



Non-spatial Visualisation

8

Karel Macků

Abstract

An enormous amount of various data is produced every day. With proper data visualisation, an information hidden in the data can be easily and quickly revealed. It is necessary to create a communication channel that could quickly and efficiently transfer the information from the data to the user. By using visual elements like charts, graphs, and maps, data visualisation is an accessible way to see and understand trends, outliers, and patterns in data. This chapter offers an overview of relevant data visualisations divided into thematic categories and supported by examples.

Keywords

Visualisation · Data · Chart · Information

In the world today, we encounter enormous amounts of data every day. To convert data into useful information, data must be presented to the user in a way that allows interpreting, analysing and applying the gained information (Yau 2011). It is necessary to create a communication channel that could quickly and efficiently transfer the information from the data to the user – this can be done with data visualisation. Tableau Software, a company offering a software platform

for interactive data presentation, briefly and comprehensively talks about data visualisation: “Data visualisation refers to the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualisation is an accessible way to see and understand trends, outliers, and patterns in data” (Tableau Software 2018). According to (Friedman 2008, p. 1) the “main goal of data visualisation is to communicate information clearly and effectively through graphical means. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects more intuitively”.

Visualisation is an important step in the whole process of data analysis. Legendary statistician John Tukey often mentions visualisation in the context of using visualisation to find meaning in the data (Tukey 1977). Despite his statistical focus, he believed that a graphic presentation of information plays an immense role. A proper visualisation based on source data can help to understand the data, improve decision making and provide a more objective preview of the problem represented by data (Yau 2013). A graphic can also reveal hidden patterns and relationships.

Visualisation methods have gone far beyond traditional data presentation with simple charts and graphs. Modern trends approach data visualisation as both a science and an art. Of course, certain standards of correctness (e.g. by

K. Macků (✉)
Department of Geoinformatics, Palacký University
Olomouc, Olomouc, Czech Republic
e-mail: karel.macku@upol.cz

choosing a method according to the characteristics of the data) are still kept, but there is an effort to make the result interesting and catchy to attract the reader's attention. Sophisticated data visualisation and infographics methods offer a variety of exciting charts and diagrams. The advantage of technologies today is the possibility of presenting outputs in the form of online interactive web tools, which makes the processing of the information, that author attempts to communicate, even more intuitive and attractive.

In this chapter, a non-spatial data and its visualisation are discussed. Non-spatial data plays an undeniable role in the field of economics and business intelligence. For that reason, an overview of most common and powerful possibilities how to visualise it will be presented on the following pages.

8.1 Software

Nowadays, a variety of software is easily available, knowledge of some of them is a part of general digital literacy. Almost everyone, who somehow uses a computer, is able to create any visualisation using some of the available software. Most of the computer users are skilled with Microsoft Office Excel – software that doesn't need to be presented (or its opensource alternatives Libre Office/Open Office). Working in these tools is relatively convenient and straightforward, as underlying data and graphical tools are integrated into one user environment, and the whole process is very intuitive. However, this approach does not always offer proper or high-quality graphical outputs and supports the user's tendency to blindly insert data into the provided graphics templates without deeper thinking. By this approach, data loses its ability to interpret the story that is stored in it (Nussbaumer Knafflic 2015). Another point is the technical maturity of the output. In the world of modern technologies, where most of the information is distributed online, it is much more professional to produce outputs that offer a degree of interactivity and support simple distribution in the digital environment. Interactivity allows the viewer to engage

with your data in ways impossible by static graphs. With an interactive plot, the viewers can zoom into the areas they care about, highlight the data points that are relevant to them and hide the information that is not (Barter 2017). For this reason, some tools will be introduced offering the possibility of creating interesting graphical outputs.

8.1.1 Tableau Software

The company Tableau Software offers a set of tools of the same name designed for exploratory analysis and data visualisation. The product is especially focused on an effective and highly aesthetic level of visualisation, which undoubtedly attracts many customers. The full version of the program is paid, but the version *Public Tableau* is freely available. In this version, a user can work with many formats such as Microsoft Excel, Access, text files, JSON files, databases and also spatial data (several data formats are supported). After loading the data, the user can easily select the attributes they want to visualise and based on the data type, a set of options is automatically offered to create a visualisation. The main idea of the Public Tableau is the interactivity and presentation of the outputs in the online environment so the result can be shared with other users as attractive interactive data visualisation. The tools are, of course, multiplatform, they can be used as a desktop, mobile or online version.

8.1.2 HTML, Javascript and CSS

HTML, Javascript and CSS are the basis of every webpage. With modern technologies represented by HTML 5, advanced data visualisation running native in the internet browser can be done. This solution is probably suitable only for technically-advanced users/developers, who can handle coding with these technologies. There are several libraries designed for building interactive/static web visualisations, for example, Javascript libraries D3.js, Charts.js or FusionCharts. These libraries offer dozens of charts; detailed information can be found on their websites.

