Revision of IBM SPSS Statistics

IBM SPSS Statistics™ (SPSS from now on) is a computer package design to cater for your every statistical need. It carries out analyses that take hours by hand in a matter of seconds. Here's a list of what SPSS can and cannot do:

SPSS will:

- Do complex statistical analysis in a few seconds.
- Produce graphs.
- Save you time.
- Do anything statistical that *you ask it to do*.

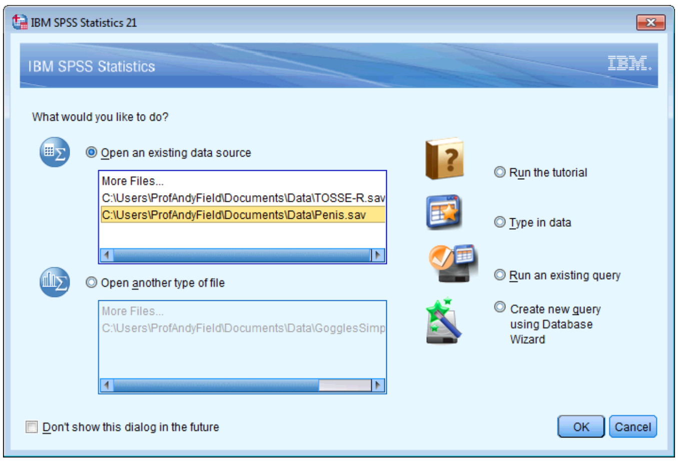
SPSS won't:

- Decide what test needs to be done in a given circumstance.
- Pass your exam for you.
- Make tea.

So, although SPSS is a very powerful tool, it works only within the limits of your own statistical knowledge. In many respects it is rather stupid because if you provide it with data and ask it to run a completely meaningless analysis, it will happily oblige. Therefore, you still need to use your brain to initiate the correct analysis.

Getting started with SPSS

There are several excellent texts that give introductions to the general environment within which SPSS operates. In my entirely unbiased opinion, the best is Field (2013)! SPSS mainly uses two windows: the data editor (this is where you input your data and carry out statistical functions) and the *viewer* (this is where the results of any analysis appear).



Once SPSS has been activated, a start-up window will appear, which allows you to select various options. If you already have a data file on disk that you would like to open then select *Open an existing data source* by clicking on the ○ so that it looks like ◉: this is the default option. In the space underneath this option there will be a list of recently used data files that you can select with the mouse. If you want to open a data file that isn't in the list then simply select *More Files…* with the mouse and click on OK.

The main SPSS window includes a data editor for entering data. This window is where most of the action happens. The data editor has two views: the data view and the variable view. The data view is for entering data into the data editor, and the variable view allows us to define various characteristics of the variables within the data editor. At the bottom of the data editor, you should notice that there are two tabs labelled 'Data View' and 'Variable View' (Data View Variable View) and all we do to switch between these two views is click on these tabs (the highlighted tab tells you which view you're in, although it will be obvious).

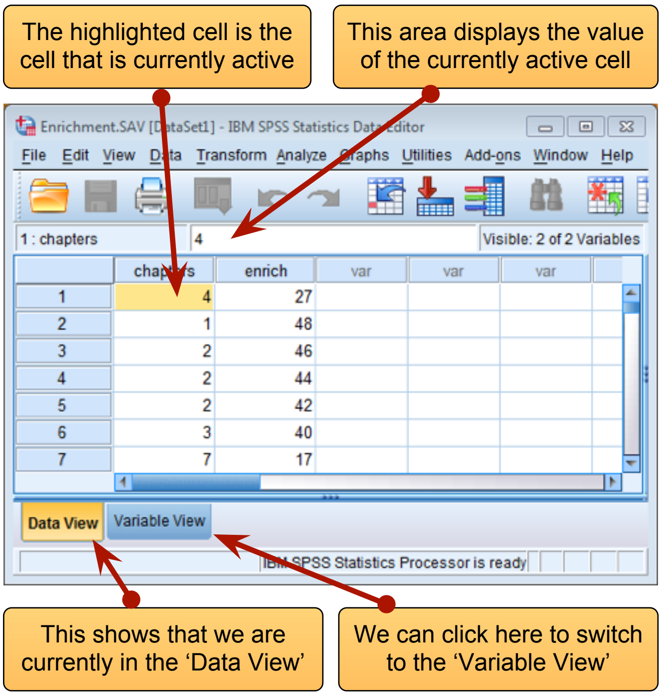


Figure 1: The SPSS Data Editor ('Data View')

Entering Data Using the Data Editor

Overview

When you first load SPSS it will provide a blank data editor (showing the 'data view') with the title *New Data*. When inputting a new set of data, you must input your data in a logical way. The SPSS data editor is arranged such that *each row represents data from one participant while each column represents a variable*. There is no discrimination between independent and dependent variables, both types should be placed in a separate column. It follows from this arrangement that any variable measured with the same participants (a repeated measure) should be represented by several columns (each column representing one level of the repeated measures variable). However, when a between-group design was used (e.g. different participants were assigned to each level of the independent variable) the data will be represented by two columns: one that has the values of the dependent variable and one that is a coding variable indicating to which group the participant belonged. This idea will become clearer as you learn about how to carry out specific procedures.

YouTube http://www.youtube.com/watch?v=b163iBByycw

The data editor is made up of lots of *cells*, which are just boxes in which data values can be placed. When a cell is active it becomes highlighted with a black surrounding box (as in Figures 1 and 2). You can move around the data editor, from cell to cell, using the arrow keys ← ↑ ↓ → (found on the right of the keyboard) or by clicking the mouse on the cell that you wish to activate. To enter a number into the data editor simply move to the cell in which you want to place the data value, type in the value, then press the arrow button appropriate to the direction in which you wish to move. So, to enter a row of data, move to the far left of the row, type the first value and then press → (this inputs the value and then moves you into the next cell on the right).

The first step in entering your data is to create some variables using the 'variable view' of the data editor, and then to input your data using the 'data view' of the data editor. We'll go through these two steps by working through an example.

Creating a Variable: The 'Variable view'

Before we input any data into the data editor, we need to create the variables. To create variables we use the 'Variable View' of the data editor. To access this view click on the 'Variable View' tab at the bottom of the data editor (Data View Variable View); the contents of the window will change (Figure 2).

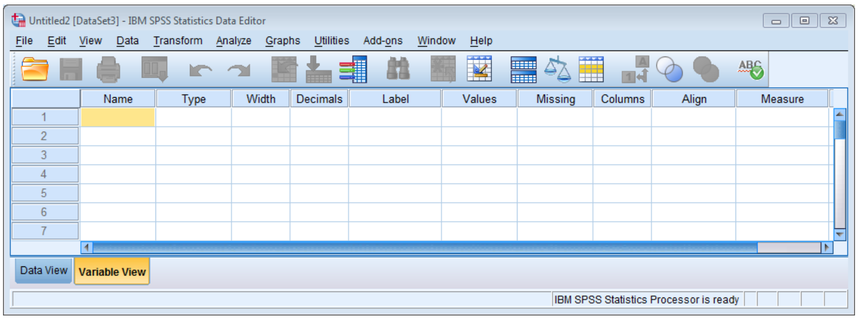


Figure 2: The 'Variable View' of the SPSS Data Editor

Every row of the variable view represents a variable, and you set characteristics of a particular variable by entering information into the labelled columns. You can change various characteristics of the variable by entering information into the following columns (play around and you'll get the hang of it):

Name You can enter a name in this column for each variable. This name will appear at the top of the corresponding column in the data view, and helps you to identify variables in the data view. There are certain symbols you can't use (mainly symbols that have other uses in SPSS such as +, −, $, &), and you can't use spaces. (Many people use a 'hard' space in variable names, which replaces the space with an underscore; for example, Andy_Field instead of Andy Field.) If you use a character that SPSS doesn't like you'll get an error message saying that the variable name is invalid when you click on a different cell, or try to move off the cell using the arrow keys.

Type You can have different types of data. Mostly you will use numeric variables (which just means that the variable contains numbers - SPSS assumes this data type). You will come across string variables, which consist of strings of letters. If you wanted to type in people's names, for example, you would need to change the variable type to be string rather than numeric. You can also have currency variables (i.e. £s, $s, euro) and date variables.

Label The name of the variable (see above) has some restrictions on characters, and you also wouldn't want to use huge long names at the top of your columns (they become hard to read). Therefore, you can write a longer variable description in this column. This may seem pointless, but is actually one of the best habits you can get into.

Values This column is for assigning numbers to represent groups of people (see below).

Measure This is where you define the level at which a variable was measured (*Nominal*, *Ordinal* or *Scale*). This is important, especially if you want to draw graphs!

Data set 1: Between-group design

To begin with, imagine that a film company director was interested in whether there was really such a thing as a 'chick flick' (a film that typically appeals to women more than men). He took 20 men and 20 women and showed half of each sample a film that was supposed to be a 'chick flick' (*Bridget Jones' Diary*), and the other half of each sample a film that didn't fall into the category of 'chick flick' (*Memento*, a brilliant film by the way). In all cases he measured their physiological arousal as an indicator of how much they enjoyed the film.

