**Lab 1: Introduction/Replication and Reproducibility**

**Today’s Lab:** Today we will be familiarizing ourselves with Github: what it is, how to set up your account, and how to use it. We will learn how to connect our Github account to RStudio, how to create a new repository (repo) on Github, how to create a new project in RStudio, and how to cloning an existing repo from Github.

**I. Introduction to Git & Github**

First, what is Git? Git is a version control system. Jenny Bryan in “Happy Git with R” says to think of the Git repository as the "Track Changes" feature from Microsoft Word on steroids. Next, let's talk about Github. Github is a hosting service. It provides a home for your Git-based projects on the internet. We want to make sure that our RStudio can talk to local Git (and therefore Github). If you haven’t already, please go to <https://github.com/> and register an account.

For setting up new or existing projects, the general workflow is as follows:

1. Dedicate a directory (AKA “folder”) to it on your local drive.

2. Make it an RStudio Project.

3. Make it a Git repository.

4. Knit your R Markdown file and make commits on the "Git" tab periodically.

5. Push commits to GitHub.

**II. Connecting Github to RStudio**

First check if you already have Git installed in RStudio by typing “which git” and “git --version” into your RStudio console. If you receive a message saying something like “git: command not found,” then we will have to install Git.

**Installing Git (from Jenny Bryan’s “Happy Git with R”):**

Windows

Install Git for Windows, also known as msysgit or “Git Bash”, to get Git in addition to some other useful tools, such as the Bash shell. Yes, all those names are totally confusing, but you might encounter them elsewhere and I want you to be well-informed.

We like this because Git for Windows leaves the Git executable in a conventional location, which will help you and other programs, e.g. RStudio, find it and use it. This also supports a transition to more expert use, because the “Git Bash” shell will be useful as you venture outside of R/RStudio.

NOTE: When asked about “Adjusting your PATH environment”, make sure to select “Git from the command line and also from 3rd-party software”. Otherwise, we believe it is good to accept the defaults.

Note that RStudio for Windows prefers for Git to be installed below C:/Program Files and this appears to be the default. This implies, for example, that the Git executable on my Windows system is found at C:/Program Files/Git/bin/git.exe. Unless you have specific reasons to otherwise, follow this convention.

Mac

1. Go to the shell and enter one of these commands to elicit an offer to install developer command line tools:

git --version

git config

1. Accept the offer! Click on “Install”.

**Introduce yourself to Git**

Type the following into the console:

git config --global user.name 'Jane Doe'

git config --global user.email 'jane@example.com'

git config --global --list

**III. Creating a New Repository**

There are two ways you can create a new repo on Github:

1. From your Github homepage:

A screenshot of a cell phone

Description automatically generated

1. By navigating from the right sidebar on the homepage:

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

After you click the green “New” button, Github will take you to the “Create a new respository” page where you will need to create a repository name (something short and straightforward) and you can add a description to explain the purpose of the repo. For this course, if you are submitting your labs and assignments via your Github link, you should select the “Public” option when creating your repo so that we can see your projects. Make sure you check the “Initialize this repository with a README” box at the bottom. Then click the green “Create repository” button.

A screenshot of a cell phone

Description automatically generated

Once the repo is created, you will be brought to a page that looks like this:

A screenshot of a cell phone

Description automatically generated

**IV. Creating a New Project in RStudio**

1. First, we want to create a folder on our local drive titled “Lab 1” (ideally this is a sub-folder within in your folder for this course).
2. Next, we want to clone the repo we created in part III to our local drive. We can do so by clicking the green “Clone or download” button and copying the HTTPS link:

A screenshot of a cell phone

Description automatically generated

1. Open RStudio on your computers. On the menu bar located at the top of the screen, click “File” 🡪 “New Project” 🡪 “Version Control”:

A screenshot of a cell phone

Description automatically generated

1. Select “Git” on the next page:

A screenshot of a cell phone

Description automatically generated

We want to paste the repo URL that we copied previously from our Github page into the “Repository URL” box. We can leave the “Project directory name” as the default option. In the “Create project as a subdirectory of” box, click “Browse” and make sure that you have selected the “Lab 1” folder that you created in step 1 above. Then, click “Create Project”.

1. Once the new project has been created, go back to the menu bar located at the top of the screen and click “File” 🡪 “New File” 🡪 “R Markdown”:

A screenshot of a cell phone

Description automatically generated

1. When creating the R Markdown file, use this naming convention and the Word output format. When we “knit” our R Markdown file later, RStudio will output a Word document with our code.

A screenshot of a cell phone

Description automatically generated

We have already created a lab R Markdown document for you for today’s lab titled “Lab 1\_student.”

**V. Cloning a Repo**

The Johns Hopkins University Center for Systems Science and Engineering (CSSE) created a visual dashboard of COVID-19 global cases using ArcGIS and has made all the data open-access. We will focus on navigating Github and learn how to pull data from Github then import the data into RStudio to use.

1. First, go to the Github repository where the data is hosted: <https://github.com/CSSEGISandData/COVID-19>.
2. We can explore the different folders in the repo: the data is contained in the folder titled “csse\_covid\_19\_data.” Within the “csse\_covid\_19\_data” folder, there are two folders of 1) daily reports and 2) time series data. We want to read the “README.md” file to understand what files are contained in the dataset. It is important to explore the dataset before starting the data analysis part in RStudio so we know what data we are dealing with and have an analysis plan in mind.

A screenshot of a social media post

Description automatically generated

We want to use the time series summary tables since those include latitude and longitude, and the number of confirmed, deaths, and recovered cases.

1. Next, we want to clone the repo to our local drive. We can do so by clicking the green “Clone or download” button:

A screenshot of a cell phone

Description automatically generated

1. Open RStudio on your computers. On the menu bar located at the top of the screen, click “File” 🡪 “New Project” 🡪 “Version Control”:

A screenshot of a cell phone

Description automatically generated

Then select “Git” on the next page:

A screenshot of a cell phone

Description automatically generated

We want to paste the repo URL that we copied previously from the JHUCSSE Github page into the “Repository URL” box. We can leave the “Project directory name” as the default option. In the “Create project as a subdirectory of” box, click “Browse” and make sure that you have selected the “Lab 1” folder that you created in step 2 above. Then, click “Create Project”.

A screenshot of a cell phone

Description automatically generated

1. In your local lab 1 repo (mine is titled “p8321\_lab1\_pt1”), create a new folder called “data.” Copy and paste all the files in the “csse\_covid\_19\_time\_series” folder to the “data” folder you just created in your local repo. We are creating local copies of the datasets so that when we push our changes to Github, it is to our own Github repo and not the JHU CSSE repo.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. From Courseworks, download the R Markdown document we already created for you titled “Lab 1\_student” and add it to your local repo folder.
2. We’ll now generate some descriptive statistics and plots with the COVID-19 data in RStudio!
3. Homework: Finish generating the plots from the labs if you have not done so already. Knit your R Markdown document to a Word document and submit on Courseworks (due April 6, 2020 by 3:59 pm).