

Chapter 5

Data Visualization



5.1 Introduction

Visualization via graphics like charts, graphs, and images is an effective and efficient way to interpret and understand data and help spot valuable information such as patterns, trends, and anomalies [1]. The reason is that, unlike tables and written text, graphs are primarily visual in nature, and approximately 70% of our sense receptors are dedicated to vision [2]. Moreover, our eyes are drawn to patterns and colors, can easily differentiate red from blue and a circle from a square, and can quickly see trends and outliers [3].

While the invention of data visualization may not be easily attributed to one individual, William Playfair (1759–1823) is generally viewed as the inventor of many common graphical forms, such as bar and pie charts [4]. One of his well-known visualizations is his balance of trade and chart of the national debt of England, one of the earliest line charts used to represent time series (Fig. 5.1).

Playfair was also one of the first people to use charts not only to educate but also to persuade and convince, for example, by comparing the “weekly wages of a good mechanic” and the “price of a quarter of wheat” from 1565 to 1821 (Fig. 5.2) [7].

Another classic example of an old visualization is the illustration of Napoleon’s failed Russian campaign of 1812 by Charles Minard (Fig. 5.3). The graph displays the number of French soldiers marching toward and then retreating from Moscow, overlaid on top of a map. The thickness of the band is representative of the number of soldiers, which decreased as the army moved from France on the right to Russia on the left. Underneath the map is a line chart displaying the temperature that soldiers faced as they moved during the campaign.

Florence Nightingale visualized in 1858 the factors affecting the lives and death rates of the British army in a graphic known as “Nightingale’s Rose” or “Nightingale’s Coxcomb” (Fig. 5.4). She showed Queen Victoria in her visual graphic that it was infections (in blue) killing the highest number of soldiers and not wounds [7].

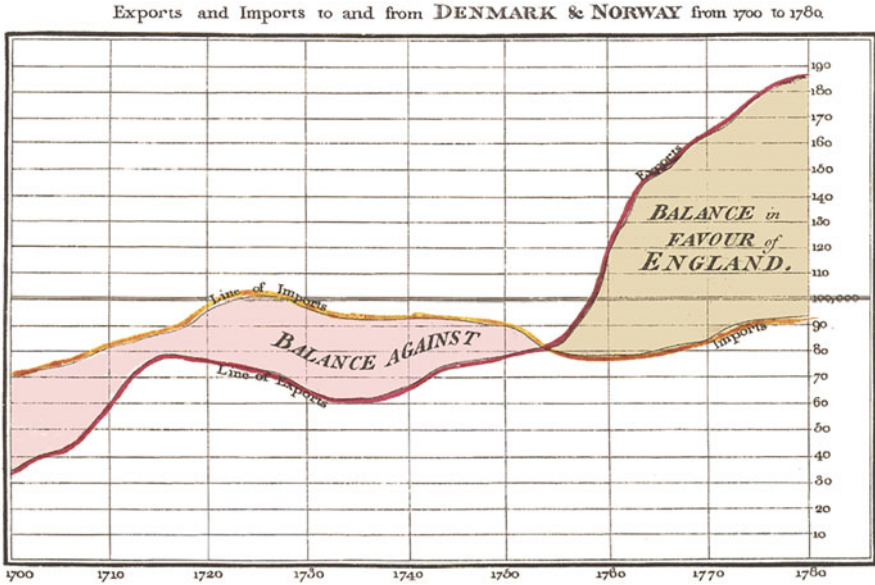


Fig. 5.1 William Playfair’s balance of trade and chart of national debt of England (Source: Wikimedia [5])

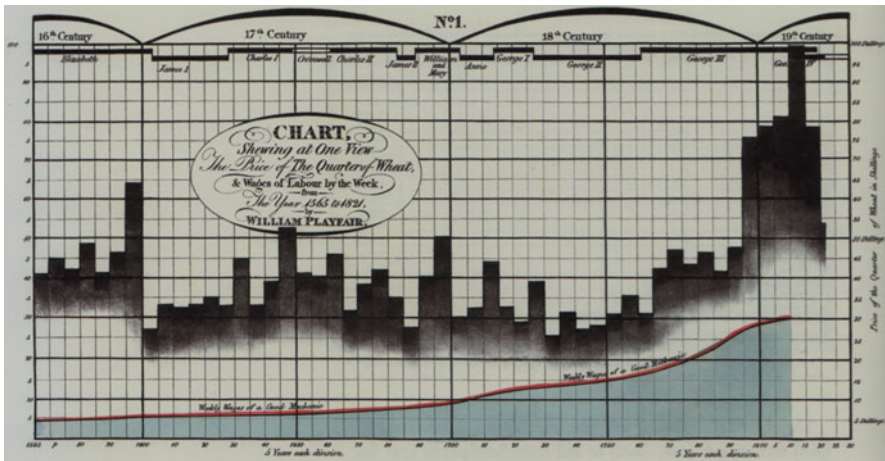


Fig. 5.2 William Playfair’s 1821 chart (Source: Wikimedia [6])

In the remainder of this chapter, we introduce the basics of data visualizations, including a taxonomy of basic graphical objects and charts and their uses. We include several visualizations from different fields generated with different software

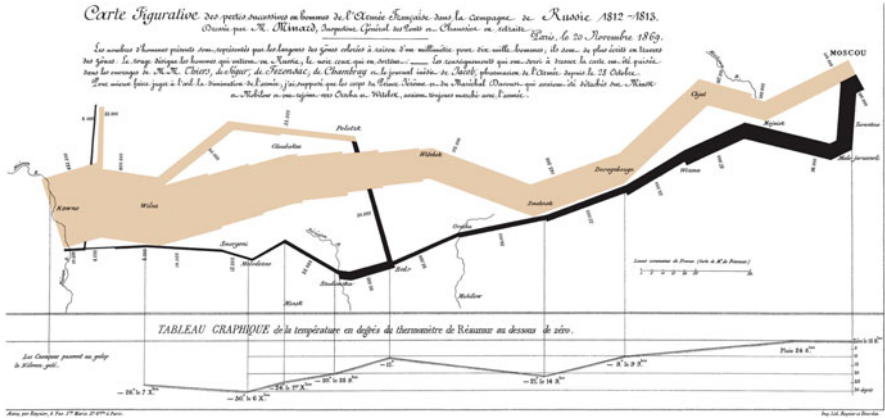


Fig. 5.3 Napoleon’s failed Russian campaign of 1812 by Charles Minard (Source: Wikimedia [8])

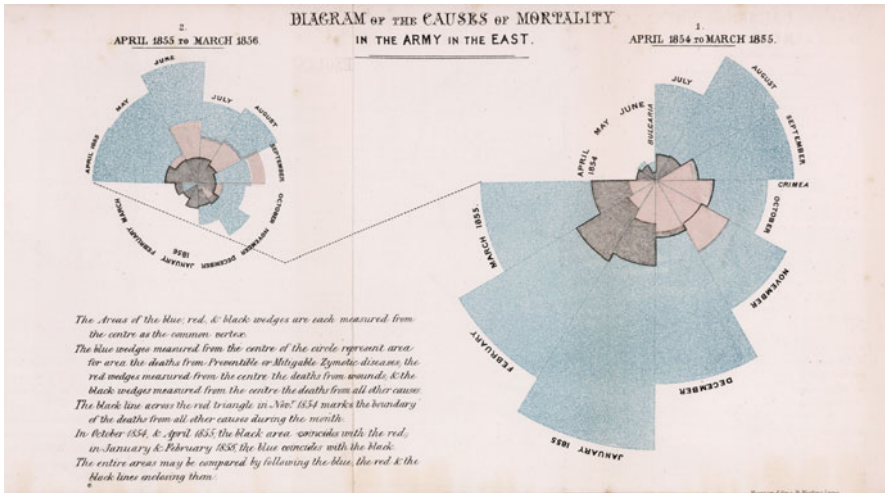


Fig. 5.4 Florence Nightingale’s 1858 diagram of the causes of mortality in the army in the East (Source: Wikimedia [9])

packages using different sources of open data. We also cover infographics and dashboards, which are visualization-rich tools that are increasingly being used in many industries. We finish with guidelines for building good visualizations.

5.2 Presentation and Visualization of Information

The type of data sometimes dictates the type of graph that can or cannot be used. As a quick reminder, data are broken into two types: quantitative and categorical. Quantitative values measure things and consist of a quantity and unit of measure (e.g., 300 km). Categorical data divide information into useful groups and are nominal (e.g., fall, winter, spring, summer), ordinal (e.g., low, medium, high), interval (e.g., 0–9, 10–19, . . .), and hierarchical (e.g., year, quarter, month, week, day). In many cases, the type of data dictates the type of graph and visualization to be used.

5.2.1 A Taxonomy of Graphs

Of the different available taxonomies of graphs, we will follow the one proposed by Stephen Few [2], a well-known expert in data visualization. According to Few, quantitative data can be basically represented in graphs by the following six basic objects: points, lines, bars, boxes, shapes with varying 2D areas, and shapes with varying color intensity.

A *point* is a simple dot on a graph representing two values, one on each axis, and a graph consisting of such points is referred to as a scatterplot (Figs. 5.5 and 5.6). Scatterplots, which are representations of many distinct data points on a single chart, give a general idea about the distribution of the data and are useful in highlighting relationships between different variables, showing if the two variables tend to vary independently or not. Scatterplots are also useful in showing correlations and in detecting data outliers [1, 2, 11].

A *line* connects a series of values or distinct points in a graph and is a good representation of how values change or evolve over time (called a time series) [1]. Line charts are used to view trends and cycles in data, usually over time or other ordinal data (Figs. 5.7 and 5.8) [11].

A *bar*, one of the most common types of data visualizations, is a rectangle that encodes quantitative information by its length (Fig. 5.9). Bars are easy to see and compare and should always begin at the value of 0 [2]. Bar charts are used to quickly compare data across categories, show trends and outliers, and highlight differences at a glance. Bar charts are especially effective when data can be split into multiple categories [11]. Graphs composed of vertical bars are referred to as column charts.

