Data visualisation

12

Edwin de Jonge

Explanation of symbols

.	data not available
*	provisional figure
**	revised provisional figure (but not definite)
x	publication prohibited (confidential figure)
–	nil
–	(between two figures) inclusive
0 (0.0)	less than half of unit concerned
empty cell	not applicable
2011–2012	2011 to 2012 inclusive
2011/2012	average for 2011 up to and including 2012
2011/'12	crop year, financial year, school year etc. beginning in 2011 and ending in 2012
2009/'10–2011/'12	crop year, financial year, etc. 2009/'10 to 2011/'12 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Table of Contents

1. Introduction to the theme

1.1 General description and reading guide

1.1.1 Description of the theme

Data visualisation is the art and skill of presenting data in a visual manner such that the information contained in the data becomes apparent.

There are many examples where visualising data provides insight that is difficult to express or cannot be expressed in another way. This report discusses a number of these examples.

Data visualisation is a useful tool for all stages of the statistical process. Traditionally, data visualisation has had at least two major areas of application in statistics: data exploration and communication.

In data exploration, data visualisation is a tool for statisticians who are analysing the data. By making the data visual, the analyst obtains insight into the structure of the data. For example, an impression can be quickly obtained into the distribution of a variable, or an outlier can be rapidly detected (while it would remain virtually invisible in a numerical summary). In addition, depicting the regression line from a dataset can give an impression of how "correctly" the regression approximates the data, or a relationship between the variables can be suggested. In this context, data visualisation is mainly an explorative and supporting tool to study the data.

In communication, data visualisation is highly suited to showing a relevant pattern in the data to a broader public in a single glance. Just like statistics, data visualisation provides a summary of the data. However, the result is not a key figure or a measure, but a diagram.

The distinction between exploration and publication is an artificial one. A good diagram that can be used for communication purposes is often the result of an analysis, for which this diagram has already been used. Conversely, developing a visualisation for presenting statistical data can result in a further analysis of the data. However, it is still useful to draw this distinction because the nature of a selected visualisation form is different. In publication, much more attention must be paid to the layout and the readability of a diagram because it is intended for non-experts. In many cases, non-experts will understand diagrams faster than the statistical story behind them.

This report mainly discusses the second application described above. It provides guidelines for creating a good diagram or maps to communicate statistical data.

The content of this report does not constitute new information; it is mainly a compilation of standard publications about data visualisation, such as Chambers et al. (1983), Tufte (1983), Tukey (1977) and Wallgren et al. (1996); international guidelines for creating diagrams and maps, such as UNECE (2009); and previous Statistics Netherlands publications (in particular, Bethlehem, 2002). The subject is broad and deep enough to serve as the subject of many books. This report is therefore limited to the main issues, and it provides references to the relevant literature where possible.

The rest of this chapter discusses the objective of this document and the subject's place in the statistical process. Chapter 2 provides a short historic overview of the use of data visualisation methods, while Chapter 3 sets out general guidelines that should be followed when designing a visualisation. We summarise these guidelines in chapter 4. Different diagram goals and types are discussed in chapter 5. And, finally, the use of thematic maps is addressed in chapter 6.

### 1.1.2 Goal, problems and solutions

The goal of data visualisation is to make the information contained in the data clear. There are many examples where visualising data provides insight that is difficult to express or cannot be expressed in another way. We discuss several of these examples in chapter 2. Another well-known example is examined in Cleveland (1993). There exists a dataset from Minnesota from the early 1930s about a field study on the growth of the barley crop. Ten types of barley were cultivated at six different sites, two years in a row. This therefore relates to 10 x 6 x 2 observations. These data were analysed by different statisticians, including the famous statistician and founder of modern statistics, Fisher, as well as by Anscombe and Daniel.

These analyses missed a striking aspect of the data: the Morris site is the only site that had better results for 1931 than for 1932. It is striking that the extent to which this happened corresponded with the difference for the other sites. It later emerged that the analyses missed an error in the data: for the Morris site, the data for the years 1931 and 1932 were reversed[1]. If the diagram from Figure 1 had been used, then this problem would have been clear: here, we can easily see that the data for the Morris site deviate from the other sites.

---

[1] A detailed description of how this was possible is provided in Cleveland (1993, p. 328-339). Briefly summarised, the agricultural experts were more focused on the analysis method than the data itself. Anscombe had combined the site and the year into a single categorical variable, as a result of which the error was not noticed. Daniel had identified a number of data anomalies in the table, but not the one for the Morris site.

Figure 1. Barley data
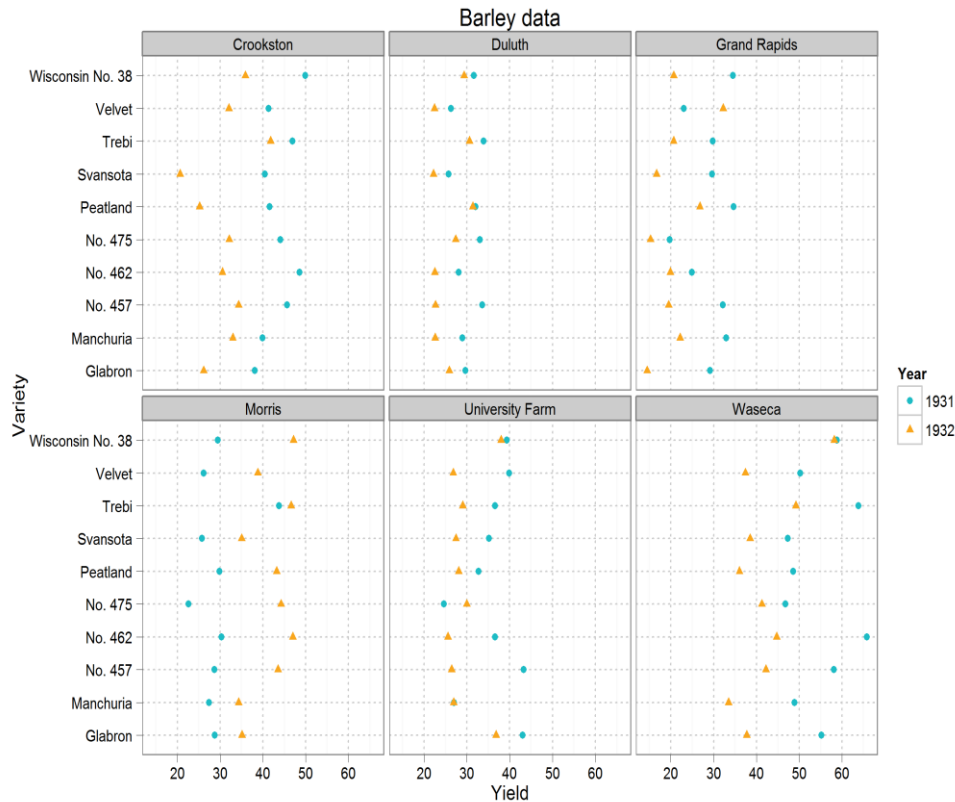
Creating a good diagram requires expertise. The first requirement is expertise about the data to be presented. After all, the goal is to communicate an interesting pattern in data. For example, data can contain artefacts that are important for presenting those data (content-related knowledge). In addition, knowledge is required about the presentation method and which pitfalls must be avoided. Issues that can affect how a visual representation is experienced are: the use of colours, colour scales, shapes, direction, size, markings, boxes and movement. An important example is the perception of "size". People interpret surface areas as "size". One of the results of this perceptual understanding is that the use of colour maps should be reserved to variables that represent density. The surface area of a region in combination with a colour provides an accurate interpretation of the value for that area. In this report, we only name a very limited number of perceptual preconditions. For a more detailed description, please refer to Few (2004), for example. This document contains guidelines for creating diagrams and maps. It is thus a tool for visualising Statistics Netherlands data.

However, things often go wrong when creating diagrams and maps. Diagrams are often confusing or do not say anything. This report contains several examples of diagrams where something has gone wrong. This is often accidental, but it is also sometimes done on purpose to exaggerate or obscure an effect. This document offers further guidelines that help prevent these errors. The incorrect representation of data is a way of misleading people. The well-known book "How to Lie with Statistics" (Huff, 1954) contains examples of many of these misleading diagrams. An incorrect

diagram can introduce a perception bias: the data are experienced in a way that is different, often less neutral, than necessary. Being aware of this has two advantages: (1) it helps us to prevent this situation when creating diagrams, and (2) it helps us to correctly interpret diagrams.

1.2 Place in the statistical process

Data visualisation can be utilised at any stage in the statistical process.

It can be used to investigate the quality of the source material, such as the amount of non-respondents and missing values. It can also provide an overview of the distribution of variables in registers in order to detect special or deviating subpopulations. Another possible use is to show differences in data before and after editing and imputation, or in the macro editing of outliers. Furthermore, it is a very useful tool in data analysis and output. This report mainly addresses data visualisation for statistical output and, to a lesser extent, for statistical analysis.

Data presentation plays the central role in statistical output. A good visualisation makes clear, in a single glance, what is going on. Diagrams and maps are accessible, and must be understandable for many end users. Diagrams and maps also tend to be reproduced in various media.

A new trend in the visualisations for communication involves offering dynamic graphics on statistical bureau websites. New aspects of such visualisations are the use of interaction and animation; however, the other rules still apply. The introduction of interactivity means that visualisations are becoming increasingly analytical: the user is given the opportunity to customise the diagram, for example, by selecting data or by choosing a different diagram type. As a result, a user becomes more of an analyst who independently decides what data is shown and how. For an overview of interactive web visualisations, please refer to Ten Bosch and De Jonge (2008). An information dashboard is a special form of dynamic graphics (such as the Inflation Dashboard ECB). An information dashboard combines multiple dynamic graphics, and each of them provides a different view of the data. If a user selects data in one of the graphics, this selection is also visible in the other graphics. This is known as linking. Dynamic graphics offer users the opportunity to explore the data themselves. These dynamic graphics also facilitate the visualisation of larger datasets, because users themselves can select the data subsets. At this point, the distinction between exploration and publication becomes fuzzy.

The functional use of animations is still in the early stages, but it is very promising. Animation is primarily used to show time series in which "time" is intuitively added to a diagram. The guidelines in this report also apply for interactive and animated visualisations, but do not comment on the use of animation and interactivity at the present time.

1.3 Definitions

Concept	Description
Bar chart	Diagram in which the values are represented using vertical or horizontal bars.
Choropleth map	A map in which the values are represented by coloured/shaded regions.
Component bar chart	Bar chart in which different components are shown alongside each other.
Data visualisation	A visual representation of statistical or other data.
Diagram	Specific form of data visualisation. Two-dimensional statistical or other data visualisation.
Dot map	Map in which the values are represented by a number of dots.
Dynamic graphics	Interactive, animated graphics, found increasingly often on websites.
Graphics	See diagram.
Information Dashboard	Combination of multiple graphics for the same data set. Often offers the option of linking.
Line chart	Diagram in which a line depicts the development of values.
Linking	Showing a selected observation in an Information Dashboard.
Map	Geographic data visualisation.
Occlusion	The problem that the representation of one observation (data point) hides other observations.
Pie chart	Diagram in which the values (in fractions of a total) are depicted as pieces of a pie. Also known as a circle graph.
Proportional symbol map	Map in which the values are represented using symbols scaled in size.
Radar plot	Diagram in which the values of an observation are shown in a structure that resembles a spider web.
Scatter plot	See scatter diagram.
Scale error	An incorrect visual representation of data in which the visual representation does not correspond to the ratio between the separate data.
Scatter diagram	Diagram in which the values are represented using points.
Stacked bar chart	Bar chart in which the bars consist of stacked components.
Treemap	Diagram to show data with a hierarchical structure.

