

INFS 5116 – Data Visualisation

Constructing graphics – Examples

Data for the examples presented here is based on a sample of 103 students who participated in a study on exam anxiety. Variables are as follows:

Variable	Description
<i>revise</i>	Time spent revising (hours)
<i>exam</i>	Exam score (out of 100)
<i>anxiety</i>	Anxiety score (out of 100)
<i>gender</i>	1='Male' 2='Female'

Table 1. Variable descriptions

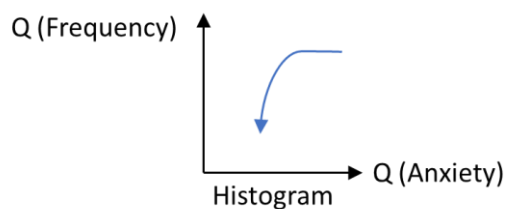
This is a simple data set that does not require any data cleaning (this should be checked and done first). In this scenario *revise*, *exam* and *anxiety* are all quantitative ratio variables (Q), while *gender* is a nominal qualitative variable (N). We can use this data set to **explore relationships** among the three quantitative variables and **check if there are differences** based on gender.

Example 1 – How much exam anxiety did the students experience?

We can address this question by examining the distribution of anxiety scores using a **histogram**.

- *Invariant*: students
- *Components*: anxiety score (Q) and frequency (Q)
- *Imposition*: orthogonal
- *Mark*: bar
- *Planar variables*: anxiety score (horizontal) and frequency (vertical)
- *Retinal variables*: position and length; colour will be chosen for aesthetic purposes only.

A schematic diagram could look like this:



This is the resulting two-component graphic:

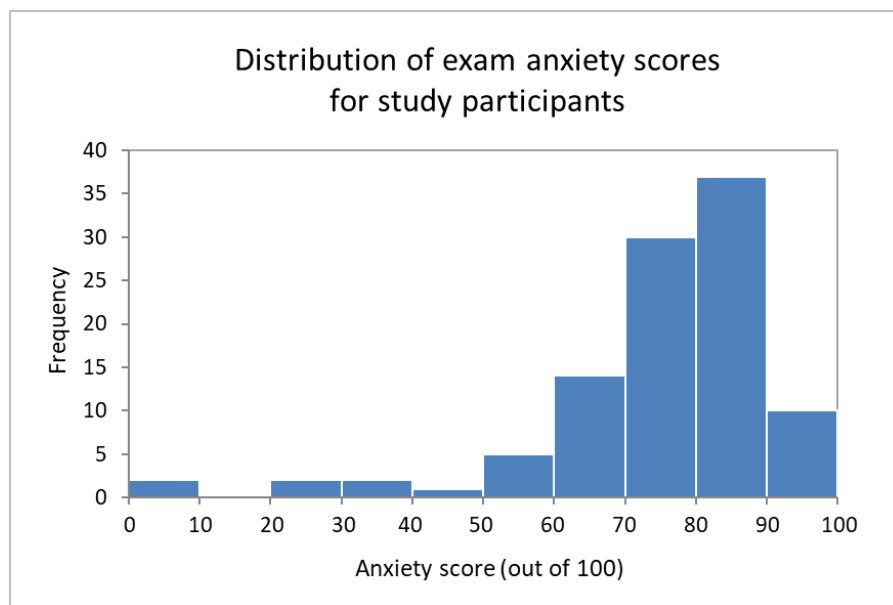


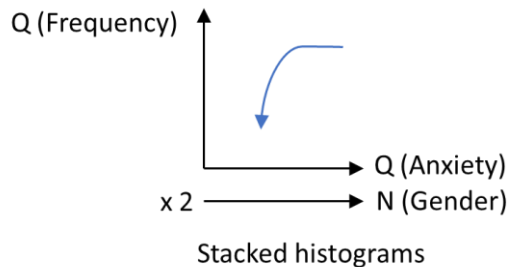
Figure 1

Questions that could be answered using Figure 1 could be as follows:

- What is the typical anxiety score? (80 to 90 points)– Elementary question
- What is the range anxiety scores for most students? (50 or higher)– Intermediate question

Example 2 – How do anxiety scores compare for male and female students?

