7. Bivariate Graphing

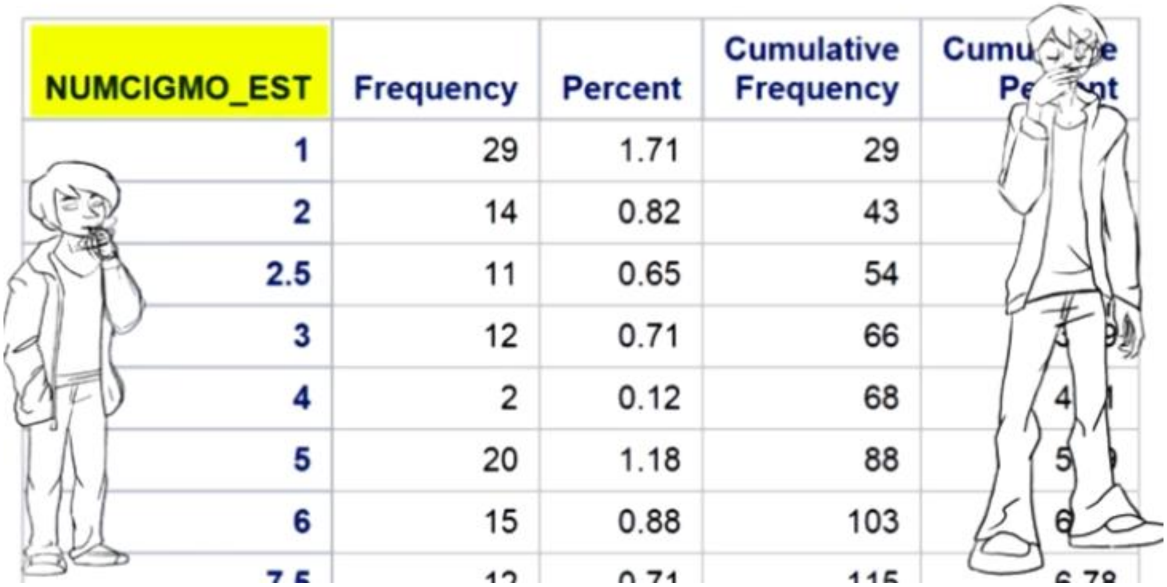
Video Link:
https://www.youtube.com/watch?v=SHZvkWWyguk&index=7&list=PL2fQHGEDK7Yyl1W9tgIo8wpYFTDumgc_j

Section 7.1: Converting a Quantitative Explanatory Variable to Categorical
Section 7.2: Graphing Categorical Explanatory and Categorical Response Relationship
Section 7.3: Graphing Categorical Response Variable with More than 2 Levels
Section 7.4: Graphing Quantitative Explanatory and Quantitative Response Relationship
Section 7.5: Graphing Categorical Explanatory and Quantitative Response Relationship

Section 7.1: Converting a Quantitative Explanatory Variable to Categorical

Since it won't be visually meaningful to examine a bar chart with a quantitative explanatory variable on the y-axis, when our response variable is categorical, before we start to graph, it's important to bin our quantitative explanatory variable into categories. That is, in order to visualize the relationship that we're interested in, we need to add some data management that will allow us to construct a C to C, or categorical to categorical bar chart.

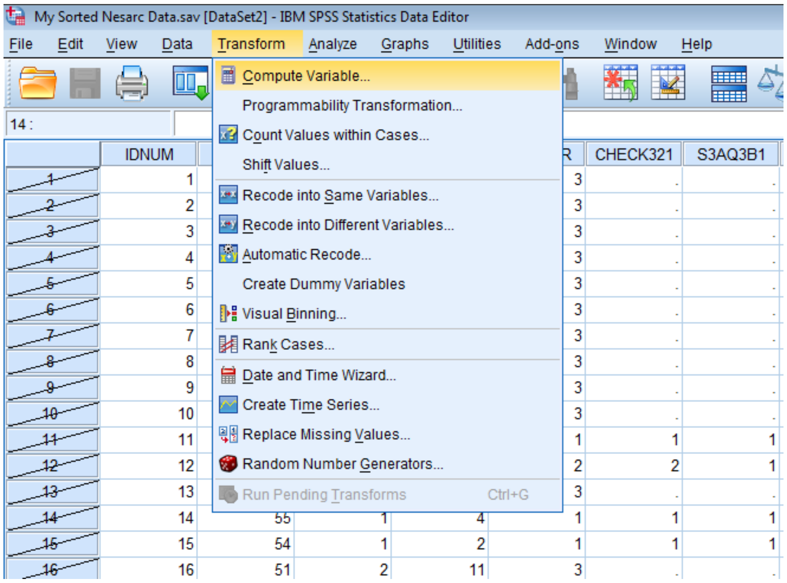
To convert a quantitative variable into a categorical variable we begin by looking at the frequency table for the explanatory variable, number of cigarettes smoked per month.



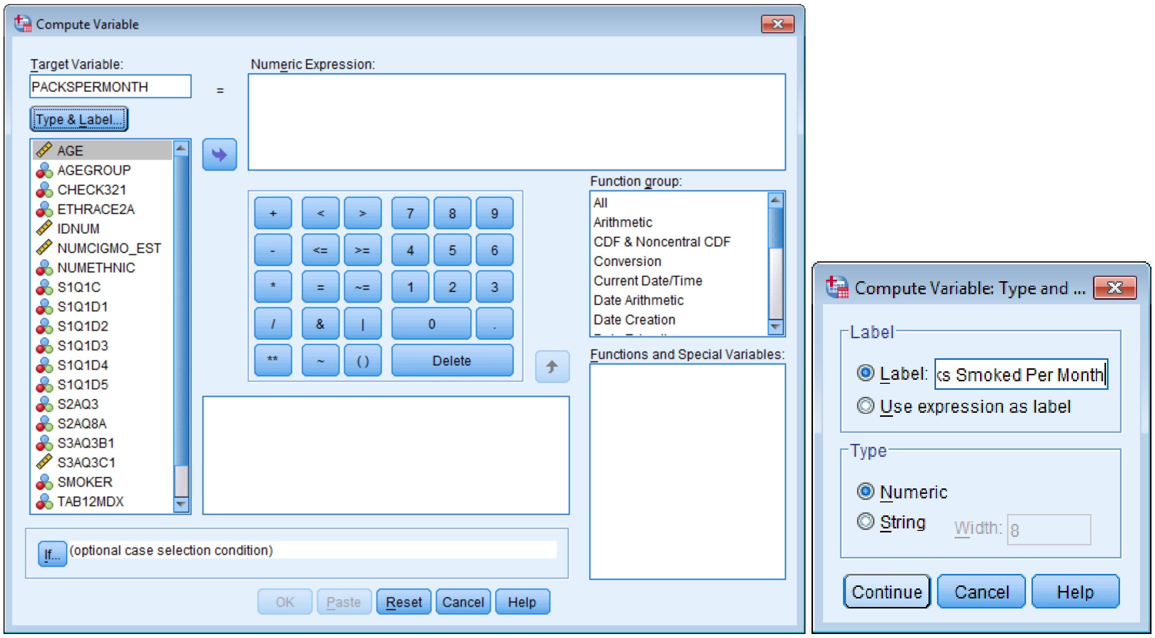
NUMCIGMO_EST	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	29	1.71	29	[illegible]
2	14	0.82	43	[illegible]
2.5	11	0.65	54	[illegible]
3	12	0.71	66	[illegible]
4	2	0.12	68	[illegible]
5	20	1.18	88	[illegible]
6	15	0.88	103	[illegible]
7.5	12	0.71	115	6.78

We could use the cumulative percent column to make decisions about grouping individuals into quartiles--roughly four equal groups in size, or even quintiles--five equal groups in size. However, in this case, it seems a better decision might be to create more meaningful smoking groups based on specific quantities. Cigarette packs contain 20 cigarettes each. We're going to create a new variable that estimates the number of packs that each individual smokes per month, rather than the number of cigarettes. This could be a step closer to a categorical variable that's meaningful.

1. To do this, click **Transform > Compute Variable**.


My Sorted Nesarc Data.sav [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Compute Variable...
Programmability Transformation...
Count Values within Cases...
Shift Values...
Recode into Same Variables...
Recode into Different Variables...
Automatic Recode...
Create Dummy Variables
Visual Binning...
Rank Cases...
Date and Time Wizard...
Create Time Series...
Replace Missing Values...
Random Number Generators...
Run Pending Transforms Ctrl+G
IDNUM
CHECK321
S3AQ3B1

2. Enter the name of the new variable under **Target Variable** on the far upper left hand corner. For our example, let's call this new variable PACKSPERMONTH, which stands for the estimate of the number of packs smoked per month. Click **Type** & **Label**... directly below the variable name you just typed in. Type in an appropriate label for the secondary variable you are creating. Click **Continue**

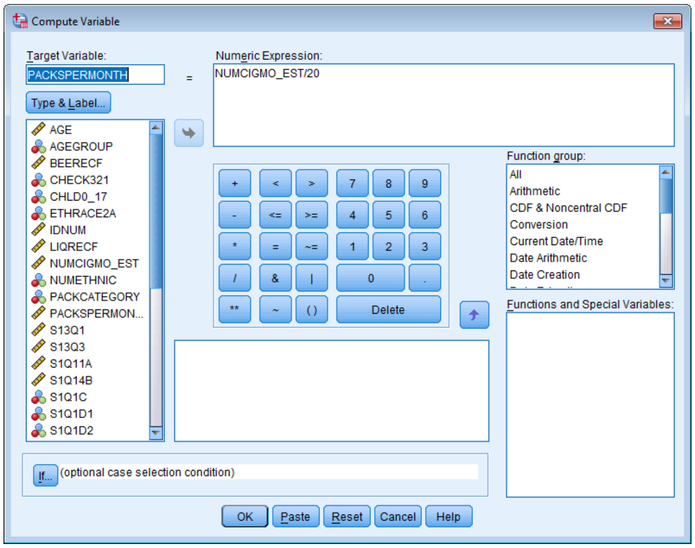

Target Variable: PACKSPERMONTH

Label: ks Smoked Per Month

Type: Numeric

Width: 8

3. In the **Numeric Expression:** window in the upper right insert the NUMCIGMO_EST variable from the left column using the arrow. The new variable is PACKSPERMONTH and it is set equal to the number of cigarettes smoked per month, NUMCIGMO_EST divided by (use a /) 20. Click **OK**.



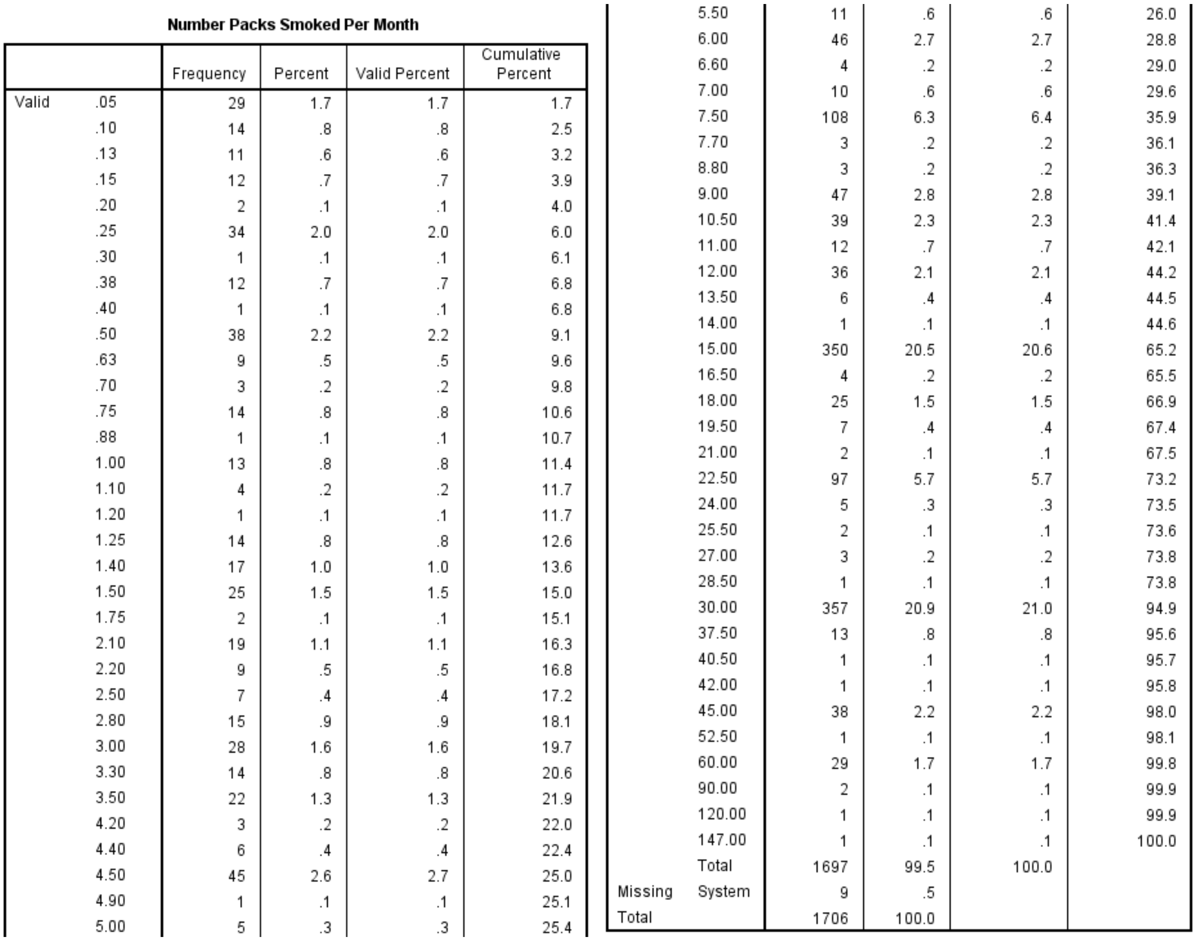
Target Variable: PACKSPERMONTH

Numeric Expression: NUMCIGMO_EST/20

4. Go through the appropriate steps to add this new variable to your **Variable Set**. These steps are shown in 05. Data Management tutorial steps 5.3 #5 through #8.

