短学期作业一

汪利军 *3140105707*

July 4, 2017

Contents

1 作业

1.1 MB

Chapter 1: 2, 4, 6, 7, 8, 12, 18;

1.2 MDL

Chapter 4 Exercise: 4.13, 4.14, 4.15, Chapter 3 Worksheet.

2 MB

2.1 Problem 1.2

orings 数据集在 DAAG 包中

```
library(DAAG)
```

```
## Loading required package: lattice
```

```
orings
```

```
##   Temperature Erosion Blowby Total
## 1          53       3      2     5
## 2          57       1      0     1
## 3          58       1      0     1
## 4          63       1      0     1
## 5          66       0      0     0
## 6          67       0      0     0
## 7          67       0      0     0
```

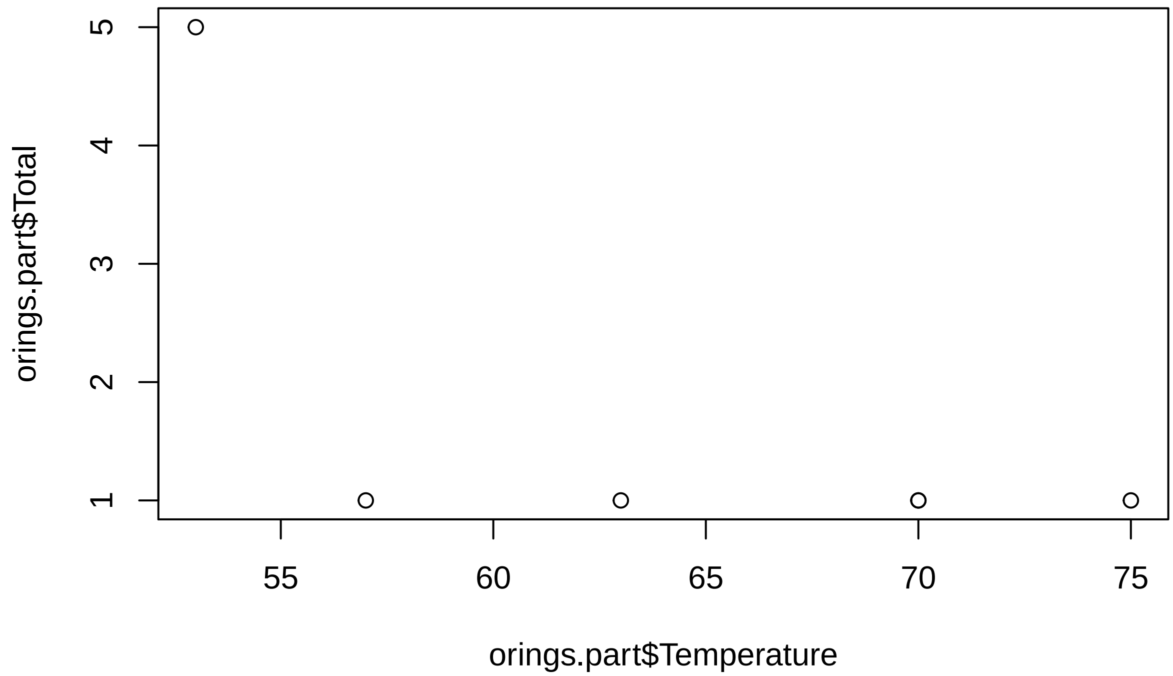
```
## 8           67     0     0     0
## 9           68     0     0     0
## 10          69     0     0     0
## 11          70     1     0     1
## 12          70     0     0     0
## 13          70     1     0     1
## 14          70     0     0     0
## 15          72     0     0     0
## 16          73     0     0     0
## 17          75     0     0     0
## 18          75     0     2     1
## 19          76     0     0     0
## 20          76     0     0     0
## 21          78     0     0     0
## 22          79     0     0     0
## 23          81     0     0     0
```

Extract row 1, 2, 4, 11, 13 and 18 from orings

```
orings.part = orings[c(1,2,4,11,13,18),]
```