This question can be answered using **histograms** like the one shown in Figure 1. We would just add gender (N) as the third component as follows:



The resulting graphic would look like this:

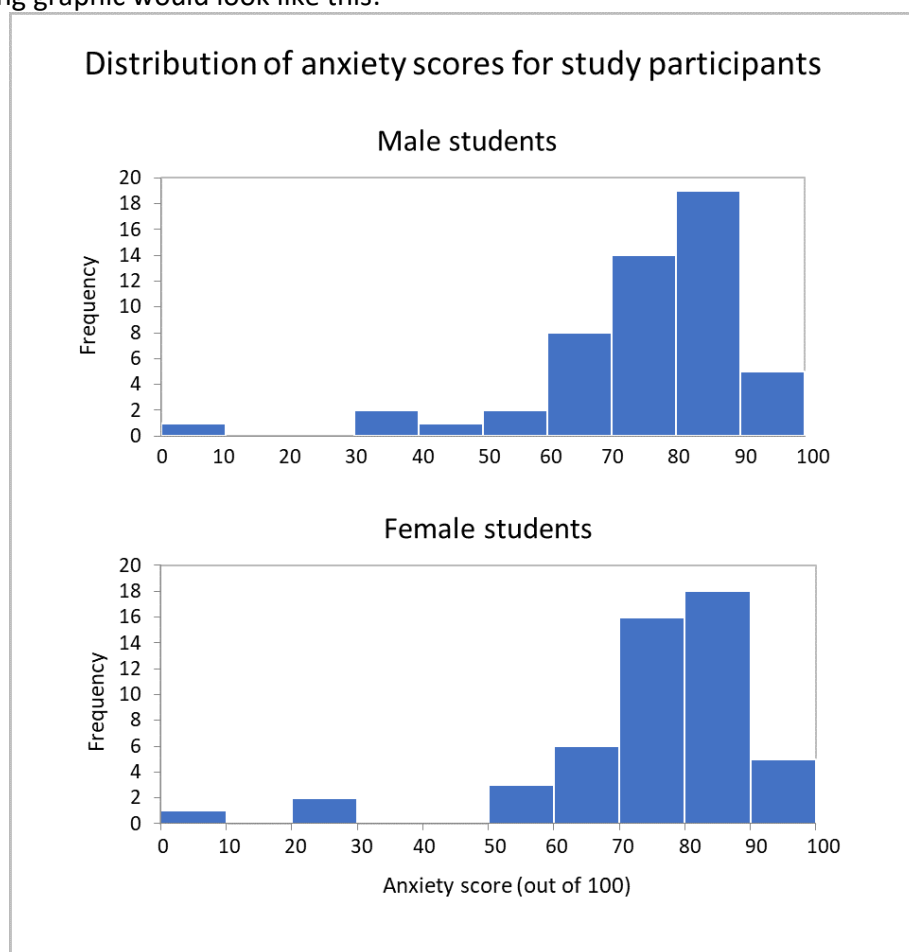


Figure 2

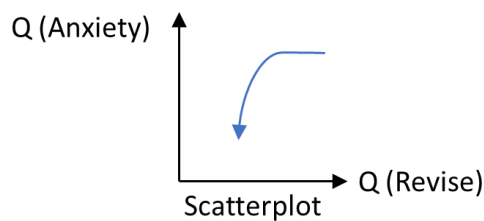
Gender is now a panel variable, and an overall question that we could answer is:

- Is there a difference in the distribution of anxiety scores for male and female students? (No significant difference)

Example 3 – Does it matter how much time was spent revising?

This question could be answered using a **scatterplot**.

- *Invariant*: students
- *Components*: anxiety score (Q) and time spent revising (Q)
- *Imposition*: orthogonal
- *Mark*: circle
- *Planar variables*: time spent revising (horizontal) and anxiety score (vertical)
- *Retinal variables*: position; colour will be chosen for aesthetic purposes only.



