

6. Univariate Graphing

Video Link:

https://www.youtube.com/watch?v=QIs7LZwhAH8&index=6&list=PL2fQHGEDK7Yyl1W9tglo8wpYFTDumgc_j

Section 6.1: Bar Chart for Categorical Variables

Section 6.2: Histogram for Quantitative Variables

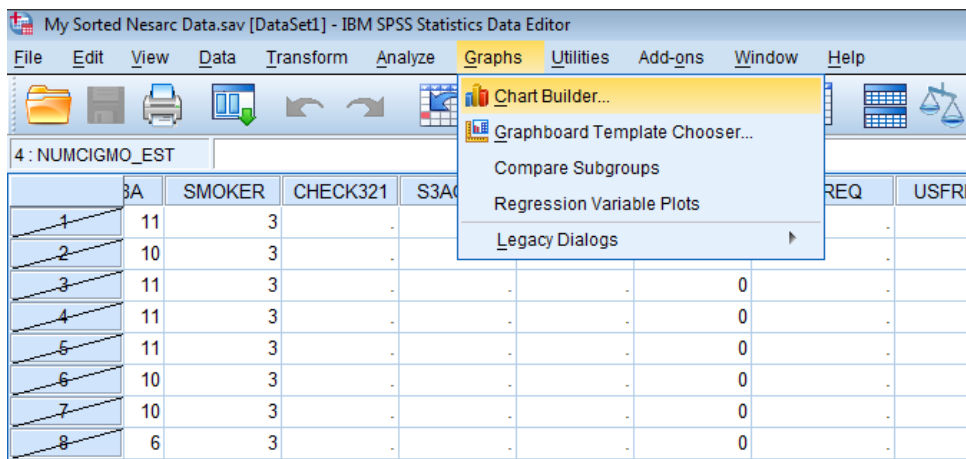
Section 6.3: Additional Descriptive Statistics for Quantitative Variables

Section 6.1: Bar Chart for Categorical Variables

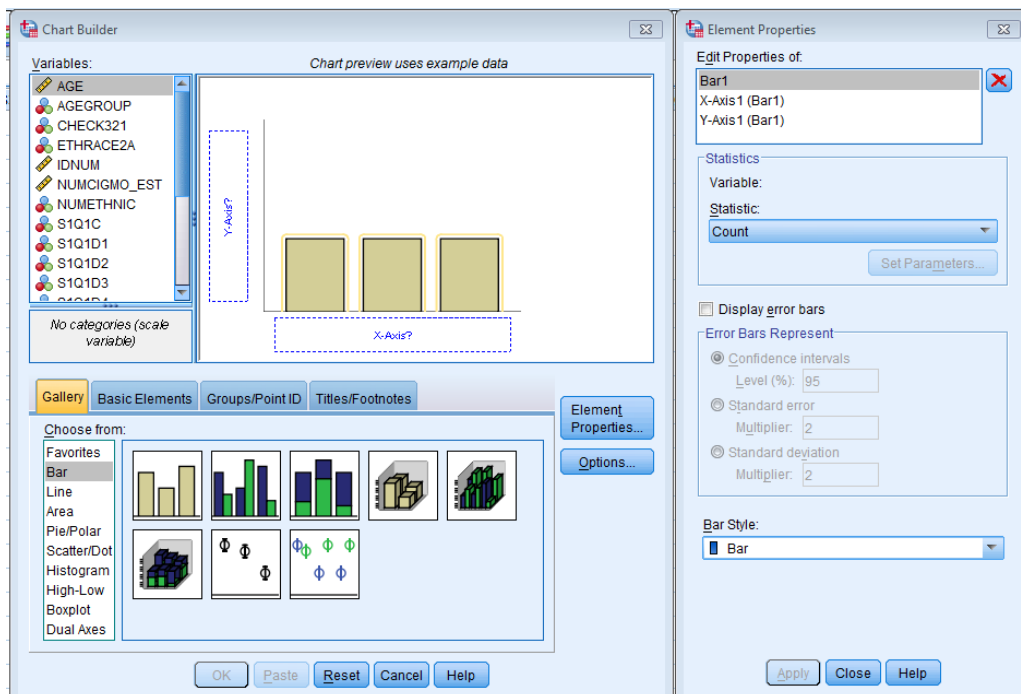
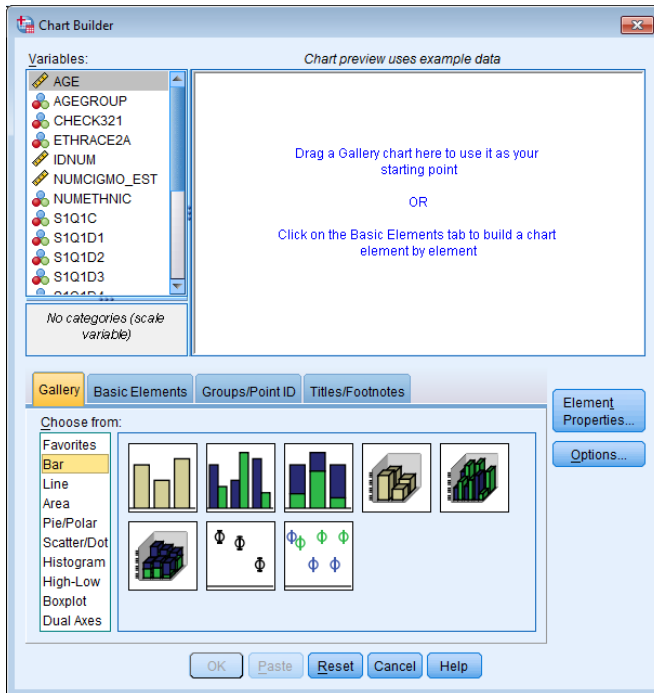
We're going to begin to visualize our variables with graphs. While we start with **graphing** one variable at a time, we'll use this as a springboard for ultimately visualizing multiple variables simultaneously within our graphs. The bar chart is one of the most frequently used graphic visualizations. Categorical variables can be visualized one at a time with univariate graphs, that is, with single variable bar charts.

