



BABELgrupo



babel_grupo



/company/babel



BABELgrupo



BABELgrupo

Making a great data
visualization using D3.js

about:me



Juan Antonio Jiménez

 juananthony

 juananthony



Contenido

- Introducción D3.js y SVG
- Introducción a la visualización de datos
- Cómo trabajar con datos
- Introducción a la percepción visual
- Ejemplo de visualización con D3.js

¿Qué es D3?

- Librería javascript
- D3 viene de Data-Driven Documents
- Hace mucho más que gráficos
- **No es una librería para hacer gráficos rápidamente.**

¿Buscas otra cosa?

Datawrapper

River Blog Academy FAQ Login

Enrich your stories with charts, in seconds.

There was a snowstorm in the Northeast and Midwest during the week that the BLS does its survey, which kept some workers at home. Additionally, the "retail apocalypse" of announced store closings meant that more jobs than normal left the economy during the month. This month, the disappointing March number was revised down from 98,000 to 79,000. But the April jobs report provides a bounce back in part because of warmer weather and fewer layoffs. The Labor Department reported gains in hospitality, mining, healthcare, and finance. Including the revisions for the February and March reports, an average of 174,000 jobs were added per month over the last three months.

Monthly Changes in U.S. Employment (Non-Farm), 2006-2017

Source: Bureau of Labor Statistics, Dept. of Labor, United States. Data as of April 2017.

[CREATE A CHART](#) [CREATE A MAP](#)

DGSAÅ: Syv i tiethus bestilte selv rapporten SSR ikke ville for Erp

In the past, the Fed has indicated that sustained growth in the U.S. labor market is one of the factors it considers when deciding interest rates. The strong April jobs

Datawrapper makes it easy to create beautiful charts.
See for yourself how different newsrooms use our charts:

NVD3.js Home Examples Live Code Source Blog Downloads ZIP TAR.GZ

NVD3 Re-usable charts for d3.js

This project is an attempt to build re-usable charts and chart components for d3.js without taking away the power that d3.js gives you. This is a very young collection of components, with the goal of keeping these components very customizable, staying away from your standard cookie cutter solutions.

[View more examples >](#) [GitHub Repo](#)

Getting Started

Download [d3.v3.js](#). This is the only required library for NVD3.

Download the latest nv.d3.js (version 1.8.1).

Checkout our [examples gallery](#) for how to code your first chart. Our examples should provide you with enough code to start making beautiful

Any questions or bug reports? Visit our [GitHub](#) page.

www.datawrapper.de

nvd3.org

¿Cómo funciona



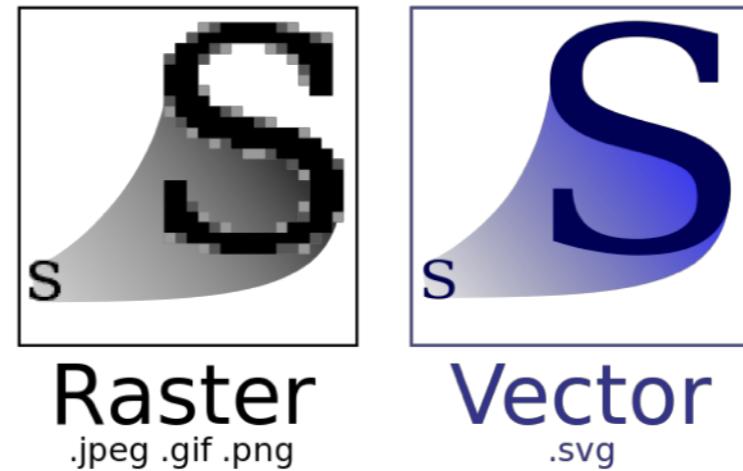
- D3 se usa para procesar y representar datos, añadir interactividad y optimizar la visualización de datos.
- D3 se basa en selections and data-binding.
- D3 usa el DOM.

Ejemplos 1 y 2



<https://github.com/juananthony/D3-Meetup-examples>

Scalable Vector Graphics



- Formato para gráficos vectoriales (bidimensionales)
- Disponemos de diferentes objetos geométricos.

```
<line x1="..." y1="..." x2="..." y2="..." ... />
```

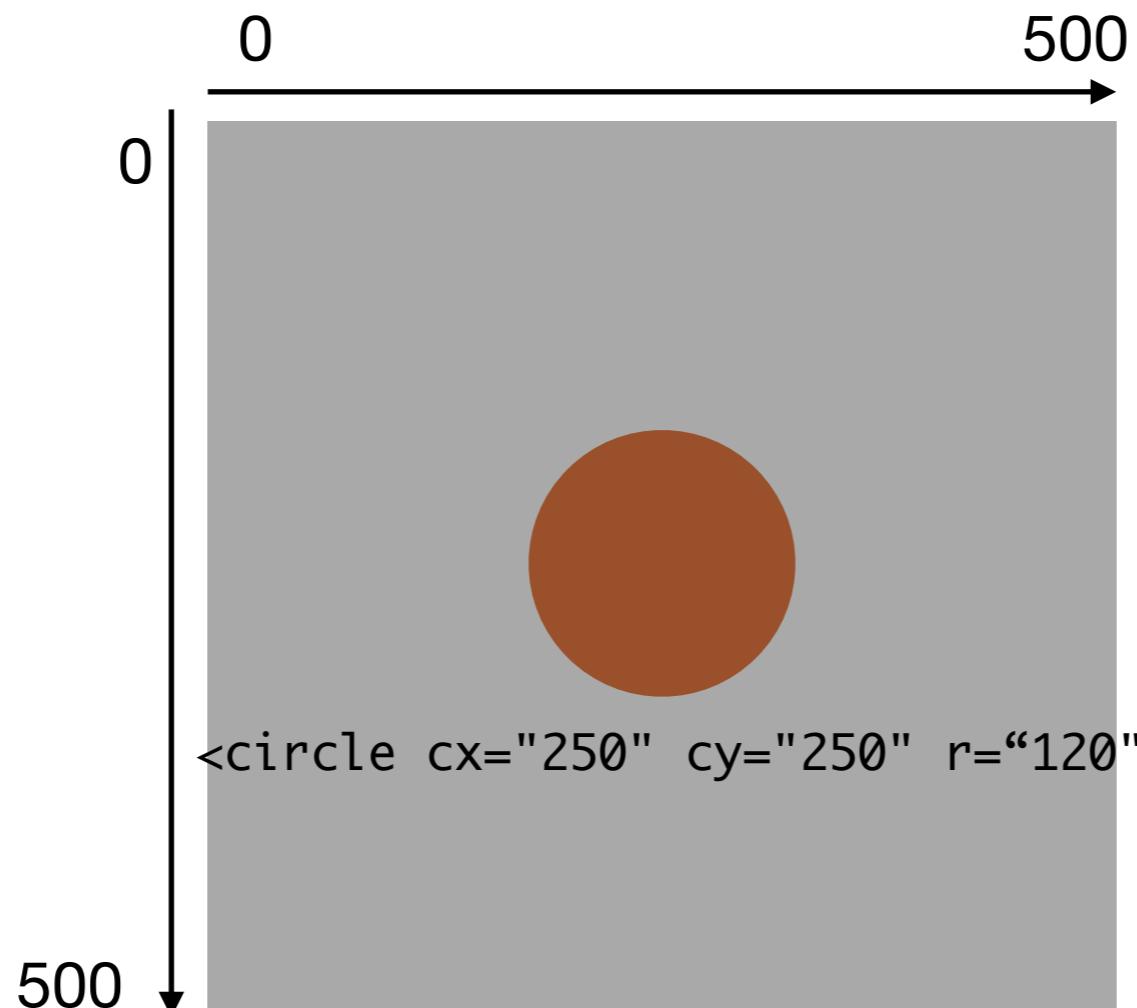
```
<rect x="..." y="..." height="..." width="..." ... />
```

```
<circle r="..." cx="..." cy="..." ... />
```

...
.