8.1.3 R

It is free and open-source statistical and mathematical computing software, primarily focused on data analysis and modelling. Since R has been developed mainly for statistical analysis, it has a solid background for different types of calculations suitable for data analysis. There is a lot of packages, which can extend the functionality of R software with just a simple code command. Thanks to the packages, R is a very mighty tool for data visualisation. Of course, a knowledge of code writing is required (as well as with HTML), which makes R for many people inapplicable. But once this obstacle is overcome, a new world of data handling and visualisation is opened. All graphics can be saved in vector formats, so it is possible to edit and refine the design of the outputs in suitable graphical software, like Adobe Illustrator or Inkscape. Except for traditional static graphic, also interactive outputs can be produced with special R packages. Sometimes, the interactivity is redeemed by complexity in the form of one extra line of code!

8.1.4 Datawrapper

Datawrapper is an online tool for making the interactive charts. It has a very simple interface; a user can upload data from a file or paste the value directly into the field. The tool generates graphics automatically; a user can choose one of

the 16 types of visualisation. Several refining steps can be done, like customising of axis, labelling or colour setting. This tool is an ideal solution when one needs a quick, simple interactive visualisation without any programming.

These examples were just a small slice of what nowadays technologies offers. Everyone is comfortable with a different level of challenge, content control and output options so everyone can find their optimal tool for creation of graphical outputs. There is an overview of another tools for visualisation in following table. Of course, this list is not complete, there are dozens of tools in offer (Table 8.1).

8.2 Charts Classification

There might be a confusion in terminology regarding the visualisation of non-spatial data. Usually, words ‘chart’ or ‘graph’ are generally used to describe any visual output. For many people, these two terms mean the same, but there is a difference. A chart is a superior term for a group of methods, how to present information. A graph is a particular graphical tool, which shows a mathematical relationship between sets of data (Blaettler 2018). With this approach, a graph is a subcategory of a chart. For this reason, the term chart will be rather used in this chapter, to keep the description of different methods more board.

Table 8.1 An example of visualisation tools

Name	Output	Difficulty	Pricing
Plotly	Interactive/static	Coding required	Free with some paid plans
Highcharts Cloud	Interactive	Easy to handle	Free
D3.js	Interactive	Coding required	Free
Charts.js	Interactive	Coding required	Free
Inforgram	Interactive/static	Easy to handle	Free, extra features paid
RAWGraphs	Interactive/static	Easy to handle	Free
DataHero	Interactive/static	Easy to handle	Paid
Visually	Interactive/static	Easy to handle	Paid
Visme	Interactive/static	Easy to handle	Free
Google charts	Interactive	Easy to handle	Free
vVisualize.me	Static	Easy to handle	Free

Source: Author

Different types of charts will be described in the following chapter. Since there are dozens of possibilities of visualisations, only the most interesting or most commonly used variants will be introduced. For better thematic logic, the individual methods were divided into thematic groups. The inspiration for this system was the book *Visualize This* (Yau 2011) and the website www.datavizproject.com (Ferdio ApS 2017).

8.2.1 Trend Over the Time

Time series are typical data for many phenomena. Things are changing in time, and this change can be easily captured and presented by suitable graphics. Talking about time series, users try to explore the trend in data. Is the value of the phenomena increasing or decreasing? Are there any repetitive cycles?

Temporal data can be divided into discrete and continuous types. The knowledge about this character of data should guide the user in a decision, which kind of graph should be used. For example, a monthly revenue report is an information referenced to a one-time step – a month, so this can be considered as a discrete phenomenon. Then, a simple bar or point graph can be used. The second type is the continuous data. This is kind of information which can be measured at any time of day during any day of the year. A typical example could be a temperature or another meteorological phenomenon; regarding the economic data, we can use stock exchange prices as an example. The structure of data is same for discrete and continuous phenomena, to distinguish the difference, the proper way of visualisation should be used. The most primitive solution is to connect discretely plotted data with any line.

8.2.1.1 Bar Chart

Bar charts are commonly used, which means the user doesn't need to 'learn', how to read the graph. The graphic element is a rectangular bar whose length represents the value. The time axis captures time points, which have to be ordered chronologically. Then every bar stands for one discrete time point. Finally, there are many

additional ways how to tune the bar graph, e.g. bars can be placed horizontally or vertically or some of the bars can be highlighted by a different colour (e.g. time points when the value is higher than set limit etc.) (Fig. 8.1).

8.2.1.2 Point Chart

Point chart works on same principle as the bar chart does, except for used geometrical element – it is a simple point here. This can sometimes be more suitable since the points do not represent such graphic content and load as bars. Point chart is also known as a scatterplot when non-temporal data is used. It is crucial to properly create an axis representing the value of the phenomenon, as there is no other way to find out the value.

