Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

Exercise 1

How many cases are there in this data set? How many variables? For each variable, identify its data type (e.g. categorical, discrete).

There are 20000 cases in this dataset.

From

```
> tail(cdc)
       genhlth exerany hlthplan smoke100 height weight wtdesire age gender
19995      good       0        1        1     69    224      224  73      m
19996      good       1        1        0     66    215      140  23      f
19997 excellent       0        1        0     73    200      185  35      m
19998      poor       0        1        0     65    216      150  57      f
19999      good       1        1        0     67    165      165  81      f
20000      good       1        1        1     69    170      165  83      m
```

Variable are 20000, and the data type are identified as following:

```
> table(cdc$genhlth)
excellent very good      good      fair      poor
     4657      6972      5675      2019       677
> table(cdc$exerany)
    0     1
 5086 14914
> table(cdc$hlthplan)
    0     1
 2524 17476
> table(cdc$smoke100)
    0     1
10559  9441
> table(cdc$height)
  48   49   50   51   52   53   54   55   56   57   58   59   60   61   62   63   64
 65   66
   2    1    1    2    2    7    3    4   17   20   51  170  613  594 1272 1368 1662 1
568 1843
  67   68   69   70   71   72   73   74   75   76   77   78   79   80   81   82   83
 84   93
```

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

```
1671 1505 1380 1500 1296 1393  784  605  321  189   80   43   15   10    3    2    1
   1    1
> table(cdc$wtdesire)
  68   77   78   80   82   85   88   90   91   92   93   94   95   96   97   98   99
100  101
   1    3    1    4    2    4    1   15    1    2    2    1   20    1    3   16    6
152    5
 102  103  104  105  106  107  108  109  110  111  112  113  114  115  116  117  118
119  120
  19   24   10  183   20   27   48   15  474    9   52   27   26  534   31   45  111
 25 1037
 121  122  123  124  125  126  127  128  129  130  131  132  133  134  135  136  137
138  139
  22   61   53   51 1031   52   55  121   29 1393    8   73   22   44  915   34   28
 67   19
 140  141  142  143  144  145  146  147  148  149  150  151  152  153  154  155  156
157  158
1183   17   42   21   13  607   23   30   49   18 1482    8   42   26   20  469   21
 31   52
 159  160  161  162  163  164  165  166  167  168  169  170  171  172  173  174  175
176  177
  12 1038   13   51   24   17  689    9   23   48   16  887   11   65   20   32  877
 24    9
 178  179  180  181  182  183  184  185  186  187  188  189  190  191  192  193  194
195  196
  48   15 1058    5   27   17   10  613   13   12   15    9  651    5   10    8   13
263   13
 197  198  199  200  201  202  203  204  205  206  207  208  209  210  211  212  213
214  215
   8   20   10  823    2    2   11   11  108    4    7    7    3  253    2   12    1
   2  112
 216  217  218  219  220  222  223  224  225  226  227  228  229  230  234  235  237
238  240
   1    1    9    2  214    5    1    2  108    1    1    1    1  105    1   44    1
   2   67
```

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

```
 242  244  245  248  249  250  252  255  260  265  267  270  273  274  275  280  285
290  298
   1    1   11    1    1   76    2    5   17    6    1    6    1    1    9   12    1
  3    1
 300  315  320  325  350  601  680
  13    1    1    2    1    1    1
> table(cdc$gender)
    m     f
 9569 10431
```

Exercise 2

Create a numerical summary for `height` and `age`, and compute the interquartile range for each. Compute the relative frequency distribution for `gender` and `exerany`. How many males are in the sample? What proportion of the sample reports being in excellent health?

Summary for height:

```
> summary(cdc$height)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  48.00   64.00   67.00   67.18   70.00   93.00
```

Interquartile range for height:

```
> 70-64
[1] 6
```

Summary for age:

```
> summary(cdc$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   31.00   43.00   45.07   57.00   99.00
```

Interquartile range for age:

```
> 57-31
[1] 26
```

the relative frequency distribution for `gender` and `exerany`:

