
IIC2026

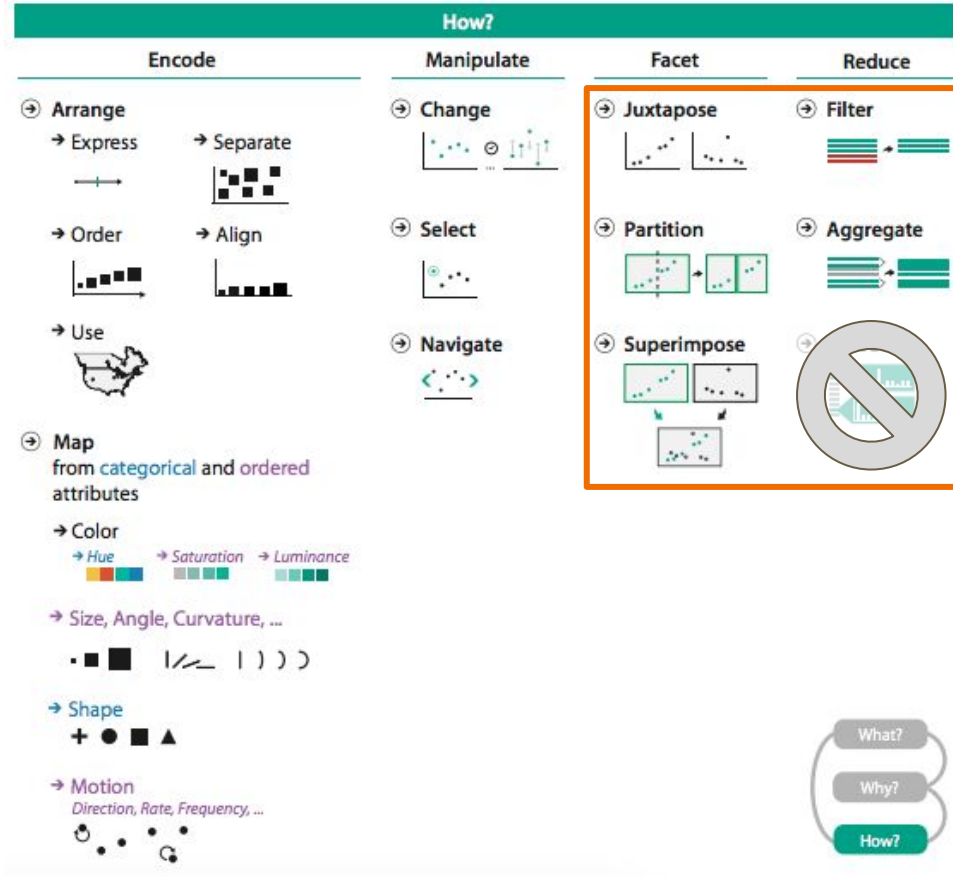
Visualización de Información

— Hernán F. Valdivieso López —
(2023 - 1 / Clase 12)

Temas de la clase - Facetas y Reducción de Datos

1. Yuxtaposición y vista coordinadas
2. Partición de vistas
3. Superposición
4. Filtrar
5. Agregación
6. Reducción de dimensionalidad

Temas de la clase - Facetas y Reducción



Yuxtaposición y vista coordinadas

Yuxtaposición

- Separar espacio en más de una vista simultánea.
- Las principales decisiones a considerar en este tema es:
 - ¿Se comparten los datos en cada vista?
 - ¿Se comparten las codificaciones visuales en cada vista?
 - ¿Cómo se organizan y visibilizan las vistas? Todas al mismo tiempo, hay que hacer scroll. Una al lado de otra o encima de otra.

Yuxtaposición

¿Se comparten los datos en cada vista?

En esta decisión, podemos encontrar 3 principales opciones:

- **Datos compartidos.** Todas las vistas presentan exactamente los mismos datos, generalmente con diferentes visualizaciones o agrupados de diferentes formas.
- **Panorámica-detalle.** Una vista muestra un subconjunto de lo que muestra la otra vista.
- **Partición de datos.** No hay datos compartidos. Cada vista presenta un subconjunto de datos diferentes.

Yuxtaposición

¿Se comparten los datos en cada vista?

- **Datos compartidos.** Todas las vistas presentan exactamente los mismos datos, generalmente con diferentes visualizaciones o agrupados de diferentes formas.
- En el ejemplo se muestran los datos agrupados en una suma y en la otra vista segmentados por país (aunque falta la leyenda)

Venta Mensual



Distribución de ventas



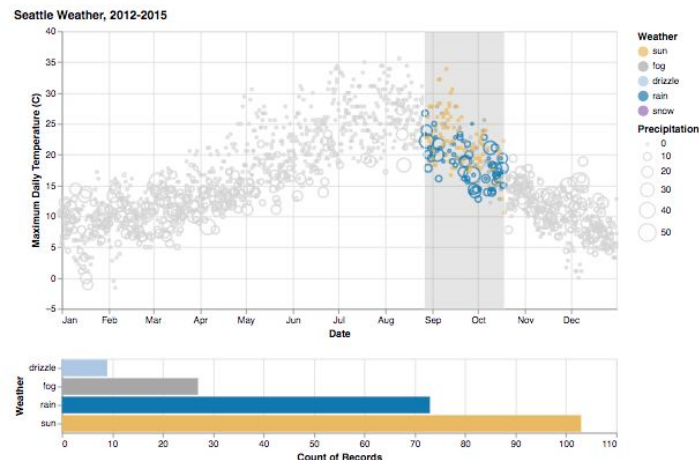
Yuxtaposición

¿Se comparten los datos en cada vista?

- **Panorámica-detalle.** Una vista muestra un subconjunto de lo que muestra la otra vista.
- En el ejemplo se muestran los datos en un gráfico de burbuja y cuando seleccionas un conjunto, se muestra un histograma de dicho conjunto.

Seattle Weather Exploration

This graph shows an interactive view of Seattle's weather, including maximum temperature, amount of precipitation, and type of weather. By clicking and dragging on the scatter plot, you can see the proportion of days in that range that have sun, rain, fog, snow, etc. Created by @jakevdp.



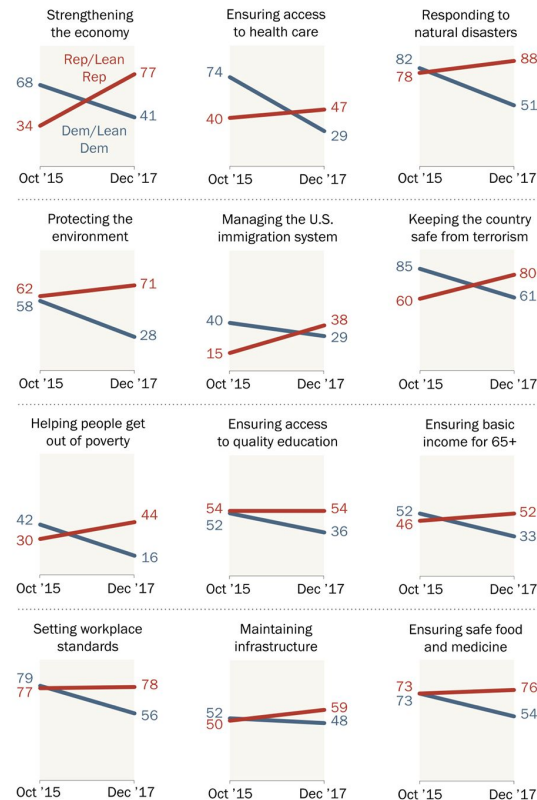
Yuxtaposición

¿Se comparten los datos en cada vista?

- **Partición de datos.** No hay datos compartidos. Cada vista presenta un subconjunto de datos diferentes.
- Cuando se empareja esta opción con la misma codificación en cada vista, suele llamarse como múltiples pequeños (*small múltiple*)

Shifting partisan views of government performance

% who say the federal government is doing a very/somewhat good job ...



Source: Survey of U.S. adults conducted Nov. 29-Dec. 4, 2017.

PEW RESEARCH CENTER

Yuxtaposición

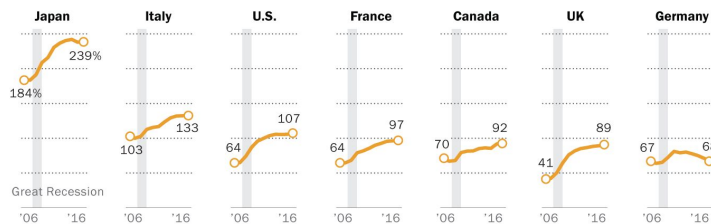
¿Se comparten las codificaciones visuales en cada vista?