Table 1: Film data

Male		Female	
Bridget Jones' Diary	Memento	Bridget Jones' Diary	Memento
22	37	3	30
13	20	15	25
16	16	5	31
10	28	16	36
18	27	13	23
24	18	20	14
13	32	11	21
14	24	19	31
19	21	15	22
23	35	7	14

We need to pay attention to the rule of the data editor that *scores from different people appear in different rows* because this is not how the data are laid out in the table above (to save space). Remember that each score above comes from a different person, therefore we will need 40 rows (not 10 as above). To enter the data above into the SPSS data editor we need to create several variables. If we begin with the variable **Arousal**, we should follow these steps:

1. Move the on-screen arrow (using the mouse) to the first white cell in the column labelled *Name*.
2. Type the word *Arousal*.
3. Move off of this cell using the arrow keys on the keyboard (you can also just click on a different cell, but this is actually a very slow way of doing it).

You've just created your first variable! You should notice that once you've typed a name, SPSS creates default settings for the variable (such as assuming it's numeric and assigning 2 decimal places).

Now because I want you to get into good habits, move to the cell in the column labelled *label* and type 'Average physiological arousal during the film'.

Once the variable has been created, you can return to the data view by clicking on the 'data view' tab at the bottom of the data editor (Data View Variable View). The contents of the window will change, and you'll notice that the first column now has the label *arousal*. To enter the data, click on the white cell at the top of the column labelled *arousal* and type the first value (22). To register this value in this cell, we have to move to a different cell and because we are entering data down a column, the most sensible way to do this is to press the ↓ key on the keyboard. This action moves you down to the next cell, and the number 22.00 should appear in the cell above. Enter the next number (13) and then press ↓ to move down to the next cell, and so on.

Creating Coding Variables

A coding variables (also known as a grouping variable) is a type of variable that you will use on numerous occasions: it is a variable consisting of a series of numbers that represent levels of an independent variable. In experiments, coding variables are used to represent independent variables that have been measured *between-groups* (i.e. different participants were assigned to different groups). So, if you were to run an experiment with one group of participants in an experimental condition and a different group of participants in a control group, you might assign the experimental group a code of 1, and the control group a code of 0. When you come to put the data into the data editor, then you would create a variable (which you might call **group**) and type in the value 1 for any participants in the experimental group, and a 0 for any participant in the control group. This tells the computer that all of the cases that have been assigned the value 1 should be treated as belonging to the same group, and likewise for the participants assigned the value 0.

There is a simple rule for how variables should be placed in the SPSS data editor: levels of the between-group variables go down the data editor whereas levels of within-subject (repeated measures) variables go across the data editor. We will practice this rule over the coming weeks.

We have two coding variable in our data: the one describing whether a person was male or female, and a second describing whether the person watched 'Bridget Jones' diary' or 'Memento'. Let's create the variable **gender** first. To create this coding variable, we follow the steps for creating a normal variable, but we also have to tell the computer which numeric codes have been assigned to which groups. So, first of all, move to the cell in the second row in the column labelled *Name* type a name (let's call it **Gender**). I'm still trying to instil good habits, so move along the third row to the column called *Label* and give the variable a full description such as 'Participant's gender' Then to define the group codes, move along the row to the column labelled Values and into this cell: None .... Click on ... to access the *Value Labels* dialog box (see Figure 3).

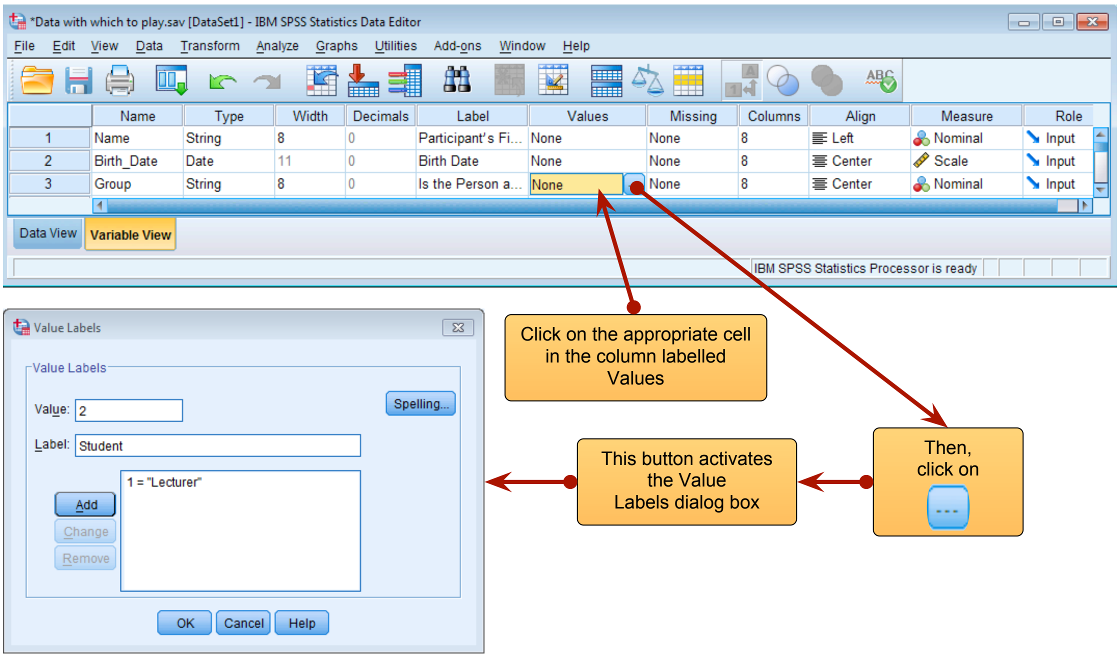


Figure 3: Defining coding values in SPSS

The *Value Labels* dialog box is used to specify group codes. This can be done in three easy steps. First, click with the mouse in the white space next to where it says *Value* (or press *Alt* and *u* at the same time) and type in a code (e.g. 1). These codes are completely arbitrary; for the sake of convention people typically use 0, 1, 2, 3, etc, but in practice you could have a code of 495 if you were feeling particularly arbitrary. The second step is to click the mouse in the white space below, next to where it says *Value Label* (or press *Tab*, or *Alt* and *e* at the same time) and type in an appropriate label for that group. In Figure 3 I have already defined a code of 1 for Males, and then I have typed in 2 as my code and given this a label of *Female.* The third step is to add this coding to the list by clicking on Add. When you have defined all of your coding values you can click on Spelling... and SPSS will check your variable labels for spelling errors (which can be very handy if you are as bad at spelling as I am). To finish, click on OK; if you click on OK and have forgotten to add your final coding to the list, SPSS will display a message warning you that any pending changes will be lost. In plain English this simply tells you to go back and click on Add before continuing. Finally, coding variables always represent categories and so the level at which they are measured is nominal (or ordinal if the categories have a meaningful order). Therefore, you should specify the level at which the variable was measured by going to the column labelled *Measure* and selecting Nominal (or Ordinal if the groups have a meaningful order) from the drop-down list.

Having defined your codes, you can then go to the data view (by clicking the tab at the bottom of the window) and type these numerical values into the appropriate column (so if a person was a male type a value of 1, if they were a female type the value 2). You can get the computer to display the numerical codes, or the value labels that you gave them by clicking on .

Having created these variables **arousal** and **gender**, try to create the variable **film** and enter the rest of the data yourself. Save these data in a file called **ChickFlick.sav** – we'll need them later.

Data set 2: repeated measures

Hiccups can be a serious problem: Charles Osborne apparently got a case of hiccups while slaughtering a hog (well, who wouldn't?) that lasted 67 years. People have many methods for stopping hiccups (a surprise, holding your breath), but actually medical science has put its collective mind to the task too. The official treatment methods include tongue-pulling manoeuvres, massage of the carotid artery, and, believe it or not, digital rectal massage (Fesmire, 1988). Let's say we wanted to put digital rectal massage to the test (as a cure of hiccups I mean). We took 15 hiccup sufferers, and during a bout of hiccups administered each of the three procedures (in random order and at intervals of 5 minutes) after taking a baseline of how many hiccups they had per minute. We counted the number of hiccups in the minute after each procedure.

Table 2: Hiccups data

Baseline	Tongue Pulling	Carotid Artery Massage	Digital Rectal Massage
15	9	7	2
13	18	7	4
9	17	5	4
7	15	10	5
11	18	7	4
14	8	10	3
20	3	7	3
9	16	12	3
17	10	9	4
19	10	8	4
3	14	11	4
13	22	6	4
20	4	13	4
14	16	11	2
13	12	8	3

These data can be entered as above because each person took part in all four conditions – therefore, each row represents a particular person, and each column represents the number of hiccups that they had during a particular treatment.

