

**1<sup>a</sup>**  
**Emisión**

# DATA SCIENCE

## **Módulo 05** Manipulación y visualización de datos con Python

*Mtro. Ricardo Daniel Alanis Tamez*



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
Dirección General de Cómputo y de Tecnologías de información y Comunicación  
Dirección de Docencia en TIC



Educación  
Continua  
1971 - 2021

# Librerías de visualización

## **Manipulación y visualización de datos con Python**

Ricardo Alanís

# Presentación

En este módulo vamos a empezar a tratar el tema de la visualización de datos. Primero conoceremos las diferentes alternativas técnicas para graficar datos, para luego tocar base en los diferentes conceptos detrás

# Objetivo

El participante identificará las distintas alternativas en librerías de visualización y los conceptos generales.

# Agenda de Hoy

- 1. Paseo por las distintas librerías de visualización de datos**
- 2. Conceptos generales de visualización y librerías disponibles**
- 3. El rol de la ciencia de datos en la visualización de la información.**

# Ejercicio 7.1: Paseo a librerías

Exploremos en equipos las diferentes librerías disponibles. Enmarca sus características, documentación y casos de uso.

- Matplotlib
- Seaborn
- Ggplot
- Plot.ly
- Bokeh
- Streamlit
- D3.JS - Javascript
- Vega y VegaLite - Javascript(Adaptador para Python)

# Introducción a la visualización de datos

La visualización de datos es la representación gráfica de datos e información. La visualización de datos utiliza tablas, gráficos, software u otras herramientas de visualización para proporcionar una descripción general rápida de los datos y mostrar las tendencias y relaciones que existen.

## Objetivos más importantes

- Visualizar la tendencia en el conjunto de datos
- Reconocer fácilmente los valores atípicos
- Reconocer patrones de datos
- Comprender la relación entre los datos

# Herramientas comunes para visualizar

Vamos a dar una búsqueda rápida a estas herramientas. Qué tienen en común, qué tienen de diferente.

- Power BI
- Tableau
- Qlik
- Adaptive Insights
- Dash
- Google Charts



# Errores comunes en la visualización de datos

- Etiquetas incorrectas a los datos
- Las proporciones gráficas están lejos de representar la realidad
- Se usó la gráfica incorrecta para la comunicación buscada

# Comunicación visual de la información

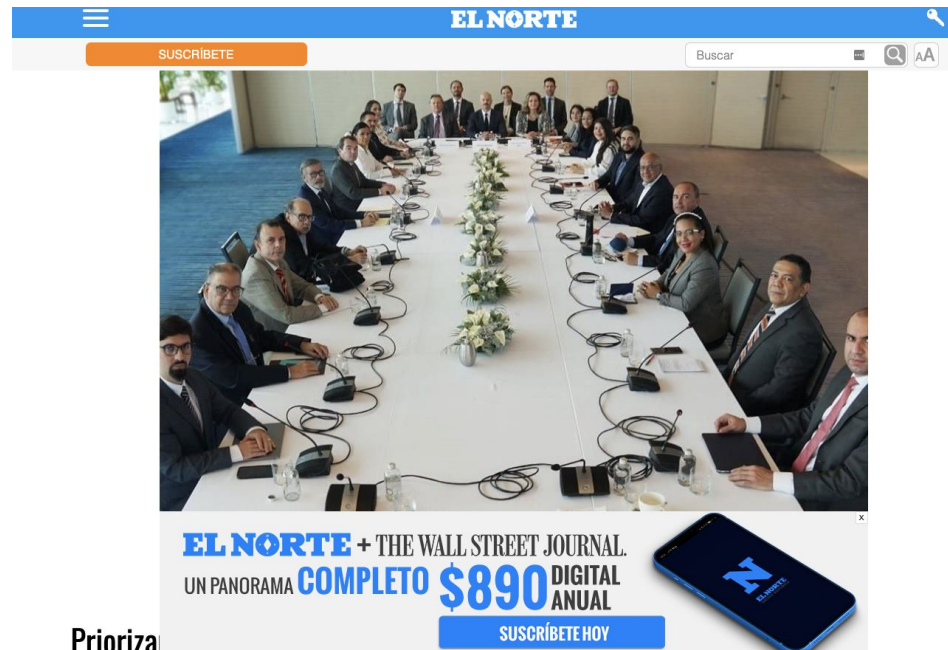
- Debemos de ser intencionales con la forma en la que presentamos la información.

