

Chapter 2 – Displaying and Summarizing Quantitative Data

The three primary rules of data analysis are:

1. Make a picture.
2. Make a picture.
3. Make a picture.

TI-calculators can help with this, although they cannot make bar graphs or pie charts for categorical data or dotplots and stem-and-leaf displays. All of these are fairly easily done by hand (at least for small data sets; for larger ones, use a full computer package).

This chapter will discuss histograms for numeric data. Other plots of quantitative data will be discussed later. Histograms are examined for shape (skewed right/left or symmetric), center, and spread. They also tell us whether or not a distribution is unimodal (one-humped) or multi-modal (many humps). We will also look at computing summary statistics for a set of data.

HISTOGRAMS

Histograms are connected barcharts. Since the data are presumed to represent particular observations on some continuous portion of the real number line and since order here matters, bars are always displayed connected to one another (unless there happens to be a gap in the values). A good histogram has equal bar widths, high and low ends not too dramatically different from the maximum and minimum values of the data, and intervals which “make sense.” As we will see, they are useful in highlighting major features of the distribution of a single variable or for comparing two distributions (if done properly); they also have a capacity for “artiness” since their shape can change, depending on the choice of beginning values and bar widths.

In this first example, we will use the data on pulse rates of 24 women to create the histogram shown in figure 4-3 of the text. For convenience, the data are below.

88	56	80	60	76	72	68	80
64	80	84	64	68	72	80	76
72	76	84	76	72	68	68	64

L1	L2	L3	Z
88			
56	█		
80		---	
60			
76			
72			
68			
L2(1)=			

The first step in making a histogram is to enter the data. The data have been entered into list L1; the first few values are seen in the accompanying figure.

TI-83/84 Steps to Create a Histogram

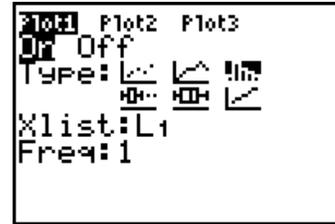
The next step is to define the plot. This is done by pressing $\boxed{2\text{nd}}\boxed{Y=}$ (Stat Plot). You will see the screen at right. Notice that there are three plots which can be displayed at any one time. For most purposes, there should be only one turned “on” at once. Notice here Plot1 is On and Plots 2 and 3 show Off. Scrolling down the menu are options 4 and 5 that turn all plots off or on with a single command. Selecting either of these will transfer the command to the home screen. Executing it requires pressing $\boxed{\text{ENTER}}$.

STAT PLOTS			
1	Plot1...On		
	▬ L1 1		
2	Plot2...Off		
	▬ L4 1 +		
3	Plot3...Off		
	▬ L3 1 □		
4	PlotsOff		

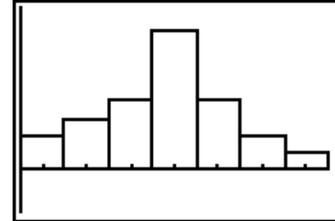
Press $\boxed{\text{ENTER}}$ to select Plot1. The cursor should be blinking over the word On. If On is not already highlighted, press $\boxed{\text{ENTER}}$ to move the highlight. Notice that there are six graphics types. Histograms are the third choice. Pressing $\boxed{\downarrow}$ will move the cursor to the first plot type. Use $\boxed{\rightarrow}$ to move the cursor to the histogram figure. Press $\boxed{\text{ENTER}}$ to move the highlight.

1	Plot1	Plot2	Plot3
	On	Off	Off
Type:	█	▬	▬
	▬	▬	▬
Xlist:	L1		
Ylist:	L2		
Mark:	█	+	.

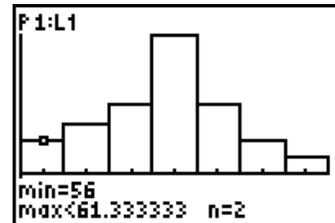
At this point, your screen probably looks like this. We're ready to display the graph, since our data was in list L1 and each data value had frequency 1 (represented one observation). If you want to graph data in other lists, move the cursor to `Xlist:` and enter the list name (`2nd` n, where n is the number of the list). We'll talk more about frequencies later. Notice if you move the cursor to `Freq:` it will flash as `█`. If you need to change this back from something else to a 1 you will need to press `ALPHA` before typing the 1.