→ *You know how to create variables and enter data, so create four new variables representing the treatments in this experiment. Label them, set the properties, and then enter the scores for the 15 participants. Save the completed data file as* ***Hiccups.sav****.*

The SPSS Viewer

Alongside the SPSS Data Editor window, there is a second window known as the *SPSS Viewer*. Figure 4 shows the basic layout of the SPSS viewer. On the right-hand side there is a large space in which the output is displayed. SPSS displays both graphs and the results of statistical analyses in this part of the viewer. It is also possible to edit graphs and to do this you simply double-click on the graph you wish to edit (this creates a new window in which the graph can be edited). On the left-hand side of the output viewer there is a tree diagram illustrating the structure of the output. This tree diagram is useful when you have conducted several analyses because it provides an easy way of accessing specific parts of the output. The tree structure is fairly self-explanatory in that every time you do something in SPSS (such as drawing a graph or running a statistical procedure), it lists this procedure as a main heading.

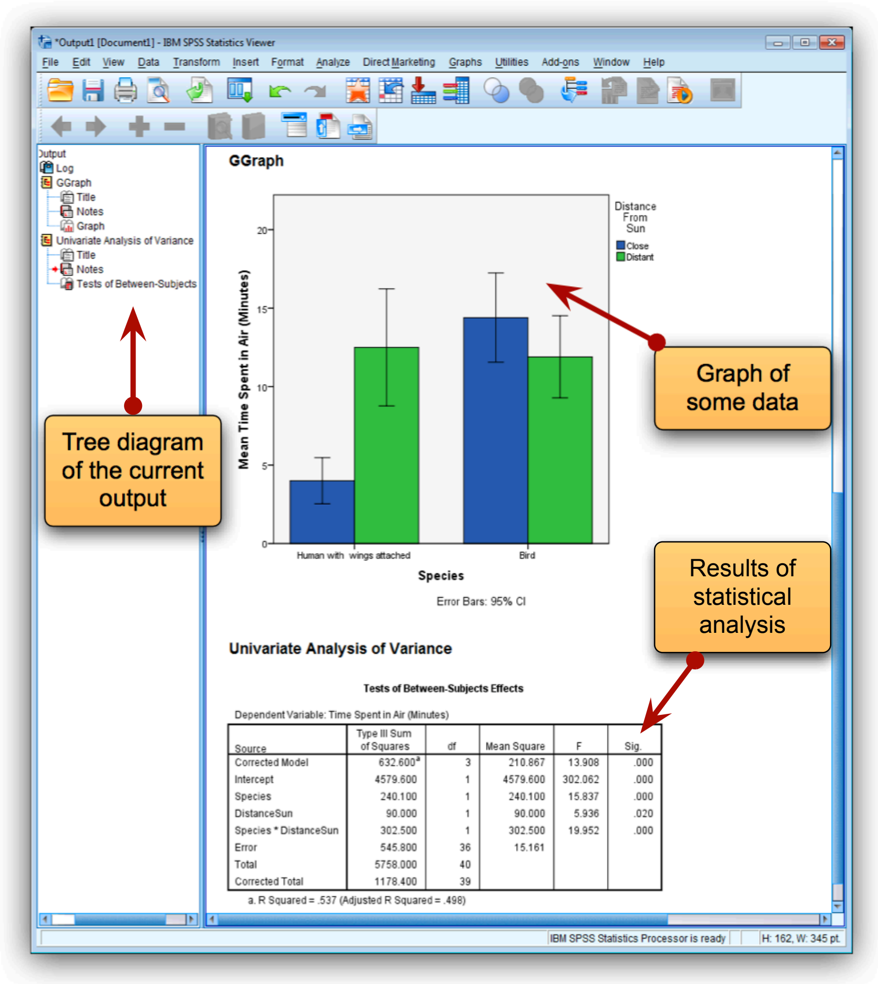


Figure 4: The Output Viewer

YouTube http://www.youtube.com/watch?v=8dDIw_oJsBs

In Figure 4 I conducted a graphing procedure followed by a univariate analysis of variance (ANOVA) and so these names appear as main headings. For each procedure there are a series of sub-headings that represents different parts of the analysis. For example, in the ANOVA procedure, which you'll learn more about later in the course, there are several sections to the output such as Levene's test and a table of the between-group effects. You can skip to any one of these sub-components of the ANOVA output by clicking on the appropriate branch of the tree diagram. So, if you wanted to skip straight to the between-group effects you should move the on-screen arrow to the left-hand portion of the window and click where it says *Tests of Between-Subjects Effects*. This action will highlight this part of the output in the main part of the viewer

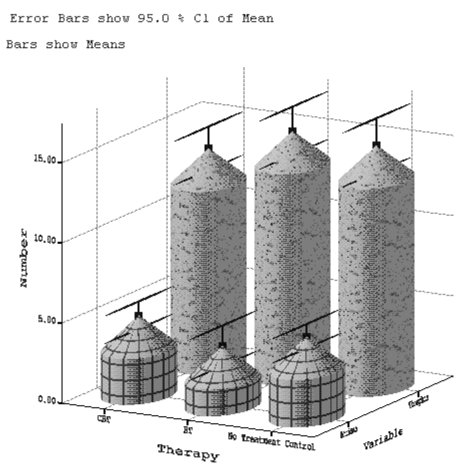
Presenting Data

Does and Don’ts of Presenting Data

The advent of computers has allowed people to spend vast amounts of time producing very snazzy looking graphs, but I hope to convince you that snazzy is not always best! Tufte has written extensively about how data should be presented (e.g., Tufte, 2001). Some of books look at various types of visual displays and discusses some of Tufte’s advice (Field, 2013, 2016); Briefly, Tufte points out that graphs should, amongst other things:

- ✓ Show the data
- ✓ Induce the reader to think about the data being presented (rather than some other aspect of the graph, like how nice the colour scheme is).
- ✓ Avoid distorting the data
- ✓ Present many numbers with minimum ink!
- ✓ Make large data sets coherent
- ✓ Encourage the reader to compare different pieces of data
- ✓ Reveal data.

However, graphs often don’t do these things. Let’s look at an example of a bad graph.



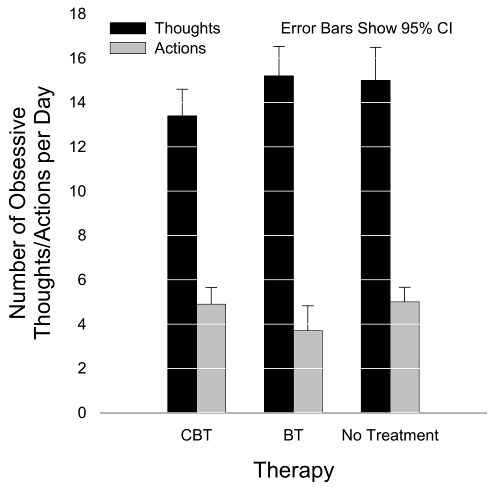


Figure 5: A cringingly bad example of a graph from the first edition of my book ‘Discovering Statistics …’ (2000), and a better example of how it should be done

One of the most truly appalling examples of bad graphics was done by me in the first edition of my book ‘Discovering Statistics …’ (2000). Overexcited by SPSS’s ability to put all sorts of useless crap on graphs (like 3-D effects, fill effects and so on) I literally went into some weird orgasmic state and produced an absolute abomination. It’s a graph showing the mean number of obsessive thoughts and behaviours experienced by groups of people suffering from obsessive compulsive disorder after three types of therapy: Cognitive Behaviour Therapy (CBT), Behaviour therapy (BT) and no treatment. Figure 5 reproduces this graph and shows an alternative display of these data. What’s wrong with the original (left)?

- ✗ The bars have a 3-D effect: Never use 3-D on a 2-D graph because all it does is obscure the data. In particular, it makes it hard to see the values of the bars because of the 3-D effect. This graph is a great example because the 3-D effect makes the error bars almost impossible to read.

- Patterns: The bars also have patterns, which although very pretty, merely distract the eye from what matters (namely the data). These are completely unnecessary!
- Cylindrical bars: what's that all about?
- Badly labelled y-axis: 'number' of what … delusions? Fish? Cabbage eating sea lizards from the 8th dimension? Idiots who don't know how to draw graphs?

Now, take a look at the redone version (right). What improvements have been made?

- ✓ 2-D: The completely unnecessary third dimension is gone making it much easier to compare the values across therapies and thoughts/behaviours.
- ✓ The y-axis has a more informative label: we now know that is was the number of obsessive thoughts or actions per day that was being measured.
- ✓ Distractions: There are fewer distractions like patterns, cylindrical bars and the like!

Tufte (2001) goes a step further and recommends trying to minimise the amount of ink used to present data.

Governments love to lie with statistics, but scientists shouldn't. How you present your data can make a huge difference to the message conveyed to the audience. Figure 6 shows two graphs that, believe it or not, display exactly the same data. The first panel shows how the graph should probably be scaled. The important thing is that the y-axis begins at 0, and this creates the correct impression: that there people have more nightmares after eating cheese. However, imagine you were the minister for cheese and you wanted to create the impression that cheese does not induce nightmares, all you would need to do is re-scale the graph (by not starting the y-axis at zero) and there suddenly seems to be only a small difference difference. Tempting as it is, don't do this (unless you plan to be a politician, in which case you might as well start practicing this kind of deceit).

