Presenting Data

Do’s and Don'ts of Presenting Data

Computers have allowed people to spend vast amounts of time producing very fancy looking graphs. However, fancy is not always best! Tufte (2001) has written an excellent text about how data should be presented; the main points that he (and others) make are summarized extremely well in a very readable article by Wright & Williams (2003).

Tufte (2001) points out that graphs should do, among other things, the following:

- 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 color 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.

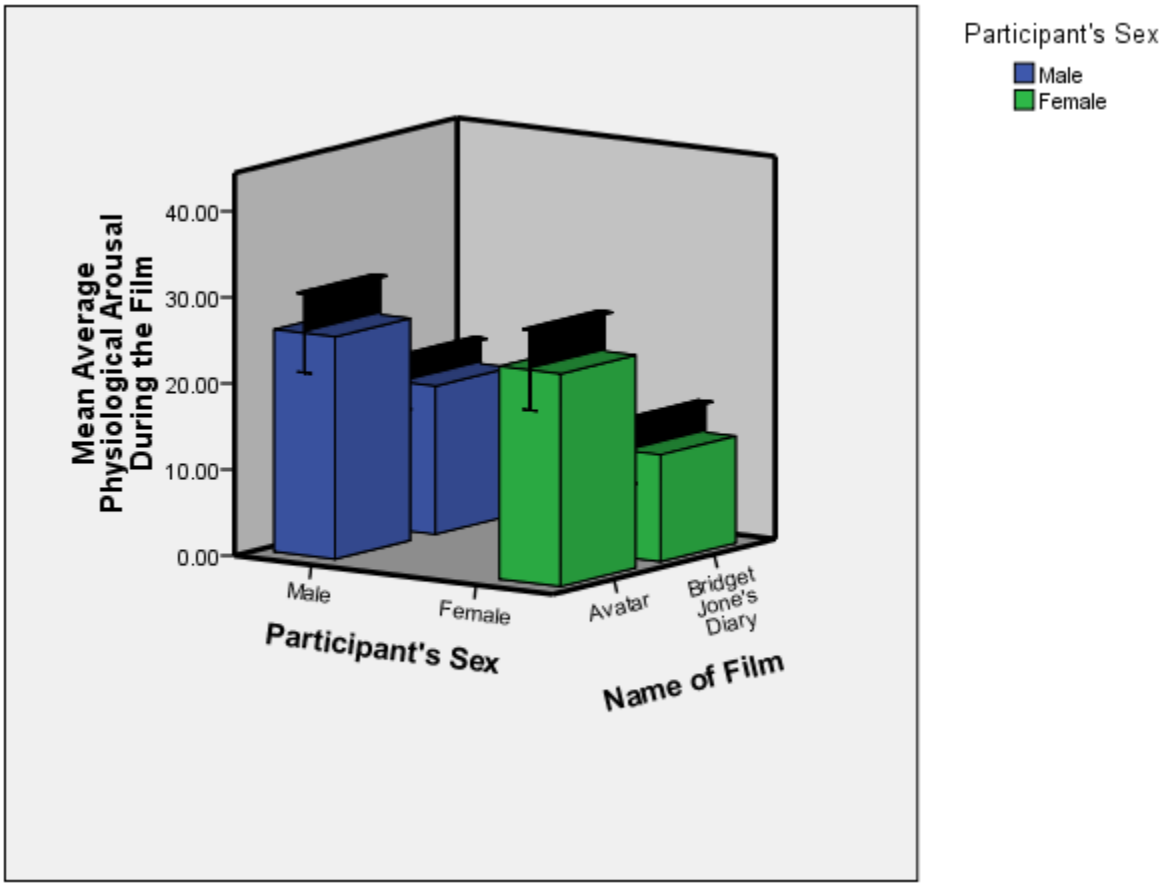
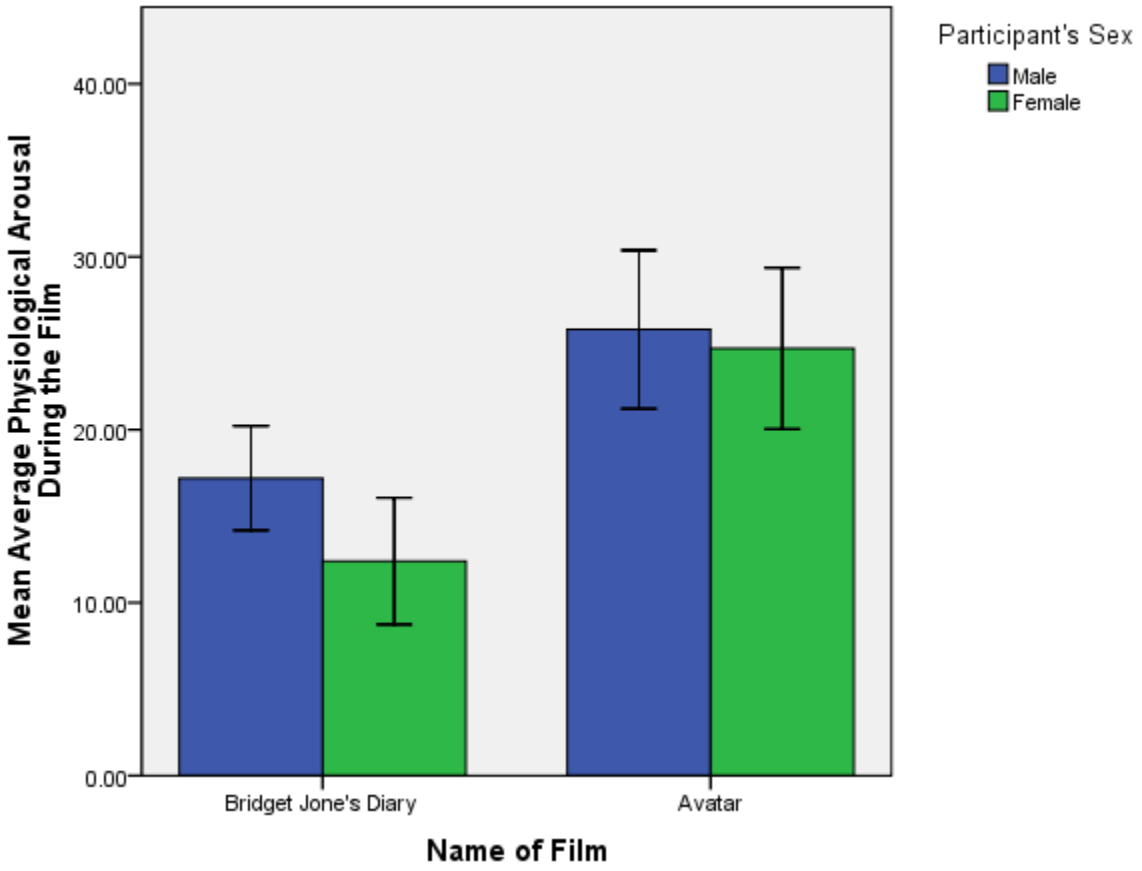


