EXPLORATIONS

Chapter 3

# Comparing Distributions of Univariate Data

Topic 9 covers comparing data and constructing multiple univariate plots.

## Topic 9—Multiple Univariate Plots

*Example*: Building heights in Philadelphia, PA were stored in list **phily** and folder **BLDTALL** in Topic 1. Store Seattle building heights (buildings 400 or more feet tall) in list **seattle**, and New York City building heights (the 24 tallest buildings) in list **nyc**.

Store the following data, in the order listed, in lists **seattle** and **nyc** in folder **BLDTALL**.

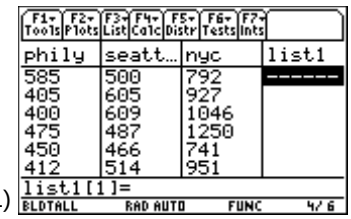
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **seattle** | 500 | 605 | 609 | 487 | 466 | 514 | 454 | 456 | 543 | 409 | 574 | 943 |
| | 493 | 730 | 580 | 743 | 722 | 448 | | | | | | |
| **nyc** | 792 | 927 | 1046 | 1250 | 741 | 951 | 850 | 813 | 808 | 730 | 750 | 750 |
| | 1368 | 1362 | 915 | 716 | 752 | 739 | 778 | 814 | 745 | 757 | 866 | 861 |

(Source: Reprinted with permission from the World Almanac and Book of Facts 2000. © 2000 World Almanac Education Group, Inc. All rights reserved.)

© 2001 TEXAS INSTRUMENTS INCORPORATED

1. Press ◆ APPS, **1:Flash Apps**, and then select the Stats/List Editor.
2. Create the list **seattle** by highlighting the **list1** heading. Press 2nd [INS] and type the name **seattle**.
3. Repeat step 2 to insert the name **nyc** in place of **list2**.
4. Enter the **seattle** and **nyc** data values from the table on page 49 under the appropriate headings (screen 1).

(1)

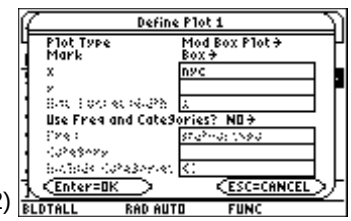



## Parallel Boxplots

Parallel boxplots are the quickest way to get a pictorial overview of the comparison between data lists on the TI-89.

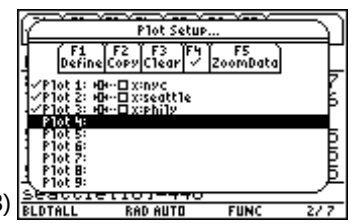
1. From the Stats/List Editor and folder **BLDTALL**, press F2 **Plots**, and select **1:Plot Setup**.
2. Highlight **Plot 1**, and press F1 **Define** to define **Plot 1** as a modified boxplot with X List: **nyc** (screen 2).
3. Press ENTER twice to return to the Plot Setup screen.

(2)



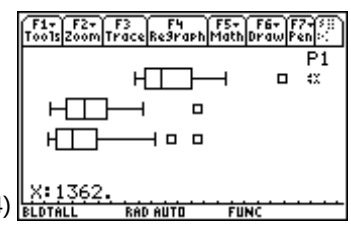

4. Repeat steps 2 and 3 for **Plot 2** defined for list **seattle** and **Plot 3** defined for list **phily** (screen 3).

(3)




5. From the Plot Setup screen, press F5 **ZoomData**. After the plots are displayed, press F3 **Trace** and ▶ four times (screen 4).

(4)




© 2001 TEXAS INSTRUMENTS INCORPORATED

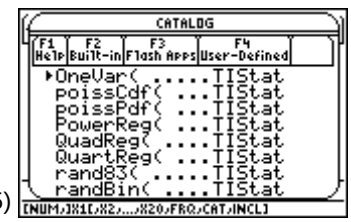
All the distributions are skewed to the right with at least one outlier. New York City (P1) has three outliers of **1250**, **1362**, and **maxX = 1368** feet (the Empire State Building, One World Trade Center, and Two World Trade Center, respectively). The most obvious difference is with New York City having taller buildings (center shifted to the right). Seventy-five percent of NYC's 24 tallest buildings are over 750 feet = $Q_1$, while Seattle has only one building that tall (the outlier), and Philadelphia has three buildings that tall (including the two outliers). Philadelphia buildings (minus the outliers) have the greatest overall spread, but NYC's interquartile range (spread of center 50% of the box) is the largest and its center box also has the most skewness. Seattle's middle 50% is almost symmetric (median line almost in the center of the box).

