

Visualización de información y analítica visual

Clase 1: Introducción al curso



Visualización de información

*Computer based visualization systems
provide visual representations of
datasets designed to help people carry
out tasks more effectively*

Tamara Munzner



¿Por qué este curso? 🙄

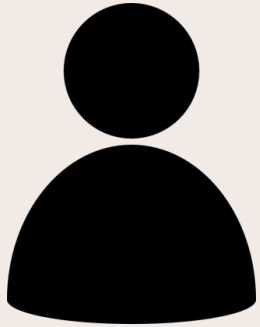


¿Por qué este curso?

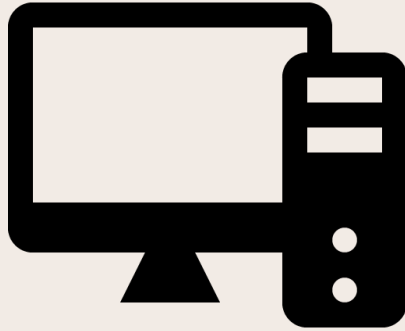
- 1. Las posibilidades para diseñar visualizaciones son muchas...**
- 2. Hay visualizaciones que no sirven para algunas tareas**
- 3. Los tres actores involucrados: el computador, la persona y la visualización presentan limitantes**



Los tres actores



El humano



El computador

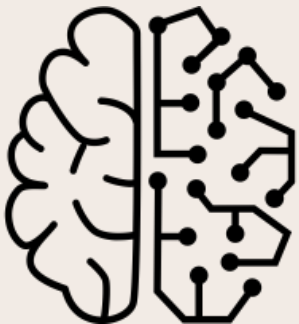


La visualización

El objetivo: llevar a cabo una tarea

¿Por qué/cuándo incluir al humano?

En general el humano se incluye cuando la tarea no es tan específica, y no es muy claro el qué preguntas hay que hacerse para llevarla a cabo.



- El uso de **machine learning/IA** nos ha permitido quitar al humano del loop en ciertas tareas (o en *partes* de ciertas tareas)

Al quitar al humano también quitamos la necesidad de visualizar. El modelo de ML no necesita una visualización de los datos



¿Por qué/cuándo incluir al humano?

Pero hay tareas (o partes de ellas) donde no hay un paso-a-paso específico

- **Nos podemos hacer distintas preguntas**
- **Podemos abordar distintos enfoques**

pero no se sabe a priori cual es la pregunta/enfoque correcto...

En este caso lo mejor es tener a un humano y usar visualizaciones para potenciar sus capacidades.



¿Por qué/cuándo incluir al humano?

Siempre que haya participación de un humano, se puede potenciar sus capacidades por medio de visualizaciones

Ejemplos:

- **Crear una visualización para monitorear el entrenamiento epoch-a-epoch de un modelo de DL**
- **Crear una visualización para interpretar datos de alguna investigación científica**
- **Crear una visualización para ayudar a una persona a explicar algo que esa persona ya sabe a otros.**



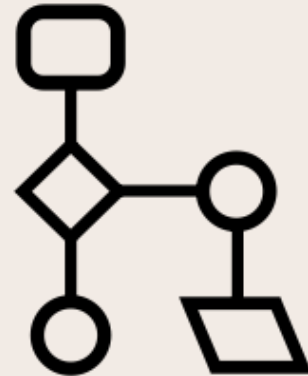
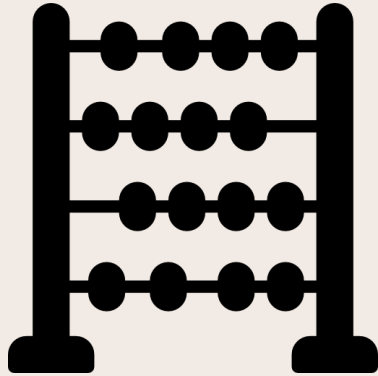
¿Por qué incluir al computador?

El uso de herramientas computacionales ha permitido tratar volúmenes de datos imposibles de tratar a mano.



¿Por qué usar representaciones de los datos?

Una **representación externa** de los datos potencia las capacidades humanas, permitiéndole ir más allá de sus limitaciones cognitivas y de memoria.



Y por qué usar representaciones visuales?

El sistema visual es el más adecuado (al menos actualmente) para transmitir información.