2. History

The early history of the use of graphics has a number of especially good examples that clearly and effectively show the information present in the numerical data. They demonstrate the power of diagrams.[2] Here we show some striking examples. John Playfair (1759 - 1823) is viewed as one of the inventors of the statistical diagram. Playfair was the first to use the most important types of diagrams, namely the line chart, bar chart and pie chart. In 1786, he published a book that contained 44 diagrams that visually depicted economic variables. See Playfair (1786). Of the total of 44 diagrams, 43 relate to time series.

Figure 2. Price of wheat and level of wages, Playfair (1786)
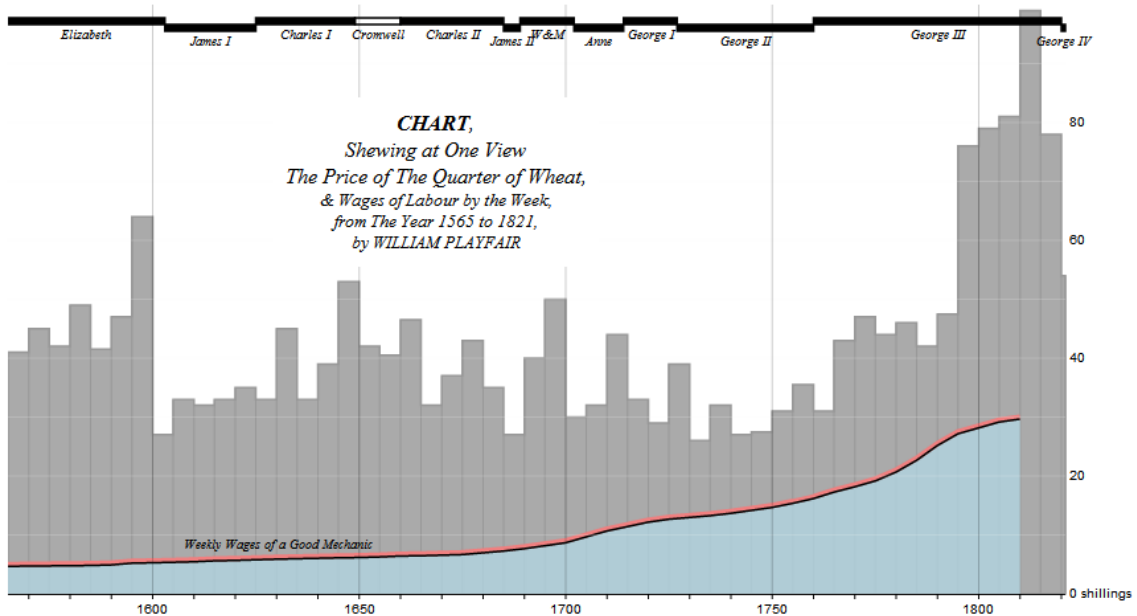
Figure 2 depicts the price of wheat and the level of wages from 1565 to 1821. The top of the diagram also shows the reign of the British monarchs. The combination of a time series for the price of wheat and the level of wages gives an impression of the purchasing power in this period.

Diagrams can be a powerful tool for the explorative research of data. An excellent example of this is the diagram made by John Snow in 1854. Using a chart of the centre of London, he used dots to plot all the deaths due to cholera. He also indicated the location of the 11 available water pumps (marked with an 'x'). This map is shown in Figure 3 (see also Tufte, 1983).

---

[2] Much of this section was taken from Bethlehem (2002).

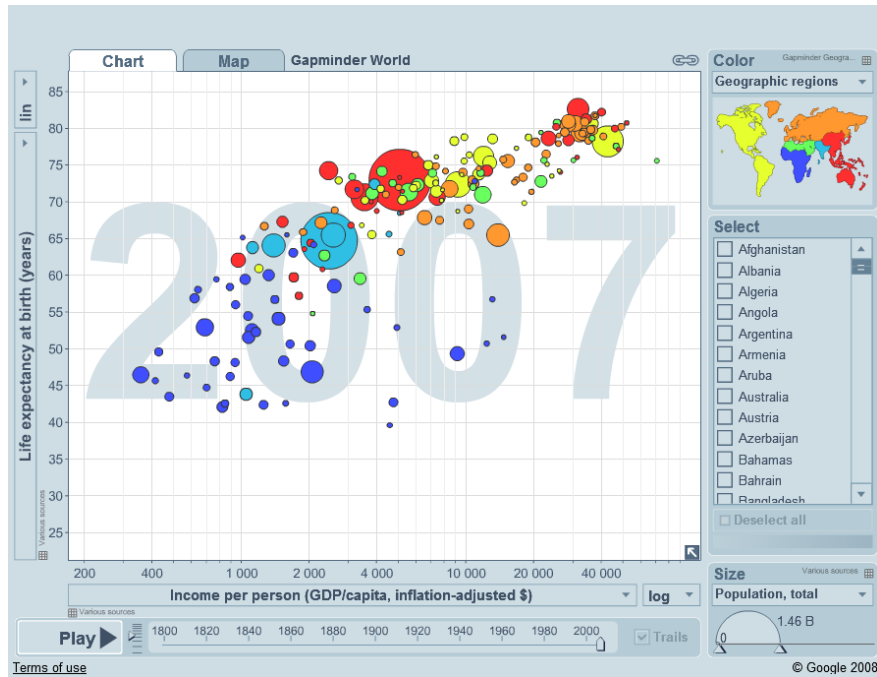*Figure 3. Deaths from cholera in London in 1864*

By studying the density distribution of the dots, Snow discovered that the deaths were concentrated mainly around the water pump in Broad Street. He suspected that the disease was related to water from this pump. The water supply for the pump ran over a cemetery. After this pump's handle was removed and it could no longer be used, the epidemic that had claimed more than 500 victims quickly came to an end.

A classic diagram is that of Napoleon's Russian campaign of 1812. This diagram was created by the Frenchman Charles Joseph Minard in 1869. The diagram, which is discussed in Tufte (1983) and Wainer (1997) and other sources, is shown in Figure 4.

The diagram depicts the fate of the French army after crossing the Polish-Russian border. The size of the army is indicated by the thickness of the grey line. The army initially consisted of 422 000 men when the campaign began. When the army reached Moscow in September, it had been reduced to 100 000 men. The army's return is shown by the black line. At the bottom of the diagram, the temperature during the return trip is also shown. The banks of the Berezina river were strewn with the bodies of 22 000 dead soldiers after the army crossed the river in November when the temperature dropped to -20 degrees Celcius. When the army reached the Polish-Russian border at the end of the year, only 10 000 men remained of the army's original 422 000 men.

Figure 4. Napoleon's campaign of 1812

Many experts consider Minard's diagram as one of the best statistical diagrams that has ever been made, not only because of the clarity with which this tragic story is told, but also because of the richness of the information that is available in a single illustration. Six variables are plotted: time, geographic position in the form of X and Y-coordinates, the size of the army, the direction in which the army was moving and the temperature. The values and the relationships between the different variables can be understood fairly easily from the diagram[3].

The introduction of the internet has given a big boost to interactive and animated visualisation. A good and well-known example is Gapminder (http://www.gapminder.org) (see figure 5). This visualisation gives the user the opportunity to examine multivariate data. The tool presents what is known as a bubble chart. On the x and y-axes of this bubble chart, we can see average income and life expectancy. The area of the bubble indicates the population size, and the colour of the bubble indicates which part of the world the country in question is located in. The diagram also has a time axis on which the data can be played out. Tracking can be used to follow an individual country over time: for each year, the associated point is plotted in this diagram. The Gapminder website also features videos in which Gapminder's designer, Hans Rossling, uses this tool to show very clearly how the different countries in the world have developed over the past 100 years.

---

[3] The compact representation of these six different variables also comes with a price: the "flow" of soldiers is not geographically accurate: the thickness and also the connections between the battle locations are not geographical. However, they do indicate the size of the army and the sequence of the battles respectively.

*Figure 5. Gapminder data visualisation on the internet*

A recent visualisation from the New York Times (2008) offers a representation of inflation in the United States (see fig. 6). What is interesting about this visualisation is that it shows how the different components of inflation individually affect the total inflation, and it also indicates what the inflation is for the separate components. For example, we can see that the price of gas (petrol) was high, and that this significantly contributed to the total inflation. The type of diagram used here is known as a treemap. A treemap uses surface area and colours to represent two variables (such as spending and inflation). In addition, a treemap shows the hierarchical tree structure of the data. This visualisation from the New York Times is an interactive visualisation: we can zoom in on separate products and their price development. The visualisation is one of the results of a doctoral research by Balzer and Deussen (2005).

Figure 6. Interactive treemap showing US inflation

An Average Consumer's Spending

Each shape below represents how much the average American spends in different categories. Larger shapes make up a larger part of spending.

Color shows change in prices from March 2007 to March 2008

-10% -2 0 +2 +4 +6 +10 +20 +40%
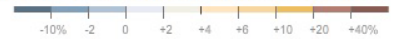
ZOOM IN   ZOOM OUT

Food and beverages 15%
The high price of oil is a factor that has made food prices rise quickly.

Miscellaneous 3%

Recreation 6%

Education/Communication 6%
Cellphones were added to the index in 1997. Because the Consumer Price Index can be slow to add new goods, which are often cheaper, it may overstate parts of inflation.

Housing 42%
In the C.P.I., home ownership costs track rent prices more closely than housing prices. This means inflation may have been understated when home prices were rising faster than rents.
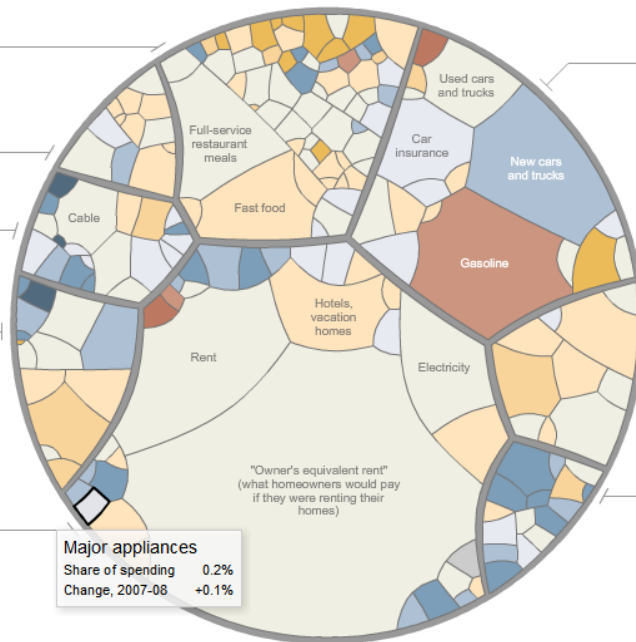
Transportation 18%
Gas is 5.2 percent of spending nationwide, but only 3.8 percent in the New York area.

