

# Using GitHub and RStudio to Facilitate Authentic Learning Experiences in a Regression Analysis Course

Maria Tackett, Duke University

## Motivation

Students often take Regression Analysis at a time when they are deciding their major and career goals. Therefore, it is important to not only keep students engaged but to also help them develop the skills they need to be successful in research, internships, and future careers. According to the 2014 *American Statistical Association Curriculum Guidelines for Undergraduate Programs in Statistical Science*,

*“...to be prepared for statistics and data science careers, students need facility with professional statistical analysis software, the ability to access and manipulate data in various ways, and the ability to perform algorithmic problem-solving.”*

The activities described in this poster are designed to help students work towards these objectives.

## Course Description

**Audience:** Students who have taken an introductory statistics course. Designed primarily for statistics majors but taken by 70 students from a range of academic disciplines each semester.

**Content:** Analysis of multivariate datasets, with an emphasis on linear regression, analysis of variance, and logistic regression. Focus on data analysis over mathematical theory.

**Computing:** RStudio Cloud, Git/GitHub

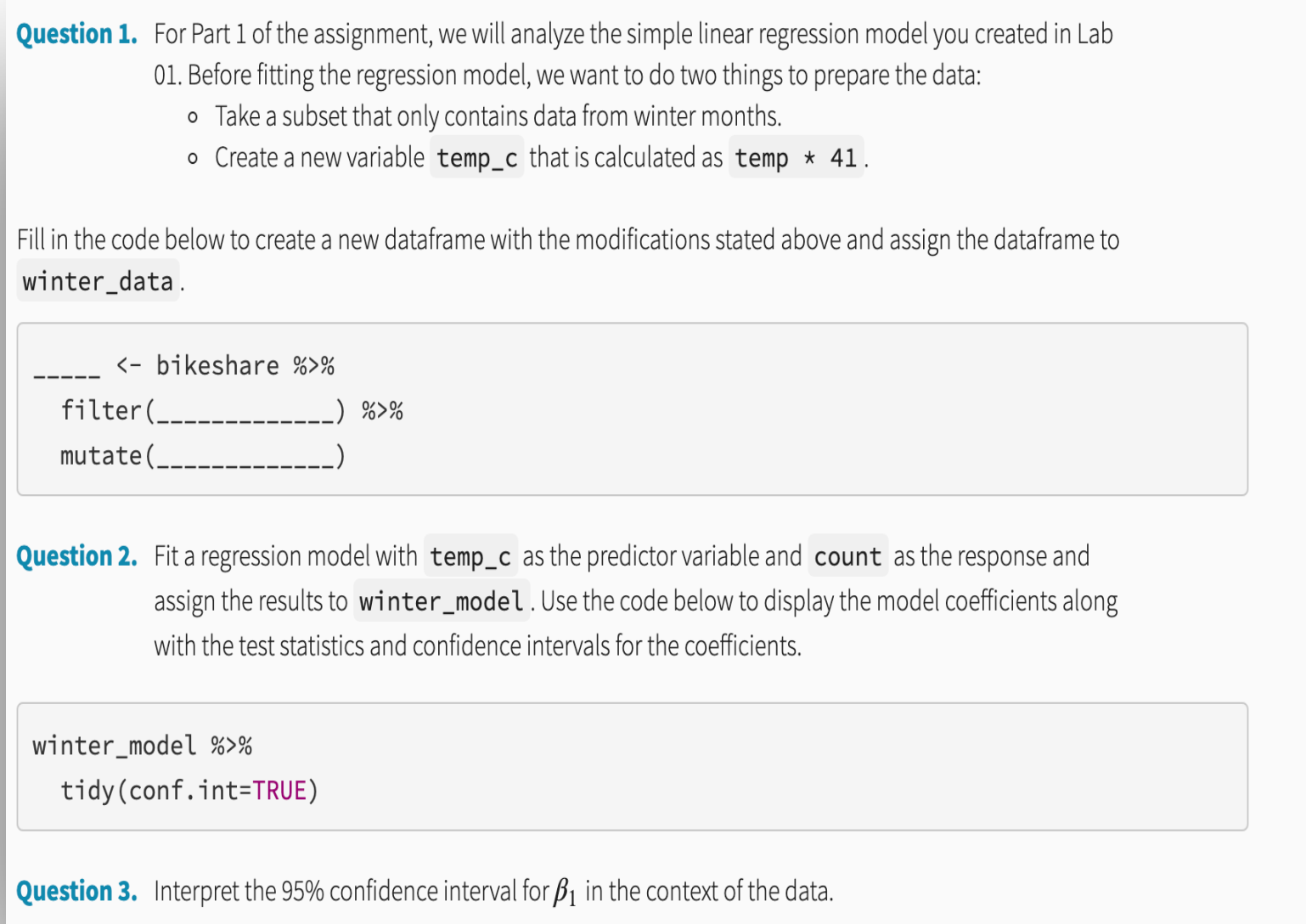
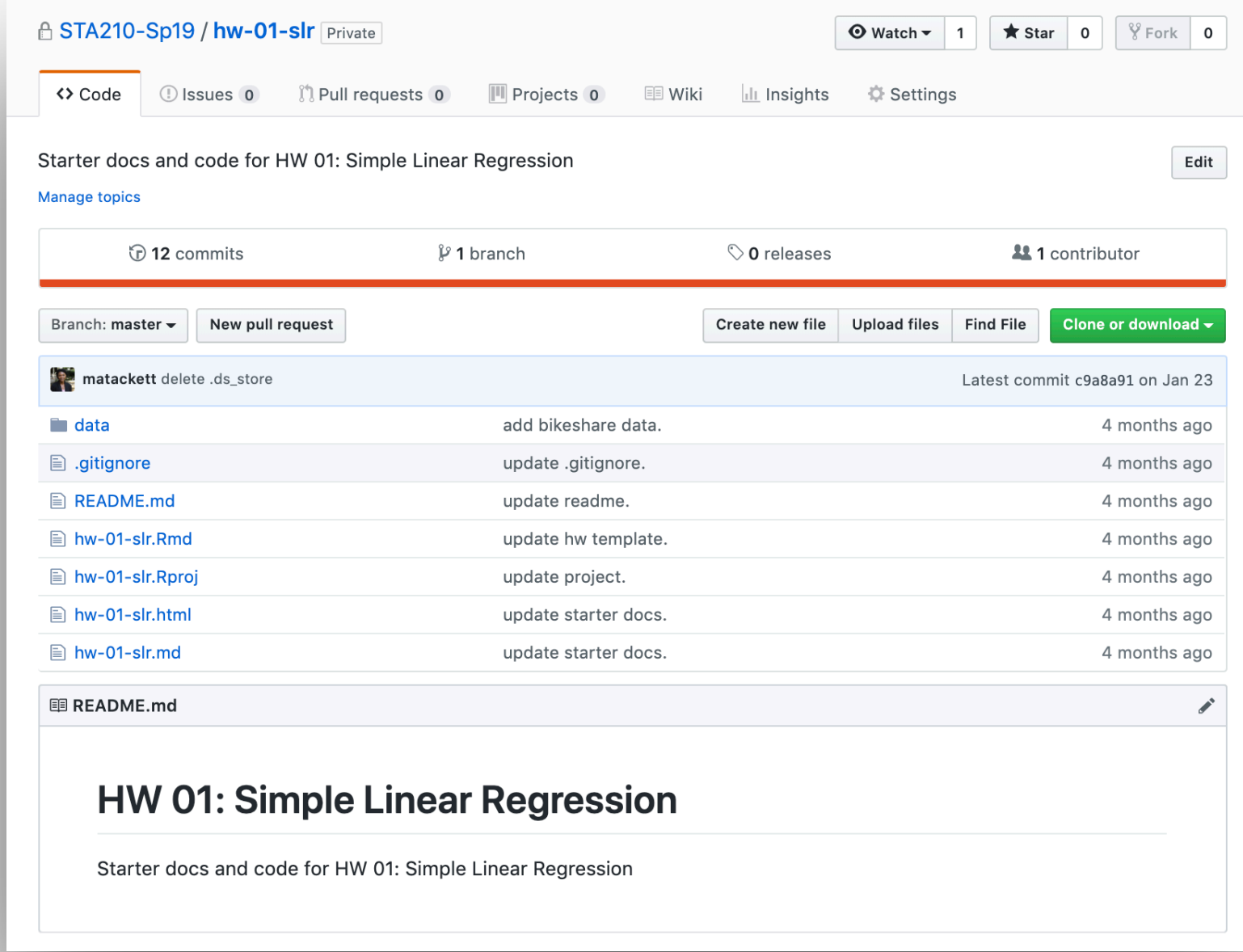
**Learning activities:**

