VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

LAB ASSIGNMENT – 2
Data Manipulation using R

Name: Tejas Rahul Rokade
Reg. No.: 20BDS0033
Lab Slot: L33 & L34

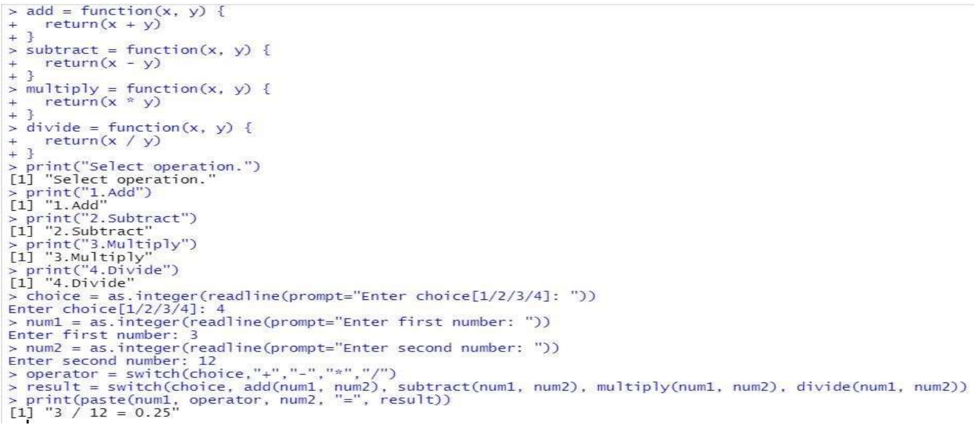
1. Create user defined functions to perform various arithmetic operations and call each functions using menu driven format (try functions with and without parameters, functions with default argument)
2. Familiarize basic statistical operation on a random vector of 100 elements
   a. Mean b. Median c. Mode d. Range
   e. IQR f. Standard deviation g. Summary h. Histogram
   i. Table
3. Perform given operations on a data frame
   a. Create a data frame
   b. Access a component ([, [[, $)
   c. Structure of data frame
   d. Add new column
   e. Add new row
   f. Delete column
   g. Delete specific row
   h. Order data frame (with, order, arrange)
4. Read Air quality dataset and handle the missing data using following technique
   a. Drop Row
   b. Drop Column
   c. Imputation (Replace with unknown, mean or Group mean)

1. Create user defined functions to perform various arithmetic operations and call each functions using menu driven format (try functions with and without parameters, functions with default argument)

Code:

```
add = function(x, y) {
return(x + y)
}
subtract = function(x, y) {
  return(x - y)
}
multiply = function(x, y) {
  return(x * y)
}
divide = function(x, y) {
  return(x / y)
}
print("Select operation.") print("1.Add") print("2.Subtract")
print("3.Multiply") print("4.Divide") choice =
as.integer(readline(prompt="Enter choice[1/2/3/4]: ")) num1 =
as.integer(readline(prompt="Enter first number: ")) num2 =
as.integer(readline(prompt="Enter second number: "))
operator = switch(choice,"+","-","*","/")
result = switch(choice, add(num1, num2), subtract(num1, num2),
multiply(num1, num2), divide(num1, num2))
print(paste(num1, operator, num2, "=", result))
```

Output:



```
> add = function(x, y) {
+   return(x + y)
+ }
> subtract = function(x, y) {
+   return(x - y)
+ }
> multiply = function(x, y) {
+   return(x * y)
+ }
> divide = function(x, y) {
+   return(x / y)
+ }
> print("Select operation.")
[1] "Select operation."
> print("1.Add")
[1] "1.Add"
> print("2.Subtract")
[1] "2.Subtract"
> print("3.Multiply")
[1] "3.Multiply"
> print("4.Divide")
[1] "4.Divide"
> choice = as.integer(readline(prompt="Enter choice[1/2/3/4]: "))
Enter choice[1/2/3/4]: 4
> num1 = as.integer(readline(prompt="Enter first number: "))
Enter first number: 3
> num2 = as.integer(readline(prompt="Enter second number: "))
Enter second number: 12
> operator = switch(choice,"+","-","*","/")
> result = switch(choice, add(num1, num2), subtract(num1, num2), multiply(num1, num2), divide(num1, num2))
> print(paste(num1, operator, num2, "=", result))
[1] "3 / 12 = 0.25"
```

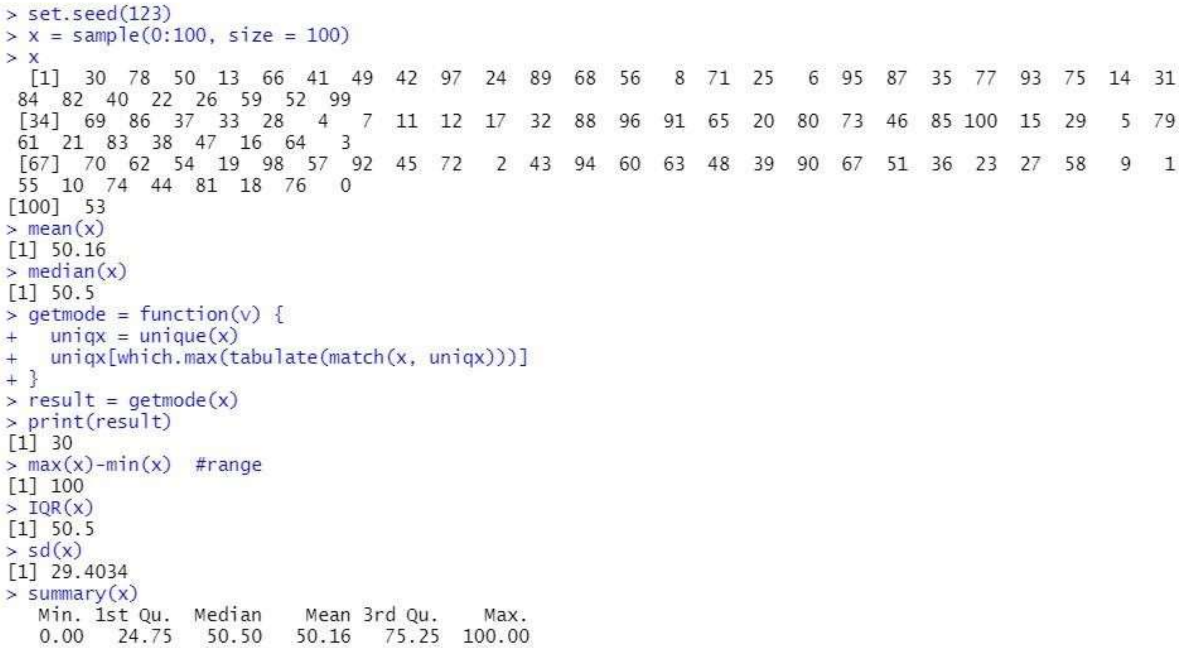
2. Familiarize basic statistical operation on a random vector of 100 elements

- ***a.*** Mean
- ***b.*** Median
- ***c.*** Mode
- ***d.*** Range
- ***e.*** IQR
- ***f.*** Standard Deviation
- ***g.*** Summary
- ***h.*** Histogram
- ***i.*** Table

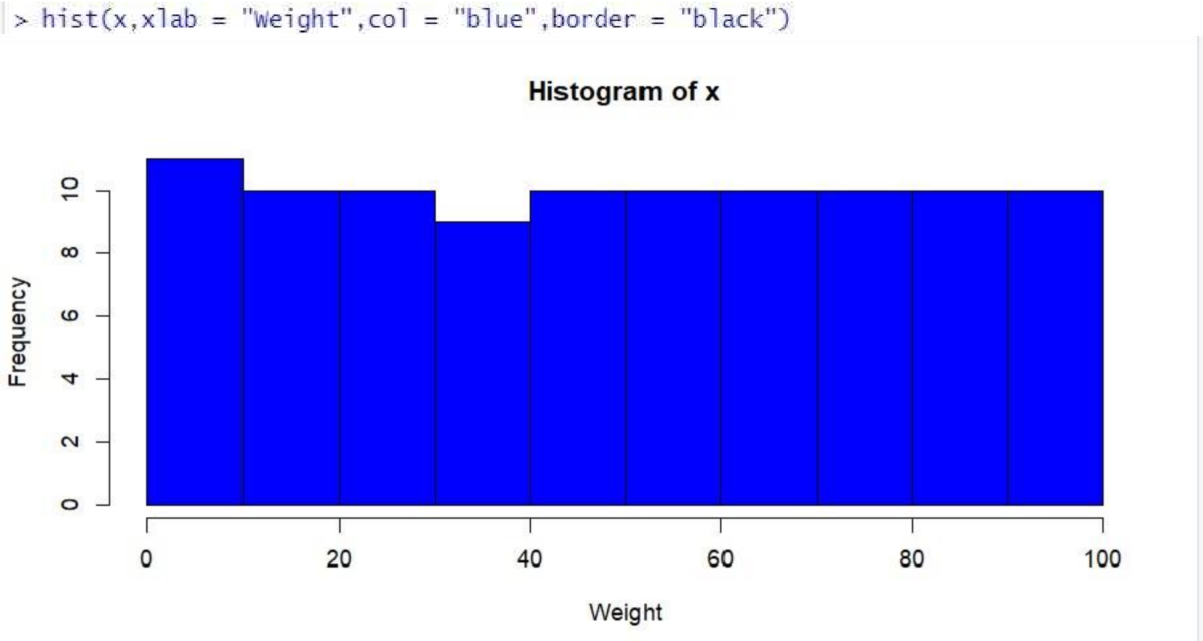
Code:

```
set.seed(123)
x = sample(0:100, size = 100)
x mean(x) median(x)
getmode = function(v) {
uniqx = unique(x)
  uniqx[which.max(tabulate(match(x, uniqx)))]
} result =
getmode(x)
print(result)
max(x)-min(x) #range
IQR(x) sd(x)
summary(x)
hist(x,xlab = "Weight",col = "blue",border = "black") table(x)
```

