In [1]:
```
library(arules)
```

```
Warning message:
"package 'arules' was built under R version 3.3.3"Loading required package: Matrix

Attaching package: 'arules'

The following objects are masked from 'package:base':

    abbreviate, write
```

Preprocessing steps required for association rule mining

In [2]:
```
bk <-read.csv("bankdata.csv",header = TRUE, sep = ',')
```

In [3]:
```
#remove id field
bk[,1] <-NULL
```

In [4]:
```
head(bk)
```

age	sex	region	income	married	children	car	save_act	current_act	mortgage	pep
54	MALE	INNER_CITY	26707.9	YES	1	NO	YES	YES	YES	YES
27	FEMALE	INNER_CITY	11604.4	YES	2	YES	YES	YES	NO	NO
42	MALE	INNER_CITY	15499.9	YES	0	YES	NO	YES	YES	YES
43	MALE	TOWN	33088.5	NO	0	NO	YES	YES	YES	NO
64	FEMALE	INNER_CITY	34513.6	YES	1	NO	YES	YES	NO	YES
43	MALE	TOWN	32395.5	YES	3	YES	YES	YES	NO	NO

In [5]:
```
#discretize continuous attribute to norminal
bk[["age"]] <- ordered(cut(bk[["age"]], c(10, 40, 100)), labels = c("Young", "Senior"))
```

In [6]:
```
bk[["income"]] <- discretize(bk[["income"]], categories = 2)
```

In [7]:
```
#bk[["children"]] <- discretize(bk[["children"]], categories = 2)
bk[["children"]] <- ordered(cut(bk[["children"]], c(-1,0,10)), labels = c("NoChildren", "HaveChildren"))
```

In [8]:
```
head(bk)
```

age	sex	region	income	married	children	car	save_act	current_act	mortgage	pep
Senior	MALE	INNER_CITY	[6294,34712)	YES	HaveChildren	NO	YES	YES	YES	YES
Young	FEMALE	INNER_CITY	[6294,34712)	YES	HaveChildren	YES	YES	YES	NO	NO
Senior	MALE	INNER_CITY	[6294,34712)	YES	NoChildren	YES	NO	YES	YES	YES
Senior	MALE	TOWN	[6294,34712)	NO	NoChildren	NO	YES	YES	YES	NO
Senior	FEMALE	INNER_CITY	[6294,34712)	YES	HaveChildren	NO	YES	YES	NO	YES
Senior	MALE	TOWN	[6294,34712)	YES	HaveChildren	YES	YES	YES	NO	NO

In [9]:
```
bk[sapply(bk, is.character)] <- lapply(bk[sapply(bk, is.character)],as.factor)
```