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

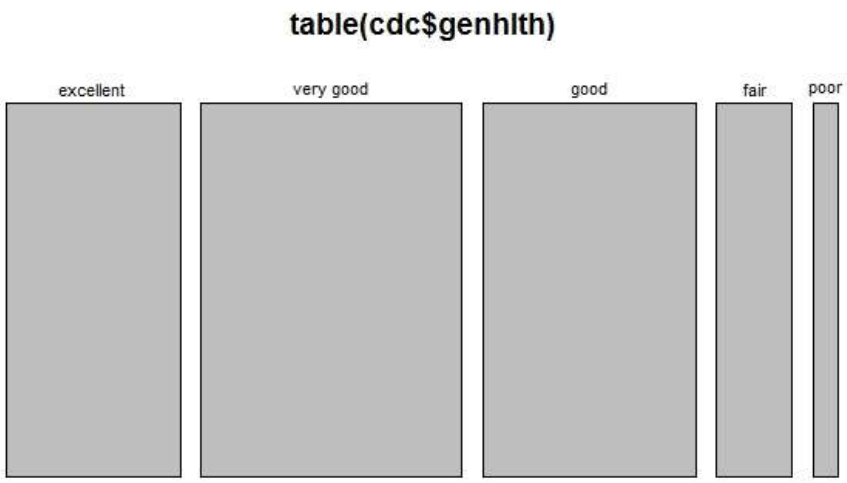
```
> table(cdc$exerany, cdc$gender)/20000

          m       f
  0 0.10745 0.14685
  1 0.37100 0.37470
```

```
> summary(cdc$gender)
    m     f
 9569 10431
```

```
> table(cdc$genhlth)

excellent very good      good      fair      poor
     4657      6972      5675      2019       677
> mosaicplot(table(cdc$genhlth))
```

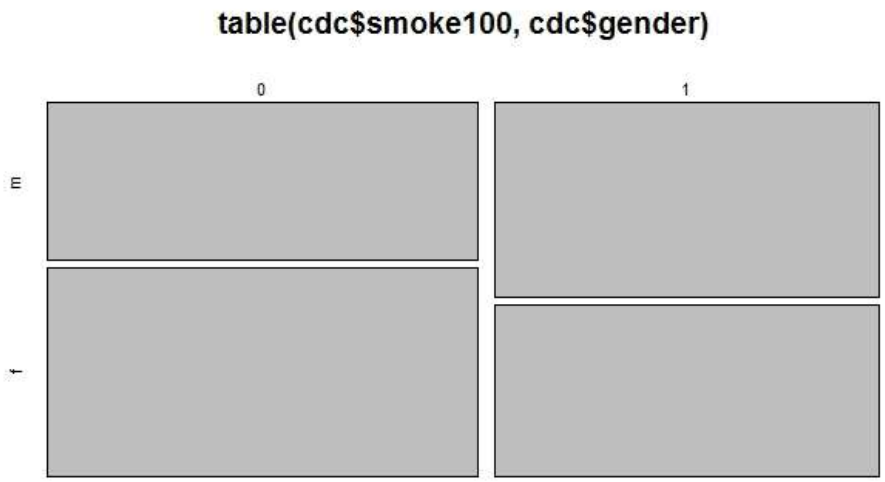


Exercise 3

What does the mosaic plot reveal about smoking habits and gender?

```
> mosaicplot(table(cdc$smoke100,cdc$gender))
```

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639



Exercise 4

Create a new object called `under23_and_smoke` that contains all observations of respondents under the age of 23 that have smoked 100 cigarettes in their lifetime. Write the command you used to create the new object as the answer to this exercise.

```
> under23_and_smoke<-subset(cdc, smoke100=="1" & age <23)
> under23_and_smoke
```

	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
13	excellent	1	0	1	66	185	220	21	m
37	very good	1	0	1	70	160	140	18	f
96	excellent	1	1	1	74	175	200	22	m
180	good	1	1	1	64	190	140	20	f
182	very good	1	1	1	62	92	92	21	f
240	very good	1	0	1	64	125	115	22	f
262	fair	0	1	1	71	185	185	20	m
296	fair	1	1	1	72	185	170	19	m
297	excellent	1	0	1	63	105	100	19	m
300	fair	1	1	1	71	185	150	18	m
306	excellent	0	1	1	66	138	138	21	f
325	fair	1	1	1	63	110	110	20	f
355	very good	1	1	1	64	150	135	20	f

