

# You don't have to be an expert to create something meaningful

Maria Tackett  
Duke University

WSDS 2020  
October 2020



[bit.ly/wsds2020-create](https://bit.ly/wsds2020-create)

Motivation

NC Fines and  
Fees

Interactive  
Learning  
Tools

Challenges  
+  
Lessons  
Learned

Motivation

NC Fines and  
Fees

Interactive  
Learning  
Tools

Challenges  
+  
Lessons  
Learned



# Intro to Data Science

Data visualization

Data wrangling

Inference

Statistical modeling

Communicating results

R + GitHub

## What comes next???

## Benefits to students

- Learning by doing
- Communication skills
- Confidence as a statistician
- Sense of professional belonging

## Benefits to faculty

- Teaching growth
- Advancing scholarship
- Interdisciplinary partnerships
- *It's fun!*

NC Fines and  
Fees

Motivation

Interactive  
Learning  
Tools

Challenges  
+  
Lessons  
Learned

# Fines and fees in North Carolina

Who

1 undergraduate student

What

Develop interactive tool to explore NC fines and fees data

When

Fall 2019 independent study

Where

On campus

# Why

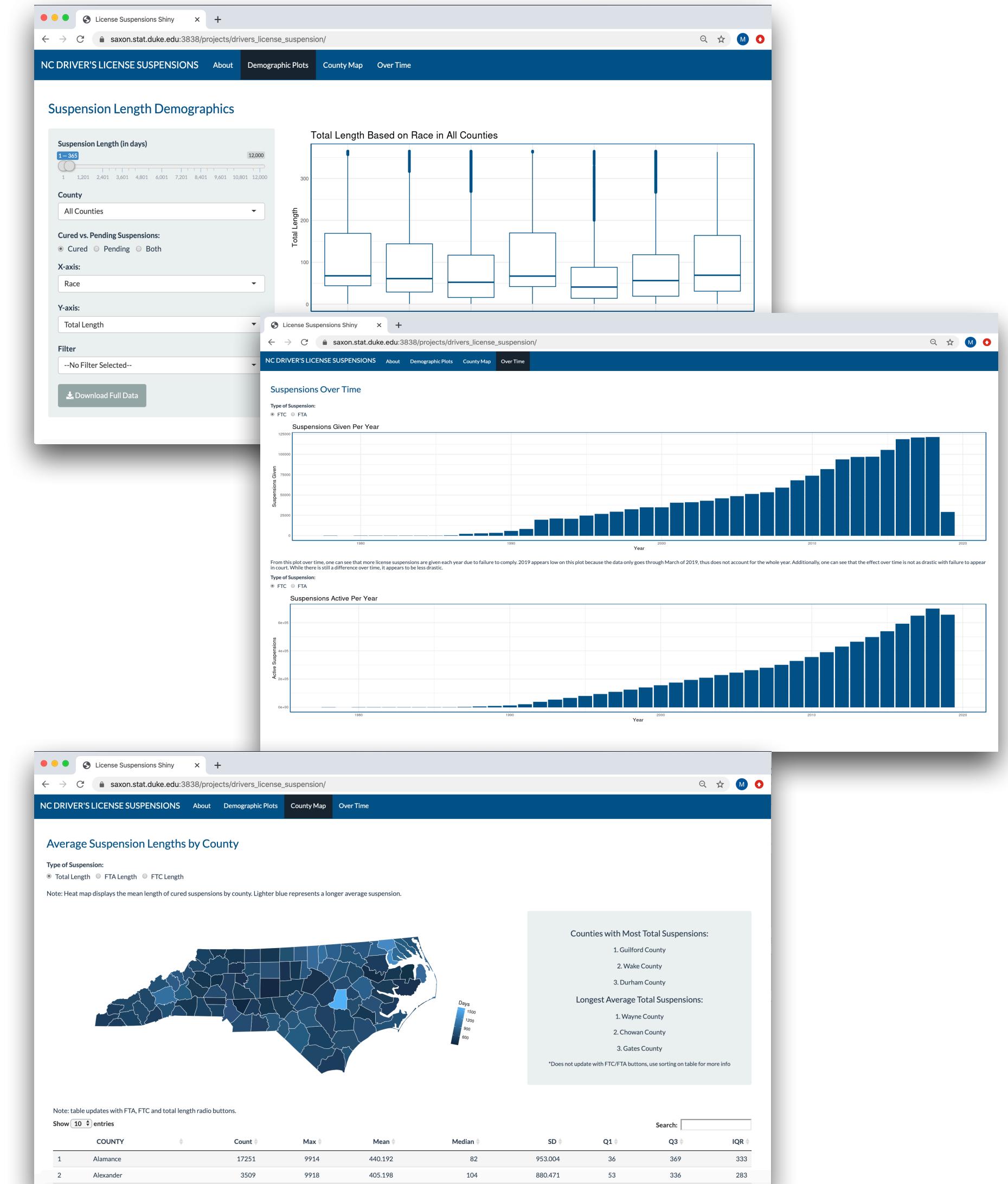
- ✓ Interesting and meaningful topic
- ✓ Data are publicly available
- ✓ Data are complex and messy
- ✓ Project has a clear audience

