

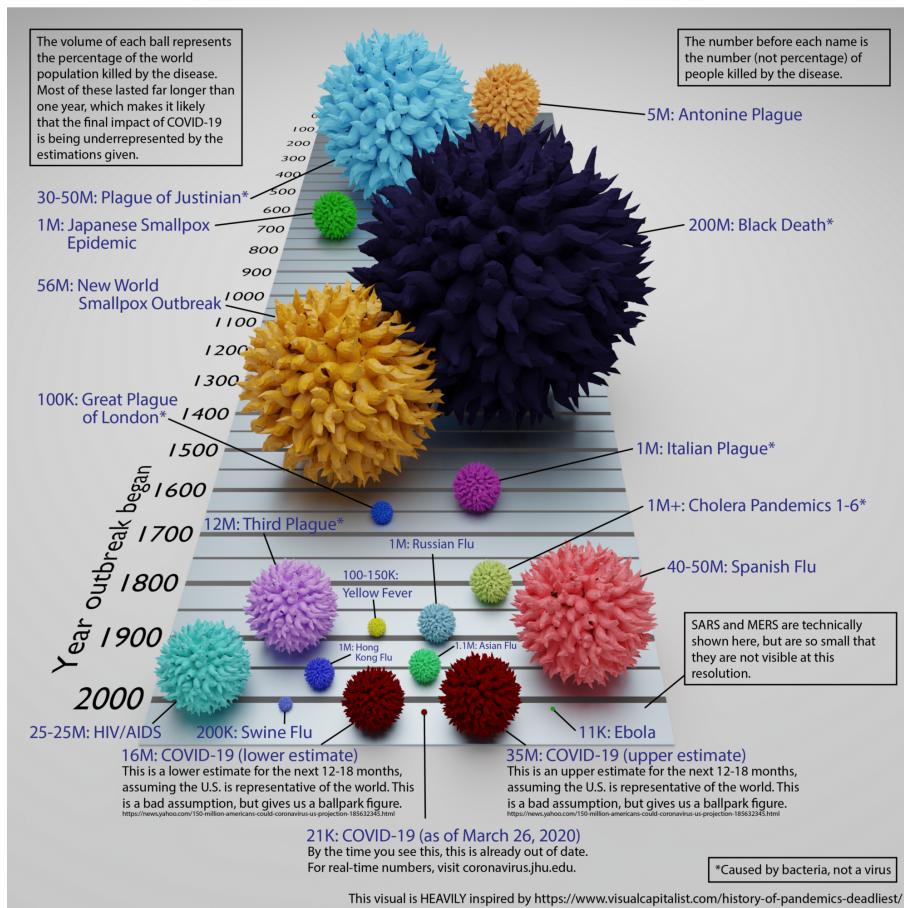
University of Trento --- M.Sc. Data Science

Data Visualization Lab Exam 11 January 2021

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 meaning of the visual encoding elements in the following infographic describing pandemics in history. Indicate also pros and cons of the infographic itself in terms of data visualization aspects.



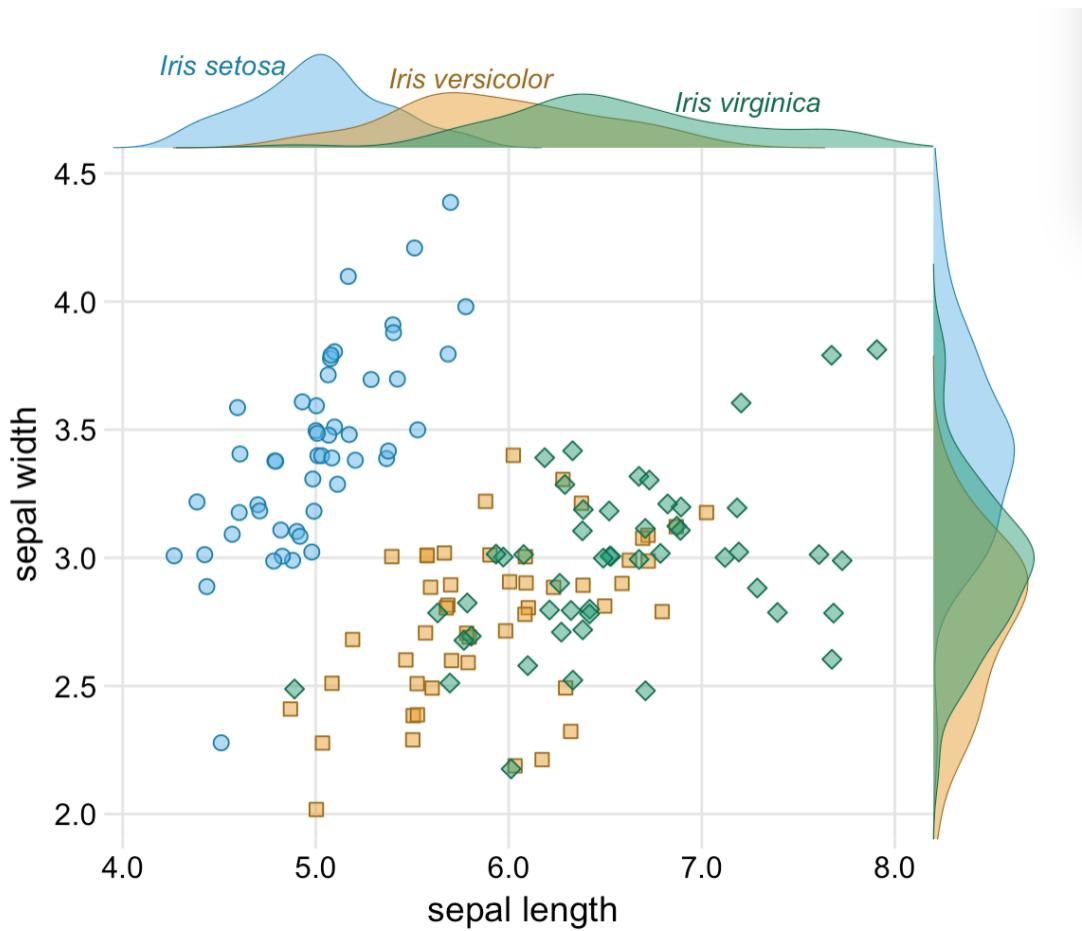
2. [0-5 points] Discuss in detail the issue of *project and chart composition*, providing graphical examples (different from those included in the lecture notes) explaining the underlying concepts.

3. [0-5 points] Describe the differences between PCA and UMAP algorithms in dimensionality reduction, and provide an example highlighting such differences.

4. [0-7 points] Using the datafiles [states_geo.tsv](#) and [states_vote.tsv](#), prepare a choropleth map of the US States including at least one socioeconomic feature from *states_geo.tsv* and an election-related variable from *states_vote.tsv*.

5. [0-7 points] Consider the datafile [Leukemia.csv](#), collecting the expression of 7129 genes of 72 leukemia patients, belonging to 4 different classes *B-cell*, *T-cell*, *BM*, *BP*, as indicated in the last column of the .csv file. Using the tSNE dimensionality reduction planar (2D) projection on the 7129 describing features, color the samples according to their class and discuss if the 4 groups of patients can be well separated into two distinct groups in the projection. Repeat the same process with the MDS algorithm and compare the two results.

6. [0-7 points] Using the datafile [iris.csv](#), try to replicate the following scatter plot with marginal distributions.



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.