Name_____

AP Statistics CED 1.1 Daily Video 1 (Skill 1.A)

Introducing Statistics: What Can We Learn from Data?

What Will We Learn?

How do we identify the question to be answered or the problem to be solved in a given context?

How can Statistics be used to help answer important, real-world questions based on data that vary?

The Flint Water Crisis

- **Where?** _____ (population about 100,000)
- **When?** _____
- **What?** The city switched its water supply from _____ to the _____.
- **Why?** _____
- **What happened?**
  - Residents said the water _____, _____, and _____ bad.
  - Some residents developed _____, _____ or _____.
  - City officials _____ the water was _____ to drink.

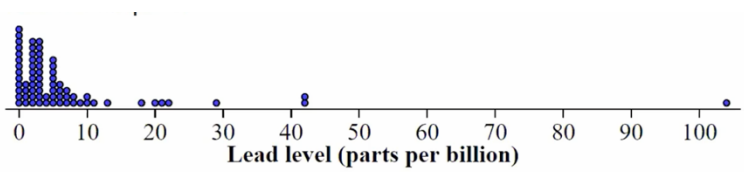
Was Flint's water safe to drink?

Collected Data: City officials measured lead levels in 71 water samples from Flint residents in January to June 2015. Here are the data (in part per billion). *Note:* Lead levels greater than 15 parts per billion are considered unhealthy.

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3
3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11 13
18 20 21 22 29 42 42 104

Analyze Data: Here is a graph of the lead levels from the 71 waters samples.

I more than 10% of water samples have lead levels greater than 15 parts per billion, the water is not safe to drink.



What percent of lead levels in these 71 water samples exceed 15 parts per billion? (Draw a back around all samples greater that 15 and then count them as done on the video.)

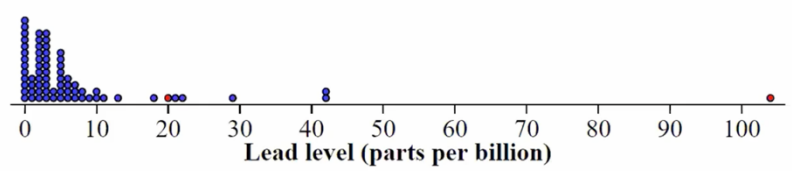
We find: _____ or _____ of the samples had lead levels > 15 ppb

Interpret Results: Is Flint's tap water safe to drink? _____, because _____ > 10%.

Name_____

Was Flint's Water Safe to Drink?

City officials omitted 2 water samples from the analysis marked in red: 20 parts per billion (came from a business) and 104 parts per billion (came from a home that used a filter).



Based on the lead levels of the 69 water samples, why did the city declare the water safe to drink? Omitting the red dots, only _____ or _____ of the remaining water samples have lead levels _____ than 15 parts per billion, and _____.

What happened next?

- Virginia Tech researchers conducted a thorough study of lead levels in Flint's drinking water. And found about _____ of samples had lead levels above 15 parts per billion.
- Flint pediatrician Mona Hanna-Attisha found elevated blood levels in children had _____ since 2014.
- June 2014 – October 2015: Legionnaire's disease killed _____ and sickened _____ Flint residents.
- 2016: _____ sued government officials to demand safe drinking water.
- Courts ruled for the _____, requiring _____ water delivery and replacement of the city's _____ pipes.
- Several government officials were _____.
- Lead levels in Flint have remained _____ action levels since July, 2016.

What Should We Take Away?

How do we identify the question to be answered or the problem to be solved in a given context?

How can Statistics be used to help answer important, real-world questions based on data that vary?

_____, _____, _____

Name_____

AP Statistics CED 1.2 Daily Video 1 (Skill 2.A)

The Language of Variation - Variables

What Will We Learn?

How do we identify the individuals and variables in a data set?

What are the two main types of variables, and how do we distinguish them?

The structure of a data set

The individuals in a data set can be _____, _____, or _____.

	ID	Type	Price	Year built	No. of bedrooms	Pool?	Distance to beach (mi)	Parking	Zip code
Individuals	001	House	649,995	1964	3	Y	0.23	Carport	29577
	002	Condo	422,750	2008	2	N	1.78	Garage	29575
	003	Townhome	399,900	1999	1	Y	4.15	Outdoor	29588
	004	Condo	550,000	2014	3	Y	6.69	Outdoor	29579
	005	House	822,000	2019	4	Y	0.88	Garage	29572
	006	House	499,900	2001	3	N	2.66	Carport	29588

In this data set, the individuals are shown in the _____ of the spreadsheet. Individual row in this data set represent: _____ for sale in Charleston, SC.

Identifying variables

A variable is a _____ that changes from _____ individual to another.

The variables in this data set are:

_____

_____

_____

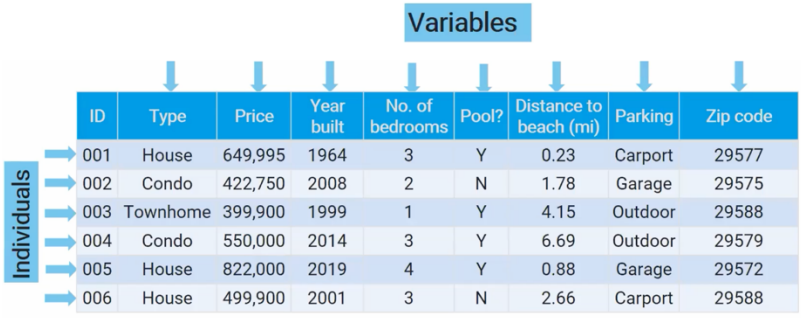
_____

_____

_____

_____

_____



Two Types of Variables

A _____ variable takes on values that are category names or group labels.

_____ _____

_____ _____

A _____ variable takes on the numerical values for a measured or counted quantity.

_____ _____

_____ _____

**Note*: We can tell a quantitative variable because it makes sense to find an _____ of those values.

Name_____

Classifying variables

- Categorical data = Values of a _____ in a data set.
- Quantitative data = Values of a _____ in a data set.
- Not all variables that take _____ values are quantitative! (e.g., zip code)
- It is possible to make a _____ variable _____ by grouping values. e.g.,
  Distance to beach = _____ (<1 mile), _____ (1 -<3 miles) , _____ (+3 miles)

Let's Practice – Identifying and Classifying Variables

An AP Statistics teacher collected data from all 30 students in class with the following survey.

Grade level: (Circle one) 9 10 11 12	Age: _____ (e.g., 16.39 years)
Favorite season: (Circle one) Fall Spring Summer Winter	Birth month: _____
Reaction time in an online test: _____ milliseconds	Height: _____ centimeters
Can you roll your tongue? _____	Number of people who live in your household: _____

Which of the following is <u>not</u> a variable in this data set?
(a) Favorite season
(b) Number of people living in household
(c) Number of students in this AP Statistics class
(d) Whether or not you can roll your tongue

Classify each variable as categorical or quantitative

Age: _____
Birth Month: _____
Grade Level: _____
Number of people living in household: _____

What Should We Take Away?

How do we identify the individuals and variables in a data set?
Individuals: _____ described by a set of data.
Variables: Characteristic that _____ from one individual to another.