Figure 1: A **really** bad example of a graph and a better example of how it should be done (below)



IBM-SPSS has the ability to put all sorts of useless crap on graphs (like 3-D effects, fill effects and so on) so the first graph is based on some type of altered experience state and the end result sucks.

- 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!

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 sex and type of movie.
- The y-axis has a more informative label: we now know that is the level of arousal that was being measured.
- Distractions: There are fewer distractions like patterns, extra angles and the like!

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

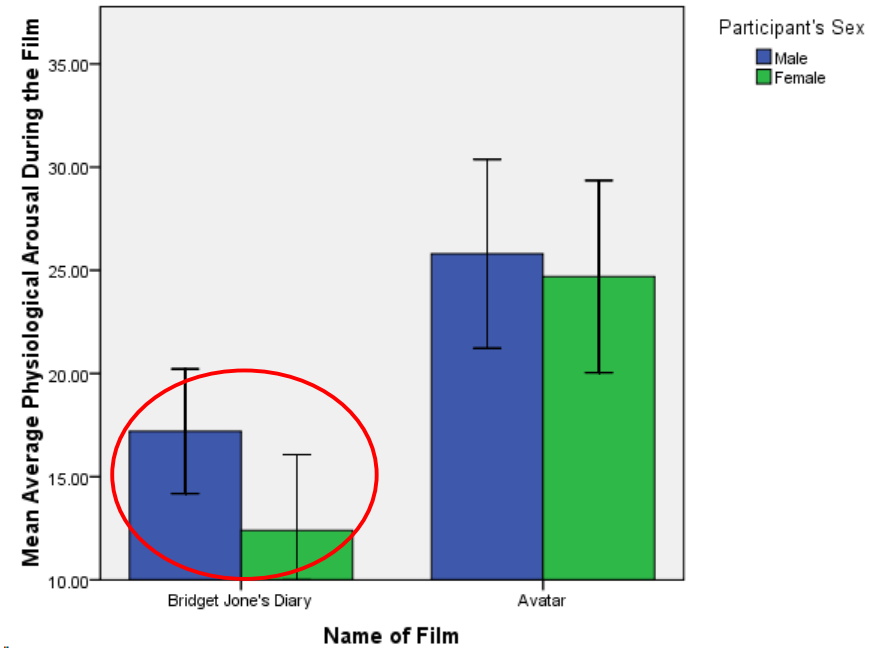


Figure 2

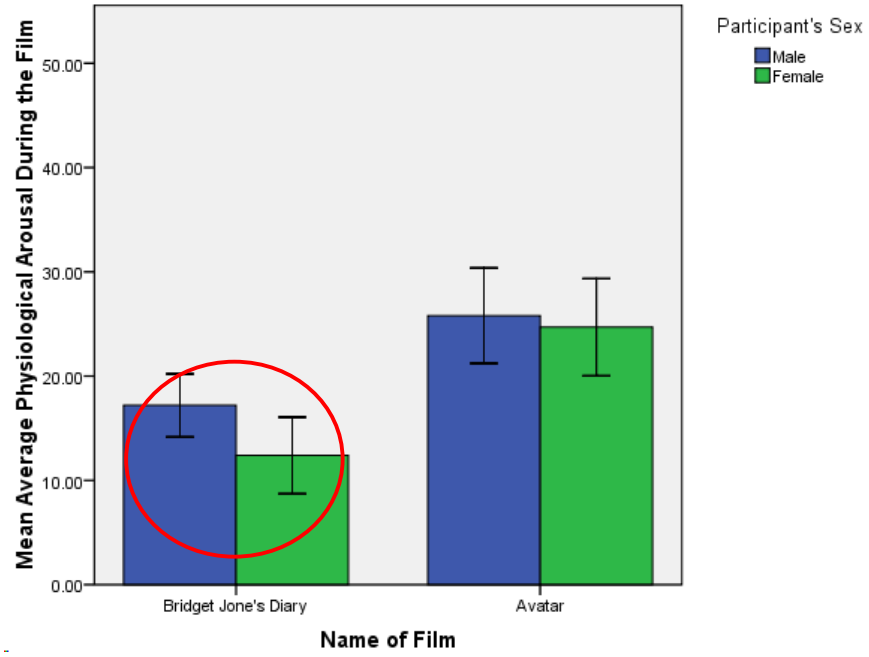


Figure 3

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 2 and 3 shows two graphs that, believe it or not, display exactly the same data. Figure 2 shows how the graph should not be scaled. The important thing is that the y-axis begins at 0 (Figure 3), and this creates the correct impression... However, imagine you wanted to create the impression that there is a sex difference, all we have to do is re-scale the graph (by not starting the y-axis at zero) and there suddenly seems to be a huge difference. In fact, the difference

is numerically very small (and nonsignificant), but the graph makes it look really big! Tempting as it is, don't do this (unless you plan to be President of the United States at some point in your future in which case you might as well start practicing this kind of deceit. Ask Ross Perot.

