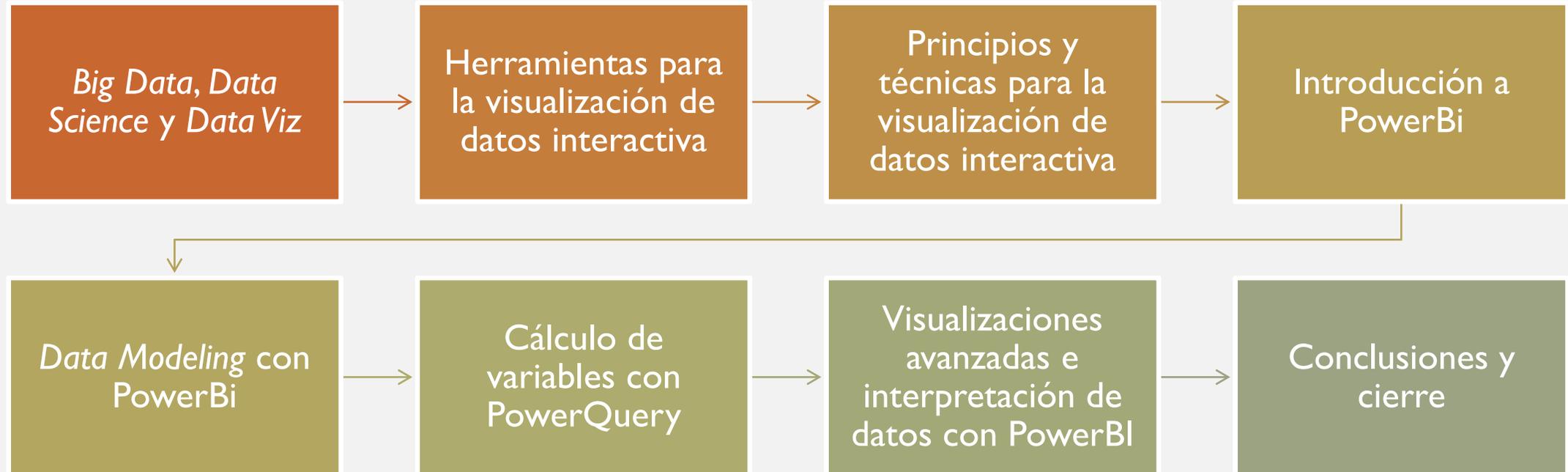


Jordi Morales i Gras
jordimorales.com
@jmoralesigras

VISUALIZACIÓN INTERACTIVA DE DATOS CON POWERBI

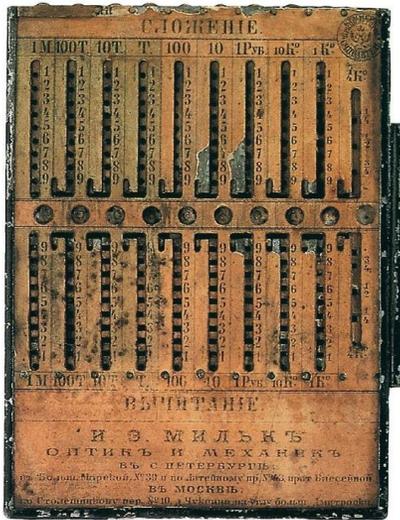
¿QUÉ VEREMOS?



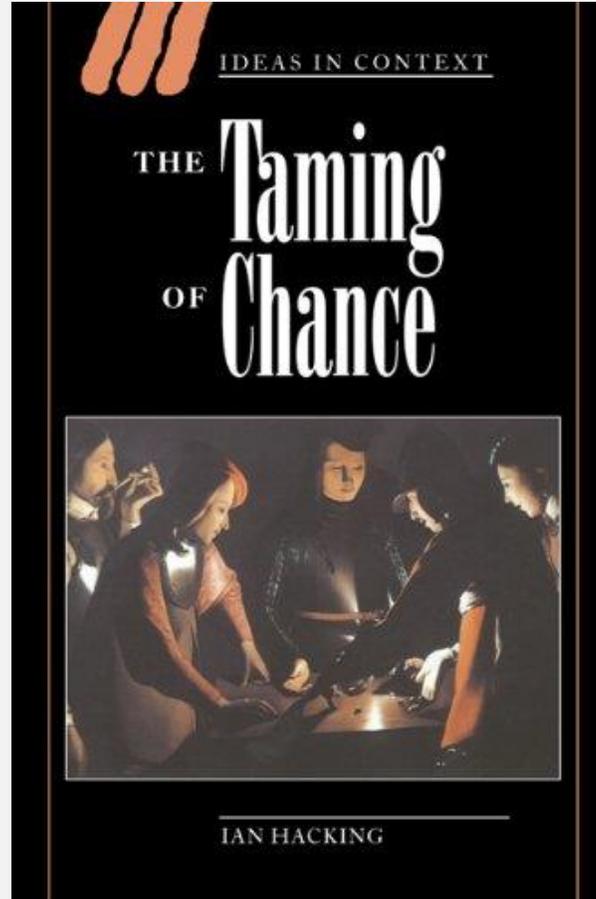
|

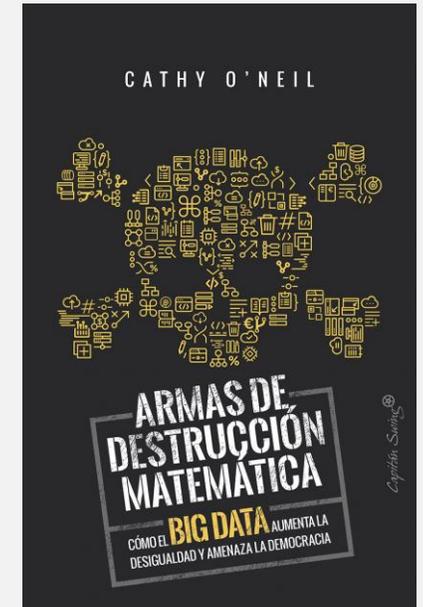
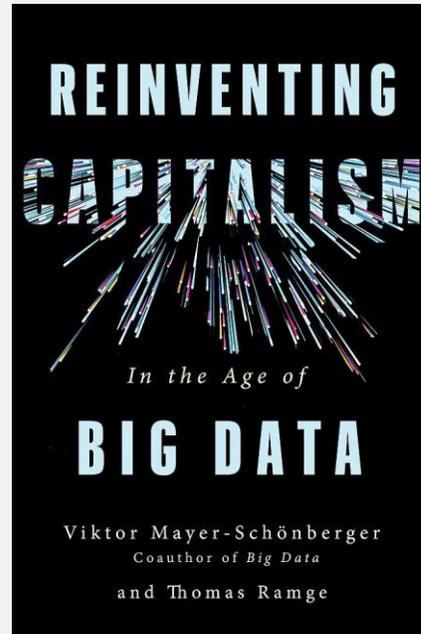
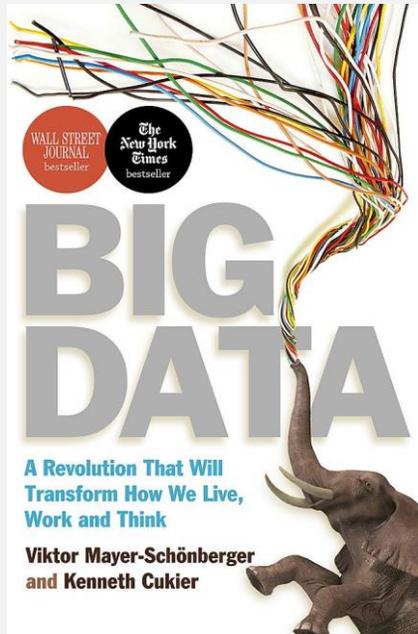
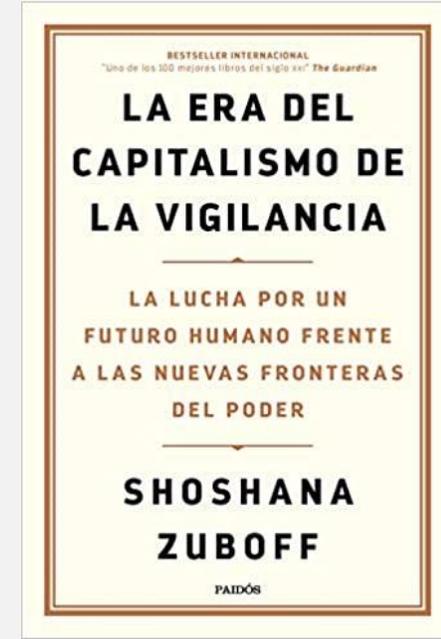
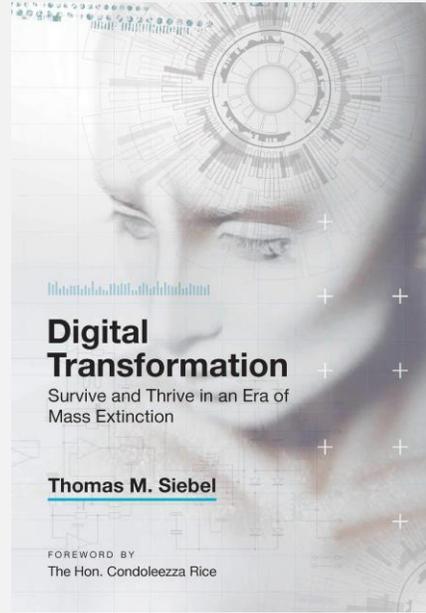
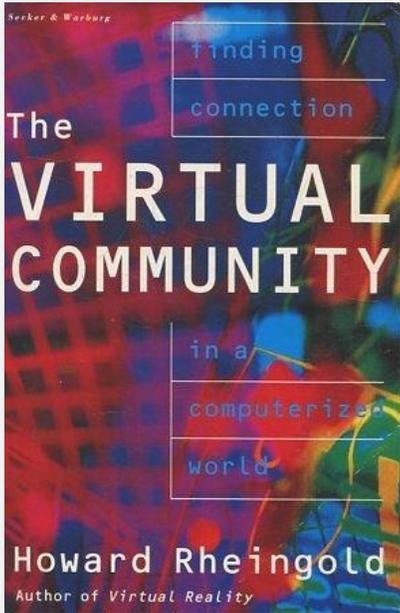
**BIG DATA, DATA
SCIENCE Y DATA VIZ**

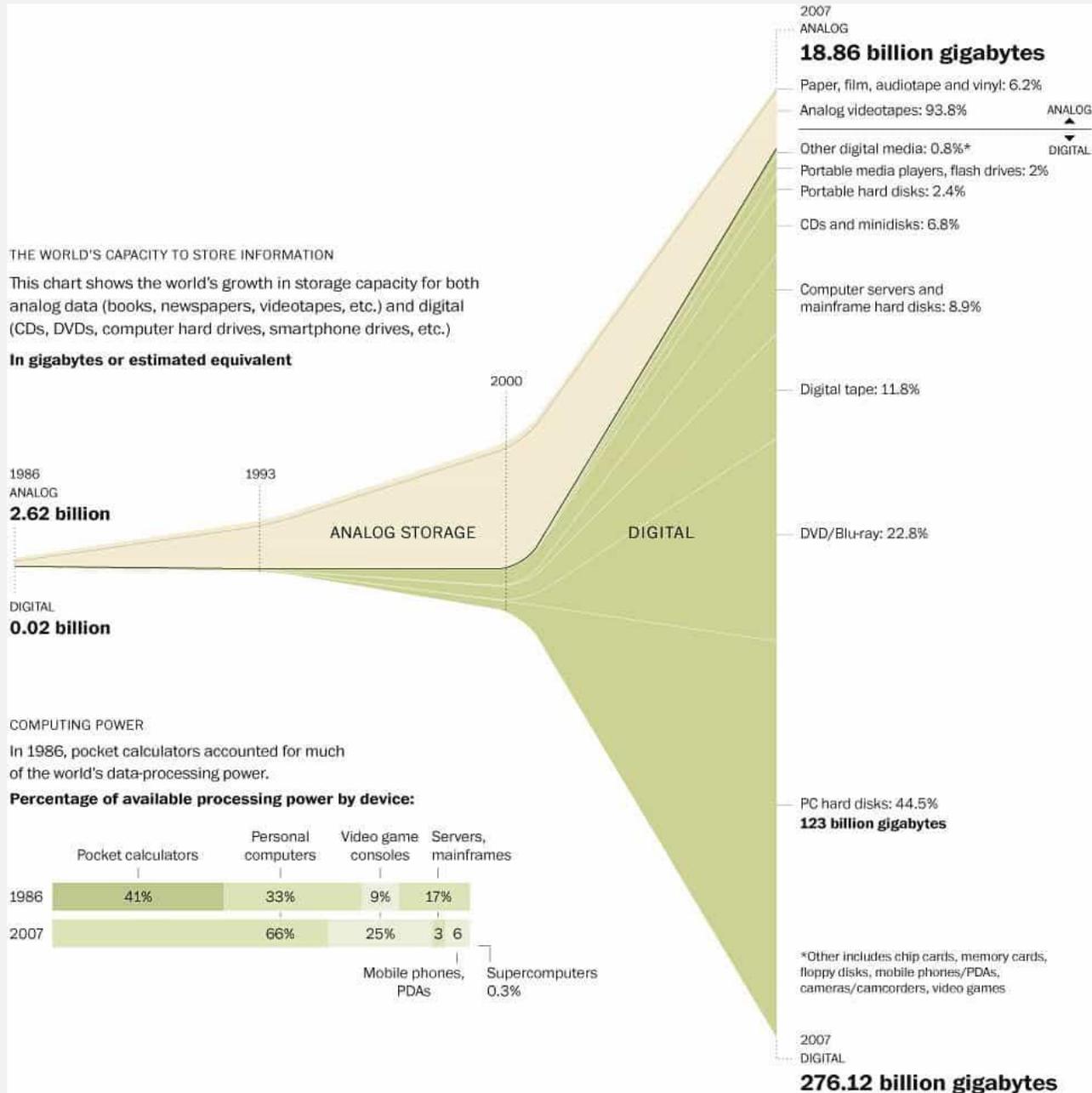
Alain Desrosières
La politique
des grands nombres
Histoire de la raison statistique



La Découverte/Poche



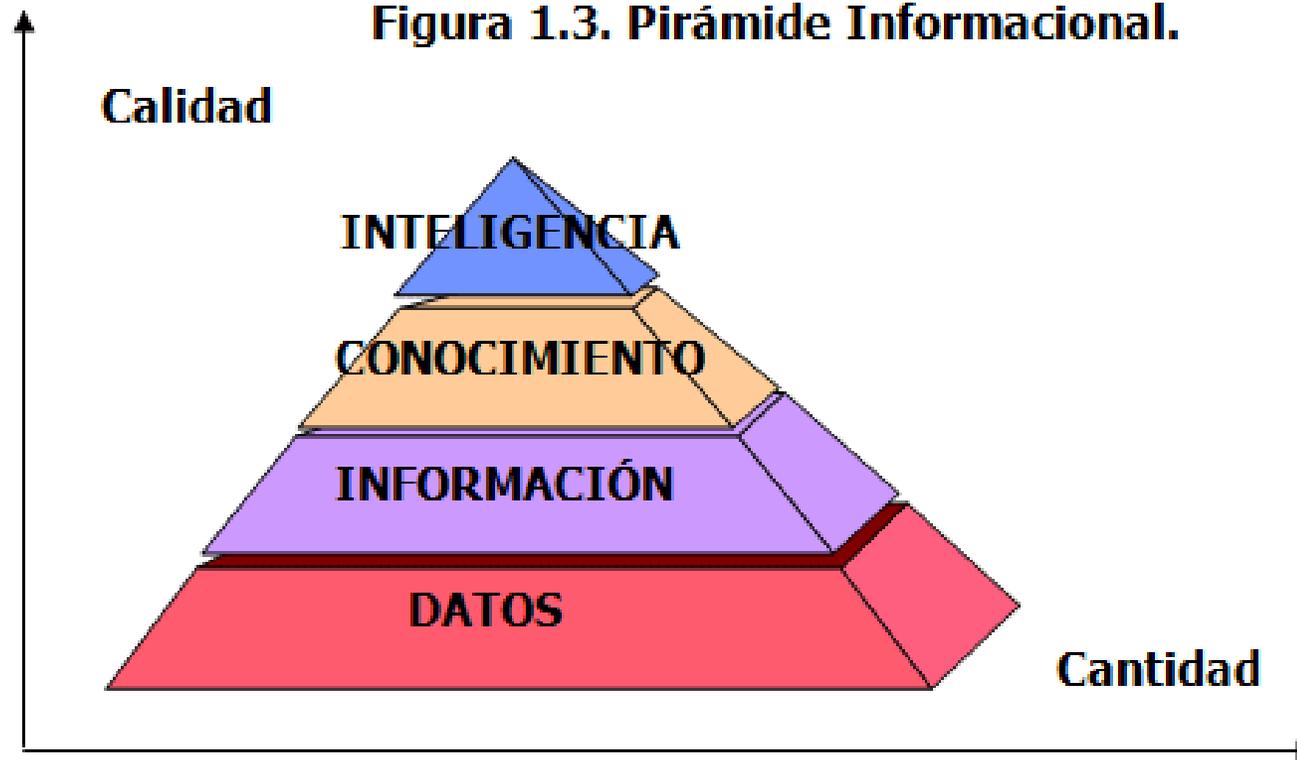




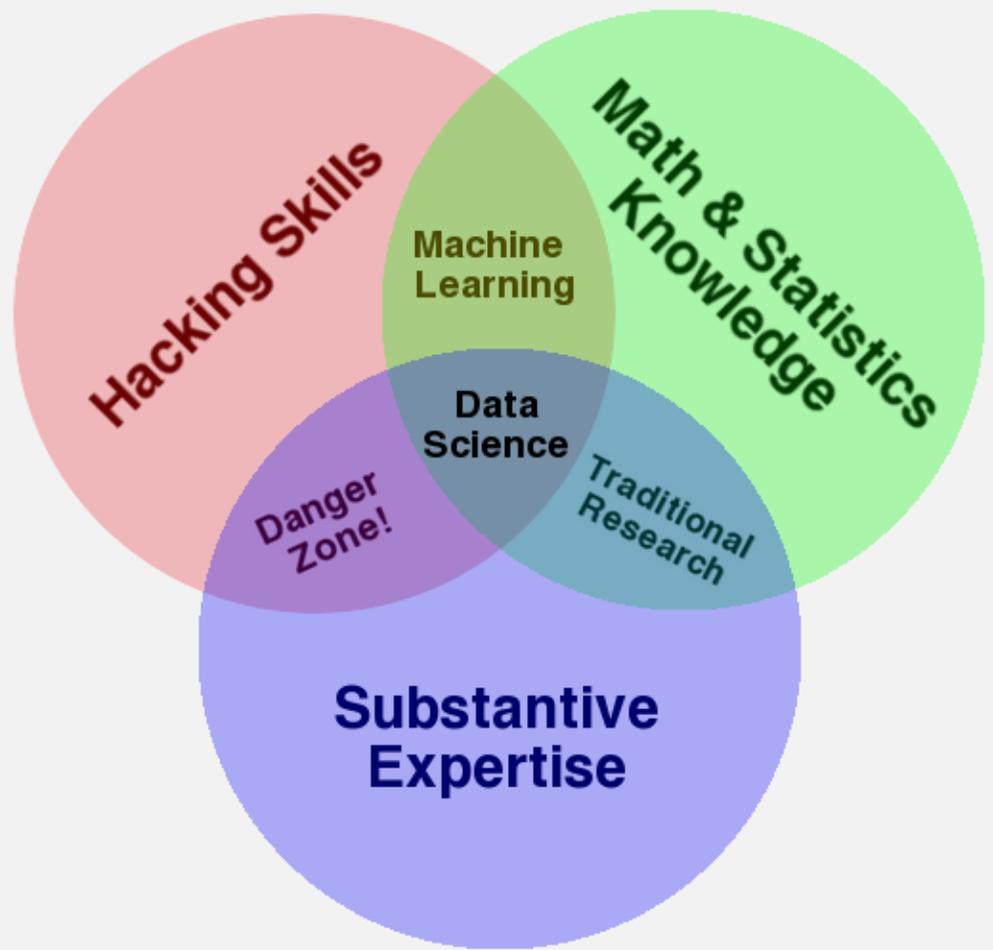
M. Hilbert and P. López, 2011. The world's technological capacity to store, communicate, and compute information.



Figura 1.3. Pirámide Informativa.



Fuente: Dante, Ponjuán, Gloria. Gestión de la información en las organizaciones. Principios, conceptos y aplicaciones. Santiago de Chile, 1998.