En esta decisión, podemos encontrar 2 opciones:

- **Si:** todos los gráficos usan la misma codificación visual.
- **No:** una o más vistas usan una codificación visual distinta. Esta también es conocida como **vistas multiforma**

After Great Recession, debt increased substantially in most G-7 economies

Total gross debt as a share of GDP in the Group of Seven nations

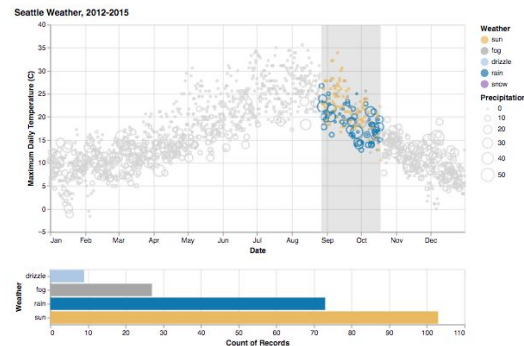


Note: Gross debt represents total liabilities of all levels and units of government — national, state/provincial and local — less liabilities held by other levels or units of government, unless otherwise noted by source.
Source: The International Monetary Fund, World Economic Outlook, accessed Sept 7, 2017.

PEW RESEARCH CENTER

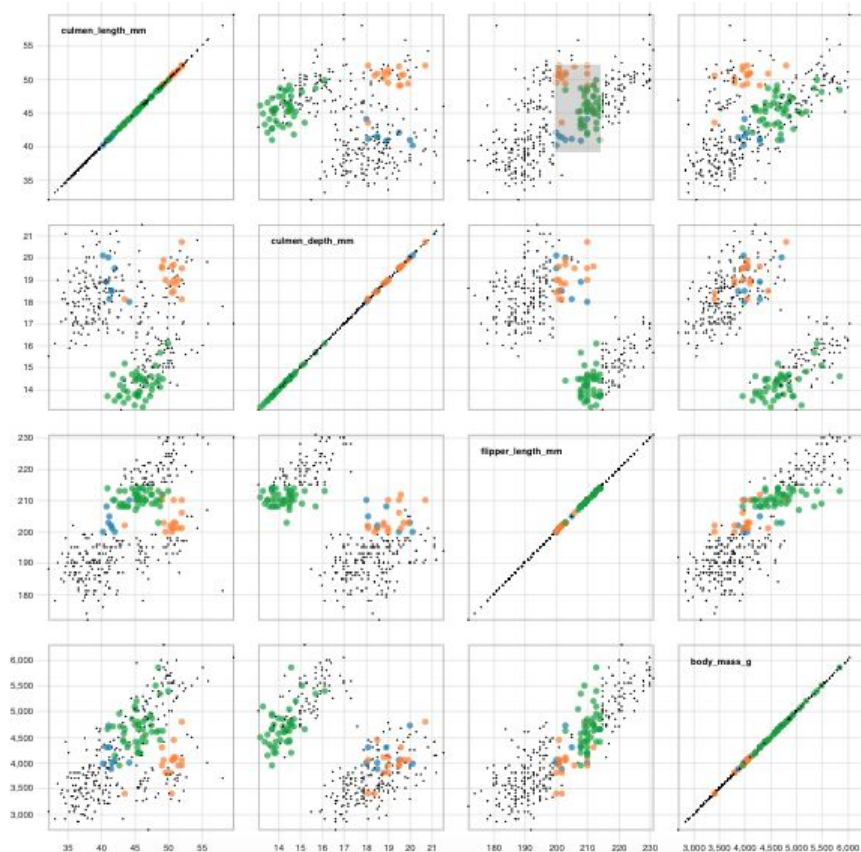
Seattle Weather Exploration

This graph shows an interactive view of Seattle's weather, including maximum temperature, amount of precipitation, and type of weather. By clicking and dragging on the scatter plot, you can see the proportion of days in that range that have sun, rain, fog, snow, etc. Created by @jakevdp.



Yuxtaposición - Vistas coordinadas

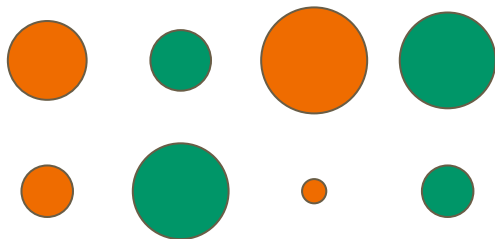
- Caso interactivo donde la acción en una vista repercute en otras vistas.
- Los casos típicos son provocar filtros o enfatizar el mismo dato en las demás vistas.



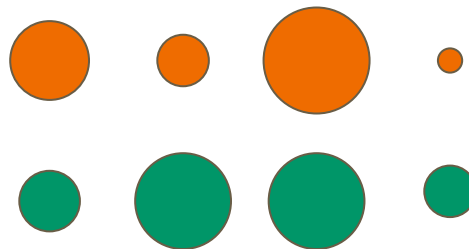
Partición de vistas

Partición de vistas

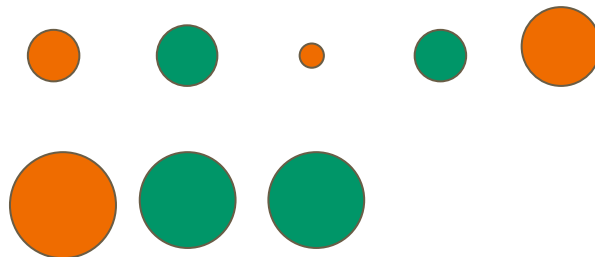
- Decisión específica de yuxtaposición donde cada vista presenta un conjunto de los datos bajo un criterio en común.
- Esta decisión consiste en organizar un dataset de múltiples atributos en grupos significativos en base a sus valores y cómo organizar dichos grupos espacialmente en vistas.



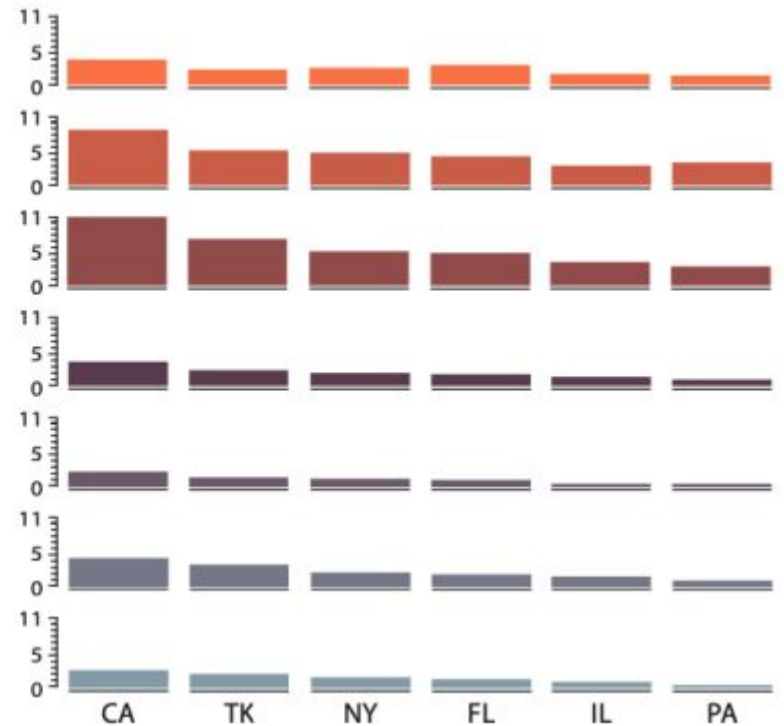
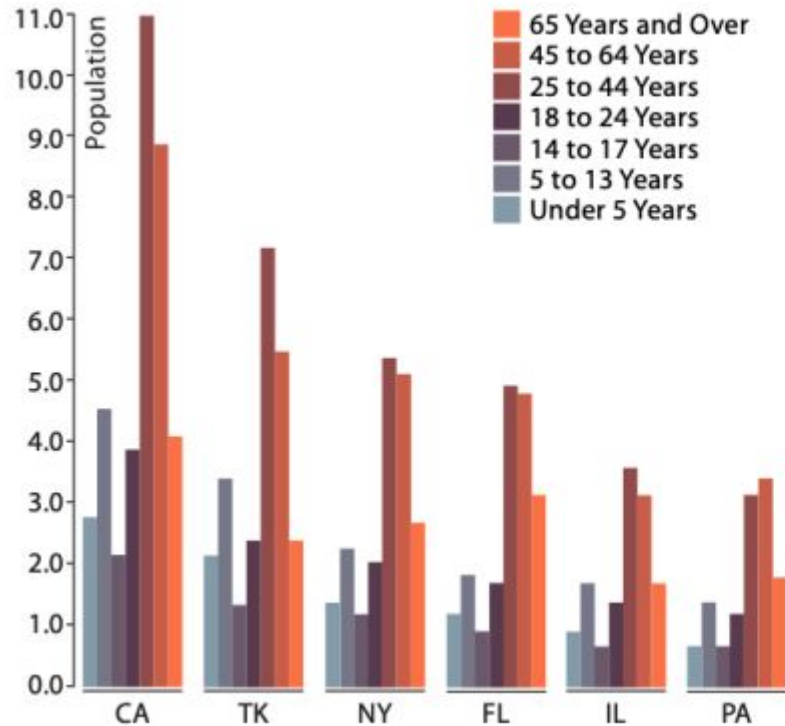
Por algún
atributo
categórico



Por algún
atributo
ordenado

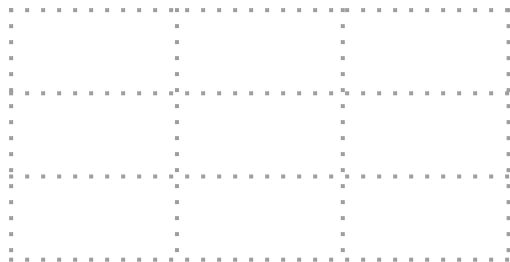


Partición de vistas



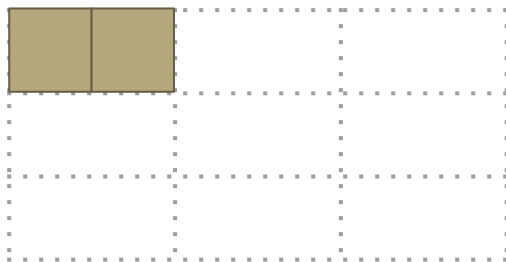
Partición de vistas

- Como se mencionó en clases anteriores, las particiones también pueden ser una sucesiva a otra



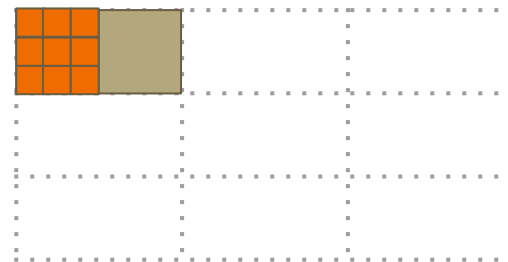
División inicial

Cada región muestra los resultados de encuestas por país.



Segunda división

Para cada país se segmentan los resultados entre fumadores y no fumadores.



Tercera división

Para cada grupo de personas se segmentan los resultados por rango de edad.

Partición de vistas

Sistema HIVE

Permite cambiar el orden de las particiones.

Forma 1:

- Primera división: 4 categorías distintas.
- Segunda división: 33 categorías distintas.
- Última división: 108 categorías distintas y dicha región se pintó según un atributo cuantitativo.



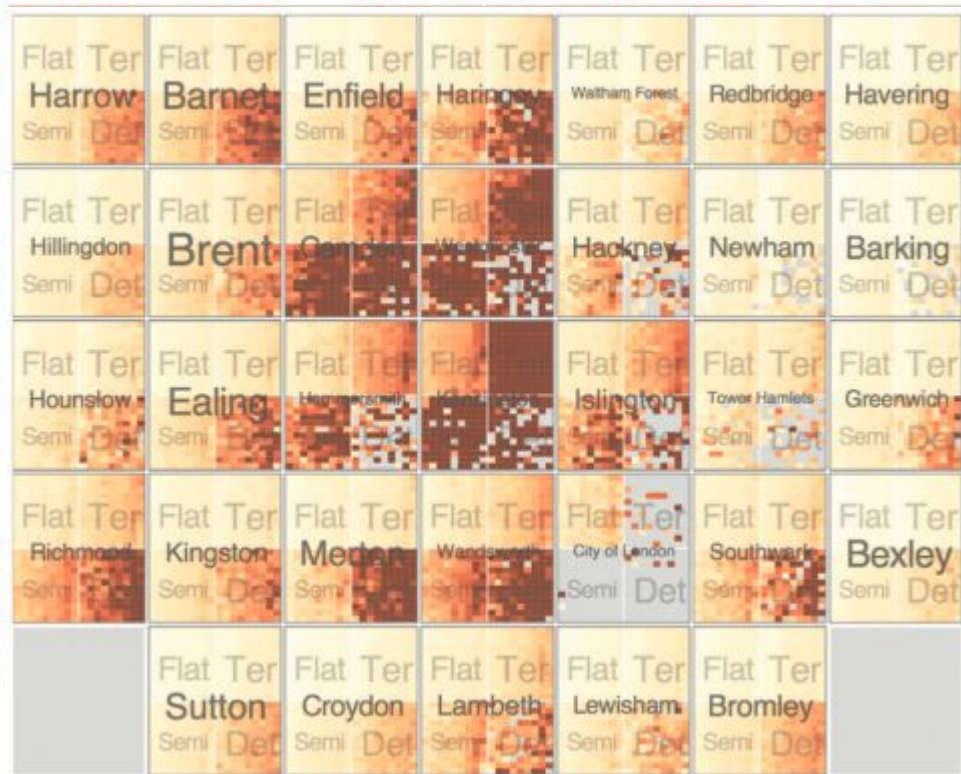
Partición de vistas

Sistema HIVE

Permite cambiar el orden de las particiones.

Forma 1:

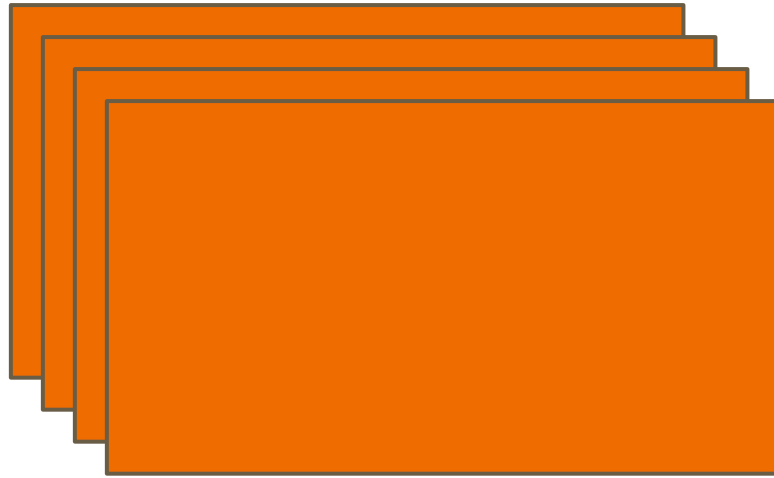
- Primera división: 33 categorías distintas.
- Segunda división: 4 categorías distintas.
- Última división: 108 categorías distintas y dicha región se pintó según un atributo cuantitativo.



Superposición

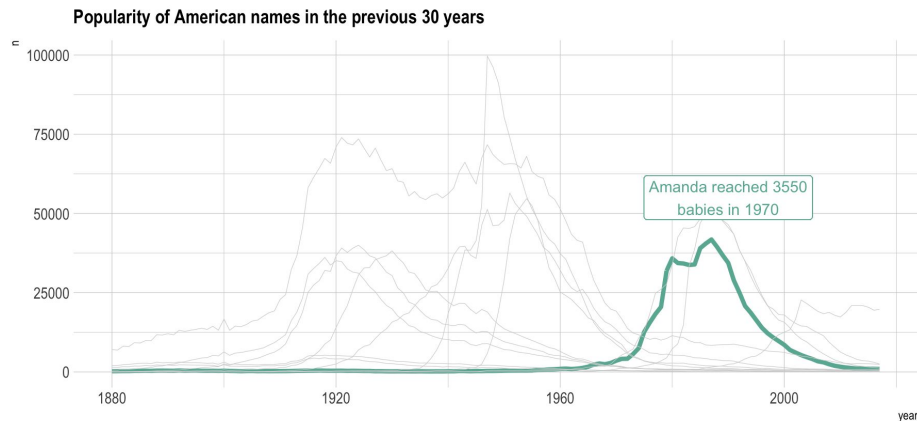
Superposición

- Generar y colocar capas una sobre la otra para producir una sola vista.
- Esta opción sirve como alternativa a la yuxtaposición de vistas separadas lado a lado.



