

Kirell Benzi

Visualización de Datos y Storytelling

Magíster en Data Science – Facultad de
Ingeniería UDD



Carlos Elías Pérez Pizarro

Agradecimientos: Eduardo Graells-Garrido

Brote de cólera

1854. Londres. Brote de cólera. En solo 10 días, más de 500 personas murieron en un barrio. La misteriosa ola de muertes es especialmente aterrador, porque nadie sabe cuál es la fuente. Nadie, excepto el epidemiólogo [John Snow](#), quien se dio cuenta de que el suministro de agua estaba propagando la enfermedad. Trazó un mapa de todas las muertes con un ingenioso gráfico de barras mapeado y pudo demostrar que, a medida que se acercaba a la bomba de agua de Broad Street, el número de muertes aumentaba. De este modo, pudo probar que el agua contaminada estaba causando el cólera. La determinación del origen de la enfermedad ayudó a poner fin a la epidemia y salvar millones de vidas.

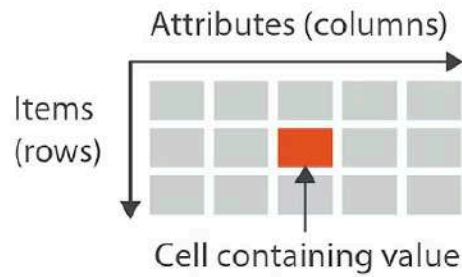


The background of the slide is an abstract, fluid pattern of light blue and white, resembling marbled paper or smoke. The pattern consists of swirling, organic shapes that create a sense of movement and depth. The colors are soft and pastel, with the white areas appearing as highlights within the blue swirls.

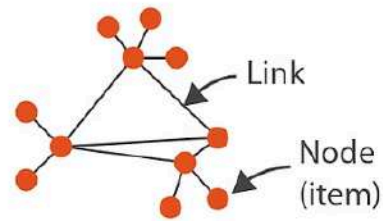
¿Qué visualizar?

Tipos de datos

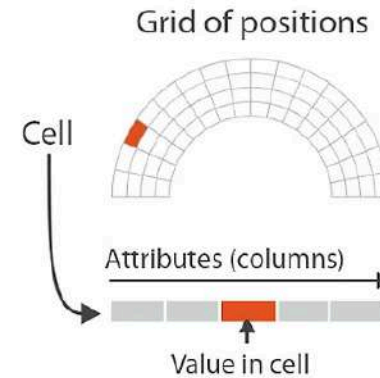
→ Tables



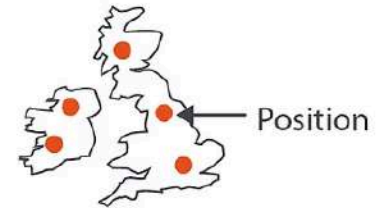
→ Networks



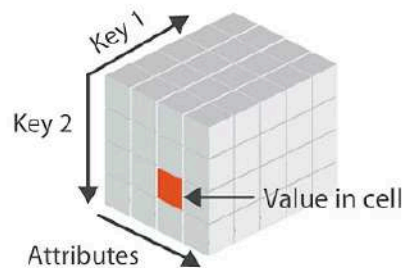
→ Fields (Continuous)



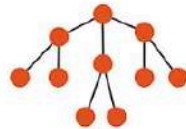
→ Geometry (Spatial)



→ *Multidimensional Table*



→ *Trees*



Tablas

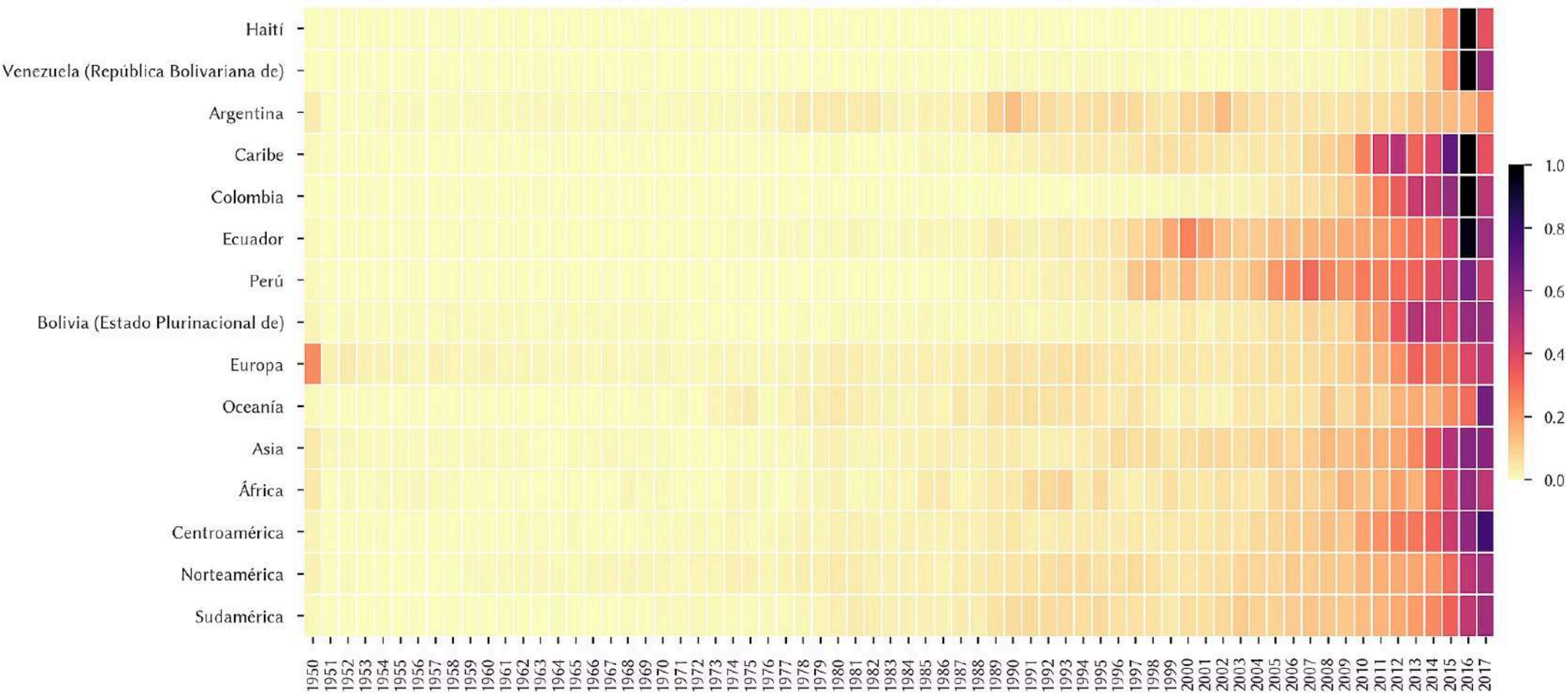
Una tabla es una colección de elementos (filas, observaciones, etc.).

Es un dataset base, en tanto todo lo podemos expresar como una tabla. Incluso una tabla multidimensional (un tensor) puede ser convertido a una tabla bidimensional (una matriz).

	REGION	PROVINCIA	COMUNA	DC	AREA	ZC_LOC	ID_ZONA_LOC	NVIV	NHOGAR	PERSONAN	P07	P08	P09	P10	P10COMUNA	P10PAIS	P11	P11COMUNA	P11PAIS	P12
0	15	152	15202	1	2	6	13225	1	1	1	1	1	73	1	98	998	3	15101	998	1
1	15	152	15202	1	2	6	13225	3	1	1	1	1	78	1	98	998	2	98	998	1
2	15	152	15202	1	2	6	13225	3	1	2	2	2	78	1	98	998	2	98	998	1
3	15	152	15202	1	2	6	13225	3	1	3	5	2	52	1	98	998	2	98	998	1
4	15	152	15202	1	2	6	13225	3	1	4	11	1	44	1	98	998	2	98	998	1

 **Operaciones:** filtrado, agrupación, agregación, ordenamiento, pivot, normalizaciones, etc.

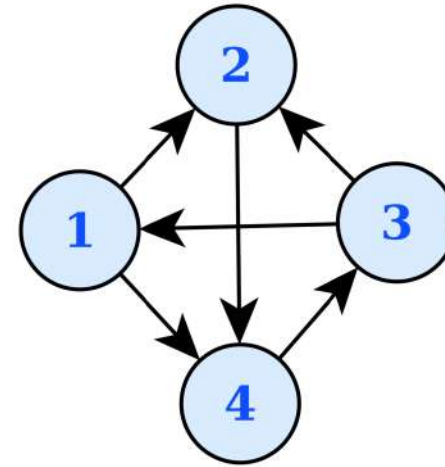
Distribución Porcentual de Inmigrantes por Año de Llegada al País, por País o Grupo de Origen (Censo 2017, P12PAIS_GRUPO)



Redes y Árboles

En una red hay nodos y aristas.

Ambas entidades se pueden representar como tablas. Lo importante es que no se pierda la información de conectividad, dependencia, relación, etc., entre nodos.



i Operaciones: cálculo de importancia de nodos y aristas, caminos de un nodo a otro, detección de comunidades, predicción de atributos.

