dm practice (10.01) solution

1. Data Import - 외부 데이터를 불러 들이는 R code 를 작성하시오.

a) File Name : Forbes2000.txt, b) Directory Path : C:\data mining

```
read.table("C:/data mining/Forbes2000.txt", sep = "\t", header = T)
```

```
## 혹은
setwd("C:/data mining")
read.table("Forbes2000.txt", sep = "\t", header = T)
```

2. Data Export - R 에서 생성된 Data Frame Object 를 txt 파일로 저장하는 R code 를 작성하시오.

a) Object : Forbes2000, b) Separator : Tab

```
write.table(Forbes2000, "Forbes2000.txt", sep = "\t")
```

3. Download Data

UCI Repository 에 있는 adult 데이터

```
url <- "http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adu
lt.data"
download.file(url, destfile = "adult.data")
```

```
## Warning: InternetOpenUrl failed: '작업 시간을 초과했습니다.'
```

```
## Error: cannot open URL
## 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.d
ata'
```

4 번 ~ 10 번 문항은 Forbes2000 데이터를 사용하여 문제 해결

```
data(Forbes2000, package='HSAUR')
```

4. Forbes2000 데이터의 각 Column 에 대한 summary 를 list 형식으로 나타내는 R code 를 작성하시오.

```
data(Forbes2000, package = "HSAUR")
```

```
lapply(Forbes2000, summary)
```

```
## $rank
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##       1     501    1000    1000    1500    2000
##
## $name
##    Length     Class      Mode
##      2000 character character
##
## $country
##                         Africa                      Australia
##                              2                             37
##    Australia/ United Kingdom                        Austria
##                              2                              8
##                        Bahamas                        Belgium
##                              1                              9
##                        Bermuda                         Brazil
##                             20                             15
##                         Canada                  Cayman Islands
##                             56                              5
##                          Chile                          China
##                              4                             25
##                Czech Republic                        Denmark
##                              2                             10
##                        Finland                         France
##                             11                             63
##       France/ United Kingdom                         Germany
##                              1                             65
##                         Greece                Hong Kong/China
##                             12                             20
##                        Hungary                          India
##                              2                             27
##                      Indonesia                        Ireland
##                              7                              8
##                        Islands                         Israel
##                              1                              8
##                          Italy                          Japan
##                             41                            316
##                         Jordan                       Kong/China
##                              1                              4
##                          Korea                        Liberia
##                              4                              1
##                     Luxembourg                       Malaysia
##                              2                             16
##                         Mexico                     Netherlands
##                             17                             28
##  Netherlands/ United Kingdom                     New Zealand
##                              2                              1
##                         Norway                       Pakistan
##                              8                              1
##       Panama/ United Kingdom                           Peru
##                              1                              1
##                    Philippines                         Poland
##                              2                              1
##                       Portugal                         Russia
##                              7                             12
```

```
##                     Singapore                   South Africa
##                            16                             15
##                   South Korea                          Spain
##                            45                             29
##                        Sweden                    Switzerland
##                            26                             34
##                        Taiwan                       Thailand
##                            35                              9
##                        Turkey                 United Kingdom
##                            12                            137
##    United Kingdom/ Australia  United Kingdom/ Netherlands
##                             1                              1
## United Kingdom/ South Africa                 United States
##                             1                            751
##                     Venezuela
##                             1
##
## $category
##               Aerospace & defense                          Banking
##                                19                              313
##     Business services & supplies                    Capital goods
##                                70                               53
##                          Chemicals                    Conglomerates
##                                50                               31
##                       Construction                Consumer durables
##                                79                               74
##            Diversified financials            Drugs & biotechnology
##                               158                               45
##              Food drink & tobacco                     Food markets
##                                83                               33
## Health care equipment & services     Hotels restaurants & leisure
##                                65                               37
##    Household & personal products                          Insurance
##                                44                              112
##                          Materials                            Media
##                                97                               61
##              Oil & gas operations                        Retailing
##                                90                               88
##                    Semiconductors              Software & services
##                                26                               31
##  Technology hardware & equipment      Telecommunications services
##                                59                               67
##                 Trading companies                   Transportation
##                                25                               80
##                         Utilities
##                               110
##
## $sales
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.01    2.02    4.36    9.70    9.55  256.00
##
## $profits
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
```

```
## -25.800   0.080   0.200   0.381   0.440  21.000       5
##
## $assets
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.3     4.0     9.3    34.0    22.8  1260.0
##
## $marketvalue
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     2.7     5.2    11.9    10.6   329.0
```

5. profits 이 높은 상위 50 개의 기업에 대한 Plot 을 그리는 R code 를 작성하시오.

sales against assets (or some suitable transformation of each variable)

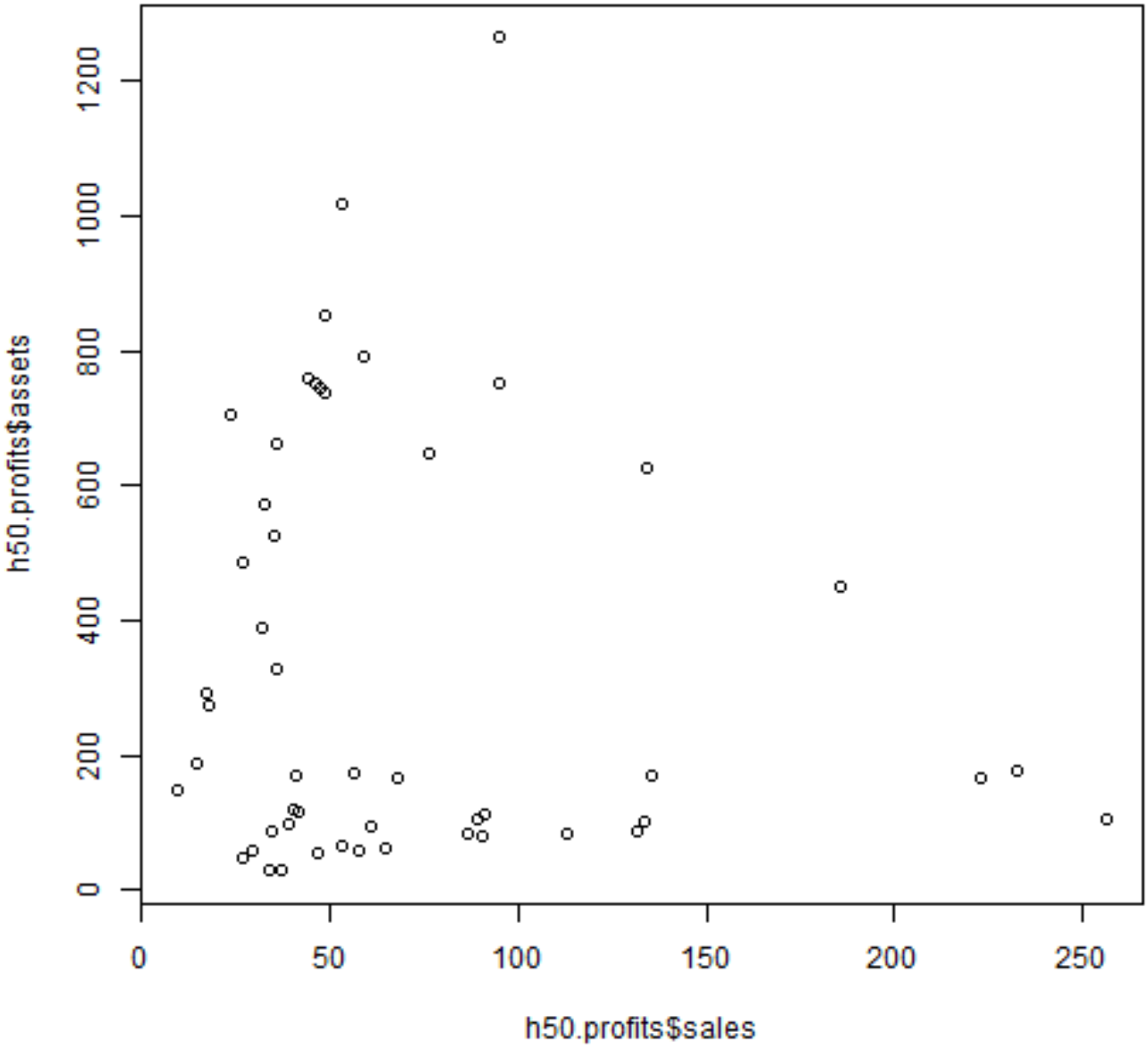
```
order.profits <- order(Forbes2000$profits, decreasing = TRUE, na.last = TR
UE)
h50.profits <- Forbes2000[which(order.profits <= 50), ]
h50.profits
```

```
##     rank                    name                    country
## 1      1                Citigroup              United States
## 2      2         General Electric              United States
## 3      3      American Intl Group              United States
## 4      4                ExxonMobil              United States
## 5      5                       BP             United Kingdom
## 6      6          Bank of America              United States
## 7      7               HSBC Group             United Kingdom
## 8      8             Toyota Motor                      Japan
## 9      9               Fannie Mae              United States
## 10    10          Wal-Mart Stores              United States
## 11    11                      UBS                Switzerland
## 12    12                ING Group                Netherlands
## 13    13 Royal Dutch/Shell Group Netherlands/ United Kingdom
## 14    14       Berkshire Hathaway              United States
## 16    16                      IBM              United States
## 17    17                    Total                     France
## 18    18               BNP Paribas                    France
## 19    19  Royal Bank of Scotland             United Kingdom
## 20    20               Freddie Mac              United States
## 22    22             Altria Group              United States
## 23    23            ChevronTexaco              United States
## 24    24                   Pfizer              United States
## 25    25              Wells Fargo              United States
## 26    26          Verizon Commun               United States
## 29    29          General Motors               United States
## 31    31               Microsoft               United States
## 32    32                  Nestle                 Switzerland
## 33    33       SBC Communications              United States
## 34    34     Deutsche Bank Group                    Germany
## 35    35           Siemens Group                    Germany
## 36    36                    HBOS             United Kingdom
```

```
## 37    37                    ENI                    Italy
## 38    38          ConocoPhillips            United States
## 40    40           Merrill Lynch            United States
## 44    44        Procter & Gamble            United States
## 48    48        ABN-Amro Holding              Netherlands
## 51    51            Nissan Motor                    Japan
## 54    54  Societe Generale Group                   France
## 55    55              PetroChina                    China
## 59    59                 MetLife            United States
## 64    64          Novartis Group              Switzerland
## 67    67       Washington Mutual            United States
## 69    69           Deutsche Post                  Germany
## 71    71        Volkswagen Group                  Germany
## 72    72    Tokyo Electric Power                    Japan
## 77    77              US Bancorp            United States
## 80    80         GlaxoSmithKline           United Kingdom
## 83    83                   Nokia                  Finland
## 103  103    Lehman Bros Holdings            United States
## 149  149   Westpac Banking Group                 Australia
##                            category  sales profits  assets marketvalue
## 1                           Banking  94.71   17.85 1264.03      255.30
## 2                     Conglomerates 134.19   15.59  626.93      328.54
## 3                         Insurance  76.66    6.46  647.66      194.87
## 4              Oil & gas operations 222.88   20.96  166.99      277.02
## 5              Oil & gas operations 232.57   10.27  177.57      173.54
## 6                           Banking  49.01   10.81  736.45      117.55
## 7                           Banking  44.33    6.66  757.60      177.96
## 8                 Consumer durables 135.82    7.99  171.71      115.40
## 9            Diversified financials  53.13    6.48 1019.17       76.84
## 10                         Retailing 256.33    9.05  104.91      243.74
## 11           Diversified financials  48.95    5.15  853.23       85.07
## 12           Diversified financials  94.72    4.73  752.49       54.59
## 13             Oil & gas operations 133.50    8.40  100.72      163.45
## 14                        Insurance  56.22    6.95  172.24      141.14
## 16  Technology hardware & equipment  89.13    7.58  104.46      171.54
## 17             Oil & gas operations 131.64    8.84   87.84      116.64
## 18                          Banking  47.74    4.73  745.09       59.29
## 19                          Banking  35.65    4.95  663.45       90.21
## 20           Diversified financials  46.26   10.09  752.25       44.25
## 22             Food drink & tobacco  60.70    9.20   96.18      111.02
## 23             Oil & gas operations 112.94    7.43   82.36       92.49
## 24            Drugs & biotechnology  40.36    6.20  120.06      285.27
## 25                          Banking  31.80    6.20  387.80       97.53
## 26      Telecommunications services  67.75    2.57  165.97      103.97
## 29                Consumer durables 185.52    3.82  450.00       27.47
## 31              Software & services  34.27    8.88   85.94      287.02
## 32             Food drink & tobacco  64.56    5.48   62.15      106.55
## 33      Telecommunications services  39.16    5.97  100.17       82.93
## 34           Diversified financials  58.85    1.53  792.49       50.23
## 35                    Conglomerates  86.62    2.81   85.47       75.77
## 36                          Banking  32.68    3.09  571.76       52.87
## 37             Oil & gas operations  53.29    4.82   67.91       76.13
## 38             Oil & gas operations  90.49    4.83   81.95       46.72
```

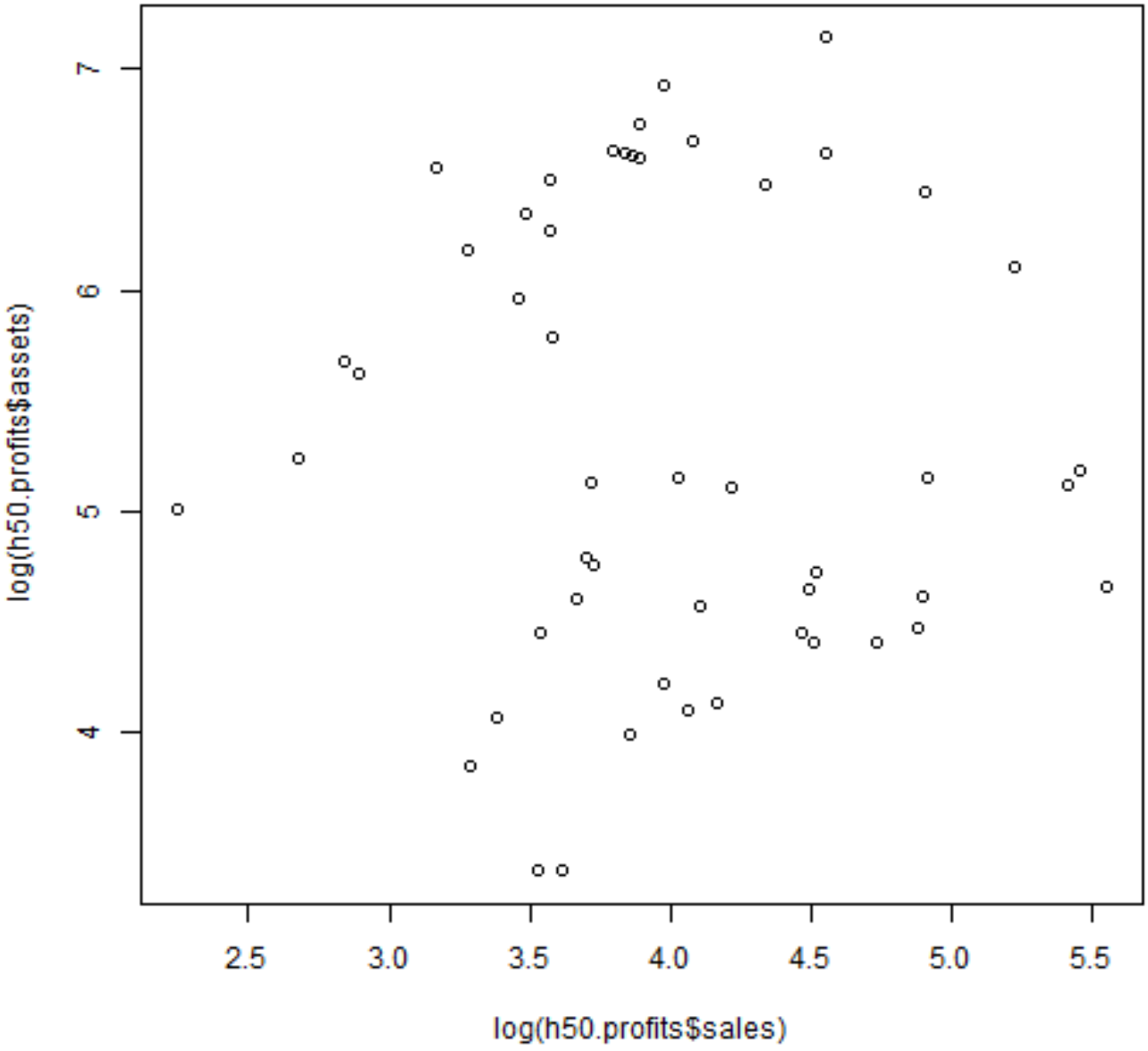
```
## 40           Diversified financials  26.64    3.47  485.77       57.52
## 44    Household & personal products  46.99    5.81   53.86      131.89
## 48                          Banking  23.64    3.98  704.95       39.29
## 51              Consumer durables  57.77    4.19   60.56       41.71
## 54                          Banking  35.52    1.61  526.54       40.61
## 55           Oil & gas operations  29.53    5.67   58.36       90.49
## 59                      Insurance  35.79    2.24  326.84       26.34
## 64          Drugs & biotechnology  26.77    5.40   46.92      116.43
## 67                          Banking  18.01    3.88  275.18       39.69
## 69                Transportation  41.23    1.64  169.33       26.83
## 71              Consumer durables  91.33    2.71  112.87       17.42
## 72                      Utilities  41.62    1.40  116.68       30.63
## 77                          Banking  14.57    3.73  189.29       52.88
## 80          Drugs & biotechnology  34.16    6.34   29.19      124.79
## 83  Technology hardware & equipment  37.05    4.52   29.15      104.30
## 103          Diversified financials  17.10    1.47  291.64       23.01
## 149                         Banking   9.45    1.49  150.08       24.44
```

```r
plot(h50.profits$sales, h50.profits$assets)
```



plot of chunk unnamed-chunk-5

```
plot(log(h50.profits$sales), log(h50.profits$assets))
```



plot of chunk unnamed-chunk-5

6. 각 나라별 sales 평균을 구하는 R code 를 작성하시오.

```
tapply(Forbes2000$sales, Forbes2000$country, mean)
```

```
##                      Africa                   Australia
##                       6.820                       5.245
##    Australia/ United Kingdom                     Austria
##                      11.595                       4.143
##                     Bahamas                     Belgium
##                       1.350                      10.114
```

```
##                       Bermuda                        Brazil
##                         6.841                         6.339
##                        Canada                Cayman Islands
##                         6.430                         1.660
##                         Chile                         China
##                         1.603                         5.100
##                Czech Republic                       Denmark
##                         1.805                         6.349
##                       Finland                        France
##                        10.292                        20.102
##       France/ United Kingdom                       Germany
##                         1.010                        20.781
##                        Greece              Hong Kong/China
##                         2.528                         2.044
##                       Hungary                         India
##                         3.370                         3.868
##                     Indonesia                       Ireland
##                         2.450                         4.765
##                       Islands                        Israel
##                         6.670                         2.060
##                         Italy                         Japan
##                        10.214                        10.191
##                        Jordan                    Kong/China
##                         1.330                         5.718
##                         Korea                       Liberia
##                        15.005                         3.780
##                    Luxembourg                      Malaysia
##                        14.185                         1.716
##                        Mexico                    Netherlands
##                         3.938                        17.021
##  Netherlands/ United Kingdom                    New Zealand
##                        92.100                         2.640
##                        Norway                       Pakistan
##                        10.780                         1.230
##       Panama/ United Kingdom                          Peru
##                         5.930                         0.170
##                   Philippines                        Poland
##                         1.565                         4.410
##                      Portugal                        Russia
##                         3.884                         7.673
##                     Singapore                  South Africa
##                         3.685                         4.124
##                   South Korea                         Spain
##                         7.969                         7.843
##                        Sweden                   Switzerland
##                         7.666                        12.457
##                        Taiwan                      Thailand
##                         2.751                         2.513
##                        Turkey                United Kingdom
##                         4.713                        10.445
##    United Kingdom/ Australia  United Kingdom/ Netherlands
##                        10.010                         7.540
## United Kingdom/ South Africa                United States
```

```
##                       2.060                       10.058
##                   Venezuela
##                       0.980
```

7. marketvalue 가 200 이상인 기업은 category 별로 몇개인지 나타내는 R code 를 작성하시오.

함수로 만들어서 결과가 나오게 만들 것. (예시 : ex6 <- function(dat){})

```
ex7 <- function(dat) {
    dat <- dat
    dat200 <- subset(dat, marketvalue >= 200)
    table(dat200$category)
}

ex7(Forbes2000)
```

```
##
##              Aerospace & defense                          Banking
##                                0                                1
##     Business services & supplies                  Capital goods
##                                0                                0
##                         Chemicals                  Conglomerates
##                                0                                1
##                      Construction               Consumer durables
##                                0                                0
##           Diversified financials           Drugs & biotechnology
##                                0                                1
##             Food drink & tobacco                     Food markets
##                                0                                0
## Health care equipment & services     Hotels restaurants & leisure
##                                0                                0
##    Household & personal products                        Insurance
##                                0                                0
##                         Materials                            Media
##                                0                                0
##             Oil & gas operations                        Retailing
##                                1                                1
##                   Semiconductors             Software & services
##                                0                                1
##  Technology hardware & equipment      Telecommunications services
##                                0                                0
##                Trading companies                   Transportation
##                                0                                0
##                        Utilities
##                                0
```

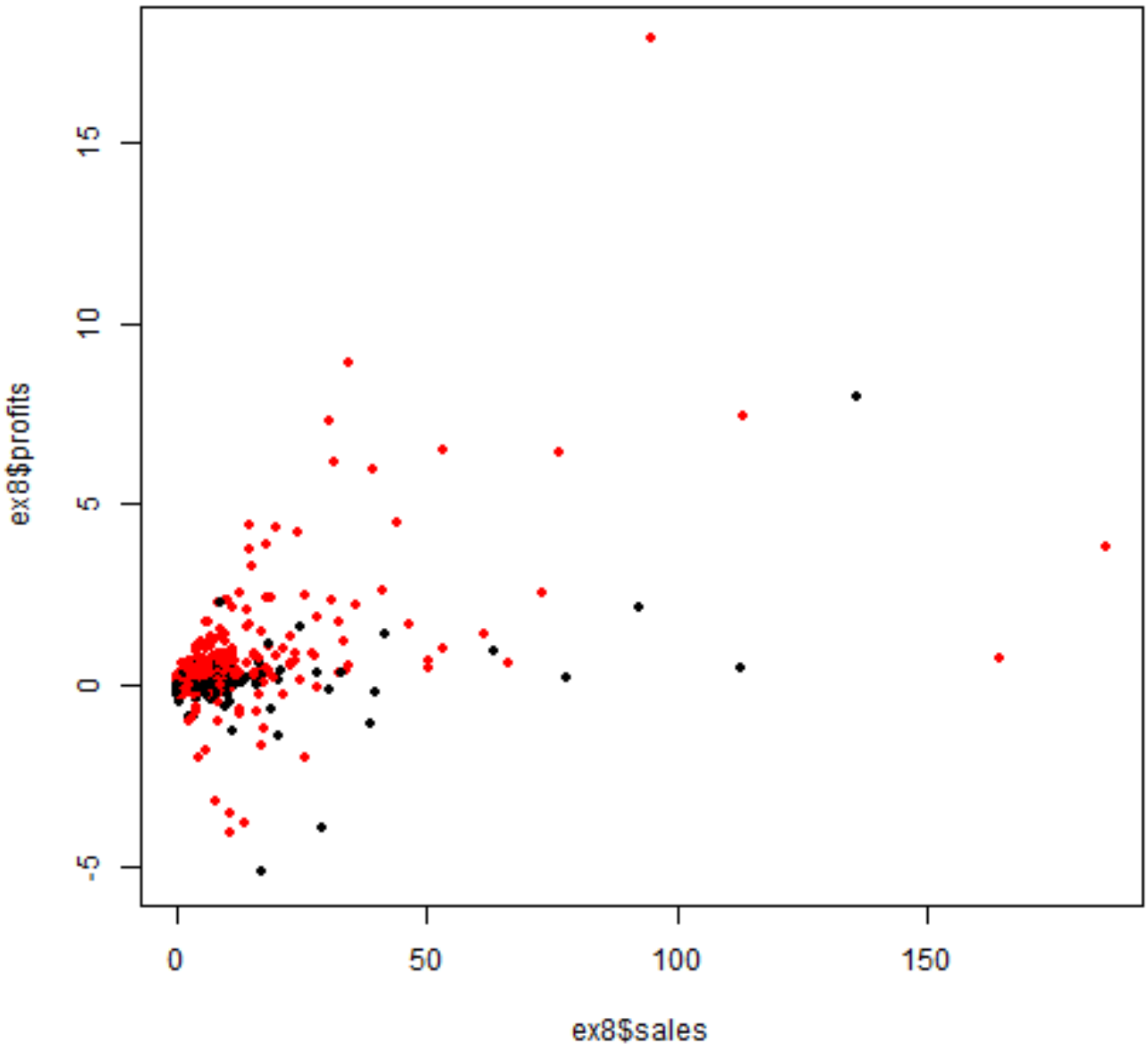
8. US 와 Japan 에 대한 Plot (sales vs. profits)을 그리는 R code 를 작성하시오.

US 와 Japan 은 각각 다른 색의 점으로 표시가 되게 할 것. (참고: help(plot) arguments)

```
ex8 <- subset(Forbes2000, country == c("United States", "Japan"))
ex8$country <- as.character(ex8$country)
ex8$country <- as.factor(ex8$country)
str(ex8)
```

```
## 'data.frame':    532 obs. of  8 variables:
##  $ rank       : int  1 3 8 9 15 23 25 29 30 31 ...
##  $ name       : chr  "Citigroup" "American Intl Group" "Toyota Motor" "
Fannie Mae" ...
##  $ country    : Factor w/ 2 levels "Japan","United States": 2 2 1 2 2 2
 2 2 1 2 ...
##  $ category   : Factor w/ 27 levels "Aerospace & defense",..: 2 16 8 9
2 19 2 8 24 22 ...
##  $ sales      : num  94.7 76.7 135.8 53.1 44.4 ...
##  $ profits    : num  17.85 6.46 7.99 6.48 4.47 ...
##  $ assets     : num  1264 648 172 1019 793 ...
##  $ marketvalue: num  255.3 194.9 115.4 76.8 81.9 ...
```

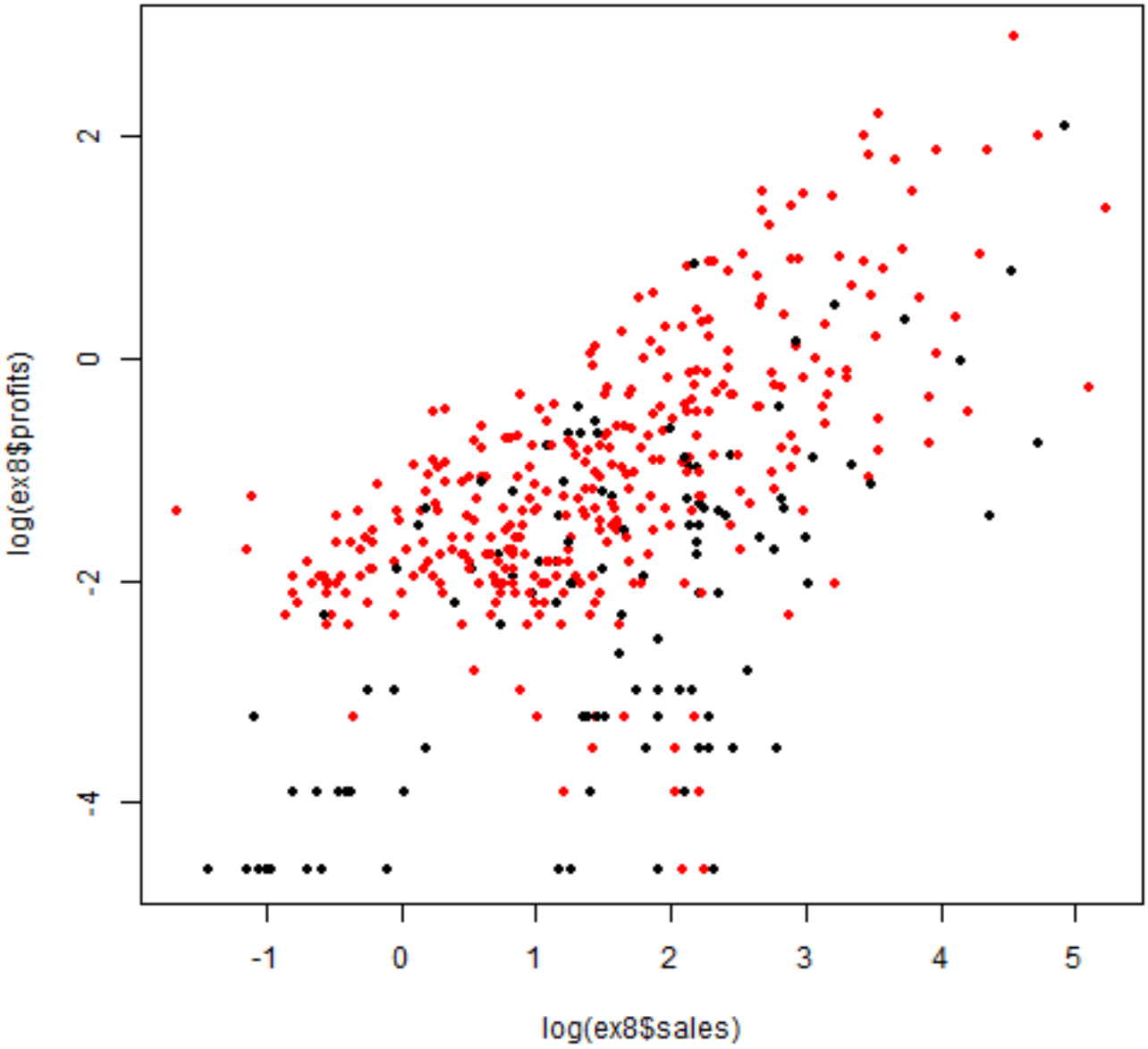
```
plot(ex8$sales, ex8$profits, col = ex8$country, pch = 20)
```



plot of chunk unnamed-chunk-8

```
# Log transformation
plot(log(ex8$sales), log(ex8$profits), col = ex8$countr, pch = 20)
```

```
## Warning: NaNs produced
```



plot of chunk unnamed-chunk-8