

# WATER DATA ANALYTICS

ENVIRON 790.02

Spring 2022



# WELCOME TO WATER DATA ANALYSIS

**Kateri Salk, PhD (she/her)**

Adjunct Assistant Professor of Water Resources

Aquatic Ecologist, Tetra Tech

kateri.salk@duke.edu



Background:

- Limnology, biogeochemistry
- Environmental informatics
- Empirical and process-based modeling

# OFFICE HOURS

I will set office hours based on student schedules. Fill out the poll!

<https://www.when2meet.com/?13981797-YJeci>

- 1 hour of virtual office hours each week
- Before & after class in person

# COURSE OBJECTIVES

1. Synthesize information on fundamental and applied topics in water resources using quantitative analysis
2. Apply the appropriate steps of the data analytics pipeline to answer questions about aquatic systems
3. Develop marketable skills in data management, analysis, and communication for the aquatic sciences field

# COURSE SCHEDULE

**Week 1:** Intro and R boot camp

**Weeks 2-4:** Physical properties of lakes & rivers

**Weeks 4-7:** Water quality in lakes & rivers

**Weeks 8-12:** Time series analysis, spatial analysis, high frequency data

**Week 13:** Final project workshop

**Finals week:** Project presentations, project report

# ASSIGNMENTS

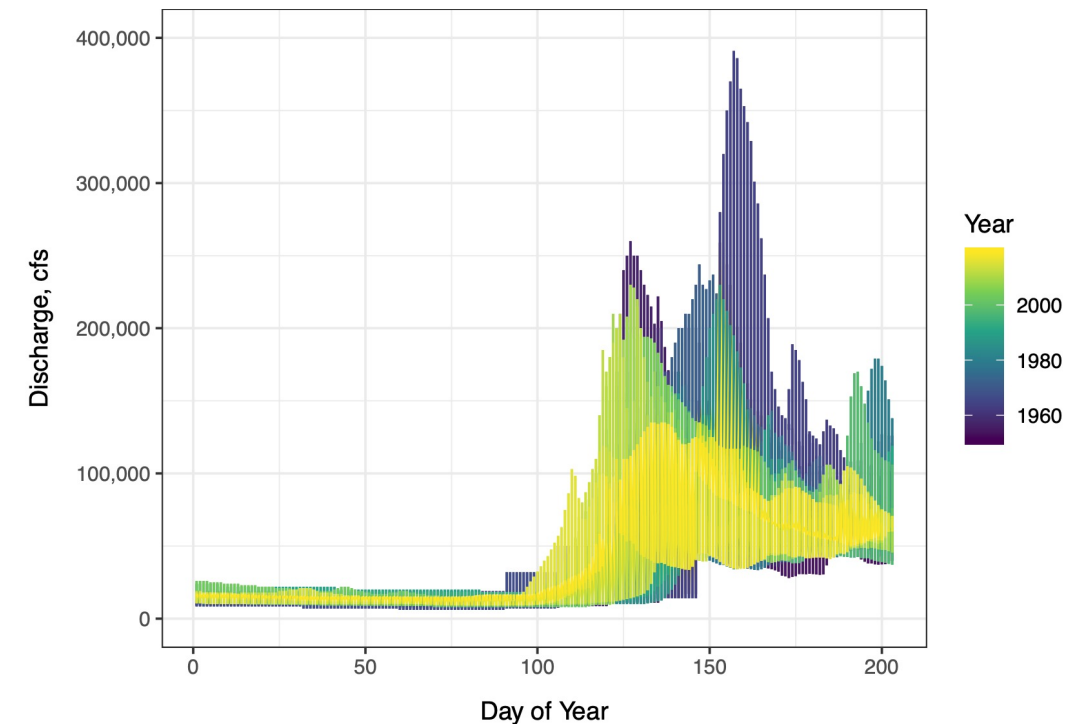
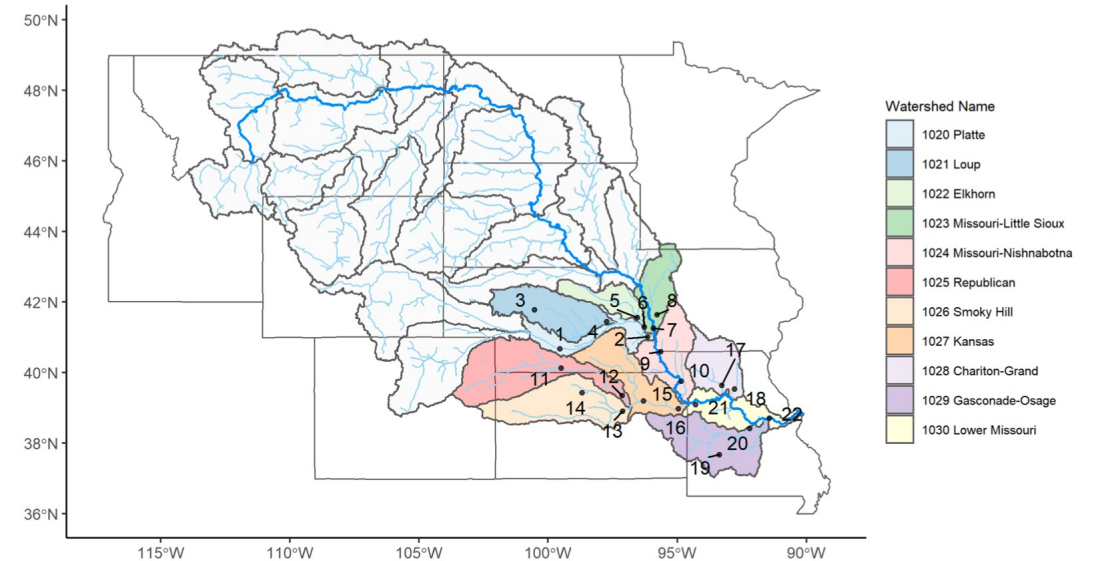
- Each week will have an assignment (homework) (80 % of grade)
- Competency based
- I will provide a key – try to complete the assignment without it
- Following completion, take the corresponding survey on Sakai
- Due dates in syllabus are targets, not deadlines

