Chapter 1: Exploring Data

1.1 Describing Distributions with Graphs

Statistics is the science of data. We begin our study of statistics by mastering the art of examining data. In this chapter of YMS, you learn how to make a number of displays including dotplots, stemplots, histograms, and ogives. It is good practice to construct these plots by hand to gain a better understanding of their meaning and connections to your data. However, once you've mastered construction by hand, our TI's are capable of making basic univariate plots such as histograms, boxplots, and modified boxplots. *Note: Do not rely on your calculator to make these plots until you have mastered constructing them by hand!*

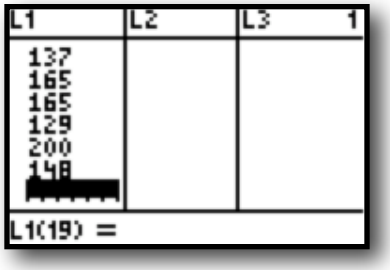
Displaying Univariate Data

Consider the following data on the Survey of Study Habits and Attitudes (SSHA) scores for 18 female college students. The test evaluates motivation, study habits, and attitudes toward school:

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	129	200	148

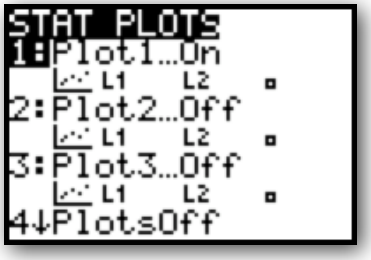
To make a histogram or boxplot on our TI, we must first enter our data. Entering data on the TI is easy. Data can be stored in Lists in a spreadsheet program under the [STAT] menu.

1. Press [STAT] **1: Edit…**
2. Enter the "SSHA" data into **L1**
3. Enter all 18 values into the list, pressing [ENTER] after each value.



To view a plot of the data, you need to set up your statistics plots:

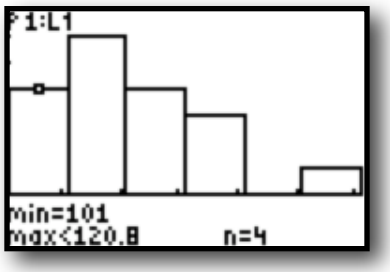
1. Press , **Y= (STAT PLOT**)
2. Select **1: Plot1…**
3. Press [ENTER] to highlight **On**
4. Select the **Histogram** option under **Type:**



You are now ready to view a histogram of the "SSHA" data. To do this, you need to set your window to the appropriate values. You can do this by changing the parameters in the **WINDOW** mode, or you can **Zoom** directly to the data.

To zoom directly to the histogram:

1. Press **ZOOM**
2. Select **9: ZoomStat**
3. Describe the plot *in the context of the problem*
4. Press **TRACE** to see categories and frequencies



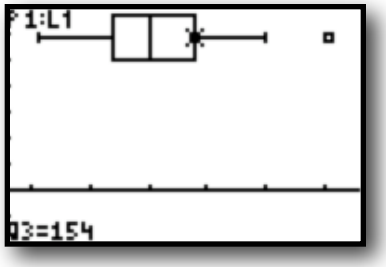
Using the "SSHA" data, select **modified boxplot** to get another view of the distribution.

1.

2.

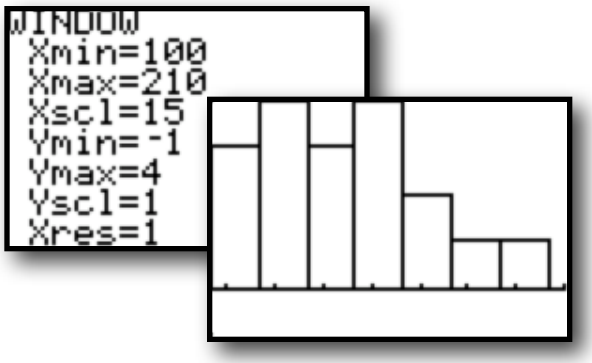
3.

