



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Resumen Final

Visualización de Información

IIC2026

Profesor: Denis Parra

Última Clase

- Material del curso en sitio web: <https://puc-infovis.github.io/>
- Además en Github: <https://github.com/PUC-Infovis/syllabus-2018>
- Tópicos de esta clase:
 - Cosas que no deberían olvidar al terminar este curso
 - Algunos aspectos importantes revisados en los casos de uso
 - ¿Cómo puedo continuar estudiando/investigando en Visualización de Información?

Responder Encuesta para Agendar Presentación Final

- <https://goo.gl/forms/n2bSYW6FQ7vabVs53>

Disponibilidad horarios presentaciones finales

Las presentaciones finales de proyecto del curso IIC2026 - Visualización de Información se llevarán a cabo el 29 de noviembre, 2018. Se realizarán en el espacio del Atelier de INRIA, en el Centro de Innovación. Dadas las restricciones de espacio es necesario dividir las presentaciones en horarios, y debido al número de grupos y a que será en época de exámenes, este cuestionario busca revisar posibles topes con otras evaluaciones que podrían tener los alumnos.

*Obligatorio

Nombre de grupo *

Tu respuesta

¿En qué horarios no pueden presentar? *

Recuerden, todos los integrantes de su grupo deben estar presentes para su presentación. Marquen aquellos horarios donde alguno de ustedes tenga tope.

- 9:00 - 11:00
- 11:00 - 13:00
- 14:00 - 16:00
- Podemos presentar en todos.

Justificación u observaciones de restricciones horarias *

Competencias del curso I

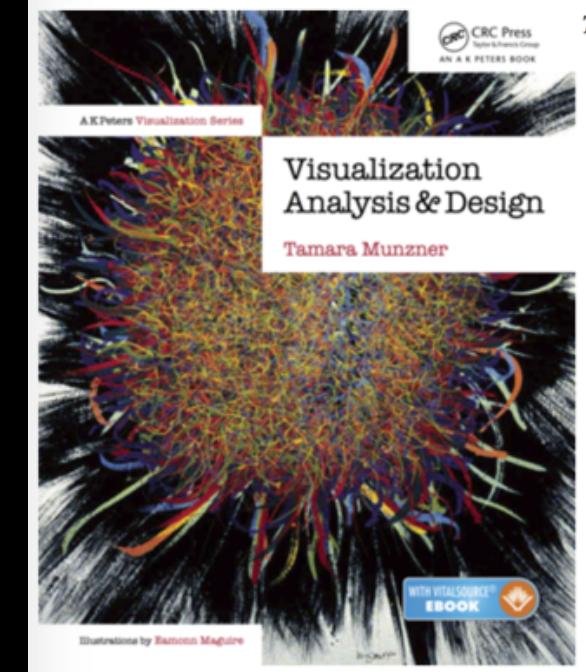
- Aplicar un proceso de diseño para crear visualizaciones efectivas,
- Llevar ideas a prototipos concretos, con la ayuda de bosquejos y software,
- Utilizar principios de percepción y cognición humana en visualización,
- Exponerse a distintos dominios de datos (e.g. redes, textos, cartografía),

Competencias del curso II

- Aplicar distintos métodos de visualización para un rango variado de *datasets*,
- Evaluar una visualización de forma crítica, pudiendo además sugerir mejoras,
- Trabajar como miembro en un equipo para sacar adelante un proyecto.

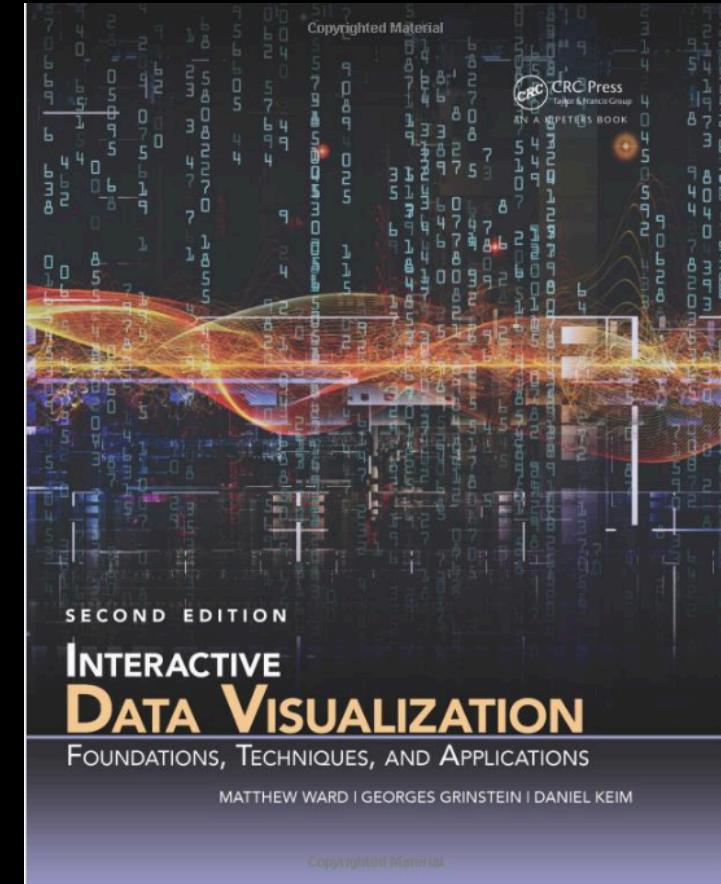
Material

- El *backbone* de este curso está basado en el trabajo de Tamara Munzner, a partir de su libro ***Visualization: Analysis & Design***.

