

An Introductory R Course for Statistics Majors

Dr. Laura Taylor, Department of Mathematics & Statistics



Abstract:

This poster discusses a course, Statistical Computing for Simulation and Theory, developed at an undergraduate, liberal arts university to teach the R language. While primarily developed for statistics majors and minors, the course has a pre-requisite of a 200-level introductory statistics course and therefore is also populated by students interested in learning R. The course explores statistical theory (such as the Central Limit Theorem, impacts of assumption violations on analyses, and the power of a test) through the lens of student-coded simulations using R. The poster covers course content, pedagogical strategies, and assessments used in this course. One assessment highlighted in the poster is the final course project where students illustrate their statistical conceptual knowledge and ability to communicate statistics through a Shiny app and blog post. Changes between the first and second offerings of the course are also presented. This interactive poster invites viewers to share their ideas about what content this type of course should cover and how this course might differ with a focus on data science.

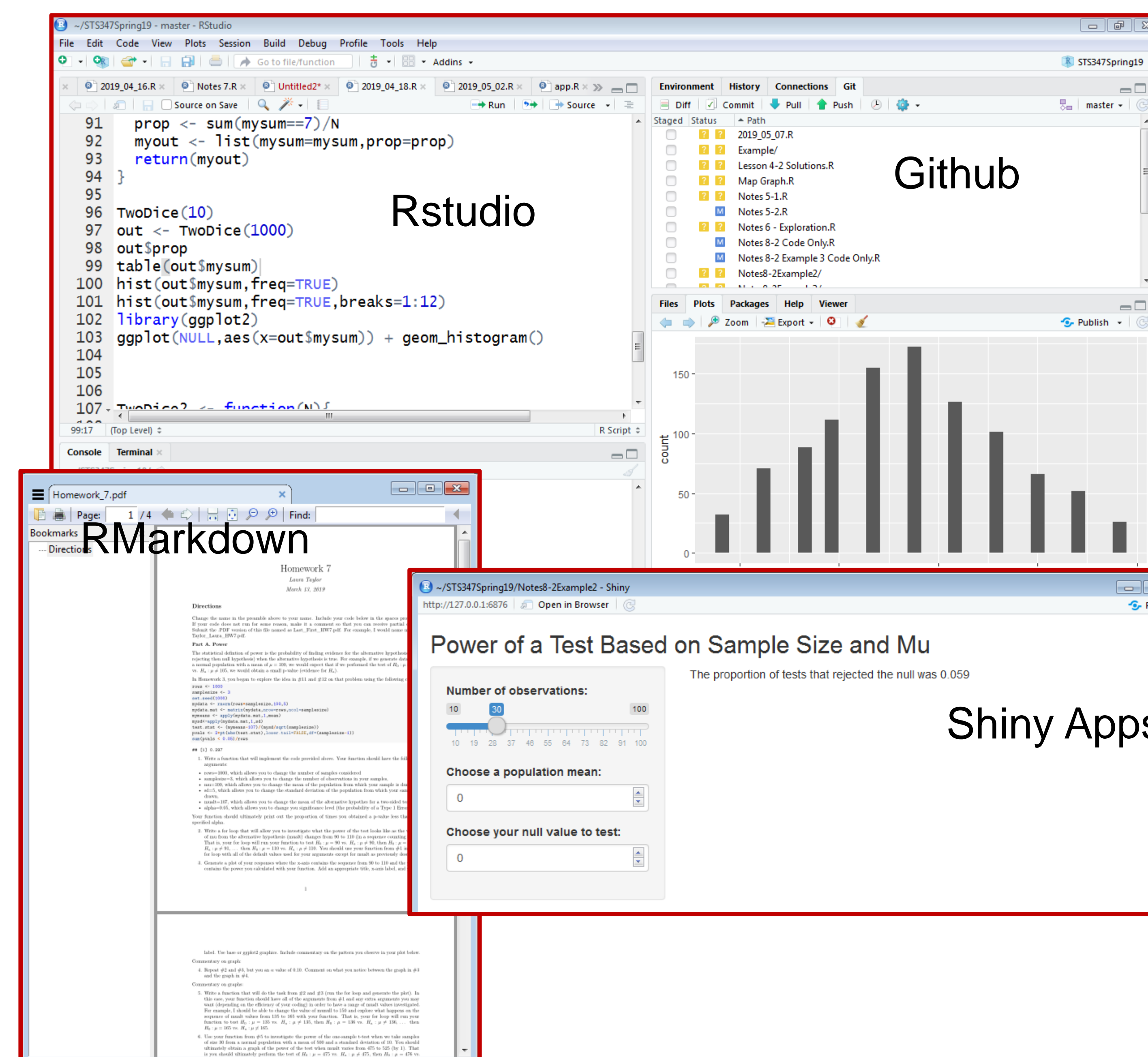
Course Description:

An intermediary course in **statistical computing and theory** using R software. This course introduces the software R with an emphasis on utilizing its powerful graphics and simulation capabilities. This course also presents R companion software, such as R Studio. This course focuses on **a non-calculus based approach to understanding theoretical statistical concepts through a simulation-based approach**. Students will gain knowledge and experience in writing both simple and more advanced simulations in R. This course has **two parallel goals of becoming fluent in R syntax and developing and understanding of statistical theory through simulation**.

