

150
ING

la Argentina celebra
su ingeniería
1870-2020

II Simposio de Informática INDUSTRIA 4.0



08 al 16 DE JUNIO

2020

Visual Analytics para Análisis Exploratorio de Datos

Disertante:

Duván Alberto Gómez Betancur

Institución:

Tecnológico de Antioquia I.U.



Attribution-NonCommercial-
NoDerivatives 4.0 International
(CC BY-NC-ND 4.0)



Sociedad de la Información

2020 *This Is What Happens In An Internet Minute*





Datos

- Estructurados



Size	#bedrooms	...	Price (1000\$)
2104	3		400
1600	3		330
2400	3		369
⋮	⋮		⋮
3000	4		540

- No Estructurados:



Audio



Image



¿Cómo darle sentido?

- Análisis estadístico
- Aprendizaje automático
- Análisis visual



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



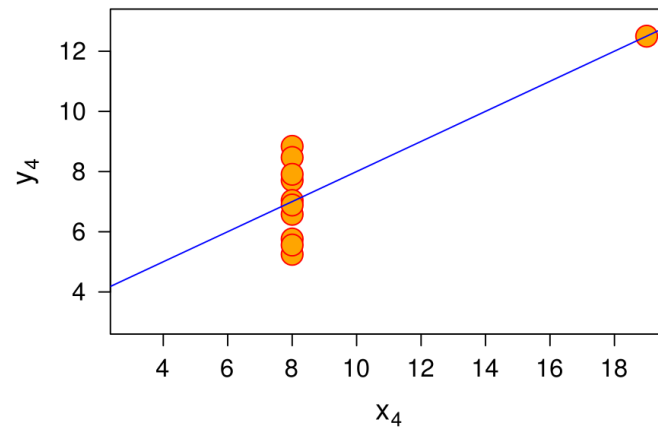
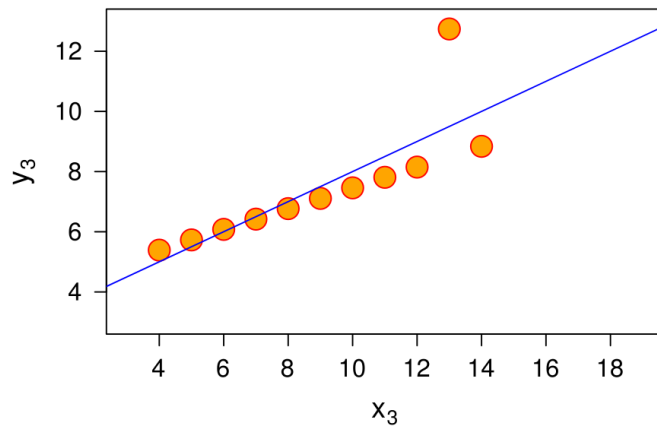
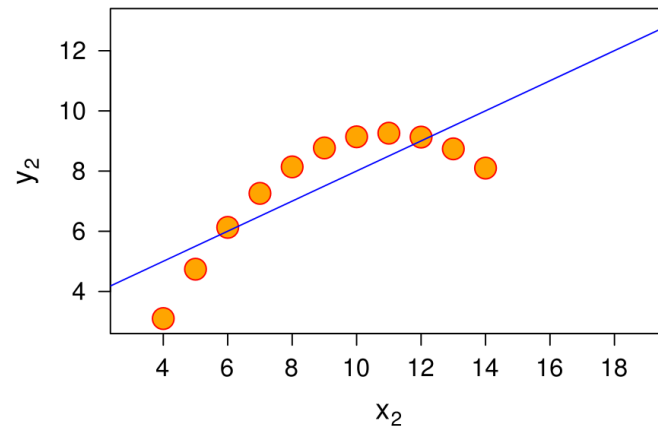
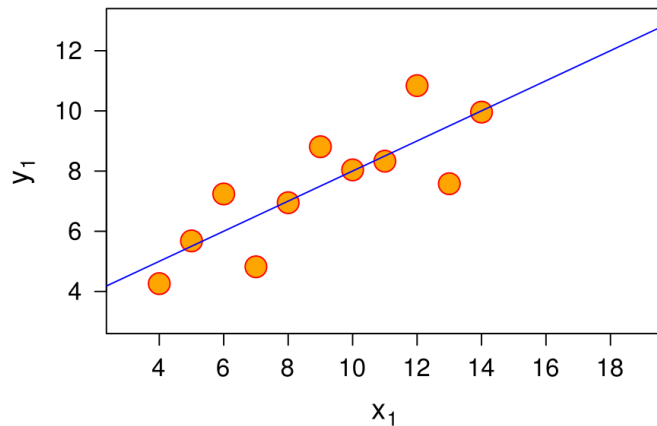
Cuarteto de Anscombe

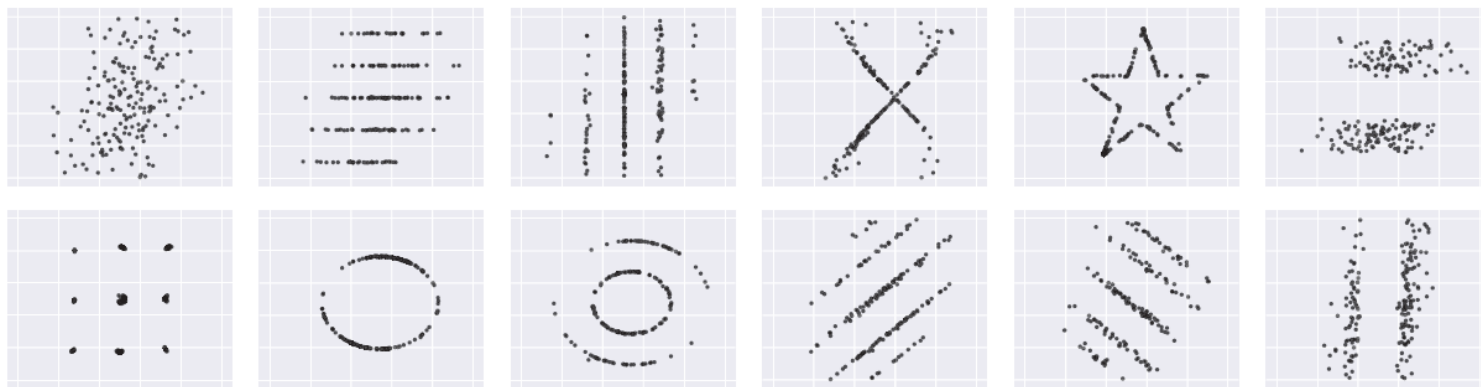
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

