

Literate Programming and Statistics (CMP595 PPGC/INF/UFRGS)

Lucas Mello Schnorr, Jean-Marc Vincent

INF/UFRGS

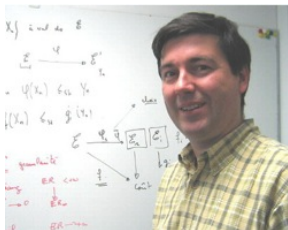
Porto Alegre, Brazil – October 2018



Teaching Staff

Jean-Marc Vincent

jean-marc.vincent@imag.fr

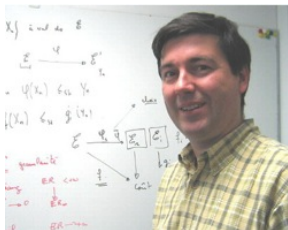


*Laboratoire d'Informatique de
Grenoble (France), Université de
Grenoble-Alpes, Inria POLARIS
Research Team, Performance
evaluation of large scale
systems/networks/platforms*

Teaching Staff

Jean-Marc Vincent

jean-marc.vincent@imag.fr



Lucas Mello Schnorr

schnorr@inf.ufrgs.br



Laboratoire d'Informatique de Grenoble (France), Université de Grenoble-Alpes, Inria POLARIS Research Team, Performance evaluation of large scale systems/networks/platforms

LPS Coordinator, INF/UFRGS (Porto Alegre, Brazil), Invited Inria POLARIS Research Team, Performance Analysis, Parallel Computing (HPC), Design of Controlled Experiments

Logistics

In case you need to contact us, write an e-mail with the subject

Subject: [LPS 2018] ...

Logistics

In case you need to contact us, write an e-mail with the subject

Subject: [LPS 2018] ...

To whom should you send your questions?

Lucas

- ▶ General LPS organization (and logistics)
- ▶ Technical installation (Rstudio + packages), R language

Jean-Marc

- ▶ Concepts about statistics

Logistics

In case you need to contact us, write an e-mail with the subject

Subject: [LPS 2018] ...

To whom should you send your questions?

Lucas

- ▶ General LPS organization (and logistics)
- ▶ Technical installation (Rstudio + packages), R language

Jean-Marc

- ▶ Concepts about statistics

But both of us are fully capable to answer any question regarding LPS

Two-fold Goal

Conceptual

To present the fundamental philosophy behind **literate programming** to conduct a faithful and reproducible data analysis, combining sound statistical procedures with modern data analytics tools.

Technical

The course is based on Rstudio as IDE and using the R programming language for data analysis. Every lecture will be backed up with practical sessions and worked out examples.

In Practice

Know how to conduct a study of a computer science object, from measurements that have been collected about it.

Scientific Method

Question + Hypothesis + Confirm/Refute

In Practice

Know how to conduct a study of a computer science object, from measurements that have been collected about it.

Scientific Method

Question + Hypothesis + Confirm/Refute

Steps

1. Know how to design a simple experiment and collect data
2. Know how to analyze the collected data (data analytics)
3. Know how to report (data visualization, comments, analysis)

Tools

- ▶ Distributed version control (git)
- ▶ Analysis, manipulation, visualization (Rstudio, dplyr, tidyr, ggplot2)
- ▶ Report (R markdown within Rstudio)

Lecture Organization

This is the scheduler for the 2018/2 edition.

Day	Date	Hour	Room
0	29/10 (Monday)	9:00 – 12:00 (3h)	Lab 43413-102
1	30/10 (Tuesday)	9:00 – 10:15 (1h15)	Lab 43413-102
2	31/10 (Wednesday)	9:00 – 12:00 (3h)	Lab 43413-101
3	01/11 (Thursday)	9:00 – 12:00 (3h)	Lab 43413-101
4	08/11 (Thursday)	9:00 – 12:00 (4h)	Lab 43413-101

Day 1 (today)

- ▶ General introduction and literate programming
- ▶ Practical use of Rstudio (basics)

LPS Evaluation

- ▶ Mid-term activity (for next week): topic is given
- ▶ Mini-project (by the end of November): choose-yourself-topic

Topic overview

Environment

- ▶ Introduction, problem
- ▶ Literate Programming, Rstudio, Rmd,

Data manipulation

- ▶ Data carpentry (or Data Wrangling)
- ▶ The `dplyr` and `tidyr` packages

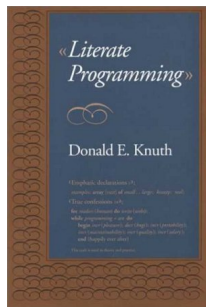
Data visualization

- ▶ The `ggplot2` package (grammar of graphics)
- ▶ Guidelines, Checklist for good graphics

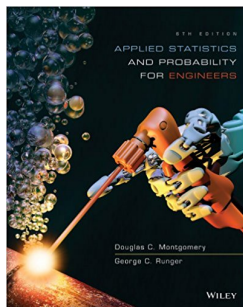
Probability and Statistics

- ▶ Introduction
- ▶ Probabilistic Modeling

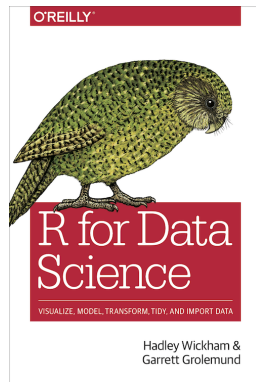
References



Literate Programming.
Donald E. Knuth
(Stanford, California)
(CSLI Lecture Notes,
no. 27.). ISBN
0-937073-80-6.



Applied Statistics and
Probability for
Engineers 6th Edition.
Douglas C.
Montgomery (Author),
George C. Runger.
Wiley.



R for Data Science.
Garrett Grolemund,
Hadley Wickham.
[http:
//r4ds.had.co.nz/](http://r4ds.had.co.nz/)