



Introducción a la Visualización de Datos



Alejandro González
@nihilistBird
alex@outliers.es

0. Materiales

demos.outliers.es/cursos/utad/dataScience/viz

0. Contenidos

1. Visualización de datos
2. Comunicando con datos
3. Tableau Public

1. Introducción

1. Introducción

De los datos a la sabiduría

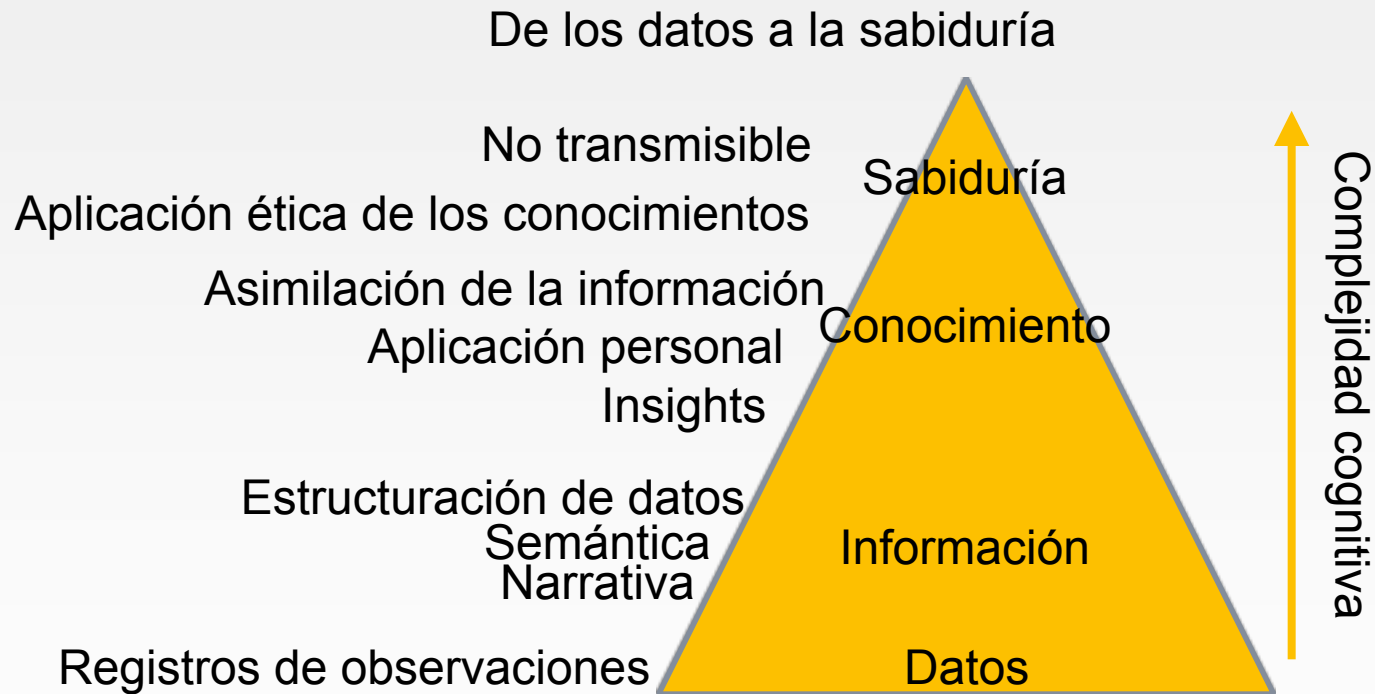
“Los datos, organizados y empleados debidamente, pueden convertirse en información.

La información, absorbida, comprendida y aplicada por las personas, puede convertirse en conocimientos.

Los conocimientos aplicados frecuentemente en un campo pueden convertirse en sabiduría, y la sabiduría es la base de la acción positiva”

Michael Cooley, "Architect or Bee?" Hogarth Press, London, UK, 1987.

1. Introducción



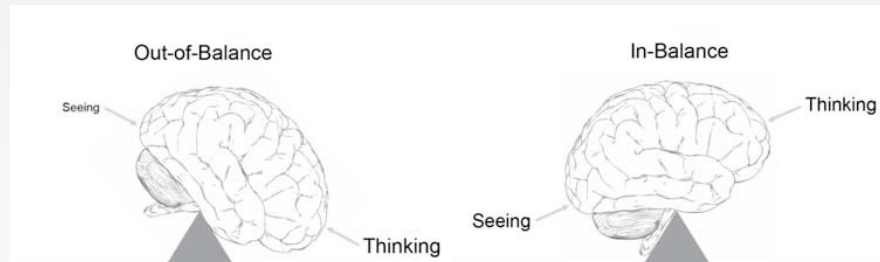
1. Introducción

¿Qué se entiende por visualizar?

“Tecnologías que transforman datos en información mediante elementos visuales”

1. Introducción

¿Por qué visualizar?



[http://www.interaction-design.org/encyclopedia/
data_visualization_for_human_perception.html](http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html)

1. Introducción

¿Por qué visualizar datos?

“Visualization is critical to data analysis. It provides a front line of attack, revealing intricate structure in data that cannot be absorbed in any other way. We discover unimagined effects, and we challenge imagined ones.”

William S. Cleveland, “Data Visualization”

1. Introducción

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

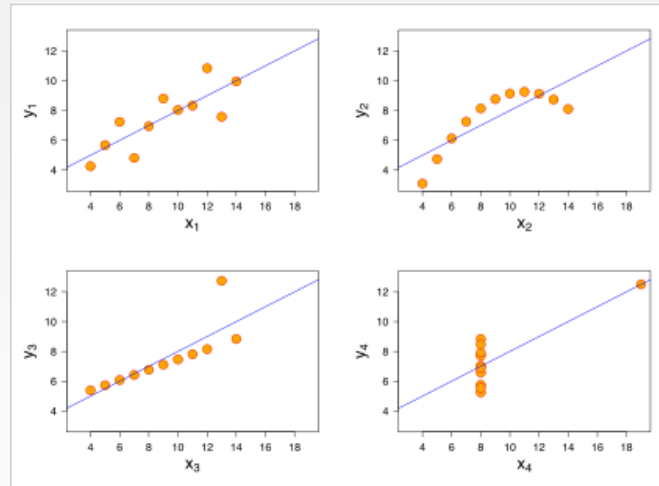
¿Por qué visualizar datos?

Property	Value
Mean of x in each case	9 (exact)
Variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

Anscombe's Quartet, http://en.wikipedia.org/wiki/Anscombe's_quartet

1. Introducción

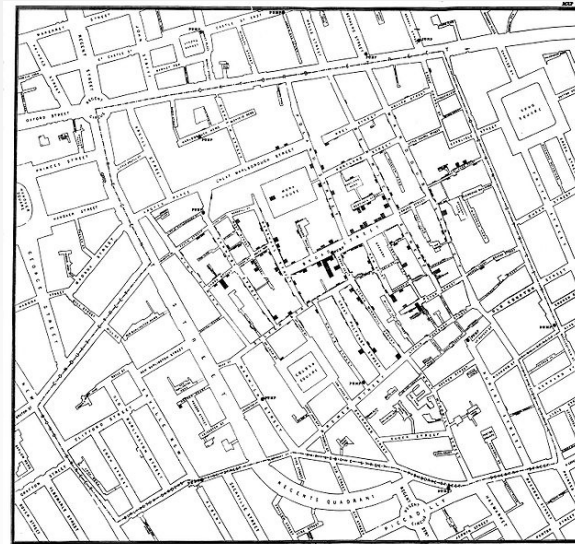
¿Por qué visualizar datos?



Anscombe's Quartet, http://en.wikipedia.org/wiki/Anscombe's_quartet

1. Introducción

¿Por qué visualizar datos?



Brote de cólera de 1854 en la calle Broad,
http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

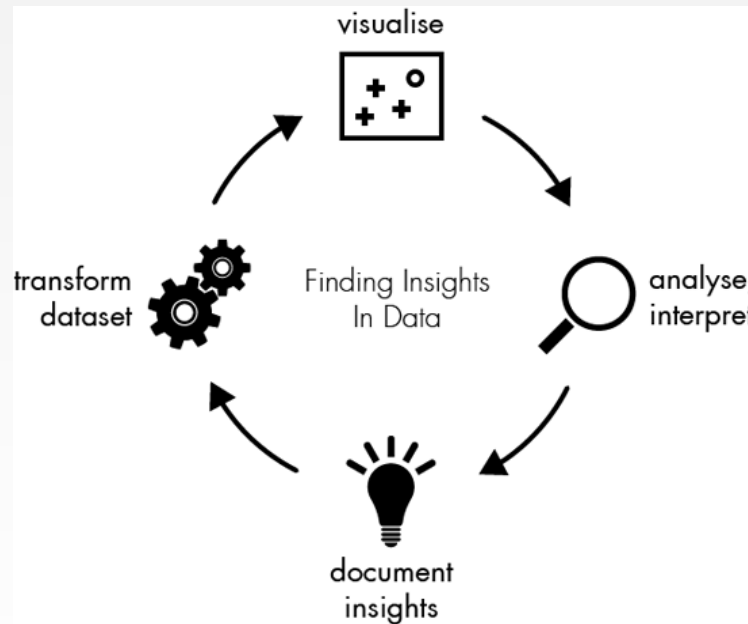
1. Introducción

¿Por qué visualizar datos en 2015?