### 1-VarStats for Multiple Lists

1. From the Home screen, press [CATALOG], and then press [F3] **Flash Apps**.
2. You are in alpha mode so you do not press the [alpha] key. Press the letter **O** (screen 5). Note the syntax at the bottom of the screen when ▶ is next to **OneVar(**. NUM is the number of lists designated as **x1**, **x2**, …, **x20**.
3. Press [ENTER] and **tistat.onevar(** is pasted in the input line of the Home screen.

***Note:*** *Lists do not need to be of equal length.*

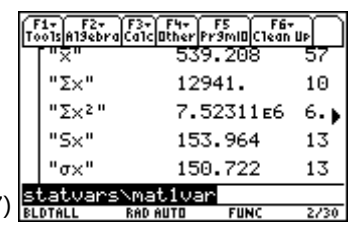
(5)




4. Type and/or paste **3, phily, seattle, nyc)** and then press [ENTER] to complete the operation (screen 6). (**Done** is displayed.)
5. Press [2nd] [VAR-LINK], scroll down to highlight the **STATVARS** folder, and press ▶ to expand the folder and highlight **mat1var**.
6. Press [ENTER] to paste **mat1var** to the Home screen input line.
7. Press [ENTER] (screen 7).
8. To view the entire matrix of values, press ▲ once to highlight the matrix. Press ▶ or ◀ to go right or left, and [↑] ▼ or [↑] ▲ to go up or down. (The [↑] key is to the right of [2nd].)

(6)


(7)




© 2001 TEXAS INSTRUMENTS INCORPORATED

Below is a table summary of seven key variables for each of the three cities. As a reminder:

$\bar{x}$ = mean

$\sigma_x$ = standard deviation

$n$ = sample size

Med = median

$Q_3$ = third quartile (75% value)

$Q_1$ = first quartile (25% value)

IQR = interquartile range

| | phily | seattle | nyc |
|---|---|---|---|
| $\bar{x}$ | 539 | 571 | 878 |
| $\sigma_x$ | 151 | 133 | 188 |
| $n$ | 24 | 18 | 24 |
| Med | 489 | 529 | 811 |
| $Q_3$ | 579 | 609 | 921 |
| $Q_1$ | 426 | 466 | 750 |
| IQR | 153 | 143 | 171 |

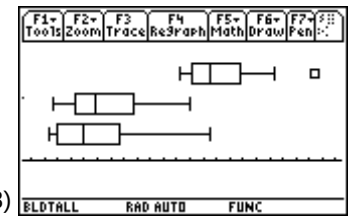
Summary measures without outliers:

| | phily | seattle | nyc |
|---|---|---|---|
| $\bar{x}_0$ | 507 | 549 | 814 |
| $\sigma_0$ | 109 | 101 | 85 |
| $n_0$ | 22 | 17 | 21 |
| $Med_0$ | 485 | 514 | 792 |
| $IQR_0$ | 155 | 146 | 116 |

© 2001 TEXAS INSTRUMENTS INCORPORATED

The summary measures in the first table confirm what you observed from the modified boxplots, but the values calculated without the outliers emphasize the extreme nature of the New York outliers to the extent that the measure of variability for New York has changed from the most variable to the least (compare $\sigma_x$ and $IQR_x$ with $\sigma_0$ and $IQR_0$). Screen 8 shows what the boxplot looks like if you delete the outlier values from the data set and regraph. Compare screen 8 with screen 4. With the reduced data set, the Chrysler Building in New York City (1046 feet) becomes a possible outlier.

