

Visualización de información y analítica visual

Clase 8: Reducción de datos / Focus+Context



Cinco maneras de manejar las complejidades visuales

1. Derivación de nuevos datos
 2. Manipulación de vista
 3. Múltiples vistas
 4. Reducción de datos
 5. Focus and context
- Clase 3
- Clase 7



Reducción de datos



Reducción de datos

Llamaremos como reducción de datos a aquellas decisiones de diseño que impliquen reducir la cantidad de información que mostramos en la visualización

- Esta decisión de diseño puede aplicarse tanto a ítems como a atributos
- Normalmente estos idioms se mezclan con cambios en vistas, seleccion y otras decisiones de diseño.

¿Por qué reducir?

Reducir es una manera de manejar complejidades en visualización

- En el **paradigma estático**, la reducción es solo *reducir* la información que el usuario ve
- En el **caso dinámico**, podemos ir controlando gradualmente la cantidad de información que mostramos y esta decisión es **bidireccional**: desde un estado inicial podemos mostrar menos información o más información



Nos vamos a enfocar principalmente en casos dinámicos, ya que son más interesantes.

¿Por qué reducir?

Reducir los datos puede parecer una manera muy obvia y tonta de manejar las complejidades de la visualización...



Si tu visualización se vuelve muy compleja porque muestra mucha información, entonces prueba mostrando menos (duh?)



Pero lo que queremos realmente es que esta reducción no tenga un efecto nocivo en la tarea del usuario, es decir, queremos evitar esconderle información relevante al usuario

Dos maneras de hacer reducción:

1. Filtración 

2. Agregación 



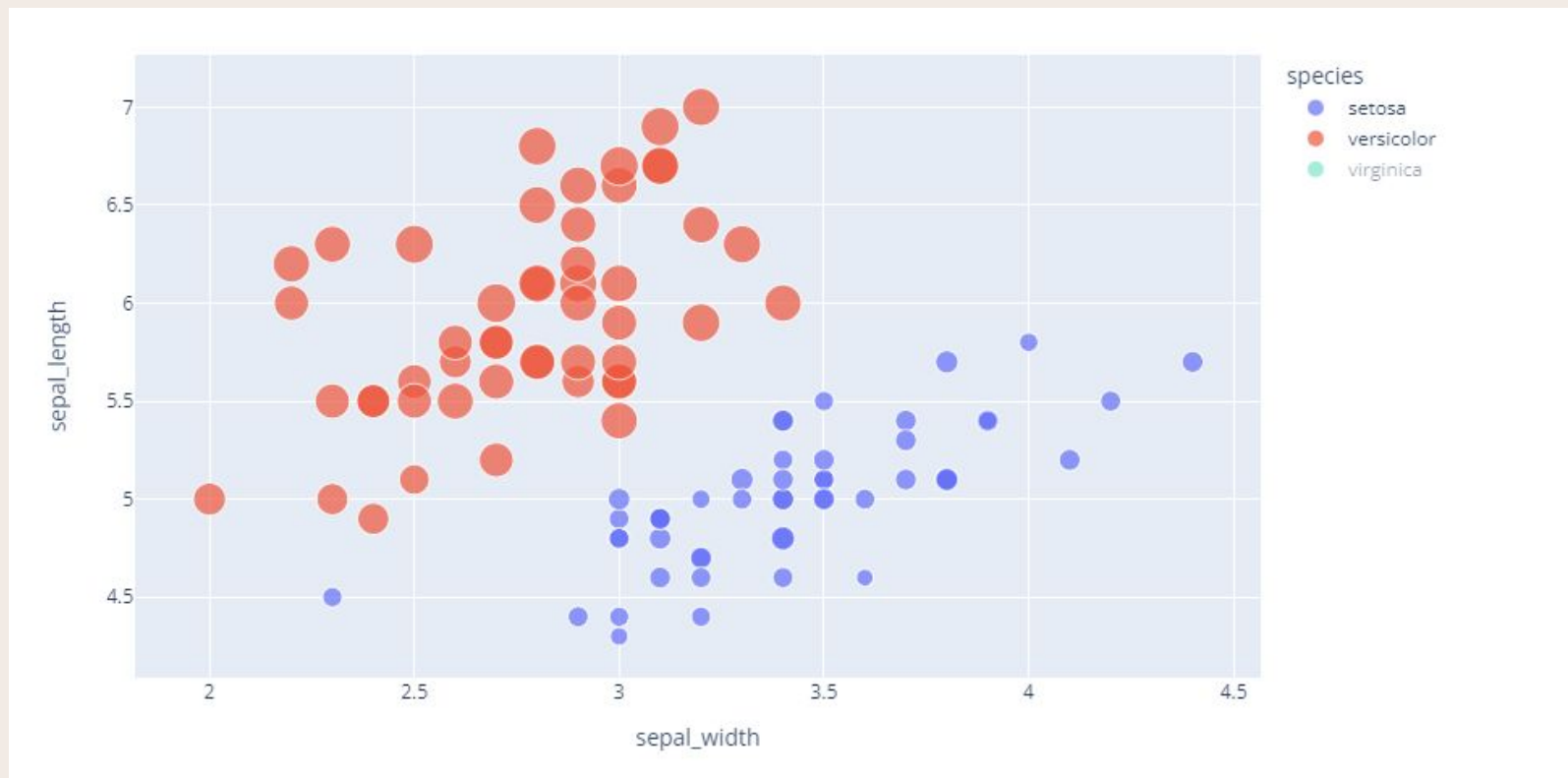
Filtros



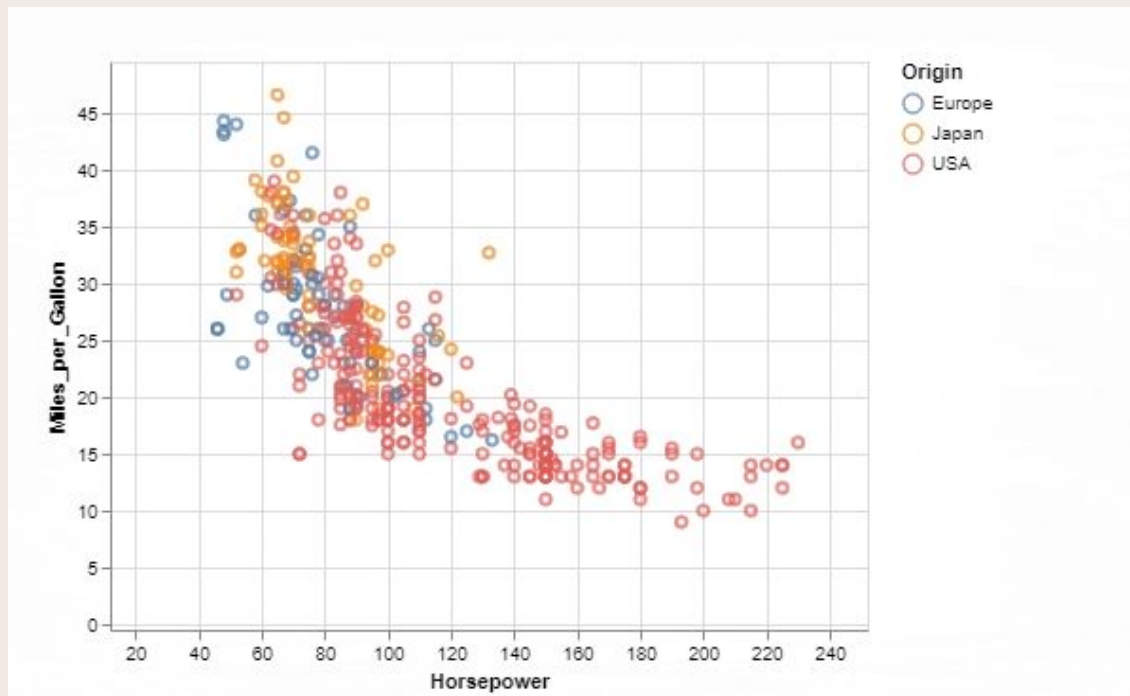
**La filtración reduce el número total de elementos
visibles ocultando un subconjunto de ellos**



Filtros: ejemplos



Filtros: ejemplos



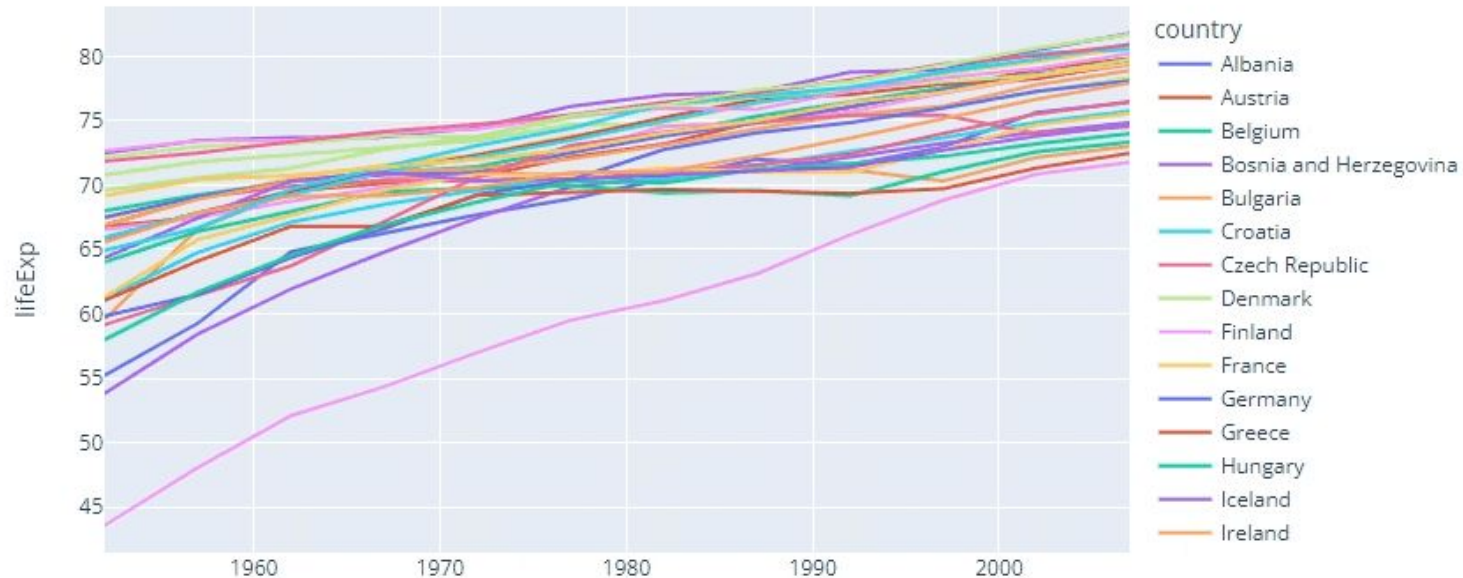
Filtración de ítems

La manera más básica de hacer filtración es simplemente omitir algunos ítems completamente de la visualización

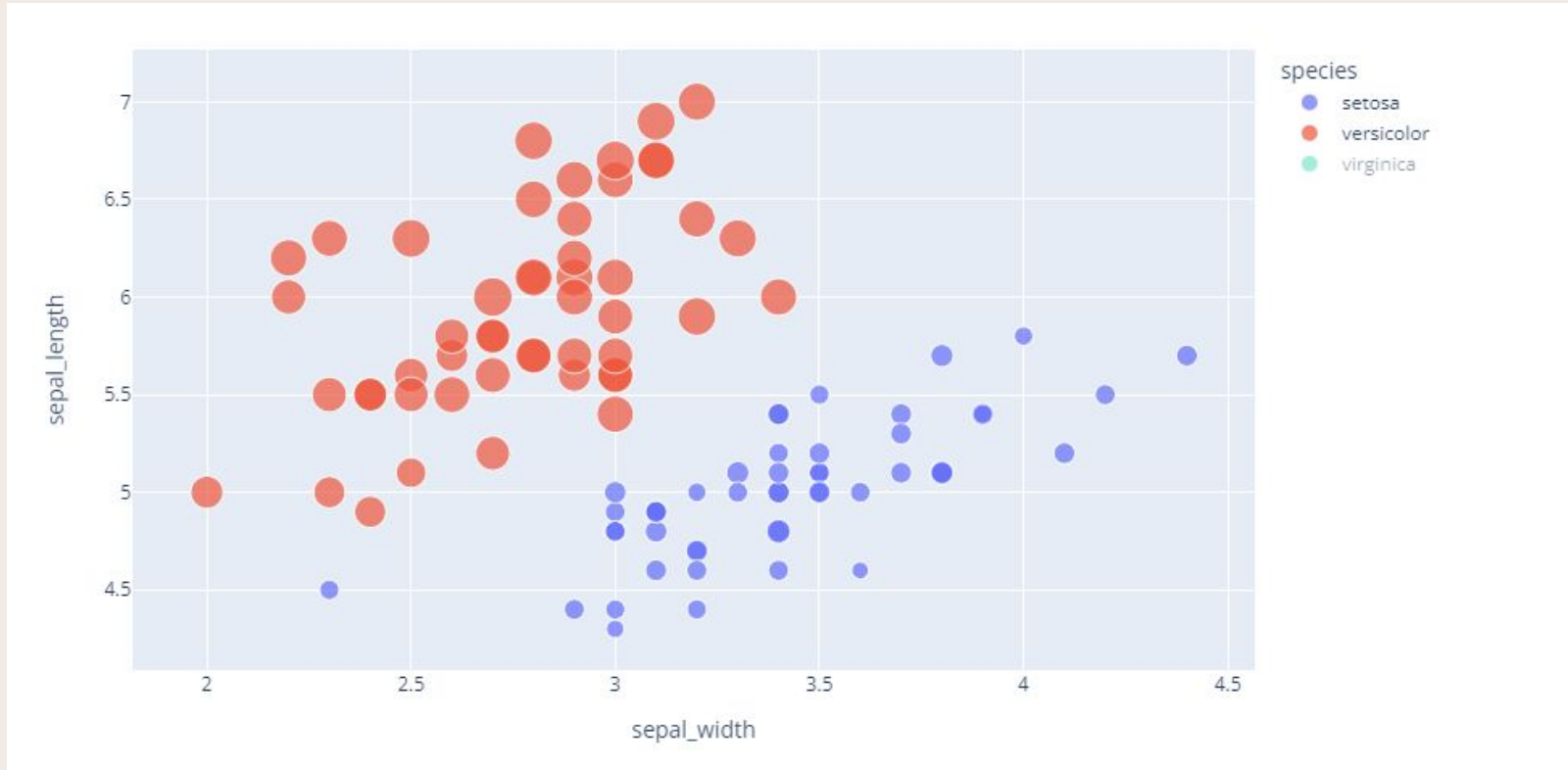
- Normalmente los ítems se filtran de acuerdo a algún criterio que puede ser su valor en alguno de sus atributos

