Effective graphical displays

Jean-luc Doumont

What?   So what?

Information ▶ Message

Interpretation
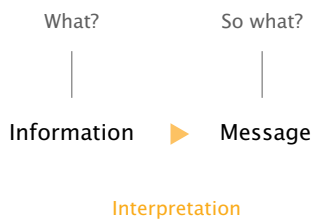
Get your audience to

- pay attention to,
- understand,
- (be able to) act upon

a maximum of messages, given constraints

To optimize communication…

First law
Adapt to your audience

Second law
Maximize the signal-to-noise ratio

Third law
Use effective redundancy

Planning the graph

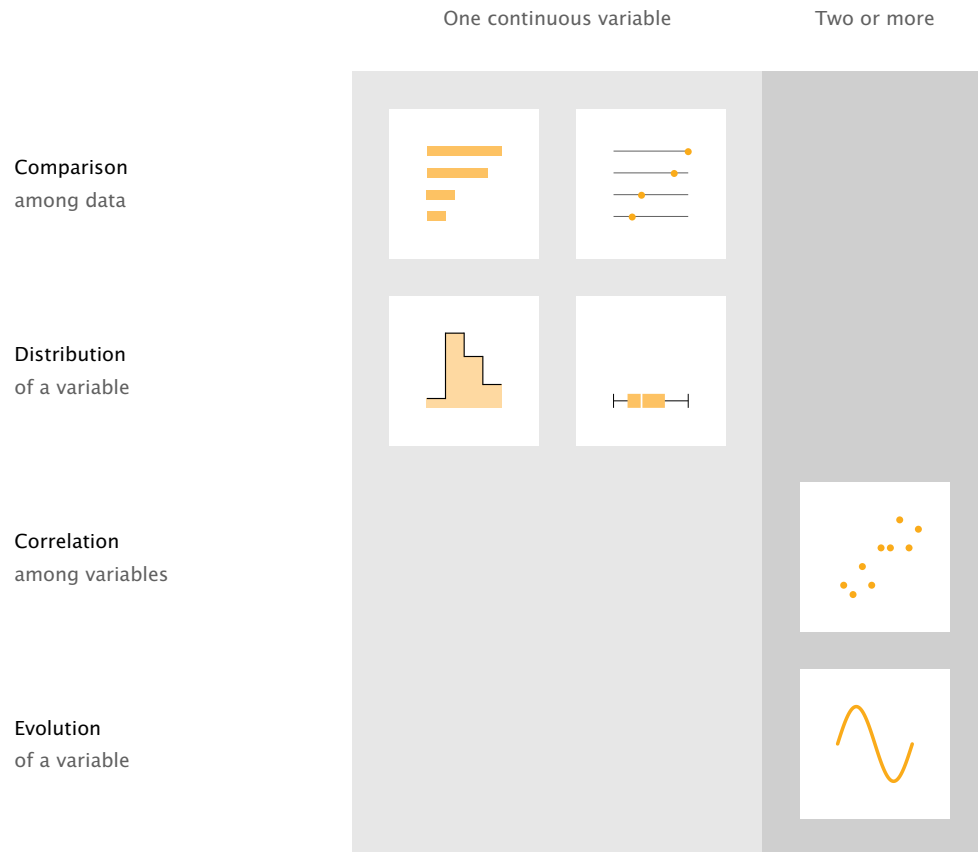Select the type of graph on the basis of

- the intended message or research question
  (comparison, distribution, correlation, evolution)

- the structure of the data set
  (number and nature of variables)

Patient	Gender	Weight [kg]	Conc. [ng/ml]
1	female	75	31.3
2	male	81	16.5
3	female	52	11.8
⋮	⋮	⋮	⋮

Discrete variable — Continuous variables

Designing the graph

One continuous variable　　Two or more

Comparison
among data

Distribution
of a variable

Correlation
among variables

Evolution
of a variable

Select the basic design on the basis of the intended message or research question and the number of continuous variables. To render discrete variables (that is, when you are comparing comparisons, distributions, correlations, or evolutions of data among subsets), distinguish the subsets within one panel or display them in as many juxtaposed panels (or combine these two approaches to represent several discrete variables).

