

**Vocabulary for properly describing statistical plots -  
ALWAYS mention these characteristics when describing a plot**

1. Shape

- Normal, uniform, symmetric, skewed left or right, bimodal, etc. Center and spread

2. Center - usually mean or median

3. Outliers and other unusual features

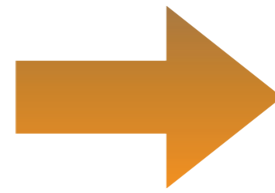
4. Spread - spread, range, interquartile range, standard deviation

- Clusters - many points closely grouped together
- Gaps - no points for a certain part of the graph

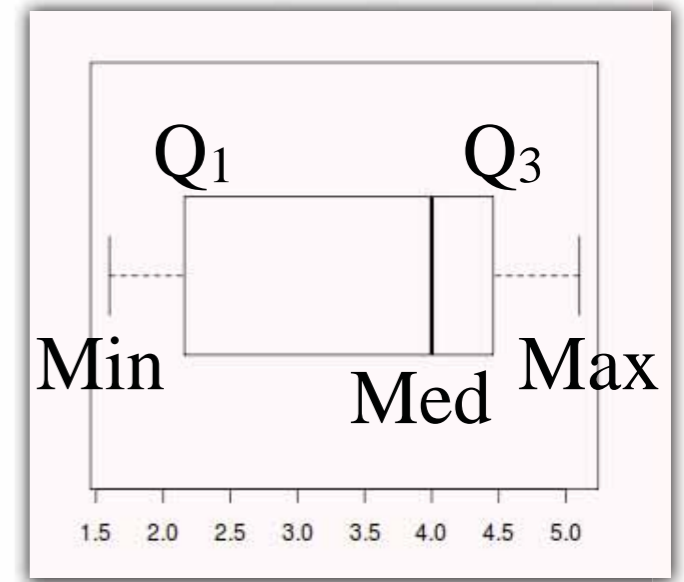
- Remember your SOCS

- If asked to compare distributions use words like “smaller than”, “larger than”, “similar to”, etc. Writing SOCS for both data sets will receive no credit.

# 5 Number Summary



# Boxplot



- Minimum

- $Q_1$  - 25% of the data lie below this marker

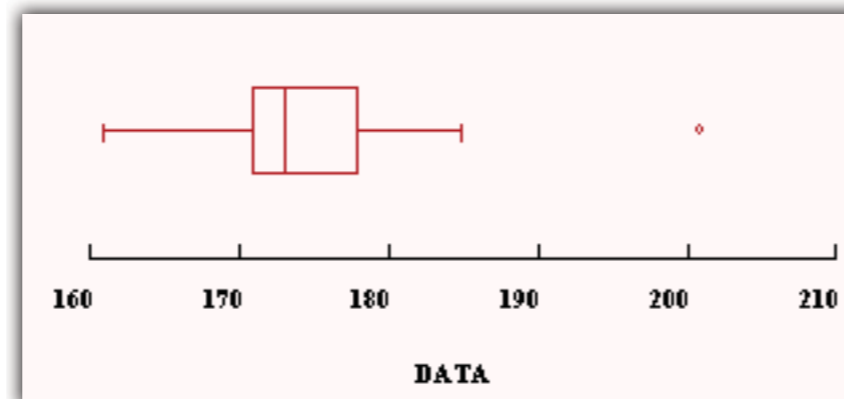
- Median - Middle score if the number of scores is odd, average of the two middle scores if the number is even

