Association Rules Mining

Rajesh K. Jat
2021kpad1001@iiitkota.ac.in

Statistics for Data Science
(Winter, 2021-22)

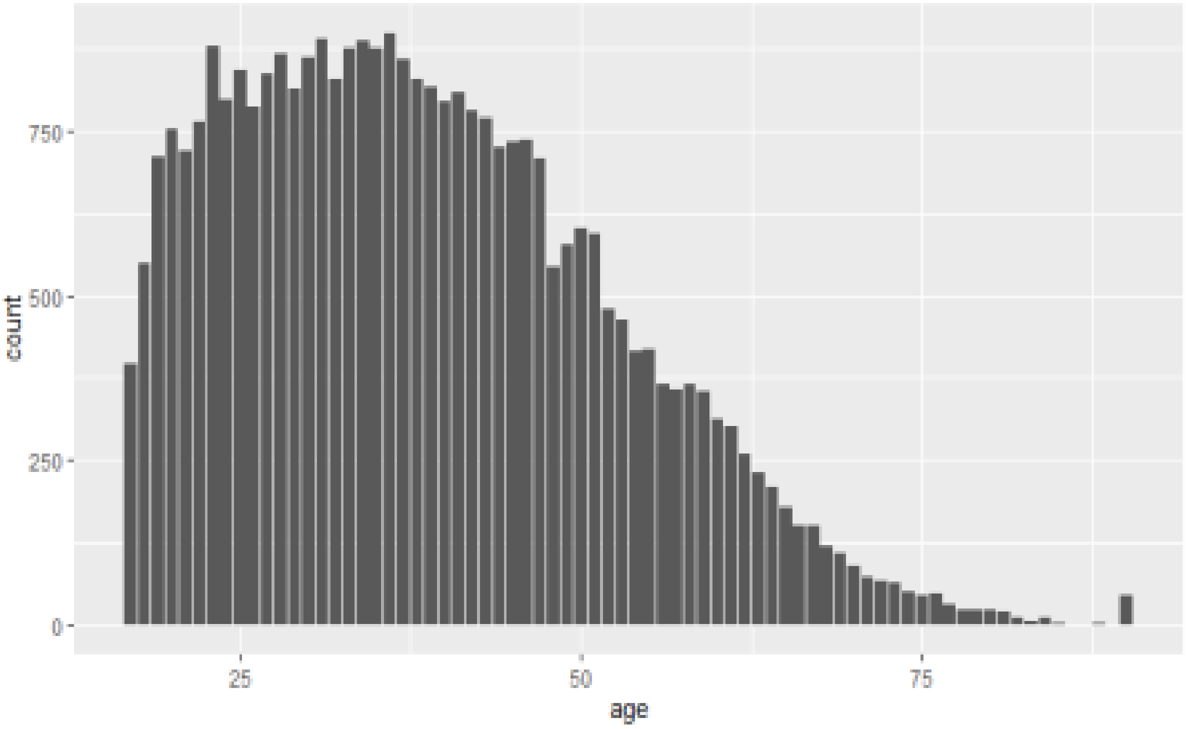
Dataset:-

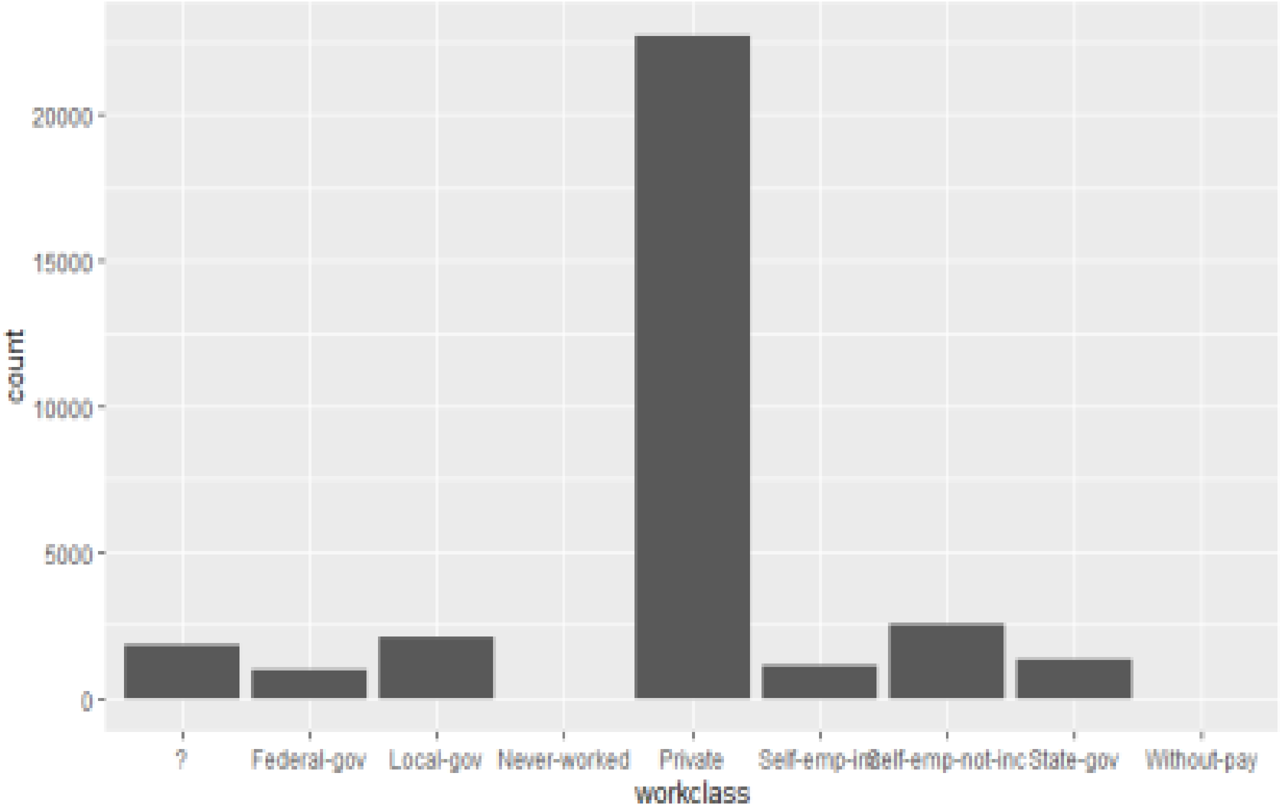
