



Data Visualization: Best Practices

Release date: February 24, 2023

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- | | |
|---|----------------|
| • Statistical Information Service | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line | 1-514-283-9350 |

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “[Standards of service to the public](#)”.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada as represented by the Minister of Industry, 2023

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Data Visualization: Best Practices

Introduction

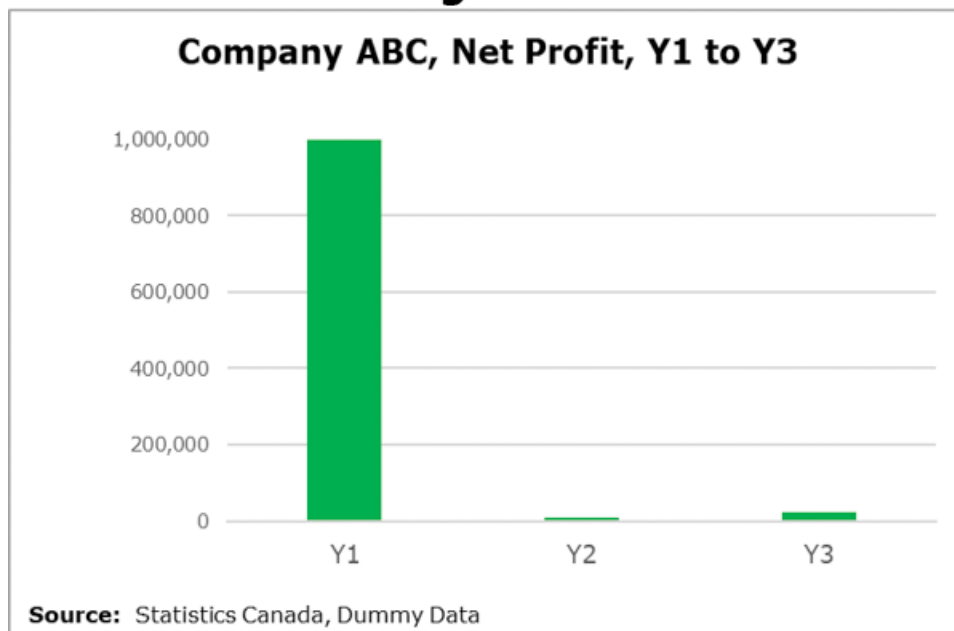
When used properly, charts (including figures and diagrams) can simplify the presentation of information and the communication of clear and precise messages. However, with the wide range of options available, creating effective charts can be complex. This reference tool is intended to provide a basic guide to creating effective charts that take advantage of the options available.

Preparation

Before creating a chart, it is important to answer a few important questions.

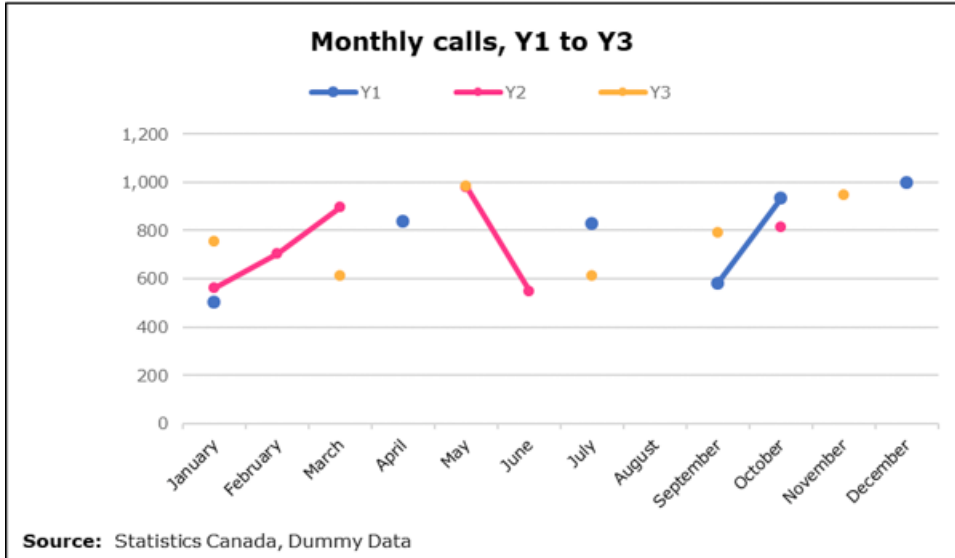
- **What data do I have access to?** The purpose of this question is to clarify the nature and status of the data that you are aiming to present through a chart. The type of data, the number of variables available, and the presence of missing data will all have an impact on the type of chart to be used, as well as the way in which the information is presented.
- **Who is my target audience?** This question is designed to determine who the chart is primarily intended for. The target audience and their level of knowledge and expertise will have an impact on the way the information is presented, as well as the type of chart and the terminology used.
- **What is the message I want to deliver?** Charts can convey a message quickly and effectively. It is therefore important to establish the message that you want to highlight. This will have an impact on the type of chart chosen, its title and many other elements.
- **Should I use a graph?** Charts can be very useful in many circumstances. However, there are some situations where it may be better not to use charts:
 - ▶ When the data are very scattered;

Figure 1



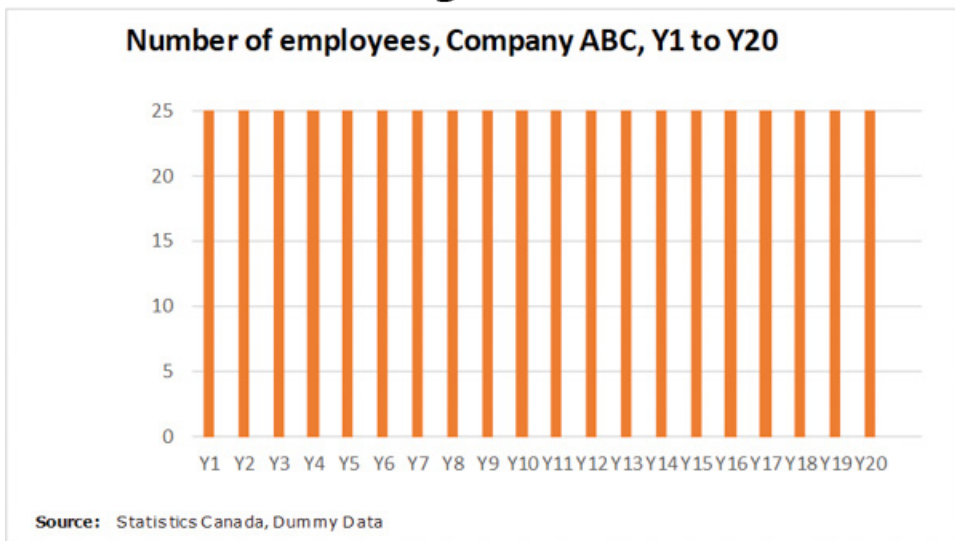
► When there is a lot of missing data;

Figure 2



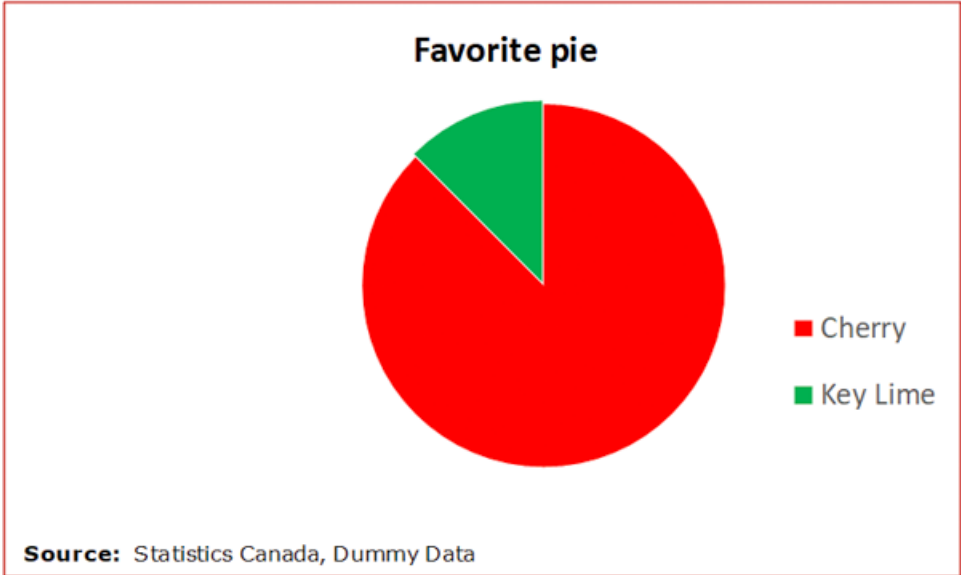
► When the data are very homogeneous;

Figure 3



► When there is little data;

Figure 4



► When there is a lot of data.

Figure 5

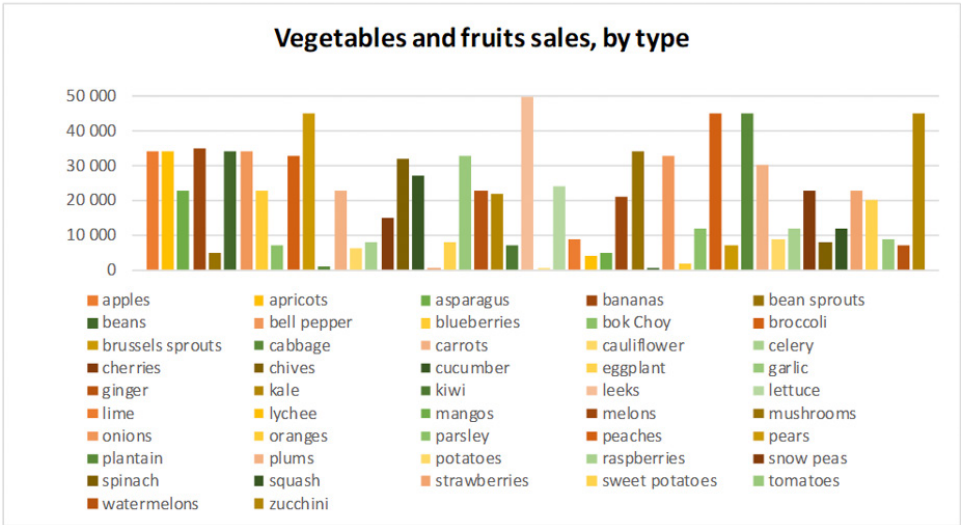
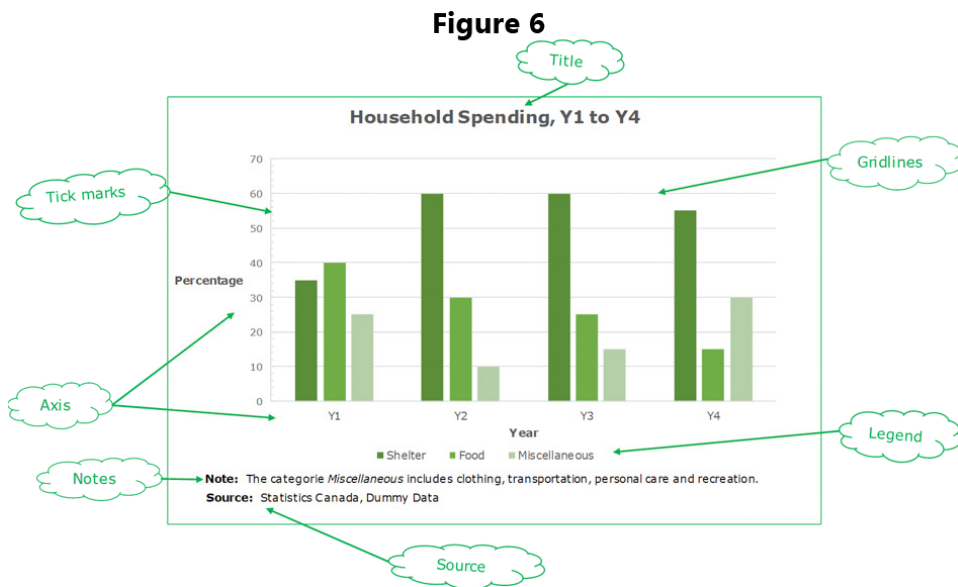


Chart components

The different types of charts have common components, for which there are best practices.



1. Title

The title of a chart usually appears at the top of the chart and is part of it. The different axis of the chart can also have their own titles.

- The title of the chart can be descriptive (e.g., include variables and time periods covered) or informative (e.g., specify the intended message of the chart).
- Axis titles should be descriptive.
- The title should be clear and help the reader better understand the data presented.
- The chart, including its title, should be considered separate from the text. If necessary, the acronyms used should be redefined.
- If the data refer to a specific unit of measurement, this can be specified in the title of the chart or the title of the axis.

2. Axis

Some types of charts use axis to present the data.

- Axis titles are not always needed, unless there is a specific unit of measurement.
- With special exceptions, axis should start at zero and not be broken.

3. Tick marks

Tick marks provide visual cues for easy reading of the data.

- There should be a balance between including too many tick marks, cluttering the axis, and too few tick marks, making it difficult to read the data.
- The intervals should be regular, such as multiples of 10, 100, 1,000 or million.
- Tick marks should always be on the inside of plot area.
- Tick marks should not be used simultaneously with grid lines.

4. Grid lines

Grid lines appear on certain types of charts, in order to facilitate the reading and comparison of values.

- The lines should be thin, light and placed behind the elements presenting the data.
- The space between the lines should be sufficient to facilitate the reading of the data, without cluttering the chart.
- Grid lines should not be used simultaneously with tick marks.

5. Legend

A legend can be used to label the different variables or categories presented in a chart. It can be particularly useful when there are multiple items.

- The legend should appear close to the objects it labels, without interfering with the data.
- The legend should be less prominent than the objects it labels.
- A legend should always be used if there are multiple categories of data (in different colours).

6. Notes

Notes can be used to provide details about the chart, such as methodological considerations, data limitations, or a description of abbreviations used.

- Footnotes should be numbered, notes do not need to be.
- Notes should be brief and specific.
- Depending on the audience, notes may be used to clarify technical terms used.

7. Sources

A chart should always include the sources of the data used.

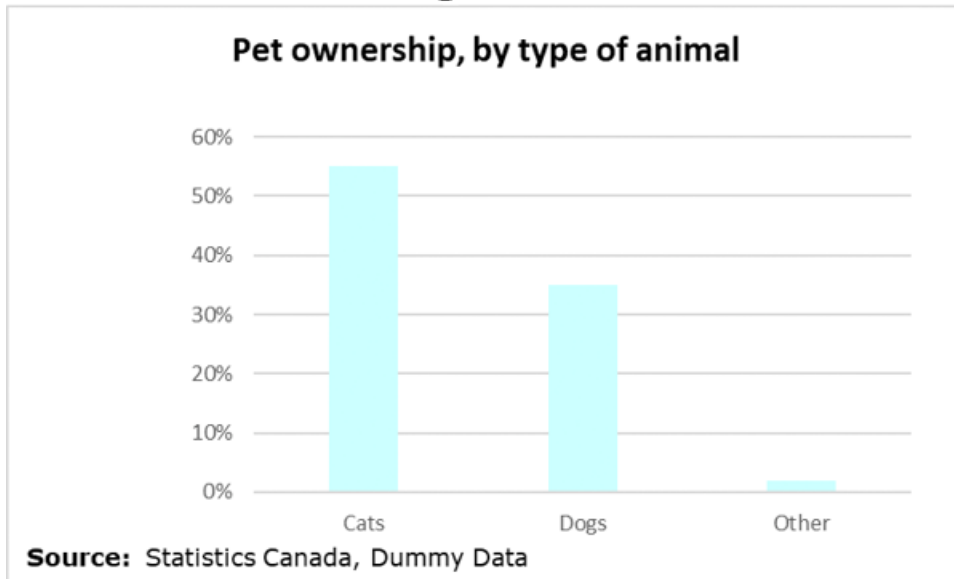
- Data sources should be specific enough to allow readers to find the data, if necessary.
- If the data source is frequently updated (e.g., exchange rates between two currencies), it is important to specify the date the data were extracted.

8. Colours

Colours can be used to make data easier to understand, to emphasise specific elements or to communicate certain messages quickly.

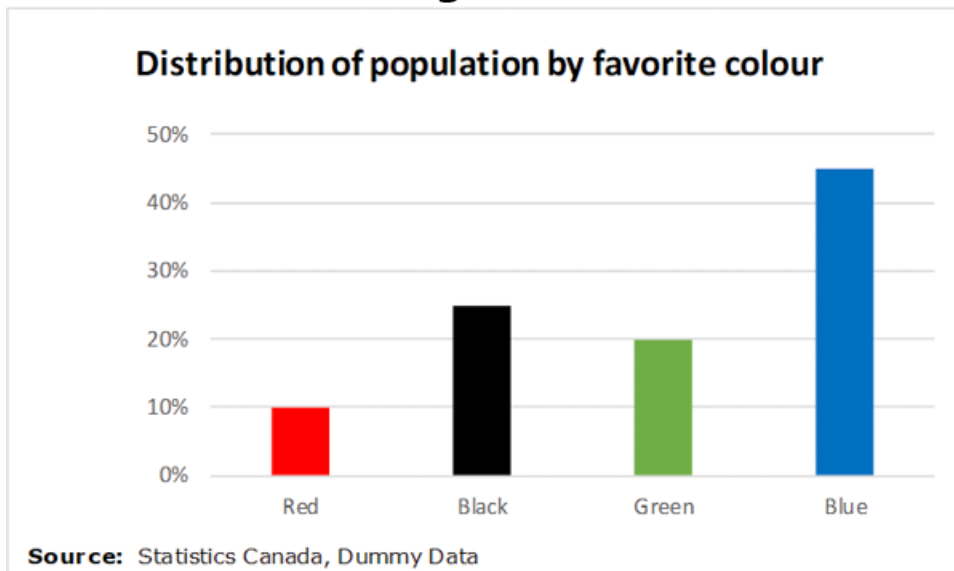
- To facilitate the visualisation of data, it is recommended to use colours that contrast with the background colour of the chart.

Figure 7



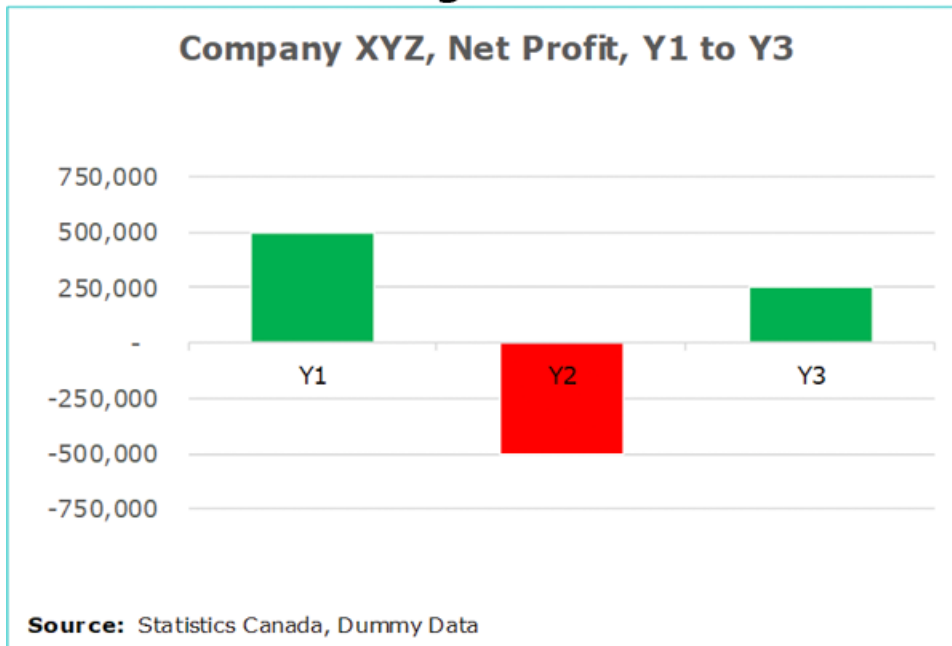
- Colours should be used only when necessary to achieve a specific communication objective.

Figure 8



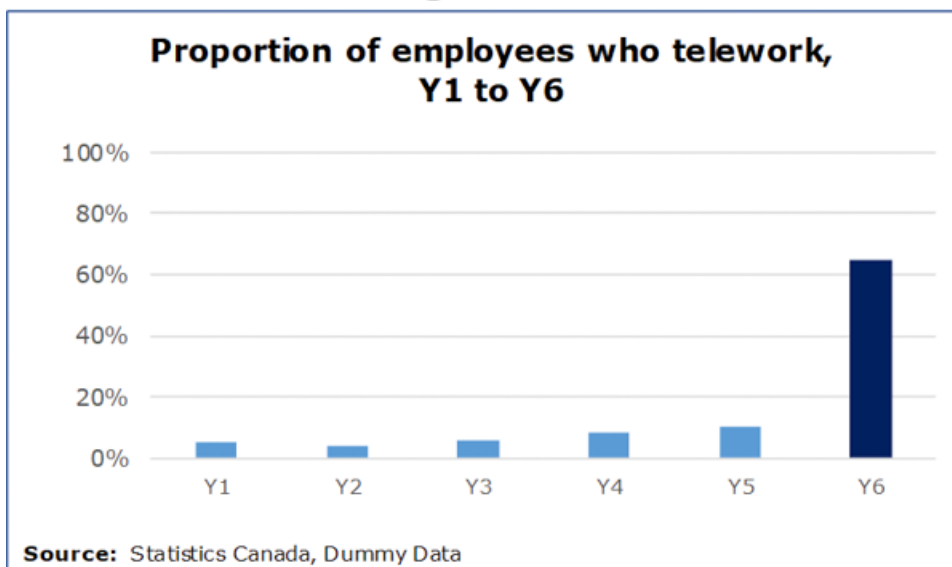
- Different colours should be used only when they have different meanings.

Figure 9



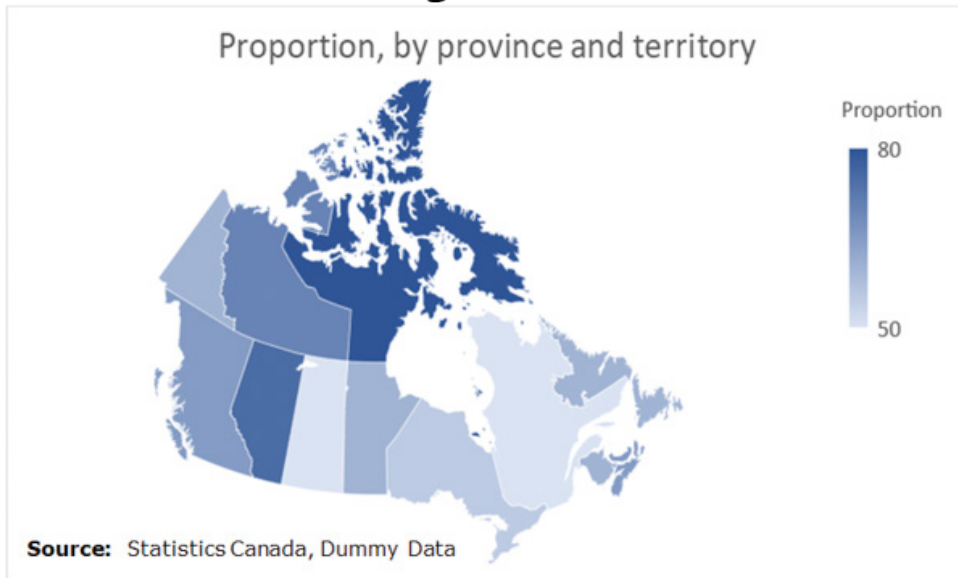
- Do not use shade of similar colours without adequate space between them.
- Light, natural colours should be used to present most data, except for data to which attention is to be drawn, for which bright or dark colours are preferred.

Figure 10



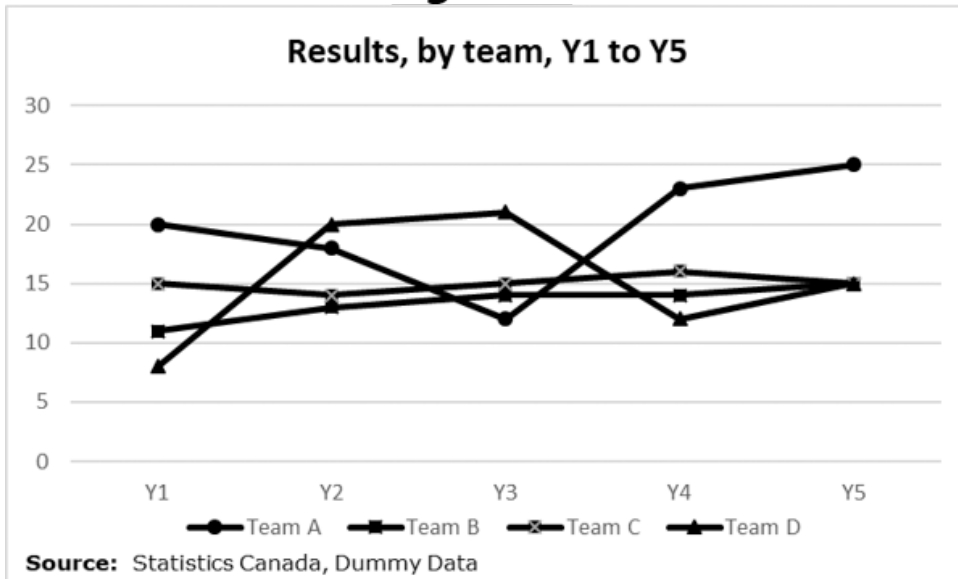
- Where different values of a variable are to be presented, it is preferable to use a single colour, varying the intensity (lighter for low values and darker for high values).

Figure 11



- Patterns can be used instead of colours when required.

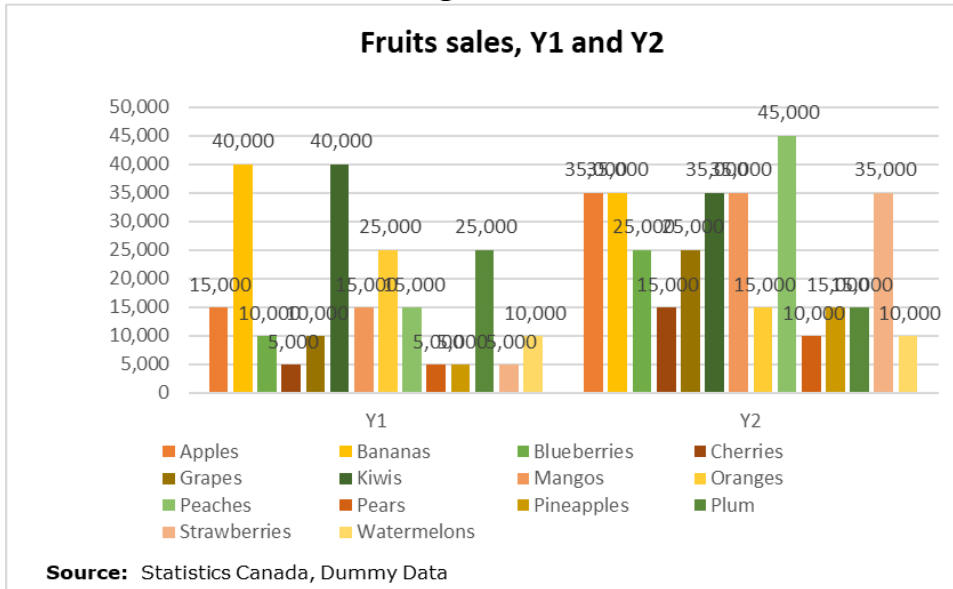
Figure 12



9. Other considerations

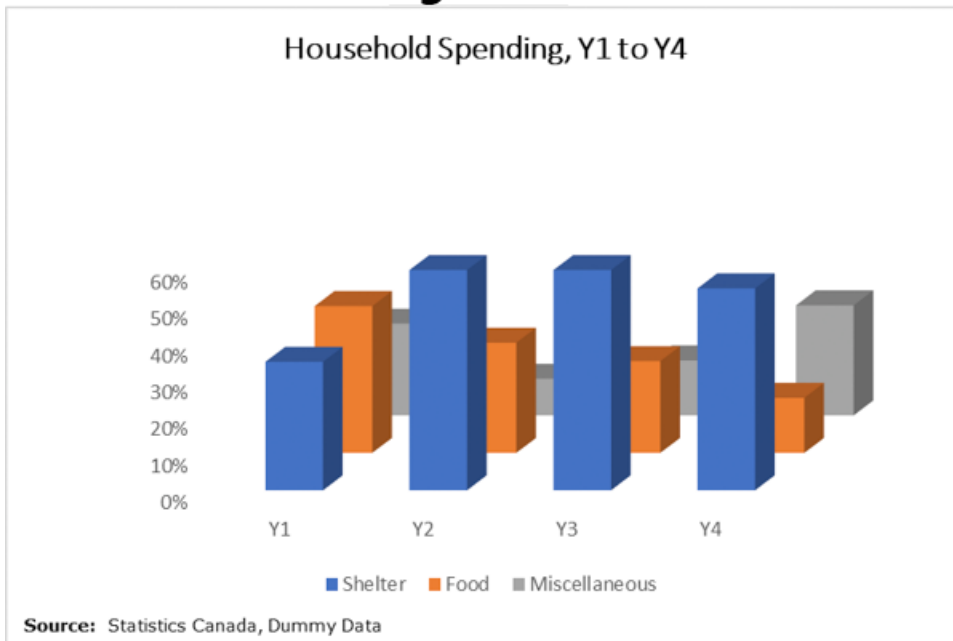
- Keep it simple and avoid overloading the charts.

Figure 13



- With some exceptions, avoid 3D charts, which are more difficult to understand.

Figure 14



- For accessibility, notes should be used if there are visual elements on the chart that are not part of the data table.

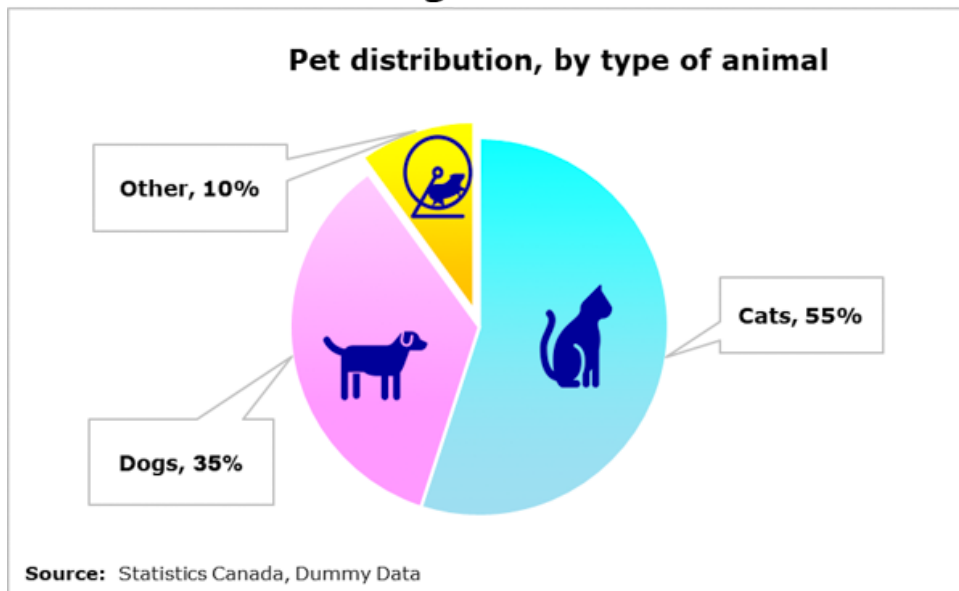
Types of charts

There are several types of charts, each with advantages and disadvantages, depending on the context and the nature of the data. The selection of the right type of chart will be influenced by different factors such as the type of data, the messages to be communicated and the target audience.

1. Pie charts

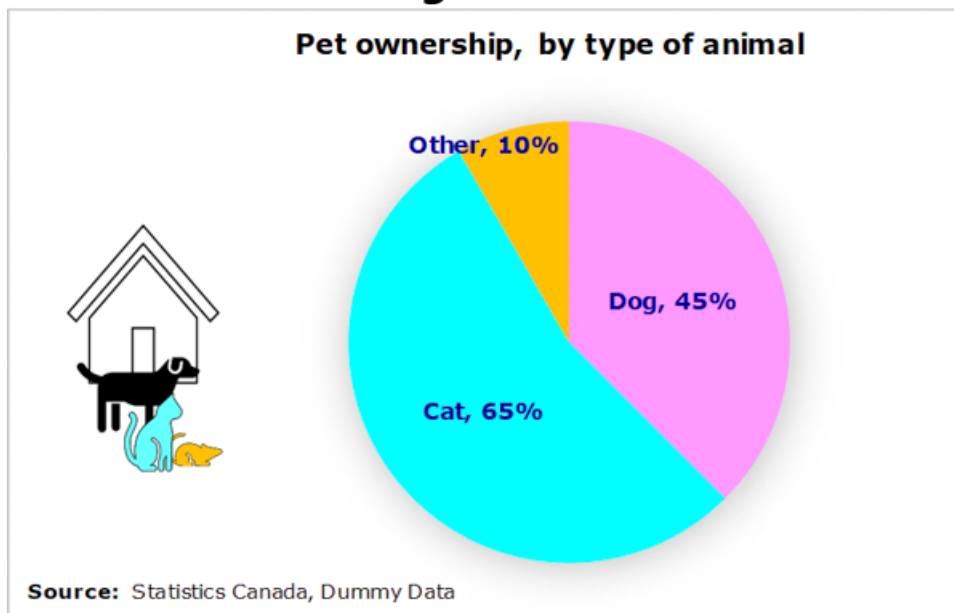
A pie chart shows the percentage distribution of a given variable. Each segment represents a category and its size is proportional to its weight in the total.

Figure 15



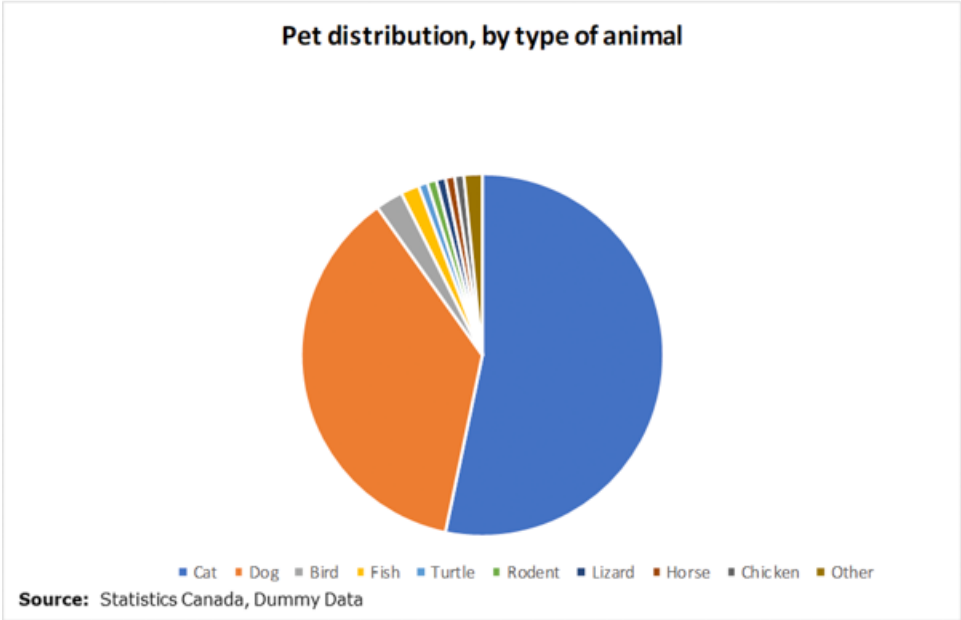
- Pie charts can only be used for data where the sum of the different categories adds up to 100%.