ESTRATEGIAS CLÁSICAS PARA LA
OBTENCIÓN DE DATOS SOCIALES



FUENTES DE DATOS MASIVOS



MEDIOS SOCIALES

Imagen, video, audio o texto de redes sociales virtuales



CLOUD

Público, privado o corporativo



WEB

Datos web, analytics



IoT

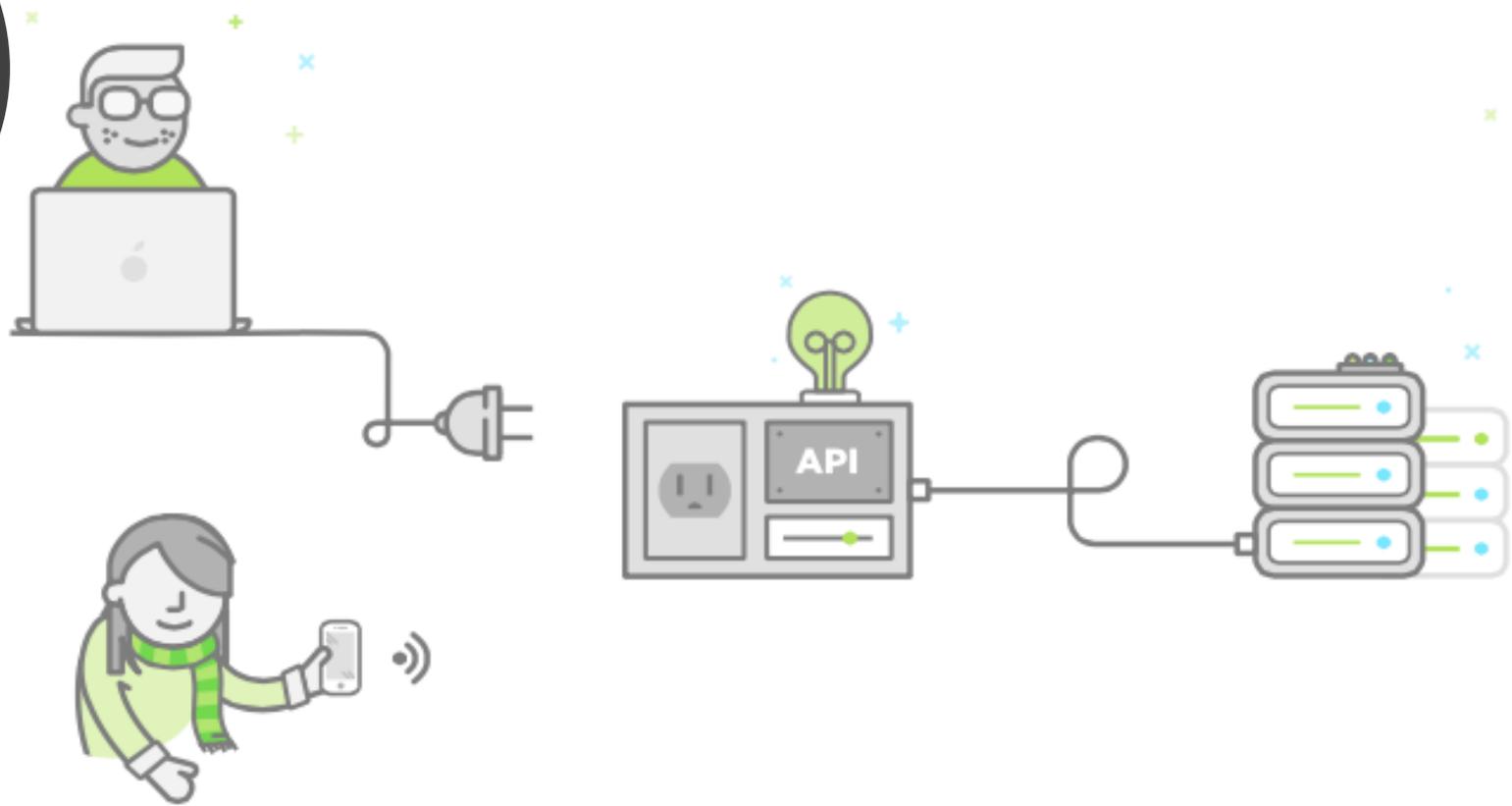
Sensores y dispositivos conectados



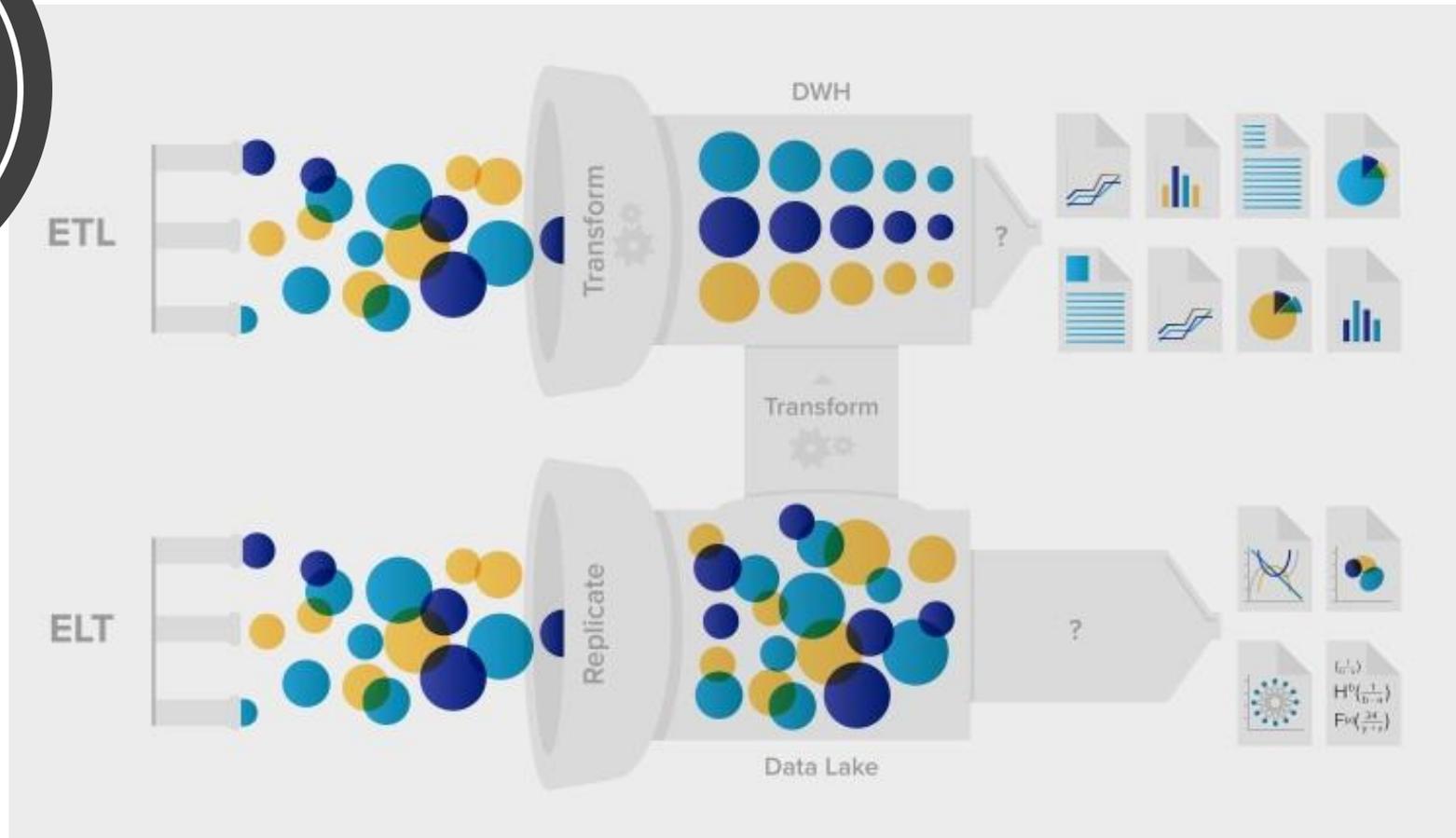
BASES DE DATOS

Datos secundarios tradicionales

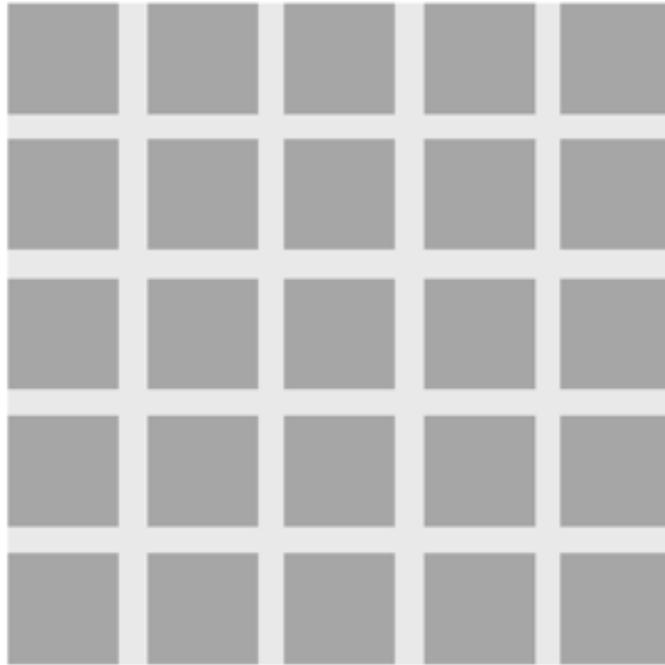
EL
NUEVO
CIRCUIT
O DEL
DATO



ETL VS ELT

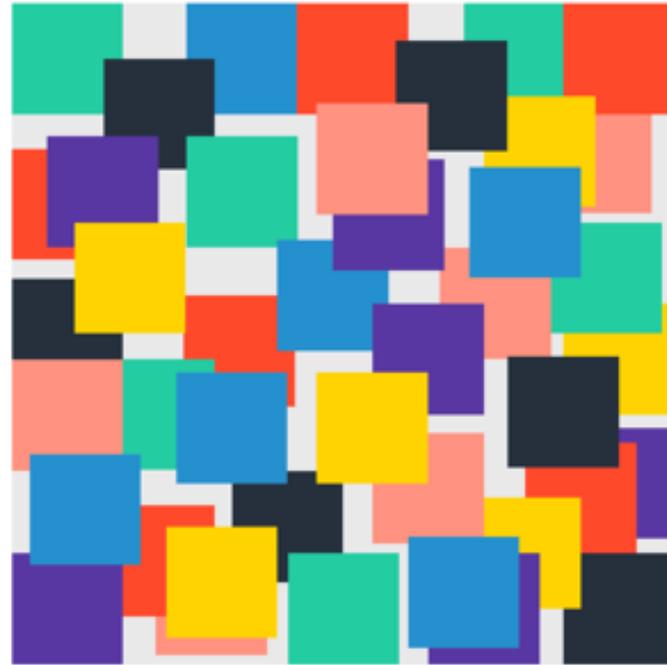


Structured data



Database, CRM, ERP

Unstructured data

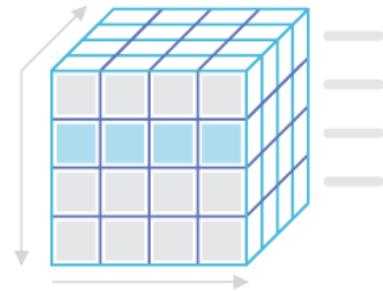
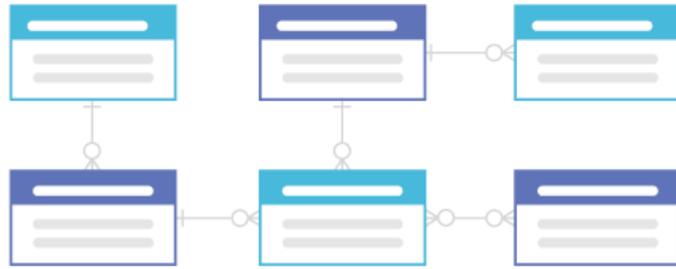


Text, audio, videos



SQL

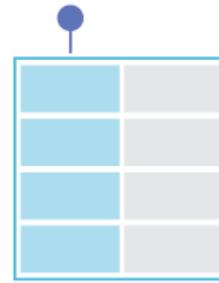
Relational Database Management Systems (RDBMS)



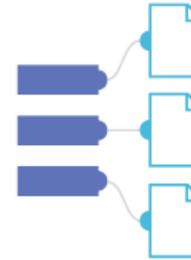
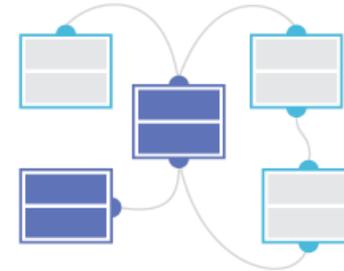
Online Analytical Processing (OLAP) Cube

NoSQL

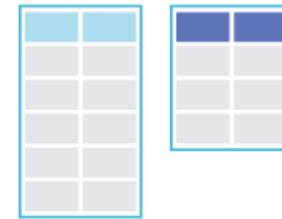
Key-Value



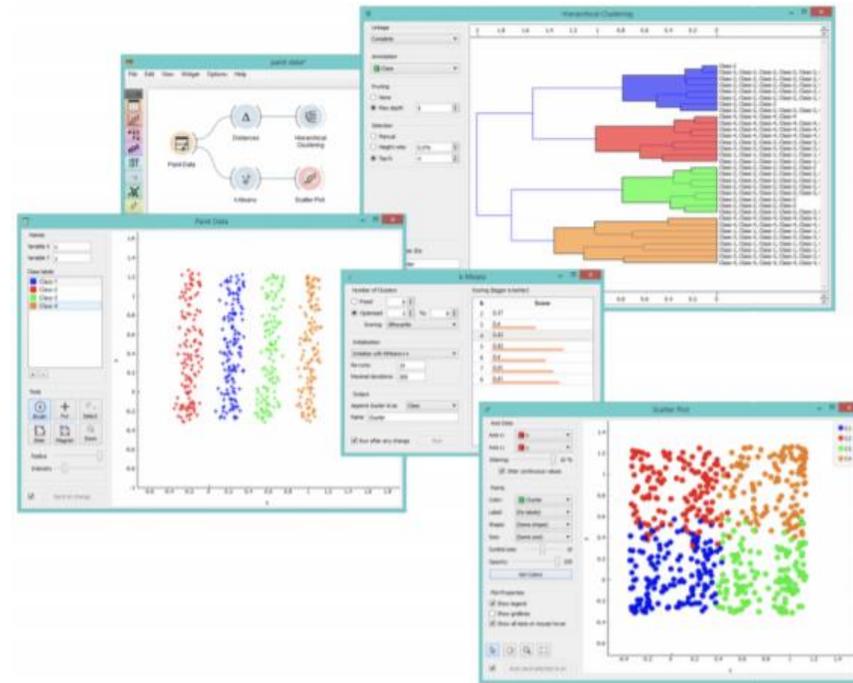
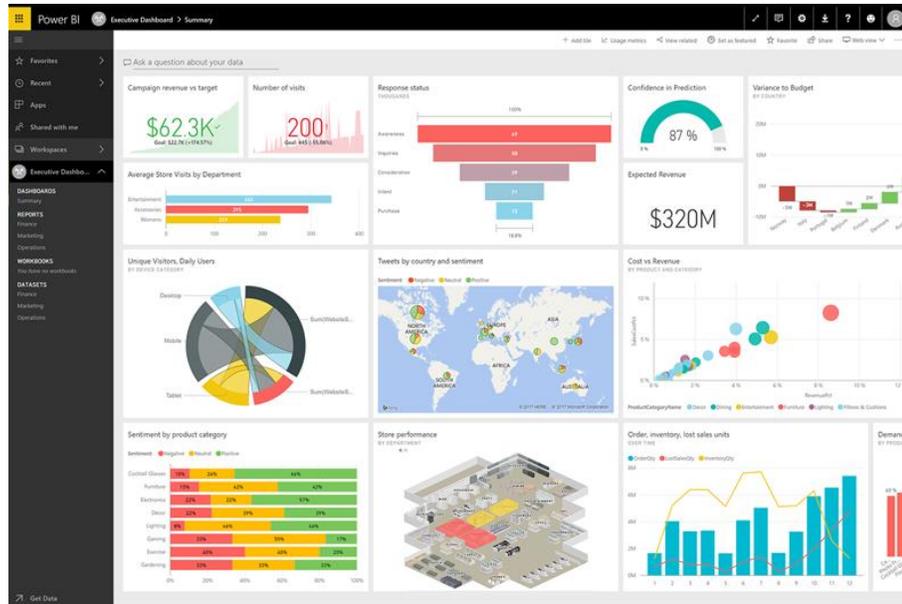
Graph



Document



Column store



BI (ETL) VS. DATAMINING (ELT)

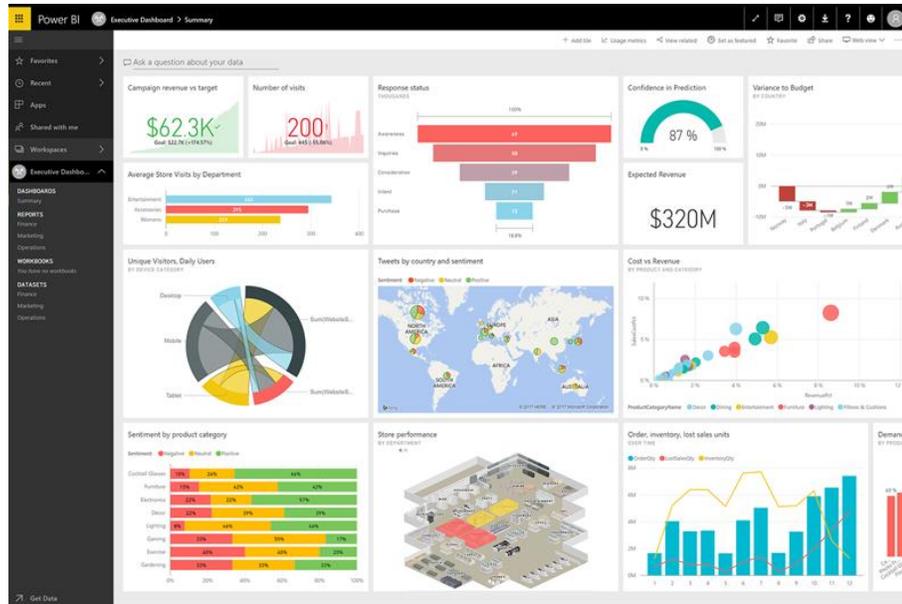
DOS ESTRATEGIAS PARA LA VISUALIZACIÓN DE DATOS

1. PARA RESPONDER UNA PREGUNTA CLAVE



2. PARA DESCUBRIR FACETAS O DIMENSIONES DE UN FENÓMENO





BI (ETL)

#2

HERRAMIENTAS
PARA LA
VISUALIZACIÓN DE
DATOS
INTERACTIVA

COVID-19

Alarma en el Reino Unido por un fallo técnico en el recuento de contagios

- Unos 16.000 casos de coronavirus no fueron notificados a tiempo para rastrear los contactos

<https://www.lavanguardia.com/internacional/20201005/483861404090/alarma-reino-unido-recuento-fallo-tecnico-contagios.html>

SOFTWARE INADECUADO



MALOS DATOS



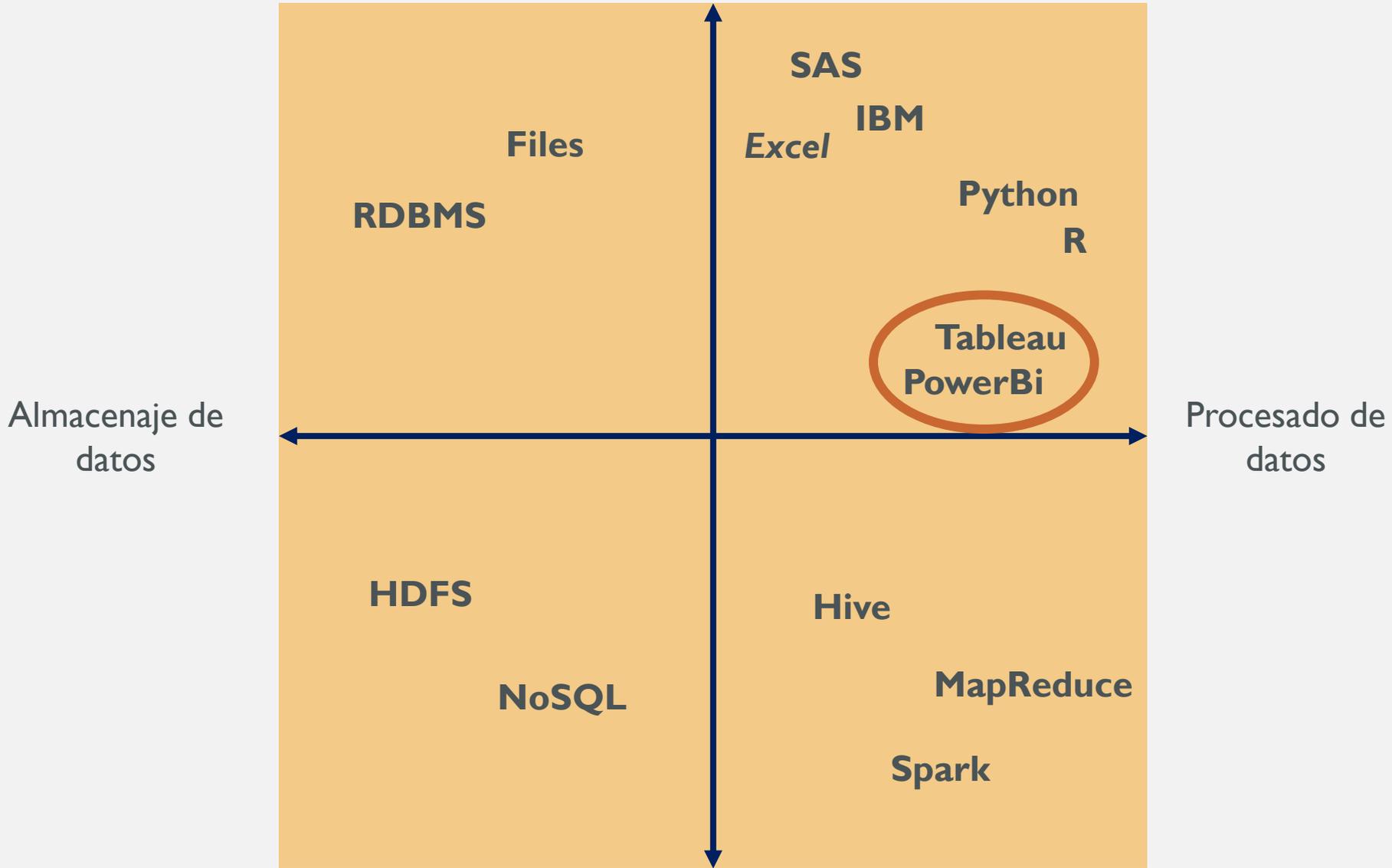
MALAS DECISIONES

Un **fallo técnico** que implicó que unos **16.000 casos de coronavirus** en el **Reino Unido** no fueron notificados a tiempo ha retrasado los esfuerzos del Gobierno británico para rastrear los contactos de esas personas que dieron positivo, informan este lunes los medios nacionales.

El error, según apuntan varios medios británicos, surgió desde un laboratorio que enviaba el reporte diario de los resultados de las PCR que realizaba en sus instalaciones en formato CSV (Coma Separated Values). Dicho formato es compatible con Excel, el programa que usa el PHE para indexar todos los casos. El archivo compartido con el departamento oficial incluía todo el histórico, no sólo los nuevos.

Día tras día, los responsables cargaban los nuevos datos al final del excel principal. Pero mientras que los archivos CSV pueden tener cualquier tamaño, los archivos de Microsoft Excel solo pueden tener una longitud de 1.048.576 filas. Cuando se abre un archivo CSV más largo que lo soportado en Excel, las filas inferiores se cortan y ya no se muestran, aunque el programa advierte al usuario que hay información que no cabe en el máximo establecido por Microsoft y, por ende, no será cargada en el documento.

Tecnologías
maduras



Tecnologías innovadoras



BI ARCHITECTURE

1

Data Source

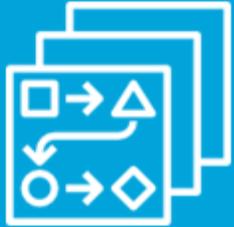


ERP



2

ETL



3

Data Warehouse



Data Marts

4

Data Viz



BI TOOLS COMPARISON 2021



Power BI



Data Studio

Microsoft Power BI

- Free for Desktop
- \$10 per user for Server
- Power Query to edit data
- Easier to use if you're familiar with Excel
- Custom with R / Python
- Direct connection with OneDrive & other Microsoft Ecosystems
- Good community & support

Easy to start with powerful data transformation capability

Tableau

- \$70 for Desktop
- \$12+ per user for Server
- Long legacy and widely adopted
- Beautiful visualization & gallery
- Can do fancy graph
- Better map support
- Strong community & support

Nice visualization and easy to explore data based on grammar of graphic

Google Data Studio

- FREE!
- Backed by Google ecosystem - Google Analytics, Google Ads, Google Form
- Fastest data refresh

Good for online data especially Google Analytics & Google AdWords

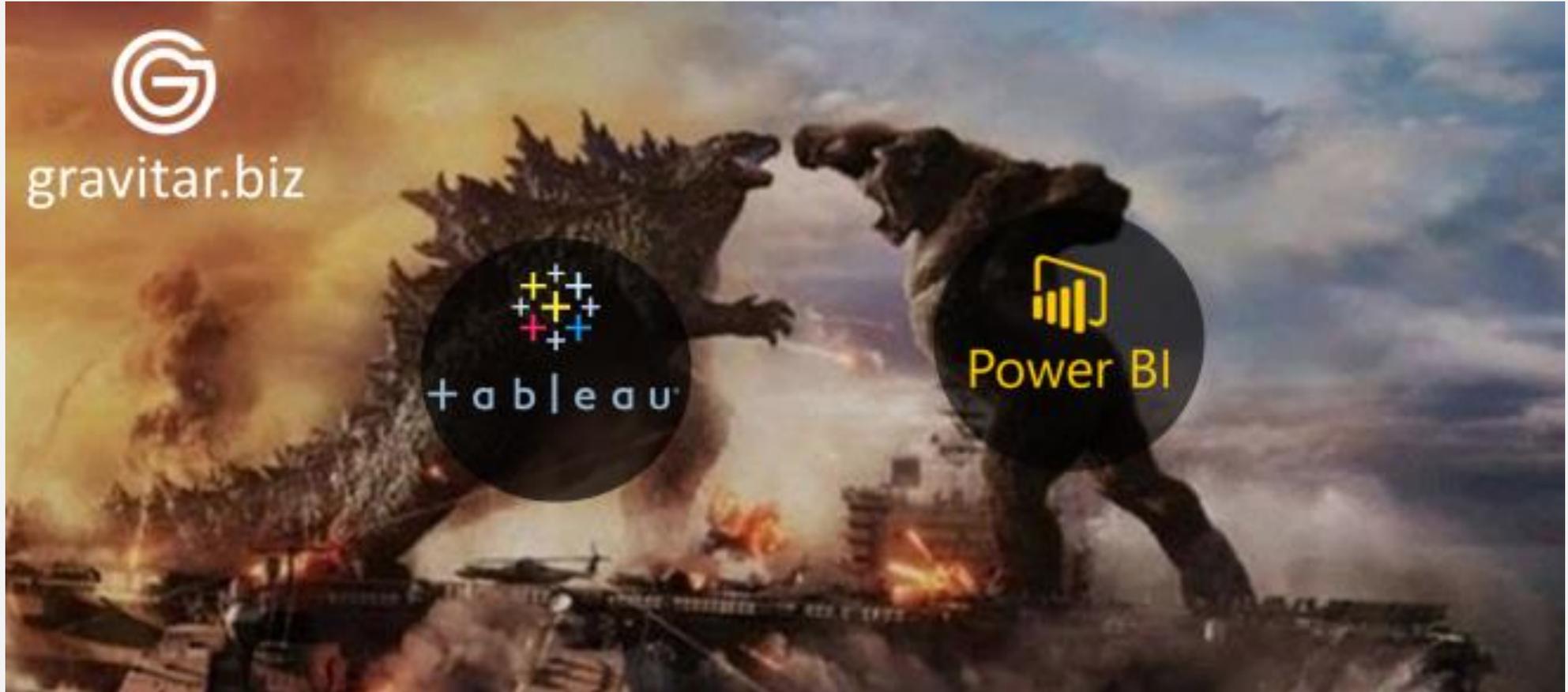




gravitar.biz



Power BI





gravitar.biz



#3

PRINCIPIOS Y
TÉCNICAS PARA LA
VISUALIZACIÓN DE
DATOS
INTERACTIVA

PRINCIPIOS DE LA VISUALIZACIÓN INTERACTIVA DE DATOS

Lima, Manuel. 2017. "Data Visualization - Material Design."
<https://material.io/design/communication/data-visualization.html>

I. HONESTIDAD Y TRANSPARENCIA



CLARIDAD > ESTÉTICA



Una buena visualización deberá
contener todos los elementos
necesarios para su correcta
interpretación

2. FACILIDAD DE LECTURA



RESPECTO HACIA LOS CÓDIGOS DEL LECTOR



Es importante considerar los hábitos y costumbres del lector de los datos, procurando así una fácil interpretación de los datos

3. LA EXPERIENCIA DEL USUARIO AL CENTRO



LA TÉCNICA AL SERVICIO DEL
RELATO



La elegancia de una visualización, así
como su velocidad de carga,
contribuirán a su capacidad
informativa

4. CLARIDAD EN EL ENFOQUE GRÁFICO



EL GRAFISMO AL SERVICIO DEL
RELATO



Los elementos gráficos han de jugar
a favor de la comprensión del
fenómeno sin distracciones
innecesarias

5. SENSIBILIDAD Y ESCALABILIDAD



DISEÑO RESPONSIVO



Hay que considerar las características de los medios y los soportes en los que se representarán los datos

6. ESTRUCTURA Y CONSISTENCIA



COHERENCIA GRÁFICA

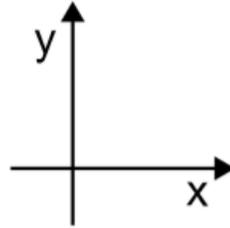


Cuanto más integración exista entre los objetos de una representación, más cómoda y familiar resultará

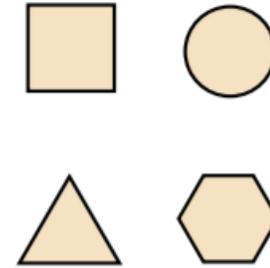
TÉCNICAS DE VISUALIZACIÓN DE DATOS

Morales i Gras. 2020. “Visualización de datos extraídos de los medios sociales.” Fundació Universitat Oberta de Catalunya (FUOC)

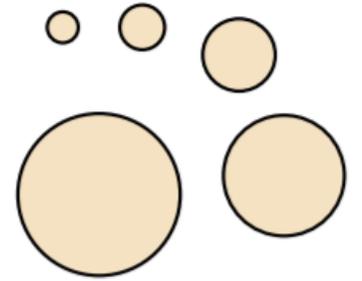
LOS ELEMENTOS BÁSICOS DE UNA VISUALIZACIÓN



posición



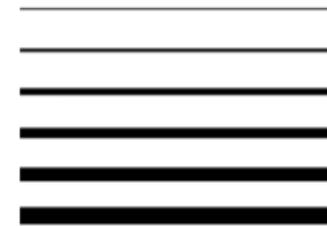
forma



medida



color



grosor de línea

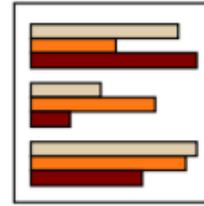


tipo de línea

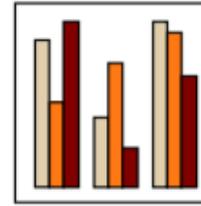
LAS VISUALIZACIONES MÁS HABITUALES

1. ¿De qué tipo de datos disponemos?
2. ¿La comprensión de qué fenómeno queremos emplazar?

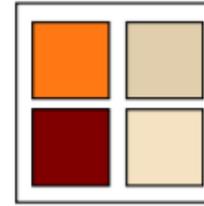
Cantidades según categorías



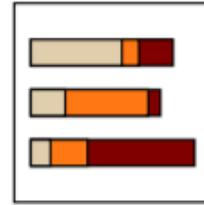
barras agrupadas



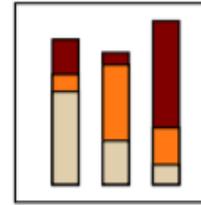
columnas agrupadas



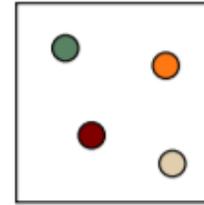
mapa de calor



barras apiladas

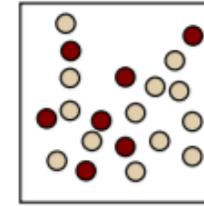


columnas apiladas

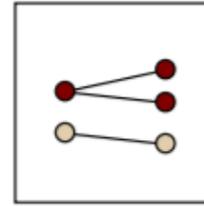


puntos

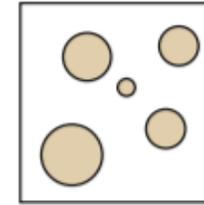
Relaciones numéricas



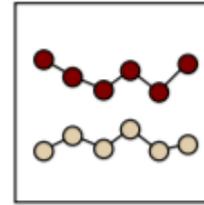
dispersión



pendiente



burbujas

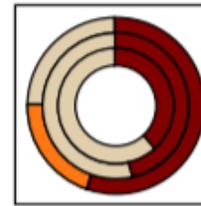


líneas

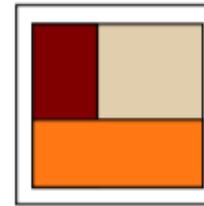
Proporciones



sectores



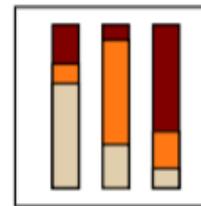
anillos



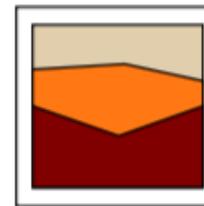
rectángulos



barras apiladas al 100 %

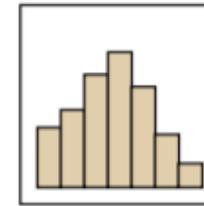


columnas apiladas 100 %

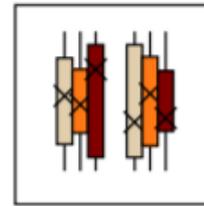


áreas aplicadas al 100 %

Distribuciones



histograma

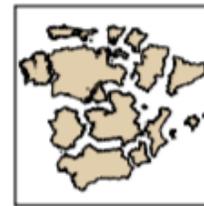


caja y bigotes

Datos geoespaciales

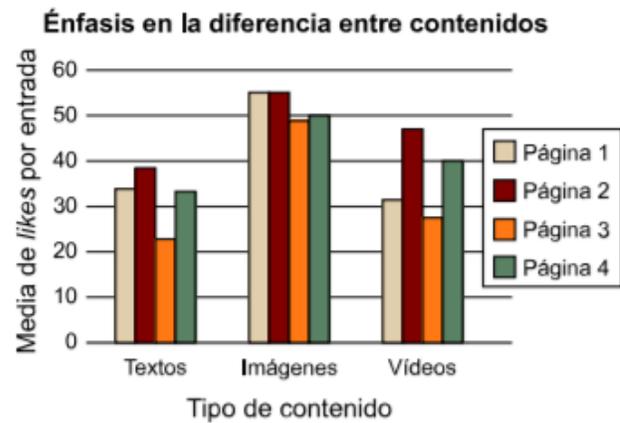
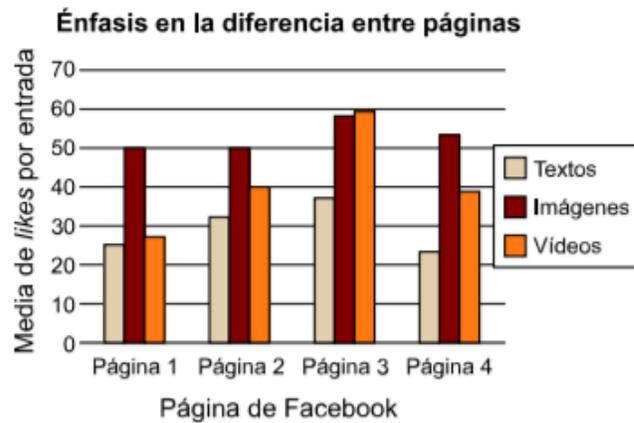


mapa coroplético



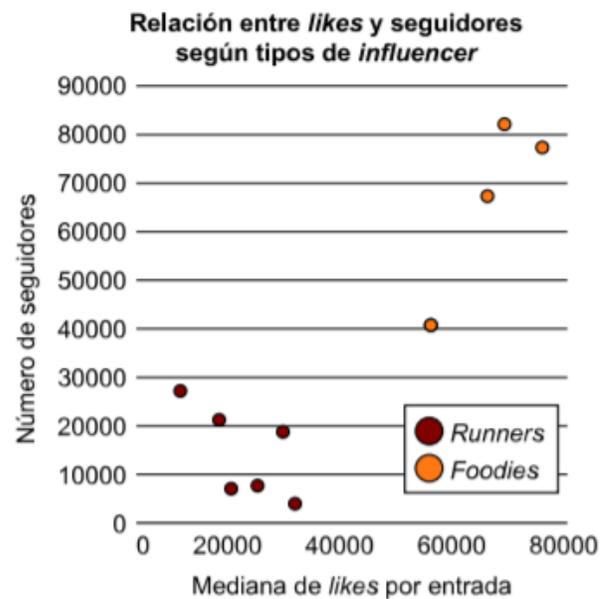
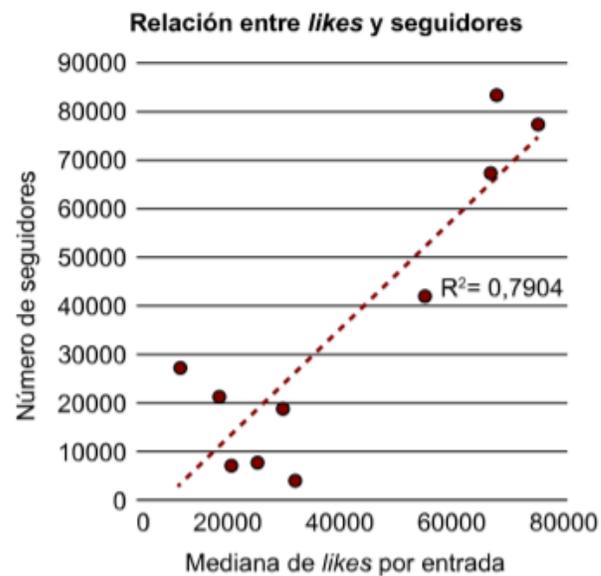
cartograma

