PROJECT REPORT

COMPARATIVE STATISTICAL ANALYSIS BETWEEN STEM AND NON-STEM STUDENTS

A REPORT SUBMITTED TO

सत्यं शिवं सुन्दरम्

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA
FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS
THIRD YEAR B.SC (2024-25)

UNDER THE GUIDANCE OF:
DR. RUPAL M. SHAH
MS. SHREYA MATHUR

PROJECT MEMBERS:
PURI ANKITKUMAR KAUSHLENDRA
KHARVA SHRUTI HARIVADAN
ACHARYA MATRA RAJESHBHAI
JHA ALOKCHANDRA SHREEPATI
DHEERAJ JOSHI

DECLARATION

We, students of Department of Statistics, Faculty of Science, Maharaja Sayajirao University, hereby declare that the project titled **“Comparative Statistical Analysis Between STEM and Non-STEM Students”** is our original work.

This project is undertaken as part of our final year academic requirement and has been completed with the guidance and supervision of **Dr. Rupal M. Shah** and **Ms. Shreya Mathur**. The study involves comparative analysis of motivating factors, academic performances and study patterns of students pursuing STEM fields and students who are not. The study also provides a comprehensive SWOT analysis between STEM and Non-STEM fields.

We confirm that this work has been carried out independently, using reliable data sources and appropriate statistical techniques. Any external references, data sources, or prior research used in this project have been duly cited and acknowledged.

This project has not been submitted, either partially or fully, to any other institution for any academic or professional purpose.

CERTIFICATE

This is to certify Puri Ankitkumar Kaushlendra, Kharva Shruti Harivadan, Acharya Matra Rajeshbhai, Jha Alokchandra Shreepati and Dheeraj Joshi have successfully and satisfactorily completed the project titled:

“COMPARATIVE STATISTICAL ANALYSIS BETWEEN STEM AND NON-STEM STUDENTS”

as a team for the academic year 2024-25, and have submitted the work to the Department of Statistics as a partial fulfilment to the requirement of the Bachelor's Degree in Statistics and have represented their original work under the supervision and guidance of Dr. Rupal M. Shah.

I wish them grand success for the future.

Prof. V. A. Kalamkar
Head of Department

Dr. Rupal M. Shah
Guide

Ms. Shreya Mathur
Guide

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have supported us throughout the course of this project.

First and foremost, I am deeply thankful to my project guide, Dr. Rupal M. Shah, and guide Ms. Shreya Mathur for their constant guidance, valuable suggestions, and encouragement at every stage of this research. Their insights and feedback played a crucial role in shaping this study.

We extend our heartfelt thanks to all the students who took the time to respond to our questionnaire and share their honest ideas and experiences. Without their cooperation, this project would not have been possible.

We want to express our sincere gratitude to our Head of Department Prof. V.A. Kalamkar for providing us with the platform and opportunity to work on this project and also want to thank our professors and peers from the Department of Statistics for their feedback, support, and motivation. A special note of thanks to our friends and seniors for their motivation and constant support.

Finally, we would like to acknowledge the use of various statistical tools and data analysis methods that enabled us to explore the differences between STEM and Non-STEM students with clarity and depth. This project has been an enriching experience, and we are grateful for the opportunity to work on a topic that is both relevant and insightful.

ABSTRACT

This project explores the key differences between **STEM** (Science, Technology, Engineering, and Mathematics) and **Non-STEM** students in terms of their motivations, study patterns, academic performance, interpersonal skills, life aspirations, and satisfaction with their chosen fields. It also includes a SWOT analysis and examines perceptions of unemployment scenarios across both fields.

Data was collected using a structured questionnaire to obtain a diverse representative sample of students. Statistical tools and visualization techniques were used to analyse the responses and identify significant patterns. The results highlight distinct perspectives and experiences between the two groups, offering valuable insights for educators, students, and policymakers to better understand and support diverse academic paths.

OBJECTIVES OF THE STUDY

1. To **investigate the motivating factors and reasons** behind students choosing STEM and Non-STEM courses, including personal interests, societal expectations, and career aspirations
2. To **compare the academic performance and study patterns** of STEM and Non-STEM students, highlighting differences in workload, learning strategies, and engagement.
3. To perform a **SWOT Analysis** (Strengths, Weaknesses, Opportunities, and Threats) of both STEM and Non-STEM fields to evaluate their current relevance, challenges, and future potential.

TABLE OF CONTENTS

1. Introduction

1.1 Background

In recent decades, the growing emphasis on Science, Technology, Engineering, and Mathematics (STEM) education has played a pivotal role in shaping national policies, academic institutions, and workforce dynamics. These fields are considered instrumental in driving economic growth, technological innovation, and global competitiveness (Marginson et al., 2013). On the other hand, Non-STEM disciplines—which include humanities, social sciences, arts, and commerce—are equally critical in developing soft skills, critical thinking, ethical awareness, and cultural understanding (Côté & Allahar, 2011). Understanding the dichotomy and interaction between these two educational streams is vital in formulating effective educational strategies, guiding student choices, and balancing the labour market.

In India, this conversation gains further relevance. With the implementation of the National Education Policy (NEP) 2020, a renewed focus has been placed on multidisciplinary education, vocational training, and the development of 21st-century skills (Government of India, 2020). Still, the educational landscape remains divided, with STEM often perceived as more prestigious or economically promising than its Non-STEM counterpart (Agarwal, 2016). This perception significantly influences student behaviour, course enrolment trends, and even parental expectations.

1.2 Significance of the Study

This research aims to explore and compare the profiles, motivations, academic behaviour, and aspirations of students enrolled in STEM versus Non-STEM courses in India. By doing so, the study seeks to shed light on prevailing stereotypes, gender gaps, and the socioeconomic factors influencing these educational paths. Moreover, it will assess the impact of these choices on students' future trajectories, satisfaction levels, and skill acquisition.

With the rising debate around employability, education reform, and holistic development, studies like this are essential. They not only help policymakers and educators identify gaps but also empower students with data-driven insights into their academic choices.

1.3 Global and National Context

Globally, the World Economic Forum (2023) continues to highlight the gender gap in STEM, with women making up less than 30% of professionals in core science and engineering sectors. In India, the All India Survey on Higher Education (AISHE, 2021) reports a similar disparity, alongside a growing enrolment in commerce and arts, especially among female students. The causes are multifaceted—ranging from socio-cultural barriers and lack of awareness to perceived difficulty and employability concerns.

1.4 Literature Review

A wealth of literature exists examining why students choose specific educational tracks. According to Eccles' Expectancy-Value Theory (1983), students are likely to pursue domains where they feel competent and expect positive outcomes. This theory has been used extensively to understand gendered patterns in STEM interest (Jacobs & Eccles, 2000). Similarly, Bourdieu's concept of 'cultural capital' suggests that family background and social environment greatly influence educational choices (Bourdieu, 1986).

Research by Wang and Degol (2013) found that while cognitive abilities may be similar across genders, identity and motivational factors often skew career interests. In contrast, Non-STEM pathways are often viewed as more accessible, relatable, or creatively fulfilling (Becher & Trowler, 2001). Further, studies have revealed a growing mismatch between graduate output and job market requirements, calling into question the long-term value of both streams unless supplemented with interdisciplinary training (Tomlinson, 2008).

This project builds upon these findings by conducting a comparative analysis between STEM and Non-STEM students through first-hand data collection. It further adds value by integrating a SWOT analysis to evaluate the broader strengths and limitations of both educational fields in today’s context.

2. Methodology

2.1 Research Design

The initial task in our study was to design a questionnaire that effectively addressed the research objectives. A single questionnaire was constructed to collect responses from students belonging to both STEM and Non-STEM fields. The questionnaire comprised **22** questions aimed at eliciting meaningful insights relevant to our study. It included items such as the **respondent's current year of study**, **reasons for choosing a particular field**, **types of resources used for learning, perceived challenges and opportunities**, and the **likelihood of switching fields** if given the opportunity, among others.

The subsequent step involved determining an appropriate sample size. For this purpose, **Faculty of Science and Faculty of Technology and Engineering** —were selected under the STEM category, as data from these were readily accessible. Similarly, to represent the Non-STEM group, data were collected from **the Faculty of Arts and the Faculty of Commerce**, thereby ensuring data availability while maintaining a manageable scope for the study.

To calculate the sample size, we first obtained the total number of students enrolled in each of the four selected faculties from the university headquarters. These figures were aggregated to determine the overall population size.

2.2 Sample size determination

2.2.1 Primary Data

Primary data refers to the data collected first-hand by the researcher for a specific research purpose. Unlike secondary data, which is obtained from existing sources, primary data is **original and directly gathered from the respondents**. In the context of this study, primary data was collected through a structured questionnaire administered to students from both STEM and Non-STEM fields. This approach allowed for the collection of relevant, up-to-date, and specific information aligned with the study's objectives.

The layout of our questionnaire was:

1. What is your gender?
   - ○ Male
   - ○ Female
   - ○ Prefer not to say

2. What is your current year of study?
   - ○ First
   - ○ Second
   - ○ Third
   - ○ Fourth
   - ○ Previous
   - ○ Final

3. What is your field of study?
   - ○ STEM _________
   - ○ Non-STEM ___________

4. Why did you choose your current field of study? (Select all that apply)
   - ○ Passion/Interest
   - ○ Career prospects
   - ○ Family influence
   - ○ Financial stability
   - ○ Flexibility of the field
   - ○ Other (please specify)

5. What is your latest academic performance?
   - ○ More than 70%
   - ○ Between 60 and 69.99%
   - ○ Between 50 and 59.99%
   - ○ Between 40 and 49.99%

6. How do you primarily study?
   - ○ Self-study
   - ○ Group discussions
   - ○ Tutoring sessions
   - ○ Online resources

7. On average, how many hours do you spend studying per day outside classes?
   - ○ Less than 3 hours
   - ○ 3 to 5 hours
   - ○ 5 to 7 hours
   - ○ More than 7 hours

8. How do you handle academic challenges?
   - ○ Seek help from peers and professors
   - ○ Use online resources for additional learning
   - ○ Work harder independently
   - ○ Joining study groups

9. What kind of extra-curricular activities do you involve in?
   - ○ Sports and Fitness

- Arts & Creativity
- Social and Community Service
- Internship
- Other (please specify)

10. How often do you involve in activities other than academics?
    - Regularly
    - Sometimes
    - Seldom
    - Never

11. What career path do you plan to pursue after your current course?
    - Industry
    - Research/Academia
    - Entrepreneurship
    - Freelancing/Independent Work
    - Other (please specify)

12. What skills do you think are most critical for success in your field?
    - Technical skills
    - Creativity
    - Problem-solving
    - Communication
    - Networking
    - Others

13. What class of economic section do you belong to?
    - Elite Class
    - Upper Middle Class
    - Lower Middle Class
    - Below Lower Middle

14. Are you pursuing your current field as your first choice?
    - Yes
    - No

15. In your opinion, how well does your field prepare you for the current job market?
    - Very poorly
    - Poorly
    - Neutral
    - Well
    - Very well

16. For the below question, provide number between 1 to 5 with respect to your field. (1- Very unsatisfied, 2-Unsatisfied, 3- Neutral, 4- Satisfied, 5- Very Satisfied)

With choice of your field of study	Real world application of your field of study

17. What are the main strengths of your field of study? (Select all that apply)
    - High earning potential
    - Flexibility and creativity

- ○ Contribution to society
- ○ Job security
- ○ Personal growth
- ○ Others (please specify)

18. What are the biggest challenges of your field of study? (Select all that apply)
    - ○ Heavy workload
    - ○ Limited job prospects
    - ○ High competition
    - ○ Lack of recognition
    - ○ Other (please specify)

19. What do you think is the biggest opportunity for growth in your field?
    - ○ Technological advancements
    - ○ Interdisciplinary applications
    - ○ Demand in the job market
    - ○ Other (please specify)

20. What do you think is the biggest threat to your field?
    - ○ Job automation
    - ○ Market saturation
    - ○ Economic downturns
    - ○ Lack of funding or resources
    - ○ Others (please specify)

21. Would you consider switching fields in the future?
    - ○ Yes
    - ○ No
    - ○ Maybe

22. Would you recommend your field of study to others?
    - ○ Yes
    - ○ No
    - ○ May be

2.2.2 Sampling Methods

Sampling is a crucial step in research methodology that involves selecting a **subset of individuals from a larger population to represent the whole**. In this study**, stratified random sampling** was employed to ensure representation from different academic disciplines. The population was divided into two strata—STEM and Non-STEM—and further into faculties (Science, Technology, Arts, and Commerce). Within each stratum**, proportional allocation** was used to determine the number of respondents based on the faculty-wise population. This method enhances the accuracy and representativeness of the sample by accounting for the variability across different groups. The final sample size was determined using **Cochran's formula**, with a **proportion (p) of 0.5** to reflect the assumption of equal likelihood of choosing either academic field.

Obtaining a comprehensible data, especially for our latest project, which relies on primary sources is a daunting task to perform. Making sure our data is accurate has been a big job. Along the way, we've learned a ton and faced lots of challenges. From figuring out how to collect primary data effectively to discovering new things, it's been a great learning experience for us.

We realized that on-ground data collection is very tough. The investment of time, effort and making strangers interested in the study, so that they will be willing to fill questionnaire is hard. Despite these challenges, we could collect data from those who did not have smartphones, and could not write or read-only due to on-ground data collection.

3. Statistical Methods Employed

In order to explore relationships among categorical variables and ordinal data within the collected dataset of STEM and Non-STEM students, several inferential statistical techniques were utilized. This section discusses the rationale and theoretical underpinnings of these techniques—**Chi-square test of independence**, **Goodman-Kruskal Gamma**, and **Odds Ratio**—and elaborates on their appropriateness for the nature of the data collected in this study.

3.1 Chi-Square Test of Independence

The **Chi-square (χ^2) test of independence** is a non-parametric statistical test used to determine whether there is a significant association between two categorical variables. It compares the observed frequencies in each category of a contingency table to the frequencies that would be expected if the variables were independent. The formula for the test statistic is:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where Oi is the observed frequency and Ei is the expected frequency under the null hypothesis.

In this study, the Chi-square test was applied to examine relationships between categorical variables such as **field of study** and **gender**, **career motivations**, **co-curricular involvement**, **career goals**, and **skills valued**. These are all nominal variables, for which the Chi-square test is well-suited. When the test yields a p-value less than the chosen level of significance (typically 0.05), the null hypothesis of independence is rejected, indicating a statistically significant association between the variables.

This test is appropriate for the current dataset as it involves multiple instances of cross-tabulated categorical variables and sufficient sample size (n = 270), meeting the assumptions required for the Chi-square approximation to be valid.

3.2 Standardized Pearson Residual

In the context of statistical analysis, particularly when using Chi-square tests of independence for categorical data, Standardized Pearson Residuals serve as a diagnostic tool to identify the strength and direction of the deviation of observed frequencies from the expected frequencies.

The Pearson residual for a cell in a contingency table is calculated as:

$$R_P = \frac{O - E}{\sqrt{E}}$$

In our project comparing STEM and Non-STEM students across various categorical variables (such as study mode, co-curricular involvement, career goals, etc.), we employed the Chi-square test of independence to examine the association between course type and each of these factors.

However, while the Chi-square test tells whether there is a significant association, it does not indicate which specific cells in the contingency table are responsible for this relationship. This is where Standardized Pearson Residuals become valuable.

By calculating the standardized residuals for each cell, we were able to:

Pinpoint which category combinations (e.g., STEM students preferring competitive exams, or Non-STEM students favouring creative fields) deviated the most from what was expected under independence.

Interpret the direction of the deviation—positive values indicating a cell had more observations than expected, and negative values indicating fewer.

This enhanced the interpretive power of my findings and added depth to my conclusions about how students in STEM and Non-STEM tracks differ in meaningful ways.

3.3 Goodman-Kruskal Gamma

The **Goodman-Kruskal Gamma (G)** is a measure of association used to analyse the strength and direction of the relationship between two **ordinal** variables. Unlike Pearson's correlation which assumes interval data and linearity, Gamma is designed for ordinal data and is particularly useful when analysing ranked data with tied pairs.

Gamma is computed as:

$$G=(P-Q)/(P+Q)$$

where P is the number of concordant pairs and Q is the number of discordant pairs. A Gamma value ranges from -1 (perfect negative association) to +1 (perfect positive association), with 0 indicating no association.

In this study, Goodman-Kruskal Gamma was applied to assess the relationship between **academic performance** and **study hours**, both of which were measured as ordered categories (e.g., performance bands and ranges of study hours). Gamma is appropriate here because it respects the ordinal nature of these variables and does not assume equal intervals or normality.

3.4 Odds Ratio (OR)

The **Odds Ratio (OR)** is a measure of association that quantifies the odds of an event occurring in one group relative to the odds of it occurring in another group. It is commonly used in logistic regression and 2x2 contingency tables.

The odds ratio is defined as:

$$OR=(a/c)\ (d/b)$$

where:

- a and b represent the number of successes and failures in the exposed group,
- c and d in the non-exposed group.

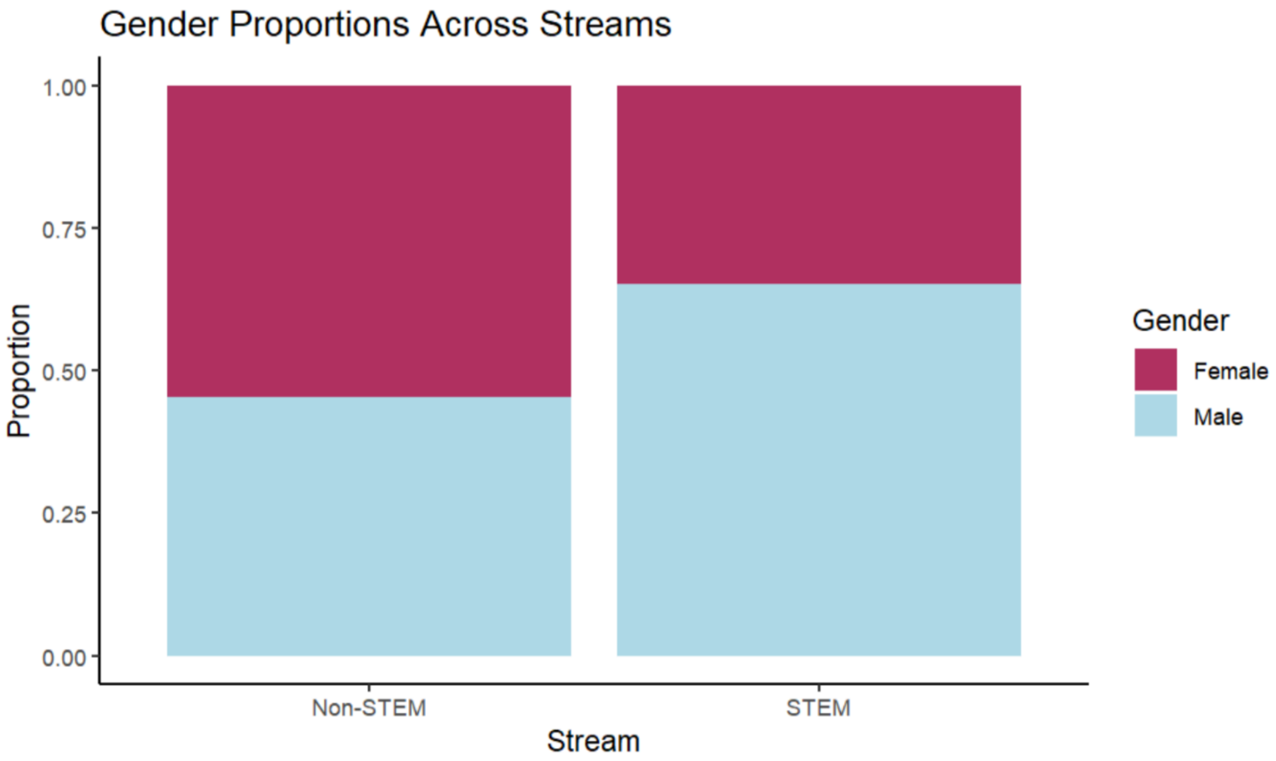
In the context of this study, the odds ratio was employed to evaluate binary outcomes such as whether a student's current field was their **first choice**, or whether they had **switched fields**. The odds of choosing STEM or Non-STEM as a first choice were compared, as were the odds of switching. While odds ratios provide an intuitive understanding of likelihood and effect size, their interpretation is contingent on the 95% confidence interval not including 1—otherwise, the result is not statistically significant.

This method is appropriate for **binary categorical outcomes** in the dataset and is particularly informative for assessing the **direction and strength** of associations where dichotomous outcomes are of interest.

4. Hypothesis Testing

4.1 Gender and Stream Association

	Stream		
Gender	**Non-STEM**	**STEM**	**Grand Total**
Female	**108**	**25**	**133**
Male	**90**	**47**	**137**
Grand Total	**198**	**72**	**270**



Question: Is there a statistically significant relationship between a student's gender and their choice of academic stream (STEM vs Non-STEM) in India?

Hypotheses:

- **H_0**: There is no association between gender and academic stream.
- **H_1**: There is an association between gender and academic stream.

