[Weka]
連續型資料轉離散型資料

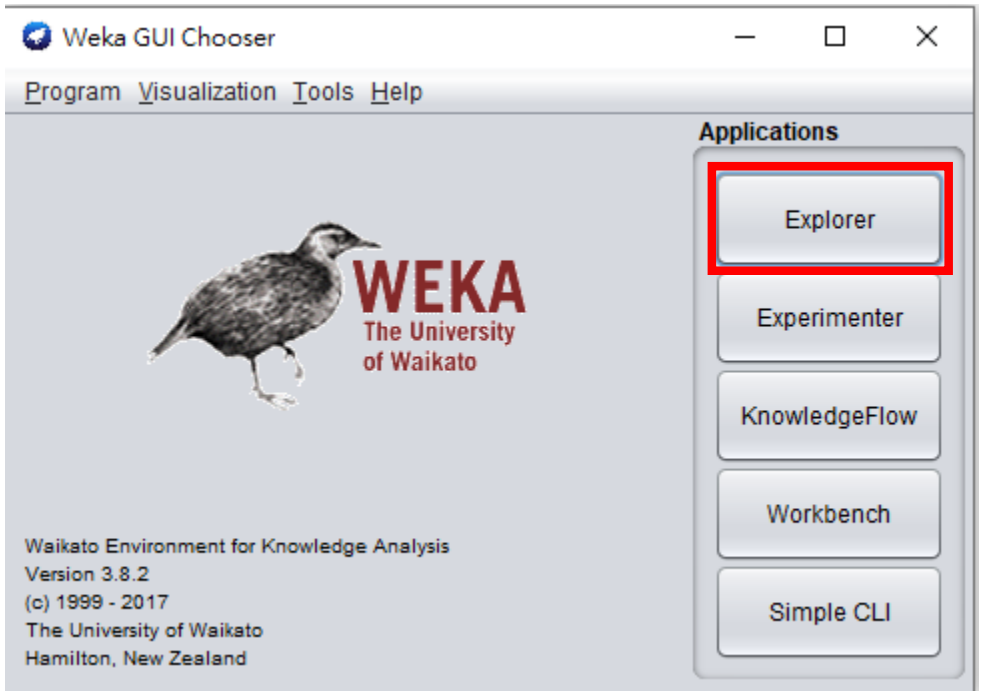
國立臺中科技大學資訊工程系
張家瑋 博士

Step 1

	A	B	C	D
1	Age	Marriage	income	buy_computer
2	24	single	High	no
3	28	single	mediate	no
4	35	single	low	yes
5	32	Married	mediate	no
6	40	Married	low	no
7	42	Married	low	no
8	38	Married	mediate	no
9	29	single	High	no
10	22	Married	low	no
11	33	Married	mediate	no
12	25	Married	High	yes
13	50	Married	mediate	no
14	35	single	mediate	yes
15	45	Married	low	no
16	37	single	mediate	yes
17	18	single	low	no

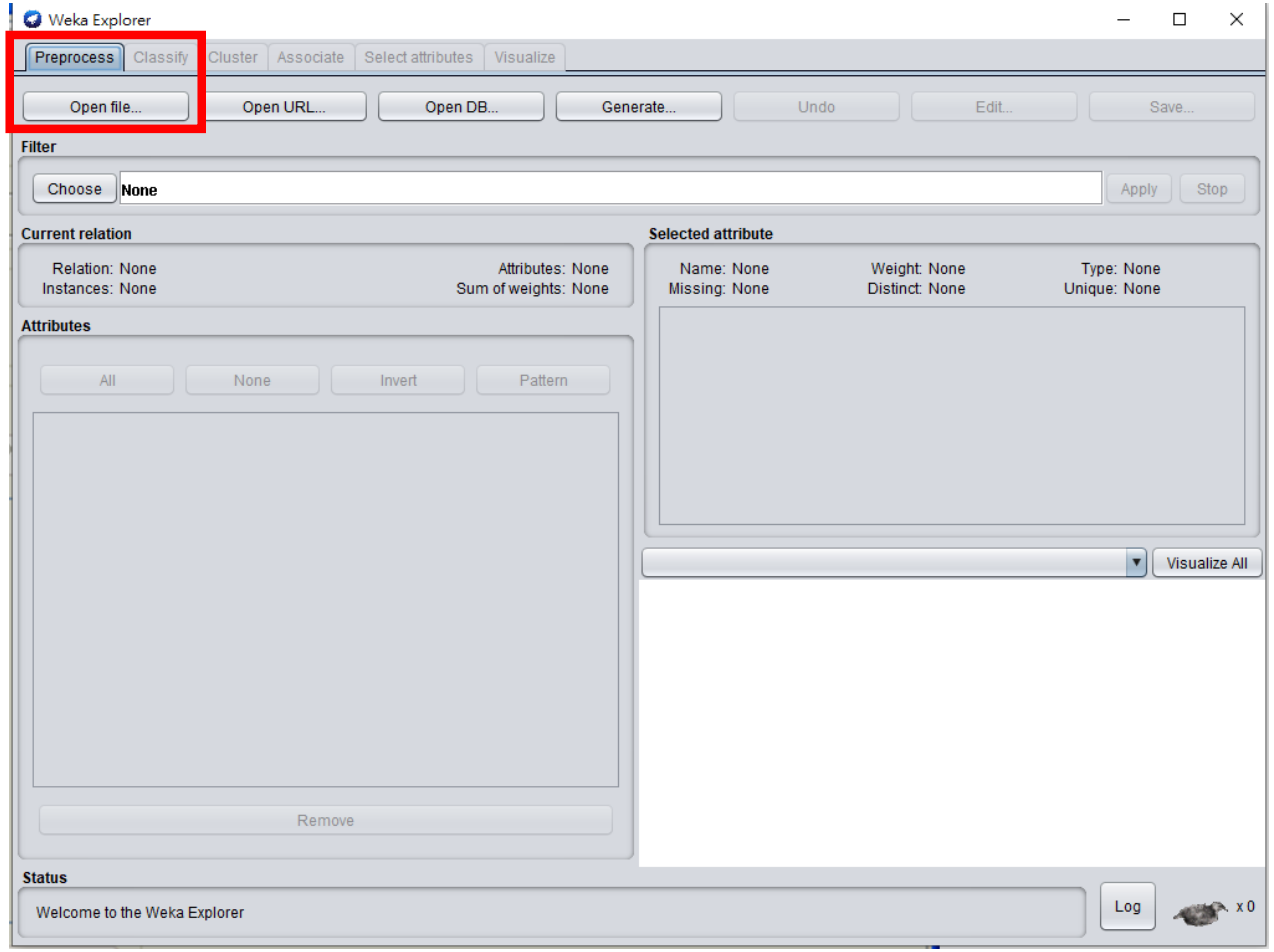
- 為了使用特定分類演算法，如決策樹系列方法時，較好的輸入的資料型態是離散型資料。
- 年齡是連續數值，如何轉成離散型資料呢？

Step 2



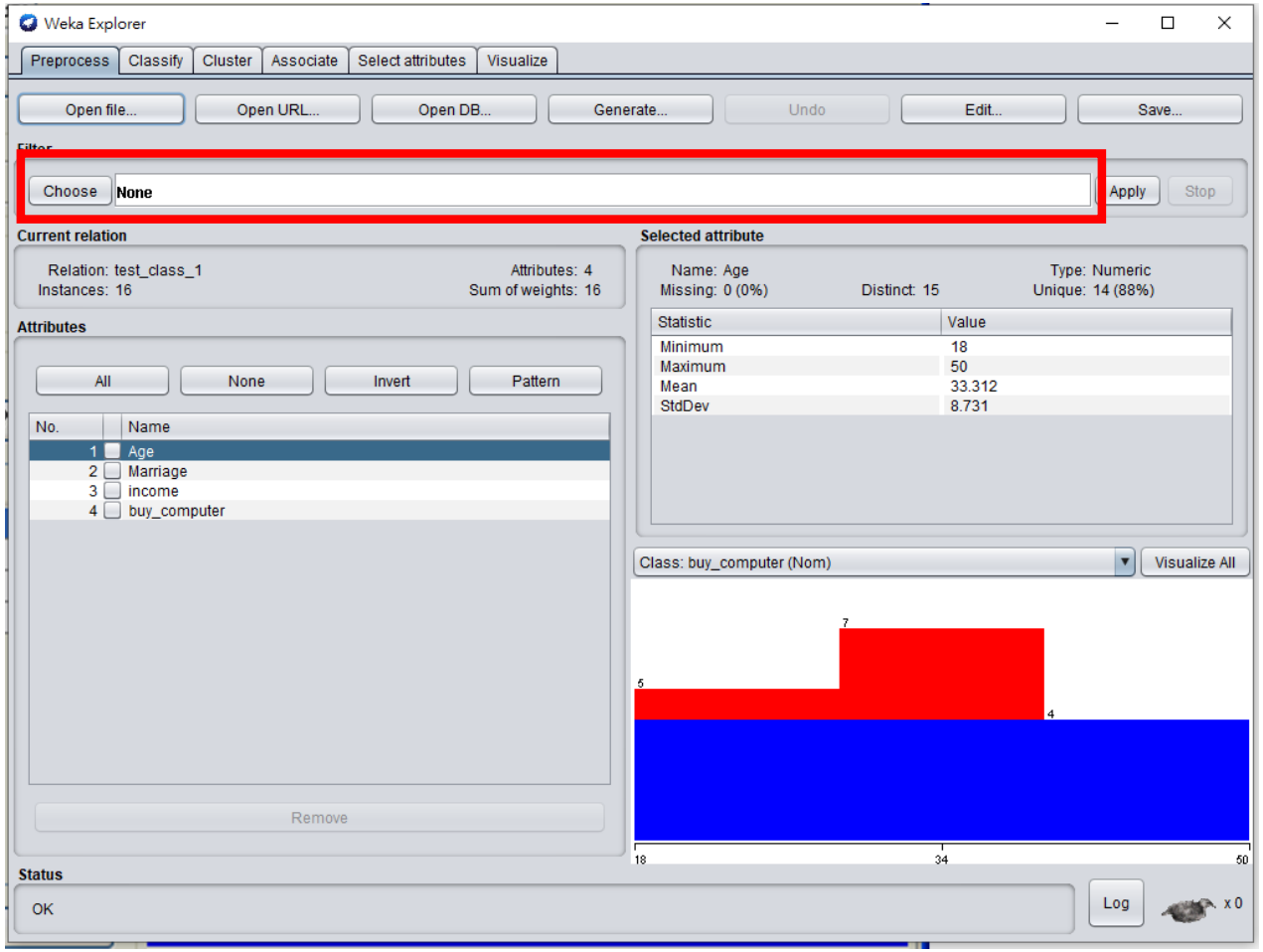
- 準備好 CSV 或 ARFF 格式的檔案。
- Weka 啟動後，點選 Explorer。

