Exercise: Adult Data Set

1. Changing the pre-processing

```
# delete unknown meaning variables
AdultUCI[["education-num"]]=NULL
AdultUCI[["fnlwgt"]]=NULL
```

`age` categorized with levels "Young", "Middle-aged", "Senior", "Old";
`hours-per-week` categorized with levels "Part-time", "Full-time", "Over-time", "Workaholic";
`workclass` reduced from 8 to 5 subgroups;
`education` reduced subgroups;
`marital-status` reduced to just "Never-Married", "Married", "Not-Married", "Widowed";
`capital-gain` categorized as "None", "Low", "High";
`capital-loss` categorized as "None", "Low", "High".

2. NAs

```
# omit observations with NA data
is.na(AdultUCI) = AdultUCI=='?'
is.na(AdultUCI) = AdultUCI==' ?'
AdultUCI = na.omit(AdultUCI)
```

3. Changing support and confidence

`support=0.002`, `conf=0.08` for most `apriori` runs;
however switched to 0.05 to get results with:

```
# 8 - capital loss low - adjusted supp an conf
capitallosslow<-apriori(data=Adult, parameter=list(supp=0.002,conf = 0.05),
                        appearance = list(default="lhs",rhs="capital-loss=Low"),
                        control = list(verbose=F))
```

4. Focus on specific RHSs

10 apriori were executed for different RHSs:

`#1 income=0`

```
> income0<-apriori(data=Adult, parameter=list(supp=0.002,conf = 0.08),
+                  appearance = list(default="lhs",rhs="income=0"),
+                  control = list(verbose=F))
> income0<-sort(income0, decreasing=TRUE,by="lift")
> inspect(income0[1:5])
    lhs                                                      rhs         support confidence lift
93  {education=Dropout,occupation=Priv-house-serv}           => {income=0} 0.0023  1          1.3
94  {marital-status=Never-Married,occupation=Priv-house-serv} => {income=0} 0.0021  1          1.3
100 {occupation=Priv-house-serv,capital-gain=None}           => {income=0} 0.0045  1          1.3
150 {relationship=Other-relative,native-country=Mexico}      => {income=0} 0.0029  1          1.3
287 {occupation=Machine-op-inspct,relationship=Other-relative} => {income=0} 0.0025  1          1.3
```

Being private house service with low education with more individuals earning less than 50k $ per year.

#2 – Doctorates

```
> doctorates<-apriori(data=Adult, parameter=list(supp=0.002,conf = 0.08),
+                  appearance = list(default="lhs",rhs="education=Doctorate"),
+                  control = list(verbose=F))
 lhs                              rhs                   support confidence lift
1 {age=Senior,
    occupation=Prof-specialty,
    sex=Male,
    capital-gain=None,
    income=1}                  => {education=Doctorate}  0.0023       0.17   14
```

3 – married

```
married<-apriori(data=Adult, parameter=list(supp=0.002,conf = 0.08),
+                    appearance = list(default="lhs",rhs="marital-status=Married"),
+                    control = list(verbose=F))
> married<-sort(married, decreasing=TRUE,by="confidence")
> inspect(married[1:5])
    lhs                                                rhs                      support confidence lift
59 {relationship=Husband,native-country=Philippines} => {marital-status=Married} 0.0025  1          2.1
```

4 - capital gain none

```
     lhs                                                rhs                 support confidence lift
22  {capital-loss=High}                              => {capital-gain=None} 0.0230  1          1.1
23  {capital-loss=Low}                               => {capital-gain=None} 0.0243  1          1.1
204 {age=Young,native-country=Mexico}                => {capital-gain=None} 0.0059  1          1.1
208 {marital-status=Never-Married,native-country=Mexico} => {capital-gain=None} 0.0073  1          1.1
245 {education=Masters,capital-loss=High}            => {capital-gain=None} 0.0033  1          1.1
```

5 - capital gain low

```
lhs                              rhs                support confidence lift
1 {age=Old,
   race=White,
   capital-loss=None,
   income=0}                   => {capital-gain=Low}  0.0022       0.12  2.5
```

6 - capital gain high

```
lhs                              rhs                 support confidence lift
1 {education=Prof-School,
   marital-status=Married,
   capital-loss=None,
   hours-per-week=Over-time,
   income=1}                   => {capital-gain=High}  0.0021       0.39   11
```

7 - capital loss none

```
lhs                                rhs                 support confidence lift
8  {native-country=Jamaica}     => {capital-loss=None} 0.0027  1          1
```

8 - capital loss low - adjusted `conf`

```
 lhs                                                                rhs                support confidence lift
1 {education=HS-grad,capital-gain=None,income=1}                  => {capital-loss=Low} 0.0023  0.053      2.2
2 {age=Middle-aged,capital-gain=None,income=1}                    => {capital-loss=Low} 0.0059  0.052      2.1
3 {age=Middle-aged,capital-gain=None,hours-per-week=Over-time,income=1} => {capital-loss=Low} 0.0028  0.053      2.2
```

9 - capital loss high

```
  lhs                              rhs                 support confidence lift
1 {education=Masters,
   marital-status=Married,
   race=White,
   capital-gain=None,
   income=1}                   => {capital-loss=High}  0.0025       0.14  6.2
2 {education=Masters,
   marital-status=Married,
   capital-gain=None,
   native-country=United-States,
   income=1}                   => {capital-loss=High}  0.0025       0.14  6.2
3 {education=Masters,
   marital-status=Married,
   race=White,
   capital-gain=None,
   native-country=United-States,
   income=1}                   => {capital-loss=High}  0.0024       0.14  6.2
```

10 - senior (25-45) profile

	lhs	rhs	support	confidence	lift
1	{workclass=**Private**, education=**Bachelors**, relationship=Wife, race=**White**, **capital-gain=None**, **capital-loss=None**, native-country=**United-States**, **income=1**}	=> {age=Middle-aged}	0.0021	0.91	1.8
2	{workclass=**Private**, education=**Bachelors**, marital-status=Married, relationship=Wife, race=**White**, **capital-gain=None**, **capital-loss=None**, native-country=**United-States**, **income=1**}	=> {age=Middle-aged}	0.0021	0.91	1.8
3	{workclass=**Private**, education=**Bachelors**, relationship=Wife, race=**White**, sex=Female, **capital-gain=None**, **capital-loss=None**, native-country=**United-States**, **income=1**}	=> {age=Middle-aged}	0.0021	0.91	1.7