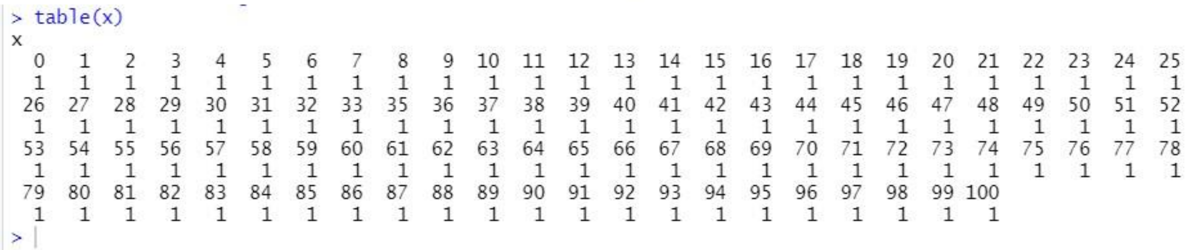
Output:



```
> set.seed(123)
> x = sample(0:100, size = 100)
> x
  [1]  30  78  50  13  66  41  49  42  97  24  89  68  56   8  71  25   6  95  87  35  77  93  75  14  31
 84  82  40  22  26  59  52  99
 [34]  69  86  37  33  28   4   7  11  12  17  32  88  96  91  65  20  80  73  46  85 100  15  29   5  79
 61  21  83  38  47  16  64   3
 [67]  70  62  54  19  98  57  92  45  72   2  43  94  60  63  48  39  90  67  51  36  23  27  58   9   1
 55  10  74  44  81  18  76   0
[100]  53
> mean(x)
[1] 50.16
> median(x)
[1] 50.5
> getmode = function(v) {
+   uniqx = unique(x)
+   uniqx[which.max(tabulate(match(x, uniqx)))]
+ }
> result = getmode(x)
> print(result)
[1] 30
> max(x)-min(x)  #range
[1] 100
> IQR(x)
[1] 50.5
> sd(x)
[1] 29.4034
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   24.75   50.50   50.16   75.25  100.00 
```



```
> hist(x,xlab = "Weight",col = "blue",border = "black")
```



```
> table(x)
x
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25 
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1 
 26  27  28  29  30  31  32  33  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52 
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1 
 53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78 
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1 
 79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1 
> |
```

3. Perform given operations on a data frame

- ***a.*** Create a data frame
- ***b.*** Access a component ([, [[, $)
- ***c.*** Structure of data frame
- ***d.*** Add new column
- ***e.*** Add new row
- ***f.*** Delete column
- ***g.*** Delete specific row
- ***h.*** Order data frame (with, order, arrange)

Code:

df = data.frame("animal"=c('koala', 'hedgehog', 'sloth', 'panda'),"country"=c('Australia', 'Italy', 'Peru', 'China'),"avg_sleep_hours"=c(21, 18, 17, 10)) print(df) df["animal"] df[["animal"]]

df$animal str(df)

cbind(df,color=c("black","grey","red","white")) rbind(df,list("penguin","Antarctica",12,"white")

