

DigitalHouse >
Coding School

DATA SCIENCE

MÓDULO 2

Visualización

Agosto de 2017

OBJETIVOS DE LA CLASE

1 Recursos visuales para distintos tipos de datos

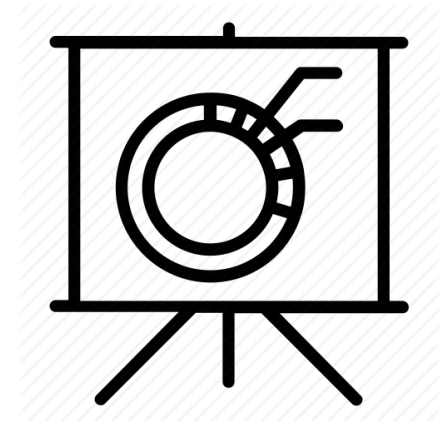
2 Aplicar conceptos del Módulo 2 y Seaborn



PRÁCTICA_GUIADA_1_Visualizacion_Seaborn.ipynb
(Reemplazo de valores, pivot_table y dummies)



PRACTICA_GUIADA_2_Visualizacion_Seaborn.ipynb



Recursos visuales



THE GRAPHIC CONTINUUM

The Graphic Continuum shows several ways that data can be illustrated individually or combined to show relationships. Use of various shapes, chart types, and colors can help identify patterns, tell stories, and reveal relationships between different sets and types of data. Bar charts, or histograms, for example, can illustrate a distribution of data over time, but they also can show categorical or geographic differences. Scatterplots can illustrate data from a single instance or for a period, but they also can be used to identify a distribution around a mean.

This set of charts does not constitute an exhaustive list, nor do the connections represent every possible pathway for linking data and ideas. Instead, the Graphic Continuum identifies some presentation methods, and it illustrates some of the connections that can bind different representations together. The six groups do not define all possibilities: Many other useful, overlapping data types and visualization techniques are possible.

This chart can guide graphic choices, but your imagination can lead the way to other effective ways to present data.