CANTIDADES SEGÚN CATEGORÍAS



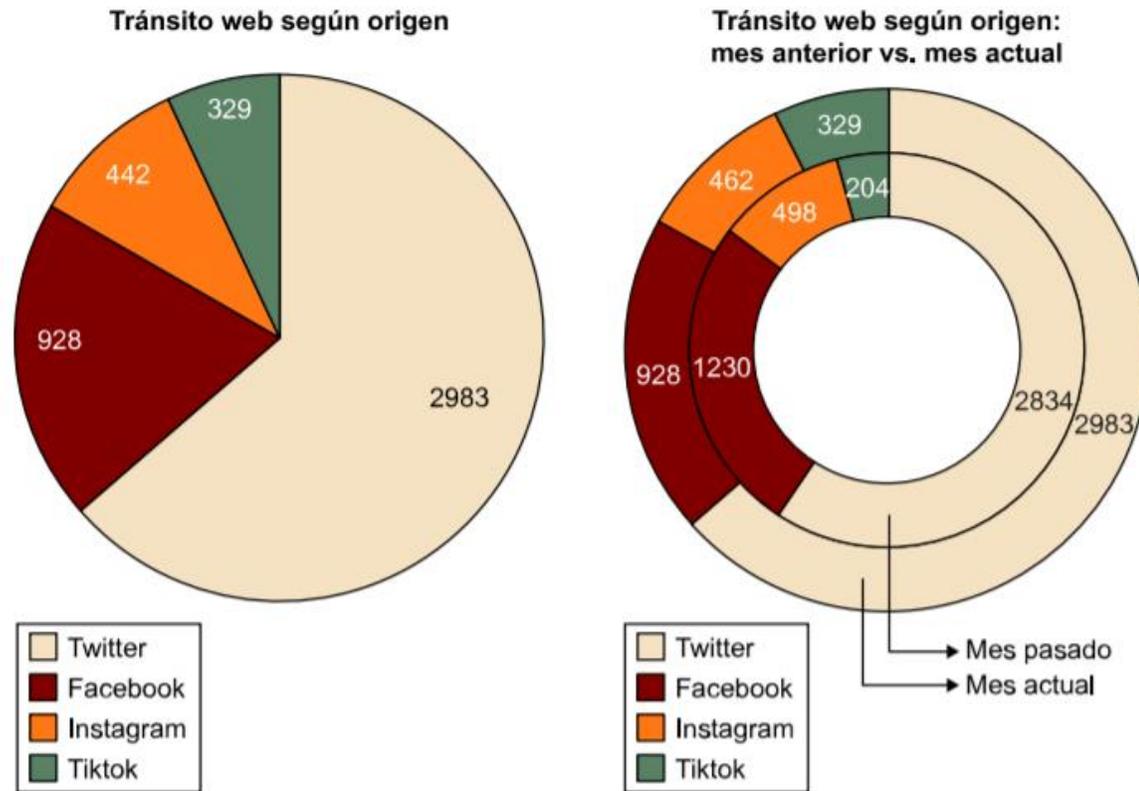
- Gráficos de barras
- Gráficos de columnas
- Gráficos de puntos
- Mapa de calor

RELACIONES NUMÉRICAS



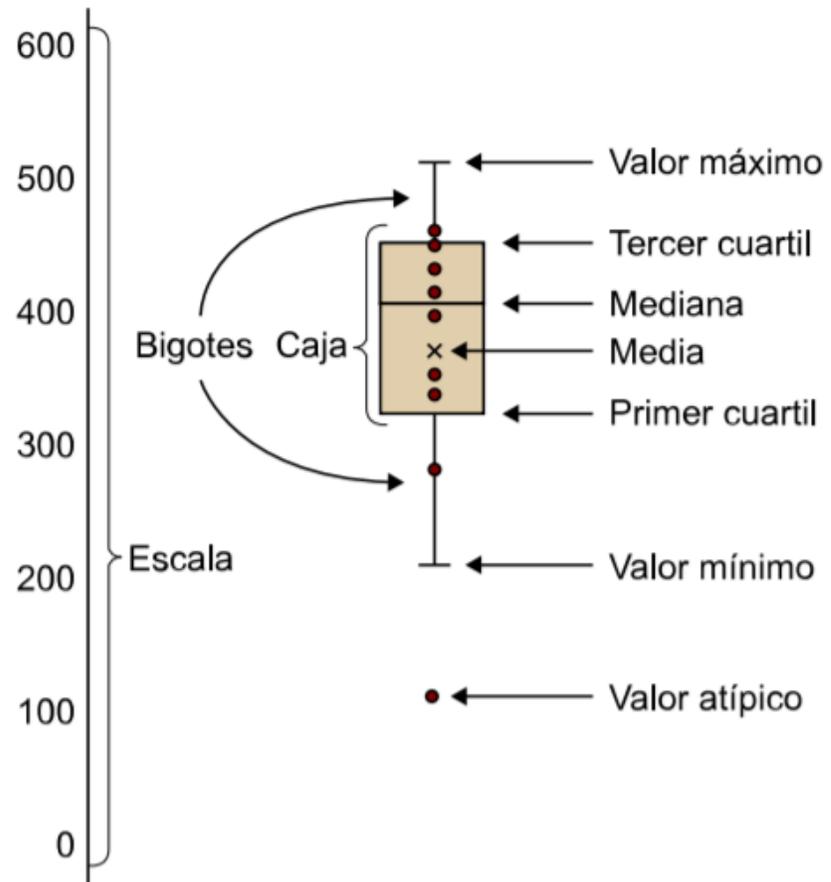
- Diagrama de dispersion
- Gráfica de burbujas
- Gráfica de pendiente
- Gráfica de líneas

PROPORCIONES



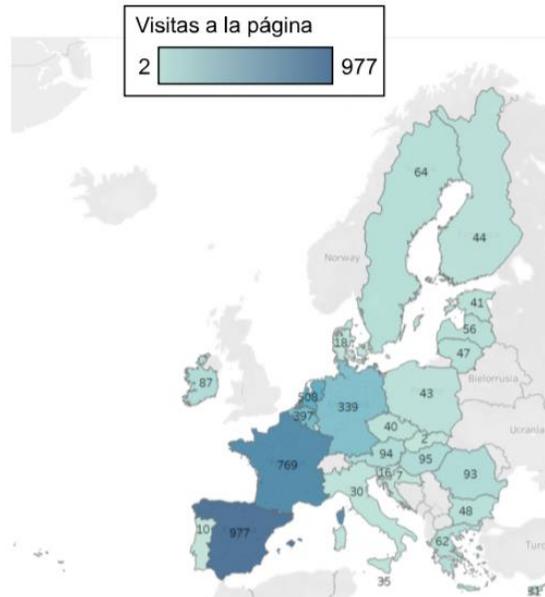
- Diagrama de sectores
- Gráfica de anillas
- Gráfica de rectángulos
- Gráfica de columnas o barras apiladas

DISTRIBUCIONES



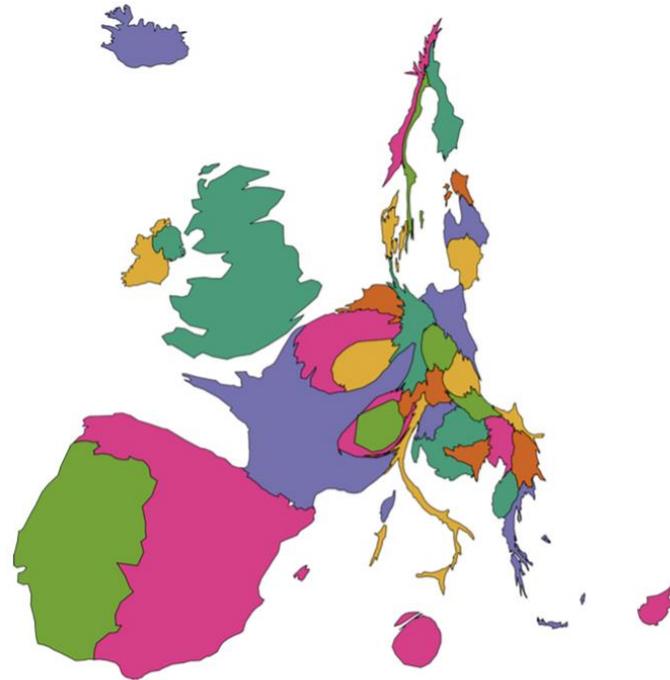
- Histograma
- Diagrama de caja y bigotes

Figura 18. Total de visitas a una página de Facebook por países



Fuente: elaboración propia con Tableau Public (datos aleatorios).

Figura 19. Total de visitas a una página de Facebook por países

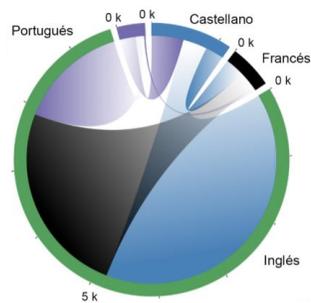


Fuente: elaboración propia con go-cart.io (Michael T. Gastner, Vivien Seguy y Pratyush More, 2018; datos aleatorios)

DATOS GEOESPACIALES

- Mapa coroplético
- Cartograma

OTRAS VISUALIZACIONES

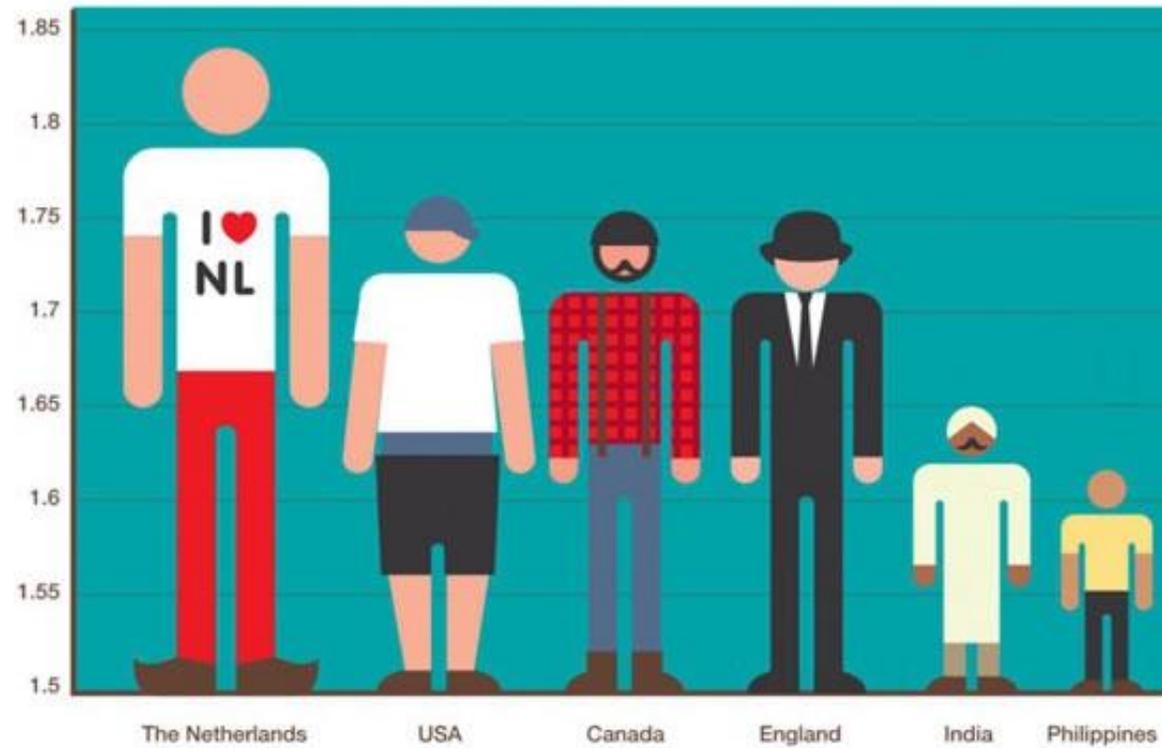


- Diagrama de Shankey
- Diagrama radial
- Diagrama de cuerdas
- Grafo
- ...

ERRORES Y TRAMPAS A EVITAR

LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



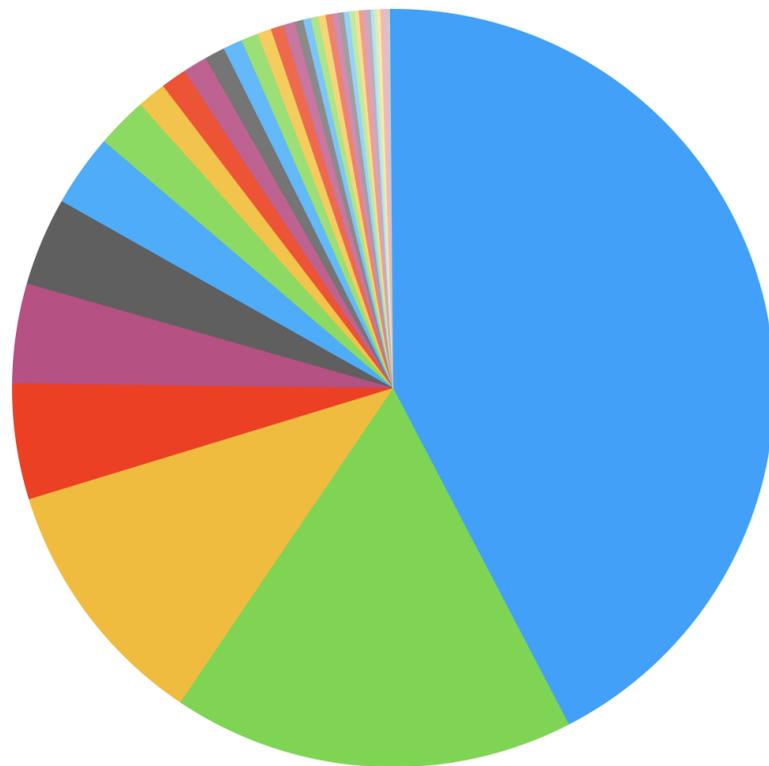
KANYE \$1.3 BILLION NET-WORTH BREAKDOWN



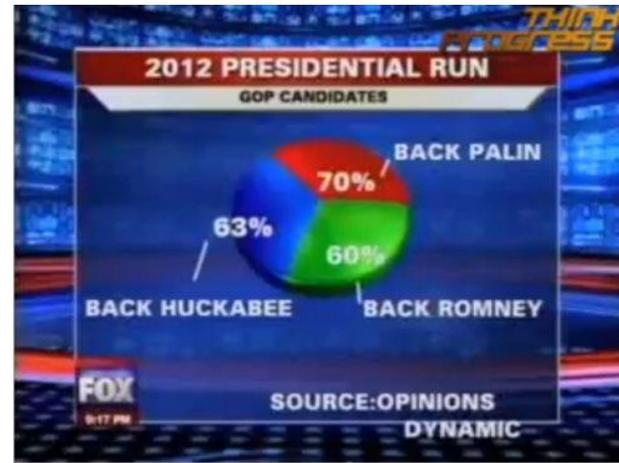
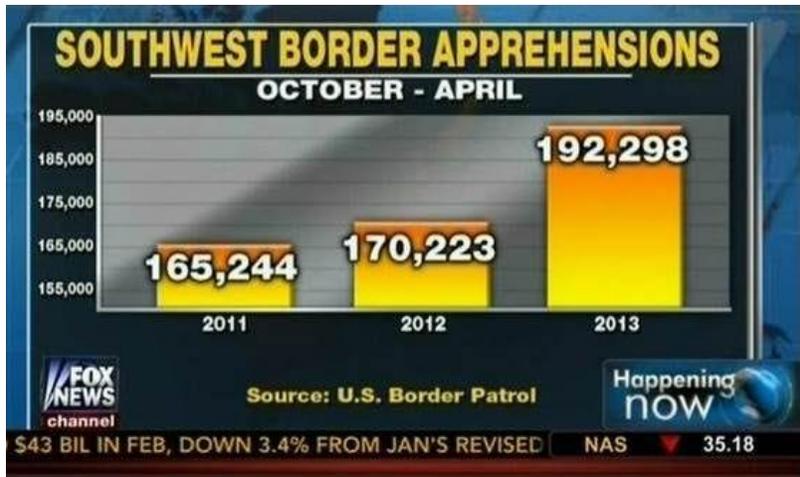
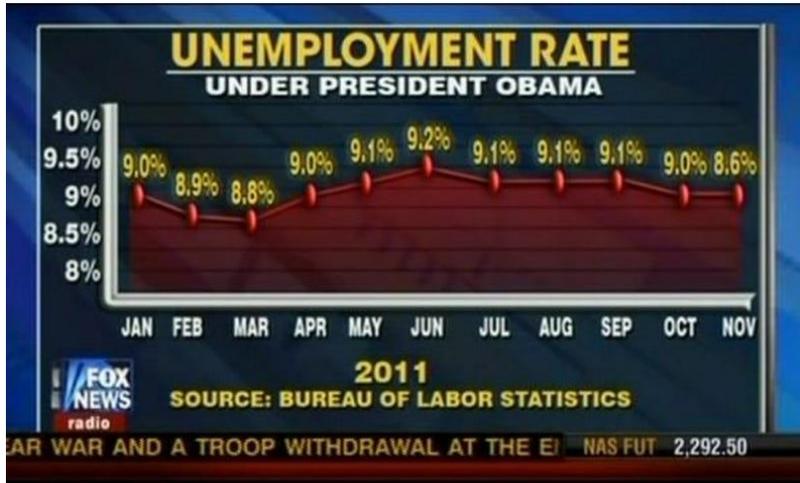


Which game(s) have you played the most?

