EXPLORATIONS

Activity 1

Describing One-Variable Data

The primary data set you will be working with in this activity is the heights (in feet) of tall buildings in Philadelphia, Pennsylvania. This was stored in list **PHILY** under “Creating a New List in the Editor” in Do This First and was entered in the order of the year completed. The data is duplicated in the table below.

Heights (ft.) of Tall Buildings in Philadelphia, PA

548	405	375	400	475	450	412
375	364	492	482	384	490	492
490	435	390	500	400	491	945
435	848	792	700	572	739	572

(Source: Reprinted with permission from the World Almanac and Book of Facts 1996. Copyright © 1995 K-III Reference Corporation. All rights reserved.)

Topic 1—Histograms and Frequency Tables from Raw Data

Setting up the Plot

1. Press [2nd] [STAT PLOT] **1:Plot1** [ENTER] to display screen 1. (Your screen may have **Off** highlighted.)

(1)

2. Turn on the plot by pressing [ENTER] with the flashing cursor on **On**. **On** will now be highlighted instead of **Off**.

3. Use [▼] to move down to **Type**.

   The cursor flashes on the first type, **Scatter** plot. Starting from the first type, pressing [▶] moves to the other five types: **xyLine**, **Histogram**, **Modified Boxplot**, **Boxplot** (regular box plot), and **Normal Probability** plot.

4. Press [▶] [▶] to move to **Histogram**, the third type, and then press [ENTER] to select it.

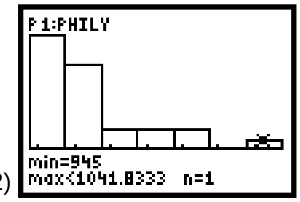
5. Use [▼] to move down to the input request **Xlist**. (Note that the input request depends on the plot type selected.)

   Paste the list **PHILY**, and set **Freq** to **1** (because this is raw data, and you will count each value once).

Activity 1, Describing One-Variable Data (cont.)

Using the Automatic Window

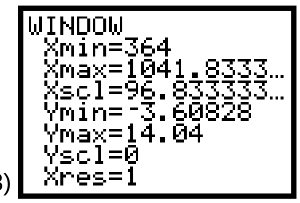
Press [ZOOM] **9:ZoomStat** [TRACE] followed by several [▶] for the display of the **Histogram** in screen 2. Notice that the lower class limit for the rightmost class is **min = 945**. There is only **n = 1** value in this last cell indicating the tallest building is 945 feet. The class width is a bit strange, but it is nice to have this automatic plot capability.



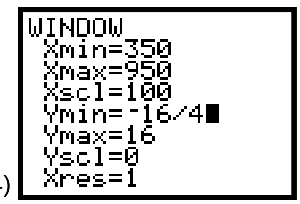
(2)

Setting Up the Window

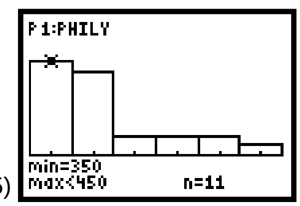
1. Press [WINDOW] for screen 3, which reveals that the shortest building (**Xmin**) is **364** feet.
2. Change the values, as shown in screen 4, to include all values with reasonable class limits and class width (**Xscl =100**). Making **Ymin** the negative of **Ymax/4** leaves enough room at the bottom of the plot screen for the class information.
3. Press [TRACE] for screen 5 with 11 buildings in the first height class. Using [▶] reveals frequencies of **10**, **2**, **2**, **2**, and **1** for the other classes.



(3)



(4)



(5)

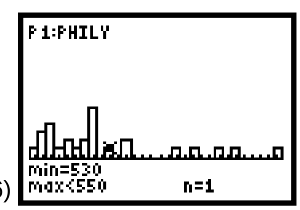
The Frequency Table

The values shown in step 3 (above) make up the *frequency table* given below.