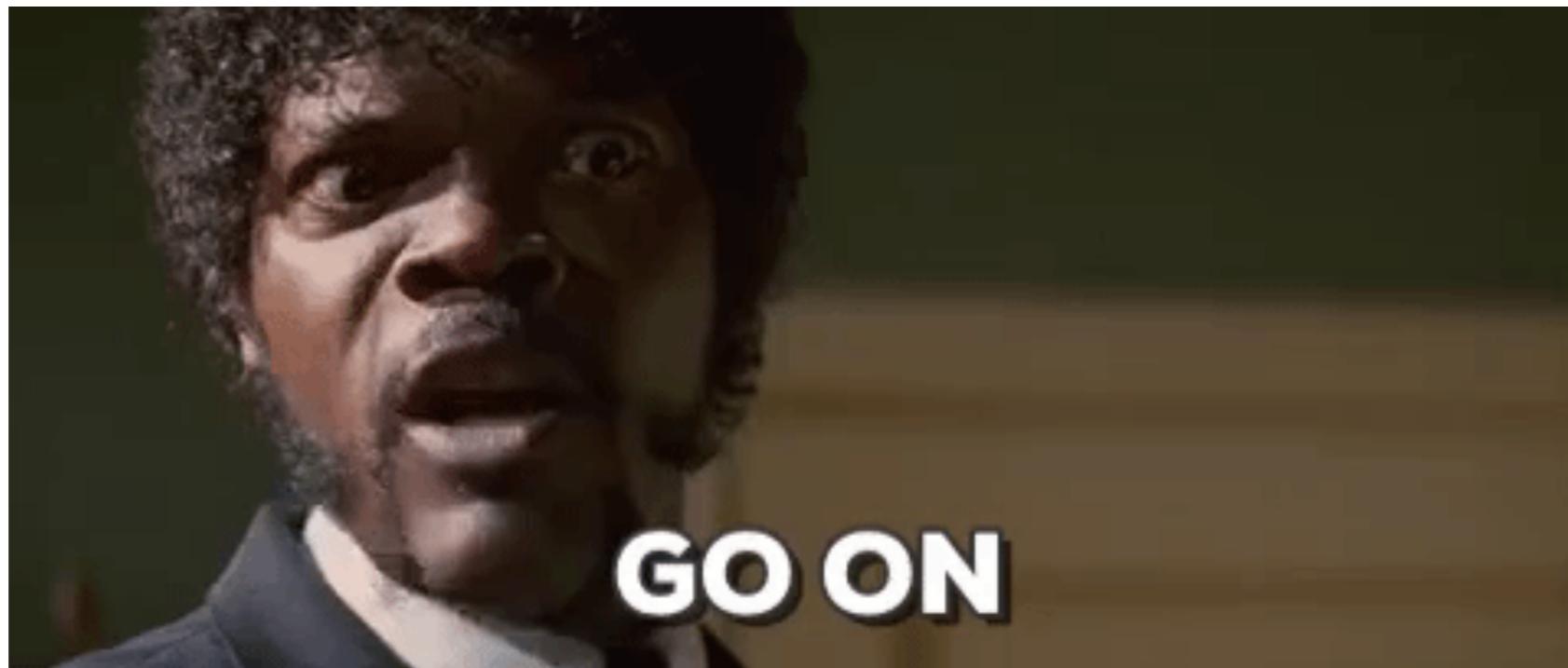
Scalable Vector Graphics

```
<svg height="500" width="500">
```



```
</svg>
```

Ejemplo 3



<https://github.com/juananthony/D3-Meetup-examples>

Introducción

Infografía

&

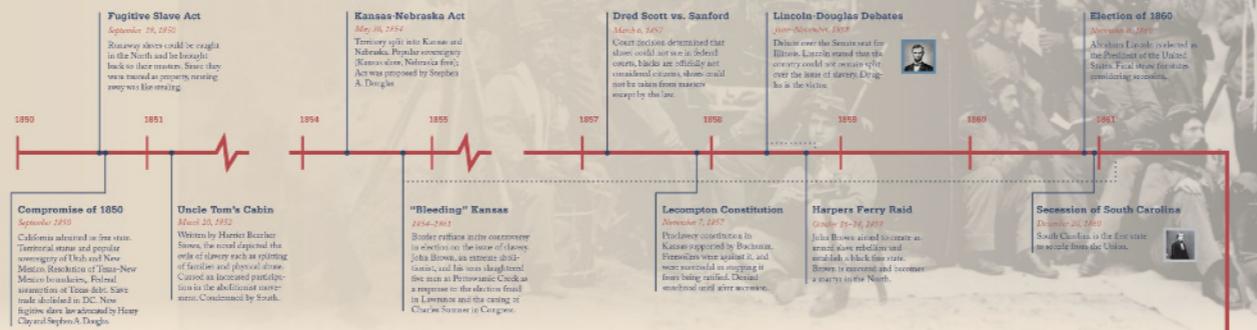
Visualización de datos

The American Civil War

From Build-Up to Resolution

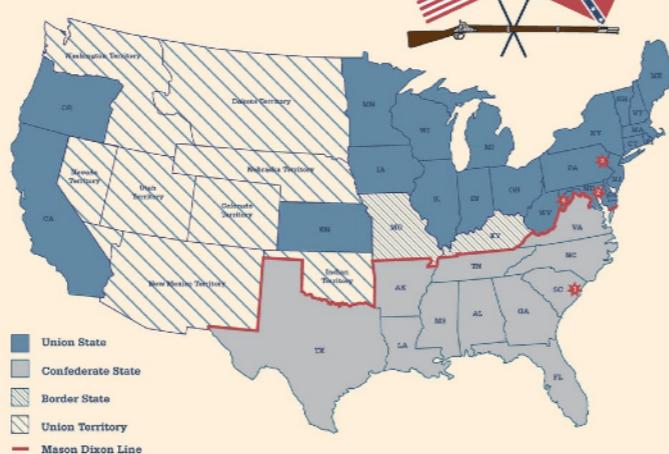
The American Civil War (1861–1865), often referred to as The War Between the States, was fought over the ceding of the Confederate States. Eleven southern slave states declared their secession from the United States and formed the Confederate States of America ("the Confederacy"); the other 25 states supported the federal government ("the Union"). After four years of warfare, mostly within the South, the Confederacy surrendered and slavery was abolished everywhere in the nation.

THE BUILD UP



THE WAR

THE STATE OF THE (DIS)UNION: 1861–1865



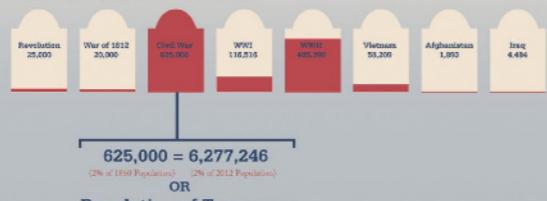
SOLDIERS IN 1861 V. 2012

A collective three million soldiers fought in the Civil War, 2 million for the Union and 1 million for the Confederacy. Our soldiers from today are not that different from those who fought in the Civil War 150 years ago.

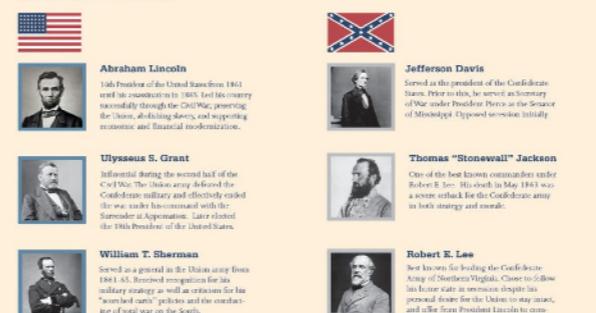


THE AFTERMATH

AMERICAN CASUALTIES BY WAR



IMPORTANT FIGURES



CAUSES OF UNION DEATHS



CIVIL WAR V. MODERN TRAGEDIES



Reconstruction

Putting the Country Back Together

Reconstruction began during the war (and continued to 1877) in an effort to solve the issues caused by reunion, specifically the legal status of the 11 breakaway states, the Confederate leadership, and the freedmen. Northern leaders during the war agreed that victory would require more than the end of fighting. It had to encompass the two war goals: secession had to be repudiated and all forms of slavery had to be eliminated. Lincoln and the Radical Republicans disagreed sharply on the criteria for these goals. They also disagreed on the degree of federal control that should be imposed on the South, and the process by which Southern states should be reintegrated into the Union. These disputes became central to the political debates after the Confederacy collapsed.

