

Before you begin, please provide your name → Solutions

MDM 4U

3.1 Graphical Displays of Information

1. Read from p141 and up to Example 1 on page 142, and then answer the following in-line with the text already provided. To distinguish your answers from the problems provided, please try to incorporate the use of either **bold-face font** or a different **colour of font**.

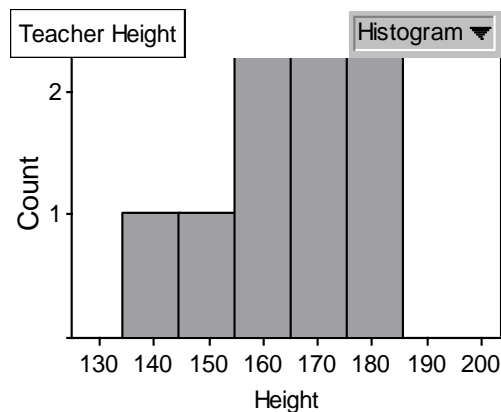
- a. What are *histograms* and why are they used?

A histogram is a visual display of data that has been organized into equal class intervals, which appear on the horizontal axis. The heights of the rectangles are proportional to the frequencies they represent. Histograms are used to show continuous data, such as earthquake magnitudes in a certain area over a certain period of time.

- b. What are *continuous data*? Provide an example of a quantity that requires measurement using continuous values.

Continuous data are data that are only measurable with real numbers. A measure of quantity will always be continuous. An example of a quantity that requires measurement using continuous values would be the heights of a group of students or the length of time a plant takes to germinate.

- c. On a histogram, like the one below, the width of each bar is called the *bin width* or *interval width*. When creating histograms, the width of each interval must be equal.



- d. What is the bin width in the histogram depicting a distribution of teacher height?
The bin width in the histogram depicting a distribution of teacher height is 10 cm.
- e. In choosing the bin width, an attempt must be made to choose one that shows a representative display of the data.

Describe what would happen to the histogram if you changed the bin width under the following restrictions:

- i. Using a small bin width?

If you were to use a smaller bin width, the histogram would not summarize the distribution of data effectively, as the data would be too spread out to make

any conclusions. Furthermore, you might run into the risk of having zero-class intervals, as you would have many short, thin bars.

ii. Using a large bin width?

If a larger bin width was used, the data wouldn't be grouped in to enough categories, and therefore you would not be able to see enough distinction between each of the categories, making it hard to find averages and trends.

f. What is the least number of intervals that is to be used when constructing histograms?
The least number of intervals that is to be used when constructing histograms is 5.

[Note: The greatest number of intervals that can be considered is 20, but recall that the goal of creating a histogram is to produce a visual that is representative of the data collected.]

g. In what situation does a frequency distribution no longer look like a histogram?
When the bin width is zero, no bar will appear and it not longer looks like a histogram.

2. Read Example 1, from page 142, and then answer the problems that follow. Please make sure that I can adequately distinguish your responses from the problems provided.

a. Write the formula that determines bin width. **Bin width = range / number of intervals**

Another rule is for the number of intervals to be approximately equal to the square root of the number of data values. For example, with 100 data values, you would use 10 intervals (i.e., square root(100) = 10).

b. Define the *range* of a data set.

The range of a data set is the difference between the maximum value and the minimum value.

c. In Example 1, why was a bin width of 0.5 used?

A bin width of 0.5 was used in Example 1 because the range (3.0) divided by the number of intervals (6.0) is 0.5, and that is the formula for finding the bin width.

d. In Example 1, how did the authors ensure that no piece of data would lie on the borderline between two intervals?

To ensure that no piece of data would lie on the borderline between two intervals, the authors used numbers that were not in the table as endpoints. They also ensured that data couldn't fall in two places by choosing intervals that did not overlap.

[Note: Whenever reasonably possible, create boundary values for intervals such that no values lie on them. If on the rare occasion this cannot be avoided, these data get included in the right-most, of the two, intervals.]

3. Using the knowledge you've gained from #1 and #2, complete the following problem.

Problem

Consider the following exam scores for a Biology class:

68, 60, 87, 77, 73, 43, 91, 57, 61, 66, 44, 90, 52, 58, 43, 58, 71, 76, 79, 78, 55, 94, 80, 75, 81, 54, 49

- a. Determine the *range* of the following data.

$$\begin{aligned} \text{Range} &= 94-43 \\ &= 51 \end{aligned}$$

- b. Determine the bin width, for this data set,

i. using 5 intervals.
$$\begin{aligned} \text{Bin width} &= 50/5 \\ &= 10 \end{aligned}$$

ii. using 20 intervals.
$$\begin{aligned} \text{Bin width} &= 50/20 \\ &= 2.5 \end{aligned}$$

iii. using 10 intervals.
$$\begin{aligned} \text{Bin width} &= 50/10 \\ &= 5 \end{aligned}$$

- c. Based on your calculations in b., choose a bin width that would be reasonable for this data set.

A bin width that would be reasonable for this data set would be 5.

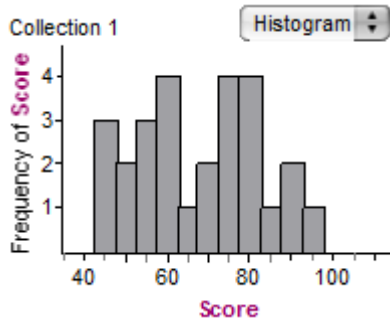
- d. Let's propose that an interval width of 5 is appropriate for constructing a histogram of the data. By choosing a starting value of 42.5, for the first interval, no data will fall on the borderlines of any of the intervals. Complete the following *frequency table* for the Biology marks provided.

Intervals	Frequency
42.5 – 47.5	3
47.5 – 52.5	2
52.5 – 57.5	3
57.5 – 62.5	4
62.5 – 67.5	1
67.5 – 72.5	2
72.5 – 77.5	4
77.5 – 82.5	4
82.5 – 87.5	1
87.5 – 92.5	2
92.5 – 97.5	1

- e. The next step in the process is to create the histogram. Instead of creating this type of visual display by hand, I would like to encourage you to use technology.

[Note: Solution 3 on p143 describes how the distribution, from Example 1 (p142), was created using *Fathom*.]

Use *Fathom* to create a histogram for the distribution of Biology marks. *Copy as picture*, in *Fathom*, and then paste your graphic into the space below.



f. Compare the bin width provided in part d. with the bin width displayed in your *Fathom* graphic.

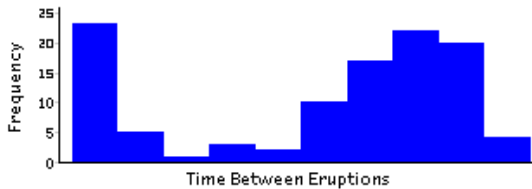
The bin widths are the same (intervals of 5 starting at 42.5) and they accurately represent the data without being too big or too small.

4. Read from “Distribution of Data” on page 143 to the end of “Key Ideas” on page 145, and then answer the problems that follow.

a. Provide another name for the distribution pictured below.

Another shape for a u-shaped distribution is a bimodal distribution.

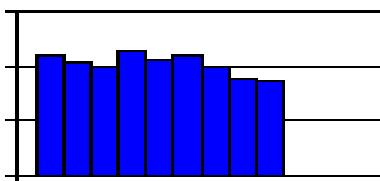
U-Shaped Distribution →



b. Describe when you would expect to see the following type of distribution.

A U-shaped distribution is seen when there are peaks at either end of the range, such as playing a card game like “spider solitaire”, where you can do either really well or poorly.

