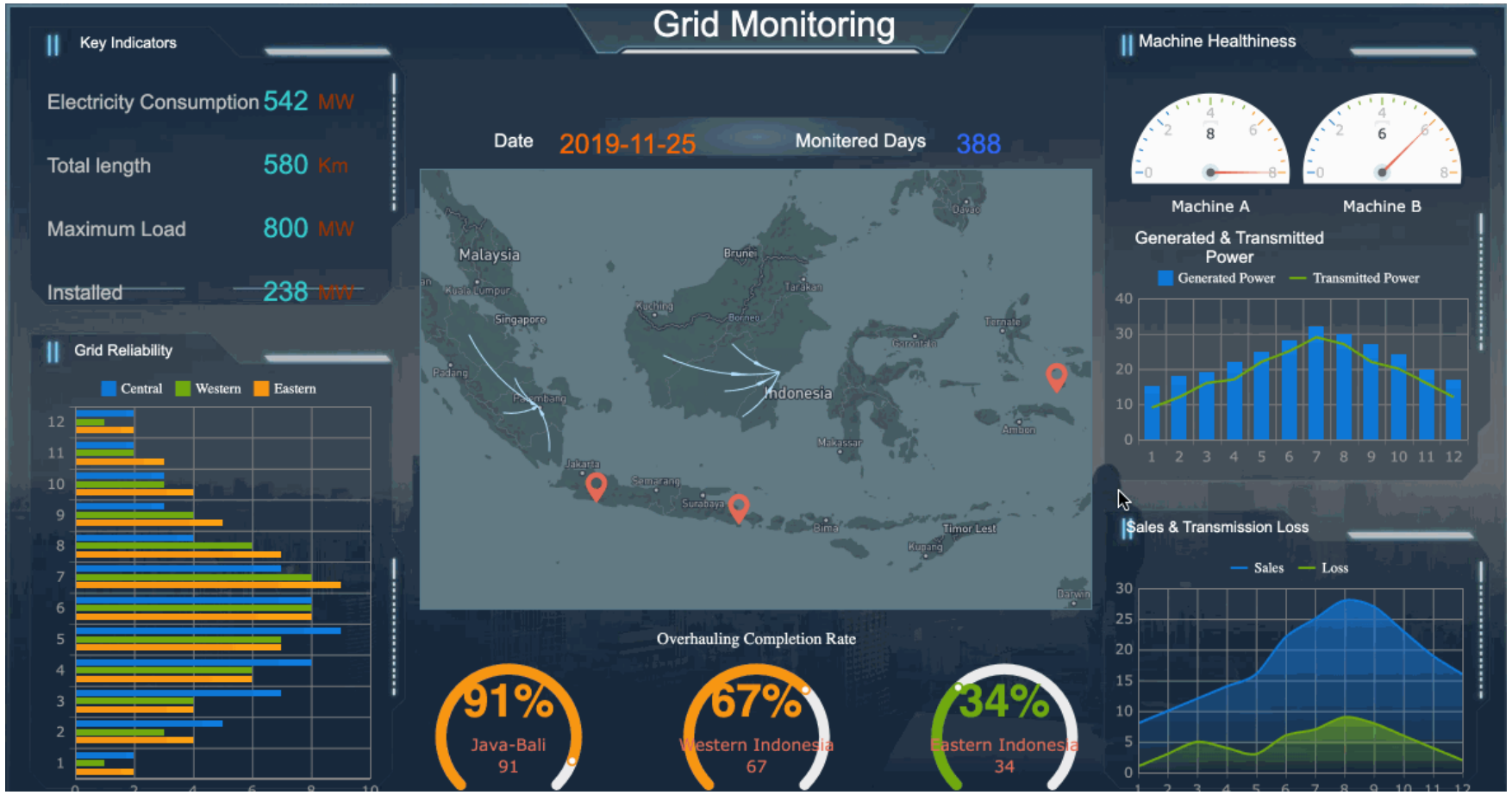
Data visualization

Kamarul Imran Musa
drkamarul@usm.my

Objectives of lecture

Understand	Understand why we need data visualization
Understand	Understand principles of data visualization
Know	Know types of data visualization/figures
Get	Get skills to perform data visualization

What are data visualization



https://towardsdatascience.com/top-7-trends-in-data-visualization-software-for-2019-c4b13e783844

Data visualization is

- … the graphical representation of information and data.
- This is done using visual elements like charts, graphs, and maps
- Data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- Objective: data visualization tools and technologies are used to analyze massive amounts of information and make data-driven decisions.

https://www.tableau.com/learn/articles/data-visualization

Infographics and visualization

Infographics = presentation

Visualization = exploration

Why visualize?

Graphics

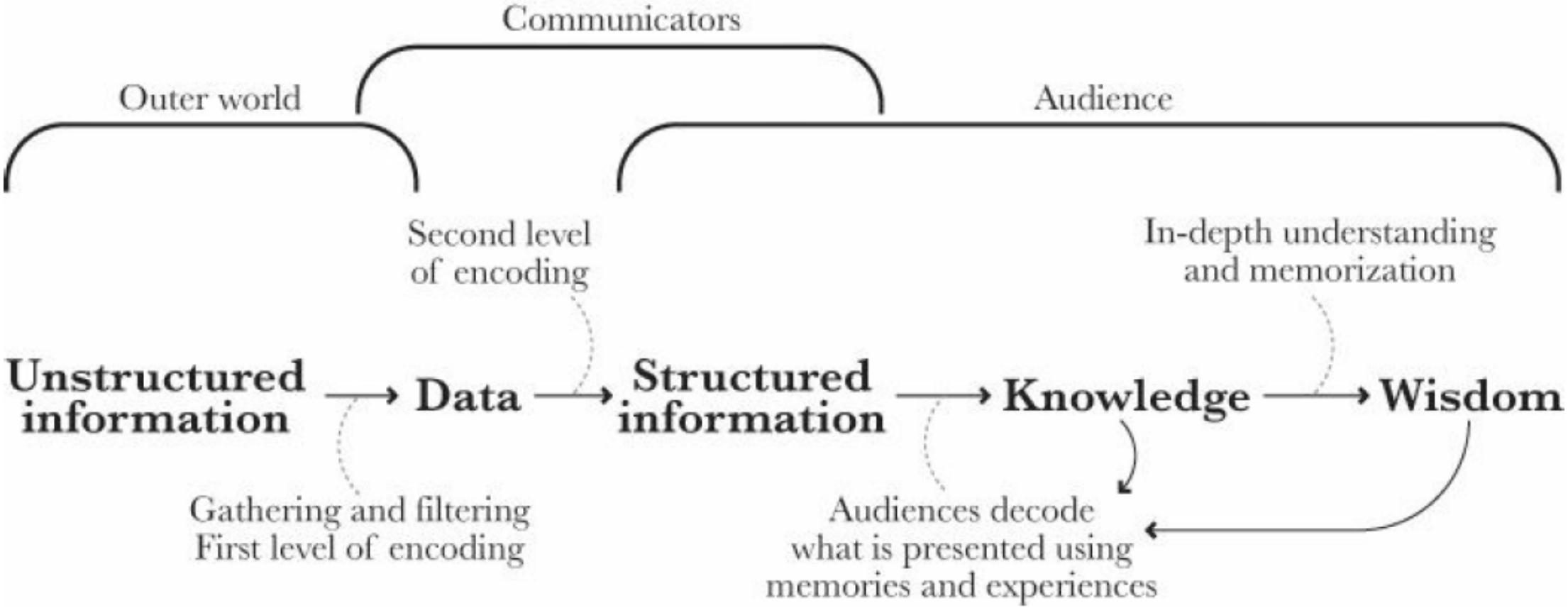
- Simple
- Clear
- Has sufficient information

