PreMiEr Precision Microbiome Engineering
NSF Engineering Research Center

Food for Thought…Statistically Speaking

Title
Food for Thought…Statistically Speaking

Overview
NC Math 1 reviews the concepts of one-variable statistics. Through the theme of food, this lesson solidifies students' understanding of statistical parameters including measures of central tendencies, measures of variability, graphs, and SOCS (shape, outliers, center, and spread).

Key Search Words
NC Math 1, mean, median, mode, range, minimum, maximum, lower quartile, upper quartile, interquartile range, standard deviation, mean absolute deviation, dot plot, bar graph, histogram, box plot, shape, outlier, center, spread

Learning Objectives
- Students will be able to calculate statistical values of 1-variable data.
- Students will be able to create graphs from 1-variable statistics.
- Students will be able to describe data sets using parameters such as shape, outliers, center, and spread.

Curriculum Alignment
NC Math 1 Standards:
- Summarize, represent, and interpret data on a single count or measurement variable. Use technology to represent data with plots on the real number line (histograms and box plots). (NC.M1.S-ID.1)
- Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. Interpret differences in shape, center, and spread in the context of the data sets. (NC.M1.S-ID.2)
- Examine the effects of extreme data points (outliers) on shape, center, and/or spread. (NC.M1.S-ID.3)

Classroom time required
Warmup 1 (10 minutes) - optional
Guided Notes Handout 1 - “Statistics: Definitions and Graphs” (40 minutes)
Exit Ticket (10 minutes)
Warmup 1 (10 minutes) - optional
Guided Notes Handout 2 - “Shape, Outliers, Center, Spread” (45 minutes)
Exit Ticket (15 minutes)

Materials & Technology

- Pencils (for students to write on the handouts)
- Colored pencils (optional)
- Statistics Handouts (4 pages) each student
- Graphing calculator (preferably Desmos)
- Exit Ticket (see **Assessment/Check for Understanding**)

Teacher Preparation for Activity

- Link the Desmos App https://www.desmos.com/calculator to your online course.
- Print out class set of handouts.

Student Preparation for Activity

Prerequisites:

These are the first two lessons of the Statistics Unit in Math 1. Although exposure to these concepts are not required, students should have seen most of the concepts of mean, median, mode, range, dot plots, and bar graphs.

Suggested Warm-up #1:

The purpose of this warm up is to help students understand the concept of 5 number summary which is putting data into FOUR groups ("buckets").

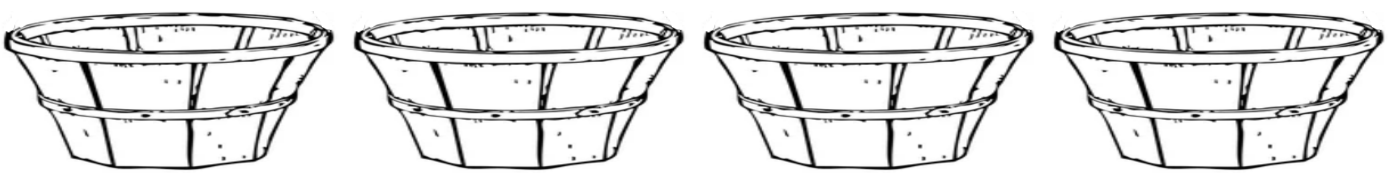
Below are a list of foods with the protein content (in grams) per serving size:

Almonds	6
Brown rice	3
Cheese	10
Chicken	29
Eggs (2)	13
Hershey's bar	3

Honey bun	2
Hummus	4
Nutrigrain bar	2
Slim Jim	6
Takis	2
Yogurt	20

Organize the foods into four groups so that:

- You write them by their protein value.
- They are placed in order from lowest protein amount to highest protein amount.
- They are placed in one of the four buckets below..
- Each bucket has the same number of food items.



Warmup #1 KEY

Below are a list of foods with the protein content (in grams) per serving size:

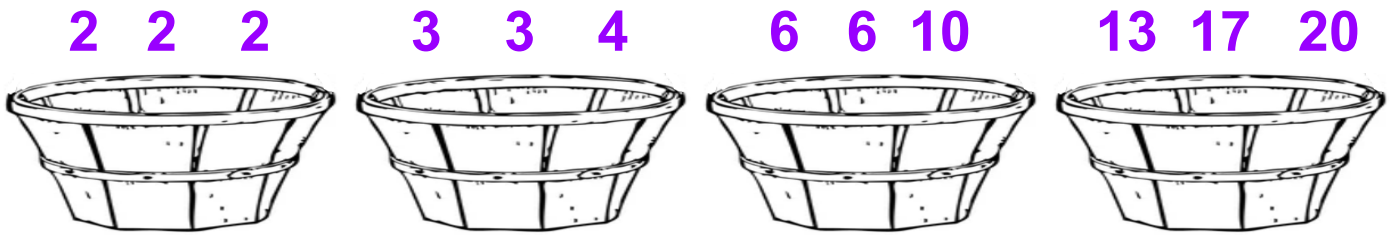
Almonds	6
Brown rice	3
Cheese	10
Eggs (2)	13
Hershey's bar	3
Honey bun	2

Hummus	4
Nutrigrain Bar	2
Salmon	17
Slim Jim	6
Takis	2
Greek Yogurt	20

Organize the foods into four groups so that:

- You write them by their protein value.
- They are placed in order from lowest protein amount to highest protein amount.
- They are placed in one of the four buckets below..
- Each bucket has the same number of food items.

Listed in ascending order: 2 2 2 3 3 4 6 6 10 13 17 20



Suggested Warm-up #2:

The purpose of this warm up is to review several 1-variable graphs.

1. Calories of fast food value meals.

450	**480**	**660**	**510**
850	**500**	**750**	**560**
270	**650**	**620**	**390**
360	**450**	**540**	**460**
650	**900**	**470**	**580**
300	**950**	**380**	**790**

Draw a histogram. (Make each bin 100 units)

2. Number of fiber (grams) per serving:

Banana	**3.5**	Choc Chip Cookie	**1.0**
Broccoli	**6.3**	Frosted Flakes	**1.0**
Carrot	**3.6**	Orange	**3.1**
Chickpeas	**6.3**	Popcorn	**1.9**
Cheerios	**3.0**	Potato Chips	**1.4**

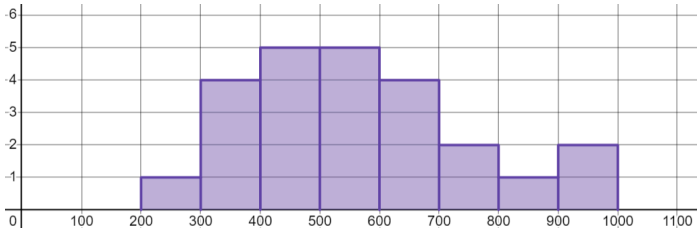
Draw a boxplot using the following data.

Warmup #2 **KEY**

1. Calories of fast food value meals.

450	**480**	**660**	**510**
850	**500**	**750**	**560**
270	**650**	**620**	**390**
360	**450**	**540**	**460**
650	**900**	**470**	**580**
300	**950**	**380**	**790**

Draw a histogram. (Make each bin 100 units)

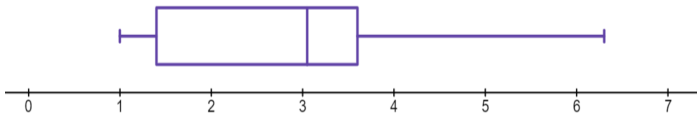


Calories in a Value Meal

2. Number of fiber (grams) per serving:

Banana	**3.5**	Choc Chip Cookie	**1.0**
Broccoli	**6.3**	Frosted Flakes	**1.0**
Carrot	**3.6**	Orange	**3.1**
Chickpeas	**6.3**	Popcorn	**1.9**
Cheerios	**3.0**	Potato Chips	**1.4**

Draw a boxplot using the following data.



Grams of Fiber

Procedure

1. Warmup #1 (suggested)
   Have students complete the answers to the warmup at the beginning.You can have cards taped to the board so students can physically put them in ascending order. After students understand the ordering and placing of the numbers. A discussion can revolve around 4 sections but 5 posts (i.e. 5 number summary). Further discussion of the 25% rule for each section can also take place.
2. Handout #1
   a. These handouts are guided notes for the Statistics Unit.
   b. As a class, have students complete a KWL chart about what they remember about statistics. (Know/Want-to-Know/ (TO BE) Learned)
   c. NOW, pass out the handout (set 1) for the students to read the leading paragraph. Discuss that statistics can be used in many areas. This unit will use the theme of food and nutrition.
   d. Guide the students through both sides of the page. This will take most of the class time. If your students seem to remember the concepts from previous years, you can flip the order on how you complete the handout - they work on the examples and then the class checks the work together.
   e. For the graphs, emphasize the TALKS components (Title, Axis, Label, Key, Scale). Not all components will be used for each graph, but it is a helpful acronym.
3. When finished, it is highly recommended to go through the questions again using a graphing calculator. Since Desmos is embedded in the NC Math 1 End-of-Course test, I recommend using the Desmos App https://www.desmos.com/calculator
4. For students who finish early, they can use the TI84 calculator to check their work. (See Differential section.)
5. Warmup #2 (suggested)
   Have students complete the answers to the warmup at the beginning. Colored pencils are useful. Emphasize that they should include the TALKS components (Title, Axis, Label, Key, Scale).
6. Handout #2
   a. These handouts are guided notes for the Statistics Unit.

b. Pass out the handout (set 2).
   Discuss that data can be described by different parameters. A useful acronym is SOCS (Shape, Outlier(s), Center, and Spread).
c. Guide the students through each of the 4 components. On both sides of the page. Emphasize vocabulary. It is recommended that some type of Anchor chart be used with these concepts.
d. For the analysis of the three sets of data at the end, have students describe SOCS. It is advised that you do one as a class and then let them work on the other two on their own before going over them. A graphing calculator tool will be useful for this section. Since Desmos is embedded in the NC Math 1 End-of-Course test, I recommend using the Desmos App https://www.desmos.com/calculator
e. If there is time, have students work on a class data to be analyzed. Some possible questions could be:
   i. Number of meals you ate from a restaurant this past week.
   ii. Number of fruits/vegetables you ate this week.
   iii. Number of packaged processed foods you are this week.