Class Limits	Frequency
350 to < 450	11
450 to < 550	10
550 to < 650	2
650 to < 750	2
750 to < 850	2
850 to < 950	1

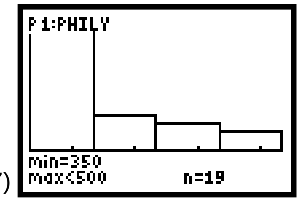
The Effect of Changing Class Width or Xscl

If you change the **WINDOW** in screen 4 so that the class width is **Xscl = 20**, press [TRACE], and then [▶], you obtain the cell that identifies the Philadelphia City Hall at 548 feet (see screen 6). (The City Hall height includes the 37-foot statue of William Penn. Until 1987 (by plan), no buildings were higher than the City Hall.)



(6)

If you change the window in screen 4 so that **Xscl = 150**, and then press [TRACE], you find that with 19 values in the first class, its rectangle does not fit the display (see screen 7). The values of **Ymin** and **Ymax** must be adjusted. (One way that works is **Ymin = -6** and **Ymax = 24**.)



(7)

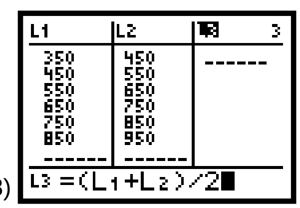
Topic 2—Histograms from Frequency and Relative Frequency Tables

A frequency table for the heights (in feet) of tall buildings in Philadelphia, PA is given below with 11 buildings from 350 up to 450 feet and only one building 850 feet or taller.

Class Limits	Frequency
350 to < 450	11
450 to < 550	10
550 to < 650	2
650 to < 750	2
750 to < 850	2
850 to < 950	1

Calculating Class Marks

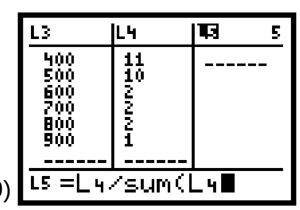
1. Store the lower class limits in **L1** and the upper class limits in **L2** in the stat editor.
2. Highlight **L3** at the top of the spreadsheet, and type [(] **L1** [+] **L2** [)] [÷] **2**, as shown in the bottom line in screen 8.
3. Press [ENTER] to reveal the class marks in **L3** (see screen 9 where the frequencies also have been added to **L4**).



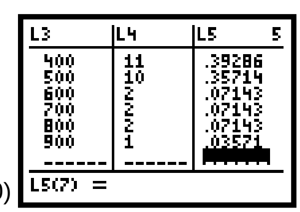
(8)

Calculating Relative Frequencies

1. Highlight **L5** at the top of the spreadsheet , as shown in screen 9.
2. Enter **L4** [÷] [2nd] [LIST] **<MATH> 5:sum(L4**, as shown in the bottom line in screen 9.
3. Press [ENTER] and the relative frequencies are placed in **L5**, as shown in screen 10, with **39.29** percent of the buildings in the first class.



(9)

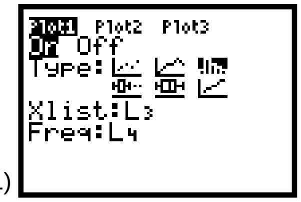


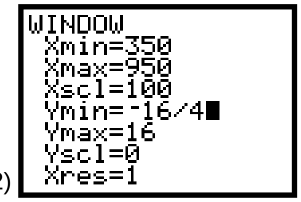
(10)

Activity 1, Describing One-Variable Data (cont.)

Setting Up Plot and Window

Set up **Plot1** for a **Histogram** as shown in screen 11 with the **WINDOW** as shown in screen 12, where
Xmin = 350, the lower limit of the first class
Xmax = 950, the upper limit of the last class
Xscl = 100, the class width or the distance between the class marks
Ymax = 16 gives room for the maximum frequency of 11, plus extra room for labels at the top of the **Histogram**.

