

Activity: Using Git and GitHub

Goal: Use Git and GitHub to put course materials on your Rivanna account.

In this activity, you will go through steps of using Git and GitHub covered in the [the reading on GitHub](#). You can also draw on the what you learned in the [Technical Orientation](#). At this point, you also should be able check off the following items:

- Understand the difference between Git and GitHub.
- Understand the purpose of Git and Github for data science work.
- Ensure Git is installed on your computer.
- Understand how to find a repository on GitHub.

Let's apply and extend this knowledge now with our course repo.

Be sure you are inside the course directory on Rivanna we created earlier.

We assume you have already created a GitHub account.

Also, before you get started, follow [these instructions](#) to set an SSH key. You can create this on both your computer and Rivanna, but for the assignment you need only create it on Rivanna.

Fork the course GitHub hosted repository (“repo”) to your GitHub account.

Go to <https://github.com/ontoligent/DS5100-2024-06-R> in your web browser.

i Note

This is the course repo — all of the course notebooks and other code will be available here. Each week, you will access your course materials here.

Click on the Fork icon in the upper right and follow the prompts to finish the process.

You should end up at the web page of your newly forked repo.

You will now have a copy of the repo in your GitHub account.

Clone the forked repo for this course inside of your course directory on Rivanna.

Find the green Code button and click on it. You should see something like this:

Make sure you have selected the SSH option.

! Important

Note: This requires that you have SSH set up.

Then click on the copy icon and paste the value into the following command:

```
git clone https://github.com:<github_user_name>/DS5100-2024-06-R.git
```

! Important

Be sure to clone the repo from *your* GitHub account, replacing `<github_user_name>` with your GitHub user name. Do not just cut-and-paste the line above!

You now have a copy the course repo to your account on Rivanna.

This will be the directory you created in your pre-class activities under Documents/.

Create a new file in your newly cloned repo.

Go to your command line window on Rivanna.

Use `cd` to move into the directory just created by the clone operation.

Move into the directory `activities/hello`

! Important

Make sure you are in this directory before proceeding.

If you get lost – for example if you moved around the file system before this step – you can `cd` to the absolute path:

```
cd ~/Documents/MSDS/DS5100/DS5100-2024-06-R/activities/hello
```

Note that the tilde sign `~` stands for the path to your home directory.

Using the file editor on Rivanna, create and save new file called `<userid>_hello.txt`, replacing `<userid>` with your actual user ID, e.g. `rca2t_hello.txt`.

In the file, introduce yourself by answering the question: What is the most recent film you watched and enjoyed?

Save the file.

Add and commit the changes you made.

Now do the following:

```
git add <userid>_hello.txt
git commit -m "Created file for class"
```

Push your new file to the repo on GitHub.

Since you have SSH set up, you can issue the following command without having to enter a password:

```
git push
```

Create a *Pull Request*

Finally, make a pull request to have your file added to the original site. To do this, follow these steps:

Click on the “Pull requests” menu item (see image below) on the web page for your repo.

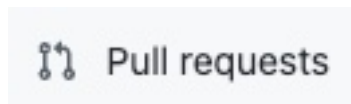


Figure 1: Image of pull request button on GitHub

Click on the green “New pull request” button.

Click on the green “Create pull request” button.

Give the request the title “In-class activity” and then press the green “Create pull request” button at the bottom of the form.

Now the ball is in the instructor’s court to merge the request with the original. If you put your file in the right place and named it properly, it will be merged.

Going Forward

During the semester, you will not be making pull requests to submit your work. We do it here to demonstrate the concept since it is so basic to working with GitHub in the real world.

Instead of making pull requests, you will be using a separate repository for your work. So, you will be working with two repositories going forward:

1. **The Course Repo**, which is where you will find course materials, such as homework keys and exercises after the classes in which they are used. This should be [updated each day](#).
2. **Your Assessments Repo**, which is where you will be your finished work as assigned.
3. **The Book Repo**, which is where you will find the textbook for this course along with all of the notebooks in it. You may choose to clone or fork this, or you may download individual notebooks from the site as needed.