

Data and Statistics

Statistical Measures and Analysis

Statistical Questions

A *statistical question* is one that anticipates a variety of different, true responses.

Ex: What color eyes do you have?

How much did it rain over the past 5 days?

What is your favorite class?

Numerical questions are statistical questions that can be responded to/answered with numerical data.

Non-statistical Questions have only one true answer.

Ex: Who is the principal of Pleasant Middle School?

How many stripes are on the United States flag?

Do you have a dog?

Measures of Center

Key Vocabulary

Measure of Center - A single number summary of the rough center of a numerical data set.

Mean - The average value of a numerical data set.

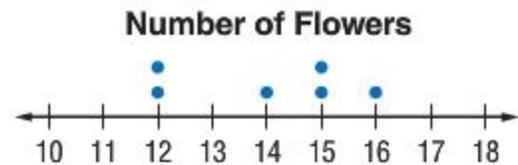
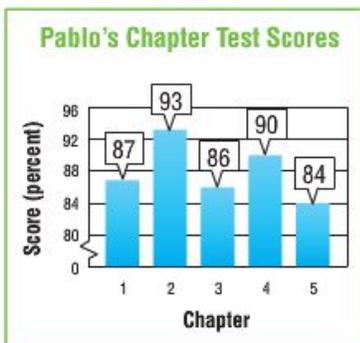
Median - The middle value of a numerical data set.

Mode - The most commonly occurring value in a numerical data set.

Calculating Mean

1. Add up all data in the given data set.
2. Divide sum (from #1) by the number of data entries.

Examples

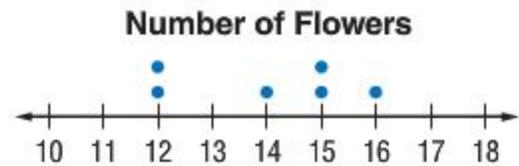


Calculating Median

1. Order the data set from least to greatest.
2. Cross out data from the left and right until you reach the middle value.
 - a. If there are *two* numbers in the middle, find the mean between the two.

Examples

18 20 22 15 19 18 18



Calculating Mode

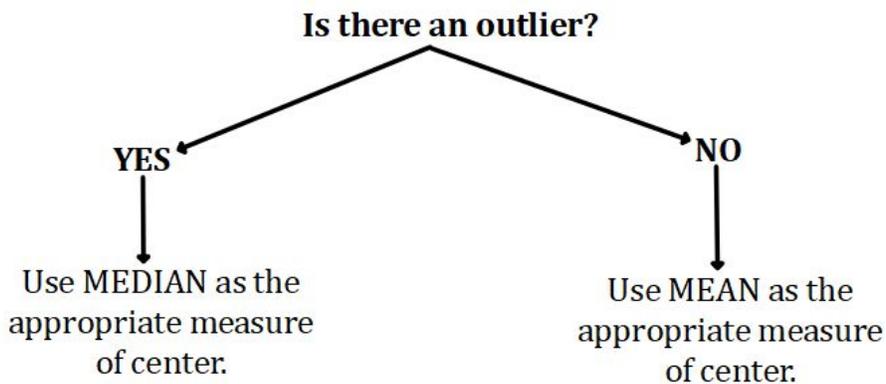
It is possible...

1. That there is NO mode. *Do not write 0 if there is no mode*
2. That there is only ONE mode.
3. That there are MULTIPLE modes.

Appropriate Measures

Key Vocabulary

Outlier(s) - Extreme data values that are either much bigger, or much smaller than the other data.



**Only use *mode* as the appropriate measure of center when there are many of the same number near the center of the data set.

Measures of Variation

Key Vocabulary

Measures of Variation - Used to describe the distribution or spread of a given numerical data set.

Quartiles - Values that divide the data set into four equal parts. (Quarters)

First Quartile - End of the first quarter of the data set (Q1)

Second Quartile - Median

Third Quartile - End of the third quarter of the data set (Q3)

Interquartile Range - The distance between the first and third quartiles of the data set.