What are the two main types of variables, and how do we distinguish them?
_____: Takes values that are category names of labels.
_____: Takes numerical values for a measured or counted quantity.

Name_____

AP Statistics CED 1.3 Daily Video 3 (Skill 2.A)

Representing a Categorical Variable with Tables

What Will We Learn?

How can we represent categorical data in tabular form?

How do these tabular representations help us describe categorical data?

Representing categorical data

An online survey asked: "Which of the following superpowers would you most like to have?

Invisibility Telepathy (read minds) Freeze time Super strength Fly

Here are data from a random sample of 50 high school students who completed the survey.

Freeze time	Invisibility	Telepathy	Super strength	Freeze time	Invisibility	Fly	Freeze time	Telepathy	Freeze time
Telepathy	Invisibility	Freeze time	Freeze time	Telepathy	Telepathy	Fly	Telepathy	Telepathy	Super strength
Invisibility	Telepathy	Super strength	Telepathy	Freeze time	Fly	Telepathy	Fly	Freeze time	Freeze time
Freeze time	Fly	Invisibility	Fly	Fly	Invisibility	Fly	Telepathy	Freeze time	Telepathy
Telepathy	Invisibility	Freeze time	Telepathy	Invisibility	Fly	Freeze time	Freeze time	Telepathy	Freeze time

Individuals: _____

Variable: _____ Type of variable: _____

How can we represent the distribution of this categorical variable??

Representing categorical data with frequencies

A _____ gives the number of individuals (cases) in each category. Count and fill in the numbers for the remaining categories

Frequency table

Superpower preference	Frequency
Fly	9
Freeze time	
Invisibility	
Super Strength	
Telepathy	

A _____ gives the proportion or percent of individuals (cases) in each category

Relative Frequency table

Superpower preference	Relative Frequency
Fly	
Freeze time	
Invisibility	
Super Strength	
Telepathy	

Describing categorical data (Use the above tables to answer.)

Which of the following statements about this data set is true?

(a) A majority of the students chose telepathy as their preferred superpower.
(b) Almost 3 times as many students chose super strength as fly for the superpower of preference
(c) Nearly half of the students picked either fly or invisibility as their preferred superpower.
(d) Exactly 50% of student chose either fly or telepathy as their preferred superpower.
(e) Invisibility is one of the more popular choices of preferred superpower.

Name_____

Let's Practice – Describing categorical data from a frequency table

The annual Monitoring the Future study surveys random sample of U.S. 8th, 10th, and 12th grade students. One question on the 2018 survey asked:

How much do you think people risk harming themselves (physically or in other ways), if they vape an e-liquid with nicotine occasionally?

Response	Frequency
No risk	501
Slight risk	782
Moderate risk	401
Great risk	377
Can't say, drug unfamiliar	191

*Total: _____

The frequency table summarizes students' responses.

Which of the following statements is <u>not</u> supported by the table?

(a) Over 1/3 of students responded "Slight risk". _____

(b) More than twice as many students responded "Slight risk" as "Great risk". _____

(c) A majority of students said that there was "No risk" or "Slight risk". _____

(d) Over 10% of students responded "Can't say, drug unfamiliar". _____

(e) The proportion of students who responded "No risk" is about 0.22. _____

*Hint: You will need to convert the data to relative frequency (percentages!)

What Should We Take Away?

How can we represent categorical data in tabular form?

With a _____ or _____.

How do these tabular representations help us describe categorical data?

Counts and relative frequencies (_____ or _____) of _____ data reveal information that can be used to _____ claims about the data in context.

Name_____

AP Statistics CED 1.4 Daily Video 1 (Skill 2.B)

Representing a Categorical Variable with Graphs

What Will We Learn?

How can we represent categorical data graphically?

How do these graphical representations help us describe categorical data?

Representing categorical data

An online survey asked: "Which of the following superpowers would you most like to have?

Invisibility Telepathy (read minds) Freeze time Super strength Fly

Here are data from a random sample of 50 high school students who completed the survey.

Freeze time	Invisibility	Telepathy	Super strength	Freeze time	Invisibility	Fly	Freeze time	Telepathy	Freeze time
Telepathy	Invisibility	Freeze time	Freeze time	Telepathy	Telepathy	Fly	Telepathy	Telepathy	Super strength
Invisibility	Telepathy	Super strength	Telepathy	Freeze time	Fly	Telepathy	Fly	Freeze time	Freeze time
Freeze time	Fly	Invisibility	Fly	Fly	Invisibility	Fly	Telepathy	Freeze time	Telepathy
Telepathy	Invisibility	Freeze time	Telepathy	Invisibility	Fly	Freeze time	Freeze time	Telepathy	Freeze time

Individuals: _____

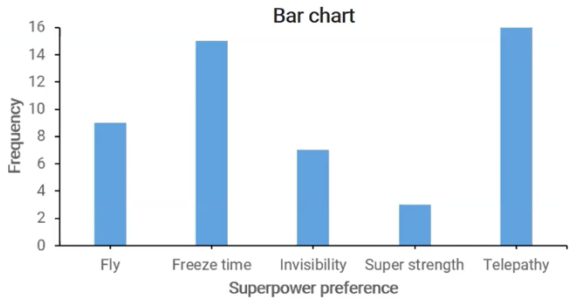
Variable: _____ Type of variable: _____

How can we represent the distribution of this categorical variable??

Representing categorical data with frequencies

Frequency table

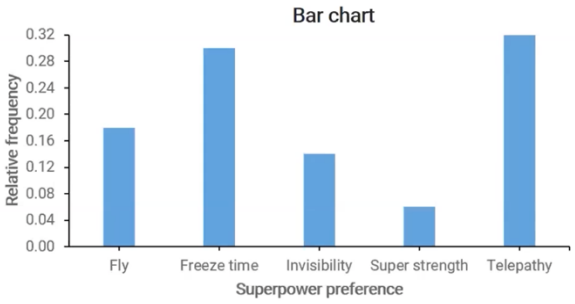
Superpower preference	Frequency
Fly	9
Freeze time	15
Invisibility	7
Super Strength	3
Telepathy	16



Representing categorical data with relative frequencies

Relative Frequency table

Superpower preference	Relative Frequency
Fly	9/50 = 0.18 = 18%
Freeze time	15/50 = 0.30 = 30%
Invisibility	7/50 = 0.14 = 14%
Super Strength	3/50 = 0.06 = 6%
Telepathy	16/50 = 0.32 = 32%



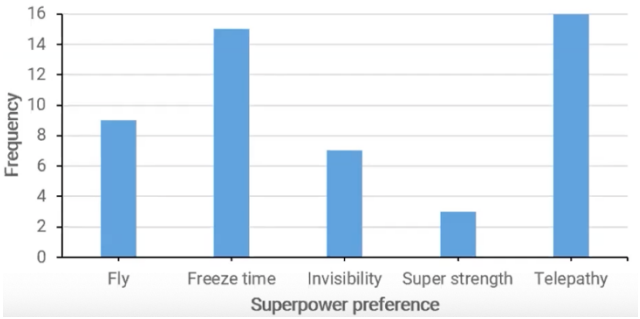
Making bar charts for categorical data

- **Label axes:** _____ name on the horizontal axis; _____/_____ on the vertical axis.
- **Scale axes:** _____ labels spread out along _____ axis; Start _____ vertical axis at ____ and go up in _____ increments until you equal or exceed maximum _____/_____.
- **Draw bars:** Make the bars _____ in width and leave _____ between them. The _____ of the bars represent the _____ frequencies or relative frequencies.

