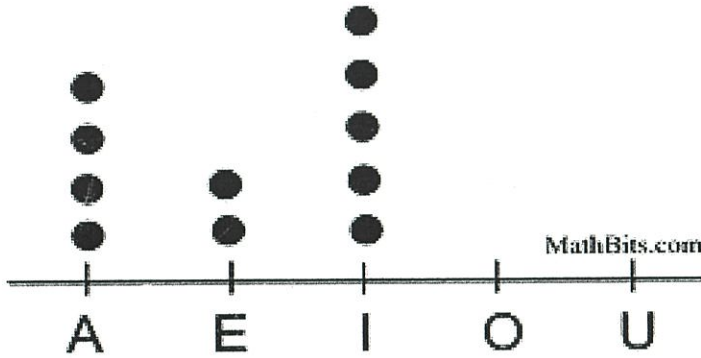


Dot Plots

Definition:

A dot plot is a type of graphic display using **filled in circles** (dots) and a simple scale to compare the counts (**frequency**) within categories or groups, when working with **small data sets with a small number of groups**.



Vowels in Antidisestablishmentarianism
(longest non-coined, non-technical word)

Dot Plot Information:

- Used with **relatively small sets of data and groups**.
- Quickly displays **clusters/gaps** of data and **outliers**.
- Each dot represents **one observation**.
(unless otherwise indicated)
- Dots are stacked in a **column**.
- Column heights represent count (**frequency**).
- May have **categories or numbers** on the x-axis.

If the **x-axis contains letters or words (non-numerical categories)**, the graph is said to be a **qualitative** dot plot. The dot plot shown above, is a qualitative dot plot, since A, E, I, O and U are not numerical values. Qualitative variables **cannot** be described in terms of center, spread, or shape. *→ Can't be measured*

If the **x-axis contains numbers (numerical categories)**, the graph is said to be a **quantitative dot** plot. This dot plot can be described in terms of center, spread or shape. *Can be measured*

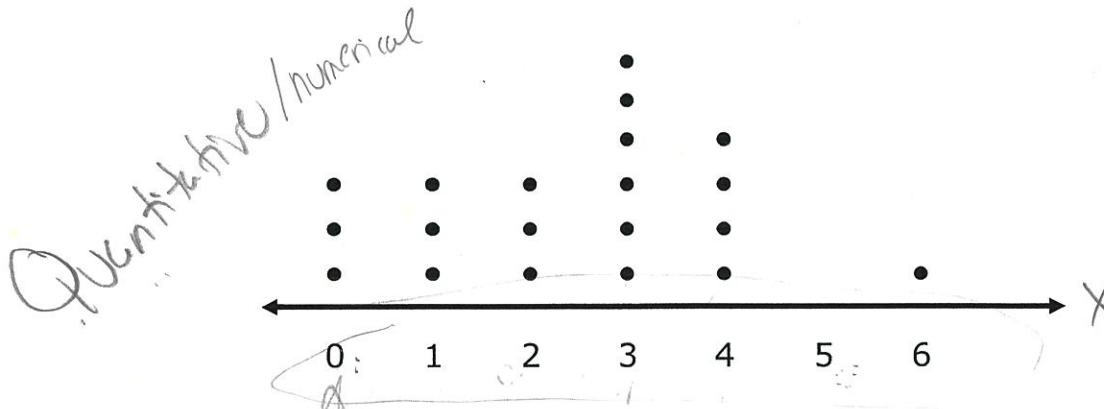
Reminder:

Dot plots can be quickly prepared for small amounts of data. They have no need for a frequency axis, and they offer a visual means of comparison. In dot plots, however, **you need to count to find the total number of data elements in each stack of dots**, and they can be difficult to construct and interpret for data sets with lots of points.

Dot plots:
Pros and Cons

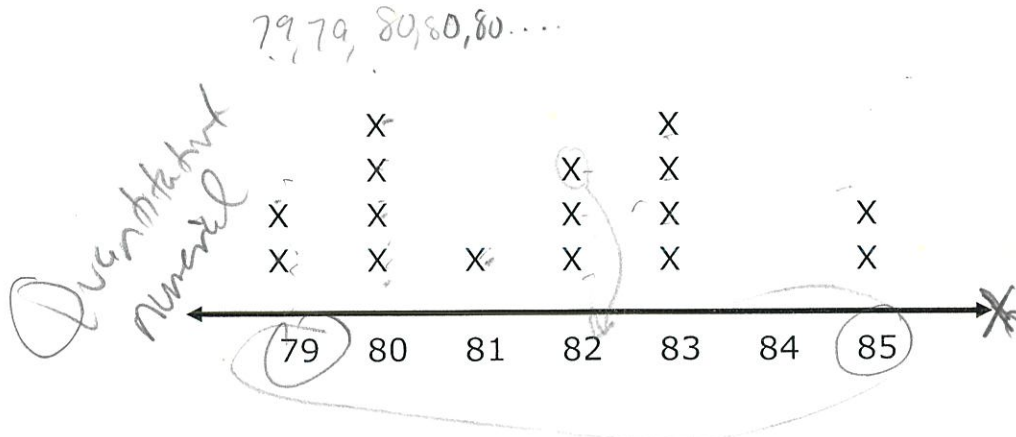
Examples:

1. The students in one social studies class were asked how many brothers and sisters (siblings) they each have. The dot plot here shows the results.



- a) How many of the students have **six** siblings?
1
- b) How many of the students have **no** siblings?
3
- c) How many of the students have **three or more** siblings?
11

2. The resting pulse rates were recorded for 16 boys in gym class before they exercised. The line plot here shows the results.



- a) What is the **range** (spread) of the pulse rates?
 $\text{max} - \text{min} = 85 - 79 = 16$
- b) What is the **mode** of the pulse rates?
80 & 83
- c) What is the **median** of the pulse rates?

$\boxed{82}$ $\frac{16}{2} = 8$ or put into Calc into L1

Shapes of Distributions

When graphed, the data in a set is arranged to show how the points are distributed throughout the set. These distributions show the **spread** (dispersion, variability, scatter) of the data. The spread may be **stretched** (covering a wider range) or **squeezed** (covering a narrower range).

The shape of a distribution is described by its number of peaks and by its possession of symmetry, its tendency to skew, or its uniformity. (Distributions that are **skewed** have more points plotted on one side of the graph than on the other.)

PEAKS: Graphs often display **peaks**, or local **maximums**. It can be seen from the graph that the data count is visibly higher in certain sections of the graph.

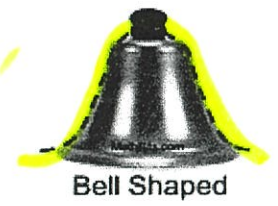
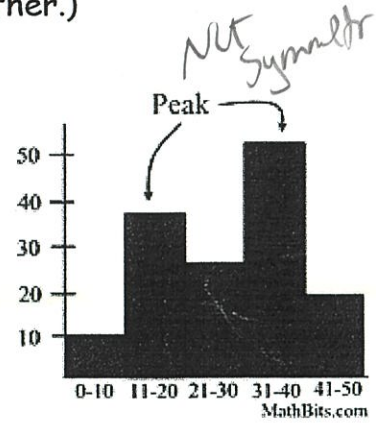
1. **one clear peak** is called a **unimodal** distribution.
2. **two clear peaks** are called a **bimodal** distribution.

(Here, the term "mode" is used to describe a local maximum in a chart (such as the midpoint of the a peak interval in a histogram). **It does not necessarily refer to the most frequently appearing score, as in the "central tendency mode"**.

3. **single peak at the center** is called **bell shaped** distribution.

Note: A bell shaped graph (bell curve), is a frequency distribution that resembles the outline of a bell when plotted on a graph.