AGE 42

FAIRYTALE IN HISTORY

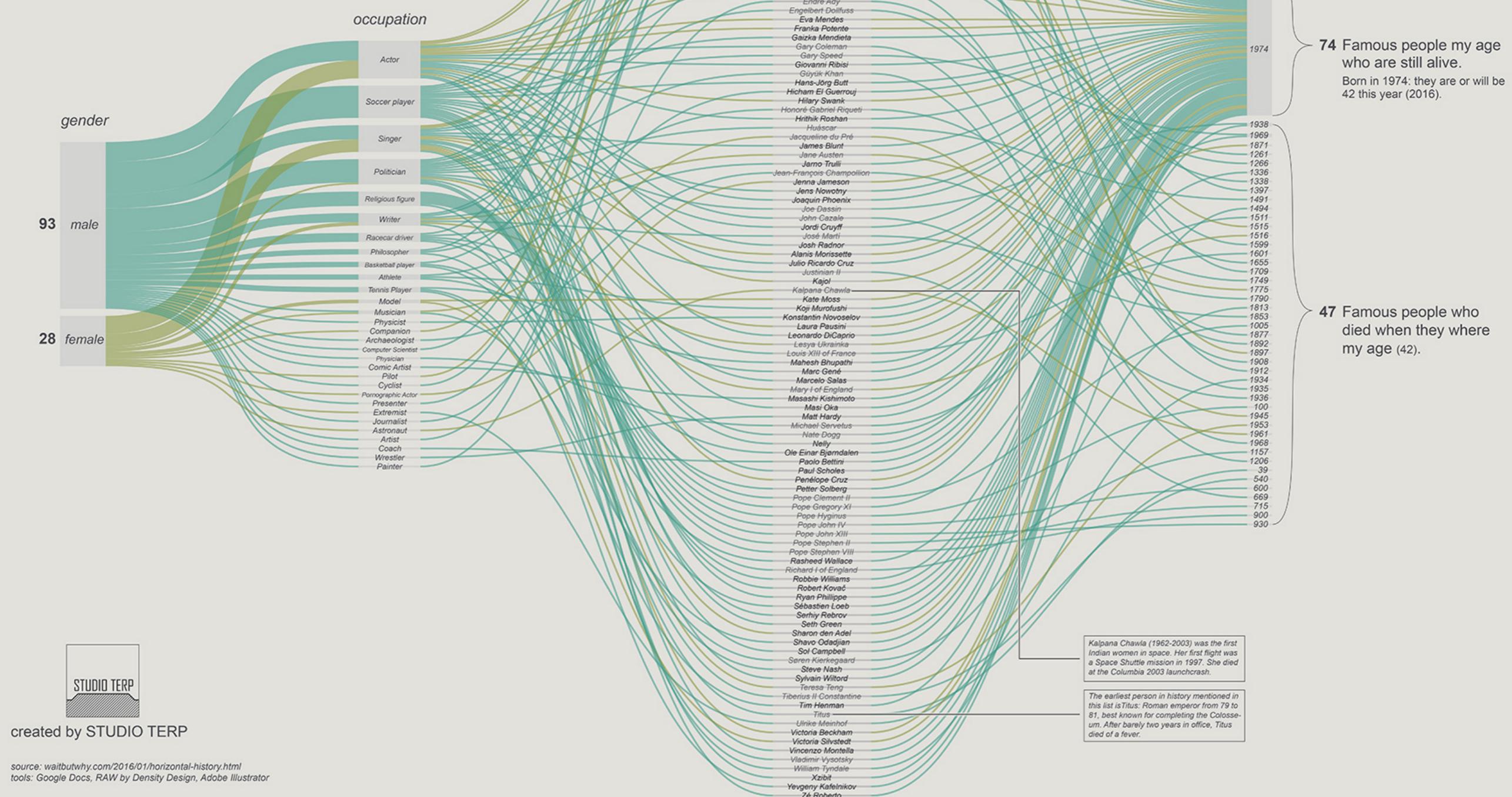
Departing from the original story and data by Tim Urban who tried to create a better understanding of history by showing a timespan of famous* people throughout history, this datavisualisation personalizes the data** by relating to the famous people who were born the same year as me (1974) and the famous people who died my age (42).

* Fame: The state of being widely known, recognized, or of great popular interest.

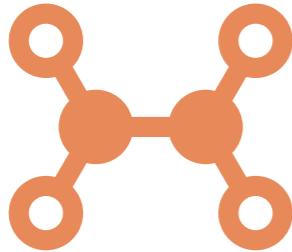
** The data was cleaned. Excluded are:

- the years BC.
- people of which data contained to much uncertainty considering their year of birth and/or death.

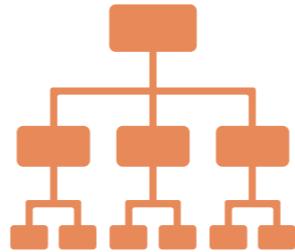
The remaining data however still is unreliable due to lack of exact dates.



¿Por qué es importante?