So, if you want to draw a good 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.

The IBM-SPSS Chart Builder

In IBM-SPSS we can use the all-singing and all-dancing Chart Builder to produce graphs. Figure 4 shows the basic Chart Builder dialog box, which is accessed through the 

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 (

).

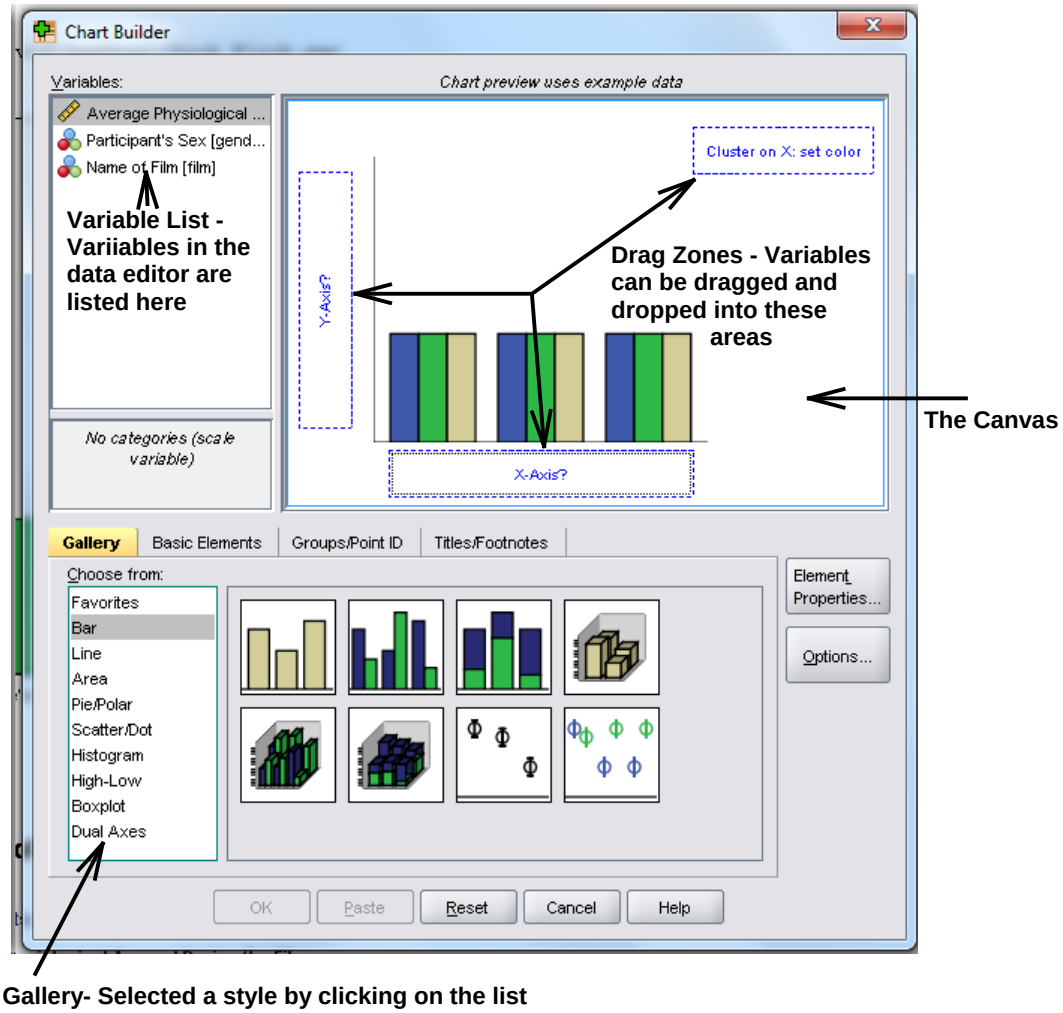
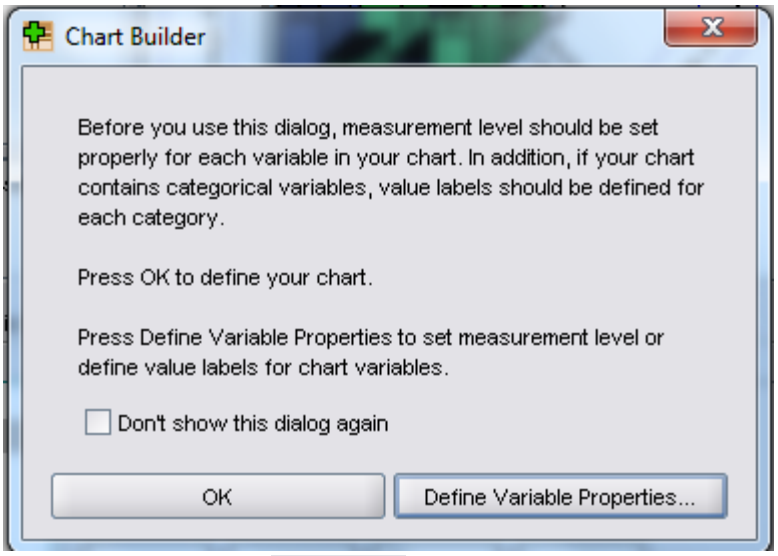


Figure 4: The IBM-SPSS Chart Builder



When you first use the chart builder to draw a graph you will see a dialog box that seems to signal the impending apocalypse (think 2012). In fact, IBM-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 IBM-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  or  to make the dialog disappear. If you forgot to

set the level of measurement for any variables then click on 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 IBM-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 11 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 colors for levels of a second grouping variable.

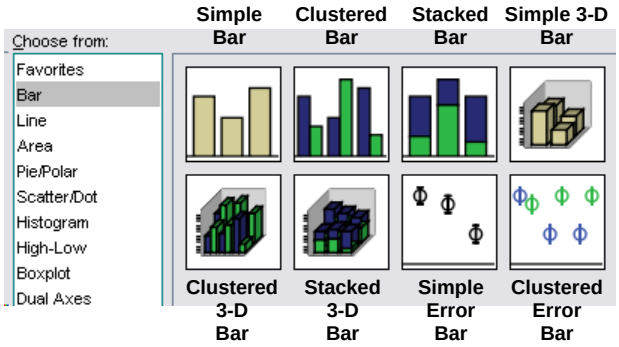


Figure 6: The bar chart gallery

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 two grouping variables (the film & sex) 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 6). 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.

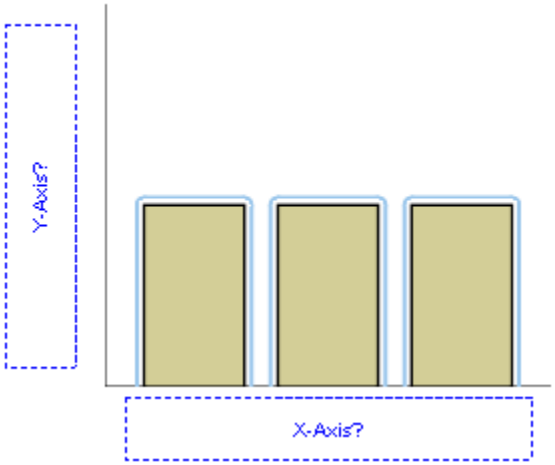
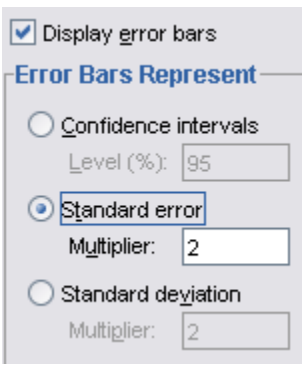
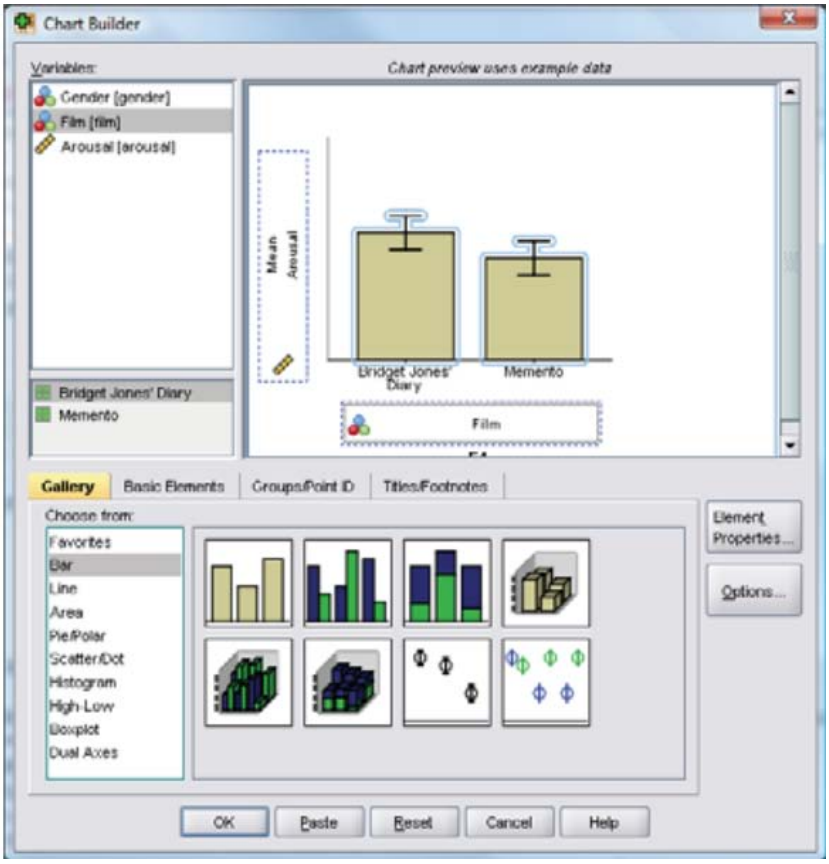


Figure 7

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. Figure 8 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 IBM-SPSS to add error bars to your bar chart to create an error bar chart by selecting Standard error. You have a choice of what your error bars represent. Normally, error bars show the standard error of the mean (SEM), and I have selected this option Note, though, that you can change the width of the SEM (or confidence interval) displayed by changing the '2' or 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.





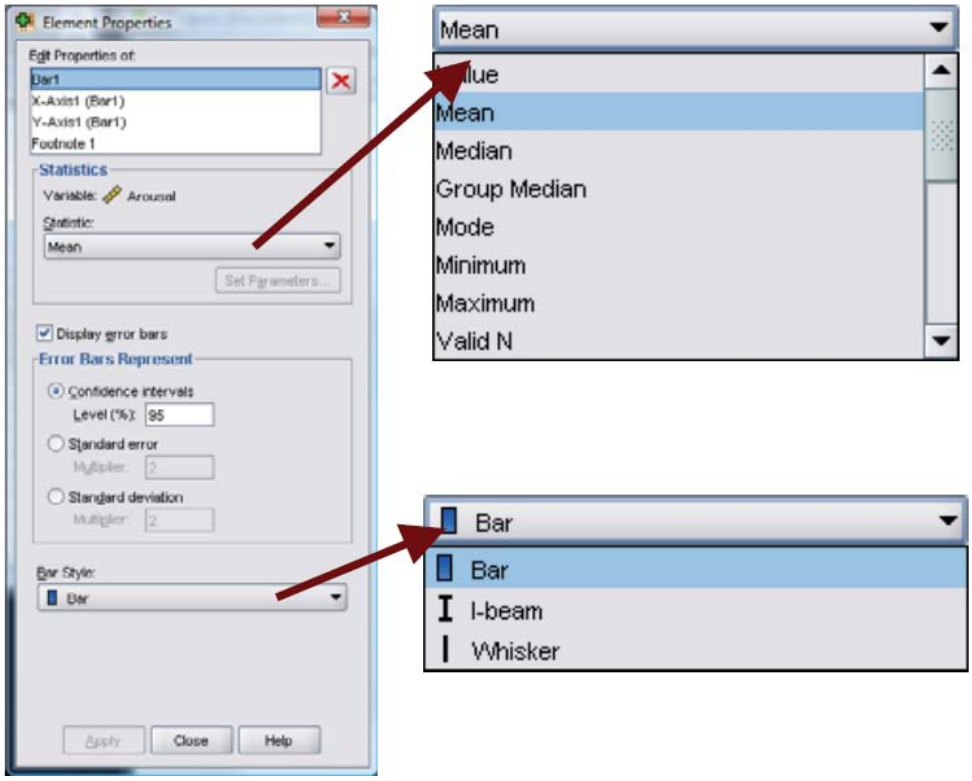


Figure 8: Dialog boxes for a simple bar chart with error bar

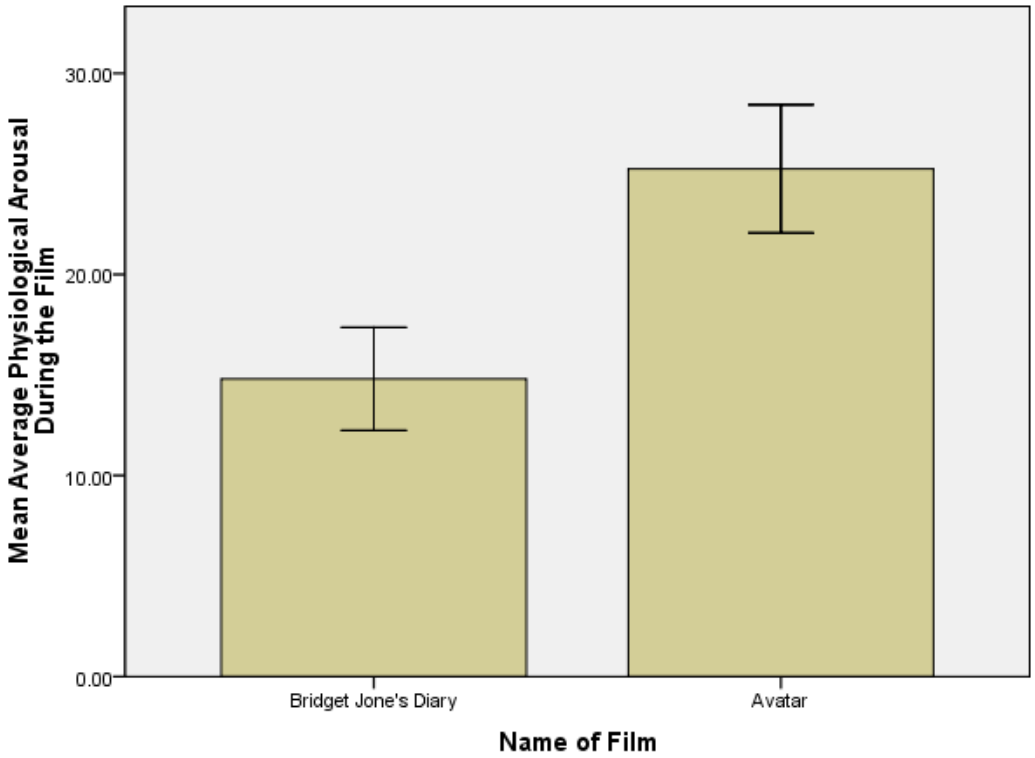


Figure 9: Bar chart of the mean arousal for each of the two films

Figure 9 shows the resulting bar chart. This graph displays the mean (and the SEM 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.

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 6). 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 Sex variable and drag it into Cluster on X: set color. This will mean that bars representing males and females will be displayed in different colors (blue & green). As in the previous graph, select error bars in the properties dialog box and click on Apply to apply them to the Chart Builder. Figure 10 shows the completed Chart Builder. Click on OK to produce the graph.

