Statistical Graphs in the TI-83/84

The TI-83/84 calculator can make the following types of graphs:

1) Scatterplot
2) Line Graph
3) Histogram
4) Box-and-Whisker Plot (with or without outliers)

And another type of graph we will not use.

The line graph tool can be used to plot frequency polygons as well.

1. One-variable graphs

a. Histogram

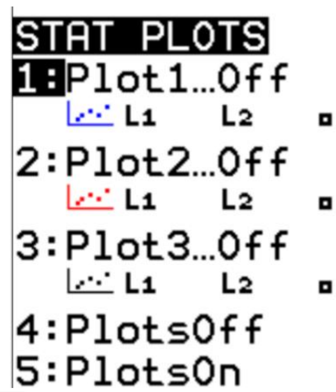
The histogram needs only a single variable. The calculator will allow you to set the class widths yourself, or allow them to be set automatically.

To begin, select STAT and then Edit to enter data into the lists.

To practice, enter the following data. Put all the data into L1.

69	64	78	74	77	100	99	60	63	90
55	85	42	73	64	77	45	76	63	77
108	77	90	86	40	104	56	67	75	88

Then select 2nd Y= to set the StatPlot settings for a histogram. Initially your screen will look like this if you've never used StatPlot before:

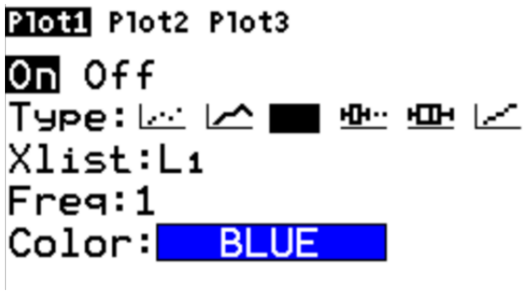


Press ENTER or 1 to set up Plot1.

On the first line OFF is highlighted. Put your cursor over ON and Press ENTER. The highlighting will move to ON.

Scroll down to the TYPE line. The third one over is the histogram setting. Put your cursor there and press ENTER to select. (Note: on older calculator screens, these six options under Type will appear on two lines.

You will still scroll right through the list. If you scroll down, it will move to the next option down (XList) for the data source.)



When you make the selection for the histogram, the bottom options will change.

XList is where the source of your data is. Since we entered the data into List 1 (L1), leave this setting alone. Also, frequency refers to how many times the data points appear. Leave this at 1 unless each item in the list represents multiple values (this is very uncommon).

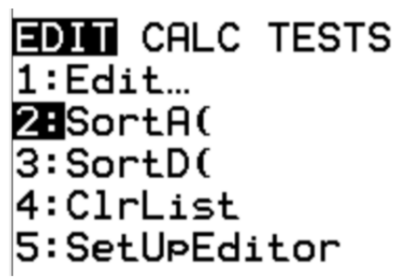
On older calculators, you won't have a color option. On newer calculators, you can leave this setting unchanged.

Now, we want to adjust the window to accommodate our data so we can see the graph. To accomplish this, we need three pieces of information from our data set:

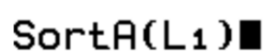
- Minimum value
- Maximum value
- Class width

The easiest way to find the minimum and maximum values, especially if the list is quite long is to sort the list. You can accomplish it this way.

Select STAT and 2 for SortA(or Sort Ascending. This will sort your list from smallest to largest.



SortA(will pop up on the main window, and press 2nd and 1 to type L1, the list we wish to short. Close the parentheses and then ENTER.



When the screen says DONE, the list is sorted. Go back to STAT and select Edit to see the sorted list.

We can see by scrolling through the list that the Minimum is 40 and the Maximum value is 108.

To calculate the class width, we need to determine how many classes we want, in other words, how many bars do we want on our histogram? There is no hard-and-fast answer to this question. The rule of thumb is between 5 and 20. Fewer for smaller lists, more for larger ones. I'm going to choose 7 for this exercise.