Activity	Focus
In-class exercises	<ul style="list-style-type: none"><li>• definitions</li><li>• concepts</li></ul>
Lab assignments	<ul style="list-style-type: none"><li>• computing</li><li>• collaboration</li></ul>
Homework	<ul style="list-style-type: none"><li>• using computing + concepts to conduct a regression analysis</li></ul>

## Assignment Workflows

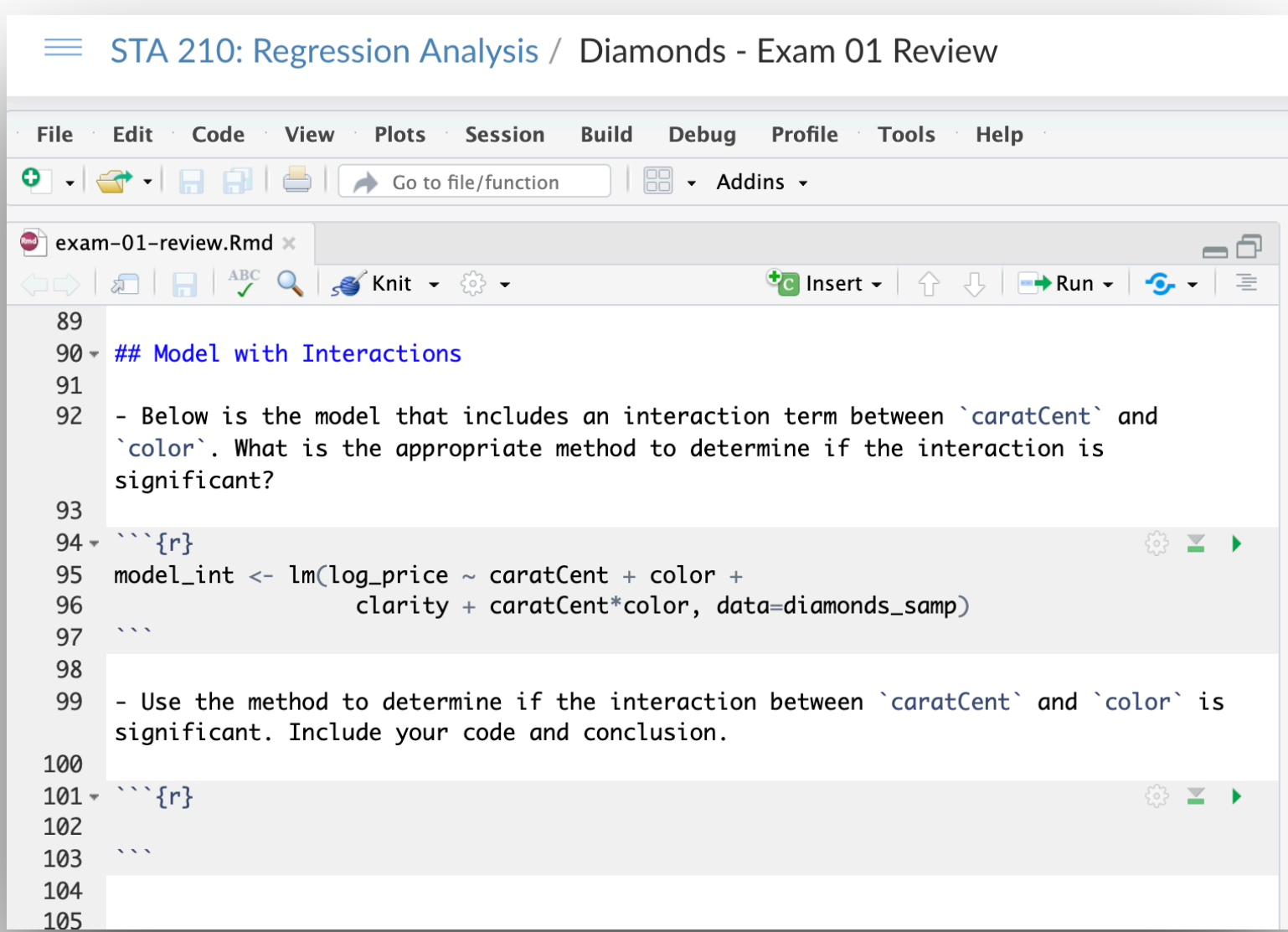
### Homework and Labs

1. Create a starter repo in GitHub containing R Markdown template and datasets.
2. Make a copy of the starter repo for each student (or team) using the `ghclass` package in R. By default the repos will be private.
3. Students clone repo into RStudio, write code and narrative in the R Markdown template, and knit to produce a Markdown document. They periodically commit and push all work to GitHub. The student submission is the most recent state of the repo prior to the deadline.
4. Write feedback to students as a “issue” in the GitHub repo and post final grades in Sakai or similar system.



### In-class Exercises

1. Create RStudio Cloud project with R Markdown template and data.
2. Students make copy of RStudio Cloud project. They write short lines of code and interpret the results using questions posted in the R Markdown template.
3. Display a student’s RStudio Cloud project on the classroom screen as the student (or group) shares their work with the class.



## Learning Outcomes

**Iterative workflow:** Students submit work by frequently pushing it to GitHub. This makes it easier for students to treat analysis as an iterative process where they can make changes, incorporate feedback, and include new ideas as they work on assignments.

**Real-world collaboration:** Git and GitHub are frequently used for collaboration in industry. By using it in class, students develop skills to better prepare them for future internships and careers.

**Preparation for higher-level courses:** By integrating computing into the course, student develop the skills required to be successful in higher-level courses that are computationally intensive.

## Dealing with Challenges

**Different computing backgrounds:** Begin semester with short in-class exercises to expose students to RStudio and GitHub early-on. Have students work in groups that include diverse computing experiences.

**Learning Curve for Git/GitHub:** Teach students only the basic Git functions (push, pull, commit) and primarily use Git through the Git pane in RStudio instead of command line.

**Setup for large classes:** Have students register for GitHub and RStudio cloud on the first day of class. Use `ghclass` R package to create assignment repos quickly.

## References

Çetinkaya-Rudel, M., & Rundel, C. (2017). Infrastructure and Tools for Teaching Computing Throughout the Statistical Curriculum. *The American Statistician*, 72, 58-65.

Course Management with `ghclass`. Retrieved from <https://rundel.github.io/ghclass/>

**Learn More!** [www.introregression.org](http://www.introregression.org)

**Questions or comments?**  
 @MT\_statistics  maria.tackett@duke.edu