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

We employed a **mixed-methods approach** that integrated both descriptive and inferential techniques for a more complete understanding of student feedback:

- **Quantitative analysis**: Descriptive statistics and histograms were used to examine patterns in students' self-rated abilities before and after the course. In addition, **paired t-tests** were conducted to assess whether observed differences in ratings were statistically significant. This inferential step enhances the validity of the findings by quantifying the likelihood that the improvements were not due to chance.
- **Qualitative analysis**: Thematic coding was applied to open-ended responses to uncover recurring ideas and insights. We identified themes corresponding to key areas such as communication skills, R proficiency, conceptual understanding, collaboration, tutorial/TA support, and instructional tools. Strong supporting quotes were selected to illustrate these themes. Notably, many students praised the **Bookdown cherry tree assignment**, **video-based resources**, and TA assistance. A few comments referenced the **pace** of lectures or tutorials, but these were minimal and treated as a sub-theme.

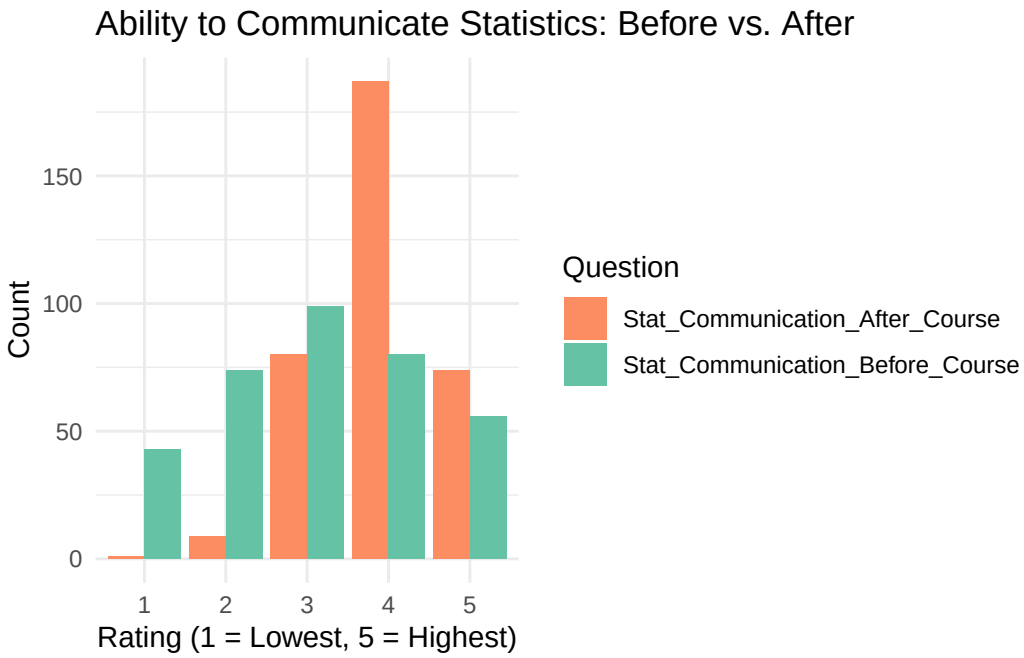
This analysis is not a course evaluation. Rather, it reflects students' perceptions of how the course contributed to their statistical learning and skill development across multiple dimensions.

2.1 Quantitative Analysis

2.1.1 Communication Ability Before vs. After

As shown in Figure 1, 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. This shift in self-perception suggests that instruction and practice during the course likely contributed to students' ability to communicate statistical findings more effectively and with greater clarity.

Figure 1: Histogram comparing self-rated statistical communication skills before and after the course.