(8)

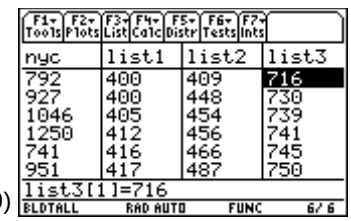



### Multiple Dotplots

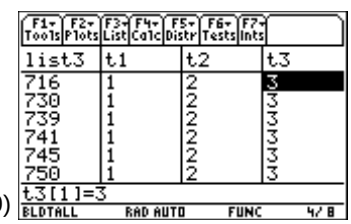
The TI-89 has no built-in dotplot function. In Topic 2 you did the plot by hand because dotplots and stemplots are most effective for small to moderate size data lists (histograms work best for longer lists). It will be helpful, however, to build multiple dotplots on the TI-89 using the following method to aid in making comparisons.

1. Copy lists **phily**, **seattle**, and **nyc** to lists **list1**, **list2**, and **list3** respectively, and sort them in ascending order (screen 9). (See Chapter 1, Topic 2, *Putting Data in Order* section.) The Stats/List Editor should resemble screen 9.
2. Replace **list4**, **list5**, and **list6** with new names **t1**, **t2**, and **t3** respectively. (See the *Do This First* chapter, *Inserting a New List Name* section.)
3. Fill list **t1**, **t2**, and **t3** with 1's, 2's, and 3's respectively, using commands **seq(1,x,1,24)**, **seq(2,x,1,18)**, and **seq(3,x,1,24)**. (See the *Do This First* chapter, *Using seq( to Generate a List* section.)
4. The screen should resemble screen 10.
5. Change the second **1** in list **t1** to **1.1**. (This corresponds to the repeated value of **400** in list **x1**.)
6. Press 2nd ⊙ to continue down list **t2** to make the 8th and 18th **t1** values have values of **1.1**.
7. List **seattle** has no repeats, but in **list3** (**nyc**) there are two **750**'s in positions 6 and 7, so make the 7th value in **t3** equal **3.1**.

(9)



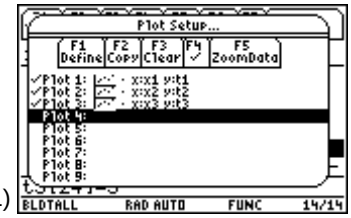

(10)




© 2001 TEXAS INSTRUMENTS INCORPORATED

8. Using F2 **Plot**, select **1:Plot Setup** and F1 **Define** to create three plots with the specifications shown in the table and in screen 11.

(11)



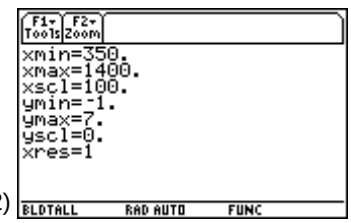

| | | | | |
|---|---|---|---|---|
| **Plot 1** | Type: Scatter | Mark: Dot | X List: list1 | Y List: t1 |
| **Plot 2** | Type: Scatter | Mark: Dot | X List: list2 | Y List: t2 |
| **Plot 3** | Type: Scatter | Mark: Dot | X List: list3 | Y List: t3 |

9. Set up the window using ◆ [WINDOW] with the following entries:

   - **xmin = 350**
   - **xmax = 1400**
   - **xscl = 100**
   - **ymin = -1**
   - **ymax = 7**
   - **yscl = 0**
   - **xres = 1**

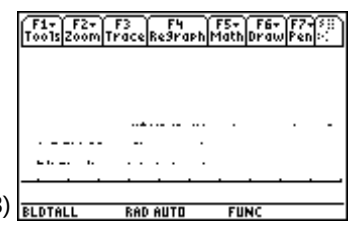
   (See screen 12.)

(12)




10. Press ◆ [GRAPH] (screen 13).

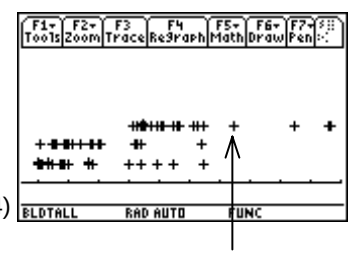
(13)




11. If the graph is difficult to see, go back to the Plot Setup screen (step 8) and change the mark in **Plot 1**, **Plot 2**, and **Plot 3** to **+** (plus) (screen 14).