Findings:
From the dataset of 270 students (198 Non-STEM and 72 STEM), females were predominantly enrolled in Non-STEM (n = 108) compared to STEM (n = 25). Males were more evenly split, with 90 in Non-STEM and 47 in STEM. A Chi-square test of independence yielded a **p-value < 0.05**, leading to the **rejection of the null hypothesis**.

Following the rejection of null hypothesis, we were interested to find out that where the gender preference for the stream lies. For this purpose, we use Standardized Pearson residual.

We then calculate expected values for the contingency table. The values for which were obtained as:

	Stream		
Gender	**Non-STEM**	**STEM**	**Grand Total**
Female	97.53	35.47	**133.00**
Male	100.47	36.53	**137.00**
Grand Total	**198.00**	**72.00**	**270**

Then we calculate the Standardised Pearson Residual using the formula discussed earlier. The values obtained are:

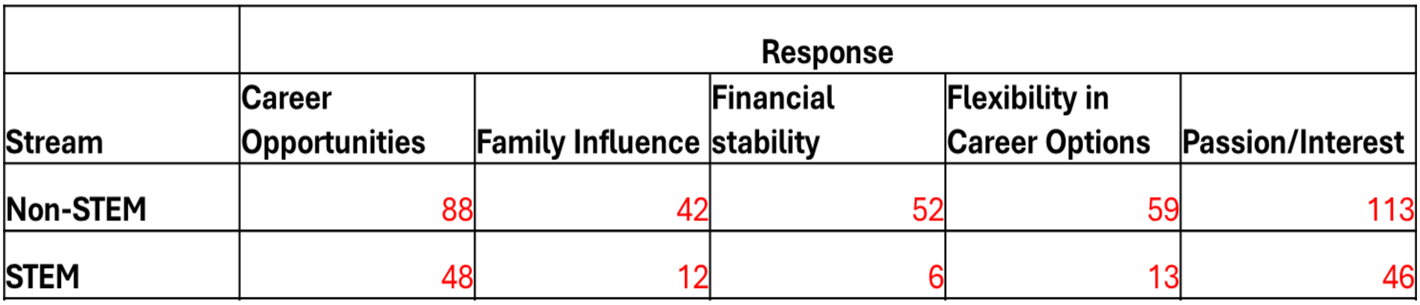
	Stream	
Gender	**Non-STEM**	**STEM**
Female	1.06	-1.76
Male	-1.04	1.73

Standardized Pearson residuals revealed that females were **more likely than expected** to choose Non-STEM and **less likely to choose STEM**, while males exhibited the reverse pattern.

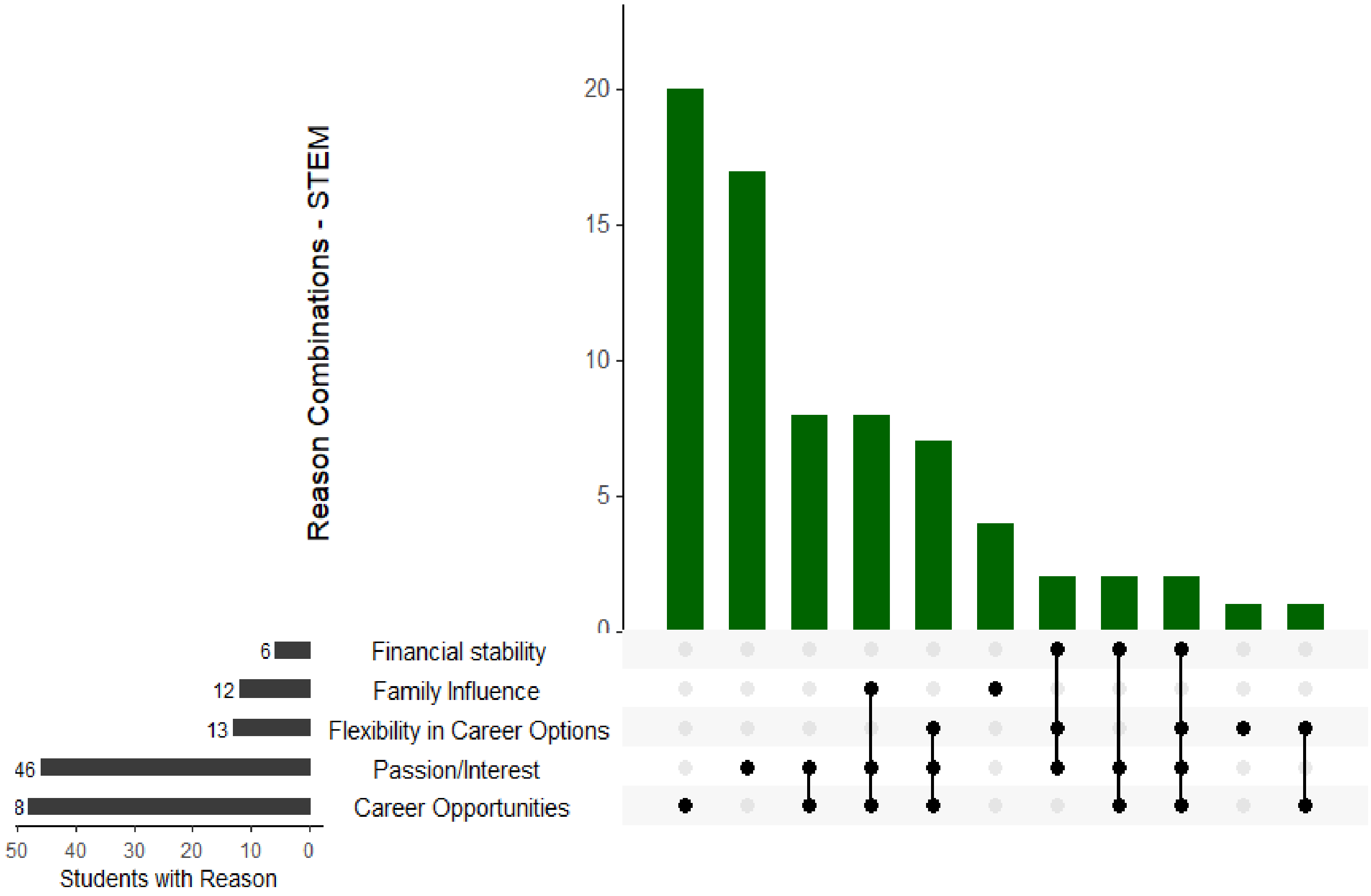
Interpretation: These results affirm the global and national findings on gender-based educational patterns. Despite policy efforts to bridge the gap, societal norms and gendered perceptions of STEM careers continue to influence educational choices in India (Marginson et al., 2013; Government of India, 2020).

Rationale: The World Economic Forum (2023) has consistently reported a persistent gender gap in STEM, with women comprising less than 30% of professionals in core science and engineering fields. In India, the AISHE (2021) similarly highlights an underrepresentation of female students in STEM disciplines. Eccles' Expectancy-Value Theory (1983) and Jacobs and Eccles (2000) suggest that such disparities may stem from perceived competence, value expectations, and social encouragement, which vary by gender.

4.2 Motivations Behind Stream Choice

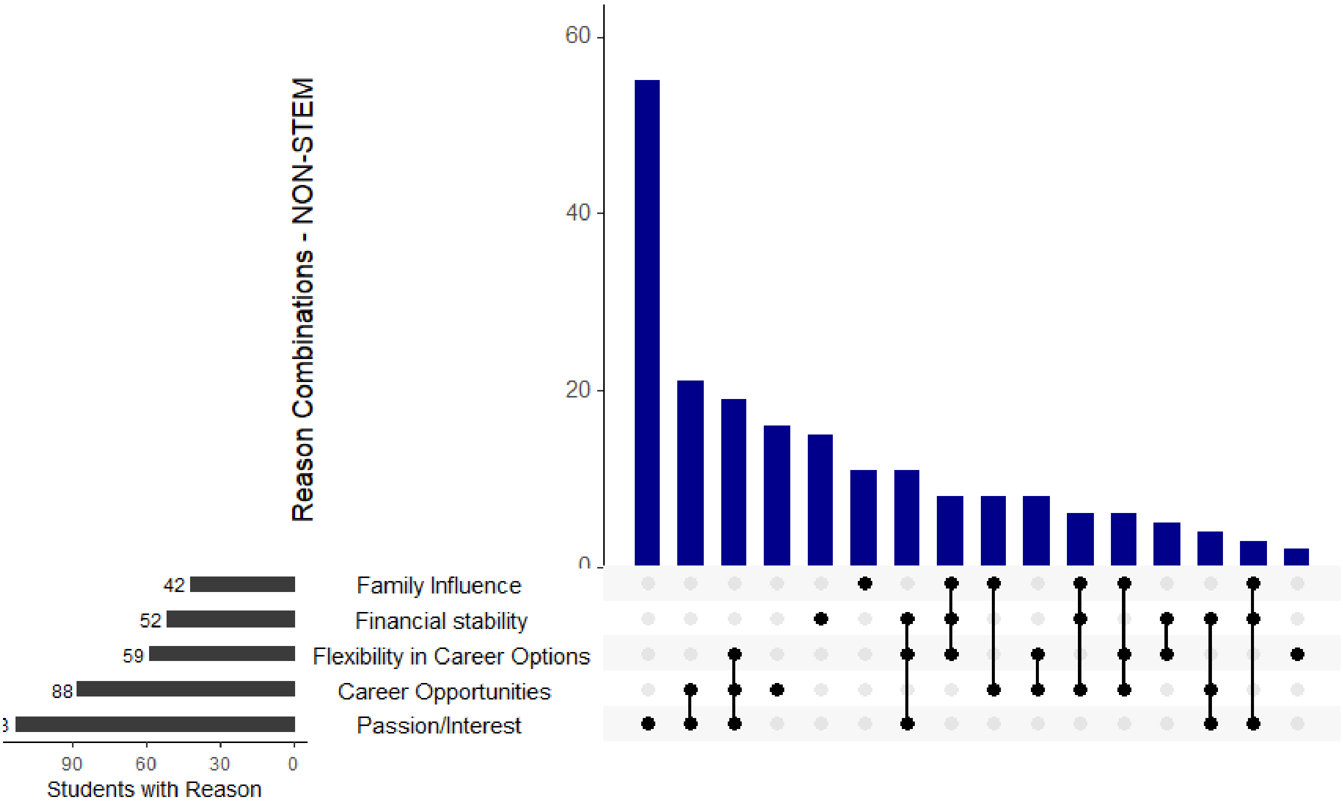


	Response				
Stream	Career Opportunities	Family Influence	Financial stability	Flexibility in Career Options	Passion/Interest
Non-STEM	88	42	52	59	113
STEM	48	12	6	13	46



Combinations: The most common combination is students choosing just Passion/Interest or Career Opportunities. Few students combine multiple reasons, but when they do, it's often Passion + Career Opportunities + Flexibility**.**

Insight: STEM students are mainly motivated by intrinsic factors (passion) and practical outcomes (career opportunities). Financial stability and family pressure are less influential for STEM students.



Combinations: Students in Non-STEM fields tend to have more diverse combinations of reasons. Passion is often combined with Career Opportunities, Flexibility, and Family Influence.

Insight: Non-STEM students are also driven heavily by passion but show more concern for financial security and family influence compared to STEM students. They seem to consider a broader range of factors before making career decisions.

Question: Do students from different academic streams report significantly different reasons for choosing their respective fields?

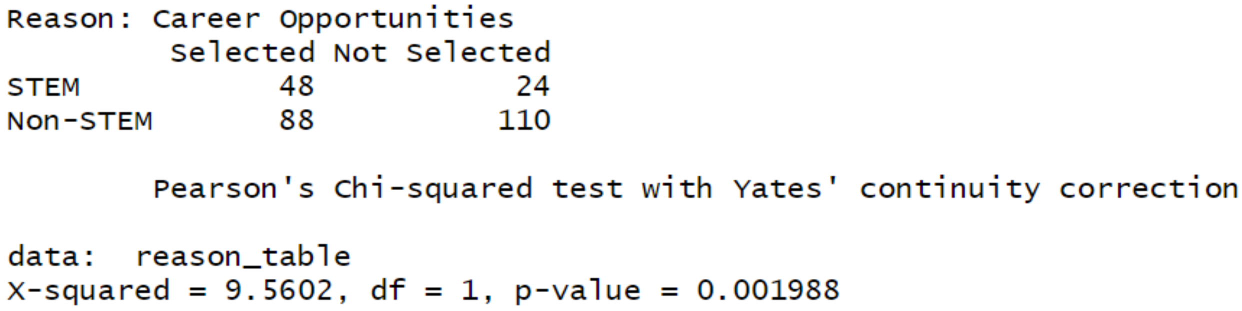
Here, the counts are from the multi-responses question where each respondent could choose more than one option irrespective of the preference of their reason or motivation.

Hence, using Chi Square test of independence to check if there is any significant association between the Stream and the reasons, directly on this frequency table would be inappropriate. Therefore, we modify our study and check if each option is a significant reason for choosing a stream or not.

2x2 table is formed for every reason so that the rows become independent and henceforth Chi Square test is applied to check if that is a significant reason or not.

Hypotheses:

- **H_0**: There is no association between academic stream and reason being career opportunities.
- **H_1**: There is an association between academic stream and reason being career opportunities.



```
Reason: Career Opportunities
         Selected Not Selected
STEM           48           24
Non-STEM       88          110

        Pearson's Chi-squared test with Yates' continuity correction

data:  reason_table
X-squared = 9.5602, df = 1, p-value = 0.001988
```

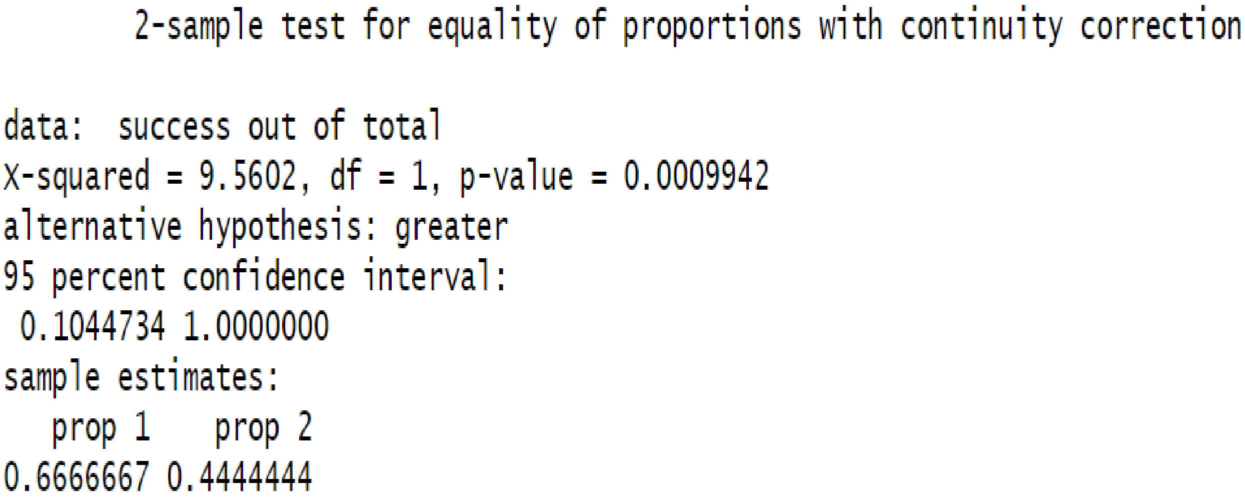
Findings:
For the reason Career Opportunities, the analysis revealed notable differences between STEM and Non-STEM students. Among STEM students, 48 selected career opportunities as a reason, while 24 did not. In contrast, among Non-STEM students, 88 selected it and 110 did not. A Pearson's Chi-squared test was conducted, yielding a chi-square statistic of 9.5602 with 1 degree of freedom and a p-value of 0.001988. Since the p-value is less than 0.05, we conclude that there is a statistically

significant association between the type of course (STEM vs Non-STEM) and citing career opportunities as a reason for their choice.

Having established a statistically significant association between career opportunities and academic stream, the next objective is to determine which stream predominantly identifies the career opportunities as a primary motivating factor. To address this, a one-sided proportion test is employed. The underlying assumption guiding this analysis is that students from STEM disciplines are more likely to prioritize career opportunities compared to their Non-STEM counterparts. Accordingly, the hypotheses for the one-sided test are formulated as follows:

- **H_0:** $\boldsymbol{P_1 \le P_2}$ (The proportion of STEM students who choose career opportunities as a primary motivation is less than or equal to the proportion of Non-STEM students.)
- **H_1:** $\boldsymbol{P_1 > P_2}$ (The proportion of STEM students who choose career opportunities as a primary motivation is greater than the proportion of Non-STEM students.)

$\boldsymbol{P_1}$ = Proportion of STEM students choosing Career Opportunities
$\boldsymbol{P_2}$ = Proportion of Non-STEM students choosing Career Opportunities



```
        2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 9.5602, df = 1, p-value = 0.0009942
alternative hypothesis: greater
95 percent confidence interval:
 0.1044734 1.0000000
sample estimates:
   prop 1    prop 2 
0.6666667 0.4444444 
```

Findings:
To investigate whether STEM and Non-STEM students differ in citing "Career Opportunities" as a motivating factor, a one-sided two-sample test for equality of proportions was conducted. Among STEM students, **48 out of 72** (approximately **66.67%**) selected "Career Opportunities," whereas among Non-STEM students, **88 out of 198** (approximately **44.44%**) selected it. The hypothesis test yielded a chi-squared value of **9.5602** with

1 degree of freedom and a p-value of **0.0009942**. Since the p-value is much smaller than the significance level of 0.05, we **reject the null hypothesis.**

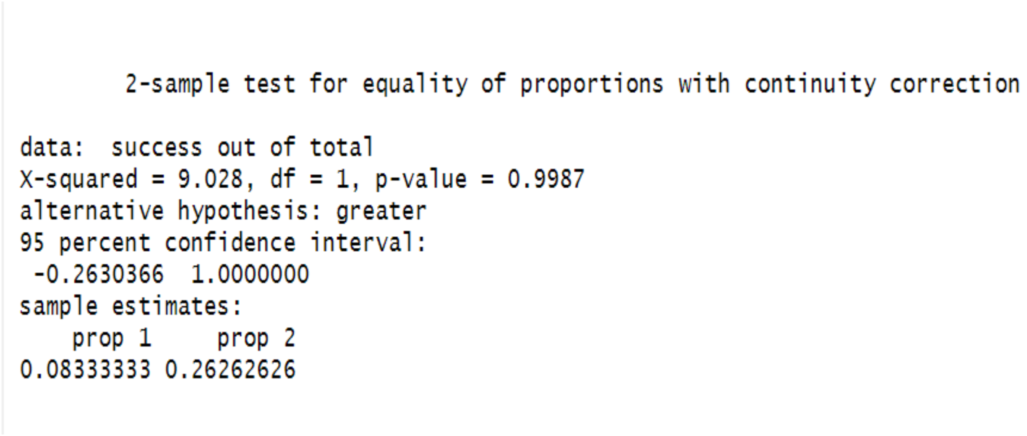
Thus, there is strong statistical evidence that STEM students are more likely than Non-STEM students to cite "Career Opportunities" as a reason for their course choice. The one-sided **95% confidence interval** for the difference in proportions ranged from **0.1047 to 1**, further supporting the finding that the proportion is significantly higher among STEM students compared to Non-STEM students.

In similar fashion, the Chi square test of independence is applied to contingency table for each reason. Below are the reasons and significant values.

Reasons	p-value	Significance
Career Opportunities	**0.001988**	**Significant**
Family Influence	**0.5133**	**Non-Significant**
Financial Stability	**0.002659**	**Significant**
Flexibility in career options	**0.07608**	**Non-Significant**
Passion/Interest	**0.3859**	**Non-Significant**

We observe that the association between stream and the reason being Financial Stability is also significant. To determine students of which stream are more influenced by financial stability, we apply one sided proportion test. The hypothesis is formulated as,

- **H_0:** $P_1 \leq P_2$ (The proportion of STEM students who choose financial stability as a primary motivation is less than or equal to the proportion of Non-STEM students.)
- **H_1:** $P_1 > P_2$ (The proportion of STEM students who choose financial stability as a primary motivation is greater than the proportion of Non-STEM students.)



```
	2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 9.028, df = 1, p-value = 0.9987
alternative hypothesis: greater
95 percent confidence interval:
 -0.2630366  1.0000000
sample estimates:
    prop 1     prop 2 
0.08333333 0.26262626 
```

Findings:
To investigate whether STEM and Non-STEM students differ in citing "Financial Stability" as a motivating factor, a one-sided two-sample test for equality of proportions was conducted. Among STEM students, 6 out of 72 (approximately 8%) selected "Financial Stability," whereas among Non-STEM students, 52 out of 198 (approximately 26.26%) selected it. The hypothesis test yielded a chi-squared value of 9.028 with 1 degree of freedom and a p-value of 0.9987. Since the p-value is larger than the significance level of 0.05, we do not have enough evidence to reject the null hypothesis.

Hence, we conclude that more students from Non-STEM group are studying in their field as compared to students in STEM group.