Filtración de ítems: ejemplo

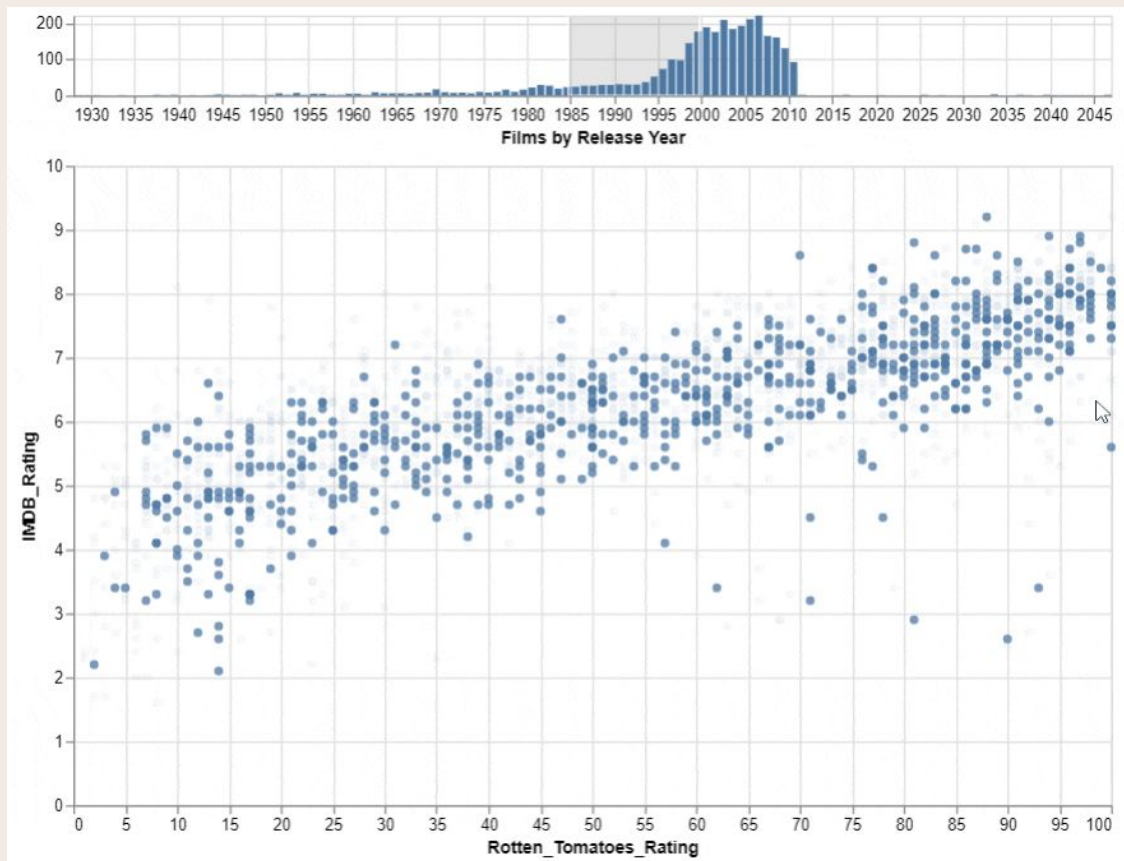
Life expentancy progression of countries per continents



Filtración de ítems: ejemplo



Filtración de ítems: ejemplo



Filtración de ítems

En términos de implementación, los filtros tienden a ser bastante sencillos (simplemente filtrar los datos)

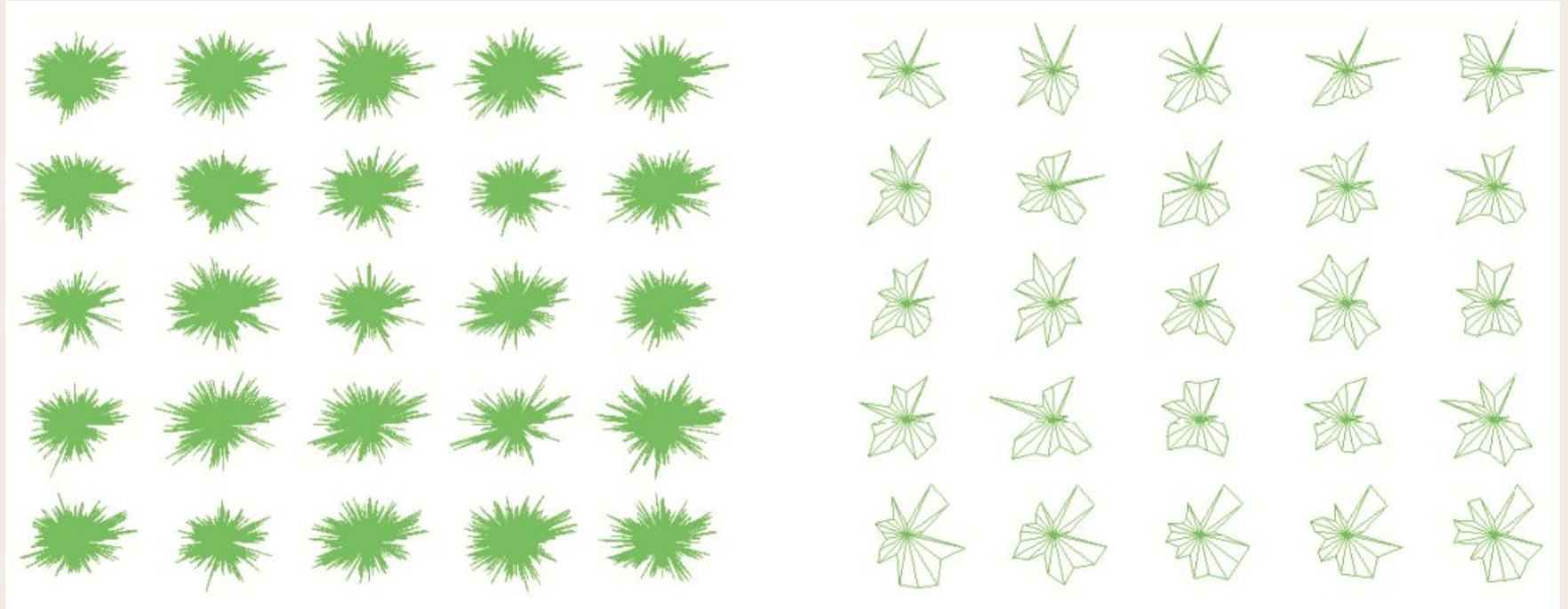
Sin embargo, para usuarios que desean **explorar/descubrir un dataset**, estos a priori no conocen los distintos valores de los elementos...

En estos casos se puede buscar herramientas de interacción que no presupongan un conocimiento previo de parte del usuario

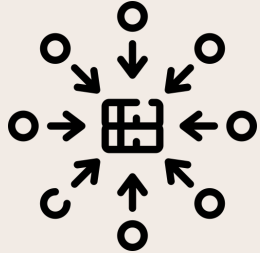
Filtración de atributos (o de dimensionalidad)

Otra manera de hacer filtro es reducir la cantidad de atributos que se muestran en una visualización

Filtración de atributos (o de dimensionalidad)