In [10]:
```
summary(bk)
```

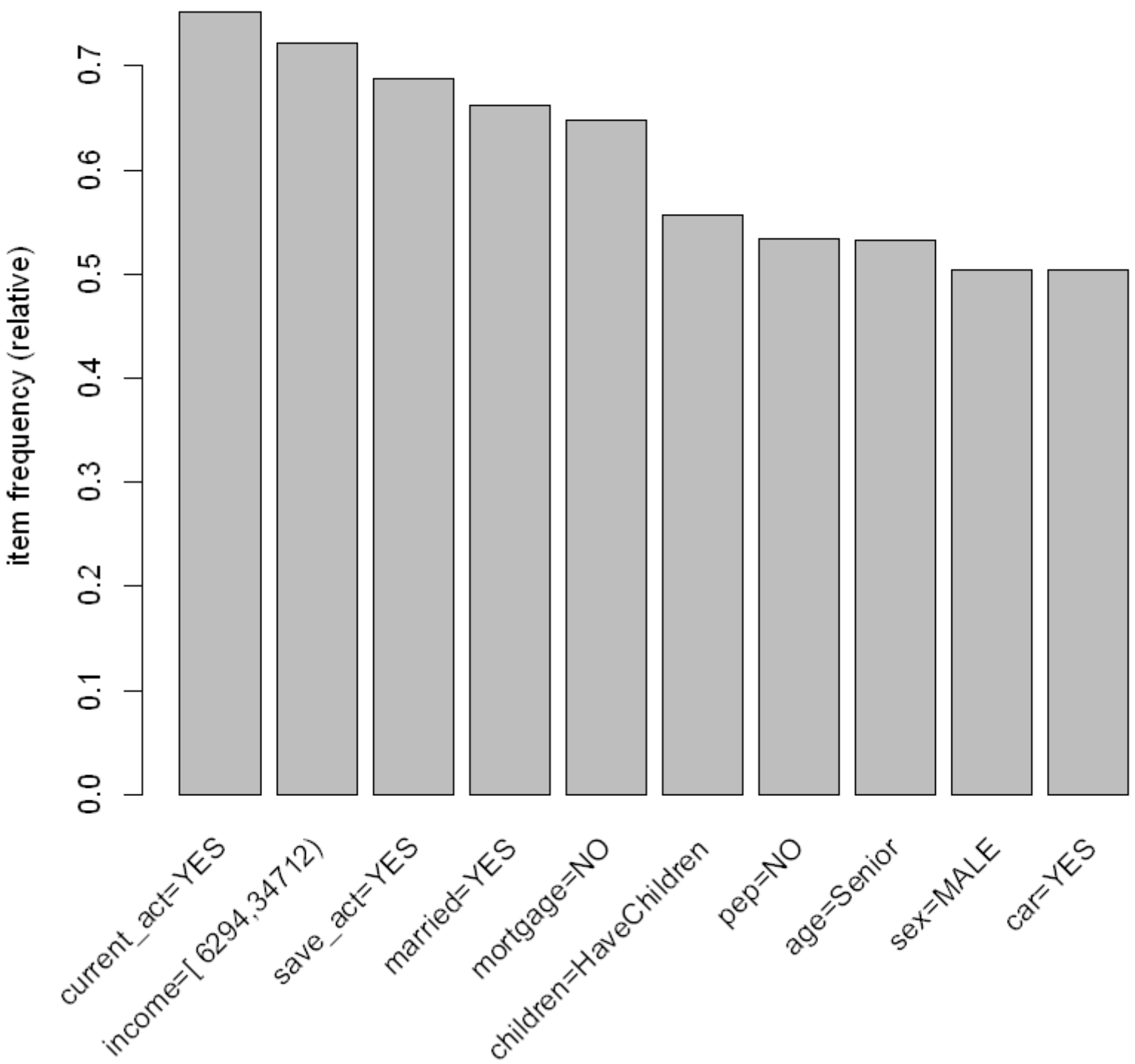
```
    age          sex            region              income    married  
 Young :234   FEMALE:248   INNER_CITY:221   [ 6294,34712):361   NO :169  
 Senior:266   MALE  :252   RURAL      : 83   [34712,63130]:139   YES:331  
                           SUBURBAN   : 55                               
                           TOWN       :141                               
        children    car      save_act  current_act mortgage   pep      
 NoChildren  :222   NO :248   NO :156   NO :124     NO :324   NO :267  
 HaveChildren:278   YES:252   YES:344   YES:376     YES:176   YES:233  
```

In [11]:
```
bk_tran <- as(bk, "transactions")
```

In [12]:
```
rules = apriori(bk_tran)
```

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.8    0.1    1 none FALSE            TRUE       5     0.1      1
 maxlen target   ext
     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 50

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[24 item(s), 500 transaction(s)] done [0.00s].
sorting and recoding items ... [24 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.01s].
writing ... [508 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

In [13]:
```
itemFrequencyPlot(bk_tran, topN = 10)
```



Experiment 0

min_suport = 0.2, confidence = 0.4, minlen = 2, pep=TRUE as RHS

In [14]:
```
rules = apriori(bk_tran, parameter = list(support = 0.2, confidence = 0.4,minlen = 2))
pep_rules = subset(rules,rhs %in% "pep=YES")
#summary(pep_rules)
inspect(sort(pep_rules, by = "lift"))
```

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.4    0.1    1 none FALSE            TRUE       5     0.2      2
 maxlen target   ext
     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 100

