Analysing Quantitative Data Computer Lab 2

Welcome to the second quantitative data analysis lab for Social Research Methods. This workbook will talk you through the activities you should undertake in the second lab session. Please work through these activities at your own pace. Please feel free to work in pairs and to discuss the activities with your classmates. Teaching staff will be on hand to answer any questions, please do not hesitate to ask for assistance.

In this session you will learn how to construct contingency tables. You will also practice some data management tasks (i.e. recoding variables).

You can download SPSS to use on your own computer here. If you have any difficulty downloading SPSS you should seek assistance from the IT Services Help Desk.

Section 1: Revision

- Open SPSS.
- Open the SRM Practice Data. These data can be downloaded from the VLE. Remember to open data you select: **FILE**, **OPEN**, **DATA**.
- Open the Lab 2 Syntax File. This file can be downloaded from the VLE. Remember that to open a syntax file you select: **FILE**, **OPEN**, **SYNTAX**.

If you cannot remember how to do these things take a look back at last week's workbook.

The syntax file contains all the code you will be asked to type throughout this workbook. Work through the syntax file in conjunction with the workbook. Remember that to 'run' a piece of code you highlight it and then click on the green triangle. Your results will appear in the Output Viewer window.

Last week you produced univariate descriptive statistics. These are simple summaries of single variables.

To warm up produce some descriptive statistics of the age variable using the code below. Highlight the code in your syntax file and click on the green triangle. SPSS will print a table with the results in the Output Viewer window.

```
FREQUENCIES age
  /FORMAT NOTABLE
  /STATISTICS=MEAN MEDIAN MODE RANGE MIN MAX.
```

Produce some descriptive statistics for the categorical variable 'walkdark' (How safe do you feel walking alone after dark?). Appropriate statistics to describe a categorical variable are the percentage of people in each category. SPSS can provide this information if you run the code below. Highlight this code in your syntax file and click on the green triangle. SPSS will print a table with the results in your Output Viewer window.

```
FREQUENCIES walkdark.
```

Repeat this process to examine the categorical variable 'cinema' (How often have you visited the cinema in the last year?).

```
FREQUENCIES cinema.
```

Section 2: Contingency Tables

So far we have only looked at one variable at a time. Most often we are interested in how different variables (i.e. different pieces of information about people) relate to each other. In this module we will undertake analyses which examine the relationship between two variables (i.e. bivariate data analysis), but much more complex analyses are possible.

To examine the relationship between two categorical variables you can use a contingency table. You can produce a contingency table using the code below. Here we investigate the association between 'sex' and 'work' (Did respondent do any paid work in last week?). Run this code by highlighting it and clicking on the green triangle.

```
CROSSTABS sex BY work
/CELLS count row.
```

The output SPSS prints in the Output Viewer in response to this code is shown below. The first table 'Case Processing Summary', tells you how many people were included in this analysis. If respondents have missing information for either of these two variables they will not be included in your contingency table. You can see that 11,640 people are included in your table, 36 people are not included because they have missing data.

The second table is your contingency table. The percentages are calculated across the rows. You can see that 61.7% of men did paid work last week, and 38.3% of men did not do paid work. You can also see that 48.5% of women did paid work last week, and 51.5% did not do paid work.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Respondent's Sex * Did respondent do any paid work in last week?	11640	99.7%	36	0.3%	11676	100.0%

Respondent's Sex * Did respondent do any paid work in last week? Crosstabulation

			Did respondent do any paid work in last week?		
			1 yes	2 no	Total
Respondent's Sex	1 male	Count	3269	2027	5296
		% within Respondent's Sex	61.7%	38.3%	100.0%
	2 female	Count	3076	3268	6344
		% within Respondent's Sex	48.5%	51.5%	100.0%
Total		Count	6345	5295	11640
		% within Respondent's Sex	54.5%	45.5%	100.0%

Decimal Places

When reporting results such as these there is no need to present percentages beyond whole numbers. We would round these numbers up and report that 62% of men did paid work last week, and 49% of women did paid work last week.

This might lead to what are called 'rounding errors' for example if we look at women we would report that 49% of women did paid work, and 52% of women did not do paid work. You will see that these two figures add up to 101%. This is not an error per se, it is just a consequence of rounding up these values.

You may then ask, why do we not just present our values including the decimal place? It is important not to present numbers which suggest more precision than they offer, in social science the things we measure are rather 'fuzzy' and prone to a little error. Given this fact, presenting percentages as whole numbers is most appropriate and conventional in social science research.

Table Order

Imagine we wanted the contingency table to be presented the other way round (i.e. we want the 'work' variable to form the rows of our contingency table, and the 'sex' variable to form the columns of our variable). We can prepare this table by just switching the order of the variables in the code. Try this, the code required is provided in the syntax file.

You should choose the order you enter your two variables based on your research question. For example in relation to these two variables we would most likely to asking whether 'sex' was associated with doing paid work. We might be interested in women's attachment to the labour market. Our question might be 'Is sex associated with participating in paid work?' It would make less sense to investigate whether participating in paid work was associated with 'sex', as paid work would not explain someone's 'sex'.

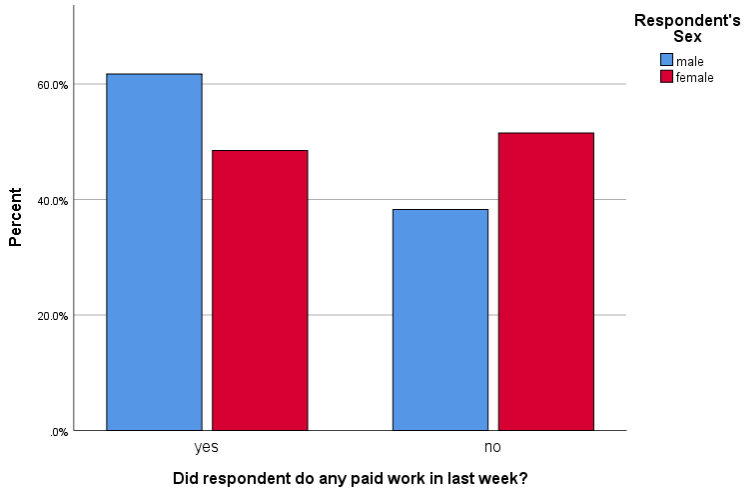
The variable that appears first in our research question should form the rows of the contingency table. You should carefully consider which way round you present your contingency table in relation to the focus of your research.

Graph of Contingency Table

Tables are not always the best way to see patterns in data. Presenting results as graphs can be very helpful. You can produce a bar chart of your contingency table using the code below. Note where the variable names for 'sex' and 'work' are placed in this code (i.e. the opposite way round than when we produce the contingency table). Highlight this code in your syntax file and click on the green triangle.

```
GRAPH
  /BAR(GROUPED)=PCT BY work BY sex.
```

The graph will be printed in your Output Viewer window (shown below).



This graph is suited to the research question 'Is sex associated with participating in paid work?'. You can see that 62% of men did paid work last week compared to 49% of women. You can also see that 38% of men did not do paid work compared to 52% of women. If you wanted your code to represent this pattern 'the other way round', you just need to change the order of your variable 'work' and 'sex' in the code. Give this a try and take a look at the slightly different graph which is produced (the code required is provided in the syntax file).

Section 3: Recoding Categorical Variables

Example 1

The code below produces a contingency table for the variables 'class' (Respondent's social class identity) and 'cinema' (How often have you visited the cinema in the last year?). You can see that 'class' is first so it will form the rows. Run this code and look at the results in your Output Viewer.

```
CROSSTABS class BY cinema
/CELLS count row.
```

Here is your contingency table of 'class' and 'cinema'. You can see that 13% of people who identify as working class visit the cinema every week, compared to 16% of people who identify as middle class, and 11% of people who identify as upper class.

Respondent's Social Class Identity * How often have you visited the cinema in the last year? Crosstabulation

			How often have you visited the cinema in the last year?					
			1 Every Week	2 Once a Month	3 Once Every Few Months	4 Once Every Six Months	5 Less than Once Every Six Months	Total
Respondent's Social Class Identity	1 Working Class	Count	691	1219	2656	512	119	5197
		% within Respondent's Social Class Identity	13.3%	23.5%	51.1%	9.9%	2.3%	100.0%
	2 Middle Class	Count	157	319	395	95	12	978
		% within Respondent's Social Class Identity	16.1%	32.6%	40.4%	9.7%	1.2%	100.0%
	3 Upper Class	Count	329	828	1544	256	34	2991
		% within Respondent's Social Class Identity	11.0%	27.7%	51.6%	8.6%	1.1%	100.0%
Total		Count	1177	2366	4595	863	165	9166
		% within Respondent's Social Class Identity	12.8%	25.8%	50.1%	9.4%	1.8%	100.0%

If you look at the column 'Less than once every six months' you will see that only a total of 165 participants fall in this category. Whilst 165 is not necessarily a small number this is much smaller than the other categories. We might consider grouping this category with the category 'Once every six months'. This would simplify our table, there are no hard and fast rules for when a researcher should do this, it is a decision the researcher has to make guided by theoretical reasoning. If you decided to group categories together you would justify why you did this when writing up your results.