An individual's annual income results from various factors.

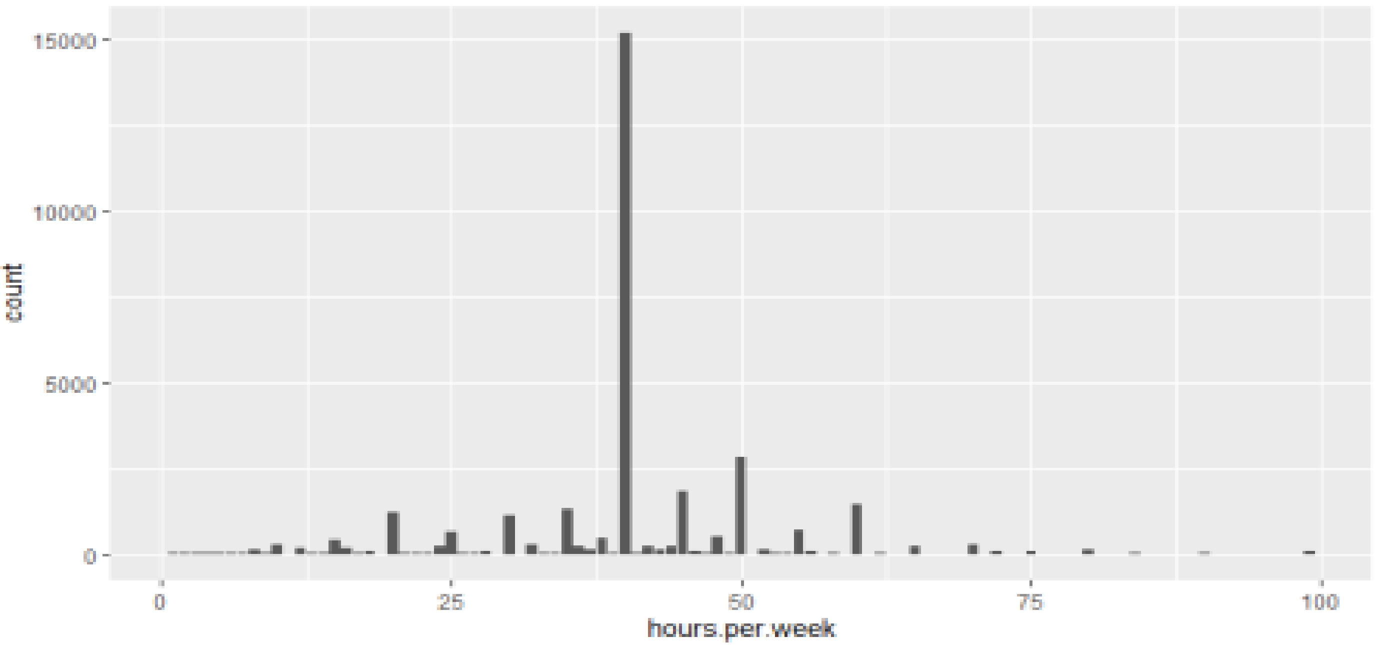
Income is influenced by the individual's education level, age, gender, occupation, etc.

