# TSCI5050: Introduction to Data Science

Alex F. Bokov, PhD, Fall 2019

## Homework #0: How to set up your online accounts

These are instructions for setting up the online accounts you will need for TSCI 5050, Introduction to Data Science. You can do everything described here using either FireFox or Chrome (this has not been tested on browsers other than these two and it may not work). You can do this on any computer as long as it has either of the supported browsers on it.

If you already have a GitHub and/or an RStudio Cloud account, you can skip the parts about creating them. Or you can create new accounts just for this class.

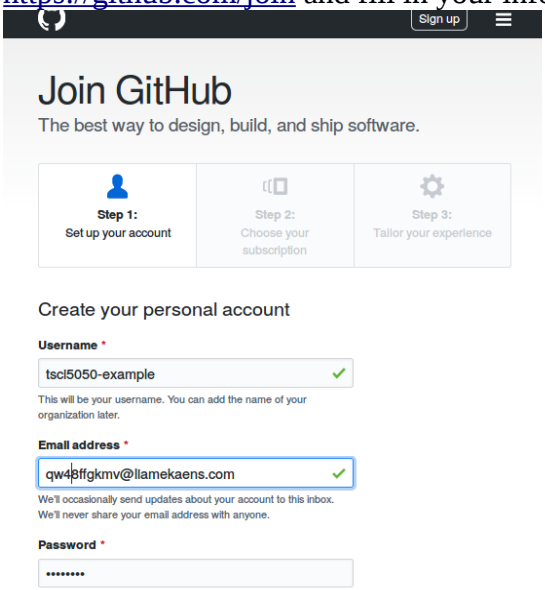
The overall steps will be:

1. Create GitHub account
2. Fork (I'll explain what that means in class) the course repository into your GitHub account
3. Create an RStudio Cloud account that uses your GitHub account to log in
4. Join the RStudio Cloud shared space we will be using for this class
5. Create a new RStudio session in that space based on the GitHub repository you forked
6. Run a script that will configure your RStudio
7. Generate a report summarizing your data.
8. Upload your own dataset that you wish to analyze during this class

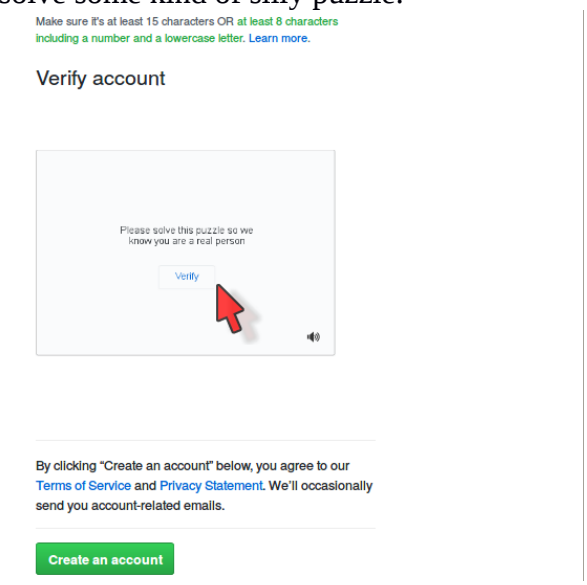
As this class progresses you will modify the starting script to the needs of your analysis. You will learn to explore, clean, and analyze your data as well as present your results.

### 1. Set up your GitHub account.

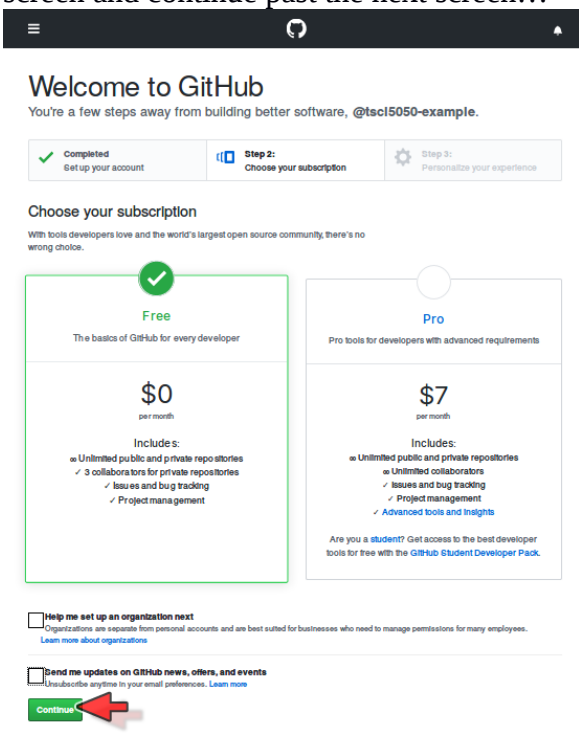
**1.1.** Go to the GitHub signup page https://github.com/join and fill in your info.