When we group together categories this is called 'recoding' a variable. Before you recode a categorical variable you should examine it using the FREQUENCIES code (see below). Run this code and look at the results in the Output Viewer window.

```
FREQUENCIES cinema.
```

The output is shown below. We can see that this variable has 5 categories and the final category 'Less than Once Every Six Months' was chosen by only 2% of respondents (1.5% really, but remember we are rounding up). We are going to group this category with the category 'Once Every Six Months', this will create a new category which represents 'Once Every Six Months or Less'.

How often have you visited the cinema in the last year?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Every Week	1229	10.5	12.9	12.9
	2 Once a Month	2471	21.2	25.9	38.7
	3 Once Every Few Months	4770	40.9	49.9	88.7
	4 Once Every Six Months	909	7.8	9.5	98.2
	5 Less than Once Every Six Months	171	1.5	1.8	100.0
	Total	9550	81.8	100.0	
Missing	999	2126	18.2		
Total		11676	100.0		

This is the code required to recode this variable. This code is telling SPSS to recode the variable 'cinema' by turning category 5 into category 4. In other words everyone who were in the category 5 box of the survey question are moved into the category 4 box of the survey question. Highlight this code in your syntax file and click on the green triangle.

```
RECODE cinema (5=4).
```

If you look in your Output Viewer it might seem that SPSS has done nothing other than print the code you have just run. This is because we are not asking SPSS to make any tables for us we are just asking it to change this variable. To see what SPSS has done you can run the FREQUENCIES code for the variable again. Highlight the code below in your syntax file and click on the green triangle.

```
FREQUENCIES cinema.
```

If you look at the output (shown below) you will see that the category 'Less than Once Every Six Months' is no longer there and that these respondents have been moved into category 'Once Every Six Months'.

How often have you visited the cinema in the last year?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Every Week	1229	10.5	12.9	12.9
	2 Once a Month	2471	21.2	25.9	38.7
	3 Once Every Few Months	4770	40.9	49.9	88.7
	4 Once Every Six Months	1080	9.2	11.3	100.0
	Total	9550	81.8	100.0	
Missing	999	2126	18.2		
Total		11676	100.0		

There is one problem. The category 'Once Every Six Months' now also contains people who visit the cinema 'Less than Once Every Six Months' but this is not reflected in the label for this category. We can change this label using the code below. This code specifies how each of the categories for the variable 'cinema' should be labelled. You can keep the labels for the first three categories the same, but the final category is changed to reflect the recoding. Highlight and run this code.

```
VALUE LABELS cinema
  1 "Every Week"
  2 "Once a Month"
  3 "Once Every Few Months"
  4 "Once Every Six Months or Less".
```

Again, if you look in the Output Viewer window you will not see anything other than this code printed. This is because we have not asked SPSS to prepare any tables, we are just asking SPSS to change the value labels for the 'cinema' variable.

You can check that a change has been made by running FREQUENCIES again for the cinema variable.

```
FREQUENCIES cinema.
```

The output is shown below. You can see that the cinema variable has only four categories and these are now appropriately labelled.

How often have you visited the cinema in the last year?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Every Week	1229	10.5	12.9	12.9
	2 Once a Month	2471	21.2	25.9	38.7
	3 Once Every Few Months	4770	40.9	49.9	88.7
	4 Once Every Six Months or Less	1080	9.2	11.3	100.0
	Total	9550	81.8	100.0	
Missing	999	2126	18.2		
Total		11676	100.0		

Now you have recoded the 'cinema' variable you can produce a contingency table again. You use the same code again to produce a contingency table of 'class' and 'cinema'. Highlight this code in the syntax file and run it.

```
CROSSTABS class BY cinema
  /CELLS count row.
```

The contingency table which is printed in the Output Viewer is shown below. There is now one less category for cinema, this simplifies the table. We can see that overall 11% of respondents visit the cinema once every six months or less. We can see that 12% of respondents who identify as working class vising the cinema once every six months or less, compared to 11% of respondents in the middle class category and 10% of respondents in the upper class category.