Fields: The dataset contains 16 columns.
Target Field: Income
The income is divided into two classes: <=50K and >50K.

Data Visualization:-







Data Discretization:

Here we discretized 4 columns: age, capital.gain, capital.loss, hours.per.week

Terms Used:

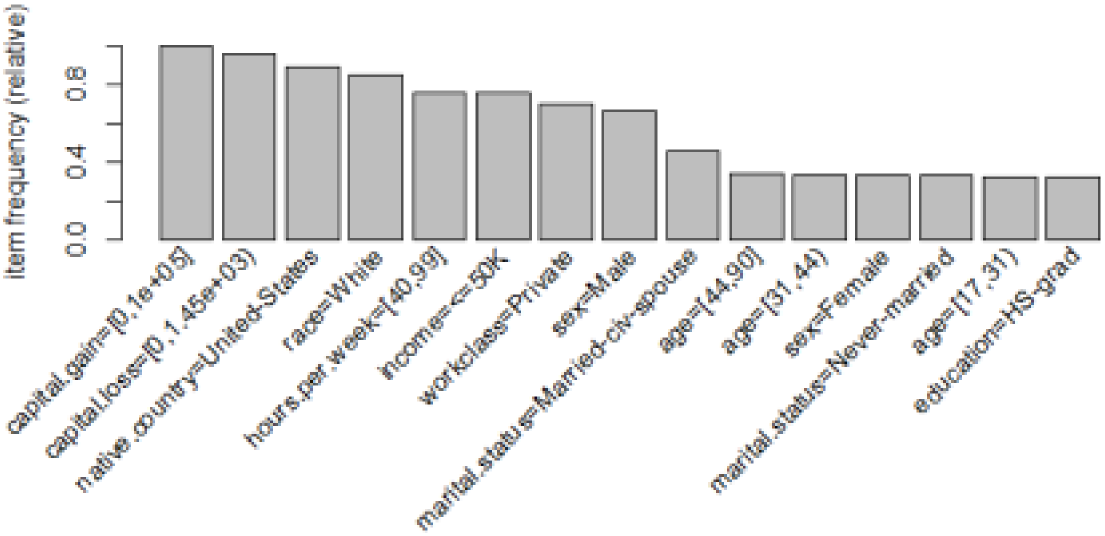
Lift →Probability of togetherness

Gini → Info Provide

Phi → Binary Correlation

Support → frequency of items occurred.

Confidence → How Reliable the rue is.



Maximal frequent itemset:

Maximal frequent itemset shows that none of its immediate supersets are frequent.

