

3190 Week 3

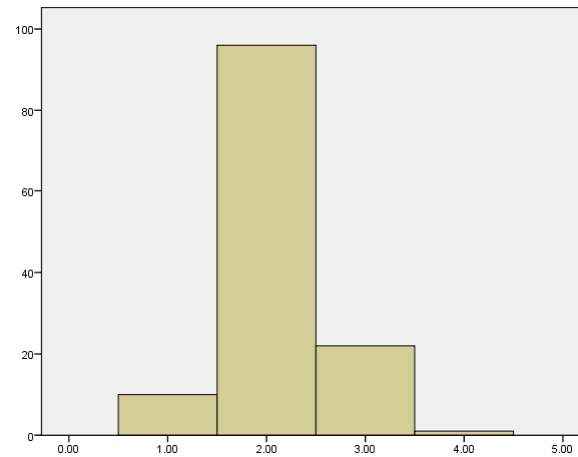
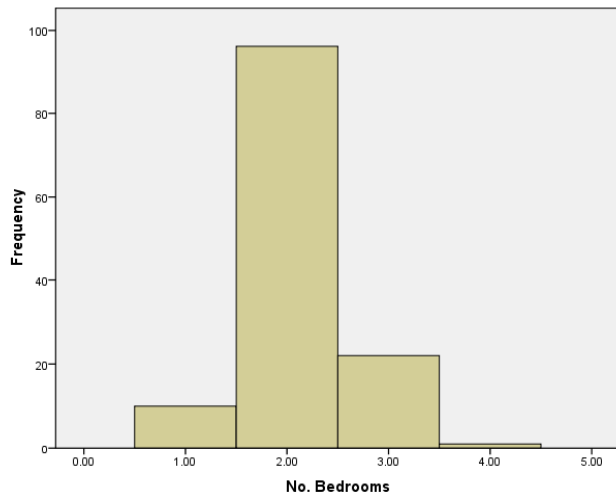
Graphing Data

Why graph?

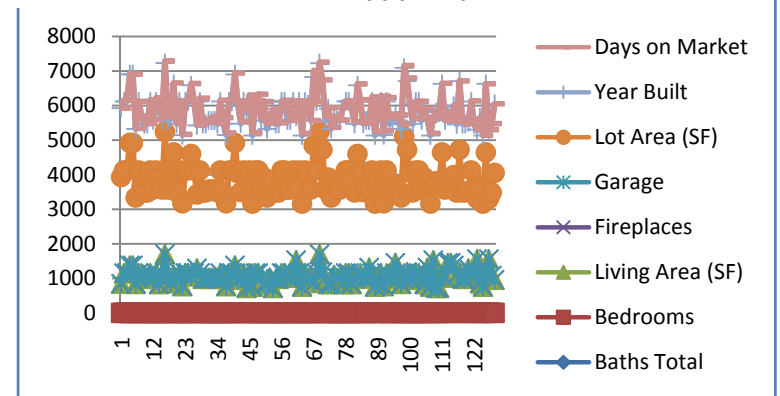
- Reports should be integrated
 - They should be narrative
 - They should be simple, yet provide detail as needed
 - They should be visual...
- Assume an ignorant audience (not stupid, ignorant)
- Anticipate your audience
 - What might they criticize
- Graphing is simply another form of communication

The Good, The Bad, & The Ugly

Townhouse Bedroom Distribution



A Crappy Graph



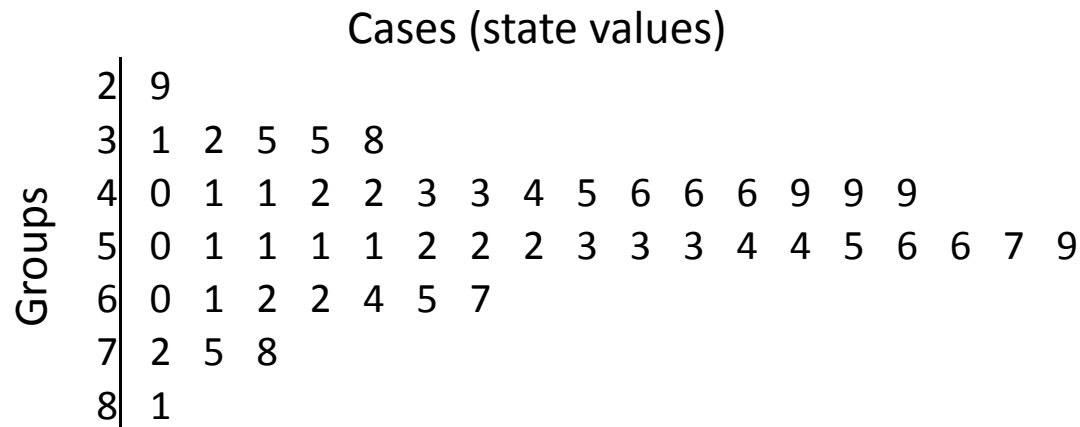
Summary

- Part of communication is to summarize
- Summary is an art
- You must make critical choices about how will make simple points that go a long way

Stem & Leaf Diagram

- An easy way to convert frequency distributions to a chart

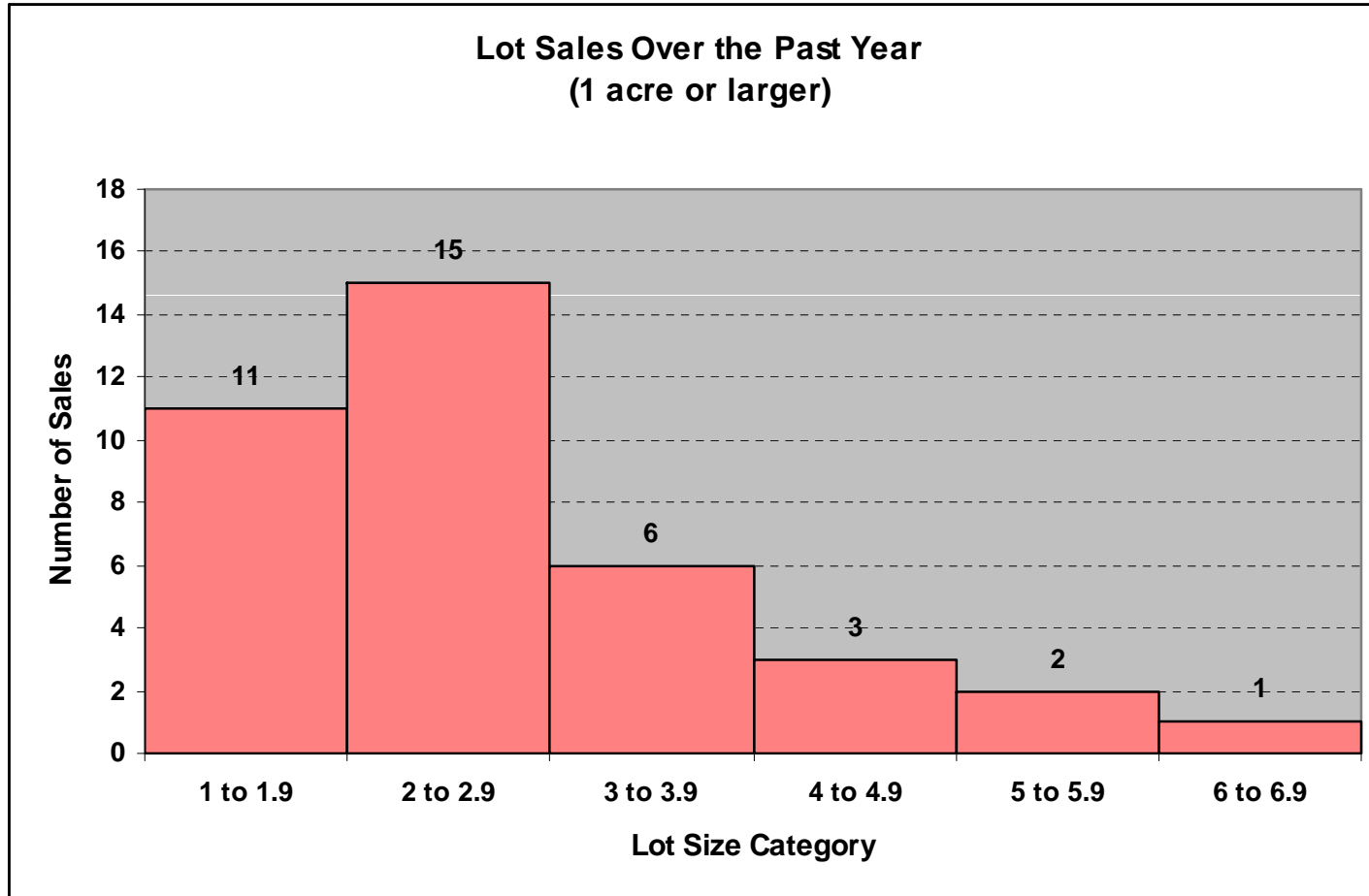
Distribution of 1996 US state unemployment rates



Histograms

- Turns a frequency distribution into a chart
- Uses bars to represent the frequency of cases in groups
- Bars are immediately adjacent to each other
 - Y axis = frequency
 - X axis = group number
 - You can label the x-axis with group boundaries or the group midpoint
 - All histograms need to have labeled axes & clear chart titles