To tell stories with data

- See what data have to say

How to visualize?

Source: Alberto Cairo. The Functional Art



How to visualize?

Unstructured information = reality

Data = records of observations

Knowledge = perception or meaning of data

Wisdom = deep understanding

More than numbers

Data can be boring

Statistics and data help you to look beyond data

Statistics is not pure analysis

What to acquire, do or have

Verified data

Reliable data

Clear, concise and nice graphics

Good

Points
Areas
Symbols
Colours
Annotation

Generate emotion

Visualization, Figures, Plots, Graphs

- Data visualization :
  - graphical representation of information and data
- Figures :
  - A figures is a visual presentation of results.
  - Come in the form of graphs, charts, drawings, photos, or maps.
  - They provide visual impact and can effectively communicate your primary finding.
- Plot:
  - A plot is a graphical technique for representing a data set
  - Usually as a graph showing the relationship between two or more variables.
  - The plot can be drawn by hand or by a computer.
- Graphs:
  - A graph is a picture designed to express words
  - Particularly the connection between two or more quantities.

Visualization and animation

- https://en.wikipedia.org/wiki/Hans_Rosling

- https://www.youtube.com/watch?v=jbkSRLYSojo

Minimalism and efficiency

- Good visual design if
  - Well-designed presentation
  - Complex idea but presented with clarity, precision and efficiency
  - Gives viewers greatest number of ideas in the shortest time with the least ink

