

# **Diplomado Big Data y Ciencias de datos Visualización de datos**

Cris Hernández

[cris@crishernandez.co](mailto:cris@crishernandez.co)

# Sobre Mi

## Trayectoria Relacionada

- Geógrafo Universidad de Chile, Experto en Ciencia de Datos Espaciales y Visualización.
- CEO Mapoteca: especializados en análisis de movilidad urbana, Chile.
- Data Scientist: Common Action Forum, España.
- Data Officer: Data Campfire, Londres.
- Becario, Stiftung Mercator: Periodismo de Datos, Alemania.
- Programación:
  - Python, JavaScript, D3js, HTML/CSS.
  - Sistemas de información geográficos.

# **Sobre ustedes:**

## **(Por qué, qué)**

# Contenido

- Introducción.
  - ¿Qué?
  - ¿Por qué?
- Tipos de visualizaciones.
- Un poco de historia.
- Composición de visualización: Marcas y canales.
- Herramientas del curso:
  - Kepler.gl
  - Tableau

# Qué esperar del curso

**... y que espero cumplir.**

- Aplicar un proceso de diseño para crear visualizaciones efectivas.
- Utilizar principios de percepción y cognición humana en visualización.
- Evaluar una visualización de forma crítica, pudiendo además sugerir mejoras.
- Implementar un dashboard con la herramienta Tableau Public.
- Comprender la interdisciplina que conlleva la visualización con otras áreas.

# Evaluaciones

## Prácticas y Teóricas

### **2 talleres evaluados (cada uno finalizando el mes)**

- Se comienza su desarrollo en clases.
- Plazo hasta justo antes de empezar la clase siguiente (18:30).
- Planificadas para ser individual.

### **1 Control**

- Preguntas de alternativas y/o desarrollo.
- Se realiza al final del curso.

### **Nota final**

- 70% Promedio de talleres + 30% Control

# Evaluaciones

- Todo se entrega en el sitio del curso.
- Todas las entregas son en PDF o en formato Tableau. En caso de no entregar en el formato solicitado, tendrán 1 punto de descuento en cada evaluación.
- En los talleres se aceptará entregas con un atraso máximo de 24 horas y se aplicará un descuento de 1 punto.

Dudas y comentarios a mi correo: [cris@crishernandez.co](mailto:cris@crishernandez.co) (Sí, termina en .CO)

# Bibliografía

Una buena parte de este curso está basado en el trabajo de Tamara Munzner, a partir de su libro **Visualization: Analysis & Design**.

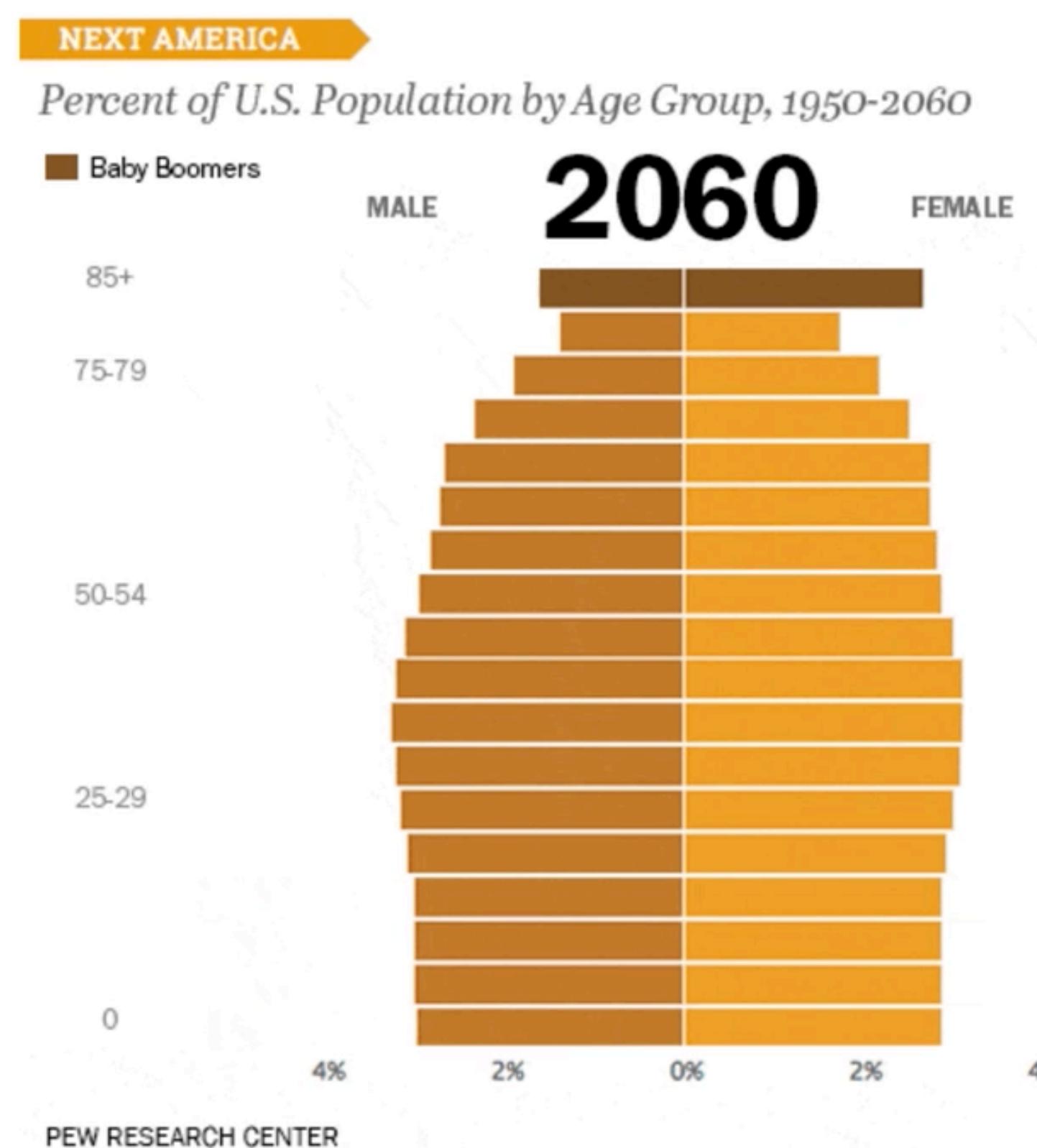
Sin embargo, también toma material de...

- Interactive Data Visualization de Matthew O. Ward, Georges Grinstein, Daniel Keim.
- The Visual Display of Quantitative Information de Edward Tufte.
- Information Visualization: Perception for Design de Colin Ware.
- Visualize This de Nathan Yau.
- Las leyes de la simplicidad de John Maeda.

# Por qué este curso

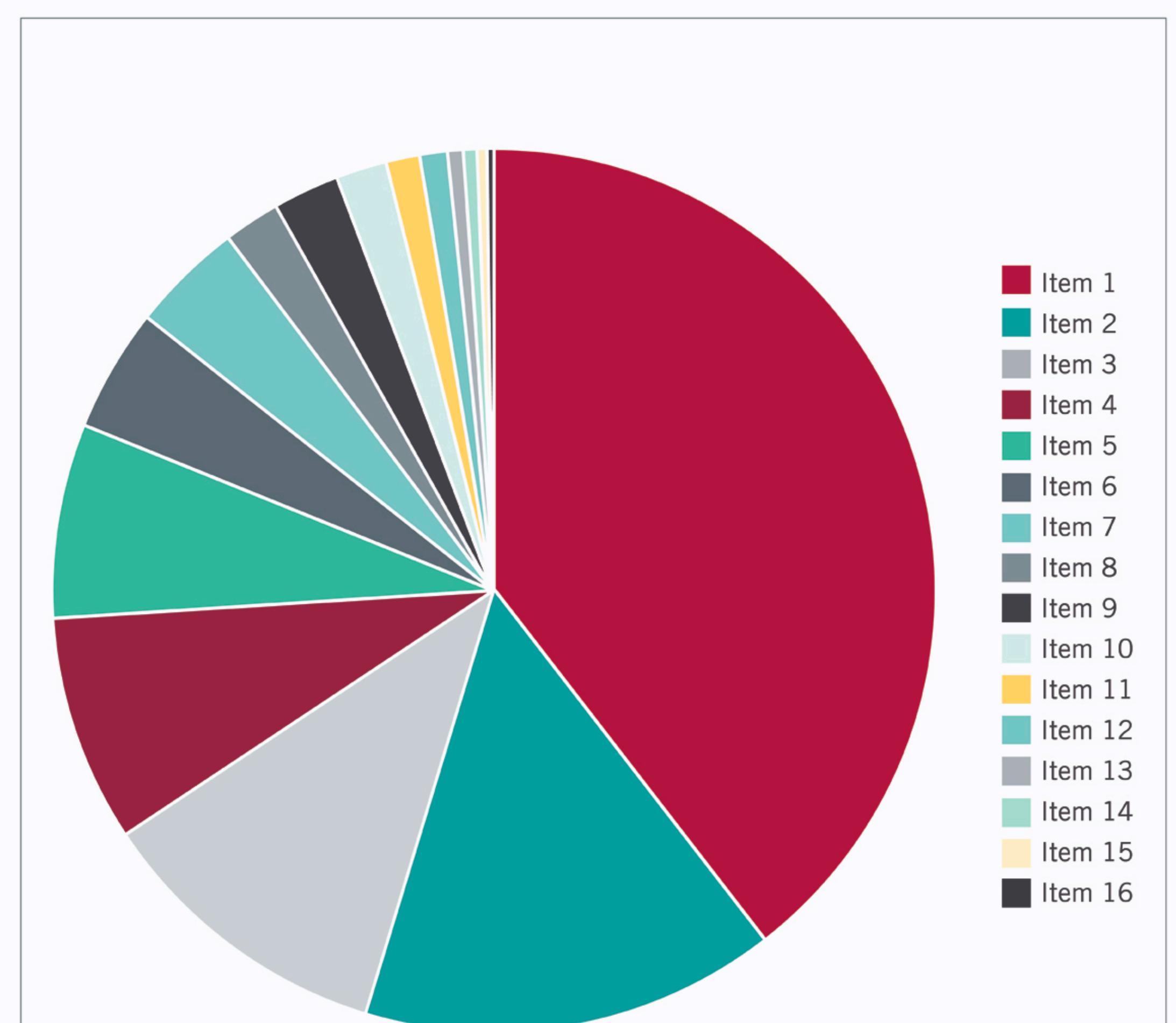
... y por qué no.

