

Day 3: Distribution of Data and their Shapes

Vocabulary

Variability:	A measurement of how different your data are
Dot Plots:	A plot of each <u>data point</u> on a scale or <u>number line</u> .
Histograms:	A graph of data that groups the data based on <u>bins</u> and represents the data in each interval by a <u>rectangle</u>
Box Plots:	A graph that provides a picture of the data ordered and divided into <u>quarters</u> that each contains approximately <u>25%</u> of the data.
Distribution:	The spread of your data
Symmetric:	Same measure (same on both sides)

Data are often summarized by graphs; the graphs are the first indicator of **variability** in the data.

Types of Graphs:

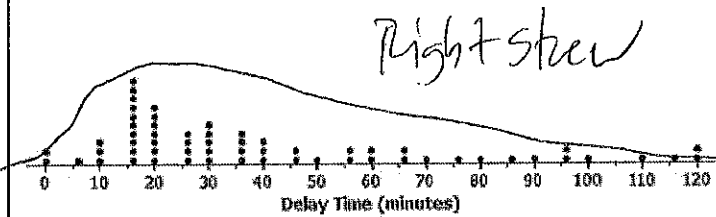
<p style="text-align: center;">Dot Plot of Viewer Age</p> <p style="text-align: center; font-size: 2em;">Left skewed</p>	<p style="text-align: center;">Histogram of Ages for Kenya</p> <p style="text-align: center; font-size: 2em;">Right skewed</p>
<p style="text-align: center;">Boxplot of Number of Pets</p> <p style="text-align: center; font-size: 2em;">Right skewed</p>	<ol style="list-style-type: none"> 1. What can you do with a random data set to help you easily create a dot plot? Organize #s, draw # line, put one dot for each data point. 2. What do you need to do with a data set to create a histogram? Organize #s, break down into bins (intervals), count the # of data in each bin. 3. What do you need to do with a data set to create a box plot? Organize #s, find median, find Q₁ & Q₃, draw box & whiskers

Analysis of Graphs:

Transportation officials collect data on flight delays (the number of minutes past the scheduled departure time that a flight takes off).

Consider the dot plot of the delay times for 60 BigAir flights during December 2012.

Dot Plot of December Delay Times



4. What can you infer about this airline and flight delays?

Most flights are delayed for 30 min or less, but some have really long delays

5. Why does this graph provide important information? Who might be interested in this graph that shows the data distribution?

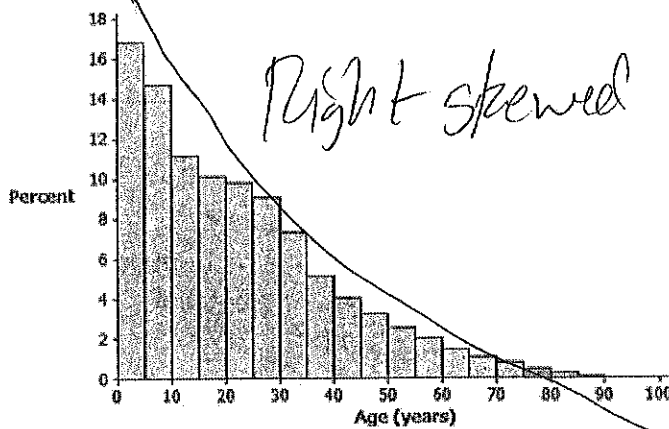
People who are flying want to know how long they expect to wait. Or airline executives

6. Based on your previous work with dot plots, would you describe this dot plot as representing a symmetric or a skewed data distribution? (Recall that a skewed data distribution is not mound shaped.) Explain your answer.

Skewed right — bump on left, tail on right.

The following histogram represents the age distribution of the population of Kenya in 2010.

Histogram of Ages for Kenya



7. What do you think this graph is telling us about the population of Kenya?

Most Kenyans are young than are old.

8. Why might we want to study the data represented by this graph?

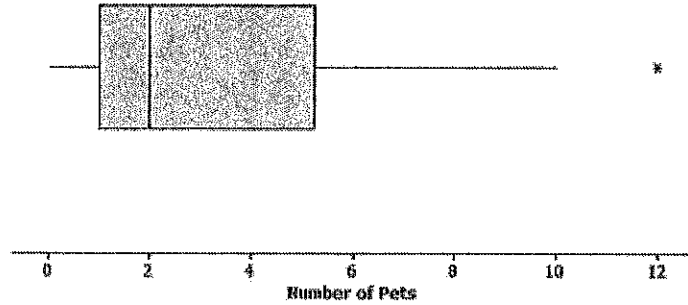
Maybe to plan health care? Retirement? Education?

9. Based on your previous work with histograms, would you describe this histogram as representing a symmetrical or a skewed distribution? Explain.

Skewed right.

30 students from River High School were asked how many pets they owned. The following box plot was prepared from their answers.

Boxplot of Number of Pets

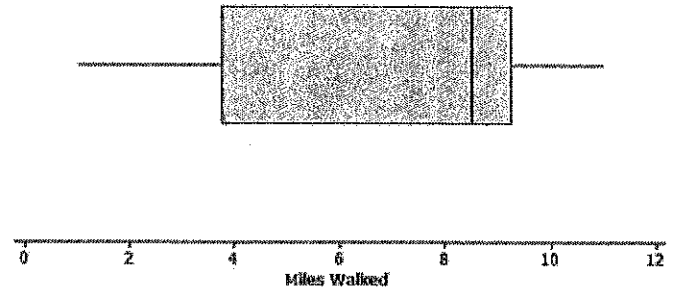


10. What does the box plot tell us about the number of pets owned by the 30 students at River HS?

$\frac{1}{2}$ own 2 or fewer pets
 $\frac{1}{2}$ own 2 or more pets

22 juniors from River High School participated in a walkathon to raise money. The following box plot was constructed using the number of miles walked by each of the 22 juniors.

Boxplot of Miles Walked for Juniors



11. What do you think the box plot tells us about the number of miles walked by the 22 juniors?

$\frac{1}{2}$ walked more than 8 miles
 $\frac{3}{4}$ walked more than 4 miles

12. Would you describe each of the box plots above as being symmetrical or skewed. Explain.

Pets is skewed right. Miles is skewed left

13) 25 people were attending an event.

The ages of the people are as follows:

[3, 3, 4, 4, 4, 4]⁶, [5, 6]¹⁴, [7, 7, 7, 7]⁴, [16, 17, 22, 22, 25]⁵

a. Create a histogram of the ages here →

b. Would you describe your graph as symmetrical or skewed? Explain your choice.

Right skewed

c. Identify a typical age of the 25 people.

Mean = 8.52, Median = 6

d. What event do you think the 25 people were attending? Use your histogram to justify your conjecture.

Little kids soccer camp.

Histogram for Event 1