Uniform Distribution →



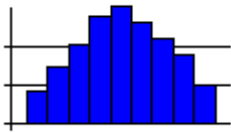
c. In what type of experiment would you expect to see a uniform distribution?

You could expect to see a uniform distribution when you are playing a game of chance (such as rolling a die) and each outcome has the same probability of occurring.

d. For the graphic provided below, describe a simulation where you would expect to see such a distribution.

A simulation where you would see a mound-shaped distribution would be rolling a pair of die and recording the sum.

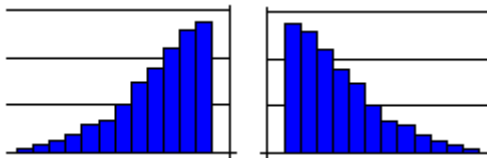
Mound-Shaped Distributions



- e. The 3 types of distributions provided on p3 are examples of *symmetric distributions*, as they nearly show symmetry about the centre of the distribution.

When the *mean* of a distribution has shifted to the right of centre, a *left-skewed* distribution results (pictured below, left); to the left of centre, a *right-skewed* distribution results (pictured below, right).

Skewed Distributions

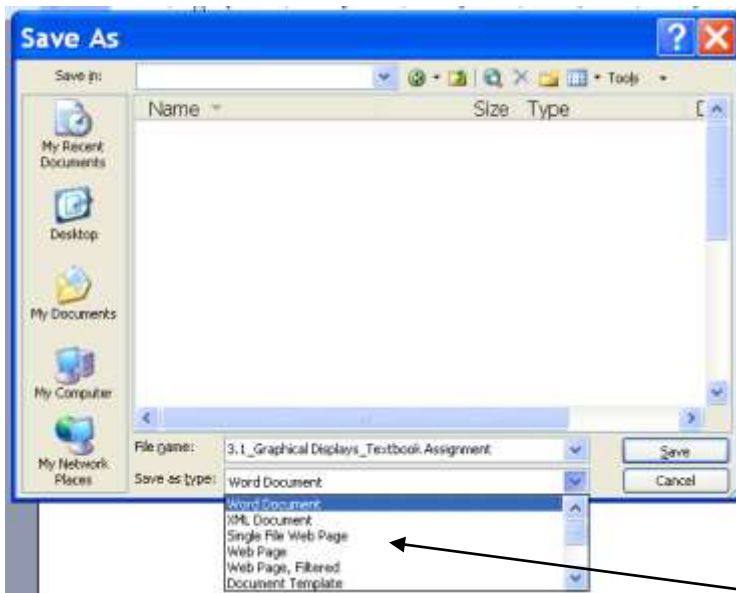


left-skewed right-skewed

- f. On page 145, a frequency table and histogram have been provided for a game of solitaire. What does the shape of this distribution tell you about the probability of getting a high score in this game of solitaire?

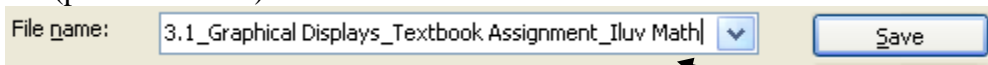
The shape of this distribution tells us that the probability of getting a high score in a game of solitaire is low because most of the games got a low score, and only a few got a high one.

5. Now that you're finished, please follow the remaining instructions very carefully such that I can arrange to complete your assessment.
- Ensure that you've typed your name at the top of page 1.
 - Save your completed work using a 1997-2003 template (pictured below).



NOTE: You should see “...1997-2003...” in the drop down menu.

- c. For the file name, please tag it by including _your name at the end of the given file name (pictured below).



Your Name

- d. Email your file to christopher.stewart@ucdsb.on.ca.
- e. If I determine that you're presenting some misconceptions regarding this section, I will provide some feedback electronically. To ensure that I have your email filed correctly, please record your email below.

My email address:

Your Homework

For consolidation and further practice, complete #1 to 3, 4ac, 5ac, 6a or c, and 8 on pp. 147-48.