Property	Value
Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.125
Correlation between x and y	0.816
Linear regression	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression	0.67



Cuarteto de Anscombe

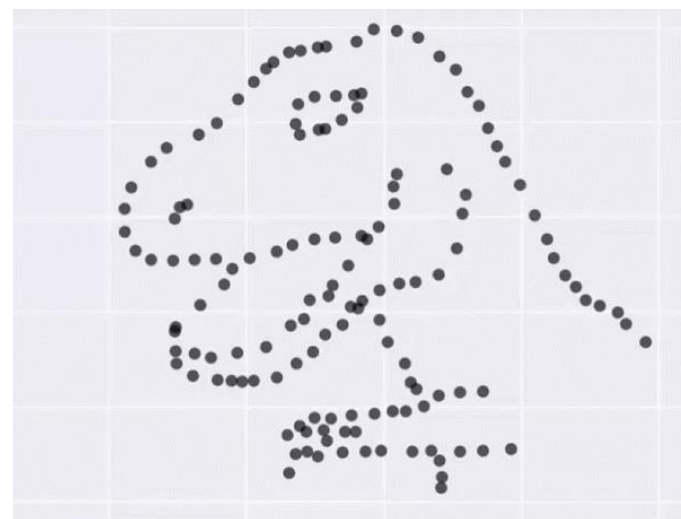
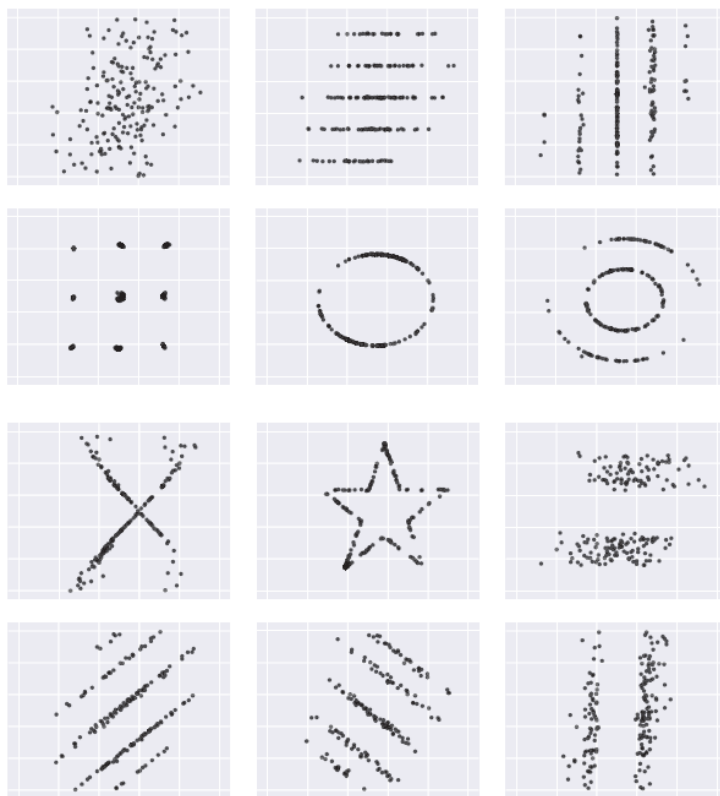




<https://dabblingwithdata.wordpress.com/2017/05/03/the-datasaurus-a-monstrous-anscombe-for-the-21st-century/>



Datasaurus

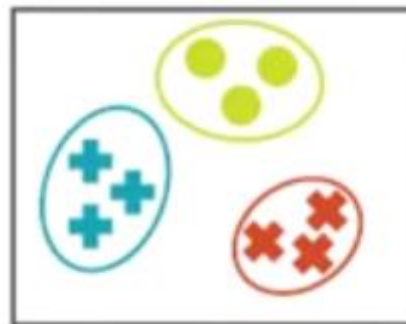
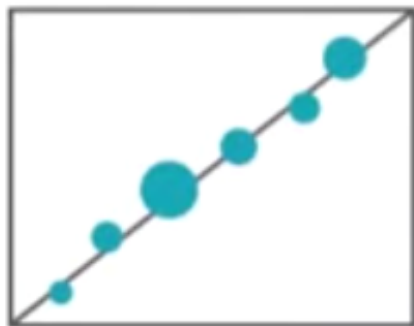


<https://dabblingwithdata.wordpress.com/2017/05/03/the-datasaurus-a-monstrous-anscombe-for-the-21st-century/>



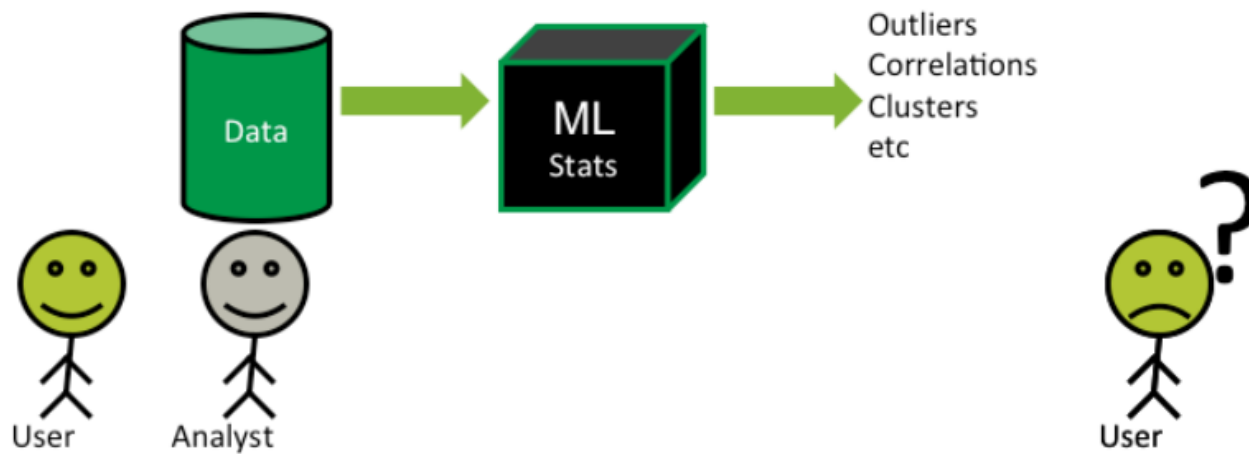
Data Mining

- Algoritmos para la obtención de reglas y modelos generales a partir de datos, para predecir comportamiento de datos nuevos.
- Descubrimiento automático de regularidades (patrones) en los datos mediante algoritmos





