1. Transaction data : nbaawards.csv.

* Collect data

Here I met a big problem, I surf lots of web sites (such as ESPN,NBA official site)and couldn’t find the data seems like ‘transaction’, there are actually lots of numeric data, but not so many categorical data can be used as items.

Finally I design the dataset and collected the data by myself, it only about hundreds of data.

* Clean data : no
* Delete data : delete some award only receive in the very few situation.
* Transform : excel->csv

1. Perform association rule analysis on above data : nbaaward.R

//PROGRAM

#an script to use Apriori to find the association rule of NBA award winner with other awards

#in the same year

# Load the libraries

library(arules)

# load the csv file nbaawards.csv

df<-read.csv("nbaawards.csv")

# get the rules

rules <- apriori(df, parameter = list(supp = 0.001, conf = 0.8,minlen=2))

# Show the top 5 rules, but only 2 digits

options(digits=2)

inspect(rules[1:5])

1. Top five interesting rules

If an super star only get ALL STAR THREE team award, it’s probably that he can’t win the MVP and championship that year.

//RULES

lhs rhs support confidence lift count

[1] {ALL.STAR.THREE=ALL-STAR THREE} => {ALL.STAR.ONE=} 0.08 1 2.9 8

[2] {ALL.STAR.THREE=ALL-STAR THREE} => {DEFENSE.ONE=} 0.08 1 1.5 8

[3] {ALL.STAR.THREE=ALL-STAR THREE} => {ALL.STAR.TWO=} 0.08 1 1.2 8

[4] {ALL.STAR.THREE=ALL-STAR THREE} => {final.MVP=} 0.08 1 1.2 8

[5] {ALL.STAR.THREE=ALL-STAR THREE} => {MVP=} 0.08 1 1.2 8