Superposición

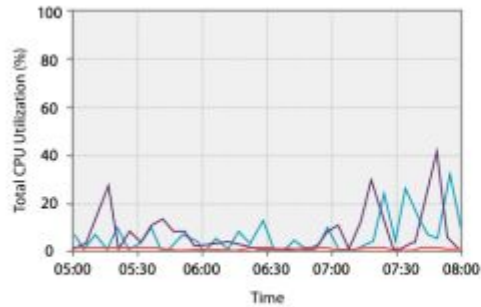
- Generar y colocar capas una sobre la otra para producir una sola vista.
- Esta opción sirve como alternativa a la yuxtaposición de vistas separadas lado a lado.
- Ideal cuando quieres resaltar una marca, pero mantener más información para complementar la visualización



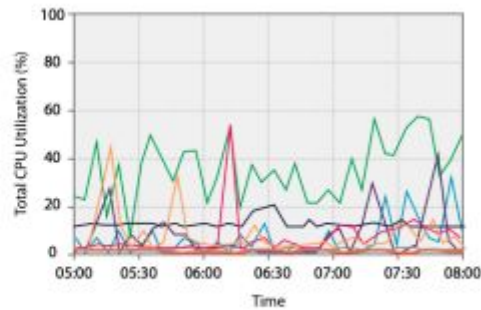
Superposición

Gráfico de múltiples líneas

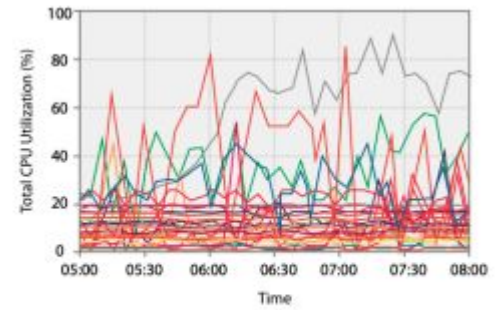
- Cada línea es una capa distinta.
- 🙄 con la intercepción de líneas y la cantidad de estas. No caer en el Spaghetti plot.



(a)



(b)

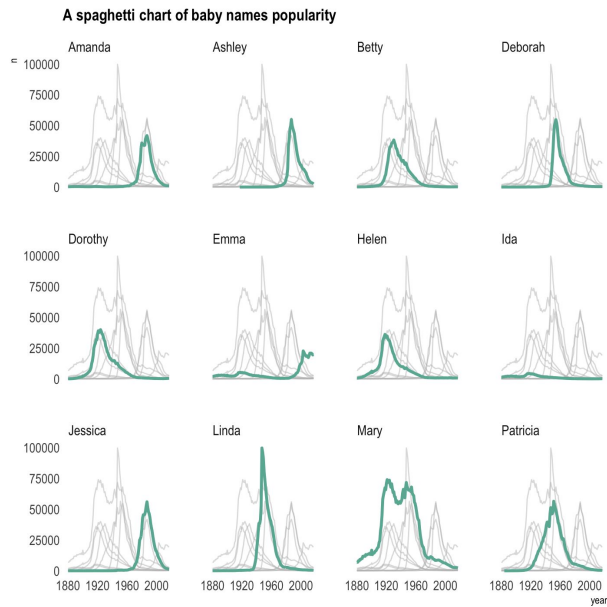
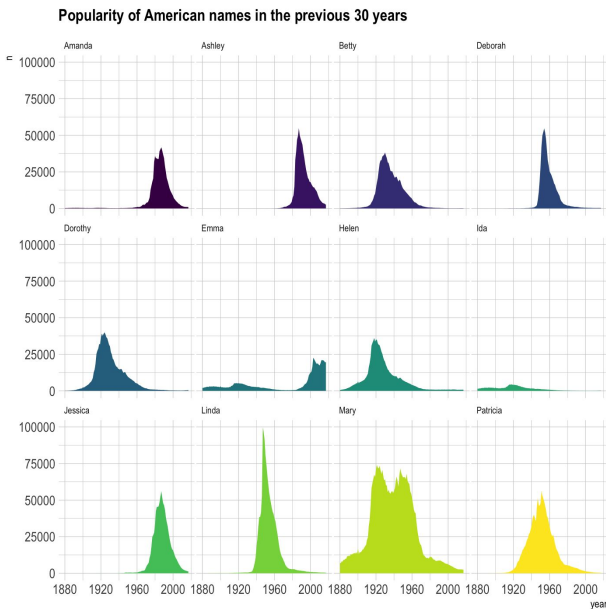
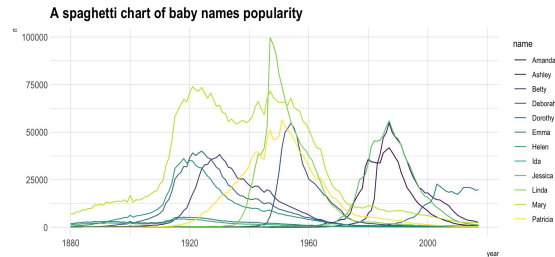


(c)

Superposición

Gráfico de múltiples líneas

- Validar si no es mejor recurrir a un *small multiples*.



Superposición

Gráfico de múltiples líneas

- Estudio empírico de múltiples líneas vs small múltiples de gráficos de área.
 - Múltiples líneas es útil para comparaciones locales.
 - Small múltiple es útil para comparaciones que implican la globalidad del rango horizontal.



(a)

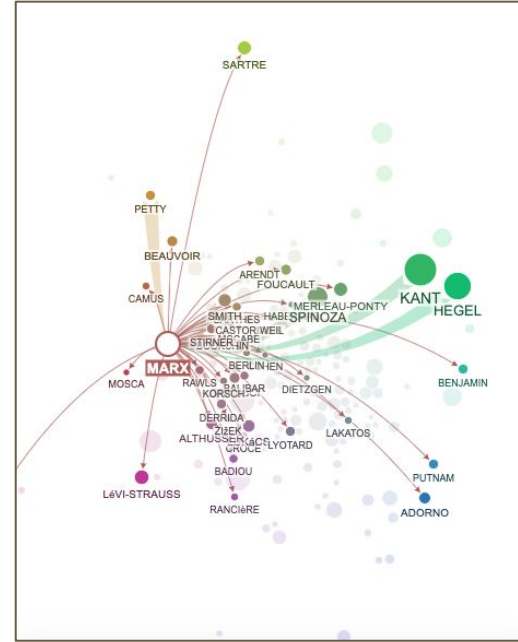


(b)

Superposición

Superposición de capas dinámicas

- Muchas veces es útil dejar que el usuario seleccione qué capa observar



Superposición

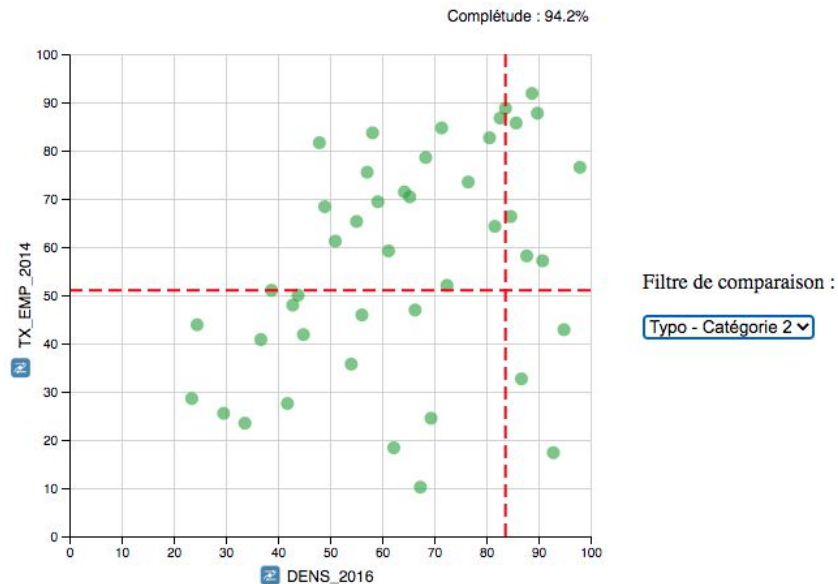
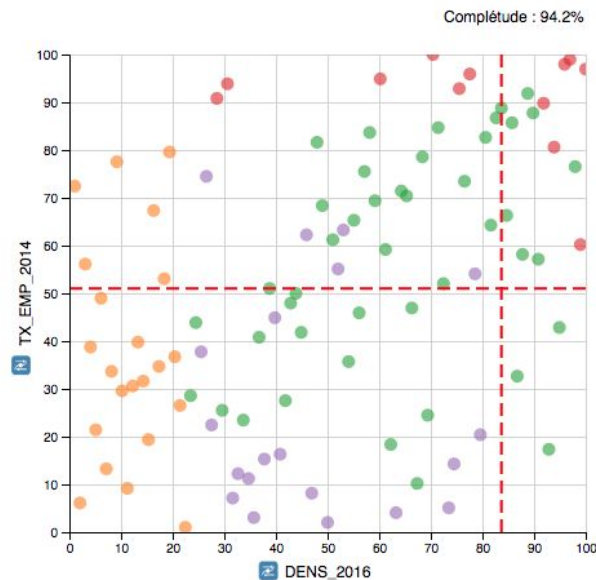
Recomendaciones

- Permitir la elección del usuario sobre las capas a visualizar deja espacio para un mayor número de capas en total.
- Revisar caso a caso si la superposición es una decisión adecuada. Lo esperado es que cada capa pueda ser distinguible.
- Verificar qué canal utilizar para distinguir las capas: ¿color? ¿forma? ¿tamaño?.

