MY472 – Lab 1: Git and GitHub

Pablo Barberá & Akitaka Matsuo

MY 472: Data for Data Scientists

October 4–5, 2018

Course website: lse-my472.github.io

Software we will use today

- R 3.5.1 – Install from https://www.r-project.org/
- RStudio – Install from https://www.rstudio.com/products/rstudio/download-server/
- GitHub Desktop – Install from https://desktop.github.com/

Introduction to Git/GitHub

Git is a type of version control.

- System that keeps records of your changes: helps track who made changes when
- Distributed (entire code and history on each machine) – Allows for collaborative development
- Possibility of reverting any changes and go back to previous state
- Git: created by Linus Torvald in 2005 to facilitate Linux kernel development
- Other options: Mercurial, Subversion
- GitHub allows you to host repositories and adds extra functionalities (UI, documentation, issues, user profiles...)

Basic concepts of git

- Code lives in a repository: collection of all files (and history)
- Every time you make changes, you need to make a commit:
  - Creates a snapshot of your code.
  - Informs how files have changes
  - You need to add a message explaining changes
- After you commit, you need to push the changes to the repository on GitHub so that others can see them
- Note – you also need to pull first to receive changes from other people
- When you start from a repository someone created, you will have to first fork it (create a copy on GitHub) and then clone it (download) to your computer

Our first test

1. Go to MY472 lectures repository: `https://github.com/lse-my472/lectures`
2. Fork it to your own GitHub profile
3. Click on Clone and use GitHub desktop to create the repository in your computer
4. Find the files and open `01-RMarkdown.rmd` using RStudio
5. Compile (knit) the RMarkdown file into an html file
6. Go back to GitHub Desktop and commit your changes
7. Then, push the changes back to GitHub
8. Go to GitHub and check that the new file is in the repo.

Our first assignment with GitHub Classroom

1. Access the GitHub classroom for assignment 1 through: `https://classroom.github.com/a/CSEdQujE` and accept the assignment
2. This will create your own version of the repository
3. Clone it using GitHub desktop
4. Write your personal information in `information.md`, commit the file and push the changes.
5. Fix the RMarkdown file `RMarkdown-practice.Rmd` so that it can be compiled. Commit and push this change.

Assessment

- 5 problem sets will be assessed (50%).
  - Submitted via GitHub (more in lab)
  - Only compiled assignments will be accepted
  - Two will be collaborative; three will be individual submissions.
- Take-home exam (50%)
  - Work with a dataset to answer a series of question
  - More open-ended format than problem sets
  - Deadline: January 18, 23:59

Assessment criteria

- 70–100: Very Good to Excellent (Distinction).
  - Perceptive, focused use of a good depth of material with a critical edge. Original ideas or structure of argument.
- 60–69: Good (Merit)
  - Perceptive understanding of the issues plus a coherent well-read and stylish treatment though lacking originality
- 50–59: Satisfactory (Pass)
  - A "correct" answer based largely on lecture material. Little detail or originality but presented in adequate framework. Small factual errors allowed.
- 30–49: Unsatisfactory (Fail)
- 0–29: Unsatisfactory (Bad fail)
  - Based entirely on lecture material but unstructured and with increasing error component. Concepts are disordered or flawed. Poor presentation. Errors of concept and scope or poor in knowledge, structure and expression.