

Datos Multivariados

Juan Morales del Olmo

24 de marzo de 2014

Datos Multivariados

Características

- ▶ Usualmente conocidos como datos **Tabulares** o **Multidimensionales**
- ▶ Formatos comunes: Excel(xls, xlsx), texto (csv, tsv, ...)
- ▶ Formado por **items**, muestras o individuos (filas) y **campos**, dimensiones o variables (columnas)

El problema

- ▶ **Muchas dimensiones que representar gráficamente de forma efectiva**
- ▶ Informalmente: “No podemos ver más de 3 dimensiones”
 - ▶ Sólo cierto si cada dimensión se *mapea*, en un mismo espacio, utilizando la variable visual “**Posición**”.

Objetos de estudio típicos:

- ▶ Correlaciones
- ▶ Regresiones
- ▶ Clasificaciones
- ▶ Agrupaciones (clustering)
- ▶ Análisis de varianza.

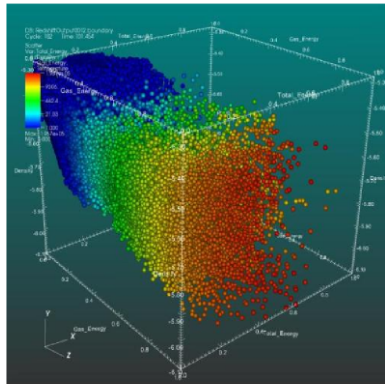
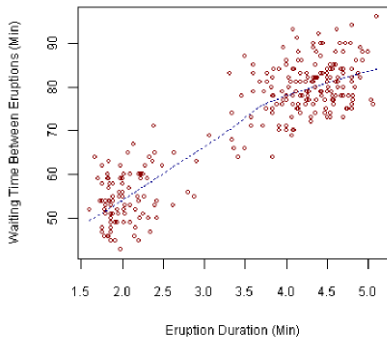
Representaciones basadas en puntos

Estrategias

- ▶ Submuestreo de dimensiones
- ▶ Múltiples vistas
- ▶ Reducción de la dimensionalidad

Scatter Plots

Old Faithful Eruptions

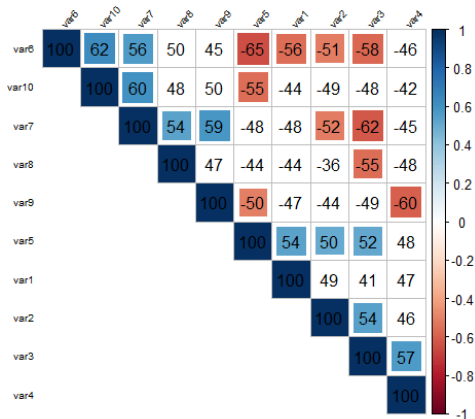


Scatter Plots Matrix

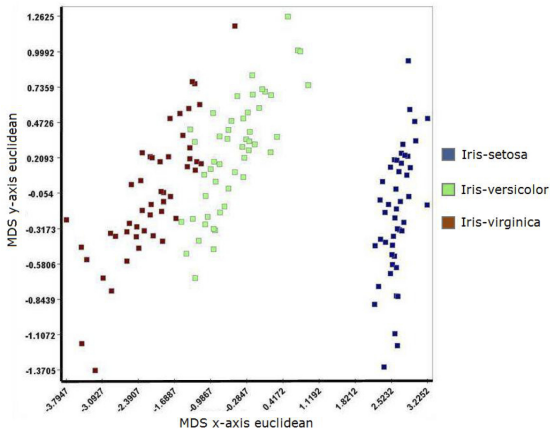


Figure: <http://bl.ocks.org/mbostock/raw/4063663/>

Correlation Matrix



Reducción de dimensionalidad (MDS)



Grand Tour



Figure: <http://www.ggobi.org/demos/tour.html>

Basados en líneas

Gráfico de líneas multivariado

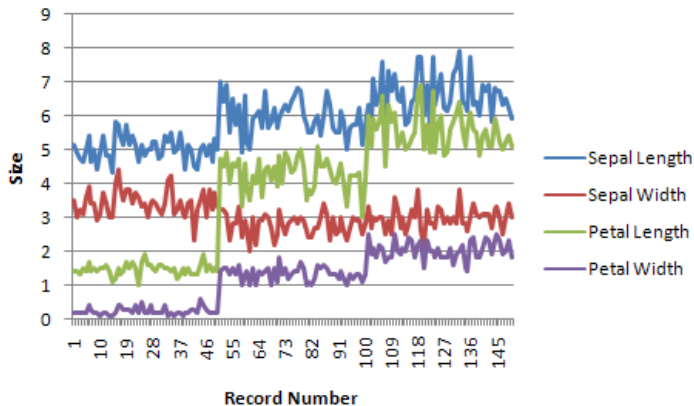


Figure: Yuxtaposición de vistas

Coordenadas Paralelas



Figure: <http://bl.ocks.org/mbostock/raw/7586334/>

Gráfica de Radar

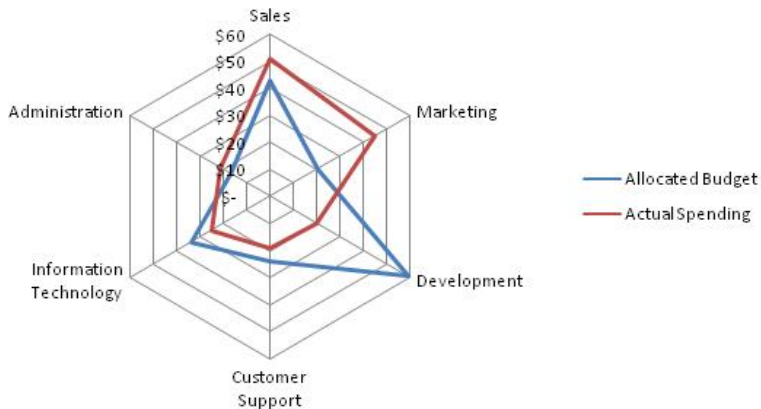


Figure: Buenas para patrones cíclicos

Gráficos de barras agrupados y apilados



Figure: <http://bl.ocks.org/mbostock/raw/3943967/>

Gráficos en mosaico

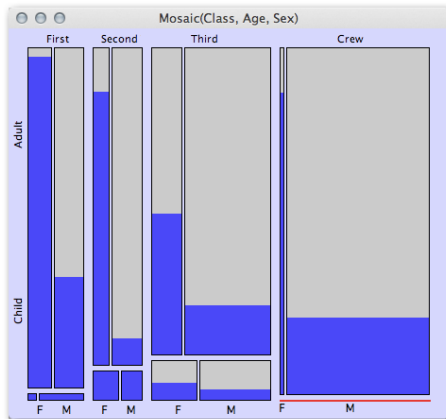


Figure: Titanic dataset

Heatmap

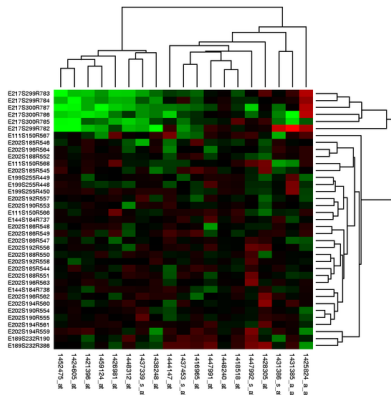


Figure: Clustering revela patrones

Explotar canales gráficos

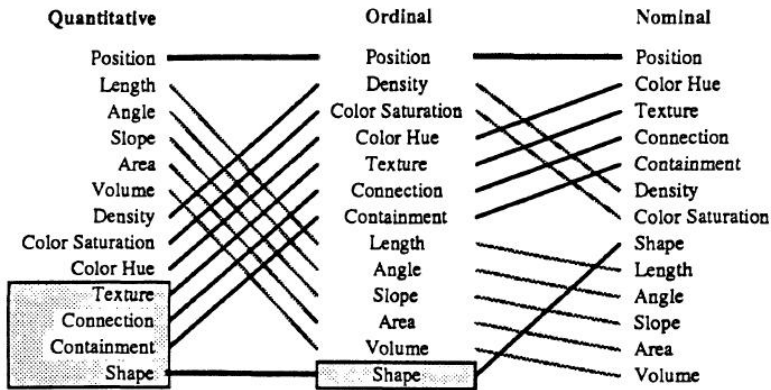
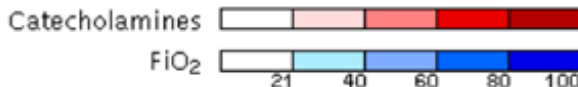
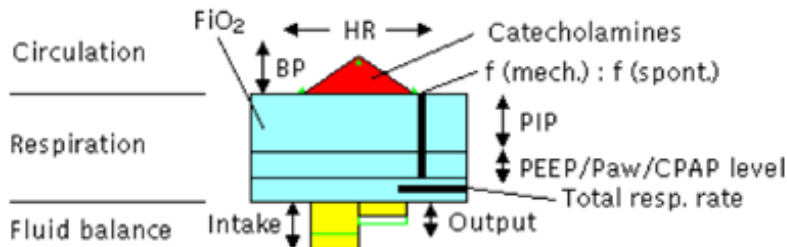
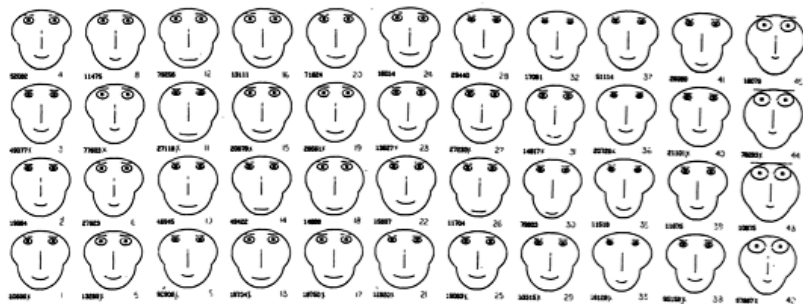


Figure: Jock D. Mackinlay's Perceptual tasks

Glifos



Caras de Chernoff



Estrategias Interactivas

Ordenado según campos

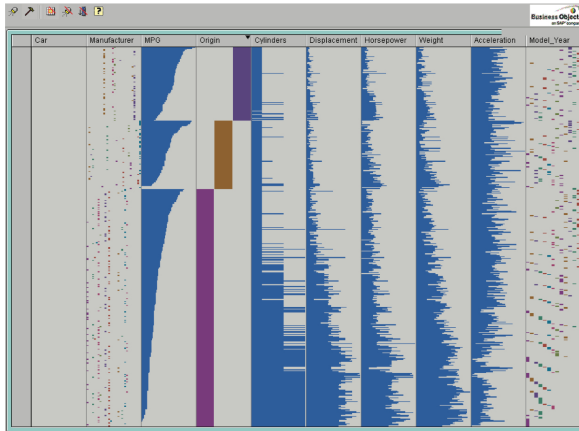


Figure: Inxight

Vistas enlazadas



Figure: <http://selfiecity.net/selfiexploratory/>