OSIR, March 2025, Volume 18, Issue 1, p.61-69

Outbreak, Surveillance, Investigation & Response (OSIR) Journal

Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, https://he02.tci-thaijo.org/index.php/OSIR

The Grammar of Science: Bar, Pie, Line, and Lie

Jaranit Kaewkungwal*

Mahidol University, Thailand

*Corresponding author email: jaranitk@biophics.org

Received: 29 Oct 2024; Revised: 18 Feb 2025; Accepted: 27 Feb 2025
https://doi.org/10.59096/osir.v18i1.272299

We are now in the world of "Big Data" which can be turned into actionable information. We thus need tools to analyze the data and present the data-driven information to convey our message for making decisions. Data visualization is a dominant tool to turn data into useful information and present it in a form that is easy to understand. Data visualization is defined as the graphical representation of information and data.[1] We can say that data visualization is a form of visual art that grabs our interest on the intended message as our eyes are drawn to colors and patterns, trends and outliers.[1] General types of data visualization include such as: tables, charts, graphs, geospatial, maps, infographics, dashboards, etc. Let's focus on the three most commonly used in data visualization: bar charts, pie charts and line charts.

Bar, Pie, Line Chart

Many people use the words "chart" and "graph" interchangeably. Both terms are ways of displaying data visually. For graphical typology, fundamental concepts behind the terms are defined. The chart is a table and/or diagram with the purpose to depict and compare multiple sets of large quantitative datasets which is the best option to analyze the data in detail. Graph can be considered as a subtype of charts but illustrates data in a more picture-like format which is often used to show more complex relationships and mathematical manipulations without impeding the readers with details allowing faster understanding.[2,3]

Bar Chart

A bar chart shows numeric values of a variable plotted as bars on one chart axis, for levels of another categorical (discrete) variable plotted on the other axis. The length of each bar corresponds to the bar's value. The numeric values may be a simple frequency count, proportion, average, total, or some other summary measures computed separately for each category.[4] For example, bars represent numbers (count) of infected persons in four risk groups (A–D), mean pain scores of four groups of cancer patients (A–D), or means of percentage change of healthcare expenditure from baseline survey among people classified in four socioeconomic levels (A–D) (Figure 1(a)–(c)). A bar chart can be used to compare discrete data or show trends over time.[5,6] Other commonly found bar chart types include: clustered or grouped bars, proportional (stacked) bars, and histograms.

Clustered bar chart

A clustered bar chart, multi-set bar chart, or grouped column chart is a group of bar graphs that is used to show a distribution of data points or perform a comparison of numeric values across multiple categories (subgroups).[6] For example, bars represent the numbers (count) of study participants in three age groups within each of the three risk groups (A–C) (Figure 1(d)).

Stacked bar chart

A stacked bar chart is an extended bar chart when a second categorical variable is added to divide each of the groups in the original categorical variable, making subgroups within a bar. It is useful for comparing proportional contributions within a category.[6] For example, bars represent numbers (count) of deaths by three causes (unintended injuries, homicide, suicide) among three risk groups (A–C) (Figure 1(e)).

Histogram

A histogram can be considered as a type of bar chart that depicts frequency values. It is different from the standard bar chart in which its primary variable is categorical in nature, whereas a histogram's primary variable is continuous and numeric.[4,7] By counting and grouping the values in a dataset into bins (intervals) based on the frequency with which they occur, a histogram then shows a distribution of the values of data within that dataset.[7,8] For example, a histogram represents the distribution of knowledge scores among students in a school (Figure 1(f)). Some have said that a histogram is not exactly a bar chart. A histogram shows the frequency of numerical data as intervals in the form of bars while a bar chart shows the numerical values of different categorical entities. The bars of a histogram cannot be reordered but the bars in a bar chart can be reordered.[5,7,9]