To draw a clear graph follow a few of Tufte's recommendations:

- ✓ Don't create false impressions of what the data actually show (likewise, don't hide effects!) by scaling the *y*-axis in some weird way.
- ✓ Abolish chartjunk: don't use patterns, 3-D effects, shadows, pictures of hippopotami, photos of your grandma or anything else.
- ✓ Avoid excess ink: this is a bit radical, but if you don't need the axis, then get rid of them.

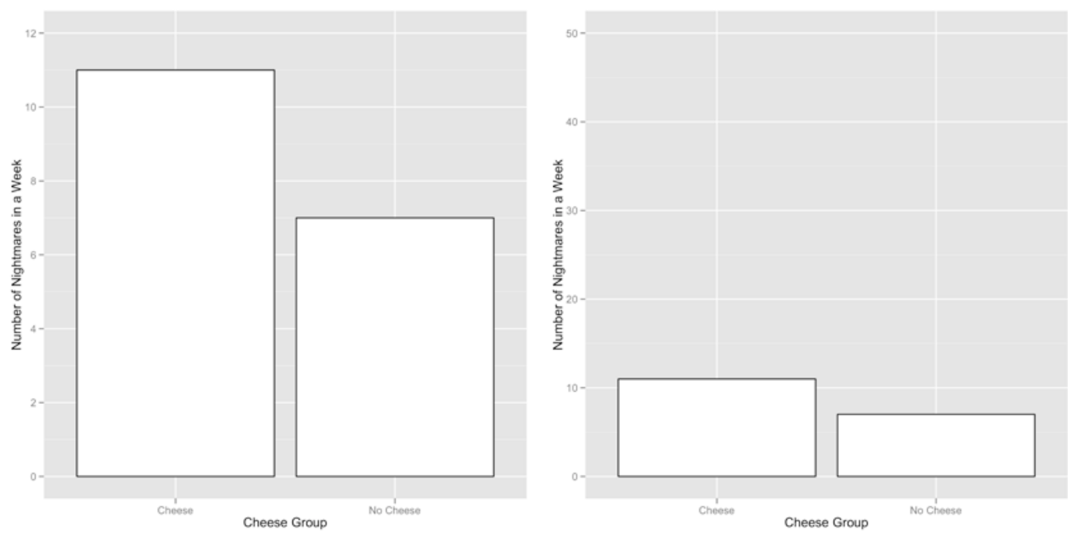


Figure 6: two graphs showing the same thing

The SPSS Chart Builder

In SPSS we can use the all-singing and all-dancing Chart Builder to produce graphs. Figure 7 shows the basic Chart Builder dialog box, which is accessed through the Graphs Chart Builder... menu. There are some important parts of this dialog box:

→ **Gallery**: For each type of graph, a gallery of possible variants is shown. Double-click on an icon to select a particular type of graph.

→ **Variable list**: The variables in the data editor are listed here. These can be dragged into drop zones to specify what is shown in a given graph.

→ **The canvas**: This is the main area in the dialog box and is where a preview of the graph is displayed as you build it.

→ **Drop zones**: These zones are designated with blue dotted lines. You can drag variables from the variable list into these zones.

There are two ways to build a graph: the first is by using the gallery of predefined graphs and the second is by building a graph on an element-by-element basis. The gallery is the default option and this tab (Gallery Basic Elements Groups/Point ID Titles/Footnotes) is automatically selected and is fine for our purposes.

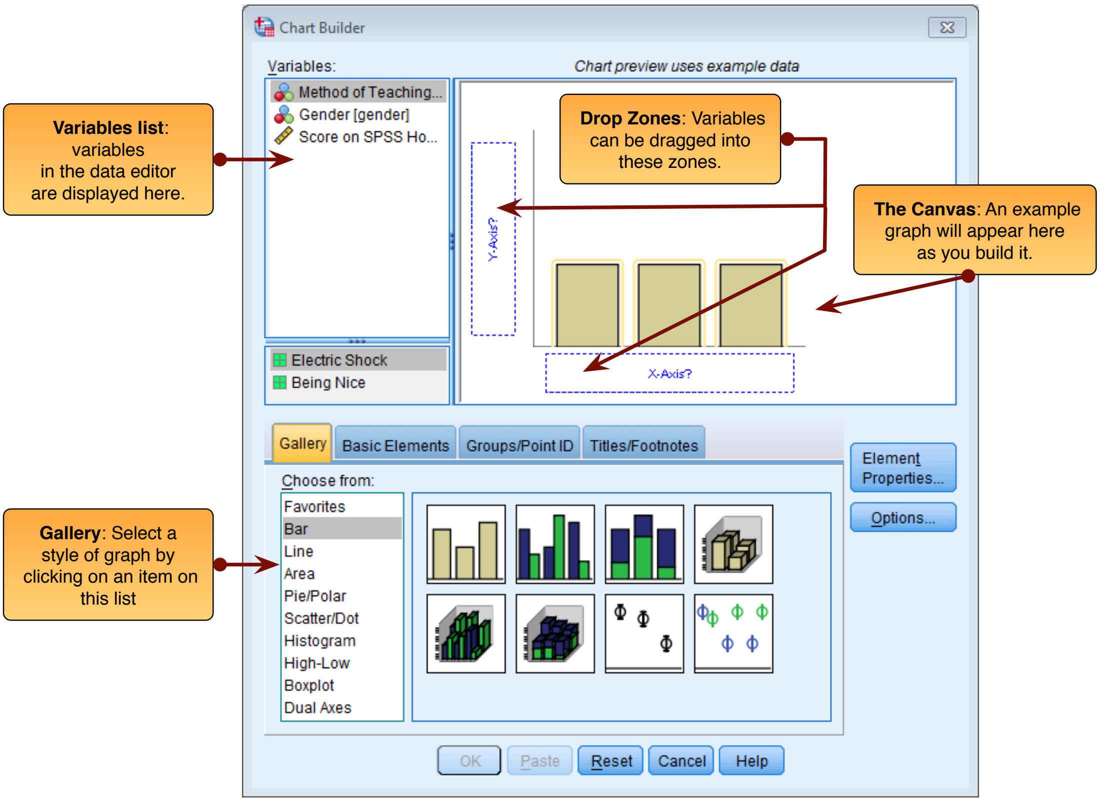
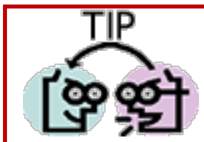
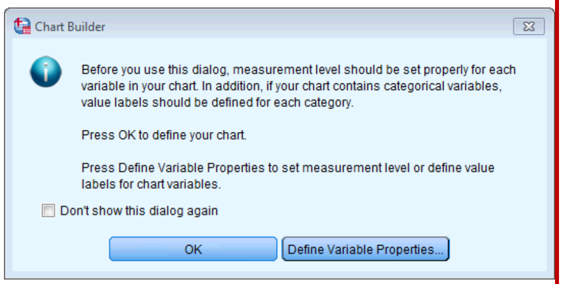
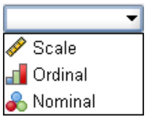


Figure 7: The SPSS Chart Builder



When you first use the chart builder to draw a graph you will see a dialog box that seems to signal an impending apocalypse. In fact, SPSS is just helpfully(?!) reminding you that for the Chart Builder to work, you need to have set the level of measurement correctly for each variable. That is, when you defined each variable you must have set them correctly to be Scale, Ordinal or Nominal. This is because SPSS needs to know whether variables are categorical (nominal) or continuous (scale) when it creates the graphs. If you have been diligent and set these properties when you entered the data then simply click on OK to make the dialog disappear. If you forgot to set the level of measurement for any variables then click on Define Variable Properties... to go to a new dialog box in which you can change the properties of the variables in the data editor.





Graphing means: bar charts and error bars

How you create bar charts in SPSS depends largely on how you collected your data (whether the means come from independent cases and are, therefore, independent, or came from the same cases and so are related). For this reason, we will look at a variety of situations. Figure 8 shows the various options in the chart builder under the option 'bar'. Given what I've said above, avoid the 3-D options. The main options that you'll use will be:

→ **Simple bar**: Use this option when you just want to see the means of scores across different groups of cases. For example, you might want to plot the mean ratings of two films.

→ **Clustered bar**: If you had a second grouping variable you could produce a simple bar chart (as above) but with bars produced in different colours for levels of a second grouping variable. For example, you could have ratings of the two films, but for each film have a bar representing ratings of 'excitement' and another bar showing ratings of 'enjoyment'.