Course Outcomes:

- Program in R proficiently.
- Import data, generate artificial data based on probabilistic models, and manage data.
- Generate appropriate and meaningful graphics and statistics.
- Implement correct syntax in order to perform standard statistical analysis techniques and identify errors in syntax.
- Create advanced functions and simulations in R to evaluate statistical theory.
- Know the basic terminology of probability theory and statistical theory.
- Generate examples of statistical theory using simulation, including the Central Limit Theorem and rules of probability theory.

Course Tools



Other Assessments

Project 1: 2018: SAS vs. R, 2019: R video tutorial (maps, R presentations, %>% , apply functions, webscraping, googlesheets)

Project 2: Build a Shiny app to explore a statistical topic (Central Limit Theorem, outliers and correlation/regression, level of confidence) and develop a blog explaining app.

Exams:

- 3 exams (2 midterms, 1 cumulative final)
- Midterms: in-class written (75-80%), in-class coding (10%), take home coding (10-15%)
- Final: in-class written (80%), in-class coding (20%)

Major Packages Used

- dplyr • ggplot2 • shiny • readr • readxl •

How does this course compare?

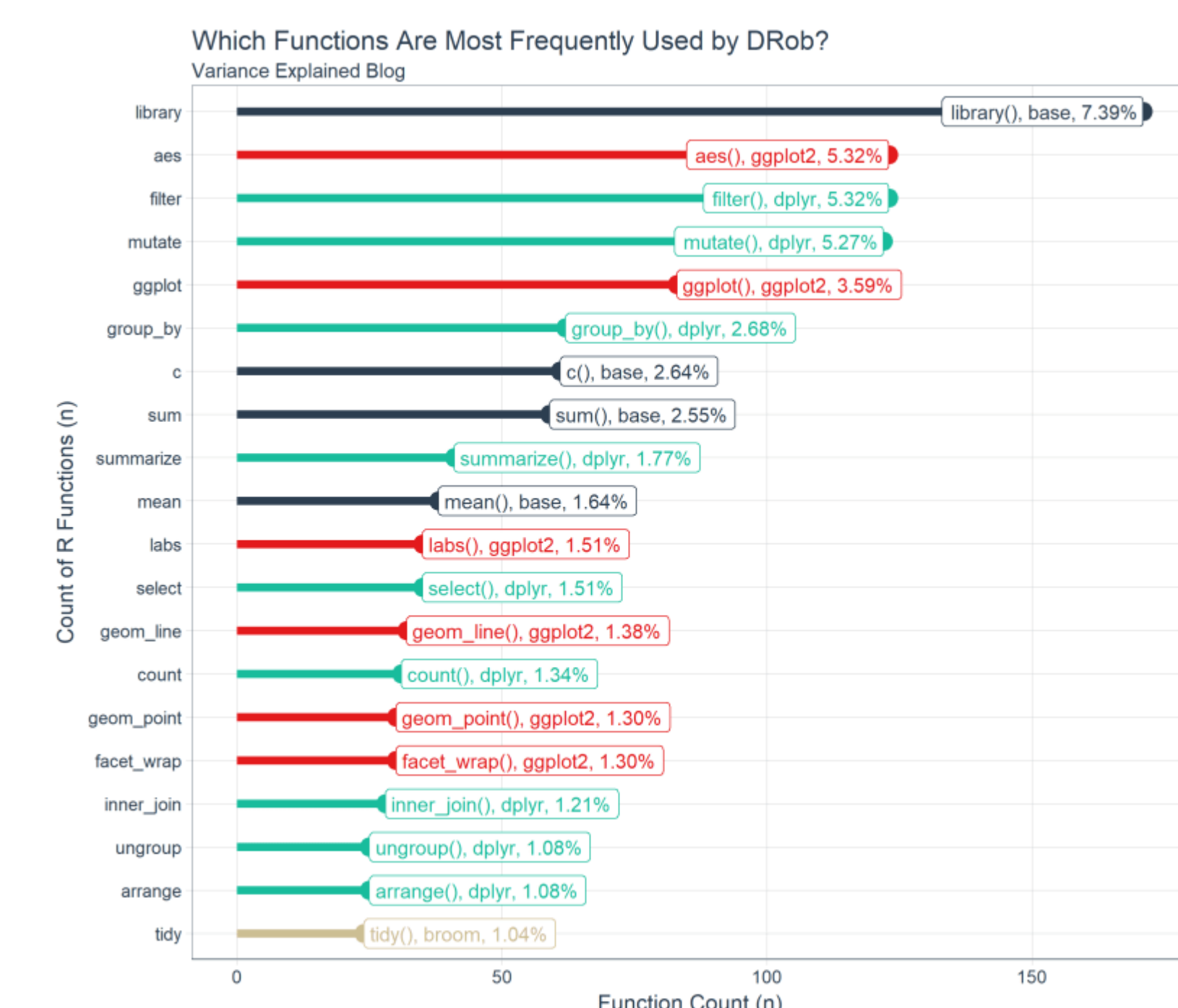
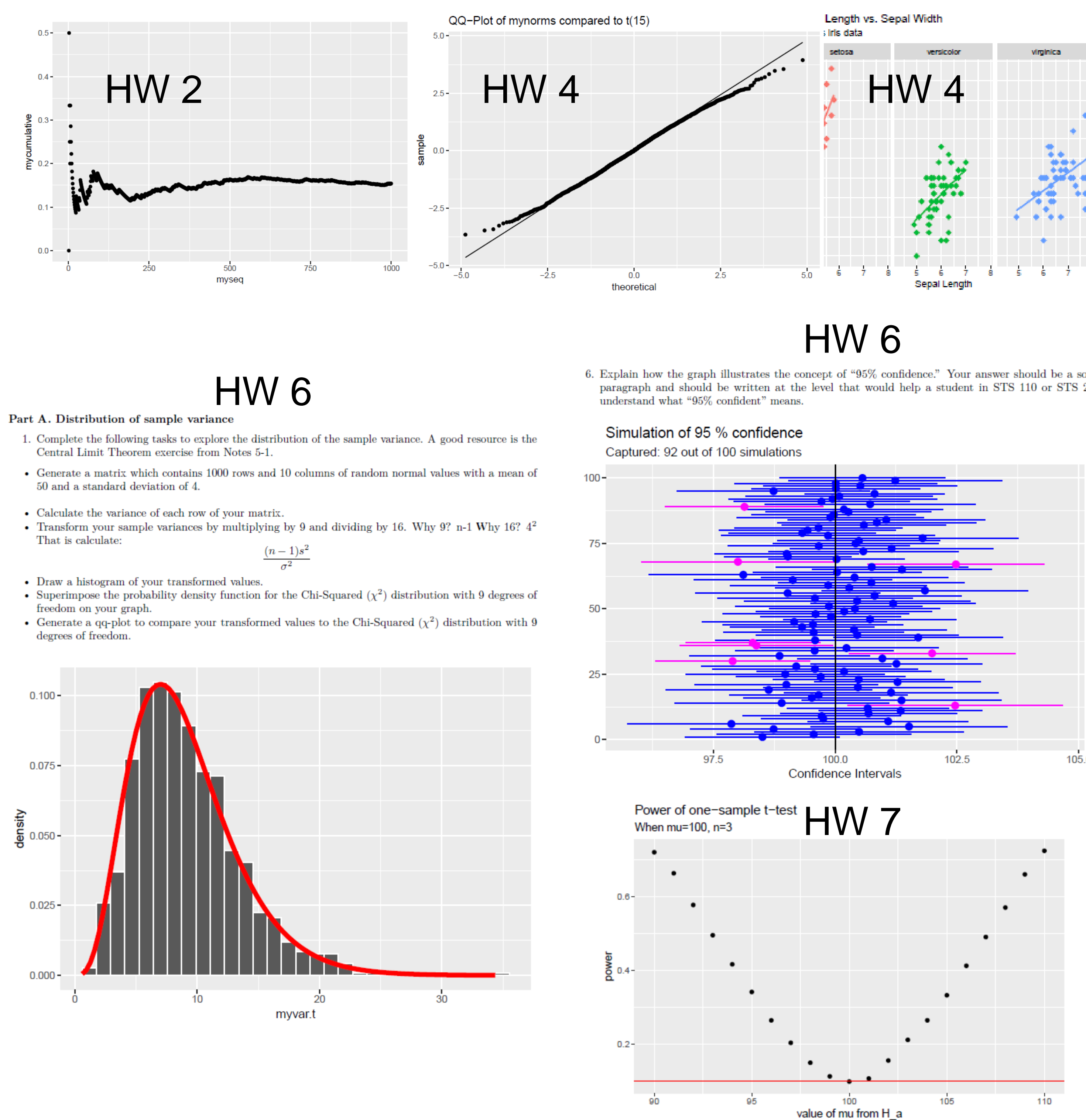


Image source: "How To Learn R, Part 1: Learn From A Master Data Scientist's Code" (March 2, 2018) <https://www.r-bloggers.com/how-to-learn-r-part-1-learn-from-a-master-data-scientists-code/>

Example Homework Tasks



Next Steps

- Use Rstudio Cloud
- Further integration of Github
- More data management, including packages like lubridate, rvest, tidy
- Use R Projects

What are your ideas – how can this course better prepare undergraduate statisticians to be included in the world of data science?

Go to <https://tinyurl.com/WSDSideas>

I'm happy to share course materials with you.
My email address is LTaylor18@elon.edu