7. For students who finish early, they can use the TI84 calculator to check their work. (See Differential section.)

Differentiation

- For students who finish early, they can use the TI84 calculator to check their work.
  - See 1-Variable statistics:
    - 'Stat' → 'Edit' → enter data in 'L1'
    - "Quit' to HOME screen.
    - 'Stat' → 'CALC' → '1-Var Stats' → 'Calculate'
  - Display graphs
    - '2nd' 'Stat Plot' → 'Plot1' → 'On'
    - Select Type
    - 'Zoom' → 'ZoomStat' → 'Graph'
- Students work on a class data to be analyzed. Some possible questions could be:
  - Number of meals you ate from a restaurant this past week.
  - Number of fruits/vegetables you ate this week.
  - Number of packaged processed foods you are this week.

Assessment/Check for Understanding

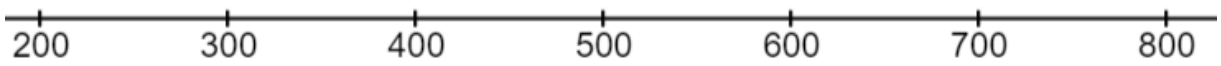
Using provided warmups, guided notes, and exit tickets.

Suggested Exit Ticket #1

Exit Ticket

Box Plot of Sodium (salt) in Foods

Canned Soup	700 mg	Bagel	400 mg	Pork Rinds	515 mg
Frozen Pizza	765 mg	American Cheese	230 mg	Biscuit	528 mg
Hot Dog	578 mg	Sausage patty	415 mg	Tortilla Chips	539 mg



What percent of numbers are between the Lower Quartile and the Upper Quartile? _____

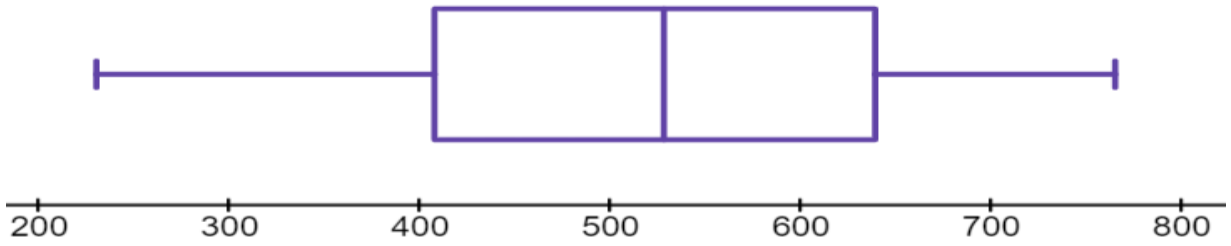
Suggested Exit Ticket #1 **KEY**

Exit Ticket

Box Plot of Sodium (salt) in Foods

Canned Soup	700 mg	Bagel	400 mg	Pork Rinds	515 mg
Frozen Pizza	765 mg	American Cheese	230 mg	Biscuit	528 mg
Hot Dog	578 mg	Sausage patty	415 mg	Tortilla Chips	539 mg

In order: **230** **400** **415** **515** **528** **539** **578** **700** **765**



What percent of numbers are between the Lower Quartile and the Upper Quartile? **50%**

Suggest Exit Ticket #2

Exit Ticket:
Use collected data from the class:

Shape

Outlier(s)

Center

Spread

Suggest Exit Ticket #2 **KEY**

Answers will vary based on the data used

Required resources
none

Supplemental resources
- Article with Recommended Daily Allowance for Teenagers: Nutrition in adolescents: physiology, metabolism, and nutrition needs, Jai Das et al, Annals of the New York Academy of Sciences, April 2017.
- Sodium Content https://www.healthline.com/nutrition/foods-high-in-sodium
- Fast Food and Obesity https://www.statcrunch.com/reports/view?reportid=53116

Author comments
You can create a Google survey for students to complete and use the class data to be used for analysis. To keep it along the same theme with Food Nutrition, here are some recommended questions you can add to the survey:
- Number of meals you ate from a restaurant this past week.
- Number of fruits/vegetables you ate this week.
- Number of packaged processed foods you are this week.
- Number of sugary desserts this week.
- Number of meals you prepared this week.
- Number of lunches you brought from home this week.
- Calories eaten at breakfast this morning.