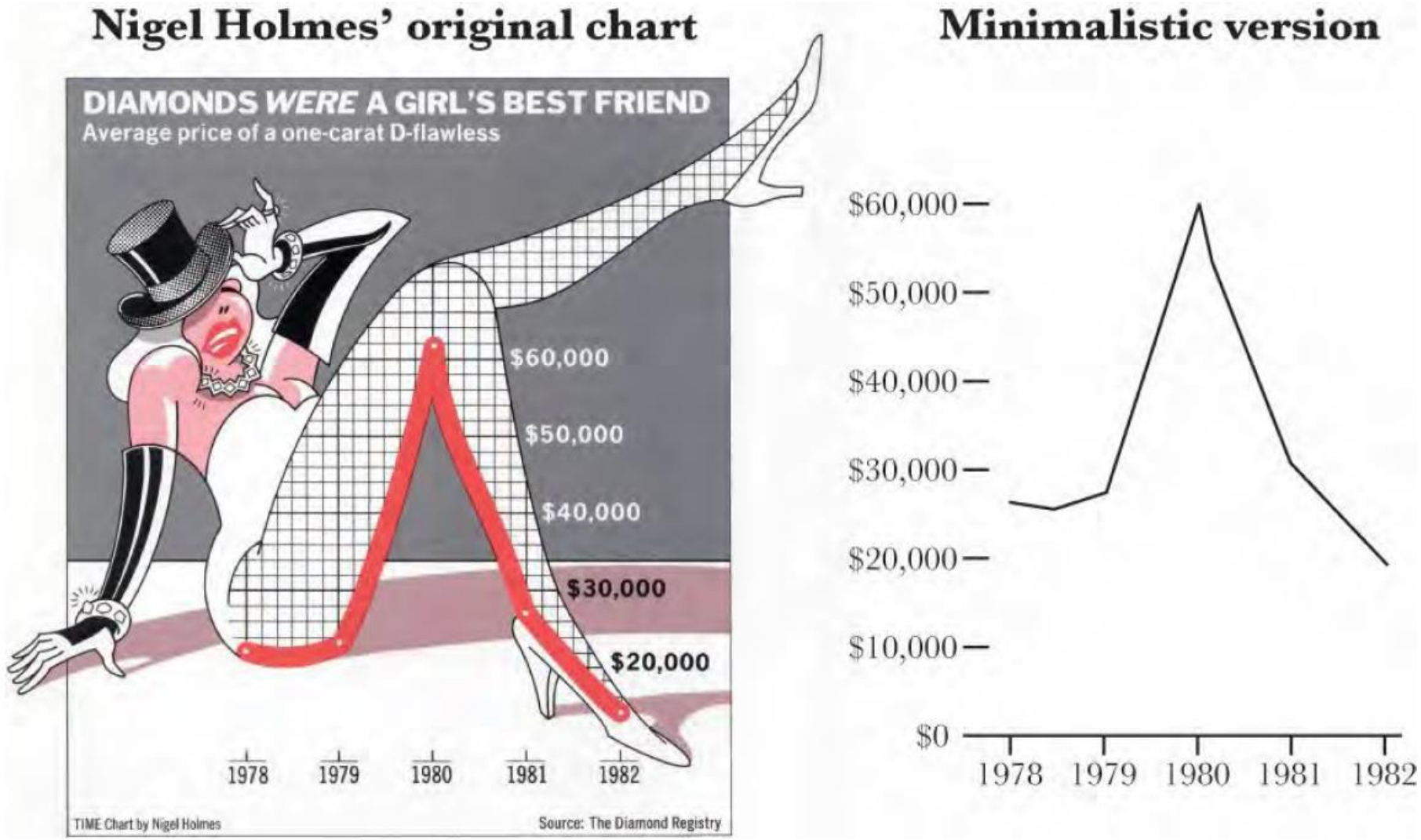


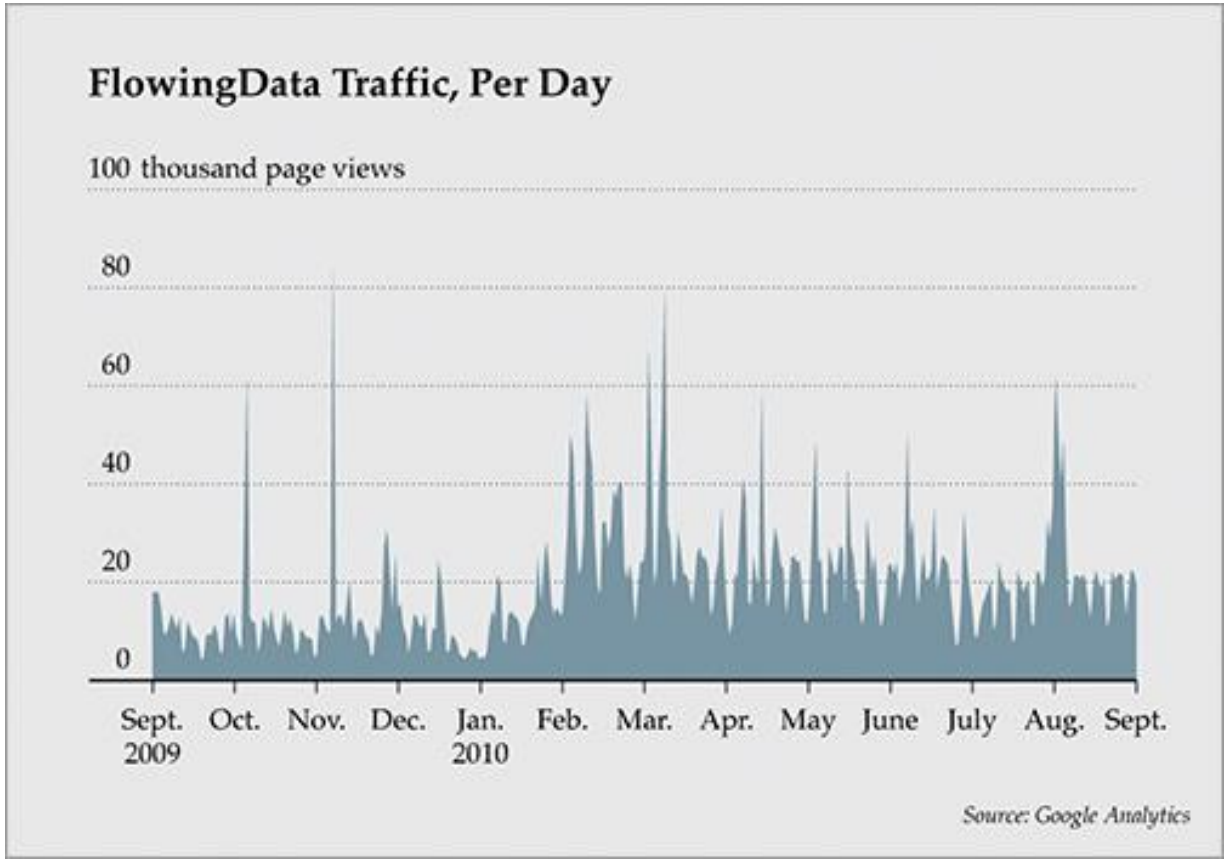
Figure 3.15. The original graphic is not very appealing, but the stripped-down version was not easily remembered.

What should I look for?

- Pattern
- Relationship

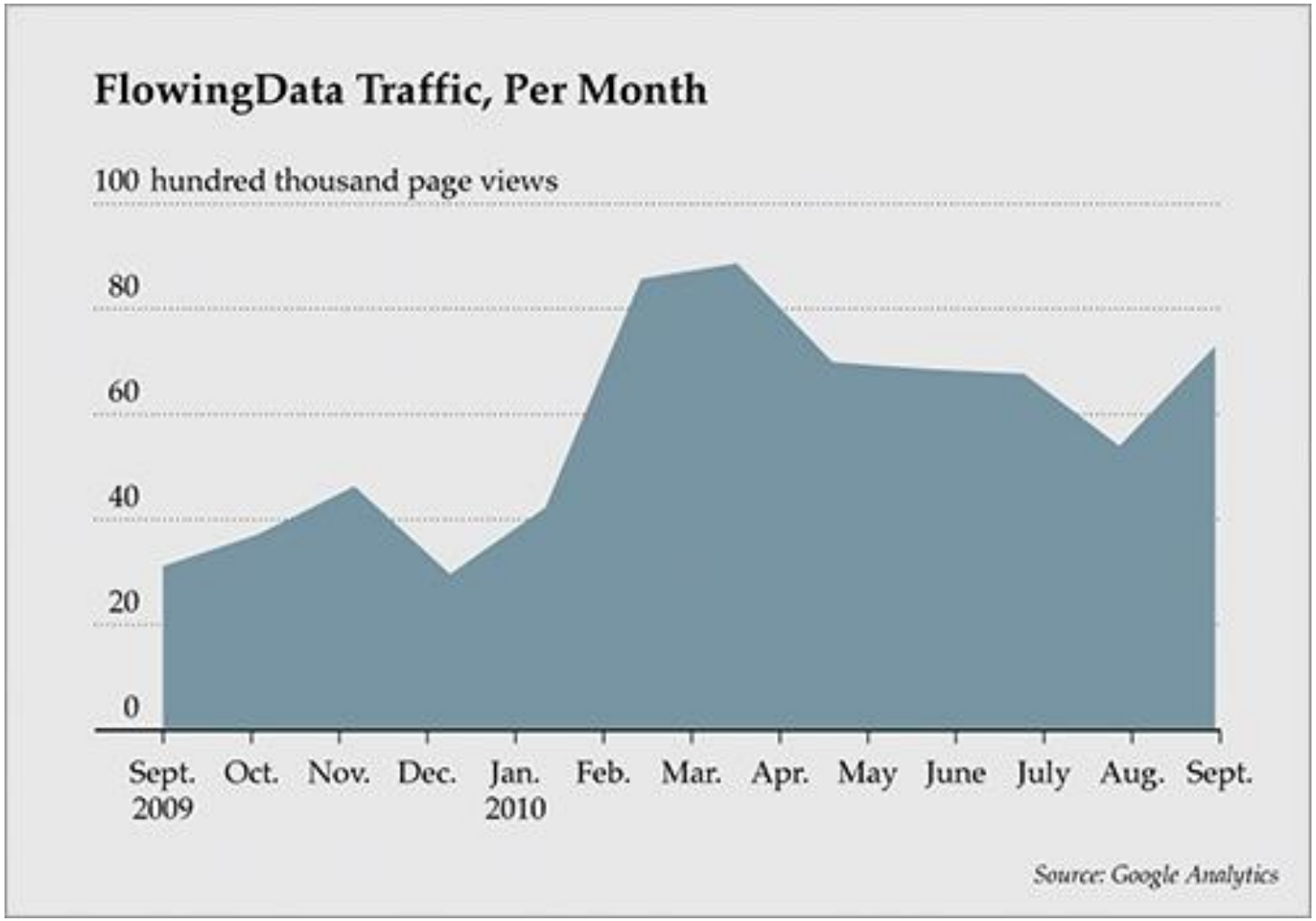
A) Pattern

- Plot based on daily data
- As time goes by
  - Bumpier graph
  - More fluctuations



