

# Manipulación y visualización de datos

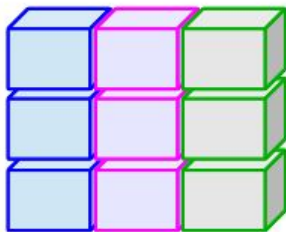
Sergio Santoyo  
Daniel Fernández

# Estructuras

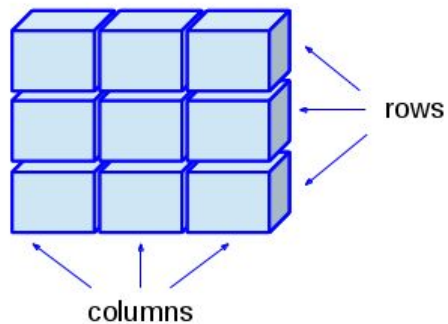
Vector



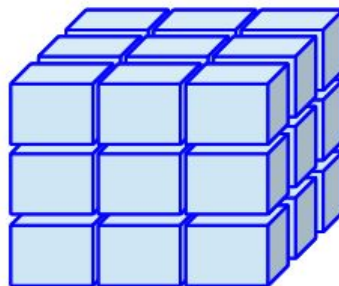
Data Frame  
(Table)



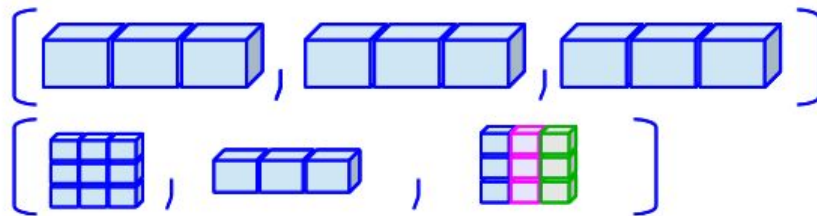
Matrix



Array



