GENERAL ASSEMBLY

DATA SCIENCE

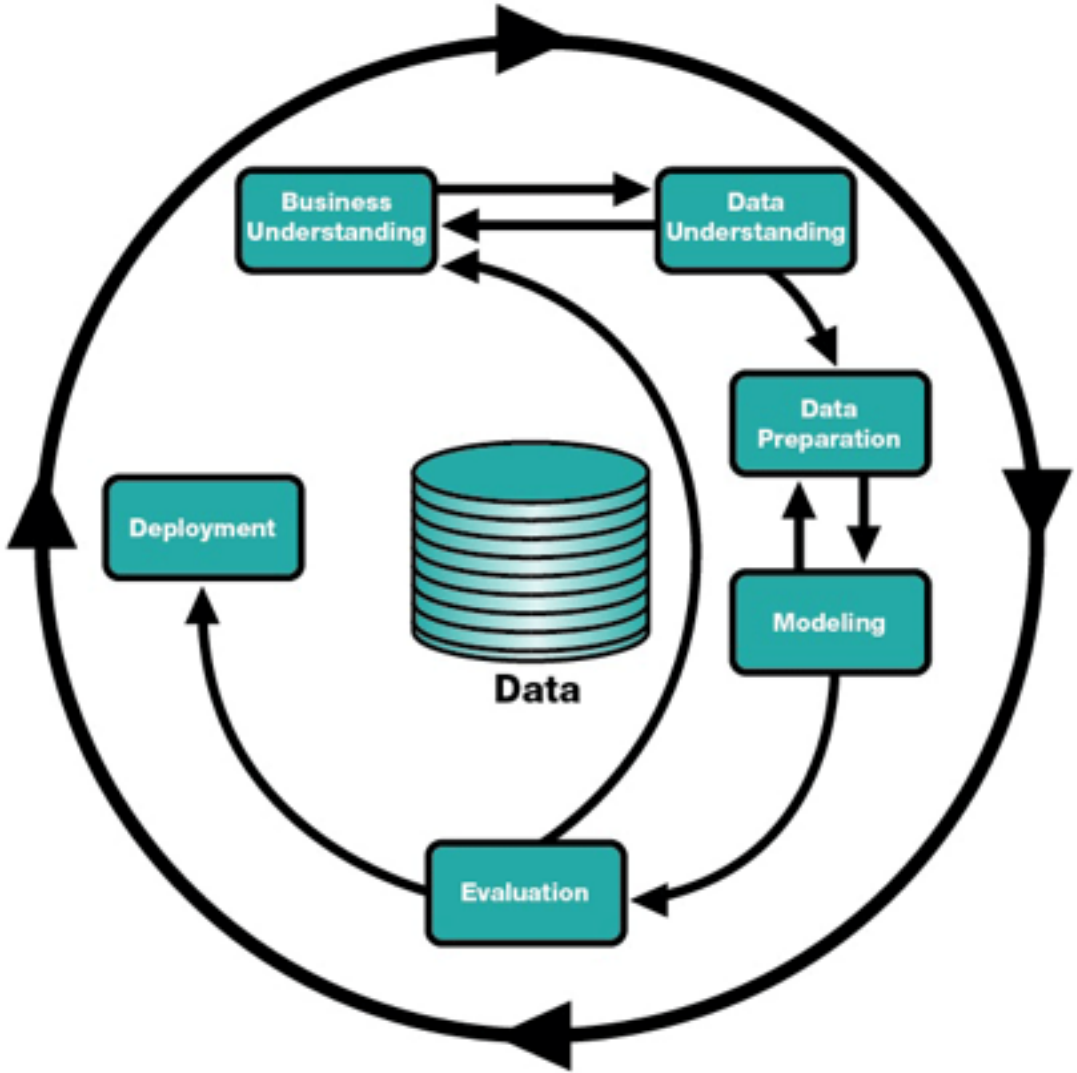
11 WEEK PART TIME COURSE

Week 1 – Data Science & Git
Wednesday 22nd March

AGENDA 2

1. What are the steps in a Data Science project?
2. What is Python and why is it Popular with Data Scientists?
3. Bit more Git
4. Git Lab
5. Discussion
6. Course Projects
7. Homework

WHAT ARE THE STEPS IN A DATA SCIENCE PROJECT?