To calculate the class width, you need to find $\frac{MaximumValue-MinimumValue}{NumberOfClasses}$. Here, that's going to be $CW = \frac{108-40}{7} \approx 9.71$ ... In general, never round your class width down. Always round UP, or the final value won't fit into the last class and you'll end up with one more class than you intended. Note that depending on how spread out your data is will change how far you should round. If you round too much, you will reduce the number of classes from your original intention. Our data is spread out enough that I'm going to round up to the nearest integer: 10.

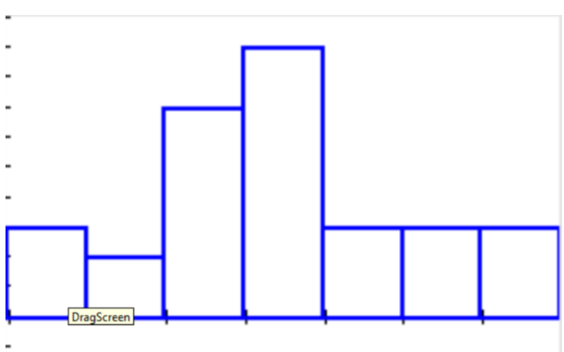
Textbooks usually choose examples that have integer values, even if they are theoretically continuous, and so class widths are chosen to mimic this, with class boundaries at the midpoint between integers, and class widths always integer values. The reality is rarely this tidy, as we will see later.

To set the Window, choose WINDOW and set the values as shown on the screenshot, with the XMIN = $MinimumValue - 0.5$ and XMAX=$MaximumValue + 0.5$ (give or take; here I used 109.5 because it's a whole number of classes) and XSCL=$ClassWidth$.

You will want the YMIN to be a bit smaller than zero so that you can see the axis clearly. You will probably need to fiddle around with YMAX to find a value that is high enough to see the tops of all the bars, but not so high that the whole graph looks smooshed.

```
WINDOW
 Xmin=39.5
 Xmax=109.5
 Xscl=10
 Ymin=-1.50345
 Ymax=10
 Yscl=1
```

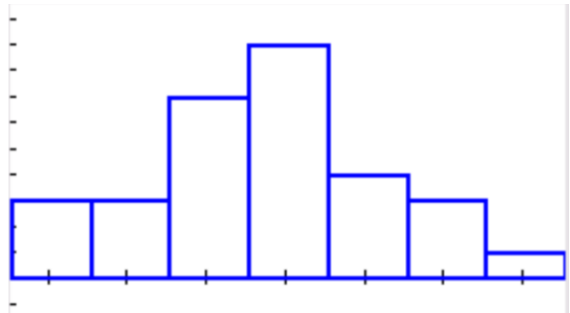
These setting produce the graph shown below upon hitting GRAPH.



The tick-marks on the left are in units of 1 (YSCL setting). If you press TRACE you can see the heights and class boundaries are shown at the bottom of the screen. Scroll right to visit each class. For instance the middle class (the tallest) has $n = 9$ (nine measurements are included in the class), and boundaries of $min = 69.5, max < 79.5$.

Alternatively, you can allow the calculator to choose an optimal number of classes for you, as well as set the window. To use this option press ZOOM 9 for ZoomStat.

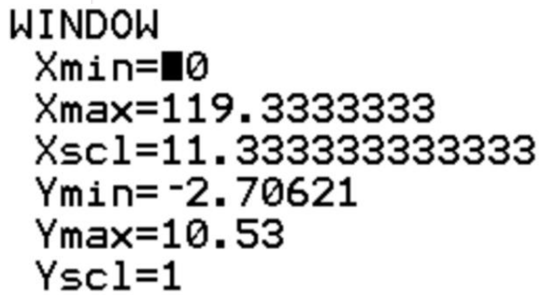
The graph produced is shown here.



Notice how much nicer this looks, even though there are still 7 classes?

The window settings are interesting:

+FThe lowest value is 40 (our minimum value), and it looks like the calculator was going for 6 classes since $\frac{108-40}{6} \approx 11.333\ldots$, but because the calculator rounded down, the last measurement of 108 didn't quite make it into class 6, and so, it is alone in the last class.



The other thing that's interesting is that the class midpoints are marked by the ticks on the x-axis, rather than the boundaries.

