Explorations: Conducting Empirical Research in Canadian Political Science (4th Edition)

IBM SPSS Statistics (SPSS) Handbook[1]

Jason Roy and Loleen Berdahl

Welcome to the *Explorations: Conducting Empirical Research in Canadian Political Science SPSS Handbook!* In this handbook, we provide you with a basic introduction to SPSS. The procedures outlined follow from the statistical methods described in *Explorations: Conducting Empirical Research in Canadian Political Science (4th Edition)*. We encourage you to work closely with the textbook as you move through this handbook; here we cover the technical "how to" of basic statistics, but we do not cover the critical issues of which statistics to use when. We use 2019 Canadian Election Study (CES) data. As we explain throughout the textbook, the CES data are a great publicly available resource for studying politics in Canada. We encourage you to practice the techniques outlined in this handbook with the CES datasets.

Please note that the screen shots included are those captured working with SPSS Version 24 on a Mac. The appearance of the SPSS work environment may differ across operating systems and versions of SPSS. Open your copy of SPSS and let's begin!

Part I: Getting Comfortable with SPSS

Destination

By the end of this section, you will be able to:

- *open a dataset in SPSS;*
- *navigate between SPSS windows;*
- *explain what a syntax file is, and why it is a valuable tool for researchers;*
- *save and export a record of your work in SPSS; and*
- *use search functions*

Like many other software programs, users can open SPSS datasets via the **File > Open > Data** option on the top menu bar (Note that a pop-up window will appear when you launch SPSS that also allows you to open files). To follow the procedures below, download and open the 2019 CES dataset (we use the telephone survey), available for free at https://doi.org/10.7910/DVN/8RHLG1. You should also download the 2019 CES technical documentation and codebooks for reference.

---

[1] We wish to acknowledge Elisabeth Gidengil's Poli 311 course (McGill University), which inspired this work.
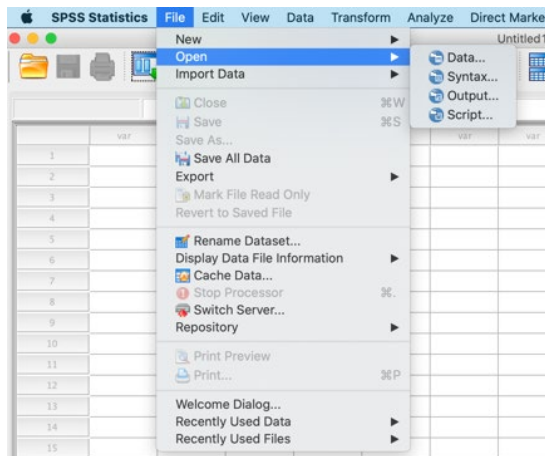
The SPSS work environment consists of two primary windows - the **Data Editor window** and the **Output window** – as well as a third commonly used window, the **Syntax window**. We explain each in turn:

- The **Data Editor window** consists of two views: **Data View** and the **Variable View**. You can switch between views by clicking on the desired view button located at the bottom of the screen (see below). In the **Data View window**, you can enter, view, or edit data in your data set. It looks like a spreadsheet. Variables are listed across the top (columns), and cases are listed down the side (rows). The **Variable View window** lists all variables in the working file along with variable information including the variable **Name**, the **Type** of variable (e.g. numeric or string), the variable **Label**, and the **Values** including the value labels for each variable.
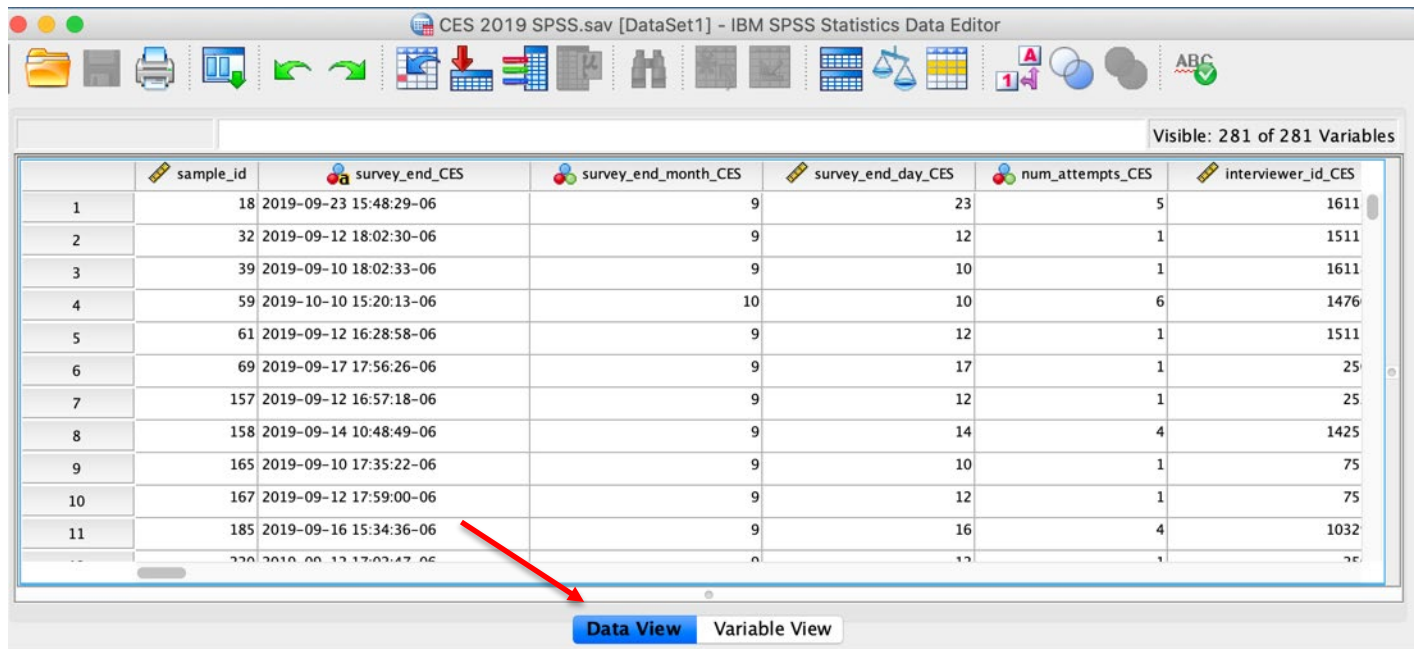
|	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
5	num_attem...	Numeric	2	0	num_attempts_CES -- Number of call attempts to ...	None	None	18	Right	Nominal
6	interviewer...	Numeric	6	0	interviewer_id_CES -- Interviewer ID for CES	None	None	20	Right	Scale
7	interviewer...	String	33	0	interviewer_gender_CES -- Interviewer gender for...	None	None	26	Left	Nominal
8	language_CES	Numeric	1	0	language_CES -- Language CES survey was compl...	{1, (1) English}...	None	14	Right	Nominal
9	phonetype_...	Numeric	1	0	phonetype_CES -- Phone type used to complete th...	{1, (1) Landline}...	None	15	Right	Nominal
10	survey_end...	String	66	0	survey_end_PES -- Date and time when PES survey...	None	None	26	Left	Nominal
11	survey_end...	Numeric	2	0	survey_end_month_PES -- Month when PES survey ...	{-7, (-7) Skipppe...	None	22	Right	Scale
12	survey_end...	Numeric	2	0	survey_end_day_PES -- Day when PES survey was c...	{-7, (-7) Skipppe...	None	20	Right	Scale
13	num_attem...	Numeric	2	0	num_attempts_PES -- Number of call attempts to ...	{-7, (-7) Skipppe...	None	18	Right	Nominal
14	interviewer...	Numeric	6	0	interviewer_id_PES -- Interviewer ID for PES (CATI ...	{-7, (-7) Skipppe...	None	20	Right	Scale
15	interviewer...	String	33	0	interviewer_gender_PES -- Interviewer gender for ...	None	None	26	Left	Nominal
16	language_PES	Numeric	1	0	language_PES -- Language PES survey was comple...	{-7, (-7) Skipppe...	None	14	Right	Nominal
17	phonetype_...	Numeric	1	0	phonetype_PES -- Phone type used to complete th...	{-7, (-7) Skipppe...	None	15	Right	Nominal
18	mode_PES	Numeric	1	0	mode_PES -- How the PES survey was completed	{-7, (-7) Skipppe...	None	10	Right	Nominal

Data View   Variable View

- The **Output window** is where every action in SPSS is recorded and all results are displayed. We recommend saving this file frequently. Note that you can open saved output files via the **File > Open > Output** option on the top menu bar. You can also use SPSS's **File > Export** option to save your output file in a different format, PDF for example, that can be viewed without the SPSS software.

To export your output file:

1. While viewing your output file, click on **File > Export** from the top SPSS menu bar
2. Select the desired document type from the drop-down **Export Output window**
3. Click **Browse** to set the save destination and the file name
4. Save the file by clicking **OK** at the bottom of the **Export Output window**
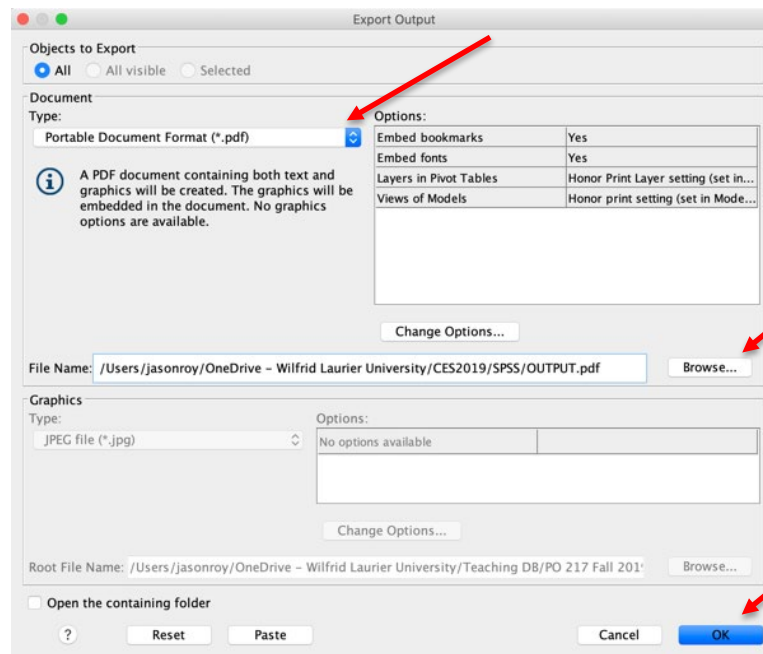
- The **Syntax window** is a workspace where you can write, edit, and save SPSS commands. The advantage is that you can easily edit, save and re-run all of your commands. We strongly recommend working with syntax files. Doing so allows you to record all of the procedures that you run and easily re-produce all results as needed. It also allows you to easily collaborate with colleagues working on the same project with you, and to share a history of your analysis with others, thus increasing research transparency. (See Chapter 3 for a discussion of ethics and data analysis.) In this handbook we primarily use the drop-down menu options to select the procedures we wish to run and then the **Paste** button to add the command to a syntax file. Note that you will then have to **Run** the command from the syntax file to complete the task. Users can add an Asterix (*) to the beginning of a line in the syntax file to make a note or add a title to the procedures that follow. You will need to add a hard return (one line space) after inserting a comment. To run a command from the syntax file, select the line(s) you wish to execute and then click on the **Play button icon** ▶ in the top center menu bar of the **Syntax window** (see below). You may also use the short-cut keys: **Command -R** on a Mac or **CTRL – R** with a Windows operating system. If you do not select the specific line(s) you wish to execute, this short cut will run the entire syntax file.

Insert hard return (one line space)
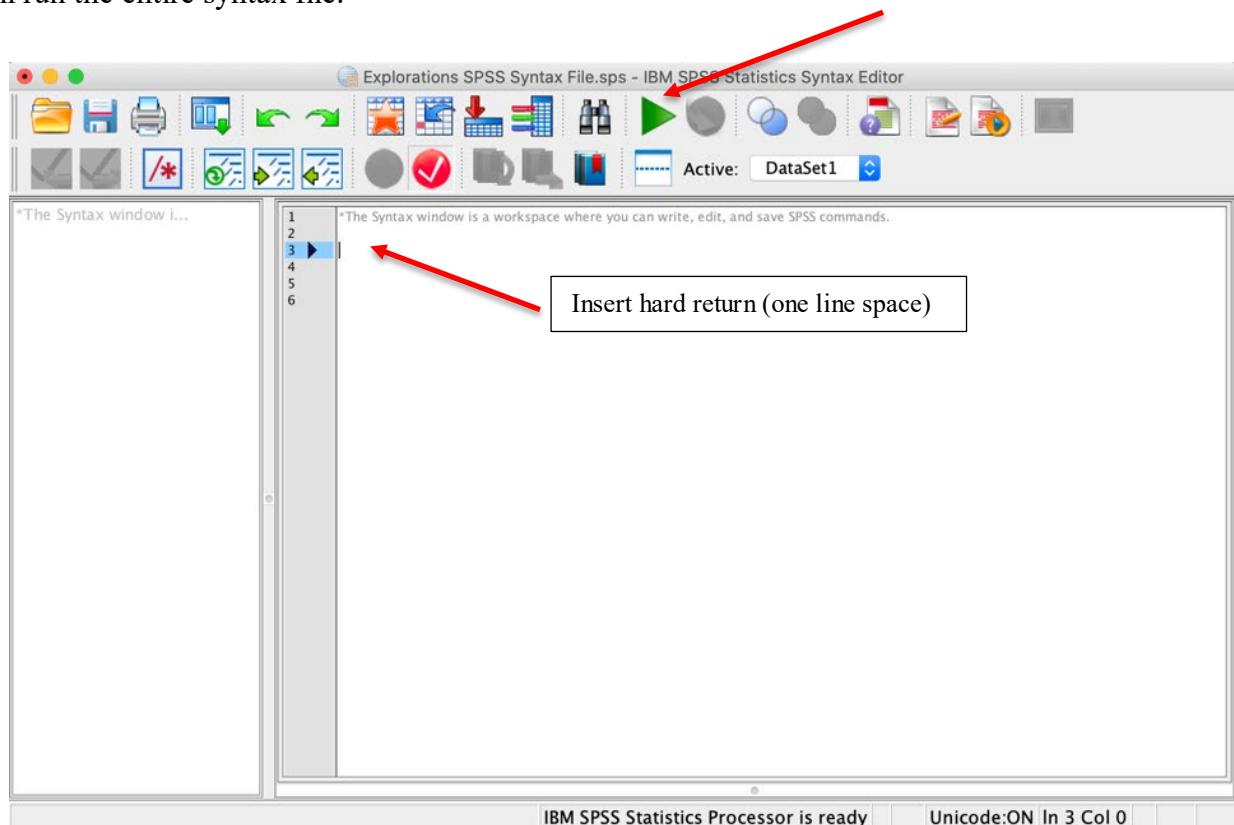
If you do not wish to work with a syntax file, you can simply **Run** the command instead of pasting it to a syntax file. We provide more detailed instruction on how to run commands in the examples below. To open a saved syntax file, choose the **File > Open > Syntax** option on the top menu bar.

Using Search functions: One of the strengths of SPSS is that it is a relatively simple program to use. We provide the steps necessary to run a range of commonly used procedures within this handbook. However, because the procedures presented here are only a small sample of what can be done in SPSS, you should familiarize yourself with the **Help** and **Search** features in SPSS. These are invaluable resources for SPSS commands, with detailed explanation of the various procedures and options available. With the **Output or**

Syntax window active, you can access the SPSS help features by selecting **Help** from the menu at the top of the screen and entering your search term into the search textbox.
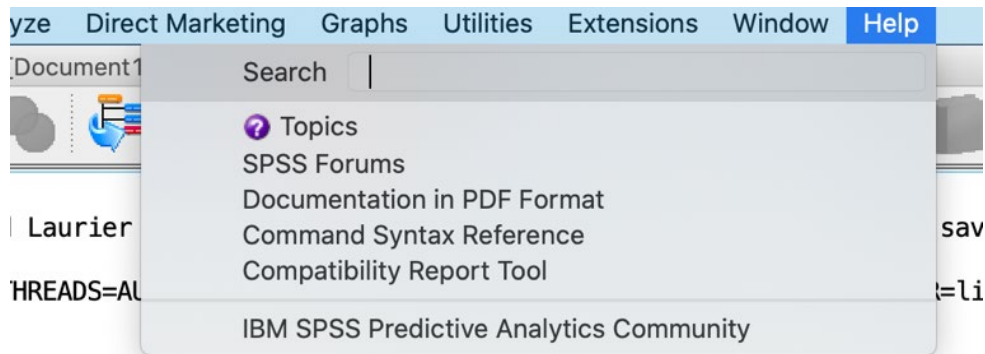
Check-In Point

If you are working alongside us, at this point you have opened the 2019 CES dataset in SPSS. You have explored a variety of SPSS windows to increase your familiarity with the various screens available to you. You now know what a syntax file is and why it is a valuable tool for researchers. You also know how to save and export your output file that records your work in SPSS. Finally, being aware that this handbook teaches you only a small amount of SPSS's capacities, you are familiar with how to search the SPSS help files.

With this foundation in place, you are ready to continue your explorations.

### Part II: Familiarizing Yourself with Variables in the Dataset

Destination

By the end of this section, you will be able to:
- *use the variable view window and codebook command to familiarize yourself with the variables in the dataset; and*
- *use the utility variables window to review variable values and value labels*

Once you have opened your dataset, you will want to take a preliminary look at the variables.[2] As discussed in Chapter 8, when using secondary datasets such as the CES, it is important to familiarize yourself with the dataset before conducting your analyses. One of the easiest ways to do so is using the **Variable View** option in the **Data Editor window** (see above). As noted above, from here you can find basic information on each of the variables in the dataset, including the variable name, variable label and the value labels. For the latter, click on the three dots located in the **Values** column of the variable you wish to explore to open a window that lists the values and corresponding labels for that variable. For example, if we wish to explore the value labels for *q6*, a variable that reports the level of satisfaction with the way democracy works in Canada, scroll down to the variable named *q6* in the **Variable View window** and then click on the three dots in the **Values** cell as shown in the screen capture image below to open a pop-up window with the value labels listed.