A *box* is also rectangular but encodes a wide range of values, such as the minimum, maximum, and median values (Fig. 5.10). Graphs of such boxes are referred to as box-and-whisker plots or boxplots and are used to show distributions of data. Typically, the box contains the median of the data, the first quartile (25% less than the median), the third quartile (25% greater than the median), and the whiskers represent data within 1.5 times the interquartile range (i.e., the range between the first and third quartiles). The whiskers can also be used to show the maximum and minimum points within the data [11]. Boxplots are often used to compare the distribution of different datasets [2].

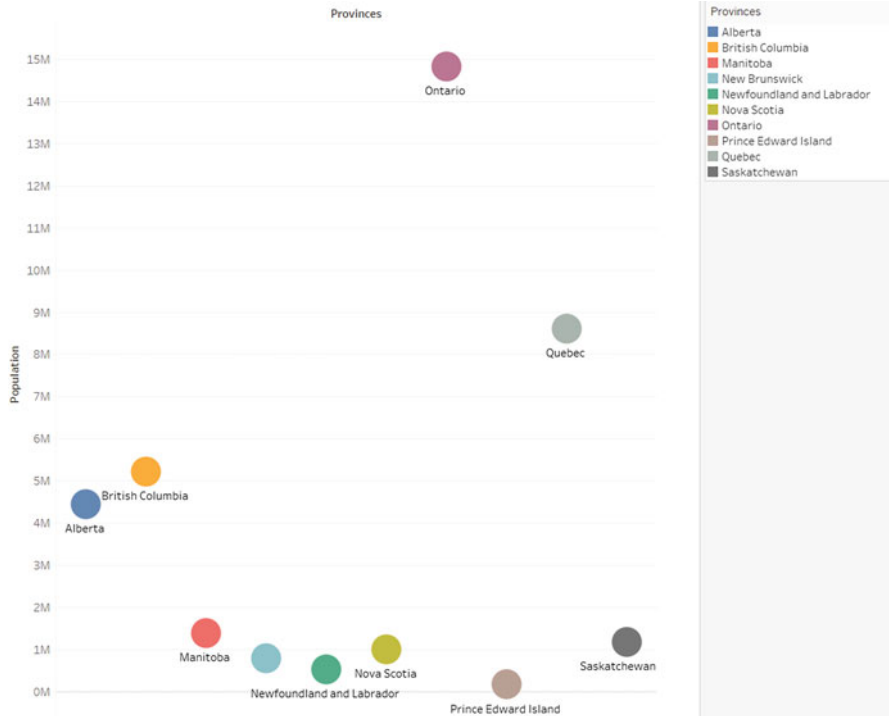


Fig. 5.5 This simple example of a *scatterplot* shows the 2021 population estimate of Canadian provinces. With this scatterplot, the significantly large size of the population in Ontario is immediately evident, as is the very low population of Prince Edward Island. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

Shapes with 2D areas represent values in proportion to their area rather than their location on the graph. A popular example is the *pie chart* (Fig. 5.11), where each sector of the pie represents a percentage of the whole. However, despite its frequent use, a pie chart is not recommended when the compared values are close or when there are many categories or sectors to compare [1, 2].

Another example of shapes with 2D areas is the *bubble*, which is a scatterplot that quantifies three values, two by their relative location on each axis and the third by the size of the bubble. A fourth variable can be quantified by applying *variable intensities of the same color* to the bubbles [2] or simply by using different colors (Fig. 5.12).

5.2.2 Relationships and Graphs

Graphs are used to display relationships in data by giving them shapes. There are eight main types of relationship graphs that are typically used: time series, ranking, part-to-whole, deviation, distribution, correlation, geospatial, and nominal comparison [2]. *Time series* graphs show how something changed (increased, fluctuated,

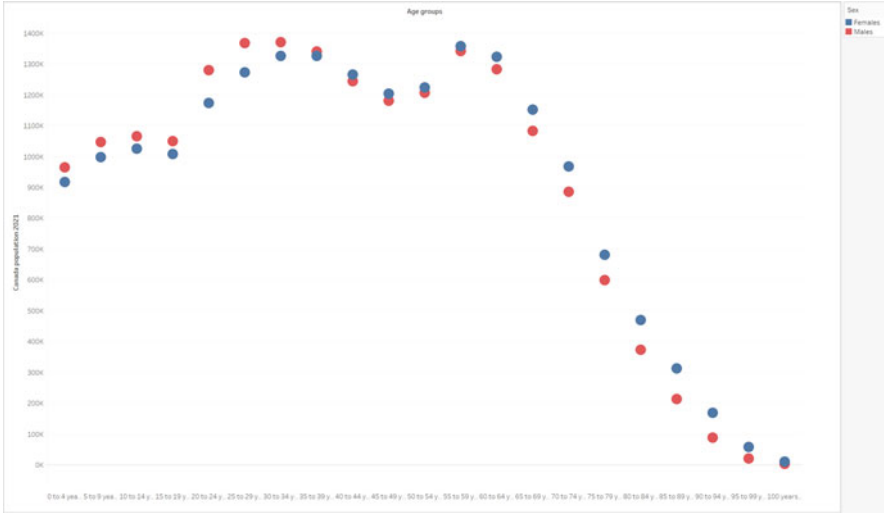


Fig. 5.6 This example of a *scatterplot* shows the population of Canada in 2021 by age group and by sex. We notice that until the age of 40, the number of males exceeds the number of females and that this is reversed afterward. We also see a dip in the population aged around 40–60 years. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

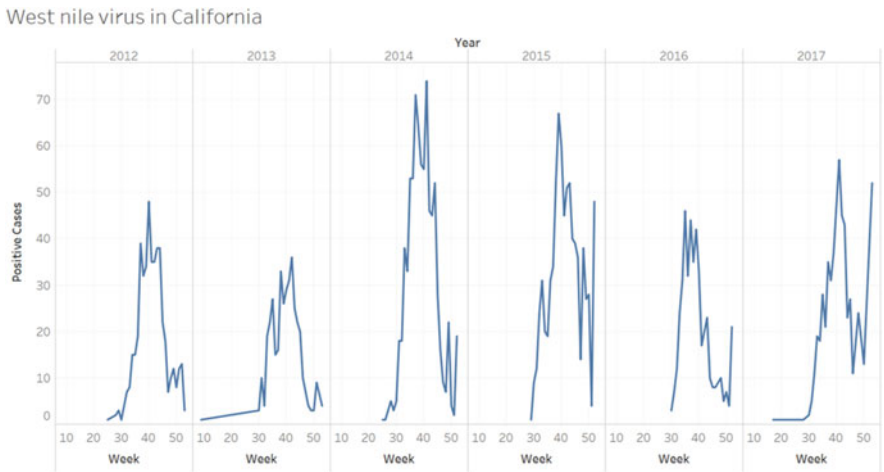


Fig. 5.7 This example of a *line chart* or time series displays the incidence of the West Nile virus in California on a weekly basis from 2012 to 2015. It clearly shows a yearly cycle where the incidence peaks between weeks 39 and 42. This graph was generated using Tableau Desktop software and the California Department of Public Health Open Data [12]

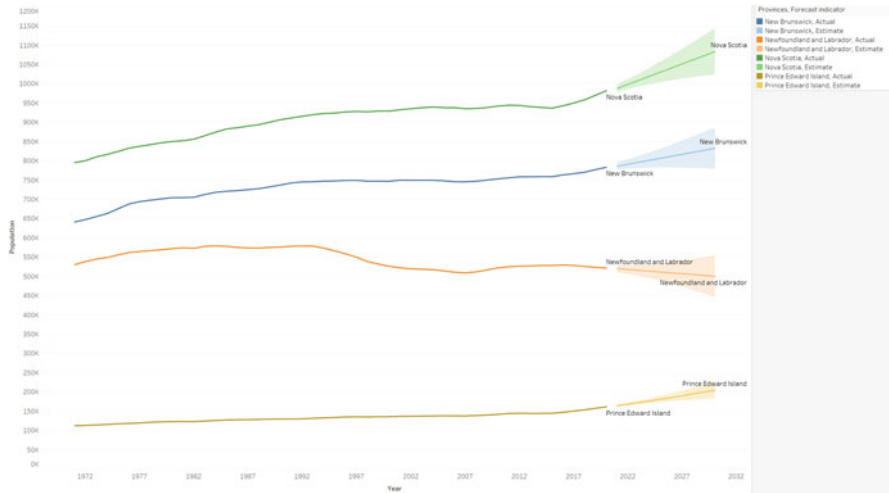


Fig. 5.8 This example of a *line chart* highlights the trend of slow population increase in three Canadian Maritime Provinces, and a decreasing trend in the fourth, from 1972 to 2021. The software generated a forecast statistic until 2032. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

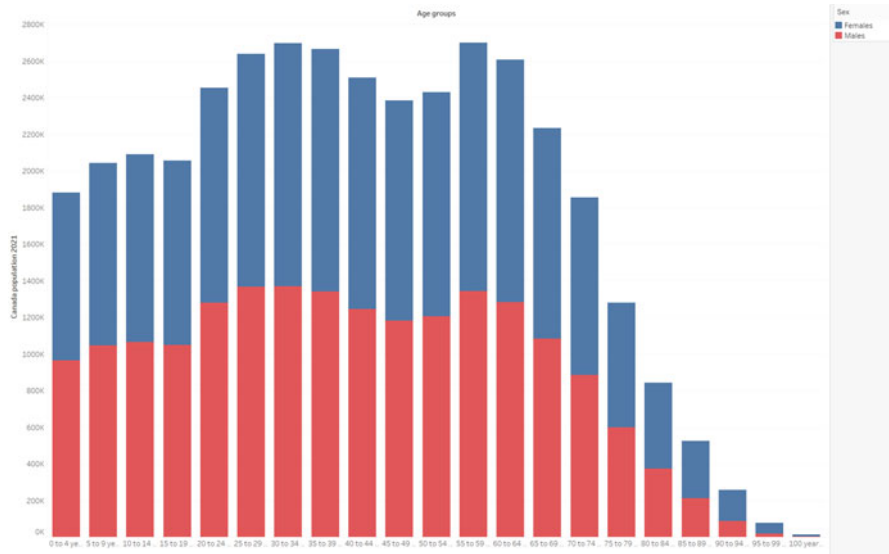


Fig. 5.9 This example of a *bar graph*, also called a column chart, or a stacked bar chart, shows the population of Canada in 2021 by age group and by sex. It is a different visualization of the same data in Fig. 5.6. This graph was generated using Tableau Desktop software Statistics Canada population open data [10]