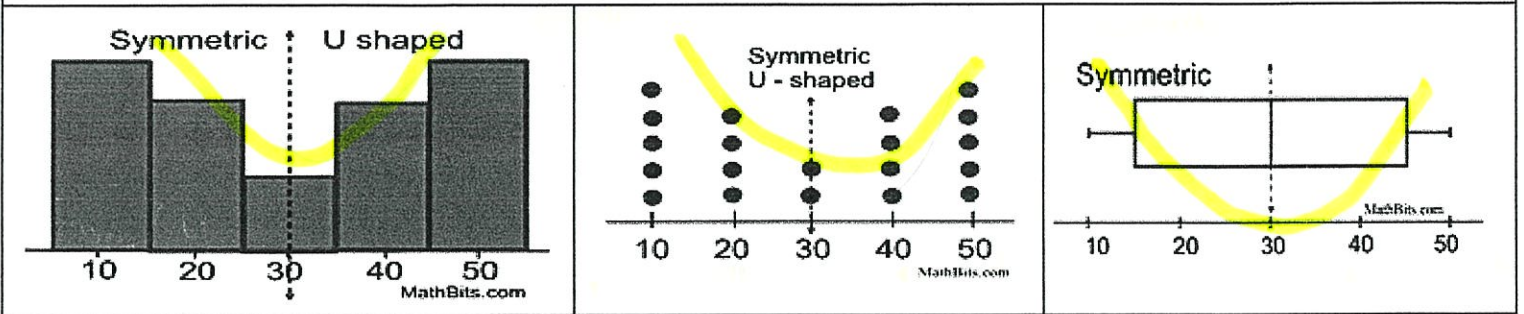
more than 2:
Multimodal



Shapes of Distributions (the graphs):

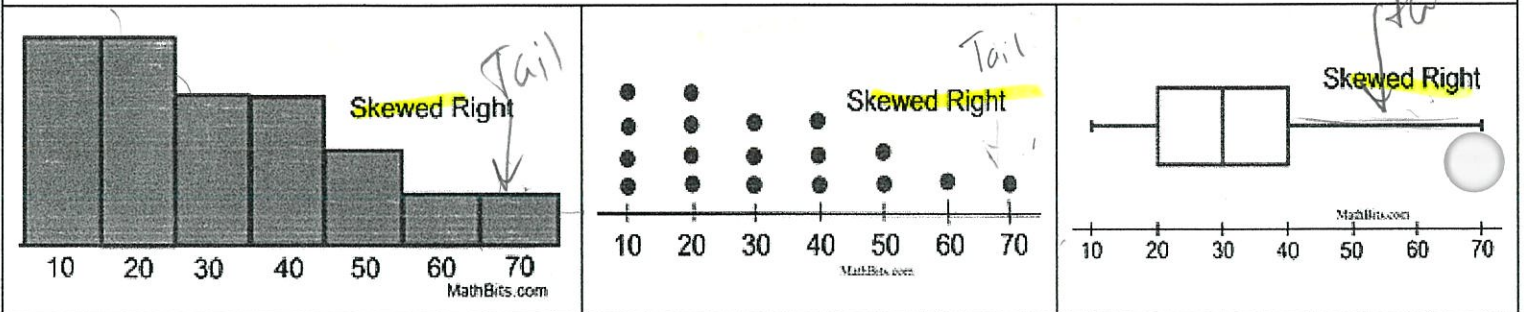
Histograms	Dot Plots	Box Plots
<p><i>Note:</i> The graphs shown below demonstrate the shapes of various sets of data. The histogram, dot plot and box plot in each separate section represent the same data set.</p>		
<p>Symmetric (bell shaped) - when graphed, a vertical line drawn at the center will form mirror images, with the left half of the graph being the mirror image of the right half of the graph. In the histogram and dot plot, this shape is referred to as being a "bell shape" or a "mound". The <i>most typical</i> symmetric histogram or dot plot has the highest vertical column in the center. This shape is often referred to as being a "normal curve" (or normal distribution). Not all symmetric graphs, however, have this shape (see Symmetric U-shaped below).</p>		
<p style="text-align: center;">Symmetric (bell shaped)</p>	<p style="text-align: center;">Symmetric (bell shaped)</p>	<p style="text-align: center;">Symmetric</p>

Symmetric (U-shaped) - as mentioned above, a symmetric graph forms a mirror image of itself when reflected in its vertical center line. Unlike the previous graphs, these histograms and dot plots have more of a U shape.



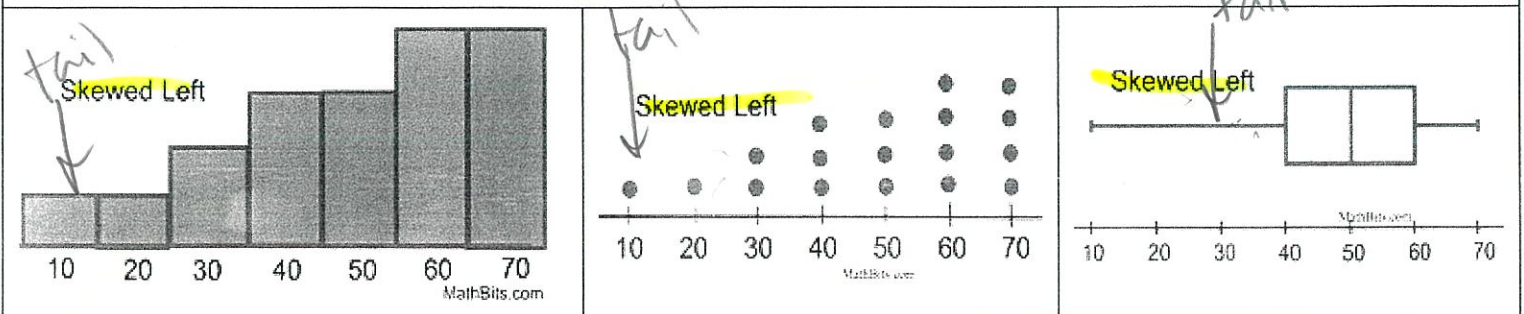
NOT symmetric (bc of outlier)