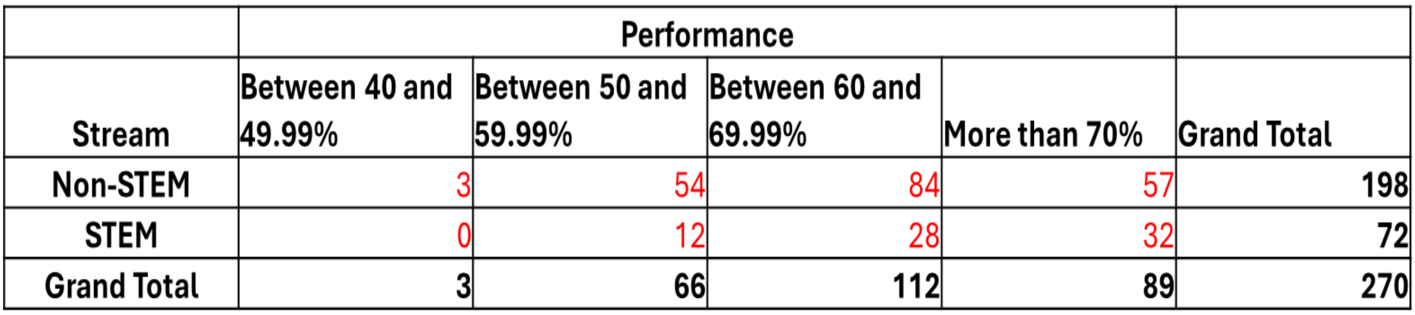
- **Interpretation**: These findings align with the motivational frameworks proposed by Eccles and further highlight the dichotomy in decision-making patterns: **extrinsically driven** choices in Non-STEM and **intrinsically motivated** ones in STEM (Wang & Degol, 2013). The above heatmap shows that for STEM students, Career Opportunities displays a significant value at 5% level of significance and hence can be considered as the prime reason. While on the other hand, Non-STEM students are more motivated by the Financial Stability that their field provides them.

STEM students are more driven by intrinsic motivation, especially passion or interest. Non-STEM students are more influenced by external factors like career prospects, financial stability, and family influence. There appears to be a statistically significant association between field of study (STEM vs Non-STEM) and career influencing factors.

Rationale: According to Bourdieu's theory of cultural capital (1986), students' academic paths are often shaped by their socio-cultural contexts. Furthermore, Côté and Allahar (2011) emphasize that instrumental motivations, such as job security and economic benefits, heavily influence student decision-making in Non-STEM fields, while intrinsic motivations dominate in STEM.

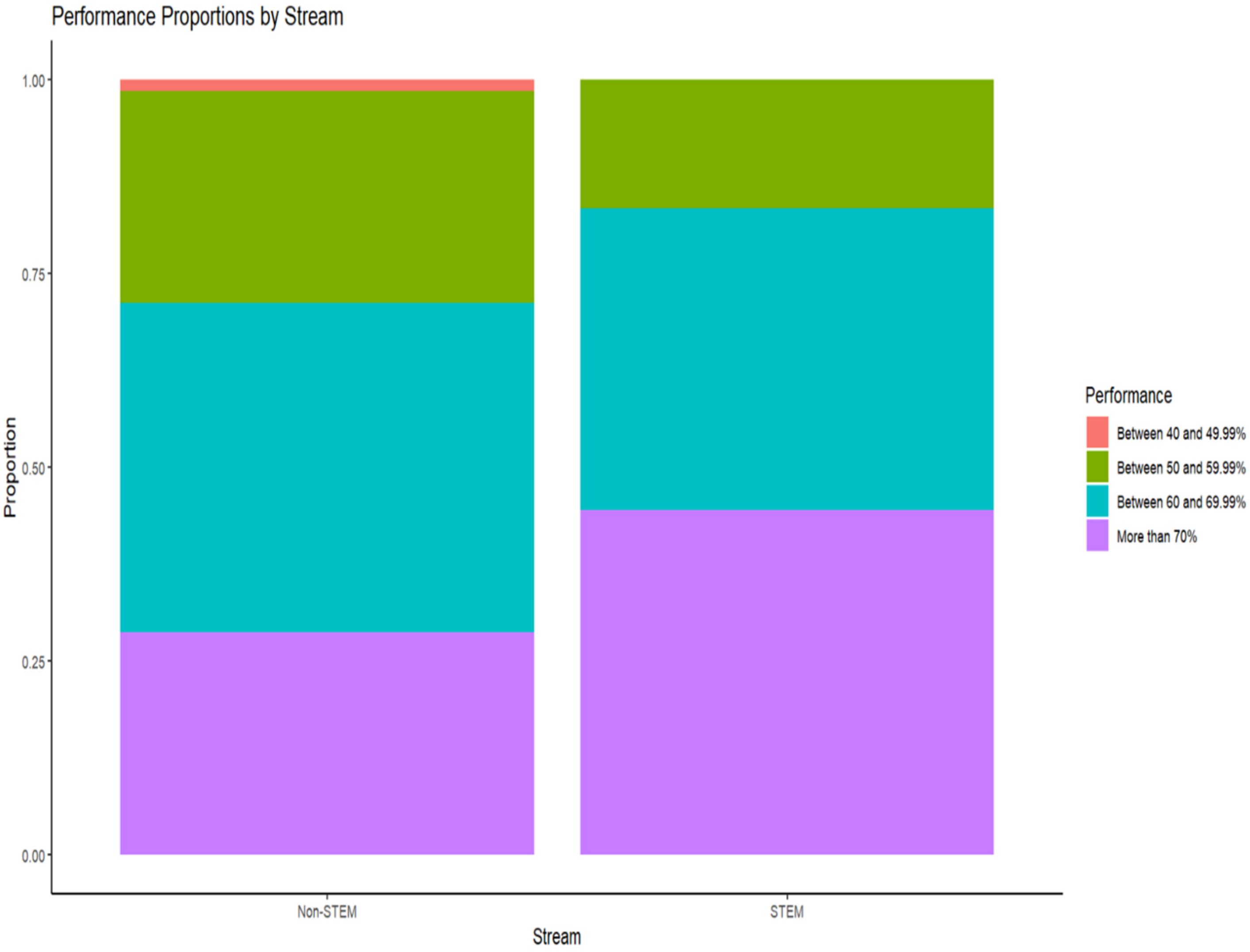
4.3 Academic Performance

Question: Is there a relationship between student's academic stream and their academic performance?



	Performance				
Stream	Between 40 and 49.99%	Between 50 and 59.99%	Between 60 and 69.99%	More than 70%	Grand Total
Non-STEM	3	54	84	57	198
STEM	0	12	28	32	72
Grand Total	3	66	112	89	270

Below shown is the performance of students by their stream.



Hypotheses:

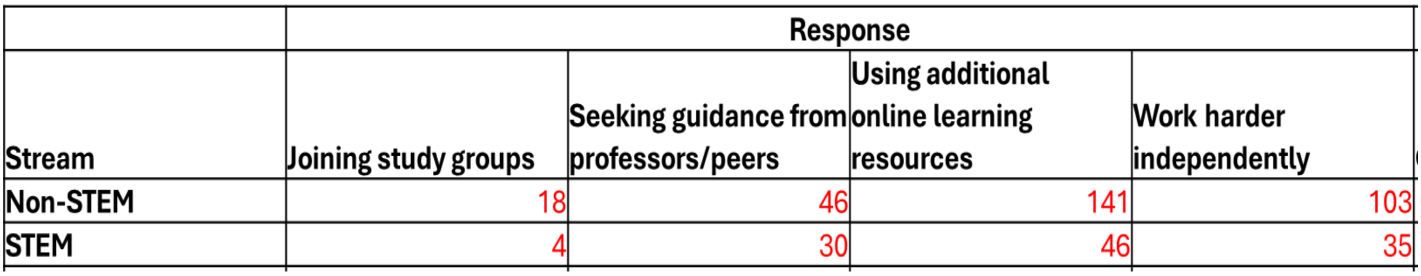
- **H_0**: There is no association between Academic stream and Academic performance.
- **H_1**: There is an association between Academic stream and Academic performance.

Findings:
Students were categorized into four performance bands. Non-STEM had an even distribution across performance brackets, whereas STEM students clustered around the 60–70% and 70%+ brackets. However, the Chi-square test yielded a **p-value > 0.05**, leading to the **non-rejection of the null hypothesis**.

Further analysis using **Goodman-Kruskal Gamma** yielded a value of **-0.019**, and **Kendall's Tau** was **0.012** with a **p-value = 0.80**, both indicating no meaningful correlation between **study hours per day** and **academic performance**.

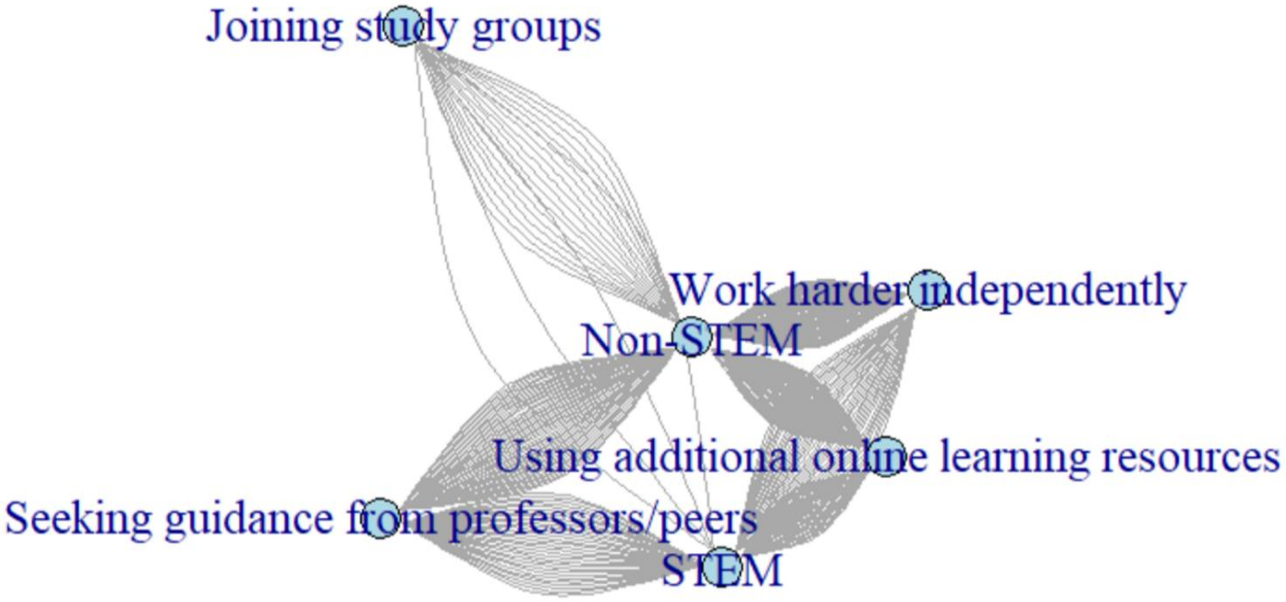
- **Interpretation**: Academic outcomes do not differ significantly by stream, suggesting that **disciplinary rigor or grading standards** may not necessarily be stream specific. This challenges stereotypes that one stream is academically superior or easier than the other.

4.4 Learning Resource Utilization



	Response			
Stream	Joining study groups	Seeking guidance from professors/peers	Using additional online learning resources	Work harder independently
Non-STEM	18	46	141	103
STEM	4	30	46	35

Network graph of stream to the resources used

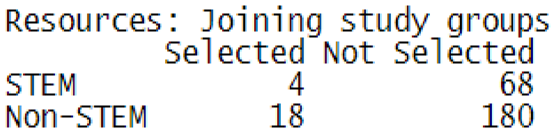

Joining study groups
Work harder independently
Non-STEM
Using additional online learning resources
Seeking guidance from professors/peers
STEM

The above diagram shows the streams as nodes connected to the resources used by the students of respective fields. Each line represents one student and the density of the lines to the same group shows how closely one stream uses the resources. The distance between Stream nodes and the resource options do not signify anything and is placed that way just for the aesthetics.

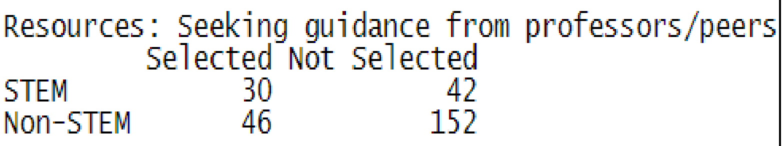
Question: Do STEM and Non-STEM students differ in the types of learning resources they use?

We then check the associations between the stream and resources used by taking each of the resource independently and making a 2x2 table.

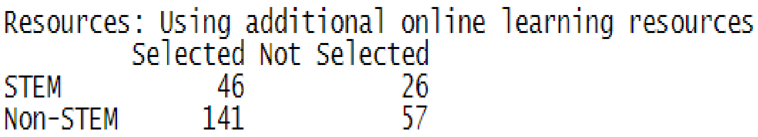
The counts for each reason are:



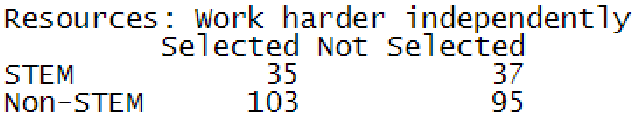
```
Resources: Joining study groups
          Selected Not Selected
STEM             4           68
Non-STEM        18          180
```



```
Resources: Seeking guidance from professors/peers
          Selected Not Selected
STEM            30           42
Non-STEM        46          152
```



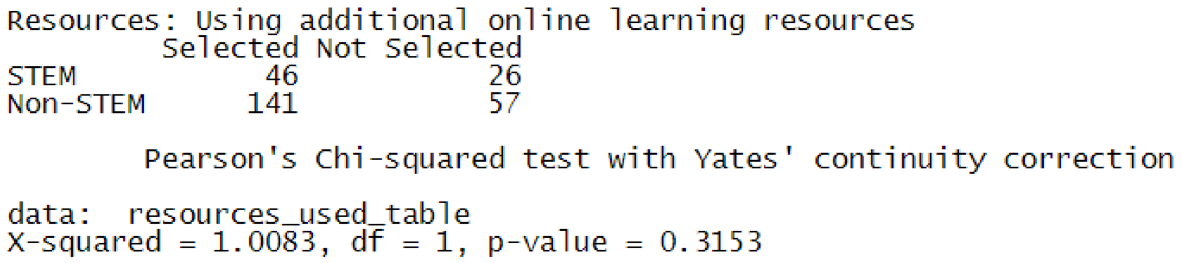
```
Resources: Using additional online learning resources
          Selected Not Selected
STEM            46           26
Non-STEM       141           57
```



```
Resources: Work harder independently
          Selected Not Selected
STEM            35           37
Non-STEM       103           95
```

We presume Using additional online resources to our variable of interest. Hence, the hypothesis is formed in the following way,

- **H_0**: There is no association between academic stream and resources used being online resources.
- **H_1**: There is an association between academic stream and resources used being online resources.



```
Resources: Using additional online learning resources
           Selected Not Selected
STEM             46           26
Non-STEM        141           57

        Pearson's Chi-squared test with Yates' continuity correction

data:  resources_used_table
X-squared = 1.0083, df = 1, p-value = 0.3153
```

Findings:
Chi-squared test of independence was conducted to examine the association between academic stream (STEM vs. Non-STEM) and the use of additional online learning resources. The observed frequencies indicated that among STEM students, 46 reported using additional online resources while 26 did not. Similarly, among Non-STEM students, 141 reported using additional resources, whereas 57 did not.

The results of the Chi-squared test yielded a test statistic of $\chi^2(1, N = 270) = 1.0083$ with a corresponding p-value of 0.3153. Since the p-value exceeds the conventional significance level of 0.05, we fail to reject the null hypothesis. This suggests that there is no statistically significant association between the academic stream and the use of additional online learning resources among the students surveyed.

Thus, the use of additional online resources appears to be independent of whether students are enrolled in STEM or Non-STEM fields.

Chi Square Test is then implemented to each of four contingency tables. The Chi-Square values and p- values thus obtained are:

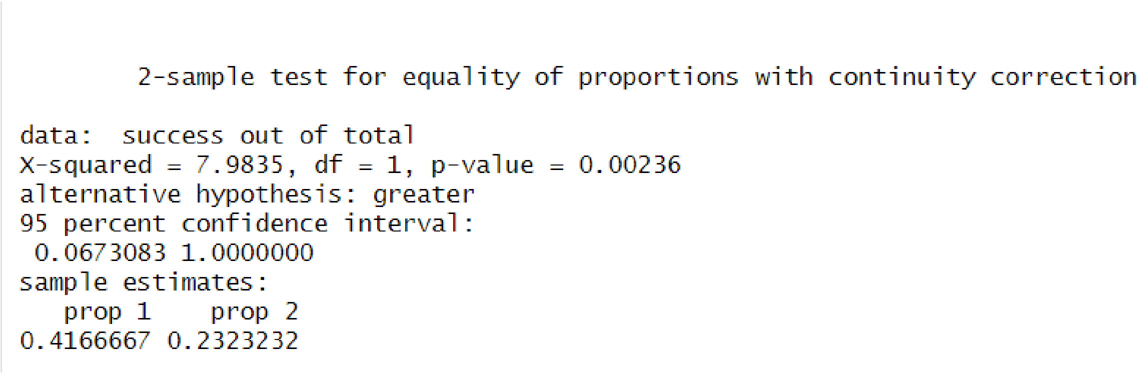
Resources Used	Chi-Square values	p-values	Significance
Joining study groups	0.47266	**0.4918**	Non-Significant
Seeking guidance from professors/peers	7.9835	**0.00472**	Significant
Using additional online learning resources	1.0083	**0.3153**	Non-Significant
Work harder independently	0.12809	**0.7204**	Non-Significant

We observe that the association between stream and the resource being seeking guidance from professors/peers is significant. All the other resources do not have a significant association with stream and thus using them is independent of the field they are studying in.

It is established that statistically significant association between seeking help from professors/peers and academic stream, the next objective is to determine which stream predominantly seeks help from professors/peers as a coping mechanism. To address this, a one-sided proportion test is employed. The underlying assumption guiding this analysis is that students from STEM disciplines are more likely to prioritize career opportunities compared to their Non-STEM counterparts. Accordingly, the hypotheses for the one-sided test are formulated as follows:

- **H_0:** $P_1 \leq P_2$ (The proportion of STEM students who choose seeking help from professors/peers as a comping mechanism is less than or equal to the proportion of Non-STEM students.)
- **H_1:** $P_1 > P_2$ (The proportion of STEM students who choose seeking help from professors/peers as a coping mechanism is greater than the proportion of Non-STEM students.)

P_1 = Proportion of STEM students select seeking help from professors/peers.
P_2 = Proportion of Non-STEM students select seeking help from professors/peers.



```
        2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 7.9835, df = 1, p-value = 0.00236
alternative hypothesis: greater
95 percent confidence interval:
 0.0673083 1.0000000
sample estimates:
   prop 1    prop 2 
0.4166667 0.2323232 
```

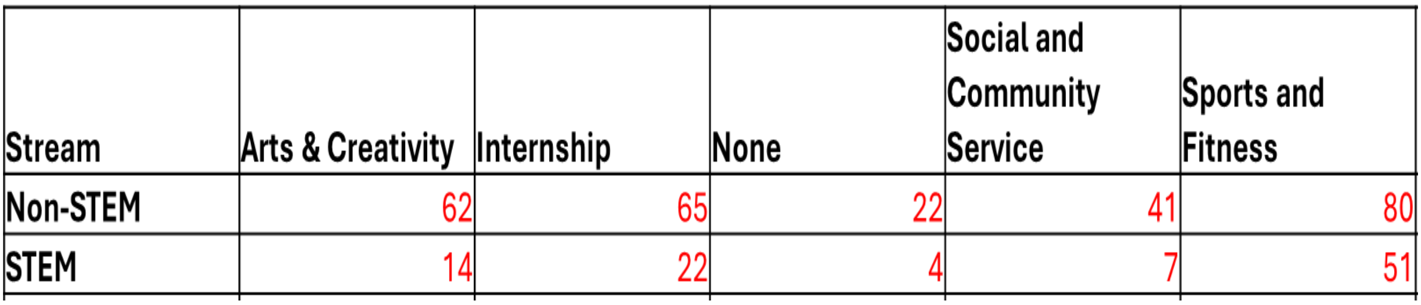
Findings:

A two-sample test for equality of proportions with continuity correction was conducted to examine whether the proportion of success differs significantly between the two groups. The results revealed a Chi-squared statistic of X² (1, N = sample size) = 7.9835 with a p-value of 0.00236. Given that the p-value is less than the conventional significance level of 0.05, we reject the null hypothesis and conclude that there is a significant difference in proportions between the two groups.

The alternative hypothesis tested was that the proportion in the first group (prop 1) is greater than that in the second group (prop 2). The sample proportion for group 1 was 0.4167, while for group 2 it was 0.2323. Furthermore, the 95% confidence interval for the difference in proportions was [0.0678, 1.0000], which does not include zero, providing additional evidence of a significant difference. Thus, it can be inferred that the proportion of seeking help from professors and peers in STEM students is significantly higher than that in Non-STEM students.

Interpretation: This suggests **digital equity** and **common study strategies** across streams, possibly reflecting the democratization of learning through online platforms and collaborative learning, regardless of disciplinary differences.

4.5 Co-curricular Involvement



Stream	Arts & Creativity	Internship	None	Social and Community Service	Sports and Fitness
Non-STEM	62	65	22	41	80
STEM	14	22	4	7	51

Here again, the counts are from the multi-responses question where each respondent could choose more than one option irrespective of the preference of their Cocurricular Activities.