Name_____

Describing categorical data

A bar chart of the data on superpower preference for a random sample of 50 high school students is shown. Which of the following statements is not supported by the graph?



(a) Super strength was chosen by less that 1/5 of the students. _____

(b) Freeze time was chosen by twice as many students as Fly. _____

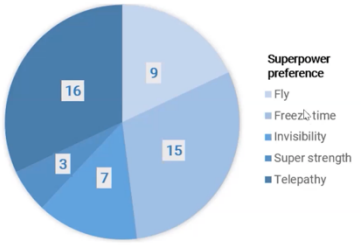
(c) Together, Freeze time and Telepathy were chosen by over 60% of the students. _____

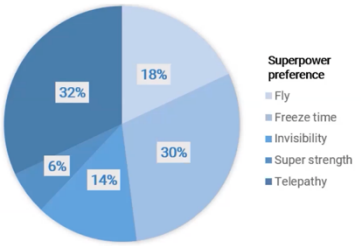
(d) The proportion of students who chose invisibility is less than half of the proportion of students who chose Telepathy. _____

(e) The most popular choice of preferred superpower was Telepathy. _____

Displaying categorical data: Pie Charts

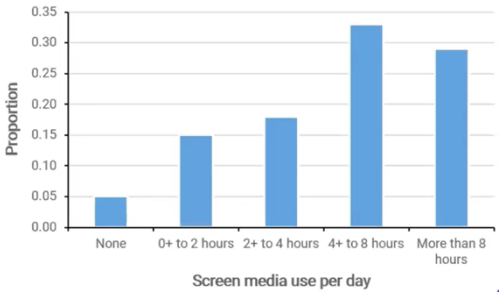
* Include a _____/_____ that connects categories to pie pieces.





Let's Practice: Describing categorical data from a bar chart

In 2019, Common Sense Media surveyed a random sample of more that 1600 U. S. 8- to 18-year-olds. The bar chart summarizes data on average daily screen media use by the teenagers (ages 13 to 18) in the sample. *Note:* 0+ to 2 hours indicates an average more than 0 hours up to and including 2 hours per day.



Which of the following statements is not supported by the graph?

(a) Over ¼ of teens reported screen media use of more that 8 hours per day. _____

(b) The proportion of teens reporting some screen media use is about 0.95. _____

(c) About 5 times as many teens reported more than 2 hours of screen media use per day as reported at most 2 hours of screen media use per day. _____

(d) A majority of teens reported over 4 hours of screen media use per day. _____

(e) About 50% of teens reported 2+ to 8 hours of screen media use per day. _____

What Should We Take Away?

How can we represent categorical data graphically?

With a _____ that displays category _____ or _____, or a _____,

How do these graphical representations help us describe categorical data?

Graphical representations of a _____ variable reveal information that can be used to _____ about the data in context.

Name_____

AP Statistics CED 1.4 Daily Video 2 (Skill 2.D)

Representing a Categorical Variable with Graphs

What Will We Learn?

How can we represent multiple sets of data for the same categorical variable in tabular form?
How can we represent multiple sets of data for the same categorical variable graphically?
How do these tabular and graphical representations help us compare multiple sets of categorical data?

Comparing categorical data

A high school teacher gave a survey to the students in all of her classes on the first day of school. The survey asked each student to record their sex (male or female). Another question asked:
"Which of the following superpowers would you most like to have?

Invisibility Telepathy (read minds) Freeze time Super strength Fly

The teacher wants to compare superpower preferences for male and female students.

Here are data for the 80 boys who completed the survey.

Freeze time	Fly	Freeze time	Invisibility	Freeze time	Freeze time	Fly	Fly	Fly	Freeze time
Invisibility	Invisibility	Freeze time	Freeze time	Invisibility	Fly	Freeze time	Invisibility	Super strength	Freeze time
Freeze time	Fly	Super strength	Super strength	Super strength	Freeze time	Fly	Freeze time	Freeze time	Fly
Freeze time	Fly	Fly	Super strength	Invisibility	Freeze time	Telepathy	Invisibility	Freeze time	Fly
Freeze time	Fly	Invisibility	Freeze time	Telepathy	Telepathy	Telepathy	Invisibility	Fly	Fly
Fly	Invisibility	Invisibility	Freeze time	Freeze time	Fly	Freeze time	Freeze time	Super strength	Invisibility
Fly	Freeze time	Super strength	Invisibility	Fly	Super strength	Freeze time	Super strength	Invisibility	Fly
Fly	Telepathy	Telepathy	Super strength	Freeze time	Freeze time	Freeze time	Freeze time	Invisibility	Telepathy

Here are data for the 125 girls who completed the survey.

Telepathy	Telepathy	Invisibility	Telepathy	Telepathy	Fly	Freeze time	Fly	Telepathy	Super strength	Fly
Telepathy	Freeze time	Freeze time	Fly	Fly	Freeze time	Fly	Freeze time	Fly	Fly	Telepathy
Invisibility	Super strength	Telepathy	Invisibility	Fly	Telepathy	Telepathy	Fly	Freeze time	Super strength	Freeze time
Invisibility	Telepathy	Fly	Fly	Telepathy	Freeze time	Telepathy	Fly	Freeze time	Invisibility	Freeze time
Freeze time	Freeze time	Invisibility	Telepathy	Telepathy	Super strength	Fly	Telepathy	Fly	Freeze time	Invisibility
Fly	Freeze time	Fly	Invisibility	Telepathy	Invisibility	Invisibility	Fly	Fly	Freeze time	
Telepathy	Telepathy	Fly	Fly	Invisibility	Invisibility	Fly	Invisibility	Freeze time	Freeze time	
Freeze time	Telepathy	Freeze time	Fly	Telepathy	Fly	Fly	Fly	Telepathy	Fly	
Telepathy	Fly	Invisibility	Invisibility	Freeze time	Fly	Invisibility	Freeze time	Freeze time	Fly	
Freeze time	Telepathy	Telepathy	Invisibility	Telepathy	Invisibility	Fly	Invisibility	Fly	Fly	
Fly	Telepathy	Freeze time	Invisibility	Fly	Fly	Fly	Fly	Telepathy	Super strength	
Freeze time	Freeze time	Telepathy	Fly	Invisibility	Freeze time	Telepathy	Fly	Telepathy	Freeze time	

How can we compare the distribution of superpower preference for these two groups?

Comparing categorical data with frequencies

The frequency table summarizes the data on the superpower preference for the 80 male and the 125 female students in this high school teacher's classes. (Copy frequencies from video)

	Fly	Freeze Time	Invisibility	Super strength	Telepathy
Male					
Female					

Which of the following statements is supported by the table?

(a) Fly was the most preferred superpower for both males and females. _____

(b) Freeze time was an equally popular choice of preferred superpower for males and females. _____

(c) Invisibility was a more popular choice of preferred superpower for females than males. _____

(d) Super strength was twice as popular among males as females for their preferred superpower. _____

(e) Telepathy was the preferred superpower for less than 10% of males but almost 25% of females. _____

Name_____

Comparing categorical data with relative frequencies

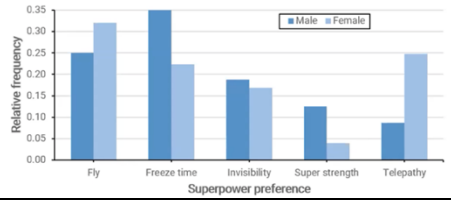
The frequency table summarizes the data on superpower preference for the 80 male and 125 female students in the high school teacher's classes.

	Fly	Freeze time	Invisibility	Super strength	Telepathy
Male	20	28	15	10	7
Female	40	28	21	5	31

The different sizes of the two groups – 80 and 125 – makes it hard to compare the distribution of superpower preference for males and females. Comparison is easier if we calculate relative frequencies within each group. (Calculate relative frequencies below.)

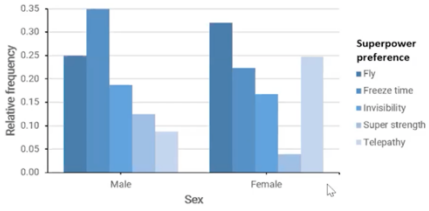
	Fly	Freeze Time	Invisibility	Super strength	Telepathy
Male					
Female					

Using bar charts to compare categorical data



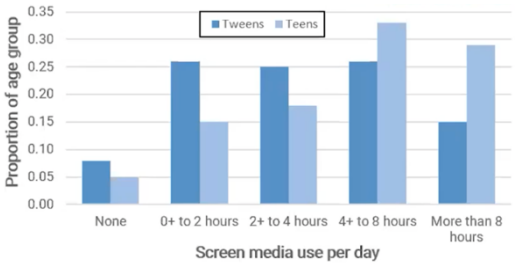
Be sure to:

- _____
- _____
- _____
- _____



Let's Practice: Describing categorical data from a bar chart

In 2019, Common Sense Media surveyed a random sample of more that 1600 U. S. 8- to 18-year-olds. The bar chart compares data on average daily screen media use by tweens (ages 8 to 12) and teens ages (13 to 18) in the sample. **Note:* 0+ to 2 hours indicates an average more than 0 hours up to and including 2 hours per day.



Which of the following statements is supported by the graph?

(a) Tweens generally report more screen media use than teens. _____

(b) About half of tweens and teens reported 2+ to 8 hours of screen media use per day. _____

(c) For both tweens and teens, the second most reported amount of screen media use per day is more than 8 hours. _____

(d) A similar proportion of tweens and teens reported 0+ to 8 hours of screen media use per day. ___

(e) Over 50% of tweens and teens report more than 4 hours of screen media use per day. _____

What Should We Take Away?

How can we represent multiple sets of data for the same categorical variable in tabular form?

With a _____ or _____ table.

How can we represent multiple sets of data for the same categorical variable graphically?

With a _____ bar graph.

How do these tabular and graphical representations help us compare multiple sets of categorical data?

Comparing _____ between and within the groups reveals information that can be used to _____ about the data _____

Name_____

AP Statistics CED 1.5 Daily Video 1 (Skill 2.B)

Representing a Quantitative Variable with Graphs

What Will We Learn?

What are the two types of quantitative variable, and how do we distinguish them?
What types of graphs can be used to represent quantitative data and what are the advantages and disadvantages of each?

A Quick Review

Categorical Variable: A variable that takes values that are _____ or _____.
Quantitative Variable: A variable that takes _____ values for a _____ or _____ quantity.

- A _____ variable can take on a _____ number of values (with gaps)
  Examples: _____
- A _____ variable can take on _____ values, but those values cannot be counted (no gaps)
  Examples: _____

Was Flint's water safe to drink?

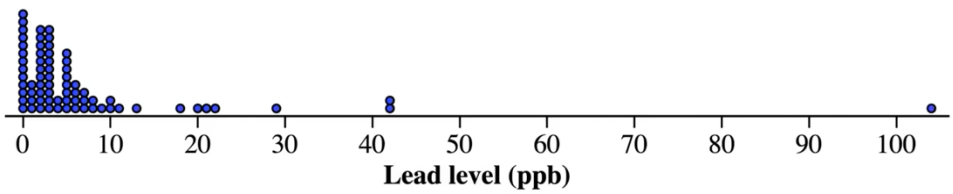
City officials measured lead levels in 71 water samples from Flint residents in January to June 2015. Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11
13 18 20 21 22 29 42 42 104

Is this discrete or continuous data?

Represent this data with a graph.

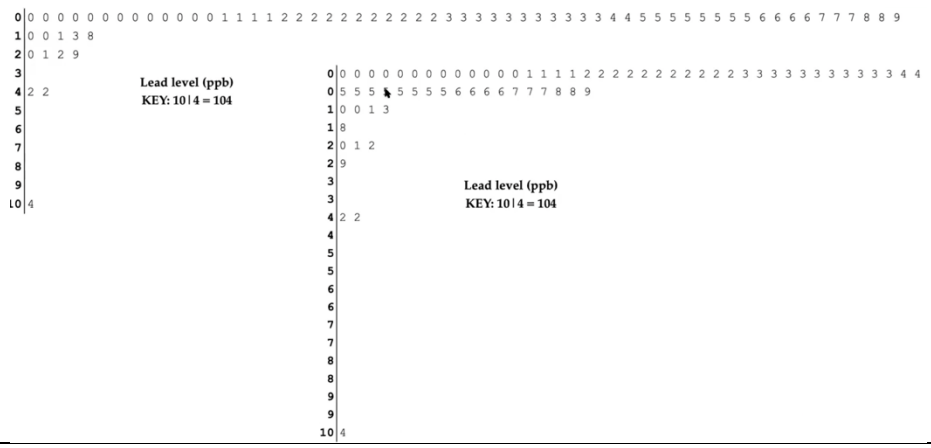
Dotplot



Advantages:
Shows every _____ in the data set. Easy to see the _____ of the distribution.

Disadvantages
Difficult to make for _____ data sets.

Stem and leaf plot

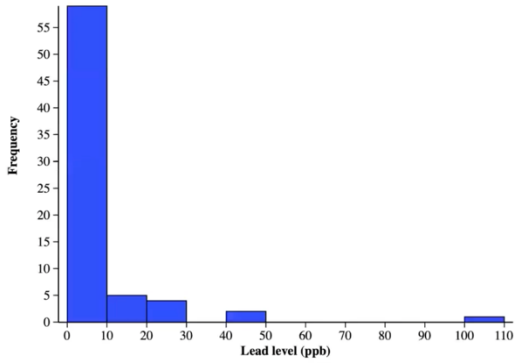


Advantages:
Shows every _____ value in the data set.
Easy to see the _____ of the distribution.