8.2.1.3 Line Chart

The line chart is a type of chart used for continuous data. The basis of the chart is the same as the basis of point chart. The continuity is added by connection of this points with line segments. Then the chart shows how data changes in the time (particular value is stored in the point), and the line segments create a feeling of continuity. It also better points to the trend between time markers (Fig. 8.2).

There is only a minor difference between the line chart and the spline chart. They differ only with the way how the points are linked. While the line graph uses straight line segments, the spline chart plots a fitted curve through each point from

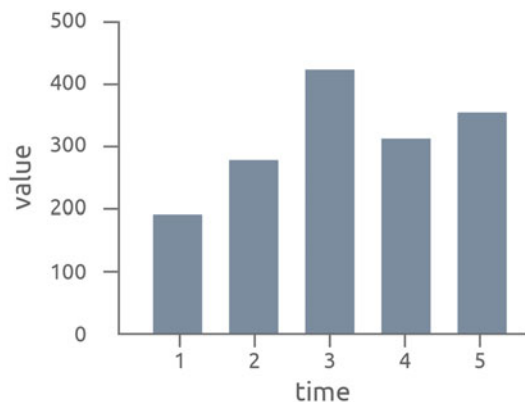


Fig. 8.1 Bar chart. (Source: Author)

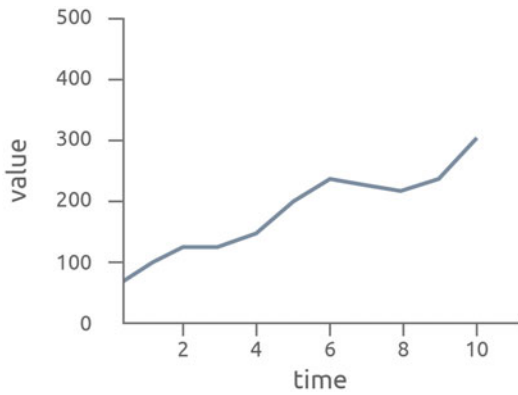


Fig. 8.2 Line chart. (Source: Author)

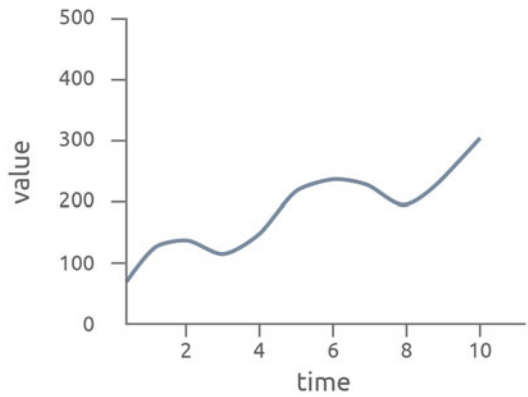


Fig. 8.3 Spline chart. (Source: Author)

the time series. This provides a more smooth and natural course (Fig. 8.3).

An attractive solution for a description of changes between two or several time point is a slope chart. It combines time-approach with multiple observed variables/categories. This helps to see differences in the development of specified categories and also the rate of change in one particular category compared to others (represented by the slope of the connecting line). At the same time, deviations in the general trend can be perfectly observed (Fig. 8.4).

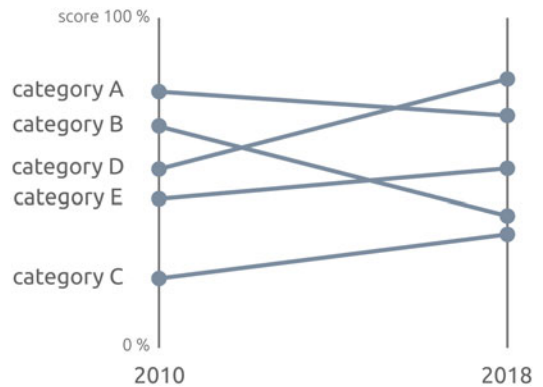


Fig. 8.4 Slope chart. (Source: Author)

8.2.1.4 Step Chart

Last modification of the line chart is a step chart. This one is formed by stepped lines between the time points. It is appropriate to use it in the situation, when the data represents a sudden change in irregular time intervals, for example, a price of any commodity which has been the same for a long time, then in one day the price increased (Fig. 8.5).

8.2.1.5 Gantt Chart

This chart visualises via bars duration of several categories in a time series. It illustrates the start and end point of occurrence of any activity/phenomenon. This chart is typically used as a project management tool for a graphical representation of the sequences of activities over time. Tasks or activities, which are parts of the whole project, are displayed in the time context (Fig. 8.6).

