

1 Collecting and Organizing Data

Primary and Secondary Data

Sometimes you collect data directly by asking people questions or observing things yourself. For example:

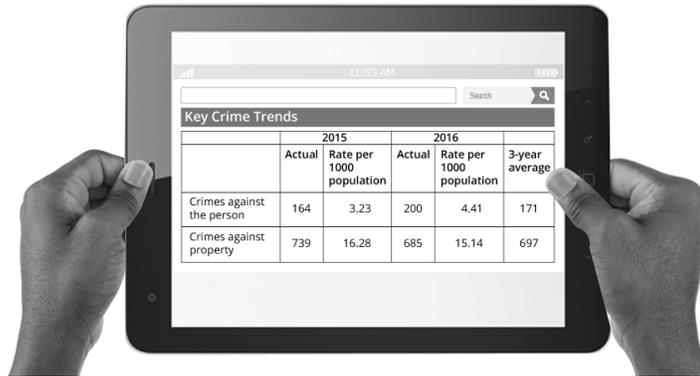
- You might stand at an intersection and count the vehicles that travel through that intersection at different times of day.

	9:00 a.m. - 9:10 a.m.	12:00 p.m. - 12:10 p.m.	5:00 p.m. - 5:10 p.m.
Number of vehicles	 		

- You might conduct a survey in which you ask people about their political opinions or their favourite pets.

When you collect data yourself, you are working with primary data.

But sometimes you gather data from elsewhere, for example, from the news media, government reports, or historical documents.



Data that have been collected by another source are known as secondary data, and their source is called a secondary source.

Data Organization

Secondary data are often organized in tables, charts, or graphs.

Primary data are not organized in advance. The person who collects the data must decide how to organize the data to create appropriate categories or appropriate graphs.

Discrete Versus Continuous Data

Some data fit into distinct categories, for example, people's favourite board games or the colours of people's eyes. Other data can be ordered, for example, the distances children travel to school or the number of siblings people have. Data that can be placed in order can be continuous or discrete.

In continuous data, values can be anything between a minimum and a maximum amount. Data about people's heights are continuous because there is always a possible height between two other heights.

Discrete data also have a particular order, but they can have only certain values. Data about phone payments are discrete. A person might pay \$45 on 1 January and \$47 on 1 February, but it doesn't make sense to think about how much the person pays in between those two dates, because he or she pays a phone bill only once a month.

Often personal judgment is involved in deciding whether a set of data is continuous or discrete. For example, the number of people in a car is discrete because only a few values are possible, but you might consider the population of Canada to be continuous because although you can include only whole numbers of people, so many different values are possible.

Reading Versus Interpreting Data

You can read information in a table, chart, or graph. You can also interpret the data and draw conclusions.

For example, if you were to ask 100 Canadians about their favourite sport to watch on TV, you might find these results:

Favourite sport to watch	Percentage of responses
hockey	35
baseball	10
football	20
soccer	15
basketball	20

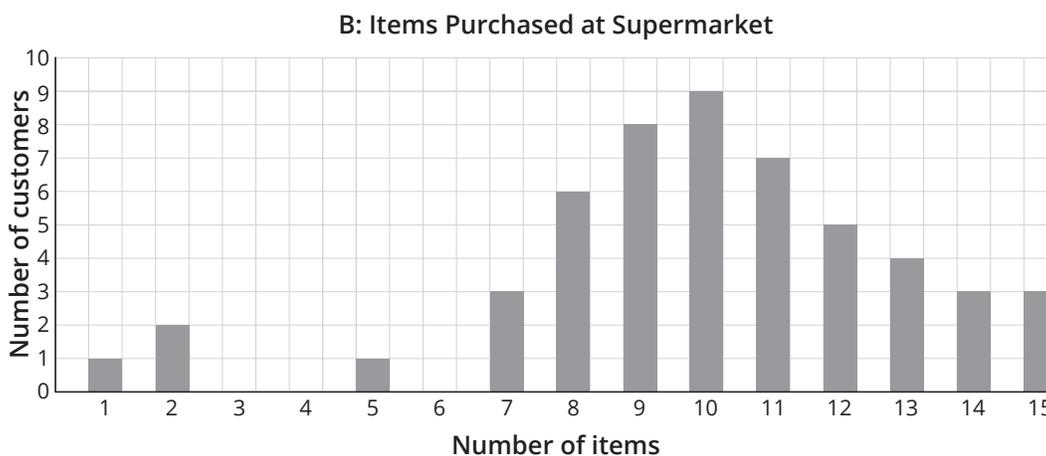
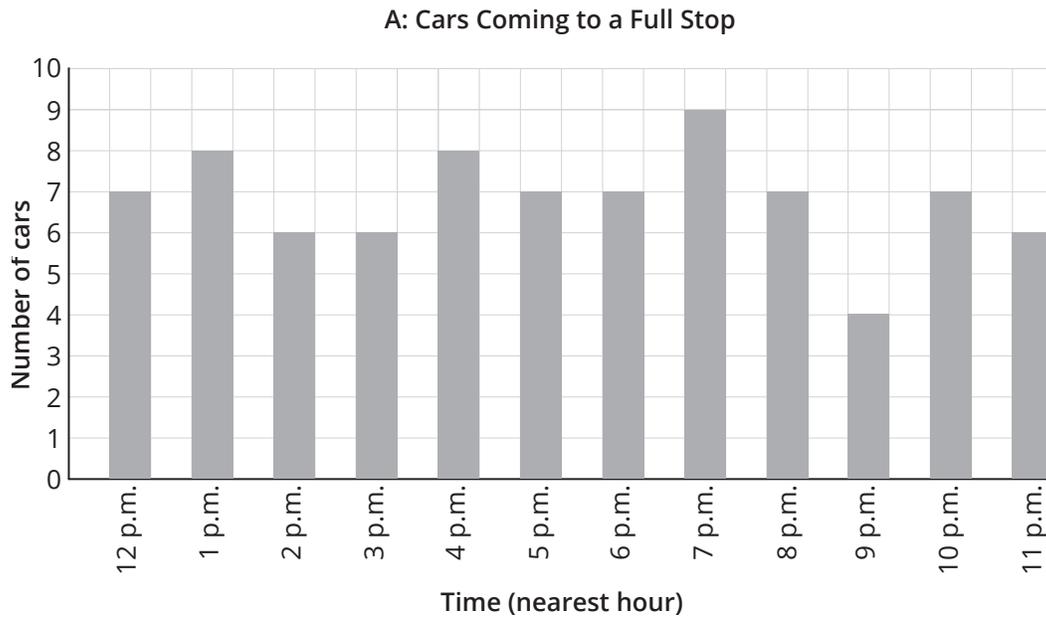
You can read the data and see that $\frac{35}{100}$ Canadians prefer to watch hockey. If you interpret the data, you might conclude that hockey is Canadians' favourite sport to watch on TV.

Data Range and Distribution

You can use certain characteristics to help you interpret data.

The range of the data tells you the difference between the least value and the greatest value in a set of data.

The distribution of the data tells you how the data values cluster, that is, whether values are very close together or evenly spread out.



The data in graph A have an even distribution. The range is 5 cars.
 The data in graph B are clustered around 10 items. The range is 9 items.

Definitions

continuous data: data with values that can be represented on a number line or a graph axis; continuous data are usually measured rather than counted; for example, children's heights are continuous but the grades children are in at school are not

discrete data: data that can have only certain fixed values; discrete data are usually counted rather than measured; for example, the grade children are in at school is discrete, but the children's heights are not

distribution: how the data values in a set of data are arranged; for example, a set of data can be clustered together or spread out

primary data: data you collect yourself

range: the difference between the greatest and least values in a set of data; for example, for the data 3, 7, 19, 20, 45, the range is $45 - 3 = 42$

secondary data: data that were collected and reported by someone else, for example, data in a newspaper or online

survey: a collection of data from a small group (a sample) to learn about characteristics of a large group (a population)