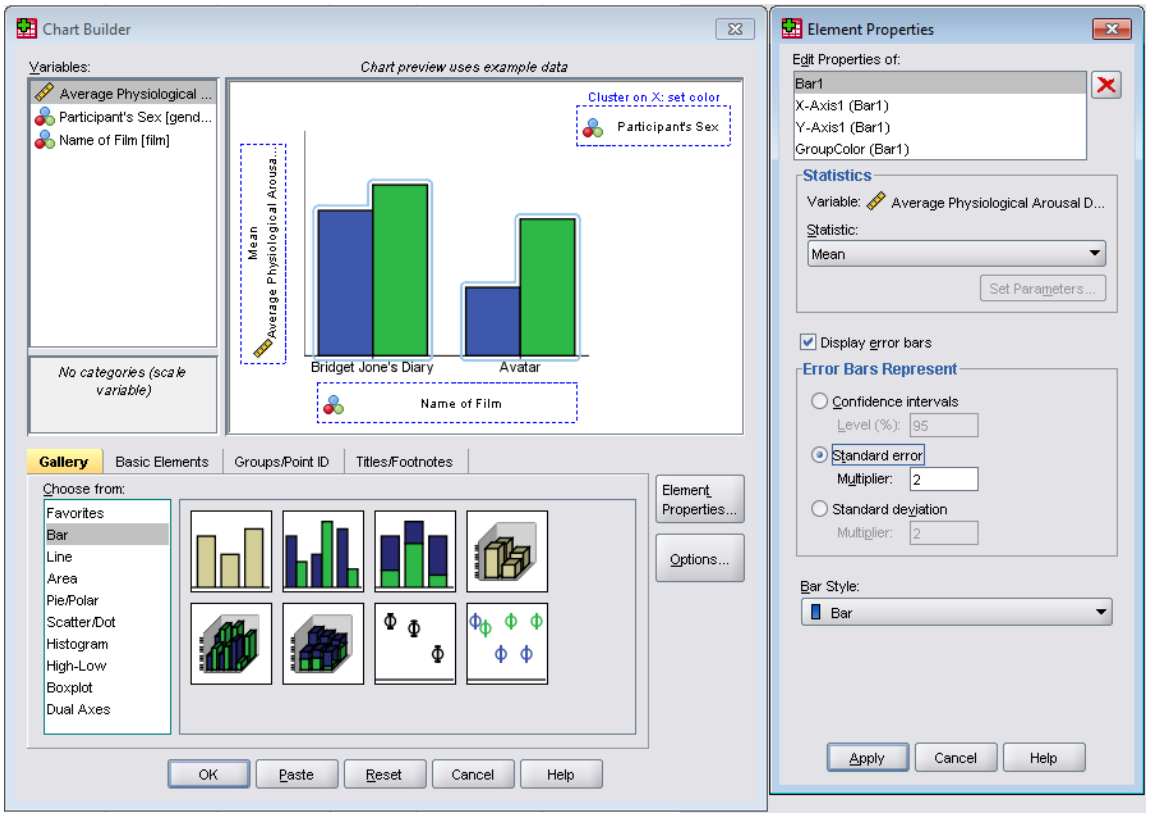


Figure 10: Dialog boxes for a clustered bar chart with error bar

Figure 11 shows the resulting bar chart. Like the simple bar chart this graph tells us that arousal was overall higher for Avatar than Bridget Jones' Diary, but it also splits this information by sex. 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 Avatar, 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!).

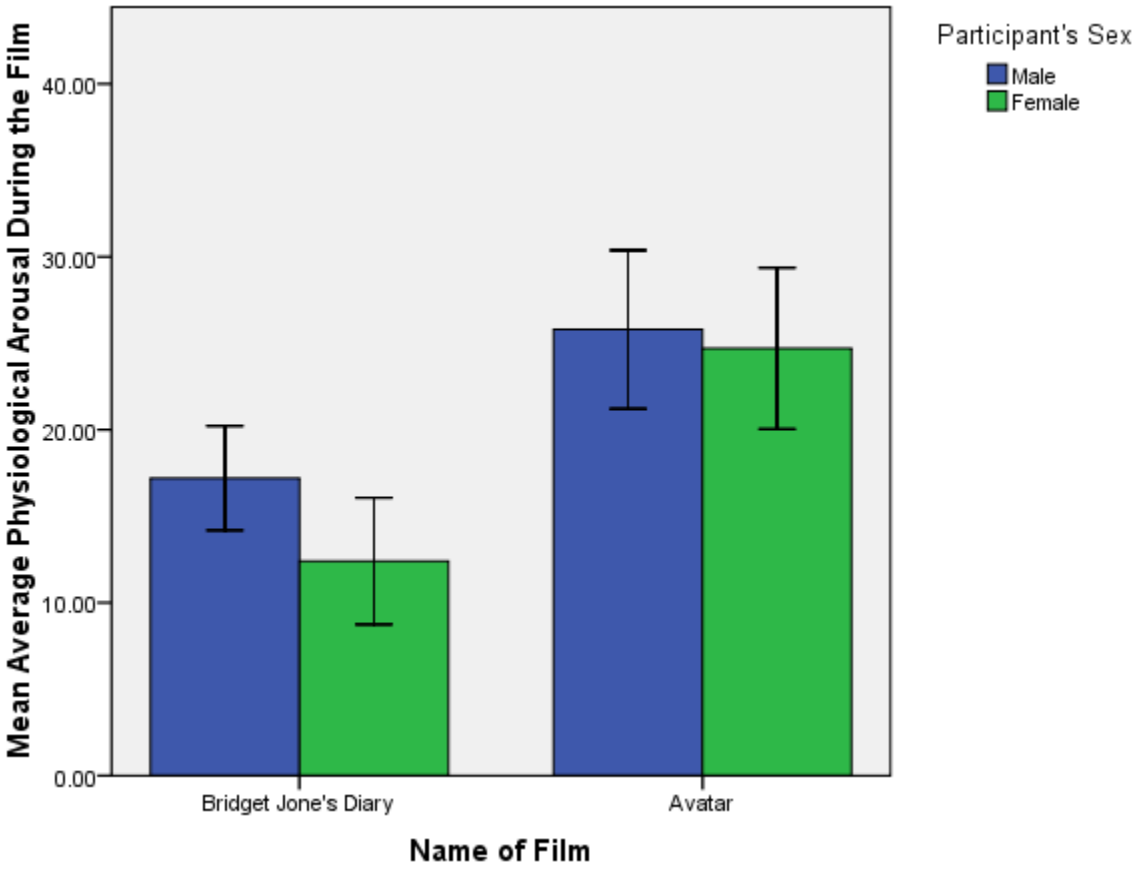


Figure 11: Bar chart of the mean arousal for each of the two films

Editing Graphs

You can edit graphs. Soon, there will be a flash movie on the course website that shows you how. For now …

Double-click anywhere on the graph. The graph becomes shaded and the Chart Editor should pop up (see Figure 12). Here you can edit elements, change colors, etc.

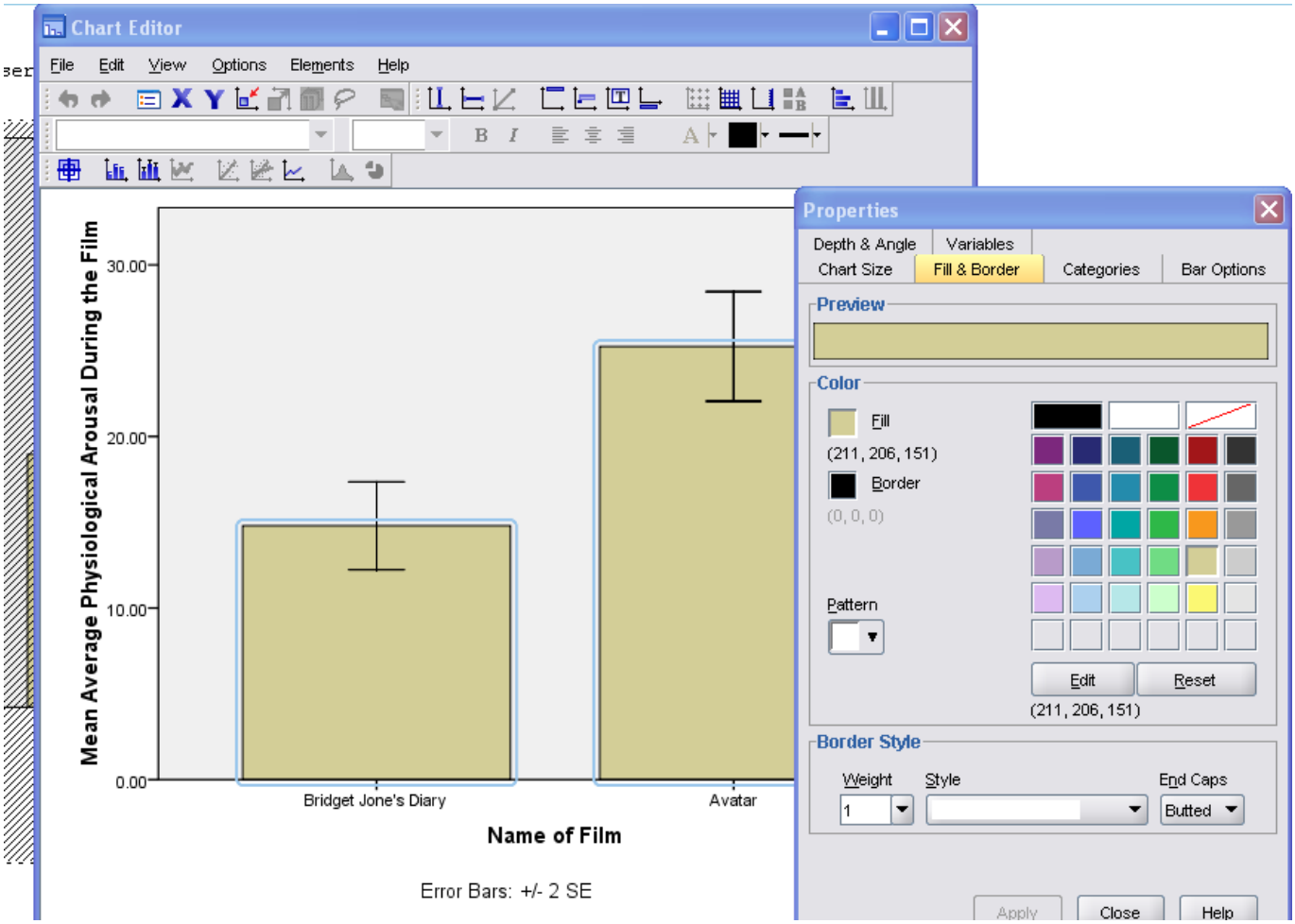
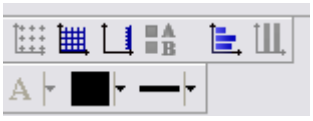
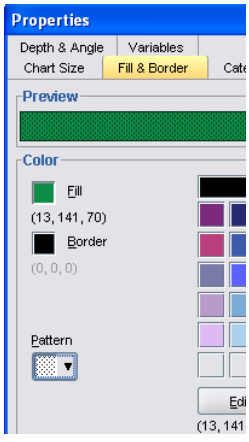
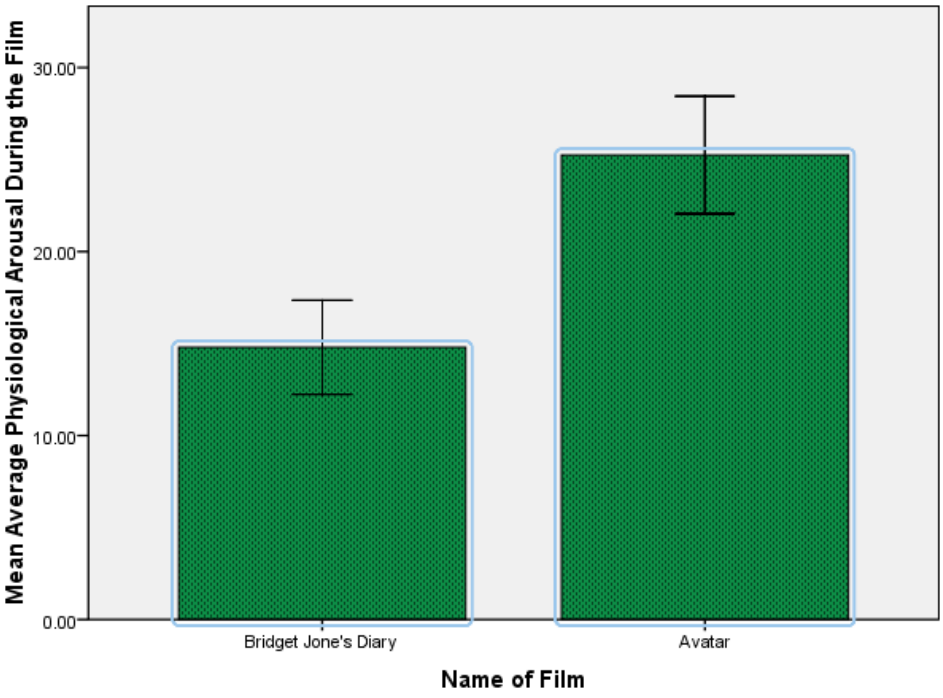


Chart Editor
File Edit View Options Elements Help
Mean Average Physiological Arousal During the Film
30.00
20.00
10.00
0.00
Bridget Jone's Diary
Avatar
Name of Film
Error Bars: +/- 2 SE
Properties
Depth & Angle
Variables
Chart Size
Fill & Border
Categories
Bar Options
Preview
Color
Fill
(211, 206, 151)
Border
(0, 0, 0)
Pattern
Edit
Reset
(211, 206, 151)
Border Style
Weight
Style
End Caps
1
Butted
Apply
Close
Help

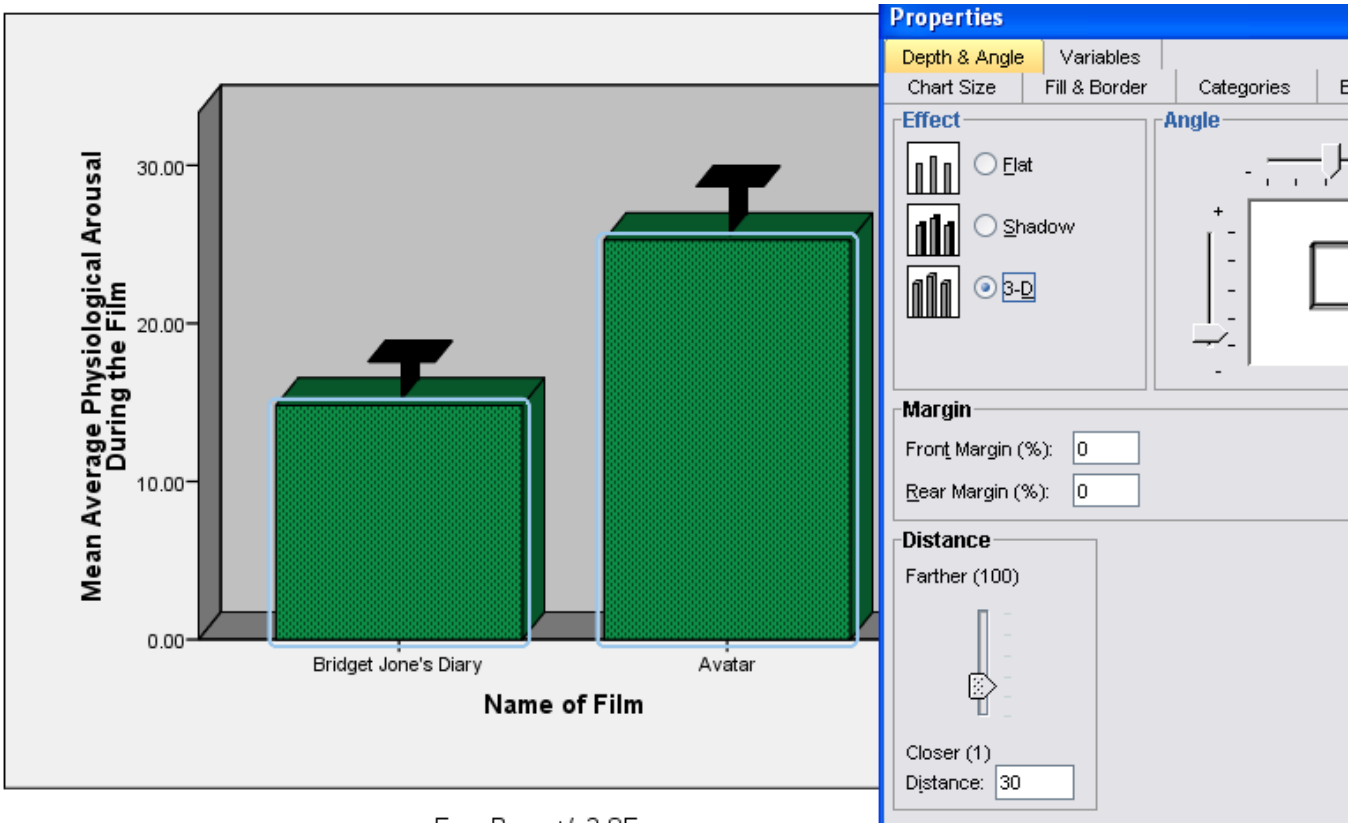
Or…




Properties
Depth & Angle
Variables
Chart Size
Fill & Border
Preview
Color
Fill
(13, 141, 70)
Border
(0, 0, 0)
Pattern


Mean Average Physiological Arousal During the Film
30.00
20.00
10.00
0.00
Bridget Jone's Diary
Avatar
Name of Film

Or …


Mean Average Physiological Arousal During the Film
30.00
20.00
10.00
0.00
Bridget Jone's Diary
Avatar
Name of Film
Properties
Depth & Angle
Variables
Chart Size
Fill & Border
Categories
Effect
Flat
Shadow
3-D
Angle
Margin
Front Margin (%): 0
Rear Margin (%): 0
Distance
Farther (100)
Closer (1)
Distance: 30

References

Tufte, E. R. (2001). *The Visual display of quantitative Information* (2nd Ed.). Cheshire, CT: Graphics Press.

Wright, D. B., & Williams, S. (2003). Producing bad results sections. *The Psychologist, l6,* 644-648.