



TOWARDS BIG DATA

Informatics tools for data manipulation

Course I

Dr. Mario Barbato

Faculty of Agriculture, Food and Environmental Sciences, Università Cattolica del Sacro Cuore

Objectives

The recent technological innovations in data-production and data-collection boosted the generation of new data at an unprecedented speed. The Big Data revolution poses new exciting opportunities and challenges to researchers, which need specific informatics knowledge to make the most out of such ever increasing datasets.

The course Towards Big Data - Informatics tools for data manipulation part 1 and 2 aims to provide the theoretical and practical background to manipulate and extract information from large datasets by means of informatics tools. Lectures are designed to allow the participants to test first-hand every new concept the lecturer presents.

The highly interdisciplinary and strongly applicative nature of the topics presented here - along with the use of freeware platforms and software - will enable the participants to include the newly acquired knowledge in their daily research routine from the first lesson.

Course outline

1. UNIX Command Line Tools [5 hours]
 - a. Basic commands and concepts.
 - b. Data exploration and extraction.
 - c. Data query and editing: regular expressions.
2. Using the data: R [3 hours]
 - a. Basics of R programming.
 - b. Data structure, exploration, and visualisation.