Hablemos de datos

Maria Isabel Serrano G. Andrés Armando Sanchez M.



Agenda



Ciclo de vida de los datos



Clasificación de datos



Niveles de abstracción de información



Funciones de agregación



Protección de datos

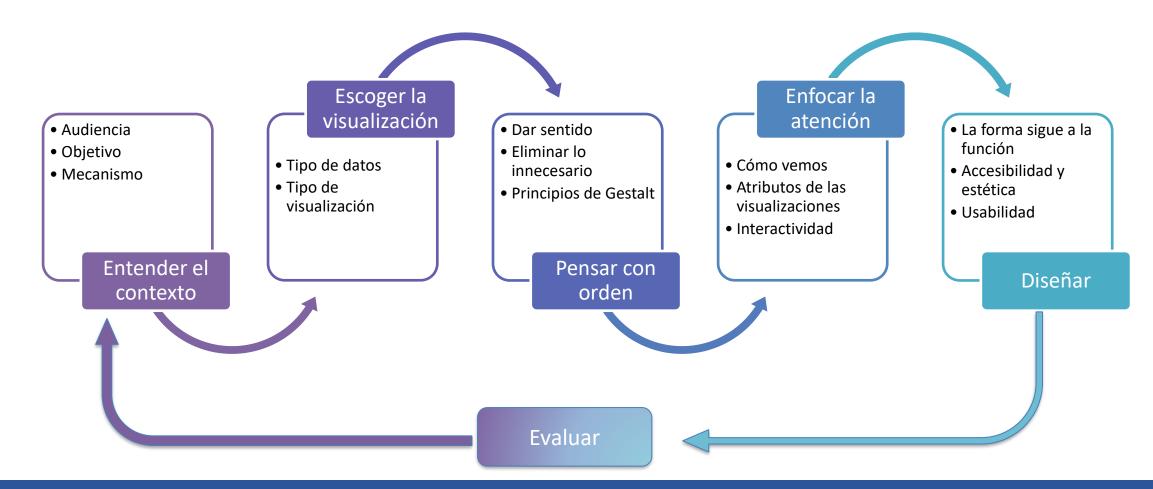


"Data! Data!... I can't make bricks without clay!"

-Sherlock Holmes via Sir Arthur Conan Doyle

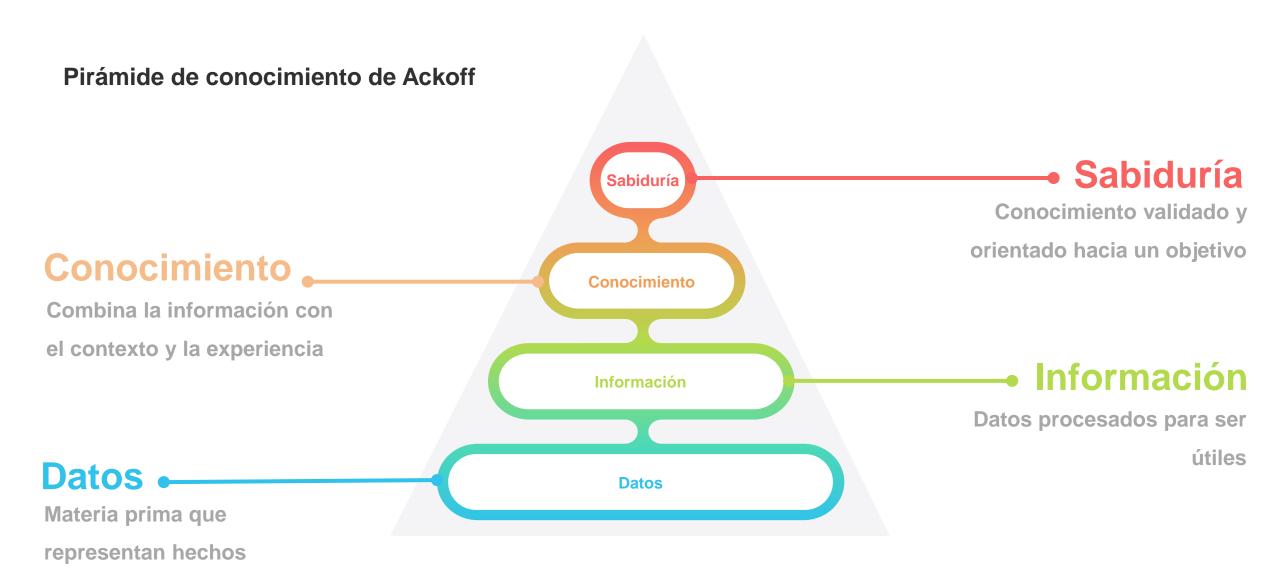


Metodología para visualizar información





Hablemos de los datos





CLASIFICACIÓN DE LOS DATOS



Abstracción de datos

 Describir los datos de forma que se pueda decidir cuál método de codificación es el más apropiado para usar con la información que se tiene.