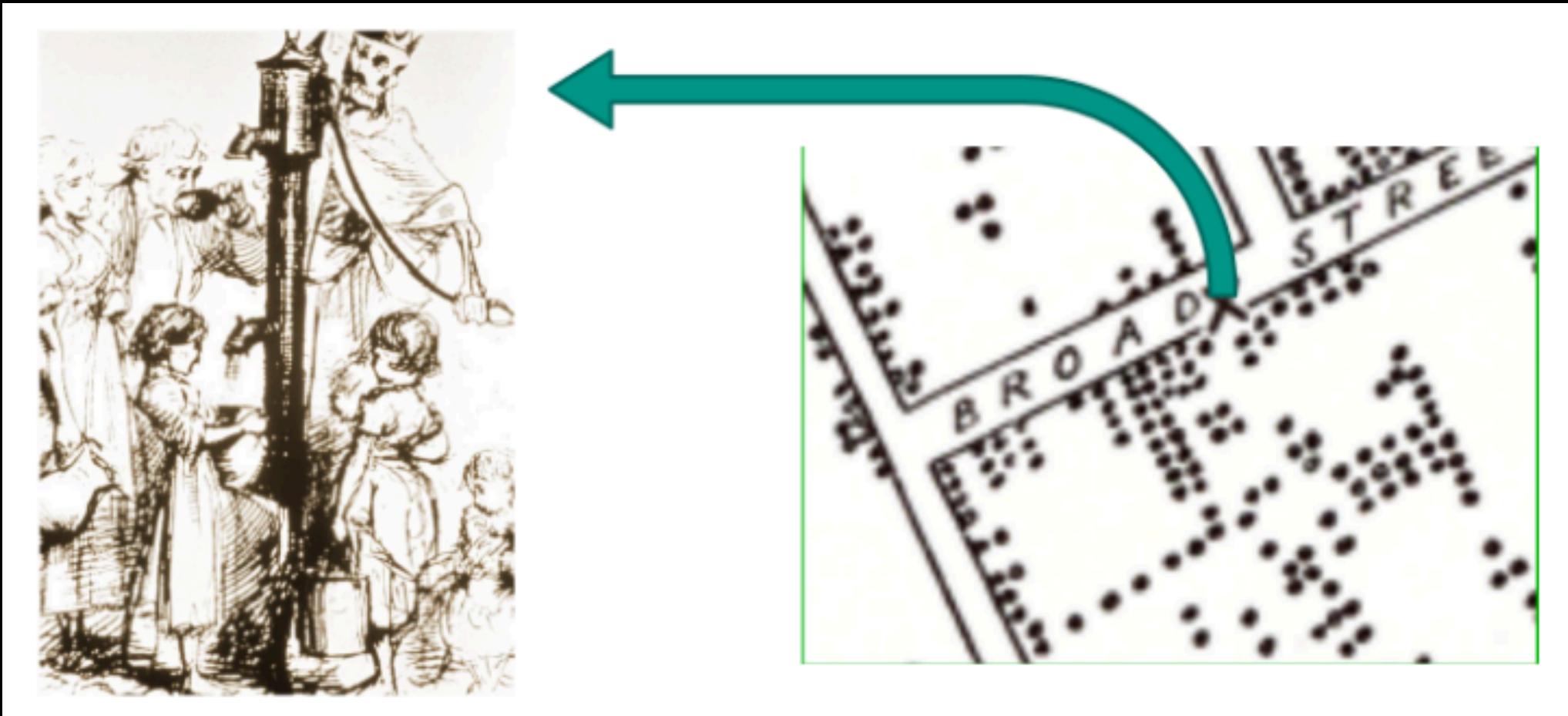


Material

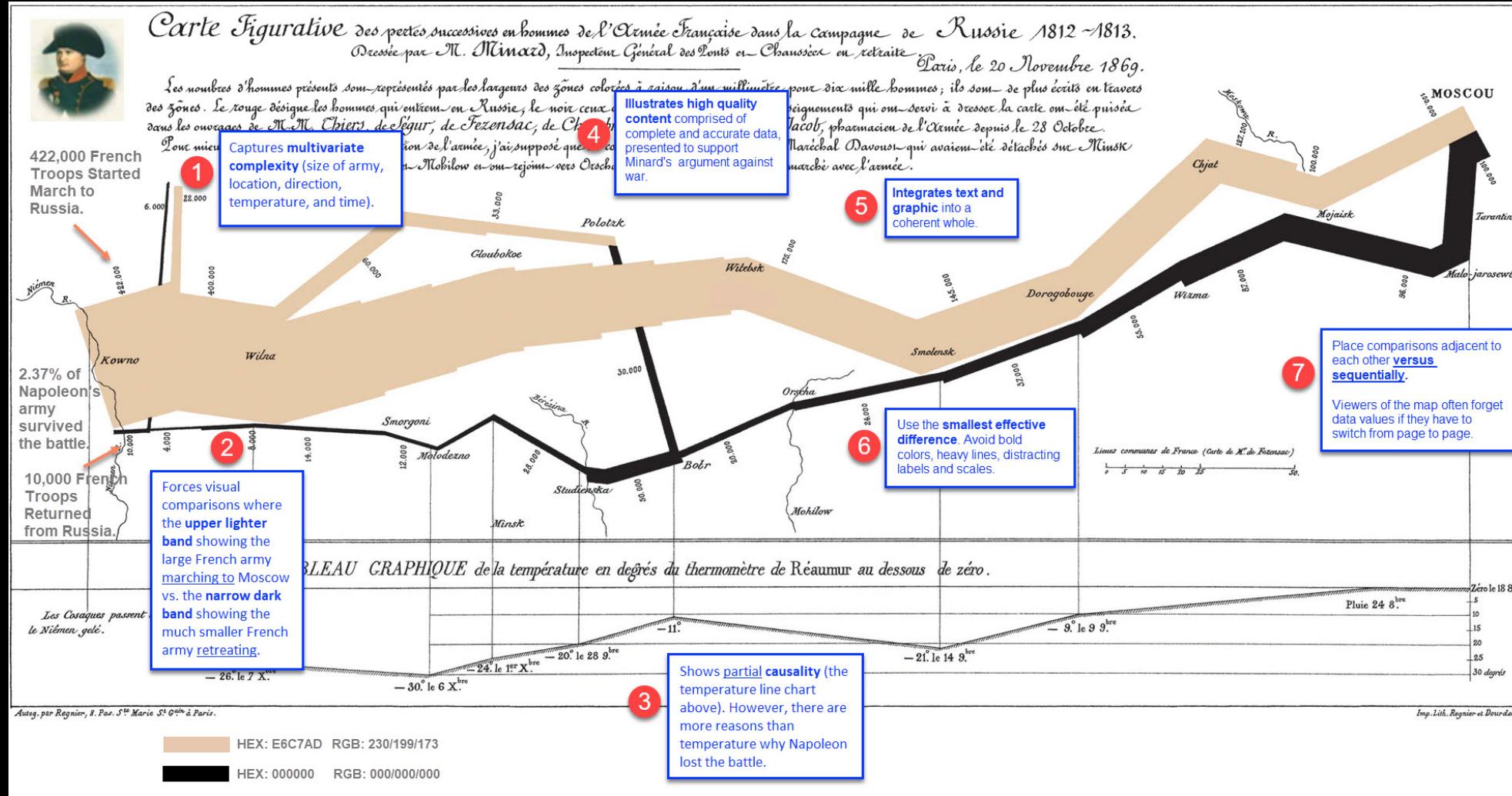
- Interactive Data Visualization: Foundations, Techniques, and Applications, de Ward, Grinstein y Keim.



Cólera en Londres: John Snow (1854)



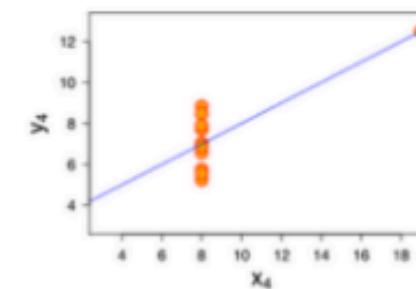
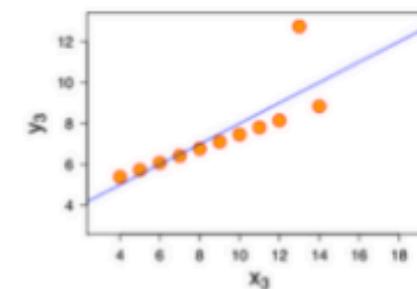
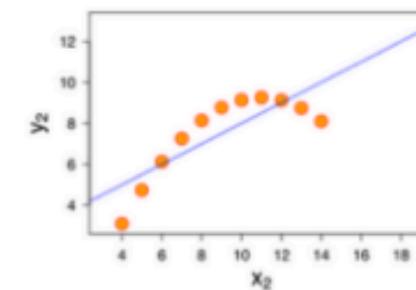
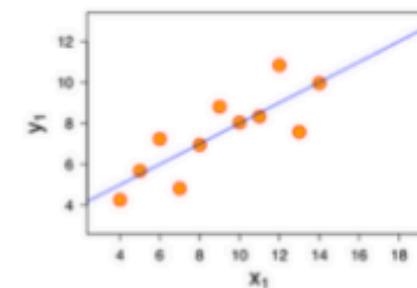
Invasión de Napoleón a Rusia: Minard (1869)



Anscombe (1973)

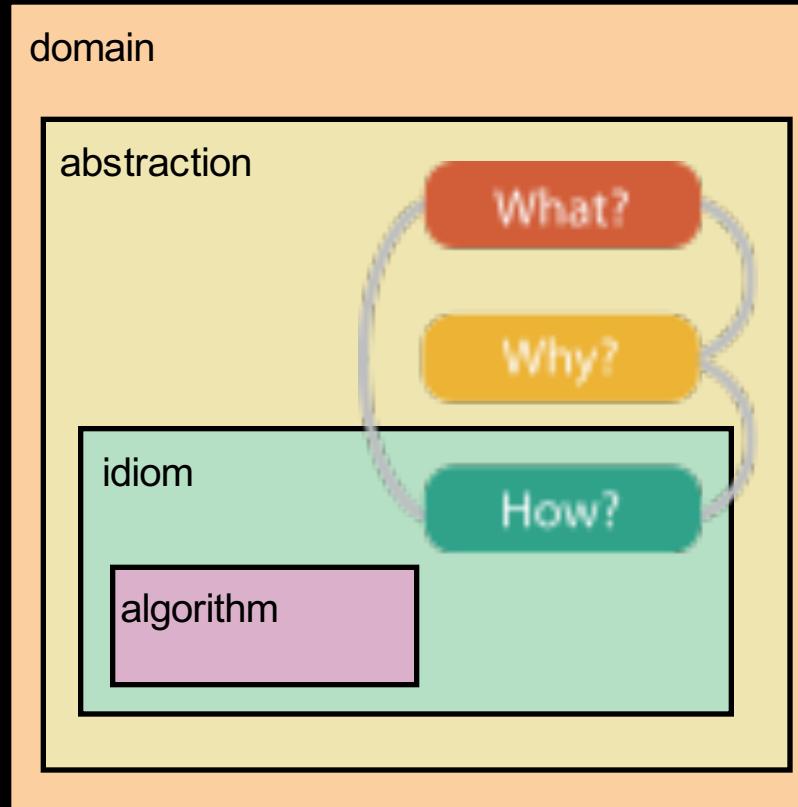
- A veces los modelos estadísticos, de predicción, nos cuentan sólo una parte de la historia

Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	plus/minus 0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places



F. J. Anscombe (1973) Graphs in Statistical Analysis, The American Statistician, 27:1, 17-21, DOI: [10.1080/00031305.1973.10478966](https://doi.org/10.1080/00031305.1973.10478966)

Framework de Tamara Munzner



[A Nested Model of Visualization Design and Validation
Munzner. *IEEE TVCG* 15(6):921-928, 2009
(Proc. InfoVis 2009).]