) df$color=NULL

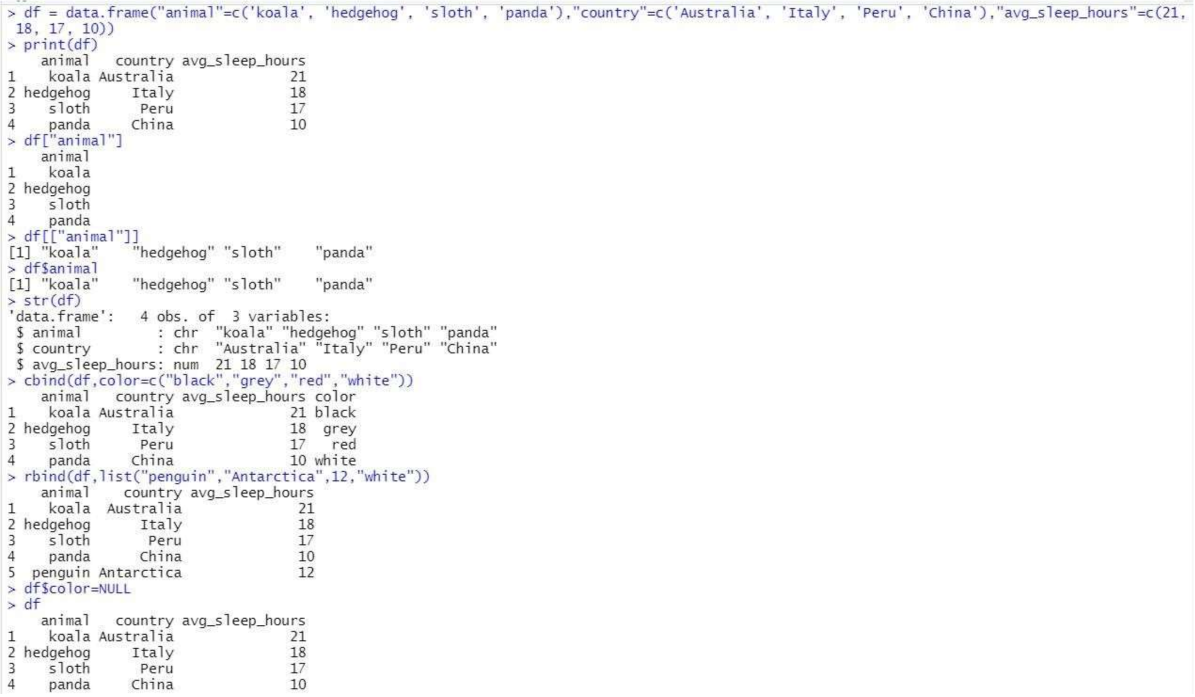
df new_df=df[-

c(2),] new_df

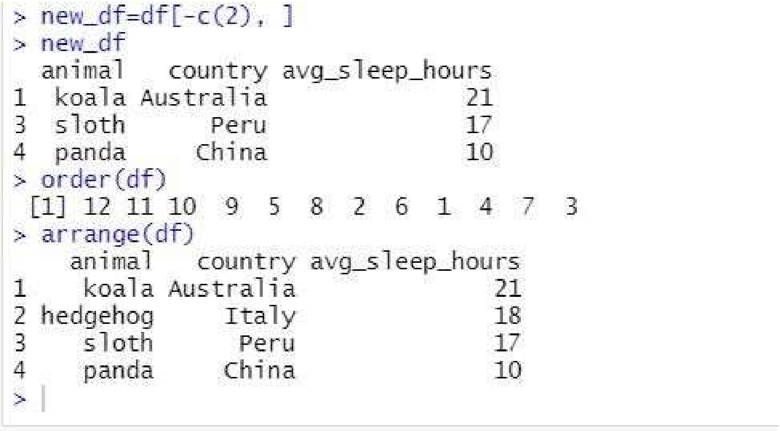
order(df)

arrange(df)

Output:



```
> df = data.frame("animal"=c('koala', 'hedgehog', 'sloth', 'panda'),"country"=c('Australia', 'Italy', 'Peru', 'China'),"avg_sleep_hours"=c(21,
 18, 17, 10))
> print(df)
   animal   country avg_sleep_hours
1    koala Australia              21
2 hedgehog     Italy              18
3    sloth      Peru              17
4    panda     China              10
> df["animal"]
   animal
1    koala
2 hedgehog
3    sloth
4    panda
> df[["animal"]]
[1] "koala"    "hedgehog" "sloth"    "panda"
> df$animal
[1] "koala"    "hedgehog" "sloth"    "panda"
> str(df)
'data.frame':	4 obs. of  3 variables:
 $ animal         : chr  "koala" "hedgehog" "sloth" "panda"
 $ country        : chr  "Australia" "Italy" "Peru" "China"
 $ avg_sleep_hours: num  21 18 17 10
> cbind(df,color=c("black","grey","red","white"))
   animal   country avg_sleep_hours color
1    koala Australia              21 black
2 hedgehog     Italy              18  grey
3    sloth      Peru              17   red
4    panda     China              10 white
> rbind(df,list("penguin","Antarctica",12,"white"))
   animal    country avg_sleep_hours
1    koala  Australia              21
2 hedgehog      Italy              18
3    sloth       Peru              17
4    panda      China              10
5  penguin Antarctica              12
> df$color=NULL
> df
   animal   country avg_sleep_hours
1    koala Australia              21
2 hedgehog     Italy              18
3    sloth      Peru              17
4    panda     China              10
```



```
> new_df=df[-c(2), ]
> new_df
  animal   country avg_sleep_hours
1  koala Australia              21
3  sloth      Peru              17
4  panda     China              10
> order(df)
 [1] 12 11 10  9  5  8  2  6  1  4  7  3
> arrange(df)
    animal   country avg_sleep_hours
1    koala Australia              21
2 hedgehog     Italy              18
3    sloth      Peru              17
4    panda     China              10
> 
```