Red de complejidad económica

Espacio del Producto

RCA

Afinidad

RCA > 1

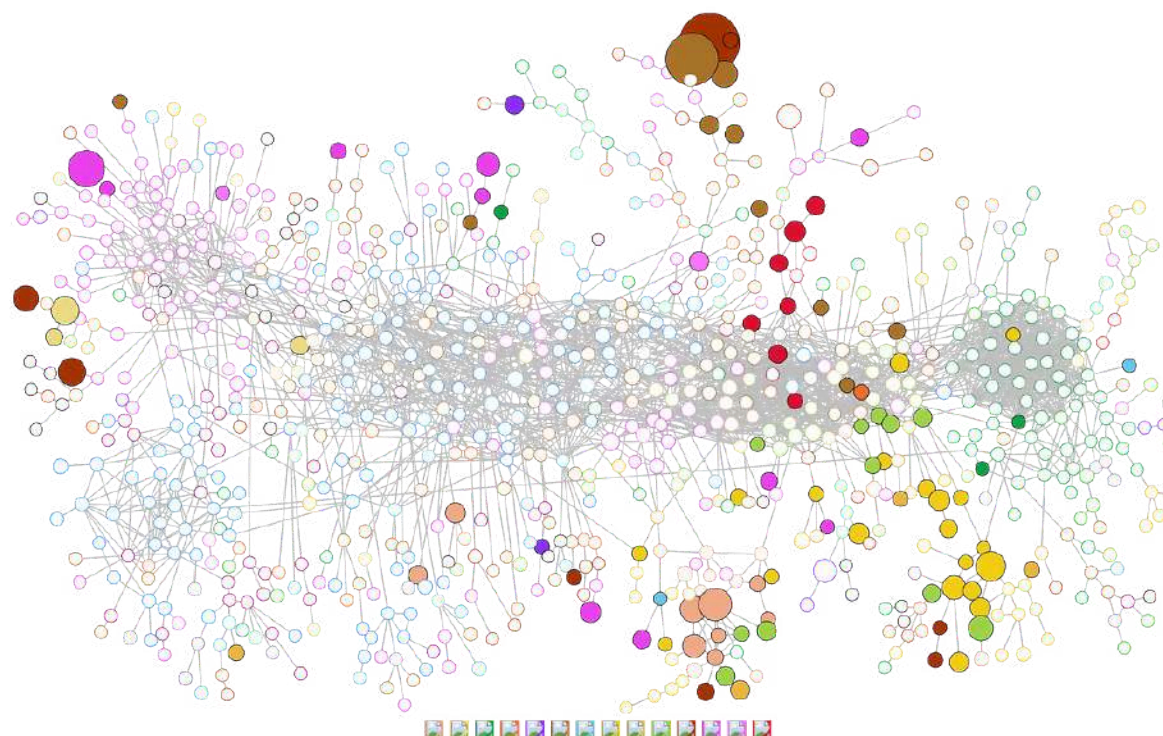
2023

El espacio de productos es una red que conecta productos que probablemente se co-exporten. El espacio de productos se puede utilizar para predecir exportaciones futuras, ya que es más probable que los países comiencen a exportar productos que están relacionados con sus exportaciones actuales. La afinidad mide la distancia entre un producto y todos los productos en los que un país se especializa actualmente.

Un nodo completamente coloreado en el Espacio de Productos indica que un país tiene una Ventaja Comparativa Revelada (RCA) mayor a 1 para un producto específico. Esto significa que el país es relativamente más competitivo en la exportación de ese producto en comparación con el promedio mundial. La RCA se calcula comparando la participación de ese producto en las exportaciones totales del país con su participación en las exportaciones globales.

Fuente de datos

Espacio de Productos de Chile (2023)



Fuente: <https://oec.world/es/profile/country/chl?selector401id=Treemap#product-space>



 Observatorio de Complejidad Económica



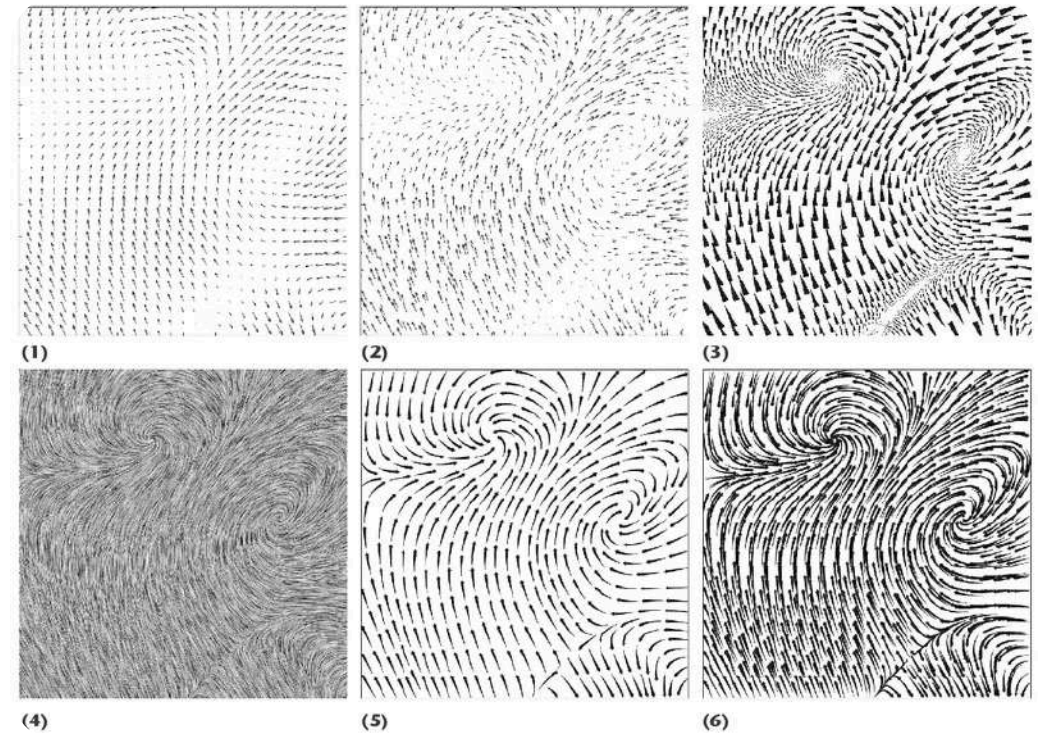
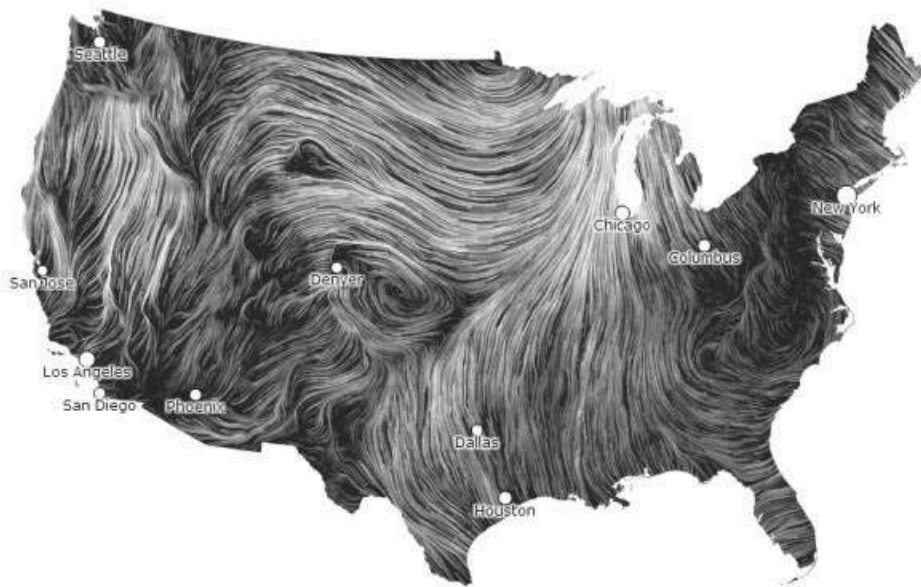
Chile (CHL) Exportaciones, Importaciones y Socios Comercia...

Encuentra las últimas estadísticas comerciales y datos de complejidad económica para Chile.

Campos

Son datos continuos en el espacio que se discretizan en grillas. Las grillas pueden ser regulares o irregulares.

