

STA 2300 and MATH398: Introduction to Data Science

Lecture

This material was abstracted from the detailed and excellent tutorial by Jenny Bryan (R Studio Black Belt).

<https://happygitwithr.com/rstudio-git-github.html>

See Chapter 6 in this document

<https://happygitwithr.com/install-git.html> to install git on your laptop.

Once installed in Rstudio configure your local github version.

```
library(usethis)
use_git_config(user.name = "Jane Doe", user.email = "jane@example.org")
```

Here the email must be the same as the one you used to create your github account.

Basic steps for setting up an R project with github as a repository for sharing and backup.

Create the new repository

Log onto github and create a new repository. Choose to add a README.md file. When you are in this repository the URL link is just the web address in your browser.

E.g. My example looks like: <https://github.com/dnychka/MowaterExample>

This is publicly accessible so you should be able to use my link and see the repository, download its contents, etc.

Here are the steps in detail:

- Go to <https://github.com> and make sure you are logged in.
- Click the green “New repository” button. Or, if you are on your own profile page, click on “Repositories”, then click the green “New” button.

How to fill this in:

- Repository name: Choose wisely!
- Description: At least a sentence what this is for.
- Public. YES Initialize this repository with a README. For everything else, just accept the default.
- Click the big green button “Create repository.”
- Copy the HTTPS clone URL to your clipboard via the green “Clone or Download” button.

You are now the owner of a new github repository!

You can share the URL, add files to this, recover previous files that were overwritten, and delete contents. It will also still be there if your laptop crashes.



Adding a new project with github backup in RStudio

The way we will use github is to keep a secure copy of your work in R. Having created the new repository now work on the R project side using tools in R studio.

In RStudio, start a new Project:

- File > New Project > Version Control > Git. In “Repository URL”, paste the URL of your new GitHub repository. It will be something like this <https://github.com/dnyska/MowaterExample>
- Accept the default project directory name, e.g. myrepo, which coincides with the GitHub repo name.
- Think carefully where you want the project folder to sit.
- Check “Open in new session”, as that’s what you’ll usually do in real life.
- Click “Create Project”.

The sandbox and the vault

The use of git for keeping track of files can be very complex, adapting to large software development efforts by many people. Here the key is to manage different *versions* of the software and that is why git is referred to as a version control system. For this course and for much of your data science work you will just have to use a few core git operations. These are **commit**, **push**, and **pull**.

Moreover, using git is simplified by the handy Git GUI that will appear as a tab in Rstudio when you open up a git based project.