Data Mining / Machine Learning



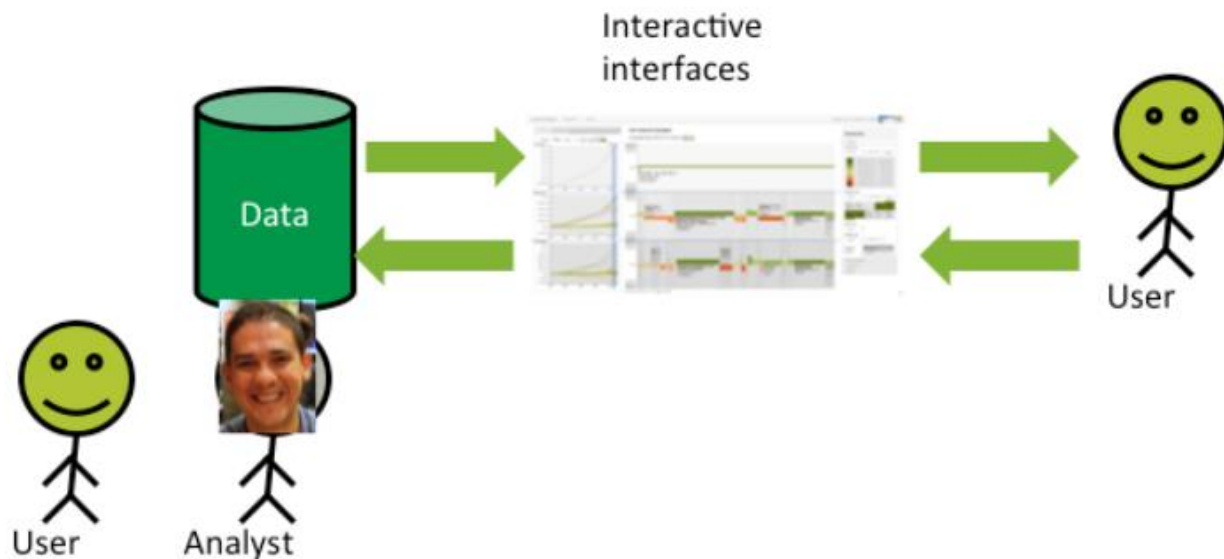


¿Cómo darle sentido?

- Análisis estadístico
- Aprendizaje automático / Data Mining
- **Análisis visual**



Information Visualization



John Alexis Guerra Gomez @guerravis

Tomado de www.johnguerra.co



Visual Analytics



Tomado de www.johnguerra.co

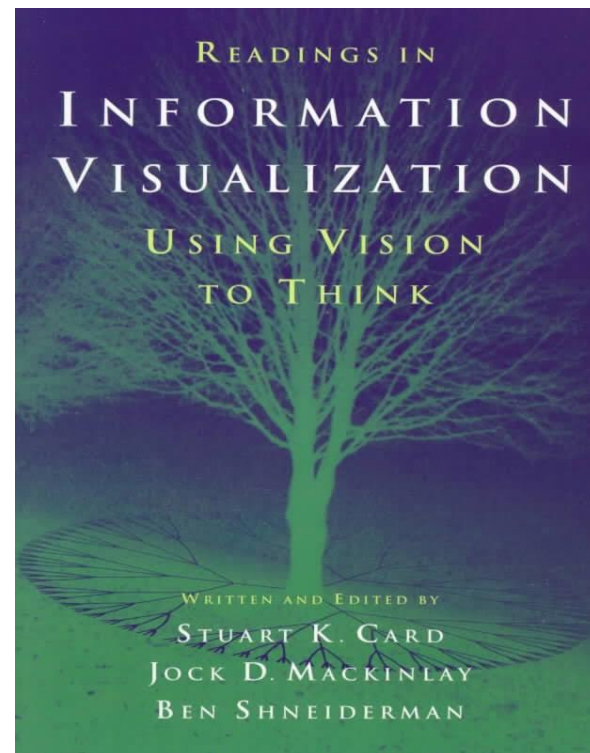


Visualización de Información

- “The use of **computer-supported, interactive, visual representations of abstract data to amplify cognition**”

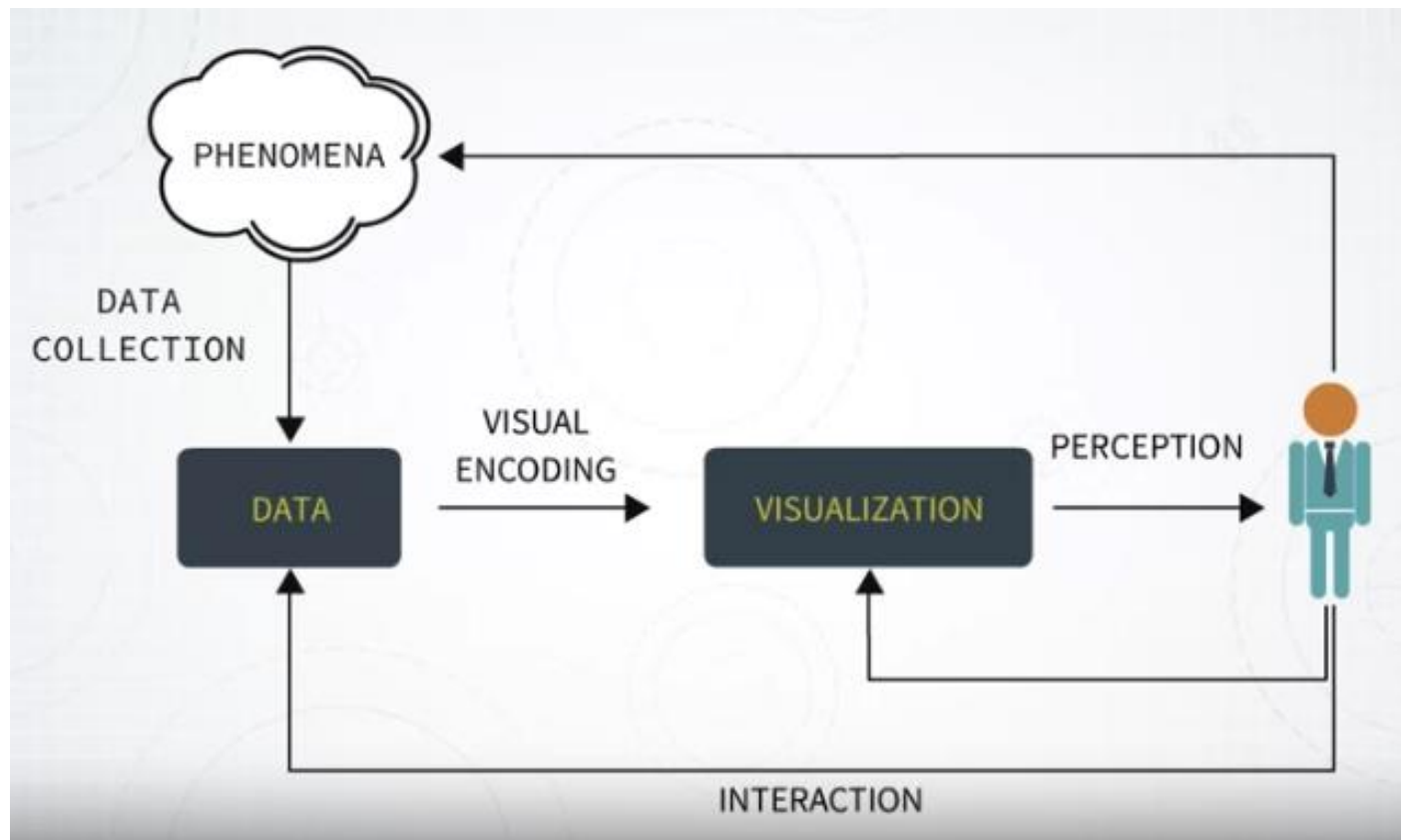
Sistemas basados en computador que permiten crear representaciones visuales de un conjunto de **datos**.

Para ayudar a las **personas** a realizar una **tarea** de manera más efectiva.





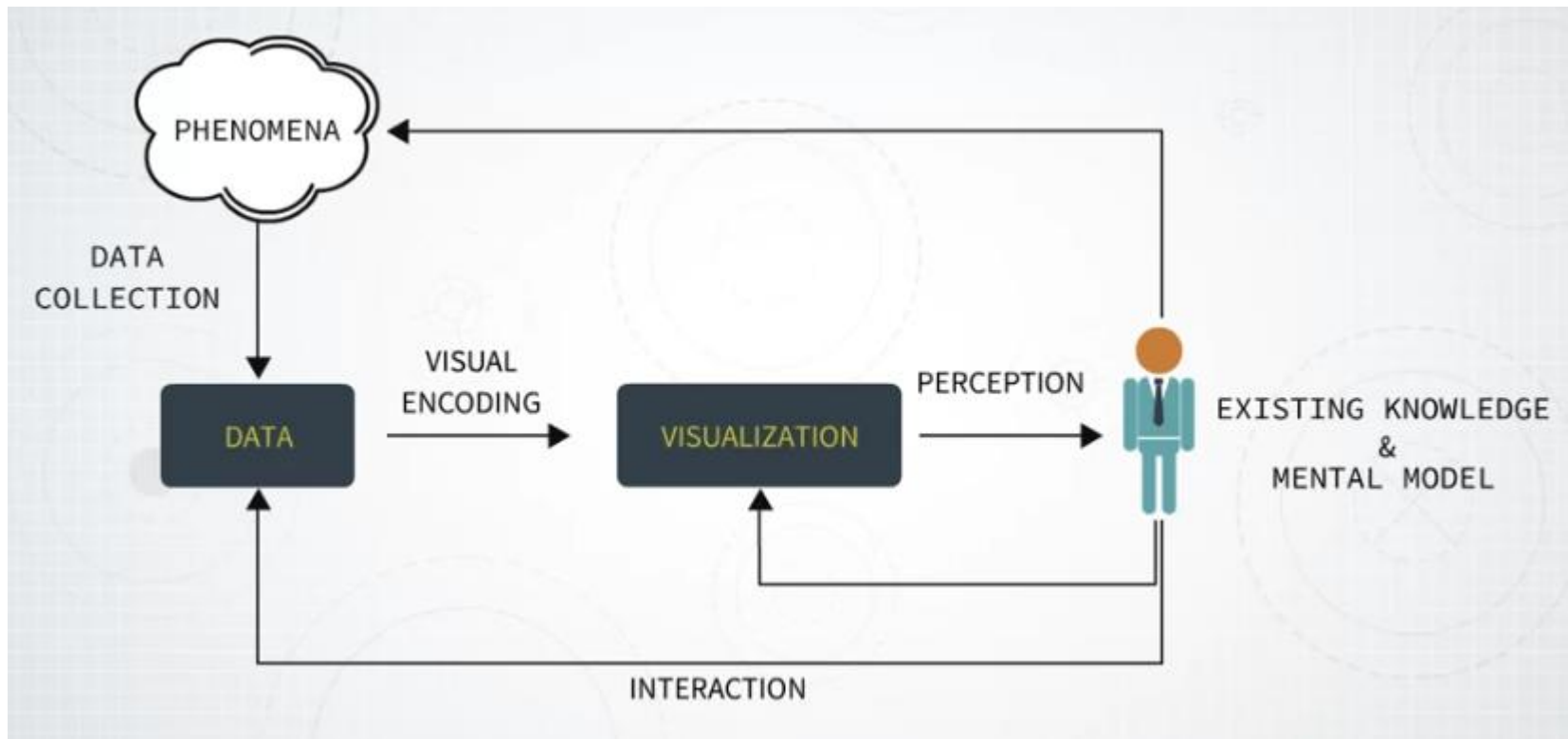
Pipeline de Visualización



(Information Visualization – Coursera, Enrico Bertini)



