



# Inteligencia de Negocios

Data Warehouse y Minería de Datos

# Reflexión

Algo peor que no tener información disponible es tener mucha información y no saber qué hacer con ella.

La Inteligencia de Negocios o Business Intelligence (BI) es la solución a ese problema

Pregunta Inteligente

Bi or not  
to Bi



# Principios de BI en las Organizaciones

# ¿Quién necesita soluciones de Business Intelligence?

¿Pasa más tiempo recolectando y preparando información que analizándola?

¿En ocasiones le frustra el no poder encontrar información que usted está seguro que existe dentro de la empresa?

¿Pasa mucho tiempo tratando de hacer que los reportes en Excel luzcan bien?

¿Quisiera tener una guía sobre las cosas que han sucedido cuando los administradores anteriores implementaban determinada estrategia?

# ¿Quién necesita soluciones de Business Intelligence?

¿No sabe qué hacer con tanta información que tiene disponible en la empresa?

¿Quiere saber qué productos fueron los más rentables durante un periodo determinado?

¿Ha perdido oportunidades de negocio por recibir información retrasada?

¿Trabaja horas extras el fin de mes para procesar documentos o reportes?

# ¿Quién necesita soluciones de Business Intelligence?

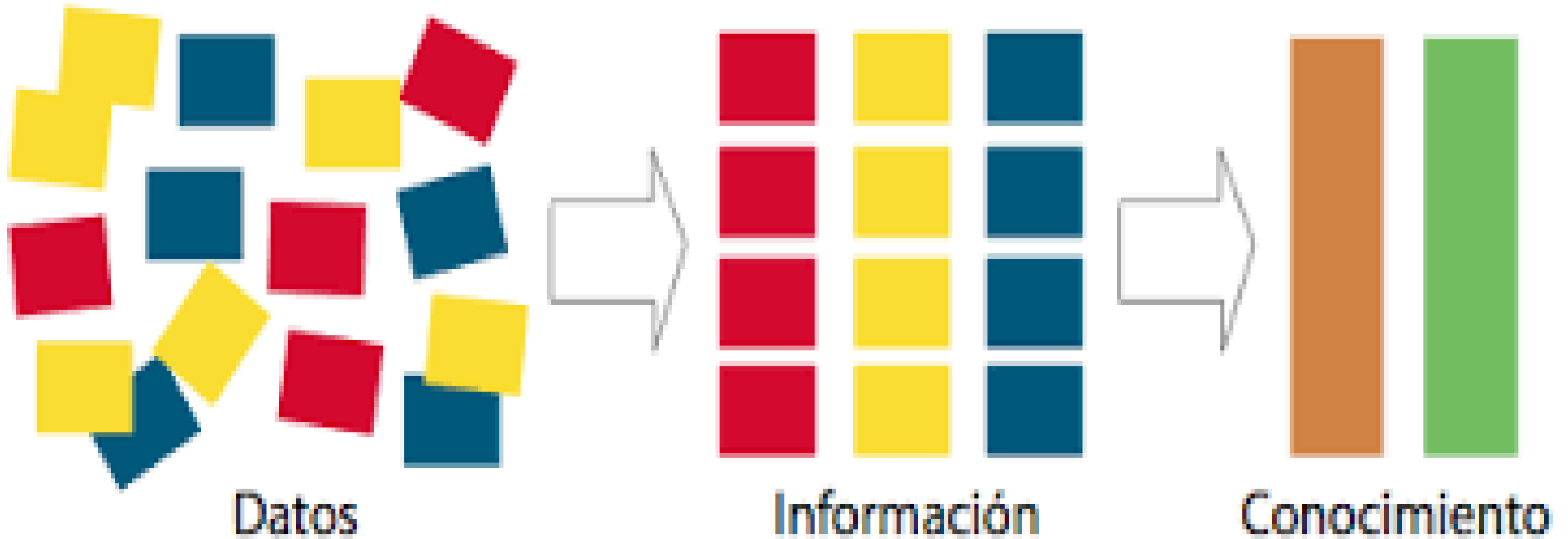
¿No sabe cuáles son los patrones de compra de sus clientes dependiendo de las zonas?

¿No sabe con certeza si su gente está alcanzando los objetivos planeados?

¿No sabe si mantiene una comunicación estrecha entre las diversas áreas de su empresa hacia una estrategia común?

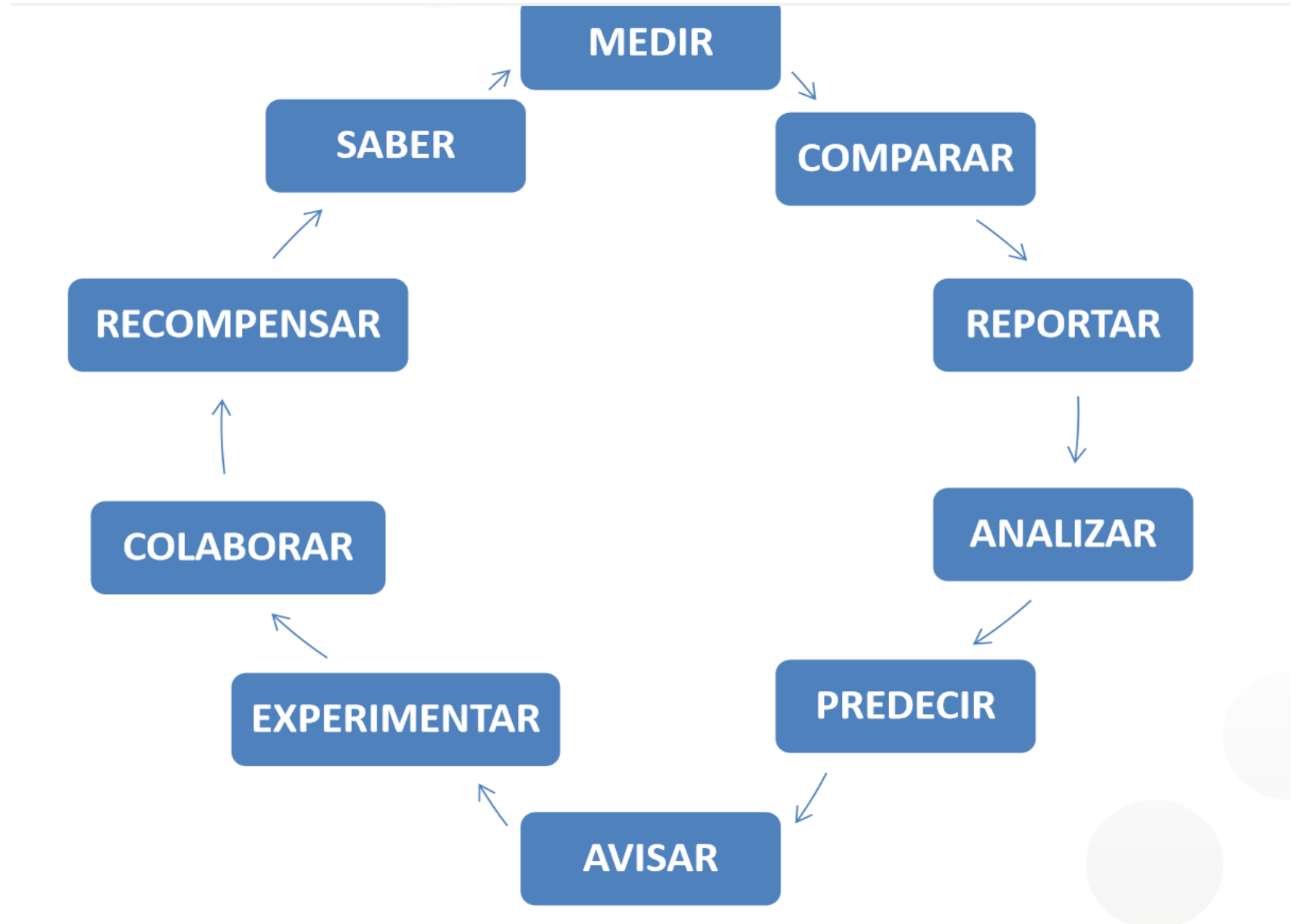
¿No tiene idea de por qué sus clientes le regresan mercancía?

# Proceso





# Datos



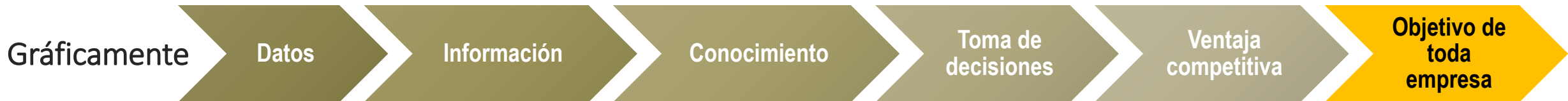
# ¿Información Estrategia?

- ✓ Información clientes, proveedores, socios.
- ✓ Añadir valor.
- ✓ Crear productos nuevos, servicios, ofertas, canales y otros.
- ✓ Nuevos negocios, modelos y mercados.
- ✓ Facilitar toma de decisiones.
- ✓ Colaboración interna e innovación.
- ✓ Colaboración con externos.

# Definición de B.I.

De acuerdo con las fuentes, inteligencia de negocios:

*“Es la habilidad para transformar los datos en información y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios para alcanzar una ventaja competitiva.”*



# Importancia del B.I.

“La misión suprema de toda empresa debe ser el mayor nivel de satisfacción para sus clientes y usuarios, pues éstos con sus compras permiten que la empresa siga existiendo y creciendo, generando de tal forma beneficios para sus integrantes (accionistas, directivos y empleados)”.

“El objetivo más importante de toda empresa es impulsar el bienestar económico y social mediante una adecuada inversión del capital.”

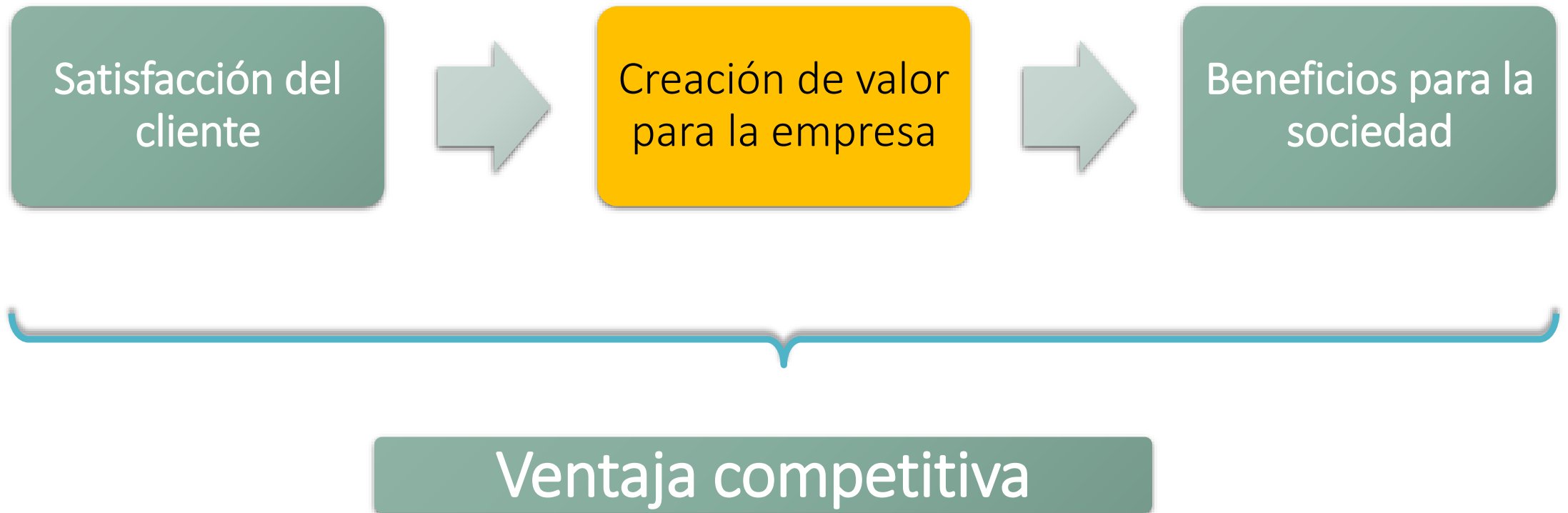
# Importancia del B.I.