Despite what textbooks usually tell us, histograms are a little bit of an art.

b. Box-and-Whisker Plots

Box-and-whisker plots (or boxplots) are also build from one-variable numerical data, so I will use the same data we used for the histogram. There are actually two types of boxplots our calculator can create: one that marks outliers, and one that does not. We will construct the one that does not mark outliers first, and then we'll return to the one that does later.

i. A boxplot from the 5-number summary.

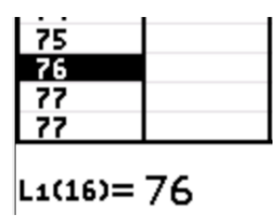
At a minimum, constructing a boxplot requires 5 numbers: the median, the first quartile, the third quartile, the minimum value and the maximum value.

We figured out the last two in the histogram example above:
Minimum value: 40
Maximum value: 108

To obtain the median, we need to sort the list (we did this in part a). The median is the "middle value" in the sorted list. If the list has an odd number of elements in it, then it will be an actual value on the list. If the list has an even number of elements in it, then we average the two closest to the middle. Since we have 30 items in the list, 1-15 are in the bottom half of the list, and items 16-30 are the top half. We need to average the 15^{th} and 16^{th} elements in the list.

Select 

and Edit, and scroll through the list to find those values. Notice as you scroll through the list the bottom line L1(#)=# changes. The first number tells you the item number in the list you are on, and the second one the value.



We can see from the screenshot that element 15 has the value 75, and element 16 has the value 76. The average of these is 75.5. This is our median.

To get the two quartiles, they mark the quarter-points of the list, or the half of each half. Repeat the median calculation on the just elements 1-15, and then again on 16-30.

Each half has 15 elements, so if we take the 7 smallest elements and the 7 biggest elements, that leave the middle one (element number 8) to be the middle value. Turns out this value is 63. The top half also has 15 elements, so we need 8 elements into that set or skip over 16, 17, 18, 19, 20, 21, 22 (the smallest 7 in the top half), and element 23 is the 8th value in that top half (and as you can see there are still 24, 25, 26, 27, 28, 29, 30—or 7 elements above that one). So the third quartile is the 23rd item in the list, which corresponds to a value of 86.

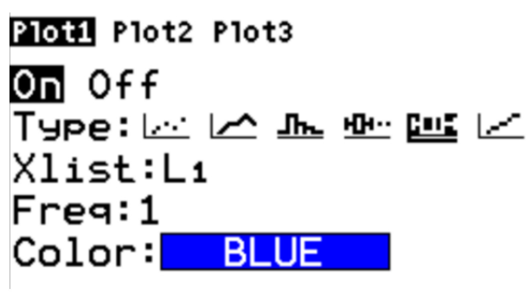
Thus the 5 number summary is:
Minimum: 40
First Quartile: 63
Median: 75.5
Third Quartile: 86
Maximum: 108

We could plot this by hand: the box extends between the first and third quartiles, with the median as the middle bar. The whiskers extend to the minimum on the left, and the maximum on the right. Be sure to use a correct scale.

We could also let the calculator do this.

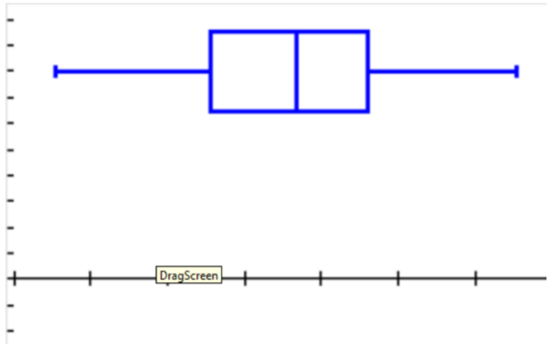
Select 2nd and Y= . We will edit Plot 1, so press ENTER .

Leave the plot ON, but scroll down to the TYPE line, and cursor over to the 5th item in the list. There are two that look quite similar, but we want the second one of those as shown.



We are stilling getting our data from List1 (L1), and leave the frequency set at 1.

Now press ZOOM 9 , so that the window adjusts automatically.



The y-axis doesn't mean anything here. It will adjust automatically to accommodate more boxplots if you choose to compare multiple data sets.