Figure 16



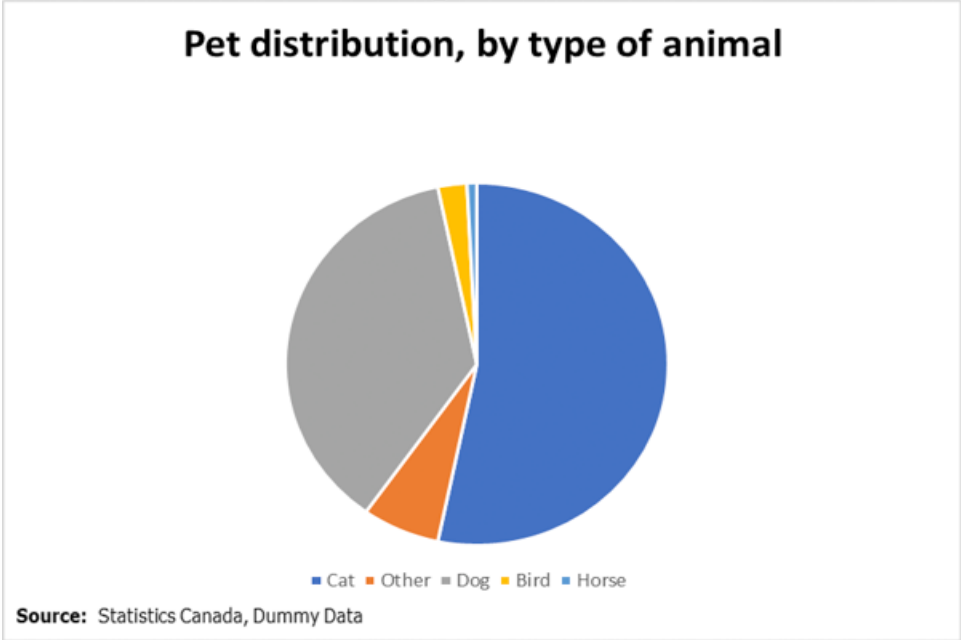
- Ideally, a pie chart has between 2 and 6 different categories.

Figure 17



- The categories should be presented in descending order, clockwise.

Figure 18

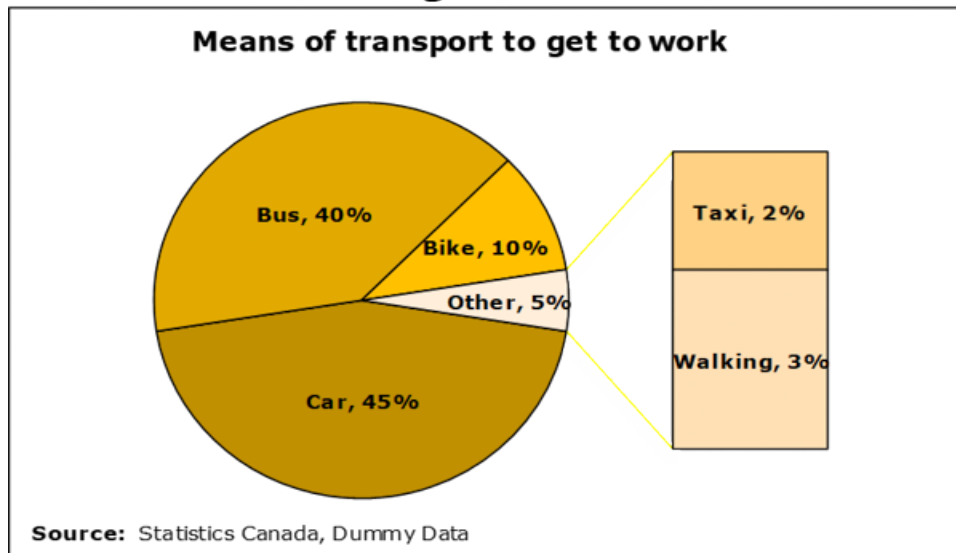


- It is not possible to present uncertainty on a pie chart.

Pie chart family

a. Bar of pie charts

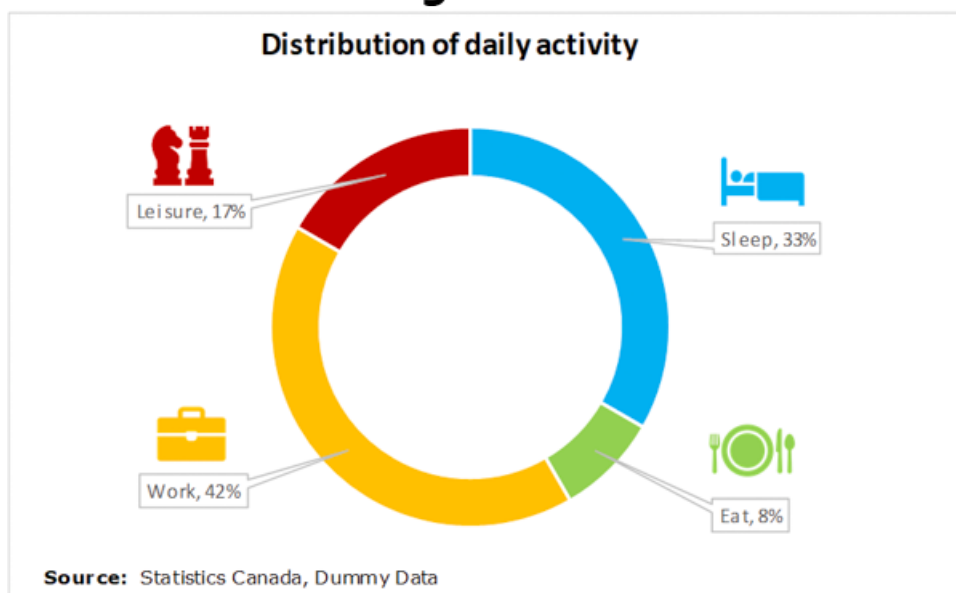
Figure 19



A bar of pie charts can be used when there are more than six categories, or when there are several small categories, which are difficult to illustrate clearly in a regular pie chart. In these cases, a new category called 'Other', the amount of which is equal to the sum of the smaller categories, is inserted into the main chart. Stacked bars show these categories next to the pie chart.

b. Donut charts

Figure 20

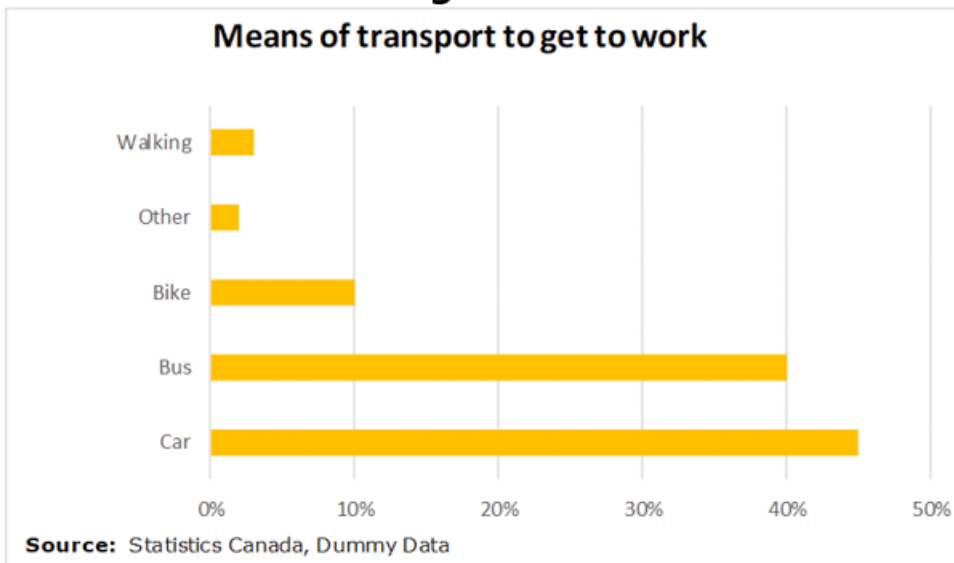


A donut chart is a pie chart with a hole in the centre. The hole makes it more difficult to estimate the relative size of categories, but can be used to present relevant information, such as a logo.

2. Bar charts

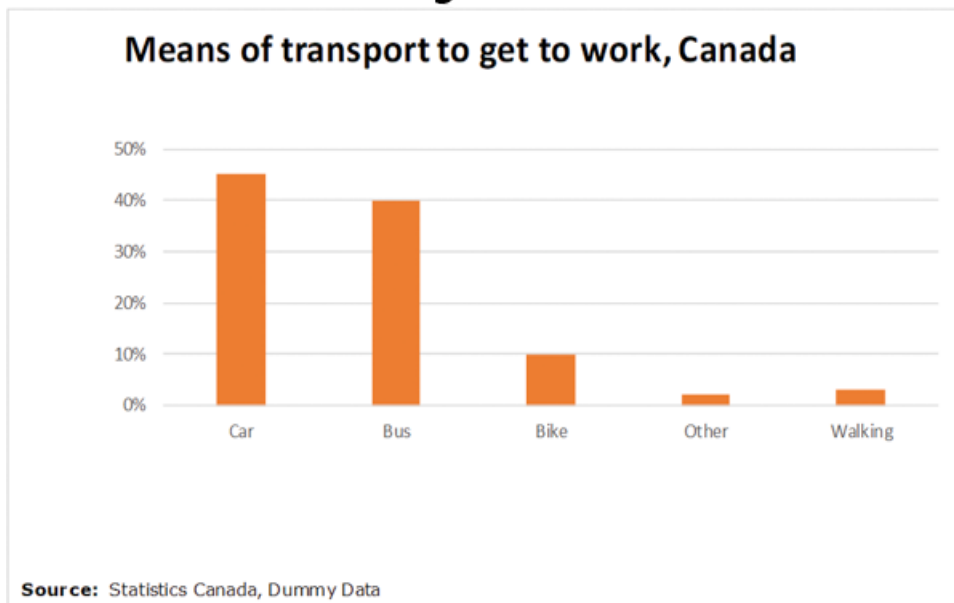
A bar chart uses bars to represent the different categories. It can be vertical or horizontal and has two axes. The names of the different categories are shown on one axis or with labels on the bars. The value of the data is shown on the other axis: this is called the scale.

Figure 21



- A vertical bar chart provides more space for the names of the categories presented.

