2. Basics of R

```
# Reference book: "Beginning R: The Statistical Programming Language"
# Author: Dr. Mark Gardener
```

Simple Vector

```
data1 = c(3, 5, 7, 5, 3, 2, 6, 8, 5, 6, 9)
data1
```

```
 [1] 3 5 7 5 3 2 6 8 5 6 9
```

```
data1[1]
```

```
[1] 3
```

```
data1[3]
```

```
[1] 7
```

```
data1[1:3]
```

```
[1] 3 5 7
```

```
data1[-1]
```

```
 [1] 5 7 5 3 2 6 8 5 6 9
```

```
data1[-1:-6]
```

```
[1] 6 8 5 6 9
```

```
data1[c(1, 3, 4, 8)]
```

```
[1] 3 7 5 8
```

```
data1[data1 > 3]
```

```
[1] 5 7 5 6 8 5 6 9
```

```
data1[data1 < 5 | data1 > 7]
```

```
[1] 3 3 2 8 9
```

```
length(data1)
```

```
[1] 11
```

```
max(data1)
```

```
[1] 9
```

```
which(data1 == max(data1))
```

```
[1] 11
```

```
which(data1 == min(data1))
```

```
[1] 6
```

```
which.max(data1)
```

```
[1] 11
```

```
which.min(data1)
```

```
[1] 6
```

Data sorting, ordering, ranking

```
data1 = c(3, 5, 7, 5, 3, 2, 6, 8, 5, 6, 9)
length(data1)
```

```
[1] 11
```

```
# seq(start, end, interval)
sq=seq(1, length(data1), 2)
sq
```

```
[1]  1  3  5  7  9 11
```

```
data1[sq]
```

```
[1] 3 7 3 6 5 9
```

```r
unique(data1)
```

```
[1] 3 5 7 2 6 8 9
```

```r
rev(data1)
```

```
 [1] 9 6 5 8 6 2 3 5 7 5 3
```

```r
diff(data1)
```

```
 [1]  2  2 -2 -2 -1  4  2 -3  1  3
```

```r
# sorting of data
sort(data1)
```

```
 [1] 2 3 3 5 5 5 6 6 7 8 9
```

```r
sort(data1,decreasing = TRUE)
```

```
 [1] 9 8 7 6 6 5 5 5 3 3 2
```

```r
# Ordering of data
order(data1) # get an index, tells the position of each item along the vector.
```

```
 [1]  6  1  5  2  4  9  7 10  3  8 11
```

```r
# Randking of data
rank(data1)
```

```
 [1]  2.5  5.0  9.0  5.0  2.5  1.0  7.5 10.0  5.0  7.5 11.0
```

```r
rank(data1, ties.method ="min")
```

```
 [1]  2  4  9  4  2  1  7 10  4  7 11
```

```r
rank(data1, ties.method ="average")
```

```
 [1]  2.5  5.0  9.0  5.0  2.5  1.0  7.5 10.0  5.0  7.5 11.0
```

```r
rank(data1, ties.method ="max")
```

```
 [1]  3  6  9  6  3  1  8 10  6  8 11
```

```r
rank(data1, ties.method ="random")
```

```
 [1]  3  5  9  4  2  1  7 10  6  8 11
```

