Programming for Data Science

Marco Beccuti

Università degli Studi di Torino

Dipartimento di Informatica

September 2021

Teacher

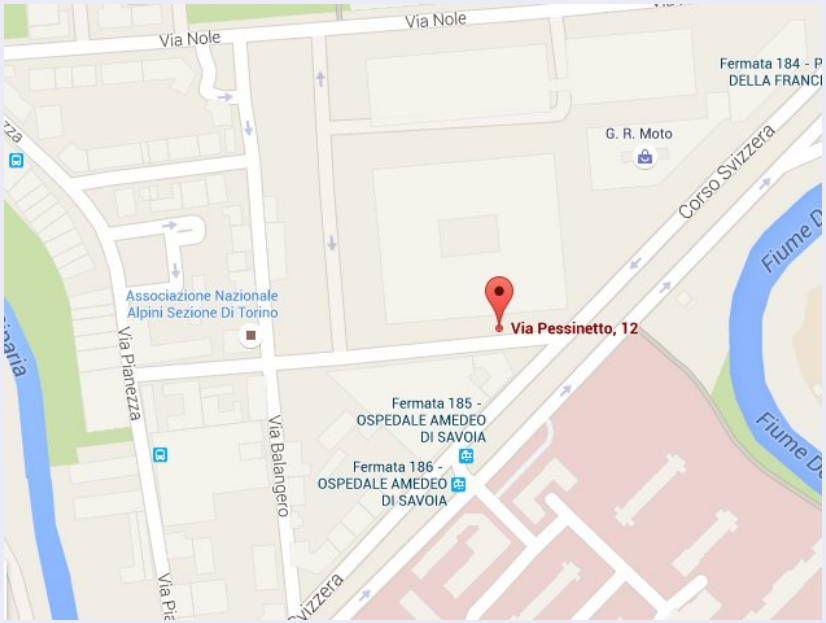
Marco Beccuti
Associate Professor in Computer Science

Office:
Dipartimento di Informatica
Corso Svizzera 185
10149 Torino

E-mail:
beccuti@di.unito.it

Home page:
http://www.di.unito.it/∼beccuti

How to reach Dipartimento di Informatica



My research is currently mainly focused on computational modeling and simulation of complex systems. In particular, I am interested in:

- Stochastic and hybrid modeling languages;
- Exact and approximated techniques to analysis the behavior of complex systems;
- Applications to computational Systems Biology.

Moreover, I work on the design of bioinformatics algorithms and workflows for the analysis of deep sequencing data (i.e. genomic, transcriptomic and single cell data) with particular emphasis on reproducibility aspect.

Course delivery

- The course consists of 24 hours of lectures and laboratories (using webex software for streaming);
- Laboratory includes exclusively practical activities;
- The slides presented during lectures are available to students as on-line materials.
- Attendance to lessons is not mandatory, but highly recommended due to the necessity of learning and employing a specific computer science instruments.

Course outline

- Introduction to Data science;
- Visualization using ggplot2;
- Data structures in R: vector, matrix, list and data frame, tibble;
- Apply in R;
- Data Transformation;
- Input and output in R;
- Function in R;
- Model in R;
- Debugging in Rstudio;
- Create a package in R.

Course examination

- Exam will consist in a oral examination;
- Practice exercise using R could be required during the examination.

Suggested readings for course

- Garrett Grolemund and Hadley Wickham, **R for Data Science**, O'Reilly Media, Inc, USA, 2017.
- The R Manuals: An Introduction to R (http://cran.r-project.org/doc/manuals/r-releas /Rintro.pdf)
- The teaching material used for lessons and a series of practical exercises are available on the web site of the course (Moodle repository)