ENVIRONMENTAL DATA ANALYTICS: WEEK 2

Spring 2022
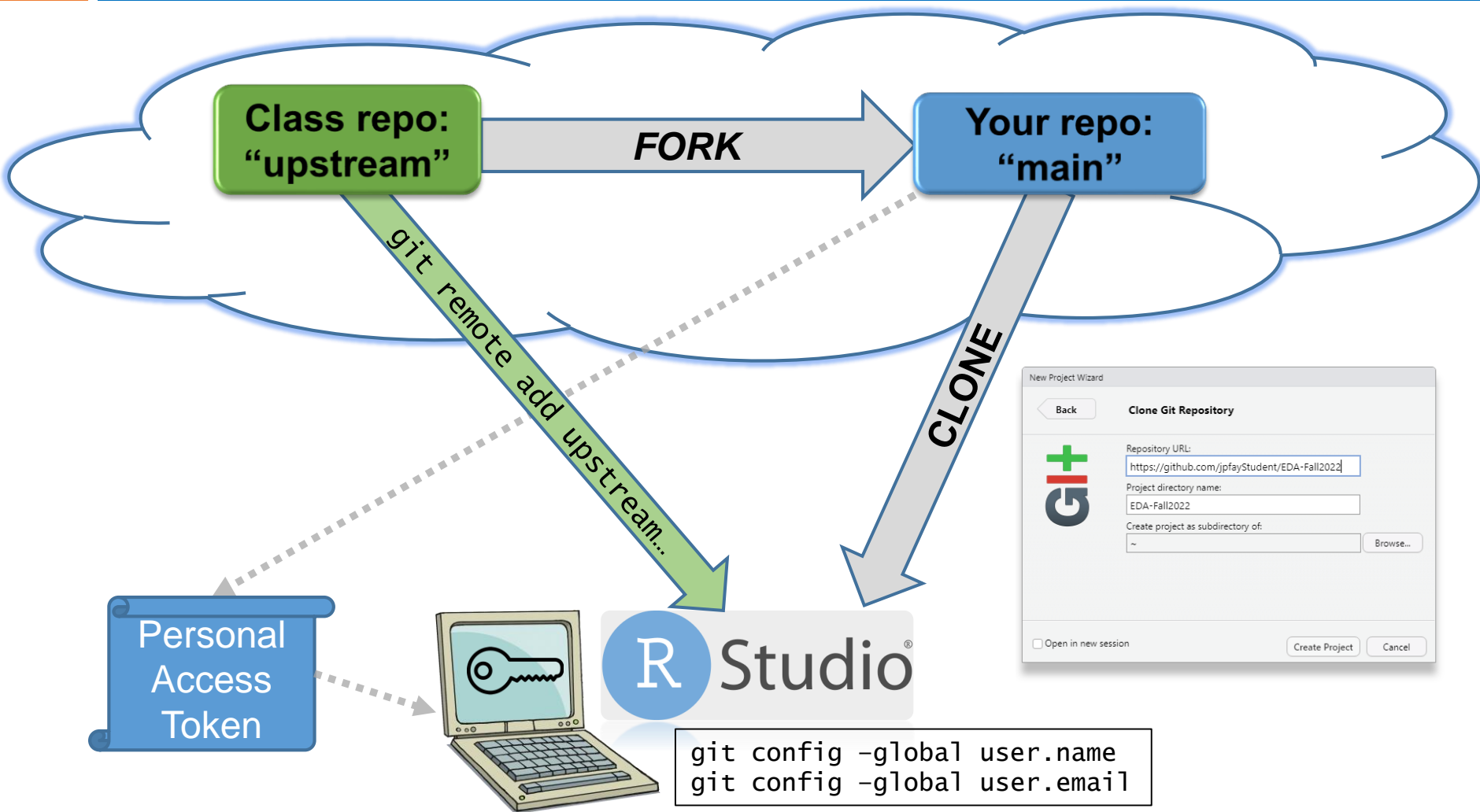
Nicholas School of the Environment - Duke University

