

# Statistical Data Analysis of Student Goals

Mateusz Zaremba

November 4, 2019

## 1 Abstract

*This should be a very brief explanation of your research paper (around 150 words). It normally includes information about the issue, why you are interested in that issue, your method/model, analysis results, discussions and conclusions.*

This paper analyses the data gathered from surveying 625 undergraduate students. The authors of the survey tried to prove two hypothesis: 1) During students' junior years, they tend to primarily focus on getting good grades while during their senior years, the focus shifts towards a deep-understanding of the subject and 2) students' enjoyment and interest tends to deteriorate as they progress through their studies. It is not obvious why this might be the case and if the student's age, sex or studied subject has any bearing. This is why the survey has 15 questions and probes 7 assessment categories. Each category consists of 3 to 1 questions and because the order of the questions is randomised, the student should not know the categories nor notice any patterns.

The data manipulation was done using R and tidyverse packages. A full analysis will be presented, including data: preparation, analysis, exploration and interpretation; calculation of confidence intervals, interpretation of the results using different kinds of graphs and an explanation of the methods used.

## 2 Introduction

*This section should explain the topic, why it is important, and how you approach the issue*

It is interesting how undergraduate students' goals change through-out their studies. They often experience various syndromes like: burnout, impostor, disheartening or even attempt a suicide. A Harvard graduate, Alex Chang, in his TEDx talk titled "The Unspoken Reality Behind the Harvard Gates" speaks about the pressure of getting the best grades; how he was called for a jasmine tea to his tutor and asked if he couldn't give it his all, while he already was doing the best he could. He also recalls one tragic night when he and his roommates were woken up at 4am, to be informed that one of his friend has taken his own life.

Because this paper is going to be talking about student's course enjoyment, expectations and his or her focus on grades vs. understanding I would like to give it another, less visible shade for there might be a lot more to say about a student who is at the bottom of the scale. A student who is enjoying the course, finds it not interesting but still primarily aims to perform better than others and is led by the fear of performing poorly. We will try to find such students and calculate the confidence interval of finding a student-at-risk in a given sample.

### 3 Data

*Explain your dataset and how the data was collected – e.g. your sampling strategy or information given by the project information.*

#### 3.1 Initial Data

This is how the data looked like before we started cleaning it:

seq	year	age	sex	subject	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12
6	3	19	1	1	7	2	2	6	7	6	7	7	5	5	7	5
7	3	20	2	1	7	2	1	7	7	6	4	4	1	6	5	2
8	3	21	1	1	1	1	5	4	7	6	1	2	3	3	7	1
9	3	NA	2	1	4	2	3	3	5	4	3	7	2	2	7	2

interest	enjoy	mastgrad
7	7	1
6	6	4
7	7	1
7	7	4

#### 3.2 Cleaning the Data

First, the *seq* column was dropped since it does not serve any purpose. Second, rows with empty cells were dropped because they could falsify our results.

The following coding information was given to us:

Sex 1=Male; 2=Female

Subject 1=Management 2=Law 3=Tourism 4=General Economics 5=Accounting 6=Statistics

Around 625 students were surveyed. To conceal the existence of the 7 categories from students, they were presented with the questions in random order.

The students answered on a 7-level scale; 1 meaning the student feels the statement asked in the question is 'Not true of him/her' and 7 meaning the student feels it is 'Very true of him/her'. See the table below for a graphical explanation:

Not true of me	1	2	3	4	5	6	7	Very true of me
----------------	---	---	---	---	---	---	---	-----------------

The students were presented with the questions in a randomized order. Because the random order was known, it was assumed that the collected data was put into a table in the The

random order was known and the questions were derandomised in the data set, i.e., *question 6* from the *Performance avoidance* category is numbered *1* in the survey. In the data set results for *question 1* were renamed to results for *question 6*. *question 12* from the *Mastery avoidance* category is numbered *2* in the survey, etc. *Table 1* presents this mapping.

Table 4: In the survey, questions from 1 to 12 were ordered randomly; questions 13, 14 and 15 remained in the same order.

Survey order	6	12	11	1	7	2	10	8	5	3	9	4	13	14	15
Category mapping	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

*Table 1* presents the category, its interpretation (Based on the category’s questions) and the questions’ numbers.

Table 5: Table 1

Category	Interpretation	Question	# of Questions
Performance approach	How important it is to students to do better than others?	1, 2, 3	3
Performance avoidance	How motivated are students by fear of performing poorly?	4, 5, 6	3
Mastery approach	Prevalence of mastery approach	7, 8, 9	3
Mastery avoidance	Student’s fear of not mastering the course	10, 11, 12	3
Interest	Student expects the course to be interesting	13	1
Enjoyment	Student expects the course to be enjoyable	14	1
Importance focus	Student’s importance focus on understanding vs. grades	15	1

Because **Interest**, **Enjoyment** and **Importance focus** categories consisted only of 1 question, no further computation was required to interpret the data from these categories. For the remaining categories, which number of questions was equal to 3, a mean for each assessment category was computed, and saved for each individual student. This resulted in 4 extra columns added to the original data set. An example of these can be seen below:

m1	m2	m3	m4
4.666667	4.333333	6.000000	3.333333

m1	m2	m3	m4
2.333333	2.333333	5.000000	2.000000
3.666667	1.333333	5.666667	1.333333
3.666667	3.666667	6.000000	5.333333
3.333333	3.333333	7.000000	4.000000

## 4 Methodology

*This section explains the statistical methods and/or your model. It is also a common practice to present the statistical model structure (i.e. equation) here as well.*

Confidence interval was used.

tidyverse packages (ggplot2, dplyr, tibble, readr, tidyr, purrr);

