Associative rule mining Document

Vivek I(IMT2013017)
Aravind K(IMT2013019)
Nikhilesh P(IMT2013031)
Arun Prasad(IMT2013037)

As part of associative rule mining on our california crime data set, we have done the following:

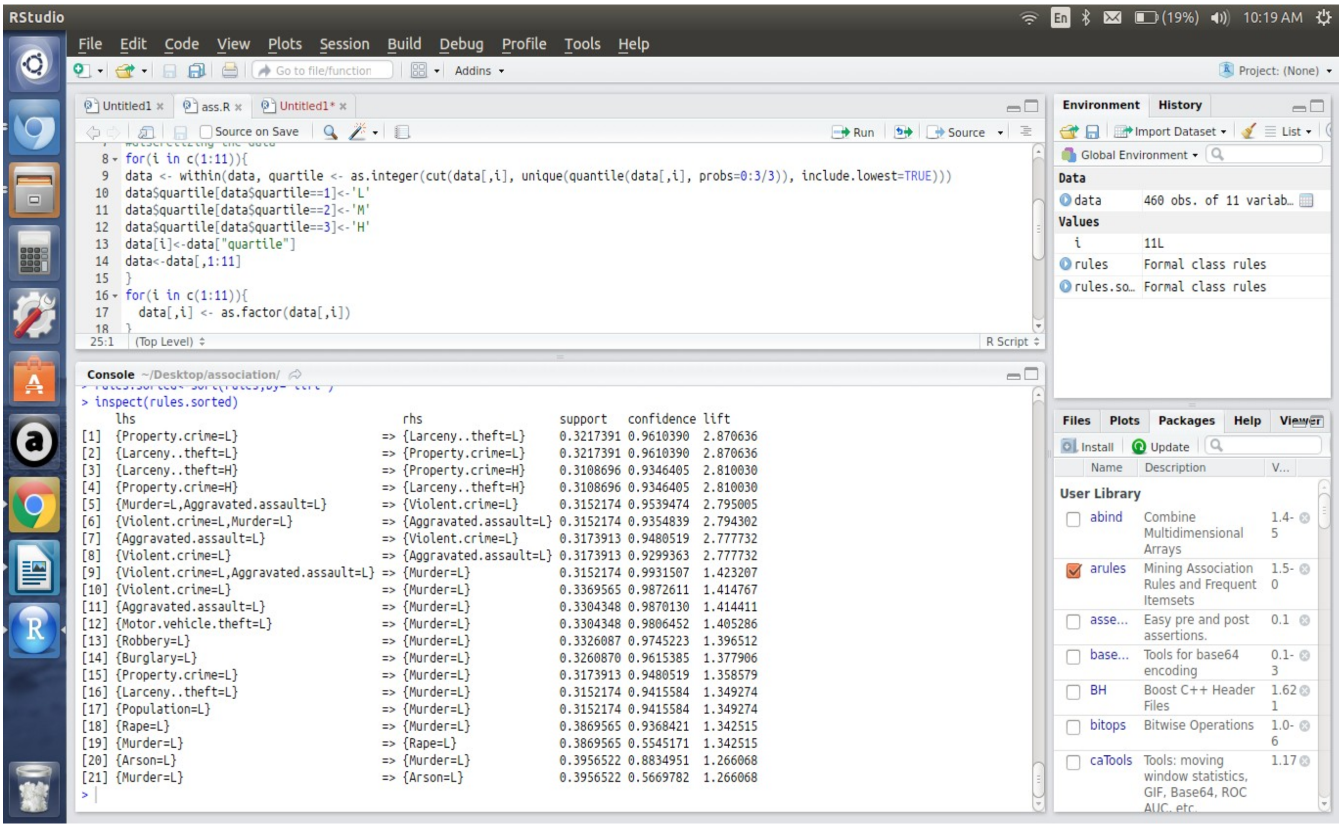
- Discretized continous attributes in our data, since attribute values must be discrete to apply associative rule mining algorithms.
- Applied apriori algorithm(by setting minsupport and minconfidence) to extract some rules i.e, doing **descriptive analysis** using these results
- Imposed some conditions in our apriori algorithm(on RHS of the implication) and arrived at certain rules i.e, **predictive analysis** using associative rule mining

<u>Discretization of continous variables</u> :

Our data set has information about number of crimes in each category(like Murder, burglary etc...) for every place in california. So, as part of discretizing these continous variables, we have converted them into three discrete variables namely L,M and H where L indicates that the number of crimes(of that category) in that place are low, M indicates that the number of crimes(of that category) in that place are medium and H indicates that the number of crimes(of that category) in that place are high.

<u>Descriptive analysis using associative rule mining</u> :

The rules obtaining after applying apriori algorithm(without imposing any conditions(on the data set(with minsupport=0.3 and minconfidence=0.7) are shown below:



Note: In our data set property crimes are divided into Burglary, Larency thefts and motor vehicle thefts and violent crimes are divided into Murder, Rape, Robbery and aggrevated assault.

Some useful and interesting rules(from the above set of 21 rules) and their interpretetion are the following:

[1], [3] – It says that, in a city(in california state) if the property crimes are low (or) high, then the larency thefts are also low (or) high respectively. Another way of interpreting this is that, among all the property crimes larency thefts mostly determine the number of property crimes. So, to reduce the number of property crimes in a city, more care can be taken towards reducing the larency thefts.

[5]- Similar interpretation as above i.e, In a city, among all the voilent crimes, murder and aggrevated assault mostly determine the number of violent crimes.

[7],[10]-Among aggrevated asault and murder, most contributing factor to violent crimes are aggrevated assaults since rule [7] has more lift value thatn rule [10]

[13]- In a city if robberies are less, then murders are less. Another interpretetion of this can be that people doing robbery are more proned to kill the people.

[15]-In a city, if the property crimes are low, then the number of murders are also low. So, we can think that people who are involved in burglary, larency thefts are more proned to kill the people.

[17]-This is a straight forward implication which says that, in a city if the population is low then the number of murdersa are also low.

[18]-In a city if number of rapes are low then the number of murders are also low. Another interpretation of the same thing is that: people who commit rapes also kill them.

<u>Predictive analysis using associative rule mining</u> :

We want to determine the factors which account for murders in a city. So, we imposed this rule on the RHS in our apriori algorithm and the result we got is the following: