

Exercise 1

#number of students(340823)

```
> dim(datstu)[[1]]
```

```
[1] 340823
```

#number of schools(898)

```
> n_distinct(datsss$schoolcode)
```

```
[1] 898
```

```
>
```

#numbers of programs(270)

```
> n_distinct(datstu$choicepgm1:datstu$choicepgm6)
```

```
[1] 270
```

```
>
```

#Number of Choices(338936)

```
> datstu = datstu %>% mutate(choice1=paste0(schoolcode1,choicepgm1),
+                               choice2=paste0(schoolcode2,choicepgm2),
+                               choice3=paste0(schoolcode3,choicepgm3),
+                               choice4=paste0(schoolcode4,choicepgm4),
+                               choice5=paste0(schoolcode5,choicepgm5),
+                               choice6=paste0(schoolcode6,choicepgm6))
> dat = select(datstu,choice1,choice2,choice3,choice4,choice5,choice6)
> n_distinct(dat$choice1,dat$choice2,dat$choice3,dat$choice4,dat$choice5,dat$choice6)
[1] 338936
```

#Missing test score(179887)

```
> sum(is.na(datstu$score))
```

```
[1] 179887
```

#Apply to less than 6 school choices

```
> subset=select(datstu,schoolcode1,schoolcode2,schoolcode3,schoolcode4,schoolcode5,schoolcode
6)
> list<-which(rowSums(is.na(subset))>0)
>
```

```
> print(list)
```

[1]	30	34	67	85	88	92	97	199	239	254	268	317	333	358
[15]	398	455	464	469	487	493	495	506	734	765	818	842	843	853
[29]	869	878	923	929	1031	1038	1079	1099	1102	1123	1137	1159	1203	1211
[43]	1238	1267	1319	1321	1421	1440	1461	1494	1506	1524	1534	1543	1570	1584
[57]	1601	1607	1611	1676	1681	1711	1716	1723	1733	1744	1773	1799	1807	1831
[71]	1845	1866	1886	1900	1949	2002	2048	2074	2076	2078	2111	2125	2139	2158
[85]	2165	2179	2204	2250	2266	2273	2281	2286	2353	2362	2365	2366	2384	2392
[99]	2401	2419	2453	2506	2508	2589	2593	2595	2628	2659	2671	2689	2692	2713

#Apply to less than 6 program choice

```
> subset1<-select(datstu,choicepgm1,choicepgm2,choicepgm3,choicepgm4,choicepgm5,choicepgm6)
> list<-which(rowSums(is.na(subset1))>0)
> print(list)
```

[1]	30	34	67	85	88	92	97	199	239	254	268	317	333	358
[15]	398	455	464	469	487	493	495	506	734	765	818	842	843	853
[29]	869	878	923	929	1031	1038	1079	1099	1102	1123	1137	1159	1203	1211
[43]	1238	1267	1319	1321	1421	1440	1461	1494	1506	1524	1534	1543	1570	1584
[57]	1601	1607	1611	1676	1681	1711	1716	1723	1733	1744	1773	1799	1807	1831
[71]	1845	1866	1886	1900	1949	2002	2048	2074	2076	2078	2111	2125	2139	2158
[85]	2165	2179	2204	2250	2266	2273	2281	2286	2353	2362	2365	2366	2384	2392
[99]	2401	2419	2453	2506	2508	2589	2593	2595	2628	2659	2671	2689	2692	2713
[113]	2717	2722	2746	2792	2797	2800	2830	2834	2875	2896	2902	2915	2972	2994

Exercise 2

Report cutoff,size, quality

	datstu1\$schoolcode	cutoff	size	quality
1	10101	213	294	287.3912
2	10102	226	621	350.7955
3	10103	214	895	306.4235
4	10104	218	53	281.6226
5	10105	205	319	325.1818
6	10106	216	482	302.0892
7	10107	209	93	277.7312
8	10108	207	90	273.5111
9	10109	194	107	271.4953
10	10110	226	1449	368.8254
11	10111	236	1471	378.8722
12	10112	217	482	331.6307
13	10114	208	520	298.8596
14	10115	203	284	280.0599
15	10116	206	173	284.6358
16	10117	206	855	313.9754
17	10118	214	318	290.6887
18	10119	217	49	278.2245
19	10120	216	77	279.1688
20	10121	210	1083	339.2992
21	10123	367	1	367.0000