5. Generate a frequency table for PACKSPERMONTH. Steps are located in .04 Working with Data tutorial steps 4.4 #1 through #3

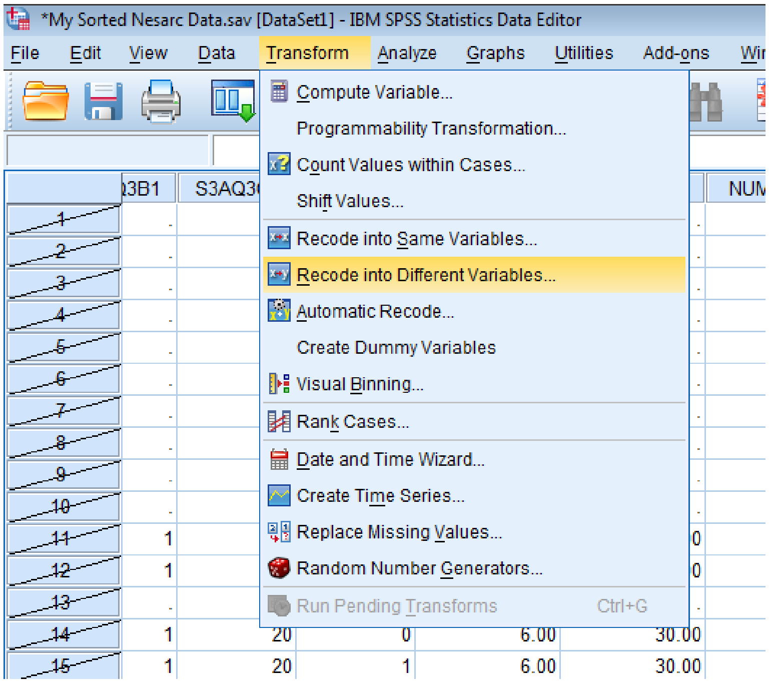
6. PACKSPERMONTH is still a quantitative variable, but now we can more easily create groups based on number of packs smoked in a month. After examining the frequency distribution, we decide to create groupings that include those who've smoked one through five packs per month, six through 10 packs per month, 11 through 20, 21 through 30, and then 30 plus packs per month.



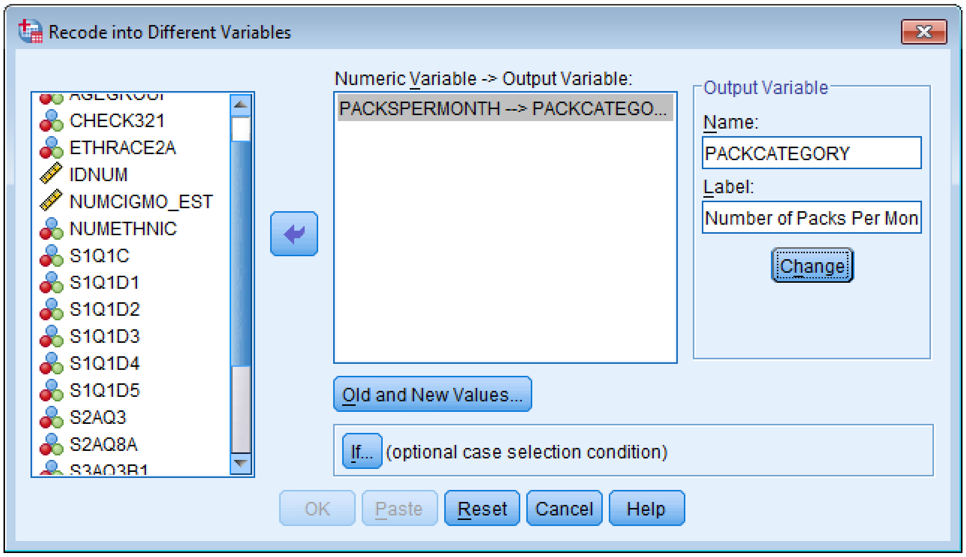
Number Packs Smoked Per Month

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.05	29	1.7	1.7	1.7
	.10	14	.8	.8	2.5
	.13	11	.6	.6	3.2
	.15	12	.7	.7	3.9
	.20	2	.1	.1	4.0
	.25	34	2.0	2.0	6.0
	.30	1	.1	.1	6.1
	.38	12	.7	.7	6.8
	.40	1	.1	.1	6.8
	.50	38	2.2	2.2	9.1
	.63	9	.5	.5	9.6
	.70	3	.2	.2	9.8
	.75	14	.8	.8	10.6
	.88	1	.1	.1	10.7
	1.00	13	.8	.8	11.4
	1.10	4	.2	.2	11.7
	1.20	1	.1	.1	11.7
	1.25	14	.8	.8	12.6
	1.40	17	1.0	1.0	13.6
	1.50	25	1.5	1.5	15.0
	1.75	2	.1	.1	15.1
	2.10	19	1.1	1.1	16.3
	2.20	9	.5	.5	16.8
	2.50	7	.4	.4	17.2
	2.80	15	.9	.9	18.1
	3.00	28	1.6	1.6	19.7
	3.30	14	.8	.8	20.6
	3.50	22	1.3	1.3	21.9
	4.20	3	.2	.2	22.0
	4.40	6	.4	.4	22.4
	4.50	45	2.6	2.7	25.0
	4.90	1	.1	.1	25.1
	5.00	5	.3	.3	25.4
	5.50	11	.6	.6	26.0
	6.00	46	2.7	2.7	28.8
	6.60	4	.2	.2	29.0
	7.00	10	.6	.6	29.6
	7.50	108	6.3	6.4	35.9
	7.70	3	.2	.2	36.1
	8.80	3	.2	.2	36.3
	9.00	47	2.8	2.8	39.1
	10.50	39	2.3	2.3	41.4
	11.00	12	.7	.7	42.1
	12.00	36	2.1	2.1	44.2
	13.50	6	.4	.4	44.5
	14.00	1	.1	.1	44.6
	15.00	350	20.5	20.6	65.2
	16.50	4	.2	.2	65.5
	18.00	25	1.5	1.5	66.9
	19.50	7	.4	.4	67.4
	21.00	2	.1	.1	67.5
	22.50	97	5.7	5.7	73.2
	24.00	5	.3	.3	73.5
	25.50	2	.1	.1	73.6
	27.00	3	.2	.2	73.8
	28.50	1	.1	.1	73.8
	30.00	357	20.9	21.0	94.9
	37.50	13	.8	.8	95.6
	40.50	1	.1	.1	95.7
	42.00	1	.1	.1	95.8
	45.00	38	2.2	2.2	98.0
	52.50	1	.1	.1	98.1
	60.00	29	1.7	1.7	99.8
	90.00	2	.1	.1	99.9
	120.00	1	.1	.1	99.9
	147.00	1	.1	.1	100.0
	Total	1697	99.5	100.0	
Missing	System	9	.5		
Total		1706	100.0		

7. Click **Transform > Recode into Different Variables…**


*My Sorted Nesarc Data.sav [DataSet1] - IBM SPSS Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Wi
Compute Variable...
Programmability Transformation...
Count Values within Cases...
Shift Values...
Recode into Same Variables...
Recode into Different Variables...
Automatic Recode...
Create Dummy Variables
Visual Binning...
Rank Cases...
Date and Time Wizard...
Create Time Series...
Replace Missing Values...
Random Number Generators...
Run Pending Transforms Ctrl+G

8. Move PACKSPERMONTH to the middle window under **Numeric Variable->Output Variable**:, **Name** and **Label** it, Click **Old and New Values** below the middle window.



Numeric Variable -> Output Variable:
PACKSPERMONTH --> PACKCATEGO...

Output Variable

Name: PACKCATEGORY

Label: Number of Packs Per Mon

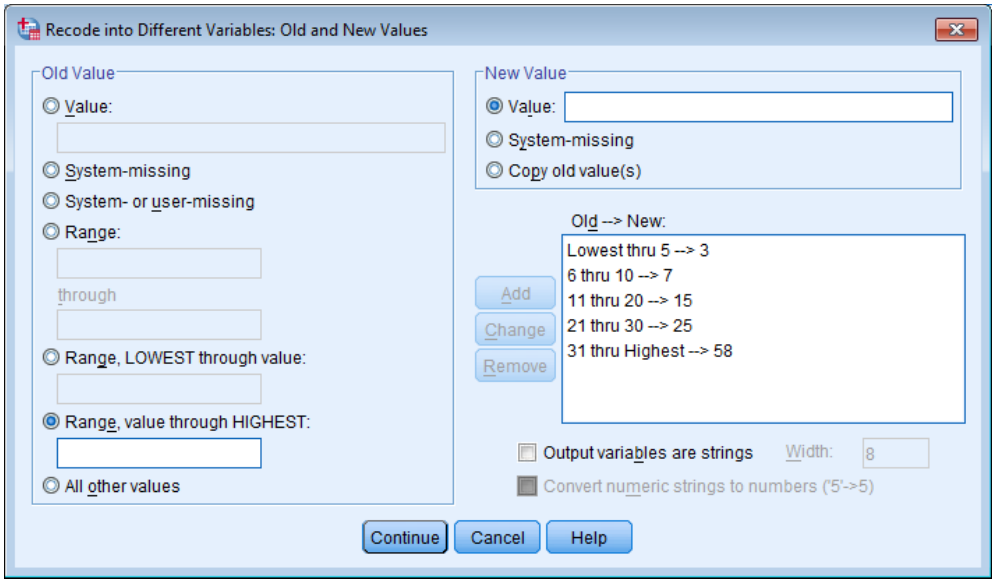
9. Under **Old Value use Range:** to bin PACKSPERMONTH for the following ranges and dummy codes.

   6-10=7
   11-20=15
   21-30=25

   Use Range, Lowest through value: for
   5 or less =3

   Use Range, value through Highest: for
   31 or more = 58

   Click **Continue > OK**.