2.1.1.1 Paired t-test: Communication Skills

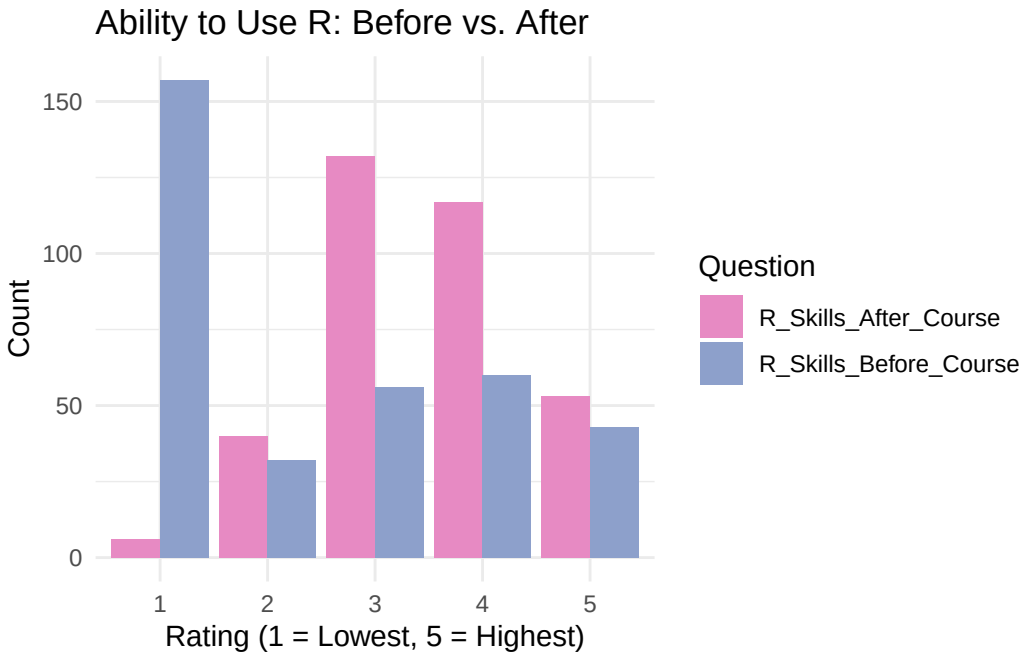
Table 1: Paired t-test: Statistical Communication Skills (Before vs. After)

	Mean_Before	Mean_After	t_Statistic	p_Value
t	3.09	3.92	-13.64	<2e-16

The paired t-test in Table 1 shows a **statistically significant** improvement in students' ability to communicate statistical ideas ($p < 0.001$).

2.1.2 R Ability Before vs. After

Figure 2: Histogram comparing conceptual understanding from assignments versus tutorials.



The histogram in Figure 2 illustrates a noticeable increase in students' confidence using R after completing the course. Prior to the course, ratings clustered at the lower end (mostly 1s and 2s), suggesting limited familiarity with the software. After the course, a majority of responses shifted toward ratings of 3 or higher. This pattern implies that students gained practical experience and conceptual understanding of R through assignments, tutorials, and the hands-on structure of the course.