---

[2] We also recommend reviewing the technical documentation and codebooks available for download with the dataset. The latter provides a listing of the survey questions asked, their corresponding variable name, and the response categories.
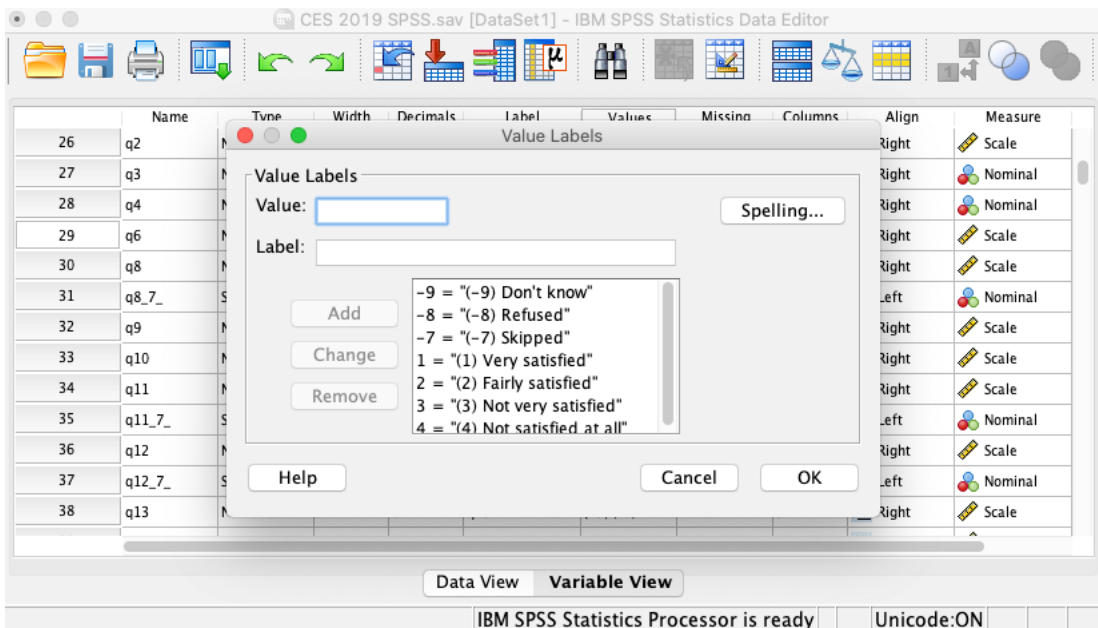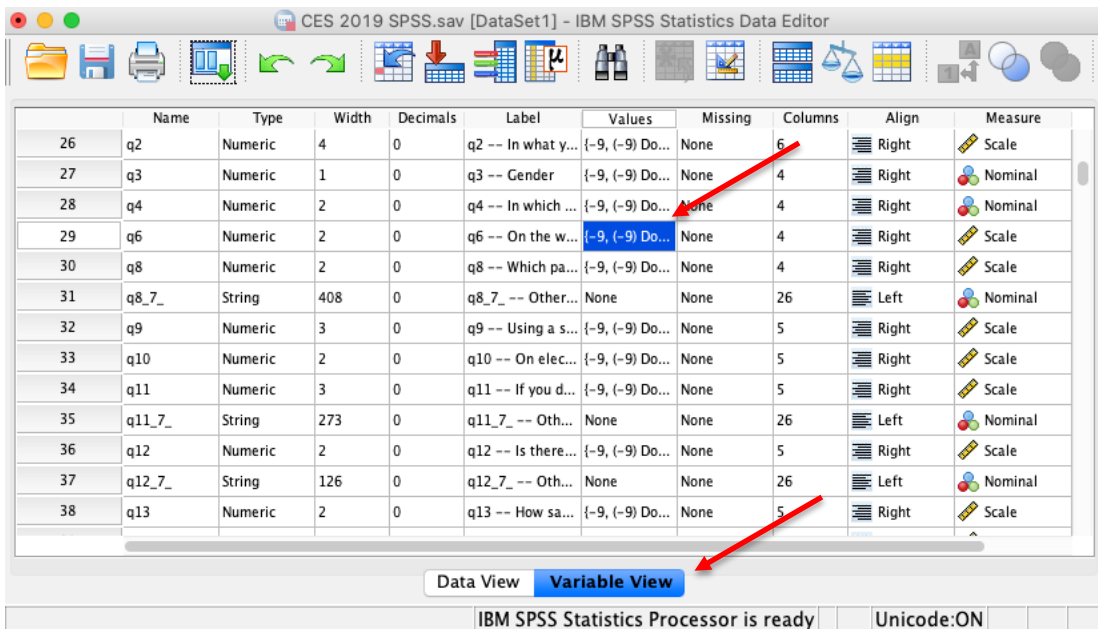
Tip: We recommend setting your SPSS preferences so that the list of variables displays the variable names in alphabetical order if they are not already. To do so, click on **SPSS Statistics > Preferences** (or **Edit > Options**) in the top menu bar to open the options window. Click on the **General** tab (default) and then select **Display names** and **Alphabetical** under **Variable Lists**, as shown in the images below. Click on **OK** to apply your preferences:
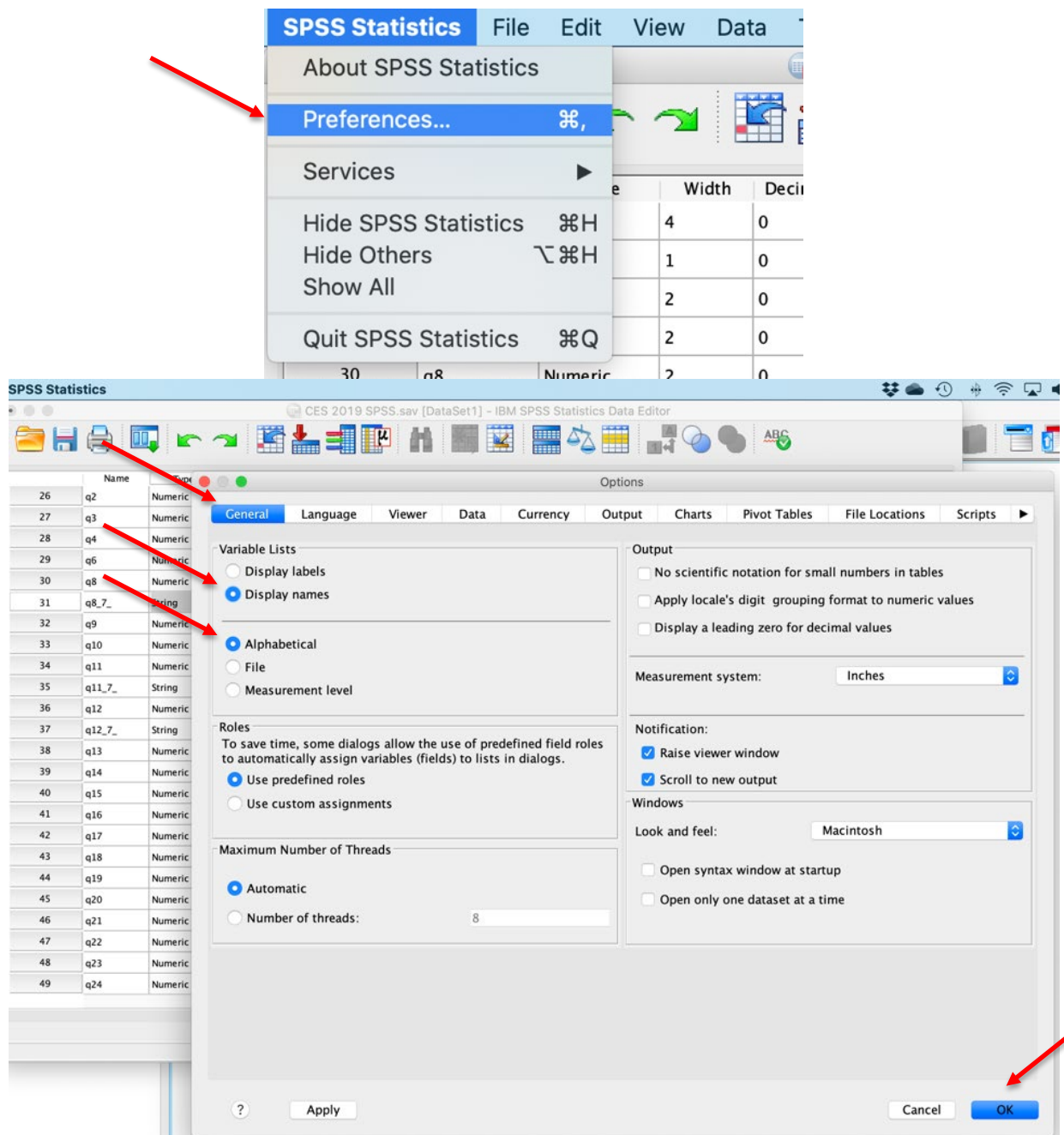
You can also view the labels assigned to each variable value by clicking **Utilities > Variables** from the top menu bar in SPSS to open the **Variables window**. From the window that opens, you can select the variable that you wish to explore by clicking on the variable name in the **Variable** column to display the variable information including the value labels. Note that this is the same value label information that we accessed in the **Variable View window**, described above:
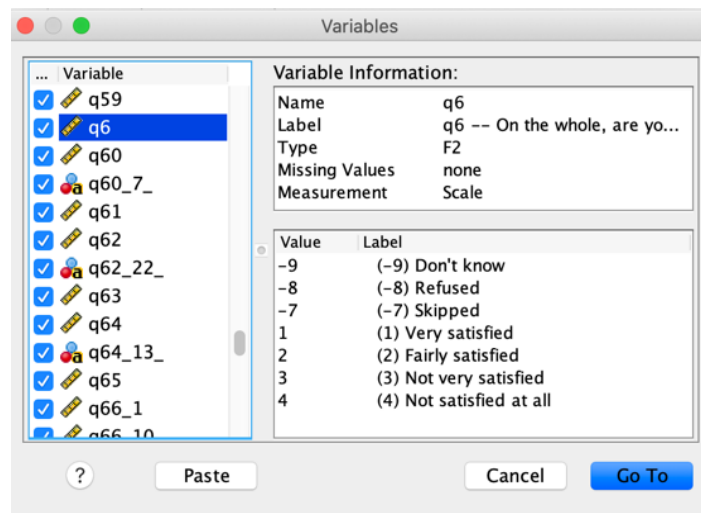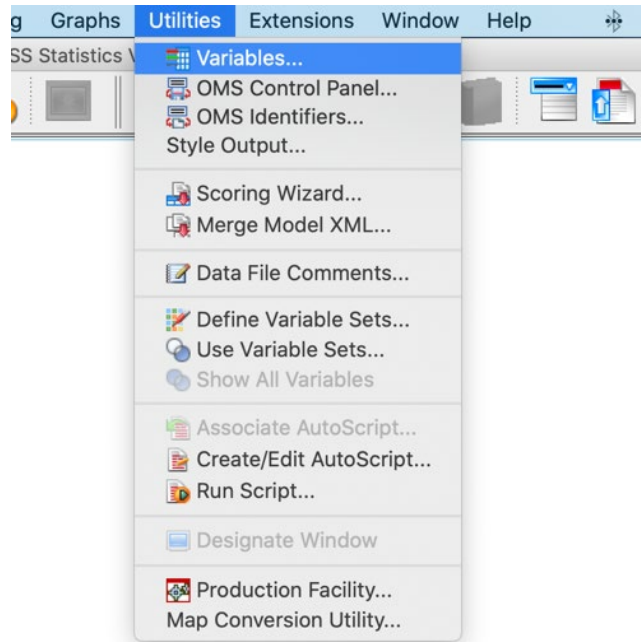
Another option for viewing the variables in the data set is to use the *codebook* command. You can access this command by clicking **Analyze > Reports > Codebook** from the top menu bar in SPSS:
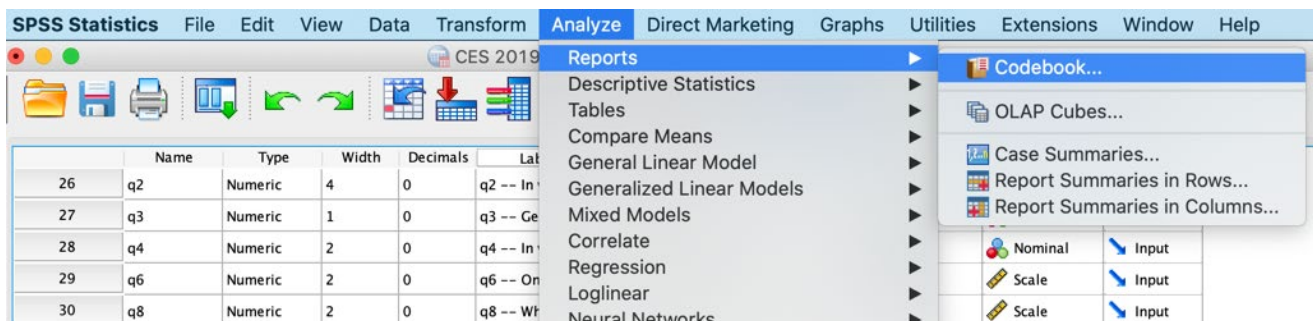
From the window that opens, you can select variables that you wish to examine by clicking on the variable name in the **Variables** column and then using the transfer arrow icon to add the variable to the **Codebook Variables** column:
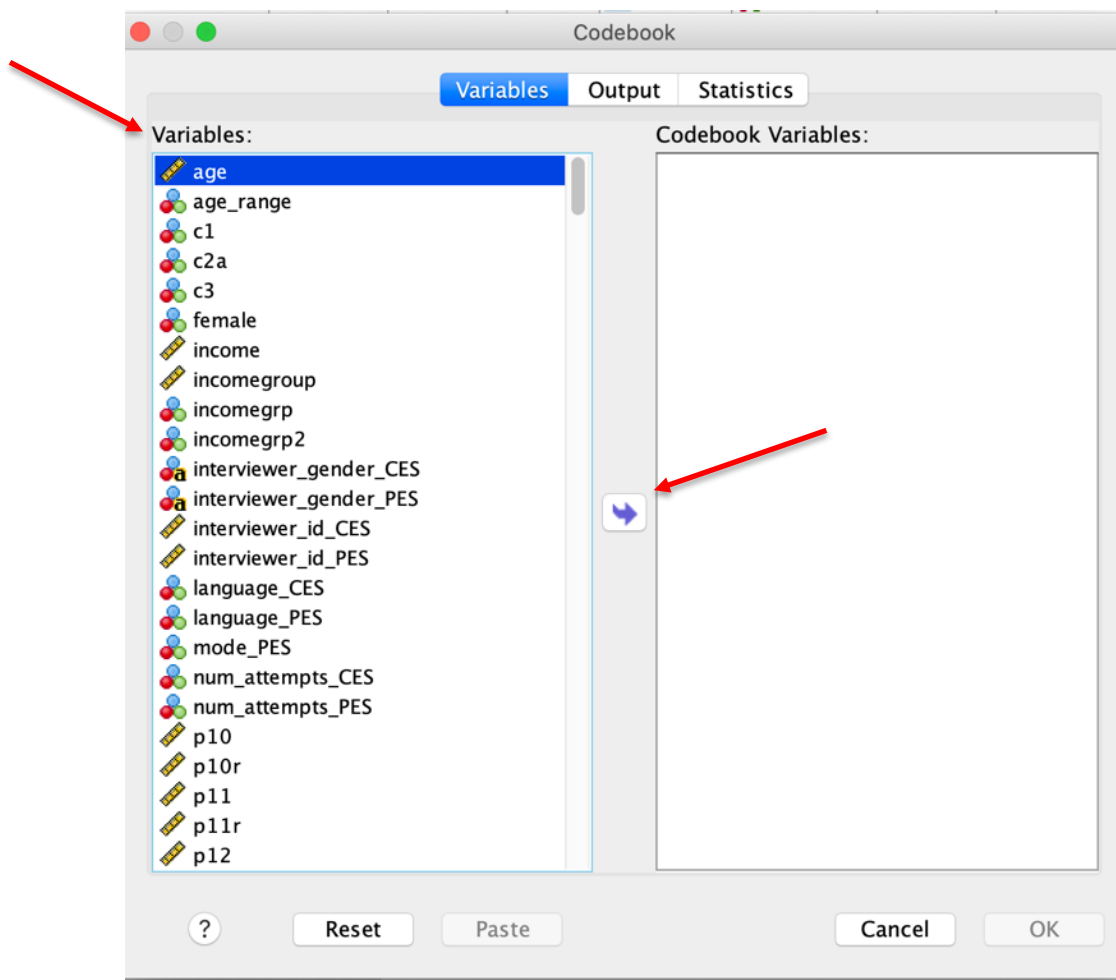
SPSS allows you to add multiple variables to the **Codebook Variables** list simultaneously. In fact, you could select all of the variables in the dataset, although this would produce a very large output file! Once you have selected the variables you wish to examine, you can click **Paste** to add the command to your syntax file and then run the command from the syntax file, as shown in the screen shots below. Click on **OK** if you do not wish to use the syntax file. Clicking the **Reset button** will clear all variables from the **Codebook Variables** list. As you will soon see, SPSS uses a very similar process for all procedures.
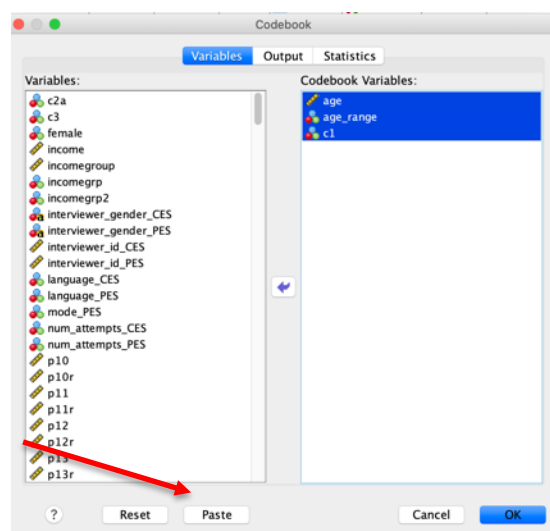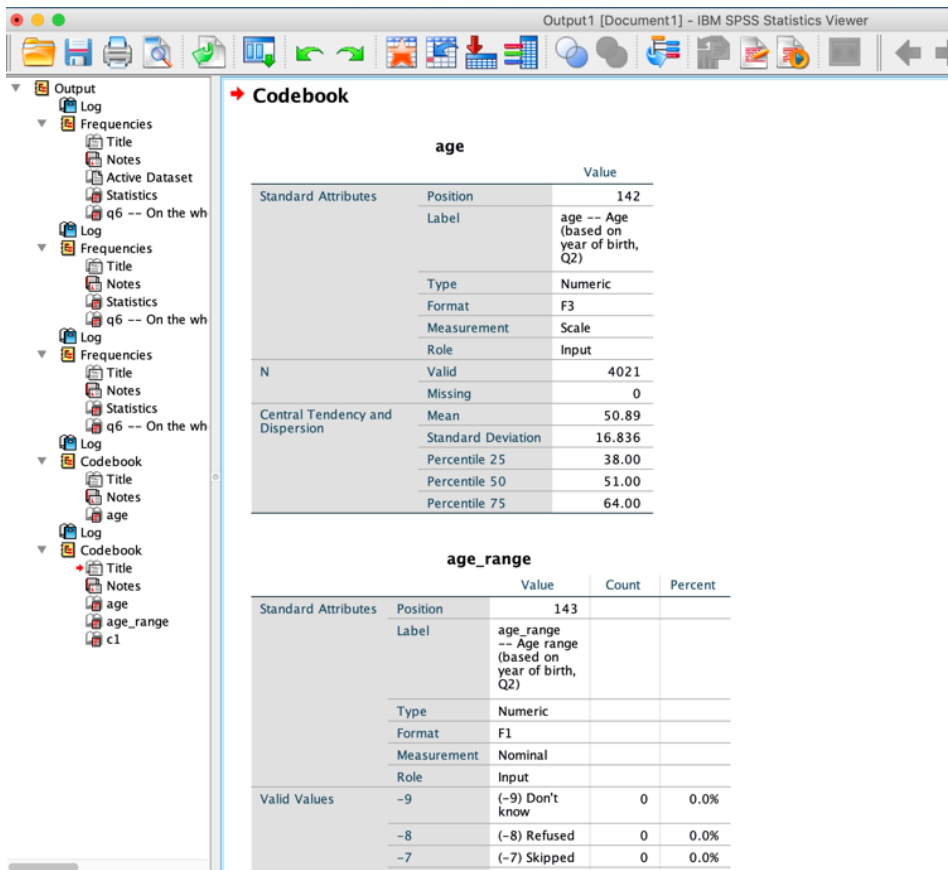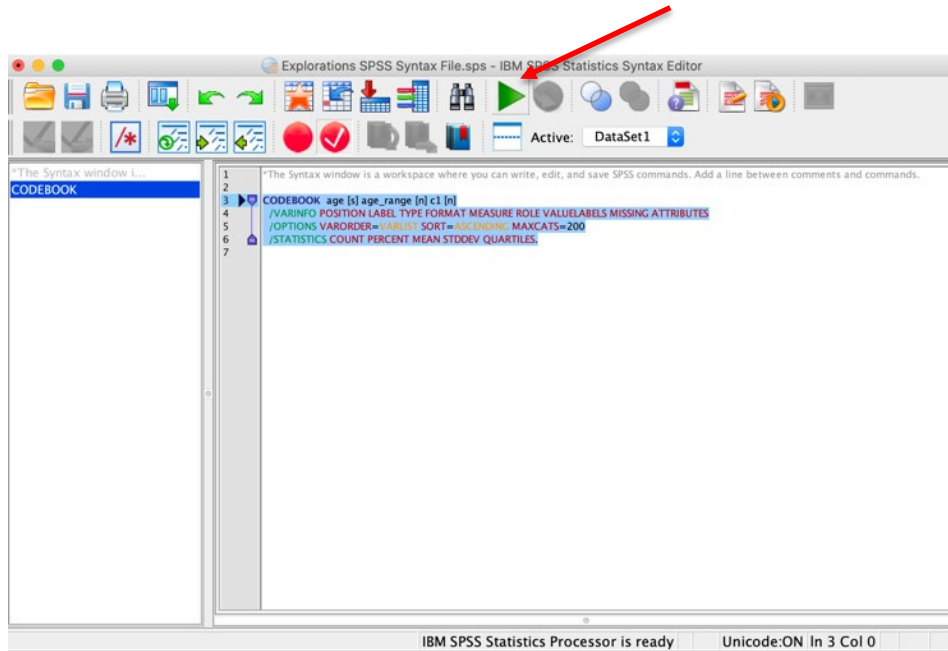
Tip: You can search for variables containing key words in the dataset via the **Variable View window** in the **Data Editor**. To do so, select the column you wish to search and then click on the **Binoculars icon** from the options menu to open a search box:
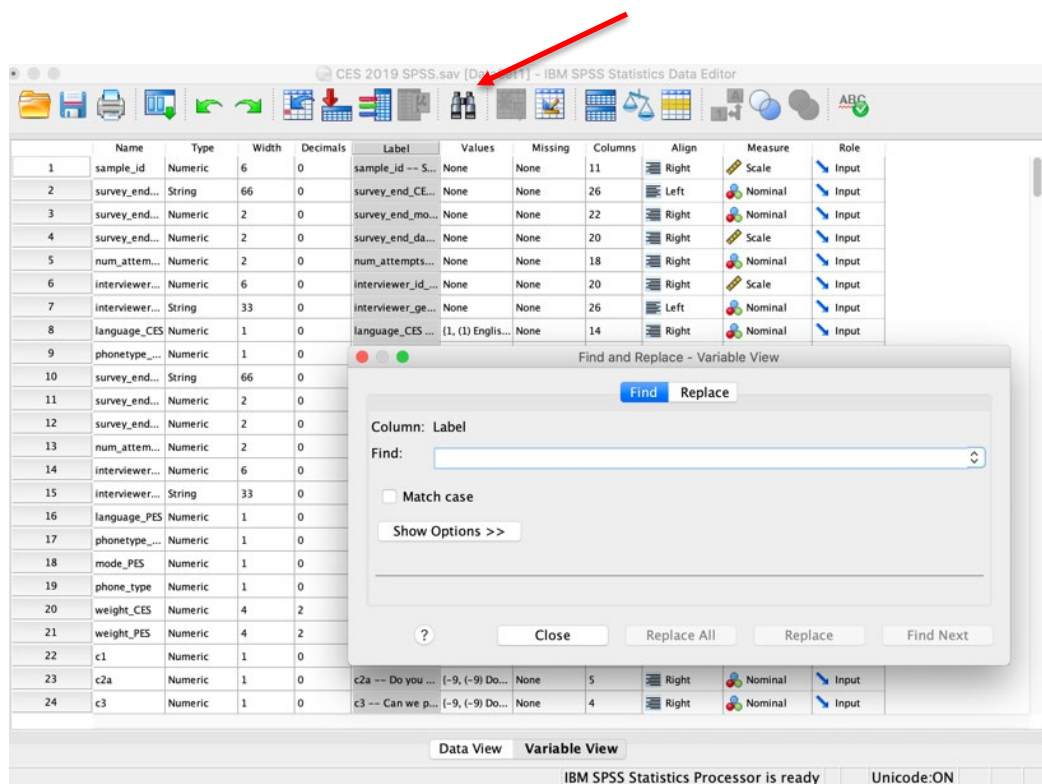
To practice, let's look at two variables in the dataset. The first, *q6*, reports the level of satisfaction with the way democracy works in Canada. The second, *p3*, reports the respondents' vote choice in the 2019 Canadian federal election. Follow these steps:

1. Click **Analyze > Reports > Codebook** from the top menu bar in SPSS.

2. Click on the **Reset button** to clear any previous work.

3. Add variables *q6* and *p3* to the **Codebook Variables** list by selecting them from the **Variables** list and using the **transfer arrow icon** between the two columns to add each variable to the **Codebook Variables** list. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

4. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

5. Compare your results with our results, reported below.

→ Codebook

q6

		Value	Count	Percent
Standard Attributes	Position	29		
	Label	q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied		
	Type	Numeric		
	Format	F2		
	Measurement	Scale		
	Role	Input		
N	Valid	4021		
	Missing	0		
Central Tendency and Dispersion	Mean	2.05		
	Standard Deviation	1.621		
	Percentile 25	2.00		
	Percentile 50	2.00		
	Percentile 75	3.00		
Labeled Values	–9	(–9) Don't know	56	1.4%
	–8	(–8) Refused	11	0.3%
	–7	(–7) Skipped	0	0.0%
	1	(1) Very satisfied	562	14.0%
	2	(2) Fairly satisfied	2248	55.9%
	3	(3) Not very satisfied	814	20.2%
	4	(4) Not satisfied at all	330	8.2%

p3

		Value	Count	Percent
Standard Attributes	Position	163		
	Label	p3 -- Which party did you vote for?		
	Type	Numeric		
	Format	F2		
	Measurement	Scale		
	Role	Input		
N	Valid	2700		
	Missing	1321		
Central Tendency and Dispersion	Mean	1.56		
	Standard Deviation	3.160		
	Percentile 25	1.00		
	Percentile 50	2.00		
	Percentile 75	3.00		
Labeled Values	–9	(–9) Don't know	8	0.2%
	–8	(–8) Refused	211	5.2%
	–7	(–7) Skipped	0	0.0%
	1	(1) Liberal Party	788	19.6%
	2	(2) Conservative Party	769	19.1%
	3	(3) NDP	460	11.4%
	4	(4) Bloc QuÃ©bÃ©cois	139	3.5%
	5	(5) Green Party	263	6.5%
	6	(6) People's Party	42	1.0%
	7	(7) Other	14	0.3%
	8	(8) Respondent cast invalid ballot	6	0.1%

The **codebook** results provide a wealth of information. For example, looking at the results for *q6*, we find the variable name (*q6*), the variable label (q6 -- On the whole, are you very satisfied, fairly satisfied, not very

satisfied), the number of missing cases and total number of cases (0/4021) as well as the frequency distribution with the corresponding values and value labels. While the other information reported in the output may be useful for some users, the information outlined here is most relevant for our purposes.

Check- In Point

If you are working alongside us, at this point you have learned how to search for and examine variables in the dataset. Before you move forward, be sure to practice these skills: Use the search function in the **Label** column of the **Variable View window** to identify variables in an area of interest to you. Once you have the variable names, generate a codebook for these variables.

Part III: Applying Survey Weights

Destination

By the end of this section, you will be able to:

- *explain why survey weights are often used in analysis; and*
- *apply a survey weight to the 2019 CES data.*

As we discuss in Chapter 5, when researchers sample from populations, they often over-sample certain population segments and then create design weights to adjust for over or under representation of certain segments of the population. When you use secondary survey datasets, be sure to consult the metadata (technical documentation, as discussed in Chapter 8) to review information on the sampling procedures and weight variables.

The 2019 CES employs a disproportionate random sampling technique that oversamples in some areas of the country, such as Quebec, while under sampling in others. The CES dataset includes two weight variables, *weight_CES* and *weight_PES*, to account for provincial over/under sampling as well as phone ownership (landline and/or cell phone. See CES technical documentation for more details). The former weights according to the full sample and the latter is a weight based on only those respondents who completed both waves of the survey (campaign period and post-election). We use the full sample weight (*weight_CES*) in our analyses.

To apply survey weights in SPSS:

1. Select **Data > Weight Cases…** from the top menu bar.
2. Click on the radio button next to **Weight cases by**
3. Select the weight variable you wish to apply from the list of variables and use the transfer arrow next to the **Frequency Variable** text box. We use the full sample weight, *weight_CES*, for the examples presented here.
4. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
5. The SPSS **Data Editor window** indicates that a weight has been applied (**Weight On**) in the bottom corner of the window.
6. You can follow the steps above to turn a weight off by selecting **Do not weight cases** in step 2.

SPSS Statistics   File   Edit   View   **Data**   Transform   Insert   Format   Analyze   Direct N

Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
Define date and time...
Define Multiple Response Sets...

Validation ▶
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...

Sort Cases...
Sort Variables...
Transpose...
Restructure...
Merge Files ▶
Aggregate...

Orthogonal Design ▶

Split File...
Select Cases...
Weight Cases...