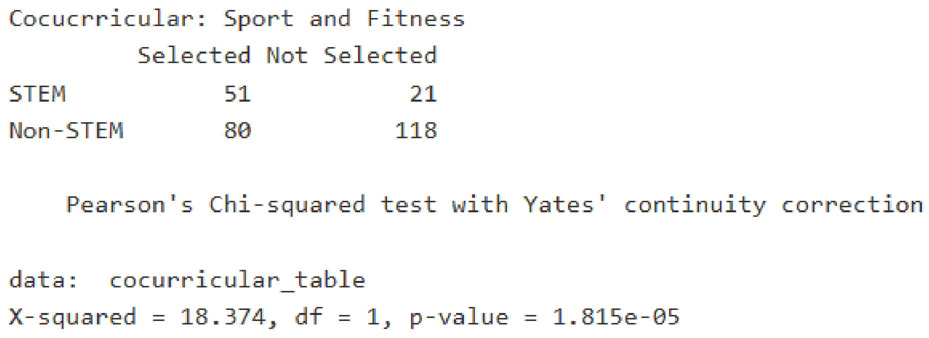
Question: Is co-curricular engagement associated with the choice of academic stream?

Using Chi Square test of independence to check if there is any significant association between the Stream and the Cocurricular Activities, directly on this frequency table would be inappropriate. Therefore, we modify our study and check if is there any association between co-curricular & the choice of academic stream or not.

2x2 table is formed for every co-curricular so that the rows become independent and henceforth Chi Square test is applied to check if that is a significant association between cocurricular activity and stream or not.

Hypotheses:

- **H_0**: There is no association between Stream and Co-curricular involvement.
- **H_1**: There is an association between Stream and Co-curricular involvement.



```
Cocucrricular: Sport and Fitness
         Selected Not Selected
STEM           51           21
Non-STEM       80          118

    Pearson's Chi-squared test with Yates' continuity correction

data:  cocurricular_table
X-squared = 18.374, df = 1, p-value = 1.815e-05
```

Findings:
For the Cocurricular activity Sports & Fitness, the analysis revealed notable differences between STEM and Non-STEM students. Among STEM students, 51 selected Sports & Fitness as a co-curricular, while 21 did not. In contrast, among Non-STEM students, 80 selected it and 118 did not. A Pearson's Chi-squared test was conducted, yielding a chi-square statistic of 18.374 with 1 degree of freedom and a p-value of 1.815e-05. Since the p-value is less than 0.05, we conclude that there is

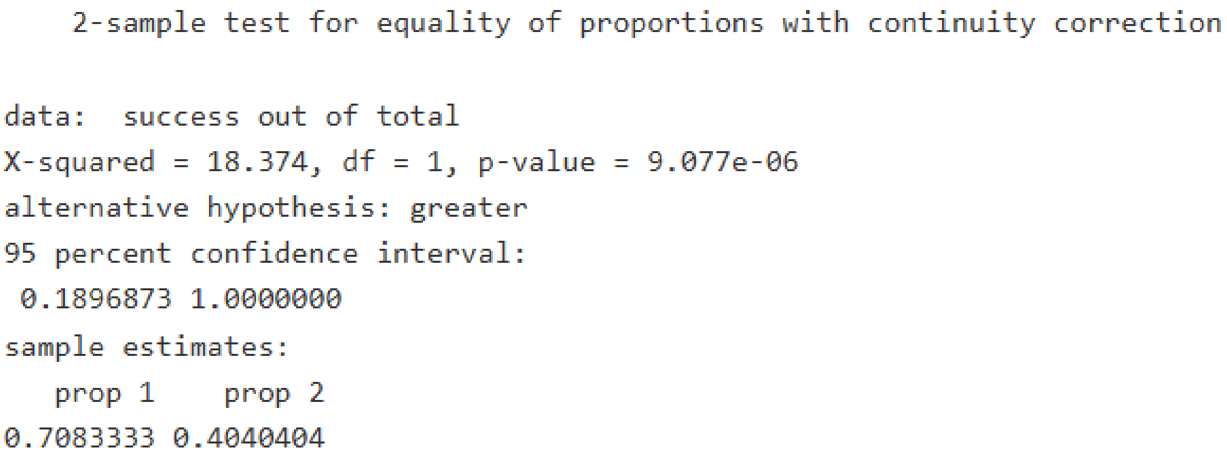
a statistically significant association between the type of course (STEM vs Non-STEM) and citing Sports & Fitness as a co-curricular activity.

Having established a statistically significant association between Sports & Fitness and academic stream, the next objective is to determine which stream select Sports & Fitness as a co-curricular activity. To address this, a one-sided proportion test is employed. The underlying assumption guiding this analysis is that students from Non-STEM disciplines are more likely to select Sports & Fitness as a co-curricular activity as compared to their STEM counterparts. Accordingly, the hypotheses for the one-sided test are formulated as follows:

- **H_0:** $P_1 \geq P_2$ (The proportion of STEM students who Sports & Fitness as a co-curricular activity is greater than or equal to the proportion of Non-STEM students.)

- **H_1:** $P_1 < P_2$ (The proportion of STEM students who select Sports & Fitness as a co-curricular activity is less than the proportion of Non-STEM students.)

P_1 = Proportion of STEM students select Sports & Fitness.
P_2 = Proportion of Non-STEM students select Sports & Fitness.



```
	2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 18.374, df = 1, p-value = 9.077e-06
alternative hypothesis: greater
95 percent confidence interval:
 0.1896873 1.0000000
sample estimates:
   prop 1    prop 2 
0.7083333 0.4040404 
```

Findings:
To investigate whether STEM and Non-STEM students differ in select “Sports & Fitness” as a co-curricular activity, a one-sided two-sample test for equality of proportions was conducted. Among STEM students, 51 out of 72 (approximately 70.83%) selected "Sports & Fitness", whereas among Non-STEM students, 80 out of 198 (approximately 40.40%) selected it. The hypothesis test yielded a chi-squared value of 18.374 with

1 degree of freedom and a p-value of 9.077e-06. Since the p-value is much smaller than the significance level of 0.05, we reject the null hypothesis.

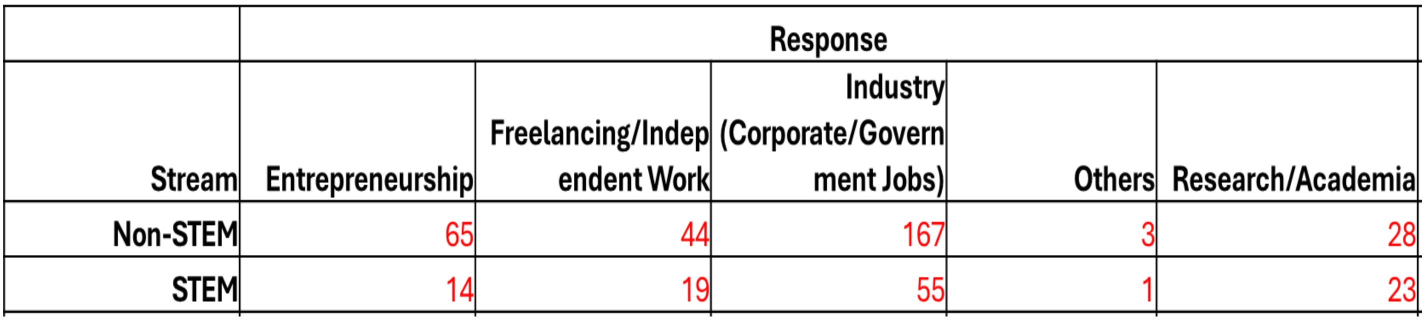
Thus, there is strong statistical evidence that Non-STEM students are more likely than STEM students to cite "Sports & Fitness" as a co-curricular activity for their course choice. The one-sided 95% confidence interval for the difference in proportions ranged from 0.1897 to 1, further supporting the finding that the proportion is significantly higher among Non-STEM students compared to STEM students.

In similar fashion, the Chi square test of independence is applied to contingency table for each reason. Below are the reasons and significant values.

Cocurricular	p-value	Significance
Arts & Creativity	**0.07762**	Non-significant
Internship	**0.8367**	Non-significant
Social & Community Service	**0.05642**	Non-significant
Sports & Fitness	**1.815e-05**	Significant
None	**0.2563**	Non-significant

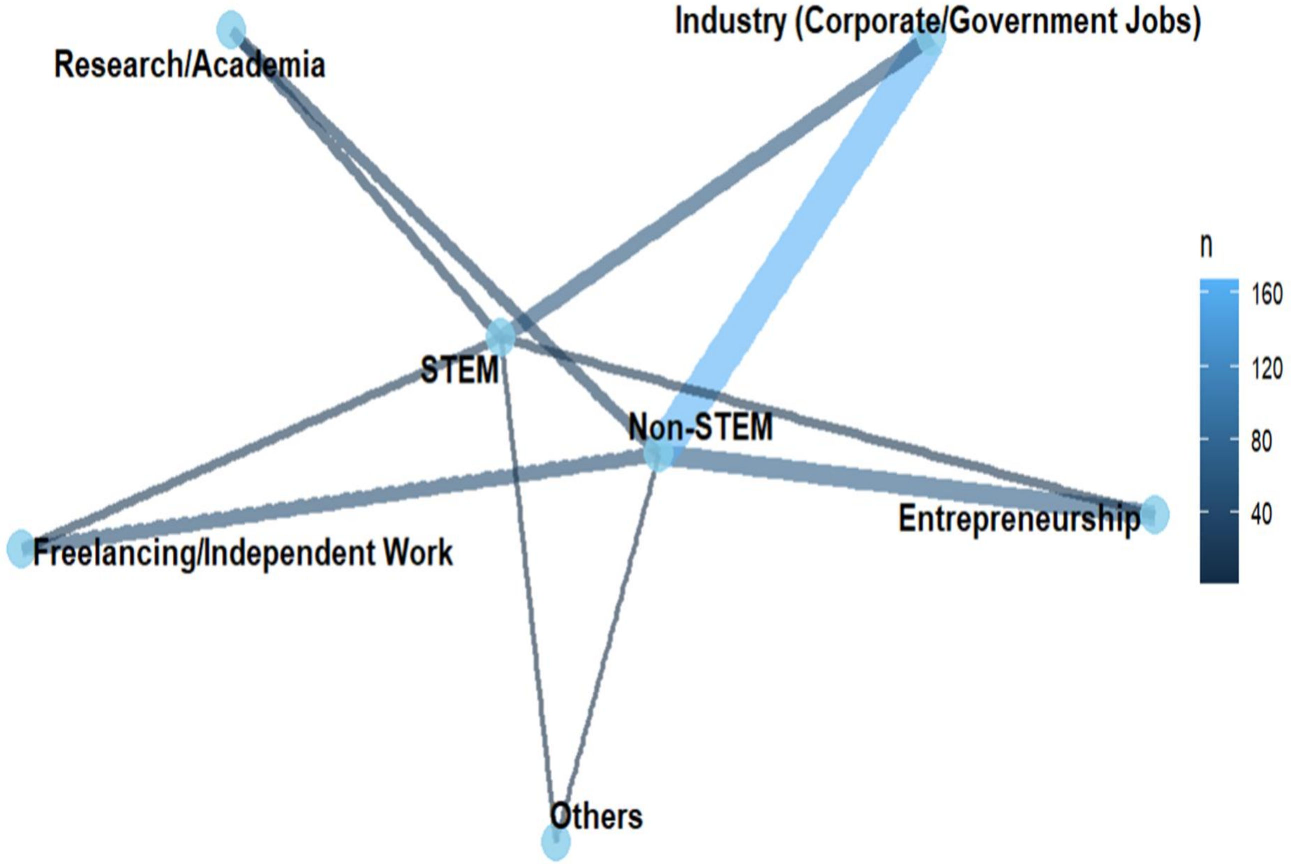
Interpretation: This aligns with Becher & Trowler's (2001) disciplinary cultures framework, where the nature of knowledge and community expectations influence broader engagement patterns. STEM's performance-based culture may align more with competitive sports, while Non-STEM's expressive and humanistic traditions align with arts and service.

4.6 Career Aspirations



	Response				
Stream	Entrepreneurship	Freelancing/Independent Work	Industry (Corporate/Government Jobs)	Others	Research/Academia
Non-STEM	65	44	167	3	28
STEM	14	19	55	1	23

Career Goals Network by Stream



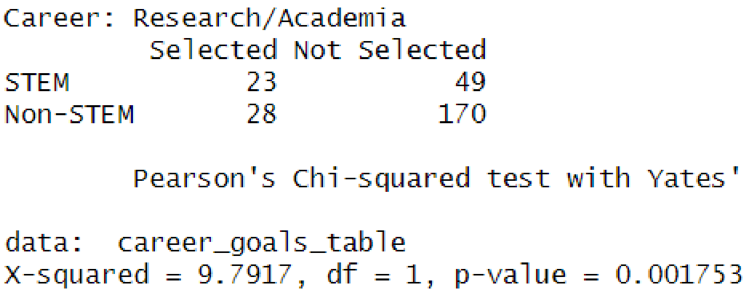
Here again, the counts are from the multi-responses question where each respondent could choose more than one option irrespective of the preference of their Career Aspirants.

Hence, using Chi Square test of independence to check if there is any significant association between the Stream and the career aspirants, directly on this frequency table would be inappropriate. Therefore, we modify our study and check if is there any association between career aspirants & the choice of academic stream or not.

2x2 table is formed for every career aspirants so that the rows become independent and henceforth Chi Square test is applied to check if that is a significant association between career aspirants and stream or not.

Hypotheses:

- **H_0**: There is no association between Stream and career aspirants.
- **H_1**: There is an association between Stream and career aspirants.



```
Career: Research/Academia
          Selected Not Selected
STEM            23           49
Non-STEM        28          170

        Pearson's Chi-squared test with Yates' continuity correction

data:  career_goals_table
X-squared = 9.7917, df = 1, p-value = 0.001753
```

Findings:
For the Career aspirants as Research/Academia, the analysis revealed notable differences between STEM and Non-STEM students. Among STEM students, 23 selected Research/Academia as career aspirants, while 49 did not. In contrast, among Non-STEM students, 28 selected it and 170 did not. A Pearson's Chi-squared test was conducted, yielding a chi-square statistic of 9.7917 with 1 degree of freedom and a p-value of 0.001753. Since the p-value is less than 0.05, we conclude that there is a statistically significant association between the type of course (STEM vs Non-STEM) and citing Research/Academia as career aspirants.

These tables allowed us to observe the distribution of choices across the two academic streams. Subsequently, a Chi-squared test of independence was performed for each 2×2 table to assess whether there was a significant association between the academic stream and the selection of specific career goals. The resulting Chi-squared values and corresponding p-values are presented as follows.

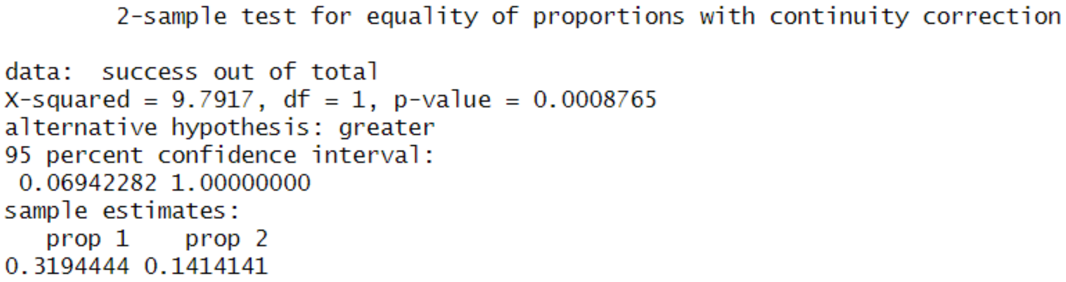
Goals	Chi-square	p-values	Significance
Entrepreneurship	3.9457	**0.04699**	Significant
Freelancing/Independent work	0.30597	**0.5802**	Non-significant
Industry (Corporate/Government jobs)	1.7738	**0.1829**	Non-significant
Research/Academia	9.7917	**0.001753**	Significant
Others	1.8804e-29	**1**	Non-significant

Having established a statistically significant association between Research/Academia and academic stream, the next objective is to determine which stream select Research/Academia as career aspirants . To address this, a one-sided proportion test is employed. The underlying assumption guiding this analysis is that students from Non-STEM disciplines are more likely to select Sports & Fitness as a co-curricular activity as compared to their STEM counterparts. Accordingly, the hypotheses for the one-sided test are formulated as follows:

- **H_0:** $\boldsymbol{P_1} \le \boldsymbol{P_2}$ (The proportion of STEM students who select Research/Academia is less than or equal to the proportion of Non-STEM students.)

- **H_1:** $\boldsymbol{P_1} > \boldsymbol{P_2}$ (The proportion of STEM students who select Research/Academia is greater than the proportion of Non-STEM students.)

$\boldsymbol{P_1}$ = Proportion of STEM students selecting Research/Academia.
$\boldsymbol{P_2}$ = Proportion of Non-STEM students selecting Research/Academia.



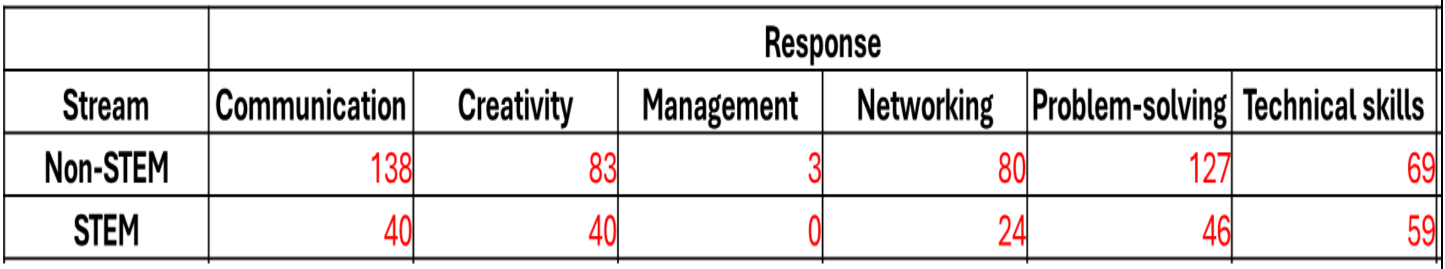
```
	2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 9.7917, df = 1, p-value = 0.0008765
alternative hypothesis: greater
95 percent confidence interval:
 0.06942282 1.00000000
sample estimates:
   prop 1    prop 2 
0.3194444 0.1414141 
```

Findings:
To investigate whether STEM and Non-STEM students differ in going for Research/Academia as a career option, a one-sided two-sample test for equality of proportions was conducted. Among STEM students, 23 out of 72 (approximately 31.94%) selected "Sports & Fitness", whereas among Non-STEM students, 28 out of 198 (approximately 14.14%) selected it. The hypothesis test yielded a chi-squared value of 9.7917 with 1 degree of freedom and a p-value of 0.0008765. Since the p-value is much smaller than the significance level of 0.05, we reject the null hypothesis.

Thus, there is strong statistical evidence that Non-STEM students are more likely than STEM students to go for Research/Academia. The one-sided 95% confidence interval for the difference in proportions ranged from 0.069 to 1, further supporting the finding that the proportion is significantly higher among Non-STEM students compared to STEM students.

