ICT Solutions for Brilliant Minds

CSC





Experiences from *High-performance R* course

Heli Juottonen & Billy Braithwaite CSC – IT Center for Science, Finland



Motivation for a new HPC-flavoured R course

- my main role: maintaining supercomputer R environment and user support
- teaching a course on basics of R for data analysis
- no trainings in supercomputer R use



Common situation to solve

X in R is taking too long. The computer is running out of memory. What can I do?

Aims of the new course

- taking into account properties of R as programming language
- understanding resource use, finding bottlenecks
- parallel and distributed computing
- using R in an HPC environment



Target audience

- R users running R on their own computer or in the supercomputer web interface RStudio
- experienced users of another programming language and/or HPC
- pre-requisites: basics of R

 $\,\circ\,$ some Linux and HPC experience useful



First course in September 2025

- one teacher with data analysis background, one with computer science background
- 2 full days
- 19 participants from universities and research institutes in Finland
- various fields and skill levels



csc

R experience

Linux & HPC experience





Course content

- short lectures, exercises inbetween
- Day 1 (web interface Rstudio):
 - introduction to computers, programming languages
 - o measuring running time and profiling R jobs
 - o how R handles objects in memory
 - o basics of parallel computing

File Edit Code View Plats Session Build Debug Profile Tools Help				heljuot 🕒 🕴
🕙 O + 🕸 🐨 + 🖬 🕼 👘 📣 Construction 📰 🔛 + Addres +				I Project: (None
ansole Terminal - Background Jobs -	0	Environment History Connections Tutorial		=0
R + R 4.4.2 + -7 -1		🔐 🔛 🖙 Import Dataset + 👘 155 MB + 🖌		🗄 List + 😒
version 4.2.2 (2005-16-10) TILE of Lement's partner: MS-2 Amountain for Statistical Computing actions: MS-2 Amountain for Statistical Computing (Toma Software and example)		 Configuration Contraction 	pry	Q.
<pre>iteriter() on new to cite a or a pocuears in policities, ge 'dee()' for some demos, 'help()' for on-line help, or elp.stort()' for an HTML browser interface to help, ge 'd()' to goil R.</pre>		Files Plots Packages Help Viewer Presentation	ne 🔕 More -	-
		C & Home		100.000
		a Name	5.24	Modified
				May 16, 2025, 0.54 PM
		P Aprofile bachjobrest hyses hor	133.8	May 16, 2025, 6:54 PM
		C C outer	2.4 XB	May 6, 2025, 6:24 PM
		pakettissenna, oma R piszettissenna, oma R piszetmodulis sistelitourse siste	284.8	Apr 3, 2025, 10:36 AM

Course content

Day 2 (web interface Rstudio, batch jobs on a cluster)

- $\,\circ\,$ getting started with parallel R (foreach)
- orunning R batch jobs on cluster
- \circ parallel batch jobs on a cluster (future)
 - \circ multiple cores
 - \circ multiple nodes
- \circ handling large datasets
- $\,\circ\,$ R and C++ integration (Rcpp)



CSC



- Welcome	
	- Tieteen tietotekniikan keskus - IT Center for Science
Puht Puhti.	i.csc.fi serving Finnish science since September 2nd 2019 ! csc.fi – Atos BullSequana X400 – 682 CPU nodes – 80 GPU nodes
Servicedesk	: 09-457 2821, servicedeskgcsc.fi Switchboard : 09-457 2001
 User Guide https://doc 	s.csc.fi
- Manage my	account
- Software -	SG(1)/
Available m	odules can be listed with command: module avail and module spider
Documenta Servicede News	tion: https://docs.csc.fi/ sk support: servicedeskgcsc.fi.
2023-10-04:	GPU monitoring has been improved and seff can now show job energy usage in Mn. Keep in mind that the data might not be complete until a few minutes after the job has ended.
2023-10-09:	Home directories are private to individual users. Home directories with incorrect permissions were secured on October 9, 2023. For file-sharing within your project group, please utilize the /projappl and /scratch folders.
2025-04-14:	SSH login to Puhti and Mahti will only work with SSH keys added in MyCSC. Password authentication and SSH keys stored in your personal -v.ssHvathorized keys file will no longer work. Instructions for setting up SSH keys in MyCSC:
	https://docs.csc.fi/computing/connecting/ssh-keys/
2025-05-15:	Puhti will have a service break on Tuesday Jun 3 8:00 - 21:00. For more info, please see the email to the "puhti-users" list on 2025-05-13 or the announcement at:

Outcome: what worked

- the course was created and held $\ensuremath{\textcircled{}}$
- basic structure of the new course ok
- positive feedback from participants
- different teacher backgrounds mentioned in feedback as a benefit – success but not easy



Outcome: what didn't work

- bringing own data for exercises

 has to be very well defined
 recommended only for advanced R users
- more alternative exercises for different skill levels

 more support for people new to HPC
 extra challenges for advanced participants





Next steps

- next course in November 2025
- same basic structure but lots of refining to do

 creating meaningful exercises that are not field-specific
- more material and exercises aimed at different skill levels
- more on handling very large data sets



Conclusions

- clear demand for an HPC-flavoured R course
- rewarding but challenging to co-teach with very different backgrounds
- one of the main challenges: diverse backgrounds of participants



CSC