



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Visualization

2021-993-qOmicS-05

Aims

The aims of this module are to be able to manage and programmatically use the grammar of graphics in the R environment, make a few plots both for general purpose and specific applications. Moreover general concepts of data communication and reproducible science will be also delivered.

Contents

- Introductory material
- General purpose data visualizations
- Specialty data visualizations
- Challenges

- Reporting tools

Detailed program

Introductory material

- Introduction to data visualization
- Purpose of visualizations in the context of data communication. Role of the "story"
- The main rules of data visualization
- Refresher on R and RStudio environments
- Concept of tidy data and the Tidyverse

Data visualization

- Colors, color blindness and color palettes
- The grammar of graphics and Ggplot2, easthetics and geometries

- General purpose plots for one single variable, categorical or quantitative
- General purpose plots for more than one variable
- Plotting statistics and uncertanties

- Miscellaneous plots

Specialty data visualizations

- Plotting fold changes in gene expression: volcano plots
- Plotting gene ontology enrichments 1: bar and dot/bubble plots
- Plotting gene ontology enrichments 2: treemaps, semantic space and external web resources
- Network visualizations with Cytoscape and related Apps

- Plotting gene ontology enrichments 3: network GO with Cytoscape

- General heatmaps and gene expression heatmaps

- The R package pheatmap

Challenges:

- Coding the measles vaccination heatmap
- Circular visualizations with circlize

Reporting in R

- Primer on markdown
- Coding and commenting in a notebook

- Building and using knitted reports
- Other tools for reproducible research and code sharing (Jupyter, Google colabs)

Prerequisites

A general knowledge of the R programming language, as well as usage of R and RStudio are assumed. Refreshers of such topics are offered in the first lessons but an independent ability to handle R, base packages and the dplyr grammar (tidyverse) is warmly suggested.

Teaching form

The module is delivered mainly by a series of videos, in which both theoretical and practical (coding) contents are shown. Along with the video, students will have access to commented R scripts and exercises. A few hours of

practicals and Q&A sessions will be delivered remotely in-person by the instructor.

Textbook and teaching resource

Semester

Assessment method

Office hours
