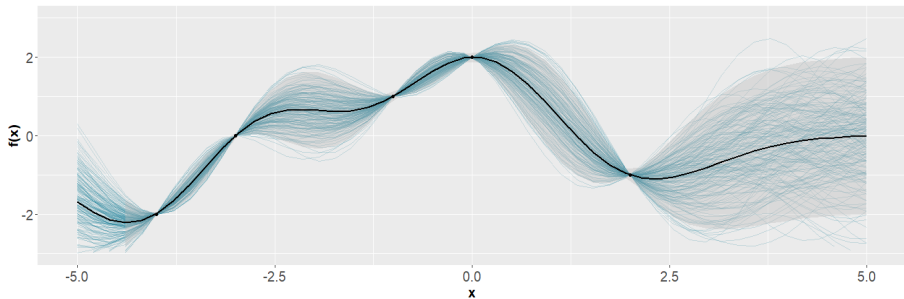


Geovisualization

Data visualization (principles)

Joaquin Cavieres

Geoinformatics, Bayreuth University



Outline

- 1 Opening thoughts
- 2 Best practices for data visualization
- 3 Misleading visualization techniques
- 4 Visualization types

Opening thoughts

Graphs are not simply just to inform you, but they also are to make you reflect.

Graphs are not simply just to inform you, but they also are to make you reflect.

- Can you explain what is a graph?
- How can you use graphs to tell people the truth in a way that they understand?
- How can you show the "uncertainty" in a graph?

On the other hand, some questions for you..

- What is your first impression when you see a graph?

On the other hand, some questions for you..

- What is your first impression when you see a graph?
- Do you think in a story when you see a graph?

Commonly a graph should represent:

- A comprehensive structure
- An easy explanation of the exhibitor
- A complete understanding of the data by the user

Considering the previous, and the exposed by Tufte, E (2001, [2]), the main principles would be:

- Graphical excellence is the well-designed presentation of interesting data
- Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.
- Graphical excellence is nearly always multivariate.
- Graphical excellence requires telling the truth.

Thus, an effective data visualization should be

- Clear
- Concise
- Visually appealing

For the above, we have to do:

- Choose the right graph type for your data
- Use appropriate labels and titles to provide context and clarify data
- Minimize clutter and distractions in the visualization
- Ensure consistency in design elements such as font size, color scheme, and layout.

Best practices for data visualization

Here are a few best practices to keep in mind when creating visualizations.

1. Have a goal in mind

Creating data visualizations is to understand the type of data you're working with and have an idea of the type of insights you would like to derive from them. It is important to keep in mind the information you want to communicate for the audience you are presenting.

2. Choose the right tool

After fix your objective for your visualization, consider the right way to help you present your data. In this case, there are code libraries and no-code/low-code platforms which could help you.

No-code/low-code platforms:

Tableau

It is a visual analytics platform that enables users to create interactive charts, maps and dashboards.

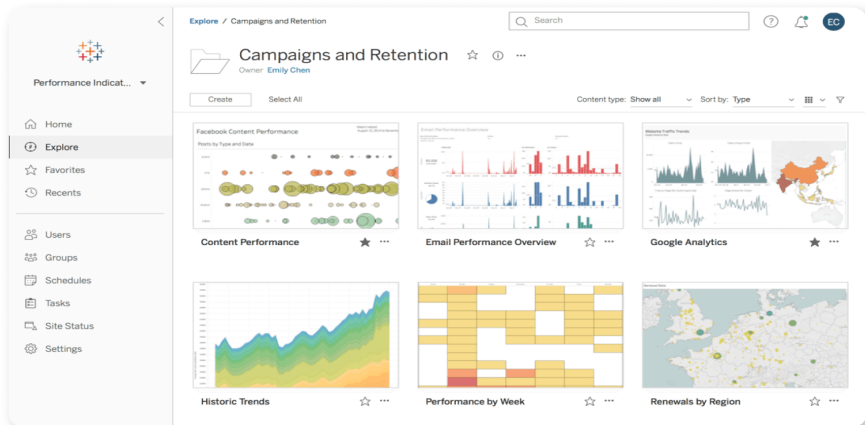


Figure: Source: <https://www.tableau.com/products/server>

No-code/low-code platforms:

Power BI

It is a popular business intelligence platform that allows professionals to create interactive dashboards, charts and graphs, offering a wide range of options and interactive elements while being user-friendly and intuitive.

Open sources (libraries) softwares

Python

Several data visualization libraries are in Python, and you can select libraries depending on your project's needs. For example,

- Matplotlib
- Seaborn
- Bokeh
- Plotly
- Streamlit

Open sources (libraries) softwares

JavaScript

It also is a popular programming language used to build web applications and interactive elements on websites and create data visualizations. A few popular libraries for creating visualizations are as follows:

- D3.js
- Chart.js
- Plotly.js
- Recharts
- Nivo

Open sources (libraries) softwares

R

In R you have access to several libraries and packages for creating a wide range of visualizations, including simple plots and interactive graphics.

- ggplot2
- plotly
- Highcharter
- Leaflet
- ggvis

3. Choose the right type of visualization

When using these visualizations, ensure that you make wise comparisons and use charts well-suited for the data type. The right visualization depends on your goals, data type and audience.

4. Use short and explicate labels and titles

It's important to use labels and titles for your visualizations and ensure they're easy to understand. These labels should provide context and inform the reader what the graph is trying to communicate.

5. Using the right colors

Colors play an important role in data visualizations as they help keep your audience engaged with the content of your dashboard, or report and draw the audience's attention to important information you're trying to point out.

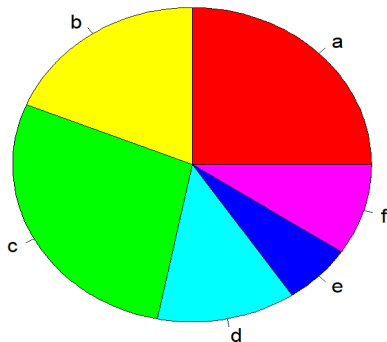
6. Avoid unnecessary visual elements

A good data visualization is easy to understand, engaging and uncluttered. Having too many things going on at once in your visualization can distract your audience from the insights they're supposed to take away from your work.

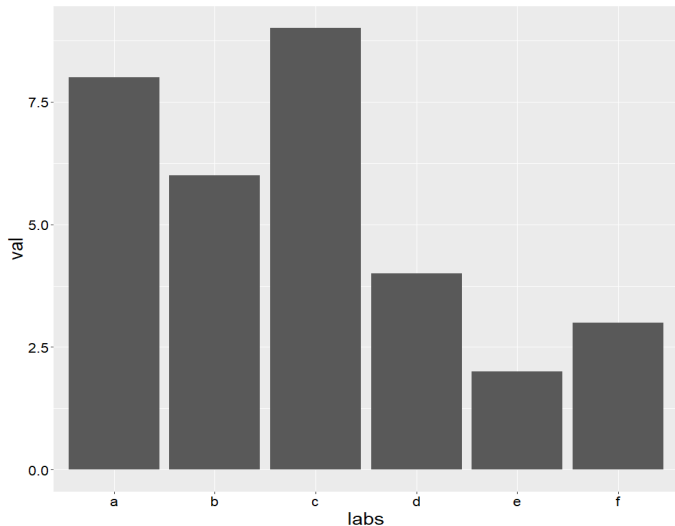
Misleading visualization techniques

A common question for beginners in R and visualization is “which library should I use?” Like most things there is no right answer. Every situation is different. For the same, there are mistakes made very frequently by visualization practitioners.

Pie charts



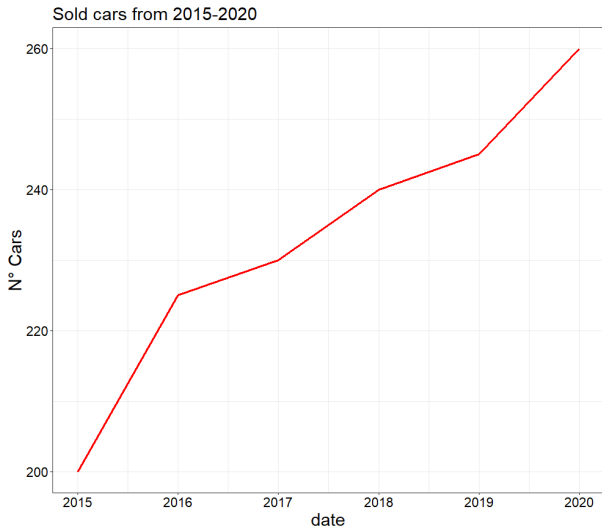
What do you think about the comparison between "f" with "d"?



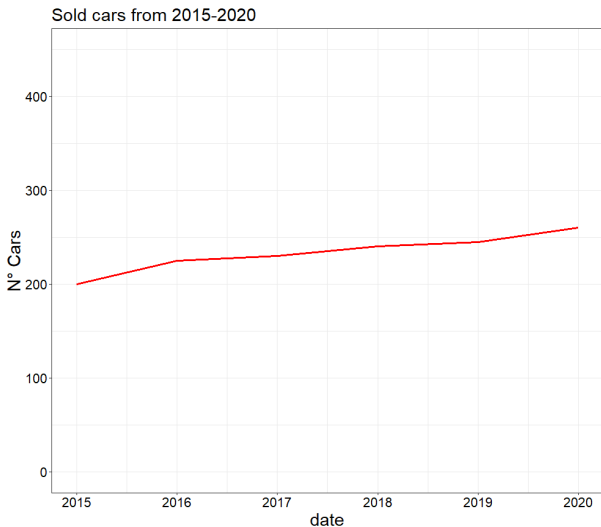
Using a bar plot we can clearly see that "f" is smaller than "d".

Bar plots are fantastic tools. It seems that more often than not they are the best visualization for the job, often out-competing more complicated flashy visualizations in terms of ease of reading/comprehension.

Truncated axis



It looks like the number of sold cars is increasing notably!

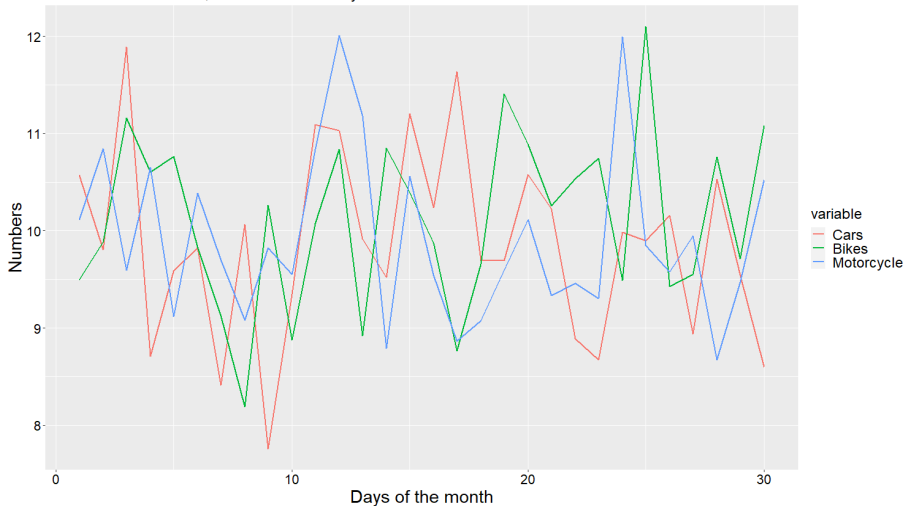


We can be calm regarding climate change! (:

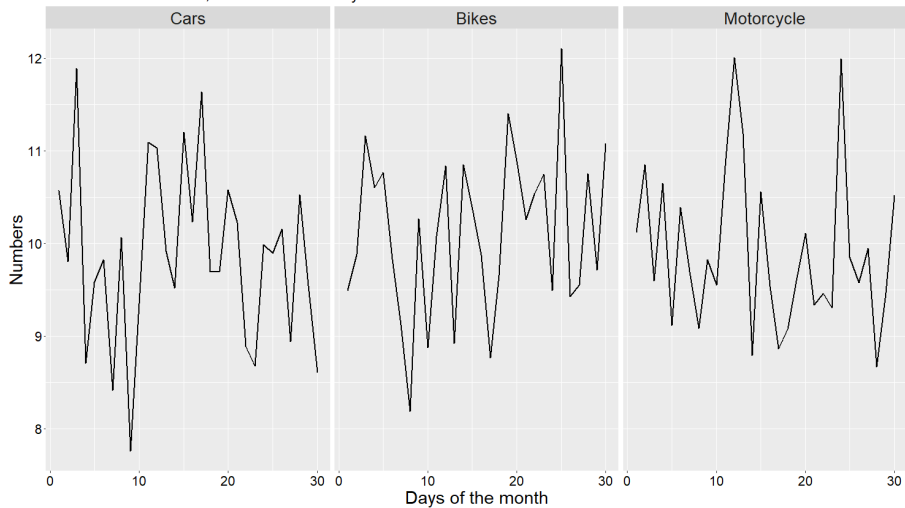
Information Overload

For example, if we have different time series data, then we could be interested in to compare temporal trends in some way.

Time series for cars, bikes and motorcycle



Time series for cars, bikes and motorcycle



You can see a more detailed explanation on this in the following page of [Nashville Biostatistics](#)!:

`http://nickstrayer.me/nashvilleBioStats//2016/05/visualization_in_r.html`

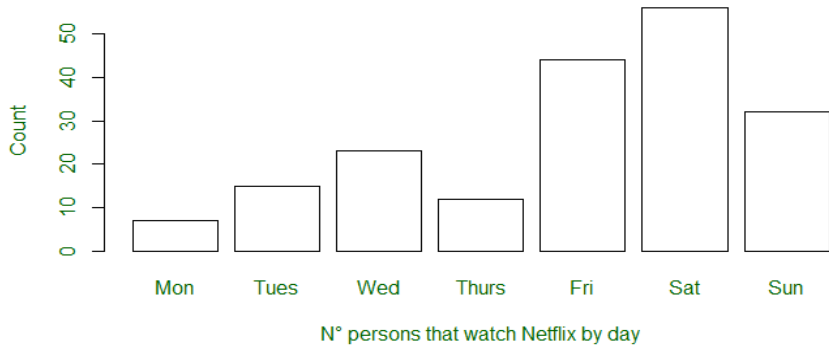
Visualization types

There are basic types of graphs than can represent your data. This makes data visualization essential for research, business, data exploration, etc.