Source: Visualize This, Figure 1-6

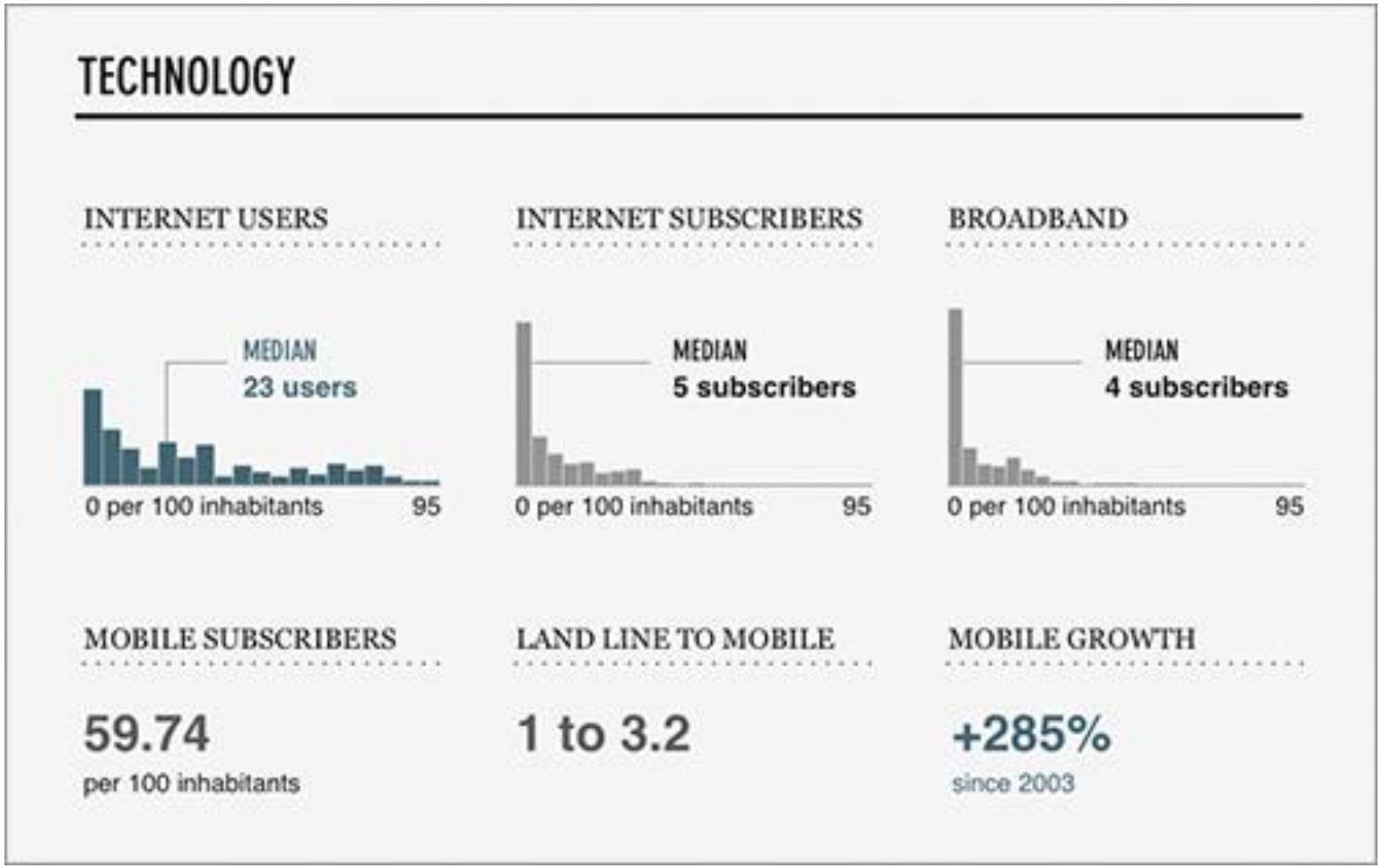
- Same data, different plot
  - Based on monthly basis
- You will get
  - Smoother plot



B) Relationship

- But unlike statistics – relationship not for answering hypothesis

- Compare distributions between groups



Source: visualize This Fig 1-8

What a graphic should contain

- A graphic
  - Must present several variables
  - Should allow comparisons
  - Should help me organize the displayed data
  - Should make correlations or relationships evident

Examples

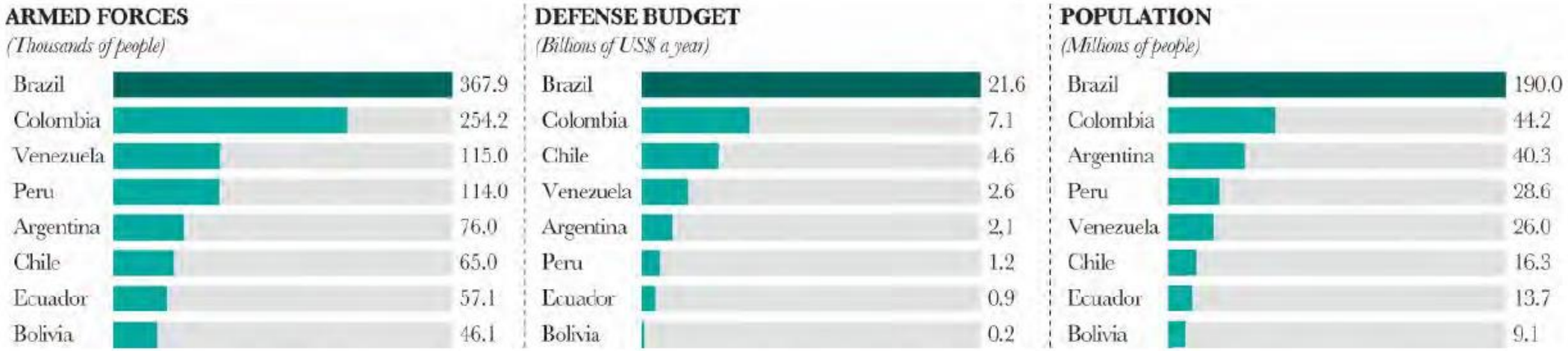


Figure 2.3. Let's get rid of the pictograms and use a traditional bar chart.

