



Pearson

# How to prepare for a job in data science?

**Małgorzata Schmidt**

Advanced Computing & Data Science Lab

WhyR, 3 July 2018



Image by Lucy Vigrass

# Studies vs internship



The image features the R logo, which consists of a large white circle centered on a dark blue background. Inside the white circle is a blue capital letter 'R'. The background is decorated with a repeating pattern of smaller, lighter blue 'R' logos.

R



## University

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



## Work

- + tidyverse
- + DT, plotly, shiny
- + tensorflow,
- + **Database/API connection:** ETLUtils, httr, jsonlite
- + **Cloud computing**



**Data**



## University

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



## Work

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



Image by Lauren Rolwing



**Version control**



## University

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

**Main use: saving files**



## Work

+ Git workflow

**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



# Only at work

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

Visit our team [blog](#) for many interesting data science related posts!



ALWAYS LEARNING