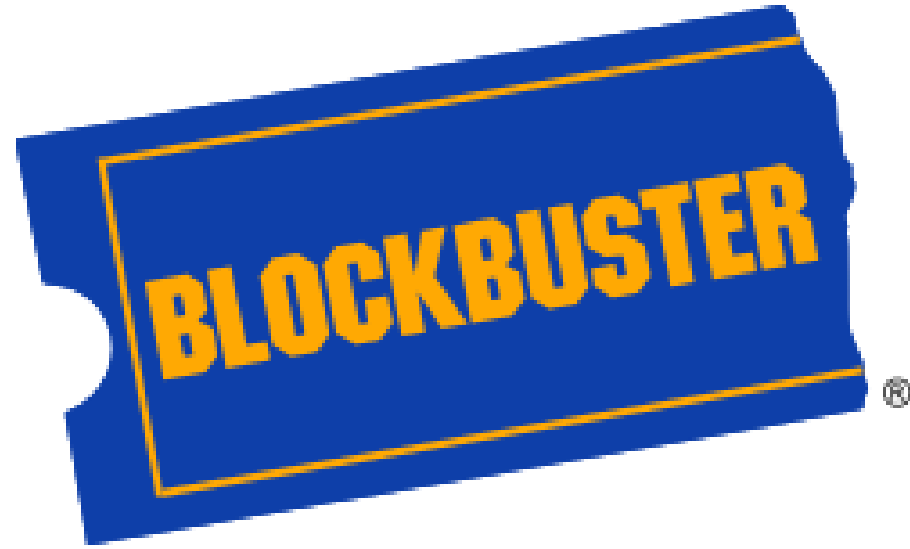
“El propósito de un negocio es lograr utilidades para poder hacer algo más. Ese algo es lo que verdaderamente justifica el negocio.”

“La primera función de toda empresa es y seguirá siendo satisfacer las necesidades de sus clientes mediante los bienes y servicios producidos. Para que esto sea sostenible es necesario invertir y obtener ganancias.”

“La meta de una empresa es ganar dinero.”

# Importancia del B.I.





**NETFLIX**

# Importancia del B.I.

La ventaja competitiva

Parte del hecho que las empresas buscan alcanzar desempeño y rentabilidad sobresaliente, a pesar de operar en industrias caracterizadas por:

- ✓ Competidores que también buscan la creación de valor para el cliente
- ✓ Productos y servicios sustitutos
- ✓ Proveedores con poder de negociación
- ✓ Clientes con poder de negociación



# Conocimiento BI

Dos grandes áreas:

1. Análisis multidimensional (OLAP)
2. Knowledge Discovery in Databases (KDD)

- Preparación de datos

- Minería de datos

  - Descriptiva

  - Predictiva

  - Para clasificación

- Modelos y patrones

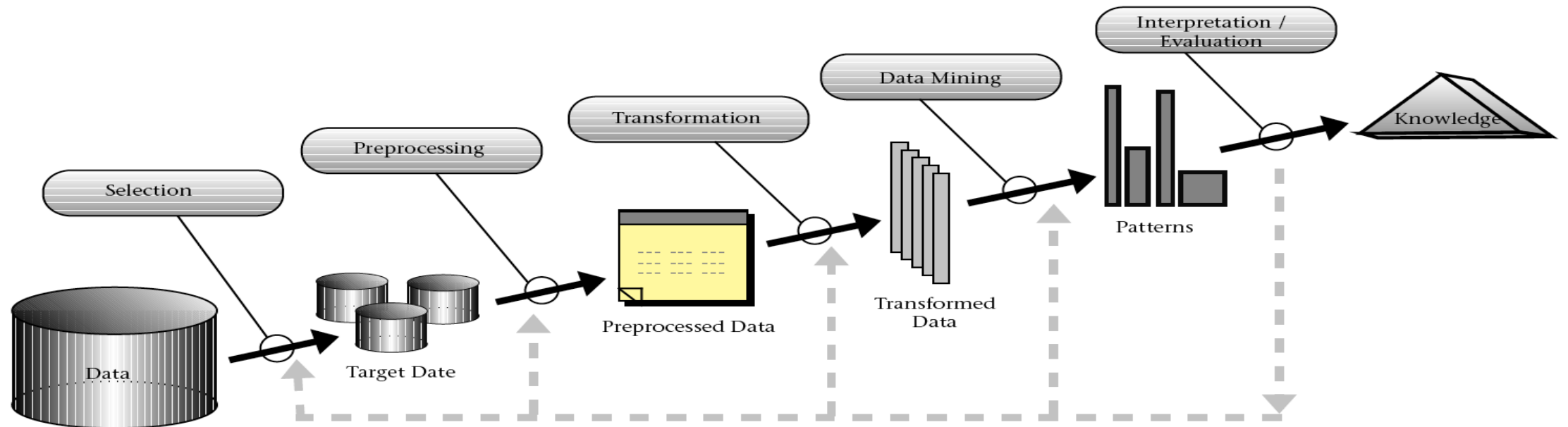
- Evaluación e interpretación

# Conocimiento B.I.

- ✓ Provee los insumos para la toma de decisiones
- ✓ Etapa donde el análisis toma especial importancia.
- ✓ Datos + Análisis = Conocimiento
- ✓ Es aquí donde entra en juego la potencia de las tecnologías de información manejadas con mucho discernimiento, análisis y conocimiento del negocio.

# Conocimiento KDD

El descubrimiento del conocimiento es todo un proceso llamado KDD (Knowledge Discovery in Databases)



# Conocimiento KDD

Enormes volúmenes de información existe información oculta y de gran importancia estratégica, pero que no puede identificarse con las técnicas clásicas de consulta de datos y reportería.

Este descubrimiento se logra con la minería de datos y sus técnicas tanto estadísticas como de inteligencia artificial para encontrar relaciones y patrones que permitan crear modelos de la realidad.

# Aplicaciones KDD

## Finanzas y Banca

- ✓ Obtención de patrones de uso fraudulento de tarjetas de crédito
- ✓ Determinación de grupos según el uso de tarjeta de crédito
- ✓ Correlaciones entre indicadores financieros
- ✓ Análisis de riesgos de crédito

# Aplicaciones KDD

## Análisis de mercado

- ✓ Análisis de la cesta de compras (compras conjuntas, secuenciales, ventas cruzadas, etc).
- ✓ Evaluación de campañas publicitarias
- ✓ Patrones de fidelidad y fuga de clientes
- ✓ Segmentación de clientes

# Aplicaciones KDD

## Seguros y salud

- ✓ Determinación de clientes que podrían ser potencialmente caros
- ✓ Procedimientos médicos solicitados conjuntamente
- ✓ Predicción de qué clientes contratan nuevas pólizas
- ✓ Identificación de patrones de comportamiento para clientes con riesgo
- ✓ Identificación de comportamiento fraudulento
- ✓ Predicción de clientes que ampliarán su póliza

# Aplicaciones KDD

## Telecomunicaciones

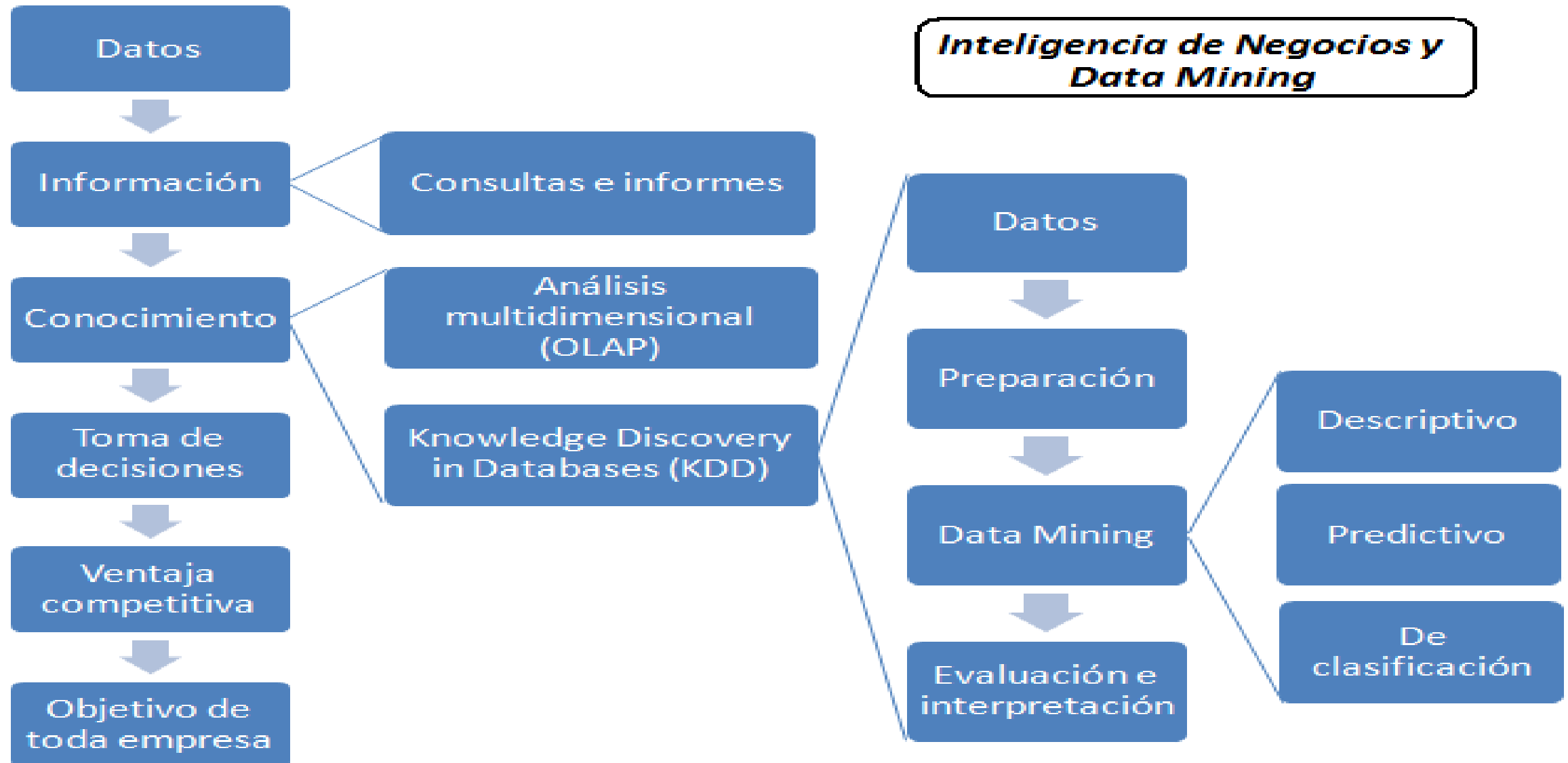
- ✓ Identificar patrones de llamadas
- ✓ Modelos de carga en redes
- ✓ Detección de fraude