You looked at the dotplot for Philadelphia buildings in Topic 2, but the additional information gathered from the multiple dotplots over the parallel boxplots is a cluster of three buildings in Seattle around 700 feet, with a gap of over 100 feet from the smaller buildings. New York City has a fourth possible outlier at 1046 feet (the Chrysler Building).

(14)




© 2001 TEXAS INSTRUMENTS INCORPORATED

## Back-to-Back Stemplots

Use the sorted values in **list1**, **list2**, and **list3** to create the following stemplots as you did in Topic 2.

***Note:*** *The back-to-back stemplots are modified to include a third list of data.*

| Philadelphia | | Seattle | | New York City |
|---|---|---|---|---|
| 44221100 | 4 | 1 | | |
| 9999885 | * | 556799 | | |
| 0 | 5 | 014 | | |
| City Hall 977 | * | 78 | | |
| | 6 | 11 Space Needle | | |
| | * | | | |
| 40 | 7 | 234 | 7 | 2344 |
| 9 | * | | * | 5555689 |
| | 8 | | 8 | 111 |
| 5 | * | | * | 567 |
| | 9 | 4 Seattle's Columbia Seafirst Center | 9 | 23 |
| One Liberty Place 5 | * | | * | 5 |
| | 10 | | 10 | |
| | * | | * | 5 Chrysler Bldg. |
| | 11 | | 11 | |
| | * | | * | |
| | 12 | | 12 | |
| | * | | * | 5 Empire State Bldg. |
| | 13 | | 13 | |
| | * | | * | 67 Two & One World Trade Center |

© 2001 TEXAS INSTRUMENTS INCORPORATED

The previous stemplots show all the data to the nearest ten feet. All cities lists are skewed to taller values, with New York City having the majority of the taller buildings and Philadelphia the majority of the smaller buildings. The variability, clusters, gaps, and outliers are consistent with what you observed in the dotplots and modified boxplots.

### Multiple (Sparse) Histograms

To combine the advantages of both the histograms and dotplots, you will compare histograms with many cells. Too many cells and a Plot Setup error will occur. Bucket widths of 25 feet will work. Using this width, the maximum frequency in any cell is **6** for the **phily** data, **4** for the **nyc** data, and **3** for the **seattle** data.

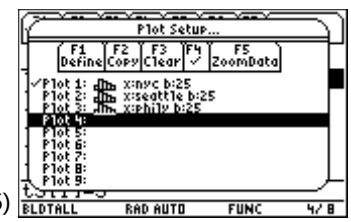
$6 + 1 = 7$, $7 * 3 = 21$, so $|ymin| + |ymax| = 21$ and you can fit three histograms on one graph screen.

1. From the Stats/List Editor, press [F2] **Plots**, **1:Plot Setup** and [F1] **Define** to create the following three plots with specifications:

| | | | |
|---|---|---|---|
| Plot 1 | Type: Histogram | X: nyc | Bucket width: 25 |
| Plot 2 | Type: Histogram | X: seattle | Bucket width: 25 |
| Plot 3 | Type: Histogram | X: phily | Bucket width: 25 |

(See screen 15.)

(15)



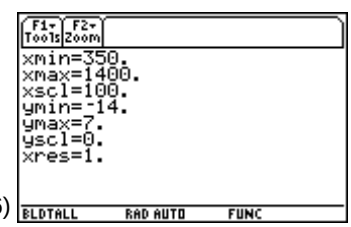

2. Highlight **Plot 2** and **Plot 3** and press [F4] (√) to deselect the plots. Observe in screen 15 that **Plot 1** is the only one checked and active.

3. Set up the window using [♦] [WINDOW] with the following entries:

   - **xmin = 350**
   - **xmax = 1400**
   - **xscl = 100**
   - **ymin = -14**
   - **ymax = 7**
   - **yscl = 0**
   - **xres = 1**

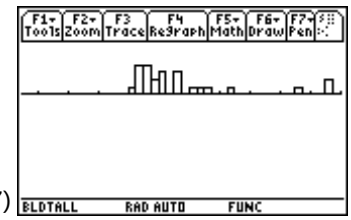
   (See screen 16. The histogram is the top third of the graph screen.)