**1.2.** Verify that you are not a bot. It will ask you to solve some kind of silly puzzle.

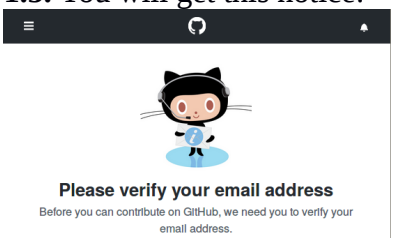


**1.3.** Click the Create an account button at the bottom of the screen and continue past the next screen…

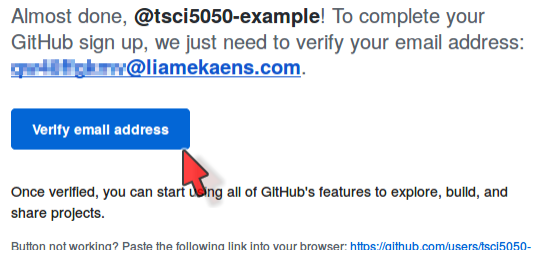



**1.4.** Fill out your interests… or not, it's optional.

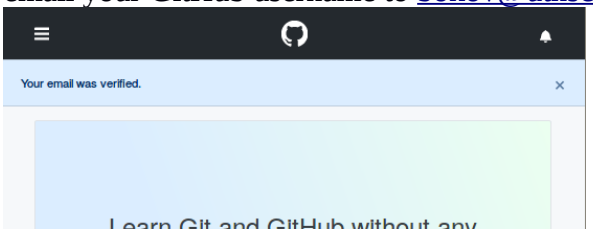

**1.5.** You will get this notice:




**1.6.** Go to your email and look for one from GitHub that looks like this. Click the verification button or link.

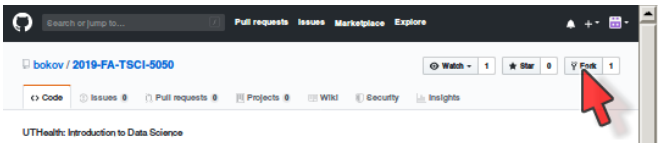



**1.7.** You will see this the screen below. Congratulations, you have a GitHub account. Please email your GitHub username to bokov@uthscsa.edu




## 2. Fork the course repository.

Go to https://github.com/bokov/2019-FA-TSCI-5050 and click the Fork button.



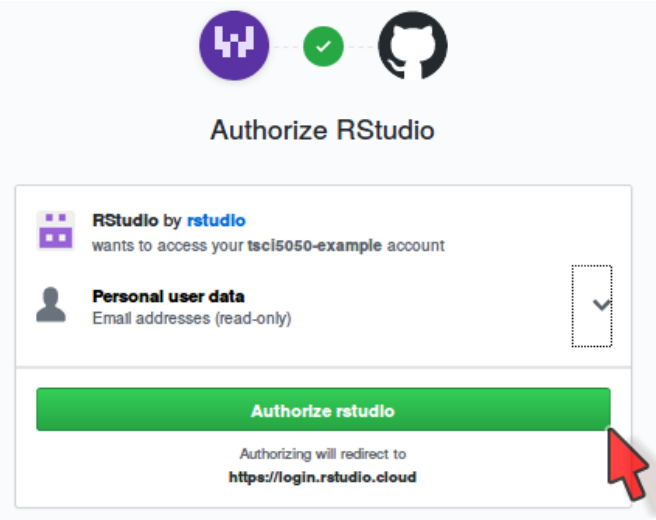

After a few moments you will see your own very own 'fork', or copy, of the class repository. We will go over in class what that means. For now, just **bookmark this page** and leave it open. Open a new tab or window in your web browser.

