



FACULTAD  
DE CIENCIAS  
ECONÓMICAS



Universidad  
Nacional  
de Córdoba

# 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.
- Hinchador 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.

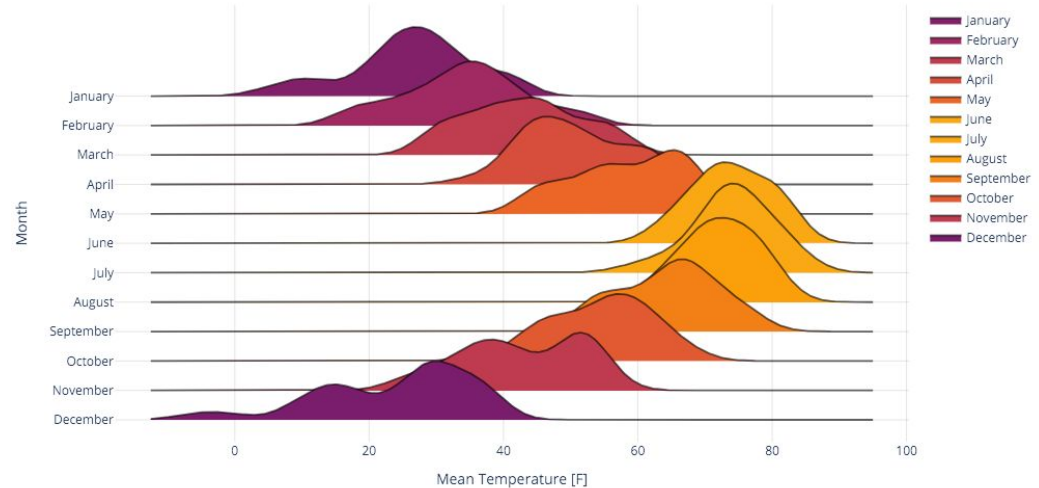
# Introducción

---

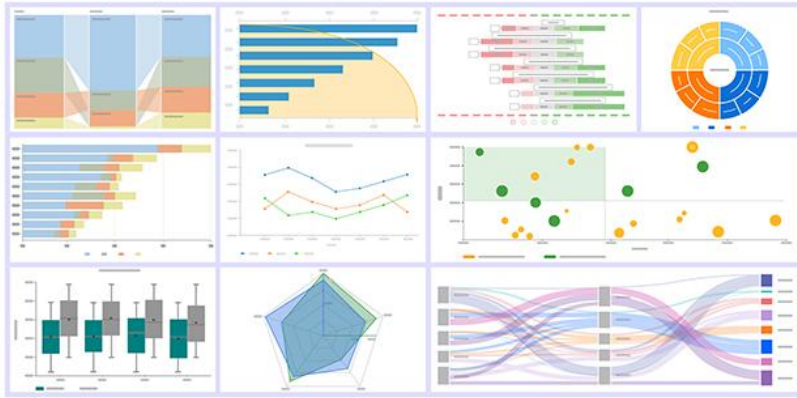
# Visualización de datos

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 cuenten con el respaldo de los datos.

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

# 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-visualization-5-key-best-practices-better-analytics/>

**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-visualization-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.

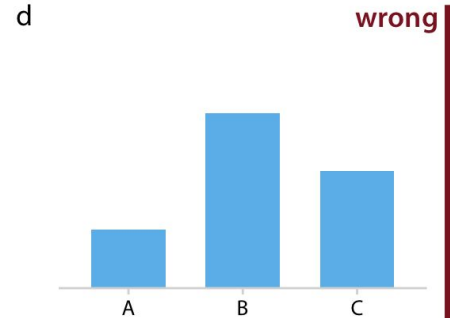
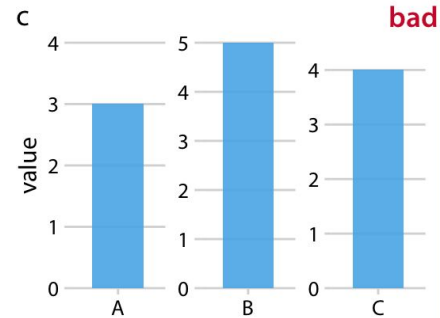
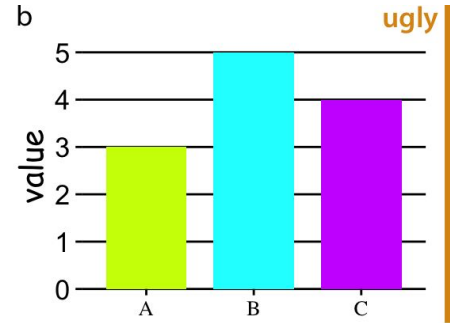
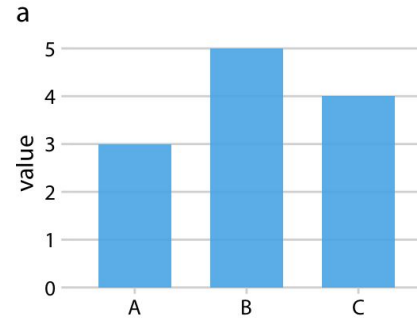
# El bueno, el malo, el feo (y el erróneo)



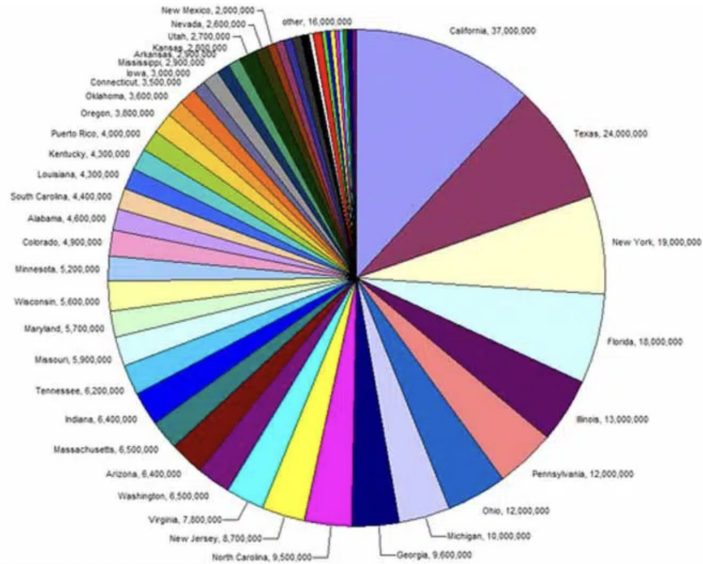
Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures.

O'Reilly Media.