Range - The difference between the greatest and least data values.

Example #1

346 250 433 369 422 298

Example #2

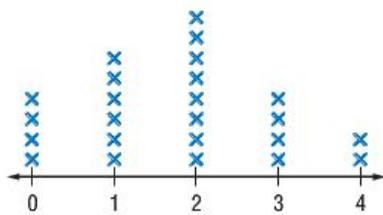
24 53 38 12 31 19 26

Statistical Displays

Line (Dot) Plots

These displays are best used to show _____.

Examples

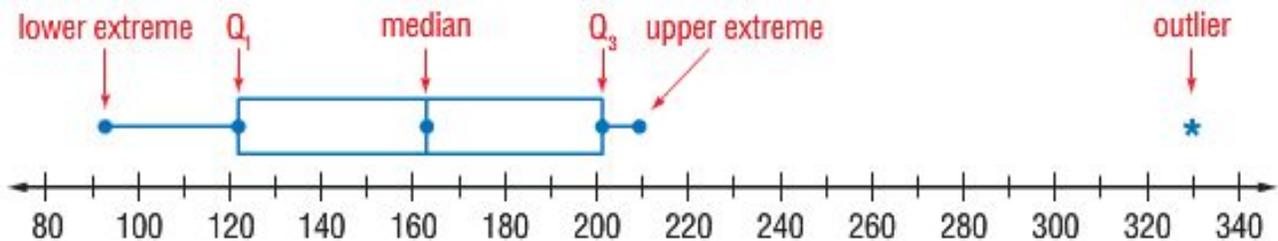


Number of Projects				
2	4	3	3	1
0	5	4	2	2
1	3	2	1	2



Box Plots - aka (Box-and-Whisker Plot)

These displays use a number line to show the _____.



Box plots separate data into four parts. Even though the parts may differ in length, each contains 25% of the data. The box shows the middle 50% of the data.

Example

Use the table. (Examples 1–3)

a. Make a box plot of the data.

Depth of Recent Earthquakes (km)						
5	15	1	11	2	7	3
9	5	4	9	10	5	7



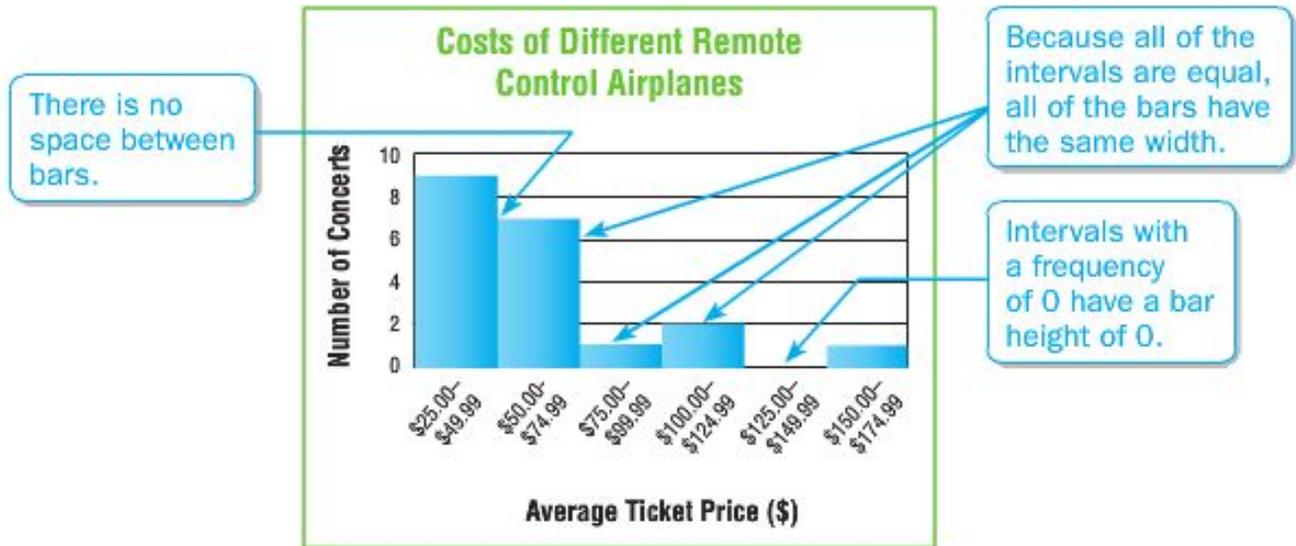
b. What percent of the earthquakes were between 4 and 9 kilometers deep? _____