## 5 Results

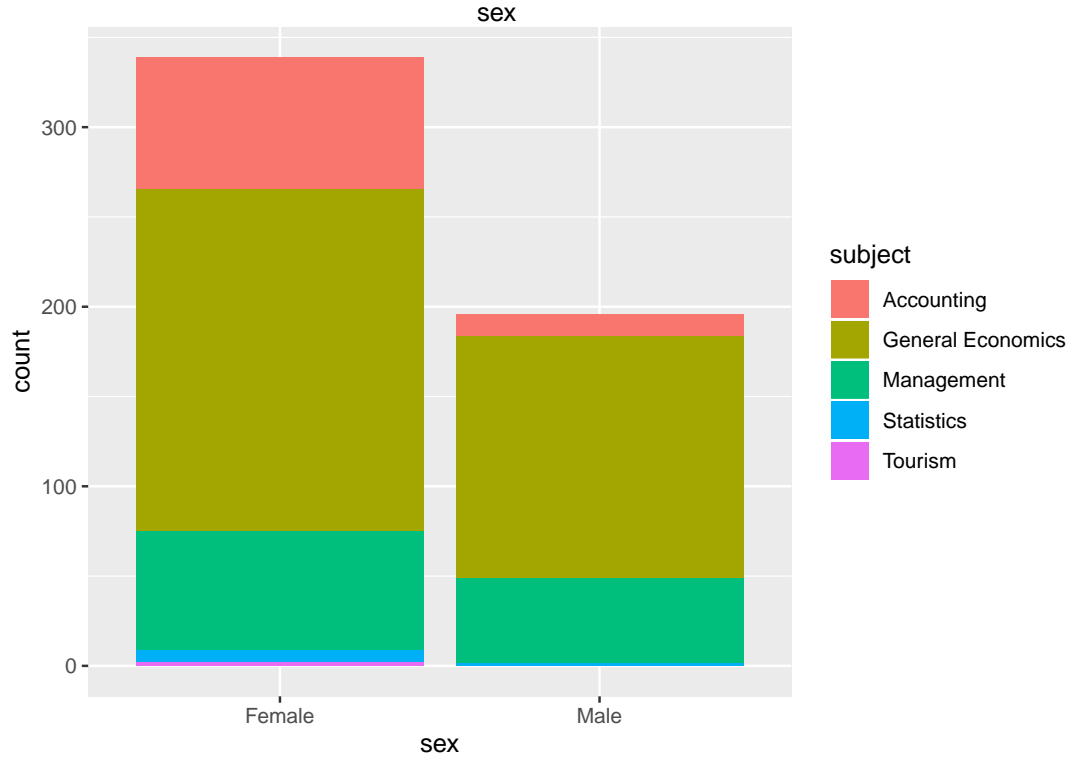
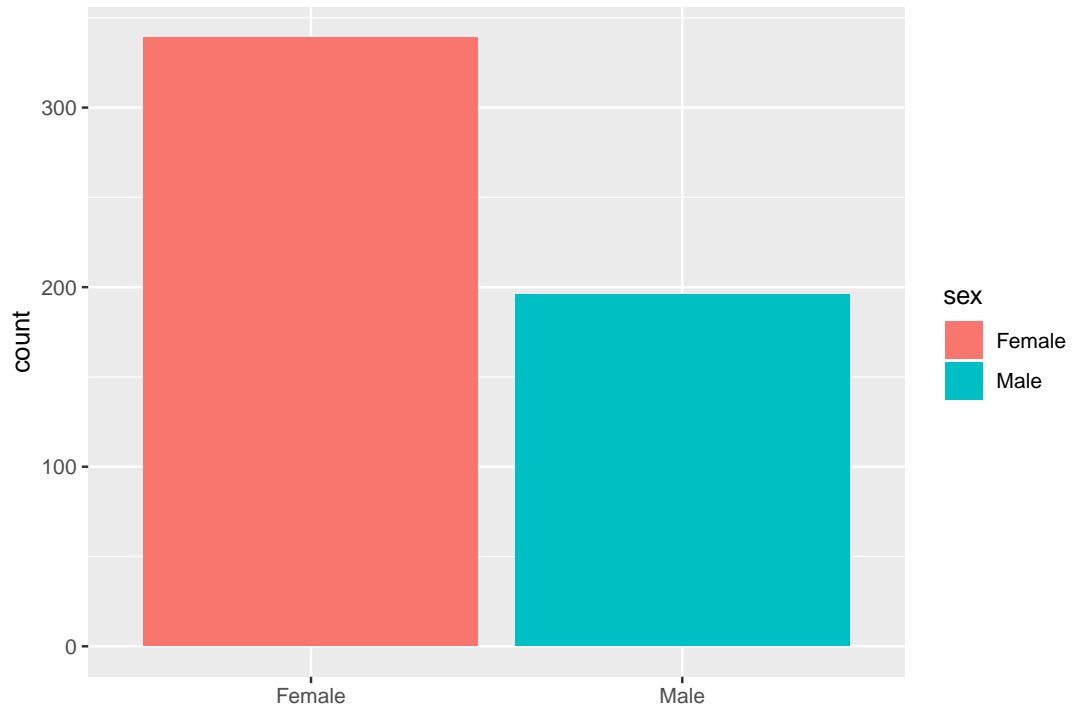
*Present both your informal and formal analyses.*

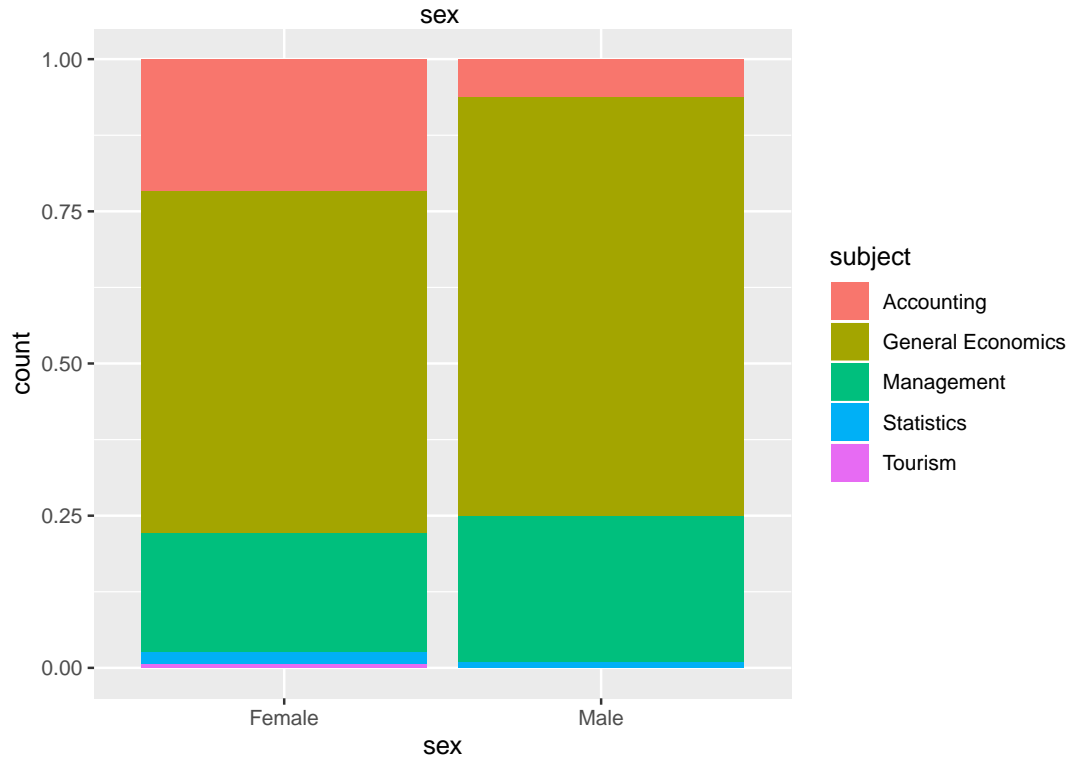
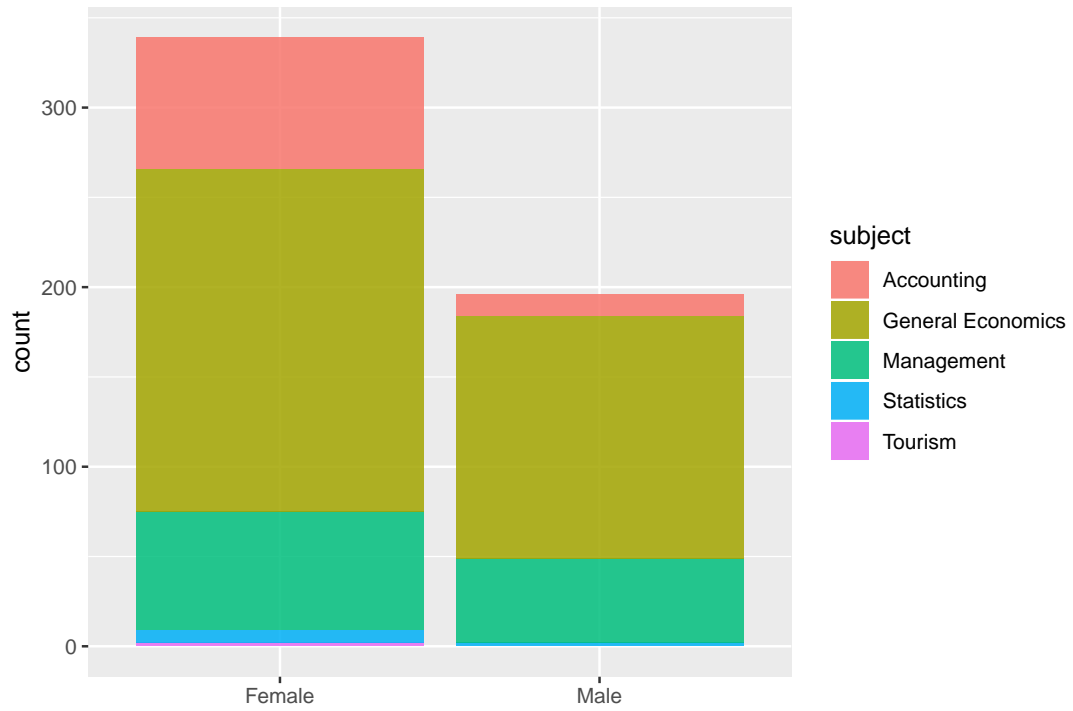
analyse the m1, m2, m3, m4 graphs bro