Framework de Tamara Munzner

- What: Tipos de Datos y de Datasets

Dataset and data types

→ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

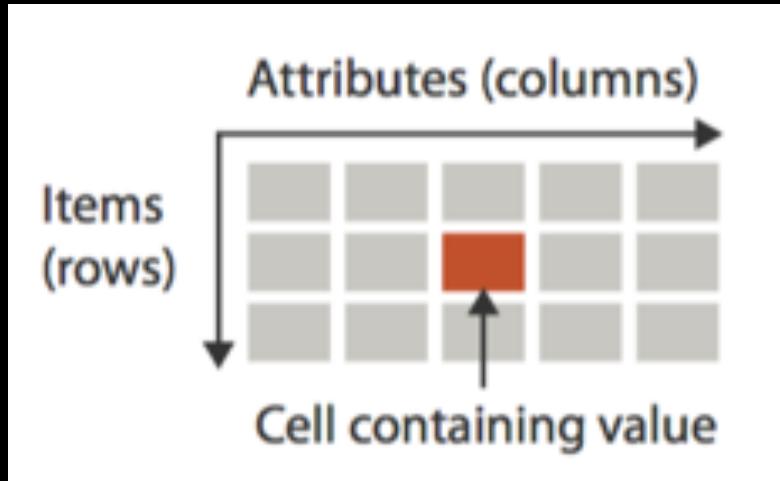
→ Data Types

→ Items → Attributes → Links → Positions → Grids

→ Dataset Availability

→ Static → Dynamic

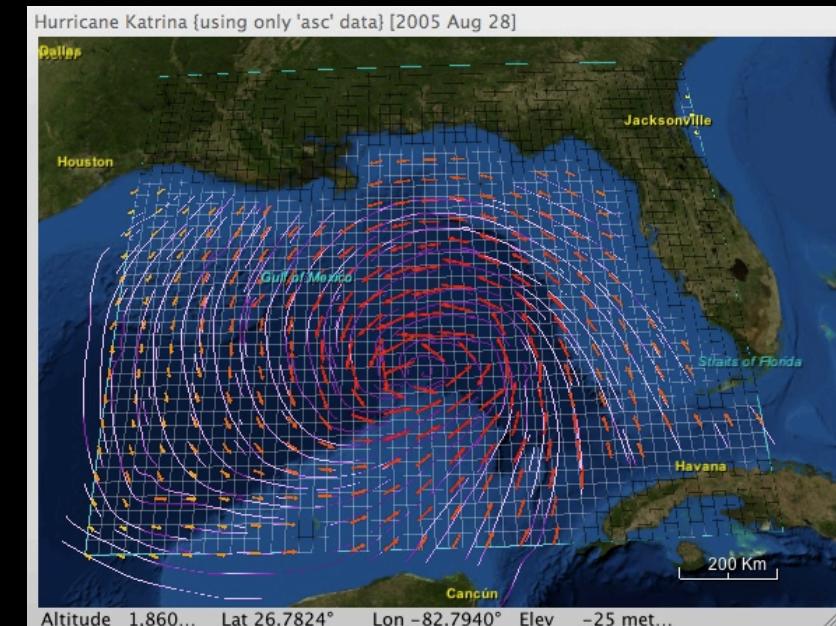
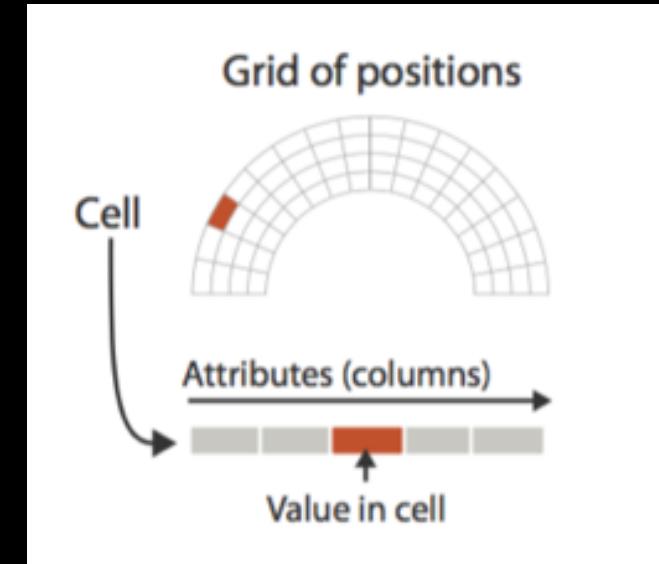
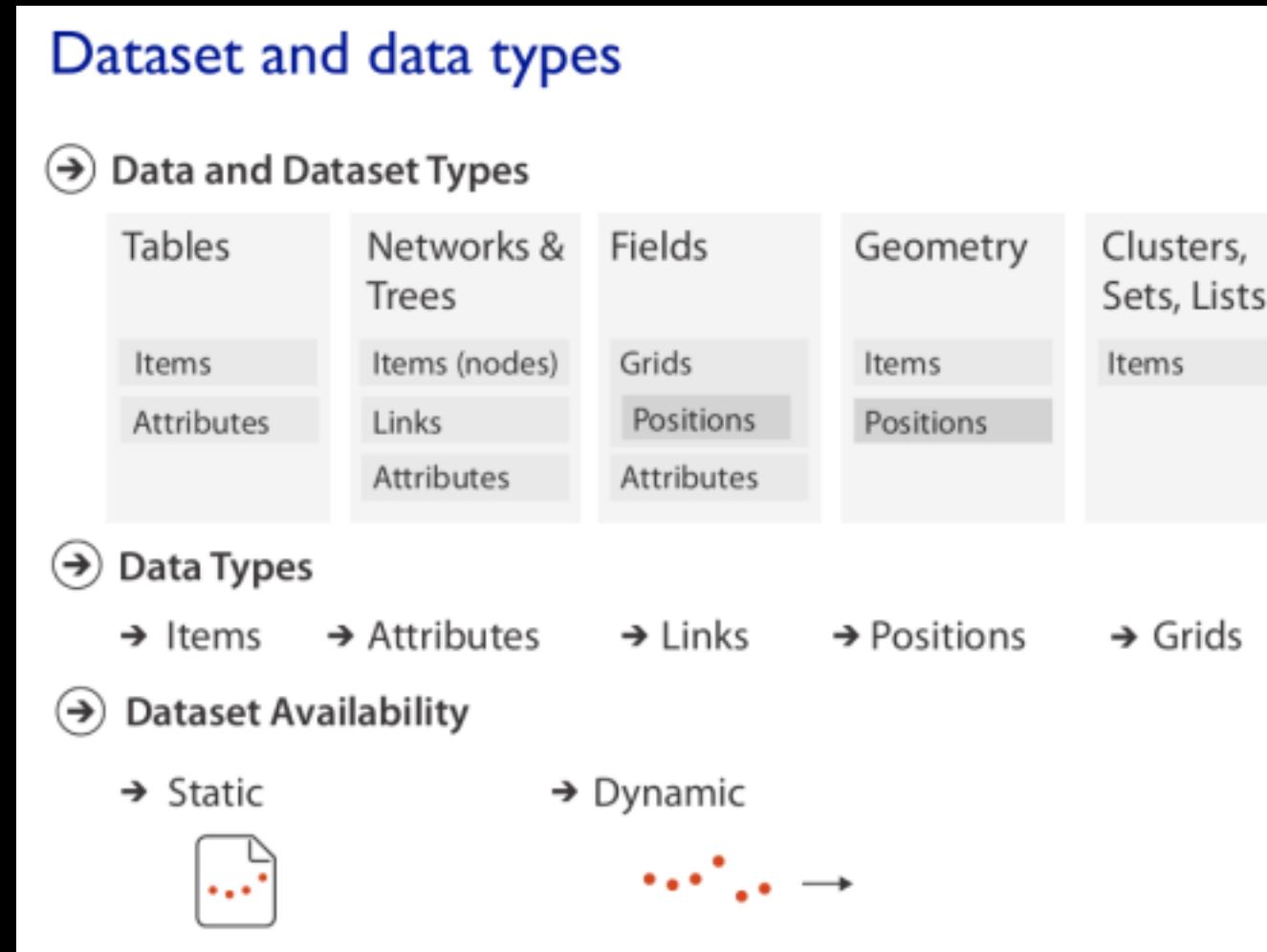
... →



The diagram illustrates a dataset structure as a grid. The vertical axis is labeled "Items (rows)" and the horizontal axis is labeled "Attributes (columns)". A single cell in the grid is highlighted in red and labeled "Cell containing value". Arrows point from the text labels to their corresponding grid axes.