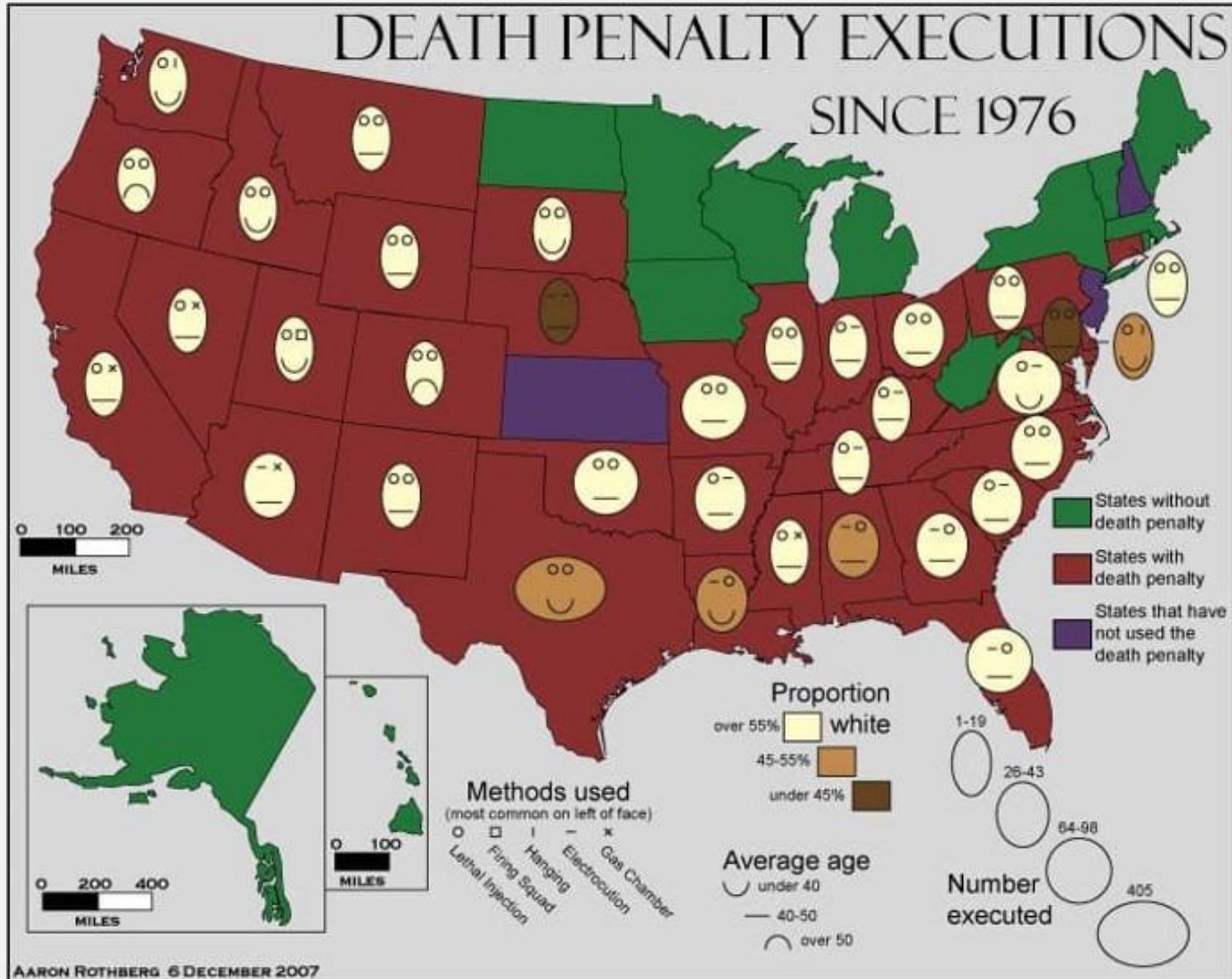
3,994 responses

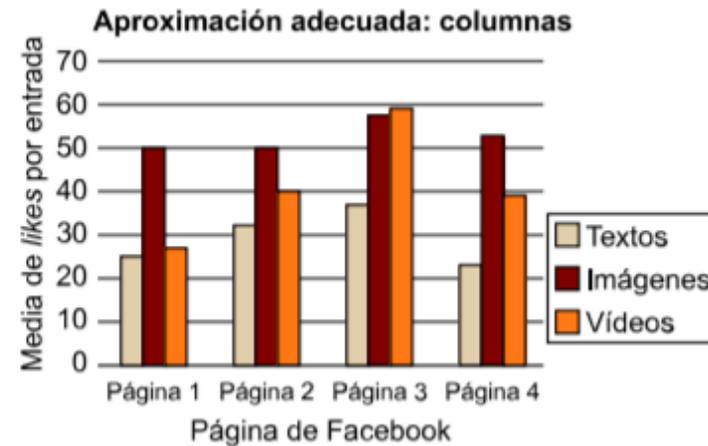
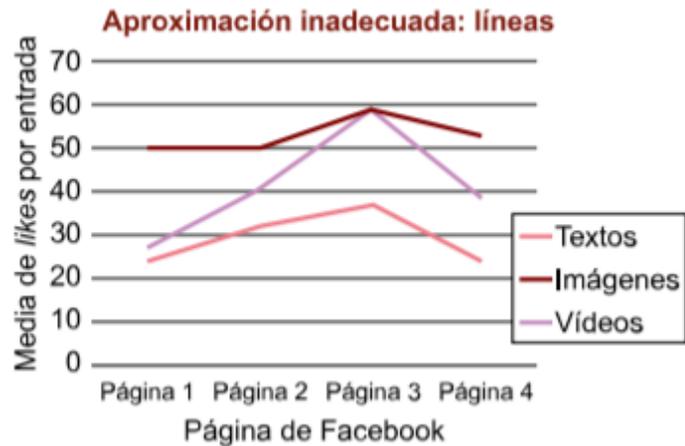


- Zelda
- The Legend of Zelda: Breath of the Wild
- Breath of the Wild
- BOTW
- Botw
- Breath of the wild
- BotW
- zelda
- Legend of Zelda: Breath of the Wild
- Legend of Zelda
- Zelda BOTW
- BoTW
- botw
- Zelda: Breath of the Wild
- Zelda BotW
- Zelda Breath of the Wild
- The Legend of Zelda
- Breath of The Wild
- The Legend of Zelda Breath of the Wild
- Zelda: BOTW
- Zelda: BotW
- Breath of the Wild
- Zelda breath of the wild
- Breath Of The Wild
- Legend of Zelda Breath of the Wild
- LoZ
- LoZ: BotW
- Zelda botw
- zelda botw
- breath of the wild
- Legend of zelda
- legend of zelda
- LoZ BOTW
- The Legend of Zelda: Breath of The Wild
- The legend of Zelda: breath of the wild
- ZELDA
- Zelda: BoTW



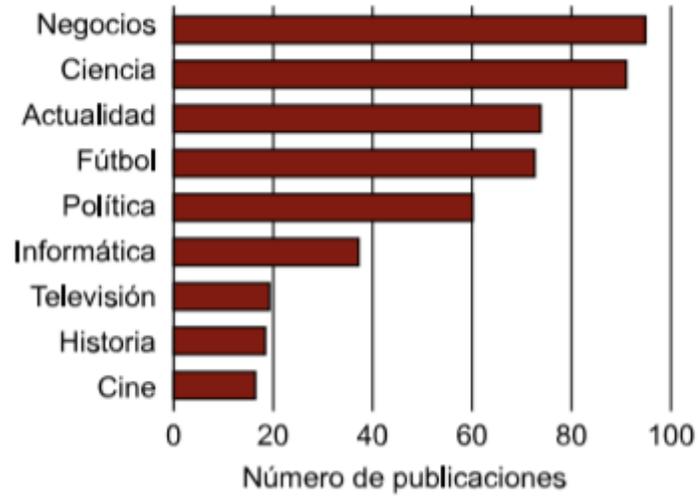
DEATH PENALTY EXECUTIONS SINCE 1976



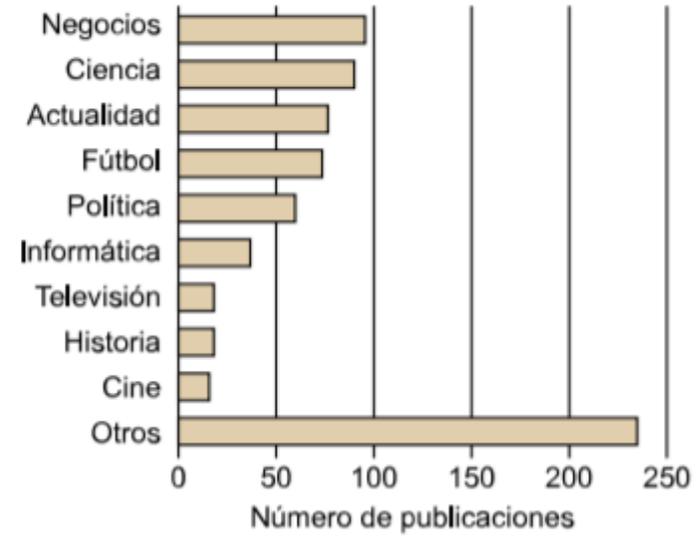


EVITAR LA CONFUSIÓN ENTRE VARIABLES CONTINUAS Y DISCRETAS

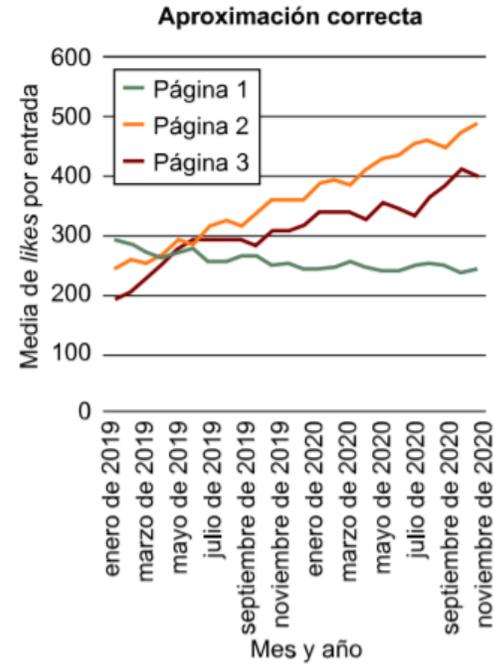
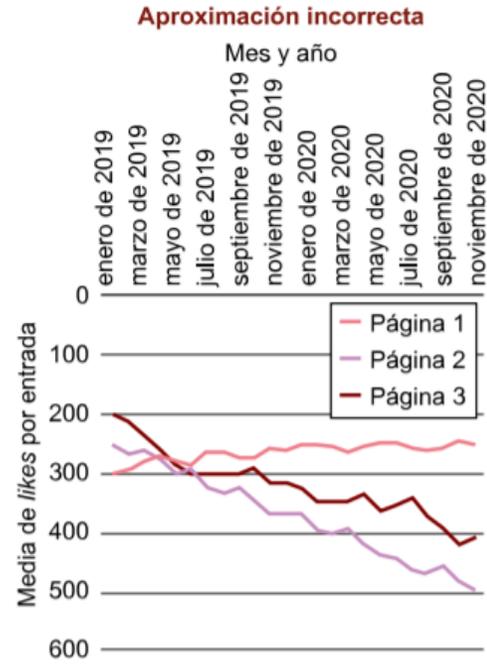
Aproximación inadecuada: sin otros