Here, the total no. of maximal frequent itemsets is 8. Some of them are:

Items	**Support**	**Count**
{race=White, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), hours.per.week=[40,99], native.country=United States}	0.5709	18592

{race=White, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), native.country=United States, income=<=50K}	0.5645	18382
{workclass=Private, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), income=<=50K}	0.5307	17281
{capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), hours.per.week=[40,99], income=<=50K}	0.5279	17192

Closest frequent itemset:

Closet frequent itemset refers that none of it's immediate superset has same support.

Here are total 48 itemsets of such type, some of them are:

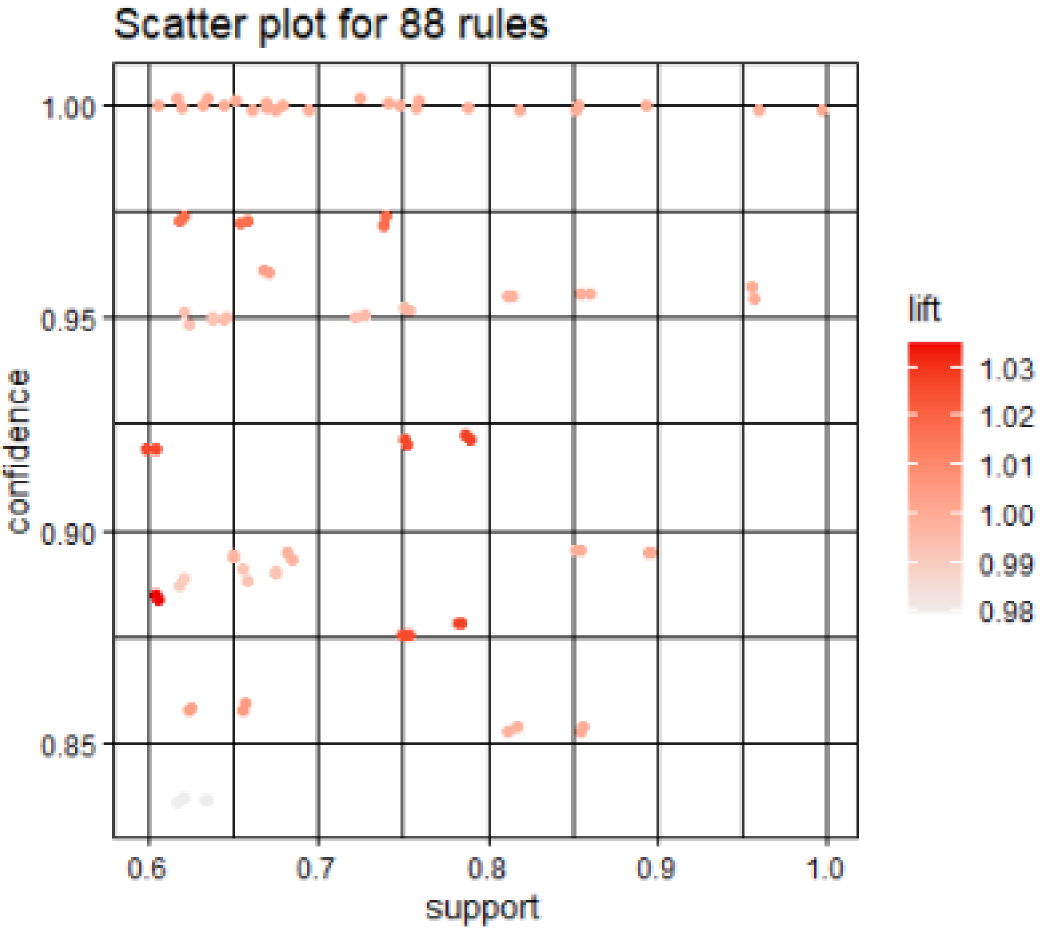
Item	**Support**	**Count**
{capital.gain=[0,1e+05]}	1.000	32561
{capital.gain=[0,1e+05], capital.loss=[0,1.45e+03)}	0.9559	31127