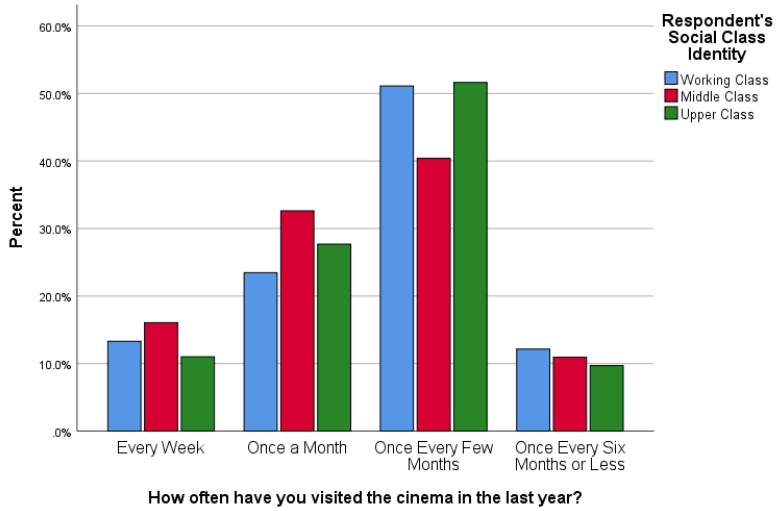
Respondent's Social Class Identity * How often have you visited the cinema in the last year? Crosstabulation

			How often have you visited the cinema in the last year?				
			1 Every Week	2 Once a Month	3 Once Every Few Months	4 Once Every Six Months or Less	Total
Respondent's Social Class Identity	1 Working Class	Count	691	1219	2656	631	5197
		% within Respondent's Social Class Identity	13.3%	23.5%	51.1%	12.1%	100.0%
	2 Middle Class	Count	157	319	395	107	978
		% within Respondent's Social Class Identity	16.1%	32.6%	40.4%	10.9%	100.0%
	3 Upper Class	Count	329	828	1544	290	2991
		% within Respondent's Social Class Identity	11.0%	27.7%	51.6%	9.7%	100.0%
Total		Count	1177	2366	4595	1028	9166
		% within Respondent's Social Class Identity	12.8%	25.8%	50.1%	11.2%	100.0%

Tables are not always the best way to see patterns in data. Presenting results as graphs can be helpful. You can produce a bar chart of your contingency table using the code below. Note where the variable names for 'cinema' and 'class' are placed in this code. Highlight this code in your syntax file and click on the green triangle.

```
GRAPH
  /BAR(GROUPED)=PCT BY cinema BY class.
```

In your output viewer you will see that this code has instructed SPSS to prepare the graph below.



Example 2

There are many ways you can recode categorical variables. Sometimes you don't want to group together different categories, you just want to get rid of a category. For example if someone answered a question and chose a response such as 'prefer not to answer' or 'don't know', this may not convey any useful information for the research question at hand. You may decide to exclude people who gave this response from your analysis. This is a decision which the researcher makes based on theoretical reasoning, and if this decision was made it should be explained when writing up the research.

Here we consider the variable 'social' (How frequently do you use social media?). Imagine you only want to study people who use social media. Examine this variable using the FREQUENCIES command, highlight the code below in your syntax file and click on the green triangle.

```
FREQUENCIES social.
```

SPSS prints the table below in the Output Viewer window. You can see that 1466 respondents (13%) do not use social media. If you wanted to conduct a piece of research which focussed only on people who use social media you would want to get rid of this category. We can do this by telling SPSS that these people should be considered missing.