# FINAL PROJECT

Choose a water-focused question and perform a quantitative analysis to answer this question

Project stages (% final grade)

- Initial idea (2%)
- Workshop session (5%)
- Final report (10%)
- Final presentation (3%)



## OTHER SYLLABUS STUFF

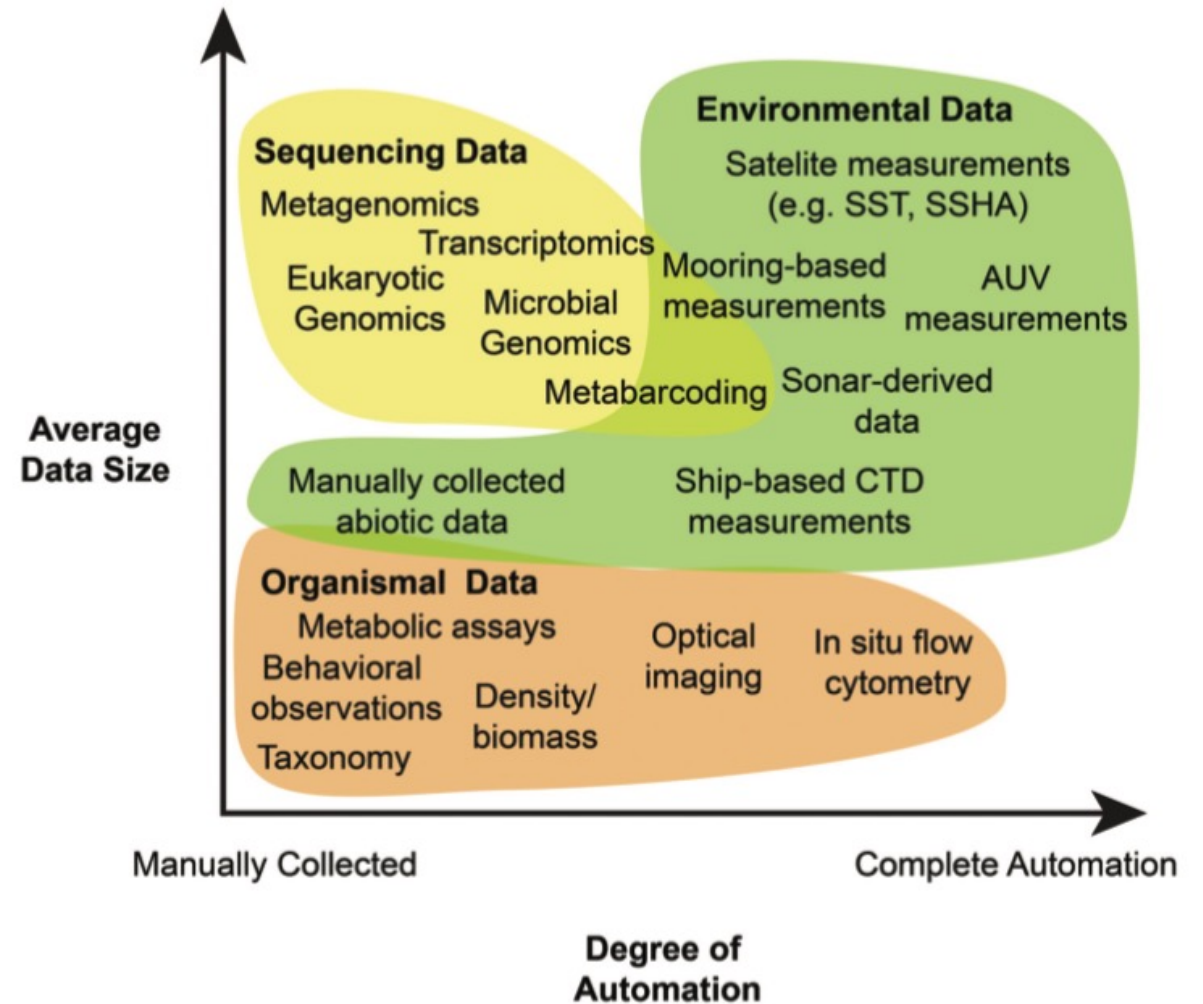
- Schedule specifics
- COVID procedures
- Accommodations



