Section 3.1 – Shapes of Distributions

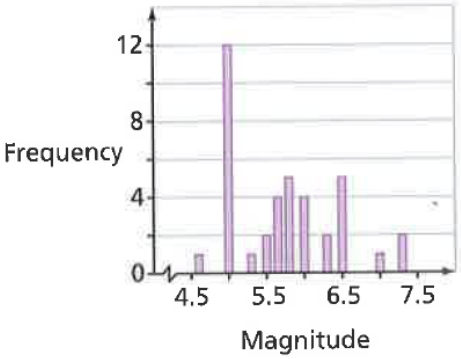
MDM4U
Jensen

Part 1: Histogram Review

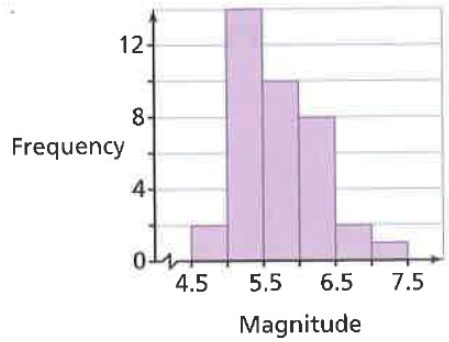
Example 1: Earthquakes are measured on a scale known as the Richter Scale. There data are a sample of earthquake magnitudes in Canada between 1960 and 1965.

5.0	5.0	6.4	5.0	6.0	5.6	6.5	6.5	5.0	5.5
6.4	7.2	5.0	5.7	5.6	5.0	5.0	5.0	5.0	5.7
5.0	7.0	5.5	5.2	4.6	6.3	7.2	6.0	5.4	5.8
6.0	5.7	6.5	5.0	5.7	5.0	5.6	6.0	5.6	6.2

What is wrong with how each of the following histograms display the above data?



The width of each bar (bin width) is too small. Too small of a bin width results in a histogram that does not effectively summarize the distribution (too many small bars). Between 5 and 10 intervals will usually produce a good bin width of the data.



Pieces of data fall on the border of two intervals. This makes it ambiguous as to where the data actually falls. Make sure to take the necessary steps when creating your intervals to make sure that this doesn't happen.

Lets Make an Effective Histogram for the Data:

a) Determine the range of the data

$$\begin{aligned} range &= highest\ value - lowest\ value \\ &= 7.2 - 4.6 \\ &= 2.6 \end{aligned}$$

b) Determine an appropriate bin (interval) width that will divide the data into 6 intervals.

$$bin\ width = \frac{range}{number\ of\ intervals}$$

$$= \frac{3.0}{6}$$

$$= 0.5$$

Note:

Round your range UP to a value that can be divided easily.

c) Determine the first value of your first interval

We added 0.4 to 2.6 when we rounded our range, therefore we should subtract 0.2 from our smallest value 4.6; which makes our starting point 4.4.

However, some data will still fall on the border of the intervals, so we should add a decimal place by subtracting .05 from our starting point.

Note:

1. If you have rounded your range up you should subtract half of the amount you rounded from the smallest value to evenly distribute the 'excess of your range'.

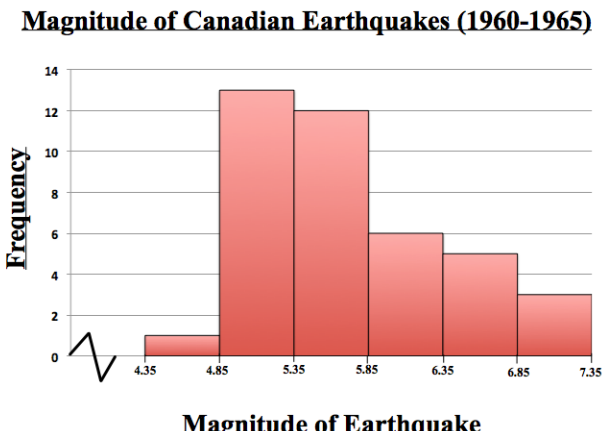
2. Make sure no data points lie on the border of two intervals. (Do this by subtracting .5 from a whole number, .05 from data with one decimal point, .005 from data with two decimal points and so on)

d) Create a frequency table using your intervals

Notice that the number one interval ends with, the next interval starts with the same number. This is because the data for a histogram is continuous!!!

Class Interval	Frequency
4.35 – 4.85	1
4.85 – 5.35	13
5.35 – 5.85	12
5.85 – 6.35	6
6.35 – 6.85	5
6.85 – 7.35	3

e) Create a histogram of the data



Part 2: The Shape of a Distribution

Step back from a histogram. What can you say about the distribution? When you describe a distribution, you should always comment about three things: its shape, center, and spread. In this lesson we will focus on shape.

The mode of a histogram is an important characteristic that is often used in describing its shape. The mode of a histogram is the interval with the highest frequency. Does the histogram have a single peak, central peak, or several separated peaks? These peaks are called modes.

The shape of a distribution is generally described in one of four ways:

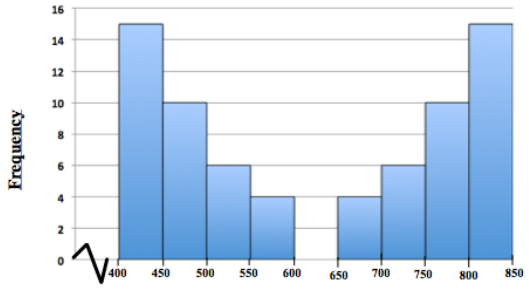
1. U-shaped
2. Uniform
3. Mound-shaped
4. Skewed

Symmetrical

Note: *A graph is roughly symmetric if the right and left sides of the graph are approximately mirror images of each other.*

1. U-Shaped Distribution

The scores from the game of spider solitaire form this type of distribution.



