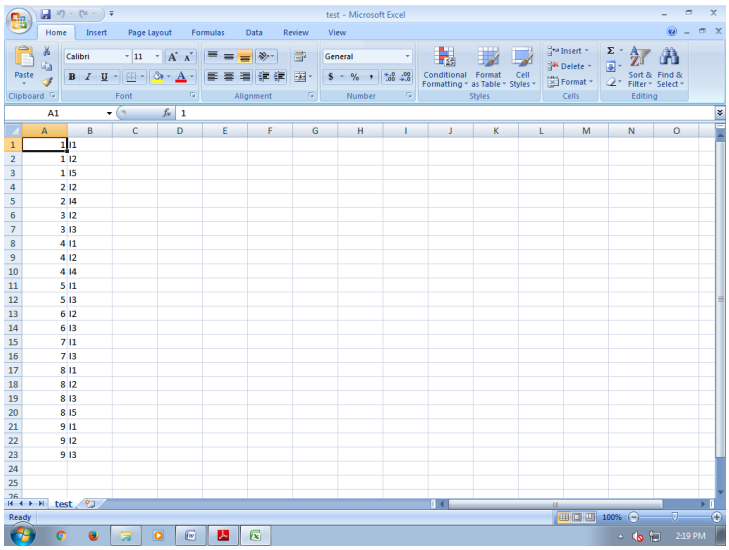
Mining Single Dimensional Boolean Association Rule in R

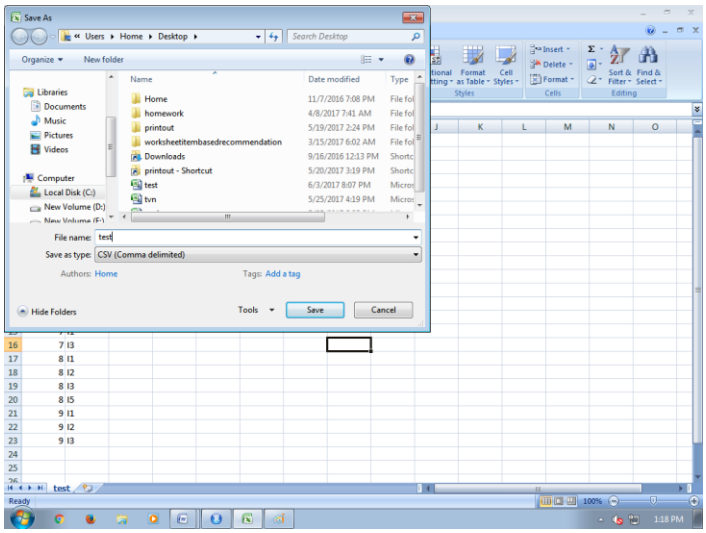
Consider the following transaction database "all_elect"

Tid	List of items
1	I1,I2,I5
2	I2,I4
3	I2,I3
4	I1,I2,I4
5	I1,I3
6	I2,I3
7	I1,I3
8	I1,I2,I3,I5

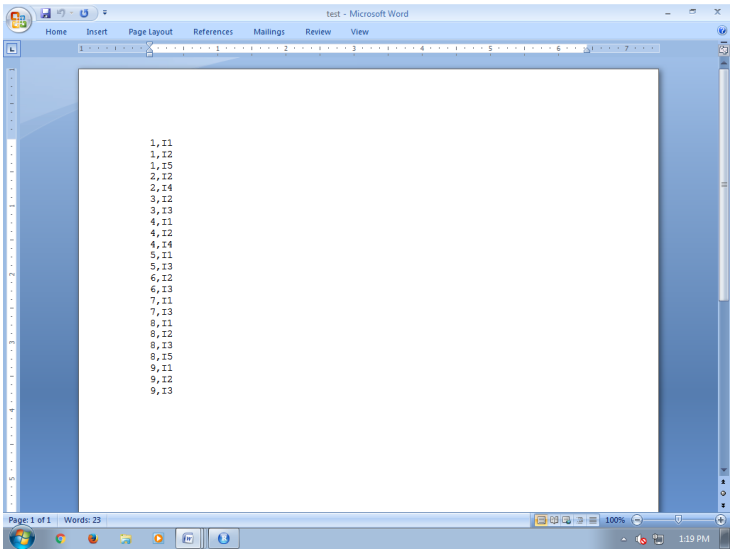
Step 1: **Making as an Excel Format**



Step 2: **Saving the File as CSV** - File->save as ->.csv(comma delimited) as shown below.