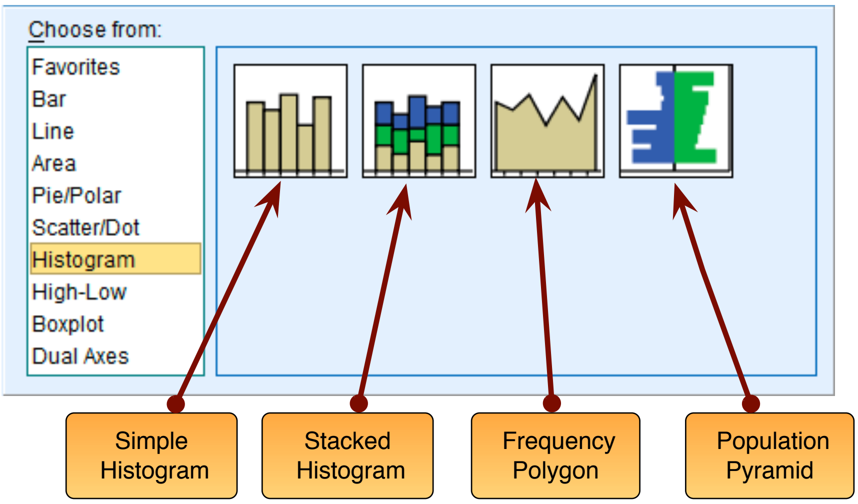


Figure 8: The bar chart gallery

Simple bar charts for independent means

First of all, let's just plot the mean rating of the two films from our film data that we typed in earlier. Open the data file that you saved earlier (**ChickFlick.sav**). We have just one grouping variable (the film) and one outcome (the arousal); therefore, we want a simple bar chart. In the Chart Builder double-click on the icon for a simple bar chart (Figure 8). On the canvas you will see a graph and two drop zones: one for the *y*-axis and one for the *x*-axis. The *y*-axis needs to be the dependent variable, or the thing you've measured, or more simply the thing for which you want to display the mean. In this case it would be **arousal**, so select arousal from the variable list and drag it into the y-axis drop zone (Y-Axis?). The *x*-axis should be the variable by which we want to split the arousal data. To plot the means for the two films, select the variable **film** from the variable list and drag it into the drop zone for the *x*-axis (X-Axis?).

Figure 9 shows the completed Chart Builder for the bar chart and some other options. The 'element properties' dialog box should appear when you select the type of graph you want, but if it doesn't click on Element Properties... in the Chart Builder. There are three important features of this dialog box. The first is that, by default, the bars will display the mean value. This is fine, but just note that you can plot other summary statistics such as the median or mode. Second, just because you've selected a simple bar chart doesn't mean that you have to have a bar chart. Also, you can ask SPSS to add error bars to your bar chart to create an error bar chart by selecting ☑ Display error bars. You have a choice of what your error bars represent. Normally, error bars show the 95% confidence interval, and I have selected this option (◉ Confidence intervals).Note, though, that you can change the width of the confidence interval displayed by changing the '95' to a different value. You can also display the standard error (the default is to show 2 standard errors, but you can change this to 1) or standard deviation (again, the default is 2 but this could be changed to 1 or another value). It's important that when you change these properties that you click on Apply: if you don't then the changes will not be applied to Chart Builder. Click on OK to produce the graph.

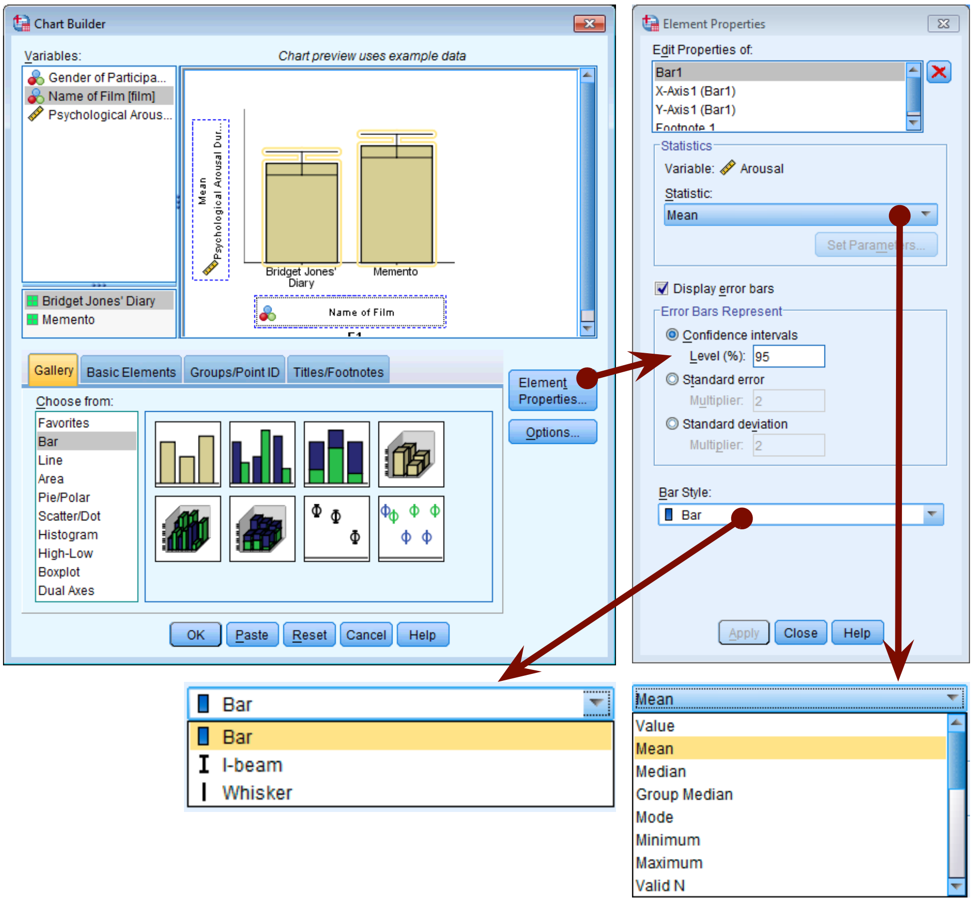


Figure 9: Dialog boxes for a simple bar chart with error bar

Figure 10 shows the resulting bar chart. This graph displays the mean (and the confidence interval of those means) and shows us that on average, people were more aroused by *Memento* than they were by *Bridget Jones' Diary*. However, we originally wanted to look for gender effects, so this graph isn't really telling us what we need to know. The graph we need is a *clustered graph*.

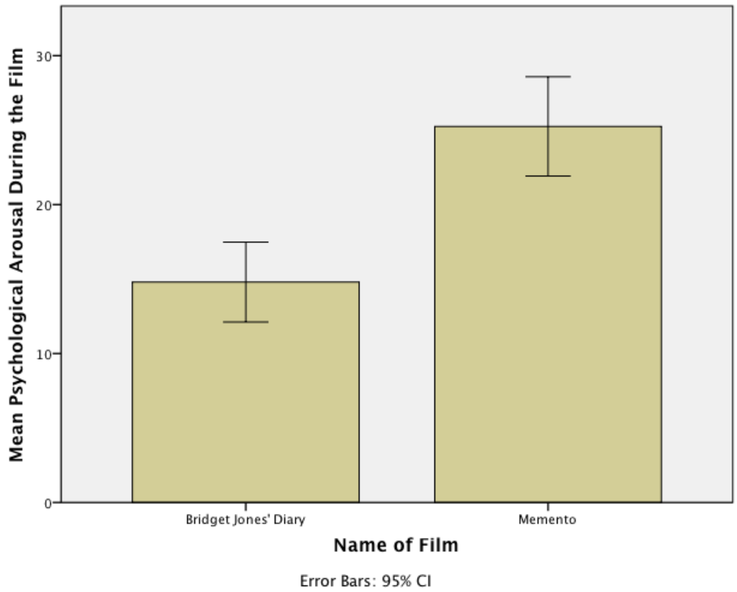


Figure 10: Bar chart of the mean arousal for each of the two films.

Clustered bar charts for independent means

To do a clustered bar chart for means that are independent (i.e. have come from different groups) we need to double-click on the clustered bar chart icon in the Chart Builder (Figure 8). On the canvas you will see a graph as with the simple bar chart but there is now an extra drop zone: Cluster on X: set color. All we need to do is to drag our second grouping variable into this drop zone. As with the previous example, select arousal from the variable list and drag it into Y-Axis?, then select film from the variable list and drag it into X-Axis?. In addition, though, select the Gender variable and drag it into Cluster on X: set color. This will mean that bars representing males and females will be displayed in different colours. As in the previous section, select error bars in the properties dialog box and click on Apply to apply them to the Chart Builder. Figure 11 shows the completed Chart Builder. Click on OK to produce the graph.

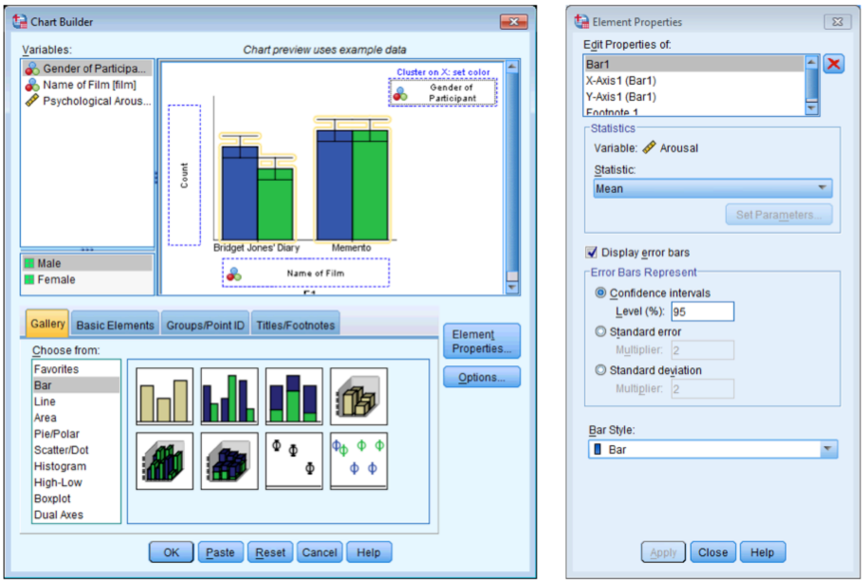


Figure 11: Dialog boxes for a clustered bar chart with error bar

Figure 12 shows the resulting bar chart. Like the simple bar chart this graph tells us that arousal was overall higher for *Memento* than *Bridget Jones' Diary*, but it also splits this information by gender. The mean arousal for *Bridget Jones'*

Diary shows that males were actually more aroused during this film than females. This indicates they enjoyed the film more than the women did! Contrast this with *Memento*, for which arousal levels are comparable in males and females. On the face of it, this contradicts the idea of a 'chick flick': it actually seems that men enjoy chick flicks more than the chicks (probably because it's the only help we get to understand the complex workings of the female mind!).

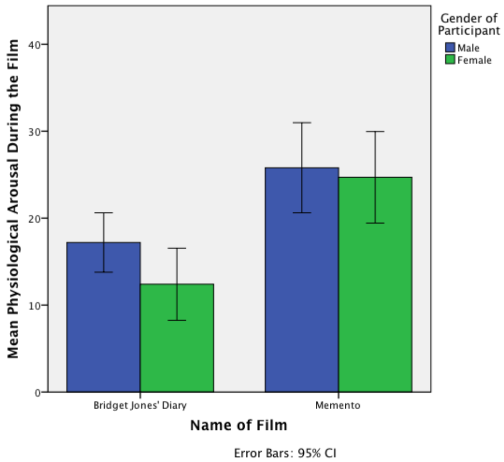


Figure 12: Bar chart of the mean arousal for each of the two films

Simple bar charts for related means

Load the file **Hiccups.sav** that you saved earlier on. In the previous two examples we have used grouping variables to specify aspects of the graph (e.g. we used the grouping variable **film** to specify the *x*-axis). For repeated-measures data we will not have these grouping variables and so the process of building a graph is a little more complicated (but not a lot more).

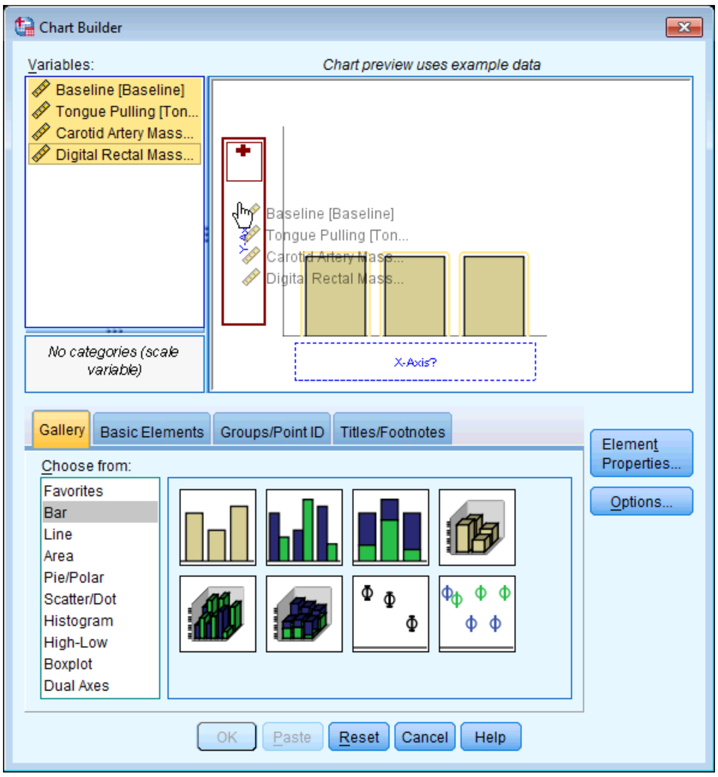


Figure 13: Specifying a simple bar chart for repeated-measures data

To plot the mean number of hiccups go to the Chart Builder and double-click on the icon for a simple bar chart (Figure 8). As before, you will see a graph on the canvas with drop zones for the *x*- and *y*-axis. Previously we specified the column in our data that contained data from our outcome measure on the *y*-axis, but for these data we have four columns containing data on the number of hiccups (the outcome variable). What we have to do then is to drag all four of these variables from the variable list into the *y-axis* drop zone. We have to do this simultaneously. First, we need to select multiple items in the variable list: to do this, select the first variable by clicking on it with the mouse. The variable will be highlighted in blue. Now, hold down the *Ctrl* key on the keyboard and click on a second variable. Both variables are now highlighted in blue. Again, hold down the *Ctrl* key and click on a third variable in the variable list and so on for the fourth. In cases in which you want to select a list of consecutive variables, you can do this very quickly by simply clicking on the first variable that you want to select (in this case **baseline**), hold down the *Shift* key on the keyboard and then click on the last variable that you want to select (in this case **digital rectal massage**); notice that all of the variables in between have been selected too. Once the four variables are selected you can drag them by clicking on any one of the variables and then dragging them into  as shown in Figure 13.

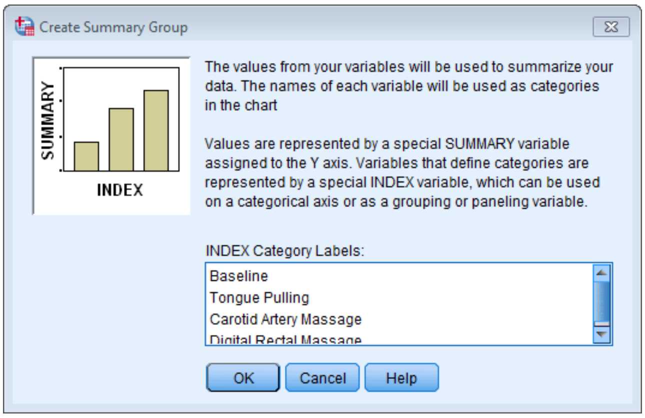


Figure 14: The *Create Summary Group* dialog box

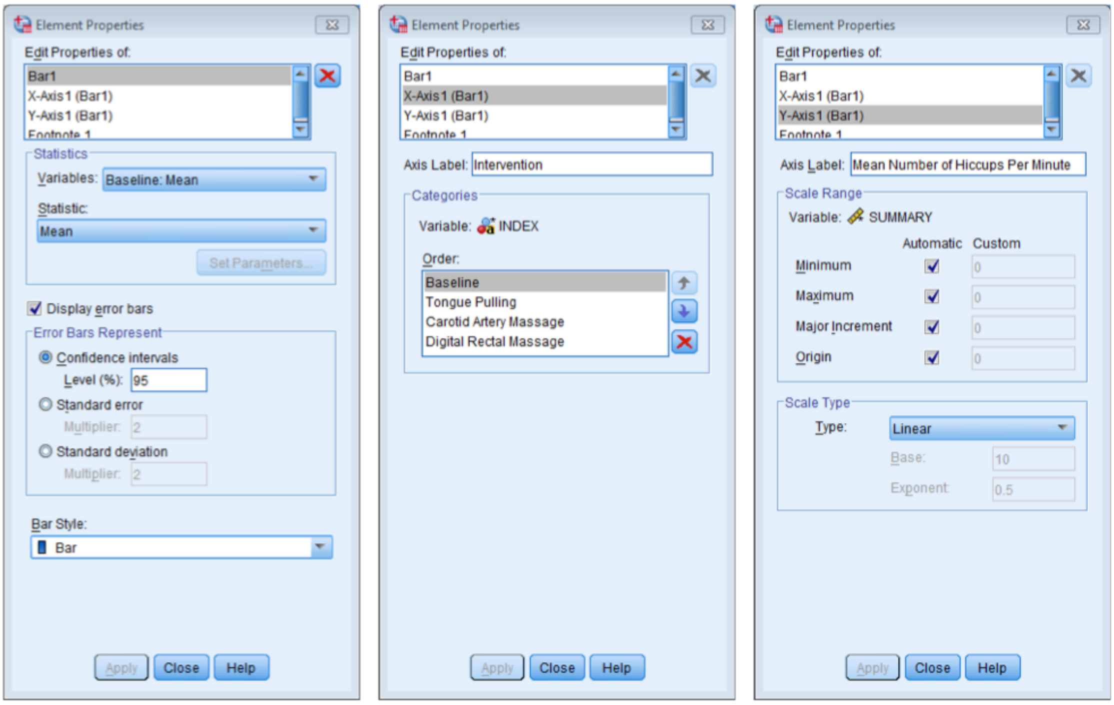


Figure 15: Setting Element Properties for a repeated-measures graph

Once you have dragged the four variables onto the y-axis drop zones a new dialog box appears (Figure 14). This box tells us that SPSS is creating two temporary variables. One is called Summary, which is going to be the outcome variable (i.e. what we measured — in this case the number of hiccups per minute). The other is called index and this variable will represent our independent variable (i.e. what we manipulated — in this case the type of intervention). SPSS uses these temporary names because it doesn't know what our particular variables represent, but we should change them to something more helpful! Just click on OK to get rid of this dialog box.