Score	400–449	450–499	500–549	550–599	600–649	650–699	700–749	750–799	800–849
Frequency	15	10	6	4	0	4	6	10	15

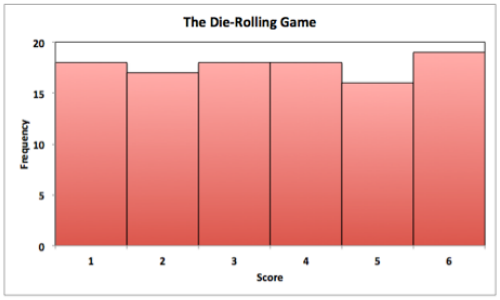
- A U-shaped distribution occurs when there are peaks at either end of the range
- Because it has two peaks, it can also be described as a bimodal distribution

Can you think of another example of a frequency distribution that would be U-shaped (bimodal)?

Marks on a test where half the class just "didn't get it".

2. Uniform Distribution

This is the distribution you would expect from an experiment such as rolling a single die.



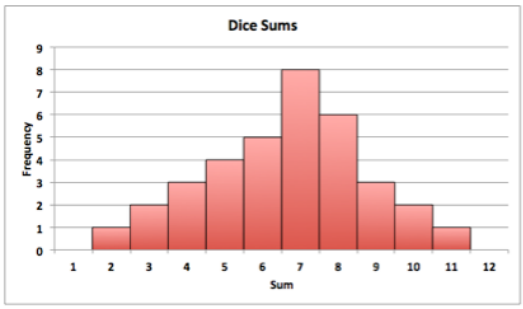
Score	1	2	3	4	5	6
Frequency	18	17	18	18	16	19

- When each outcome has a similar frequency, it is called a uniform distribution. The height of each bar is roughly equal.

- Notice, there doesn't appear to be any one single mode.

3. Mound Shaped Distribution

Rolling a pair of dice and recording the sum results in this type of distribution.



Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	8	6	3	2	1	0

- In this distribution, there is an interval with the greatest frequency in the middle, and the frequencies of all other intervals decrease on either side of that

- The frequency distribution takes on a mound (or bell) shape. It can also be described as unimodal since it has one clear peak (mode).

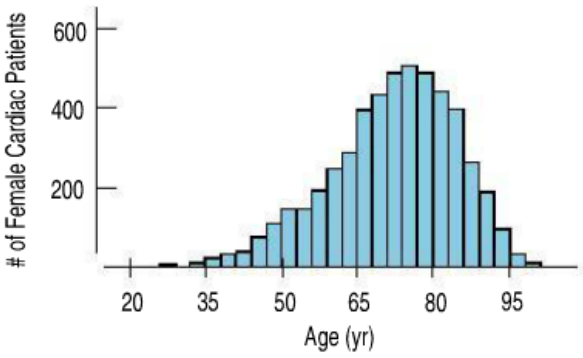
Do you notice any similarities between the first 3 shapes of distributions?

They are all roughly symmetric

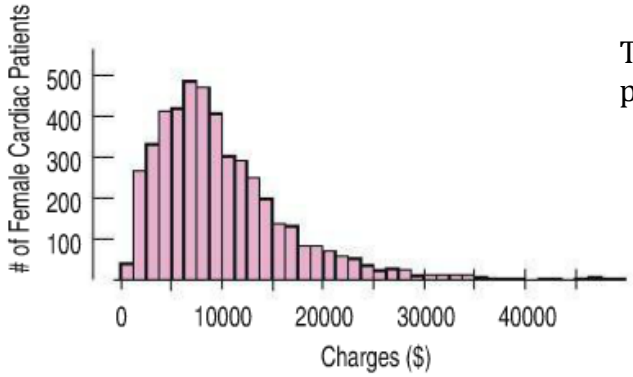
4. Skewed

The thinner ends of a distribution are called the tails. If one tail stretches out farther than the other, the histogram is said to be skewed to the side of the longer tail.

Another way to say it is that the interval or group of intervals with the highest frequencies are near one end of the histogram. As a result, the distribution seems to tail off to the left or right.



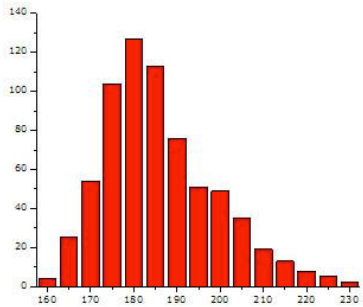
This distribution of ages of female heart attack patients is left skewed.



This distribution of cost of treatment for heart attack patients is right skewed.

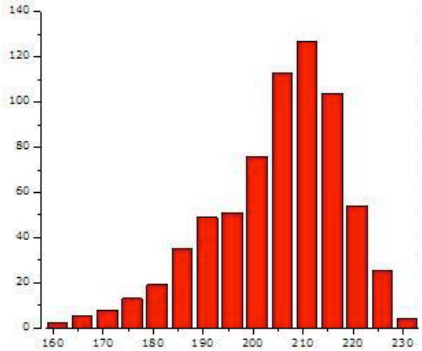
Tip: *If you get mixed up between left and right skewed.....look at your toes!*

Why call it left or right skewed?



Right Skewed

The mean is skewed (pulled) to the right of the mode



Left Skewed

The mean is skewed (pulled) to the left of the mode