

# ENVIRONMENTAL DATA ANALYTICS

ENV 872L  
Spring 2020



# WELCOME TO ENVIRONMENTAL DATA ANALYTICS

**Kateri Salk, PhD**

Visiting Assistant Professor of Water Resources

[kateri.salk@duke.edu](mailto:kateri.salk@duke.edu)

Grainger Hall 3115



Office hours: Thursdays 3:00-4:00, Fridays 10:30-11:30

# WELCOME TO ENVIRONMENTAL DATA ANALYTICS

**Teaching Assistant: Njeri Kara**

2<sup>nd</sup> year MEM, Energy & Environment

[ruth.kara@duke.edu](mailto:ruth.kara@duke.edu)

Office hours: TBD

# WELCOME TO ENVIRONMENTAL DATA ANALYTICS

## **29 students**

- 7 – Water Resources Management
- 6 – Coastal Environmental Management
- 6 – Environmental Science & Conservation
- 4 – International Development & Policy
- 3 – Environmental Economics & Policy
- 2 – Ecotoxicology & Environmental Health
- 2 – Energy & Environment
- 1 – Business & Environment
- 1 – PhD

What types of data are common in your field?

# DATA PRACTICES ARE IN DEMAND

- Software skills are necessary prerequisites for employment
- Reproducible data pipelines are efficient and consistent → increased productivity
- Troubleshooting skills increase the capacity for independence
- Effective reporting → clear and efficient communication of outcomes/findings

# COURSE OBJECTIVES

1. Gain proficiency in the language and application of R software
2. Synthesize information from datasets, working from start to finish in the “data pipeline”
3. Develop skills to identify and apply appropriate statistical and graphical approaches for environmental datasets
4. Integrate multiple technological applications involved in contemporary data analysis, workflow, and management

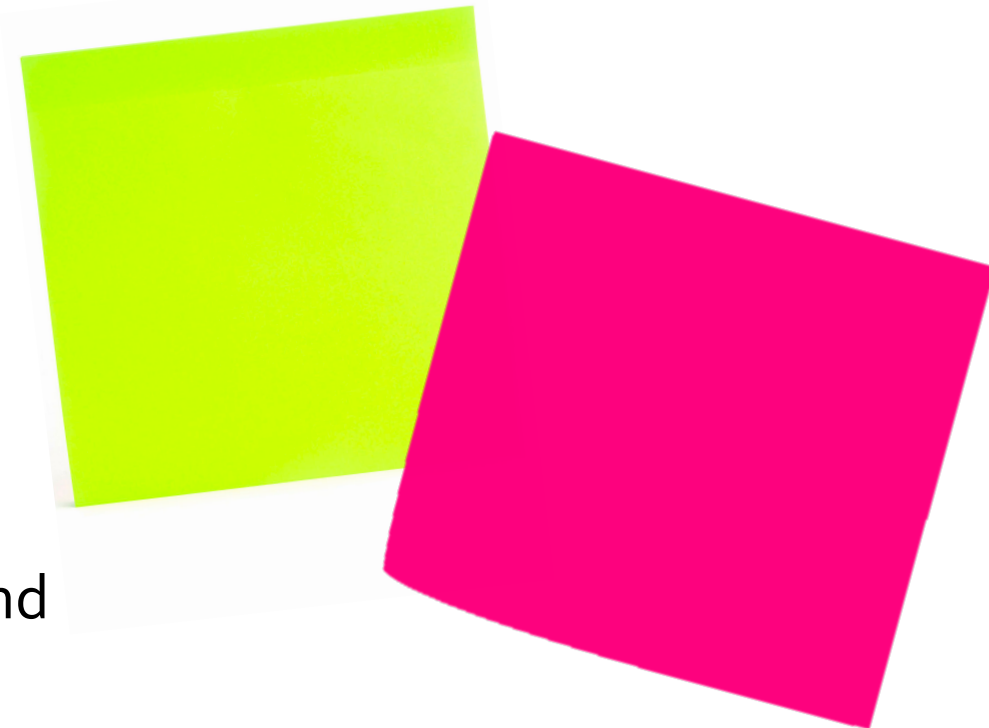
# TECHNICAL LOGISTICS

## Computers

- Option 1: Lab computer, mapped to CIFS server
- Option 2: Personal computer, download all necessary software
- Software installation guide provided

## During class

- Green sticky note:
  - I am keeping up
  - I am ready to move on
- Pink sticky note:
  - I am stuck
  - my task is still in progress
- 1 % bonus if you keep your stickies until the end



# COURSE LOGISTICS

- Course materials in Github repository
  - Pull updates each class
  - Save your own progress inside files
- Helpful tools
  - R for Data Science
  - Stack Overflow
  - Google
- Assignments submitted to Sakai



# INQUIRY-BASED LEARNING

Construction of knowledge through scientific practices

Involves:

- Problem solving skills
- Active participation
- Knowledge discovery by the learner
- Inductive and/or deductive approach

Outcomes: inquiry based learning > traditional instruction

## Engage

The purpose of the **ENGAGE** stage is to pique student interest and get them personally involved in the lesson, while preassessing prior knowledge.

1

## Explore

The purpose of the **EXPLORE** stage is to get students involved in the topic; providing them with a chance to build their own understanding.

2

## Explain

The purpose for the **EXPLAIN** stage is to provide students with an opportunity to communicate what they have learned so far and figure out what it means.

3

## Extend

The purpose for the **EXTEND** stage is to allow students to use their new knowledge and continue to explore its implications.

4

## Evaluate

The purpose for the **EVALUATION** stage is for both students and teachers to determine how much learning and understanding has taken place.

5

# EXPECTATIONS

- Software up to date (outside class)
- Troubleshooting
- Missing class
- Cooperative learning

# GRADING

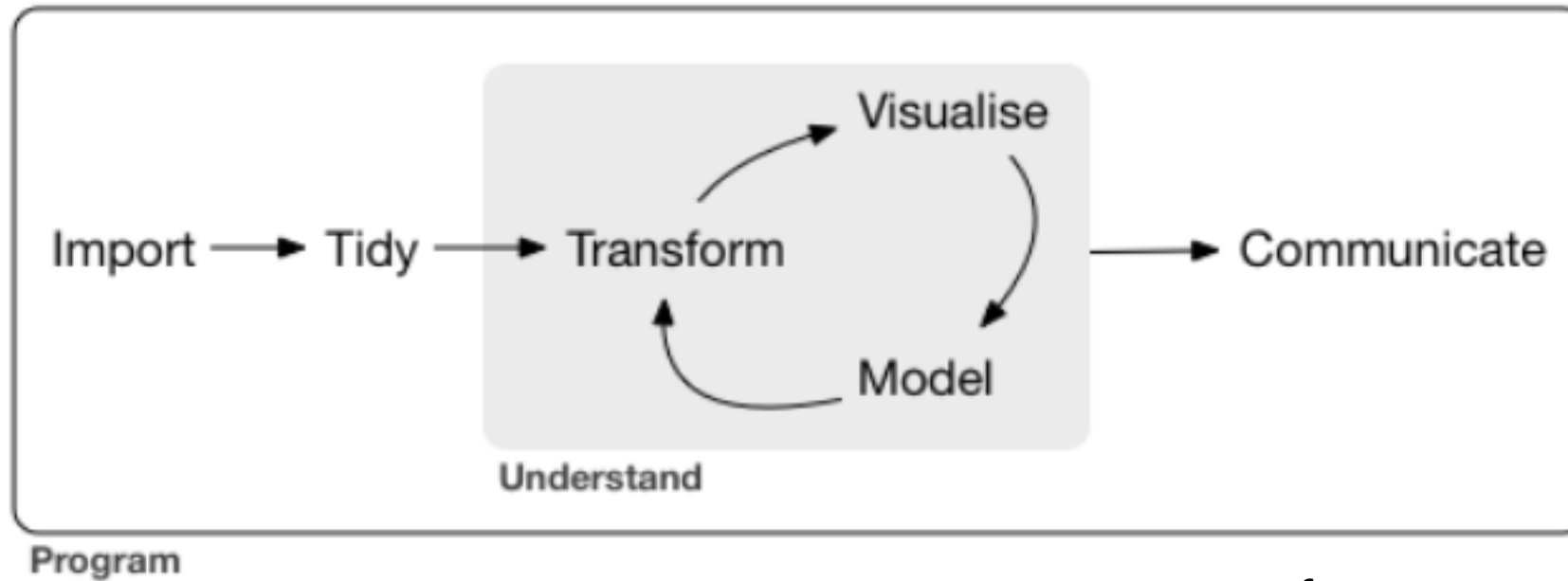
## Assignments 75 %

- 11 assignments, 1 per unit
- Evaluated for completion, application of concepts/code, discussion

## Project 25 %

- Choose example dataset and research question
- Run through data workflow
- Report

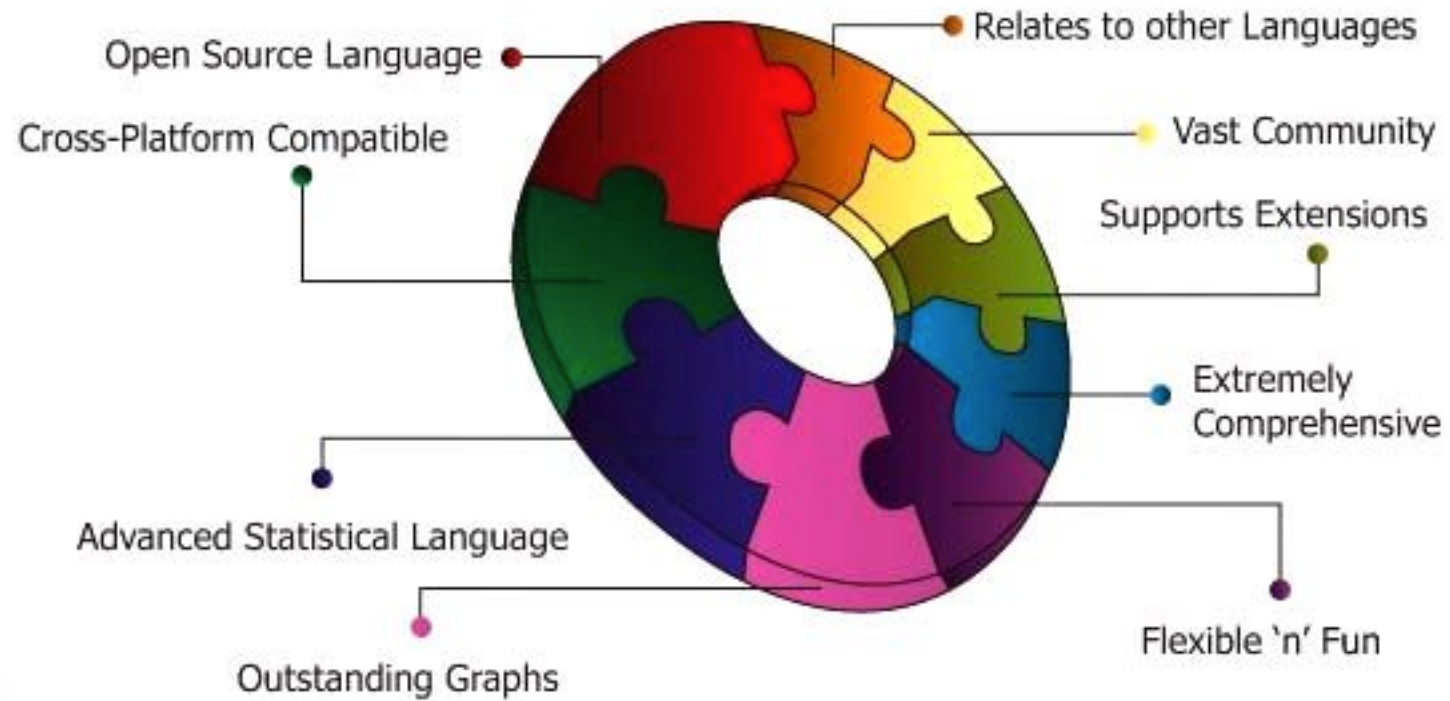
# THE DATA SCIENCE PIPELINE



From: R for Data Science

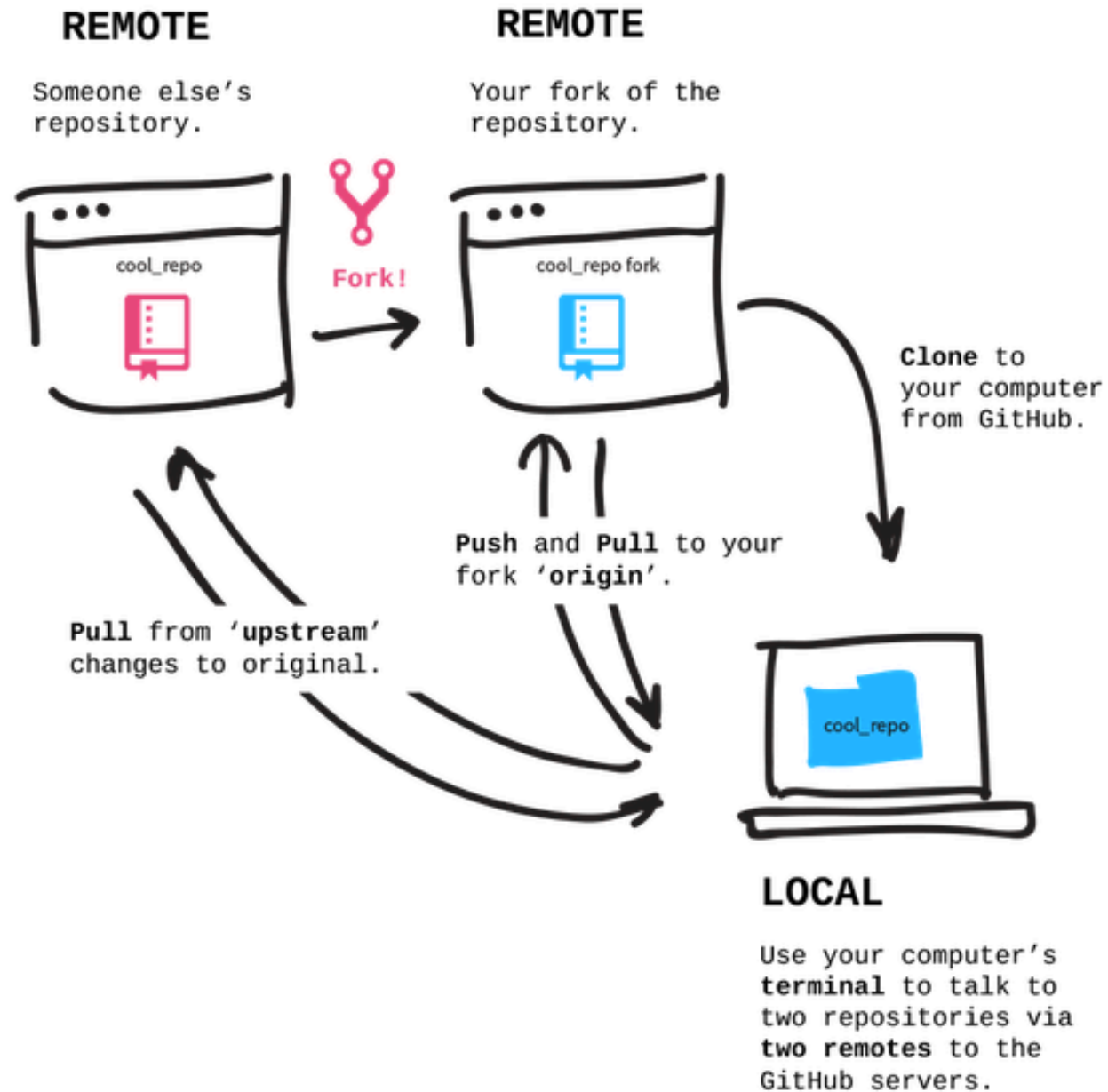
Which components of data science have you practiced before?

# WHY R?



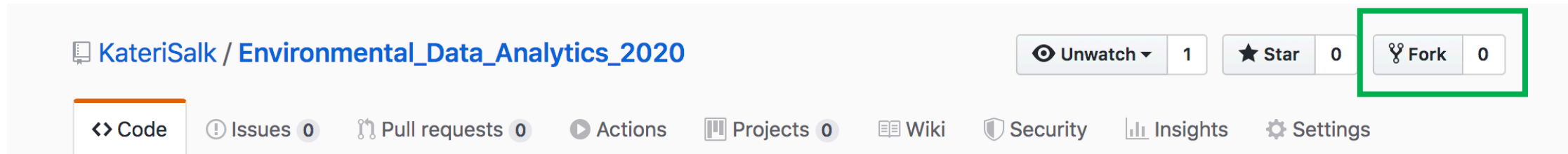
[imarticus.org](http://imarticus.org)

