Efficient screen real estate management:
improving data visualization for small screens.

Jorge Camoes (jorge.camoes@wisevis.com)[[1]](#footnote-1)

**Keywords:** data visualization, Eurostat, statistical books, small screens

# Introduction

Even the simplest data set can be visualized from multiple perspectives. Not only our options are situation-dependent, but the criteria we use to evaluate them are subject to discussion. More than a set of algorithms to be applied in each situation, data visualization is a flexible visual language with multiple purposes, where “speakers” can develop their unique style.

When paired up with technology, this openness somewhat blurred the traditional divide between exploration and explanation / communication in data visualization. Other fault lines emerged in recent years: tools (programming languages vs. point-and-click applications), interaction (dynamic vs. static), medium (analog vs. digital).

In a context of an ever-increasing volume of data, and more types of data that could be visualized, more people are entering the field. Some have a less traditional profile (graphic designers, journalists), while others feel the need to add visualization and design to their core skills (statisticians).

## Visual encoding

Data visualization encodes data into geometric primitives like points, lines, areas or volumes. Their characteristics (what Jacques Bertin calls “retinal variables” [1]), including position, size, or color, vary with data and design choices.

Researchers often try to establish the effectiveness and accuracy of retinal variables. In a well-known study, Cleveland and McGill [2] ranked them by how accurately people could read them. Jock D. Mackinlay [3] expanded Bertin’s list and showed their accuracy is task -dependent.

Recent research seems to be less interested in measuring accuracy, shifting focus to interest and memory: which of the suggested designs will be more effective at attracting interest and be retained in memory?

It can be argued that these perspectives complement each other: while the former can help improve visualization in a mass production environment (business, statistical offices), the latter benefits a more handcrafted/artistic data visualization perspective (the media).

## Managing page real estate

Since humans enjoy a large field of vision, data visualizations could be designed to take advantage of this. Large, high resolution screens can in fact display hundreds or thousands of data points. This is less promising once we factor in the fact that only a small section of our field of vision receives our attention at any given moment. The eye moves constantly from one fixation point to the next. Still, there are advantages (speed, working memory) in using a large surface, when compared to scrolling or flipping pages.

But large screens tend to be supported by large and heavy monitors, a problem in an always-connected, mobile society. While we wait for a technological solution, we must rely on high-resolution small screens that we can carry in our pockets.

We can take advantage of higher resolutions to scale objects down and increase the number of objects displayed on a screen, but there are practical limits: text can look better (smoother curves) but it can’t be reduced below a certain threshold.

We can also see this when displaying data points. Each data point in a table can be encoded into a point, in a vector or in a Cartesian space. Strictly speaking, this point can be displayed using a single pixel but, similarly to font size, other factors (visual acuity, task, or design choices…), prevent us from reaching this theoretical minimum.

## Terminology

Defining a few of the terms used in this discussion:

* **Efficiency.** In a general sense, efficiency is the ability to achieve a result while minimizing resource allocation. Here, the resource we care about is screen or page real estate, the available space to display data visually, and the goal is to compare similar messages (for example, how population age structure changes over time) using designs that have a larger or a smaller graphical footprint. Alternatively, a chart is more efficient if, for the same area, it allows for higher data density.
* **Graphical footprint.** Borrowing the concept from “ecological footprint”: the amount of resources required to perform a certain action. In this case, the total area used by visual objects.
* **Chart area.** The total area of a chart, including title, legend, labels and annotations.
* **Plot area.** The rectangular area defined by the axes (or surrounding a circular chart) where the data is plotted.

## Sources of inefficiency

This discussion will examine a few typical sources of inefficiency. It is not meant to be an exhaustive list, and with the necessary caveat that their nature can change:

* **Geometric primitives.** Representing a data point using a dot takes up much less space than representing it using a thick line (a bar) or an area (a slice of a circular chart).
* **Decoration.** decorative objects and effects add to the total area of the chart. The most common type of decoration is the pseudo-3D effect.
* **Textures.** Unlike decoration, textures don’t add extra elements to the representation of a point, but the object itself must be bigger so that the texture can be perceived. A simple, 2-slice pie chart can be very small, but it must be bigger if a 3D effect is added, and even bigger, if a glossy texture is used to paint the slices.
* **Chart type.** A chart type is often associated with a geometric primitive, and thus inherits its graphical footprint. The total area may not change, but switching to a more efficient primitive can free up space, allowing us to add more data points.
* **Chart size.** Chart size is often pre-defined by style guidelines as a fixed proportion of page width or height, with the net effect of creating charts that are unnecessarily large.
* **Labels and legends.** Labels, especially the labels in the categorical axis, can have a dramatic effect on the total chart size, due to position and string and font sizes. Legends are often unnecessary and placed outside the plot area, adding to the total area.
* **Data.** The scales of quantitative axes should be designed to allow for good discriminability. To maintain it in the presence of strong outliers often means increasing chart size.
* **Metrics.** A wrong combination of metrics and chart types can lead to a very inefficient visualization. If the distribution is essentially flat (for example, annual GDP in euros) and a bar chart (which requires starting the scale at zero) is used, the result is a flat chart with no apparently meaningful variation. Changing from absolute values to percent change over time improves the message, even if the chart type remains the same.