Pipeline de Visualización

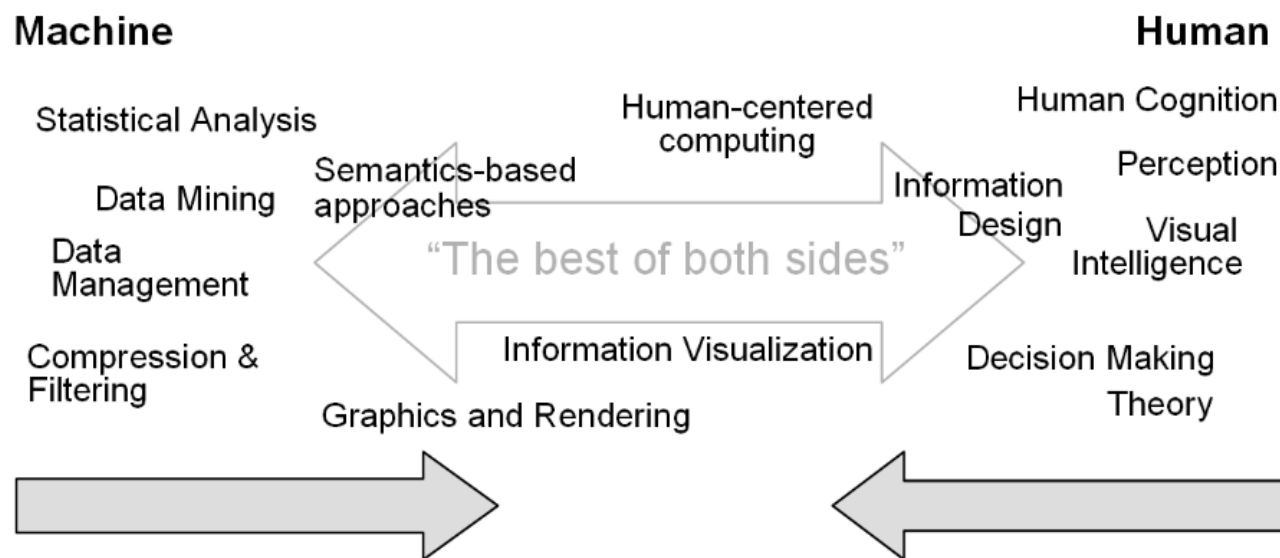


(Information Visualization – Coursera, Enrico Bertini)



El Humano en el “loop”

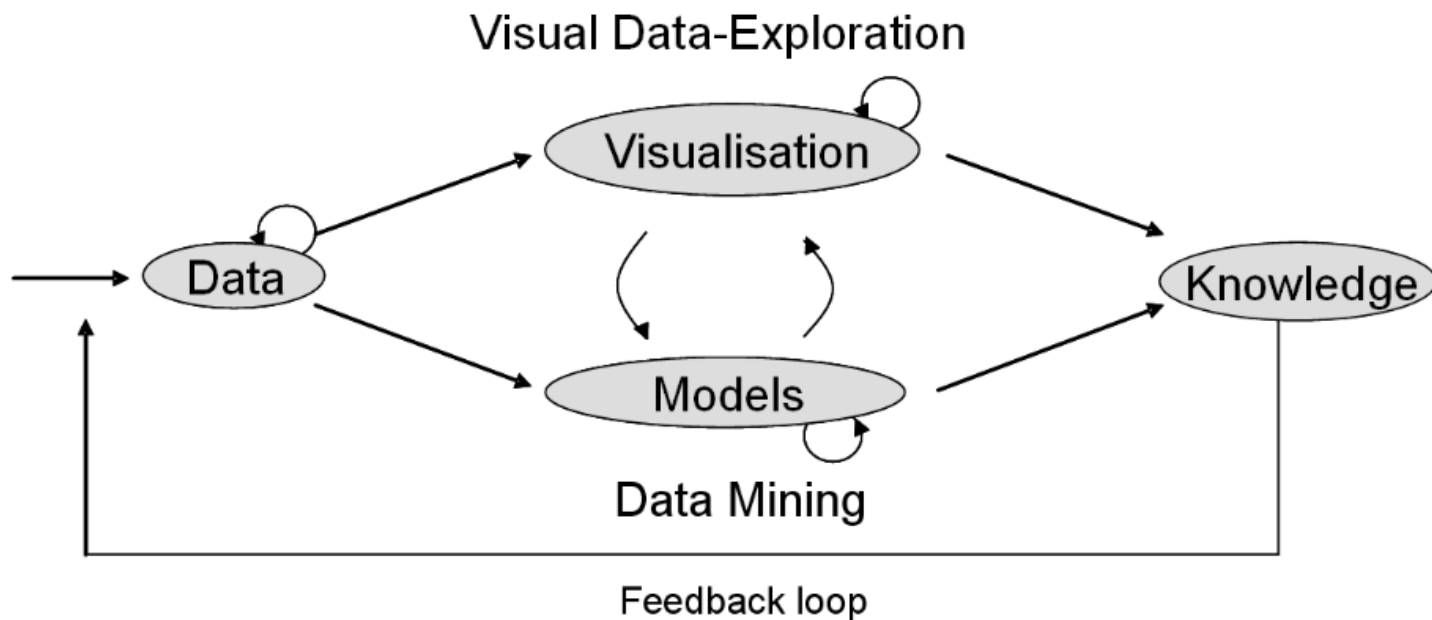
La Analítica Visual busca **aumentar la capacidad del humano** en lugar de reemplazarlo con métodos de decisión computacional.



Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4950 LNCS, pp. 154–175). Berlin, Heidelberg: Springer Berlin Heidelberg.



Visual Data-Exploration



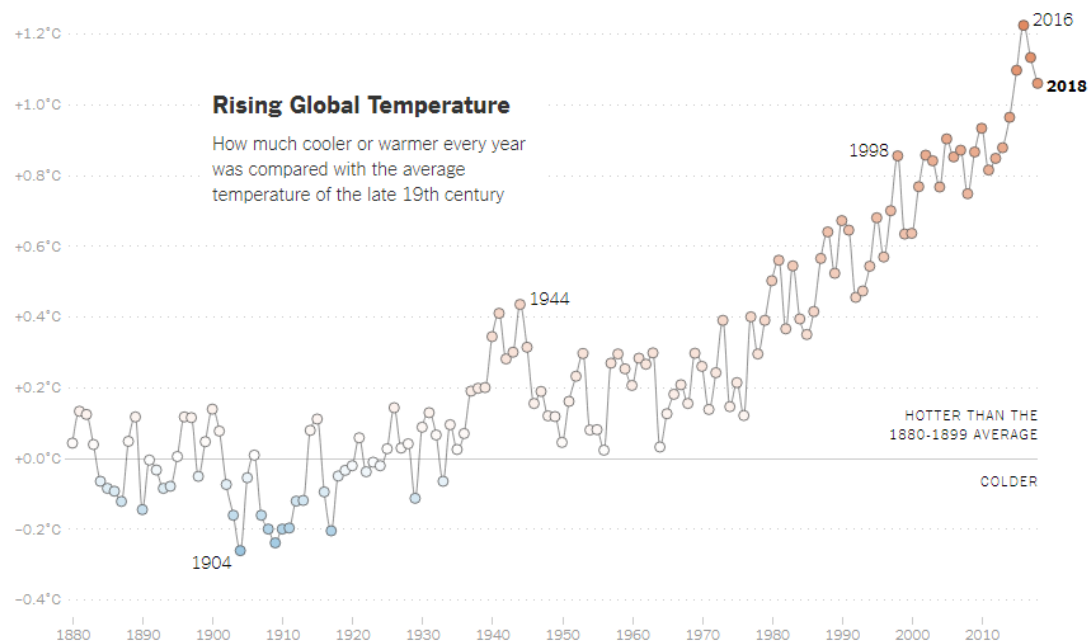
Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4950 LNCS, pp. 154–175). Berlin, Heidelberg: Springer Berlin Heidelberg.