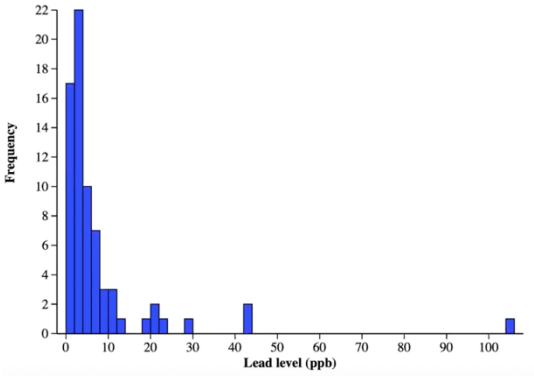
Disadvantages:
Difficult to make for _____ data sets.

Name_____

Histogram With a histogram you are going to use _____ of values.



Intervals of width 10



Intervals of width 2

Advantages

Easier to make for _____.

Easy to see _____ of the distribution.

Disadvantage

Does not show _____ individual value in the data set.

What Should We Take Away?

What are the two types of quantitative variable, and how do we distinguish them?

_____: countable (with gaps)

_____: not countable (no gaps)

What types of graphs can be used to represent quantitative data and what are the advantages and disadvantages of each?

_____ and _____: (+) shows _____ value, easy to see _____, (-) hard to make for _____ data sets.

_____: (+) easier to make for _____ data sets, easy to see _____, (-) does not show _____ value.

Name_____

AP Statistics CED 1.6 Daily Video 1 (Skill 2.A)

Describing the distribution of a Quantitative Variable

What Will We Learn?

What are the important characteristics to discuss when describing the distribution of quantitative data?

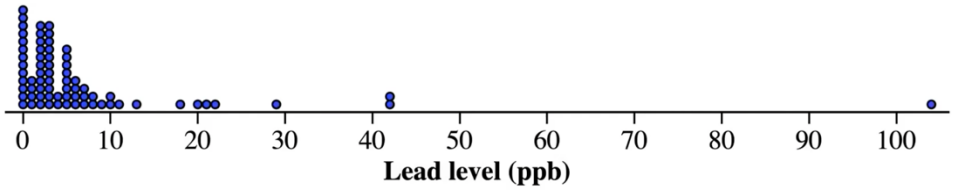
What are the best ways to discuss the important characteristics when describing a distribution of quantitative data?

Was Flint's water safe to drink?

City officials measured lead levels in 71 water samples from Flint residents in January to June 2015.

Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11
13 18 20 21 22 29 42 42 104



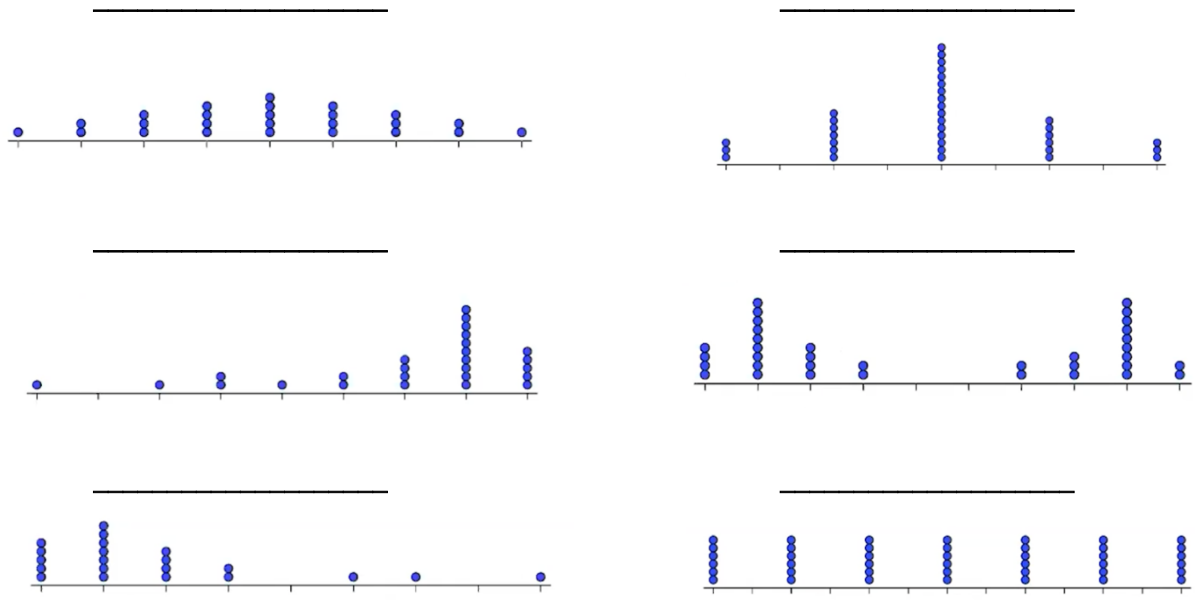
Describe the distribution of lead level for the 71 water samples in Flint.

How to Describe a Distribution of Quantitative Data

There are four important characteristics you have to include in the description:

* _____ * _____ * _____ * _____

Shape (label each as you watch the video)



Center

Which value in the distribution best describes the _____?

Variability

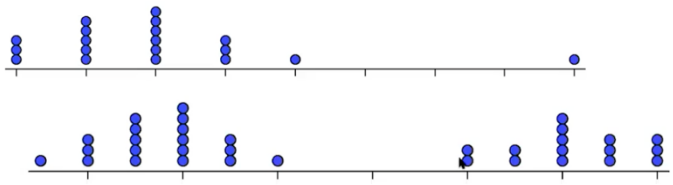
Are the values in the distribution _____ together or are they _____ out?

Name_____

Unusual Features

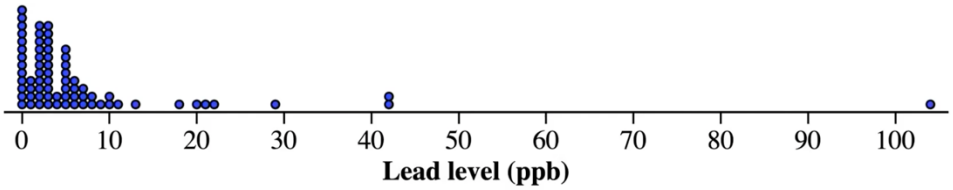
_____

_____



Was Flint's water safe to drink?

Describe the distribution of lead level for the 71 water samples of Flint.



Shape: The distribution of lead levels is _____ and _____ to the right.
Center: The _____ lead level is around _____.
Variability: Lead levels _____ from a _____ value of 0 to a _____ value of 104 parts per billion.
Unusual Features: There is a _____ of values between _____, a large _____ between _____, and several possible _____.

*Note: Make sure that you include the _____ of the problem in your solution!!!

What Should We Take Away?

What are the important characteristics to discuss when describing the distribution of quantitative data?

There are four characteristics to include in your solution.

_____, _____, _____, and

_____

What are the best ways to discuss the important characteristics when describing a distribution of quantitative data?

Shape: _____

Center and Variability: More in Daily Video 1.7

Unusual features: _____

Name_____

AP Statistics CED 1.7 Daily Video 1 (Skill 2.C)

Summary Statistics for a Quantitative Variable

What Will We Learn?

What summary statistics can be used to describe the center and position of a distribution of quantitative data?

What summary statistics can be used to describe the variability of a distribution of quantitative data?

Was Flint's water safe to drink?

City officials measured lead levels in 71 water samples from Flint residents in January to June 2015. Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11
13 18 20 21 22 29 42 42 104

Use summary statistics to describe the center and variability of the distribution of lead levels.

