Lesson: Organizing and Graphing Data (One-variable statistics)

Raw Data: the unprocessed information collected for a study

Example from Textbook Page 92:

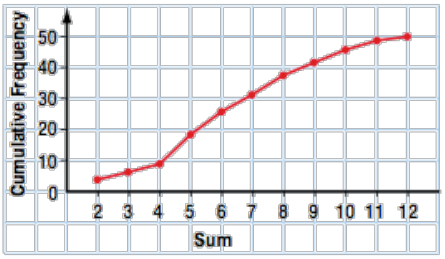
Here are the sums of the two numbers from 50 rolls of a pair of standard dice.

11	4	4	10	8	7	6	6	5	10	7	9	8	8
4	7	9	11	12	10	3	7	6	9	5	8	6	8
2	6	7	5	11	2	5	5	6	6	5	2	10	9
6	5	5	5	3	9	8	2						

Frequencies:

Sum	Tally	Frequency	**Cumulative Frequency**
2	IIII	4	4
3	II	2	4+2=6
4	III	3	6+3=9
5	𝍸 IIII	9	9+9=18
6	𝍸 III	8	18+8=26
7	𝍸	5	26+5=31
8	𝍸 I	6	31+6=37
9	𝍸	5	37+5=42
10	IIII	4	42+4=46
11	III	3	46+3=49
12	I	1	49+1=50

Always the same as the number of rolls!!



Cumulative Frequency Graph will always be INCREASING (⇧)

Relative Frequency: shows the frequency of a data group as a fraction or percent of the whole data set

Example from Textbook Page 97: Marks on data management test:

$$interval\ size = \frac{possible\ range\ of\ data}{\#\ of\ intervals}$$

e.x. $\frac{96-39}{13} = 4.38 \sim 5$ **(always round up)**

78	81	55	60	65	86	44	90
77	71	62	39	80	72	70	64
88	73	61	70	75	96	51	73
59	68	65	81	78	67		

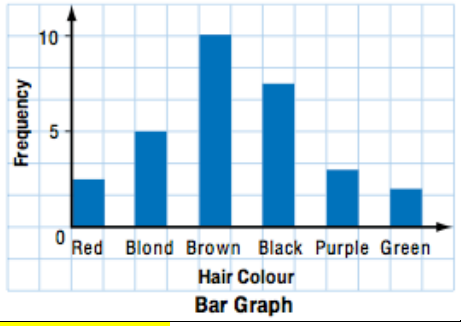
Ways to represent intervals of discrete data:
1) 34.5 – 39.5 2) 34 – 39
39.5 – 44.5 40 – 44
44.5 – 49.5 45 – 49

Ways to represent intervals of continuous data:
1) 34.5 – 39.5 2) $34.5 \leq x < 39.5$
39.5 – 44.5 $39.5 \leq x < 44.5$
44.5 – 49.5 $44.5 \leq x < 49.5$
inclusive for lower fence exclusive for upper fence.

Marks(%)	Midpoint	Tally	Frequency	Relative Frequency (round to 3 decimal places)
34.5-39.5	37	I	1	1 ÷ 30 = 0.033
39.5-44.5	42	I	1	0.033
44.5-49.5	47	-	0	0
49.5-54.5	52	I	1	0.033
54.5-59.5	57	II	2	0.067
59.5-64.5	62	IIII	4	0.133
64.5-69.5	67	IIII	4	0.133
69.5-74.5	72	𝍸 I	6	0.200
74.5-79.5	77	IIII	4	0.133
79.5-84.5	82	III	3	0.100
84.5-89.5	87	II	2	0.067
89.5-94.5	92	I	1	0.033
94.5-99.5	97	I	1	0.033
Total			30	

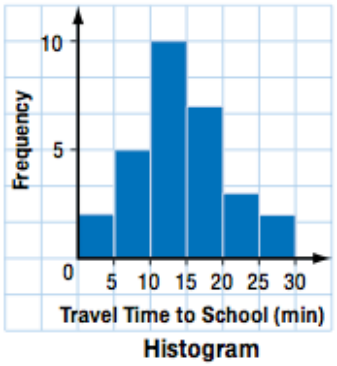
Bar Graph (Qualitative Variables)

- a chart or diagram that represents quantities with horizontal or vertical bars whose lengths are proportional to the quantities
- represents all kinds of variables that may not have set order such as hair colour or citizenship



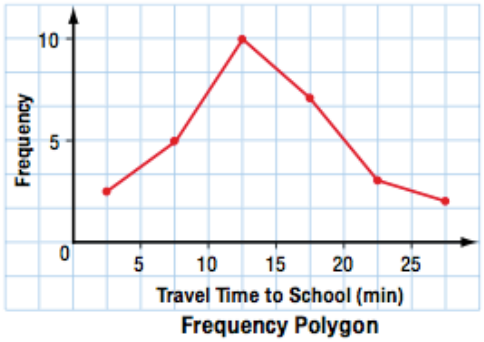
Histogram (Quantitative Variables)

- a special form of bar graph which the areas of the bars are proportional to the frequencies of the values of the variable
- used for variables whose values can be arranged in numerical order especially continuous variables such as weight, temperature

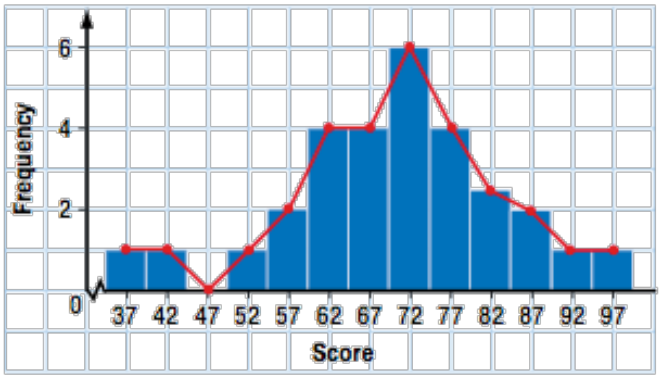


Frequency Polygon

- illustrates the same information as a histogram or bar graph by plotting frequencies versus variable values and then joining the points with straight lines



Superimposed frequency polygon and histogram



Pictograph

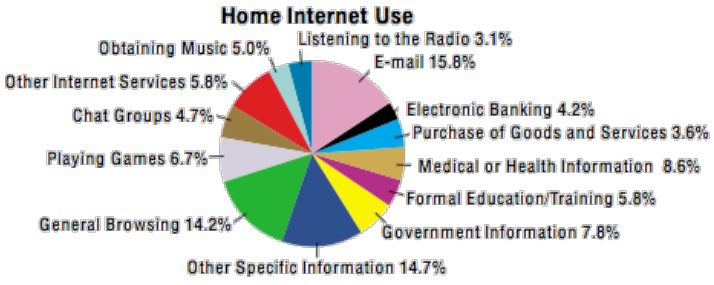
Home Internet Use

E-mail	💻💻💻💻💻💻💻💻
Electronic Banking	💻💻
Purchase of Goods and Services	💻💻
Medical or Health Information	💻💻💻💻
Formal Education/Training	💻💻💻
Government Information	💻💻💻💻
Other Specific Information	💻💻💻💻💻💻💻
General Browsing	💻💻💻💻💻💻💻
Playing Games	💻💻💻
Chat Groups	💻💻
Other Internet Services	💻💻💻
Obtaining Music	💻💻
Listening to the Radio	💻

Each 💻 represents 2% of households.

Circle/Pie Graph

Example from Textbook Page 98:

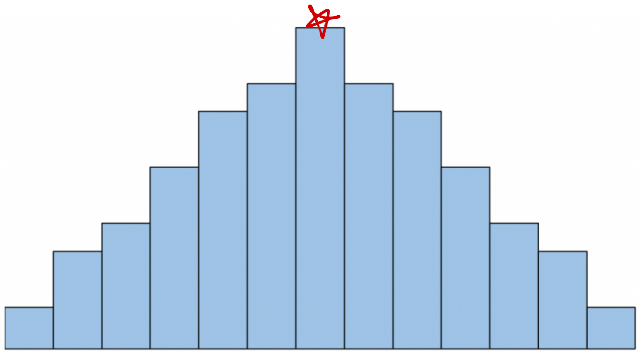


More about Histograms…

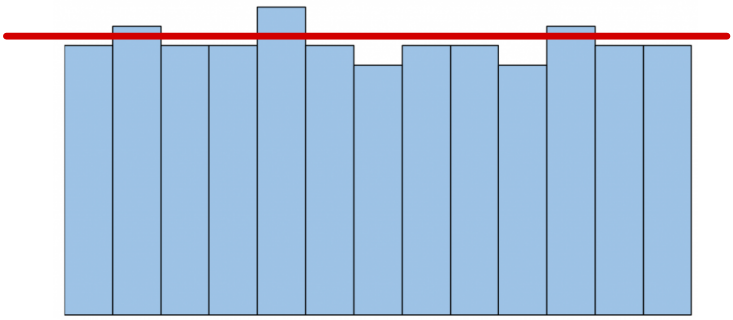
How to describe the shape of histograms:

- **x-axis** displays the **values** of the dataset
- **y-axis** displays the **frequencies** of each value

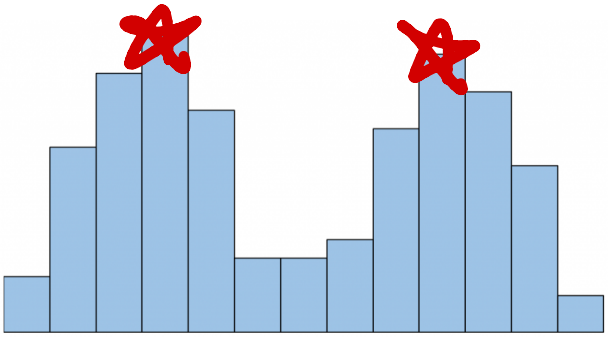
1) **Bell-Shaped**: a histogram is bell-shaped if it resembles a "bell" curve and has one single peak in the middle of the distribution. The most common real-life example of this type of distribution is the normal distribution.



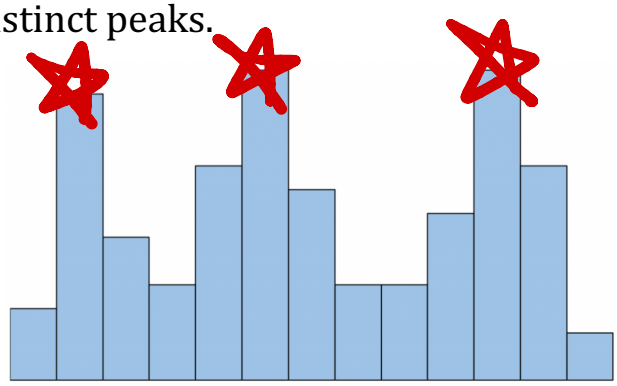
2) **Uniform**: A histogram is described as "uniform" if every value in a dataset occurs roughly the same number of times. This type of histogram often looks like a rectangle with no clear peaks.



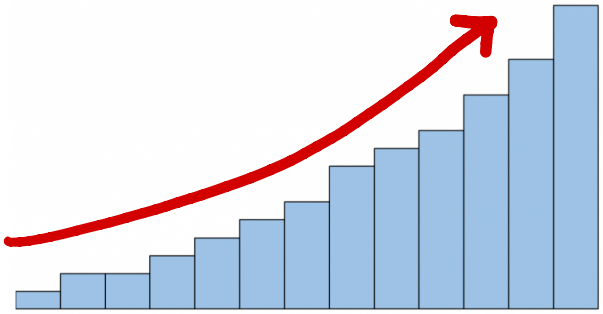
3) **Bimodal**: A histogram is described as "bimodal" if it has two distinct peaks.



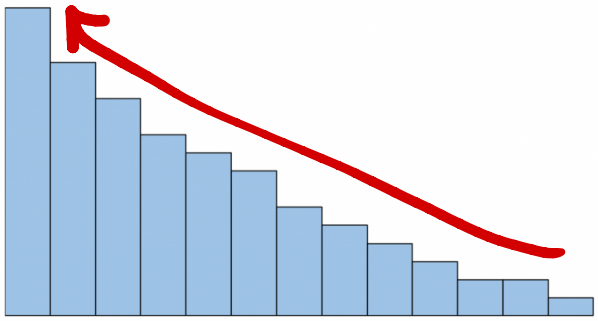
4) **Multimodal**: A histogram is described as "multimodal" if it has more than two distinct peaks.



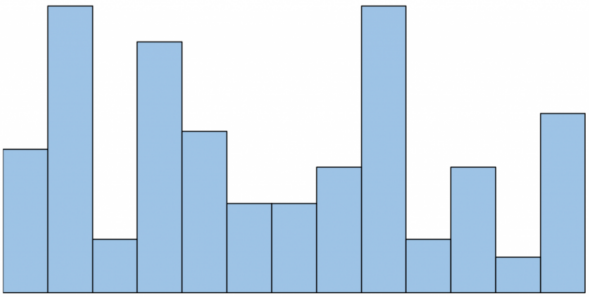
5) **Left Skewed**: A histogram is left skewed if it has a "tail" on the left side of the distribution. Sometimes this type of distribution is also called "negatively" skewed.



6) **Right Skewed**: A histogram is right skewed if it has a "tail" on the right side of the distribution. Sometimes this type of distribution is also called "positively" skewed.



7) **Random**: The shape of the distribution can be described as "random" if there's no clear pattern in the data at all.



Organizing and Graphing Data Practice Assignment

Continuous Data

$$interval\ size = \frac{possible\ range\ of\ data}{\#\ of\ intervals}$$

Relative Frequency Distribution

1. Given the final marks from last year's Grade 12 Data Management class. (continuous data)

42	49	50	51	63	66	67	68	69	70
73	74	77	78	80	83	84	85	85	88
89	90	91	93	95	96	97	98	99	100

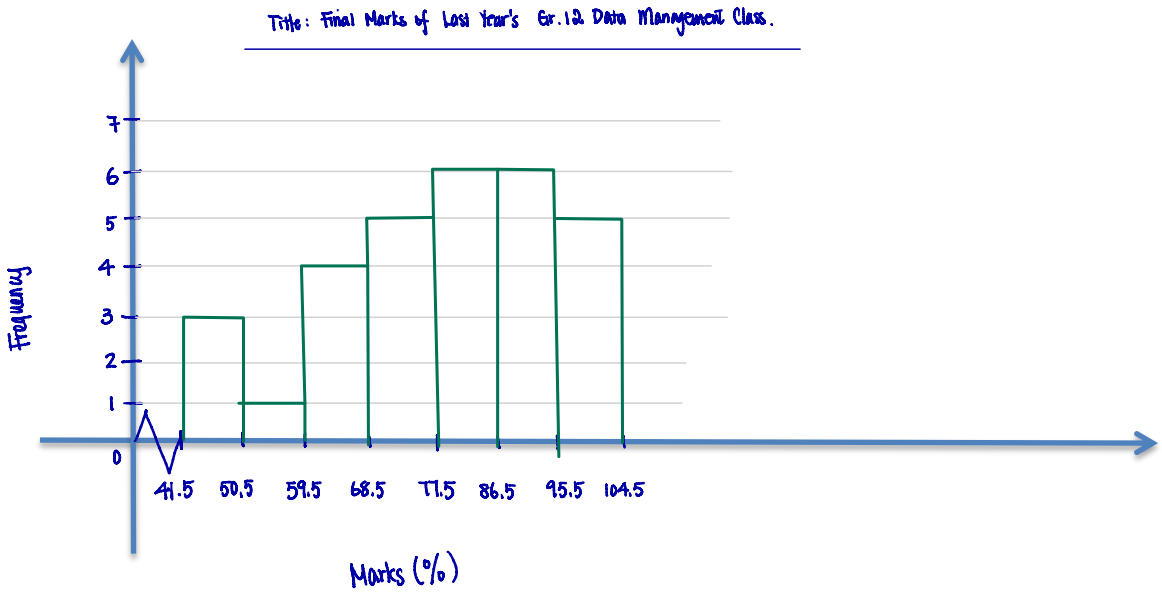
$\frac{100-42}{7} \doteq 8.2858$

$\doteq 9$ marks

a) Using the formula of determining the interval size above, construct a frequency table with 7 intervals, includes the column for the relative frequency

Intervals for Marks (%)	MIDPOINT	TALLY	FREQUENCY	RELATIVE FREQUENCY
41.5 – 50.5	(41.5+50.5)÷2 = 46	III	3	3 ÷ 30 = $\frac{1}{10}$ OR 0.1
50.5 – 59.5	55	I	1	1 ÷ 30 = $\frac{1}{30}$ OR 0.03
59.5 – 68.5	64	IIII	4	4 ÷ 30 = $\frac{2}{15}$ OR 0.13
68.5 – 77.5	73	~~IIII~~	5	5 ÷ 30 = $\frac{1}{6}$ OR 0.17
77.5 – 86.5	82	~~IIII~~ I	6	6 ÷ 30 = $\frac{1}{5}$ OR 0.2
86.5 – 95.5	91	~~IIII~~ I	6	0.2
95.5 – 104.5	100	~~IIII~~	5	0.17

b) Construct a histogram. (Be sure to label the graph clearly)



c) What proportion of students had marks between 60% and 69%?

This means exclusively

$\frac{4}{30} = \frac{2}{15}$ OR 0.13 OR 13%

Approximately 13% of the students in last year's data management class had marks between 60% and 69%.

d) Do you find the method of determining the interval from a) effective for analyzing this scenario? Why?

The intervals determined by the formula is not effective for analyzing this scenario because most students/teachers would like to know the results in away 10s of the intervals such as 40s, 50s, 60s, 70s, and 80s as our levels are determined using the same intervals.

e) If you are going to re-do the frequency table, how would you do it this time?

Intervals for Marks (%)	MIDPOINT	TALLY	FREQUENCY	RELATIVE FREQUENCY
39.5 – 49.5	44.5	II	2	$\frac{2}{30} = \frac{1}{15}$ OR 0.07
49.5 – 59.5	54.5	II	2	0.07
59.5 – 69.5	64.5	𝍸	5	$\frac{5}{30} = \frac{1}{6}$ OR 0.17
69.5 – 79.5	74.5	𝍸	5	0.17
79.5 – 89.5	84.5	𝍸 II	7	$\frac{7}{30} \doteq 0.23$
89.5 – 99.5	94.5	𝍸 III	8	$\frac{8}{30} = \frac{4}{15}$ OR 0.27
99.5 – 109.5	104.5	I	1	$\frac{1}{30} \doteq 0.03$

f) Why would you choose to organize the data using the selected interval style from e)?

I choose to organize the data using the selected interval style because it accommodates the normal practice in the grading system. Often, we determine the way how intervals are defined by identifying the target users of the analysis and cater to the needs for their perspectives.