Debemos de tener presente:

- Quién es el público
- Cuál es la motivación a comunicar
  - Objetivo del trabajo
- Cuál es el contexto a compartir
  - Datos, información alrededor
- Porqué debería de ser relevante el mensaje que se manda
  - Conclusiones e insights

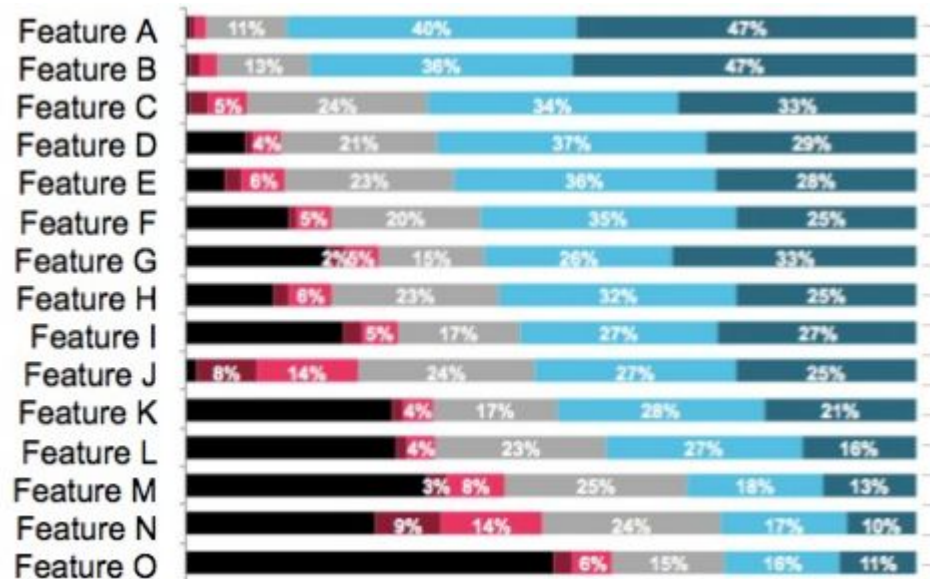
# Generando un marco

- El público debería de tener claro cuál es el objetivo a entender
- Se debe de ganar la atención
- Y si parece obvio, qué bueno. Estas haciendo un buen trabajo en enmarcar el problema.



Prioriza

# No ocultar el insight, dejarlo claro.



## User satisfaction varies greatly by feature

### Product X User Satisfaction: Features

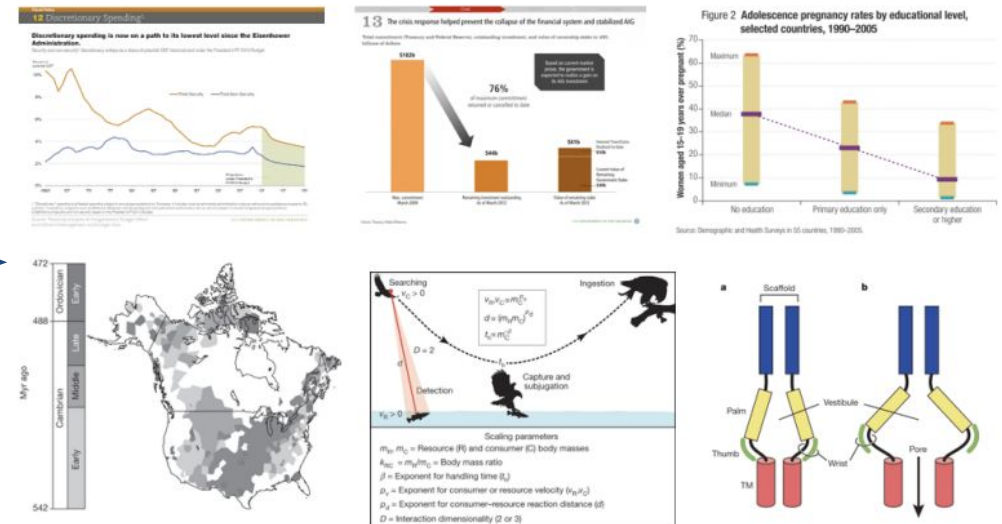
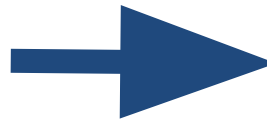
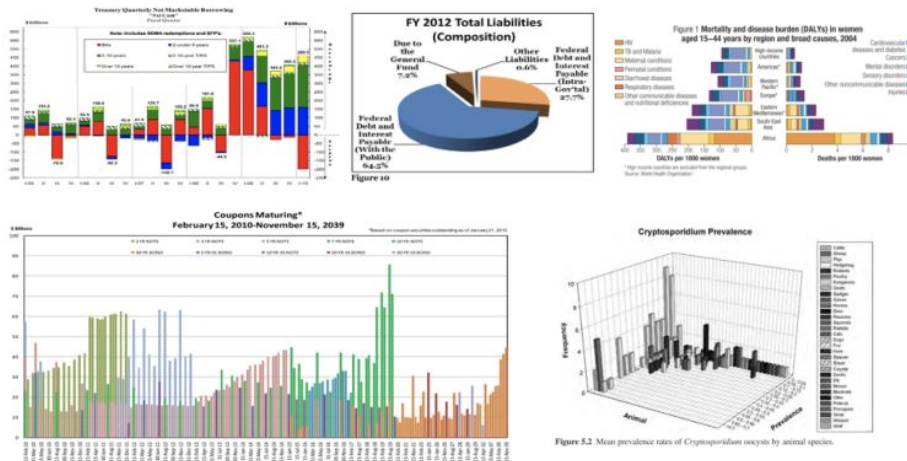
• Completely satisfied • Very satisfied • Somewhat satisfied • Not very satisfied • Not satisfied at all • Have not used