Summary Statistics for Center

Mean:

Sum of _____ the data values _____ by the _____ of values.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Median:

_____ value of an _____ data set (_____ number of values)

_____ of the _____ middle value of an ordered data set (even number of values)

Q1 and Q3:

The first quartile, _____, is the _____ of the _____ half of the ordered data set.

The third quartile, _____, is the _____ of the _____ half of the ordered data set.

Was Flints water safe to drink? (See data and graph above)

Use summary statistics to describe the center and variability of the distribution of lead levels.

Mean: $\bar{x} = \frac{(0+0+0+\ldots+42+42+104)}{71} = \frac{519}{71} = 7.31$

Median: Because there is an odd (71) number of values, the _____ value is the median.

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5
6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104

The median is _____. Interpreting this median: About _____ of the lead levels are _____ 3 ppb and about _____are above 3 ppb.

Q1: The first quartile is the _____ of the first half of the data (don't include median).

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 **(Circle Q1)**

Q3: The third quartile is the _____ of the second half of the data (don't include median)

3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104 **(Circle Q3)**

Name_____

Summary Statistics for Variability

Range: The difference between the _____ value and the _____ data value.

Interquartile Range (IQR): Difference between the _____ and _____ quartiles. **(Q3 – Q1)**

Standard Deviation: _____ distance that each value is away from the _____. The _____ of standard deviation is the _____.

$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

Was Flint's water safe to drink?

Using summary statistics to describe the variability of the distribution of lead level (ppb) for the samples of 71 water samples in Flint.

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5
6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104

Range: _____ = _____ = _____

The range of the distribution of lead levels is _____. (Range is a **SINGLE VALUE!)**

Interquartile Range: _____.

The _____ of the values for lead level has a range of _____.

Standard deviation:

$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

Steps:

1. Find the _____ between each individual value and the mean and square.
2. _____ up all of these values and _____ by (*n*-1). (What you have here is called the variance.)
3. Take the _____ of the variance to get the standard deviation.

$$s_x = \sqrt{\frac{1}{71-1}\left((0-7.31)^2 + (0-7.31)^2 + \ldots + (42-7.31)^2 + (104-7.31)^2\right)} = \sqrt{205.84} = 14.35$$

The lead level from each sample _____ by about _____ from the mean of _____.

What Should We Take Away?

What summary statistics can be used to describe the center and position of a distribution of quantitative data?

Center: _____

Position: _____

What summary statistics can be used to describe the variability of a distribution of quantitative data?

Variability: _____

Name_____

AP Statistics CED 1.7 Daily Video 2 (Skill 4.B)

Summary Statistics for Quantitative Variable

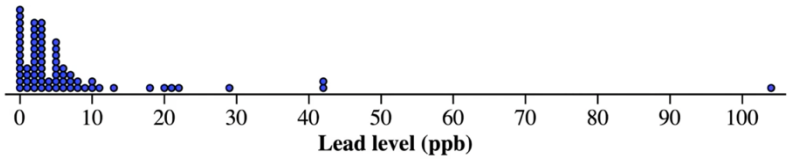
What Will We Learn?

How can we determine if a value in a data set is an outlier?
Which summary statistics are resistant and which are nonresistent?
Which measures of center and variability are best for describing a distribution?

Was Flint's water safe to drink?

City officials measured lead levels in 71 water samples from Flint residents in January to June 2015. Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11
13 18 20 21 22 29 42 42 104



Summary Statistics from previous videos:

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

Determining Outliers – Two Methods

Method 1 – An outlier is a value more that _____ below the _____ and more that _____ above the _____.

Low outlier < _____; < _____; < _____
There are ____ low outliers!

High outlier > _____; > _____; > _____
There are _____ high outliers! They are: _____

Method 2 – An outlier is a value located 2 or more _____ above, or below the mean.

Low outlier < _____; < _____; < _____
There are ____ low outliers.

High outlier > _____; > _____; > _____
There are _____ high outliers! They are: _____

Name_____

How do outliers influence summary statistics?

Here are the original summary statistics.

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

range = 104 – 0 = 104
IQR = 7 – 2 = 5

Here are the summary statistics with the highest outlier _____ removed.

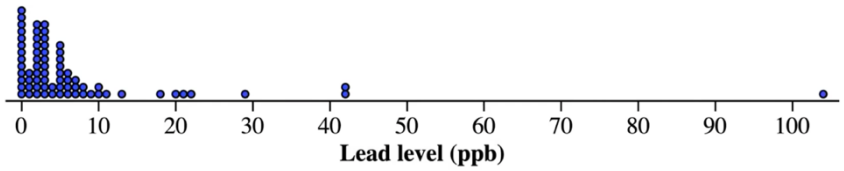
n	mean	SD	min	Q_1	med	Q_3	max
70	5.929	8.45	0	2	3	6	42

range = 42 – 0 = 42
IQR = 6 – 2 = 4

The _____, _____, and _____ were highly influenced by removing the outlier (_____)

The _____ and the _____ were not greatly affected. (_____)

What measures of center and variability are best?



The answer depends on the shape of the distribution.

For a _____ distribution, use _____ for center and _____ for variability.
For a _____ distribution, use _____ for center and _____ for variability.

What Should We Take Away?

How can we determine if a value in a data set is an outlier?

- **Less the _____ below Q1 or more the _____ above Q3**
- **_____ away from the mean.**

Which summary statistics are resistant and which are nonresistent?

Resistant: _____, _____
Nonresistant: _____, _____, _____

Which measures of center and variability are best for describing a distribution?

For _____ distributions: _____ and _____
For _____ distributions: _____ and _____

Name_____

AP Statistics CED 1.8 Daily Video 1 (Skill 2.B, 2.A)

What Will We Learn?

What is the five-number summary and how do we use it to make a boxplot?

How does the shape of the graph influence the relative relationship of the mean and median?

Was Flint's water safe to drink?

City officials measured lead levels in 71 water samples from Flint residents in January to June 2015.

Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11
13 18 20 21 22 29 42 42 104

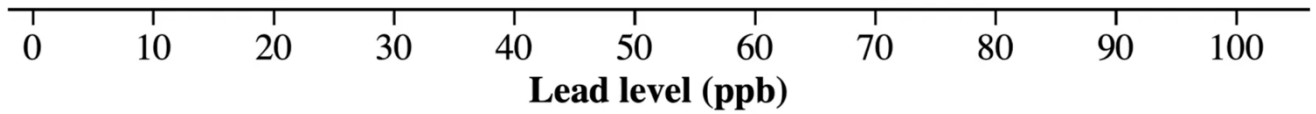
Here is the summary statistics from previous video:

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

Find the five-number summary and use it to make a boxplot.

Five-number summary and boxplot (Circle the data and label the line below as you watch video.)

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5
6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104



Boxplot

Advantages:

Shows the _____ and _____

Splits the data into _____

Disadvantages:

Does not show _____ value.

Can hide certain features of the _____ of the distribution.

Measures of center – what do you notice?

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

The _____ is much larger than the median!

The mean is _____.

The median is _____.