¿Para qué usar visualizaciones?

- Explicativo
- Exploratorio
- Confirmatorio

Teach About Climate Change With These 24 New York Times Graphs



Source: NASA | By The New York Times

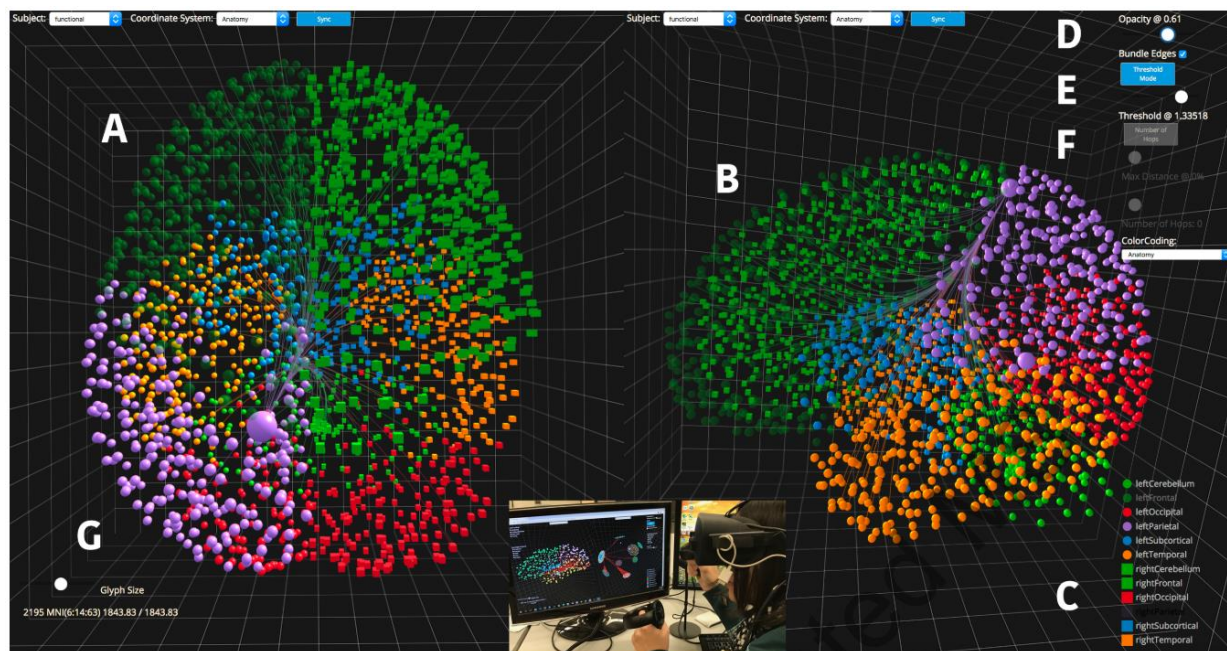


¿Para qué usar visualizaciones?

- Explicativo

- Exploratorio

- Confirmatorio





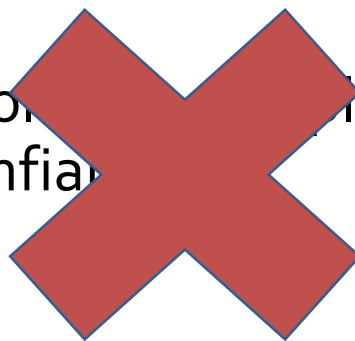
¿Para qué usar visualizaciones?

- Fines explicativos.
- Fines exploratorios.
- Carácter confirmatorio.
- Cuando existe una solución completamente automática y dicha solución es confiable.



¿Para qué usar visualizaciones?

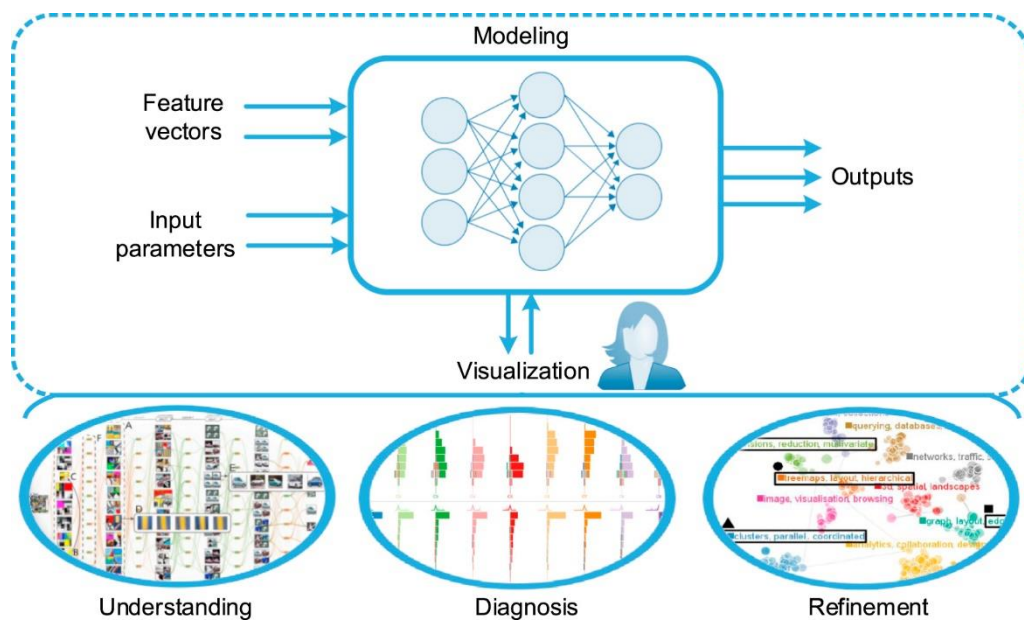
- Fines explicativos.
- Fines exploratorios.
- Carácter confirmatorio.
- Cuando existe una solución completamente automática y dicha solución es confiable.





Una visualización permite ...

- Análisis exploratorio de datos.
- Presentación de resultados.
- Ayudar a los usuarios a entender lo que están haciendo los modelos automáticos.





Una visualización permite ...

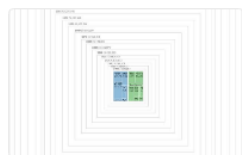
- Análisis exploratorio de datos.
- Presentación de resultados.
- Ayudar a los usuarios finales a entender lo que están haciendo los modelos automáticos.