If you'll recall the data-managed variables of interest in our sample project are now tab12mdx' representing a diagnosis of nicotine dependence in the past 12 months, and NUMCIGMO_EST, representing the average number of cigarettes smoked per month.

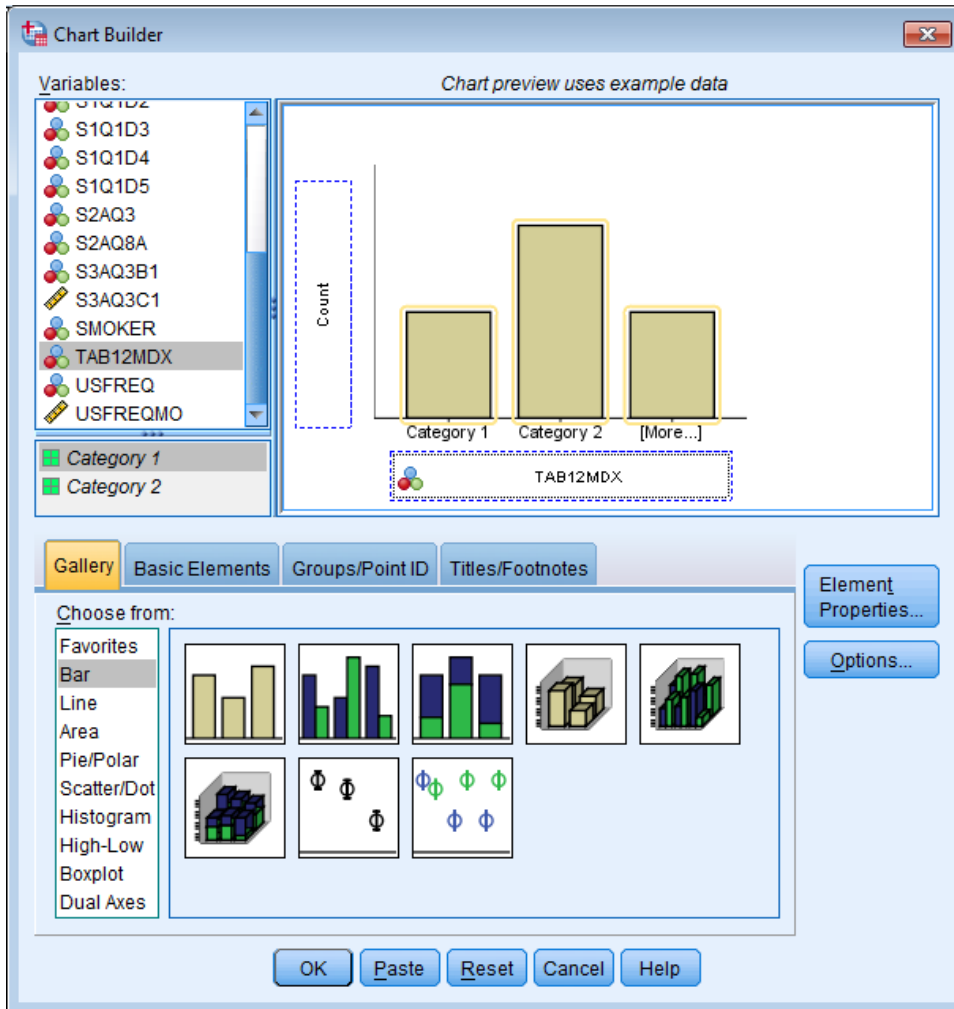
1. To generate a univariate graph click **Graphs > Chart Builder**.



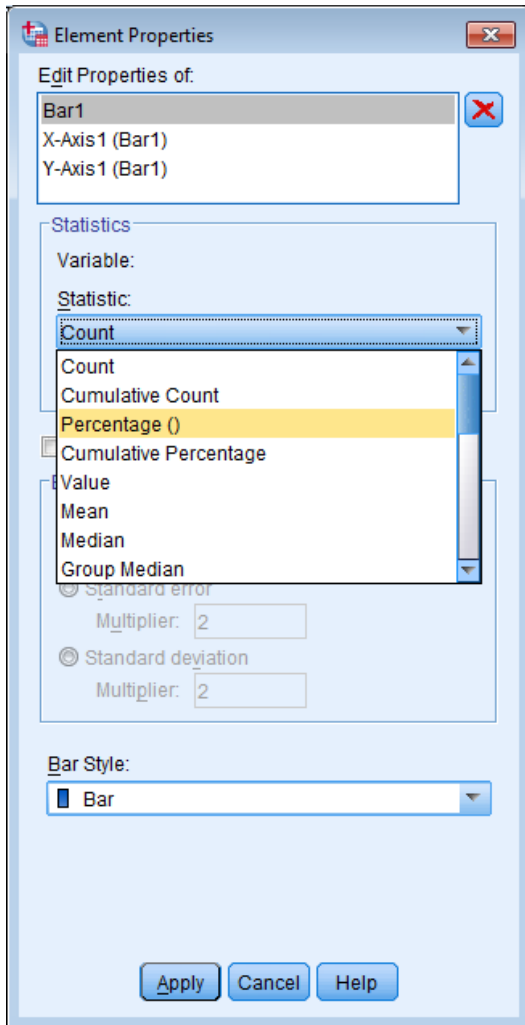
- In the lower portion of the window, directly below **Choose from:** click **Bar**. Drag the top left picture of the **Simple Bar** graph to the **Chart preview** box directly above it.