- Easy to compare – sorted
- Source: Alberta Cairo. The Function Art

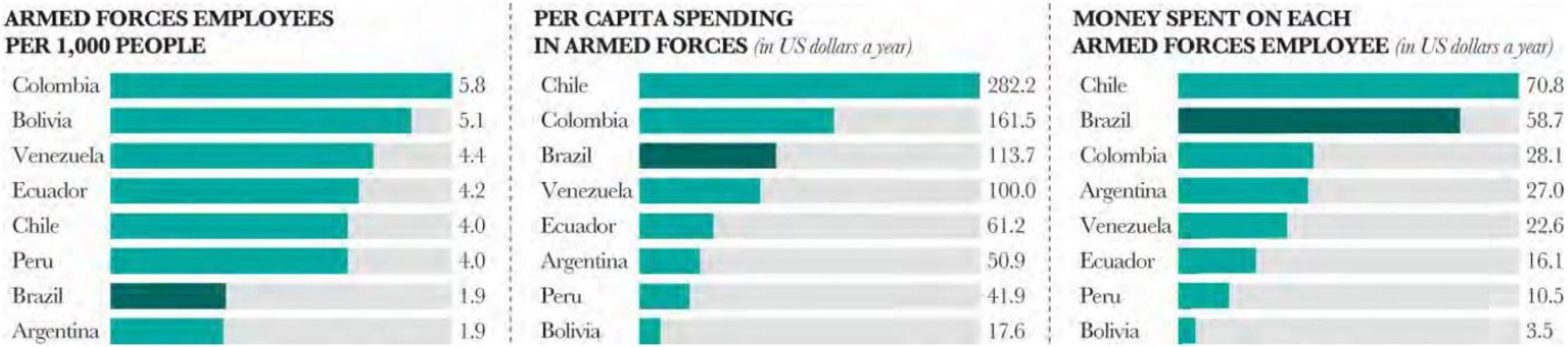
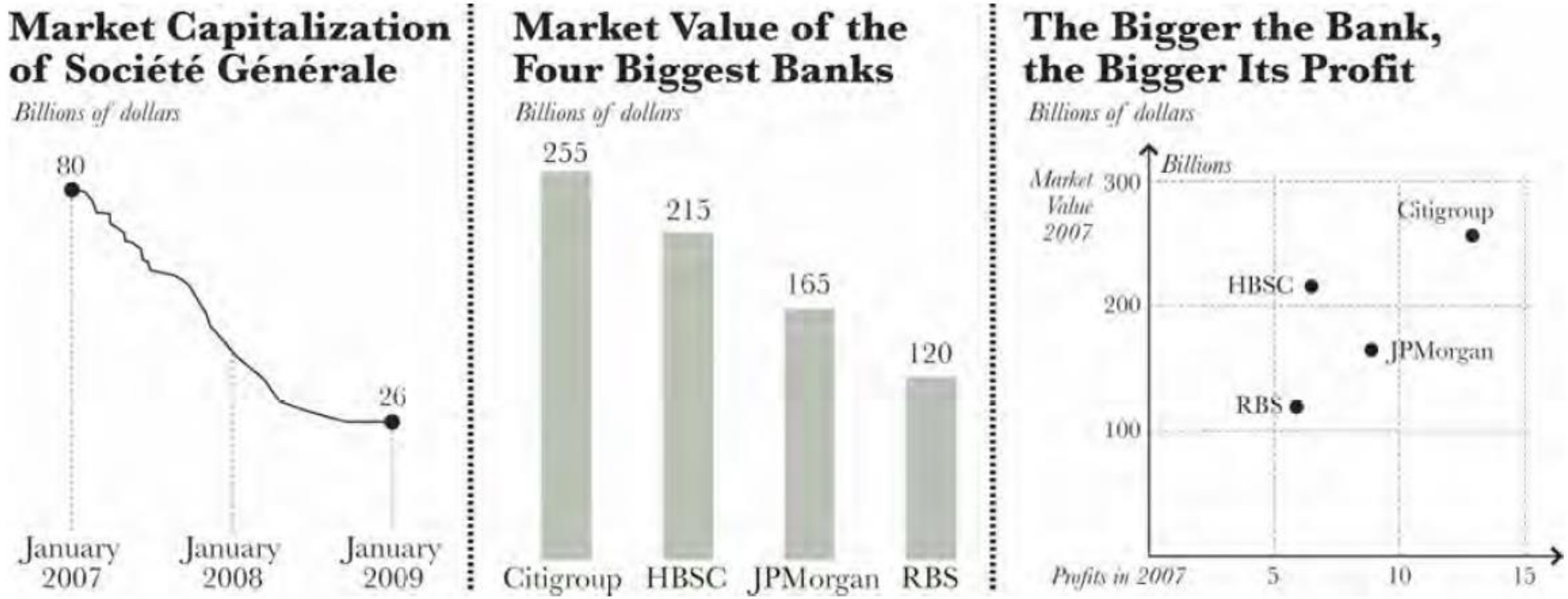


Figure 2.4. The country rankings change quite a bit when you encode secondary variables instead of absolute numbers.

- Absolute vs relative terms
  - Venezuela and Bolivia are more militarized
  - Columbia is far ahead because war on narco-guerillas

Change through time, comparison and correlation



- Time series chart, bar chart and scatter-plot chart

Data visualization, figures and plots

- Data visualization is the graphical representation of information and data.
- Represented by visual elements
- Visual elements
  - charts, graphs, and maps
- Objectives of data visualization tools
  - See and understand trends, outliers, and patterns in data.

https://www.tableau.com/learn/articles/data-visualization

Figures

- the symbol for a number or an amount expressed in numbers

- https://dictionary.cambridge.org/dictionary/english/figure

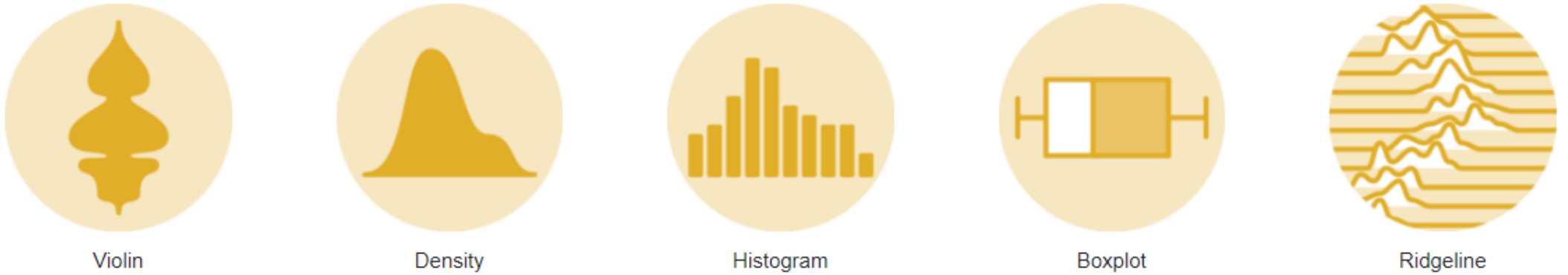
Plots

- A plot is a graphical technique for representing a data set
- For example, as a graph showing the relationship between two or more variables.
- It can be drawn by hand or by a mechanical or electronic plotter.
- Graphs are a visual representation of the relationship between variables
- This is useful for humans who can then quickly derive an understanding which may not have come from lists of values.