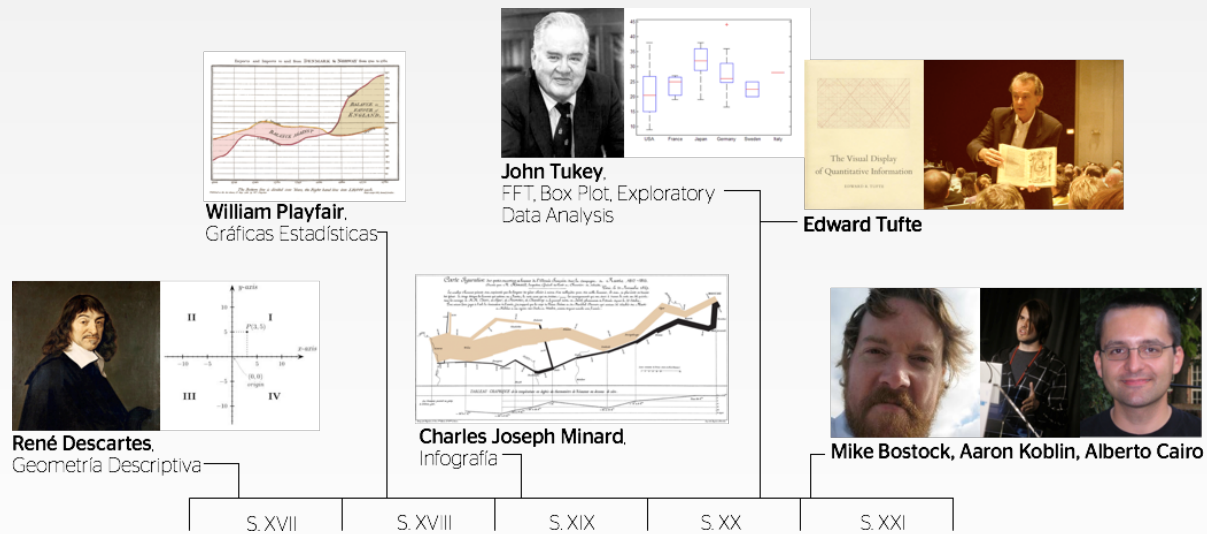
- “Big Data” necesita “Big Insights”
- Explosión de herramientas
- Reconocimiento como disciplina imprescindible en el análisis de datos
- Necesidades en múltiples campos
- Tendencia fuerte: Tiempo Real
- El ciudadano como productor de datos
- “Quantified Self”
- Business Analytics
- Necesidad de consumo rápido

1. Introducción

Fases del trabajo de visualización



1. Introducción



1. Introducción

Diseño de la Visualización

Proceso adaptado

Maqueta de datos

Maqueta de interacción

Test de usabilidad / interacción

Capa visual

1. Introducción

Capa visual

- Coherencia de fuentes, colores y tamaños
- Percepción: Comparación
- Percepción: Agrupación


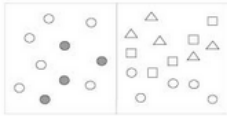
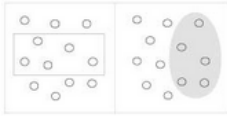

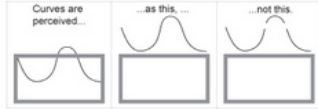

1. Introducción

Capa visual: Comparación

1. Position along a common scale
2. Position along identical, nonaligned scales
3. Length
4. Angle — slope
5. Area
6. Volume
7. Colour hue — Colour saturation — Density

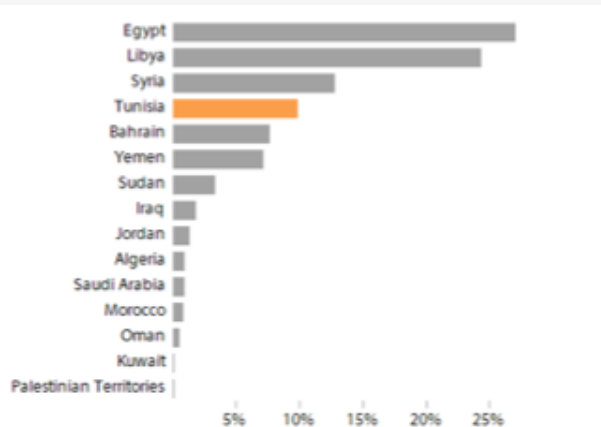
1. Introducción

Capa visual: Agrupación

Proximity	Objects that are close together are perceived as a group.	
Similarity	Objects that share similar attributes (e.g., color or shape) are perceived as a group.	
Enclosure	Objects that appear to have a boundary around them (e.g., formed by a line or area of common color) are perceived as a group.	
Closure	Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.	
Continuity	Objects that are aligned together or appear to be a continuation of one another are perceived as a group.	
Connection	Objects that are connected (e.g., by a line) are perceived as a group.	

1. Introducción

¿Qué gráfico escojo? Gráfico de Barras



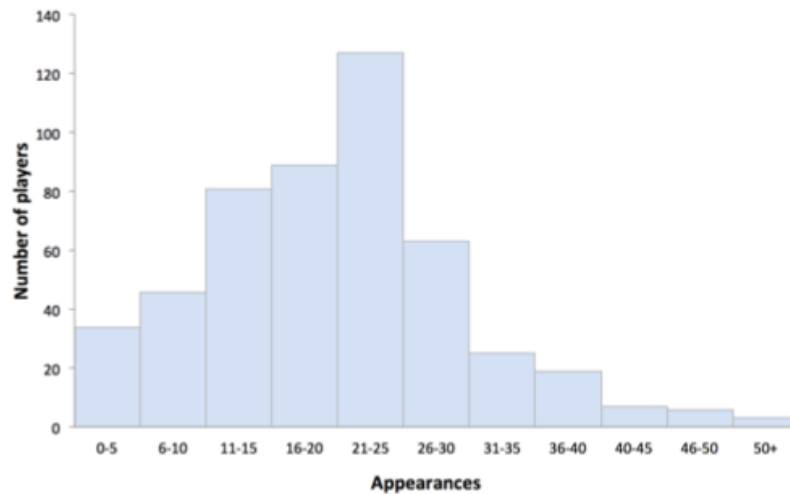
Comparación de una variable entre categorías

La disposición lineal facilita la comparación

Siempre hay que arrancar el rango desde cero, para evitar distorsiones perceptuales

1. Introducción

¿Qué gráfico escojo? Histograma

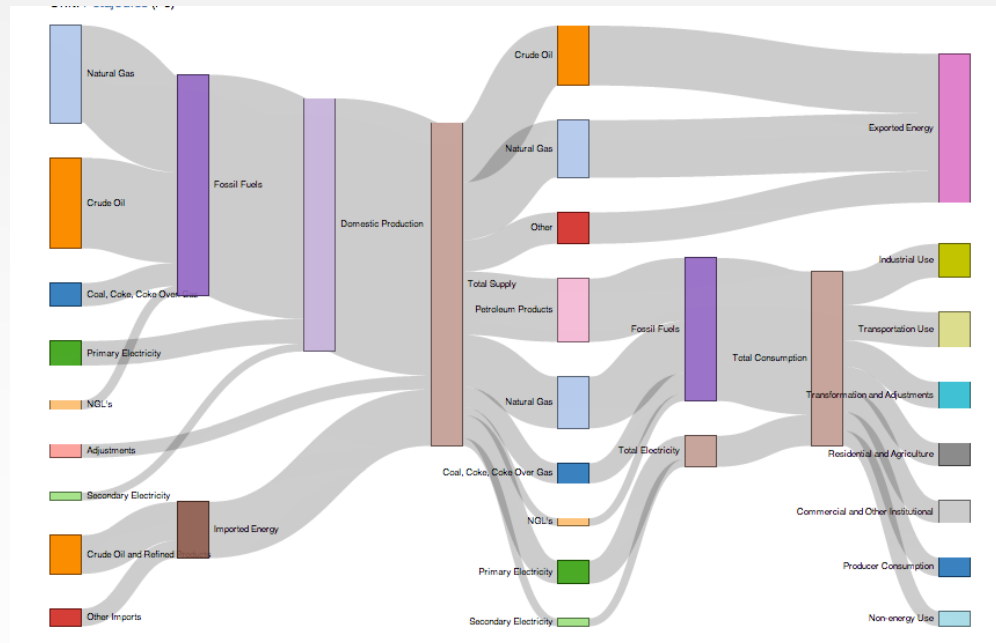


Revela la distribución estadística de un valor cuantitativo (eje Y) vs un valor cuantitativo discretizado

Desvela propiedades estadísticas de las distribuciones asociadas

1. Introducción

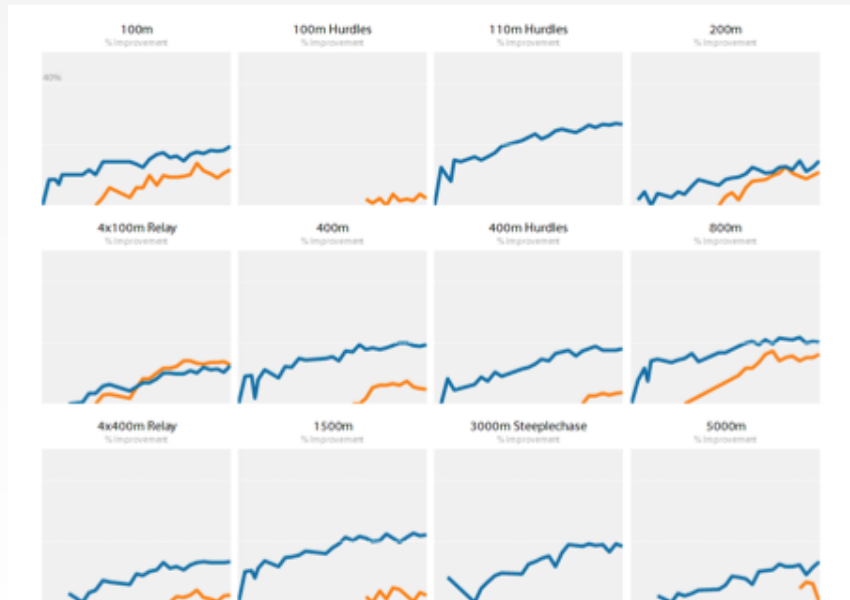
¿Qué gráfico escojo? Flujos



Flujos con diagrama de Sankey. Energía en Canadá
http://gnewton.ca/gn/sankey/sankey_canada_energy_flow_2007.html

1. Introducción

¿Qué gráfico escojo? Small multiples



Aprovechan la capacidad de nuestro sistema visual de detección de patrones

1. Introducción

¿Qué gráfico escojo? Treemap



Revelan jerarquías y magnitudes de un sólo vistazo

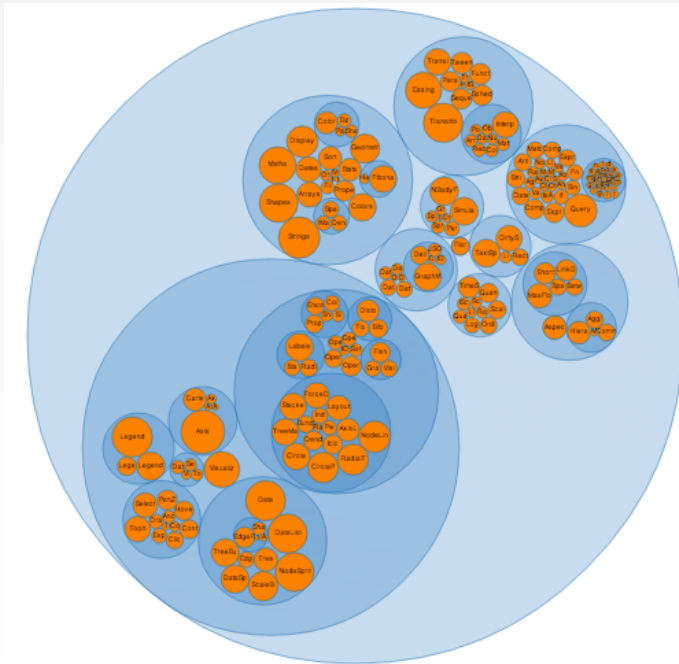
Las áreas ayudan a mejorar la percepción de la distribución.

Se pueden usar colores para superponer otra categorización

¿Qué gráfico escojo? “Circle Packing”

1. Introducción

¿Qué gráfico escojo? ‘Circle Packing’



Revelan jerarquías y magnitudes de un sólo vistazo

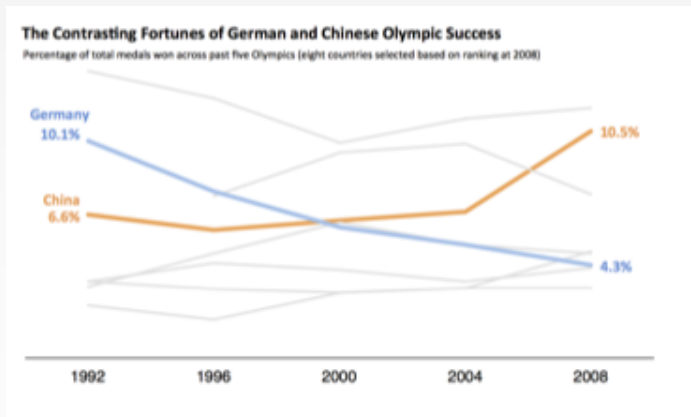
Las áreas ayudan a mejorar la percepción de la distribución.

Se pueden usar colores para superponer otra categorización o reforzar la existente

Aunque ocupan más espacio, se perciben mejor las jerarquías que con el treemap

1. Introducción

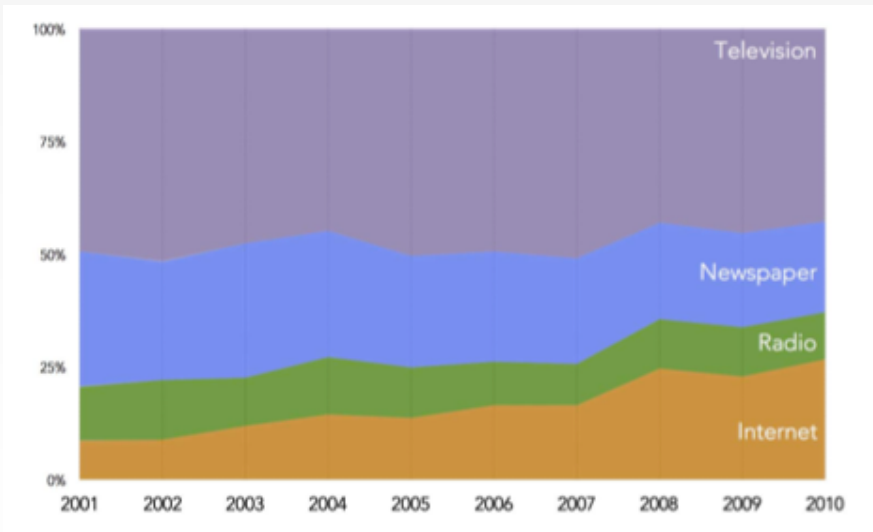
¿Qué gráfico escojo? Diagrama de “Fiebre”



Comparación de varias series temporales

1. Introducción

¿Qué gráfico escojo? “Stacked Area”



Comparación de varias series temporales

Alternativa al gráfico de tarta, con una dimensión más

Los valores de cada categoría NO son su altura absoluta, sino relativa a la categoría inferior