To open the CSV file(test.csv) in Ms-Word



>setwd("C:/Users/home/Desktop")

>install.packages('arules')
>install.packages('arulViz')

> library('arules')
> library('arulesViz')
Loading required package: grid
>

> txn = read.transactions(file="test.CSV",format = "single",sep=",",col = c(1,2))

> summary(txn)

```
transactions as itemMatrix in sparse format with
 9 rows (elements/itemsets/transactions) and
 5 columns (items) and a density of 0.5111111
most frequent items:
    I2     I1     I3     I4     I5 (Other)
     7      6      6      2      2      0
element (itemset/transaction) length distribution:
sizes
2 3 4
5 3 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   2.000   2.000   2.556   3.000   4.000
includes extended item information - examples:
  labels
1     I1
2     I2
3     I3
includes extended transaction information - examples:
  transactionID
1             1
2             2
3             3
```

> inspect(txn)

```
   items        transactionID
[1] {I1,I2,I5}    1
[2] {I2,I4}       2
[3] {I2,I3}       3
[4] {I1,I2,I4}    4
[5] {I1,I3}       5
[6] {I2,I3}       6
[7] {I1,I3}       7
[8] {I1,I2,I3,I5} 8
[9] {I1,I2,I3}    9
```

> class(txn)

```
[1] "transactions"
attr(,"package")
[1] "arules"
```

> txn@itemInfo[1:5,]

```
[1] "I1" "I2" "I3" "I4" "I5"
```

Frequent ltem set Generation

Single Item set

> itemsets <- apriori(txn,parameter=list(minlen=1,maxlen=1,support=0.22, target="frequent itemsets"))

```
Apriori
Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime
         NA    0.1    1 none FALSE           TRUE       5
 support minlen maxlen           target   ext
    0.22      1      1 frequent itemsets FALSE
Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
Absolute minimum support count: 1
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[5 item(s), 9 transaction(s)] done [0.00s].
sorting and recoding items ... [5 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 done [0.00s].
writing ... [5 set(s)] done [0.00s].
creating S4 object  ... done [0.00s].
Warning message:
In apriori(txn, parameter = list(minlen = 1, maxlen = 1, support = 0.22,  :
  Mining stopped (maxlen reached). Only patterns up to a length of 1 returned!
```

> summary(itemsets)

```
set of 5 itemsets
most frequent items:
    I1     I2     I3     I4     I5 (Other)
```

```
    1      1      1      1      1      0

element (itemset/transaction) length distribution:sizes
1
5
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1       1       1       1       1       1
summary of quality measures:
    support
 Min.   :0.2222
 1st Qu.:0.2222
 Median :0.6667
 Mean   :0.5111
 3rd Qu.:0.6667
 Max.   :0.7778
includes transaction ID lists: FALSE
mining info:
 data ntransactions support confidence
  txn           9    0.22          1
> inspect(head(sort(itemsets, by = "support"), 10))
    items support
[1] {I2}  0.7777778
[2] {I3}  0.6666667
[3] {I1}  0.6666667
[4] {I4}  0.2222222
[5] {I5}  0.2222222
```

Two item set

```
> itemsets <- apriori(txn,parameter=list(minlen=2,maxlen=2,support=0.22,
target="frequent itemsets"))
Apriori
Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime
         NA    0.1    1 none FALSE           TRUE       5
 support minlen maxlen           target   ext
    0.22      2      2 frequent itemsets FALSE

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

Absolute minimum support count: 1
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[5 item(s), 9 transaction(s)] done [0.00s].
sorting and recoding items ... [5 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [6 set(s)] done [0.00s].
creating S4 object  ... done [0.00s].
Warning message:
In apriori(txn, parameter = list(minlen = 2, maxlen = 2, support = 0.22,  :
  Mining stopped (maxlen reached). Only patterns up to a length of 2 returned!
```