Aproximación adecuada: con otros

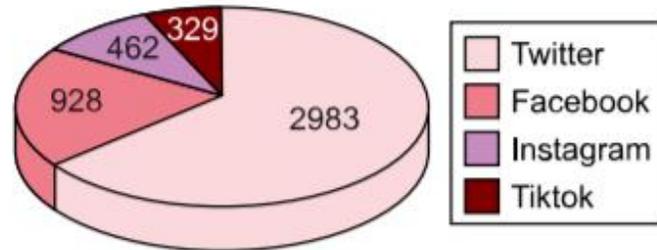


EVITAR LA INVISIBILIZACIÓN DE CASOS

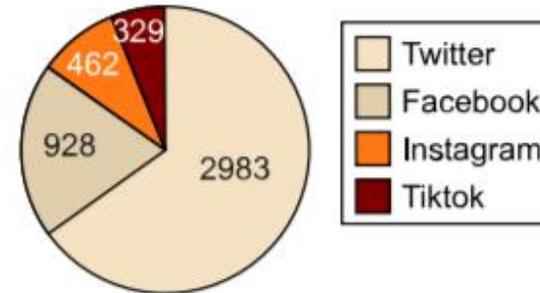


EVITAR LA RUPTURA INNECESARIA DE CONVENCIONES

Aproximación inadecuada al tráfico web según origen

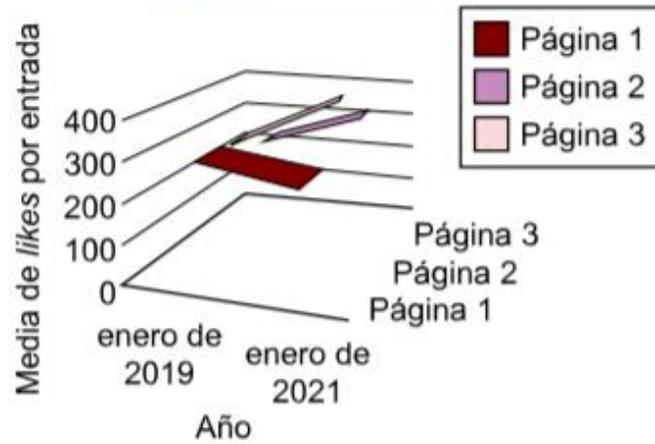


Aproximación adecuada al tráfico web según origen

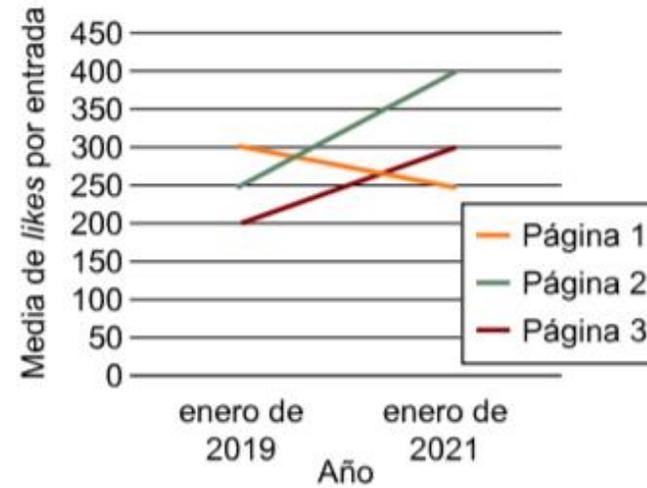


EVITAR LAS TRES DIMENSIONES

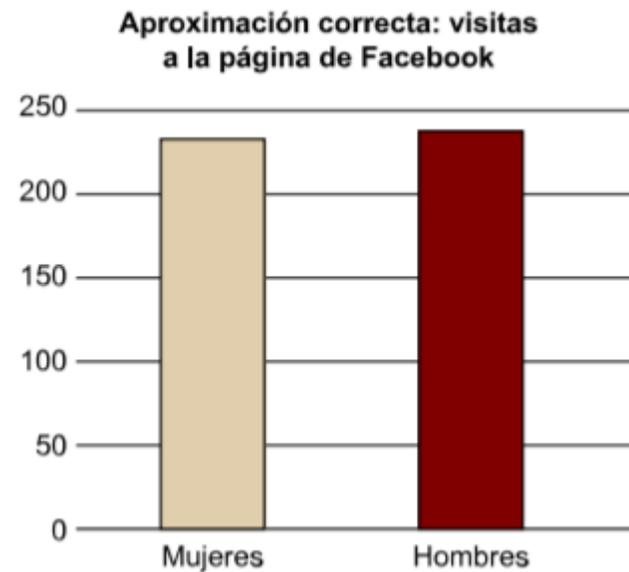
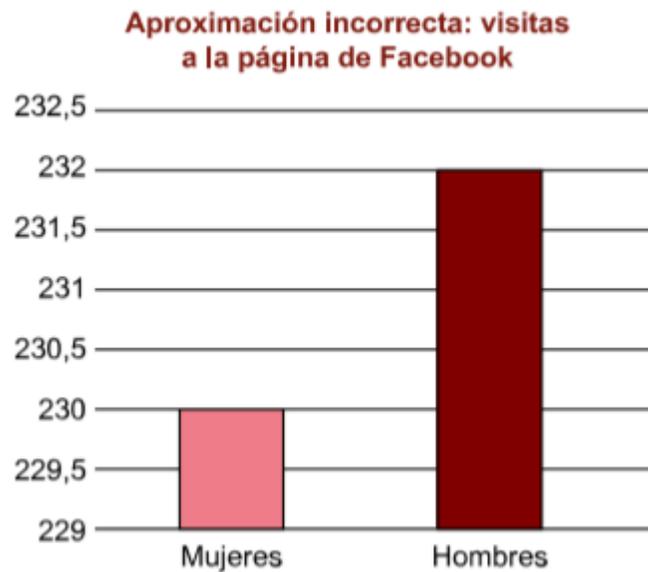
Aproximación inadecuada a la evolución del compromiso en tres páginas de Facebook



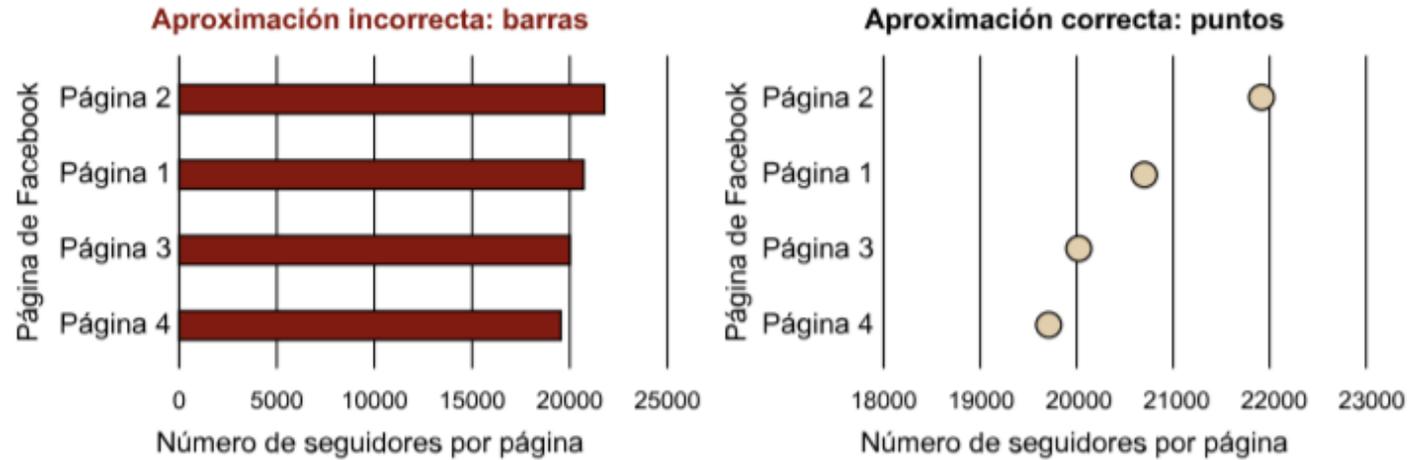
Aproximación adecuada a la evolución del compromiso en tres páginas de Facebook



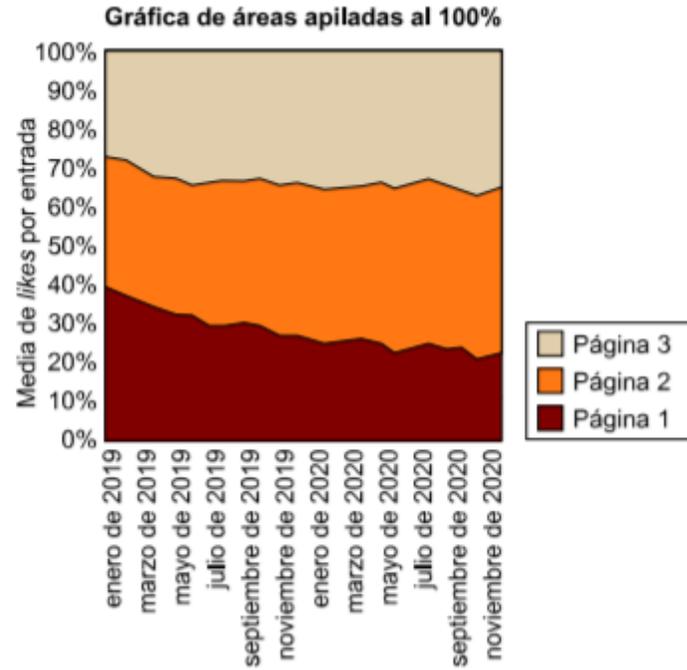
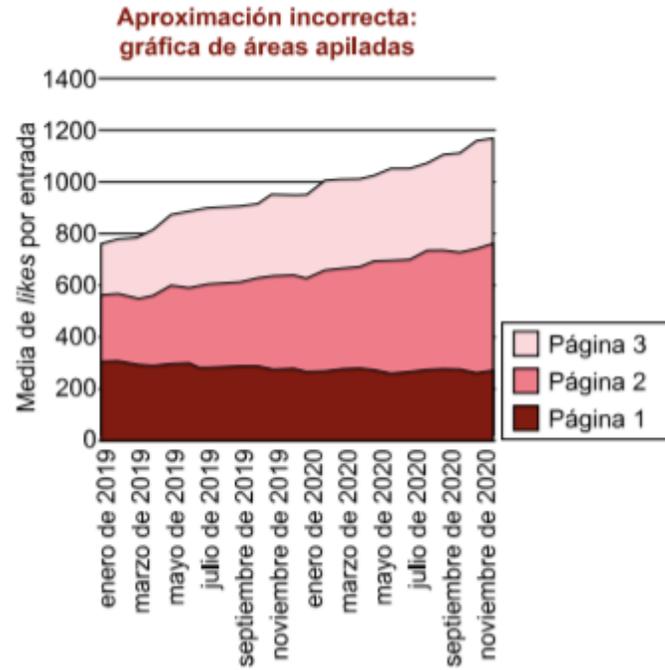
!!!EVITAR LAS TRES DIMENSIONES!!!



EVITAR EL RECORTE DE EJES EN LAS REPRESENTACIONES CON BARRAS O COLUMNAS



SI QUEREMOS ENFATIZAR DIFERENCIAS PEQUEÑAS: PUNTOS > BARRAS/COLUMNAS



**ELEGIR LA VISUALIZACIÓN
ADECUADA PARA EL RELATO**

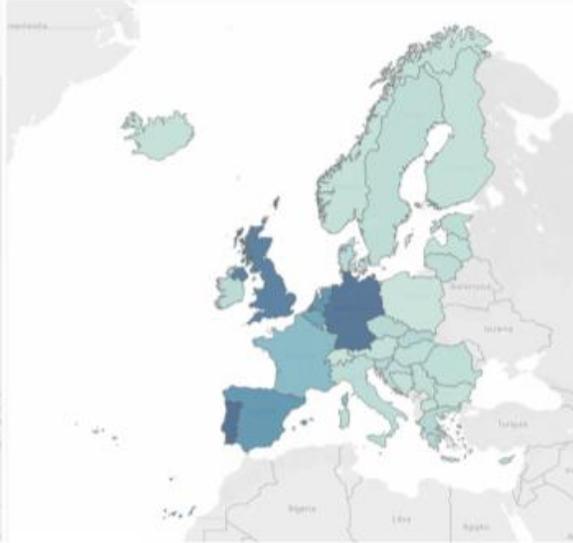
Aproximación incorrecta

Visitas a la web
5  809



Aproximación correcta

Visitas a la web
5  809



UTILIZAR UN CRITERIO LÓGICO PARA LA
PALETA DE COLOR

BIBLIOGRAFÍA Y RECURSOS

- Morales i Gras, Jordi. 2020. *Visualización de datos extraídos de los medios sociales*. Fundació Universitat Oberta de Catalunya (FUOC).
- Lima, Manuel. 2017. Material Design. Data Visualization: <https://material.io/design/communication/data-visualization.html>.
- Wilke, Claus O. 2019. *Fundamentals of Data Visualization. A Primer on Making Informative and Compelling Figures*. Sebastopol, CA: O'Reilly Media.

#4

INTRODUCCIÓN A POWERBI

PRÁCTICA I
LOS PRIMEROS PASOS CON EL SOFTWARE

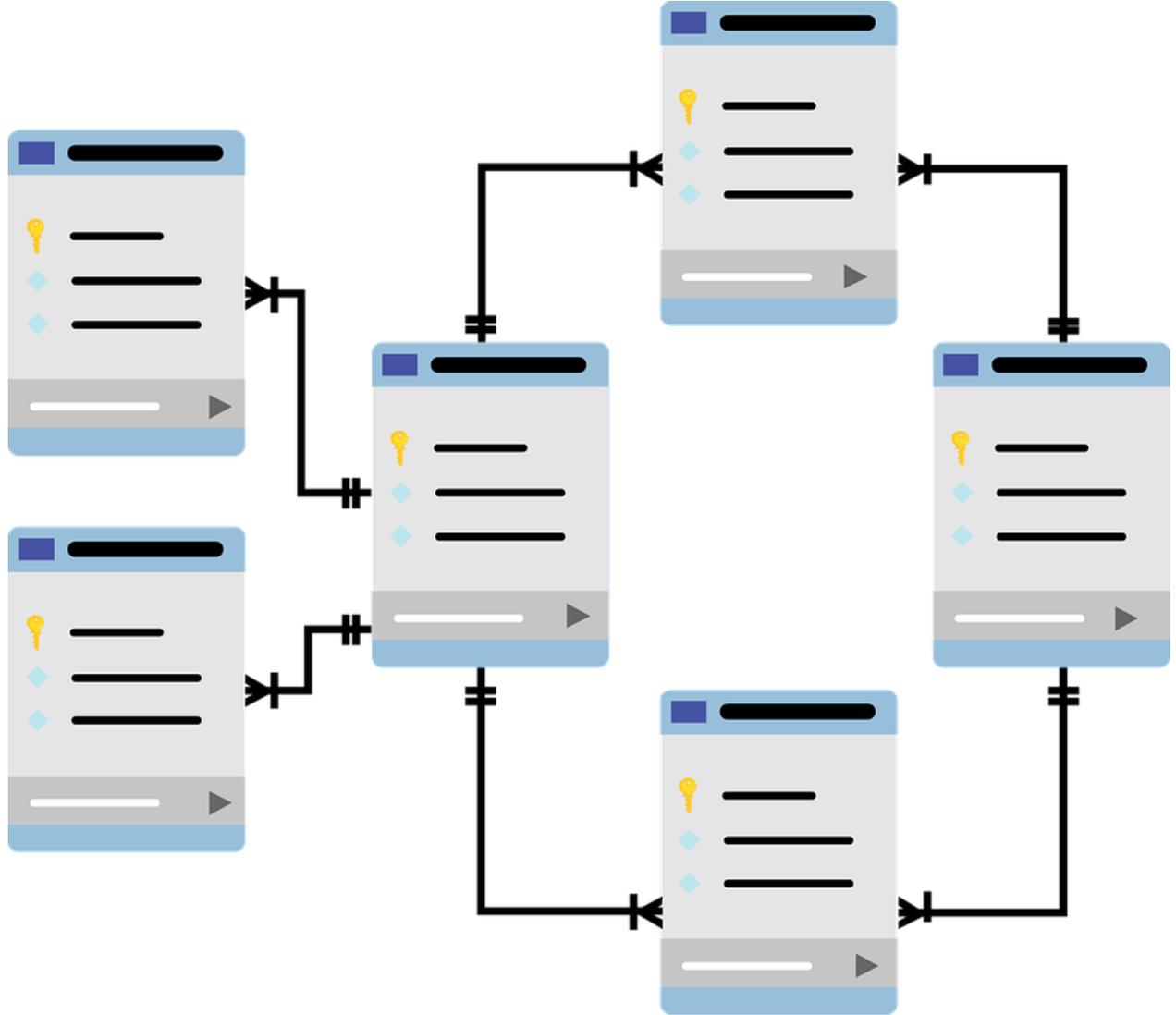
RETO I

- Crea un dashboard con las visualizaciones necesarias y responde las siguientes preguntas:
 - ¿Cuál fue el post con más reacciones durante el mes de abril?
 - ¿Cuál fue el día con más posts publicados?
 - ¿Cuántas comparticiones tuvieron los posts publicados el 6 de marzo?