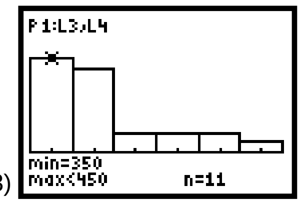
(11)


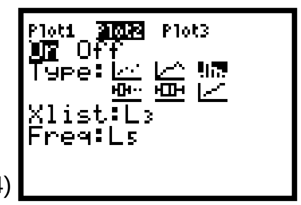
(12)


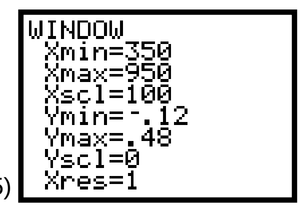
The Frequency Histogram

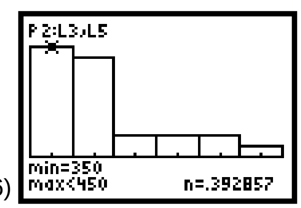
Pressing [TRACE] reveals the frequency **Histogram** displayed in screen 13.

The *relative frequency Histogram* is given in screen 16 (with the setup in screens 14 and 15) with **n = .3929** percent of the buildings in the first class. (Note that the **Freq** input is actually the relative frequencies in **L5**.)

(13)


(14)


(15)


(16)


Topic 3—Stem-and-Leaf Plots and Dot Plots

The stem-and-leaf plot and dot plot are easy to make by hand if the data is in order.

1. Make a copy of the list **PHILY** (the Philadelphia, PA building heights that you saved. See the first page of Activity 1) by storing it in **L4**.
2. From the home screen, paste **ʟPHILY** and press [STO▶] **L4**; or, from the spreadsheet, highlight **L4** in the top line, paste **ʟPHILY** in the bottom line, and press [ENTER].

Putting Data In Order

Sort the values in ascending order from low to high value by pressing STAT **2:SortA(L4** ENTER. **Done** is displayed, as shown in screen 17.

(17)

Stem-and Leaf Plot

Entering the stat editor reveals the data for easy reading. The first seven values shown in screen 18, when rounded to the nearest ten and given in tens, are **36, 38, 38, 38, 39, 40, 40**. These values start the stem-and-leaf plot shown below. The complete plot is given below the start-up plot with data from St.Louis.

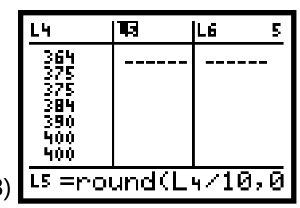
3|68889
4|00

St. Louis   Philadelphia
98| 3 |68889
320| 4 |00114458899999
994| 5 |0577
3| 6 |
| 7 |049
| 8 |5
| 9 |5

St. Louis, Missouri has nine buildings over 350 feet tall as follows.

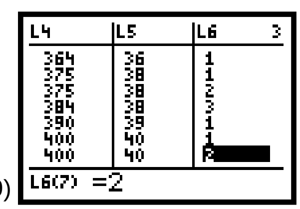
375 390 398 420 434 540 588 593 630

(Source: Reprinted with permission from the World Almanac and Book of Facts 1996. Copyright © 1995 K-III Reference Corporation. All rights reserved.)



(18)

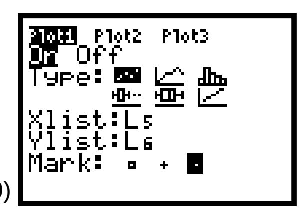
Above is the complete back-to-back stem-and-leaf plots for the two data sets. St. Louis has a fairly symmetric distribution of tall building heights with no buildings taller than 630 feet (the arch). Philadelphia has many more tall buildings with a distribution skewed to taller values. Until 1986, however,when the seven tallest buildings were built in Philadelphia, St. Louis had the taller buildings.



(19)

Dot Plot

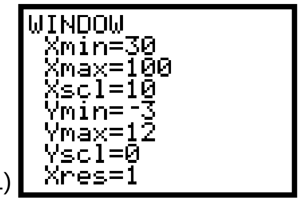
With the data in order, a dot plot is easy to construct by hand with one dot for each building above its height scale value, but it can also be plotted easily on the TI-83.

