The Maharaja Sayajirao University of Baroda
Faculty of Science
Department of Statistics
Final Year B.Sc (2024-25)

COMPARATIVE STATISTICAL ANALYSIS OF STEM AND NON-STEM STUDENTS

Under the guidance of:
Dr. Rupal M. Shah
Ms. Shreya Mathur

A presentation by:
Puri Ankitkumar Kaushlendra
Kharva Shruti Harivadan
Acharya Matra Rajeshbhai
Jha Alokchandra Shreepati
Dheeraj Joshi

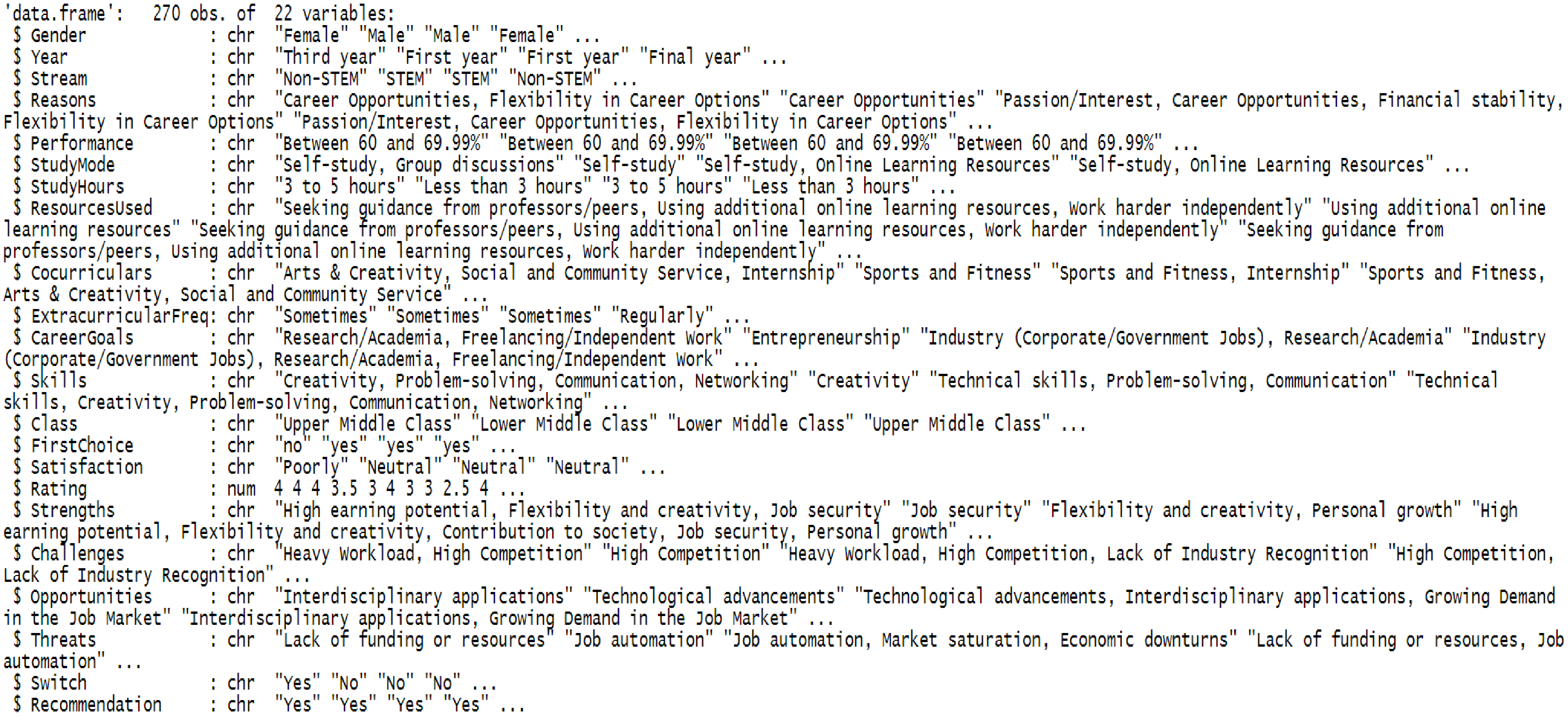
install.packages(“Picking_where_we_left”)

In Mid-Semester presentation, we displayed our findings based on pilot data.

Since then, we have collected **270** samples, out of which **198** are from **Non-STEM** students and **72** are from **STEM** students.

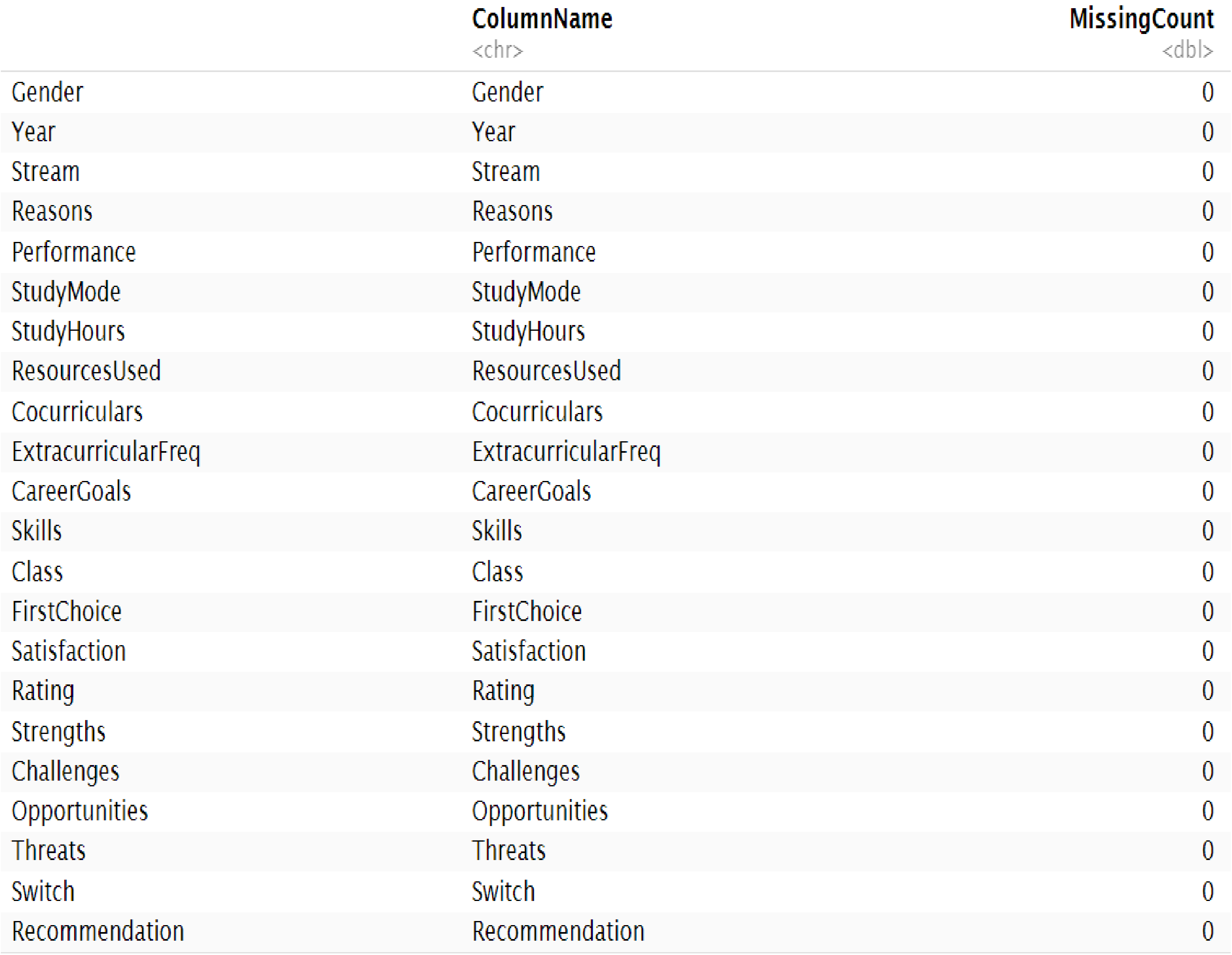
From the sample data, **49.26%** are from **females** and **50.74%** are from **males.**

?data



```
'data.frame':	270 obs. of  22 variables:
 $ Gender            : chr  "Female" "Male" "Male" "Female" ...
 $ Year              : chr  "Third year" "First year" "First year" "Final year" ...
 $ Stream            : chr  "Non-STEM" "STEM" "STEM" "Non-STEM" ...
 $ Reasons           : chr  "Career Opportunities, Flexibility in Career Options" "Career Opportunities" "Passion/Interest, Career Opportunities, Financial stability,
Flexibility in Career Options" "Passion/Interest, Career Opportunities, Flexibility in Career Options" ...
 $ Performance       : chr  "Between 60 and 69.99%" "Between 60 and 69.99%" "Between 60 and 69.99%" "Between 60 and 69.99%" ...
 $ StudyMode         : chr  "Self-study, Group discussions" "Self-study" "Self-study, Online Learning Resources" "Self-study, Online Learning Resources" ...
 $ StudyHours        : chr  "3 to 5 hours" "Less than 3 hours" "3 to 5 hours" "Less than 3 hours" ...
 $ ResourcesUsed     : chr  "Seeking guidance from professors/peers, Using additional online learning resources, Work harder independently" "Using additional online
learning resources" "Seeking guidance from professors/peers, Using additional online learning resources, Work harder independently" "Seeking guidance from
professors/peers, Using additional online learning resources, Work harder independently" ...
 $ Cocurriculars     : chr  "Arts & Creativity, Social and Community Service, Internship" "Sports and Fitness" "Sports and Fitness, Internship" "Sports and Fitness,
Arts & Creativity, Social and Community Service" ...
 $ ExtracurricularFreq: chr  "Sometimes" "Sometimes" "Sometimes" "Regularly" ...
 $ CareerGoals       : chr  "Research/Academia, Freelancing/Independent Work" "Entrepreneurship" "Industry (Corporate/Government Jobs), Research/Academia" "Industry
(Corporate/Government Jobs), Research/Academia, Freelancing/Independent Work" ...
 $ Skills            : chr  "Creativity, Problem-solving, Communication, Networking" "Creativity" "Technical skills, Problem-solving, Communication" "Technical
skills, Creativity, Problem-solving, Communication, Networking" ...
 $ Class             : chr  "Upper Middle Class" "Lower Middle Class" "Lower Middle Class" "Upper Middle Class" ...
 $ FirstChoice       : chr  "no" "yes" "yes" "yes" ...
 $ Satisfaction      : chr  "Poorly" "Neutral" "Neutral" "Neutral" ...
 $ Rating            : num  4 4 4 3.5 3 4 3 3 2.5 4 ...
 $ Strengths         : chr  "High earning potential, Flexibility and creativity, Job security" "Job security" "Flexibility and creativity, Personal growth" "High
earning potential, Flexibility and creativity, Contribution to society, Job security, Personal growth" ...
 $ Challenges        : chr  "Heavy Workload, High Competition" "High Competition" "Heavy Workload, High Competition, Lack of Industry Recognition" "High Competition,
Lack of Industry Recognition" ...
 $ Opportunities     : chr  "Interdisciplinary applications" "Technological advancements" "Technological advancements, Interdisciplinary applications, Growing Demand
in the Job Market" "Interdisciplinary applications, Growing Demand in the Job Market" ...
 $ Threats           : chr  "Lack of funding or resources" "Job automation" "Job automation, Market saturation, Economic downturns" "Lack of funding or resources, Job
automation" ...
 $ Switch            : chr  "Yes" "No" "No" "No" ...
 $ Recommendation    : chr  "Yes" "Yes" "Yes" "Yes" ...
```

?missing.values



	ColumnName	MissingCount
	<chr>	<dbl>
Gender	Gender	0
Year	Year	0
Stream	Stream	0
Reasons	Reasons	0
Performance	Performance	0
StudyMode	StudyMode	0
StudyHours	StudyHours	0
ResourcesUsed	ResourcesUsed	0
Cocurriculars	Cocurriculars	0
ExtracurricularFreq	ExtracurricularFreq	0
CareerGoals	CareerGoals	0
Skills	Skills	0
Class	Class	0
FirstChoice	FirstChoice	0
Satisfaction	Satisfaction	0
Rating	Rating	0
Strengths	Strengths	0
Challenges	Challenges	0
Opportunities	Opportunities	0
Threats	Threats	0
Switch	Switch	0
Recommendation	Recommendation	0

We find no missing observation in our data, and can safely move onto the next part.

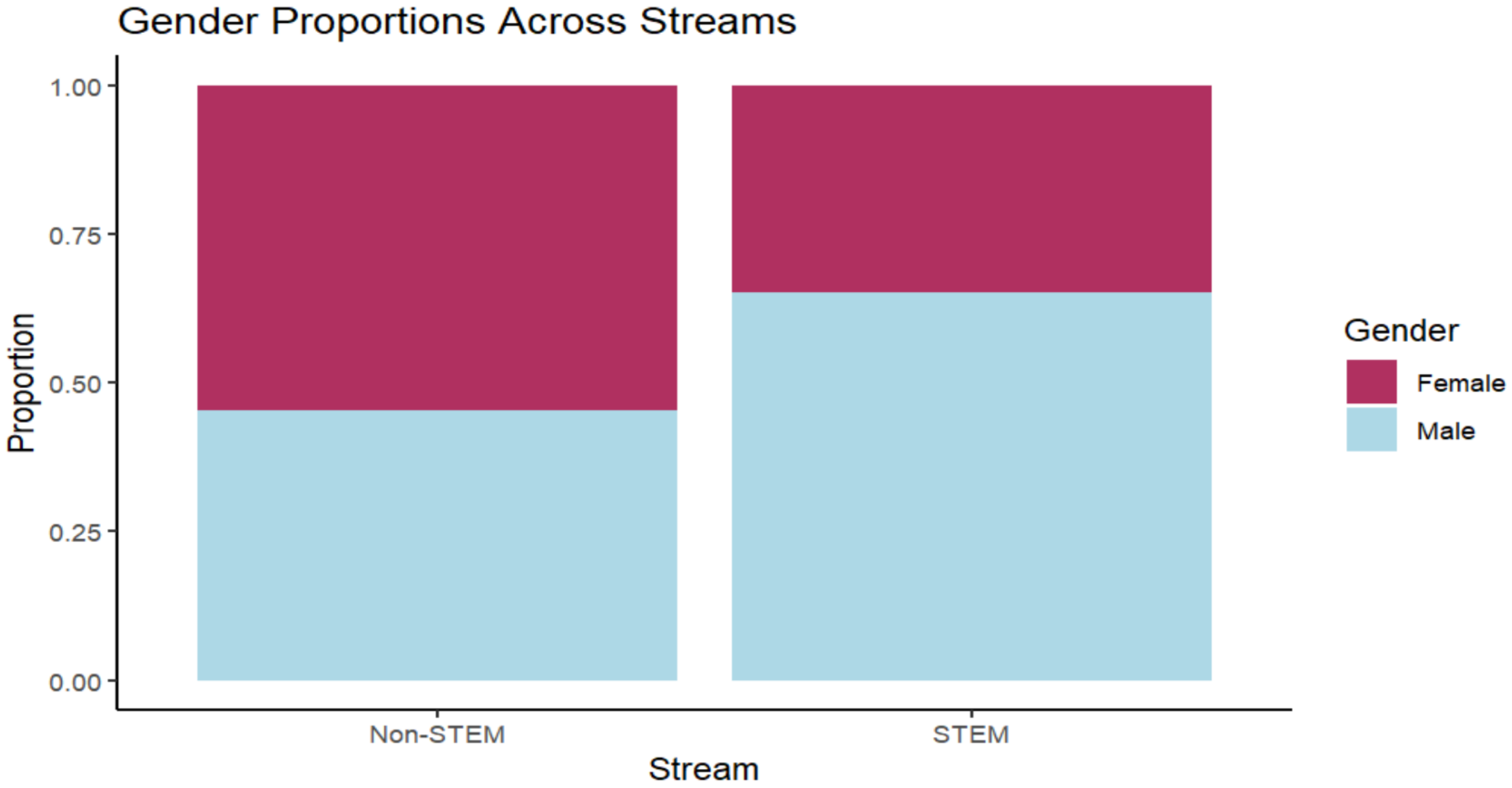
#Objective_1:

# To investigate the motivating factors and reasons behind choosing STEM and Non-STEM courses

data$Stream, data$Gender

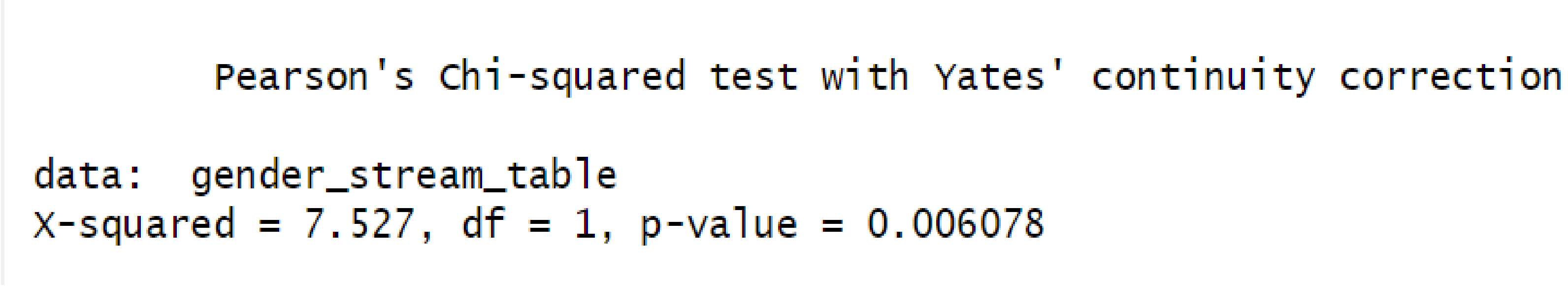
	Stream		
Gender	Non-STEM	STEM	Grand Total
Female	108	25	133
Male	90	47	137
Grand Total	198	72	270

Do we have a preference of the study fields based on gender?


