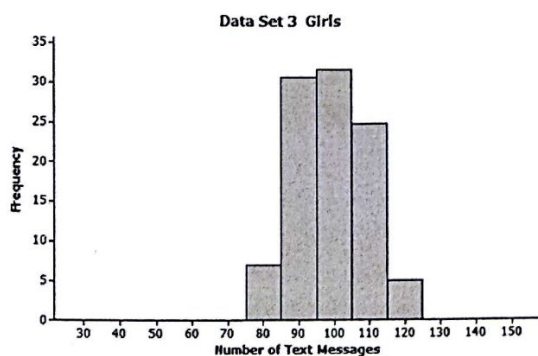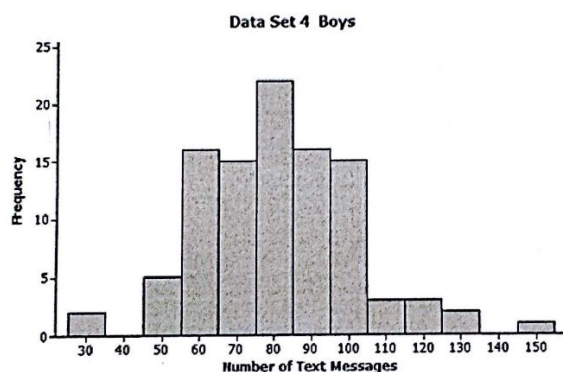# Warm-Up Day 2   Key

Data Set 3 consists of data on the number of text messages sent in one month for 100 teenage girls who have a cell phone. Data Set 4 consists of data on the number of text messages sent in one month for 100 teenage boys who have a cell phone. Histograms of the two data sets are shown below.

Data Set 3 Girls

Data Set 4 Boys

Shape
Outliers
Center
Spread

5. Describe the data distribution of number of text messages for the girls (Data Set 3). Be sure to comment on center, spread and overall shape.

approx normal
symmetric
range 75 to 125
center around 100
no outliers
short tails
uni – modal

6. Are Data Set 3 and Data Set 4 centered in about the same place? If not, which one has the greater center?

7. Which of Data Set 3 and ⟨Data Set 4⟩ has greater spread?
   ranges 30 – 150

8. On average, did the girls (Data Set 3) or the boys (Data Set 4) send more text messages?

girls –
avg looks
higher

6) Set 3 – centered around 100
   Set 4 – centered around 80

STAT → CALC
1-var Stats

1. Suppose your six exam grades in a course are 52, 69, 75, 86, 86, and 92. Compute your final course grade using a 10-point grading scale (90-100 A, 80-89 B, etc.) by calculating:
   a. Mean $\bar{X}$ 76.7
   b. Median 80.5
   c. Mode 86 (appears most)
   d. Midrange $\dfrac{max + min}{2} = \dfrac{92+52}{2} = 72$

2. A magazine gave the following data for the number of calories in a meal hot dog for each of 17 brands:

   173, 191, 182, 190, 177, 147, 146, 138, 175, 136, 179, 153, 107, 195, 135, 140, 138

   Find the mean, median, mode, and midrange for the number of calories in a meat hot dog in the 17 brands. If necessary, round answers to the nearest tenth of a calorie.

   mean: 158.9   midrange: 151
   median: 153
   mode: 138

3. Find all four measures of central tendency of salary for the top 10 paid Baseball Players. (*Wikipedia*)

Name	Position	Team(s)	Salary	Ref
Alex Rodriguez	3B	New York Yankees	$32,000,000	[16]
Vernon Wells	OF	Los Angeles Angels	$26,187,500	[17]
CC Sabathia	SP	New York Yankees	$24,285,714	[18]
Mark Teixeira	1B	New York Yankees	$23,125,000	[19]
Joe Mauer	C	Minnesota Twins	$23,000,000	[20]
Johan Santana	SP	New York Mets	$21,644,707	[21]
Todd Helton	1B	Colorado Rockies	$20,275,000	[22]
Miguel Cabrera	1B	Detroit Tigers	$20,000,000	[23]
Roy Halladay	SP	Philadelphia Phillies	$20,000,000	[24]
Ryan Howard	1B	Philadelphia Phillies	$20,000,000	[25]

   mean: 23,051,792
   median: 22,322,354
   mode: 20,000,000
   midrange: 26,000,000

4. Use the following frequency distribution to find each measure of central tendency.

Score	Frequency	
1	2	II
2	4	IIII
3	3	III
4	1	I

   min: 1 $\dfrac{1+4}{2}$

   Find the mean, median, mode, and midrange of each set of data below. Which measure of central tendency is the most telling for the data? Are any of them the same?

   mean: 2.3
   median: 2
   mode: 2
   midrange: 2.5

The *range* indicates the total spread of the data, and is calculated by finding the difference between the highest and lowest data values in a data set.

The *standard deviation* for a data set is a measure of the dispersion of a set of data from its mean. The more spread apart the data, the higher the standard deviation. Standard deviation is calculated as the square root of the variance.