Filtrar

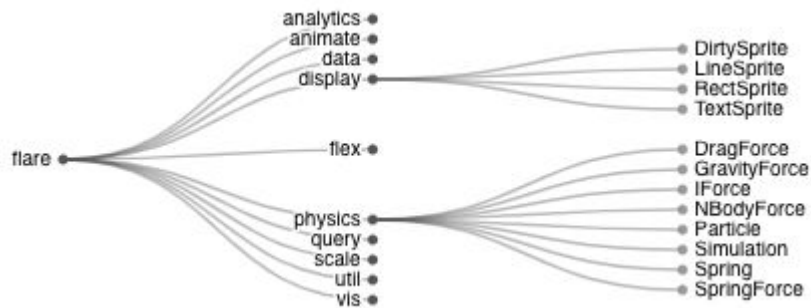
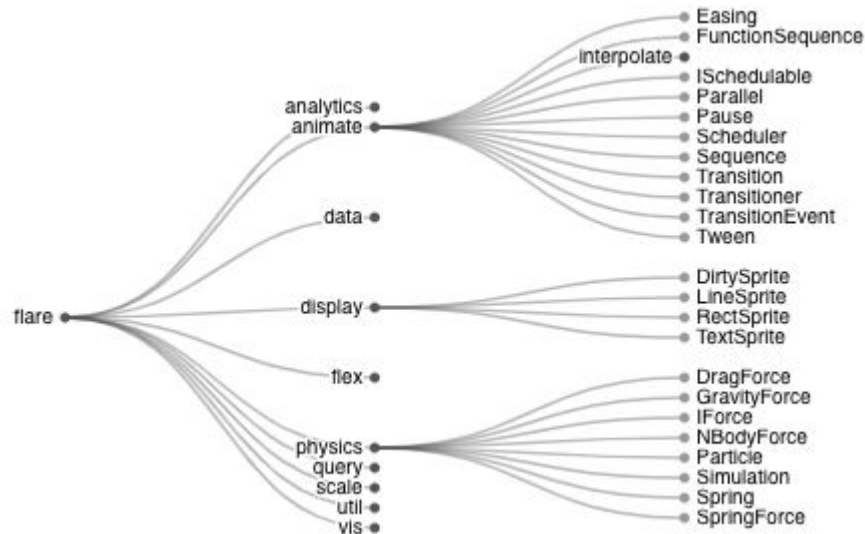
Filtrar

- Reducir el total de elementos visibles mediante **eliminación de parte de ellos**.



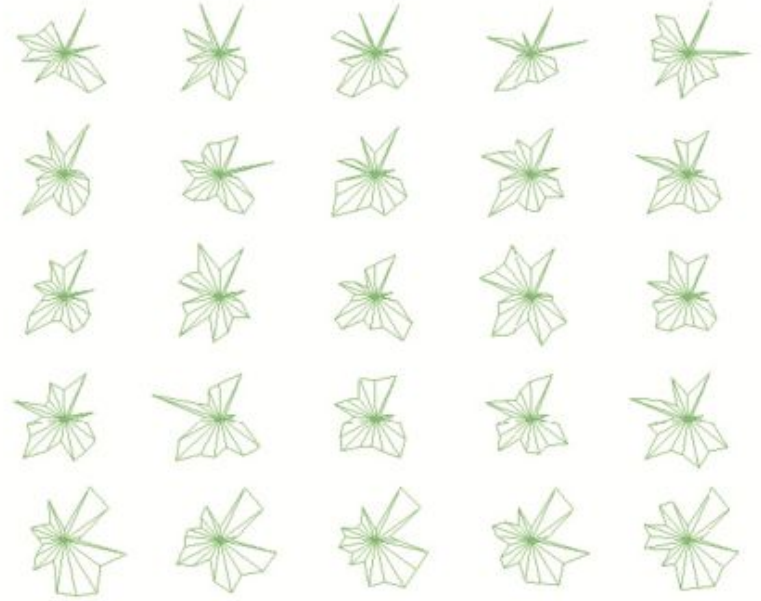
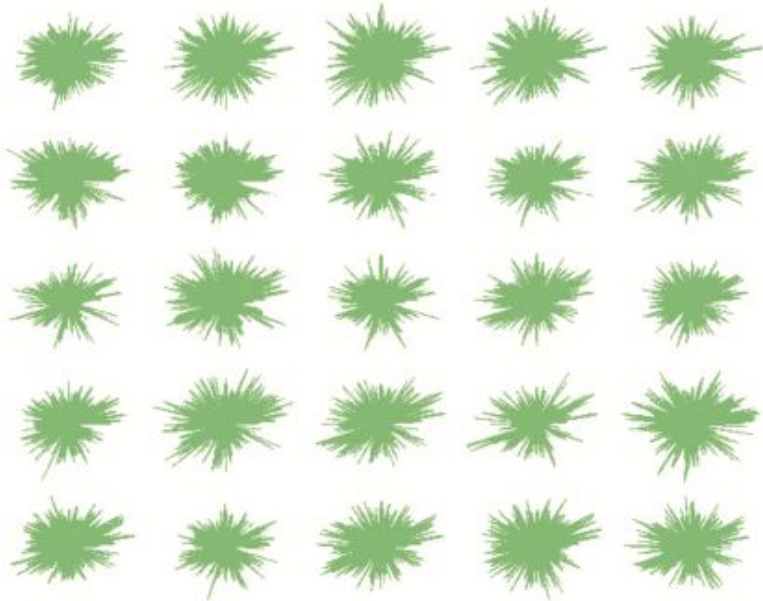
Filtrar

- Reducir el total de elementos visibles mediante **eliminación de parte de ellos**.



Filtrar

- Reducir el total de elementos visibles mediante **eliminación de parte de ellos**.
- También puede ser aplicado a atributos. Cada línea representa el valor de un atributo. Cada estrella es un ítem.



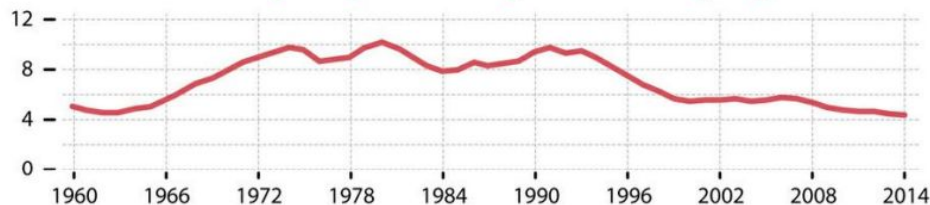
Filtrar

- La idea de filtración es bastante simple.
- 🙄 Recordar que en la mayoría de los casos, el usuario no conoce el dataset.
 - No sabe cuales son todos los atributos disponibles.
 - No sabe el rango de los atributos numéricos.
- El reto aparece al diseñar un sistema de visualización donde **la filtración permite explorar el dataset de forma efectiva.**

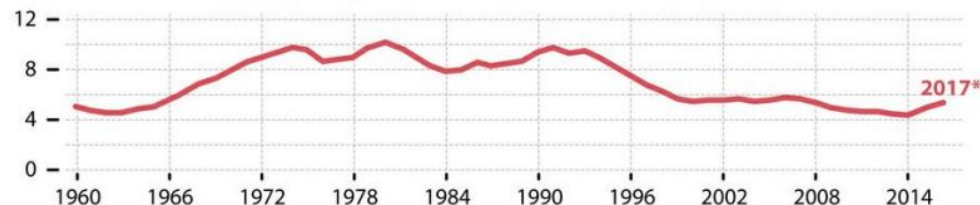
Filtrar 🙄

- **Cuidado con qué datos se van a filtrar.** Puede generar conclusiones diferentes.
- Este ejemplo del 2018, inicialmente se filtraron los datos hasta 2014 para indicar que había una tendencia a la baja en la cantidad de asesinatos.... pero faltaban los datos de los últimos 3 años, lo cual ya no reflejaba una tendencia total a la baja.