# Weekly meetings

- Technical knowledge and skills development
- Learning about NC legal system
- Project updates
- Preparation for meetings with collaborators
- Set goals for the week

# Activities

- R Shiny app for data exploration
- Written report
- Presentation to local stakeholders
- Review of existing work
- Preliminary data analysis + modeling
- Written reflection



# After the independent study

 Duke Law | Center for Science and Justice

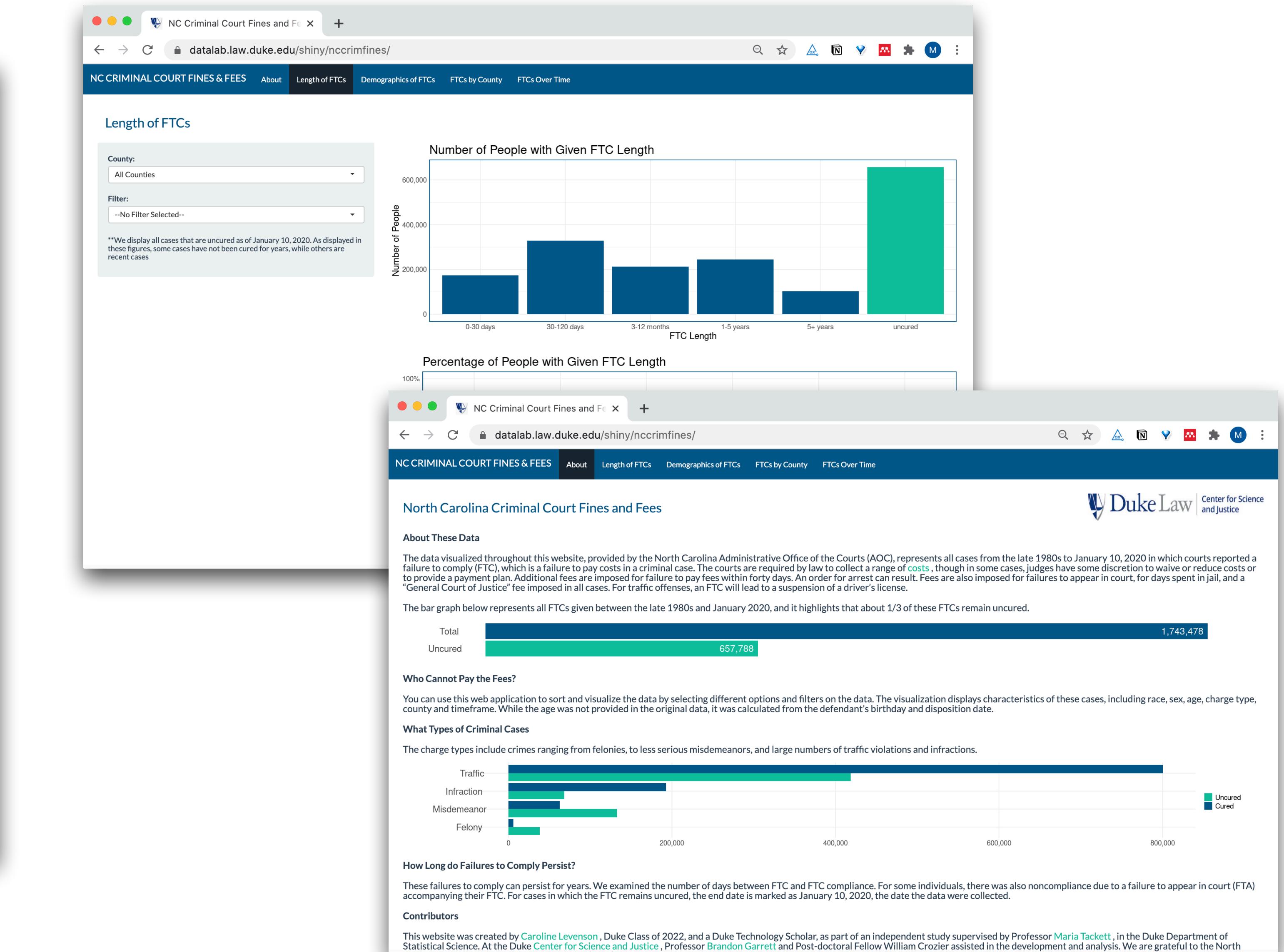
A Center for Science and Justice Report:

## The Explosion of Unpaid Criminal Fines and Fees in North Carolina

April 22, 2020



NC Fines and Fees



Motivation

NC Fines and  
Fees

Interactive  
Learning  
Tools

Challenges  
+  
Lessons  
Learned

# Project inspirations

**ShinyEd**

Statistics education apps created with Shiny. Source code available on [GitHub](#).

Developed in collaboration with [Brittany Cohen](#).

*Fork me on GitHub*

Diagnostics for simple linear regression  
Diagnostic plots for simple linear regression where the variables have a linear up/down, curved up/down, and fan-shaped relationship.

Distribution calculator  
Calculate probabilities under the normal, binomial, t, F, and chi-square distributions.

Central Limit Theorem for Proportions  
Central limit theorem for distribution of sample proportions for varying sample sizes and proportions of success.

Central Limit Theorem for Proportions  
Central limit theorem for distribution of sample proportions for varying sample sizes and proportions of success.

ShinyEd

PennState  
Eberly College of Science

- I. Introduction
- II. Introductory Apps
  - Chapter 1: Data Gathering
  - Chapter 2: Data Description
  - Chapter 3: Basic Probability
  - Chapter 4: Statistical Inference
- III. Upper Division Apps
  - Chapter 5: Probability
  - Chapter 6: Regression
  - Chapter 7: ANOVA
  - Chapter 8: Time Series
  - Chapter 9: Sampling

## Introduction to the Book Of Apps for Statistics Teaching

The apps in this collection are the work of undergraduate students majoring in Statistics and Data Science at Penn State University. These students took part in the **BOAST** program in 2017, 2018, 2019, and 2020. (BOAST = Book Of Apps for Statistics Teaching). The program included either a DataCamp course on Shiny (2018, 2019, 2020) or a full day workshop at the U.S. Conference On Teaching Statistics on Developing R Shiny Applications to Enhance Teaching and Learning presented by Justin Post and Herle McGowan from North Carolina State University (2017). This was followed by a full-time summer research experience developing the apps in this collection throughout the months of June and July. Finally, the **BOAST** students take part in a one-credit research

BOAST

# Creating interactive learning tools

Who

4 undergraduate students + project manager

What

Develop interactive learning tools to explore statistical concepts

When

6 weeks in Summer 2020

Where

Zoom + Slack

# Why

- ✓ Develop deeper understanding of statistical concepts
- ✓ Focus on data visualization + communication
- ✓ Project has a clear audience
- ✓ Feasible to do remotely