Health care 6%
As a group, the elderly spend about twice as much of their budget on medical care.

Apparel 4%
The ratio of spending on women's clothes to that on men's clothes is about 2 to1.

Used cars and trucks

Car insurance

New cars and trucks

Full-service restaurant meals

Fast food

Gasoline

Cable

Hotels, vacation homes

Electricity

Rent

"Owner's equivalent rent" (what homeowners would pay if they were renting their homes)

Major appliances

Share of spending	0.2%
Change, 2007-08	+0.1%

Sources: Bureau of Labor Statistics; Michael Balzer, University of Konstanz (Germany)

Matthew Bloch, Shan Carter and Amanda Cox/The New York Times

13

3. General guidelines

This report describes the use of visualisation methods to show the information that is hidden in the data. When creating a visual presentation, the following preconditions should be kept in mind.[4]

- Determine the target group:
    - Expert users will understand the data better than non-experts.
    - A different audience probably needs a different presentation form. A different audience will often need additional context and will not be familiar with the statistics' major variables.
- How and where the message will be presented: will it be an extensive, detailed analysis in a report, or an interactive web page?
- Consistency over multiple data visualisations: ensure that the elements of multiple visualisations are consistent in a presentation and are presented in a clear and understandable manner. This particularly applies for the use of colour. In the event of multiple diagrams, use the same colour for the same subject.
- Possibility of misinterpretation: test the visualisation with colleagues or members of the target group to check whether the message comes across as intended.

An important name in the data visualisation field is that of Edward Tufte. He wrote, among other things, a standard work (Tufte, 1983) on data visualisation. His beautifully designed books contain his analysis of visualisation forms and his formulations of concepts and visualisation principles for this purpose. We will examine two of these concepts that help in determining whether a visualisation offers sufficient information.

The first concept is the percentage of ink in the diagram that is designated to the data (known as the data-ink ratio). Diagrams are often created by graphic designers who mainly want to make the diagram visually interesting and therefore add additional elements. However, the most important element in the diagram is the data. A good design has a high data-ink ratio, and dedicates only a minimum amount of ink to unnecessary elements.

The second concept is data density: the amount of data shown per cm^2. If there is only a small amount of data, then a table or text is often clearer than a diagram. A good design has high data density (see figure 7).

---

[4] These points were largely taken from UNECE (2009).

3.1 Determine the message

The most important step in designing a visualisation is to ask the question of what message should be transmitted to the users. What must be communicated, and how does the data visualisation help in this process? It may seem like we are stating the obvious here, but, unfortunately, it is a regular occurrence that diagrams and maps are designed only because they look nice.

Before designing a diagram or map, ask the following questions.

- What message should this diagram transmit? What do I want to say?

- What variables are important?

- What is the main pattern that is present in the data?

- Is there an unexpected or striking pattern in the data that is worth showing?

- Who is the data visualisation intended for?

The answers to these questions can help you to formulate a message that must be transmitted to the users through visualisation.

A difficulty in this context is that a dataset will often have multiple variables and a number of interesting patterns. Many analysts and subject matter experts fall into the trap of wanting to show all the possible patterns in the data. They know the data inside and out, but do not fully comprehend that their audience has less of an overview and less context to understand all the nuances of the data. The basic rule of effective communication is to limit yourself to the essential meaning of the data and then to transmit that particular message. So, during the design phase, make a choice as to which variable or which pattern is the most important to depict.

After completing the design, check whether you have answered the question that was asked at the start of the process. If not, redesign the visualisation or formulate a new message that fits in with the visualisation you have designed.
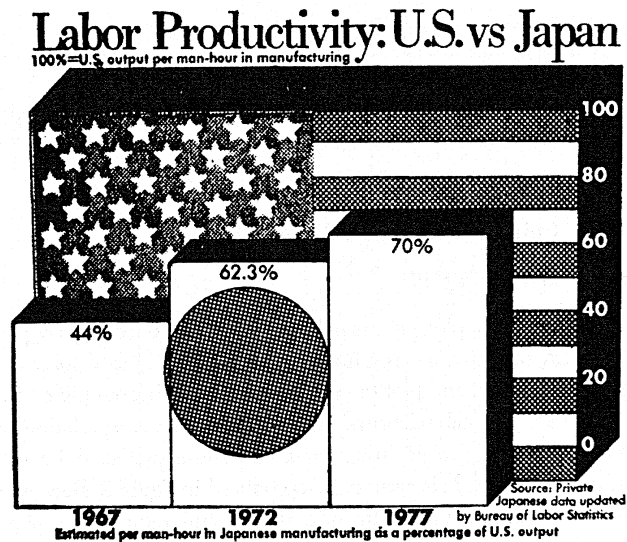
3.2 Focus on the data

There are many examples of diagrams in which the aesthetic aspect has determined the design. The most important underlying principle of data visualisation is:

Focus on the data and avoid anything that distracts from this.

While designing the visualisation, try to only show data, and limit anything that is not data to a minimum. In other words, limit yourself visually to the main idea of the data. It is a pitfall to make the diagram more visually attractive by adding all sorts of unnecessary yet visually interesting elements. Tufte (1983) refers to this as "chart junk". An example of chart junk is the diagram below that shows the labour productivity of the US vs. Japan.

Figure 7. Chart Junk

Labor Productivity: U.S. vs Japan
100%=U.S. output per man-hour in manufacturing

100
80
70%
60
62.3%
40
44%
20
0

Source: Private
Japanese data updated
by Bureau of Labor Statistics

1967   1972   1977

Estimated per man-hour in Japanese manufacturing as a percentage of U.S. output

This diagram has a low data-ink ratio and a low data density. The diagram contains only three numbers, but the majority of the ink is wasted on unnecessary visual elements that are distracting and do not add anything to the diagram.

3.3 Presentation form

Think about in advance which presentation form is best suited to transmitting the message: text, table, diagram and/or map? Sometimes a table is more suitable than a diagram or map. If the dataset is very small and only includes a few observations, then a table is generally preferred over a diagram or map. A diagram with one or two observations is a waste of paper (or of disk space). It is also makes no sense to make a map or a diagram if there is too little variation in the data. In this case, the message in such data can often be expressed more compactly and more clearly in words.

The following chapters delve into the design aspects of diagrams and maps.

4. Diagrams

4.1 Motivation

A diagram is a visual representation of statistical data, in which the data are represented using visual symbols. Statistical data are often understood more easily if they are shown in a diagram instead of a table. Human perception is better at recognising geometric figures than at interpreting a large amount of numbers.

Diagrams can often compactly reveal relevant patterns in large amounts of data. Diagrams are therefore a good method of communicating important findings in the data to the end users of the statistical data.

Diagrams can be used for:

- **Comparisons** of statistical data. Which population parameters of subpopulations are larger? Are there more men than women? Are employees in the metal industry younger on average than those in the wood industry?

- **Developments** over **time**: How does the value of a variable change? Does the inflation increase?

- **Frequency distributions**: How is a variable distributed? What is the age composition of the Dutch population?

- Showing the **composition**: Part-Whole: How do the parts compare to the total? What percentage of high school students go on to university?

- Showing possible **relationships** and **correlations** between variables. Do the assets of Dutch households increase with the average age?

- Showing **deviations**: How does the growth of the Dutch economy differ from average European growth?

In this chapter, we examine errors that can arise when making diagrams. We also describe a number of frequently used diagrams and offer guidelines for creating good diagrams.

4.2 Guidelines

A good diagram:

- catches the reader's attention;

- does not mislead the reader;

- depicts the data in a compact manner;

- makes it easy to compare the data and clearly shows trends and differences;

- illustrates the message that is hidden in the data.

These rules are largely based on the work of Tufte (1983), Schmid (1983), Wainer (1997), Few (2004) and UNECE (2009).

There are two fundamental principles that apply to diagrams:

- Ensure that the visual representation of quantities corresponds to their numerical value. In other words, ensure that the visual encoding is accurate. This means that the natural sequence and relationships of numerical values must be respected: a value that is twice as high must also be twice as large visually[5].

- Avoid representing data in three dimensions. The reason for this is that a 3D representation often introduces visual bias because of the use of perspective. In addition, large values can "block" the view of the smaller values.

It is good to keep these principles in mind: they are violated far too frequently and, as a result, it is easy for the audience or the users to interpret the diagram incorrectly. Of course, this may actually be the intention of the person who made the diagram. However, as a statistical bureau, it is our responsibility to be as objective as possible in the information we communicate, and we therefore should avoid such diagrams.

The two principles are further elaborated in the subsections below. We take a look at the different components of a diagram that can be employed incorrectly, and which can even be used to present data in such a way that the diagram is interpreted differently. Huff (1954) offers multiple examples in which diagrams are used to manipulate statistics.

4.2.1 Scale errors

Scale errors are errors in which the visual representation of the quantity to be represented does not correspond to the actual quantity. There are different types of scale errors.

The best known scale error is a scale in which the zero line is omitted from the diagram. If the minimum of the visual scale in a diagram is set to a value higher than 0, then the differences in values in the diagram are magnified. This can result in totally different diagrams, as becomes clear in Figure 8. By omitting the zero line from this figure, the development compared to the total volume is magnified.

---

[5] In other words: the visual representation must correspond to the scale used. If a diagram uses a logarithmic scale, then the visual representation must correspond to the logarithmic scale.

Figure 8. Example of omitting the zero line
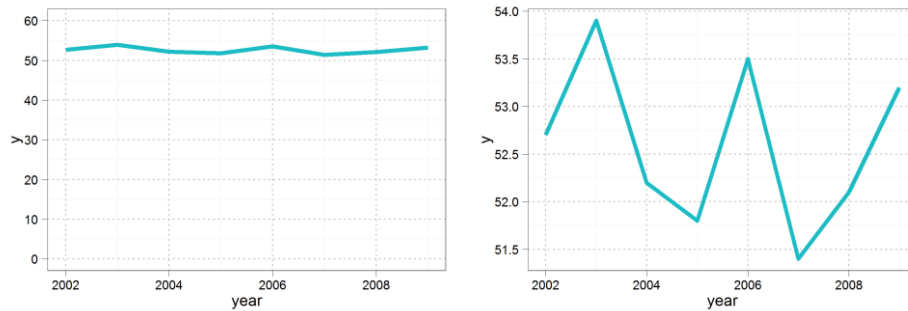
If the variation of the values is small compared to the absolute values, then it may seem as if a variable shows huge variations, while the variable is actually almost constant. For bar charts, omitting the zero line is definitely not recommended (Few, 2004) because users of the diagram interpret the surface area of the bars as values. If the zero line is omitted, the surface area of the bars does not correspond with the values that they represent. For line charts, omitting a zero line can be useful in isolated cases, primarily if the focus lies on the changes in a variable.

Another known scale error is when a scale does not correspond to the distance between values on the scale. This error often occurs in time scales. On the x-axis, periods are marked with different step sizes: the scale is not equidistant. A good example of the bias that this can produce can be seen in the diagram where the income of doctors is compared with that of technical workers. The diagram originally appeared in the Washington Post, and is taken from Wainer (1997).