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[24 item(s), 500 transaction(s)] done [0.00s].
sorting and recoding items ... [22 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 done [0.00s].
writing ... [758 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
     lhs                                          rhs       support confidence
[1]  {age=Senior,children=HaveChildren}       => {pep=YES} 0.200   0.6622517
[2]  {married=NO}                             => {pep=YES} 0.204   0.6035503
[3]  {children=HaveChildren,save_act=YES}     => {pep=YES} 0.228   0.5757576
[4]  {children=HaveChildren,mortgage=NO}      => {pep=YES} 0.208   0.5621622
[5]  {children=HaveChildren,current_act=YES}  => {pep=YES} 0.228   0.5402844
[6]  {age=Senior,current_act=YES}             => {pep=YES} 0.214   0.5297030
[7]  {children=HaveChildren}                  => {pep=YES} 0.294   0.5287770
[8]  {age=Senior}                             => {pep=YES} 0.278   0.5225564
[9]  {age=Senior,save_act=YES}                => {pep=YES} 0.208   0.5098039
[10] {save_act=YES,mortgage=NO}               => {pep=YES} 0.224   0.4977778
[11] {sex=MALE}                               => {pep=YES} 0.244   0.4841270
[12] {car=YES}                                => {pep=YES} 0.244   0.4841270
[13] {current_act=YES,mortgage=NO}            => {pep=YES} 0.242   0.4840000
[14] {mortgage=NO}                            => {pep=YES} 0.306   0.4722222
[15] {current_act=YES}                        => {pep=YES} 0.354   0.4707447
[16] {region=INNER_CITY}                      => {pep=YES} 0.206   0.4660633
[17] {save_act=YES,current_act=YES}           => {pep=YES} 0.240   0.4545455
[18] {save_act=YES}                           => {pep=YES} 0.308   0.4476744
[19] {car=NO}                                 => {pep=YES} 0.222   0.4475806
[20] {sex=FEMALE}                             => {pep=YES} 0.222   0.4475806
[21] {income=[ 6294,34712),current_act=YES}   => {pep=YES} 0.224   0.4258555
[22] {income=[ 6294,34712)}                   => {pep=YES} 0.306   0.4238227
     lift
[1]  1.4211409
[2]  1.2951723
[3]  1.2355313
[4]  1.2063566
[5]  1.1594085
[6]  1.1367017
[7]  1.1347145
[8]  1.1213656
[9]  1.0939998
[10] 1.0681927
[11] 1.0388991
[12] 1.0388991
[13] 1.0386266
[14] 1.0133524
[15] 1.0101817
[16] 1.0001359
[17] 0.9754194
[18] 0.9606747
[19] 0.9604735
[20] 0.9604735
[21] 0.9138530
[22] 0.9094908
```

Experiment 1

min_suport = 0.2, confidence = 0.4, minlen = 2, sex=MALE & married=NO & pep=TRUE as RHS

In [15]:
```
rules = apriori(bk_tran, parameter = list(support = 0.1, confidence = 0.4,minlen = 2))
pep_rules = subset(rules,lhs %ain% c("sex=MALE","married=NO")& rhs %in% "pep=YES")
#summary(pep_rules)
inspect(sort(pep_rules, by = "lift"))
```

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.4    0.1    1 none FALSE            TRUE       5     0.1      2
 maxlen target   ext
     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 50

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[24 item(s), 500 transaction(s)] done [0.00s].
sorting and recoding items ... [24 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.00s].
writing ... [5667 rule(s)] done [0.59s].
creating S4 object  ... done [0.01s].
    lhs                       rhs       support confidence lift
[1] {sex=MALE,married=NO} => {pep=YES} 0.1     0.5952381  1.277335
```

Experiment 2

min_suport = 0.3, confidence = 0.4, minlen = 2, pep=TRUE as RHS

6 rules were generated

In [16]:
```
rules = apriori(bk_tran, parameter = list(support = 0.1, confidence = 0.04,minlen = 2))
pep_rules = subset(rules, lhs %in% "age=Young" & rhs %in% "pep=YES")
#summary(pep_rules)
inspect(sort(pep_rules, by = "support"))
```

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
       0.04    0.1    1 none FALSE            TRUE       5     0.1      2
 maxlen target   ext
     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 50

