

- 1- "Discretize" used equal-width binning method to discretize attribute values.
- 2- Since equal-width partitioning divides the range into intervals of equal size, the distribution was not affected too much. The distinctness was decreased from 17 to 10 since range was divided into 10.
- 3- a) $\text{Interval} = (\text{Max. value} - \text{Min. value}) / \text{bin size} = (34-4)/3 = 10$
Bin 1: 4, 8, 9 (values in interval $(-\infty, 14]$)
Bin 2: 15, 21, 21, 24 (values in interval $(14, 24]$)
Bin 3: 25, 26, 28, 29, 34 (values in interval $(24, 34]$)
b) We will divide range into intervals that contain equal numbers of value.
Bin 1: 4, 8, 9, 15
Bin 2: 21, 21, 24, 25
Bin 3: 26, 28, 29, 34
- 4- Equal-width binning method is used for discretization about equal weights. On the other hand, equal-depth binning method is used for discretization about equal frequencies. If I'm looking for anomalies, I can use equal-width method since outliers are also in presentation of discretization. For data scaling, equal-depth method is more useful.