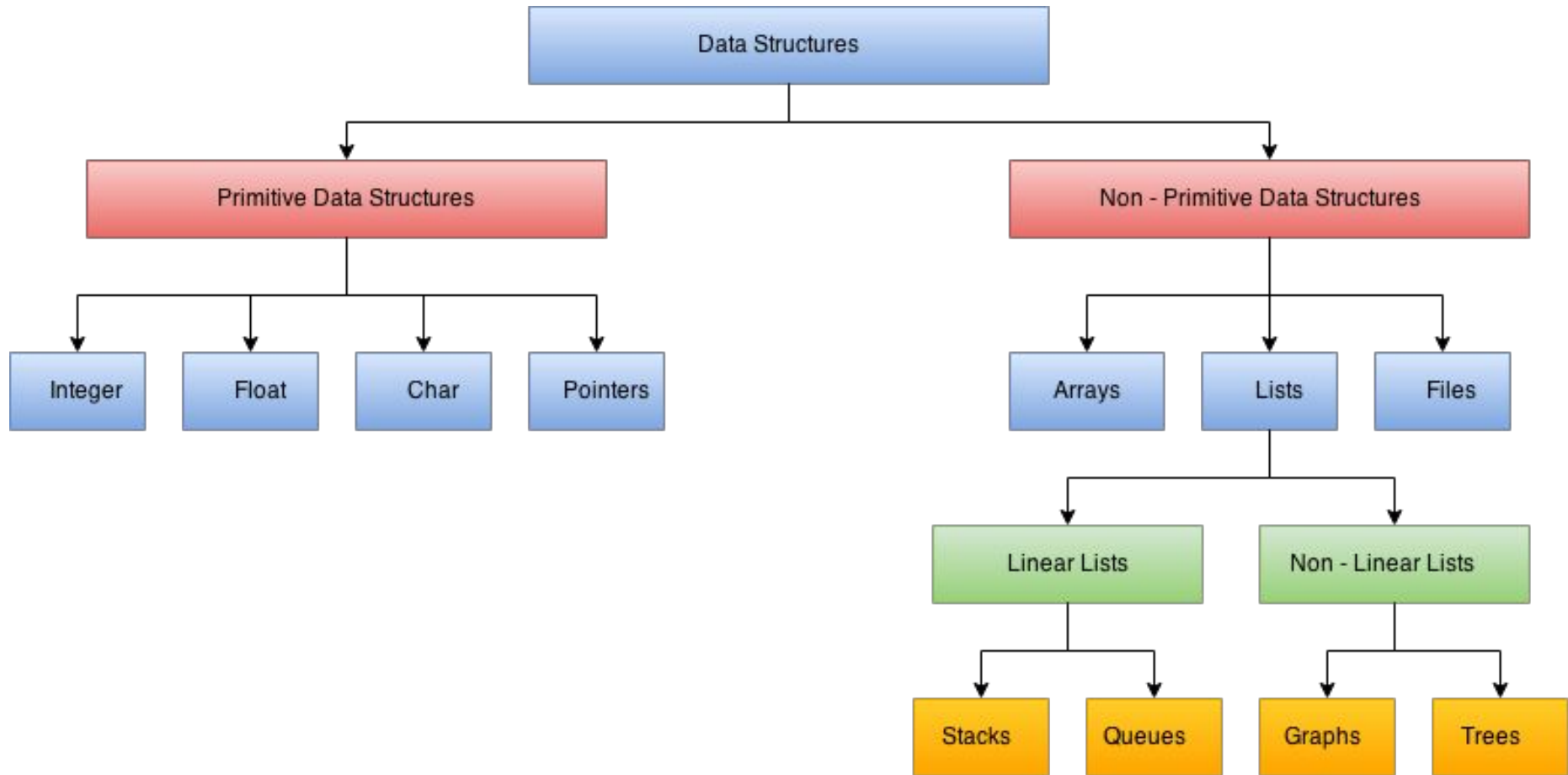
Lists



# Tipos de datos

- Cuantitativo
- Ordinal
- Nominal
- Intervalos (series de tiempo)
- Secuencias (texto)
- Imágenes





# Bases de datos

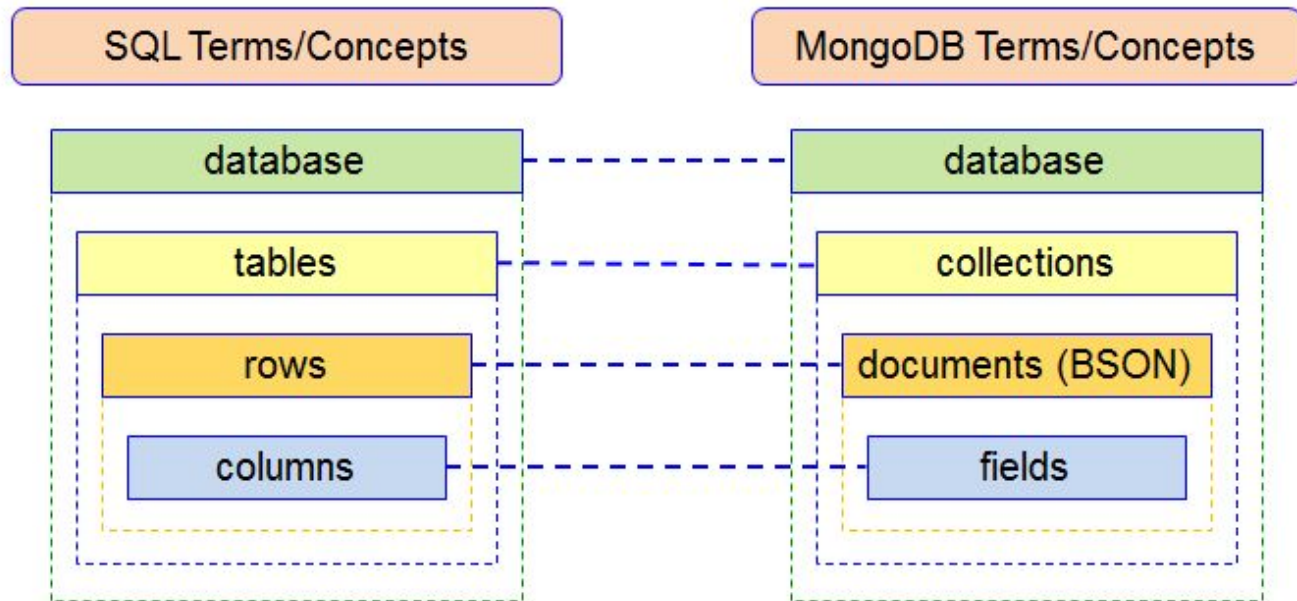
- Relacionales

- columnas → atributos
- filas → observaciones
- matrices  $\longleftrightarrow$  llaves
- .csv