<https://clauswilke.com/dataviz/>

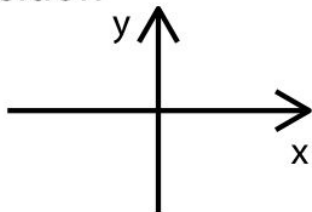


## Más ejemplos de gráficos malos/feos/erróneos:

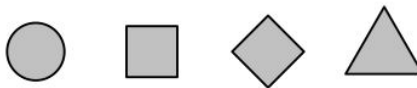


## ¿Qué podemos modificar? #aesthetics

position



shape



size



color



line width

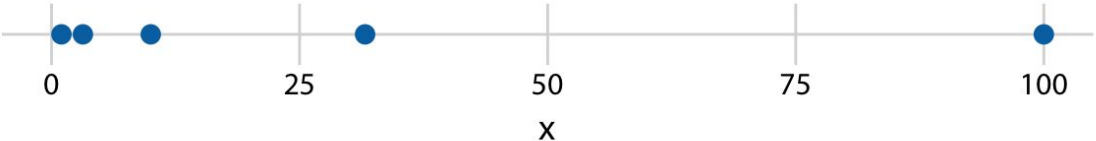


line type

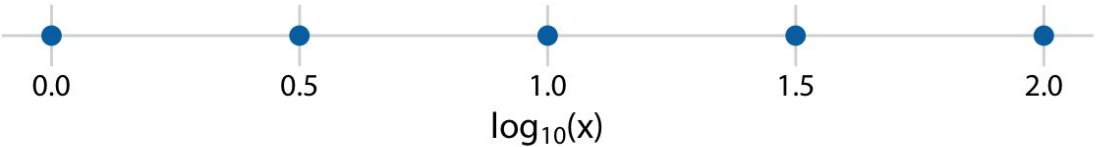




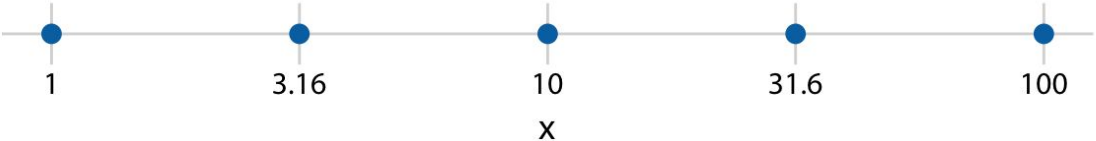
original data, linear scale



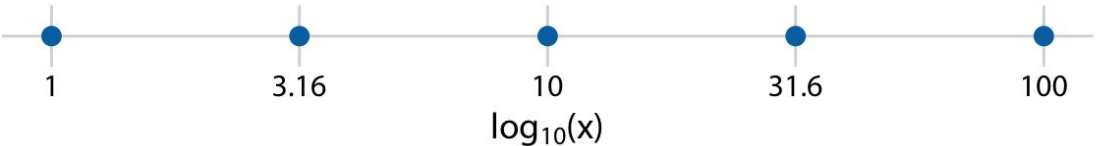
log-transformed data, linear scale



original data, logarithmic scale

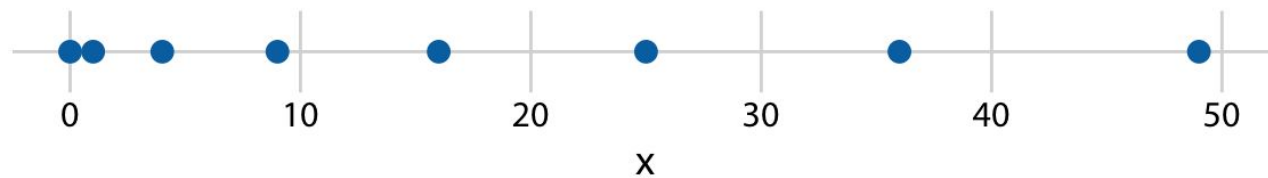


logarithmic scale with incorrect axis title

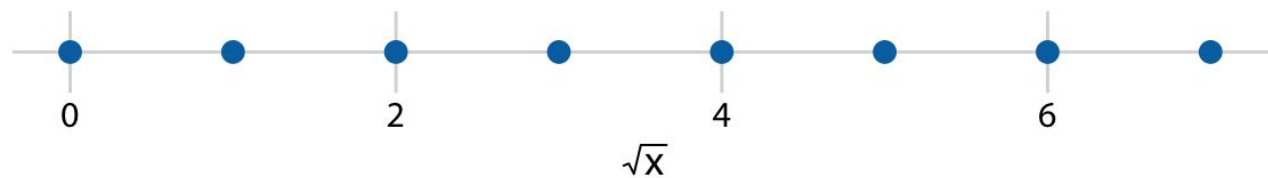


wrong

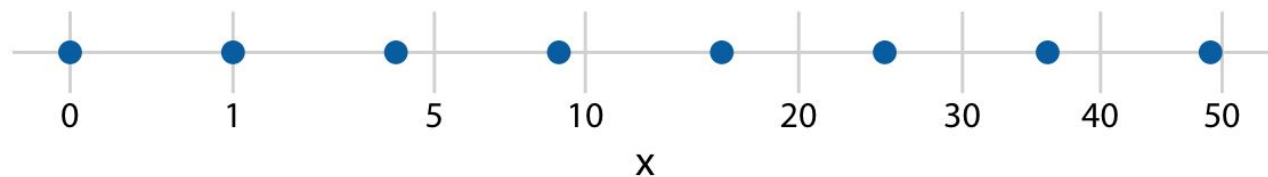
original data, linear scale



square-root-transformed data, linear scale



original data, square-root scale



**Remove**  
to improve  
(the **data-ink** ratio)



Antes, repasemos tipos de datos/variables:



# Tipos de gráficos

---





# ¡Son 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

---

*matplotlib*



seaborn



 pandas



bokkeh



# iA programar!

---