# Team workflow



Primarily asynchronous work

- Communication using Slack



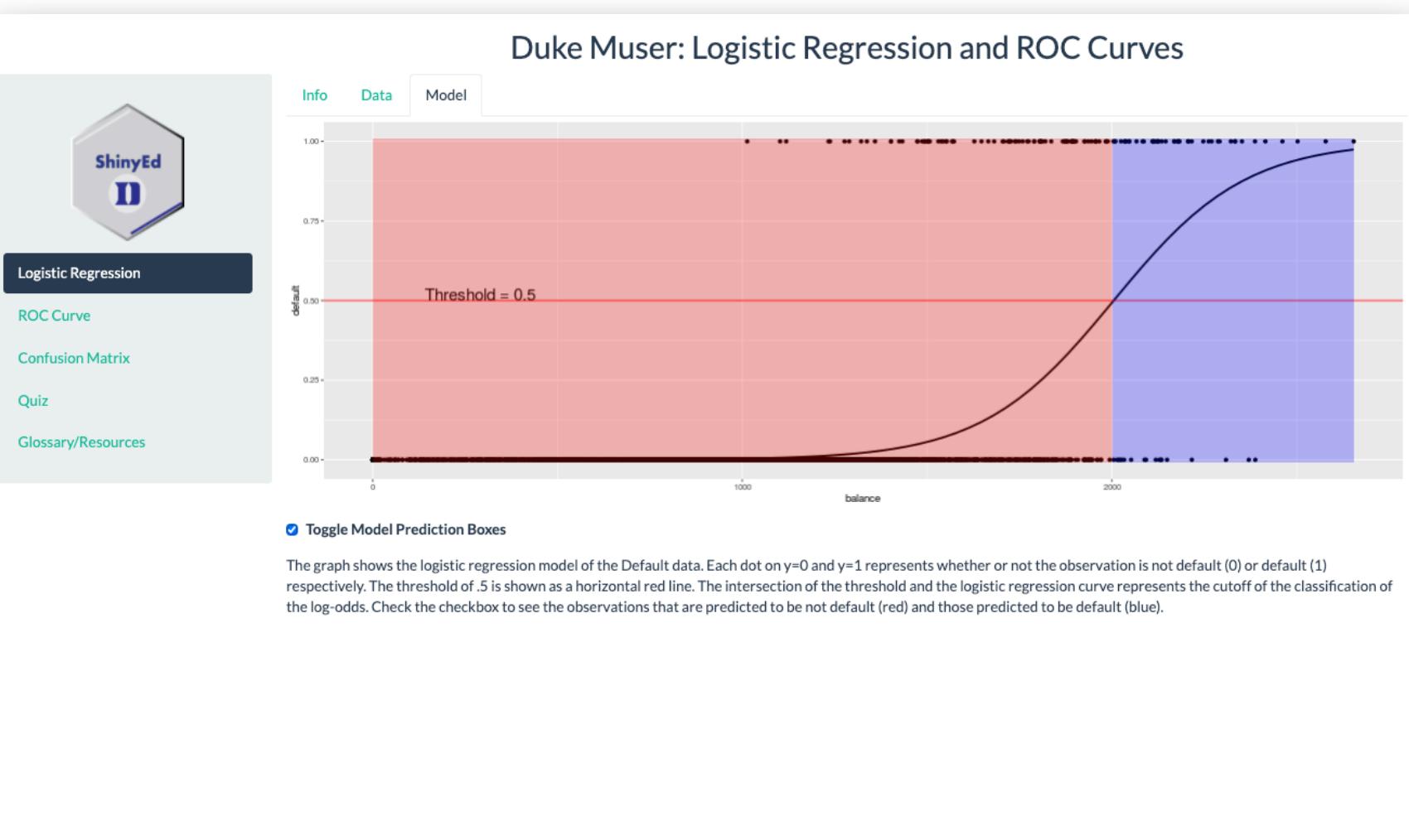
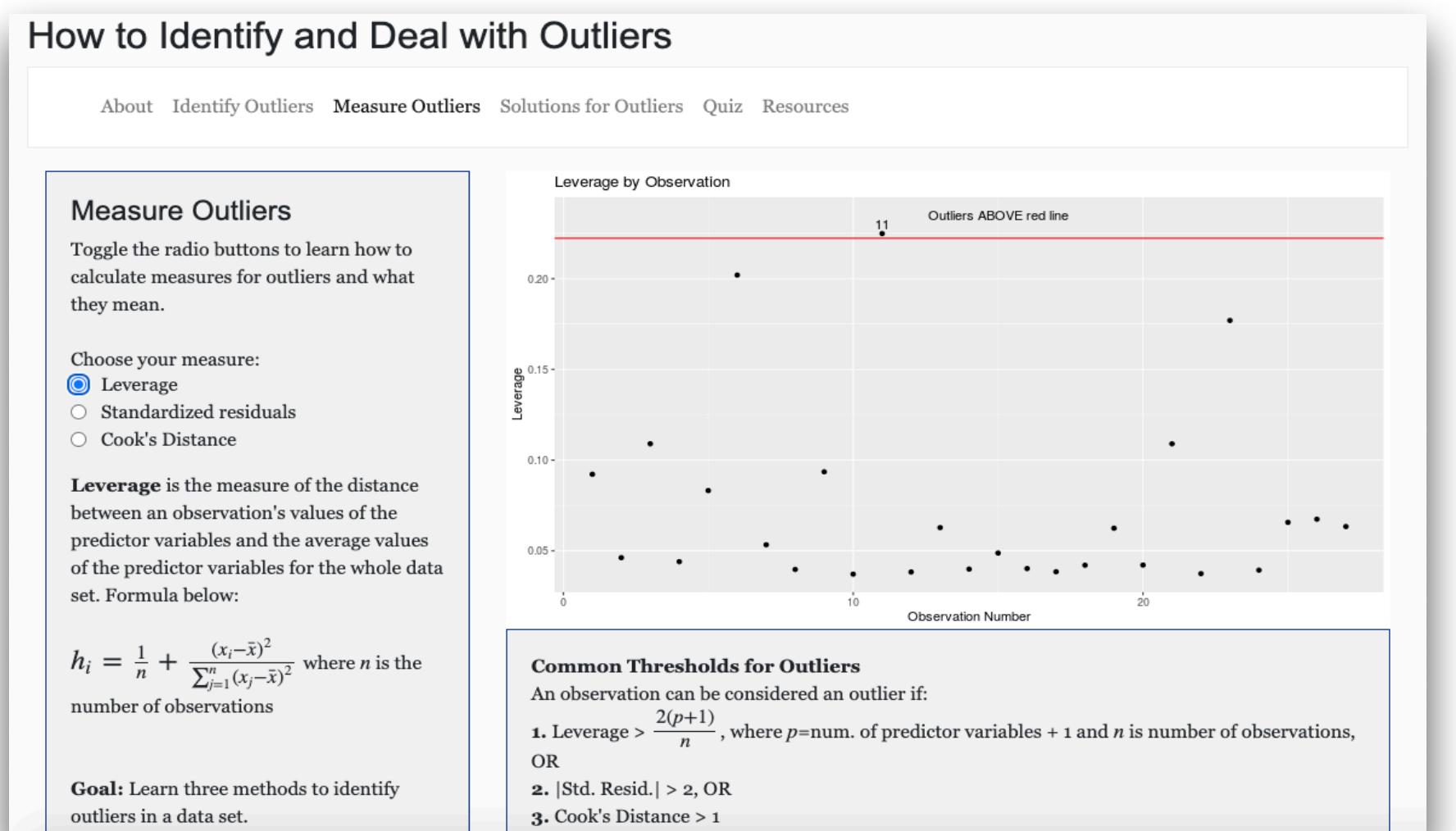