U.S. murder rate (yearly murders per 100,000 people)



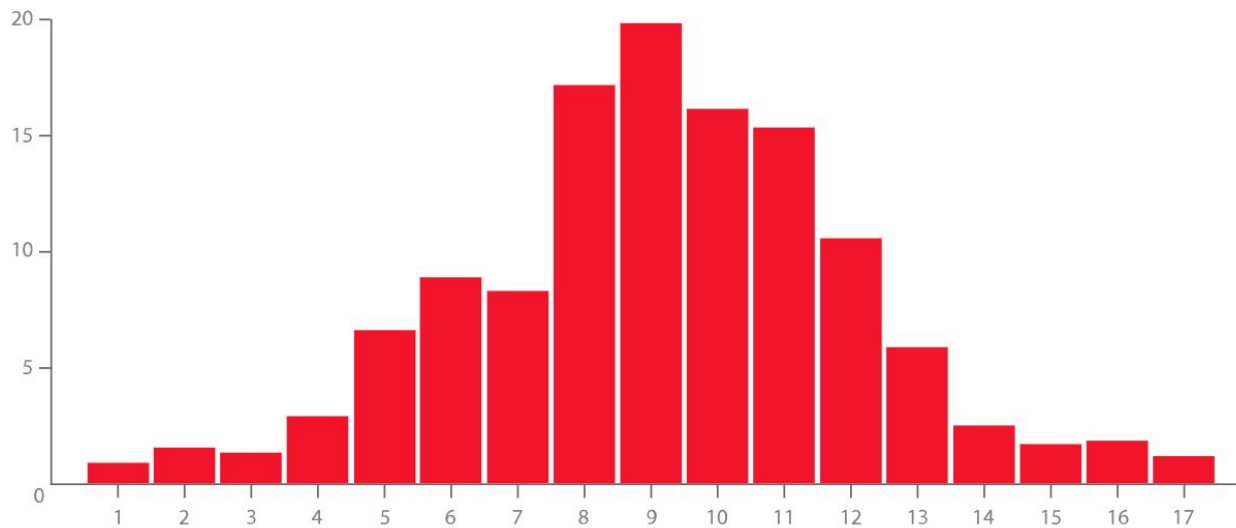
U.S. murder rate (yearly murders per 100,000 people)



Agregación

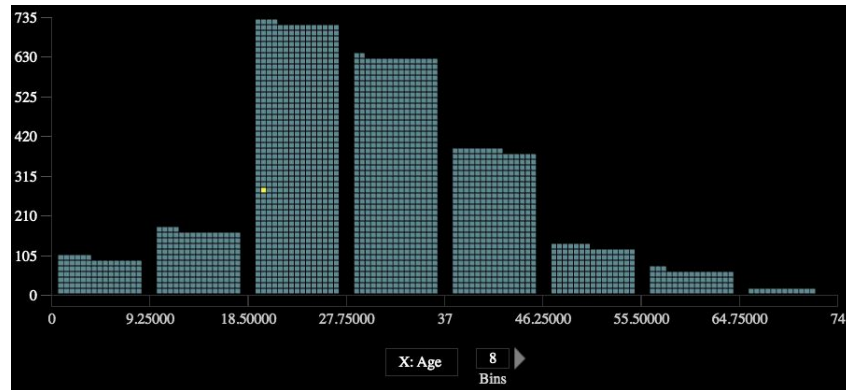
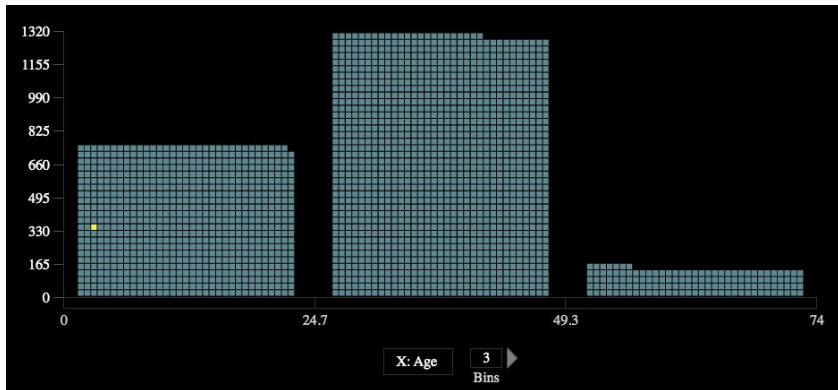
Agregación

- Reducir el total de elementos visibles mediante **agrupación y reemplazo de elementos**.
- Suele relacionarse con operaciones de agregación como promedio, conteo, mínimo y máximo.



Agregación

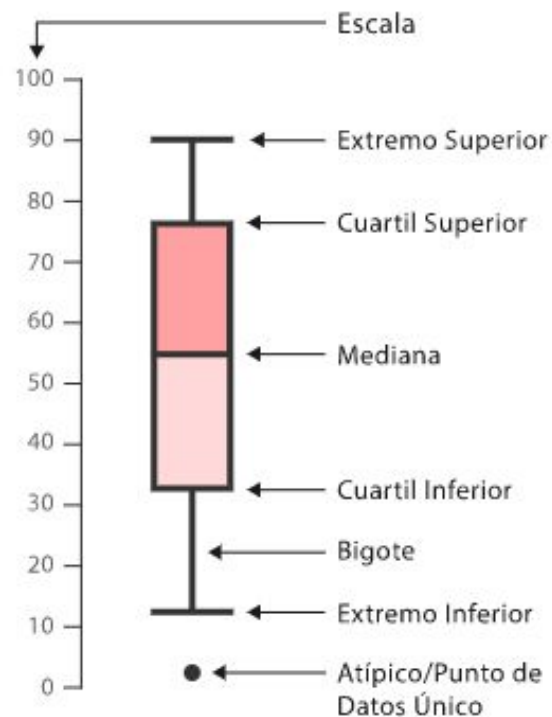
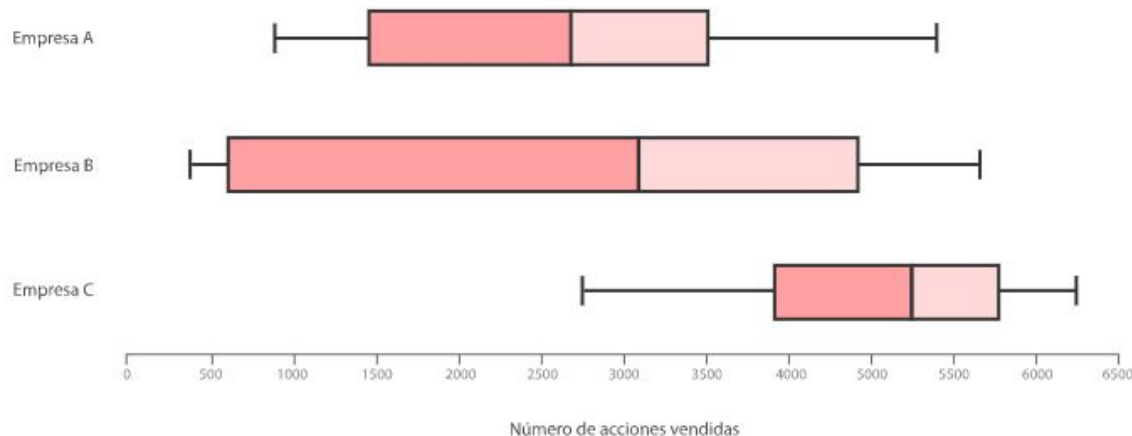
- Reducir el total de elementos visibles mediante **agrupación y reemplazo de elementos**.
- Suele relacionarse con operaciones de agregación como promedio, conteo, mínimo y máximo.
- Podemos apoyarnos de otros aspectos visuales, como la proximidad, para dar la sensación de agrupamiento sin reemplazar los elementos.



Agregación

Gráfico de caja o bigote

- Comúnmente ocupado para entregar un resumen estadístico agregado de los valores que presenta un atributo cuantitativo