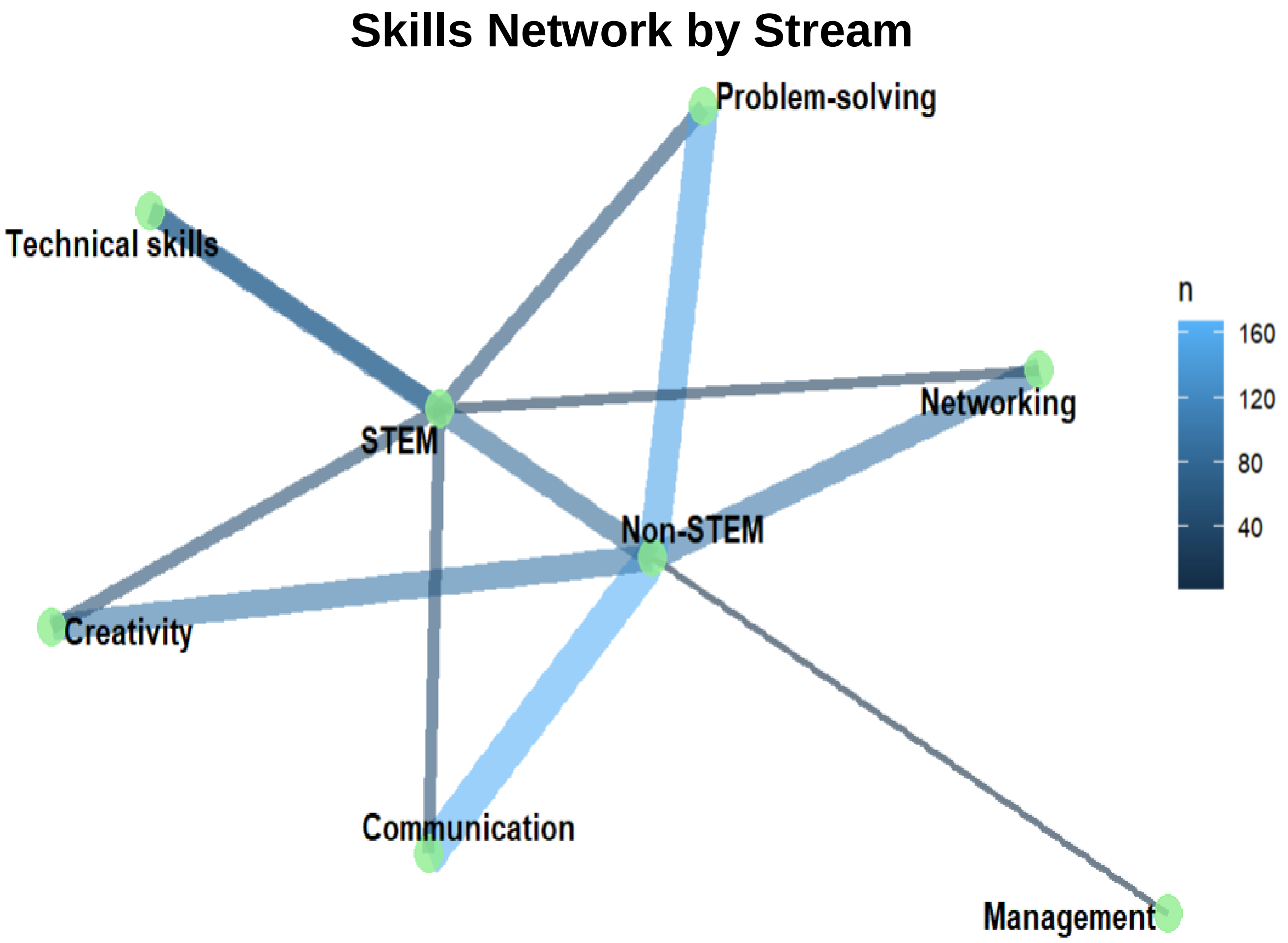
Interpretation: This may reflect the broader range of perceived opportunities in Non-STEM disciplines, which are often more flexible in career transitions. The findings also support Tomlinson's (2008) assertion that credentialism and employability shape student perceptions of career pathways.

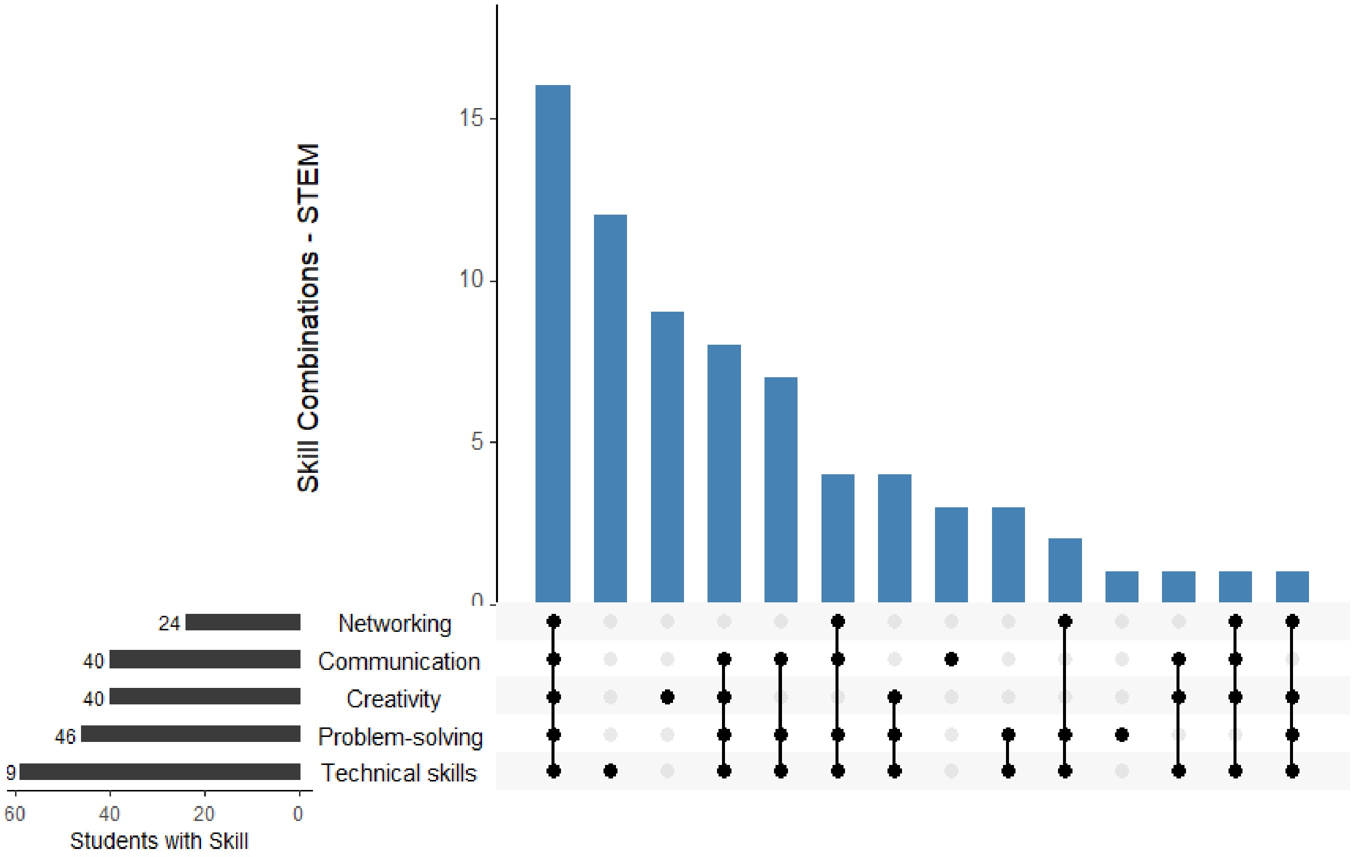
4.7 Skills Perceived as Valuable



	Response					
Stream	Communication	Creativity	Management	Networking	Problem-solving	Technical skills
Non-STEM	138	83	3	80	127	69
STEM	40	40	0	24	46	59

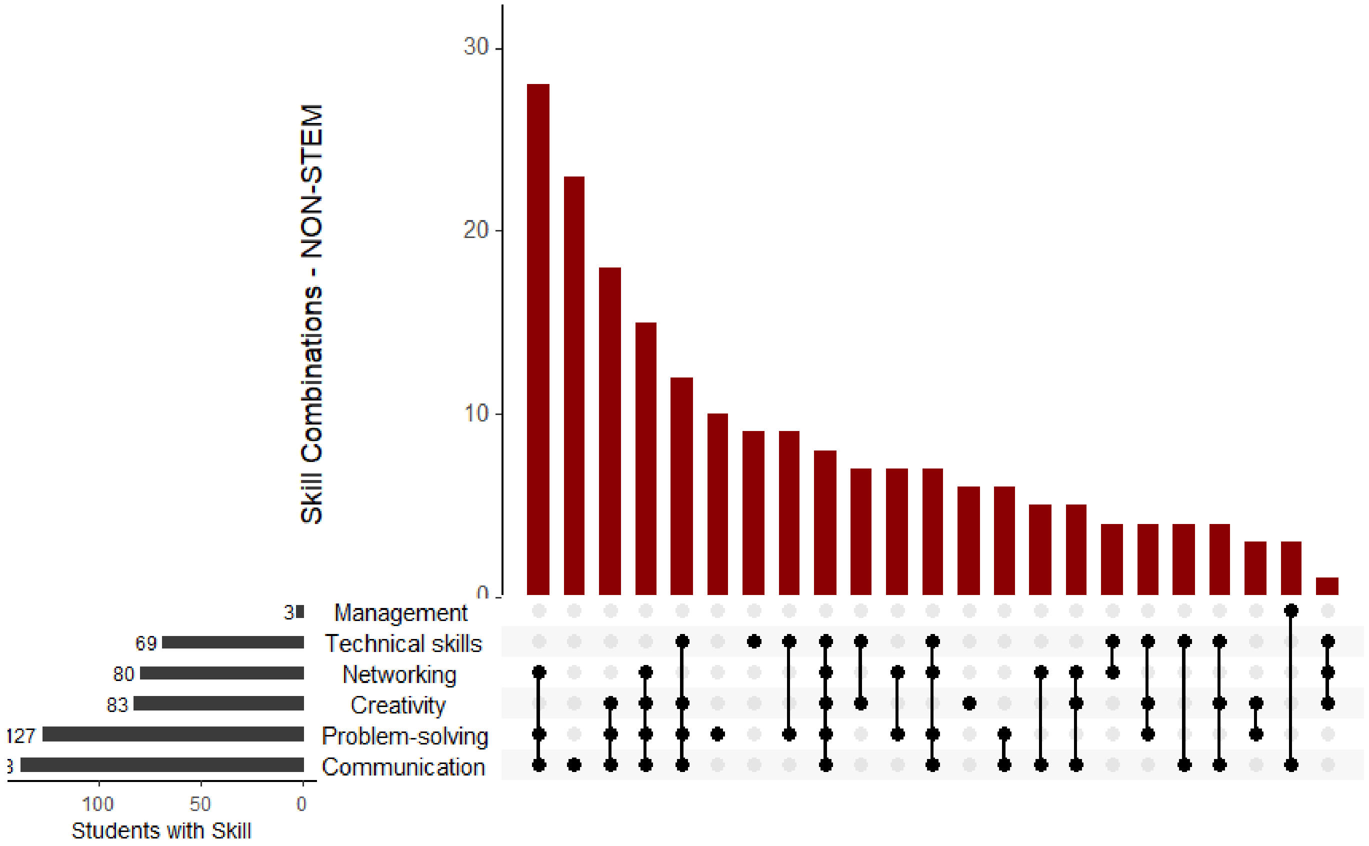
Here, the counts are from the multi-responses question where each respondent could choose more than one option irrespective of the preference of their Skills.


Skills Network by Stream



Combinations: The strongest combinations involve Technical Skills + Problem-solving. Communication and Creativity are present but secondary compared to technical expertise.

Insight: STEM students focus heavily on technical and problem-solving abilities, with soft skills like communication and networking being relatively less prioritized. This highlights a technical specialization mindset typical of STEM fields.



Combinations: Non-STEM students report strong combinations of soft skills like Communication, Problem-solving, Creativity, and Networking. Technical skills are less emphasized compared to soft skills.

Insight: Non-STEM careers seem to emphasize communication, creativity, and networking abilities much more than technical skills. Management is a rare skill among them, suggesting most are still early in their career journey.

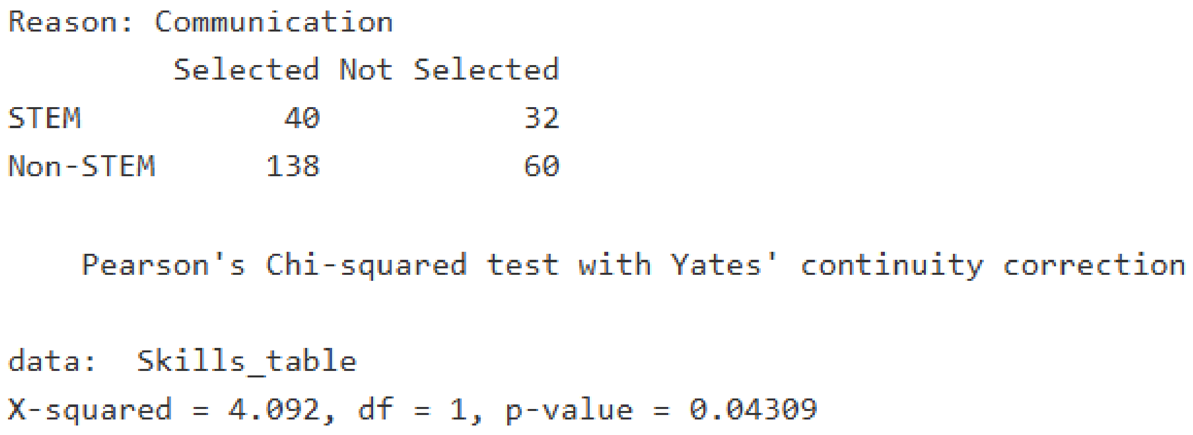
Question: Are there differences in skill prioritization between STEM and Non-STEM students?

Using Chi Square test of independence to check if there is any significant association between the Stream and the Skills Perceived as Valuable, directly on this frequency table would be inappropriate. Therefore, we modify our study and check if is there any association between skills & the choice of academic stream or not.

2x2 table is formed for every skill so that the rows become independent and henceforth Chi Square test is applied to check if that is a significant association between skills and stream or not.

Hypotheses:

- **H_0**: There is no association between stream and communication as a valued skills.
- **H_1**: There is an association between stream and communication as a valued skills.



```
Reason: Communication
        Selected Not Selected
STEM          40           32
Non-STEM     138           60

    Pearson's Chi-squared test with Yates' continuity correction

data:  Skills_table
X-squared = 4.092, df = 1, p-value = 0.04309
```

Findings:
For the Skills Communication, the analysis revealed notable differences between STEM and Non-STEM students. Among STEM students, 40 selected Communication as a Skill, while 32 did not. In contrast, among Non-STEM students, 138 selected it and 60 did not. A Pearson's Chi-squared test was conducted, yielding a chi-square statistic of 4.092 with 1 degree of freedom and a p-value of 0.04309. Since the p-value is less than 0.05, we conclude that there is a statistically significant association between the type of course (STEM vs Non-STEM) and citing Communication as a Skills Perceived as Valuable.

Skills	Chi-square	p-values	Significance
Communication	4.092	**0.04309**	Significant
Creativity	3.4278	**0.06411**	Non-significant
Management	0.15513	**0.6937**	Non-significant
Networking	0.83609	**0.3605**	Non-significant

Problem Solving	1.1077e-30	1	Non-significant
Technical Skills	45.101	1.871e-11	Significant

Interpretation: This finding reflects discipline-specific skill orientation and supports the notion of disciplinary epistemologies shaping student identities (Becher & Trowler, 2001).

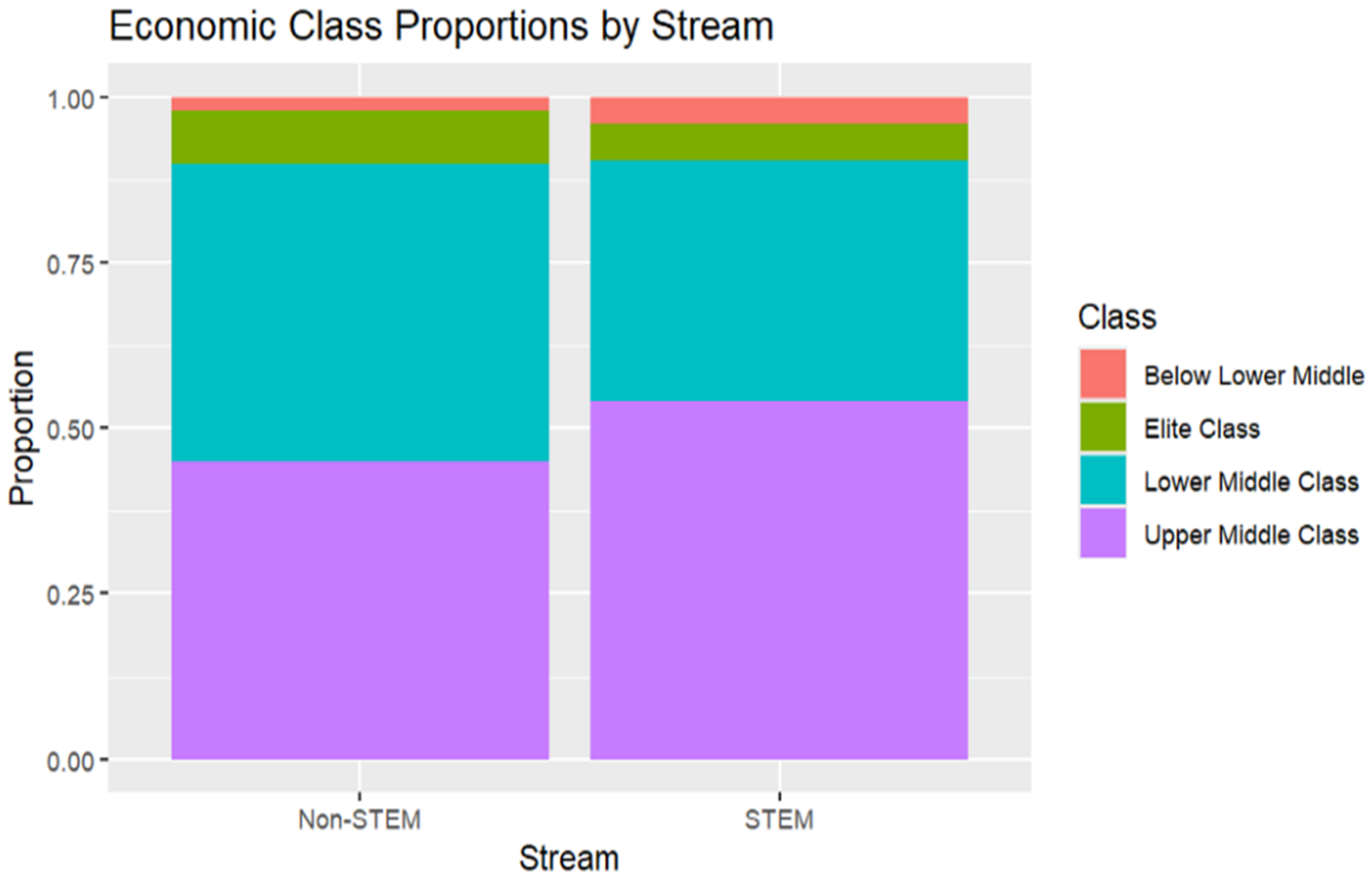
4.8 Socioeconomic Background

Stream	Below Lower Middle	Elite Class	Lower Middle Class	Upper Middle Class	Grand Total
Non-STEM	4	16	89	89	198
STEM	3	4	26	39	72
Grand Total	7	20	115	128	270

Hypotheses:

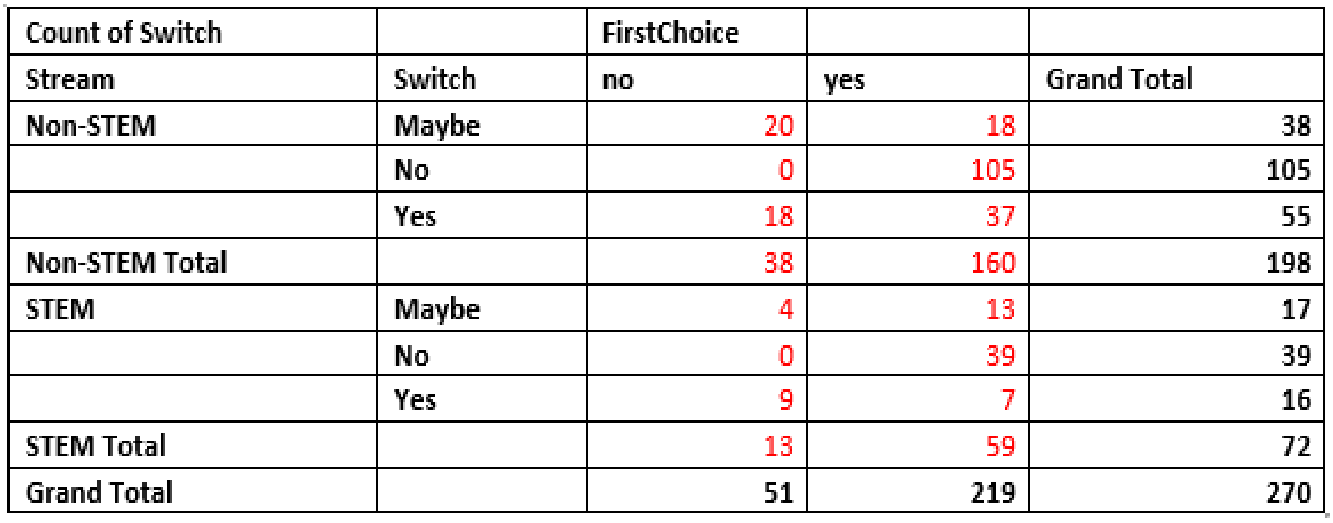
- **H_0**: There is no association between Stream and Economic class.
- **H_1**: There is an association between Stream and Economic class.

Findings:
Distribution across economic classes was even, and Chi-square analysis showed a **p-value > 0.05**.

