STA107 Post-Course Survey Analysis

Evaluating Student Reflections on R and Statistical Learning

Sinan Ma Jaiditya Dev

2025-04-13

Table of contents

1 Introduction

This report presents an analysis of the anonymous post-course survey completed by students enrolled in **STA107: Introduction to Statistics** at the University of Toronto Mississauga. The survey was voluntary, open for multiple submissions, and contributed **2% to the course grade**. Its goal was to evaluate students' experiences with R-based activities, the integration

of normal distribution concepts, and overall satisfaction with the course components. This feedback will help refine future course offerings and improve the learning experience.

2 Survey Questions

This section outlines the structure of the survey and highlights key question types:

- **Quantitative items**: Likert-scale responses on course clarity, usefulness of R exercises, and understanding of statistical concepts.
- **Qualitative items**: Open-ended prompts for students to share what they found most/least helpful and suggestions for improvement.

The post-course survey comprised 17 open- and close-ended questions. Among these, 11 were qualitative in nature (`Q3`, `Q6`, `Q8`, `Q10`, `Q11`, `Q12`, `Q13`, `Q14`, `Q15`, `Q16`, `Q17`), focusing on free-text responses reflecting students' thoughts on the content, instruction, and overall experience.

3 Data Analysis

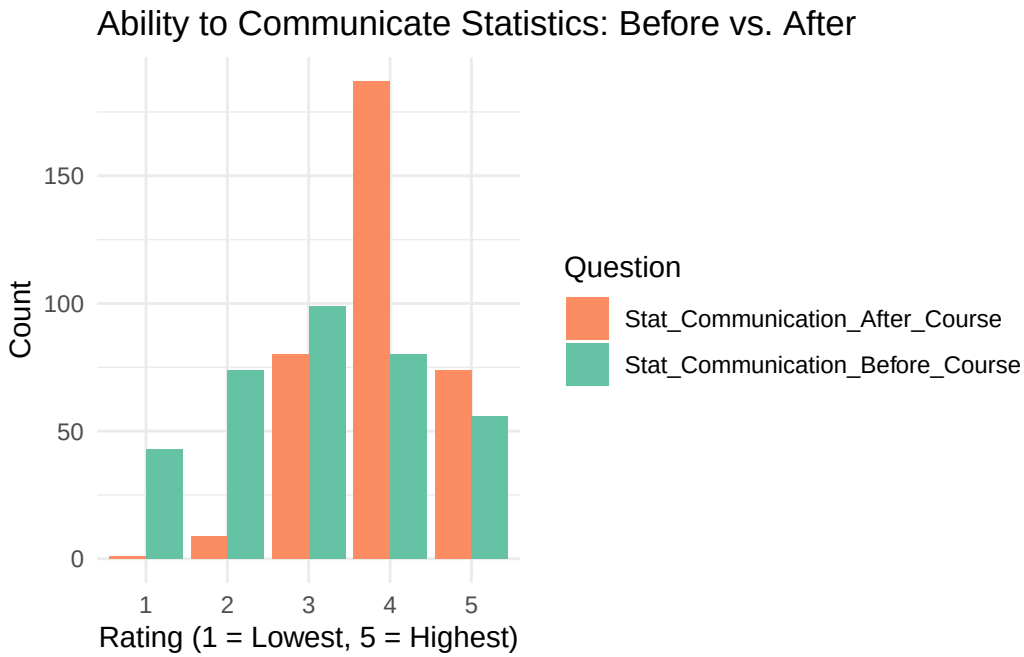
3.1 Methodology

We employed a **mixed-methods approach**:

- **Quantitative analysis**: Descriptive statistics and visualizations for Likert-scale questions.
- **Qualitative analysis**: Thematic coding of open-ended responses to identify common sentiments and suggestions.