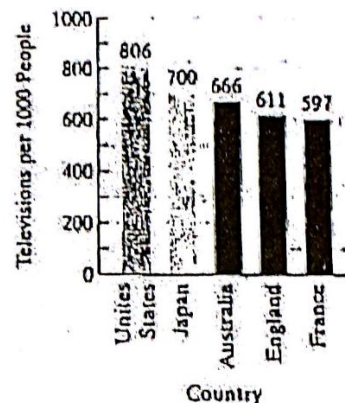To compute the standard deviation for a data set, do the following:

1. Enter the data into L_1
2. 2nd-List-Math-stdDev(
3. Put L_1 into the parentheses and hit enter
4. If you just want the variance, choose that one from the list or square the standard deviation.

Problem Set

1. The average person living in the United States watches about 1551 hours of television per year. That's equal to nearly 65 straight days, or 18% of one year. Figure 12.11 shows the five countries with the most televisions per 1000 people. Find the range of televisions per 1000 people for the five countries with the most televisions per capita. max − min
   $$806 - 597 = 209$$

2. a. Find the range for this set of data: 16, 22, 28, 28, 34   18

   b. Find the range for this set of data: 312, 783, 219, 312, 426, 219
   564

Countries With the Most Televisions per Capita

806 700 666 611 597

Country

Source: International Telecommunications Union

FIGURE 12.11

3. Figure 12.11, showing the five countries with the most televisions per capita, is repeated in the margin. Find the standard deviation for the number of televisions per 1000 people. $S_x = 83.7$

4. Find the standard deviation of the data items in each of the samples shown below.

   Sample A: 17, 18, 19, 20, 21, 22, 23   2.16
   Sample B: 5, 10, 15, 20, 25, 30, 35   10.8

5. Two fifth grade classes have nearly identical mean scores on an aptitude test, but one class has a standard deviation three times that of the other. All other factors being equal, which class is easier to teach and why?

   lower standard deviation
   no outliers
   smaller spread

   20%   100%

Univariate Data Analysis – Super Bowl Points!   Name: **Key**

Super Bowl XLV	Super Bowl XLIV	Super Bowl XLIII	Super Bowl XLII
Feb. 6, 2011 Cowboys Stadium North Texas MVP: Aaron Rodgers, QB, Green Bay » Green Bay Packers 31 Pittsburgh Steelers 25 Go To: Select...	Feb. 7, 2010 Miami Gardens Florida MVP: Drew Brees, QB, New Orleans » New Orleans Saints 31 Indianapolis Colts 17 Go To: Select...	Feb. 1, 2009 Raymond James Stadium Tampa, Florida MVP: Santonio Holmes, WR, Pittsburgh » Pittsburgh Steelers 27 Arizona Cardinals 23 Go To: Select...	Feb. 3, 2008 U. of Phoenix Stadium Glendale, Arizona MVP: Eli Manning, QB, New York » New York Giants 17 New England Patriots 14 Go To: Select...
Super Bowl XLI	Super Bowl XL	Super Bowl XXXIX	Super Bowl XXXVIII
Feb. 4, 2007 Dolphin Stadium Miami, Florida MVP: Peyton Manning, QB, Indianapolis » Indianapolis Colts 29 Chicago Bears 17 Go To: Select...	Feb. 5, 2006 Ford Field Detroit, Michigan MVP: Hines Ward, WR, Pittsburgh Seattle Seahawks 10 » Pittsburgh Steelers 21 Go To: Select...	Feb. 6, 2005 Alltel Stadium Jacksonville, Florida MVP: Deion Branch, WR, New England » New England Patriots 24 Philadelphia Eagles 21 Go To: Select...	Feb. 1, 2004 Reliant Stadium Houston, Texas MVP: Tom Brady, QB, New England Carolina Panthers 29 » New England Patriots 32 Go To: Select...

I. List the scores of the winning teams, in order least to greatest:

17	21	24	27	29	31	31	32

Find the middle number (if there is not a middle, take the average of the middle 2): this is the <u>median</u>.

Median = **28**
**Q2**

$$(3 \quad 4 \quad 5) \quad 6 \quad (7 \quad 8 \quad 9)$$
↑ ↑ ↑

Draw brackets around the lower set of the data (4 numbers), and the upper set of the data (4 numbers), leaving out the median.

We need to find the <u>lower quartile</u> and the <u>upper quartile</u> of the data. These are the medians of the lower and upper sets of data. Since there are an even number of data points, we have to find the average of the two middle data points.

Lower Quartile = **22.5**   Upper Quartile = **31** $\ast IQR = \dfrac{31 - 22.5}{= 8.5}$
Q1   **Q3**

Draw a <u>box-and-whisker</u> plot below, labeling the points with the appropriate values.