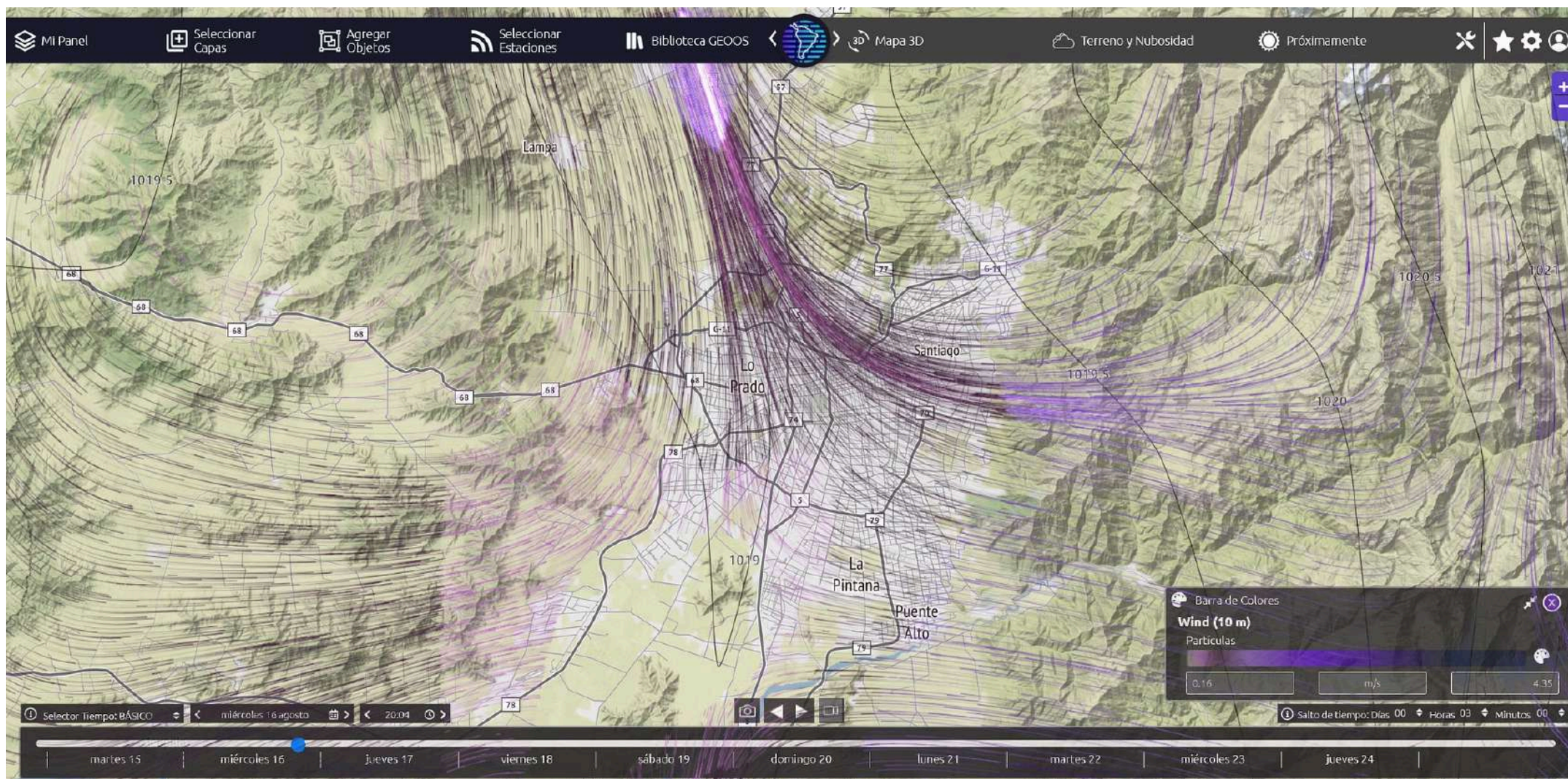
Usualmente son datos provenientes de simulaciones (ej., física), de sensores (ej., viento), de exámenes médicos (ej., escáner).



Visualización de Campos vectoriales.

R. Moorhead, P. Rheingans, C. Johnson, T. S. Yoo, H. Pfister and T. Munzner, "[Visualization Research Challenges: A Report Summary](#)," in Computing in Science & Engineering, vol. 8, no. , pp. 66-73, 2006.

i Operaciones: Interpolación espacial, gradientes y derivadas, comparación temporal o multivariante, líneas de flujo o corriente, etc

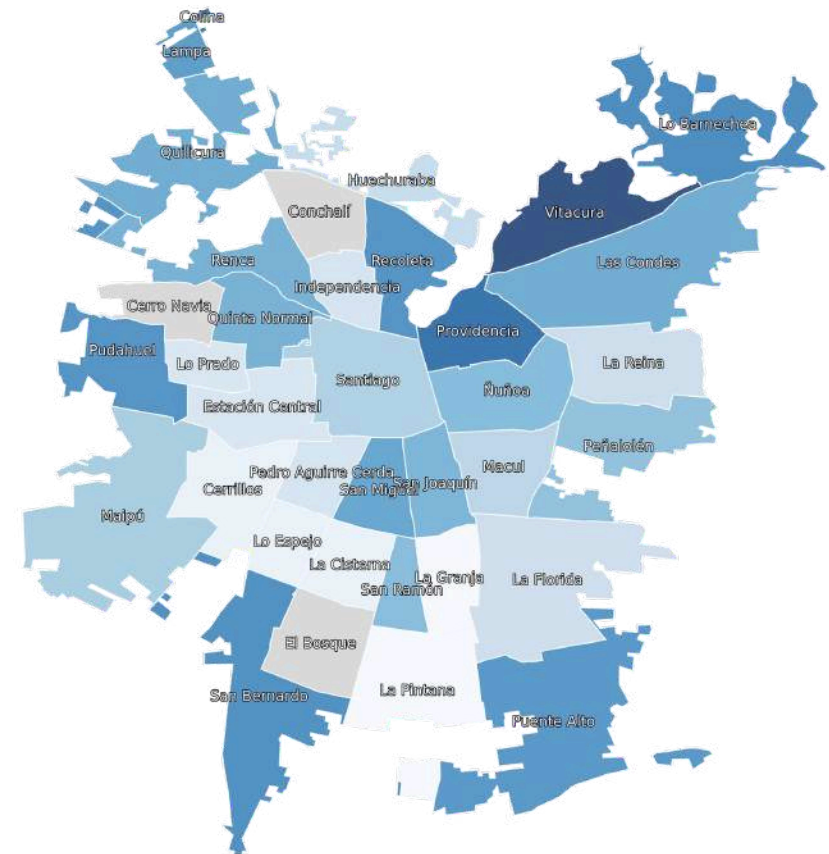


Geografía

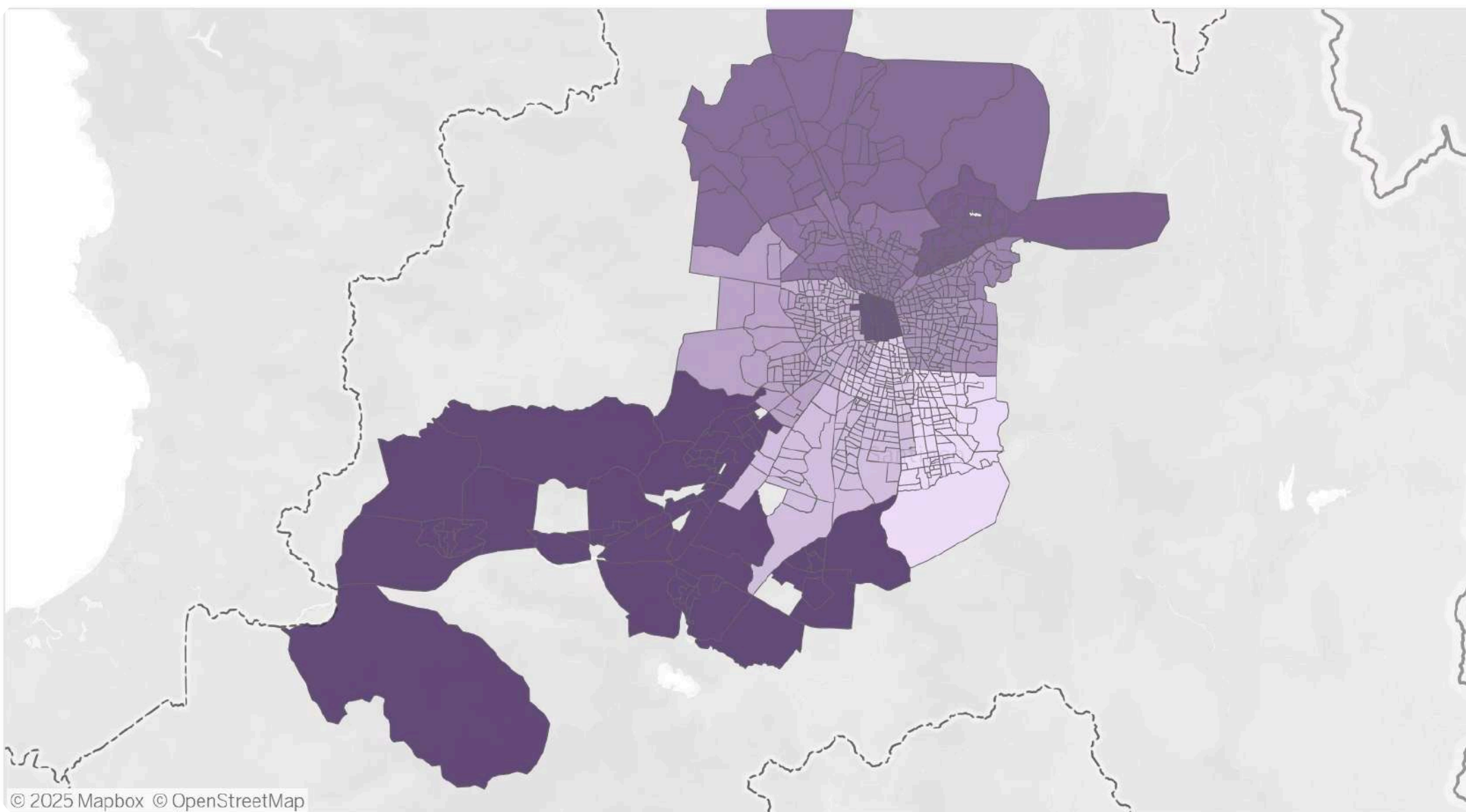
Los tipos de datos anteriores en su mayoría debían ser abstraídos visualmente también.

Los datos geográficos usualmente los podemos representar de manera directa (lo que no significa que sea la mejor manera).

