<u>Create a Frequency (Pivot) Table of the Qualification and Gender variables.</u> <u>Compare the modal Qualification for each Gender.</u>

The mode can be defined as the most frequently occurring individual observation For females the most frequently occurring qualification is school as this appears 35 times within the sample of female participants

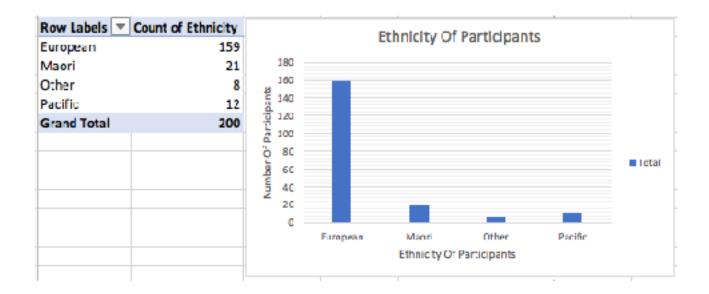
Whereas for males the most frequently occurring qualification is vocational as this appears 35 times within the sample of male participants

This tells us the most common qualification gained by females is school whereas the most common qualification gained by males is vocational

Count of Gend	ler Column Labe	ls ▽		
Row Labels	▼ Female		Male	Grand Total
Degree		16	16	32
None		28	14	42
School		35	26	61
Vocational		30	35	65
Grand Total		109	91	200

Draw a suitable graph of the Ethnicity variable, and comment on what it shows

The mode can be defined as the most frequently occurring individual observation. The mode for the ethnicity of the participants in the sample is European with 159 of 200 participants being of European ethnicity. This indicates that the participants in this sample is of European descent - the amount of European participants in the sample compared to other ethnicities may indicate bias in the sampling method of the data



<u>Draw boxplots of Hours Worked by Qualification.</u> Ensure the ordinal nature of <u>Qualification is reflected in the graph.</u> Use your graph to compare the hours worked <u>for the four groups, i.e. explain what the graph shows.</u>

This is the middle value of the sample, with 50% of the data less than the median, and 50% of the data greater than the median

Comparing the median of the 4 qualification groups we can see that school vocational and degree have an average hours worked of 40 hours whereas none qualifications have an average hours worked of 36.5.

This indicates that school vocational and degree qualification group tend to work slightly longer hours than groups with no qualification although the small difference of 3.5hours does not make this a strong claim

lower and upper quartiles

Lower quartile: average of lowest 50% Upper quartile: average of highest 50%

In the lower quartile We can observe that the none group has the lowest number of hours worked (16.75 hours) when compared to the school (18.5 hours) vocational (17.5 hours) and degree (26.75) group which means that on average the lowest 50% of the none group works 0.75 hours less than the lowest 50% of any other group

In the upper quartile We can observe that the school group has the lowest number of hours worked (42.5 hours) when compared to the none (44.25 hours) vocational (45 hours) and degree (45 hours) which means that on average the highest 50% of the school group works 1.75 hours less than the highest 50% of any other group

minimum and maximums

Minimum: lowest value Maximum: highest value

The school vocational and degree group all have a minimum number of hours worked of 2 hours whereas the none group has a minimum of 4 hours

The degree school and vocational group both have the same maximum of 70 hours but it appears that the maximum of the degree group is an outlier which means that if we were to exclude this data the degree's group maximum value would be 50 hours making it the lowest in the group but with the data included in the sample the none group would have the lowest maximum of 60 hours



Calculate the mean and standard deviation of the Income data, and the 85th percentile. What does the latter number describe about the incomes?

A percentile indicates a particular value's position in the sample

The 85th percentile is the value for which 85% of the sample values are less than this value So 85% of the sample makes an income below \$952

Calculate the mean and standard deviation of the Income variable for males and females separately. (Hint: consider using the Sort functionality) Draw boxplots of the Income variable by Gender. Do the means and standard deviations agree with the information shown by the boxplots? Explain.

F Mean 424.266055 M Mean 756.549451 F median 390 M median 710

The mean and median are different but is in a relatively similar range - this is caused by the different method of calculation and the outliers within the sample. Therefore whilst the value of the mean doesn't correlate with those on the graph we can see in the small difference in the outcome that the data is presented accurately

	Question 4	Questi	on 5		
Income		nale mean "			
mean	575.455		424.266055		
Standard	tem				
deviation		roland International			
854h	348.375149 dox	le mean	343/1		
percentile	952	756.54	10AE1		
percente	932	_	19-131		
		ndard			
	dec	istion 324.83	35352		
	Inco	me by Gende	er		
1800					
1000	-		-		
1400					
1200			2		
800					
_	- 8				
600					
400			~		
200	_				
	Male		Female		
		Gender			
	-				
	Females Ma	les			
upper qua		952			
lower quar	t 228.5	515			
minimum	11	107			
maximums		1614			
median	390	710			

F Std 291.534371 M Std 324.835352 F IQR 544- 228.5=315.5 M IQR 952- 515=437

The female STD and IQR is different but in a relatively similar range whereas the Male STD and IQR is significantly different - the outcome has likely been effected by the outliers in both samples. Therefore the STD doesn't agree with the graph but would likely be more similar if outliers are removed

Draw a histogram of the Income variable. Summarise the sample distribution. If the histogram is bimodal, can you explain the source of this?

The histogram is unimodal with a majority of people in the sample earning between \$431-\$641, the graph is skewed to the right which indicates that a majority of sample lie towards a lower income range