## 7. The Next US



Visualization by: Pew Research Center

Learn more: [Next America](#)



- Item 1
- Item 2
- Item 3
- Item 4
- Item 5
- Item 6
- Item 7
- Item 8
- Item 9
- Item 10
- Item 11
- Item 12
- Item 13
- Item 14
- Item 15
- Item 16

# Por qué este curso

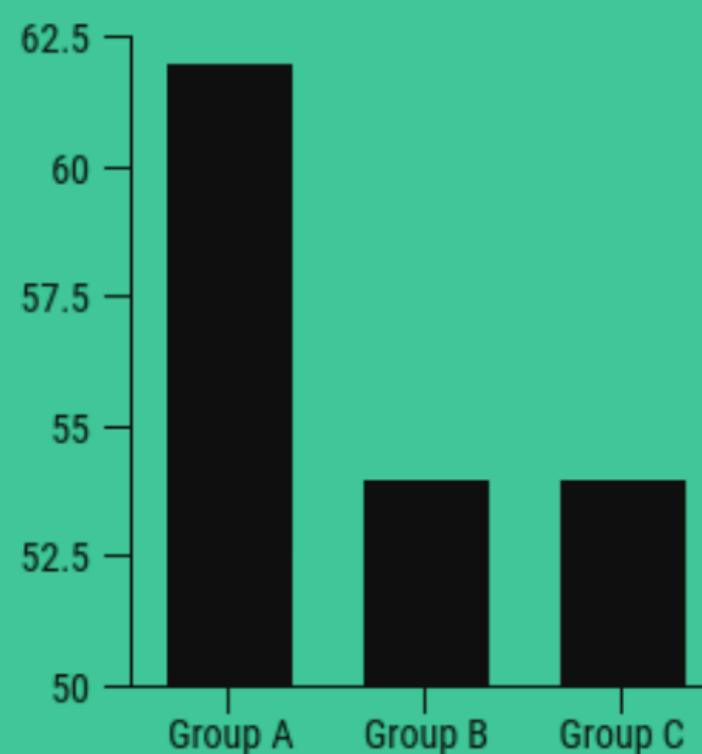
... y por qué no.

<https://venngage.com/blog/misleading-graphs/>

1

## OMITTING THE BASELINE

In most cases, the baseline for a graph is 0. But writers can skew how data is perceived by making the baseline a different number. This is known as a “truncated graph”.



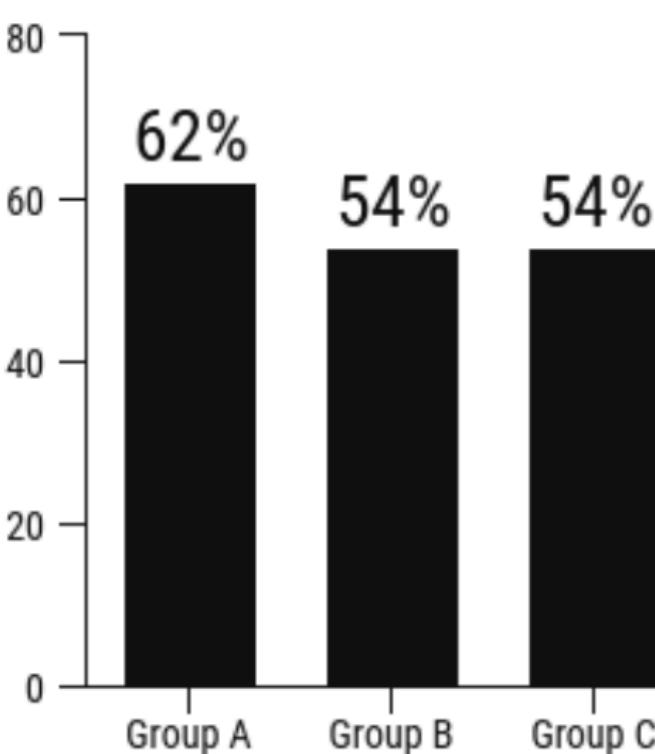
MISLEADING

- Starting the vertical axis at 50 makes a small difference between groups seem massive
- Group A looks much larger than Groups B and C

VS

ACCURATE

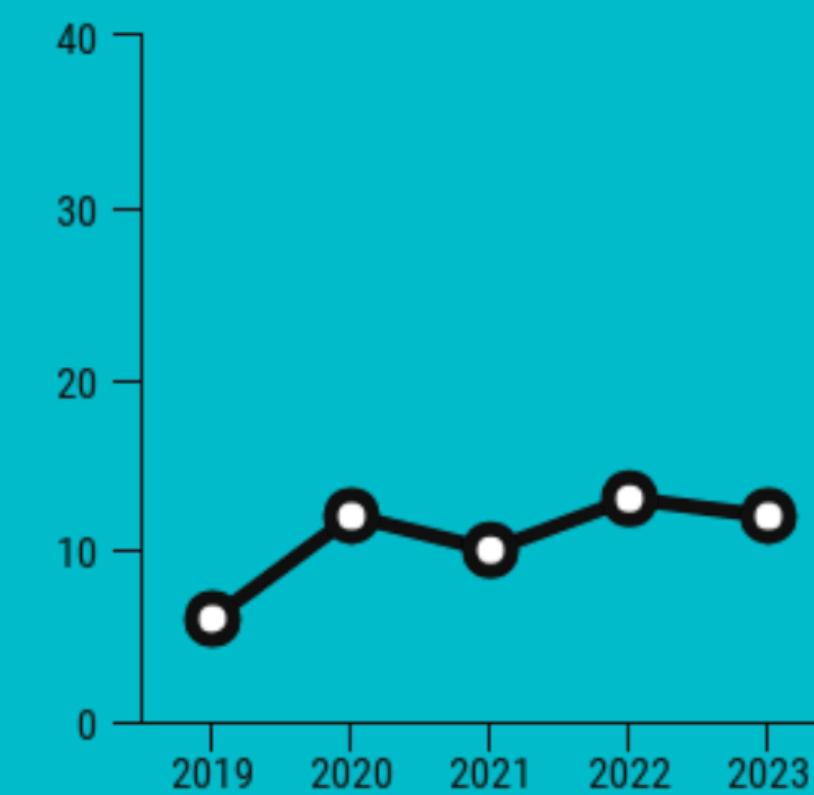
- Starting the vertical axis at 0 offers a more accurate depiction of the data
- The difference between the groups does not seem as dramatic



2

## MANIPULATING THE Y-AXIS

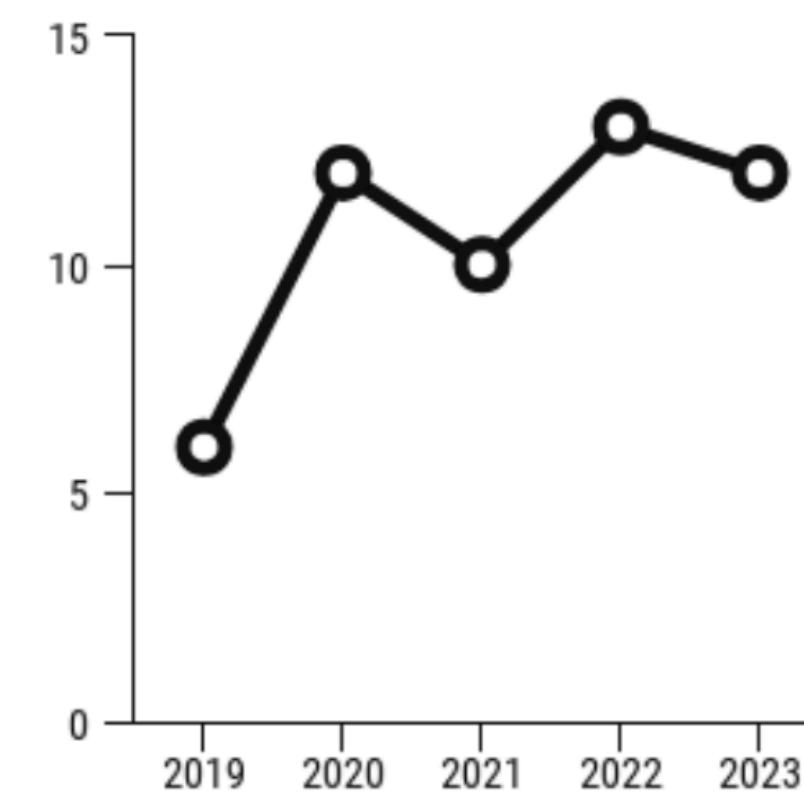
Expanding or compressing the scale on a graph can make changes in data seem more or less significant than they actually are.