#5

DATA MODELING
CON POWERBI

BI \approx SQL



RELACIONES Y
CARDINALIDAD

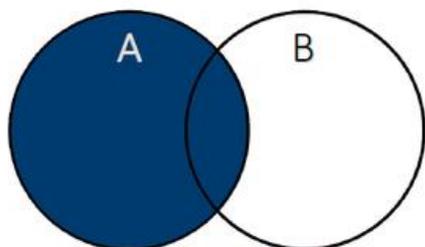
EMPLEADOS

Id	DNI	Nombre	Apellido
1	30094089S	Eugeni	Belda
2	78208028R	Gemma	Pulido
3	74579803A	Izaskun	Alfonso

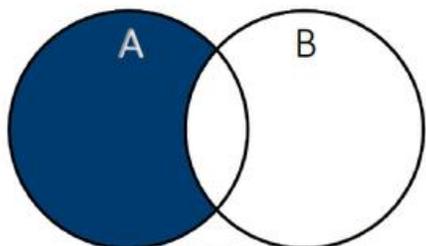
VENTAS

Empleado	Producto	Unidades	Precio
1	abc001	8	340,00 €
1	def001	6	120,00 €
2	def001	9	180,00 €
3	xyz003	8	80,00 €

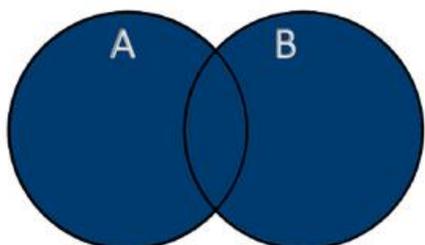




LEFT INCLUSIVE



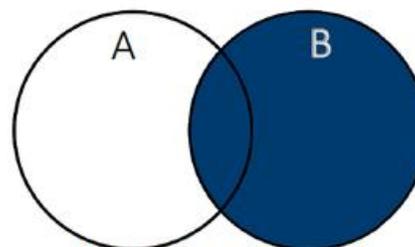
LEFT EXCLUSIVE



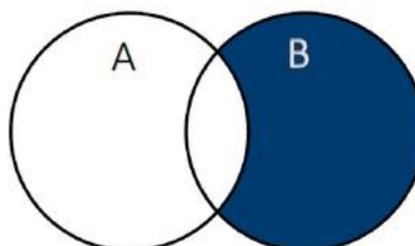
FULL OUTER INCLUSIVE

SQL JOINS

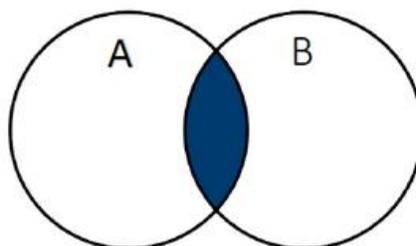
LEFT INCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key	RIGHT INCLUSIVE SELECT [Select List] FROM TableA A RIGHT OUTER JOIN TableB B ON A.Key= B.Key
LEFT EXCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key WHERE B.Key IS NULL	RIGHT EXCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key WHERE A.Key IS NULL
FULL OUTER INCLUSIVE SELECT [Select List] FROM TableA A FULL OUTER JOIN TableB B ON A.Key = B.Key	FULL OUTER EXCLUSIVE SELECT [Select List] FROM TableA A FULL OUTER JOIN TableB B ON A.Key = B.Key WHERE A.Key IS NULL OR B.Key IS NULL
INNER JOIN SELECT [Select List] FROM TableA A INNER JOIN TableB B ON A.Key = B.Key	



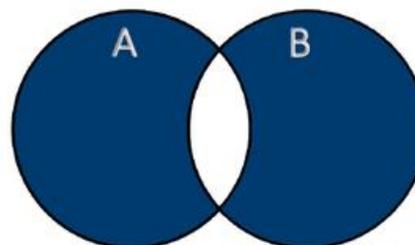
RIGHT INCLUSIVE



RIGHT EXCLUSIVE



INNER JOIN



FULL OUTER EXCLUSIVE

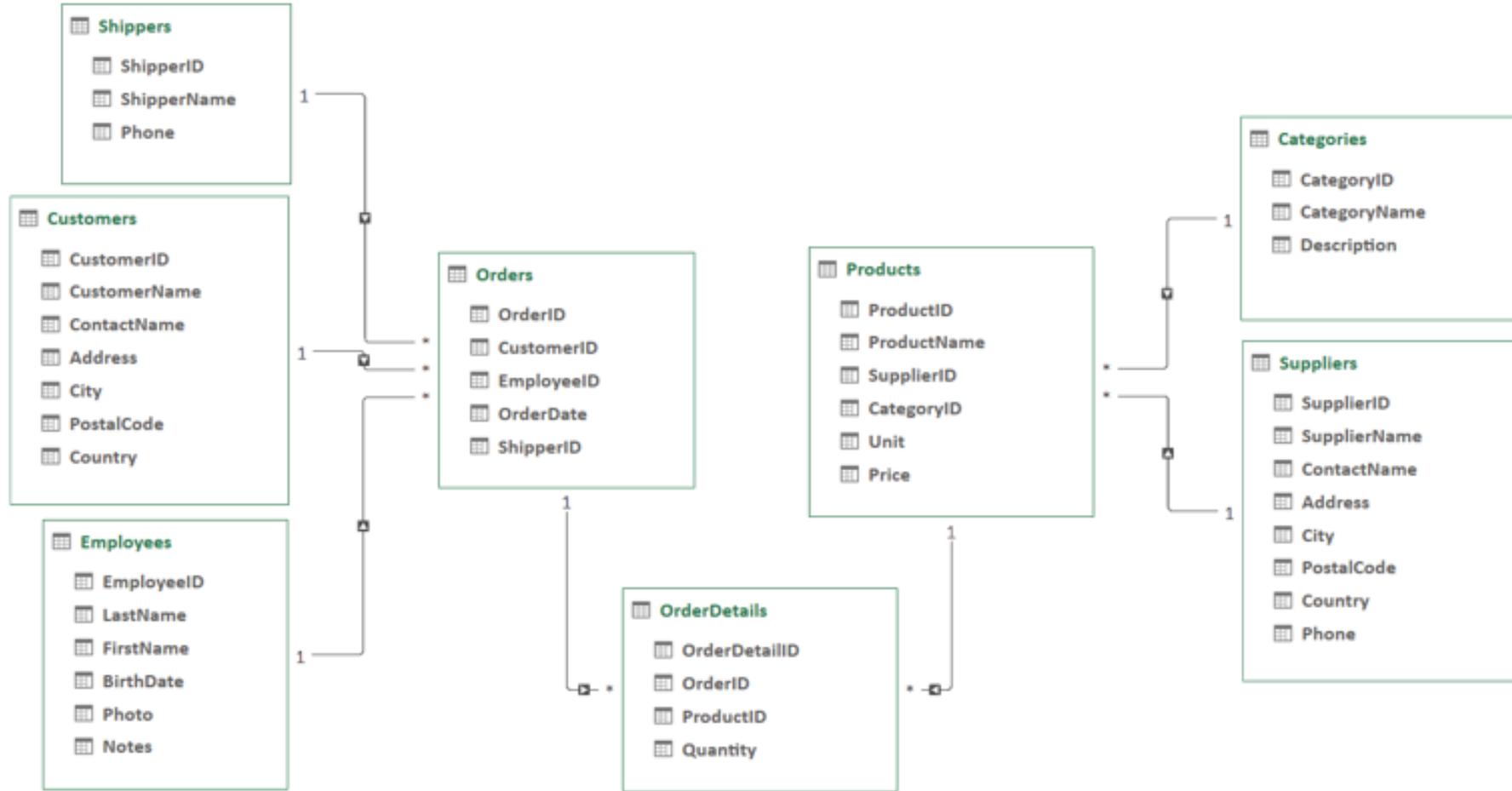
PRÁCTICA 2

CRUZANDO DATOS

RETO 2

- Crea un diagrama de sectores con la proporción de comentarios según el género de sus autores.
- Crea un dashboard interactivo que contenga el ranking de los posts más comentados y las proporciones de género de sus autores y de sentimiento detectado en los textos.
- Identifica el 2º post con más comentarios y responde a las siguientes preguntas:
 - ¿Qué porcentaje de comentarios tienen un sentimiento positivo?
 - ¿Qué porcentaje de comentarios han sido escritos por mujeres?

PRÁCTICA 3
MODELAJE Y VISUALIZACIÓN



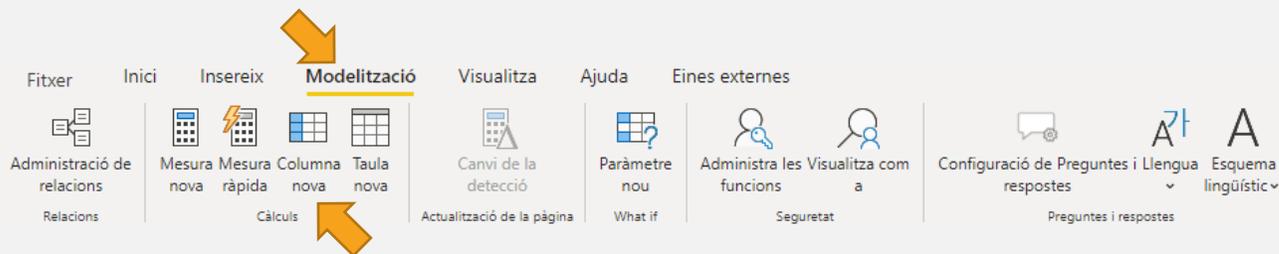
RETO 3

Crea un dashboard interactivo con las visualizaciones necesarias para responder a estas preguntas:

- ¿Cuál es el país que más pedidos ha hecho? ¿Y el que menos?
- ¿Cuál es el principal transportista para los pedidos de Alemania? ¿Y para los de Austria?
- ¿Cuánto ha facturado la empresa, en total?

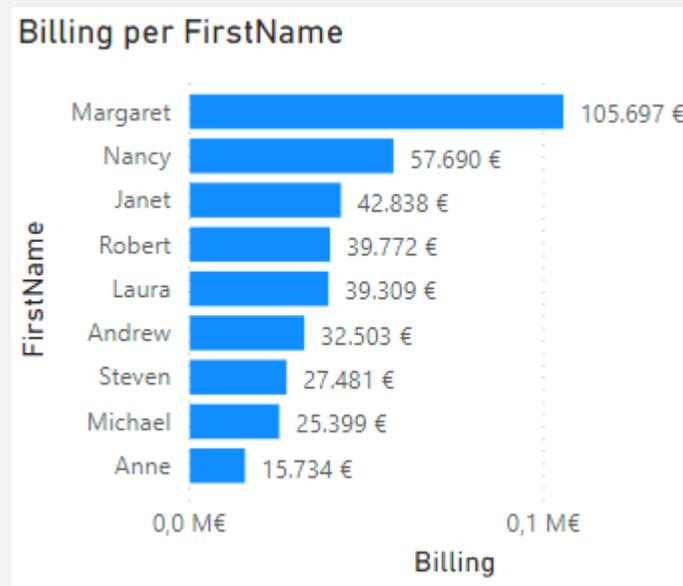
#6

CÁLCULO DE
VARIABLES CON
DAX Y
POWERQUERY



COLUMNAS CALCULADAS (DAX)

```
1 Billing = RELATED(PRODUCTS[Price]) * 'ORDER DETAILS'[Quantity]
```



PRÁCTICA 4
CÁLCULO DE VARIABLES

RETO 4

- Genera una variable para clasificar las categorías de productos veganos o no veganos
- Crea un dashboard interactivo con las visualizaciones necesarias para responder a estas preguntas:
 - Gráfica de con la facturación de cada empleado para productos veganos y no veganos
 - Gráfica de barras con la facturación de cada producto, pintando los productos en función de si son veganos o no
 - Timeline mixto (columnas y barras) con la cantidad de productos no veganos vendidos y la facturación mes a mes

#7

VISUALIZACIONES
AVANZADAS E
INTERPRETACIÓN E
DATOS CON
POWERBI

PRÁCTICA 5
VISUALIZACIONES AVANZADAS EN INTERPRETACIÓN
DE DATOS

RETO 5

Crea un dashboard interactivo con las visualizaciones necesarias para responder a estas preguntas:

- ¿Cuál fue el día con más tuits del periodo? ¿Observas algún patrón en la cronología?
- ¿Cuál ha sido el tuit con más retuits? ¿Y el tuit con más respuestas?
- ¿Cuál es la palabra más habitual en las descripciones de los autores?
- En base a la información anterior, ¿qué perfil de usuario crees que abunda en la conversación analizada?

#8

CONCLUSIONES Y
CIERRE

NOTAS FINALES

- La visualización de datos no es un aspecto menor en la cadena de valor de los datos masivos, sino que es un punto central e incluso crítico. La visualización de datos es fuente de valor.
- Una visualización funcional puede definirse como una representación visual de datos honesta, fácil de leer, agradable, clara, sensible al medio y gráficamente consistente.
- Representar mal los datos visualmente es la forma más fácil y rápida de manipular, y una de las más eficaces. Una mayor cultura de representación del dato permite una mayor y mejor fiscalización de los fenómenos de desinformación.

POWERBI Y SUS PRINCIPALES COMPETIDORES

Software	Ventajas	Desventajas
PowerBi	<ul style="list-style-type: none"> • Gran comunidad de usuarios • Versión gratis muy generosa • Compatibilidad Microsoft • Permite guardar archivos en la versión gratuita • Licencias académicas fácilmente accesibles • Facilidad de uso 	<ul style="list-style-type: none"> • Software privativo • Software de pago • Limitaciones en la versión gratuita • No apto para Mac ni Linux (salvo Server \$\$)
Tableau	<ul style="list-style-type: none"> • Largo desarrollo • Gran comunidad de usuarios • Versión pública bastante generosa • Licencias académicas fácilmente accesibles • Facilidad de uso 	<ul style="list-style-type: none"> • Software privativo • Software de pago • Limitaciones en la versión pública • No permite guardar archivos en la versión pública • Lenguaje propio • No apto para Linux (salvo Server \$\$)
DataStudio	<ul style="list-style-type: none"> • Software gratuito • Gran comunidad de usuarios • Compatibilidad Google • Facilidad de uso 	<ul style="list-style-type: none"> • Software privativo • Limitaciones importantes en el volumen de datos • Sin opción premium
Grafana	<ul style="list-style-type: none"> • Software libre y gratuito • Versión server + Cloud (\$\$) • Compatibilidad SQL + NoSQL 	<ul style="list-style-type: none"> • Dificultad de uso: curva de aprendizaje prolongada