


School on Multivariate data analysis methods for biological data using the R package mixOmics, 2nd edition

Are you struggling with understanding and analyzing large and complex biological data sets? The Toulouse GenoToul Biostat platform and the Unité de Glycobiologie Structurale et Fonctionnelle UMR 8576 CNRS/Université Lille 1 are organizing an Autumn School within the EU COST Action "The quest for tolerant varieties: phenotyping at plant and cellular level (FA1306) (https://www.plant-phenotyping.org/home_costfa1306). The course will provide both theory and hands-on training focusing on multivariate data analysis and dimension data reduction for biological studies.



The objective of this three days Autumn School is to introduce the fundamental concepts of multivariate dimension reduction methodologies. Those methods are particularly useful for data exploration and integration of large data sets, and especially in the context of systems biology, or in research areas where statistical data integration is required. Each methodology that will be presented during the course will be applied on biological "omics" studies including transcriptomics, metabolomics and proteomics data sets using the R package mixOmics (<http://mixomics.org/>).

Organized by GenoToul Biostat platform (<http://perso.math.univ-toulouse.fr/biostat/>)
Unité de Glycobiologie Structurale et Fonctionnelle UMR 8576 CNRS/Université Lille 1
(UGSF, <http://ugsf-umr-glycobiologie.univ-lille1.fr/?lang=en>)

Dates 12 September until 14 September 2017

Practical information

The course fee is 300 € for the academia and 600 € for the industry. It covers tuition, course material, coffee/tea, lunches and a dinner downtown. Please note that no fee will be counted for 12 PhD students and/or early postdocs (up to 8 years after Ph.D) selected by COST Action FA1306, and their travel and living expenses will be take in charge up to 650 €.

Location: Lille, France

Maximum number of participants: 20

Deadline for application: 15 June 2017

Decision for attendee selection and grant allowance: 30 June 2017

Registration: 4 to 15 July 2017

Apply for this Summer School at:

<https://www.surveymonkey.com/r/GDRYJWI>

In addition, send a CV to E. Goulas (Estelle.goulas@univ-lille1.fr) and do not forget to mention if you want to apply for a grant.

Contact in Lille

E Goulas, UGSF

Estelle.goulas@univ-lille1.fr

Prerequisite and requirements

The audience is expected to have a good working knowledge in R (e.g. handling data frames and perform simple calculations). Attendees are requested to bring their own laptops, having installed the software RStudio <http://www.rstudio.com/> and the R package mixOmics (instructions provided prior to the training).

More details on the covered topics

1. *Key methodologies in mixOmics and their variants*
 - A. Exploration of one data set and how to estimate missing values
 - B. Identification of biomarkers to discriminate different treatment groups
 - C. Integration of two data sets and identification of biomarkers
 - D. Repeated measurements design
 - E. Introduction to the integration of more than two data sets
2. *Review on the graphical outputs implemented in mixOmics*
 - A. Sample plot representation
 - B. Variable plot representation for data integration
 - C. Other useful graphical outputs
3. *Case studies and applications*

The following statistical concepts will be introduced: covariance and correlation, multiple linear regression, classification and prediction, cross-validation, selection of diagnostic or prognostic markers, cross-validation, l_1 and l_2 penalties in a regression framework. Each methodology will be illustrated on a case study (we will alternate theory and application).

Note that mixOmics is not limited to biological data only and can be applied to other type of data where integration is required.

Target group

The course is intended for data analysts in the fields of bioinformatics, computational biology and applied statistics with a good statistical knowledge and **a good working knowledge in R**. It will be particularly useful to those interested in:

1. *Exploring large data sets.*
2. *Selecting features with methods implementing LASSO-based penalizations.*
3. *Using graphical techniques to better visualize data.*
4. *Understanding and/or applying multivariate projection methodologies to large data sets.*

Results

After completion of this workshop, participants will be able to

1. *Understand fundamental principles of multivariate projection-based dimension reduction technique.*
2. *Perform statistical integration and feature selection using recently developed multivariate methodologies.*
3. *Apply those methods to high throughput biological studies, including their own studies.*

Course leader



Sébastien Déjean was awarded his PhD in Applied Statistics in 2002 at Université de Toulouse, France after spending 4 years in a Biometry lab at INRA (French National Institute for Agronomic Research). He then has been working at the Toulouse Mathematics Institute as a research engineer. He works in close collaboration with researchers working in different areas, such as high-throughput biology, chemistry and information retrieval. Sébastien is an expert in statistical data analysis and he contributes to the development of several R packages including mixOmics as a core member.

More details on Sébastien's current research projects: <http://perso.math.univ-toulouse.fr/dejean/>

Additional teachers

Dr David Rengel (data analyst, LIPM)

One or two other additional teachers will be present depending on the number of participants

Local organizing committee

Dr Estelle Goulas (UGSF)

Dr Anne-Sophie Blervacq (UGSF)

Dr Anne Creach (UGSF)

Dr Brigitte Huss (UGSF)

Pr Simon Hawkins (UGSF)

"Vieille bourse" in Lille