The easiest way to display a histogram (or any statistics plot) is to press `ZOOM` 9 (`Zoom Stat`). The resulting graph is seen at right. Notice the y-axis penetrating one of the bars. The x-axis "floats" a little way up from the bottom of the screen. This is so that values as seen in the next picture do not interfere with the plot.



To see exactly what the graph shows, press `TRACE`. A blinking cursor will appear in the leftmost bar. At the bottom of the screen the minimum and maximum values for the bar, and number of observations in the bar are displayed. This first bar goes from 56 to 61.3333. There are two observations in this interval, indicated by the `n=2` at the lower right. Pressing the right arrow key (`▶`) will allow you to continue through the graph.

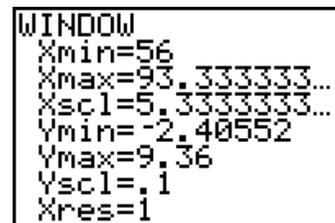


At this point, we can see the distribution of pulse rates appears to be unimodal and relatively symmetric (bars fall roughly equally from the center peak.) We see the center is around a 75 beats per minute with 8 observations in that bar.

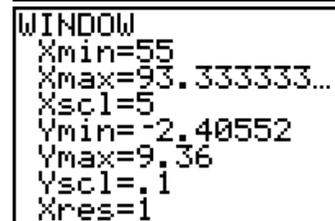
There is a downfall to using simply `ZOOM` 9 for histograms. Look at the first interval. It doesn't really make sense in a natural way. The bar width represents a difference between the low and high ends of each bar of 5.333333... which is unnatural. We'd like to fix this.

Manipulating Windows

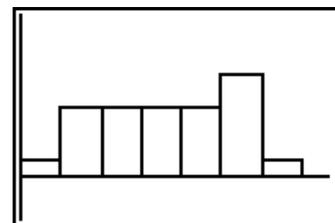
To force particular minimums, maximums and scaling we will press `WINDOW`. This displays a screen like the one at right. Notice the `Xmin` was the smallest value shown on the plot and `Xscl` was the bar width. These are quantities we'd like to change. You generally won't have to change any of the `Y` variables here (unless a scaling change loses the top of a bar – then increase `Ymax`). Another reason to possibly change `Y` variables is to increase resolution.



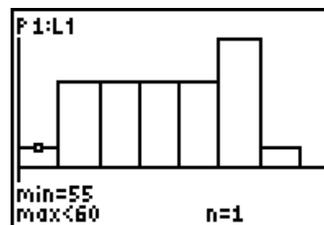
Change `Xmin` to 55 and `Xscl` to 5 (sounds pretty reasonable, and reproduces the scaling used in the text.)



To display the new graph, press `GRAPH`. NEVER press `ZOOM` 9 after changing a window. You'll just go back to the one you had before! This looks better, but changing the scaling left some room at the top of the graph. Let's change the window and lower the `Ymax` to 7.



Here's the new graph (remember, press **[GRAPH]** after changing the window again). This graph points out one of the pitfalls of histograms as mentioned above – you notice that in this scaling, the central peak seen in the default histogram has been lost. This graph looks much more uniform (no real peak in the graph).



TI-89 Histograms

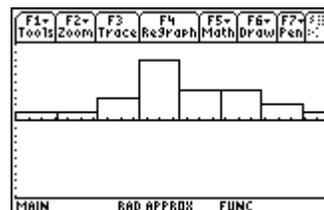
Once data has been entered in the Statistics/List Editor, the next step is to define the plot. This is done from the Statistics Editor by pressing **[F2]** (**Plots**) followed by **[ENTER]** to select **Plot Setup**. You will see the screen at right. Notice that there are nine plots which can be displayed at any one time. For most purposes, there should be only one active (checked) plot at once.



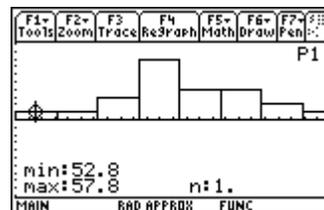
Press **[F1]** to select defining **Plot 1** since it was highlighted. The cursor should be blinking over the plot type. If the plot type is not already set to a histogram, pressing the right arrow gives a menu of five plot types. Move the cursor to highlight choice **4:Histogram** and press **[ENTER]** to select it or press **[4]**. Press the down arrow to the box labeled **x**. Press **[2nd]**-**[VAR-LINK]** to get the list of list names. Move the cursor to highlight the one you wish to use, then press **[ENTER]** to select it. The TI-89 then wants the histogram bucket width which is the bar width. Press the down arrow to move to this box. You may need to use trial and error to get a good picture. Here, I have set the bar width to 5. Since we don't have a separate list of frequencies, **Use Freq and Categories** is set to **NO**. Press **[ENTER]** to complete the plot definition. You will be returned to the Plot Setup menu.



The easiest way to start displaying a histogram (or any statistics plot) is to press **[F5]** (**Zoom Data**). The resulting graph is seen at right. Notice the y-axis penetrating one of the bars. The x-axis "floats" up from the bottom of the screen. This is so that values seen in the next picture do not interfere with the plot. In many cases, the TI-89 will not show the full heights of the bars due to the tabs at the top of the screen. It does not always get the windowing correct. We don't see the full heights of the bars. We'll change that later.



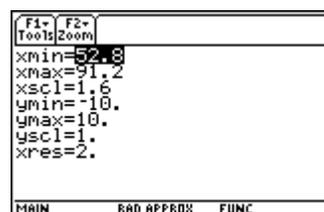
To see exactly what the graph shows, press **[F3]** (**Trace**). A blinking cursor will show in the first bar at the left of the graph. At the bottom of the screen are displayed the minimum value and maximum value for the bar, and the number of observations in the bar. This bar goes from 52.8 to 57.8. There is one observations in this interval. Pressing the right arrow key (**[▶]**) will allow you to continue through the graph seeing the interval ranges and numbers of observations in each interval.



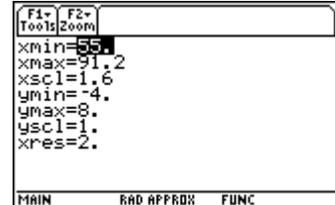
At this point, we can see the distribution of pulse rates appears to be unimodal and relatively symmetric (bars fall roughly equally from the center peak.) We see the center is around 73 beats per minute. There is another downfall to using simply **Zoom Data** for histograms. Look at the intervals. They really do not make sense in a natural way. Well need to fix this.

Manipulating Windows

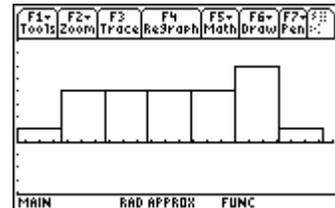
To force particular minimums, maximums and scaling we will press **[♦]****[F2]** (**Window**). This displays the screen at right. Notice the **xmin** was the smallest value shown on the plot; **xmax** is the largest. **ymin** and **ymax** are analogous. **xsc1** and **ysc1** are the distances between axis "tick marks." If your plot has lost the tops of the bars, you will need to increase **ymax** until they can be seen.



Change x_{min} to 55 (just slightly smaller than the smallest data value) y_{min} to -4 and y_{max} to 8. y_{min} is set low so that the legends which appear after pressing **Trace** don't obscure portions of the graph, but don't set it so low that most of the window area is blank space.

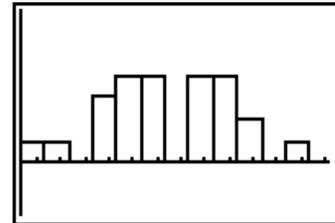


To display the graph with the new window settings, press \blacklozenge **F3**.

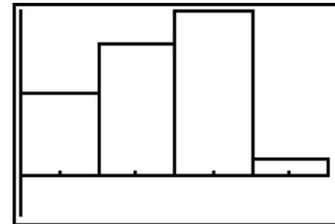


Manipulating the number of bars

The number of bars in a histogram is controlled by the value of X_{scl} on the 83/84 calculators and by the bucket width on the 89 series. In this graph, X_{scl} (the barwidth) was set to 3. The distribution looks unimodal again, but there is an interesting gap in the center. There are also gaps between the two smallest bars and the highest bar. It looks like we might have outliers.