Histogram, example

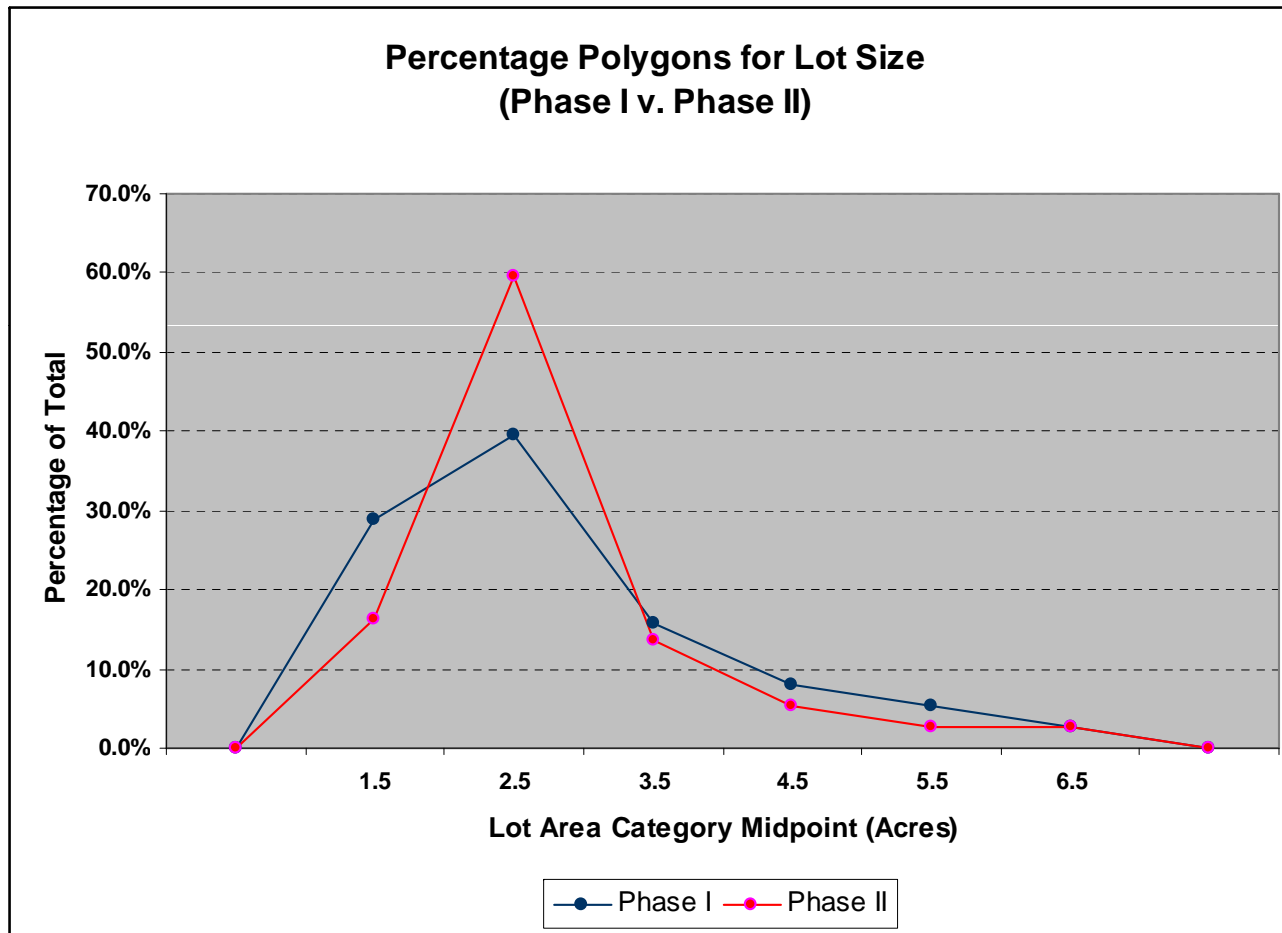


Does this frequency show a raw data or relative frequency distribution?

Frequency Polygon

- Converts a histogram to a line graph
- Allows easy expression of multiple frequency distributions on the same graph
 - Which is cumbersome with a histogram

Frequency Polygon, example



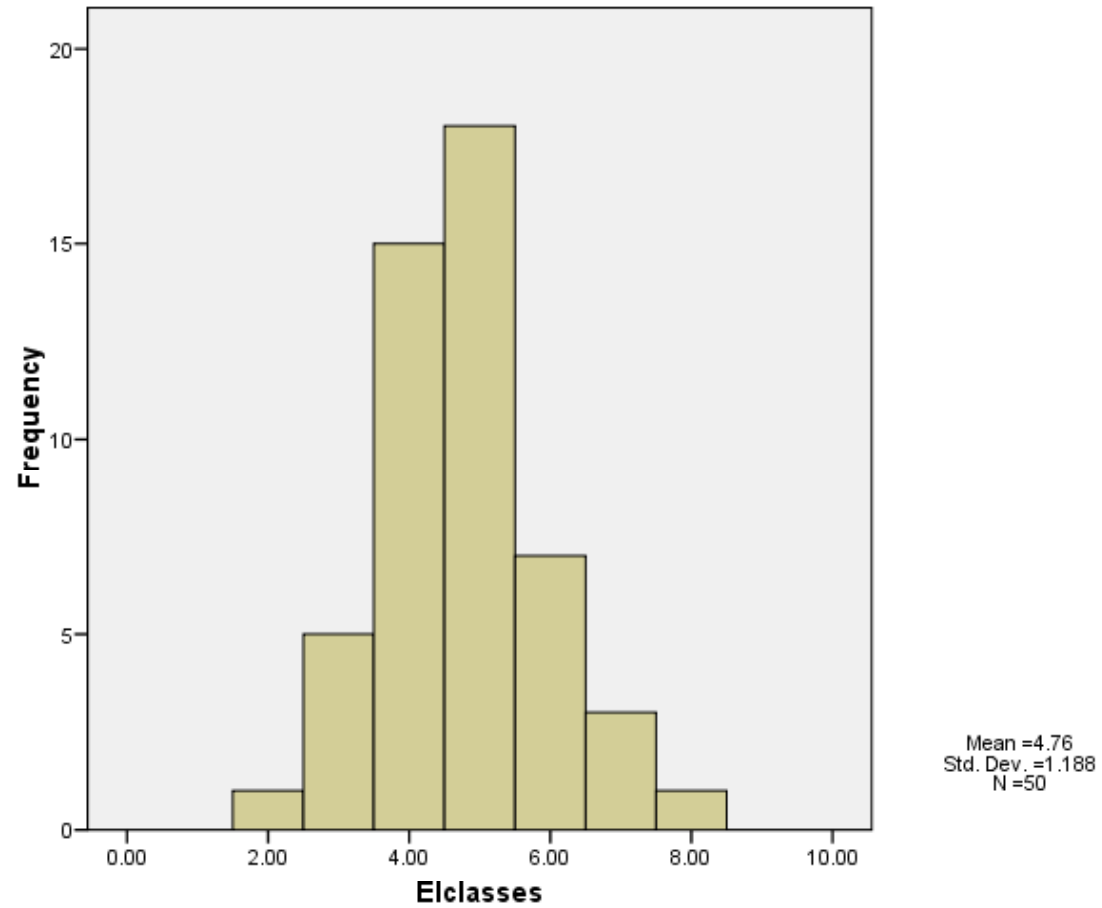
Does this frequency show a raw data or relative frequency distribution?

Cumulative Frequency Polygon

- A frequency polygon that graphs frequency of cases in a group and frequency of cases in any lower groups
- Its is additive
- For each group you graph its frequency and add the frequency of all lower groups to it.