Skewed Right (positively skewed) - fewer data plots are found to the right of the graph (toward the larger numeric values). The "tail" of the graph is pulled toward higher positive numbers, or to the right. The mean typically gets pulled toward the tail, and is greater than the median. (There may be exceptions to the this last statement.)

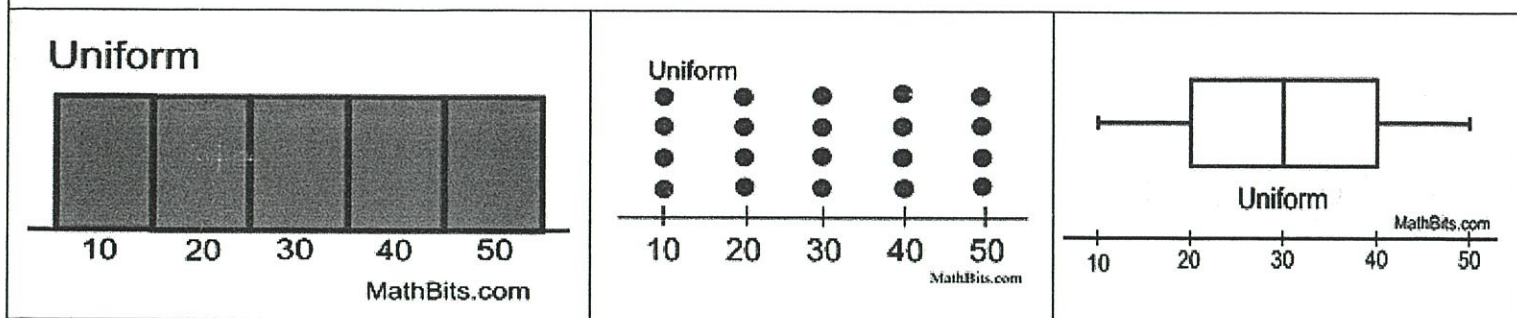


NOT symmetric (bc of outlier)

Skewed Left (negatively skewed) - fewer data plots are found to the left of the graph (toward the smaller numeric values). The "tail" of the graph is pulled toward the lower or negative numbers, or to the left. The mean typically gets pulled toward the tail, and is less than the median. (There may be exceptions to the this last statement.)



Uniform - The data is spread equally across the range. There are no clear peaks in these graphs, since each data entry appears the same number of times in the set. Notice in the boxplot how each section is of equal length: min to Q_1 , Q_1 to median, median to Q_3 , and Q_3 to max. These graphs are also symmetric.