Tipo de datos → Operación o transformación

Tipo de datos → Representación visual



Clasificación de los datos

Estructura

Datos estructurados

Datos no estructurados

Datos semiestructurados







txt, img, videos, clicks, redes sociales, IoT, eventos en tiempo real, fuentes de negocio, etc.

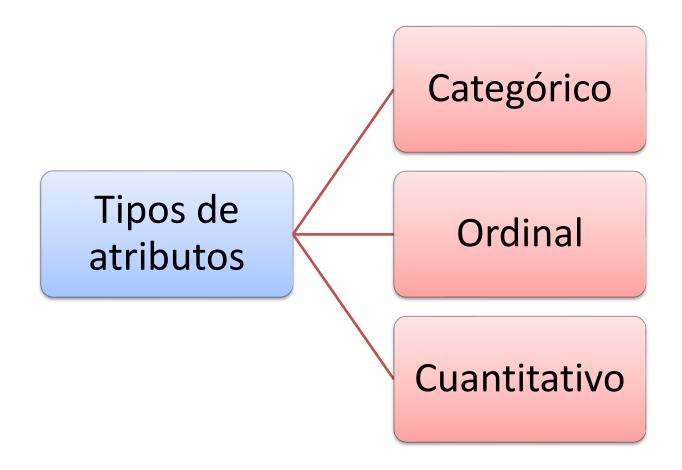


Atributos de los datos - calidad



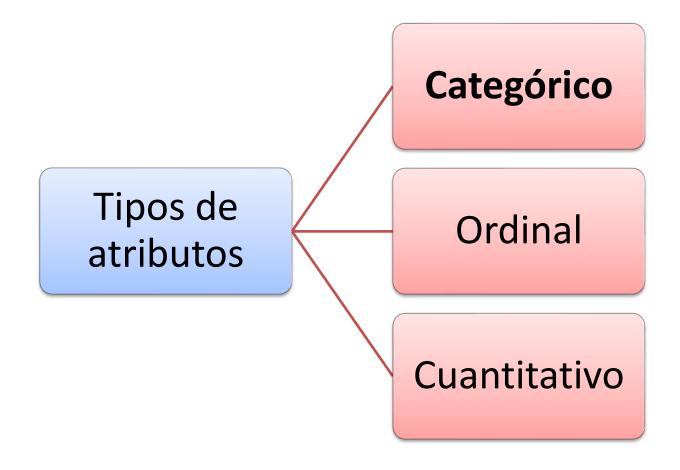


Tipos de atributos





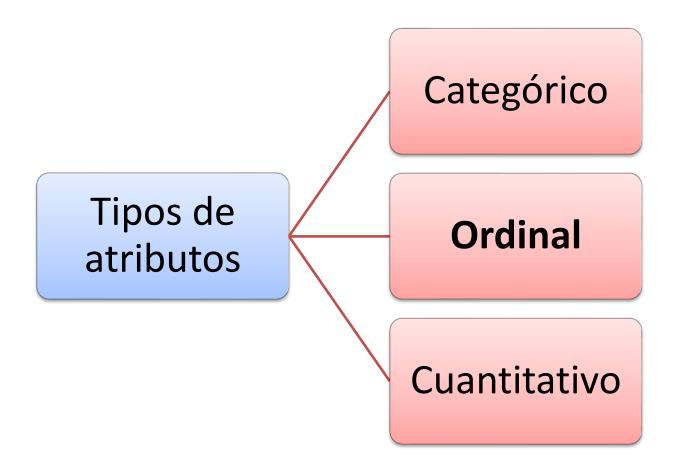
Recordemos tipos de atributos



- ✓ Sus valores describen categorías
- ✓ No tienen orden
- ✓ Ej: color cabello, tipo de comida



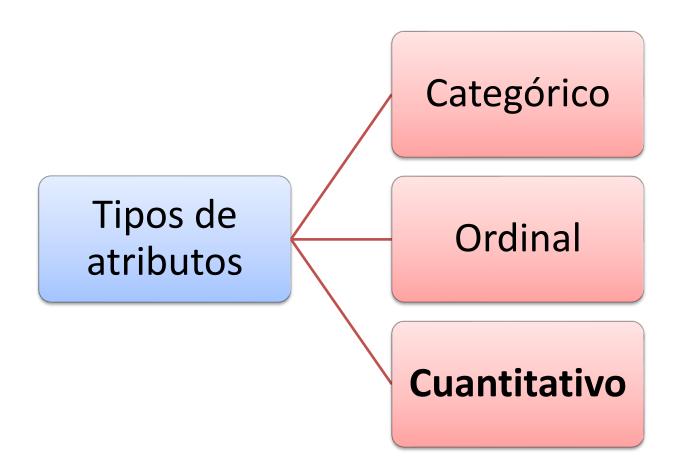
Recordemos tipos de atributos



- ✓ Sus valores describen categorías
- ✓ Que pueden ser ordenadas (tiene sentido hacerlo)
- ✓ Ej: estatus económico, nivel educativo
- ✓ No se conoce la "distancia" entre categorías → no aplican operaciones aritméticas