Step 3



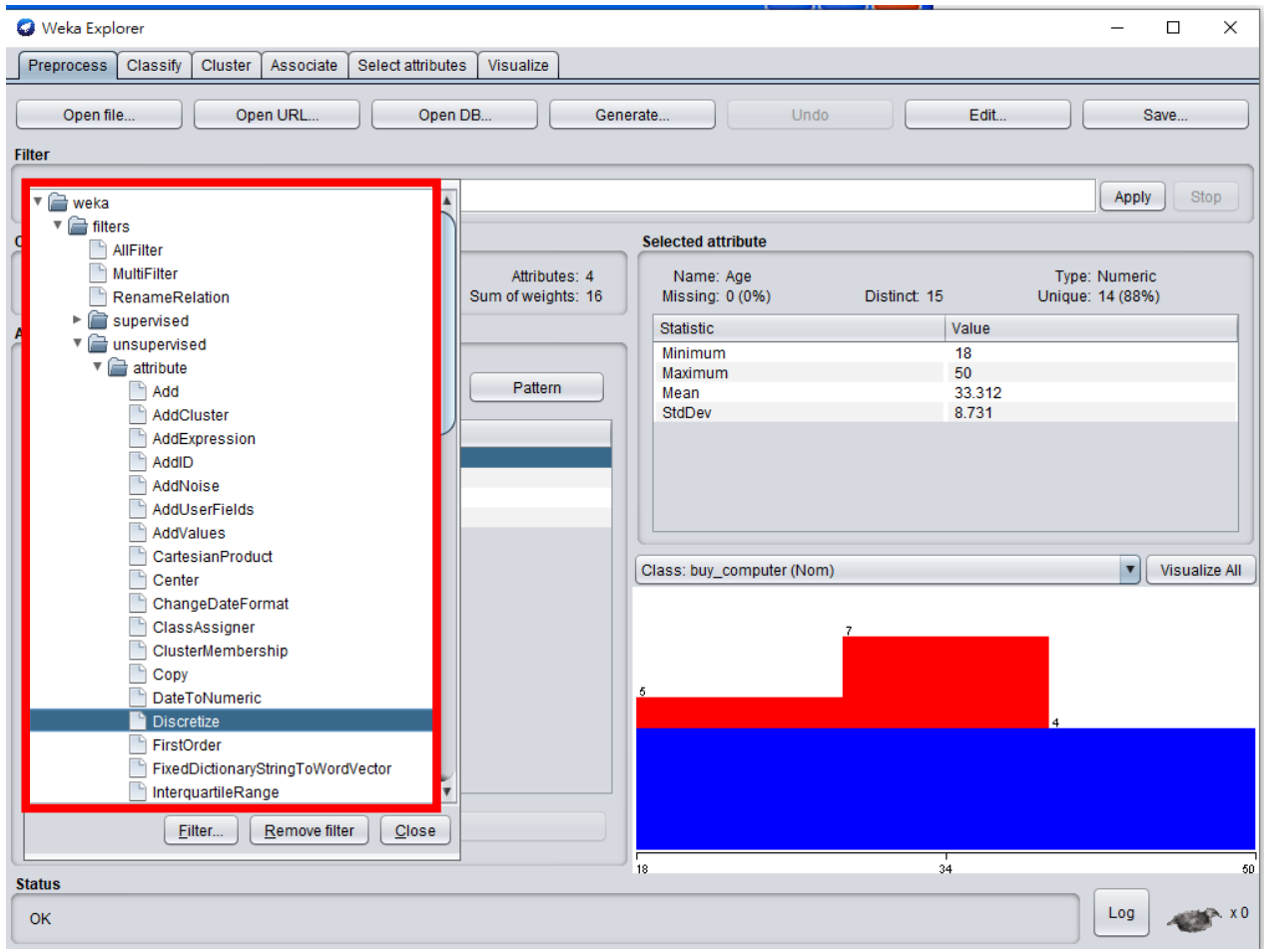
- 選擇 Preprocess。
- 點擊 Open files ，選取欲處理的 csv, arff 資料集。

