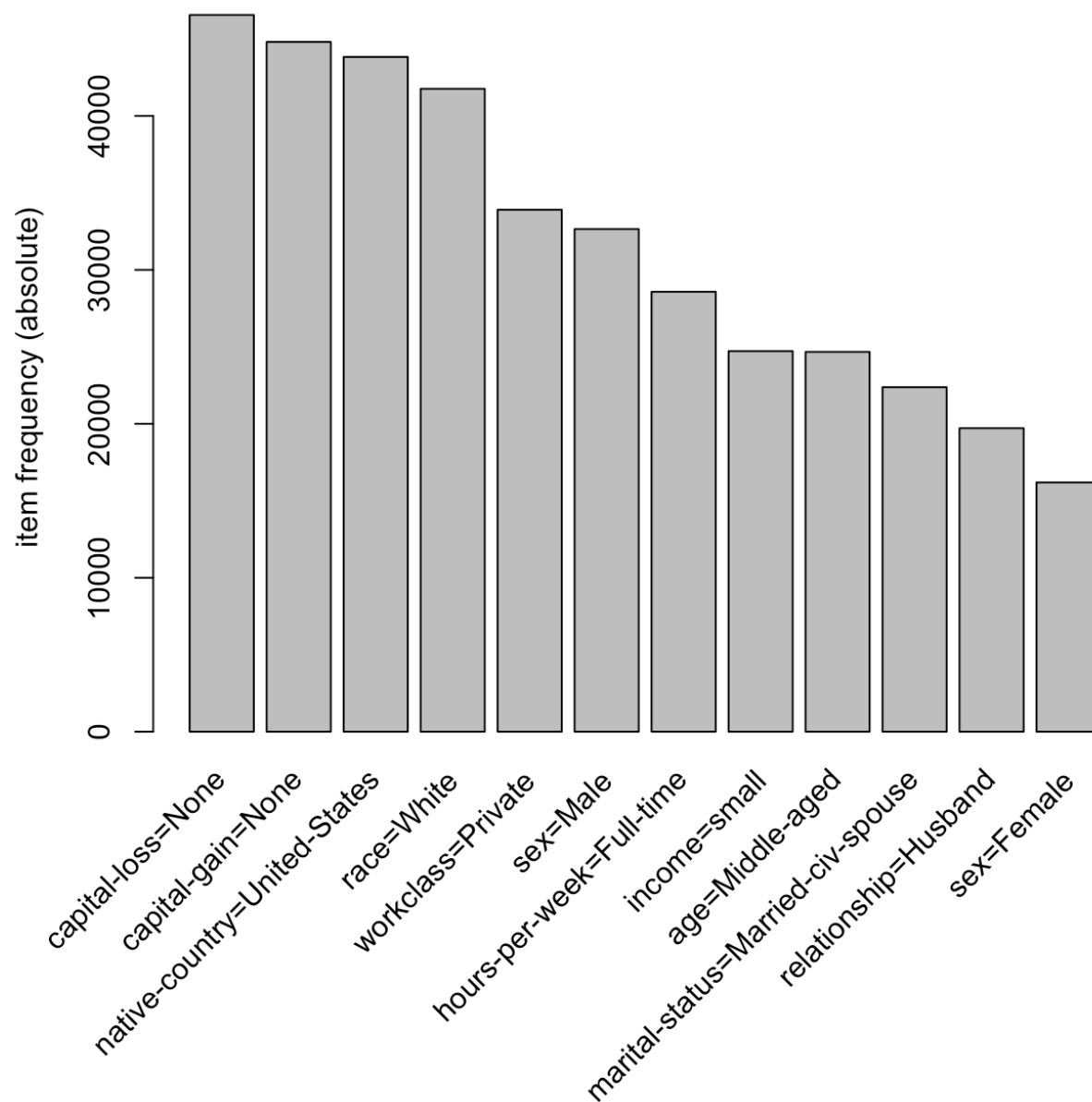


# Hw14

20150056 국윤범

## Problem 1

(a)

