

VISUALIZACIÓN REDUCCIÓN DE LA DIMENSIÓN

Visualización

Reducción Dimensional

- En varias circunstancias es necesario reducir las dimensiones de los datos porque exceden las capacidades de la técnica de visualización. En este proceso es necesario preservar, lo más posible, la información contenida en los datos.
- Existen varias técnicas, incluyendo la reducción manual, tales como Análisis de Componentes Principales (PCA), Escalamiento MultiDimensional (MDS), Mapas Auto-Organizativos de Kohonen (SOM), Incrustación Lineal Local.

Análisis de Componentes Principales

PCA

- El Análisis de Componentes Principales es un procedimiento estadístico que usa una transformación ortogonal para convertir un conjunto de observaciones de variables que están posiblemente correlacionadas a un conjunto de valores de variables que están linealmente sin correlación que se llaman componentes principales.
- Es un método no paramétrico para extraer información significativa de conjuntos de datos “confusos”.
- Un método para descubrir las estructuras de baja dimensión y ocultas que forman un conjunto de datos.
- Estas estructuras se visualizan más fácil.

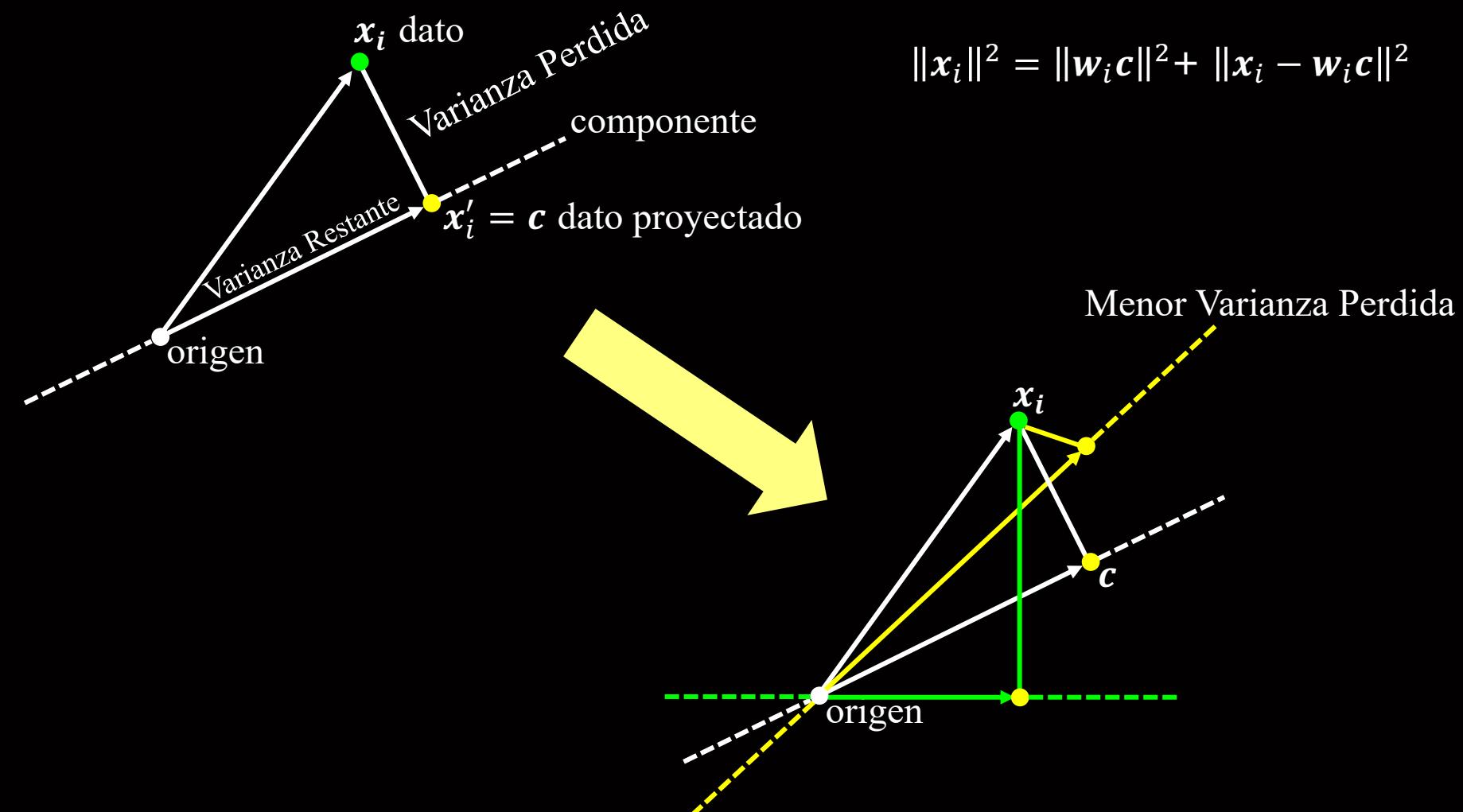
Análisis de Componentes Principales

PCA

- El objetivo del Análisis de Componentes Principales es calcular la base más significativa para expresar el conjunto de datos “ruidosos”.
- La esperanza es que la base nueva filtrará el ruido y hará evidente la estructura oculta de los datos.

Análisis de Componentes Principales

PCA



Análisis de Componentes Principales

PCA

- Lo que se desea es realizar un mapeo $W: \mathbb{R}^n \rightarrow \mathbb{R}^q$, donde $q < n$, de tal forma que
$$\mathbf{y} = W^T \mathbf{x}, \mathbf{x} \in \mathbb{R}^n \text{ y } \mathbf{y} \in \mathbb{R}^q.$$