Step 4



- Choose 前處理方法

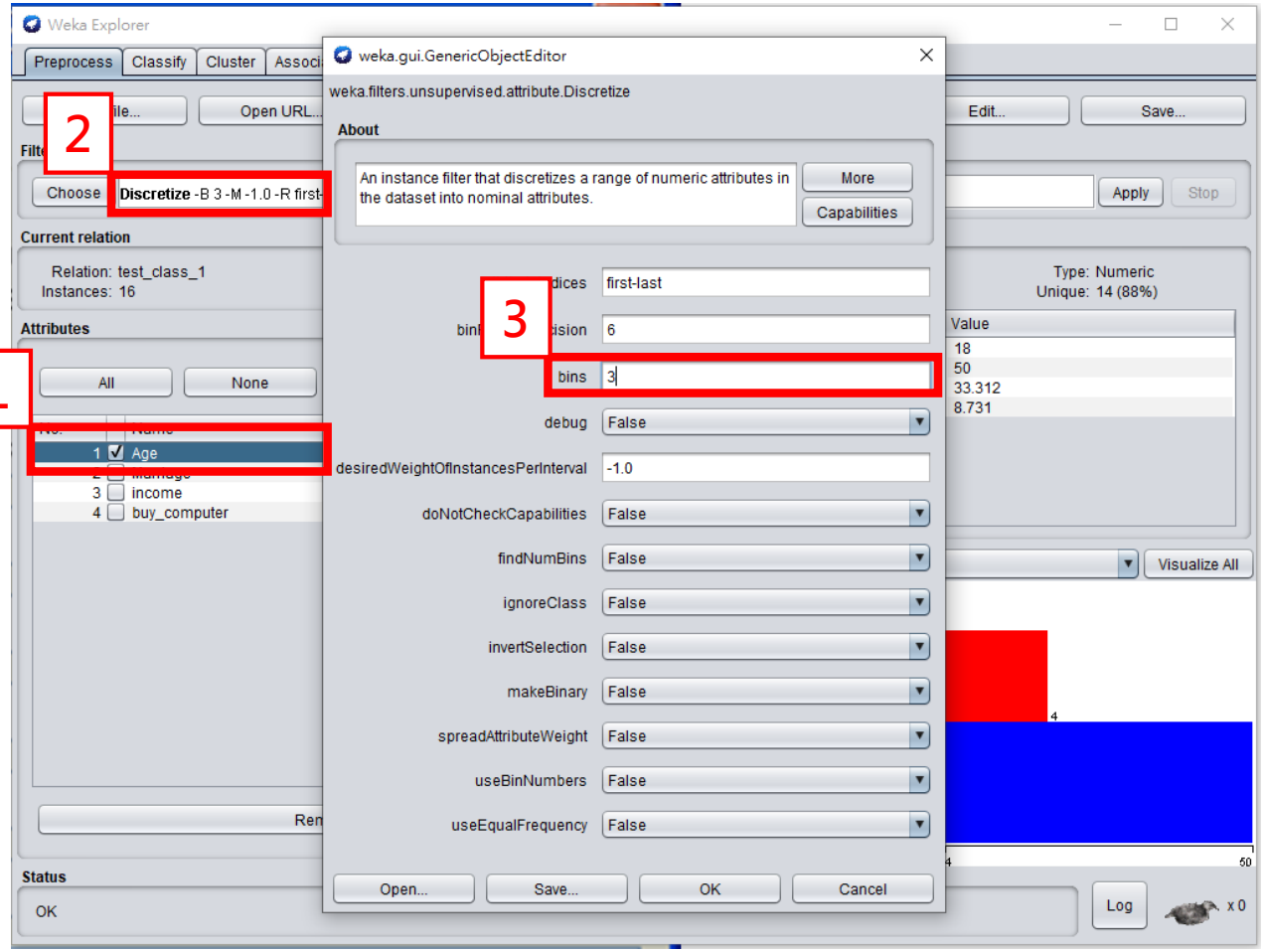
Step 5



- 請依照下列階層找到 Discretize 方法