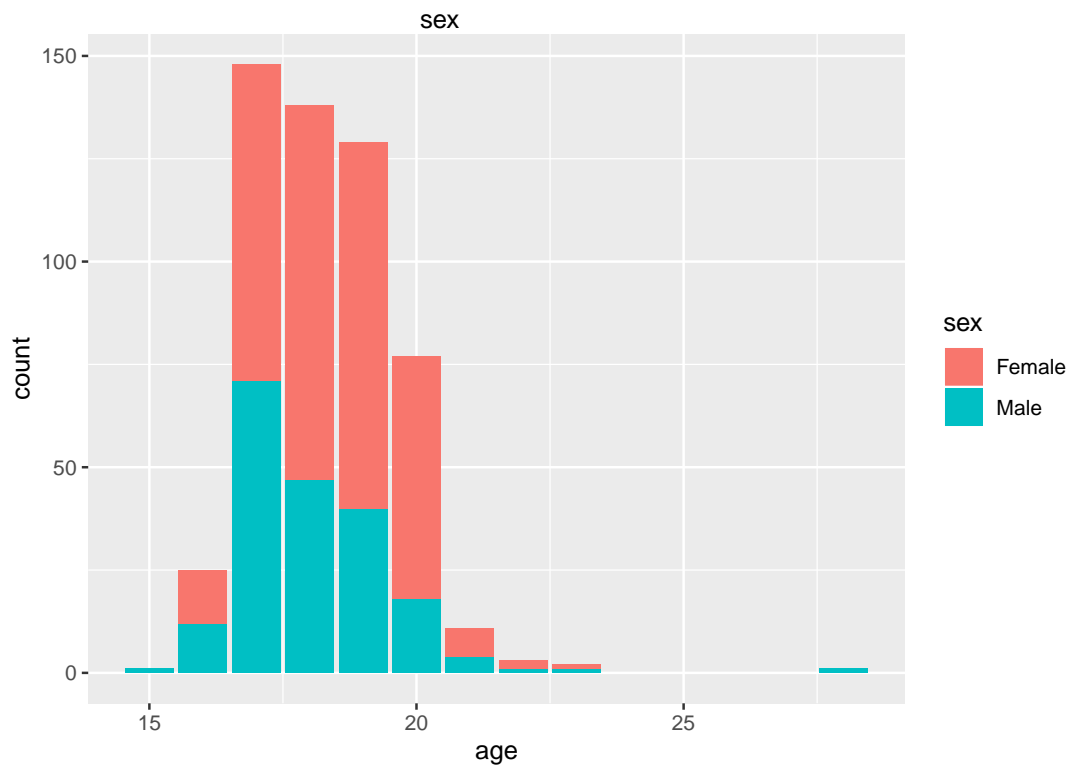
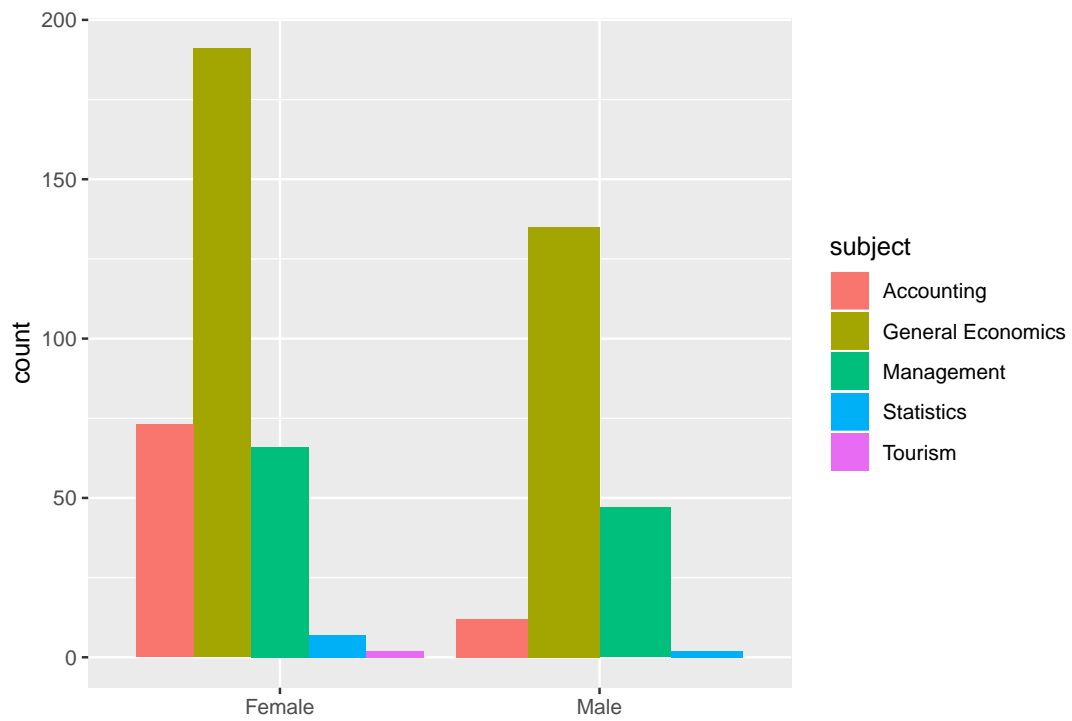
## 6 Conclusion / Discussion