Figure 9. Incomes of doctors vs. technicians (Washington Post)
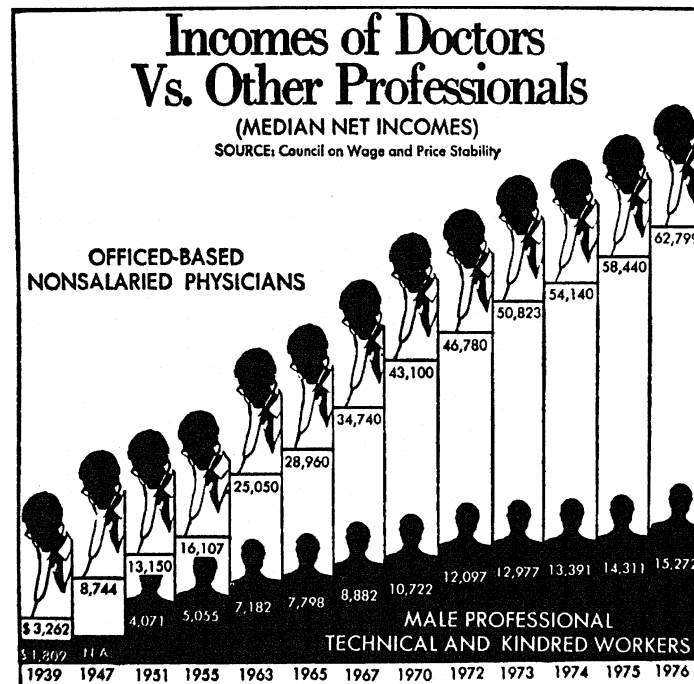
There seems to be a reasonable linear increase of salaries over the years. However, the time axis is not equidistant: the values on the axis make big jumps in the beginning and end with one-year steps.

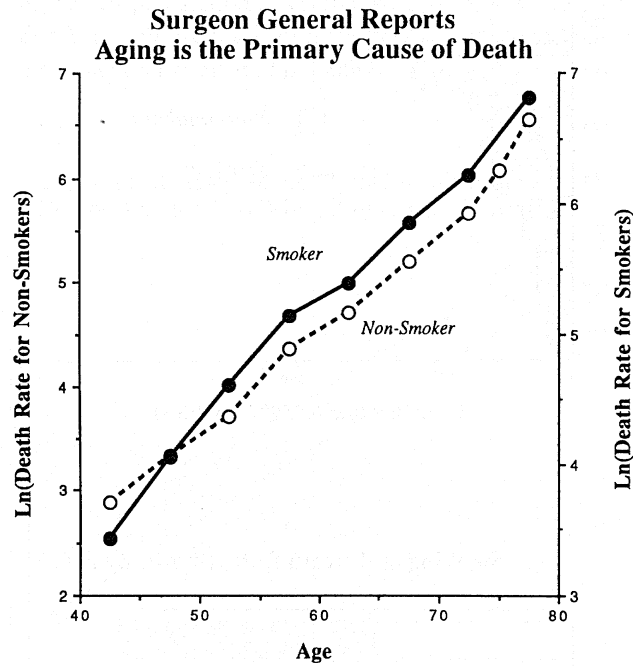*Figure 10. Correct representation of the income of doctors vs. technicians*
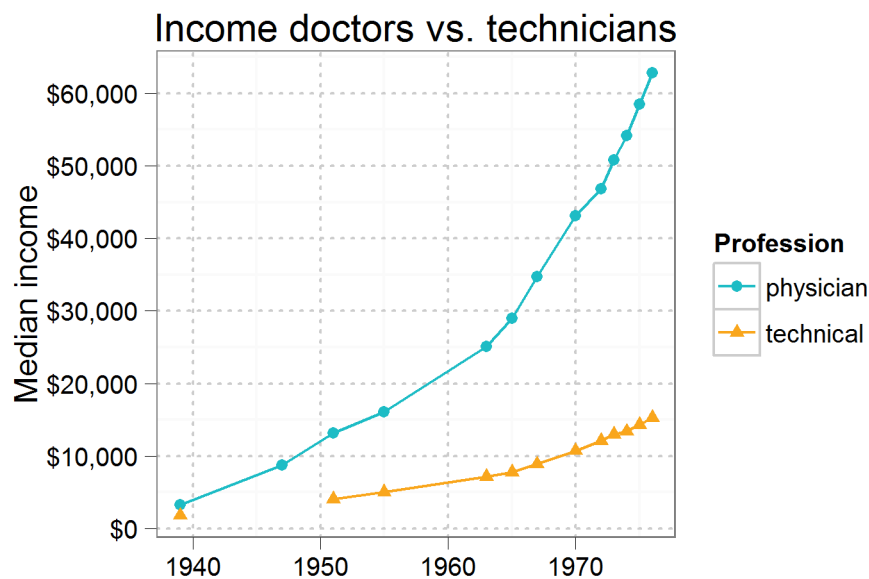
This becomes clear if we correctly represent the data as shown in Figure 10. Here, we see that the steps in time were incorrectly distributed in the original diagram, and also that the doctors' incomes have actually skyrocketed.

Another case of scale errors is the use of multiple scales in the same diagram. This should be avoided because such a diagram is difficult to interpret and can even lead to misinterpretations. Wainer (1997) provides a known example of this, which studies the relationship between death and age for smokers and non-smokers.

Figure 11 shows that the likelihood of death increases with age. There does not appear to be a difference between smokers and non-smokers. However, this diagram makes use of a double Y-axis. The Y-axis on the left is used to show the data for the non-smokers, and the Y-axis on the right is for the data for the smokers. Note that the scale distribution of the Y-axis on the left does not correspond to the Y-axis on the right.

Figure 11. Incorrect use of double y-axis

Surgeon General Reports
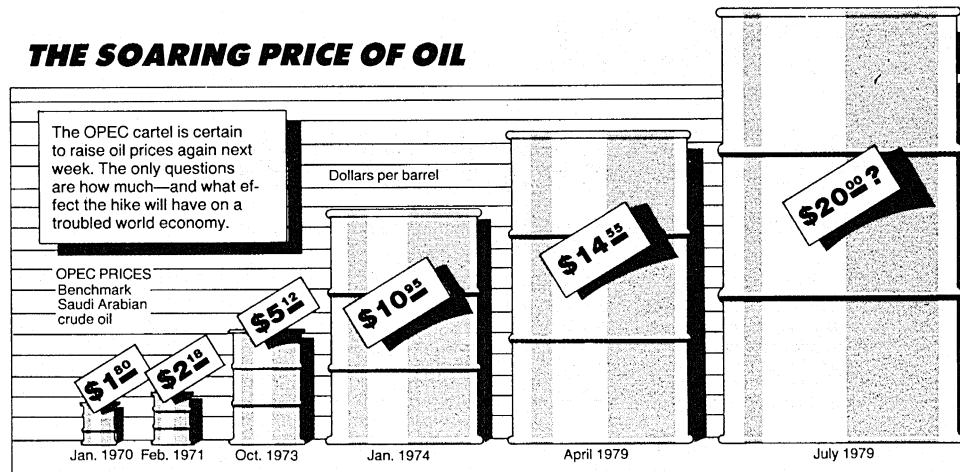**Aging is the Primary Cause of Death**

If the same scale is used, then the following diagram is the result (Figure 12). This diagram clearly shows a difference between smokers and non-smokers.

Figure 12. Correct representation of smokers and non-smokers

Death Rate (Log Scale) Plotted Against Age
Prospective Study of Mortality in U.S. Veterans
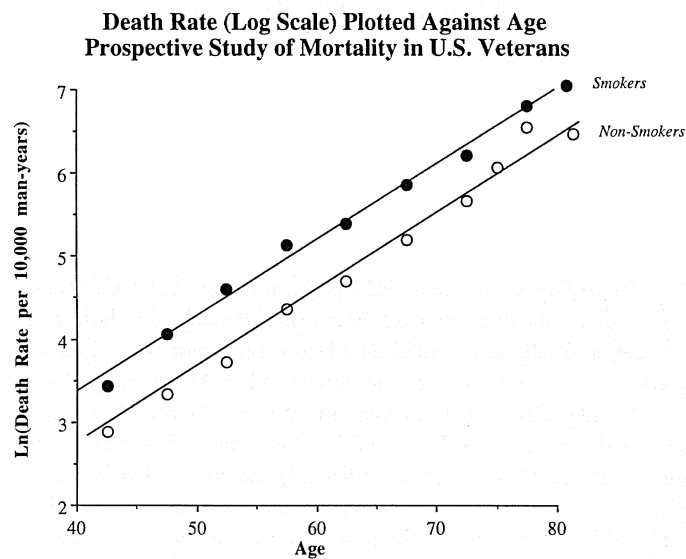
Another major scale error is the use of symbols in a bar chart. Instead of bars, a symbol is scaled in height with the value of a variable. A typical example of this is discussed in Schmid (1983). The diagram, shown in figure 13 below, attempts to depict the rise in the oil price in the years 1970 to 1979.

Figure 13. Incorrect scales of symbols

THE SOARING PRICE OF OIL

The OPEC cartel is certain to raise oil prices again next week. The only questions are how much—and what effect the hike will have on a troubled world economy.

Dollars per barrel

OPEC PRICES
Benchmark
Saudi Arabian
crude oil

$1.80 $2.18 $5.12 $10.95 $14.55 $20.00?

Jan. 1970   Feb. 1971   Oct. 1973   Jan. 1974   April 1979   July 1979

The height of the barrels increases proportionally with the price of oil. However, the problem in this diagram is that the user looks at the surface area of the barrels and not at their height. The ratios between the barrel surface areas do not correspond with the ratios between the associated oil prices.

Also note that there is clearly no equidistant scale distribution on the time axis. This produces an incorrect impression of the development of the oil price over time.

4.2.2 Avoid data objects with 3D perspective

Oftentimes, diagrams are considered boring, and designers feel it is necessary to make the diagrams more appealing. One of the ways chart junk is created is by adding a three-dimensional perspective. The use of this perspective does not create any added value whatsoever. In fact, it causes the data-ink ratio to drop: more ink is used for the same amount of data. The values in the diagram are also more difficult to read than in the normal bar chart variation. For example, in Figure 14, it is difficult to compare the values of A and C: the initial impression is that the bars are the same height. However, by looking more closely at the horizontal lines, we can see that A is larger than C.

Figure 14. Bias due to a 3D diagram
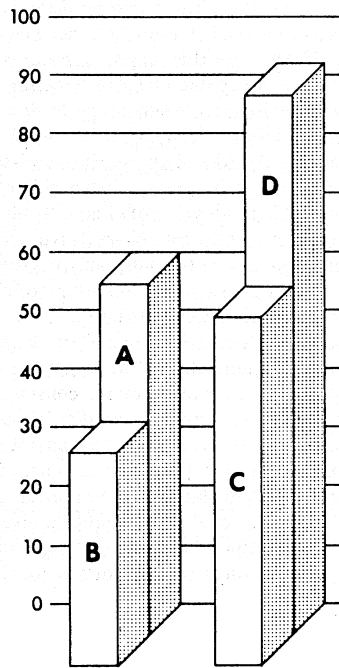
Two other problems with 3D are occlusion and perspective. In occlusion, a larger value in the foreground hides smaller values that are positioned more towards the back. Compare this to a mountain landscape: if we are standing next to a mountain, it is impossible to see the valleys and the smaller mountains on the other side of the mountain. Occlusion mainly occurs in 3D bar charts. Perspective ensures that values that are farther away appear smaller. Again, compare the situation with a mountain landscape: it is hard to estimate the height of mountain tops that appear in the distance. A contour map is always the preferred choice for obtaining a good impression of a mountain landscape, and this type of map is two dimensional (2D).