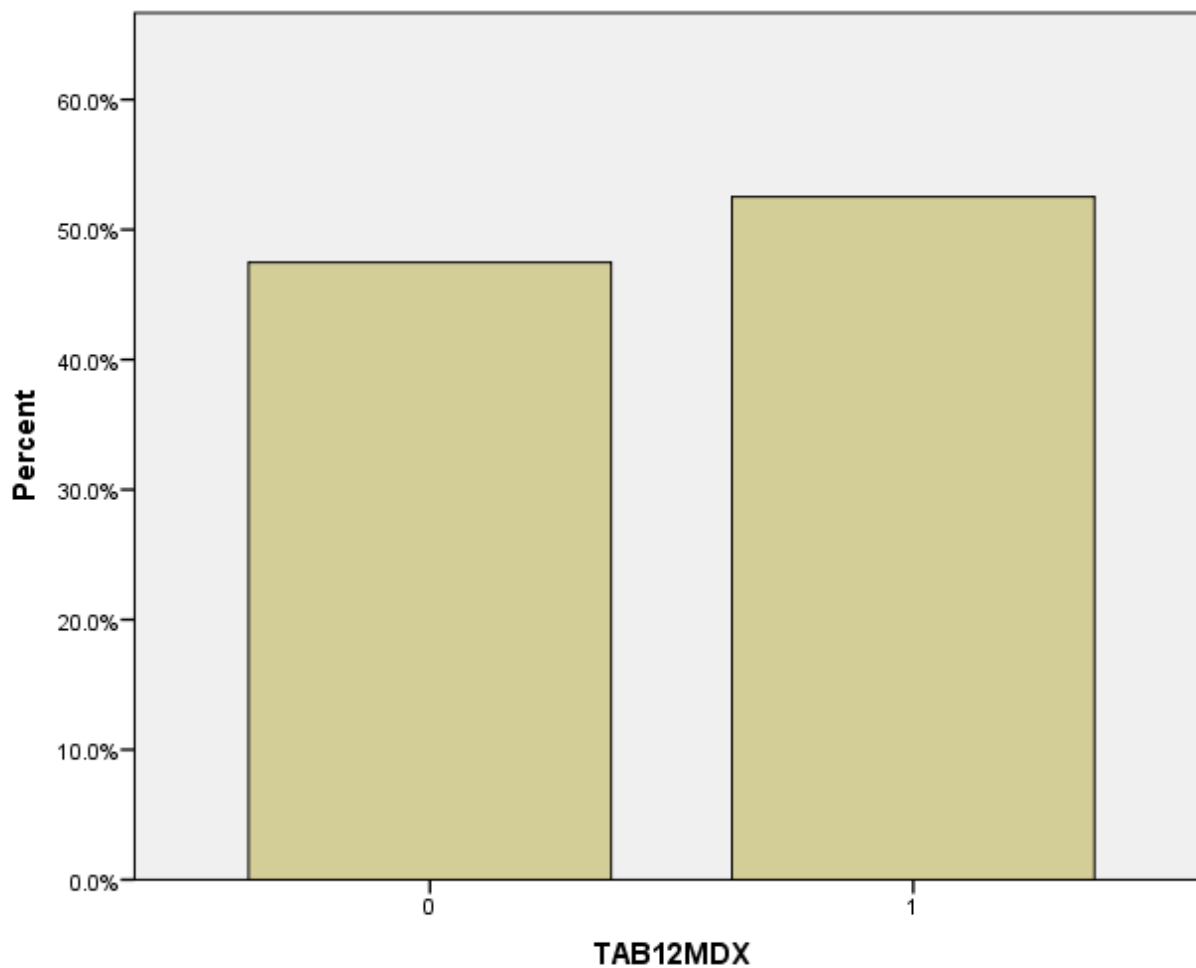
3. Select the variable you would like to graph from the **Variables:** list in the upper left of the window. Drag it to the **X- Axis?** located in the **Chart preview** box.



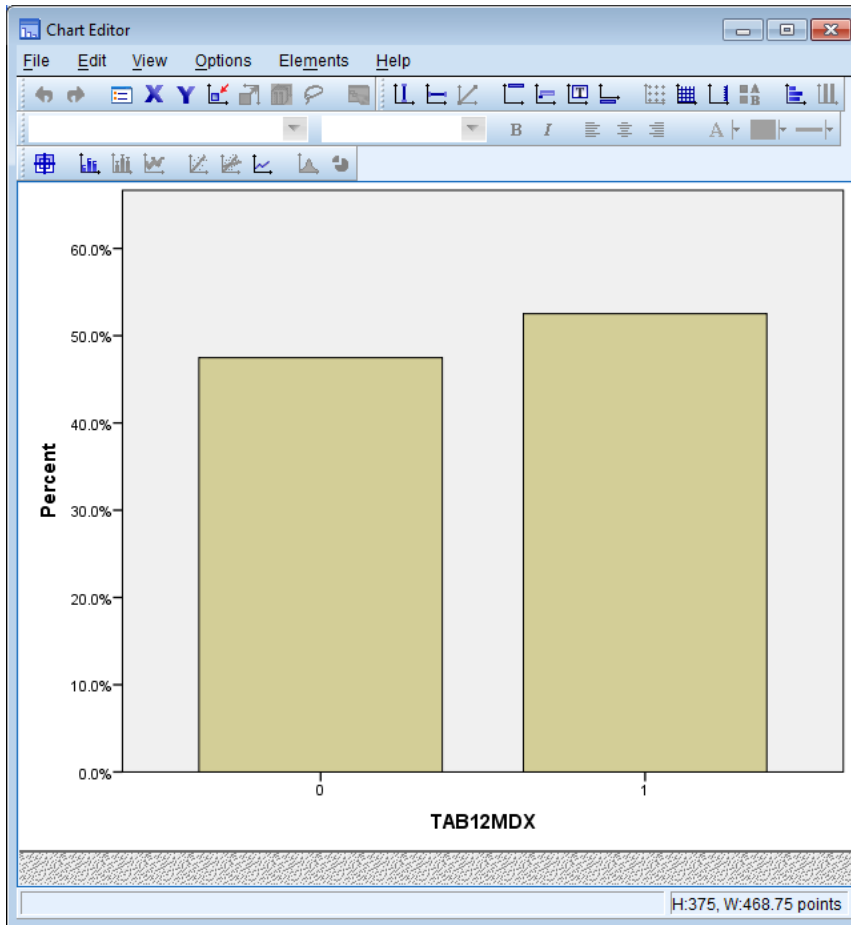
4. Directly right of the **Chart Builder** window is the **Element Properties** window. In the second box click the dropdown menu under **Statistic:**, choose **Percentage(?)**. At the very bottom of the **Element Properties** window Click **Apply**. At the bottom of the **Chart Builder** window Click **OK**. The graph will generate in your output window.