Por el contrario, son difíciles de leer, requieren esfuerzo cognitivo

1. Introducción

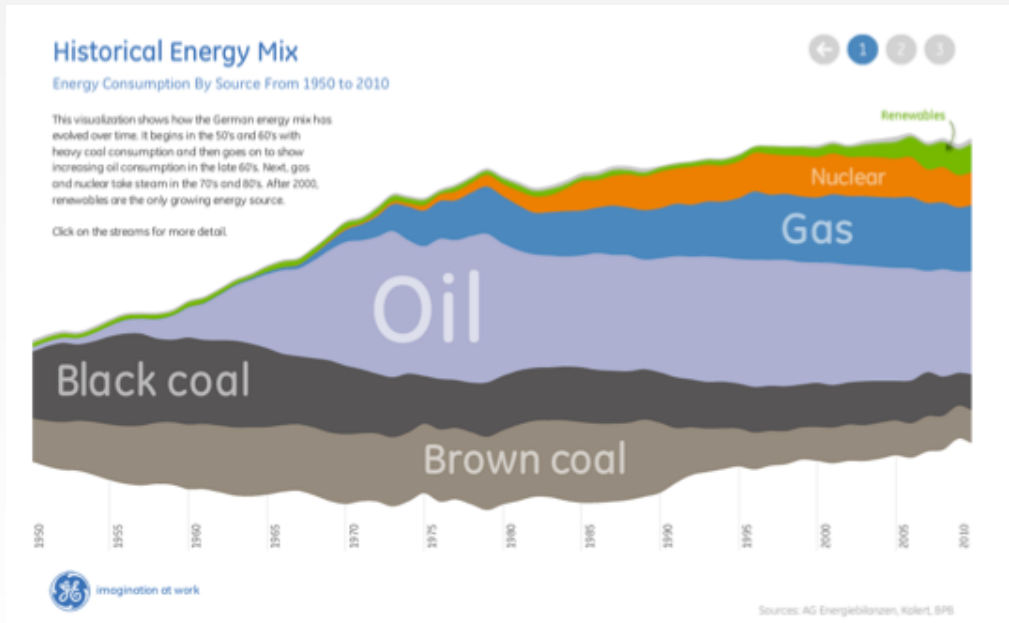
¿Qué gráfico escojo? StreamGraph

Comparación de varias series temporales

Los valores de cada categoría NO son su altura absoluta, sino relativa a la categoría inferior

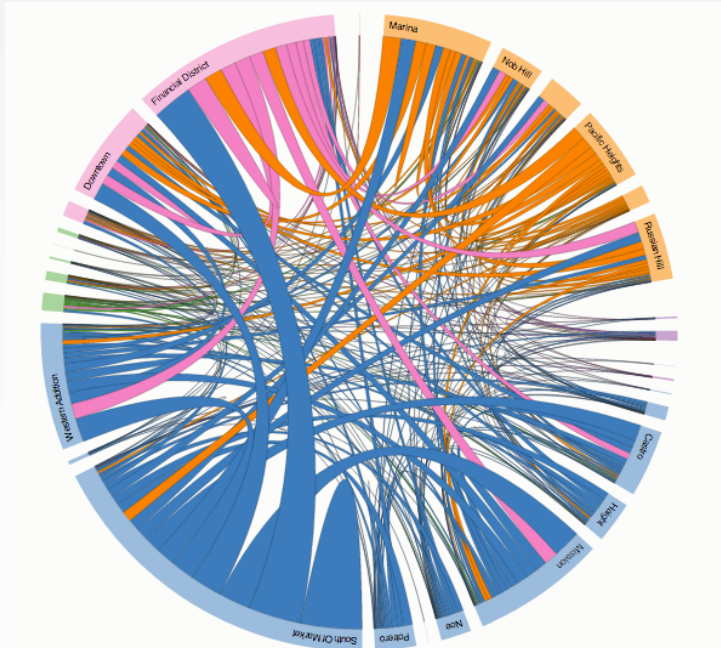
Tienen utilidad para ver el patrón global (suma de las categorías) y el individual

Por el contrario, son difíciles de leer, requieren esfuerzo cognitivo



1. Introducción

¿Qué gráfico escojo? Diagrama de Cuerdas



Relaciones entre varias categorías

Las relaciones se sacan a partir de una matriz de coocurrencias, simétrica o no

En realidad dibuja una red, pero con énfasis en las conexiones, no en la disposición espacial

<http://bost.ocks.org/mike/uberdata/>

1. Introducción

¿Qué gráfico escojo? Network Layout

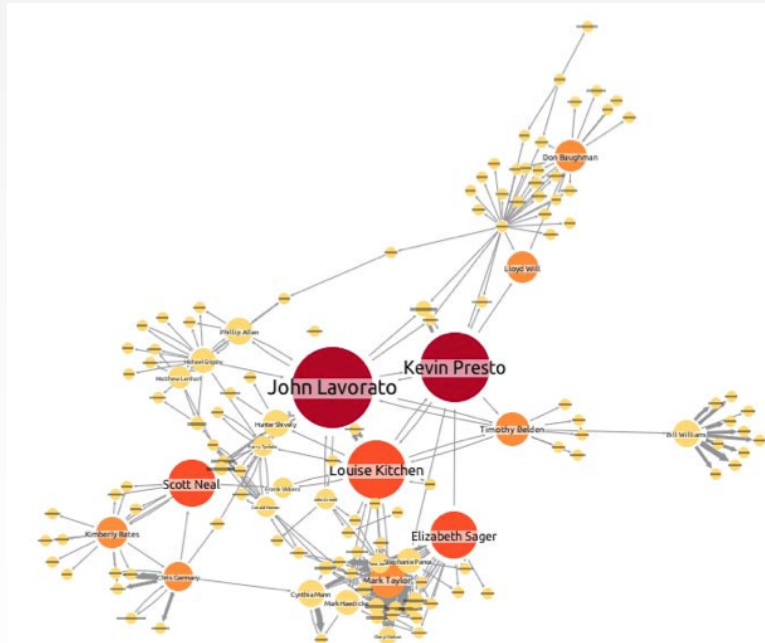


Image created by Joe Parry (<http://key-lines.com/>)

Relaciones entre varias categorías

El énfasis es en la posición espacial (layout)

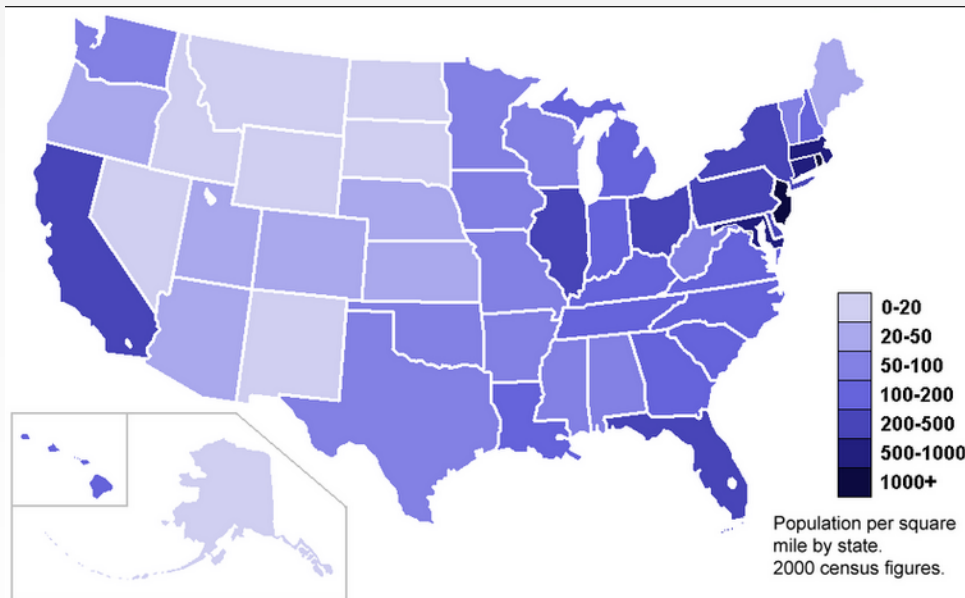
Se suele perder el valor de la conexión

Con muchos nodos se hacen complicadas de leer ('hairballs')

Si no es una red de mundo pequeño, la distribución espacial deja de aportar sentido

1. Introducción

¿Qué gráfico escojo? Mapas de calor



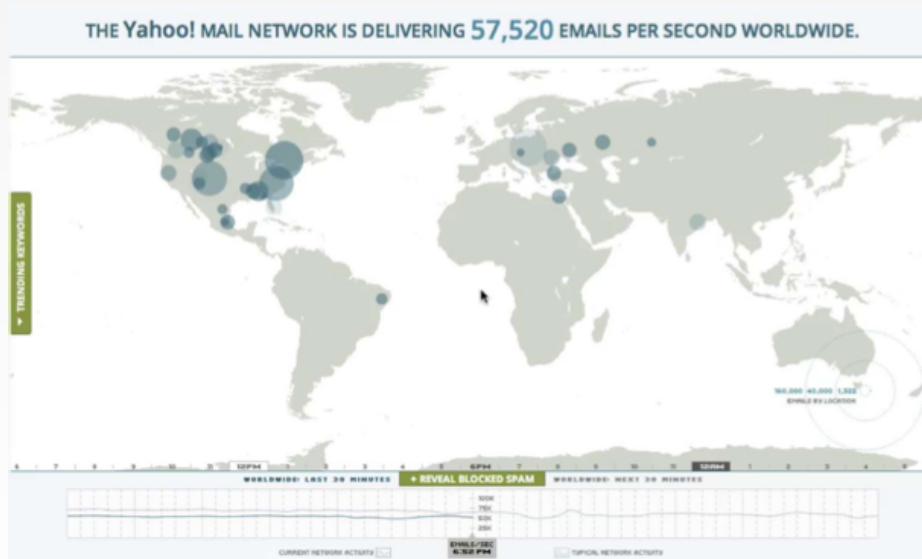
Distribución de una variable cuantitativa sobre un territorio segmentado en categorías (normalmente políticas)