We need to edit some of the properties of the graph. Figure 15 shows the options that need to be set: if you can't see this dialog box then click on Element Properties... in the Chart Builder. In the left panel of Figure 15 just note that I have selected to display error bars (see the previous two sections for more information). The middle panel is accessed by clicking on *X-Axis1 (Bar1)* in the list labelled *Edit Properties of* which allows us to edit properties of the horizontal axis. The first thing we need to do is give the axis a title and I have typed *Intervention* in the space labelled *Axis Label*. This label will appear on the graph. Also, we can change the order of our variables if we want to by selecting a variable in the list labelled *Order* and moving it up or down using and . If we change our mind about displaying one of our variables then we can also remove it from the list by selecting it and clicking on . Click on Apply for these changes to take effect. The right panel of Figure 15 is accessed by clicking on *Y-Axis1 (Bar1)* in the list labelled *Edit Properties of* which allows us to edit properties of the vertical axis. The main change that I have made here is to give the axis a label so that the final graph has a useful description on the axis (by default it will just say Mean, which isn't very helpful). I have typed 'Mean Number of Hiccups Per Minute' in the box labelled *Axis Label*. Also note that you can use this dialog box to set the scale of the vertical axis (the minimum value, maximum value and the major increment, which is how often a mark is made on the axis). Mostly you can let SPSS construct the scale automatically and it will be fairly sensible — and even if it's not you can edit it later. Click on Apply to apply the changes.

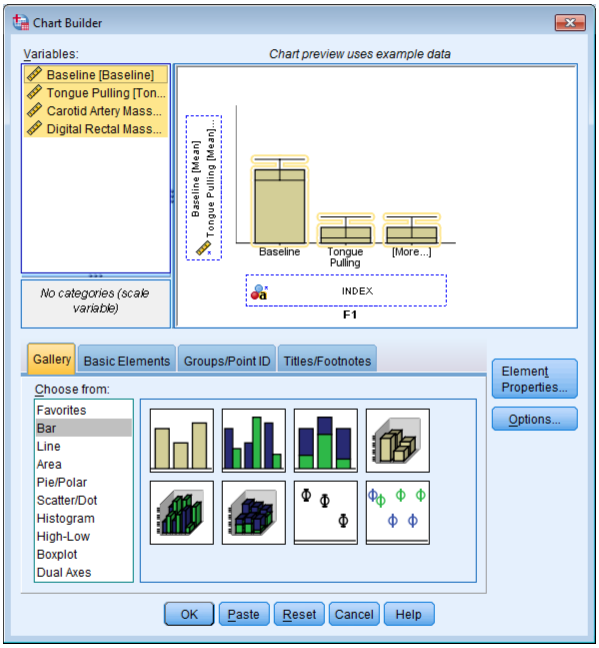


Figure 16: Completed Chart Builder for a repeated-measures graph

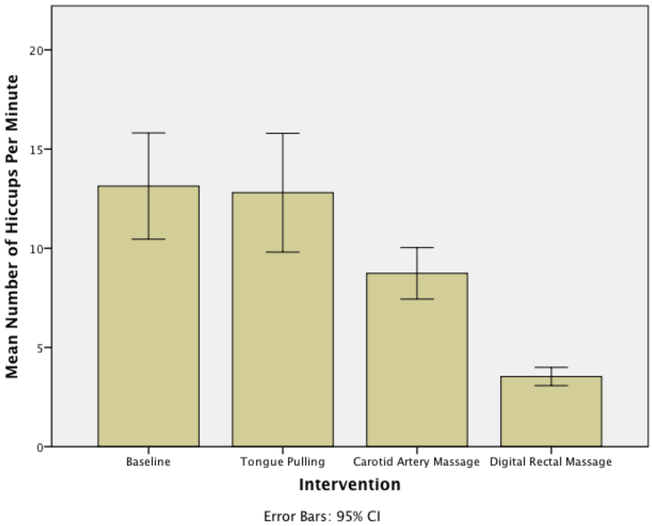


Figure 17: Bar chart of the mean number of hiccups at baseline and after various interventions

Figure 16 shows the completed Chart Builder. Click on OK to produce the graph. The resulting bar chart in **Figure** 17 displays the mean (and the confidence interval of those means) number of hiccups at baseline and after the three interventions. Note that the axis labels that I typed in have appeared on the graph. The error bars on graphs of repeated-measures designs aren't actually correct (see Field, 2013, Chapter 9) but you'll have to read the book if you want to sort that problem out! We can conclude that the amount of hiccups after tongue pulling was about the same as at baseline; however, carotid artery massage reduced hiccups, but not by as much as a good old fashioned digital-rectal massage. The moral here is: if you have hiccups, find something digital and go amuse yourself for a few minutes.

Clustered bar charts for related means

Now we have seen how to plot means that are related (i.e. show different conditions applied to the same group of cases), you might well wonder what you do if you have a second independent variable that had been measured in the same sample. You'd do a clustered bar chart, right? Wrong? Actually, the SPSS Chart Builder doesn't appear to be able to cope with this situation at all — at least not that I can work out from playing about with it.

Clustered bar charts for 'mixed' designs

The Chart Builder might not be able to do charts for multiple repeated-measures variables, but it can graph what is known as a mixed design. This is a design in which you have one or more independent variables measured using different groups, and one or more independent variables measured using the same sample. Basically, the Chart Builder can produce a graph provided you have only one variable that was a repeated measure.

We all like to text-message. What will happen to the children, though? Not only will they develop super-sized thumbs, they might not learn correct written English. Imagine we conducted an experiment in which a group of 25 children was encouraged to send text messages on their mobile phones over a six-month period. A second group of 25 children was forbidden from sending text messages for the same period. To ensure that kids in this latter group didn't use their phones, this group was given armbands that administered painful shocks in the presence of microwaves (like those emitted from phones). The outcome was a score on a grammatical test (as a percentage) that was measured both before and after the intervention. The first independent variable was, therefore, text message use (text messagers versus controls) and the second independent variable was the time at which grammatical ability was assessed (baseline or after six months). The data are in the file **Text Messages.sav**.

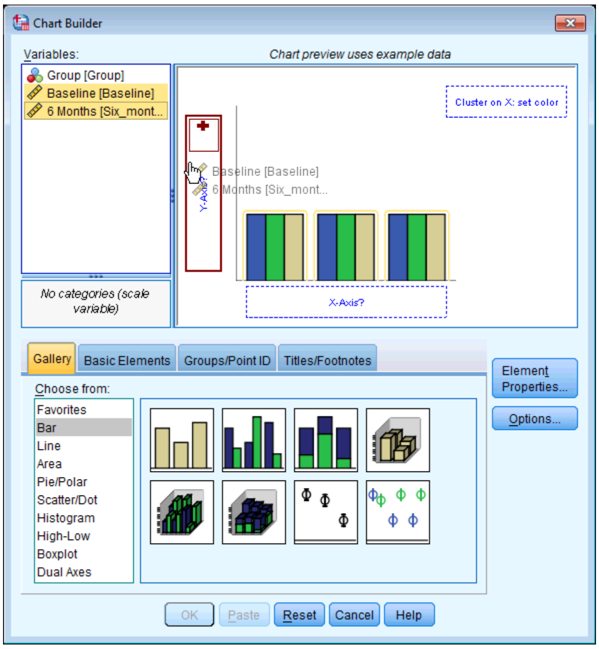
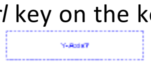
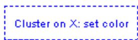


Figure 18: Selecting the repeated-measures variable in the Chart Builder

To graph these data, follow the procedure for graphing related means. Our repeated-measures variable is time (whether grammatical ability was measured at baseline or six months) and is represented in the data file by two columns, one for the baseline data and the other for the follow-up data. In the Chart Builder select these two variables simultaneously by clicking on one and then holding down the *Ctrl* key on the keyboard and clicking on the other. When they are both highlighted click on either one and drag it into  as shown in Figure 18. The second variable (whether children text messaged or not) was measured using different children and so is represented in the data file by a grouping variable (**group**). This variable can be selected in the variable list and dragged into 

. The two groups will be displayed as different-coloured bars.