Gender Proportions Across Streams

#Hypothesis:

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



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

data:  gender_stream_table
X-squared = 7.527, df = 1, p-value = 0.006078
```

As the p-value < 0.05, we reject the null hypothesis at 5% level of significance.

Therefore, we conclude that there exists an association between Gender and Stream of students.

The study published in Global Gender Gap Report 2023, states that women make up only 29.2 per cent of all STEM (science, technology, engineering and mathematics) workers across 146 countries.

Standardized Pearson Residual

- **Expected Values**

	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

- **Pearson's Residual**

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

All residuals are between -2 and +2, so none of the individual cells stand out strongly, but they do contribute together to the overall significant Chi-square result.

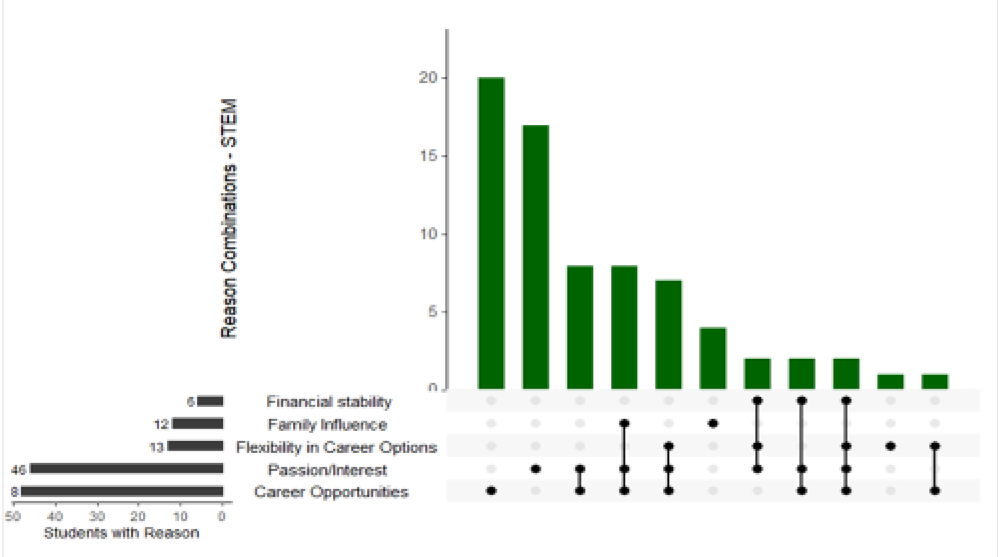
Females are more likely expected to choose Non-STEM while males are more likely expected to choose STEM.

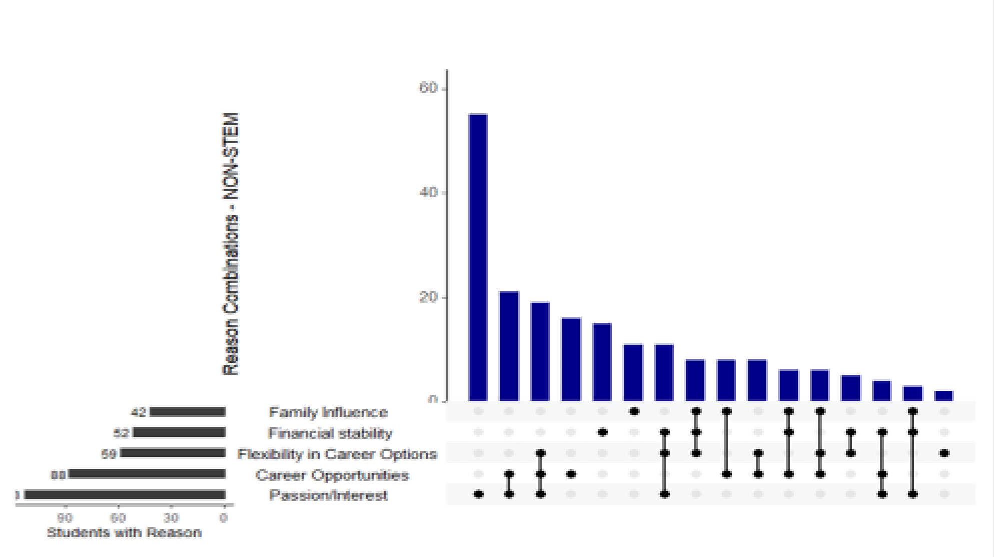
Although no individual cell exceeds ±1.96 (the usual cutoff for strong residuals at 5% level of significance), the overall pattern still supports a gender-stream association.

data$Stream, data$Reasons

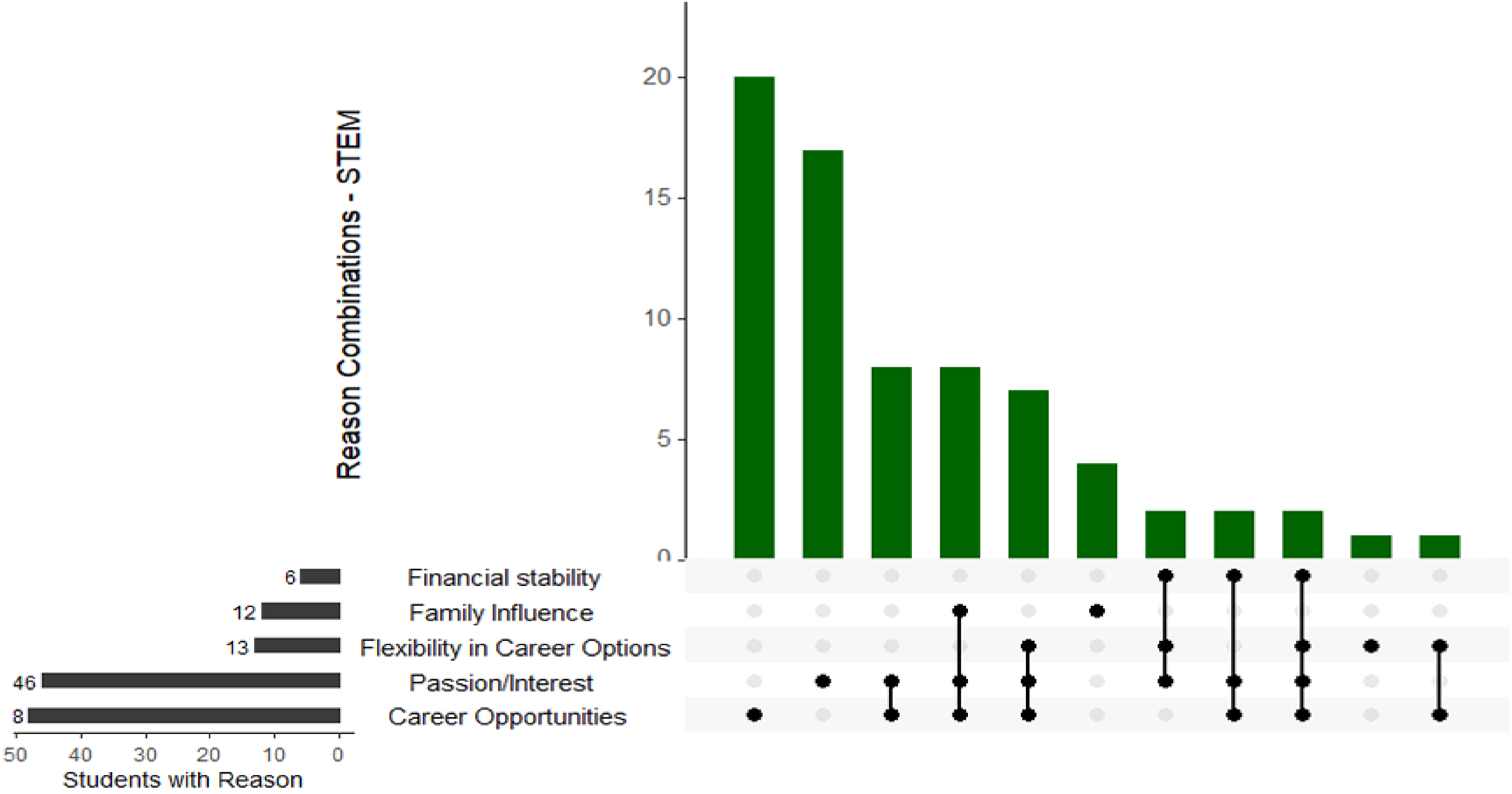
	Response				
Stream	Career Opportunities	Family Influence	Financial stability	Flexibility in Career Options	Passion/Interest
Non-STEM	88	42	52	59	113
STEM	48	12	6	13	46

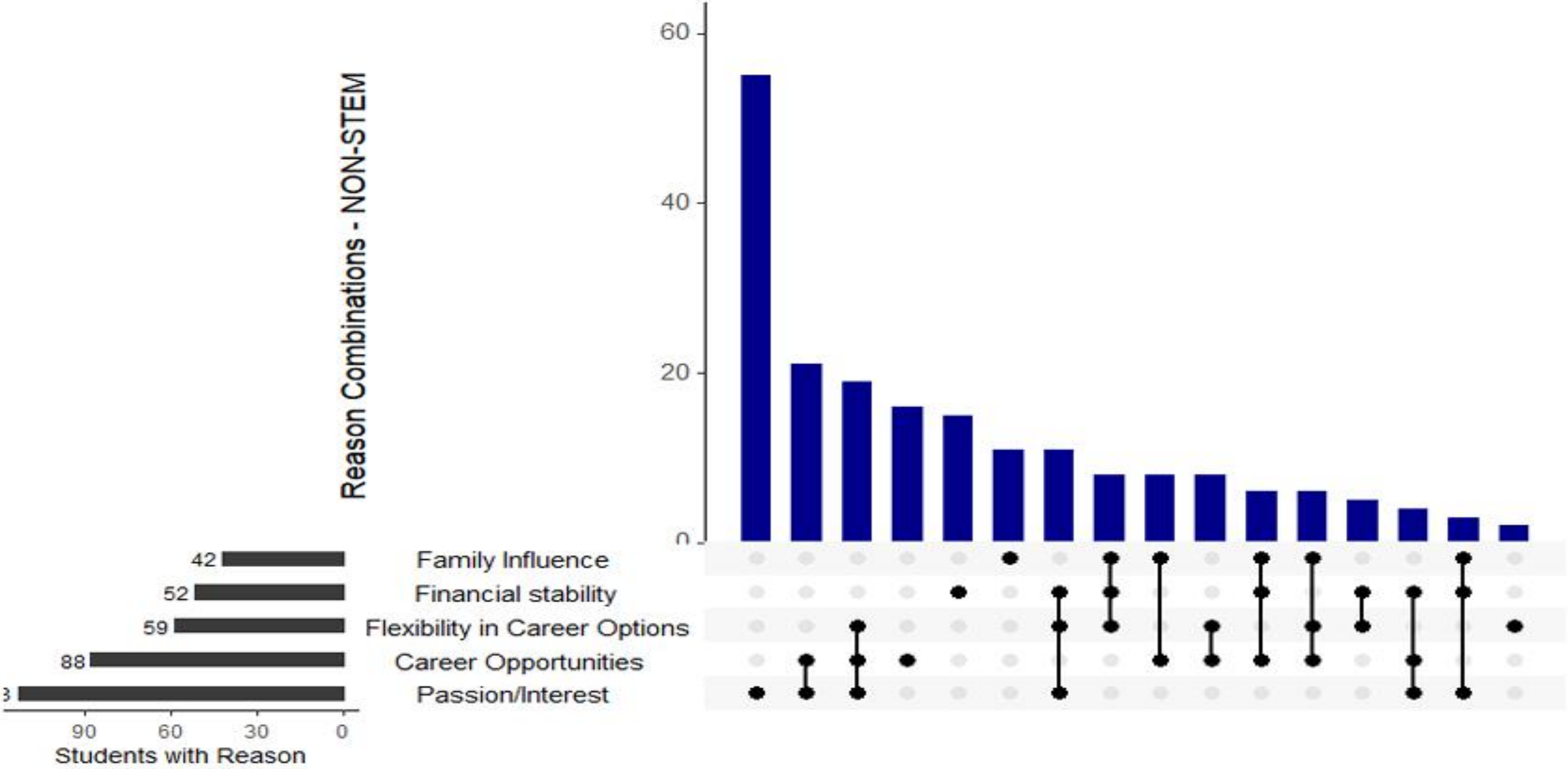
UpSet Plot





Comparative statistical analysis of STEM and Non-STEM students

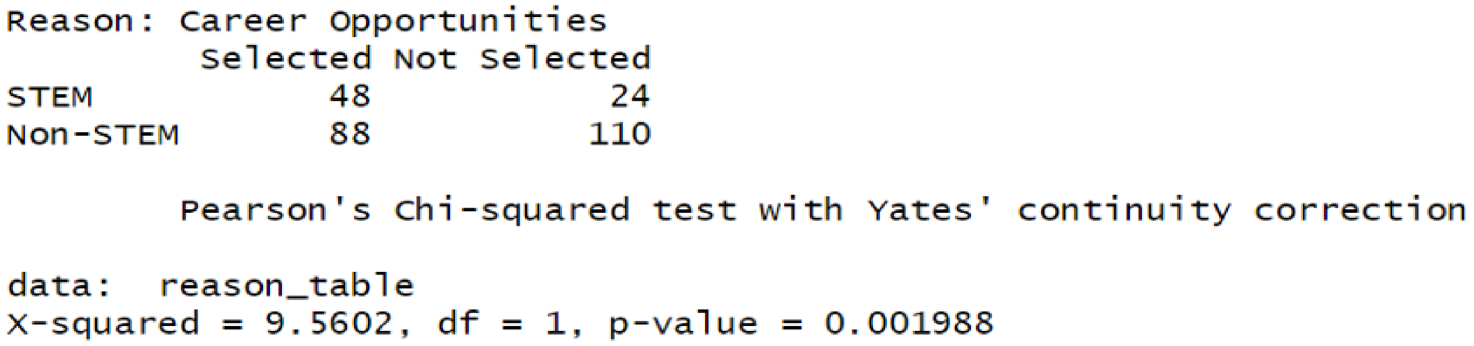

Reason Combinations - STEM
20
15
10
5
0
6
12
13
46
8
Financial stability
Family Influence
Flexibility in Career Options
Passion/Interest
Career Opportunities
50
40
30
20
10
0
Students with Reason


Reason Combinations - NON-STEM
60
40
20
0
42
52
59
88
3
Family Influence
Financial stability
Flexibility in Career Options
Career Opportunities
Passion/Interest
90
60
30
0
Students with Reason

As the counts are from the multi-responses question, Chi Square Test of independence cannot be directly applied to check for association. Hence, a 2x2 contingency table is formed for each reason and subsequently, Chi Square Test is applied.

#Hypothesis:

- 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.



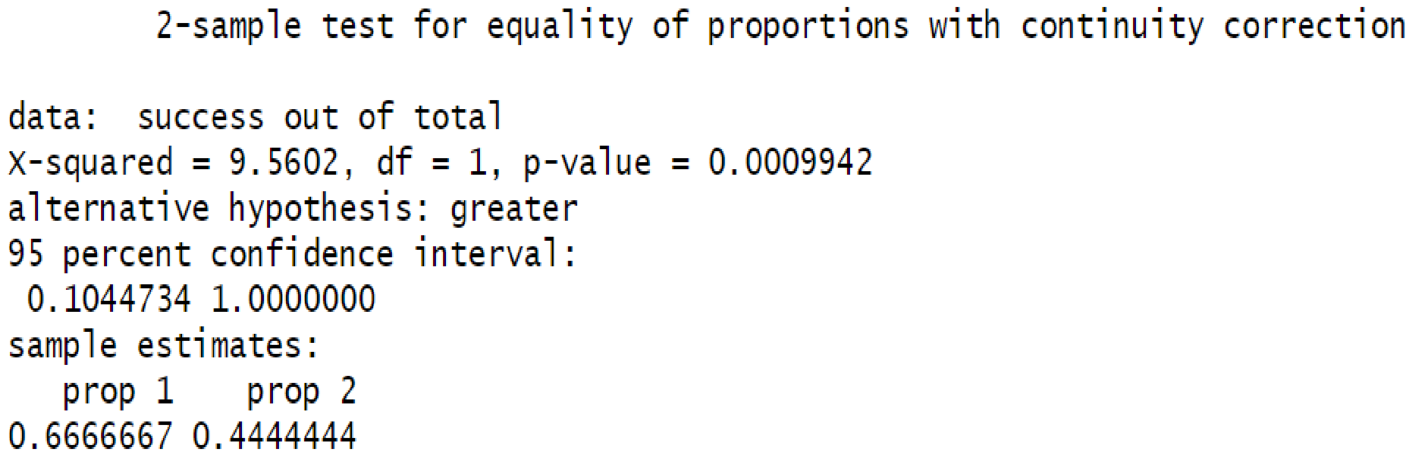
```
Reason: Career Opportunities
         Selected Not Selected
STEM           48           24
Non-STEM       88          110

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

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

As the p-value < 0.05, we reject the null hypothesis at 5% level of significance.
Therefore, we conclude that there exists an association between Stream and Career Opportunity being the reason of students to select a stream.

Next up, is the task to determine that which stream gives more preference to “career opportunity” as a reason to choose their field. For that purpose, one-sided proportion test is applied.

#Hypothesis:

- H_0: $P_1 \leq P_2$ (The proportion of STEM students who choose career opportunity 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 career opportunity as a primary motivation is greater than the proportion of Non-STEM students.)

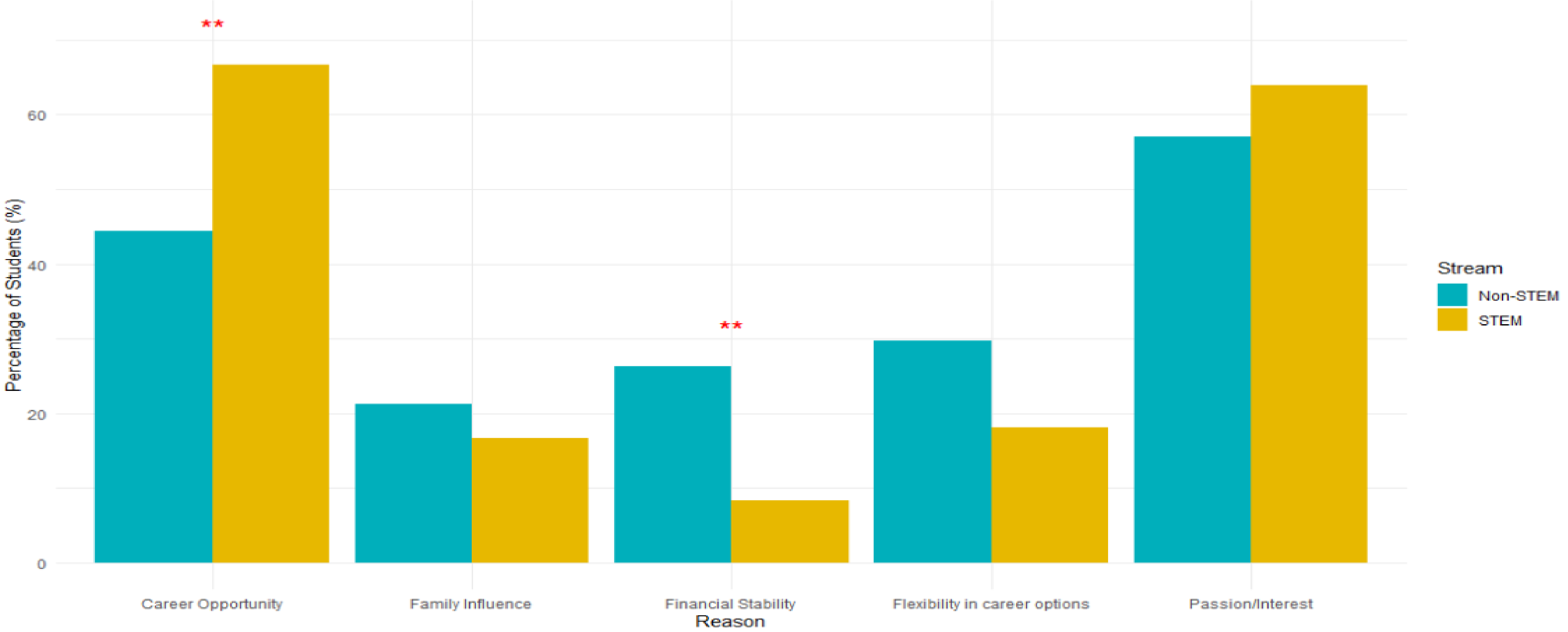


```
	2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 9.5602, 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 
```

As the p-value < 0.05, we reject the null hypothesis at 5% level of significance.
Therefore, we conclude that STEM students are more likely than Non-STEM students to cite “Career Opportunity" as a reason for their course choice.

Similarly, 2x2 contingency table is formed for each reason and following that Chi Square test is applied and proportion test is done for significant reasons.

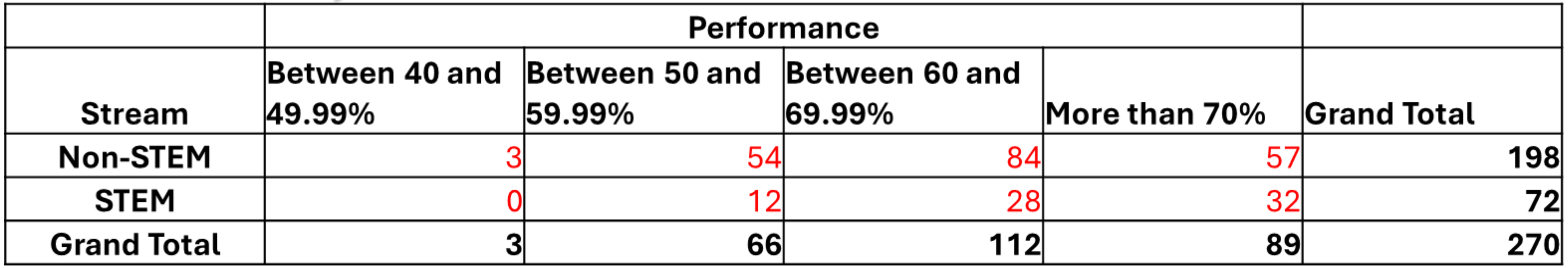
Reasons	p-value	Significance
Career Opportunities	**0.001988**	Significant
Family Influence	**0.5133**	Non-Significant
Financial Stability	**0.002659**	Significant
Flexibility in career options	**0.07608**	Non-Significant
Passion/Interest	**0.3859**	Non-Significant


Reasons for Stream Choice with Significance (Percentage)

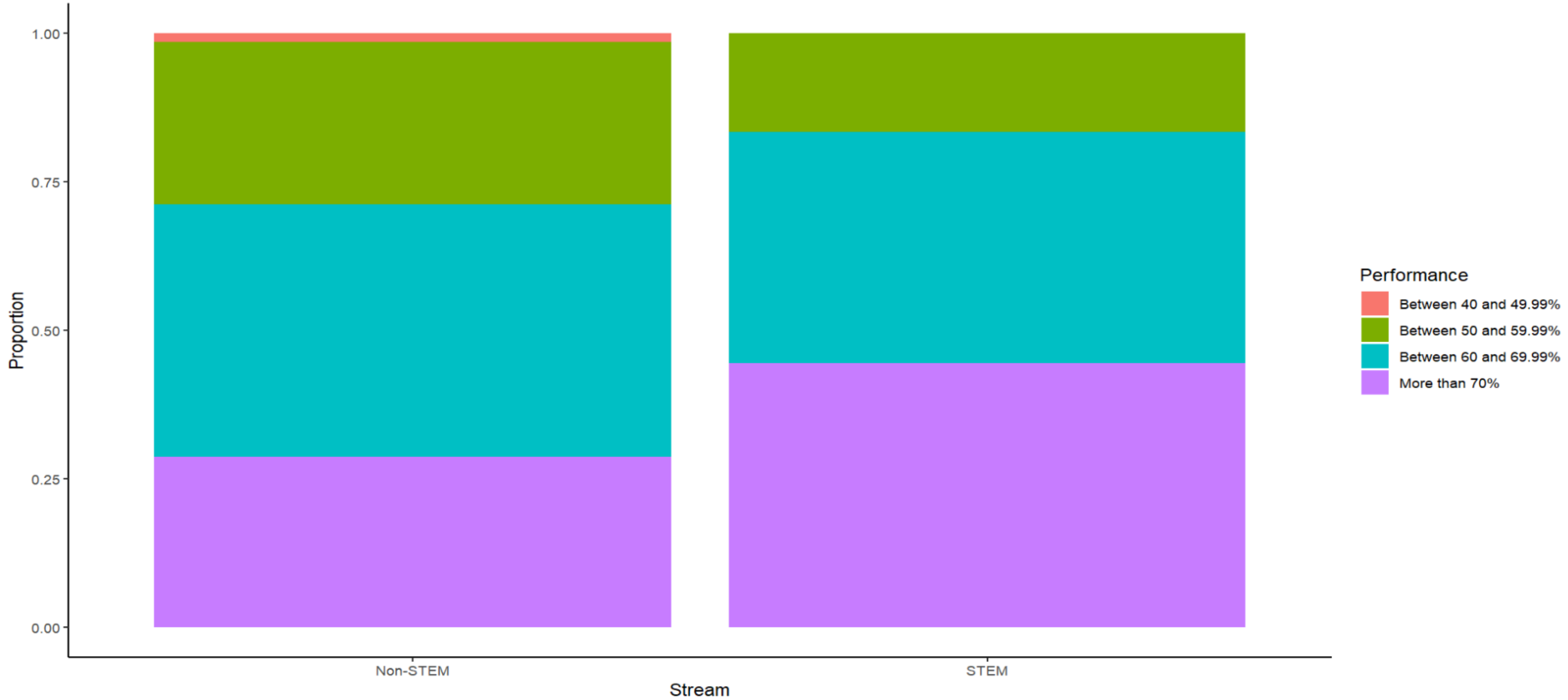
#Objective_2:

# To compare the academic performance and study pattern of STEM and Non-STEM students

data$Stream, data$Performance



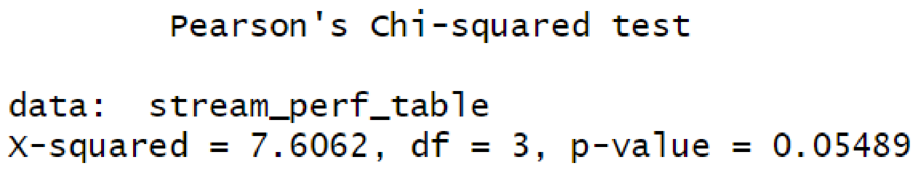
	Performance				
Stream	Between 40 and 49.99%	Between 50 and 59.99%	Between 60 and 69.99%	More than 70%	Grand Total
Non-STEM	3	54	84	57	198
STEM	0	12	28	32	72
Grand Total	3	66	112	89	270



Performance Proportions by Stream

#Hypothesis:

- Ho : There is **no** association between Stream and Performance of students.
- H1: There is an association between Stream and Performance of students.

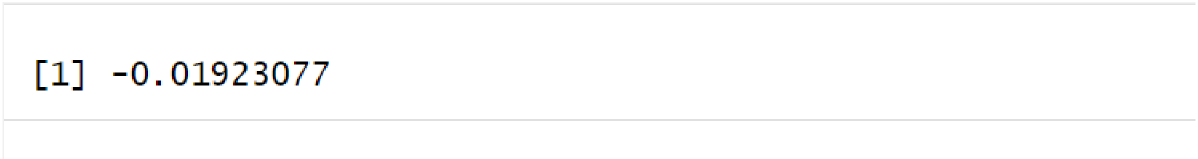


```
        Pearson's Chi-squared test

data:  stream_perf_table
X-squared = 7.6062, df = 3, p-value = 0.05489
```

As the p-value > 0.05, we do not have enough evidence to reject the null hypothesis at 5% level of significance.
Therefore, we conclude that there exists no association between Stream and Performances of students.

The performance is not affected by the stream chosen, but can we say that there is a relationship between performance and study hours?

We move forward using **Goodman Kruskal Gamma** to check if there is a meaningful relationship between Performance and StudyHours per day.

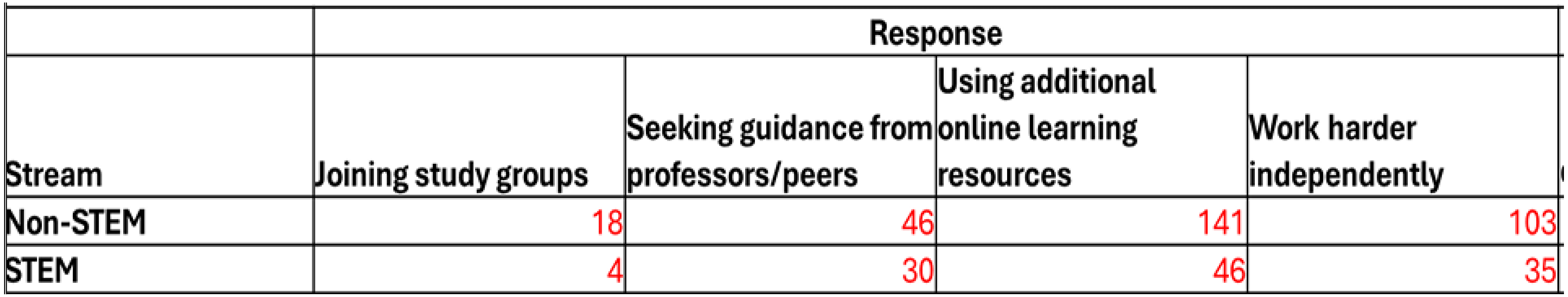


```
[1] -0.01923077
```

The Goodman-Kruskal Gamma value of -0.019 indicates a **very weak and essentially negligible negative association between Study Hours per day and Performance.**

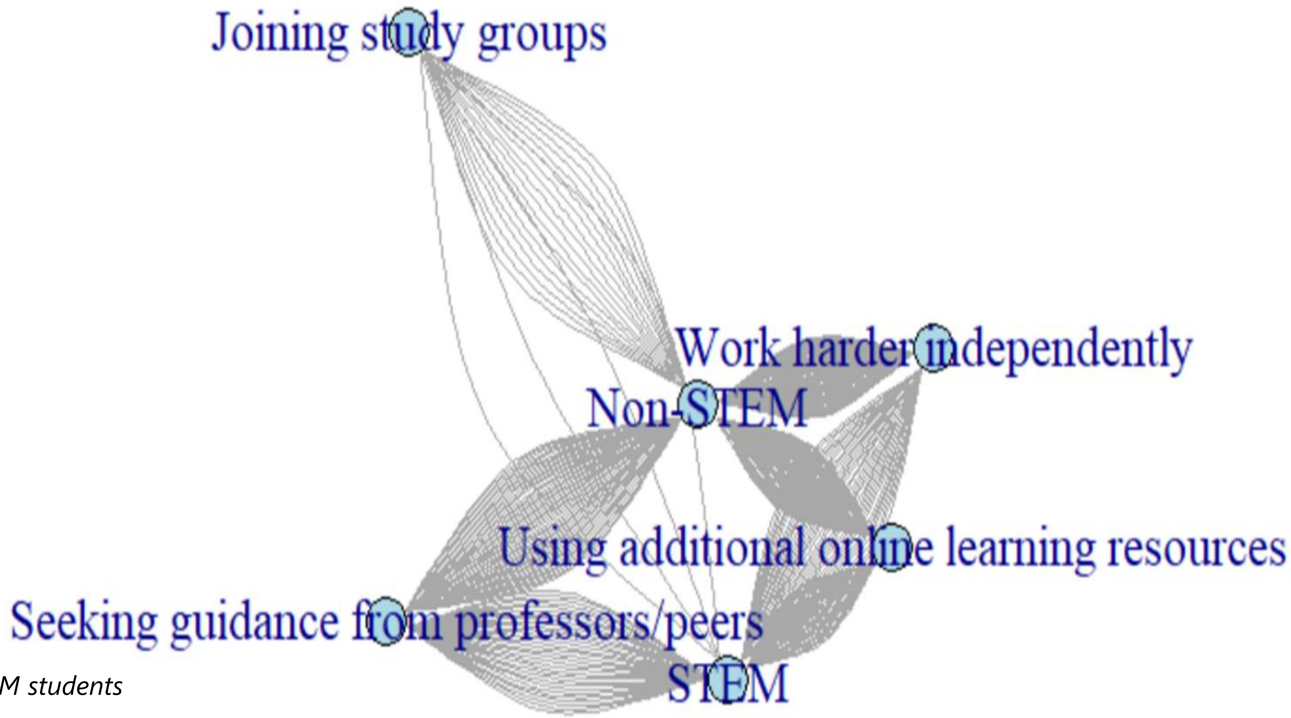
The negative sign suggests a slight tendency to perform better even with minimal but required hours of study — but the value is so close to 0 that it's likely due to **random variation or noise** in the data.

