Lab Exercises

1. Consider the dataset *survey.csv* from datasets folder. Using the appropriate functions in **dplyr** package create the following data objects:

Name	Object description	Columns
MaleNonSmokers	Males who are have never smoked	All
PulseGT80	Students whose pulse rate is greater than 80	Sex, Exer, Smoke, Pulse
RtHand	Create a variable of ratio of variables Wr.Hnd / NW.Hnd as Ratio_Hnd	Ratio_Hnd, Clap, Age
DescStats	Calculate the mean and standard deviation for the variables Age	
DescGrp	Calculate the mean and standard deviation for the variables Age grouped by Sex	

2. Given the files ***Items.csv***, ***Orders.csv*** and ***Ord_Details.csv*** in the folder datasets, merge them with appropriate keys to form a combined data.

3. Combine the data in the files ***Courses.csv*** and ***CourseSchedule.csv*** with appropriate keys

4. Consider a dataset comb1 in the datasets folder. Reshape the dataset to the following structure:

```
> glimpse(dt_comb1)
Observations: 76
Variables: 3
$ District <fctr> Ahmednagar, Aurangabad, Buldhana, Dhule, Goa, Kolhapur, Mumbai, Nadurbar, Nagpur, Nan...
$ ItemType <chr> "Highlighter", "Highlighter", "Highlighter", "Highlighter", "Highlighter", "Highlighte...
$ qty      <int> 225, 162, 30, 193, 81, 108, 119, 162, 107, 111, 152, 154, 174, 326, 121, 72, 75, 164, ...
```

Data should be like this:

```
     District  ItemType    qty
1   Ahmednagar Highlighter  225
2   Aurangabad Highlighter  162
```

3	Buldhana	Highlighter	30
4	Dhule	Highlighter	193
5	Goa	Highlighter	81
6	Kolhapur	Highlighter	108
7	Mumbai	Highlighter	119
8	Nadurbar	Highlighter	162
9	Nagpur	Highlighter	107
10	Nanded	Highlighter	111
11	Nasik	Highlighter	152
12	Pune	Highlighter	154
13	Raigad	Highlighter	174
14	Ratnagiri	Highlighter	326
15	Sangli	Highlighter	121
16	Satara	Highlighter	72
17	Sindhudurga	Highlighter	75
18	Solapur	Highlighter	164
19	Thane	Highlighter	224
20	Ahmednagar	Marker	51
21	Aurangabad	Marker	159
22	Buldhana	Marker	32
23	Dhule	Marker	390
24	Goa	Marker	243
25	Kolhapur	Marker	118
26	Mumbai	Marker	160
27	Nadurbar	Marker	281
28	Nagpur	Marker	121
29	Nanded	Marker	240
30	Nasik	Marker	239
31	Pune	Marker	309
32	Raigad	Marker	436
33	Ratnagiri	Marker	398
34	Sangli	Marker	110
35	Satara	Marker	156
36	Sindhudurga	Marker	217
37	Solapur	Marker	160
38	Thane	Marker	246
39	Ahmednagar	Pen	674
40	Aurangabad	Pen	755
41	Buldhana	Pen	616
42	Dhule	Pen	553
43	Goa	Pen	518
44	Kolhapur	Pen	352
45	Mumbai	Pen	391
46	Nadurbar	Pen	386
47	Nagpur	Pen	280
48	Nanded	Pen	761
49	Nasik	Pen	619
50	Pune	Pen	768
51	Raigad	Pen	545
52	Ratnagiri	Pen	1504
53	Sangli	Pen	221
54	Satara	Pen	373
55	Sindhudurga	Pen	1253
56	Solapur	Pen	391
57	Thane	Pen	733
58	Ahmednagar	Refill	69
59	Aurangabad	Refill	50
60	Buldhana	Refill	38
61	Dhule	Refill	40
62	Goa	Refill	NA
63	Kolhapur	Refill	NA
64	Mumbai	Refill	29
65	Nadurbar	Refill	156
66	Nagpur	Refill	33

67	Nanded	Refill	28
68	Nasik	Refill	65
69	Pune	Refill	122
70	Raigad	Refill	40
71	Ratnagiri	Refill	74
72	Sangli	Refill	47
73	Satara	Refill	NA
74	Sindhudurga	Refill	77
75	Solapur	Refill	15
76	Thane	Refill	58

5. Consider the dataset comb2 in the datasets folder. PatientID in the data is to be broken up into the format projectID-SiteID/PatientNumber. Create the dataset with the following structure:

```
> glimpse(dt_comb2)
Observations: 10
Variables: 5
$ projectID     <int> 123, 123, 134, 190, 190, 234, 231, 231, 134, 234
$ SiteID        <int> 984, 345, 93, 32, 309, 92, 902, 309, 92, 984
$ PatientNumber <int> 45, 34, 43, 34, 32, 23, 21, 21, 34, 43
$ Lab_test      <chr> "A", "A", "A", "B", "B", "B", "B", "A", "B", "A"
$ Result        <chr> "Present", "Absent", "Present", "Present", "Absent",...
```

Data View Should be:

	projectID	SiteID	PatientNumber	Lab_test	Result
1	123	984	45	A	Present
2	123	345	34	A	Absent
3	134	093	43	A	Present
4	190	032	34	B	Present
5	190	309	32	B	Absent
6	234	092	23	B	Present
7	231	902	21	B	Absent
8	231	309	21	A	Absent
9	134	092	34	B	Present
10	234	984	43	A	Present