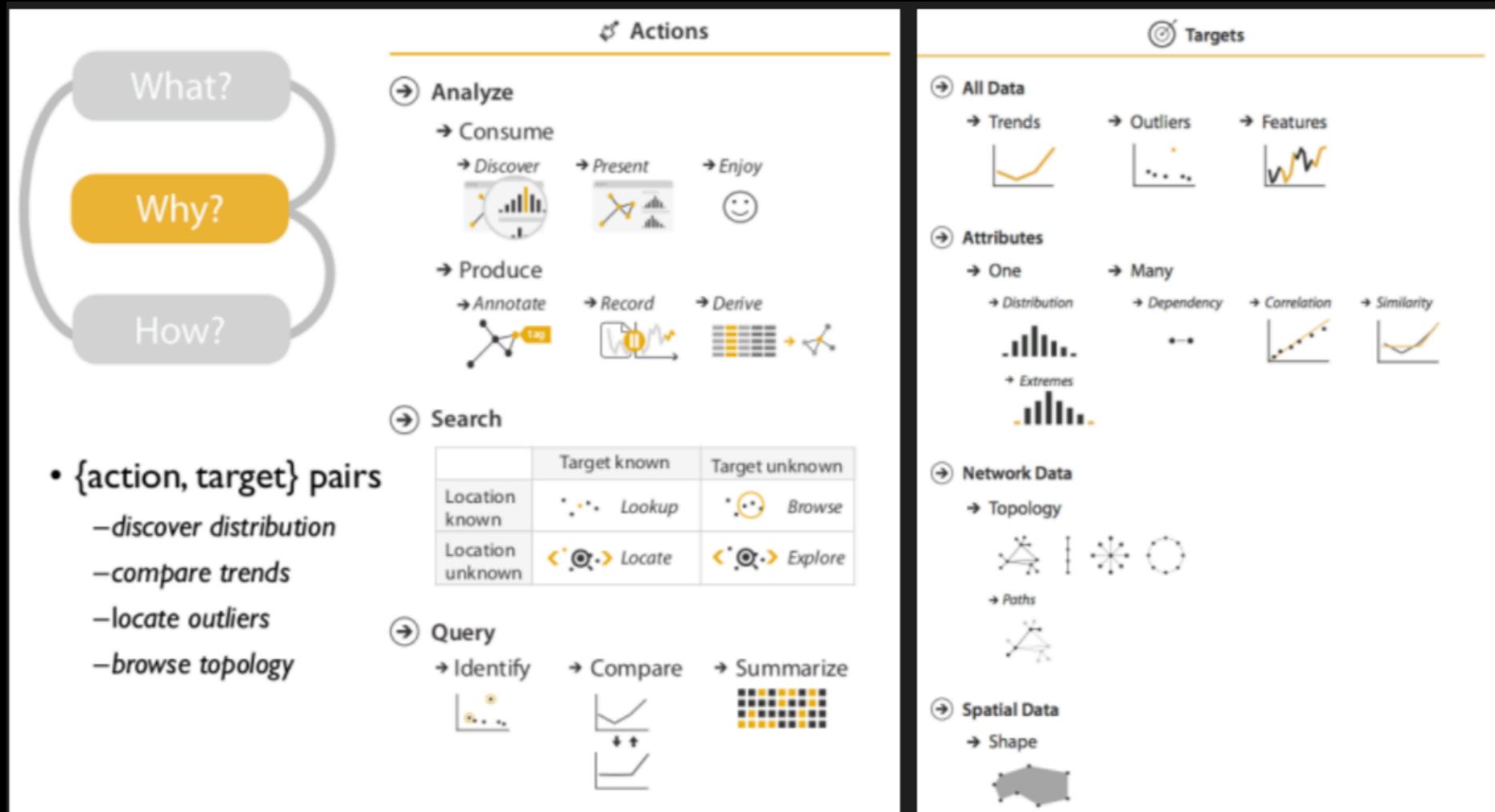
Framework de Tamara Munzner

- What: Tipos de Datos y de Datasets



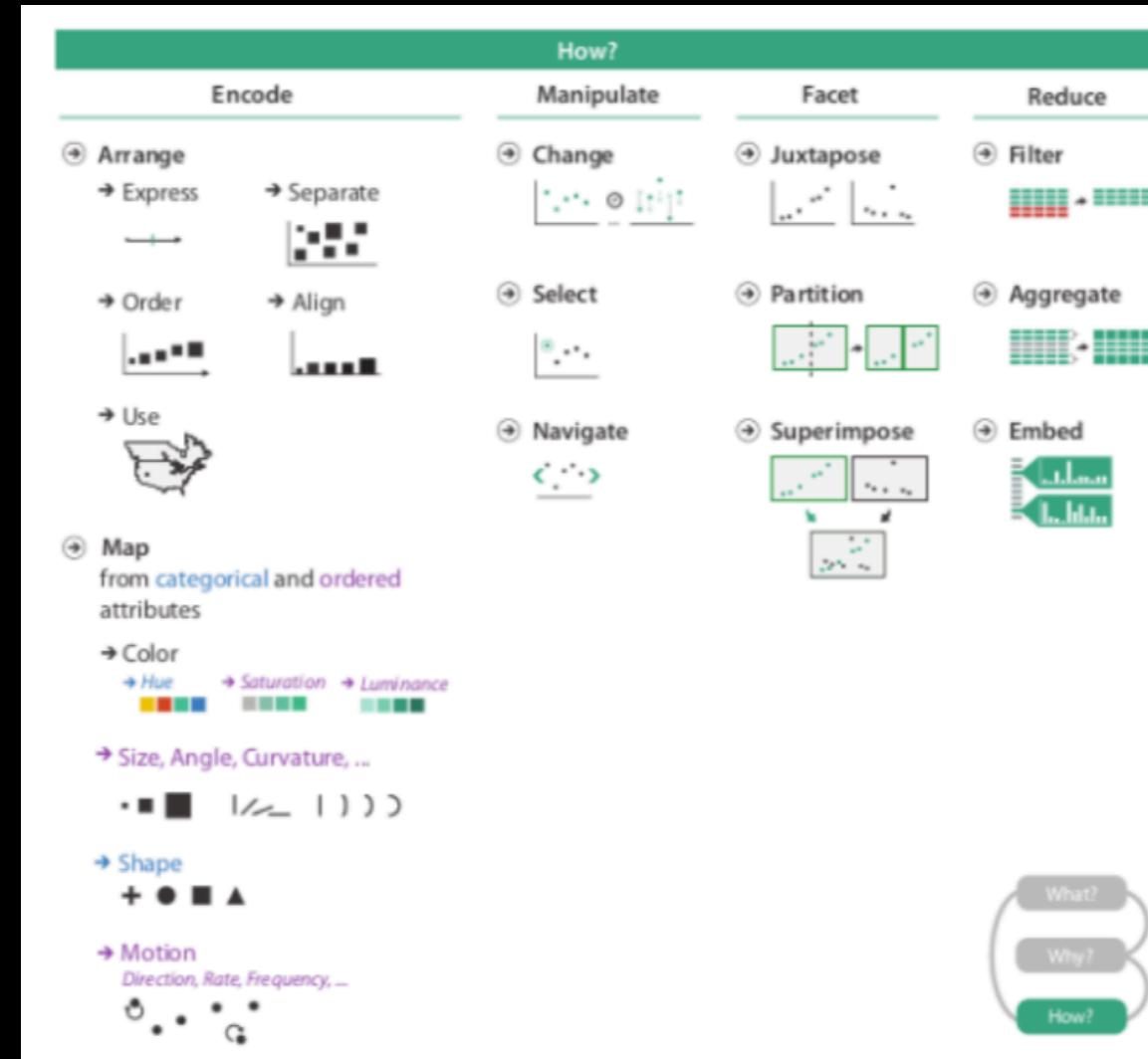
Framework de Tamara Munzner

- Why: tareas codificadas como {acciones, objetivos}



Conceptos del framework de Tamara Munzner

- How: Encoding Visual y de Interacción



Percepción y Psicología Cognitiva

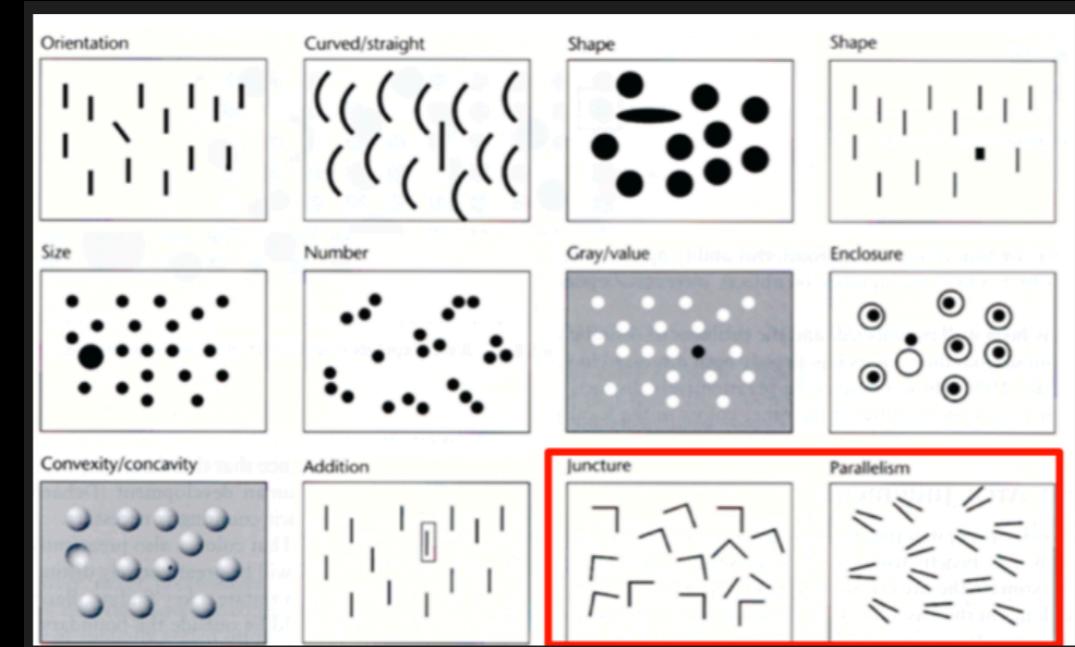
– Identificar aspectos de visión de bajo y de alto nivel

– Identificar concepto de Gestalt

– Modelo de estructura de memoria

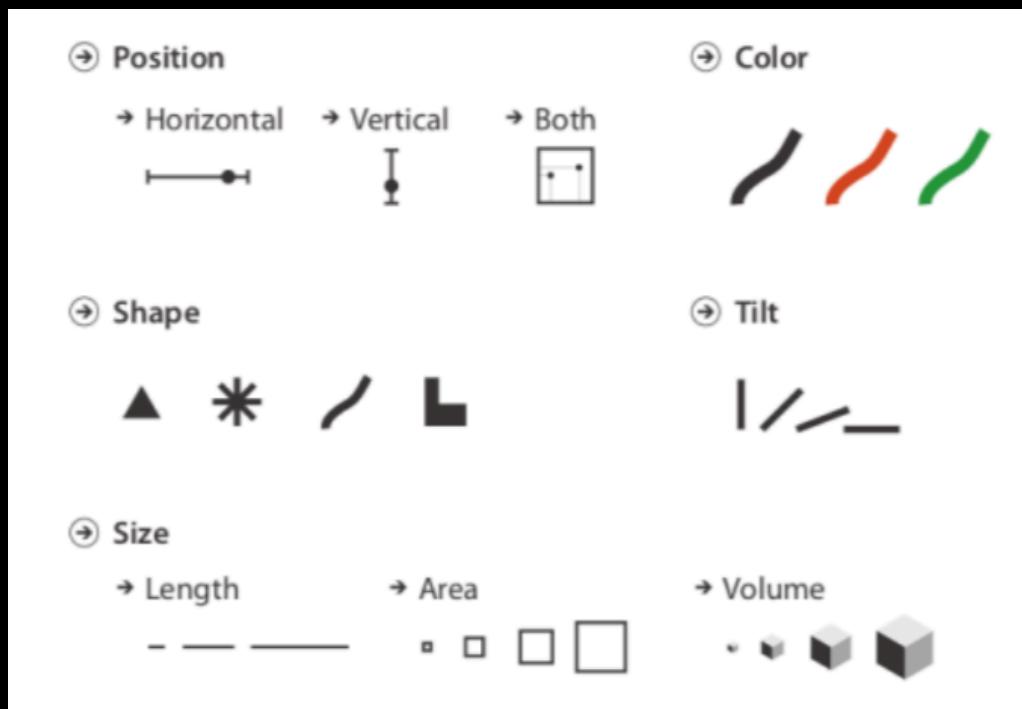
- Pre-atentiva (sensorial)
- De corto plazo
- De largo plazo

