R PROGRAMMING & MATLAB LAB

(ACADEMIC YEAR : 2017-2018)

I SEMESTER

ASSIGNMENT 1

ASHISH CHANDRAKANT DUSANE

M. TECH. (ACDS)

COMPUTER ENGG. DEPARTMENT

{ PRN : 170101261004 }

Section I : Importing Files to R :-

1. Import the ClinicalTrail dataset (data is shared) and find the following details.
After importing go through the dataset and see all column names.
For Import :

```
> Clinical<-read.table("/home/student/Downloads/Lab/ClinicalTrail.txt",header = TRUE)
> Clinical
```

i. How many subjects (patients) are enrolled at each center in a clinical trial.

```
> table(Clinical$center)
Center A Center B Center C Center D Center E
      15       18       26       16       25
```

ii. How many subjects (patients) are under the age of 60 in a clinical trial.

```
> Clinical[Clinical$age <60 ,]
   patient      age treatment   center
1        1 48.99590 Treatment Center E
3        3 56.49530 Treatment Center D
6        6 58.86249 Treatment Center A
9        9 54.17409 Treatment Center E
10      10 59.72465 Treatment Center B
11      11 59.76640 Treatment Center E
14      14 55.34836 Treatment Center B
16      16 55.64419 Treatment Center D
22      22 46.22340 Treatment Center C
24      24 58.90856 Treatment Center E
27      27 50.39020 Treatment Center C
28      28 57.67435 Treatment Center D
32      32 52.71358 Treatment Center D
35      35 53.66986 Treatment Center B
41      41 58.73375 Treatment Center E
43      43 54.74381 Treatment Center E
48      48 56.04094 Treatment Center A
49      49 58.99353 Treatment Center C
52      52 56.97833   Control Center A
53      53 52.57280   Control Center E
57      57 55.20245   Control Center B
58      58 58.80887   Control Center E
62      62 54.57935   Control Center B
63      63 57.19065   Control Center C
64      64 59.28596   Control Center B
65      65 56.40973   Control Center D
66      66 56.03066   Control Center C
67      67 50.73804   Control Center B
69      69 53.72042   Control Center D
```

```
70       70 51.06824   Control Center B
72       72 46.41414   Control Center D
73       73 52.37945   Control Center B
74       74 58.60910   Control Center A
76       76 55.28715   Control Center C
77       77 53.71898   Control Center A
79       79 51.71612   Control Center A
81       81 58.13902   Control Center E
82       82 59.14607   Control Center C
83       83 54.71910   Control Center E
86       86 54.49040   Control Center B
87       87 56.14297   Control Center C
88       88 56.34445   Control Center E
89       89 49.85930   Control Center D
91       91 57.29272   Control Center E
92       92 52.86771   Control Center C
96       96 51.64860   Control Center C
98       98 53.75690   Control Center B
99       99 59.63808   Control Center E
```

iii. Which center has the most subjects with a missing value for age in the clinical trial?

```
> table(is.na(Clinical$age))

FALSE
  100
```

2. Import dataset access_log file (file is shared) and in to R. Import only and make Date, Month, Year, Hours, Min and Seconds columns

```
> access<-read.fwf("D:/Lab/Lab/access_log",width = c(-18,2,-1,3,-1,4,-1,2,-1,2,-1,2),col.names = c("Date","Months","Year","Hours","Mins","Seconds"))
> access
    Date Months Year Hours Mins Seconds
1      7    Mar 2004    16    5      49
2      7    Mar 2004    16    6      51
3      7    Mar 2004    16   10       2
4      7    Mar 2004    16   11      58
5      7    Mar 2004    16   20      55
6      7    Mar 2004    16   23      12
7      7    Mar 2004    16   24      16
8      7    Mar 2004    16   29      16
9      7    Mar 2004    16   30      29
10     7    Mar 2004    16   31      48
11     7    Mar 2004    16   32      50
12     7    Mar 2004    16   33      53
13     7    Mar 2004    16   35      19
14     7    Mar 2004    16   36      22
15     7    Mar 2004    16   37      27
16     7    Mar 2004    16   39      24
```

17	7	Mar 2004	16	43	54
18	7	Mar 2004	16	45	56
19	7	Mar 2004	16	47	12
20	7	Mar 2004	16	47	46
21	7	Mar 2004	16	49	4
22	7	Mar 2004	16	50	54
23	7	Mar 2004	16	52	35
24	7	Mar 2004	16	53	46
25	7	Mar 2004	16	54	55
26	7	Mar 2004	16	56	39
27	7	Mar 2004	16	58	54
28	7	Mar 2004	17	9	1
29	7	Mar 2004	17	10	20
30	7	Mar 2004	17	13	50
31	7	Mar 2004	17	16	0
32	7	Mar 2004	17	17	27
33	7	Mar 2004	17	21	44
34	7	Mar 2004	17	22	49
35	7	Mar 2004	17	23	54
36	7	Mar 2004	17	26	30
37	7	Mar 2004	17	27	37
38	7	Mar 2004	17	28	45
39	7	Mar 2004	17	29	59
40	7	Mar 2004	17	31	39
41	7	Mar 2004	17	35	35
42	7	Mar 2004	17	39	39
43	7	Mar 2004	17	42	15
44	7	Mar 2004	17	46	17
45	7	Mar 2004	17	47	43
46	7	Mar 2004	17	50	44
47	7	Mar 2004	17	53	45
48	7	Mar 2004	17	56	54
49	7	Mar 2004	17	58	0
50	7	Mar 2004	18	0	9
51	7	Mar 2004	18	2	10
52	7	Mar 2004	18	4	5
53	7	Mar 2004	18	9	0
54	7	Mar 2004	18	10	9
55	7	Mar 2004	18	17	26
56	7	Mar 2004	18	19	1
57	7	Mar 2004	18	22	52
58	7	Mar 2004	18	26	32
59	7	Mar 2004	18	32	39
60	7	Mar 2004	18	34	42
61	7	Mar 2004	18	42	29
62	7	Mar 2004	18	46	0
63	7	Mar 2004	18	47	6
64	7	Mar 2004	18	48	15
65	7	Mar 2004	18	52	30
66	7	Mar 2004	18	53	55
67	7	Mar 2004	18	57	7
68	7	Mar 2004	18	58	52
69	7	Mar 2004	18	59	52
70	7	Mar 2004	19	1	48
71	7	Mar 2004	19	3	58
72	7	Mar 2004	19	8	55
73	7	Mar 2004	19	10	13