- Si formamos la matriz \mathbf{A} como el agrupamiento de los m datos ($[\mathbf{x}_1 \dots \mathbf{x}_m]$), entonces el problema de optimización es

$$\max_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{w}^T \mathbf{A}\|^2 = \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{A}^T \mathbf{w}$$

de tal forma que $\frac{1}{2} \mathbf{w}^T \mathbf{w} = 1$.

Análisis de Componentes Principales

PCA

Solución por Eigenvalores

$$L(\mathbf{w}, \lambda) = \frac{1}{2}\mathbf{w}^T \mathbf{A} \mathbf{A}^T \mathbf{w} - \lambda\left(\frac{1}{2}\mathbf{w}^T \mathbf{w} - 1\right)$$

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, \lambda) = 0 \Rightarrow \mathbf{A} \mathbf{A}^T \mathbf{w} = \lambda \mathbf{w}$$

Esto produce el primer componente (primer eje). Para el segundo, se resuelve el siguiente problema

$$\max_{\mathbf{w}_2 \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{w}_2^T \mathbf{A}\|^2$$

de tal forma que $\frac{1}{2}\mathbf{w}_2^T \mathbf{w}_2 = 1$ y $\mathbf{w}_1^T \mathbf{w}_2 = 0$.

Que se soluciona

$$L(\mathbf{w}_2, \lambda, \beta) = \frac{1}{2}\mathbf{w}_2^T \mathbf{A} \mathbf{A}^T \mathbf{w}_2 - \lambda\left(\frac{1}{2}\mathbf{w}_2^T \mathbf{w}_2 - 1\right) - \beta \mathbf{w}_1^T \mathbf{w}_2$$

$$\frac{\partial}{\partial \mathbf{w}_2} L(\mathbf{w}_2, \lambda, \beta) = 0 \Rightarrow \mathbf{A} \mathbf{A}^T \mathbf{w}_2 - \lambda \mathbf{w}_2 - \beta \mathbf{w}_1 = 0$$

$$\mathbf{A} \mathbf{A}^T \mathbf{w}_2 = \lambda \mathbf{w}_2$$

Por lo tanto, λ_1 y λ_2 corresponden a las soluciones y a los eigenvectores PC.

Análisis de Componentes Principales

PCA

Solución general

$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_q \in \mathbb{R}^n} \frac{1}{2} \sum_{j=1}^q \left\| \mathbf{w}_j^T \mathbf{A} \right\|^2$$

de tal forma que $\frac{1}{2} \mathbf{w}_i^T \mathbf{w}_i = 1$ y $\mathbf{w}_i^T \mathbf{w}_j = 0$ ($i \neq j$).