– Reconocer características que se procesan de forma pre-atentiva

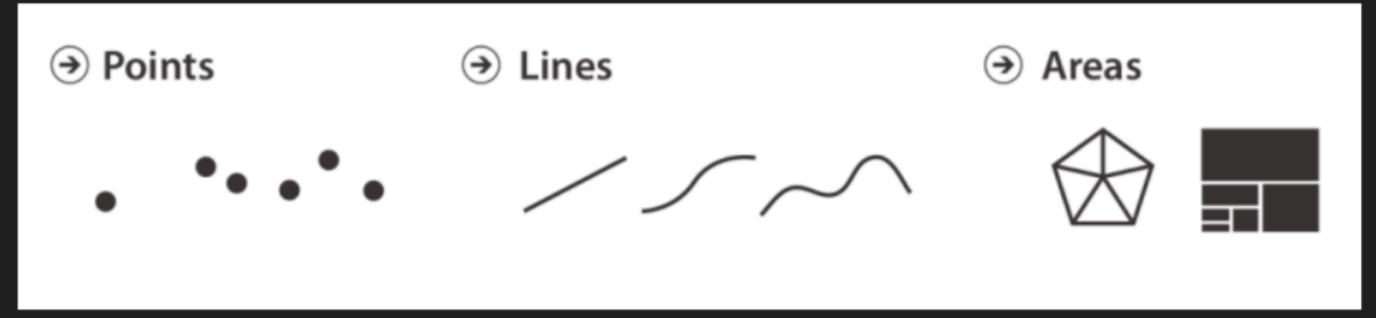


Marcas y Canales

- Definiciones
- Tipos
- Criterios para rankear su efectividad

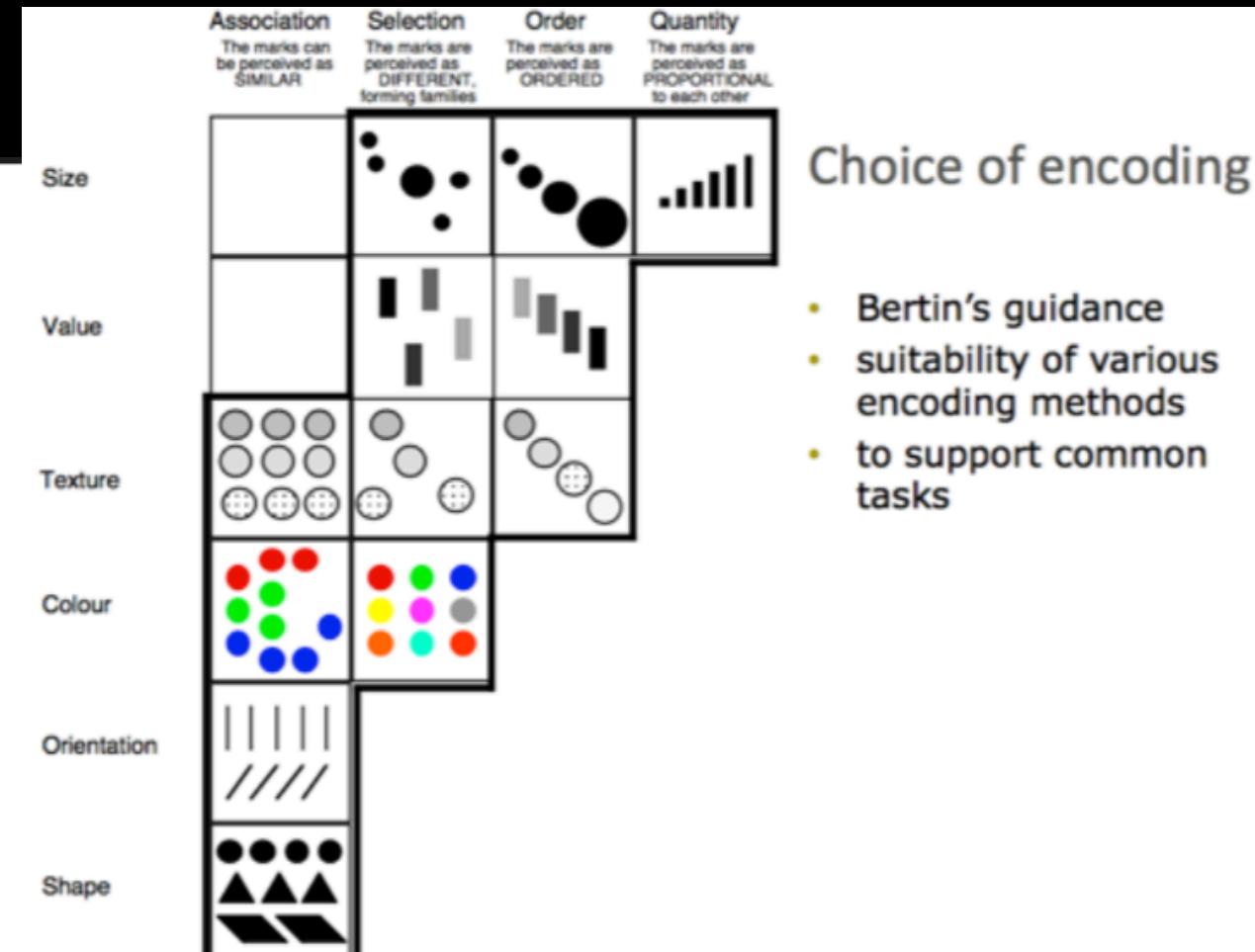
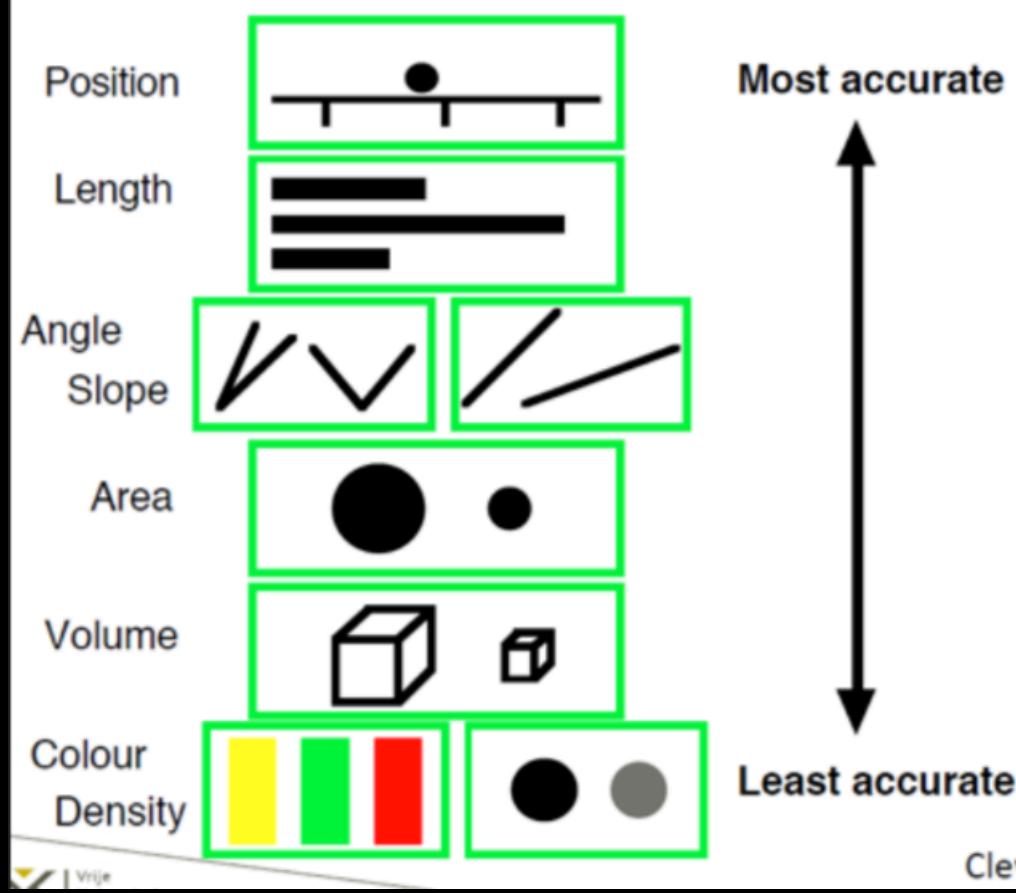


- Una **marca** es un elemento geométrico básico, que puede ser clasificado según el número de dimensiones espaciales que requiera.



Un **canal** visual permite controlar la **apariencia** de las marcas, independientemente de la dimensionalidad de este elemento primitivo.

Precisión de Canales y Guía de Bertin



Cleveland and McGill (1984)

Conceptos del framework de Tamara Munzner

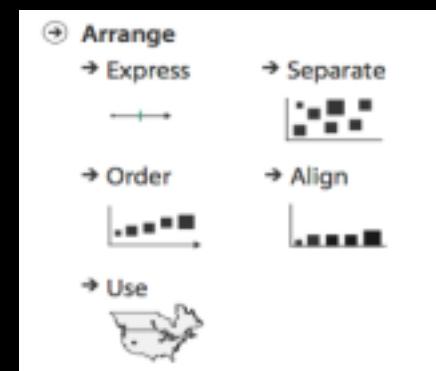
- How: Encoding Visual

Las cuatro familias de cómo implementar

- *Encode*
- *Manipulate*
- *Facet*
- *Reduce*

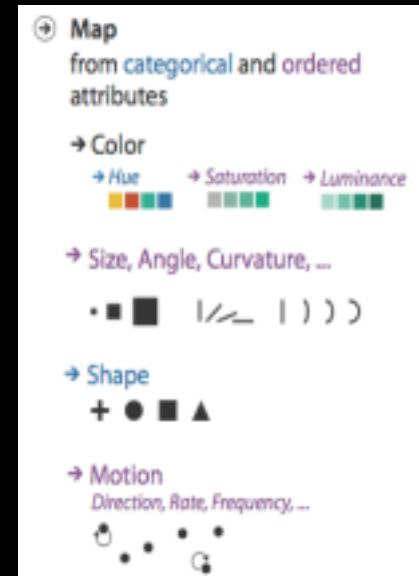
Encode I

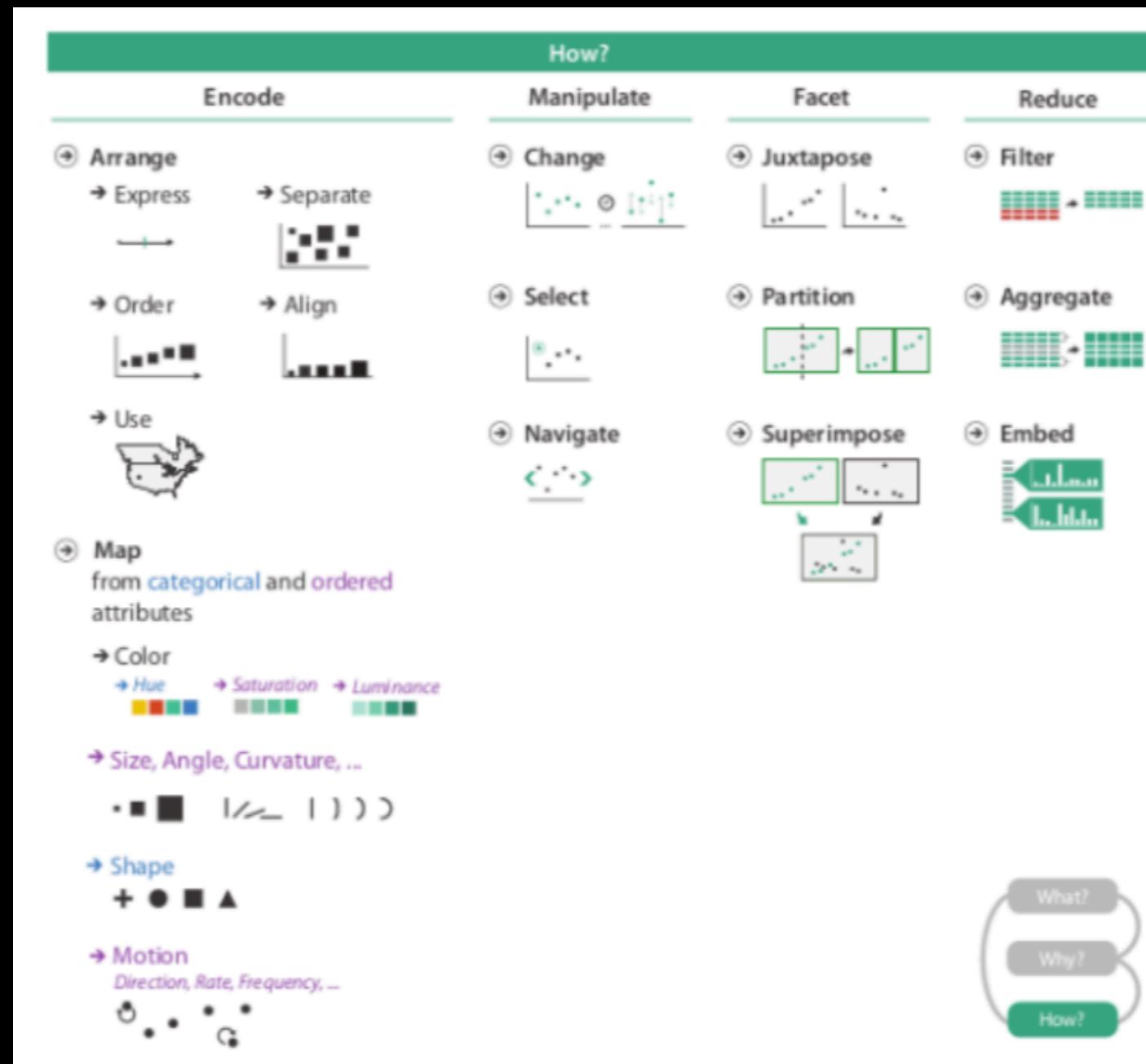
- Esta clase se divide en dos subfamilias: **arrange** y **map**.
- Con **arrange**, buscamos saber cómo organizar los **datos en el espacio**.
 - De todos los *encodings*, es el más crucial porque el uso del espacio domina el modelo mental que tiene el usuario de los datos.
 - Queremos saber cómo expresar los valores, cómo separar, ordenar y alinear las regiones, y cómo usar un espacio dado (e.g. *dataset geográfico*)



Encode II

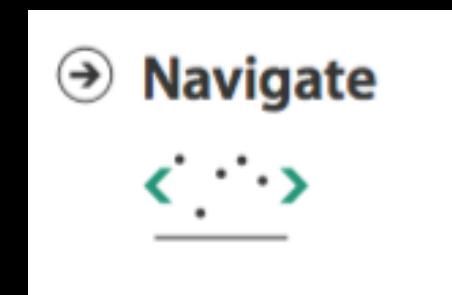
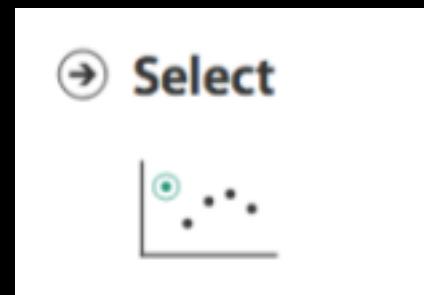
- Esta clase se divide en dos subfamilias: **arrange** y **map**.
- Con **map**, buscamos aprovechar los **canales visuales no-espaciales**.
 - Podemos trabajar con color (*hue, saturation, luminance*), tamaño, ángulo, curvatura, formas.
 - Pero también con atributos dinámicos: dirección, frecuencia, tasa de aparición.





