Unit 2 – Data Visualization
Homework

Solutions

#1. This exercise gives you practice with the ideas of Unit 2, Section 5 (*"Graphical Summaries for Quantitative Data"*)

Using the data below (*source: Daniel, 6th edition page 30, problem 2.3.5* note: *You do NOT need to buy this book.*),

7	10	12	4	8	7	3	8	5
12	11	3	8	1	1	13	10	4
4	5	5	8	7	7	3	2	3
8	13	1	7	17	3	4	5	5
3	1	17	10	4	7	7	11	8

1a. By any means you like (by hand, StatKey, Stata, R), construct a <u>stem and leaf display</u>.

BY HAND:

Stem	Leaf
0	**1 1 1 1**
0	**2 3 3 3 3 3 3**
0	**4 4 4 4 4 5 5 5 5 5**
0	**7 7 7 7 7 7 7**
0	**8 8 8 8 8 8**
1	**0 0 0 1 1**
1	**2 2 3 3**
1	
1	**7 7**

1b. **By any means you like (by hand, StatKey, Stata, R),** construct a histogram.

Solution using StatKey

__Step 1. Launch http://www.lock5stat.com/StatKey/ The following should appear:

__Step 2. At left, under Descriptive Statistics and Graphs, click on **ONE QUANTITATIVE VARIABLE**

__Step 3. StatKey provides you an example using data called tonail arsenic. To answer this homework question, you need to clear the tonail arsenic data and replace it with the homework data. From the top pale blue bar, click on **Edit Data.** You will see the following at this point:

Edit data   ✖

```
Arsenic
0.119
0.118
0.099
0.118
0.275
0.358
0.08
0.158
0.31
0.105
0.073
0.832
0.517
0.851
0.269
0.433
0.141
0.135
0.175
```

☐ First column is identifier

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only one column (or two if there is an identifier).

Ok
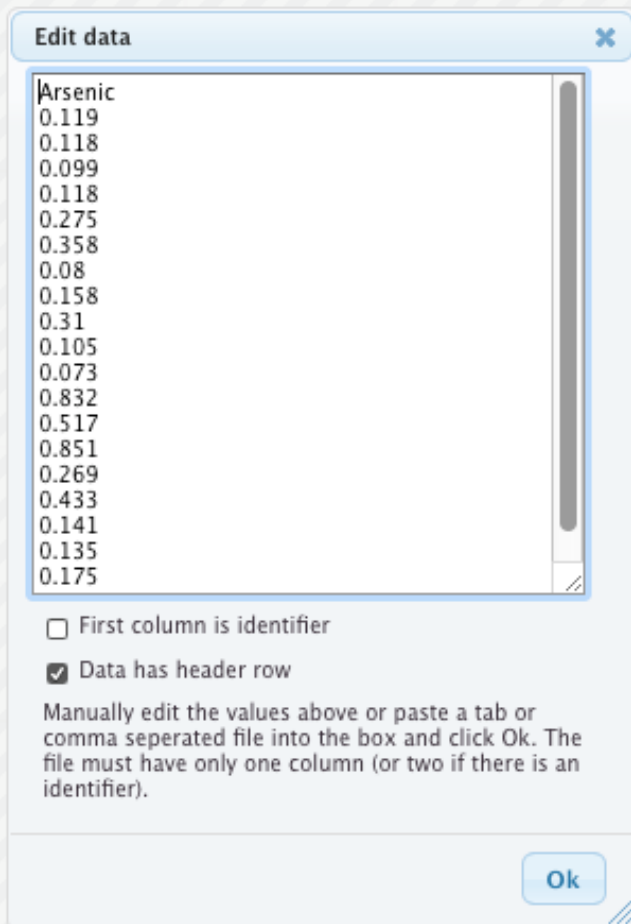
__step 5.   Edit the header row "Arsenic" so that it reads "X" instead.   Then delete the data that is below. Finally, enter the data values from question 1 in the homework. You should then have the following. At bottom right, click **OK**

EXCEL USERS – I find it faster to put my data into Excel first and then paste it into StatKey.

Edit data   ✖

X
7
10
12
4
8
7
3
8
5
12
11
3
8
1
1
13
10
4
4

☐ First column is identifier

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only one column (or two if there is an identifier).