MISLEADING

- The scale is disproportionate to the data, making the change over time seem small

VS



ACCURATE

- The scale is proportionate to the data, showing a greater change over time

# Qué es Visualización de Datos (VdD)

- 1."Transformación de lo simbólico a lo geométrico".  
[McCormick et al. 1987]
- 2.“[...] encontrar la memoria artificial que mejor apoya a nuestros medios naturales de percepción.” [Bertin 1967]
- 3.“El uso de representaciones visuales de datos, generados por computador, interactivos, para amplificar nuestra cognición.” [Card, Mackinlay, & Shneiderman 1999]

# Qué es la VdD

“Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions” (Tableau)

# Qué es la VdD

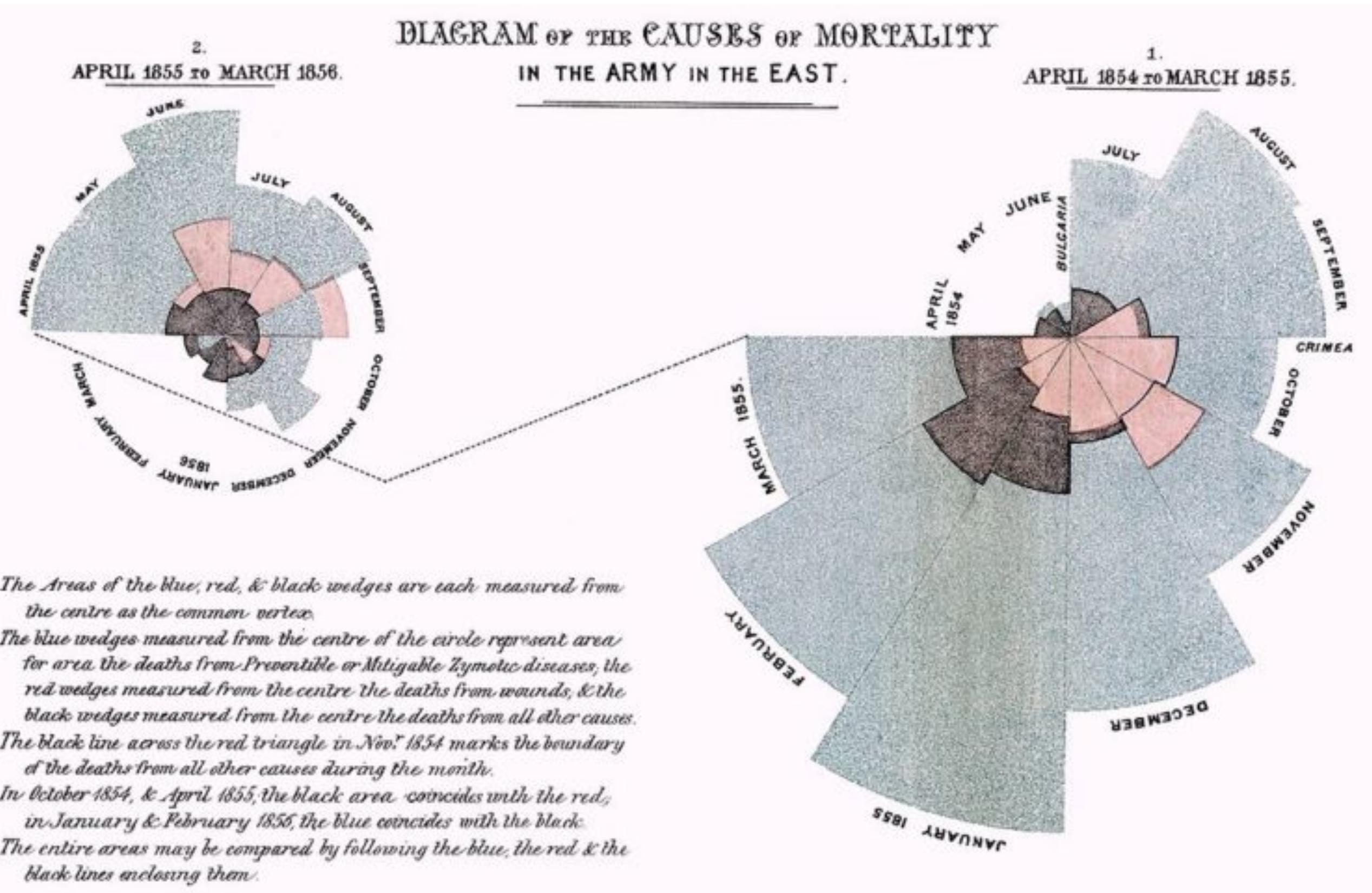


En 1854, durante una epidemia de Córlera en Londres, el Dr. John Snow usa un análisis espacial para apoyar su hipótesis.

Encontró que las muertes eran principalmente por una bomba de agua contaminada.



# Qué es la VdD

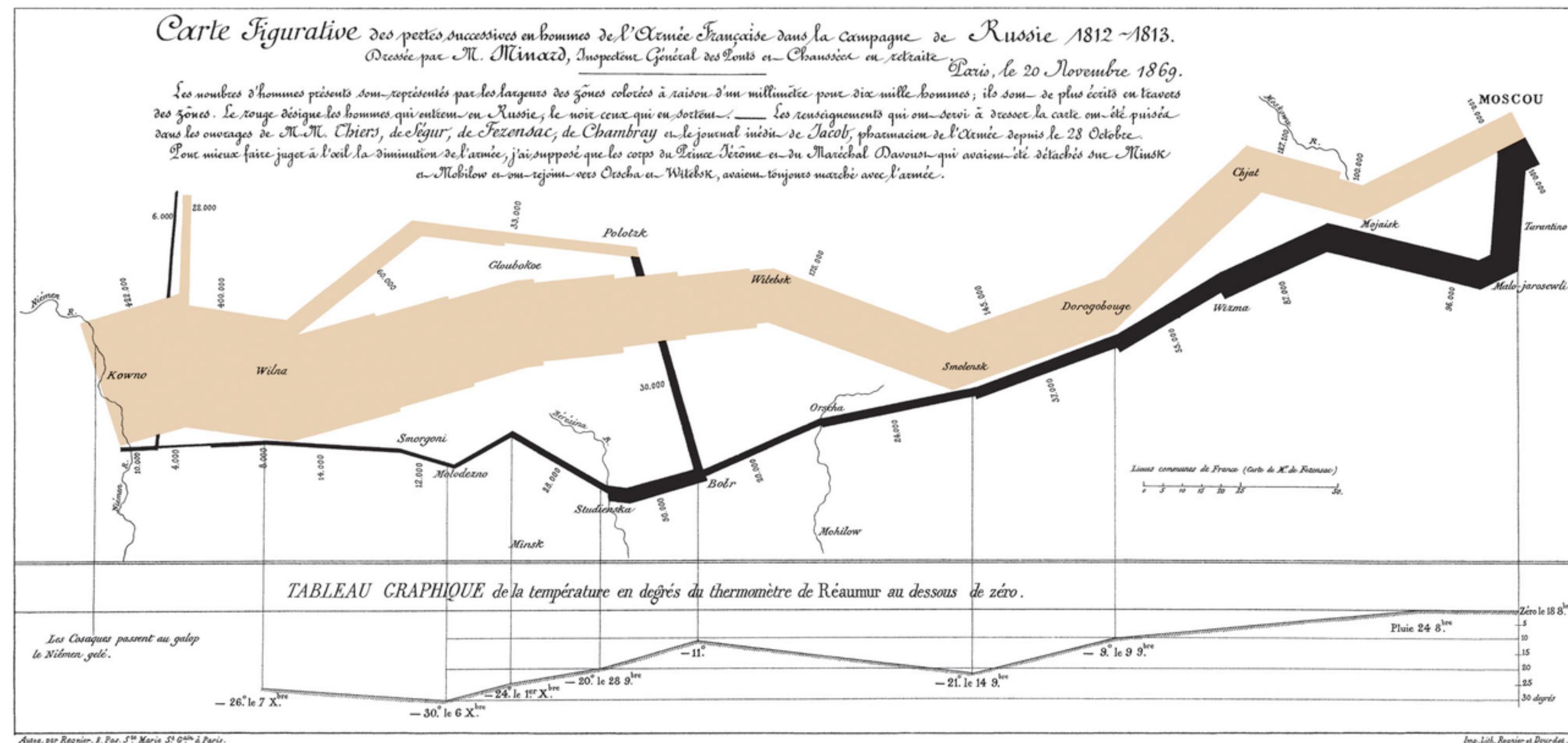


En 1858, Florence Nightingale, dibujó este gráfico para fin de ilustrar las causas de la mortalidad de los soldados en el hospital militar que dirigía durante la guerra de Crimea.