## 3. Sign up for RStudio Cloud

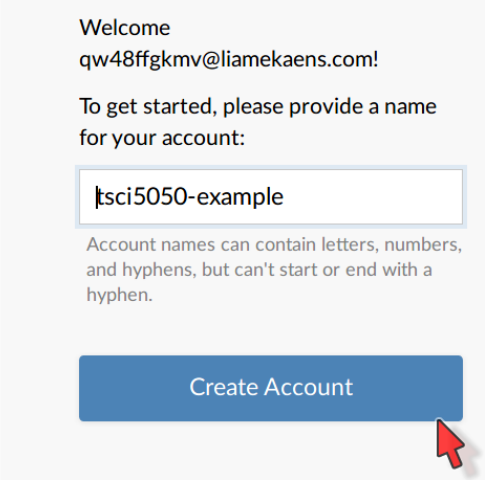
**3.1.** Now you will need an RStudio Cloud account. Go to https://login.rstudio.cloud/register and at the bottom, click 

.

**3.2.** Authorize RStudio to use GitHub for authentication.



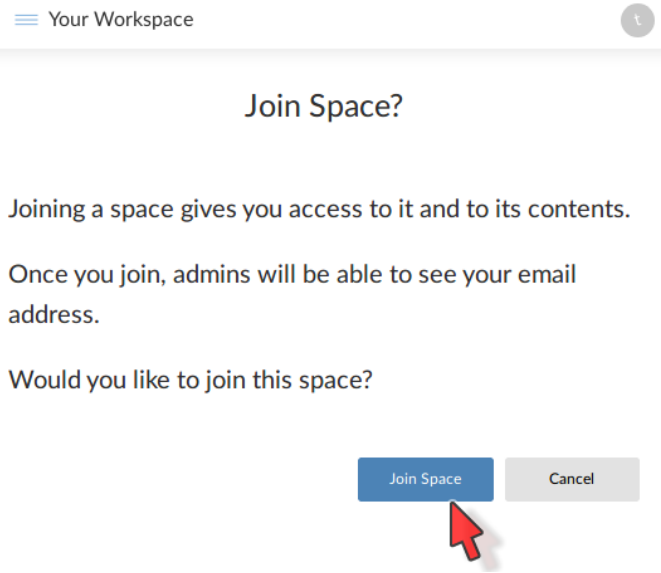

**3.3.** Create your RStudio account




## 4. Join our shared space on RStudio Cloud

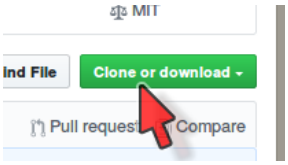
**4.1.** Go to this link (it has to be the full link)…
https://rstudio.cloud/spaces/13223/join?access_code=nHiV3SQw9QgJ2Puf9NG9uWM6KxPepEceX7bvlQgr

**4.2.** Click "Join Space".

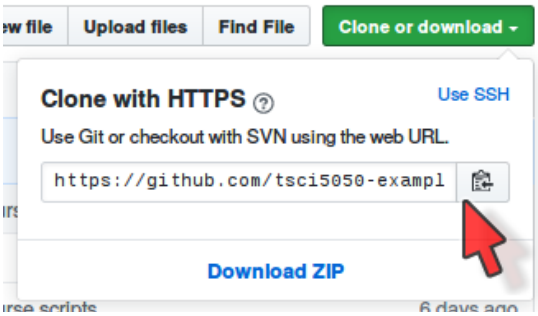


## 5. Create a new project using the repository you forked on GitHub in part 2.

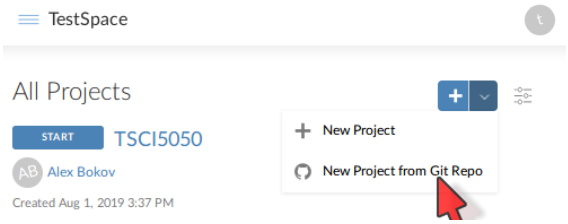
**5.1.** Go back to the GitHub browser tab and click here…