Feature O is least-used feature; what steps can we proactively take with existing users to increase use?

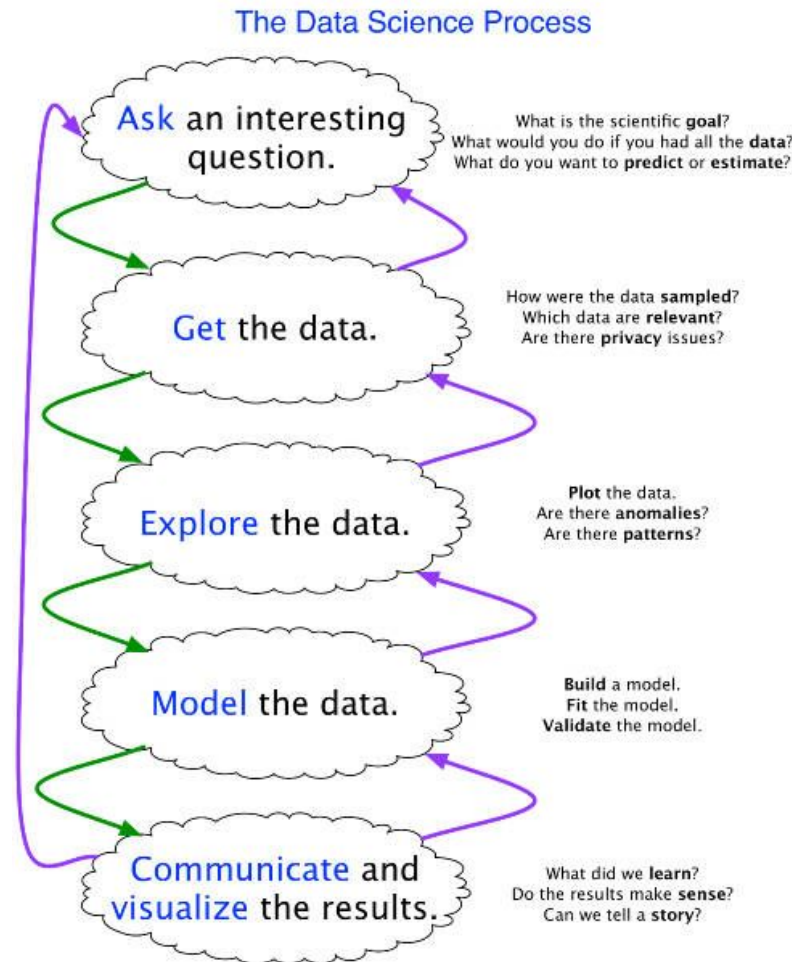
# Tomar en cuenta

- Determinar objetivos lo más pronto posible
- Generar una comunicación concreta
- La gente recuerda más si se termina en una nota positiva
- Menos es más, KISS (Keep it simple, stupid)