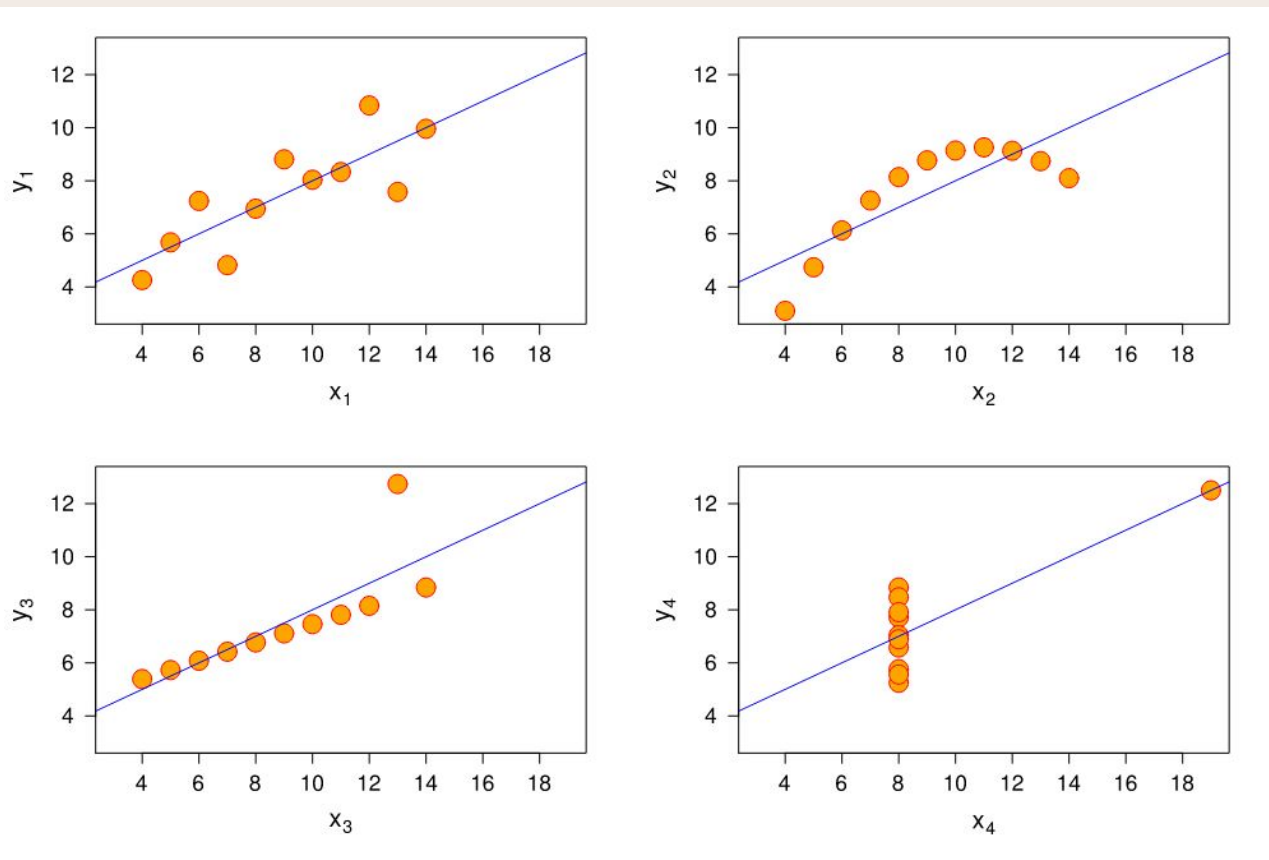
Una visualización nos permite ver los datos en detalle

I		II		III		IV	
X1	Y1	X2	Y2	X3	Y3	X4	Y4
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

Para todos los datasets:

- **Media de x: 9**
- **Varianza de x: 11**
- **Media de y: 7.5**
- **Varianza de y: 4.125**
- **Correlación x-y: 0.816**
- **Regresión lineal**
 $y = 3 + 0.5x$
con $R^2 = 0.67$

Una visualización nos permite ver los datos en detalle



**Ahora ahondemos un poco en
ciertos aspectos de visualización
de información**

Tenemos muchas opciones disponibles

Existen muchas opciones de representar un mismo conjunto de datos:

- Tipo de gráfico
- Color/es
- Formas
- Interactividad
- Estilo
- Combinaciones de gráficos en una misma



**¿Qué criterio usan
ustedes para
diseñar una
visualización?**

Puede ser un poco overwhelming tener que elegir una visualización y justificar que es la visualización adecuada para una tarea en específico

Qué debemos tomar en cuenta para diseñar una visualización

Lo más básico debería ser enfocarnos en los datos que tenemos... cierto?

1. Foco en la tarea

2. Foco en la efectividad



Foco en la tarea

Una herramienta que sirve muy bien para una tarea puede no ser adecuada para otra. En general la tarea a la que sirve la visualización es igual de importante que el tipo de dato que tenemos.

En general vamos a tener que reformular la tarea: pasarla de una tarea de un dominio específico y **abstraerla (Clase 3)**



Foco en la efectividad

Definimos una visualización como un apoyo para las tareas de usuario, de ahí que sea importante enfocarnos en la efectividad.

Hay un compromiso con la correctitud, precisión y veracidad de nuestra visualización y los datos que representa.

It's not just about making pretty pictures

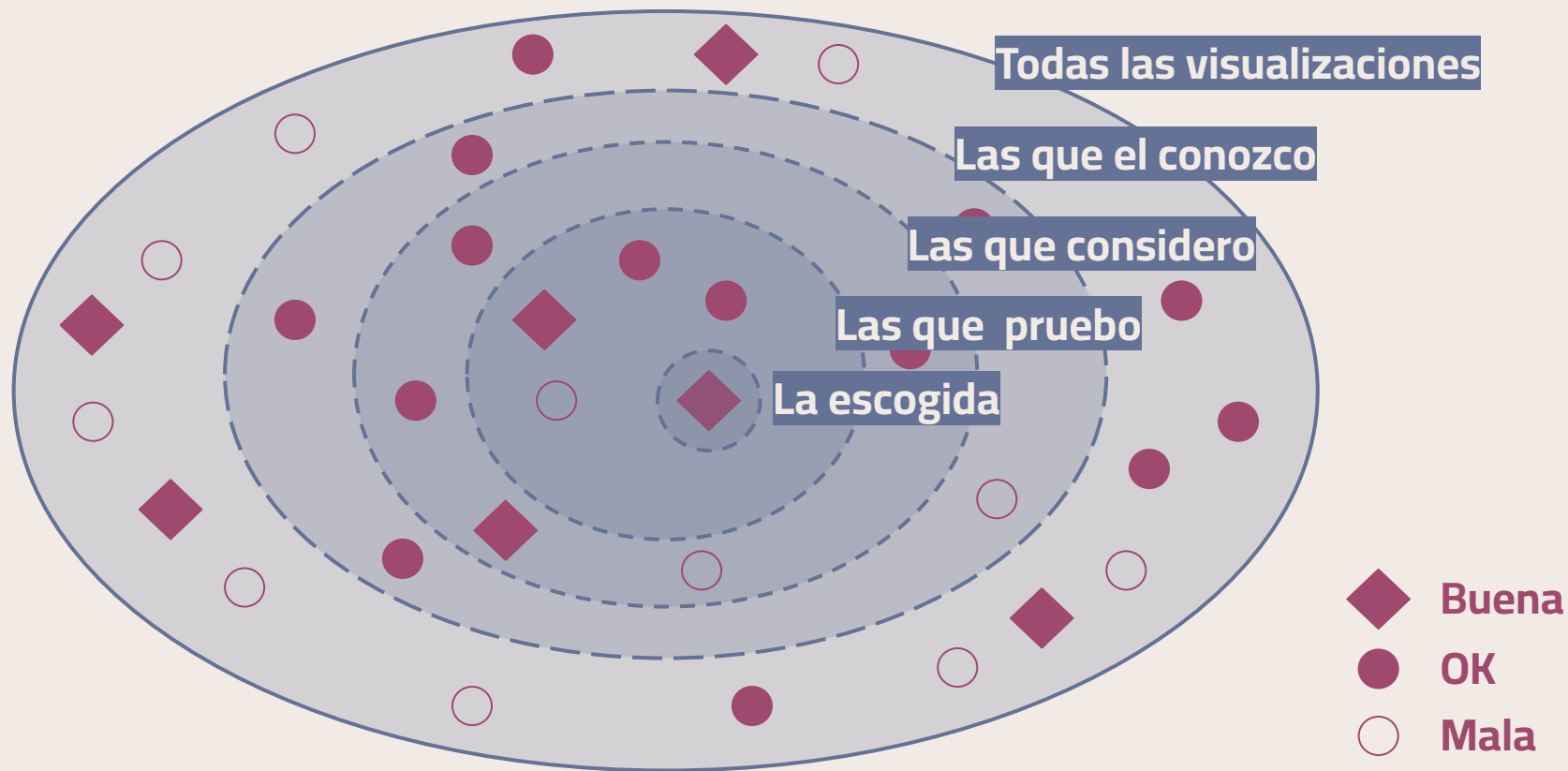


Por qué algunas visualizaciones son tan malas...

La principal razón por la cual el diseño de visualizaciones es complejo es debido a la enorme cantidad de posibilidades en el espacio de diseño, donde la gran mayoría de ellas no van a ser efectivas para un contexto de uso específico.

En general, uno no tiene que enfocarse en buscar la mejor solución de todas (optimizar), sino en encontrar la mejor solución entre varias que sean buenas (satisfacer).