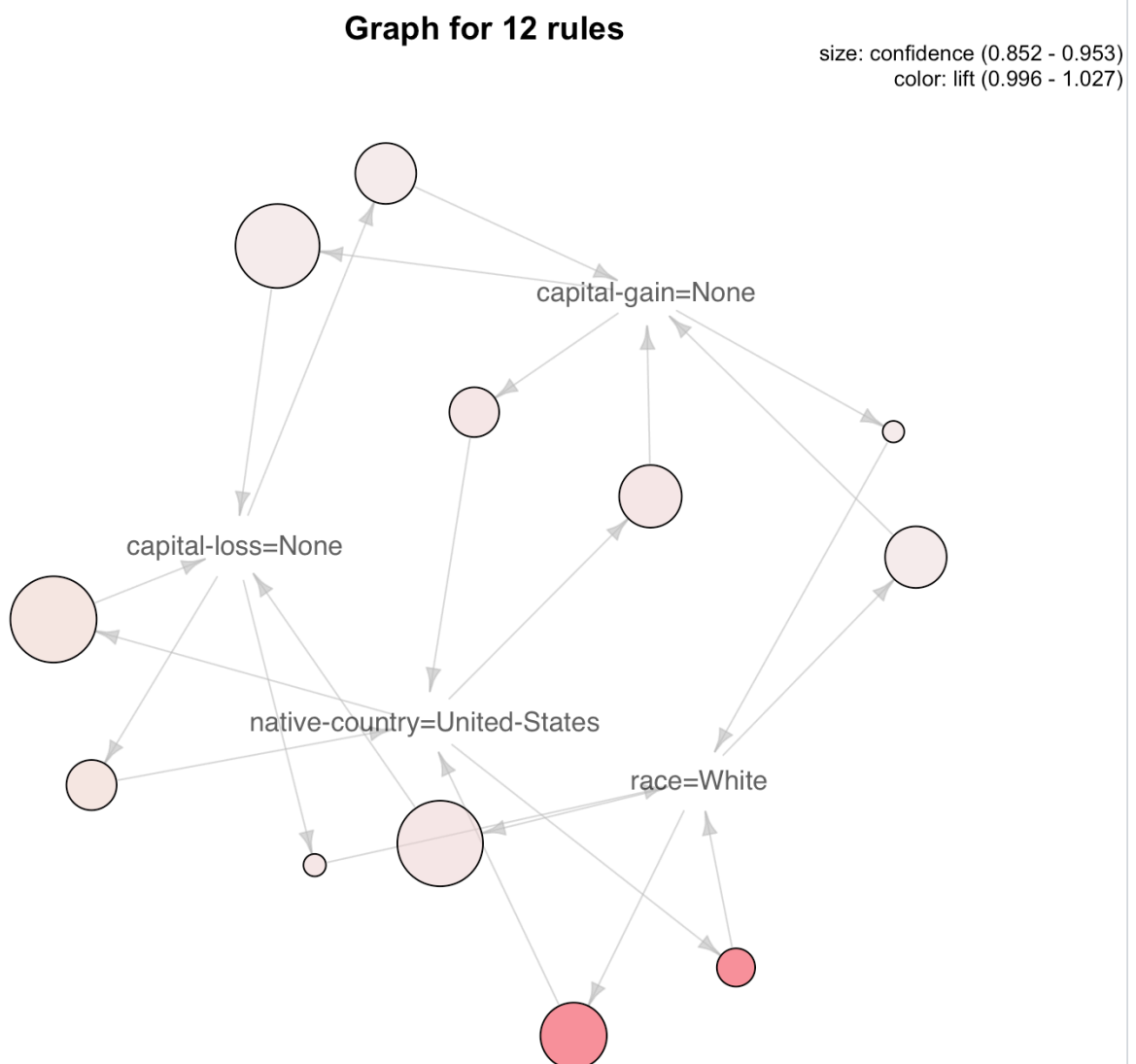


(b) The following is the predominant 12 rules base on support.

```
> inspect(rules_sorted[1:12])
```

	lhs	rhs	support	confidence	lift	count
[1]	{capital-gain=None}	=> {capital-loss=None}	0.8707	0.9491	0.9956	42525
[2]	{capital-loss=None}	=> {capital-gain=None}	0.8707	0.9133	0.9956	42525
[3]	{native-country=United-States}	=> {capital-loss=None}	0.8548	0.9525	0.9992	41752
[4]	{capital-loss=None}	=> {native-country=United-States}	0.8548	0.8967	0.9992	41752
[5]	{native-country=United-States}	=> {capital-gain=None}	0.8220	0.9159	0.9984	40146
[6]	{capital-gain=None}	=> {native-country=United-States}	0.8220	0.8960	0.9984	40146
[7]	{race=White}	=> {capital-loss=None}	0.8137	0.9516	0.9983	39742
[8]	{capital-loss=None}	=> {race=White}	0.8137	0.8536	0.9983	39742
[9]	{race=White}	=> {native-country=United-States}	0.7881	0.9217	1.0271	38493
[10]	{native-country=United-States}	=> {race=White}	0.7881	0.8782	1.0271	38493
[11]	{race=White}	=> {capital-gain=None}	0.7818	0.9143	0.9967	38184
[12]	{capital-gain=None}	=> {race=White}	0.7818	0.8522	0.9967	38184

Then, the following is the graph to visualize above mined rules.



From the above, the items which has the most number (6) of arrows are “capital-loss=None”, “capital-gain=None”, and “native-country=United-States”. From (a), they actually rank from 1<sup>st</sup> to 3<sup>rd</sup>.

(c) I calculate the crosstable for the dataset Adult. Since it's too huge, I only print out for index=1,...,6.

```
> tab[1:6,1:6]
```

	age=Young	age=Middle-aged	age=Senior	age=Old	workclass=Federal-gov	workclass=Local-gov
age=Young	9627	0	0	0	101	302
age=Middle-aged	0	24671	0	0	754	1659
age=Senior	0	0	12741	0	552	1070
age=Old	0	0	0	1803	25	105
workclass=Federal-gov	101	754	552	25	1432	0
workclass=Local-gov	302	1659	1070	105	0	3136

Then, the joint probability that someone's age is Middle-aged and someone's workclass happens to be Private can computed as follows:

```
> tab['age=Middle-aged', 'workclass=Private'] / tab['age=Middle-aged', 'age=Middle-aged']
[1] 0.7239
```

That is 0.7239.