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

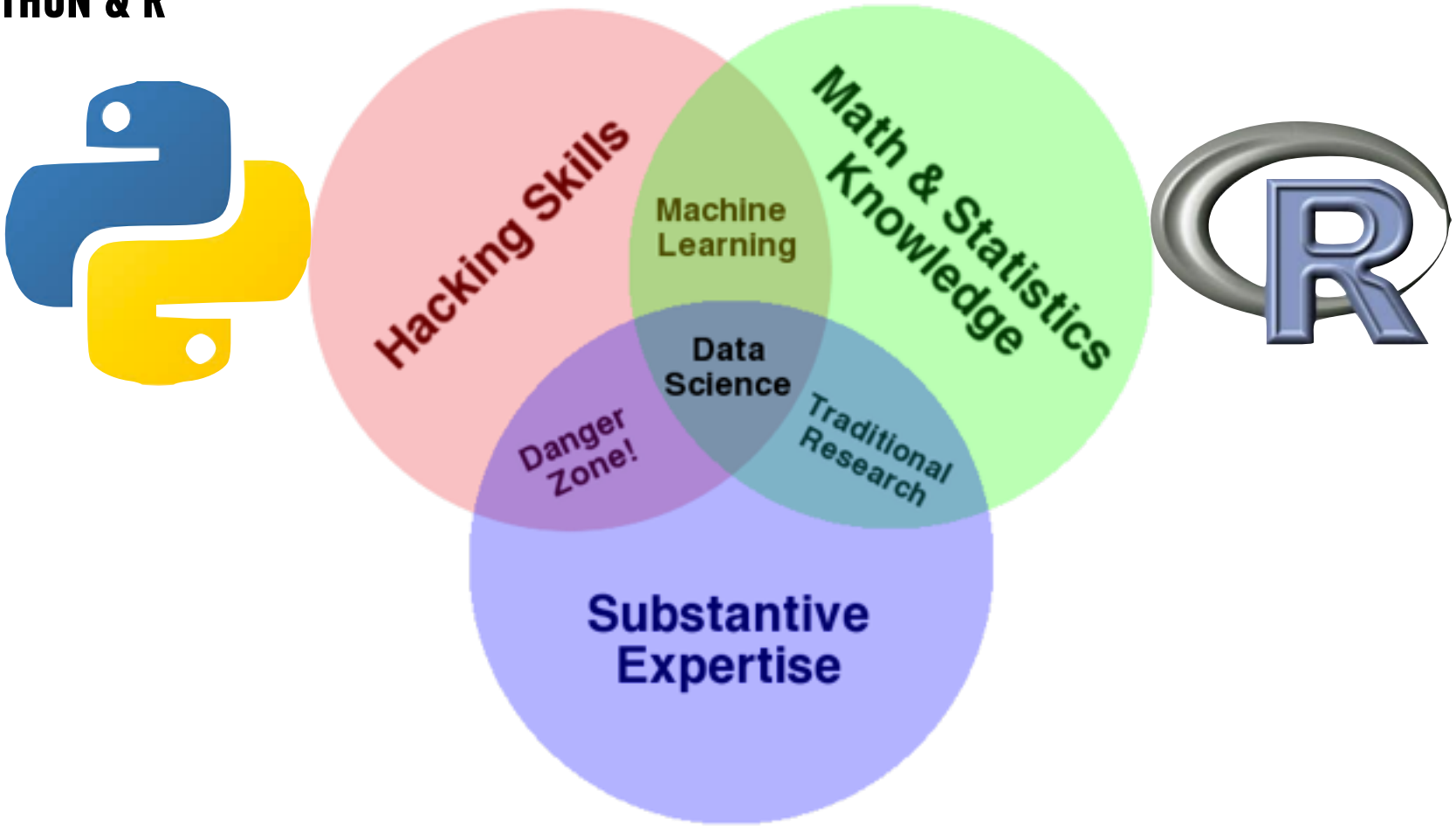
What is Data Science?

"Data analysis projects today rely on databases, computer and network hardware, and computer and network software. A collection of models and methods for data analysis will be used only if the collection is implemented in a computing environment that makes the models and methods sufficiently efficient to use"

- Created by Guido Van Rossem in 1991 and emphasizes productivity and code readability
- Version 3 (but 2.7 is still very popular)
- “Python is an interpreted, object-oriented, high-level programming language with dynamic semantics”

- Batteries Included: Large collection of built in libraries e.g. SciKit, Pandas, Theano, etc
- Simple and clean syntax
- General purpose language: lots of people outside of data science will be able to work with it

PYTHON & R


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise
R

DATA SCIENCE PART TIME COURSE

USING DATA SCIENCE PACKAGES

- Packages are libraries of code written to solve a particular set of problems
- In Python there are many related to data science including Pandas, SciKit Learn, Numpy
- These are installed and managed with PIP (Pip Installs Packages)