Espacio de posibilidades



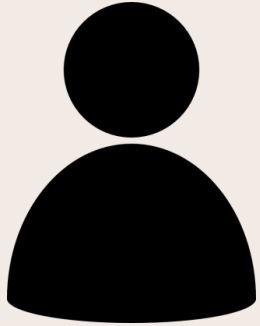
Es importante poder validar las visualizaciones

pero también es muy difícil...

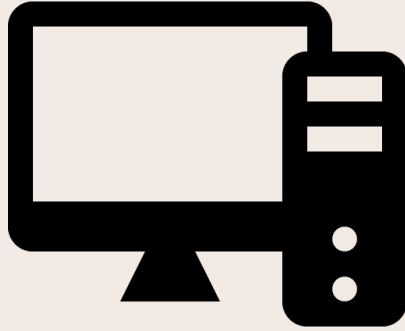
- cómo sé que la visualización funciona?
- cómo sé qué diseño es mejor para el usuario objetivo?
 - el que entienden más rápido?
 - el que les queda grabado en la memoria?
 - el que encuentren más bonito/entretenido?
 - el que les permite ser más *eficientes*?
- quién es este usuario?



Y hay que tener en consideración ciertas limitantes



**Capacidades
cognitivas y
perceptuales de los
humanos**

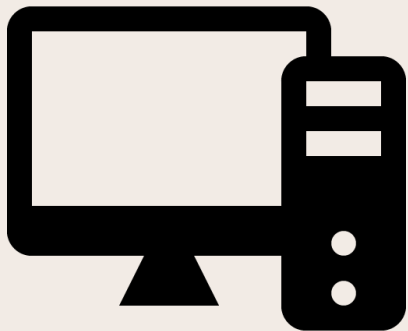


**Capacidad
computacional**



**Capacidad de
display**

Y hay que tener en consideración ciertas limitantes



**Capacidad
computacional**

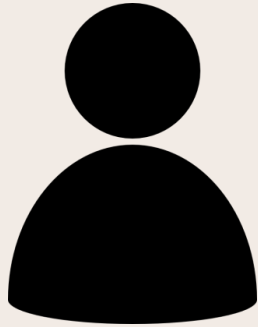
- **Escalabilidad**

Diseñar visualizaciones que puedan manejar grandes volúmenes de datos

- **Uso de memoria**
- **Tiempo de ejecución**

Hay que tener especial consideración en esto cuando la visualización es interactiva o dinamica

Y hay que tener en consideración ciertas limitantes



**Capacidades cognitivas
y perceptuales de los
humanos**

- La memoria y la atención son recursos limitados
- Algunas diferencias visuales son imperceptibles para los humanos

Clase 2 (próxima clase)!

Y hay que tener en consideración ciertas limitantes

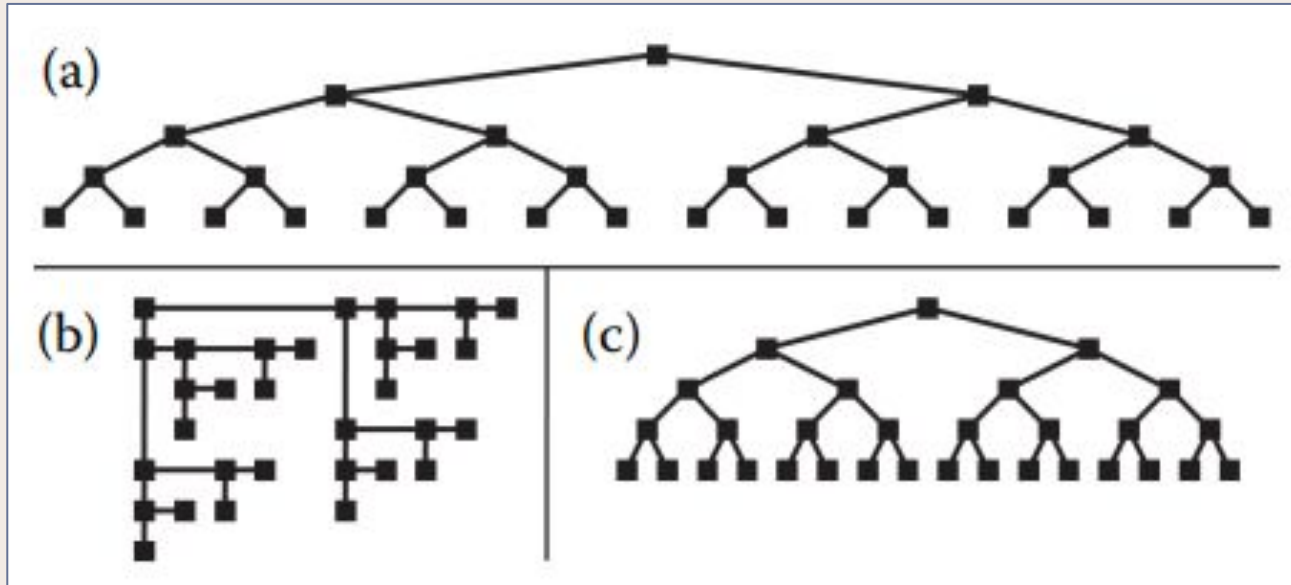


Capacidad de
display

Finalmente, el display donde se va a ver nuestra visualización es limitado:

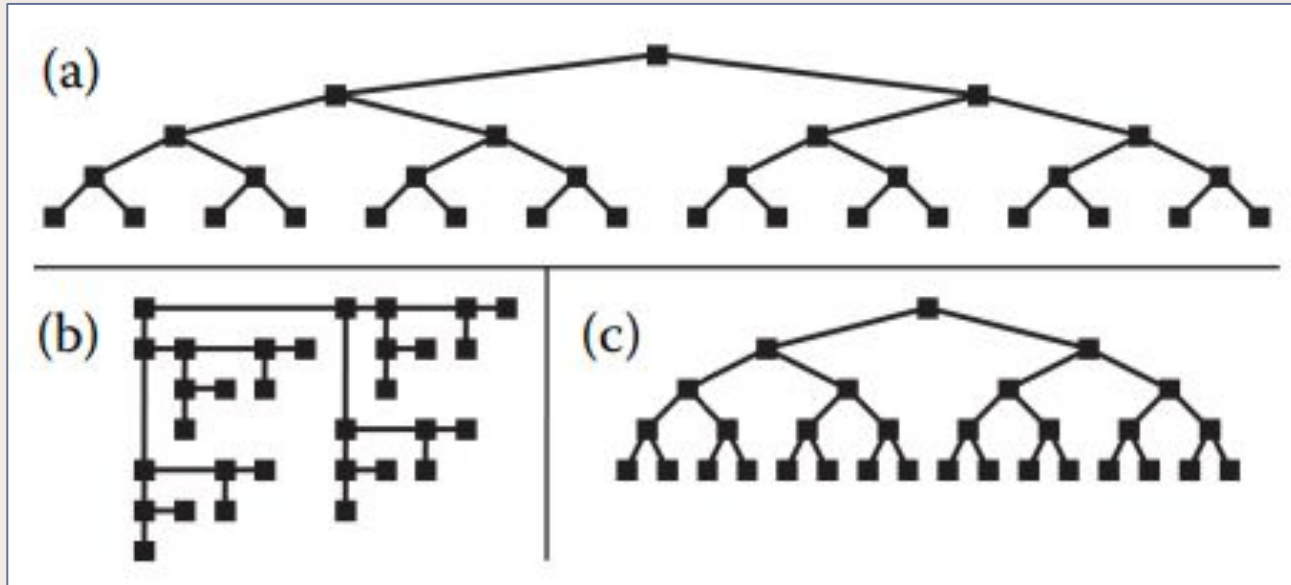
- hay que cuidar la densidad de información
- Trade-off: mostrar toda la información en una misma visualización y la navegación por múltiples visualizaciones

Y hay que tener en consideración ciertas limitantes



- a) Baja densidad de información
- b) Alta densidad, pero se hace difícil visualizar el concepto de profundidad
- c) Una buena alternativa

Y hay que tener en consideración ciertas limitantes



- a) Baja densidad de información
- b) Alta densidad, pero se hace difícil visualizar el concepto de profundidad
- c) Una buena alternativa

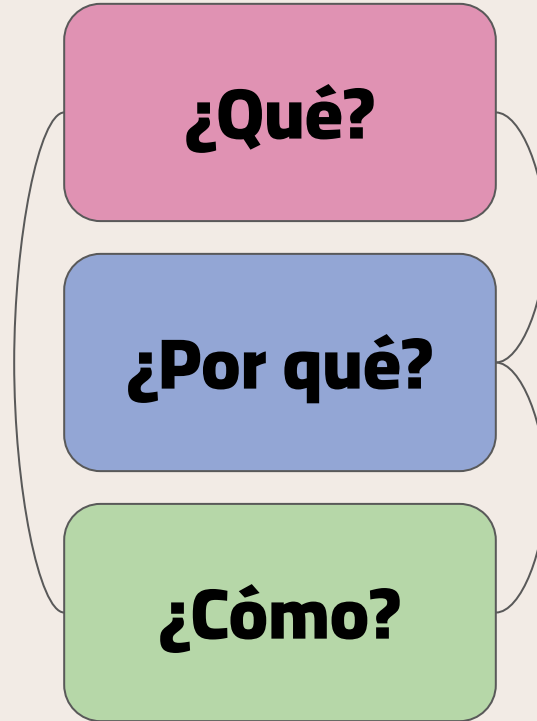
**En este curso el enfoque estará
en el análisis**