- No Relacionales

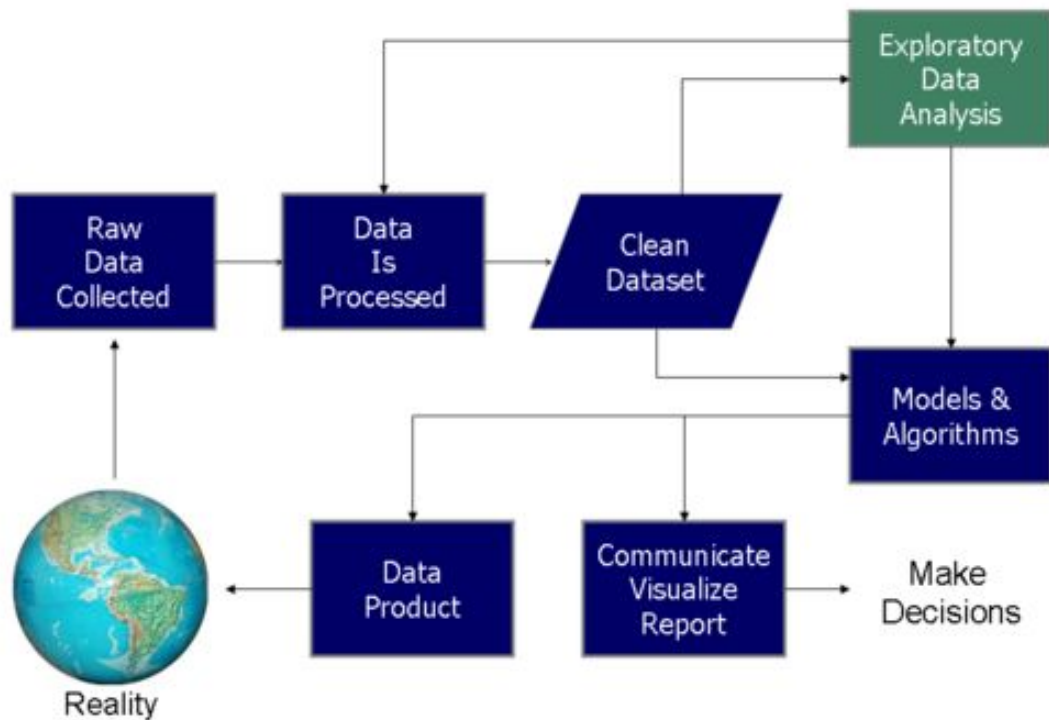
- Documentos
- Texto
- Sin formato o formato flexible
- .json





<b>DATA WAREHOUSE</b>	<b>vs.</b>	<b>DATA LAKE</b>
structured, processed	<b>DATA</b>	structured / semi-structured / unstructured, raw
schema-on-write	<b>PROCESSING</b>	schema-on-read
expensive for large data volumes	<b>STORAGE</b>	designed for low-cost storage
less agile, fixed configuration	<b>AGILITY</b>	highly agile, configure and reconfigure as needed
mature	<b>SECURITY</b>	maturing
business professionals	<b>USERS</b>	data scientists et. al.