Agregación



**La agregación reduce el número total de elementos
visibles mediante agrupación y reemplazo de
elementos**



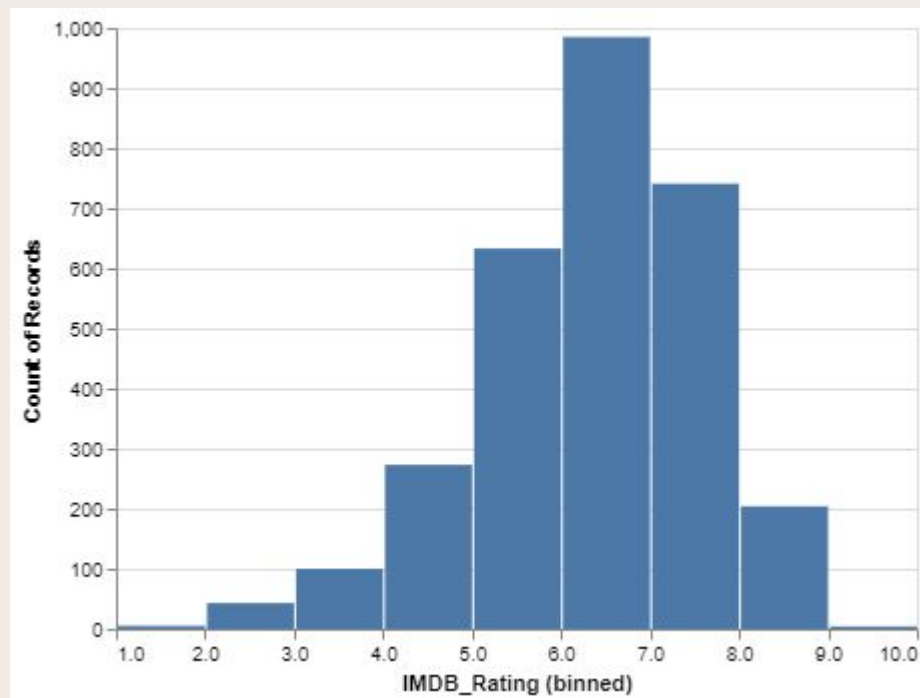
Agregación

Existen varias estrategias para agrupar datos, pero la mayoría usa datos derivados del tipo agregados, tales como:

- promedio
- conteo
- mínimo
- máximo

 Estas decisiones inevitablemente hacen que se pierda información

Histograma



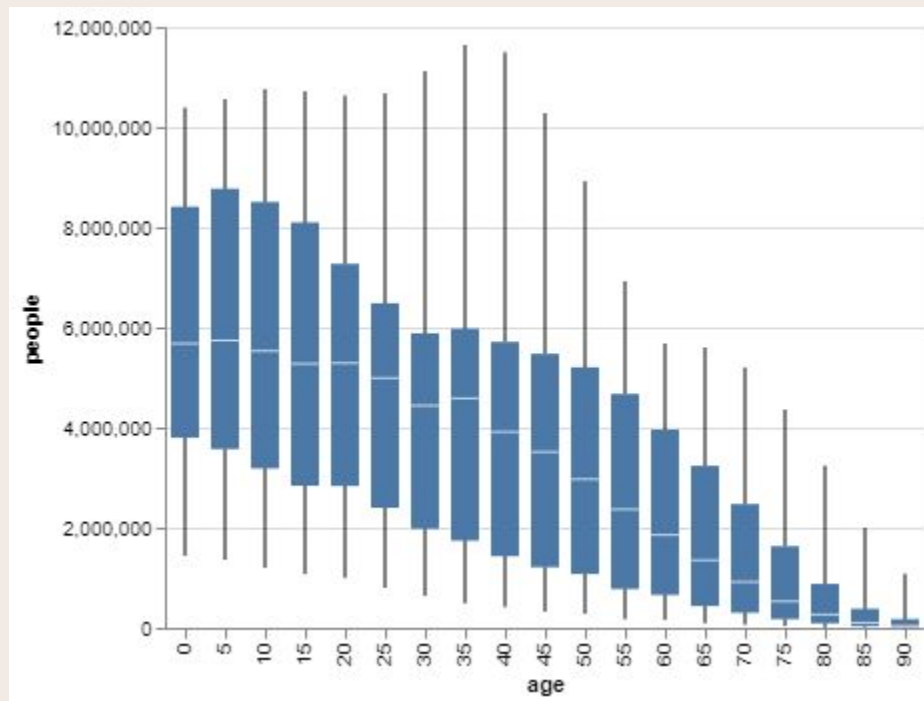
https://altair-viz.github.io/gallery/simple_histogram.html

Histograma

Idiom	Histograma
Qué: datos	Datos tabulares: un atributo cuantitativo
Qué: datos derivados	Tabla derivada: un atributo llave ordenado (bins), un atributo cuantitativo (conteo por bin)
Cómo: codificación	Layout rectilíneo. Marcas de línea con posición alineada para expresar el valor derivado. En la posición en el eje x se codifica el atributo llave
Por qué: tarea	Descubrir distribución de datos, outliers.



Boxplot



Boxplot

Idiom	Boxplot
Qué: datos	Tabla atributos cuantitativos
Qué: datos derivados	Indicadores para los atributos cuantitativos originales, que representan su distribución
Cómo: codificación	Un glifo por atributo original, alineación 1D de un glifo por atributo que agrupa muchos ítems.
Cómo: reducción	Agregación de ítems a través de indicadores
Por qué: tarea	Caracterizar la distribución, encontrar outliers, valores extremos, promedios, identificar skew
Escala	Docenas de atributos, ítems (ilimitados)

Agregación espacial

Cuando los datos son espaciales, podemos hacer agregación tomando en cuenta esta naturaleza espacial de los datos.

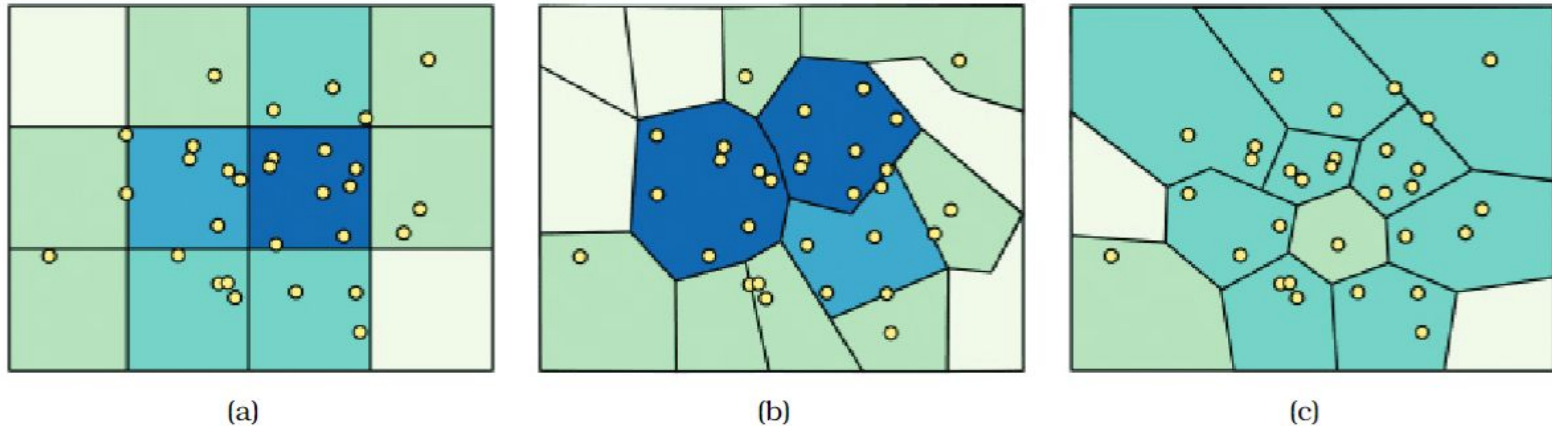


Figure 13.10. Modifiable Areal Unit Problem (MAUP) example, showing how different boundaries for aggregation regions lead to very different visual patterns on choropleth maps. (a) Central region is high density. (b) Central region is medium density. (c) Central region is low density. From http://www.e-education.psu.edu/geog486/l4_p7.html, Figure 4.cg.6.