The resulting graphic would look like this:

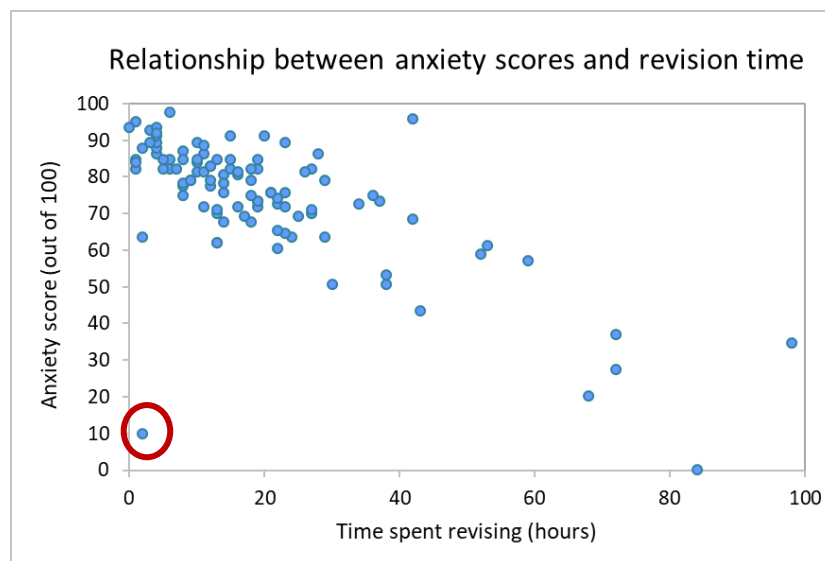


Figure 3

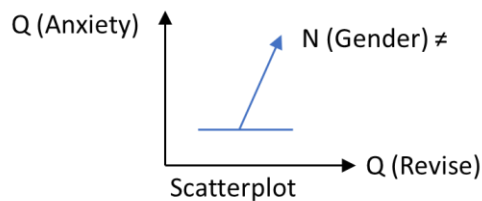
There are a number of elementary level questions that can be addressed using this scatterplot, mostly to do with marks that stand out from the rest, for example the mark that has been circled in red.

- What was the anxiety score of the student who did virtually no revision? (10 points) – *Elementary question*
- What is the relationship between time spent revising and anxiety scores? (Students who spent more time revising generally had lower anxiety scores) – *Overall question*
- Does spending more time revising have a positive effect on anxiety scores? (Yes) – *Overall question*

Example 4 – Was there a gender difference in the relationship between exam results and anxiety levels?

This question can be answered using a **scatterplot** with three components:

- *Invariant*: students
- *Components*: anxiety score (Q), exam score (Q) and gender (N)
- *Imposition*: orthogonal
- *Mark*: circle
- *Planar variables*: anxiety score (horizontal) and exam score (vertical)
- *Retinal variables*: position and colour hue (to differentiate marks based on gender).



The resulting graphic is shown in Figure 4.