4. Read Air quality dataset and handle the missing data using following technique

a. Drop Row

b. Drop Column

c. Imputation (Replace with unknown, mean or Group mean)

Code:

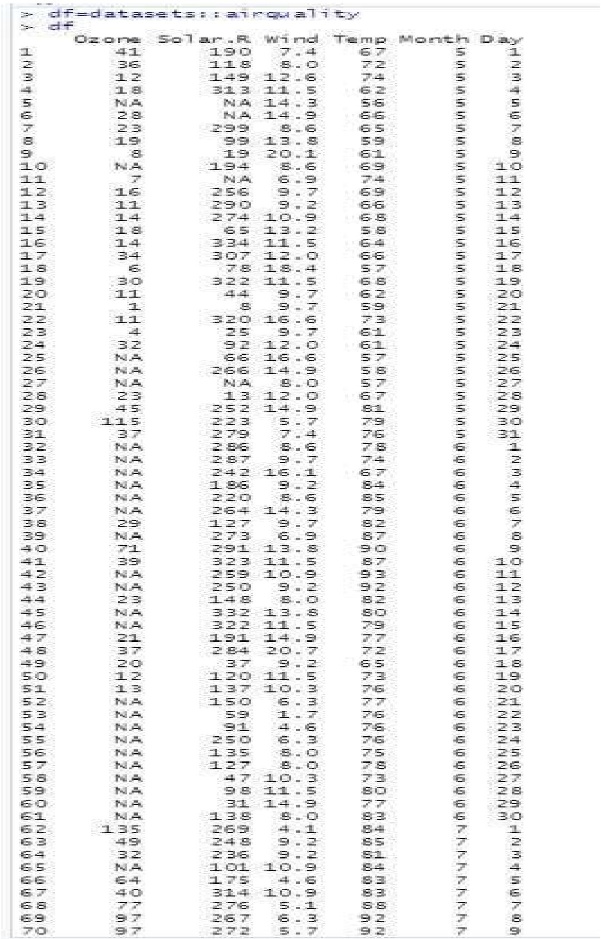
df=datasets::airquality

df y=na.omit(df) y df[,

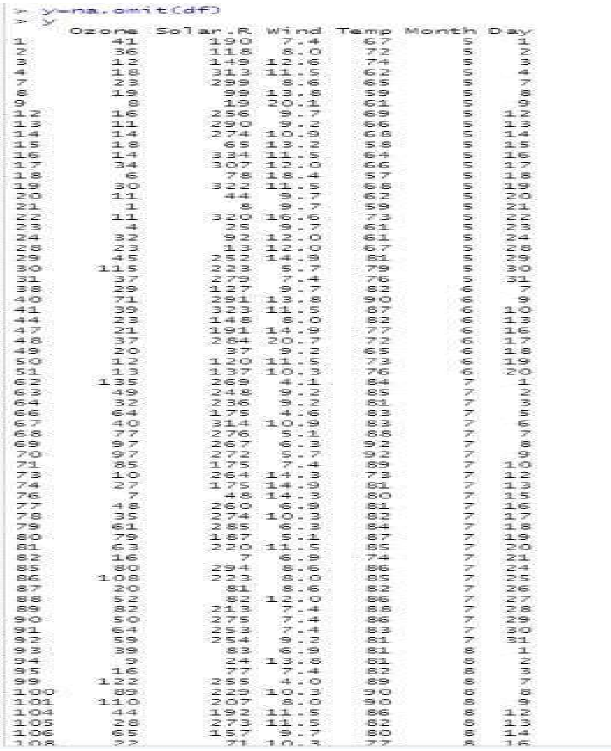
colSums(is.na(df))==0]

