# Making Histograms

**Histograms** are like bar graphs, but the bars are drawn so they touch each other. Histograms are used only with numerical data.

**Example.** These are prices of hair dryers in three stores (in dollars). Make a histogram.

14  15  19  20  20  20  21  24  25  34  34  35  35  37  42  45  55

We need to decide how many bins to make and how "wide" they are. For that, we first calculate the **range**, or the difference between the greatest and smallest data item. It is $55 - 14 = 41$. Then we divide the range into equal parts (bins) to get the *approximate* bin width.

If we make five bins, we get $41 \div 5 = 8.2$ for the bin width. The bins would be 8.2 units apart. However, in this case it is nice to have the bins go by whole numbers, so we round 8.2 up to 9 and use 9 for the bin width.

The important part is that *each data item needs to be in one of the bins*. You may have to try out slightly different bin widths and starting points to see how it works. This time, starting the first bin at 13 makes the last bin to end at 57, which works, because the data will "fit" into the bins. (Starting at 14 would work, too.)

The **frequency** describes *how many data items fall into that bin*. Lastly, all we need is to draw the histogram, remembering that the bars touch each other.

| Price ($) | Frequency |
|-----------|-----------|
| 13..21    | 7         |
| 22..30    | 2         |
| 31..39    | 5         |
| 40..48    | 2         |
| 49..57    | 1         |
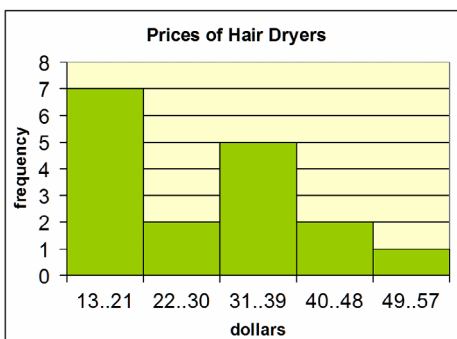


This is a **double-peaked** distribution and **skewed to the right** (the direction of skewness is where the "long tail" of the distribution is, in this case to the right). Since it definitely is *not* bell-shaped, the mean is *not* a good measure of center. Therefore, the *median* is the better choice for measure of center.

Consequently, we need to use *interquartile range*, and not mean absolute deviation, as a measure of variation.

The median is underlined below:

14  15  19  20  20  20  21  24  **25**  34  34  35  35  37  42  45  55

The 1st quartile is $20 and the 3rd quartile is $36 (verify those). So the interquartile range is $16.

This means that half of the data is found between 20 and 36 dollars. This price range is quite large for devices with a median price of only $25. A large range compared to the median describes data that is widely scattered. (We can also see that from the dot plot.)

Since the median is $25, which is nearer the low end of the interval from $20 to $36, the prices are somewhat more concentrated in the lower end of that interval.
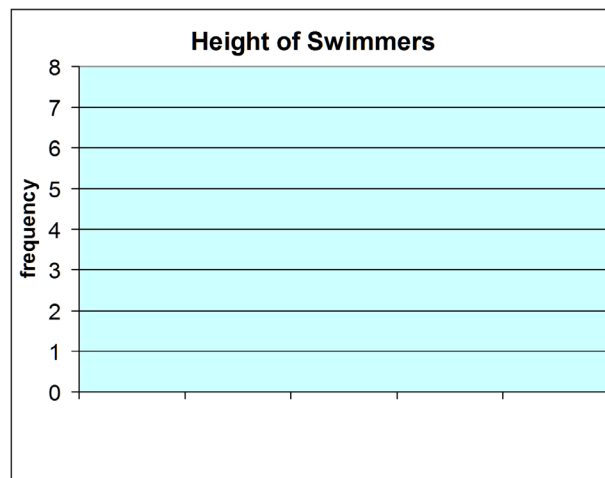
For a comparison, look at the dot plot as well. It has a similar shape to the histogram.
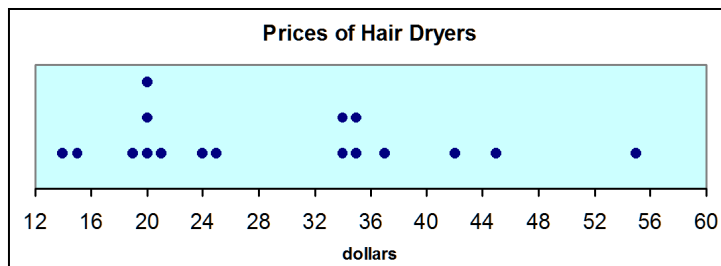
1. This data lists the heights of 24 swimmers in centimeters. Make a histogram with <u>five</u> bins.