data$Stream, data$ResourcesUsed

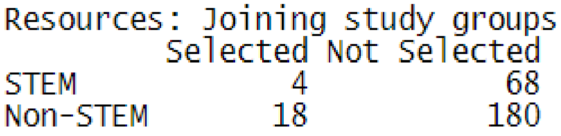


	Response			
Stream	Joining study groups	Seeking guidance from professors/peers	Using additional online learning resources	Work harder independently
Non-STEM	18	46	141	103
STEM	4	30	46	35

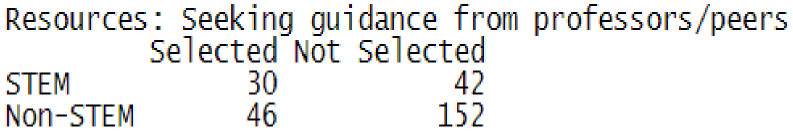
Network graph of stream to the resources used


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

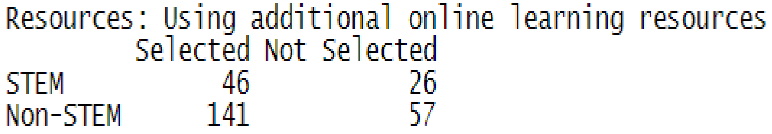
Question: Do STEM and Non-STEM students differ in the types of learning resources they use?



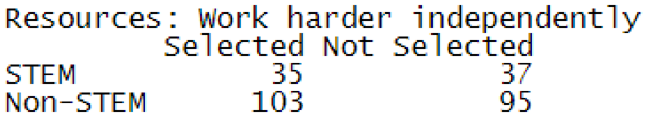
```
Resources: Joining study groups
         Selected Not Selected
STEM            4           68
Non-STEM       18          180
```



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



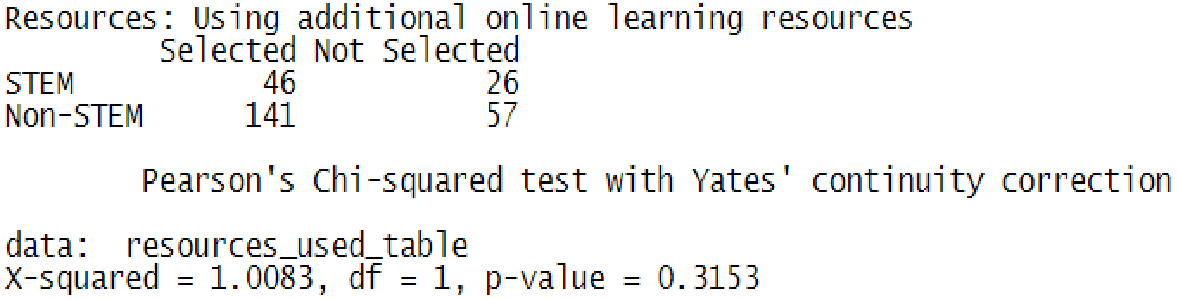
```
Resources: Using additional online learning resources
         Selected Not Selected
STEM           46           26
Non-STEM      141           57
```



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

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

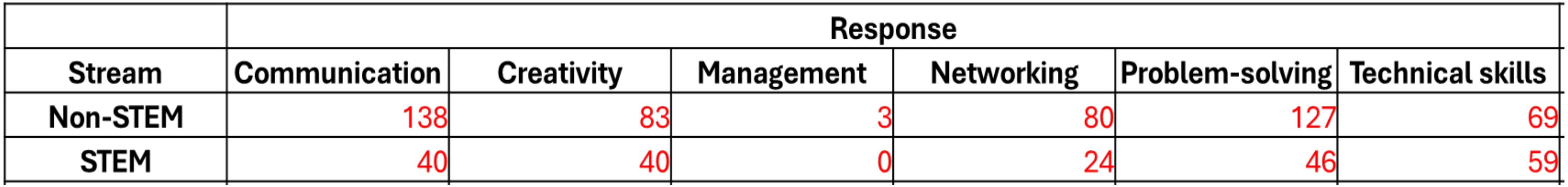
- **H_0**: There is no association between academic stream and resources used being online resources.
- **H_1**: There is an association between academic stream and resources used being online resources.



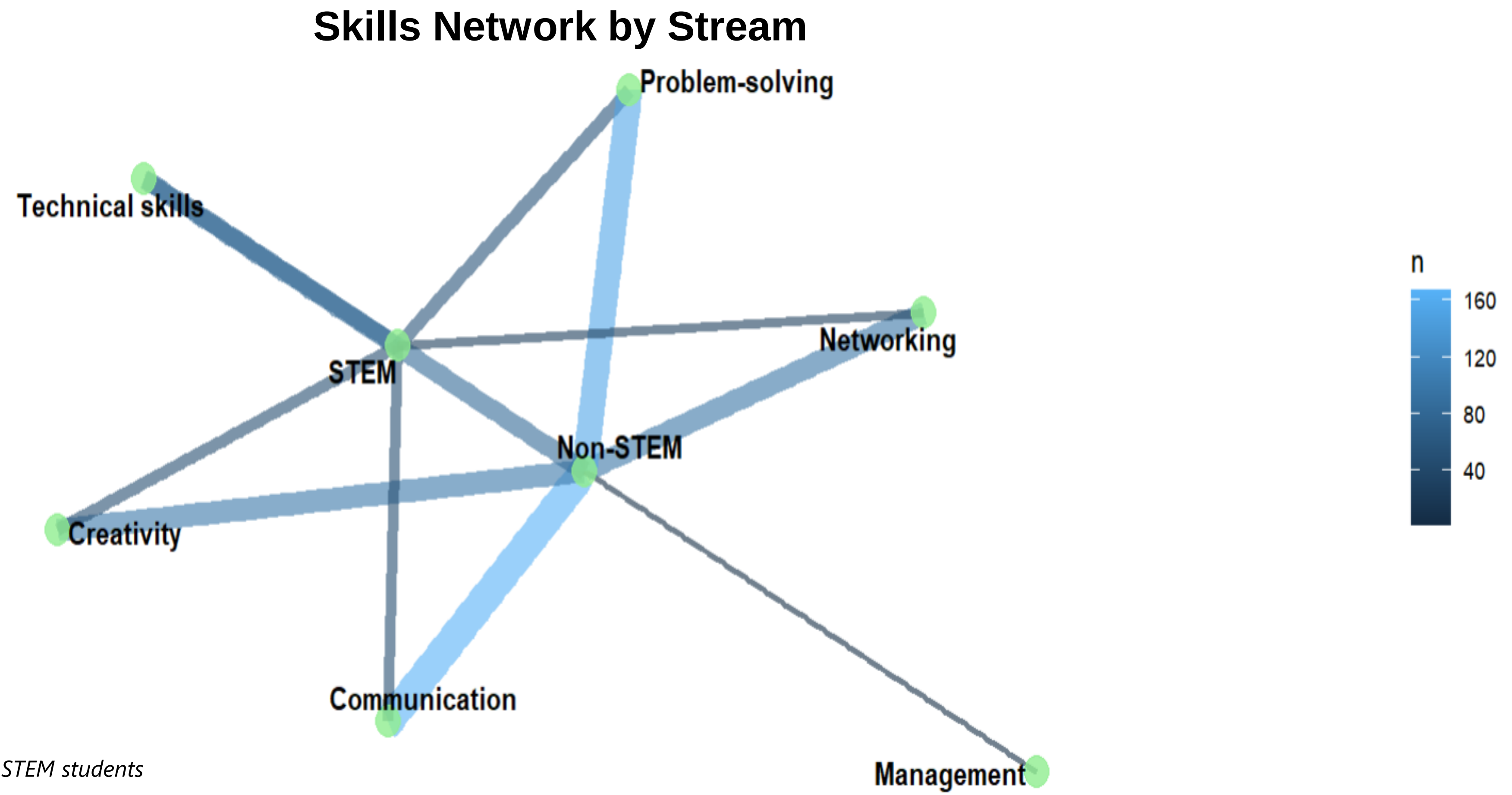
```
Resources: Using additional online learning resources
         Selected Not Selected
STEM           46           26
Non-STEM      141           57

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

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

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

data$Stream, data$Skills

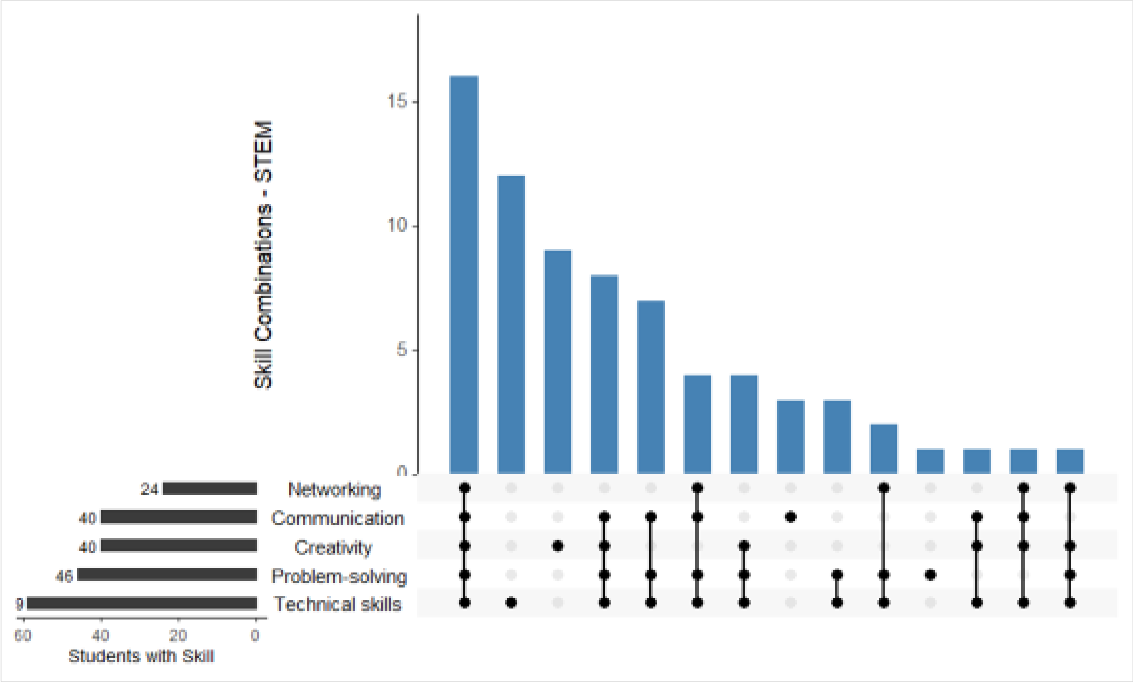


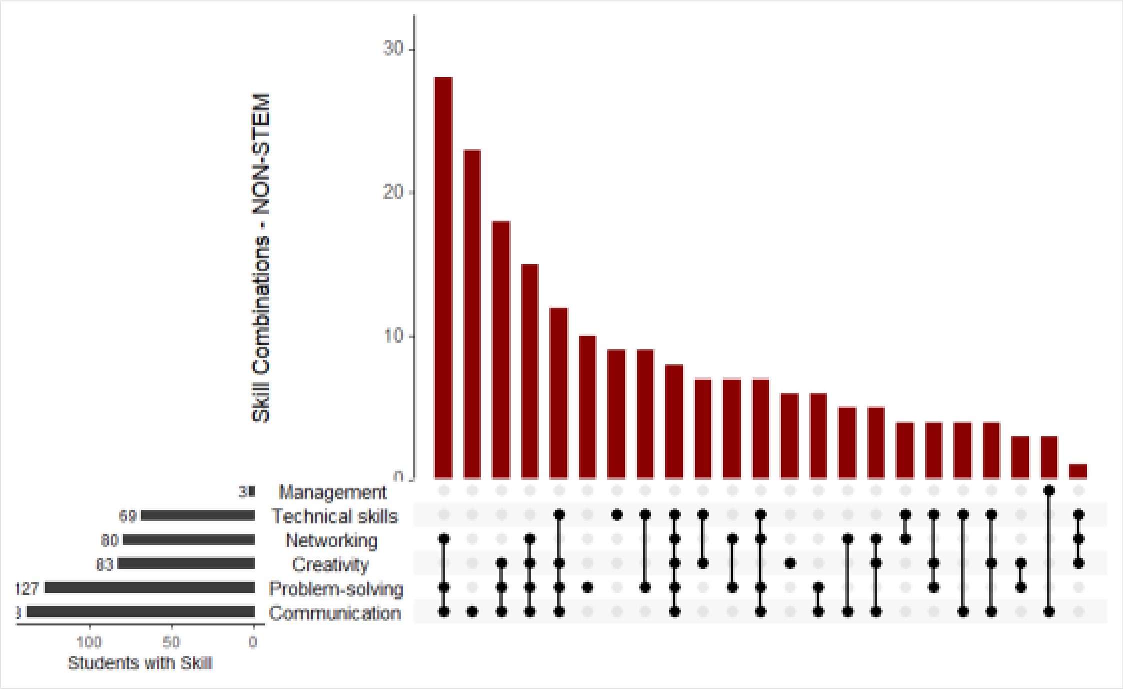
	Response					
Stream	Communication	Creativity	Management	Networking	Problem-solving	Technical skills
Non-STEM	138	83	3	80	127	69
STEM	40	40	0	24	46	59



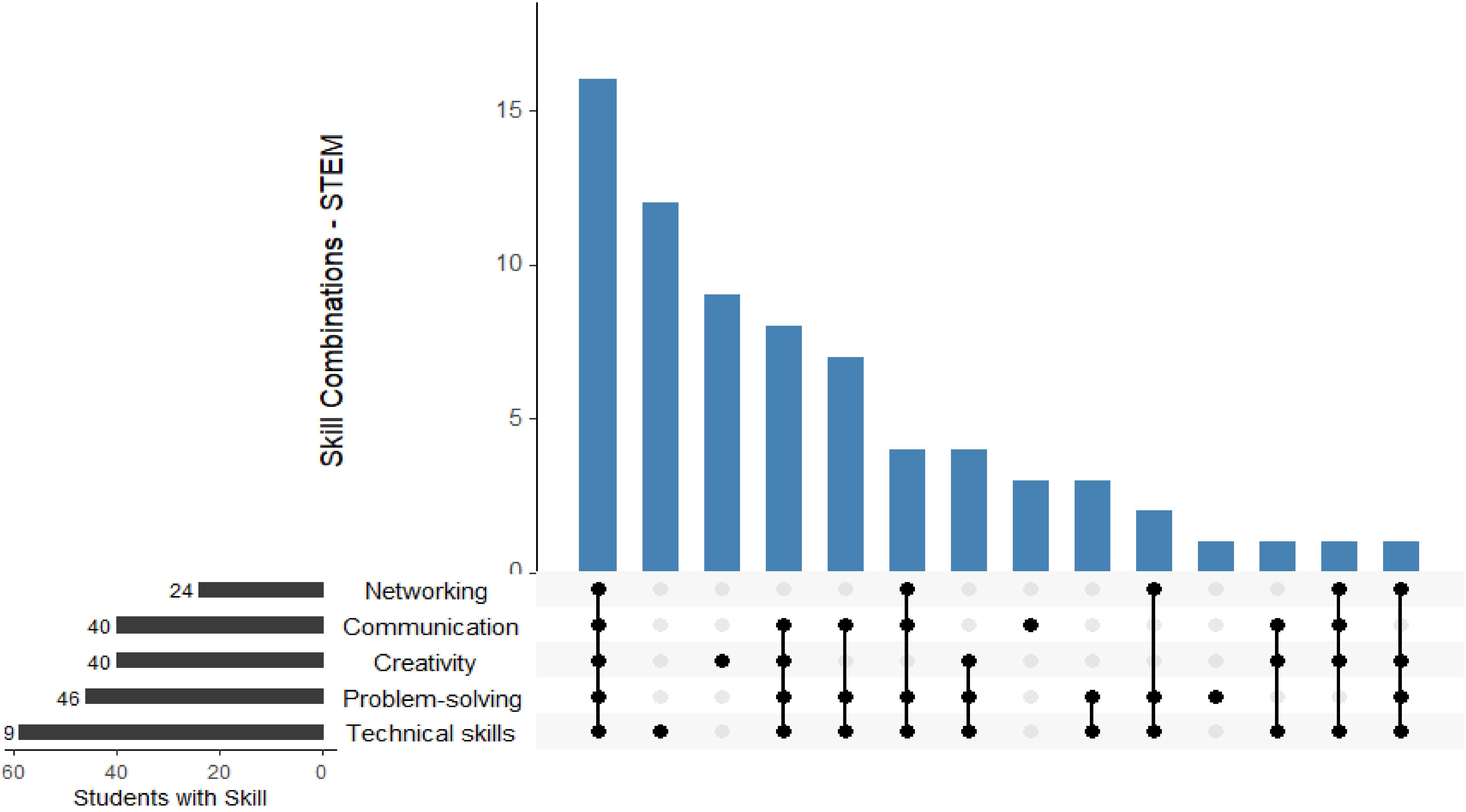
Skills Network by Stream

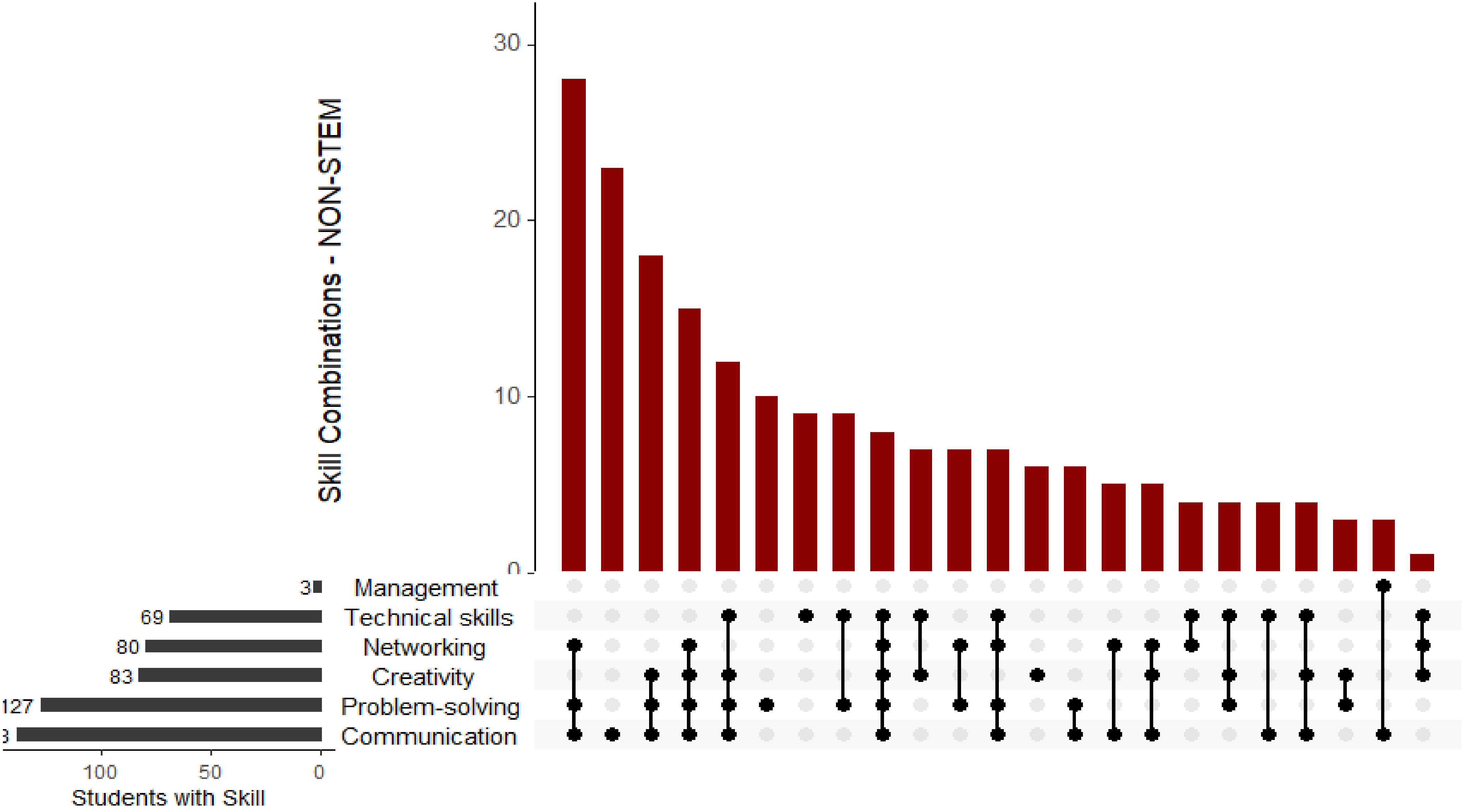
Upset Plot





Comparative statistical analysis of STEM and Non-STEM students

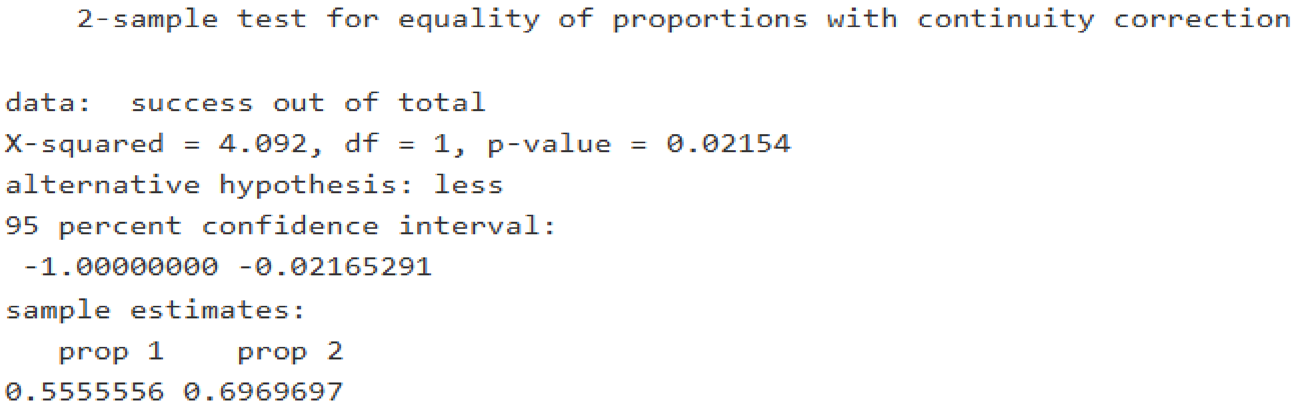

Skill Combinations - STEM
15
10
5
0
24
Networking
40
Communication
40
Creativity
46
Problem-solving
9
Technical skills
60
40
20
0
Students with Skill


Skill Combinations - NON-STEM
30
20
10
0
Management
Technical skills
Networking
Creativity
Problem-solving
Communication
3
69
80
83
127
3
100
50
0
Students with Skill

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

As communication shows a significant value for difference in proportion, we conduct a test presuming that students in Non-STEM fields require more communication skill comparatively.

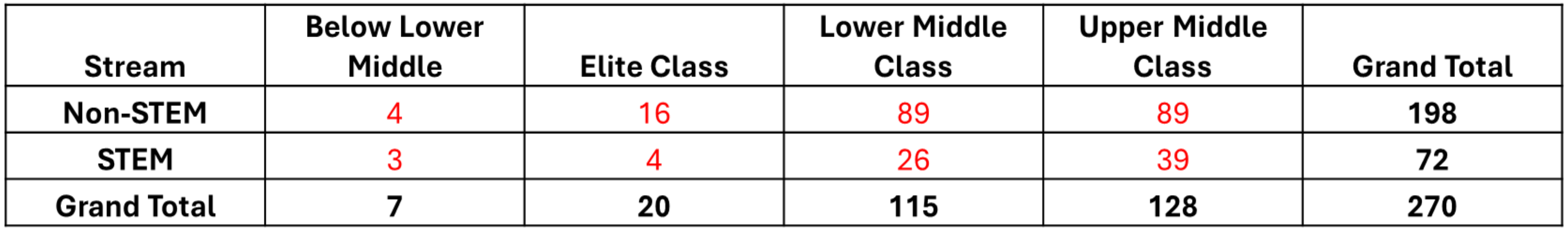
- H_0: $P_1 \geq P_2$(The proportion of STEM students require Communication skill is greater than or equal to the proportion of Non-STEM students.)
- H_1: $P_1 < P_2$ (The proportion of STEM students require Communication skill is less than the proportion of Non-STEM students.)



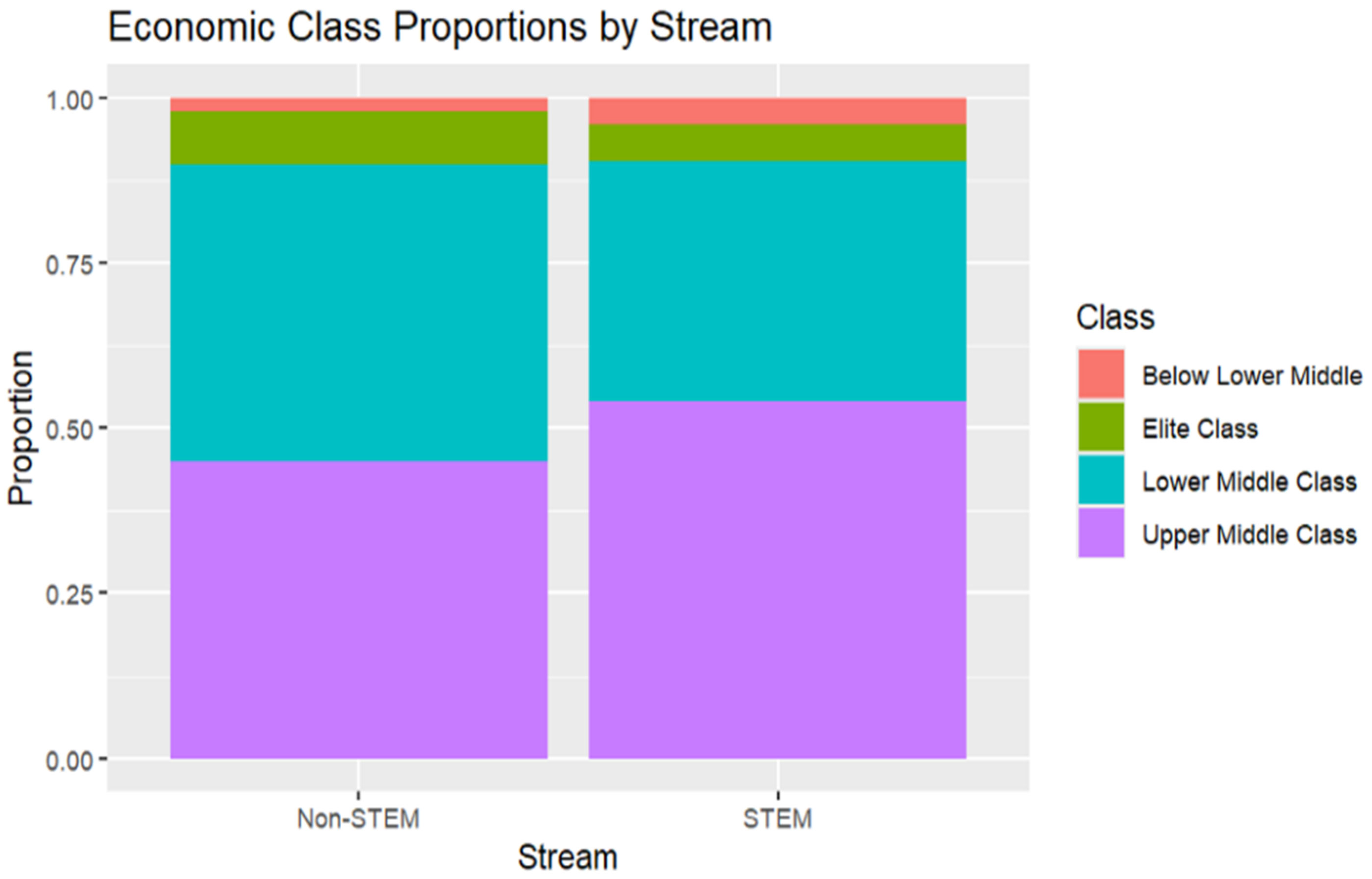
```
	2-sample test for equality of proportions with continuity correction

data:  success out of total
X-squared = 4.092, df = 1, p-value = 0.02154
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.02165291
sample estimates:
   prop 1    prop 2 
0.5555556 0.6969697 
```

Here, p-value < 0.05, hence we reject null hypothesis at 5% level of significance. Therefore, we conclude that Non-STEM student have more requirement of communication skill than STEM student.

data$Stream, data$EconomicClass



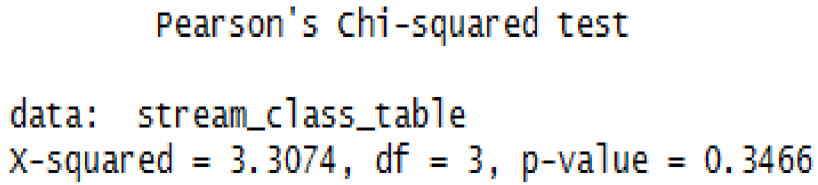
Stream	Below Lower Middle	Elite Class	Lower Middle Class	Upper Middle Class	Grand Total
Non-STEM	4	16	89	89	198
STEM	3	4	26	39	72
Grand Total	7	20	115	128	270



Do we observe significant difference in students choosing STEM or Non-STEM fields by their economic status?

#Hypotheses:

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



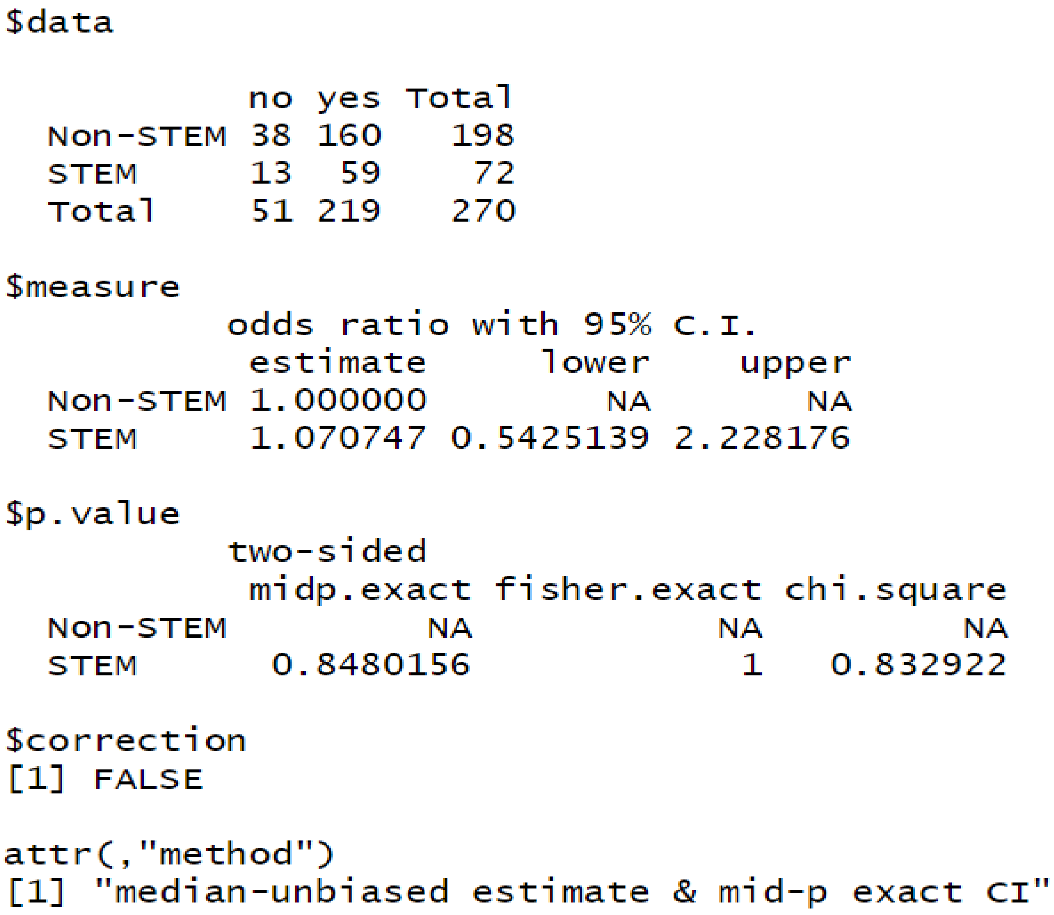
```
        Pearson's Chi-squared test

