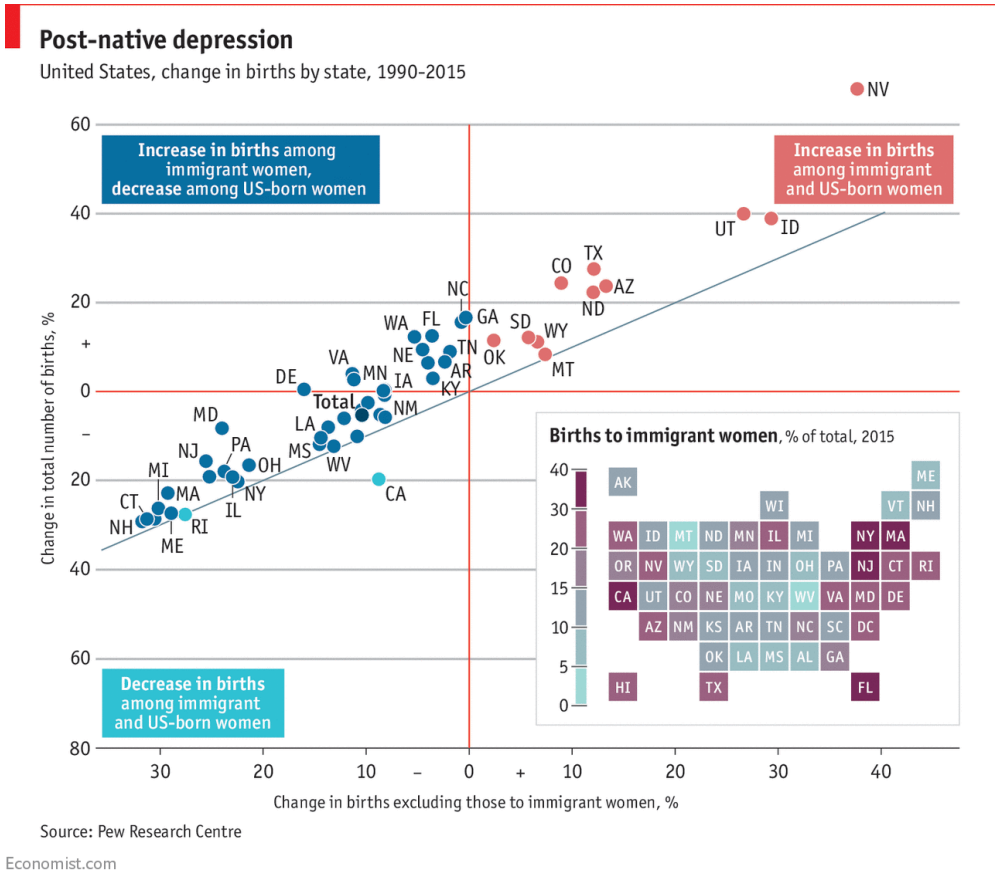
University of Trento --- M.Sc. Data Science

Data Visualization Lab Exam
08 February 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 incidence of post-native depression in the US between 1990-2015. 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 *accessible design* for a data visualization, providing graphical examples (different from those included in the lecture notes) explaining the underlying concepts.

3. [0-5 points] Describe the differences between tSNE and UMAP algorithms in dimensionality reduction, and provide an example (different from those included in the lecture notes) highlighting such differences.

4. [0-7 points] The datafile census.xslx collects the regional italian population data for the years 1951-2011. Using this data (or a subset of) prepare a data visualization composed of at least two panels, one being a choropleth map and the other being any statistical chart of your choice.

5. [0-7 points] Consider the datafile microbiome.csv, collecting the abundances of 6696 bacterial species (OTU) of 675 patients. Prepare a set of at least 5 plots with the 2D projection of the dataset by using the UMAP algorithm, varying the *number of neighbours* parameter in the range 5-200 and the *minimum distance* parameter and discuss in details the different shapes resulting in the projected plots.

6. [0-7 points] Using the datafile migration.xslx, try to replicate the following circos plot.

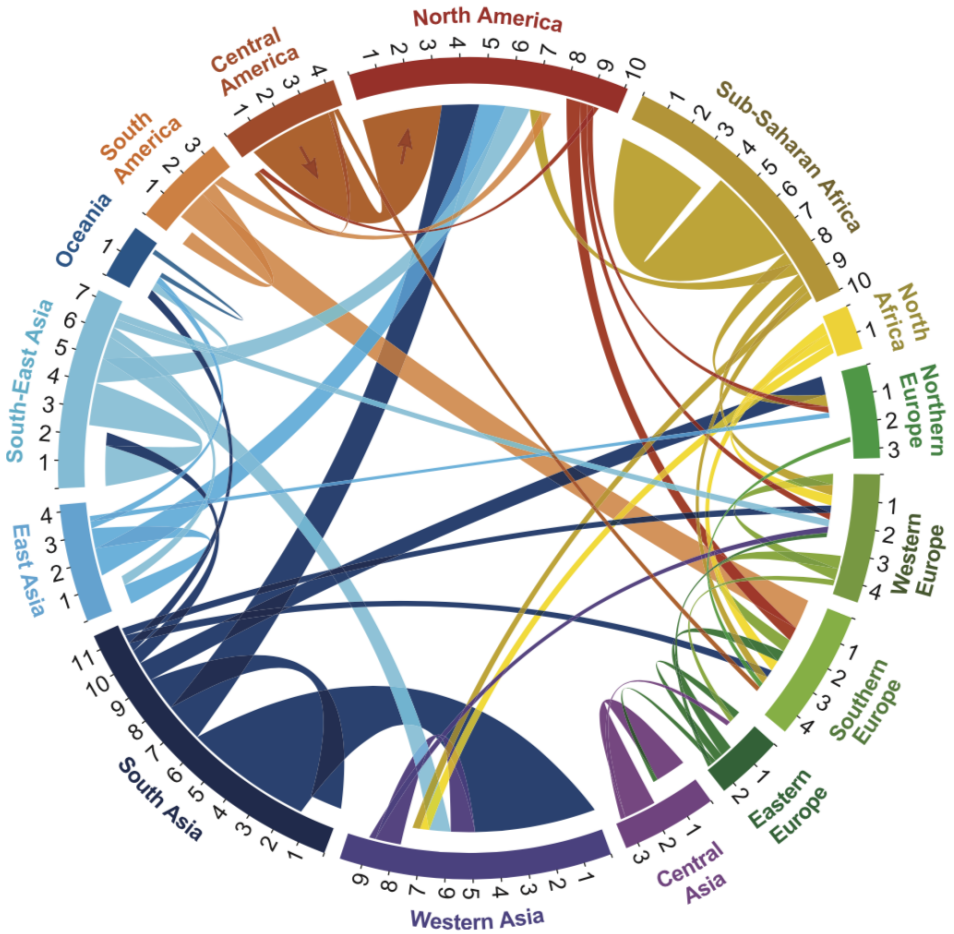


Fig. 4. Circular plot of migration flows between and within world regions during 2005 to 2010. Tick marks show the number of migrants (inflows and outflows) in millions. Only flows containing at least 170,000 migrants are shown.

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.