



# Data-Mining

Ingeniería Informática  
Departamento de Lenguajes y  
Sistemas Informáticos



## Miembros del grupo:

- José Carlos Pavón Montañez
- Pablo Pérez Luna

## Práctica 1.1

*Abra el fichero base\_datos\_mineria.arff y corrija los errores que existentes. Enumere a continuación cada uno de los errores que ha encontrado, a qué se debían y cómo los ha corregido.*

- *Error 1: No estaba comentado la linea “Base de datos de alumnos”, hemos añadido %*
- *Error 2: En la linea 3 se ha añadido el carácter @ y corregido relacion por relation*
- *Error 3: En la linea 5 falta el tipo STRING*
- *Error 4: En la linea 6 falta el nombre del atributo (hemos puesto sexo)*
- *Error 5: En la linea 7 cambiamos Entero por Integer*
- *Error 6: En la linea 8 falta el nombre del atributo (hemos puesto peso)*
- *Error 7: En la linea 11 cambiamos date por data*
- *Error 8: En la linea 13 el valor V no está contemplado en sexo, ponemos H*
- *Error 9: En las lineas 13, 14, 17, 20 y 41 al campo nombre le faltaba un carácter \_ o bien suprimir de alguna manera el espacio en blanco.*
- *Error 10: Omisiones de datos en las lineas 16, 34 y 39, ponemos ?.*

*Una vez corregido el fichero de datos, abra el fichero con el interfaz gráfico Explorer y conteste a las siguientes preguntas:*

- *¿Hay algún atributo que no pueda visualizar? Indique cual y cómo lo podría eliminar sin borrarlo del fichero .arff*

Respuesta:

El atributo nombre que es de tipo string, y por tanto no podemos hacer el histograma de los datos. Para eliminarlo hacemos click en el campo check asociado al atributo nombre y a continuación en el boton REMOVE.

- *¿Qué utilidad tiene el cuadro combinado que aparece a la derecha (junto al botón etiquetado como “Visualize All”)?*

Respuesta:

El combobox sirve para clasificar una distribución según de la clase indicada por el combobox. En este caso sería conveniente que la clase seleccionada tuviera valores discretos, para poder clasificar de manera correcta.

- *Indique el n° de instancias y atributos*

Respuesta:

Instancias hay 33 y Atributos 5 (si contamos el que hemos eliminado previamente).

- *Haga una descripción lo más detallada posible de cada atributo (no contar el que ha sido eliminado en el primer punto): nombre, tipo, n° de valores omitidos, n° de valores que no se repiten, ...*

Respuesta:

Atributo **sexo**,

Tipo nominal (enumerado)

