Assignment 4

Saket Choudhary http://rpubs.com/saketkc/164963

March 20, 2016

Problem 1

```
x1 <- rexp(n=100)
m.x1 <- mean(x1)
m.x1
```

```
## [1] 0.8577574
```

```
sd.x1 <- sd(x1)
sd.x1
```

```
## [1] 0.8133896
```

Mean: 0.8577574

Standard Deviation: 0.8133896

Problem 2

```
x0.1 <- rexp(n=100, rate=0.1)
x0.5 <- rexp(n=100, rate=0.5)
x5 <- rexp(n=100, rate=5)
x10 <- rexp(n=100, rate=10)
```

`x0.1`: Mean = 9.1864982 SD = 8.6377849

`x0.5`: Mean = 2.3717696 SD = 2.0429888

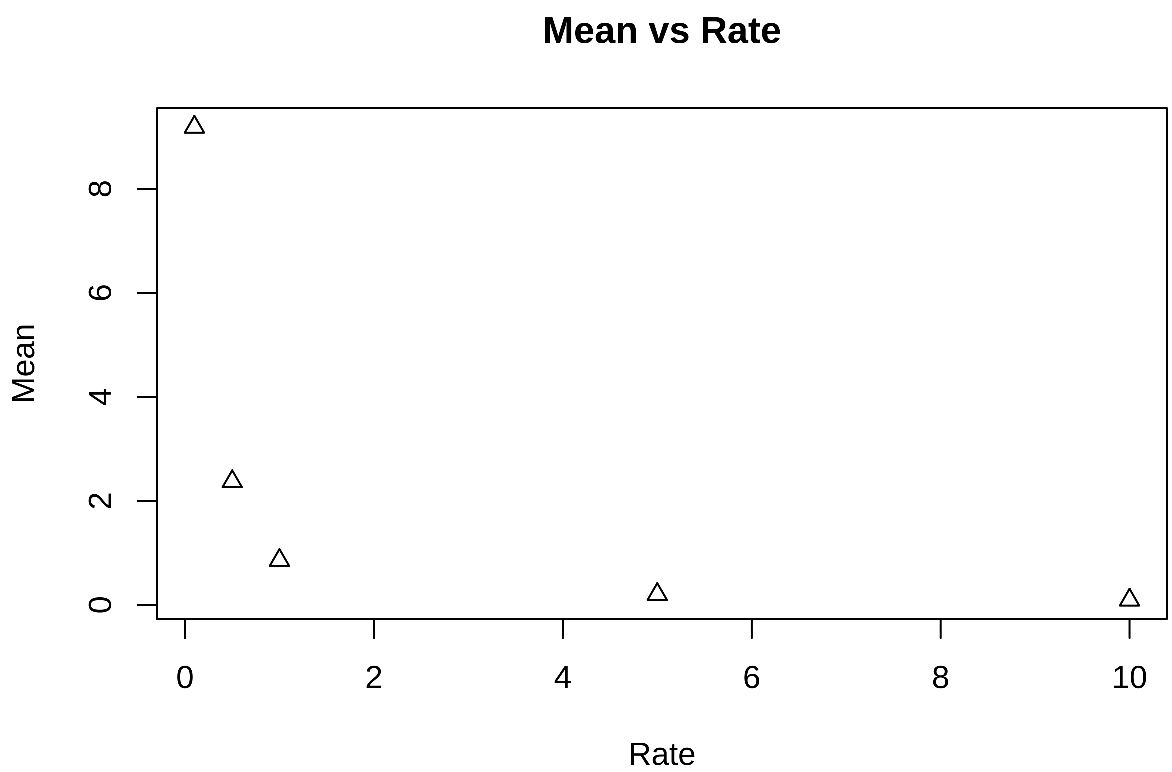
`x5`: Mean = 0.2024976 SD = 0.2166985

`x10`: Mean = 0.0942307 SD = 0.1002465

Problem 3

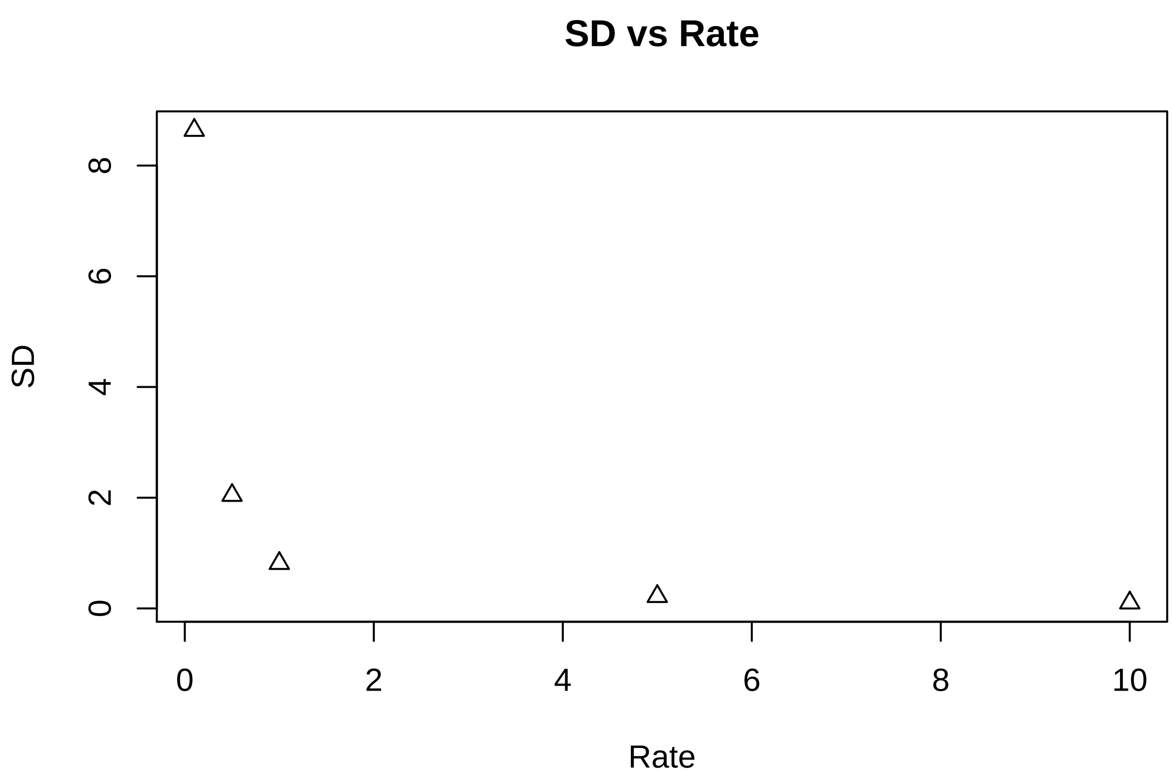
Problem 3.a

```
plot(c(0.1,0.5,1,5,10), c(mean(x0.1), mean(x0.5), mean(x1), mean(x5), mean(x10)), pch=2, xlab = 'Rate',
```



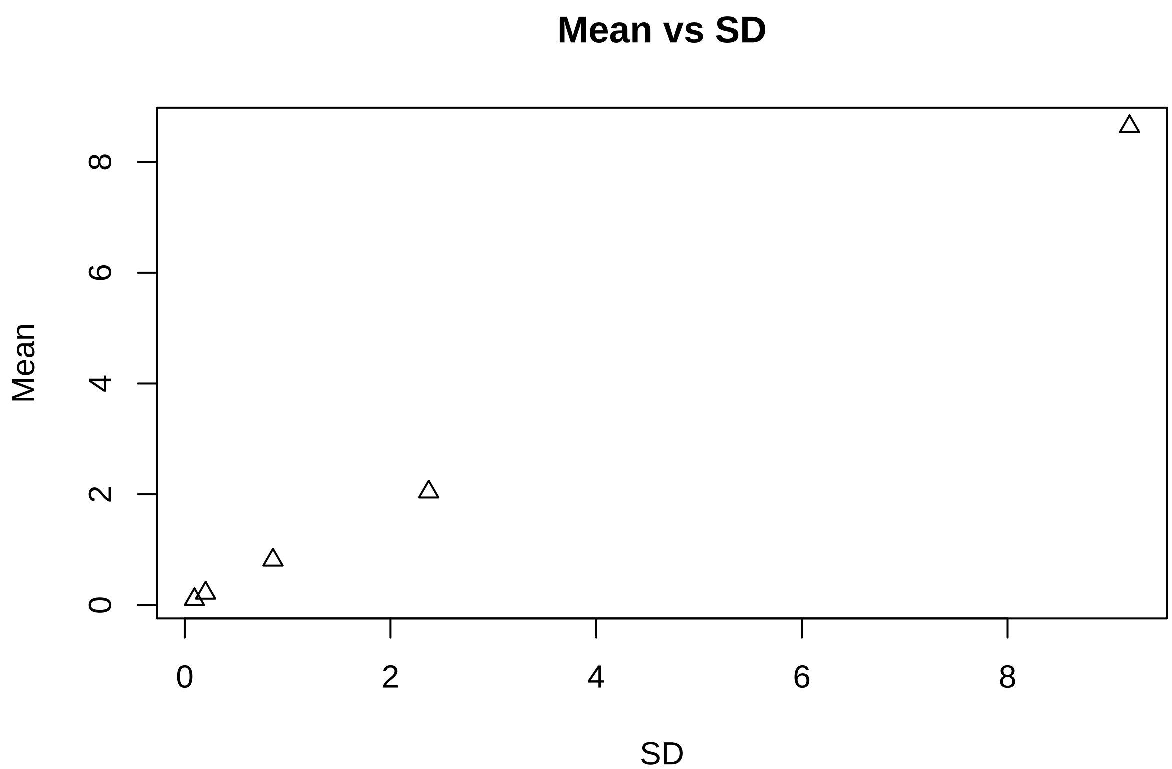
Problem 3.b

```
plot(c(0.1,0.5,1,5,10), c(sd(x0.1), sd(x0.5), sd(x1), sd(x5), sd(x10)), pch=2, xlab = 'Rate', ylab='SD'
```



Problem 3.c

```
plot(c(mean(x0.1), mean(x0.5), mean(x1), mean(x5), mean(x10)), c(sd(x0.1), sd(x0.5), sd(x1), sd(x5), sd
```



$E[X] = \frac{1}{\lambda}$ and $Var[X] = \frac{1}{\lambda^2}$ hence $E[X] vs \sqrt{Var[X]}$ follows a linear trend

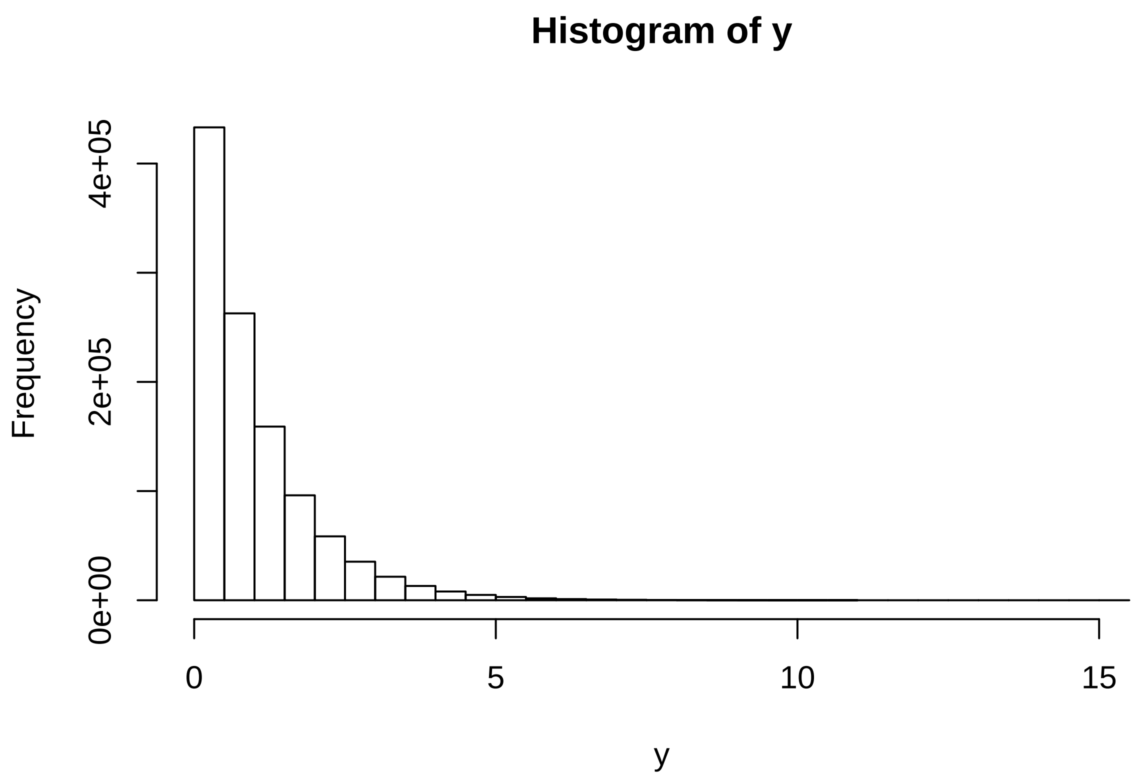
Problem 4

```
y <- rexp(n=1100000)
```

Mean of y: 0.9996713 SD of y: 1.0018827

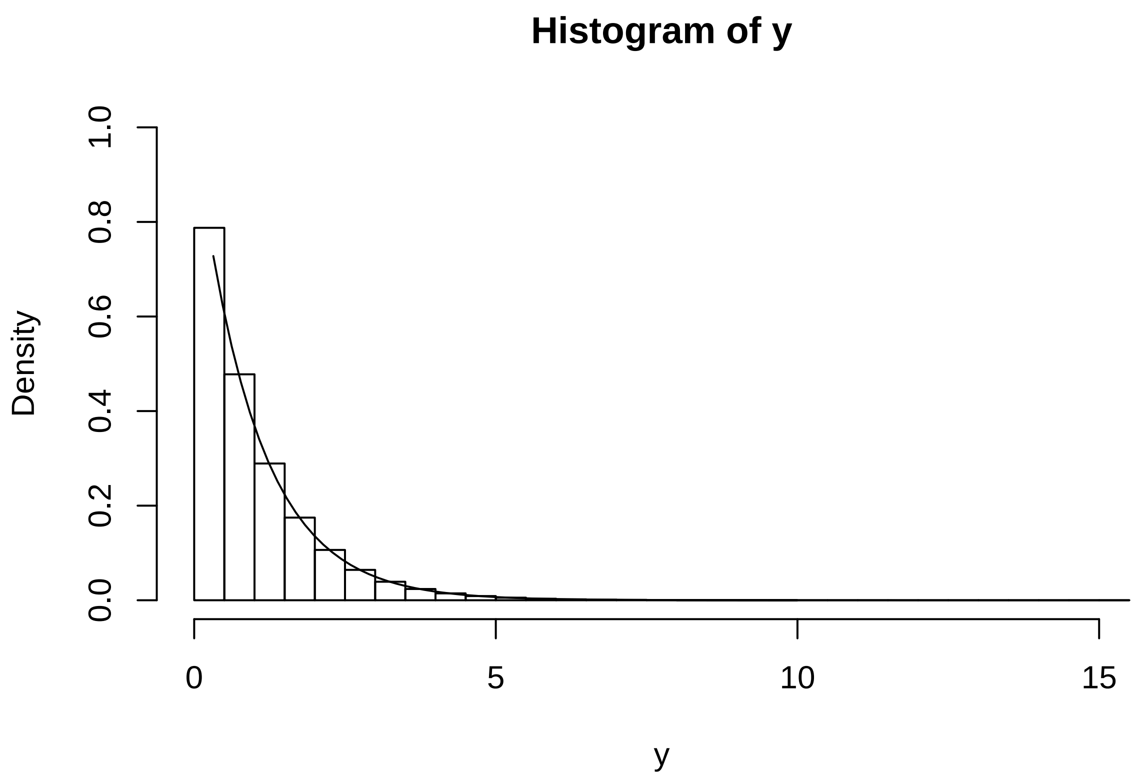
Problem 5

```
hist(y)
```

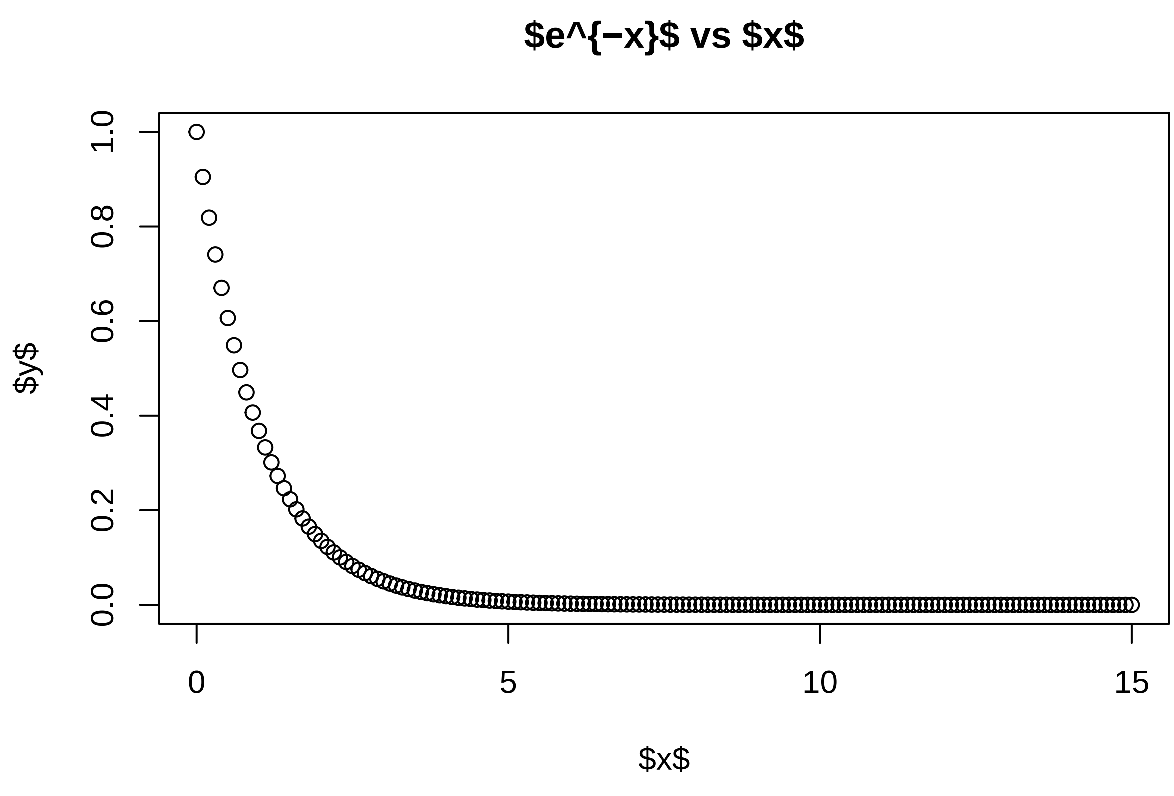