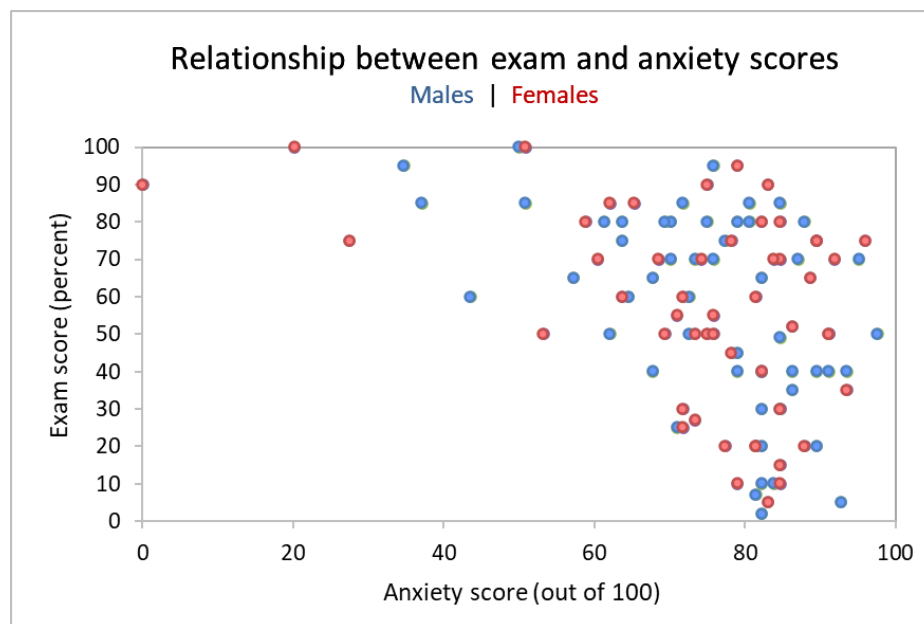


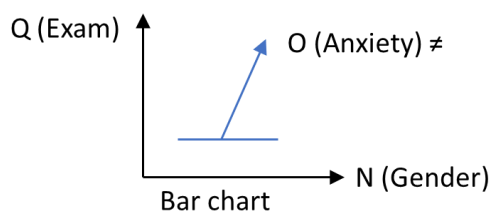
Figure 4

Overall level question could whether there are any obvious differences between male and female students. The answer is there aren't, the relationship between exam anxiety and exam performance appears to be the same for male and female students. It is a complex relationship in that there were students with high anxiety scores who performed poorly in the exam but also those who performed very well.

Example 5 – How do exam scores compare based on anxiety levels and gender?

We are going to create a **bar chart** to try and address this question, and to do so we are going to create a new ordinal categorical variable based on anxiety scores. This new variable will have three levels, *low*, *medium* and *high*. Anxiety scores classified as 'low' are those below the lower quartile (Q1), 'medium' anxiety scores lie between the lower quartile (Q1) and the upper quartile (Q3), and 'high' anxiety scores are values above the upper quartile (Q3).

- *Invariant*: students
- *Components*: anxiety level (O), exam score (Q) and gender (N)
- *Imposition*: orthogonal
- *Mark*: bar
- *Planar variables*: mean exam score (vertical) and gender (horizontal)
- *Retinal variables*: position, length and colour saturation (to differentiate marks based on anxiety levels).



This is the resulting graphic:

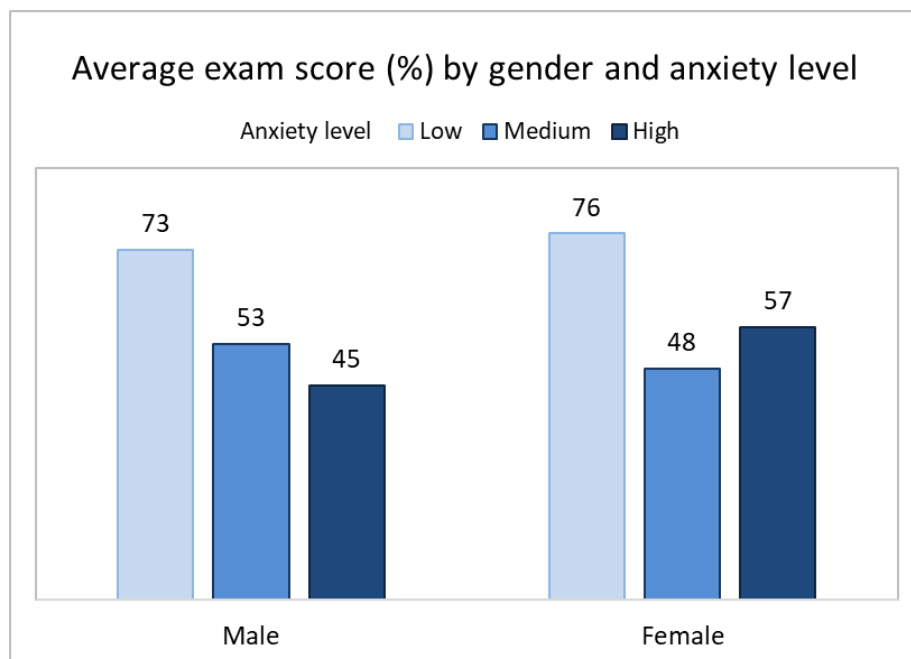


Figure 5

Possible questions are:

- What is the relationship between average exam score and anxiety level for male students? (Average exam score decreases as the anxiety level increases) – *Intermediate level question*

- Is there a difference in the relationship between average exam scores and anxiety levels for male and female students? (Yes) – *Overall level question*

Are these effective and efficient graphics? Is there a way to improve them? Are there other graphics that could be designed to represent this data? You be the judge!