

# How to prepare for a job in data science?

#### **Małgorzata Schmidt**

Advanced Computing & Data Science Lab



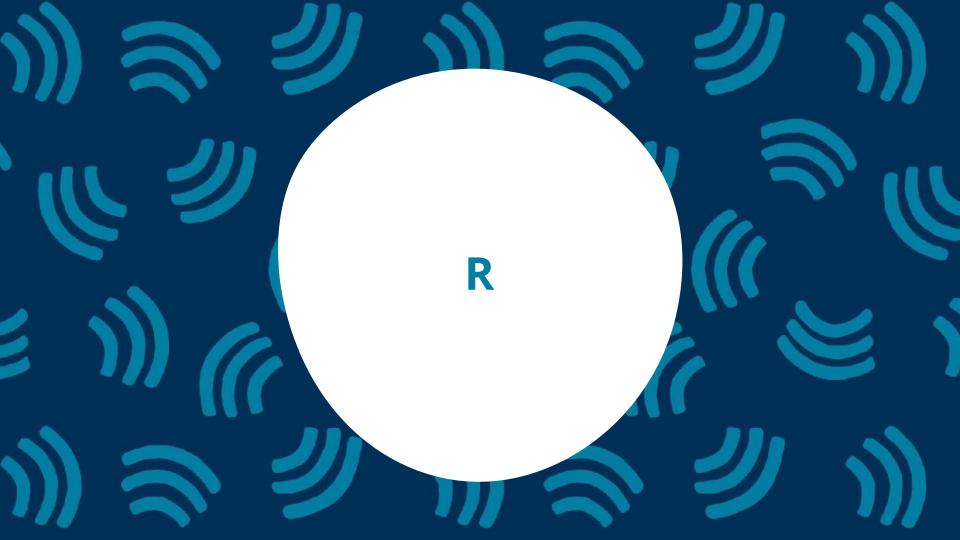
#### Studies vs internship







ı







- Data manipulation: dplyr
- Data visualisation: ggplot2
- Machine learning: glmnet, nlrwr, MASS, klaR, class, rpart, tree, randomForest, caret

- + tidyverse
- + DT, plotly, shiny
- + tensorflow,

- Database/API connection:
  ETLUtils, httr, jsonlite
- Cloud computing









- Data ready for analysis...
- ... quite small...
- ... and created to test models.

- Data stored in different formats which requires preprocessing...
- ... multidimensional (often with multiple meaningless variables)...
- with a lot of missing/test values









- University's own Git-based system
- Github introduced
- Four core commands: git add, git commit, git pull, git push

+ Git workflow

Main use: saving files

Main use: collaboration and keeping track of various versions of files



How to best prepare for a job in data science?





### Only at a university

- Focus on theory:
  - models assumptions
  - math behind various models



1

## **P** Only at work

- Hard skills:
  - latest packages, methods, solutions
  - variety of tools used
- Soft skills:
  - communication with clients
  - managing time
  - self-organisation
  - o ...



Visit our team <u>blog</u> for many interesting data science related posts!