Selecting 

, will take you through the values that mark each of the 5 key points on the graph, which should correspond to the values we obtained by hand earlier.

ii. Boxplots with outliers

If we want to plot outliers on the boxplot, by hand, we will need to find the IQR or interquartile range. This is the distance between the first and third quartiles. Here, that's $86 - 63 = 23$. For an element in the list to be classified as an outlier, we need to see if there are any elements more than 1.5 time the IQR past either quartile.

$$1.5IQR = 34.5$$

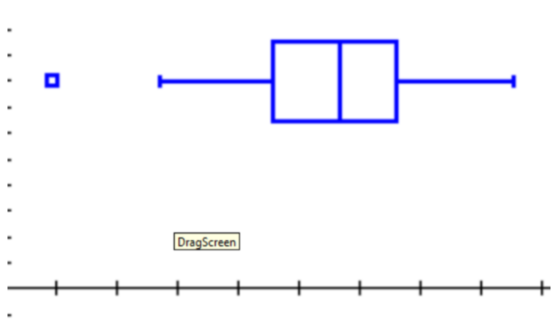
$$Q1 - 34.5 = 63 - 34.5 = 28.5$$
$$Q3 + 34.5 = 86 + 34.5 = 120.5$$

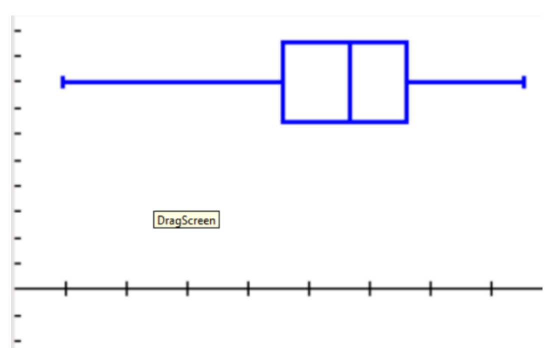
If we have any elements that are less than 28.5 or more than 120.5, they would be considered outliers (very unusual values). Our list doesn't have any of these, so if we got back to StatPlot and select the 4th option under TYPES as shown:

The two graphs will look exactly the same. (Notice that there is an extra line for how we'd like to label the outliers on the graph.)

To see what happens, we need to change the data in the list. What if our smallest value was 22 instead of 40?

The two boxplot settings will now produce the following graphs:





As you can see, a bit different. The left tail is longer on the 5-number summary version (the right graph, TYPE #5), whereas, the tail is about the same length on the left graph (the outlier graph, TYPE #4), but the dot marks the unusual value. Outliers are by definition unusual, and so this could be a legitimate low-probability event, or it could be a symptom of a processing error. These value can affect results substantially, and so they are sometimes removed prior to further analysis.

2. Two-variable graphs

a. Line graph

Line graphs are typically used with time data (sales over time, or grades over time, for instance), where the x-values are time related data, and the y-variable is a numerical measurement. Line graph tools can also be used to create frequency polygons.

Let's look at the case of a frequency polygon, since frequency distributions come up more often in our textbook.

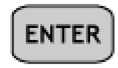
Consider the following frequency distribution of the ages of U.S. Presidents at inauguration.

Age(in years)	Frequency
42	1
43	1
46	2
47	2
48	1
49	2
50	1
51	5
52	2
54	5
55	4
56	3
57	3
58	2
60	1
61	3
62	1
64	2
65	1
68	1
69	1

First, enter the data into the calculator. Put the ages in List1 (L1) and the frequency in List2 (L2). Recall, you enter the lists by selecting 

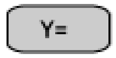
then Edit. If you need to clear data

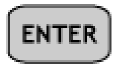
from the list, Scroll UP to the name of the list and select 

, then 

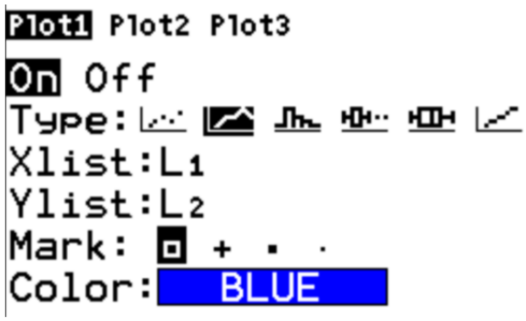
. (Don't hit Delete!) Then enter the data into the blank list. Then move over to L2 by scrolling right, and entering the data.