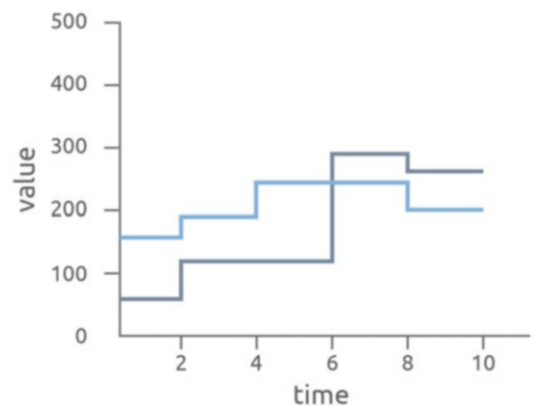


Fig. 8.5 Step chart. (Source: Author)

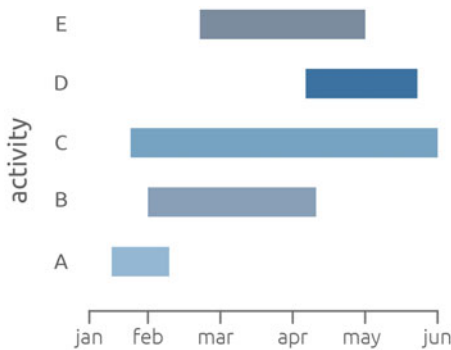


Fig. 8.6 Gantt chart. (Source: Author)

8.2.2 Proportions

Proportion data is grouped by categories/types. Each category represents a possibility, which is part of the certain unit. This distribution of proportions is the most important information for comparing groups between themselves. With proportional visualisation, questions like “Are all of the categories equally represented? Is there any category which dominates?” can be answered.

For this type visualisation, a data needs to have a form of proportions that add up to 1 or 100%. Every part could be stored relatively (as a proportion) and absolutely – total values allow to compare not only proportional part but also total size/amount in different categories.

8.2.2.1 Pie Chart

A pie chart is one of the most often used charts and is typical for an explanation of proportions. The circle which is representing the whole is divided into sectors. The arc length of each segment (or interior central angle, or area) is illustrating the proportion of individual categories. All categories together must form a unit/100%.

8.2.2.2 Doughnut Chart (Fig. 8.7)

The doughnut chart is just a modification of a pie chart, only the blank centre is added. That allows presenting of multiple information at the same time since the inner blank space could be filled with additional related data.

According to some of the resources (Nussbaumer Knaflic 2015), the pie or doughnut chart is an inappropriate way how to visualise

proportional data. This is caused by the greater difficulty of perceiving angles or area than distances (which are the key information regarding, e.g. bar charts), it is a common property of human eye perception. In a situation when two or more categories are represented by an approximately same value, it’s difficult to decide which one is greater. This issue can be solved by adding labels. Still, several authors recommend using different proportional methods, like a stacked/simple bar charts.

8.2.2.3 Stacked Bar Chart

Instead of pie/doughnut charts, simple bar chart ordered from highest value to least can be used. All bars have the same baseline; the endpoint is easier to compare. Even small differences can be distinguished. The length of bars is recalculated in that way that their sum equals to whole/100%.

A stacked bar chart is a perfect solution for visualising proportion and comparing several classes at the same time. Because of their geometrical representation, they are even more space-saving than pie charts. Stacked bar chart contains multiple values on top of each other, which shows the division of the whole into categories. Concurrently, individual bars represent the different level of categories or even time points. For example – the stacked bar chart can represent sale strategies: every bar signifies a particular strategy (A-E in the Fig. 8.8), different colour shades represent a type of product, and on the y-axis, total sales are displayed.

8.2.2.4 Tree Map/Area Chart

This type of charts uses a structure of rectangles and their area to express the proportion of the whole part. Size of every rectangle represent the metrics. The outer rectangle represents parent categories, and rectangles within the parent are subcategories (Yau 2011). Therefore, primary requirement is that data has to have a tree-based structure (Fig. 8.9).

There is a similar alternative, which doesn’t require tree structure in the input dataset. Simple square area chart, also called a waffle chart, uses a regular grid of small cells. If the value of the cell is set, then the proportion is expressed by a number of cells (Fig. 8.10).

