



Data Visualization

La visualización de los datos es tan importante como su recopilación y su análisis.

Autor: Massin Laaouaj Madrouni

Medium Post:

<https://medium.com/@massin.laaouaj/data-visualization-73819cb5c212>

2023/24

Para mí, la visualización de los datos es como contar un viaje que tuviste, o un festival de música al que fuiste.

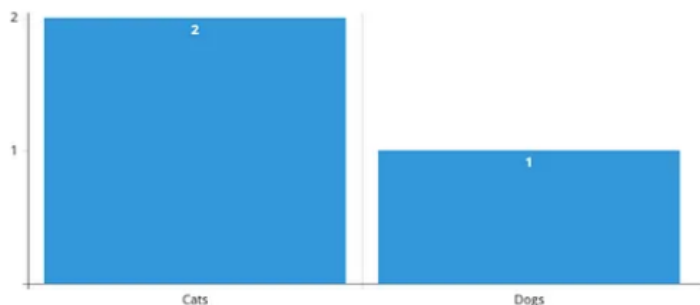
Al igual que para una buena comunicación, se ha de saber expresar bien las ideas y saber escuchar de forma activa a las otras partes, y adaptar el mensaje al público objetivo utilizando un lenguaje claro y adaptativo para asegurar que el mensaje se transmita de forma efectiva (ya que no hablas de la misma forma con tu amigo, que con tu jefe).

También pasa con a la hora de crear una visualización de los datos, ya que sin una buena comunicación de los datos puede ser una herramienta poderosa para entender que reflejan, el impacto, ...

Vale... pero porque me ha de importar?

Bien, por ejemplo queremos hacer un recuento y visualizar la gente que ha votado gatos o perros, como su mascota preferida.

Con este ejemplo tan sencillo podríamos pensar de representarlo de muchas formas distintas, pero nos acotamos a dos únicas opciones la grafica “Donuts” y la de “barras”.



Podemos ver de una vista cuál es la que transmite mejor los votos: Donuts

Pero porque...¿?

Fijate bien,

los colores el Donuts representa muy bien cuál es cuál,

la información en el de donuts te presenta el porcentaje, el número total de votos, y el título de lo que estamos viendo

Ahora ves la diferencia del porqué representar bien la información es importante.

Contexto

Cuando empezamos a pensar en la visualización de nuestra tarea, necesitamos definir su objetivo principal o el contexto de la visualización. Hay dos casos de uso importantes para la creación de gráficos: análisis exploratorio y explicativo.

- **Análisis exploratorio;** es el hecho de coger un papel y lápiz, y empezar a entender el análisis, es decir, una conversa interior para al final conseguir explicar los datos de forma efectiva.
- **Análisis explicativo;** cuando ya has conversado interiormente y sabes cómo debes explicarlo (an. exploratorio), ahora toca saber a quién te diriges:
¿Quién es mi audiencia?
¿Qué quiero lograr?
¿Estoy mostrando demasiado?

En este análisis debes decidir colores, adaptación, que grafico utilizar,...

Psicología del ser humano

Según el siguiente [estudio](#), muestra que tan bien entienden el mensaje que se representa en los diferentes gráficos, concluyendo el siguiente orden (de más entendedor a menos):