Labeled Values

ties

s
ve Dataset
stics
-- On the wh

ties

s
stics
-- On the wh

ties

s
stics
-- On the wh

ok

Standard Attributes

[t1] – IBM S

1.4%

0.3%
0.0%
14.0%

55.9%

20.2%

8.2%

Percent

vote for?

Weight Cases

q77
q77eng
q77fr
q8
q9
r1
sample_id
satdem
satdemocracy
survey_end_day_CES
survey_end_day_PES
survey_end_month_CES
survey_end_month_PES
unigrad
vote
vote2019
weight_CES

○ Do not weight cases
● Weight cases by
    Frequency Variable:

Current Status: Do not weight cases

?   Reset   Paste   Cancel   OK

14

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
263	p16r	Numeric	3	0	p16r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
264	p17r	Numeric	3	0	p17r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
265	p36r	Numeric	3	0	p36r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
266	p37r	Numeric	3	0	p37r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
267	p38r	Numeric	3	0	p38r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
268	p39r	Numeric	3	0	p39r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
269	p40r	Numeric	3	0	p40r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
270	p41r	Numeric	3	0	p41r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
271	p42r	Numeric	3	0	p42r -- (numer...	{-9, (-9) Do...	None	6	Right	Scale	Input
272	satdem	Numeric	1	0	Satisfaction wi...	{1, Very Sati...	None	8	Right	Nominal	Input
273	vote2019	Numeric	1	0	Vote choice in ...	{-9, (-9) Do...	None	10	Right	Nominal	Input
274	satdemocracy	Numeric	1	0	Satisfaction wi...	{1, Very Sati...	None	14	Right	Nominal	Input
275	female	Numeric	1	0		None	None	8	Right	Nominal	Input
276	incomegroup	Numeric	7	0		None	None	13	Right	Scale	Input
277	income	Numeric	7	0		None	None	9	Right	Scale	Input

Data View   **Variable View**

IBM SPSS Statistics Processor is ready   Unicode:ON   Weight On

Check-In Point

Survey weights can be a challenging idea for many new researchers. Before you move forward, ensure that you are comfortable with your understandings. Why do researchers use survey weights? How can you as a user of secondary survey datasets determine how the original researchers constructed their survey weights? We encourage you to review both Chapter 5 and Chapter 8 of the *Explorations* textbook before moving forward to the next section.
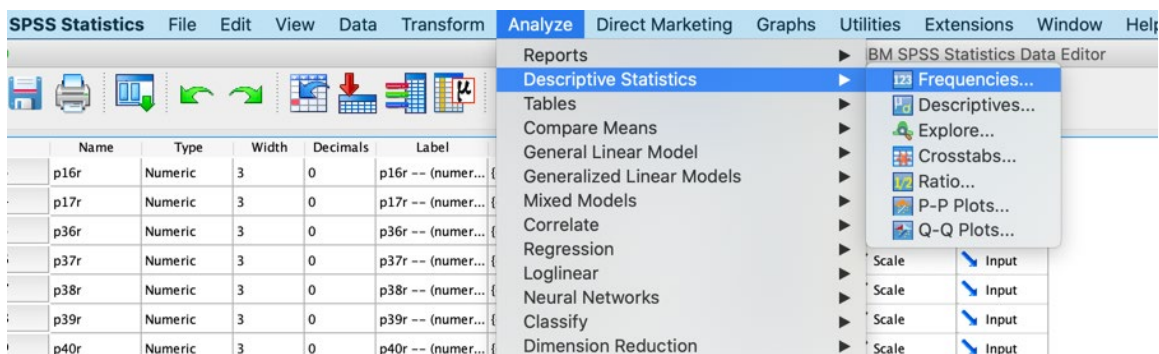
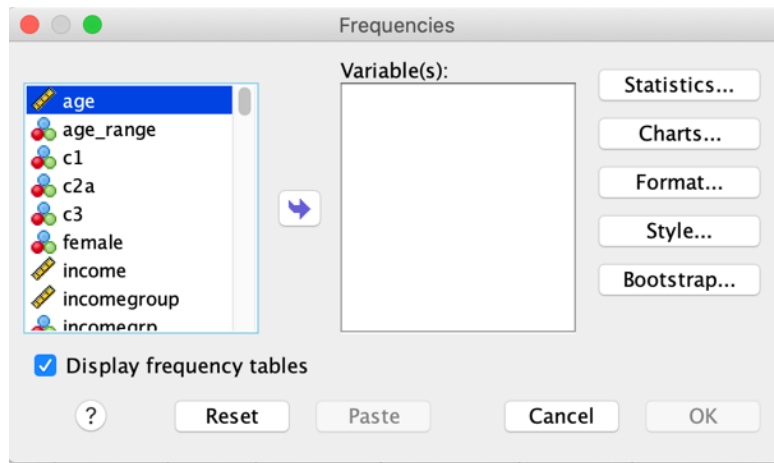## Part IV: Examining Frequency Distributions and Univariate Statistics

Destination

By the end of this section, you will be able to:

- *generate frequency distributions; and*
- *generate univariate statistics.*

In Chapter 12, we discuss how researchers start their analyses by examining the frequency distributions and summary statistics for each individual variable in their analysis. These can be generated in SPSS by selecting **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar. This will open a new window where you can select the variable(s) that you wish to examine.

SPSS Statistics	File	Edit	View	Data	Transform	Analyze	Direct Marketing	Graphs	Utilities	Extensions	Window	Help

Reports ▶
Descriptive Statistics ▶ — Frequencies...
Tables ▶ — Descriptives...
Compare Means ▶ — Explore...
General Linear Model ▶ — Crosstabs...
Generalized Linear Models ▶ — Ratio...
Mixed Models ▶ — P-P Plots...
Correlate ▶ — Q-Q Plots...
Regression ▶
Loglinear ▶
Neural Networks ▶
Classify ▶
Dimension Reduction ▶

Name	Type	Width	Decimals	Label
p16r	Numeric	3	0	p16r -- (numer...
p17r	Numeric	3	0	p17r -- (numer...
p36r	Numeric	3	0	p36r -- (numer...
p37r	Numeric	3	0	p37r -- (numer...
p38r	Numeric	3	0	p38r -- (numer...
p39r	Numeric	3	0	p39r -- (numer...
p40r	Numeric	3	0	p40r -- (numer...

Try this for yourself to generate a frequency table for the satisfaction with the way democracy works in Canada variable (*q6*):

1. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q6* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Ensure that the box beside **Display Frequency Tables** is checked.
5. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon in** the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
6. Compare your results with our results, reported below.

➡ **Frequencies**

Statistics

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

N	Valid	4021
	Missing	0

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	56	1.4	1.4	1.4
	(−8) Refused	11	.3	.3	1.7
	(1) Very satisfied	562	14.0	14.0	15.6
	(2) Fairly satisfied	2248	55.9	55.9	71.5
	(3) Not very satisfied	814	20.2	20.2	91.8
	(4) Not satisfied at all	330	8.2	8.2	100.0
	Total	4021	100.0	100.0	

Note that the results include the unweighted raw frequency (Frequency), relative frequency (Percent and Valid Percent; the later reports the relative frequency *without* missing cases and is the relative frequency that we recommend reporting) and the cumulative frequency (Cumulative Percent).

In the previous section, we discussed survey weights, and how to apply them. Let's do this now, by apply the survey weight, *weight_CES* (see above or apply the weight using your syntax file) and then re-running the frequency command to account for the disproportionate random sample by weighting the data. Compare your weighted and unweighted results. How does weighing the data alter the results? **For subsequent analyses in this handbook, we continue to work with weighted data.**

➡ **Frequencies**

Statistics

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

N	Valid	4021
	Missing	0

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	54	1.3	1.3	1.3
	(−8) Refused	8	.2	.2	1.5
	(1) Very satisfied	565	14.1	14.1	15.6
	(2) Fairly satisfied	2248	55.9	55.9	71.5
	(3) Not very satisfied	828	20.6	20.6	92.1
	(4) Not satisfied at all	318	7.9	7.9	100.0
	Total	4021	100.0	100.0	

In examining a variable, you will also want to consider the appropriate measures of central tendency and dispersion. (Review Chapter 12 if you need refreshing on the appropriate measures of central tendency and dispersion by variable level.) As noted earlier, the frequency distribution results include the raw frequency, relative frequency and the cumulative frequency. This is an ordinal variable, and from these results you can visually identify the appropriate measure of central tendency (median) and dispersion (range). To move beyond a visual assessment, you can specify the summary statistics that you wish to view. Note that SPSS does not include the variation ratio statistic. Fortunately, this is easily identified with the information reported in the frequency distribution table (again, see Chapter 12).

For example, to obtain the median and range for this variable we would do the following to produce the results reported below:

1. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q6* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Ensure that the box beside **Display Frequency Tables** is checked.
5. Click on the **Statistics** button to open the statistics option window (see below).

6. Select the measure of **Central Tendency** and **Dispersion** that you wish to report. In this example, we will request the **Median** and the **Range** since this is an ordinal-level variable. Once you have made your selection, click on **Continue** (see below).
7. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
8. Compare your results with our results, reported below.
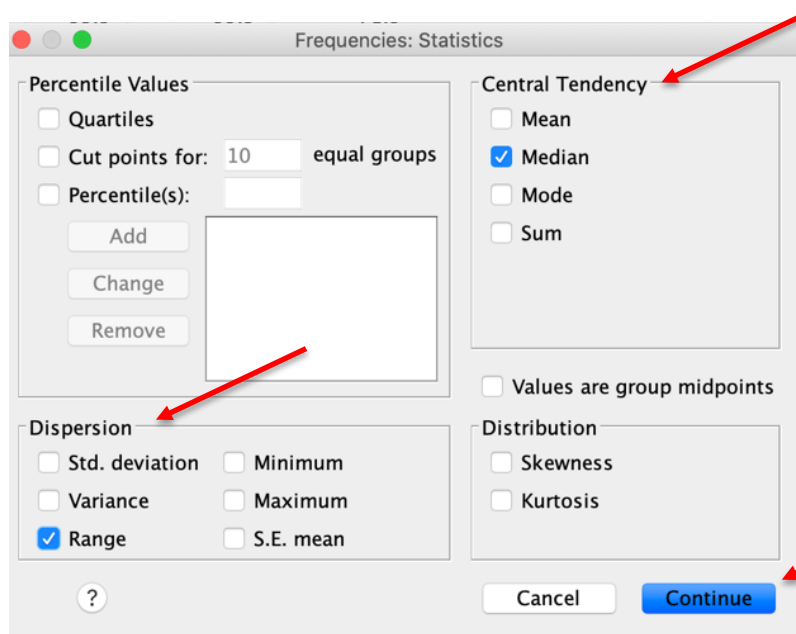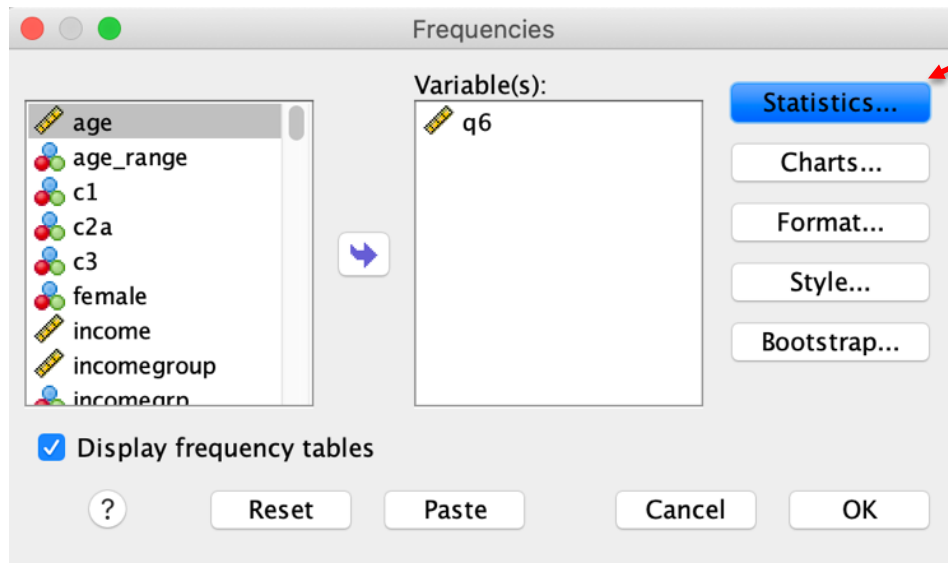
→ Frequencies

Statistics

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

N	Valid	4021
	Missing	0
Median		2.00
Range		13

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	54	1.3	1.3	1.3
	(−8) Refused	8	.2	.2	1.5
	(1) Very satisfied	565	14.1	14.1	15.6
	(2) Fairly satisfied	2248	55.9	55.9	71.5
	(3) Not very satisfied	828	20.6	20.6	92.1
	(4) Not satisfied at all	318	7.9	7.9	100.0
	Total	4021	100.0	100.0	

Let's interpret these univariate statistics, starting with the median. SPSS reports the median as "2.00". Note that these are the actual values for the variable, not the value labels as reported in the frequency distribution. To determine the value labels associated with these values, you need to view the value labels for the variable. We describe how to do this above using either the **Variable View window** in the **Data Editor** or by accessing the **Variable window** through the **Utilities > Variables…** option in the top menu bar.

Based on the value labels, we see that the value label for 2 is "Fairly satisfied". But how do we interpret the range value of 13? Recall from Chapter 12 that the range is estimated by subtracting the lowest value from the highest value. When you look at the range result, the number should strike you as a bit curious – how does a variable with four possible response categories have a range of 13? When you see results like this, you should always ask questions and seek out the answer. In this case, the answer lies with the coding. In the CES dataset "Don't know" is coded as -9 (see above). Thus, SPSS calculated the range as (-9) – (4) for a range of 13. In this example, some recoding is necessary to estimate the range. We will return to this topic shortly; for now, simply know that it is always important to critically assess your results, as statistical software will not catch such issues for you!

Check-In Point

At this point, you should understand how to use SPSS to generate frequency distributions and univariate statistics (specifically the median and the range). Before moving forward, be sure to practice these skills with other variables, using either the **Variable View window** in the **Data Editor** or the **Variable window** accessed through **Utilities** to help interpret categories and to identify curious results that may reflect coding.

**Part V: Creating and Recoding Variables**

Destination

By the end of this section, you will be able to:

- *explain why you should never recode original variables;*
- *create and recode new variables; and*
- *rename variables and add or alter variable labels.*

As we have already observed, it is often necessary to recode variables before you can work them. As a rule, we recommend never altering original variables within a dataset. We will repeat this, in case you are reading quickly: **never alter original variables in a dataset**. Instead, you should generate a new variable from the original and then make the transformations you need to your new variable. There are two reasons for this: (1) it allows you to check your work by comparing the recoded variable against the original one, and (2) maintaining the original variable allows you to use a different transformation processes if you need to do so at a later time.

Creating and Recoding New Variables

To create a new variable, we use the **Transform > Recode into Different Variables...** option from the top menu bar:
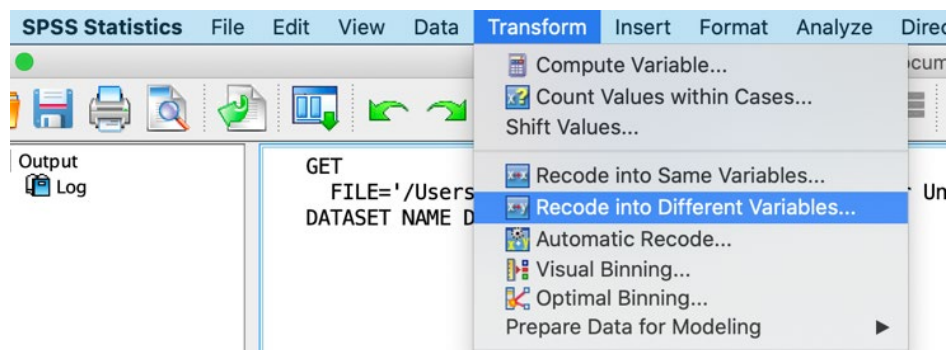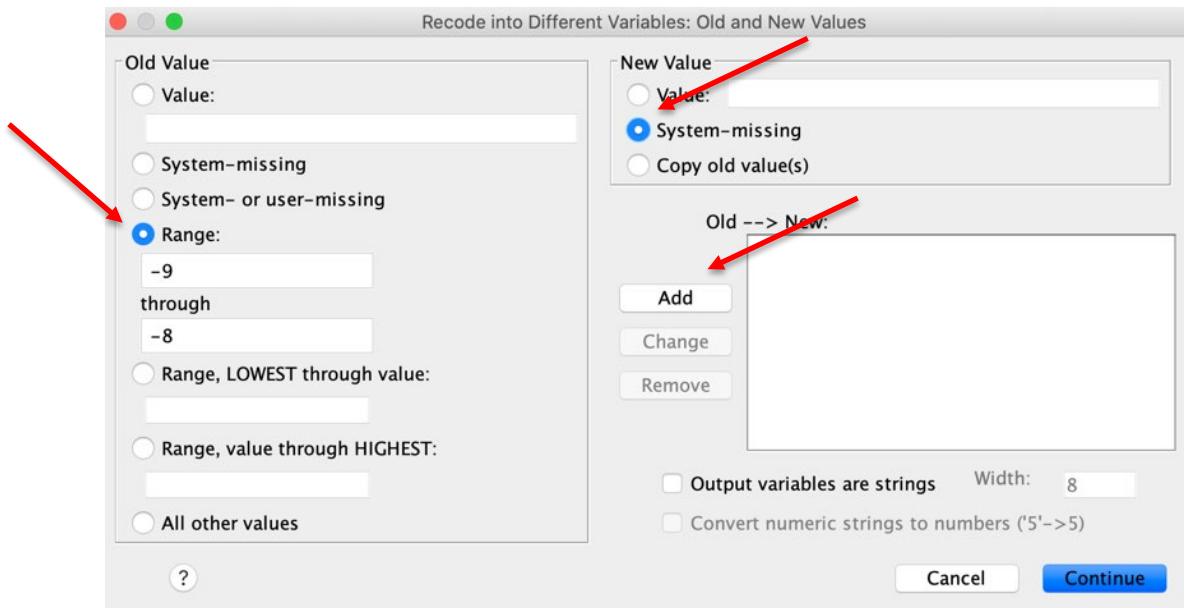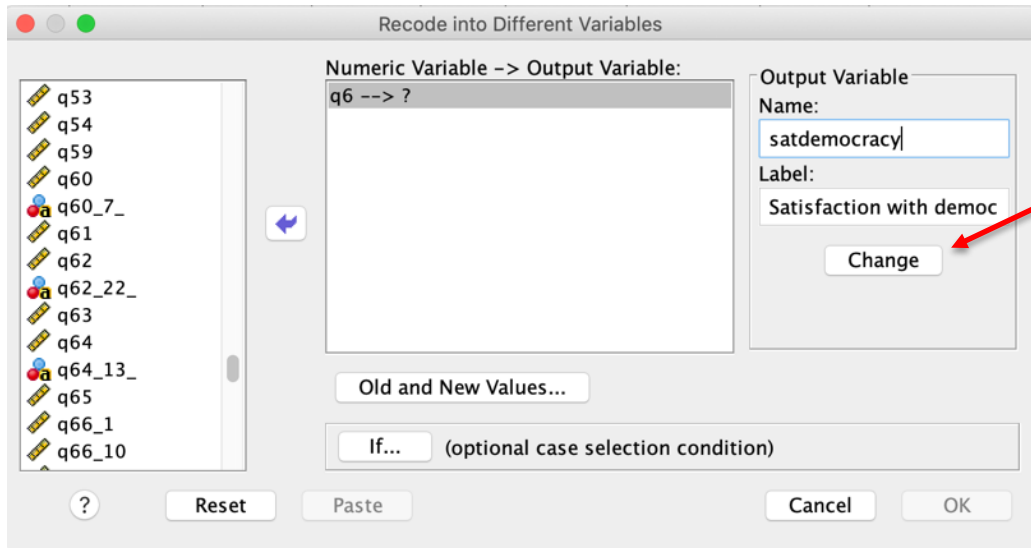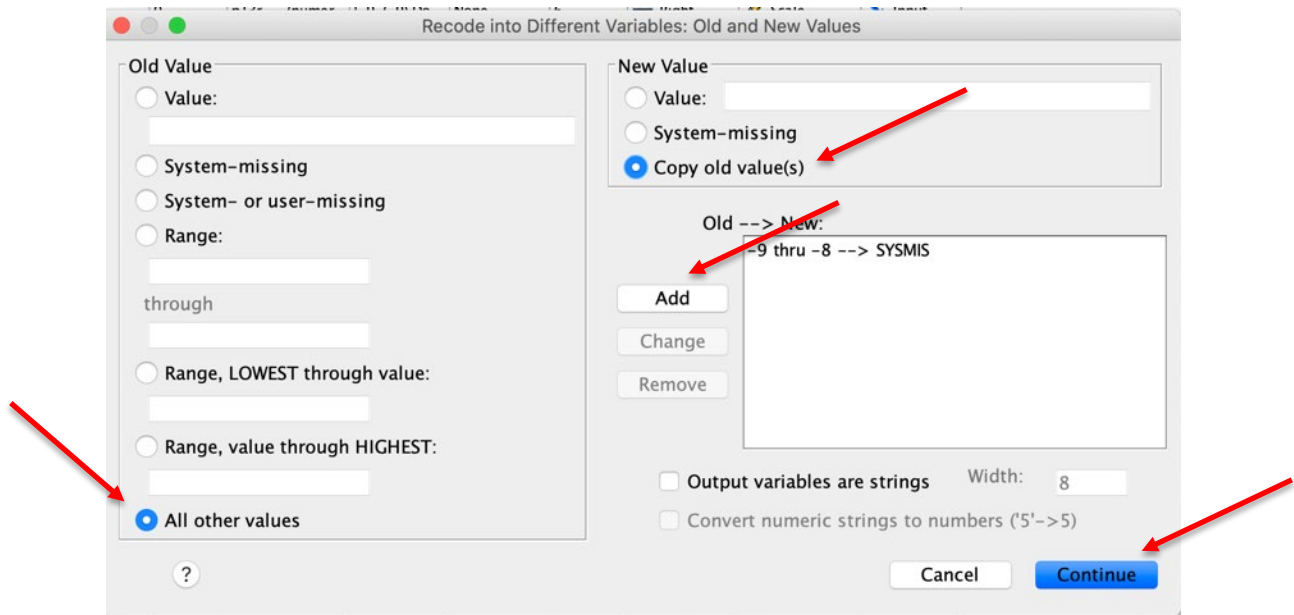

Let's consider this with variable *q6* from the 2019 CES. You are going to create a new variable named *satdemocracy* and label it *Satisfaction with democracy* (remember that the variable looks at satisfaction with democracy).

1. Select **Transform > Recode into Different Variables...** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q6* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply (see below).
5. Click on the button titled **Old and New Values** to recode the new variable. For example, let's say you want to transform *satdemocracy* to remove cases that report "Don't know" or "Refused". Recall from the last section that the range for *q6* was nonsensical given the inclusion of these responses, which are coded as -9 and -8, respectively. Given that we cannot be sure how individuals who answered "Don't know" or "Refused" feel about the way democracy works in Canada, we want to exclude these cases from our new variable for our analysis. To recode values, you need to enter the original value(s) under the **Old Value** column and the corresponding new value(s) under the **New Value** column and then click on the **Add** button to add the transformation to the recode command. Given that we want to recode -9 and -8 to **System-missing**, we can enter each value separately or use the **Range** option by

clicking the radio button next to that option and adding the values and then select the radio button next to **System-missing** in the **New value** column (see below).
6. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.
7. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**
8. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon in** the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

When you generate a new variable, you can confirm your work by comparing the frequency distributions of the original variable and the new variable. To produce a frequency distribution table for multiple variables in a single command, simply add both variables to the **Variable(s)** list:

1. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q6* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add variable *satdemocracy* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Ensure that the box beside **Display Frequency Tables** is checked**.**
6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
7. Compare your results with ours, noting especially the difference in the number of valid cases and missing cases for the two variables as reported in the **Statistics box**:

Statistics

		q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied	Satisfaction with democracy
N	Valid	4021	3959
	Missing	0	62

Frequency Table

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(-9) Don't know	54	1.3	1.3	1.3
	(-8) Refused	8	.2	.2	1.5
	(1) Very satisfied	565	14.1	14.1	15.6
	(2) Fairly satisfied	2248	55.9	55.9	71.5
	(3) Not very satisfied	828	20.6	20.6	92.1
	(4) Not satisfied at all	318	7.9	7.9	100.0
	Total	4021	100.0	100.0	

Satisfaction with democracy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	565	14.1	14.3	14.3
	2.00	2248	55.9	56.8	71.0
	3.00	828	20.6	20.9	92.0
	4.00	318	7.9	8.0	100.0
	Total	3959	98.5	100.0	
Missing	System	62	1.5		
Total		4021	100.0		

The distribution of the two variables should be identical, although you will note that the value labels are not attached to the new variable. You will need to add these, which we return to shortly.