Reemplazar cognición (pensamiento) con percepción

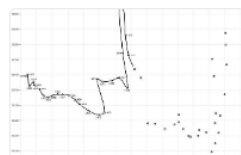


Idioms (Modismos)

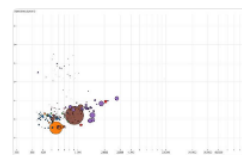
- Representaciones basadas en computador, externas, que utilizan los conceptos de percepción para crear las visualizaciones.



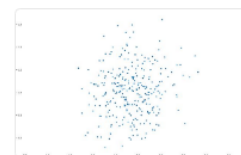
Animated treemap



Connected scatterplot



The wealth & health of nations

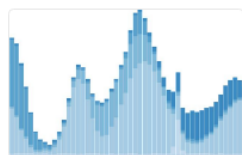


Scatterplot tour

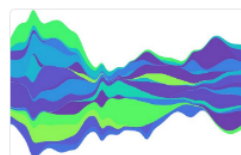


Bar chart race

2000



Stacked-to-grouped bars



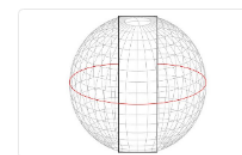
Streamgraph transitions



Smooth zooming



Zoom to bounding box



Orthographic to equirectangular



World tour



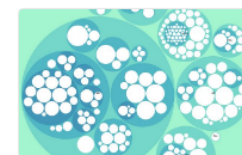
Walmart's growth



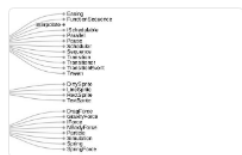
Hierarchical bar chart



Zoomable treemap



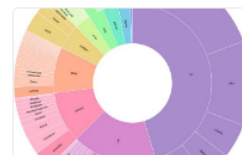
Zoomable circle packing



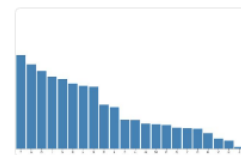
Collapsible tree



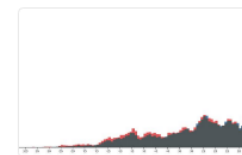
Zoomable icicle



Zoomable sunburst



Sortable bar chart



Icelandic population by age, 18...



Idioms (Modismos)

- Representaciones basadas en computador, externas, que utilizan los conceptos de percepción para crear las visualizaciones.
- El espacio de búsqueda de soluciones para visualización es **infinito**
- **Se requieren técnicas para seleccionar la mejor visualización:**
 - ¿Cómo dibujarlo?
 - ¿Cómo manipularlo?

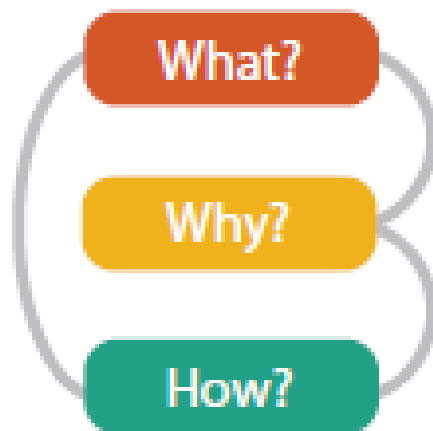


Limitaciones en Recursos

- Limitaciones computacionales:
 - Tiempo de procesamiento.
 - Memoria del sistema.
- Limitaciones humano:
 - Atención.
 - Memoria.
 - Retención.
- Limitaciones de los display (pantallas):
 - Pixeles son un recurso precioso, el recurso máspreciado.
 - Densidad de la información



Framework de Visualización



- ¿Qué?
 - Abstracción de los datos.
- ¿Por qué?
 - Abstracción de la tarea.
- ¿Cómo se muestra?
 - Idiom: “codificación visual” e interacción

(Framework de Tamara Munzner)



What?

Datasets

Attributes

Data Types

→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

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

Attribute Types

→ Categorical



→ Ordered

→ Ordinal

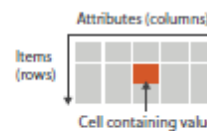


→ Quantitative

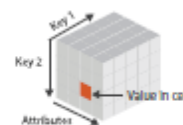


Dataset Types

→ Tables



→ Multidimensional Table



→ Networks



→ Trees



→ Fields (Continuous)



→ Geometry (Spatial)



Dataset Availability

→ Static



→ Dynamic



Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



What?

Why?

How?



Why?

Actions

Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



Search

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

Query

→ Identify



→ Compare



→ Summarize



Targets

All Data

→ Trends



→ Outliers



→ Features



Attributes

→ One

→ Distribution



→ Extremes

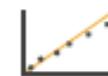


→ Many

→ Dependency



→ Correlation



→ Similarity



Network Data

→ Topology



→ Paths



Spatial Data

→ Shape



Figure 3.1. Why people are using vis in terms of actions and targets.