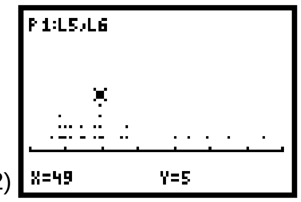


(20)

1. Highlight **L5**, as shown in the top line of screen 18.
2. Press MATH **<NUM> 2:round(L4** ÷ **10** , **0**), as shown in the bottom line of screen 18.

3. Press [ENTER] for the heights of the buildings to the nearest 10 feet (given in tens), as shown in screen 19.
4. In **L6**, type the counting numbers next to each height in **L5**. There is **1** value of **36**, but **1**, **2**, **3** values of **38**, and so on, as shown in screen 19.
5. Set up **Plot1** for a **Scatter** plot (the first type), as shown in screen 20.
6. Set up the **WINDOW** as shown in screen 21.
7. Press [TRACE] and then [▸] a few times for screen 22, which shows there are five buildings approximately 490 feet tall in Philadelphia.

(21)


(22)


Topic 4—Measures of Central Tendency and Variability

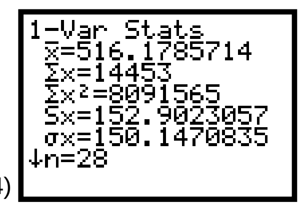
The most common measures of central tendency (the *mean* and the *median*) and measures of spread or variability of a distribution (the *standard deviation*, *variance*, *interquartile range*, and *five-number summary*) are given for both a list of raw data and for data grouped in a frequency table.

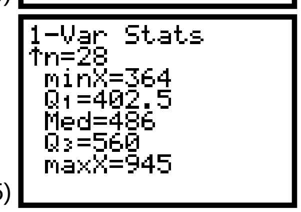
Using Raw Data

You will use the heights of tall Philadelphia buildings as stored in list **ʟPHILY** in Do This First.

1. Press [STAT] **<CALC> 1:1-Var Stats ʟPHILY** to display screen 23.
2. Press [ENTER] for screen 24, and then use [▾] to see screen 25.

(23)


(24)


(25)


mean - The first value in screen 24 is **$\bar{x}$ = 516.1785714**, which is equal to the sum of the data (the second value of the output), or **Σx = 14453** divided by the number of data values, or **n = 28**.

median - **Med = 486** is found in the second screen of the output (screen 25), the middle value of the five-number summary.

five-number summary - From the second screen of the output, **minX = 364**, **Q1 = 402.5**, **Med = 486**, **Q3 = 560**, **maxX = 945**.

standard deviation - **Sx = 152.902** and **σx = 150.147** are the last values of the first output screen above **n**. (See screen 24.)

Sx = $\sqrt{((\Sigma(x - \bar{x})^2)/(n - 1))} = \sqrt{((\Sigma x^2 - (\Sigma x)^2/n)/(n - 1))}$

σx = $\sqrt{((\Sigma (x - \mu)^2)/n)} = \sqrt{((\Sigma x^2 - (\Sigma x)^2/n)/n)}$

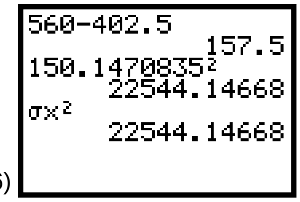
***Note**: Because the set of interest is a population of the heights of all the tall buildings in Philadelphia (not a sample of heights), the mean is usually signified as μ but its calculation and value are the same as $\bar{x}$.*

***Note:** Because the set of interest is a population of the heights of all the tall buildings in Philadelphia (not a sample of heights), σx is appropriate (Sx is used to estimate σx from a sample).*

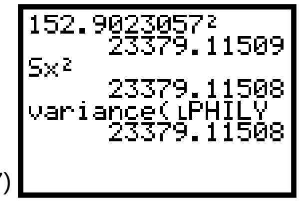
interquartile range - It is easy enough to calculate the interquartile range from **Q3** - **Q1** = **157.5**, as shown in the first two lines of screen 26.

variance - It is a bit tedious to calculate the variance by typing in the digits of the standard deviation and then squaring it using $\boxed{x^2}$. You may want to paste **σx** by pressing $\boxed{\text{VARS}}$ **5:Statistics <XY> 4: σx** and then squaring this, as shown in the last two lines of screen 26.

(26)



(27)

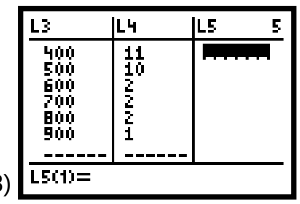


If your data was a sample from a larger population, you could calculate the sample variance (division by **n-1**) in a similar way with $\boxed{\text{VARS}}$ **5:Statistics <XY> 3:Sx**. Another possibility is to press $\boxed{\text{2nd}}$ [LIST] **<MATH> 8:variance(ʟPHILY**, as shown in the last two lines of screen 27. This gives the same answer, **23379.11508**.

Using Grouped Data

You will use the frequency table of the Philadelphia data as grouped in Topic 2, including the class marks below. The class marks are stored in **L3** and the frequencies in **L4**. (See screen 28.)

(28)



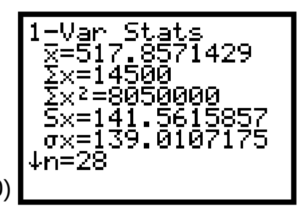
Class Limits	Class Mark	Frequency
350 to < 450	400	11
450 to < 550	500	10
550 to < 650	600	2
650 to < 750	700	2
750 to < 850	800	2
850 to < 950	900	1

1. Press $\boxed{\text{STAT}}$ **<CALC> 1:1-Var Stats L3** $\boxed{,}$ **L4** for screen 29.
2. Press $\boxed{\text{ENTER}}$ and then use $\boxed{\downarrow}$ to see screens 30 and 31.

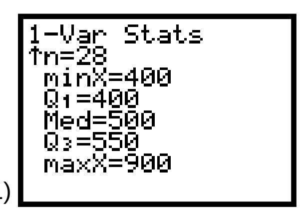
   The output is similar to the raw data output because basically, it comes from the same data. However, instead of 11 different values from 350 up to 449, you treat all 11 values as 400. Likewise, the largest value is treated as 900 instead of the actual 945. A mean of **517.9** feet for the grouped data compared to the actual **516.2** (in screen 24) feet is not bad.

(29)

(30)



(31)



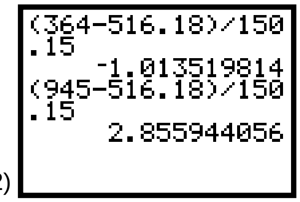
If the actual data is such that there are in fact eleven 400s, ten 500s, two 600s, two 700s, two 800s, and one 900, then the output using the grouped procedure above is correct and not an approximation.

If you have the raw data, it makes sense to use it. But if you only have a table of grouped data to summarize raw data, this is the option to use.

Topic 5—Measures of Position and Cumulative Relative Frequency Distributions (Ogives): Percentiles and Quartiles, Standard Scores (z-scores)

You will work with the heights (in feet) of tall Philadelphia buildings that you stored in list **ʟPHILY** in Do This First. This data is duplicated at the beginning of Activity 1.

(32)



Finding Standard scores (z-scores)

Standard scores (z scores) are numbers that represent how many standard deviations X is above or below the mean, or $z = (X - \mu)/\sigma x$.

From Topic 4 you found the mean = **μ = 516.18**, and the standard deviation = **σx = 150.15**. Use these values to find the z score of the smallest (364 feet) and the tallest (945 feet) buildings in your list. (See screen 32.)

Smallest Building: z = **(364 - 516.18)/150.15 = ⁻1.01**
Largest Building: z = **(945 - 516.18)/150.15 = +2.86**

The [2nd] [ENTRY] (last entry) feature works nicely here. Pasting μ and **σx** (as shown in Topic 4) is also a possibility.

The smallest value is about one standard deviation below the mean, and the largest value is almost three standard deviations above the mean. This indicates the non-symmetric nature of the distribution (it is skewed to the right).

Finding Percentiles and Quartiles

The *percentile* of a score indicates what percent of the data values are less in magnitude. Not all textbooks use the same procedure for finding percentiles, but all methods give similar answers as the list size increases.

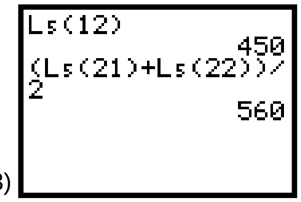
The first quartile is the same as the twenty-first percentile (**Q1= P25**); also **Q3 = P75**, and the median is the same as **P50**. These three values are part of the five-number summary obtained with [STAT] **<CALC> 1:1-Var Stats ʟPHILY**, as shown in Topic 4, with **Q1 = 402.5**, **Med = 486**, **Q3 = 560** and **n = 28**.

1. Put the **PHILY** data in order in **L5** (from smallest to largest value) as was done in Topic 3.

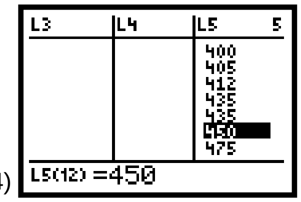
2. Find **P40** and **P75** (remember **n = 28**).

   P40: 0.40 ∗ 28 = 11.2. Round up to 12. The twelfth value of **L5** or **L5(12) = 450**, as shown in screen 33.

   P75: 0.75 ∗ 28 = 21. Because this is an integer, take the mean of the twenty-first and twenty-second values, as shown in the display (**Q3 = 560**).



(33)



(34)

Given a value of 450 feet for a building's height, what percentile value is this in your list?

Going down the list in the stat list editor, you find this is the twelfth value, as shown in screen 34, so there are 11 tall buildings less than 450 feet. Because 11/28 = 0.393, this is approximately **P39**. Because you found **P40 = 450** (above) and now have **P39 = 450**, at least one percentile value is indeed approximate, but these two values would be closer as the size of the data list increased.

Finding Cumulative Relative Frequency Distributions (Ogives) and Percentiles

Using the frequency tables of the building heights in Topic 2 (at right), do the following.

Class Limits	Frequency
350 to < 450	11
450 to < 550	10
550 to < 650	2
650 to < 750	2
750 to < 850	2
850 to < 950	1

1. Put the lower class limits and the last upper class limit in **L1**.

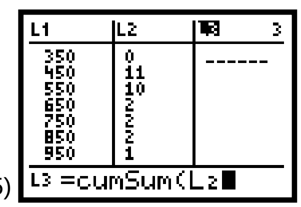
2. Put the frequencies in **L2**, but with the first value being an extra value zero. (See screen 35.)

3. Highlight **L3** and paste **cumSum(L2**, as shown in the last line of screen 35. (**cumSum** from [2nd] [LIST] **<OPS> 6**).

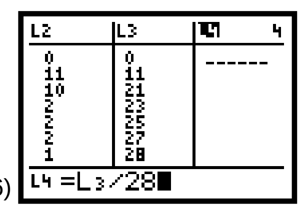
4. Press [ENTER] for screen 36 with the cumulative frequencies in **L3**.

5. Highlight **L4** and put **L3** [÷] **28**, as shown in the last line of screen 36.

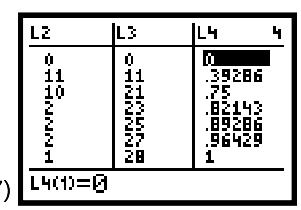
6. Press [ENTER] for screen 37 with the cumulative relative frequencies in **L4**.



(35)



(36)



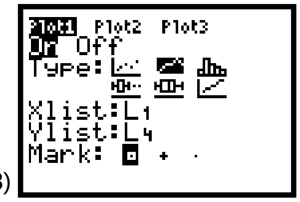
(37)

7. Set up **Plot1** for an **xyLine** plot (the second type), as shown in screen 38.

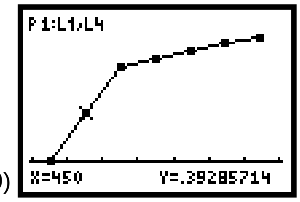
8. Plot with [ZOOM] **9:ZoomStat**, and then press [TRACE] and [▶] for the plot in screen 39.

   Notice that **450 = P39** as before and in the second row of the spreadsheet in screen 37. About 39 percent of the tall buildings (11 buildings) are less than 450 feet.

(38)



(39)



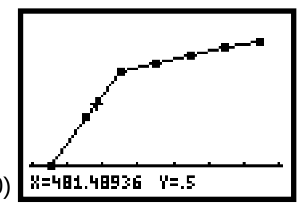
9. Press [▶] again to move up to the next point and the third row of the spreadsheet with 21 or 75 percent of the values below 550 feet (the third quartile).

   Earlier, you calculated **Q3 = 560** feet (the mean of the twenty-first and twenty-second value). Again, as the sample size increases (such as the results of the Scholastic Aptitude Test (SAT) where all values are represented), these differences will diminish.

10. To approximate other values that are not points in the spreadsheet, press [GRAPH]**,** and then the cursor control keys to move the cross hairs as close to the desired point as possible.

The example in screen 40 approximates **P50** as 481 compared to the previous value of 486. You are limited by using straight lines to connect points, but this limitation is reduced as the data set and number of points increase. You are also limited by technology and the width of a pixel.

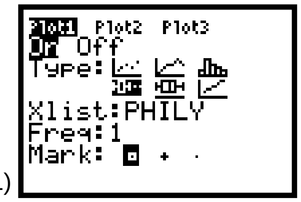
(40)



The ogive in screen 40 indicates the skewed nature of the distribution of the building heights. The slope is very steep up to **P75** (between 350 and 550 feet), but the top 25 percent of the data is more spread out (from 550 to 950 feet).

Topic 6—Box Plots and Five-Number Summary

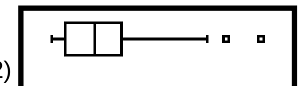
You will start with the heights (in feet) of tall Philadelphia buildings stored in list **ʟPHILY** in Do This First. This data is duplicated at the beginning of Activity 1.



(41)

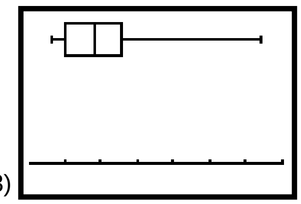
Setting up Box plots

1. Set up **Plot1** for a **Modified Boxplot** (the fourth type), as shown in screen 41. It is called modified to distinguish it from the next choice, a regular box plot , called just **Boxplot**.



(42)

2. Press ZOOM **9:ZoomStat** for screen 42. If you were set up for a regular **Boxplot**, you would get the next display in screen 43. The only difference is that the right "whisker" (these are also sometimes called "box and whisker" plots) is extended out to the maximum value.

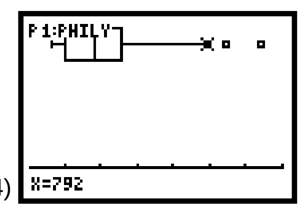


(43)

Calculating a Five-Number Summary

1. Press TRACE and the median shows a flashing box and value of **486** at the bottom of the screen.
2. Press ◄ and you have **Q1 = 402.5**.

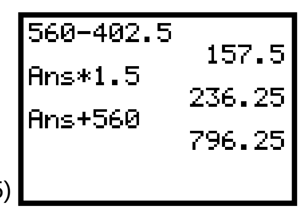
   Press ◄ again and you have **minX = 364**.
3. Go in the other direction and you find **Q3 = 560**, and then skip all the way to the right box for **maxX = 945**.



(44)

These make up the five-number summary of data, which was also covered in Topic 4 and available from the second screen of STAT **<CALC> 1:1-Var Stats ʟPHILY**.

The two points skipped were **X = 792** (shown in screen 44) and **X = 848**, the third and second largest values. You separated the top two values from the whiskers to identify them as possible "outliers," values that are far away from the rest of the data and that may have a special story.



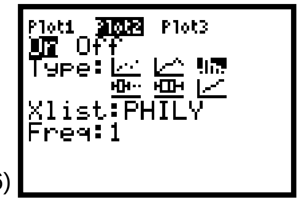
(45)

(To separate these points, take **Q1 - 1.5 * IQR** and **Q3 + 1.5 * IQR**, where **IQR** stands for the interquartile range or **IQR = Q3 - Q1 = 560 - 402.5 = 157.5** (as shown in Topic 4). Thus, **1.5 * IQR = 1.5 * 157.5 = 236.25** and **Q3 + 236.25 = 796.25**. (See screen 45.) The first value below 796.25 is 792, but 848 and 945 are above it. **Q1 - 236.25 = 166.25,** and none of the buildings in our list were less than that; therefore, no possible outliers are identified off to the left.)

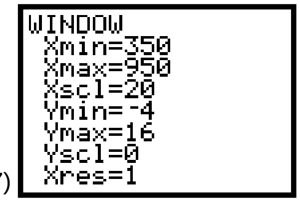
Activity 1, Describing One-Variable Data (cont.)

Box Plot with Histogram

After setting up **Plot1** as above in screen 40, set up **Plot2** for a **Histogram** as shown in screen 46 and **WINDOW** as shown in screen 47.

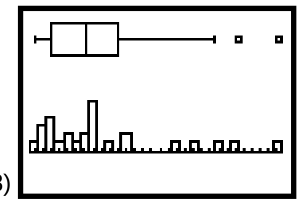


(46)



(47)

Pressing GRAPH gives screen 48. The added **Histogram** (as shown in Topic 1) indicates you might want to treat the five upper values as special.


(48)

Side-By-Side Box Plots

Up to three box plots can be plotted on the same screen.

The 28 tallest buildings in New York, New York are stored in list **NYNY** and listed below. (Twenty-eight buildings were identified as tall in Philadelphia. In New York City, 131 buildings were so identified.)

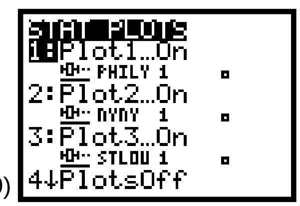
ʟNYNY: 700 705 707 716 724 725 730 739 741 743 745 750 750 752 757 778 792 808 813 814 850 915 927 950 1046 1414 1362 1368

(Source: Reprinted with permission from the World Almanac and Book of Facts 1996. Copyright © 1995 K-III Reference Corporation. All rights reserved.)

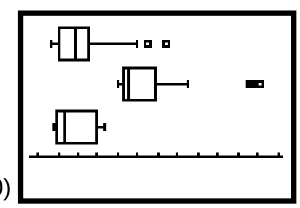
Nine buildings are identified as tall in St. Louis, Missouri. The heights are stored in **ʟSTLOU**.

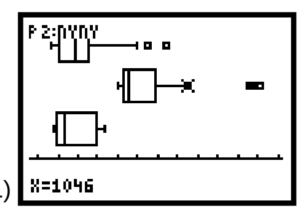
ʟSTLOU: 375 390 398 420 434 540 588 593 630

1. Set up all three plots as **Modified Boxplots**, as shown in screen 49.
2. Press ZOOM **9:ZoomStat** for the next display (screen 50).
3. Press TRACE and ▼ to investigate the different plots.



(49)


(50)



(51)

New York (in the middle) is the city of skyscrapers, and Philadelphia seems much more like St. Louis. Because of the difference in sample size, it is easer to make this later comparison with back-to-back stem-and-leaf plots, as shown in Topic 3.

