Name:_____ Class:_____ Date:_____

One-Variable Statistics Notes

Part I: Types of Data

The two main uses of statistics:

1. _____
2. _____

Important Vocabulary:

1. Data: a _____ of _____ in context.
2. Population: a set of individuals that we wish to _____ and/or make _____ about.
3. Sample: a _____ of the population that data is collected from.
4. Individual: _____ of a _____.
5. Variable: _____ recorded about each individual in a data set.

Types of Variables:

Categorical: records _____ or _____ of an individual.

Examples: _____

Quantitative: variable that _____ a characteristic of an individual.

Examples: _____

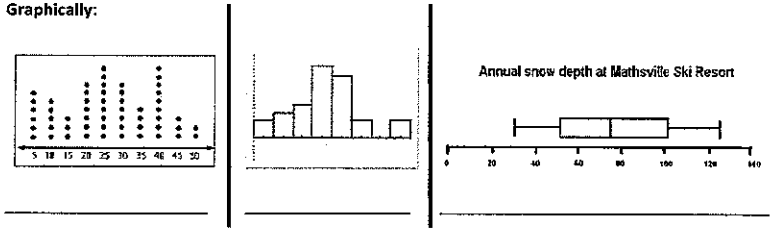
Categorical	Quantitative

Name:_____ Class:_____ Date:_____

Two ways to describe data:

1. _____
2. _____

Graphically:



Numerically:

Measures of Center - _____, _____

Measures of Spread - _____, _____, _____

Part 2: Frequency Tables and Histograms

Activity: Link-Up

Using the paperclips at your table, see how many you can link together in 30 seconds. Record the data for each member of your group below:

Student 1: _____ Student 4: _____

Student 2: _____ Student 5: _____

Student 3: _____ Student 6: _____

Creating a Frequency Table:

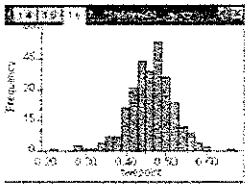
Intervals (MUST BE EQUAL)	Tally Marks	Frequency

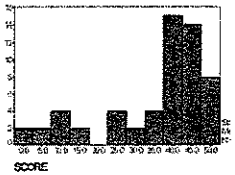
Name:_____ Class:_____ Date:_____

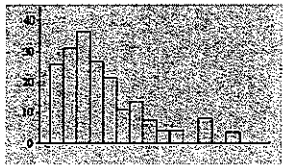
Creating a Histogram:

Describing Distributions:

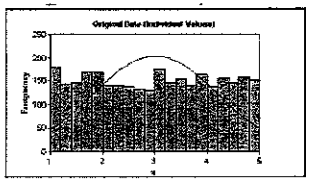
1. _____







_____ _____ _____



3 More Ways to Describe Distributions

1. _____
2. _____
3. _____

_____

Center	Spread	Outliers

Common Core Math 1

One-Variable Data
Homework: Introduction to Statistics

Name _____

Date _____

1. Determine whether the following data is **categorical (C)** or **quantitative (Q)**

   a. The candidate a survey respondent will support in an upcoming election.

   b. The length of time of people's drive to work.

   c. The number of televisions in a household.

   d. The distance kickers for a football team can kick a football.

   e. The number of pages copied in the copy room each day.

   f. The kind of tree in each person's front yard in a neighborhood.

   g. The type of blood a person has.

   h. The jersey numbers of the football team.

   i. The heights of the tallest buildings in the world.

   j. The language spoken by 2000 people coming in to JFK Airport.

2. A math student is interested in figuring out the average price of vehicles at Glentown High School. She takes a sample of 50 cars in the school's parking lot and finds the average value to be $13, 400.

   a. What is the population?

   b. What are the individuals?

   c. What data is being collected? (Include units if applicable)

   d. What type of data is it (categorical or quantitative)? How do you know?