Un error muy común es usar diferentes tonos en la escala de color, en lugar de uno

La deficiencia fundamental son los errores perceptuales asociados a que las poblaciones no están distribuidas de manera uniforme en el territorio

1. Introducción

¿Qué gráfico escojo? Mapa de burbujas



Superpone una variable cuantitativa a su posición en el mapa

Ojo con los solapamientos entre burbujas, hacen difícil de leer el mapa

A veces es mejor codificar la variable en la transparencia en lugar del radio

1. Introducción

¿Qué gráfico escojo? Mapa de cuerdas



Image from "Toronto Flight Lines" (<http://www.biodiaspora.com/>), created by Bio.Diaspora 2012

Superpone un diagrama de red sobre un mapa, cuando se conectan posiciones geográficas

Típicamente el mapa puede desaparecer si hay muchas conexiones y se seguirá viendo si se dibujan los orígenes/destinos de los enlaces con un color diferente



1. Introducción

¿Qué gráfico escojo? Wordclouds

Están ya muy quemadas

Es especialmente importante:

Ponderar bien los tamaños (aplicando TF-IDF)

No abusar de la variedad de color

Eliminar bien el ruido

Extraer bien las entidades de más de un token

Que haya una historia detrás, si no, es mejor descartarlas

1. Introducción

Metáforas de visualización

Comparación



Stacked Area
Graph



Line Graph



Bubble Chart



Radar Chart

1. Introducción

Metáforas de visualización

Proporciones



Pie Chart



Treemap



Circle Packing

1. Introducción

Metáforas de visualización

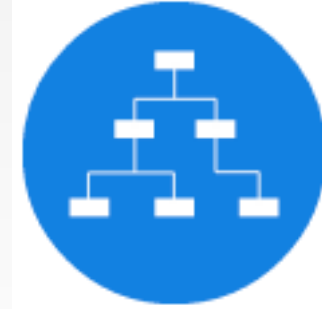
Relaciones



Arc Diagram



Venn Diagram



Tree Diagram

1. Introducción

Metáforas de visualización

Geo



Choropleth Map



Flow Map

1. Introducción

Metáforas de visualización

Flujos



Flow Map



Sankey Diagram

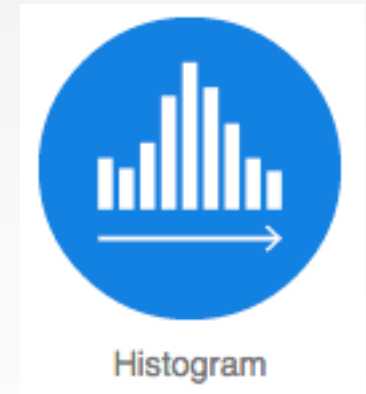
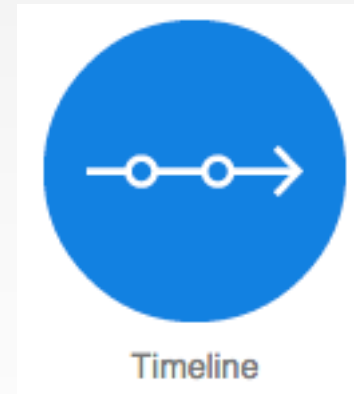
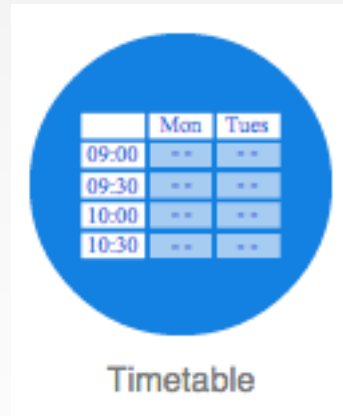
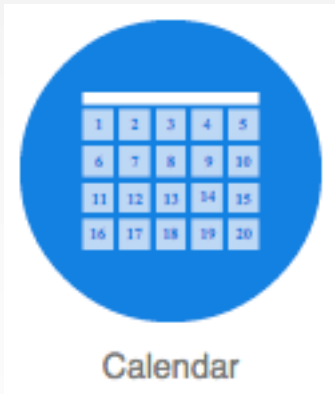


Parallel Sets

1. Introducción

Metáforas de visualización

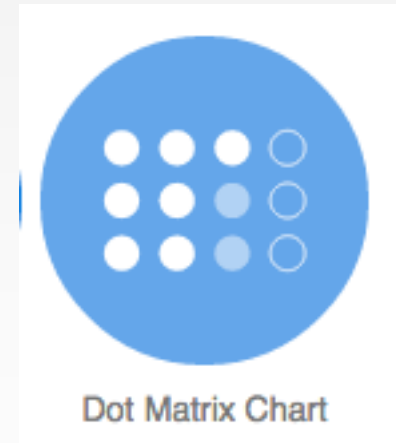
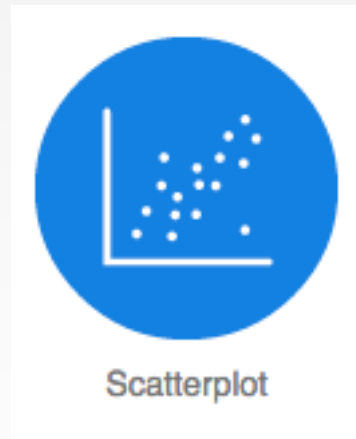
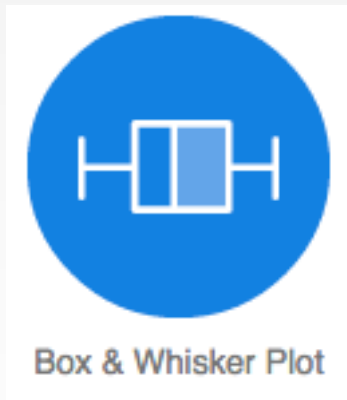
Datos en el tiempo



