Homework-01

DingHuan,3170102085

1. a. Noticing that the data in iowa.csv is seperated by ";", I choose to add an argument `sep = ";"` to the funtion `read.csv()`.

```
iowa.df <- read.csv("data/iowa.csv", header = T, sep = ";")
head(iowa.df)
```

```
##   Year Rain0 Temp1 Rain1 Temp2 Rain2 Temp3 Rain3 Temp4 Yield
## 1 1930 17.75  60.2  5.83  69.0  1.49  77.9  2.42  74.4  34.0
## 2 1931 14.76  57.5  3.83  75.0  2.72  77.2  3.30  72.6  32.9
## 3 1932 27.99  62.3  5.17  72.0  3.12  75.8  7.10  72.2  43.0
## 4 1933 16.76  60.5  1.64  77.8  3.45  76.4  3.01  70.5  40.0
## 5 1934 11.36  69.5  3.49  77.2  3.85  79.7  2.84  73.4  23.0
## 6 1935 22.71  55.0  7.00  65.9  3.35  79.4  2.42  73.6  38.4
```

   b. There are 33 rows and 10 columns.

```
dim(iowa.df)
```

```
## [1] 33 10
```

   c. The names of the columns are shown below.

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

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

   d. The value of row 5, column 7 of `iowa.df` is 79.7.

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

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

   e. The second row of `iowa.df` is shown below.

```
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. a. The first line creates a vector of characters(or strings) named `vactor1`.

```
vector1 <- c("5", "12", "7", "32")
```

The second line calculate the maximum of the strings in `vector1` according to the lexicographical order. Given that `"1" < :3" < "5" < "7"`, we get the answer of `max(vector1)` `"7"`.

```
max(vector1)
```

```
## [1] "7"
```

The third line arranges strings in `vector1` from small to large. Reconsidering the analysis above, we get the answer below.

```r
sort(vector1)
```

```
## [1] "12" "32" "5"  "7"
```

The fourth line gets an error, because `sum()` can only calculate the numeric arguments, rather than strings.

```r
sum(vector1)
```

```
## Error in sum(vector1): invalid 'type' (character) of argument
```

b. The first series of commands produce an error, because the first line create an character variable named `vector2`, whose components are characters `"5","7","12"` as a result of the inability of a vector variable to save multiple types of data. However, the addition operation of characters `+` is not defined in R.

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

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

The second series of commands returns 19, because a dataframe variable can save different types of data, so the addition of numeric data 7 and 12 is equal to `19`.

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

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

The third series of commands also produce an error at the third line. The first line create a list variable, whose 2nd and 4th components are numerical. The second line add the 2th component to the 4th one as values, returning an answer 168. However, the syntax `list4[2]` returns `z2 = 42` as a list, thun an error occurs because the addition operation of lists `+` is not defined in R.

```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. a. Use arguments `by` and `length.out` in `seq()` to achieve the target.

```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. `rep(1:3, times = 3)` repeat the total vector `1:3` for 3 times, returning an vector `1 2 3 1 2 3 1 2 3`. `rep(1:3, each = 3)` repeat each components in vector `1:3` for 3 times separately, returning an vector `1 1 1 2 2 2 3 3 3`.

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

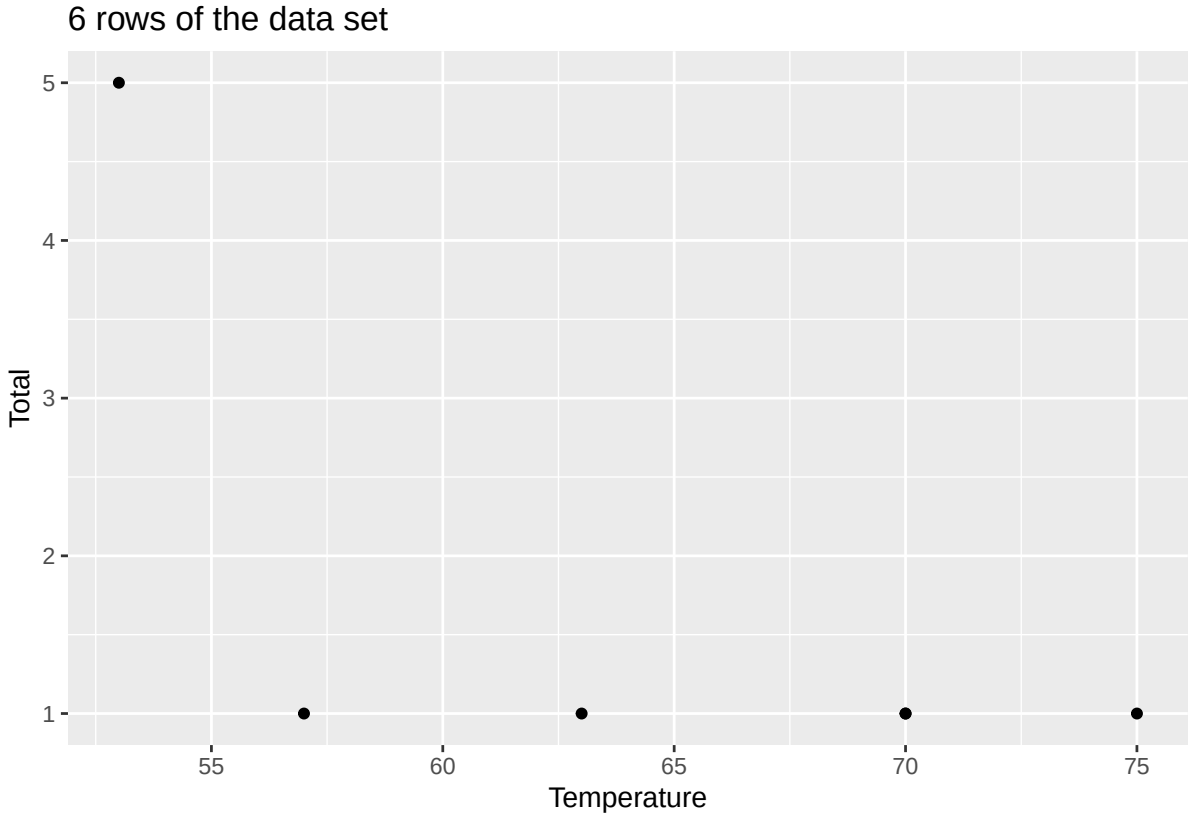
```
## [1] 1 2 3 1 2 3 1 2 3
```

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

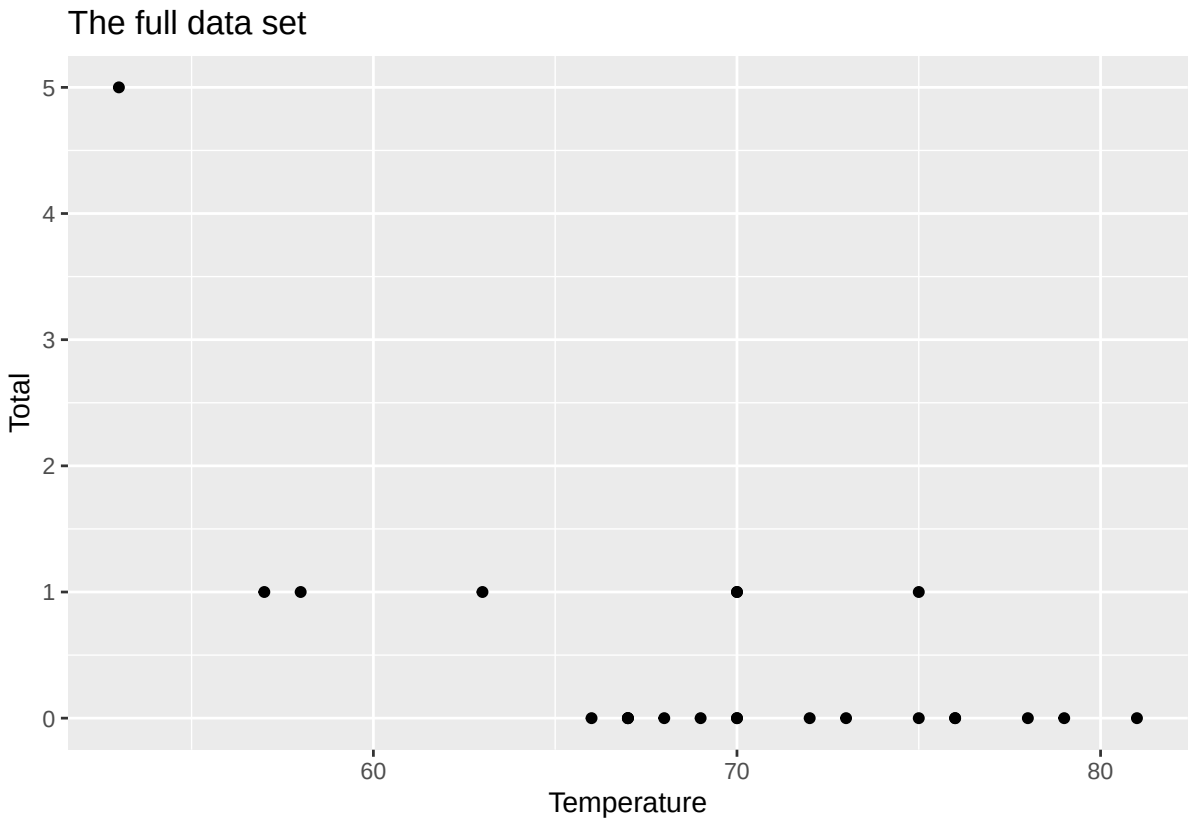
```
## [1] 1 1 1 2 2 2 3 3 3
```

4. (MB.CH1.2)

```
library(DAAG)
library(tidyverse)
data(orings)
pre.launch <- orings[c(1,2,4,11,13,18),]
pre.launch %>% ggplot(aes(x = Temperature, y = Total)) +
  geom_point() +
  labs(title = "6 rows of the data set")
```




```
orings %>% ggplot(aes(x = Temperature, y = Total)) +
  geom_point() +
  labs(title = "The full data set")
```



5. (MB.CH1.4) a. There's no column holds missing values.

```
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 ...
```

```
which(!complete.cases(ais))
```

```
## integer(0)
```

b. There is a large imbalance in the numbers of the two sexes in gym, netball, T_sport and W_polo.

```
Sports <- levels(ais$sport)
ST = matrix(nrow = 2, ncol = length(Sports),
            dimnames = list(c('male', 'female'),Sports))
for (i in 1:length(Sports)){
```

```
  ST[1,i] <- sum((ais$sex == 'm') & (ais$sport == Sports[i]))
  ST[2,i] <- sum((ais$sex == 'f') & (ais$sport == Sports[i]))
}
print(ST)
```

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

```
balance = ST['male',] / ST['female',]
names(which(balance < 0.5 | balance > 2))
```

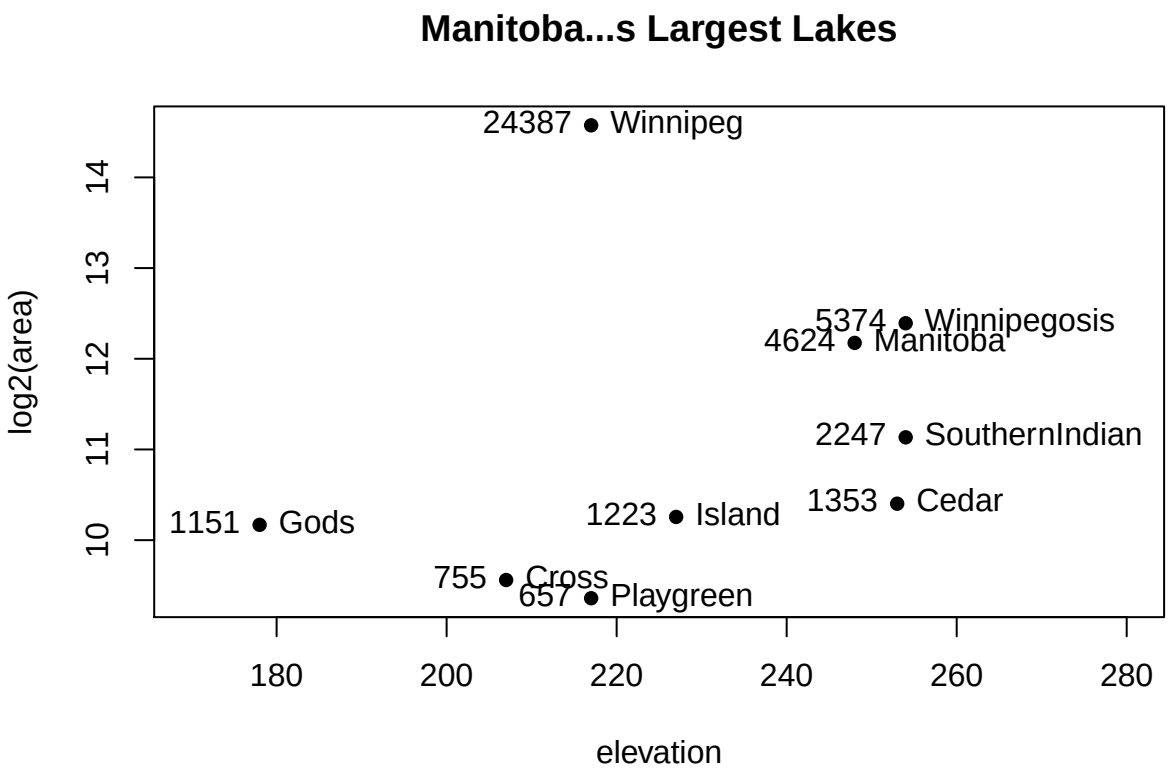
```
## [1] "Gym"     "Netball" "T_Sprnt" "W_Polo"
```

6. (MB.CH1.6)

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

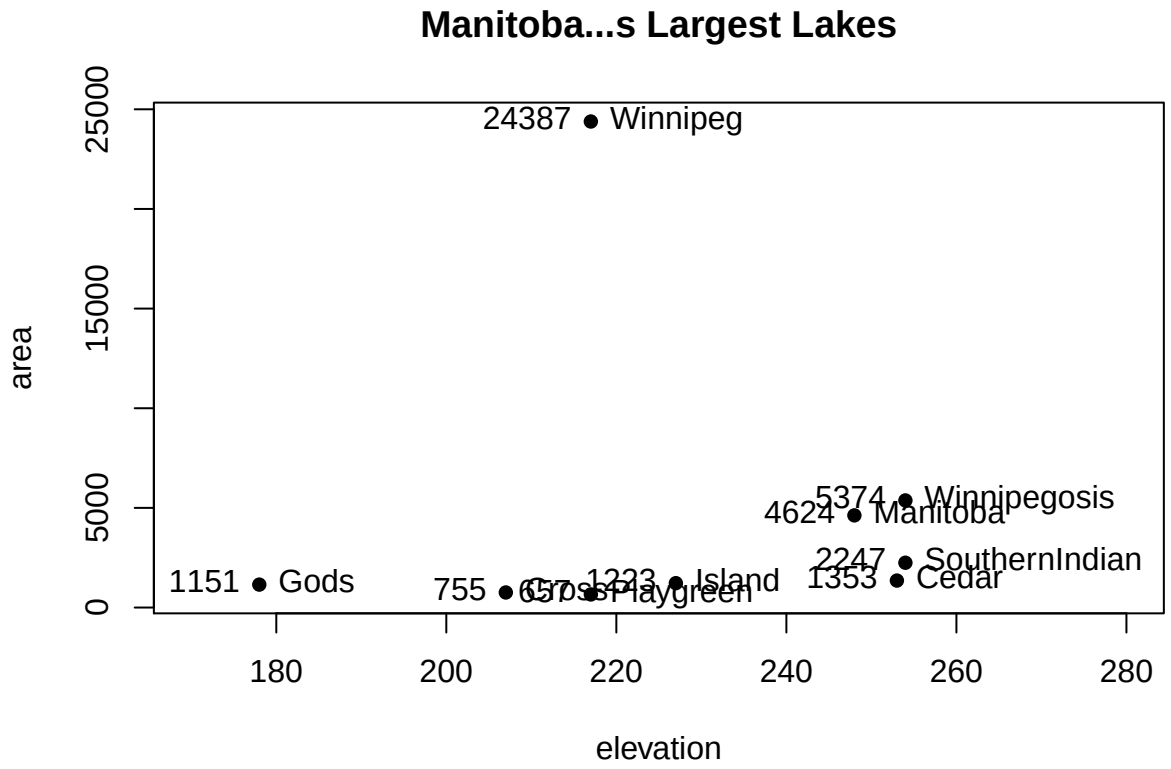
a.

```
attach(Manitoba.lakes)
plot(log2(area) ~ elevation, pch=16, xlim=c(170,280))
text(log2(area) ~ elevation, labels=row.names(Manitoba.lakes), pos=4)
text(log2(area) ~ elevation, labels=area, pos=2)
title("Manitoba's Largest Lakes")
```



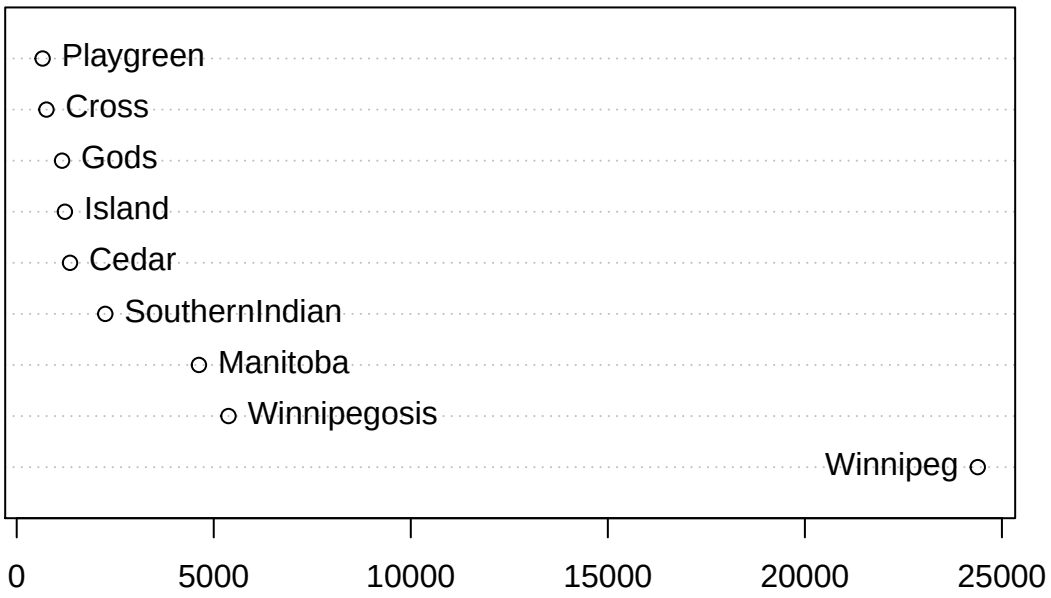
b.

```
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. (MB.CH1.7)

```
dotchart(Manitoba.lakes$area, xlab = "area")
text(Manitoba.lakes$area[-1],2:9,labels=row.names(Manitoba.lakes[-1,]), pos = 4)
text(Manitoba.lakes$area[1],1,labels=row.names(Manitoba.lakes[1,]), pos = 2)
title("Manitoba's Largest Lakes")
```




```
dotchart(Manitoba.lakes$area, xlab = "area (on a logarithmic scale)", log = "x")
text(Manitoba.lakes$area[-1],2:9,labels=row.names(Manitoba.lakes[-1,]), pos = 4)
text(Manitoba.lakes$area[1],1,labels=row.names(Manitoba.lakes[1,]), pos = 2)
title("Manitoba's Largest Lakes")
```