Mapa de Transacciones por Comuna



Operaciones: encontrar, comparar, explorar, contextualizar, manipular.



 reportedemovilidad.ing.uc.cl



Reporte de Movilidad

Muestra la cantidad total promedio de viajes generados durante cada hora en un día laboral normal (lunes a viernes no festivo).

Combiaciones



Tipos de atributo

¿Perro o gato? No hay operaciones aritméticas, ni orden.

➔ Attribute Types

➔ Categorical

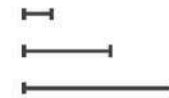


➔ Ordered

➔ Ordinal



➔ Quantitative



Aunque hay un orden, no podemos hacer operaciones aritméticas. Ejemplo: tallas de poleras, o equipos de fútbol.

Aquí sí podemos operar. Podemos calcular algo como: $(a + b) * 0.5$.

➔ Ordering Direction

➔ Sequential



➔ Diverging



➔ Cyclic

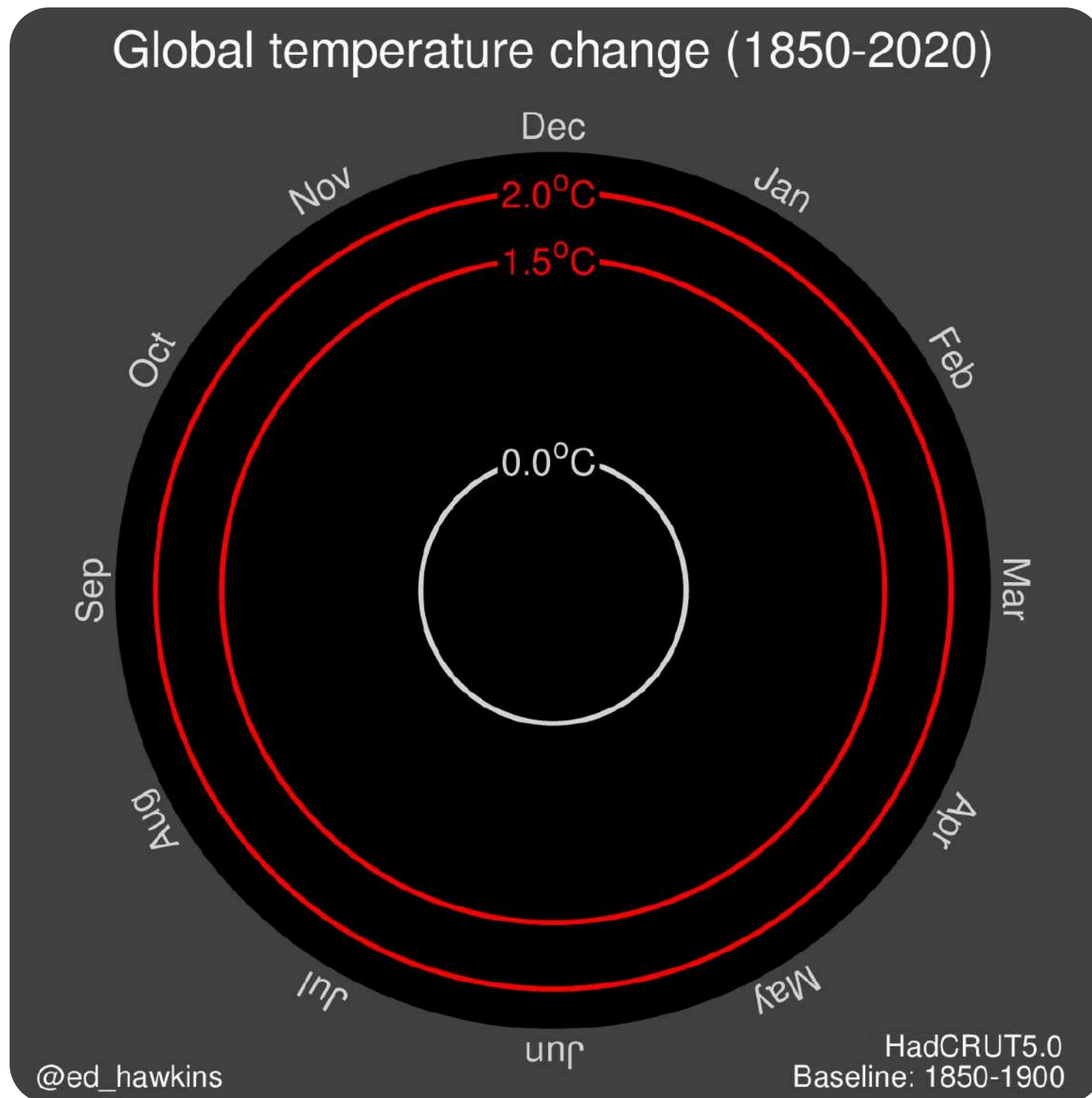


Hay un punto de origen, y de ahí sólo se crece. Ejemplo: edad.

Hay un punto de origen, y de ahí se puede avanzar en dos direcciones opuestas. Ejemplo: temperatura en °C.

Hay un punto de origen, y de ahí se puede avanzar hacia adelante, pero hay atributos que se repiten en ciclos (ejemplo: los meses del año).

Ejemplo de atributo cíclico



Estos atributos se pueden desplegar en gráficos con coordenadas polares, o usando espirales.

La fecha es un atributo interesante, puesto que puede ser **secuencial** (desde el *Big Bang*), **divergente** (en algún punto específico, como el año 0 en occidente), o **cíclica**, poniendo énfasis en los meses más que en los años.

 Climate Lab Book



Climate spirals

Effectively
communicating climate...

¿Por qué es importante conocer los tipos de datasets?

Nos permitirá elegir herramientas para trabajar con los datos

Hay herramientas y modelos para cada tipo de dataset.

Ej.: algoritmos de detección de comunidades (redes) y algoritmos de clustering (tablas) generan resultados similares: grupos de elementos.

En general, los datos tenderán a estar representados como tablas. Pero los datasets **fluyen entre representaciones.**

Ej.: es posible derivar una red a partir de datos de tabla.

The background of the image is an abstract, fluid pattern of light blue and white, resembling marbled paper or smoke. The pattern consists of swirling, organic shapes that create a sense of movement and depth. The colors are soft and pastel, with the white areas appearing as highlights within the blue swirls.

¿Para qué visualizar?

Why?

Actions

→ Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive

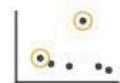


→ Search

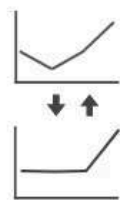
	Target known	Target unknown
Location known	••• Lookup	••• Browse
Location unknown	<•••> Locate	<•••> Explore

→ Query

→ Identify



→ Compare



→ Summarize



Targets

→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One

→ Distribution



→ Extremes



→ Many

→ Dependency



→ Correlation

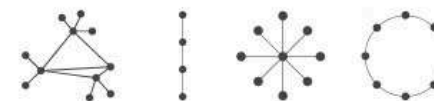


→ Similarity



→ Network Data

→ Topology



→ Paths



→ Spatial Data

→ Shape



What?

Why?

How?

Analizar -> Consumir

➔ Analyze

➔ Consume

➔ Discover

➔ Present

➔ Enjoy

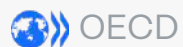
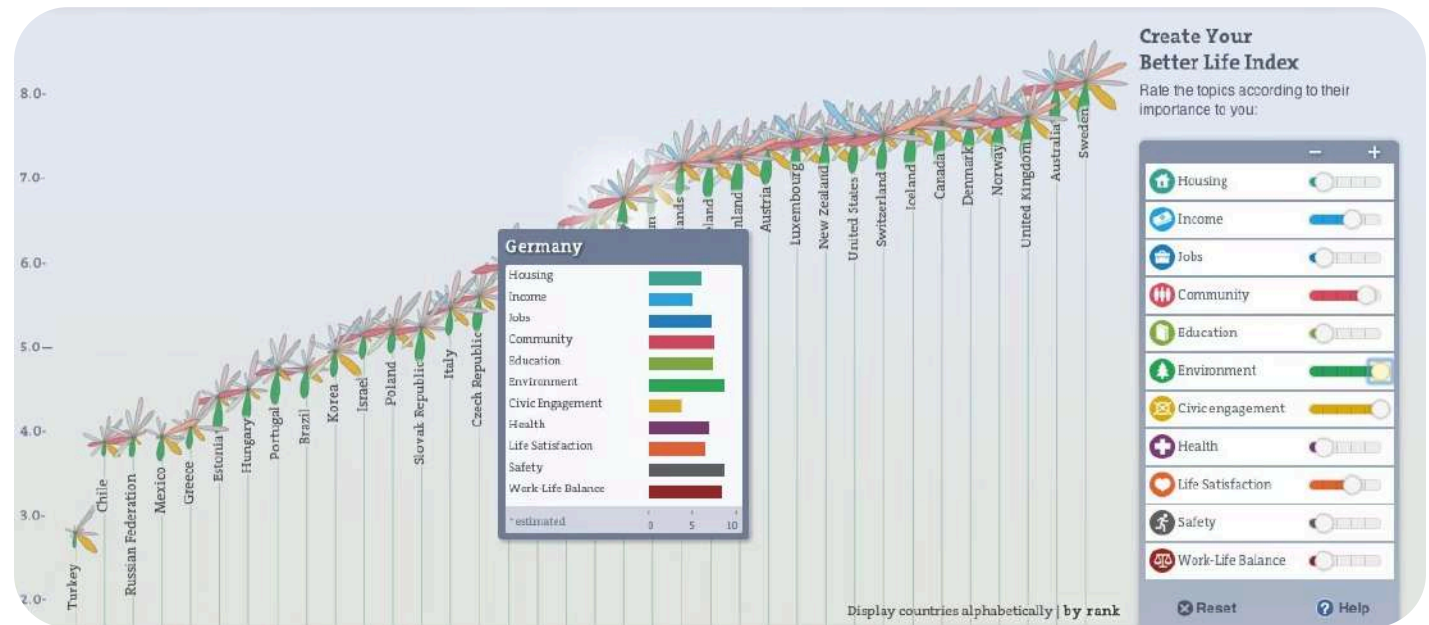


➔ Produce

➔ Annotate

➔ Record

➔ Derive



OECD Better Life Index

There is more to life than the cold numbers of GDP and economic statistics – this Index allows you to compare people's well-being across countries, based on 11 dimensions the OECD has identified as essential, in the areas of...