From the results, we can confirm that we have not made any errors by comparing the number of cases in each category of the new variable against those of the original variable. You should note that the number of missing cases includes the 54 respondents who responded "Don't Know" and the 8 respondents who refused to answer (note that we continue to use weighted data, see above).

What would you do if you discovered that you made a mistake? (For example, say you accidently set the value of 4 as missing, when it is in fact not missing). First, you would congratulate yourself for creating new variables rather than transforming an original variable, as your problem can be easily fixed. Second, you can simply delete a mis-transformed variable in SPSS by selecting it in the **Variable View** of the **Data Editor Window** and deleting the row (see below). Note that new variables are added to the end of the variable list in **Variable View**. You can then use the **Create and Recode New Variables** procedure outlined above to correct the error. This should be even easier if you can work from a **Syntax file** to make the change. (You might also discover that the error can be fixed using different options, such as transforming the new variable

using the **Transform - Recode into Same Variables…** option). Please don't actually do this – we are going to continue working with *satdemocracy!*
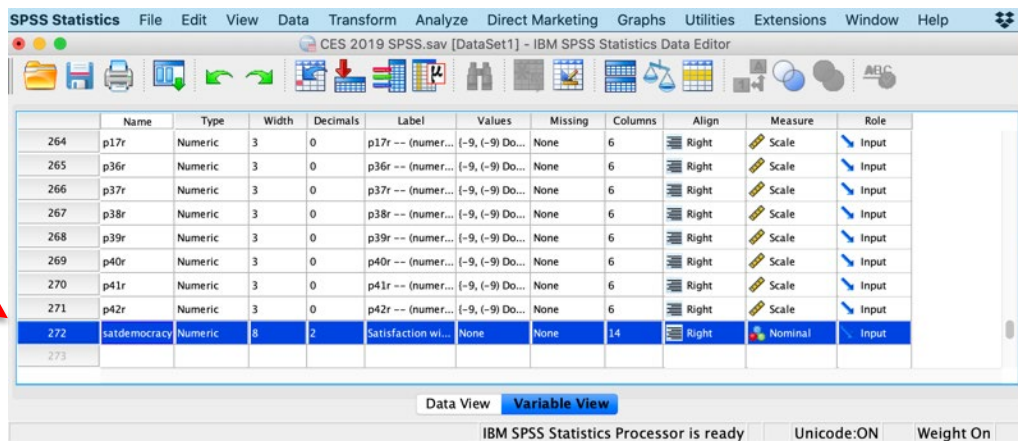

An alternative to transforming an existing variable is to create a new variable using the **Transform – Compute Variable…** command. This option allows for more advanced transformations including the option of combining multiple variables to generate a single measure. This is part of a host of options that can be used with the **Compute Variable** command. We do not cover these more advanced procedures in detail, although we do provide an example of how **Compute** can be used to combine variables in an example below. We encourage interested users to seek out additional information on the various options that can be used with the **Compute Variable** command via the SPSS help files.

Renaming variables and adding or altering value labels

The majority of the variables in the 2019 CES dataset include variable and value labels – descriptions of each variable, and descriptions of the values for categorical variables. However, for any new variables that you generate, you will need to either create or modify variable names and labels. One way to do so is through the **Variable View window** of the **Data Editor**.
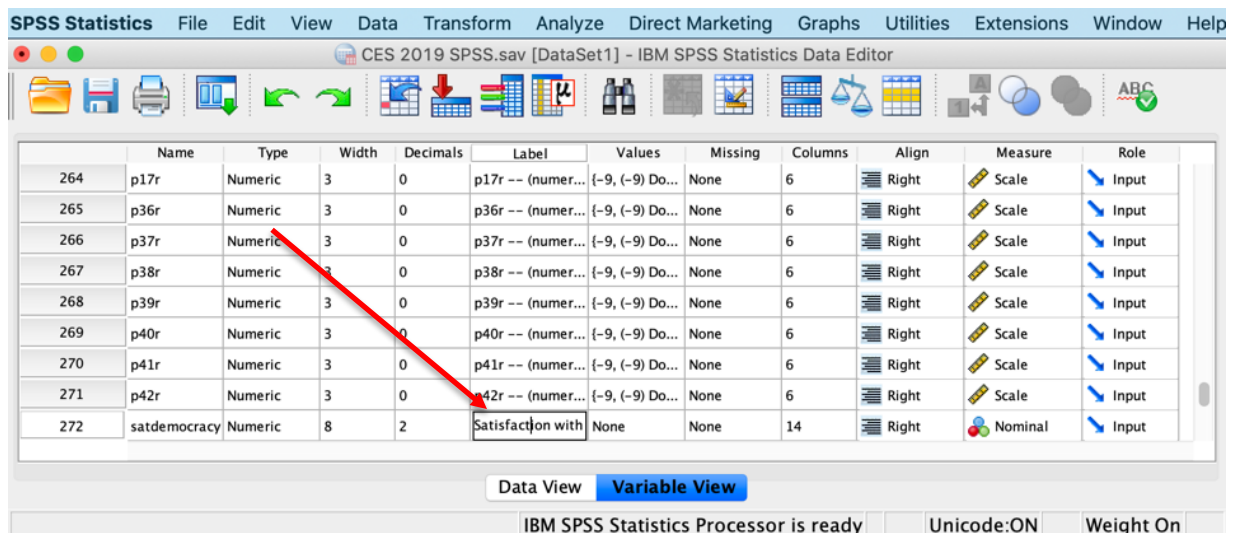
To change the name of an existing variable:

1. Open the **Variable View window** of the **Data Editor.**

2. Double - click on the variable name you wish to change in the **Name column.**

3. Type the new variable name.
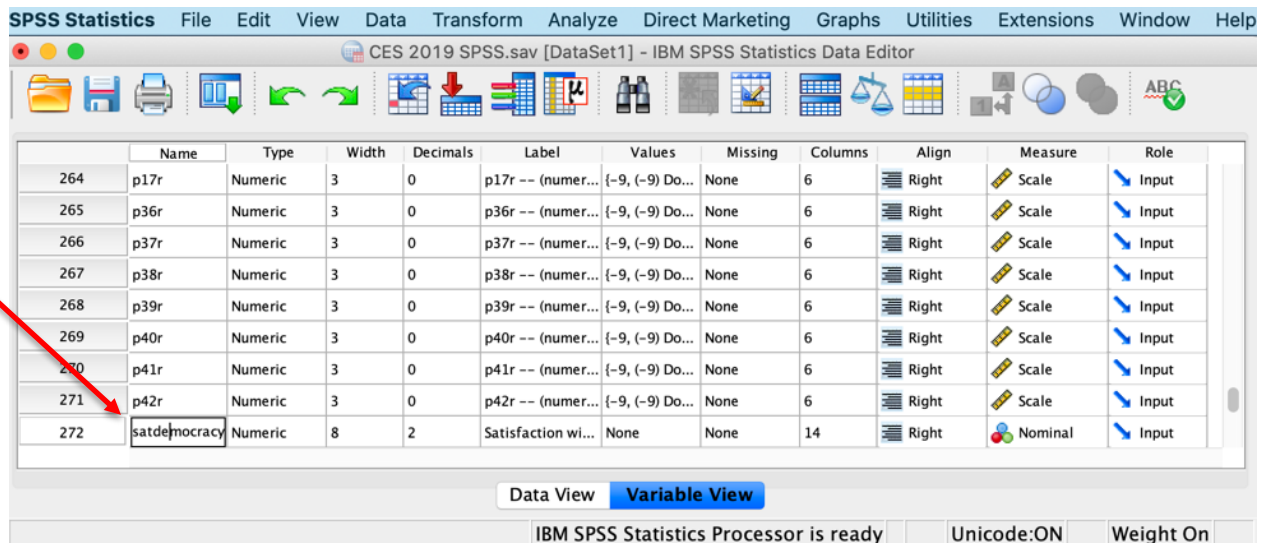
To add or change a variable label to an existing variable:

1. Open the **Variable View window** of the **Data Editor.**

2. Double - click on the variable label you wish to change / add in the **Label column.**
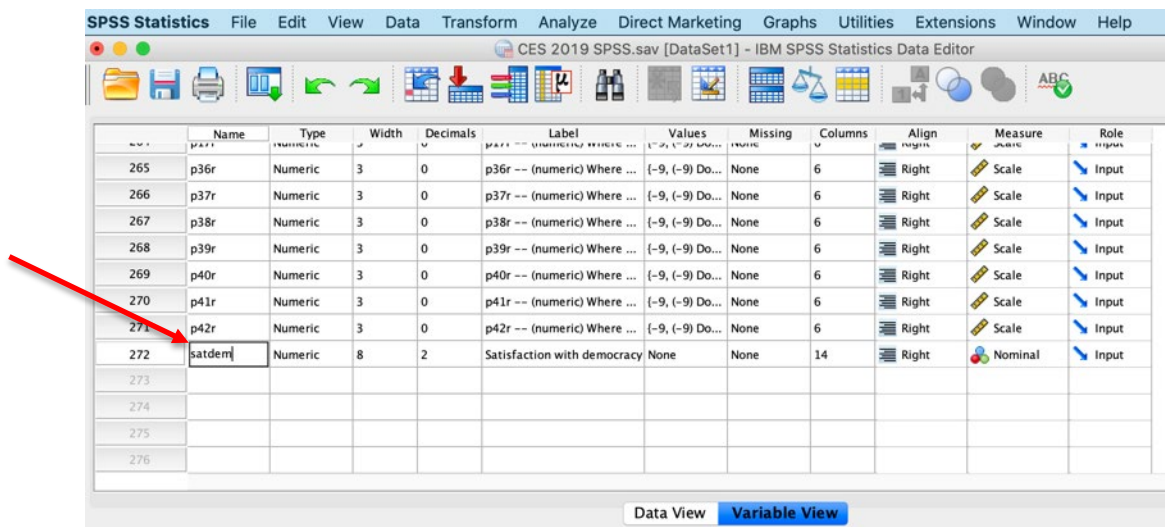
3. Type the new variable label.

To add value labels to an existing variable:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** of the variable you wish to change / add value labels.

3. In the **value labels window** that opens, enter the first numerical value you wish to assign a label to in the **Value: textbox** and the corresponding label in the **Label: textbox**.

4. Click the **Add** button to add the value label to the list of labels and then repeat steps 3 and 4 for each additional value.

5. Once all of the values have been assigned labels, click the **OK button** to add the labels.
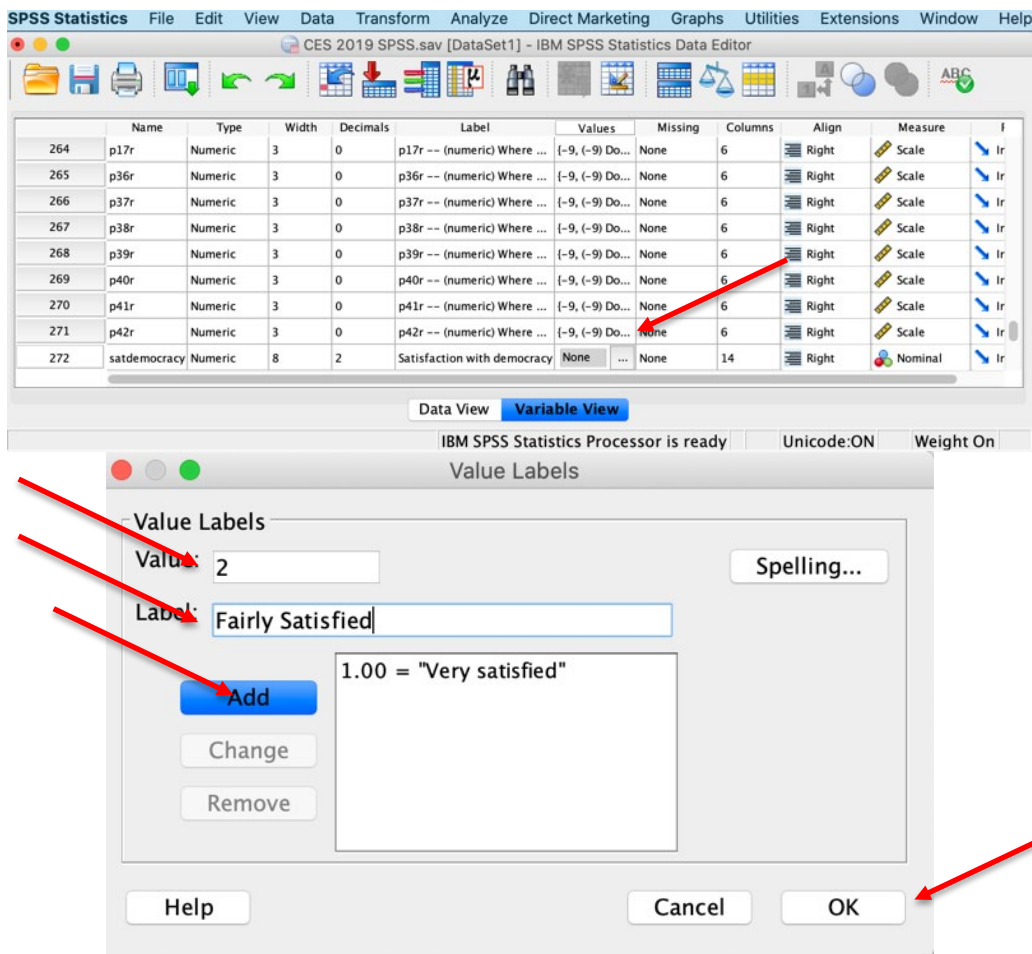
For example, we can use the steps listed above to rename our newly generated variable, *satdemocracy*, to *satdem* and add the appropriate value labels:

Next, use the steps outlined above to add the following value labels: 1 "Very satisfied" 2 "Fairly satisfied" 3 "Not very satisfied" 4 "Not satisfied at all"
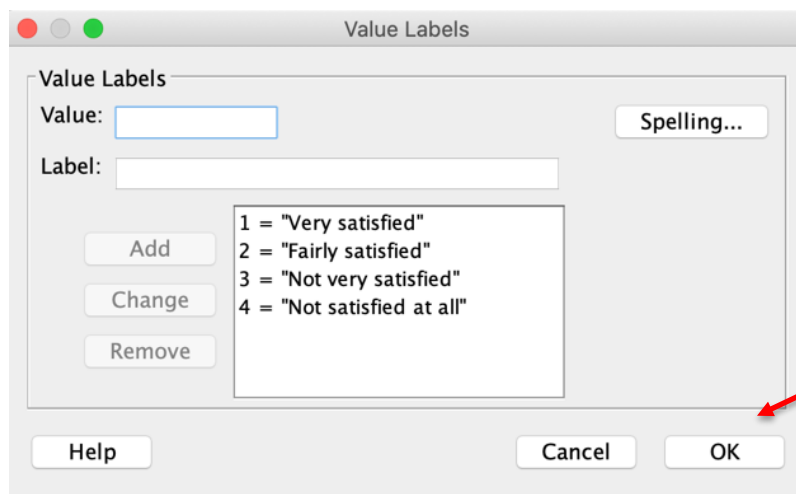
You should once again check your work by comparing the new variable against the original variable using the steps outlined above to produce a frequency table for each or by re-running the frequency command for the old and new variable from your syntax file. Note that you will need to update the new variable name to reflect the change! (*satdem*)

Statistics

		q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied	Satisfaction with democracy
N	Valid	4021	3959
	Missing	0	62

Frequency Table

q6 -- On the whole, are you very satisfied, fairly satisfied, not very satisfied

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	54	1.3	1.3	1.3
	(−8) Refused	8	.2	.2	1.5
	(1) Very satisfied	565	14.1	14.1	15.6
	(2) Fairly satisfied	2248	55.9	55.9	71.5
	(3) Not very satisfied	828	20.6	20.6	92.1
	(4) Not satisfied at all	318	7.9	7.9	100.0
	Total	4021	100.0	100.0	

Satisfaction with democracy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	'Very satisfied'	565	14.1	14.3	14.3
	'Fairly satisfied'	2248	55.9	56.8	71.0
	'Not very satisfied'	828	20.6	20.9	92.0
	'Not satisfied at all'	318	7.9	8.0	100.0
	Total	3959	98.5	100.0	
Missing	System	62	1.5		
Total		4021	100.0		

27

You should note that the raw frequencies for each category of the two variables are the same, even though the relative (percent) and cumulative frequencies differ. This is due to the exclusion of respondents who answered "don't know" or "refused" in the new variable. Repeat the steps outlined in Part IV to see how the removal of these response categories affects the summary statistics.

Tip: SPSS does not let you paste the rename command or the value labels command to the syntax file. If you wish to make the same changes using the syntax file, you would use the following syntax:

> RENAME VARIABLE (satdemocracy = satdem).
>
> ADD VALUE LABELS *satdem*
> 1 'Very satisfied'
> 2 'Fairly satisfied'
> 3 'Not very satisfied'
> 4 'Not satisfied at all'.

Be sure to include the "." at the end of the command! Using copy and paste, you can modify this syntax as needed for adding other variable labels in your syntax file.

Check-In Point

We covered a bit of ground in this section: you now know the reasons why you should never recode original variables (and thus will avoid future despair when discovering recoding errors). You know how to create new variables, and recode, rename, and add variable and value labels to those new variables. These are likely to be some of the most frequently used procedures when working with data.

Part VI: Creating Bar Graphs and Pie Charts

Destination

By the end of this section, you will be able to:
- *create bar graphs and pie charts.*

SPSS offers a range of graphing options. While the numerous options available for graphing within SPSS are beyond the scope of this handbook, we do outline the steps to create basic graphs and charts that you may use to display univariate frequency data.

The graphing options in SPSS are listed under **Graphs** in the top menu bar. Let's report the frequency distribution of our recoded satisfaction with democracy variable in a bar chart and a pie chart using the following steps:

Bar chart:

1. Select **Graphs > Legacy Dialogs > Bar…** from the top menu bar.
2. Select **Simple** and then click the **Define button** from the graph options window (note the other options available in this window).
3. Choose what the bars should represent. In this example, we will display **% of cases**.
4. Add variable *satdem* to the **Category Axis** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** beside the **Category Axis** textbox. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Click on the **Titles…** button to add a title, subtitle and / or footnote to your graph and then click on the

Continue button.

6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

7. Compare your results with ours.
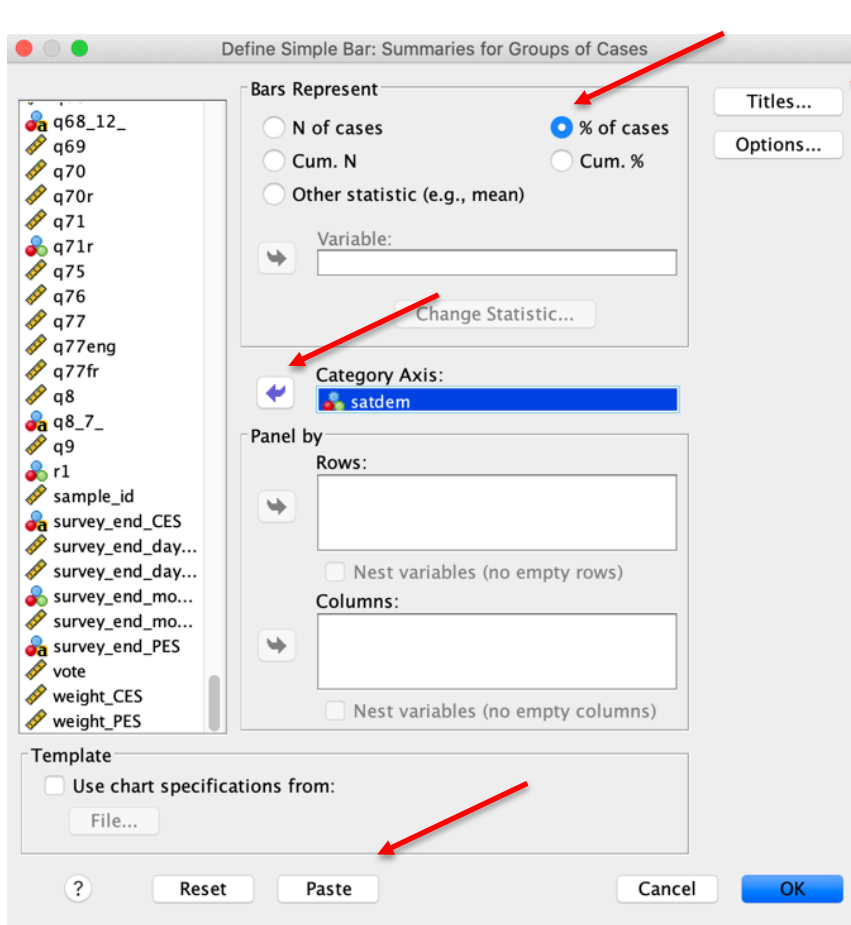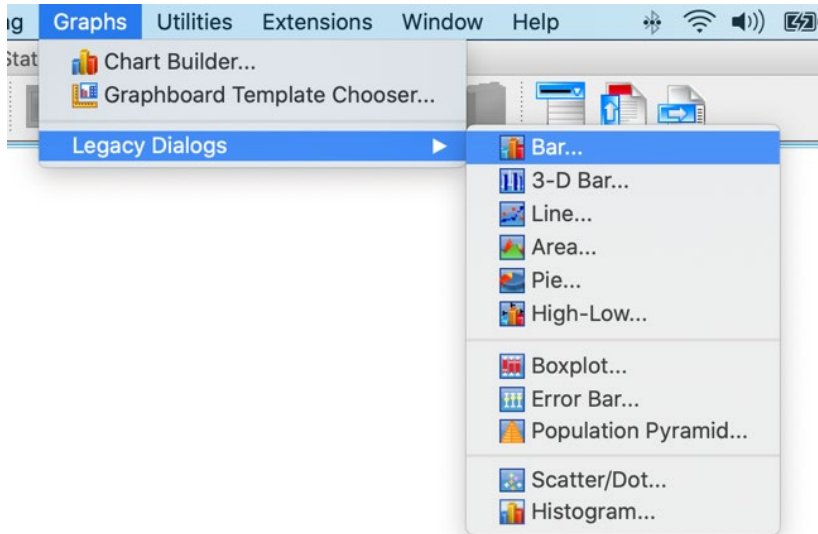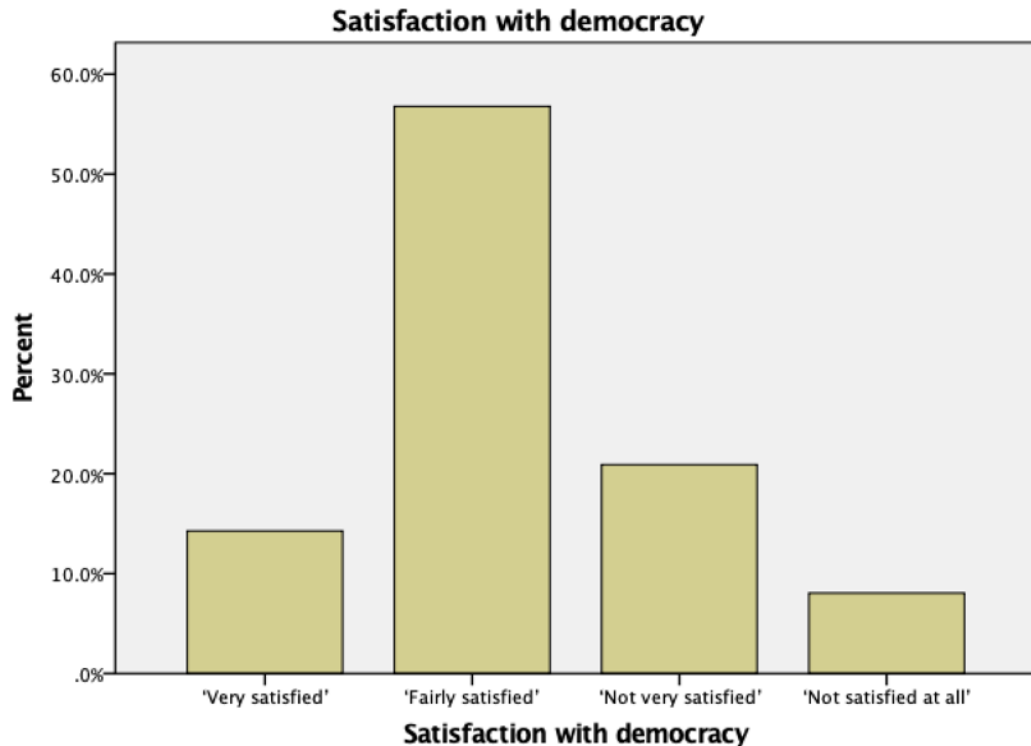
➔ Graph