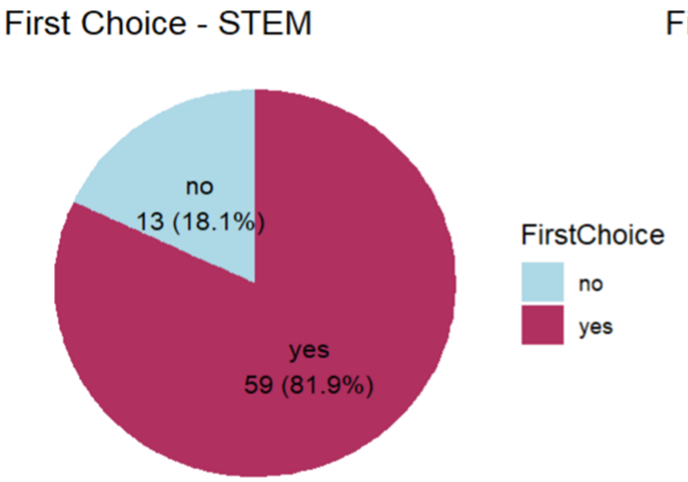

Economic Class Proportions by Stream

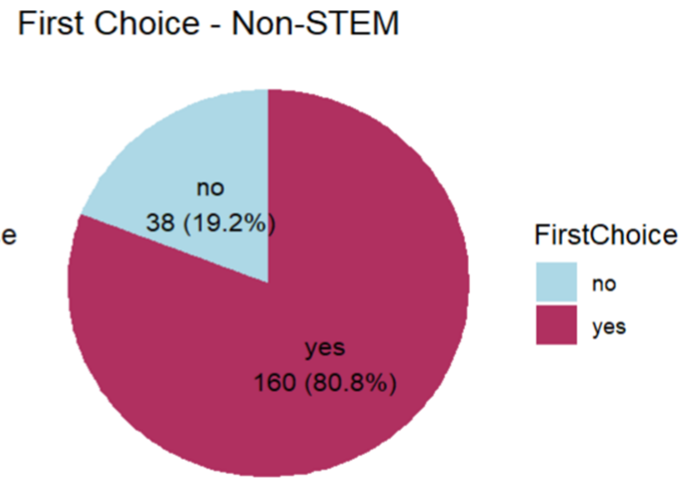
Interpretation: This suggests that **economic background does not act as a barrier** to stream choice in this sample. It may reflect improved access to education across socioeconomic strata in urban India, though this warrants deeper investigation.

4.9 First Choice and Switching



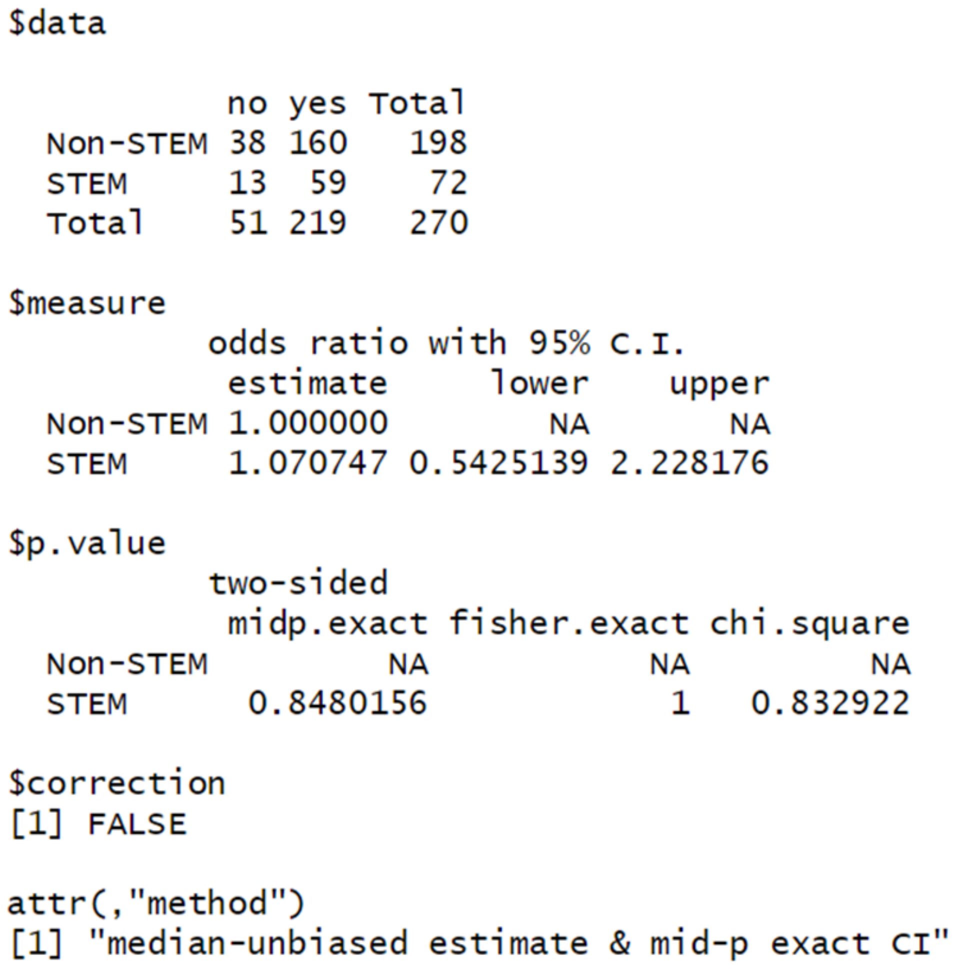
Count of Switch		FirstChoice		
Stream	Switch	no	yes	Grand Total
Non-STEM	Maybe	20	18	38
	No	0	105	105
	Yes	18	37	55
Non-STEM Total		38	160	198
STEM	Maybe	4	13	17
	No	0	39	39
	Yes	9	7	16
STEM Total		13	59	72
Grand Total		51	219	270





Hypotheses: For **First Choice**

- **H_0**: The odds of choosing a STEM stream as the first choice are equal to the odds of choosing a non-STEM stream (odds ratio = 1).
- **H_1**: The odds of choosing a STEM stream as the first choice differ from the odds of choosing a non-STEM stream (odds ratio ≠ 1).

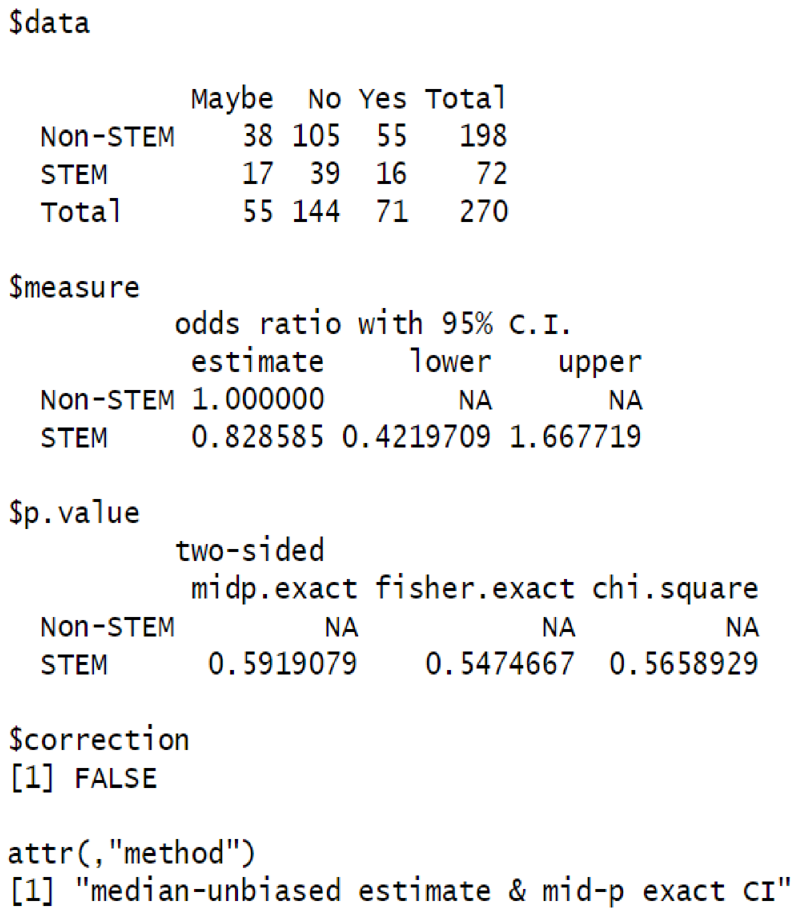


```
$data

          no yes Total
 Non-STEM 38 160   198
 STEM     13  59    72
 Total    51 219   270

$measure
         odds ratio with 95% C.I.
          estimate     lower    upper
 Non-STEM 1.000000        NA       NA
 STEM     1.070747 0.5425139 2.228176

$p.value
         two-sided
          midp.exact fisher.exact chi.square
 Non-STEM         NA           NA         NA
 STEM      0.8480156            1   0.832922

$correction
[1] FALSE

attr(,"method")
[1] "median-unbiased estimate & mid-p exact CI"
```

Findings:
For **First Choice**, the odds ratio for STEM was 1.071, with a 95% CI of 0.543 to 2.228 → **Not significant**

Hypotheses: For **Switching**

- **H_0**: The odds of switching streams are the same for STEM and non-STEM students (odds ratio = 1).
- **H_1**: The odds of switching streams differ between STEM and non-STEM students (odds ratio ≠ 1).



```
$data

          Maybe  No Yes Total
  Non-STEM    38 105  55   198
  STEM        17  39  16    72
  Total       55 144  71   270

$measure
          odds ratio with 95% C.I.
           estimate     lower    upper
  Non-STEM 1.000000        NA       NA
  STEM     0.828585 0.4219709 1.667719

$p.value
          two-sided
           midp.exact fisher.exact chi.square
  Non-STEM         NA           NA         NA
  STEM      0.5919079    0.5474667  0.5658929

$correction
[1] FALSE

attr(,"method")
[1] "median-unbiased estimate & mid-p exact CI"
```

Findings:
For **Switching**, the odds ratio was 0.8285 (95% CI: 0.421–1.667) → **Also not significant**.

Interpretation: The choice of stream appears **largely incidental or constrained**, rather than driven by strong personal preference. This aligns with Côté and Allahar’s (2011) critique of modern education systems prioritizing employability over student agency.

5. Associations

5.1 Cramer’s V association

Cramér’s V is a statistic used to measure the strength of association between two nominal categorical variables. It is an extension of the Chi-square test of independence, which tells whether two variables are related—but not how strong that relationship is. Cramér’s V fills that gap by quantifying the effect size of the relationship.

Formula:

$$V = \sqrt{\frac{x^2}{n\,(k-1)}}$$

Where:

X^2 = Chi-square statistic
n = Total number of observations
k = The smaller of the number of rows or columns in the contingency table

Cramér’s V ranges from 0 to 1, where:

0 = No association between variables
1 = Perfect association

Values closer to 1 indicate a stronger relationship, but there’s no universal benchmark—interpretation often depends on context and sample size.

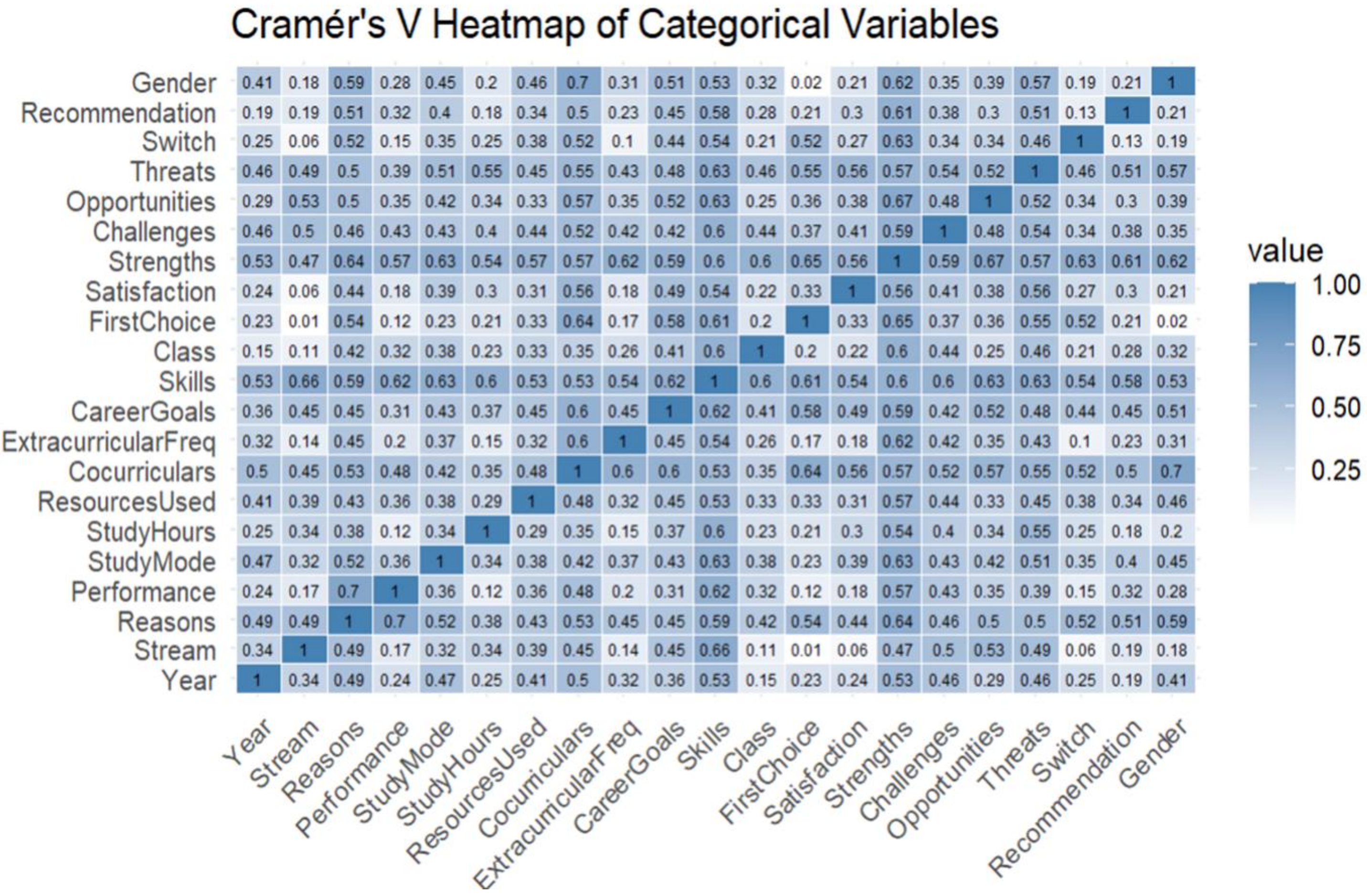
0.1 = Weak
0.3 = Moderate
0.5 = Strong

In my project comparing STEM and Non-STEM students across various categorical factors, I used the Chi-square test to detect significant relationships. However, while the Chi-square test told me whether a relationship exists, it didn’t tell me how strong that relationship was. For that, I used Cramér’s V.

Cramér’s V allowed us to quantify the strength of associations found in the Chi-square tests and compare multiple categorical relationships to see which factors are more strongly associated with students’ field of study.

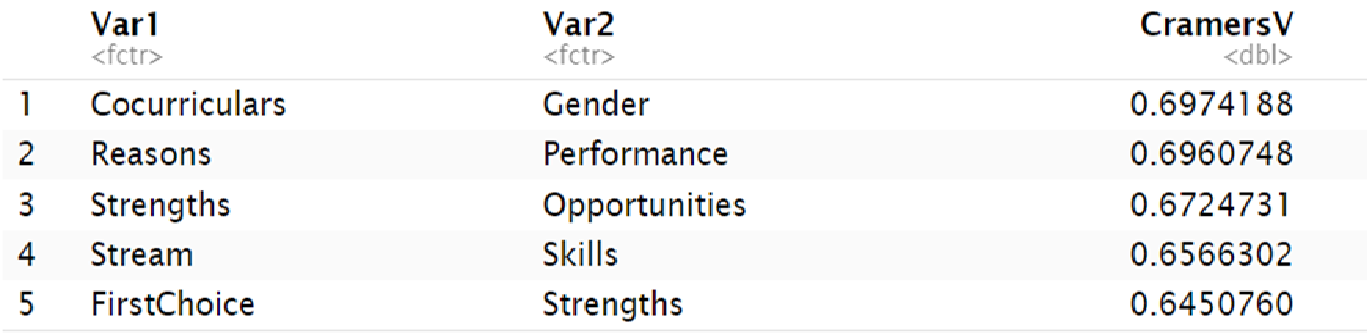
For example, a Cramér’s V value of 0.45 between course type and career goal indicates a moderate-to-strong association, suggesting course choice may significantly influence long-term aspirations.

5.2 Associations



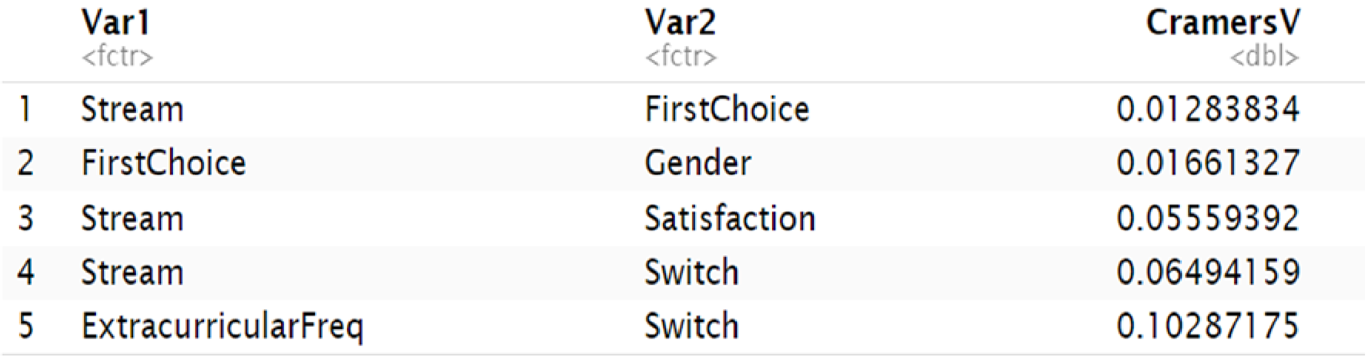
This is a heatmap which shows the strength of associations between all the categorical variables. But as it is difficult to point out the associations, here are the top 5 and bottom 5 by strength of Cramer’s V association.

Top 5:



	Var1	Var2	CramersV
	<fctr>	<fctr>	<dbl>
1	Cocurriculars	Gender	0.6974188
2	Reasons	Performance	0.6960748
3	Strengths	Opportunities	0.6724731
4	Stream	Skills	0.6566302
5	FirstChoice	Strengths	0.6450760

Bottom 5:



	Var1	Var2	CramersV
	<fctr>	<fctr>	<dbl>
1	Stream	FirstChoice	0.01283834
2	FirstChoice	Gender	0.01661327
3	Stream	Satisfaction	0.05559392
4	Stream	Switch	0.06494159
5	ExtracurricularFreq	Switch	0.10287175

High association between Co-curriculars and Gender may suggest that Males and Females do have a preference for cocurricular activities. Infact, reasons for choosing the field of study does affect the performance.

There is low association between Stream and the variable that a student is studying their field as their first choice or not. This is likely due to the fact that the proportions of students choosing their field of study is same irrespective of the stream.

6. SWOT Analysis

6.1 SWOT Analysis and Its Relevance to the Study

SWOT analysis is a strategic planning tool used to systematically identify and evaluate the internal and external factors that influence the success of an individual, organization, or field. The acronym SWOT stands for Strengths, Weaknesses, Opportunities, and Threats. Strengths and weaknesses are internal factors—elements within the system's control—while opportunities and threats are external factors, shaped by the environment in which the system operates.

In the context of this study, which aims to compare STEM and Non-STEM fields, SWOT analysis serves as a valuable framework for synthesizing the qualitative and quantitative data collected. By organizing findings into the four categories, the analysis can provide a clear and comprehensive overview of the characteristics, advantages, limitations, external prospects, and challenges faced by students in both academic streams.

Strengths refer to the inherent advantages or capabilities of each field, such as high employability rates for STEM students or diverse skill sets developed by Non-STEM students.

Weaknesses highlight internal limitations, such as a potential lack of interdisciplinary exposure for STEM students or limited technical training in Non-STEM programs.