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[24 item(s), 500 transaction(s)] done [0.00s].
sorting and recoding items ... [24 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.01s].
writing ... [6092 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
     lhs                          rhs       support confidence      lift
[1]  {age=Young}               => {pep=YES}   0.188  0.4017094 0.8620373
[2]  {age=Young,
      income=[ 6294,34712)}    => {pep=YES}   0.176  0.3911111 0.8392942
[3]  {age=Young,
      current_act=YES}         => {pep=YES}   0.140  0.4022989 0.8633023
[4]  {age=Young,
      income=[ 6294,34712),
      current_act=YES}         => {pep=YES}   0.130  0.3892216 0.8352394
[5]  {age=Young,
      mortgage=NO}             => {pep=YES}   0.114  0.3851351 0.8264702
[6]  {age=Young,
      sex=MALE}                => {pep=YES}   0.110  0.4365079 0.9367123
[7]  {age=Young,
      income=[ 6294,34712),
      mortgage=NO}             => {pep=YES}   0.108  0.3776224 0.8103484
[8]  {age=Young,
      car=NO}                  => {pep=YES}   0.100  0.4000000 0.8583691
[9]  {age=Young,
      save_act=YES}            => {pep=YES}   0.100  0.3571429 0.7664010
[10] {age=Young,
      sex=MALE,
      income=[ 6294,34712)}    => {pep=YES}   0.100  0.4237288 0.9092893
```

Experiment 3

min_suport = 0.1, confidence = 0.8, minlen = 2, income=[34712,63130] & pep=TRUE as RHS & lift > 1.0

10 rules were generated

In [17]:
```
rules = apriori(bk_tran, parameter = list(support = 0.1, confidence = 0.8,minlen = 3))
pep_rules = subset(rules, lhs %in% "income=[34712,63130]" & rhs %in% "pep=YES" & lift > 1.0)
#summary(pep_rules)
inspect(sort(pep_rules, by = "lift"))
```

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.8    0.1    1 none FALSE            TRUE       5     0.1      3
 maxlen target   ext
     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 50

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[24 item(s), 500 transaction(s)] done [0.00s].
sorting and recoding items ... [24 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.00s].
writing ... [503 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
    lhs                          rhs       support confidence     lift
[1] {income=[34712,63130],
     children=HaveChildren,
     mortgage=NO}             => {pep=YES}    0.102  0.8644068 1.854950
[2] {income=[34712,63130],
     children=HaveChildren,
     save_act=YES,
     current_act=YES}         => {pep=YES}    0.102  0.8360656 1.794132
[3] {income=[34712,63130],
     children=HaveChildren,
     save_act=YES}            => {pep=YES}    0.124  0.8266667 1.773963
[4] {age=Senior,
     income=[34712,63130],
     children=HaveChildren,
     current_act=YES}         => {pep=YES}    0.104  0.8253968 1.771238
[5] {income=[34712,63130],
     children=HaveChildren,
     current_act=YES}         => {pep=YES}    0.112  0.8235294 1.767230
[6] {income=[34712,63130],
     children=HaveChildren}   => {pep=YES}    0.134  0.8170732 1.753376
[7] {age=Senior,
     income=[34712,63130],
     children=HaveChildren,
     save_act=YES}            => {pep=YES}    0.116  0.8169014 1.753007
[8] {age=Senior,
     income=[34712,63130],
     children=HaveChildren}   => {pep=YES}    0.124  0.8157895 1.750621
```

Three interesting rules

lhs		rhs	support	confidence	lift
{sex=MALE,married=NO}	=>	{pep=YES}	0.1	0.5952381	1.277335
{age=Senior, children=HaveChildren,save_act=YES}	=>	{pep=YES}	0.212	0.6235294	1.3380459
{income=[34712,63130],children=HaveChildren,mortgage=NO}	=>	{pep=YES}	0.102	0.8644068	1.854950

Observations

1. From **experiment 0**, we can see that compares with female, male has more interest in the PEP product, also, senior people have higher change to buy PEP.
2. From **expriment 1**, we can see a sinlge male has significant relationship with the PEP purchase behavior.
3. From **expriment 2** and **experiment 3**, we can see that the younger people(age < 35) doesn't have much interest buying our product.
4. **Experiment 4** investigated the customers who have a higher income, since those people have much interest in purchasing the PEP product.

Recommendations

The target customer should have the following feathers: **1.single male; 2.having no mortgage; 3.having a saving account; 4.senior people who has children; 5.having higher income(income > 34712$)**.

Calculate support, confidence and lift

In [18]:
```
bk2<-as.data.frame(bk)
```

In [19]:
```
summary(bk2)
```

```
     age          sex             region               income    married
 Young :234   FEMALE:248   INNER_CITY:221   [ 6294,34712):361   NO :169
 Senior:266   MALE  :252   RURAL      : 83   [34712,63130]:139   YES:331
                           SUBURBAN   : 55
                           TOWN       :141
         children    car      save_act  current_act mortgage   pep
 NoChildren  :222   NO :248   NO :156   NO :124     NO :324   NO :267
 HaveChildren:278   YES:252   YES:344   YES:376     YES:176   YES:233
```

In [20]:
```
## this doesnot work in jupyter notebook, but works in RStudio!
sigma_male_marryNo <- nrow(bk2[ bk2$sex=="MALE" & bk2$married == "NO",]) # = 83
sigma_male_marryNo_pepYes <-nrow (bk[( bk$sex=="MALE" & bk$married == "NO" & bk$pep=="YES"),]) # = 50
```

now caculate the support, confidence and lift for the following rule;

{sex=MALE,married=NO} => {pep=YES}

support = 50/500 = 0.1

confidence = 50/84 = 0.5952381

lift = (50x500)/(233x84) = 1.277335

In [ ]:
```

```