Pie chart:

1. Select **Graphs > Legacy Dialogs > Pie…** from the top menu bar.
2. Select **Summaries for groups of cases** and then click the **Define button** from the graph options window (note the other options available in this window).
3. Choose what the slices should represent. In this example, we will display **% of cases**.
4. Add variable *satdem* to the **Define slices by:** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** beside the **Define slices by:** textbox. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Click on the **Titles…** button to add a title, subtitle and / or footnote to your graph and then click on the **Continue** button.
6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
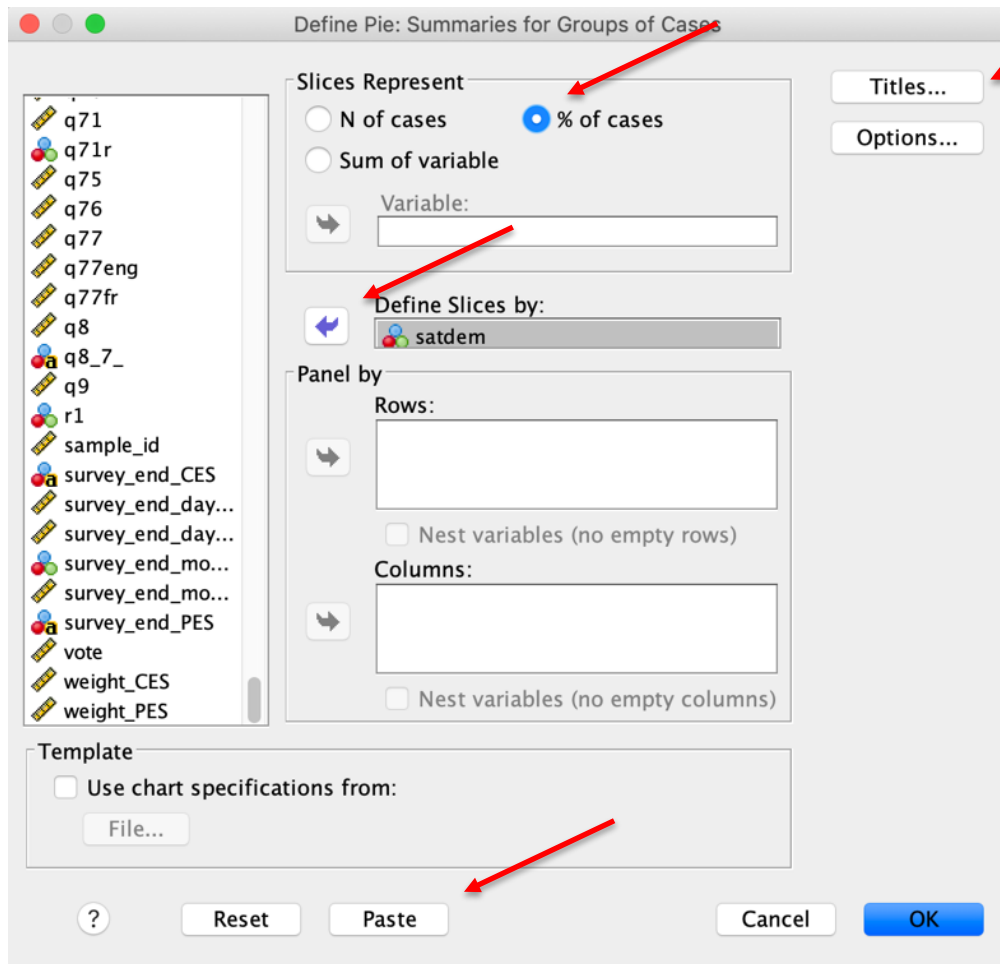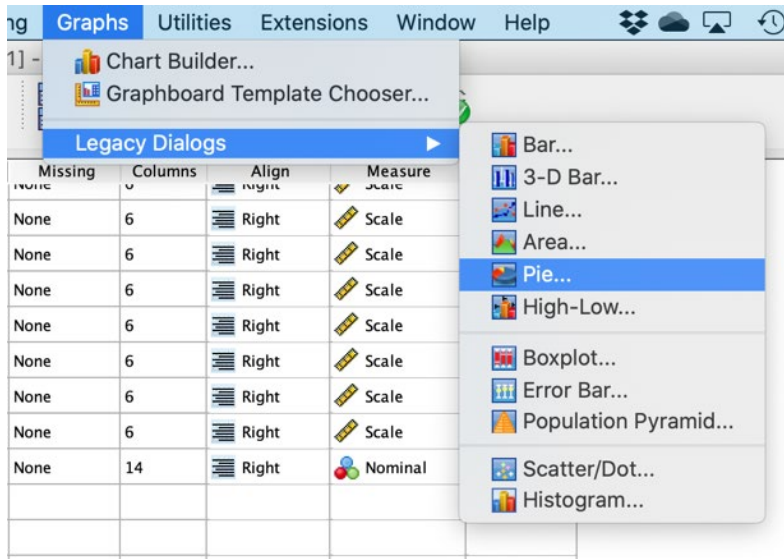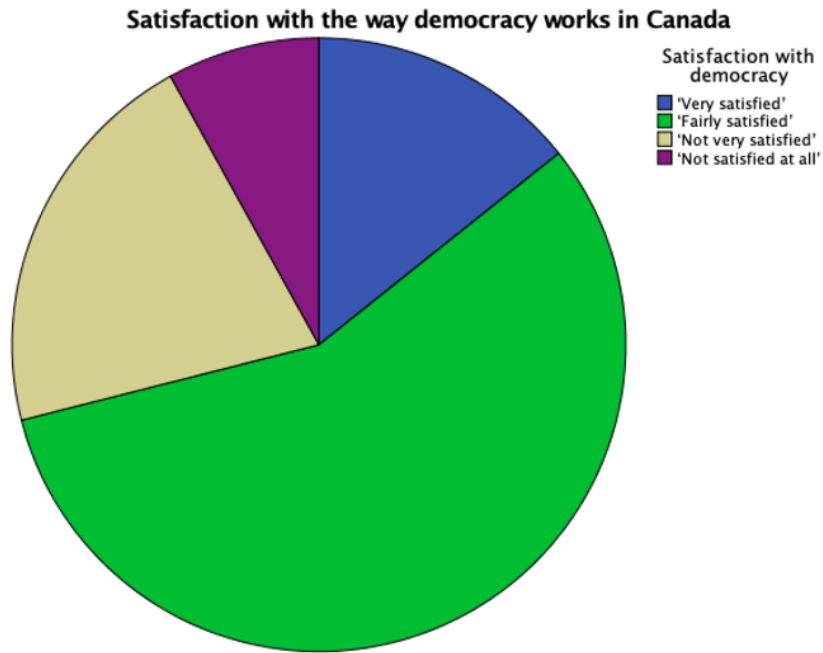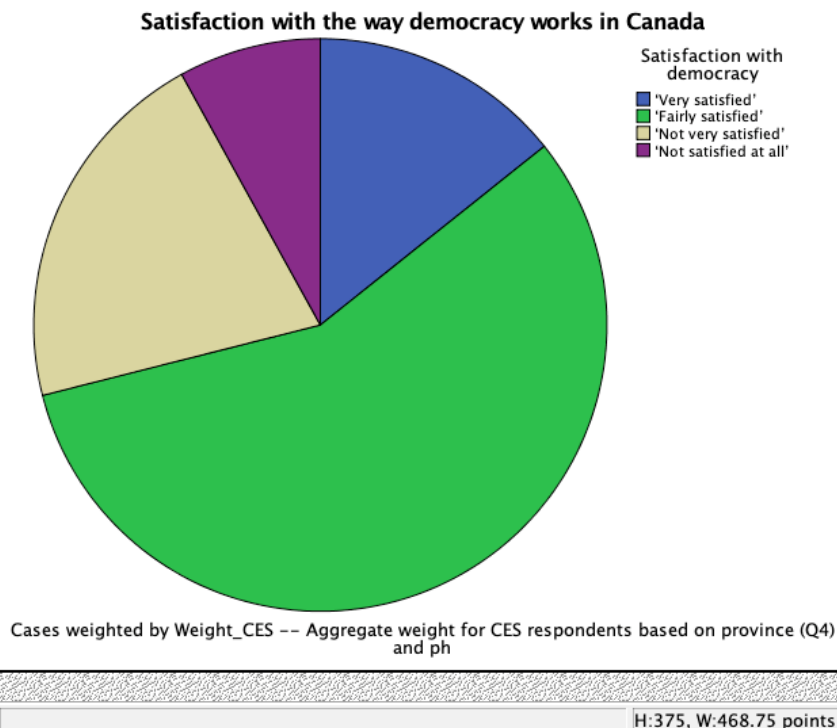7. Compare your results with ours.

➡ **Graph**

Satisfaction with the way democracy works in Canada

Satisfaction with democracy
- 🟦 'Very satisfied'
- 🟩 'Fairly satisfied'
- ⬜ 'Not very satisfied'
- 🟪 'Not satisfied at all'

There are a number of ways to modify how you graph displays and what information is included. In addition to the SPSS help file for graphing, you can make changes to the graph after it has been created by double-clicking on the graph in the **Output window** to open the **Chart Editor** feature. Once your changes have been made in the **Chart Editor window**, close the window to display your updated graph in the **Output window**.

Chart Editor

Satisfaction with the way democracy works in Canada

Satisfaction with democracy
- 🟦 'Very satisfied'
- 🟩 'Fairly satisfied'
- ⬜ 'Not very satisfied'
- 🟪 'Not satisfied at all'

Cases weighted by Weight_CES -- Aggregate weight for CES respondents based on province (Q4) and ph

H:375, W:468.75 points
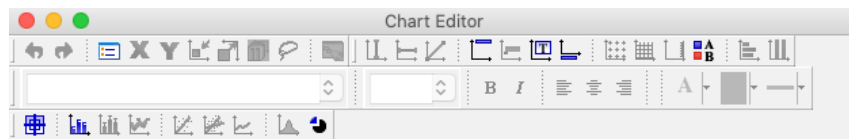
Check-In Point

Congratulations – you are now able to create basic univariate graphs. If you have an interest in more advanced graphing features, be sure to review the SPSS graph help files and other online resources.

<center><h3>Part VII: Comparing Two Independent Samples</h3></center>
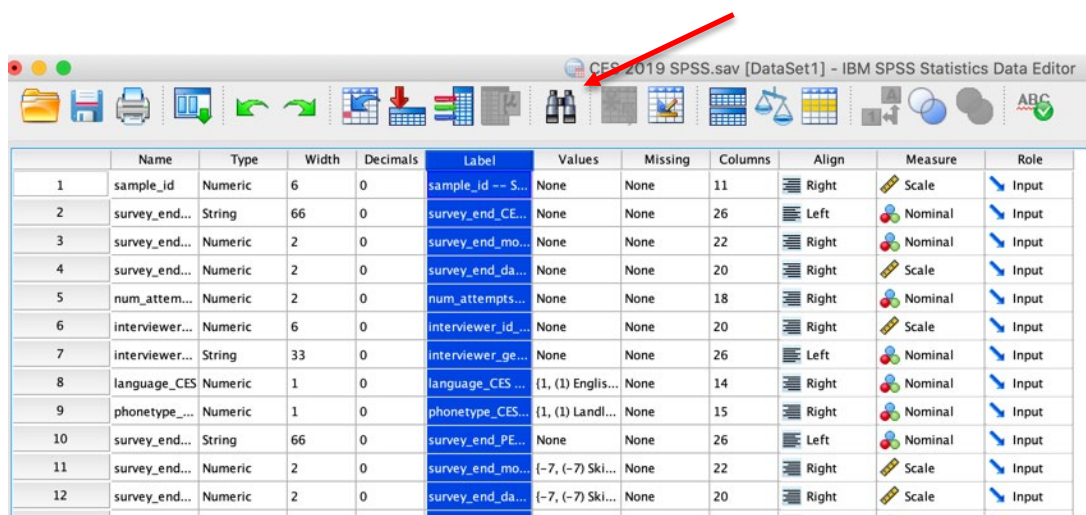
Destination

By the end of this section, you will be able to:

- *use a t-test to assess differences of means between two independent samples.*

In Chapter 13, we consider how we often wish to compare the means of two independent groups to see if they differ. To assess differences of means between two groups, we can use a t-test. To do this, we need a variable with our two groups of interest (for example, a treatment group and a control group from an experimental study) and an interval/ratio variable for which we expect a difference between the two groups. (Recall that means should only be used with interval/ratio level variables.)

For example, let's compare the average income (interval/ratio variable) of men and women (dichotomous variable).[3] We can use the following steps to find, examine, recode, and check the variables that we will use for this analysis:

1. In the **Variable View window** of the **Data Editor**, click on the **Label** column heading to highlight the **Label** column.
2. Click on the **Binocular** icon in the top row of the **Variable View window** or select **Edit – Find** from the top menu bar to open a search window.
3. Type "income" into the search box and click the **Find Next** button to locate all variable labels that contain the keyword.
4. Your search should bring you to variable *q69*, the respondent's reported income.
5. Type "gender" into the search box and click the **Find Next** button to locate all variable labels that contain the keyword.
6. Your search should bring you to variable *q3*, the respondent's reported gender.

---

[3] Note that the CES asks about household income. We have used this as a proxy for personal income in this example.
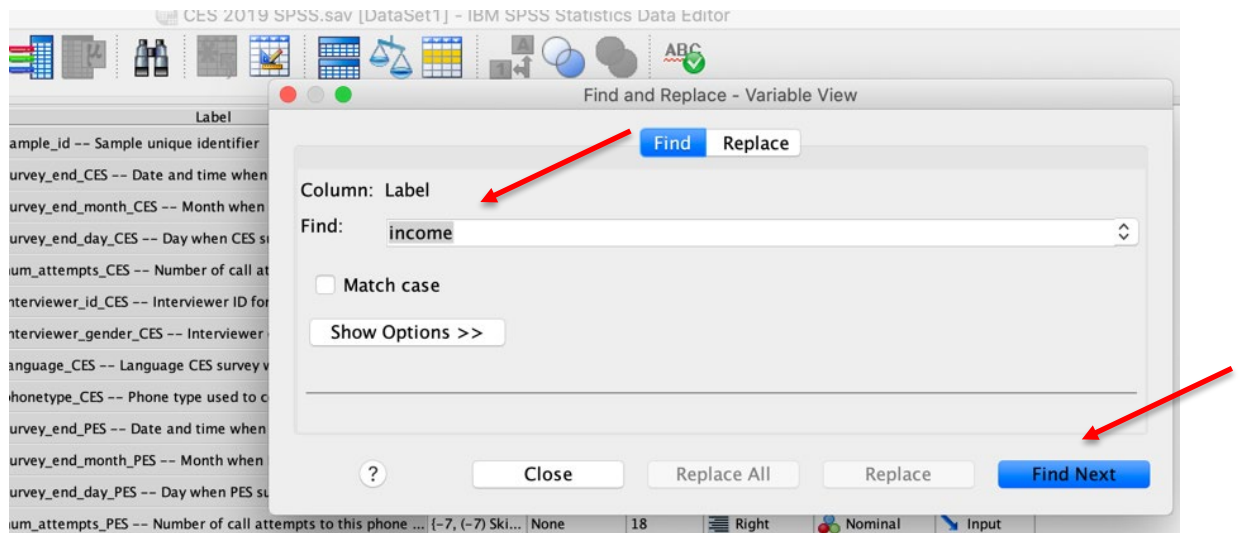
The next step in the process is to examine the value labels and distribution of the original variable:

1. Click on **Utilities > Variables** from the top menu bar in SPSS to open the **Variables** window.
2. Click on variable *q69* from the **Variable** column to display the variable information including the value labels. You can then click on variable *q3* to view information for this variable. Note that this window can remain open for reference while you generate a frequency distribution for the variable.
3. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
4. Click on the **Reset button** to clear any previous work.
5. Add variables *q69* and *q3* to the **Variable(s)** list by selecting them from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add them. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Ensure that the box beside **Display Frequency Tables** is checked**.**
7. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
8. Compare your results with our results, reported below. Note that we only include partial results for *q69*.

q69 -- Could you please tell me your total household income before taxes for the

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	378	9.4	9.4	9.4
	(−8) Refused	599	14.9	14.9	24.3
	0	115	2.9	2.9	27.2
	1000	2	.0	.0	27.2
	2000	1	.0	.0	27.2
	3000	2	.0	.0	27.3
	4000	1	.0	.0	27.3
	5000	1	.0	.0	27.3
	6000	6	.2	.2	27.5
	7000	1	.0	.0	27.5
	8000	1	.0	.0	27.5
	9000	8	.2	.2	27.7
	10000	14	.4	.4	28.1
	11000	3	.1	.1	28.2
	12000	19	.5	.5	28.6
	13000	11	.3	.3	28.9
	14000	7	.2	.2	29.1
	15000	18	.4	.4	29.5
	16000	7	.2	.2	29.7
	17000	8	.2	.2	29.9
	18000	17	.4	.4	30.3
	19000	9	.2	.2	30.5
	20000	39	1.0	1.0	31.5

q3 -- Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(1) Male	2239	55.7	55.7	55.7
	(2) Female	1781	44.3	44.3	100.0
	(3) Other	1	.0	.0	100.0
	Total	4021	100.0	100.0	

Given that there are individuals who refused to answer the income question or did not know their income, we will want to recode the income variable to set these cases to "System Missing". For the gender variable, there are not enough individuals for a meaningful analysis who choose "Other" as their gender. As such, we will recode this variable to create a dichotomous gender variable. To so, we begin by creating a new variable, *income*, from *q69* using the following steps:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q69* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it.

(Note that the variable names should be in alphabetical order if you followed our recommendation above).

4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *income* and label it "Income in dollars".
5. Click on the button titled **Old and New Values** to recode the new variable.
6. Given that we want to recode -9 and -8 to **System-missing**, we can enter each value separately or use the **Range** option by clicking the radio button next to the option and adding the values and then select the radio button next to **System-missing** in the **New value** column
7. Click on the **Add** button to add the transformation to the recode command.
8. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.
9. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**
10. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
11. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
12. Click on the **Reset button** to clear any previous work.
13. Add variable *q69* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
14. Add variable *income* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
15. Ensure that the box beside **Display Frequency Tables** is checked**.**
16. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
17. Compare your results to our abbreviated results, noting especially the difference in the number of valid cases and missing cases for the two variables as reported in the **Statistics box**:

Frequencies

Statistics

		q69 -- Could you please tell me your total household income before taxes for the	Income in dollars
N	Valid	4021	3044
	Missing	0	977

Frequency Table

q69 -- Could you please tell me your total household income before taxes for the

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(−9) Don't know	378	9.4	9.4	9.4
	(−8) Refused	599	14.9	14.9	24.3
	0	115	2.9	2.9	27.2
	1000	2	.0	.0	27.2
	2000	1	.0	.0	27.2
	3000	2	.0	.0	27.3
	4000	1	.0	.0	27.3
	5000	1	.0	.0	27.3
	6000	6	.2	.2	27.5
	7000	1	.0	.0	27.5
	8000	1	.0	.0	27.5
	9000	8	.2	.2	27.7
	10000	14	.4	.4	28.1
	11000	3	.1	.1	28.2
	12000	19	.5	.5	28.6

Income in dollars

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	115	2.9	3.8	3.8
	1000.00	2	.0	.1	3.8
	2000.00	1	.0	.0	3.9
	3000.00	2	.0	.1	3.9
	4000.00	1	.0	.0	4.0
	5000.00	1	.0	.0	4.0
	6000.00	6	.2	.2	4.2
	7000.00	1	.0	.0	4.2
	8000.00	1	.0	.0	4.2
	9000.00	8	.2	.3	4.5
	10000.00	14	.4	.5	5.0
	11000.00	3	.1	.1	5.1
	12000.00	19	.5	.6	5.7

We can repeat the same steps for the second variable, *q3*:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.

3. Add variable *q3* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *gender* and label it "Dichotomous Gender".

5. Click on the button titled **Old and New Values** to recode the new variable.

6. Add 3 (the value for the Other category of the original variable) to the **Value** text box under the **Old Value** column and then select the radio button next to **System-missing** in the **New value** column. Click on the **Add** button to add the transformation to the recode command.

7. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.

8. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**

9. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

10. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.

11. Click on the **Reset button** to clear any previous work.

12. Add variable *q3* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

13. Add variable *gender* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

14. Ensure that the box beside **Display Frequency Tables** is checked**.**

15. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

18. Compare your results with ours, noting especially the difference in the number of valid cases and missing cases for the two variables as reported in the **Statistics box**:

Statistics

		q3 -- Gender	Dichotomous Gender
N	Valid	4021	4020
	Missing	0	1

Frequency Table

q3 -- Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(1) Male	2239	55.7	55.7	55.7
	(2) Female	1781	44.3	44.3	100.0
	(3) Other	1	.0	.0	100.0
	Total	4021	100.0	100.0	

Dichotomous Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	2239	55.7	55.7	55.7
	Female	1781	44.3	44.3	100.0
	Total	4020	100.0	100.0	
Missing	System	1	.0		
Total		4021	100.0		

Once you have confirmed that your new variable has been recoded correctly, you can add value labels to the variables in the **Variable View window** of the **Data Editor (**or using your syntax code, as explained above). For our income variable, we do not need to add value labels since it is an interval/ratio level measures (the values *are* the labels!). To add labels to the gender variable:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** for the new variable (*gender*).

3. In the **value labels window** that opens, enter 1 in the **Value: textbox** and "Male" in the **Label: textbox**.

4. Click the **Add** button to add the value label to the list of labels and then repeat steps 3 and 4 to assign the label "Female" to value 2.

5. Click the **OK button** to add the labels and check your work by re-running the frequency distribution for the new and original variables.

With the variables now recoded, you can compare the average income of the men and women using a t-test.

To do so:

1. Select **Analyze > Compare Means > Independent Sample T-test …** from the top menu bar.
2. Click on the **Reset button** if any variables appear in the **Test Variable(s):** or **Grouping Variable:** list.

3. Add variable *income* to the **Test Variable(s):** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

4. Add variable *gender* to the **Grouping Variable:** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above). You should note the two question marks (? ?) beside the **Grouping Variable**. This is because you need to indicate which groups are to be compared. This is done by providing the category values for the groups we are comparing.

5. Click on the **Define Groups** button and add the value 1 for Group 1 and 2 for Group 2 (these are the values assigned to males and females, respectively). Click the **Continue button** once these values have been added.