Logical values and Data extraction

```r
data1 = c(3, 5, 7, 5, 3, 2, 6, 8, 5, 6, 9)
data1
```

```
 [1] 3 5 7 5 3 2 6 8 5 6 9
```

```r
which(data1 == 6) # gives index
```

```
[1]  7 10
```

```r
data1 == 6 # logical values
```

```
 [1] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
```

```r
data1 > 5 # logical values
```

```
 [1] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
```

```r
data1 > 5 & data1 < 8 # logical values
```

```
 [1] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
```

```r
data1[length(data1)]
```

```
[1] 9
```

```r
data1[-c(1,length(data1))]
```

```
[1] 5 7 5 3 2 6 8 5 6
```

```r
data1[data1 != 5]
```

```
[1] 3 7 3 2 6 8 6 9
```

```r
data1[data1 %in% c(5,6,8)]
```

```
[1] 5 5 6 8 5 6
```

```r
data1[!(data1 %in% c(5,6,8))]
```

```
[1] 3 7 3 2 9
```

```r
data1[data1==7]=10
data1
```

```
 [1]  3  5 10  5  3  2  6  8  5  6  9
```

```
duplicated(data1)
```

```
 [1] FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE
```

```
temp =subset(data1,data1>5)
temp
```

```
[1] 10  6  8  6  9
```

Basic Statistics

```
data1 = c(3, 5, 7, 5, 3, 2, 6, 8, 5, 6, 9)
data1
```

```
 [1] 3 5 7 5 3 2 6 8 5 6 9
```

```
sum(data1)
```

```
[1] 59
```

```
mean(data1)
```

```
[1] 5.363636
```

```
sd(data1)
```

```
[1] 2.15744
```

```
summary(data1)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   4.000   5.000   5.364   6.500   9.000
```

```
cumsum(data1)
```

```
 [1]  3  8 15 20 23 25 31 39 44 50 59
```

```
example(cumsum)
```

```
cumsum> cumsum(1:10)
 [1]  1  3  6 10 15 21 28 36 45 55

cumsum> cumprod(1:10)
 [1]       1       2       6      24     120     720    5040   40320  362880
[10] 3628800

cumsum> cummin(c(3:1, 2:0, 4:2))
[1] 3 2 1 1 1 0 0 0 0

cumsum> cummax(c(3:1, 2:0, 4:2))
[1] 3 3 3 3 3 3 4 4 4
```

Shortcut for letters and months

```
LETTERS
```

```
 [1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L" "M" "N" "O" "P" "Q" "R" "S"
[20] "T" "U" "V" "W" "X" "Y" "Z"
```

```
letters
```

```
 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s"
[20] "t" "u" "v" "w" "x" "y" "z"
```

```
month.abb
```

```
 [1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
```

More on vectors

```
vec1=c(1,2,3,14,15)
vec2=c(11,12,13,14,15)
rev(vec1)
```

```
[1] 15 14  3  2  1
```

```
vec1^2
```

```
[1]   1   4   9 196 225
```

```
vec1*vec2
```

```
[1]  11  24  39 196 225
```

```
vec2/vec1
```

```
[1] 11.000000  6.000000  4.333333  1.000000  1.000000
```

```
vec2 - vec1
```

```
[1] 10 10 10  0  0
```

```
vec1 + vec2
```

```
[1] 12 14 16 28 30
```

```r
sum(vec1)+sum(vec2)
```

```
[1] 100
```

```r
mean(sqrt(vec1+vec2))
```

```
[1] 4.394897
```

```r
intersect(vec1,vec2)
```

```
[1] 14 15
```

```r
setdiff(vec1,vec2)
```

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

```r
union(vec1,vec2)
```

```
[1]  1  2  3 14 15 11 12 13
```

Data Frame

```r
marks = c(12, 15, 17, 13,  8, 9, 8, 7)
grade = c("first", "first", "first","first", "second", "second", "second", "second")
grade = c(rep(c("first","second"),each=4))
grade = c(rep(c("first","second"),times=4))
result = data.frame(marks, grade)
result
```

```
  marks  grade
1    12  first
2    15 second
3    17  first
4    13 second
5     8  first
6     9 second
7     8  first
8     7 second
```

```r
str(result)
```

```
'data.frame':   8 obs. of  2 variables:
 $ marks: num  12 15 17 13 8 9 8 7
 $ grade: chr  "first" "second" "first" "second" ...
```

```r
names(result)
```

```
[1] "marks" "grade"
```

```r
attributes(result)
```

```
$names
[1] "marks" "grade"

$class
[1] "data.frame"

$row.names
[1] 1 2 3 4 5 6 7 8
```