74	7	Mar 2004	19	15	38
75	7	Mar 2004	19	16	44
76	7	Mar 2004	19	18	5
77	7	Mar 2004	19	19	19
78	7	Mar 2004	19	21	1
79	7	Mar 2004	19	22	11
80	7	Mar 2004	19	24	57
81	7	Mar 2004	19	26	22
82	7	Mar 2004	19	29	46
83	7	Mar 2004	19	31	25
84	7	Mar 2004	19	32	45
85	7	Mar 2004	19	36	14
86	7	Mar 2004	19	39	40
87	7	Mar 2004	19	41	33
88	7	Mar 2004	19	42	45
89	7	Mar 2004	19	49	28
90	7	Mar 2004	19	52	28
91	7	Mar 2004	19	54	33
92	7	Mar 2004	19	55	40
93	7	Mar 2004	19	56	41
94	7	Mar 2004	19	58	24
95	7	Mar 2004	20	0	6
96	7	Mar 2004	20	2	13
97	7	Mar 2004	20	3	29
98	7	Mar 2004	20	4	35
99	7	Mar 2004	20	7	12
100	7	Mar 2004	20	11	33
101	7	Mar 2004	20	12	55
102	7	Mar 2004	20	23	35
103	7	Mar 2004	20	25	31
104	7	Mar 2004	20	31	40
105	7	Mar 2004	20	35	28
106	7	Mar 2004	20	38	14
107	7	Mar 2004	20	40	41
108	7	Mar 2004	20	42	9
109	7	Mar 2004	20	44	48
110	7	Mar 2004	20	55	43
111	7	Mar 2004	20	56	56
112	7	Mar 2004	20	58	27
113	7	Mar 2004	21	3	48
114	7	Mar 2004	21	6	5
115	7	Mar 2004	21	7	24
116	7	Mar 2004	21	14	32
117	7	Mar 2004	21	20	14
118	7	Mar 2004	21	21	40
119	7	Mar 2004	21	23	38
120	7	Mar 2004	21	31	12
121	7	Mar 2004	21	33	51
122	7	Mar 2004	21	39	55
123	7	Mar 2004	21	41	4
124	7	Mar 2004	21	42	47
125	7	Mar 2004	21	44	10
126	7	Mar 2004	21	50	22
127	7	Mar 2004	21	52	5
128	7	Mar 2004	22	3	19
129	7	Mar 2004	22	4	44
130	7	Mar 2004	22	6	16

```
131     7    Mar 2004     22     7       33
132     7    Mar 2004     22     8       43
133     7    Mar 2004     22     9       44
134     7    Mar 2004     22    10       55
135     7    Mar 2004     22    12       28
136     7    Mar 2004     22    15       57
137     7    Mar 2004     22    17       40
138     7    Mar 2004     22    27       18
139     7    Mar 2004     22    29       10
140     7    Mar 2004     22    31       25
141     7    Mar 2004     22    35       53
142     7    Mar 2004     22    36       58
143     7    Mar 2004     22    39        0
144     7    Mar 2004     22    45       46
145     7    Mar 2004     22    47       19
146     7    Mar 2004     22    48       55
147     7    Mar 2004     22    51       55
148     7    Mar 2004     22    53       36
149     7    Mar 2004     22    54       43
150     7    Mar 2004     22    58       24
151     7    Mar 2004     23     9        7
152     7    Mar 2004     23    10       44
153     7    Mar 2004     23    13       51
154     7    Mar 2004     23    15       51
155     7    Mar 2004     23    16       57
156     7    Mar 2004     23    19        1
157     7    Mar 2004     23    20       26
158     7    Mar 2004     23    23        0
159     7    Mar 2004     23    27       26
160     7    Mar 2004     23    30       23
161     7    Mar 2004     23    34       31
162     7    Mar 2004     23    36       48
163     7    Mar 2004     23    37       48
164     7    Mar 2004     23    42       44
165     7    Mar 2004     23    47       58
166     7    Mar 2004     23    50        3
```

3. Import dataset error_log file (file is shared) into R and add suitable column names to it.

```
> access<-read.fwf("D:/Lab/Lab/error_log",width = c(-1,3,-1,3,-2,1,-1,
2,-1,2,-1,2,-1,4),col.names = c("Day","Months","Date","Hours","Mins","
Seconds","Year"))
> access
     Day Months Date Hours Mins Seconds Year
1    Sun    Mar    7    16    2       0 2004
2   <NA>   <NA>   NA    NA   NA      NA   NA
3    Sun    Mar    7    16    2       0 2004
4   <NA>   <NA>   NA    NA   NA      NA   NA
5    Sun    Mar    7    16    2       0 2004
6   <NA>   <NA>   NA    NA   NA      NA   NA
7    Sun    Mar    7    16    5      49 2004
8   <NA>   <NA>   NA    NA   NA      NA   NA
```

9	Sun	Mar	7	16	45	56	2004
10	<NA>	<NA>	NA	NA	NA	NA	NA
11	Sun	Mar	7	17	13	50	2004
12	<NA>	<NA>	NA	NA	NA	NA	NA
13	Sun	Mar	7	17	21	44	2004
14	<NA>	<NA>	NA	NA	NA	NA	NA
15	Sun	Mar	7	17	23	53	2004
16	<NA>	<NA>	NA	NA	NA	NA	NA
17	Sun	Mar	7	17	23	53	2004
18	<NA>	<NA>	NA	NA	NA	NA	NA
19	Sun	Mar	7	17	27	37	2004
20	<NA>	<NA>	NA	NA	NA	NA	NA
21	Sun	Mar	7	17	31	39	2004
22	<NA>	<NA>	NA	NA	NA	NA	NA
23	Sun	Mar	7	17	58	0	2004
24	<NA>	<NA>	NA	NA	NA	NA	NA
25	Sun	Mar	7	18	0	9	2004
26	<NA>	<NA>	NA	NA	NA	NA	NA
27	Sun	Mar	7	18	10	9	2004
28	<NA>	<NA>	NA	NA	NA	NA	NA
29	Sun	Mar	7	18	19	1	2004
30	<NA>	<NA>	NA	NA	NA	NA	NA
31	Sun	Mar	7	18	42	29	2004
32	<NA>	<NA>	NA	NA	NA	NA	NA
33	Sun	Mar	7	18	52	30	2004
34	<NA>	<NA>	NA	NA	NA	NA	NA
35	Sun	Mar	7	18	58	52	2004
36	<NA>	<NA>	NA	NA	NA	NA	NA
37	Sun	Mar	7	19	3	58	2004
38	<NA>	<NA>	NA	NA	NA	NA	NA
39	Sun	Mar	7	19	8	55	2004
40	<NA>	<NA>	NA	NA	NA	NA	NA
41	Sun	Mar	7	19	22	11	2004
42	<NA>	<NA>	NA	NA	NA	NA	NA
43	Sun	Mar	7	19	31	25	2004
44	<NA>	<NA>	NA	NA	NA	NA	NA
45	Sun	Mar	7	19	39	40	2004
46	<NA>	<NA>	NA	NA	NA	NA	NA
47	Sun	Mar	7	19	41	33	2004
48	<NA>	<NA>	NA	NA	NA	NA	NA
49	Sun	Mar	7	19	42	45	2004
50	<NA>	<NA>	NA	NA	NA	NA	NA
51	Sun	Mar	7	20	2	13	2004
52	<NA>	<NA>	NA	NA	NA	NA	NA
53	Sun	Mar	7	20	4	35	2004
54	<NA>	<NA>	NA	NA	NA	NA	NA
55	Sun	Mar	7	20	11	33	2004
56	<NA>	<NA>	NA	NA	NA	NA	NA
57	Sun	Mar	7	20	12	55	2004
58	<NA>	<NA>	NA	NA	NA	NA	NA
59	Sun	Mar	7	20	25	31	2004
60	<NA>	<NA>	NA	NA	NA	NA	NA
61	Sun	Mar	7	20	44	48	2004
62	<NA>	<NA>	NA	NA	NA	NA	NA
63	Sun	Mar	7	20	58	27	2004
64	<NA>	<NA>	NA	NA	NA	NA	NA
65	Sun	Mar	7	21	16	17	2004