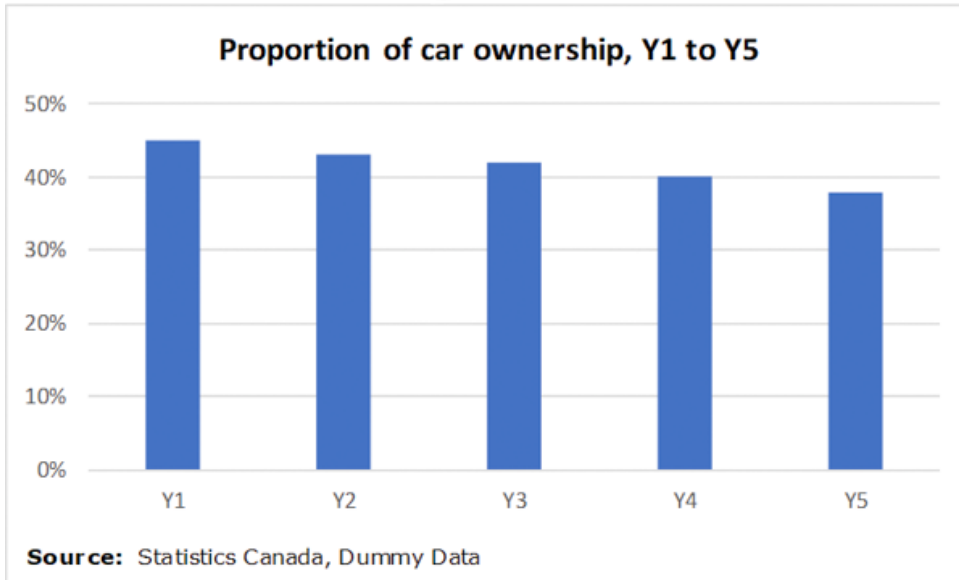
Figure 22



- If the aim is to emphasise the order of magnitude of different categories, present them in ascending or descending order. Do not change the order of categories that have a natural order, such as months or years.
- Ideally, a bar chart would have between 2 and 10 different categories.

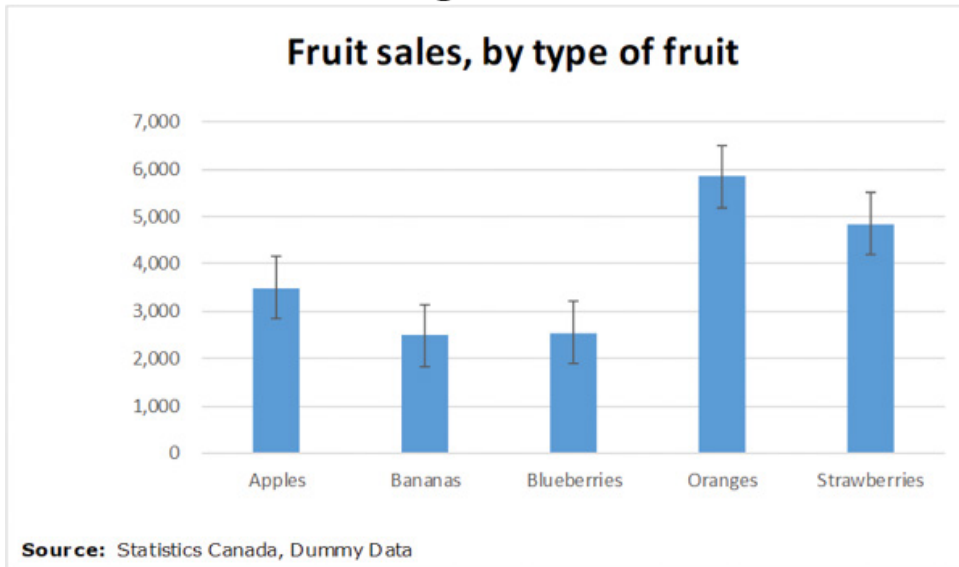
- Where the data have a time element, this should be presented chronologically on the X-axis, from left to right.

Figure 23



- It is possible to present uncertainty on a bar chart.

Figure 24

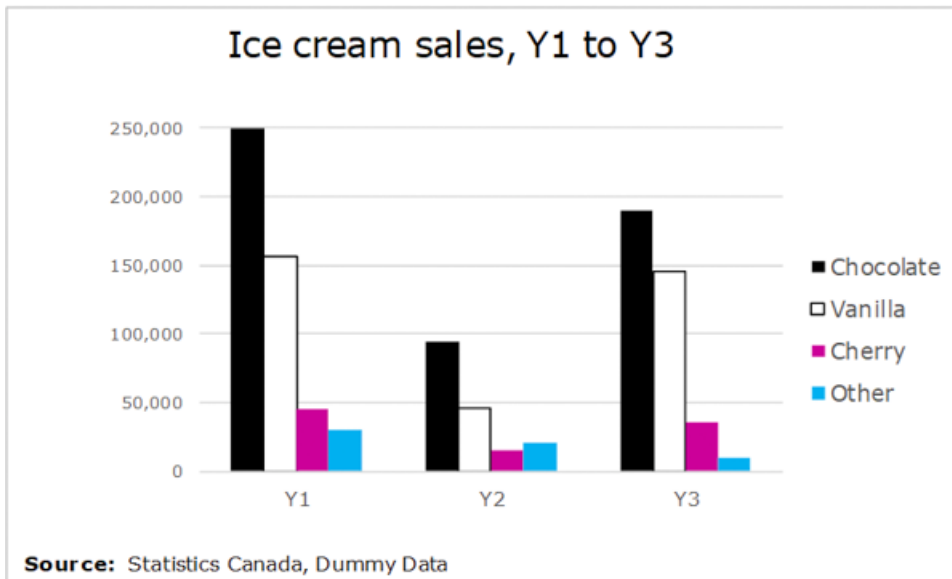


Bar chart family

a. Grouped bar charts

It is possible to present two or more series of data in a grouped bar chart. However, the more series there are, the more difficult it is to focus on one at a time.

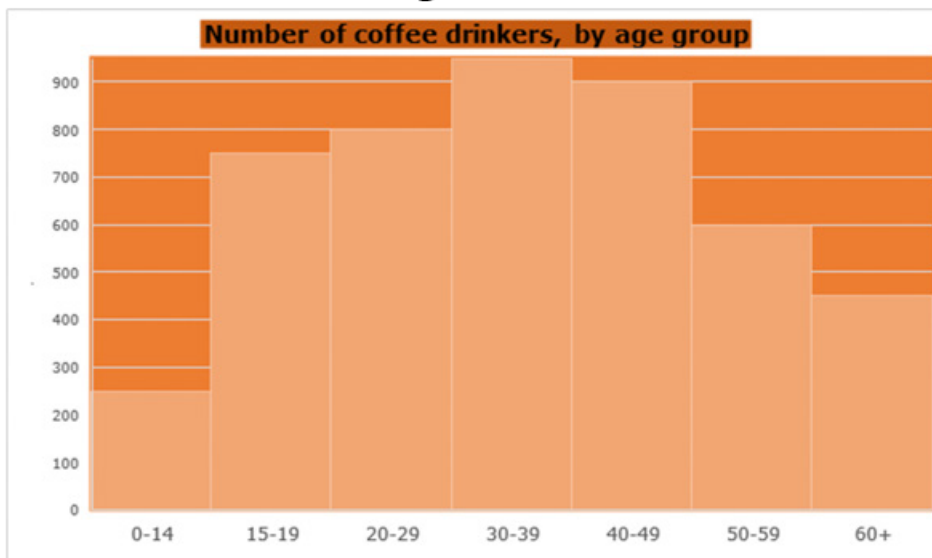
Figure 25



b. Histograms

Histograms are used to illustrate the summary of a continuous variable measured on an interval scale. In a histogram, the bars are connected to each other, with no space between them.

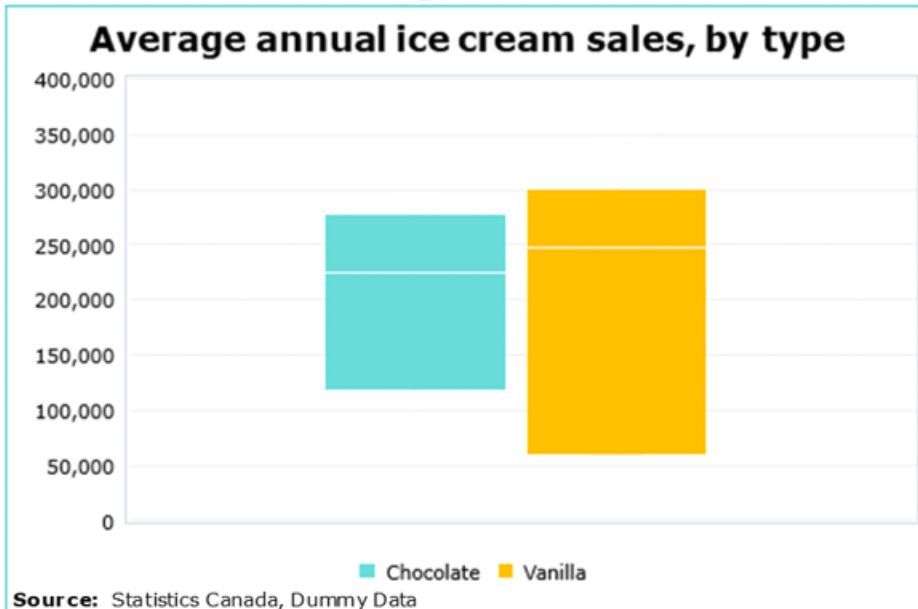
Figure 26



c. Box plots

Box plots are used to illustrate the distribution of different categories of a variable. Each bar starts at the minimum value and ends at the maximum value of the category. There is usually a thick line inside each bar that shows the center of the distribution, usually the median.

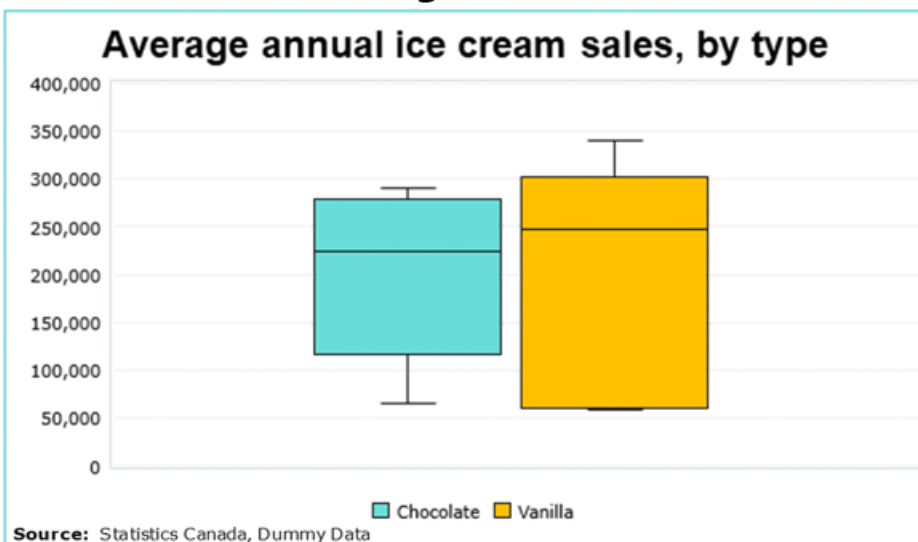
Figure 27



d. Box and whisker plots

This chart is one of the most effective charts for visualizing information about the frequency distribution of variables and the distribution of a continuous variable. It displays minimum, first quartile, median, third quartile and maximum value of a category of a variable.

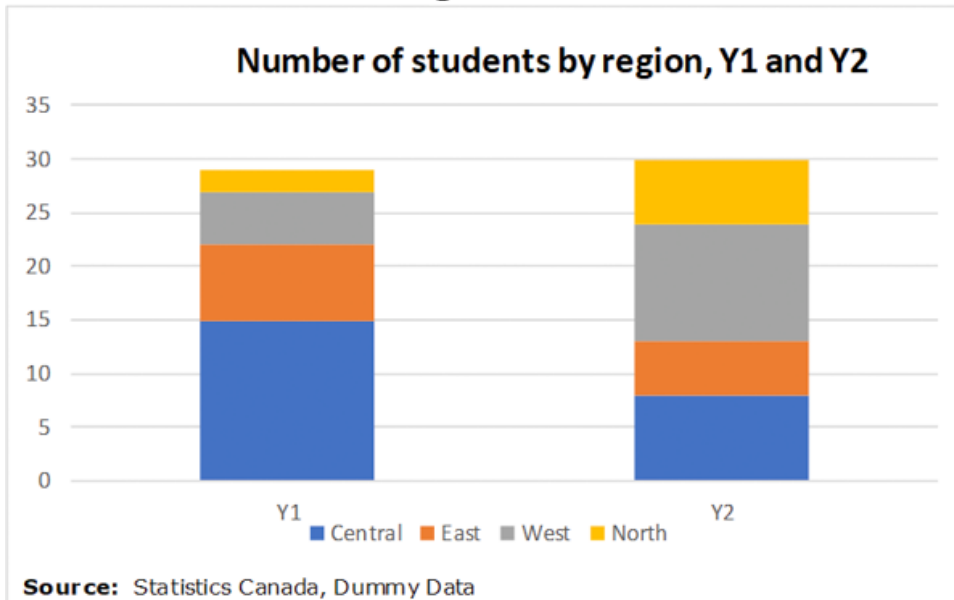
Figure 28



e. Stacked bar chart

Stacked bar charts are used to illustrate the total values of different categories. Additionally, each bar is broken down to show subcomponents in each category. Since the baseline of subcategories varies between bars, only the first subcomponent can be visualized efficiently.