Relacionar datos



**Organizar y filtrar
la información**



**Enfoque analítico
para dar respuestas**

Etapas visualización



Investigación



Selección de datos



Boceto



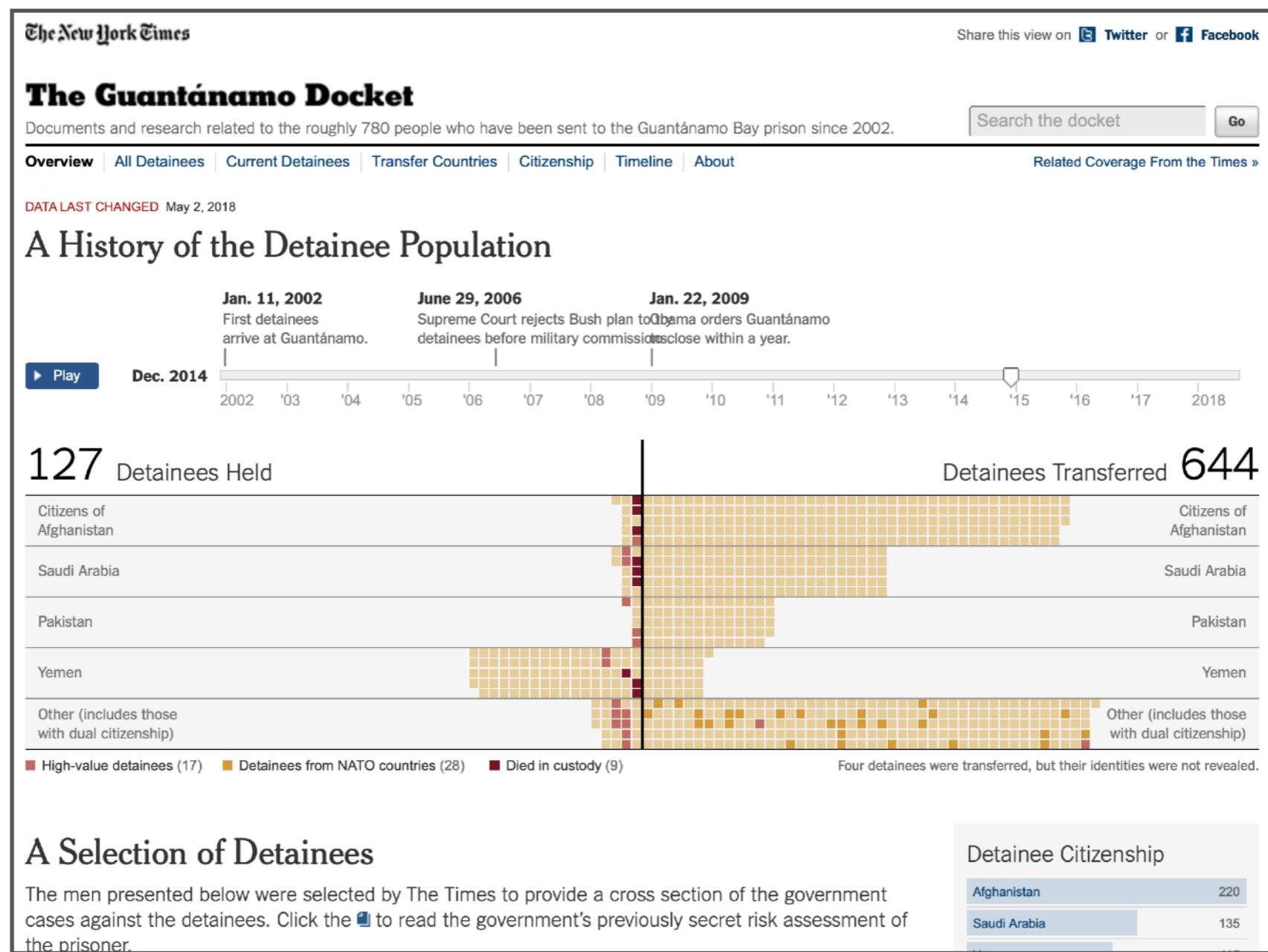
Elaboración



Percepción del usuario

Ejemplo

Enlace



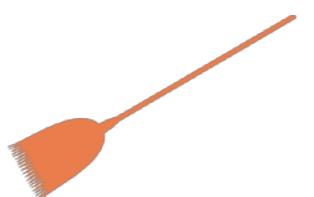
Trabajar con datos



Recolección



Preparación



Limpieza

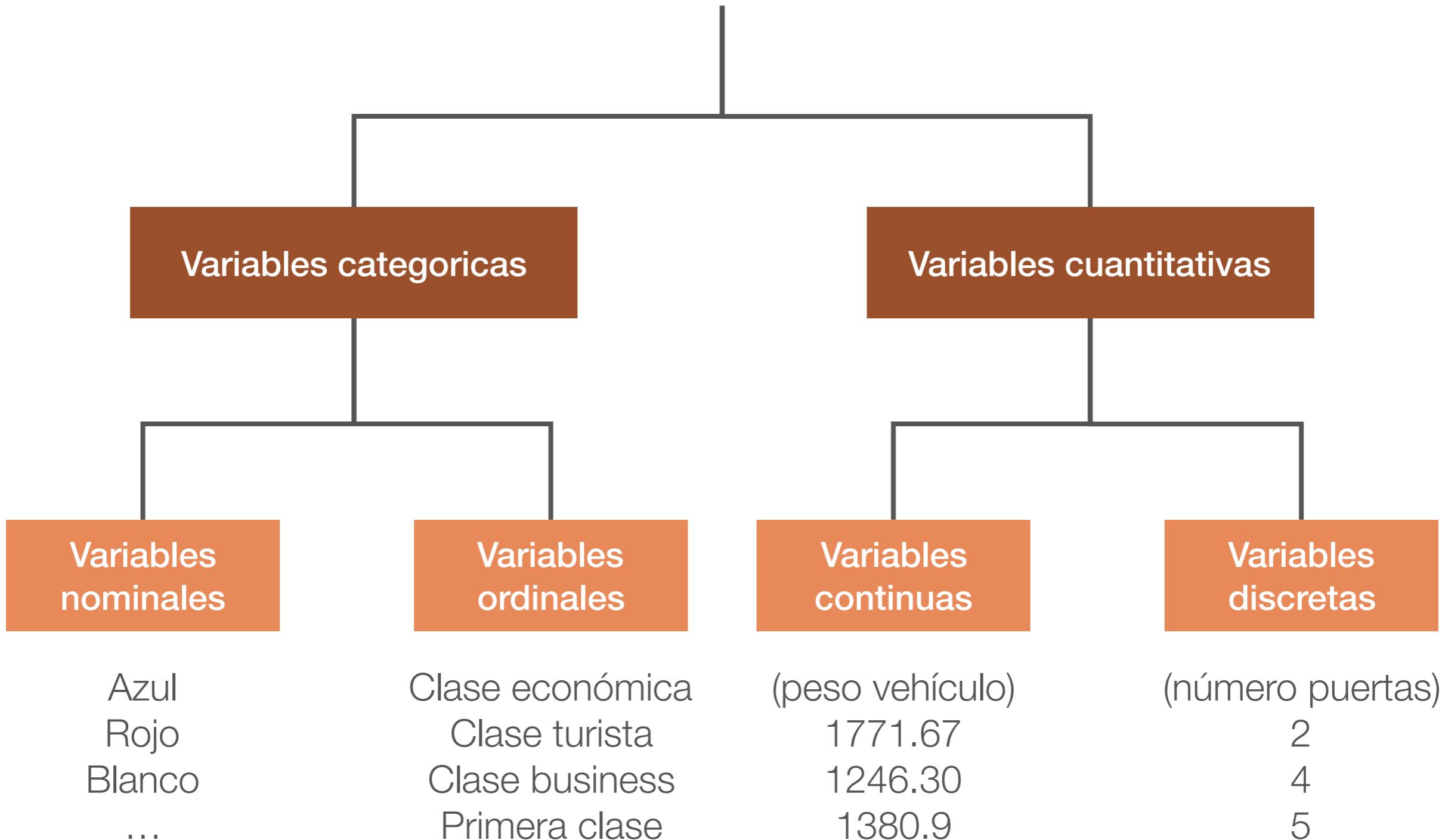


Transformación



Visualización

Tipos de variables

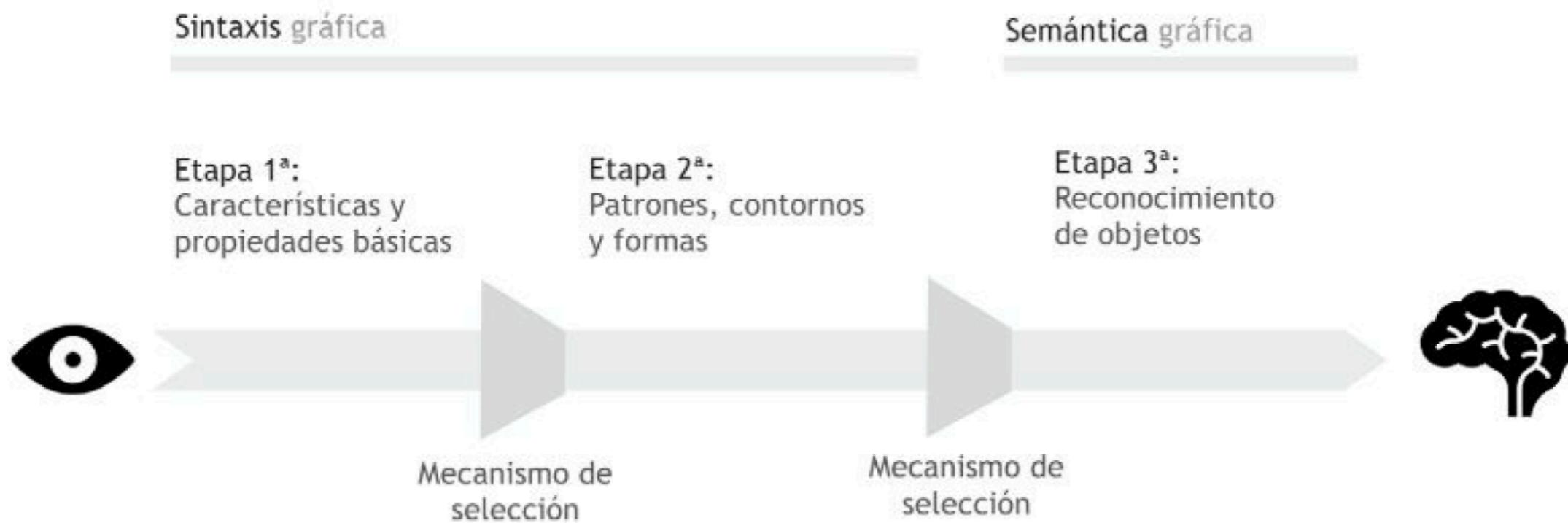


Percepción visual

- La atención es selectiva.
- La percepción es activa y no selectiva

Etapas del proceso de percepción

1. Percepción pre-atentiva.
2. Percepción atentiva.
3. Percepción post-atentiva.



1 - Percepción pre-atentiva

- Color



- Orientación



- Forma