2011 NFL Rushing Statistics for Top 50 Rushers

Rank	Player	Team	Position	Rushing Attempts	Total Yards for the Season	Average Yards per Attempt	Average Yards per Game	Number of Rushing Touchdowns	Longest Run of the Season*	Total Number of Fumbles
1	Maurice Jones-Drew	JAC	RB	343	1606	4.7	100.4	8	56	3
2	Arian Foster	HOU	RB	278	1224	4.4	94.2	10	43	4
3	Fred Jackson	BUF	RB	170	934	5.5	93.4	6	80T	2
4	Darren McFadden	OAK	RB	113	614	5.4	87.7	4	70T	1
5	LeSean McCoy	PHI	RB	273	1309	4.8	87.3	17	60	1
6	Ray Rice	BAL	RB	291	1364	4.7	85.2	12	70T	2
7	Michael Turner	ATL	RB	301	1340	4.5	83.8	11	81T	3
8	Matt Forte	CHI	RB	203	997	4.9	83.1	3	46	1
9	Adrian Peterson	MIN	RB	208	970	4.7	80.8	12	54	1
10	Marshawn Lynch	SEA	RB	285	1204	4.2	80.3	12	47	3
11	Willis McGahee	DEN	RB	249	1199	4.8	79.9	4	60T	3
12	Ryan Mathews	SD	RB	222	1091	4.9	77.9	6	39	5
13	Steven Jackson	STL	RB	260	1145	4.4	76.3	5	47T	1
14	Frank Gore	SF	RB	282	1211	4.3	75.7	8	55	2
15	Beanie Wells	ARI	RB	245	1047	4.3	74.8	10	71	4
16	Reggie Bush	MIA	RB	216	1086	5	72.4	6	76T	4
17	Cedric Benson	CIN	RB	273	1067	3.9	71.1	6	42	5
18	DeMarco Murray	DAL	RB	164	897	5.5	69	2	91T	1
19	Shonn Greene	NYJ	RB	253	1054	4.2	65.9	6	31	1
20	Chris Johnson	TEN	RB	262	1047	4	65.4	4	48T	3
21	Jahvid Best	DET	RB	84	390	4.6	65	2	88T	0
22	Tim Hightower	WAS	RB	84	321	3.8	64.2	1	22	0
23	Ben Tate	HOU	RB	175	942	5.4	62.8	4	56	4
24	Chris Ivory	NO	RB	79	374	4.7	62.3	1	35T	0
25	Rashard Mendenhall	PIT	RB	228	928	4.1	61.9	9	68	1
26	Michael Bush	OAK	RB	256	977	3.8	61.1	7	44	1
27	Peyton Hillis	CLE	RB	161	587	3.6	58.7	3	24T	1
28	LeGarrette Blount	TB	RB	184	781	4.2	55.8	5	54T	3
29	Ahmad Bradshaw	NYG	RB	171	659	3.9	54.9	9	37	0
30	Evan Royster	WAS	RB	56	328	5.9	54.7	0	28	0

NFL Team Name Abbreviations:

ARI: Arizona Cardinals
ATL: Atlanta Falcons
BAL: Baltimore Ravens
BUF: Buffalo Bills
CAR: Carolina Panthers
CHI: Chicago Bears
CIN: Cincinnati Bengals
CLE: Cleveland Browns
DAL: Dallas Cowboys
DEN: Denver Broncos
DET: Detroit Lions
GB: Green Bay Packers
HOU: Houston Texans
IND: Indianapolis Colts
JAX: Jacksonville Jaguars
KC: Kansas City Chiefs
MIA: Miami Dolphins
MIN: Minnesota Vikings
NE: New England Patriots
NO: New Orleans Saints
NYG: New York Giants
NYJ: New York Jets
OAK: Oakland Raiders
PHI: Philadelphia Eagles
PIT: Pittsburgh Steelers
SD: San Diego Chargers
SEA: Seattle Seahawks
SF: San Francisco 49ers
STL: Saint Louis Rams

Common Core Math 1 Name _____

Univariate Data Date _____

Homework: Frequency Tables and Histograms

1) The following frequency table lists the ages of patrons dining at a local restaurant at 7:00 PM.

Age	Frequency
1-10	1
11-20	4
21-30	5
31-40	7
41-50	6
51-60	5
61-70	1

a) Construct a histogram for the data. Use graph paper!

b) Describe the data distribution in context (shape, center, spread, and any outliers).

2) A bank wants to improve its customer service. Before deciding to hire more workers, the manager decides to get some information on the waiting times customers currently experience. During a week, 50 customers were randomly selected, and their waiting times, in minutes, were recorded.