Agregación

Variaciones

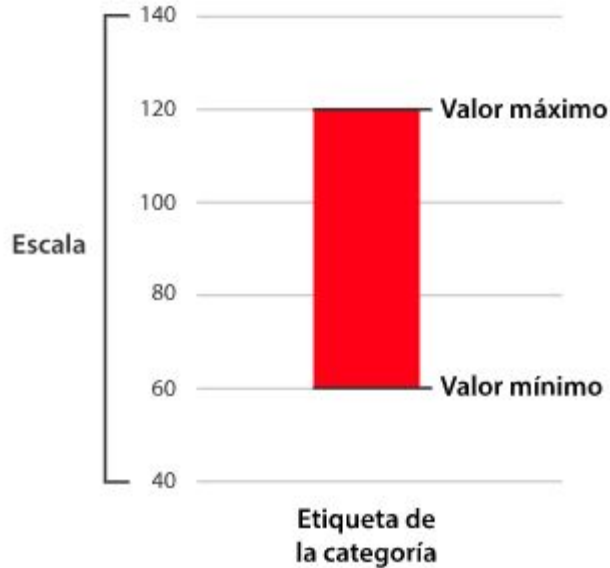


Gráfico de barras flotantes

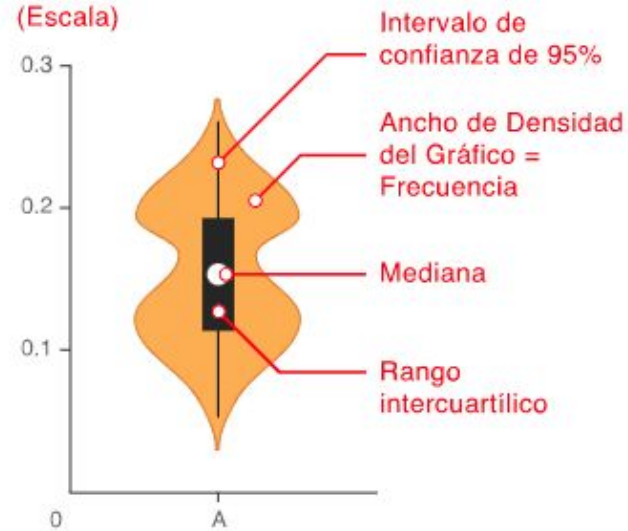
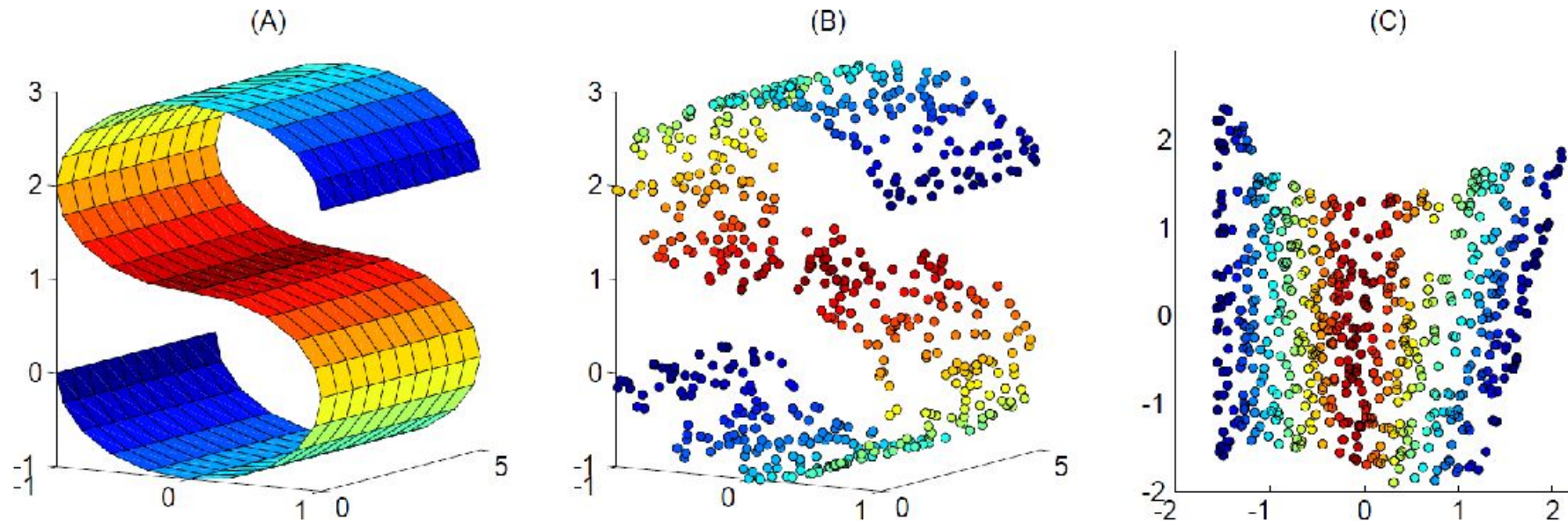


Gráfico de violín

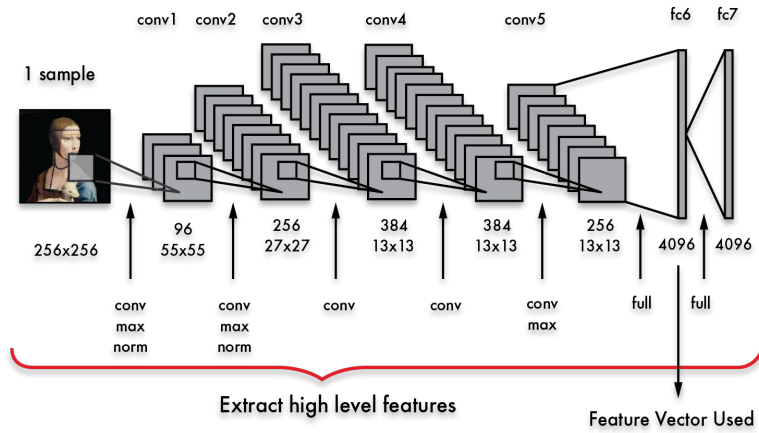
Reducción de dimensionalidad

Reducción de dimensionalidad

- Transformación de datos multidimensionales a datos de pocas dimensiones sin perder las caracterizaciones originales.
- Son un campo de estudio importante en *data science* y estadística.



Reducción de dimensionalidad

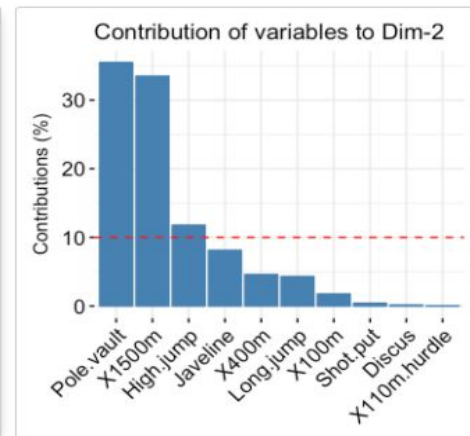
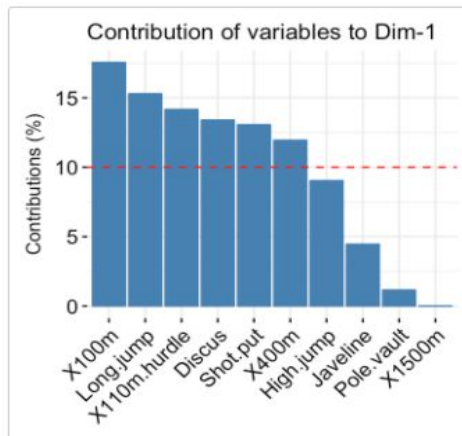
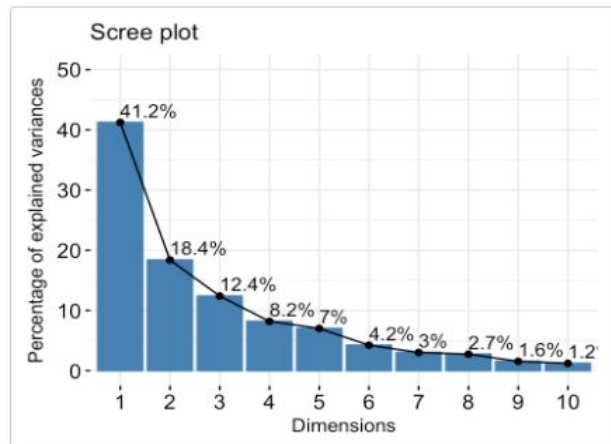


De 4.096 dimensiones
a 2 dimensiones

Reducción de dimensionalidad

🤔 ¿Cómo funcionan estas "dimensiones"?

- En conjunto, todas las nuevas dimensiones tratan de guardar la misma información que el ítem en su alta dimensionalidad.
- En varios algoritmos, estas nuevas dimensiones se construyen a partir de combinar información de distintos atributos originales.



Reducción de dimensionalidad

Algoritmos existentes:

- PCA
- UMAP
- T-SNE
- Entre otros ...

Reducción de dimensionalidad

Algoritmos existentes:

- PCA
- **UMAP**
- **T-SNE**
- Entre otros ...