x=is.na(df) df[x]=0 df

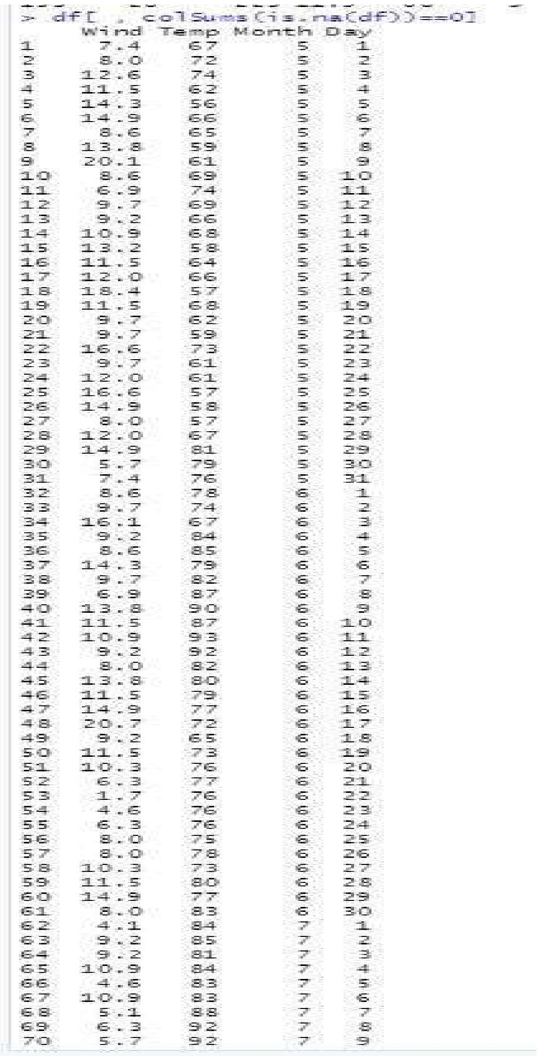
Output:



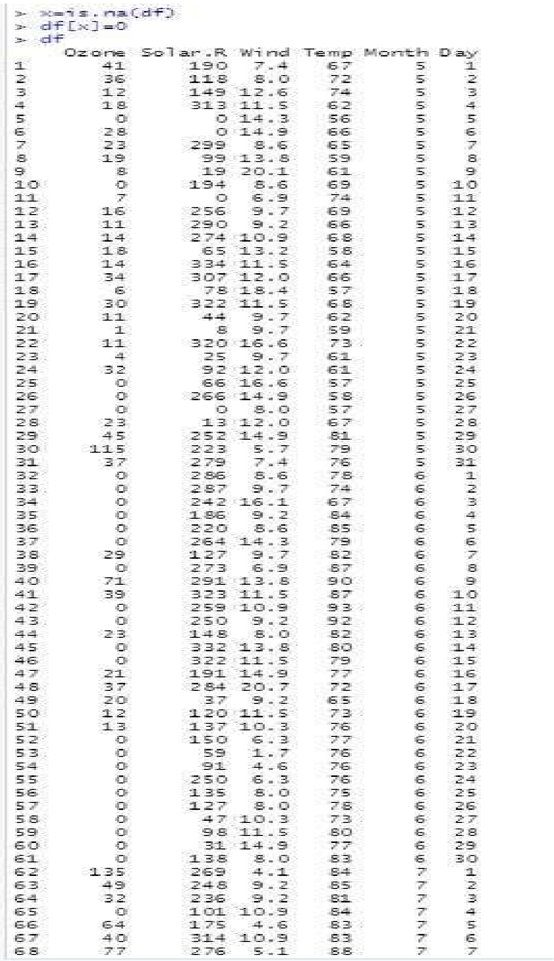
```
> df=datasets::airquality
> df
   Ozone Solar.R Wind Temp Month Day
1     41     190  7.4   67     5   1
2     36     118  8.0   72     5   2
3     12     149 12.6   74     5   3
4     18     313 11.5   62     5   4
5     NA      NA 14.3   56     5   5
6     28      NA 14.9   66     5   6
7     23     299  8.6   65     5   7
8     19      99 13.8   59     5   8
9      8      19 20.1   61     5   9
10    NA     194  8.6   69     5  10
11     7      NA  6.9   74     5  11
12    16     256  9.7   69     5  12
13    11     290  9.2   66     5  13
14    14     274 10.9   68     5  14
15    18      65 13.2   58     5  15
16    14     334 11.5   64     5  16
17    34     307 12.0   66     5  17
18     6      78 18.4   57     5  18
19    30     322 11.5   68     5  19
20    11      44  9.7   62     5  20
21     1       8  9.7   59     5  21
22    11     320 16.6   73     5  22
23     4      25  9.7   61     5  23
24    32      92 12.0   61     5  24
25    NA      66 16.6   57     5  25
26    NA     266 14.9   58     5  26
27    NA      NA  8.0   57     5  27
28    23      13 12.0   67     5  28
29    45     252 14.9   81     5  29
30   115     223  5.7   79     5  30
31    37     279  7.4   76     5  31
32    NA     286  8.6   78     6   1
33    NA     287  9.7   74     6   2
34    NA     242 16.1   67     6   3
35    NA     186  9.2   84     6   4
36    NA     220  8.6   85     6   5
37    NA     264 14.3   79     6   6
38    29     127  9.7   82     6   7
39    NA     273  6.9   87     6   8
40    71     291 13.8   90     6   9
41    39     323 11.5   87     6  10
42    NA     259 10.9   93     6  11
43    NA     250  9.2   92     6  12
44    23     148  8.0   82     6  13
45    NA     332 13.8   80     6  14
46    NA     322 11.5   79     6  15
47    21     191 14.9   77     6  16
48    37     284 20.7   72     6  17
49    20      37  9.2   65     6  18
50    12     120 11.5   73     6  19
51    13     137 10.3   76     6  20
52    NA     150  6.3   77     6  21
53    NA      59  1.7   76     6  22
54    NA      91  4.6   76     6  23
55    NA     250  6.3   76     6  24
56    NA     135  8.0   75     6  25
57    NA     127  8.0   78     6  26
58    NA      47 10.3   73     6  27
59    NA      98 11.5   80     6  28
60    NA      31 14.9   77     6  29
61    NA     138  8.0   83     6  30
62   135     269  4.1   84     7   1
63    49     248  9.2   85     7   2
64    32     236  9.2   81     7   3
65    NA     101 10.9   84     7   4
66    64     175  4.6   83     7   5
67    40     314 10.9   83     7   6
68    77     276  5.1   88     7   7
69    97     267  6.3   92     7   8
70    97     272  5.7   92     7   9
```



```
> y=na.omit(df)
> y
    Ozone Solar.R Wind Temp Month Day
1      41     190  7.4   67     5   1
2      36     118  8.0   72     5   2
3      12     149 12.6   74     5   3
4      18     313 11.5   62     5   4
7      23     299  8.6   65     5   7
8      19      99 13.8   59     5   8
9       8      19 20.1   61     5   9
12     16     256  9.7   69     5  12
13     11     290  9.2   66     5  13
14     14     274 10.9   68     5  14
15     18      65 13.2   58     5  15
16     14     334 11.5   64     5  16
17     34     307 12.0   66     5  17
18      6      78 18.4   57     5  18
19     30     322 11.5   68     5  19
20     11      44  9.7   62     5  20
21      1       8  9.7   59     5  21
22     11     320 16.6   73     5  22
23      4      25  9.7   61     5  23
24     32      92 12.0   61     5  24
28     23      13 12.0   67     5  28
29     45     252 14.9   81     5  29
30    115     223  5.7   79     5  30
31     37     279  7.4   76     5  31
38     29     127  9.7   82     6   7
40     71     291 13.8   90     6   9
41     39     323 11.5   87     6  10
44     23     148  8.0   82     6  13
47     21     191 14.9   77     6  16
48     37     284 20.7   72     6  17
49     20      37  9.2   65     6  18
50     12     120 11.5   73     6  19
51     13     137 10.3   76     6  20
62    135     269  4.1   84     7   1
63     49     248  9.2   85     7   2
64     32     236  9.2   81     7   3
66     64     175  4.6   83     7   5
67     40     314 10.9   83     7   6
68     77     276  5.1   88     7   7
69     97     267  6.3   92     7   8
70     97     272  5.7   92     7   9
71     85     175  7.4   89     7  10
73     10     264 14.3   73     7  12
74     27     175 14.9   81     7  13
76      7      48 14.3   80     7  15
77     48     260  6.9   81     7  16
78     35     274 10.3   82     7  17
79     61     285  6.3   84     7  18
80     79     187  5.1   87     7  19
81     63     220 11.5   85     7  20
82     16       7  6.9   74     7  21
85     80     294  8.6   86     7  24
86    108     223  8.0   85     7  25
87     20      81  8.6   82     7  26
88     52      82 12.0   86     7  27