How frequently do you use social media?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Daily	1369	11.7	11.7	11.7
	2 Several Times a Week	3742	32.0	32.1	43.8
	3 Less Frequently	5093	43.6	43.6	87.4
	4 Does Not Use	1466	12.6	12.6	100.0
	Total	11670	99.9	100.0	
Missing	999	6	.1		
Total		11676	100.0		

SPSS already knows that the value 999 should be considered missing. We can see that in the table above SPSS identifies 6 values as missing because they have the value 999. We want to tell SPSS that as well as the value 999, the value 4 (i.e. the 'Does Not Use' category) should also be considered missing. We can tell SPSS this using the code below. Highlight this code in the syntax file and click on the green triangle.

```
MISSING VALUES social (999 4).
```

If you look in the Output Viewer window it appears that SPSS has not done anything. But you have not asked SPSS to create any tables you have asked it to make a change to the 'social' variable. To look at the changes we can examine the 'social' variable using the FREQUENCIES command. Highlight the piece of code below in the syntax file and click on the green triangle.

```
FREQUENCIES social.
```

The output SPSS prints following this command is shown below. You can see that SPSS now considers category 4 'Do Not Use' of the variable 'social' as missing. You will see that category 4 is no longer included in the calculation of 'Valid Percent'. If you produce a contingency table of the 'social' variable and another variable, category 4 will not be included, try this.

How frequently do you use social media?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Daily	1369	11.7	13.4	13.4
	2 Several Times a Week	3742	32.0	36.7	50.1
	3 Less Frequently	5093	43.6	49.9	100.0
	Total	10204	87.4	100.0	
Missing	4 Does Not Use	1466	12.6		
	999	6	.1		
	Total	1472	12.6		
Total		11676	100.0		

Of course in many circumstances keeping the category 'Does not Use' would be completely appropriate for this variable, this was just an example of the recoding process.

Section 4: Recoding Scale Variables

Now we will examine the variables 'age' and 'benefits' (Agreement with Statement: 'Cutting benefits damages too many people's lives'). If you take a moment to look at the information describing these variables under the Variable View tab of your Data Editor, you will see that age is a scale variable. This presents a problem for producing a contingency table as this technique is used to examine two categorical variables.

This is not a major problem as scale variables can all be turned into categorical variables. The ages of the respondents in our sample vary between 16 and 101 years, and the median age is 49 years. There is no hard a fast rule as to how scale variables should be categorised, this a decision made by the researcher. The decisions made should be sensible given the variable at hand. For example if most people were aged in their 20s and very few people were aged over 30 in your sample it would be sensible to create an 'old category' for those aged 30+. In this data set there are many people aged over 30 so this would not be a sensible decision. Whatever decision is made it is important that the researcher describes and justifies the decision they have made when writing up their research. You can also take guidance from other published studies when making decisions such as these.

In our case we will recode age into three categories:

- Category 1 "Young" – Respondents age 16 to 35.
- Category 2 "Middle" – Respondents age 36 to 59.
- Category 3 "Old" – Respondents age 60 and over.

This decision was made by examining the mean, median, mode and range of these data. As the ages of our respondents have a wide range we are recoding them into three simple groups that capture the overall variation in ages parsimoniously.

We are going to create a new variable called 'agegrp' (age group), and we are going to place everyone aged 16 to 35 in category 1, everyone age 36 to 59 in category 2, and everyone age 60 and over in category 3. The code to do this is shown below. This is the most complex code you will be shown in this module. Before you run it we will look at what each line does (below).

```
COMPUTE agegrp = age.
RECODE agegrp (16 THRU 35 =1) (36 thru 59 = 2) (60 thru 101 = 3).
VARIABLE LABELS agegrp 'Age Group'.
VALUE LABELS agegrp
   1 "Young"
   2 "Middle"
   3 "Old".
MISSING VALUES agegrp (999).
VARIABLE LEVEL agegrp (ORDINAL).
```

The first line starting COMPUTE creates a new variable called 'agegrp' which is exactly the same as the variable 'age'. This allows us to make the new 'agegrp' variable whilst keeping he age variable in the data set unchanged.

The second line starting RECODE puts the values of age into three categories. People age from 16 to 35 are put in category 1, people aged from 36 to 59 are put in category 2, and people aged 60 to 101 (the oldest age) are put in category 3.

The third line starting VARIABLE LABELS provides a label for the variable 'agegrp'.