1. Introducción

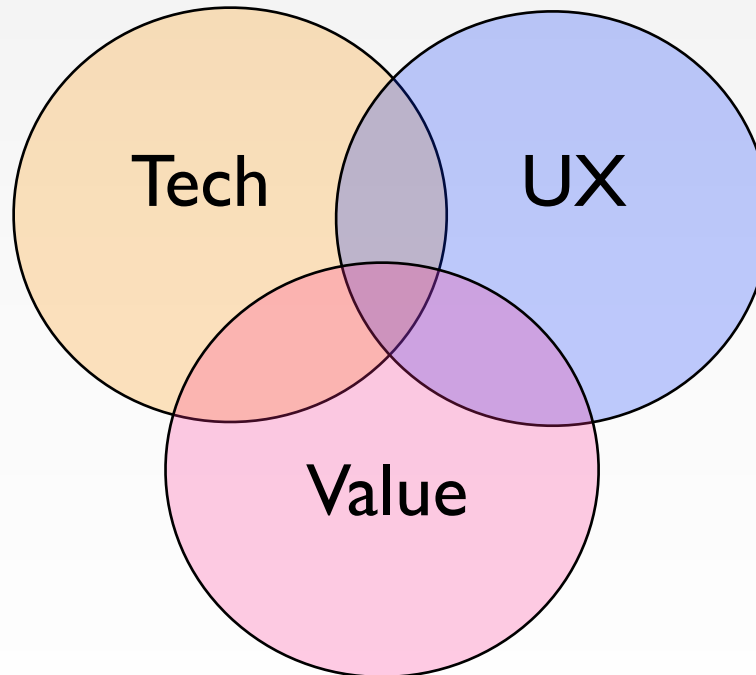
Metáforas de visualización

“Patrones” y correlaciones



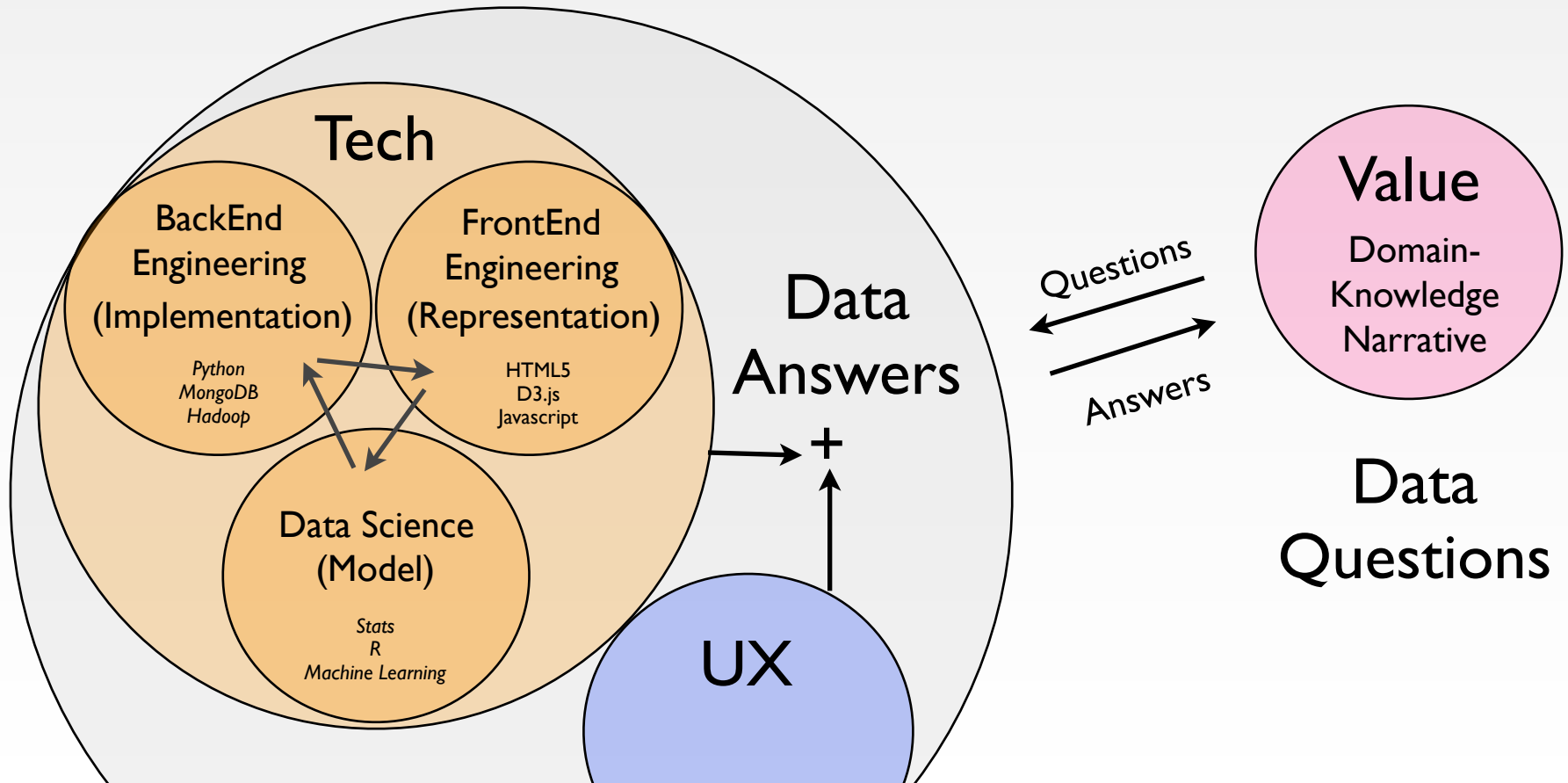
1. Introducción

El equipo de Datos (I)



1. Introducción

El equipo de Datos (II)