## Sistemas de información

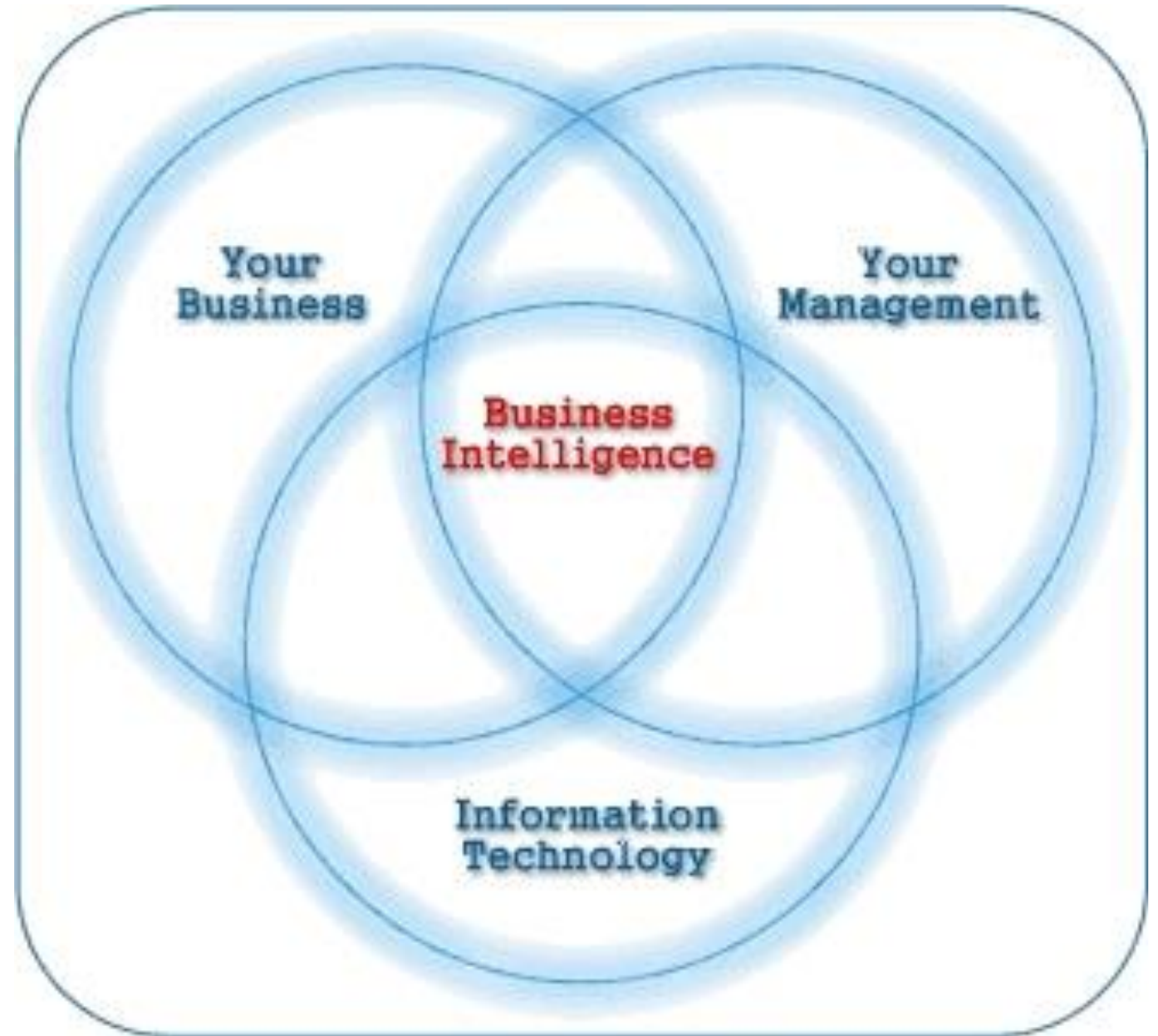
- ✓ Clasificación de correo electrónico
- ✓ Patrones de visitas en la Web
- ✓ Análisis de logs



# Visión de un BI



# Visión de un BI



# Toma de decisiones

# ¿Qué es?

La toma de decisiones es un proceso sistemático y racional a través del cual se selecciona una alternativa de entre varias, siendo la seleccionada la optimizadora (la mejor para nuestro propósito).

Tomar una decisión es resolver diferentes situaciones de la vida en diferentes contextos: a nivel laboral, familiar, sentimental, empresarial, etc.

# ¿Qué es?

Tomar la correcta decisión en un negocio o empresa es parte fundamental del administrador ya que sus decisiones influirán en el funcionamiento de la organización, generando repercusiones positivas o negativas según su elección.

¿Sabemos escoger bien nuestras decisiones?

¿Poseemos algún patrón para tomar nuestras decisiones?

¿El azar es parte de una decisión?

¿Es normal guiarse de la intuición para tomar una decisión?

# Etapas del proceso de toma de decisiones.

- ✓ Identificar y analizar el problema.
- ✓ Investigación u obtención de información.
- ✓ Determinación de parámetros.
- ✓ Construcción de una alternativa.
- ✓ Aplicación de la alternativa.
- ✓ Especificación y evaluación de las alternativas.
- ✓ Implantación.

# Tipos de Decisiones





# Estratégicas

¿Quién será nuestro mercado meta?

¿Qué productos o servicios debemos ofrecer?

¿Dónde localizamos las sucursales o puntos de venta?

¿Qué procesos deberíamos subcontratar y cuáles dejarse en la compañía?

¿Deberíamos implementar un nuevo sistema de información corporativo?

# Tácticas

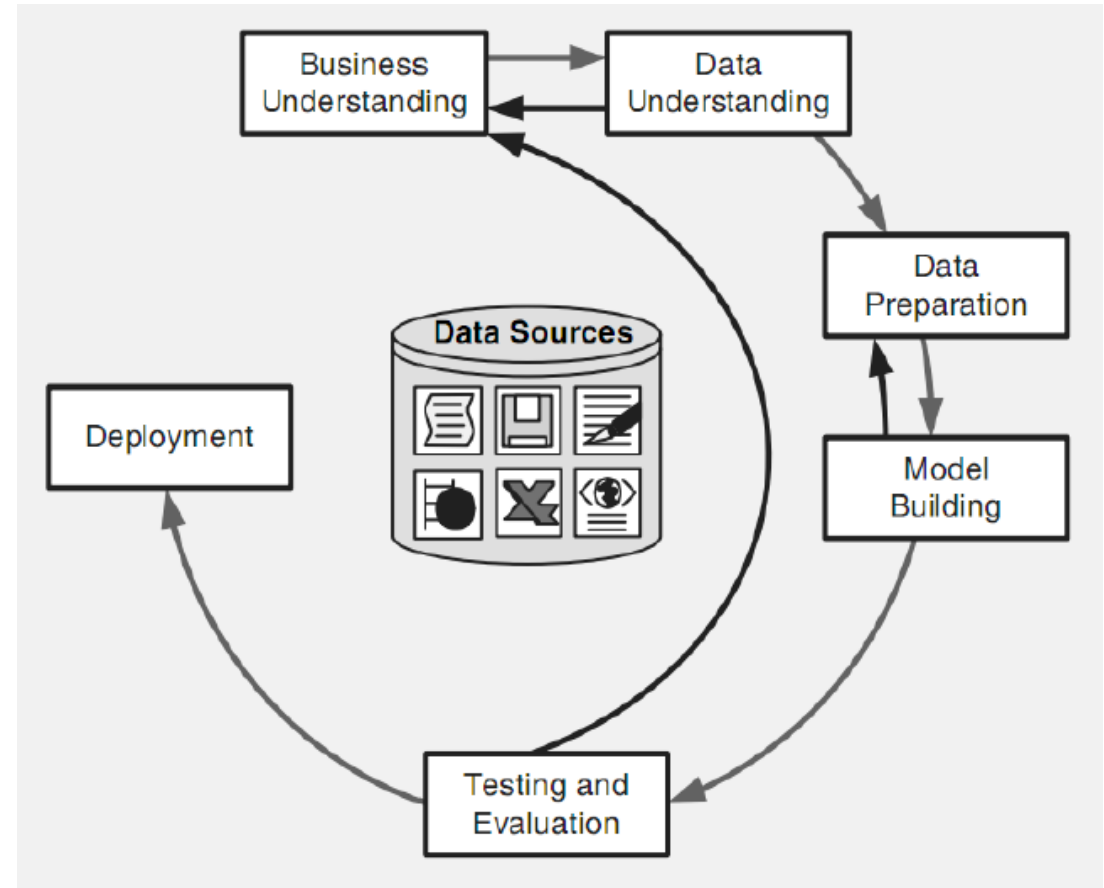
- ¿A quién dirigir la publicidad y a través de qué canales?
- ¿Qué acciones tomar para retener a nuestros clientes?
- ¿Cuánto comprar al proveedor y con qué frecuencia?
- ¿Cómo distribuir los productos en un supermercado?
- ¿Qué políticas de crédito implementar?
- ¿Qué tipo de beneficios para empleados conviene implementar?
- ¿Cuál debería ser el grado de disponibilidad óptimo de los sistemas?

# Explotación

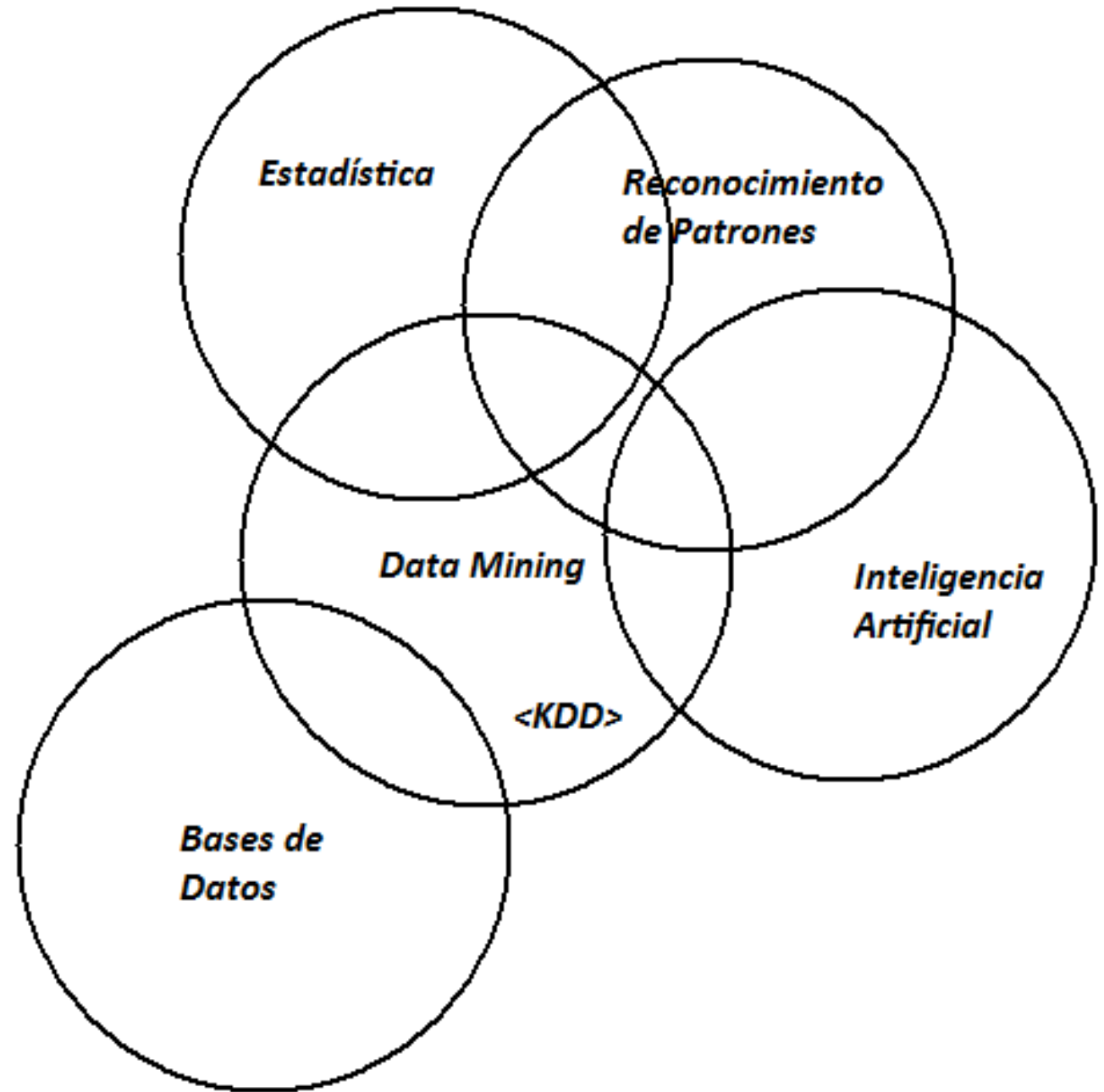
Introducción al BI

# Knowledge Discovery in Databases (KDD)

El proceso de  
Data Mining



# Knowledge Discovery in Databases (KDD)



# Limpieza y transformación

- ✓ Valores atípicos (outliers) y faltantes (missing)
- ✓ Creación de características
- ✓ Discretización y numerización
- ✓ Normalización

# Algoritmo

“Conjunto de instrucciones bien definidas, ordenadas y finitas, que permiten realizar una actividad mediante pasos sucesivos”.