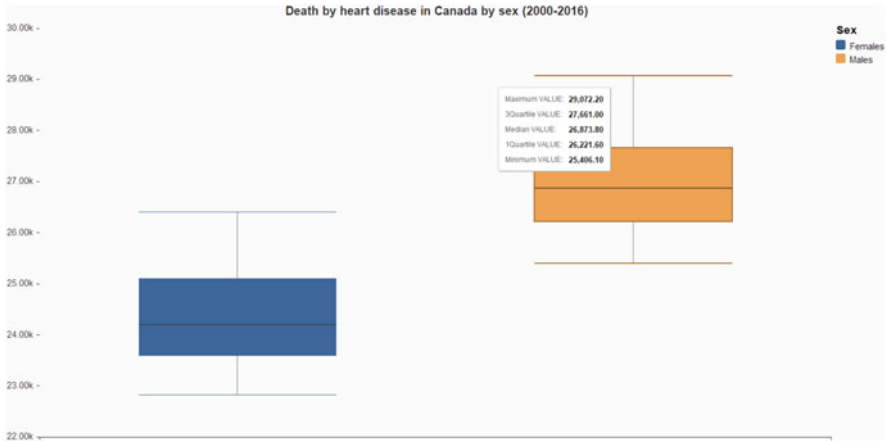


Fig. 5.10 This example of a box graph, also known as a *box plot*, represents death by heart disease by sex in Canada (2000–2016). Each box displays the minimum, first quartile, median, third quartile, and maximum values in the dataset. This graph was generated using SAP Lumira software and Statistics Canada leading cause of death open data [13]

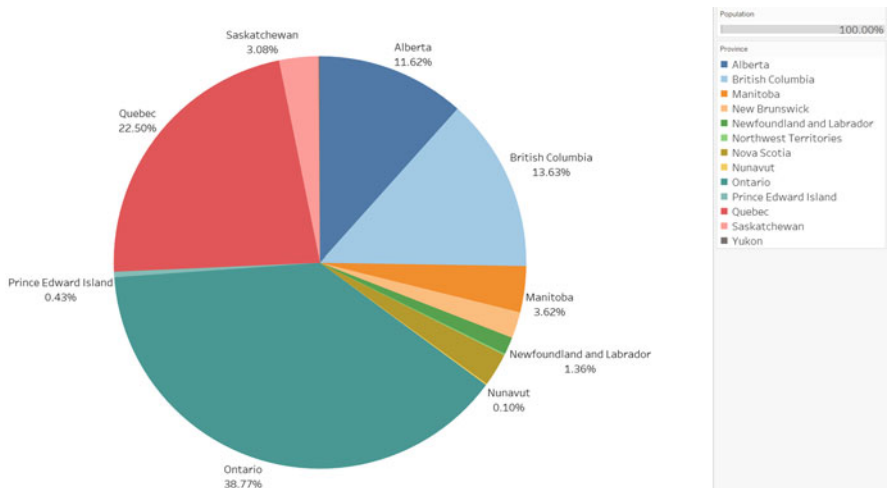


Fig. 5.11 This example of a *pie chart* depicts the distribution of the Canadian population by province, clearly highlighting the relative population size in each. It is a different visualization of the same data in Fig. 5.5. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

declined, etc.) over time (e.g., Figs. 5.7 and 5.8). Graphs display *ranking* relationships such as larger than, smaller than, and equal to, sorted in increasing or decreasing order (Fig. 5.9, though, is not sorted). Graphs display *part-to-whole*

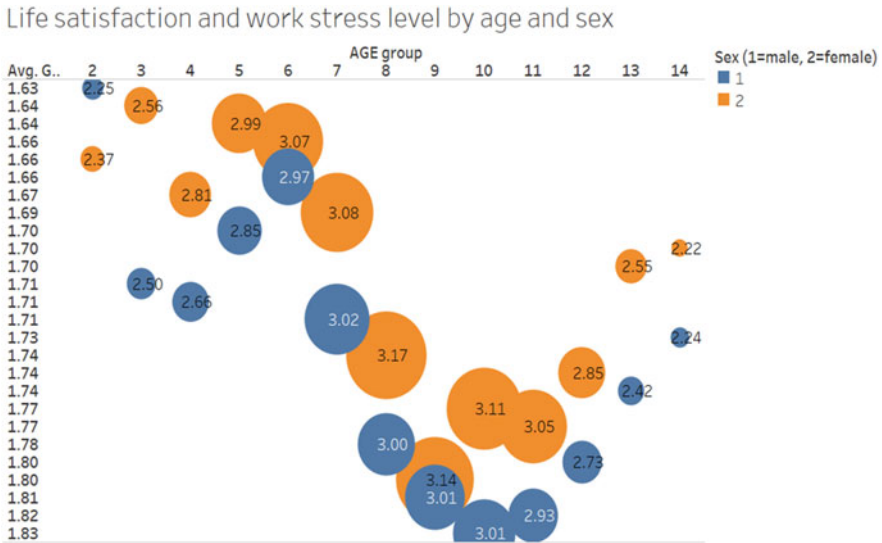


Fig. 5.12 This example of a *bubble chart* displays the average life satisfaction level (y-axis) and work stress level (bubble size), by age group (x-axis) and sex (color) in Canada (2012). Among the visual observations are that after a certain age, females have a lower life satisfaction and higher stress level, but that there is no clear relationship between life satisfaction and work stress in general. This graph was generated using Tableau Desktop software and Statistics Canada Community Health open data [14]

relationships by showing how individual values make up the whole of something (for example, by percentage or rate of total) and how they compare to each other (Figs. 5.9 and 5.11). *Deviations* represent how one or more sets of values differ from a reference set of values (Fig. 5.13) [2].

A *distribution* represents how values are distributed across an entire range, from the lowest to the highest, and is called a *frequency distribution* when it shows the number of times something occurs. When bars are used, it is referred to as a histogram (Fig. 5.14) [2]. Histograms group the data into specific categories known as bins and assign a bar size that is proportional to the number of records in each bin [11].

A graph displays a *correlation* when it shows whether two sets of values vary (increase, decrease, follow) in relation to each other, positively or negatively, and to what degree (e.g., Figs. 5.6, 5.9, and 5.12). *Geospatial* relationships between values are displayed by plotting them on a map (Fig. 5.15). Finally, a *nominal comparison* is the simple display of a set of discrete quantitative values so that they can be easily read and compared (e.g., Fig. 5.13) [2].

To display a specific relationship graphically, different objects and types of graphs can be used, with some being more adequate for the task than others, while others should be avoided. Table 5.1 is a summary of the recommended graphical objects used to display each type of relationship described above.

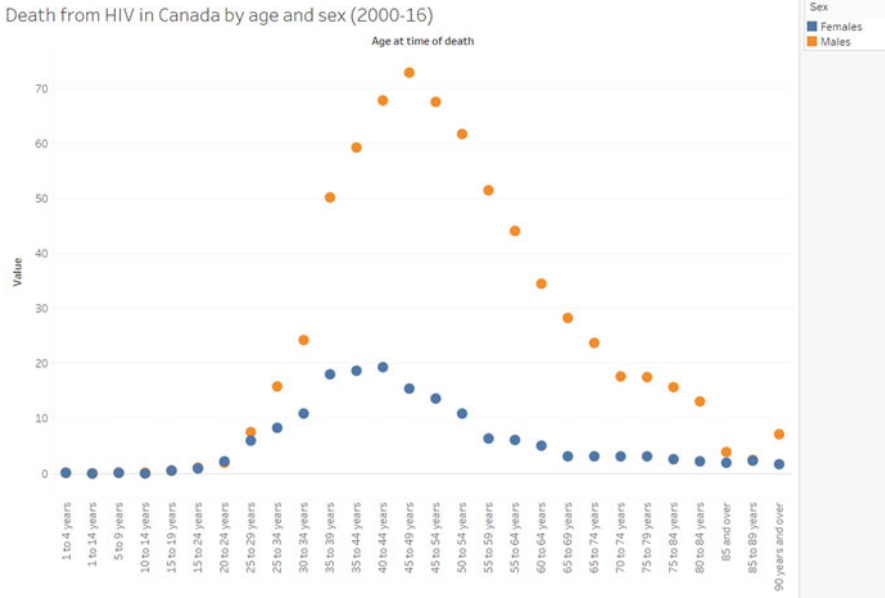


Fig. 5.13 This graph represents the number of deaths per 100,000 citizens from HIV in Canada by age group and sex (2000–2016). It shows a significant and very clear *deviation* for males, compared to females, after the age of 25. This graph was generated using Tableau Desktop software and Statistics Canada leading cause of death open data [13]

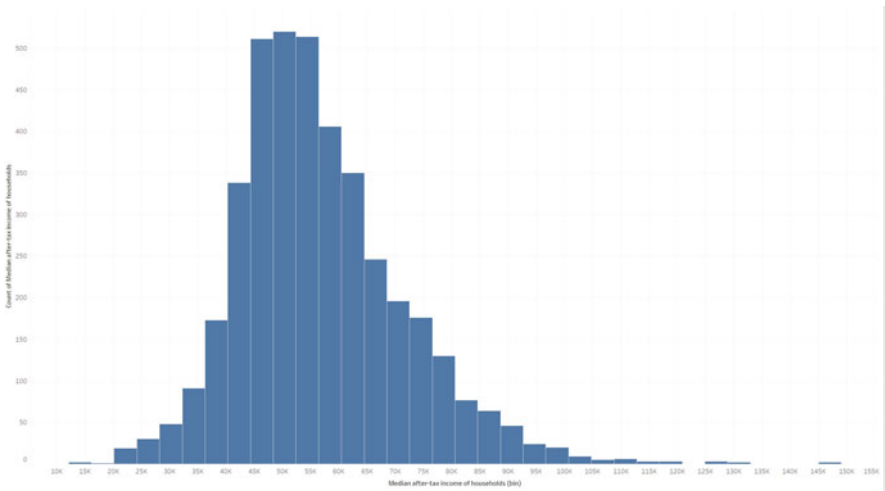


Fig. 5.14 This example of a *histogram* displays the *distribution* of households’ median after-tax income (2016). This graph was generated using Tableau Desktop software and Statistics Canada 2016 Census open data [15]



Fig. 5.15 This example of a *geospatial map* displays the population by province in Canada, where a larger font indicates a larger population in the different Canadian provinces and territories in 2016. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

In addition to the visualizations shown so far, there is a large number of possible visualizations, many of which are very popular and useful. Below are some additional examples of popular or interesting advanced visualizations (Figs. 5.16, 5.17, 5.18, 5.19, 5.20 and 5.21).