The use of three-dimensional diagrams is therefore not recommended. A possible exception is an interactive three-dimensional scatter plot, where the user can navigate through the "diagram space". However, the added value of such a visualisation is unclear in the professional literature: it mainly depends on whether the users understand the navigation aspect.

4.3 Diagram components

The most important component of a diagram is the data. Besides the data shown, a diagram also contains components that are needed to understand the data: title, legend, labels, grid lines, footnotes, etc.

For a diagram to be well understood, it must contain the following elements:

- The title of the diagram must indicate clearly, briefly and concisely what this diagram is about. A title can be informative or descriptive.

  - An informative title provides all the information that is needed to understand the data.

- A descriptive title offers a description of the main pattern that is visible in the data.

- The names of the axes must indicate what is found on these axes. Horizontal titles are desirable because they improve readability.

- Grid lines can be added so that users can easily read the values in the diagram. Do this in moderation, and ensure that the grid lines are not too prominent. They are a support element: the data must receive the most visual attention.

- Legends and texts must clearly indicate which data elements are represented in the diagram. Use a legend only if this is necessary to understand the diagram.

Avoid the use of decorative elements that do not improve the comprehension of the data.

5. Diagram objective and type

There are many types of diagrams. In recent years, new diagram types have arisen. Which diagram type should be used depends on the goal of the visualisation. We discuss the most frequently used types of diagrams in this section. The breakdown we use here comes from Few (2004). The appendix of this document includes a diagram that uses a variation of this breakdown.

The rest of this chapter explains the different goals and types of diagrams used for these purposes.

For the reader's convenience, the table below provides a summary of the goals with the associated diagram types.

Goal	Diagram type	Key points
Comparison	Bar chart (horizontal and vertical)	• Zero line scale errors • Sort bars by size
Time series	Line chart, bar chart, radar plot	• Equidistant time steps • Focus on absolute value or on development?
Composition	Stacked bar chart, multiple bar chart, pie chart, treemap	• Is comparing part vs. whole important? (pie chart) • Or is comparing components important? (bar chart) • Is the development of the whole or of the components important? • Is there a hierarchical structure? (treemap)
Deviation	Bar chart + line chart	• Plot deviation vs. reference instead of variable • Plot a reference line
Distribution	Histogram	• Select a good step size for the classes
Correlation	Scatter diagram, bubble chart	• If possible, add a regression line (scatter diagram) • Occlusion occurs if there is a large amount of data.

5.1 Comparison

In comparison, the goal is to compare the values of classes (or variables of classes) with each other. For example: "Are there more women than men with a certain characteristic in a certain reference period?"

This is a goal that occurs frequently in official statistics. A large amount of the Statistics Netherlands data contains qualitative variables. An example of a qualitative variable is gender, which leads diagram that shows the difference in values for men and women. Another example is a diagram in which the number of residents is shown for the 12 provinces of the Netherlands. There is often no natural order that can be assigned to the values in the qualitative variable.

Typical forms to show these are a vertical column chart and a horizontal bar chart.

First try to make a horizontal bar chart. In most cases, this is the clearest way to represent the values of a variable.
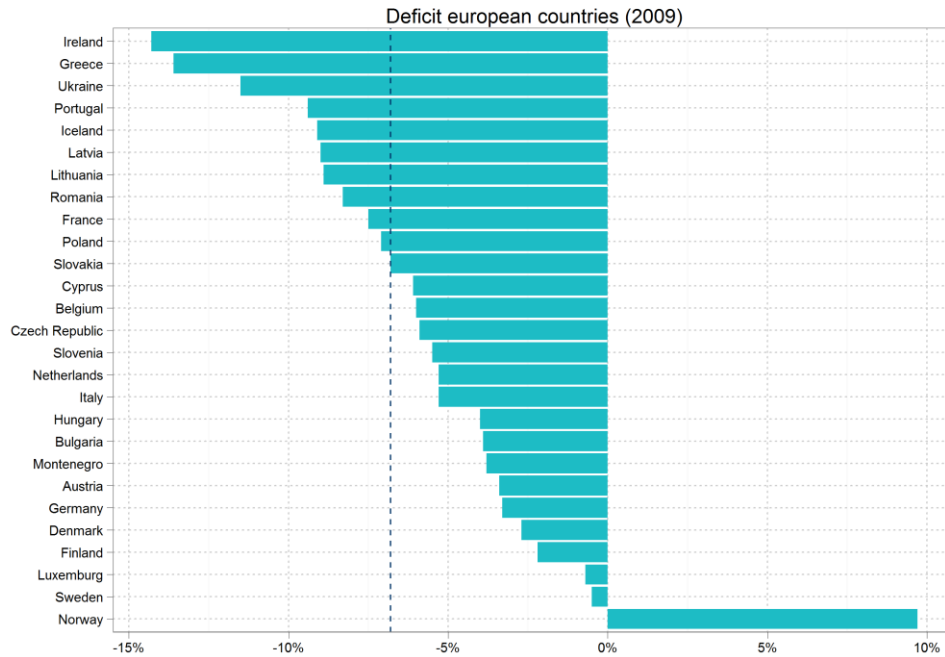
5.1.1 Bar chart

In a bar or column chart, bars are used to represent the value of variables. The length of a column indicates the value of the variable.

A column or bar chart enables the different values to be quickly compared. We can see in a single glance, for example, that the number of women in the Netherlands is larger than the number of men. A horizontal bar chart is preferred because the names of classes are easier to read in this form, and the diagram does not tend to be interpreted as a histogram (see later in this document).

If possible, it is good practice to sort the bars by size. This is a form of ranking. The order in the diagram shows immediately which value is larger. An example of such a diagram is presented in Figure 15. In this diagram, the reader can immediately see how the budget deficit in each country compares with the others.

Sorting the diagram by the value of the variable is recommended in most cases, except if there is a natural order; for example, an age classification.

Figure 15. Bar chart with budget deficits for EU countries, 2009

If there are multiple variables and many categories, a bar chart tends to become unreadable. In that case, it is better to use a table with values of the variables along the x- and y-axes and with a bar chart in each cell of the table.
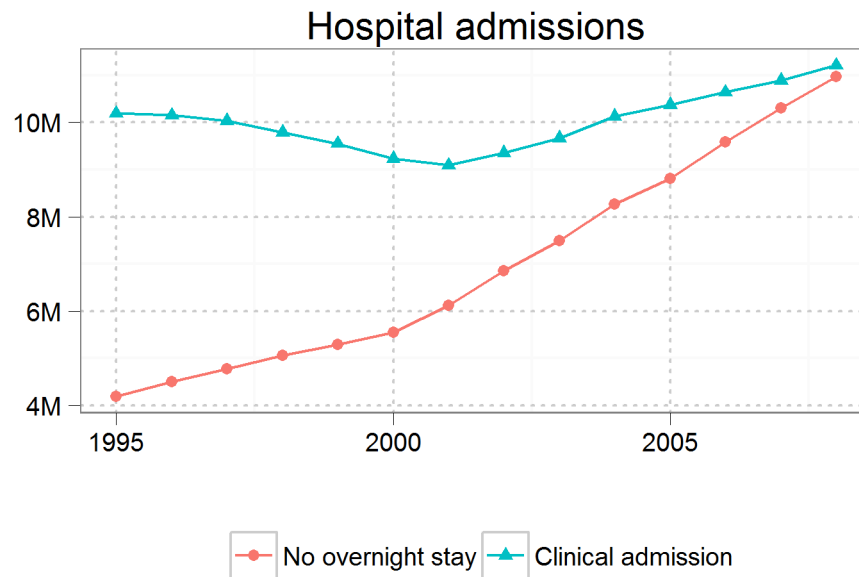
5.2 Time series

The goal of a time series is to clearly depict the development of (the value of) a variable.

A time series shows the development of the values of a variable over time. It is a comparison for the same variable over multiple periods. This means that one of the axes of the diagram has a time scale that indicates which time units are represented: years, quarters, etc. An important aspect of time is that points in time have a specific order, and that the time series must show the time chronologically. Furthermore, it is important that a consistent step size is used on the time axis (see section 4.2.1).

5.2.1 Line chart

The most frequently used diagram for representing a time series is a line chart. The primary goal of a line in a time series is to show the order of the data points: the line forms a timeline which contains the points in the diagram. This is also the reason that a scatter diagram (see section 4.2.1) is not recommended for a time series: it is less effective. The convention is to use the x-axis to show the time scale, where the time runs from left to right.

Figure 16. Time series of hospital admissions

Hospital admissions

5.2.2 Bar chart

Another frequently used diagram type for a time series is a bar chart. A bar chart places more emphasis on the size of the values measured than on the development of the values over time. If there are only a few values (< 4) and there is little to say about the development of the variables, then it is a good idea to use a bar chart. If there are many observations over time, then a line chart is preferred.

5.2.3 Radar plot

If a cyclical period is involved (months of the year, days of the week) then a radar plot can be helpful. Radar plots are also known as spider plots, spider charts, web charts or star charts. In a radar plot, the values of multiple variables are plotted in a web-like structure. Each variable forms a spoke, in which the value moves from 0 to 1. The observed value for this variable is standardised and plotted, and the points are then connected by a line.

The radar plot works if there are outliers, and if two data sets are being compared in which one of the two consistently scores lower than the other. In other cases, a radar plot can be confusing.

A radar plot has other disadvantages. The variables in a radar plot do not have a natural order, but the order used is very important when drawing the radar plot: it determines the form of the spider web. A high value followed by a large value draws attention to both values. A different order leads to a totally different form: for example, the values 1, 1, 1, 9, 9, 9 vs. 1, 9, 1, 9, 1, 0, 9 (see Figure 17).

Figure 17. A radar plot is sensitive to the order of the variables
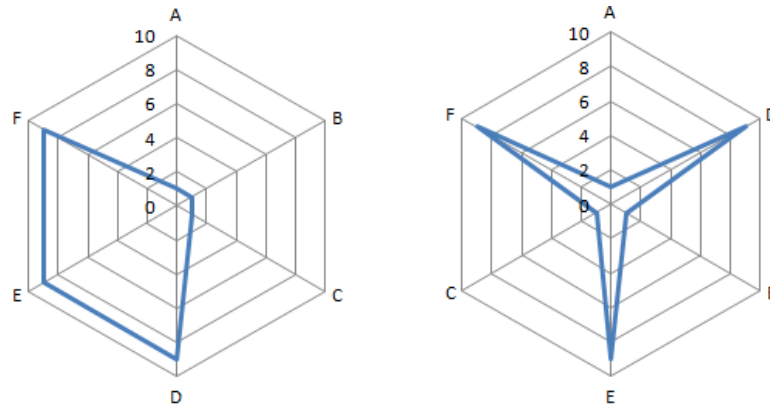
5.3 Composition: Part-whole

The goal of composition is to show how parts compare to the whole. This is often done by using percentages, for which the total is 100%. It is therefore clear how each part relates to the total.