La filosofía del curso

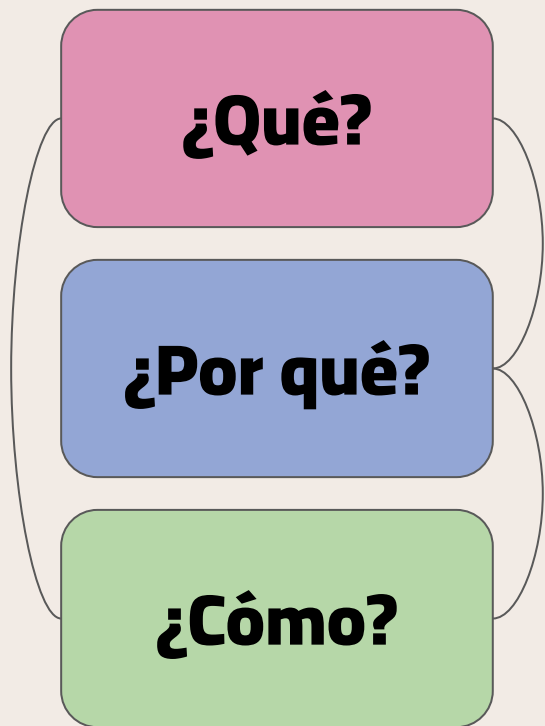
**¡Analizar visualizaciones existentes es una buena base
para diseñar nuevas 🌈!**

Para esto vamos a trabajar con el framework de análisis propuesto por Tamara Munzner en el libro Visualization Analysis and Design (libro guía del curso) que busca darle estructura al espacio de diseño de visualizaciones.

La filosofía del curso



La filosofía del curso



¿Qué datos está viendo el usuario?

¿Por qué el usuario está usando esta visualización?

¿Cómo están contruidos los elementos y codificaciones visuales?

Framework anidado de Tamara Munzner

Por las razones que conversamos anteriormente, es necesario tener un modelo de trabajo (Clase 3)