Considerando una matriz $\mathbf{W}_{n \times q}$ cuyas columnas son los vectores \mathbf{w}_i y $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_q)$, donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$ son los q eigenvalores más grandes de $\mathbf{A}\mathbf{A}^T$. Entonces el problema es

$$\max_{\mathbf{W} \in \mathbb{R}^{n \times q}} trace(\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W})$$

de tal forma que $\mathbf{W}^T \mathbf{W} = \mathbf{I}$.

Análisis de Componentes Principales

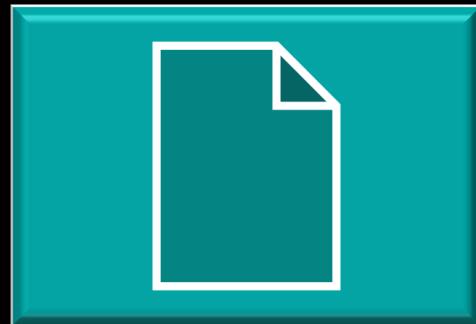
PCA

Un algoritmo para PCA es:

1. Obtener la media de los datos y restarlos a cada uno de los datos. Esto resulta en una media igual a cero.
2. Calcular la matriz de covariancia $\mathbf{S} = \frac{1}{n}(\mathbf{A} - \bar{\mathbf{A}})^T(\mathbf{A} - \bar{\mathbf{A}})$, donde $\bar{\mathbf{A}} = I_{n \times m}(\bar{x}_1, \dots, \bar{x}_m)^T$.
3. Calcular los eigenvalores y eigenvectores de la matriz \mathbf{S} .
4. Ordenar los eigenvalores (preservando la relación con sus eigenvectores) en orden mayor a menor.
5. Seleccionar los q eigenvalores más grandes (representan las dimensiones que se desean).
6. Calcular $\mathbf{S}^T \mathbf{A} = \mathbf{Y}$. La matriz \mathbf{Y} representa los datos transformados.