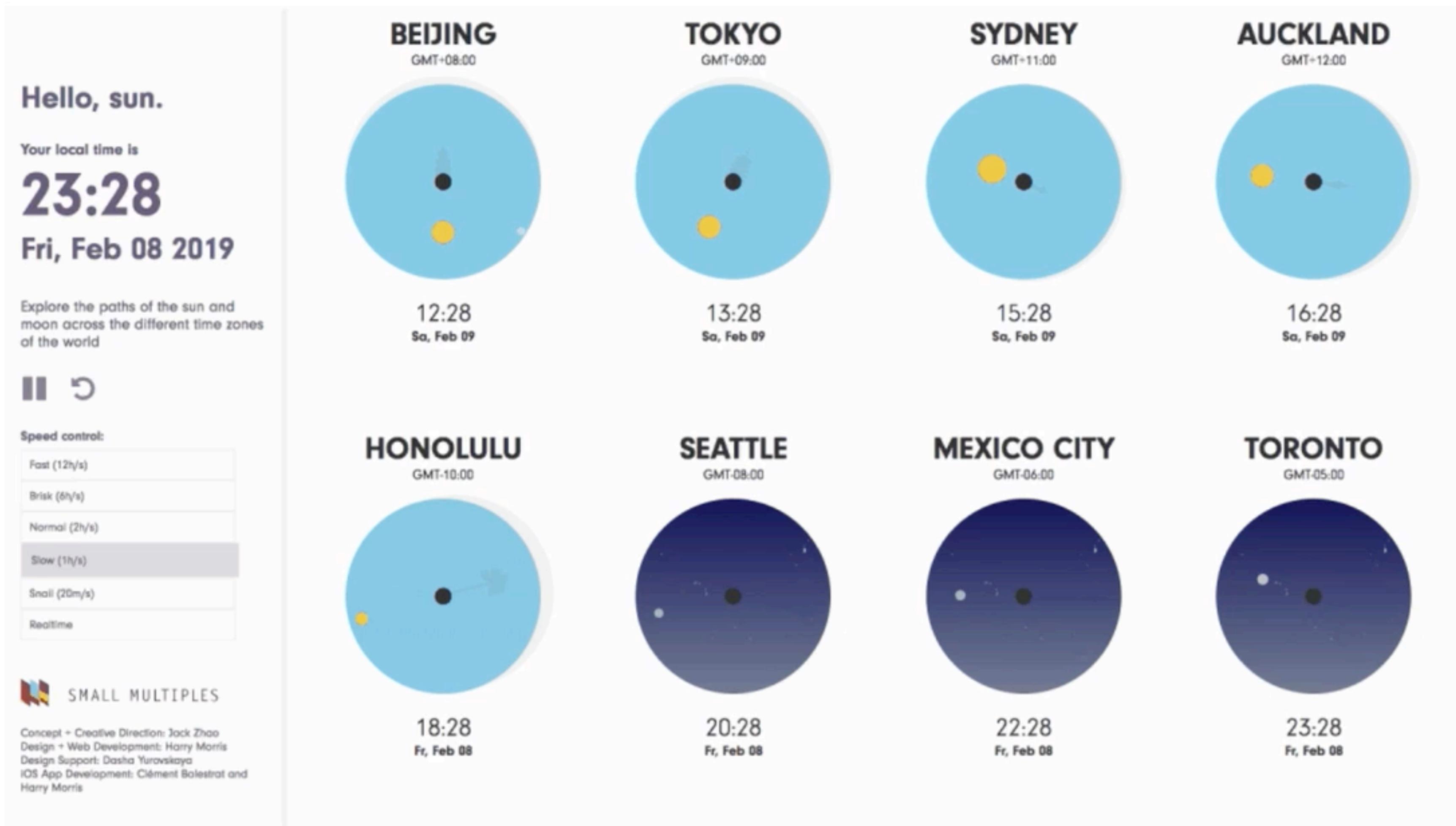
Se dio cuenta de que el 80% de los muertos eran víctimas de los deficientes tratamientos sanitarios.

# Qué es la VdD

La visualización espacio-temporal más antigua y famosa es el mapa de Minard (1869) que representa la invasión de Napoleón a Rusia. Se puede ver en el mapa de flujos información de la ida, del regreso, con eventos como batallas y otros hitos de la campaña.



# Hello Sun app



By Small Multiples

# Por qué es la relevante la VdD

- Representación de un concepto abstracto, como datos.
- Uso de imágenes visuales generadas por computador.
- Ayudar a personas: a entender, a convencer o a realizar acciones eficientemente

# Por qué es la relevante la VdD

- In 2020, people created 1.7 MB of data every second.
- By 2025, 200+ zettabytes of data will be in cloud storage around the globe.
- By the end of 2020, 44 zettabytes will make up the entire digital universe.
- Every day, 306.4 billion emails are sent, and 500 million Tweets are made.

<https://techjury.net/blog/how-much-data-is-created-every-day/#gref>

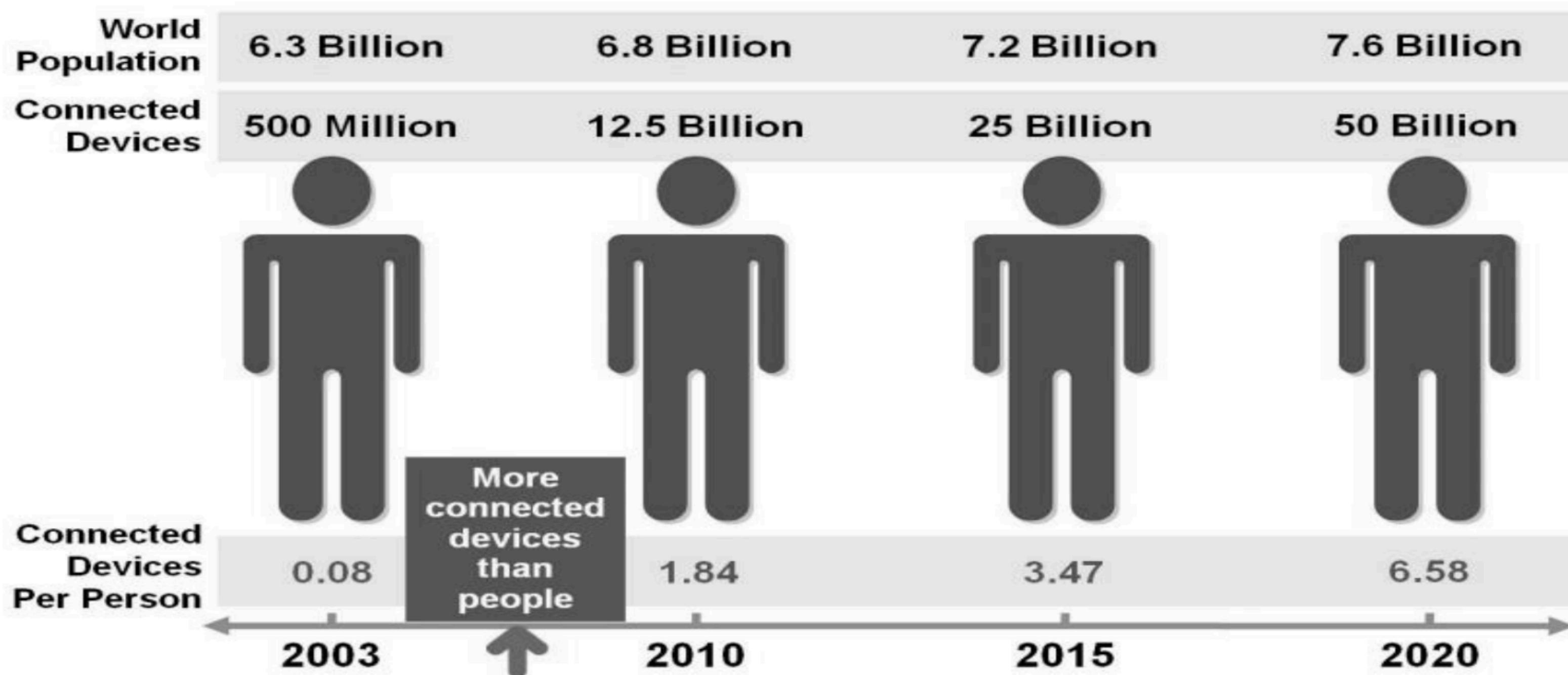
MÉTRICA	VALOR	BYTES
BYTE (B)	1	1
KILOBYTE (KB)	1.024 <sup>1</sup>	1.024
MEGABYTE (MB)	1.024 <sup>2</sup>	1.048.576
GIGABYTE (GB)	1.024 <sup>3</sup>	1.073.741.824
TERABYTE (TB)	1.024 <sup>4</sup>	1.099.511.627.776
PETABYTE (PB)	1.024 <sup>5</sup>	1.125.899.906.842.624
EXABYTE (EB)	1.024 <sup>6</sup>	1.152.921.504.606.846.976
ZETTABYTE (ZB)	1.024 <sup>7</sup>	1.180.591.620.717.411.303.424
YOTTABYTE (YB)	1.024 <sup>8</sup>	1.208.925.819.614.629.174.706.176

**Cuando me preguntan cuantas décimas me faltan para aprobar el ramo.**



# De dónde provienen los datos?

Billones de sensores aportan con datos pasivos. Mientras que millones de dispositivos móviles, computadores y otros, aportan con datos activos.