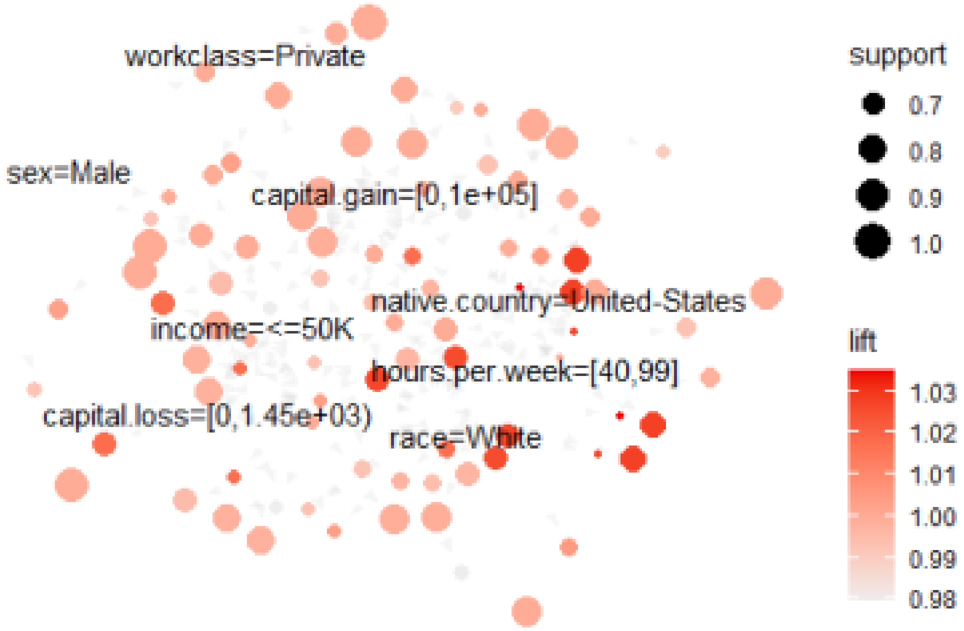
Association rule:

support = 0.6

confidence = 0.8

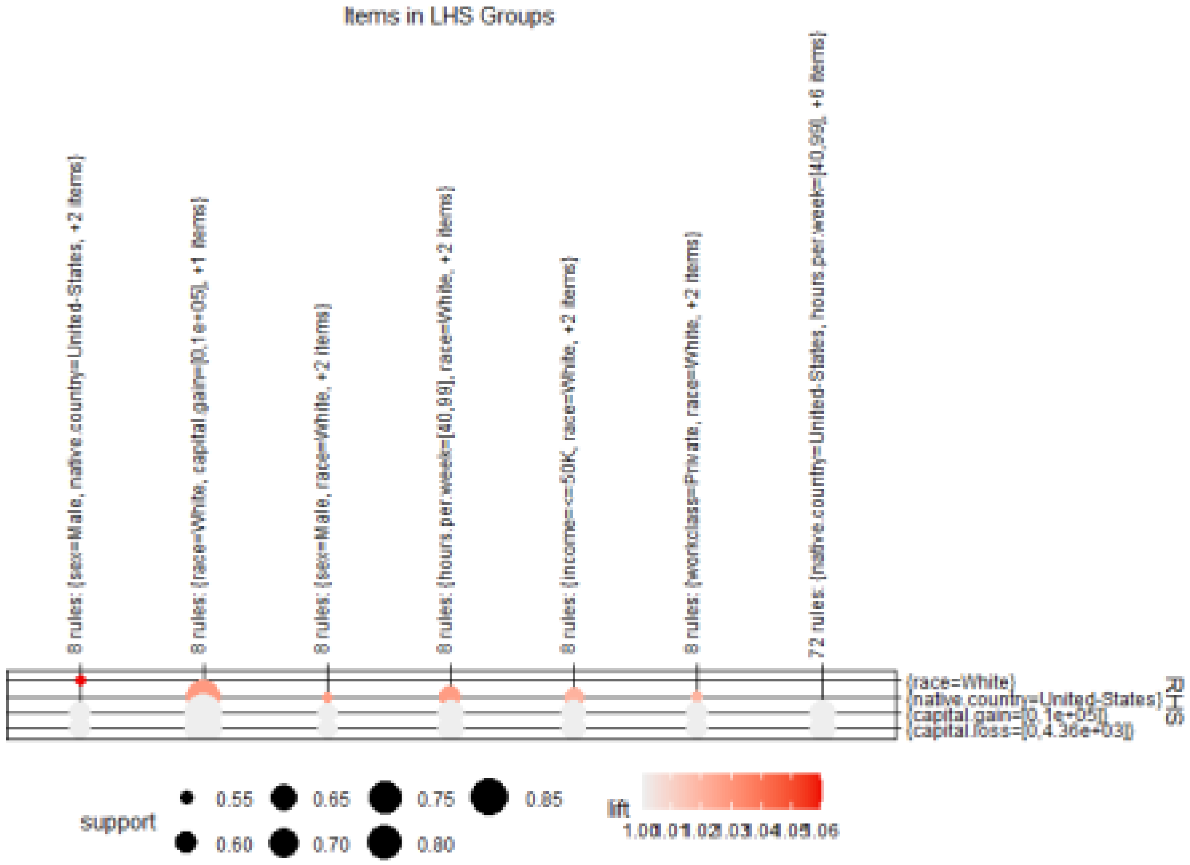
Total number of association rules is 88.