Data extraction from data frame

```r
nrow(result)
```

```
[1] 8
```

```r
ncol(result)
```

```
[1] 2
```

```r
result$grade # use $ to extract a particular part of a data item.
```

```
[1] "first"  "second" "first"  "second" "first"  "second" "first"  "second"
```

```r
result$grade[1:3]
```

```
[1] "first"  "second" "first"
```

```r
result[3, 1]
```

```
[1] 17
```

```r
result[4, 1:2]
```

```
  marks  grade
4    13 second
```

```r
result[, 1]
```

```
[1] 12 15 17 13  8  9  8  7
```

```
result[c(1, 3, 5), ]
```

```
  marks grade
1    12 first
3    17 first
5     8 first
```

```
result[c(1, 3, 5), -2]
```

```
[1] 12 17  8
```

```
result[c(1,3,5),"grade"]
```

```
[1] "first" "first" "first"
```

```
result$grade
```

```
[1] "first"  "second" "first"  "second" "first"  "second" "first"  "second"
```

Sorting, ordering of data frame

```
sort(result[, 1])
```

```
[1]  7  8  8  9 12 13 15 17
```

```
sort(result[, "grade"])
```

```
[1] "first"  "first"  "first"  "first"  "second" "second" "second" "second"
```

```
sort(result$grade)
```

```
[1] "first"  "first"  "first"  "first"  "second" "second" "second" "second"
```

```
order(result)
```

```
 [1]  8  5  7  6  1  4  2  3  9 11 13 15 10 12 14 16
```

Viewing and setting names for data frame

```
result
```

```
  marks  grade
1    12  first
2    15 second
3    17  first
4    13 second
5     8  first
6     9 second
7     8  first
8     7 second
```

```
names(result)
```

```
[1] "marks" "grade"
```

```
colnames(result) = c("score", "class")
rownames(result) = c("s1", "s2", "s3", "s4", "s5", "s6", "s7", "s8")
result
```

```
   score  class
s1    12  first
s2    15 second
s3    17  first
s4    13 second
s5     8  first
s6     9 second
s7     8  first
s8     7 second
```

Matrix

```
data3 = matrix(c(1, 2, 3, 10, 11, 12, 21, 22, 23), nrow = 3, ncol = 3)
data3
```

```
     [,1] [,2] [,3]
[1,]    1   10   21
[2,]    2   11   22
[3,]    3   12   23
```

```
data3 = matrix(c(1, 2, 3, 10, 11, 12, 21, 22, 23), nrow = 3, ncol = 3, byrow = T)
data3
```

```
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]   10   11   12
[3,]   21   22   23
```

```
data3 = matrix(c(1, 2, 3, 10, 11, 12, 21, 22, 23), nrow = 3, ncol = 3, byrow = FALSE, dimnames = list(c
data3
```

```
   c1 c2 c3
r1  1 10 21
r2  2 11 22
r3  3 12 23
```

Data extraction from matrix

```
data3[2, ]
```

```
c1 c2 c3
 2 11 22
```

```
data3[, 3]
```

```
r1 r2 r3
21 22 23
```

```
data3[2,3]
```

```
[1] 22
```

```
data3[5]
```

```
[1] 11
```

```
data3[, "c2"]
```

```
r1 r2 r3
10 11 12
```

```
data3["r1", ]
```

```
c1 c2 c3
 1 10 21
```

Sorting, ordering and ranking of matrix

```
sort(data3)
```

```
[1]  1  2  3 10 11 12 21 22 23
```

```
sort(data3[, 1])
```

```
r1 r2 r3
 1  2  3
```

```
order(data3)
```

```
[1] 1 2 3 4 5 6 7 8 9
```

```r
order(data3[, 1])
```