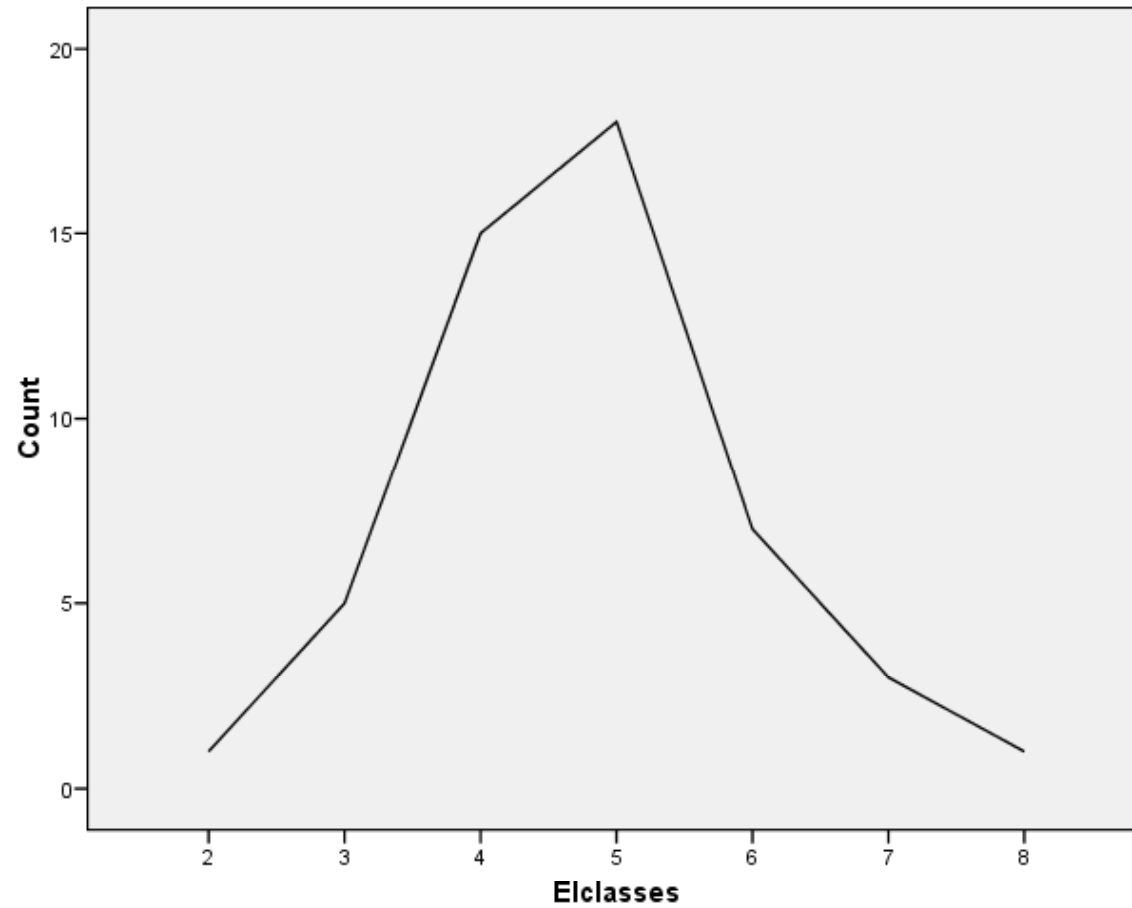
Histogram

Distribution of 1996 US state unemployment rates



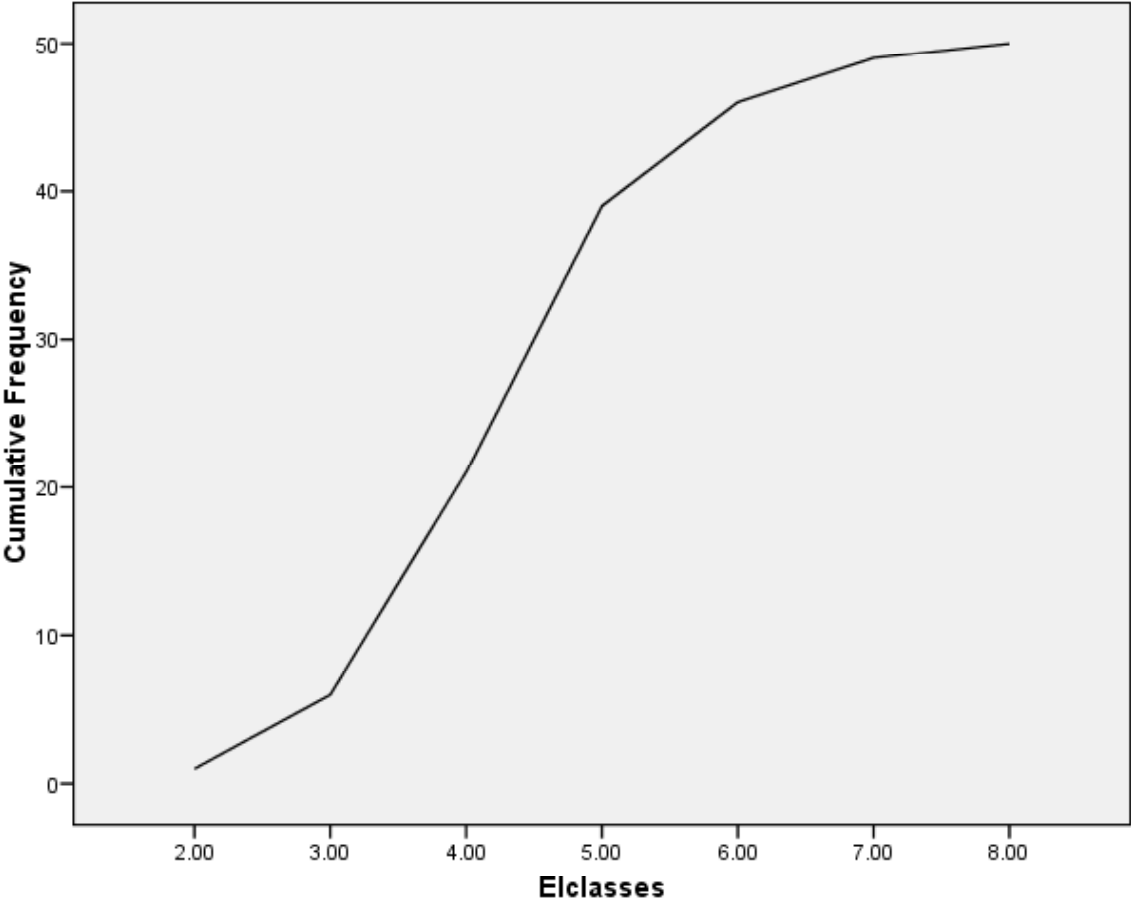
Frequency Polygon

Distribution of 1996 US state unemployment rates



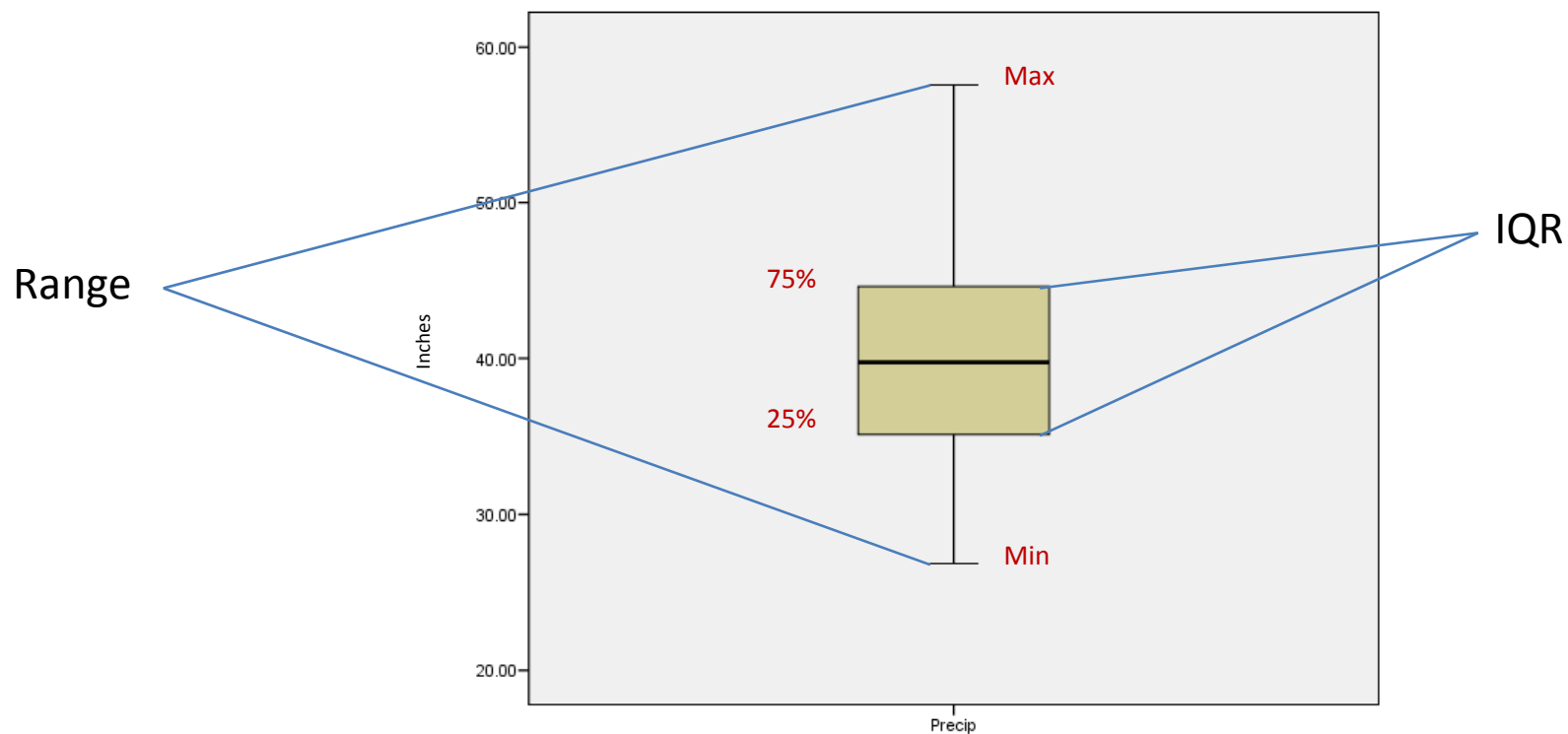
Cumulative Frequency Polygon

Cumulative Distribution of 1996 US state unemployment rates

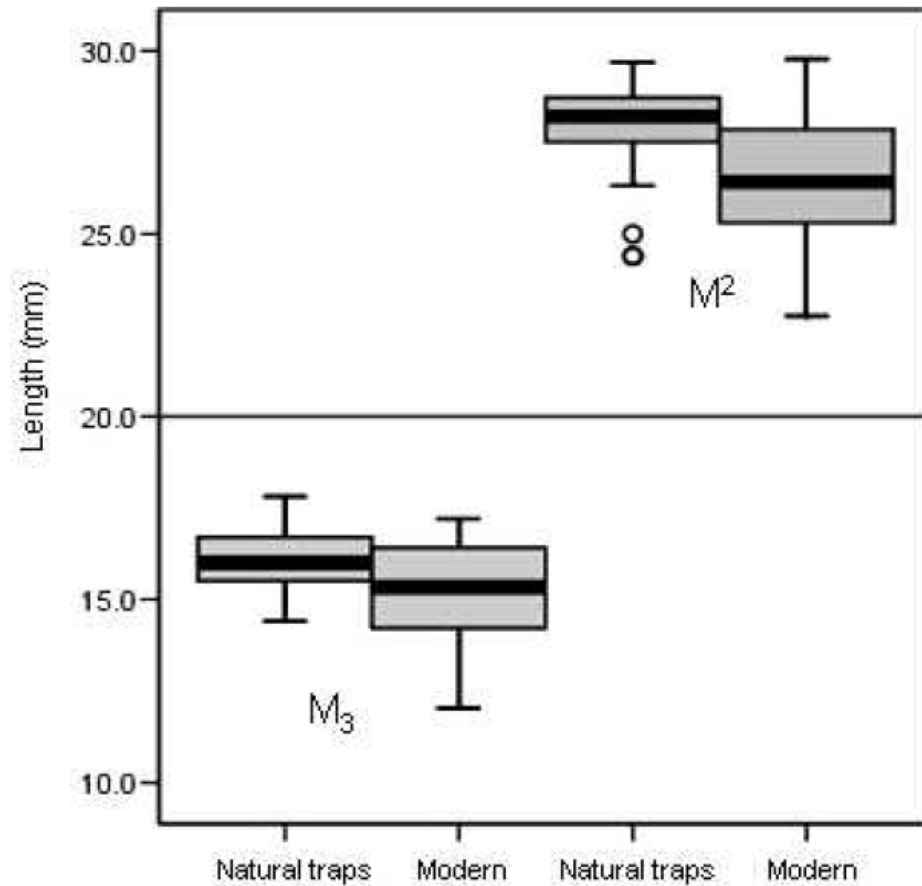


Boxplots

- Simply a graphic representation of the 5 points of data summary
- Tells you central tendency, dispersion, and symmetry



Boxplots & symmetry

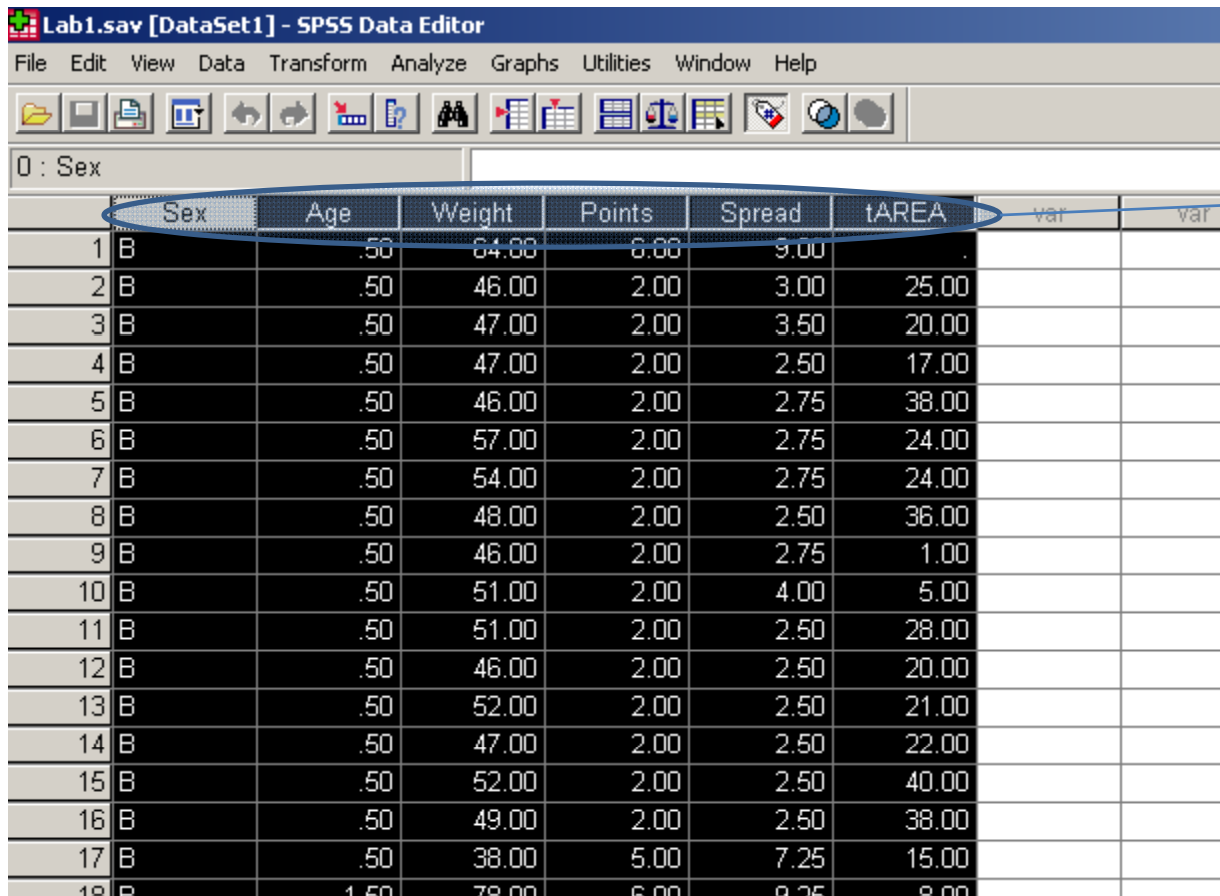


Graphs in SPSS

Sorting & Grouping

- The easiest way to group data is to sort data into an ordered array
- Then insert a new variable and use it as a “grouping variable”
- This is important for graphs and some inferential tests later in the class

Sorting



Lab1.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

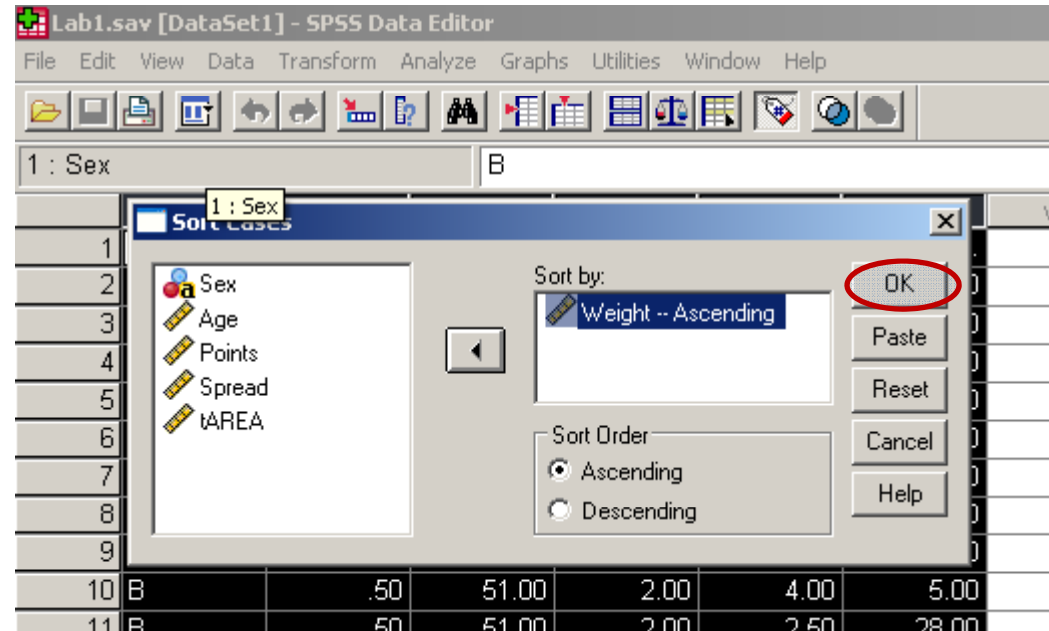
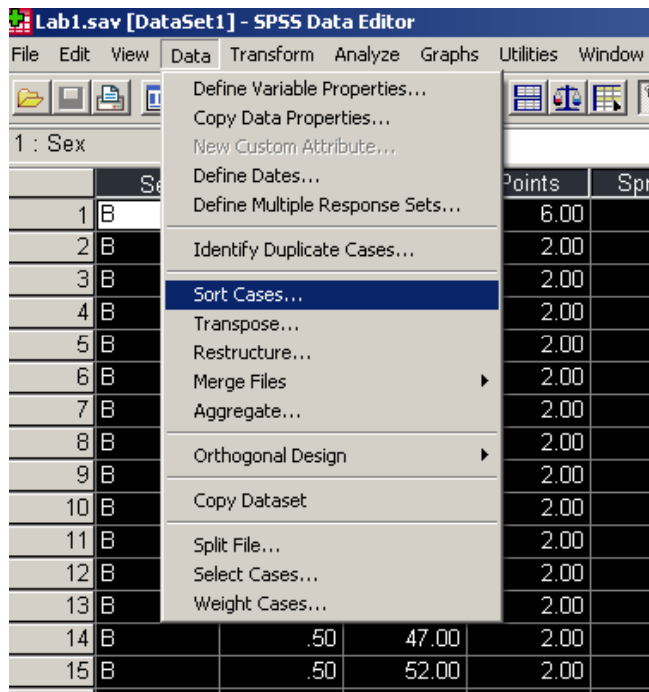
0 : Sex

	Sex	Age	Weight	Points	Spread	tAREA	var	var
1	B	.50	64.00	6.00	9.00			
2	B	.50	46.00	2.00	3.00	25.00		
3	B	.50	47.00	2.00	3.50	20.00		
4	B	.50	47.00	2.00	2.50	17.00		
5	B	.50	46.00	2.00	2.75	38.00		
6	B	.50	57.00	2.00	2.75	24.00		
7	B	.50	54.00	2.00	2.75	24.00		
8	B	.50	48.00	2.00	2.50	36.00		
9	B	.50	46.00	2.00	2.75	1.00		
10	B	.50	51.00	2.00	4.00	5.00		
11	B	.50	51.00	2.00	2.50	28.00		
12	B	.50	46.00	2.00	2.50	20.00		
13	B	.50	52.00	2.00	2.50	21.00		
14	B	.50	47.00	2.00	2.50	22.00		
15	B	.50	52.00	2.00	2.50	40.00		
16	B	.50	49.00	2.00	2.50	38.00		
17	B	.50	38.00	5.00	7.25	15.00		
18	B	.50	78.00	6.00	8.25	8.00		

Click & drag to the right to highlight each column

This is a very important step; you do not want to sort only 1 column, which mixes up your case data

Sorting



This will sort your cases by weight from lightest to heaviest rearranging all of the other data (e.g., age, sex) appropriately