Ok
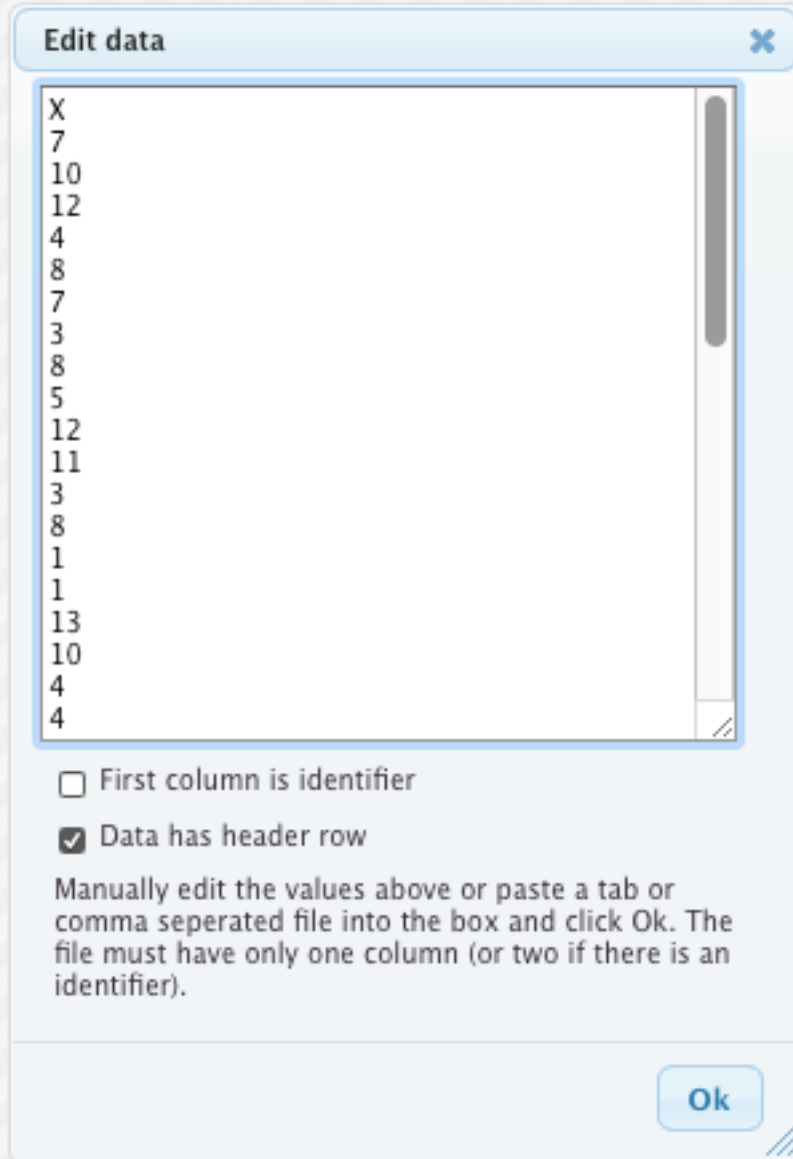
Note – The above "screen capture" shows only the first few rows.

__step 6. You should now see your data identified as "Custom Dataset" at upper right. And you may be shown a dot plot.   To get your histogram, click on the tab for **HISTOGRAM**

