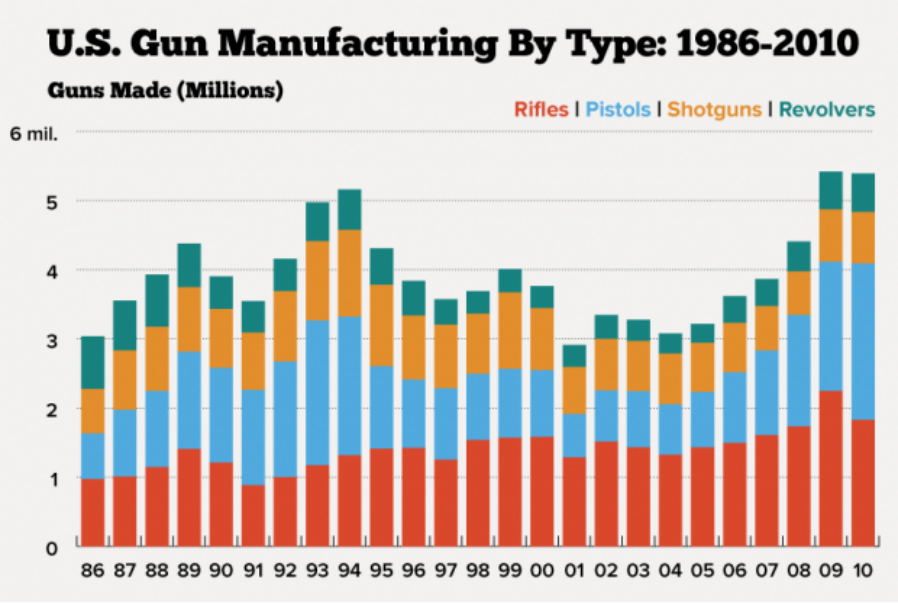
University of Trento --- M.Sc. Data Science

Data Visualization Lab Exam
10 January 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 three elements of graph understanding (perceiving/interpreting/comprehending), first in general and then applying these principles to the reading of the following plot



2. [0-5 points] Describe the *chrts* rule for graph classification and show and comment one example for each category.

3. [0-5 points] Describe the PCA algorithm, highlighting its pros and cons; find a case study where PCA is more appropriate than MDS.

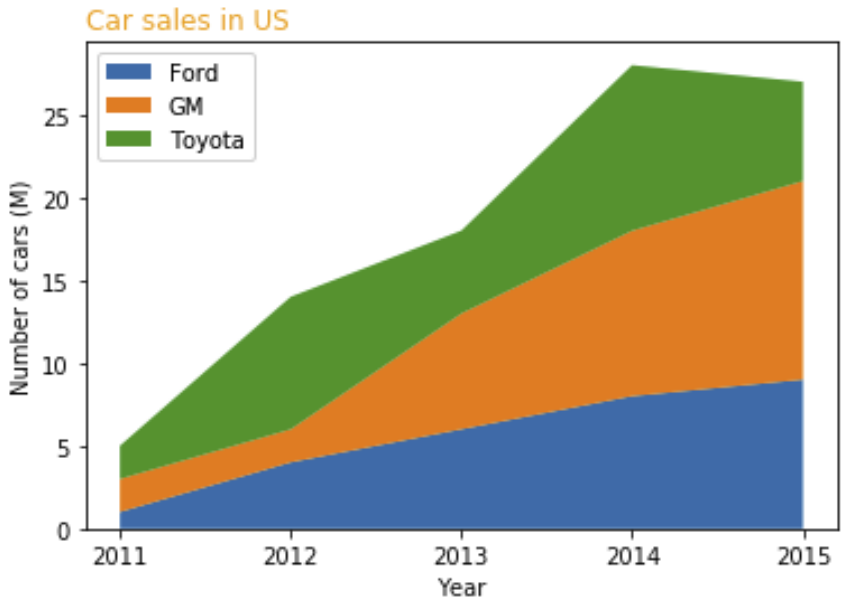
4. [0-7 points] Find a data visualization (one or more plots) for the dataset in https://archive.ics.uci.edu/ml/datasets/Forest+Fires (or a subset of the data); justify your choice and find/comment one or more non-trivial relations among the data visible in the plot(s).

5. [0-7 points] Using the dataset in Ex.4, consider variables *x* and *y* as coordinates and plot the heatmap of the average temperature in each (x,y) cell for the months of January and July. Then, using only the four columns FFMC, DMC, DC, ISI as variables, plot an UMAP embedding of the dataset coloring the datapoints according to the *month* column. Discuss the two obtained visualizations.

6. [0-7 points] Using the following data

	2011	**2012**	**2013**	**2014**	**2015**
Ford	1	4	6	8	9
GM	2	2	7	10	12
Toyota	2	8	5	10	6

replicate the plot



Email the notebook(s) to *giuseppe.jurman@unitn.it* and please wait for correct reception 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.