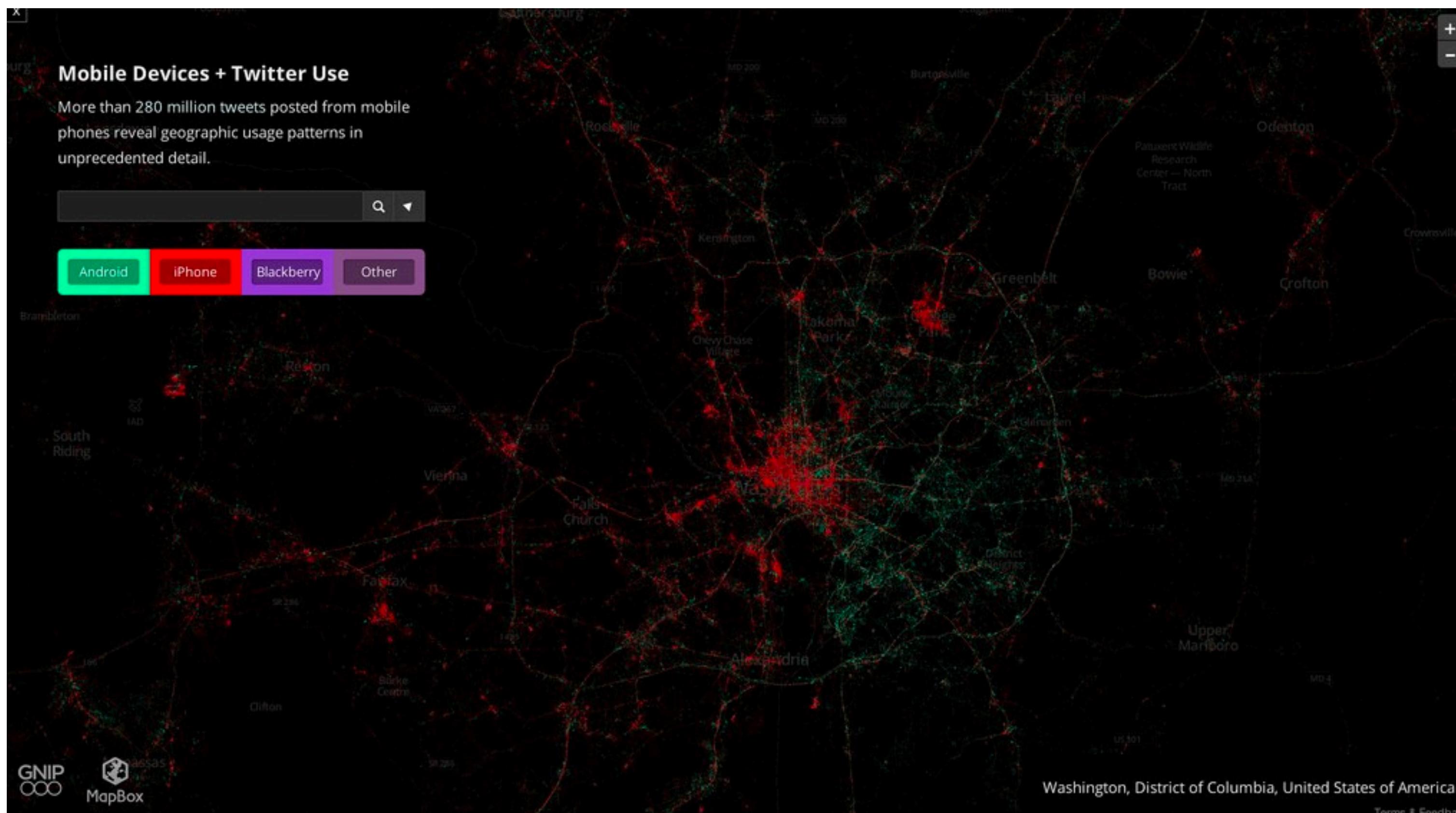


Fuente: Cisco, 2011.

# De dónde provienen los datos?

Billones de sensores aportan con datos pasivos. Mientras que millones de dispositivos móviles, computadores y otros, aportan con datos activos.

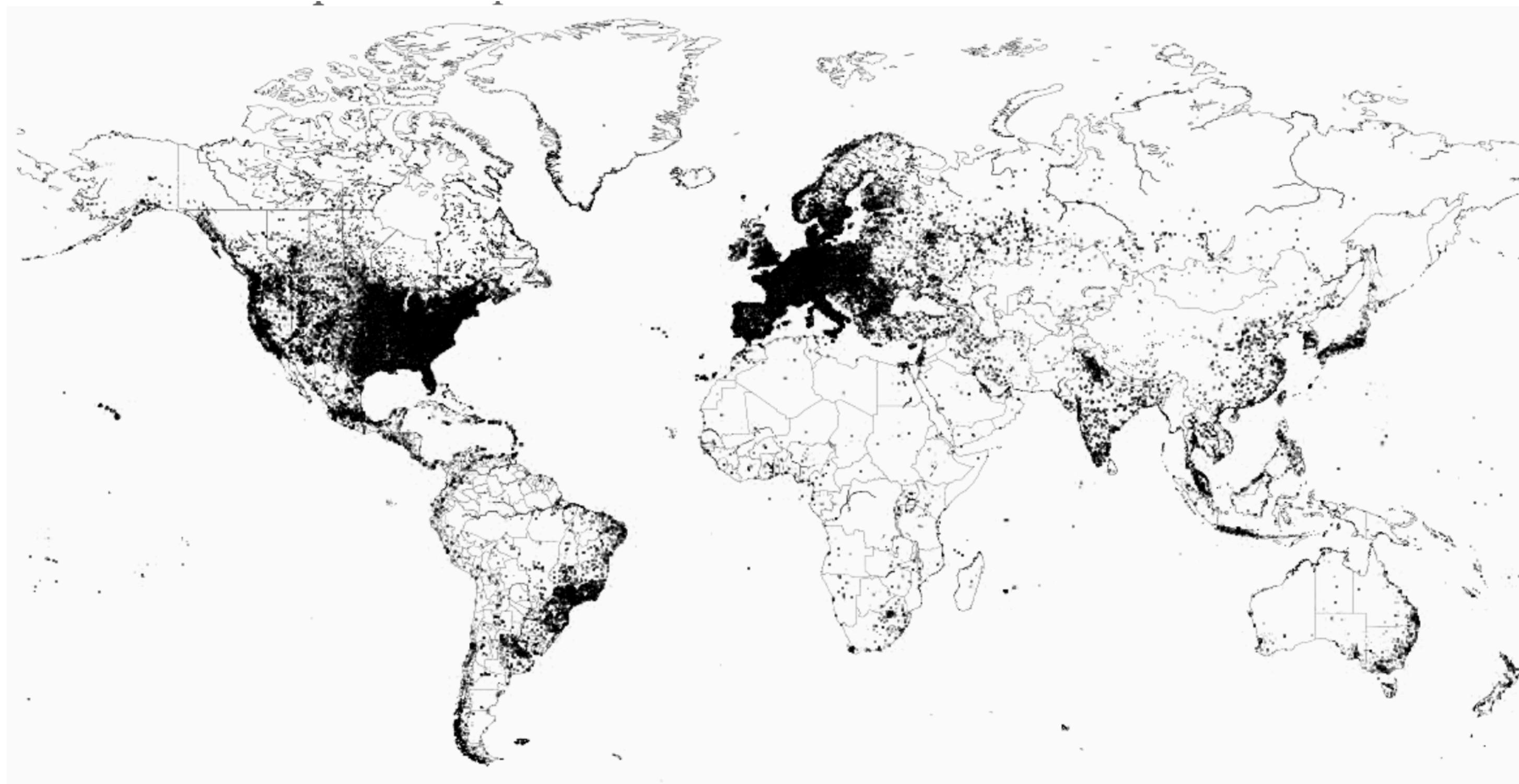
<https://www.wired.com/2013/07/design-06272013-tweetmaps/>



# De dónde provienen los datos?

Billones de sensores aportan con datos pasivos. Mientras que millones de dispositivos móviles, computadores y otros, aportan con datos activos.