- https://en.wikipedia.org/wiki/Plot_(graphics)

Types of graphs

Distribution



https://www.r-graph-gallery.com/

Correlation

Scatter

Heatmap

Correlogram

Bubble

Connected scatter

Density 2d

Ranking

Barplot

Spider / Radar

Wordcloud

Parallel

Lollipop

Circular Barplot

https://www.r-graph-gallery.com/

Grouped and Stacked barplot

Treemap

Doughnut

Pie chart

Dendrogram

Circular packing

Evolution

Line plot

Area

Stacked area

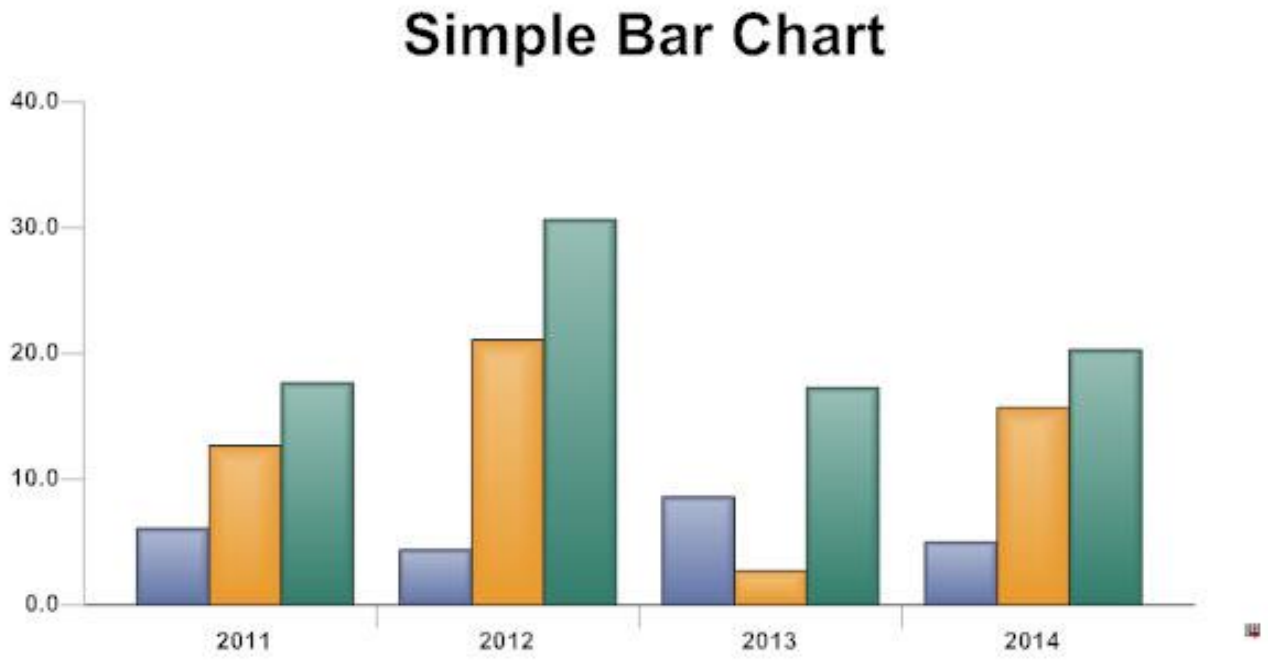
Streamchart

Time Series

https://www.r-graph-gallery.com/

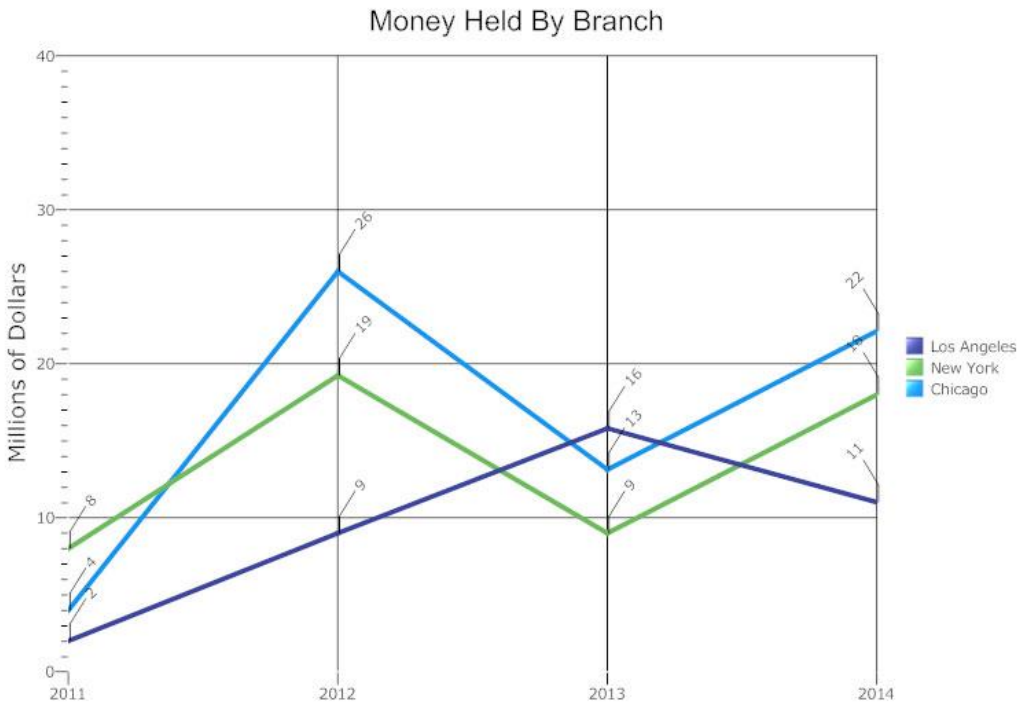
Different types of graphs

- Bar chart :
  - is a chart or graph that presents categorical data
  - It has rectangular bars with heights or lengths proportional to the values that they represent.



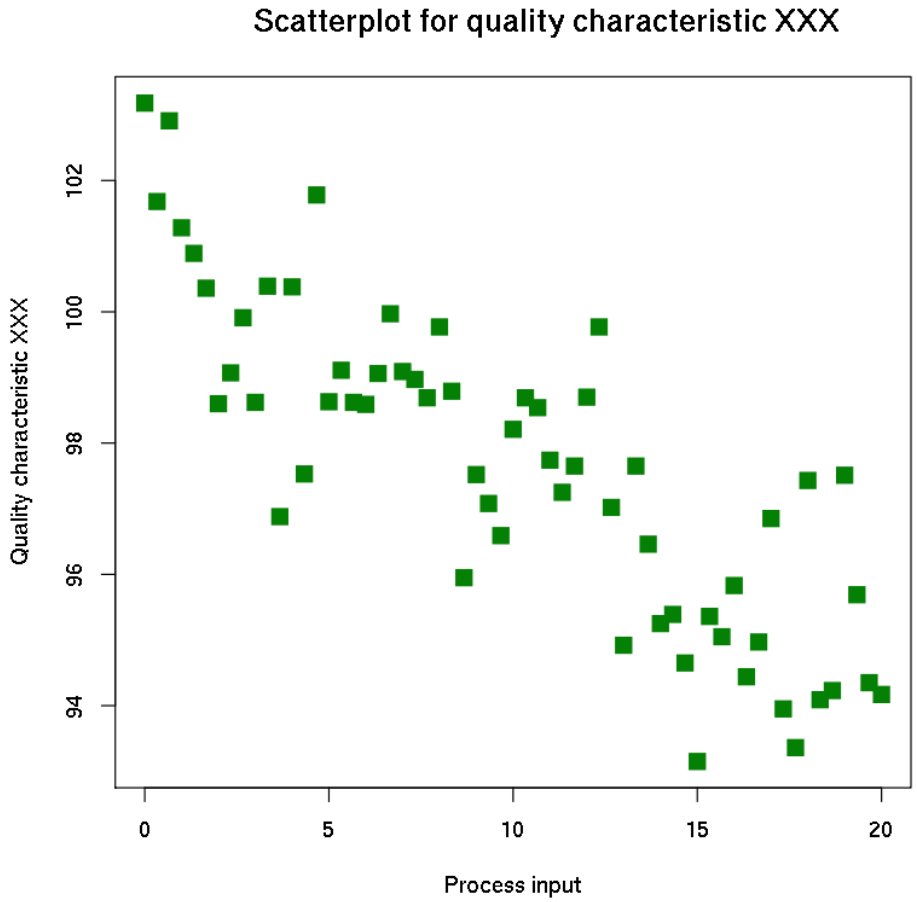
https://www.smartdraw.com/bar-graph/

- Line graphs :
  - A line graph, also known as a line chart
  - It is a type of chart used to visualize the value of something over time.
  - For example, a finance department may plot the change in the amount of cash the company has on hand over time.



https://www.smartdraw.com/line-graph/

- Scatter plot :
  - A scatter plot (also called a scatterplot, scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates
  - It displays values for typically two variables for a set of data.