data:  stream_class_table
X-squared = 3.3074, df = 3, p-value = 0.3466
```

Here, p-value > 0.05, therefore we do not have enough evidence to reject the null hypothesis at 5% level of significance.
Economic status of a student is not a measure that significantly determines choice of field.

According to Family Socioeconomic Status and Choice of STEM Major in College, students from low socio-economic background have a lower representation in STEM, but the same is not reflected in our sample.

data$Stream, data$FirstChoice

Here, we aim to study that is there any evidence to conclude that students of any specific stream are studying the field as their first choice.



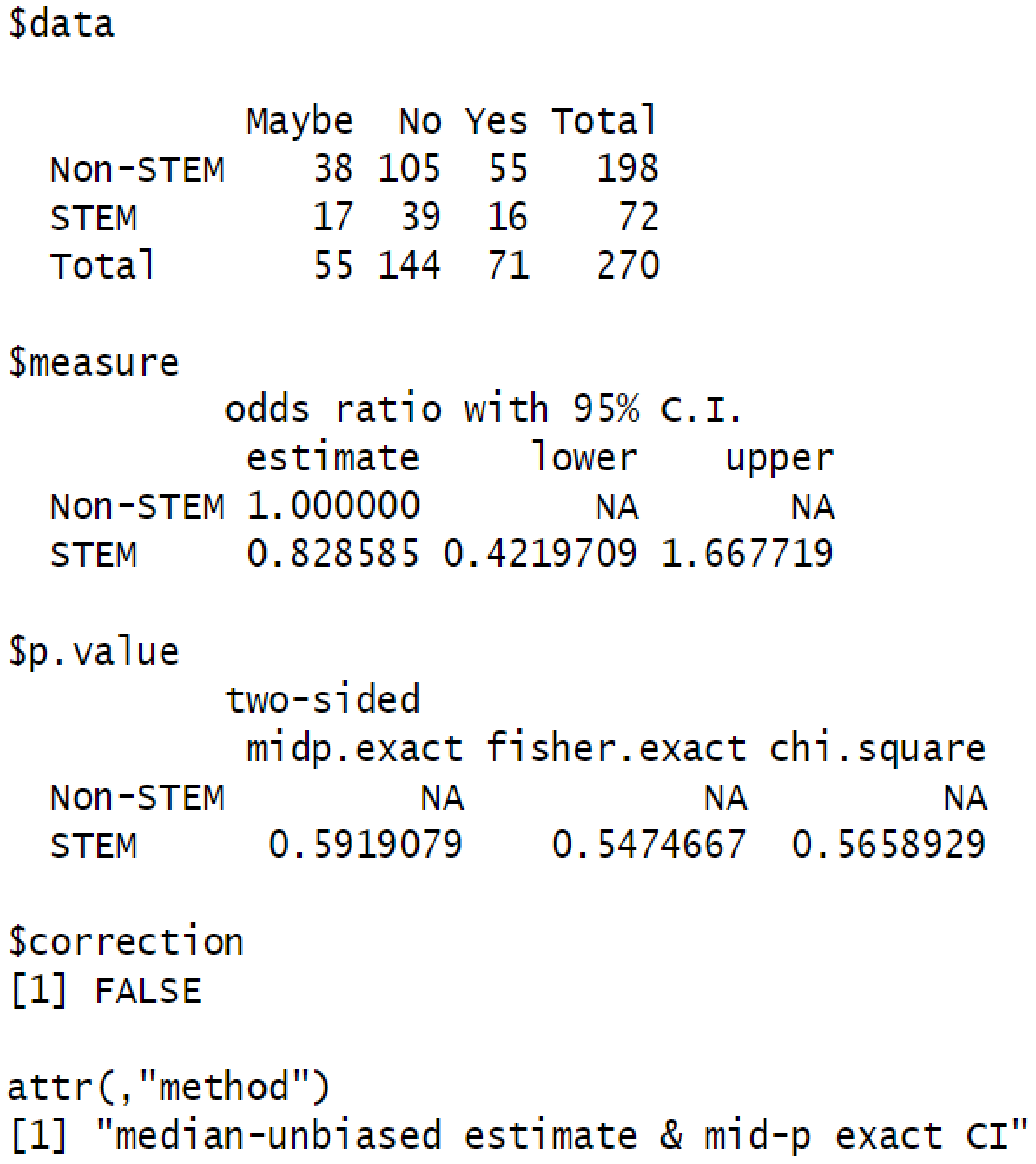
```
$data

          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"
```

All p-values are much higher than 0.05, so no statistically significant association between stream (STEM vs Non-STEM) and response to FirstChoice.

Odds ratio (OR) for STEM = 1.071
This means students in STEM are 7.1% more likely to say “Yes” to FirstChoice compared to Non-STEM students.

But here's the catch, the 95% Confidence Interval is from 0.543 to 2.228. Since 1 falls within the CI, the result is not statistically significant.

But is there any association between Stream and Switching?



