

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

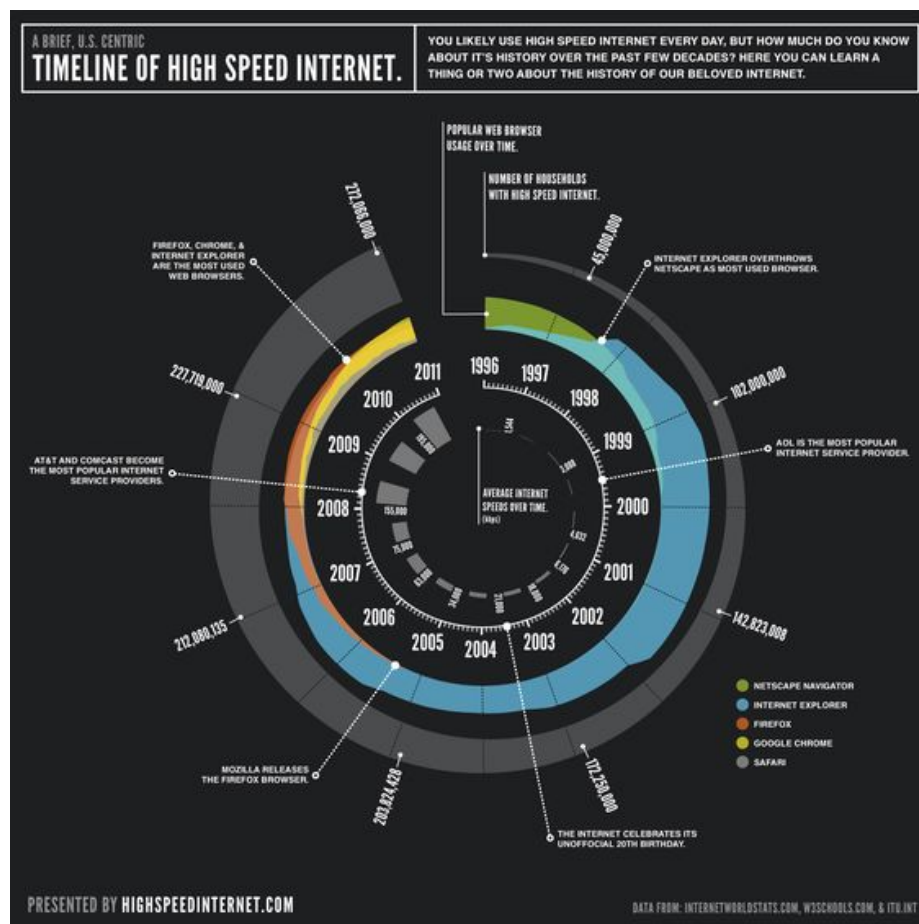
Data Visualization Lab Exam

14 February 2022

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:

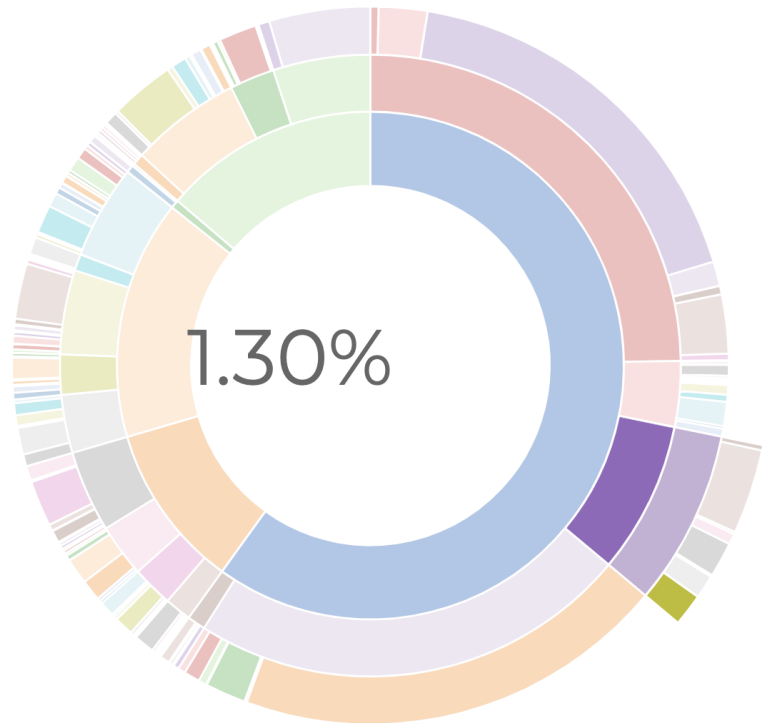
1. [0-5 points] Describe in detail the meaning of the visual encoding elements in the following infographic reporting data concerning the timeline of high speed internet.



2. [0-5 points] Discuss in detail the importance of the work of William Playfair in the history of Data Visualization.
3. [0-5 points] Discuss similarities and differences between the PCA and the classical MDS algorithms, providing at least two graphical examples different from those shown in the course slides.
4. [0-7 points] The datafile [worldwide_crop_consumption.csv](#) collects information about consumption of different crops across several years at country level. Prepare an infographic with two panels describing some patterns inferred by (a subset of) the dataset, including a choropleth map.
5. [0-7 points] Consider the datafile [Setlur.csv](#) collecting the gene expressions (features) of 472 prostate cancer patients (samples). Each column is a sample, while features are on rows. First row is sample ID, first column is gene name, last column is cancer type: 1 indicates Prostate_Cancer_with_Fusion, 0 indicates Prostate_Cancer_without_Fusion, -1 indicates Unknown_Prostate_Cancer. Furthermore, patients with sample ID between GSM208029 and GSM208392 are Swedish, while the remaining patients (from GSM208404 to GSM208512) come from the US. Project the samples on a bidimensional plane using PCA, MDS, UMAP and tSNE and color the samples according to the combination of cancer type and provenience (6 colors in total). What algorithm better separates the six classes?

6. [0-7 points] Using the dataset [world_data.csv](#) try to replicate the following plot.

Asia > South > Eastern Asia > Vietnam 1.30%



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.