y is the PDF of exponential distribution and hence matches the following distribution e^{-x}

```
h <- hist(y, plot = FALSE)
ylim <- range(0, h$density, dexp(0))
hist(y, freq = FALSE, ylim = ylim)
curve(dexp, y, add=TRUE)
```




```
plot(seq(0,15,0.1), exp(-seq(0,15,0.1)), main='$e^{-x}$ vs $x$', xlab = '$x$', ylab='$y$')
```



Problem 6

```
y.mat  <- matrix(y, nrow=1100,ncol=1000)
```

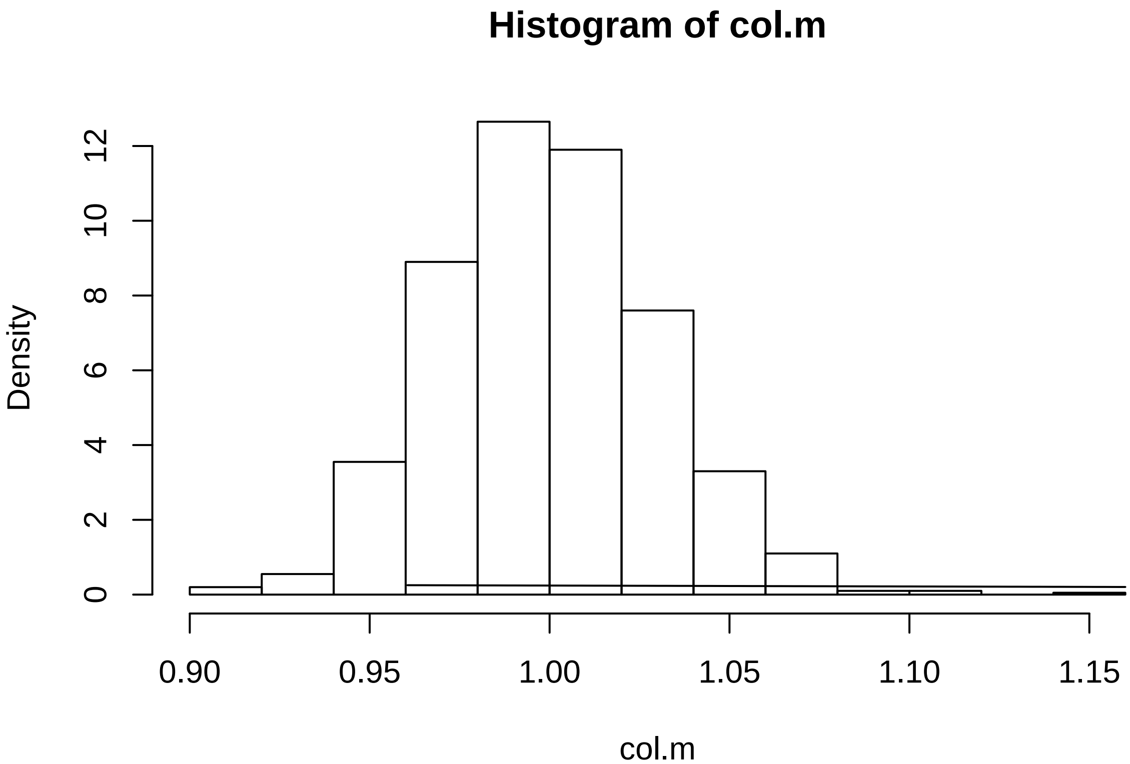
Problem 7

```
mean(y.mat[,371])
```

```
## [1] 1.060127
```

Problem 8

```
col.m <- colMeans(y.mat)
h <- hist(col.m, plot = FALSE)
ylim <- range(0, h$density, dnorm(0))
hist(col.m, freq = FALSE, ylim = ylim)
curve(dnorm, col.m, add=TRUE)
```



The shape of the column means soes not match that of Problem 2 because of central limit theorem, which states that the distribution of mean of large number of iterants(columns in this case) is approximately normal