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

370	good	1	0	1	64	200	120	21	f
414	excellent	1	1	1	63	130	130	22	f
416	very good	1	1	1	75	220	220	18	m
421	good	0	1	1	62	160	130	20	f
439	good	1	1	1	62	102	110	19	f
446	excellent	1	1	1	62	130	120	22	f
484	good	1	1	1	67	165	148	22	m
492	very good	1	1	1	76	253	226	19	m
521	very good	1	1	1	63	100	100	20	f
545	excellent	1	0	1	70	175	150	21	f
588	excellent	1	1	1	69	164	135	21	f
628	excellent	1	1	1	72	165	175	21	m
674	excellent	1	1	1	69	175	175	21	m
693	excellent	1	0	1	77	208	200	21	m
698	excellent	1	0	1	67	140	150	22	m
699	good	1	0	1	69	238	170	18	f
733	excellent	1	0	1	74	173	200	22	m
752	good	1	1	1	68	210	180	20	m
817	very good	1	1	1	70	230	200	22	m
882	good	1	1	1	72	160	160	21	m
958	good	0	0	1	71	120	120	19	f
962	excellent	1	1	1	68	165	180	21	m
1039	very good	0	0	1	68	180	165	21	m
1116	fair	1	0	1	66	163	140	18	f
1147	good	0	0	1	63	150	120	22	f
1172	excellent	1	1	1	73	197	230	18	m
1183	very good	0	0	1	72	135	142	18	m
1200	good	1	1	1	68	150	175	22	m
1241	good	1	1	1	70	180	160	21	m
1256	good	1	1	1	62	150	130	18	f
1328	very good	0	0	1	60	110	110	19	f
1339	excellent	0	1	1	64	138	150	22	m
1370	good	1	1	1	73	180	180	19	m
1379	excellent	1	1	1	64	180	150	20	f
1382	excellent	1	1	1	67	160	160	22	m
1401	good	1	0	1	59	163	130	19	f

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

1413	fair	0	1	1	64	168	140	18	f
1417	good	0	0	1	59	103	103	21	f
1441	very good	1	1	1	67	115	115	18	f
1488	good	1	1	1	65	130	130	22	f
1489	excellent	1	1	1	74	200	190	22	m
1507	fair	1	0	1	62	110	110	21	f
1513	very good	1	0	1	65	160	120	21	f
1555	very good	1	1	1	73	160	175	19	m
1557	good	1	0	1	65	140	133	21	f
1565	good	1	1	1	72	170	210	19	m
1574	good	1	1	1	69	140	160	19	m
1579	good	0	1	1	72	150	165	19	m
1602	very good	0	1	1	61	120	120	19	f
1638	very good	1	1	1	77	236	225	18	m
1670	good	1	1	1	62	98	110	21	f
1713	excellent	1	1	1	72	180	185	22	m
1751	fair	1	1	1	65	130	115	18	f
1762	excellent	1	0	1	71	165	250	21	m
1790	good	1	0	1	74	150	175	21	m
1806	very good	1	1	1	65	140	120	22	f
1841	very good	1	1	1	70	145	145	20	m
1844	very good	0	1	1	66	130	120	20	f
1880	good	1	1	1	74	155	175	22	m
1920	very good	1	1	1	60	130	120	19	f
1946	very good	1	0	1	63	140	135	20	m
2011	excellent	1	0	1	62	90	115	20	f
2019	fair	1	0	1	72	240	200	22	m
2024	very good	1	1	1	73	210	210	20	m
2071	good	1	1	1	64	120	120	20	f
2106	good	1	1	1	66	145	130	20	f
2119	excellent	1	0	1	63	125	120	21	f
2120	excellent	1	0	1	67	185	155	20	f
2124	excellent	1	1	1	72	175	175	21	m
2143	good	0	0	1	65	120	130	19	f
2161	good	0	1	1	72	135	135	19	m
2163	excellent	1	0	1	75	227	227	21	m

Course: DATA 606
Lab 1
Student Name: Lung Tze Fung
ID: 23637639

2199	excellent	1	1	1	64	135	135	21	f
2206	excellent	1	0	1	63	125	125	20	f
2209	good	1	1	1	69	140	130	20	f
2244	good	0	0	1	66	145	145	19	m
2252	very good	0	1	1	71	145	180	18	m
2256	very good	1	1	1	69	150	150	20	m
2366	very good	1	0	1	70	150	150	20	m
2391	very good	1	1	1	76	205	205	21	m
2404	excellent	1	1	1	67	175	150	20	f
2409	excellent	1	1	1	66	155	155	18	m
2422	excellent	0	1	1	74	165	165	20	m
2461	very good	1	1	1	74	265	230	20	m
2466	very good	1	1	1	72	160	160	20	m
2476	good	1	1	1	76	225	250	21	m
2495	very good	1	1	1	75	210	200	19	m
2505	fair	1	1	1	70	180	180	19	m
2569	excellent	1	1	1	73	175	195	18	m
2634	very good	1	1	1	69	140	160	19	m
2719	very good	1	1	1	74	220	220	20	m
2728	excellent	1	1	1	71	150	160	19	m
2783	good	1	1	1	59	145	145	22	f
2828	good	1	1	1	69	175	200	20	m
2842	fair	0	0	1	70	135	135	21	m
2880	excellent	1	0	1	71	140	140	18	m
2901	poor	1	0	1	68	125	140	20	f
2918	fair	1	0	1	64	200	165	22	f

[reached getOption("max.print") -- omitted 509 rows]