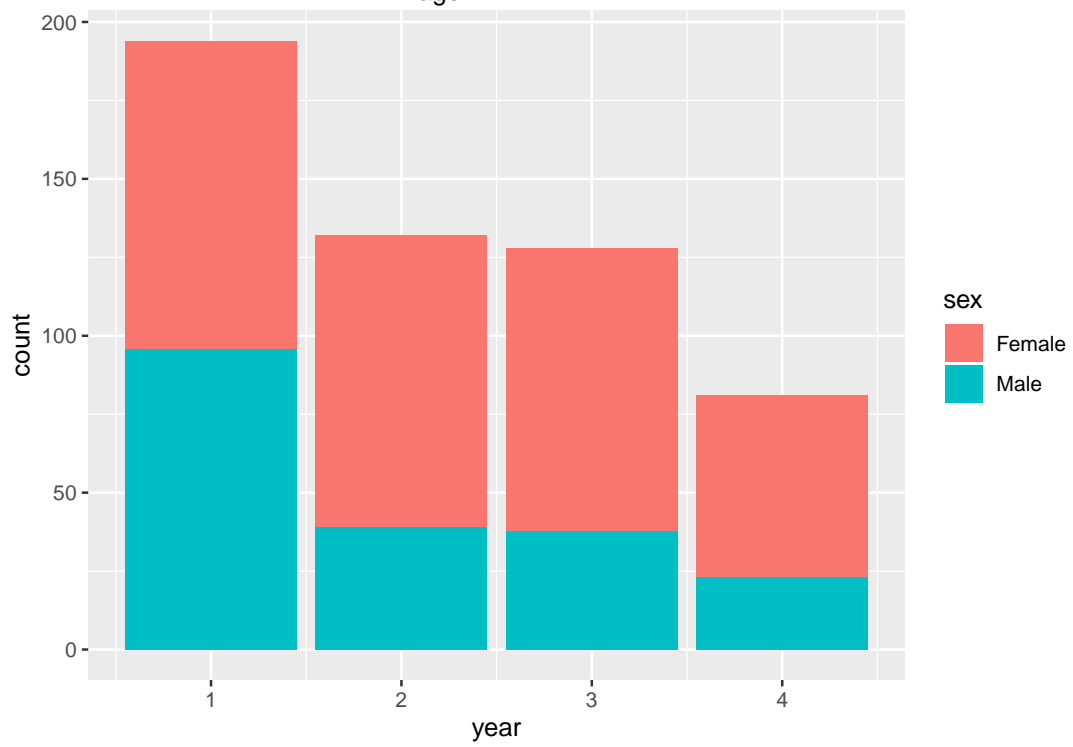
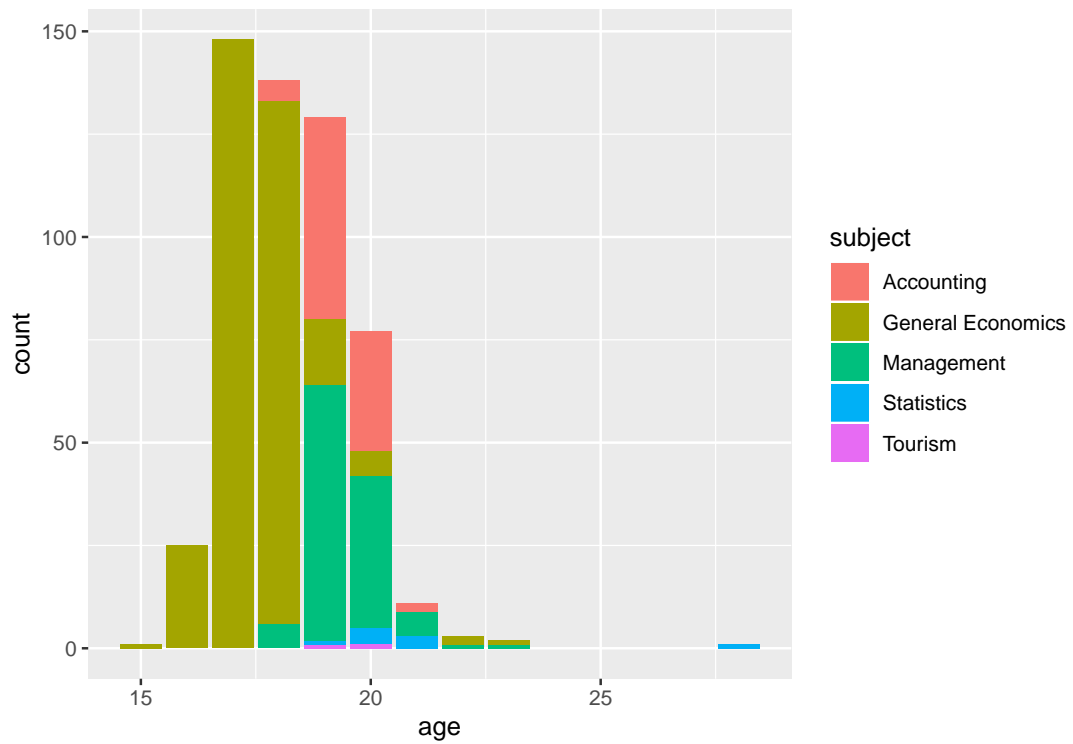
*You need to conclude your project, discuss the results, discuss any reservations that you have about the study and list any future work.*

so, what conclusions do you have bruh?