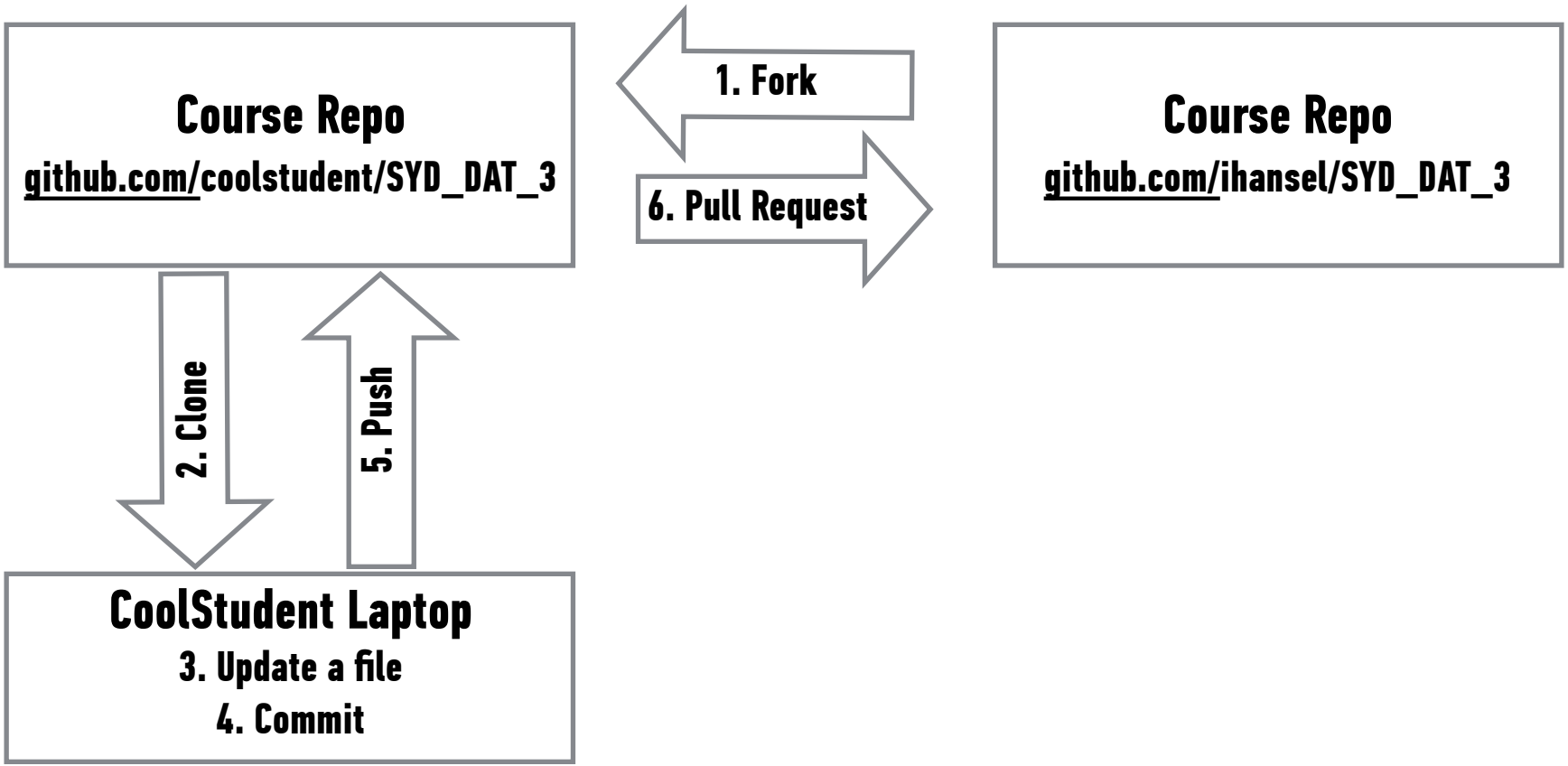
```
pip install some-package-name
```

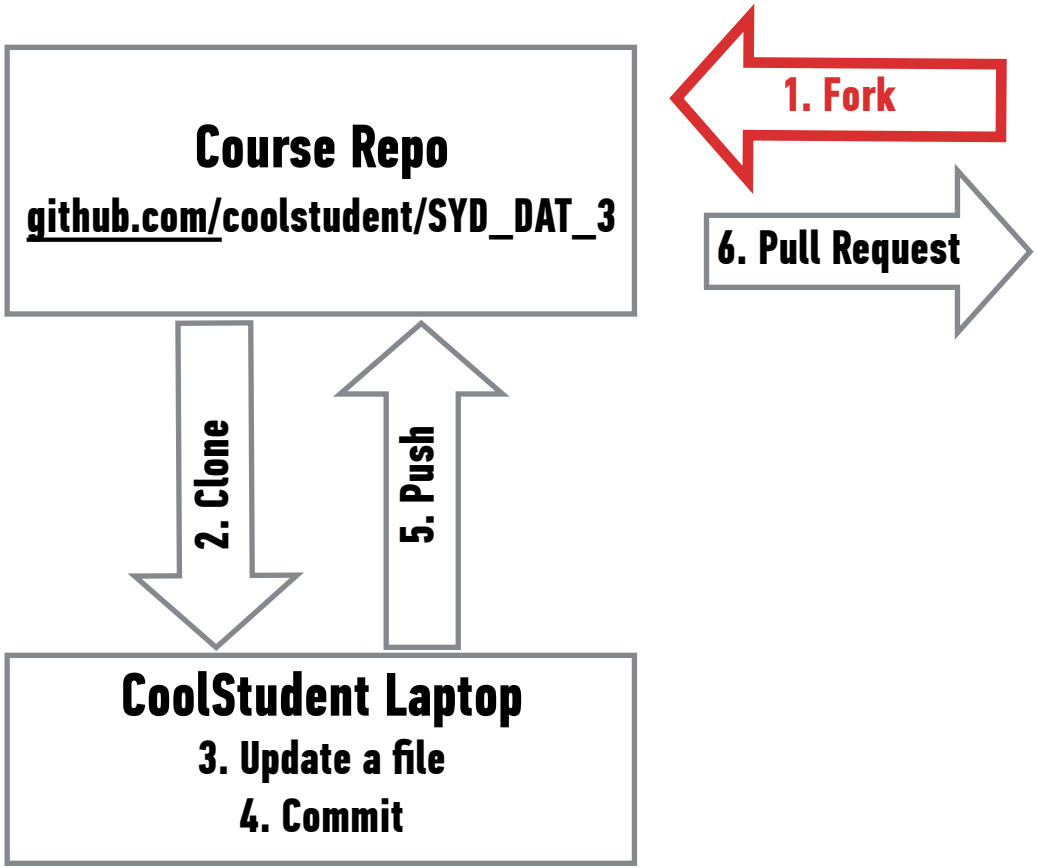
- pandas: manipulate data
- SciPy / NumPy: scientific computing and numerical calculations
- scikit-learn: use machine learning methods
- matplotlib: visualise data
- statsmodels: perform statistical tests
- Beautiful Soup: read in XML and HTML data
- iPython: interactive programming