Resumen: filtros y agregación

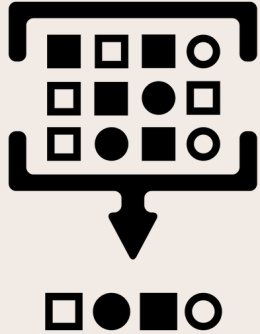
Filtros

- Algunos elementos se ocultan
- Simple de entender
- Puede generar un sesgo (el usuario puede olvidarse que los elementos filtrados existen)

Agregación

- No se oculta ningún elemento
- El usuario debe entender el criterio bajo el cual se agregan los datos
- Se muestran todos los ítems/atributos, pero con menos detalle

Agregación de atributos



La agregación de atributos reduce el número total de elementos visibles mediante combinación de múltiples atributos reemplazados por un atributo representante

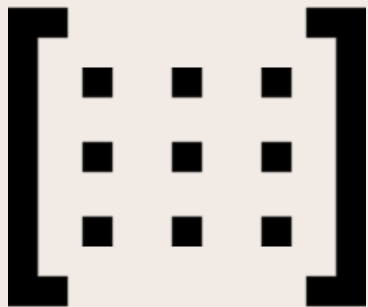


Agregación de atributos y reducción de dimensionalidad

La manera más simple de hacer agregación de atributos es agrupándolos en base a una medida de **similaridad** y tal vez calcular algún promedio por grupo

Hay técnicas especiales de reducción de dimensionalidad creadas especialmente para que la agregación de atributos no pierda información...

Reducción de dimensionalidad

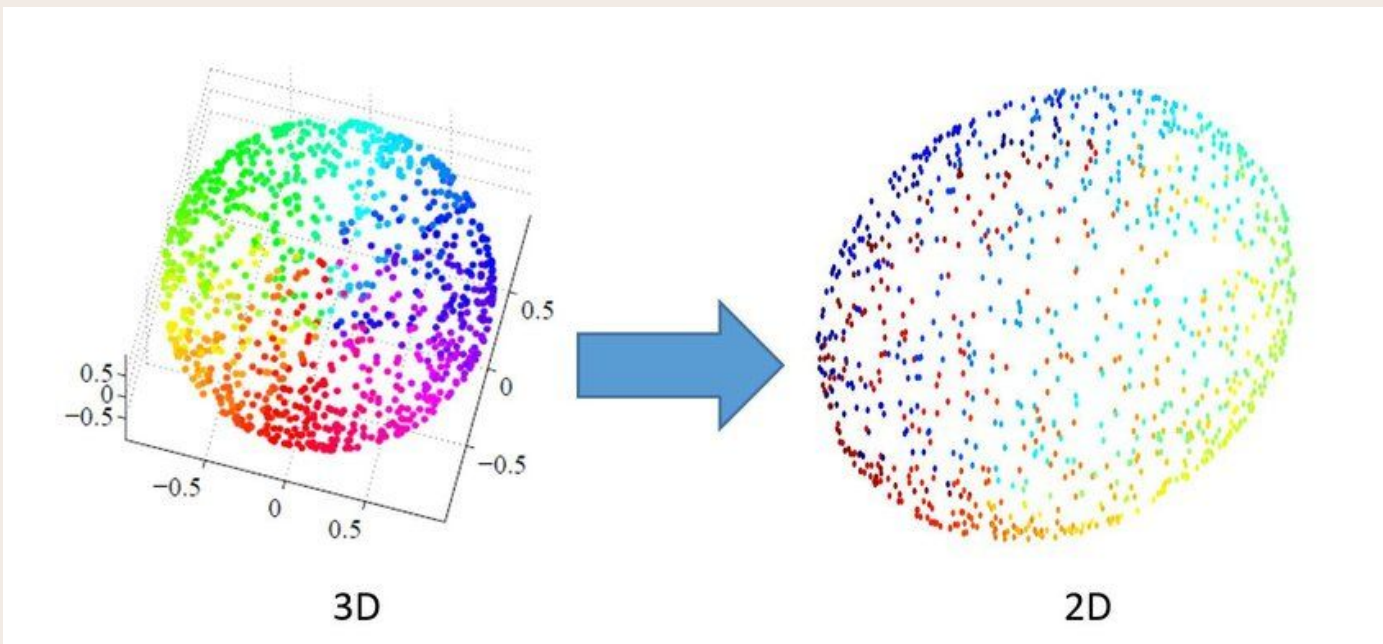


La reducción de dimensionalidad reduce el número total de elementos visibles mediante la transformación de datos multidimensionales a datos de pocas dimensiones que preservan las caracterizaciones originales

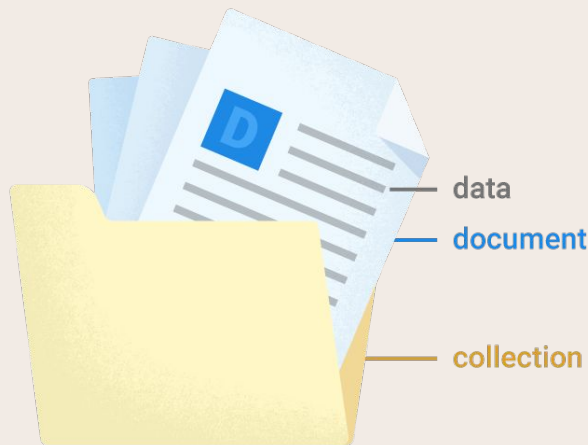


Reducción de dimensionalidad

Las técnicas de **reducción de dimensionalidad** son creadas y utilizadas en los campos de estadística y minería de datos



Reducción de dimensionalidad: ejemplo



Supongamos que tenemos una colección de miles de documentos.

Si bien la manera natural de tratar un documento es leerlo, cuando tenemos un gran volumen esto no es factible.

Los documentos no son directamente visualizables, primero tenemos que pasarlos a vectores de alta dimensionalidad, y después ver cómo visualizarlos