Sources
David Lab website http://el.ladlab.org/
Inspirational article explaining FoodSeq www.pnas.org/doi/10.1073/pnas.2304441120
Amy Gross (amy.gross@orange.k12.nc.us)

Appendices
Student handouts and keys.

Math 1 Statistics: Definitions and Graphs Name _____

Everyday, you make choices about the foods you eat. You decide where and what to buy, how it's prepared and how much you eat. Foods are highly regulated and studied. These studies require a field in mathematics called statistics. **Statistics** is the collection and analysis of data which can be used to predict future events. To understand statistics, we first need to learn some terms often used.

Let's start off with the **Measures of Central Tendency** (in the center)**:**

	Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**
Mean (average, 'balance point') $\overline{X}$ *add all number and divide by number of data*	Mean:
Median ('middle' value) *order numbers from least to greatest; then find the middle value*	Median:
Mode (most frequent) *number(s) that occurs the most frequent*	Mode:

If we want to organize our data into four sections, we can use the **5 Number Summary**

Minimum (min)	**Lower Quartile** (1st quartile Q1)	**Median** (med)	**Upper Quartile** (1st quartile Q3)	**Maximum** (max)
smallest value	middle of bottom half	middle value	middle of upper half	largest value

Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**

We can also determine how much the data spreads away from the center. These are the **Measures of Variability**:

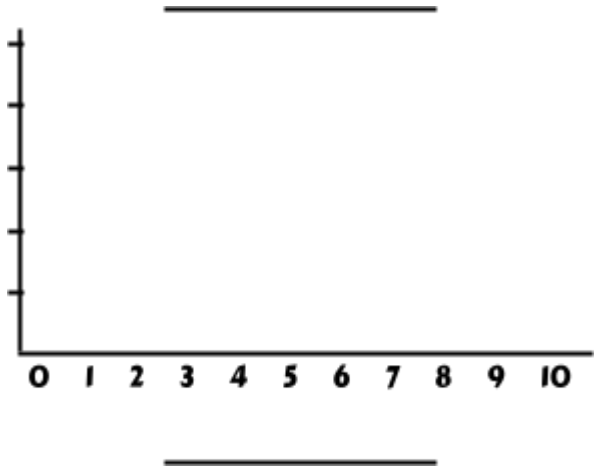
	Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**
Range *maximum - minimum*	
Interquartile Range *upper quartile - lower quartile*	
Mean Absolute Deviation (M.A.D.) *use a calculator*	
Standard Deviation (S.D.) σ *use a calculator*	

Diets are a contributing factor to people's health. What if doctors want to know what specific foods patients have recently eaten but there are no written records of their meals? Scientists are working on analyzing human stool to detect food genetic material (called DNA and RNA). It is helpful to visualize the data. Here are a few examples of **graphs** we could use:

Dot Plot

A scientist records the number of plant taxa (species) that showed up in 14 tested samples.

2	2	4	2	3	8	3
4	5	2	1	7	3	5

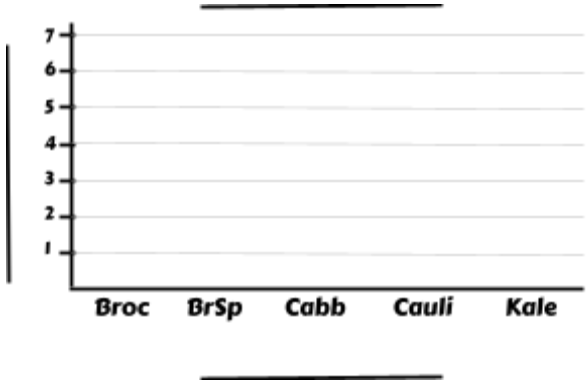


Key: **X = 1 sample**

Bar Graph

Species of plants and animals are classified into different taxonomy genuses based on shared characteristics and genetic sequences. One common group of plants detected in healthy diets are in the *Brassica* genus. Some common names of these species are: cabbage, broccoli, cauliflower, kale, brussels sprouts, and collard greens. During a study, scientists matched given samples to the following species:

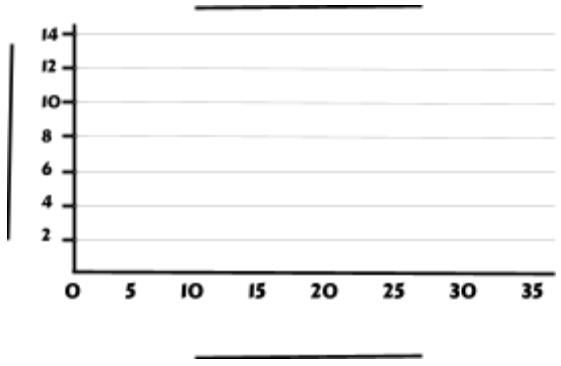
Broc	Kale	Broc	Cauli	Broc
Kale	Broc	Cauli	BrSp	BrSp
Broc	Kale	Cauli	Cabb	Broc



Histogram

Over one week, 30 high school students wrote down the foods they ate. They then counted the number of non-processed fruits/vegetables.

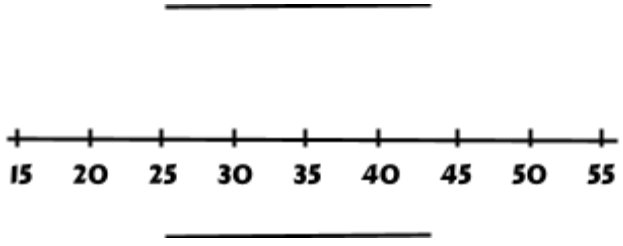
5	13	2	4	19	6	8	10	34	6
4	11	3	9	11	7	2	19	16	2
8	10	8	7	20	9	6	22	11	7



Box Plot

Often, samples are tested from different groups collected at multiple sites. A research lab tested batches of samples from 14 different sites. The number of participants in each group are:

19	21	24	27	27	28	28
32	33	35	36	39	43	51

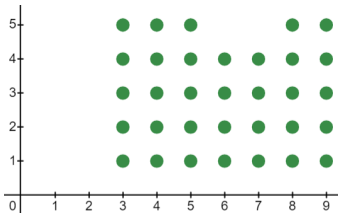
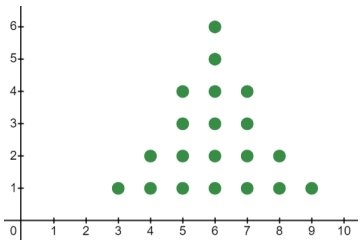
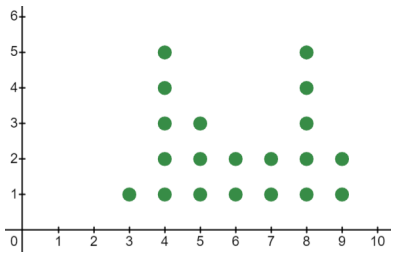
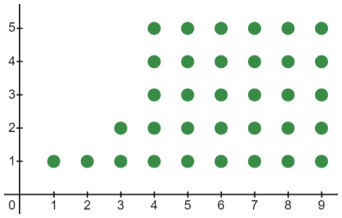
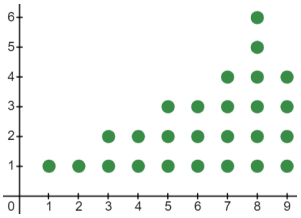
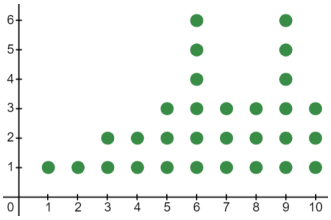
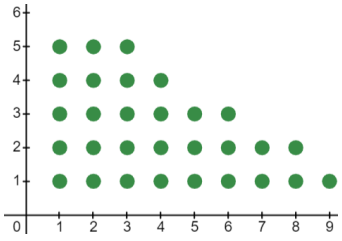
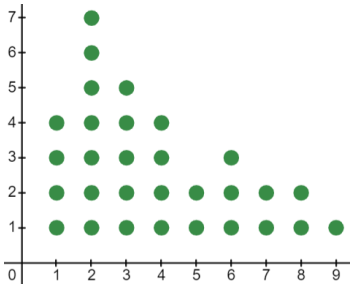
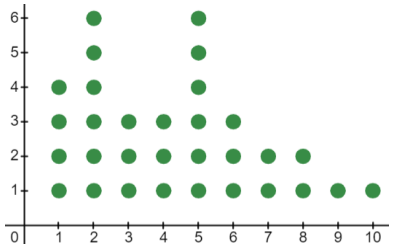


Notice that most graphs/plots have **'TALKS'** components: **t**itle, **a**xis, **l**abel, **k**ey, and **s**cale

Math 1 Shape, Outliers, Center, Spread ( ) Name _____

Shape

Once data is run through statistics and organized in a graph, we can start to see the parameters. The first component is the **shape**. We will use the dot plots below to help us describe the shape of a distribution.

SHAPES	No peaks	_____ (one-peak)	_____ (two-peaks)
_____		 Special name: _____	
_____ (tail to left)			
_____ (tail to right)			

Outliers

Sometimes, there are data values that are much smaller or much larger than the other values. These are called **outliers**. There is a formula that is used to determine if there are any outliers: **1.5 x IQR** which is subtracted from the lower quartile and added to the upper quartile. This 1.5 x IQR is a “fence” that is used to determine which data points are in the acceptable range and which data points are outliers because they are “outside the fence”.

Example: **47 49 52 55 59 61 81** IQR = 1.5 x IQR = Outliers? _____	Example: **4 24 26 26 28 30 42 55** IQR = 1.5 x IQR = Outliers? _____

Center

The **center** is used when describing the mean, median, and mode (as discussed yesterday).

If data is symmetrical and there are NO outliers, then we represent the center using the…	_____
If data is skewed or if we have outliers, then we represent the center using the…	_____

Spread

When data is "**spread** out" it is described as variation. These are the distances away from the center (standard deviation and M.A.D) and from other data (range and interquartile range).

If data is symmetrical and there are NO outliers, then we represent the variation with…	_____
If data is skewed or if we have outliers, then we represent the variation with…	_____

Statistically describe each set of data using Shape, Outlier(s), Center, and Spread (**SOCS**)

A. Twenty Olympic athletes tracked the number of fruits/veggies for a week. 25 13 22 34 19 25 40 49 34 21 25 28 25 22 19 22 31 28 27 28	B. Popular teenage drinks' sugar content (in grams per serving). 39 23 20 31 23 39 38 29 55 35 23 66 45 40 63 45 39 68	C. Ultra-processed foods (UPF) are packaged foods that are high in preservatives, fillers, sugars and fats. Gamers were asked how many UPF they ate over the weekend. 4 10 14 16 2 15 14 15 8 14 10 14
Shape	Shape	Shape
Outlier(s)	Outlier(s)	Outlier(s)
Center	Center	Center
Spread	Spread	Spread

Math 1 Statistics: Definitions and Graphs **KEY**

Everyday, you make choices about the foods you eat. You decide where and what to buy, how it's prepared and how much you eat. Foods are highly regulated and studied. These studies require a field in mathematics called statistics. **Statistics** is the collection and analysis of data which can be used to predict future events. To understand statistics, we first need to learn some terms often used.

Let's start off with the **Measures of Central Tendency** (in the center)**:**

	Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**
Mean (average, 'balance point') $\overline{X}$ *add all number and divide by number of data*	Mean: **sum of all numbers = 71** **71 ÷ 9 = 7.89**
Median ('middle' value) *order numbers from least to greatest; then find the middle value*	Median: **2 4 5 8 8 10 11 11 12** ↑ **8**
Mode (most frequent) *number(s) that occurs the most frequent*	Mode: **8 and 11**

If we want to organize our data into four sections, we can use the **5 Number Summary**

Minimum (min)	**Lower Quartile** (1st quartile Q1)	**Median** (med)	**Upper Quartile** (1st quartile Q3)	**Maximum** (max)
smallest value	middle of bottom half	middle value	middle of upper half	largest value

Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**

2 4 5 8 8 10 11 11 12

MIN = **2** Q1 = **4.5** MED = **8** Q3 = **11** MAX = **12**

We can also determine how much the data spreads away from the center. These are the **Measures of Variability**:

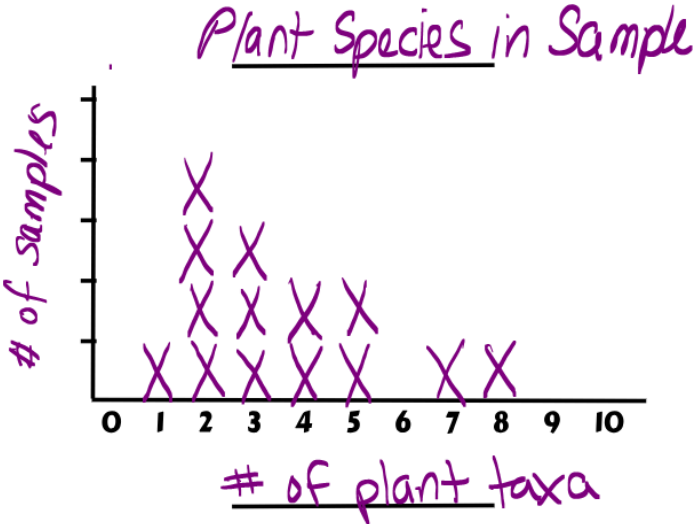
	Number of fruits/vegetables consumed daily: **12 5 8 11 10 11 2 8 4**
Range *maximum – minimum*	**12 – 8 = 4**
Interquartile Range *upper quartile – lower quartile*	**11 – 4.5 = 6.5**
Mean Absolute Deviation (M.A.D.) *use a calculator*	(calculator) **2.8**
Standard Deviation (S.D.) σ *use a calculator*	(calculator) σ = **3.3**

Diets are a contributing factor to people's health. What if doctors want to know what specific foods patients have recently eaten but there are no written records of their meals? Scientists are working on analyzing human stool to detect food genetic material (called DNA and RNA). It is helpful to visualize the data. Here are a few examples of **graphs** we could use:

Dot Plot

A scientist records the number of plant taxa (species) that showed up in 14 tested samples.

2	2	4	2	3	8	3
4	5	2	1	7	3	5

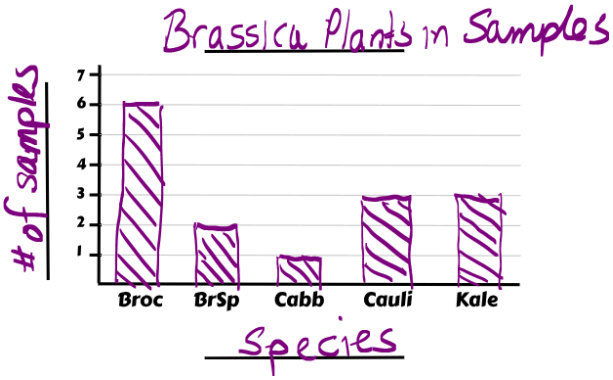


Key: **X = 1 sample**

Bar Graph

Species of plants and animals are classified into different taxonomy genuses based on shared characteristics and genetic sequences. One common group of plants detected in healthy diets are in the *Brassica* genus. Some common names of these species are: cabbage, broccoli, cauliflower, kale, brussels sprouts, and collard greens. During a study, scientists matched given samples to the following species:

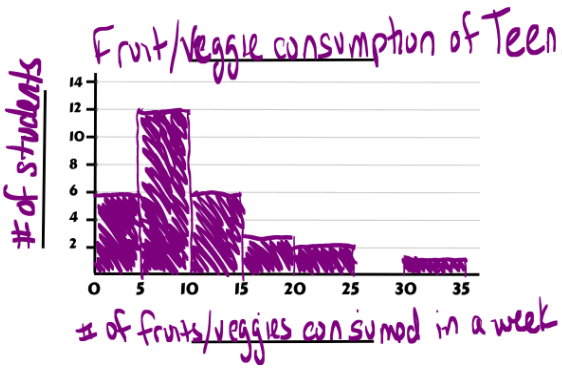
Broc	Kale	Broc	Cauli	Broc
Kale	Broc	Cauli	BrSp	BrSp
Broc	Kale	Cauli	Cabb	Broc



Histogram

Over one week, 30 high school students wrote down the foods they ate. They then counted the number of non-processed fruits/vegetables.

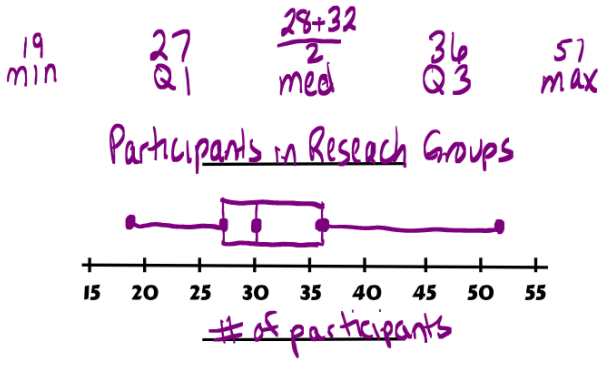
5	13	2	4	19	6	8	10	34	6
4	11	3	9	11	7	2	19	16	2
8	10	8	7	20	9	6	22	11	7



Box Plot

Often, samples are tested from different groups collected at multiple sites. A research lab tested batches of samples from 14 different sites. The number of participants in each group are:

19	21	24	27	27	28	28
32	33	35	36	39	43	51

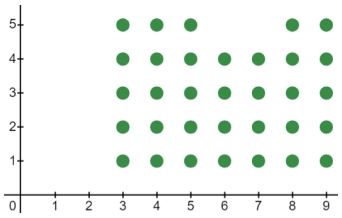
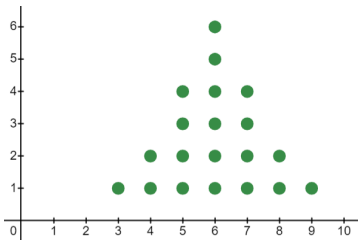
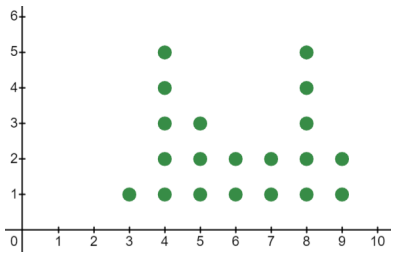
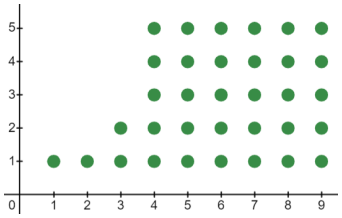
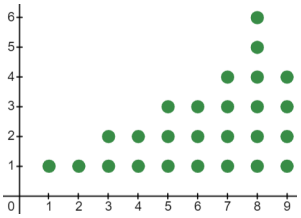
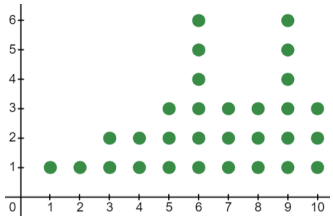
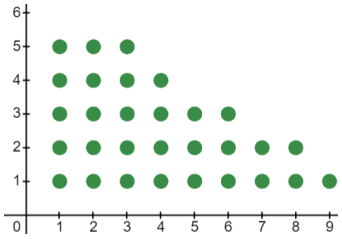
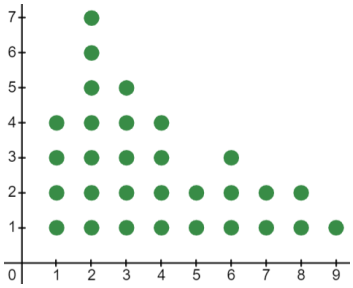
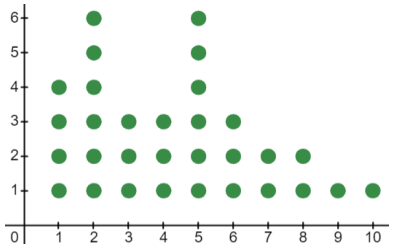


Notice that most graphs/plots have **'TALKS'** components: **t**itle, **a**xis, **l**abel, **k**ey, and **s**cale

Math 1 Shape, Outliers, Center, Spread (**SOCS**) **KEY**

Shape

Once data is run through statistics and organized in a graph, we can start to see the parameters. The first component is the **shape**. We will use the dot plots below to help us describe the shape of a distribution.

SHAPES	No peaks	**Unimodal** (one-peak)	**Bimodal** (two-peaks)
Symmetric		 Special name: **Normal**	
Left-skewed (tail to left)			
Right-skewed (tail to right)			

Outliers

Sometimes, there are data values that are much smaller or much larger than the other values. These are called **outliers**. There is a formula that is used to determine if there are any outliers: **1.5 x IQR** which is subtracted from the lower quartile and added to the upper quartile. This 1.5 x IQR is a "fence" that is used to determine which data points are in the acceptable range and which data points are outliers because they are "outside the fence".

Example: **47 49 52 55 59 61 81**
Q1 **Q3**

IQR = **61 – 49 = 12** 1.5 x IQR = **18**
Fences **49 – 18 = 31** 61 + 18 = **79**
Outliers? **81**

Example: **4 24 26 26 28 30 42 55**
Q1 = 25 **Q3 = 36**

IQR = **36 – 25 = 11** 1.5 x IQR = **16.5**
Fences 25 – 16.5 =8.5 **36 + 16.5 = 52.5**
Outliers? **4 and 55**

Center

The **center** is used when describing the mean, median, and mode (as discussed yesterday).

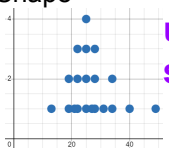
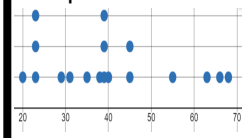
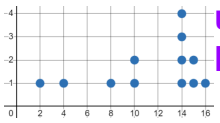
If data is <u>symmetrical</u> and there are NO outliers, then we represent the center using the…	**mean**
If data is <u>skewed</u> or if we have outliers, then we represent the center using the…	**median**

Spread

When data is "**spread** out" it is described as variation. These are the distances away from the center (standard deviation and M.A.D) and from other data (range and interquartile range).

If data is <u>symmetrical</u> and there are NO outliers, then we represent the variation with…	**standard deviation**
If data is <u>skewed</u> or if we have outliers, then we represent the variation with…	**Interquartile range**

Statistically describe each set of data using Shape, Outlier(s), Center, and Spread (**SOCS**)

A. Twenty Olympic athletes tracked the number of fruits/veggies for a week. 25 13 22 34 19 25 40 49 34 21 25 28 25 22 19 22 31 28 27 28	B. Popular teenage drinks' sugar content (in grams per serving). 39 23 20 31 23 39 38 29 55 35 23 66 45 40 63 45 39 68	C. Ultra-processed foods (UPF) are packaged foods that are high in preservatives, fillers, sugars and fats. Gamers were asked how many UPF they ate over the weekend. 4 10 14 16 2 15 14 15 8 14 10 14
Shape  **unimodal** **symmetrical**	Shape  **bimodal** **right-skewed**	Shape  **unimodal** **left-skewed**
Outlier(s) IQR: 29.5 - 22 = 6.5 Fences: 10.75 and 40.75 **<u>Yes: 49</u>**	Outlier(s) IQR: 45 - 29 = 16 Fences: 5 and 69 **<u>None</u>**	Outlier(s) IQR: 14.5 - 9 = 5.5 Fences: .75 and 15.25 **<u>None</u>**
Center mean = 26.85 **<u>median = 25</u>** mode = 25	Center **<u>mean = 40.1</u>** median = 39 mode = 23, 39	Center **<u>mean = 11.3</u>** median = 14 mode = 14
Spread s.d. = 7.8 M.A.D.= 5.7 range 49-13= 36 **<u>IQR= 6.5</u>**	Spread **<u>s.d. = 14.5</u>** M.A.D.= 11.3 range 68-20= 48 IQR= 16	Spread **<u>s.d. = 4.4</u>** M.A.D.= 3.8 range 16 - 2= 14 IQR= 5.5