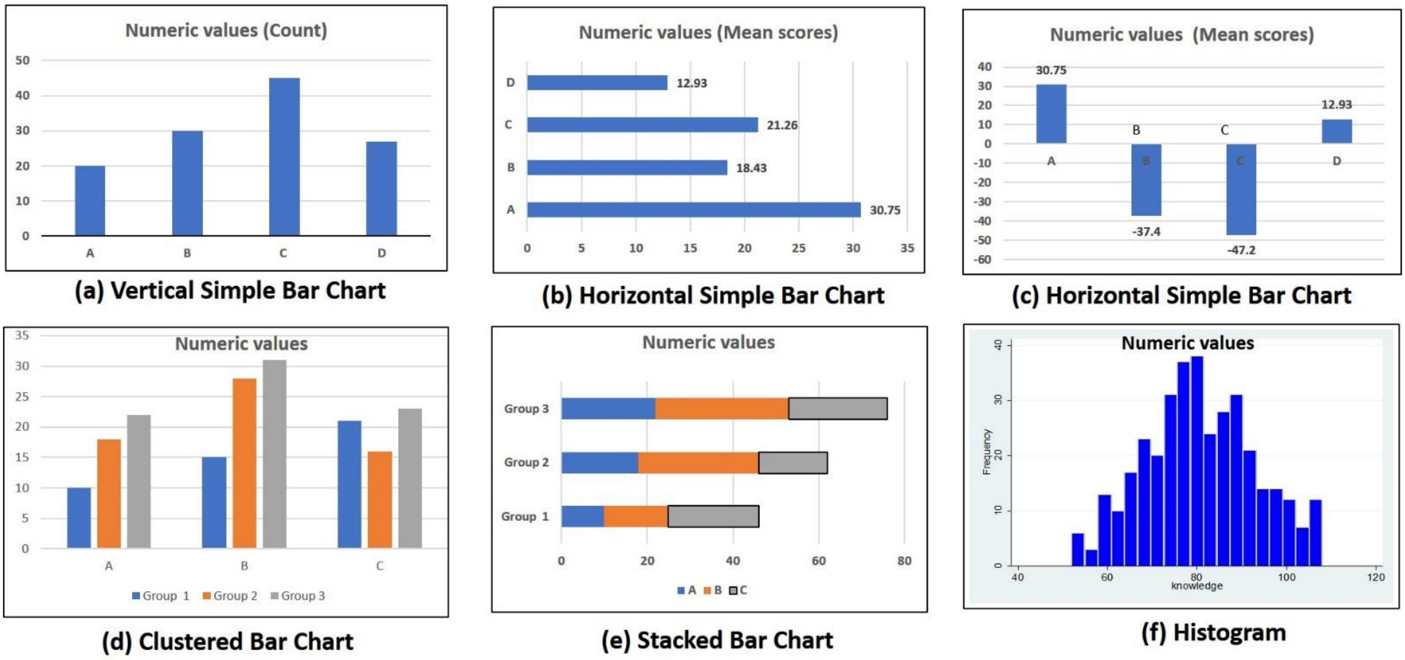


Figure 1. Different types of bar charts

Pie Chart

A pie chart shows the total amount divided into separate categories as radial slides of a circle (pie), in which the size of each slide represents a proportion of a category of the total (100%). The pie chart is effective in presenting the idea of a part-to-whole comparison. The pie chart is sometimes presented in a "donut" form. For example, a pie chart depicts percentages of seven types of food consumption (A–F) among people in a community (Figure 2). There are shortcomings of the pie chart. It is different from the bar chart, because a single pie chart is not good for comparison purposes. It is difficult to comprehensively interpret the pie slices, particularly when there are too many slices or too many similarly-sized slices.[10] When slices are too small, it is difficult to read and color distinctly.[5]

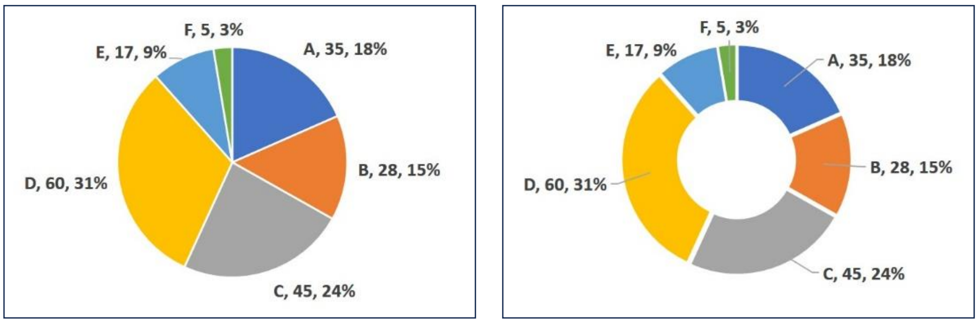


Figure 2. Different types of pie charts

Line Chart

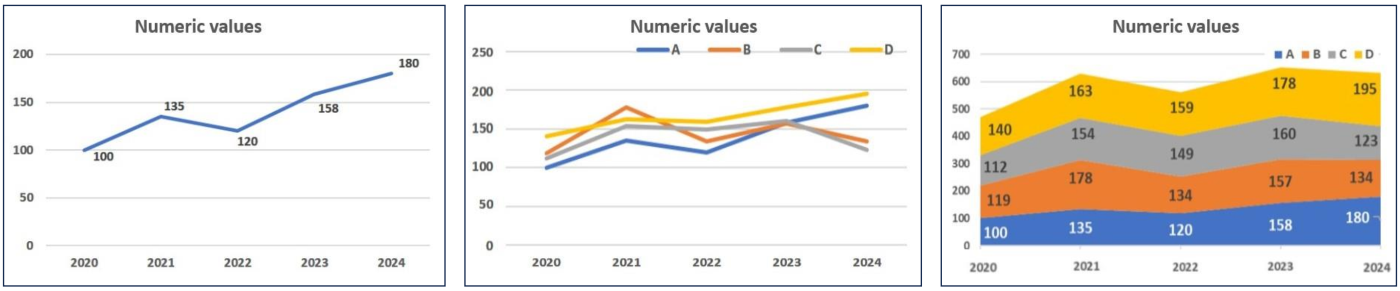
A line chart, also called a line graph, presents a series of data points connected by straight-line segments. The line chart is commonly used to show trends over time and compare several data sets.[6] The line chart is commonly used due to its simplicity which make it universally understandable. For example, a line chart shows the number of emergency department visits at a hospital during 2020–2024 (Figure 3(a)). Besides simple line charts, different types of line charts can be used in various contexts for different analytical purposes including multiple line charts and compound line charts.[11]

Multiple line chart

Multiple line chart involves plotting several lines on the same chart, each representing different data series. For example, a line chart shows the number of emergency department visits at four hospitals during 2020–2024 (Figure 3(b)).