The data are as follows: 18.5, 9.1, 3.1, 6.2, 1.3, 0.5, 4.2, 5.2, 0.0, 10.8, 5.8, 1.8, 1.5, 1.9, 0.4, 3.5, 8.5, 11.1, 0.3, 1.2, 4.4, 3.8, 5.8, 1.9, 3.6, 2.5, 4.5, 5.8, 1.5, 0.7, 0.8, 0.1, 9.7, 2.6, 0.8, 1.2, 2.9, 3.0, 3.2, 2.8, 10.9, 0.1, 5.9, 1.4, 0.3, 5.5, 4.8, 0.9, 1.6, and 2.2.

a) Construct a frequency table of the data. (Remember to define the classes so that there are approximately 5-7 groupings.)

b) Construct a histogram. Use graph paper!

c) Describe the data distribution in context.

Homework – Representing Data Graphically

Answer the following questions on your own paper. All graphs must be done on graph paper.

1. Suzie surveyed her class with the following question: "How many hours of TV do you typically watch in one week." Below is the data she collected:

8	2	6	4	4	5	6	1	11	10
4	2	1	10	6	10	7	6	2	9
11	7	8	6	9	2	4	2	2	7

a) Create a dot plot of this data.

b) Describe the data distribution in context.

2. The number of representatives each state has in the US Congress depends on the population of the state. Below are the numbers of representatives in each state in the 100th Congress (1988-1990).

7	1	5	4	45	6	6	1	19	10
2	2	22	10	6	5	7	8	2	8
11	18	8	5	9	2	3	2	2	14
3	34	11	1	21	6	5	23	2	6
1	7	27	3	1	10	8	4	9	1

a. State the minimum number of representatives.

b. State the maximum number of representatives.

c. Find the range.

d. Make a frequency table showing the distribution of the number of representatives per state. Use intervals of length 5 (ie 1-5, 6-10...)

e. Using your calculator, create a histogram to represent the data. The x-axis scale should match the intervals in the frequency table above. Use the trace function to help you sketch an accurate histogram below.

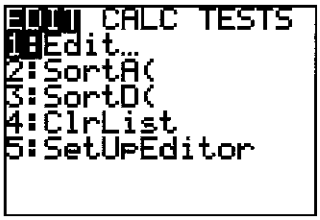
f. What is the window of the histogram you created?

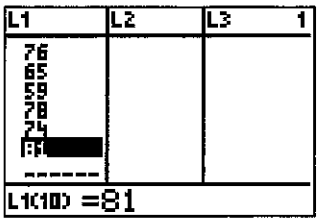
Make a Histogram on Your Calculator!

I'll use the following heights of a group of seventh graders as an example.

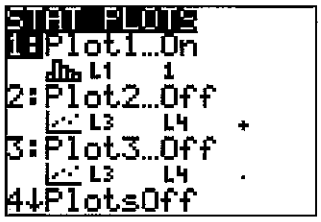
Student	Height (in.)
Megan	61
Morgan	65
Kyle	70
Darren	71
Angie	76
Cassie	65
Brady	59
Chris	78
Cody	74
Maria	81

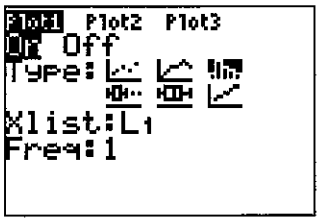
Step 1: Enter the height data in **L1**.
Commands: **STAT → EDIT**
Use 2nd QUIT to exit



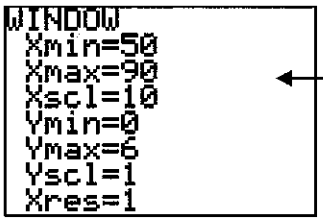


Step 2: Setup the graph.
Commands: **2nd → Y= → 1**
then select "On" and press enter



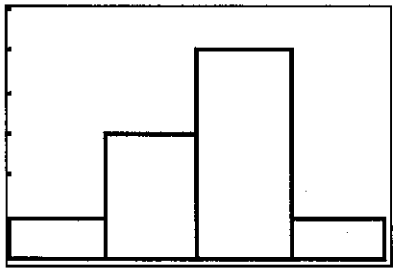


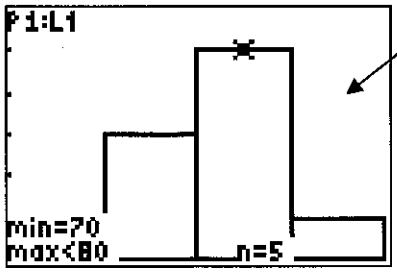
Step 3: Adjust the window to fit the data. Use **Xscl** to adjust the width of the bars. I used 10 here because I want to see how many people are in each category of 10 inches.
Commands: **ZOOM → 9**
Window



Using this window setting, we will have sorted the data into four categories: Heights between 50 and 59 inches, 60 to 69, 70 to 79 and 80 to 89.

Step 4: Look at your graph!
Commands: **Graph**





Using the **Trace** button you can see that there are 5 people in the 70 to 79 inch category.

Mean vs. Median Review

Use the following data set:

Name:_____

Class:_____

Date:_____

32, 43, 38, 33, 41, 39, 40, 35, 36

1) Find the Mean:

2) Find the Median:

Median Example 2) 14, 19, 15, 17, 12, 10

Investigation

We can use the mean and the median of a set of data to describe what is typical about the distribution. Let's use these *measures of center* to describe the distribution of names in a class. Below are twelve names. Count the number of letters in each name and write that number in the column labeled "Number of Letters". Do not count spaces.

Name	Number of Letters
Pete Thomas	
Shaquana Smith	
Stewart Hughes	
Huang Mi	
Richard Lewis	
Virginia Bates	
Angel Mendoza	
Mary Wall	
Danielle Duncan	
Will Jones	
Ana Romero	
Jana Turner	

Use the area below to calculate median and mean.

Median: _____

Mean: _____

1) William Arthur Philip Louis Mountbatten-Windsor is joining the class!! Recalculate the median and mean with him in the class.

Median:_____ Mean: _____

What happened to the median?

What happened to the mean?

Adapted from Data About Us, *Connected Mathematics 2, Grade 6*

2) William Arthur Philip Louis Mountbatten-Windsor decided to move back to England so remove him from the class. However, Ty Lee is now going to join the class. Recalculate the median and mean with him in the class.

Median:_____ Mean: _____

What happened to the median?

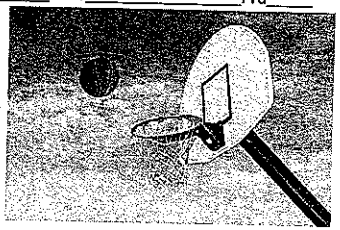
What happened to the mean?

3) The names that we added acted as _____. An _____ is a value that is much larger or much smaller than the rest of the data values.

4) You should use the mean when _____.

5) You should use the median when _____.

Name _____ Date_____ Prd____



Homework – Graded Assignment!

Investigation 4: Mean vs. Median

Discuss the following with your partner or group. Be prepared to share your answers with the class.

The heights of Washington High School's basketball players are: 5 ft 9in, 5 ft 4in, 5 ft 7 in, 5ft 6 in, 5 ft 5 in, 5 ft 3 in, and 5 ft 7 in. A student transfers to Washington High and joins the basketball team. Her height is 6 ft 10in.

1) What is the mean height of the team before the new player transfers in? What is the median height?

2) What is the mean height after the new player transfers? What is the median height?

3) What effect does her height have on the team's height distribution and stats (center and spread)?

4) How many players are taller than the new mean team height? How many players are taller than the new median team height?

5) Which measure of center more accurately describes the team's typical height? Explain.

Measures of Spread: Standard Deviation

So far in our study of numerical measures used to describe data sets, we have focused on the mean and the median. These measures of center tell us the most typical value of the data set. If we were asked to make a prediction about a member of a data set, we would use a measure of center to predict that value. However, measures of center do not give us the complete picture.

Consider the following test scores:

Student	Test 1	Test 2	Test 3	Test 4
Johnny	65	82	93	100
Will	82	86	88	84
Anna	80	99	73	88

Who is the best student? How do you know?

Thinking about the Situation

Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.

1) What is the mean test score for each student?

2) Based on the mean, who is the best student?

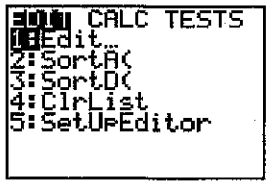
3) If asked to select one student, who would you pick as the best student? Explain.

Finding the mean, median and standard deviation on the calculator.

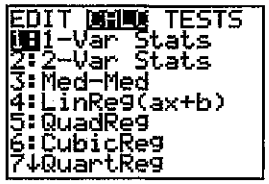
Enter the following data into L_1 in the calculator.

28, 48, 53, 25, 38, 23, 49, 32

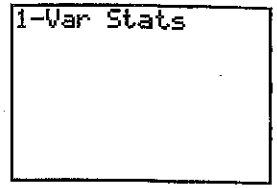
Step 1: The data should already be entered as lists in the calculator. Press STAT.



Step 2: Press right arrow button so that CALC is highlighted and the CALC menu appears (as shown). 1: 1-VAR STATS should already be highlighted.



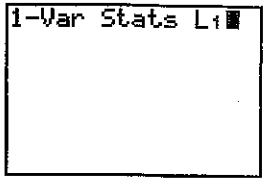
Step 3: Press ENTER. 1-VAR STATS will appear on the Home Screen.



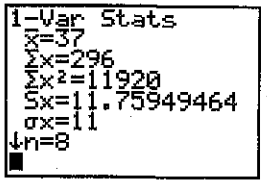
Step 4: Now enter the list names.

For this example, the data are stored in L1.

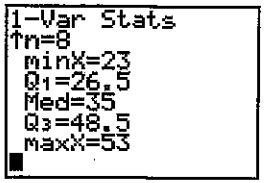
Notice above "1" on the number keypad is "L1" in yellow. Press the yellow 2ND button, then "1" on the keypad. L1 should appear on the screen.



Step 5: Press ENTER. 1-VAR STATS makes many calculations, more than can fit on a screen. Use the down arrow keys to see all the information.



- **Mean**
  - Notation: x with a bar above (the first thing listed under 1-VAR STATS
  - In this example, the mean is 37



- **Median**
  - Notation: MED
  - In this example, the median is 35

- **Standard Deviation**
  - Notation: Sx (Sample)
  - In this example, the standard deviation is 11.759....
  - Notations: σx (Population)
  - We will be using Sx (Sample Standard Deviation) for standard deviation, this is used more frequently because in order to use σx (Population Standard Deviation) we need to have the data for the entire population and that is not often the case. Sample Deviation is the larger of the two values.

Using your calculator, determine Standard Deviation for Johnny, Will, and Anna.

Johnny:_____

Will: _____

Anna: _____

What does this mean? _____

_____

_____

Name: _____

Date: _____

All information taken from:

surveystar.com/startips/jan2013.pdf

* Standard Deviation

Standard Deviation (often abbreviated as "Std Dev" or "SD") provides an indication of how far the individual responses to a question vary or "deviate" from the mean. SD tells the researcher how spread out the responses are – are they concentrated around the mean, or scattered far & wide? Did all of your respondents rate your product in the middle of your scale, or did some love it and some hate it?

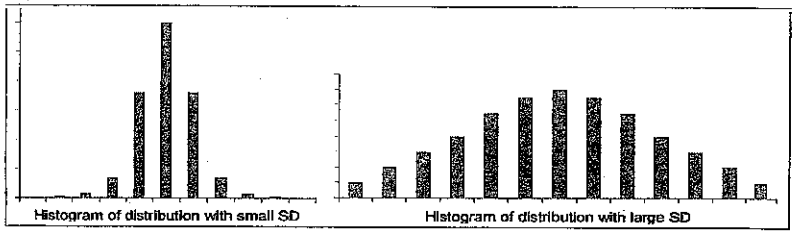
Let's say you've asked respondents to rate your product on a series of attributes on a 5-point scale. The mean for a group of ten respondents (labeled 'A' through 'J' below) for "good value for the money" was 3.2 with a SD of 0.4 and the mean for "product reliability" was 3.4 with a SD of 2.1. At first glance (looking at the means only) it would seem that reliability was rated higher than value. But the higher SD for reliability could indicate (as shown in the distribution below) that responses were very polarized, where most respondents had no reliability issues (rated the attribute a "5"), but a smaller, but important segment of respondents, had a reliability problem and rated the attribute "1". Looking at the mean alone tells only part of the story, yet all too often, this is what researchers focus on. The distribution of responses is important to consider and the SD provides a valuable descriptive measure of this.

Respondent:	Good Value for the Money:	Product Reliability:
A	3	1
B	3	1
C	3	1
D	3	1
E	4	5
F	4	5
G	3	5
H	3	5
I	3	5
J	3	5
Mean	3.2	3.4
Std Dev	0.4	2.1

Two very different distributions of responses to a 5-point rating scale can yield the same mean. Consider the following example showing response values for two different ratings. In the first example (Rating "A") the Standard Deviation is zero because ALL responses were exactly the mean value. The individual responses did not deviate at all from the mean. In Rating "B", even though the group mean is the same (3.0) as the first distribution, the Standard Deviation is higher. The Standard Deviation of 1.15 shows that the individual responses, on average", were a little over 1 point away from the mean.

Respondent:	Rating "A"	Rating "B"
A	3	1
B	3	2
C	3	2
D	3	3
E	3	3
F	3	3
G	3	3
H	3	4
I	3	4
J	3	5
Mean	3.0	3.0
Std Dev	0.00	1.15

Another way of looking at Standard Deviation is by plotting the distribution as a histogram of responses. A distribution with a low SD would display as a tall narrow shape, while a large SD would be indicated by a wider shape.



Histogram of distribution with small SD

Histogram of distribution with large SD

SD generally does not indicate "right or wrong" or "better or worse" – a lower SD is not necessarily more desireable. It is used purely as a descriptive statistic. It describes the distribution in relation to the mean.

Task:

The free throw shooting percentages for the final five years of three NBA legends are shown below.

Michael Jordan: 82%, 79%, 78%, 83%, 83%
Magic Johnson: 86%, 91%, 89%, 91%, 85%
Shaquille O'Neil 56%, 50%, 60%, 51%, 50%

a. Use the data to calculate standard deviation for each of the players.

b. Which player was the most consistent free throw shooter? Explain.

c. Which player was the least consistent free throw shooter? Explain.

d. If you had to pick one of these players for your team, based on free throw stats only, which player would you chose and why?

e. If you had to pick one of these players for your team, based on who you wanted to play on a team with, which player would you chose and why?

Common Core Math 1
One-Variable Data
Measures of Center and the Standard Deviation

Name _____
Date _____

1) Superbowl XLIII featured 2 of the NFL's most unknown offensive linemen. The data sets give the name of the players and their weights (lbs).

Cardinals		Steelers	
Mike Gandy	316	Max Starks	345
Reggie Wells	308	Chris Kemoeata	344
Lyle Sendlein	300	Justin Hartwig	312
Duece Luti	332	Darnell Stapelton	305
Levi Brown	322	Willie Colon	315

a) Find the mean and median weights of both the Steelers' and Cardinals' offensive lines.

b) Calculate the standard deviation of the weights of both the Steelers' and Cardinals' offensive lines.

c) Compare the Steelers' and the Cardinals' offensive lines. How are they different? How are they alike?

d) Assume that the Cardinals' offensive linemen each put on 15 pounds. Calculate the mean, median, and standard deviation for this new group of data.

e) Which statistical values changed compared to the original group of Cardinals' linesmen? Which stayed the same? Why do you think this happened?

Common Core Math 1
One-Variable Data
Measures of Center and the Standard Deviation

Name _____
Date _____

1) Superbowl XLIII featured 2 of the NFL's most unknown offensive linemen. The data sets give the name of the players and their weights (lbs).

Cardinals		Steelers	
Mike Gandy	316	Max Starks	345
Reggie Wells	308	Chris Kemoeata	344
Lyle Sendlein	300	Justin Hartwig	312
Duece Luti	332	Darnell Stapelton	305
Levi Brown	322	Willie Colon	315

a) Find the mean and median weights of both the Steelers' and Cardinals' offensive lines.

b) Calculate the standard deviation of the weights of both the Steelers' and Cardinals' offensive lines.

c) Compare the Steelers' and the Cardinals' offensive lines. How are they different? How are they alike?

d) Assume that the Cardinals' offensive linemen each put on 15 pounds. Calculate the mean, median, and standard deviation for this new group of data.

e) Which statistical values changed compared to the original group of Cardinals' linesmen? Which stayed the same? Why do you think this happened?

Common Core Math 1
Univariate Data
Boxplots and Outliers

Name _____
Date _____

1. Given the following grades on an English 11 test:

91, 98, 87, 76, 100, 45, 72, 85, 92, 88, 87, 90, 91, 66, 100, 99, 67, 85, 79, 80, 85

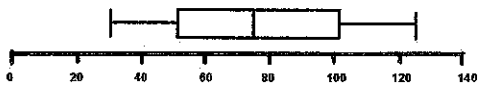
Do you think there will be an outlier?

Calculate the 5 number summary and any outliers.

Remove the outlier and recalculate the 5 number summary. What changes have occurred?

Why would it be useful to remove the outlier to look at this data? Explain a scenario in which we would want to leave the outlier in and a scenario in which we would want to remove the outlier.

2. Given the box plot below, create a scenario that would produce this data.



Boxplots and Outliers Practice

Below is a stem and leaf plot of the amount of money spent by 25 shoppers at a grocery store.

0	3 6
1	0 1 7 8 9
2	0 0 3 6 8
3	1 3 4 7
4	2 5 5
5	0
6	5
7	2 6
8	
9	7
10	
11	3

Key: 4|2 = $42

a. Calculate the mean and median.

b. Calculate the lower and upper quartiles and IQR.

c. Determine which, if any, values are outliers.

d. Write several sentences to describe this data set in context.

e. Name some factors that might account for the extreme values, and the much lower measure of center.

Common Core Math 1
Univariate Data
Boxplots and Outliers

Name _____
Date _____

Show all work on your own paper!

The 1984 Winter Olympics were held in Sarajevo, Yugoslavia. The table below lists the total number of gold, silver, and bronze medals won, by country.

Country	Total Medals
Austria	1
Canada	4
Czechoslovakia	6
Finland	13
France	3
Germany, East	24
Germany, West	4
Great Britain	1
Italy	2
Japan	1
Liechtenstein	2
Norway	9
Sweden	8
Switzerland	5
USSR	25
United States	8
Yugoslavia	1

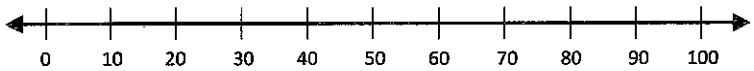
1. Find the five-number summary of the data. Show all work!
2. Use the 1.5IQR rule to find the outliers. Show all work!
3. Which countries are outliers? Why do you think these countries are outliers?
4. Construct a box plot of the data. Make a sketch on graph paper.
5. Which measures of the center and spread would be most appropriate to use to describe this data set? Explain.
6. Organize the data into a frequency table. Use interval width of 5.
7. Construct a histogram of the data. Make a sketch on graph paper.
8. Describe the data in context. Why do you think the data is skewed?

Common Core Math I Name_____
Unit 1: One-Variable Data
Comparing Data Sets Practice

Listed below are the quality rating values of natural peanut butters:
34, 40, 52, 57, 57, 60, 60, 63, 67, 69, 69, 69, 71, 89

The data values for regular peanut butter are as follows:
11, 23, 23, 26, 29, 31, 31, 33, 34, 34, 35, 40, 40, 43, 45, 46, 49, 54, 54, 60, 76, 83, 83

1) Construct side-by-side boxplots for the two types of peanut butters. Make a sketch below.



2) Record the five-number summary for each in the table below. Calculate the IQR for each set of data and record in the table.

Statistic	Natural Peanut Butter	Regular Peanut Butter
Min		
Q_1		
M		
Q_3		
Max		
IQR		

3) Calculate the mean and standard deviation for each set of data and record below.

Statistic	Natural Peanut Butter	Regular Peanut Butter
$\bar{x}$		
s		

4) Which measure of center and spread would be most appropriate to use to describe these two sets of data? Explain.

5) Compare the two data sets in context. Be sure to address shape, center, spread, and outliers. Which type of peanut butter is better?

Common Core Math 1
Unit 1: One-Variable Data
Comparing Data Sets

Name _____

Date _____

The miles per gallon for city travel of ten cars and ten SUVs are given below:

Cars	MPG	SUVs	MPG
Geo Metro	46	Jeep Grand Cherokee	14
Honda Civic CX	42	Ford Explorer	21
Hyundai Excel GS	29	Chevy Silverado	21
Mazda 323	29	Toyota Tacoma	22
Plymouth Sundance	26	Nissan Frontier	16
Saturn SL	28	Chevy Suburban	12
Eagle Summit	31	GMC Yukon	12
Nissan Sentra E	29	Ford F150	13
Ford Festiva GL	35	Jeep Wrangler	18

Construct histograms for each type of car, using the same window. Make a sketch below:

For each type of automobile, find the following information:

Statistic	Cars	SUVs
Min		
Q_1		
M		
Q_3		
Max		
Range		
IQR		
$\bar{x}$		
s		
Outliers		