

# Data visualization I

## Intro to Data Science

Shawn Santo

# Today's agenda

- Grammar of graphics
- Data visualization with `ggplot2`
- Reinforce the version control workflow

# Getting started

# GitHub to RStudio

You are the owner of a repository on GitHub that you'd like to work on locally

1. Create your personal private repository by clicking <https://classroom.github.com/a/pnpmg3zG>
2. Navigate to the repository you just created using the newly provided link
3. Click on the green Code button in that repository, and copy the git URL
4. Go to your RStudio Docker container at <https://vm-manage.oit.duke.edu/containers/rstudio>
5. In RStudio, go to *File* → *New Project* → *Version Control* → *Git*
6. Copy and paste the git URL of your assignment repo into the dialog box labeled *Repository URL*. Adjust the folder to where you want this assignment located.
7. Click *Create Project*, and the files from your GitHub repo will be displayed in the *Files* pane in RStudio.

# Push your work to GitHub

You've modified some files, saved them, and are ready to update your remote repository on GitHub

1. Prepare your modified files for **commit** by **staging** them in the `Git` tab.
2. Click the **commit** button in the `Git` tab, a new window will open
3. Verify that you're happy with all the changes
4. Enter a short informative message in the **commit** dialog box and hit the **commit** button, close this window
5. In the `Git` tab, click the green up arrow to **push** your work to GitHub

# Reproducible data analysis

# Reproducibility checklist

What does it mean for a data analysis to be "reproducible"?

## Near-term goals:

- Are tables and figures reproducible from the code and data?
- Does code actually do what you think it does?
- In addition to what was done, is it clear **why** it was done? (e.g., how were parameter settings chosen?)

## Long-term goals:

- Can the code be used for updates to the current data?
- Can the code be used for other data?
- Can you extend the code to do other things?

# Toolkit



- Scriptability → R
- Literate programming (code, narrative, output in one place) → R Markdown
- Version control → Git / GitHub