> summary(itemsets)

```
set of 6 itemsets
most frequent items:
    I2     I1     I3     I5     I4 (Other)
     4      3      2      2      1      0
element (itemset/transaction) length distribution:sizes
2
6
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     2       2       2       2       2       2
summary of quality measures:
   support
 Min.   :0.2222
 1st Qu.:0.2222
 Median :0.3333
 Mean   :0.3333
 3rd Qu.:0.4444
 Max.   :0.4444
includes transaction ID lists: FALSE
mining info:
 data ntransactions support confidence
  txn           9    0.22          1
```

> inspect(head(sort(itemsets, by = "support"), 10))

```
   items   support
[1] {I1,I3} 0.4444444
[2] {I2,I3} 0.4444444
[3] {I1,I2} 0.4444444
[4] {I2,I4} 0.2222222
[5] {I1,I5} 0.2222222
[6] {I2,I5} 0.2222222
```

Three item set

> itemsets <- apriori(txn,parameter=list(minlen=3,maxlen=3,support=0.22, target="frequent itemsets"))

```
Apriori
Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime
         NA    0.1    1 none FALSE            TRUE       5
 support minlen maxlen           target   ext
    0.22      3      3 frequent itemsets FALSE
Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
Absolute minimum support count: 1
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[5 item(s), 9 transaction(s)] done [0.00s].
sorting and recoding items ... [5 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [2 set(s)] done [0.00s].
```

```
creating S4 object  ... done [0.00s].
Warning message:
In apriori(txn, parameter = list(minlen = 3, maxlen = 3, support = 0.22,  :
  Mining stopped (maxlen reached). Only patterns up to a length of 3 returned!
```

> summary(itemsets)

```
set of 2 itemsets
most frequent items:
    I1     I2     I3     I5     I4 (Other)
     2      2      1      1      0      0
element (itemset/transaction) length distribution:sizes
3
2
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      3       3       3       3       3       3
summary of quality measures:
    support
 Min.   :0.2222
 1st Qu.:0.2222
 Median :0.2222
 Mean   :0.2222
 3rd Qu.:0.2222
 Max.   :0.2222
includes transaction ID lists: FALSE
mining info:
 data ntransactions support confidence
  txn           9    0.22          1
```

> inspect(head(sort(itemsets, by = "support"), 10))

```
    items      support
[1] {I1,I2,I5} 0.2222222
[2] {I1,I2,I3} 0.2222222
```

Rule Generation

> c <- apriori(txn,parameter = list (support = 0.22,confidence = 0.90,target = "rules"))

```
Apriori
Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime
        0.9    0.1    1 none FALSE           TRUE       5
 support minlen maxlen target   ext
    0.22      1     10  rules FALSE
Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
Absolute minimum support count: 1
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[5 item(s), 9 transaction(s)] done [0.00s].
sorting and recoding items ... [5 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [5 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

> inspect(txn)

```
   items        transactionID
[1] {I1,I2,I5}    1
[2] {I2,I4}       2
[3] {I2,I3}       3
[4] {I1,I2,I4}    4
[5] {I1,I3}       5
[6] {I2,I3}       6
[7] {I1,I3}       7
[8] {I1,I2,I3,I5} 8
[9] {I1,I2,I3}    9
```

> inspect(c)

```
   lhs       rhs  support   confidence lift
[1] {I4}    => {I2} 0.2222222 1         1.285714
[2] {I5}    => {I1} 0.2222222 1         1.500000
[3] {I5}    => {I2} 0.2222222 1         1.285714
[4] {I1,I5} => {I2} 0.2222222 1          1.285714
[5] {I2,I5} => {I1} 0.2222222 1          1.500000
```