Recode into Different Variables: Old and New Values

Old Value

○ Value:

○ System-missing

○ System- or user-missing

○ Range:

through

○ Range, LOWEST through value:

● Range, value through HIGHEST:

○ All other values

New Value

● Value:

○ System-missing

○ Copy old value(s)

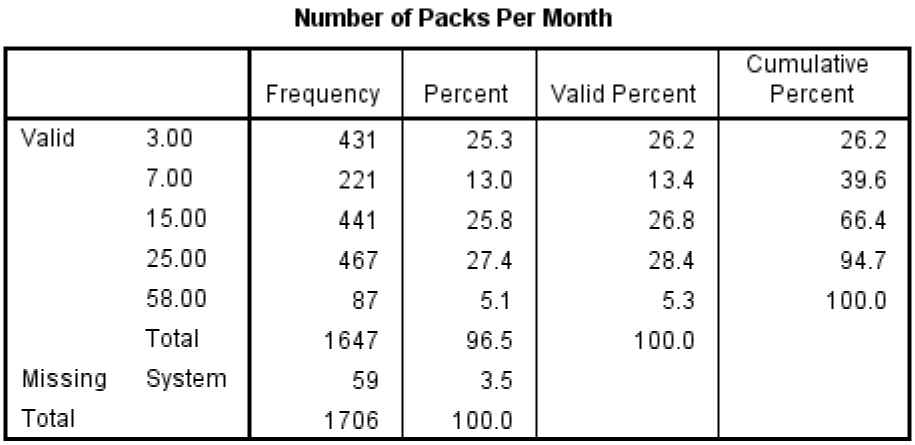
Old --> New:

Lowest thru 5 --> 3
6 thru 10 --> 7
11 thru 20 --> 15
21 thru 30 --> 25
31 thru Highest --> 58

□ Output variables are strings Width: 8

□ Convert numeric strings to numbers ('5'->5)

10. Complete the appropriate steps to add the new variable PACKSCATEGORY to your **Variable Set**.



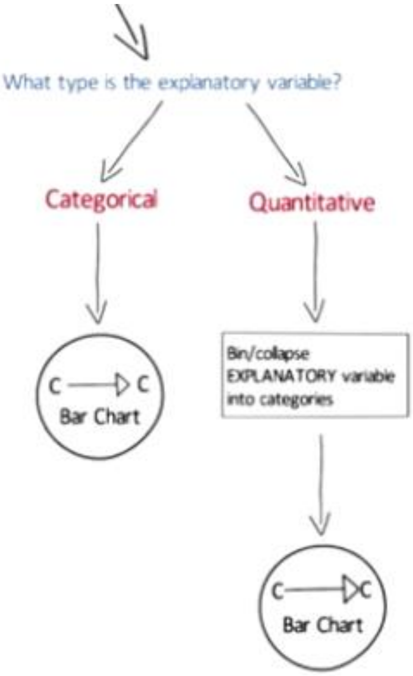
Number of Packs Per Month

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	431	25.3	26.2	26.2
	7.00	221	13.0	13.4	39.6
	15.00	441	25.8	26.8	66.4
	25.00	467	27.4	28.4	94.7
	58.00	87	5.1	5.3	100.0
	Total	1647	96.5	100.0	
Missing	System	59	3.5		
Total		1706	100.0		

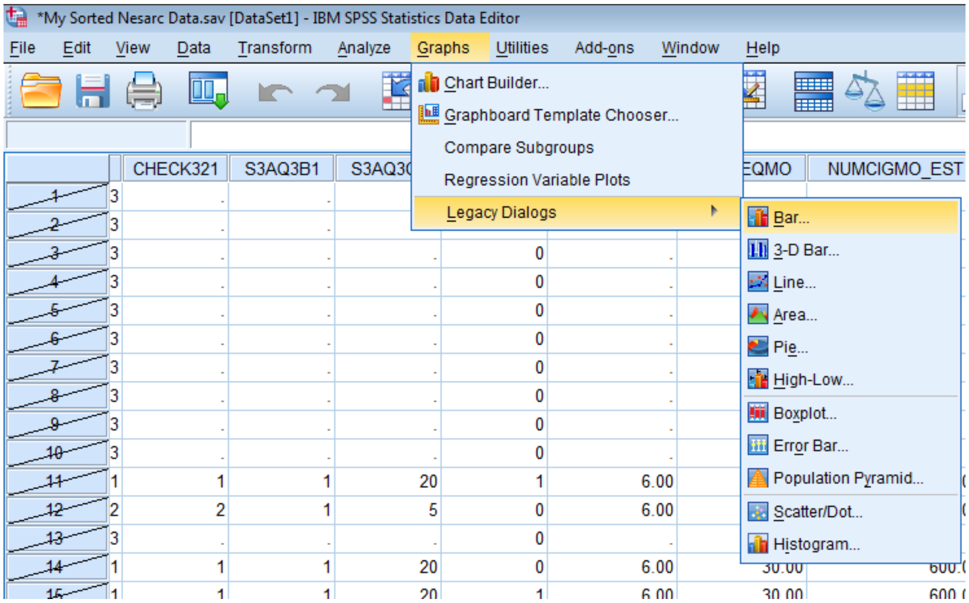
With this new categorical variable representing packs of cigarettes smoked per month, we've retained as much of the quantitative features of the original variable we could manage, while also assuring the graph will be interpretable, now that the explanatory variable is categorical.

Section 7.2 Graphing Categorical Explanatory and Categorical Response Relationships

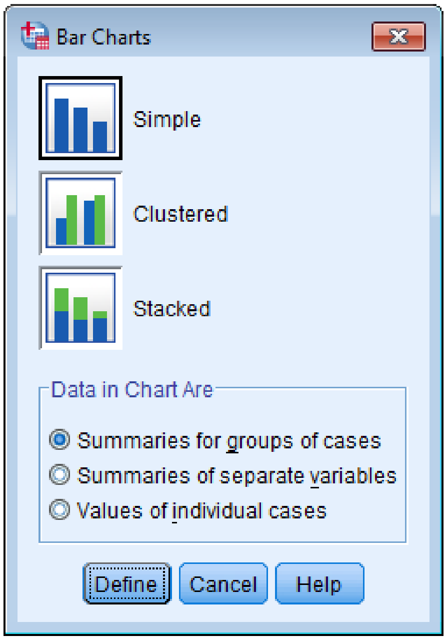
Now that we've collapsed our explanatory quantitative variable into categories, we're ready to make our C to C, or category to category bar chart.


What type is the explanatory variable?
Categorical
Quantitative
C C
Bar Chart
Bin/collapse EXPLANATORY variable into categories
C C
Bar Chart

1. Click **Graphs > Legacy Dialogs**



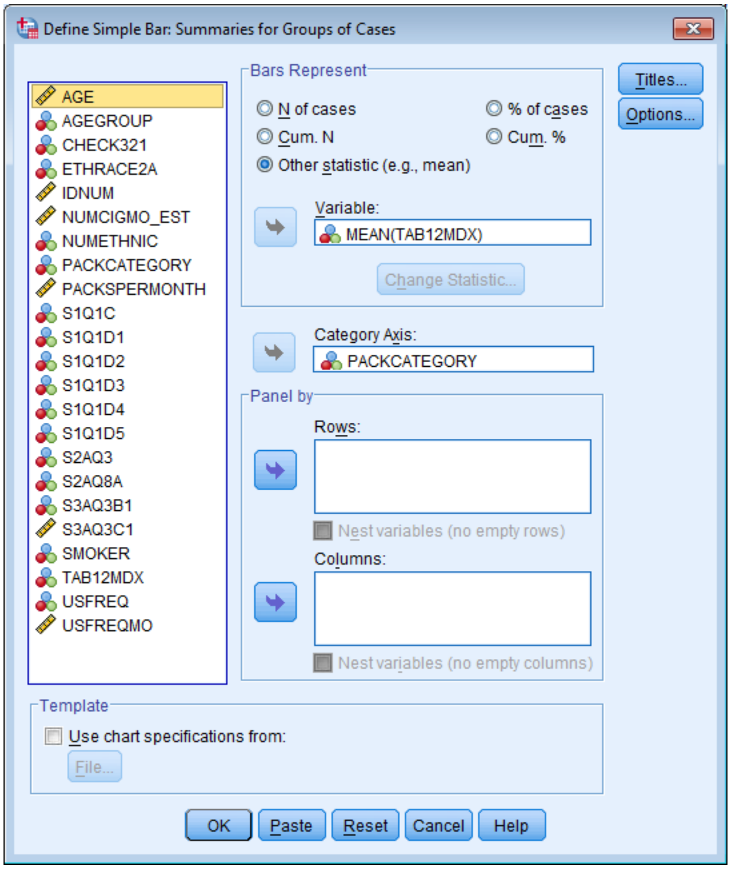
2. Click on the graph left of **Simple > Define.**



Data in Chart Are

◉ Summaries for groups of cases
○ Summaries of separate variables
○ Values of individual cases

3. In the top middle **Bars Represent** box Click **Other statistic (e.g., mean)**. Using the arrow directly below, move the **Categorical Response Variable** from the left window to below **Variable:**. Use the next arrow down to move the Categorical Explanatory Variable to the **Category Axis:** window. Click **Titles…** in the upper right corner.



Define Simple Bar: Summaries for Groups of Cases

Bars Represent

○ N of cases ○ % of cases
○ Cum. N ○ Cum. %
● Other statistic (e.g., mean)

Variable: MEAN(TAB12MDX)

Category Axis: PACKCATEGORY

Panel by

Rows:

□ Nest variables (no empty rows)

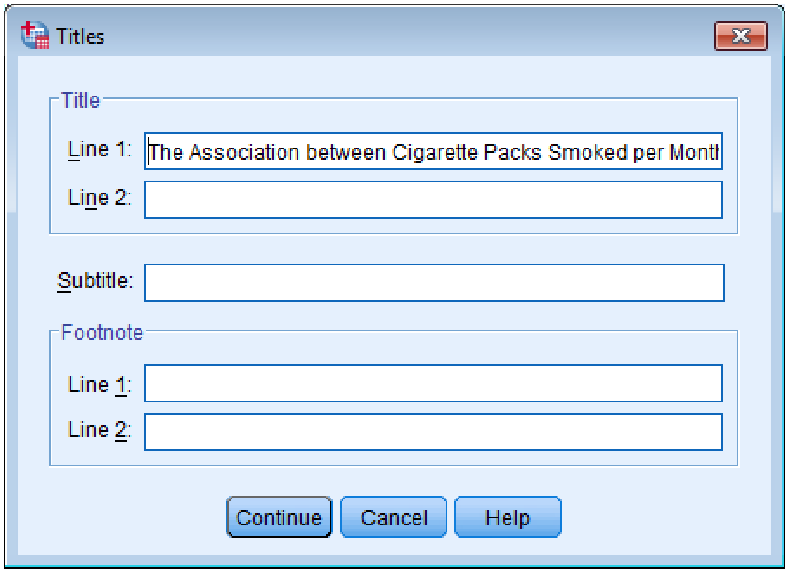
Columns:

□ Nest variables (no empty columns)

Template

□ Use chart specifications from:

4. In the top window, **Line 1:**, appropriately title your graph. Click **Continue > OK**.



Titles

Title

Line 1: The Association between Cigarette Packs Smoked per Montl

Line 2:

Subtitle:

Footnote

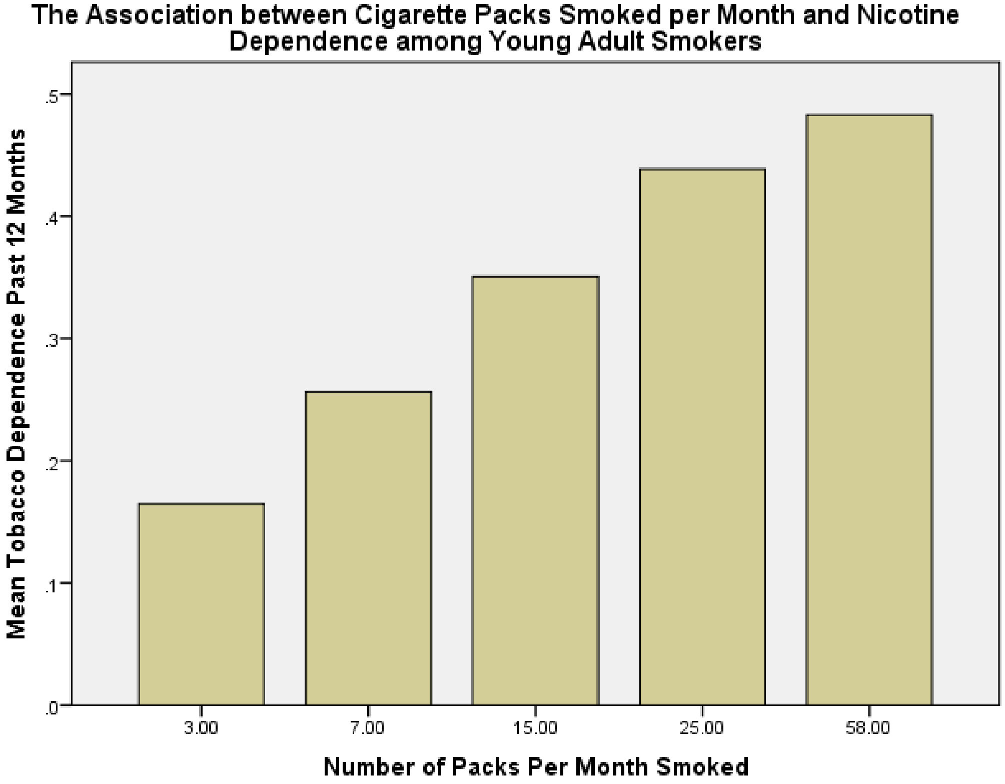
Line 1:

Line 2:

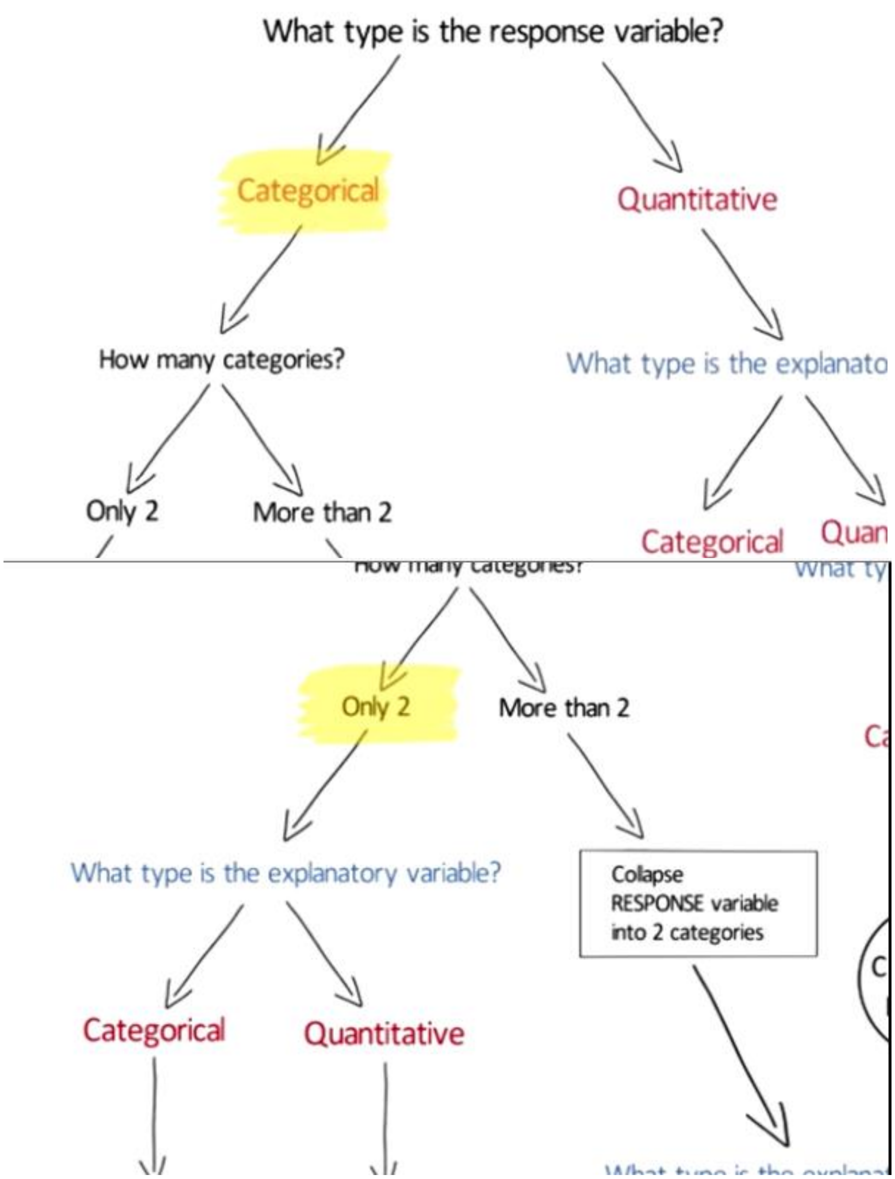
Here is our categorical by categorical bar chart. PACKCATEGORY, our explanatory variable, is on the x-axis. And this is by the rate or proportion of nicotine dependence along the y-axis. So you can see from this graph, among those smoking one to five packs a month, about 25% of those individuals are nicotine dependent.

Among those smoking six to ten packs a month, 50% are nicotine dependent. Among those smoking 11 to 20 packs a month, 58% are nicotine dependent. Among those smoking 21 to 30 packs per month, almost 70% are nicotine dependent. And among those smoking more than 30 packs a month, more than 70% are nicotine dependent, around 77% here.

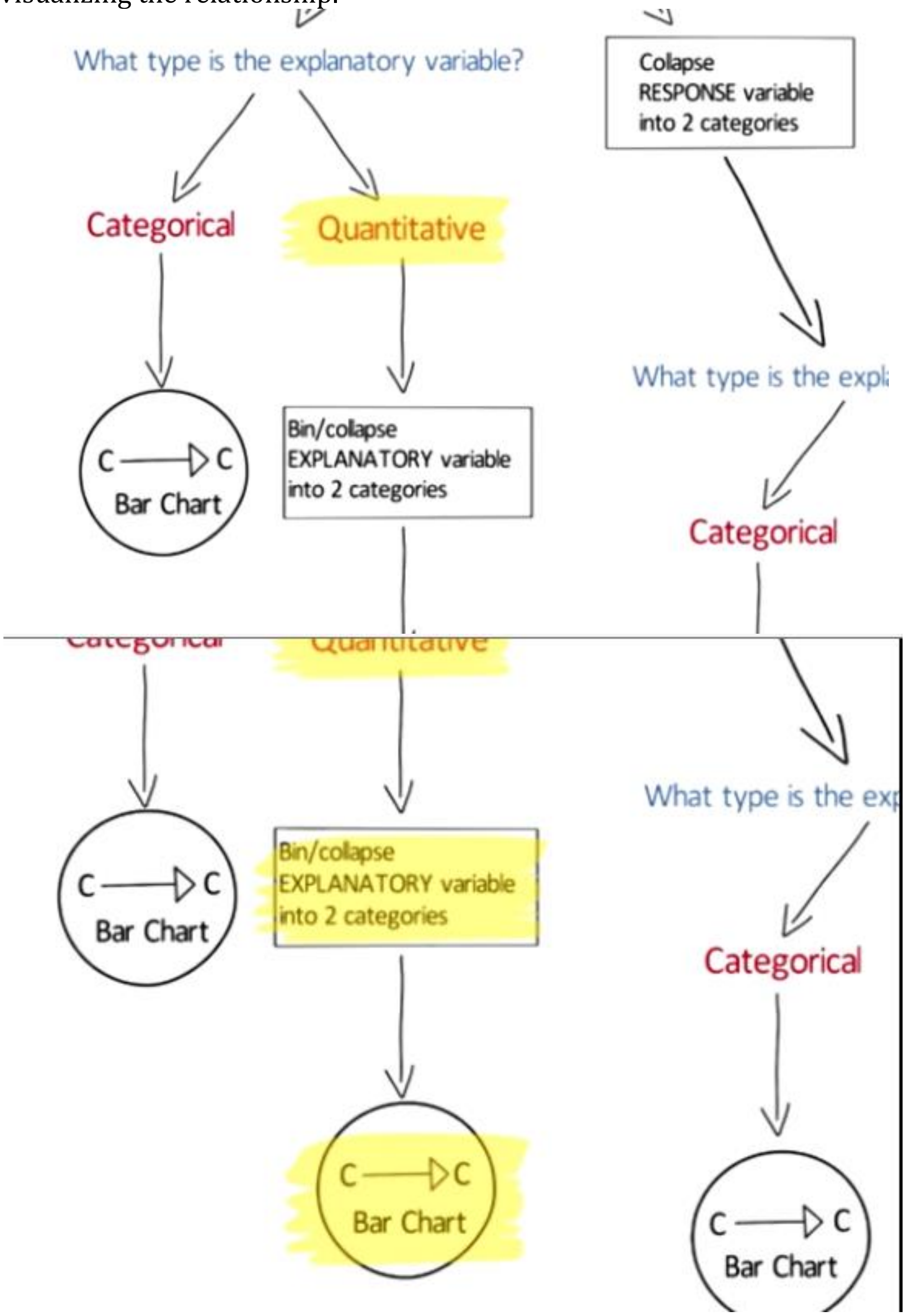
We can also see that these rates form a pattern. That is, the more packs smoked per month, the higher the rate of nicotine dependence. So, in a graphical way, we're already seeing that there seems to be a relationship between smoking and nicotine dependence, as we hypothesized.



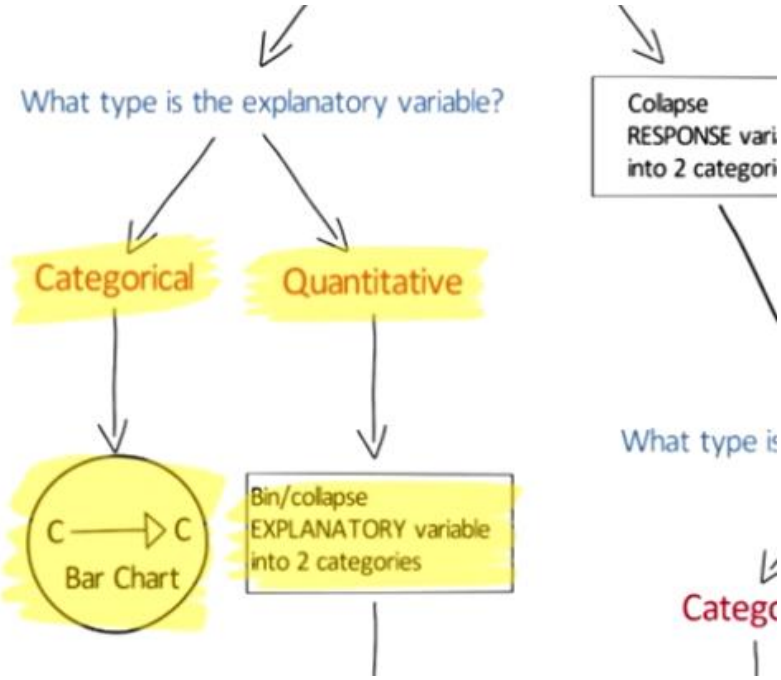
Looking at our graphing decisions chart, we can see the steps we've taken to generate a bivariate graph with a categorical response variable that has two categories



and a quantitative explanatory variable. We also discussed how to convert the quantitative explanatory variable to a categorical variable--a step which must be taken for the purposes of visualizing the relationship.

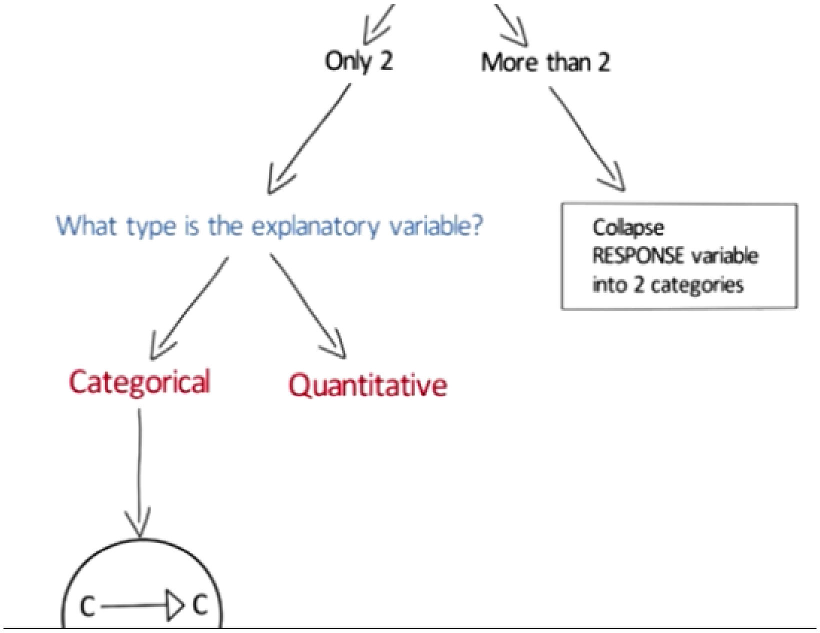

What type is the explanatory variable?
Collapse RESPONSE variable into 2 categories
Categorical
Quantitative
C → C Bar Chart
Bin/collapse EXPLANATORY variable into 2 categories
What type is the expl
Categorical
Categorical
Quantitative
C → C Bar Chart
Bin/collapse EXPLANATORY variable into 2 categories
What type is the exp
Categorical
C → C Bar Chart
C → C Bar Chart

If our explanatory variable was originally categorical rather than quantitative, we could have skipped this step and just moved on to a categorical by categorical bar chart.