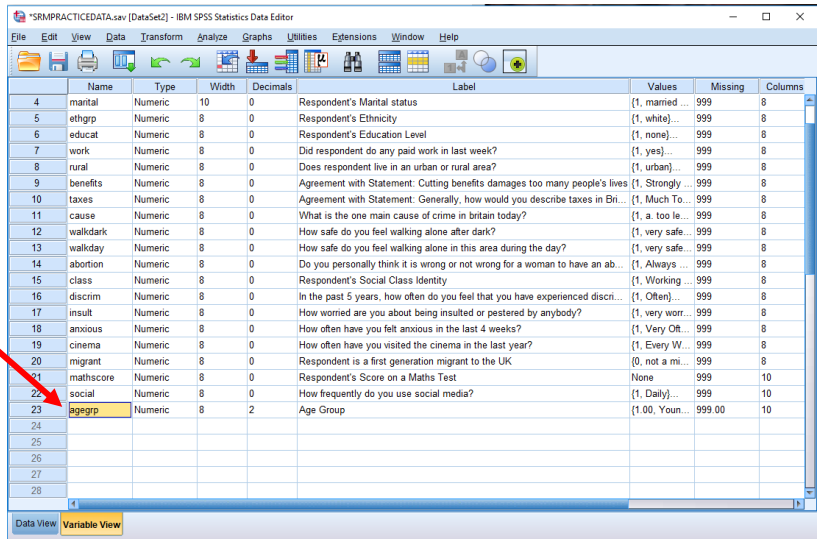
The next line starting VALUE LABELS tells SPSS what each of the new categories for this variable are called.

The line starting MISSING VALUES tells SPSS what the missing values are for the variable 'agegrp'.

Finally, the line starting VARIABLE LEVEL tells SPSS what type of variable this is (i.e. it is ordinal).

Take a look in the Variable View tab of your Data Editor. You will see that all of the details you specified above are details that will be filled into the columns in Variable View. Run the code above by highlighting it in your syntax and clicking on the green triangle.

After you have run this code you will see that no tables are printed in your Output Viewer, this is because you are just making changes to a variable and not asking for any tables to be made. Take a look in the Variable View of your Data Editor Window. Look at the bottom of the sheet and you will see the new 'agegrp' variable has been added (see below). All the details you can see in the Variable View window for this variable will match what you have asked SPSS to input for you.



Now you can examine the new 'agegrp' variable using the FREQUENCIES code. Highlight the code below in your syntax file and click on the green triangle.

```
FREQUENCIES agegrp.
```

The output printed in the Output Viewer window is shown below. You can see that in the new categorical age group variable 24% of respondents are in the 'Young' category, 41% of respondents are in the 'Middle' category, and 34% of respondents are in the 'Old' category.

Age Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Young	2852	24.4	24.5	24.5
	2.00 Middle	4797	41.1	41.1	65.6
	3.00 Old	4012	34.4	34.4	100.0
	Total	11661	99.9	100.0	
Missing	System	15	.1		
Total		11676	100.0		

Now we will be able to examine the variables 'agegrp' and 'benefits' using a contingency table. The code to produce this contingency table is shown below. Run this code and take a look at the output.

```
CROSSTABS agegrp BY benefits
/CELLS count row.
```

The output is shown below. We can see that 3% of young people strongly disagree that 'cutting benefits damages too many people's lives', 17% disagree, 44% agree and 36% strongly agree. Amongst 'old' people, 1% strongly disagree with the statement, 7% disagree, 34% agree and 58% strongly agree.

Age Group * Agreement with Statement: Cutting benefits damages too many people's lives Crosstabulation

			Agreement with Statement: Cutting benefits damages too many people's lives				
			1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree	Total
Age Group	1.00 Young	Count	85	461	1194	964	2704
		% within Age Group	3.1%	17.0%	44.2%	35.7%	100.0%
	2.00 Middle	Count	80	503	1669	2307	4559
		% within Age Group	1.8%	11.0%	36.6%	50.6%	100.0%
	3.00 Old	Count	39	278	1288	2185	3790
		% within Age Group	1.0%	7.3%	34.0%	57.7%	100.0%
Total		Count	204	1242	4151	5456	11053
		% within Age Group	1.8%	11.2%	37.6%	49.4%	100.0%

What you might notice about this table is that one category of the 'benefits' variable has only a relatively small number of respondents. We might be concerned that our conclusions concerning people who 'strongly disagree' are based on too few respondents.

