





Unidad 4. Herramientas para la visualización de datos

Docente: Mauricio Morán

Septiembre - Octubre 2023

Un poquito sobre mi

- Soy de La Calera, Cba pero hace nueve años que vivo en Bariloche.
- Estudié en la UNC y luego en el IB.
- Investigador de CONICET. Hago investigación en materiales.
- Cuatro cinco años de experiencia con Python.
- Estudiante de ciencia de datos.
- Hincha de Talleres.
- Caminatas, bicicleta, esquí, leer y videojuegos.

Programa tentativo de la Unidad 4:

- Clase 1: Introducción y conceptos fundamentales. Tipos de datos. Tipos de gráficos. Buenas prácticas para la visualización de datos. Principales paquetes Python. Introducción a Matplotlib.
- Clase 2: Visualización de datos numéricos: datos, mapas, gráficos, uso de glyphs, gráficos apilados. Matplotlib-Pandas y Seaborn.
- Clase 3: Visualización de datos no numéricos: gráficos, redes, gráficos de componente-principal, árboles, escalado multidimensional. Presentación de scikit-learn: escalado - estandarización - normalización, PCA, árboles de decisión para clasificación y regresión.
- Clase 4: Reportes, dashboards y otros recursos. Comunicación de datos. Business Intelligence. Storytelling. Dashboards con Plotly.
- Tutoria: Desarrollo de actividad práctica.

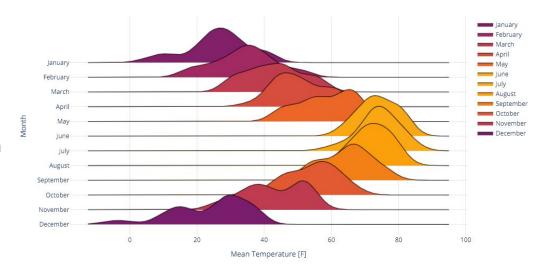
Clase 1: Introducción y conceptos fundamentales.

Introducción

Visualización de datos

Se refiere a la presentación de datos en un formato ilustrado o gráfico.

- Facilitar interpretación.
- Captar conceptos.
- Identificar patrones.
- Comunicación.



¿Para qué sirven?



Panorama general. Identificar un contexto más amplio en los datos. Percibir tendencias y patrones en los datos.

Tomar decisiones. Las cifras concretas y los conocimientos tangibles permiten que las decisiones cuentan con el respaldo de los datos.

Seguir las tendencias. Realizar seguimiento del progreso, detectar tendencias y utilizar los conocimientos para tomar decisiones estratégicas fundamentadas.

https://powerbi.microsoft.com/es-es/data-visualization/

Buenas prácticas

Una buena visualización debe ser:

Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.

https://www.linkedin.com/pulse/data-visu alization-5-key-best-practices-better-analy tics/

Veraz: debe basarse en datos exactos. Debe ser honesta y transparente, y representar los datos con claridad y precisión.

Funcional: debe diseñarse con un objetivo claro. Debe transmitir su mensaje con eficacia, y su público debe ser capaz de entenderlo con facilidad.

Atractiva: un gráfico visualmente atractivo capta la atención y hace más interesante la información. Sin embargo, la belleza nunca debe ir en detrimento de la funcionalidad o la veracidad.

Una buena visualización debe ser:

Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.

https://www.linkedin.com/pulse/data-visu alization-5-key-best-practices-better-analytics/

Perspicacia: no sólo debe presentar los datos con claridad, sino que también debe proporcionar información que vaya más allá de las cifras.

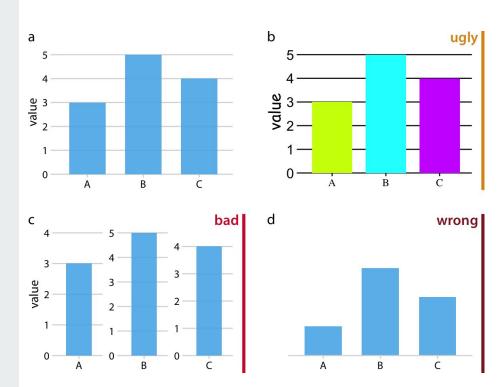
Esclarecedora: proporcionando a la audiencia una visión nueva o inesperada de los datos. Para ello hay que ir más allá de lo obvio y explorar distintos ángulos y perspectivas para revelar ideas ocultas.

El bueno, el malo, el feo (y el erróneo)

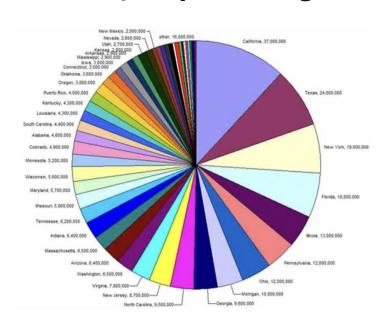


Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures.

O'Reilly Media. https://clauswilke.com/dataviz/

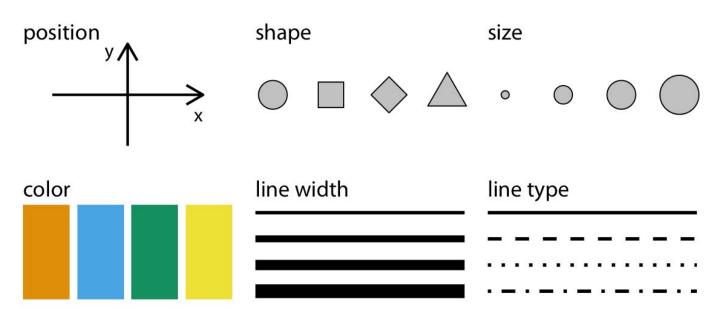


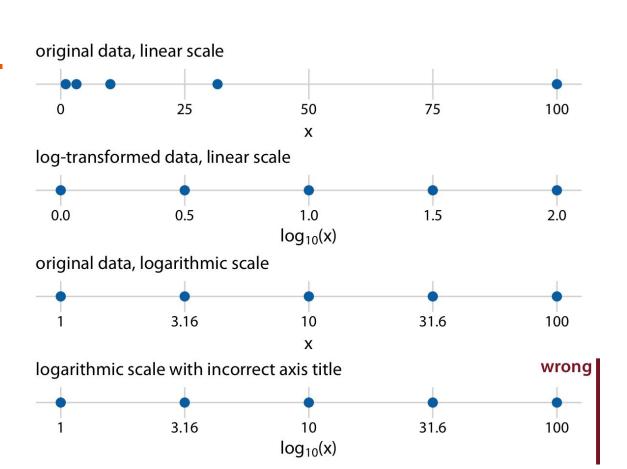
Más ejemplos de gráficos malos/feos/erróneos:

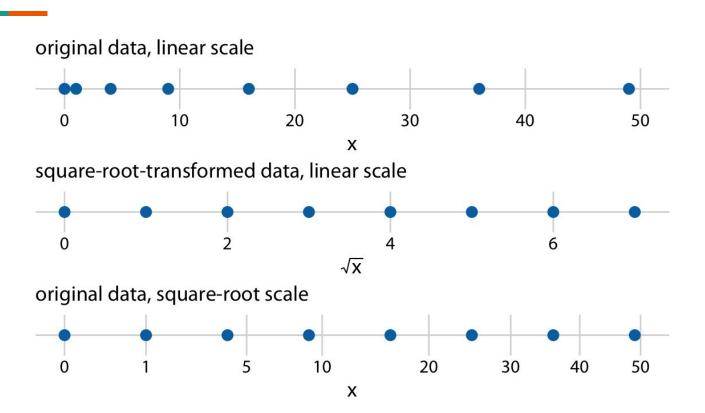




¿Qué podemos modificar? #aesthetics







Remove to improve (the data-ink ratio)

Antes, repasemos tipos de datos/variables:



Tipos de gráficos



Emphasise variations CHZ from a food reference point Typically for reference point a cere bur it can also be a forget or a long-from annuage. Can also be used to allow sentment (post variation).

Example FT uses Trade sarplaside (c): climate change





Correlation

Store the reset analytic between two or more vertebles. De mindfull that, unest you tell them otherwise, many moders will assume the habitary size should have to be coupalitie, one causes the term to be coupalitie.



the patterns between 2 categories of data, less effective at phowing fine differences in emounts.

Ranking

Use where an item's position in an ordered list is more important than its abook more relative value. Don't be aftered to highlight the points of innerest

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

....

Change over Time

Example FT uses Share price movements, economic time series, second changes in a merber

A good way of strong changing can for his veryibles whenever filter is a relatively claim pattern of in strongers.

Latipos charts delay nore after too to the data value has stendard benhalum — doss nor have to shart at zero duit preferable).

Magnitude

Show size comparisons: These can be infloring (set being other to set larger/stager) or about, to know for set fine of freezes. Usually meas show a coertain number (for exemple, barreb, collect or secondo) mithor than a chicalitation specially mithor than a chicalitation according to the or and control of the or and can be called the or and can

The standard way to compare the size of things Must always after at Dion the acco.

Stow hore a single entity can be broke down into its component elements. If the resider's interest is solely in the size of the components, consider a magnitude-type charminated.

Part-to-whole

Example FT uses
"scal budgets, company shuctures,
refored metition results.

Spatial

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

Flow

Example FT uses
Movement of funds, trade originals, larger halfor; relationship graphs.

Example FT uses
Population dentity, satural recourse locations, natural classifier discharged, cardinated stees, seciation in election results.

The standard approach for parting data on a map – should always be rated unto a condition to the probability of one or condition to the condition of the condit

https://python-graph-gallery.com/

https://www.data-to-viz.com/

https://www.dataviz-inspiration.com/

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 meaninoful data visualisations.

FT graphs, Controlly Circ Complete British La Fouring Collect Participation Congressioners Paid Pt Collect Notice States States (Marine States States



Distribution

Short-values in a dataset and hort other they occur. The sliepe (or falses) of a citation from the bight lighting the lock of antiforming or causily in the data.

Good far showing inclvide all values in a clistriberion, can be a problem when the many dots have the

iSon muchísimos!

Vamos a ver unos pocos, ¡pero de los más usados! Veremos gráficos de:

- Línea.
- Puntos.
- Barras.
- Torta.
- Histogramas.
- Cajas, violín
- Mapas de calor.
- Geoespaciales.
- etc.

Lo importante no es saberlos a todos sino saber que tipo de datos se ajustan mejor a nuestros datos y a lo que queremos presentar.

Paquetes de Python para visualización

matplatlib

seaborn



bokeh

iA programar!