report district, latitude, longitude and combine with “cutoff,size, quality”

	datstu1\$schoolcode	cutoff	size	quality	sssdistrict	ssslong	ssslat
1	10101	213	294	287.3912	Accra Metropolitan	-0.19711526	5.607396
2	10102	226	621	350.7955	Accra Metropolitan	-0.19711526	5.607396
3	10103	214	895	306.4235	Accra Metropolitan	-0.19711526	5.607396
4	10104	218	53	281.6226	Accra Metropolitan	-0.19711526	5.607396
5	10105	205	319	325.1818	Accra Metropolitan	-0.19711526	5.607396
6	10106	216	482	302.0892	Accra Metropolitan	-0.19711526	5.607396
7	10107	209	93	277.7312	Accra Metropolitan	-0.19711526	5.607396
8	10108	207	90	273.5111	Accra Metropolitan	-0.19711526	5.607396
9	10109	194	107	271.4953	Accra Metropolitan	-0.19711526	5.607396
10	10110	226	1449	368.8254	Accra Metropolitan	-0.19711526	5.607396
11	10111	236	1471	378.8722	Ga West (Amasaman)	-0.39751053	5.664688
12	10112	217	482	331.6307	Accra Metropolitan	-0.19711526	5.607396
13	10114	208	520	298.8596	Accra Metropolitan	-0.19711526	5.607396
14	10115	203	284	280.0599	Accra Metropolitan	-0.19711526	5.607396
15	10116	206	173	284.6358	Accra Metropolitan	-0.19711526	5.607396
16	10117	206	855	313.9754	Accra Metropolitan	-0.19711526	5.607396
17	10118	214	318	290.6887	Ga West (Amasaman)	-0.39751053	5.664688
18	10119	217	49	278.2245	Accra Metropolitan	-0.19711526	5.607396
19	10120	216	77	279.1688	Accra Metropolitan	-0.19711526	5.607396
20	10121	210	1083	339.2992	Accra Metropolitan	-0.19711526	5.607396
21	10123	367	1	367.0000	Accra Metropolitan	-0.19711526	5.607396
22	10124	229	3	250.6667	Accra Metropolitan	-0.19711526	5.607396

Showing 1 to 22 of 573 entries. 7 total columns

create rows corresponds with (school,program)

	key	value
1	choice1	30905General Science
2	choice1	31201General Arts
3	choice1	30105Business
4	choice1	31201General Arts
5	choice1	30403General Arts
6	choice1	30902General Arts
7	choice1	10502Visual Arts
8	choice1	30403General Arts
9	choice1	30109General Arts
10	choice1	30303Agriculture
11	choice1	30702General Science
12	choice1	30403General Arts
13	choice1	31201Business
14	choice1	30403General Arts
15	choice1	30305General Arts
16	choice1	21103Home Economics
17	choice1	60301General Arts
18	choice1	31201Business
19	choice1	30601General Arts
20	choice1	30105General Arts
21	choice1	31201Visual Arts
22	choice1	30801General Arts

Exercise 3

Distance

```
> z=sqrt((69.172 * (sssjss$ssslong-sssjss$point_x) * cos(sssjss$point_y/57.3))^2+(69.172 * (sssjss$ssslat-sssjss$point_y))^2)
> print(z)
[1] 134.096471 302.887571 149.911415 384.024735 204.682638 81.015011 292.417252
[8] 167.107839 380.540341 66.122956 44.526311 106.052990 42.203469 103.949462
[15] 128.928600 132.106895 98.519871 85.030567 26.494148 117.911328 74.615532
[22] 79.051800 98.813356 107.660006 56.252070 0.000000 124.130355 167.365322
[29] 119.472803 78.252814 109.692085 193.810773 177.893780 227.641645 75.144875
[36] 67.407523 139.116217 194.370882 30.451392 105.551403 178.011348 153.254922
[43] 142.314449 67.901202 160.916337 152.270995 72.871858 175.039090 37.879170
[50] 91.983016 49.692823 13.848205 56.822337 51.161667 74.825318 96.087569
[57] 9.056901 43.736152 20.939331 76.759309 53.429613 109.890170 108.947873
[64] 27.057917 50.748625 54.356008 24.259177 86.988323 55.111036 141.109343
[71] 111.369981 92.220317 119.081588 13.056425 47.960275 18.670382 167.916144
[78] 25.622759 129.600753 157.950860 93.177469 109.407552 11.622429 225.085515
```