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

```r
rank(data3)
```

```
[1] 1 2 3 4 5 6 7 8 9
```

```r
rank(data3[, 1])
```

```
r1 r2 r3
 1  2  3
```

Setting names for matrix

```r
rownames(data3) = c("Row1", "Row2", "Row3")
colnames(data3) = c("Col1", "Col2", "Col3")
```

Another way of setting names for matrix

```r
rown = c("rowno1", "rowno2", "rowno3")
coln = c("colno1", "colno2", "colno3")
dimnames(data3) = list(rown, coln) #Changing names in Matrix
data3
```

```
       colno1 colno2 colno3
rowno1      1     10     21
rowno2      2     11     22
rowno3      3     12     23
```

More on matrix

```r
mat1=matrix(c(4,5,6,7,8,9,10,11,12),nrow = 3,ncol=3)
mat2=matrix(c(13,14,15,16,17,18,19,20,21),nrow = 3,ncol = 3)
mat1;mat2
```

```
     [,1] [,2] [,3]
[1,]    4    7   10
[2,]    5    8   11
[3,]    6    9   12

     [,1] [,2] [,3]
[1,]   13   16   19
[2,]   14   17   20
[3,]   15   18   21
```

```
mat1+10
```

```
     [,1] [,2] [,3]
[1,]   14   17   20
[2,]   15   18   21
[3,]   16   19   22
```

```
mat1+mat2
```

```
     [,1] [,2] [,3]
[1,]   17   23   29
[2,]   19   25   31
[3,]   21   27   33
```

```
mat1-mat2
```

```
     [,1] [,2] [,3]
[1,]   -9   -9   -9
[2,]   -9   -9   -9
[3,]   -9   -9   -9
```

```
mat1/mat2
```

```
          [,1]      [,2]      [,3]
[1,] 0.3076923 0.4375000 0.5263158
[2,] 0.3571429 0.4705882 0.5500000
[3,] 0.4000000 0.5000000 0.5714286
```

```
diag(mat1)
```

```
[1]  4  8 12
```

```
mat3 =rbind(mat1,mat2)
mat3
```

```
     [,1] [,2] [,3]
[1,]    4    7   10
[2,]    5    8   11
[3,]    6    9   12
[4,]   13   16   19
[5,]   14   17   20
[6,]   15   18   21
```

```
mat4 =cbind(mat1,mat2)
mat4
```

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    4    7   10   13   16   19
[2,]    5    8   11   14   17   20
[3,]    6    9   12   15   18   21
```

Matrix using rbind and cbind command

```
sample1 = c(5, 6, 9, 12, 8)
sample2 = c(7, 9, 13, 10)
#cmat = cbind(sample1, sample2) # not working? why?
length(sample2) = 5
length(sample2) = length(sample1)
cmat = cbind(sample1, sample2)
cmat
```

```
     sample1 sample2
[1,]       5       7
[2,]       6       9
[3,]       9      13
[4,]      12      10
[5,]       8      NA
```

```
cmat1 = rbind(sample1, sample2)
cmat1
```

```
        [,1] [,2] [,3] [,4] [,5]
sample1    5    6    9   12    8
sample2    7    9   13   10   NA
```

```
sample3 = c("a", "b", "c", "d", "e")
mix.mat = cbind(sample1, sample2, sample3)
mix.mat #all the data items are converted to characters.
```

```
     sample1 sample2 sample3
[1,] "5"     "7"     "a"
[2,] "6"     "9"     "b"
[3,] "9"     "13"    "c"
[4,] "12"    "10"    "d"
[5,] "8"     NA      "e"
```

Lists

```
data5=c(month.abb)
mylist = list(marks, data3, grade, data5)
mylist
```

```
[[1]]
[1] 12 15 17 13  8  9  8  7

[[2]]
       colno1 colno2 colno3
rowno1      1     10     21
rowno2      2     11     22
rowno3      3     12     23
```