- Tamaño



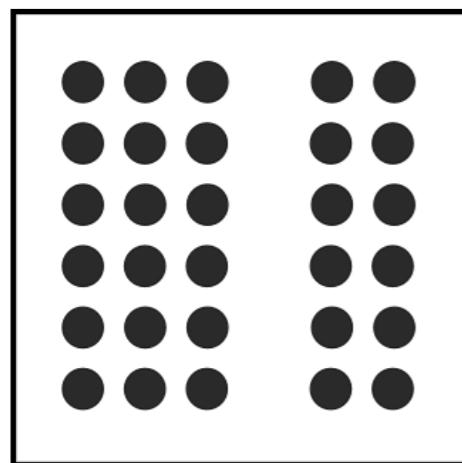
- Brillo



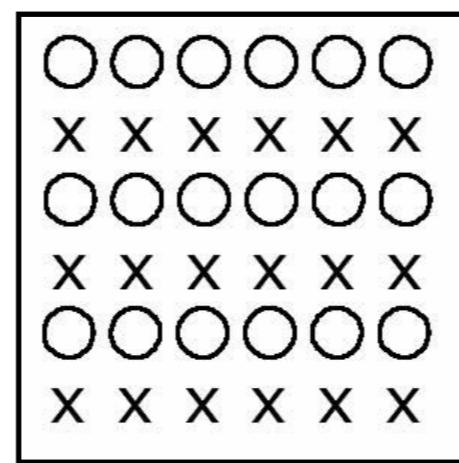
- Movimiento

2 - Percepción atentiva

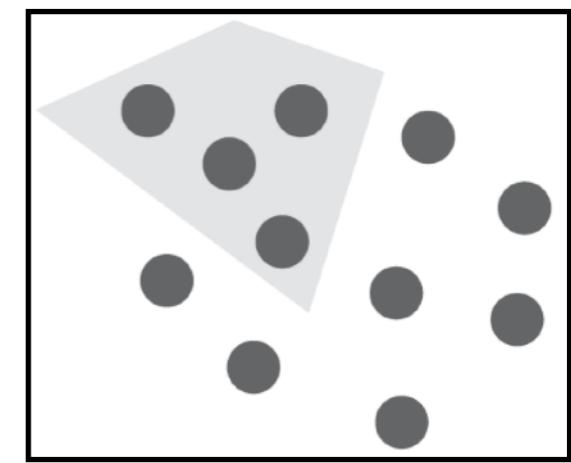
Psicología de la Gestalt



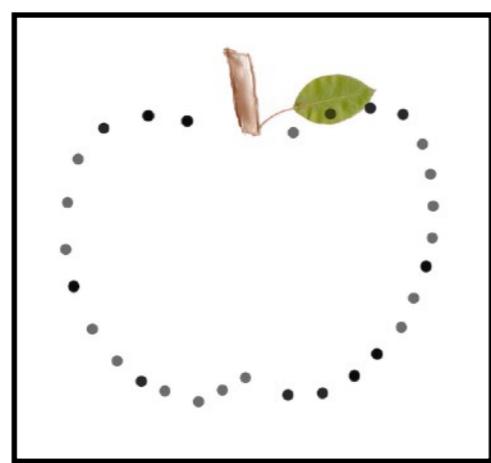
Principio de proximidad



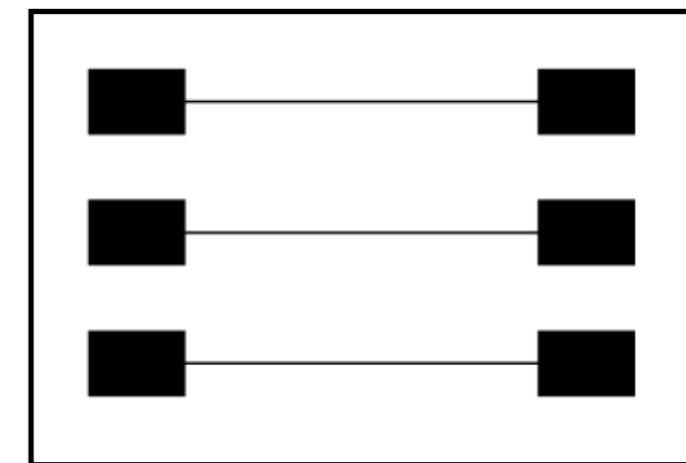
Principio de similitud



Principio de cercado



Principio de continuidad



Principio de conexión

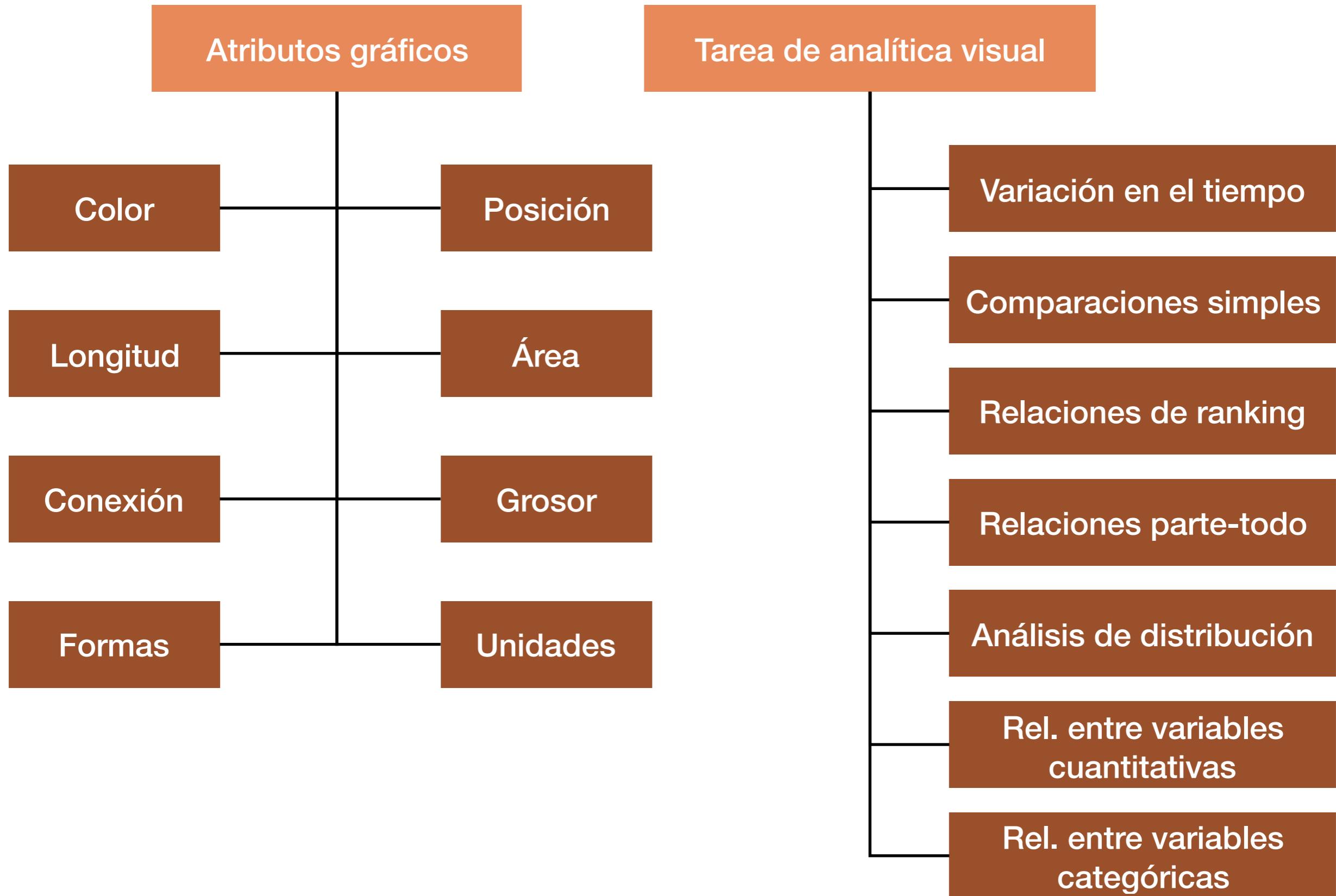
3 - Percepción post-atentiva

- Establecer un vínculo entre lo que se ve y lo que se conoce.
- Vinculada a la cognición y la memoria.
- Tener cuidado con:
 - Lenguaje textual
 - Lenguaje iconográfico

Recomendaciones para el proceso percepción visual

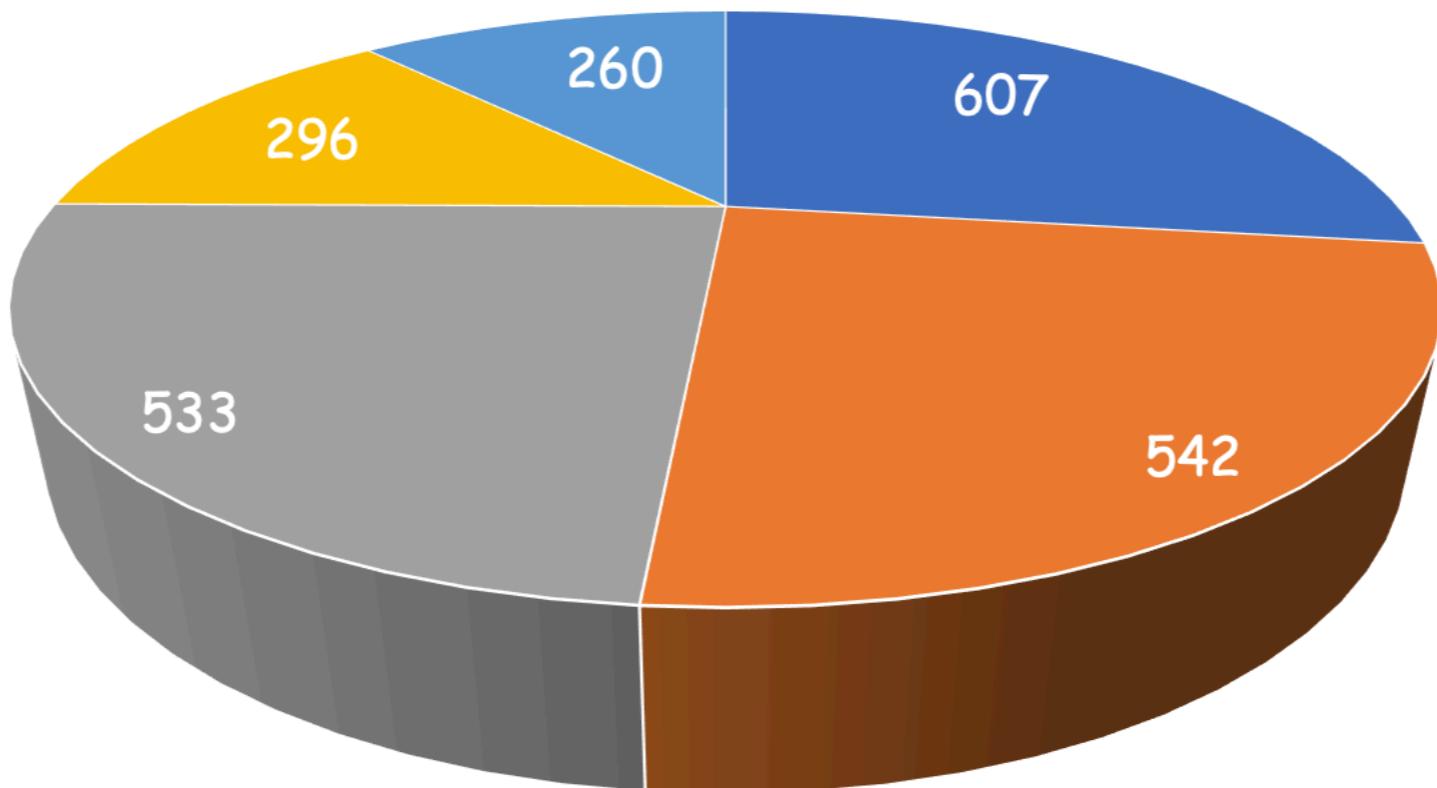
- Enfatizar
- Organizar
- Hacer reconocible