Problem 9

```r
mean(y.mat[y.mat>1])
```

```
## [1] 2.002016
```

Part 2

Problem 2.a

```r
temp.data <- read.csv("Temperature.csv", header=TRUE)
temp.data$DateNr <- as.Date(temp.data$DateNr, format='%m/%d/%Y')
month <- format(temp.data$DateNr, '%m')
year <- format(temp.data$DateNr, '%Y')

temp.df <- data.frame(month)
temp.df$month <- as.factor(month)
temp.df$year <- as.factor(year)
temp.df$Station <- temp.data$Station
temp.df$temperature <- temp.data$Temperature
head(temp.df)
```

```
##   month year Station temperature
## 1    10 1990    DANT         4.0
## 2    06 1990    DANT         6.0
## 3    08 1990    DANT         7.3
## 4    04 1990    DANT         8.2
## 5    09 1990    DANT        17.4
## 6    06 1990    DANT        18.1
```

```r
agg <- aggregate(temperature~year+month, data=temp.df, mean, na.rm=TRUE)
head(agg)
```

```
##   year month temperature
## 1 1990    01    6.788889
## 2 1991    01    6.600000
## 3 1992    01    6.013953
## 4 1993    01    8.511111
## 5 1994    01    6.147059
## 6 1995    01    7.902857
```

```r
tempeture.list <- agg$temperature
```

```r
print(agg[, c('year', 'month', 'temperature')])
```

```
##     year month temperature
## 1   1990    01    6.788889
## 2   1991    01    6.600000
## 3   1992    01    6.013953
```

```
## 4   1993    01    8.511111
## 5   1994    01    6.147059
## 6   1995    01    7.902857
## 7   1996    01    4.038462
## 8   1997    01    5.416000
## 9   1998    01    9.254324
## 10  1999    01    9.961724
## 11  2000    01    8.637727
## 12  2001    01    7.035714
## 13  2002    01   10.958636
## 14  2003    01    7.190741
## 15  2004    01    9.739167
## 16  2005    01    8.729143
## 17  1990    02    8.475000
## 18  1991    02   10.207143
## 19  1992    02    6.212500
## 20  1993    02    5.750000
## 21  1994    02    7.489189
## 22  1995    02   10.011429
## 23  1996    02    4.086364
## 24  1997    02   10.378182
## 25  1998    02   10.388333
## 26  1999    02    7.201600
## 27  2000    02    7.294865
## 28  2001    02    9.698947
## 29  2002    02   12.540385
## 30  2003    02    9.150000
## 31  2004    02    7.656818
## 32  2005    02    9.139355
## 33  1990    03    8.330769
## 34  1991    03    8.220000
## 35  1992    03    8.747826
## 36  1993    03    7.324242
## 37  1994    03   10.267347
## 38  1995    03    8.495385
## 39  1996    03    5.297222
## 40  1997    03    6.688200
## 41  1998    03   10.344444
## 42  1999    03    8.962500
## 43  2000    03    8.725490
## 44  2001    03    7.328378
## 45  2002    03   10.792128
## 46  2003    03    8.006786
## 47  2004    03    8.062955
## 48  2005    03    7.916500
## 49  1990    04    8.777419
## 50  1991    04    9.218182
## 51  1992    04   10.968333
## 52  1993    04    9.685714
## 53  1994    04    9.011905
## 54  1995    04   10.160000
## 55  1996    04    6.688235
## 56  1997    04    8.038293
## 57  1998    04   11.072727
```

```
## 58  1999   04  11.264138
## 59  2000   04  11.291333
## 60  2001   04   8.908235
## 61  2002   04  10.095111
## 62  2003   04   9.900816
## 63  2004   04  10.198491
## 64  2005   04  10.725918
## 65  1990   05  12.775758
## 66  1991   05   8.166667
## 67  1992   05  11.140313
## 68  1993   05  13.000000
## 69  1994   05  13.661538
## 70  1995   05  11.135135
## 71  1996   05   9.795312
## 72  1997   05  12.238462
## 73  1998   05  13.465116
## 74  1999   05  14.098378
## 75  2000   05  14.380909
## 76  2001   05  