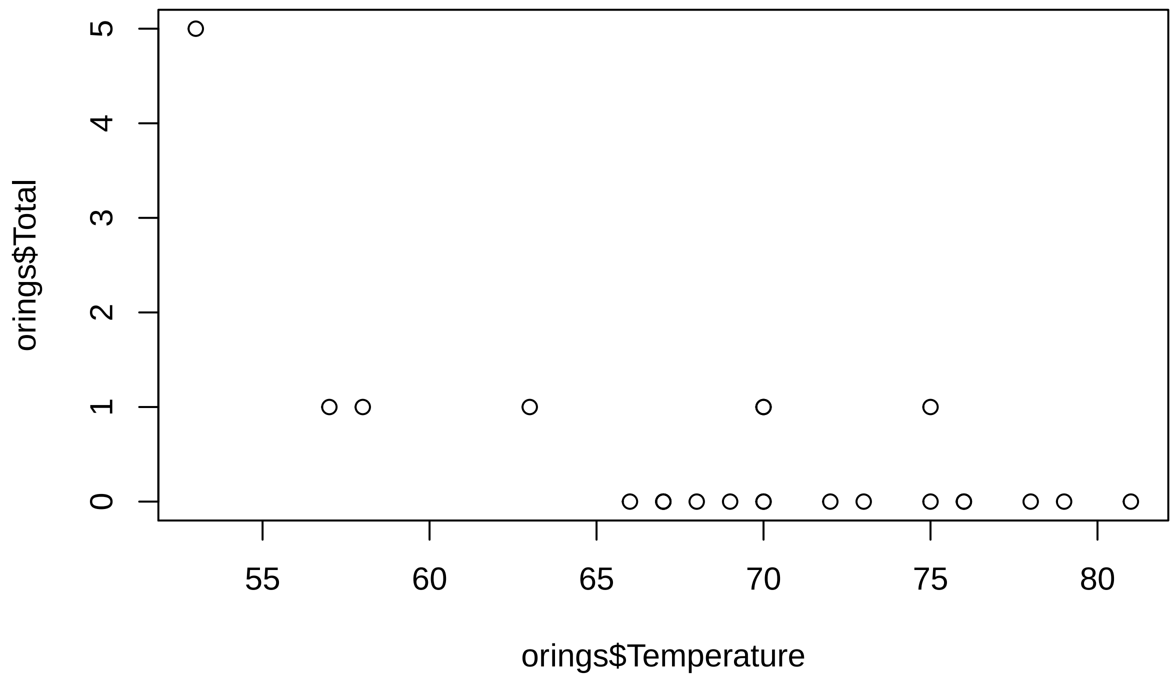
Plot total incidents against temperature for the new data frame

```
plot(orings.part$Temperature, orings.part$Total)
```



The similar plot for the full data set.

```
plot(orings$Temperature, orings$Total)
```



2.2 Problem 1.4

```r
library(DAAG)
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 ...
```

判断缺失值

```r
which(is.na(ais))
```

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

也就是没有缺失值，所以每一列都无缺失值。The table that shows the numbers of males and females for each different sport is as follows.

```r
table(ais$sex, ais$sport)
```

```
##
```