# GITHUB SETUP



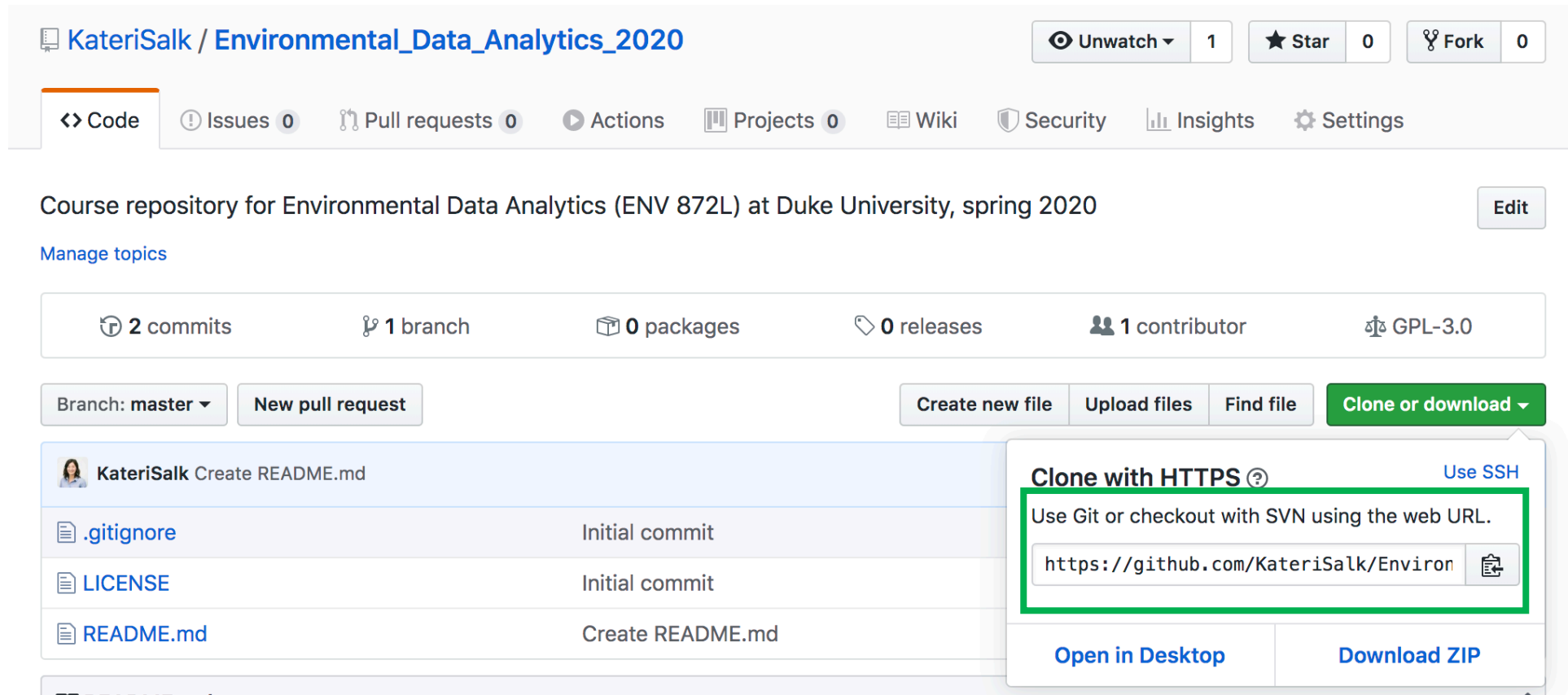
# GITHUB SETUP: FORKING

1. Navigate to [https://github.com/KateriSalk/Environmental\\_Data\\_Analytics\\_2020](https://github.com/KateriSalk/Environmental_Data_Analytics_2020)
2. Fork the repository to your GitHub account



# GITHUB SETUP: CLONING

## 3. Copy the link to your forked repository



The screenshot shows the GitHub interface for the repository 'KateriSalk / Environmental\_Data\_Analytics\_2020'. The repository description is 'Course repository for Environmental Data Analytics (ENV 872L) at Duke University, spring 2020'. The repository has 2 commits, 1 branch, 0 packages, 0 releases, 1 contributor, and is licensed under GPL-3.0. The 'Clone or download' button is highlighted in green, and a dropdown menu is open, showing the 'Clone with HTTPS' option selected. The URL 'https://github.com/KateriSalk/Environ' is displayed in the dropdown, along with a copy icon. The dropdown also includes options for 'Open in Desktop' and 'Download ZIP'.