Press 



and 

, to set up Plot1. Scroll down to the TYPE line, and select the second graph type in that row. This is the line graph tool.

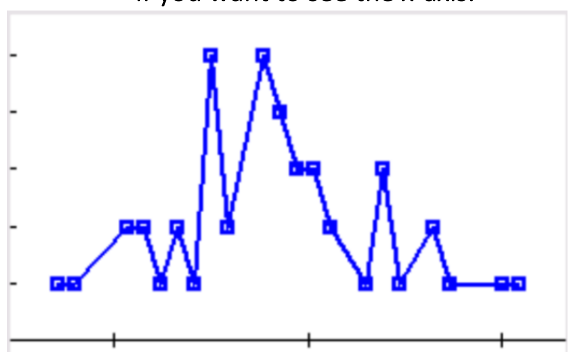


The XList is the values that were measured (or time in a traditional line graph), and the YList are the frequencies here. We have entered the data to match the defaults so just press 

to adjust the window. You may want to adjust your YMIN value under 

if you want to see the x-axis.

You can also enter values outside the data range and set the frequency to be zero to close the polygon with the x-axis. This process works similarly if you convert the counts to frequencies.



b. Scatterplots

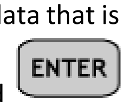
A study looked at the relationship between two variables related to soil quality and plant growth. Their research produced the following data.

x	3	8	10	11	13	16	27	30	35	37	38	44	103	142
y	4	7	8	8	10	11	16	26	21	9	31	30	75	90

To create a scatterplot from hand, the process is similar to a line graph, only the dots are not connected sequentially. And you may have more than one y-value associated with a given x-value.

In the calculator, enter the data under 

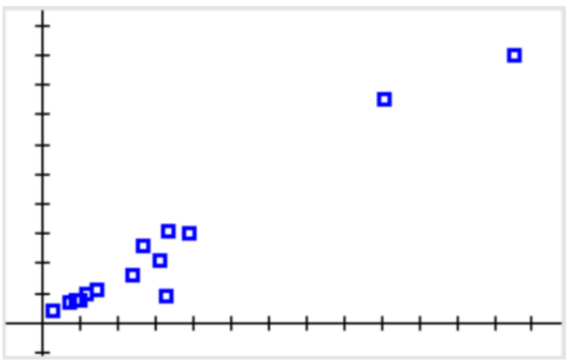
and Edit. As with the above section, you may need to clear the data that is already in the list. You can do this by Scrolling up, and then selecting 

and 

. Then repeat on List2 (L2). Put the x-values (explanatory variable) into L1 and the y-values (response variable) in L2.

Press 2nd Y= and ENTER to set up Plot1. On the TYPE line choose the first graph type for the Scatterplot.

We've entered the data to correspond to the defaults for the calculator, so now press ZOOM 9 to adjust the Window and see the graph.



We will discuss in a later section linear regression lines, and then we'll add to this, and discuss interpretations of this graph type.

Practice Problems. Graph each of the following datasets (or distributions) with the indicated graph type.

1. Create a histogram of the dataset shown below.

12	**31**	**35**	**40**	**41**	**44**	**47**	**48**	**51**	**53**
54	55	55	57	58	62	65	73	81	119

2. Create a boxplot for the dataset shown below. Are there any outliers?

81	90	90	92	93	94	95	96	96	102
103	105	106	111	113	116	119	122	122	123
125	125	128	137	158					

3. Create a line graph of the following data.

Year	1999	2000	2001	2002	2003	2004	2005	2005	2006
Graduation Rate	50.3%	55.5%	55.4%	55.9%	60.6%	65.0%	71.9%	72.6%	91.0%

4. Create a scatterplot of the data below.

Name	Test Mark (x)	Homework Mark (y)
Rob	61	35
Thomas	95	50
Mark	44	5
Wanda	93	50
Judy	63	15
Haydn	80	34
Barbara	62	16
Karen	95	50
Marilyn	65	7
Phil	88	38