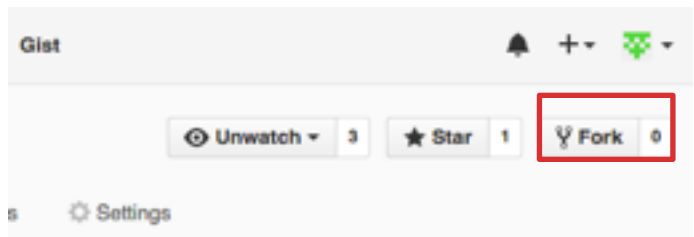
DATA SCIENCE PART TIME COURSE

A BIT MORE GIT!

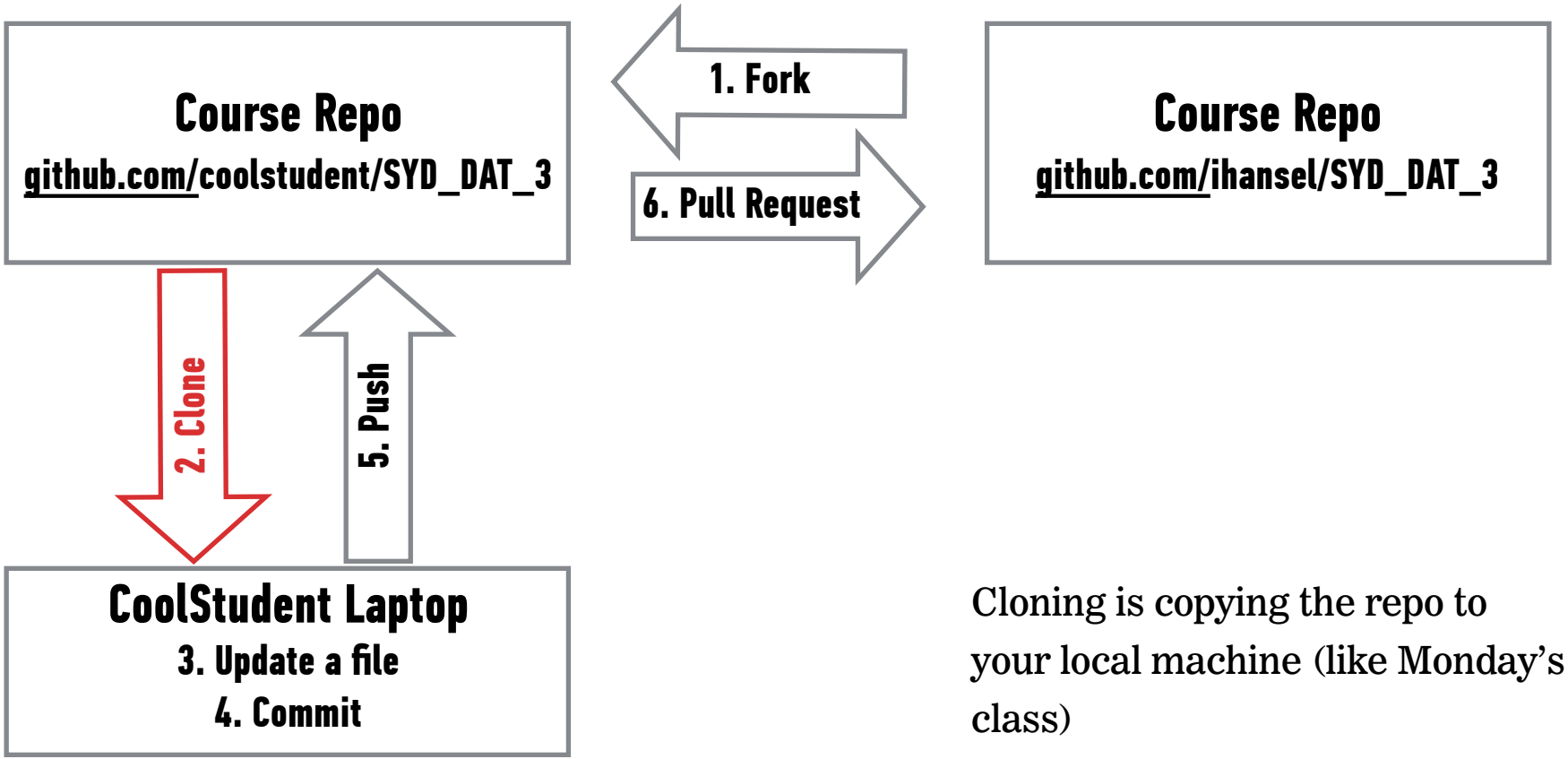
- Forking, making a copy of someone else's repository so you can work on it


