



# UL HPC School 2014

## PS6: Using R on the UL HPC Platform

---

**J. Emeras**

University of Luxembourg, Luxembourg

## Latest versions available on Github:

UL HPC tutorials:

<https://github.com/ULHPC/tutorials>

UL HPC School:

<http://hpc.uni.lu/hpc-school/>

PS6tutorial sources:

<https://github.com/ULHPC/tutorials/tree/devel/advanced/R>



# Summary

---

- 1 Pre-requisites
- 2 Objectives
- 3 Practical Session



# Summary

- 1 Pre-requisites**
- 2 Objectives
- 3 Practical Session



# Install and Run R

## 1 On your local machine:

- Find a release that fits your distribution at CRAN Archive
- Install and launch R-Studio

<http://cran.r-project.org/>

<https://www.rstudio.com/>

## 2 On the cluster

First connect to the cluster, then submit a job to run R.

```
(localhost)$> ssh chaos-cluster  
(frontend)$> oarsub -I -l core=1,walltime="00:30:00"  
(node)$> module load R/3.0.2-ictce-5.3.0  
(node)$> R
```

## 3 Install and Load a Package

```
(R-shell)$> install.packages("ggplot2")  
(R-shell)$> library(ggplot2)
```



# Summary

---

- 1 Pre-requisites
- 2 Objectives**
- 3 Practical Session



# Objectives of this Practical Session

- Being able to plot data
  - histogram for data distribution
  - plot in different colors from different data sources
- Know some tips to organize your data
  - aggregate a dataset by column and apply an aggregation function
  - data.table package for binary search in datasets
  - performance in R operations
- R in parallel
  - on one machine
  - on a cluster with socket communications



# Summary

- 1 Pre-requisites
- 2 Objectives
- 3 Practical Session**



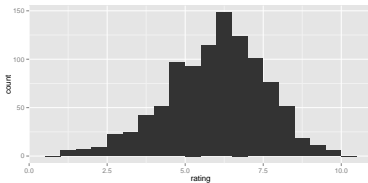


## Exercises

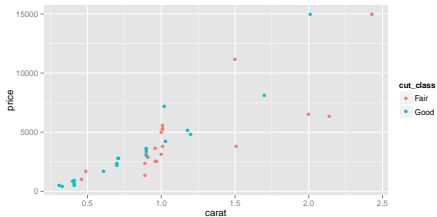
- Start the tutorial <https://github.com/ULHPC/tutorials/tree/devel/advanced/R>
  - Plot 2 graphs in section Simple Plotting
  - Answer 2 questions at the end of section Organizing your Data
  - Compare performance of aggregation operations w/wo parallelization
- Plot a speedup graph
  - with different number of cores and/or machines
  - needs: ggplot, parallel R

# Simple Plotting

Movies Histogram:



Diamonds Plot with 2 colours:





## PS Questions

Question: use `ddply` instead of `tapply` in the first example

```
ddply(DT, .(x), summarize, sum(v))
```

Question: return the min and max instead of the sum.

```
min_max = function(data){  
  c(min(data), max(data))  
}  
DT[,min_max(v),by=x]  
  
## or  
DT[,c(min(v), max(v)),by=x]
```



Thank you for your attention...

## Questions?



- 1 Pre-requisites
- 2 Objectives
- 3 Practical Session



Thank you for your attention...

## Usefull links

- CRAN Archive <http://cran.r-project.org/>
- ggplot2 Documentation <http://docs.ggplot2.org/current/>
- CRAN HPC Packages <http://cran.r-project.org/web/views/HighPerformanceComputing.html>
- Advanced R programming by Hadley Wickham <http://adv-r.had.co.nz/>