<https://www.wired.com/2013/07/design-06272013-tweetmaps/>



Fuente: *Shodan*, 2014

# Tipos de VdD

Interactive

- D3.js
- Plotly
- R
- Processing
- Unity

Static

- Illustrator
- Adobe Spark
- Infogram
- Visme
- Piktochart

# Tipos de VdD



# Tipos de VdD

The Data Visualisation Catalogue

About • Blog • Shop • Resources

中文 Espanol Русский Türkçe

Logre el éxito con nuevas habilidades



Codifique, experimente y cree su siguiente aplicación con 12 meses de servicios gratis

Search by Function

View by List



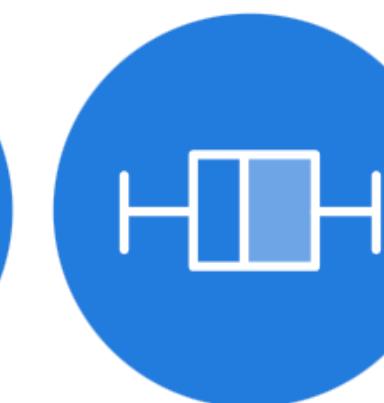
Arc Diagram



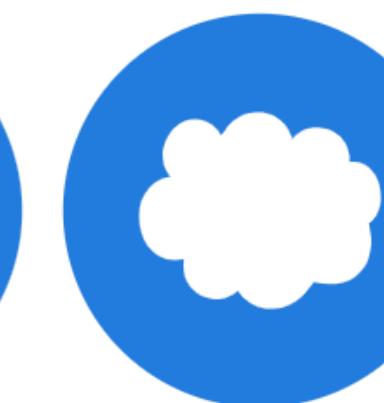
Area Graph



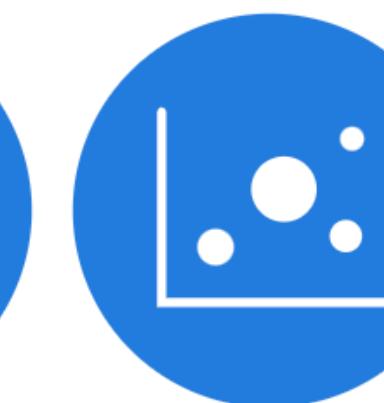
Bar Chart



Box & Whisker Plot



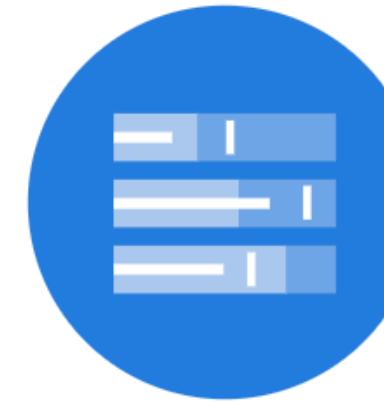
Brainstorm



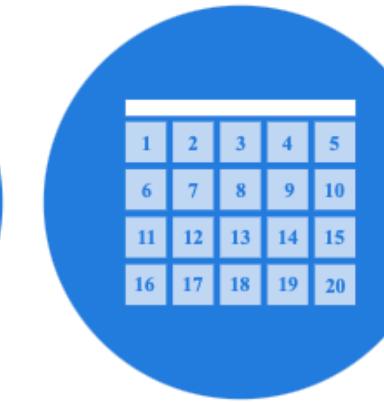
Bubble Chart



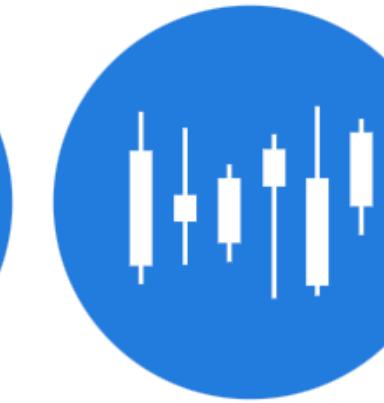
Bubble Map



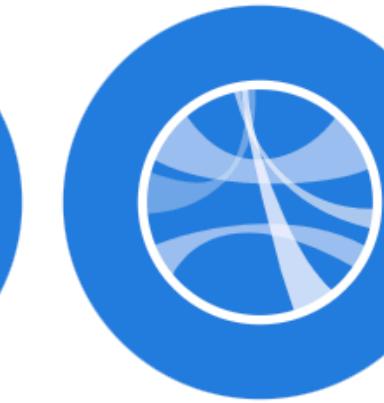
Bullet Graph



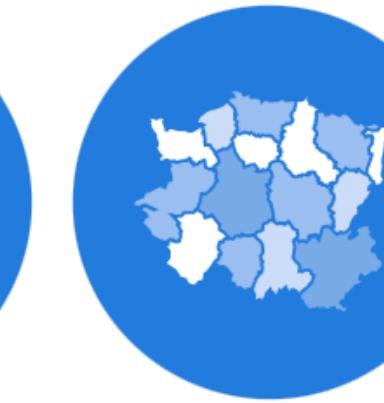
Calendar



Candlestick Chart



Chord Diagram



Choropleth Map

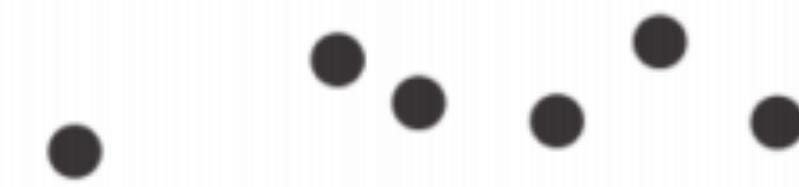


PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE

# Marcas y Canales

**Marca:** elemento geométrico básico, que puede ser clasificado según el número de dimensiones espaciales que requiera.

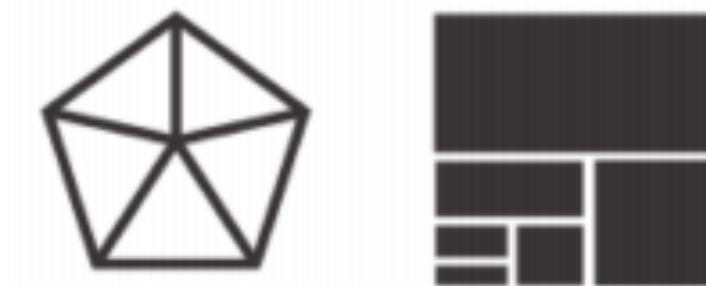
→ Points



→ Lines



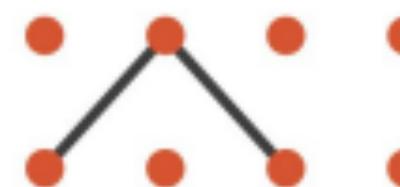
→ Areas



→ Containment



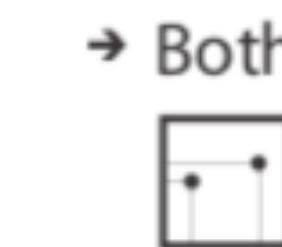
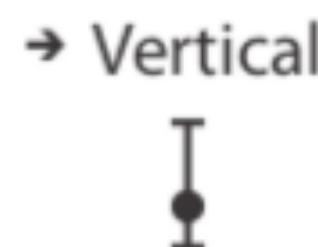
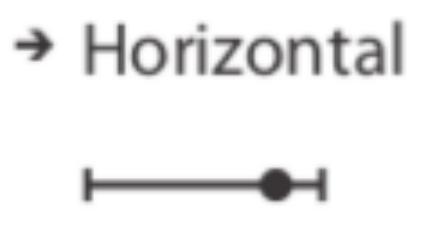
→ Connection



# Marcas y Canales

**Canal:** permite controlar la apariencia de las marcas, independientemente de la dimensionalidad de este elemento primitivo