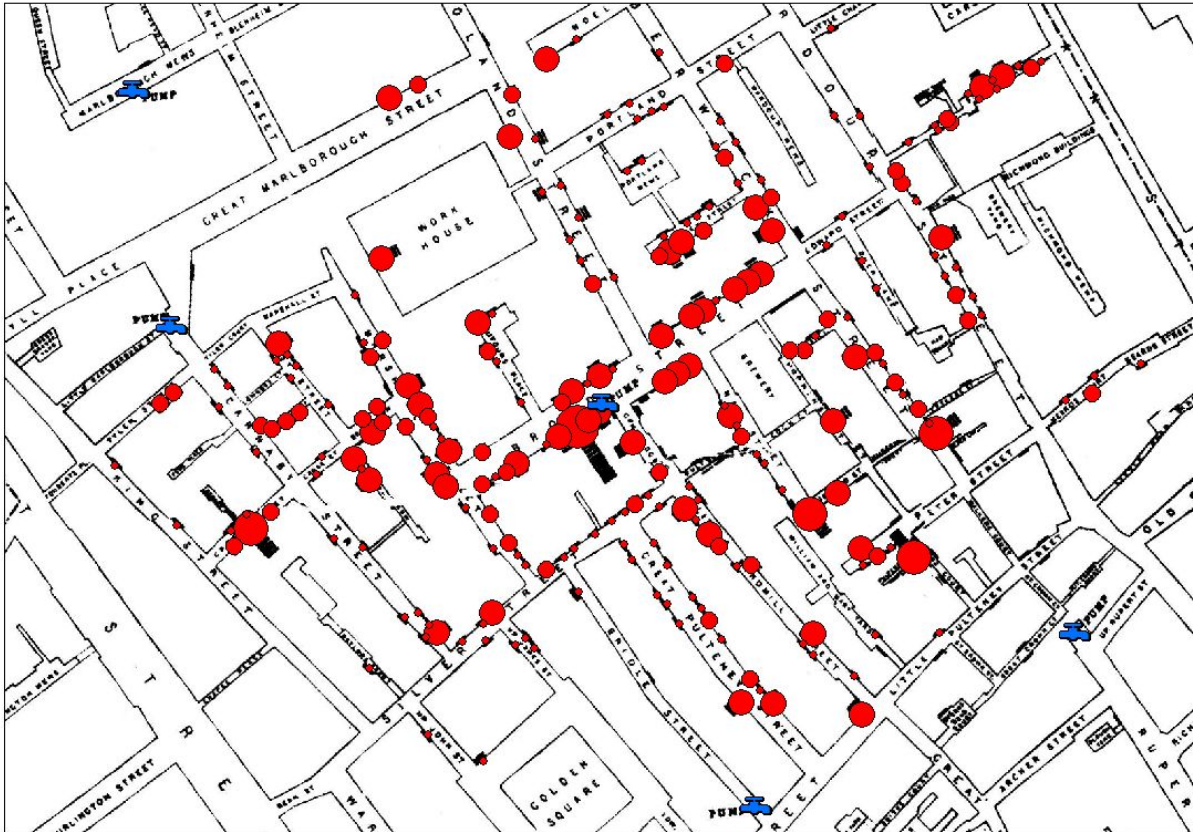
Caracterización del dominio

Abstracción de datos y tareas

Codificación visual y de interacción

Implementación algorítmica

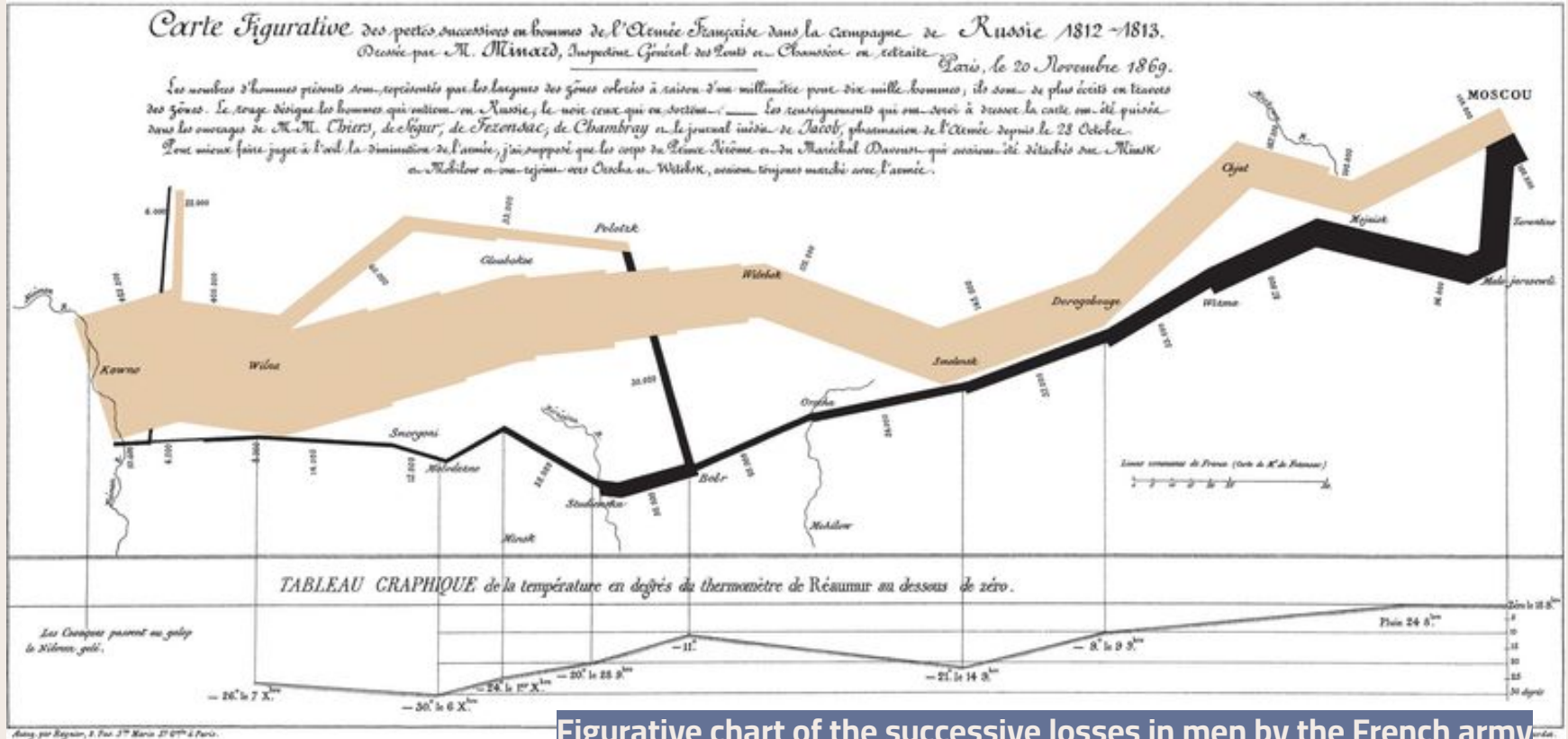
Ejemplos de visualizaciones



**Mapa del brote de
cólera en Londres,
1854.**

Dr. John Snow

Ejemplos de visualizaciones



Figurative chart of the successive losses in men by the French army in the Russian campaign 1812-1813. Charles Joseph Minard