Fig. 8.7 Pie and doughnut chart. (Source: Author)

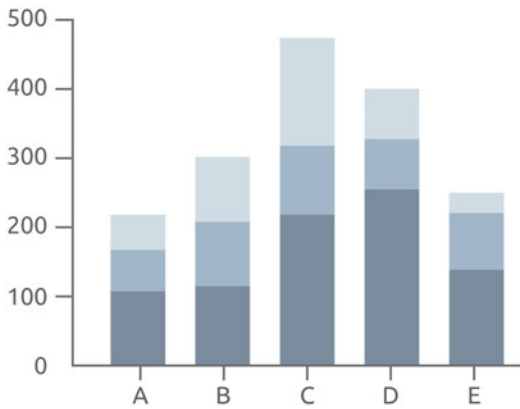
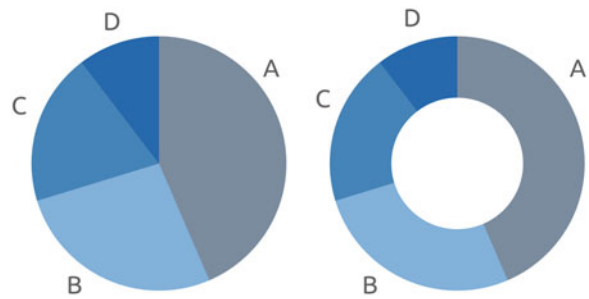


Fig. 8.8 Stacked bar chart. (Source: Author)

Regarding the tree map or area charts, there is the same issue with the perception of two-dimensional object as was discussed in pie chart paragraph. In this case, if the area map has a cell-based regular structure, the perception of information can be done correctly by simple counting of cells. Nussbaumer Knaflic (2015, p. 59) describes another situation when area charts are quite helpful: “when visualisation of numbers of vastly different magnitude is needed. The second dimension you get using a square for this (which has both height and width, compared to a bar that has only height or width) allows this to be done in a more compact way than possible with a single dimension”.

8.2.3 Relations and Correlation

There are many ways how to quantify relations between several variables/group. A statistical approach provides mathematical tools, such as

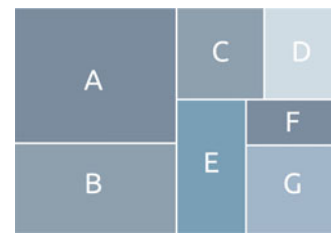


Fig. 8.9 Tree map. (Source: Author)

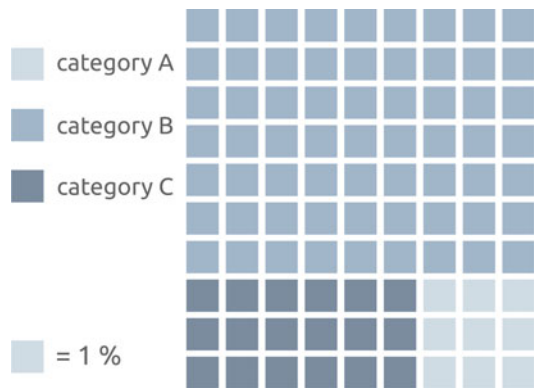


Fig. 8.10 Area chart. (Source: Author)

correlation or regression (if a conditions regarding the characteristics of variables are fulfilled). Sometimes it is much easier just to plot the data to reveal the hidden relations. A correlation simply describes, how two variables change together. Sometimes it is forgotten that correlation doesn't equal causation. Basic correlation of two variables expressed with chart can quickly describe the behaviour of the data, a rate of relation can be estimated, maybe a clustering tendency can be discovered.

8.2.3.1 Scatterplot

A scatterplot is one of the fundamental charts used for plotting of relations and dependencies. The data is displayed as a set of points placed in a Cartesian coordinate system. Therefore, the chart is limited for displaying of relations between only two variables. The placement of points in the chart helps to easily estimate the correlation between variables – if they are positively correlated, points are formed in the line-shaped group, rising with the value of the represented phenomenon. If the correlation is negative, this line group has a decreasing trend. With no significant correlation, points are not grouped in the line-shaped and spread in the field randomly. Another categorical information can be added into the chart in the form of point colour or different shapes. Then, it is possible to observe differences between particular types; they might tend to create a cluster which indicates their similarity, or controversy, points might be overlapping, then there is no clear pattern in the categorical groups (Fig. 8.11).

8.2.3.2 Bubble Plot

It is possible to add a third variable into the scatterplot and compare more information at the same time. The size of the bubbles expresses the third variable – the measure here is an area, not radius nor diameter. The only area can accurately represent differences related to original number: if the displayed value is doubled than another, the area of a graphical element must also be double size. If another measure is used (e.g. diameter), the ratio between value and the area of a graphic element wouldn't be the same. Of course, the chart can be modified regarding shapes, squares or triangles could also be used. Attention must be paid to the position of a particular graphics element – the bigger element must not overlap the smaller one, so it would be not visible. This rule should be implemented in software (Fig. 8.12).