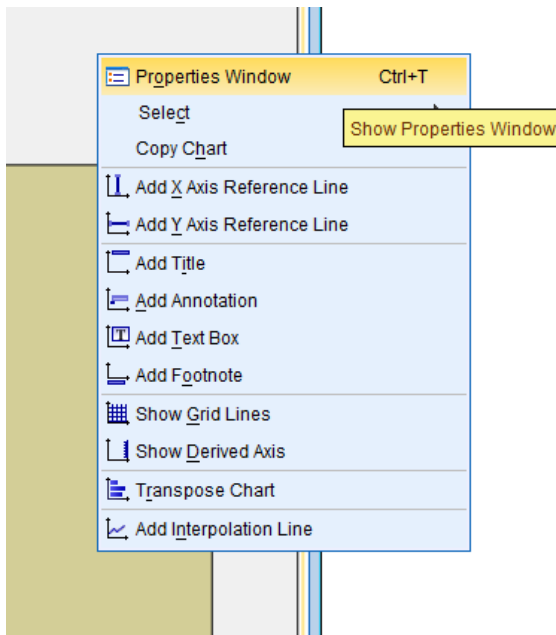
Here is the bar chart we told SPSS to generate. It shows the percent of young adult smokers (our sample) with nicotine dependence, about 52.5%, as indicated by the response code of one, and those without nicotine dependence, about 47.5%, indicated by a zero.



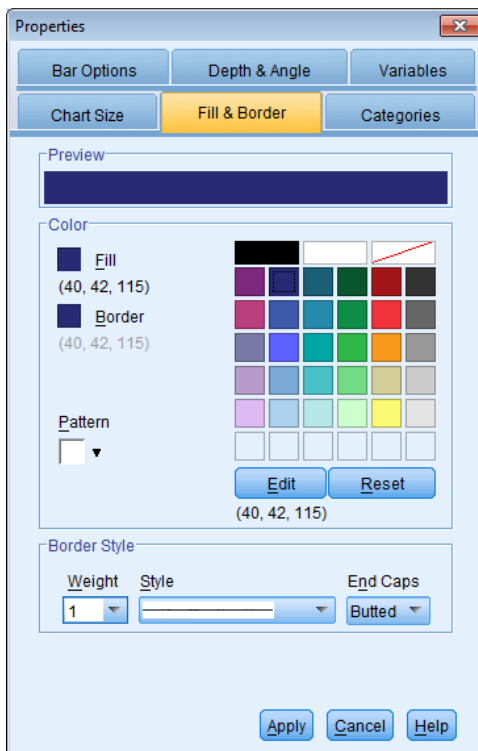
5. In your output window if you double click on the graph you will get the **Chart Editor** that will allow you to customize the univariate graph. If you edit the graph in some way and decide you want to undo it. Click the arrow facing left directly under **File** in the upper left corner.



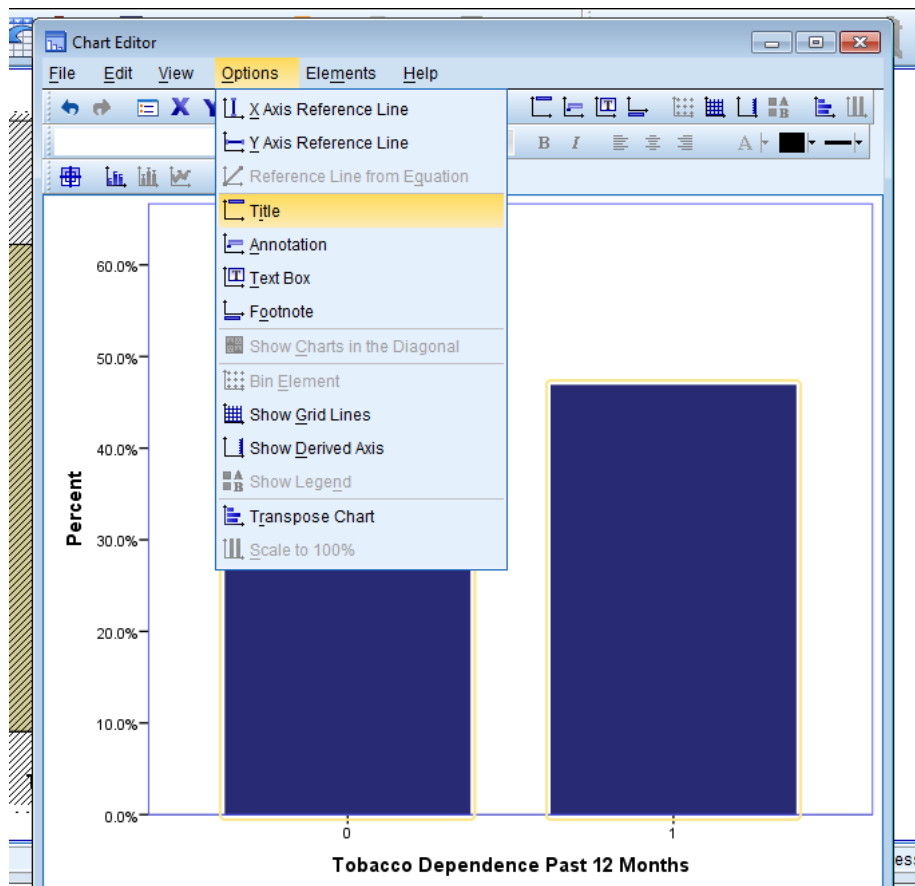
6. If you right click on any part of the graph you will get options to customize that portion of the graph.