Subsets displayed in juxtaposed panels
(with identical scales)

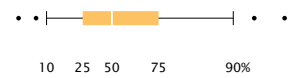t_0　Faster diffusion　　　　　　　　　Faster drift
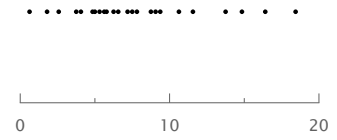
t_1

t_2
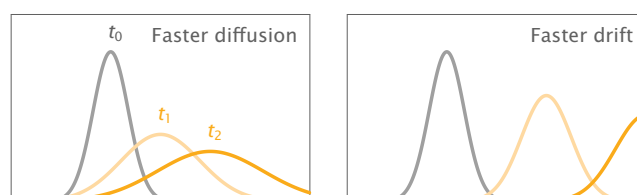
Subsets distinguished
within one panel

To compare data, consider a length representation (horizontal bars, starting necessarily from zero) or, for closely grouped data, a position representation (dots along a scale that need not start from zero). Both of these afford more accuracy than a pie chart.
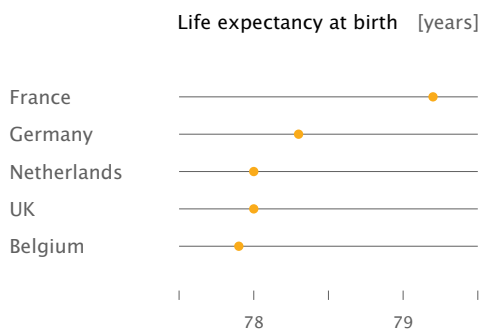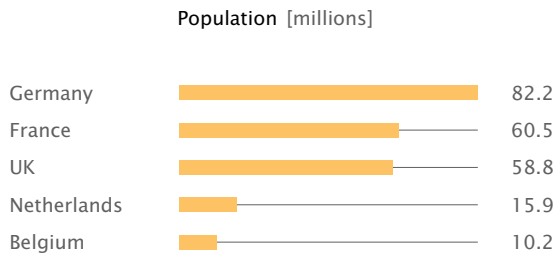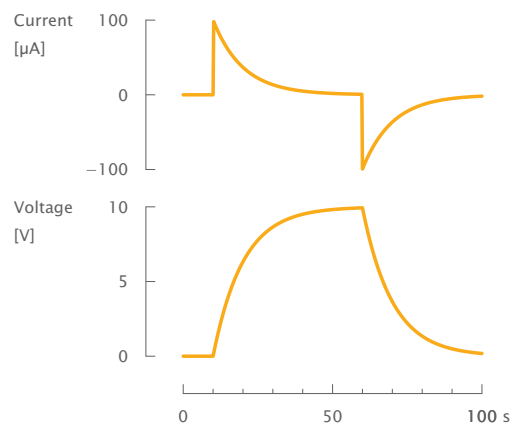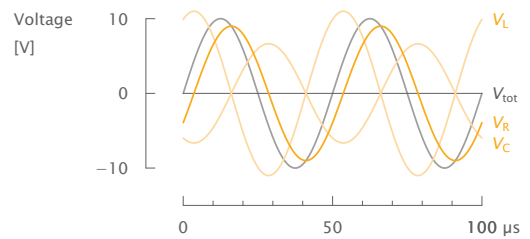
Showing the entire data set, as points along a scale, is the most accurate way to convey its distribution. For large data sets, it may be useful to summarize the distribution with a histogram or with a box plot, showing five percentiles plus the outliers. Box plots allow an easy comparison between subsets of data, each summarized by one box, along the same scale.