What type is the explanatory variable?
Categorical
Quantitative
C → C
Bar Chart
Bin/collapse EXPLANATORY variable into 2 categories
Collapse RESPONSE vari into 2 categori
What type is
Categ

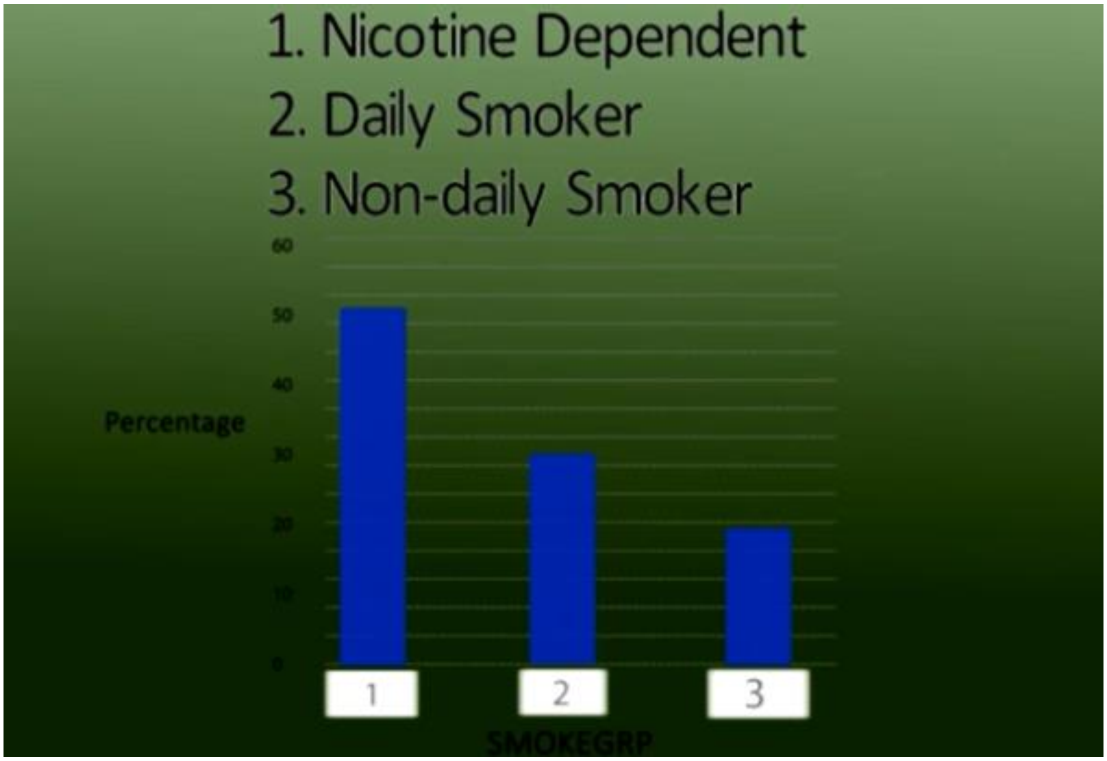
Section 7.3: Graphing Categorical Response Variable with More than 2 Levels

What decisions need to be made if the response variable has more than two categories? In this case, we would need to collapse our response variable categories into two categories. To demonstrate this, we'll have to modify the research question.


Only 2
More than 2
What type is the explanatory variable?
Collapse RESPONSE variable into 2 categories
Categorical
Quantitative
C → C

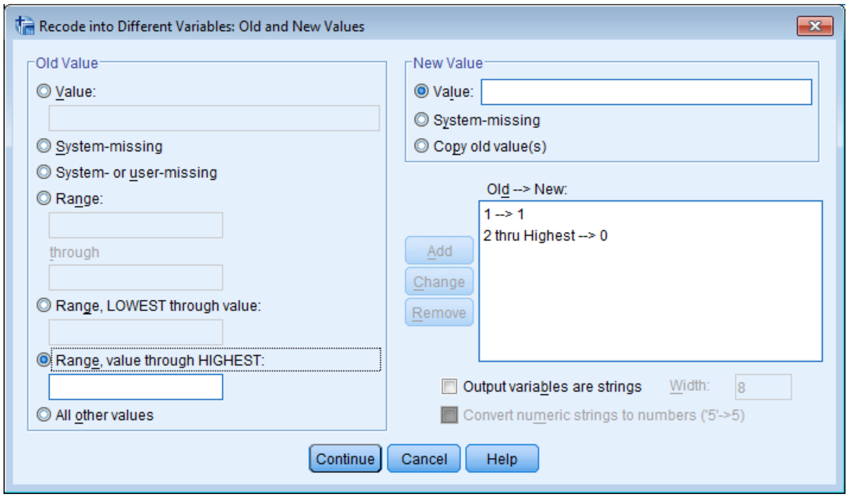
Let's modify the research question to look at the association between ethnicity and smoking stage. We will create a response variable that categorizes young adult smokers into three groups: non-daily smokers, daily smokers, and those with nicotine dependence.

This sample can be described with these three smoking categories. This univariate bar chart shows that about 50% of the young adults sampled are nicotine dependent. About 30% are daily smokers without nicotine dependence. And almost 17% are non-daily smokers.



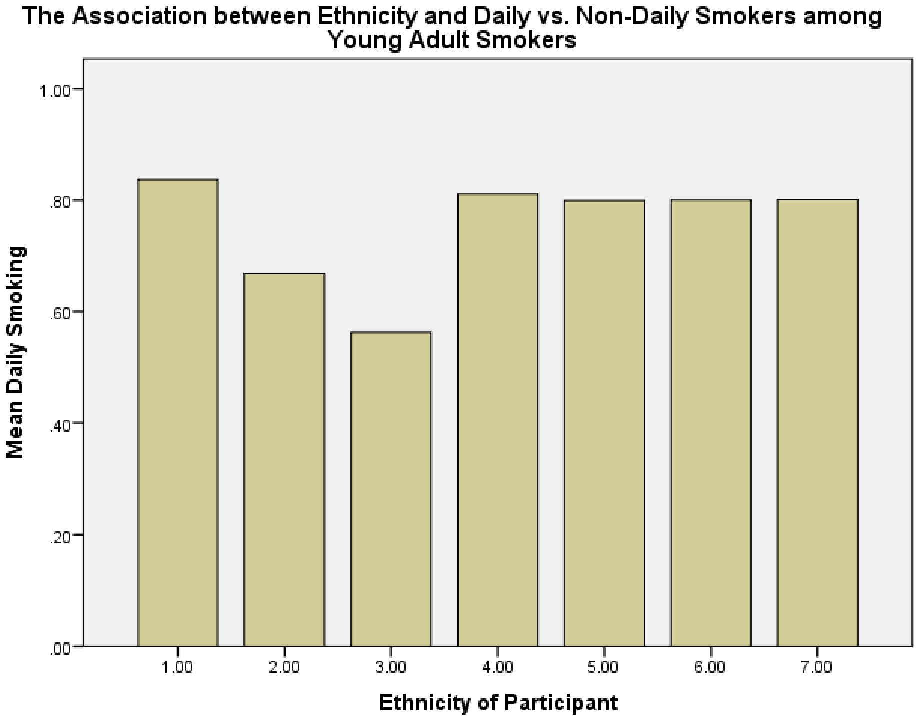
However, to examine a relationship between this variable as my response variable and another, we need to collapse this to only two categories. To do this we'd need to make some decisions. Here are two perfectly reasonable decisions that we could make. We could examine the association between ethnicity and daily versus non-daily smokers, or we could examine the association between ethnicity and nicotine dependent versus non-nicotine dependent individuals thereby collapsing across these categories in some way. In either case, some data management needs to be added to the program.

1. To collapse the response variable into daily versus non-daily smokers, repeat steps 5.6 Data Management #1 through #3. For this example, to collapse the response variable into daily versus non-daily smokers set S3AQ3B1 equal to1, that is, if the individual smokes 30 days a month, **DAILY,** our new variable set equal to 1. Then S3AQ3B1 greater than or equal to 2 then **DAILY** equal to 0.


Recode into Different Variables: Old and New Values
Old Value
○ Value:
○ System-missing
○ System- or user-missing
○ Range:
through
○ Range, LOWEST through value:
● Range, value through HIGHEST:
○ All other values
New Value
● Value:
○ System-missing
○ Copy old value(s)
Add
Change
Remove
□ Output variables are strings
Width: 8
▨ Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

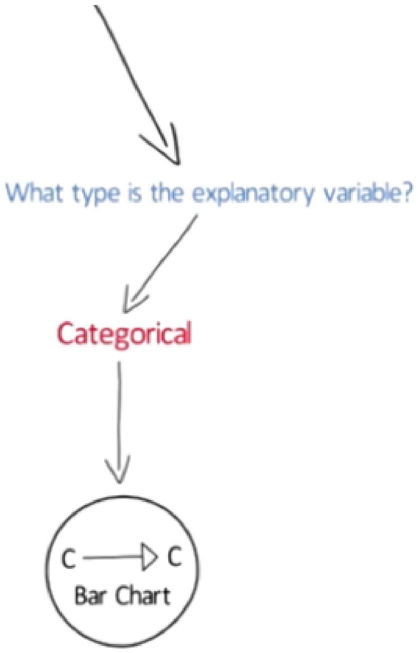
Old --> New:
1 --> 1
2 thru Highest --> 0

2. To graph the relationship between a categorical explanatory variable, **ETHNICITY**, and a categorical response variable, **DAILY**, we use the same steps for graphing the relationship between a categorical explanatory variable and a quantitative response variable--a response variable that has been binned into two categories. That is, repeat 7.2 steps #1 through #4.



Remember our categorical response variable should not have more than two categories or levels. And those two categories should be coded as 0 and 1--0 represents 'no' or negative observations and 1 represents 'yes' or positive observations. In this format, requesting the mean of our categorical response variable actually gives us the proportion of 1s or positive observations.

Because our response variable was categorical with more than two categories, we needed to collapse it into only two categories. And because our explanatory variable, ethnicity, was categorical, we created a categorical by categorical bar chart. Had our explanatory variable been quantitative, we would have needed to bin or collapse that variable into categories before creating the categorical by categorical bar chart.



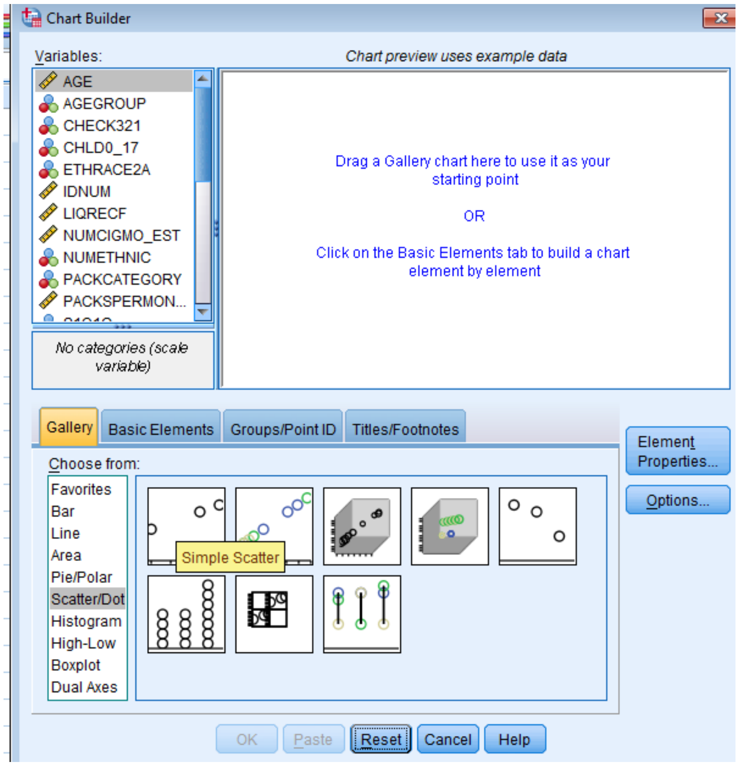
From the two bivariate graphing examples that we've covered, we've filled in the left side of our graphing decisions flow chart. Each example showed situations when our response variable was categorical. Let's talk now about the right side of our flow chart, when the response variable is quantitative.