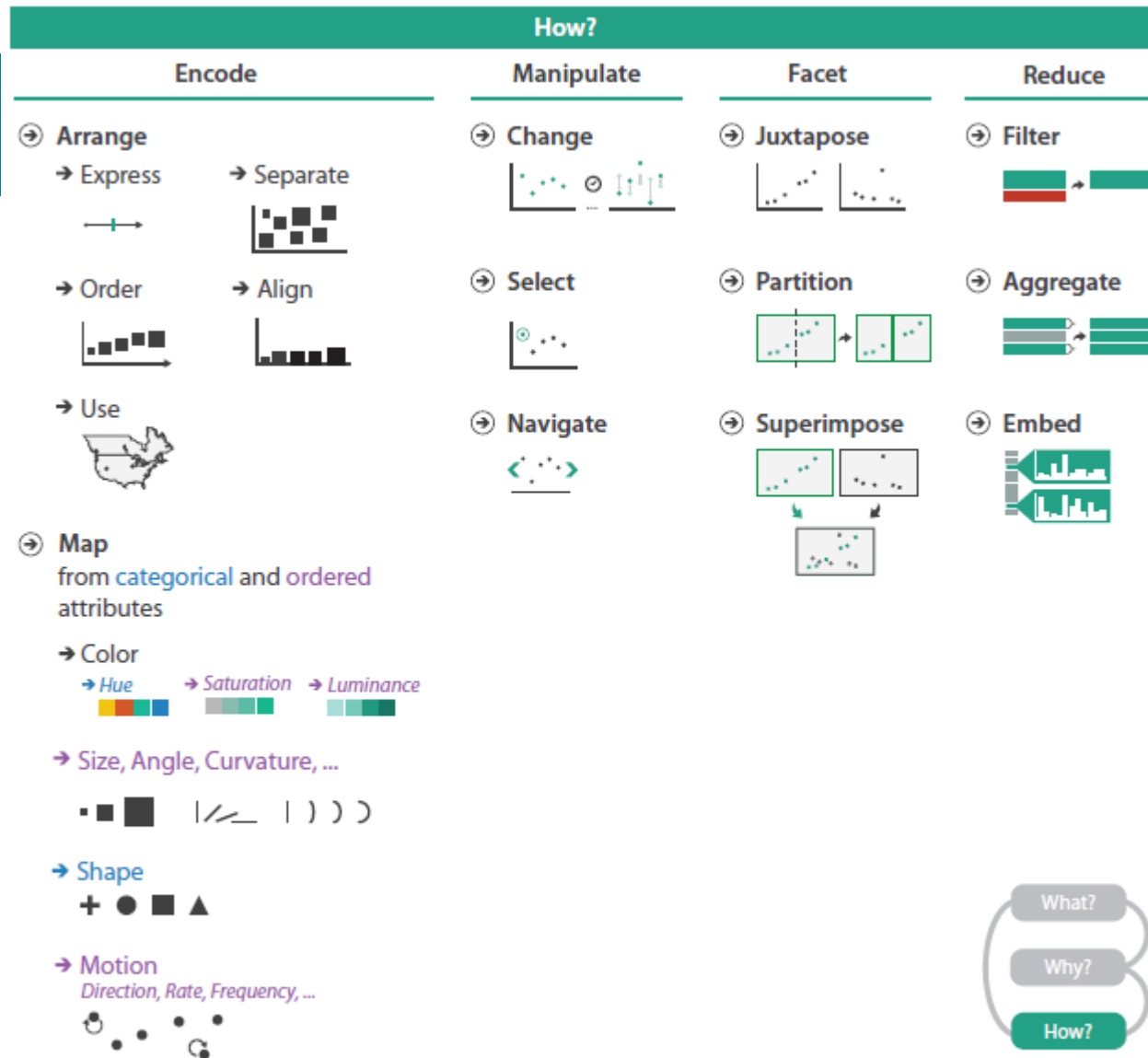
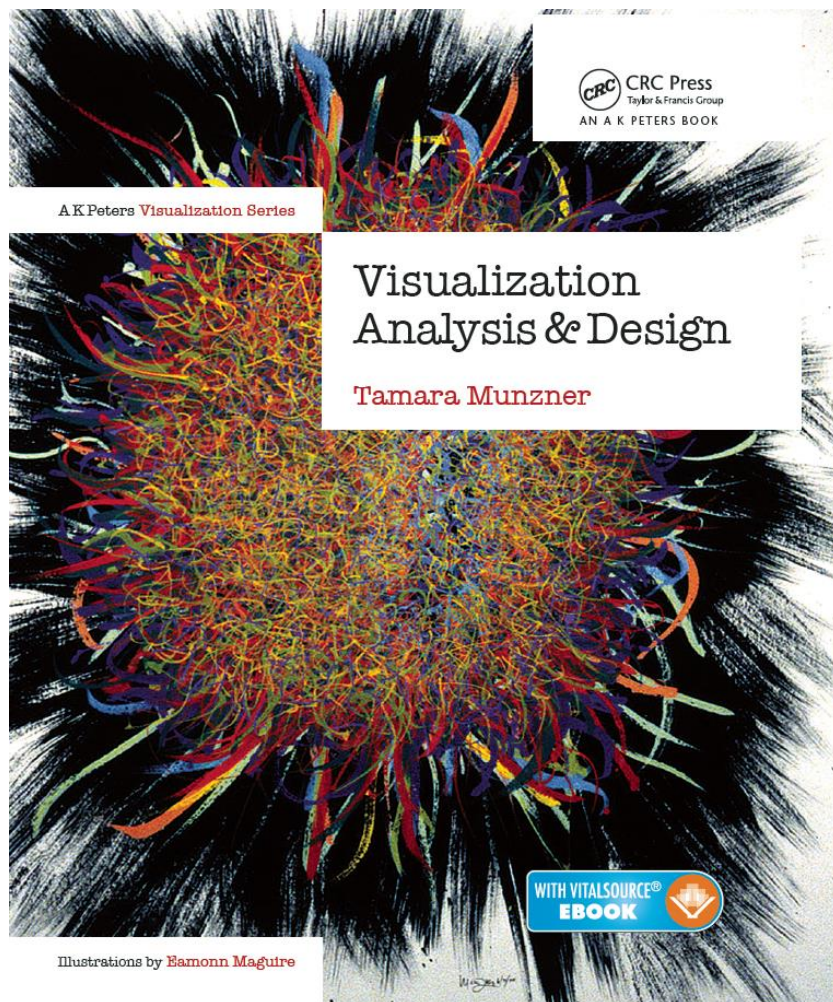


Figure 3.7. How to design vis idioms: encode, manipulate, facet, and reduce.



Visualization Analysis and Design



Tamara Munzner



Insights

- Conocimiento profundo del problema.
- A partir de los datos.
- No obvio.
- Que provoque acciones o decisiones.



150
ING

la Argentina celebra
su ingeniería
1870-2020

II Simposio de Informática INDUSTRIA 4.0



08 al 16 DE JUNIO

2020

Para recordar:

- Datos, Usuarios y Tarea.
- En el análisis exploratorio de datos lo que importa es encontrar **insights**.
- Visual Analytics, es una buena forma de encontrar **insights**

Gracias!



Attribution-NonCommercial-
NoDerivatives 4.0 International
(CC BY-NC-ND 4.0)

150 ING

la Argentina celebra
su ingeniería
1870-2020

II Simposio de Informática INDUSTRIA 4.0



Contacto:

Duván Alberto Gómez Betancur

duvan.gomez34@tdea.edu.co



@dagomezb



Enlace a la disertación virtual

<https://www.youtube.com/watch?v=Za3oZaiUFgg>

¿Preguntas?



Este obra está bajo
una [licencia de Creative
Commons Reconocimiento-
NoComercial-
SinObraDerivada 4.0
Internacional](#).

Esta licencia permite copiar y
distribuir libremente la obra
pero obliga a atribuir la autoría
y prohíbe la creación de obras
derivadas (remezcla) y el uso
comercial

UNCA
UNIVERSIDAD NACIONAL DE CATAMARCA



**Tecnológico
de Antioquia**
Institución Universitaria



POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID