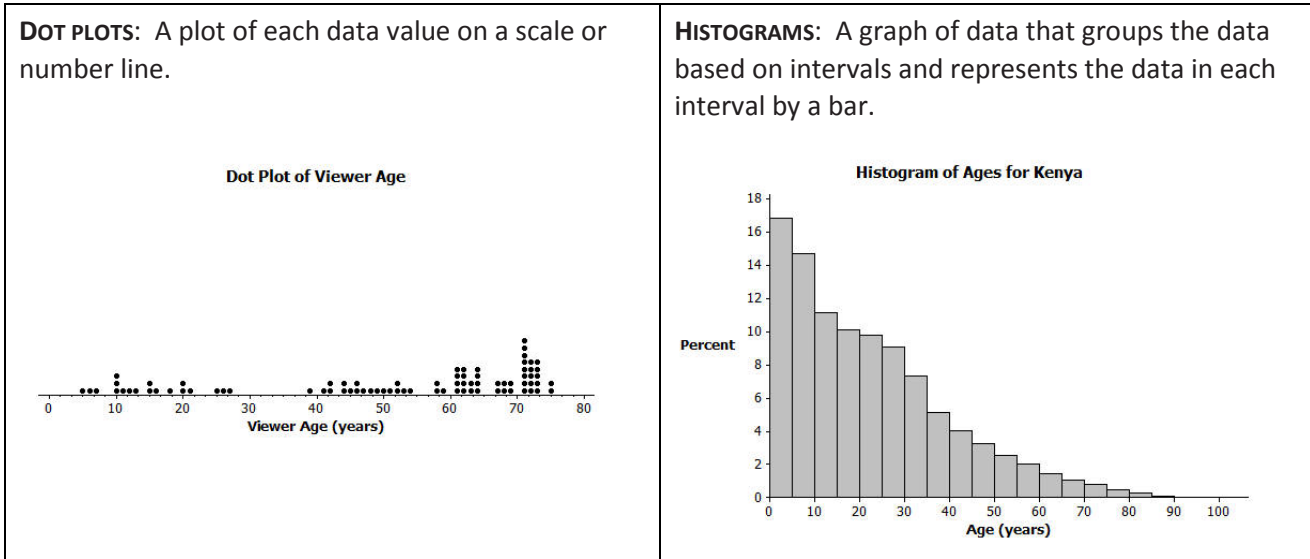


Module 4, Unit 9 Lesson 8 – Distributions: Shape, Center, and Spread

Distributions - Data are often summarized by graphs. We often refer to the group of data presented in the graph as a **distribution**. Below are examples of the two types of graphs we will be using most.



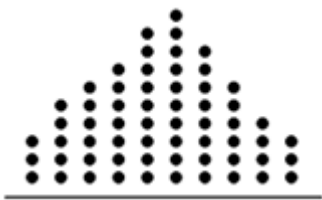
Once we have the graph of a distribution, we often want to describe the graph. We generally describe a distribution by commenting on its **shape, center and spread**.

To describe the **shape**, we will generally use *symmetric, skewed left, or skewed right*