Codificación gráfica de datos



Ejemplo de lo que NO hay que hacer

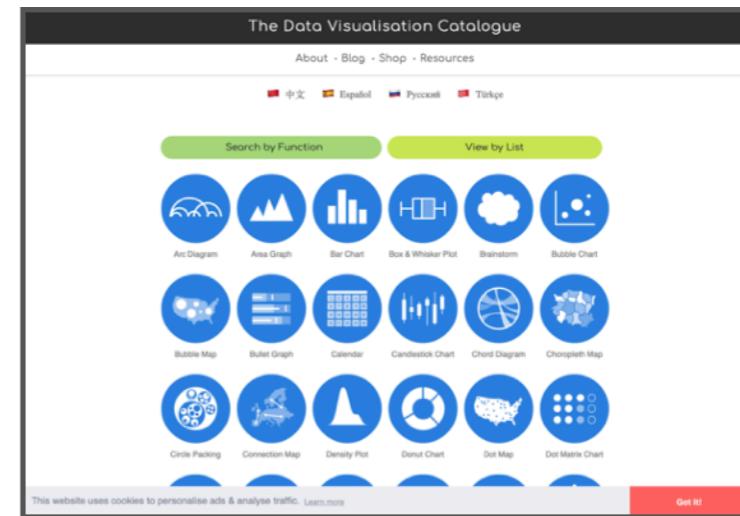
- Gráficos 3D
- Comic Sans
- Diagrama circular (o de tarta/pastel)



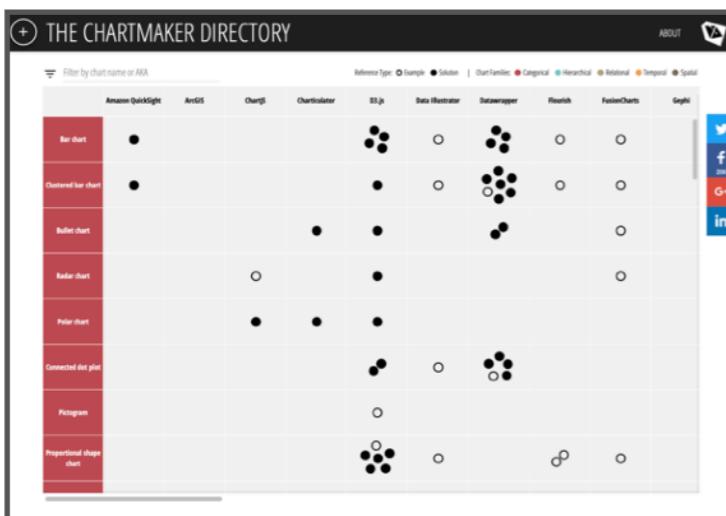
Catálogos de visualización



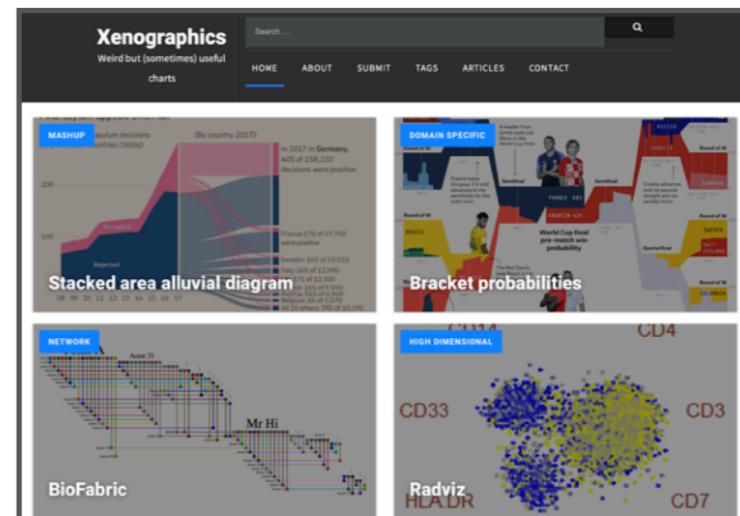
datavizproject.com



datavizcatalogue.com



[chartmaker directory](http://chartmaker.directory)



[Xenographics](http://xenographics.com)

Deviation

Correlation

Ranking

Distribution

Change over Time

Magnitude

Part-to-whole

Spatial

Flow

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population Categorical distribution, revealing inequality

Give emphasis to changing trends. These can be short (intra-day) movements or extended series (months/years/centuries). Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine-grained details). If these show a location number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information, relationship graphs.

Diverging bar
 A simple standard bar chart that can handle both negative and positive magnitude values.

Scatterplot
 The standard way to show the relationship between two continuous variables, each of which has its own axis.

Ordered bar
 Standard bar charts the ranks of values much more easily when sorted into order.

Histogram
 The standard way to show a distribution - keep the gaps between columns small to highlight the shape of the data.

Line
 The standard way to show changing time series. If data are irregular, consider markers to represent data points.

Column
 The standard way to compare the size of things. Must always start at 0 on the axis.

Stacked column/bar
 A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Basic choropleth (rate/ratio)
 The standard approach for putting data on a map. It should always be rates rather than totals and use a sensible base geography.

Sankey
 Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Diverging stacked bar
 Perfect for presenting survey results which involve sentiment (eg disagree/heutral/agree).

Column + line timeline
 A good way of showing the relationship between an amount (columns) and a rate (line).

Ordered column
 See above.

Dot plot
 A simple way of showing the change or range (min/max) of data across multiple categories.

Column
 Columns work well for showing change over time - but usually best with only one series of data at a time.

Bar
 See above. Good when the data are not time series and labels have long category names.

Marimekko
 A good way of showing the size and proportion of data at the same time - as long as the data are not too complicated.

Proportional symbol (count/magnitude)
 Use for totals rather than rates - be wary that small differences in data will be hard to see.

Waterfall
 Used to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Spine
 Spins a single value into two contrasting components (eg male/female).

Connected scatterplot
 Usually used to show how the relationship between 2 variables has changed over time.

Ordered proportional symbol
 Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot
 Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Column + line timeline
 A good way of showing the relationship over time between an amount (columns) and a rate (line).

Paired column
 As per standard column but allows for multiple values. Can become tricky to read with more than 2 series.

Pie
 A common way of showing part-to-whole data - but be aware that it's difficult to accurately compare the size of the segments.

Flow map
 For showing unambiguous movement across a map.

Chord
 A complex but powerful diagram which can illustrate 2-way flows (and net wins/losses) in a matrix.

Surplus/deficit filled line
 The shaded area of these charts allows a balance to be shown either against a baseline or between two series.

Bubble
 Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

Dot strip plot
 Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Barcode plot
 Like dot strip plots, good for displaying all the data in a table; they work best when highlighting individual values.

Slope
 Good for showing changing totals as long as the data can be simplified into 2 or 3 points without missing a key part of the story.

Paired bar
 See above.

Donut
 Similar to a pie chart - but the centre can be a good way of making space to include more information about the data (eg total).

Treemap
 Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Equalized cartogram
 Converting each unit on a map to a regular and equal-sized shape - good for representing voting regions with equal values.

XY heatmap
 A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Slope
 Perfect for showing how ranks have changed over time or vary between categories.

Lollipop
 Lollipops draw more attention to the data value than standard bar/columns and can also show rank and value effectively.

Violin plot
 Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple averages).

Area chart
 Use with care - these are good at showing changes to total, but seeing change in components can be very difficult.

Marimekko
 A good way of showing the size and proportion of data at the same time - as long as the data are not too complicated.

Proportional symbol
 Use when there are big differences between values and/or seeing fine differences between data is not so important.