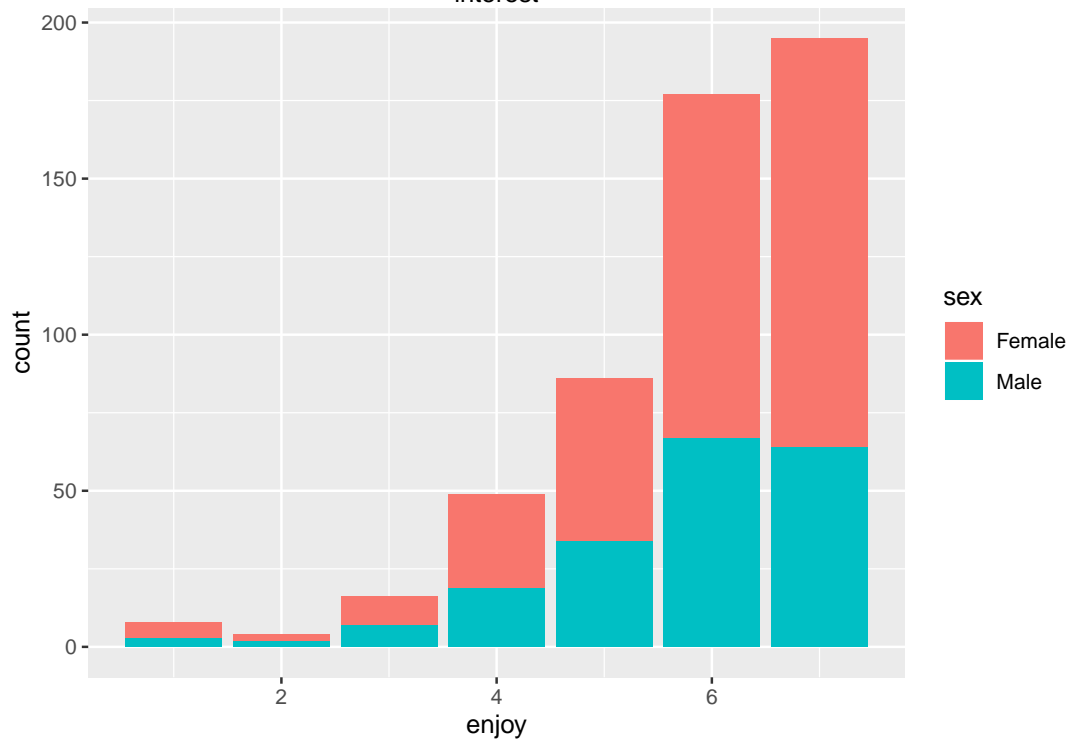
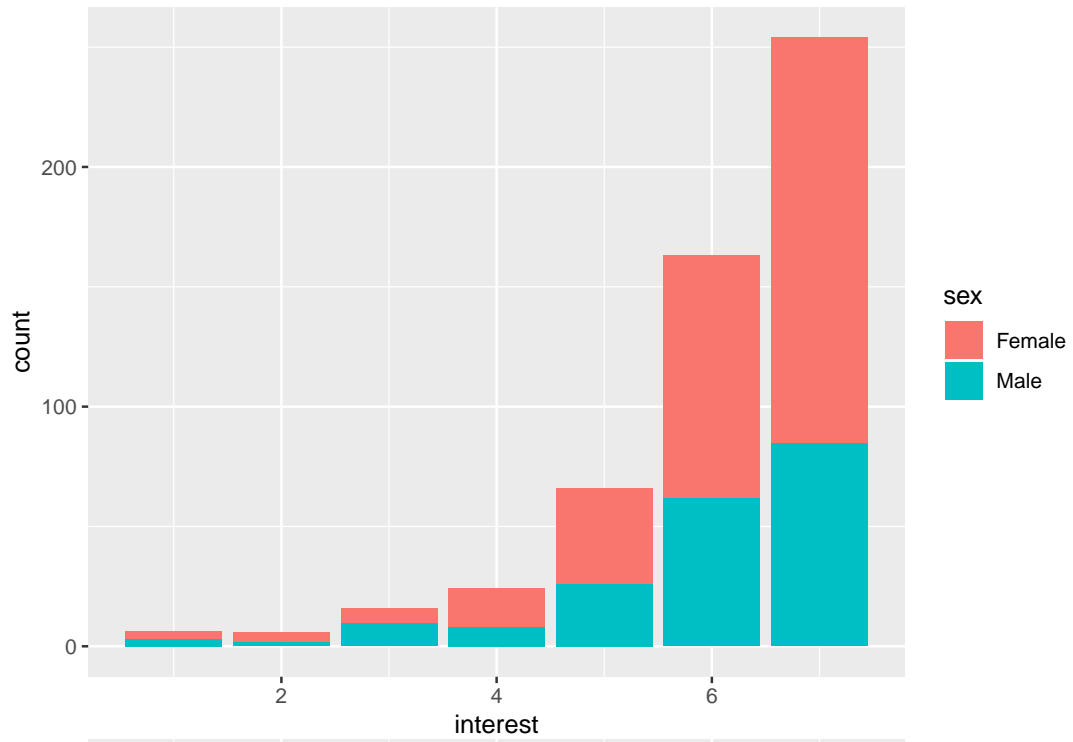


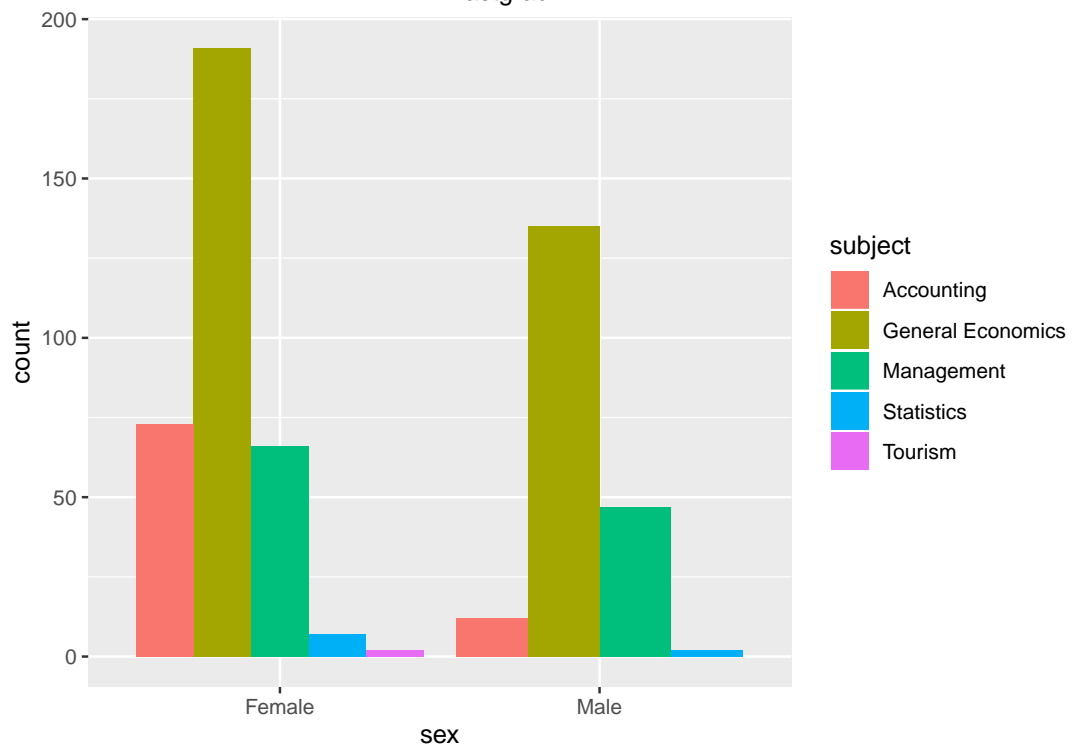
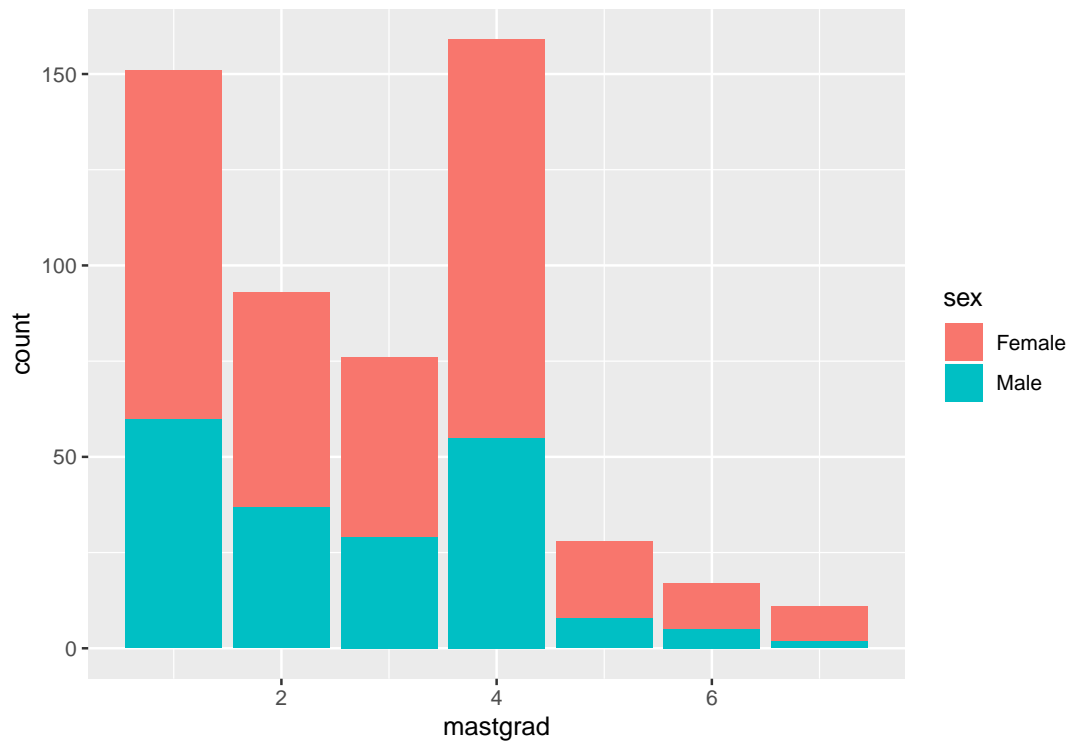