Often, when creating the diagram, an attempt is made to show the relationship of the parts to the whole, and also the relationship between the various parts. This is sometimes possible. However, always keep in mind which of the two is the most important, because it is often better to make two separate diagrams instead.

For official statistics, a part-whole diagram is often interesting. For example: What percentage of Dutch residents lives in the ten largest cities? Or: How much energy do households use compared with the total energy consumption?

5.3.1 Pie chart

A pie chart is a frequently used diagram to show part-whole relationships. The pie pieces or sectors show the share of the total.

Figure 18. Pie chart of revenue from national taxation
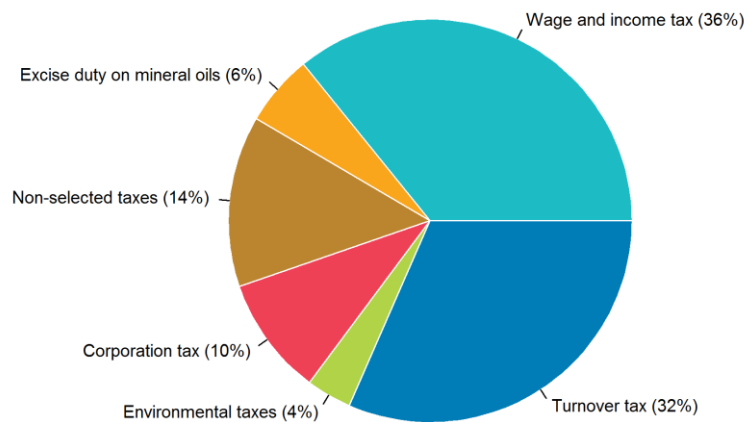
Pie charts are used often but they do have several problems, which is why they are not popular with visualisation experts. A problem of pie charts is that the sectors are difficult to compare with one another. This means that pie charts cannot be used effectively for comparative purposes[6]. For example, in Figure 18, it is difficult to see if the Wage and Income taxes are larger or smaller than the Turnover tax. Another problem is that the pie pieces with extremely small values are very difficult to depict.

5.3.2 Bar chart

In most cases, it is better to use a bar chart with percentages. This allows the parts to be clearly compared and it is easier to see what the separate components are.

5.3.3 Stacked bar chart

Another frequently used presentation form is the stacked bar chart. This is a bar chart in which the bars are divided into parts. This shows both the totals and the separate components. A stacked bar chart also has a number of problems. For example, interpreting the separate values in a stacked bar chart is less direct: the values of the separate bars are not immediately clear. The same as for the pie chart, it is difficult to compare the different parts. Because the bars are not on the same line, the user must compare the amount of surface area for each one.

[6] This can be partially adressed by sorting the slices on their sizes.

If we are depicting time series where parts of a total should be shown, then a stacked diagram can be useful. The advantage of a stacked diagram in this case is that the variation of the total can also be shown, as can be seen in Figure 19.

*Figure 19. Use of sustainable energy, stacked bar chart with a time series*
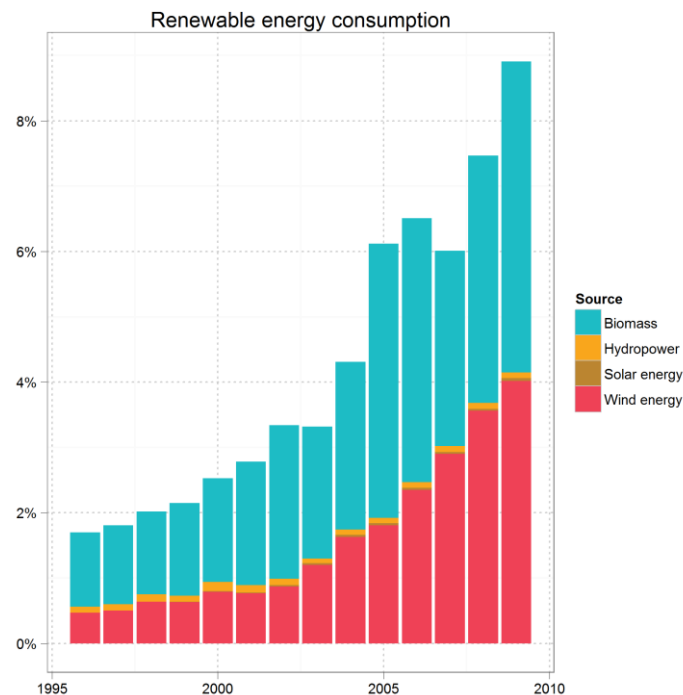
In dynamic graphics, it is increasingly common to be able to switch between a 100% scale (see Figure 20) and an absolute scale. This offers two perspectives: change of relationships between the parts over time vs. the absolute development over time.

Figure 20. Stacked 100% bar chart (Use of renewable energy)
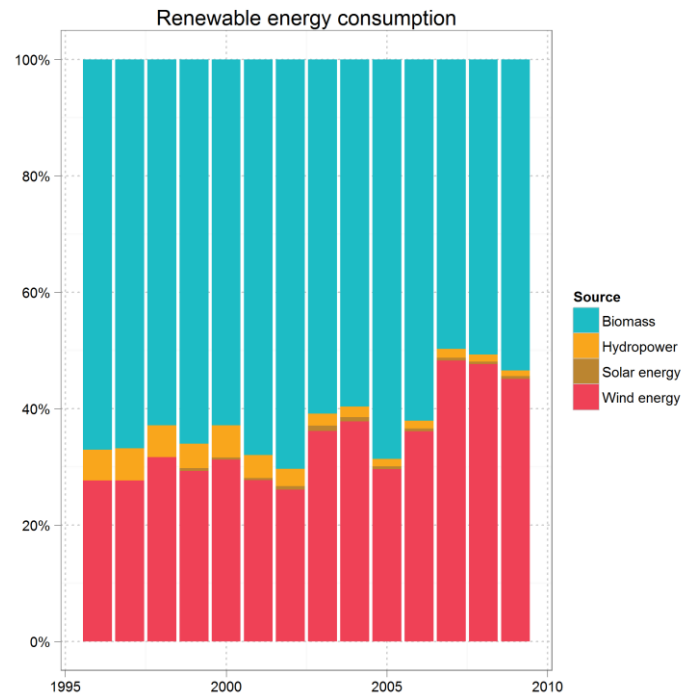
An alternative is a component bar chart. Here, the bars are not stacked, but are instead positioned next to one another (see Figure 21). This makes the relationships between the parts clear. Their separate development is somewhat visible, but it is less clear how the whole develops over time.

Figure 21. Use of sustainable energy, component bar chart
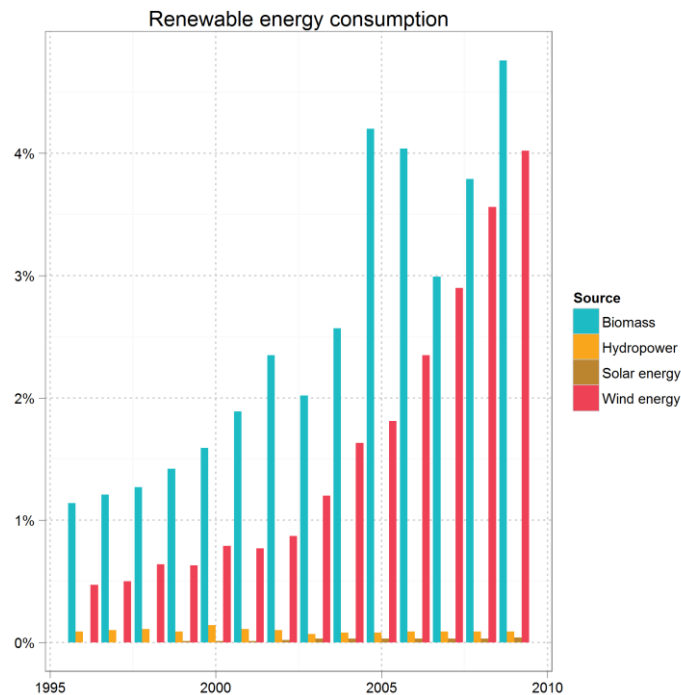
Part-whole only applies for quantitative variables in which the total is the sum of the parts. For quantitative variables that are aggregated in another way, such as averages or index numbers, another method is needed to compare the parts with the total.

A frequently occurring design for such data is a comparison in which the total is included as a separate bar. However, this design has a number of limitations. It is more difficult to directly compare each part with the total. The design is also somewhat "skewed": the parts and the total are not shown in relation to one another, but as equal components.

Figure 22. Economic growth in the third quarter of 2009 vs. the third quarter of 2008: part-whole diagram: EU is included as a separate, vertical line
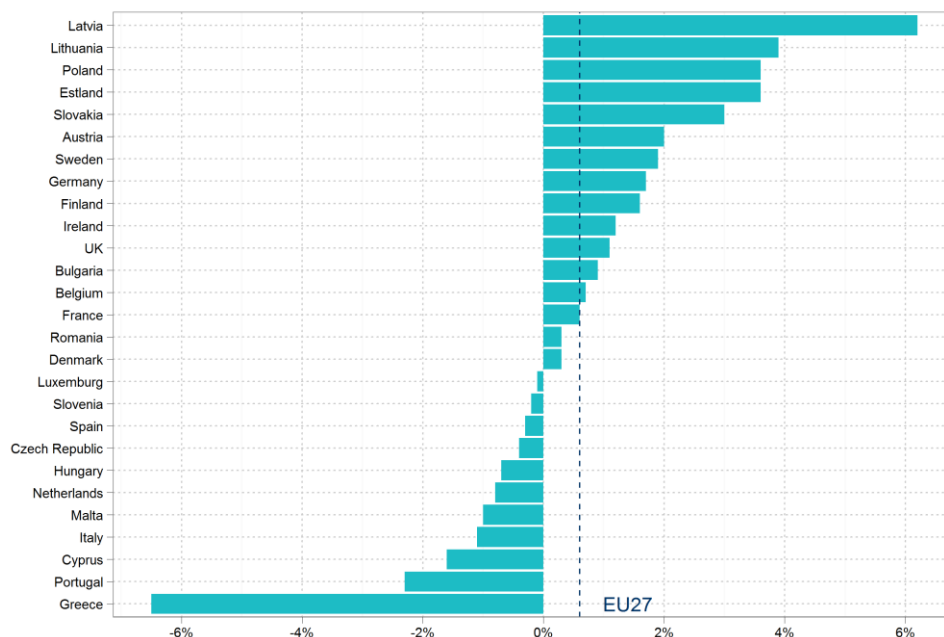

Figure 22 shows a better design: here, the average of the EU27 is added as a separate line. This clearly shows how the different countries in the EU have performed compared to the EU as a whole.

The most frequently used diagram type for part-whole is a vertical or horizontal bar chart.