Voronoi
 A way of turning points into areas - any point within each area is closer to the central point than any other centroid.

Scaled cartogram (value)
 Stretching and shrinking a map so that each area is sized according to a particular value.

Bump
 Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Boxplot
 Summarise multiple distributions by showing the median (centre) and range of the data.

Population pyramid
 A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Cumulative curve
 A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

Fan chart (projection)
 Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Iratype (pictogram)
 Excellent solution in some instances - use only with whole numbers (as it's not nice off an arm to represent a decimal).

Arc
 A hemisphere, often used for visualising parliamentary composition by number of seats.

Dot density
 Used to show the location of individual events/locations - make sure to annotate any patterns the reader should see.

Hear map
 Grid-based data values mapped with an intensity colour scale. As choropleth map - but not snapped to an admin/political unit.

XY heatmap
 A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Slope
 Perfect for showing how ranks have changed over time or vary between categories.

Lollipop
 Lollipops draw more attention to the data value than standard bar/columns - does not have to start at zero (but preferable).

Connected scatterplot
 A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap
 A great way of showing temporal patterns (daily/weekly/monthly) - at the expense of showing precision in quantity.

Radar
 A space-efficient way of showing value of multiple variables - but make sure they are organised in a way that makes sense to the reader.

Gridplot
 Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Parallel coordinates
 An alternative to radar charts - again, the arrangement of the variables is important. Usually benefits from highlighting values.

Venn
 Generally only used for schematic representation.

Bump
 Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Frequency polygons
 For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm
 Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Rest with medium-sized datasets.

Circle timeline
 Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline
 Projects time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Waterfall
 Good for showing part-to-whole relationships where some of the components are negative.

Grouped symbol
 An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Seismogram
 Another alternative to the circle timeline for showing series where there are big variations in the data.

FT

Designing with data

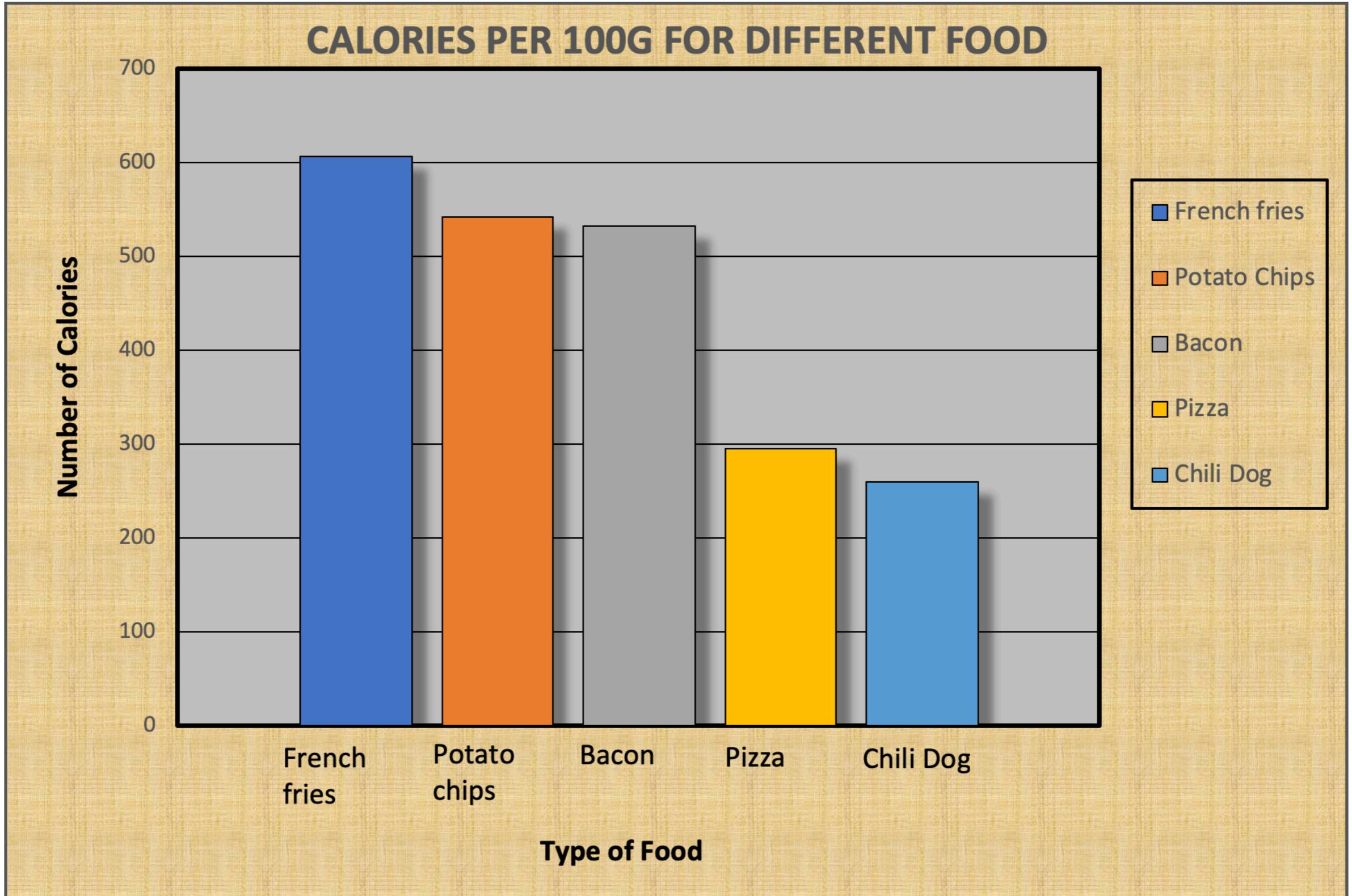
There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.



ft.com/vocabulary

Ejemplo visualización

¿Cuántas calorías tiene el Bacon con respecto al resto?



Data-ink ratio

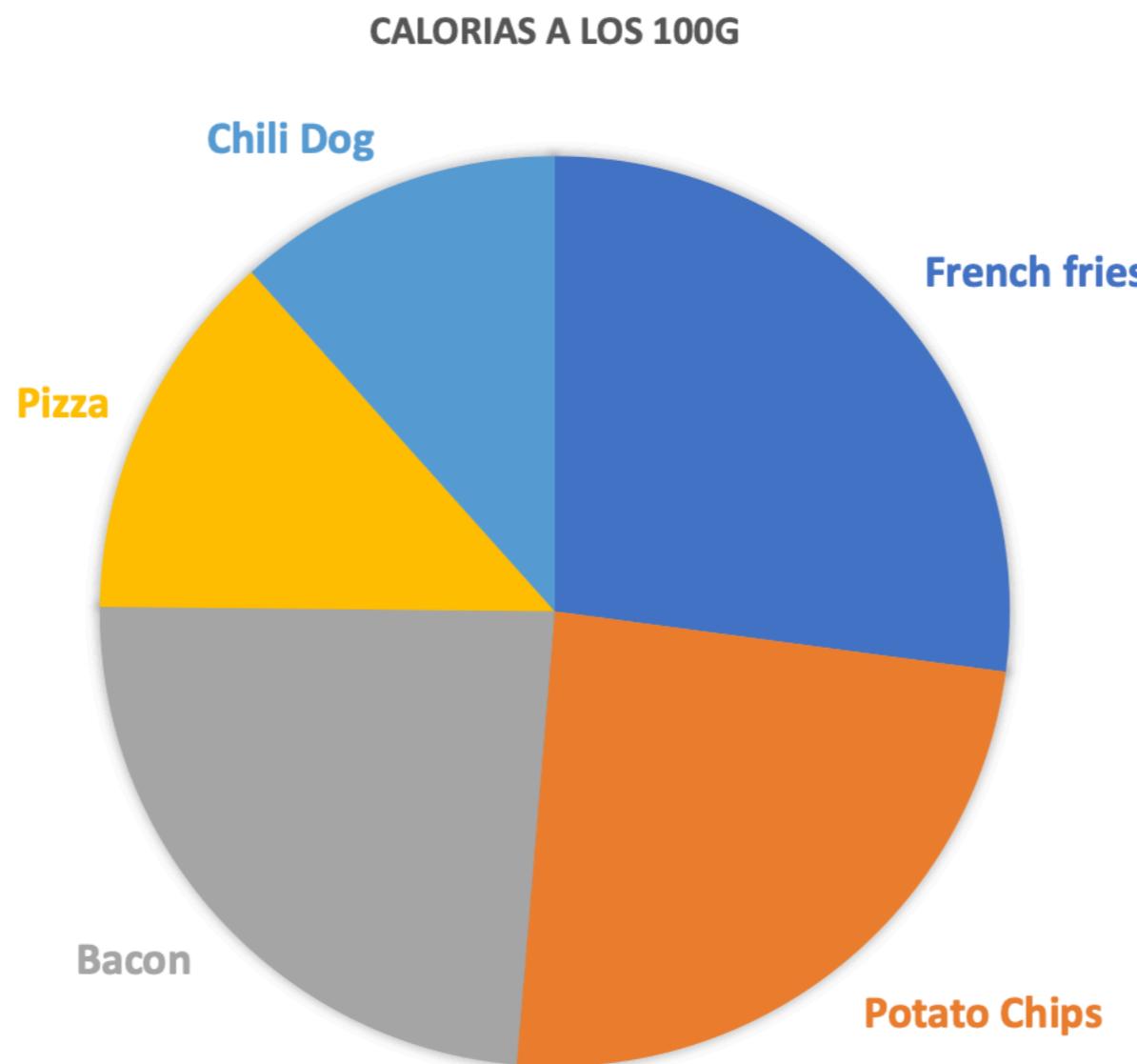
Data-ink ratio =
$$\frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