Compound line chart

A compound line chart or stacked line chart shows the cumulative effect of several data sets stacked on top of each other. For example, a line chart shows stacked numbers of emergency department visits at four hospitals during 2020–2024 (Figure 3(c)).



(a) Simple line chart **(b) Multiple line chart** **(c) Compound line chart**

Figure 3. Different types of line charts

Presenting data in graphic format can be misrepresented or misinterpreted when placed in the wrong style of data visualization.[1] Upon choosing the wrong chart type, we may end up not only failing to present our message clearly but also misleading our readers. So, which chart should we pick? There is no clear-cut answer as it depends on several factors including types of data, subject matter, desired format, intended message, intended audience, and personal preferences.[12] Here are some suggestions.[8] When we want to compare data among different categories, the bar chart is a better choice. A part-to-whole can be presented as a pie chart or a stacked bar chart. If we want to convey change over a period of time as a trend, the line chart is the choice.

Chart Attributes

Unique graphical attributes should be applied to both quantitative and qualitative data in terms of shape, color, size, angle and position.[8] Good charts should have clear labels, legends, and annotations. Elements of a visualization can be modified to emphasize or diminish the impact of the data. Some recommendations are as follows.[1,8,10,13–16]

Legends and Annotations

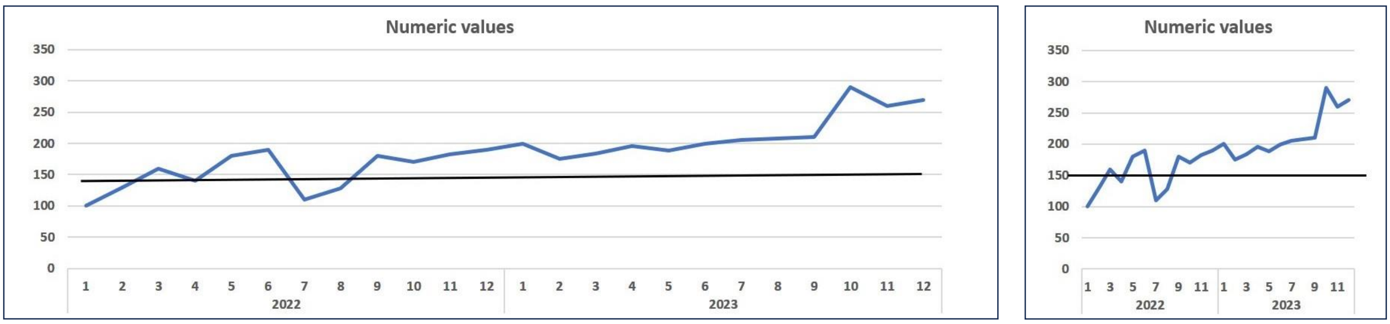
Legends and annotations describe a chart's information. Annotations should highlight data points, data outliers, and any noteworthy content. Text can be used to label different chart elements including chart titles, data labels, axis labels, and legends. Applying "bold text" on too many elements can make it harder to identify important elements. Unclear label wording can induce bias.

Colors

Colors can stimulate subconscious emotional reactions. While colors are used to provide contrast and emphasis, too many colors in a single chart can hinder focus.

Tick Mark Spacing

Tick mark spacing can emphasize granularity. The line can be styled in different ways, such as using colors, dashes or varied opacities. Be careful that increasing the length of the horizontal scale can flatten the perception of the rising line. In the opposite way, squeezing the axis can be misleading because it exaggerates change (Figure 4).



Expand horizontal axis **Compress horizontal axis**

Figure 4. Comparison of trend graph in expand and compress horizontal axis format

Axis Scaling

Axis scaling can be a problem. Most software used to create charts often automatically scale to "best fit" the data values within the dataset. When comparing data values from different datasets in separate charts, be cautious to take notice of whether both datasets have the same scale or not. As shown in the example, different axis scaling could give a wrong impression about the length of bars in bar charts (Figure 5).

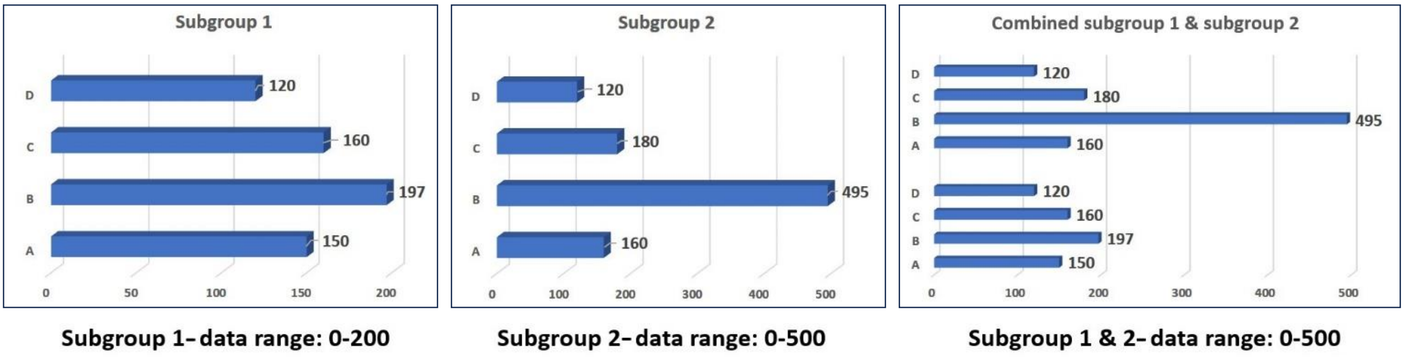


Figure 5. Different axis scaling among subgroups

Blocking Size

The blocking size of a bar chart or histogram depends on the data intervals, so-called binning. Dissimilar bin values will give a different impression about data distribution. As shown in the example, distributions of PCR values of 200 patients look different in different bins (Figure 6).

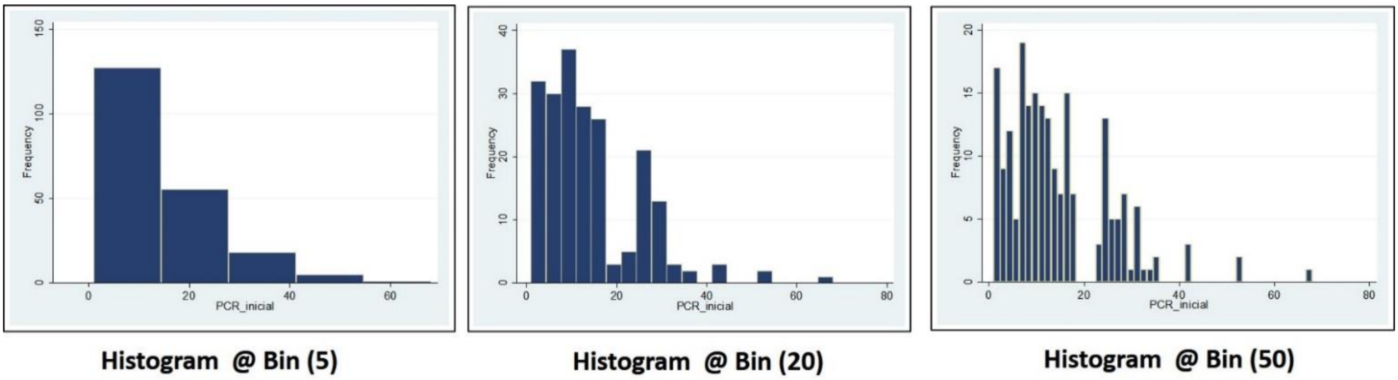


Figure 6. Different blocking size (binning)

Data Overload

Data overload can occur when there is too much information in a chart because it will make the chart look unfocused, fuzzy and difficult to interpret. Crossing too many variables will make the chart difficult to understand, and the readers may make inaccurate assumptions due to bias and confusion. A good chart should actually say more with less. Removing unnecessary formatting and clutter can make the graphics more impactful and easier to interpret.

Misleading and Lie

The main principle of data visualization is to present the information with respect to accuracy, clarity, and integrity in a way that doesn't distort it.[8] It is easy to lie with any kind of charts. Deceptive or misleading charts typically center on visual tricks.[8,10,14,16–18]

Axis Cropping

Axis cropping may be done either on the X- or Y-axis (or both) to show a subset of the data. A truncated axis is a common violation of data visualization design. There must be a good reason to do the cropping. Manipulating the Y-axis by zooming in so that it doesn't start at zero (0) may be done when the baseline value is further away from the zero baseline; but zooming in too much can misrepresent the trend, or exaggerate small differences to appear larger in the reader's eye (Figure 7).

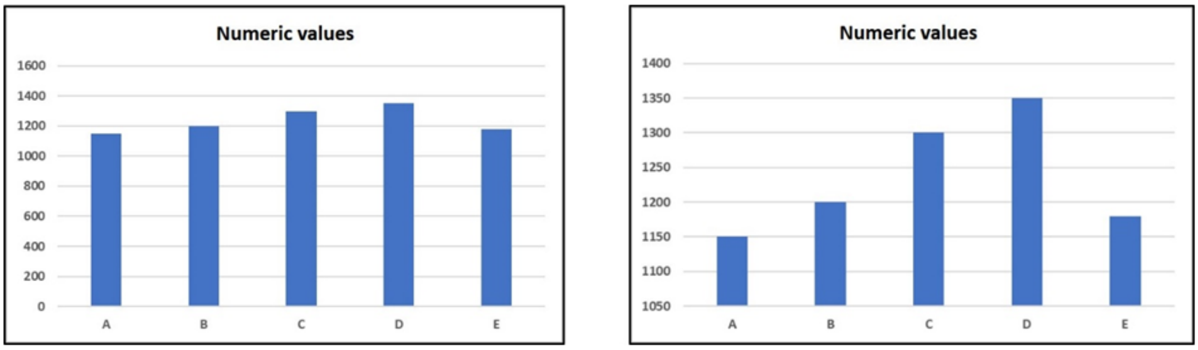


Figure 7. Axis cropping (Truncated Y-axis) showing the differences among the 5 groups

Inverted Vertical Axis

Inverted vertical axis should be rarely used as it can easily introduce the incorrect conclusion. A classic example of a chart showing gun deaths in the inverted vertical axis misleads the readers to think that gun deaths have been declining (Figure 8).

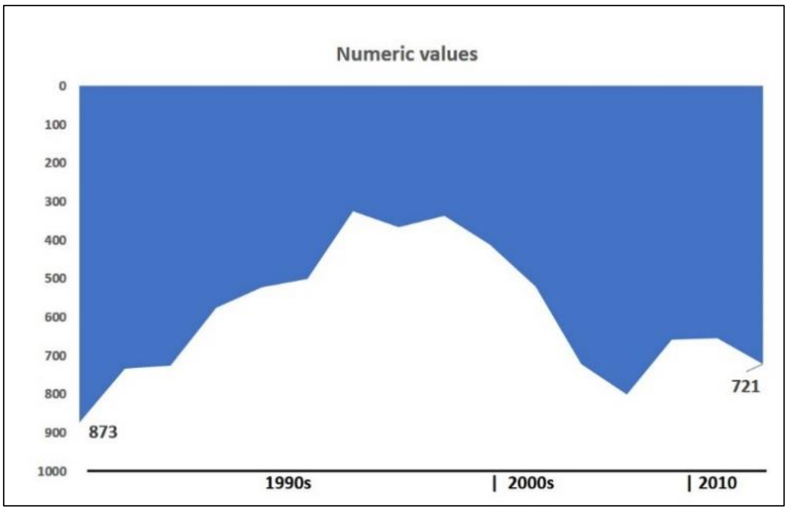


Figure 8. Inverted vertical axis showing numbers of gun deaths over the years

Broken Bar Chart

A broken bar chart is often done by adding a symbol // to denote that the length of the bar is "broken" when it actually extends much longer. However, it is still misleading, so the readers should be informed to interpret the results with caution (Figure 9).

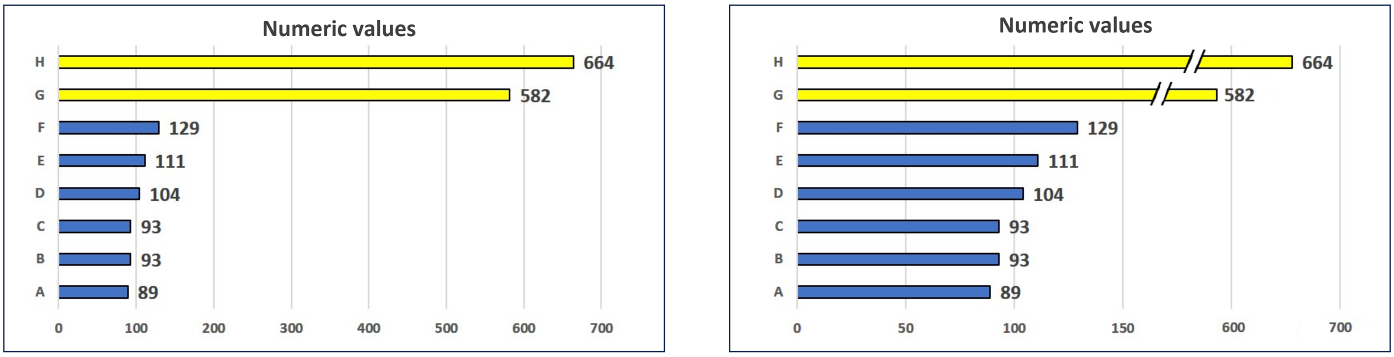


Figure 9. A broken bar chart cutting the length of two long bars

Dual Axis Chart

A dual axis chart is sometimes used to present two data series in one chart. When a vertical axis shows the scale of the first data series on the left side and a second vertical axis on the right side of a chart, it may be confusing, difficult to interpret, and mislead the readers to conclude the trends of the two lines as the causal effect, rather than association (Figure 10).

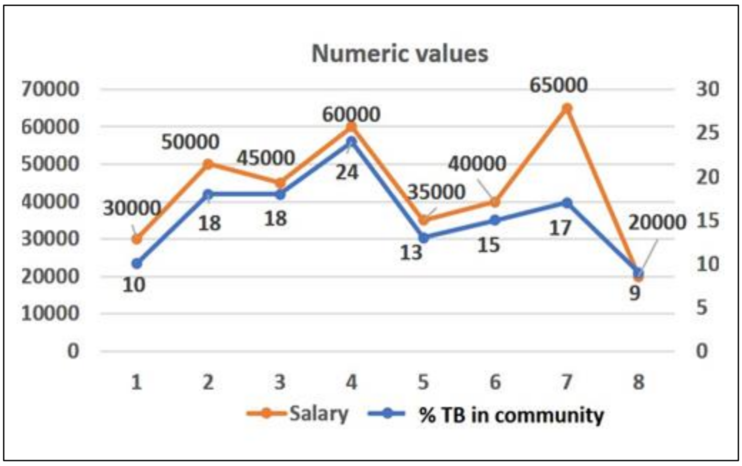


Figure 10. Dual axis chart showing average salary and percentage of TB cases in eight communities

Incorrect Curve-fitting Chart

An incorrect curve-fitting chart can occur when attempting to show the correlation or relationship between two variables with a scatter plot together with its regression line. To correctly explain the trend of the data points, make sure that the data suggest a linear or non-linear trend, and perform the proper "curve-fitting" model accordingly. Do not simply draw an arbitrary line and claim that the data trend corresponds to the drawn line (Figure 11).

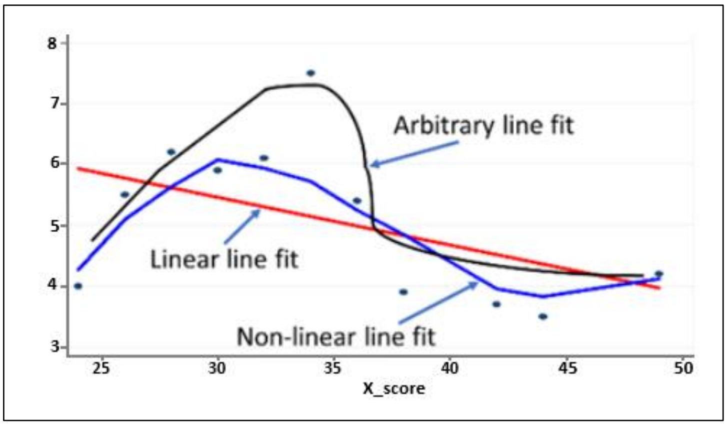
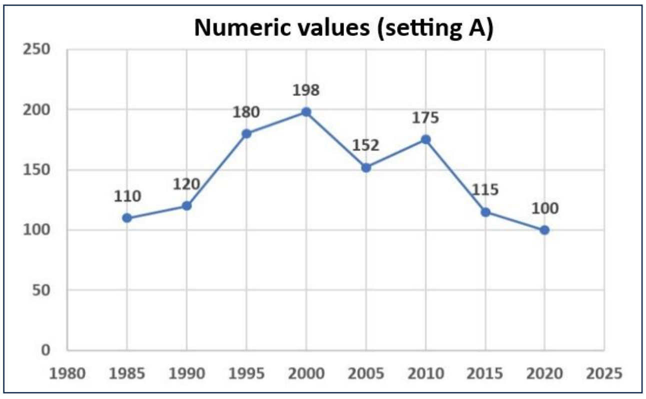


Figure 11. Curve fitting line chart with correct non-linear trend (not the incorrect linear or arbitrary line fit)

Unequal Time Spacing Charts

Unequal time spacing charts are not comparable. When comparing trends over time, take notice that the charts are not based on different time frame or have unequal time intervals (i.e., month over month, quarter over quarter, year over year, etc.). It is not a good idea to collate two charts that are based on unequal time spacing because it can make the readers perceive a false impression about the data trends (Figure 12).



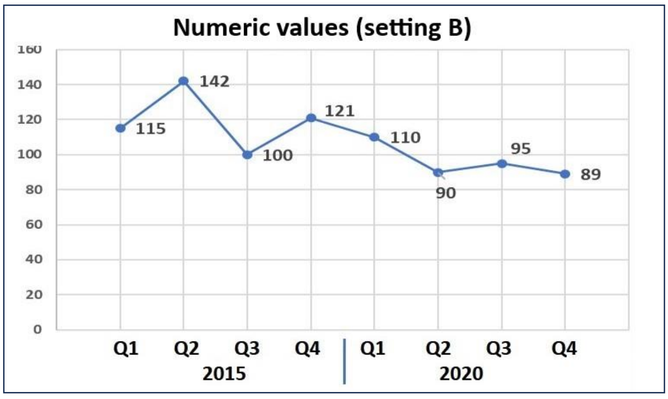
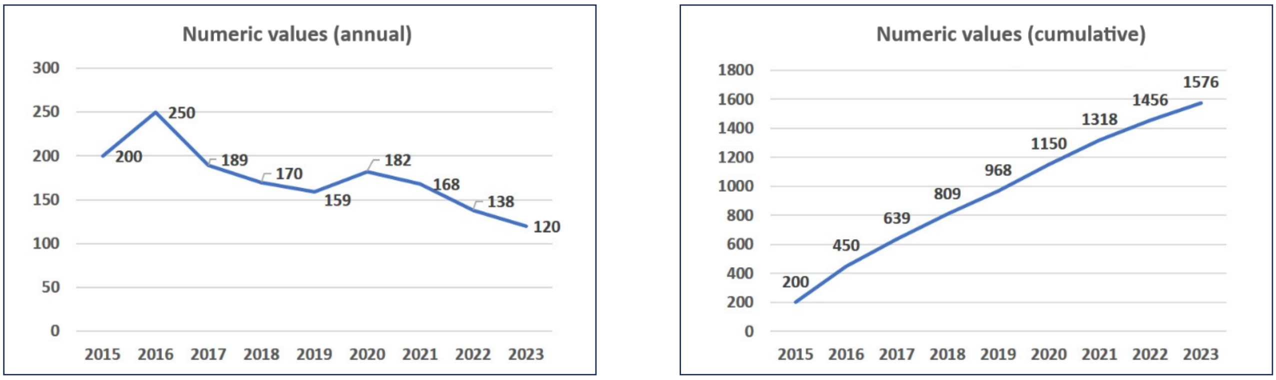


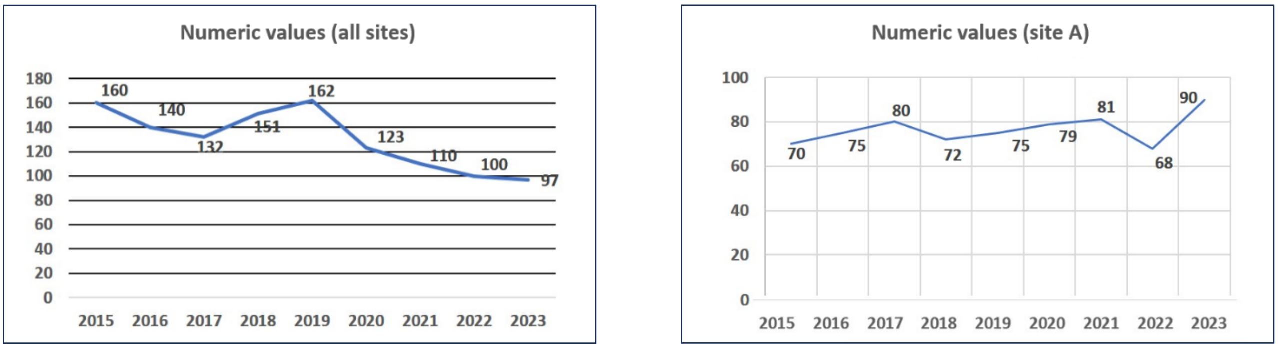
Figure 12. Incomparable charts due to different time-spacing

Reasoning Errors

Reasoning errors can occur when comparing charts with unsupported assertions or logical fallacies. As shown in Figure 13(a), a confusing and incorrect conclusion may occur when comparing two charts showing the opposite direction, the trend of cumulative income (aggregated data) over the years vs. the trend of annual income over the years. The cumulative graph usually makes things seem a lot more positive than they actually are. On the other hand, "cherry-picking" occurs when presenting a chart with incomplete evidence which, in fact, could not be generalizable with more representative evidence. As shown in Figure 13(b), a chart presents an outcome of a subset group but falsely claims it as if it represents the whole group.



(a) Annual data vs. cumulative data



(b) All sites data vs. specific site data

Figure 13. Incomparable charts due to logical fallacy or cherry-picking

Misleading Comparisons

Misleading comparisons can occur when comparing data values across charts, but the data values are based on different attributes or units. It is also important to understand the numerator and denominator of the calculated statistics. As an example, percentage changes over time among different groups can be calculated in different ways. The percentage changes over time can be calculated from the baseline denominator of each group, or from different denominators (number of subjects) at each time point. With different calculations, the interpretation of the charts will be distinctively different (Figure 14).

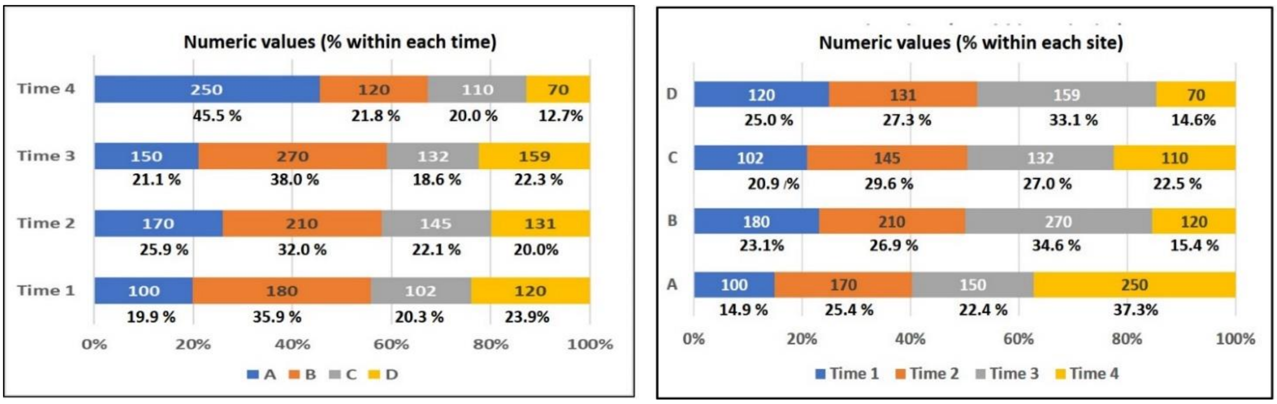


Figure 14. Different conclusions from different percentage calculations

Incomplete Presentation

Incomplete presentation occurs when there are missing categories, unintentionally or intentionally, in a chart. This will make the readers misinterpret the chart. A frequent mistaking is seen in a pie chart that the overall percentage does not add up to 100% due to missing certain categories or rounding up the numbers.

Final Thoughts

Bar, pie and line can lie. In designing a chart, we should take into consideration both scientific and ethical viewpoints. A chart must *tell the truth* and be presented without biases and deception. Presenters, should avoid creating misleading data visualizations, and think carefully about choosing the right kind of chart to communicate the data in a truthful way. Readers, should appraise the data visualization carefully before interpreting and drawing any conclusion.

Suggested Citation

Kaewkungwal J. The grammar of science: bar, pie, line, and lie. OSIR. 2025 Mar;18(1):61–9. doi:10.59096/osir.v18i1.272299.

References

1. Tableau. What is data visualization? definition, examples, and learning resources [Internet]. San Francisco (CA): Salesforce, Inc; [cited 2024 Oct 20]. <https://www.tableau.com/learn/articles/data-visualization>
2. University Libraries. Data visualization: charts and graphs [Internet]. Las Vegas: University of Nevada Las Vegas: [cited 2024 Nov 10]. <https://guides.library.unlv.edu/c.php?g=991477&p=9562156#:~:text=Charts%20are%20the%20best%20option,with%20details%20allowing%20faster%20understanding>
3. Sheeba M. Graphs vs charts [Internet]. Mumbai: WallStreetMojo Team; [cited 2024 Nov 10]. <https://www.wallstreetmojo.com/graphs-vs-charts/>
4. Yi M. A complete guide to bar charts [Internet]. Sydney: Atlassian; [cited 2024 Oct 20]. <https://www.atlassian.com/data/charts/bar-chart-complete-guide>
5. Yi M. How to choose between a bar chart and pie chart [Internet]. Sydney: Atlassian; [cited 2024 Oct 20]. <https://www.atlassian.com/data/charts/how-to-choose-pie-chart-vs-bar-chart>
6. Carbon Design System. Data visualization, Simple charts [Internet]. New York: IBM; [cited 2024 Oct 20]. <https://carbondesignsystem.com/data-visualization/simple-charts/>
7. Pupkin J. Histogram [Internet]. [place unknown]: Wall Street Oasis; [updated 2023 Nov 24; cited 2024 Nov 10]. <https://www.wallstreetoasis.com/resources/excel/study/histogram>
8. Material Design. Data visualization [Internet]. Mountain View (CA): Google; [cited 2024 Oct 20]. <https://m2.material.io/design/communication/data-visualization.html>
9. Moore DS, McCabe GP, Craig BA. Introduction to the Practice of Statistics. 10th ed. New York: W. H. Freeman; 2021.
10. Fortney K. 5 ways data visualizations can lie [Internet]. San Francisco: Towards Data Science; 2017 May 27 [cited 2024 Oct 20]. <https://towardsdatascience.com/5-ways-data-visualizations-can-lie-46e54f41de37>
11. Thompson C. Line chart: definition, types, examples. New York: Investopedia; [updated 2024 Aug 16; cited 2024 Oct 20]. <https://www.investopedia.com/terms/l/linechart.asp>
12. Math Worksheets Center. Bar graphs, pie charts, and line graphs: how are they similar and how are they different? [Internet]. [place unknown]: Math Worksheets Center; [cited 2024 Oct 20]. <https://www.mathworksheetscenter.com/mathtips/barlinepiegraph.html#google_vignette>
13. Tracy S. How to lie with charts [Internet]. San Francisco: Medium; 2016 Sep 1 [cited 2024 Oct 20]. <https://stephen-tracy.medium.com/how-to-lie-with-charts-97396e4642>
14. PolicyViz. 10 ways to mislead with data visualization. [place unknown]: PolicyViz; 2023 Feb 7 [cited 2024 Oct 20]. <https://policyviz.com/2023/02/07/10-ways-to-mislead-with-data-visualization/>
15. Pop R. Lies, damned lies, and visualizations [Internet]. Bergen: Everviz; 2021 Nov 8 [cited 2024 Oct 20]. <https://www.everviz.com/blog/lies-damned-lies-and-visualizations/>
16. Dougherty J, Ilyankou I. Hands-on data visualization: interactive storytelling from spreadsheets to code [Internet]. [place unknown]: Jack Dougherty and Ilya Ilyankou; 2021.

Chapter 14.1, How to lie with charts; [cited 2024 Oct 20]; [about 12 p.]. <https://handsondataviz.org/how-to-lie-with-charts.html>

17. Parikh R. How to lie with data visualization [Internet]. San Francisco: Heap by Contentsquare; 2014 Apr 14 [cited 2024 Oct 20]. <https://www.heap.io/blog/how-to-lie-with-data-visualization>

18. Lisnic M, Polychronis C, Lex A, Kogan M. Misleading beyond visual tricks: how people actually lie with charts. Proceedings of the CHI'23: CHI Conference on Human Factors in Computing Systems; 2023 April 23–28; Hamburg, Germany. New York: Association for Computing Machinery; 2023. 21 p. doi:10.1145/3544548.3580910.