Tech Roundup
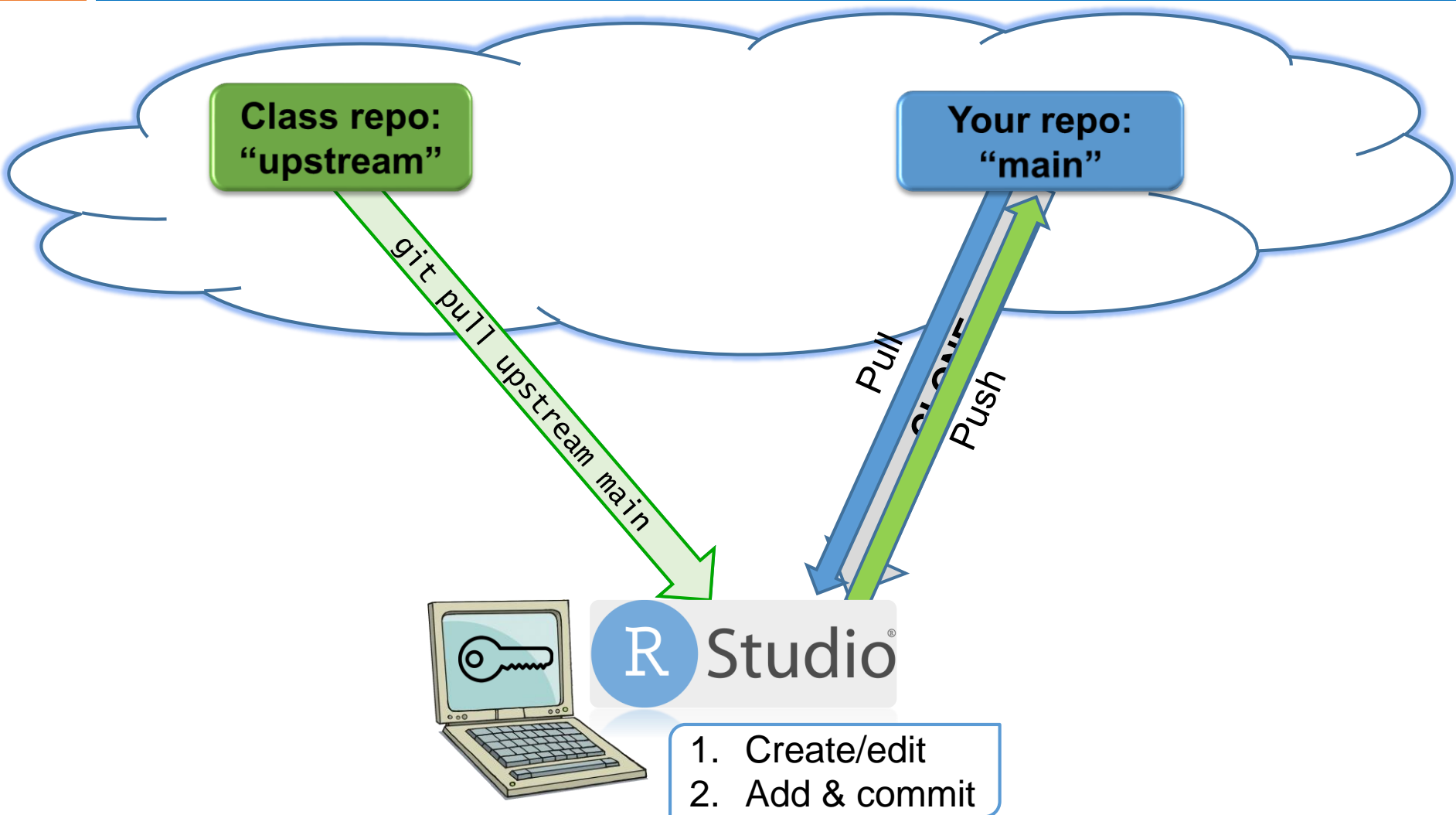
- Issues joining Slack? Navigating Slack?