**5.2.** Make sure it says "**Clone with HTTPS**". Then click the little clipboard icon (📋) in the pop-up window.

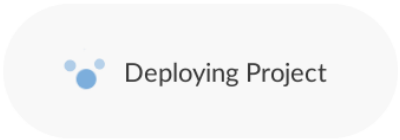



**5.3.** Switch back to the RStudio tab. Create a "New Project from Git Repo".

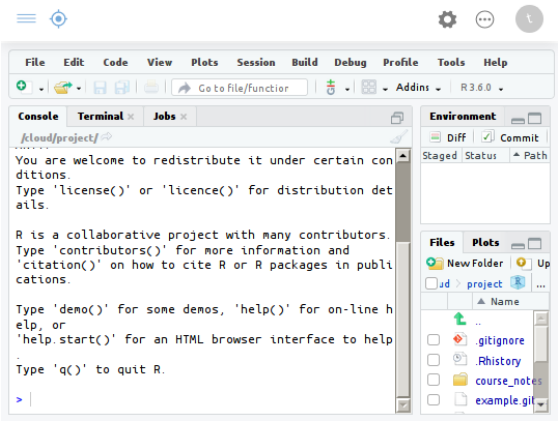



**5.4.** Paste the URL you copied in step 5.2 into the space where it asks you for the URL of your Git repository. Notice that the "Add packages from the base project" is checked in the picture. That's how it should be.


**5.5.** Wait for it to deploy.




**5.6.** There it is! The remote RStudio session you will be using for most of this class and perhaps beyond. **Bookmark this page**, you'll be going here a lot.

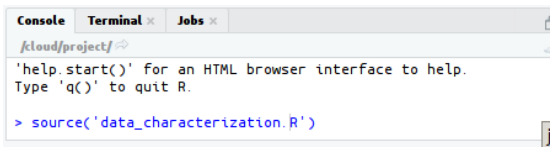



## 6. Run your scripts for the first time.

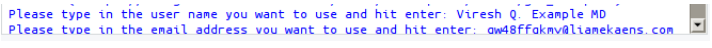
Running this script will configure some settings on your project and download some additional files you will need.

**6.1.** Copy-paste the following command into your R console, but don't walk away yet!

```
source('data_characterization.R')
```

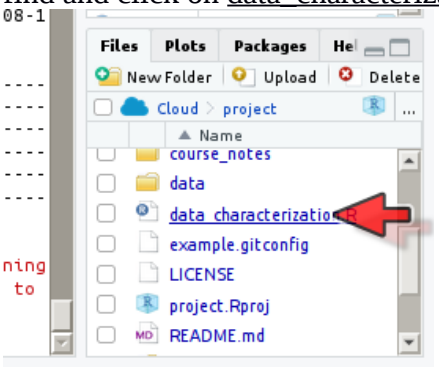



**6.2.** You will be prompted for your name and email. Put in the name and email you want to use hitting "Enter" after each. The user input below is just an example, fill in your own information.



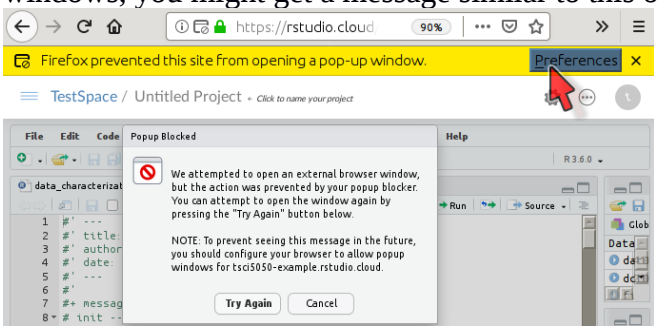
# 7. Build your first RMarkdown report.

**7.1.** Wait for the previous step to finish running so that you again see the "**>**" prompt the R console. In the lower-right pane of your window, in the "Files" tab, find and click on data_characterization.R.

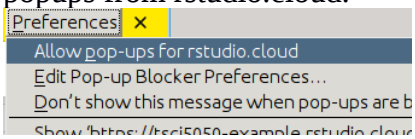



**7.2.** It's the same script you ran from the console. It will open in a new tab in the top-left panel of your screen. At the top you will see a small notepad icon like this: . Click on it to compile a report.

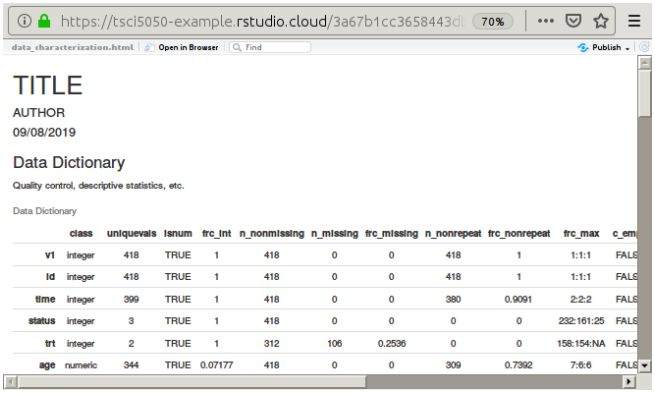

Note: If your browser is configured to prohibit popup windows, you might get a message similar to this one:




If that happens, configure your browser to allow popups from rstudio.cloud.