155  155  156  157  158  159  159  160  162  162  163  163
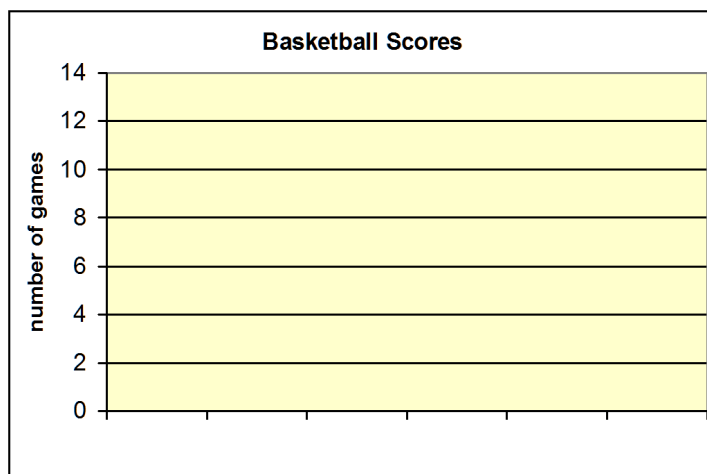164  165  166  167  167  168  168  170  172  174  175  177

| Height (cm) | Frequency |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Height of Swimmers**

2. Make a histogram from this data, which lists all the scores a basketball team had in the games in one season. Make six bins.

60  62  68  71  72  72  73  74  74  74  75  75  76  77  77  77
78  78  78  79  79  81  81  82  83  83  85  86  88  90  92  95

| Score | Frequency |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Basketball Scores**

3. **a.** As this distribution in problem 2 has its peak near the center, you could use either mean or median as a measure of center. This time, find the median and the interquartile range.

median _____     interquartile range _____

**b.** Describe the variation in the data. Is it very scattered (varied), somewhat so, or not very much so?

4. This data is the finishing times of a group of athletes for the 100-meter dash. Make two histograms, one with four bins and the other with five, and compare them.

Notice the differences in how they look. Both show a peak, but the one with more bins shows more detail. The look of a histogram changes depending on the number of bins and the starting point of the first bin. For this reason, if the data set is small, a dot plot or a stem-and-leaf plot may work better than a histogram.

**a.** Make a histogram with 4 bins.

11.8  12.0  12.1  12.1  12.3  12.4  12.5  12.5  12.6  12.6  12.7
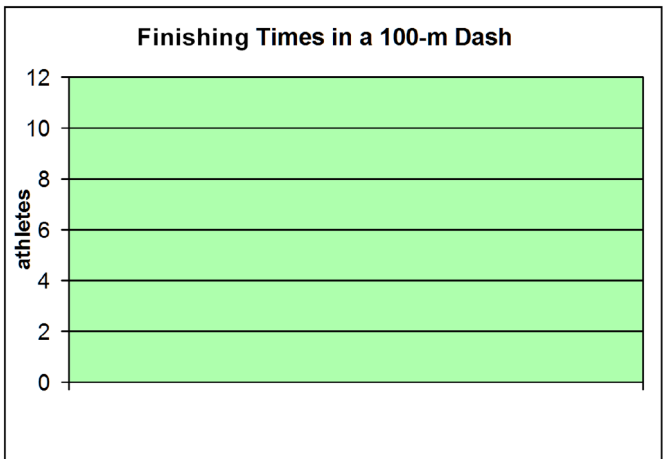12.7  12.7  12.7  12.7  12.8  12.8  12.8  12.8  12.9  12.9  13.1

| Finishing times (seconds) | Frequency |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |

**Finishing Times in a 100-m Dash**
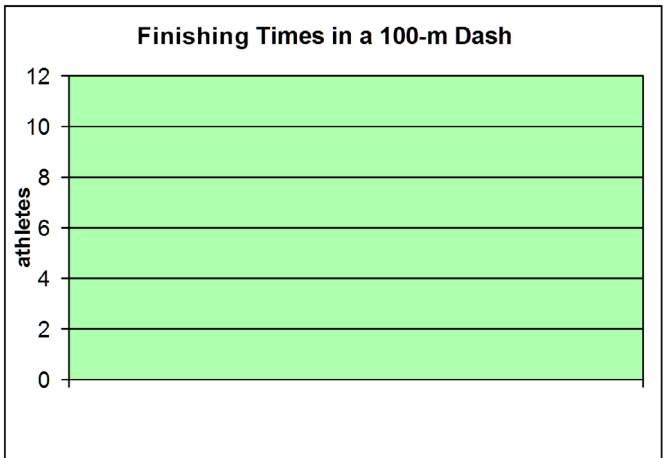
**b.** Make a histogram with 5 bins (from the same data).

11.8  12.0  12.1  12.1  12.3  12.4  12.5  12.5  12.6  12.6  12.7
12.7  12.7  12.7  12.7  12.8  12.8  12.8  12.8  12.9  12.9  13.1

| Finishing times (seconds) | Frequency |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Finishing Times in a 100-m Dash**

**Puzzle Corner**  Find a quick, *mental math* method for calculating the mean for this data set.  102, 94, 99, 105, 96, 107, 101, 104 (the weights of a litter of kittens at birth, in grams)