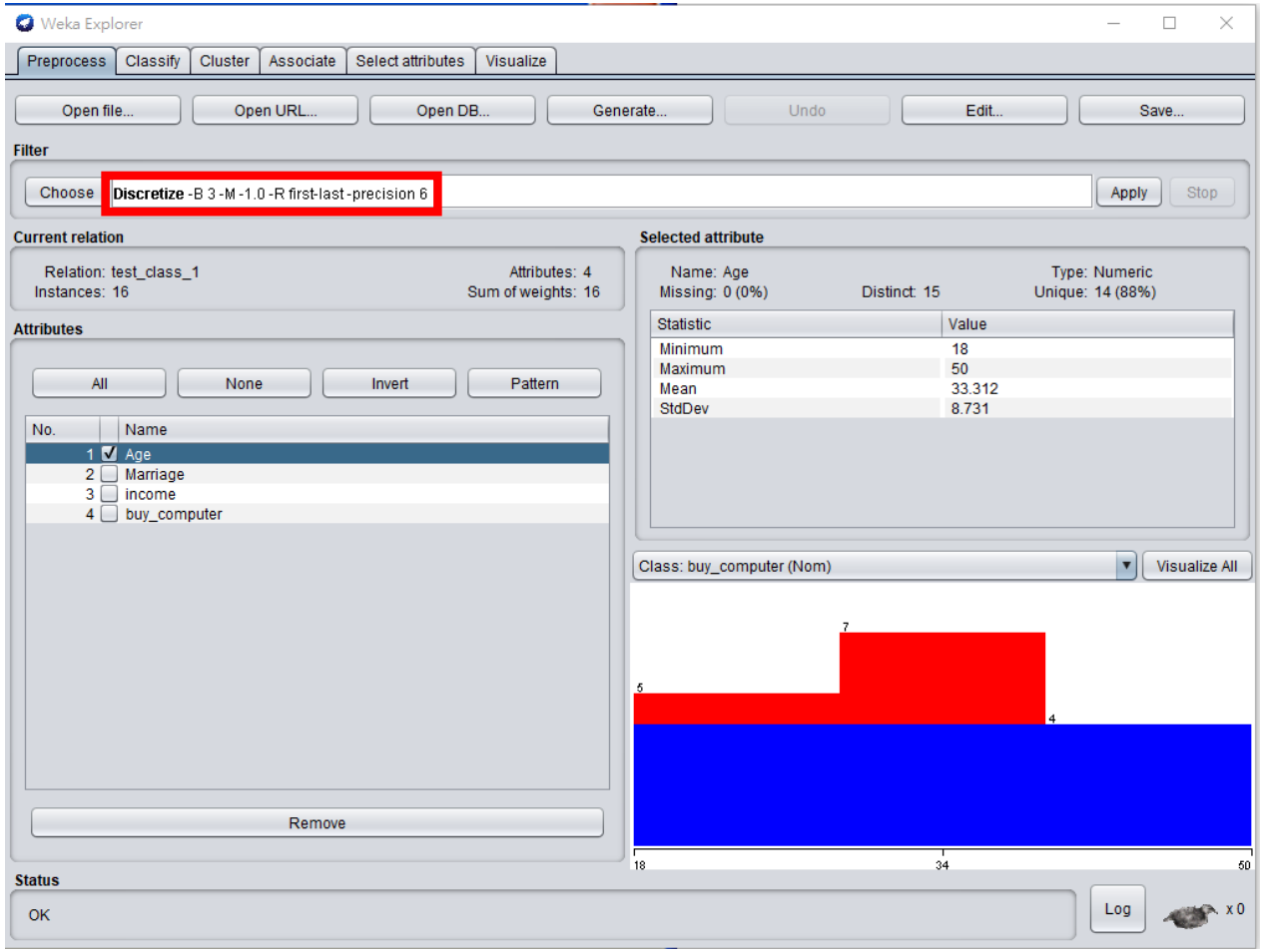
- weka
  - unsupervised
    - attribute
      - Discretize