Recordemos tipos de atributos



- ✓ Sus valores representan una cantidad medible
- ✓ Ej: altura de personas
- ✓ Se conoce y puede calcular la "distancia" entre ellos.
- ✓ Se aplican operaciones aritméticas



Ejercicio – Reconocer tipos de atributos

Para el dataset de ventas:

<u>Sample - Superstore Sales</u>

(Excel) - Google Sheets

 Clasifique cada uno de los atributos según su tipo y justifique si es necesario

Order ID	Order Priority	Order Quantity	Sales	Region	Customer Segment	Product Category
3	Low	6	261,54	Central	Small Business	Office Supplies
6	Not Specified	2	6,93	West	Corporate	Office Supplies
32	High	26	2808,08	West	Corporate	Furniture
32	High	24	1761,4	West	Corporate	Furniture
32	High	23	160,2335	West	Corporate	Technology
32	High	15	140,56	South	Corporate	Technology
35	Not Specified	30	288,56	Central	Corporate	Office Supplies
35	Not Specified	14	1892,848	Central	Corporate	Technology
36	Critical	46	2484,7455	West	Home Office	Technology
65	Critical	32	3812,73	Central	Corporate	Technology
66	Low	41	108,15	West	Consumer	Office Supplies
69	Not Specified	42	1186,06	South	Corporate	Furniture
69	Not Specified	28	51,53	South	Corporate	Office Supplies
70	Low	48	90,05	East	Home Office	Office Supplies
70	Low	46	7804,53	East	Home Office	Technology
96	High	37	4158,1235	South	Home Office	Technology
97	Medium	26	75,57	West	Consumer	Office Supplies
129	Low	4	32,72	Central	Corporate	Office Supplies
130	High	3	461,89	South	Corporate	Technology
130	High	29	575,11	South	Corporate	Office Supplies
130	High	23	236,46	South	Corporate	Office Supplies
132	Medium	27	192,814	East	Consumer	Technology
132	Medium	30	4011,65	East	Consumer	Furniture
134	Not Specified	11	1132,6	Central	Home Office	Office Supplies
135	Not Specified	25	125,85	East	Consumer	Technology
166	High	10	567,936	South	Consumer	Technology



Semántica de los atributos

¿Qué es semántica?

El significado de los atributos.

Características que se agregan a los tipos de atributos y permiten entender su significado

- Espacial geográfico
- Temporal relacionado > Divergente: se puede con el tiempo
- Secuencial
 - identificar un valor de cero.
 - Cíclico: información cíclica en el tiempo. Ej: día de la semana

Jerárquico: permite crear jerarquías entre atributos



Ejercicio – Semántica de atributos

Para el dataset de ventas:

Sample - Superstore Sales

(Excel) - Google Sheets

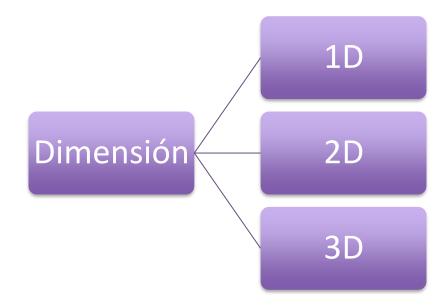
 Completar la clasificación con la semántica

Order ID	Order Priority	Order Quantity	Sales	Region	Customer Segment	Product Category
3	Low	6	261,54	Central	Small Business	Office Supplies
6	Not Specified	2	6,93	West	Corporate	Office Supplies
32	High	26	2808,08	West	Corporate	Furniture
32	High	24	1761,4	West	Corporate	Furniture
32	High	23	160,2335	West	Corporate	Technology
32	High	15	140,56	South	Corporate	Technology
35	Not Specified	30	288,56	Central	Corporate	Office Supplies
35	Not Specified	14	1892,848	Central	Corporate	Technology
36	Critical	46	2484,7455	West	Home Office	Technology
65	Critical	32	3812,73	Central	Corporate	Technology
66	Low	41	108,15	West	Consumer	Office Supplies
69	Not Specified	42	1186,06	South	Corporate	Furniture
69	Not Specified	28	51,53	South	Corporate	Office Supplies
70	Low	48	90,05	East	Home Office	Office Supplies
70	Low	46	7804,53	East	Home Office	Technology
96	High	37	4158,1235	South	Home Office	Technology
97	Medium	26	75,57	West	Consumer	Office Supplies
129	Low	4	32,72	Central	Corporate	Office Supplies
130	High	3	461,89	South	Corporate	Technology
130	High	29	575,11	South	Corporate	Office Supplies
130	High	23	236,46	South	Corporate	Office Supplies
132	Medium	27	192,814	East	Consumer	Technology
132	Medium	30	4011,65	East	Consumer	Furniture
134	Not Specified	11	1132,6	Central	Home Office	Office Supplies
135	Not Specified	25	125,85	East	Consumer	Technology
166	High	10	567,936	South	Consumer	Technology