Reducción de dimensionalidad: ejemplo

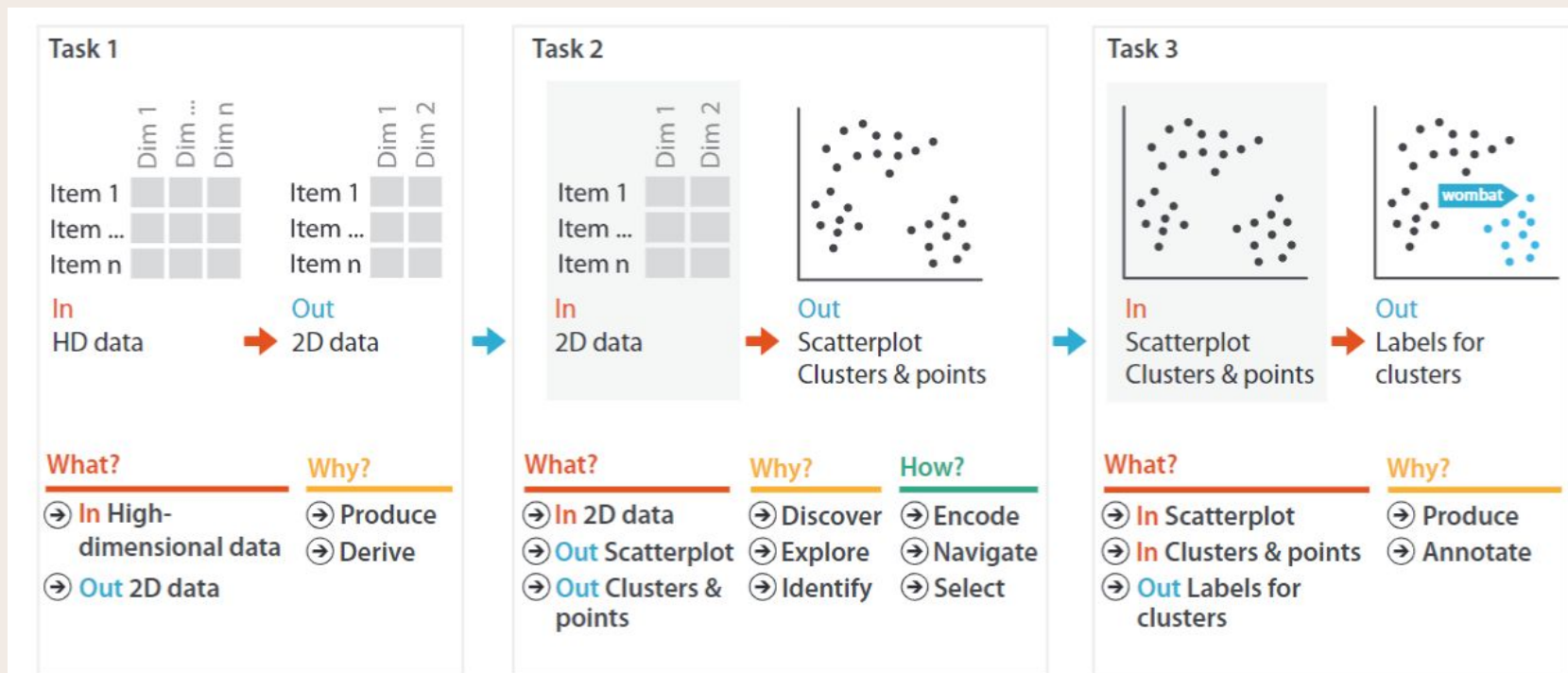


Figure 13.13. A chained sequence of what–why–how analysis instances for the scenario of dimensionality reduction of document collection data.

Reducción de dimensionalidad: ejemplo

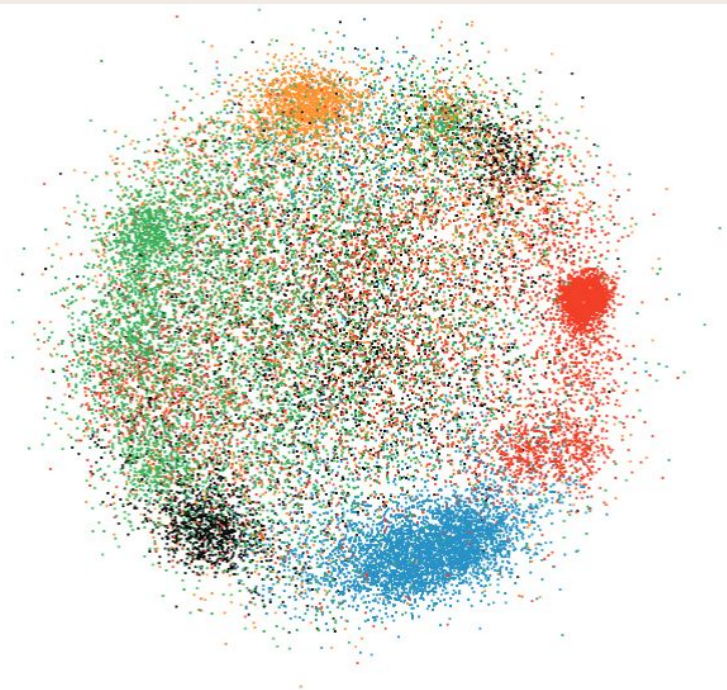


Figure 13.12. Dimensionality reduction of a large document collection using Glimmer for multidimensional scaling. The results are laid out in a single 2D scatterplot, allowing the user to verify that the conjectured clustering shown with color coding is partially supported by the spatial layout. From [Ingram et al. 09, Figure 8].

Reducción de dimensionalidad para documentos

Idiom	Reducción de dimensionalidad para documentos
Qué: datos	Colección de documentos de texto
Qué: datos derivados	Bag of words, 10.000 atributos. Tabla con 2 atributos, reducción de dimensionalidad
Cómo: codificación	Scatterplot, coloreado por clusters
Cómo: reducción	Agregación de atributos mediante reducción de dimensionalidad
Por qué: tarea	Encontrar clusters, visualizar elementos de alta dimensión
Escala	Atributos originales: 10.000, atributos derivados: 2. Items: cientos de miles

¿Cómo mostrar datos reducidos en dimensión?

En las técnicas clásicas de reducción de dimensionalidad, el usuario elige en cuantos atributos quedará su dataset.

- Cuando reducimos a dos atributos, entonces el scatterplot es la visualización estándar para visualizarlos
- Cuando hay más de dos atributos, podemos crear una matriz de scatterplots

¿Por qué scatterplots?

Si bien, el uso general de los scatterplots es para reconocer relaciones (o correlaciones) entre un par de atributos, este no es el uso que se les da cuando se usan para representar datos reducidos dimensionalmente.

De hecho, las técnicas de reducción de dimensionalidad están hechas de manera que los atributos resultantes **NO ESTÉN CORRELACIONADOS** (como por ejemplo PCA).

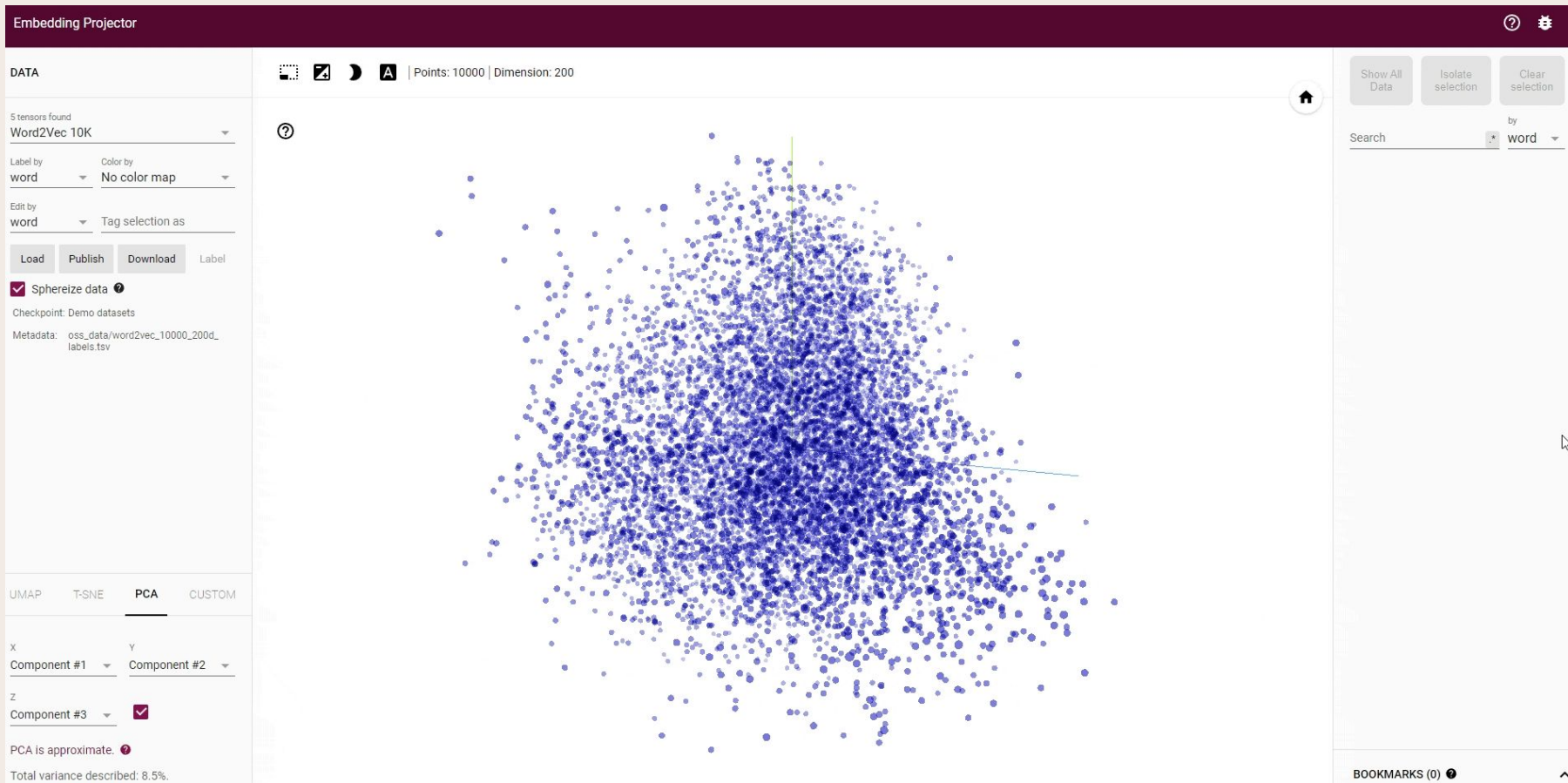
¿Por qué scatterplots?