## ④ Position



## ④ Color



## ④ Shape



## ④ Tilt



## ④ Size

### → Length



### → Area



### → Volume



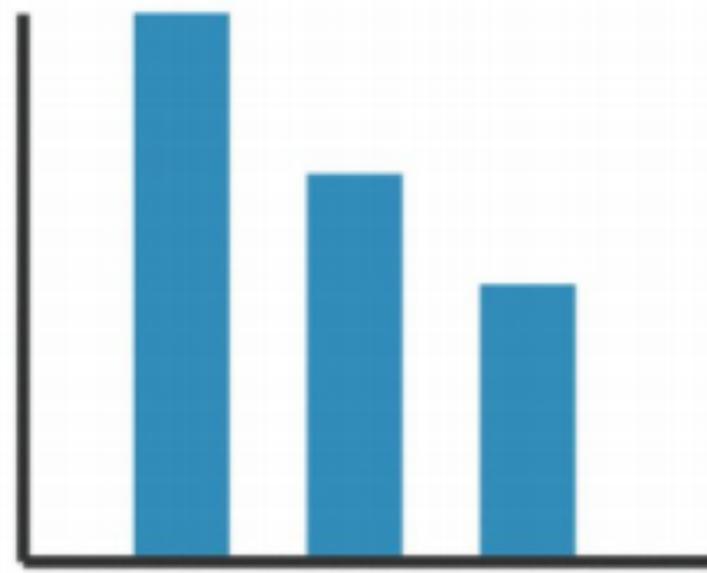
# Marcas y Canales

## Por qué son importantes

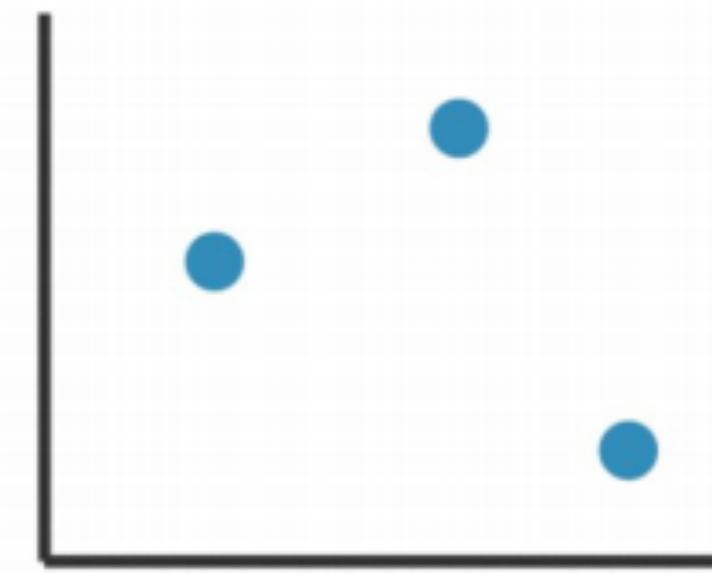
- Pensar en marcas y canales nos entrega los bloques elementales para analizar los encodings visuales.
- El diseño de estos encodings pueden ser descritos como una combinación de: elementos gráficos (marcas) y sus apariencias (canales).
- De esta forma, una visualización muy compleja puede ser desglosada en componentes más simples, que, a su vez, pueden ser analizados en términos de marcas y canales.

# Marcas y Canales

Por qué son importantes



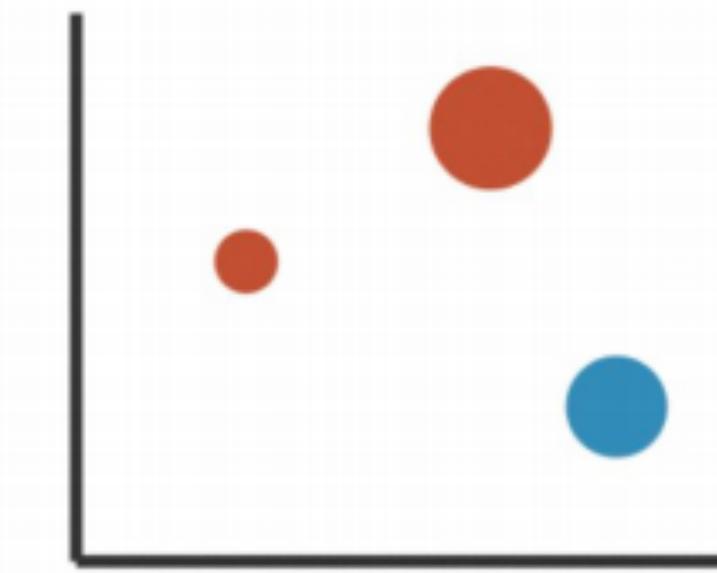
**Marca:** línea  
**Canal:** posición y largo



**Marca:** punto  
**Canal:** posición



**Marca:** punto  
**Canal:** posición y color (hue)



**Marca:** punto  
**Canal:** posición, tamaño y color (hue)

# Marcas y Canales

## Por qué son importantes

### Tipos de canales

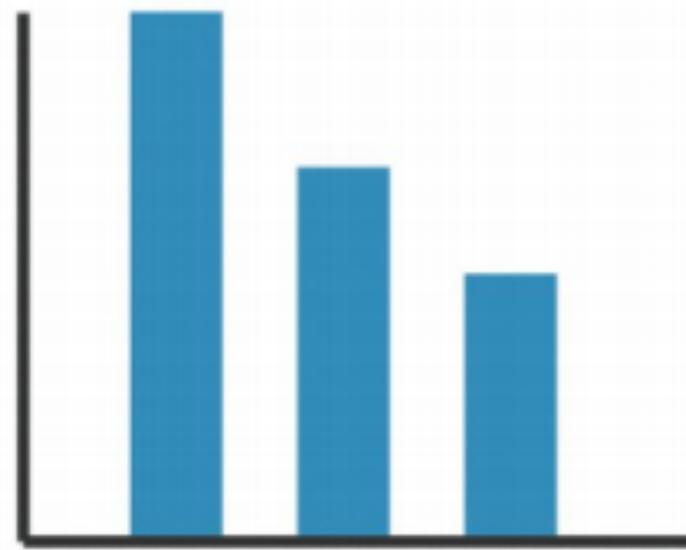
- El sistema de percepción humano tiene dos tipos de modalidades:
- El identity channel permite discernir información sobre qué es algo o dónde se encuentra,
- El magnitude channel, por otra parte, nos permite saber cuánto de ese algo existe.

Con estas dos modalidades, podemos saber, por ejemplo:

- ¿Qué figura es? ¿un círculo, un triángulo, una cruz o un heptágono? **[identity]**
- ¿De qué hue es? ¿rojo, verde, caqui o gris? **[identity]**
- ¿Cuánta saturación tiene ese azul? ¿celeste, azul marino o turquesa? **[magnitude]**
- ¿Dónde está? ¿en qué región se encuentra la marca? **[identity]**
- ¿Qué tan larga es aquella línea con respecto a esta? **[magnitude]**
- ¿Cuánto espacio hay entre ambos rectángulos? **[magnitude]**

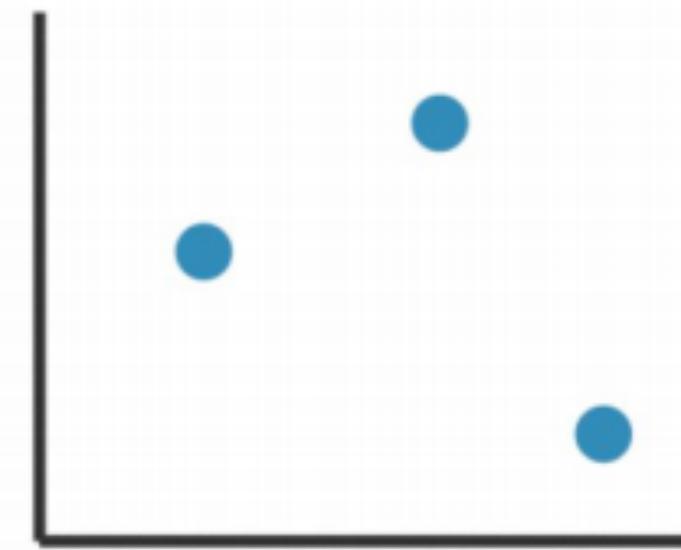
# Marcas y Canales

Por qué son importantes



**Marca:** línea  
**Canal:** posición y largo

1 attr cuantitativo  
1 attr categórico



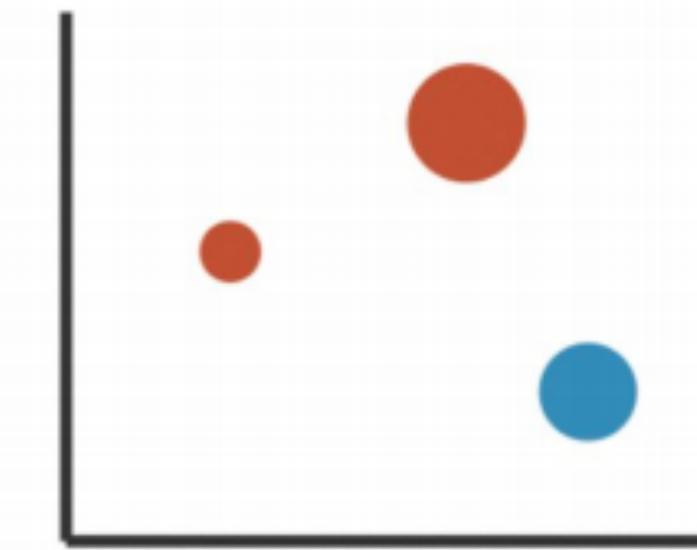
**Marca:** punto  
**Canal:** posición

2 attrs cuantitativos



**Marca:** punto  
**Canal:** posición y color (hue)

2 attrs cuantitativos  
1 attr categórico



**Marca:** punto  
**Canal:** posición, tamaño y color (hue)

3 attrs cuantitativos  
1 attr categórico

# Marcas y Canales

Por qué son importantes

④ Position

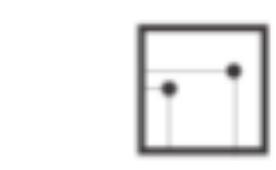
→ Horizontal



→ Vertical



→ Both



④ Color



④ Shape



④ Size

→ Length



→ Area



→ Volume



¿Cómo usarlos?

