

Adventures in Supercomputing with R

Virtual lectures at Charles University Department of
Probability and Mathematical Statistics

George Ostrouchov

Oak Ridge National Laboratory and University of Tennessee

2021/12/31 (updated: 2022-02-23)

The Virtual Lectures

- Originally planned in person via *Fulbright Distinguished Chair* grant
 - Declined due to personal travel restriction
- Current Virtual Lectures are not affiliated with Fulbright in any way
- Lecture exercises conducted at the Czech supercomputing center IT4I

The Lecturer

- Czechia, Praha: through grade 9
- Canada: high school and University of Waterloo
- USA: Iowa State University
- Joined Oak Ridge National Laboratory, USA, in 1983
 - now Distinguished Member of R&D Staff at ORNL
 - and Joint Faculty Professor of Statistics at the University of Tennessee
- Visited IT4I in 2015 and 2016 (also Charles U in 2016)

IT4Innovations

Czech National Supercomputing Center



Technical University of Ostrava

Lecture Notes and Exercises

- Lecture notes on Github "RBigData/KPMS-IT4I" repository
 - <https://github.com/RBigData/KPMS-IT4I>
- Exercises on Github "RBigData/KPMS-IT4I-EX" repository
 - <https://github.com/RBigData/KPMS-IT4I-EX>
- Each lecture posted the day before in `lectures` directory of KPMS-IT4I and replaced with annotated version after
- Weekly exercises in `exercise` directory of KPMS-IT4I-EX

Objectives

Understanding of parallel hardware and software, and their access from R

Ability to manage multicore and multinode computing from R

Ability to use these technologies for statistical computing and large data processing on large parallel systems

Gaining knowledge when multiple nodes are needed and when faster code or only a few cores are enough

References:

1. R: **Advanced R** by Hadley Wickham. *Course assumes you already know R, but this is a good reference, especially for explaining deeper nuances of the R language.*
2. HPC: Introduction to HPC with MPI for Data Science by Nielsen (2016)
3. HPC: **Introduction to High Performance Scientific Computing** by Eijkhout (2022)
4. HPC & R: Mastering Parallel Programming with R by Chapelle Et Al. (2016)
5. HPC & R: Parallel Computing for Data Science, With Examples in R, C++ and CUDA by Matloff (2016)

Slides created via the R package **xaringan** and converted to pdf with **xaringanBuilder**.
Presentation and annotation via reMarkable.