= proportion of a graphic's ink devoted to the
non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased

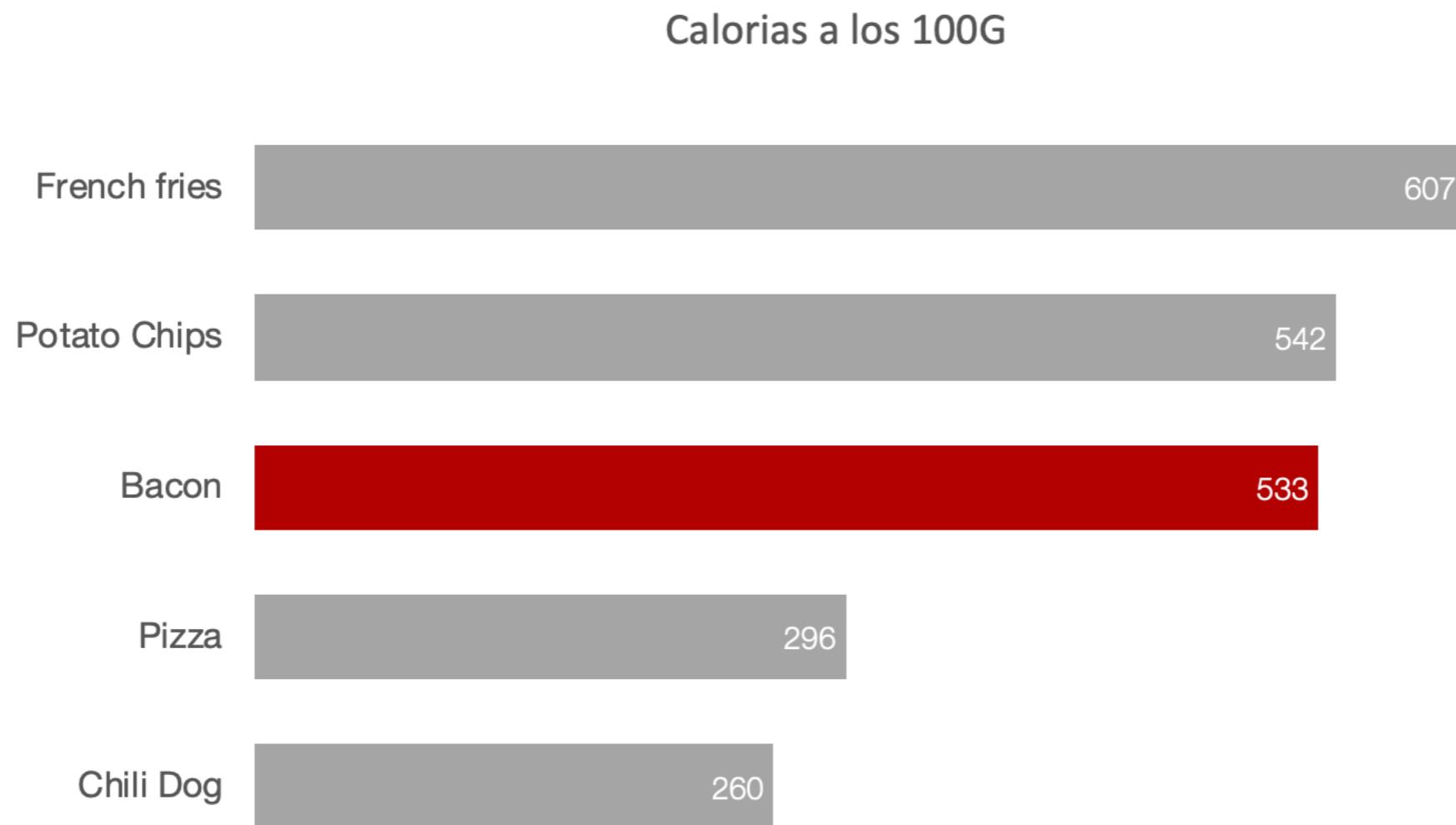
Possible propuesta

¿Cuántas calorías tiene el Bacon con respecto al resto?



Propuesta mejora

¿Cuántas calorías tiene el Bacon con respecto al resto?

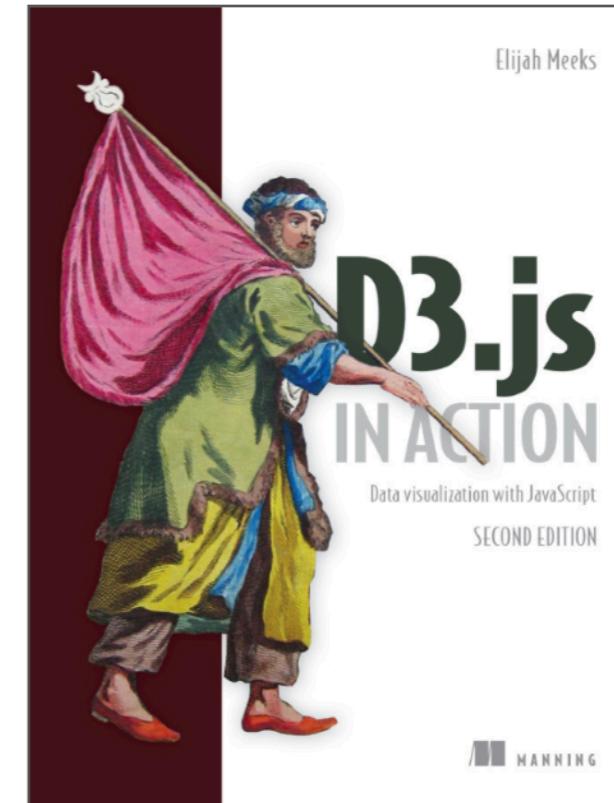
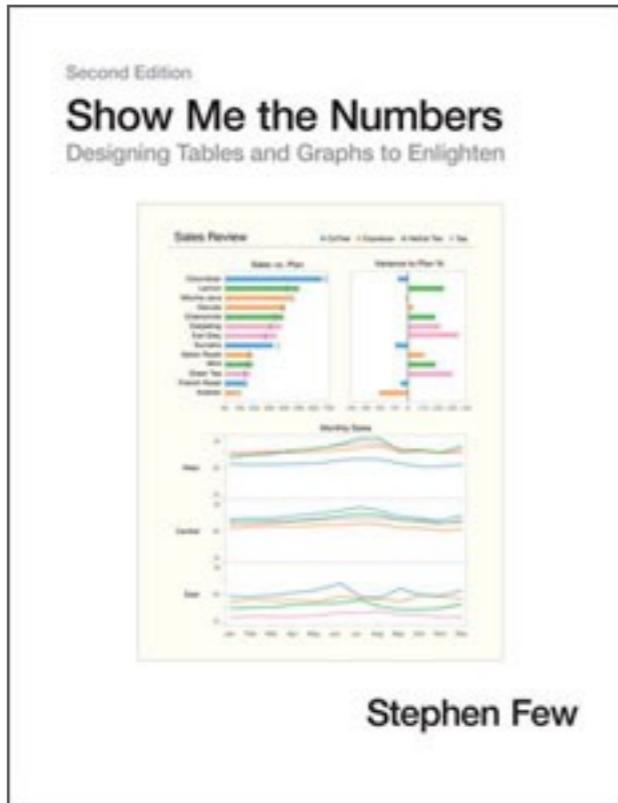


Ejemplo visualización con



- Dataset preparado
- Diagrama de líneas
- Uso de escalas (rango y dominio)
- Creando ejes
- <https://github.com/juananthony/D3-Meetup-examples>
<https://github.com/juananthony/D3-Meetup-examples>

Lecturas recomendadas



Few, S. (2012). *Show me the Numbers*.
Burlingame, California: Analytics Press.

Meeks, E. (2017). *D3.js in Action*.
Second Edition. Manning Publications.

¡Gracias!