1. Introducción

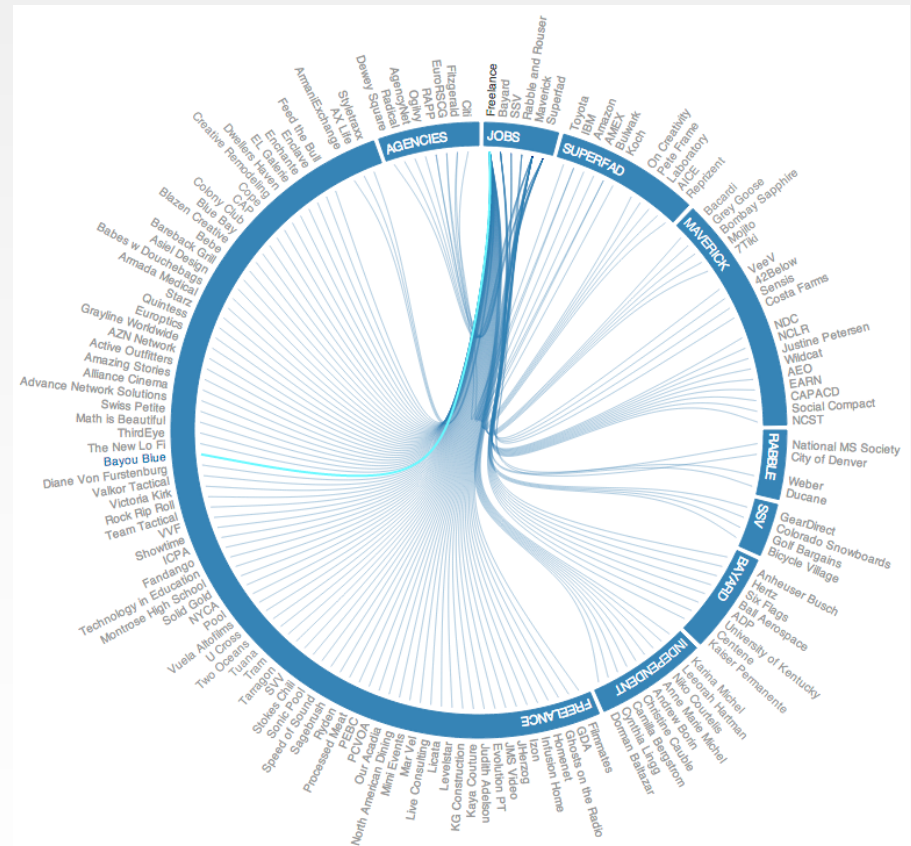
Processing

Para escritorio

Basado en Java

Bueno para artistas
multimedia





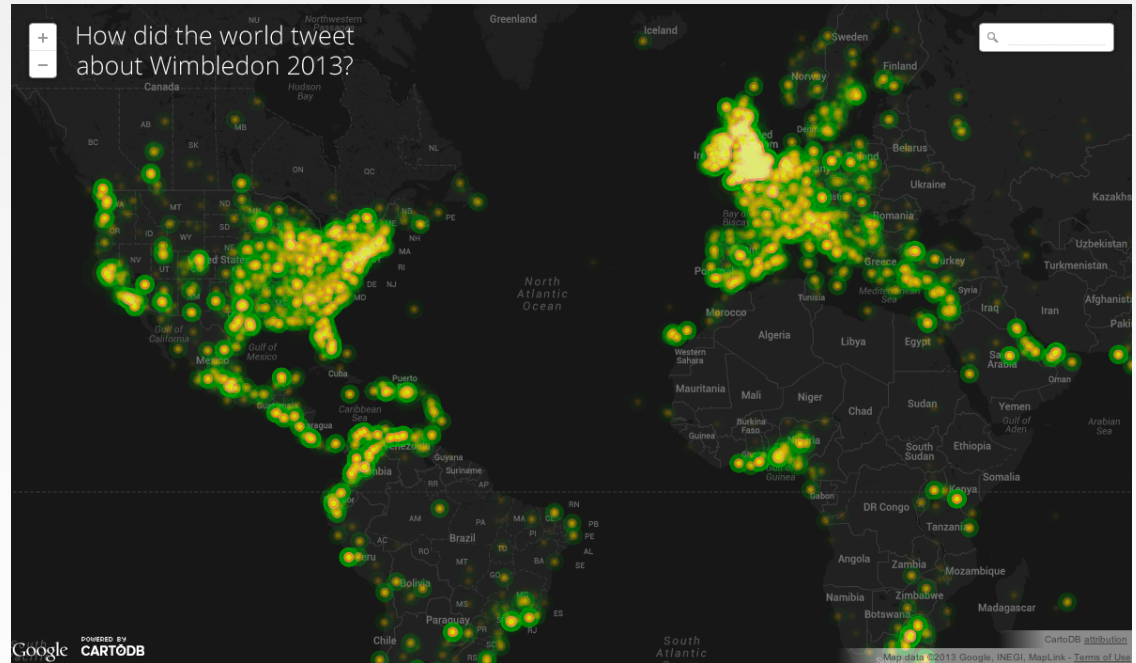
1. Introducción

CartoDB

‘Excel’ para mapas

Versión libre

Mejora continua



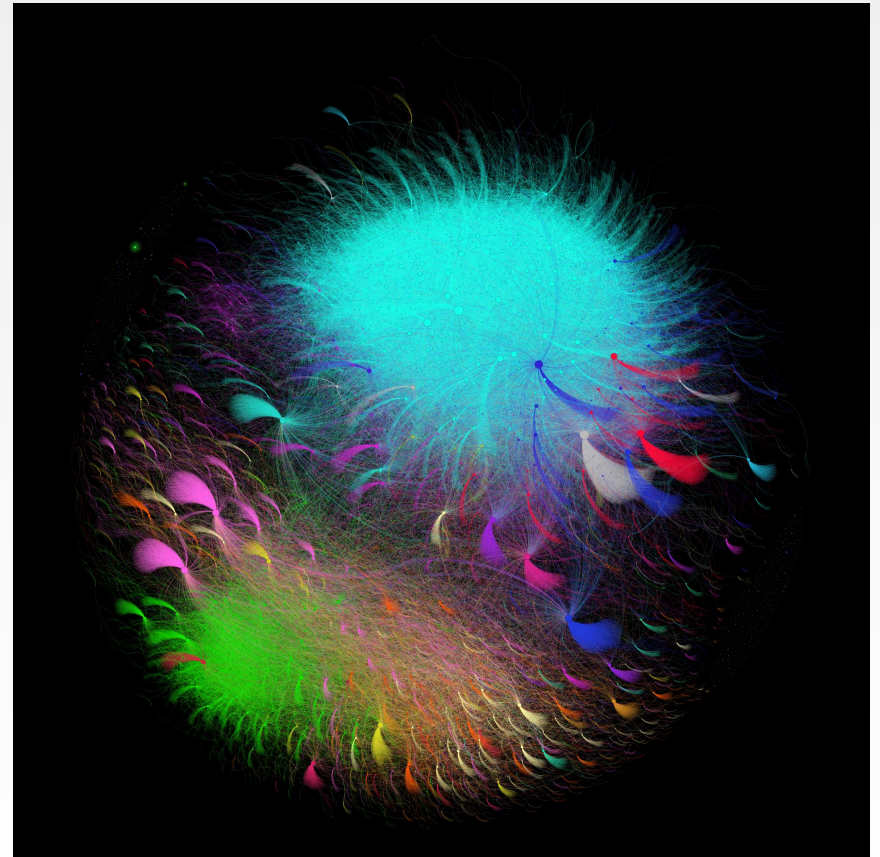
1. Introducción

Gephi

‘Excel’ para redes

Análisis de redes también

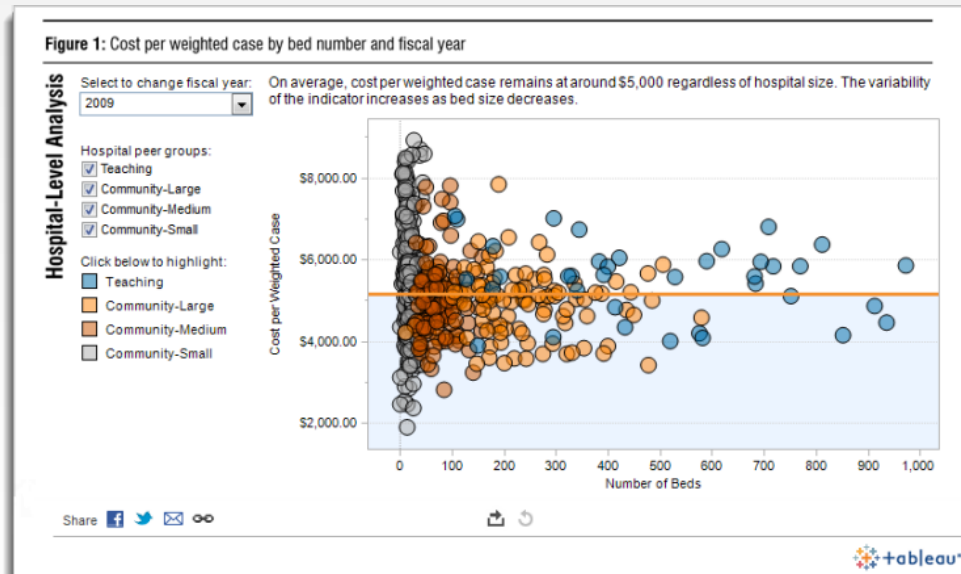
100% gratis



[Link](#)

1. Introducción

Tableau



‘Excel’ ‘para todo’

Visualizaciones interactivas

Dashboards

Versión libre: ‘Tableau Public’

1. Introducción

Datamatic



Visualización online

Visualizaciones interactivas

Libre

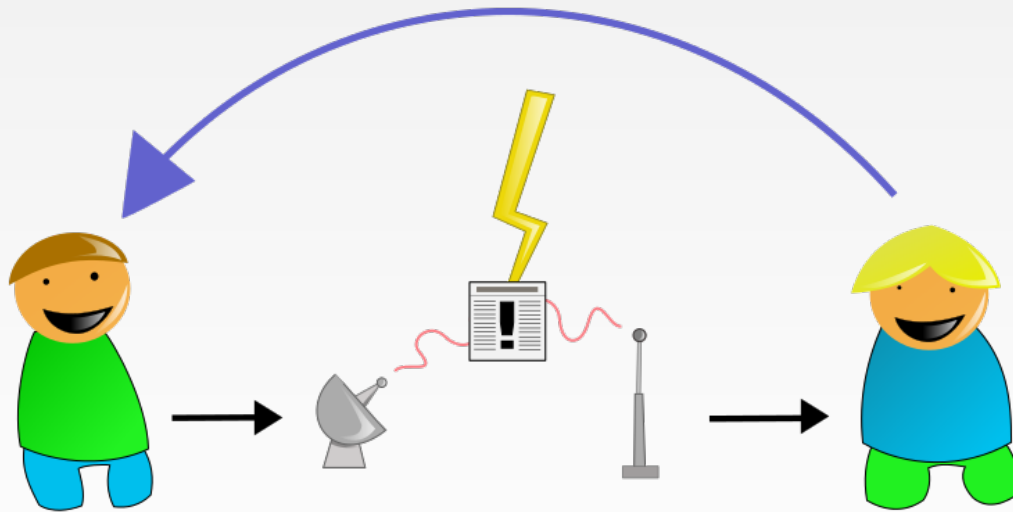
Genera vizs en D3.js

<http://datamatic.io>

2. Comunicando con Datos

2. Comunicando con Datos

Modelo de comunicación
(Shannon and Weaver)



http://en.wikipedia.org/wiki/File:Communication_shannon-weaver2.svg

2. Comunicando con Datos

Los 3 tipos de problemas de la Comunicación
(Shannon and Weaver)

El problema técnico:

¿Con qué nivel de corrección se pueden transmitir el mensaje?

El problema semántico:

¿Con cuanta precisión se transmite el significado adecuado?

El problema de la efectividad:

¿Con qué nivel de efectividad afecta el significado transmitido al comportamiento?

2. Comunicando con Datos

6 principios de comunicar con datos

1. Objetivo
2. Datos adecuados
3. Visualizaciones adecuadas
4. Estética adecuada
5. Canal y medio adecuados
6. Comprueba el resultado

2. Comunicando con Datos

Objetivo

¿Qué?

¿A quién?

¿Para qué?

2. Comunicando con Datos

Datos adecuados

Menos es más

Datos correctos y completos

Ética de la comunicación con datos

2. Comunicando con Datos

Visualizaciones adecuadas

Conocer el tipo de datos a representar

Elegir el tipo de variables/elementos que
representarán los datos y harán emerger insights

2. Comunicando con Datos

Estética adecuada

La Visualización ha de ser estética, pero no es solo estética

Ayuda a mejorar la recepción del mensaje

La estética no puede condicionar el significado

2. Comunicando con Datos

Canal y medios adecuados

¿Cómo?

¿Cuándo?

¿Dónde?

2. Comunicando con Datos

Comprueba el resultado

¡Proceso iterativo!

Alcance

Entendimiento

Impacto

2. Comunicando con Datos

Tips a tener en cuenta

Contexto

No usar 3d si no aporta

Los % suman 100

Cuidado con los diagramas radiales (radar,...)

Evitar etiquetados insuficientes o malos

Elegir escalas de color adecuadas

3. Tableau Public

3. Tableau Public

¿Porqué Tableau?