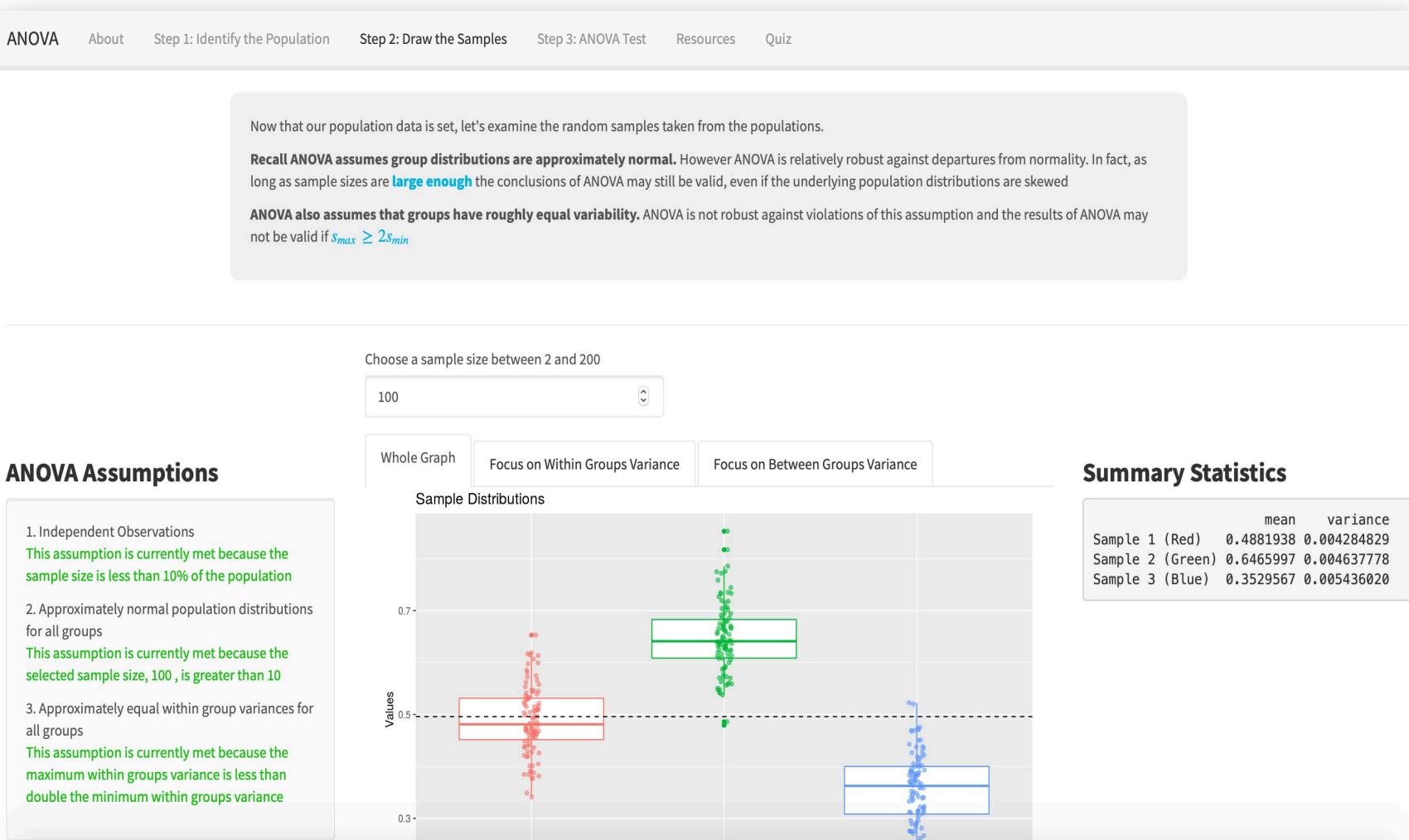
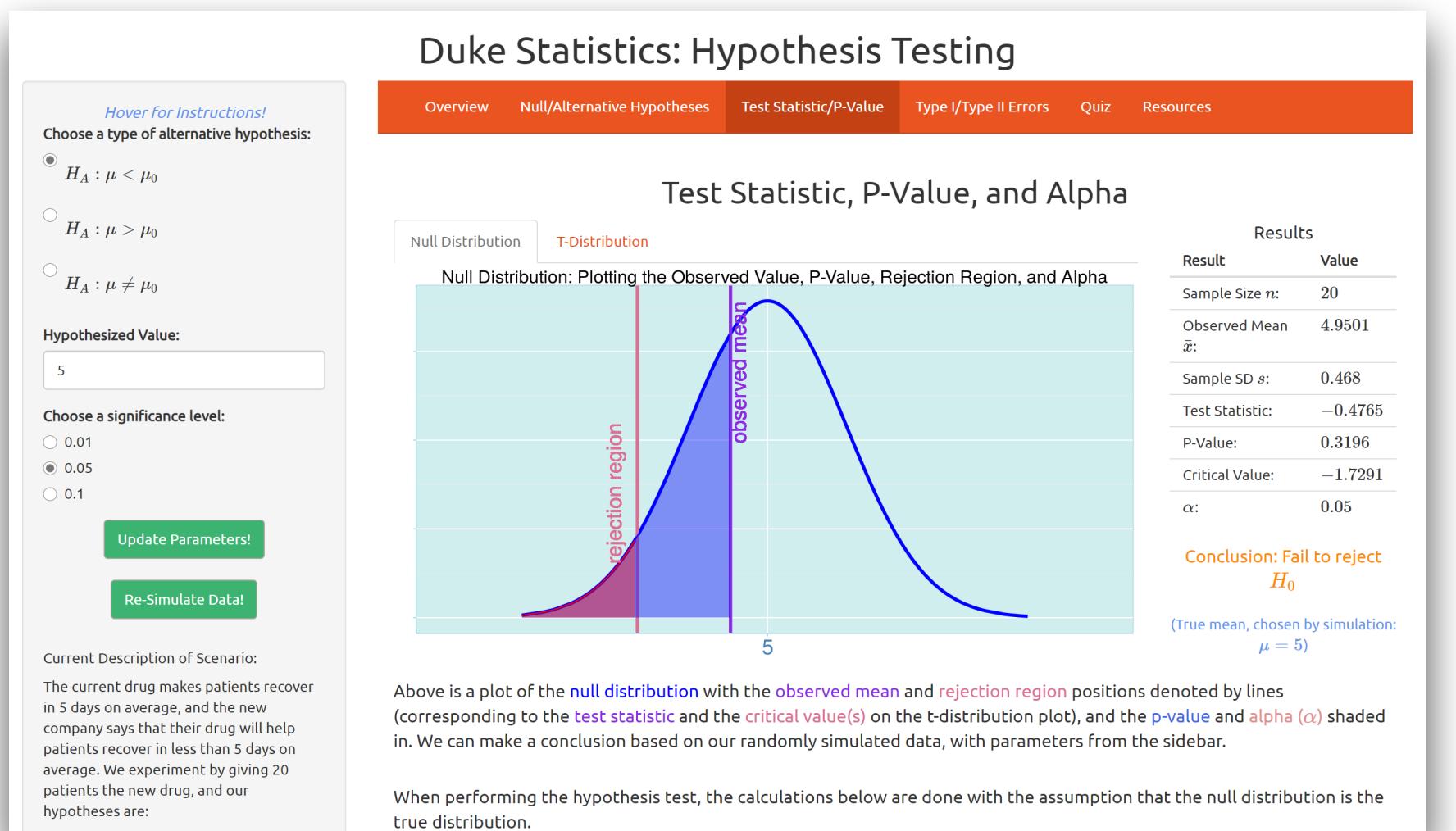
Meetings on Zoom