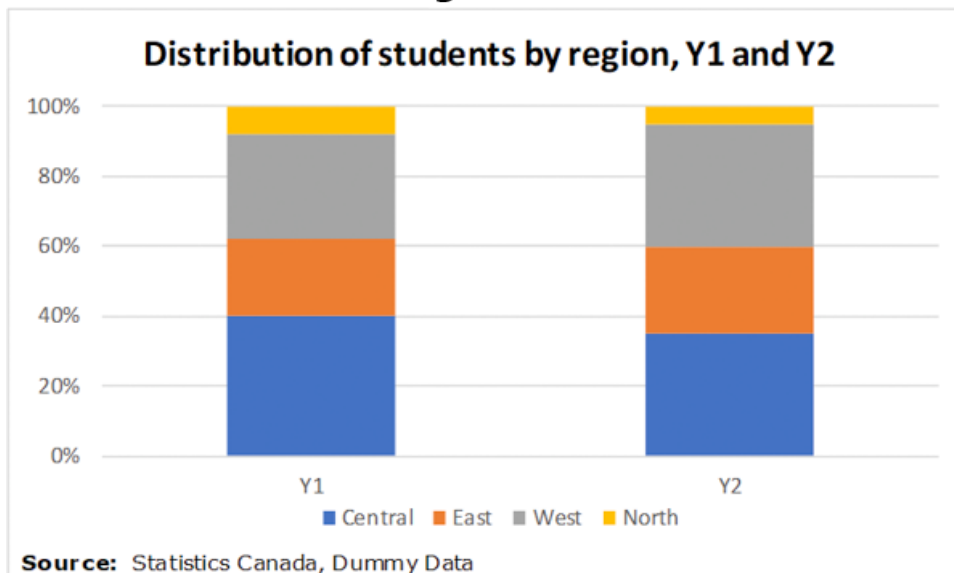
Figure 29



f. 100% stacked bar charts

100% stacked bar charts are used to illustrate the ratio of subcategories. They are similar to stacked bar charts, but show the relative value of each category rather than the absolute value.

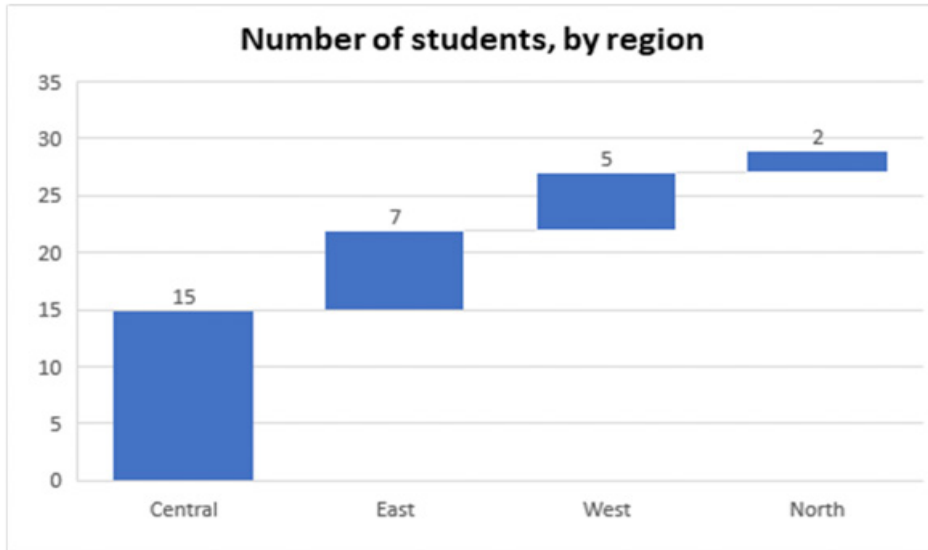
Figure 30



g. Waterfall charts

Waterfall charts pull apart the pieces of a stacked bar chart and show each subcomponent separately. The first bar starts from its natural base value and the rest of the bars start at the value of the previous bar and can have a positive or a negative value.

Figure 31

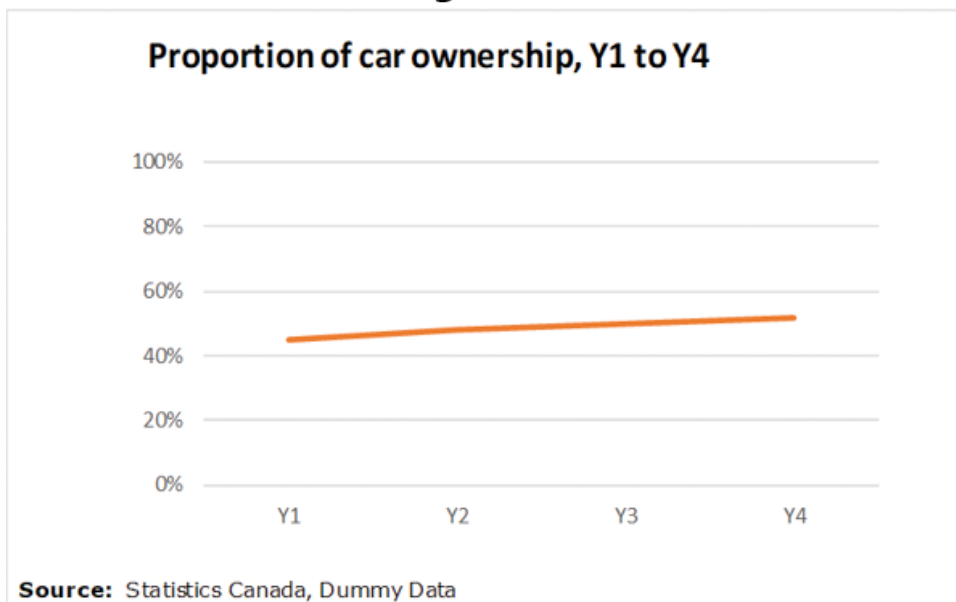


Source: Statistics Canada, Dummy Data

3. Line charts

Unlike bar charts which emphasize individual values, line charts emphasize continuity and evolution from point to point. They are commonly used to show changes and trends over time.

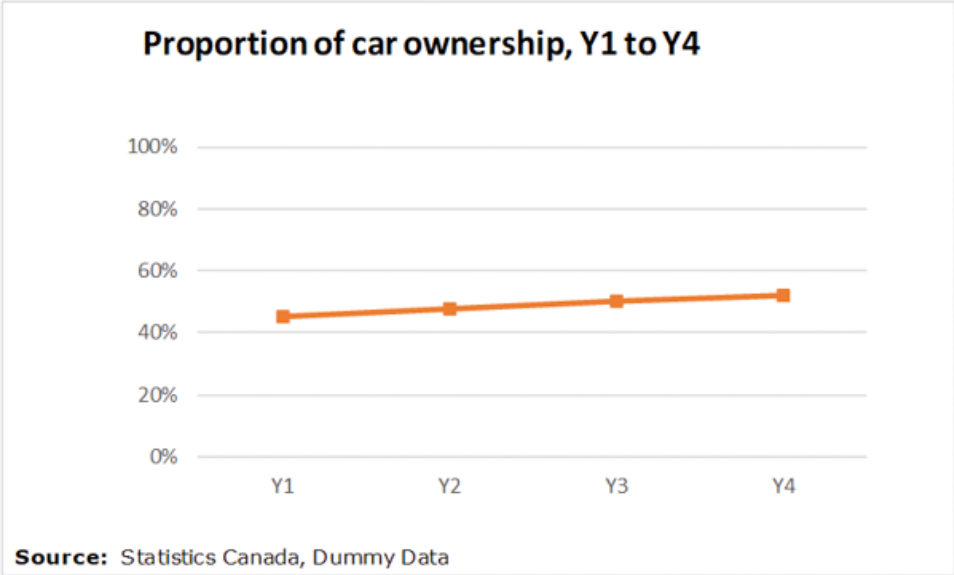
Figure 32



Source: Statistics Canada, Dummy Data

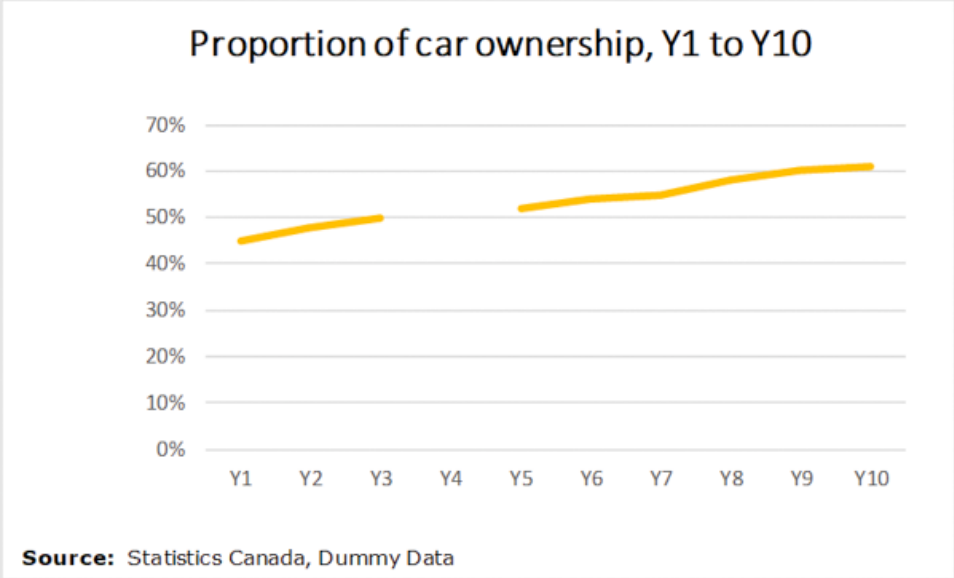
- Dots can be used on the line to emphasize values.

Figure 33



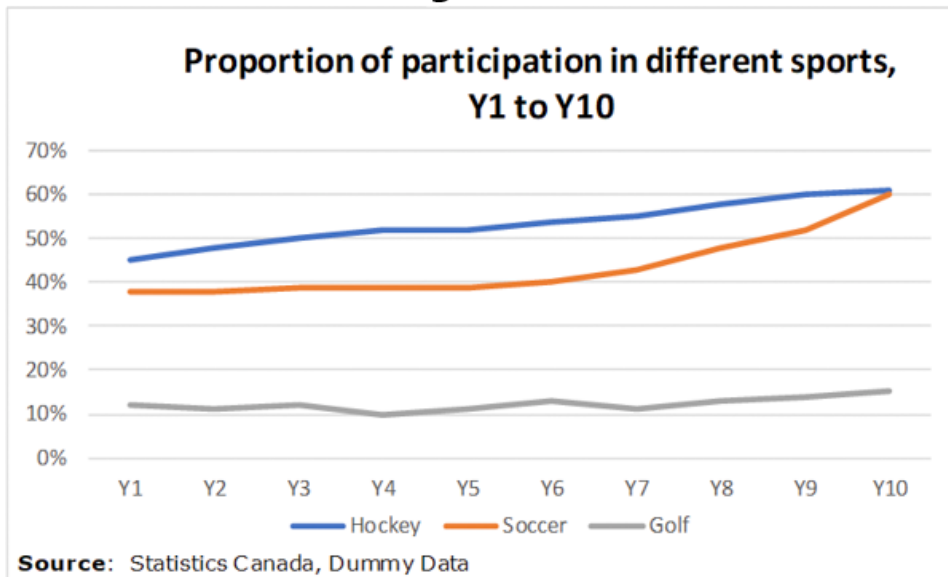
- Spaces or dotted lines can be used when data are missing.

Figure 34



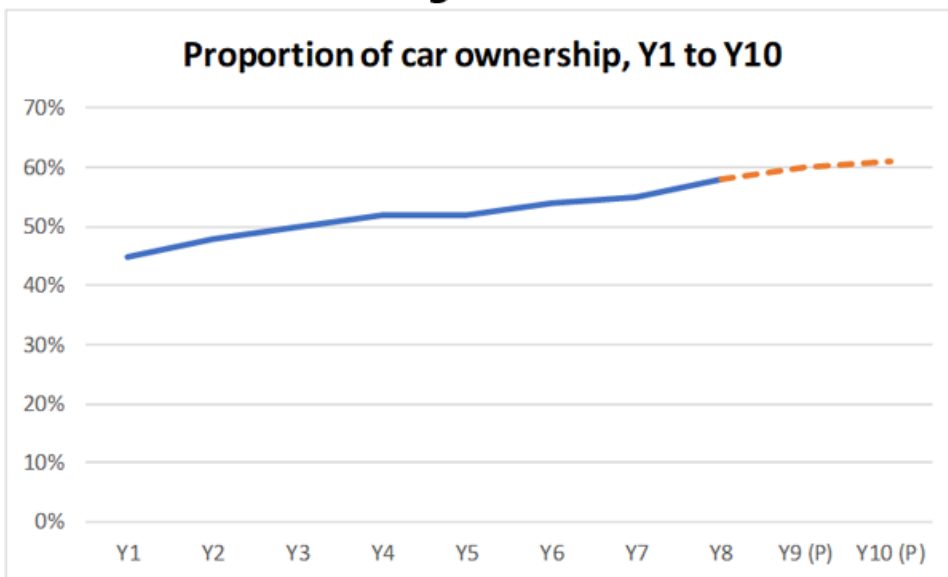
- When multiple items are presented on the same chart, they should have the same units of measure; different colors should be used to distinguish them; and the lines should be visually distinct.

Figure 35



- A different line (e.g., dotted or different color) should be used to distinguish actual data from trends, projections, and targets.

Figure 36



Note: (P) = Projections

Source: Statistics Canada, Dummy Data

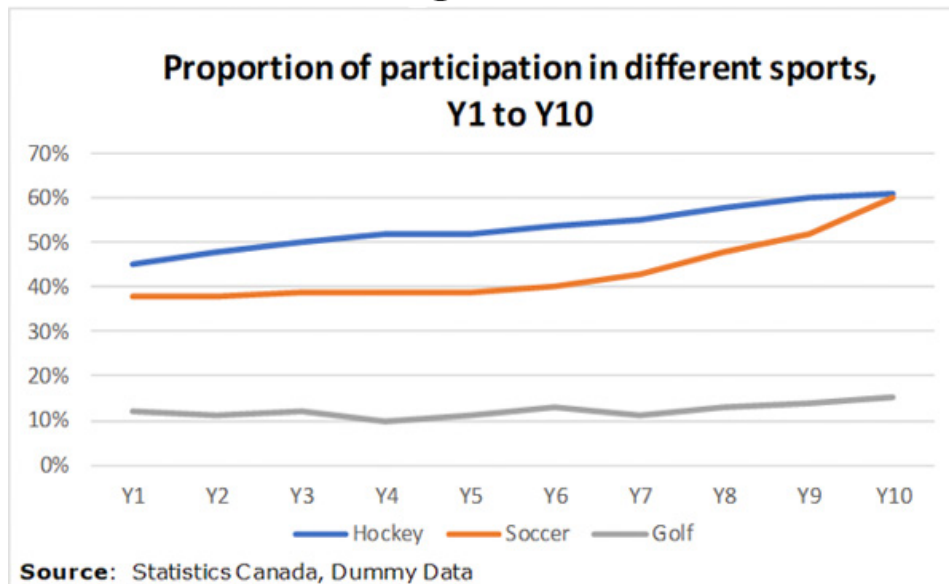
- Shading can be used to show uncertainty.

Line chart family

a. Grouped line charts

It is possible to present two or more series of data in a grouped line chart. However, the more series there are, the more difficult it is to focus on one at a time.

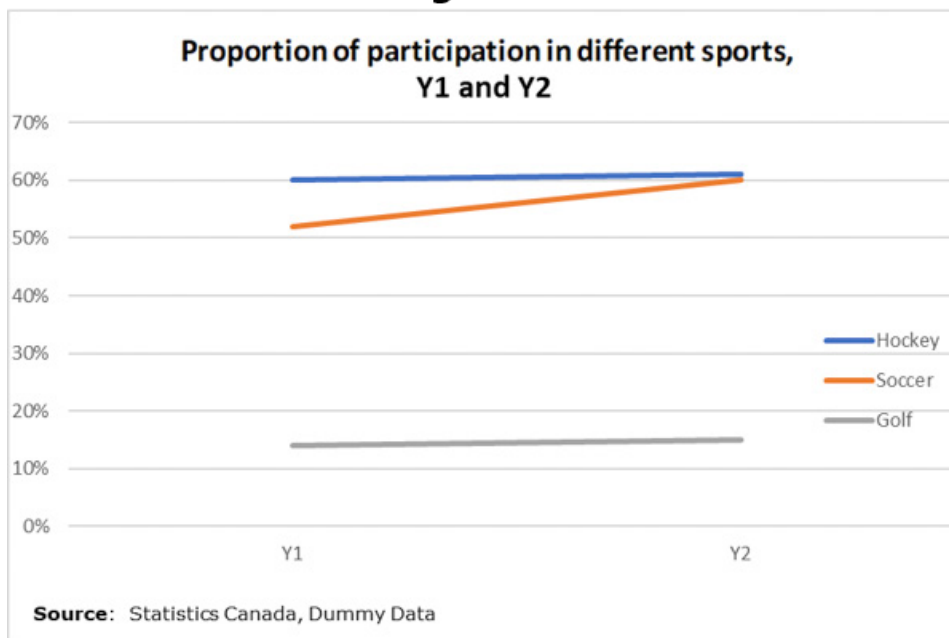
Figure 37



b. Slope graphs

Slope graphs illustrate the relative increase or decrease in a set of variables between two data points. They provide a clear visual ordering among variables and can be used to visualize a ranking.

Figure 38

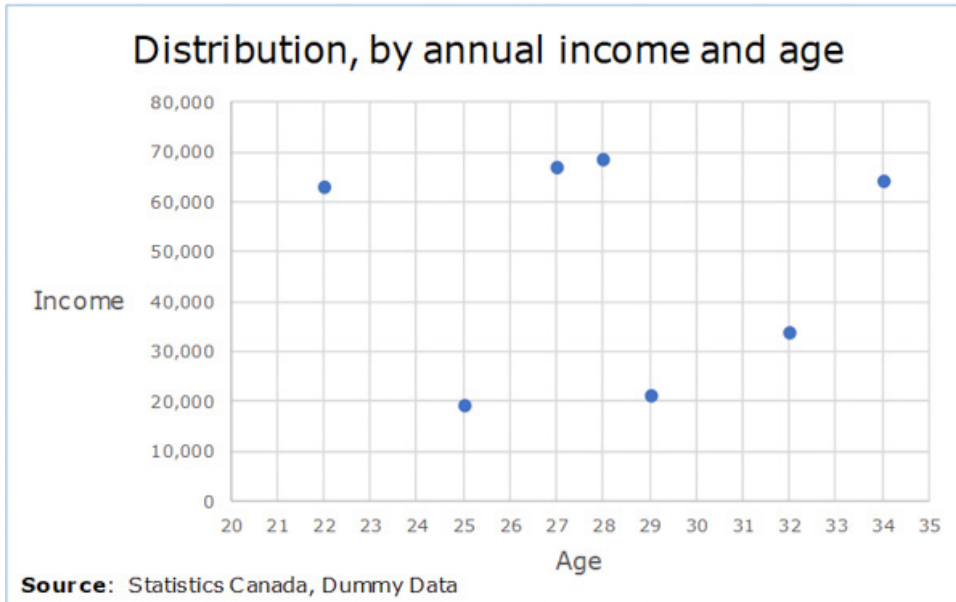


4. Point charts

Point charts are typically used to illustrate the trend or pattern of frequency distribution of variables. They usually have an additional element, i.e. a regression line which show the estimated slope of a model.

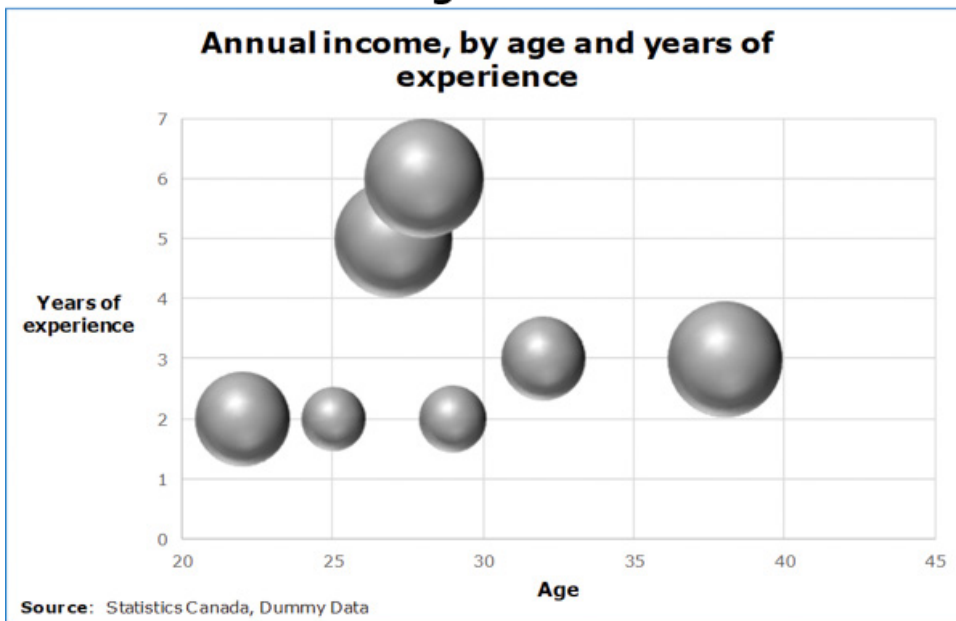
- It is important that the different dots can be clearly distinguished from each other.

Figure 39



- Dots of different sizes can be used to represent different values. If there is overlap, the smaller dots should be placed in front of the others so that they are visible.

Figure 40



- The relative size of dots can be difficult for readers to estimate.

