Data wrangling I

Intro to Data Science

Shawn Santo

Today's agenda

- Manipulating data to extract meaning

- Introduce a new member of the `tidyversse-dplyr`

- Tidy data, what it is, why it is useful

- Reinforce the version control workflow

Announcements

- Soft deadline for Lab 02 is February 4 at 11:59pm ET
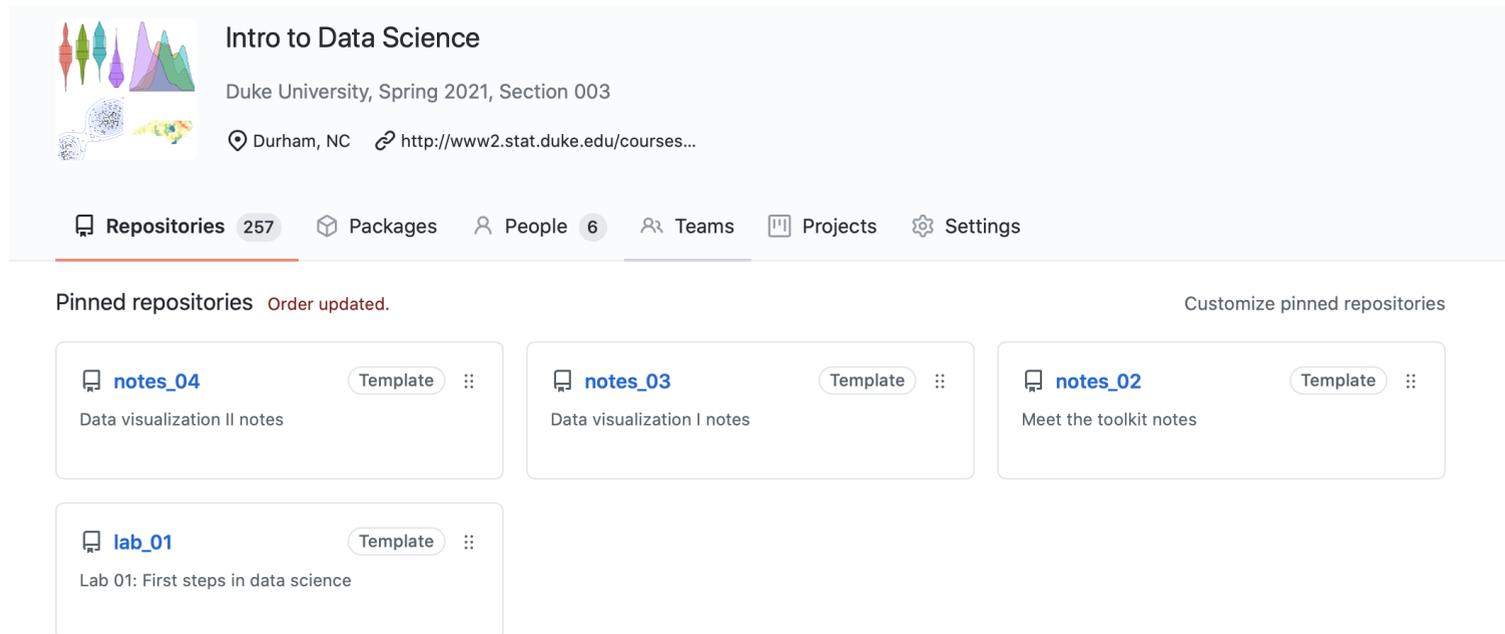
- Keep working on Homework 01

Getting started

- Create your personal private repository by clicking https://classroom.github.com/a/-LXARqNZ

- Follow the steps as we have done previously to clone this and create a new RStudio project in the RStudio Docker containers.

Tip: When navigating between assignments (RStudio project), click the project icon in your files pane for that project before working in the Rmd file.

Completed notes

I'll post completed notes in the template repository. For example, you'll find `notes_05_solutions.Rmd` in the `notes_05` repository in our GitHub organization.

Intro to Data Science

Duke University, Spring 2021, Section 003

⊙ Durham, NC   🔗 http://www2.stat.duke.edu/courses…

🖥 **Repositories** 257   📦 **Packages**   👤 **People** 6   👥 **Teams**   🗔 **Projects**   ⚙ **Settings**

Pinned repositories   Order updated.   Customize pinned repositories

🖥 **notes_04** `Template` ⠿	🖥 **notes_03** `Template` ⠿	🖥 **notes_02** `Template` ⠿
Data visualization II notes	Data visualization I notes	Meet the toolkit notes

🖥 **lab_01**   `Template` ⠿
Lab 01: First steps in data science