Análisis de Componentes Principales

PCA



Ejemplo Gráfico

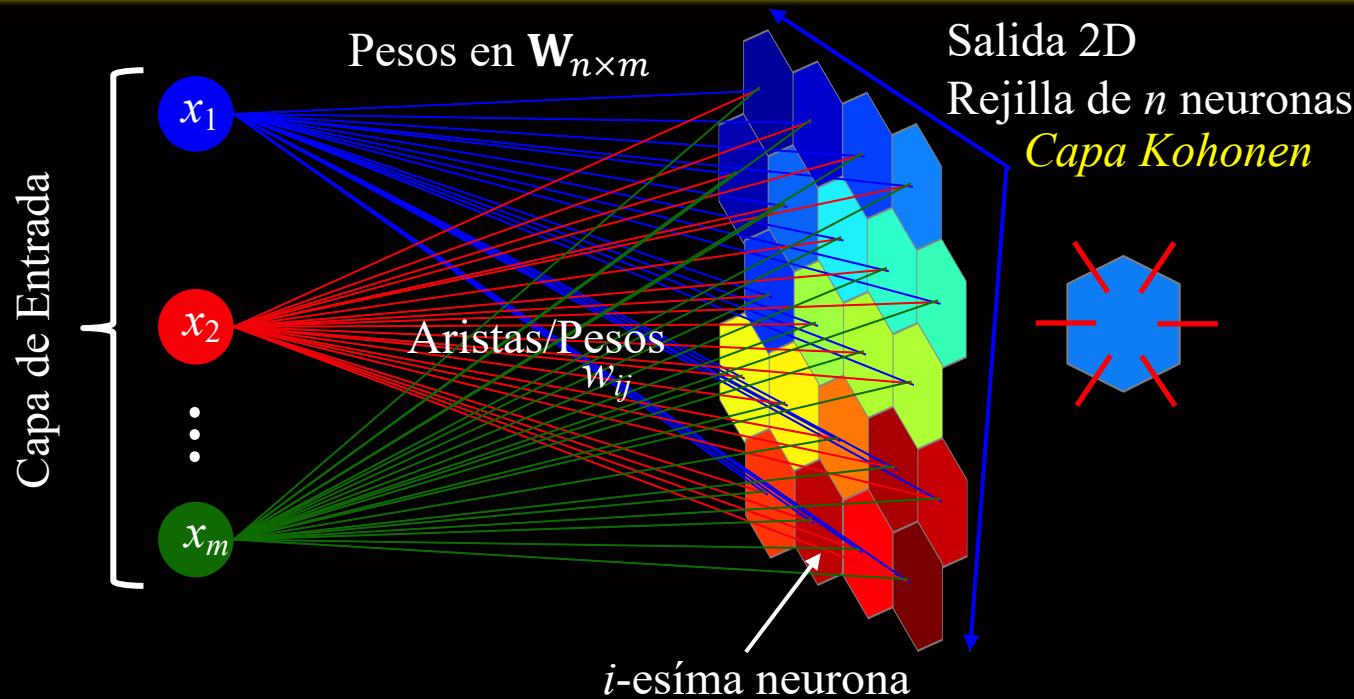
Mapas Auto-Organizativos

SOM o Mapas de Kohonen

- Propuesto por Teuvo Kohonen, es uno de los modelos de redes neuronales no supervisadas más comunes para realizar agrupamiento (*clustering*), detección de características o reducción de dimensiones.
- Para visualización, los SOMs se usan para proyectar datos de altas dimensiones a un espacio de menos dimensiones manteniendo propiedades topológicas similares.
- En el proceso del desarrollo del modelo, las SOMs utiliza técnicas de aprendizaje competitivo (otras redes neuronales usan un proceso de reducción de error a través de retro-propagación con gradiente descendiente).
- Se puede decir, entonces, que una SOM reduce las dimensiones de los datos y despliega similitudes entre los datos.

Mapas Auto-Organizativos

SOM o Mapas de Kohonen



Mapas Auto-Organizativos

SOM o Mapas de Kohonen

Método de Aprendizaje:

1. Inicializar los pesos (aleatoriamente, componentes principales).
2. $t \leftarrow 0$
3. Repetir
4. Seleccionar una entrada \mathbf{x}_k aleatoriamente.
5. Encontrar la neurona ganadora (Best Matching Unit) que cumple:

$$c_i^t = \arg \min_j \|d(\mathbf{x}_k, \mathbf{w}_j^t)\|$$

6. Encontrar el vecindario de la neurona c_i^t con el radio $\rho^t = \rho_0 e^{-\frac{t}{\lambda}}$.
7. Actualizar los pesos de BMU:

$$\mathbf{w}_j^{t+1} = \mathbf{w}_j^t + \eta^t L^t (\mathbf{x}_k - \mathbf{w}_j^t)$$

donde L es la tasa de aprendizaje y η es la tasa de influencia. Que se calculan $L^t = L_0 e^{-\frac{t}{\lambda}}$ y $\eta^t = e^{-\frac{1}{2} \left(\frac{d(\mathbf{x}_k, \mathbf{w}_j^t)}{\rho^t} \right)}$

8. $t \leftarrow t + 1$

Escalamiento MultiDimensional

MDS

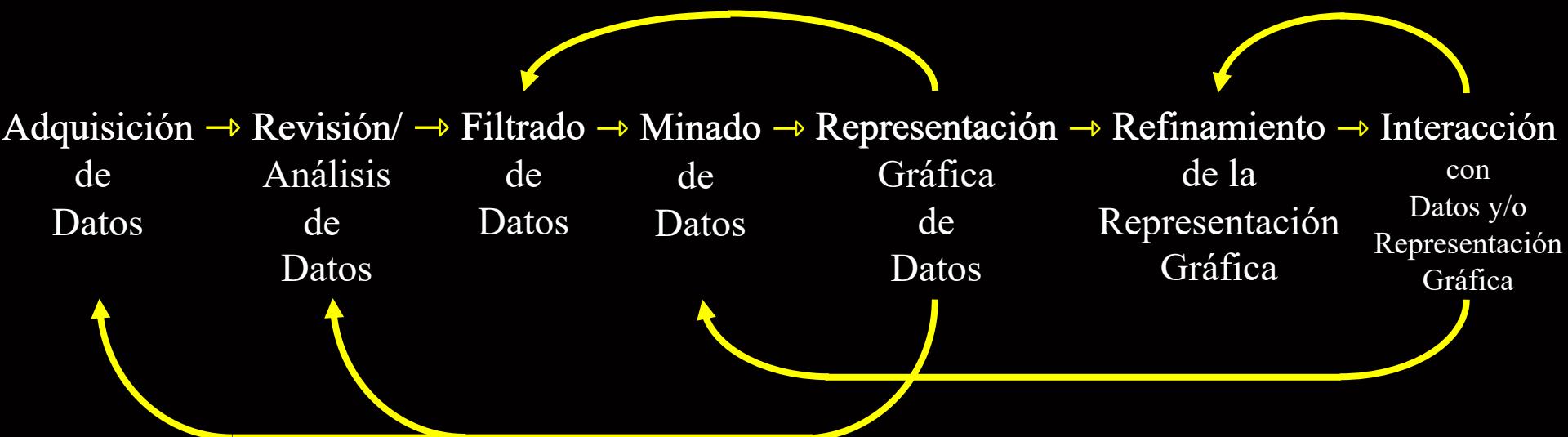
- Es una familia de métodos que tienen su origen en el desarrollo de un modelo psicológico para juicios de discrepancia/similitud.
- En MDS, las entidades pueden ser diferentes ciudades, productos manufacturados, servicios, programas, modos de producción o participantes en una competencia (entidades que pueden ser presentadas a un juez o panel de jueces quienes pueden calificar o asignar calificaciones con respecto a alguna característica que revela similitudes o distancias entre entidades).

Escalamiento MultiDimensional

MDS

- MDS es un método que crea un mapa que muestra las posiciones relativas de un número de objetos, basándose únicamente en una tabla de similitudes (distancias) entre los objetos. El mapa puede tener varias dimensiones (1, 2, 3 o más).
- El método calcula la solución métrica o no-métrica.
- La tabla de similitudes (distancias) se conoce como la matriz de proximidad y se puede producir directamente de experimentos o indirectamente usando una matriz de correlación.