KateriSalk / Environmental\_Data\_Analytics\_2020

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security Insights Settings

Course repository for Environmental Data Analytics (ENV 872L) at Duke University, spring 2020

Manage topics

2 commits 1 branch 0 packages 0 releases 1 contributor GPL-3.0

Branch: master New pull request Create new file Upload files Find file Clone or download

KateriSalk Create README.md

.gitignore	Initial commit
LICENSE	Initial commit
README.md	Create README.md

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

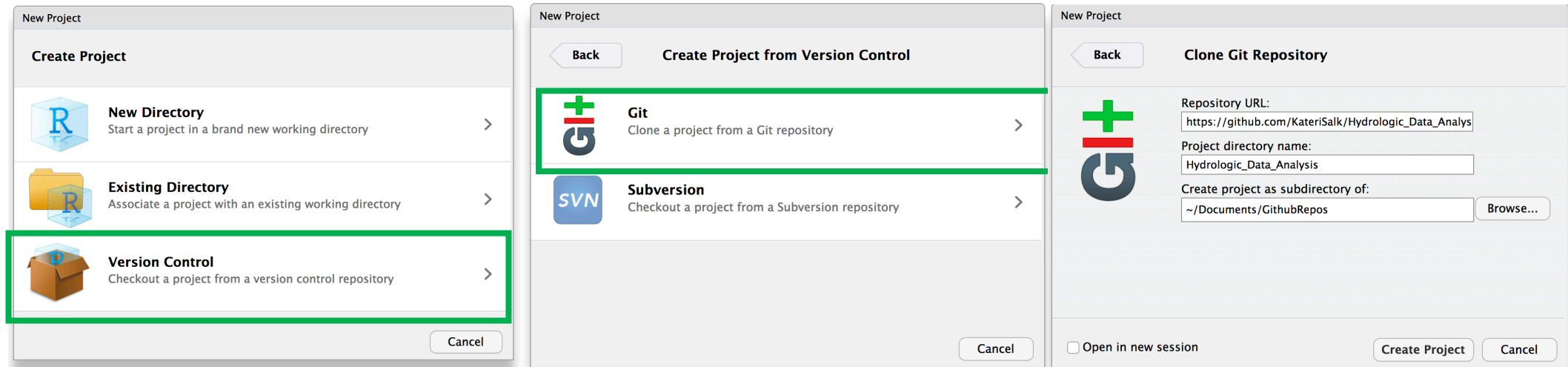
https://github.com/KateriSalk/Environ

Open in Desktop Download ZIP



# GITHUB SETUP: CLONING

4. Open RStudio and go to File > New Project...
5. Select "Version Control", then "Git"
6. Paste your forked repo URL and choose a folder where the local repo will be saved



# GITHUB SETUP: COMMIT AND PUSH

Open the **Git\_Help file** and follow the instructions in the Editing, Committing, Pushing section.

1. Navigate to the "README.md" file in the Files tab and open it.
2. Type your name after "Student:" and save.
3. Now that you have edited a file, it should now appear in the Git tab. Click the box to the left of the file, where a check mark should now appear.
4. Press the **"Commit" button**. A new window should appear that shows the changes that have been made to the file.
5. Write a message detailing the edits you've made to the README file. You should always include a commit message to your commits so that your future self and/or your collaborators will know what changes were made. Click "Commit".
6. Click the green upward facing arrow: **the "Push" button**. Your remote repository is now up to date with **your local repository**.

# GITHUB SETUP: ADD AND PULL FROM THE REMOTE

Follow the instructions in the “Pulling from the upstream remote” section of the Git\_Help file:

1. In the Terminal tab, type: `git remote add upstream https://github.com/KateriSalk/Environmental_Data_Analytics_2020`
2. In the Terminal tab, type: `git pull upstream master`