66	<NA>	<NA>	NA	NA	NA	NA	NA
67	Sun	Mar	7	21	20	14	2004
68	<NA>	<NA>	NA	NA	NA	NA	NA
69	Sun	Mar	7	21	31	12	2004
70	<NA>	<NA>	NA	NA	NA	NA	NA
71	Sun	Mar	7	21	39	55	2004
72	<NA>	<NA>	NA	NA	NA	NA	NA
73	Sun	Mar	7	21	44	10	2004
74	<NA>	<NA>	NA	NA	NA	NA	NA
75	Sun	Mar	7	22	6	16	2004
76	<NA>	<NA>	NA	NA	NA	NA	NA
77	Sun	Mar	7	22	8	43	2004
78	<NA>	<NA>	NA	NA	NA	NA	NA
79	Sun	Mar	7	22	9	44	2004
80	<NA>	<NA>	NA	NA	NA	NA	NA
81	Sun	Mar	7	22	12	28	2004
82	<NA>	<NA>	NA	NA	NA	NA	NA
83	Sun	Mar	7	22	27	18	2004
84	<NA>	<NA>	NA	NA	NA	NA	NA
85	Sun	Mar	7	22	45	46	2004
86	<NA>	<NA>	NA	NA	NA	NA	NA
87	Sun	Mar	7	23	30	23	2004
88	<NA>	<NA>	NA	NA	NA	NA	NA
89	Sun	Mar	7	23	42	44	2004
90	<NA>	<NA>	NA	NA	NA	NA	NA
91	Mon	Mar	8	0	11	22	2004
92	<NA>	<NA>	NA	NA	NA	NA	NA
93	Mon	Mar	8	0	32	45	2004
94	<NA>	<NA>	NA	NA	NA	NA	NA
95	Mon	Mar	8	0	40	10	2004
96	<NA>	<NA>	NA	NA	NA	NA	NA
97	Mon	Mar	8	1	4	5	2004
98	<NA>	<NA>	NA	NA	NA	NA	NA
99	Mon	Mar	8	1	19	18	2004
100	<NA>	<NA>	NA	NA	NA	NA	NA
101	Mon	Mar	8	1	35	13	2004
102	<NA>	<NA>	NA	NA	NA	NA	NA
103	Mon	Mar	8	1	47	6	2004
104	<NA>	<NA>	NA	NA	NA	NA	NA
105	Mon	Mar	8	1	59	13	2004
106	<NA>	<NA>	NA	NA	NA	NA	NA
107	Mon	Mar	8	2	12	24	2004
108	<NA>	<NA>	NA	NA	NA	NA	NA
109	Mon	Mar	8	2	54	54	2004
110	<NA>	<NA>	NA	NA	NA	NA	NA
111	Mon	Mar	8	3	46	27	2004
112	<NA>	<NA>	NA	NA	NA	NA	NA
113	Mon	Mar	8	3	48	18	2004
114	<NA>	<NA>	NA	NA	NA	NA	NA
115	Mon	Mar	8	3	52	17	2004
116	<NA>	<NA>	NA	NA	NA	NA	NA
117	Mon	Mar	8	3	55	9	2004
118	<NA>	<NA>	NA	NA	NA	NA	NA
119	Mon	Mar	8	4	22	55	2004
120	<NA>	<NA>	NA	NA	NA	NA	NA
121	Mon	Mar	8	4	24	47	2004
122	<NA>	<NA>	NA	NA	NA	NA	NA

```
123  Mon    Mar    8     4    40       32 2004
124 <NA>   <NA>   NA    NA   NA       NA   NA
125  Mon    Mar    8     4    55       40 2004
126 <NA>   <NA>   NA    NA   NA       NA   NA
127  Mon    Mar    8     4    59       13 2004
128 <NA>   <NA>   NA    NA   NA       NA   NA
129  Mon    Mar    8     5    22       57 2004
130 <NA>   <NA>   NA    NA   NA       NA   NA
131  Mon    Mar    8     5    24       29 2004
132 <NA>   <NA>   NA    NA   NA       NA   NA
133  Mon    Mar    8     5    31       47 2004
134 <NA>   <NA>   NA    NA   NA       NA   NA
135  Mon    Mar    8     6    23       52 2004
136 <NA>   <NA>   NA    NA   NA       NA   NA
137  Mon    Mar    8     6    43       32 2004
138 <NA>   <NA>   NA    NA   NA       NA   NA
139  Mon    Mar    8     6    49       27 2004
140 <NA>   <NA>   NA    NA   NA       NA   NA
141  Mon    Mar    8     7     7       13 2004
142 <NA>   <NA>   NA    NA   NA       NA   NA
 [ reached getOption("max.print") -- omitted 73 rows ]
```

4. Import input.json file (file is shared) into R and convert it into dataframe and store it in object.

```
> json_data <- fromJSON(file="D:/Lab/Lab/input.json")
> json_data
$ID
[1] "1" "2" "3" "4" "5" "6" "7" "8"

$Name
[1] "Rick"     "Dan"      "Michelle" "Ryan"     "Gary"     "Nina"     "
Simon"
[8] "Guru"

$Salary
[1] "623.3"  "515.2"  "611"    "729"    "843.25" "578"    "632.8"  "72
2.5"

$StartDate
[1] "1/1/2012"   "9/23/2013"  "11/15/2014" "5/11/2014"  "3/27/2015"  "
5/21/2013"
[7] "7/30/2013"  "6/17/2014"

$Dept
[1] "IT"         "Operations" "IT"         "HR"         "Finance"    "
IT"
[7] "Operations" "Finance"


> json<-as.data.frame(json_data)
> json
  ID     Name Salary  StartDate       Dept
1  1     Rick  623.3   1/1/2012         IT
2  2      Dan  515.2  9/23/2013 Operations
```

```
3  3 Michelle    611 11/15/2014         IT
4  4     Ryan    729  5/11/2014         HR
5  5     Gary 843.25  3/27/2015    Finance
6  6     Nina    578  5/21/2013         IT
7  7    Simon  632.8  7/30/2013 Operations
8  8     Guru  722.5  6/17/2014    Finance
```

i. Extract the names starting with alphabet "R" and whose salary is less than 500.

ii. Extract employee information whose salary is less than 800 and working in IT and Operations Dept.

6. Consider a dataset iris (in-built dataset) and store it in different object.
Find the following details.
i. How many variety of species are there and at frequency those species are appearing the dataset.

```
> table(iris$Species)

    setosa versicolor  virginica
        50         50         50
```

ii. Find the species name whose "Sepal.Length" is equal to 6.9 and at what frequency those species are appearing in the dataset.

```
> library(dplyr)

> filter(iris,iris$Sepal.Length==6.9)
  Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1          6.9         3.1          4.9         1.5 versicolor
2          6.9         3.2          5.7         2.3  virginica
3          6.9         3.1          5.4         2.1  virginica
4          6.9         3.1          5.1         2.3  virginica

> table(iris$Sepal.Length==6.9)
FALSE  TRUE
  146     4
```

7. Consider a dataset mtcars (in-built dataset) and store it in different object. Second column in the dataset is "mpg" which is Miles/(US) gallon. Add an extra column called "Condition" to the mtcars and fill the values as follows (Example is shown below)
i. Multiply mpg and cyl

```
> cars$Condition<-cars$mpg*cars$cyl
> cars
                  mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4        21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag    21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
```

Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

	Condition
Mazda RX4	126.0
Mazda RX4 Wag	126.0
Datsun 710	91.2
Hornet 4 Drive	128.4
Hornet Sportabout	149.6
Valiant	108.6
Duster 360	114.4
Merc 240D	97.6
Merc 230	91.2
Merc 280	115.2
Merc 280C	106.8
Merc 450SE	131.2
Merc 450SL	138.4
Merc 450SLC	121.6
Cadillac Fleetwood	83.2
Lincoln Continental	83.2
Chrysler Imperial	117.6
Fiat 128	129.6
Honda Civic	121.6
Toyota Corolla	135.6
Toyota Corona	86.0
Dodge Challenger	124.0
AMC Javelin	121.6
Camaro Z28	106.4

```
Pontiac Firebird        153.6
Fiat X1-9               109.2
Porsche 914-2           104.0
Lotus Europa            121.6
Ford Pantera L          126.4
Ferrari Dino            118.2
Maserati Bora           120.0
Volvo 142E               85.6
```

9. After adding an extra column, export it to local drive in both csv and text files

For Text Files :-

```
> write.table(cars,"D:/Lab/Lab/Condition.txt")

> car<-read.table("D:/Lab/Lab/condition.txt",header=TRUE)
> car
                   mpg cyl  disp  hp drat    wt  qsec vs am gear car
b Condition
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
126.0
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
126.0
Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
91.2
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
128.4
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
149.6
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
108.6
Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
114.4
Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
97.6
Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
91.2
Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
115.2
Merc 280C         17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
106.8
Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
131.2
Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
138.4
Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
121.6
Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
83.2
Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
83.2
Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
117.6
```

```
Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
129.6
Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
121.6
Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
135.6
Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
86.0
Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
124.0
AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
121.6
Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
106.4
Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
153.6
Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
109.2
Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
104.0
Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
121.6
Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
126.4
Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
118.2
Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
120.0
Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
85.6
```