To solve this problem we could also recode the 'benefits' variable to group together 'strongly disagree' and 'disagree' responses. The code below is telling SPSS to recode the variable 'benefits' by turning category 1 into category 2. In other words everyone who were in the category 1 box of the survey question are moved into the category 2 box of the survey question. Highlight this code in your syntax file and click on the green triangle.

```
RECODE benefits (1=2).
```

You will also need to update the value labels for the benefits variable to reflect the fact that category 2 now includes both strongly disagree and disagree. Highlight this code in your syntax file and click on the green triangle.

```
VALUE LABELS benefits
  2 "Strongly Disagree & Disagree"
  3 "Agree"
  4 "Strongly Agree".
```

Remember that there will be nothing to see in your Output Viewer window, but you can see the changes you have made to be benefits variable using the FREQUENCIES command.

```
FREQUENCIES benefits.
```

The table below will be printed in your Output Viewer window. You can see there are now only three categories.

Agreement with Statement: Cutting benefits damages too many people's lives

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 Strongly Disagree & Disagree	1448	12.4	13.1	13.1
	3 Agree	4154	35.6	37.5	50.6
	4 Strongly Agree	5463	46.8	49.4	100.0
	Total	11065	94.8	100.0	
Missing	999	611	5.2		
Total		11676	100.0		

Now we can produce the contingency table again, this time with the simplified 'benefits' variable. This can be done by running the code below.

```
CROSSTABS agegrp BY benefits
/CELLS count row.
```

The contingency table is shown below. We can see that 20% of young people disagree with the statement, 13% of respondents in the 'middle' age group disagree with the statement, and 8% of respondents in the 'old' group disagree with the statement.

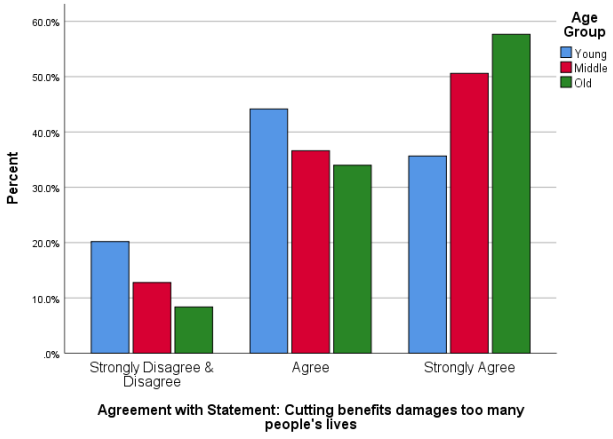
Age Group * Agreement with Statement: Cutting benefits damages too many people's lives Crosstabulation

			Agreement with Statement: Cutting benefits damages too many people's lives			
			2 Strongly Disagree & Disagree	3 Agree	4 Strongly Agree	Total
Age Group	1.00 Young	Count	546	1194	964	2704
		% within Age Group	20.2%	44.2%	35.7%	100.0%
	2.00 Middle	Count	583	1669	2307	4559
		% within Age Group	12.8%	36.6%	50.6%	100.0%
	3.00 Old	Count	317	1288	2185	3790
		% within Age Group	8.4%	34.0%	57.7%	100.0%
Total		Count	1446	4151	5456	11053
		% within Age Group	13.1%	37.6%	49.4%	100.0%

You can also make a graph of this contingency table using the code below.

```
GRAPH
  /BAR(GROUPED)=PCT BY benefits BY agegrp.
```

Here is the graph SPSS will produce (below). Overall you can see that respondents to the survey tend to agree with this statement. You can also see that older respondents are the most likely to strongly agree with the statement compared to the other age groups, and younger respondents are the most likely to strongly disagree or disagree compared to the other age groups.



Section 5: Independent Practice

We could simplify the 'benefits' variable further by grouping together the 'Agree' and 'Strongly Agree' categories. Try to recode the variable again to group together these two categories. Use the code from when we recoded the variable previously and adapt it. Remember to carefully check details in your code such as spelling and where the full stop is placed.

When you have recoded the 'benefits' variable produce the contingency table of 'class' and 'benefits' again, and produce the graph of the contingency table.

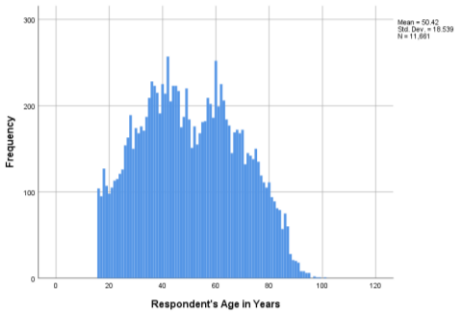
Pick some further variables from the data set and practice creating contingency tables and graphs of these contingency tables. Consider whether it would make sense to recode the variables before you produce your contingency tables. Remember that scale variables must be recoded into categories, as contingency tables are only used for categorical variables.