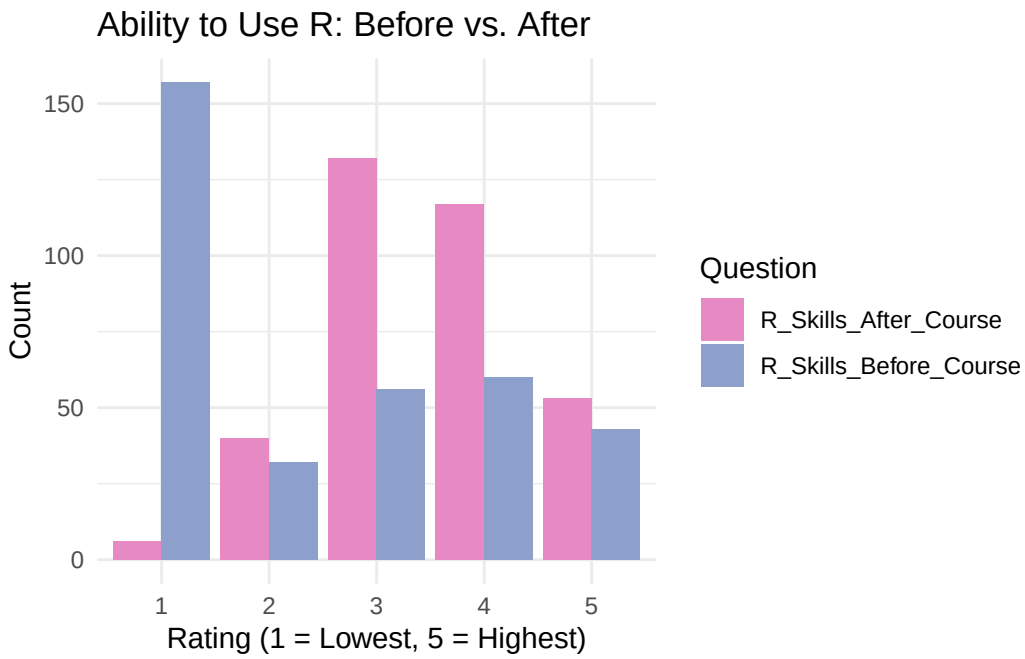
3.2 Quantitative Analysis

3.2.1 Communication Ability Before vs. After



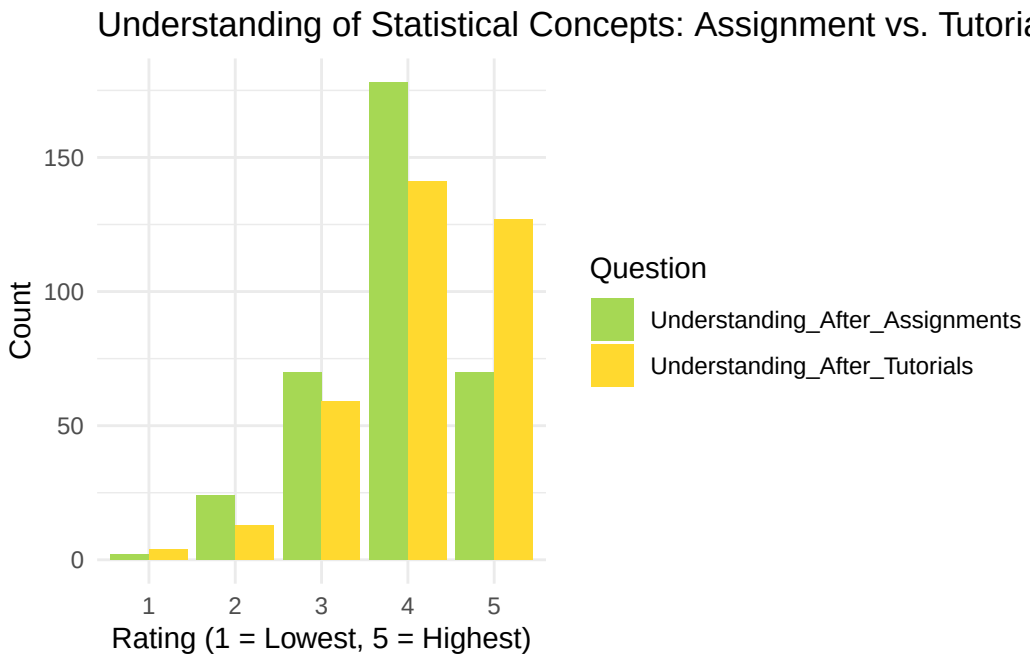
The histogram shows a clear improvement in students' perceived communication ability. More students rated themselves at level 4 or 5 after the course compared to before, indicating that the course helped them gain confidence in articulating statistical ideas.

3.2.2 R Ability Before vs. After



Students entered the course with relatively low self-reported R skills (mostly 1s and 2s), but after completing the course, a majority rated themselves at 3 or higher. This shift demonstrates the course's success in introducing R programming effectively.

3.2.3 Understanding After Assignments vs. Tutorials



Students found both tutorials and assignments useful for understanding statistical concepts, with a slight preference for tutorials. The higher number of level 4 and 5 responses for tutorials suggests that real-time, interactive support may have been especially beneficial.

3.3 Qualitative Analysis

4 Results

4.1 Quantitative Results

The survey included six Likert-scale questions aimed at capturing students' self-perceived development in statistical understanding and communication, both before and after the course. The questions also addressed students' ability to use R software and their comprehension of statistical concepts through tutorials and weekly assignments.

Key findings include:

- **Improved Communication Skills**: Students reported a notable improvement in their ability to communicate statistics, with the average post-course rating (M = 3.92) higher than the pre-course rating (M = 3.09).