Manipulate

- Esta clase tiene tres grandes operaciones: **change**, **select**, **navigate**.
- *Change* servirá para cambiar, por ejemplo, el *encoding*, la disposición de los datos, el nivel de agregación, etc.
- *Select* servirá para escoger los ítems o atributos a observar.
- *Navigate* servirá para cambiar el *viewpoint*. Hace sentido utilizar la metáfora de una cámara mirando una escena desde un punto de vista móvil
(e.g. *panning and zooming*)



Ejemplo: Zoom

Tipos de *zooming*

- Existen dos tipos de *zooming*: geométrico y semántico.
- La forma más intuitiva es la geométrica, ya que corresponde casi exactamente a nuestra experiencia de la vida real.

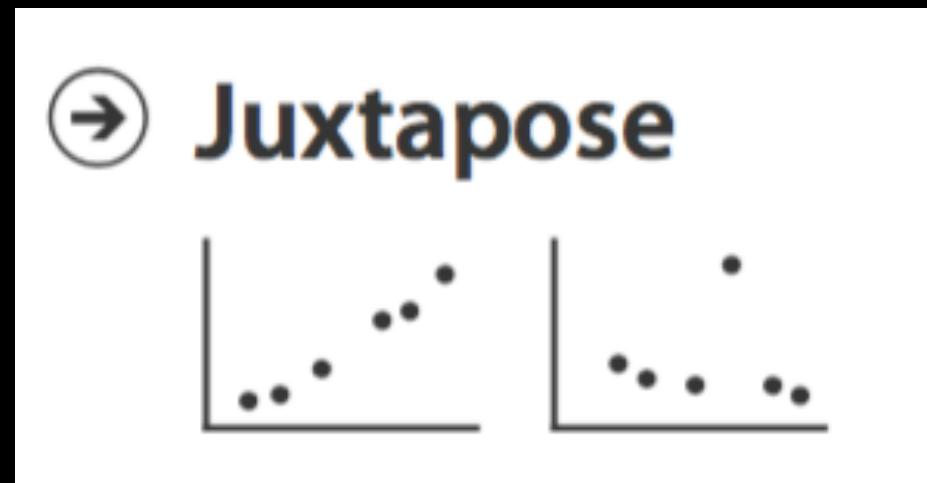


→ *Zoom*
Geometric or Semantic



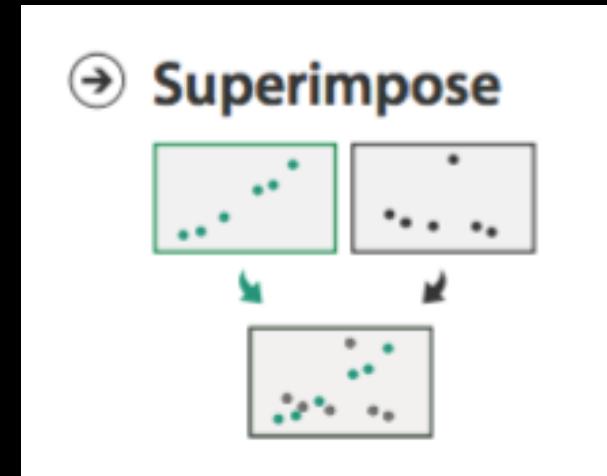
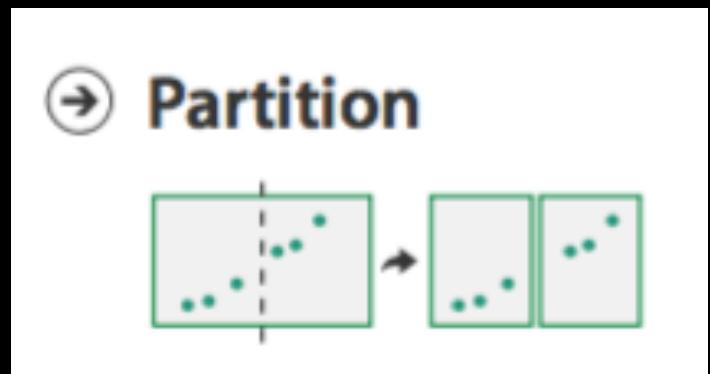
Facet I

- Esta clase tiene tres operaciones: **juxtapose**, **partition**, **superimpose**.
- La idea de *facet* es mostrar diferentes ángulos de un *dataset*, dividiendo la visualización en diferentes vistas.
- Por ejemplo, en **juxtapose**, se debe elegir cómo coordinar las vistas entre ellas, cuántos datos compartir, qué canales utilizar, etc.



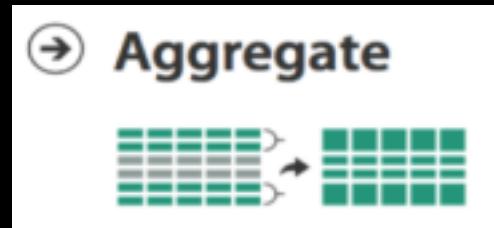
Facet II

- En ***partition***, se debe elegir cuántas regiones utilizar, cómo dividir los datos entre ellas, o el orden de los atributos al utilizar.
- En ***superimpose***, se debe elegir cómo los elementos serán particionados en las distintas capas, cuántas capas usar, etc.



Reduce

- Esta familia tiene por objetivo manejar la complejidad del *dataset*.
- Tenemos tres grandes operaciones:
 - Filtrar es la más simple, pues elimina la cantidad de elementos mostrados (e.g. por uno o más rangos de interés)
 - Agregar elementos busca que un grupo de elementos sea representado por un nuevo elemento que los represente; de esta forma, se hace un *merge* (e.g. obtener el promedio es el caso más simple)
 - Embeber permite reducir la cantidad de elementos mediante una sofisticada combinación de filtrado y agregación.



Ejemplo: Estimar dimensionalidad correcta

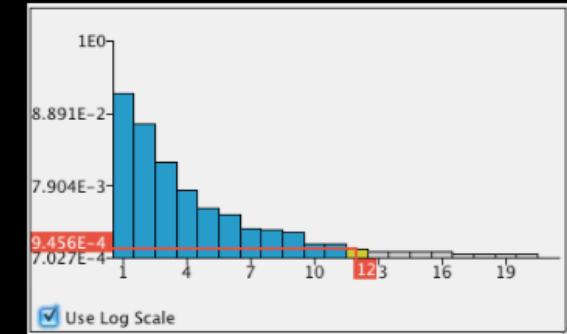
Estimating true dimensionality

- how do you know when you would benefit from DR?
 - consider error for low-dim projection vs high-dim projection
- no single correct answer; many metrics proposed
 - cumulative variance that is not accounted for
 - strain: match variations in distance (vs actual distance values)
 - stress: difference between interpoint distances in high and low dims

$$\text{stress}(D, \Delta) = \sqrt{\frac{\sum_{ij} (d_{ij} - \delta_{ij})^2}{\sum_{ij} \delta_{ij}^2}}$$

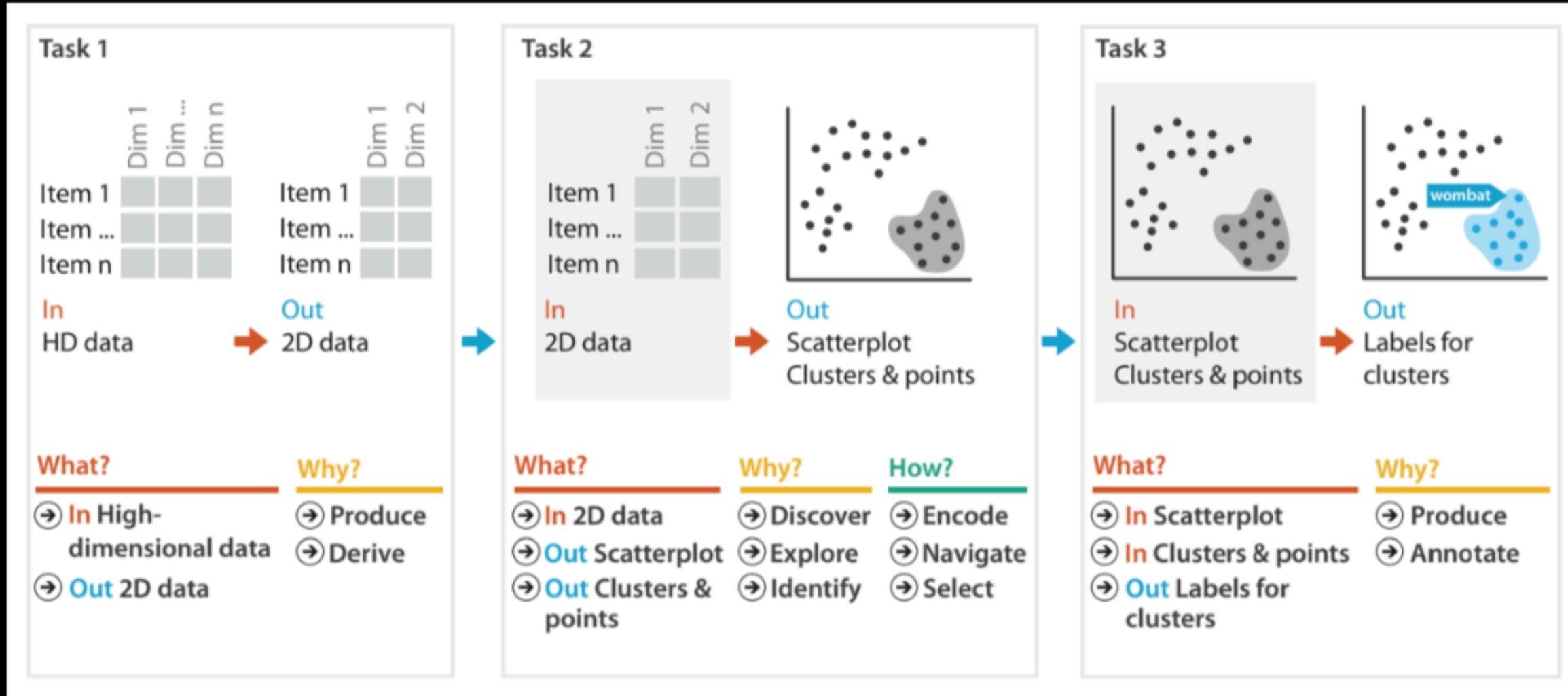
- D : matrix of lowD distances
- Δ : matrix of hiD distances δ_{ij}