- $Q_3$  - 75% of the data lie below this marker

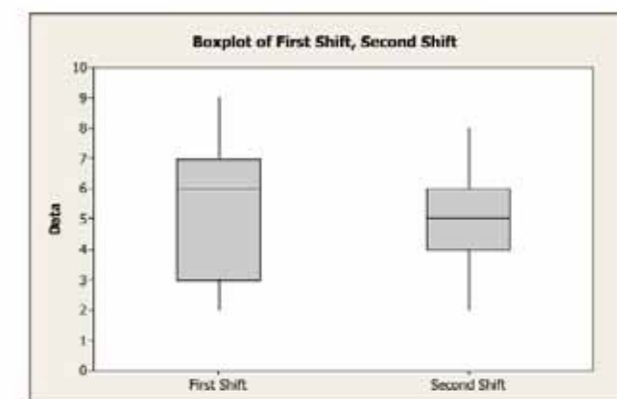
- Maximum

## Modified Boxplot

**50% of the data lie within the IQR**



## Parallel Boxplots



**Percentiles** - value such that \_\_\_\_ % of the observations in the data set fall *below* that value

$Q_1 = 25\text{th percentile}$

Median = 50th percentile

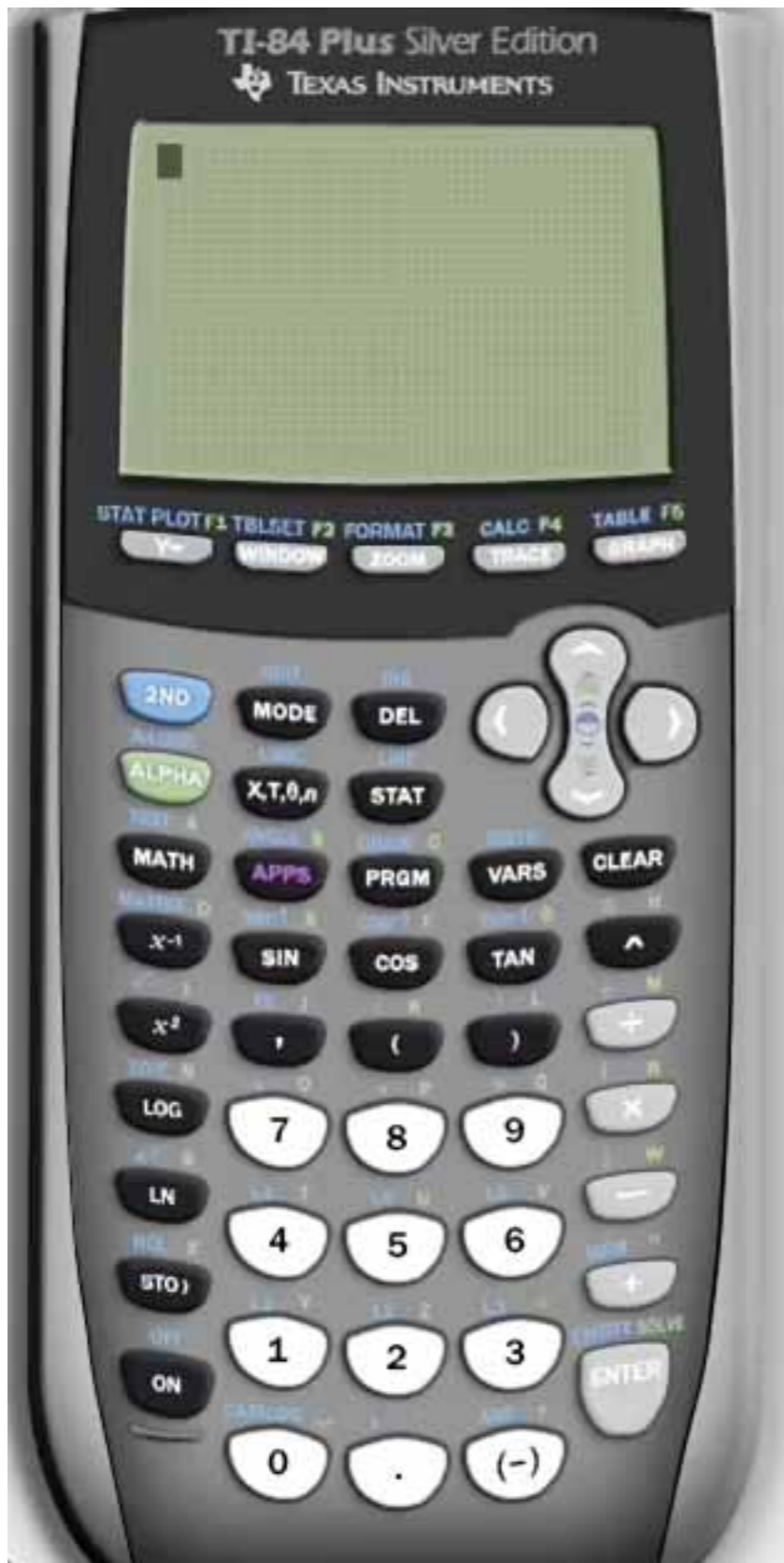
$Q_3 = 75\text{th percentile}$

## How to Calculate Outliers

1. Calculate  $IQR = Q_3 - Q_1$
2. Multiply IQR by 1.5
3. 
$$\left. \begin{array}{l} Q_1 - 1.5 * IQR \\ Q_3 + 1.5 * IQR \end{array} \right\} \text{Safety Zone}$$