Exercise 4

Average Cutoff(178.2), /size(23204) /quality (280.9)

```
datstu1$rankplace      cutoff      size      quality
Min.   :1.00      Min.   :158.0   Min.   : 2878   Min.   :251.2
1st Qu.:2.25      1st Qu.:167.0   1st Qu.: 9308   1st Qu.:258.8
Median :3.50      Median :179.0   Median :27458   Median :282.7
Mean   :3.50      Mean   :178.2   Mean   :23204   Mean   :280.9
3rd Qu.:4.75      3rd Qu.:188.8   3rd Qu.:31514   3rd Qu.:299.0
Max.   :6.00      Max.   :198.0   Max.   :45306   Max.   :313.6
> |
```

average standard deviation

sd(cutoff=15.40671), sd(size)=16732.4, sd(quality)=25.64774

```
> sd(groupby_set_2$cutoff)
[1] 15.40671
> sd(groupby_set_2$size)
[1] 16732.4
> sd(groupby_set_2$quality)
[1] 25.64774
> |
```

Average Distance(149.85)

> summary(z)

```
Min. 1st Qu. Median      Mean 3rd Qu.      Max.
 0.00  75.95 126.31 149.85 194.09 387.67
> |
```

Redo the same table differentiating by test score quantiles

```
> quantile(datstu$score,na.rm=TRUE,probs = seq(0,1, 0.25))
 0%  25%  50%  75% 100%
158  256  289  330  469
```

choice3	choice4	choice5	choice6	quantile
0902General Arts	30903Home Economics	30403General Arts	30801General Science	0%~25%
0304General Arts	30402Business	30402General Arts	30303Agriculture	0%~25%
0402Business	30403Home Economics	30303Business	30201General Arts	25%~50%
0403General Arts	30403General Arts	30110General Arts	30305General Arts	0%~25%
0203Technical	9030401Mech. Eng. Craft Pract.	30504General Arts	31204Technical	0%~25%
1101General Arts	31202General Arts	30903Business	30403Agriculture	25%~50%
0503Visual Arts	10111Visual Arts	30403General Arts	30403General Arts	0%~25%
0304General Arts	30302Home Economics	30701General Arts	30702General Arts	0%~25%
0305General Arts	30110General Arts	30108General Arts	30403General Arts	0%~25%
0403Agriculture	30504Agriculture	30403Agriculture	30902Agriculture	0%~25%
0403Agriculture	30504Agriculture	30902Agriculture	30403Agriculture	0%~25%
1201Agriculture	30403General Arts	30110General Arts	31102Agriculture	0%~25%
0403Agriculture	30402Business	30304Visual Arts	31203General Arts	0%~25%
0110General Arts	31201General Arts	30304General Arts	31203Business	0%~25%
0402General Arts	30403General Arts	30109General Arts	30201General Arts	0%~25%
1201Agriculture	30403General Arts	30903Home Economics	31203General Arts	25%~50%

Exercise 5

Values	
error	num [1:10000] 1.55 2.38 1.47 2.21 1.17 ...
list	int [1:17734] 30 34 67 85 88 92 97 199 239 254...
row_name	chr [1:8] "X" "score" "age" "male" "schoolcode..."
x1	num [1:10000] 1.56 2.81 2.81 2.02 1.7 ...
x2	num [1:10000] 2.64 4.58 1.34 2.57 3.66 ...
x3	int [1:10000] 0 0 1 0 0 0 0 1 0 1 ...
y_mean	-0.480709834423024

Exercise 6

Calculate the correlation between Y and X1=0.1808476

```
> cor(cal_df$ydum,x1)
[1] 0.1808476
```

The outcome of regression Y on X and calculate the coefficients

```

> X= c(1,x1,x2,x3)
> cal_df$constant=1
> X=as.matrix(cal_df[,c(7,1,2,3)])
> solve(t(X)%*%X)%*%t(X)%*%as.matrix(cal_df$y)
      [,1]
constant 2.50799995
x1        1.20669146
x2       -0.90310975
x3         0.08653004
> |

```