# Methods

To test for efficiency, a chart is used as a reference and alternative designs using the same data are suggested, subject to the following constraints and defaults:

* Typical Eurostat charts: in general, examples are picked from common chart types used in Eurostat publications, like population pyramids, bar and line charts, but there can be references to chart types (gauges) or design choices (3D effects) not used by Eurostat.
* Typical Eurostat data: especially for bar charts, the dataset includes 28 or more countries.
* Comfortable reading of the entire chart in a 5.2” screen with 1080x1920 resolution.
* Font: common sans serif font, 11 p.p.
* Total area should not increase.
* Alternative designs should require similar skills and be created using similar tools.
* More data will be added if the redesign frees up enough space and if it improves the message.

Note that what a chart says is shaped, not only by the data, but also by any of the design choices. This means that every alternative design communicates a slightly different message.

# Results

This is still a work in progress, but it’s clear that being aware of the common sources of inefficiencies, and examining and evaluating each object in a chart, can improve efficiency. For example:

* A simple change like moving a legend into the plot area can decrease the chart area size by a third or more.
* Some regular-sized charts can in some cases be replaced by sparklines.
* A chart heavily optimized for efficiency, like a horizon chart, can display in a single screen data that would otherwise be spread over multiple pages.

If a single chart could summarize the benefits of an efficiency-oriented approach to data visualization, the population pyramid would be a strong candidate:

* By displaying the data in both sides of the vertical axis, it unnecessarily duplicates the size of the plot area.
* Using bars makes it harder adding more data (to compare two populations, for example).

When both series are placed on the same side of the vertical axis and lines are used instead of bars:

* The size of plot area is cut by half;
* The traditional chart makes it hard to compare Male and Female structures. This change improves effectiveness, because comparisons are easier.
* Since no other changes are made, traditional analysis of its overall shape is retained.
* Because lines are used instead of bars, more populations can be displayed in a single chart.
* This change doesn’t seem to require higher literacy to make sense of the chart.

These short examples show that there is a wide range of choices that can impact chart efficiency. Some of these choices are simple and cost-free but, in some cases, we must assume a higher graphic literacy, or we must decide on tradeoffs (improved comparability versus some loss of detail).

# Conclusions

Designing visual representations with effectiveness in mind helps us finding ways of reducing their graphical footprint, with several advantages:

* More data can be juxtaposed or superimposed, creating a more complex and complete portrait of the topic.
* Avoid sequential data display, often not desirable because of the burden on the working memory.
* Can be placed near a text discussing the topic.
* Fit into small screens.
* Create richer dashboards for desktop monitors.

Designing for efficiency, like designing for effectiveness, is a rational choice that should create a better visual transcription of the data table. But it must be recognized that we often have different and conflicting goals in data visualization design. Like effectiveness, efficiency should inform our designs, but we should be flexible enough to adjust to a specific situation. Effectiveness shouldn’t be reduced to banning pie charts, and efficiency shouldn’t be reduced to counting pixels.

Quoting Amanda Cox, then graphics editor at *The* *New York Times*, “[t]here’s a strand of the data viz world that argues that everything could be a bar chart. That’s possibly true but also possibly a world without joy.” [4] This emotional dimension must also be considered, and a tradeoff where some effectiveness / efficiency is lost, and some attention is gain, is often the right solution.

# References

[1] Jacques Bertin, *Semiology of Graphics: Diagrams, Networks, Maps*. Translated by W. J. Berg. University of Wisconsin Press 1983

[2] William S. Cleveland and Robert McGill: “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods” in *Journal of the American Statistical Association*, Vol. 79, No. 387 (Sep., 1984), pp. 531-554

[3] Jock Mackinlay: “Automating the design of graphical presentations of relational information”, in *ACM Transactions on Graphics*, Volume 5 Issue 2, April 1986, Pages 110-141, ACM New York, NY, USA

[4] Amanda Cox, interviewed by Scott Berinato, *The Power of Visualization’s “Aha!” Moments*, Harvard Business Review, March 9, 2013 < https://hbr.org/2013/03/power-of-visualizations-aha-moment >

1. Wisevis Unipessoal. [↑](#footnote-ref-1)