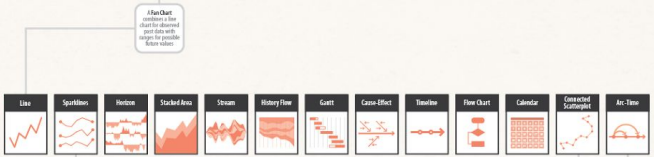
DISTRIBUTION

Graphical representations of the distribution of data



TIME

Track changes over time



COMPARING CATEGORIES

Compare values across categories



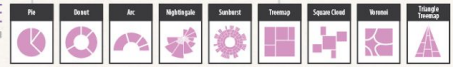
GEOSPATIAL

Relate data to its geography



PART-TO-WHOLE

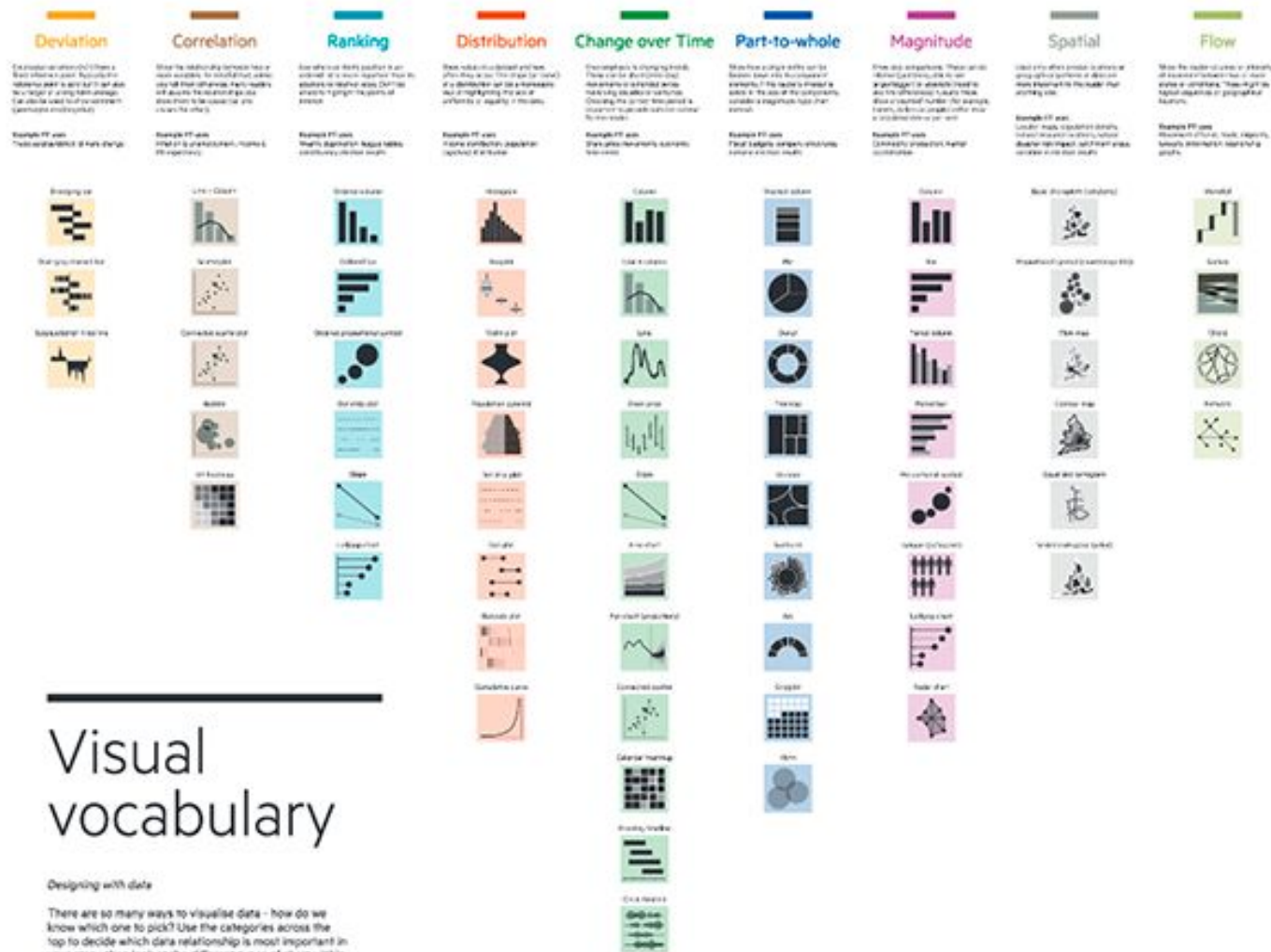
Visualizations that relate the part of a variable to its total



RELATIONSHIP

Illustrates correlations or relationships between variables



[Ver pdf](#)[Visitar](#)

Visual vocabulary

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.

Inspired by the Creative Commons by-nc-sa/ShareAlike license

Correlation

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)

Examples of use

Inflation & unemployment, income & life expectancy

Chart types

scatterplot



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

line-column



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

scatterplot-connected



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

Bubble



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

XY-heatmap



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

Distribution

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

Examples of use

Income distribution, population (age/sex) distribution

Chart types

histogram



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

boxplot



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

violin



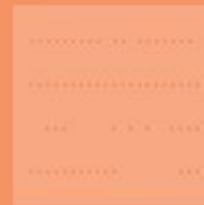
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

population-pyramis



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

dot-plot-strip



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

dot-plot



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

Change v Time

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

Examples of use

Share price movements, economic time series

Chart types

line



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

column-timeline



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

column-line-timeline



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

stock-price



Usually focused on day-to-day activity, these charts show opening/closing and hi/low points of each day

slope



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

area



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

Part to whole

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

Examples of use

Fiscal budgets, company structures, national election results

Chart types

column-
stacked



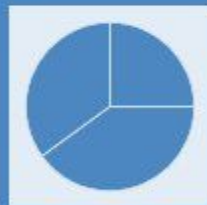
A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

bar-stacked-
proportional



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

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.

doughnut



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

Voronoi



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

Referencias



