all three moments of data processing, data representation and data presentation, providing graphical examples for some of the discussed points (different from those included in the lecture notes).
3. [0-5 points] Describe the PCA algorithm, and comment similarities and differences with MDS, tSNE and UMAP. Then provide a detailed example where using PCA is more adequate than using the other three methods, motivating such choice.
4. [0-7 points] Using the datafile [covid.csv](#), prepare two maps showing the difference at country level of the Covid19 cases in Latin America on April, the 1st and on July, the 1st.
5. [0-7 points] Consider the dataset [parkinson.csv](#), including 756 samples (rows) described by 754 features (columns), plus the last binary column (*class*) indicating whether the sample is a patient (class 1) or an healthy control (class 0). Using a dimensionality reduction planar projection on the 754 describing features (using the method you rate as the most adequate), color the samples according to their class and discuss if the patients and controls can be separated into two distinct groups in the projection.

6. [0-7 points] Using the datafile [covid.csv](#), try to replicate the following plot, describing the distribution of COVID-19 cases worldwide, as of 4 July 2020.



Email the notebook(s) to giuseppe.jurman@unitn.it and please **wait for confirmation of correct receipt of the files before leaving the room.**

Notes:

- Exam is passed when at least 18 points are earned.
- If more than 30 points are achieved, the corresponding mark will be "30 cum laude"
- Use of the internet is allowed, but the candidate is expected to work individually.