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

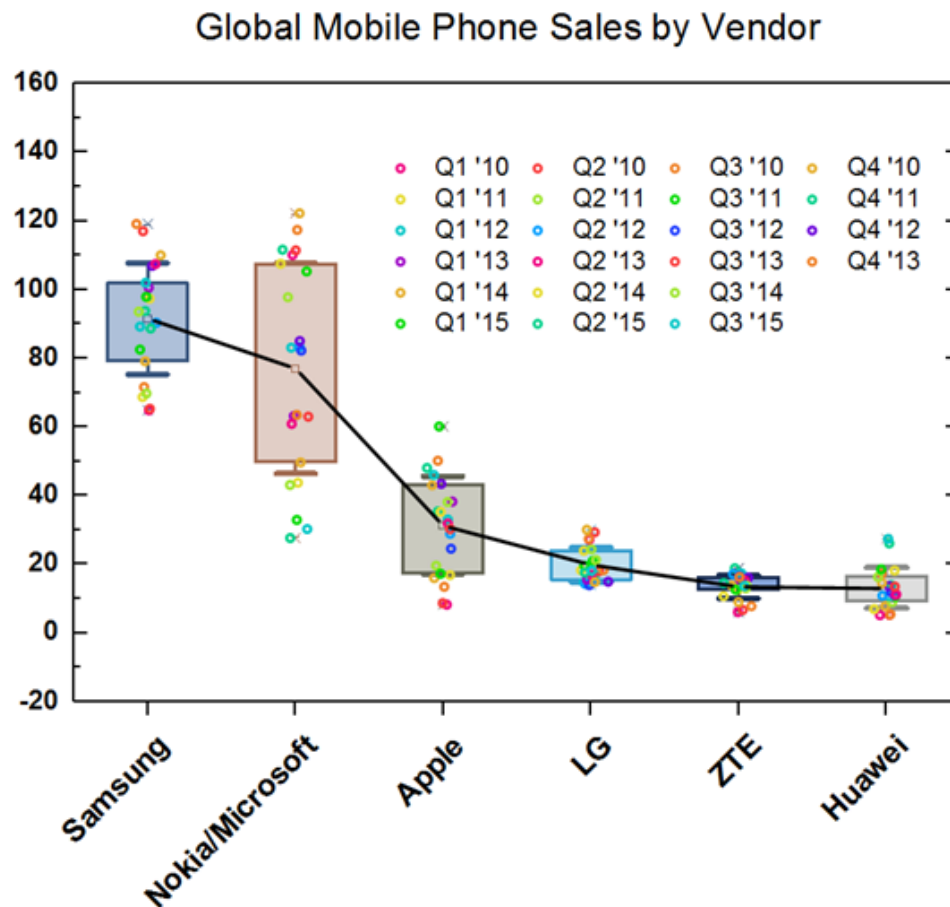**Data Visualization Lab Exam**
14 June 2019

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 elements of visual encoding (marks and annotations), their meaning and their use. Explain and comment the meaning of the visual encoding elements in the following picture

2. [0-5 points] Describe similarities and differences between a barplot and a histogram, and find two examples where only one of the two charts is the correct visualization option.

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

4. [0-7 points] Find a data visualization (one or more plots) for the data in http://tiny.cc/ju777y (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] Starting from the same dataset in http://tiny.cc/ju777y , select the data subset S corresponding to the countries Australia, Canada, Greece, Hungary, Portugal and United States. Build a two-dimensional UMAP projection and a two-dimensional PCA projection of S coloring the points according to the country and comment the resulting visualization.

6. [0-7 points] Using the datafile http://tiny.cc/lwz77y , replicate the following plot

Fold Change and P-value in Nanostring/SEQC data

(Hint: the color of the bars depends on the sign of the FoldChange; the number of the stars is *floor(-ln(Pvalue)/8)*)

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.