12.855172
## 77  2002   05  13.092093
## 78  2003   05  14.544118
## 79  2004   05  12.832250
## 80  2005   05  13.721176
## 81  1990   06  13.361290
## 82  1991   06  11.088889
## 83  1992   06  15.712069
## 84  1993   06  15.340741
## 85  1994   06  13.222222
## 86  1995   06  12.572917
## 87  1996   06  14.558621
## 88  1997   06  15.856154
## 89  1998   06  15.580000
## 90  1999   06  15.377321
## 91  2000   06  14.906923
## 92  2001   06  14.370750
## 93  2002   06  14.962667
## 94  2003   06  17.653333
## 95  2004   06  15.159000
## 96  2005   06  15.702692
## 97  1990   07  15.877143
## 98  1991   07  15.838889
## 99  1992   07  14.817544
## 100 1993   07  15.163415
## 101 1994   07  15.886441
## 102 1995   07  15.657143
## 103 1996   07  17.524242
## 104 1997   07  18.232982
## 105 1998   07  15.273778
## 106 1999   07  18.252000
## 107 2000   07  16.334894
## 108 2001   07  17.797200
## 109 2002   07  17.302041
## 110 2003   07  18.684694
## 111 2004   07  16.724909
```

```
## 112 2005    07   17.469459
## 113 1990    08   16.892308
## 114 1991    08   16.489286
## 115 1992    08   13.987500
## 116 1993    08   12.525536
## 117 1994    08   16.296154
## 118 1995    08   17.843860
## 119 1996    08   17.036508
## 120 1997    08   18.162222
## 121 1998    08   15.752500
## 122 1999    08   16.624792
## 123 2000    08   18.524043
## 124 2001    08   18.885500
## 125 2002    08   17.885455
## 126 2003    08   17.482857
## 127 2004    08   15.713750
## 128 2005    08   16.060000
## 129 1990    09   14.446154
## 130 1991    09   12.973333
## 131 1992    09   12.848039
## 132 1993    09   13.252037
## 133 1994    09   13.223382
## 134 1995    09   15.120000
## 135 1996    09   13.481034
## 136 1997    09   15.949583
## 137 1998    09   14.495000
## 138 1999    09   18.681364
## 139 2000    09   15.409459
## 140 2001    09   13.563158
## 141 2002    09   16.697838
## 142 2003    09   15.632609
## 143 2004    09   14.387222
## 144 2005    09   16.324286
## 145 1990    10   12.666667
## 146 1991    10   11.809091
## 147 1992    10   11.894615
## 148 1993    10   13.316667
## 149 1994    10   12.417647
## 150 1995    10   13.741463
## 151 1996    10   14.212000
## 152 1997    10   13.689000
## 153 1998    10    8.994375
## 154 1999    10   13.111842
## 155 2000    10   12.390465
## 156 2001    10   13.095246
## 157 2002    10   11.058649
## 158 2003    10   11.120244
## 159 2004    10   13.198525
## 160 2005    10   13.311389
## 161 1990    11   11.070968
## 162 1991    11    8.824444
## 163 1992    11   11.472927
## 164 1993    11    9.250000
## 165 1994    11   12.021951
```

```
## 166 1995    11   11.784615
## 167 1996    11   10.106727
## 168 1997    11   12.714565
## 169 1998    11    8.992941
## 170 1999    11    7.147619
## 171 2000    11   10.396981
## 172 2001    11   11.143158
## 173 2002    11    9.605217
## 174 2003    11    9.591622
## 175 2004    11   12.233158
## 176 2005    11   11.864054
## 177 1990    12    7.913636
## 178 1991    12    9.121622
## 179 1992    12    8.122188
## 180 1993    12    8.975610
## 181 1994    12   11.083636
## 182 1995    12   11.168889
## 183 1996    12    8.547500
## 184 1997    12    9.422000
## 185 1998    12    9.570000
## 186 1999    12    9.077955
## 187 2000    12    8.494400
## 188 2001    12    9.220488
## 189 2002    12    8.426596
## 190 2003    12    9.460000
## 191 2004    12   10.121579
## 192 2005    12   10.462500
```

