Data Mining Assignment 1

Lina Mi

Student ID: @01377283

1. Read the *usedcars.csv* file into R studio:

```
>usedcars<-read.csv('M:/milina/Registered_Courses_2018_Spring/Data Mining/week1/usedcars.csv', header=TRUE, sep=',')
```

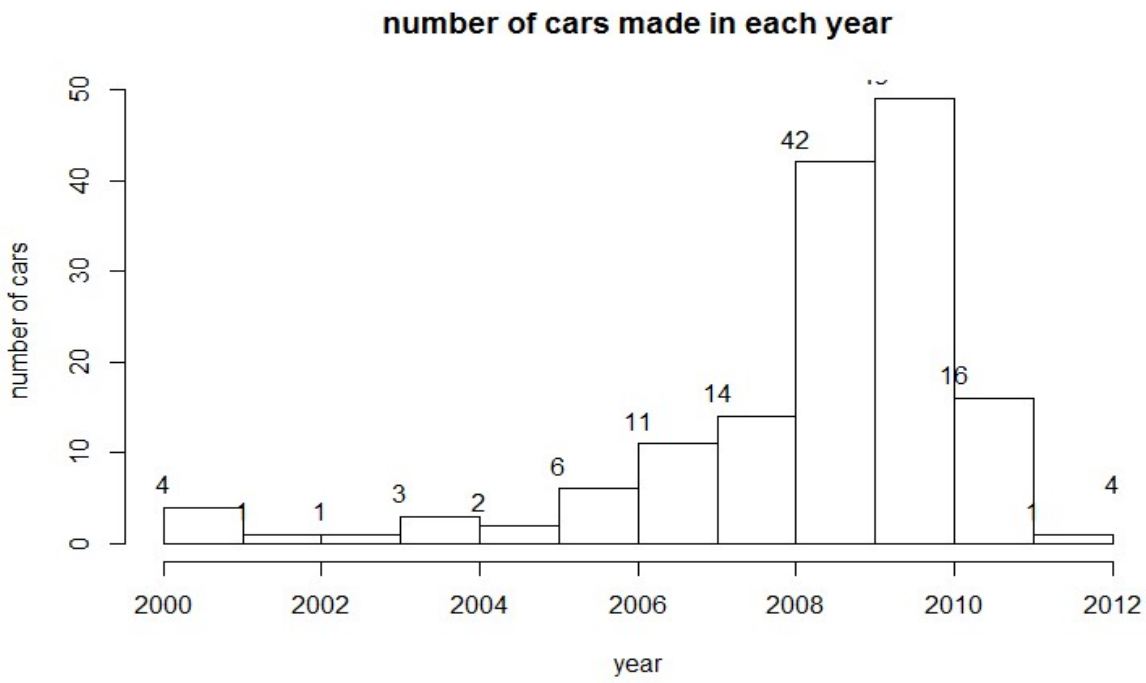
2. Using *str()* command to investigate the structure of data frame *usedcars*

```
>str(usedcars)
'data.frame': 150 obs. of  6 variables:
 $ year        : int  2011 2011 2011 2011 2012 2010 2011 2010 2011 2010 ...
 $ model       : Factor w/ 3 levels "SE","SEL","SES": 2 2 2 2 1 2 2 2 3 3 ...
 $ price       : int  21992 20995 19995 17809 17500 17495 17000 16995 16995 16995
 ...
 $ mileage     : int  7413 10926 7351 11613 8367 25125 27393 21026 32655 36116 ...
 $ color       : Factor w/ 9 levels "Black","Blue",..: 9 4 7 4 8 7 2 7 7 7 ...
 $ transmission: Factor w/ 2 levels "AUTO","MANUAL": 1 1 1 1 1 1 1 1 1 1 ...
```

3. Use *his()* to display the number of cars made in each year

```
a<-hist(usedcars$year, freq=TRUE, breaks=c(2000,2001,2002,2003,2004,2005,2006,2007,
2008,2009,2010,2011,2012),xlab="year", ylab="number of cars", main="number of cars
 made in each year" )
> text(a$breaks, a$counts,labels=a$counts,pos=3)
```

The resulting plot is show as following:



From this plot, it can be seen that the used cars that were made in 2008 and 2009 are most welcomed by people, since among the sold used cars, the numbers of cars made in 2008 and 2009 are much more than that of cars made in other years, 42 and 49 separately. the cars made in 2001, 2002, 2011 are much less favorite by people, because the number of used cars made those years is only one .