Atributos de los datos - Dimensión

- Valores escalares 1D: peso
- Valores vectoriales 2D: dirección de movimiento
- Tensores 3D o más: describen fuerzas de dirección y de corte.
 Fuerza gravitacional.





Operaciones con los datos

- Operaciones matemáticas y de comparación
- Combinación
- Inversión de valores
- Creación / eliminación de entidades o relaciones
- Transformación de entidades
- Dividir una entidad en sus componentes

Difíciles de visualizar. Se pueden trabajar con animaciones.



Transformación de datos

- Funciones de agregación
 - Suma, promedio, max, ... Típico funciones estadísticas
 - Agregaciones por fecha y hora (resolución) o espaciales
- Geo-codificación (nombre → coordenadas)
- Agrupaciones binning
 - Transformar datos cuantitativos en ordinales o discretos
- Re-escalar o re-expresar datos cuantitativos
 - Normalización: representar valores en una escala diferente. Ej. [-1, 1]
 - Porcentajes
 - Distancia a una referencia (ej. promedio)



Transformación de datos

¿Por qué es importante?

- Parte del proceso de diseño
- Puede facilitar la comunicación y el entendimiento





¿Porqué es útil conocer los tipos de atributos antes de hacer una visualización?



¿Cuál es la mejor visualización?

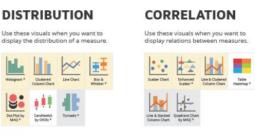














Depende

- Del usuario que toma la decisión
- De la tarea o problema de negocio que se quiere resolver
- 3. De los datos que se tienen disponibles

Preferred for Dashboards Not Recommended (there is a Better Alternative or it is just a Clone) * Custom Visual

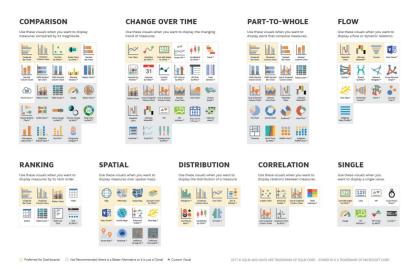
2017 © SQLBI AND OKVIZ ARE TRADEMARK OF SQLBI CORP. - POWER BI IS A TRADEMARK OF MICROSOFT CORP.



Investigación

Por grupos mostrar ejemplos prácticos de visualizaciones según el objetivo







En resumen

- Los datos son la fuente de todo proceso de visualización de datos y representan el contexto y características
- Conocer la clasificación y características de los datos nos guía para seleccionar representaciones gráficas apropiadas, estrategias de codificación.
- Seleccionar su tipo y definir la semántica son parte del proceso de diseño y pueden facilitar la comunicación y el entendimiento



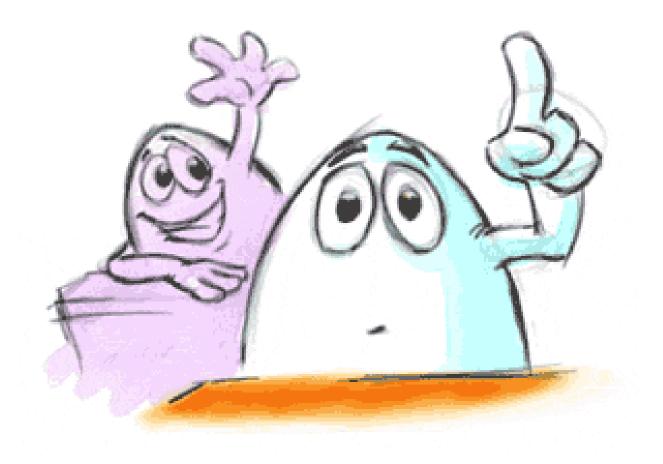
Datos Abiertos

Todos aquellos datos accesibles y reutilizables, sin exigencia de permisos específicos.

- Disponibilidad y acceso
- Reutilización y redistribución
- Participación universal







¿Preguntas?



Hablemos de datos

Maria Isabel Serrano G. Andrés Armando Sanchez M.