Problem 2.b

```r
count <- as.data.frame(table(temp.df$Station))
print(count)
```

```
##    Var1 Freq
## 1  DANT  300
## 2  DREI  293
## 3    G6  278
## 4  GROO  296
## 5  HAMM  295
## 6  HANS  309
## 7  HUIB  296
## 8  LODS  294
## 9  MARS  296
## 10  N02  402
## 11  N10  665
## 12  N20  266
## 13  N70  268
## 14  R03  161
## 15  R50  106
## 16  R70  106
## 17 SOEL  295
## 18 T004  339
```

```
## 19 T010  261
## 20 T100  258
## 21 T135  259
## 22 T175  258
## 23 T235  258
## 24 VLIS  421
## 25  W02  272
## 26  W20  191
## 27  W70  190
## 28 WISS  296
## 29 ZIJP  296
## 30 ZUID  303
```

```
sorted <- count[order(-count$Freq),]
top10 <- sorted[1:10,]
```

Top 10 stations(with most number of readings):

```
print(top10)
```

```
##    Var1 Freq
## 11  N10  665
## 24 VLIS  421
## 10  N02  402
## 18 T004  339
## 6  HANS  309
## 30 ZUID  303
## 1  DANT  300
## 4  GROO  296
## 7  HUIB  296
## 9  MARS  296
```

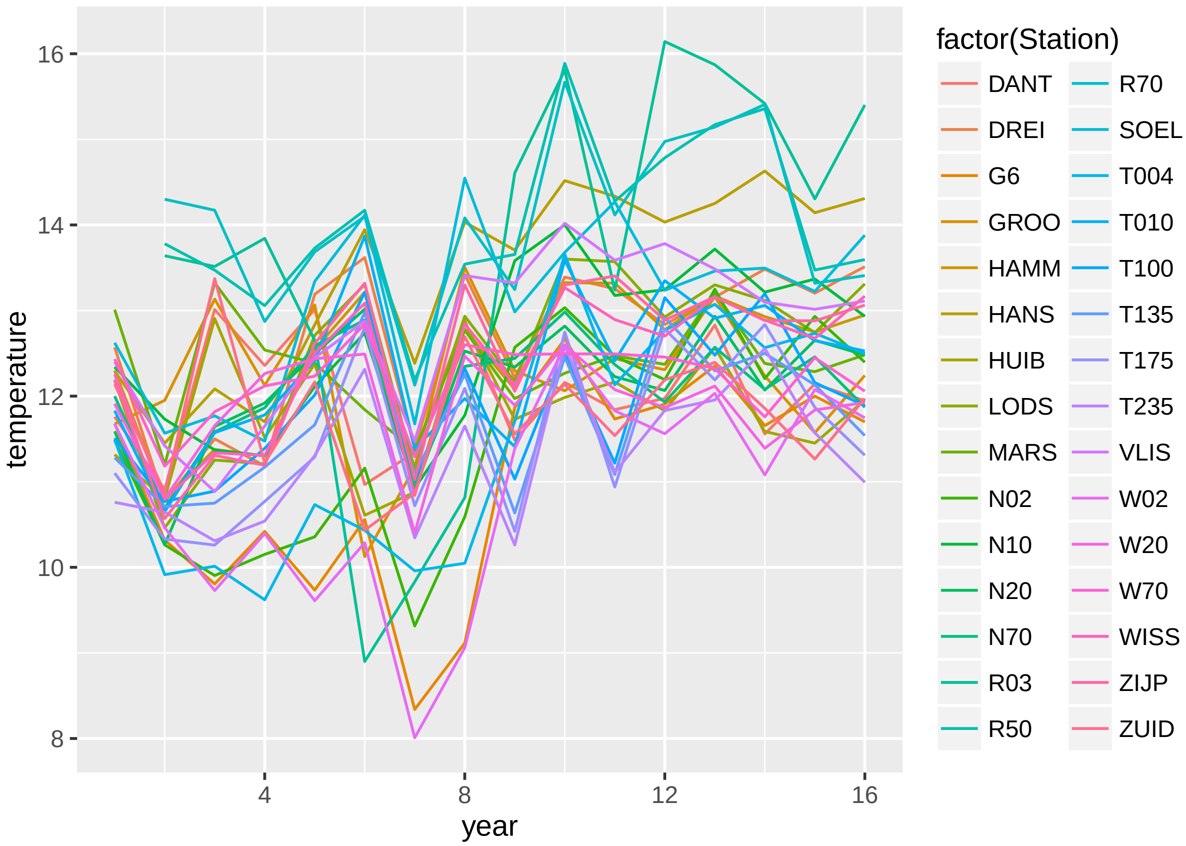
```
top10.stations <- sorted$Var1
```

```
agg<-aggregate(temperature~Station+year+month, data=temp.df, mean)
head(agg)
```

```
##   Station year month temperature
## 1    HAMM 1990    01    5.800000
## 2    HANS 1990    01    5.900000
## 3    LODS 1990    01    5.400000
## 4     N10 1990    01    8.766667
## 5    VLIS 1990    01    6.200000
## 6    WISS 1990    01    5.900000
```

In the following part to draw the time series, I simply aggregate by years (since including month leads to a lot of data points on the X axis)

```
library(ggplot2)
agg<-aggregate(temperature~Station+year, data=temp.df, mean)
agg$year <- as.numeric(agg$year)
ggplot(agg, aes(x=year, y=temperature)) + geom_line() + aes(color=factor(Station))
```


factor(Station)
DANT
DREI
G6
GROO
HAMM
HANS
HUIB
LODS
MARS
N02
N10
N20
N70
R03
R50
R70
SOEL
T004
T010
T100
T135
T175
T235
VLIS
W02
W20
W70
WISS
ZIJP
ZUID
temperature
16
14
12
10
8
4
8
12
16
year