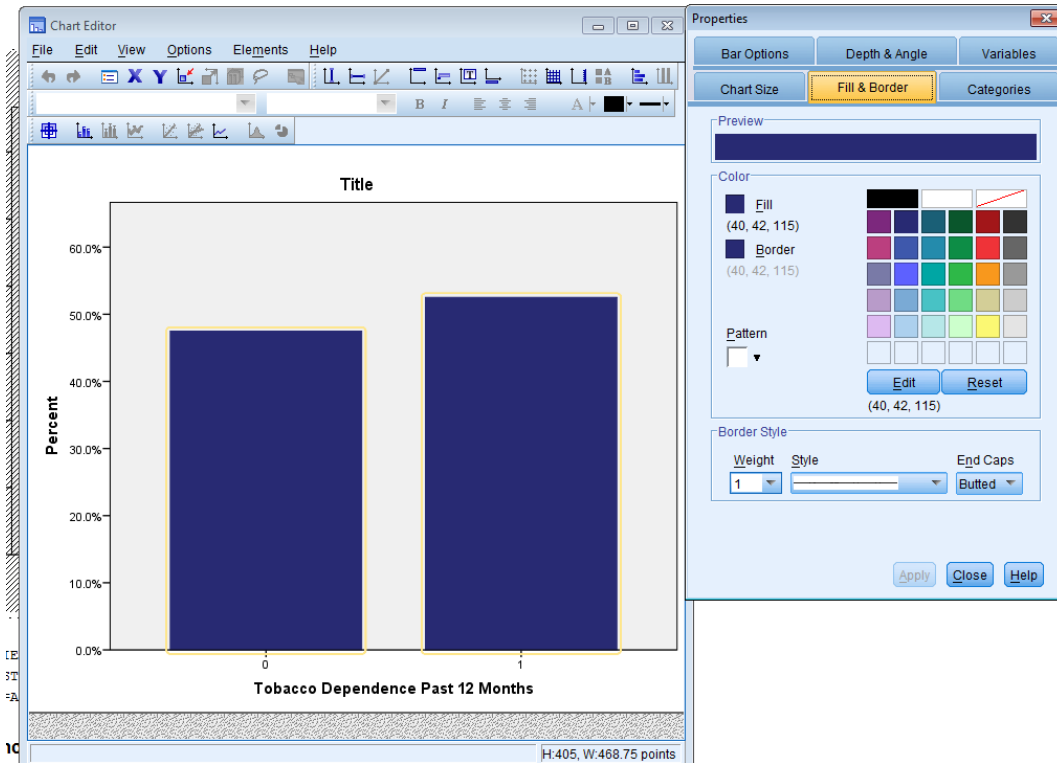
For example, I could right click on one of the bars in the graph, click **Properties Window** then change the **Fill** and **Border** of both bars by select a different color then Click **Apply**. Each of the top tabs allows you to customize a different portion of your graph.



7. It is important to appropriately label any graph. Click **Options > Title**.

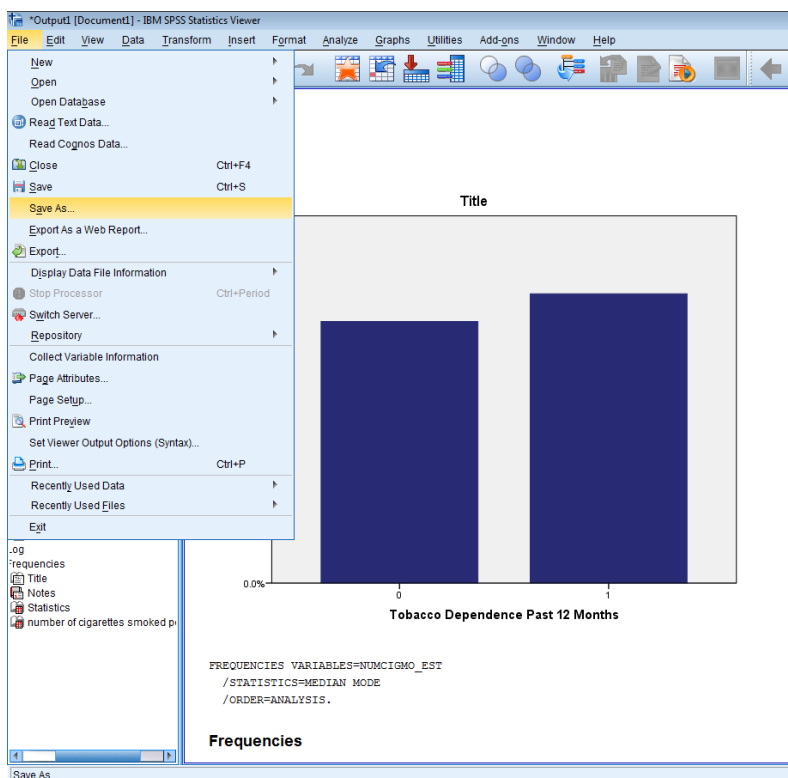


8. Above the graph Click on **Title** and then type an appropriate title for the variable you are graphing.



Once done customizing your graph, close the Chart Editor window. The edited graph will show in your output window. You can come back to the graph at any point to edit it. When time allows explore the different customizable options.

9. Since you have made changes to your graph it is important to save your SPSS Output so you do not lose all those changes. In the upper left corner Click **File > Save As**. Follow normal steps to save a file in the desired location.