Symmetric

Skewed ~~left~~ **right**

Skewed ~~right~~ **Left**



The mean and median are both near the middle



The mean is "pulled" to the right of the median

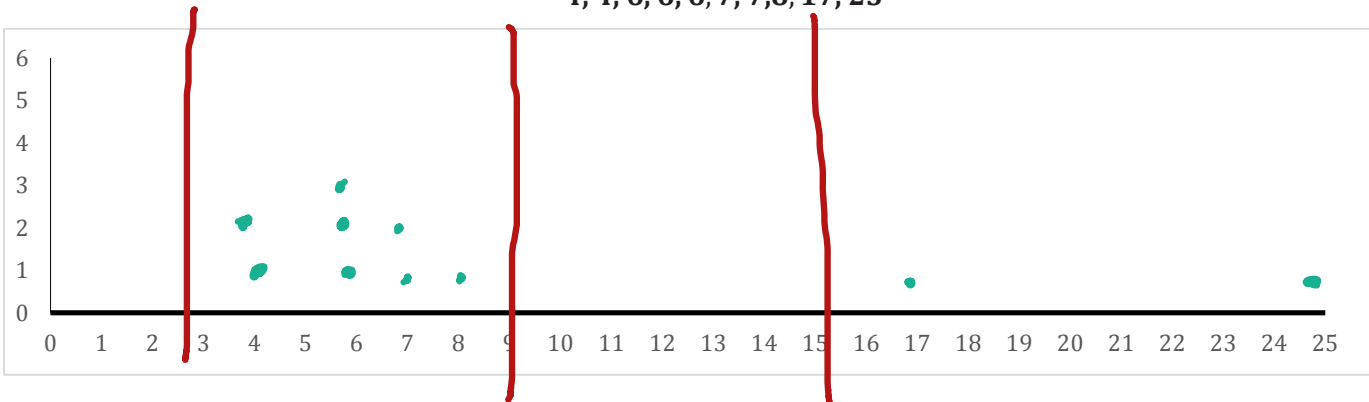


The mean is "pulled" to the left of the median

Example 1: Use the data below to create a dot plot. The comment on its shape.

Ten people were attending an event. The ages of the people are as follows:

4, 4, 6, 6, 6, 7, 7, 8, 17, 25



To describe the **center**, we calculate (or approximate) the *mean* or *median*. Calculate the mean and median of the data.

<p>Mean:</p> $\frac{2(4) + 3(6) + 2(7) + 8 + 17 + 25}{10} = \frac{90}{10} = 9$	<p>Median:</p> $\frac{6 + 7}{2} = 6.5$
---	---

How do the mean and the median in our data set compare?

To describe the **spread**, we will calculate (or approximate) the *standard deviation* or *range*.

Range is the difference between the largest and smallest values in the data set. So range = max - min
 In this data set range = $25 - 4 = 21$.

But wait, what is a standard deviation?

The standard deviation of a set of numbers tells us:
 the distance between a _____ number in set and the mean.

It will always be a positive number. How it is calculated varies slightly, depending on whether the standard deviation is describing numbers from a population or numbers from a sample. Because most instances of standard deviation describe samples, we will focus on the sample formula.

	Let's calculate the standard deviation of our data.											
<p>Standard Deviation of a Sample Step 1: Calculate the mean of the sample</p> <p>Step 2: For each number in the data set, calculate its deviation from the mean.</p> <p>Step 3: Square each of the deviations in Step 2.</p> <p>Step 4: Take the sum of the values in Step 3, and divide by the sample size minus 1.</p> <p>Step 5: Take the square root</p>	Data	4	4	6	6	6	7	7	8	17	25	Mean
												9
	2.	5	5	3	3	3	2	2	1	8	16	0
	3.	25	25	9	9	9	4	4	1	64	256	0
	4.	40.6					5.	$\sigma_x = 6.37$				

This math is a lot of work, so we will sometimes call upon Desmos to calculate it for us.

Interpretation: A typical age in this group is 6.37 years away from the mean.

Back on the dot plot, mark the mean and the ages 1 standard deviation away in each direction.

Example 2:

A local baseball club, the Manatees, has twelve players. The batting averages for those players are as follows :0.255, 0.275, 0.275, 0.275, 0.280, 0.280, 0.280, 0.280, 0.285, 0.285, 0.285, 0.290, 0.295

a) Create a dot plot for the batting averages:



b) Describe the shape of the dot plot

c) Find the mean and median of the data set. Which is larger?

$$\bar{x} = 0.280$$

$$\text{median} = 0.280$$

d) Find the standard deviation of the batting averages using Desmos.

$$\sigma_x = 0.009$$

Interpret:

Example 3:

Another local team is the Snappers. The batting averages for those players are as follows: 0.270, 0.275, 0.275, 0.280, 0.280, 0.280, 0.285, 0.285, 0.285, 0.290, 0.290, 0.295.

a) Create a dot plot for the batting averages:



b) Describe the shape of the dot plot

Symmetric

d) Find the standard deviation of the batting averages using Desmos.

$$0.007$$

Interpret:

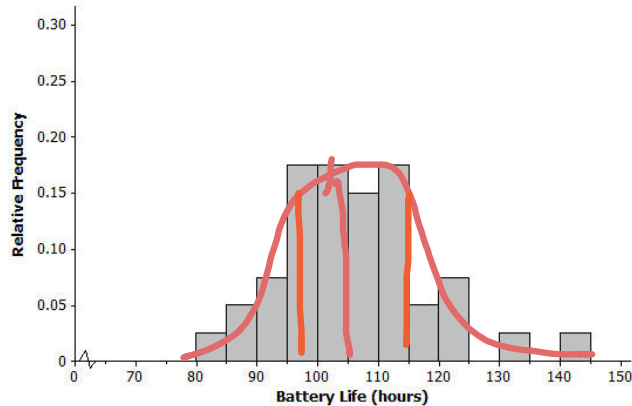
e) Out of the 2 teams, which one has the more consistent batting averages? How do you know?

Snappers

Example 3:

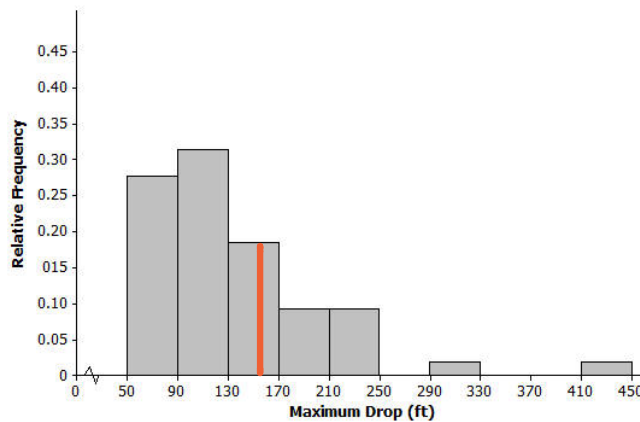
Have you ever noticed how sometimes batteries seem to last a long time, and other times the batteries seem to last only a short time?

The histogram below shows the distribution of battery life (hours) for a sample of 40 batteries of the same brand.



1. How would you describe its shape? *symmetric*
2. Is the mean of the battery life distribution closest to 95, 105, or 115 hours? *105*
3. Consider 5, 10, or 25 hours as an estimate of the standard deviation for the battery life distribution. Which seems to be the most reasonable description of a typical distance from the mean? Explain your answer. *10*

The histogram below shows the distribution of the greatest drop (in feet) for 55 major roller coasters in the United States.

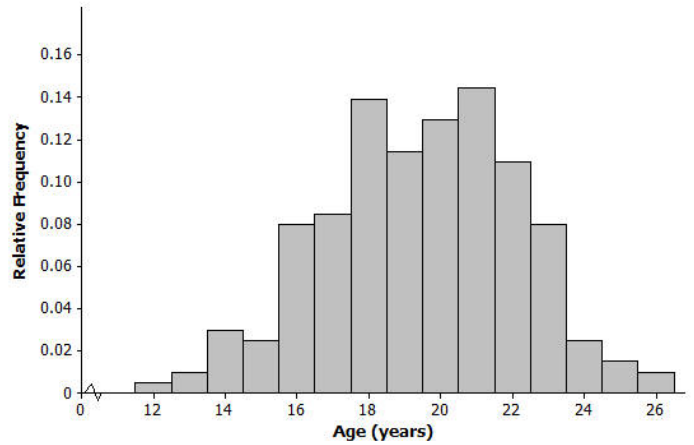


4. Would you describe this distribution of roller coaster maximum drop as approximately symmetric or as skewed? Explain your answer. *skewed right*
5. Is the mean of the maximum drop distribution closer to 90, 135, or 240 feet? Explain your answer. *135*
6. Is the standard deviation of the maximum drop distribution closer to 40, 70, or 100 ~~hours~~ feet? Explain your answer. *70 or 100*

Practice

A local utility company wanted to gather data on the age of air conditioners that people have in their homes. The company took a random sample of 200 residents of a large city and asked if the residents had an air conditioner, and if they did, how old it was. Below is the distribution in the reported ages of the air conditioners.

1. Describe the shape.
2. Is the mean of the age distribution closer to 15, 20, or 25 years? Explain your answer.
3. Is the standard deviation of the age distribution closer to 3, 6, or 9 years? Explain your answer.

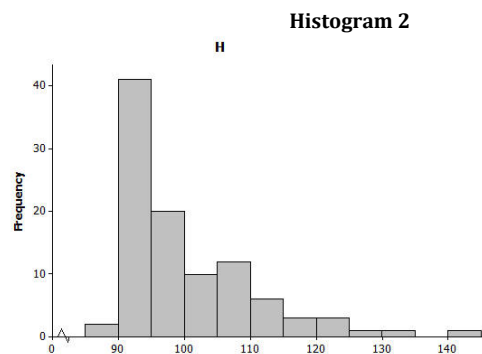
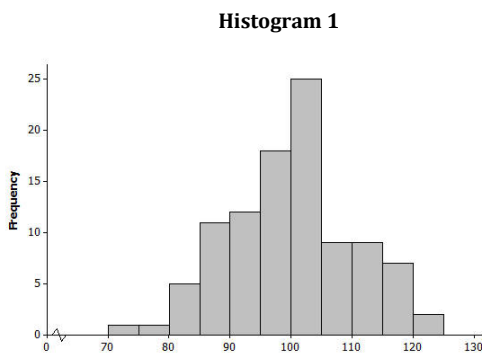


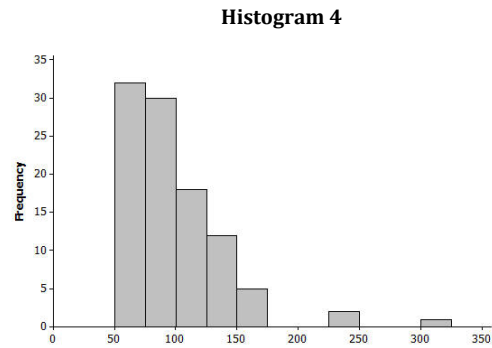
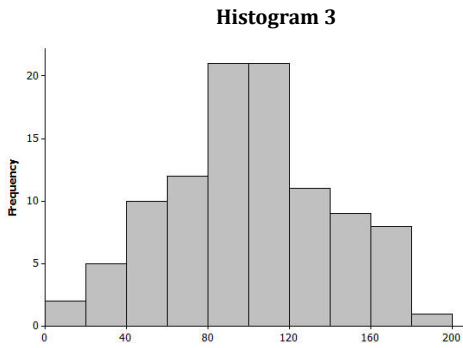
Consider the following histograms. Descriptions of the four distributions are also given. Match the description of a

Distribution	Shape	Mean	Standard Deviation
A	Skewed to the right	100	10
B	Approximately symmetric,	100	10
C	Approximately symmetric,	100	40
D	Skewed to the right	100	40

Histogram	Distribution
1	
2	
3	
4	

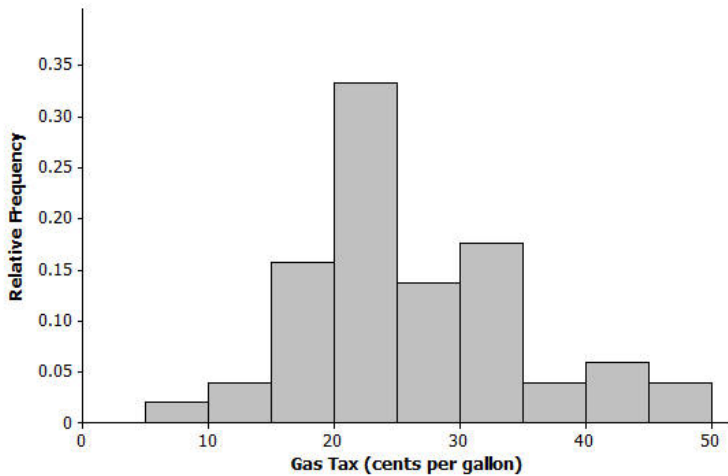
distribution with the appropriate histogram.





Example 4:

- The histogram below shows the distribution of gasoline tax per gallon for the 50 states and the District of Columbia in 2010. Describe the shape, center, and spread of this distribution. Interpret the standard deviation in context.



- The histogram below shows the distribution of the number of automobile accidents per year for every 1,000 people in different occupations. Describe the shape, center, and spread of this distribution. Interpret the standard deviation on context.