5.4 Deviation

The goal of the deviation design is to depict how a set of values deviates. Deviation shows the extent to which the values of a variable differ from reference values. These reference values may be values from a previous reporting period. In that case, the goal is to show the change. A deviation diagram can also be used for a part-whole relationship or a time series in which a trend line is plotted.

If the goal is to show the deviation very clearly, then it is often useful to transform the original data: calculate and then plot the deviations (instead of the original

variables). Differences above and below the reference values are made clear through this process.

It can also be useful to sort the data and to split it into same-size components (quantiles) and then to give these groups a colour. This indicates what the order of the data is.

Figure 22 is an example of a deviation diagram in which the colour arrangement is used to make a distinction. The diagram also includes a reference line to clearly show the deviation compared to the average. However, no data translation has been performed, so the focus lies more on the performance of the countries compared to one another.

The diagram types used are bar and line charts.

5.5 Distribution

A histogram or distribution diagram shows the distribution or frequency over different classes. An example of a distribution diagram is shown in Figure 23.

Figure 23. Income distribution 2010

Income distribution of Dutch households 2010
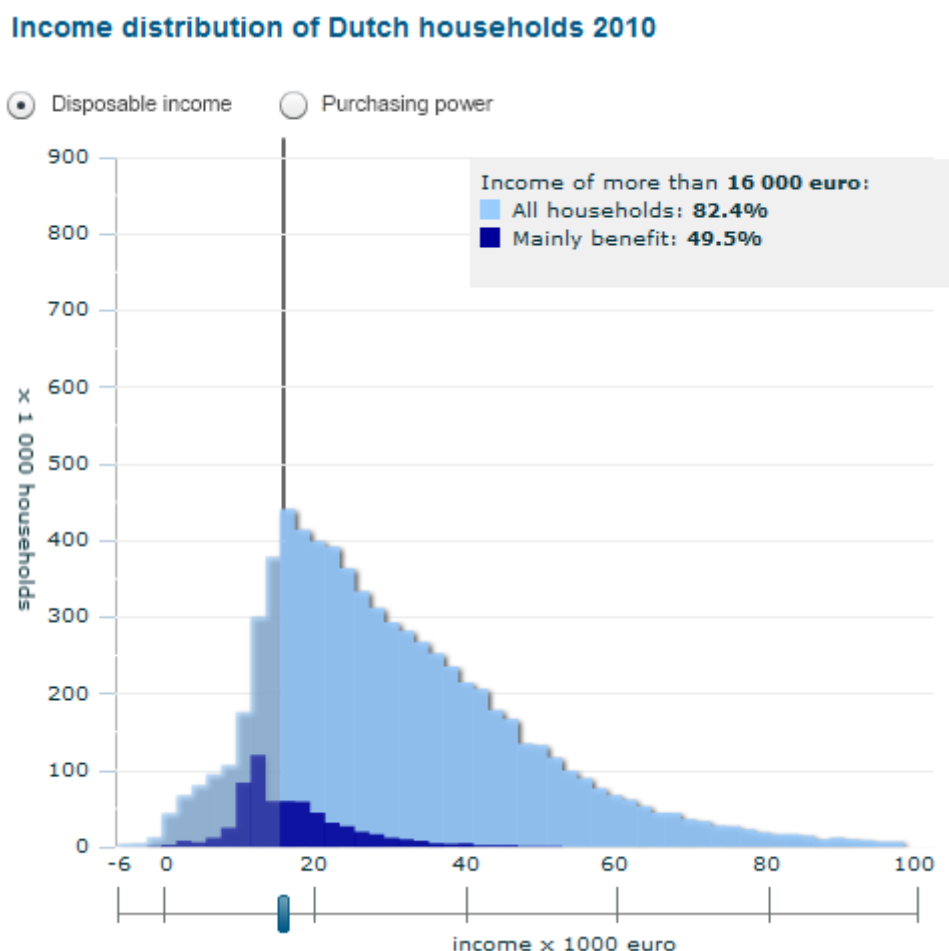
This diagram shows the classes (income) on the x-axis and the frequency on the y-axis. It therefore represents how income is distributed in the population and/or subpopulations. The vertical line shows what percentage of households has an

34

income of more than 16 000 euros. Another example of a distribution diagram is a population pyramid. This shows the age structure of the Dutch population, split by gender. (This diagram also contains an extra line that allows the user to look at which part of the population has an income that is higher than the current position.)

The diagram type used for this is the histogram.

5.6 Correlation

The goal of correlation is to show whether variables are related, if they are positively or negatively related, and the extent to which they are related.

A frequently used diagram is scatter plot, in which two variables are plotted against each other, and the points are then plotted.

Figure 24. Scatter plot, number of required procedures and days needed to establish a company

Figure 24 shows an example of a scatter plot in which the number of procedures and the number of days needed to establish a new company are plotted against each other. This diagram suggests that these two variables are correlated.

A scatter plot can often be improved by adding a trend or regression line.

5.7 New diagram types

This report lists only a selection of the most frequently used types of diagrams. New diagram types are, for example, the bubble chart and the treemap, as shown in the examples presented earlier in the historical overview.

In a bubble chart (see Figure 5), three numerical variables and a categorical variable can be plotted. Two of the numerical variables are marked on the axes and the third variable is represented by the surface area of the bubble. The colour of the circle is used to represent the categorical variable. As such, it is possible to combine many variables in a single diagram and also show whether they are related. The goal of the bubble chart is to show correlation.

A disadvantage of the bubble chart is that it often suffers from occlusion: data points that are next to each other can overlap, which makes the diagram harder to read.

Moreover, time series are less clear in this form than when they are presented as a line chart.

A treemap is often used to show a hierarchical structure in the data. Two numerical variables are plotted for each class in the data. One of the two variables is used for the surface area and the other for the colour of the area. In the previous example of inflation (see Figure 6), we can see the composition of inflation. Here, the extent of spending is represented by the surface area, and the level of the inflation is indicated using colour.

The goal of a treemap is to show part-whole relationships.

6. Thematic maps

6.1 Motivation

A large amount of statistical data has a regional component. Statistics Netherlands publishes many of its statistics split into a regional[7] classification. At Statistics Netherlands, the most frequently used and best-known of these is the NUTS-3 classification required by Europe: provinces (12), COROP (40) and municipalities (± 400). Furthermore, Statistics Netherlands is one of the few statistical bureaus that compiles and publishes detailed district and neighbourhood statistics.

Oftentimes, regional data are shown in diagrams. If the number of regions is limited (<10), then that is often the best method of displaying the data. However, a diagram does not take account of the spatial aspect of the data: how are the data distributed geographically? If there is a regional pattern and the data contains sufficient regions, then a thematic map is the preferred choice.

The goal of a thematic map is not to show the topography, but to show the characteristics of a theme or subject on the map. A thematic map can have topographical elements, such as rivers, mountains and cities, but these are provided for orientation or to provide clarity to the map's "reader".

A map visualisation method offers several advantages. It is, for example, recognisable to many users: they are familiar with the map of the Netherlands and can locate cities or particular areas. In addition, a map depicts a geographic distribution pattern that is otherwise totally lacking. An important example was discussed earlier in this document: the distribution of the victims of cholera in London (see Figure 3). This pattern clearly pointed towards the source of the contamination.

Maps can be used for the following purposes:

- To show the geographic location and the spatial distribution/density of the data (**distribution**). Where do the residents of the Netherlands live? This is the most important use of maps. A map is the only way to represent this in a recognisable and accessible manner.

- To compare different areas (**comparison**). Which areas are ageing, and which are relatively young?

- To **summarise** large amounts of data. What is the population density of all the districts (11,000) in the Netherlands?

Well-known standard texts with a lot of detail about the design of thematic maps are MacEachren (1995) and Kraak and Ormeling (2002). In this chapter, we discuss the

---

[7] Regions are geographical areas that have borders, names and other geographical characteristics. These characteristics can be used to match statistical information to a map.

most important types of thematic maps. However, we recommend consulting these texts for a more extensive and detailed description of how to create thematic maps.

6.2 Guidelines

A good thematic map shows detailed information for each region and depicts the geographic spread of the element that is being plotted. It is not always possible to satisfy both requirements.

In addition, a well-designed thematic map enables maps to be compared or combined.

A good map:

- is simple and easy to understand;

- transmits a clear and objective message;

- shows the data in an accurate manner and is not misleading;

- speaks for itself and does not require an explanation.

Do not make a map if:

- the data do not have a regional component (because, without the regional aspect, making a map makes no sense);

- the data relate to too few regions (<10);

- the map is not recognisable or is too complicated for the users;

- there is not enough room to create a map that can be read and understood.

6.3 Map types

There are different types of thematic maps, and the main ones are discussed in the following sections. The table below gives a brief summary of the options.

Goal	Absolute figures	Percentage/Density
Geographic distribution	Dot map, Proportional symbol map	Choropleth map
Regional comparison	Proportional symbol map	Choropleth map

6.3.1 Dot maps

In a dot map, the values of a variable are divided "into portions": each dot counts for a fixed amount. For each location or region, the number of dots is plotted that corresponds to the value of the variable. The previously discussed map by John

Snow about the cholera outbreak in London is an example of a dot map. A dot map mainly indicates the distribution of a variable, which can lead to problems if the regions are close together: dots may be plotted on top of each other. Moreover, a dot map makes it difficult to interpret values for regions because it is difficult for users to add up the dots.

*Figure 25. Dot map of the number of residents in the US (source Wikipedia)*
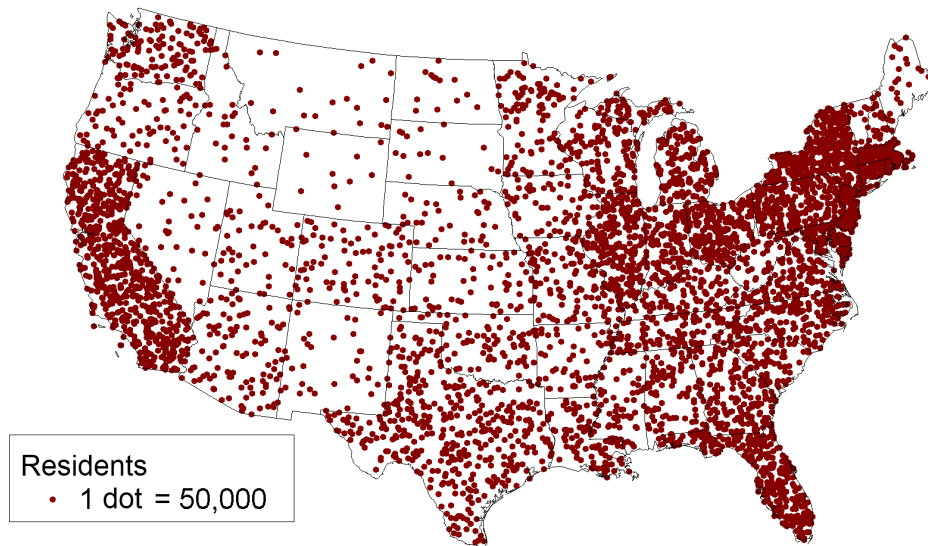
6.3.2 Proportional symbol maps

In a proportional symbol map, a symbol is plotted on the centre[8] of a region, and the surface area of the symbol is scaled with the value of the variable. An example of a geographic pattern is the number of votes cast by each municipality for the SGP political party (a Protestant party) in the 2003 elections for the Dutch House of Representatives. That pattern shows the "bible belt" of the Netherlands: a strip where a relatively large amount of Dutch Reformed Church members runs from Zeeland to the Veluwe.

---

[8] This is often – but not necessarily – the geographic centre, known as the centroid.

Figure 26. "Bible belt", votes cast for the SGP in 2003 (StatLine)
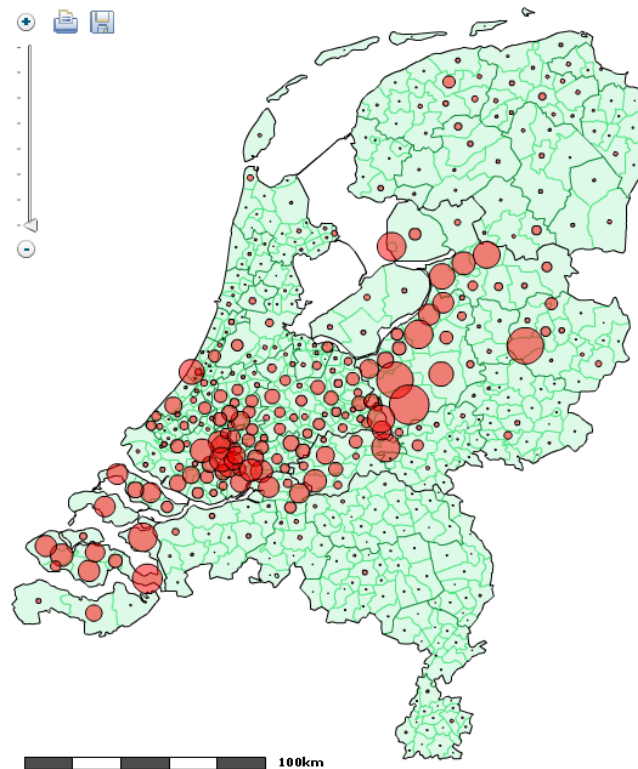
The map shows the value of the number of votes cast by linearly scaling the surface area of the circle. This map indicates per municipality how many votes were cast for the SGP, and also shows the geographic distribution. Note that this concerns the absolute number of votes, and not the percentages of the total number of votes cast. To represent the percentages, a choropleth map would be a better option (see below). It is common practice to use a circle as the symbol, but squares or triangles are also possible.

This map is a good example of the use of a proportional symbol map because, even though it is a good thematic map, it also shows the weakness of proportional symbol maps: from Rotterdam to Dordrecht, many circles overlap and the map becomes harder to read. A proportional symbol map is less suitable for data in which a variable is reasonably uniformly distributed for adjacent regions.

### 6.3.3 Colour maps / Choropleth maps

A frequently used type of map is a colour map or choropleth map. In this map type, a region or area is coloured in with a colour that corresponds to the value of a quantitative variable that is scaled with the surface area of the region. In this way, density is shown. It is also possible to show percentage figures using a choropleth map.

Choropleth maps are appealing maps because they clearly illustrate both the value of a region and the geographic distribution.

However, colour maps are frequently used incorrectly. Often, absolute numbers are used instead of densities. This results in a perceptual bias. Large regions receive more perceptual attention than then they actually have a right to from a data perspective. It is therefore not a good idea to show the number of residents per municipality using a choropleth map; the correct representation is the population density per municipality.

Figure 27. Average distance by road to the closest general practitioner office, 2007
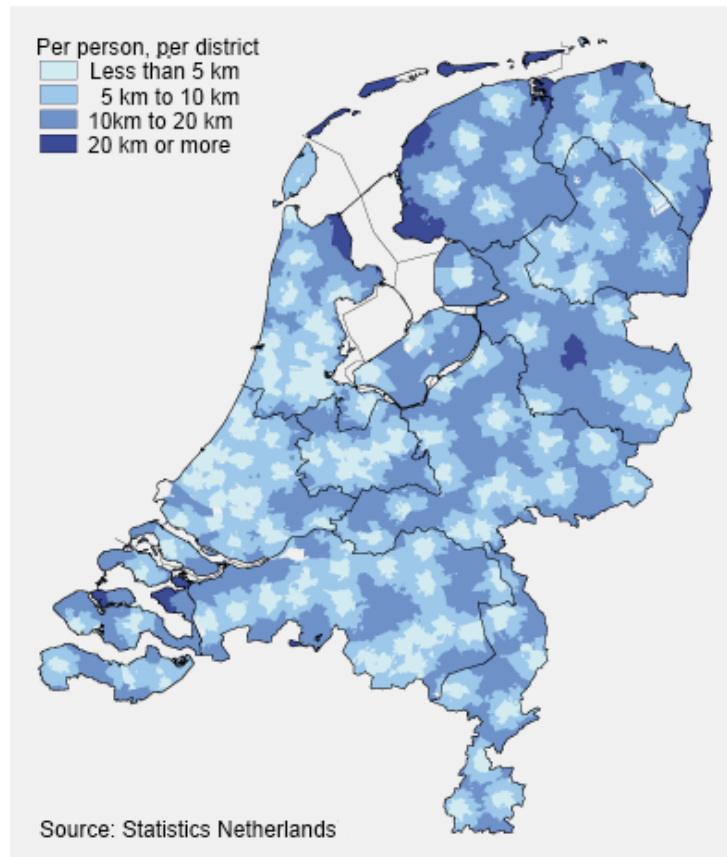
Depending on the dataset, the data can be shown by means of a breakdown by class. This could be a cluster[9], quantiles or a linear classification. Do not use more than five different classes. More classes are difficult to distinguish between in a map. The method of classification is very important for the message transmitted by the data. Geographic data are often asymmetrically distributed. Figure 28 shows the exact same data but with a different scale classification. Select a scale classification that corresponds to the distribution of the data. Create a histogram, for example, for this purpose.

---

[9] The standard practice for a random dataset is to use a k-means cluster algorithm to determine the clusters.

41

Figure 28. Linear and quantile, and cluster classification of population density per municipality (2009)

Depending on the message, different scales are possible in a choropleth map. An ordinal colour scale expresses a certain order: which class ranks higher and which class ranks lower. It must be possible to distinguish between the separate colours. In an ordinal colour scale, it is important that the colours run from light to dark. One advantage of this is that these colour scales can still be distinguished between if the map is printed in black and white

A divergent scale expresses how regions deviate from a norm. Examples of this are negative growth and positive growth or below average and above average. In a divergent scale, the data is split into two parts. For the two separate parts, a separate colour scale is used, in which the end colour (usually white) forms the link between the two colour scales.

For some data, it is a good idea to make maps in which the extent to which the area is coloured in/shaded corresponds to the value of the area. A continuous scale is used for this purpose. Research has shown that the most suitable colour scale is a scale in which a colour runs from white (transparent) to a saturated colour. Such a colour scale is experienced as a linear scale and therefore can also be used effectively to represent continuous variables.

6.4 Map components

The most important part of a map is the data. Therefore make sure that most of the visual attention is focused on the data.

Other components are also needed for the map to be properly understood. However, ensure that 80% to 85% of the surface area of a map visualisation is dedicated to the map/data.

The necessary components are:

- The title of the map clearly indicates what the map is about. Another possibility is to add a subtitle.

- The legend must indicate what the colours and symbols used in the map mean.

- The regions that are used must be indicated, for example, in the title or the legend. Is it a map about municipalities, COROPs or districts?

Other possible options:

- Text labels that help readers to understand the map or the data.

- A map scale that indicates the size and the distances.

6.4.1 Legend

A legend must clearly indicate how the map should be read and understood.

To make a useful legend, it is a good idea to observe the following rules:

- The data intervals of the classes used must be clear. Therefore, do not use 100-200, 200-300 and 300-400, but 100-199, 200-299 and 300-399 instead. Avoid intervals where the end points overlap.

- Avoid data intervals where there is a gap between the ends of the classes.

- Give a different colour to areas for which no data is available, and state this in the legend.

7. Conclusion and expansions

Diagrams and maps are a good way of presenting statistical data. This does not apply only for the general public, but also for other statistics colleagues. A well-designed diagram or map provides insight, while a poorly designed one can lead to the wrong conclusions.

Diagrams form an important tool for transferring information that is present in statistical data material. However, the use of diagrams is not without its dangers. We therefore recommend taking account of the guidelines set out in this document.

Developments in the field of visualisation are constantly advancing. New types of diagrams – albeit small in number – are still being created. An area that is currently undergoing serious development is the field of interactive visualisations. Animation is used increasingly often to present statistical data, although this medium is still in its early stages.

Possible expansions of this document are the deployment of visualisation techniques for analysis purposes, the use of animation, an elaboration of the use of colour scales and a representation of the reliability of figures (such as by using nomograms).

8. References

Balzer, M. and Deussen, O. (2005), Voronoi treemaps. In: *Proceedings IEEE Symposium on Information Visualisation*, IEEE Press.

Bethlehem, J. (2002), *Gebruik van grafieken*. Internal report, Statistics Netherlands, Voorburg.

Bosch, O. ten and Jonge, E. de (2008), Visualising Official Statistics. *Statistical Journal of the IAOS* 25, 103-116.

Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. (1983), *Graphical Methods for Data Analysis*. Wadsworth International Group, Belmont / Duxbury Press, Boston.

Cleveland, William S. (1993), *Visualizing Data*. Hobart Press, Summit, NJ.

Few, Stephen (2004), *Show Me The Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, Oakland.

Gapminder, www.gapminder.org.

Huff, Daryll (1954), *How to lie with statistics*. Norton, New York.

Kraak, Menno-Jan and Ormeling, Ferjan (2002), *Cartography: Visualization of Spatial Data*. Prentice Hall.

MacEachren, Alan M. (1995), *How maps work: representation, visualization, and design*. Guilford Press, New York.

New York Times (2008), *All of Inflation Little Parts*. www.nytimes.com.

Playfair, J. (1786), *The Commercial and Political Atlas*. London.

Schmid, C.F. (1983), *Statistical Graphics, Design Principles and Practices*. Wiley, New York.

Tufte, E. (1983), *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

Tukey, J.W. (1977), *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

UNECE (2009), *Making Data Meaningful, Part 2, A guide to presenting statistics*.

Wainer, H. (1997), *Visual revelations: graphical tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. Copernicus / Springer Verlag, New York.

Wallgren, A., Wallgrens, B., Persson, R., Jorner, U. and Haaland, J-A. (1996), *Graphing Statistics and Data, Creating Better Charts*. Sage Publications, London.

Wikipedia, nl.wikipedia.org/wiki/Thematische_kaart.

Version history

Version	Date	Description	Authors	Reviewers
Dutch version: Datavisualisatie				
1.0	13-04-2011	First Dutch version	Edwin de Jonge	Jelke Bethlehem Olav ten Bosch Peter Corbey
1.1	Forthcoming	Some new figures	Edwin de Jonge	
English version: Data visualisation				
1.1E	26-07-2012	First English version	Edwin de Jonge	