Course Repo
github.com/coolstudent/SYD_DAT_3
1. Fork
6. Pull Request
Course Repo
github.com/ihansel/SYD_DAT_3
2. Clone
5. Push
CoolStudent Laptop
3. Update a file
4. Commit

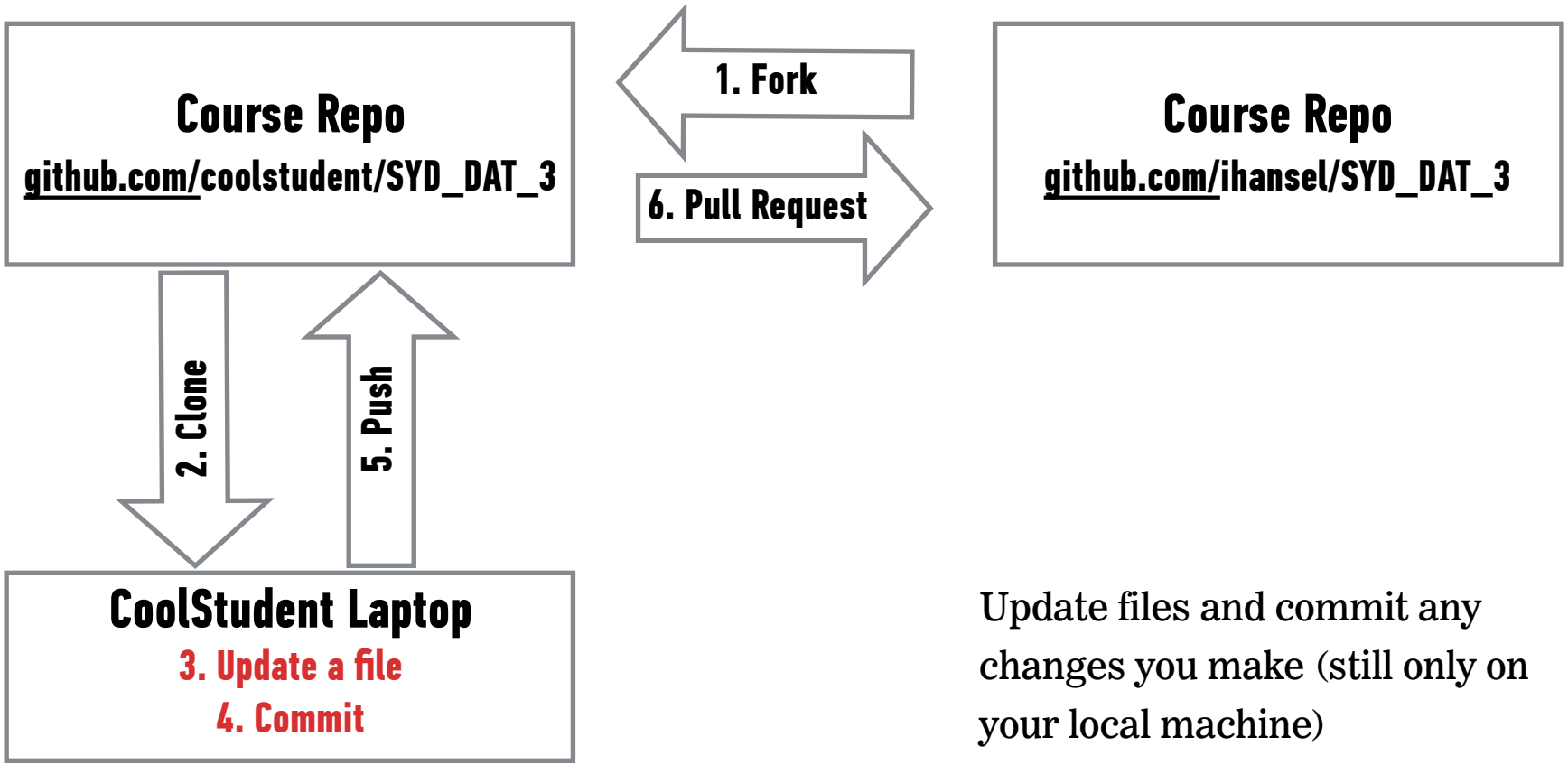




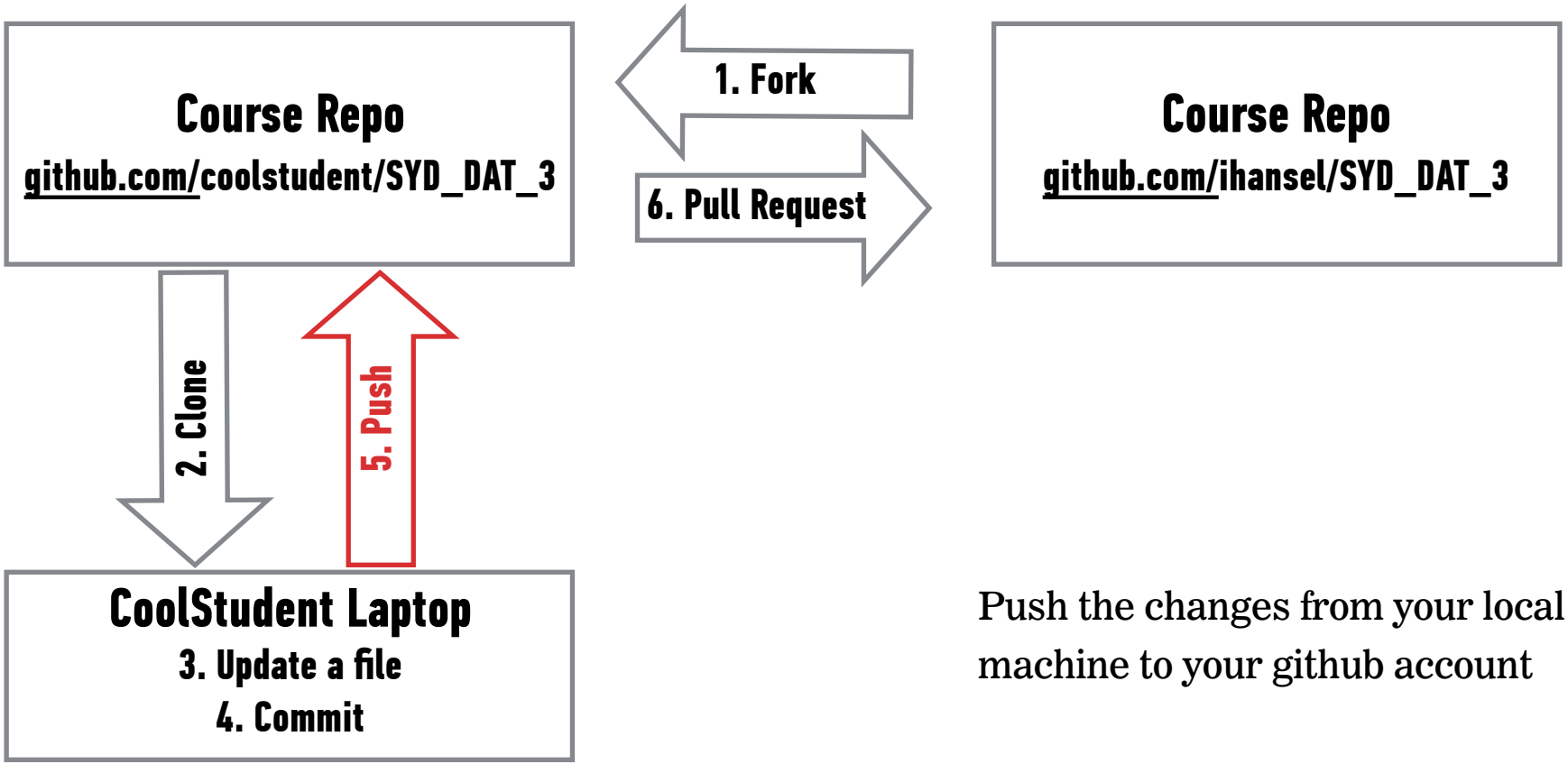
Forking, making a copy of someone else's repository so you can work on it