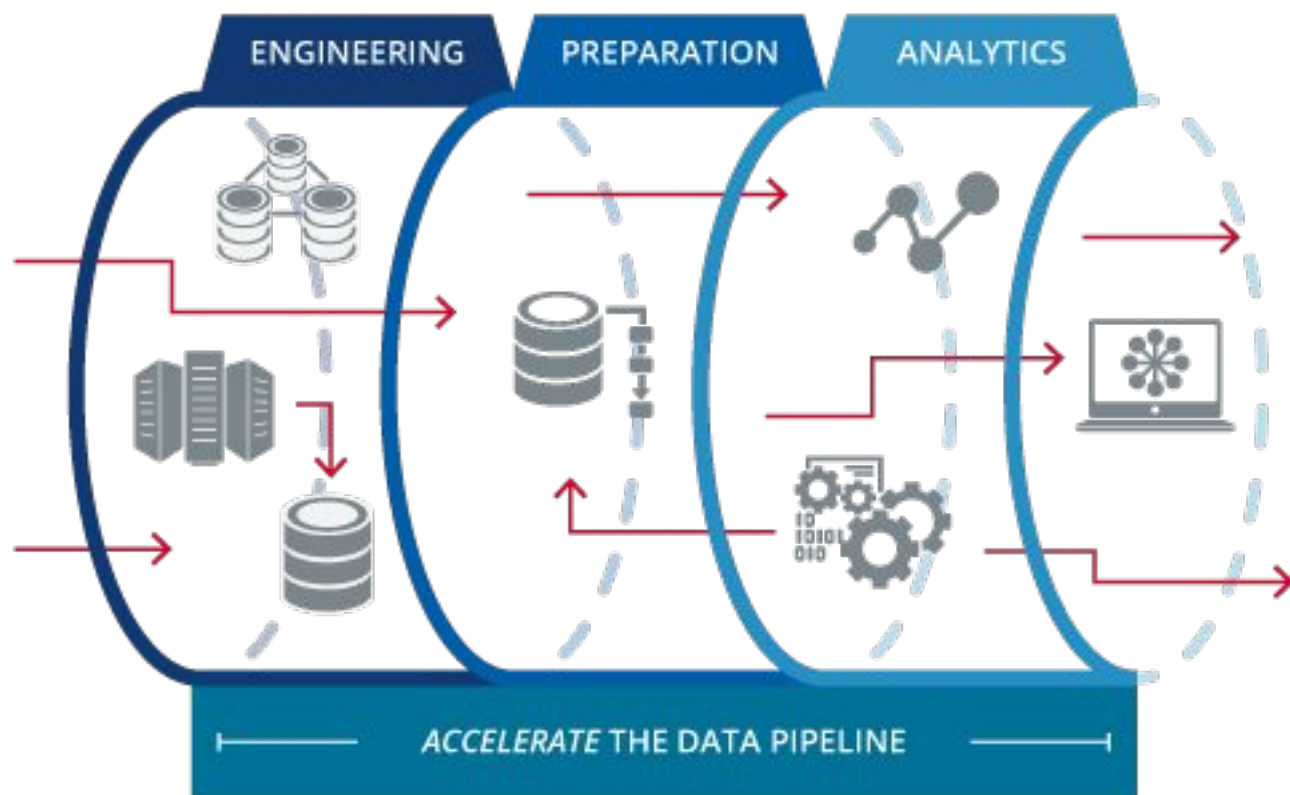
## Data Science Process



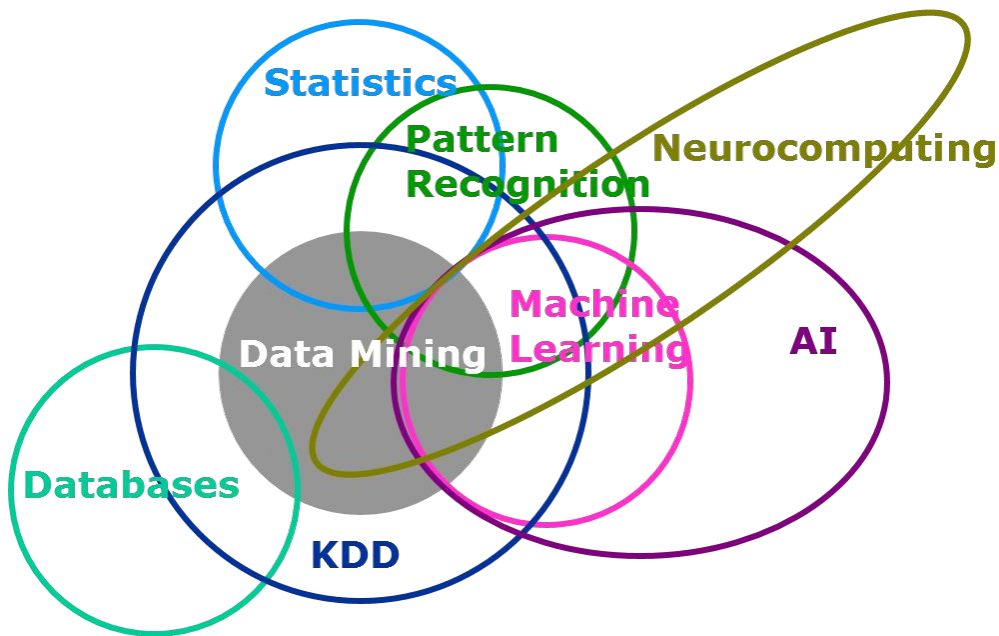


# Data Engineering

- Cómputo: C++, Python, R, Java, MatLab
- Almacenamiento: HDD, SQL, MongoDB, RAM, Git
- DevOps: AWS, GCloud, Docker
- Producto: Base de datos, análisis, estudio, servicio, plataforma, app, dashboard



# Modelos estadísticos



ML

Modelos estadísticos

weights

parameters

learning

fitting

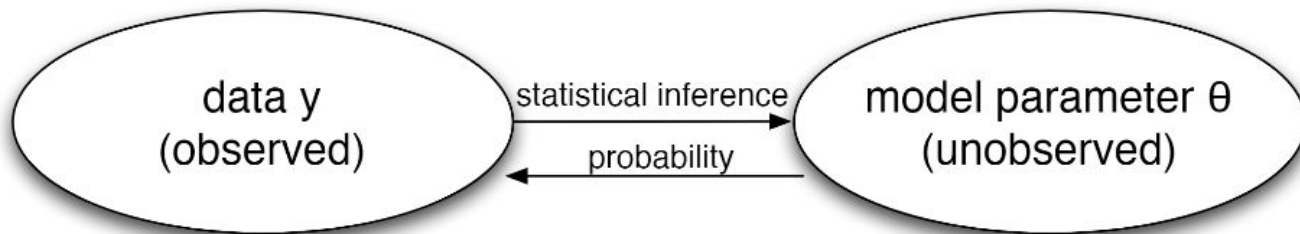
generalization

test set performance

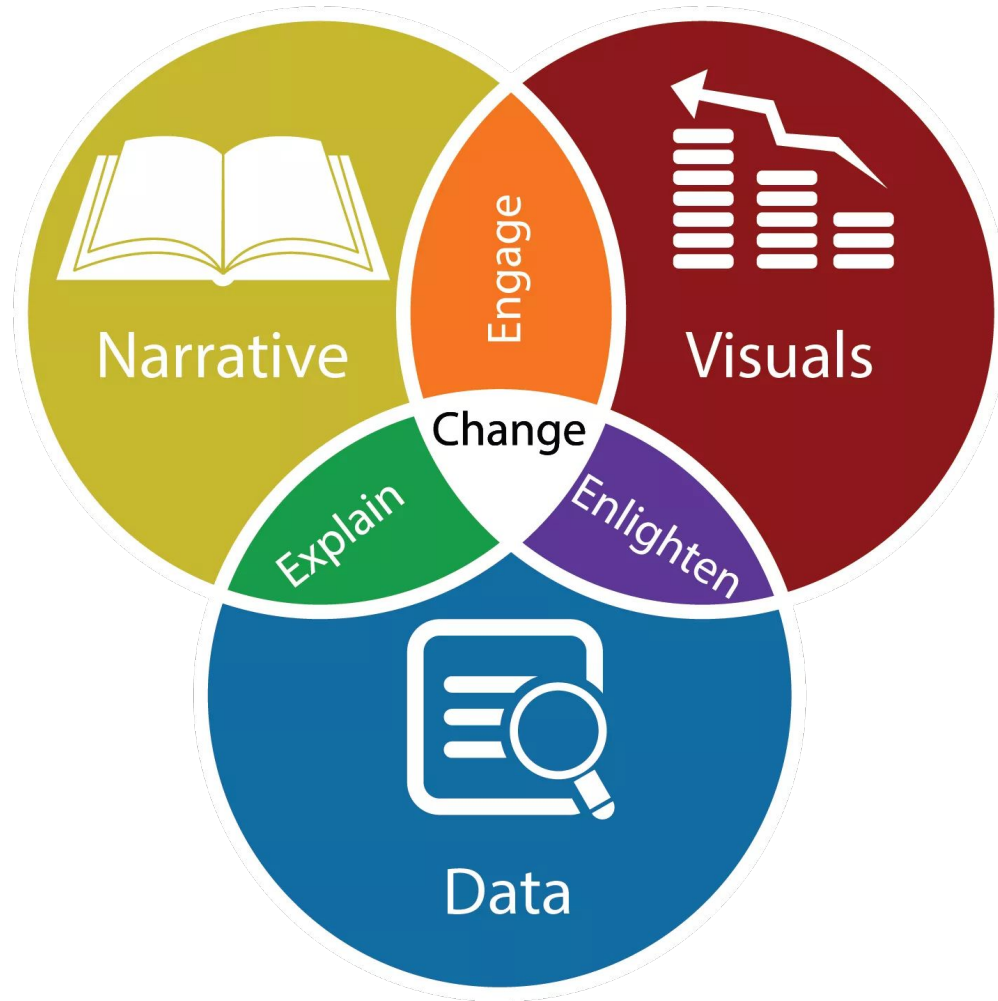
supervised learning

regression/classification

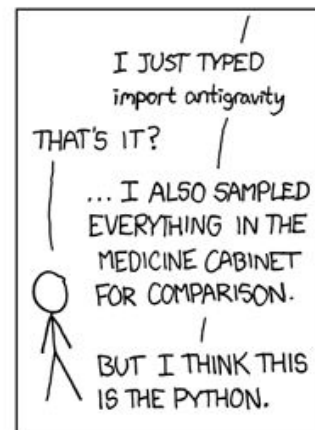
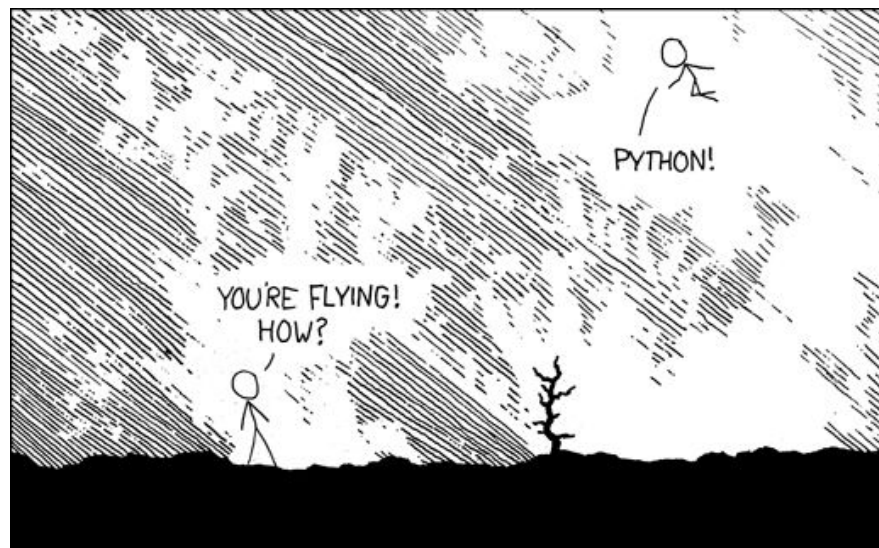
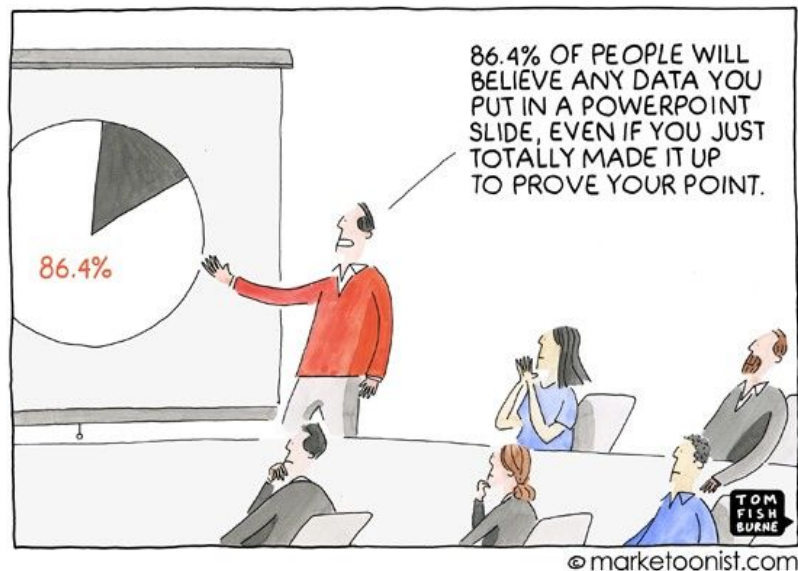
**ALL MODELS ARE WRONG, BUT SOME ARE USEFUL**



# Storytelling



# Notebook 1:



# Notebook Reto 0:

- Dataset: Titanic Machine Learning From Disaster
- Pasos:
  - Descargar el dataset de: <https://www.kaggle.com/c/titanic/data>
  - Descargar el notebook: <https://github.com/ldfo/ds-ua-2018/tree/master/clase2>
  - Contestar las preguntas en el notebook con un análisis exploratorio. (Gráficas, tablas, rellenar nulos, etc)
  - Cualquier duda preguntar al canal de Slack