Sorting

*Lab1.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

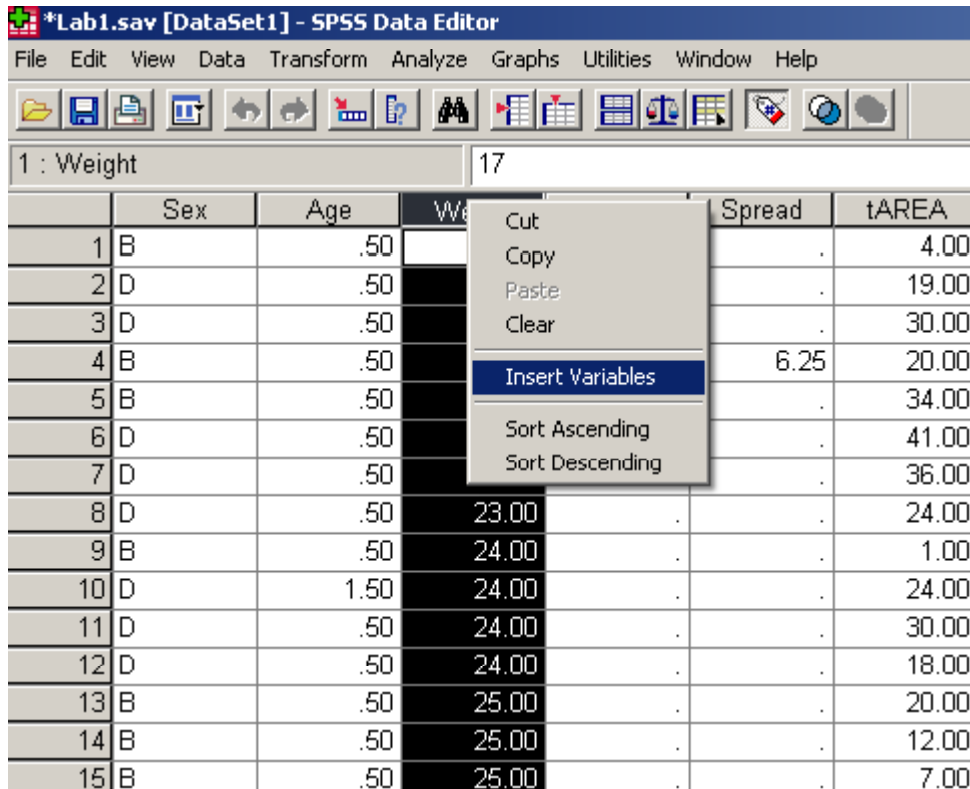
1 : Sex B

	Sex	Age	Weight	Points	Spread	tAREA
1	B	.50	17.00	.	.	4.00
2	D	.50	20.00	.	.	19.00
3	D	.50	21.00	.	.	30.00
4	B	.50	22.00	2.00	6.25	20.00
5	B	.50	23.00	.	.	34.00
6	D	.50	23.00	.	.	41.00
7	D	.50	23.00	.	.	36.00
8	D	.50	23.00	.	.	24.00
9	B	.50	24.00	.	.	1.00
10	D	1.50	24.00	.	.	24.00
11	D	.50	24.00	.	.	30.00
12	D	.50	24.00	.	.	18.00
13	B	.50	25.00	.	.	20.00
14	B	.50	25.00	.	.	12.00
15	B	.50	25.00	.	.	7.00
16	D	2.50	25.00	.	.	40.00
17	D	.50	25.00	.	.	41.00
18	D	.50	25.00	.	.	32.00
19	D	.50	25.00	.	.	12.00
20	D	.50	25.00	.	.	3.00
21	D	.50	25.00	.	.	1.00
22	D	.50	25.00	.	.	14.00
23	D	.50	25.00	.	.	1.00
24	D	.50	25.00	.	.	25.00
25	B	3.50	26.00	4.00	11.75	.
26	B	.50	26.00	.	.	32.00
27	B	.50	26.00	.	.	25.00
28	B	.50	26.00	.	.	23.00
29	B	.50	26.00	.	.	8.00
30	D	.50	26.00	.	.	2.00
31	B	.50	27.00	.	.	15.00
32	R	.50	27.00	.	.	6.00

Now each deer is ordered from lowest to highest in terms of weight; notice many of these are does

The heaviest deer (if you scroll down) will be bucks

Grouping



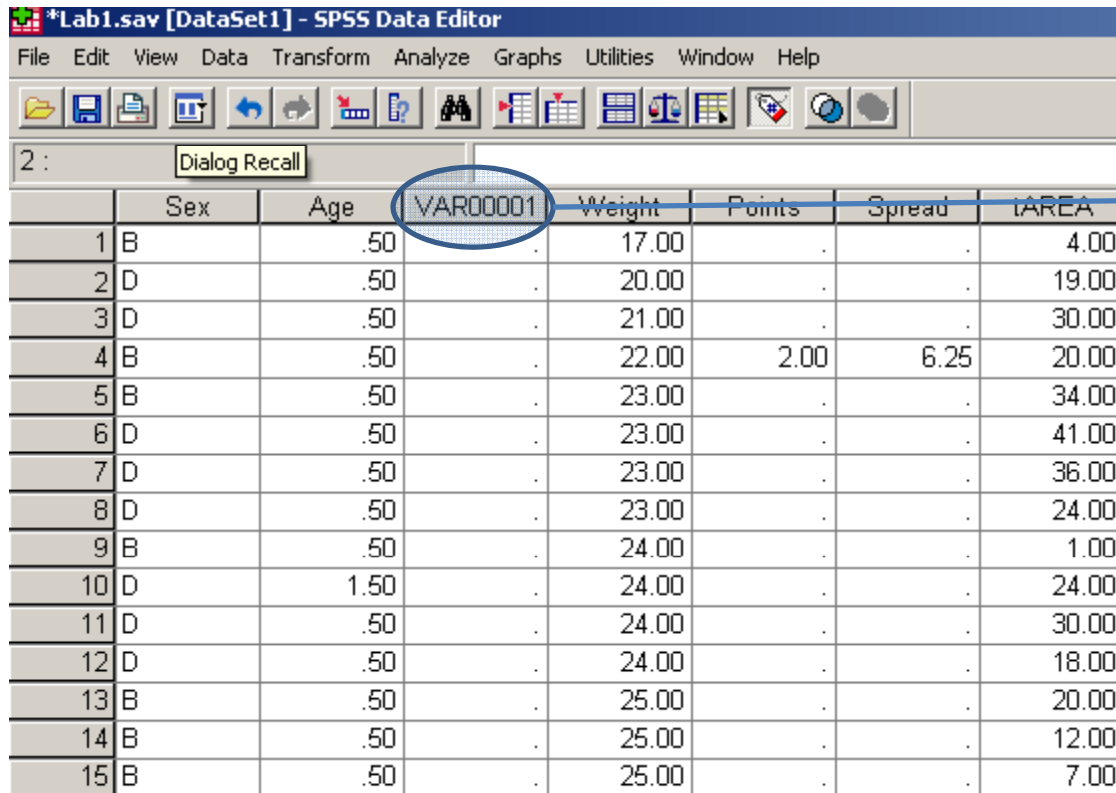
The screenshot shows the SPSS Data Editor interface for a file named '*Lab1.sav [DataSet1]'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The data view shows a table with columns: Sex, Age, Weight, Spread, and tAREA. The 'Weight' column is highlighted, and a context menu is open over it, showing options: Cut, Copy, Paste, Clear, Insert Variables (highlighted), Sort Ascending, and Sort Descending. The data rows are numbered 1 to 15.

	Sex	Age	Weight	Spread	tAREA
1	B	.50		.	4.00
2	D	.50		.	19.00
3	D	.50		.	30.00
4	B	.50		6.25	20.00
5	B	.50		.	34.00
6	D	.50		.	41.00
7	D	.50		.	36.00
8	D	.50	23.00	.	24.00
9	B	.50	24.00	.	1.00
10	D	1.50	24.00	.	24.00
11	D	.50	24.00	.	30.00
12	D	.50	24.00	.	18.00
13	B	.50	25.00	.	20.00
14	B	.50	25.00	.	12.00
15	B	.50	25.00	.	7.00

Highlight the weight variable

Then right click and select
insert variables

Grouping

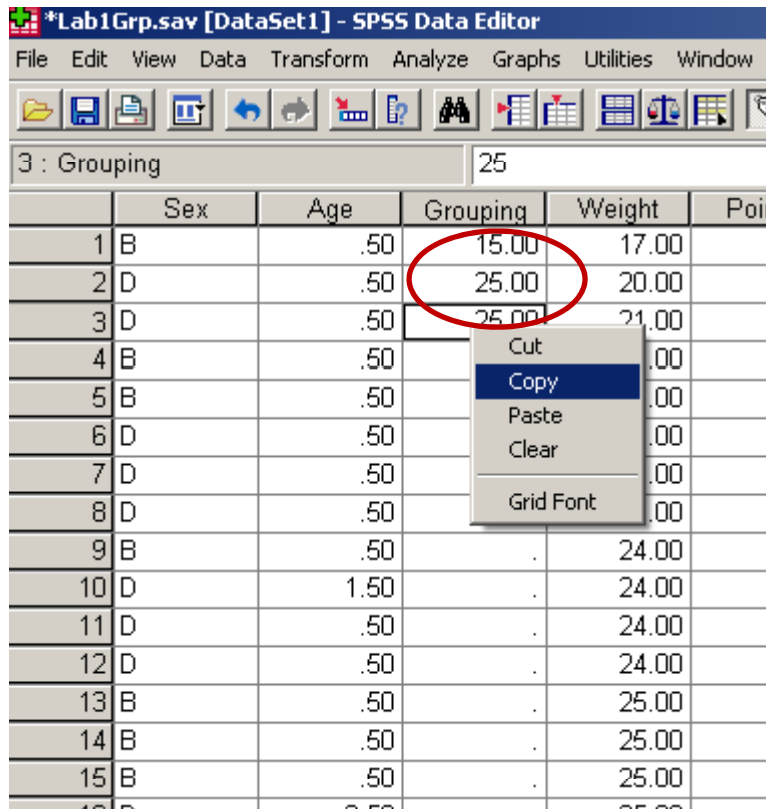


The screenshot shows the SPSS Data Editor interface for a file named '*Lab1.sav [DataSet1]'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. Below the menu bar is a toolbar with various icons. The main window displays a data table with 15 rows and 8 columns. The columns are labeled Sex, Age, VAR00001, Weight, Points, Spread, and tAREA. The 'VAR00001' column is circled in blue. The data rows are as follows:

	Sex	Age	VAR00001	Weight	Points	Spread	tAREA
1	B	.50	.	17.00	.	.	4.00
2	D	.50	.	20.00	.	.	19.00
3	D	.50	.	21.00	.	.	30.00
4	B	.50	.	22.00	2.00	6.25	20.00
5	B	.50	.	23.00	.	.	34.00
6	D	.50	.	23.00	.	.	41.00
7	D	.50	.	23.00	.	.	36.00
8	D	.50	.	23.00	.	.	24.00
9	B	.50	.	24.00	.	.	1.00
10	D	1.50	.	24.00	.	.	24.00
11	D	.50	.	24.00	.	.	30.00
12	D	.50	.	24.00	.	.	18.00
13	B	.50	.	25.00	.	.	20.00
14	B	.50	.	25.00	.	.	12.00
15	B	.50	.	25.00	.	.	7.00

New variable; go to “variable view” and name it “grouping”

Grouping



*Lab1Grp.sav [DataSet1] - SPSS Data Editor

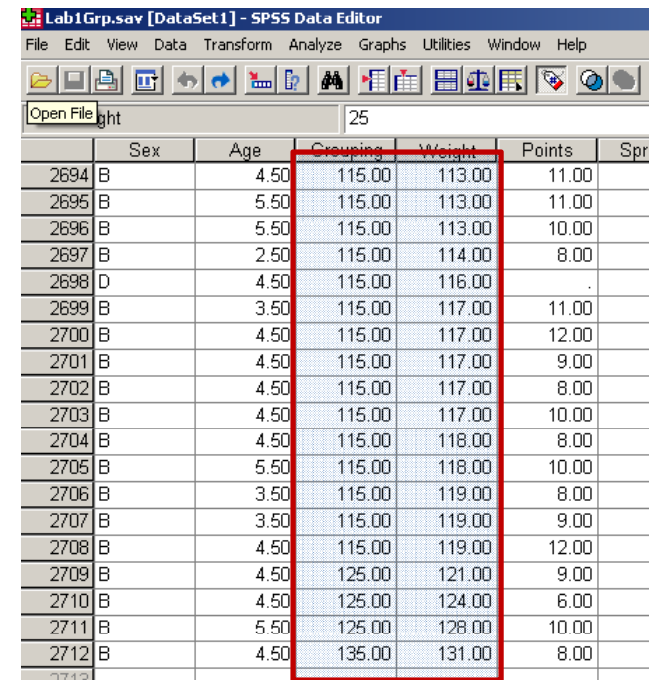
File Edit View Data Transform Analyze Graphs Utilities Window

3 : Grouping 25

	Sex	Age	Grouping	Weight	Poi
1	B	.50	15.00	17.00	
2	D	.50	25.00	20.00	
3	D	.50	25.00	21.00	
4	B	.50	.	.00	
5	B	.50	.	.00	
6	D	.50	.	.00	
7	D	.50	.	.00	
8	D	.50	.	.00	
9	B	.50	.	24.00	
10	D	1.50	.	24.00	
11	D	.50	.	24.00	
12	D	.50	.	24.00	
13	B	.50	.	25.00	
14	B	.50	.	25.00	
15	B	.50	.	25.00	

Here, I have labeled any deer with weight group mid points $10 - 19.9 = 15$, $20 - 29.9 = 25$ and so forth.

Go through and label the rest in the grouping variable



*Lab1Grp.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

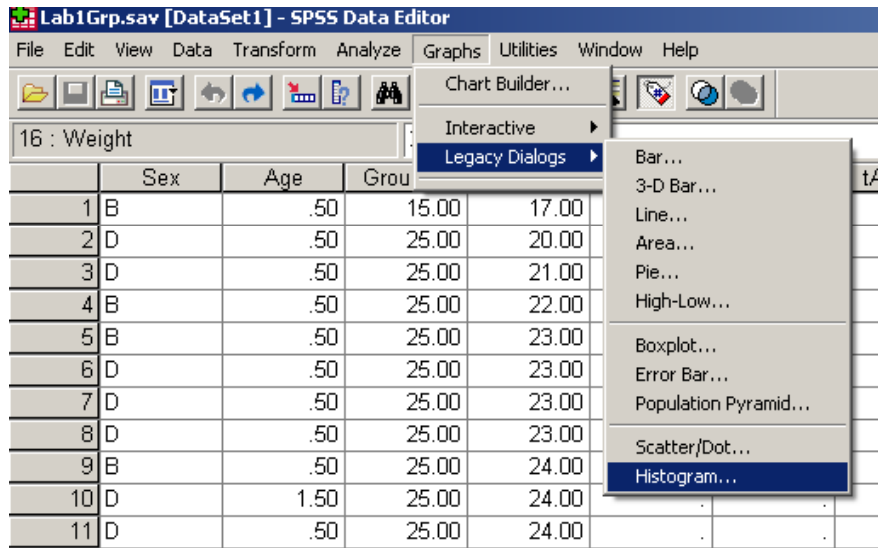
Open File ght 25

	Sex	Age	Grouping	Weight	Points	Spr
2694	B	4.50	115.00	113.00	11.00	
2695	B	5.50	115.00	113.00	11.00	
2696	B	5.50	115.00	113.00	10.00	
2697	B	2.50	115.00	114.00	8.00	
2698	D	4.50	115.00	116.00	.	
2699	B	3.50	115.00	117.00	11.00	
2700	B	4.50	115.00	117.00	12.00	
2701	B	4.50	115.00	117.00	9.00	
2702	B	4.50	115.00	117.00	8.00	
2703	B	4.50	115.00	117.00	10.00	
2704	B	4.50	115.00	118.00	8.00	
2705	B	5.50	115.00	118.00	10.00	
2706	B	3.50	115.00	119.00	8.00	
2707	B	3.50	115.00	119.00	9.00	
2708	B	4.50	115.00	119.00	12.00	
2709	B	4.50	125.00	121.00	9.00	
2710	B	4.50	125.00	124.00	6.00	
2711	B	5.50	125.00	128.00	10.00	
2712	B	4.50	135.00	131.00	8.00	

Graphing

- There are many ways to graph in SPSS
- I prefer the “Legacy Dialogs” method
- It is simple and will provide basic graphs
- “Chart Builder” is more complicated, but also more flexible... it is covered in Chapter 3 (Cronk)

Histogram

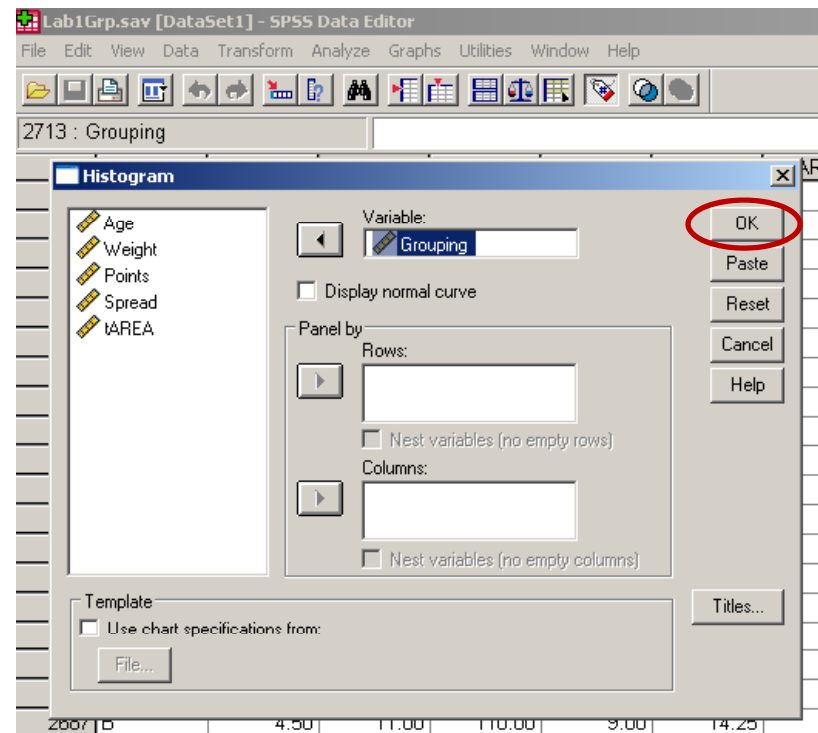


Go to Graphs, Legacy Dialogs, Histogram

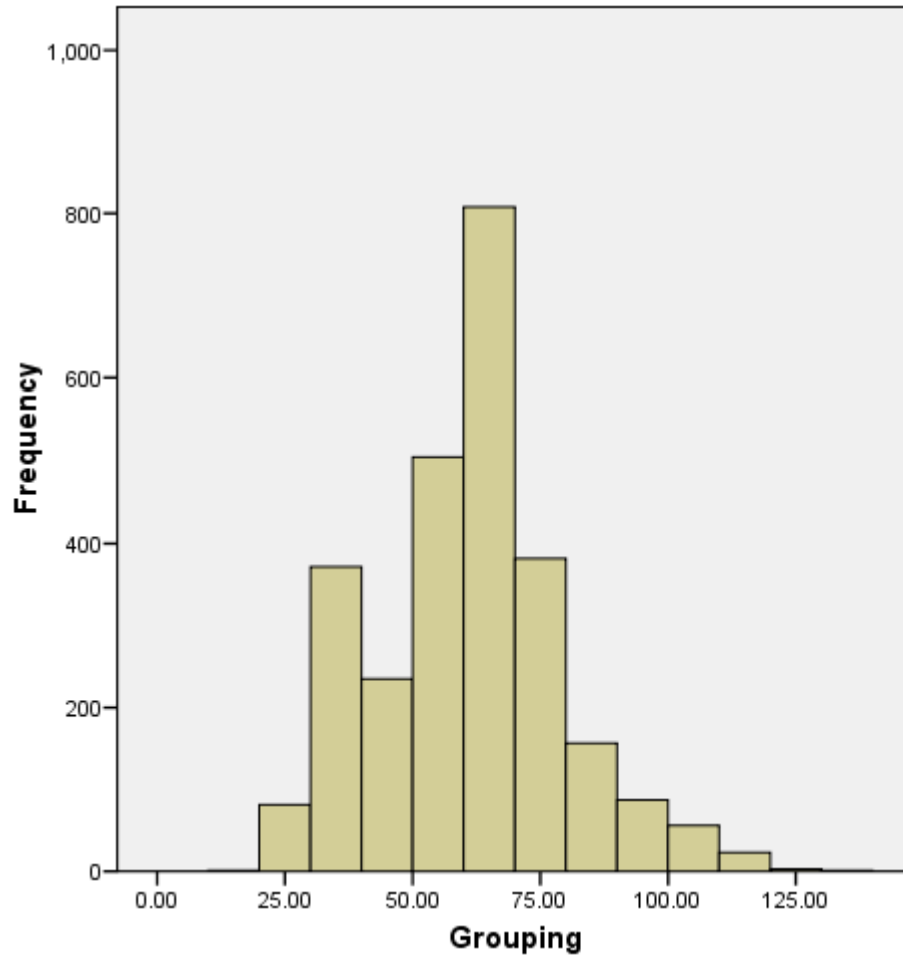


We are going to make a histogram based on our weight groups

You should make a frequency distribution on your own to provide a complementary table



Histogram



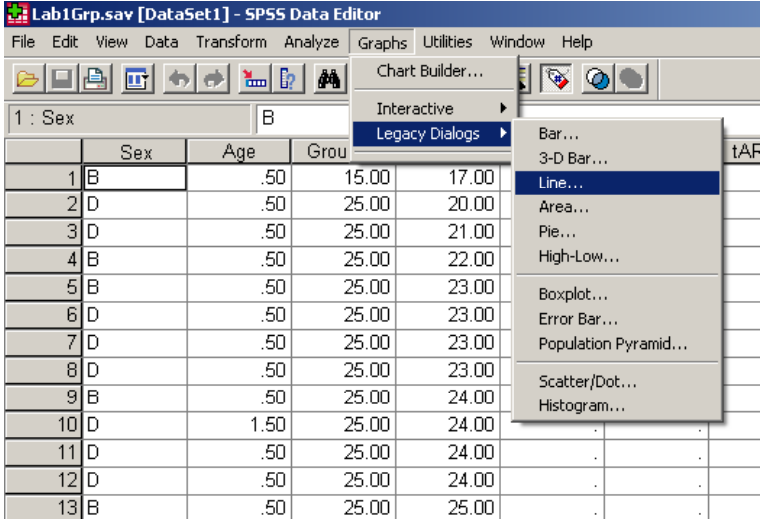
Unweighted statistics

Mean	60.3521
Std. Deviation	18.14458
Variance	329.226

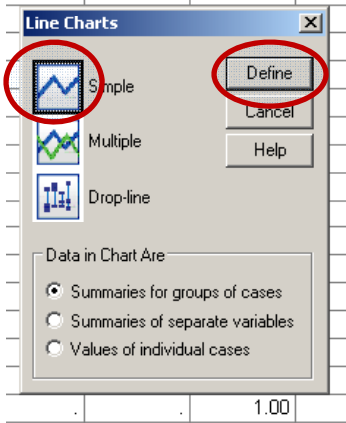
Mean =60.96
Std. Dev. =18.322
N =2,712

Weighted statistics

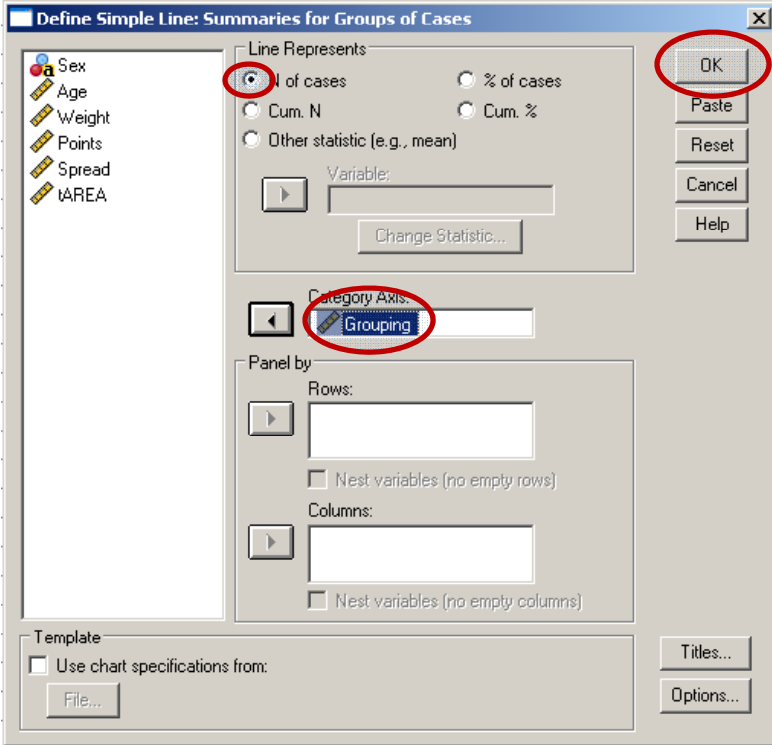
Frequency Polygon



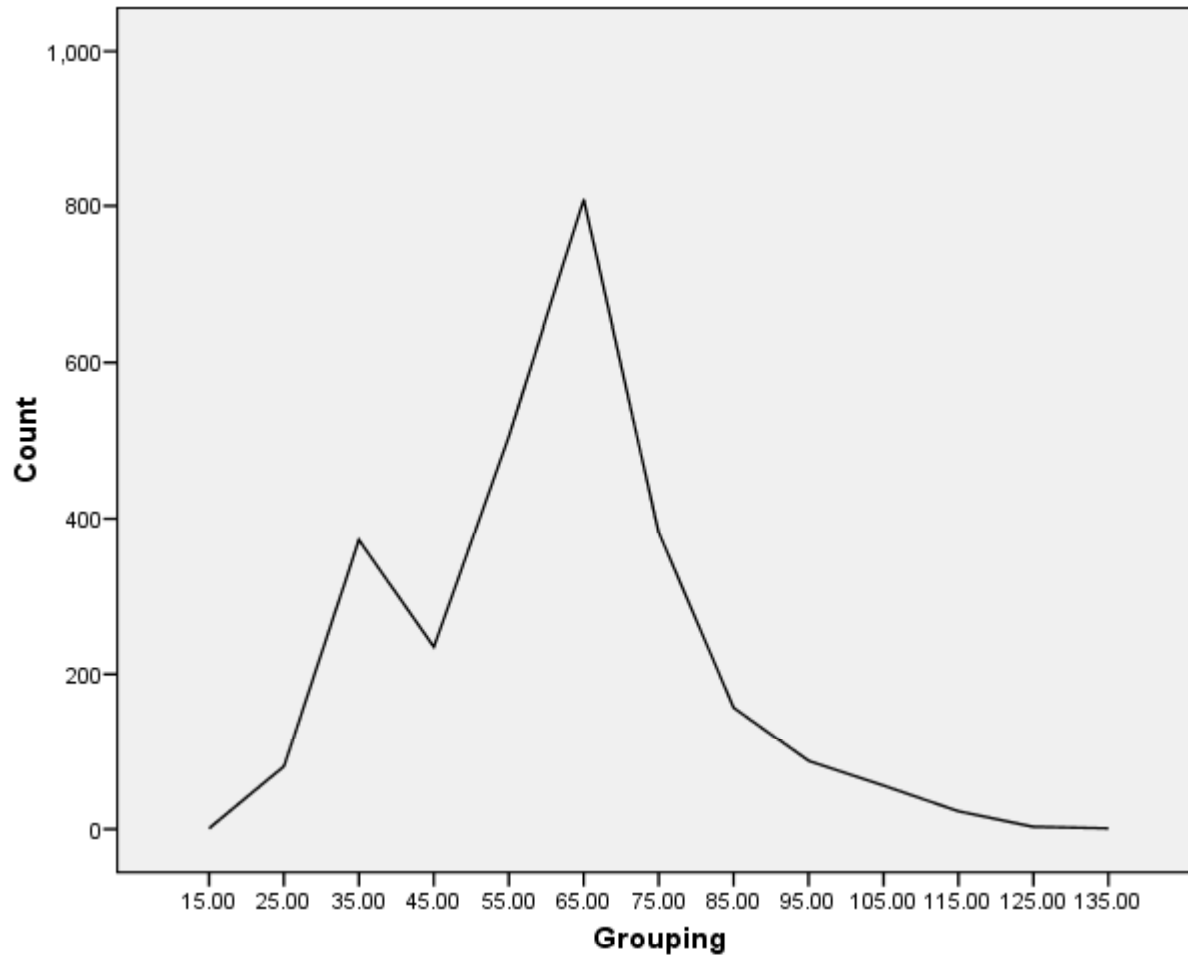
Go to Graphs, Legacy Dialogs, Line



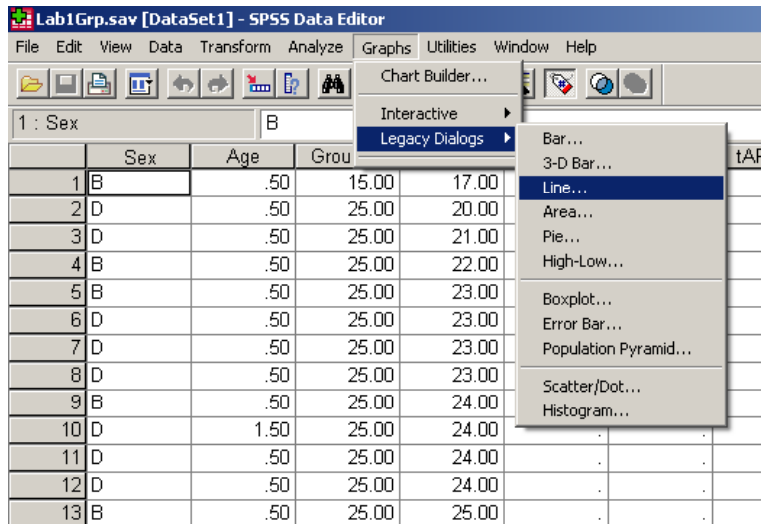
Select "simple" & push "define"



Frequency Polygon



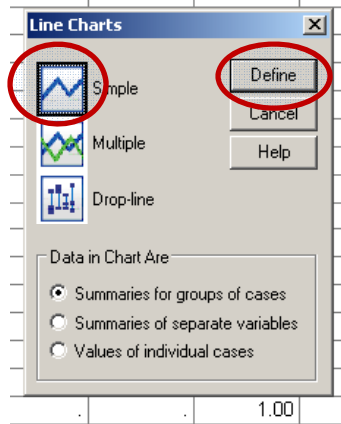
Cumulative Frequency Polygon



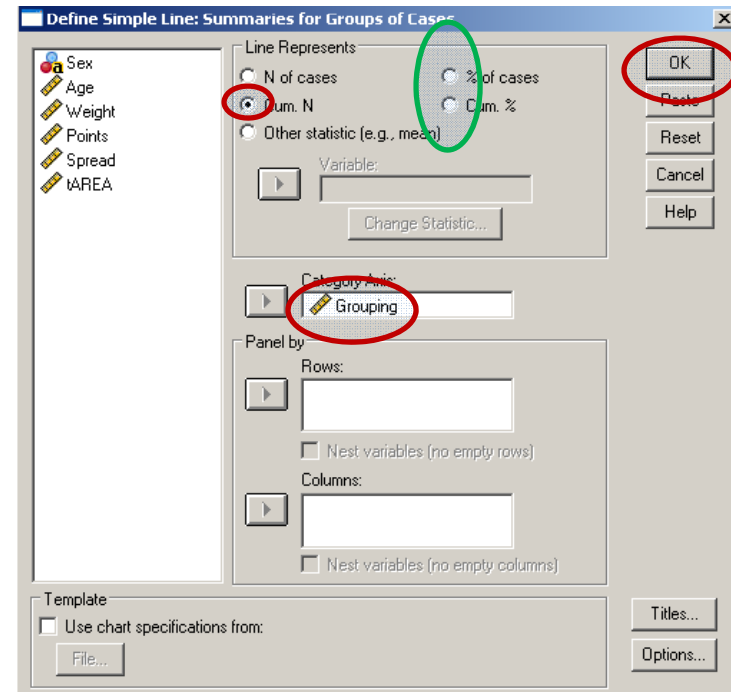
Go to Graphs, Legacy Dialogs, Line



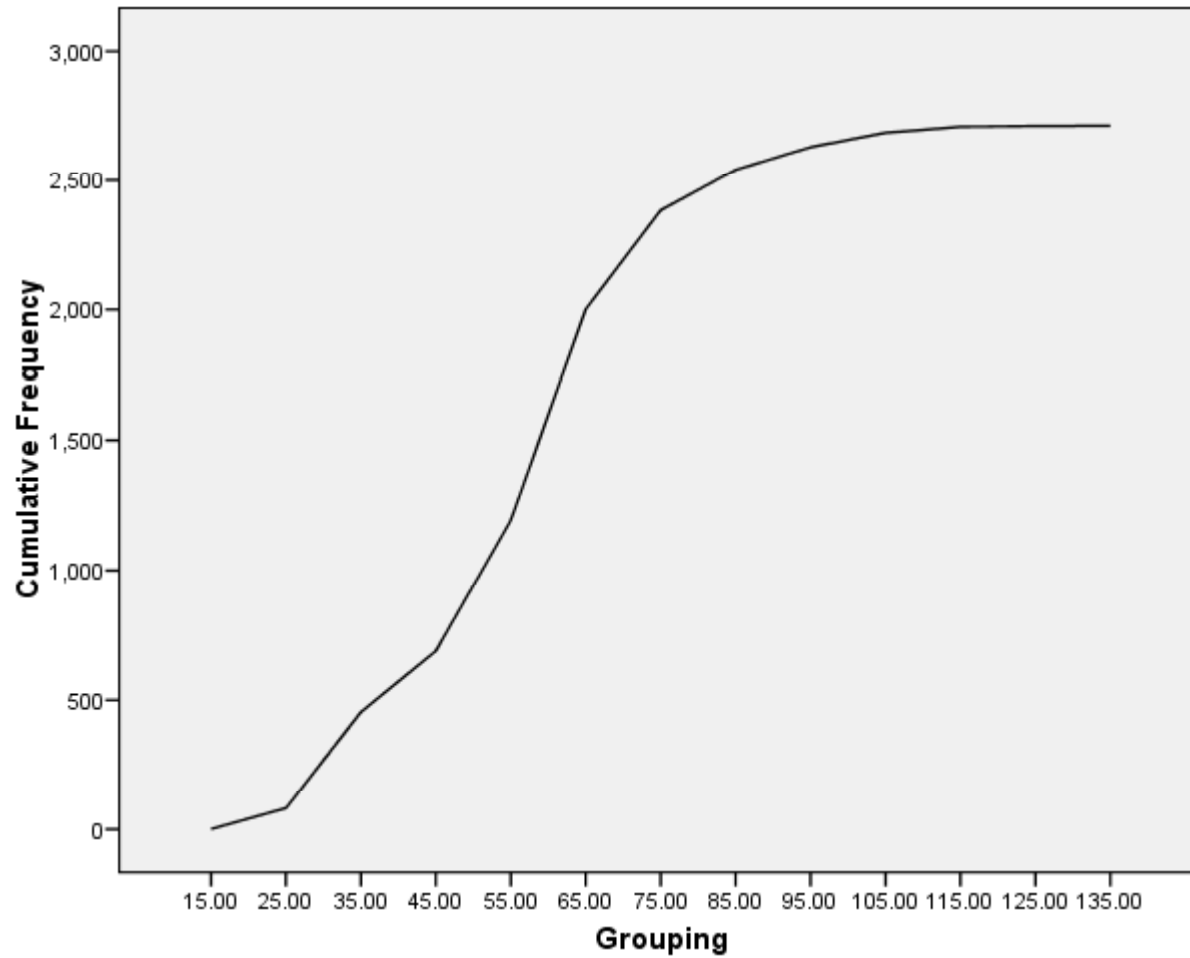
If you want relative frequency or cumulative frequency polygons click those in the green area



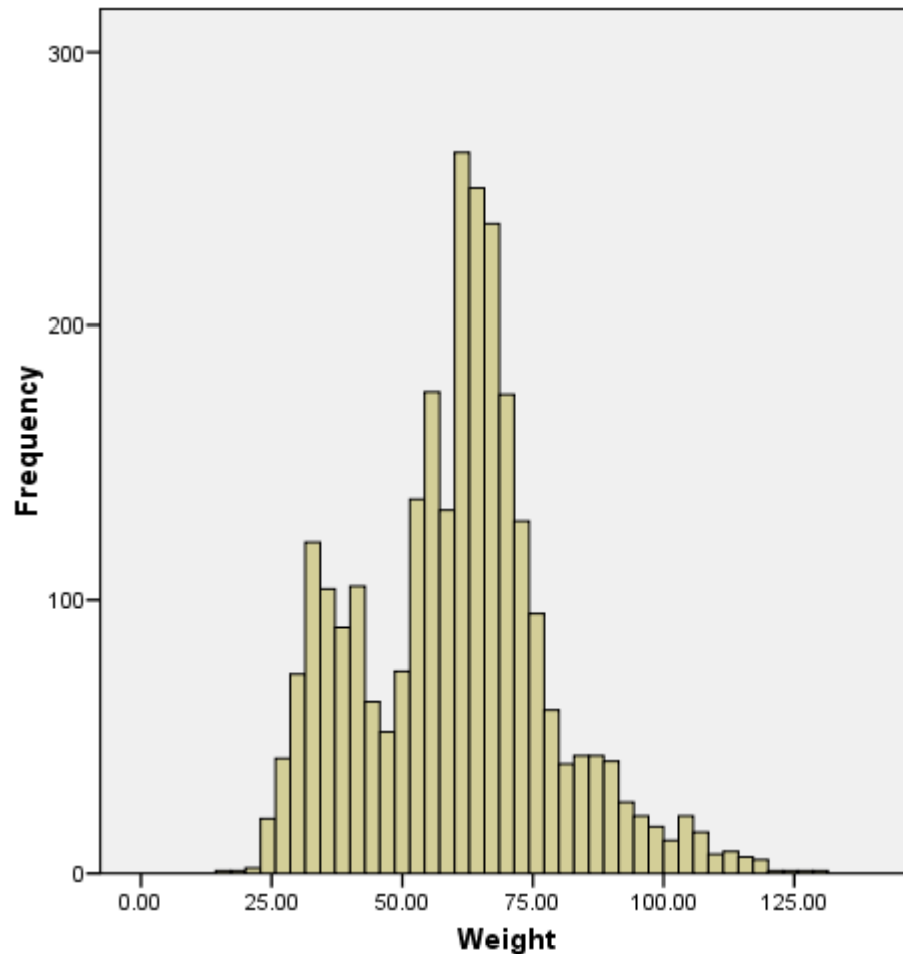
Select "simple" & push "define"



Cumulative Frequency Polygon



Histogram of our raw data



Histograms with such small groups as derived from raw data (here “weight” instead of “grouping”) may not be useful with small samples

But here we have such a large sample that it is quite meaningful

Mean =60.35
Std. Dev. =18.145
N =2,712