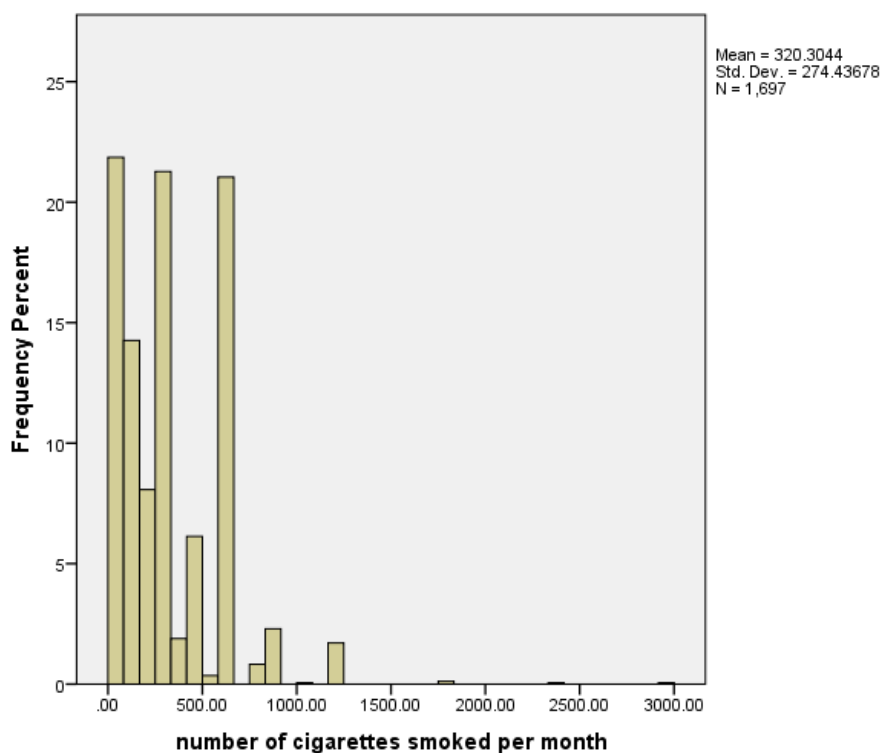
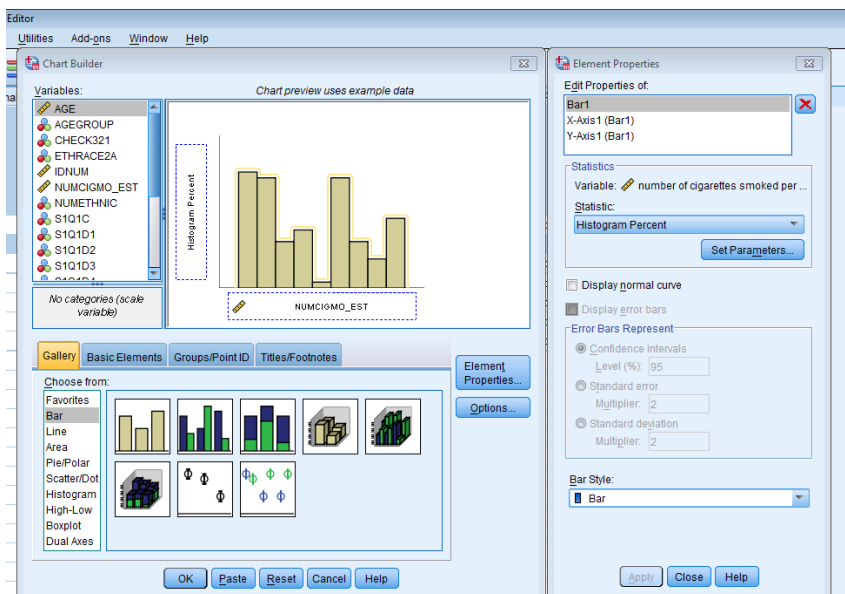
Return to Video - 06. Univariate Graphing

Section 6.2: Histogram for Quantitative Variables

You will see that SPSS generates a graphical distribution of a quantitative variable around its midpoint or median. In a histogram, intervals of values are plotted on the x-axis, rather than discrete or separate values as in a bar chart created for a categorical variable.

1. To generate a Histogram repeat steps 6.1 #1 through #3.

2. Directly right of the **Chart Builder** window is the **Element Properties** window. In the second box click the dropdown menu under **Statistic:**, choose **Histogram Percent**. At the very bottom of the **Element Properties** window Click **Apply**. At the bottom of the **Chart Builder** window Click **OK**. The graph will generate in your output window.