- scree plots as simple way: error against # attrs



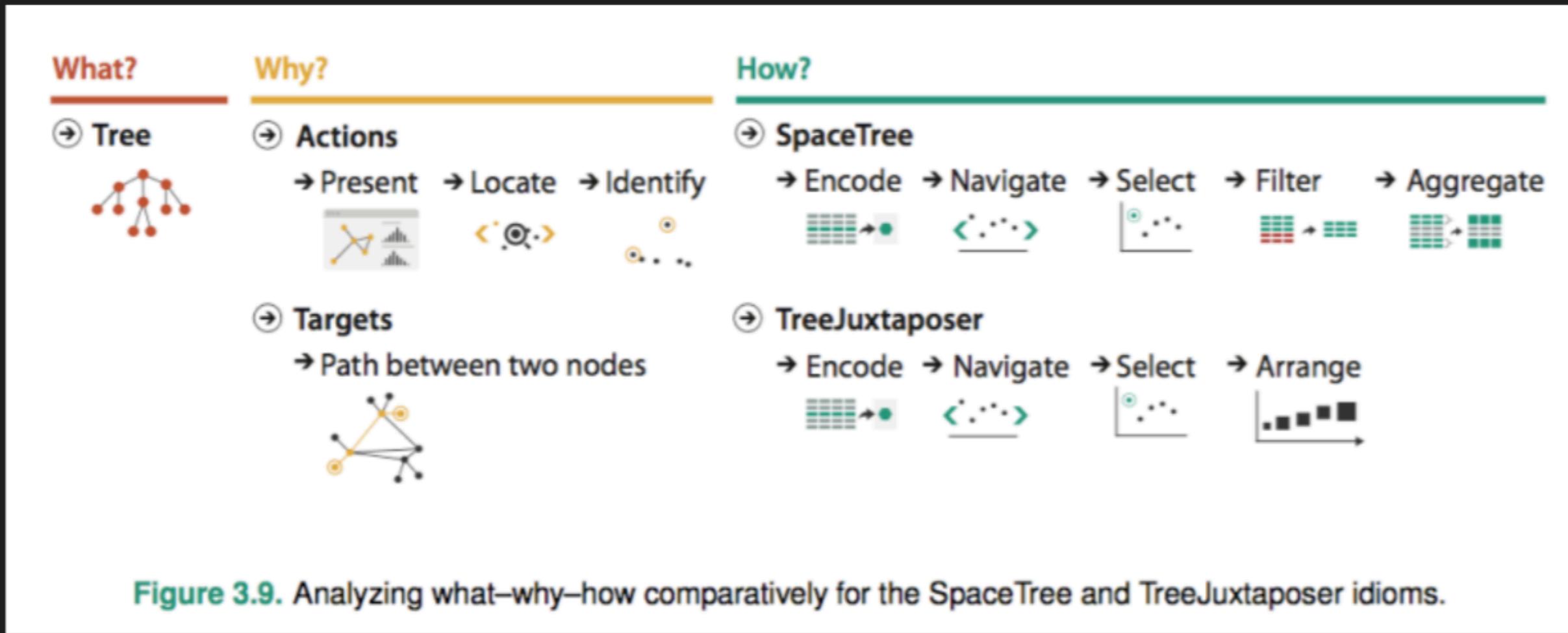
- original dataset: 294 dims
- estimate: almost all variance preserved with < 20 dims

Uso del framework



Uso del framework

Comparación de dos "idioms"



Uso del framework

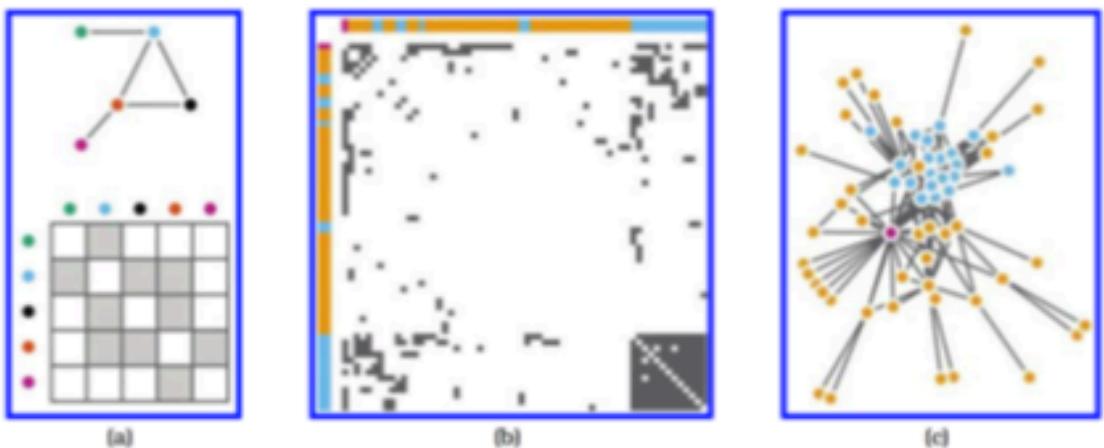
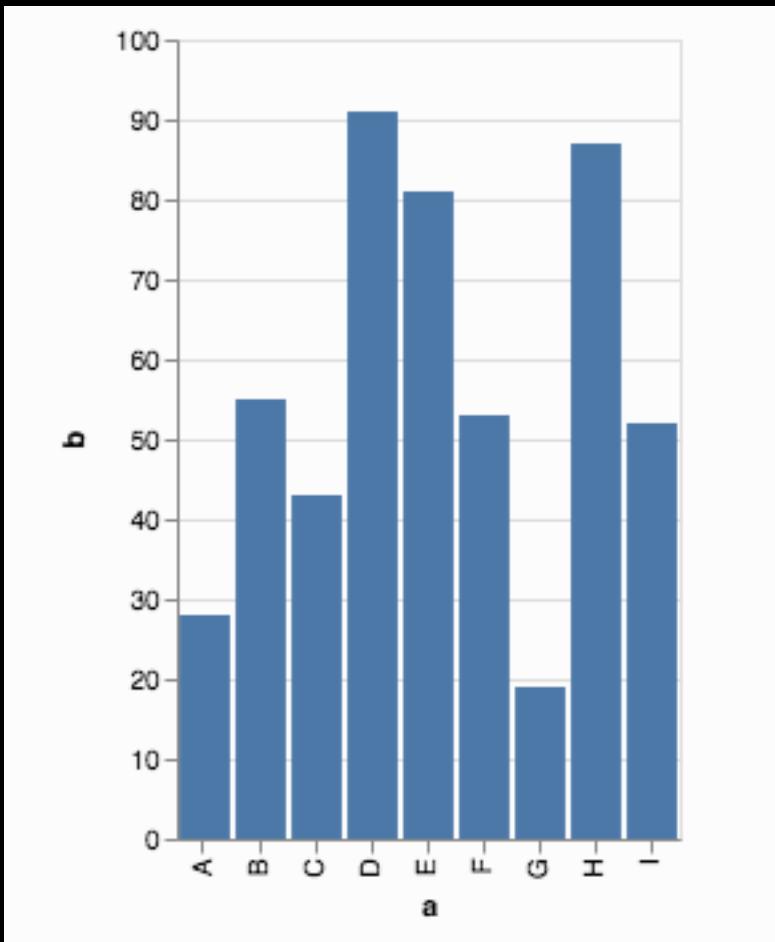


Figure 9.6. Comparing node-link matrix and matrix views of a network. (a) Node-link and matrix views of small network. (b) Matrix view of larger network. (c) Node-link view of larger network. From [Gehlenborg and Wong 12, Figures 1 and 2].

Idiom	Adjacency Matrix View
What: Data	Network.
What: Derived	Table: network nodes as keys, link status between two nodes as values.
How: Encode	Area marks in 2D matrix alignment.
Scale	Nodes: 1000. Links: one million.

Casos de Uso

- Herramientas: Vega, Vega-lite, Altair



```
import altair as alt
import pandas as pd

data = pd.DataFrame({
    'a': ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I'],
    'b': [28, 55, 43, 91, 81, 53, 19, 87, 52]
})

alt.Chart(data).mark_bar().encode(
    x='a',
    y='b'
)
```

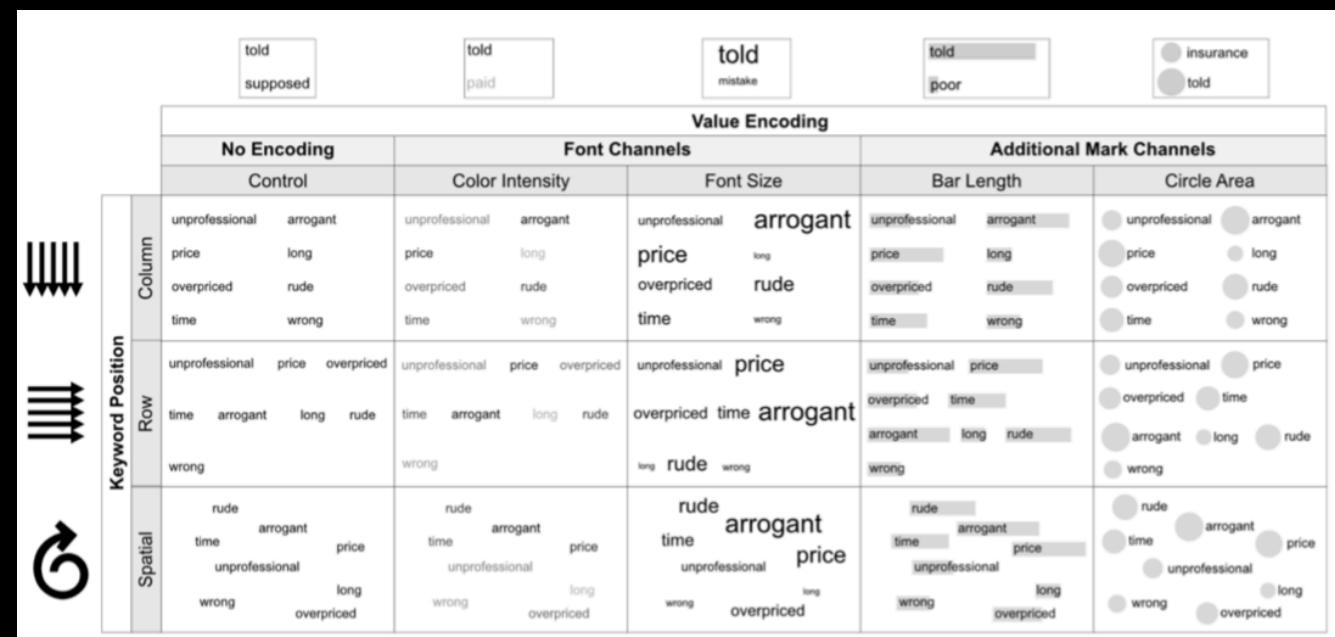
Aplicaciones

- Ejemplo: Visualización de texto
- Cuál es el mejor diseño para wordclouds?

Tareas

- **Comparar Valores:** Comparar magnitudes de dos palabras seleccionadas
- **Buscar Palabras:** Buscar por una palabra específica en el cloud
- **Identificar Tópicos:** Elegir a qué tópico corresponde una palabra (de una lista)
- **Construir Tópicos:** Describir qué tópicos hay incluidos en una word cloud.

Felix, C., Franconeri, S., & Bertini, E. (2017).
Taking Word Clouds Apart: An Empirical Investigation of the Design
Space for Keyword Summaries.
IEEE Transactions on Visualization and Computer Graphics.