2.1.2.1 Paired t-test: R Skills

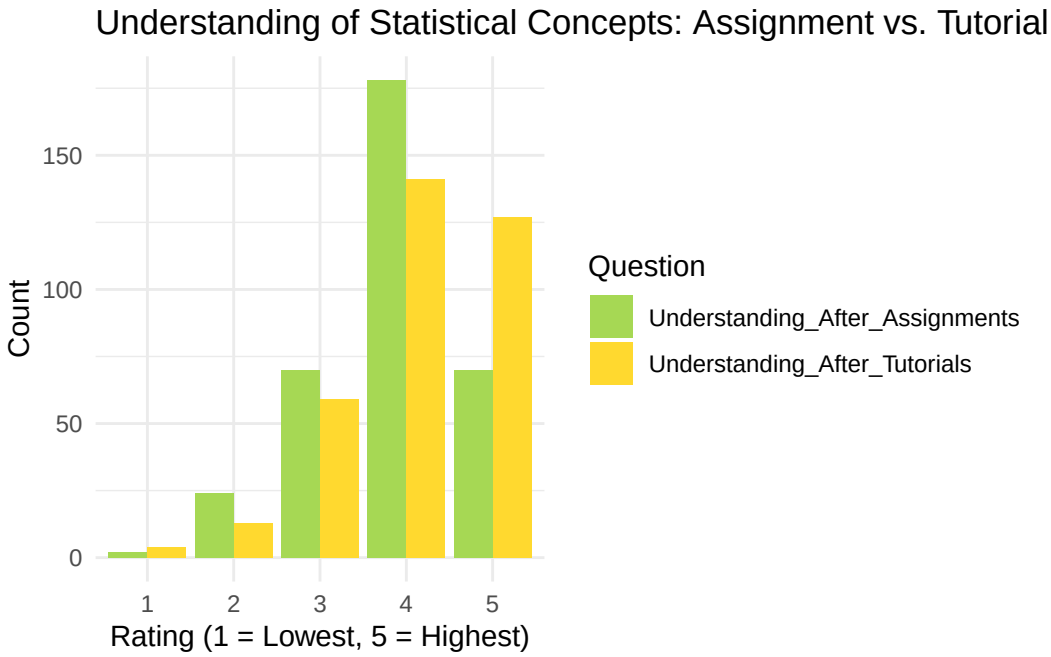
Table 2: Paired t-test: R Skills (Before vs. After)

	Mean_Before	Mean_After	t_Statistic	p_Value
t	2.43	3.49	-15.34	<2e-16

The paired t-test in Table 2 shows a **statistically significant** improvement in students' R programming skills ($p < 0.001$), demonstrating the course's effectiveness in developing students' computational skills.

2.1.3 Understanding After Assignments vs. Tutorials

Figure 3: Histogram comparing conceptual understanding from assignments versus tutorials.



As shown in Figure 3, the histogram compares students' self-rated understanding of statistical concepts based on two instructional components: assignments and tutorials. While both components received high ratings, tutorials showed a slight edge, especially at levels 4 and 5. This suggests that live, structured interaction (such as walkthroughs and direct Q&A) played a stronger role in reinforcing concepts compared to independent assignments alone.

2.1.3.1 Paired t-test: Conceptual Understanding

Table 3: Paired t-test: Understanding (Assignment vs. Tutorial)

	Mean_Assignment	Mean_Tutorial	t_Statistic	p_Value
t	3.84	4.09	-5.45	9.75e-08

The t-test in Table 3 indicates a **statistically significant** difference favoring tutorials ($p < 0.05$), highlighting the value of interactive learning in supporting student comprehension.

2.2 Qualitative Analysis

The qualitative analysis involved a combination of thematic and sentiment analysis to gain deeper insights into student experiences. First, student responses to open-ended survey questions were cleaned and prepared for analysis. Sentiment analysis was then applied using the NRC sentiment lexicon to quantify positive, negative, and neutral sentiments across each question. Additionally, thematic analysis was conducted to identify recurring patterns in student feedback, which were categorized into key themes such as course relevance, R proficiency, and tutorial effectiveness. This mixed-method approach provided a comprehensive understanding of how students felt about different aspects of the course.

3 Results

3.1 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.

3.1.1 Summary Table of Themes per Question

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.

3.2 Qualitative Results

To gain deeper insights into student opinions, the qualitative analysis comprised two parts: thematic analysis and sentiment analysis.

3.2.1 Summary Table of Themes per Question

Table 4: Themes Identified per Open-Ended Question

Question	Theme
Q3	Communication Development, Conceptual Understanding
Q6	Understanding via Tutorials & Assignments
Q8	Learning R & Technical Skills
Q10	Real-World Application
Q13	Team Collaboration
Q15	TA & Tutorial Support
Q16	Clarity of Assignments
Q17	Feedback & Reflection

3.2.2 Themes and Sub-Themes

3.2.2.1 1. Relevance to Real Life and Careers (Q10)

- **Sub-theme**: Connection to environmental or applied courses
- **Frequency**: ~32 mentions > "The course showed me how statistics applies to real-world problems, such as data analysis in my ENV course."