For CSV Files :-

```
> write.csv(cars,"D:/Lab/Lab/conditioncsv.txt")
> carcsv<-read.csv("D:/Lab/Lab/conditioncsv.txt",header=TRUE)
> carcsv
```

```
                   X  mpg cyl  disp  hp drat    wt  qsec vs am gear c
arb
1          Mazda RX4 21.0   6 160.0 110 3.90 2.620 16.46  0  1    4
4
2      Mazda RX4 Wag 21.0   6 160.0 110 3.90 2.875 17.02  0  1    4
4
3         Datsun 710 22.8   4 108.0  93 3.85 2.320 18.61  1  1    4
1
4     Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44  1  0    3
1
5  Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3
2
6            Valiant 18.1   6 225.0 105 2.76 3.460 20.22  1  0    3
1
7         Duster 360 14.3   8 360.0 245 3.21 3.570 15.84  0  0    3
4
```

```
8             Merc 240D 24.4   4 146.7  62 3.69 3.190 20.00  1  0     4
2
9              Merc 230 22.8   4 140.8  95 3.92 3.150 22.90  1  0     4
2
10             Merc 280 19.2   6 167.6 123 3.92 3.440 18.30  1  0     4
4
11            Merc 280C 17.8   6 167.6 123 3.92 3.440 18.90  1  0     4
4
12           Merc 450SE 16.4   8 275.8 180 3.07 4.070 17.40  0  0     3
3
13           Merc 450SL 17.3   8 275.8 180 3.07 3.730 17.60  0  0     3
3
14          Merc 450SLC 15.2   8 275.8 180 3.07 3.780 18.00  0  0     3
3
15  Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0     3
4
16 Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0     3
4
17   Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42  0  0     3
4
18            Fiat 128 32.4   4  78.7  66 4.08 2.200 19.47  1  1     4
1
19         Honda Civic 30.4   4  75.7  52 4.93 1.615 18.52  1  1     4
2
20      Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90  1  1     4
1
21       Toyota Corona 21.5   4 120.1  97 3.70 2.465 20.01  1  0     3
1
22    Dodge Challenger 15.5   8 318.0 150 2.76 3.520 16.87  0  0     3
2
23         AMC Javelin 15.2   8 304.0 150 3.15 3.435 17.30  0  0     3
2
24          Camaro Z28 13.3   8 350.0 245 3.73 3.840 15.41  0  0     3
4
25    Pontiac Firebird 19.2   8 400.0 175 3.08 3.845 17.05  0  0     3
2
26            Fiat X1-9 27.3   4  79.0  66 4.08 1.935 18.90  1  1     4
1
27       Porsche 914-2 26.0   4 120.3  91 4.43 2.140 16.70  0  1     5
2
28        Lotus Europa 30.4   4  95.1 113 3.77 1.513 16.90  1  1     5
2
29      Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.50  0  1     5
4
30        Ferrari Dino 19.7   6 145.0 175 3.62 2.770 15.50  0  1     5
6
31       Maserati Bora 15.0   8 301.0 335 3.54 3.570 14.60  0  1     5
8
32          Volvo 142E 21.4   4 121.0 109 4.11 2.780 18.60  1  1     4
2

   Condition
1      126.0
2      126.0
3       91.2
4      128.4
5      149.6
```

```
6       108.6
7       114.4
8        97.6
9        91.2
10      115.2
11      106.8
12      131.2
13      138.4
14      121.6
15       83.2
16       83.2
17      117.6
18      129.6
19      121.6
20      135.6
21       86.0
22      124.0
23      121.6
24      106.4
25      153.6
26      109.2
27      104.0
28      121.6
29      126.4
30      118.2
31      120.0
32       85.6
```

10. Import dataset bnames2 (dataset is shared), which consists of top 1000 male and female baby names in US from 1800 to 2008. Dataset is 258000 x5.

```
> babyname<-read.csv("D:/Lab/Lab/bnames2.csv",header=TRUE)
> babyname
       year        name     prop  sex soundex
1      1880        John 0.081541  boy    J500
2      1880     William 0.080511  boy    W450
3      1880       James 0.050057  boy    J520
4      1880     Charles 0.045167  boy    C642
5      1880      George 0.043292  boy    G620
6      1880       Frank 0.027380  boy    F652
7      1880      Joseph 0.022229  boy    J210
8      1880      Thomas 0.021401  boy    T520
9      1880       Henry 0.020641  boy    H560
10     1880      Robert 0.020404  boy    R163
11     1880      Edward 0.019965  boy    E363
12     1880       Harry 0.018175  boy    H600
13     1880      Walter 0.014822  boy    W436
14     1880      Arthur 0.013504  boy    A636
15     1880        Fred 0.013251  boy    F630
16     1880      Albert 0.012609  boy    A416
17     1880      Samuel 0.008648  boy    S540
18     1880       David 0.007339  boy    D130
19     1880       Louis 0.006993  boy    L200
20     1880         Joe 0.006174  boy    J000
21     1880     Charlie 0.006165  boy    C640
22     1880    Clarence 0.006165  boy    C465
```

23	1880	Richard	0.006148	boy	R263
24	1880	Andrew	0.005439	boy	A536
25	1880	Daniel	0.005430	boy	D540
26	1880	Ernest	0.005194	boy	E652
27	1880	Will	0.004966	boy	W400
28	1880	Jesse	0.004805	boy	J200
29	1880	Oscar	0.004594	boy	O260
30	1880	Lewis	0.004366	boy	L200
31	1880	Peter	0.004189	boy	P360
32	1880	Benjamin	0.004138	boy	B525
33	1880	Frederick	0.004079	boy	F636
34	1880	Willie	0.004020	boy	W400
35	1880	Alfred	0.003961	boy	A416
36	1880	Sam	0.003860	boy	S500
37	1880	Roy	0.003716	boy	R000
38	1880	Herbert	0.003581	boy	H616
39	1880	Jacob	0.003412	boy	J210
40	1880	Tom	0.003370	boy	T500
41	1880	Elmer	0.003150	boy	E456
42	1880	Carl	0.003142	boy	C640
43	1880	Lee	0.003049	boy	L000
44	1880	Howard	0.003015	boy	H630
45	1880	Martin	0.003015	boy	M635
46	1880	Michael	0.002990	boy	M240
47	1880	Bert	0.002939	boy	B630
48	1880	Herman	0.002931	boy	H655
49	1880	Jim	0.002914	boy	J500
50	1880	Francis	0.002905	boy	F652
51	1880	Harvey	0.002905	boy	H610
52	1880	Earl	0.002829	boy	E640
53	1880	Eugene	0.002770	boy	E250
54	1880	Ralph	0.002677	boy	R410
55	1880	Ed	0.002618	boy	E300
56	1880	Claude	0.002610	boy	C430
57	1880	Edwin	0.002610	boy	E350
58	1880	Ben	0.002576	boy	B500
59	1880	Charley	0.002576	boy	C640
60	1880	Paul	0.002542	boy	P400
61	1880	Edgar	0.002390	boy	E326
62	1880	Isaac	0.002314	boy	I220
63	1880	Otto	0.002289	boy	O300
64	1880	Luther	0.002196	boy	L360
65	1880	Lawrence	0.002170	boy	L652
66	1880	Ira	0.002103	boy	I600
67	1880	Patrick	0.002094	boy	P362
68	1880	Guy	0.002018	boy	G000
69	1880	Oliver	0.001976	boy	O416
70	1880	Theodore	0.001959	boy	T360
71	1880	Hugh	0.001892	boy	H200
72	1880	Clyde	0.001866	boy	C430
73	1880	Alexander	0.001782	boy	A425
74	1880	August	0.001774	boy	A223
75	1880	Floyd	0.001740	boy	F430
76	1880	Homer	0.001731	boy	H560
77	1880	Jack	0.001723	boy	J200
78	1880	Leonard	0.001689	boy	L563
79	1880	Horace	0.001681	boy	H620

80	1880	Marion	0.001596	boy	M650
81	1880	Philip	0.001571	boy	P410
82	1880	Allen	0.001554	boy	A450
83	1880	Archie	0.001546	boy	A620
84	1880	Stephen	0.001486	boy	S315
85	1880	Chester	0.001419	boy	C236
86	1880	Willis	0.001402	boy	W420
87	1880	Raymond	0.001394	boy	R553
88	1880	Rufus	0.001377	boy	R120
89	1880	Warren	0.001334	boy	W650
90	1880	Jessie	0.001301	boy	J200
91	1880	Milton	0.001258	boy	M435
92	1880	Alex	0.001241	boy	A420
93	1880	Leo	0.001241	boy	L000
94	1880	Julius	0.001208	boy	J420
95	1880	Ray	0.001199	boy	R000
96	1880	Sidney	0.001199	boy	S350
97	1880	Bernard	0.001182	boy	B656
98	1880	Dan	0.001182	boy	D500
99	1880	Jerry	0.001149	boy	J600
100	1880	Calvin	0.001132	boy	C415
101	1880	Perry	0.001132	boy	P600
102	1880	Dave	0.001106	boy	D100
103	1880	Anthony	0.001098	boy	A535
104	1880	Eddie	0.001089	boy	E300
105	1880	Amos	0.001081	boy	A520
106	1880	Dennis	0.001081	boy	D520
107	1880	Clifford	0.001073	boy	C416
108	1880	Leroy	0.001047	boy	L600
109	1880	Wesley	0.001039	boy	W240
110	1880	Alonzo	0.001030	boy	A452
111	1880	Garfield	0.001030	boy	G614
112	1880	Franklin	0.001013	boy	F652
113	1880	Emil	0.001005	boy	E540
114	1880	Leon	0.000997	boy	L500
115	1880	Nathan	0.000963	boy	N350
116	1880	Harold	0.000954	boy	H643
117	1880	Matthew	0.000954	boy	M300
118	1880	Levi	0.000946	boy	L100
119	1880	Moses	0.000937	boy	M220
120	1880	Everett	0.000929	boy	E163
121	1880	Lester	0.000921	boy	L236
122	1880	Winfield	0.000912	boy	W514
123	1880	Adam	0.000878	boy	A350
124	1880	Lloyd	0.000878	boy	L300
125	1880	Mack	0.000878	boy	M200
126	1880	Fredrick	0.000870	boy	F636
127	1880	Jay	0.000870	boy	J000
128	1880	Jess	0.000870	boy	J200
129	1880	Melvin	0.000870	boy	M415
130	1880	Noah	0.000870	boy	N000
131	1880	Aaron	0.000861	boy	A650
132	1880	Alvin	0.000861	boy	A415
133	1880	Norman	0.000861	boy	N655
134	1880	Gilbert	0.000853	boy	G416
135	1880	Elijah	0.000845	boy	E420
136	1880	Victor	0.000845	boy	V236

137	1880	Gus	0.000836	boy	G200
138	1880	Nelson	0.000836	boy	N425
139	1880	Jasper	0.000828	boy	J216
140	1880	Silas	0.000828	boy	S420
141	1880	Christopher	0.000819	boy	C623
142	1880	Jake	0.000811	boy	J200
143	1880	Mike	0.000802	boy	M200
144	1880	Percy	0.000794	boy	P620
145	1880	Adolph	0.000785	boy	A341
146	1880	Maurice	0.000785	boy	M620
147	1880	Cornelius	0.000777	boy	C654
148	1880	Felix	0.000777	boy	F420
149	1880	Reuben	0.000777	boy	R150
150	1880	Wallace	0.000769	boy	W420
151	1880	Claud	0.000760	boy	C430
152	1880	Roscoe	0.000760	boy	R200
153	1880	Sylvester	0.000752	boy	S412
154	1880	Earnest	0.000743	boy	E652
155	1880	Hiram	0.000743	boy	H650
156	1880	Otis	0.000743	boy	O320
157	1880	Simon	0.000743	boy	S550
158	1880	Willard	0.000743	boy	W463
159	1880	Irvin	0.000726	boy	I615
160	1880	Mark	0.000726	boy	M620
161	1880	Jose	0.000709	boy	J200
162	1880	Wilbur	0.000693	boy	W416
163	1880	Abraham	0.000684	boy	A165
164	1880	Virgil	0.000684	boy	V624
165	1880	Clinton	0.000667	boy	C453
166	1880	Elbert	0.000667	boy	E416
167	1880	Leslie	0.000667	boy	L240
168	1880	Marshall	0.000659	boy	M624
169	1880	Owen	0.000659	boy	O500
170	1880	Wiley	0.000659	boy	W400
171	1880	Anton	0.000650	boy	A535
172	1880	Morris	0.000650	boy	M620
173	1880	Manuel	0.000633	boy	M540
174	1880	Phillip	0.000633	boy	P410
175	1880	Augustus	0.000625	boy	A223
176	1880	Emmett	0.000625	boy	E530
177	1880	Eli	0.000617	boy	E400
178	1880	Nicholas	0.000617	boy	N242
179	1880	Wilson	0.000608	boy	W425
180	1880	Alva	0.000591	boy	A410
181	1880	Harley	0.000591	boy	H640
182	1880	Newton	0.000591	boy	N350
183	1880	Timothy	0.000591	boy	T530
184	1880	Marvin	0.000583	boy	M615
185	1880	Ross	0.000583	boy	R200
186	1880	Curtis	0.000574	boy	C632
187	1880	Edmund	0.000566	boy	E355
188	1880	Jeff	0.000557	boy	J100
189	1880	Elias	0.000549	boy	E420
190	1880	Harrison	0.000549	boy	H625
191	1880	Stanley	0.000549	boy	S354
192	1880	Columbus	0.000541	boy	C451
193	1880	Lon	0.000541	boy	L500

```
194    1880         Ora 0.000541   boy    O600
195    1880      Ollie 0.000532   boy    O400
196    1880     Russell 0.000532   boy    R240
197    1880      Pearl 0.000524   boy    P640
198    1880     Solomon 0.000524   boy    S455
199    1880       Arch 0.000515   boy    A620
200    1880        Asa 0.000507   boy    A200

 [ reached getOption("max.print") -- omitted 257800 rows ]
```

Section II : Merging data frames :~

Using merge function to combine two data frames based on column names

11. Create the following dataframes (Dataframe 1, Dataframe 2) as shown below.

```
> dataframe1<-data.frame(Died.At=c(22,40,72,41),Writer.At=c(16, 18, 36, 36),First.Name=c("John", "Edgar",
 "Walt", "Jane"),Second.Name=c("Doe", "Poe", "Whitman", "Austen"),Sex=c("MALE", "MALE", "MALE", "FEMALE")
,Date.Of.Death=c("2015-05-10", "1849-10-07", "1892-03-26","1817-07-18"))
> dataframe1
  Died.At Writer.At First.Name Second.Name    Sex Date.Of.Death
1      22        16       John         Doe   MALE    2015-05-10
2      40        18      Edgar         Poe   MALE    1849-10-07
3      72        36       Walt     Whitman   MALE    1892-03-26
4      41        36       Jane      Austen FEMALE    1817-07-18
>
```

```
> dataframe2 <- data.frame(Age.At.Death=c(22,40,72,41), Location=5:8)
> dataframe2
  Age.At.Death Location
1           22        5
2           40        6
3           72        7
4           41        8
```

12. Extract the names of Dataframe 1 and Dataframe 2 created in the previous question

```
> colnames(dataframe1)
 [1] "Died.At"                    "Writer.At"                  "First.Name"
"Second.Name"
 [5] "Sex"                        "Date.Of.Death"              "Writers_df.
Died.At"       "Writers_df.Writer.At"
 [9] "Writers_df.First.Name"     "Writers_df.Second.Name"    "Writers_df
.Sex"            "Writers_df.Date.Of.Death"
```

13. Change the names of Dataframe 1. (Use indexing). Here we are changing the column name in order to make column names same in both the vector.

```
>  colnames(dataframe1)<-c("Age.At.Death","Writer.At","First.Name","Sec
ond.Name","Sex","Date.Of.Death")
> dataframe1
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death
1           22        16       John          Doe   MALE    2015-05-10
2           40        18      Edgar          Poe   MALE    1849-10-07
3           72        36       Walt     Whitman   MALE    1892-03-26
4           41        36       Jane      Austen FEMALE    1817-07-18
```

14. We can also merge the dataframes without changing the column name also.

```
> merge(dataframe1,dataframe2, by="Age.At.Death")
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death L
ocation
1           22        16       John          Doe   MALE    2015-05-10
5
2           40        18      Edgar          Poe   MALE    1849-10-07
6
3           41        36       Jane      Austen FEMALE    1817-07-18
8
4           72        36       Walt     Whitman   MALE    1892-03-26
7
```

15. Write a command to merge the above two dataframes. Use by.x and by.y arguments separately and check the output.

```
> merge(dataframe1,dataframe2, by="Age.At.Death")
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death L
ocation
1           22        16       John          Doe   MALE    2015-05-10
5
2           40        18      Edgar          Poe   MALE    1849-10-07
6
3           41        36       Jane      Austen FEMALE    1817-07-18
8
4           72        36       Walt     Whitman   MALE    1892-03-26
7
```

16. Reduce the age by 1 in each row of Dataframe 1

```
> dataframe1
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death
1           22        16       John          Doe   MALE    2015-05-10
2           40        18      Edgar          Poe   MALE    1849-10-07
3           72        36       Walt     Whitman   MALE    1892-03-26
4           41        36       Jane      Austen FEMALE    1817-07-18
```

```
> dataframe1$Age.At.Death<-dataframe1$Age.At.Death-1
> dataframe1
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death
1           21        16       John          Doe   MALE    2015-05-10
2           39        18      Edgar          Poe   MALE    1849-10-07
3           71        36       Walt     Whitman   MALE    1892-03-26
4           40        36       Jane      Austen FEMALE    1817-07-18
```

17. Add the extra column to the Dataframe 1. Column name should be ID and the values are 1:4

```
> dataframe1$ID<-c(1:4)
> dataframe1
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death I
D
1           21        16       John          Doe   MALE    2015-05-10  1
2           39        18      Edgar          Poe   MALE    1849-10-07  2
3           71        36       Walt     Whitman   MALE    1892-03-26  3
4           40        36       Jane      Austen FEMALE    1817-07-18  4
```

18. Change any one value of Age.At.Death in one of the dataframe and trying to merge command and check the output.

```
> dataframe2
  Age.At.Death Location
1           22        5
2           40        6
3           72        7
4           41        8
> dataframe2$Age.At.Death[2]<-dataframe2$Age.At.Death[2]-1

> dataframe2
  Age.At.Death Location
1           22        5
2           39        6
3           72        7
4           41        8

> merge(dataframe1,dataframe2, by="Age.At.Death")
  Age.At.Death Writer.At First.Name Second.Name  Sex Date.Of.Death ID L
ocation
1           39        18      Edgar          Poe MALE    1849-10-07  2
6
```

19. Try the merge command on the datasets created with the argument `all.x=TRUE` and `all.y=TRUE` and check the output

```
> merge(dataframe1,dataframe2, all.x = "TRUE", all.y = "TRUE")
  Age.At.Death Writer.At First.Name Second.Name    Sex Date.Of.Death I
D Location
1           21        16       John          Doe   MALE    2015-05-10  1
NA
```

```
2           22         NA       <NA>        <NA>   <NA>          <NA> NA
5
3           39         18     Edgar         Poe   MALE    1849-10-07  2
6
4           40         36       Jane      Austen FEMALE    1817-07-18  4
NA
5           41         NA       <NA>        <NA>   <NA>          <NA> NA
8
6           71         36       Walt     Whitman   MALE    1892-03-26  3
NA
7           72         NA       <NA>        <NA>   <NA>          <NA> NA
7
```

Using merge function to combine two data frames based on rownames

20. Create the following dataframe.

```
> dataframe3<-data.frame(Address <- c("50 West 10th", "77 St. Marks Pl
ace", "778 Park Avenue"),Maried <- c("YES", "NO", "YES"))

> dataframe3
Address....c..50.West.10th....77.St..Marks.Place....778.Park.Avenue..
Maried....c..YES....NO....YES..
1                                                          50 West 10th
YES
2                                                       77 St. Marks Place
NO
3                                                          778 Park Avenue
YES
```

Apply the merge function on "writers_df" and "limited_writers_df" datasets and check the output. To merge rows, use by=0 argument.

```
> merge(writers_df,limited_writers_df, by=0)
  Row.names Died.At Writer.At First.Name Second.Name  Sex Date.Of.Deat
h            Address Maried
1         1      22        16       John          Doe MALE    2015-05-10
50 West 10th    YES
2         2      40        18      Edgar          Poe MALE    1849-10-07
77 St. Marks Place     NO
3         3      72        36       Walt     Whitman MALE    1892-03-26
778 Par
k Avenue    YES
```

Section III : Transforming data frames :~

21. Create the following dataframe. You can copy paste the commands on R console

```
> companiesData<-data.frame(fy <- c(2010,2011,2012,2010,2011,2012,2010
,2011,2012),company <-c("Apple","Apple","Apple","Google","Google","Goo
gle","Microsoft","Microsoft","Microsoft"),revenue <- c(65225,108249,15
6508,29321,37905,50175,62484,69943,73723),profit <- c(14013,25922,4173
3,8505,9737,10737,18760,23150,16978))
> companiesData
```

```
> companiesData<-data.frame(fy <- c(2010,2011,2012,2010,2011,2012,2010,2011,2012),company <-c("Apple","Apple","Apple","Google","Google"
,"Google","Microsoft","Microsoft","Microsoft"),revenue <- c(65225,108249,156508,29321,37905,50175,62484,69943,73723),profit <- c(14013,
25922,41733,8505,9737,10737,18760,23150,16978))
> companiesData
  fy....c.2010..2011..2012..2010..2011..2012..2010..2011..2012. company....c..Apple....Apple....Apple....Google....Google....Google...
1                                                          2010                                                                  Apple
2                                                          2011                                                                  Apple
3                                                          2012                                                                  Apple
4                                                          2010                                                                 Google
5                                                          2011                                                                 Google
6                                                          2012                                                                 Google
7                                                          2010                                                              Microsoft
8                                                          2011                                                              Microsoft
9                                                          2012                                                              Microsoft
  revenue....c.65225..108249..156508..29321..37905..50175..62484.. profit....c.14013..25922..41733..8505..9737..10737..18760..23150..
1                                                         65225                                                                 14013
2                                                        108249                                                                 25922
3                                                        156508                                                                 41733
4                                                         29321                                                                  8505
5                                                         37905                                                                  9737
6                                                         50175                                                                 10737
7                                                         62484                                                                 18760
8                                                         69943                                                                 23150
9                                                         73723                                                                 16978
>
```

Find the following :~

I. **Transform the data by adding an extra column Margin (You have to frame a formula how margin is calculated)**

```
> companiesData$margin <- (companiesData$profit / companiesData$revenue) * 100
> companiesData
    fy    company revenue profit   margin
1 2010      Apple   65225  14013 21.48409
2 2011      Apple  108249  25922 23.94664
3 2012      Apple  156508  41733 26.66509
4 2010     Google   29321   8505 29.00651
5 2011     Google   37905   9737 25.68790
6 2012     Google   50175  10737 21.39910
7 2010 Microsoft   62484  18760 30.02369
8 2011 Microsoft   69943  23150 33.09838
9 2012 Microsoft   73723  16978 23.02945
>
```

II. **Find which company is having highest margin**

```
> highestM <- companiesData[companiesData$margin == max(companiesData$margin),]
> highestM
    fy   company revenue profit   margin
8 2011 Microsoft   69943  23150 33.09838
>
```

III. Find which company is having lowest margin

```
> LowestM <- companiesData[companiesData$margin == min(companiesData$margin),]
> LowestM
    fy company revenue profit  margin
6 2012  Google   50175  10737 21.3991
>
```

SECTION IV :-

22. Create the following data frame.
Subject <- c(1,2,1,2,2,1)
Gender <- c("M", "F", "M", "F", "F","M")
Test <- c("Read", "Write", "Write", "Listen", "Read", "Listen")
Result <- c(10, 4, 8, 6, 7, 7)

```
> observations_long
  Subject Gender   Test Result
1       1      M   Read     10
2       2      F  Write      4
3       1      M  Write      8
4       2      F Listen      6
5       2      F   Read      7
6       1      M Listen      7
>
```

- **Convert the above data as shown below :**

Subject	Gender	Listen	Read	Write
1	M	7	10	8
2	F	6	7	4

```
> dcast(pew,Subject+Gender~Test)

  Subject Gender Listen Read Write
1       1      M   1529   96    54
2       2      F   1489  116    76
```

23. Import the Pew dataset (It is shared) in table format and go through the data. This dataset is not ready for analysis, columns are of same variables ("Income").

