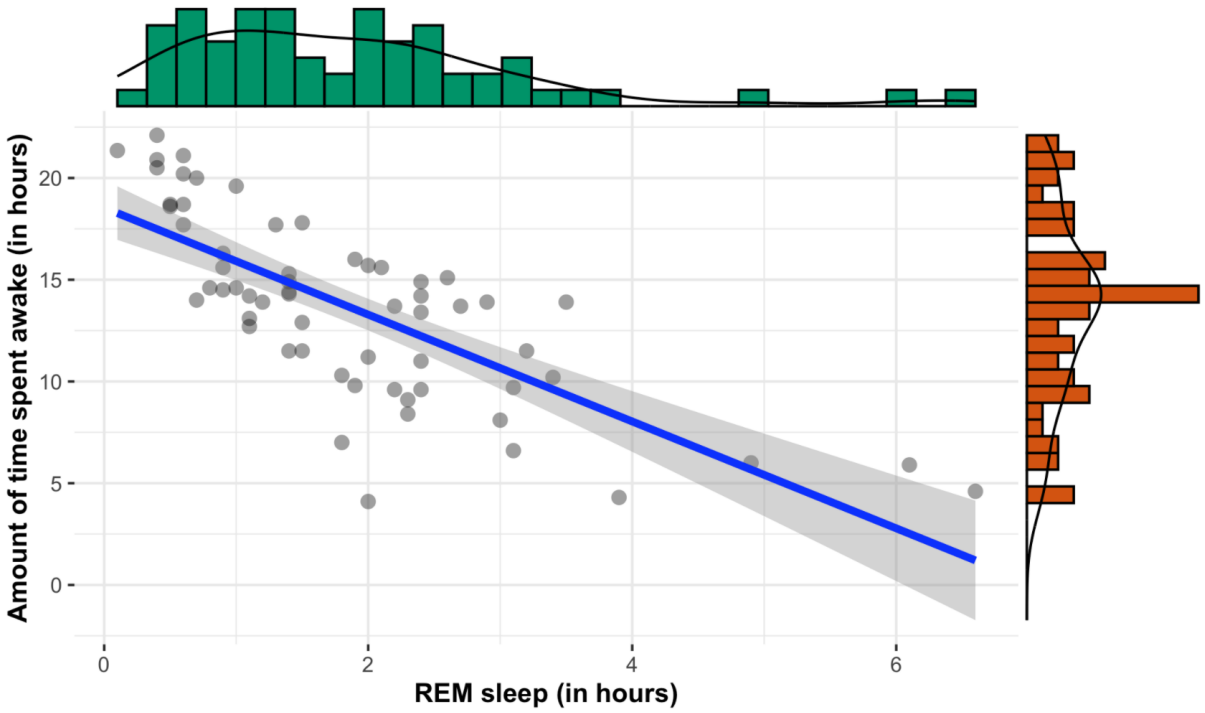
University of Trento --- M.Sc. Data Science

Data Visualization Lab Exam
09 July 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 reporting data concerning the sleeping time of different individuals. The blue line represents a linear model fitting the data.



2. [0-5 points] Discuss in detail the issue of chartjunks in infographics as opposed to useful decorations, and present two examples, different from those shown in the course slides.

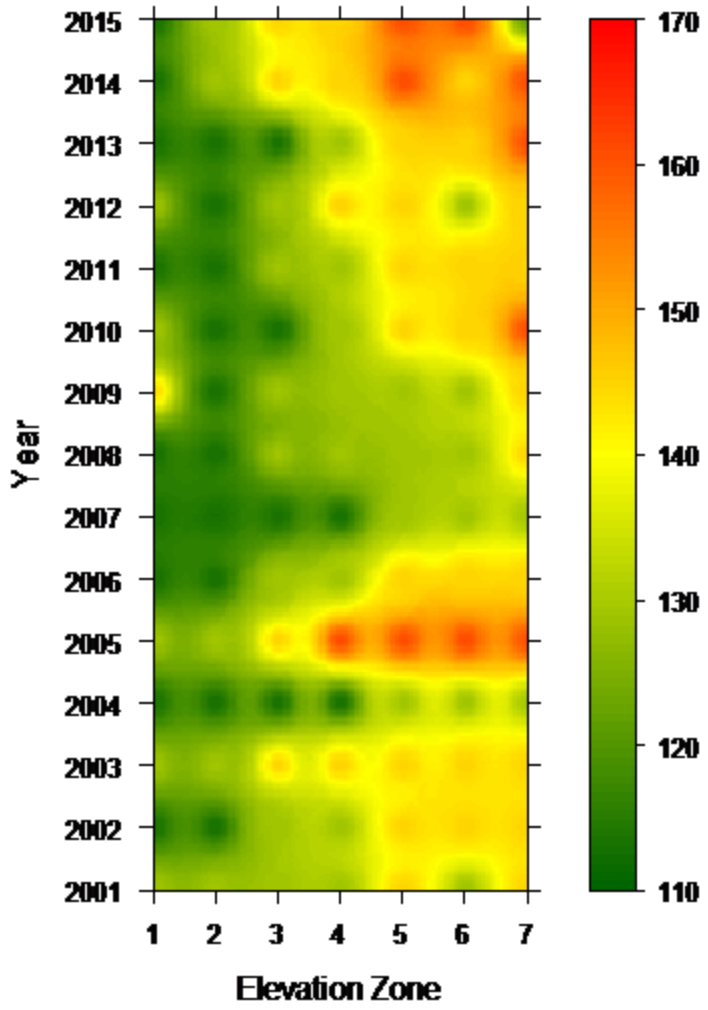
3. [0-5 points] Describe the different versions of MDS (classical, metric, non-metric) and provide some examples different from those shown in the course slides.

4. [0-7 points] The datafile socioeconomy_time_series.csv collects 4 indicators (inflation rate, unemployment rate, health expenditure rate as GDP percentage and public education expenditure rate as GDP percentage) for a number of countries in the time span 1960-2018. Using this data or a subset of, (*e.g.*, only the European countries) prepare a data visualization including
   - at least two choropleth maps showing the dynamics of one or more indicators in different years and
   - a (non-geographic) statistical chart of your choice.

   (Note that the dataset has several missing values.)

5. [0-7 points] Consider the datafile Colposcopies.csv collecting data about quality assessment of digital colposcopies for 287 individuals. Every colposcopy is described by 62 variables, while the last column ("*label*") classifies the colposcopy as of good (1) or bad (0) quality. Using the 62 describing features, prepare a 2-dim tSNE and a 2-dim UMAP plot, with the 287 samples colored according to their *label* and discuss what dimensionality reduction algorithm provides better separation between the two different classes of samples (0 and 1).

6. [0-7 points] Using the dataset heatmap.csv try to replicate the following plot.



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.