Comienza a imponerse como estándar en la visualización de datos para no programadores

El uso en medios debería forzar mejoras y desarrollo más frecuentes

Es un buen compromiso entre simplicidad de uso y capacidades

3. Tableau Public

Qué es Tableau

Software para crear visualizaciones de datos complejas y publicables sin necesidad de conocimientos de programación.

Manera rápida y simple de visualizar datos

Permite publicar los dashboards e integrarlos en páginas propias

3. Tableau Public

Qué no es Tableau

No permite el formateo ni adaptación en profundidad de los datos.

No permite realizar visualizaciones diferentes de las ofrecidas.

No llega al “estado del arte”

Si algo no funciona no puedes hacer nada

3. Tableau Public

“Tipos” de Tableau

Desktop

Server

Online

Public

Reader

3. Tableau Public

Connect

In a file

Tableau Data Extract

Microsoft Excel

Text File

On a server

Amazon Redshift

Cloudera Hadoop

Firebird

Google Analytics

Google BigQuery

Hortonworks Hadoop Hive

HP Vertica

Microsoft SQL Server

MySQL

OData

Oracle

Pivotal Greenplum Database

PostgreSQL

Salesforce

Teradata

Windows Azure Marketplace

Fuentes de Datos

Tableau Public solo permite importar datos desde excel, archivos tipo csv (comma separated values), fuentes de odata o Windows Azure Marketplace

Con las licencias de pago permite muchas mas fuentes de datos

3. Tableau Public

Interfaz de Tableau

datasetUN 0 Add...

Connected to Text File

Directory
/Users/nihilist/codigo/cursos/utad...

Files
Enter file name

- datasetUN.csv
- i+dWorldBank.txt
- testKK.csv
- WDI_Country.csv
- WDI_CS_Notes.csv
- WDI_Data.csv
- WDI_Description.csv
- WDI_Footnotes.csv
- WDI_MiniData.csv
- WDI_Series.csv
- WDI_ST_Notes.csv

datasetUN.csv

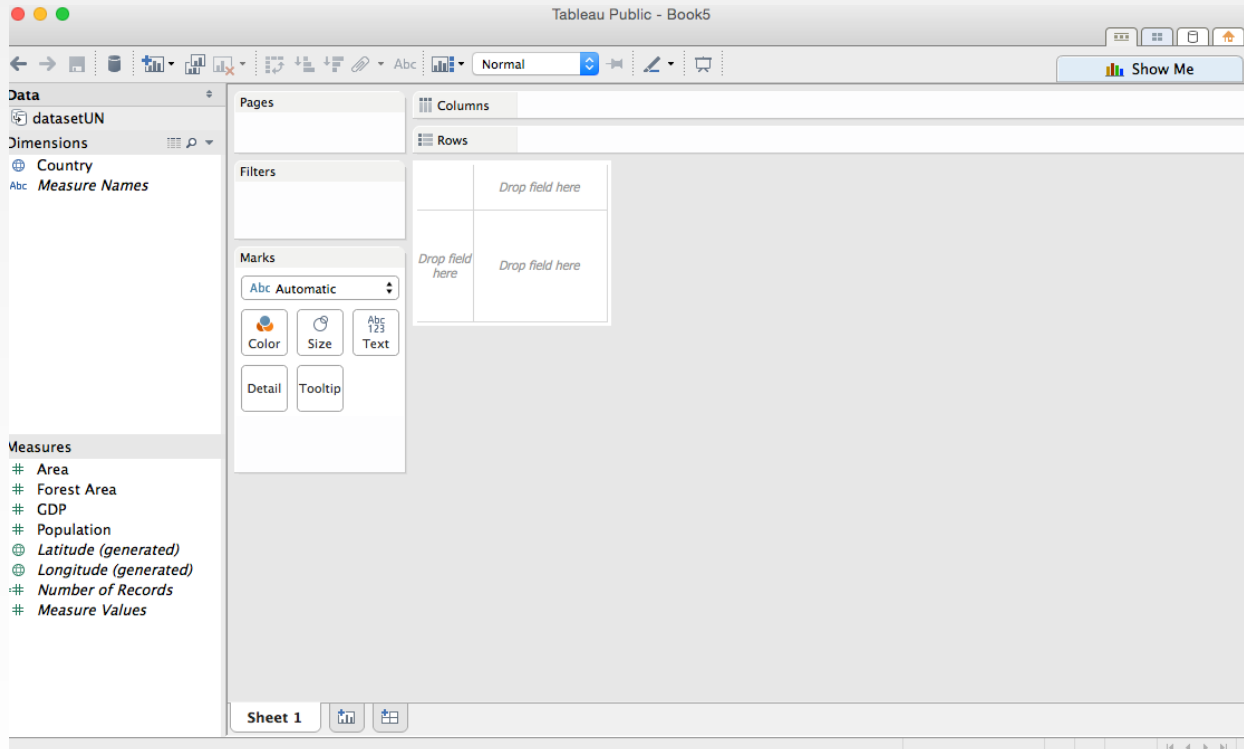
[Go to Worksheet](#)

☐ Show hidden fields Rows 111

Country	Population	Area	GDP	Forest Area
	#	#	#	#
Canada	34,754,300.00	9,984,6...	52,283.33	3,101,340.00
Sao Tome a...	188,098.00	960.00	1,386.33	270.00
Turkmenistan	5,172,930.00	488,100...	6,469.39	41,270.00
Montenegro	621,081.00	13,810.00	6,514.20	5,430.00
Lithuania	2,987,770.00	65,300.00	13,984.22	21,756.00

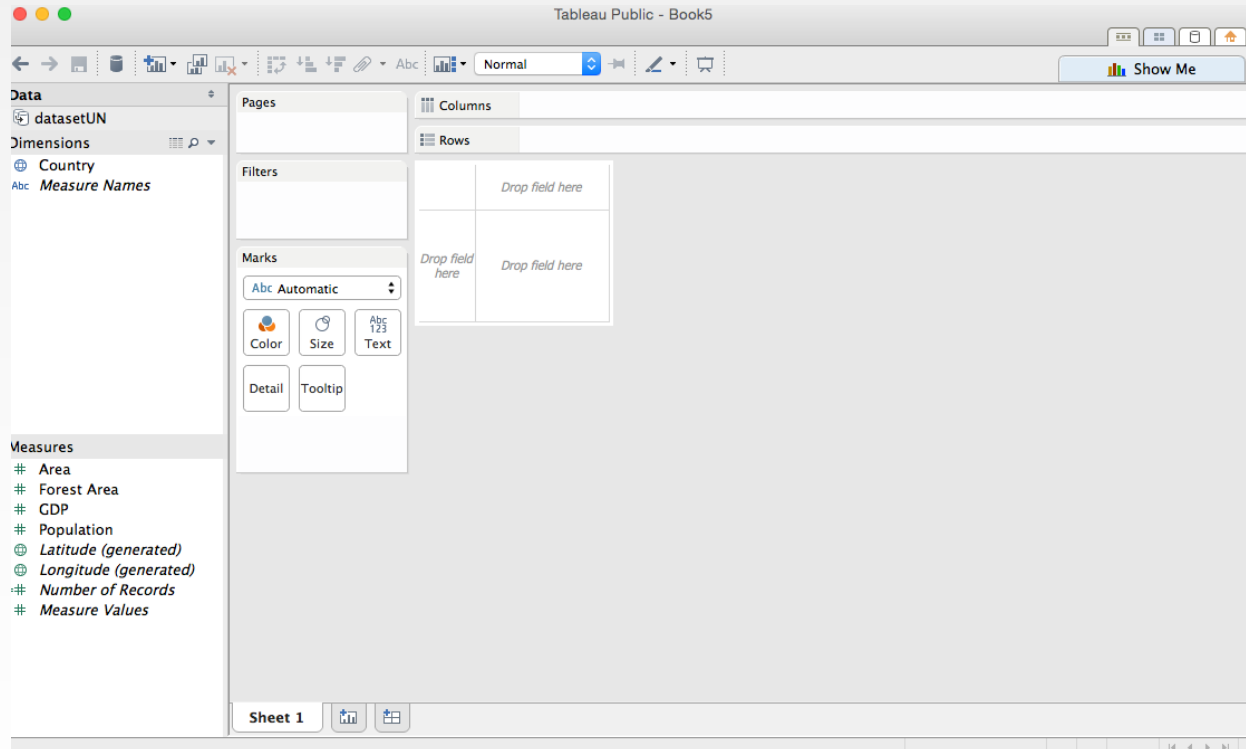
3. Tableau Public

Interfaz de Tableau



3. Tableau Public

Interfaz de Tableau



3. Tableau Public

Tipos de Datos

Dimensiones ó Medidas

Cadenas, números, fechas o campos geo

Discretos o continuos