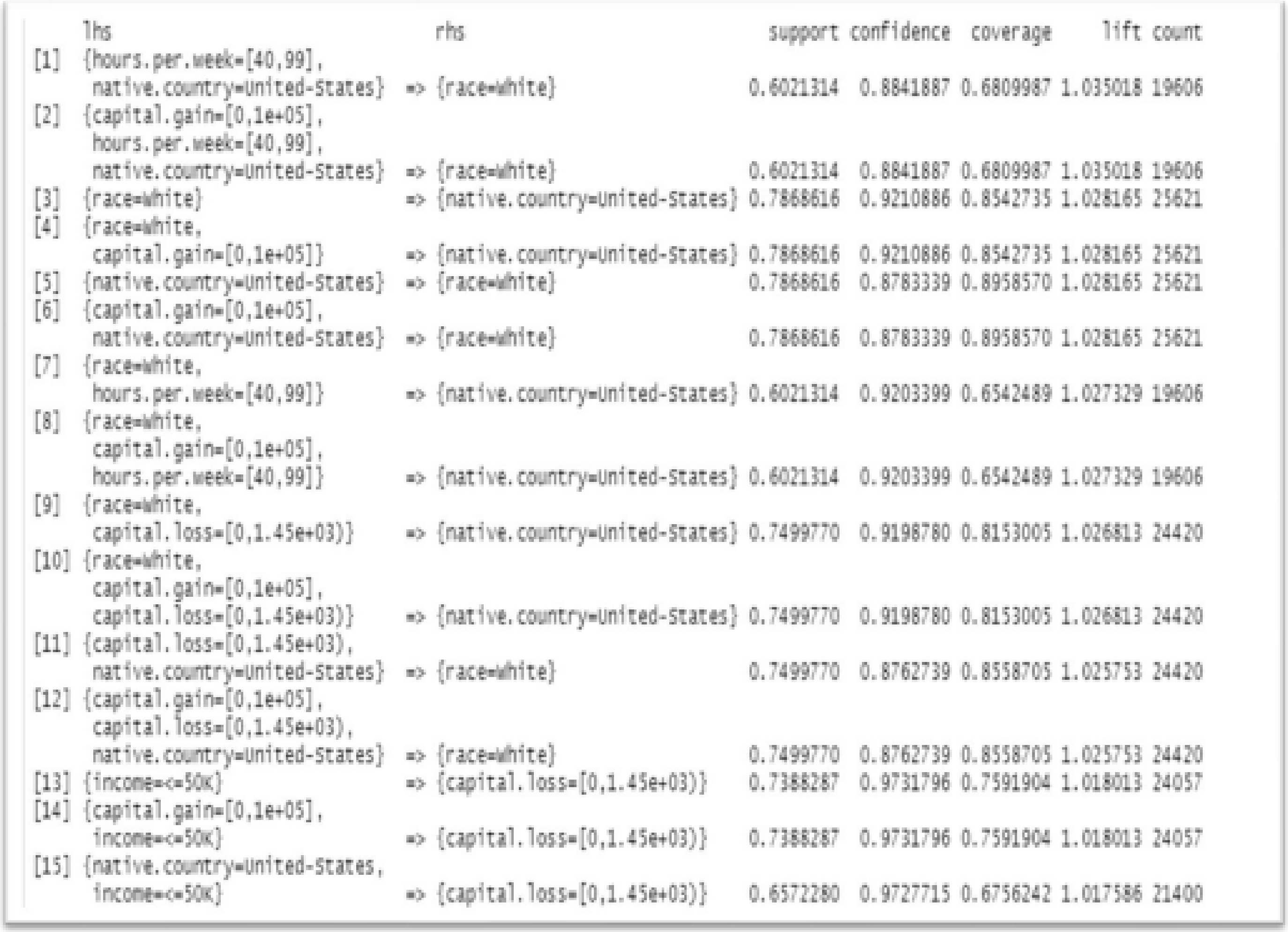




Bigger the circle, higher the support.

Higher the intensity, higher the lift(probability).





```
     lhs                                   rhs                                support confidence  coverage     lift count
[1]  {hours.per.week=[40,99],
      native.country=United-States}  => {race=White}                      0.6021314  0.8841887 0.6809987 1.035018 19606
[2]  {capital.gain=[0,1e+05],
      hours.per.week=[40,99],
      native.country=United-States}  => {race=White}                      0.6021314  0.8841887 0.6809987 1.035018 19606
[3]  {race=White}                    => {native.country=United-States} 0.7868616  0.9210886 0.8542735 1.028165 25621
[4]  {race=White,
      capital.gain=[0,1e+05]}        => {native.country=United-States} 0.7868616  0.9210886 0.8542735 1.028165 25621
[5]  {native.country=United-States}  => {race=White}                      0.7868616  0.8783339 0.8958570 1.028165 25621
[6]  {capital.gain=[0,1e+05],
      native.country=United-States}  => {race=White}                      0.7868616  0.8783339 0.8958570 1.028165 25621
[7]  {race=White,
      hours.per.week=[40,99]}        => {native.country=United-States} 0.6021314  0.9203399 0.6542489 1.027329 19606