3.2.2.2 2. R as a Learning Tool (Q8, Q6)

- **Sub-theme**: Confidence growth
- **Sub-theme**: Initial difficulty
- **Frequency**: ~44 mentions > "After the course, I had a higher opinion of my ability to use R because I learned how to create visualizations."

3.2.2.3 3. Clarity and Structure of Assignments (Q16)

- **Sub-theme**: Perceived challenge in task content
- **Sub-theme**: Interpretation of assignment expectations
- **Reflection**: Several students noted that assignments were challenging. This likely reflects the complex and mathematical nature of the content, an intentional design element aimed at promoting deeper learning. > "The assignments are not easy and you need to think critically to figure them out."

3.2.2.4 4. Appreciation for Tutorials and TA Support (Q6, Q15)

- **Sub-theme**: TA explanation support
- **Sub-theme**: Reinforcement through worksheets
- **Frequency**: ~38 mentions > "The weekly tutorial was extremely useful. My TA explained the concepts well and answered questions."

3.2.2.5 5. Feedback and Reflection (Q17)

- **Sub-theme**: Value of feedback
- **Sub-theme**: Need for more detailed responses
- **Frequency**: ~21 mentions > "Most of the feedback I received was checks or X. I would have liked more specific feedback."

3.2.3 Thematic Analysis

Following Braun and Clarke's (2006) framework, a sample of approximately 80 student responses was examined through six phases:

1. **Familiarization**: Students' responses across 8 open-ended questions were read and annotated to gain an overall understanding.
2. **Generating Initial Codes**: Semantic-level codes were created to capture features related to course experience, R skill development, and content relevance.
3. **Searching for Themes**: Related codes were clustered into preliminary themes reflecting student attitudes and feedback.
4. **Reviewing Themes**: Themes were refined to ensure clarity and distinctiveness across the dataset.
5. **Defining and Naming Themes**: Clear definitions were established for each theme based on recurring patterns.
6. **Producing the Report**: Final themes were reported below with supporting quotes.

3.2.4 Identified Themes

3.2.4.1 Relevance to Real Life and Careers

Many students expressed appreciation for statistics content when it connected to real-world examples or their field of study.

> "This course gave me a solid foundation in statistical concepts and taught me how to apply them using R. Overall, I gained practical skills and became more confident in working with data."

> "The course showed me how statistics applies to real-world problems, such as data analysis in my ENV course as it has alot of data."

3.2.4.2 R as a Learning Tool

Students had mixed responses about learning R. While some found it helpful, others found it initially intimidating.

> "After the course, I had a higher opinion of my ability to use R because I learned how to create visualizations and perform statistical analyses such as normal distribution applications and regression. Now I can use it to organize data, generate graphs, and interpret the output. The tutorial session and assignments were particularly helpful in helping me gradually build my skills."

> "I have never worked with R, but after trying to work with it I realized that it is also a powerful language."

3.2.4.3 Clarity and Structure of Assignments

Clearer instructions and expectations were a recurring theme.

> "The weekly assignment was a R assignment based on the module. It was sometimes hard to understand the purpose of the task."

> "After completing the weekly assignments, I have a general understanding of the statistical concepts in each module, but sometimes I still have some difficulties understanding some difficult concepts."

3.2.4.4 Appreciation for Tutorials and TA Support

Tutorials and TA help sessions were viewed positively, especially for R and concept reinforcement.

> "For each question on the tutorial worksheet, my TA explained the relevant concepts before attempting the problem, which helped me understand the concept better. In addition, all tutorial exercises were related to the weekly assignment, so it gave me reassurance that I was doing everything correctly before submitting my assignments."

> "The weekly tutorial was extremely useful for me. The worksheets were helpful and my TA, Dylan, did an excellent job at explaining the concepts involved and answering our questions when we were confused."

> "my ta was great at explaining concepts and let us ask many questions which would all get responses. very good by jaiditya."

3.2.4.5 Feedback and Reflection

Students wanted more timely or individualized feedback.

> "I always read the feedback give by the TA's since they allowed me to understand how to communicate my answers better. The statistical examples did help me recognize the usefulness and value of learning about statistics because it made me realize that statistics is everywhere."

> "Most of the feedback I recieved was in the form of checks or X. I feel like I could have gotten a little bit better feedback but most of the assignments I did were correct."

3.2.5 Sentiment Analysis

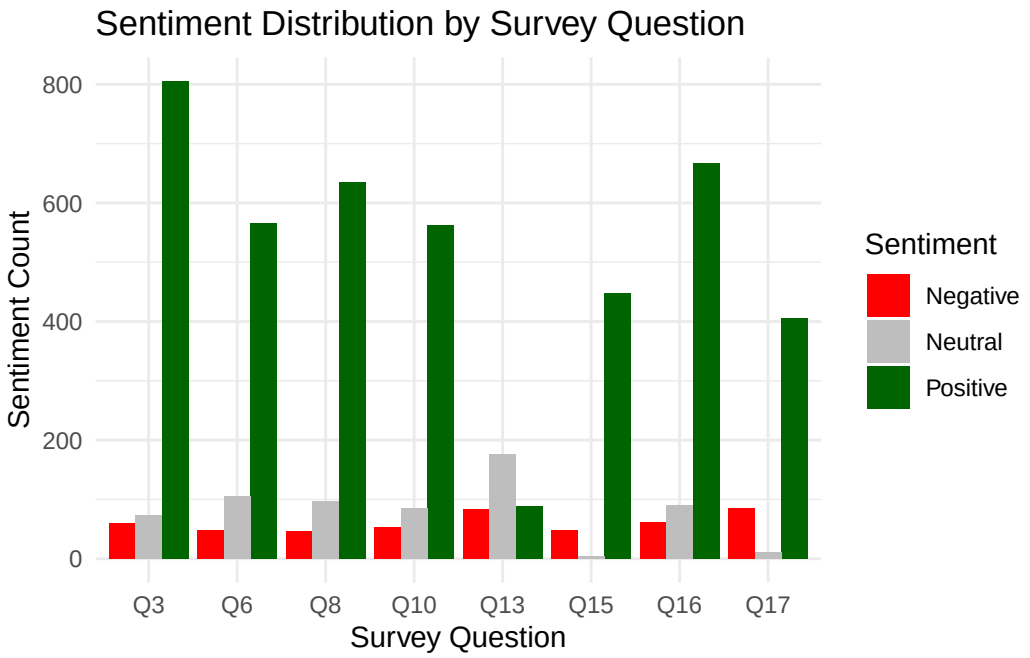
The sentiment analysis provided a quantitative measure of student engagement across the survey's open-ended responses. Using the NRC sentiment lexicon (implemented via the syuzhet package), we quantified positive, negative, and neutral sentiments for each question.

The analysis revealed:

- **Predominantly Positive Sentiment**: Most responses expressed positive sentiments, especially regarding structured content delivery, effective use of R, and clear statistical explanations.
- **Mixed Feedback on Clarity and Pace**: While many responses were positive overall, some negative sentiments were associated with challenges such as rapid topic progression and difficulties with R syntax.
- **Critical Comments**: Negative terms like "confusing," "fast," and "examples" frequently appeared in responses about areas needing clearer instructional support.

The bar chart above (generated in the sentiment analysis code chunk) illustrates the distribution of sentiment (positive, negative, neutral) for each of the 8 survey questions.

Figure 4: Sentiment Distribution by Survey Question



This qualitative analysis reveals that students valued tutorials, TA support, and real-world connections. Many experienced growth in using R, while also noting areas for improvement such as assignment clarity and feedback quality. These insights complement the quantitative findings and can help inform future course enhancements.

4 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.

5 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