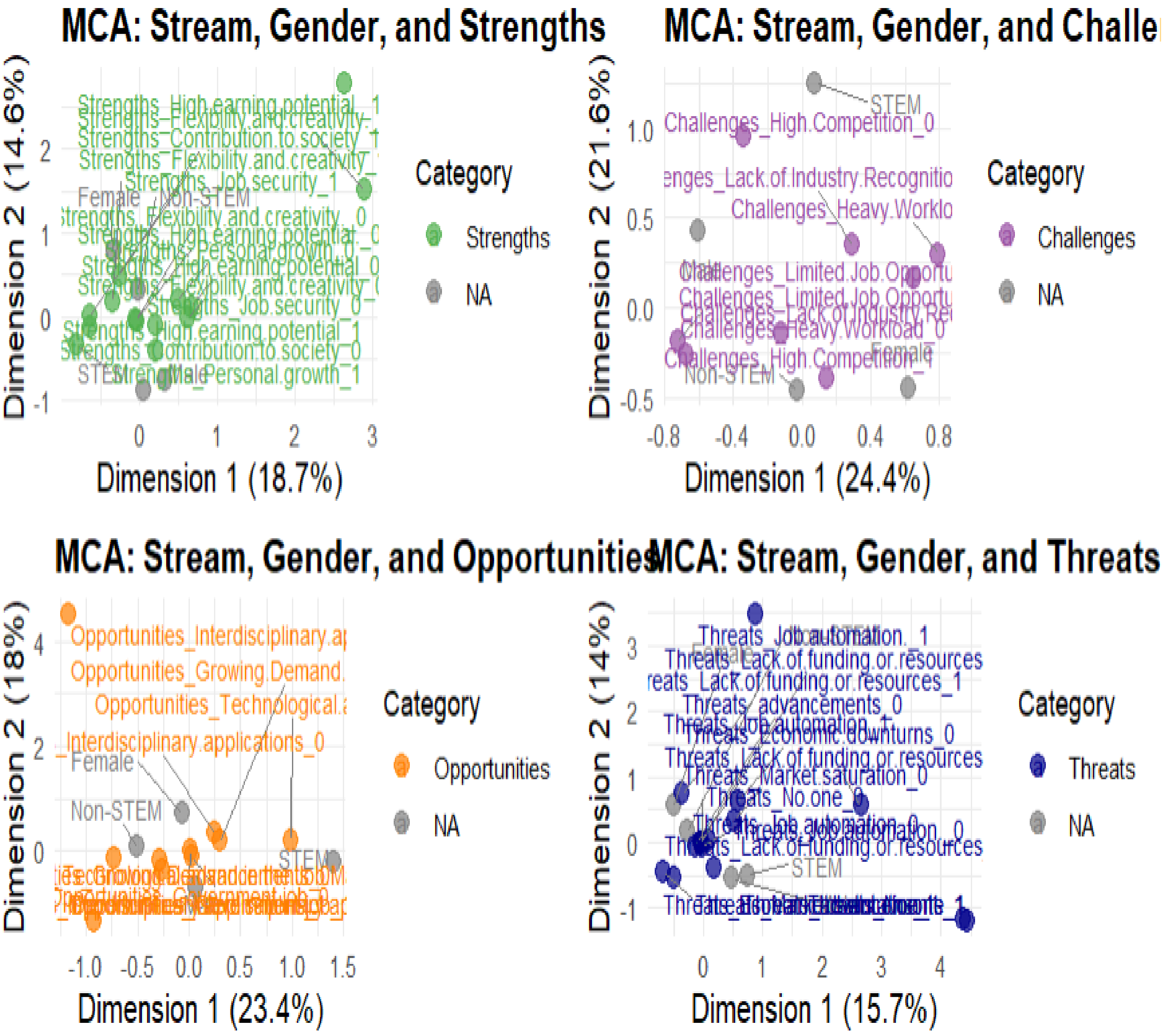
Opportunities encompass favorable external trends or conditions, like emerging industries creating new roles for STEM graduates or the rising demand for critical thinking and creativity from Non-STEM graduates.

Threats identify external challenges, including rapid technological changes that could render certain STEM skills obsolete, or job market saturation affecting Non-STEM career paths.

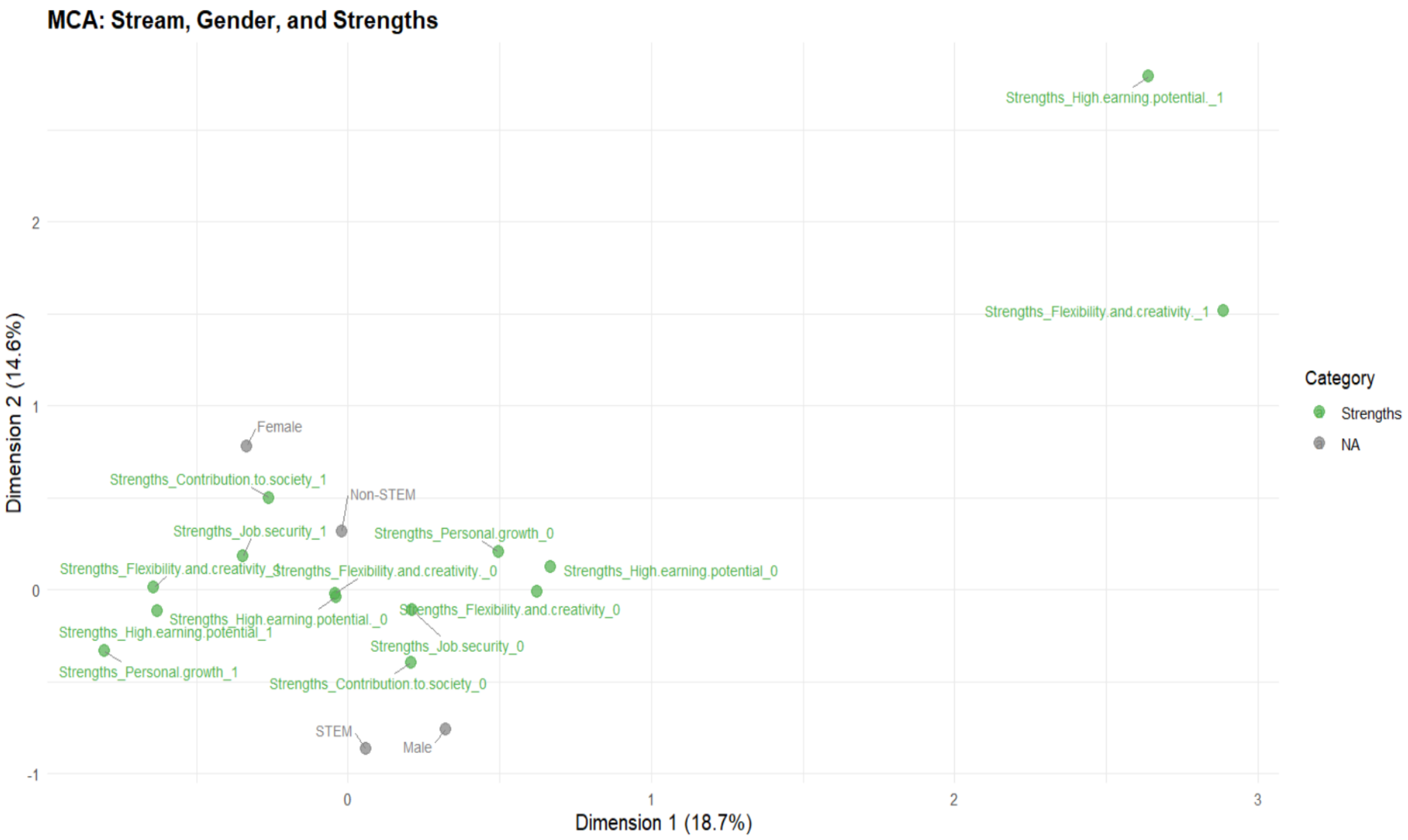
Integrating SWOT analysis into the project not only enhances the depth of the comparative study but also facilitates strategic recommendations. It helps stakeholders—including students, educators, and policymakers—understand the areas where interventions may be needed, recognize potential career opportunities, and make informed decisions based on a balanced evaluation of internal competencies and external conditions.

Thus, by applying SWOT analysis, this study moves beyond basic comparisons and provides a structured, strategic perspective on the academic and professional outlooks of students pursuing STEM and Non-STEM courses.

We perform the SWOT analysis using Multiple Correspondence Analysis (MCA), where variables having close associations are grouped together.



6.2 Strengths



Female and Non-STEM students cluster near attributes like Contribution to Society and Personal Growth, suggesting that these groups prioritize careers that offer meaningful societal impact and opportunities for personal development. Their responses are aligned closely with these strengths.

Male and STEM students are located closer to the center but slightly nearer to Job Security. This suggests that, while they do not show extreme preferences, they may place relatively more value on stable and secure career prospects compared to other strengths.

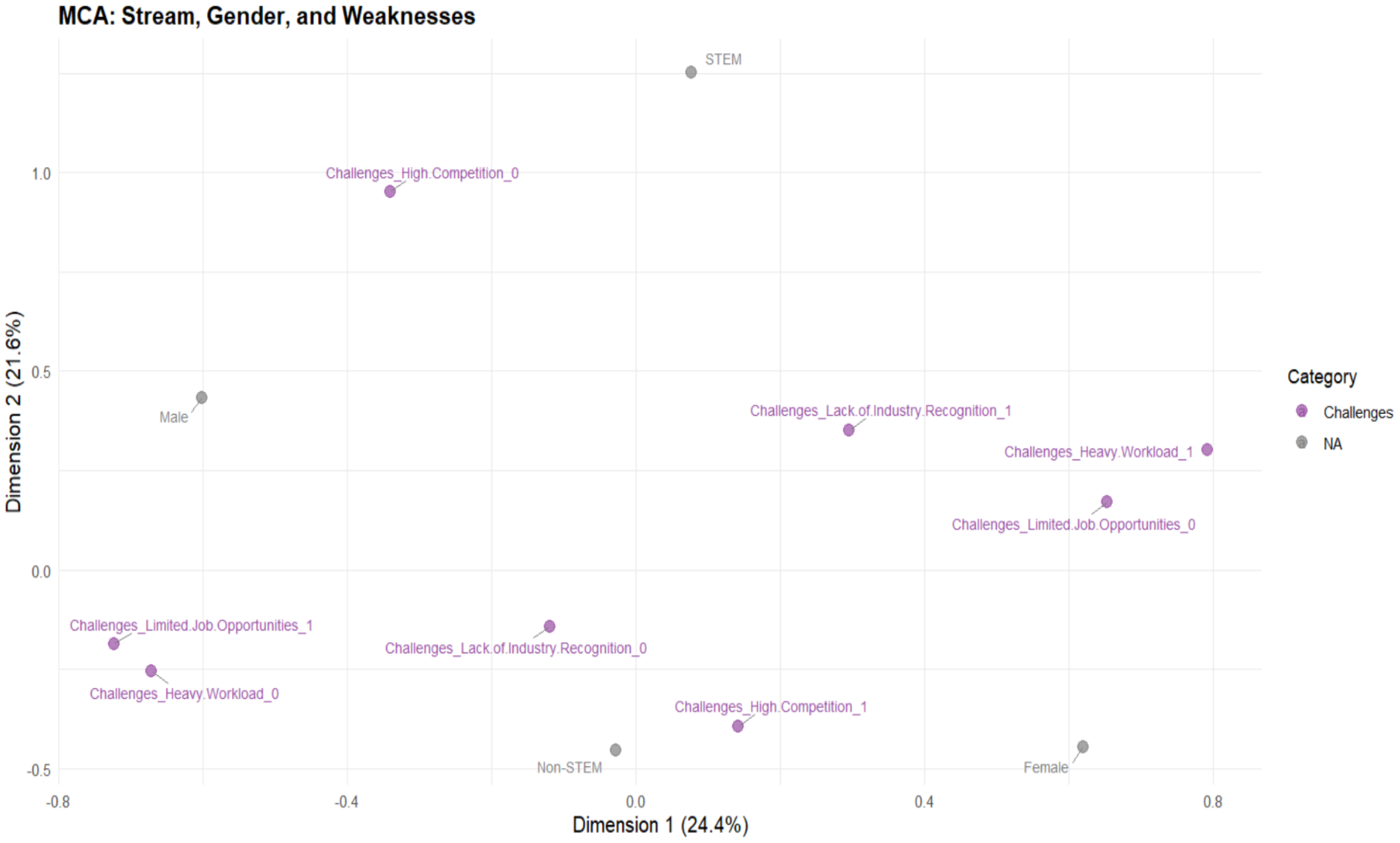
High Earning Potential and Flexibility and Creativity stand apart from the center and are associated with individuals who place a distinct emphasis on these attributes (indicated by the "_1" suffix). The separation suggests that valuing high income or flexibility is not strongly linked to gender or stream, but represents a distinct subgroup of students across categories.

Strengths marked "_0" (meaning not selected or less important) are generally clustered closer to the origin, indicating a more neutral or common distribution among respondents.

The greater spread of points for strengths like Flexibility and Creativity and High Earning Potential along Dimension 1 and 2 indicates higher variability in how different students prioritize these attributes.

In summary, the MCA shows that while some career motivations such as job security are commonly valued across groups, others such as societal contribution, personal growth, and high earning potential reveal clear distinctions based on gender and academic background. This highlights important differences in career aspirations between STEM and Non-STEM students and between male and female students.

6.3 Weaknesses



STEM and Male are positioned toward the top of the graph, closer to Challenges_High_Competition_0 (meaning they do not strongly perceive high competition as a challenge).

This suggests that male and STEM students feel relatively less concerned about high competition compared to others.

Found toward the lower-right of the graph, near Challenges_High_Competition_1, Challenges_Heavy_Workload_1, and Challenges_Lack_of_Industry_Recognition_1.

This suggests that female and Non-STEM students perceive more challenges, particularly: High competition, Heavy workloads, Lack of recognition from the industry.

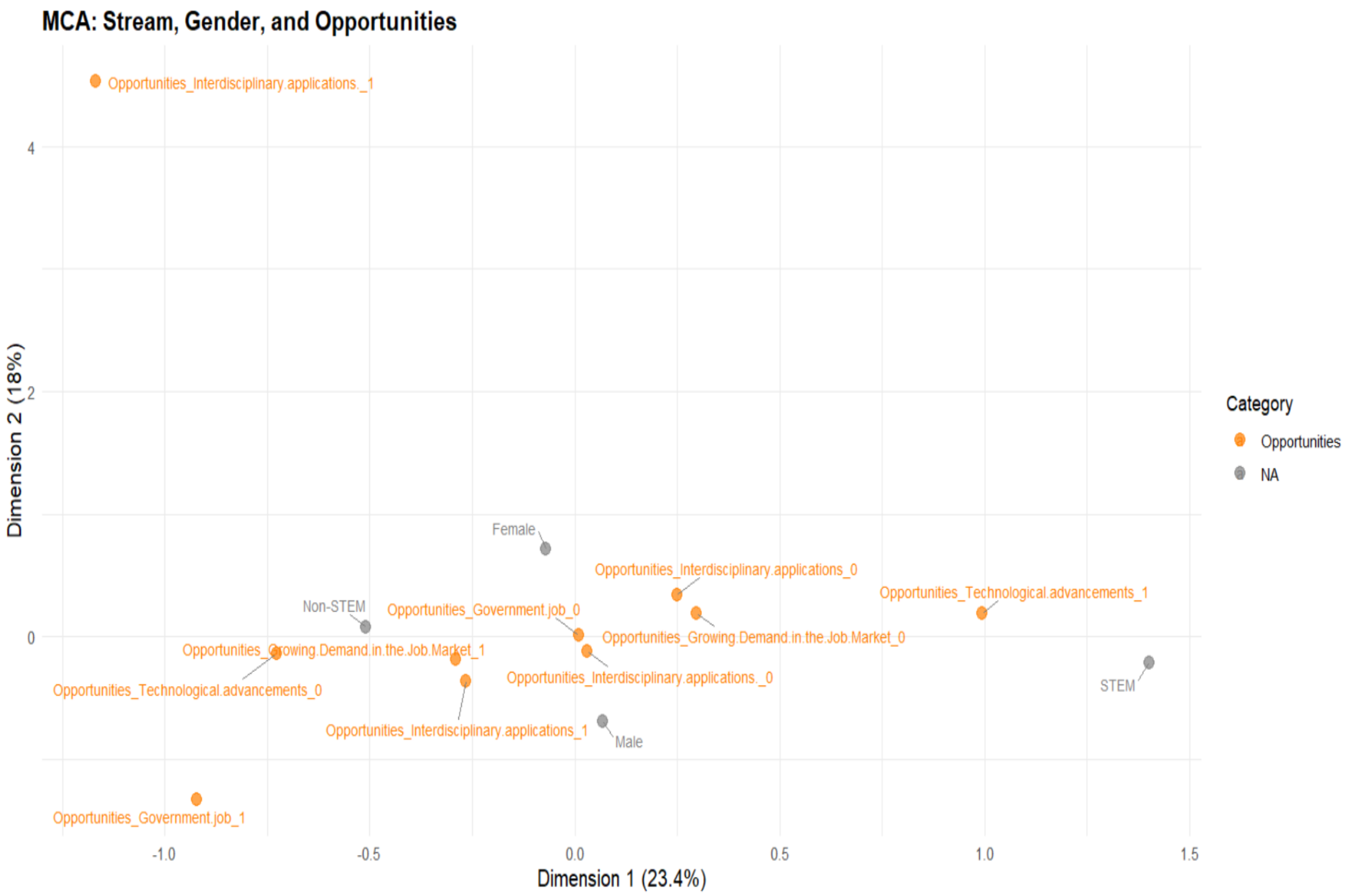
Challenges_Limited_Job_Opportunities and Challenges_Lack_of_Industry_Recognition cluster closer to Non-STEM and Female students. Meaning these challenges are more frequently perceived by Non-STEM and Female students.

Challenges_Heavy_Workload_1 is also nearby, reinforcing that workload stress is a bigger concern for these groups.

Some points, like Challenges_Limited_Job_Opportunities_0, are closer to the center — suggesting a general spread (not strongly associated with one gender or stream).

The separation between "_0" (no challenge perceived) and "_1" (challenge perceived) indicates clear differences among student groups in how they perceive the career environment.

6.4 Opportunities



Positioned toward the right side of the graph, closer to opportunities like Technological Advancements and Growing Demand in the Job Market.

This indicates that male STEM students are more optimistic or aware of the growing demand and technology-driven opportunities.

Found on the left side, closer to Opportunities_Government_Job_1 and Opportunities_Interdisciplinary_Applications_0.

Suggests that female and Non-STEM students are more inclined towards or value government job opportunities, but show less emphasis on interdisciplinary applications compared to their STEM counterparts.

Opportunities_Technological_Advancements_1 and Opportunities_Growing_Demand_in_the_Job_Market_1 are pulled toward the STEM and Male side.

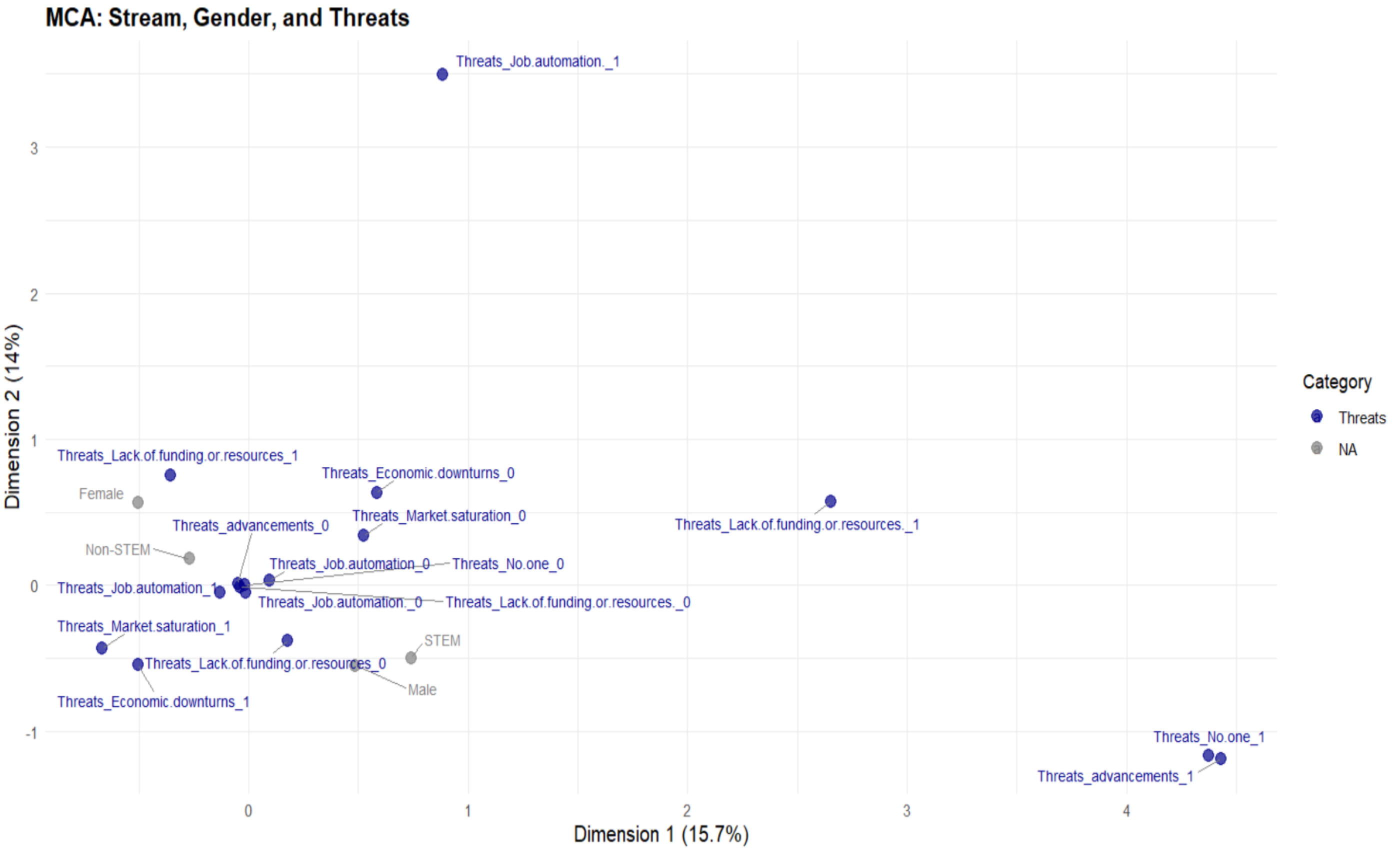
Opportunities_Government_Job_1 is located closer to Non-STEM and Female side, showing government sector as a more perceived opportunity for these students.