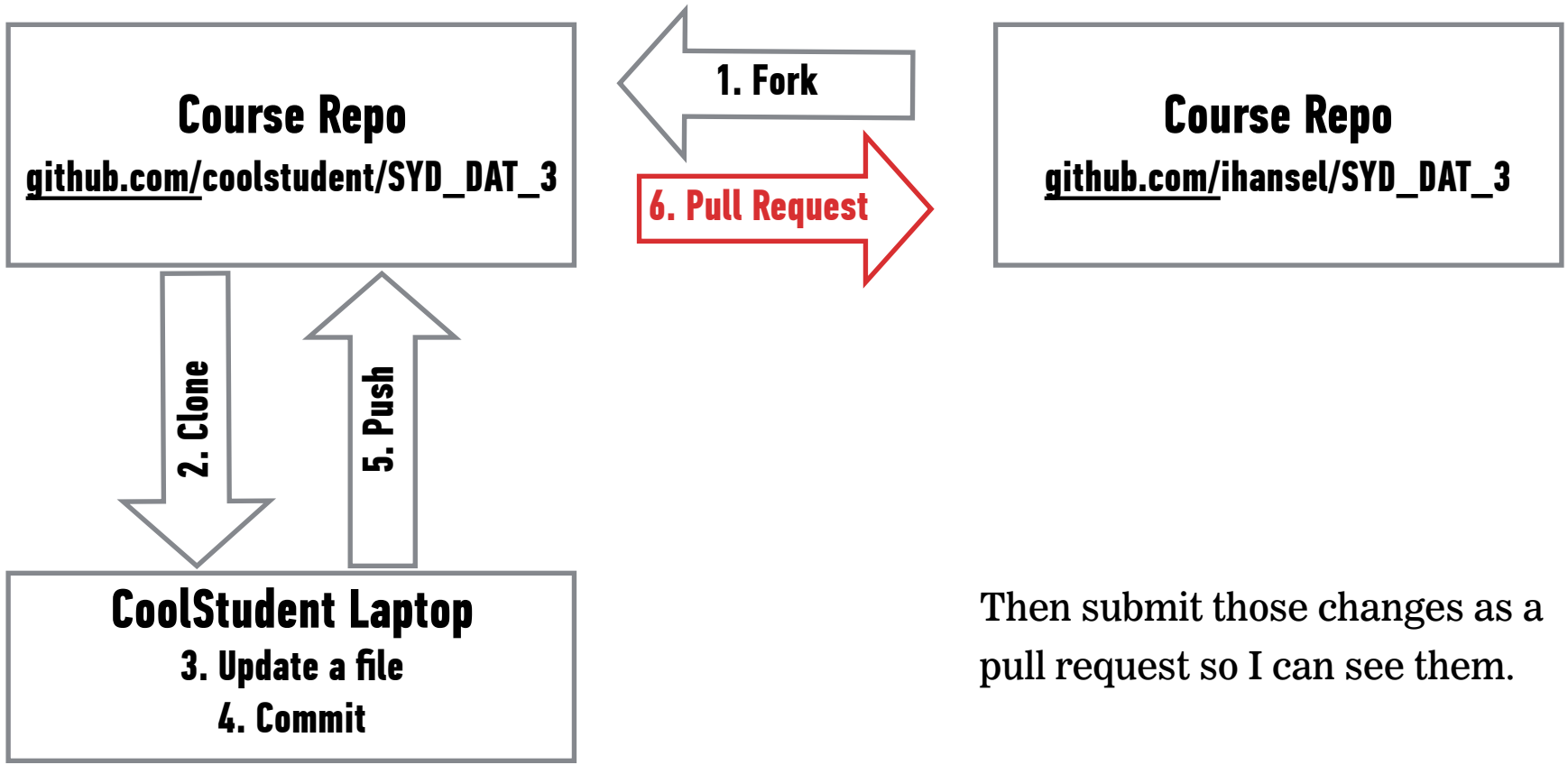
Cloning is copying the repo to your local machine (like Monday's class)



Update files and commit any changes you make (still only on your local machine)



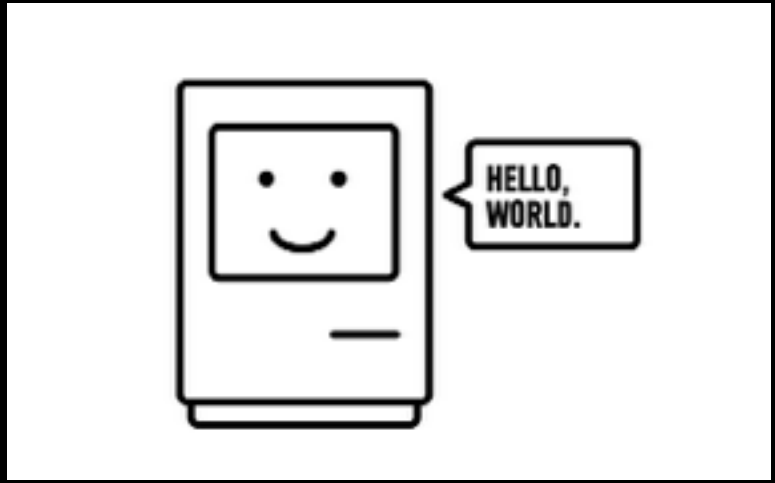
Push the changes from your local machine to your github account



Then submit those changes as a pull request so I can see them.