**7.3.** Either way, in the end you should get a new window that looks like this:

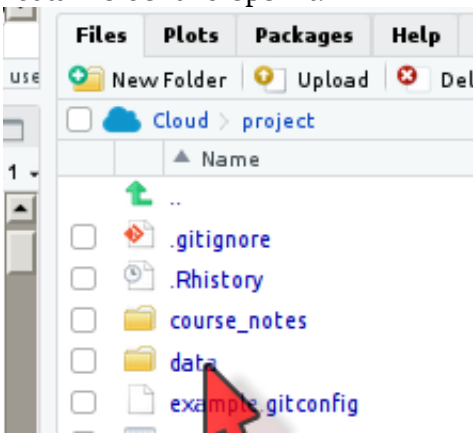



We will go over in class how to use this information.

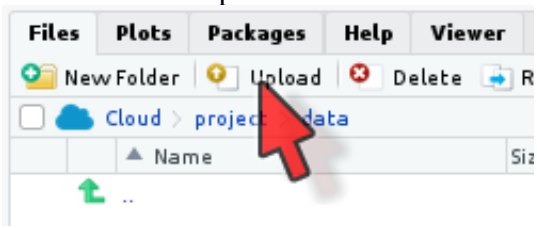
## 8. Upload your data.

It's time to think about what dataset you would like to analyze during this course. If possible, please try to use real data that's part of your intended research-- you're going to be spending effort on this, might as well choose something that will also advance your goals beyond this class.

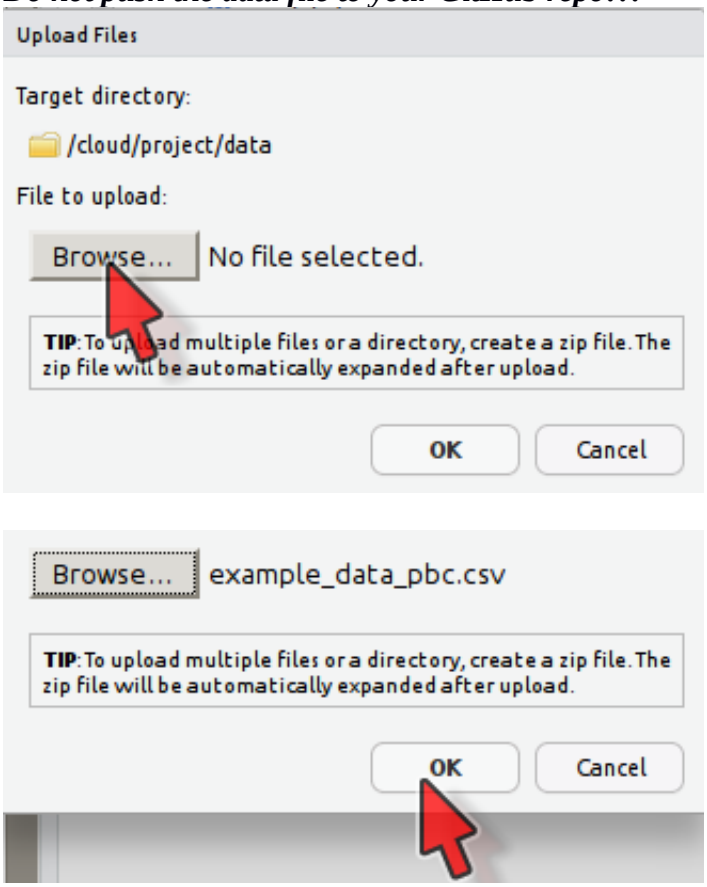
**8.1.** In the "Files" tab of the lower-right panel, find the "data" folder and open it.




**8.2.** Click the "Upload" button.




**8.3.** Click on the "Browse" button in the resulting dialogue and find the data file on your local computer that you want to upload and click "OK" to upload it. ***Do not upload PHI-containing data. Do not upload data to any directory other than the one called 'data'. Do not push the data file to your GitHub repo!!!***







**That's it! Congratulations, you are all set for TSCI 5050. See you in class where I will explain what all these steps did, what to do next, and answer any other questions that you might have by then.**