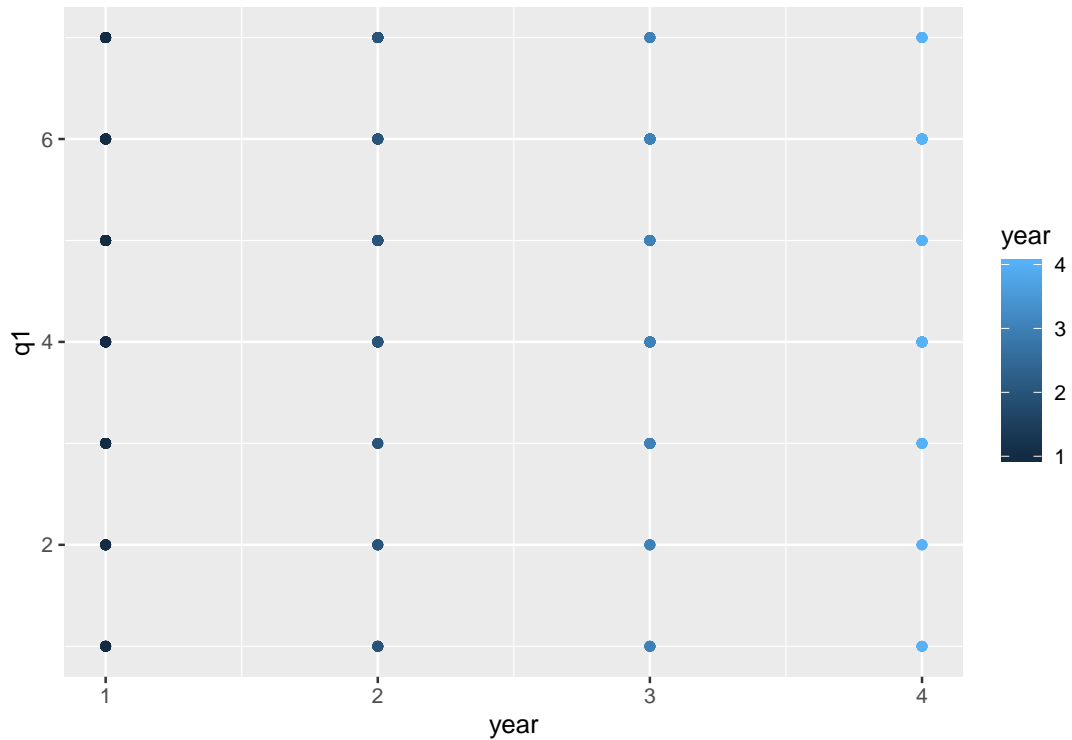






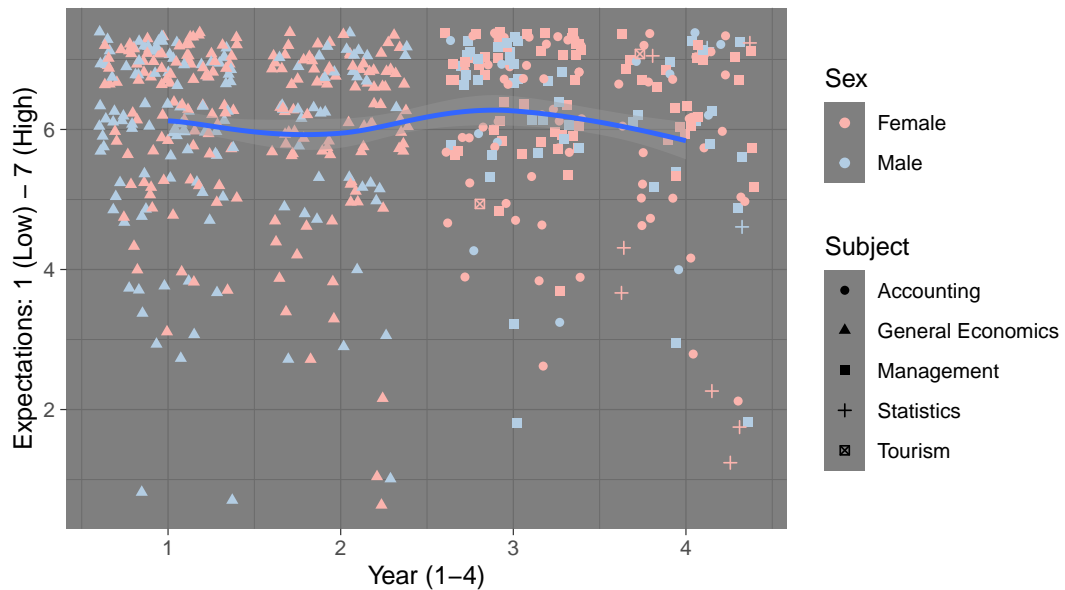






Student's course interestedness expectations set on basis of:  
different years of study, sexes and subjects.

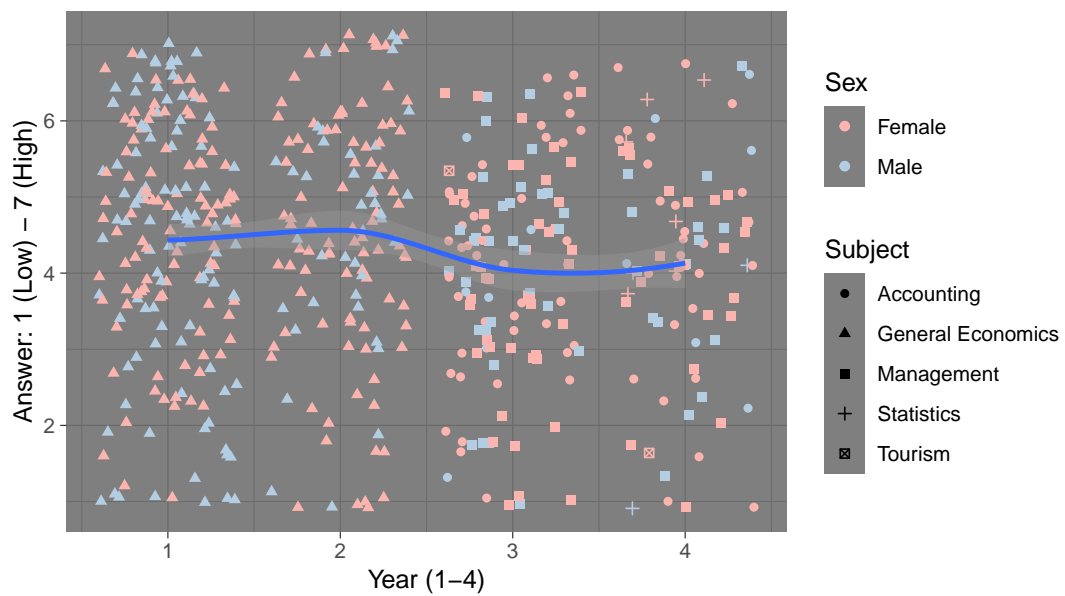
'I expect my courses this semester to be very interesting'



Data source: Elliot, A. J. and McGregor, H. A. (2001)

Student's grade-orientation focus set on basis of:  
different years of study, sexes and subjects.