- Full team meeting 2x per week
- Meeting with project manager + students



Daily check-in

# The apps!



# What students learned

- ✓ Wrangling messy real-world data
- ✓ Ethical considerations of working with sensitive data
- ✓ Communicating statistical results to a general audience
- ✓ Collaborating with subject matter experts
- ✓ Reproducibility + version control + collaboration using R Markdown and GitHub
- ✓ How to work on a long-term, open-ended project

Motivation

NC Fines and  
Fees

Interactive  
Learning  
Tools

Challenges  
+  
Lessons  
Learned

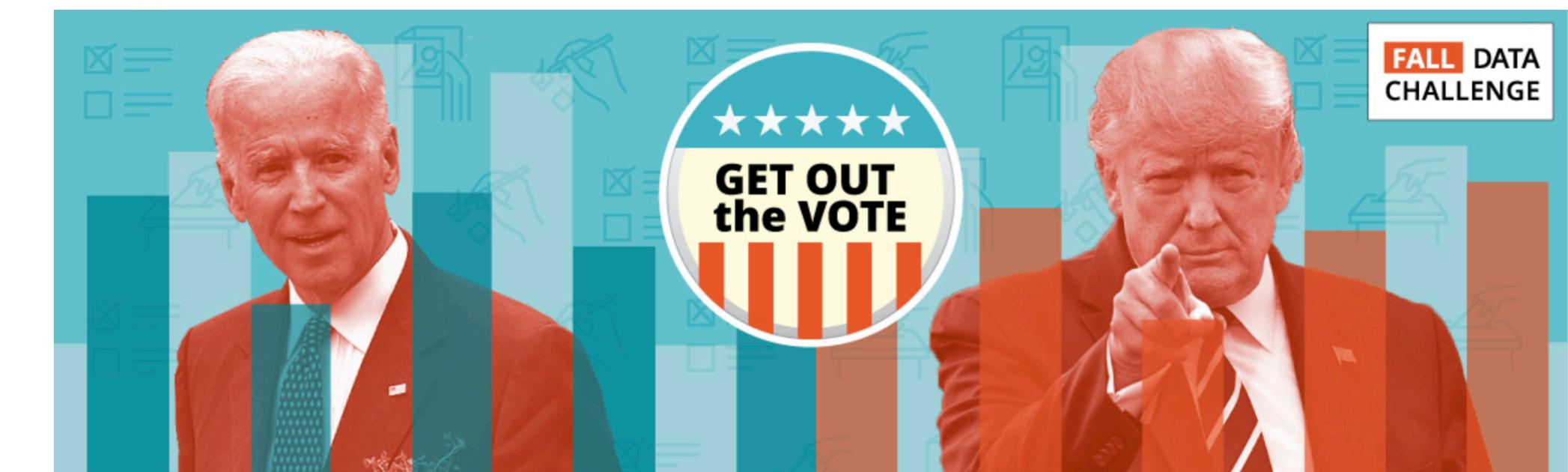
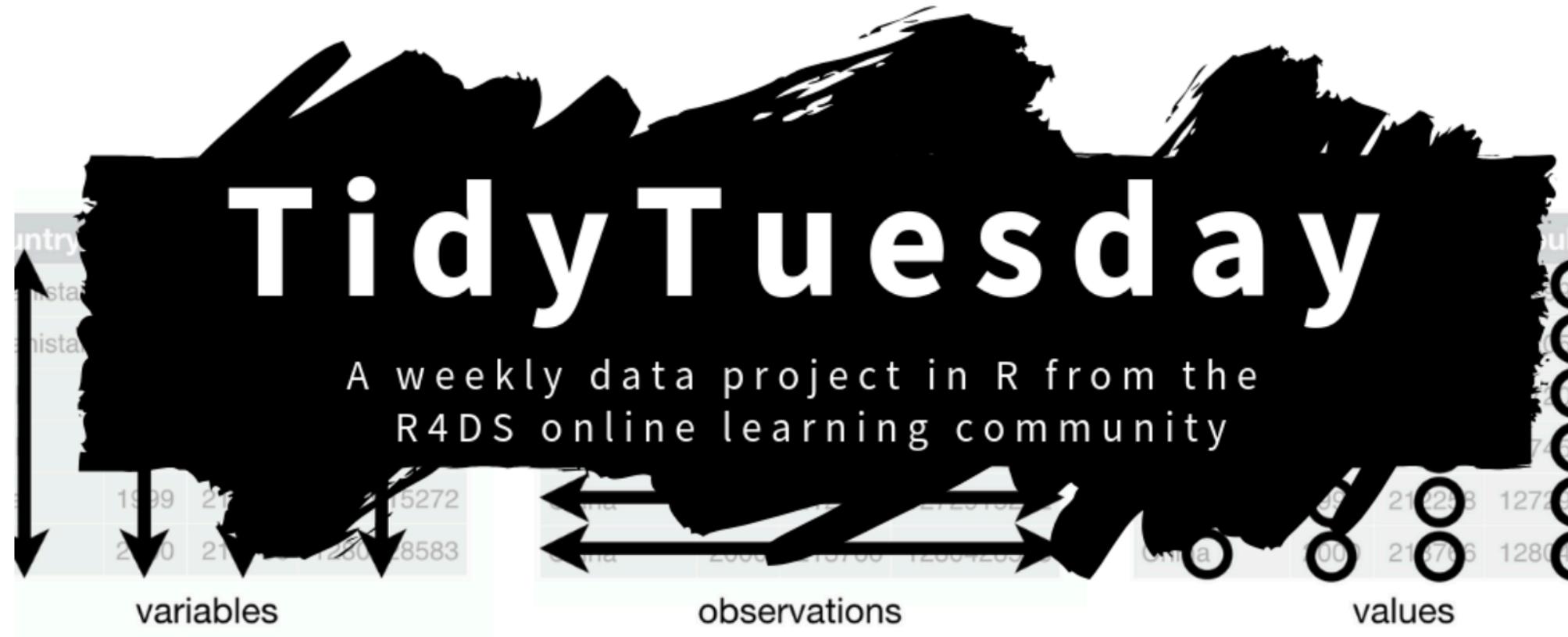
# Challenges

- ⚠ Time - especially getting started
- ⚠ Balancing exploration and focus
- ⚠ Building community in a remote environment
- ⚠ Funding

# Lessons learned

- 💡 Allow (more) time for preparation and skill development
- 💡 Have students work together (Nolan et al., 2020)
- 💡 Advertise widely to make opportunities more accessible
- 💡 Set clear goals with regular accountability
- 💡 Learn from others

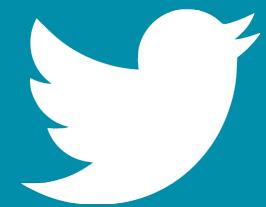
# Mentoring students in short-term projects



# Thank You!



maria.tackett@duke.edu



@MT\_statistics



[bit.ly/wsds2020-create](https://bit.ly/wsds2020-create)

# References

Joseph R. Nolan, Kelly S. McConville, Vittorio Addona, Nathan L. Tintle & Dennis K. Pearl (2020) *Mentoring Undergraduate Research in Statistics: Reaping the Benefits and Overcoming the Barriers*, Journal of Statistics Education, 28:2, 140-153.