Analizar -> Producir

→ Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

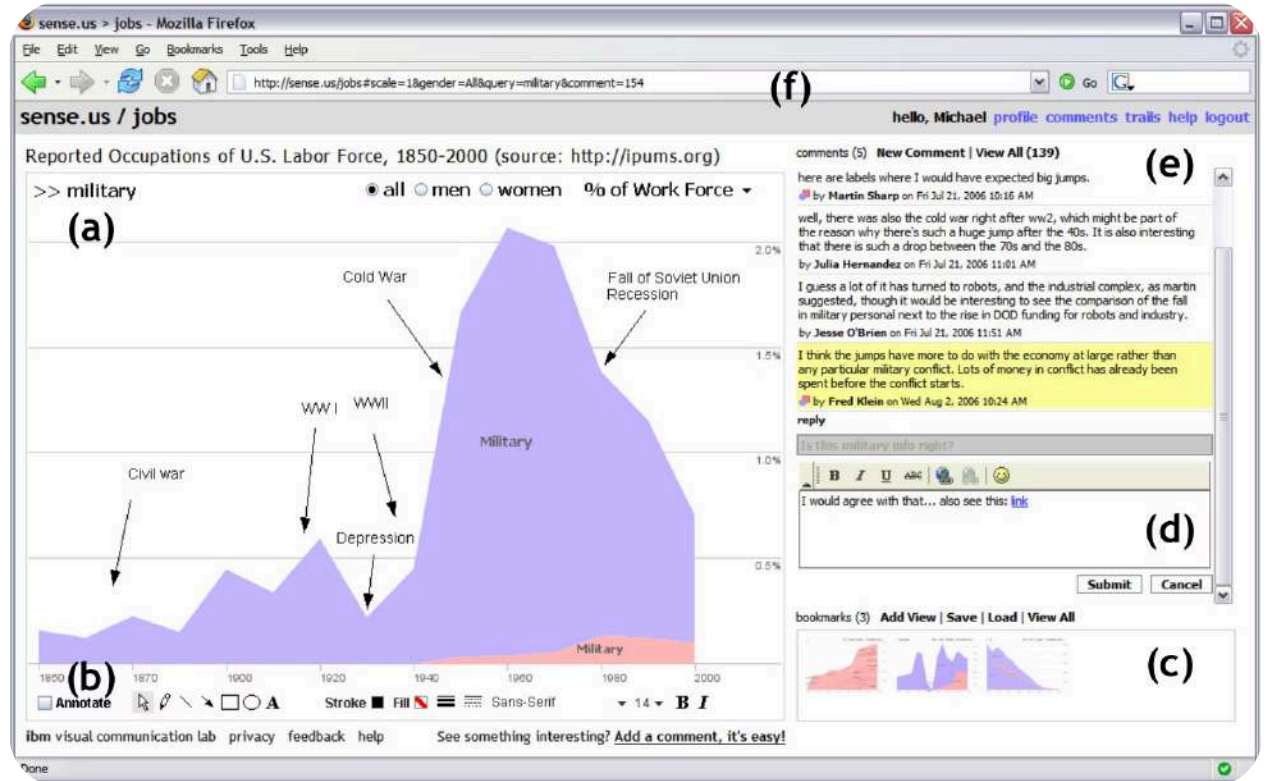
→ Annotate



→ Record



→ Derive



Buscar

Sabemos dónde vamos y qué haremos, pero queremos saber qué hay alrededor.



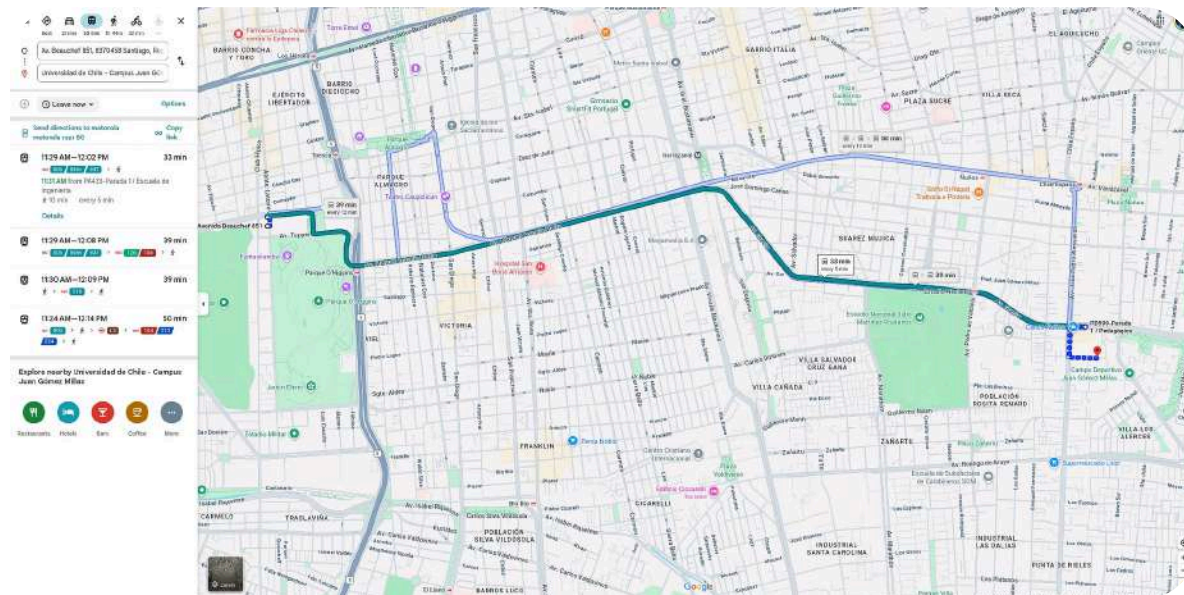
Search

	Target known	Target unknown
Location known	 Lookup	 Browse
Location unknown	 Locate	 Explore

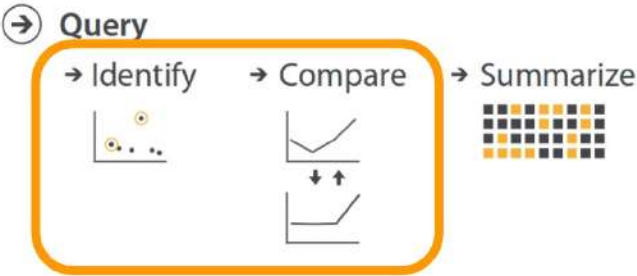
¿Dónde queda el campus TGM de la Universidad?

Sabemos el barrio al que iremos, pero no el lugar.
¡No lo elegimos aún!

¿Qué te tinca hacer y dónde?

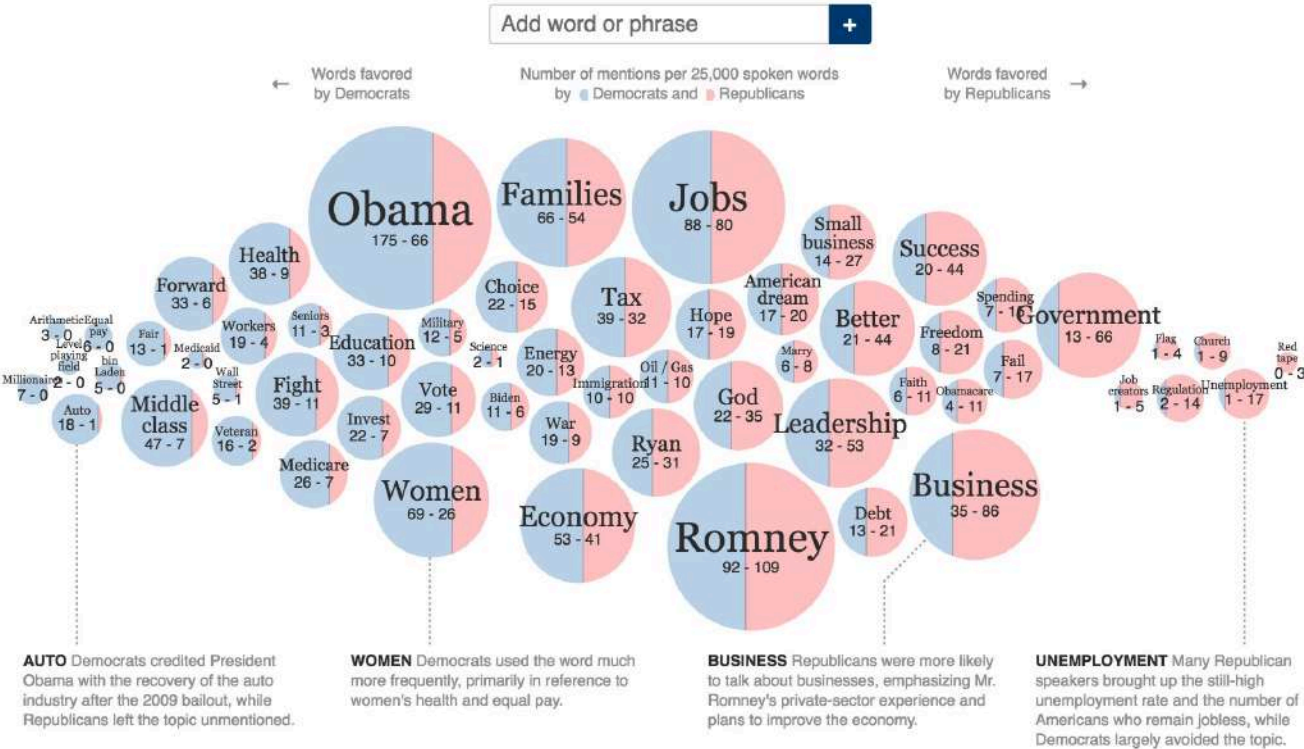


Consultar



At the National Conventions, the Words They Used

A comparison of how often speakers at the two presidential nominating conventions used different words and phrases, based on an analysis of transcripts from the Federal News Service.