4.



To set the window parameters for a histogram or boxplot yourself:

1. Press **WINDOW**
2. Set **Xmin** and **Xmax** to reflect the minimum and maximum of your dataset
3. Set **Ymin** to –1 and **Ymax** to the largest frequency
4. Set **Yscl** to equal your desired category width
5. Press **Graph** to see your plot



Comparing Data Displays

Throughout the course of your studies, you may be asked to compare sets of univiarate data. Your calculator has the ability to display two boxplots on the same screen to allow for easy comparison. *Again, do not rely on your calculator until you understand how to do it by hand!*

Consider the following data on home run counts for Barry Bonds and Hank Aaron.

Barry Bonds

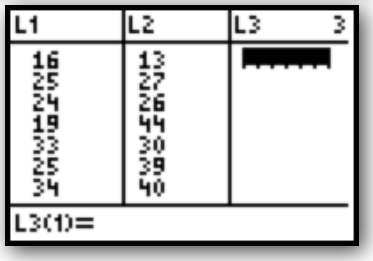
16	25	24	19	33	25	34	46	37	33	42	40	37	34	49	73	46	45	45

Hank Aaron

13	27	26	44	30	39	40	34	45	44	24	32	44	39	29	44	38	47	34	40

1. Enter the Bonds data into **L1**
2. Enter the Aaron data into **L2**

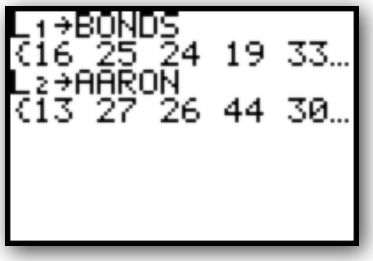
   You do not need to put the data in order. If this is desired, you can "sort" the list using the **SortA(** command in the **LIST** menu.



3. Save the data for future reference
4. On your "homescreen", enter the following

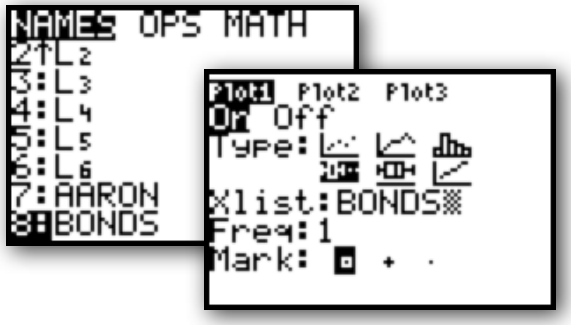
   2ND [1] **(L1)** [STO▸] **ALPHA B O N D S** [ENTER]
   2ND [2] **(L2)** [STO▸] **ALPHA A A R O N** [ENTER]

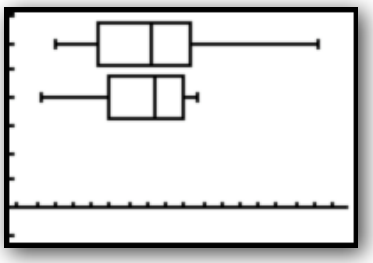
   Your lists have now been stored for future reference.



5. Press **2ND Y= (STAT PLOT)**
6. Set **Plot1 On Type: Modified Boxplot**
7. Set **Xlist: to 2ND STAT (LIST) X:BONDS**

8. Press **2ND Y= (STAT PLOT)**
9. Set **Plot2 On Type: Modified Boxplot**
10. Set **Xlist: to 2ND STAT (LIST) X:AARON**





11. Press **ZOOM 9:ZoomStat**

12. Compare the home run counts for Bonds (top) and Aaron (bottom). Don't forget to interpret the *SOCS* (Shape, Outliers, Center, Spread) for each batter!