- Issues installing applications:
  R… | RStudio… | Git…

- Issues creating your R project?
  Forking… | Cloning… | PAT… | Committing… | Pushing…

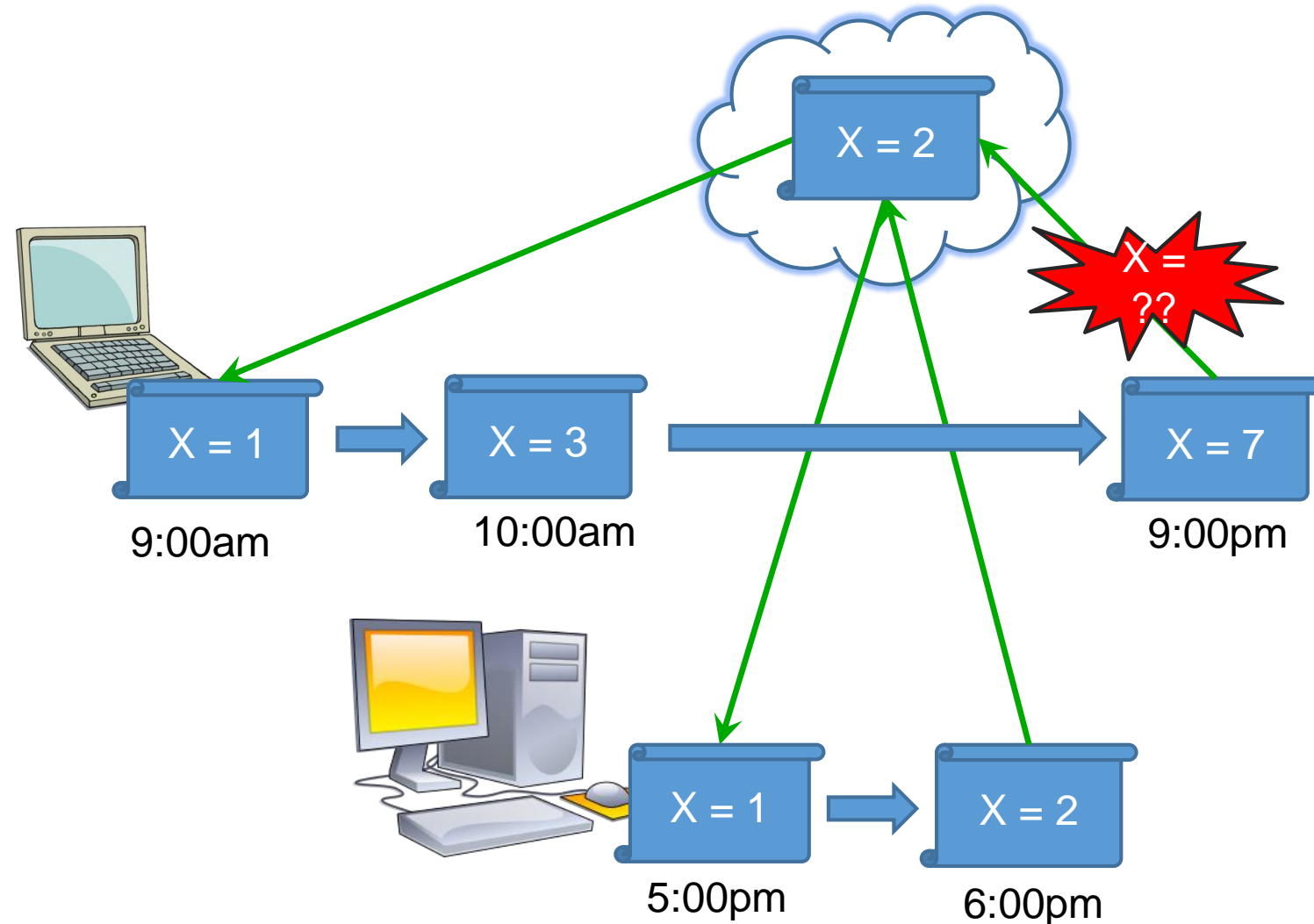# Explaining Git/GitHub: Setup

Explaining Git/GitHub: Use

Git: Merge conflicts

X = 2

X = ??