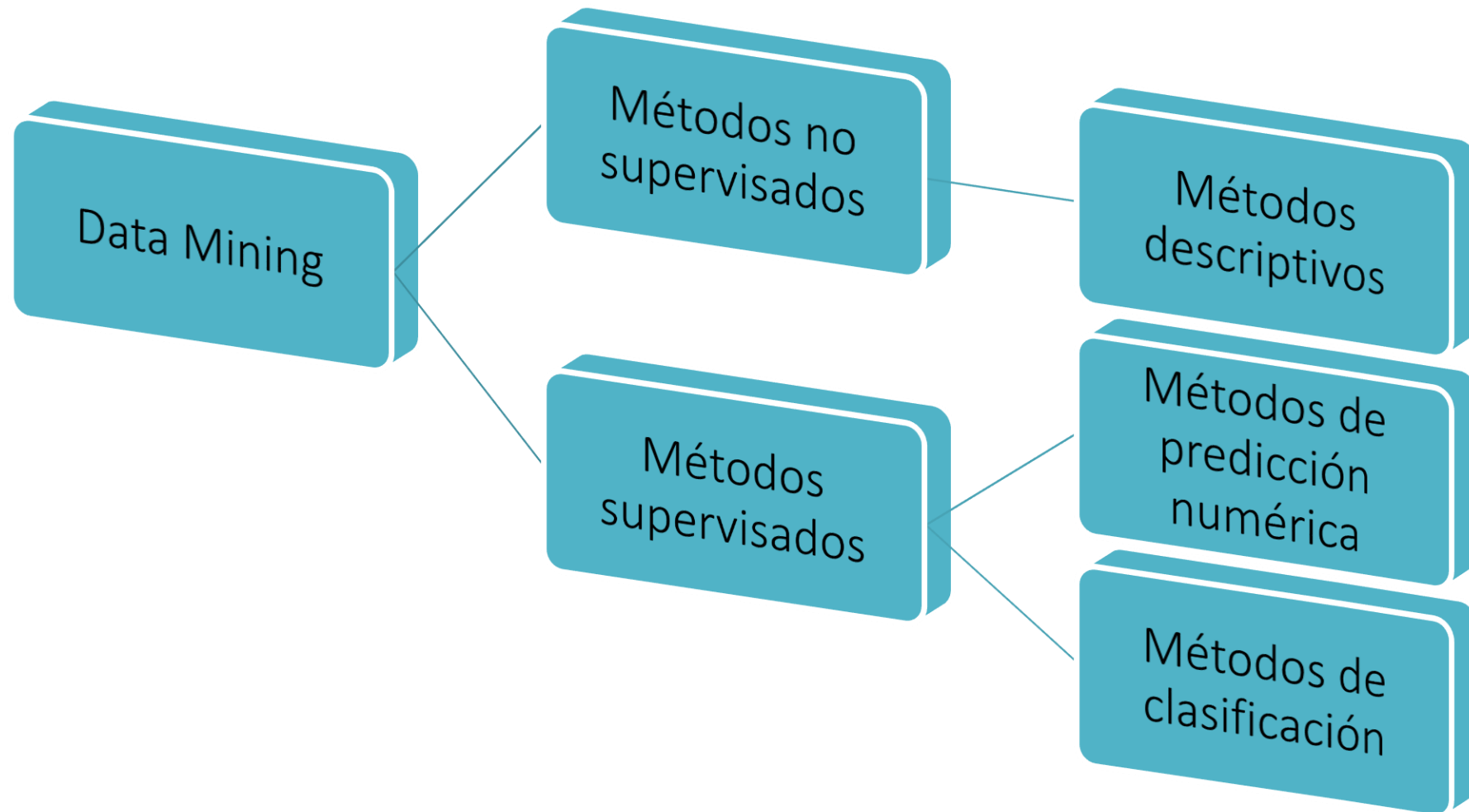
Thomas Cormen ( 2009). Introduction to algorithms. The MIT Press.

# Creación de Modelos de Datos

Si el scoring consiste en aplicar un algoritmo a un juego de datos con el objetivo de predecir una clase objetivo o encontrar un patrón desconocido, el modelado comprende todas las tareas directamente involucradas en el proceso de construcción y ajuste del mencionado algoritmo.



# Knowledge Discovery in Databases (KDD)



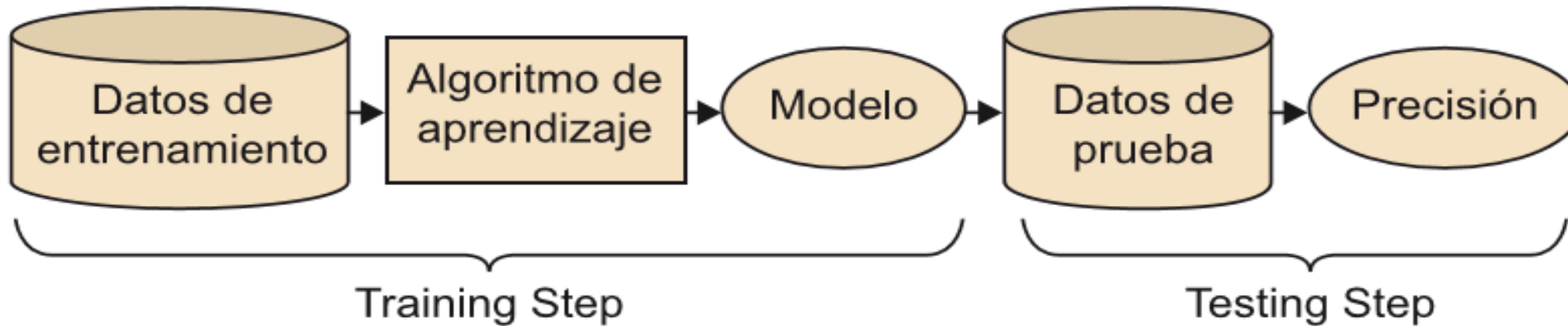
# Métodos supervisados

La clasificación supervisada persigue la obtención de un modelo válido para predecir casos futuros a partir del aprendizaje de casos conocidos.

A partir de un conjunto de objetos descritos por un vector de características y del que conocemos la clase a la que pertenece cada objeto, se construye un juego de datos llamado de entrenamiento o de aprendizaje, que nos servirá para construir un modelo o regla general que nos va a permitir clasificar objetos nuevos de los que no conocemos la clase a la que pertenecen.

Para los algoritmos de ***aprendizaje supervisado***, una secuencia típica de aprendizaje sería la siguiente:

Figura 1. Aprendizaje supervisado



Fuente: Bing Liu. *Web Data Mining*

Donde vemos que a partir de un juego de ***datos de entrenamiento*** se ajusta un ***modelo de aprendizaje***, que acaba estableciendo un modelo, cuyo nivel de precisión es evaluado a partir de un juego de ***datos de prueba***, distinto al de aprendizaje.

# Métodos no supervisados

La clasificación no supervisada persigue la obtención de un modelo válido para clasificar objetos a partir de la similitud de sus características.

A partir de un conjunto de objetos descritos por un vector de características y a partir de una métrica que nos defina el concepto de similitud entre objetos, se construye un modelo o regla general que nos va a permitir clasificar todos los objetos.

No se trata de modelos predictivos sino de modelos de descubrimiento de patrones.

# Métodos no supervisado

Existen dos grandes familias de algoritmos de clasificación no supervisada:

- Los *algoritmos jerárquicos* construyen nodos de forma jerárquica, unos a partir de otros. La representación de los resultados se hace habitualmente mediante de diagramas.

Estos se dividen en dos subcategorías:

- ✓ Aglomerativos o bottom up:
- ✓ Divisivos o top down:

# Métodos no supervisado

- Los *algoritmos particionales*, también llamados algoritmos de optimización, obtienen los nodos a partir de la optimización de una función adecuada para el propósito del estudio.

Esta función suele estar relacionada con la métrica seleccionada para establecer el concepto de similitud entre objetos.

# Visualización

Introducción al BI

# Ejemplos de visualización de datos

<https://data.london.gov.uk/>

<http://www.eia.gov/>

<http://hdr.undp.org/es/content/explorador-de-datos-publicos-0>

<http://datos.bancomundial.org/pais/el-salvador?view=chart>

<http://www.enigma.io/>

<https://www.quandl.com/>



# Exploración mediante visualización

0.55013990	0.48953730	-0.47760050	0.26873980	0.19144250	0.64966080	0.00000000	0.34171810
0.53868300	0.12009240	-0.04613400	0.25846950	-0.78261670	-0.34335560	0.20486080	1.05938110
1.10115710	-0.32305700	-0.84267060	0.33619720	0.71667910	-0.31420390	-0.03708170	-0.00927386
0.27354790	0.15458760	0.43580900	-0.30735850	-0.04533910	-1.38800690	-1.33394660	-1.10955710
0.03747770	0.29034370	0.50429590	1.28228950	0.03669720	-1.27476160	-0.26939150	-0.23286140
0.25990900	0.65734650	0.32192790	-0.22921060	-0.22973720	1.00396060	-0.15502460	0.82199290
0.63392860	0.73640320	-0.12330460	0.99647270	0.39291020	0.51313270	-0.50186030	-0.00869641
0.73504270	0.56847110	-0.93647180	0.79202860	-0.58301650	-0.18697940	-0.35762940	0.05127330
0.53539560	0.10989010	-0.26175800	-0.86352860	-0.27327070	-0.53091280	-0.58539940	-0.43297540
-0.18222840	-1.40832830	-0.67013490	0.40834440	0.69843520	-0.75504830	-0.05307860	-0.57532310
-0.37443400	0.62931560	-0.13897330	-0.26963560	-0.86342230	-1.19644590	0.18698250	0.75542130
0.11500350	-0.60086600	0.05333810	-1.03953800	-0.18854370	-0.11693800	-1.15273780	-0.95663270
-0.24834440	0.49792530	1.59647670	0.46057980	0.59331180	0.04468280	-0.08039300	1.08170930
0.16567840	-0.13058240	0.10460250	-0.01741550	0.83609130	0.62186910	0.49785410	0.07687050
-0.74500510	-0.61412490	0.11156880	-0.30004290	0.35253650	0.04284120	1.10483040	-0.08470990
-1.33517100	0.37541470	-0.20875010	-0.48810250	0.77078040	0.08691870	1.14633090	0.54949770
-1.21715770	0.14445960	1.12855320	0.64608160	-0.32513550	1.07059220	0.07447860	-0.13230790
0.04884400	0.30919450	0.42991560	-0.04038450	0.03232060	0.28271410	-0.07249300	0.46751570
0.74138340	-0.84553350	0.84484800	0.63419980	-0.39679450	-0.52335570	-0.07067140	-1.08439420
0.72498410	-0.18191880	-1.83042790	0.50851560	-0.20880180	1.08643170	-0.07165030	0.43021030
0.76072460	-0.28019930	-0.99125820	-0.96176590	0.04775930	0.97859810	-1.48912700	-0.47188670
-0.24191600	0.46075500	-1.62536210	0.34353020	-0.62764920	0.42654420	0.22053420	-0.59494700
-0.64092030	0.40522660	0.50242980	0.73758400	0.53693460	-2.01488910	-0.83409030	0.67455030
-0.21775540	0.93167700	-0.04989600	-0.90689740	-0.47858940	0.06749350	-0.49742850	0.41518390

# Exploración mediante visualización

Visualización puede entenderse como un proceso de “**reducción de dimensiones**” de los datos

¿Cómo hacemos **interpretables** a los resultados que nos ofrece la **minería de datos**?, con el ojo humano.