How important it is to students to do better than others?

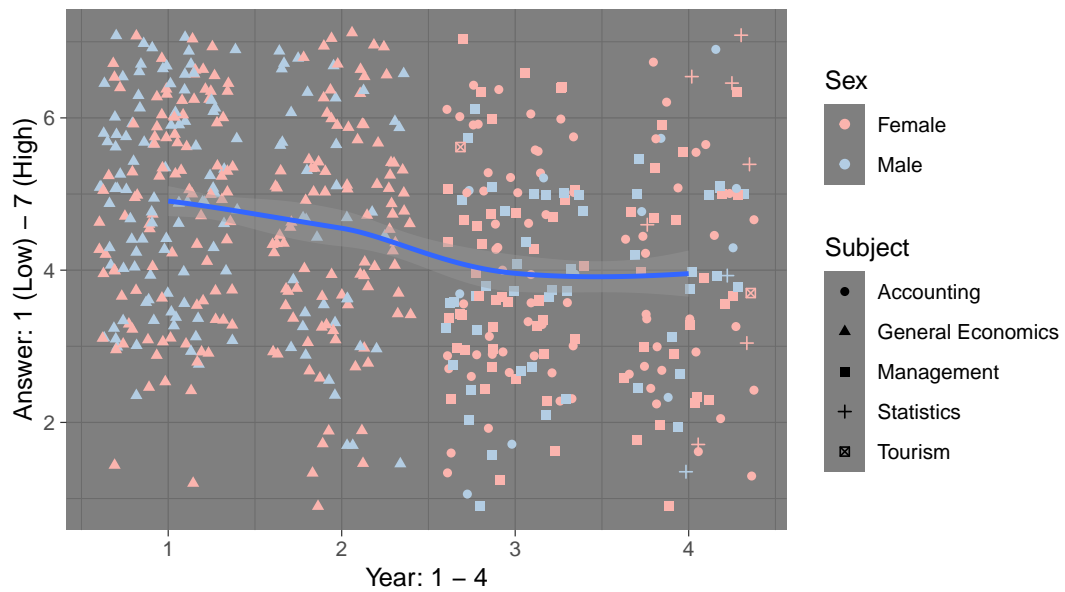


Data source: Elliot, A. J. and McGregor, H. A. (2001)

Figure 1: ...

Student's grade-orientation focus set on basis of:  
different years of study, sexes and subjects.

How motivated are students by fear of performing poorly?

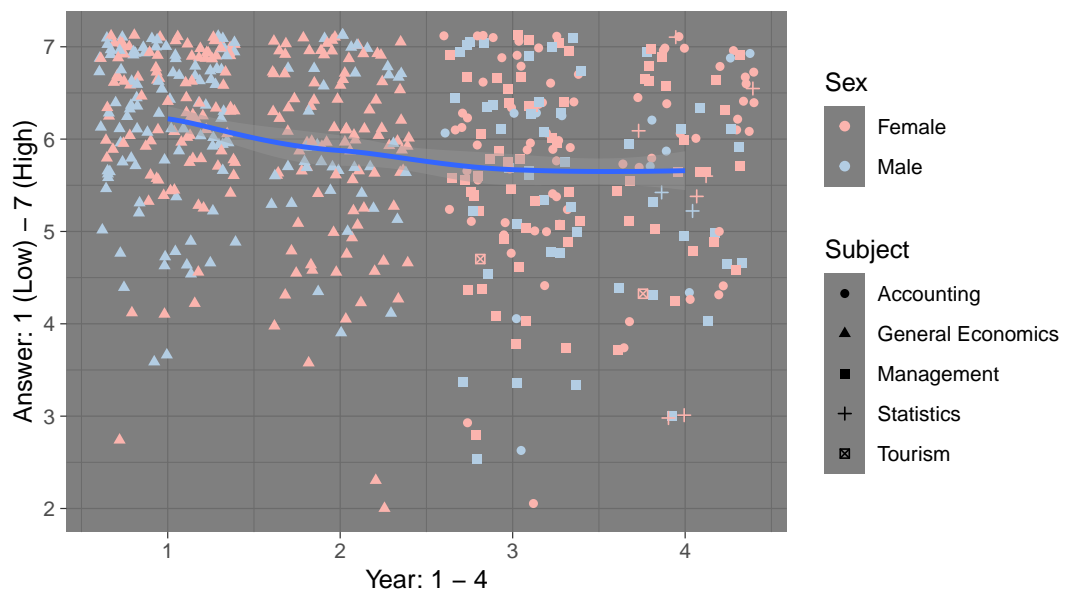


Data source: Elliot, A. J. and McGregor, H. A. (2001)

Figure 2: ...

Student's focus on understanding set on basis of:  
different years of study, sexes and subjects.

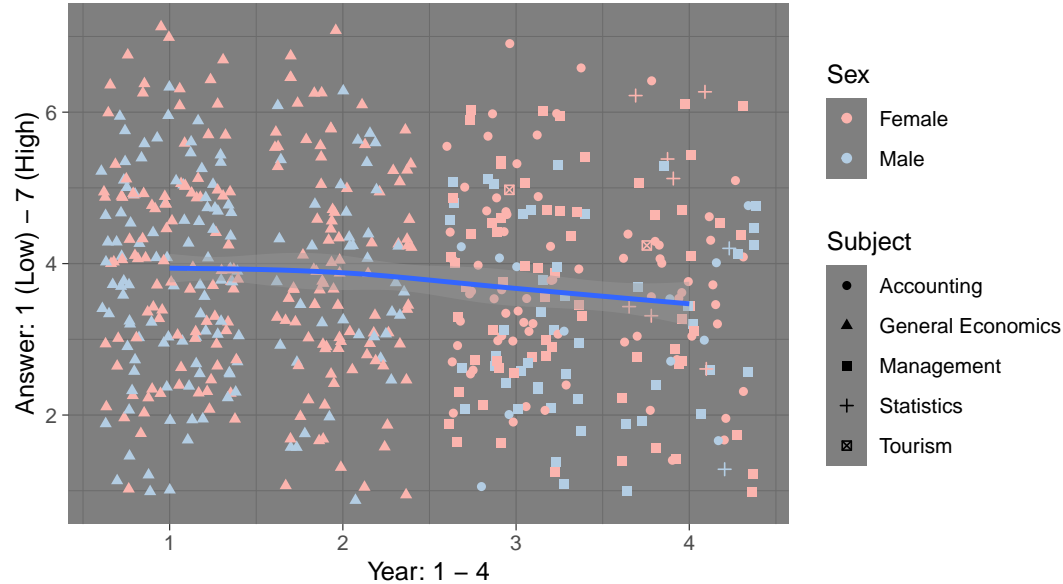
Prevalence of mastery approach.