Section 6: More Graphs

This section introduced you to some further graphs. Below the code to produce various graphs is shown, try running this code. Try applying this code to different variables.

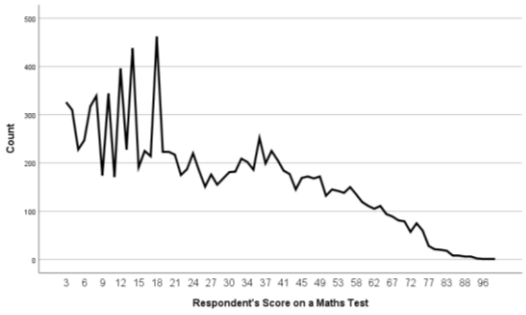
Histogram: Histograms are used to describe scale variables. This code produces a histogram for the scale variable 'age'.

```
GRAPH
  /HISTOGRAM age.
```



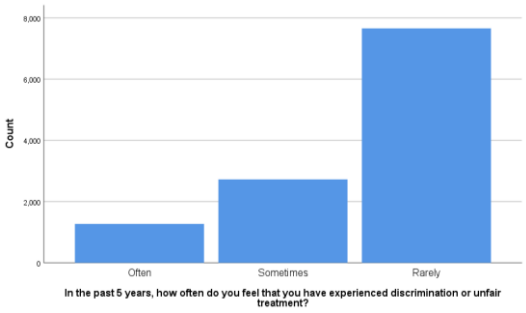
Line Graph: Line graphs can be used to describe scale variables. This code produces a line graph for the scale variable 'mathscore'.

```
GRAPH
  /LINE mathscore.
```



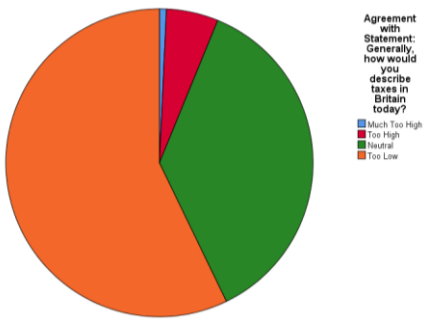
Bar Chart: Bar charts can be used to describe categorical variables. This code produces a bar chart for the categorical variable 'discrim'.

```
GRAPH
  /BAR discrim.
```



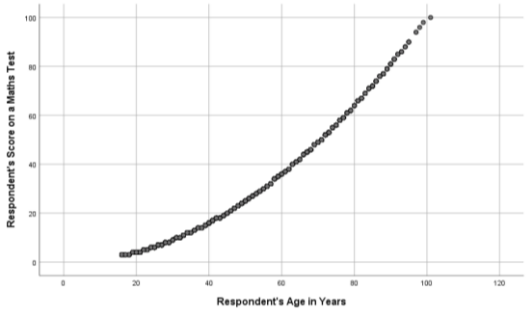
Pie Chart: Pie charts can be used to describe categorical variables. This code produces a pie chart for the categorical variable 'taxes'.

```
GRAPH
  /PIE taxes.
```



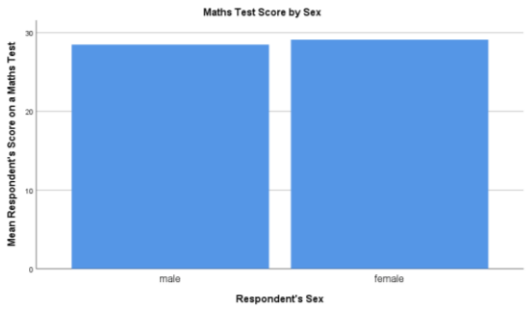
Scatterplot: Scatterplots show the relationship between two scale variables. The code below produces a scatterplot for the two scale variables 'age' and 'mathscore'.

```
GRAPH
  /SCATTERPLOT age WITH mathscore.
```



Barchart (more complex): This bar chart shows the relationship between the mean of a scale variable (maths test score) and a categorical variable (sex).

```
GRAPH
  /BAR=mean(mathscore) BY sex
  /TITLE 'Maths Test Score by Sex'.
```



You have now completed the second SPSS computer lab, well done!

Please remember that additional supporting resources are available on the VLE, and please do not hesitate to ask if you require any additional support or assistance.