Any data points outside the safety zone are considered outliers

# Let's finish with a look at a box plot on the calculator

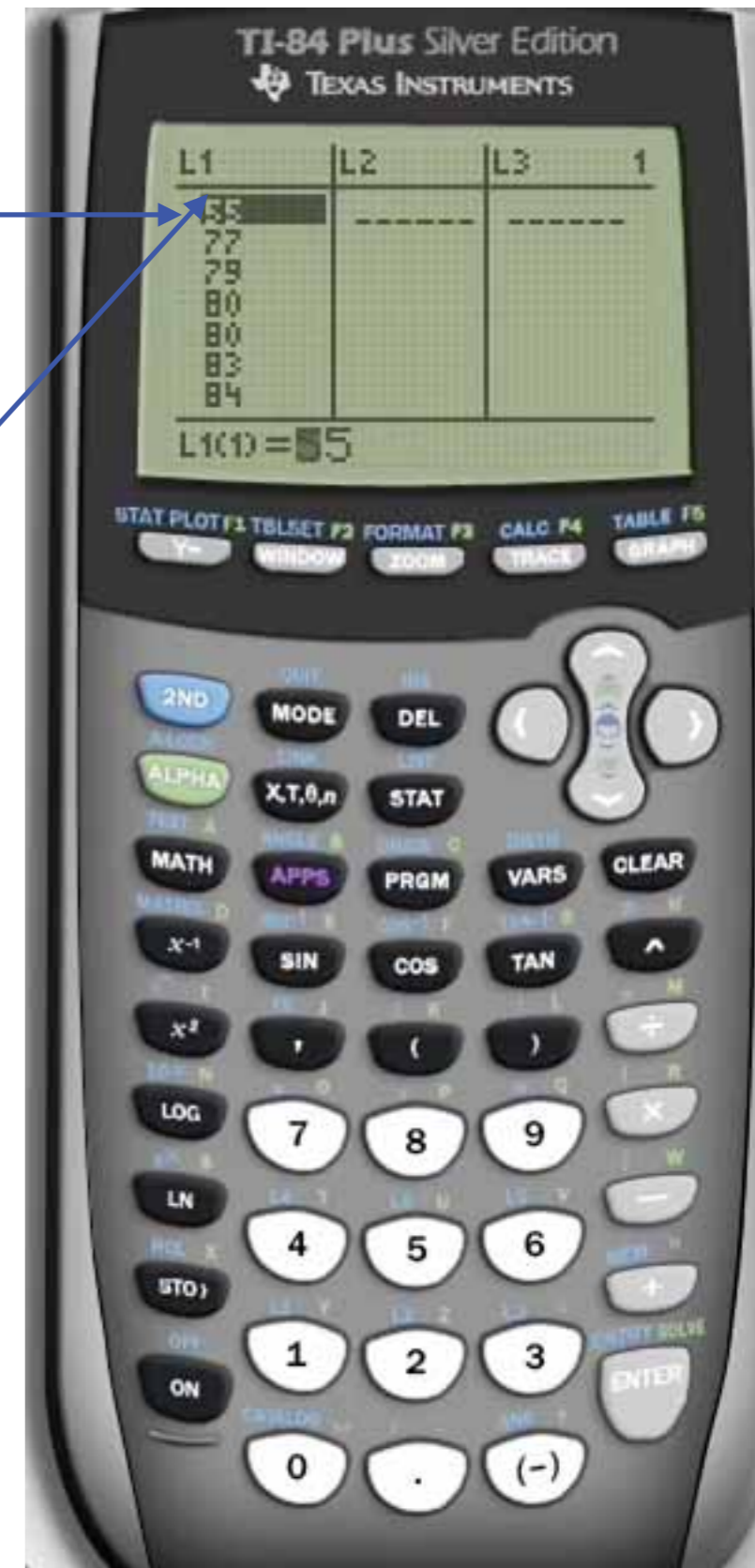


Let's try changing the minimum value from 75 to 55 to see how it changes the graph

Notice that 55 is an outlier because it is separate from the rest of the box plot.

Notice also that the interquartile range is still 9. That is,  $Q_1 = 80$  and  $Q_3 = 89$

Despite the calculator's ability to determine the outlier you will have to know the formula to determine outliers. Let's do this example by hand



# Let's finish with a look at a box plot on the calculator

Notice that 55 is an outlier because it is separate from the rest of the box plot.

Notice also that the interquartile range is still 9. That is,  $Q_1 = 80$  and  $Q_3 = 89$