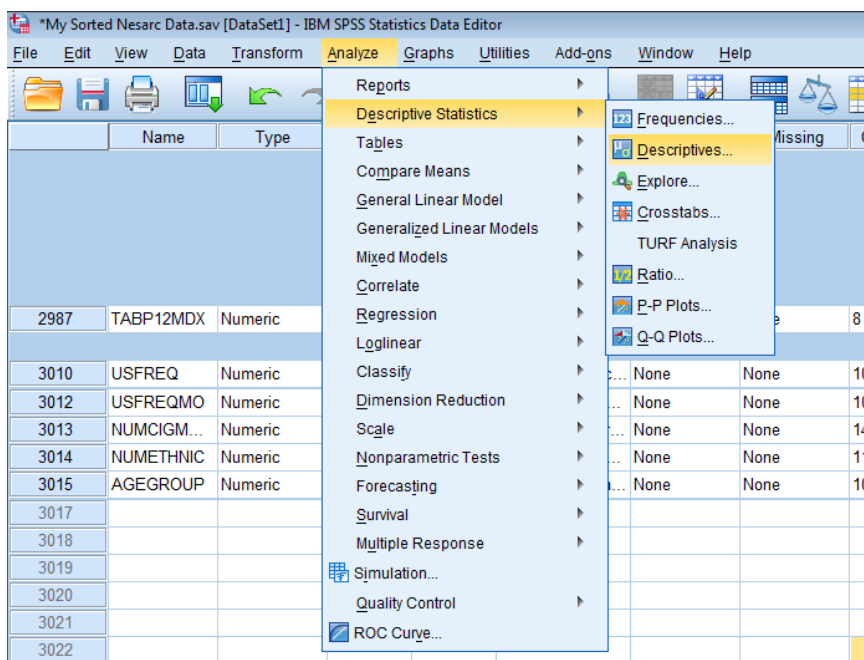


3. Repeat steps 6.1 #5 through #8 to edit your Histogram.

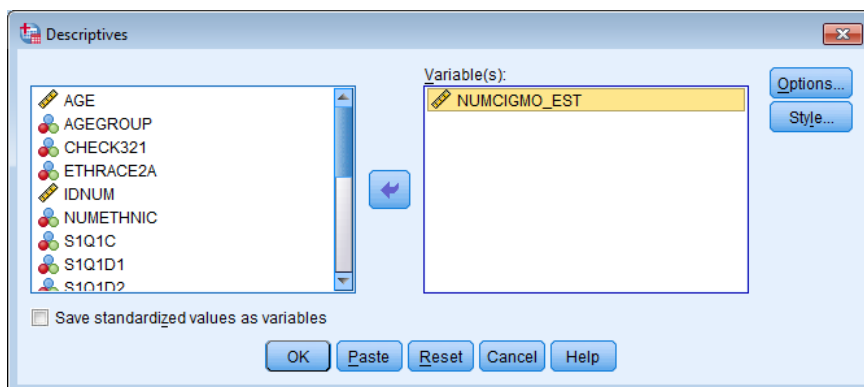
Section 6.3: Descriptive Statistics for Quantitative Variables

For quantitative variables we can also generate a **Descriptive Statistics** table that will display the mean, standard deviation, variance, minimum, maximum, and range. This is not appropriate to do for a categorical variable as the values are typically dummy codes and not meaningful numeric values.

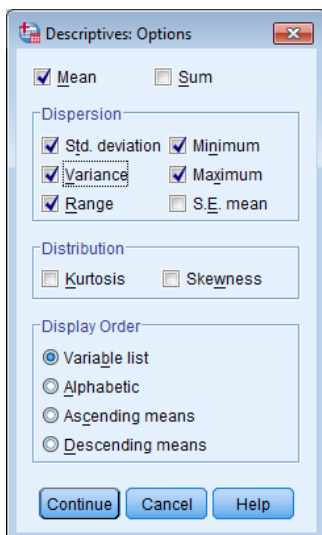
1. Click **Analyze > Descriptive Statistics**.



2. In the left hand window select the variable(s) you want to generate a descriptive statistics table for and move to the right window using the arrow. Click **Options...**



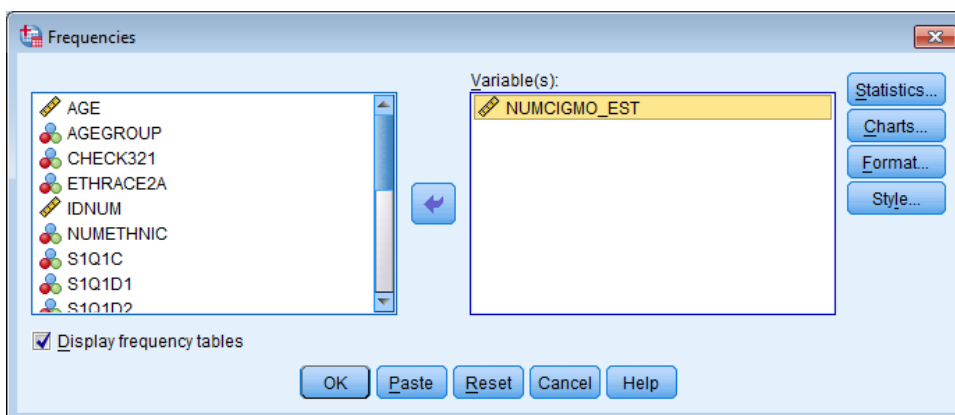
- Confirm that Mean, Std. deviation, Minimum, and Maximum are checked. Check Variance and Range. Click **Continue > OK**.



The descriptive statistics table will generate in your output window.

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
number of cigarettes smoked per month	1697	2939.00	1.00	2940.00	320.3044	274.43678	75315.545
Valid N (listwise)	1697						

- To generate the Mode and Median, Click **Analyze > Descriptive Statistics > Frequencies**. Remove any categorical variables listed under **Variable(s):** in the left hand window using the arrow and move the appropriate quantitative variable(s) to the **Variable(s):** window on the right. Click **Statistics...**



5. Check **Median & Mode** then Click **Continue > OK**. This will generate 2 tables. One with the Median and Mode and then a frequency distribution. To save space the second half of the frequency table was omitted from this tutorial.

Statistics

number of cigarettes smoked per
month

N	Valid	1697
	Missing	9
Median		300.0000
Mode		600.00

number of cigarettes smoked per month

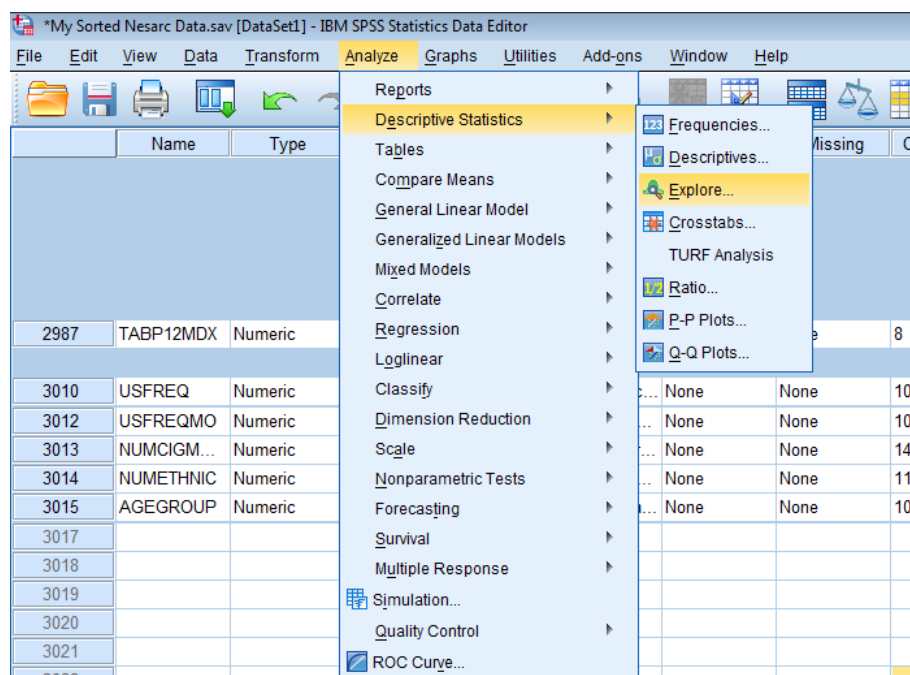
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	29	1.7	1.7	1.7
	2.00	14	.8	.8	2.5
	2.50	11	.6	.6	3.2
	3.00	12	.7	.7	3.9
	4.00	2	.1	.1	4.0
	5.00	34	2.0	2.0	6.0
	6.00	1	.1	.1	6.1
	7.50	12	.7	.7	6.8
	8.00	1	.1	.1	6.8
	10.00	38	2.2	2.2	9.1
	12.50	9	.5	.5	9.6
	14.00	3	.2	.2	9.8
	15.00	14	.8	.8	10.6
	17.50	1	.1	.1	10.7
	20.00	13	.8	.8	11.4
	22.00	4	.2	.2	11.7
	24.00	1	.1	.1	11.7
	25.00	14	.8	.8	12.6
	28.00	17	1.0	1.0	13.6
	30.00	25	1.5	1.5	15.0
	35.00	2	.1	.1	15.1
.	42.00	19	1.1	1.1	16.3
	44.00	9	.5	.5	16.8

Descriptive Statistics are extremely useful in better understanding important characteristics of the cigarettes smoked per month variable.

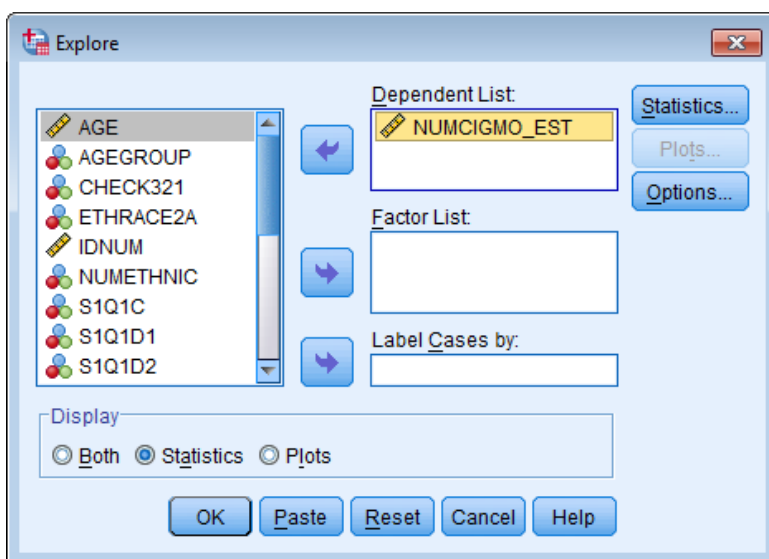
We now know that the young adult smokers in our sample smoke on average 320 cigarettes a month. That the median of the amount smoked, is 300 cigarettes per month. And that the mode, or the most common number of cigarettes smoked per month, is 600. In that the standard deviation is about 274, we can say that, on average, young adult smokers smoked 320 per

month, plus or minus 274 cigarettes. So as you can see, there's an extremely large range in terms of cigarettes smoked, and a lot of variability on this variable.

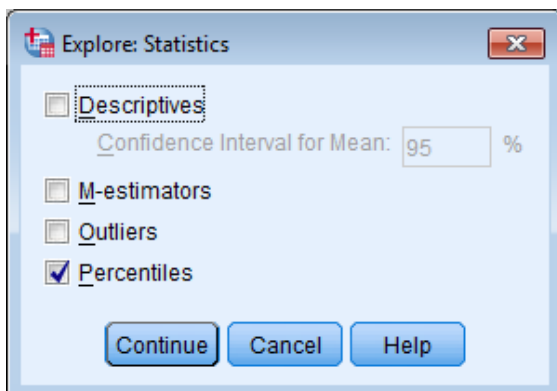
6. To generate the Quantiles, cut points for specific percentiles on a variable, Click **Analyze > Descriptive > Explore**.



7. Move the quantitative variable(s) you are analyzing from the left side window to the **Dependent List:** window in the upper right using the top arrow. At the very bottom (in the **Display** box), **Both** will be clicked, but we need to Click **Statistics** instead. Then Click **Statistics...** in the upper right corner.



8. Uncheck **Descriptives** as we have already generated these calculations. Check **P**ercentiles. Click **C**ontinue > **O**K.



In the table below we want to look at the **Weighted Average** row. The values in that row represent the quantitative value below which a given percentile of observations in a group of observations fall. The given percentiles are listed directly above that row under the word **Percentiles**. So for example, we can see that 50% of the observations fall at or below 300 cigarettes smoked per month.

		Percentiles						
		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	number of cigarettes smoked per month	5.0000	15.0000	90.0000	300.0000	600.0000	600.0000	750.0000
Tukey's Hinges	number of cigarettes smoked per month			90.0000	300.0000	600.0000		

It's very important to remember that most of the univariate statistics are not appropriate to calculate for categorical variables, particularly those that are represented with dummy codes. If you'll recall, the nicotine dependence variable is represented with dummy codes, that is, a yes is indicated with a 1, and a no is indicated with a zero. If we were complete the above steps for the categorical variable of nicotine dependence, tab12mdx, SPSS would still generate univariate tables. However, the statistics wouldn't make any sense. As you can see, we've got a mean and a standard deviation, based on dummy codes.

Categorical variables can often be described well with frequency tables or with a bar chart. For quantitative variables, it's best to examine histograms, and then to supplement these with exact measures of shape, center, and spread.