DATA SCIENCE PART TIME COURSE

GIT LAB



DISCUSSION TIME

DISCUSSION TIME

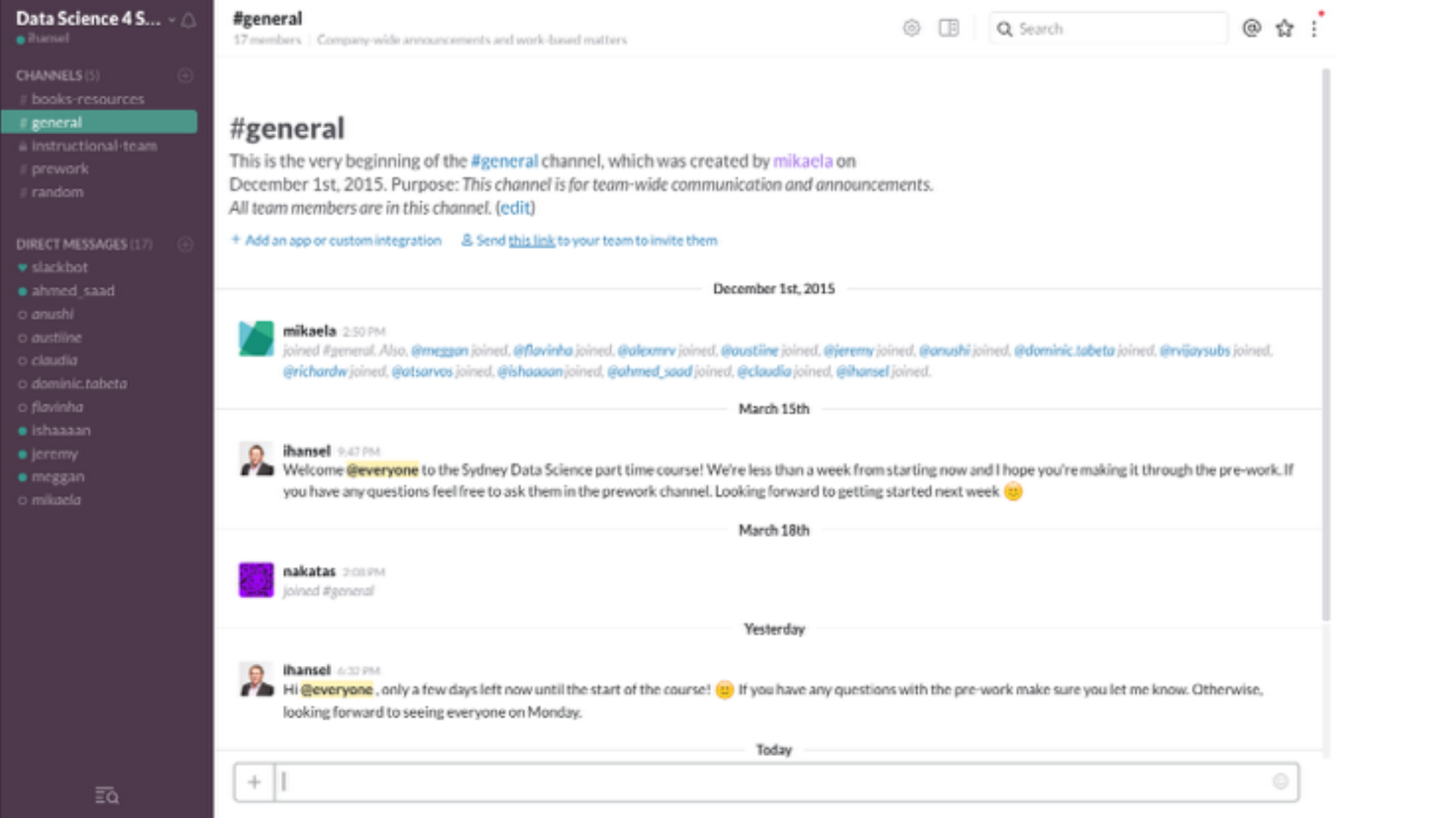
Prework

- Readings
  - Metacademy Learning Plan
  - Data Science Handbook
  - An Introduction to Statistical Learning

DISCUSSION TIME

Homework

- Due next Friday
- I will review within 7 days
- Counts to letter of completion
- (Will be easy)


Data Science 4 S...
ihansel
CHANNELS (5)
books-resources
general
instructional-team
prework
random
DIRECT MESSAGES (17)
slackbot
ahmed_saad
anushi
austiine
claudia
dominic.tabeta
flavinha
ishaaaan
jeremy
meggan
mikaela
#general
17 members | Company-wide announcements and work-based matters
Search
#general
This is the very beginning of the #general channel, which was created by mikaela on December 1st, 2015. Purpose: This channel is for team-wide communication and announcements. All team members are in this channel. (edit)
+ Add an app or custom integration
Send this link to your team to invite them
December 1st, 2015
mikaela 2:50 PM
joined #general. Also, @meggan joined, @flavinha joined, @alexmrv joined, @austiine joined, @jeremy joined, @anushi joined, @dominic.tabeta joined, @rvijaysubs joined, @richardw joined, @atsarvos joined, @ishaaaan joined, @ahmed_saad joined, @claudia joined, @ihansel joined.
March 15th
ihansel 9:47 PM
Welcome @everyone to the Sydney Data Science part time course! We're less than a week from starting now and I hope you're making it through the pre-work. If you have any questions feel free to ask them in the prework channel. Looking forward to getting started next week
March 18th
nakatas 2:08 PM
joined #general
Yesterday
ihansel 6:30 PM
Hi @everyone, only a few days left now until the start of the course! If you have any questions with the pre-work make sure you let me know. Otherwise, looking forward to seeing everyone on Monday.
Today