En el caso de data reducida dimensionalmente, lo que buscamos es que items próximos sean similares (respecto de sus atributos originales) y que items lejanos sean distintos.

En este sentido, las tareas que podríamos querer realizar (y para las cuales son buenos los scatterplots) podrían ser

- Analizar la distribución de los datos
- Identificar clusters
- Identificar outliers
- Chequear que la reducción de dimensionalidad haya sido correcta

¿Por qué scatterplots?



Embed: Focus + Context



Embed: focus + context

Ahora veremos las decisiones de diseño que involucran *incrustar* (embed) información detallada de un conjunto de datos (focus) en una vista que también contenga información general acerca de los datos (context)



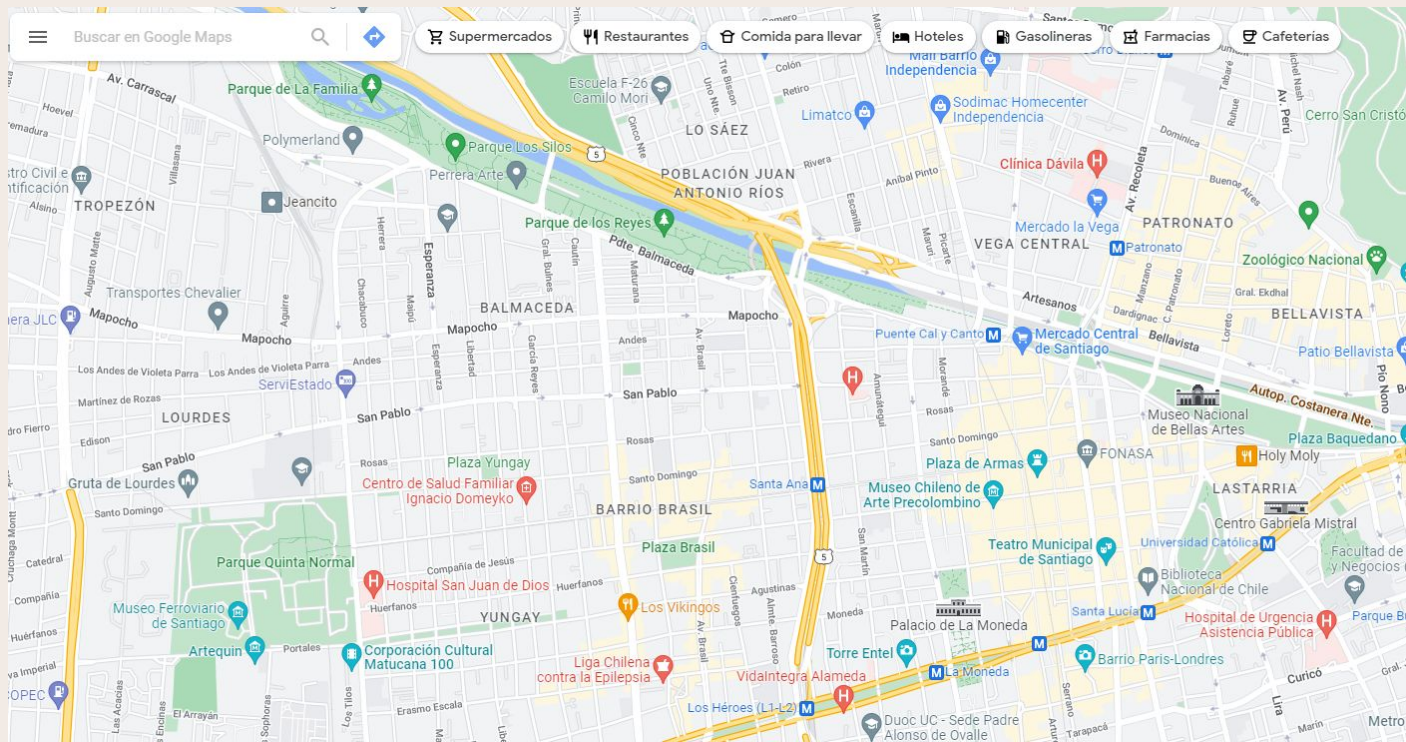
¿Por qué hacer focus+context?

El objetivo principal de hacer focus+context es disminuir la posibilidad de que el usuario se **desorienta**, dado un **idiom de navegación** como podría ser la traslación (panning) o el zooming.

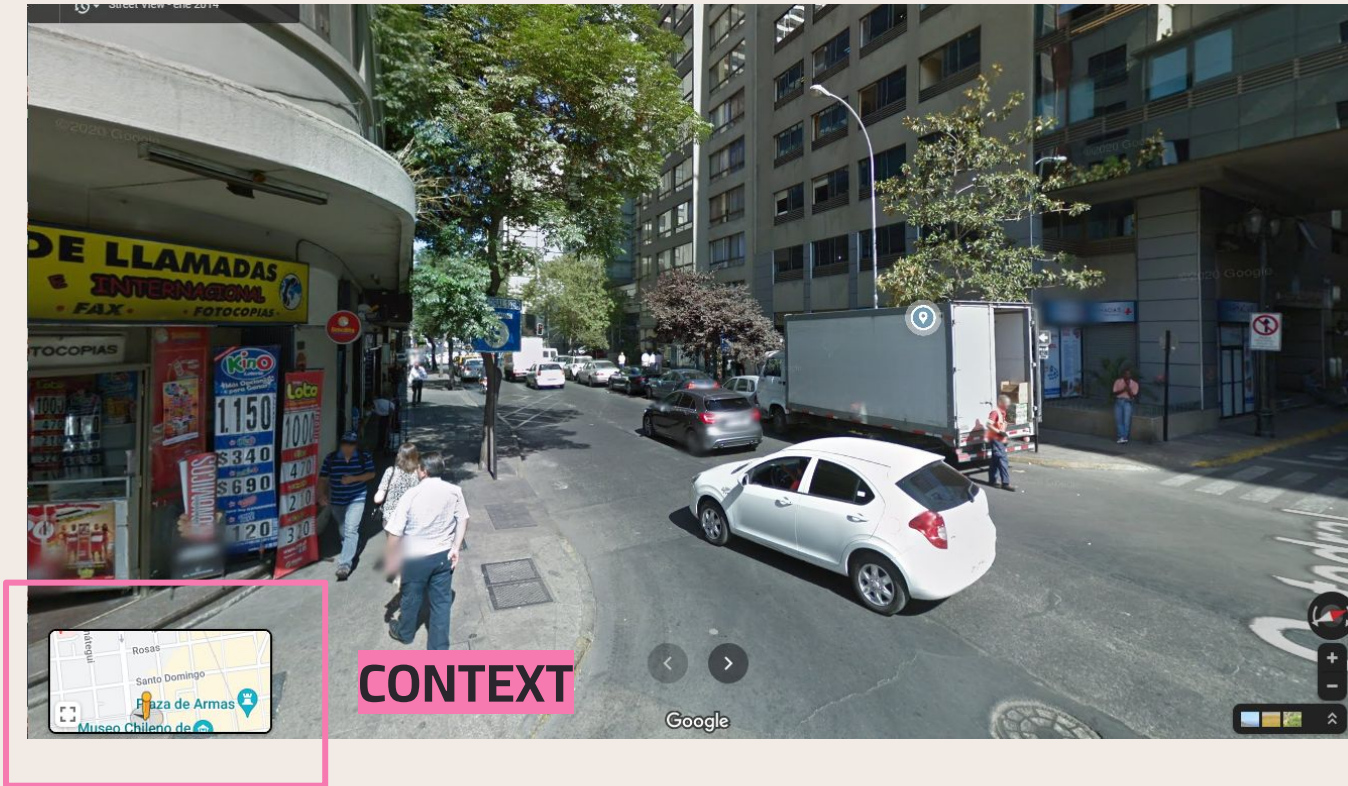
Los idioms focus+context mantienen en las **vistas detalladas**, **información contextual** que permite al usuario ubicarse, sin necesidad de apelar a su propia memoria, dejándole esta información disponible en la vista.