1.2 Describing Distributions with Numbers

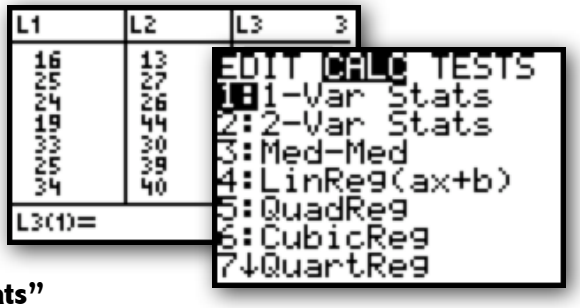
When you first encounter a dataset, it is a good habit to study a graphical display and estimate the SOCS. However, for a more detailed understanding of data, we must calculate numeric summaries of the center and spread. *Note: Be sure you understand how the following measures are calculated before relying on the TI to do the mechanics for you.*

The most common measures of center for a dataset are **mean** ($\bar{x}$) and **median** (Q2). The most common measures of spread/variability for a dataset are **range** (max-min), **interquartile range "IQR"** (Q3-Q1), and **standard deviation** (s_x).

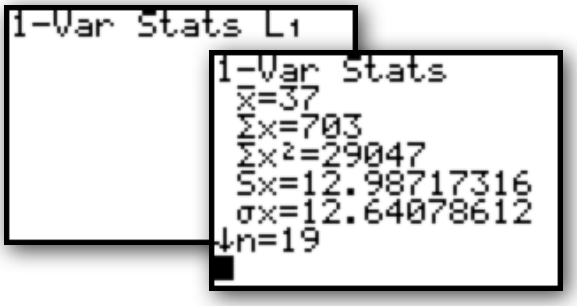
Calculating Numeric Summaries

The calculation of each of these measures, especially the standard deviation, can be quite tedious. Thankfully, the TI can automate those calculations for us. Like plotting data, the calculator requires that you enter the dataset before it can report a numeric summary. If you haven't done so already, enter the Bonds and Aaron data into [STAT] **Edit… L1** and **L2**, respectively.

1. Enter data in to into [STAT] **1:Edit…**
2. Press [STAT] **CALC 1:1-Var Stats** [ENTER]

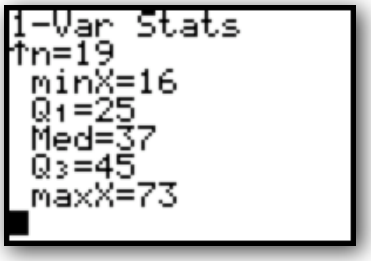


3. Your homescreen should read **"1-Var Stats"**
4. Press **2ND** [1] **(L1)** [ENTER]
5. A numeric summary of the Bonds data should appear.
6. Repeat Steps 2 through 4 for **L2** to get a numeric summary of the Aaron data.



7. Scroll down on each numeric summary to see the 5-number summary.

Remember to interpret the numeric summary in the context of the problem!



AP® Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to perform an exploratory data analysis. Remember, the calculator can be used to automate your calculations and provide basic data displays…however, it is your job to provide the contextual interpretation! Never answer a question by just copying a calculator plot or by simply listing the 1-variable statistics…be sure to label and interpret your analysis!

When making a plot:

- Be careful inputting your data
- Choose an appropriate plot
- Use the modified boxplot if you want to see if outliers exist
- Sketch the plot and LABEL axes!
- Interpret the SOCS of the graph in the context of the problem
- For comparisons, be sure to label each dataset on your plot

When calculating numeric summaries:

- Be careful inputting your data
- Choose the appropriate measures of center and spread
    - Mean and standard deviation, or 5-number summary
    - Be sure to refer to the sample standard deviation s_x, not
- Interpret the measures in the context of the problem
- When comparing datasets, be sure to compare the center and spread for each dataset in the context of the problem
- Know how to use the 1.5 IQR rule to determine outliers
- Be able to justify outliers on a modified boxplot by using this rule