Step 6



1. 選取 Age
2. 點擊 Discretize 處
3. 將 bins 設定成 3

Step 7



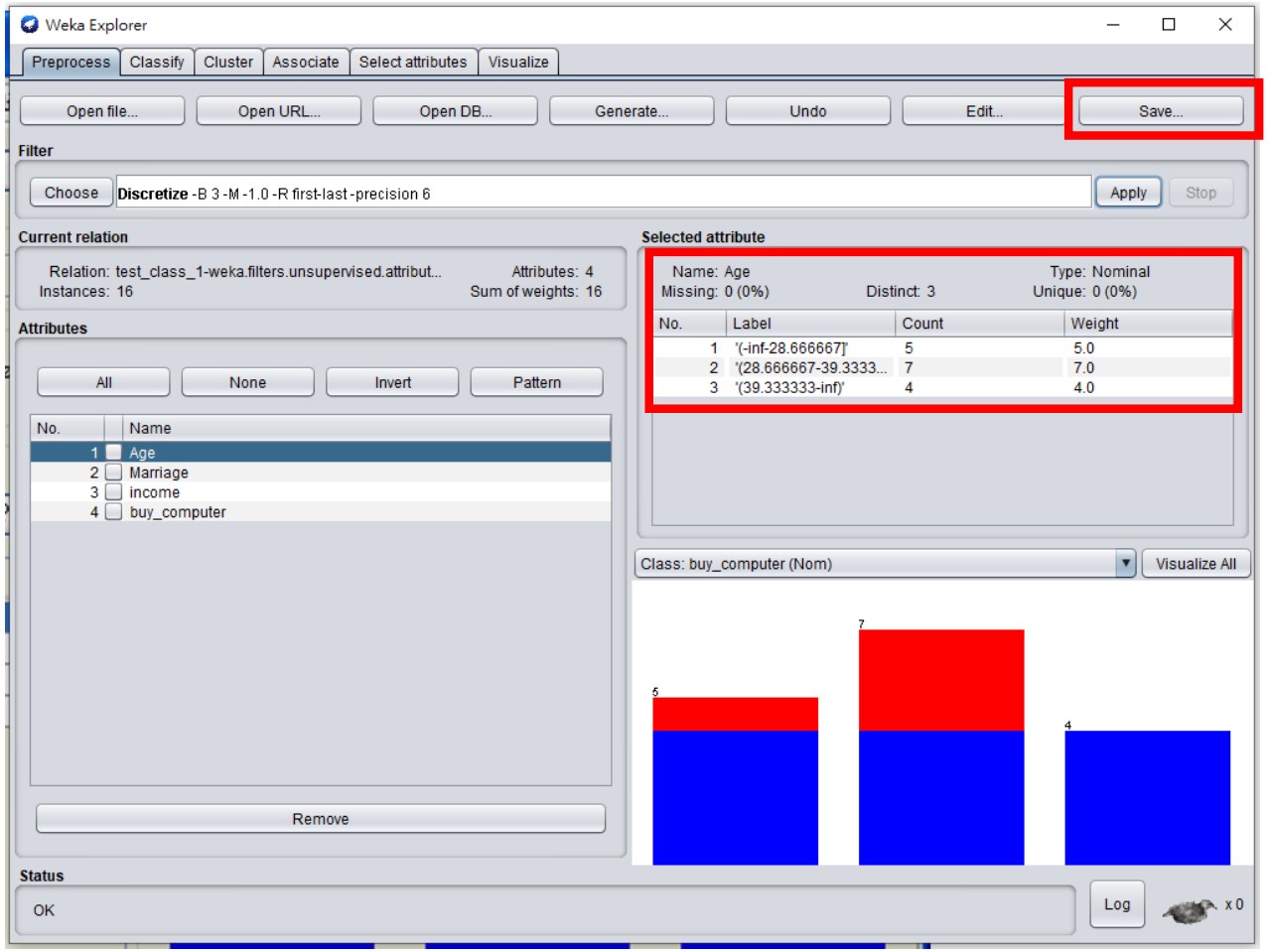
1. 確認

Discretize –B 3 –M -1.0 –R first-last-precision 6

2. 確認 Age 有勾選

3. 點擊 Apply

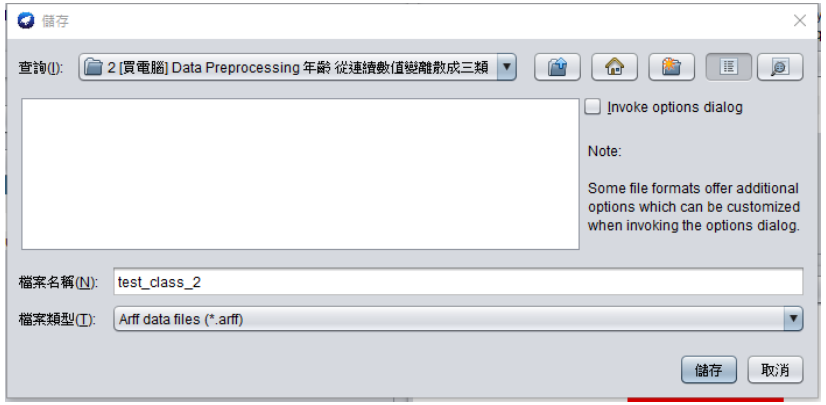
Step 8