6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on **the Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

7. Compare your results with ours.
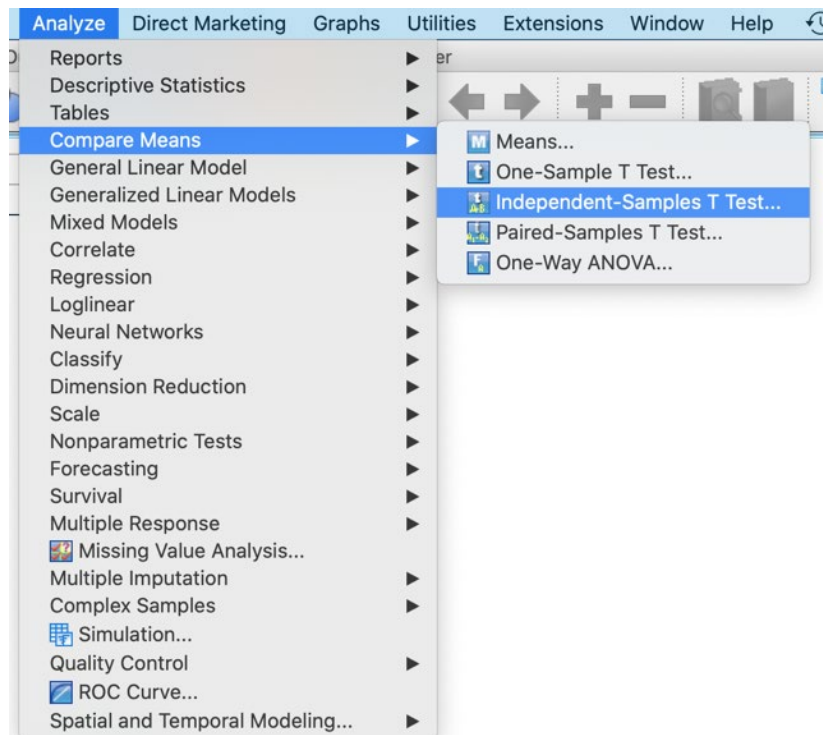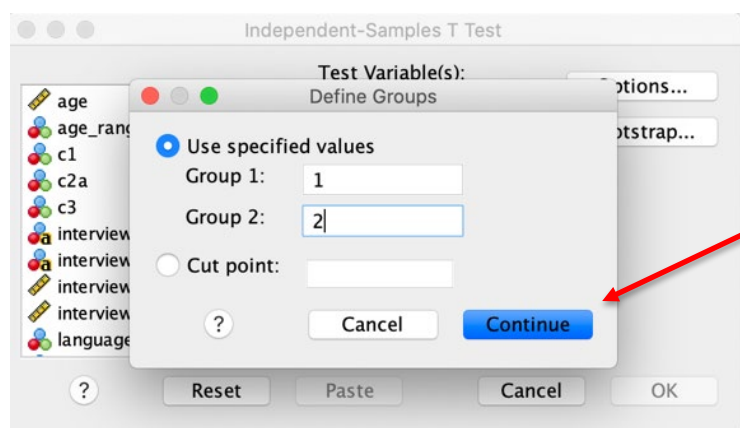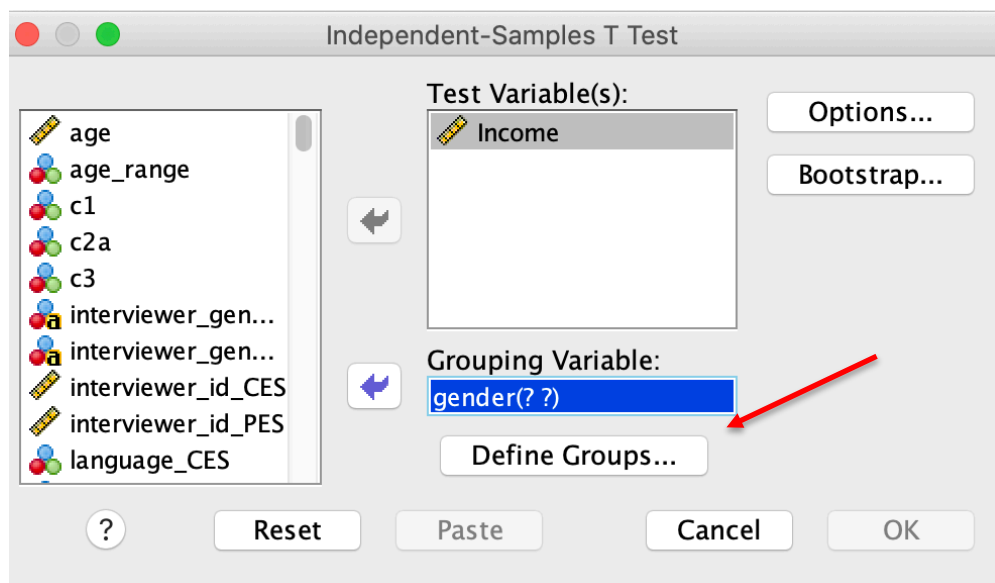
T-Test

Group Statistics

	Dichotomous Gender	N	Mean	Std. Deviation	Std. Error Mean
Income in dollars	Male	1754	116243.583	128803.507	3075.87062
	Female	1290	97642.4991	111024.968	3091.63394

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Income in dollars	Equal variances assumed	4.396	.036	4.170	3041	.000	18601.0843	4460.28797	9855.59996	27346.5687
	Equal variances not assumed			4.265	2965.712	.000	18601.0843	4361.09854	10049.9984	27152.1702

Do the incomes of men and women differ? The results suggest that they do. From the table above we see that the mean income for men is \$116,243.58 compared to a mean income of \$97,642.50 for women, a difference of \$18,601.08 (the value reported as **Mean Difference** in the table). In other words, the results indicate that, on average, men earn \$18,601.08 a year more than women. We provide a more detailed discussion of this test and the results reported here in the text (see Chapter 13).

In Chapter 13, we also discuss statistical significance. You will recall that researchers use a one-tailed test

when they hypothesize a specific direction for a relationship and use a two-tailed test when they do not hypothesize a direction. The SPSS t-test automatically reports the level of statistical significance of the difference of the two means, for a two-tailed test. Given that we did not make any assumptions about the direction of the relationship in advance, we would use the results from this two-tailed test, a result that indicates a statistically significant relationship at $p<0.001$. As a final point, note that SPSS reports the results for Levene's Test for Equality of Variance. This is a test that assesses whether or not the two groups (men and women in this example) have the same variance on the dependent variable (income). If the result is statistically significant (e.g. the value in the Sig. cell is <0.05), the we can NOT assume equal variances and we must read the results from the second row (Equal variances not assumed). If the value is >0.05, we can assume equal variances and read the results from the top row. In this example, the significance value indicates unequal variances.

Check-In Point

If you are following along with the examples by running your own data, you should now be able to use a t-test to assess differences of means between two independent samples. This is a useful skill, and we encourage you to practice by exploring other income differences between other dichotomous groups in the CES dataset. For example, do the average incomes of university and non-university graduates differ?

<center>**Part VIII: Examining Bivariate Relationships for Nominal and/or Ordinal Variables**</center>

Destination

By the end of this section, you will be able to:

- *create crosstabulations;*
- *calculate measures of association; and*
- *calculate Chi Square.*

In Chapter 12, we introduce you to the four questions we must answer when assessing whether there is a relationship between two variables:

1. What is the form/direction of the relationship?
2. How strong is the relationship?
3. Is the relationship statistically significant?
4. What happens to the relationship when we control for other variables?

In this section, we will focus on how to use SPSS to help answer the first three of these questions for nominal and/or ordinal variables. In the section that follows, we will look at interval/ratio variables. The final section of this handbook will consider the fourth and final question.

To examine the relationship between two nominal and/or ordinal variables, we create a cross-tabulation (contingency table), calculate the appropriate measure of association, and calculate the appropriate inferential statistic. With SPSS, we can do all of this easily using the **Analyze** > **Descriptive Statistics > Crosstabs…** command.

For example, let's test the hypothesis that those with higher levels of income will also be more interested in politics. Before we can do so, we must recode our variables. For this example, we will begin by recoding our dependent variable (DV), political interest, into a three-point measure ranging from low to high (terciles), as follows:

1. In the **Variable View window** of the **Data Editor**, select the **Label** column and click on the **binoculars icon** from the options menu to open a search box and search for "interest". This should bring you to variable *q9* a measure of political interest.
2. Click on the three dots in the **Values** cell in the **Variable View window** of the **Data Editor** to view the value labels for *q9*. Note that the values range from 0 (No interest at all) to 10 (Great deal of interest) with -9, -8, and -7 used to denote *Don't Know*, *Refused*, and *Skipped* responses respectively. You will begin by recoding this variable to remove these cases.
3. Select **Transform > Recode into Different Variables…** from the top menu bar.
4. Click on the **Reset button** to clear any previous work.
5. Add variable *q9* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *polinterest* and label it "Political interest".
7. Click on the button titled **Old and New Values** to recode the new variable.
8. Use the **Range** option to recode values -9 through -7 to system missing by selecting the radio button next to **System-missing** in the **New value** column. Click on the **Add** button to add the transformation to the recode command.
9. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.
10. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**
11. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
12. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
13. Click on the **Reset button** to clear any previous work.
14. Add variable *q9* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
15. Add variable *polinterest* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
16. Ensure that the box beside **Display Frequency Tables** is checked**.**
17. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
19. Compare your results with ours, noting especially the difference in the number of valid cases and missing cases for the two variables as reported in the **Statistics box**: [4]

---

[4] Note that there are only 5 cases reported as system missing in the new variable when, based on the frequencies reported for the original variable, it appears that there should be 6 missing cases. The inconsistency is due to the use of the weight variable in SPSS.

Statistics

		q9 -- Using a scale from 0 to 10, where 0 means no interest at all and 10 means	Political interest
N	Valid	4021	4016
	Missing	0	5

Frequency Table

q9 -- Using a scale from 0 to 10, where 0 means no interest at all and 10 means

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(-9) Don't know	5	.1	.1	.1
	(-8) Refused	1	.0	.0	.1
	(0) No interest at all	83	2.1	2.1	2.2
	(1)	38	1.0	1.0	3.1
	(2)	70	1.7	1.7	4.9
	(3)	92	2.3	2.3	7.2
	(4)	124	3.1	3.1	10.3
	(5)	366	9.1	9.1	19.4
	(6)	345	8.6	8.6	27.9
	(7)	566	14.1	14.1	42.0
	(8)	780	19.4	19.4	61.4
	(9)	427	10.6	10.6	72.0
	(10) Great deal of interest	1125	28.0	28.0	100.0
	Total	4021	100.0	100.0	

Political interest

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	83	2.1	2.1	2.1
	1.00	38	1.0	1.0	3.0
	2.00	70	1.7	1.8	4.8
	3.00	92	2.3	2.3	7.1
	4.00	124	3.1	3.1	10.1
	5.00	366	9.1	9.1	19.2
	6.00	345	8.6	8.6	27.8
	7.00	566	14.1	14.1	41.9
	8.00	780	19.4	19.4	61.3
	9.00	427	10.6	10.6	72.0
	10.00	1125	28.0	28.0	100.0
	Total	4016	99.9	100.0	
Missing	System	5	.1		
Total		4021	100.0		

The results show that we no longer include cases that did not provide an actual value for their level of political interest. To transform this variable into terciles, we will use the cumulative percent column to determine the values for approximately 33%, 66% and 99% of the sample. In this case the ranges would be 0-

6, 7-8, and 9-10, respectively.[5] We will use these values to create a political interest tercile variable as follows:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *polinterest* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *polinteresttercile* and label it "Political interest terciles".
5. Click on the button titled **Old and New Values** to recode the new variable.
6. Use the **Range** option to recode values 0 through 6 to 1 by selecting the radio button next to **Value** in the **New value** column and entering 1 in the text box. Click on the **Add** button to add the transformation to the recode command.
7. Use the **Range** option to recode values 7 through 8 to 2 by selecting the radio button next to **Value** in the **New value** column and entering 2 in the text box. Click on the **Add** button to add the transformation to the recode command.
8. Use the **Range** option to recode values 9 through 10 to 3 by selecting the radio button next to **Value** in the **New value** column and entering 3 in the text box. Click on the **Add** button to add the transformation to the recode command.
9. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**
10. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
11. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
12. Click on the **Reset button** to clear any previous work.
13. Add variable *polinterest* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
14. Add variable *polinteresttercile* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
15. Ensure that the box beside **Display Frequency Tables** is checked**.**
16. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
17. Compare your results with ours.

---

[5] An alternative way to recode this variable would be to set respondents that choose 0-4 as low interest, 5 as the mid-point, and all responses over 5 as high interest. We opt to use the cut-off points reported in the cumulative percent to produce three roughly equal sized groups.

Frequency Table

Political interest

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	83	2.1	2.1	2.1
	1.00	38	1.0	1.0	3.0
	2.00	70	1.7	1.8	4.8
	3.00	92	2.3	2.3	7.1
	4.00	124	3.1	3.1	10.1
	5.00	366	9.1	9.1	19.2
	6.00	345	8.6	8.6	27.8
	7.00	566	14.1	14.1	41.9
	8.00	780	19.4	19.4	61.3
	9.00	427	10.6	10.6	72.0
	10.00	1125	28.0	28.0	100.0
	Total	4016	99.9	100.0	
Missing	System	5	.1		
Total		4021	100.0		

Political interest terciles

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1118	27.8	27.8	27.8
	2.00	1346	33.5	33.5	61.3
	3.00	1552	38.6	38.7	100.0
	Total	4016	99.9	100.0	
Missing	System	5	.1		
Total		4021	100.0		

Assuming that your recode was successful, you now need to add value labels to the new political interest tercile variable that you created. To do so, you can use the Syntax file, as we have described in the tip above, or do the following:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** next to the *polinteresttercile* variable to add value labels.

3. In the **value labels window** that opens, enter 1 in the **Value:** textbox and "Low interest" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

4. Enter 2 in the **Value:** textbox and "Middle interest" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

5. Enter 3 in the **Value:** textbox and "High interest" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

6. Once all of the values have been assigned labels, click the **OK button** to add the labels. It is advisable to confirm that the value labels have been assigned correctly by rerunning a frequency distribution for this variable.

We also will recode our independent variable (IV), income, into terciles (low, middle, and high income). It is important to note that many respondents (approximately 25% in the 2019 CES) did not report their actual income. To help reduce the number of non-responses, individuals who refuse or report that they do not know

their actual income are asked a follow-up question that provides income categories for the respondent to choose instead of stating their actual income. In this example, we want to combine the responses from the two questions to generate a new income category variable. The challenge of doing this is that SPSS will not allow us to add variables that have SYSMIS values on any of the variables that we wish to combine. One work around is to set the missing cases (SYSMIS) to "0".[6] However, you must ensure that "0" is NOT an actual variable value. We outline this process in the steps below that first transforms our income variable (we use the income variable that we generated for the t-test above) into income categories to match the income categories for variable, *q70*, which reports the response to the follow-up income questions for individuals who did not provide their actual income (you can search the **Labels** column of the **Variable View window** in the **Data Editor** to find all variables that include the key word "income"). We begin by comparing our two income measures as follows:

1. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *income* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add variable *q70* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Ensure that the box beside **Display Frequency Tables** is checked.
6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

You should see from the **Statistics** results dialogue box that the number of missing cases on the *Income* variable matches the number of Valid cases on *q70*. Note that none of the income categories for *q70* are assigned a value of "0" (you can confirm this my checking the value labels for *q70*). You will use the income value ranges from *q70* (use **Utilities > Variables** from the top menu bar in SPSS to open the **Variables window** to view the value labels) to transform the income variable into income categories as follows:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *income* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *incomegrptmp* and label it "Temporary income group variable".
5. Click on the button titled **Old and New Values** to recode the new variable.
6. To recode all cases with a value of 0 to 1, click on the radio button next to the **Value** option under **Old Value** and enter 0 in the text box. In the **New value** column, select the radio button next to **Value** and enter 1 in the text box. Click on the **Add** button to add the transformation to the recode command.
7. Use the **Range** option to recode values 1 through 30000 to 2 by selecting the radio button next to **Value** in the **New value** column and entering 2 in the text box. Click on the **Add** button to add the transformation to the recode command.

---

[6] Another option is to use a series of IF…THEN type statements in the SPSS syntax file.

8. Use the **Range** option to recode values 30001 through 60000 to 3 by selecting the radio button next to **Value** in the **New value** column and entering 3 in the text box. Click on the **Add** button to add the transformation to the recode command.

9. Use the **Range** option to recode values 60001 through 90000 to 4 by selecting the radio button next to **Value** in the **New value** column and entering 4 in the text box. Click on the **Add** button to add the transformation to the recode command.

10. Use the **Range** option to recode values 90001 through 110000 to 5 by selecting the radio button next to **Value** in the **New value** column and entering 5 in the text box. Click on the **Add** button to add the transformation to the recode command.

11. Use the **Range** option to recode values 110001 through 150000 to 6 by selecting the radio button next to **Value** in the **New value** column and entering 6 in the text box. Click on the **Add** button to add the transformation to the recode command.

12. Use the **Range** option to recode values 150001 through 200000 to 7 by selecting the radio button next to **Value** in the **New value** column and entering 7 in the text box. Click on the **Add** button to add the transformation to the recode command.

13. Use the **Range** option to recode values 200001 through 2130000 to 8 by selecting the radio button next to **Value** in the **New value** column and entering 8 in the text box. Click on the **Add** button to add the transformation to the recode command.

14. Check the radio button next to System-missing under **Old Value** to recode SYSMIS to 0 by selecting the radio button next to **Value** in the **New value** column and entering 0 in the text box. Click on the **Add** button to add the transformation to the recode command.

15. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**

16. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

17. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.

18. Click on the **Reset button** to clear any previous work.

19. Add variable *income* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

20. Add variable *incomegrptmp* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

21. Ensure that the box beside **Display Frequency Tables** is checked**.**

22. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

23. Compare your results for *incomegrptmp* with ours.

Temporary income group variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	977	24.3	24.3	24.3
	1.00	115	2.9	2.9	27.2
	2.00	362	9.0	9.0	36.1
	3.00	618	15.4	15.4	51.5
	4.00	552	13.7	13.7	65.2
	5.00	313	7.8	7.8	73.0
	6.00	494	12.3	12.3	85.3
	7.00	340	8.5	8.5	93.8
	8.00	251	6.2	6.2	100.0
	Total	4021	100.0	100.0	

You now have a new income group variable that is matched to the income category variable (*q70*) categories with missing responses from the actual income variable (*income*) set to "0". You now need to recode *q70* missing cases to "0" by generating a second temporary income group variable as follows:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *q70* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *incomegrptmp2* and label it "Second temporary income group variable".
5. Click on the button titled **Old and New Values** to recode the new variable.
6. Check the radio button next to System-missing under **Old Value** to recode SYSMIS to 0 by selecting the radio button next to **Value** in the **New value** column and entering 0 in the text box. Click on the **Add** button to add the transformation to the recode command.
7. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.
8. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**
9. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
10. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
11. Click on the **Reset button** to clear any previous work.
12. Add variable *q70* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
13. Add variable *incomegrptmp2* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
14. Ensure that the box beside **Display Frequency Tables** is checked**.**
15. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a

Mac or CTRL – R with a Windows operating system).

16. Compare your results for *incomegrptmp2* with ours. You should note that all SYSMIS cases in the original variable are now assigned a value of "0".

Second temporary income group variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	−9.00	69	1.7	1.7	1.7
	−8.00	238	5.9	5.9	7.6
	.00	3044	75.7	75.7	83.3
	1.00	12	.3	.3	83.6
	2.00	107	2.7	2.7	86.3
	3.00	178	4.4	4.4	90.7
	4.00	147	3.7	3.7	94.4
	5.00	69	1.7	1.7	96.1
	6.00	67	1.7	1.7	97.8
	7.00	35	.9	.9	98.6
	8.00	55	1.4	1.4	100.0
	Total	4021	100.0	100.0	

The next step is to combine the two temporary income group variables using the following steps:

1. Select **Transform > Compute variable…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. In the **Target Variable:** text box, enter a name for your new variable. We will call ours *incomegrpmerged*.
4. Click on the **Type & Label** button to add a label for your new variable. We will label ours "Merged income groups". Click Continue after you have labeled your new variable.
5. Add variable *incomegrptmp* to the **Numeric Expression** text box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Click on the "+" symbol below the **Numeric Expression** text box or type "+" after the *incomegrptmp* variable name in the expression box to add it to *incomegrptmp2*.
7. Add variable *incomegrptmp2* to the **Numeric Expression** text box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above). The expression should appear as: *incomegrptmp + incomegrptmp2*
8. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
9. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
10. Click on the **Reset button** to clear any previous work.
11. Add variable *incomegrptmp* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
12. Add variable *incomegrptmp2* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
13. Add variable *incomegrpmerged* to the **Variable(s)** list by selecting it from the list of variables in the

left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

14. Ensure that the box beside **Display Frequency Tables** is checked**.**
15. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
16. Compare your results for *incomegrpmerged* with ours.

Merged income groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-9.00	69	1.7	1.7	1.7
	-8.00	238	5.9	5.9	7.6
	1.00	126	3.1	3.1	10.8
	2.00	469	11.7	11.7	22.4
	3.00	796	19.8	19.8	42.2
	4.00	699	17.4	17.4	59.6
	5.00	381	9.5	9.5	69.1
	6.00	561	14.0	14.0	83.1
	7.00	375	9.3	9.3	92.4
	8.00	306	7.6	7.6	100.0
	Total	4021	100.0	100.0	

You should note that the new merged income group variable (*incomegrpmerged*) no longer has any cases with a value of "0". However, there are cases with values of -8 and -9, representing "Don't know" and "Refused" from the original income group variable (*q70*). You will need to convert these values to SYSMIS and then assign value labels to the merged income group variable. In this case, since *incomegrpmerged* is NOT an original dataset variable, we will use the **Recode into Same Variables…** option:

1. Select **Transform > Recode into Same Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *incomegrpmerged* to the **Numeric Variables** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Click on the button titled **Old and New Values** to recode the variable.
5. Use the **Range** option to recode values -9 through -8 to SYSMIS by selecting the radio button next to **SYSMIS** in the **New value** column. Click on the **Add** button to add the transformation to the recode command.
6. Click on **Continue** to return to the main **Recode into Same Variables window.**
7. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
8. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
9. Click on the **Reset button** to clear any previous work.
10. Add variable *incomegrpmerged* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
11. Ensure that the box beside **Display Frequency Tables** is checked**.**

12. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
13. Compare your results with ours. Note that the 307 missing cases are those respondents who did not report their actual income or provide a response to the follow-up income category question (*q70*).

➜ Frequencies

Statistics

Merged income groups

N	Valid	3714
	Missing	307

Merged income groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	126	3.1	3.4	3.4
	2.00	469	11.7	12.6	16.0
	3.00	796	19.8	21.4	37.5
	4.00	699	17.4	18.8	56.3
	5.00	381	9.5	10.3	66.6
	6.00	561	14.0	15.1	81.7
	7.00	375	9.3	10.1	91.8
	8.00	306	7.6	8.2	100.0
	Total	3714	92.4	100.0	
Missing	System	307	7.6		
Total		4021	100.0		

The final transformation is to convert the merged income group variable into income terciles. We use the same approach for recoding income groups as we did for our political interest tercile variable, referring to the cumulative percent column to determine the values for approximately 33%, 66% and 99% of the sample. In this case the ranges would be 1-3, 4-5, and 6-8, respectively. We will use these values to create an income tercile variable as follows:

1. Select **Transform > Recode into Different Variables…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *incomegrpmerged* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *incometercile* and label it "Income terciles".
5. Click on the button titled **Old and New Values** to recode the new variable.

6. Use the **Range** option to recode values 1 through 3 to 1 by selecting the radio button next to **Value** in the **New value** column and entering 1 in the text box. Click on the **Add** button to add the transformation to the recode command.

7. Use the **Range** option to recode values 4 through 5 to 2 by selecting the radio button next to **Value** in the **New value** column and entering 2 in the text box. Click on the **Add** button to add the transformation to the recode command.

8. Use the **Range** option to recode values 6 through 8 to 3 by selecting the radio button next to **Value** in the **New value** column and entering 3 in the text box. Click on the **Add** button to add the transformation to the recode command.

9. Once you have defined all of the recodes, click on **Continue** to return to the main **Recode into Different Variables window.**

10. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

11. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.