# INTEGRATING BIG DATA INTO AQUATIC ECOLOGY

6 key challenges:

1. Recognizing big data
2. Data handling
3. Analytical techniques
4. Verification
5. Data sharing
6. Developing knowledge infrastructure



*Durden et al. 2017*

# 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.

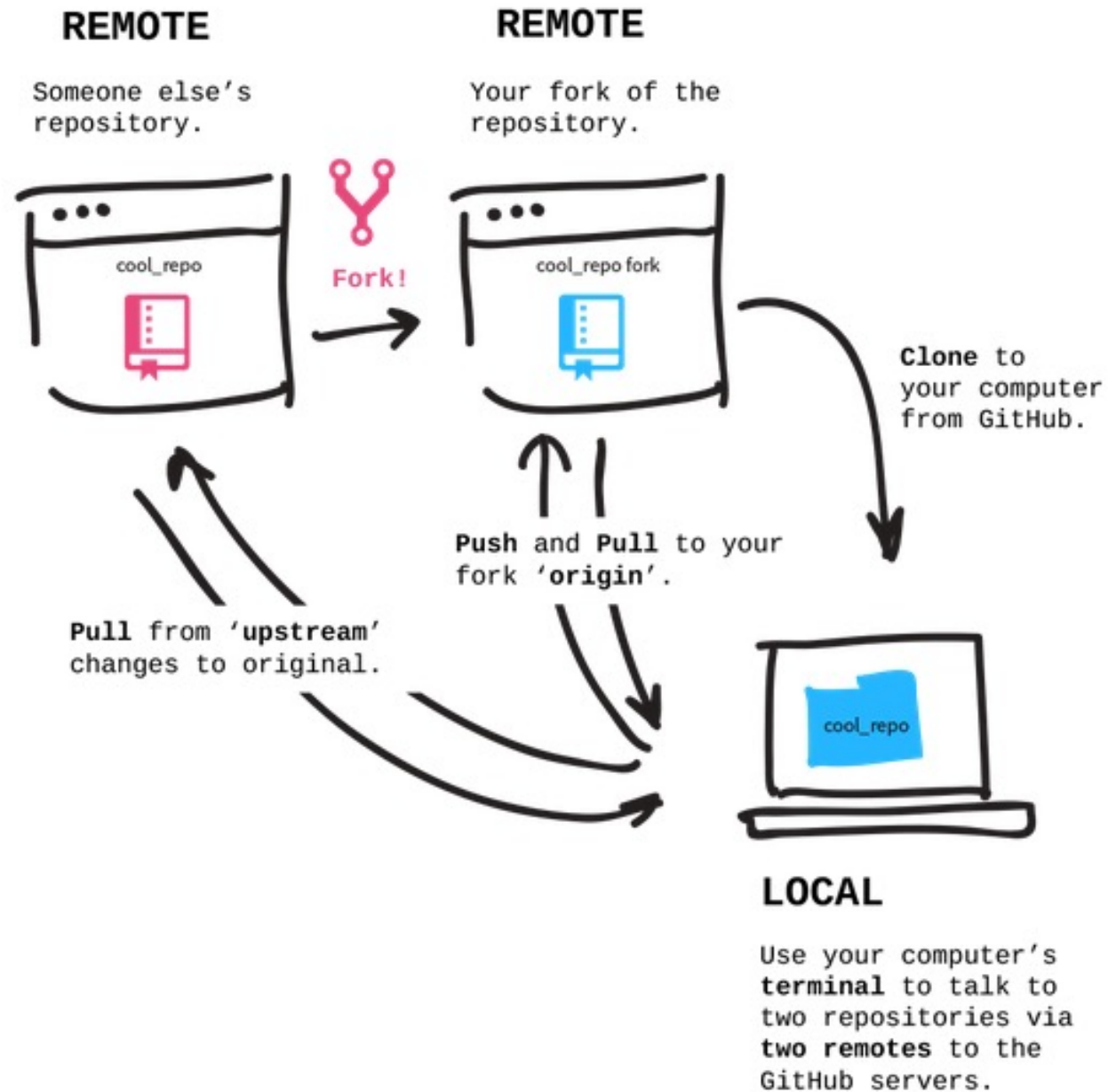
4

## Evaluate

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

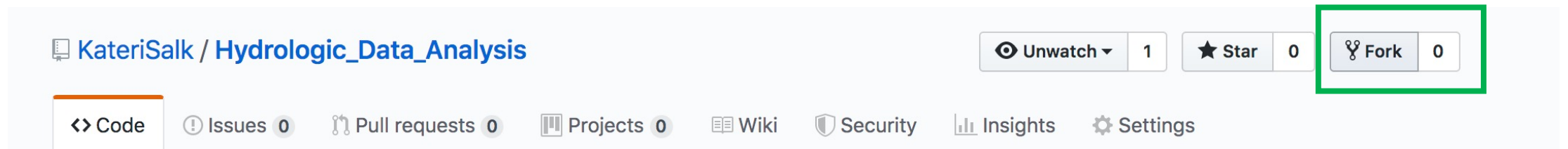
5

# GITHUB SETUP



# GITHUB SETUP: FORKING

1. Navigate to [https://github.com/KateriSalk/Water\\_Data\\_Analytics\\_2022](https://github.com/KateriSalk/Water_Data_Analytics_2022)
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 / Hydrologic\_Data\_Analysis'. The repository has 1 commit, 1 branch, 0 releases, 1 contributor, and is licensed under GPL-3.0. The 'Clone or download' button is highlighted, and the dropdown menu is open, showing the 'Clone with HTTPS' option selected. The URL 'https://github.com/KateriSalk/Hydrologic\_Data\_Analysis' is displayed in the input field, and a copy icon is visible next to it. The 'Open in Desktop' and 'Download ZIP' buttons are also visible at the bottom of the dropdown.

KateriSalk / Hydrologic\_Data\_Analysis

Unwatch 1 Star 0 Fork 0

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

Duke University course: Hydrologic Data Analysis (Fall 2019) Edit

Manage topics

1 commit 1 branch 0 releases 1 contributor GPL-3.0

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

KateriSalk Initial commit

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

Clone with HTTPS Use SSH

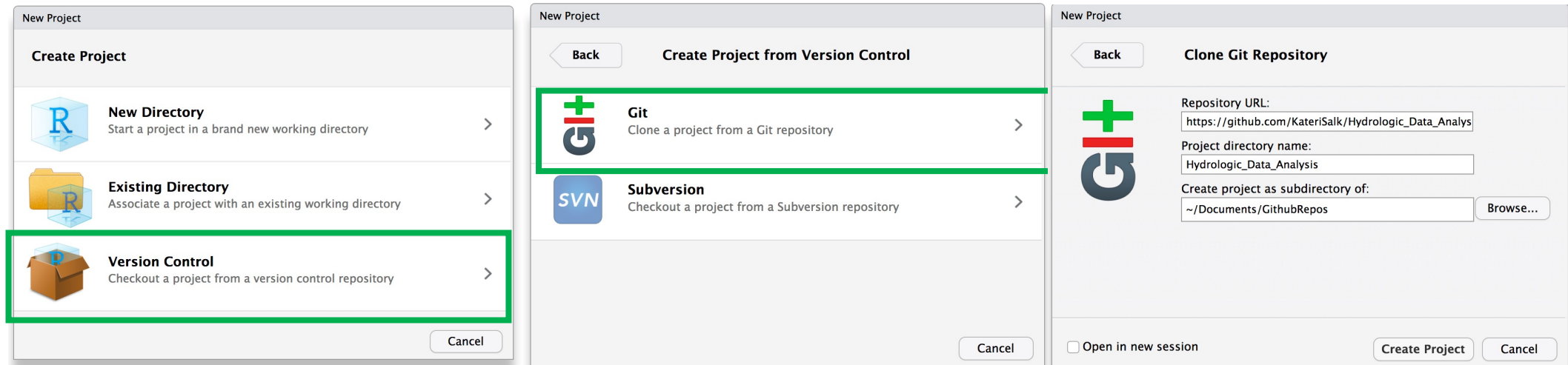
Use Git or checkout with SVN using the web URL.

https://github.com/KateriSalk/Hydrologic\_Data\_Analysis

Open in Desktop Download ZIP

# GITHUB SETUP: CLONING

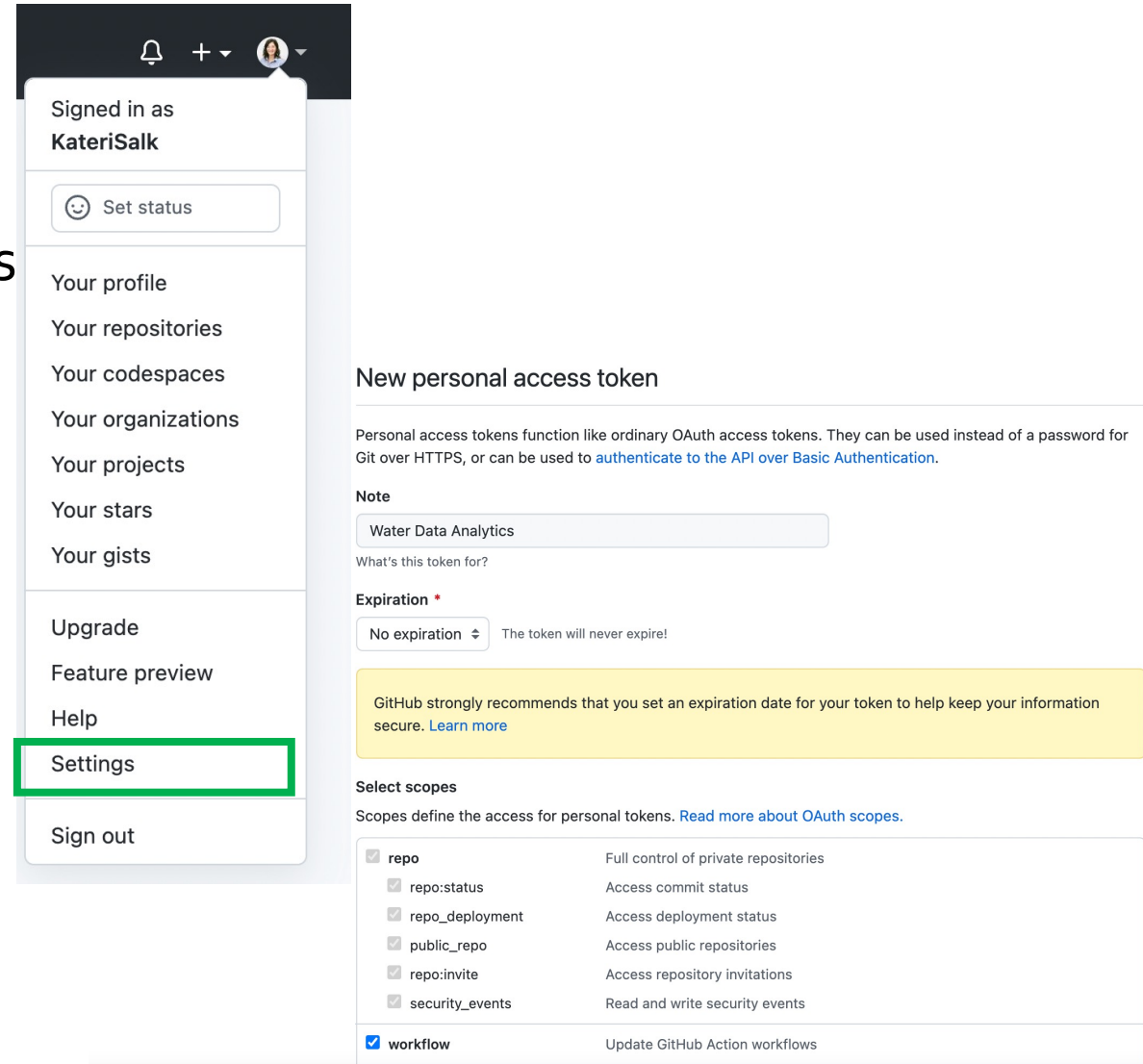
4. Open RStudio and go to File > New Project...
5. Select "Version Control", then "Git"
6. Paste your forked repo URL and choose a folder where the local repo will be saved



# GITHUB SETUP: ERROR ON CLONING?

Set up a personal access token

1. From your profile menu, select Settings
2. Go to Developer Settings
3. Go to Personal access tokens
4. Click "Generate new token"
5. Add info
  - Note: suggest "Water Data Analytics"
  - Expiration
  - Repo, workflow, and user scope



The image shows a screenshot of the GitHub web interface. On the left, a profile menu is open, displaying the user's name 'KateriSalk' and a list of options: 'Set status', 'Your profile', 'Your repositories', 'Your codespaces', 'Your organizations', 'Your projects', 'Your stars', 'Your gists', 'Upgrade', 'Feature preview', 'Help', 'Settings' (highlighted with a green box), and 'Sign out'. On the right, the 'New personal access token' page is visible. It includes a text input field with the value 'Water Data Analytics', a note about token expiration, and a section for selecting scopes. The 'repo' scope is checked, and the 'workflow' scope is also checked. The 'expiration' dropdown is set to 'No expiration'.

Signed in as  
**KateriSalk**

Set status

Your profile  
Your repositories  
Your codespaces  
Your organizations  
Your projects  
Your stars  
Your gists

Upgrade  
Feature preview  
Help  
**Settings**  
Sign out

### New personal access token

Personal access tokens function like ordinary OAuth access tokens. They can be used instead of a password for Git over HTTPS, or can be used to [authenticate to the API over Basic Authentication](#).

**Note**

Water Data Analytics

What's this token for?

**Expiration** \*

No expiration ⌵ The token will never expire!

GitHub strongly recommends that you set an expiration date for your token to help keep your information secure. [Learn more](#)

**Select scopes**

Scopes define the access for personal tokens. [Read more about OAuth scopes](#).

<input checked="" type="checkbox"/> repo	Full control of private repositories
<input checked="" type="checkbox"/> repo:status	Access commit status
<input checked="" type="checkbox"/> repo_deployment	Access deployment status
<input checked="" type="checkbox"/> public_repo	Access public repositories
<input checked="" type="checkbox"/> repo:invite	Access repository invitations
<input checked="" type="checkbox"/> security_events	Read and write security events
<input checked="" type="checkbox"/> workflow	Update GitHub Action workflows

# GITHUB SETUP: PERSONAL ACCESS TOKEN

Copy access token from GitHub, then enter the following into R:

```
install.packages("gitcreds")  
library(gitcreds)  
gitcreds_set()
```

Then, paste the access token when prompted

From: <https://happygitwithr.com/https-pat.html>



# GITHUB SETUP: COMMIT AND PUSH

- Open the Git\_Help file and follow the instructions in the Editing, Committing, Pushing section.
- Familiarize yourself with how to keep the local, remote, and upstream remote repositories up to date with each other.