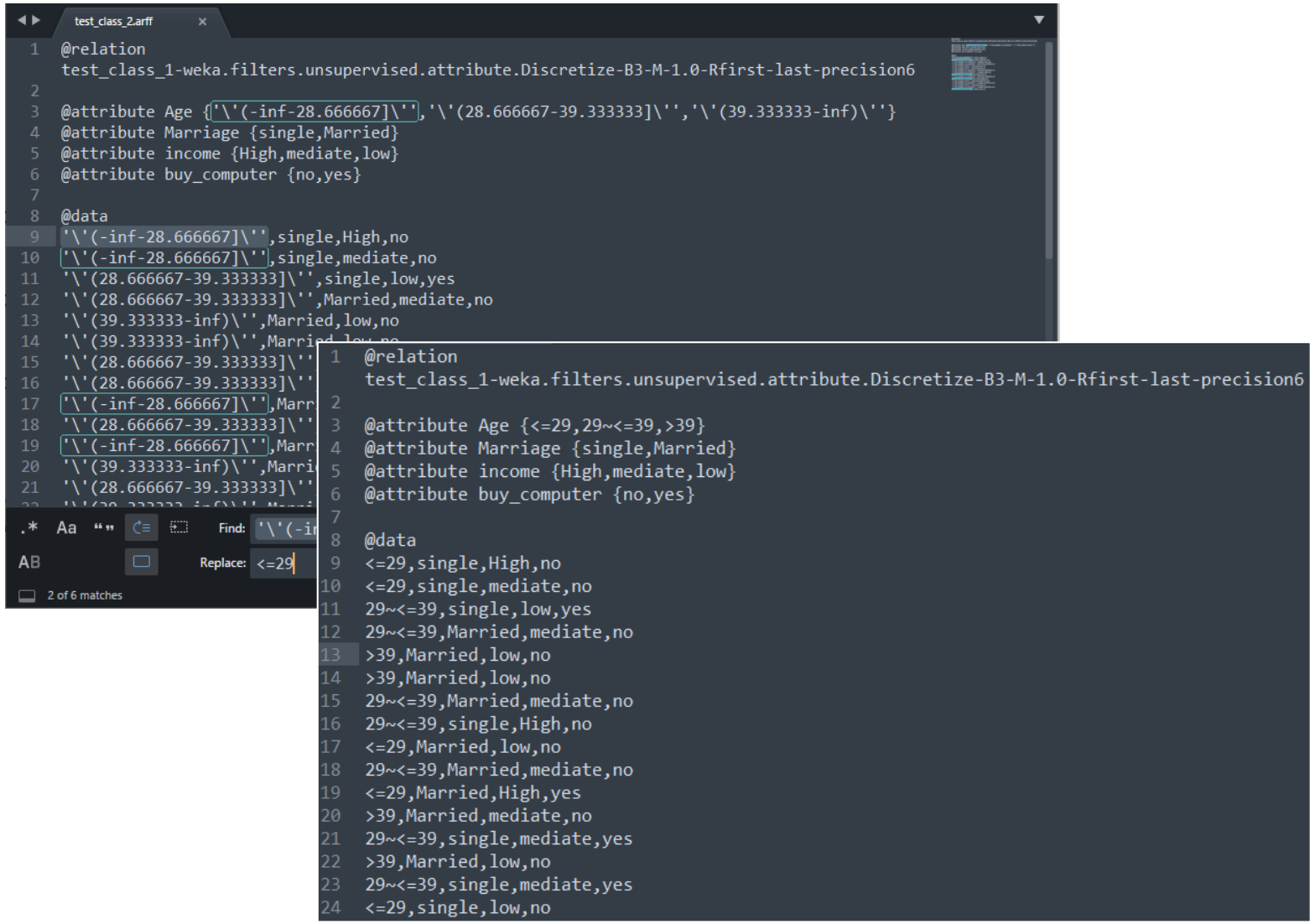
- 處理結果顯示出三個區間

1. -inf-28.666667
2. 28.66667-39.3333...
3. 39.333333-inf

- 點擊 Save 存檔



Step 9



- 更換三個區間的表示法

1. 以 <=29 取代
   '\'(-inf-28.666667]\''
2. 以 29~<=39 取代
   '\'(28.666667-39.333333]\''
3. 以 >39 取代
   '\'(39.333333-inf)\''