X = 1
9:00am

X = 3
10:00am

X = 7
9:00pm

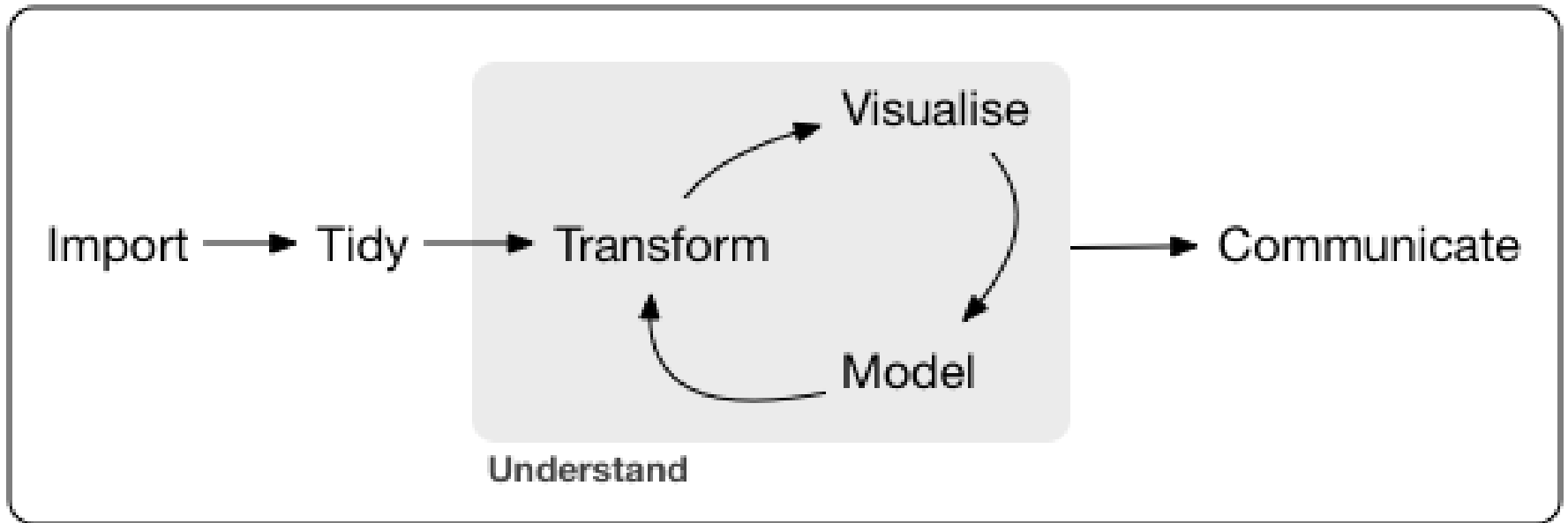X = 1
5:00pm

X = 2
6:00pm

Exercise…

If Git goes totally sideways…

- ☐ Make a back up (e.g. zip) of you project folder

- ☐ Quit RStudio and rename project folder

- ☐ Restart RStudio and create a new project linked to your forked repo

- ☐ Copy over any missing items from your renamed folder to your newly cloned repository

Q&A: What is Data Analytics

https://vita.had.co.nz/papers/tidy-data.pdf

Data Types

- Understand what it means to "tidy" data

- Differentiate "primary" and "secondary" data

- Differentiate "qualitative" and "quantitative" data

- Identify different file types used in data analytics and discuss why some formats are better than others in terms of transparency and reproducibility

- Describe the various data structures used in data analytics and what each are used for: *Vectors*, *matrices*, *arrays*, *data frames*, *lists*

- Understand the difference between *R* and *RStudio*

- Become familiar with the typical layout of an RStudio session

Up next: Module 2

Reproducibility & Coding Basics

- ☐ What is "reproducibility"? Why is it important?

- ☐ Working with RMarkdown files…

- ☐ Writing R code…