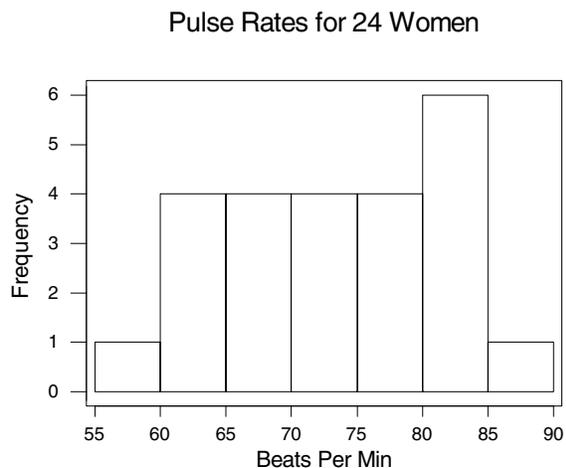
It's possible to have too few bars. Here, X_{scl} has been set to 10; y_{min} is -3 and y_{max} is 10. (These were changed for picture resolution.)



You will need to use your own judgment to decide how many bars to include and where. Your instructor may give you some guidelines. One rule of thumb for many years was to have somewhere between 5 and 20 bars; for most small data sets dividing the number of observations by 5 gives a good estimate of how many bars will give a decent picture.

“Printing” the Picture

Unfortunately, calculators do not have printers. To make a hard copy of the graph once you are satisfied, use the **TRACE** key to examine the entire graph. Make a picture of the histogram, clearly labeling each axis and giving the graph a title. Remember that the intervals given are the endpoints of the intervals. Label them as such. When you are finished, you should have a picture like the one at right. If you have TI-Connect software, you can use the screen capture application to save the picture on the computer for printing directly from the application, or use copy and paste to include the graph as a part of a word processing document. However, notice in the plots above that the screen capture does not include proper axis labels or titles!



HISTOGRAMS WITH FREQUENCIES SPECIFIED

Sometimes data are given in the form of tables with both the data value and the number of times each value was observed. The frequency table on the side of the page shows the heights (in inches) of 130 members of a choir, as given in Problem 32 of Chapter 4. Entering 130 numbers could be tiresome, but there is a way to use the counts given.

Height	Count
60	2
61	6
62	9
63	7
64	5
65	20
66	18
67	7
68	12
69	5
70	11
71	8
72	9
73	4
74	2
75	4
76	1

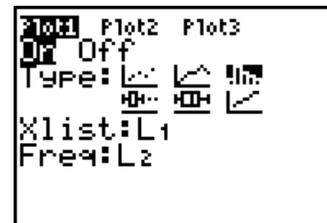
We want to make a histogram to display this distribution. The procedure (allowing for the basic difference in defining the plot) is analogous on the 89 series calculators.) Enter the heights in one list and the observed counts in a second list. (We will put the heights in L1 and the counts in L2.) Make sure the lists are the same length, and that data values match with the given counts.

The first part of the lists looks like this.

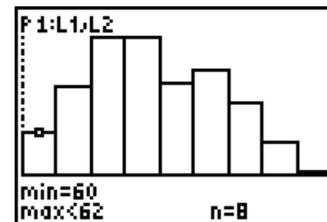
L1	L2	L3	1
60	2		----
61	6		
62	9		
63	7		
64	5		
65	20		
66	18		

L1(1)=60

From $2^{nd} Y=$ (Stat Plot) we will define Plot1 as at right. Notice Xlist is L1 (where the actual values are) and Freq is L2 (where the counts are.)

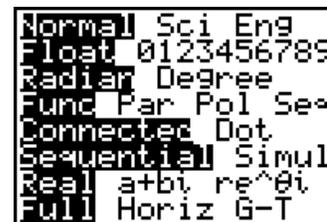


Pressing $ZOOM 9$ gives the following graph. Notice that in this case the intervals and bar width (2) seem reasonable. This is again a right-skewed distribution (from the major peak, the right-hand side is longer than the left). Visually, the center is around 67 or 68 inches; from the data table the data range between 60 and 76 inches.



CALCULATING NUMERICAL SUMMARIES

Before computing any numeric summaries, we want to ensure that the calculator is using all possible decimal places in its intermediate calculations. It is best to use all digits possible, rounding any results only at the end. To do this, press $MODE$. The screen at right will be shown. The calculator should have Float highlighted. If not, move the cursor to highlight Float and press $ENTER$ to move the highlight.