Histograms

These displays are a type of _____ that is used to display numerical data that have been organized into equal intervals.

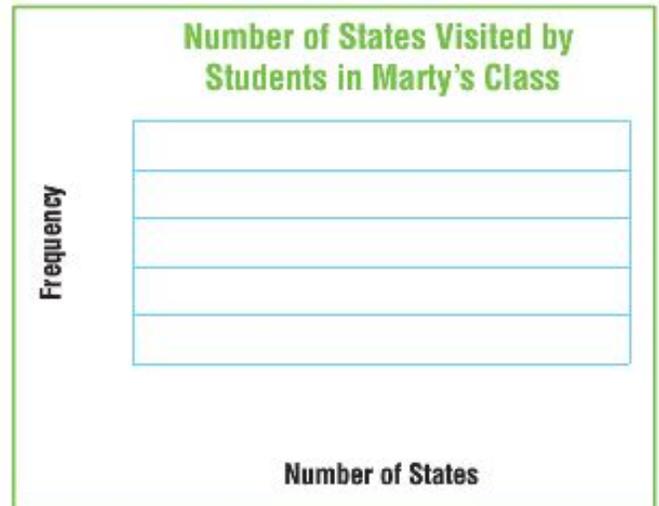
Intervals - Numerical data that is split up into equivalent 'ranges' of value.

Frequency Tables allow us to see how many pieces of data exist within each interval.



Example

Number of States Visited by Students in Marty's Class		
Number of States	Tally	Frequency
0-4		9
5-9		3
10-14		5
15-19		3
20-24		6
25-29		1



Line Graphs

These displays are used to show a _____.

On a line graph, the _____ is almost always labeled with some unit of _____.

Shape of Data Distribution

The *distribution* of a data set shows the arrangement of data values. This is done using the following:

Cluster - Data that is grouped closely together.

Gap - Numbers that have no data value(s) create gaps in the data.

Peak - The most frequently occurring values; this is also known as the *mode*.

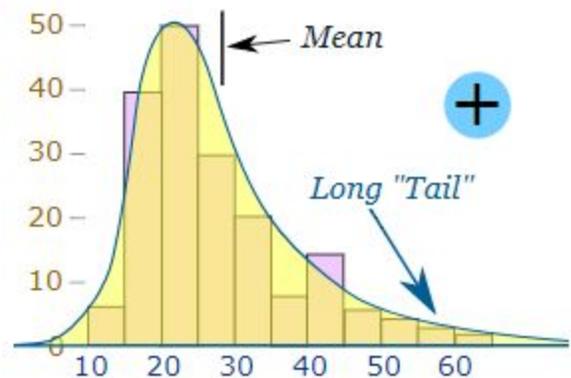
Symmetry - The left side of the distribution looks exactly like the right side.

Outliers - Data that exists much smaller or bigger than the accompanying data.

*Box plots can only be described using *symmetry*, *clusters*, and *outliers*.

Skewed Data - Data can be skewed to the left or to the right. To determine this direction, if any, look at the long tail in the distribution.

In the example shown to the right, the data would be considered *skewed to the right*.



Examples