→ Use what you learnt earlier in this handout to add error bars to this graph and to label both the *x*- (I suggest 'Time') and *y*-axis (I suggest 'Mean Grammar Score (%)').

The finished Chart Builder is in Figure 19. Click on OK to produce the graph. Figure 20 shows the resulting bar chart. It shows that at baseline (before the intervention) the grammar scores were comparable in our two groups; however, after the intervention, the grammar scores were lower in the text messagers than in the controls. Also, if you compare the two blue bars you can see that text messagers' grammar scores have fallen over the six months; compare this to the controls (green bars) whose grammar scores are fairly similar over time. We could, therefore, conclude that text messaging has a detrimental effect on children's understanding of English grammar and civilization will crumble, with Abaddon rising cackling from his bottomless pit to claim our wretched souls. Maybe.

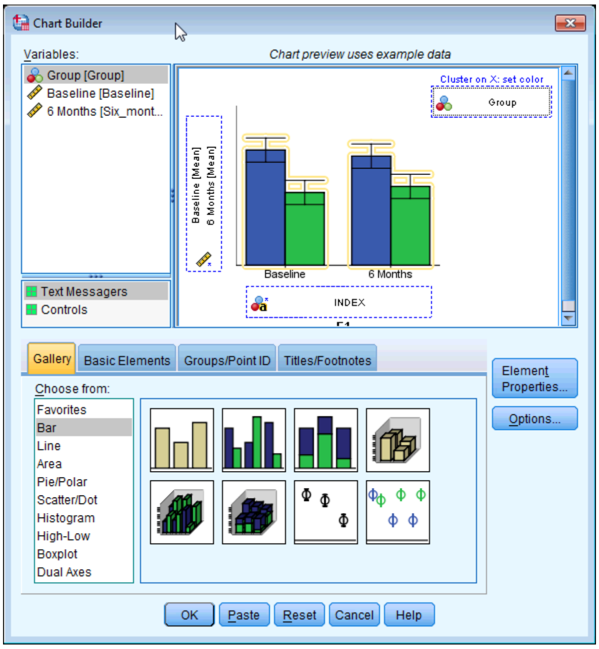


Figure 19: Completed dialog box for an error bar graph of a mixed design

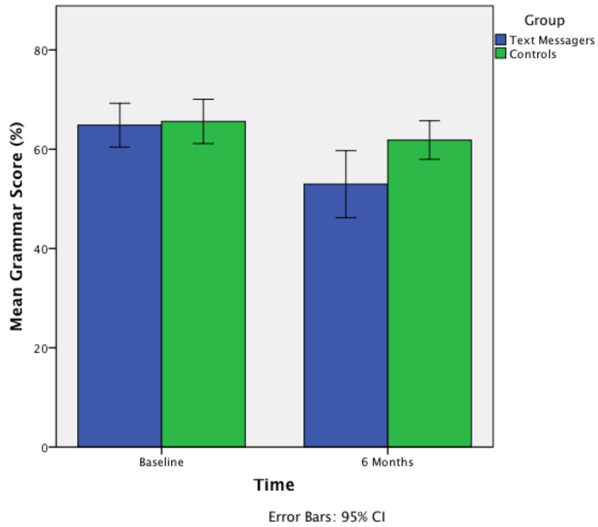


Figure 20: Error bar graph of the mean grammar score over six months in children who were allowed to text-message versus those who were forbidden

Editing Graphs

You can edit graphs, there's a movie that shows you how.

YouTube http://www.youtube.com/watch?v=en0t0QS9uo8

Tasks

Task 1

A neurologist carried out an experiment to investigate the depressant effects of certain recreational drugs. She tested 20 clubbers in all: 10 were given an ecstasy tablet to take on a Saturday night and 10 were allowed only to drink alcohol. Levels of depression were measured using the Beck Depression Inventory (BDI) the day after and midweek.

Using what you know about entering data below into SPSS, Enter the data above into the SPSS Data Editor. Draw two bar charts; one to see whether ecstasy makes you more depressed than alcohol the day after (Sunday) and the other showing whether ecstasy makes you more depressed than alcohol midweek (Wednesday).

Table 3: Drug data

Participant	Drug	BDI (Sunday)	BDI (Wednesday)
1	Ecstasy	15	28
2	Ecstasy	35	35
3	Ecstasy	16	35
4	Ecstasy	18	24
5	Ecstasy	19	39
6	Ecstasy	17	32
7	Ecstasy	27	27
8	Ecstasy	16	29
9	Ecstasy	13	36
10	Ecstasy	20	35
11	Alcohol	16	5
12	Alcohol	15	6
13	Alcohol	20	30
14	Alcohol	15	8
15	Alcohol	16	9
16	Alcohol	13	7
17	Alcohol	14	6
18	Alcohol	19	17
19	Alcohol	18	3
20	Alcohol	18	10

Task 2

Statistics and maths anxiety are common and affect people's performance on maths and stats assignments; women in particular can lack confidence in mathematics (Field, 2010). Zhang, Schmader, and Hall (2013) did a study did an intriguing study in which students completed a maths test in which some put their own name on the test booklet, whereas others were given a booklet that already had either a male or female name on. Participants in the latter two conditions were told that the names on the booklets were to protect their anonymity and that they would be using this other person's name for the purpose of the test. Women who completed the test using a different name performed better than those who completed the test using their own name. (There were no such effects for men.) Zhang et al. concluded that performing under a different name freed women from fears of self-evaluation, allowing them to perform better.

Table 4 contains a small subset of the data from Zhang et al.'s study.

Using what you know about entering data in to SPSS, enter the data below into the SPSS Data Editor. Draw a clustered bar chart to show the mean maths test score (out of 100) for males and females completing the test under the three different types of names.

Table 4: A subsample of Zhang et al.'s (2013) data

Male			Female		
Female Fake Name	Male Fake Name	Own Name	Female Fake Name	Male Fake Name	Own Name
33	69	75	53	31	70
22	60	33	47	63	57
46	82	83	87	34	33
53	78	42	41	40	83
14	38	10	62	22	86
27	63	44	67	17	65
64	46	27	57	60	64
62	27			47	37
75	61			57	80
50	29				

Task 3

According to some highly unscientific research done by a UK department store chain and reported in Marie Clare magazine (http://ow.ly/9Dxvy) shopping is good for you: they found that the average women spends 150 minutes and walks 2.6 miles when she shops, burning off around 385 calories. In contrast, men spend only about 50 minutes shopping, covering 1.5 miles. This was based on strapping a pedometer on a mere 10 participants. Although I don't have the actual data, some simulated data based on these means are below. Enter these data into SPSS and save them as **Shopping Exercise.sav**.

Table 5: Shopping data

Male		Female	
Distance	Time	Distance	Time
0.16	15	1.40	22
0.40	30	1.81	140
1.36	37	1.96	160
1.99	65	3.02	183
3.61	103	4.82	245

Task 4

I was taken by two new stories. The first was about a Sudanese man who was forced to Marry Goat after being caught having sex with it (http://ow.ly/9DyyP). I'm not sure he treated the goat to a nice dinner in a posh restaurant before taking advantage of her, but either way you have to feel sorry for the goat. I'd barely had time to recover from that story when another appeared about an Indian man forced to marry a dog to atone for stoning two dogs and stringing them up in a tree 15 years earlier http://ow.ly/9DyFn. Why anyone would think it's a good idea to enter a dog into matrimony with a man with a history of violent behaviour towards dogs is beyond me. Still, I wondered whether a goat or dog made a better spouse. I found some other people who had been forced to marry goats and dogs and measured their life satisfaction and, also, how much they like animals. Enter these data into SPSS and save as **Goat or Dog.sav**.

Table 6: Goat and dog data

Goat	Dog

Animal Liking	Life Satisfaction	Animal Liking	Life Satisfaction
69	47	16	52
25	6	65	66
31	47	39	65
29	33	35	61
12	13	19	60
49	56	53	68
25	42	27	37
35	51	44	72
51	42		
40	46		
23	27		
37	48		

References

Fesmire, F. M. (1988). Termination of intractable hiccups with digital rectal massage. *Annals of Emergency Medicine, 17*(8), 872.

Field, A. P. (2010). Teaching Statistics. In D. Upton & A. Trapp (Eds.), *Teaching Psychology in Higher Education* (pp. 134-163). Chichester, UK: Wiley-Blackwell.

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.

Field, A. P. (2016). *An adventure in statistics: the reality enigma*. London: Sage.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Zhang, S., Schmader, T., & Hall, W. M. (2013). L'eggo My Ego: Reducing the Gender Gap in Math by Unlinking the Self from Performance. *Self and Identity, 12*(4), 400-412. doi:10.1080/15298868.2012.687012

Terms of Use