Consultar

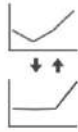


Query

→ Identify



→ Compare



→ Summarize

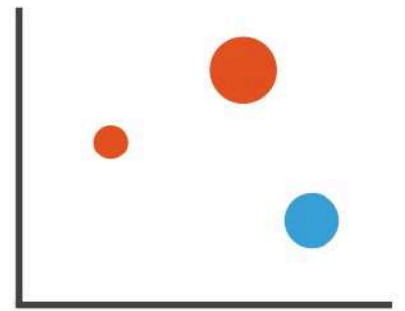
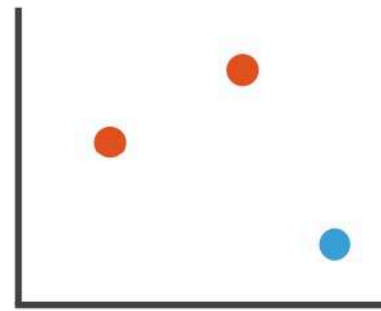
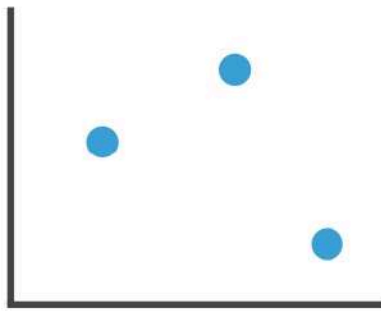
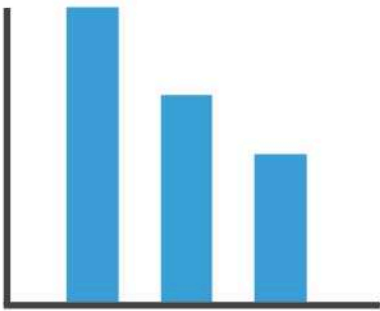


The background of the image is an abstract, fluid pattern of light blue and white, resembling marbled paper or smoke. The pattern consists of swirling, organic shapes that create a sense of movement and depth. The colors are soft and pastel, with the white areas appearing as highlights within the blue swirls.

¿Cómo visualizar?

Codificación visual

La codificación visual es el proceso en el que asignamos propiedades gráficas (formas, colores) a los atributos de nuestros datos.



Este proceso se hace a través de **marcas** y **canales**.

Marcas

Primitivas geométricas que sirven de **bloques de construcción** de elementos gráficos.

➔ Points



➔ Lines



➔ Areas



Canales

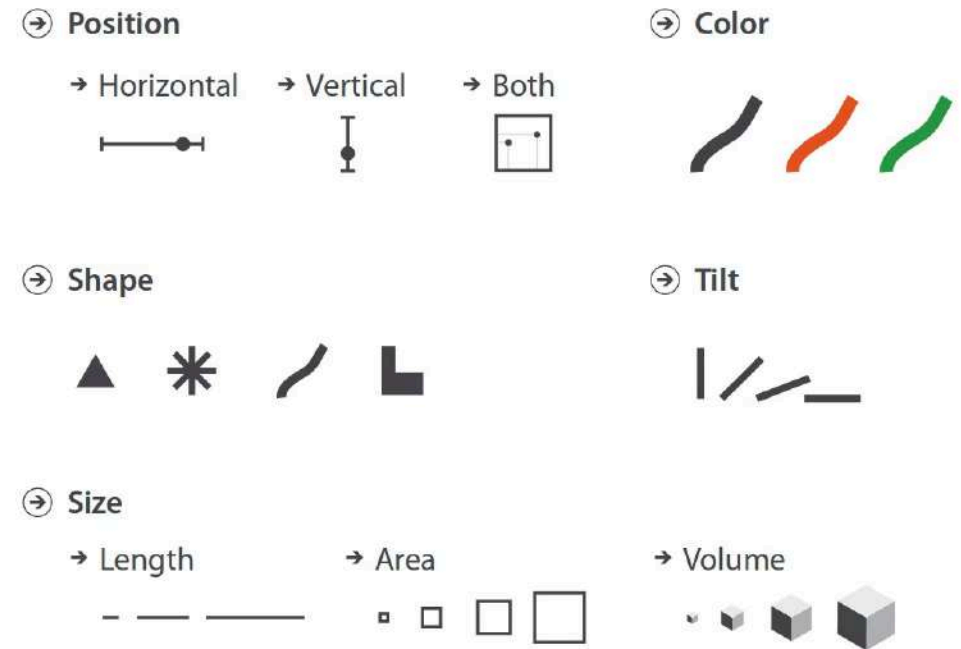
Controlan la **apariencia** de las marcas.

Pueden **combinarse**: a una marca se le pueden aplicar múltiples canales.

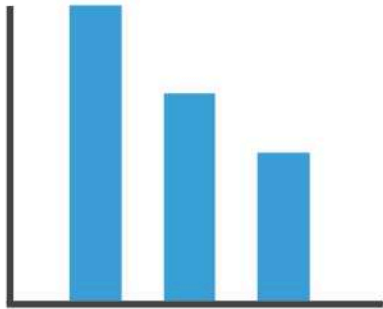
El tipo de marca determina los canales que se pueden utilizar:

Un punto solamente tiene posición (es 0D), pero puede tener un tamaño o forma específica. No puede, por ejemplo, tener inclinación.

Pueden ser **redundantes**: distintos canales en una misma marca codifican el mismo atributo.

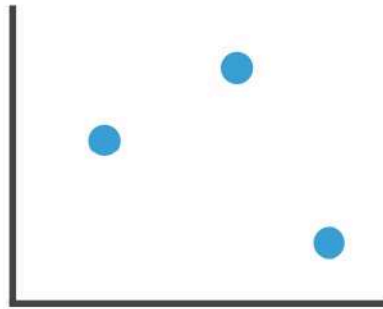


Marcas y Canales



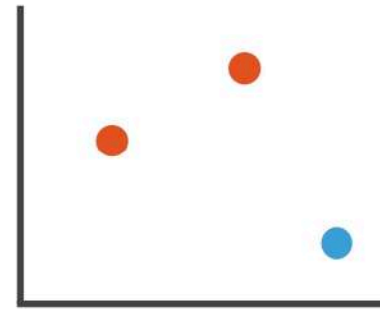
2 canales:
posición horizontal
Largo vertical

marca: línea



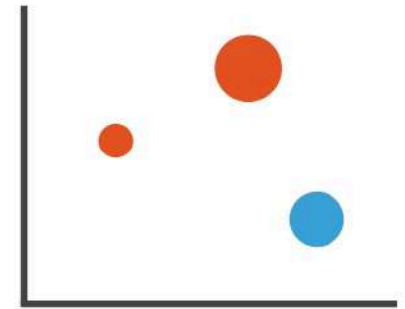
2 canales:
posición vertical
posición horizontal

marca: punto



3 canales:
posición vertical
posición horizontal
tono de color (hue)

marca: punto



4 canales:
posición vertical
posición horizontal
tono de color (hue)
tamaño (área)

marca: punto



Efectividad

Codificar los atributos más importantes de los datos con los canales más efectivos.

Los canales se dividen de acuerdo a lo que expresan: **magnitud** o **identidad**.

Podemos ordenarlos de acuerdo a su efectividad para realizar tareas.

Podemos clasificarlos de acuerdo a su expresividad para datos ordinales o categóricos.

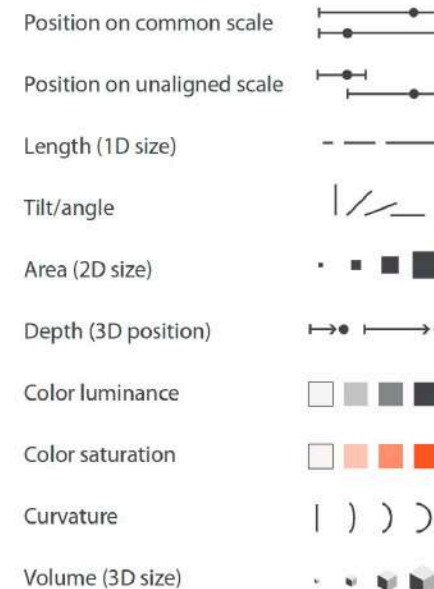


Expresividad

Debe haber coherencia entre el tipo de canal (magnitud, identidad) con la semántica del atributo (cuantitativo, ordinal, categórico).