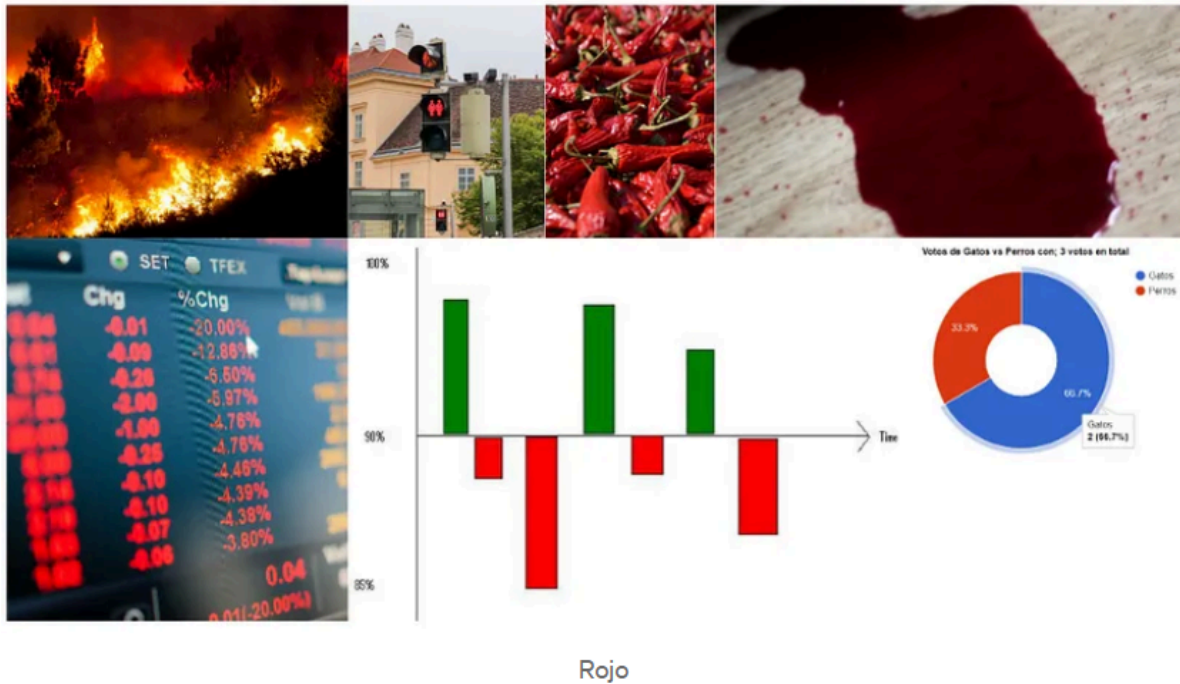
- gráficos de posición — por ejemplo; scatter plot
- gráficos de longitud— por ejemplo; bar chart
- gráficos de dirección — por ejemplo; line chart
- gráficos de ángulo — por ejemplo; pie chart
- gráficos de área — por ejemplo; bubble chart
- gráficos de volumen — por ejemplo; 3D chart
- gráficos de tono y saturación del color — por ejemplo; heat map

El objetivo principal de la visualización es transmitir bien la información y centrarnos en nuestra audiencia y en cómo perciben el mensaje. Por eso nos interesa adaptar.

Psicología del color

La selección de los colores en la visualización de los datos se utiliza principalmente para categorizar y distinguir los grupos de datos.

La mente relaciona/asocia los colores con [situaciones reales o eventos](#) de la naturaleza, por ejemplo, el uso de colores cálidos como el rojo o el naranja representa una relación con el calor o la energía, o el uso de colores fríos como el azul o el verde representa relaciones con la calma o la naturaleza.



Los colores son una herramienta muy útil y potente a utilizar, para representar mejor y transmitir mejor la información, a sí mismo cada color de forma inconsciente lo asociamos con situaciones y conseguimos que el público objetivo reciba mejor el mensaje.

Sisi... pero qué hay de malo en ponerlo todo en un color que me guste a mí?

Te pongo un ejemplo, imagina que tengamos un conjunto de números aleatorios y queremos identificarlos en un conjunto.

La parte izquierda lo pondremos sin color (que se busquen la vida, saben contar...), y en la parte derecha con color:

Cuántos 7's hay ?

Cuántos 4's hay ?

56146807431345038465	56146807431345038465
3527384869796858747	3527384869796858747
38272427175768685838	38272427175768685838
37802737475676878483	37802737475676878483
8382822737475718452	8382822737475718452
6092642969804926264	6092642969804926264
86984516471395767463	86984516471395767463
631802430715697017860	631802430715697017860

Vaya... ahora se ve cómo de importante es el color!!

Los datos no únicamente se encuentran representados en gráficos, también en forma de texto, con la combinación de colores y el tamaño de letras podemos destacar/puntualizar datos, por ejemplo:

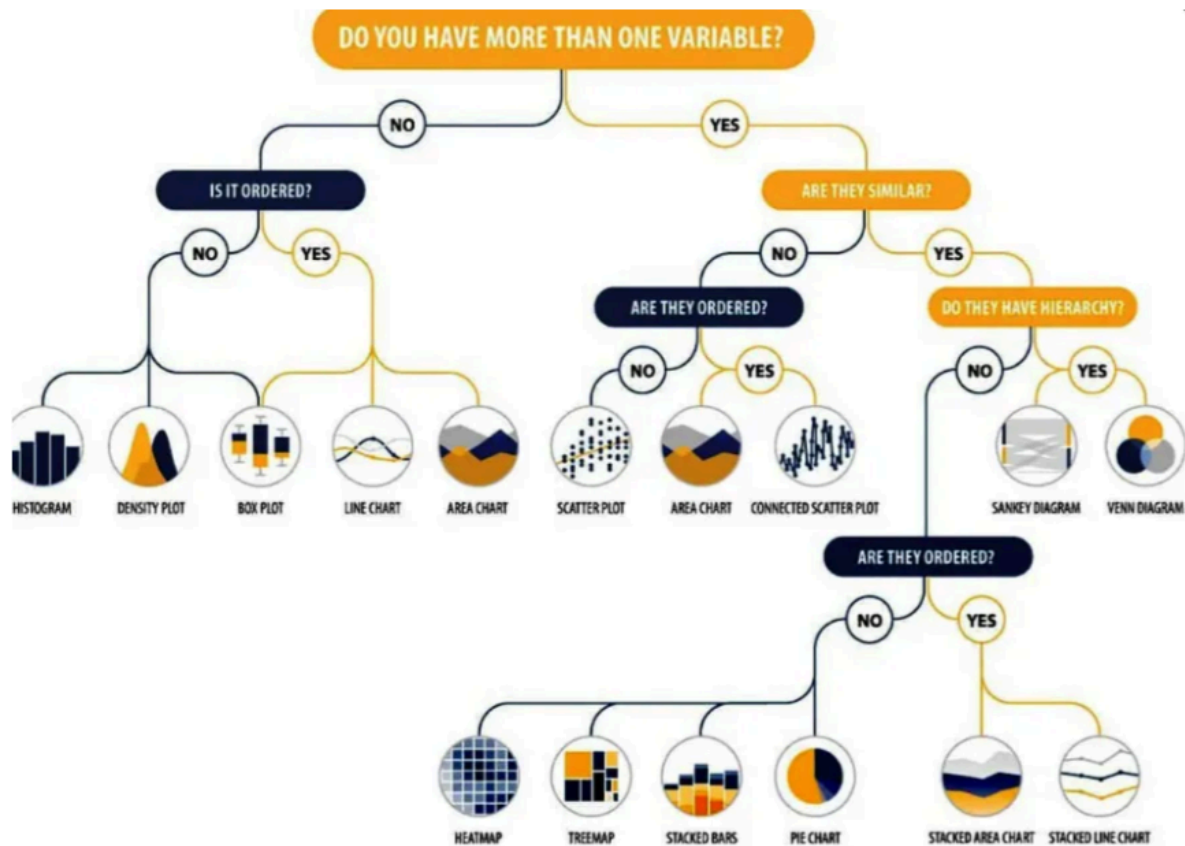
The average annual GMV per customer in Switzerland is **\$2,341**
while for other regions, this value is just **\$1,158**

¿Qué herramientas existen?

- Google Charts
- Oracle Analysis
- Power BI
- Tableau
- ChartJS
- ...