Usando visualización para entender más algunos algoritmos:

- [How to Use t-SNE Effectively](#)
- [Understanding UMAP](#)

Reducción de dimensionalidad

Algoritmos existentes:

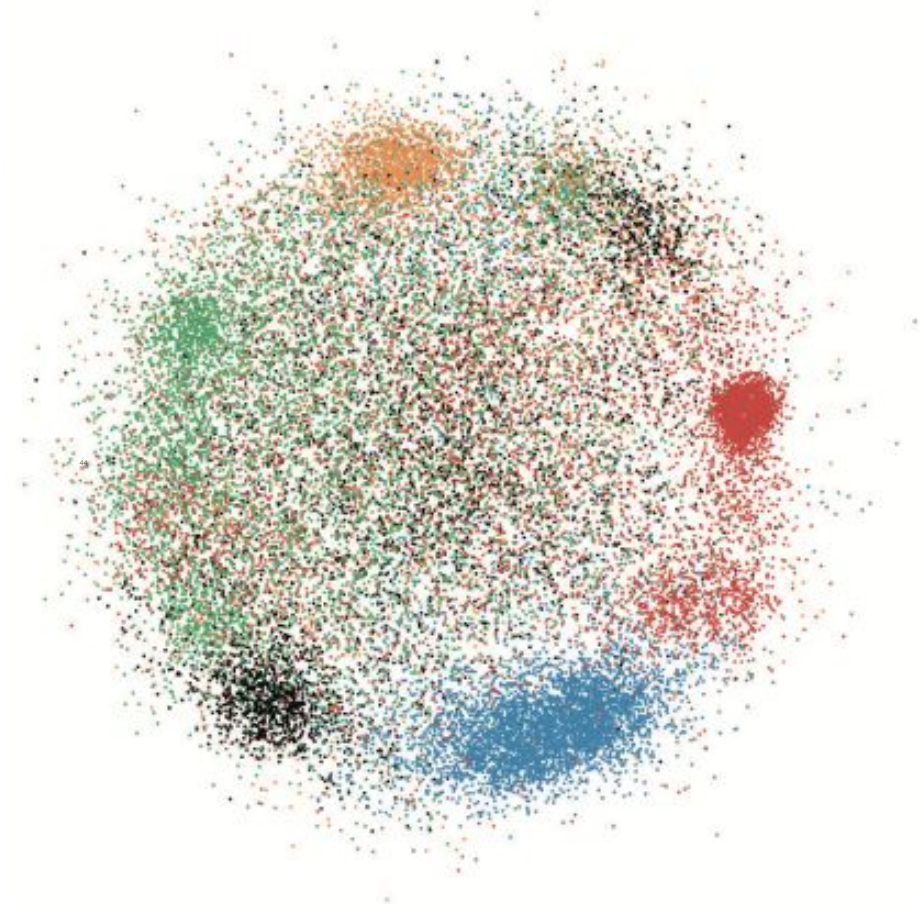
- PCA
- UMAP
- T-SNE
- Entre otros ...

En visualización se ocupan muchas veces para mostrar espacialmente y en dos dimensiones un *dataset*. De tal forma que ítems próximos realmente sean similares según sus todos sus atributos originales, y qué ítems lejanos sean realmente diferentes.

Reducción de dimensionalidad

Casos de aplicación - Libros similares

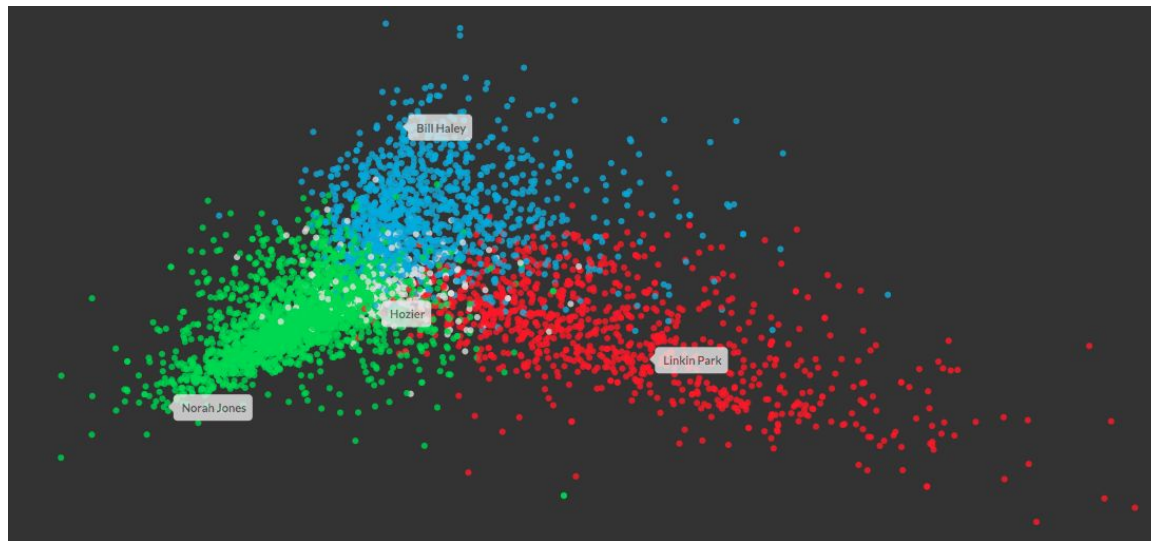
- Un documento se puede describir por el conteo de todas las palabras que contiene. Cada palabra distinta es una dimensión.
- Se utiliza reducción de dimensionalidad para obtener un vector de (x,y) por documento.



Reducción de dimensionalidad

Casos de aplicación - Cantantes similares

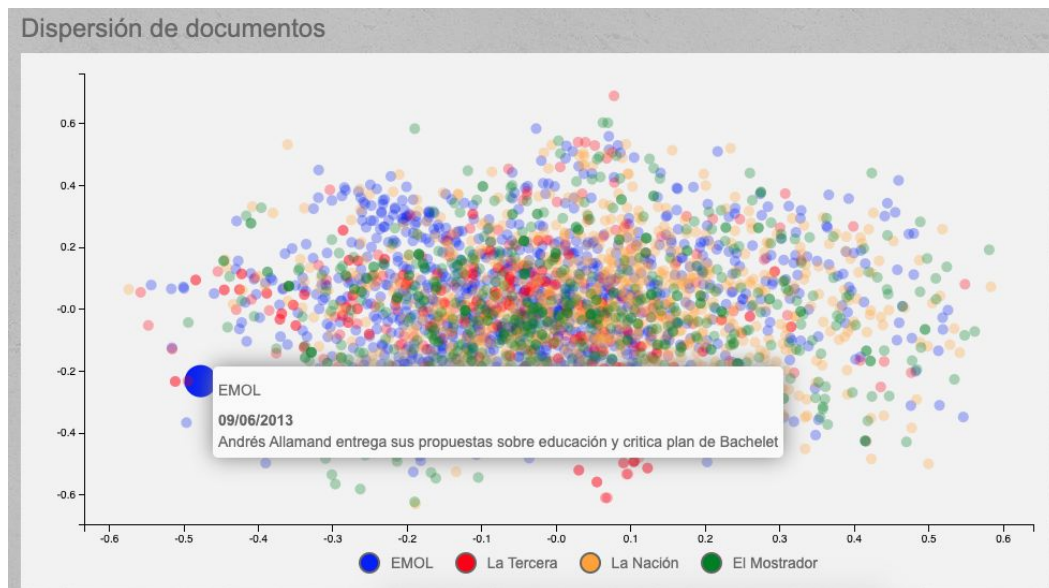
- Cada cantante se representa con información de sus canciones.
- Se construye un vector (x, y) por cantante mediante reducción de dimensionalidad.



Reducción de dimensionalidad

Casos de aplicación - Noticias elecciones 2013

- Cada noticia representa un nodo. Su color presenta quién lanzó la noticia.
- Se utiliza reducción de dimensionalidad para obtener un vector de (x,y) por noticia.



Reducción de dimensionalidad (Bonus)

Reconstrucción de datos luego de aplicar reducción

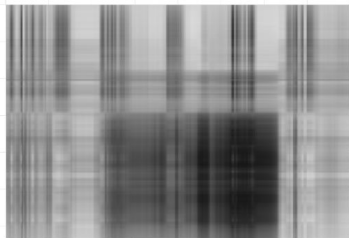


Imagen de 800x600
(480000 pixels)

Reducción de dimensionalidad (Bonus)

Reconstrucción de datos luego de aplicar reducción

First 2 Components



First 3 Components



First 10 Components



First 20 Components



First 4 Components



First 5 Components



First 50 Components



First 100 Components



Imagen de 800x600
(480000 pixels)

Antes de salir... Revisión de contenidos (RC)

Se acaba de publicar un mini control de alternativas en Canvas sobre lo que **vimos en la clase de hoy**.

- **Duración:** 2 semanas (sin considerar semana de receso) para realizarlo a partir de hoy. Una vez terminado el plazo, tendrán retroalimentación por cada pregunta.
- **Intentos para responder:** ilimitados.
- **Extensión:** 7 preguntas de 1 punto c/u.
- **Condición para obtener el punto RC:** Al menos 6 puntos de 7.
- Cada vez que respondan, verán el puntaje total logrado, pero no cuáles preguntas están correctas e incorrectas.

Próximos eventos

Próxima clase

- Brush y agregación de datos con D3.
- Traer *notebook* 

Ayudantía de mañana

- Sala de ayuda para la T3

Tarea 3

- Se entrega el próximo jueves a las 20:00.

IIC2026

Visualización de Información

— Hernán F. Valdivieso López —
(2023 - 1 / Clase 12)