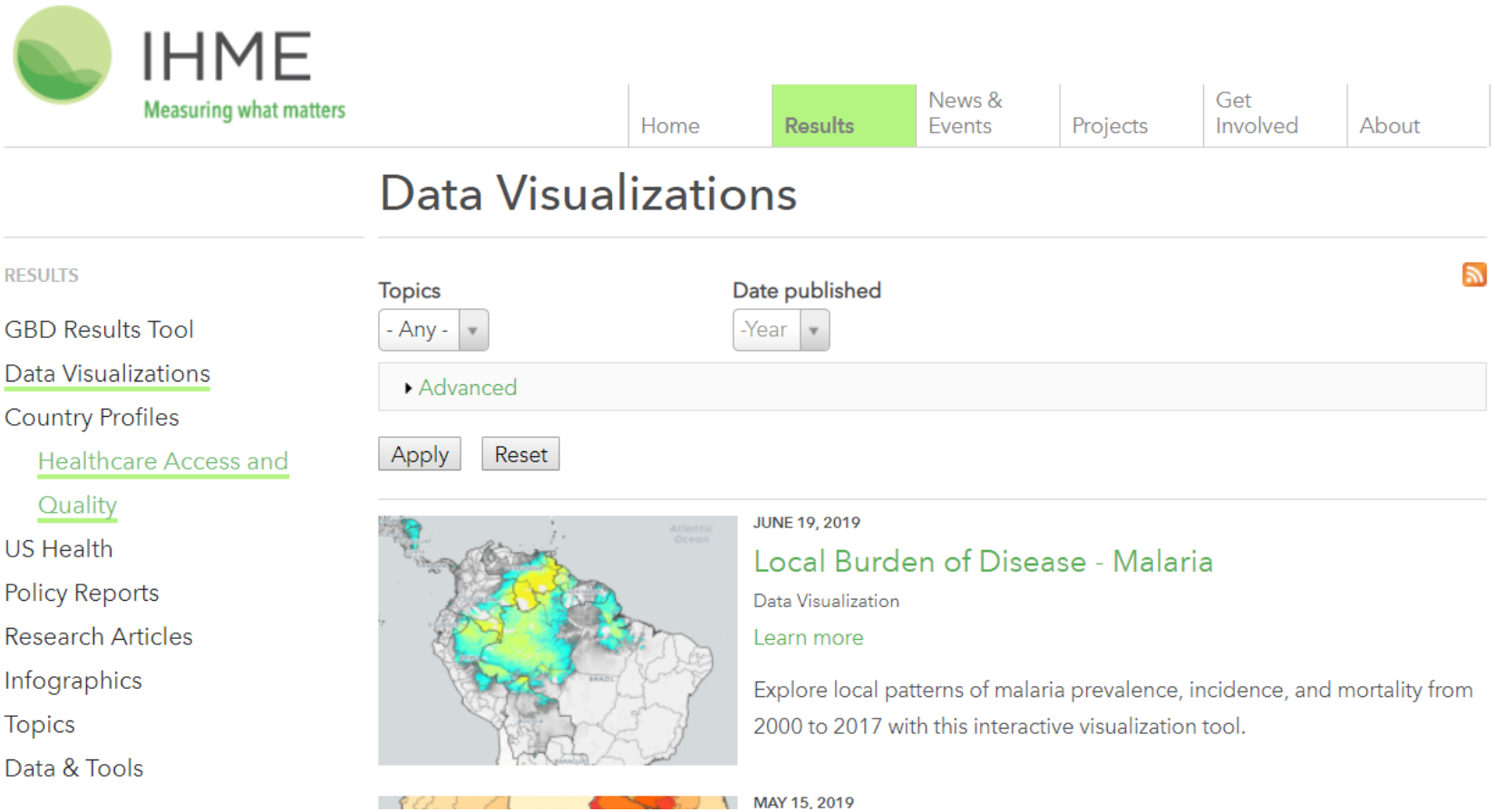


https://en.wikipedia.org/wiki/Scatter_plot

Communicating data visualization

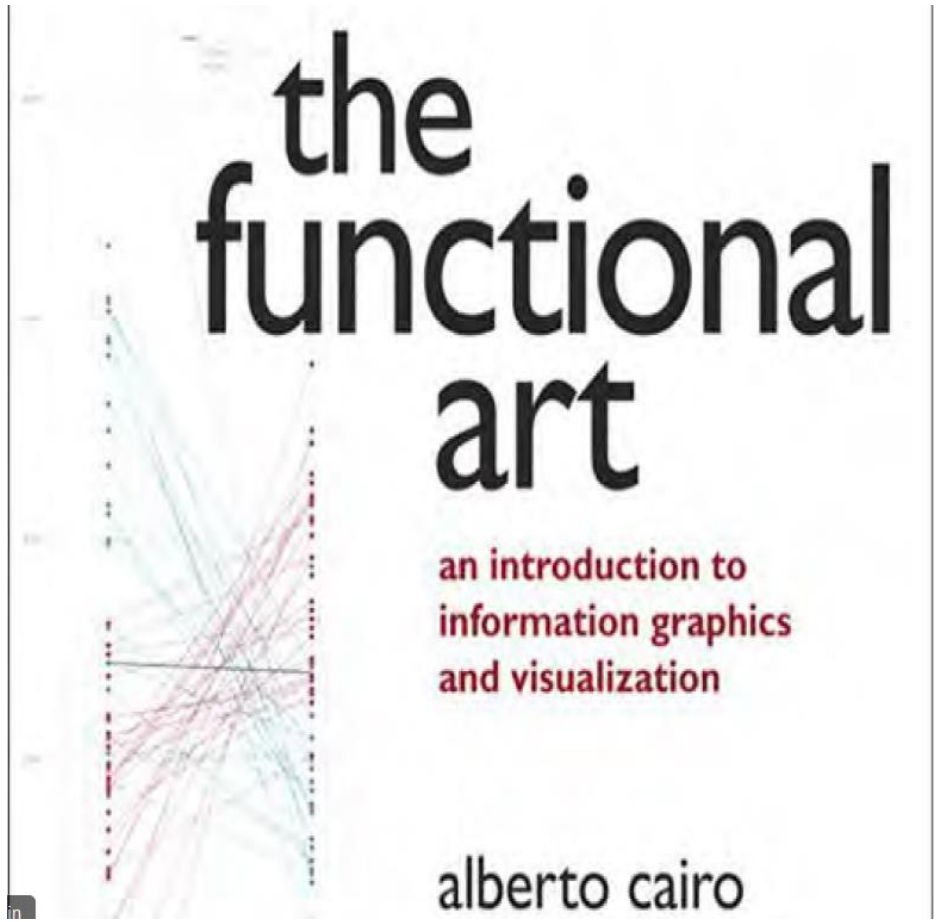
- https://www.gapminder.org/

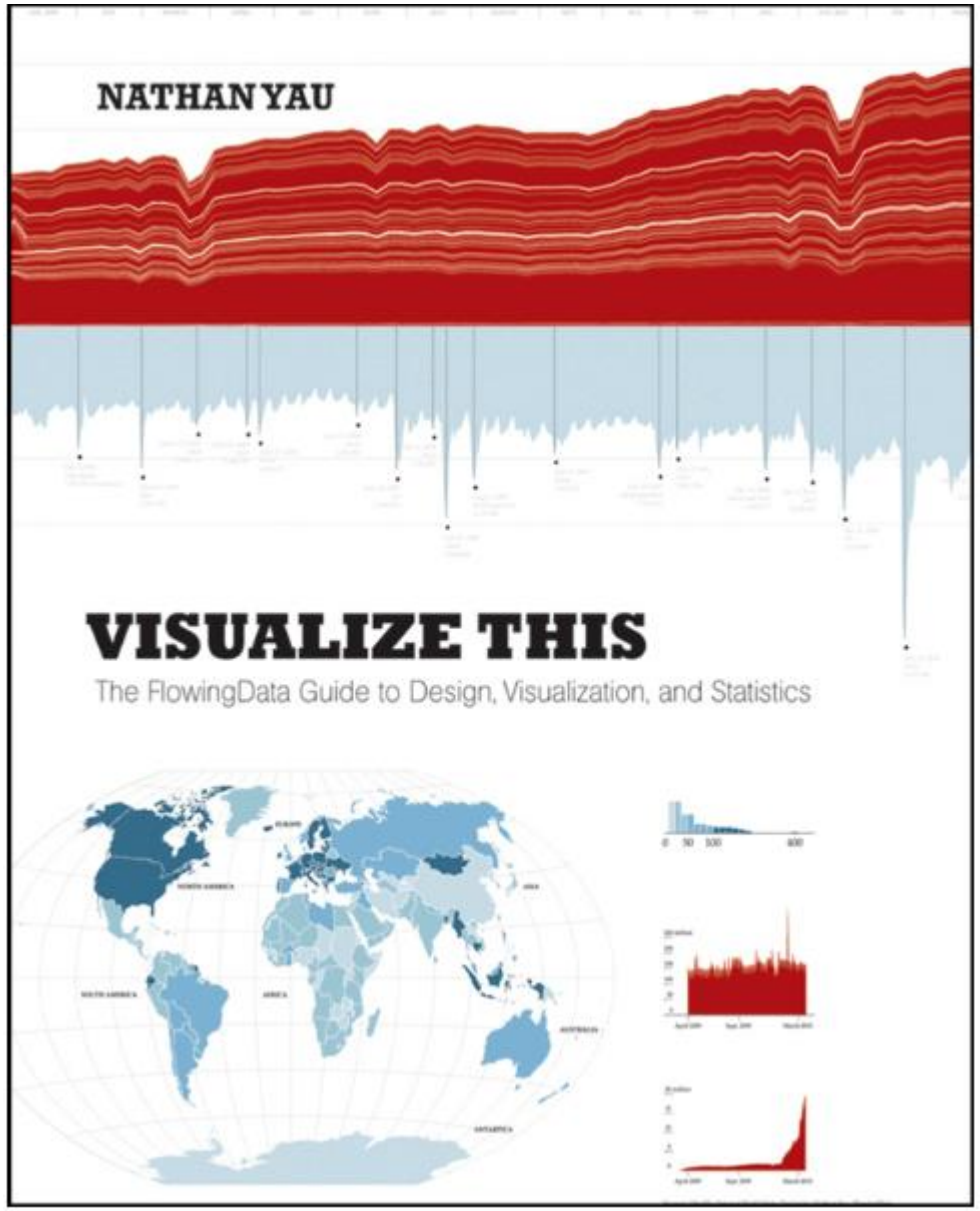
- http://www.healthdata.org/results/data-visualizations



References

- Edward R Tufte. The Visual Display of Quantitative Information Graphics.


the
functional
art
an introduction to
information graphics
and visualization
alberto cairo


NATHAN YAU
VISUALIZE THIS
The FlowingData Guide to Design, Visualization, and Statistics

References

- https://r-graphics.org/
- https://r4ds.had.co.nz/data-visualisation.html
- http://www.ggplot2-exts.org/gallery/