Cheatsheet

Esta chuleta la puedes imprimir, o puede visitar esta [página](#):



Data Visualization Hierarchical Structure

COLORS IN DATA VISUALIZATION

RULES

INTUITIVENESS

Use intuitive colors. When choosing them, consider what associations do they evoke. If possible, use colors that audience will associate with your data anyway.



MODERATION

Use colors in moderation. For a simple dataset, a single color is preferable. Use color as a strategic tool to highlight the important parts of your visual.



CONSISTENCY

Use colors consistently. Change colors if you want your audience to feel the change for the specific reason, but never simply for the sake of novelty.



CLARITY

Use colors to make the data easier to read. Make sure your audience will be able to distinguish between the items shown in the visualization.



CLASSIFICATION

Don't use a gradient color palette for categories. And the other way round - different colors for same measurement.



EXPLAINABILITY

Make sure to explain to your audience what exactly used colors mean. Remember to create a color key.



COLOR SCHEMES

Monochromatic - the simplest formula for harmony is monochromatic. Consists of different shades of one hue. Not a good choice if we want to highlight something.



Analogous - this scheme is composed of colors that are next to each other on the wheel. Usually they match up pretty well, making elegant and clear look.



Complementary - uses two colors which are opposite on the color wheel. With saturated colors makes very vibrant look. Try to tone down colors to avoid overvibrance, by adjusting saturation and lightness/darkness. Do not use with text with saturated colors.



Triadic - uses three colors that evenly spaced on the color wheel. Makes that none of colors is dominant and quite vibrant look.



Split-Complementary - variation of complementary scheme. Uses base color and two adjacent to its complementary color. Often this scheme is more pleasant to the eye than usual complementary scheme.



Tetradic - this scheme consists of four colors, two of them are complementary to other two. Choosing one color as dominant and the rest as accents, gives the best result.



Remember! Do not stick strictly to colors imposed by a scheme. These patterns are just starting points, you can create your own variations based on schemes above. Check also: paletton.com

PRO TIPS

The color grey is the most important color in data visualization.

The use of color should always be an intentional decision. Never let your tool make this important decision for you!

After creating your visualization, close your eyes and then look back at it, taking note of where your eyes are drawn first. Is it where you want your audience to focus?

When picking colors consider the connotations colors have in other cultures. You can check: informationisbeautiful.net/visualizations/colors-in-cultures

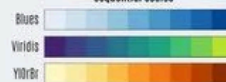
Remember about color deficiency issues (color blindness). You can check: projects.susielu.com/viz-palette

COLOR PALETTES

QUANTITATIVE DATA - SEQUENTIAL OR DIVERGING COLORS

Color is used to show variations in the data. The palette contains a sequence of colors that clearly indicate which values are larger or smaller than which other ones (sequential scale). It can also visualize the deviation of data values in one of two directions relative to a neutral midpoint (diverging scale). Diverging scale can be viewed as two merged sequential scales.

Sequential scales



Diverging scales



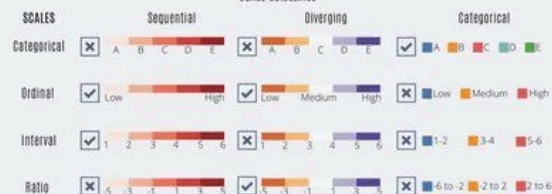
CATEGORICAL DATA - QUALITATIVE COLORS

Color is used to separate areas into distinct categories. The palette should consist of colors as distinct from one another as possible. The maximum number of categories that can be displayed is about 12 (practically speaking, probably fewer).



All examples are available in Seaborn library. Check also: matplotlib.org/3.1.1/whatsnew/

USAGE GUIDELINES



Authors:

Mateusz Krzyżniński
Paweł Wojciechowski
Artur Ziółkowski



Warsaw University of Technology
Faculty of Mathematics and Information Science