[8]  {race=White,
      capital.gain=[0,1e+05],
      hours.per.week=[40,99]}        => {native.country=United-States} 0.6021314  0.9203399 0.6542489 1.027329 19606
[9]  {race=White,
      capital.loss=[0,1.45e+03)}     => {native.country=United-States} 0.7499770  0.9198780 0.8153005 1.026813 24420
[10] {race=White,
      capital.gain=[0,1e+05],
      capital.loss=[0,1.45e+03)}     => {native.country=United-States} 0.7499770  0.9198780 0.8153005 1.026813 24420
[11] {capital.loss=[0,1.45e+03),
      native.country=United-States}  => {race=White}                      0.7499770  0.8762739 0.8558705 1.025753 24420
[12] {capital.gain=[0,1e+05],
      capital.loss=[0,1.45e+03),
      native.country=United-States}  => {race=White}                      0.7499770  0.8762739 0.8558705 1.025753 24420
[13] {income=<=50K}                  => {capital.loss=[0,1.45e+03)}    0.7388287  0.9731796 0.7591904 1.018013 24057
[14] {capital.gain=[0,1e+05],
      income=<=50K}                  => {capital.loss=[0,1.45e+03)}    0.7388287  0.9731796 0.7591904 1.018013 24057
[15] {native.country=United-States,
      income=<=50K}                  => {capital.loss=[0,1.45e+03)}    0.6572280  0.9727715 0.6756242 1.017586 21400
```