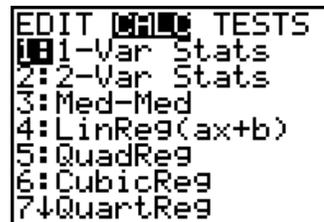


To calculate the numerical summary statistics for a single variable, first enter them into a list. Here the pulse rate data have been entered into list L1. The first few values are shown at right.

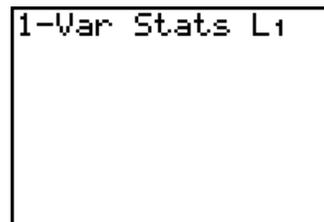
L1	L2	L3	2
88			----
56			
80			
60			
76			
72			
68			

L2(1)=

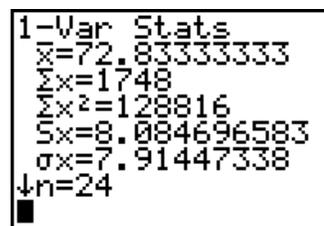
Press **[STAT]** then arrow to **CALC**. The menu at right will be displayed. The menu is organized so that the most frequently used options are at the top. Notice that **1:1-Var Stats** is highlighted. Press **[ENTER]** to select that option (or press **[1]**). The command will be transferred to the home screen.



Now you need to tell the calculator which list to use as input. Press **[2nd][1]** (L1). If no list name is given, the calculator will default to use L1, but it's good practice to specify the list name. Press **[ENTER]** to carry out the command.



The first page of results is displayed at right. The arrow at the bottom left indicates more results are available and can be found by using **[down arrow]**. We first see the mean pulse rate is 72.833333. The calculator does not know if the data you are using represent a sample or a population. It has only one symbol for the mean (\bar{x}). If your data represents a population, you should report the mean using proper notation, and call it μ . The two values displayed next are the sum of the data values and the sum of squared data values. These are intermediate quantities used in computations, and are generally not of interest. Two different standard deviations are also reported as measures of spread.

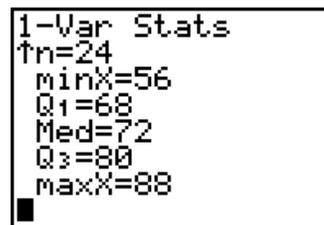


$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}}$$
 is the sample standard deviation and σ_x is the population standard deviation (the formula

is almost the same; the divisor is n). This data does not represent all possible pulse rates even for the women studied, so we will use s_x (8.084696583) as the standard deviation. Which is the correct value to use depends on data you have – is it from a sample or is it for a population? The last value on the screen is the number of items in the data list; in this example, n is 24.

One thing to bear in mind is that calculators (and computers) will use (and report) many more digits than really make sense to use. It comes from division (in which, as we know, things don't always come out evenly) and taking square roots (which also aren't usually whole numbers). How many digits to report should be decided by your instructor, but a good rule of thumb is to report one more place than in the original data. Our data was in beats per minute, so we'll use one decimal place. Also, since we don't have all the possible pulse rates, we will report $\bar{x} = 72.8$ and $s = 8.1$.

Using the down arrow, we find the five-number summary. The median (another measure of center) is 72, which is close to the mean in this data set (as it should be since the data were roughly symmetric – at least in our initial histogram). We can use the other values in this summary to compute two other measures of spread: the Interquartile Range (IQR) which is the spread of the middle half of the data, and the Range. The IQR is $Q_3 - Q_1$, or $80 - 68$ which is 12. This means the central half of the data had a spread of 12 beats per minute. The Range is $\text{max} - \text{min}$, or $88 - 56$ which is 32 beats per minute.



The procedure on a TI-89 is similar. Press **[F4]** (**Calc**). The menu at right will be displayed. The menu is organized so that the most often used options are at the top. Notice that **1:1-Var Stats** is highlighted. Press **[ENTER]** to select that option (or press



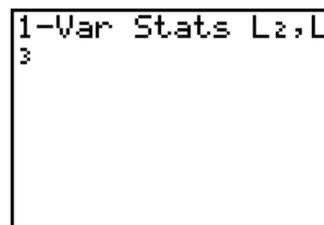
This is the input screen. You need to tell the calculator which list to use as input. Press $\text{2nd} \square$ ([VAR-LINK]). Move the cursor to the correct list and press ENTER to select it. Since each value in our list occurred once, leave Freq at 1. Press ENTER to execute the command.