There are infinite additional ways to visualize data and information. What has been covered so far in this chapter is an introduction to the basic and most popular visualizations. To view additional examples of interesting and rich visualizations, you can explore numerous sources such as the Information Is Beautiful website (<https://informationisbeautiful.net/>), the Data Visualization Catalogue (<https://datavizcatalogue.com/>), and Tableau’s public gallery (<https://public.tableau.com/en-us/s/gallery>).

A popular guide for selecting a chart based on the number of variables, the kind of comparison needed, and the time frame can be found at the Extreme Presentation website (https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html).

A number of excellent interactive visualizations can be found at different websites, such as Statistics Canada’s Interact with Data site (<https://www.statcan.gc.ca/en/interact>). Two examples can be found in Figs. 5.22 and 5.23. Another good example is the website of the Institute for Health Metrics and Evaluation (IHME), which is an independent global health research center at the University of Washington (<http://www.healthdata.org/results/data-visualizations>). Most visualizations are interactive with filters that allow the viewer to select from a variety of graph

Table 5.1 Graphical object to use for each type of relationship (adapted from Few (2012) [2])

Relationship	Graphical Objects			
	Points	Lines	Bars	Boxes
Time series (categorical data on the <i>x</i> -axis and quantitative values on the <i>y</i> -axis)	Dot plot only when values were not collected at consistent intervals of time	For emphasis on the overall pattern (e.g., Figs. 5.7 and 5.8)	For emphasis on individual values	Only when showing distributions that change over time
Ranking	Dot plot when the quantitative scale does not start at 0; otherwise, use bar	Avoid	Horizontal or vertical (e.g., Fig. 5.9, but preferably sorted)	Only when ranking multiple distributions; horizontal or vertical
Part-to-whole	Avoid	To display how parts of a whole change over time	Horizontal or vertical (e.g., Fig. 5.9)	Avoid
Deviation	Dot plot when the quantitative scale does not start at 0	Useful when combined with time series	Horizontal or vertical; always vertical if combined with time series	Avoid
Distribution (single)	Known as a strip plot; emphasis on individual values	Known as a frequency polygon; emphasis on overall pattern	Known as a histogram; emphasis on individual intervals (e.g., Fig. 5.14)	Avoid
Distribution (multiple)		Known as a frequency polygon. Limit to a few lines	Avoid	Known as a box plot (e.g., Fig. 5.10)
Correlation	Known as a scatterplot (e.g., Fig. 5.6)	Avoid	Horizontal or vertical	Avoid
Geospatial	Different point sizes encode values (e.g., Fig. 5.16)	Used to mark routes (e.g., Fig. 5.16)	Avoid	Avoid
Nominal comparison	Use a dot plot if scales do not start at 0	Avoid	Horizontal or vertical; use when the scale starts at 0 (e.g., Fig. 5.9)	Avoid

types, regions, dates, measures, indicators, etc. Some are also dynamic and display data evolving over a period of time. Another very common type of interactive visualization is the dashboard, which is introduced next.

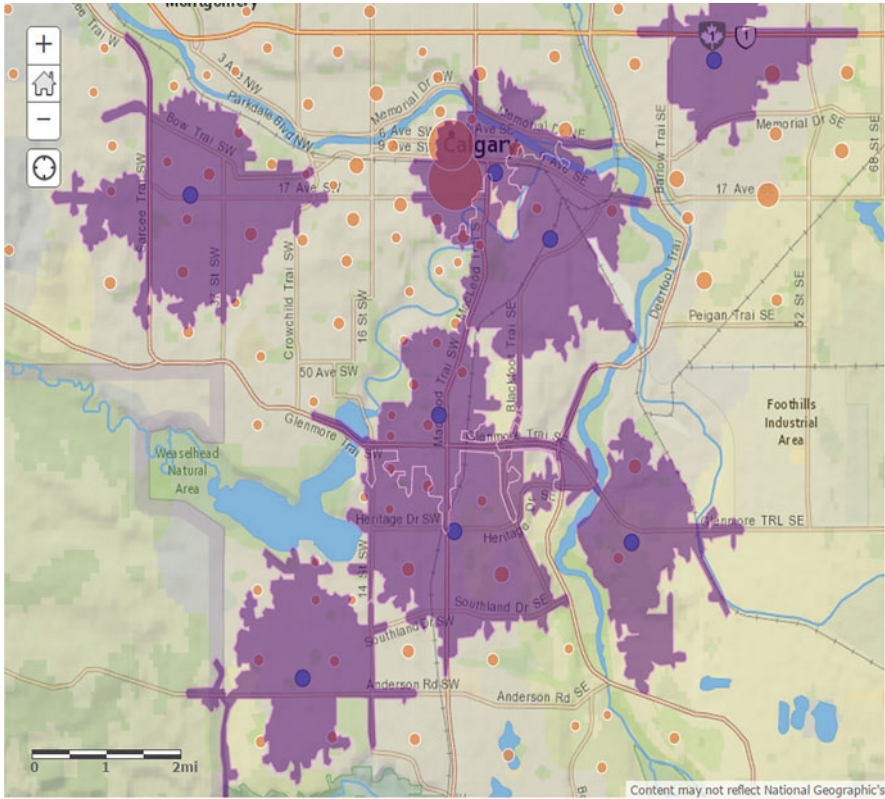


Fig. 5.16 This example of a *geospatial graph* shows the typical distance that can be reached in 5 minutes from the police stations in the city of Calgary in Canada (blue dots). The orange circles in different sizes represent the relative crime rate in the different municipalities of the city. This graph was generated using the ESRI ArcGIS Online tool [16] and the city of Calgary open data [17]

5.2.3 Dashboards

The dashboard is “a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance” [20]. It is a collection of related visualizations that are tied together through interactivity and are displayed on a single page and can combine multiple different types of data in a single location [21]. Dashboards can be used, for example, to monitor marketing campaigns’ landing pages, conversion rates, visitors by location, lead by campaign source, and other key performance indicators (KPIs) (Fig. 5.24). This interactive analytical dashboard allows the user to select dates and regions for analysis. Another example is an e-commerce dashboard (Fig. 5.25), which can be used to analyze and monitor sales and revenue from different perspectives. More examples of dashboards from different industries and

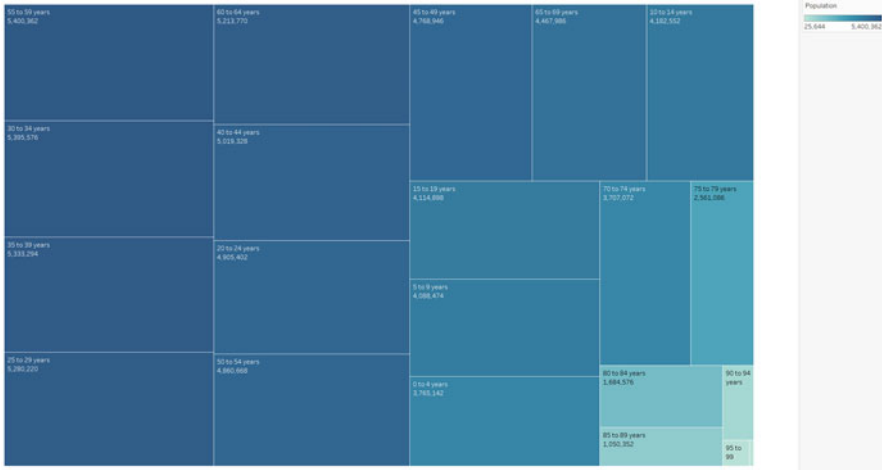


Fig. 5.17 This is an example of a *heat map*, displaying the distribution of the Canadian population by age group (2021). The larger the rectangle and the darker (or “hotter”) the color, the higher the population. We see here that the 55–59 years old is the largest group in Canada. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]

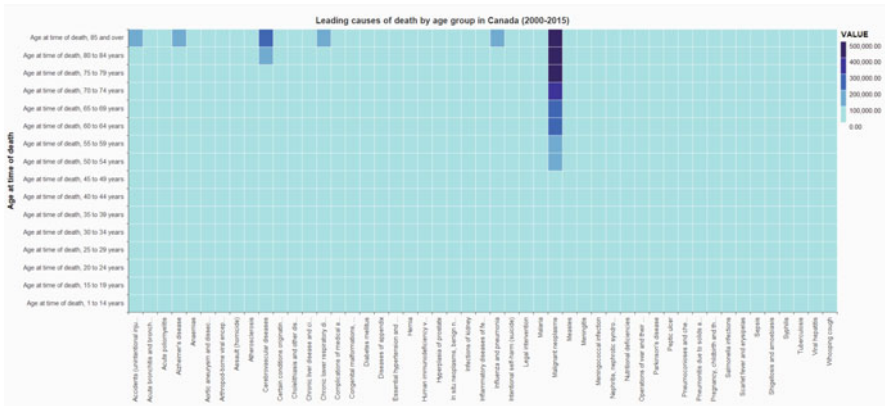


Fig. 5.18 This is an example of a *heat map* displaying the leading causes of death by age in Canada (2000–2015). The darker the color, the higher the total number of deaths. The main difference between a heat map and a treemap is that the latter can enable a hierarchical presentation of additional variables [1]. This graph was generated using SAP Lumira software and Statistics Canada leading cause of death open data [13]

job functions can be found on the websites of business intelligence companies such as Qlik (<https://www.qlik.com/us/dashboard-examples>) and Sisense (<https://www.sisense.com/dashboard-examples/>).

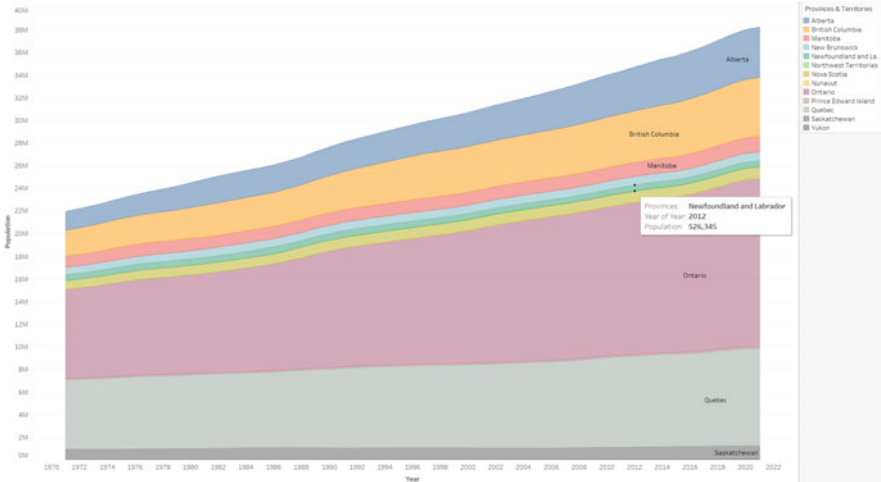


Fig. 5.19 This example of an *area chart* displays the growth of the population by province in Canada between 1971 and 2021. This graph was generated using Tableau Desktop software and Statistics Canada population open data [10]



Fig. 5.20 This is an example of a *tag cloud* used to display the leading causes of death in Canada (2000–2015). The larger the text and the darker its color, the higher the total number of incidents. Tag clouds are useful for displaying words or phrases based on their frequency and hence importance [1]. This graph was generated using SAP Lumira software and Statistics Canada leading cause of death open data [13]

Dashboards can be broken down into three roles: strategic, analytical, and operational. At the executive level of an organization, dashboards support long-term strategic decisions and focus on high-level measures of performance, including forecasts. They tend to be simple and not interactive and do not require real-time data updates. Dashboards that support data analysis demand rich comparisons, more extensive history, and interaction with data, such as drilling down for more details.

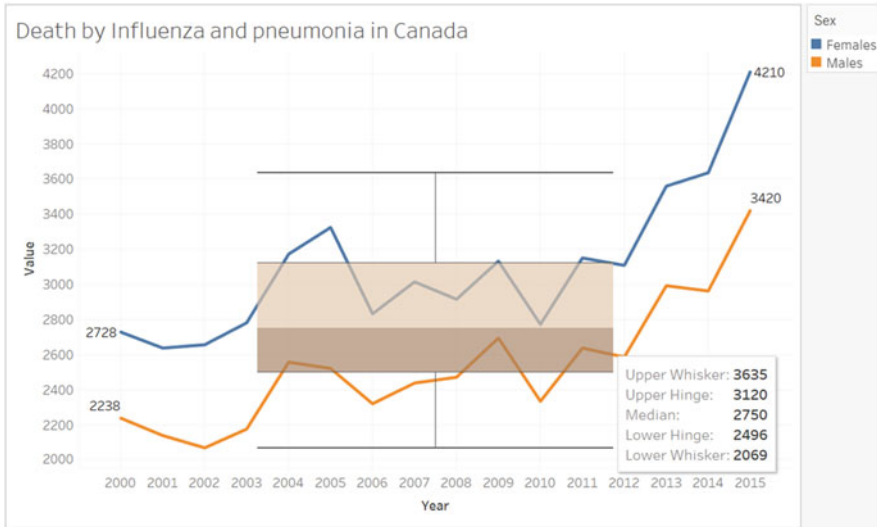


Fig. 5.21 This is an example of a chart combining different graphical objects. It displays via lines the number of deaths by influenza and pneumonia by sex from 2000 to 2015. It also includes a box plot displaying the upper and lower whiskers, median, and upper and lower hinges. This graph was generated using Tableau Desktop software and Statistics Canada leading cause of death open data [13]

They can help detect patterns in the data to identify the causes of problems, for example. Similar to strategic dashboards, analytical dashboards work with static, not real-time, data. Finally, operational dashboards are dynamic and immediate in their nature. They present real-time data in a simple way but also have the means to attract attention in cases when an operation falls outside the range of the acceptable threshold of performance [20].

A good dashboard should be designed with the most important view in the top left corner, not include too many views, use a consistent color scheme or compatible ones, and have all the filters grouped together [21].

5.2.4 Infographics

The term “infographics” is an abbreviation of “information graphics.” They are a combination of data visualizations, text, and images, presented in a logical manner similar to storytelling, and are used to convey information and messages in an attractive and easy-to-understand format [1, 24]. Infographics use many different visual cues to convey information. With the overwhelming amount of data and content generated and shared online, infographics have become very important due to their ability to present information to an audience in a way that can capture

Distributions of Household Economic Accounts, Wealth: Interactive tool

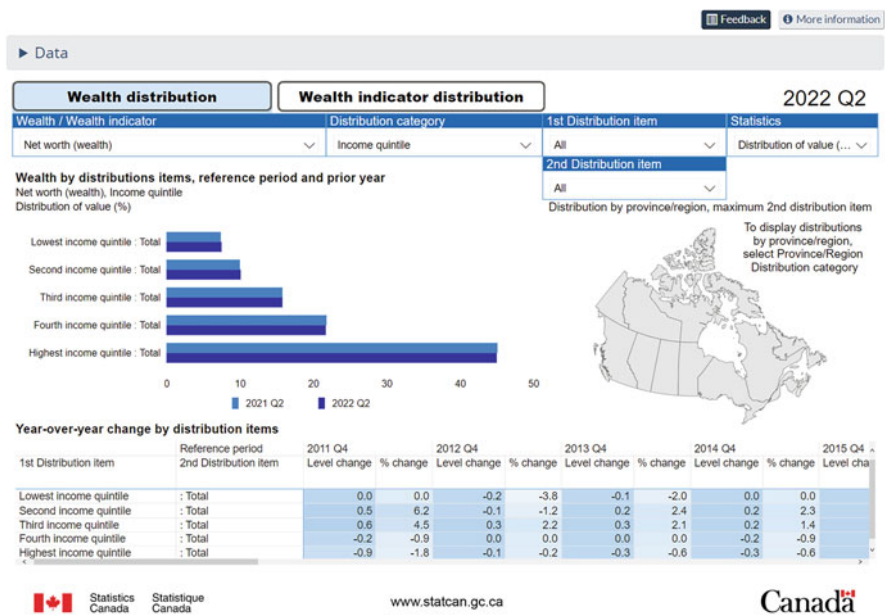


Fig. 5.22 This is an example of an *interactive visualization* consisting of a table, bar chart, and map representing the distribution of household wealth in Canada. The site allows users to select wealth indicator, distribution, and statistics. The graph was generated on Statistics Canada’s Interact with Data—Data Visualization site [18]

and keep the audience’s attention, engage them, and aid in their comprehension and retention of the material [25]. Infographics are used in multiple disciplines, such as public policy, journalism, business, and politics. In the healthcare field, infographics are used for health communication and engagement, particularly to support comprehension among individuals with low health literacy [26]. They are helpful tools for communicating key messages clearly, challenging people’s thinking, and changing behaviors and attitudes [24].

Figure 5.26 shows two examples of infographics by Statistics Canada. The first one was created to inform the public of its new 24-hour movement guidelines for children and youth [27]. It includes general information, basic statistics about the current physical activity levels in the country, and the factors that can increase them, all in a simple, clear, easy-to-understand, and visually stimulating fashion for both parents and young people. The second example is an infographic on rail transportation in Canada in 2020 [28]. Its role is to inform the public of the expenses and revenues, the origin and destination of shipments, and the volume and type of products in a simple and clear way that is easy to understand.

New Housing Price Index: Interactive Dashboard

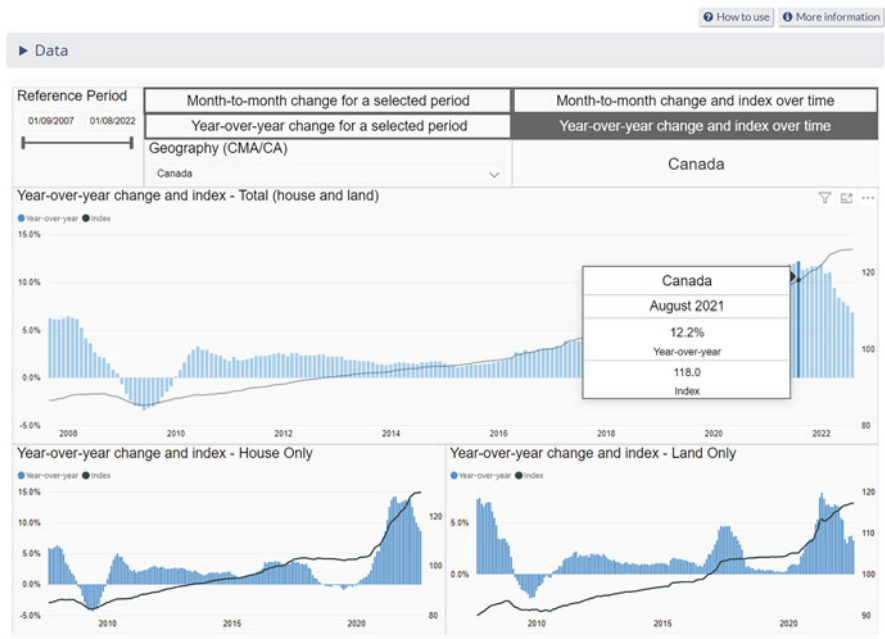


Fig. 5.23 This is an example of *interactive visualizations* consisting of combined bar (change) and line (index) charts representing the new housing price index for houses and land in Canada. The site allows users to select the reference period, region, and type of change. The graph was generated on Statistics Canada’s Interact with Data—Data Visualization site [19]

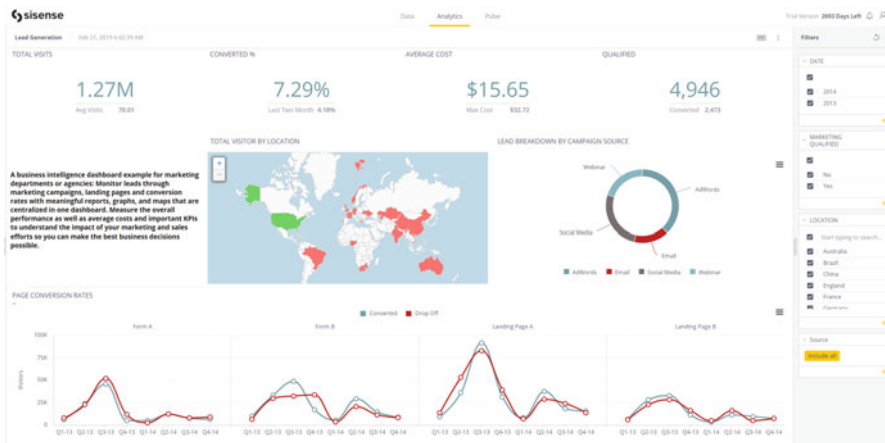


Fig. 5.24 Example of a lead generation (marketing) dashboard (Source: Sisense.com with permission [22])

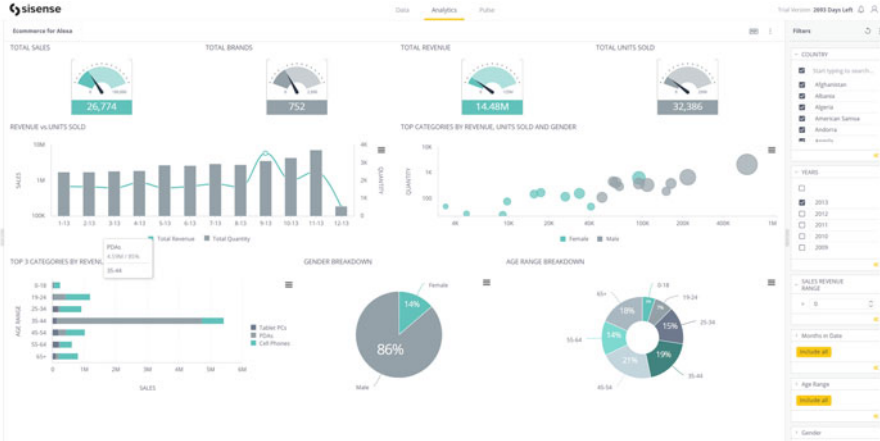


Fig. 5.25 Example of an e-commerce dashboard (Source: Sisense.com with permission [23])

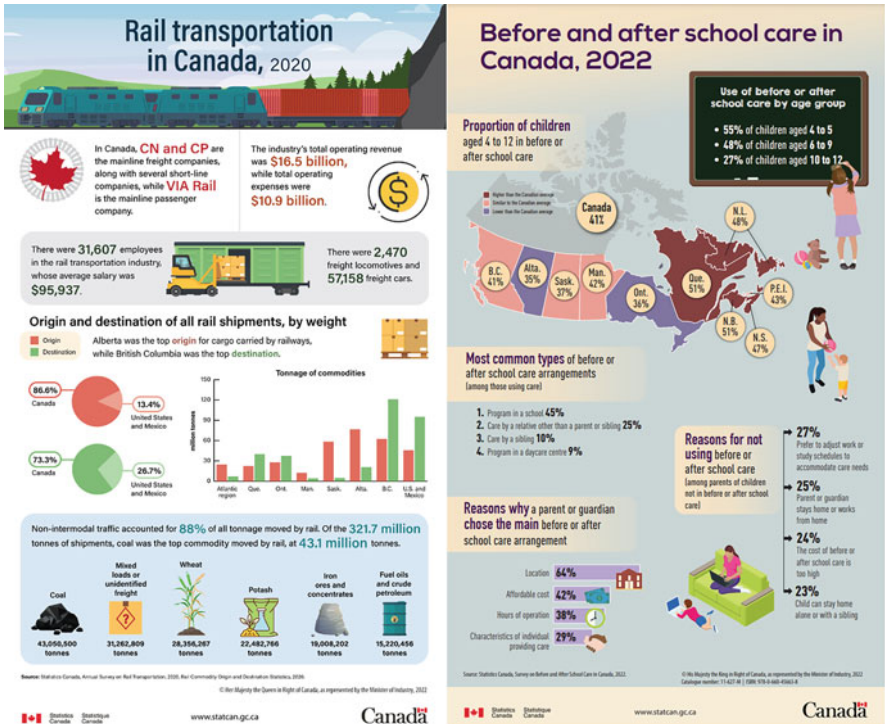


Fig. 5.26 Examples of infographics (Source: Statistics Canada [27, 28])

In general, infographics are designed without complex terminology, allowing the public to understand the message without explanation from professionals [24]. A study on the design of health-related infographics for engaging community members with varying levels of health literacy found that successful designs are rich in information but without distracting details. They support comparison, between treatments, for example, with a clear recommendation. They provide valuable contextual information and use familiar colors and symbolic analogies such as the battery charging level to represent a patient's sleep and energy levels [26].

5.3 Building Effective Visualizations

Clear communication of quantitative information is the essence of a graph. Six principles, known as ACCENT, are at the basis of effective visual display of data [4]:

- **Apprehension:** the ability to correctly recognize relationships between variables
- **Clarity:** the ability to visually differentiate the different elements of a graph
- **Consistency:** the ability to build your understanding of a graph on similarities with previous ones
- **Efficiency:** the ability to identify a complex relationship in a simple manner
- **Necessity:** the need for the graph
- **Truthfulness:** the ability to determine the true value represented by the graph

It is important to remember that despite the richness of graphs, sometimes the use of a simple table is more efficient and effective in achieving the goal of data interpretation. Tables are recommended when you want to look up individual values, compare individual values, use precise values, or use summary and detailed information in a single display [2]. Graphs are best used when the message is contained in the shape of the values, such as trends and patterns, and to reveal relationships between sets of values [2].

While graphs and charts are excellent tools for conveying information and telling stories, you should be aware of many bad visualizations that are encountered regularly. Among the sources of deficiency is the use of certain types of popular graphs that should be avoided and the bad design of appropriate graphs. Among the charts to avoid are donut charts, radar charts, circle charts, funnel charts, 3D charts, and in some cases, pie charts [2, 29]. These somewhat popular charts are visually appealing and are available in many software packages, but they fail to present information accurately, clearly, accessibly, and efficiently [2]. In terms of design, one of the most important guidelines is to avoid clutter, which includes visual elements that use space and do not increase our understanding, which increases our cognitive load or the mental effort required to learn new information [29]. Examples of clutter are too many elements, such as lines and bars, axes, data labels, colors, and text. In addition to simplicity, the data need to be put in context, so the reader can understand its meaning. The numbers that give a more faithful representation, be it percentage change or absolute value, should be used. Color and fonts should convey

information and not be used for decoration. Natural increments for the y-axis scale are required, as is a zero baseline in all bar charts. Finally, it is best to use as few graphical elements as possible to keep the visualization crisp and clean [30].

5.4 Data Visualization Software

The creation of appealing and beautiful visualizations can be achieved with a very large number of software tools, starting with the common Microsoft Excel. Today, Excel can create many different basic charts and graphs, such as columns and lines, and complex charts and graphs, such as treemaps and waterfalls. It can also create combinations of charts, such as clustered columns and lines (Fig. 5.27). For the novice user, Excel recommends the most appropriate charts to use based on the data selected in the spreadsheet.

Business analytics tools such as Tableau Desktop, Microsoft Power BI, and Lumira by SAP provide a very large number of visualizations that are easy to use and do not require advanced technical knowledge. Most of the work done is via simple pointing, clicking, and dragging with the mouse. These advanced analytics tools can also interpret the data to identify dimensions and measures, which are comparable to the categorical and quantitative data discussed earlier in the book. The

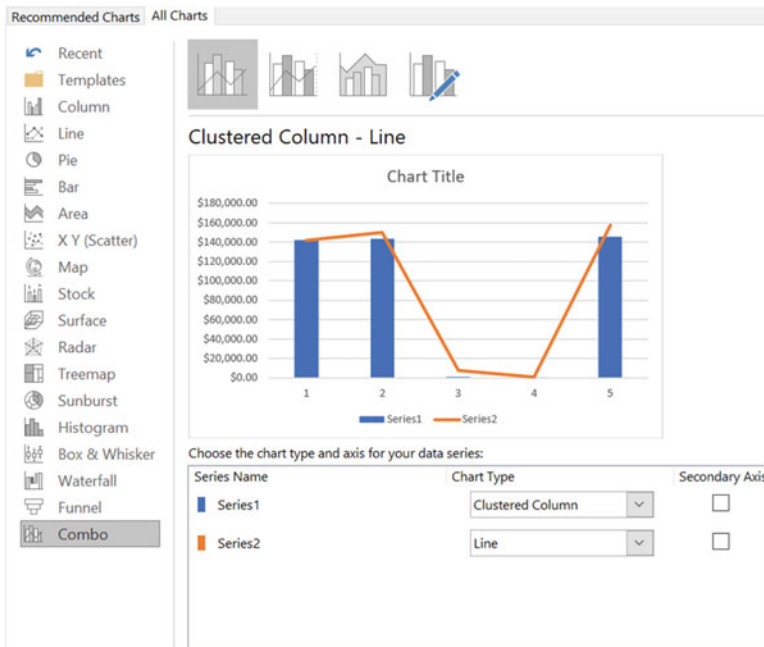


Fig. 5.27 Example of a visualization with Microsoft Excel

charts to use are recommended based on the data available. Tableau Desktop, for example, has a Show Me button (Fig. 5.28) that highlights the available charts that can be used based on the data and makes suggestions for using them. Such tools also allow you to easily create presentations, infographics combining different charts, and dashboards connected to dynamic data sources.

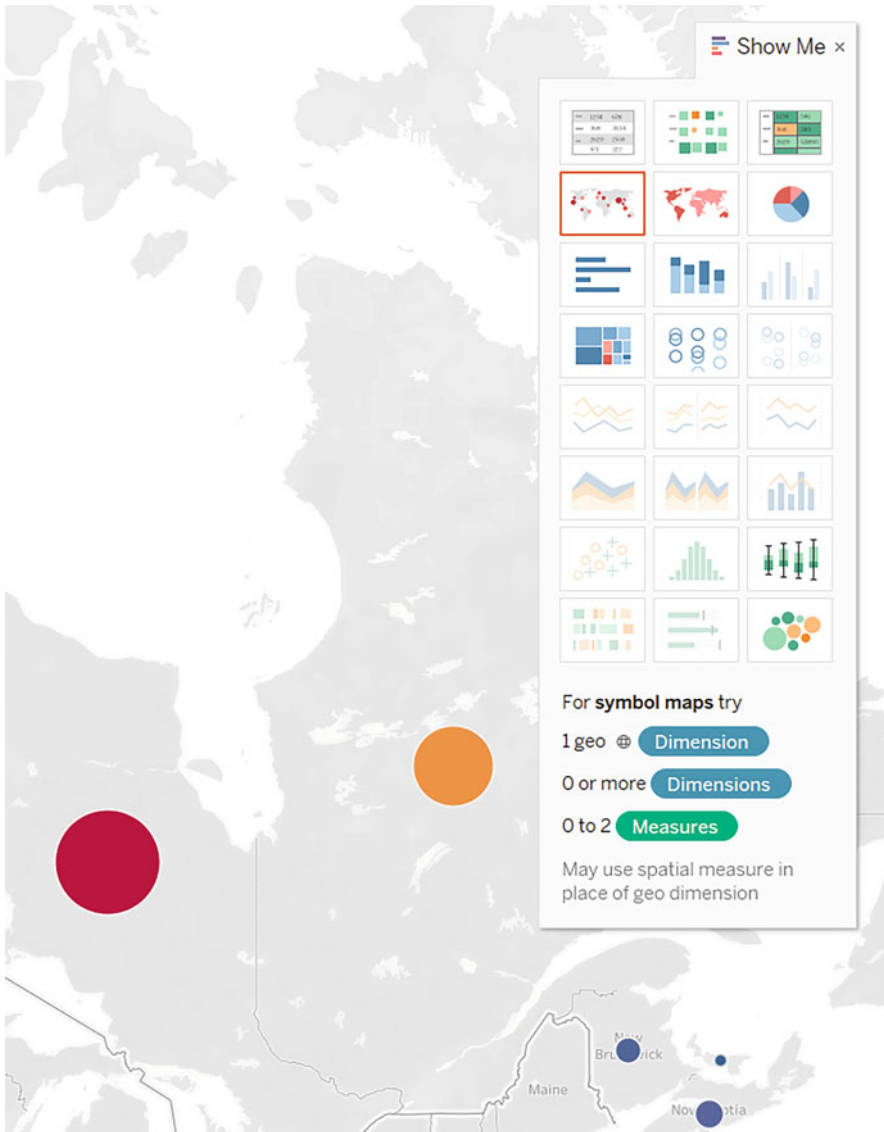


Fig. 5.28 Tableau analytics tool interface with the Show Me button

While we have discussed Tableau, SAP Lumira, and Microsoft Excel in this chapter and used them to generate the visualizations above, it is important to note that there are many large software companies, such as SAS and IBM, and relatively smaller niche players, such as QlikView and Sisense, that have software with remarkable visualization capabilities. According to Gartner’s 2022 Magic Quadrant, the leading platforms in analytics and business intelligence are Microsoft, Salesforce (Tableau), and Qlik [31]. An article by *PC Magazine* lists Microsoft Power BI, Tableau Desktop, Sisense, Domo, Google Analytics, Salesforce Einstein, Zoho, SAP Analytics Cloud, and Chartio as the nine best data visualization tools [32].

In addition to the visualization tools or applications mentioned above, there are many open-source libraries that allow analysts to present data in an interactive way and engage a broad audience with new data [33]. An example of open-source visualization libraries is D3.js (<https://d3js.org/>), where D3 stands for “Data Driven Documents.” It is a JavaScript library for producing dynamic, interactive data visualizations in web browsers, with features for interactions and animations. D3.js uses HTML, CSS, and SVG to create data visualizations to be viewed on any browser [33]. Another example is Google Charts (<https://developers.google.com/chart>), which provides interactive charts for browsers and mobile devices. It uses JavaScript and has a rich gallery of charts, is customizable, connects to dynamic data, and provides interactivity and dashboards.

5.5 Conclusion

Data visualization is a critical capability for understanding and interpreting complex data and relationships. Graphs and charts can tell a story, highlight trends, identify outliers and deviations, make comparisons, and more in a simple and effective way. There are many types of graphs and charts available, and selecting the one that best matches the data and the questions you are trying to answer is crucial. Bad visualizations are difficult to understand and can distort what the data are trying to tell us. Today, it is easy to create very rich visualizations using modern analytics tools with simple pointing and clicking; however, it remains critical to have a good understanding of the data to select the best visualization and be able to interpret it.

5.6 Key Terms

1. Scatterplot
2. Line charts
3. Bar charts
4. Box-and-whisker plots
5. Boxplot
6. Pie chart

7. Bubble chart
8. Time series graphs
9. Ranking relationships
10. Part-to-whole relationships
11. Deviations
12. Distribution
13. Frequency distribution
14. Histograms
15. Correlation
16. Geospatial relationships
17. Nominal comparison
18. Dashboard
19. Infographics
20. Analytics tools
21. Tableau Desktop
22. Power BI
23. Lumira

5.7 Test Your Understanding

1. Cite an example where would you use a scatterplot.
2. When would a line chart be a better fit to the objectives than a bar chart?
3. What is the benefit of a boxplot? Draw two boxplot examples to make your point clear.
4. Your city is looking for the best way to plot poverty levels on a map. What type of visualization tools would you suggest?
5. What is the difference between a bar chart and a histogram? Give a few examples that illustrate the difference.

5.8 Read More

1. Birnbaum, D. (2021). Regarding data visualization. *Infect Control Hosp Epidemiol*, 42(9), 1154–1155. <https://doi.org/10.1017/ice.2020.457>
2. Byrd, V., & Dwenger, N. (2021). Activity Worksheets for Teaching and Learning Data Visualization. *IEEE Comput Graph Appl*, 41(6), 25–36. <https://doi.org/10.1109/mcg.2021.3115396>
3. Min, S. H., & Zhou, J. (2021). smplot: An R Package for Easy and Elegant Data Visualization. *Front Genet*, 12, 802,894. <https://doi.org/10.3389/fgene.2021.802894>
4. Nguyen, V. T., Jung, K., & Gupta, V. (2021). Examining data visualization pitfalls in scientific publications. *Vis Comput Ind Biomed Art*, 4(1), 27. <https://doi.org/10.1186/s42492-021-00092-y>

5. Pakarinen, T., & Ojala, J. (2021). Profeel-An open source dosimetry data visualization and analysis software. *Comput Methods Programs Biomed*, 212, 106,457. <https://doi.org/10.1016/j.cmpb.2021.106457>
6. Park, S., Bekemeier, B., Flaxman, A., & Schultz, M. (2021). Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. *Inform Health Soc Care*, 1–19. <https://doi.org/10.1080/17538157.2021.1982949>
7. Park, S., Bekemeier, B., & Flaxman, A. D. (2021). Understanding data use and preference of data visualization for public health professionals: A qualitative study. *Public Health Nurs*, 38(4), 531–541. <https://doi.org/10.1111/phn.12863>
8. Senanayake, D. A., Wang, W., Naik, S. H., & Halgamuge, S. (2021). Self-Organizing Nebulous Growths for Robust and Incremental Data Visualization. *IEEE Trans Neural Netw Learn Syst*, 32(10), 4588–4602. <https://doi.org/10.1109/tnnls.2020.3023941>
9. Shee, K., Pal, S. K., Wells, J. C., Ruiz-Morales, J. M., Russell, K., Dudani, S., Choueiri, T. K., Heng, D. Y., Gore, J. L., & Odisho, A. Y. (2021). Interactive Data Visualization Tool for Patient-Centered Decision Making in Kidney Cancer. *JCO Clin Cancer Inform*, 5, 912–920. <https://doi.org/10.1200/cci.21.00050>
10. Wang, Q., Chen, Z., Wang, Y., & Qu, H. (2021). A Survey on ML4VIS: Applying MachineLearning Advances to Data Visualization. *IEEE Trans Vis Comput Graph*, Pp. <https://doi.org/10.1109/tvcg.2021.3106142>
11. Wu, A., Wang, Y., Shu, X., Moritz, D., Cui, W., Zhang, H., Zhang, D., & Qu, H. (2021). AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization. *IEEE Trans Vis Comput Graph*, Pp. <https://doi.org/10.1109/tvcg.2021.3099002>

5.9 Lab

5.9.1 Working Example in Tableau

In this chapter, you will learn how to perform basic visualizations using Tableau Desktop. Tableau Desktop is one of the few major data visualization applications that work on both Windows and Mac computers. You will be provided the instructions to get a student copy of Tableau Desktop and be guided to a number of tutorials where you will follow up with the demonstration videos and practice with the provided data files.

5.9.1.1 Getting a Student Copy of Tableau Desktop

Go to Tableau for students (<https://www.tableau.com/academic/students>) and click on the “Get Tableau for Free” button. You will be asked to provide your student information in order to get a one-year free Tableau Desktop license. Until your

credentials are verified, you can use a free 14-day license. You will be provided instructions to download Tableau Desktop for Windows or Mac.

Once your one-year student license expires, if you are still a student, you may request an extension by resubmitting a request at www.tableau.com/studentlicense.

5.9.1.2 Learning with Tableau’s how-to Videos and Resources

The simplest way to learn how to create visualizations is to use the how-to videos provided by Tableau at <https://public.tableau.com/en-us/s/resources>. While the videos refer to Tableau Public, the cloud-based visualization application, you can follow the instruction and use the provided data files with Tableau Desktop, which you have downloaded and installed. The videos, which mostly range between 3 and 7 minutes, include step-by-step instructions that you can follow using the respective data file.

The how-to-videos teach you how to connect to data in Excel and CSV formats, Google Sheets, Web Data Connectors, spatial files (for maps), and PDFs. They will teach you how to work with the data and prepare it by cleaning, structuring, pivoting, and merging. You will learn to understand the logic of charts, how to create them, and how to use the “Show Me” feature in Tableau. You will learn how to create and format dashboards and stories. Finally, you will learn how to make visualizations for multiple devices and for sharing on the web.

5.9.2 Do It Yourself

5.9.2.1 Assignment 1: Introduction to Tableau

Go to <https://public.tableau.com/en-us/s/resources> and watch the first video, entitled “Tableau Public Overview.” Follow the instructions and apply them Tableau Desktop. There’s no need to create a Tableau Public account and upload your file. Insert the two required screenshots below and submit this answer sheet and your saved Tableau file (yourname.twb) to your instructor. Use a different color or format than the one in the video demo.

1. Insert a screenshot of the map of **Europe** showing the CO2 emission per

capita (*Hint: use these buttons:* ).

2. Insert a screenshot of your dashboard (two charts) showing **Canada’s** emission trend. (*Hint: click on Canada on the map.*)

5.9.2.2 Assignment 2: Data Manipulation and Basic Charts with Tableau

Go to <https://public.tableau.com/en-us/s/resources> and watch videos 7–12 inclusive. Follow the instructions and apply them in Tableau Desktop. Insert the required screenshots below and describe in a couple of sentences what you learned from each video. Use a different color or format than the video demo. Submit this answer sheet and your saved Tableau file (yourname.twb) to your instructor.

1. Insert a screenshot and summary from video #7: Data Preparation—The Data Interpreter.
2. Insert a screenshot and summary from video #8: Data Preparation—Pivoting your Data.
3. Insert a screenshot and summary from video #9: Data Preparation—Splitting your Data.
4. Insert a screenshot and summary from video #10: Data Preparation—Joins and Unions.
5. Insert a screenshot and summary from video #11: Creating Your First Chart.
6. Insert a screenshot and summary from video #12: Using the Show Me Tool Bar.

5.9.3 Do More Yourself

5.9.3.1 Assignment 3: Charts and Dashboards with Tableau

Go to <https://public.tableau.com/en-us/s/resources> and watch videos 13–16 inclusive. Follow the instructions and apply them in Tableau Desktop. Insert the required screenshots below and describe in a couple of sentences what you learned from each video. Use a different color or format than the video demo. Submit this answer sheet and your saved Tableau file (yourname.twb) to your instructor.

1. Insert a screenshot and summary from video #13: Understanding the Logic of Charts.
2. Insert a screenshot and summary from video #14: Combining Sheets on a Dashboard.
3. Insert a screenshot and summary from video #15: Combining Sheets on a Dashboard.
4. Insert a screenshot and summary from video #16: Dashboard Formatting.

5.9.3.2 Assignment 4: Analytics with Tableau

Watch the four Tableau videos mentioned below and follow along with the provided workbook. For each video, you follow along with on your computer using Tableau, provide two screenshots showing that you did the work. Use a different color or format than the video demo. Submit this answer sheet to your instructor.

Note: the first time you access the videos, you may need to create a free online Tableau account if you do not have one.

1. Watch the Tableau Trend Lines video, follow along, and insert two screenshots below.
2. Watch the Tableau Reference Lines video, follow along, and insert two screenshots below.
3. Watch the Tableau Forecasting video, follow along, and insert two screenshots below.
4. Watch the Tableau Clustering video, follow along, and insert two screenshots below.

References

1. N. Kalé, N. Jones, *Practical Analytics* (Epistemy Press, 2015)
2. S. Few, *Show me the Numbers: Designing Tables and Graphs to Enlighten* (Analytics Press, 2012)
3. Tableau, *What Is Data Visualization? Definition, Examples, and Learning Resources*. <https://www.tableau.com/learn/articles/data-visualization#:~:text=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data>. Accessed 1 April 2022
4. M. Friendly, *Gallery of Data Visualization*. [Online]. Available: www.datavis.ca/gallery/
5. W. Playfair, Exports and Imports to and from Denmark & Norway from 1700 to 1780., P. TimeSeries.png, Ed., ed. (Wikimedia, 1786)
6. W. Playfair, Chart showing at one view the price of the quarter of wheat, and wages of labour by the week, from 1565 to 1821. a. W. o. L. b. t. W. Chart showing at one view the price of the quarter of wheat, from 1565 to 1821.png, Ed., ed. (Wikimedia Commons, 1821)
7. E. Fink, *Three of History's Best Charts Ever According to the Economist*, vol. 2008, ed. (Tableau, 2008)
8. C. Minard, *Carte Figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812–1813*. Minard.png, Ed., ed. (Wikimedia Commons, 1869)
9. F. Nightingale, *Diagram of the Causes of Mortality in the Army in the East*. Nightingale-mortality.jpg, Ed., ed. (Wikimedia Commons, 1858)
10. Statistics Canada, *Table 17-10-0005-01 Population Estimates on July 1st, by Age and Sex*. (2021)
11. Tableau, *Which Chart or Graph Is Right for You?* (Tableau, 2021). [Online]. Available: <https://www.tableau.com/asset/which-chart-or-graph-is-right-for-you>
12. CHHS, *West Nile Virus Cases, 2006-Present*. California Department of Public Health. <https://data.chhs.ca.gov/dataset/west-nile-virus-cases-2006-present>. Accessed 8 August 2018

13. S. Canada, *Statistics Canada. Table 13-10-0394-01 Leading Causes of Death, Total Population, by Age Group.* <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310039401>. Accessed 28 June 2018
14. S. Canada, *Canadian Community Health Survey, 2012: Annual Component.* [Online]. Available: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=135927>
15. Statistics Canada, Table 11-10-0239-01 Income of Individuals by Age Group, Sex and Income Source, Canada, Provinces and Selected Census Metropolitan Areas. (2016)
16. ESRI, *ESRI ArcGIS Online.* <https://www.esri.com/en-us/arcgis/products/arcgis-online/overview>. Accessed
17. C. o. Calgary, *Open Calgary.* <https://data.calgary.ca/>. Accessed
18. Statistics Canada, *Distributions of Household Economic Accounts, Wealth: Interactive Tool.* <https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2020006-eng.htm>. Accessed 8 April 2022
19. Statistics Canada, New housing price index: Interactive dashboard. <https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2019013-eng.htm>. Accessed 8 April 2022
20. S. Few, *Information Dashboard Design* (O'Reilly, 2006)
21. Tableau, Visual analysis best practices - simple techniques for making every data visualization useful and beautiful. (2021)
22. Sisense, *Lead Generation Sample Dashboard.* <https://trial.sisense.com/app/main#/dashboards/62518c886d94960036e1cb59>. Accessed 9 April 2022.
23. Sisense, *E-commerce Sample Dashboard.* <https://trial.sisense.com/app/main#/dashboards/625191e36d94960036e1cbeb>. Accessed 9 April 2022
24. H. Scott, S. Fawcner, C. Oliver, A. Murray, *Why Healthcare Professionals Should Know a Little about Infographics*, vol 50 (Ed: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2016), p. 1104
25. J. Lankow, J. Ritchie, R. Crooks, *Infographics: The Power of Visual Storytelling* (John Wiley & Sons, 2012)
26. A. Arcia et al., Sometimes more is more: Iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *J. Am. Med. Inform. Assoc.* **23**(1), 174–183 (2015)
27. S. Canada, Physical activity of Canadian children and youth, Ed. (2017)
28. Statistics Canada, *Rail Transportation in Canada.* (2020). <https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2022024-eng.htm>. Accessed
29. C.N. Knaflc, *Storytelling with Data: A Data Visualization Guide for Business Professionals* (John Wiley & Sons, 2015)
30. D. Wong, *Guide to Information Graphics (The Wall Street Journal)* (W. W. Norton & Company, New York, 2010)
31. A. Kronz, K. Schlegel, J. Sun, D. Pidsley, A. Ganeshan, Magic quadrant for analytics and business intelligence platforms. (2022). [Online]. Available: <https://www.gartner.com/document/4012759?ref=solrAll&refval=322141776>
32. P. Baker, The best data visualization tools. https://www.pcmag.com/picks/the-best-data-visualization-tools?test_uuid=06r4MYCu5PZzCkufjQSV3po&test_variant=a. Accessed 9 April 2022
33. IBM, *Data Visualization.* https://www.ibm.com/cloud/learn/data-visualization#toc-open-source-b_jcRWNy. Accessed 20 July 2022