Return to Video - 07. Bivariate Graphing

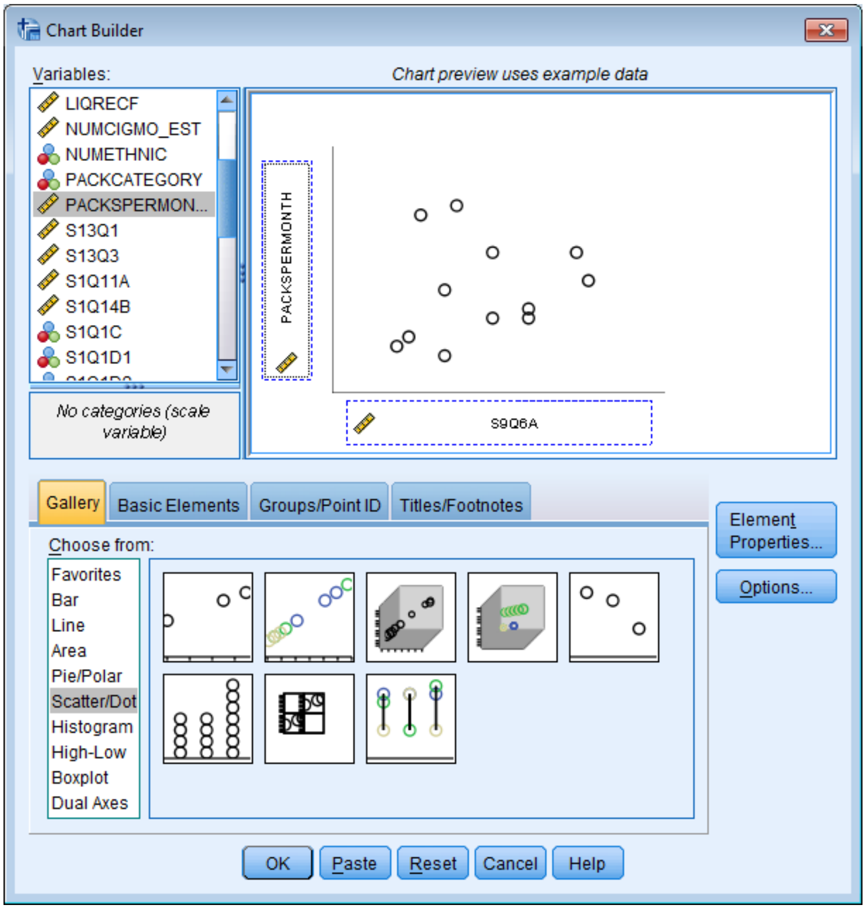
Section 7.4: Graphing Quantitative Explanatory and Quantitative Response Relationships

As an example let's examine the relationship between age onset first general anxiety episode, S9Q6A, and number of cigarette packs per month, PACKSPERMONTH. S9Q6A dummy code 99 was set to missing. Repeat 05 Data Management 5.1 steps #1 and #2 to treat for missing.

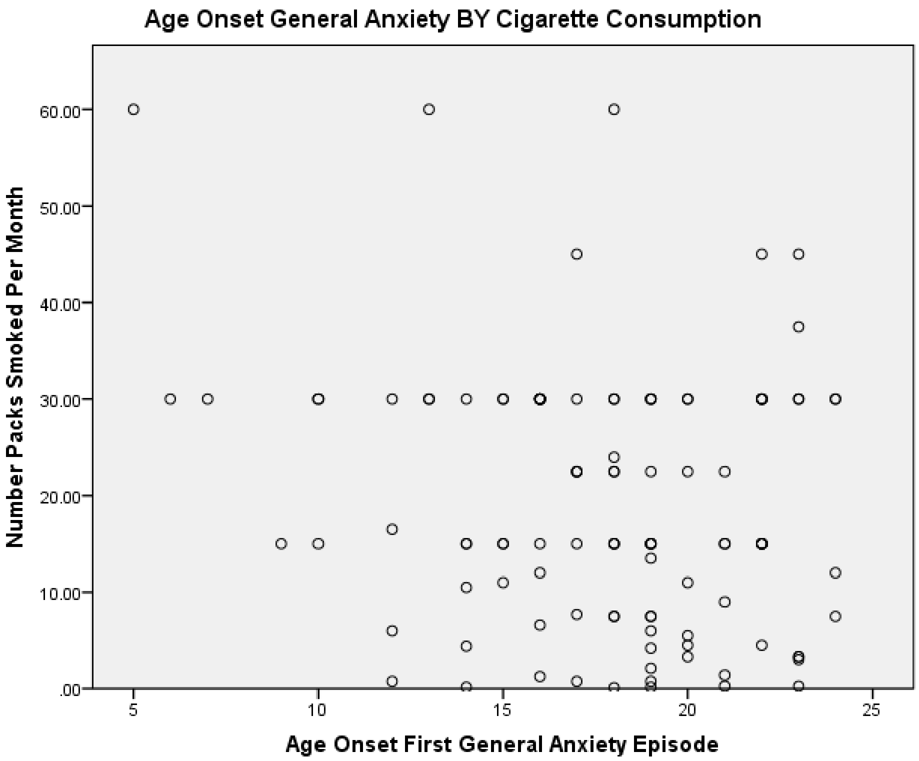
1. Click **Graphs > Chart Builder**. In the lower window under **Choose from:**, Click **Scatter/Dot**, drag the top left Simple Scatter up to the Chart preview window.



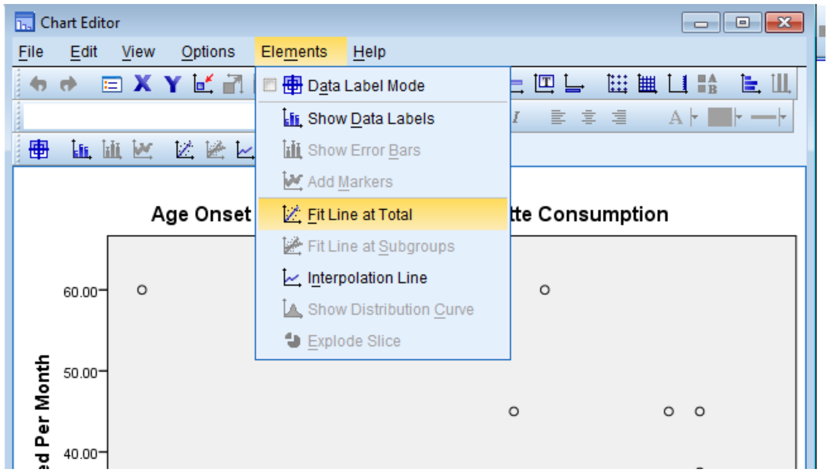
2. From the upper left **Variables:** window, drag your Quantitative Explanatory variable, S9Q6A, to the X-Axis and your Quantitative Response variable, PACKSPERMONTH, to the Y-Axis.



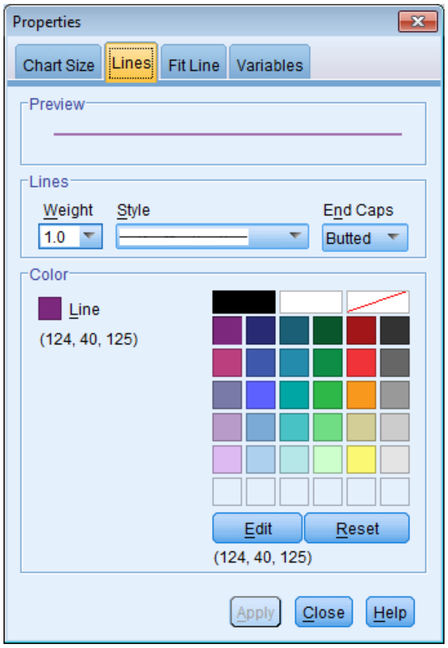
3. In the output window double click on the graph then use **Options** and the **Element Properties** to set a graph title and customize the graph.

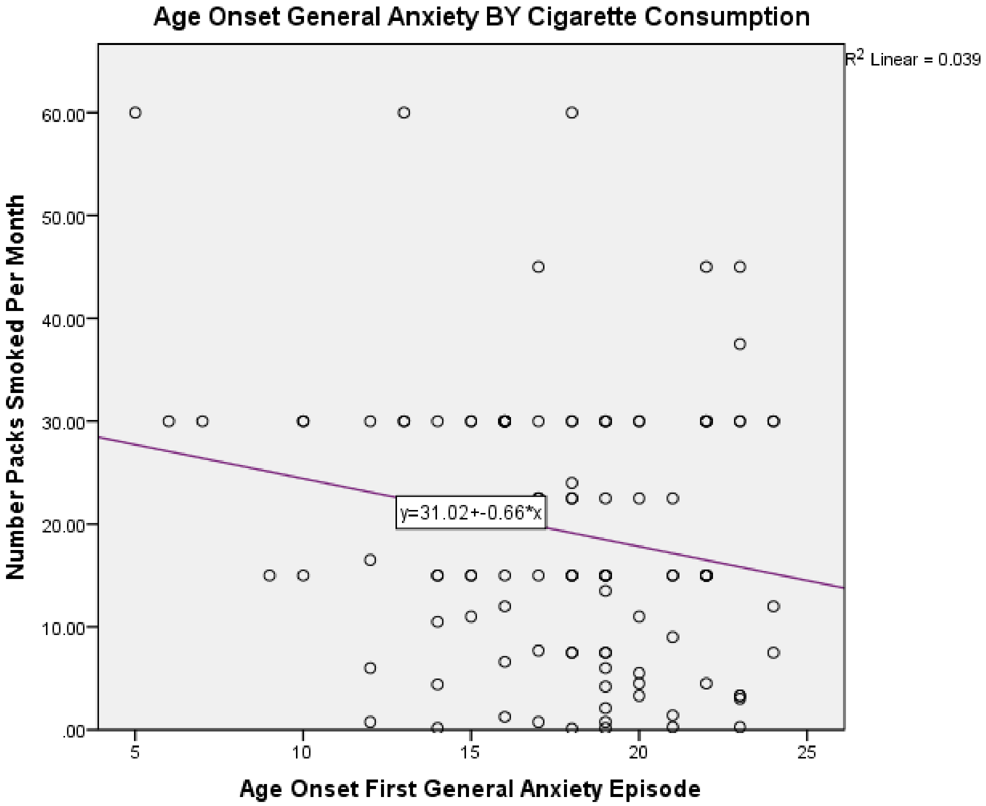


4. To characterize the relationship that we see in a scatter plot, it can be helpful to draw a line of best fit through the observations, as a way of trying to determine how the dots line up. That is, do they seem to line up in a positive or a negative direction, or with a positive or negative slope? In the output window double click the graph. Click **Element** > **Fit Line at Total**.



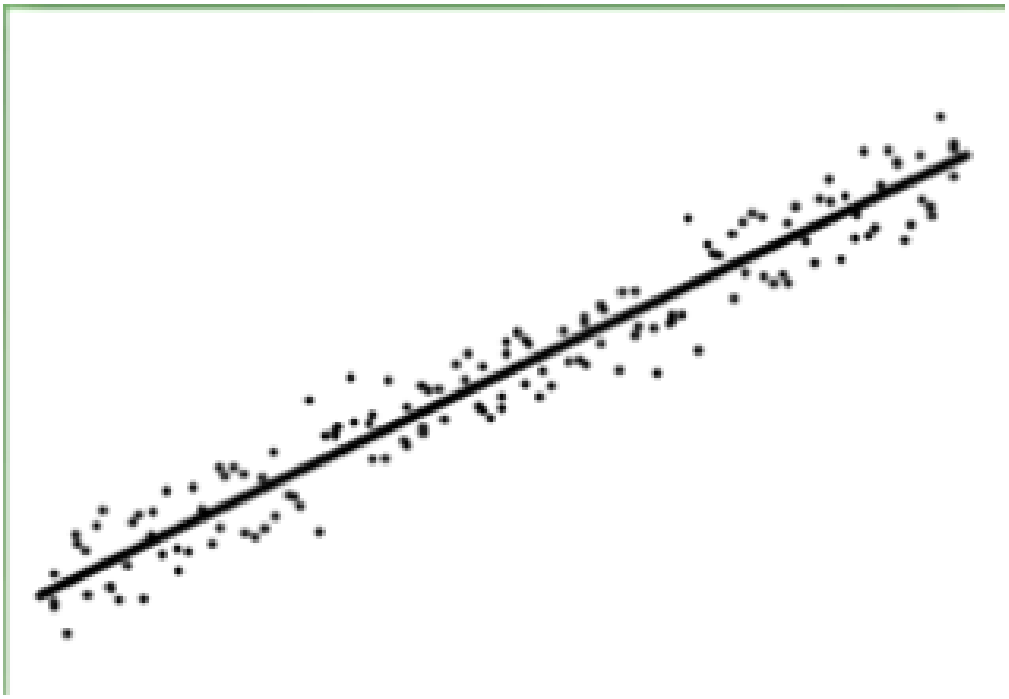
5. A Properties window will appear. There is no need to make changes, but you can Click the **Lines** tab to change the line of best fit size, color, etc. If you make it changes Click **Apply**. Close the **Properties** window and **Chart Editor** window.



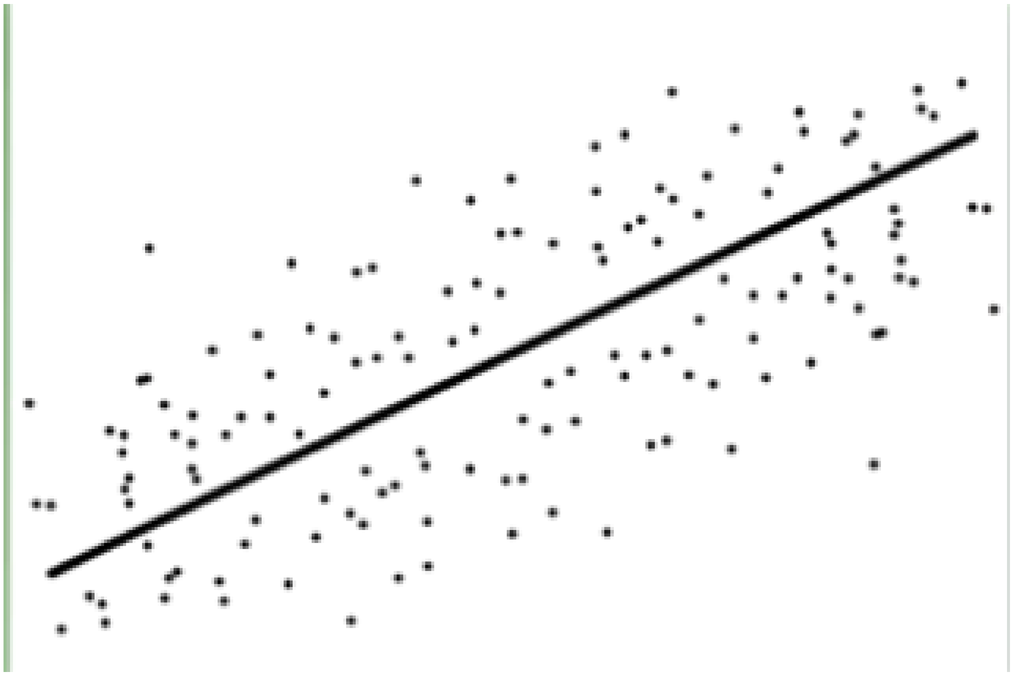


A decreasing slope, as we have here, between age onset first general anxiety episode and number of packs cigarettes smoked per month, indicates the relationship is negative. That is, an increase in one of the variables seems to be associated with a decrease in the other.

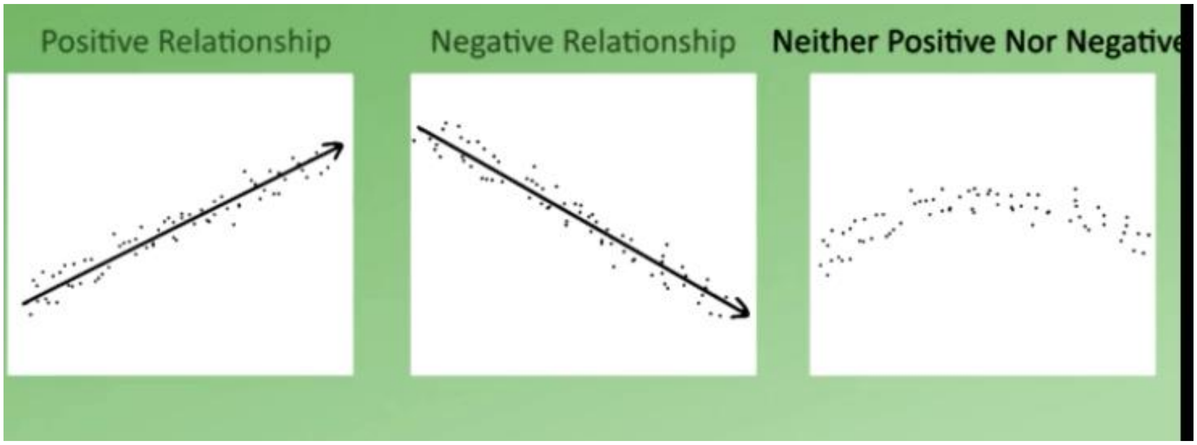
The strength of the relationship in a scatter plot is determined by how closely the data points follow the form. In this scatter plot the data points follow the linear pattern quite closely. This is an example of a very strong, positive relationship.



In this other scatter plot, the points also follow the linear pattern, but much less closely. And therefore, we can say that this is a weaker relationship.



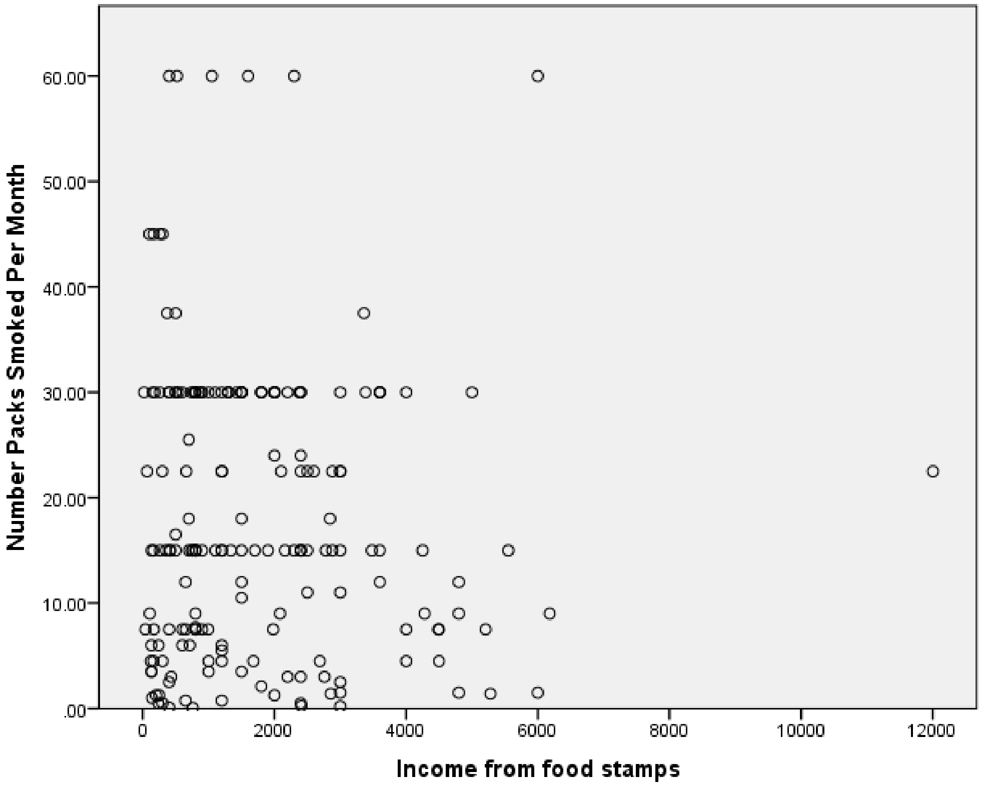
The form of the relationship is its general shape. When identifying the form, we try to find the simplest way to describe the shape of the scatter plot. There are many possible forms.
A positive or increasing relationship means that an increase in one of the variables is associated with an increase in the other. A negative or decreasing relationship means that an increase in one of the variables is associated with a decrease in the other, as shown in this central scatter plot. Not all relationships can be classified as either positive or negative. Further, if you can't plausibly put a line through the dots, if the dots are just an amorphous cloud of specks on the graph, then there may be no relationship.



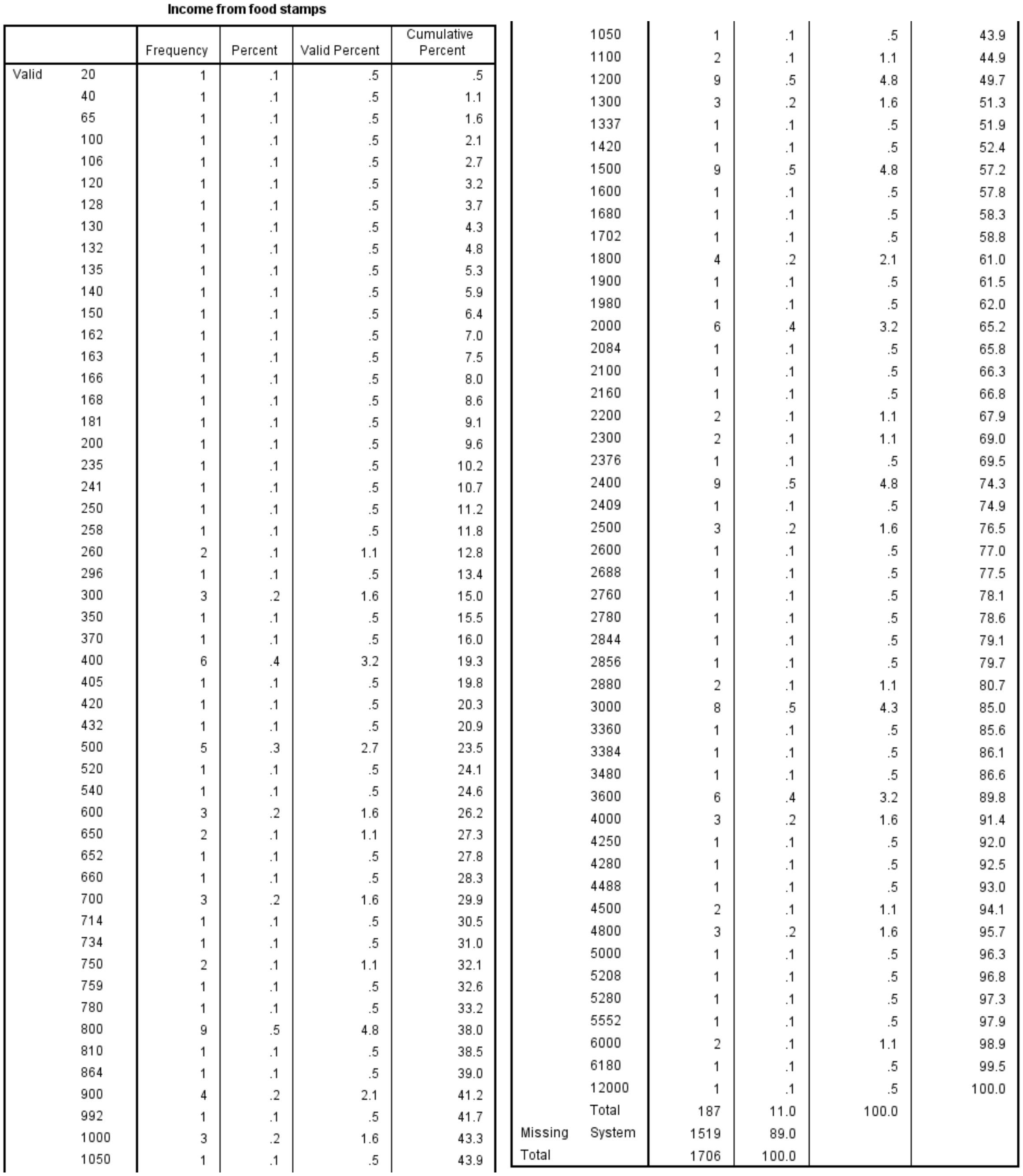
For various reasons, a scatter plot is sometimes limited in its ability to allow us to evaluate a relationship visually.

Section 7.5: Graphing Categorical Explanatory and Quantitative Response Relationships

Here is a scatterplot of income from food stamps by number of packs smoked per month. Most participants reported between 0 and $3400.00, the dots on this scatter plot seem to clump in the lower left hand corner of the graph. So to try to get a better sense of whether or not there is a relationship between these two variables, we would try to categorize or group the explanatory variable income.



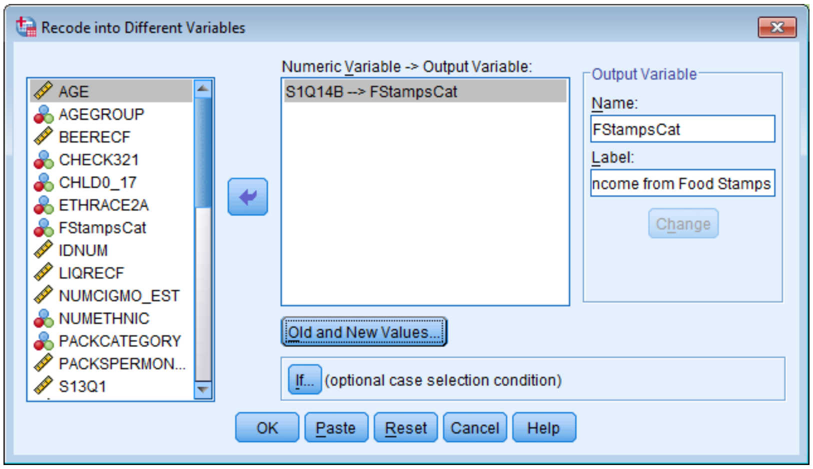
1. Generate a frequency table for S1Q14B, income from food stamps. Using the cumulative percent column we can determine the value at which the 25th, 50th, and 75th quartiles fall. That is at which value at 25%, 50%, and 75% of the data falls at or below. For the 25% = 540, 50% = 1200, and the 75% = 2409.



Income from food stamps

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20	1	.1	.5	.5
	40	1	.1	.5	1.1
	65	1	.1	.5	1.6
	100	1	.1	.5	2.1
	106	1	.1	.5	2.7
	120	1	.1	.5	3.2
	128	1	.1	.5	3.7
	130	1	.1	.5	4.3
	132	1	.1	.5	4.8
	135	1	.1	.5	5.3
	140	1	.1	.5	5.9
	150	1	.1	.5	6.4
	162	1	.1	.5	7.0
	163	1	.1	.5	7.5
	166	1	.1	.5	8.0
	168	1	.1	.5	8.6
	181	1	.1	.5	9.1
	200	1	.1	.5	9.6
	235	1	.1	.5	10.2
	241	1	.1	.5	10.7
	250	1	.1	.5	11.2
	258	1	.1	.5	11.8
	260	2	.1	1.1	12.8
	296	1	.1	.5	13.4
	300	3	.2	1.6	15.0
	350	1	.1	.5	15.5
	370	1	.1	.5	16.0
	400	6	.4	3.2	19.3
	405	1	.1	.5	19.8
	420	1	.1	.5	20.3
	432	1	.1	.5	20.9
	500	5	.3	2.7	23.5
	520	1	.1	.5	24.1
	540	1	.1	.5	24.6
	600	3	.2	1.6	26.2
	650	2	.1	1.1	27.3
	652	1	.1	.5	27.8
	660	1	.1	.5	28.3
	700	3	.2	1.6	29.9
	714	1	.1	.5	30.5
	734	1	.1	.5	31.0
	750	2	.1	1.1	32.1
	759	1	.1	.5	32.6
	780	1	.1	.5	33.2
	800	9	.5	4.8	38.0
	810	1	.1	.5	38.5
	864	1	.1	.5	39.0
	900	4	.2	2.1	41.2
	992	1	.1	.5	41.7
	1000	3	.2	1.6	43.3
	1050	1	.1	.5	43.9
	1100	2	.1	1.1	44.9
	1200	9	.5	4.8	49.7
	1300	3	.2	1.6	51.3
	1337	1	.1	.5	51.9
	1420	1	.1	.5	52.4
	1500	9	.5	4.8	57.2
	1600	1	.1	.5	57.8
	1680	1	.1	.5	58.3
	1702	1	.1	.5	58.8
	1800	4	.2	2.1	61.0
	1900	1	.1	.5	61.5
	1980	1	.1	.5	62.0
	2000	6	.4	3.2	65.2
	2084	1	.1	.5	65.8
	2100	1	.1	.5	66.3
	2160	1	.1	.5	66.8
	2200	2	.1	1.1	67.9
	2300	2	.1	1.1	69.0
	2376	1	.1	.5	69.5
	2400	9	.5	4.8	74.3
	2409	1	.1	.5	74.9
	2500	3	.2	1.6	76.5
	2600	1	.1	.5	77.0
	2688	1	.1	.5	77.5
	2760	1	.1	.5	78.1
	2780	1	.1	.5	78.6
	2844	1	.1	.5	79.1
	2856	1	.1	.5	79.7
	2880	2	.1	1.1	80.7
	3000	8	.5	4.3	85.0
	3360	1	.1	.5	85.6
	3384	1	.1	.5	86.1
	3480	1	.1	.5	86.6
	3600	6	.4	3.2	89.8
	4000	3	.2	1.6	91.4
	4250	1	.1	.5	92.0
	4280	1	.1	.5	92.5
	4488	1	.1	.5	93.0
	4500	2	.1	1.1	94.1
	4800	3	.2	1.6	95.7
	5000	1	.1	.5	96.3
	5208	1	.1	.5	96.8
	5280	1	.1	.5	97.3
	5552	1	.1	.5	97.9
	6000	2	.1	1.1	98.9
	6180	1	.1	.5	99.5
	12000	1	.1	.5	100.0
	Total	187	11.0	100.0	
Missing	System	1519	89.0		
Total		1706	100.0		

2. From 05 Data Management repeat steps 5.6 #1 through #4 to collapse S1Q14B into 4 categories, FSTAMPSCAT, based on the percentiles cuts we determined in the previous step.

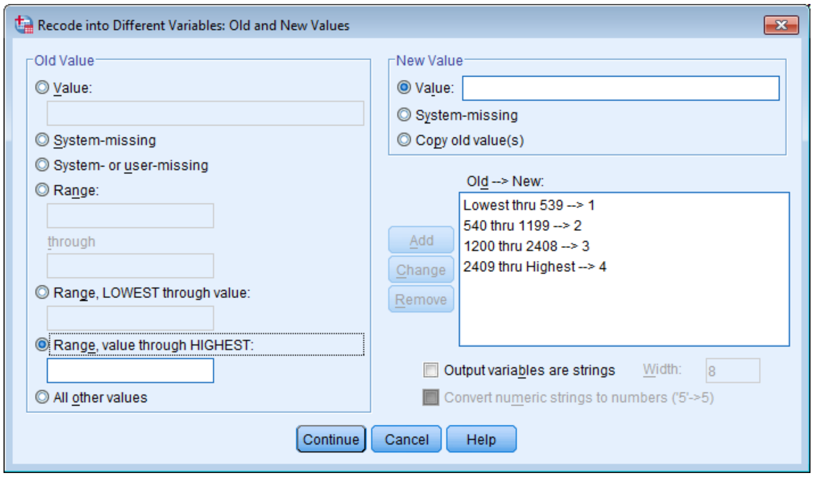


Numeric Variable -> Output Variable: S1Q14B --> FStampsCat

Output Variable

Name: FStampsCat

Label: ncome from Food Stamps



Old Value

○ Value:

○ System-missing

○ System- or user-missing

○ Range:

through

○ Range, LOWEST through value:

● Range, value through HIGHEST:

○ All other values

New Value

● Value:

○ System-missing

○ Copy old value(s)

Old --> New:

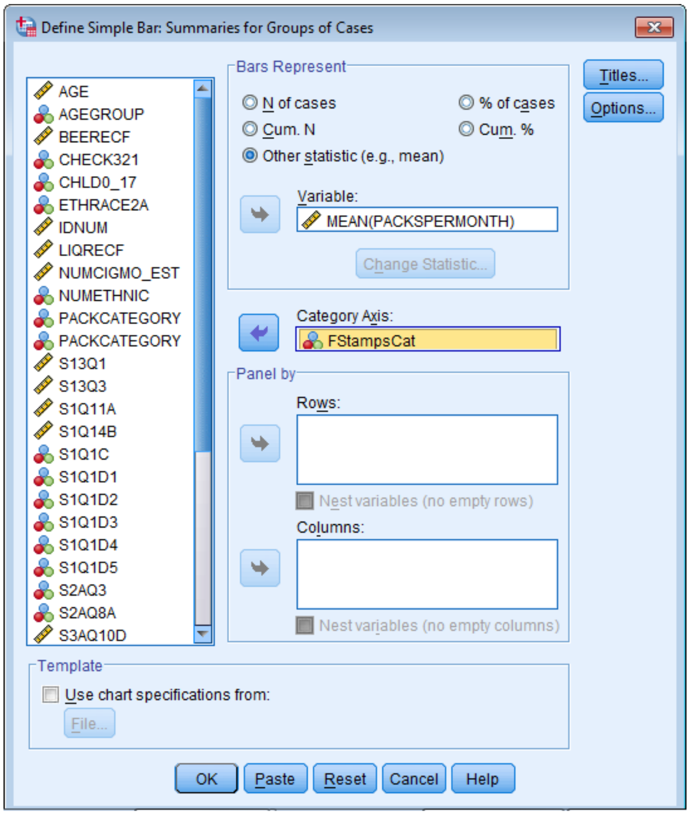
```
Lowest thru 539 --> 1
540 thru 1199 --> 2
1200 thru 2408 --> 3
2409 thru Highest --> 4
```

□ Output variables are strings Width: 8

□ Convert numeric strings to numbers ('5'->5)

Now we are ready to graph a Categorical Explanatory variable, FSTAMPSCAT, and a Quantitative Response variable, PACKSPERMONTH.

3. Repeat steps 7.2 #1 through #4 using the appropriate variables.



Define Simple Bar: Summaries for Groups of Cases

Bars Represent

○ N of cases ○ % of cases

○ Cum. N ○ Cum. %

● Other statistic (e.g., mean)

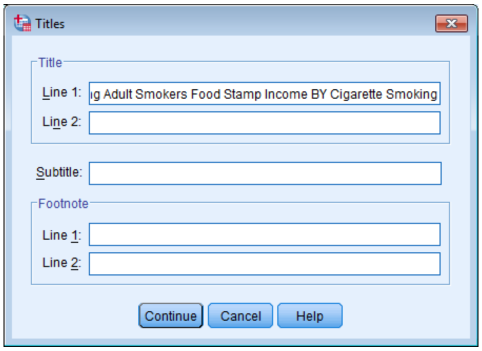
Variable: MEAN(PACKSPERMONTH)

Category Axis: FStampsCat

Panel by

Template

□ Use chart specifications from:



Titles

Title

Line 1: ıg Adult Smokers Food Stamp Income BY Cigarette Smoking

Line 2:

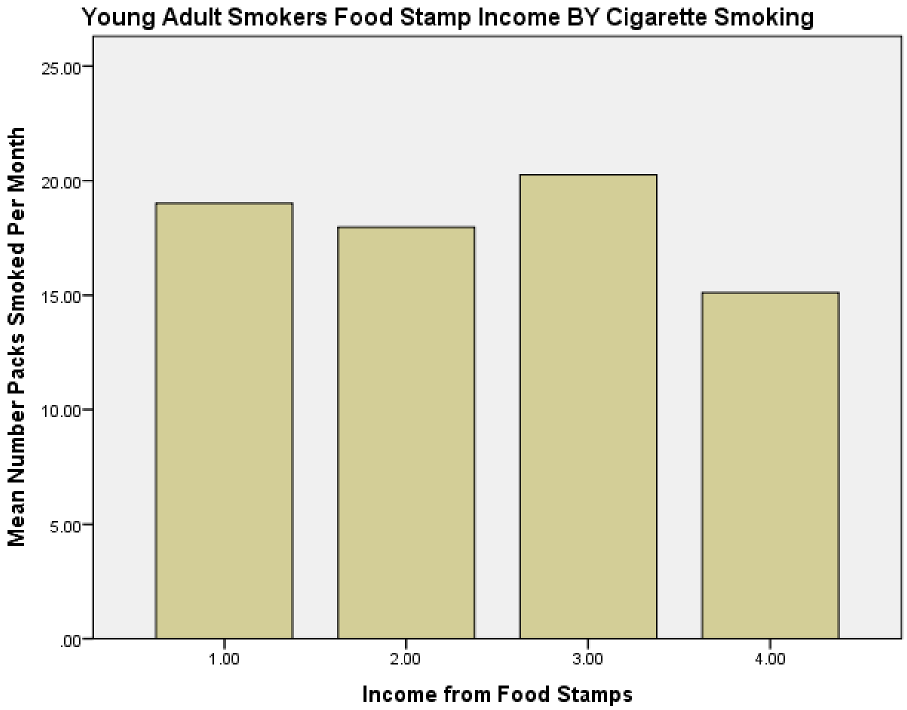
Subtitle:

Footnote

Line 1:

Line 2:

4. Complete the appropriate edits to your graph.



In this bar chart, while we can see differences in packs smoked per month on income from food stamps, the relationship does not seem to be linear. One might have expected that as income from food stamps goes up the number of packs smoked per month would go down, that is a negative linear relationship. However, group 3 does not follow that linear trend we might have expected.

We've worked through each type of bivariate that is 2-variable graph, highlighting when and how each should be used to visualize a relationship. Now let's just very briefly summarize.

When visualizing a categorical to categorical relationship, we use a bar chart with explanatory categories on the x-axis, and the proportion of our response variable on the y-axis.

When visualizing a categorical to quantitative relationship, we use a bar chart with explanatory categories on the x-axis, and the mean of our response variable on the y-axis.

When visualizing a quantitative to quantitative relationship, we use a scatter plot, in which each observation is displayed according to the values of the explanatory and response variables.

Use these basic guidelines, as well as the graphing decisions flowchart to visualize the relationships between your own variables of interests.