12. Click on the **Reset button** to clear any previous work.

13. Add variable *incomegrpmerged* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

14. Add variable *incometercile* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

15. Ensure that the box beside **Display Frequency Tables** is checked**.**

16. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

17. Compare your results with ours

Statistics

		Merged income groups	Income terciles
N	Valid	3714	3714
	Missing	307	307

Frequency Table

Merged income groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	126	3.1	3.4	3.4
	2.00	469	11.7	12.6	16.0
	3.00	796	19.8	21.4	37.5
	4.00	699	17.4	18.8	56.3
	5.00	381	9.5	10.3	66.6
	6.00	561	14.0	15.1	81.7
	7.00	375	9.3	10.1	91.8
	8.00	306	7.6	8.2	100.0
	Total	3714	92.4	100.0	
Missing	System	307	7.6		
Total		4021	100.0		

Income terciles

|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1391	34.6	37.5	37.5
	2.00	1081	26.9	29.1	66.6
	3.00	1242	30.9	33.4	100.0
	Total	3714	92.4	100.0	
Missing	System	307	7.6		
Total		4021	100.0		

Assuming that your recode was successful, you now need to add value labels to the income tercile variable that you created. To do so, you can use the Syntax file, as we have described in the tip above, or do the following:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** next to the *incometercile* variable to add value labels.

3. In the **value labels window** that opens, enter 1 in the **Value:** textbox and "Low income" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

4. Enter 2 in the **Value:** textbox and "Middle income" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

5. Enter 3 in the **Value:** textbox and "High income" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

6. Once all of the values have been assigned labels, click the **OK button** to add the labels. It is advisable to confirm that the value labels have been assigned correctly by rerunning a frequency

distribution for this variable.

With our variables prepared for analysis, we can now use the steps below to generate a cross tabulation with the appropriate statistics to test the relationship:

1. Select **Analyze > Descriptive Statistics > Crosstabs…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add your dependent variable (*polinteresttercile*) to the **Row(s)** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add your independent variable (*incometercile*) to the **Column(s)** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Click on the **Statistics…** button to open the **Statistics** window.
6. Given that both variables are ordinal, we can assess the strength of the association by looking at the Gamma or the Tau value (see Chapter 12). Since gamma tends to inflate the strength of the relationship, we will opt for the more conservative Tau estimate. Since there are an equal number of categories for both variables, we choose Tau-b by checking the box beside it.
7. Recall from Chapter 13 that we can assess whether or not the relationship is statistically significant by looking at the Pearson Chi-square value. To request this statistic, check the box beside Chi-square in the **Crosstabs Statistics window** and click **Continue**.
8. Click on the **Cells…** button to open the **Cell Display** window.
9. Select **Column** under **Percentages** in order to compare values of the DV across categories of the IV (see Chapter 12). Note that you can request that SPSS report multiple percentages, but we advise against this to help simplify interpretation of the results. Click **Continue** to return to the main Crosstabs window.
10. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
11. Compare your results to ours, reported below.
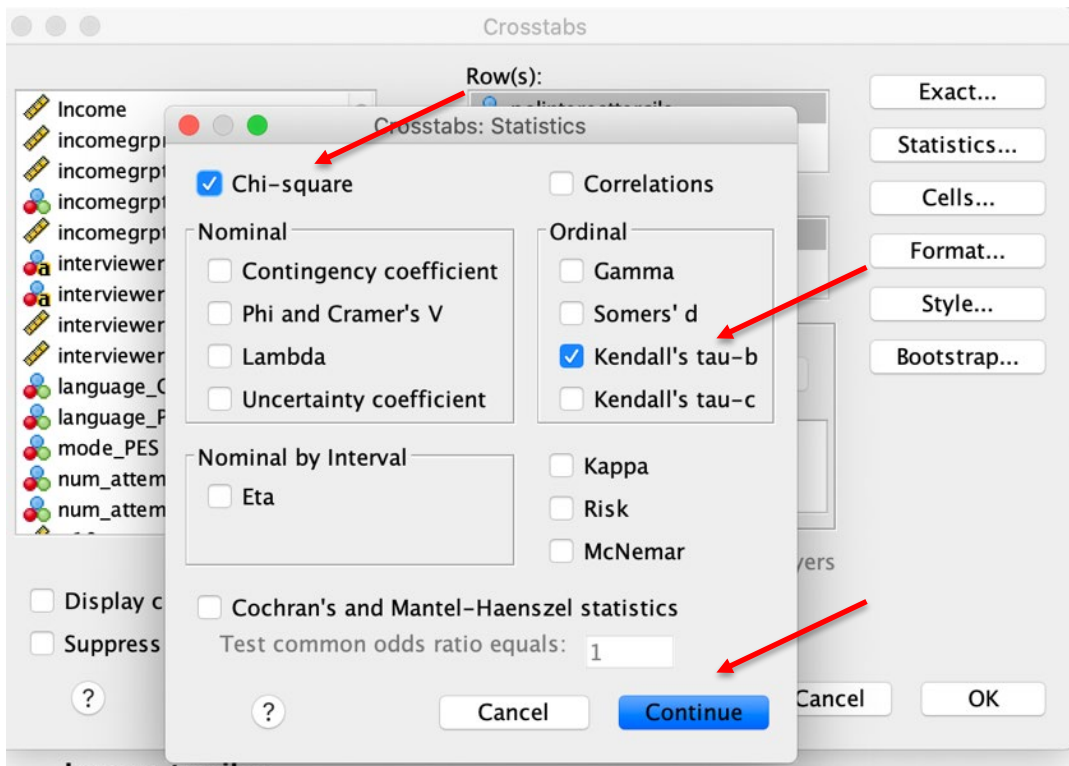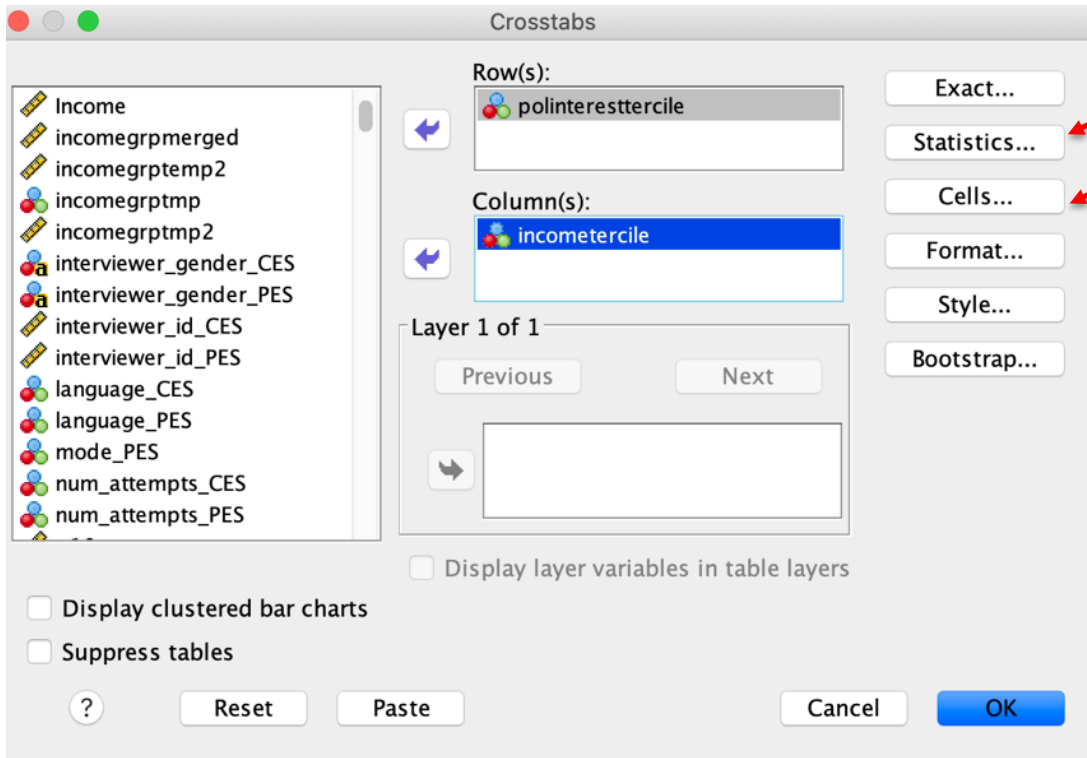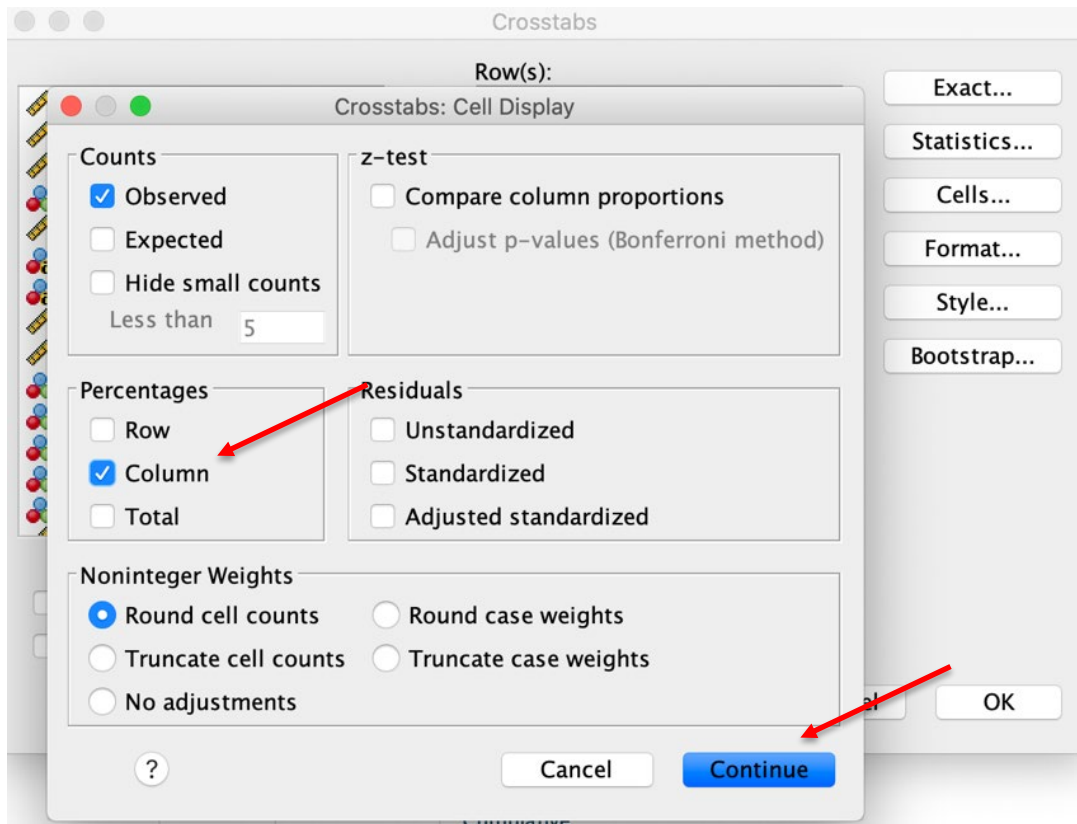
➜ Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Political interest terciles * Income terciles	3709	92.2%	312.000	7.8%	4021.000	100.0%

Political interest terciles * Income terciles Crosstabulation

			Income terciles			Total
			Low income	Middle income	High income	
Political interest terciles	Low interest	Count	482	264	281	1027
		% within Income terciles	34.8%	24.4%	22.6%	27.7%
	Middle interest	Count	428	391	432	1251
		% within Income terciles	30.9%	36.2%	34.8%	33.7%
	High interest	Count	477	425	529	1431
		% within Income terciles	34.4%	39.4%	42.6%	38.6%
Total		Count	1387	1080	1242	3709
		% within Income terciles	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	58.011[a]	4	.000
Likelihood Ratio	57.187	4	.000
Linear-by-Linear Association	42.305	1	.000
N of Valid Cases	3709		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 299.05.

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Ordinal by Ordinal	Kendall's tau-b	.094	.015	6.381	.000
N of Valid Cases		3709			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Is there a relationship between income and political interest? Recall from Chapter 12 that the first step when assessing the results from a contingency table with two ordinal level variables is to look for a consistent increase/decrease in the percentage of respondents across categories of the IV in the top row and the opposite pattern in the bottom row. In this example, reading across the top row ("Low interest"), we find that the percentages decrease as we move from left to right (low to high income): high income earners are approximately 12 percentage points less likely to indicate low political interest compared to their low income counterparts. Looking at the bottom row ("High interest") we find the reverse pattern, with high income earners approximately 8 percentage points more likely to indicate high political interest relative to those in the low income category.

As noted in Chapter 12, our next step is to consider the correct correlation coefficient. Given that both variables are ordinal, we can assess the strength of the association by looking at the Gamma or the Tau value. As indicated above, since gamma tends to inflate the strength of the relationship, we will opt for the more conservative Tau estimate. In this example, we find an extremely weak, positive (as income increases, political interest increases) relationship with a Tau-b value of 0.09.

Finally, recall from Chapter 13 that we can assess whether or not the relationship is statistically significant by looking at the Pearson Chi-square value. The results indicate that the relationship is statistically significant at $p<0.001$. As such, we would conclude that our results support our hypothesis: those with higher levels of income appear to be more interested in politics.

Check-In Point

This section covered an incredible amount of information – creating crosstabulations, how to calculate measures of association, and calculating Chi Square (not to mention some complicated variable recoding!). This is powerful, but it is always critical to keep in mind that your decisions with respect to recoding have great influence on the results, so always check your work carefully before assessing relationships.

Part IX: Examining Bivariate Relationships between Interval/Ratio Variables

Destination

By the end of this section, you will be able to:

- *create scatterplots;*
- *calculate Pearson's Correlation Coefficient; and*
- *conduct basic linear regression.*

To assess the relationship between two continuous (interval/ratio) variables, we continue to ask the same four questions noted in the last section (and, of course, in Chapter 12), but we use different statistical techniques.

For example, we might theorize that younger individuals are more apt to like the Green party. To test this, we first recode the variables for analysis. To do this, you can use the following steps:

1. In the **Variable View window** of the **Data Editor**, select the **Label** column and click on the **binoculars icon** from the options menu to open a search box and search for "green". This should bring you to variable *q18* a measure of feelings about the Green party.
2. Click on the three dots in the **Values** cell in the **Variable View window** of the **Data Editor** to view the value labels for *q18*. Note that the values range from 0 (Really dislike) to 100 (Really like) with -9, -8, -7, and -6 used to denote "Aware of party, but don't know", "Refused", "Skipped", and "Don't know party" responses respectively. You will begin by recoding this variable to remove these cases.
3. Select **Transform > Recode into Different Variables…** from the top menu bar.
4. Click on the **Reset button** to clear any previous work.
5. Add variable *q18* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *greenfeelings* and label it "Feelings about the Green Party".
7. Click on the button titled **Old and New Values** to recode the new variable.
8. Use the **Range** option to recode values -9 through -6 to system missing by selecting the radio button next to **System-missing** in the **New value** column. Click on the **Add** button to add the transformation to the recode command.

9. To copy the other values to the new variable without recoding them, select **All other values** from the **Old Value** column and then **Copy old values** from the **New value** column and click on the **Add** button to add the transformation to the recode command.
10. Click on **Continue** to return to the main **Recode into Different Variables window.**
11. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
12. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
13. Click on the **Reset button** to clear any previous work.
14. Add variable *q18* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
15. Add variable *greenfeelings* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
16. Ensure that the box beside **Display Frequency Tables** is checked**.**
17. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
20. Check your work, noting especially the difference in the number of valid cases and missing cases for the two variables as reported in the **Statistics box**.

Assuming that your recode was successful, you can add value labels to the lowest (really dislike) and highest (really like) values of your newly coded variable for reference. All other values do not require labels. To add the value labels, you can use the Syntax file, as we have described in the tip above, or do the following:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** next to the *greenfeelings* variable to add value labels.

3. In the **value labels window** that opens, enter 0 in the **Value:** textbox and "Really dislike" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

4. Enter 100 in the **Value:** textbox and " Really like" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

5. Click the **OK button** to add the labels. It is advisable to confirm that the value labels have been assigned correctly by rerunning a frequency distribution for this variable.

Next, we need to find the variable for the respondents' age. To do this, you can use the following steps:

1. In the **Variable View window** of the **Data Editor**, select the Label column and click on the **binoculars icon** from the options menu to open a search box and search for "age". This should bring you to variable *age* (after some searching). Note that there are no value labels assigned to this variable.
2. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
3. Click on the **Reset button** to clear any previous work.
4. Add variable *age* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Ensure that the box beside **Display Frequency Tables** is checked**.**

6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

7. Note that the age variable does not require recoding.

With our variables ready for analysis, we can produce a scatterplot to visually inspect whether or not there appears to be a linear relationship between age and feelings towards the Green party using the following steps to produce the scatterplot shown below:

1. Select **Graphs > Legacy Dialogs > Scatter/Dot…** from the top menu bar.
2. Select **Simple Scatter** from the window that opens and then click the **Define button** from the graph options window (note the other options available in this window).
3. Add variable *greenfelings* to the **Y Axis** by selecting it from the list of variables in the left side column and using the **transfer arrow icon** beside the **Y Axis:** textbox. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add variable *age* to the **X Axis** by selecting it from the list of variables in the left side column and using the **transfer arrow icon** beside the **X Axis:** textbox. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Click on the **Titles…** button to add a title, subtitle and / or footnote to your graph and then click on the **Continue** button.
6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
7. Compare your results with ours.
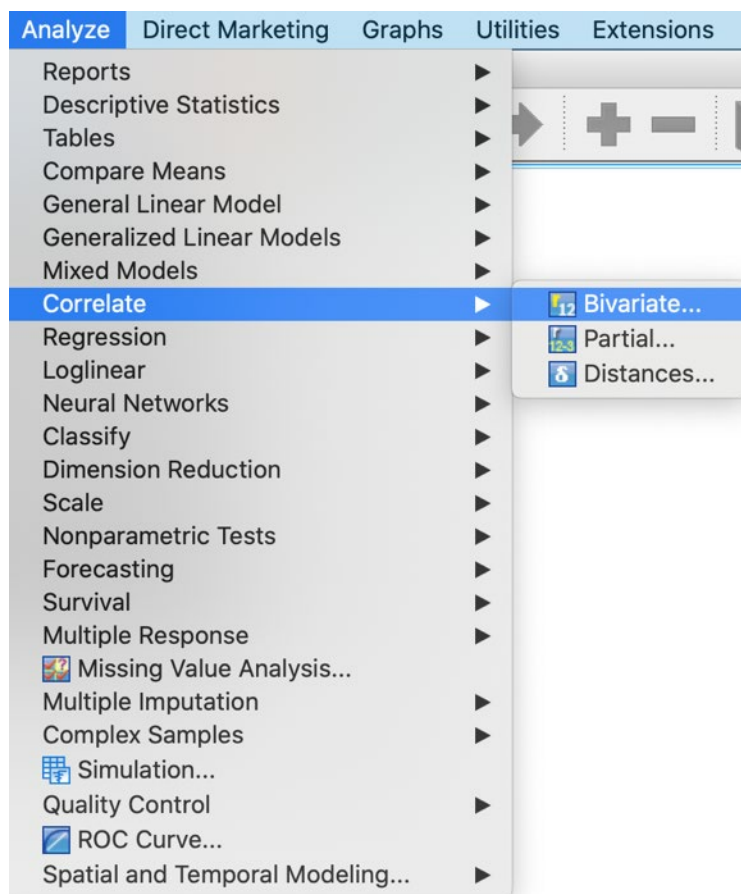
Feelings about the Green party according to Age
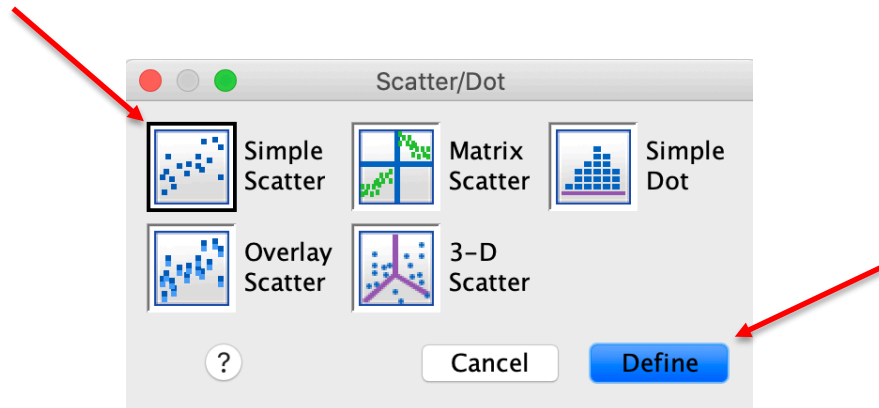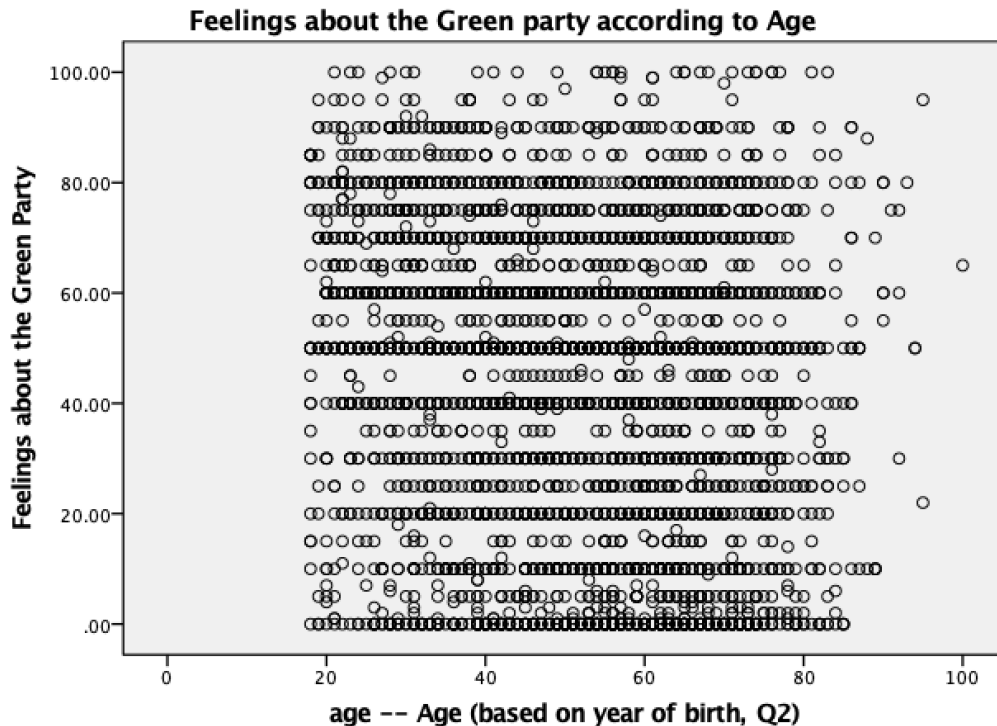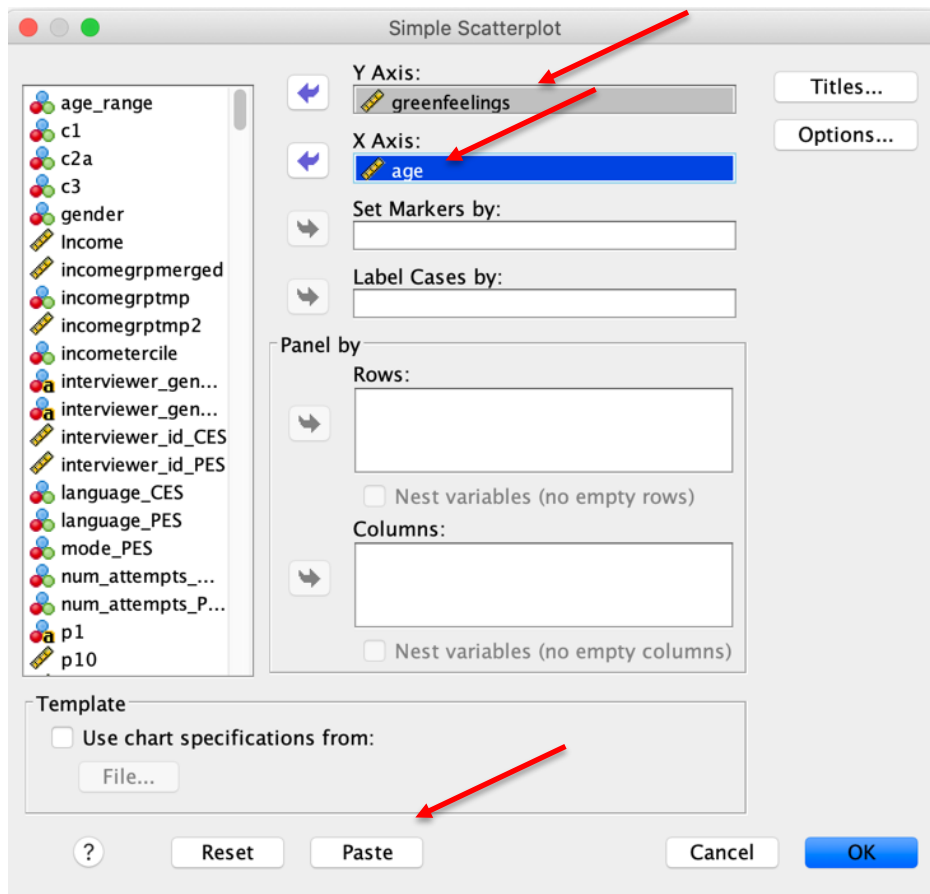
Now, we assess the scatter plot. What do you see? Don't panic – we don't see anything either. Based on the graph, it is difficult to interpret any type of relationship! It may be that age is *not* associated with feelings

about the Green party. To be sure, we need to look to further, either using Pearson's *r* or basic linear regression.