Note: After reshaping the dataset, save the output in different object

We have 11 columns and their names are as follows

```
[1] "religion" "X.10k" "X.10.20k" "X.20.30k" "X.30.40k"
[6]  "X.40.50k"  "X.50.75k"  "X.75.100k"  "X.100.150k"
"X.150k"
[11] "Don.t.know.refused"
```

```
> pew<-read.table("D:/Lab/Lab/pew.txt",header=TRUE)
> pew
```

	religion	X..10k	X.10.20k	X.20.30k	X.30.40k	X.40.50k	X.50.75k	X.75.100k	X.100.150k	X.150k	Don.t.know.refused
1	Agnostic	27	34	60	81	76	137	122	109	84	96
2	Atheist	12	27	37	52	35	70	73	59	74	76
3	Buddhist	27	21	30	34	33	58	62	39	53	54
4	Catholic	418	617	732	670	638	1116	949	792	633	1489
5	Don't know/refused	15	14	15	11	10	35	21	17	18	116
6	Evangelical Prot	575	869	1064	982	881	1486	949	723	414	1529
7	Hindu	1	9	7	9	11	34	47	48	54	37
8	Historically Black Prot	228	244	236	238	197	223	131	81	78	339
9	Jehovah's Witness	20	27	24	24	21	30	15	11	6	37
10	Jewish	19	19	25	25	30	95	69	87	151	162
11	Mainline Prot	289	495	619	655	651	1107	939	753	634	1328
12	Mormon	29	40	48	51	56	112	85	49	42	69
13	Muslim	6	7	9	10	9	23	16	8	6	22
14	Orthodox	13	17	23	32	32	47	38	42	46	73
15	Other Christian	9	7	11	13	13	14	18	14	12	18
16	Other Faiths	20	33	40	46	49	63	46	40	41	71
17	Other World Religions	5	2	3	4	2	7	3	4	4	8
18	Unaffiliated	217	299	374	365	341	528	407	321	258	597

Change the column names from second column to tenth column as shown below

```
[1] "religion" "10k" "10-20k" "20-30k" "30-40k"
[6] "40-50k" "50-75k" "75-100k" "100-150k" "150k"
[11] "Don.t.know.refused"
```

```
> colnames(pew)<-c("religion","10k","10-20k","20-30k","30-40k","40-50k","50-75k","75-100k","100-150k","150k","Don.t.know.refused")
> pew
```

	religion	10k	10-20k	20-30k	30-40k	40-50k	50-75k	75-100k	100-150k	150k	Don.t.know.refused
1	Agnostic	27	34	60	81	76	137	122	109	84	96
2	Atheist	12	27	37	52	35	70	73	59	74	76
3	Buddhist	27	21	30	34	33	58	62	39	53	54
4	Catholic	418	617	732	670	638	1116	949	792	633	1489
5	Don't know/refused	15	14	15	11	10	35	21	17	18	116
6	Evangelical Prot	575	869	1064	982	881	1486	949	723	414	1529
7	Hindu	1	9	7	9	11	34	47	48	54	37
8	Historically Black Prot	228	244	236	238	197	223	131	81	78	339
9	Jehovah's Witness	20	27	24	24	21	30	15	11	6	37
10	Jewish	19	19	25	25	30	95	69	87	151	162
11	Mainline Prot	289	495	619	655	651	1107	939	753	634	1328
12	Mormon	29	40	48	51	56	112	85	49	42	69
13	Muslim	6	7	9	10	9	23	16	8	6	22
14	Orthodox	13	17	23	32	32	47	38	42	46	73
15	Other Christian	9	7	11	13	13	14	18	14	12	18
16	Other Faiths	20	33	40	46	49	63	46	40	41	71
17	Other World Religions	5	2	3	4	2	7	3	4	4	8
18	Unaffiliated	217	299	374	365	341	528	407	321	258	597

Now, put all the columns[2:10] in one column (can be named as Income and it corresponding values).

```
> car<-melt(pew,id.vars = c("religion","Don.t.know.refused"),variable.name = "Income")
> car
```

	religion	Don.t.know.refused	Income	value
1	Agnostic	96	10k	27
2	Atheist	76	10k	12
3	Buddhist	54	10k	27
4	Catholic	1489	10k	418
5	Don’t know/refused	116	10k	15
6	Evangelical Prot	1529	10k	575
7	Hindu	37	10k	1
8	Historically Black Prot	339	10k	228
9	Jehovah's Witness	37	10k	20
10	Jewish	162	10k	19
11	Mainline Prot	1328	10k	289
12	Mormon	69	10k	29
13	Muslim	22	10k	6
14	Orthodox	73	10k	13
15	Other Christian	18	10k	9
16	Other Faiths	71	10k	20
17	Other World Religions	8	10k	5
18	Unaffiliated	597	10k	217
19	Agnostic	96	10-20k	34
20	Atheist	76	10-20k	27
21	Buddhist	54	10-20k	21
22	Catholic	1489	10-20k	617
23	Don’t know/refused	116	10-20k	14
24	Evangelical Prot	1529	10-20k	869
25	Hindu	37	10-20k	9
26	Historically Black Prot	339	10-20k	244
27	Jehovah's Witness	37	10-20k	27
28	Jewish	162	10-20k	19
29	Mainline Prot	1328	10-20k	495
30	Mormon	69	10-20k	40
31	Muslim	22	10-20k	7
32	Orthodox	73	10-20k	17

33	Other Christian	18	10-20k	7
34	Other Faiths	71	10-20k	33
35	Other World Religions	8	10-20k	2
36	Unaffiliated	597	10-20k	299
37	Agnostic	96	20-30k	60
38	Atheist	76	20-30k	37
39	Buddhist	54	20-30k	30
40	Catholic	1489	20-30k	732
41	Don’t know/refused	116	20-30k	15
42	Evangelical Prot	1529	20-30k	1064
43	Hindu	37	20-30k	7
44	Historically Black Prot	339	20-30k	236
45	Jehovah's Witness	37	20-30k	24
46	Jewish	162	20-30k	25
47	Mainline Prot	1328	20-30k	619
48	Mormon	69	20-30k	48
49	Muslim	22	20-30k	9
50	Orthodox	73	20-30k	23
51	Other Christian	18	20-30k	11
52	Other Faiths	71	20-30k	40
53	Other World Religions	8	20-30k	3
54	Unaffiliated	597	20-30k	374
55	Agnostic	96	30-40k	81
56	Atheist	76	30-40k	52
57	Buddhist	54	30-40k	34
58	Catholic	1489	30-40k	670
59	Don’t know/refused	116	30-40k	11
60	Evangelical Prot	1529	30-40k	982
61	Hindu	37	30-40k	9
62	Historically Black Prot	339	30-40k	238
63	Jehovah's Witness	37	30-40k	24
64	Jewish	162	30-40k	25
65	Mainline Prot	1328	30-40k	655
66	Mormon	69	30-40k	51
67	Muslim	22	30-40k	10
68	Orthodox	73	30-40k	32
69	Other Christian	18	30-40k	13
70	Other Faiths	71	30-40k	46
71	Other World Religions	8	30-40k	4
72	Unaffiliated	597	30-40k	365
73	Agnostic	96	40-50k	76
74	Atheist	76	40-50k	35
75	Buddhist	54	40-50k	33
76	Catholic	1489	40-50k	638
77	Don’t know/refused	116	40-50k	10
78	Evangelical Prot	1529	40-50k	881
79	Hindu	37	40-50k	11
80	Historically Black Prot	339	40-50k	197
81	Jehovah's Witness	37	40-50k	21
82	Jewish	162	40-50k	30
83	Mainline Prot	1328	40-50k	651
84	Mormon	69	40-50k	56
85	Muslim	22	40-50k	9
86	Orthodox	73	40-50k	32
87	Other Christian	18	40-50k	13
88	Other Faiths	71	40-50k	49
89	Other World Religions	8	40-50k	2
90	Unaffiliated	597	40-50k	341
91	Agnostic	96	50-75k	137
92	Atheist	76	50-75k	70
93	Buddhist	54	50-75k	58
94	Catholic	1489	50-75k	1116
95	Don’t know/refused	116	50-75k	35

96	Evangelical Prot	1529	50-75k	1486
97	Hindu	37	50-75k	34
98	Historically Black Prot	339	50-75k	223
99	Jehovah's Witness	37	50-75k	30
100	Jewish	162	50-75k	95
101	Mainline Prot	1328	50-75k	1107
102	Mormon	69	50-75k	112
103	Muslim	22	50-75k	23
104	Orthodox	73	50-75k	47
105	Other Christian	18	50-75k	14
106	Other Faiths	71	50-75k	63
107	Other World Religions	8	50-75k	7
108	Unaffiliated	597	50-75k	528
109	Agnostic	96	75-100k	122
110	Atheist	76	75-100k	73
111	Buddhist	54	75-100k	62
112	Catholic	1489	75-100k	949
113	Don’t know/refused	116	75-100k	21
114	Evangelical Prot	1529	75-100k	949
115	Hindu	37	75-100k	47
116	Historically Black Prot	339	75-100k	131
117	Jehovah's Witness	37	75-100k	15
118	Jewish	162	75-100k	69
119	Mainline Prot	1328	75-100k	939
120	Mormon	69	75-100k	85
121	Muslim	22	75-100k	16
122	Orthodox	73	75-100k	38
123	Other Christian	18	75-100k	18
124	Other Faiths	71	75-100k	46
125	Other World Religions	8	75-100k	3
126	Unaffiliated	597	75-100k	407
127	Agnostic	96	100-150k	109
128	Atheist	76	100-150k	59
129	Buddhist	54	100-150k	39
130	Catholic	1489	100-150k	792
131	Don’t know/refused	116	100-150k	17
132	Evangelical Prot	1529	100-150k	723
133	Hindu	37	100-150k	48
134	Historically Black Prot	339	100-150k	81
135	Jehovah's Witness	37	100-150k	11
136	Jewish	162	100-150k	87
137	Mainline Prot	1328	100-150k	753
138	Mormon	69	100-150k	49
139	Muslim	22	100-150k	8
140	Orthodox	73	100-150k	42
141	Other Christian	18	100-150k	14
142	Other Faiths	71	100-150k	40
143	Other World Religions	8	100-150k	4
144	Unaffiliated	597	100-150k	321
145	Agnostic	96	150k	84
146	Atheist	76	150k	74
147	Buddhist	54	150k	53
148	Catholic	1489	150k	633
149	Don’t know/refused	116	150k	18
150	Evangelical Prot	1529	150k	414
151	Hindu	37	150k	54
152	Historically Black Prot	339	150k	78
153	Jehovah's Witness	37	150k	6
154	Jewish	162	150k	151
155	Mainline Prot	1328	150k	634
156	Mormon	69	150k	42
157	Muslim	22	150k	6
158	Orthodox	73	150k	46

```
159          Other Christian                 18     150k    12
160             Other Faiths                 71     150k    41
161    Other World Religions                  8     150k     4
162             Unaffiliated                597     150k   258
```

24. Import GDP data (Data is shared). The dataset is having 14 rows and 13 columns. Reshape the data in such a way that it should have four columns: Country, Variable, Year, Var1. After reshaping store the new data in different object.

```
> gdp<-melt(GDP,id.vars = c("Country","Variable"),variable.name = "newGDP",
value.name = "var1")

> gdp
    Country Variable newGDP     var1
1         A     var1  X1995       ..
2         A     var2  X1995       ..
3         B     var1  X1995 18268.01
4         B     var2  X1995     2.87
5         C     var1  X1995 21088.14
6         C     var2  X1995     1.60
7         D     var1  X1995   313.74
8         D     var2  X1995     2.66
9         E     var1  X1995 21123.66
10        E     var2  X1995     2.69
11        F     var1  X1995 29941.64
12        F     var2  X1995     1.32
13        G     var1  X1995  4891.60
14        G     var2  X1995    -7.86
15        A     var1  X1996       ..
16        A     var2  X1996       ..
17        B     var1  X1996 18738.99
18        B     var2  X1996     2.58
19        C     var1  X1996 21608.14
20        C     var2  X1996     2.47
21        D     var1  X1996   321.36
22        D     var2  X1996     2.43
23        E     var1  X1996 21659.55
24        E     var2  X1996     2.54
25        F     var1  X1996 30703.73
26        F     var2  X1996     2.55
27        G     var1  X1996  5063.81
28        G     var2  X1996     3.52
29        A     var1  X1997  8000.01
30        A     var2  X1997     6.83
31        B     var1  X1997 19360.46
32        B     var2  X1997     3.32
33        C     var1  X1997 21988.64
34        C     var2  X1997     1.76
35        D     var1  X1997   331.76
36        D     var2  X1997     3.24
37        E     var1  X1997 22299.13
38        E     var2  X1997     2.95
39        F     var1  X1997 31716.04
40        F     var2  X1997      3.3
```

41	G	var1	X1997	5328.88
42	G	var2	X1997	5.23
43	A	var1	X1998	8212.9
44	A	var2	X1998	2.66
45	B	var1	X1998	20151.42
46	B	var2	X1998	4.09
47	C	var1	X1998	22739.28
48	C	var2	X1998	3.41
49	D	var1	X1998	342.12
50	D	var2	X1998	3.12
51	E	var1	X1998	22972.31
52	E	var2	X1998	3.02
53	F	var1	X1998	32671.27
54	F	var2	X1998	3.01
55	G	var1	X1998	5512.59
56	G	var2	X1998	3.45
57	A	var1	X1999	7847.36
58	A	var2	X1999	-4.45
59	B	var1	X1999	20715.54
60	B	var2	X1999	2.8
61	C	var1	X1999	23436.61
62	C	var2	X1999	3.07
63	D	var1	X1999	351.7
64	D	var2	X1999	2.8
65	E	var1	X1999	23613.87
66	E	var2	X1999	2.79
67	F	var1	X1999	33748.21
68	F	var2	X1999	3.3
69	G	var1	X1999	5647.06
70	G	var2	X1999	2.44
71	A	var1	X2000	7702.89
72	A	var2	X2000	-1.84
73	B	var1	X2000	20866.9
74	B	var2	X2000	0.73
75	C	var1	X2000	24194.85
76	C	var2	X2000	3.24
77	D	var1	X2000	365.33
78	D	var2	X2000	3.87
79	E	var1	X2000	24150.86
80	E	var2	X2000	2.27
81	F	var1	X2000	34599.47
82	F	var2	X2000	2.52
83	G	var1	X2000	5934.98
84	G	var2	X2000	5.1
85	A	var1	X2001	7288.48
86	A	var2	X2001	-5.38
87	B	var1	X2001	21364.02
88	B	var2	X2001	2.38
89	C	var1	X2001	24300.57
90	C	var2	X2001	0.44
91	D	var1	X2001	377.15
92	D	var2	X2001	3.24
93	E	var1	X2001	24788.69
94	E	var2	X2001	2.64
95	F	var1	X2001	34483.98
96	F	var2	X2001	-0.33

97	G	var1	X2001	5864.12
98	G	var2	X2001	-1.19
99	A	var1	X2002	6430.98
100	A	var2	X2002	-11.77
101	B	var1	X2002	21801.41
102	B	var2	X2002	2.05
103	C	var1	X2002	24411.48
104	C	var2	X2002	0.46
105	D	var1	X2002	386.26
106	D	var2	X2002	2.42
107	E	var1	X2002	25368.87
108	E	var2	X2002	2.34
109	F	var1	X2002	34669.47
110	F	var2	X2002	0.54
111	G	var1	X2002	5852.99
112	G	var2	X2002	-0.19
113	A	var1	X2003	6932.45
114	A	var2	X2003	7.8
115	B	var1	X2003	22404.59
116	B	var2	X2003	2.77
117	C	var1	X2003	24650.02
118	C	var2	X2003	0.98
119	D	var1	X2003	398.86
120	D	var2	X2003	3.26
121	E	var1	X2003	25885.48
122	E	var2	X2003	2.04
123	F	var1	X2003	35312.75
124	F	var2	X2003	1.86
125	G	var1	X2003	5872.29
126	G	var2	X2003	0.33
127	A	var1	X2004	7486.24
128	A	var2	X2004	7.99
129	B	var1	X2004	22676.26
130	B	var2	X2004	1.21
131	C	var1	X2004	25076.01
132	C	var2	X2004	1.73
133	D	var1	X2004	415.96
134	D	var2	X2004	4.29
135	E	var1	X2004	26582.19
136	E	var2	X2004	2.69
137	F	var1	X2004	36450.55
138	F	var2	X2004	3.22
139	G	var1	X2004	6055.92
140	G	var2	X2004	3.13
141	A	var1	X2005	8094.17
142	A	var2	X2005	8.12
143	B	var1	X2005	23039.43
144	B	var2	X2005	1.6
145	C	var1	X2005	25346.01
146	C	var2	X2005	1.08
147	D	var1	X2005	432.63
148	D	var2	X2005	4.01
149	E	var1	X2005	26890.73
150	E	var2	X2005	1.16
151	F	var1	X2005	37267.33
152	F	var2	X2005	2.24

```
153       G     var1  X2005  6162.84
154       G     var2  X2005     1.77
```