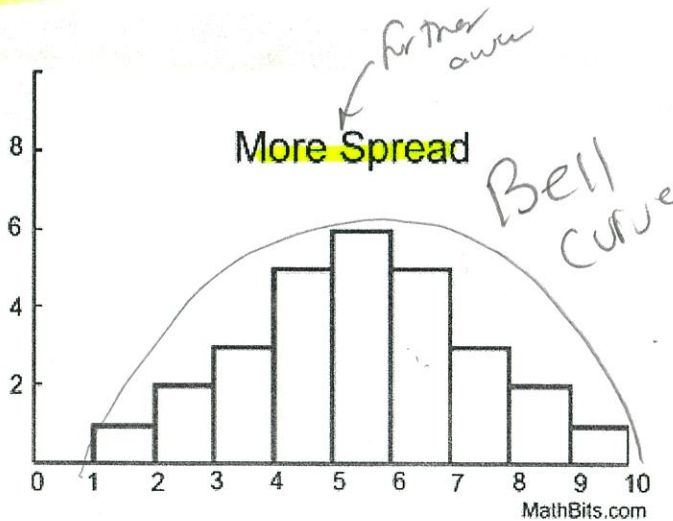
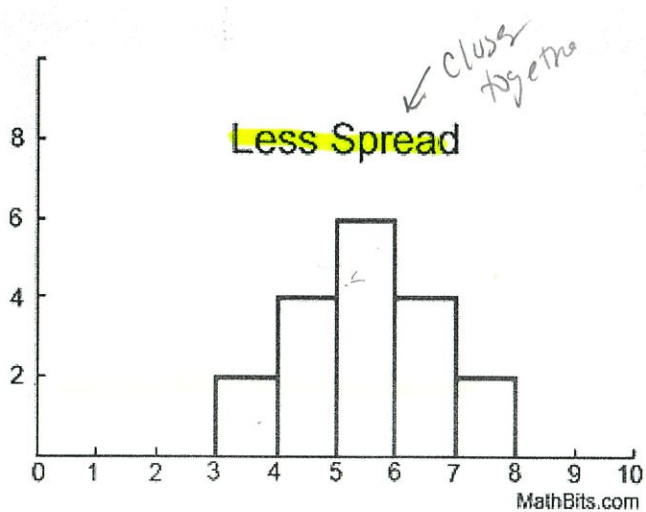
Asymmetric - Neither symmetric, nor skewed
Measures of Spread

Definition:

A measure of spread (variability, dispersion, scatter) refers to how the data within the set is "spread out" (or "dispersed", or "scattered") about the mean

If the data is clustered around the center value, the "spread" is small.

The further the distances of the data values from the center value, the greater the "spread".



Measures of Spread

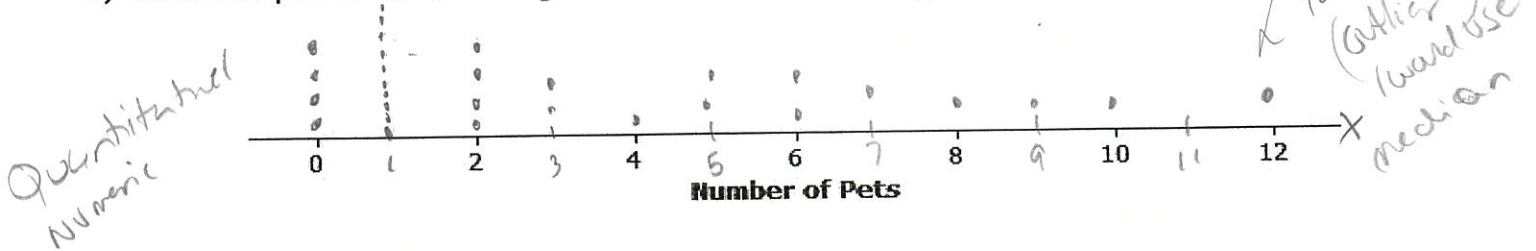
May also be called: Measures of Variability, Measures of Dispersion, or Measures of Scatter

Example

1. Students from River City High School were randomly selected and asked, "How many pets do you currently own?" The results are recorded below:

0	0	0	0	1	1	1	1	1	1	1	1	1	1	2
2	2	2	3	3	4	5	5	6	6	7	8	9	10	12

a) Make dot plot of the following data set. Use the following scales:



b) Describe the distribution of the number of pets.

Unimodal, skewed right

c) Calculate the mean number of pets owned by the 30 students from River City High School.

$$\bar{x} = \frac{\sum x}{n} \text{ sum} \rightarrow \frac{96}{30} = 3.2$$

d) Calculate the median number of pets owned by the thirty students.

$$\frac{30}{2} = 15$$

e) What do you think is a typical number of pets for students from River City High School? Explain how you made your estimate.

mode 1

mode
B/c pets are non-numeric use the mode & the mode is 1!

- Important Facts For This Lesson:
1. A Dot plot provides a graphical representation of a data distribution, helping us to visualize the distribution.
 2. The mean and the median of the distribution are numerical summaries of the center of a data distribution.
 3. When the distribution is nearly symmetrical → (NO outlier), the mean and the median of the distribution are approximately equal.
 4. When the distribution is not symmetrical / skewed → (outlier) affects, the mean and the median are not the same.
 5. For symmetrical distributions, the mean/range is an appropriate choice for describing a typical value for the distribution.
 6. For skewed data distributions, the median/IQR is a better description of a typical value.

outlier