Homework 1

1. code for assignment 1:

   a. read **iowa.csv**

```
iowa.df = read.csv("iowa.csv", header=T, sep=";")
```

   b. how many rows and columns

```
sprintf("numbers of rows: %d", dim(iowa.df)[1])
```

```
## [1] "numbers of rows: 33"
```

```
sprintf("numbers of columns: %d", dim(iowa.df)[2])
```

```
## [1] "numbers of columns: 10"
```

   c. names of the columns

```
colnames(iowa.df)
```

```
##  [1] "Year"  "Rain0" "Temp1" "Rain1" "Temp2" "Rain2" "Temp3" "Rain3" "Temp4"
## [10] "Yield"
```

   d. the value of row 5, column 7

```
iowa.df[5,7]
```

```
## [1] 79.7
```

   e. display the second row

```
iowa.df[2,]
```

```
##   Year Rain0 Temp1 Rain1 Temp2 Rain2 Temp3 Rain3 Temp4 Yield
## 2 1931 14.76  57.5  3.83    75  2.72  77.2   3.3  72.6  32.9
```

2. code for assignment 2:

   a.

   - `vector1 <- c("5", "12", "7", "32")` is correct;
   - `max(vector1)` is correct, and the result is `"7"`;
   - `sort(vector1)` is correct, and the result is `"12" "32" "5"  "7"`;
   - `sum(vector1)` is wrong, because strings can't be summed.

   b.

   - ```
     vector2 <- c("5",7,12)
     vector2[2]+vector2[3]
     ```

     ```
     ## Error in vector2[2] + vector2[3]: non-numeric argument to binary operator
     ```
     ```

- ```r
  dataframe3 <- data.frame(z1="5", z2=7, z3=12)
  dataframe3[1,2]+dataframe3[1,3]
  ```

  ```
  ## [1] 19
  ```

- ```r
  list4 <- list(z1="6", z2=42, z3="49", z4=126)
  list4[[2]]+list4[[4]]
  ```

  ```
  ## [1] 168
  ```

  ```r
  list4[2]+list4[4]
  ```

  ```
  ## Error in list4[2] + list4[4]: non-numeric argument to binary operator
  ```

3. code for the assignment 3:

   a.

```r
seq(1, 10000, by=372)
```

```
##  [1]    1  373  745 1117 1489 1861 2233 2605 2977 3349 3721 4093 4465 4837 5209
## [16] 5581 5953 6325 6697 7069 7441 7813 8185 8557 8929 9301 9673
```

```r
seq(1, 10000, length.out=50)
```

```
##  [1]     1.0000   205.0612   409.1224   613.1837   817.2449  1021.3061
##  [7]  1225.3673  1429.4286  1633.4898  1837.5510  2041.6122  2245.6735
## [13]  2449.7347  2653.7959  2857.8571  3061.9184  3265.9796  3470.0408
## [19]  3674.1020  3878.1633  4082.2245  4286.2857  4490.3469  4694.4082
## [25]  4898.4694  5102.5306  5306.5918  5510.6531  5714.7143  5918.7755
## [31]  6122.8367  6326.8980  6530.9592  6735.0204  6939.0816  7143.1429
## [37]  7347.2041  7551.2653  7755.3265  7959.3878  8163.4490  8367.5102
## [43]  8571.5714  8775.6327  8979.6939  9183.7551  9387.8163  9591.8776
## [49]  9795.9388 10000.0000
```

   b.

```r
rep(1:3, times=3)
```

```
## [1] 1 2 3 1 2 3 1 2 3
```

```r
rep(1:3, each=3)
```

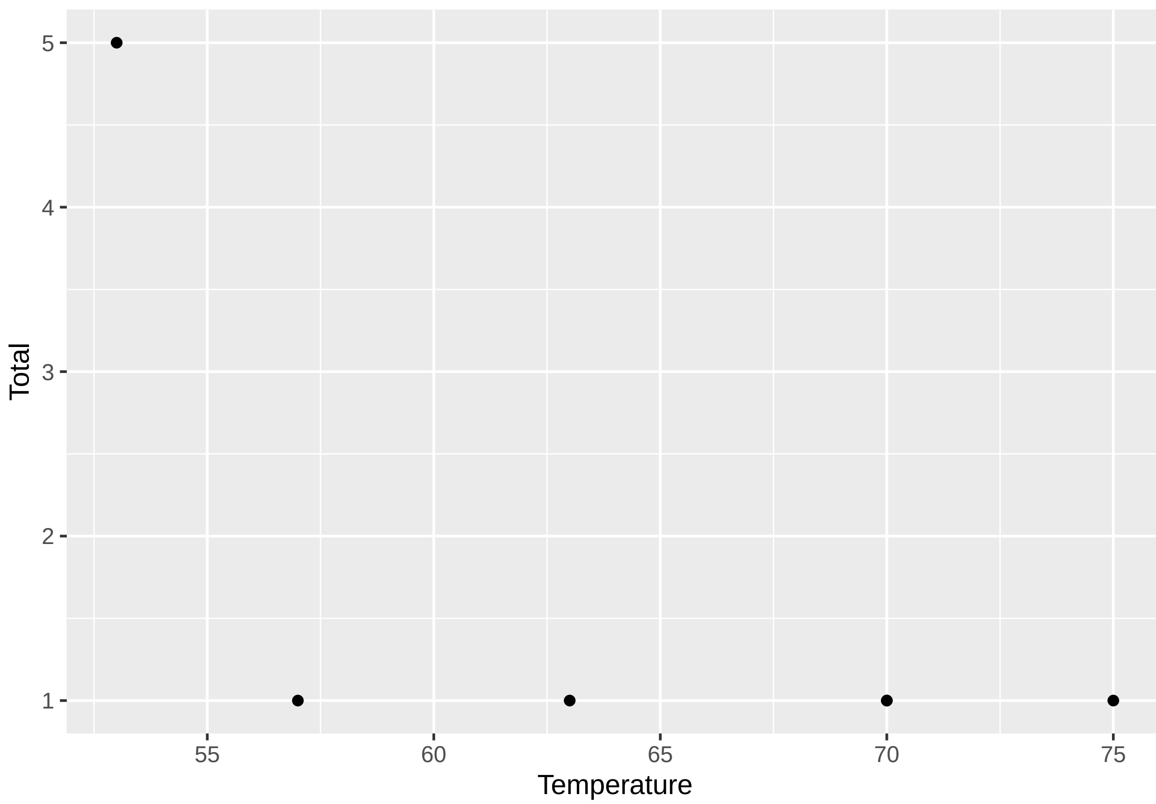
```
## [1] 1 1 1 2 2 2 3 3 3
```

4. code for assignment MB.Ch1.2:

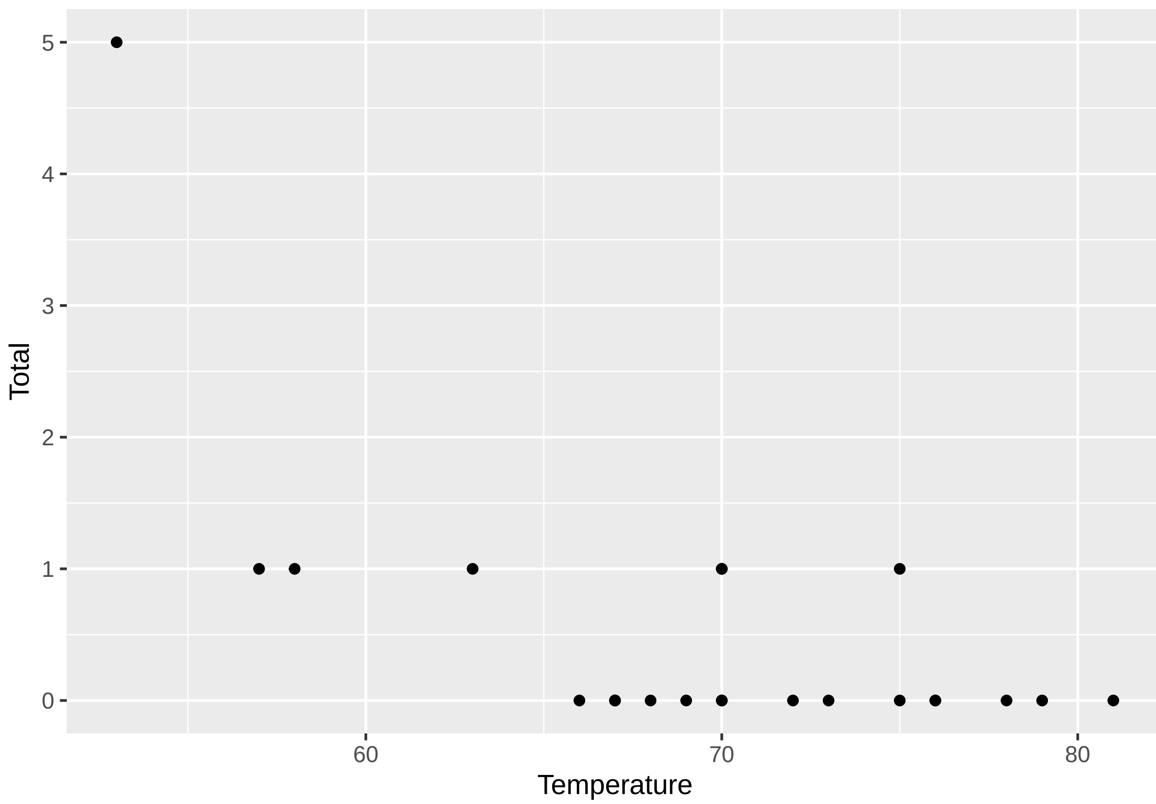
```r
library(DAAG)
library(ggplot2)
data("orings")
df <- data.frame(orings[c(1,2,4,11,13,18),])
head(df)
```

```
##    Temperature Erosion Blowby Total
## 1           53       3      2     5
## 2           57       1      0     1
## 4           63       1      0     1
## 11          70       1      0     1
## 13          70       1      0     1
## 18          75       0      2     1
```

```
ggplot(data=df,aes(x=Temperature,y=Total))+ geom_point()
```




```
ggplot(data=orings,aes(x=Temperature,y=Total))+ geom_point()
```



5. code for assignment MB.Ch1.4:

   a.

```
library(DAAG)
data("ais")
str(ais)
```

```
## 'data.frame':    202 obs. of  13 variables:
##  $ rcc    : num  3.96 4.41 4.14 4.11 4.45 4.1 4.31 4.42 4.3 4.51 ...
##  $ wcc    : num  7.5 8.3 5 5.3 6.8 4.4 5.3 5.7 8.9 4.4 ...
##  $ hc     : num  37.5 38.2 36.4 37.3 41.5 37.4 39.6 39.9 41.1 41.6 ...
##  $ hg     : num  12.3 12.7 11.6 12.6 14 12.5 12.8 13.2 13.5 12.7 ...
##  $ ferr   : num  60 68 21 69 29 42 73 44 41 44 ...
##  $ bmi    : num  20.6 20.7 21.9 21.9 19 ...
##  $ ssf    : num  109.1 102.8 104.6 126.4 80.3 ...
##  $ pcBfat: num  19.8 21.3 19.9 23.7 17.6 ...
##  $ lbm    : num  63.3 58.5 55.4 57.2 53.2 ...
##  $ ht     : num  196 190 178 185 185 ...
##  $ wt     : num  78.9 74.4 69.1 74.9 64.6 63.7 75.2 62.3 66.5 62.9 ...
##  $ sex    : Factor w/ 2 levels "f","m": 1 1 1 1 1 1 1 1 1 1 ...
##  $ sport : Factor w/ 10 levels "B_Ball","Field",..: 1 1 1 1 1 1 1 1 1 1 ...
```

   b.

```r
kinds <- table(ais["sport"])
col <- names(kinds)
# initialize the dataframe
freq_table_genders = data.frame(
    male   = seq(1,10),
    female = seq(1,10)
)
row.names(freq_table_genders) <- col
for(i in seq(1,10)){
    freq_table_genders[col[i],"male"] <-
            sum(ais$sex == "m" & ais$sport == col[i])
    freq_table_genders[col[i],"female"] <-
            sum(ais$sex == "f" & ais$sport == col[i])
}
# print the dataframe to show numbers of
# different genders in each sports
print(freq_table_genders)
```

```
##          male female
## B_Ball     12     13
## Field      12      7
## Gym         0      4
## Netball     0     23
## Row        15     22
## Swim       13      9
## T_400m     18     11
## T_Sprnt    11      4
## Tennis      4      7
## W_Polo     17      0
```

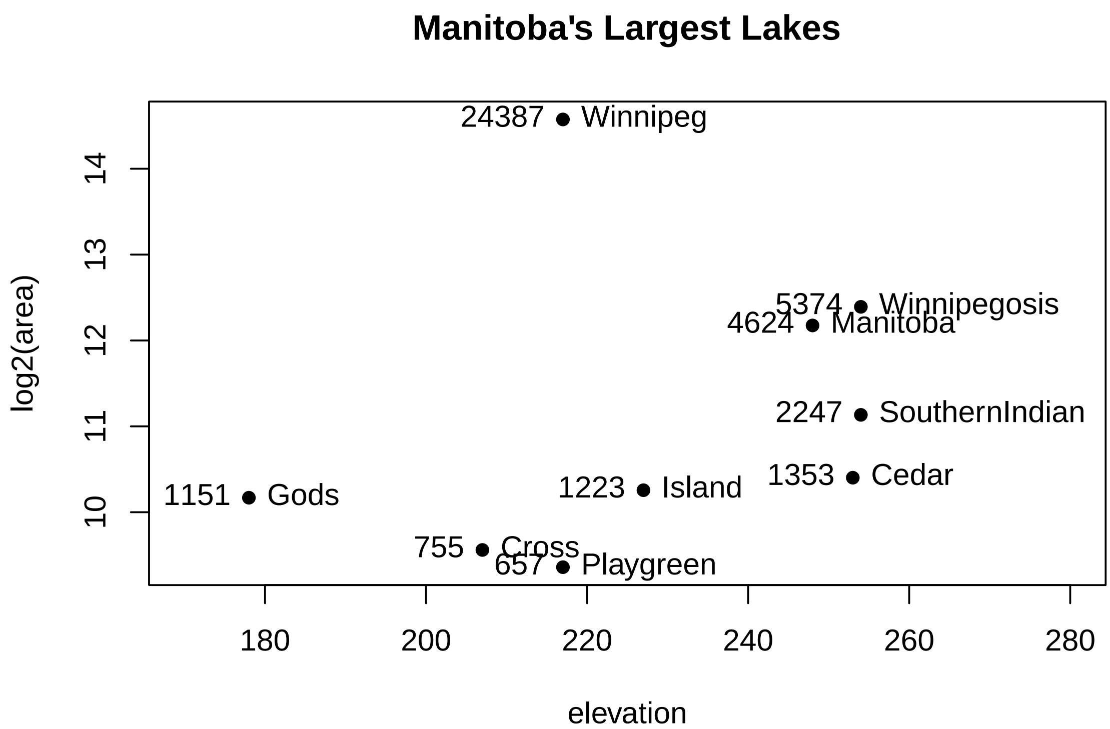
From above, we see that in **netball**, there's a large imbalance in numbers of the two sexes.

6. code for assignment MB.Ch1.6:

```r
elevation <- c(217,254,248,254,253,227,178,207,217)
area      <- c(24387, 5374, 4624, 2247, 1353,
                  1223, 1151, 755, 657)
Manitoba.lakes <- data.frame(elevation, area)
row.names(Manitoba.lakes) <- c("Winnipeg", "Winnipegosis",
        "Manitoba", "SouthernIndian", "Cedar", "Island",
        "Gods", "Cross", "Playgreen")
```

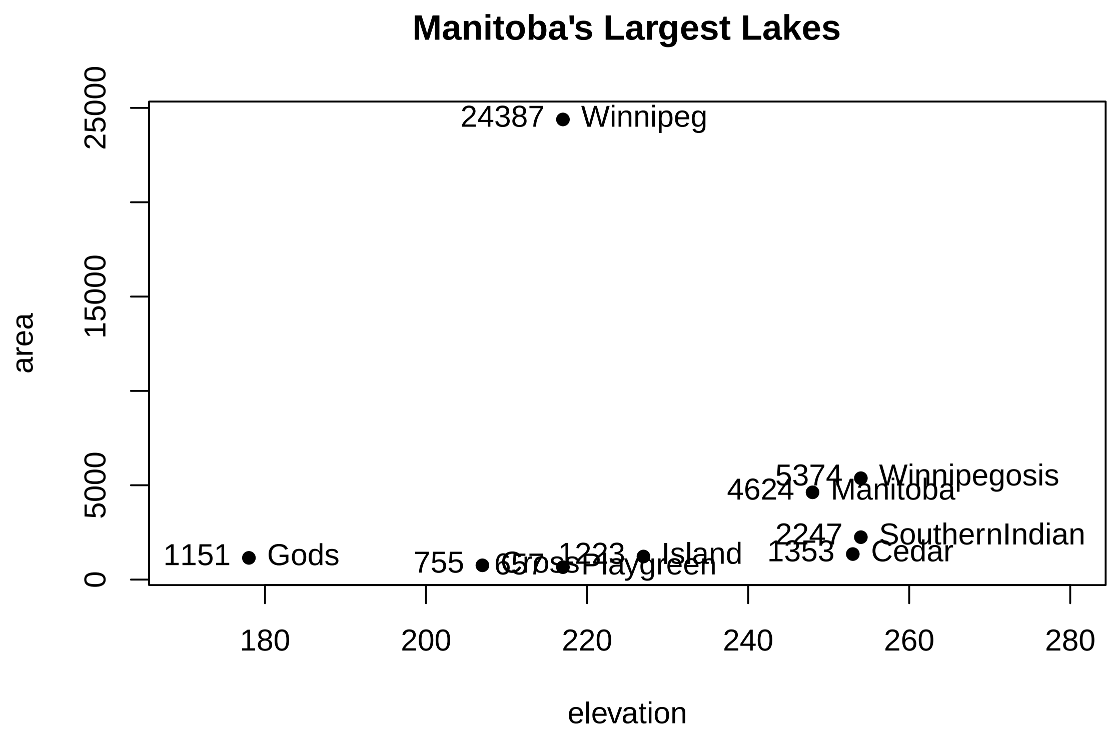
a.

```r
attach(Manitoba.lakes)
plot(log2(area)~elevation, pch=16, xlim=c(170,280))
# NB: Doubling the area increases log2(area) by 1.0
text(log2(area)~elevation, labels=row.names(Manitoba.lakes), pos=4)
text(log2(area)~elevation, labels=area, pos=2)
title("Manitoba's Largest Lakes")
```



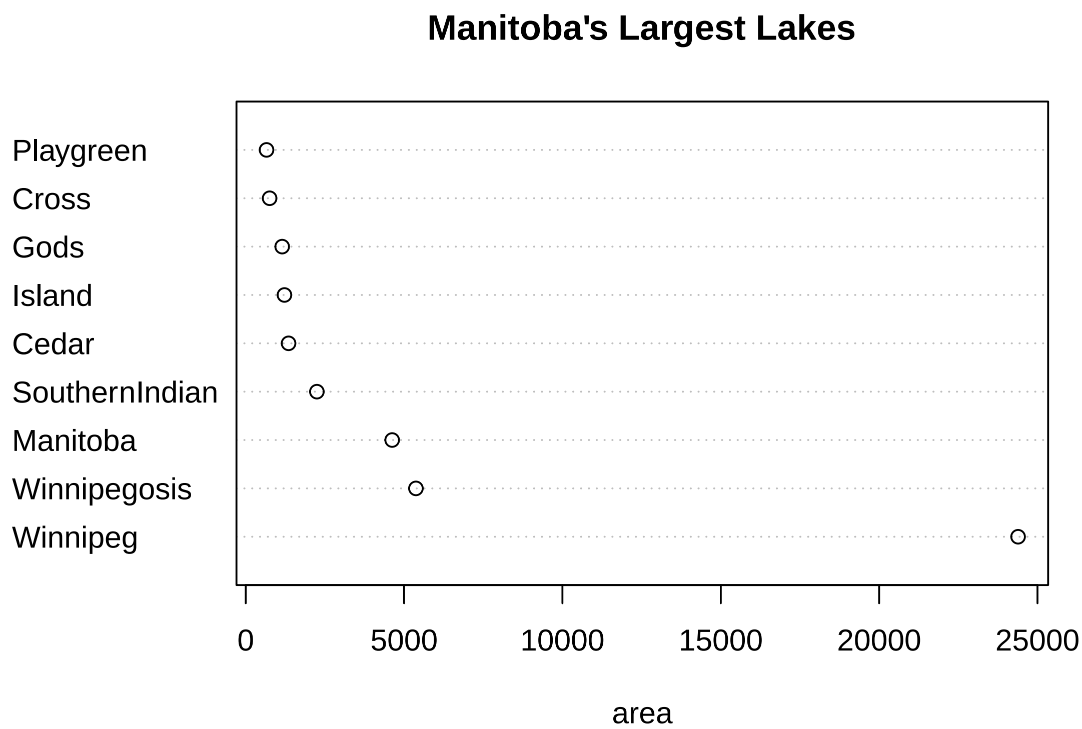
b.

```
attach(Manitoba.lakes)
plot(area~elevation, pch=16, xlim=c(170,280), ylog=T)
text(area~elevation, labels=row.names(Manitoba.lakes), pos=4, ylog=T)
text(area~elevation, labels=area, pos=2, ylog=T)
title("Manitoba's Largest Lakes")
```

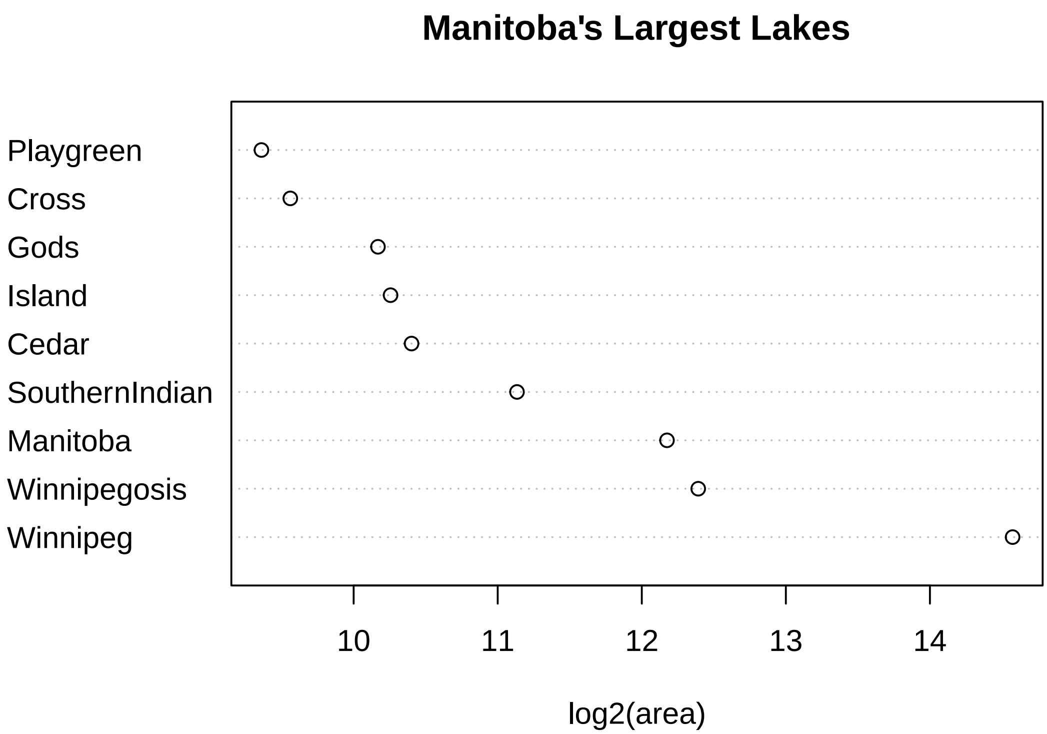


7. code for assignment MB.Ch1.7:

```
dotchart(area, labels=row.names(Manitoba.lakes),
         main = "Manitoba's Largest Lakes",
         xlab="area")
```




```
dotchart(log2(area), labels=row.names(Manitoba.lakes),
         main = "Manitoba's Largest Lakes",
         xlab="log2(area)")
```



8. code for assignment MB.Ch1.8:

```
sum(area)
```

```
## [1] 41771
```