Summary Statistics	
Statistic	Value
Sample Size	45
Mean	6.711
Standard Deviation	4.026
Minimum	1
Q_1	4.000
Median	7.000
Q_3	8.000
Maximum	17
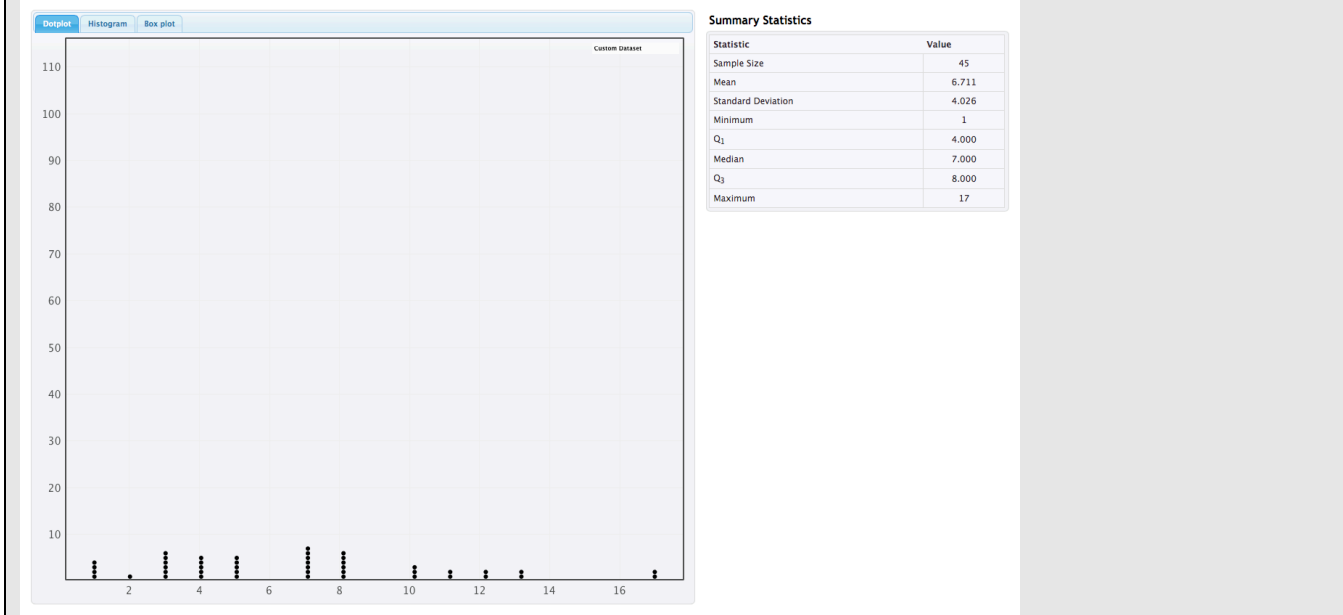
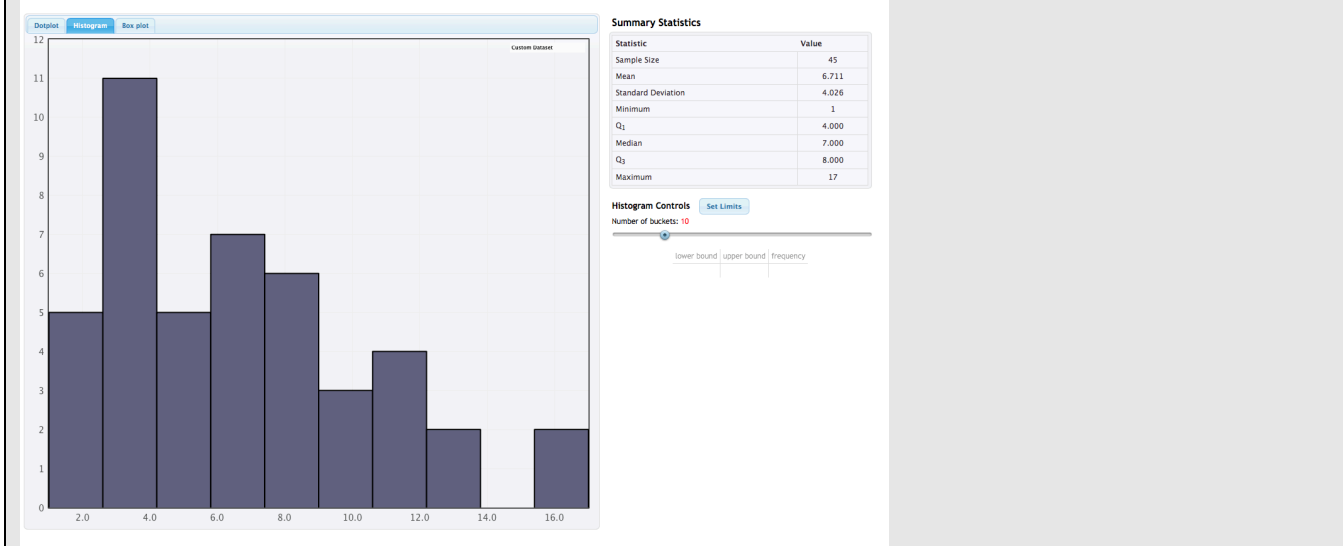__step 7. You should now see your data identified as "Custom Dataset" at upper right. And you may be shown a dot plot.   To get your histogram, click on the tab for **HISTOGRAM.** Play with the histogram controls at right as you like!

Summary Statistics	
Statistic	Value
Sample Size	45
Mean	6.711
Standard Deviation	4.026
Minimum	1
Q_1	4.000
Median	7.000
Q_3	8.000
Maximum	17

Histogram Controls   Set Limits
Number of buckets: 10

lower bound	upper bound	frequency

#2. This exercise gives you practice with the ideas of Unit 2, Section 5 (*"Graphical Summaries for Quantitative Data"*)

The following are two sets of behavioral ratings as measured by the Zang Anxiety Scale (ZAS)

SET 1 - 26 persons with a diagnosis of panic disorder:

53	51	46	45	40	35
59	51	45	60	35	
45	38	53	43	31	
36	40	41	41	38	
69	41	46	38	36	

SET 2 - 21 healthy controls:

26	26	25	25	25
28	26	26	25	
34	30	31	28	
26	34	25	25	
25	28	25	25	

2a. Construct Box and Whisker plots using the data from sets #1 and #2. **Feeling brave?** Produce a side-by-side box plot that permits direct comparison of the two groups!

Solution using StatKey
The following assumes that you have launched StatKey successfully

__step 1. Click on **One Quantitative Variable and One Categorical Variable**

Descriptive Statistics and Graphs	E
One Quantitative Variable	(
One Categorical Variable	(
One Quantitative and One Categorical Variable	(
Two Categorical Variables	(
Two Quantitative Variables	(
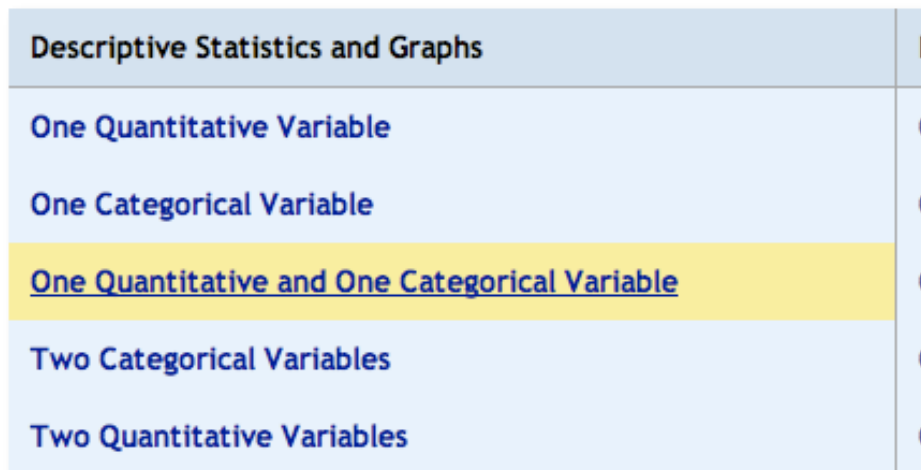
__step 2. Click on **Edit Data**
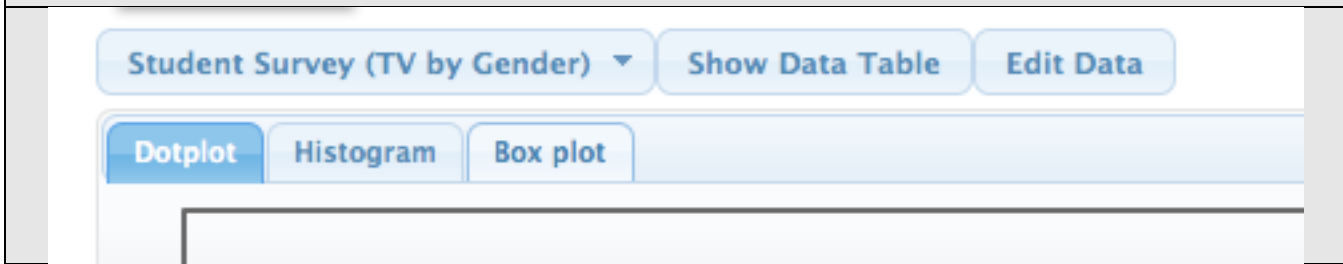
__step 3. Delete the data that is shown, taking care to keep the header "label, value"
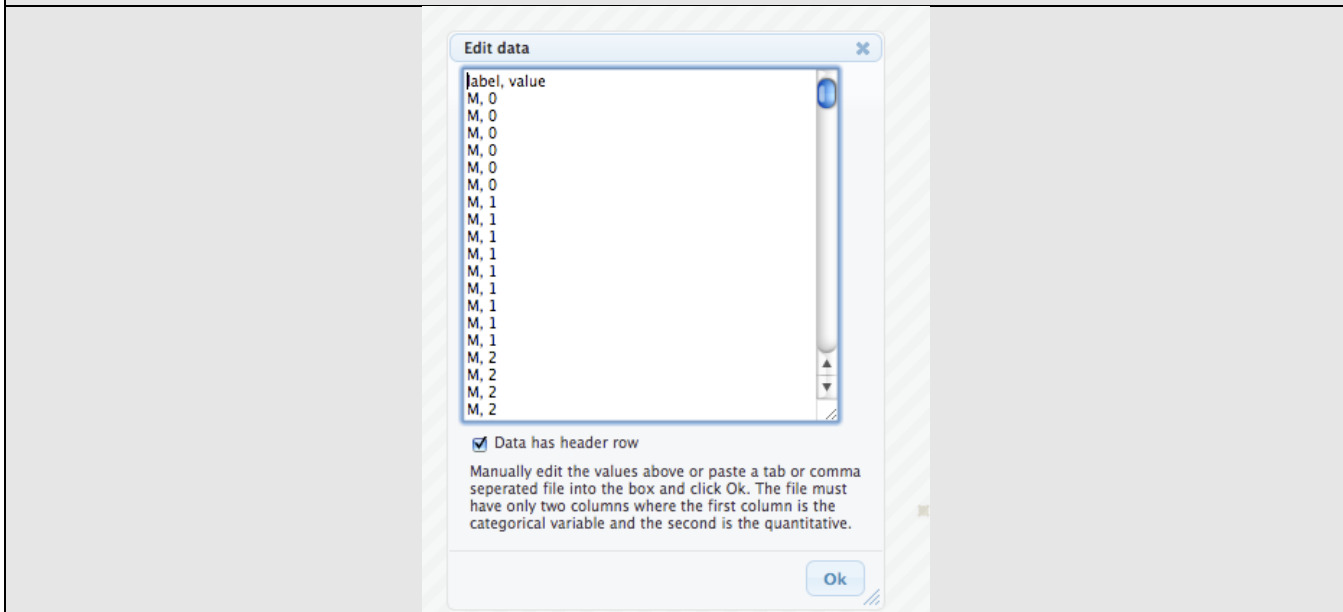
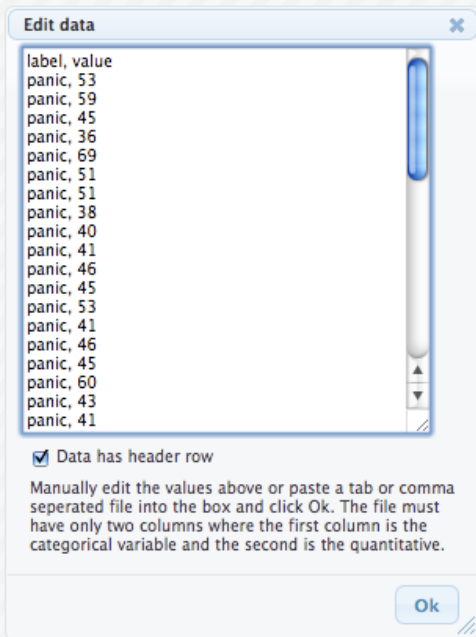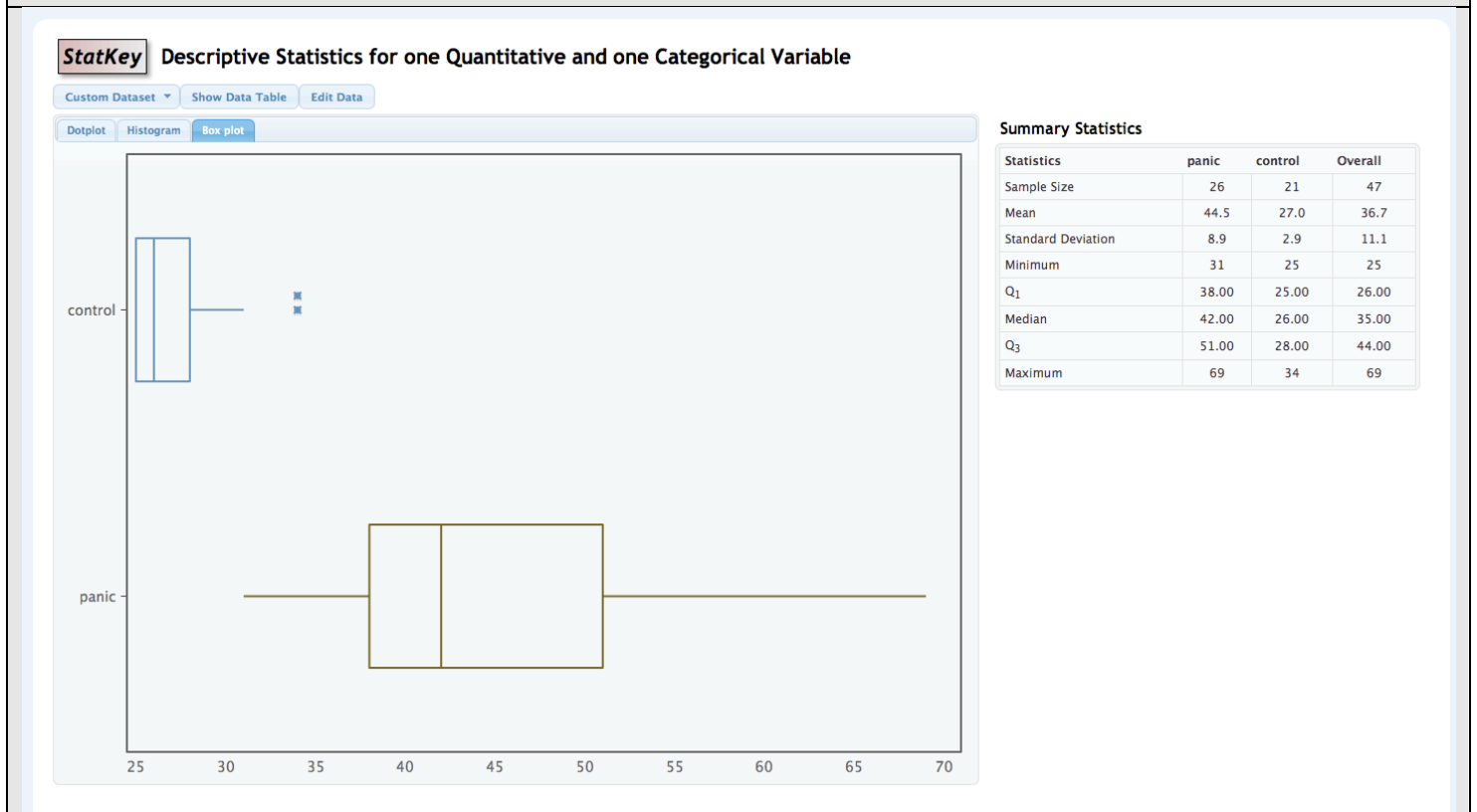__step 4 Enter the data for the panic disorder group, row by row. Then enter the data for the controls, row by row. Check your work before clicking the OK. Then click **OK.**

EXCEL USERS – Again, it might be easier and faster to put the data into Excel first and then paste it into StatKey. To do this: 1) enter group information into column "A"; 2) enter ZAS scores into column "B; 3) in column "C", in the formula box, type =concatenate(A1,",",B1) and repeat to fill out column C; and finally 4) paste your column "C" data into StatKey.

Edit data   ✖

```
label, value
panic, 53
panic, 59
panic, 45
panic, 36
panic, 69
panic, 51
panic, 51
panic, 38
panic, 40
panic, 41
panic, 46
panic, 45
panic, 53
panic, 41
panic, 46
panic, 45
panic, 60
panic, 43
panic, 41
```

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only two columns where the first column is the categorical variable and the second is the quantitative.

Ok

Note – The above "screen capture" shows only the first few rows.

__step 5. StatKey will return a dot plot. Click on **BOX PLOT**

StatKey Descriptive Statistics for one Quantitative and one Categorical Variable

Custom Dataset ▾   Show Data Table   Edit Data

Dotplot   Histogram   Box plot

Summary Statistics

Statistics	panic	control	Overall
Sample Size	26	21	47
Mean	44.5	27.0	36.7
Standard Deviation	8.9	2.9	11.1
Minimum	31	25	25
Q_1	38.00	25.00	26.00
Median	42.00	26.00	35.00
Q_3	51.00	28.00	44.00
Maximum	69	34	69

2b. In one or two sentences, compare the two groups. As we discussed, a good strategy is to first state the facts shown in the graph and then state your conclusions.

The distributions of ZAS scores are different for the 26 persons with panic disorder, compared to those among 21 healthy controls.   Compared to ZAS values among healthy controls, ZAS scores tend to be higher and more variable among persons with panic disorder.   The mean ZAS is 44.5 in the panic group versus 27.0 in controls and the median is 42.0 in the panic group versus 26.0 in controls.   The standard deviation of ZAS scores is 8.9 in the panic group versus 2.9 in controls.