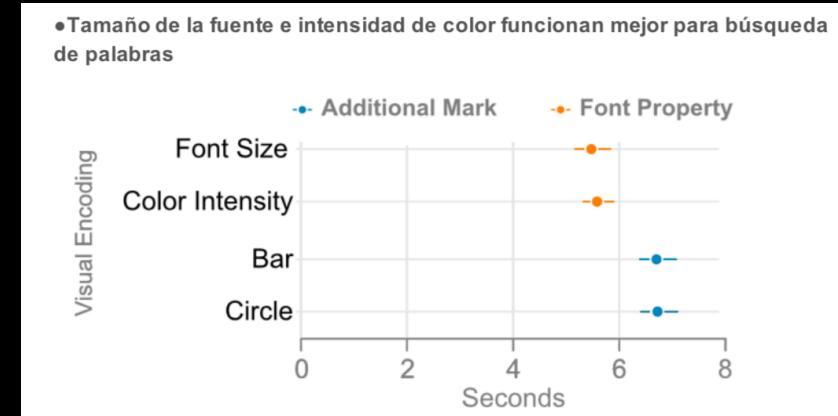
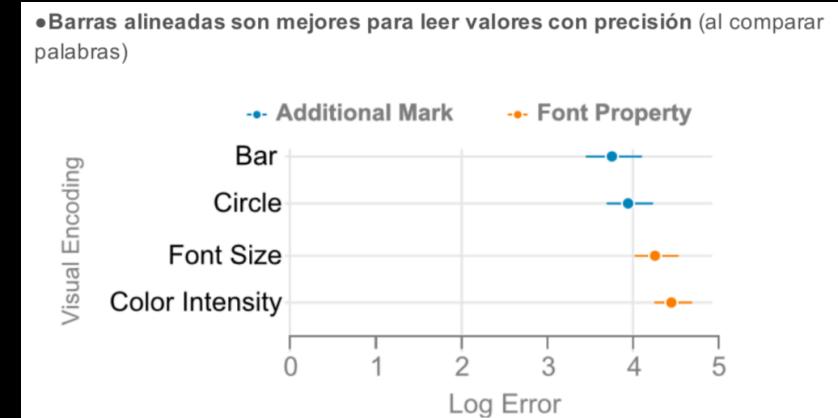


Aplicaciones

- Ejemplo: Visualización de texto
- Cuál es el mejor diseño para wordclouds?

WORD CLOUDS

- Experimentar con diseños diferentes.
- Give column bars a chance. Agregar barras a las palabras, alinear en columnas y permitir al usuarios cierta interacción.
- Probar otras disposiciones espaciales. Manteniendo tamaño de fuente e intensidad de color parece funcionar bien para algunas tareas, pero se puede experimentar con otros *layout*.
- Listas simples de palabras funcionan bien. No se ven muy fashion, pero especialmente en tareas como encontrar tópicos o temas, parecen funcionar bien.
- Pensar en ordenar. En una aplicación interactiva, puede ser buena idea agregar funcionalidad para ordenar alfabéticamente o por tamaño. Esto puede ayudar a responder diferentes preguntas en un ambiente de análisis de datos.



Charlas invitadas

- ¿Qué encodings visuales y de interacción se usaron en GALEAN, la herramienta presentada por Vanesa Peña ?
- ¿Qué encodings visuales y de interacción se usaron en la visualización de emociones a través de emoticones que presentó Eduardo Graells?

IEEE Vis 2018

- Best Paper: Draco
- Visualización espacial de histogramas multiclasé
- Redes neuronales profundas como modelo de percepción visual humana

Tendencias Generales en Investigación de InfoVis

keyword	topic (2000–2015)						keyword	PCS (2008–2015)							
	chart	#	slope	SE	df	p-val.	t-val.	chart	#	slope	SE	df	p-val.	t-val.	
interaction techniques—general		175	1.12	0.15	14	<.001	7.503	quantitative evaluation		266	4.40	0.73	6	.001	7.457
evaluation general		128	0.81	0.20	13	.001	4.068	visual knowledge discovery		416	4.33	1.07	6	.007	6.296
machine learning and statistics		85	0.73	0.11	14	<.001	6.881	time series data		276	4.31	0.76	6	.001	6.190
timeseries, time-varying data and techniques		109	0.60	0.18	14	.005	3.294	geographic/geospatial vis.		269	4.20	1.05	6	.007	4.000
multidim./multivar./multifield data and techn.		109	0.57	0.12	14	<.001	4.555	coordinated & multiple views		334	4.17	1.47	6	.030	2.962
analysis process—general		110	0.47	0.18	12	.024	2.580	data transf. and repres.		267	3.75	1.20	6	.020	3.354
graph/network data and techniques		134	0.46	0.15	14	.007	3.169	interaction design		246	1.45	1.21	6	.275	1.951
visual encoding and layout—general		78	0.32	0.09	13	.004	3.455	graph/network data		373	0.89	1.03	6	.419	1.467
data clustering and aggregation		83	0.21	0.08	12	.030	2.465	multidimensional data		255	0.75	0.32	6	.056	2.633
visualization techniques and tools—general		82	0.18	0.10	13	.085	1.853	user interfaces		270	0.60	1.93	6	.768	1.983
biomedical science and medicine		122	-0.02	0.13	14	.855	-0.187	high-dimensional data		236	0.45	1.61	6	.788	1.572
flow visualization, data, and techniques		114	-0.11	0.16	14	.488	-0.713	biomedical and medical vis.		284	0.07	1.07	6	.949	0.204
numerical methods / mathematics		94	-0.42	0.15	14	.016	-2.752	vis. system and toolkit design		262	-0.31	1.78	6	.868	-0.174
meshes, grids, and lattices		86	-0.68	0.10	14	<.001	-6.869	time-varying data		263	-0.77	0.71	6	.315	0.486
volume rendering, modeling, and vis.		232	-0.92	0.19	14	<.001	-4.776	volume rendering		278	-2.52	1.33	6	.107	-1.870

Herramienta: keyvis.org

 Getting started Search [Topics](#) About

Search for VIS paper keywords

Explore all topic clusters:

Abstraction, Simplification, Approximation

Topic cluster containing **24** keywords :

simplification 14x mesh simplification 9x edge bundling 4x quadric error metrics 4x model simplification 3x
visual abstraction 2x data abstraction 1x data simplification 1x denoising 1x density reduction 1x edge clustering 1x
edge congestion 1x edge contraction 1x generalization 1x graph simplification 1x hierarchical approximation 1x
high-dimensional approximation 1x multiphase simplification 1x polygonal surface simplification 1x schematisation 1x
sparse approximation 1x thinning 1x triangle decimation 1x visual abstraction language 1x

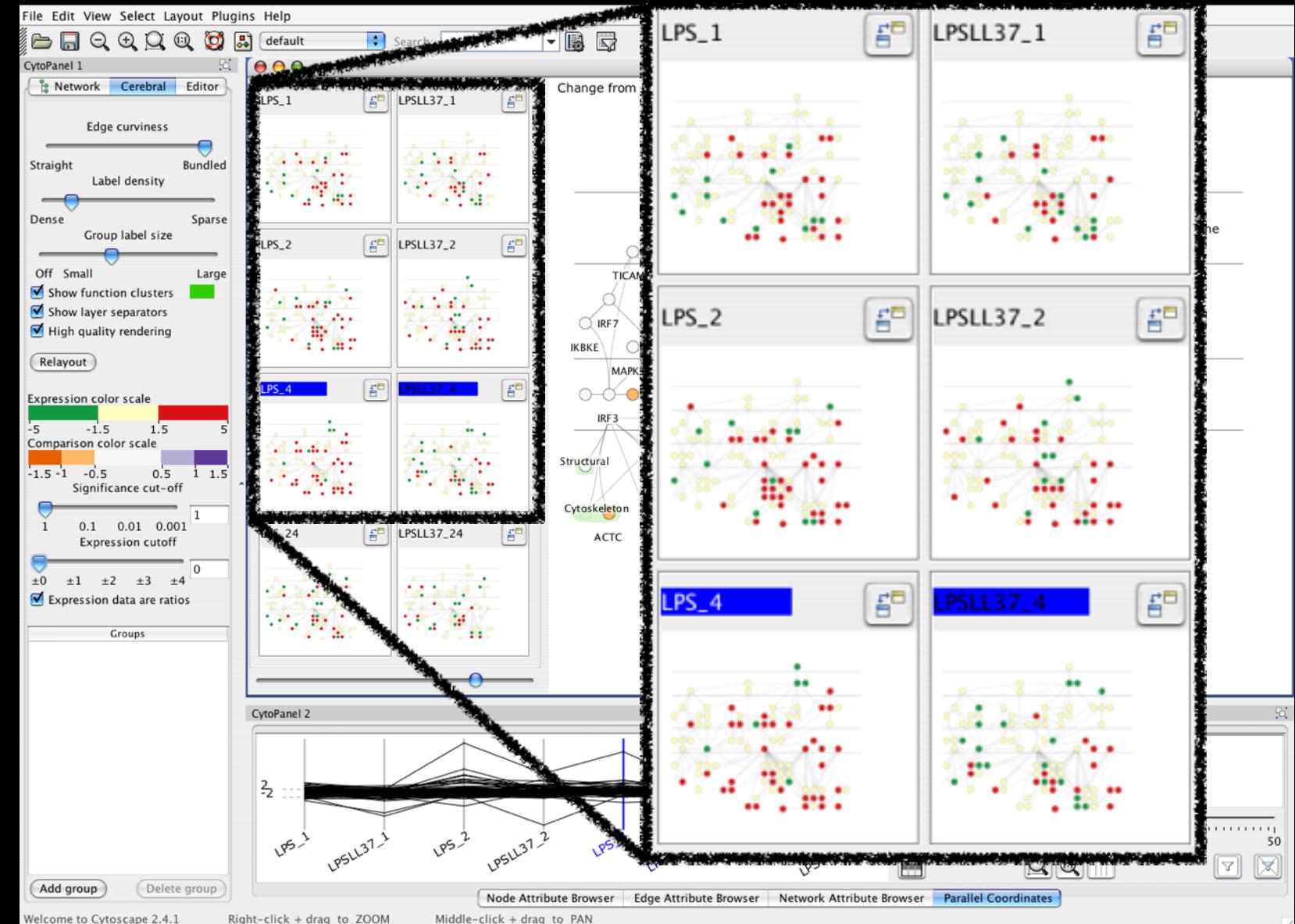
All papers that include at least one of these keywords:

Conf.	Year ▲	Title
SciVis	2015	Mining Graphs for Understanding Time-Varying Volumetric Data

Idiom: Small multiples

- encoding: same
- data: none shared
 - different attributes for node colors
 - (same network layout)
- navigation: shared

System:



Small multiples

- http://projects.flowingdata.com/tut/linked_small_multiples_demo/