- ✓ ¿Cómo hacemos los datos “digeribles” al cerebro humano?
- ✓ Gráficos para Variables Individuales
- ✓ Gráficos para Pares de Variables
- ✓ Gráficos en Mas de 2 dimensiones

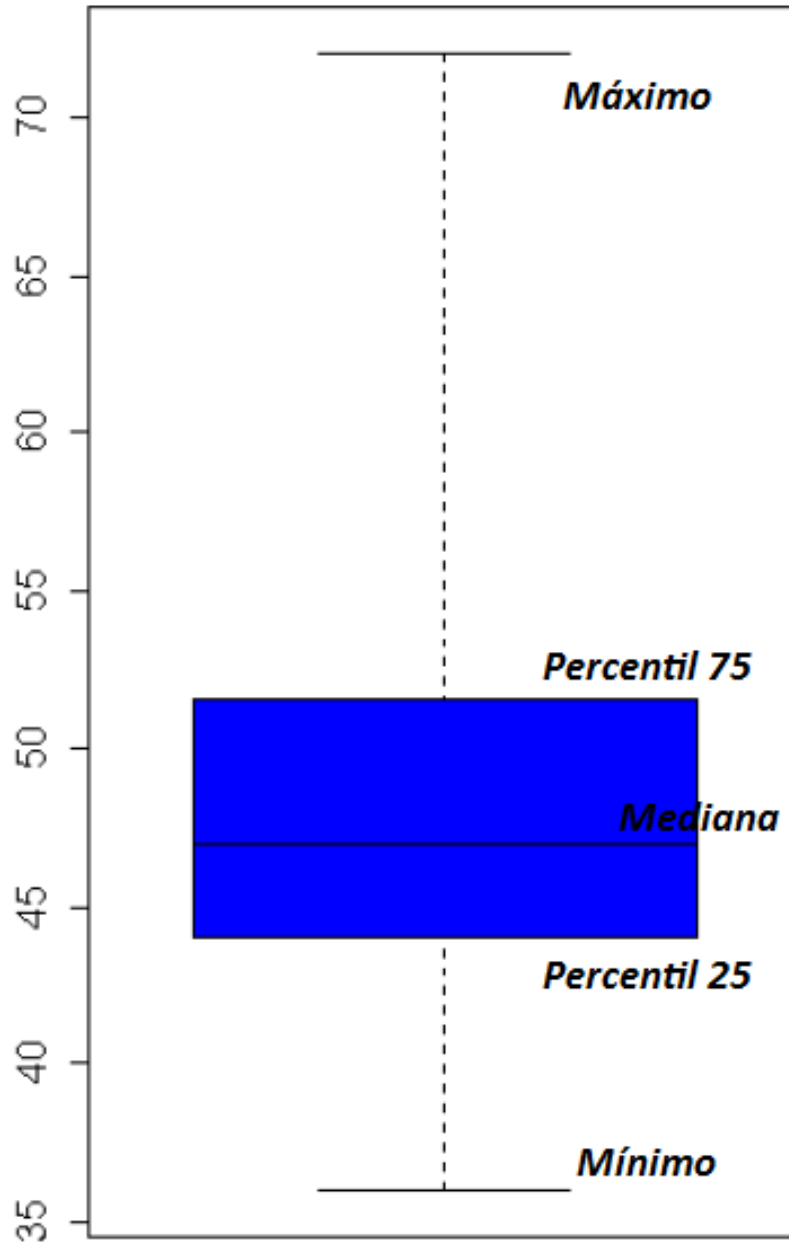
# Boxplots.

## Gráficos para Variables Individuales.

También conocido como el diagrama de los 5 números: mínimo, máximo, mediana, percentil 25, percentil 75.

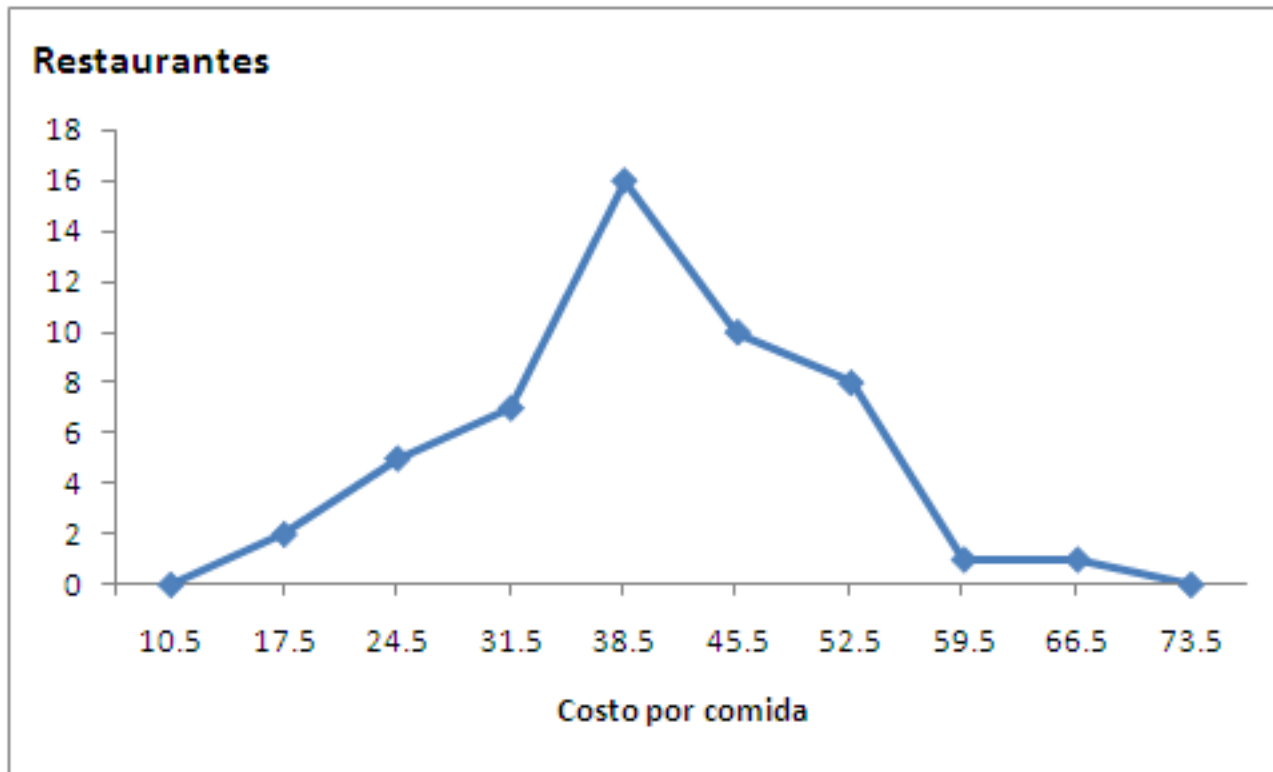
Divide a la muestra (después de ordenarla de mayor a menor) en cuatro grupos.

Para poder construirlo necesitamos definir primero percentil.



# Ojivas

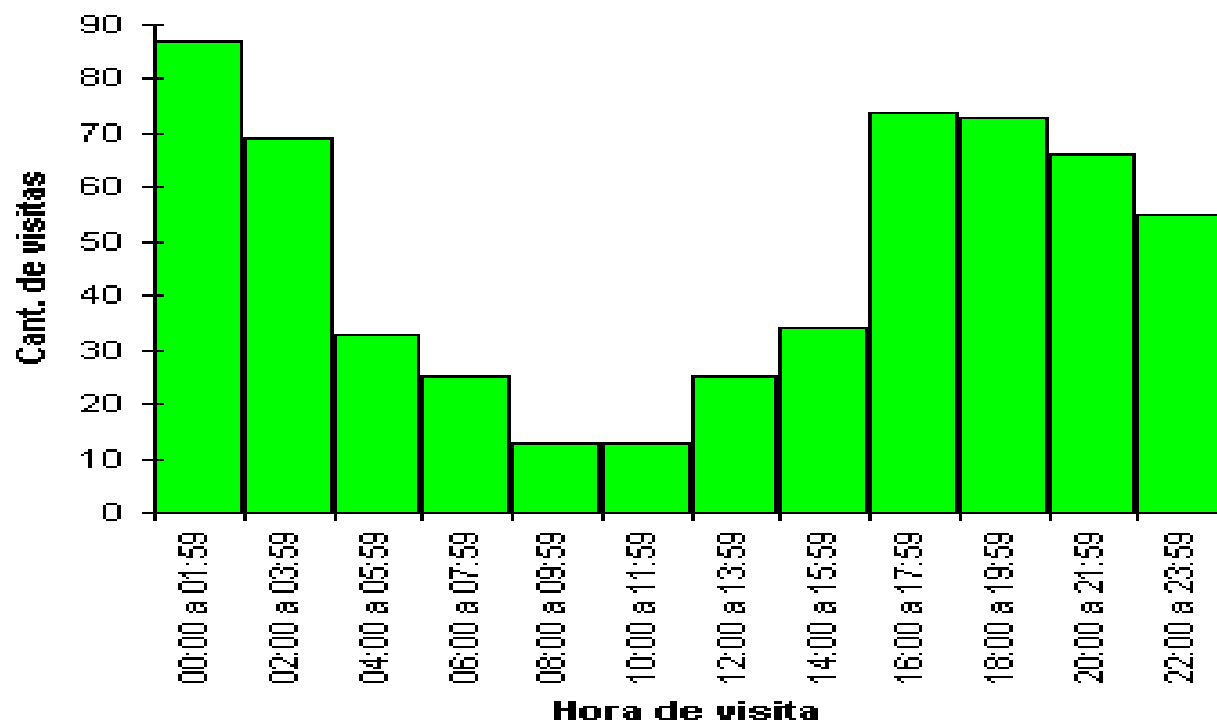
## Gráficos para Variables Individuales



También conocidas como distribuciones acumuladas empíricas, grafican los valores que toma cada una de las observaciones en la muestra contra el percentil que le corresponde a cada observación.

# Histogramas

## Gráficos para Variables Individuales



Son también conocidos como distribuciones empíricas de los datos.

El histograma muestra el número de valores de una variable que cae en intervalos consecutivos.

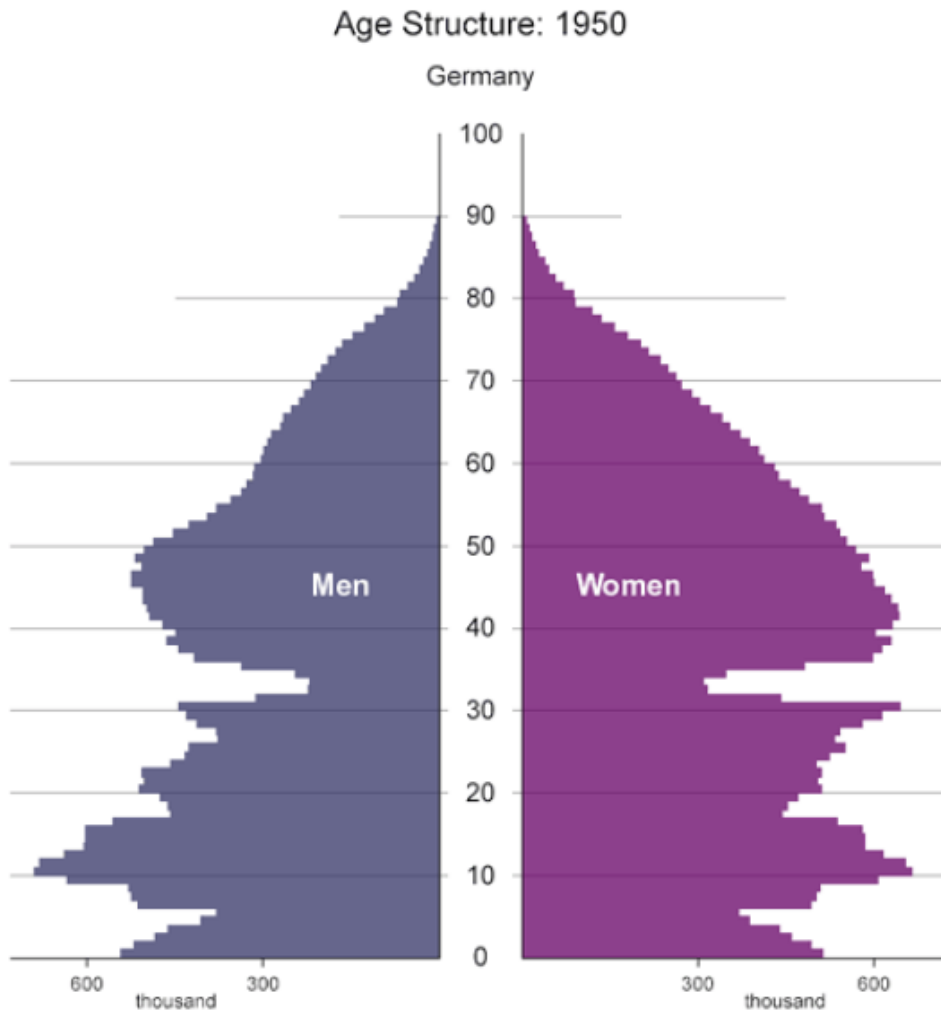
# Exploración mediante visualización

## Gráficos para Variables Individuales (Histogramas)

- ✓ La forma del histograma nos dará información adicional sobre el comportamiento de la variable
- ✓ Tomen por ejemplo la pirámide poblacional de Alemania en 1950 y la proyectada para 2050
- ✓ Ambos son histogramas para edades en el país en dos puntos en el tiempo.

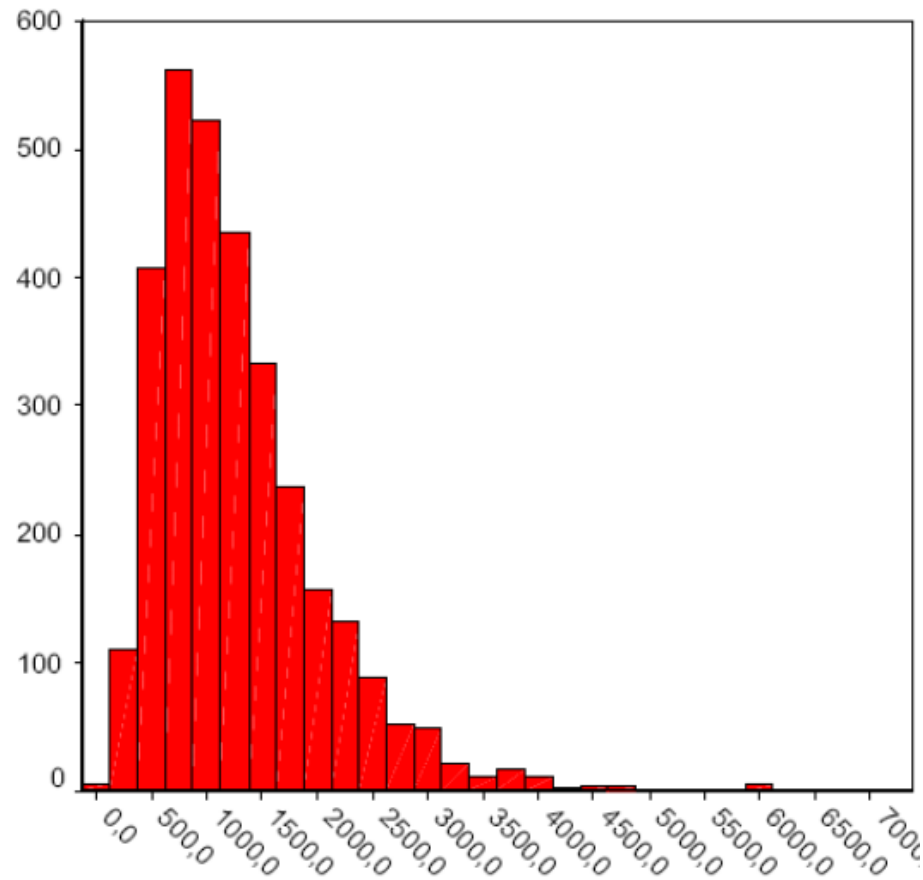
# Exploración mediante visualización

## Gráficos para Variables Individuales (Histogramas)

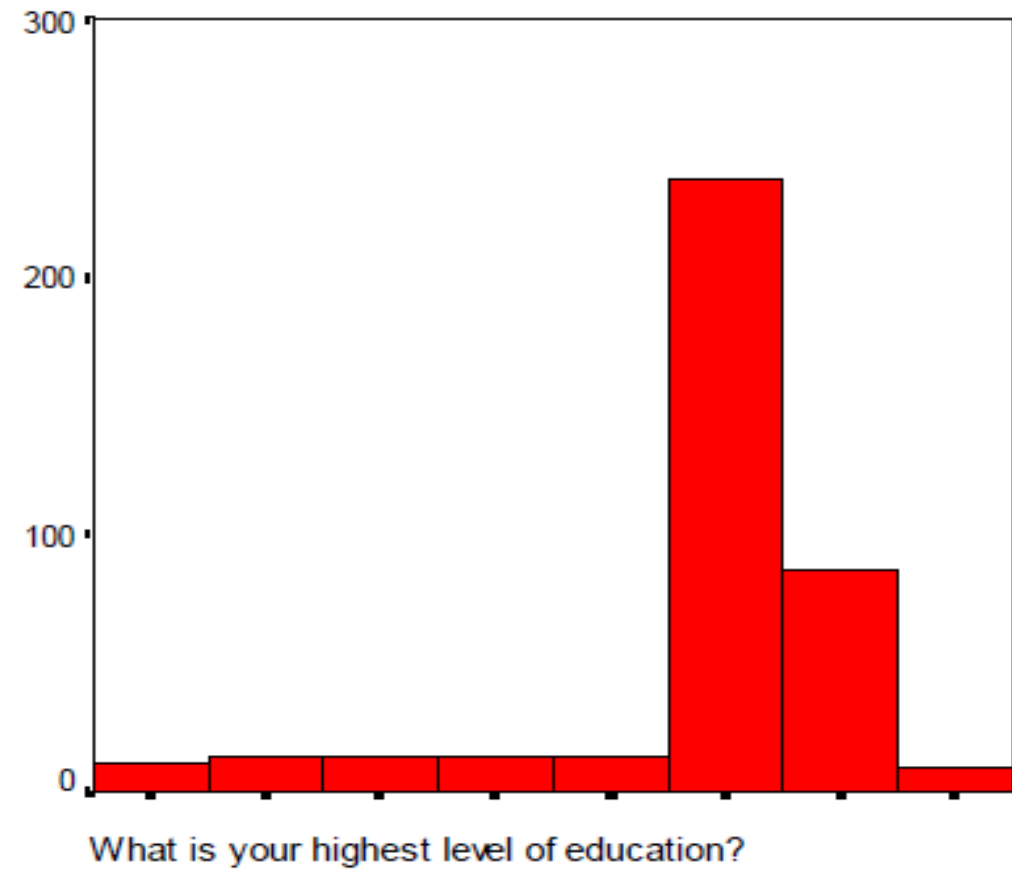


# Exploración mediante visualización

Gráficos para Variables Individuales (Histogramas)



¿Cuánto compran por visita?

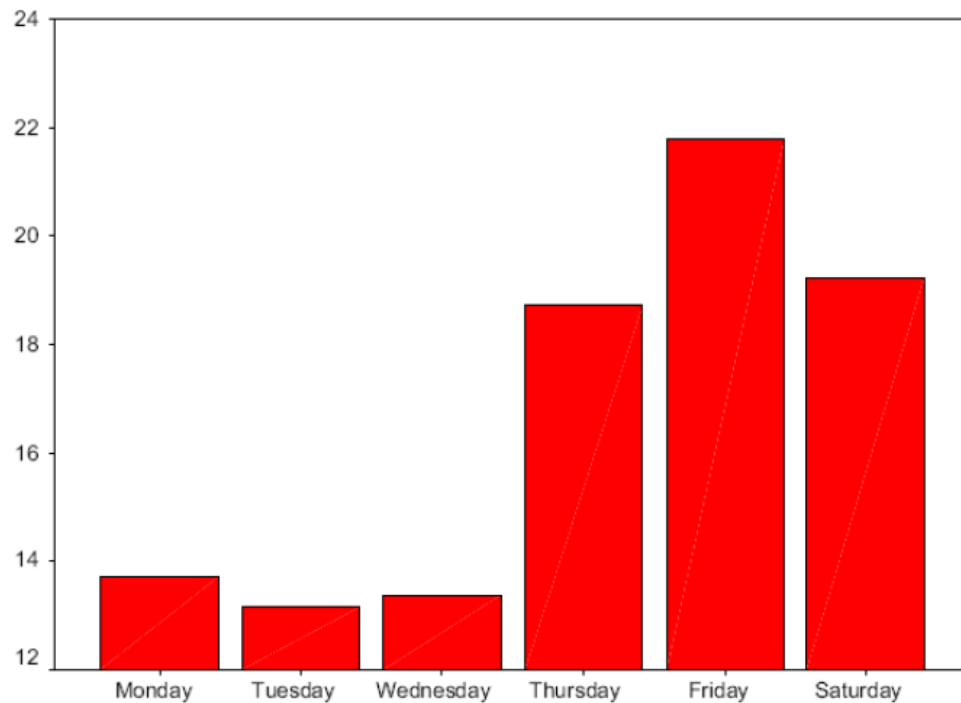


¿Qué nivel de educación tienen?

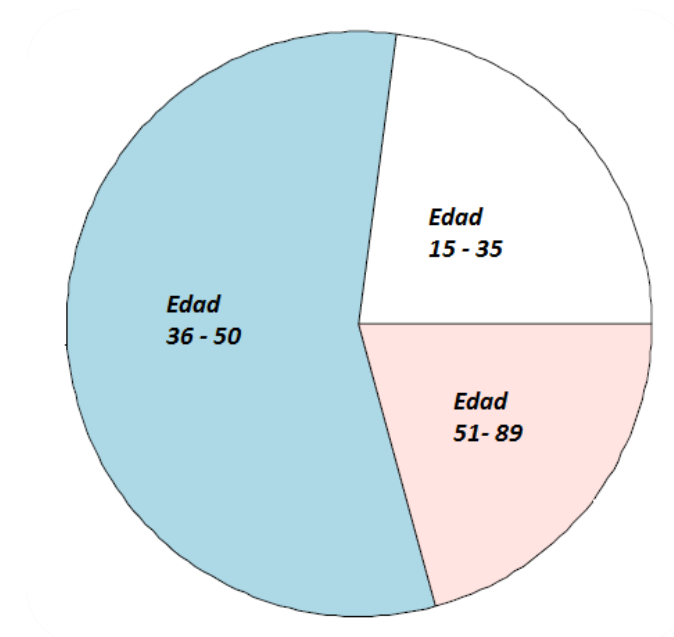


# Exploración mediante visualización

## Gráficos para Variables Individuales (Histogramas y Circular)



¿Qué día nos visita?



¿Qué edades tienen?

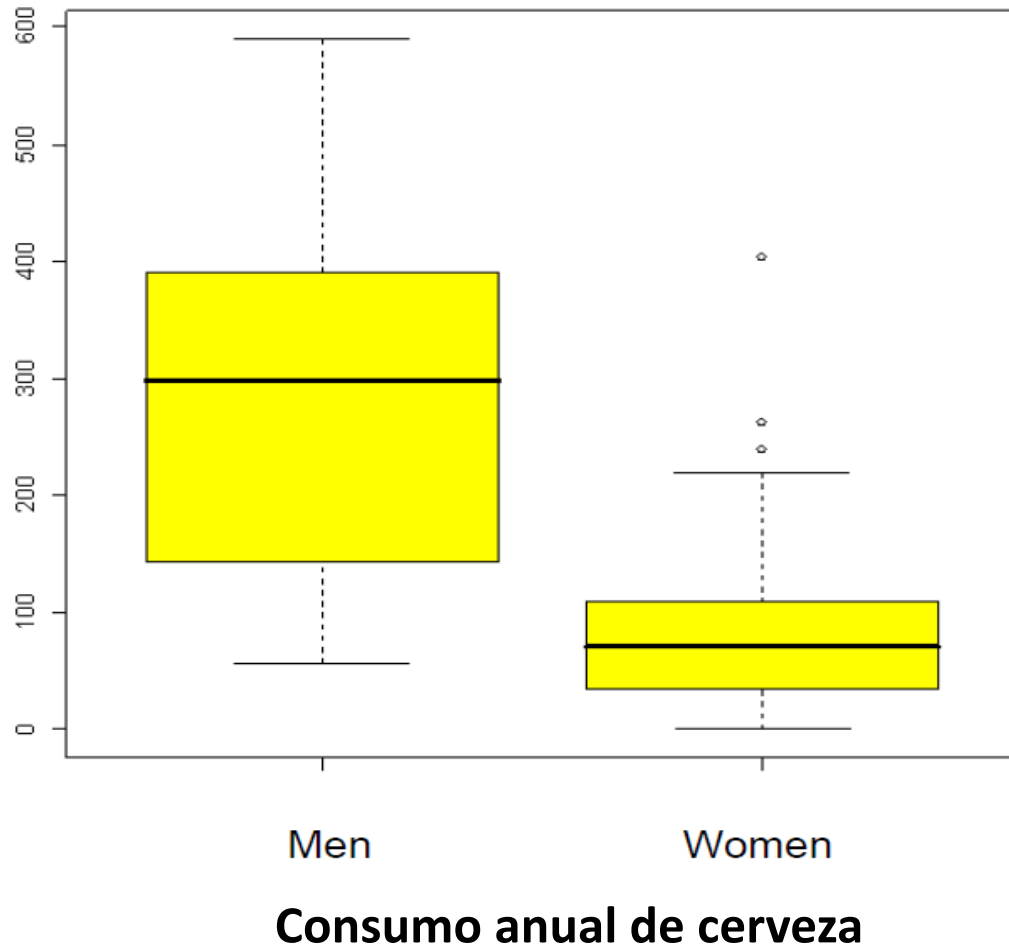
# Exploración mediante visualización

## Gráficos para Pares de Variables

Tipo de Variables	Gráficos	Resumen Numérico
Continua, Categórica	Diagramas de Caja	
Continua, Continua	Scatter Plot Matrix Plots Gráficos de Correlaciones	Covarianza Correlación
Categórica, Categórica	Diagrama de Barras Diagrama de Mosaico	Tablas de Contingencia

# Diagramas de caja.

Gráficos para Pares de Variables (continua–categórica)



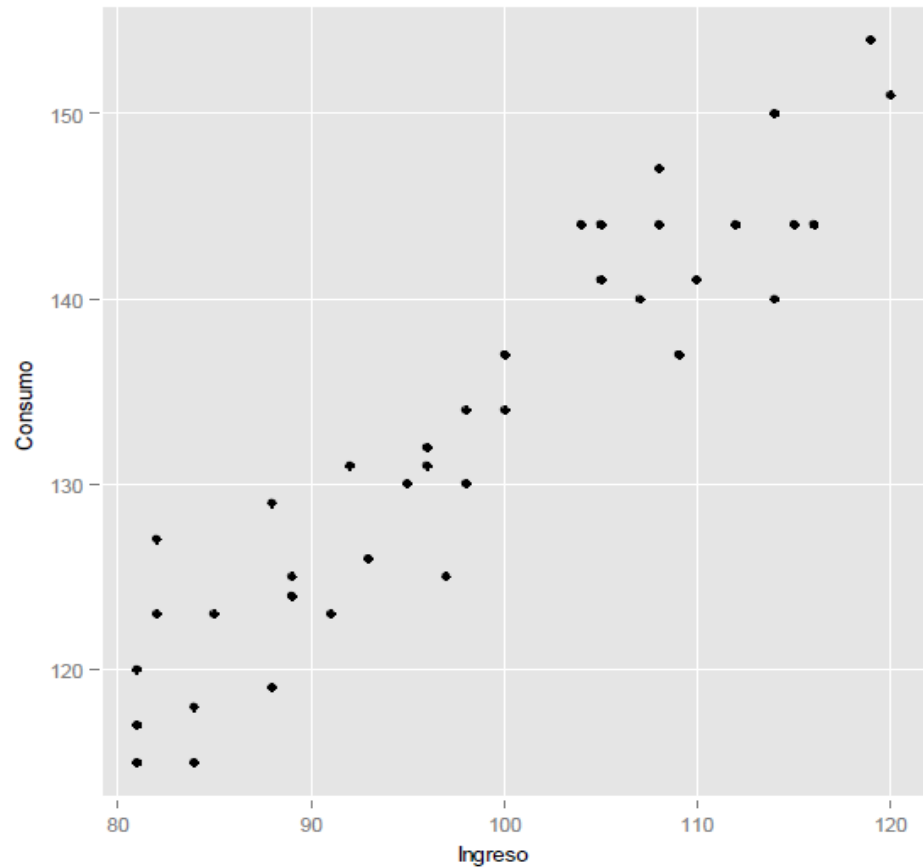
Los diagramas de caja son excelentes gráficos para comparar el desempeño de dos variables.

Queremos comparar el centro del diagrama de caja para las variables comparadas.

Queremos comparar su variabilidad.

# Scatter Plot

## Gráficos para Pares de Variables (continua—continua)



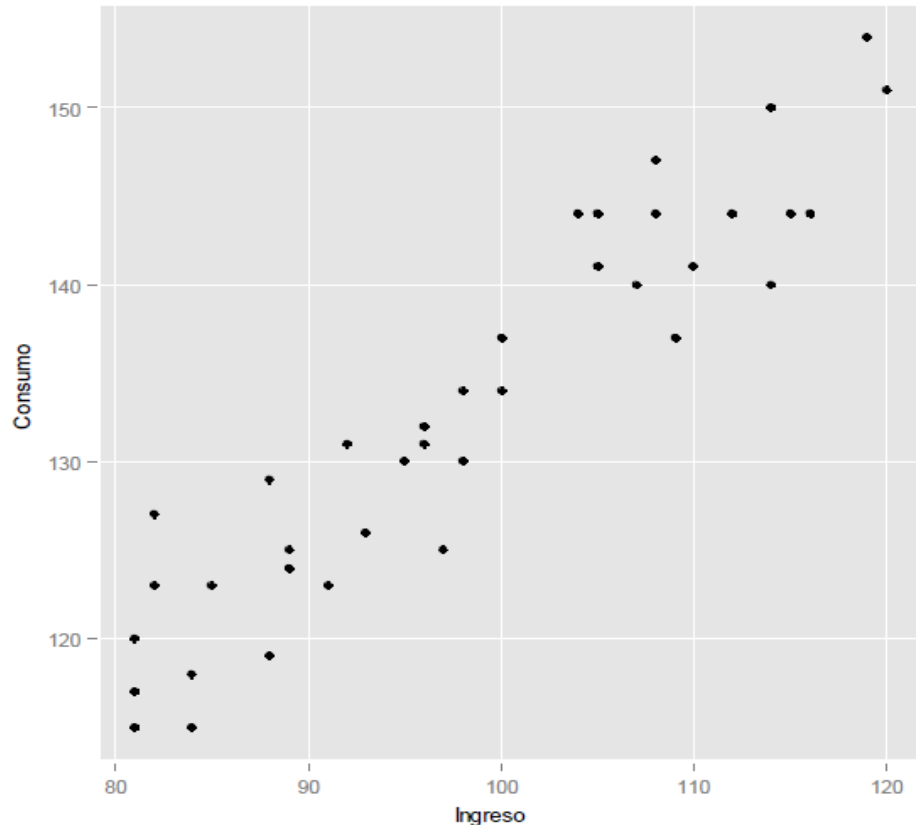
**Ingreso vs. Consumo**

Una vez que hemos descrito cada variable por separado, estamos interesados en descubrir patrones de movimiento entre las variables.

Scatter Plot son útiles para revelar si este movimiento existe entre dos variables.

# Scatter Plot

## Gráficos para Pares de Variables (continua–continua)



**Ingreso vs. Consumo**

Generalmente queremos trazar una línea (la línea de regresión de las próximas clases) a través de los puntos

¿Es la pendiente de esta línea positiva o negativa?

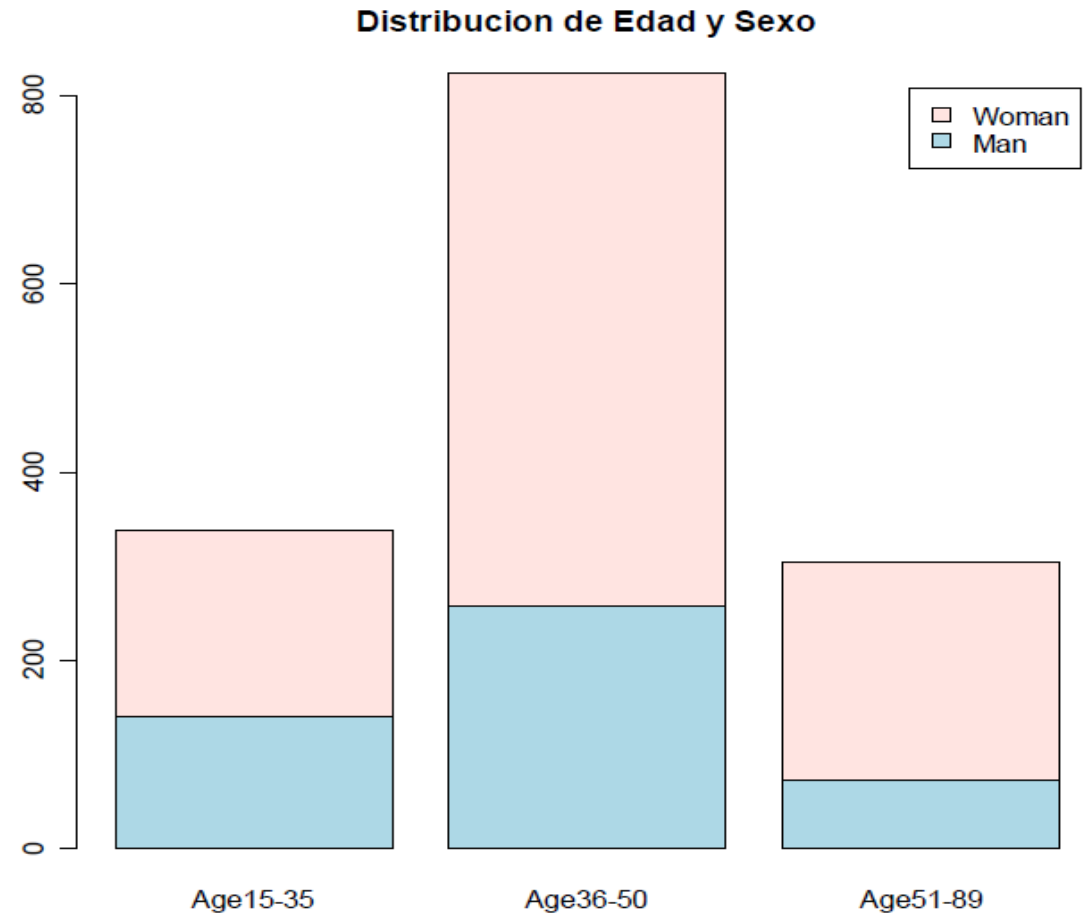
¿Es casi horizontal?

¿Es la relación no-lineal?

# Diagramas de Barras.

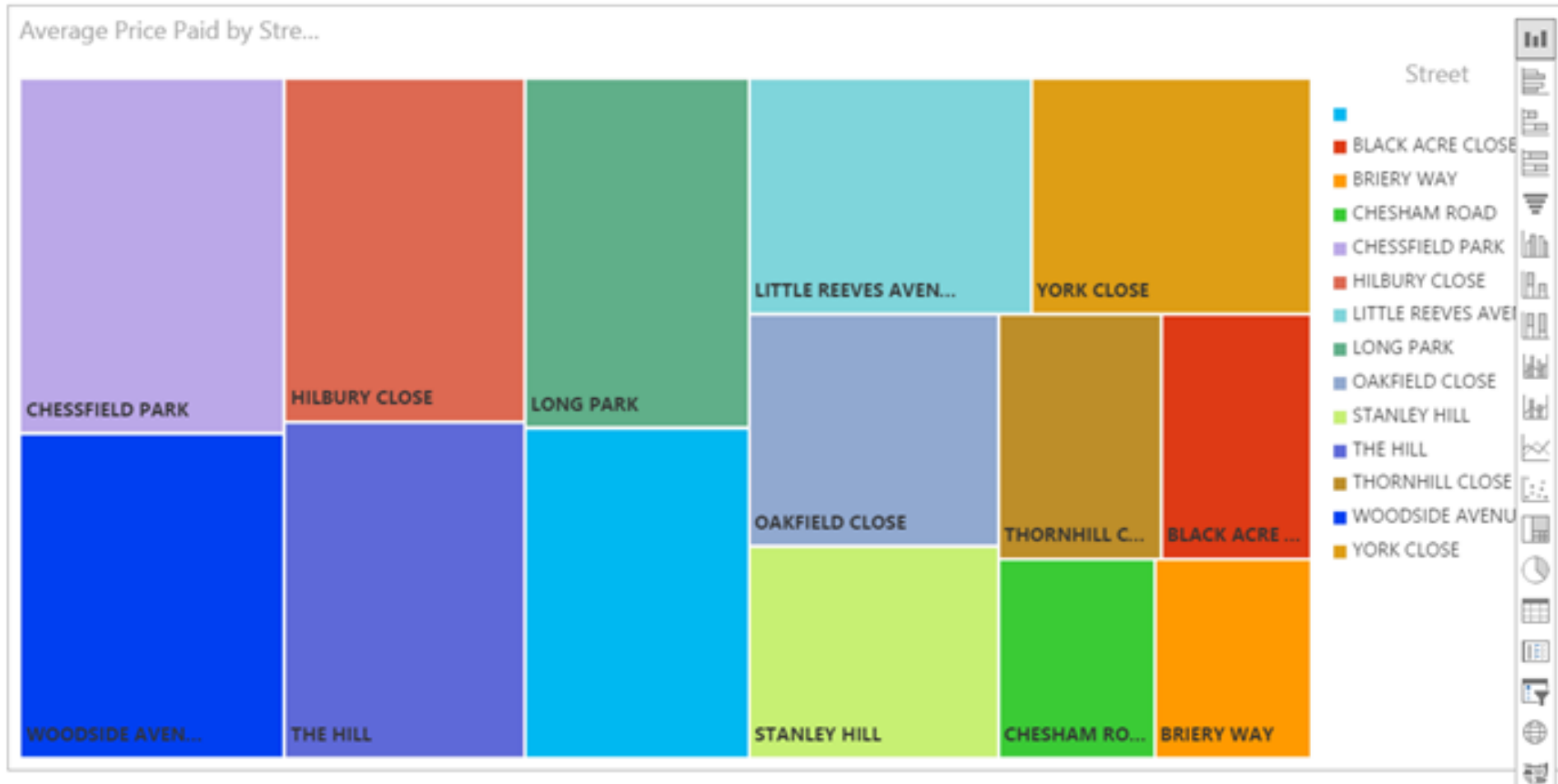
Gráficos para Pares de Variables (categórica-categórica)