Q1 _____ **Q2** _____ **Q3**

| 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
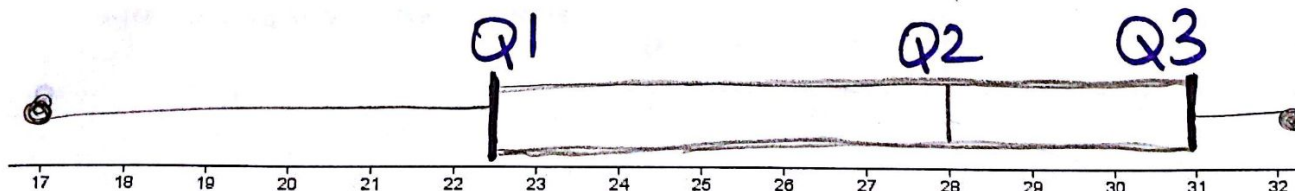
anything outside is an outlier ← 1.5 IQR Rule : $(8.5)(1.5) = 12.75$
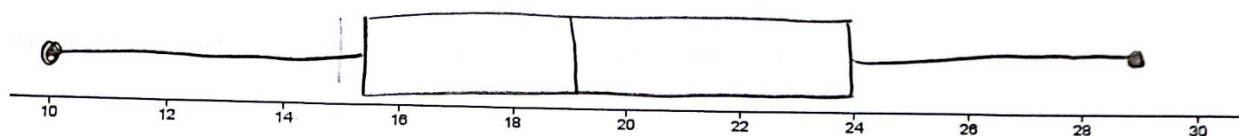$\{$ Q3 + 12.75 = 43.75
$\{$ Q1 − 12.75 = 9.75

Using the scores of the losing teams over 8 consecutive years find the points of interest and make another box-and-whisker plot.

Data, least to greatest:

10	14	17	17	21	23	25	29

Median = __19__   Lower Quartile = __15.5__   Upper Quartile = __24__

Box and Whisker Plot:

III.   Use the data below to do a box-and-whisker plot on game attendance for the first 9 weeks of Giants games.

REGULAR SEASON

Wk	Date	Game	Time (ET)	Attendance
1	Sep 11	NYG 14 @ WAS 28	Final	80,121
2	Sep 19	STL 16 @ NYG 28	Final	78,290
3	Sep 25	NYG 29 @ PHI 16	Final	69,144
4	Oct 02	NYG 31 @ ARI 27	Final	60,496
5	Oct 09	SEA 36 @ NYG 25	Final	78,650
6	Oct 16	BUF 24 @ NYG 27	Final	79,243
7	Bye			
8	Oct 30	MIA 17 @ NYG 20	Final	79,302
9	Nov 06	NYG 24 @ NE 20	Final	68,756

Data (rounded in thousands), least to greatest:

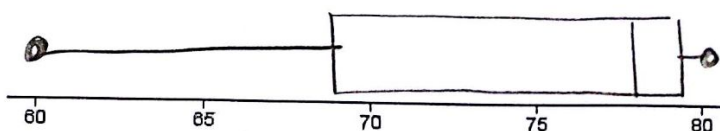| 60	69	69	78	79	79	79	80	X

Median = __78.5__

Draw brackets around the lower sets of numbers and the higher sets of numbers excluding the median and find the following:

Lower Quartile = __69__   Upper Quartile = __79__

Box and Whisker Plot:

and Leaf Plots help organize data to show trends. Typically the tens digit will be the stem, and the ones digit will be
: leaf. Fill in the Stem-and-leaf plots below for the Winning and Losing Super Bowl scores.

Winning

Stem	Leaf
3	2 1 1
2	9 7 4 1
1	7

Losing

Stem	Leaf
2	9 5 3 1
1	7 7 4 0

$3|2 = 32$

Side-by-side Stem and Leaf

	W				L	
Leaf		Stem		Leaf
1 1	2	3	
1 4 7 9	2	9 5 3 1
7	1	7 7 4 0

Range of a set of data is the difference between the highest and lowest points. Find the range of each set above.

Range of Winning Scores = 15

Range of Losing Scores = 19

Midrange is the average of the highest and lowest points of a data set. Find the midrange of each set above.

Midrange of Winning Scores = 24.5

Midrange of Losing Scores = 24

19.5

Sometimes a set of data has more than two digits. When this is the case, look at the data to decide what to use as your
"stem" – then, list the rest of the digits as the "leaf" putting a comma in between.

REGULAR SEASON

Wk	Date	Game	Time (ET)	Attendance
1	Sep 11	NYG 14 @ WAS 28	Final	80,121 ✓
2	Sep 19	STL 16 @ NYG 28	Final	78,290 ✓
3	Sep 25	NYG 29 @ PHI 16	Final	69,144 ✓
4	Oct 02	NYG 31 @ ARI 27	Final	60,496 ✓
5	Oct 09	SEA 36 @ NYG 25	Final	78,650 ✓
6	Oct 16	BUF 24 @ NYG 27	Final	79,243 ✓
7	Bye			
8	Oct 30	MIA 17 @ NYG 20	Final	79,302 ✓
9	Nov 06	NYG 24 @ NE 20	Final	68,756 ✓
10	Nov 13	NYG 20 @ SF 27	Final	69,732 ✓

In this case, it is probably best to use the thousands digit as
the stem, and the hundreds, tens, and ones as the leaf.

Stem	Leaf
6	0496, 8756, 9144, 9732
7	8290, 8650, 9243, 9302
8	0121