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

从上表可以看出，Gym，Netball，T_Sprnt 和 W_Polo 存在 large imbalance，前两项运动是女性居多，后两项运动是男性居多。

2.3 Problem 1.6

构造数据集

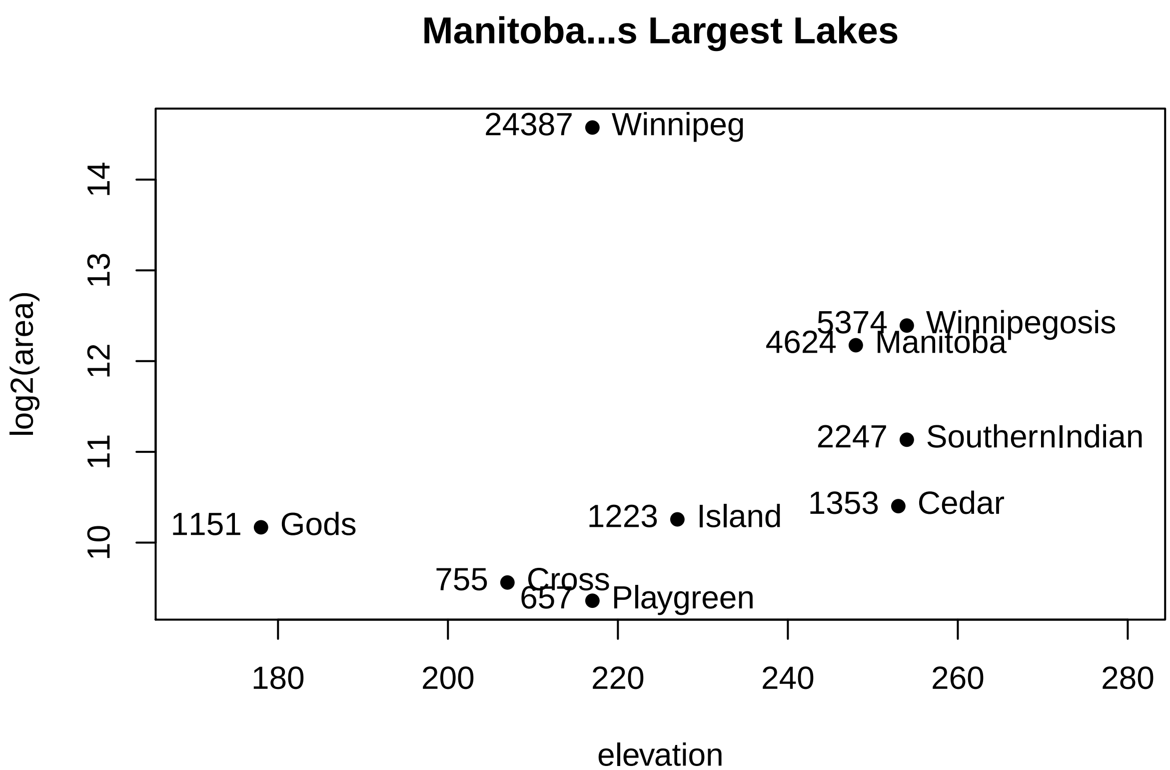
```
elevation = c(217, 254, 248, 254, 253, 227, 178, 207, 217)
area = c(24387, 5374, 4624, 2247, 1353, 1223, 1151, 755, 657)
my.Manitoba.lakes = data.frame(elevation, area)
row.names(my.Manitoba.lakes) = c("Winnipeg", "Winnipegosis",
                                 "Manitoba", "SouthernIndian",
   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

2.3.1 (a)

```
attach(my.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=rownames(my.Manitoba.lakes), pos=4)
```

```r
text(log2(area) ~ elevation, labels=area, pos=2)
title("Manitoba’ s Largest Lakes")
```

```
## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <e2>
## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <80>
## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <99>
```