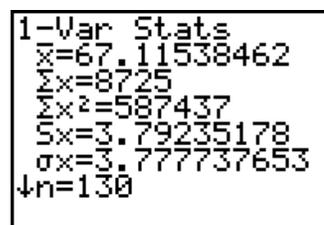
STATISTICS FOR TABULATED DATA

Earlier in this chapter, we looked at the distribution of heights of members in a choir. They were presented in a table of heights along with how many choir members there were of a given height. With these data in lists L2 and L3, we would like to know the average height for the choir.

Just as before, press STAT then arrow to CALC, press ENTER to select choice 1:1-Var Stats. On the home screen, you will specify not just one list, but two. The first list is the list of values (L2) and the second is the list of counts (L3). Your command will look like the screen at right. Don't worry that it didn't all fit on a single line. Press ENTER to execute the command. If you are using a TI-89 series calculator, proceed as in the example above, but set Use Freq and Categories to YES and specify the list of frequencies in the Freq box.



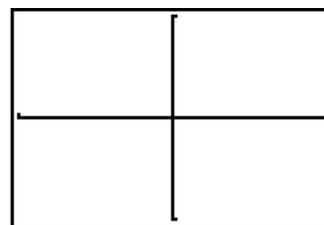
Here are the results. The average (mean) height for the choir members was 67.1 inches. The standard deviation (assuming these are not all the members possible for the choir) is 3.8 inches. If we were to consider this the entire membership of the choir, we would report $\sigma = 3.8$ (there is no practical difference between the two values here since n is large; the impact of subtracting 1 is not big.) Paging down, we find the median height was 66 inches. It is not surprising the median would be somewhat less than the mean for these data since the histogram indicated a right skewed distribution.



WHAT CAN GO WRONG?

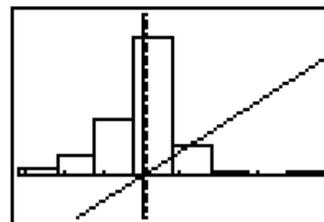
Help! I can't see the picture!

Seeing something like this (or a blank screen) is an indication of a windowing problem. This is usually caused by pressing GRAPH using an old setting. Try pressing $\text{ZOOM} \square$ to display the graph with the current data. This error can also be due to having failed to turn the plot "On."



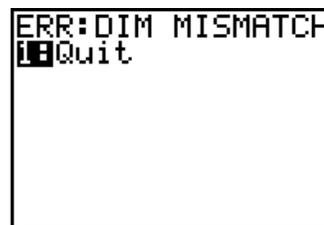
What's that weird line (or curve)?

There was a function entered on the Y= screen. The calculator graphs everything it possibly can at once. To eliminate the line, press Y= . For each function on the screen, move the cursor to the function and press CLEAR to erase it. Then redraw the desired graph by pressing GRAPH .



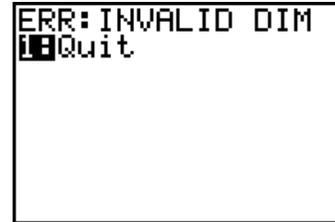
What's a Dim Mismatch?

This common error results from having two lists of unequal length. Here, it pertains either to a histogram with frequencies specified or a time plot. Press ENTER to clear the message, then return to the statistics editor and fix the problem.



What's an Invalid Dim?

This problem is generally caused by reference to an empty list. Check the statistics editor for the lists you intended to use, then go back to the plot definition screen and correct them.



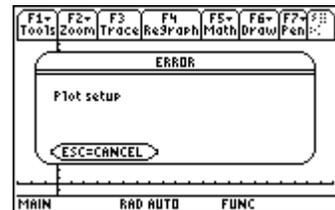
What does Stat mean?

This error is caused by having two stat plots turned on at the same time. What happened is the calculator tried to graph both, but the scalings are incompatible. Go to the STAT PLOT menu and turn off any undesired plot.



Plot setup?

This is the TI-89 equivalent of the STAT error above. It is caused by having two stat plots turned on at the same time. The calculator tried to graph both plots, but the scalings are incompatible. Go to the Stat plots menu and turn off any undesired plot by moving the cursor to that plot, and pressing [F4].

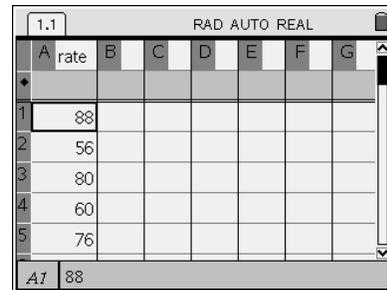


Commands for the TI-Nspire™ Handheld Calculator

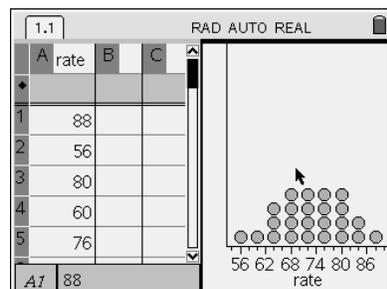
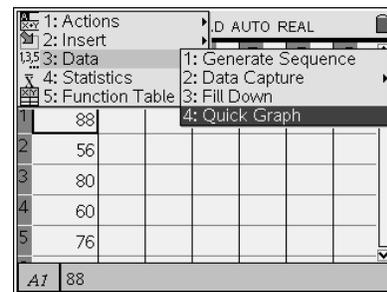
The first step in making a dotplot or histogram is to enter the data. We will use the data on pulse rates of 24 women to create a histogram. Here are the data.

88	56	80	60	76	72	68	80
64	80	84	64	68	72	80	76
72	76	84	76	72	68	68	64

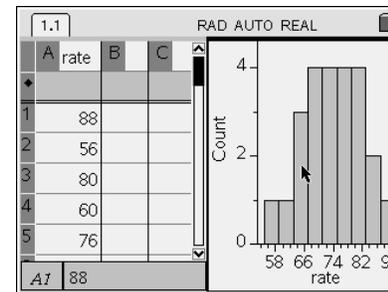
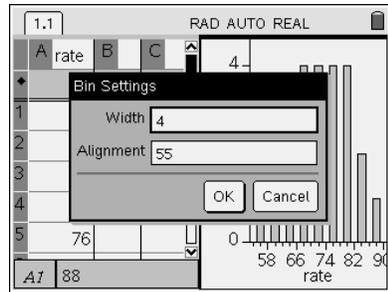
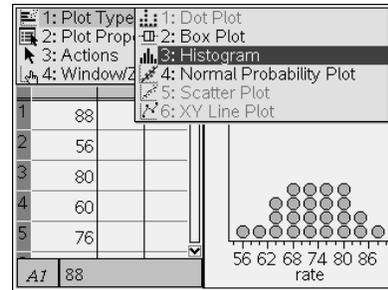
The data have been entered into list named *rate*; the first few values are seen in the accompanying figure.



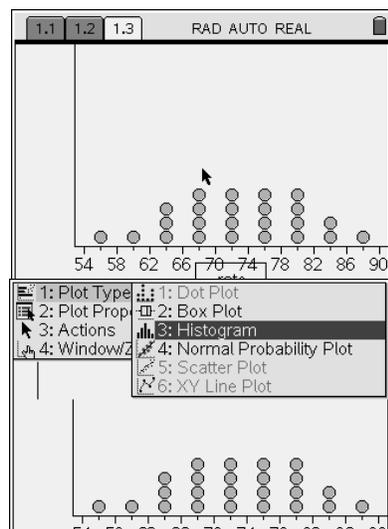
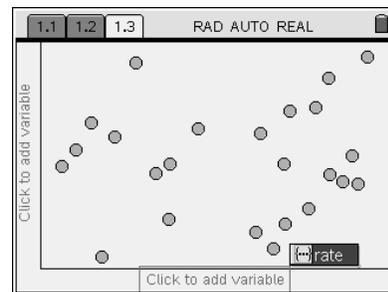
The next step is to create the plot. This is done by pressing $\left(\text{menu}\right)$, and then selecting Data and then Quick Graph. The screen is split and a dotplot is shown.



To change to a histogram, press 2nd , and then select Plot Type and Histogram. We can change the width of the rectangles. Press 2nd , select Plot Properties and then Histogram Properties, and then select Bin Settings. You may then select a new width. Let's try 4. Notice you can also adjust the beginning value for the first interval. It's called Alignment here. We'll leave it as 55. tab to OK and enter .

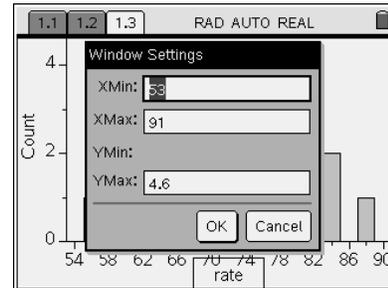


If you prefer to see the plot on an entire screen, rather than a split screen, press 2nd and then select Data & Statistics. At first you will see a plot of dots. Use the arrows to move to the bottom of the display until "Click to add variable" appears. Press enter , highlight the variable, in this case *rate*, and enter again. You will see the dotplot.

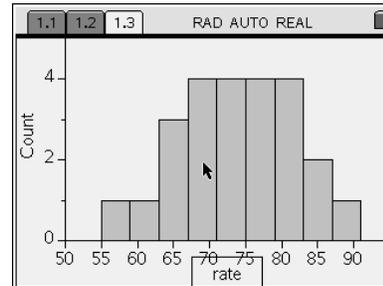
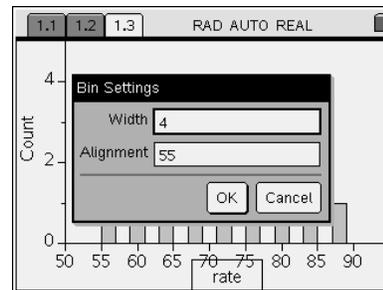


To change to a histogram, press $\text{\textcircled{MENU}}$, and then select Plot Type and Histogram.

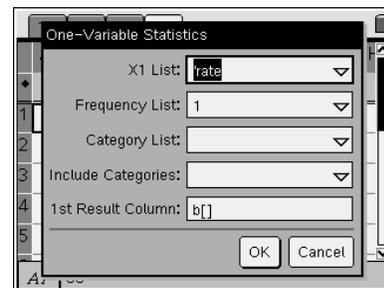
To change the minimum and/or maximum values for the Window, press $\text{\textcircled{MENU}}$, then select Window/Zoom, and then Window Settings. This displays a screen like the one at right. We would not change YMin as it would distort the shape of the distribution.



To alter the width of the rectangles, press $\text{\textcircled{MENU}}$ and then select Plot Properties, Histogram Properties, and Bin Settings. Change Width to 4, $\text{\textcircled{TAB}}$ to OK and $\text{\textcircled{ENTER}}$.



To calculate the numerical summary statistics for a single variable, first enter them into a list. We have entered the pulse rate data into the list *rate*. Press $\text{\textcircled{2ND}}$ and insert a Lists and Spreadsheet page. Move the cursor to the top of the column A and type the name “rate”. You should see the list of pulse rates appear on column A. Move the cursor to cell A1.



Press $\text{\textcircled{m}}$, then select Statistics, followed by Stat Calculations, and then select One-Variable Statistic. Type 1 in the dialogue box, $\text{\textcircled{t}}$ to OK, and $\text{\textcircled{e}}$. The screen labeled "One-Variable Statistics" at the right will be displayed. Choose the list, either by name or column name, such as a[]. Using this method the results will be displayed in the spreadsheet. You may select the column, such as b[], $\text{\textcircled{t}}$ to OK, and $\text{\textcircled{e}}$. The results will be displayed using two columns of the spreadsheet. Remember you may use $\text{\textcircled{m}}$ and (Actions) to resize the columns.

	A	B	C
1	88	Title	One-Variable Statisti...
2	56	\bar{x}	72.8333
3	80	Σx	1748.
4	60	Σx^2	128816.
5	76	$s_x := s_{n-1}x$	8.0847

You may prefer to see the summary statistics on a full screen rather than within the spreadsheet. If so, press $\text{\textcircled{2nd}}$, and then $\text{\textcircled{1}}$ for calculator. Then, similar to the method just described press $\text{\textcircled{m}}$, select Statistics, Stat Calculations, and then One-Variable Statistic. Type 1 in the dialogue box, $\text{\textcircled{t}}$ to OK, and $\text{\textcircled{e}}$. Choose the list by name, $\text{\textcircled{t}}$ to OK, and $\text{\textcircled{e}}$. Press \blacktriangle to scroll to the top of the screen.

Label	Value
"Title"	"One-Variable Statistics"
" \bar{x} "	72.8333
" Σx "	1748.
" Σx^2 "	128816.
" $s_x := s_{n-1}x$ "	8.0847
" $o_x := o_{n}x$ "	7.91447
"n"	24.