Pearson's *r*

We can find the measure of association between the two variables, Pearson's *r*. To estimate this value, we can do the following:

1. Select **Analyze > Correlate > Bivariate…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *age* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add variable *greenfeelings* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Ensure that the boxes beside **Pearson**, **Two-tailed** and **Flag significant correlations** are checked**.**
6. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system)..
7. Compare your results with ours.
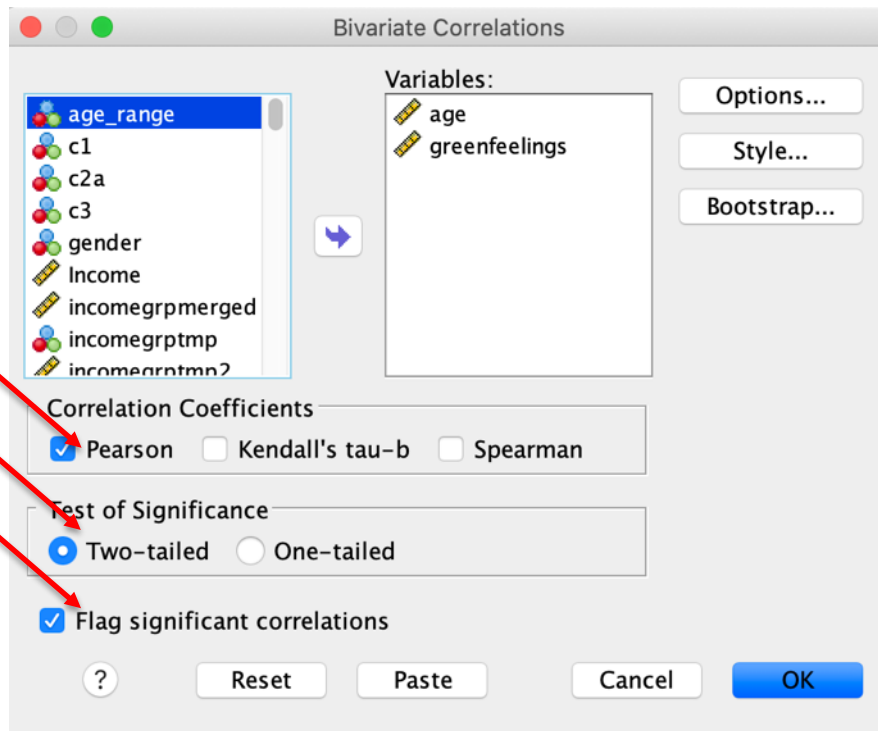
Correlations

		age -- Age (based on year of birth, Q2)	Feelings about the Green Party
age -- Age (based on year of birth, Q2)	Pearson Correlation	1	-.150**
	Sig. (2-tailed)		.000
	N	4021	3737
Feelings about the Green Party	Pearson Correlation	-.150**	1
	Sig. (2-tailed)	.000	
	N	3737	3737

**. Correlation is significant at the 0.01 level (2-tailed).

In interpreting the results, we will start with the correlation coefficient. The output includes the correlation between each variable with itself (a value of 1 since they are obviously perfectly correlated) and the measure of association between age and feelings about the Green party. The results indicate a weak, negative relationship between age and feelings about the Green party (-0.15): as age increases, feelings about the Green party decrease.

We next turn to the inferential statistic to see if this weak, negative relationship is statistically significant. The value below the correlation coefficient is the probability of observing a relationship of this strength in the sample if a similar relationship *did not* exist in the population from which the sample was drawn. In this case, the relationship is found to be statistically significant at $p<0.001$ (Note that the significance level is displayed in the "Sig. (2 -tailed)" row of the table, ".000" in this example, reported as $p<0.001$ in the text).

Basic linear regression

While the measure of association and the strength of the relationship between two variables is informative, we can also use information about the independent variable to predict scores on the dependent variable using basic linear regression. SPSS allows us to easily produce regression models. To continue with our example, we can estimate how feelings for the Green party changes for every year increase in age with the following steps:

1. Select **Analyze > Regression > Linear…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add variable *greenfeelings* as the **Dependent** variable by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add variable *age* to the **Block 1 of 1** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
6. Check your results against ours.
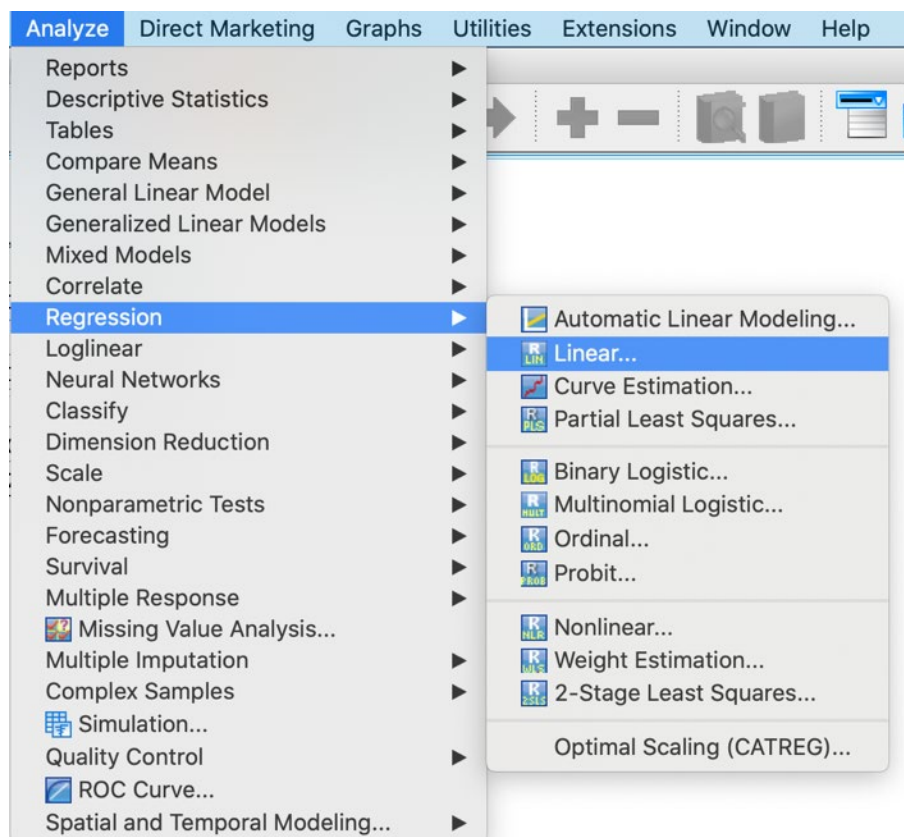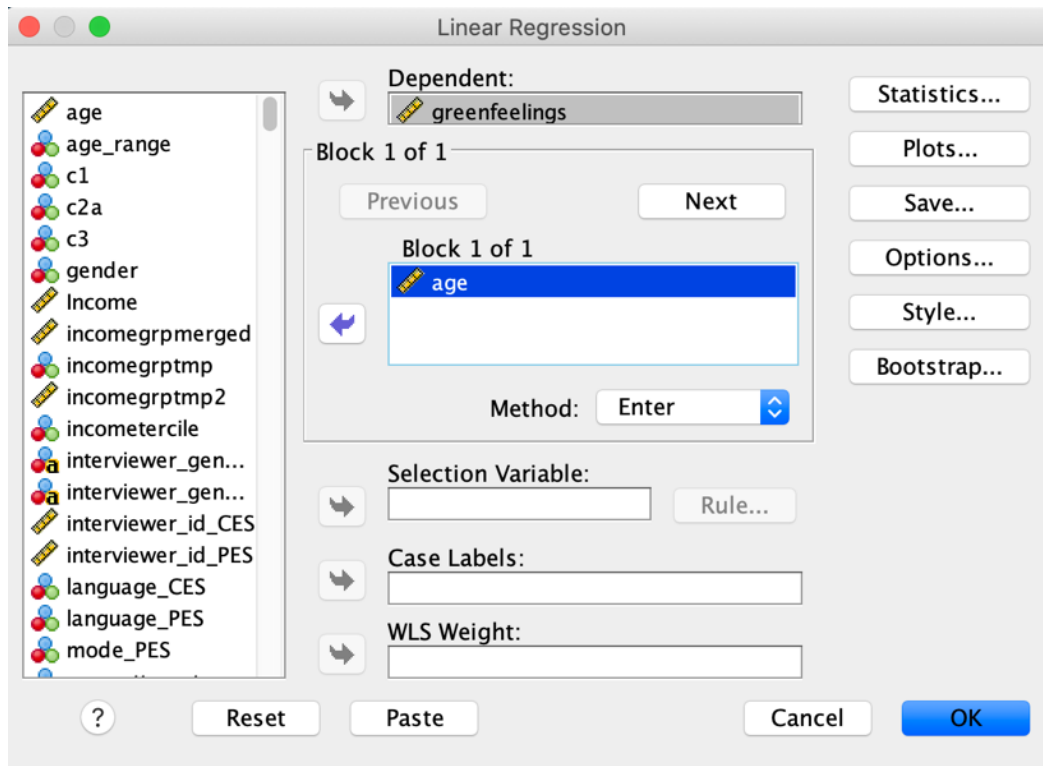
Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	age -- Age (based on year of birth, Q2)[b]	.	Enter

a. Dependent Variable: Feelings about the Green Party

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.150[a]	.023	.022	26.78678

a. Predictors: (Constant), age -- Age (based on year of birth, Q2)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	62003.529	1	62003.529	86.412	.000[b]
	Residual	2679792.65	3735	717.531		
	Total	2741796.18	3736			

a. Dependent Variable: Feelings about the Green Party

b. Predictors: (Constant), age -- Age (based on year of birth, Q2)

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	55.203	1.417		38.964	.000
	age -- Age (based on year of birth, Q2)	-.243	.026	-.150	-9.296	.000

a. Dependent Variable: Feelings about the Green Party

There is considerably more information presented here than in the Pearson's *r* results. Let's walk through some of it:

- Age coefficient. The age coefficient is -0.24. This indicates that for every year increase in age, feelings about the Green party decrease by 0.24 points. We also find that there is a statistically significant relationship between age and feelings about the Green party based on the value reported under the Sig. column (0.000). We would report this a $p < 0.001$ in our written interpretation of the results.
- Intercept: The constant (intercept), 55.20, is the value on the feelings about the Green party variable when age is equal to 0 (an impossible value given that respondents for the CES are a minimum of 18 years of age).
- R Square: How much of the variance in the dependent variable does our independent variable explain? Not much. Recall from Chapter 12 that we can determine the proportion of the dependent variable that can be explained by the independent variable by squaring the

Pearson's *r* value, producing a result known as r^2. This value is reported with the SPSS output as "R square", 0.02 in this example. In other words, using age as a predictor of feelings about the Green party reduces our prediction errors by two percent.

Check-In Point

If you are continuing to follow along by running all of the examples in your own dataset, you now have the ability to create a scatter plot, calculate Pearson's *r*, and conduct basic linear regression with SPSS. You have come a long way!

Part X: Assessing Relationships Using Control Variables

Destination

By the end of this section, you will be able to:

- *add control variables to your analyses.*

The final question when assessing a relationship between two variables is to consider what happens to the relationship once other important variables are controlled. We discuss this question fully in Chapter 13, and in this section of the handbook we direct you to the appropriate SPSS syntax.

Cross tabs and control variables

Recall from Chapter 13 that to test a control variable using a crosstabulation, you assess the IV-DV relationship separately for each category of the control and compare these results against those obtained in the original (full) model.

Let's consider the relationship between income and political interest (recall that we recoded these variables previously), controlling for education (university graduate versus non-university graduate). To do so, we instruct SPSS to create crosstabulations for each category of the dichotomous education variable using the following steps:

1. In the **Variable View window** of the **Data Editor**, select the **Label** column and click on the **binoculars icon** from the options menu to open a search box and search for "education". This should bring you to variable *q61* a measure of highest level of education completed.
2. Click on the three dots in the **Values** cell in the **Variable View window** of the **Data Editor** to view the value labels for *q61*. Note that the values range from 1 (No Schooling) to 11 (Professional degree or doctorate) with -9, -8, and -7 used to denote Don't know, Refused, and Skipped responses respectively. You will begin by recoding this variable to remove these cases and create two groups; those without a university degree (values 1 – 8) and those who have completed university (values 9-11).
3. Select **Transform > Recode into Different Variables…** from the top menu bar.
4. Click on the **Reset button** to clear any previous work.
5. Add variable *q61* to the **Numeric Variable > Output Variable** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Add a **Name** and **Label** for the new variable in the textbox below the **Output Variable** heading and click on the **Change** button to apply. We will name our new variable *universitygrad* and label it

"University graduate versus non-university graduate".

7. Click on the button titled **Old and New Values** to recode the new variable.
8. Use the **Range** option to recode values -9 through -7 to system missing by selecting the radio button next to **System-missing** in the **New value** column. Click on the **Add** button to add the transformation to the recode command.
9. Use the **Range** option to recode values 1 through 8 to 0 by selecting the radio button next to **Value** in the **New value** column and entering 0 in the text box. Click on the **Add** button to add the transformation to the recode command.
10. Use the **Range** option to recode values 9 through 11 to 1 by selecting the radio button next to **Value** in the **New value** column and entering 1 in the text box. Click on the **Add** button to add the transformation to the recode command.
11. Click on **Continue** to return to the main **Recode into Different Variables window.**
12. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon in** the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
13. Select **Analyze > Descriptive Statistics > Frequencies…** from the top menu bar.
14. Click on the **Reset button** to clear any previous work.
15. Add variable *q61* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
16. Add variable *universitygrad* to the **Variable(s)** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
17. Ensure that the box beside **Display Frequency Tables** is checked**.**
18. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
19. Check your results for *universitygrad* against ours. Note that the number of missing cases reported for the new variable (n=11) in the **Statistics** results dialogue box will not match the sum of respondents who refused to answer or did not know their highest level if education (n=12) due to weighting.

University graduate versus non-university graduate

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	2205	54.8	55.0	55.0
	1.00	1805	44.9	45.0	100.0
	Total	4010	99.7	100.0	
Missing	System	11	.3		
Total		4021	100.0		

Assuming that your recode was successful, you can add value labels to your newly coded variable. To add the value labels, you can use the Syntax file, as we have described in the tip above, or do the following:

1. Open the **Variable View window** of the **Data Editor.**

2. Click on the **…** in the **Values column** next to the *universitygrad* variable to add value labels.

3. In the **value labels window** that opens, enter 0 in the **Value:** textbox and "Non-university grad" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

4. Enter 1 in the **Value:** textbox and " University grad" to the **Label:** textbox and then click the **Add** button to add the value label to the list of labels.

5. Click the **OK button** to add the labels. It is advisable to confirm that the value labels have been assigned correctly by rerunning a frequency distribution for this variable.

With the control variable recoded, we can now use the following steps to generate a cross tabulation with a control variable including the appropriate statistics to test the relationship:

1. Select **Analyze > Descriptive Statistics > Crosstabs…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.
3. Add your dependent variable (*polinteresttercile*) to the **Row(s)** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
4. Add your independent variable (*incometercile*) to the **Column(s)** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
5. Add your control variable (*universitygrad*) to the **Layer 1 of 1** dialogue box by selecting it from the list of variables in the left side column and using the **transfer arrow icon** between the two columns to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).
6. Click on the **Statistics…** button to open the **Statistics** window.
7. Given that both variables are ordinal, we can assess the strength of the association by looking at the Gamma or the Tau value (see Chapter 12). Since gamma tends to inflate the strength of the relationship, we will opt for the more conservative Tau estimate. Since there are an equal number of categories for both variable, we choose Tau-b by checking the box beside it.
8. Recall from Chapter 13 that we can assess whether or not the relationship is statistically significant by looking at the Pearson Chi-square value. To request this statistic, check the box beside Chi-square in the **Crosstabs Statistics window** and click **Continue**.
9. Click on the **Cells…** button to open the **Cell Display** window.
10. Select **Column** under **Percentages** in order to compare values of the DV across categories of the IV (see Chapter 12). Note that you can request that SPSS report multiple percentages, but we advise against this to help simplify interpretation of the results. Click **Continue** to return to the main Crosstabs window.
11. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon in** the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).
12. Compare your results to ours, reported below.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Political interest terciles * Income terciles * University graduate versus non–university graduate	3701	92.0%	320.000	8.0%	4021.000	100.0%

Political interest terciles * Income terciles * University graduate versus non–university graduate Crosstabulation

University graduate versus non–university graduate				Income terciles			
				Low income	Middle income	High income	Total
Non–university grad	Political interest terciles	Low interest	Count	358	157	131	646
			% within Income terciles	37.0%	28.6%	25.9%	32.0%
		Middle interest	Count	281	181	164	626
			% within Income terciles	29.1%	33.0%	32.5%	31.0%
		High interest	Count	328	210	210	748
			% within Income terciles	33.9%	38.3%	41.6%	37.0%
	Total		Count	967	548	505	2020
			% within Income terciles	100.0%	100.0%	100.0%	100.0%
University grad	Political interest terciles	Low interest	Count	119	107	150	376
			% within Income terciles	28.7%	20.2%	20.4%	22.4%
		Middle interest	Count	147	208	268	623
			% within Income terciles	35.4%	39.2%	36.4%	37.1%
		High interest	Count	149	215	318	682
			% within Income terciles	35.9%	40.6%	43.2%	40.6%
	Total		Count	415	530	736	1681
			% within Income terciles	100.0%	100.0%	100.0%	100.0%
Total	Political interest terciles	Low interest	Count	477	264	281	1022
			% within Income terciles	34.5%	24.5%	22.6%	27.6%
		Middle interest	Count	428	389	432	1249
			% within Income terciles	31.0%	36.1%	34.8%	33.7%
		High interest	Count	477	425	528	1430
			% within Income terciles	34.5%	39.4%	42.5%	38.6%
	Total		Count	1382	1078	1241	3701
			% within Income terciles	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

University graduate versus non-university graduate		Value	df	Asymptotic Significance (2-sided)
Non-university grad	Pearson Chi-Square	23.207[b]	4	.000
	Likelihood Ratio	23.268	4	.000
	Linear-by-Linear Association	18.631	1	.000
	N of Valid Cases	2020		
University grad	Pearson Chi-Square	14.360[c]	4	.006
	Likelihood Ratio	13.885	4	.008
	Linear-by-Linear Association	9.738	1	.002
	N of Valid Cases	1681		
Total	Pearson Chi-Square	55.231[a]	4	.000
	Likelihood Ratio	54.465	4	.000
	Linear-by-Linear Association	40.478	1	.000
	N of Valid Cases	3701		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 297.68.

b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 156.50.

c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 92.83.

Symmetric Measures

University graduate versus non-university graduate			Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Non-university grad	Ordinal by Ordinal	Kendall's tau-b	.086	.020	4.364	.000
	N of Valid Cases		2020			
University grad	Ordinal by Ordinal	Kendall's tau-b	.065	.022	2.917	.004
	N of Valid Cases		1681			
Total	Ordinal by Ordinal	Kendall's tau-b	.092	.015	6.240	.000
	N of Valid Cases		3701			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Note that the results are reported for non-graduates, university graduates and the total sample, respectively. The interpretation of the relationship with the inclusion of a control variable is the same as the process you followed to interpret the original relationship, only you need to do so for each category of the control variable. You then compare the results from each of the control variable categories to that of the original relationship to assess whether or not the control variable affects the relationship as anticipated (see Chapter 13 for a full interpretation of the results with the addition of the control variable). Recall that we observed an extremely weak, statistically significant relationship (Tau = 0.09; p<0.001) in the original model, with levels of political interest increasing with income. When we control for education, we find that the relationship is essentially replicated for non-university graduates and weakens when we assess the university graduate group (Tau = 0.06; p<0.01). Accordingly, education does not appear to be a source of spuriousness.

Multivariate linear regression

The procedure for multivariate linear regression in SPSS is the same as that for basic linear regression, only we add the additional independent and/or control variables to the model. For example, in addition to age, we can assess how education and income influence feelings about the Green party with the following:

1. Select **Analyze > Regression > Linear…** from the top menu bar.
2. Click on the **Reset button** to clear any previous work.

3. Add variable *greenfeelings* as the **Dependent** variable by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

4. Add variable *age* to the **Block 1 of 1** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

5. Add variable *income* to the **Block 1 of 1** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

6. Add variable *universitygrad* to the **Block 1 of 1** list by selecting it from the list of variables in the left side column and using the **transfer arrow icon** to add it. (Note that the variable names should be in alphabetical order if you followed our recommendation above).

7. Click **Paste** to add the command to your syntax file and then run the command from the syntax file (Reminder: To run a command from the syntax file, select the line and then click on the **Play button icon** in the top center menu bar of the **Syntax window** or use the short-cut keys: Command -R on a Mac or CTRL – R with a Windows operating system).

8. Check your results against ours.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Income in dollars, age -- Age (based on year of birth, Q2), University graduate versus non–university graduate[b]	.	Enter

a. Dependent Variable: Feelings about the Green Party

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.256[a]	.065	.064	25.88786

a. Predictors: (Constant), Income in dollars, age -- Age (based on year of birth, Q2), University graduate versus non–university graduate

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	134236.104	3	44745.368	66.766	.000[b]
	Residual	1919071.63	2864	670.181		
	Total	2053307.73	2867			

a. Dependent Variable: Feelings about the Green Party

b. Predictors: (Constant), Income in dollars, age -- Age (based on year of birth, Q2), University graduate versus non–university graduate

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	54.293	1.753		30.967	.000
	age -- Age (based on year of birth, Q2)	-.260	.030	-.156	-8.621	.000
	University graduate versus non–university graduate	10.078	.982	.188	10.264	.000
	Income in dollars	-2.376E-5	.000	-.109	-5.940	.000

a. Dependent Variable: Feelings about the Green Party

The results show that, holding education and income equal, for every year increase in age, feelings about the Green party decrease by 0.26 points (p<0.001). In the case of education, given that we are using a dichotomous variable, we would interpret the results as indicating that university graduates are more likely (10 points) to have more positive feelings about the Green party than those who have not completed university, net of age and income. While the coefficient for income is statistically

significant, the impact on feelings about the Greens is marginal, a decrease of 0.00002 points for every unit increase in income, holding age and education constant. How much of the variance in the dependent variable does our model explain? We can use the value reported as the adjusted R-Square, which takes into account the number of variables in the model, to determine the proportion of the dependent variable that can be explained by the independent variables. In this example, our model reduces our prediction errors by 6 percent.

Check-In Point and Conclusion

As we come to the end of this introductory SPSS handbook, we hope that the procedures outlined here have provided you with the basic skills necessary to conduct your own statistical analyses. We also hope that this introduction has encouraged you to learn more about the many possibilities to use this type of statistical program for your research. As we have noted, this is only a very small sampling of the many options available in SPSS, which we hope serves as the starting point for your continued exploration of the possibilities that this software has to offer.