Channels: Expressiveness Types and Effectiveness Ranks

➔ Magnitude Channels: Ordered Attributes



➔ Identity Channels: Categorical Attributes



Percepción de canales

¿Cómo se determina que un canal es más efectivo que otro?

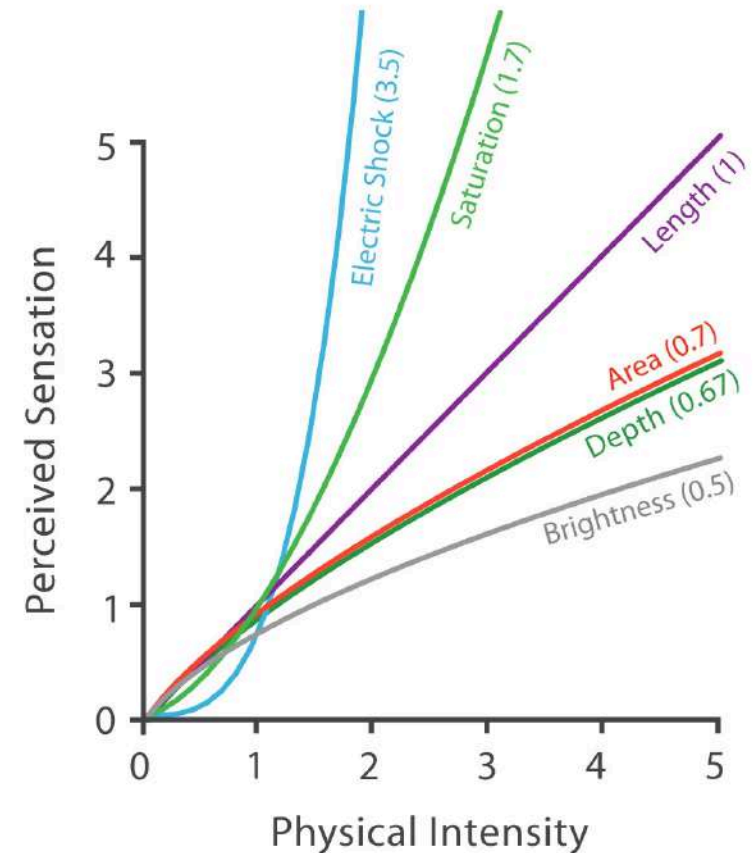
A través de estudios de:

- Precisión
- Discriminabilidad
- Separabilidad
- Saliencia

Imagen: [Ley de Steven](#) sobre percepción de estímulos (precisión).

una formulación empírica de cómo la **intensidad física de un estímulo** se traduce en la **sensación percibida** por el sistema humano.

Steven's Psychophysical Power Law: $S = I^N$



Precisión

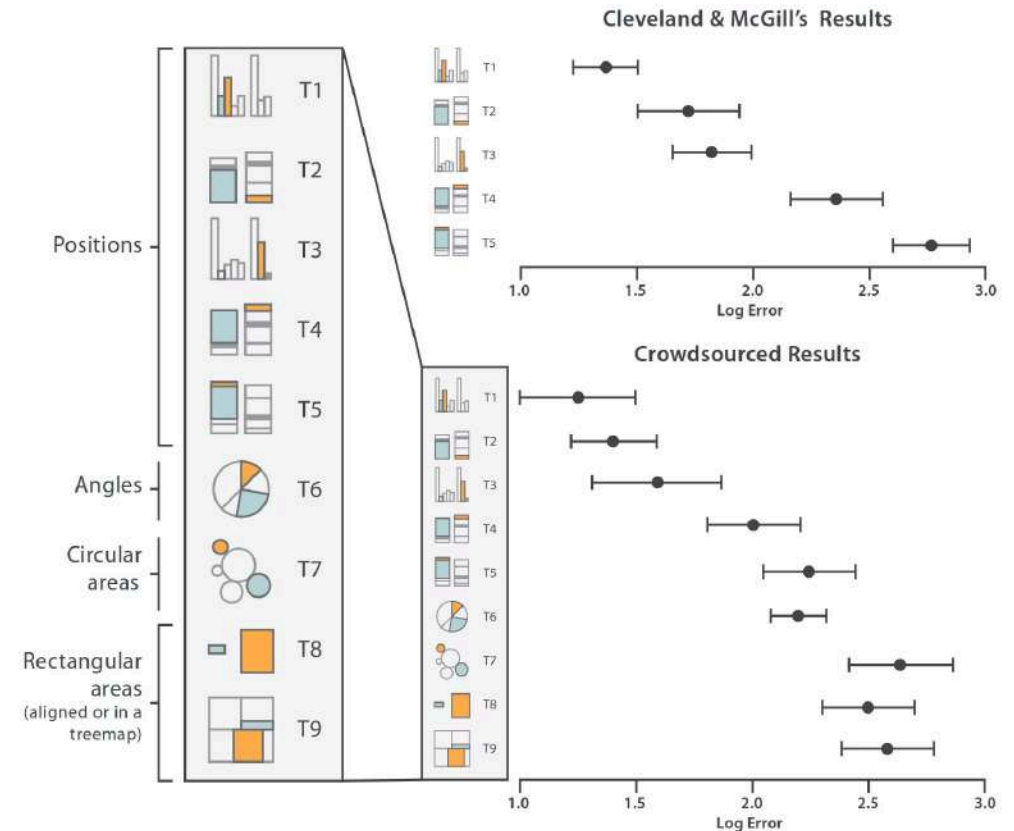
Precisión: Grado en que un canal permite **estimar valores con exactitud**.

Se pueden realizar experimentos en los que se pide a las personas que resuelvan tareas específicas. **Luego se compara lo que calcularon y se compara con el valor real.**

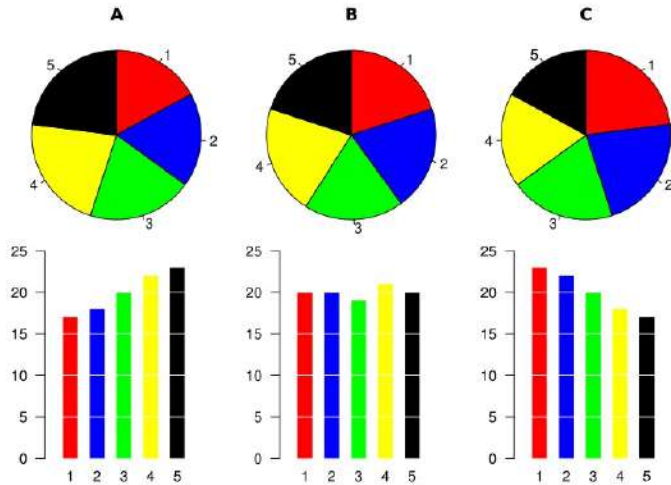
La medida del error obtenido nos permite ordenar características y técnicas.

[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design](#). Heer and Bostock. Proc ACM Conf.

Human Factors in Computing Systems (CHI) 2010, p. 203–212.



Cuidado!



Los ángulos y las áreas no son fáciles de diferenciar.

Estos *pie charts* o gráficos de torta son comunes y se parecen entre sí, sin embargo, codifican datos con distribuciones distintas.

Fuente:

https://en.wikipedia.org/wiki/Pie_chart

❗ Error uploading image.



Distinguibilidad

Capacidad de un canal para **distinguir categorías o niveles distintos sin confusión**. Es una medida de cuántos pasos o niveles perceptivos se pueden diferenciar claramente.

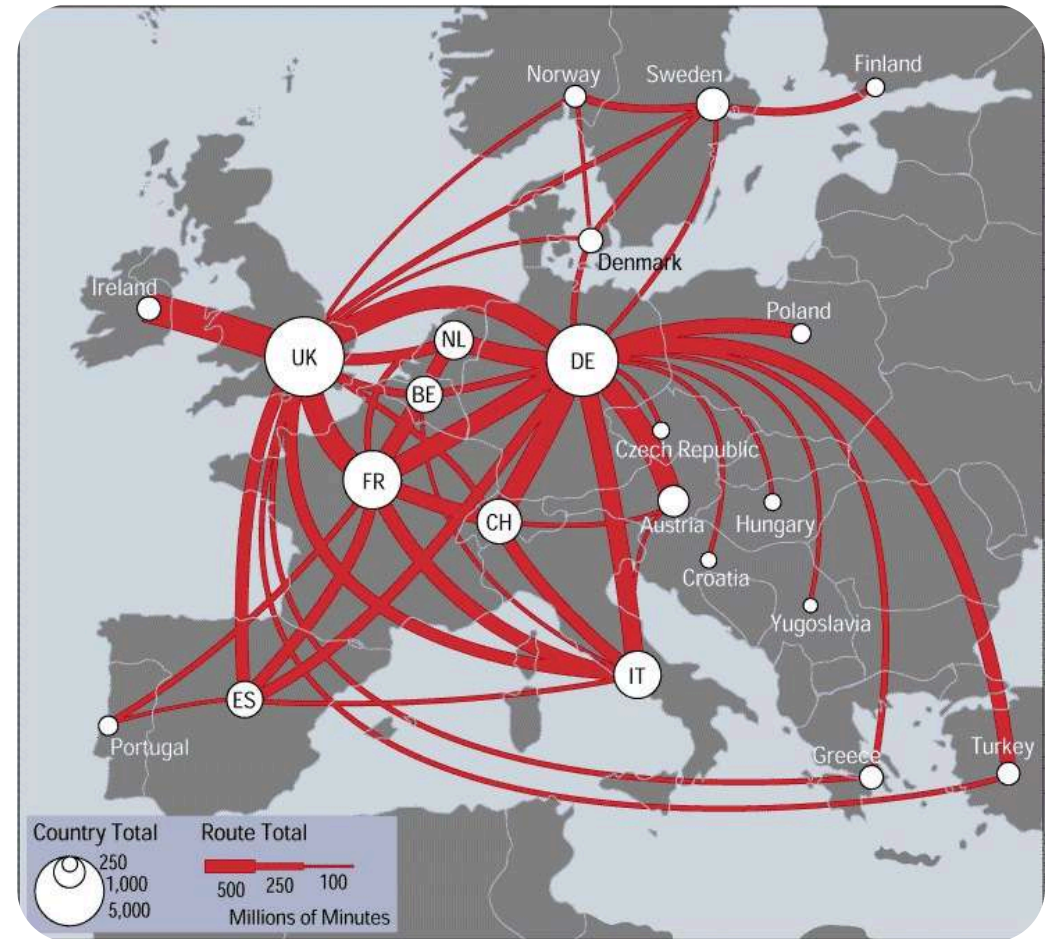
¿Cuánto podemos distinguir un valor de otro?