Pasos para Generar una Visualización

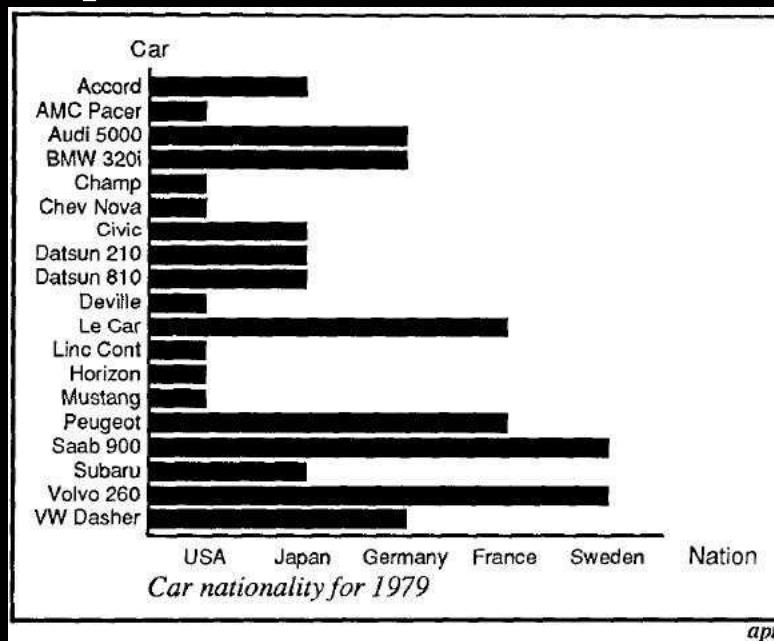


El Proceso de Visualización a Detalle

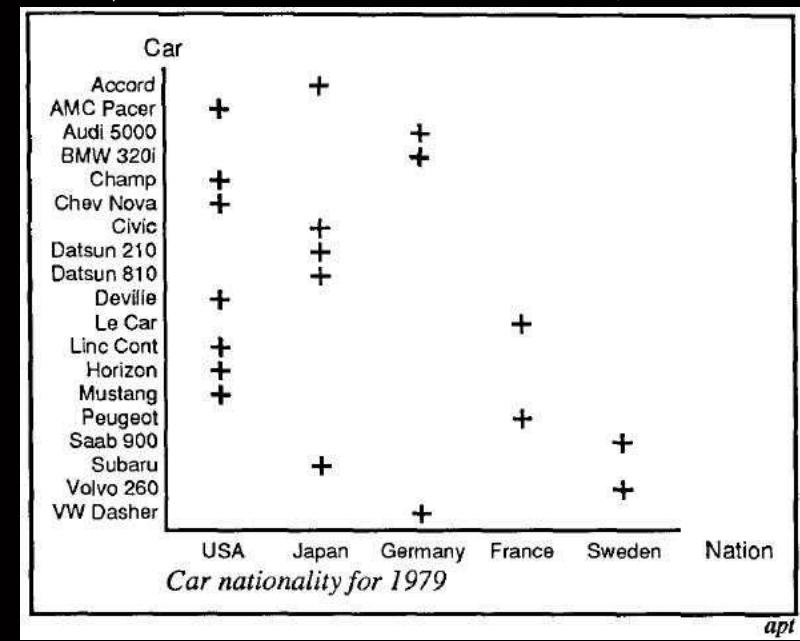
- Procesamiento y Transformación de Datos. La visualización comienza con el procesamiento de los datos “crudos” (puros) para conseguir datos útiles para el sistema de visualización.
 1. Hay que asegurarse de que los datos son mapeados a tipos de datos que puedan ser procesados por una computadora.
 2. Dependiendo de la aplicación, hay que lidiar con datos faltantes, errores en las entradas, conjuntos de datos muy grandes. Hay que tomar acciones para resolver estas situaciones (p.ej., descartar, interpolar, filtrado, dividir los datos).

El Proceso de Visualización a Detalle

- **Mapeo para visualizar.** Una vez que los datos han sido limpiados, se puede decidir sobre el tipo de representación visual que se desea o necesita. Esto requiere mapeos de representación tales como geometría, color o sonido. Cualquier visualización está crucialmente influenciada por expresividad y efectividad (en caso de poseer medidas para ellas, es necesario utilizarlas para mejorar una visualización).



Mala Visualización



Mejor Visualización

El Proceso de Visualización a Detalle

- Transformaciones para *renderizar*. La etapa final consiste en mapear de la información geométrica a una imagen, lo cual incluye interactuar con una Interfaz del Programador de Aplicaciones (API) de graficación por computadora. Para esto es necesario seleccionar los parámetros adecuados para generar la representación adecuada (volumen de visualización, colores, posición de la cámara).
- Es posible medir que tan bien se transfiere la información en cada etapa usando varias medidas, tales como expresividad o efectividad.

El Proceso de Visualización a Detalle

- **Expresividad.** Una visualización expresiva presenta toda la información y únicamente la información pertinente. La expresividad mide la concentración de la información. Dada la información que se muestra al observador, se **puede** definir una medida de expresividad como

$$M_{exp} = \frac{\text{Info Presentada}}{\text{Info Deseada}}$$

- Esta relación varia en el rango [0,1], donde 1 representa la expresividad ideal. Si la información desplegada es menor a la deseada, entonces la tasa es menor a uno y cuando hay mucha información presentada la tasa es mayor a uno. Expresar información adicional es potencialmente peligroso porque puede ser incorrecto y puede interferir con la interpretación de la información esencial. Esta medida de expresividad se puede extender para incluir varios conjuntos de información (lo cual la convierte en una función sobre conjuntos).

El Proceso de Visualización a Detalle

- Efectividad. Una visualización es efectiva cuando puede ser interpretada exactamente y rápidamente y cuando puede ser *renderizada* en una forma que cueste lo suficiente. Por lo tanto, la efectividad mide un costo específico de la percepción de información. La efectividad se puede definir como

$$M_{eff} = \frac{1}{1 + t_{\text{interpretación}} + t_{\text{renderizado}}}$$

- Entre mayor es M_{eff} , mayor es la efectividad. Si la relación es pequeña se puede deber a que el tiempo de interpretación o el de renderizado es muy alto. Si la relación es cercana a uno, entonces los tiempos son muy pequeños.

El Proceso de Visualización a Detalle

Efectividad.