Bar plot

A bar graph should be used to avoid clutter when one data label is long or if you have more than 10 items to compare.

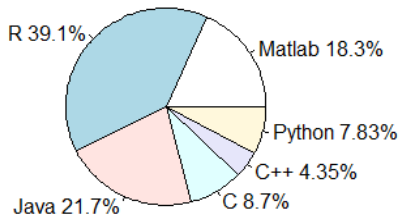
Bar plot is used to represent the values in data vector as height of the bars. The data vector passed to the function is represented over y-axis of the graph.



Pie Chart

Pie chart is a circular chart divided into different segments according to the ratio of data provided.

The total value of the pie is 100 and the segments tell the fraction of the whole pie.



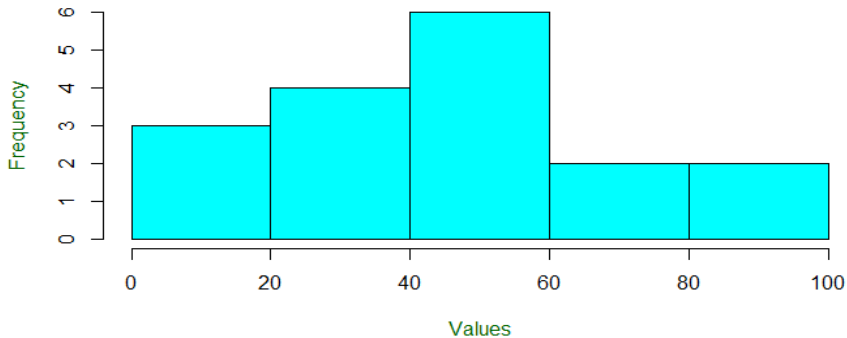
This graph is only as an example!

Histogram

Histogram is a graphical representation used to create a graph with bars representing the frequency of grouped data in vector.

Histogram is same as bar chart but only difference between them is histogram represents frequency of grouped data rather than data itself.

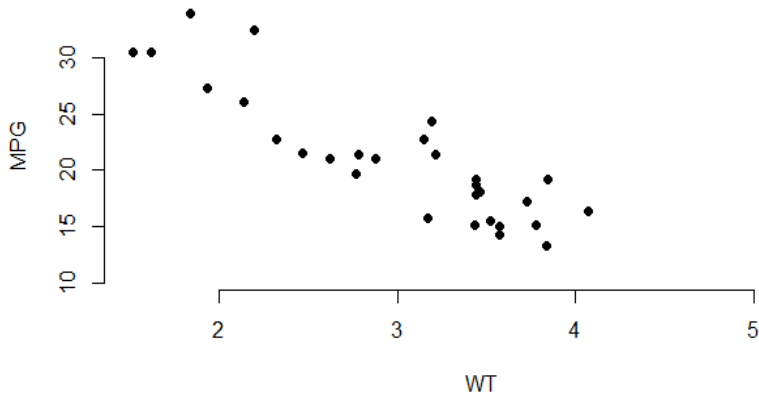
Histogram of Vector x



Scatter plot

A scatter plot is another type of graphical representation used to plot the points to show relationship between two data vectors.

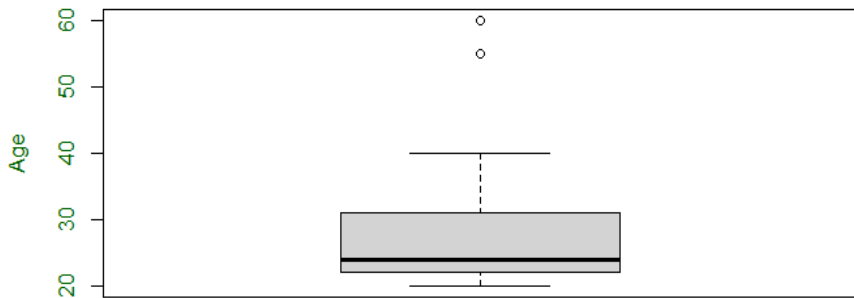
Scatter plot WT vs MPG



Box plot

Box plot shows how the data is distributed in the data vector.

It represents five values in the graph i.e., minimum, first quartile, second quartile(median), third quartile, the maximum value of the data vector.



Box Plot

See you next class!...



Lamigueiro, O. P. (2014). Displaying time series, spatial, and space-time data with R. CRC Press.



Tufte, Edward R. (2001). The Visual Display of Quantitative Information. Second. Cheshire, CT: Graphics Press.