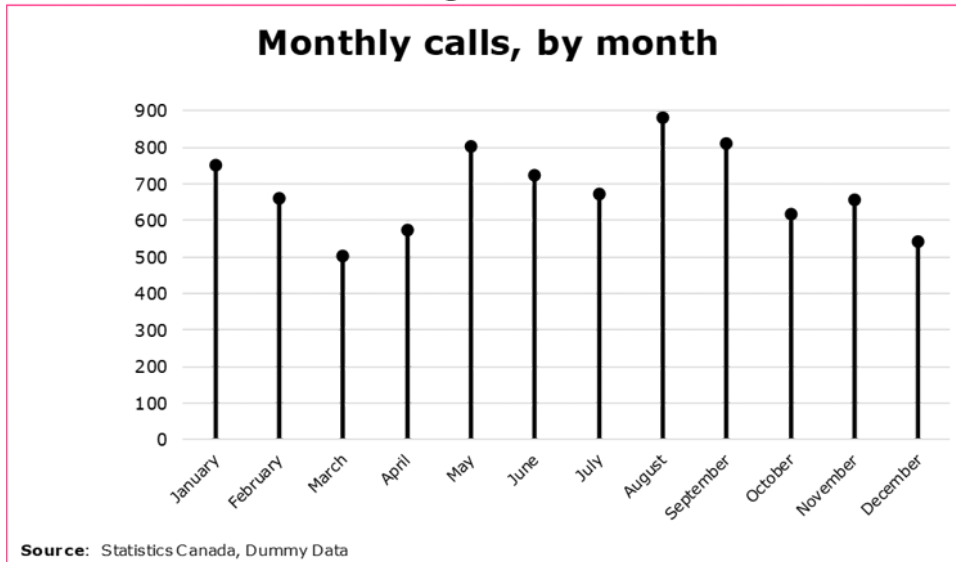
- Point charts draw the reader's attention to the dots (and their values). If your goal is to illustrate a trend, it is best to connect the dots and use a line chart instead.
- Shading around the points can be used to show uncertainty.

Point charts family

a. Lollipop graph

This chart is similar to bar charts. It uses a line for visualizing the values of each variable instead of a bar.

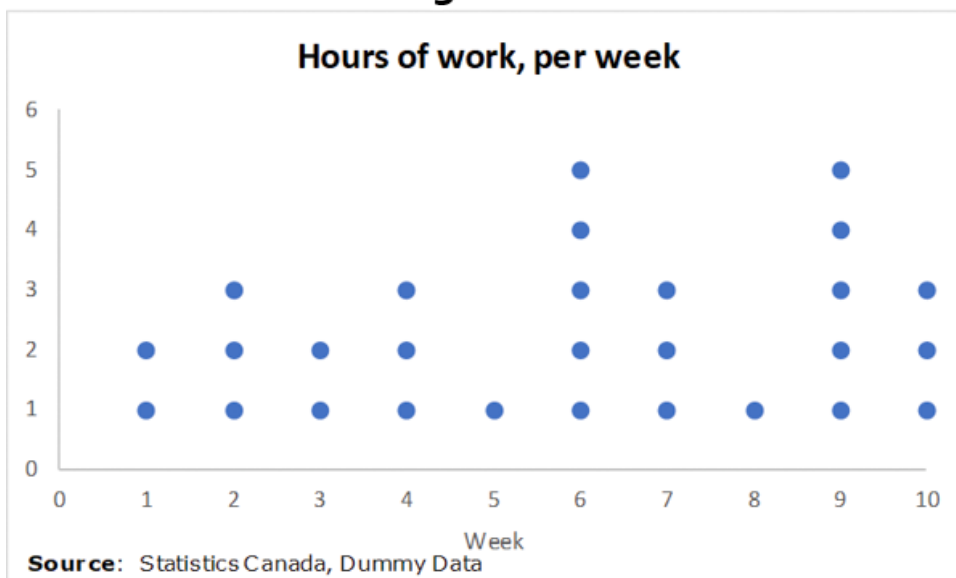
Figure 41



b. Strip plots

Strip plots display the value of each points in a data set. They are useful for visualizing the precise value of each elements in a small data set.

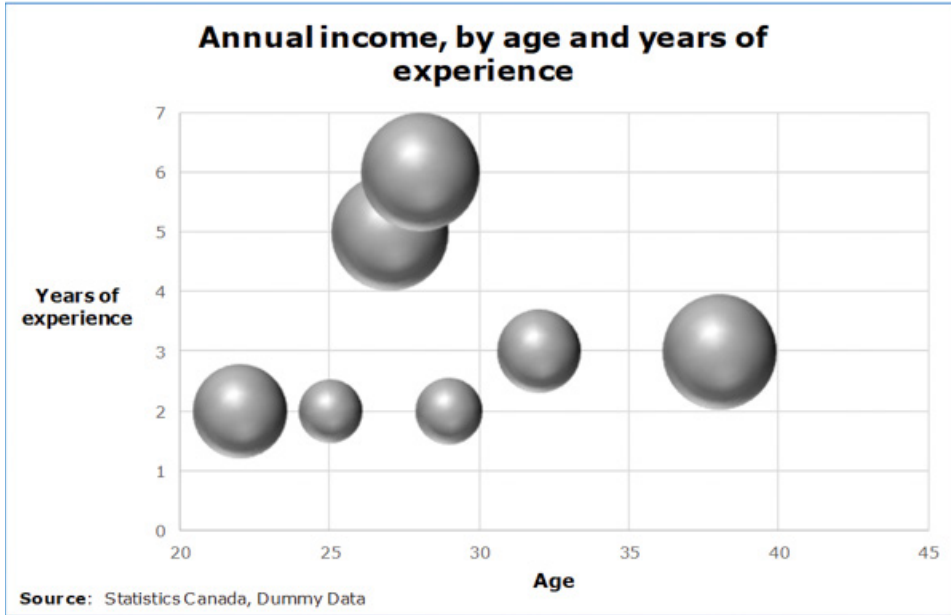
Figure 42



c. Bubble plots

With bubble plots, it is possible to illustrate a third element on the same chart, using bubbles that vary in size depending on the value.

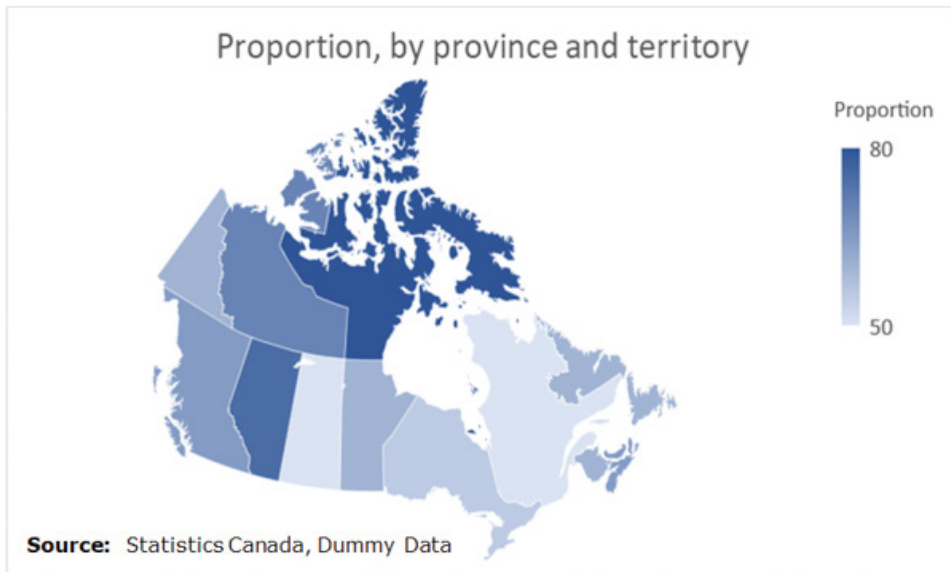
Figure 43



5. Maps

A map is a geospatial design that displays information on geographical locations.

Figure 44



The “choropleth map” uses colors to provide information: different shades of color are assigned to defined regions such as countries, provinces, and cities.

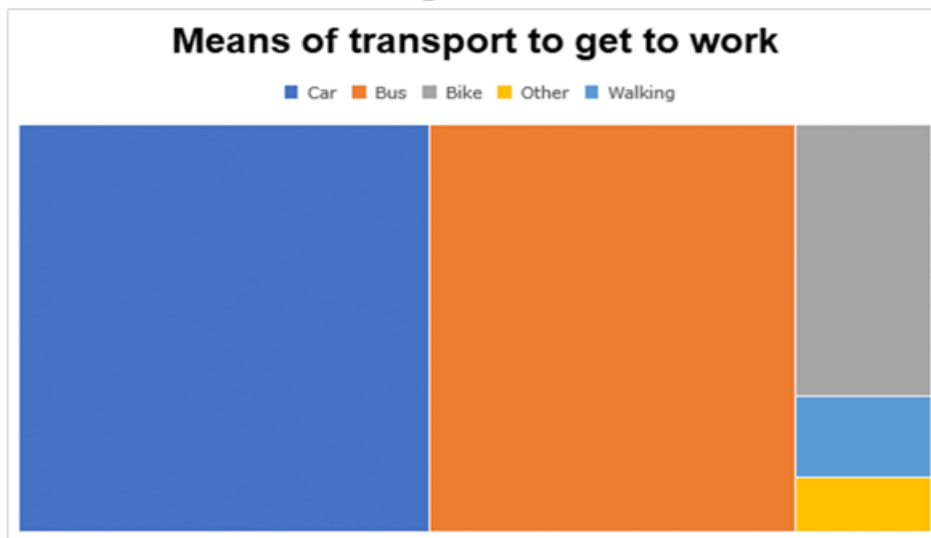
- With choropleth maps, it is preferable to use relative measurements to assign colors, rather than absolute measurements.
- On choropleth maps, it can be difficult to distinguish small areas, such as Prince Edward Island on a map of Canada.
- It is not possible to show uncertainty on the map.

Maps family

a. Tree maps

This type of maps uses rectangles proportional to the relative size of each category to illustrate them.

Figure 45

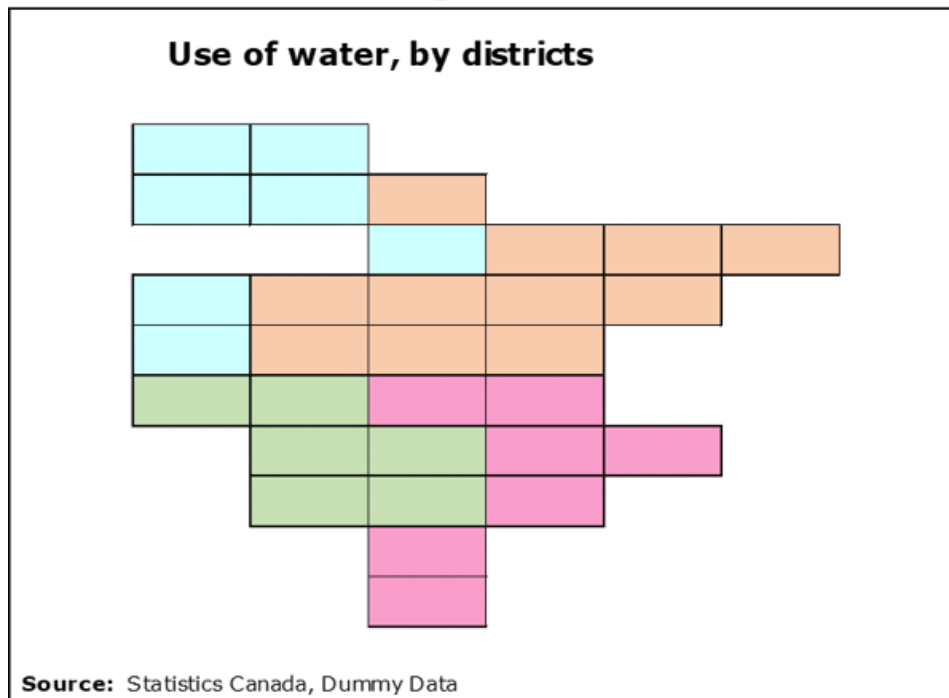


Source: Statistics Canada, Dummy Data

b. Tile grid maps

In a traditional geographical map, the size of each area has some effects on how we process the information. In tile grid maps, elements with same sizes and shapes are used and the audience can see and process the information without the side effect of element size on their judgement.

Figure 46



Glossary

Base value

The base value (also called “baseline”) is the natural starting point of a variable. Usually the base value of variables is zero, but there are cases where the logical base value is different (e.g. the price index has a base value of 1).

Uncertainty

In statistics, uncertainty refers to the fact that estimates based on a sample or projections may not reflect the true value.

Continuous variable

A continuous variable can take all possible values in a predefined range, as opposed to discrete variables, which can only take certain values in a range, usually integers.