Data source: Elliot, A. J. and McGregor, H. A. (2001)

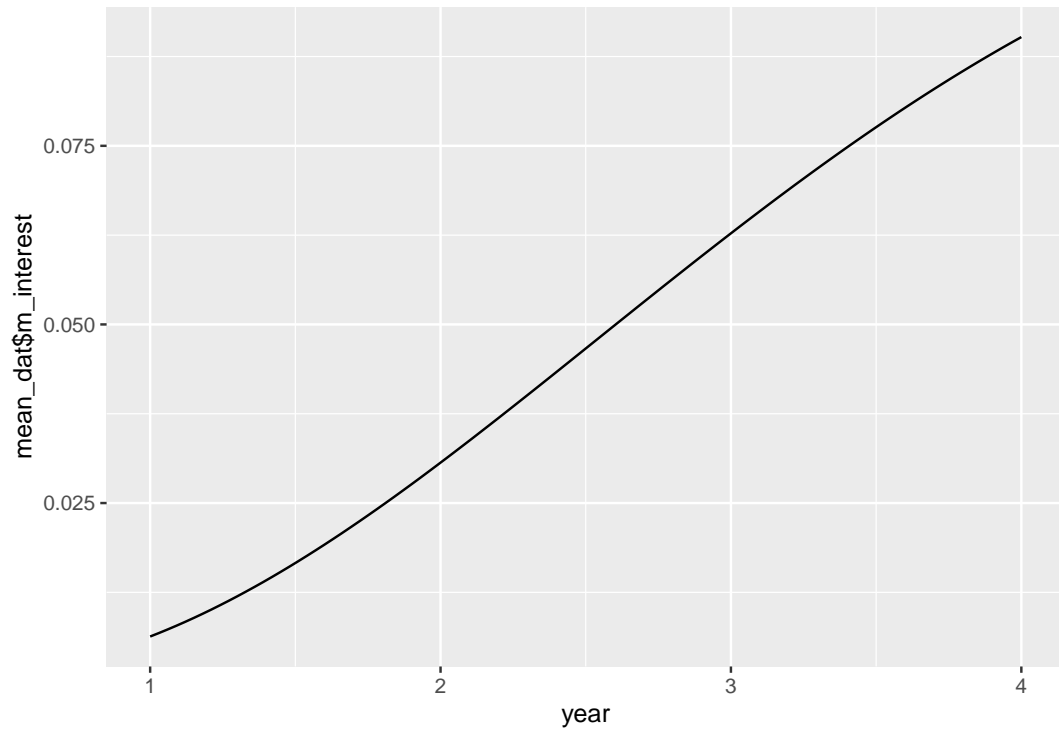
Figure 3: ...

Student's focus on understanding set on basis of:  
different years of study, sexes and subjects.  
Student's fear of not mastering the course.



Data source: Elliot, A. J. and McGregor, H. A. (2001)

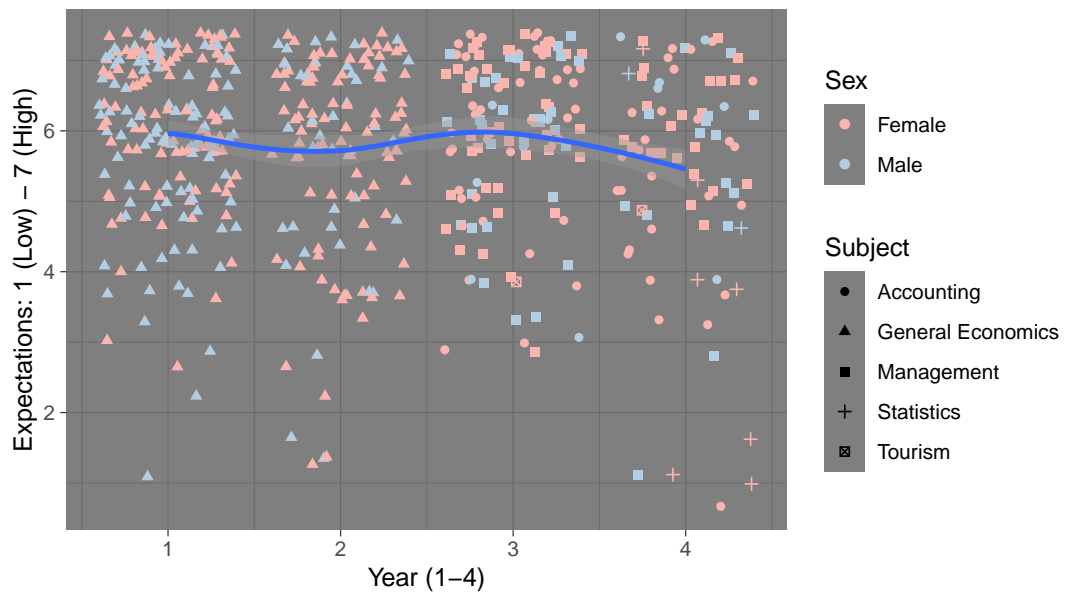
Figure 4: ...





Student's course enjoyment expectations set on basis of:  
different years of study, sexes and subjects.

'I expect my courses this semester to be very enjoyable'

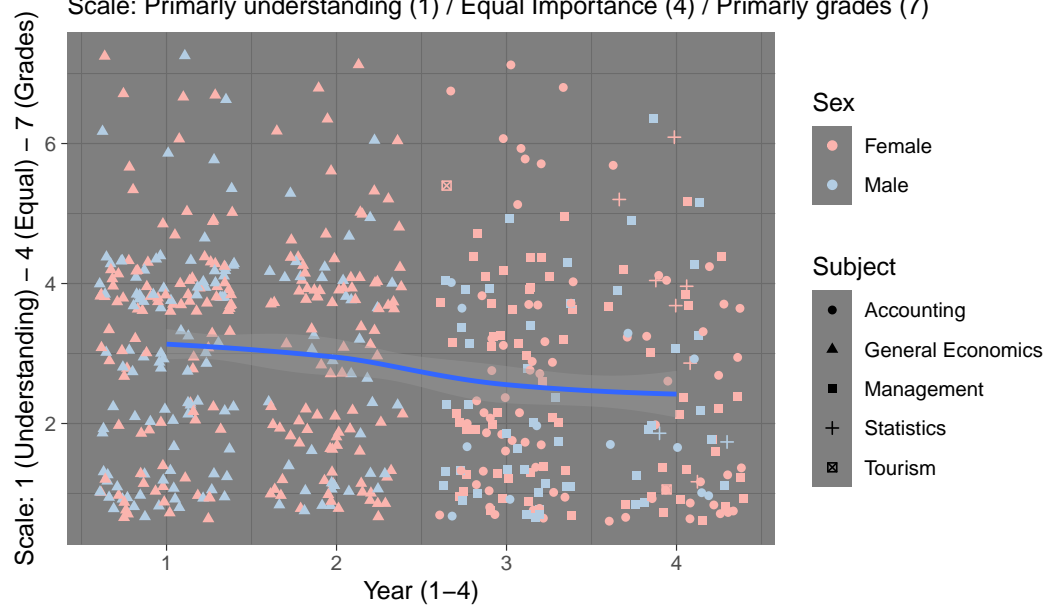


Data source: Elliot, A. J. and McGregor, H. A. (2001)

Figure 5: ...

Student's importance scale between understanding and grades set on basis different years of study, sexes and subjects.

Scale: Primarily understanding (1) / Equal Importance (4) / Primarily grades (7)



Data source: Elliot, A. J. and McGregor, H. A. (2001)

Figure 6: ...