```r
detach(my.Manitoba.lakes)
```

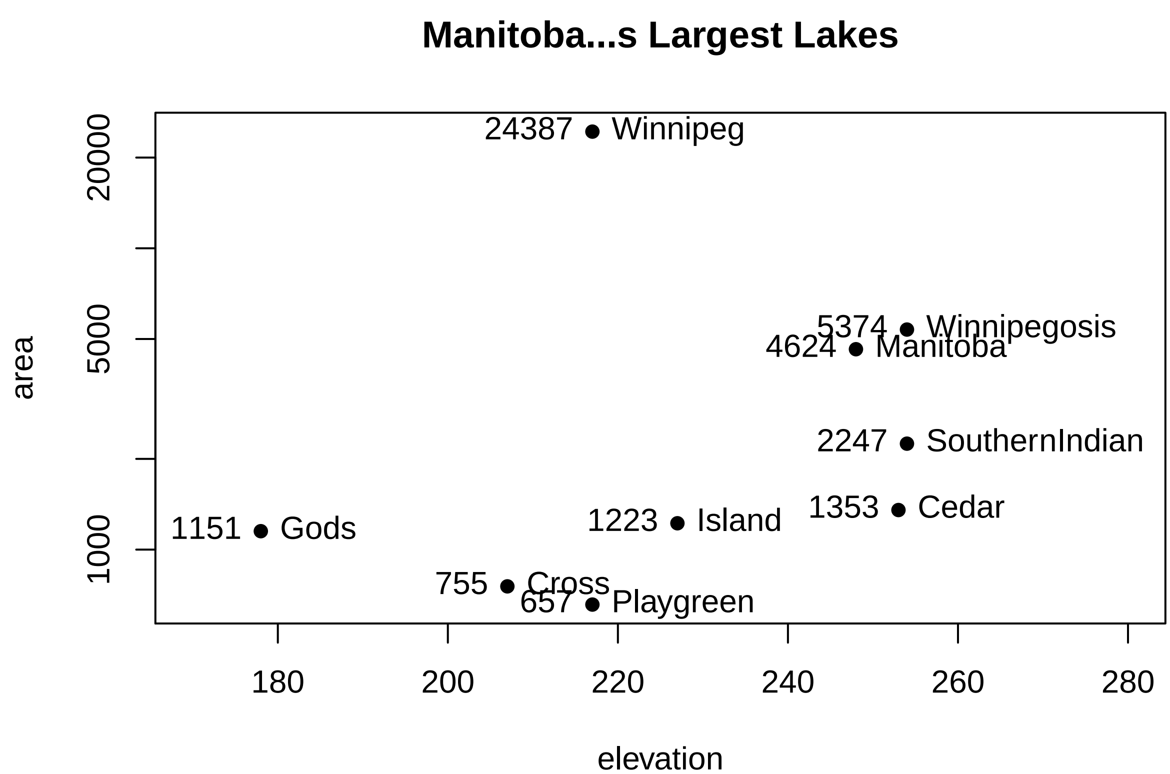
2.3.2 (b)

```
attach(my.Manitoba.lakes)
plot(area ~ elevation, log = "y", pch = 16, xlim = c(170, 280))
text(area ~ elevation,
     labels=rownames(Manitoba.lakes), pos=4)
text(area ~ elevation, labels=area, pos=2)
title("Manitoba’ s Largest Lakes")
```

```
## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <80>

## Warning in title("Manitoba’ s Largest Lakes"): conversion failure on
## 'Manitoba’ s Largest Lakes' in 'mbcsToSbcs': dot substituted for <99>
```



```r
detach(my.Manitoba.lakes)
```

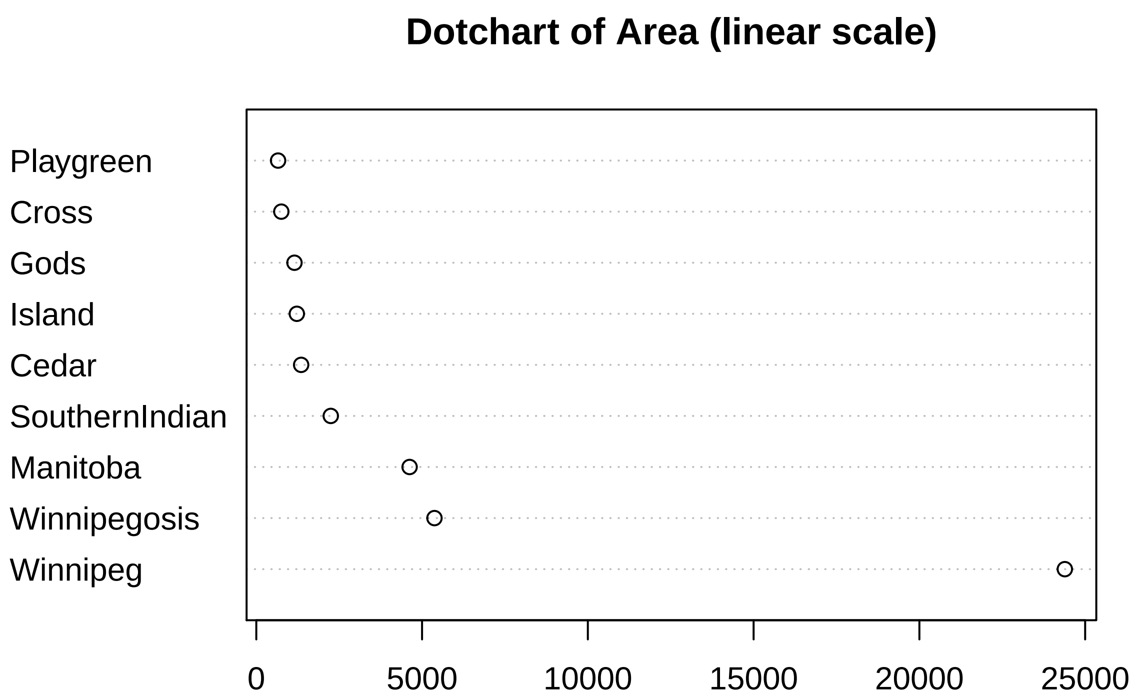
2.4 Problem 1.7

查看 dotchart 的帮助文档

```r
?dotchart
```

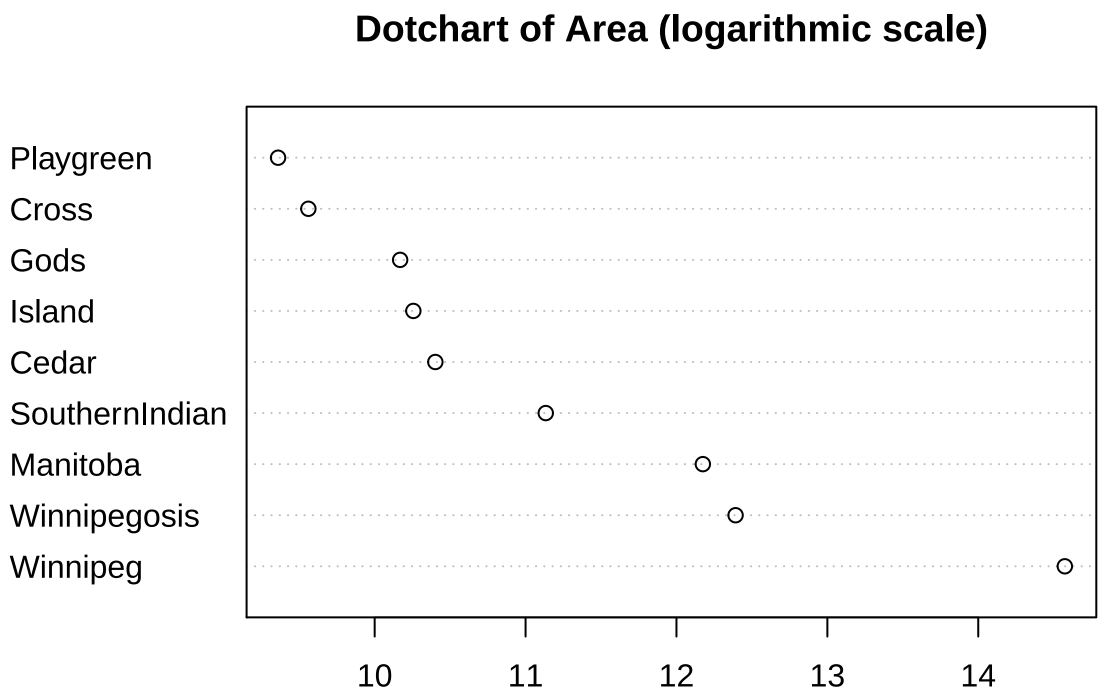
2.4.1 (a)

```r
dotchart(Manitoba.lakes$area,
         main = "Dotchart of Area (linear scale)",
         labels = rownames(Manitoba.lakes))
```



2.4.2 (b)

```
dotchart(log2(Manitoba.lakes$area),
         main = "Dotchart of Area (logarithmic scale)",
         labels = rownames(Manitoba.lakes))
```



2.5 Problem 1.8

The lower bound for the area of Manitoba covered by water is

```
sum(Manitoba.lakes$area)
```

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

2.6 Problem 1.12

```
cutoff.prop <- function(x, cutoff)
{
  ## coerce list to vector
  if (is.list(x))
    x = as.numeric(unlist(x))
  n = length(x)
  ne = sum(x>cutoff)
  return(ne/n)
}
```

2.6.1 (a)

```
x = c(1:100)
## 超过 90 的比例
cutoff.prop(x, 90)
```

```
## [1] 0.1
```

```
## 超过 80 的比例
cutoff.prop(x, 80)
```

```
## [1] 0.2
```

2.6.2 (b)

```
library(Devore7)
```

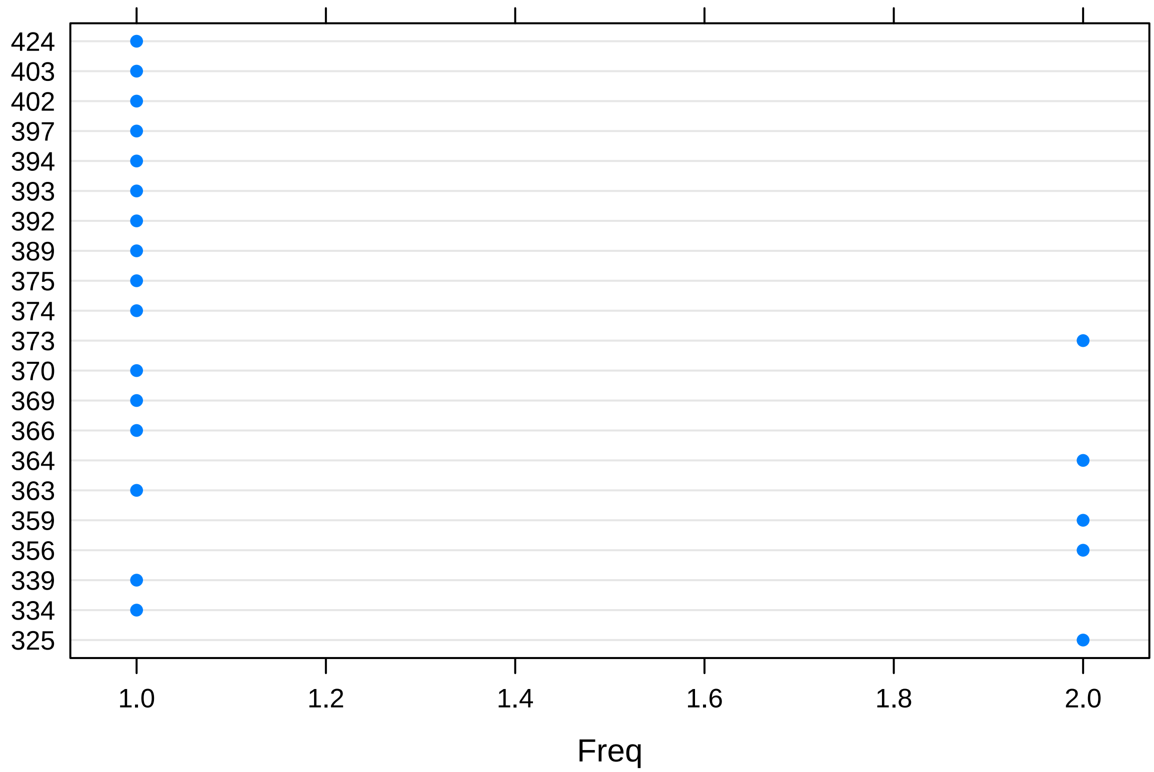
```
## Loading required package: MASS

##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:DAAG':
##
##     hills
```

```r
dotplot(ex01.36)
```



数据集中未给出数据的单位，猜测为秒，则超过 7 分钟的比例为

```r
cutoff.prop(ex01.36, 60*7)
```

```
## [1] 0.03846154
```

2.7 Problem 1.18

```r
library(MASS)

## 原格式
```

```
Rabbit
```

```
##    BPchange   Dose Run Treatment Animal
## 1      0.50   6.25  C1   Control     R1
## 2      4.50  12.50  C1   Control     R1
## 3     10.00  25.00  C1   Control     R1
## 4     26.00  50.00  C1   Control     R1
## 5     37.00 100.00  C1   Control     R1
## 6     32.00 200.00  C1   Control     R1
## 7      1.00   6.25  C2   Control     R2
## 8      1.25  12.50  C2   Control     R2
## 9      4.00  25.00  C2   Control     R2
## 10    12.00  50.00  C2   Control     R2
## 11    27.00 100.00  C2   Control     R2
## 12    29.00 200.00  C2   Control     R2
## 13     0.75   6.25  C3   Control     R3
## 14     3.00  12.50  C3   Control     R3
## 15     3.00  25.00  C3   Control     R3
## 16    14.00  50.00  C3   Control     R3
## 17    22.00 100.00  C3   Control     R3
## 18    24.00 200.00  C3   Control     R3
## 19     1.25   6.25  C4   Control     R4
## 20     1.50  12.50  C4   Control     R4
## 21     6.00  25.00  C4   Control     R4
## 22    19.00  50.00  C4   Control     R4
## 23    33.00 100.00  C4   Control     R4
## 24    33.00 200.00  C4   Control     R4
```

```
## 25     1.50   6.25  C5   Control    R5
## 26     1.50  12.50  C5   Control    R5
## 27     5.00  25.00  C5   Control    R5
## 28    16.00  50.00  C5   Control    R5
## 29    20.00 100.00  C5   Control    R5
## 30    18.00 200.00  C5   Control    R5
## 31     1.25   6.25  M1       MDL    R1
## 32     0.75  12.50  M1       MDL    R1
## 33     4.00  25.00  M1       MDL    R1
## 34     9.00  50.00  M1       MDL    R1
## 35    25.00 100.00  M1       MDL    R1
## 36    37.00 200.00  M1       MDL    R1
## 37     1.40   6.25  M2       MDL    R2
## 38     1.70  12.50  M2       MDL    R2
## 39     1.00  25.00  M2       MDL    R2
## 40     2.00  50.00  M2       MDL    R2
## 41    15.00 100.00  M2       MDL    R2
## 42    28.00 200.00  M2       MDL    R2
## 43     0.75   6.25  M3       MDL    R3
## 44     2.30  12.50  M3       MDL    R3
## 45     3.00  25.00  M3       MDL    R3
## 46     5.00  50.00  M3       MDL    R3
## 47    26.00 100.00  M3       MDL    R3
## 48    25.00 200.00  M3       MDL    R3
## 49     2.60   6.25  M4       MDL    R4
## 50     1.20  12.50  M4       MDL    R4
## 51     2.00  25.00  M4       MDL    R4
```

```
## 52     3.00  50.00  M4        MDL     R4
## 53    11.00 100.00  M4        MDL     R4
## 54    22.00 200.00  M4        MDL     R4
## 55     2.40   6.25  M5        MDL     R5
## 56     2.50  12.50  M5        MDL     R5
## 57     1.50  25.00  M5        MDL     R5
## 58     2.00  50.00  M5        MDL     R5
## 59     9.00 100.00  M5        MDL     R5
## 60    19.00 200.00  M5        MDL     R5
```

```r
TA = unstack(Rabbit, Treatment~Animal)
TA
```

```
##         R1      R2      R3      R4      R5
## 1  Control Control Control Control Control
## 2  Control Control Control Control Control
## 3  Control Control Control Control Control
## 4  Control Control Control Control Control
## 5  Control Control Control Control Control
## 6  Control Control Control Control Control
## 7      MDL     MDL     MDL     MDL     MDL
## 8      MDL     MDL     MDL     MDL     MDL
## 9      MDL     MDL     MDL     MDL     MDL
## 10     MDL     MDL     MDL     MDL     MDL
## 11     MDL     MDL     MDL     MDL     MDL
## 12     MDL     MDL     MDL     MDL     MDL
```

```r
BA = unstack(Rabbit, BPchange~Animal)
BA
```

```
##       R1    R2    R3    R4   R5
## 1   0.50  1.00  0.75  1.25  1.5
## 2   4.50  1.25  3.00  1.50  1.5
## 3  10.00  4.00  3.00  6.00  5.0
## 4  26.00 12.00 14.00 19.00 16.0
## 5  37.00 27.00 22.00 33.00 20.0
## 6  32.00 29.00 24.00 33.00 18.0
## 7   1.25  1.40  0.75  2.60  2.4
## 8   0.75  1.70  2.30  1.20  2.5
## 9   4.00  1.00  3.00  2.00  1.5
## 10  9.00  2.00  5.00  3.00  2.0
## 11 25.00 15.00 26.00 11.00  9.0
## 12 37.00 28.00 25.00 22.00 19.0
```

```r
DA = unstack(Rabbit, Dose~Animal)
DA
```

```
##        R1     R2     R3     R4     R5
## 1    6.25   6.25   6.25   6.25   6.25
## 2   12.50  12.50  12.50  12.50  12.50
## 3   25.00  25.00  25.00  25.00  25.00
## 4   50.00  50.00  50.00  50.00  50.00
## 5  100.00 100.00 100.00 100.00 100.00
## 6  200.00 200.00 200.00 200.00 200.00
## 7    6.25   6.25   6.25   6.25   6.25
## 8   12.50  12.50  12.50  12.50  12.50
## 9   25.00  25.00  25.00  25.00  25.00
## 10  50.00  50.00  50.00  50.00  50.00
```

```
## 11 100.00 100.00 100.00 100.00 100.00
## 12 200.00 200.00 200.00 200.00 200.00
```

因 TA 和 DA 对于五种 Animal 都相同，则取一列即可

```
TA1 = TA[1]
DA1 = DA[1]
names(TA1) = "Treatment"
names(DA1) = "Dose"
```

然后进行 cbind 即可

```
cbind(TA1, DA1, BA)
```

```
##    Treatment   Dose    R1    R2    R3    R4   R5
## 1    Control   6.25  0.50  1.00  0.75  1.25  1.5
## 2    Control  12.50  4.50  1.25  3.00  1.50  1.5
## 3    Control  25.00 10.00  4.00  3.00  6.00  5.0
## 4    Control  50.00 26.00 12.00 14.00 19.00 16.0
## 5    Control 100.00 37.00 27.00 22.00 33.00 20.0
## 6    Control 200.00 32.00 29.00 24.00 33.00 18.0
## 7        MDL   6.25  1.25  1.40  0.75  2.60  2.4
## 8        MDL  12.50  0.75  1.70  2.30  1.20  2.5
## 9        MDL  25.00  4.00  1.00  3.00  2.00  1.5
## 10       MDL  50.00  9.00  2.00  5.00  3.00  2.0
## 11       MDL 100.00 25.00 15.00 26.00 11.00  9.0
## 12       MDL 200.00 37.00 28.00 25.00 22.00 19.0
```

3 MDL

3.1 Problem 4.13

```r
seq(1.0, 2.0, by = 0.1)
```

```
##  [1] 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
```

3.2 Problem 4.14

```r
rep(1:3, each = 2)
```

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

3.3 Problem 4.15

```r
rep(1:3,2)
```

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

3.4 Chapter 3 Worksheet

3.4.1 构造数据

```r
Gender <- factor(c("G", "G", "B", "G", "B", "B", "B", "B", "B", "B"))
ZEP <- c(T, T, T, T, F, T, F, T, T, T)
Weight <- c(16, 14, 13.5, 15.4, 16.5, 16, 17, 14.8, 17, 16.7)
Years <- c(3, 3, 3, 4, 3, 4, 3, 3, 4, 3)
Months <- c(5, 10, 5, 0, 8, 0, 11, 9, 1, 3)
```

```r
Height <- c(100.0, 97.0, 95.5, 101.0, 100.0, 98.5, 103.0, 98.0, 101.5, 100.0)
df <- data.frame(Gender, ZEP, Weight, Years, Months, Height)
```

3.4.2 计算均值

```r
colMeans(df[c(3,4,5)])
```

```
## Weight  Years Months
##  15.69   3.30   5.20
```

3.4.3 计算 BMI

```r
BMI = Weight/(Height*0.01)^2
BMI
```

```
##  [1] 16.00000 14.87937 14.80223 15.09656 16.50000 16.49102 16.02413
##  [8] 15.41025 16.50125 16.70000
```