(16)

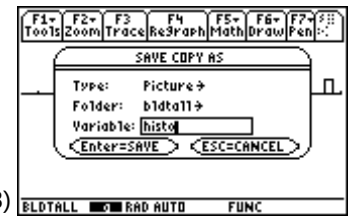



© 2001 TEXAS INSTRUMENTS INCORPORATED

4. Press ◆ [GRAPH] (screen 17).

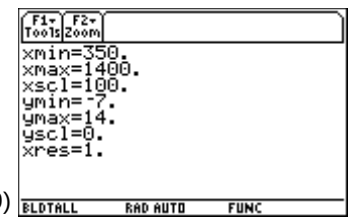
(17)



5. Press F1 **Tools** and select **2:Save Copy As** (screen 18).
6. Select Type: **Picture** and Folder: **BLDTALL**. In the Variable: field, type **histo**. Press ENTER ENTER.
7. Return to the Plot Setup screen and deselect **Plot 1**. Highlight **Plot 1** and press F4 (√) to deselect it.

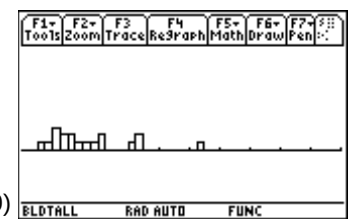
(18)



8. Select **Plot 2** (F4 (√)) with **seattle** data and change the window (◆ [WINDOW]) to the following entries:
   - **xmin = 350**
   - **xmax = 1400**
   - **xscl = 100**
   - **ymin = -7**
   - **ymax = 14**
   - **yscl = 0**
   - **xres = 1**

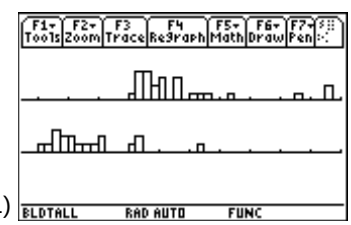
   (See screen 19.)

(19)



9. Press ◆ [GRAPH] for the middle histogram (screen 20).
10. Press F1 **Tools**, select **1:Open picture histo**, and then select Type: **Picture**.

(20)



11. Press ENTER and the top two graphs are displayed (screen 21).
12. Repeat steps 5 and 6 corresponding to screen 18 to save these graphs in place of the old histogram.

(21)


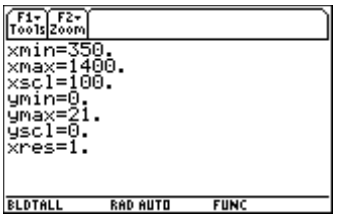

© 2001 TEXAS INSTRUMENTS INCORPORATED

13. From the Plot Setup menu, deselect **Plot 2**, select **Plot 3** with **phily** data, and change the window ([◆] [WINDOW]) to the following entries:

    - **xmin = 350**
    - **xmax = 1400**
    - **xscl = 100**
    - **ymin = 0**
    - **ymax = 21**
    - **yscl = 0**
    - **xres = 1**

    (See screen 22.)

(22)



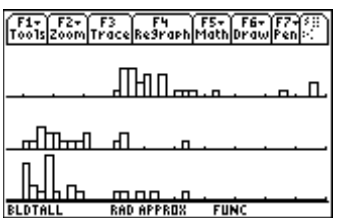

14. Press [◆] [GRAPH] for the bottom histogram.

15. Press [F1] **Tools**, select **1:Open picture histo**, and then select Type: **Picture**.

16. Press [ENTER] to view all three histograms (screen 23).

Skewness, clusters, gaps, and outliers are all shown in relationship to the other data sets.

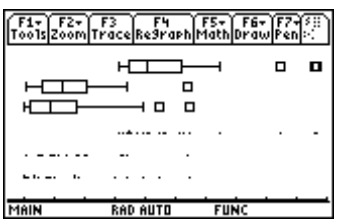
(23)




### Parallel Boxplots with Multiple Dotplots

Screen 24 gives two type comparisons on the same screen. Can you duplicate it?

(24)




© 2001 TEXAS INSTRUMENTS INCORPORATED