The positioning of Opportunities_Interdisciplinary_Applications_1 quite high along Dimension 2 suggests a separate group of students (not strictly gender or stream-bound) who highly recognize interdisciplinary opportunities, but not necessarily tied directly to gender or stream differences.

The cluster at the center (with several "_0" suffixes) means that many students, regardless of background, did not strongly emphasize certain opportunities. STEM students (especially males) are more linked with emerging technological and market-driven opportunities. Non-STEM students (especially females) show a traditional or stable preference toward government jobs.

Interdisciplinary opportunities are recognized by some, but not strongly tied to mainstream gender or stream patterns.

6.5 Threats



Threats_Job_Automation_1 is located far up along Dimension 2. This means that fear of job automation is a major concern for a significant portion of respondents, more distinctly than other threats. Those who selected job automation as a threat are clustered separately, indicating it is a standout concern.

Non-STEM and Female students are located towards the left side.

STEM and Male students are towards the right but closer to the center, suggesting less polarized perception of threats compared to Non-STEM/Female students. Females and Non-STEM students seem more aligned with concerns like lack of funding or economic downturns.

Threats_Lack_of_Funding_or_Resources and Threats_Economic_Downturns are clustered in the center-left part.

This means that concerns about lack of funding and economic instability are shared relatively equally across genders and streams, but slightly more among Non-STEM/Female students.

Threats_Market_Saturation appears near the center, meaning it’s a common threat felt moderately by all groups.

Threats_No_one_1 is located to the far bottom-right, away from most points. It suggests that some students feel that no major threat exists, and this belief is least common.

Threats_Advancements_1 is also placed far to the right. It shows that some students see technological advancements not as a threat but possibly even an opportunity.

Automation is perceived as the biggest, distinct threat, especially for some groups (more likely Non-STEM and possibly females). Lack of resources and economic uncertainty are generalized threats shared across groups. Market saturation is moderately felt.

A minority believes no real threats exist.

7. Limitations of the study and further scope

7.1 Limitations of the study

Despite the comprehensive approach adopted in this study, several limitations must be acknowledged.

First, the sample size, although sufficient for preliminary analysis, may not be large enough to generalize the findings across the broader population of STEM and Non-STEM students nationally or internationally. The participants were selected through convenience sampling, which introduces potential sampling bias and may not fully capture the diversity of experiences, backgrounds, and educational settings that exist among all students.

Furthermore, self-reported data was used extensively in the survey instrument, making the results vulnerable to biases such as social desirability bias, recall bias, and subjective interpretation of questions by respondents. Some variables, such as "career goals" or "coping strategies," are inherently subjective and could vary greatly in interpretation depending on individual experiences, which may limit the consistency and comparability of responses.

Additionally, the study focuses primarily on quantifiable aspects through statistical testing and SWOT analysis, potentially overlooking deeper qualitative nuances that interviews or open-ended survey responses might have captured. The temporal nature of the data also poses a limitation; the educational and employment landscapes are continually evolving, especially with the rapid advancement of technology and changes in global economic conditions, and thus, the findings represent a snapshot in time rather than a long-term trend.

Lastly, the scope of factors considered in distinguishing STEM and Non-STEM experiences was broad but not exhaustive; aspects such as cultural expectations, family background, and psychological factors were not explicitly measured and could be influential in shaping academic choices and career trajectories.

7.2 Scope for further research

Building on the insights and limitations of this study, there is substantial scope for further research to deepen and broaden the understanding of differences between STEM and Non-STEM student experiences.

Future studies could employ larger, more representative samples drawn from multiple regions or countries to enhance the generalizability of the results and to explore potential cultural, geographic, and institutional differences in academic and career experiences.

Incorporating a mixed-methods approach that combines quantitative surveys with qualitative interviews or focus groups would provide richer, more nuanced insights into students' motivations, challenges, and decision-making processes.

Longitudinal studies tracking students from the beginning of their undergraduate education into their early careers could offer valuable information on how aspirations, competencies, and outcomes evolve over time and how the realities of the job market interact with educational backgrounds.

Further research could also specifically investigate the role of emerging interdisciplinary fields, such as data science, environmental studies, and digital humanities, which increasingly blur the lines between traditional STEM and Non-STEM categories.

Comparative studies focusing on the effectiveness of different educational interventions—such as mentorship programs, career counseling, and skill development workshops—in enhancing student outcomes across both streams would be highly beneficial.

Moreover, future work could explore the intersectionality of factors such as gender, socioeconomic status, and ethnicity in shaping the STEM vs. Non-STEM divide, offering a more holistic and inclusive perspective.

Finally, as the global economy continues to evolve with advancements like AI, automation, and sustainability initiatives, ongoing research will be crucial to understanding how academic preparation in different fields must adapt to prepare students for the careers of the future.

7.3 Recommendations

Based on the findings and insights gathered through this study, several recommendations can be proposed to better support students in both STEM and Non-STEM fields and to bridge the gaps identified through analysis.

Firstly, institutions should consider implementing tailored career guidance programs early in the undergraduate journey to help students make informed academic and professional choices aligned with their strengths, interests, and the realities of the job market.

Special emphasis should be placed on exposing students from both streams to interdisciplinary opportunities, as the modern workforce increasingly demands a blending of technical expertise with soft skills like critical thinking, communication, and creativity. STEM programs could benefit from integrating more humanities and social sciences courses to foster well-rounded professionals, while Non-STEM programs should consider incorporating basic technical skills, data literacy, and digital competencies to enhance their students' employability.

Additionally, initiatives aimed at improving access to mentorship, internships, and research opportunities should be expanded, particularly for underrepresented groups, to ensure equitable development of career capital.

Universities and colleges are also encouraged to regularly update their curriculum in collaboration with industry experts to ensure that the skills imparted remain relevant in a rapidly changing economic and technological landscape.

Finally, fostering a campus culture that values diverse career paths—not just those perceived as traditionally lucrative or prestigious—would help reduce unnecessary stigma associated with certain fields and empower students to pursue their genuine passions with confidence. Policymakers and educators must work collaboratively to create an educational ecosystem that supports flexibility, innovation, and lifelong learning, preparing students from both STEM and Non-STEM backgrounds to thrive in a complex, interconnected, and dynamic global environment.

8. References

8.1References

1. Gottfried, M. A., Polikoff, M. S., & Stein, M. L. (2025). *Why students pursue STEM careers: A review of motivations and barriers. Humanities and Social Sciences Communications, 12(1), Article 446.* https://www.nature.com/articles/s41599-025-04446-2

2. Kelley, T. R. (2024). *Understanding career aspirations through student perceptions in STEM education. International Journal of STEM Education, 11(1), Article 66.* https://stemeducationjournal.springeropen.com/articles/10.1186/s40594-024-00466-7

3. Karabenick, S. A., & Zusho, A. (2022). *STEM students' self-regulated learning and academic performance: A comparative analysis. ERIC.* https://files.eric.ed.gov/fulltext/EJ1325557.pdf

4. Chauhan, P. (2024). *STEM vs Non-STEM career outcomes: A data-driven comparison. ResearchGate*. https://www.researchgate.net/publication/384899947_STEM_vs_Non-STEM_Career_Outcomes

5. King, H. (2018, September 30). *Girls are just as good as boys at STEM — so why aren't there more women in the field? Axios.* https://www.axios.com/2018/09/30/stem-career-paths-students-boys-girls

6. Agarwal, P. (2016). *Indian Higher Education: Envisioning the Future*. SAGE Publications.

7. Becher, T., & Trowler, P. R. (2001). *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. Open University Press.

8. Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education*. Greenwood.

9. Côté, J. E., & Allahar, A. L. (2011). *Lowering Higher Education: The Rise of Corporate Universities and the Fall of Liberal Education*. University of Toronto Press.

10. Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives* (pp. 75–146). W.H. Freeman.
11. Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM: Country comparisons*. Australian Council of Learned Academies.
12. Wang, M. T., & Degol, J. L. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, 33(4), 304–340.
13. World Economic Forum. (2023). *Global Gender Gap Report 2023*. WEF.
14. All India Survey on Higher Education (AISHE). (2021). *Ministry of Education, Government of India*.

9. Appendix

9.1 Missing Data

~~~
```{r}
missing_summary = data.frame(
  ColumnName = names(data),
  MissingCount = colSums(is.na(data))
)
print(missing_summary)
```
~~~

9.2 Gender, Performance and Economic Status Proportions Across Stream

~~~
```{r}
gender_stream_table = table(data$Gender, data$Stream)


chisq.test(gender_stream_table)


ggplot(data, aes(x = Stream, fill = Gender)) +
  geom_bar(position = "fill") +
  scale_fill_manual(values = c("Female" = "maroon", "Male" = "lightblue"))+
  labs(title = "Gender Proportions Across Streams",
       x = "Stream", y = "Proportion") +
  theme_classic()

```
~~~

9.3 Reasons, Co-curricular, Resources Used, Career goals and skills proportions Across Streams

~~~
```{r}
stem_selected <- c(48, 12, 6, 13, 46)
nonstem_selected <- c(88, 42, 52, 59, 113

stem_total <- 72
nonstem_total <- 198


reasons <- c("Parental Pressure", "Peer Influence", "Job Prospects",
"Interest", "Others")
for (i in 1:length(reasons)) {
  reason_table <- matrix(c(stem_selected[i], stem_total - stem_selected[i],
                           nonstem_selected[i], nonstem_total -
nonstem_selected[i]),

                         nrow = 2,
                         byrow = TRUE)
  colnames(reason_table) <- c("Selected", "Not Selected")
  rownames(reason_table) <- c("STEM", "Non-STEM")
  cat("\n\nReason:", reasons[i], "\n")
  print(reason_table)

  test_result <- chisq.test(reason_table)
~~~

~~~
  print(test_result)


  if (any(test_result$expected < 5)) {
    cat("Small expected count detected. Running Fisher's Exact Test
instead.\n")
    print(fisher.test(reason_table))
  }
}
```

```{r}
success <- c(48, 88)
total <- c(72, 198)
prop.test(x = success, n = total, alternative = "greater")
```
~~~

9.4 Network Graphs

~~~
```{r}
library(igraph)
edge_list = data %>% separate_rows(ResourcesUsed, sep = ", ") %>%
dplyr::select(Stream, ResourcesUsed)
graph = graph_from_data_frame(edge_list, directed = FALSE)
plot(graph, vertex.color = "lightblue")

library(ggraph)
library(tidyverse)
career_edges = data %>%
  separate_rows(CareerGoals, sep = ", ") %>%
  count(Stream, CareerGoals)

ggraph(career_edges, layout = "fr") +
  geom_edge_link(aes(width = n, color = n), alpha = 0.6) +
  geom_node_point(size = 5, color = "skyblue", alpha = 0.8) +
  geom_node_text(aes(label = name), repel = TRUE, color = "black", fontface
= "bold") +

  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    plot.margin = margin(20, 20, 20, 20)  # Add margin around the plot
  ) +
  ggtitle("Career Goals Network by Stream")
```
~~~

9.5 Pie Chart

~~~
```{r}
stream_choice_table = table(data$Stream, data$FirstChoice)
chisq.test(stream_choice_table)
stem_data = data %>%
  filter(Stream == "STEM") %>%
  count(FirstChoice) %>%
  mutate(percentage = round(n / sum(n) * 100, 1),
~~~

~~~
         label = paste0(FirstChoice, "\n", n, " (", percentage, "%)"))

nonstem_data = data %>%
  filter(Stream == "Non-STEM") %>%
  count(FirstChoice) %>%
  mutate(percentage = round(n / sum(n) * 100, 1),
         label = paste0(FirstChoice, "\n", n, " (", percentage, "%)"))

pie1 = ggplot(stem_data, aes(x = "", y = n, fill = FirstChoice)) +
  geom_bar(stat = "identity", width = 1) +
  scale_fill_manual(values = c("yes" = "maroon", "no" = "lightblue"))+
  coord_polar(theta = "y") +
  labs(title = "First Choice - STEM") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5),
size = 3.5) +
  theme_void()

pie2 = ggplot(nonstem_data, aes(x = "", y = n, fill = FirstChoice)) +
  geom_bar(stat = "identity", width = 1) +
  scale_fill_manual(values = c("yes" = "maroon", "no" = "lightblue"))+
  coord_polar(theta = "y") +
  labs(title = "First Choice - Non-STEM") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5),
size = 3.5) +
  theme_void()

library(patchwork)
pie1 + pie2
```
~~~

9.6 Goodman Kruskal Gamma and Odds Ratio

~~~
```{r}
GoodmanKruskalGamma(data$Performance, data$StudyHours)
library(epitools)
oddsratio(table(data$Stream, data$Switch))
```
~~~

9.7 Multiple Correspondence Analysis

~~~
```{r}
library(FactoMineR)
library(factoextra)
library(dplyr)
library(tidyr)
library(stringr)
library(ggplot2)
library(gridExtra)

# Function to split multiple responses and create dummy variables
create_dummies <- function(data, column_name) {
  # Extract unique responses from column
  all_responses <-
unique(unlist(strsplit(as.character(data[[column_name]]), ", ")))
  all_responses <- all_responses[!is.na(all_responses) & all_responses !=
""]
~~~

```
  # Create a matrix to hold dummy variables
  dummy_matrix <- matrix(0, nrow = nrow(data), ncol =
length(all_responses))
  colnames(dummy_matrix) <- paste0(column_name, "_",
make.names(all_responses))

  # Fill the dummy matrix
  for(i in 1:nrow(data)) {
    if(!is.na(data[[column_name]][i])) {
      responses <- unlist(strsplit(as.character(data[[column_name]][i]), ",
"))
      for(resp in responses) {
        col_idx <- which(all_responses == resp)
        if(length(col_idx) > 0) {
          dummy_matrix[i, col_idx] <- 1
        }
      }
    }
  }

  return(as.data.frame(dummy_matrix))
}

# Create dummy variables for SWOT categories
strengths_dummies <- create_dummies(data, "Strengths")
challenges_dummies <- create_dummies(data, "Challenges")
opportunities_dummies <- create_dummies(data, "Opportunities")
threats_dummies <- create_dummies(data, "Threats")

# Create combined data for overall MCA
# Assume we have Stream and Gender columns in the data
mca_data <- cbind(
  stream = data$Stream,
  gender = data$Gender,  # Include gender if available
  strengths_dummies,
  challenges_dummies,
  opportunities_dummies,
  threats_dummies
)
mca_data[] <- lapply(mca_data, as.factor)
mca_result <- MCA(mca_data, graph = FALSE)

create_enhanced_mca_plot <- function(mca_result, title, category_colors) {
  # Extract variable coordinates
  var_coords <- as.data.frame(mca_result$var$coord)
  var_coords$Variable <- rownames(var_coords)

  # Create category column for coloring
  var_coords$Category <- NA
  for(cat in names(category_colors)) {
    var_coords$Category[grep(paste0("^", cat), var_coords$Variable)] <- cat
  }

  # Plot with custom colors by category
  ggplot(var_coords, aes(x = Dim 1, y = Dim 2, color = Category, label =
Variable)) +
    geom_point(size = 3, alpha = 0.7) +
    geom_text_repel(size = 3, max.overlaps = 40, box.padding = 0.5,
segment.color = "grey50") +
    scale_color_manual(values = category_colors) +
    theme_minimal() +
    labs(title = title,
         x = paste0("Dimension 1 (", round(mca_result$eig[1,2], 1), "%)"),
         y = paste0("Dimension 2 (", round(mca_result$eig[2,2], 1), "%)"))
+
    theme(legend.position = "right",
          plot.title = element_text(size = 14, face = "bold"),
          axis.title = element_text(size = 12))
}

# Define color schemes that make it easy to differentiate categories
swot_colors <- c(
  "stream" = "#E41A1C",           # Red for Stream
  "gender" = "#377EB8",           # Blue for Gender
  "Strengths" = "#4DAF4A",        # Green for Strengths
  "Challenges" = "#984EA3",       # Purple for Challenges
  "Opportunities" = "#FF7F00",    # Orange for Opportunities
  "Threats" = "#FFFF33"           # Yellow for Threats
)

# Create individual SWOT MCA plots with clear color differentiation
# 1. Stream and Strengths
strengths_data <- cbind(stream = data$Stream, gender = data$Gender,
strengths_dummies)
strengths_data[] <- lapply(strengths_data, as.factor)
mca_strengths <- MCA(strengths_data, graph = FALSE)

# 2. Stream and Challenges
challenges_data <- cbind(stream = data$Stream, gender = data$Gender,
challenges_dummies)
challenges_data[] <- lapply(challenges_data, as.factor)
mca_challenges <- MCA(challenges_data, graph = FALSE)

# 3. Stream and Opportunities
opportunities_data <- cbind(stream = data$Stream, gender = data$Gender,
opportunities_dummies)
opportunities_data[] <- lapply(opportunities_data, as.factor)
mca_opportunities <- MCA(opportunities_data, graph = FALSE)

# 4. Stream and Threats
threats_data <- cbind(stream = data$Stream, gender = data$Gender,
threats_dummies)
threats_data[] <- lapply(threats_data, as.factor)
mca_threats <- MCA(threats_data, graph = FALSE)

# Create the plots with clear differentiation
strengths_plot <- create_enhanced_mca_plot(
  mca_strengths,
  "MCA: Stream, Gender, and Strengths",
  c("stream" = "#E41A1C", "gender" = "#377EB8", "Strengths" = "#4DAF4A")
)

challenges_plot <- create_enhanced_mca_plot(
  mca_challenges,
  "MCA: Stream, Gender, and Challenges",
  c("stream" = "#E41A1C", "gender" = "#377EB8", "Challenges" = "#984EA3")
)

opportunities_plot <- create_enhanced_mca_plot(
  mca_opportunities,
  "MCA: Stream, Gender, and Opportunities",
  c("stream" = "#E41A1C", "gender" = "#377EB8", "Opportunities" =
"#FF7F00")
)

threats_plot <- create_enhanced_mca_plot(
  mca_threats,
  "MCA: Stream, Gender, and Threats",
  c("stream" = "#E41A1C", "gender" = "#377EB8", "Threats" = "darkblue")
)

# Show individual plots
print(strengths_plot)
print(challenges_plot)
print(opportunities_plot)
print(threats_plot)

# Arrange all plots in a grid
grid.arrange(
  strengths_plot,
  challenges_plot,
  opportunities_plot,
  threats_plot,
  ncol = 2
)

# If you want to save the plots
ggsave("mca_strengths.png", strengths_plot, width = 10, height = 8)
ggsave("mca_challenges.png", challenges_plot, width = 10, height = 8)
ggsave("mca_opportunities.png", opportunities_plot, width = 10, height = 8)
ggsave("mca_threats.png", threats_plot, width = 10, height = 8)

# Save the combined grid
combined_plot <- arrangeGrob(
  strengths_plot,
  challenges_plot,
  opportunities_plot,
  threats_plot,
  ncol = 2
)
ggsave("mca_combined.png", combined_plot, width = 16, height = 12)
```

9.8 Upset Plot

~~~
```{r}
library(UpSetR)
library(tidyverse)
~~~

~~~
student_data <-
read.csv("C:\\Users\\ALOK\\Desktop\\project\\cleaneddata.csv")

cat("Unique stream values:\n")
print(unique(student_data$Stream))

student_data$Stream <- toupper(trimws(student_data$Stream))

# Filter to NON-STEM stream only
nonstem_data <- student_data %>% filter(Stream == "NON-STEM")

# DEBUG: Check how many NON-STEM rows
cat("\nNumber of NON-STEM rows found:", nrow(nonstem_data), "\n")

# Function to process multiple-response column (e.g., Reasons)
prepare_multiresponse_matrix <- function(data, column_name) {
  responses <- data[[column_name]]
  response_split <- strsplit(responses, ",\\s*")
  all_options <- unique(unlist(response_split))
  all_options <- all_options[all_options != "" & !is.na(all_options)]

  result <- data.frame(ID = 1:nrow(data))
  for (option in all_options) {
    result[[option]] <- sapply(response_split, function(x)
as.integer(option %in% x))
  }
  return(result)
}

reasons_matrix_nonstem <- prepare_multiresponse_matrix(nonstem_data,
"Reasons")
reasons_matrix_nonstem <-
reasons_matrix_nonstem[rowSums(reasons_matrix_nonstem[, -1]) > 0, ]

reason_cols_nonstem <- colnames(reasons_matrix_nonstem)[-1]
reasons_matrix_nonstem <- reasons_matrix_nonstem[, c("ID",
reason_cols_nonstem[colSums(reasons_matrix_nonstem[, reason_cols_nonstem])
> 0])]

if (nrow(reasons_matrix_nonstem) > 0 && ncol(reasons_matrix_nonstem) > 1) {
  upset(
    reasons_matrix_nonstem[, -1],
    sets = colnames(reasons_matrix_nonstem)[-1],
    order.by = "freq",
    mainbar.y.label = "Reason Combinations - NON-STEM",
    sets.x.label = "Students with Reason",
    show.numbers = TRUE,
    set_size.show = TRUE,
    number.angles = 0,
    text.scale = c(1.3, 1.3, 1, 1, 1.3, 1),
    main.bar.color = "darkblue",
    matrix.color = "black"
  )
} else {
  cat("No valid reason combinations found for NON-STEM stream.\n")
}
```
~~~