89     82     213  7.4   88     7  28
90     50     275  7.4   86     7  29
91     64     253  7.4   83     7  30
92     59     254  9.2   81     7  31
93     39      83  6.9   81     8   1
94      9      24 13.8   81     8   2
95     16      77  7.4   82     8   3
99    122     255  4.0   89     8   7
100    89     229 10.3   90     8   8
101   110     207  8.0   90     8   9
104    44     192 11.5   86     8  12
105    28     273 11.5   82     8  13
106    65     157  9.7   80     8  14
108    22      71 10.3   77     8  16
```



```
> df[ , colSums(is.na(df))==0]
    Wind Temp Month Day
1    7.4   67     5   1
2    8.0   72     5   2
3   12.6   74     5   3
4   11.5   62     5   4
5   14.3   56     5   5
6   14.9   66     5   6
7    8.6   65     5   7
8   13.8   59     5   8
9   20.1   61     5   9
10   8.6   69     5  10
11   6.9   74     5  11
12   9.7   69     5  12
13   9.2   66     5  13
14  10.9   68     5  14
15  13.2   58     5  15
16  11.5   64     5  16
17  12.0   66     5  17
18  18.4   57     5  18
19  11.5   68     5  19
20   9.7   62     5  20
21   9.7   59     5  21
22  16.6   73     5  22
23   9.7   61     5  23
24  12.0   61     5  24
25  16.6   57     5  25
26  14.9   58     5  26
27   8.0   57     5  27
28  12.0   67     5  28
29  14.9   81     5  29
30   5.7   79     5  30
31   7.4   76     5  31
32   8.6   78     6   1
33   9.7   74     6   2
34  16.1   67     6   3
35   9.2   84     6   4
36   8.6   85     6   5
37  14.3   79     6   6
38   9.7   82     6   7
39   6.9   87     6   8
40  13.8   90     6   9
41  11.5   87     6  10
42  10.9   93     6  11
43   9.2   92     6  12
44   8.0   82     6  13
45  13.8   80     6  14
46  11.5   79     6  15
47  14.9   77     6  16
48  20.7   72     6  17
49   9.2   65     6  18
50  11.5   73     6  19
51  10.3   76     6  20
52   6.3   77     6  21
53   1.7   76     6  22
54   4.6   76     6  23
55   6.3   76     6  24
56   8.0   75     6  25
57   8.0   78     6  26
58  10.3   73     6  27
59  11.5   80     6  28
60  14.9   77     6  29
61   8.0   83     6  30
62   4.1   84     7   1
63   9.2   85     7   2
64   9.2   81     7   3
65  10.9   84     7   4
66   4.6   83     7   5
67  10.9   83     7   6
68   5.1   88     7   7
69   6.3   92     7   8
70   5.7   92     7   9
```



```
> x=is.na(df)
> df[x]=0
> df
   Ozone Solar.R Wind Temp Month Day
1     41     190  7.4   67     5   1
2     36     118  8.0   72     5   2
3     12     149 12.6   74     5   3
4     18     313 11.5   62     5   4
5      0       0 14.3   56     5   5
6     28       0 14.9   66     5   6
7     23     299  8.6   65     5   7
8     19      99 13.8   59     5   8
9      8      19 20.1   61     5   9
10     0     194  8.6   69     5  10
11     7       0  6.9   74     5  11
12    16     256  9.7   69     5  12
13    11     290  9.2   66     5  13
14    14     274 10.9   68     5  14
15    18      65 13.2   58     5  15
16    14     334 11.5   64     5  16
17    34     307 12.0   66     5  17
18     6      78 18.4   57     5  18
19    30     322 11.5   68     5  19
20    11      44  9.7   62     5  20
21     1       8  9.7   59     5  21
22    11     320 16.6   73     5  22
23     4      25  9.7   61     5  23
24    32      92 12.0   61     5  24
25     0      66 16.6   57     5  25
26     0     266 14.9   58     5  26
27     0       0  8.0   57     5  27
28    23      13 12.0   67     5  28
29    45     252 14.9   81     5  29
30   115     223  5.7   79     5  30
31    37     279  7.4   76     5  31
32     0     286  8.6   78     6   1
33     0     287  9.7   74     6   2
34     0     242 16.1   67     6   3
35     0     186  9.2   84     6   4
36     0     220  8.6   85     6   5
37     0     264 14.3   79     6   6
38    29     127  9.7   82     6   7
39     0     273  6.9   87     6   8
40    71     291 13.8   90     6   9
41    39     323 11.5   87     6  10
42     0     259 10.9   93     6  11
43     0     250  9.2   92     6  12
44    23     148  8.0   82     6  13
45     0     332 13.8   80     6  14
46     0     322 11.5   79     6  15
47    21     191 14.9   77     6  16
48    37     284 20.7   72     6  17
49    20      37  9.2   65     6  18
50    12     120 11.5   73     6  19
51    13     137 10.3   76     6  20
52     0     150  6.3   77     6  21
53     0      59  1.7   76     6  22
54     0      91  4.6   76     6  23
55     0     250  6.3   76     6  24
56     0     135  8.0   75     6  25
57     0     127  8.0   78     6  26
58     0      47 10.3   73     6  27
59     0      98 11.5   80     6  28
60     0      31 14.9   77     6  29
61     0     138  8.0   83     6  30
62   135     269  4.1   84     7   1
63    49     248  9.2   85     7   2
64    32     236  9.2   81     7   3
65     0     101 10.9   84     7   4
66    64     175  4.6   83     7   5
67    40     314 10.9   83     7   6
68    77     276  5.1   88     7   7
```