- **Increased R Proficiency**: The mean rating for R software usage rose from 2.43 before the course to 3.49 after completing it, indicating significant growth in computational skills.
- **Conceptual Understanding**: Understanding of statistical concepts improved through both tutorials (M = 4.09) and weekly assignments (M = 3.84), with tutorials being rated slightly more helpful on average.

The graphs above illustrate the distribution of responses across the three themes: communication, R proficiency, and statistical understanding.

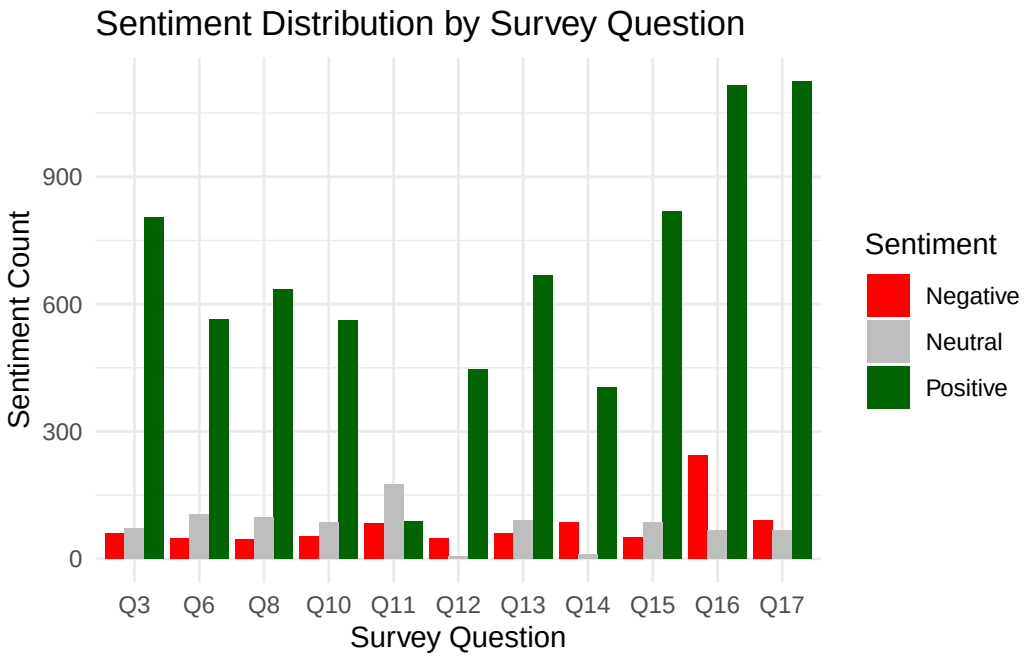
4.2 Qualitative Results

The sentiment analysis revealed a clear trend of positive student engagement across most open-ended responses. When asked about what aspects of the course were most helpful, students consistently expressed appreciation for the structured content delivery, use of R, and clarity in statistical explanations. Similarly, reflections on the usefulness of the normal model and applied examples elicited strong positive sentiment, highlighting the effectiveness of concept integration throughout the term.

Feedback on the clarity of teaching and accessibility of course material was more mixed. While many students praised the instructor's effort and teaching style, others noted inconsistencies in understanding due to the pace of delivery or difficulty following complex R assignments.

The most critical responses were seen in prompts asking what aspects of the course were least helpful and what improvements could be made. Here, students frequently cited the fast pace of certain topics, confusion around R syntax, and a need for more real-world or scaffolded examples. Words like "confusing," "fast," and "examples" frequently appeared in these responses, pointing to the need for clearer instructional supports.

Figure 1: Sentiment Distribution by Survey Question



5 Conclusion

The STA107 post-course survey provides valuable insights into how students perceive the course structure and its technical components. Findings highlight both effective practices and opportunities for enhancement. Continued evaluation and iteration will ensure the course remains responsive to student needs and pedagogical best practices.

From a qualitative perspective, student feedback revealed strong appreciation for several key elements of the course—especially the hands-on R activities, the integration of real-world statistical examples, and the clarity of normal model explanations. These components were described as engaging and instructive, affirming their value in helping students apply theoretical knowledge in practical settings.

However, when students were asked about areas of improvement, many raised concerns regarding the speed at which complex topics were introduced, as well as the difficulty of interpreting R-based assignments without more structured guidance. Themes of pacing, clarity, and the need for more scaffolded instruction were particularly prominent. Addressing these concerns by slowing down the delivery of technical material, providing more detailed examples, and incorporating mid-semester feedback opportunities could significantly enhance the learning experience in future offerings of STA107.

6 References

- Jockers, M. (2017). *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text* (R package v1.0.4). CRAN Documentation
- Jockers, M. (2020). *Introduction to the Syuzhet Package*. CRAN Vignette
- Kim, H. (2022). Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons. *Digital Humanities Quarterly, 16*(2). Article