Omitidos: 2

Distintos: 2

Únicos: 0

Atributo **edad**,

Tipo numérico

Omitidos: 2

Distintos: 10

Únicos: 6

Máx: 42

Min: 21

Media: 23.93

Desviación Típica: 4.65

Atributo **peso**,

Tipo numérico

Omitidos: 0

Distintos: 24

Únicos: 18

Máx: 110

Min: 50

Media: 71.74

Desviación Típica: 13.7

Atributo **estatura**,

Tipo numérico

Omitidos: 1

Distintos: 19

Únicos: 9

Máx: 2.05

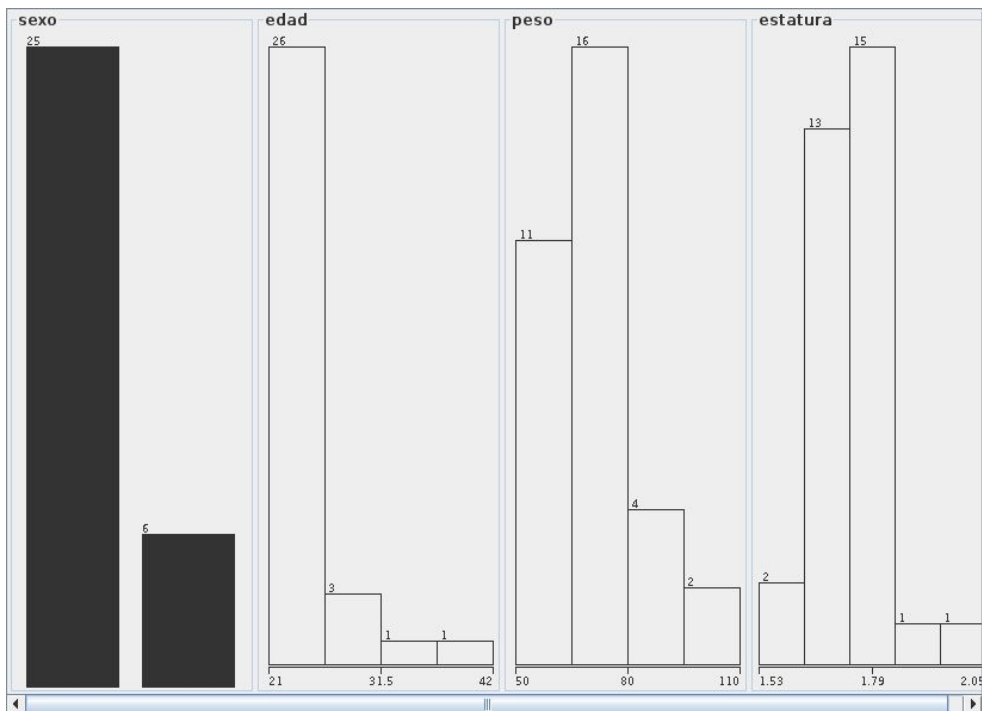
Min: 1.53

Media: 1.64

Desviación Típica: 0.094

- *Muestre gráficamente todos los valores de los atributos (sin hacer referencia a ningún atributo como clase)*

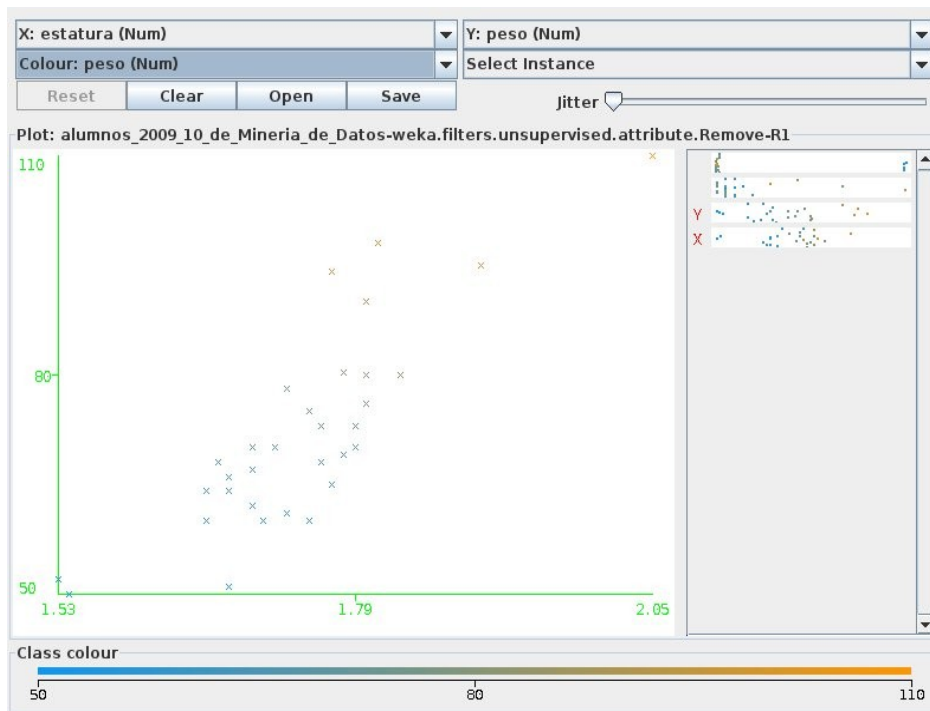
Respuesta:



- *Visualice si hay algún tipo de relación entre pares de atributos. Si encuentra alguna, inserte el gráfico donde se muestra y descríbala.*

Respuesta:

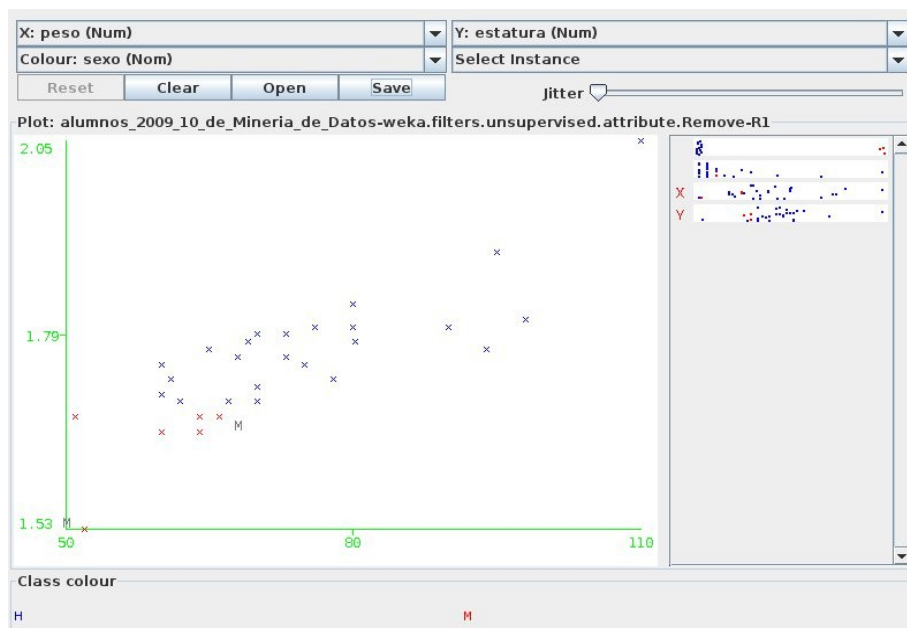
Podemos observar que el peso y la estatura crecen de manera directamente proporcional, es decir a mayor estatura el sujeto tiene más peso. Podemos observarlos en la siguiente gráfica:



- Inserte una gráfica en la que se refleje el peso en el eje X, la estatura en el eje Y y el color corresponda al sexo. ¿Qué conclusiones podemos obtener de la observación de esta gráfica?

Respuesta:

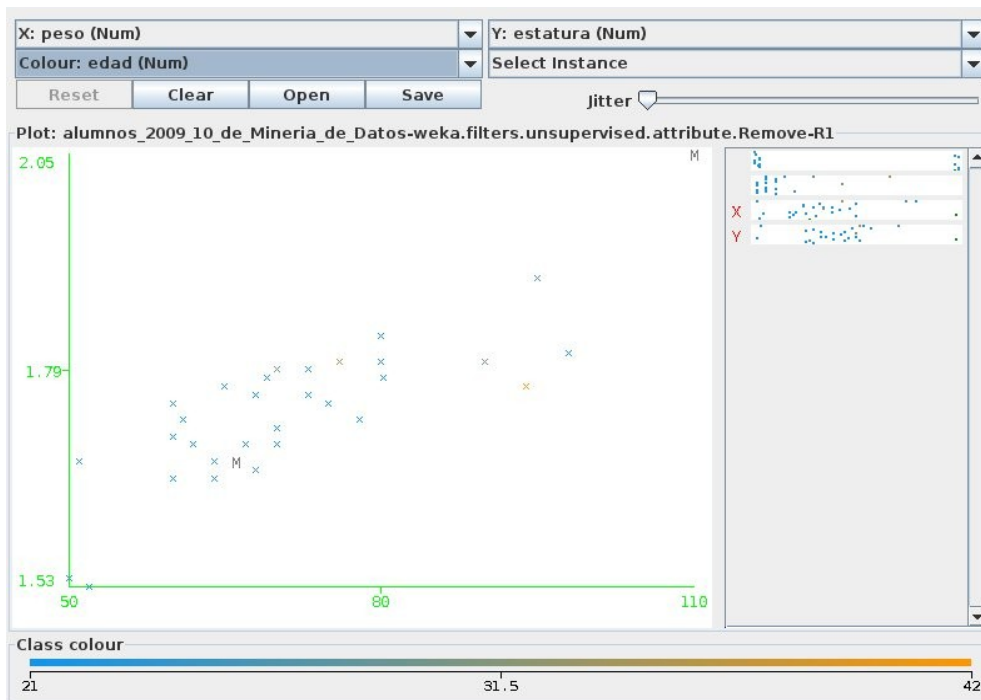
Podemos observar de la gráfica, que de la muestra obtenida, los sujetos de sexo femenino tienen menor estatura y menor peso. (Mujeres Rojo, Hombre Azul)



- Repita el paso anterior mostrando como color la edad.

Respuesta:

De esta gráfica podemos extraer que la mayoría de sujetos son de la misma edad exceptuando algunos sujetos de mayor edad de manera excepcional.



- Queremos guardar en un fichero .arff los datos de los alumnos que pesan menos de 80 kg. ¿Cómo podemos hacerlo? Describa detalladamente la secuencia de pasos.

Respuesta:

- 1 – Vamos al entorno proporcionado por la pestaña VISUALIZE.
- 2 – Seleccionamos la relación Peso-Peso (ya que es el atributo a extraer).
- 3 – En el combobox que tiene el valor por defecto “Select Instance”, seleccionamos la opción Rectangle.
- 4 – Hacemos la selección de aquellos sujetos que cumplen la condición.
- 5 – Hacemos click en el botón SUBMIT para quedarnos solo con los sujetos que hemos seleccionado.
- 6 – Por último hacemos click en el botón SAVE y ya tenemos los datos guardados.