Despite the calculator's ability to determine the outlier you will have to know the formula to determine outliers. Let's do this example by hand

Remember that you may be asked to do these by hand when working with Boxplots

$$IQR = Q_3 - Q_1 = 89 - 80 = 9$$

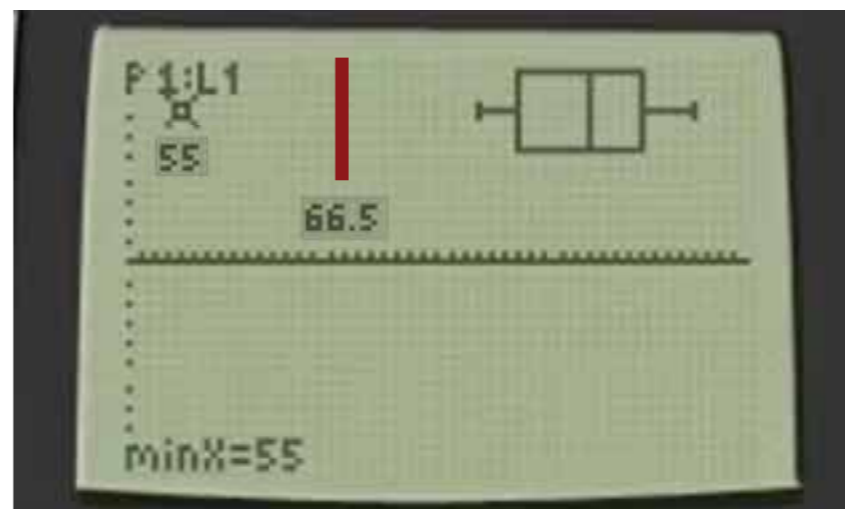
$$1.5(IQR) = 1.5 \cdot 9 = 13.5$$

So anything to the left of  $80 - 13.5 = 66.5$

And anything to the right of  $89 + 13.5 = 102.5$

Would be considered an outlier

Since  $55 < 66.5$  it is considered an outlier



Mean -  $\frac{\sum x_i}{n}$  *not scary - just means to add all of the x values*

Median - Middle #

Mode - # occurring most

Midrange

Measures of Center

Resistant ← to outliers → Non-Resistant Statistics

- Median
- IQR
- Mode

- Range
- Minimum
- Maximum
- Mean

These two will be dealt with in the next section

- Standard Deviation
- Variance