```
$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"
```

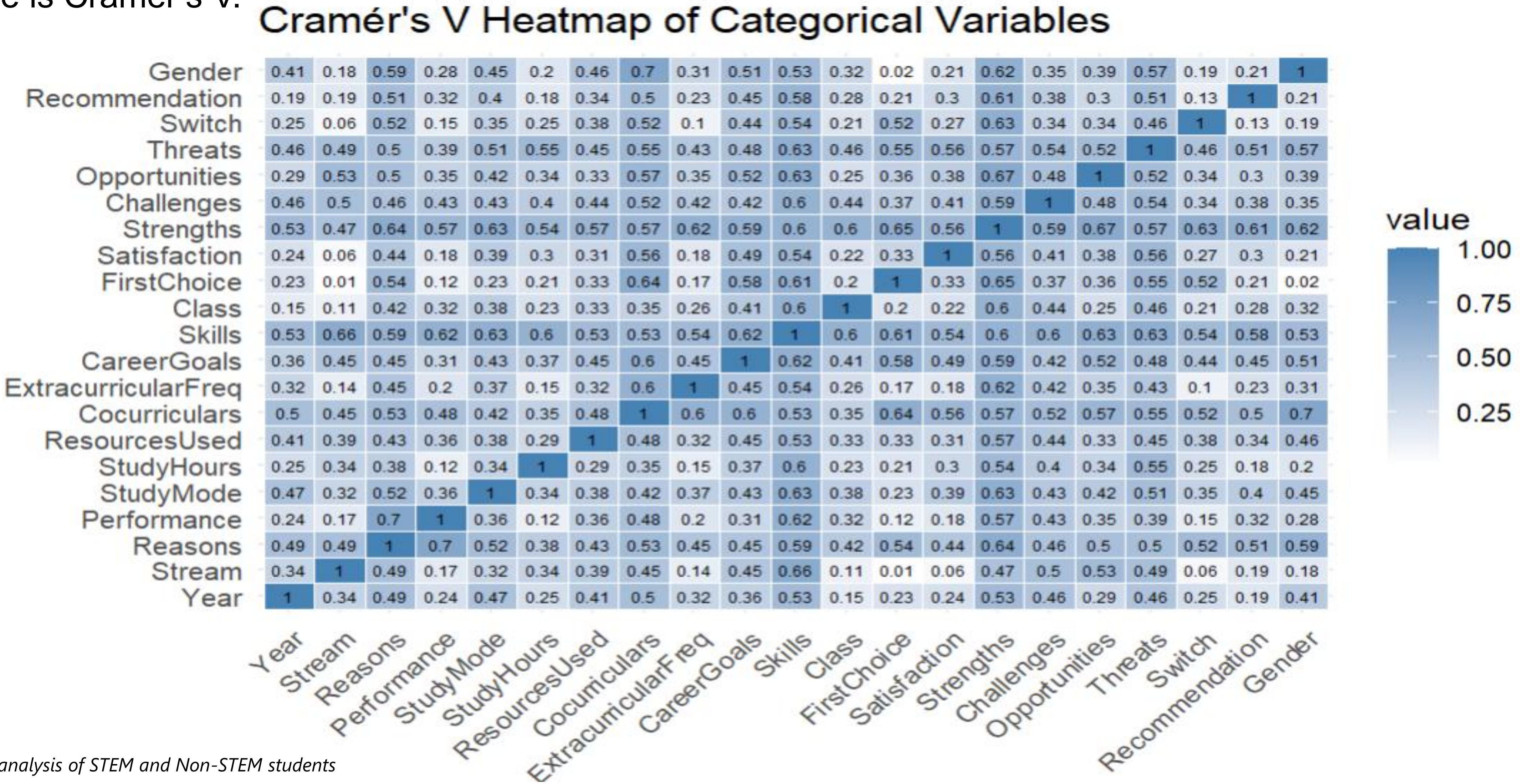
Odds ratio (OR) for STEM = 0.8285
This means students in STEM are 17.18% less likely to switch their field of study compared to Non-STEM students.

But the 95% Confidence Interval is from 0.421 to 1.667. Since 1 falls within the CI, the result is not statistically significant.

Although the odds ratio suggests STEM students are slightly less likely to switch than Non-STEM students, this difference is not statistically significant (p > 0.5). So, we can’t confidently say there’s a real difference between stream and switching.

?associations

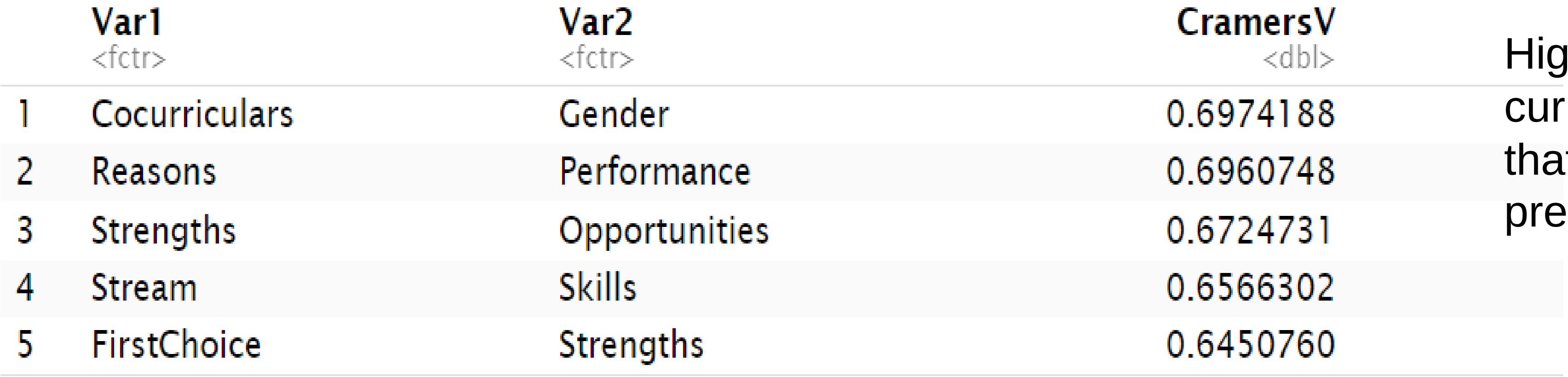
We now aim to find the association between all the variables present in the data. The tool we use for this purpose is Cramer’s V.



Cramér's V Heatmap of Categorical Variables

	Year	Stream	Reasons	Performance	StudyMode	StudyHours	ResourcesUsed	Cocurriculars	ExtracurricularFreq	CareerGoals	Skills	Class	FirstChoice	Satisfaction	Strengths	Challenges	Opportunities	Threats	Switch	Recommendation	Gender
Gender	0.41	0.18	0.59	0.28	0.45	0.2	0.46	0.7	0.31	0.51	0.53	0.32	0.02	0.21	0.62	0.35	0.39	0.57	0.19	0.21	1
Recommendation	0.19	0.19	0.51	0.32	0.4	0.18	0.34	0.5	0.23	0.45	0.58	0.28	0.21	0.3	0.61	0.38	0.3	0.51	0.13	1	0.21
Switch	0.25	0.06	0.52	0.15	0.35	0.25	0.38	0.52	0.1	0.44	0.54	0.21	0.52	0.27	0.63	0.34	0.34	0.46	1	0.13	0.19
Threats	0.46	0.49	0.5	0.39	0.51	0.55	0.45	0.55	0.43	0.48	0.63	0.46	0.55	0.56	0.57	0.54	0.52	1	0.46	0.51	0.57
Opportunities	0.29	0.53	0.5	0.35	0.42	0.34	0.33	0.57	0.35	0.52	0.63	0.25	0.36	0.38	0.67	0.48	1	0.52	0.34	0.3	0.39
Challenges	0.46	0.5	0.46	0.43	0.43	0.4	0.44	0.52	0.42	0.42	0.6	0.44	0.37	0.41	0.59	1	0.48	0.54	0.34	0.38	0.35
Strengths	0.53	0.47	0.64	0.57	0.63	0.54	0.57	0.57	0.62	0.59	0.6	0.6	0.65	0.56	1	0.59	0.67	0.57	0.63	0.61	0.62
Satisfaction	0.24	0.06	0.44	0.18	0.39	0.3	0.31	0.56	0.18	0.49	0.54	0.22	0.33	1	0.56	0.41	0.38	0.56	0.27	0.3	0.21
FirstChoice	0.23	0.01	0.54	0.12	0.23	0.21	0.33	0.64	0.17	0.58	0.61	0.2	1	0.33	0.65	0.37	0.36	0.55	0.52	0.21	0.02
Class	0.15	0.11	0.42	0.32	0.38	0.23	0.33	0.35	0.26	0.41	0.6	1	0.2	0.22	0.6	0.44	0.25	0.46	0.21	0.28	0.32
Skills	0.53	0.66	0.59	0.62	0.63	0.6	0.53	0.53	0.54	0.62	1	0.6	0.61	0.54	0.6	0.6	0.63	0.63	0.54	0.58	0.53
CareerGoals	0.36	0.45	0.45	0.31	0.43	0.37	0.45	0.6	0.45	1	0.62	0.41	0.58	0.49	0.59	0.42	0.52	0.48	0.44	0.45	0.51
ExtracurricularFreq	0.32	0.14	0.45	0.2	0.37	0.15	0.32	0.6	1	0.45	0.54	0.26	0.17	0.18	0.62	0.42	0.35	0.43	0.1	0.23	0.31
Cocurriculars	0.5	0.45	0.53	0.48	0.42	0.35	0.48	1	0.6	0.6	0.53	0.35	0.64	0.56	0.57	0.52	0.57	0.55	0.52	0.5	0.7
ResourcesUsed	0.41	0.39	0.43	0.36	0.38	0.29	1	0.48	0.32	0.45	0.53	0.33	0.33	0.31	0.57	0.44	0.33	0.45	0.38	0.34	0.46
StudyHours	0.25	0.34	0.38	0.12	0.34	1	0.29	0.35	0.15	0.37	0.6	0.23	0.21	0.3	0.54	0.4	0.34	0.55	0.25	0.18	0.2
StudyMode	0.47	0.32	0.52	0.36	1	0.34	0.38	0.42	0.37	0.43	0.63	0.38	0.23	0.39	0.63	0.43	0.42	0.51	0.35	0.4	0.45
Performance	0.24	0.17	0.7	1	0.36	0.12	0.36	0.48	0.2	0.31	0.62	0.32	0.12	0.18	0.57	0.43	0.35	0.39	0.15	0.32	0.28
Reasons	0.49	0.49	1	0.7	0.52	0.38	0.43	0.53	0.45	0.45	0.59	0.42	0.54	0.44	0.64	0.46	0.5	0.5	0.52	0.51	0.59
Stream	0.34	1	0.49	0.17	0.32	0.34	0.39	0.45	0.14	0.45	0.66	0.11	0.01	0.06	0.47	0.5	0.53	0.49	0.06	0.19	0.18
Year	1	0.34	0.49	0.24	0.47	0.25	0.41	0.5	0.32	0.36	0.53	0.15	0.23	0.24	0.53	0.46	0.29	0.46	0.25	0.19	0.41

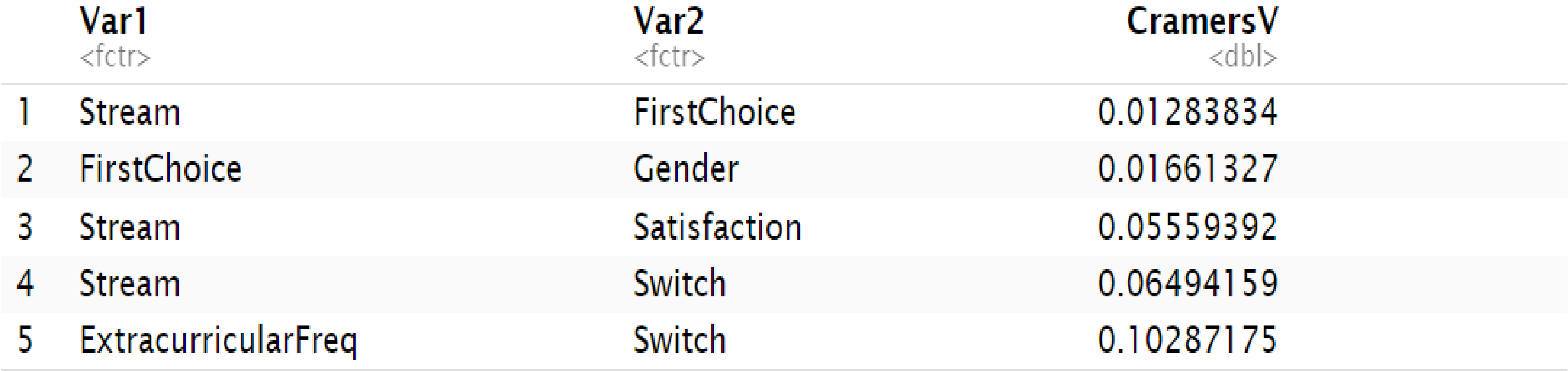
Top 5 associations



	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

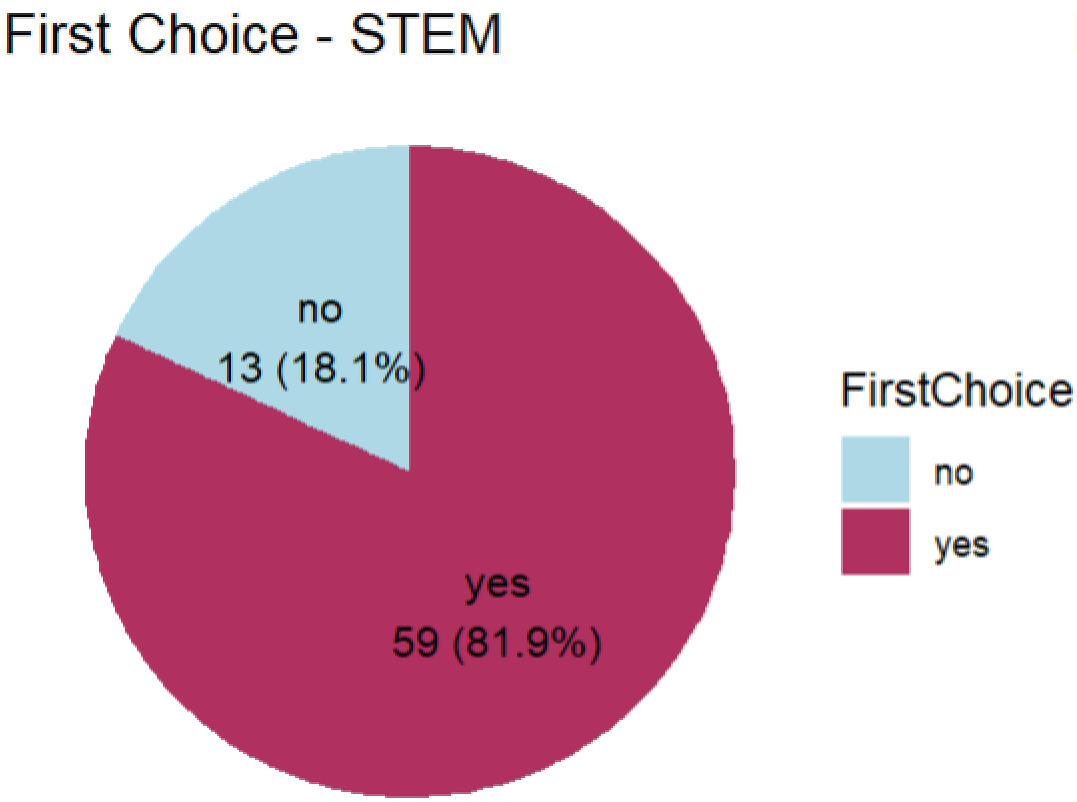
High association between Co-curriculars and Gender may suggest that Males and Females do have a preference for co-curricular activities.

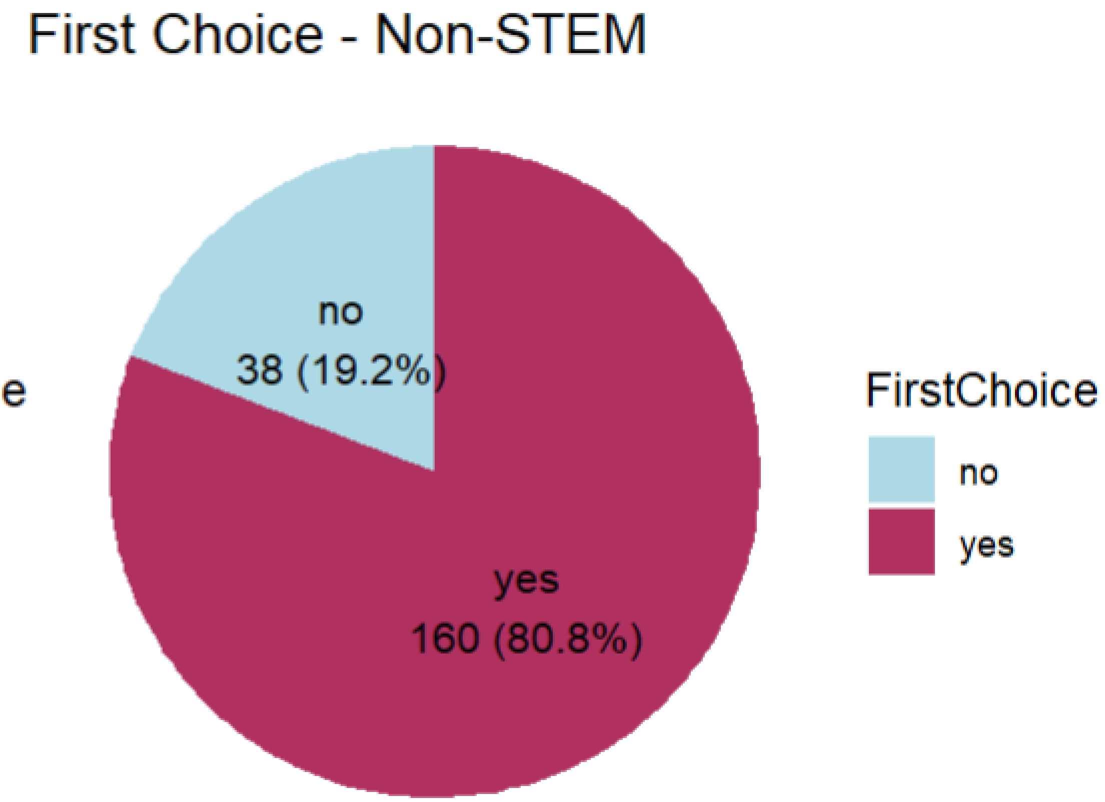
Bottom 5 associations



	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

But wait, there is low association between Stream and FirstChoice? Does this mean that most of students are not studying according to their first choice?