Population [millions]

Germany		82.2
France		60.5
UK		58.8
Netherlands		15.9
Belgium		10.2

Life expectancy at birth [years]

France
Germany
Netherlands
UK
Belgium

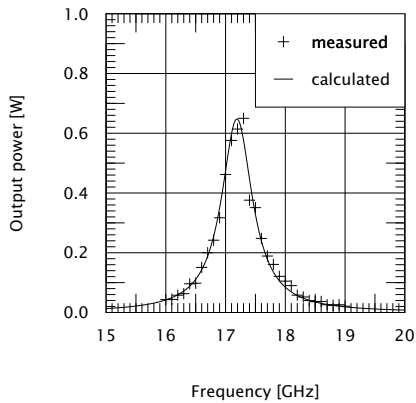78   79

Subset 1
Subset 2

10 25 50 75   90%

Scatter plots are a powerful way to reveal correlation or simply to explore bivariate data by mapping it out. For more than two variables, they can be combined in arrays (one scatter plot for each pair of variables).
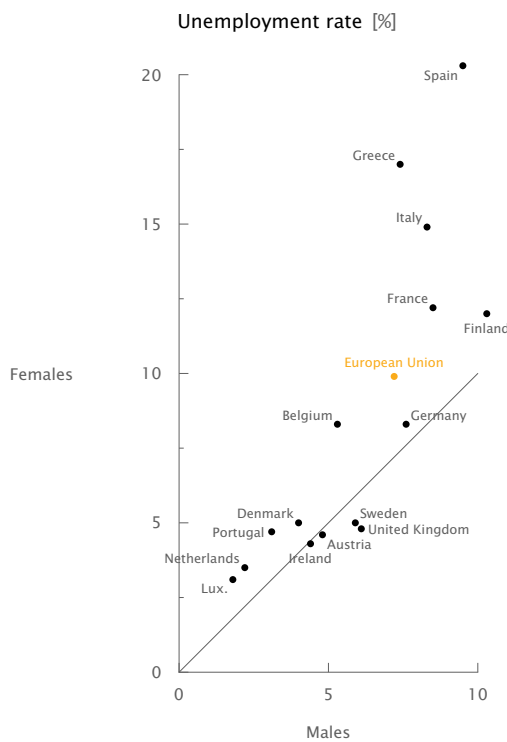
The evolution of one or more dependent variables versus an independent one is best shown by lines. Variables expressed in different units must be drawn in different panels, with a common horizontal scale.

Unemployment rate [%]

Voltage [V]

V_L
V_{tot}
V_R
V_C

0   50   100 μs

Current [μA]

0   50   100 s

Voltage [V]

Constructing the graph

A poor graph

The graph exhibits a very low signal-to-noise ratio, with excessive tick marks and uncalled-for grid lines, and comparatively little ink to represent the data.

The graph is not intuitive, for the separate legend (a key to the symbols) prevents global processing. In a sense, it is a graph to *read*, not a graph to *view*.
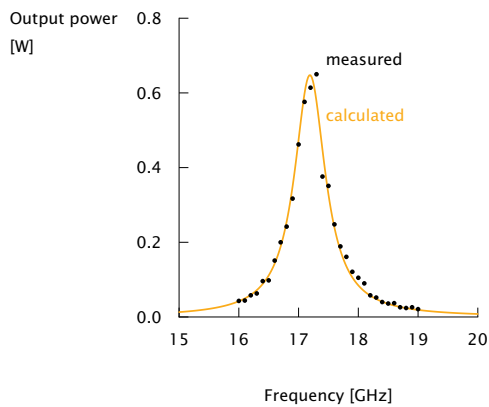
A good graph

The graph is plainer and therefore better contrasted: the background no longer interferes with the data, yet it provides sufficient information about them.

The graph is more intuitive: the labels, positioned next to the data, provide the required clarification where it is needed (when viewers look at the data).
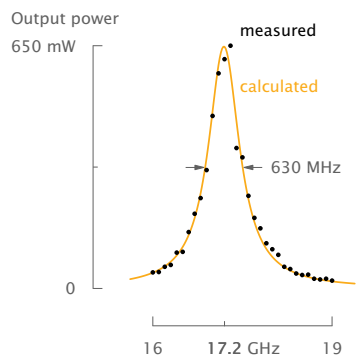
A better graph

The graph shows the data and nothing but the data: tick marks are relevant, not arbitrarily equidistant; nondata lines are gray, to make the data prominent.

The graph readily answers questions about the peak (position, height, and full width at half maximum) and about the range over which data were acquired.