# Conexión con la ciencia de datos

En estos puntos, es crítico el rol de la visualización de datos

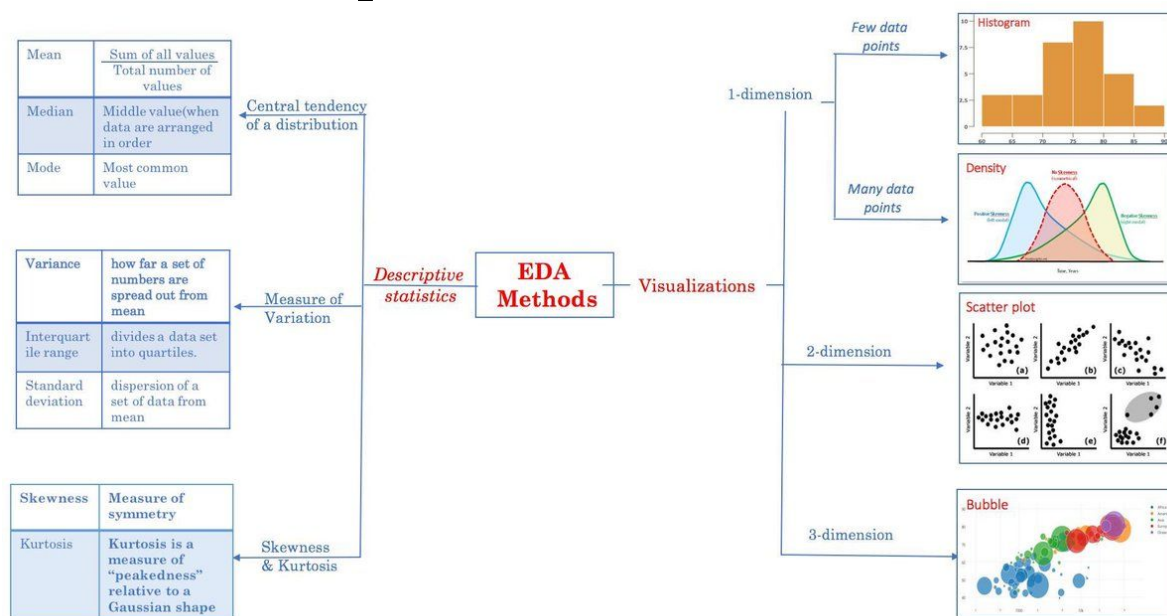


Joe Blitzstein and Hanspeter Pfister, created for the Harvard data science course <http://cs109.org/>.



# Conexión: Análisis exploratorio de datos

El análisis exploratorio de datos (EDA) es el proceso de resumir los aspectos importantes de los datos mediante la visualización. EDA es una parte importante del proceso de ciencia de datos. EDA busca verificar suposiciones y probar hipótesis, así como descubrir patrones / tendencias en los datos.



# Conexión: Compartir Conclusiones

Existen distintos productos, tanto visuales como centrados en texto para compartir los resultados

En qué circunstancias crees que deberíamos optar por uno u otro

- Productos de datos
- Presentación de datos en texto
- Blog Posts
- Presentaciones
- Código disponible en github

Qué tipos de gráficos usarías dependiendo del canal de comunicación



# ¿Preguntas?

# Referencias

- “Matplotlib: Visualization with Python.” Matplotlib. Aug 13, 2021. [matplotlib.org](https://matplotlib.org)
- “seaborn: statistical data visualization.” Seaborn. [seaborn.pydata.org](https://seaborn.pydata.org)
- Scheithauer, Dr. Gregor. “Data Visualization in Python like in R’s ggplot2.” Towardsdatascience. 14 May, 2018. [towardsdatascience.com/data-visualization-in-python-like-in-rs-ggplot2-bc62f8debbf5](https://towardsdatascience.com/data-visualization-in-python-like-in-rs-ggplot2-bc62f8debbf5)
- Bostock, Mike. “Data-Driven Documents.” D3js. [d3js.org](https://d3js.org)
- “Vega – A Visualization Grammar.” Vega. [vega.github.io/vega/](https://vega.github.io/vega/).
- “CS109 Data Science.” CS109. Harvard, 2015. <http://cs109.github.io/2015/>

# Contacto

*Mtro. Ricardo Daniel Alanis Tamez*

[ricardo@codeandomexico.org](mailto:ricardo@codeandomexico.org)

*LinkedIn: Ricardo Alanís*