```
[[3]]
[1] "first"  "second" "first"  "second" "first"  "second" "first"  "second"

[[4]]
 [1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
```

```
names(mylist) = c("mr", "d3", "gd", "d5")
mylist
```

```
$mr
[1] 12 15 17 13  8  9  8  7

$d3
       colno1 colno2 colno3
rowno1      1     10     21
rowno2      2     11     22
rowno3      3     12     23

$gd
[1] "first"  "second" "first"  "second" "first"  "second" "first"  "second"

$d5
 [1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
```

```
mylist$mr
```

```
[1] 12 15 17 13  8  9  8  7
```

View few data and Descriptive Statistics

```
result
```

```
   score  class
s1    12  first
s2    15 second
s3    17  first
s4    13 second
s5     8  first
s6     9 second
s7     8  first
s8     7 second
```

```
head(result)
```

```
   score  class
s1    12  first
s2    15 second
s3    17  first
s4    13 second
s5     8  first
s6     9 second
```

```r
head(result, n = 3)
```

```
   score  class
s1    12  first
s2    15 second
s3    17  first
```

```r
tail(result)
```

```
   score  class
s3    17  first
s4    13 second
s5     8  first
s6     9 second
s7     8  first
s8     7 second
```

```r
tail(result, n = 4)
```

```
   score  class
s5     8  first
s6     9 second
s7     8  first
s8     7 second
```

```r
summary(result$marks)
```

```
Length  Class   Mode
     0   NULL   NULL
```

```r
result.t = t(result) #Rotating Data Tables
temp = unstack(result)
temp
```

```
  first second
1    12     15
2    17     13
3     8      9
4     8      7
```

```r
stack(temp) # command to combine the values into a data frame.
```

```
  values    ind
1     12  first
2     17  first
3      8  first
4      8  first
5     15 second
6     13 second
7      9 second
8      7 second
```

More examples

```
height = c(9, 11, 6, 14, 17, 19, 28, 31, 32, 7, 6, 5, 14, 17, 15, 44, 38, 37)
plant = c(rep("vulgaris" ,9), rep("sativa", 9))
water = c(rep(c("lo","mid","hi"),each=3,times=2 ))
pw = data.frame(height, plant, water)
pw
```

```
   height    plant water
1       9 vulgaris    lo
2      11 vulgaris    lo
3       6 vulgaris    lo
4      14 vulgaris   mid
5      17 vulgaris   mid
6      19 vulgaris   mid
7      28 vulgaris    hi
8      31 vulgaris    hi
9      32 vulgaris    hi
10      7   sativa    lo
11      6   sativa    lo
12      5   sativa    lo
13     14   sativa   mid
14     17   sativa   mid
15     15   sativa   mid
16     44   sativa    hi
17     38   sativa    hi
18     37   sativa    hi
```

Using Statck and Unstack command

```
unstack(pw, form = height ~ plant)
```

```
  sativa vulgaris
1      7        9
2      6       11
3      5        6
4     14       14
5     17       17
6     15       19
7     44       28
8     38       31
9     37       32
```

```
pw.us = unstack(pw, form = height ~ water)
pw.us
```

```
  hi lo mid
1 28  9  14
2 31 11  17
```

```
3 32  6  19
4 44  7  14
5 38  6  17
6 37  5  15
```

```
stack(pw.us, select = c(hi, lo))
```

```
   values ind
1      28  hi
2      31  hi
3      32  hi
4      44  hi
5      38  hi
6      37  hi
7       9  lo
8      11  lo
9       6  lo
10      7  lo
11      6  lo
12      5  lo
```

```
stack(pw.us, select = -lo)
```

```
   values ind
1      28  hi
2      31  hi
3      32  hi
4      44  hi
5      38  hi
6      37  hi
7      14 mid
8      17 mid
9      19 mid
10     14 mid
11     17 mid
12     15 mid
```