8.2.3.3 Scatterplot Matrix

In case that exploration of more than three variables is required, a scatterplot matrix is a solution. In this case, every possible combination

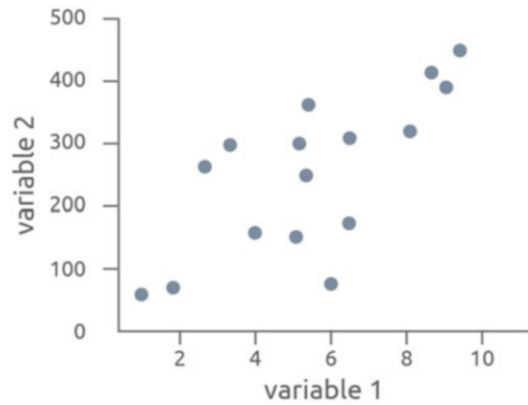


Fig. 8.11 Scatterplot. (Source: Author)

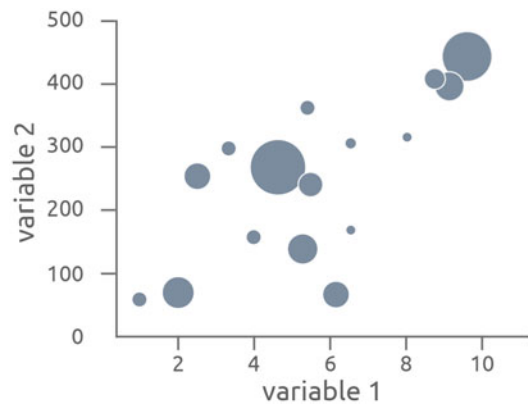


Fig. 8.12 Bubble plot. (Source: Author)

of pairs is plotted by a single scatterplot; subsequently they are all organised into a matrix. This visualisation could be a first step in an exploratory data analysis when the analyst has no clue what is the data about and what is its behaviour. With an increasing number of variables, interpretation of that kind of graphic presentation is more complicated, and the information is still kept on the elementary level of variable pairs (Fig. 8.13).

8.2.4 Differences and Comparison

Comparing a single variable is not a demanding task, the value of every record is displayed by one of the previous-mentioned methods and analysed. Bar charts or simple point charts may well serve

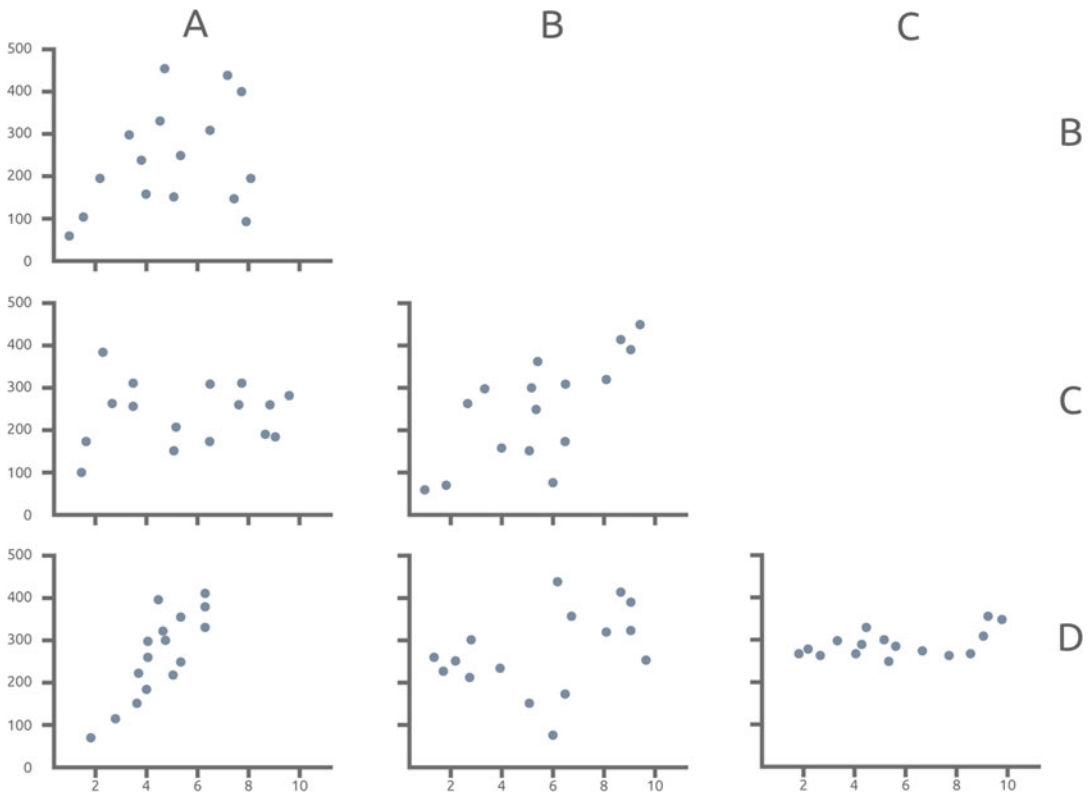


Fig. 8.13 Scatterplot matrix. (Source: Author)

to this task. Considering two or three variables, several charts for this type of visualisation have been introduced in the previous sections. Regarding the data with more variables, known as multi-dimensional data, different graphical methods have to be used.