Focus + context



Focus + context



Elipsis

La **elipsis** se refiere a omitir ciertos ítems de la vista completamente (filtro), agregar otro conjunto de ítems y mostrarlos de manera resumida para entregar contexto y finalmente **solo mostrar el subconjunto de ítems del focus en detalle**

Elipsis

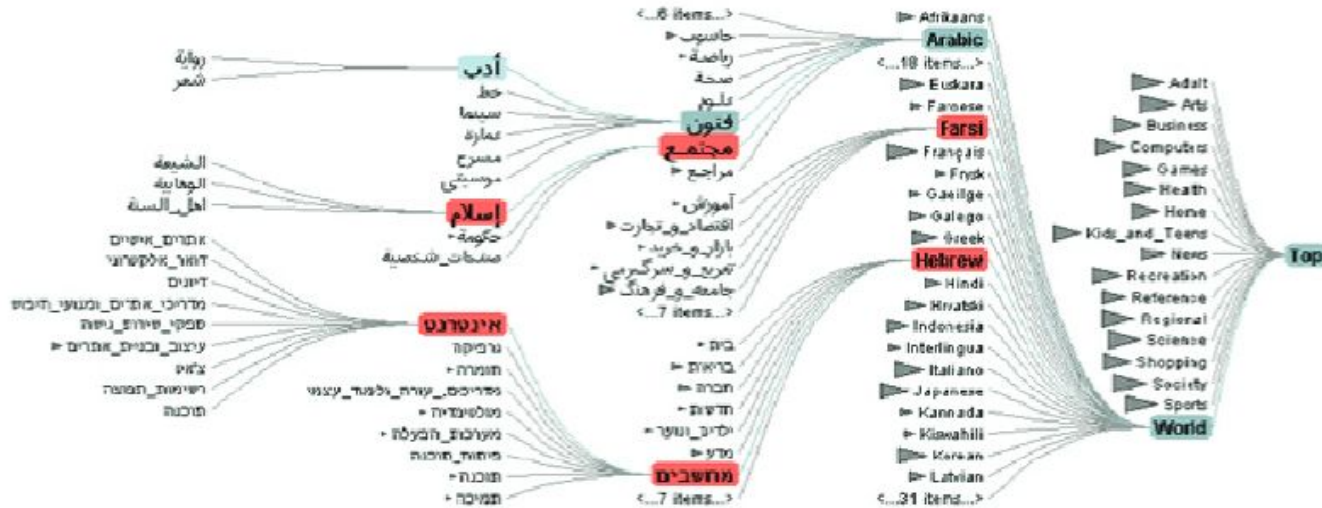


Figure 14.2. DOI Trees Revisited uses elision to show multiple focus nodes within context in a 600,000 node tree. From [Heer and Card 04, Figure 1].



Superponer

Otra manera de integrar un idiom de focus+context en una visualización es usar **capas superpuestas**. En este caso, habrá una **capa de focus**, que estará limitada a una región particular de la visualización. La capa principal, será la capa de contexto.



Superponer

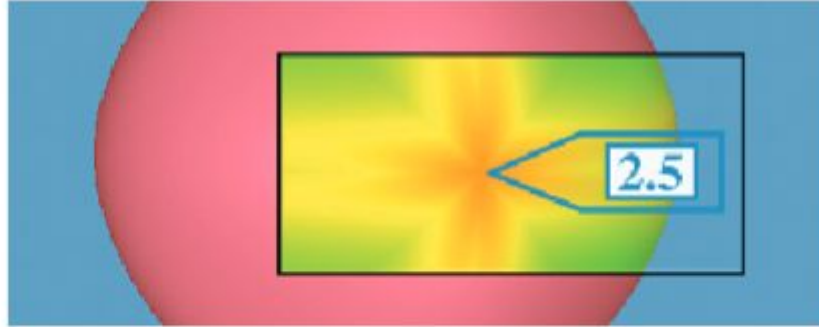


Figure 14.3. The Toolglass and Magic Lenses idiom provides focus and context through a superimposed local layer: the see-through lens color codes the patchwork sphere with Gaussian curvature information and provides a numeric value for the point at the center. From [Bier et al. 93, Figure 12].

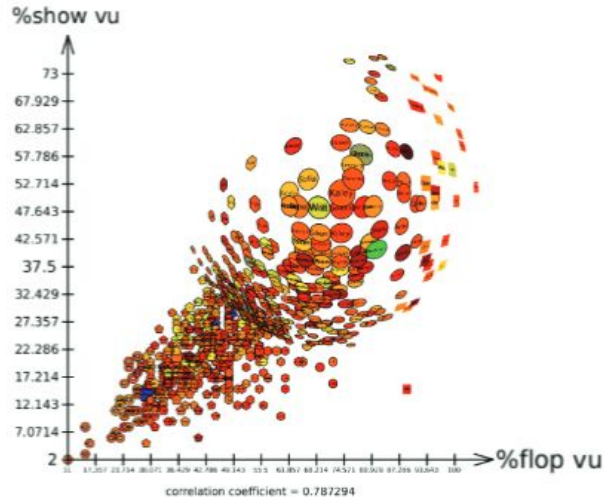


Distorsionar

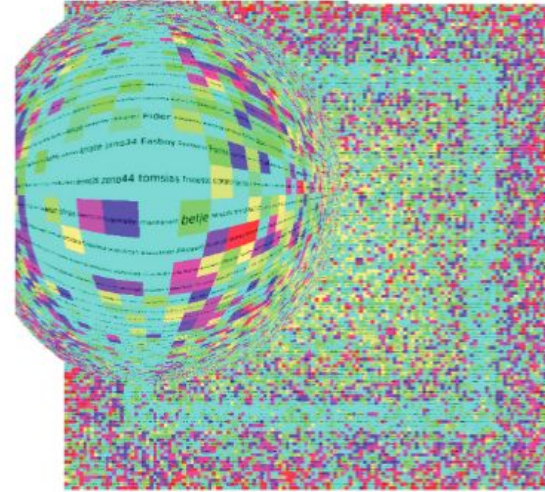
Finalmente estudiaremos la técnica de distorsionar, que se refiere a hacer una **distorsión geométrica** de regiones particulares para poder hacer espacio de **incluir detalles en el área de focus**



Distorsionar: ejemplo lente ojo de pez



(a)



(b)

Figure 14.5. Focus+context with interactive fisheye lens, with poker player data-set. (a) Scatterplot showing correlation between two strategies. (b) Dense matrix view showing correlation between a specific complex strategy and the player's winning rate, encoded by color.

Lente ojo de pez

Idiom	Lente ojo de pez
Qué: datos	Cualquier conjunto de datos
Cómo: codificación	Cualquier layout
Cómo: reducción	Distorsionamos con ojo de pez, un único focus. Región radial local, con posibilidad de que el usuario la desplace a través de la visualización.

Control 2:

Crítica de un paper de infovis



Visualización de información y analítica visual

Clase 8: Reducción de datos / Focus+Context