As a general rule:

- Skewed _____ distribution, mean _____ median
- Skewed _____ distribution, mean _____ median
- _____ distribution; mean _____ median

Name_____

What Should We Take Away?

What is the five-number summary and how do we use it to make a boxplot?

_____, _____, _____, _____, _____

Use the _____ to split the data into _____.

How does the shape of the graph influence the relative relationship of the mean and median?

- Skewed _____ distribution, mean _____ median
- Skewed _____ distribution, mean _____ median
- _____ distribution; mean _____ median

Name_____

AP Statistics CED 1.9 Daily Video 1 (Skill 2.D)

Comparing distributions of a quantitative variable

What Will We Learn?

What are the important characteristics to discuss when comparing distributions of quantitative data? What is needed for a complete response when comparing distributions of quantitative data?

A Quick Review

How to Describe a Distribution of Quantitative Data

Shape: _____, _____, _____, _____, _____, _____

Center: _____, _____

Variability: _____, _____, _____

Unusual features: _____, _____, _____,

Example: How to Compare Distributions of Quantitative Data (Free Response Question: 2015 #1)

Two large corporations, A and B, hire many new college graduates as accountants at entry-level positions. In 2009 the starting salary for an entry-level accountant position was $36,000 a year at both corporations. At each corporation, data were collected from 30 employees who were hired in 2009 as entry-level accountants and were still employed at the corporation five years later. The yearly salaries of the 60 employees in 2014 are summarized in the boxplots to the right.

(a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

Model Solution

The _____ of the distribution of yearly salary appears to be _____ (fairly symmetric) for _____ corporations A and B. The _____ salary is approximately the same for _____ corporations. The _____ and _____ of the salaries are greater for Corporation A than for Corporation B. The _____ salaries for Corporation A are outliers while Corporation B has _____.

Note that this Model Solution address **all four important characteristics**. (Highlight these key words, as you watch the video!) Additionally, highlight the **comparative words** used in the solution. Finally, highlight the words that supply the **context** to the solution.

Name_____

(b) Suppose both corporations offered you a job for $36,000 a year as an entry-level accountant.

(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

Five years after starting, as least 3 out of 30 (10%) of the salaries at Corporation A are greater than the maximum salary at Corporation B. If I accept the offer from Corporation A, I might be able to make a higher salary at Corporation A than at Corporation B.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

Five years after starting, the minimum salary at Corporation B is greater than at corporation A. It fact, at Corporation A it looks like some people are still making the starting salary of $36,000 and never received a raise in the five years since they were hired. So, if I work at Corporation A, I might never receive a raise in salary.

(Highlight the key parts of the responses as you watch the video. Notice that each response refers to a key feature of the boxplot!)

What Should We Take Away?

What are the important characteristics to discuss when comparing distributions of quantitative data?

_____, _____, _____, and _____

What is needed for a complete response when comparing distributions of quantitative data?

Address the _____
Use _____
Include _____

Name_____

AP Statistics CED 1.10 Daily Video 1 (Skill 2.D, 3.A)

The Normal Distribution

What Will We Learn?

How can we use percentile to describe the position of a value in a quantitative data set?

How can we use standardized scored to describe the position of a value in a quantitative data set?

Determining relative position

Percentile:

Percentile is the _____ of data values _____ to a given value.

Interpret: "The value of _____ is at the p^{th} percentile. About (p) percent of the values are less than or equal to _____."

Standardized score:

A standardized score for a given data value is calculated as

standardized score = $z\text{-score} = \frac{x_i - \mu}{\sigma}$

Interpret: "The value of _____ is (z-score) standard deviations above/below the mean."

CAUTION: _____ and _____ can be calculated for distributions with any shape!

Was Flint's water safe to drink?

City officials measured lead levels in 71water samples from Flint residents in January to June 2015. Here are the data (in part per billion):

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104

Summary Statistics from previous videos:

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

Find the percentile and z-score for the water samples of 20 ppb and 2 ppb.

For the water sample of 20 ppb

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104

Percentile: _____ = _____

"The value of 20 ppb is at the _____ percentile. About _____ of the values are _____ to 20 ppb.

z-score:

The value of 20 ppb is _____ standard deviations _____ the mean.

Name_____

For the water sample of 2 ppb

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 4 4 5 5 5 5 5
5 5 5 6 6 6 6 7 7 7 8 8 9 10 10 11 13 18 20 21 22 29 42 42 104

n	mean	SD	min	Q_1	med	Q_3	max
71	7.31	14.347	0	2	3	7	104

Percentile: _____ = _____

The value of 2 ppb is at the _____ percentile. About _____ of the values are _____ to 2 ppb

z-score:

The value of 2 ppb is _____ standard deviations _____ the mean.

What Should We Take Away?

How can we use percentile to describe the position of a value in a quantitative data set?

Percentile is the _____ of data values _____ a given value.

How can we use standardized scored to describe the position of a value in a quantitative data set?

A z-score tells us the number of standard deviations _____ or _____ the mean.

Name_____

AP Statistics CED 1.10 Daily Video 2 (Skill 2.D, 3.A)

The Normal Distribution

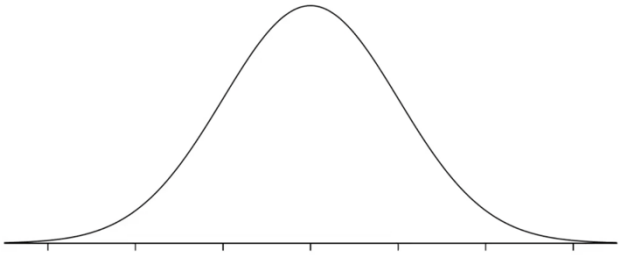
What Will We Learn?

What is a normal distribution?

How can we use the empirical rule to find the percent of data values in a given interval for a normal distribution?

How can we use the z-scores to find the percent of data values in a given interval for a normal distribution?

Normal Distribution:



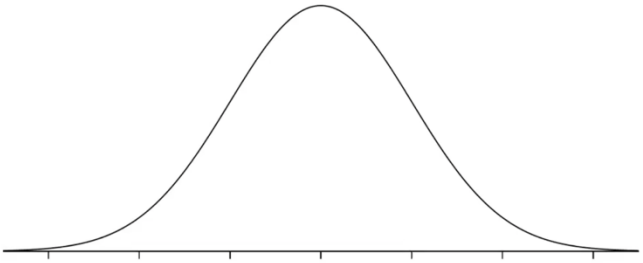
- A normal distribution is _____ (sometimes called a _____) and _____.
- Many _____ variables in the real world can be _____ by a normal distribution (height, temperature, blood pressure).
- Normal distributions are determined by the _____ (___) and the _____ _____ (___).
- Label the mean and indicate one standard deviation as done in the video.

Example: What is normal blood pressure?

Systolic blood pressure for adults can be modeled with a normal distribution with a mean of 110 mmHg and a standard deviation of 10 mmHg.

You always want to start with a picture. Then, label the mean at the very center and then label the values that are 1, 2, and 3 standard deviations above and below the mean.

(Shade the areas as you watch the video.)



Within 1 standard deviation of the mean: _____
Within 2 standard deviations of the mean: _____
Within 3 standard deviations of the mean: _____

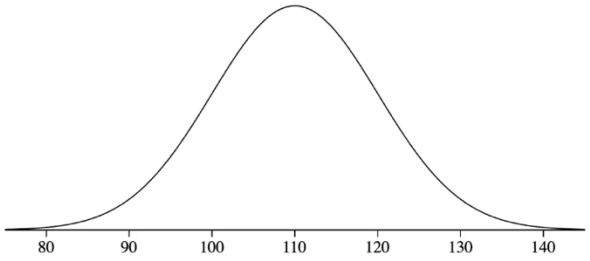
This is often referred to at the **68-95-99.7 Rule** or often referred to as **The Empirical Rule.**

Name_____

What is normal blood pressure?

What percent of adults have a systolic blood pressure below 100 mmHg?

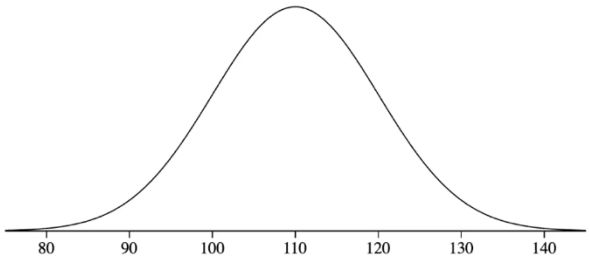
(Shade the distribution as you watch the video.)



_____

What percent of adults have a systolic blood pressure below 130 mmHg?

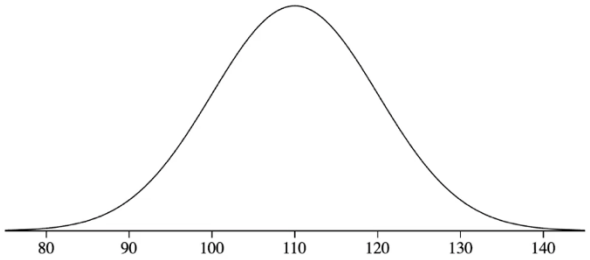
(Shade the distribution as you watch the video.)



_____

What percent of adults have a systolic blood pressure below 125 mmHg?

(Shade the distribution as you watch the video.)



Because 125 is not exactly 1, 2, or 3 standard deviations from the mean, so we cannot use the Empirical Rule. We need a new tool…The z-score!!

$$z\text{-score} = \frac{x_i - \mu}{\sigma}$$

_____

(Once we know the z-score, we will use Table A to determine the area. Pay attention to the video to learn how to use Table A.)

A z-score of _____ = _____; This means that about _____ of adults have a systolic blood pressure _____ 125 mmHg.

What Should We Take Away?

What is a normal distribution?

A model for _____ that often appears in the real world.

How can we use the empirical rule to find the percent of data values in a given interval for a normal distribution?

About _____ of the data is within ___ SD of the mean. About _____ of the data is withing ___ SD of the mean. About _____ of the data is within ___ SD of the mean.

How can we use the z-scores to find the percent of data values in a given interval for a normal distribution?

Calculate a _____ and then use _____!

Name_____

AP Statistics CED 1.10 Daily Video 3 (Skill 2.D, 3.A)

The Normal Distribution

What Will We Learn?

How can we use the z-scores to find the percent of data values in a given interval for a normal distribution (left, right, between)?

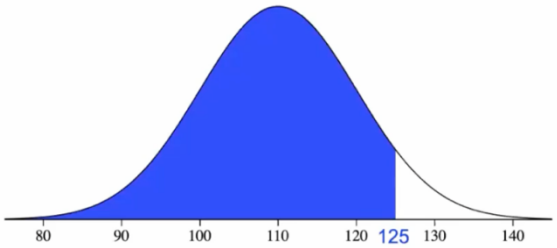
How can we find a value, given an area (proportion) for a normal distribution?

Quick Review – Area to the left

Systolic blood pressure for adults can be modeled with a normal distribution with a mean of 110 mmHg and a standard deviation of 10 mmHg.

What percent of adults have a systolic blood pressure below 125 mmHg?

$$z\text{-score} = \frac{x_i - \mu}{\sigma} \qquad z = \frac{125 - 110}{10} = 1.50$$

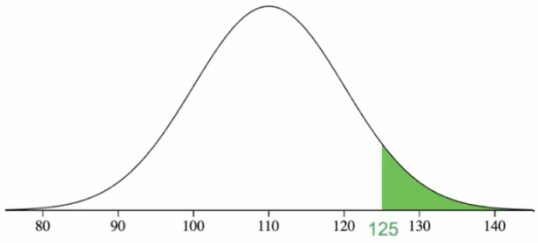


Using Table A, we found the z-score of 1.50 = _____ Which tells us that about _____ of adults have a systolic blood pressure less than 125 mmHg.

Example: Area to the right

Some experts consider a systolic blood pressure above 125 mmHg to be considered as "elevated". What proportion of adults have an "elevated" systolic blood pressure?

(If the area to the left is known, you can easily find the area to the right by simply subtracting it from 1.)



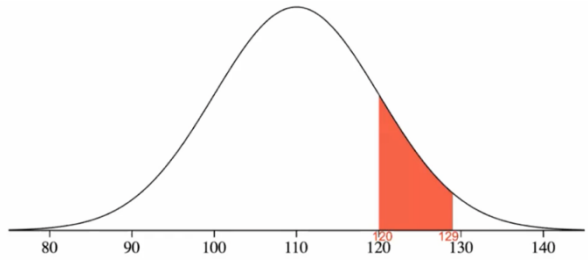
Area = _____

The proportion of adults with an "elevated" blood pressure is _____.

Example: Area between

Some experts consider a systolic blood pressure between 120 and 129 mmHg to be hypertension stage 1. What proportion of adults are hypertension stage 1?

(Always start by drawing a picture. Then, label and shade the region you are trying to find.)



Find z-score of 120 and area from Table A:

Find z-score of 129 and area from Table A:

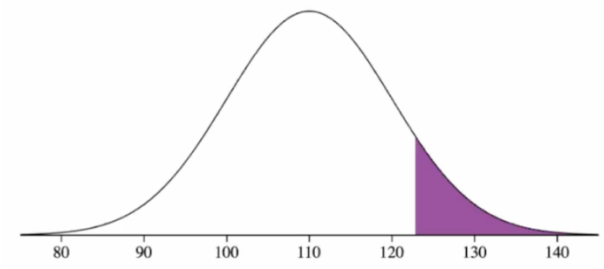
Area = _____

The proportion of adults that are hypertension stage 1 is about _____.

Name_____

Example: Working backwards

A person is considered high risk for they are in the highest 10% of all systolic blood pressures. What level of blood pressure is considered high risk?



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

Use Table A as demonstrated in the video and circular the indicated value.

Then using the formula for z-scores, plug in the values that you know and solve for x_i.

$$z\text{-score} = \frac{x_i - \mu}{\sigma}$$

x = _____

About 10% of adults are high risk, with a blood pressure of more than _____.

What Should We Take Away?

How can we use the z-scores to find the percent of data values in a given interval for a normal distribution (left, right, between)?

Left: get the area from _____
Right: 1 - _____
Between: _____ two areas from Table A

How can we find a value, given an area (proportion) for a normal distribution?

Use _____ to find the z-score.
Set up equation and _____.