En el mapa, el grosor de cada línea codifica el tamaño del flujo entre un lugar y otro.

Se utilizan tres grosores distintos, fácilmente identificables.

¿Qué pasaría si fuesen 10 grosores?

¿Diferenciaríamos valores intermedios?



Saliencia

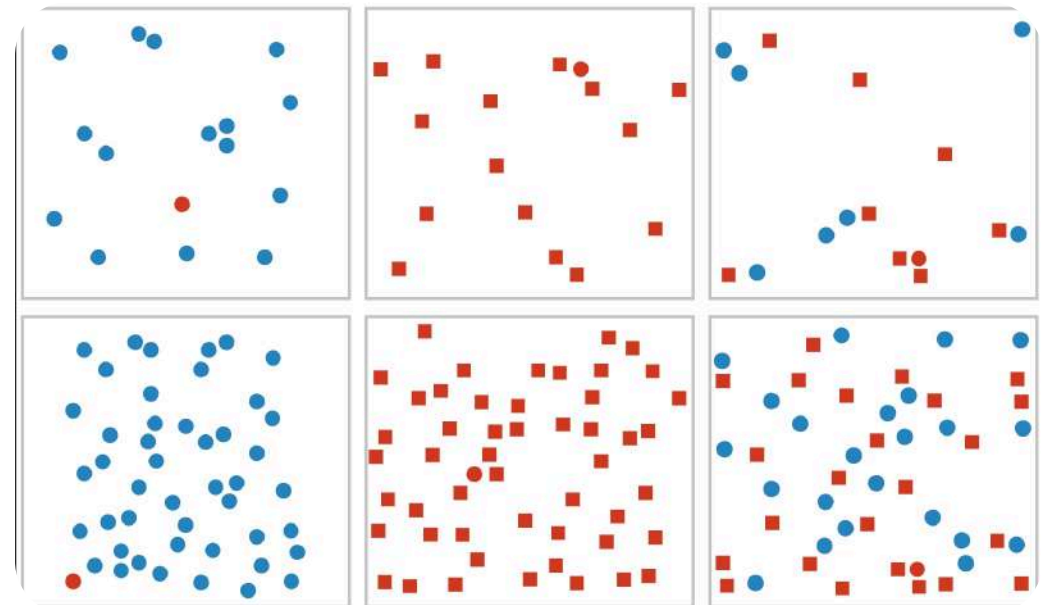
Medida de cuán **fácilmente un canal atrae la atención visual** del observador. Está asociada al concepto de **“pop-out” perceptivo**: elementos que destacan espontáneamente.

¡Encuentra el punto rojo!

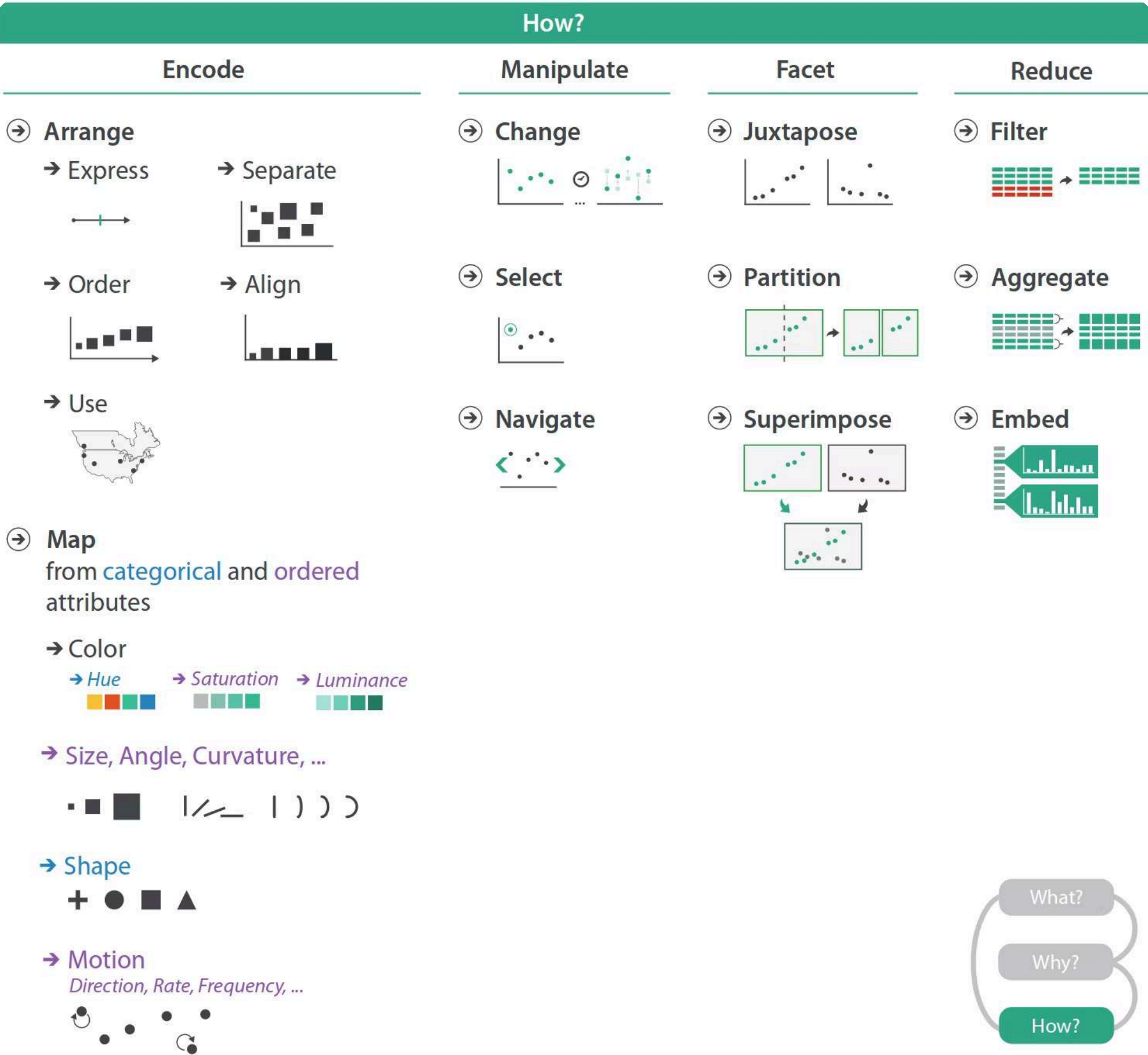
Procesamos canales en paralelo, y la velocidad de proceso suele ser independiente de distractores (primera columna).

La velocidad sí depende del canal y de cómo interactúa con otros canales (segunda columna).




















Cuando hay varios canales, comenzamos a procesar de manera serial. En ese caso la velocidad depende de los distractores (tercera columna).



Codificación visual



Conceptual building blocks of visualization – the three types of *visual encoding techniques*

arranging into meaningful configurations			linking by connectors or boundaries	varying of visual properties
<div>↔ horizontal</div> <div>↑↓ vertical</div> <div>↗↖ third dimension</div> <div>↻ angular</div> <div>↘ radial</div> <div>🌀 spiral</div> <div>≡ vertical array of horizontal axes</div> <div>≡ horizontal array of vertical axes</div> <div>Y angular array of radial axes</div>	<div> picturing</div> <div> mapping (of spatial locations)</div> <div> positioning along a coordinate axis</div> <div> extending along a coordinate axis</div> <div> diverging along a coordinate axis</div> <div> ranging along a coordinate axis</div>	<div> proportional space-filling</div> <div> positioning into category slots</div> <div> positioning into ordered slots</div> <div> spatial ordering</div> <div> nesting</div> <div> coupling by adjacency</div>	<div> connecting</div> <div> grouping by boundary</div>	<div> colour coding</div> <div> gradient coding</div> <div> shape coding</div> <div> sizing</div> <div> unit-based tallying</div>

The **orientations** shown above can be used with **arranging** techniques marked with ●