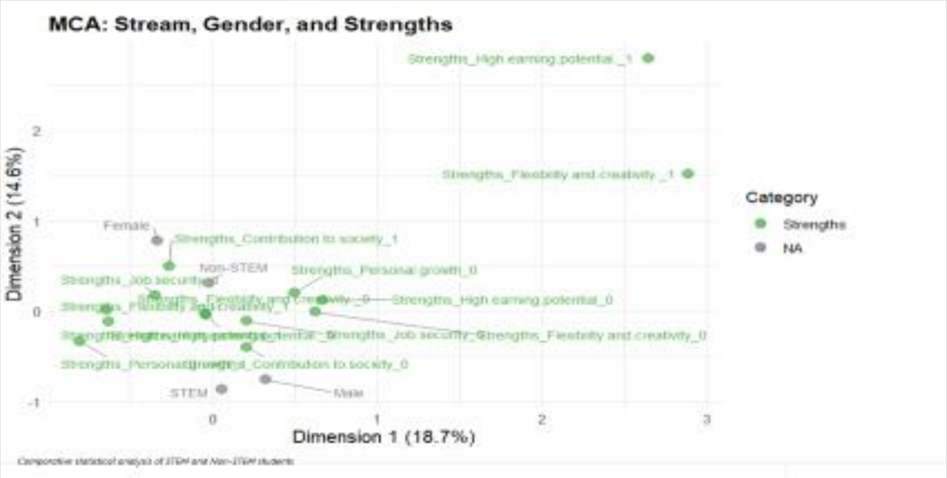


We observe from the pie graph that the proportion of students studying STEM or Non-STEM as their first choice is nearly the same. This explains the low value of association between the variables.

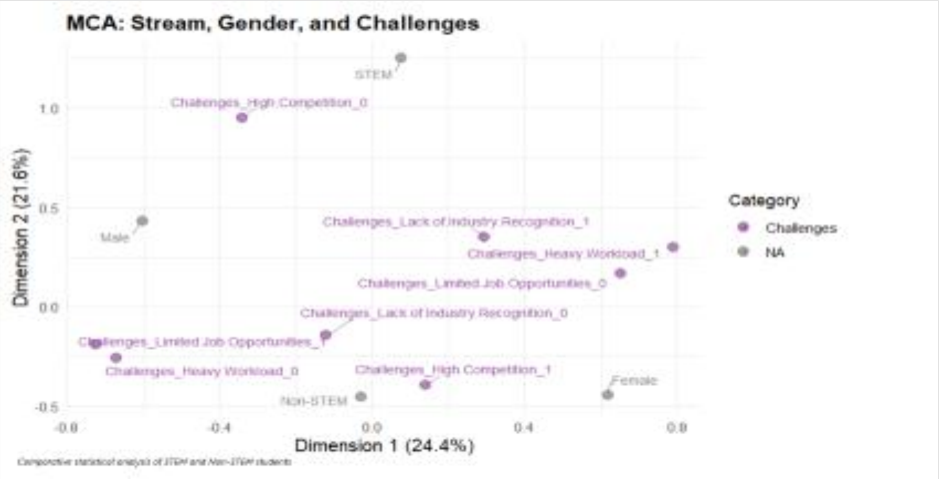
#Objective_3:

# To perform SWOT analysis for STEM and Non-STEM fields

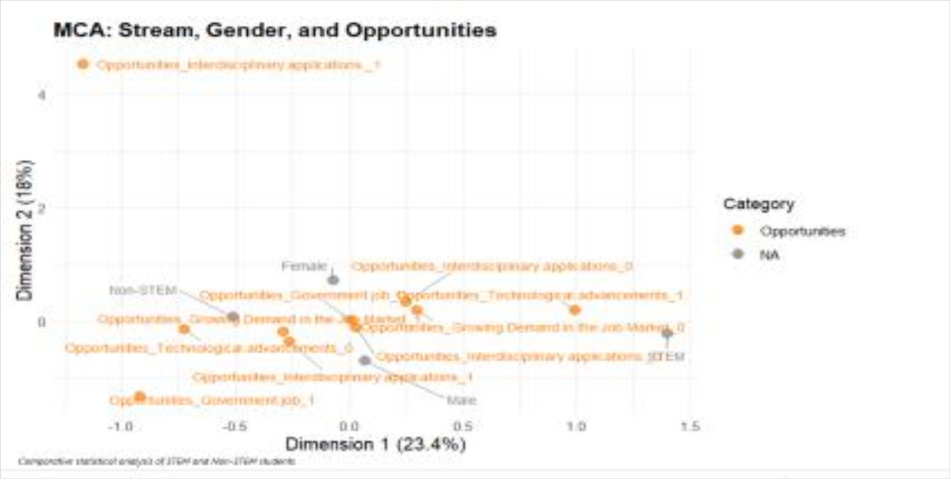
SWOT Analysis by MCA Plot


MCA: Stream, Gender, and Strengths

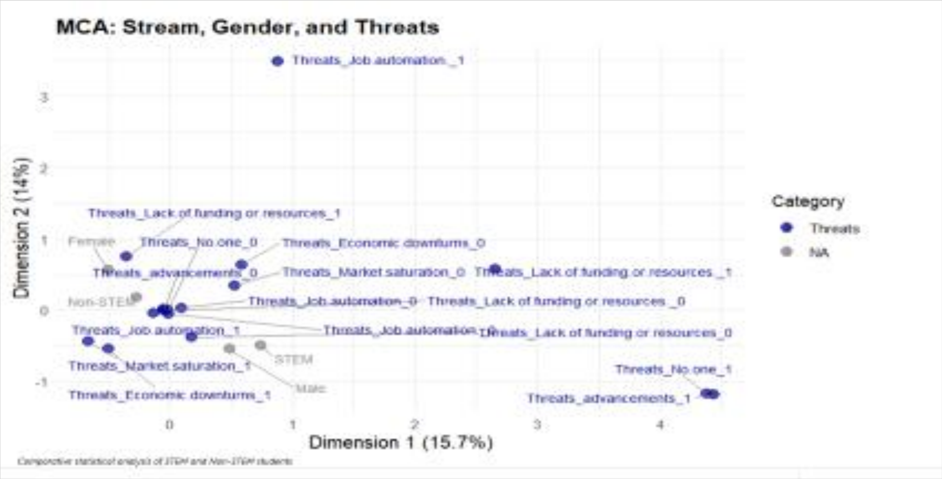
Comparative statistical analysis of STEM and Non-STEM students


MCA: Stream, Gender, and Challenges

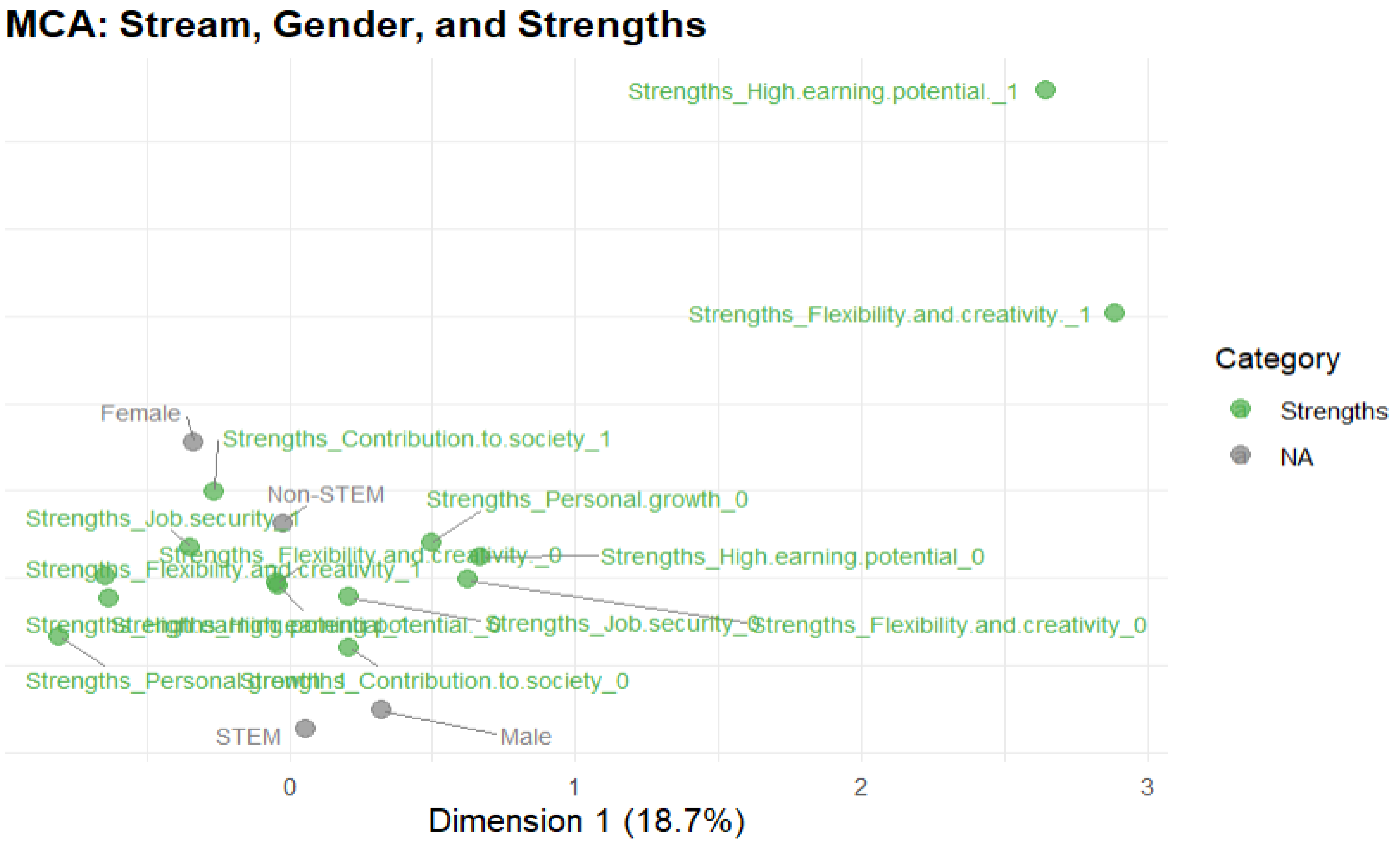
Comparative statistical analysis of STEM and Non-STEM students

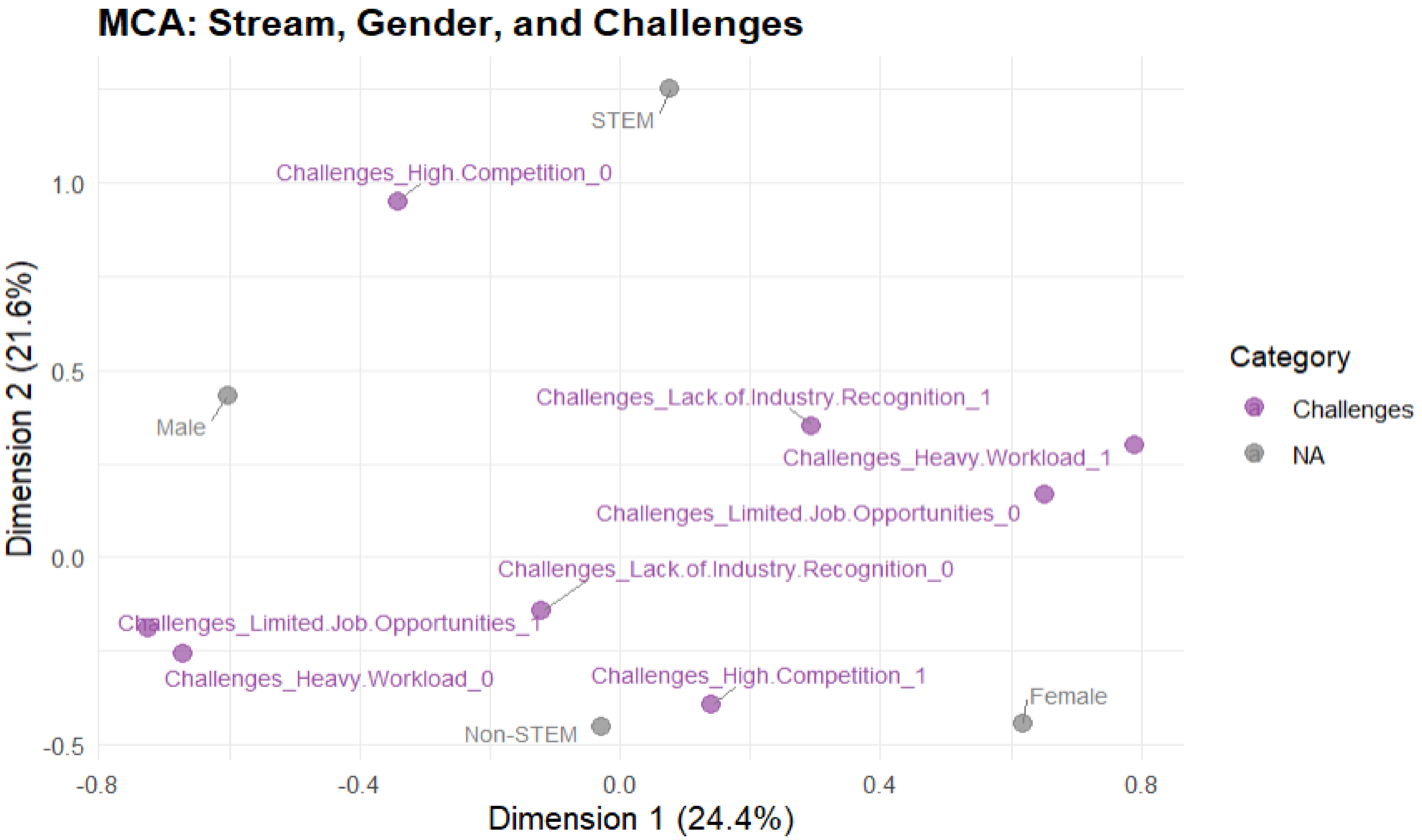

MCA: Stream, Gender, and Opportunities

Comparative statistical analysis of STEM and Non-STEM students

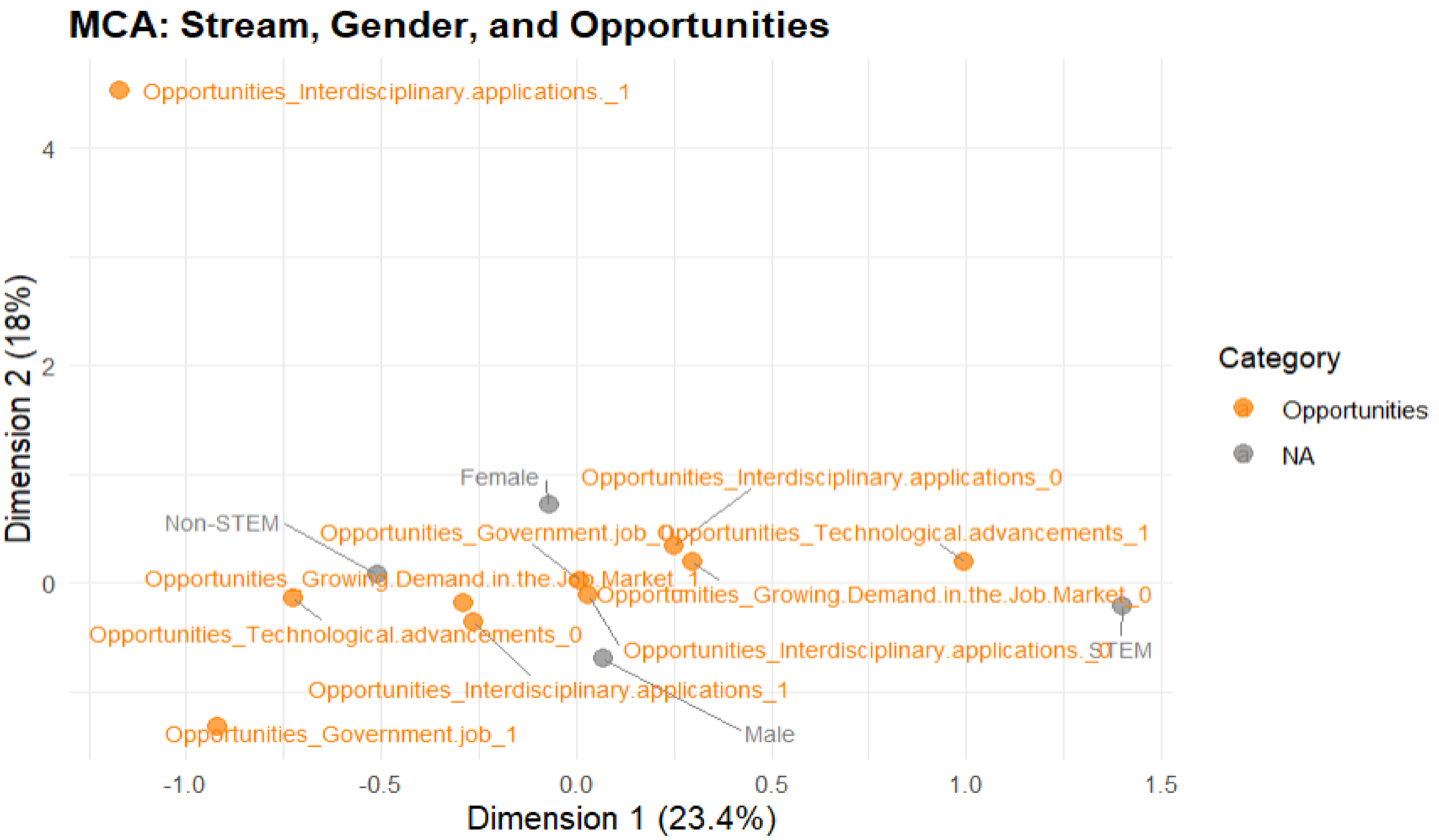

MCA: Stream, Gender, and Threats

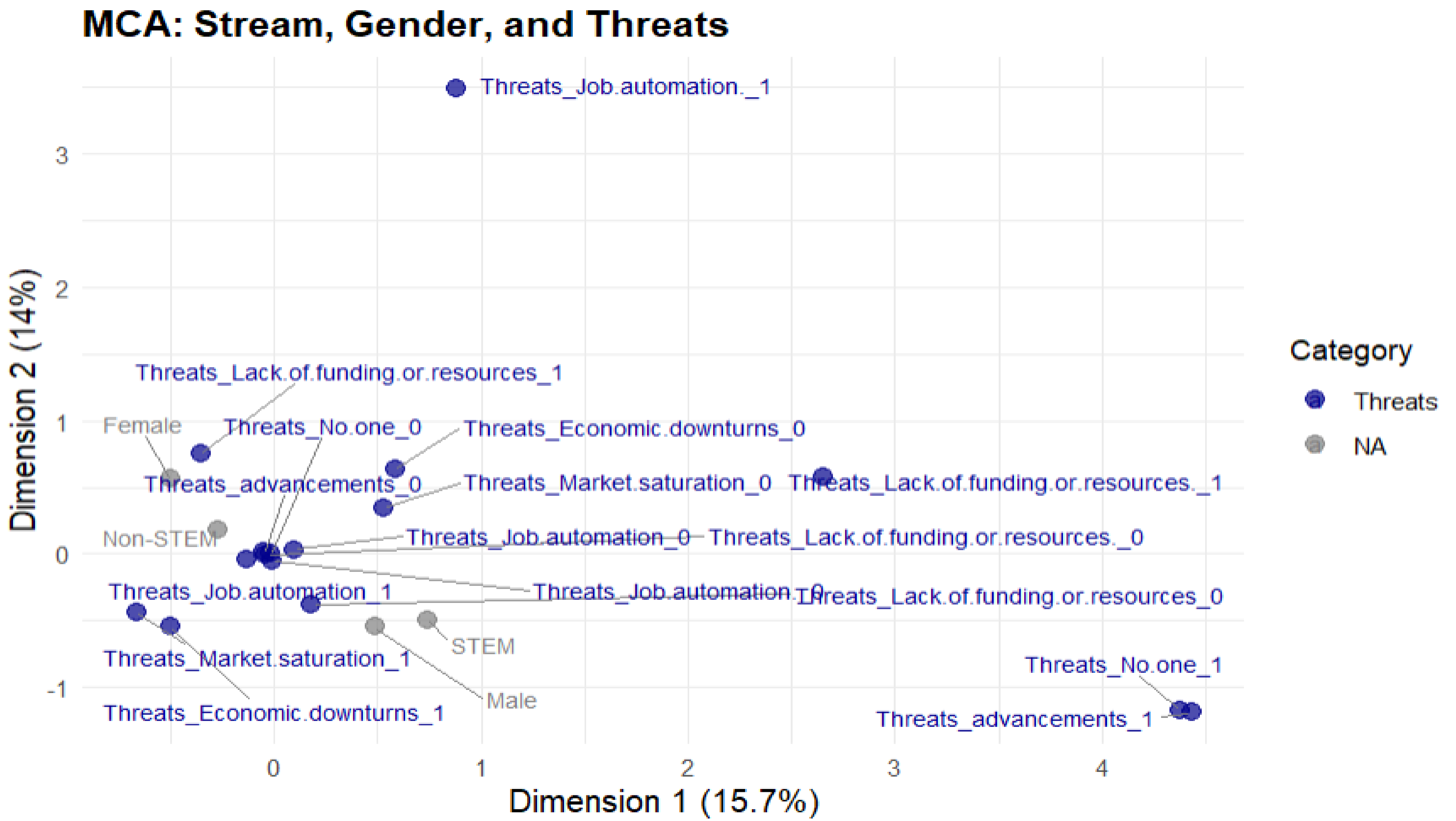
Comparative statistical analysis of STEM and Non-STEM students


MCA: Stream, Gender, and Strengths
Strengths_High.earning.potential._1
Strengths_Flexibility.and.creativity._1
Female
Strengths_Contribution.to.society_1
Non-STEM
Strengths_Personal.growth_0
Strengths_Job.security_1
Strengths_Flexibility.and.creativity._0
Strengths_High.earning.potential_0
Strengths_Flexibility.and.creativity_1
Strengths_High.earning.potential_1
Strengths_High.earning.potential._0
Strengths_Job.security_0
Strengths_Flexibility.and.creativity_0
Strengths_Personal.growth_1
Strengths_Contribution.to.society_0
STEM
Male
Dimension 2 (14.6%)
2
1
0
-1
Dimension 1 (18.7%)
0
1
2
3
Category
Strengths
NA



Comparative statistical analysis of STEM and Non-STEM students


MCA: Stream, Gender, and Opportunities
Opportunities_Interdisciplinary.applications._1
Female
Opportunities_Interdisciplinary.applications_0
Non-STEM
Opportunities_Government.job_0
Opportunities_Technological.advancements_1
Opportunities_Growing.Demand.in.the.Job.Market_1
Opportunities_Growing.Demand.in.the.Job.Market_0
Opportunities_Technological.advancements_0
Opportunities_Interdisciplinary.applications._0
STEM
Opportunities_Interdisciplinary.applications_1
Male
Opportunities_Government.job_1
4
2
0
-1.0
-0.5
0.0
0.5
1.0
1.5
Dimension 1 (23.4%)
Dimension 2 (18%)
Category
Opportunities
NA


MCA: Stream, Gender, and Threats
Threats_Job.automation._1
Threats_Lack.of.funding.or.resources_1
Female
Threats_No.one_0
Threats_Economic.downturns_0
Threats_advancements_0
Threats_Market.saturation_0
Threats_Lack.of.funding.or.resources._1
Non-STEM
Threats_Job.automation_0
Threats_Lack.of.funding.or.resources._0
Threats_Job.automation_1
Threats_Job.automation._0
Threats_Lack.of.funding.or.resources_0
STEM
Threats_Market.saturation_1
Male
Threats_Economic.downturns_1
Threats_No.one_1
Threats_advancements_1
Dimension 2 (14%)
Dimension 1 (15.7%)
3
2
1
0
-1
0
1
2
3
4
Category
Threats
NA

summarize(key_findings)

GENDER & STREAM

Significant association found. Female more likely to choose Non-STEM while Male more likely to go for STEM

MOTIVATION DIFFER

STEM: Driven by passion/ interest(intrinsic)

Non-STEM: Influenced by external factors like career prospects

PERFORMANCE

No significant difference between STEM and Non-STEM performances

STUDY HOURS & PERFORMANCE

No meaningful correlation found.

RESOURCES USED

Similar patterns across both streams, no stream exclusive usage

CO-CURRICULARS

STEM: Sports & Fitness

Non-STEM: Arts, Creativity and Social Service

CAREER GOALS

STEM leans toward research

Non-STEM prefers Industry and Entrepreneurship

SKILLS VALUED

Significant association found. Female more likely to choose Non-STEM while Male more likely to go for STEM

FIRST CHOICE AND SWITCHING

No significant difference.

ECONOMIC BACKGROUND

No significant stream-based relation found with economic background