No todos los canales son iguales: los mismos datos codificados con dos canales visuales distintos resultará en información diferente.

Dos principios guían el uso de canales visuales: expresividad y efectividad.

# Marcas y Canales

Por qué son importantes

## ¿Cómo usarlos? Principio de expresividad

Dicta que el encoding visual debe representar toda (y sólo) la información de los atributos del dataset.

Los datos ordenados deben ser mostrados de tal forma que nuestro sistema perceptual los perciba como ordenados; inversamente, debe ocurrir lo mismo con los datos no ordenados.

Los canales de magnitud funcionan bien con los atributos ordenados, mientras que los de identidad son el match correcto con los atributos categóricos.

# Marcas y Canales

Por qué son importantes

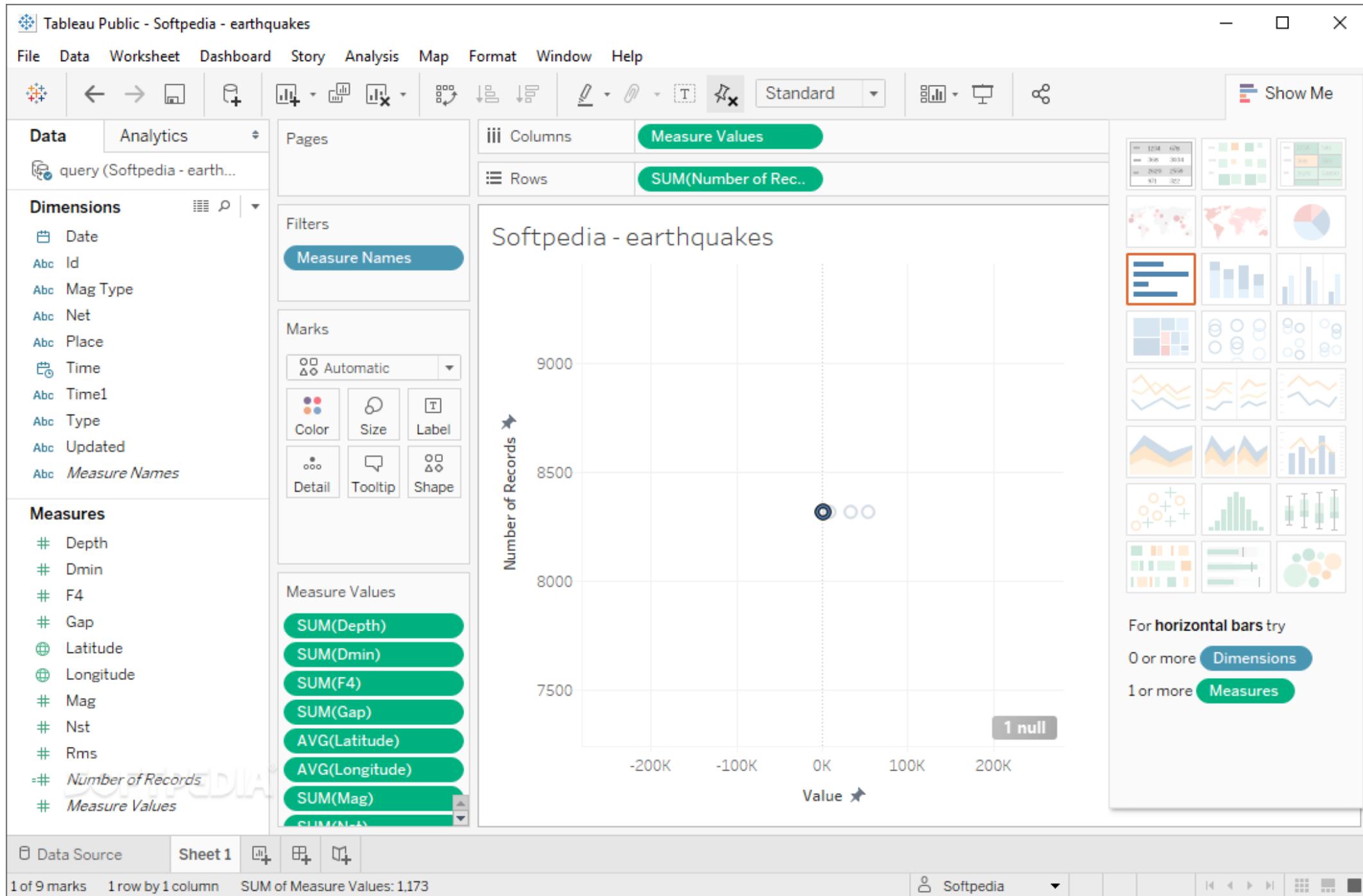
## ¿Cómo usarlos? Principio de expresividad

Dicta que el encoding visual debe representar toda (y sólo) la información de los atributos del dataset.

Si queremos comparar la diferencia de magnitudes, ¿cuál canal es más efectivo?

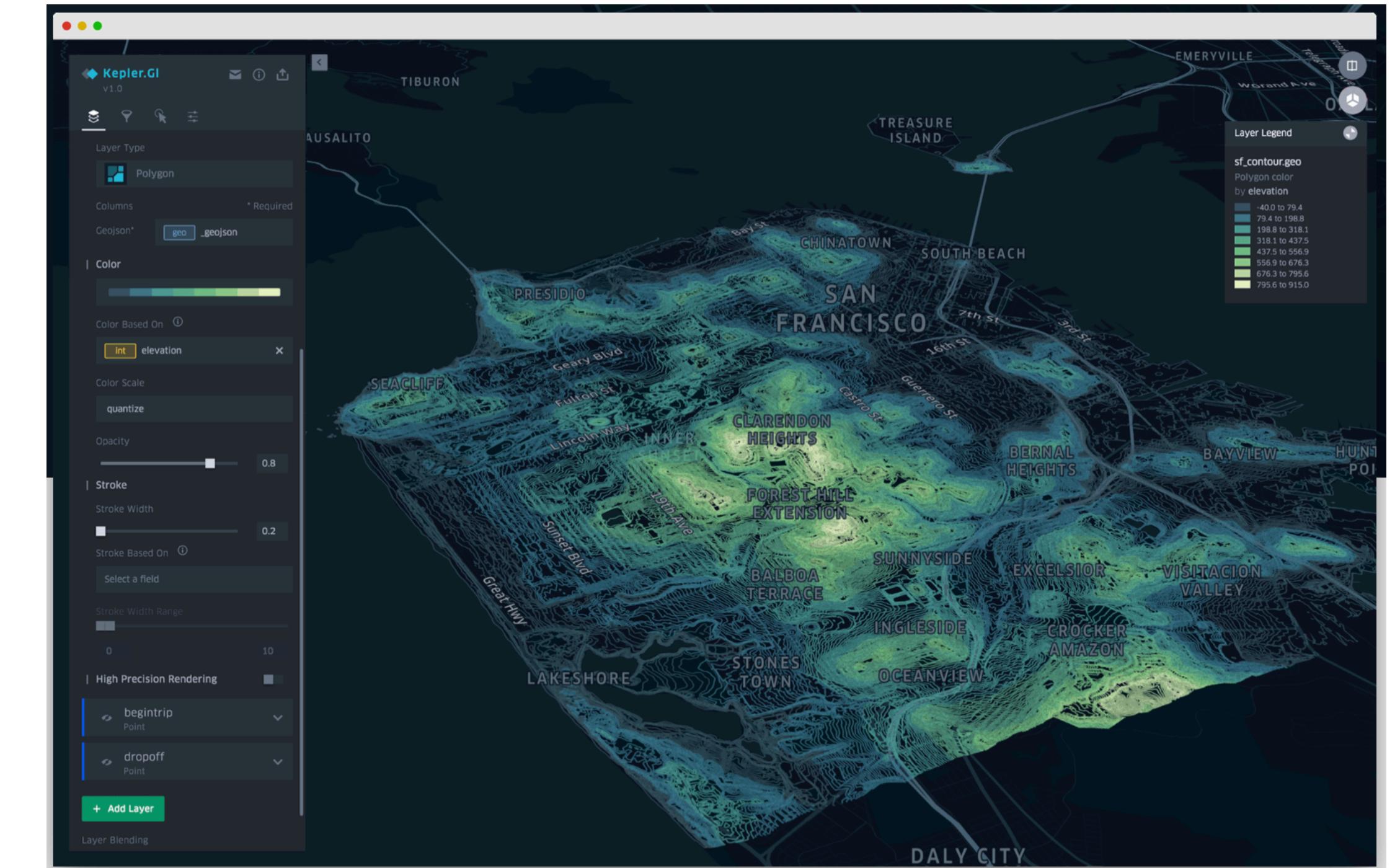


# Herramientas del curso



Tableau

<https://public.tableau.com/s/>



Keplergl

<https://kepler.gl/>

# **Diplomado Big Data y Ciencias de datos: visualización de datos**

Cris Hernández

[cris@crishernandez.co](mailto:cris@crishernandez.co)