	Age15-35	Age36-50	Age51-89	Total
Man	139	257	72	468
Woman	198	566	232	996
Total	337	823	304	1464



# Diagramas de Mosaico

Gráficos para Pares de Variables (categórica-categórica)



# Exploración mediante visualización

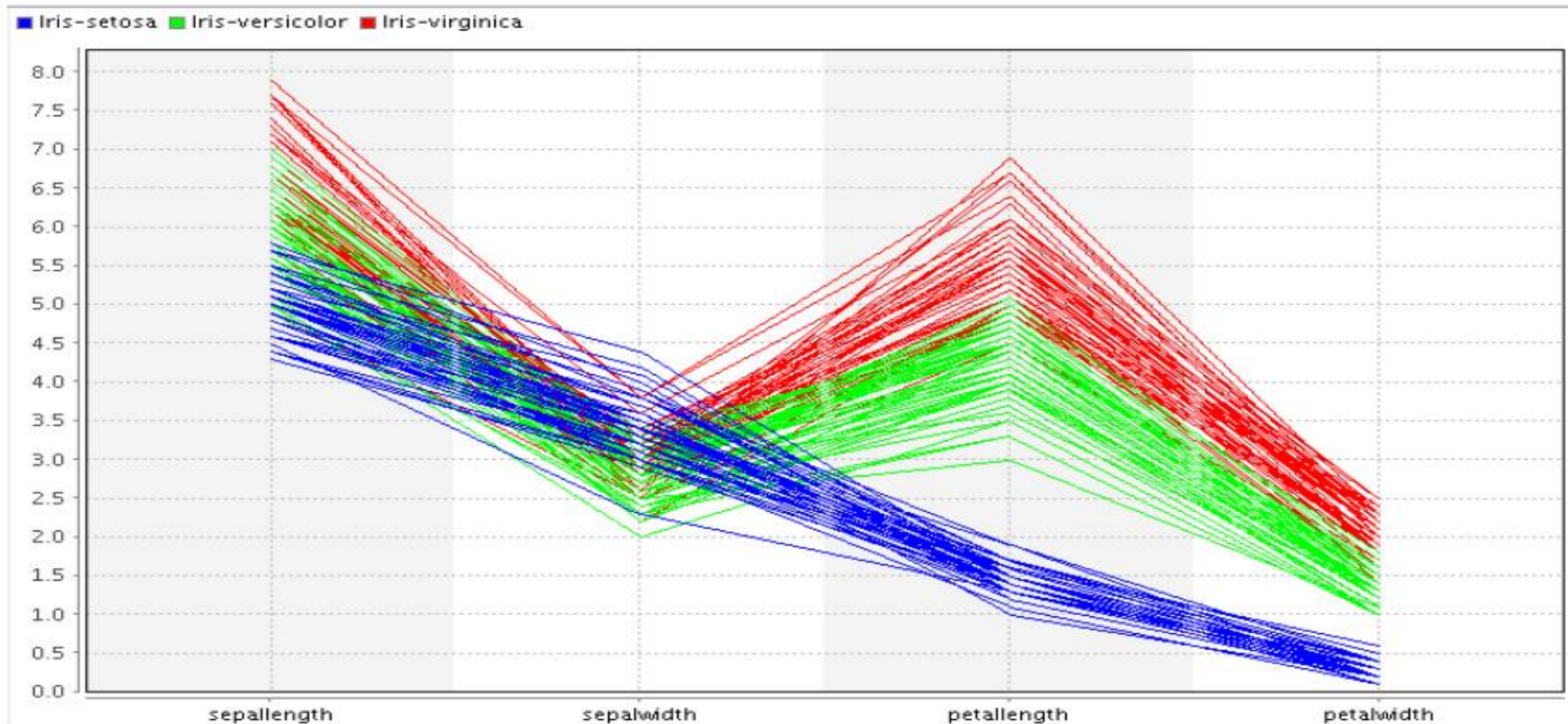
## Gráficos para más de 2 variables

Tipo de Variables	Gráficos	Resumen Numérico
Varias combinaciones	Scatter plots Diagramas de Burbuja Coordinadas Paralelas	



# Coordinadas Paralelas

Gráficos para más de 2 variables.



# Coordinadas Paralelas

**Gráficos para más de 2 variables.**

- ✓ Este grafico nos permitirá representar más de tres dimensiones en un mismo gráfico
- ✓ Cada línea representa una observación
- ✓ Cada eje vertical (hay tantos como variables hay) representa una variable

# Burbujas

Gráficos para más de 2 variables.

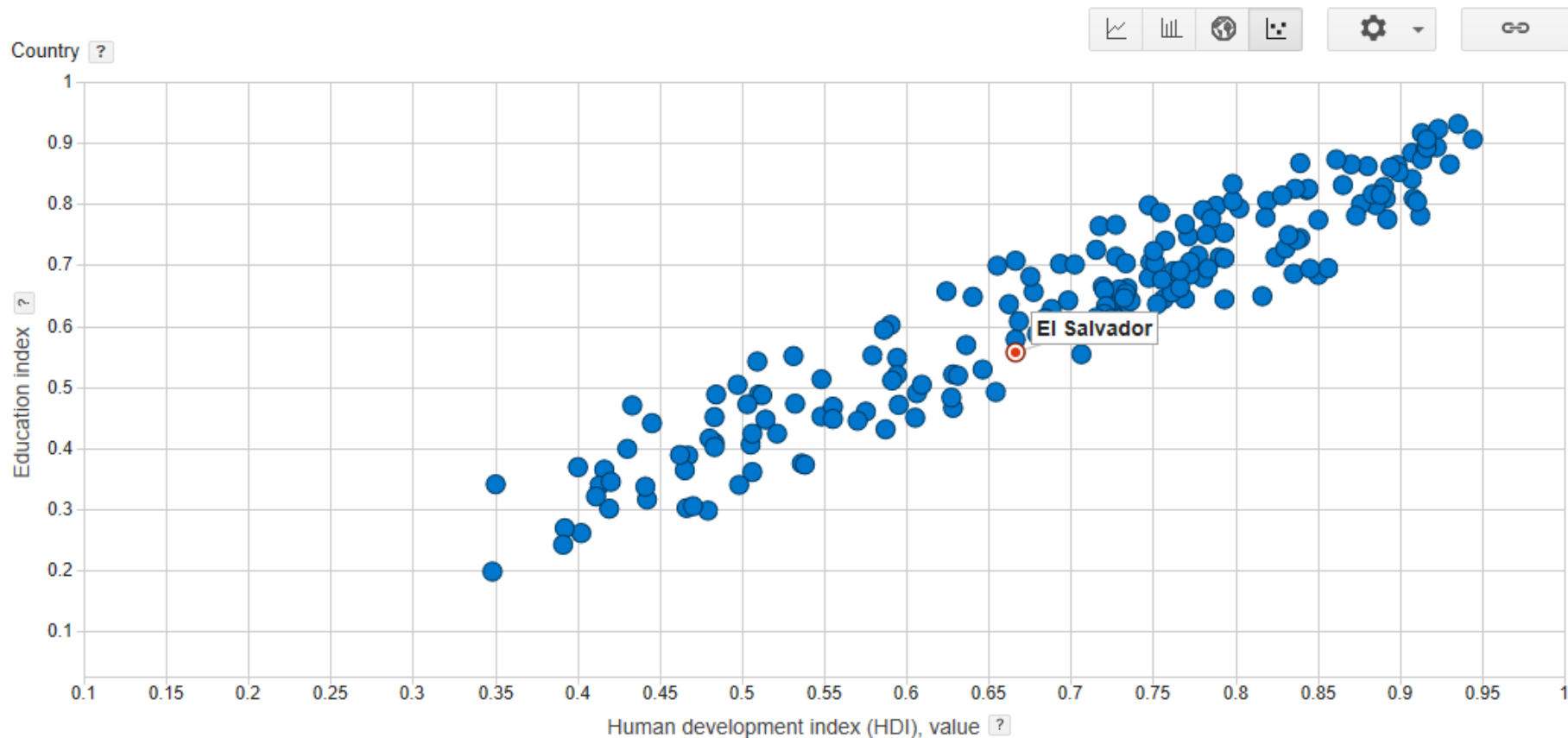
## Public Data

- ▼ Human Development Indicators
  - ▼ Composite indices
    - Human development index (HDI...)
  - ▶ Demography
  - ▼ Education
    - Adult literacy rate (% ages 15 and...)
    - Education index
    - Expected Years of Schooling
    - Mean years of schooling (of adults)
    - Primary school dropout rate (% of...)
  - ▶ Environmental sustainability

Clear

## ▼ Compare by Country

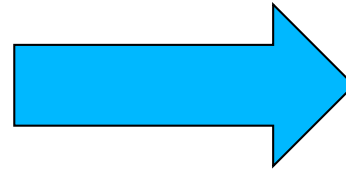
- ☐ Cyprus
- ☐ Czech Republic
- ☐ Denmark
- ☐ Djibouti
- ☐ Dominica
- ☐ Dominican Republic
- ☐ Ecuador
- ☐ Egypt
- ☒ El Salvador
- ☐ Equatorial Guinea



# Exploración y selección

## Selección de características

CLIENTE
Código
DUI
Nombres
Apellidos
Ciudad
Teléfono
FechaNacimiento
TipoLicencia
FechaLicencia
MatrículaVehículo
TipoVehículo



CLIENTE
Ciudad
FechaNacimiento
TipoLicencia
FechaLicencia
TipoVehículo

# Exploración y selección

## Análisis correlacional

	Edad	Tensión	Obesidad	Colesterol	Tabaquismo	Alcoholismo	Pulsaciones	Hierro
Edad		0.63	0.34	0.42	-0.02	0.15	0.12	-0.33
Tensión	0.63		0.22	0.56	0.72	0.43	0.27	-0.08
Obesidad	0.34	0.22		0.67	0.72	0.32	0.32	-0.21
Colesterol	0.42	0.56	0.67		0.52	0.27	0.4	0.45
Tabaquismo	-0.02	0.72	0.72	0.52		0.58	0.39	-0.12
Alcoholismo	0.15	0.43	0.32	0.27	0.58		0.23	-0.22
Pulsaciones	0.12	0.27	0.32	0.4	0.39	0.23		-0.15
Hierro	-0.33	-0.08	0.21	0.45	-0.12	-0.22	-0.15	

Selección a través de algoritmos: Featured Selection

Algoritmo de fuerza bruta

# Librerías para visualizar datos.

<http://polymaps.org/>

<http://www.flotcharts.org/>

<https://d3js.org/>

<https://carto.com/>

<https://developers.google.com/chart/>

<http://flare.prefuse.org/>

<http://arborjs.org/>

<http://visualizefree.com/>

Gracias