8.2.4.1 Heatmap

A heatmap is a simple way how to look at all data at once. Information is displayed in a matrix of regular graphical elements, mostly rectangles or squares. The value for every record’s attribute is indicated by colour intensity. The size of the heatmap is defined by the number of rows times the number of attributes, heatmap has the same number of elements as the input table does. This type of visualisation is not sophisticated for the reading of accurate values in a particular record but provides a great overall view on the complete dataset. Some characteristic patterns in data can

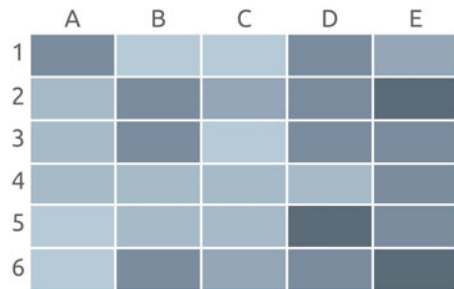


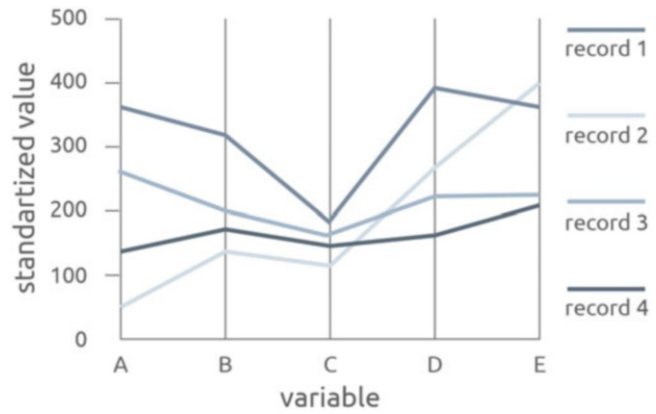
Fig. 8.14 Heatmap. (Source: Author)

be revealed, e.g. tendency to clustering (Fig. 8.14).

8.2.4.2 Paralel Coordinates

Parallel coordinates is another common chart for visualisation of multidimensional data. The number of attributes defines a number of used vertical axes – every single of them represents one variable. It follows that different axes have a different

Fig. 8.15 Parallel coordinates. (Source: Author)



scale. To avoid labelling and adding more graphical ballast into the chart, data can be scaled; then all axes have the same scale. Parallel coordinates are suitable for revealing similarities between records. For that reason, labels are not always necessary, it is enough if the plot can define groups of records with a similar pattern or trend on the observed variables (Fig. 8.15).

8.2.5 Statistical Charts

8.2.5.1 Histogram

A histogram is a specific statistical chart which describes the frequency of occurrence of values. The geometric element is a bar again, the height of the bar represent a frequency, i.e. the number of occurrences in the category, which belongs to the bar. The category here doesn't stand for a different type, it rather defines the range of values, in which are data are binded. It follows that both the horizontal and vertical axes are continuous (Fig. 8.16).

8.2.5.2 Distribution Plot

Although the horizontal axis of the histogram is continuous, the distribution is still divided into intervals. If the interval size is not set properly, a lot of information about inter-interval variation is lost. On the other hand, plotting of every single record would make the chart messy and confusing. A compromise between this approaches is a distribution plot, which is able to capture the

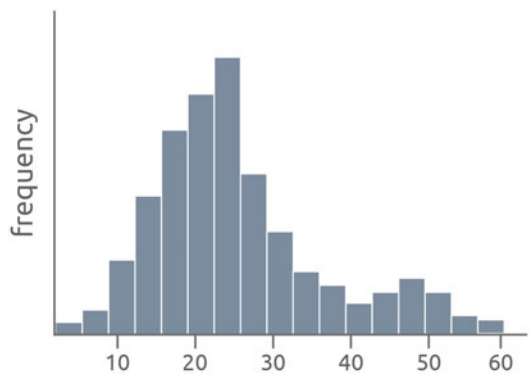


Fig. 8.16 Histogram. (Source: Author)

smaller variation within the distribution and also smoothen the detailed original data. The vertical axis represents the probability of occurrence of value from the sample population. The area under the curve has to be equal to one (or 100%) (Fig. 8.17).

8.2.5.3 Boxplot

A boxplot is an important graphical tool for descriptive statistics. In one picture, it can describe several numeric information – median, first and third quartiles, and minimum and maximum (sometimes minimum and maximum is replaced by the value calculated: $\text{mean} + -1.5 * \text{interquartile range of lower/upper quartile}$). Outliers (values out of this range) are plotted as points. The spacings between parts of boxplots describe how the raw data is dispersed or

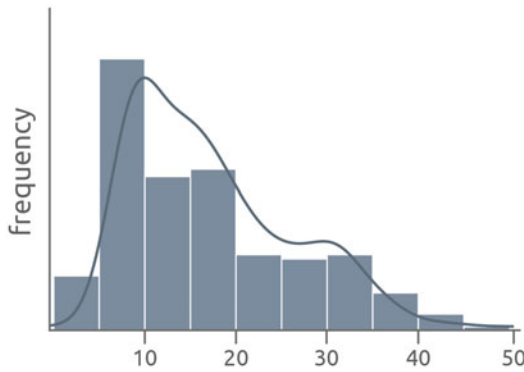


Fig. 8.17 Distribution plot. (Source: Author)

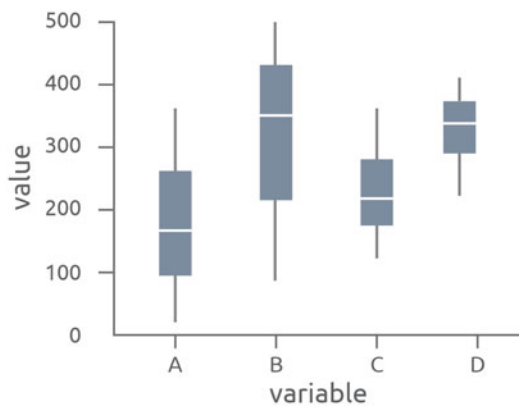


Fig. 8.18 Boxplot. (Source: Author)

condensed. With boxplots, several groups of data can be quickly compared (Fig. 8.18).

8.3 A Good Design

In the beginning, the data analyst has to know the data in detail. Once the analyst understands what kind of information is hidden in the data, what is the data type and character, he can decide which type of chart is the best solution for proper visualisation. Then another step of chart designing follows. The raw default output from the software is not wrong, but usually, it is also not the most attractive result. With an additional improvement of the graphics, information which the author tries to deliver with the chart might be easier to perceive.

It must be always considered, who is the audience, the reader of the created chart, for which purpose is the chart created. By design, the author can manipulate with the way how the chart is read. If it wants to focus on a significant trend, the axis and labels can be de-emphasised with grey colour, and the primary trend line is highlighted. Then, the trend is the information which draws the attention at first. On the other hand, a chart designed with the purpose of reading exact values must have readable and accurate labels of all axes.

Generally, some recommendation can be made. Mostly modest colours should be used. Some of the colour schemes can evoke emotions or feel (e.g. red colours indicate activity that should be addressed; neutral pastel colours means that all features in the chart are equal etc.). Proper labelling should be done – through user might know the context in which is the chart placed, he doesn't know the meaning of every single element of the chart. Therefore, a title of the chart, axes names and value labels and legend with explanations of colours should be a part of the visualisation. Geometrical aspects are also important. Sometimes it's more suitable just to rotate the chart, what makes it much easier to read (e.g. bar chart with long category names – rotation to horizontal is more natural, because it follows the way how we read the common text). A different spatial arrangement of geometrical features can solve the issues with blank space or can fit better into a whole graphic design (text or poster). Transforming the geometrical elements into pseudo-3D and displaying data that way should be avoided (the only exception is plotting a three-dimensional data with a 3D plot). Unfortunately, for example, visualisation of the pie chart in 3D is quite popular. As discussed above, the pie chart is not always a good choice for visualisation of proportional data; the combination with 3D makes it much more difficult to read or compare with others pie charts because the third dimension is problematic for perception. A perspective in 3D visualisation can also be misused for promotion – a segment of pie chart placed in the foreground looks larger than a segment of similar size in the background.

References

- Barter, R. (2017). *Interactive visualization in R*. <http://www.rebeccabarter.com/blog/2017-04-20-interactive/>. Accessed 13 Nov 2018.
- Blaettler K (2018). *The difference between charts & graphs*. *Sciencing*. <https://sciencing.com/difference-between-charts-graphs-7385398.html>. Accessed 2 Jan 2019.
- Ferdio ApS. (2017). *Data viz project*. <https://datavizproject.com/>
- Friedman, V. (2008). Data visualization and infographics. *Smashing Mag*.
- Nussbaumer Knaflic, C. (2015). *Storytelling with data*. Hoboken: Wiley.
- Tableau Software. (2018). *What is data visualization? A definition, examples, and resources*. <https://www.tableau.com/learn/articles/data-visualization>. Accessed 30 Oct 2018.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley Publishing.
- Yau, N. (2011). *Visualize this: The flowingdata guide to design, visualization, and statistics*. Indianapolis: Wiley.
- Yau, N. (2013). *Data points – Visualization that means something*. Indianapolis: Wiley.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