Ejemplos de visualizaciones

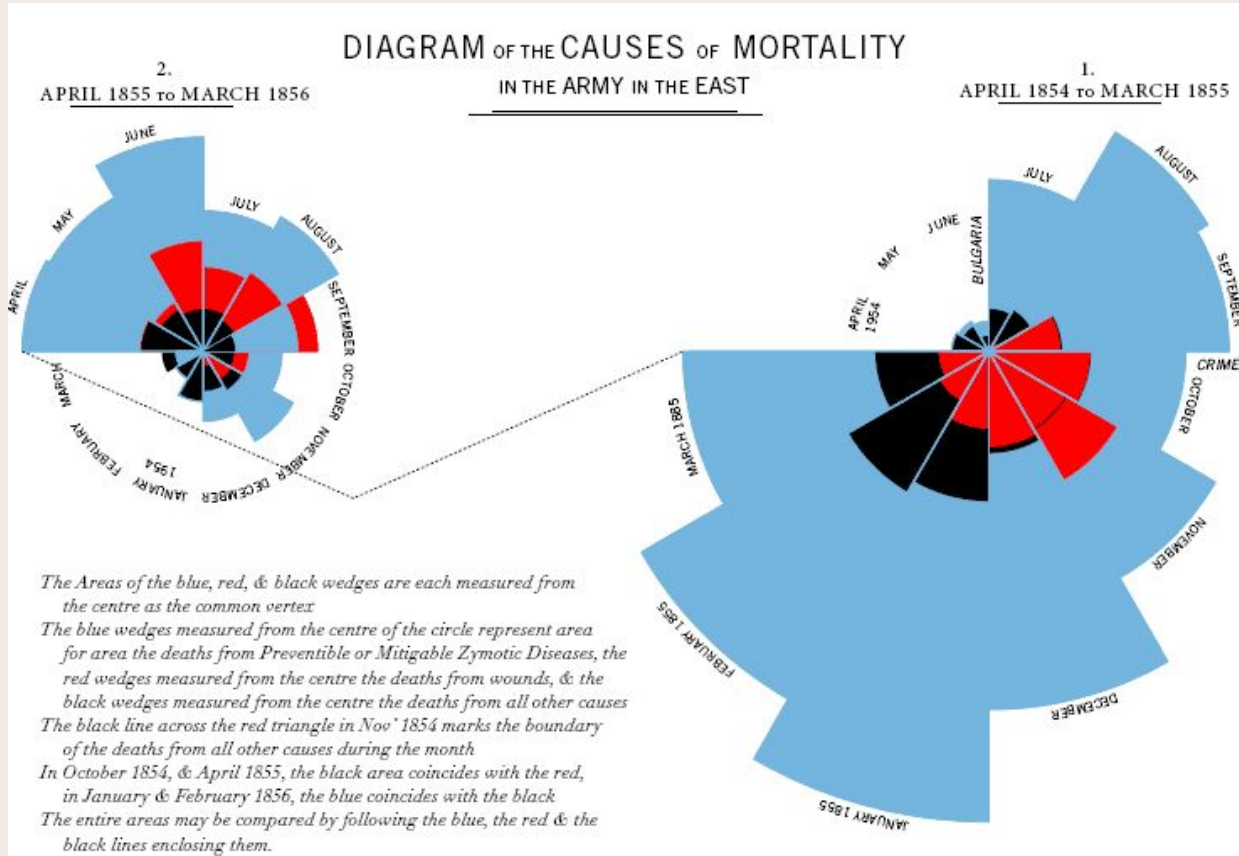


Diagram of the
causes of mortality
in the army in the
east.

Florence Nightingale

Evaluaciones del curso

- **Laboratorios (30%)**

- 4 actividades en horario de clases (ver calendario, el primero es hoy). De estos se evalúan 2.
- 1 participación en foro (todo el semestre)

- **Controles (30%)**

- 3 actividades en horario de clases (ver calendario). Se evaluará 2.

- **Proyecto final (40%)**

- Entrega última semana de clases

Cronograma del curso

1	Introducción al curso	Laboratorio
2	Percepción, marcas, canales y color	Altair
3	Framework, abstracción de datos y tareas	Streamlit
4	Rules of thumb	Laboratorio
5	Datos tabulares y manipulación de vistas	-Control-
6	Grafos, jerarquías, y otros tipos de datos	Laboratorio
7	Interactividad en visualizaciones	Laboratorio
8	Validación de visualizaciones (*)	-Control-
9	Herramientas de infovis en la industria	-Control-

Laboratorio 1:

Actividad en RawGraphs

Visualización de información y analítica visual

Clase 1: Introducción al curso

