Exercise1 output

> datjss <- read.csv("~/Desktop/dat/datjss.csv")

> datsss <- read.csv("~/Desktop/dat/datsss.csv")

> datstu <- read.csv("~/Desktop/dat/datstu.csv")

>

> #exercise 1

>

> length(unique(datstu[,1]))

[1] 340823

> # the number of students is 340823

> y<-as.vector(as.matrix(datstu[,11:16]))

> length(unique(y[complete.cases(y)]))

[1] 33

> # The number of programs is 33 according to the result

>

> x=as.vector(as.matrix(datstu[,5:10]))

> length(unique(x[complete.cases(x)]))

[1] 640

> #The number of schools is 640

>

> x<-as.matrix(x)

> y<-as.matrix(y)

> code.program<-matrix(cbind(x,y),nrow=340823\*6,ncol=2)

> # The original combination of 6 school choices and programs.

>

> z<-code.program[complete.cases(code.program),]

> codeprogram<-as.matrix(z[order(z[,1]),])

> unique(codeprogram)

[,1] [,2]

[1,] "100101" "Technical"

[2,] "100101" "General Arts"

[3,] "100101" "Home Economics"

[4,] "100101" ""

[5,] "100102" "General Arts"

[6,] "100102" "Business"

[7,] "100102" "Agriculture"

[8,] "100102" "Home Economics"

[9,] "100102" "General Science"

[10,] "100102" "Visual Arts"

[11,] "100103" "General Arts"

[12,] "100103" "Agriculture"

[13,] "100103" "Business"

[14,] "100104" "General Science"

[15,] "100104" "Home Economics"

[16,] "100104" "General Arts"

[17,] "100104" "Business"

[18,] "100105" "Home Economics"

[19,] "100105" "Business"

[20,] "100105" "General Arts"

[21,] "100105" ""

[22,] "100106" "Agriculture"

[23,] "100106" "Business"

[24,] "100106" "General Arts"

[25,] "100201" "General Science"

[26,] "100201" "Business"

[27,] "100201" "General Arts"

[28,] "100202" "General Arts"

[29,] "100202" "General Science"

[30,] "100202" "Business"

[31,] "100203" "Home Economics"

[32,] "100203" "Agriculture"

[33,] "100203" "Technical"

[34,] "100204" "Home Economics"

[35,] "100204" "Business"

[36,] "100204" "General Arts"

[37,] "100301" "Agriculture"

[38,] "100301" "General Arts"

[39,] "100301" "Home Economics"

[40,] "100302" "General Arts"

[41,] "100302" "Business"

[42,] "100302" "Home Economics"

[43,] "100302" "General Science"

[44,] "100303" "Agriculture"

[45,] "100303" "Home Economics"

[46,] "100303" "General Arts"

[47,] "100304" "Visual Arts"

[48,] "100304" "Home Economics"

[49,] "100304" "Business"

[50,] "100304" ""

[51,] "100401" "Home Economics"

[52,] "100401" "General Arts"

[53,] "100401" "Visual Arts"

[54,] "100401" "Business"

[55,] "100401" "General Science"

[56,] "100401" "Agriculture"

[57,] "100402" "Agriculture"

[58,] "100402" "Home Economics"

[59,] "100402" "Technical"

[60,] "100402" ""

[61,] "100501" "Home Economics"

[62,] "100501" "Agriculture"

[63,] "100501" "Technical"

[64,] "100501" ""

[65,] "100502" "Agriculture"

[66,] "100502" "General Arts"

[67,] "100502" "Home Economics"

[68,] "100502" ""

[69,] "100503" "General Arts"

[70,] "100503" "Business"

[71,] "100503" "General Science"

[72,] "100503" ""

[73,] "10101" "Home Economics"

[74,] "10101" "General Arts"

[75,] "10101" "Business"

[76,] "10101" "Visual Arts"

[77,] "10101" "Agriculture"

[78,] "10101" "General Science"

[79,] "10101" "Technical"

[80,] "10102" "General Arts"

[81,] "10102" "Home Economics"

[82,] "10102" "General Science"

[83,] "10102" "Visual Arts"

[84,] "10103" "General Arts"

[85,] "10103" "Business"

[86,] "10103" "Visual Arts"

[87,] "10103" "Home Economics"

[88,] "10103" "Agriculture"

[89,] "10103" "General Science"

[90,] "10104" "Visual Arts"

[91,] "10104" "General Arts"

[92,] "10104" "Home Economics"

[93,] "10104" "General Science"

[94,] "10105" "Visual Arts"

[95,] "10105" "General Science"

[96,] "10105" "General Arts"

[97,] "10105" "Agriculture"

[98,] "10105" "Business"

[99,] "10106" "Business"

[100,] "10106" "Visual Arts"

[101,] "10106" "General Arts"

[102,] "10106" "Home Economics"

[103,] "10106" "Technical"

[104,] "10106" "General Science"

[105,] "10107" "Home Economics"

[106,] "10107" "Visual Arts"

[107,] "10107" "Business"

[108,] "10107" "General Science"

[109,] "10107" "General Arts"

[110,] "10108" "Home Economics"

[111,] "10108" "General Arts"

[112,] "10108" "General Science"

[113,] "10108" "Business"

[114,] "10108" "Visual Arts"

[115,] "10109" "Visual Arts"

[116,] "10109" "Technical"

[117,] "10109" "Home Economics"

[118,] "10109" ""

[119,] "10109" "General Arts"

[120,] "10110" "General Arts"

[121,] "10110" "General Science"

[122,] "10110" "Home Economics"

[123,] "10110" "Visual Arts"

[124,] "10110" "Agriculture"

[125,] "10111" "Visual Arts"

[126,] "10111" "Agriculture"

[127,] "10111" "General Arts"

[128,] "10111" "Business"

[129,] "10111" "General Science"

[130,] "10112" "General Arts"

[131,] "10112" "Home Economics"

[132,] "10112" "Business"

[133,] "10112" "Visual Arts"

[134,] "10112" "General Science"

[135,] "10114" "General Arts"

[136,] "10114" "Business"

[137,] "10114" "General Science"

[138,] "10114" "Home Economics"

[139,] "10114" "Visual Arts"

[140,] "10115" "Technical"

[141,] "10115" "General Arts"

[142,] "10115" "Home Economics"

[143,] "10115" "General Science"

[144,] "10115" "Visual Arts"

[145,] "10115" ""

[146,] "10116" "Business"

[147,] "10116" "Visual Arts"

[148,] "10116" "Home Economics"

[149,] "10116" "General Arts"

[150,] "10116" "General Science"

[151,] "10116" ""

[152,] "10117" "Home Economics"

[153,] "10117" "Business"

[154,] "10117" "Visual Arts"

[155,] "10117" "General Arts"

[156,] "10117" "General Science"

[157,] "10117" "Agriculture"

[158,] "10118" "Visual Arts"

[159,] "10118" "Home Economics"

[160,] "10118" "Business"

[161,] "10118" "Agriculture"

[162,] "10118" "General Arts"

[163,] "10118" "General Science"

[164,] "10118" "Technical"

[165,] "10118" ""

[166,] "10119" "Business"

[167,] "10119" "Home Economics"

[168,] "10119" "Visual Arts"

[169,] "10119" "General Arts"

[170,] "10119" "General Science"

[171,] "10119" ""

[172,] "10120" "Home Economics"

[173,] "10120" "Technical"

[174,] "10120" "Visual Arts"

[175,] "10120" "General Arts"

[176,] "10120" ""

[177,] "10121" "Visual Arts"

[178,] "10121" "Business"

[179,] "10121" "Agriculture"

[180,] "10121" "General Arts"

[181,] "10121" "General Science"

[182,] "10122" "Business"

[183,] "10123" "Visual Arts"

[184,] "10123" "Home Economics"

[185,] "10123" "General Science"

[186,] "10123" "Business"

[187,] "10123" "General Arts"

[188,] "10124" "Business"

[189,] "10124" "Home Economics"

[190,] "10124" "General Arts"

[191,] "10125" "General Arts"

[192,] "10125" "Business"

[193,] "10125" "Visual Arts"

[194,] "10125" ""

[195,] "10126" "Business"

[196,] "10126" "General Arts"

[197,] "10126" "General Science"

[198,] "10128" "Visual Arts"

[199,] "10128" "Home Economics"

[200,] "10128" "General Arts"

[201,] "10128" "General Science"

[202,] "10128" "Agriculture"

[203,] "10128" ""

[204,] "10131" "General Arts"

[205,] "10131" "Business"

[206,] "10131" "Home Economics"

[207,] "10131" "Visual Arts"

[208,] "10133" "Home Economics"

[209,] "10133" "Business"

[210,] "10134" "Business"

[211,] "10135" "Home Economics"

[212,] "10135" "Business"

[213,] "10135" "General Arts"

[214,] "10138" "Business"

[215,] "10138" "General Arts"

[216,] "10138" "Visual Arts"

[217,] "10140" "General Arts"

[218,] "10140" "General Science"

[219,] "10141" "General Arts"

[220,] "10141" "General Science"

[221,] "10141" "Business"

[222,] "10142" "Business"

[223,] "10142" "General Arts"

[224,] "10142" "Visual Arts"

[225,] "10142" "General Science"

[226,] "10143" "General Arts"

[227,] "10144" "General Arts"

[228,] "10144" "General Science"

[229,] "10146" "Business"

[230,] "10146" "General Arts"

[231,] "10146" "General Science"

[232,] "10147" "General Science"

[233,] "10147" "Business"

[234,] "10148" "General Arts"

[235,] "10148" "Business"

[236,] "10149" "Visual Arts"

[237,] "10149" "Home Economics"

[238,] "10149" "General Arts"

[239,] "10149" "Business"

[240,] "10201" "Business"

[241,] "10201" "Home Economics"

[242,] "10201" "Visual Arts"

[243,] "10201" "General Arts"

[244,] "10201" "General Science"

[245,] "10201" ""

[246,] "10202" "Technical"

[247,] "10202" "Home Economics"

[248,] "10202" "General Arts"

[249,] "10202" "Visual Arts"

[250,] "10202" "General Science"

[251,] "10202" ""

[252,] "10203" "General Arts"

[253,] "10203" "Visual Arts"

[254,] "10203" "Home Economics"

[255,] "10203" "Agriculture"

[256,] "10203" ""

[257,] "10203" "Technical"

[258,] "10204" "Technical"

[259,] "10204" "Agriculture"

[260,] "10204" "Visual Arts"

[261,] "10204" "Home Economics"

[262,] "10204" "General Arts"

[263,] "10204" ""

[264,] "10205" "Home Economics"

[265,] "10205" "Visual Arts"

[266,] "10205" "General Arts"

[267,] "10205" "Business"

[268,] "10205" "General Science"

[269,] "10205" ""

[270,] "10206" "Visual Arts"

[271,] "10206" "Business"

[272,] "10206" "General Arts"

[273,] "10206" "General Science"

[274,] "10206" "Agriculture"

[275,] "10208" ""

[276,] "10208" "Visual Arts"

[277,] "10208" "General Arts"

[278,] "10210" "General Arts"

[279,] "10210" "Visual Arts"

[280,] "10210" "Home Economics"

[281,] "10210" "Business"

[282,] "10210" ""

[283,] "10212" "General Arts"

[284,] "10212" "Home Economics"

[285,] "10212" "Business"

[286,] "10212" ""

[287,] "10213" "Visual Arts"

[288,] "10213" "Home Economics"

[289,] "10213" "Business"

[290,] "10213" "General Arts"

[291,] "10213" "General Science"

[292,] "10214" "Home Economics"

[293,] "10214" "Technical"

[294,] "10214" "Business"

[295,] "10214" "General Arts"

[296,] "10214" "General Science"

[297,] "10215" "Business"

[298,] "10215" "General Arts"

[299,] "10216" "General Arts"

[300,] "10218" "Business"

[301,] "10218" "General Arts"

[302,] "10301" "Home Economics"

[303,] "10301" "Visual Arts"

[304,] "10301" "General Arts"

[305,] "10301" "Agriculture"

[306,] "10301" "General Science"

[307,] "10301" ""

[308,] "10302" "Technical"

[309,] "10302" "Business"

[310,] "10302" "Home Economics"

[311,] "10302" "General Arts"

[312,] "10302" "Agriculture"

[313,] "10302" ""

[314,] "10302" "Visual Arts"

[315,] "10401" "General Arts"

[316,] "10401" "Visual Arts"

[317,] "10401" "Home Economics"

[318,] "10401" "Business"

[319,] "10401" "Agriculture"

[320,] "10401" "General Science"

[321,] "10402" "Technical"

[322,] "10402" "Business"

[323,] "10402" "Agriculture"

[324,] "10402" "Visual Arts"

[325,] "10402" "Home Economics"

[326,] "10402" "General Arts"

[327,] "10402" ""

[328,] "10403" "General Arts"

[329,] "10403" "Business"

[330,] "10403" "Agriculture"

[331,] "10403" "Home Economics"

[332,] "10403" "General Science"

[333,] "10403" ""

[334,] "10501" "General Arts"

[335,] "10501" "Visual Arts"

[336,] "10501" "Home Economics"

[337,] "10501" "Business"

[338,] "10501" "General Science"

[339,] "10501" "Agriculture"

[340,] "10502" "Visual Arts"

[341,] "10502" "General Arts"

[342,] "10502" "General Science"

[343,] "10502" "Business"

[344,] "10502" "Agriculture"

[345,] "10502" ""

[346,] "10503" "Business"

[347,] "10503" "Home Economics"

[348,] "10503" "Agriculture"

[349,] "10503" "Visual Arts"

[350,] "10503" ""

[351,] "10503" "General Arts"

[352,] "10504" "General Arts"

[353,] "10504" "Visual Arts"

[354,] "10504" "Technical"

[355,] "10504" "Home Economics"

[356,] "10504" "Agriculture"

[357,] "10504" ""

[358,] "10506" "General Arts"

[359,] "10506" "Agriculture"

[360,] "10506" "Home Economics"

[361,] "10508" "General Arts"

[362,] "10509" "General Arts"

[363,] "10509" "Business"

[364,] "10509" "Visual Arts"

[365,] "10509" "Home Economics"

[366,] "10509" "Technical"

[367,] "10509" "General Science"

[368,] "10510" "General Arts"

[369,] "10510" "Business"

[370,] "10511" "Home Economics"

[371,] "20101" "Technical"

[372,] "20101" "Visual Arts"

[373,] "20101" "General Arts"

[374,] "20101" "General Science"

[375,] "20101" ""

[376,] "20102" "General Arts"

[377,] "20102" "General Science"

[378,] "20102" "Business"

[379,] "20102" "Visual Arts"

[380,] "20103" "General Arts"

[381,] "20103" "Business"

[382,] "20103" "Home Economics"

[383,] "20103" "General Science"

[384,] "20104" "Home Economics"

[385,] "20104" "General Arts"

[386,] "20104" "Visual Arts"

[387,] "20104" "Agriculture"

[388,] "20104" "Business"

[389,] "20104" "General Science"

[390,] "20105" "Agriculture"

[391,] "20105" "Home Economics"

[392,] "20105" "General Arts"

[393,] "20105" "Business"

[394,] "20105" "Visual Arts"

[395,] "20105" "General Science"

[396,] "20105" ""

[397,] "20106" "Home Economics"

[398,] "20106" "Business"

[399,] "20106" "General Arts"

[400,] "20106" "Visual Arts"

[401,] "20106" "Agriculture"

[402,] "20106" "General Science"

[403,] "20106" ""

[404,] "20107" "Business"

[405,] "20107" "Home Economics"

[406,] "20107" "General Arts"

[407,] "20107" "Visual Arts"

[408,] "20108" "General Arts"

[409,] "20108" "Business"

[410,] "20108" "Home Economics"

[411,] "20108" "General Science"

[412,] "20201" "Business"

[413,] "20201" "General Arts"

[414,] "20201" "Technical"

[415,] "20201" "General Science"

[416,] "20201" "Visual Arts"

[417,] "20201" ""

[418,] "20202" "Home Economics"

[419,] "20202" "General Arts"

[420,] "20202" "Agriculture"

[421,] "20202" "Business"

[422,] "20202" "Visual Arts"

[423,] "20202" ""

[424,] "20203" "Agriculture"

[425,] "20203" "Technical"

[426,] "20203" "Visual Arts"

[427,] "20203" "General Arts"

[428,] "20203" ""

[429,] "20204" "General Arts"

[430,] "20204" "Home Economics"

[431,] "20204" "Business"

[432,] "20204" ""

[433,] "20301" "General Arts"

[434,] "20301" "General Science"

[435,] "20301" "Business"

[436,] "20301" "Home Economics"

[437,] "20301" "Visual Arts"

[438,] "20302" "Visual Arts"

[439,] "20302" "General Arts"

[440,] "20302" "Home Economics"

[441,] "20302" "Agriculture"

[442,] "20302" "Business"

[443,] "20302" "General Science"

[444,] "20303" "Home Economics"

[445,] "20303" "Agriculture"

[446,] "20303" "Business"

[447,] "20303" "General Arts"

[448,] "20303" "General Science"

[449,] "20303" ""

[450,] "20304" "Visual Arts"

[451,] "20304" "General Arts"

[452,] "20304" "Home Economics"

[453,] "20304" "Agriculture"

[454,] "20304" ""

[455,] "20305" "Technical"

[456,] "20305" "Visual Arts"

[457,] "20305" "Home Economics"

[458,] "20305" "Agriculture"

[459,] "20305" "General Arts"

[460,] "20305" "General Science"

[461,] "20305" ""

[462,] "20401" "General Arts"

[463,] "20401" "Home Economics"

[464,] "20401" "General Science"

[465,] "20401" "Agriculture"

[466,] "20401" "Visual Arts"

[467,] "20401" "Business"

[468,] "20402" "Agriculture"

[469,] "20402" "Business"

[470,] "20402" "General Science"

[471,] "20402" "Home Economics"

[472,] "20402" "General Arts"

[473,] "20402" "Visual Arts"

[474,] "20403" "General Arts"

[475,] "20403" "Agriculture"

[476,] "20403" "Business"

[477,] "20403" "Home Economics"

[478,] "20403" "General Science"

[479,] "20403" ""

[480,] "20404" "Home Economics"

[481,] "20404" "General Arts"

[482,] "20404" "Technical"

[483,] "20404" "Visual Arts"

[484,] "20405" "Technical"

[485,] "20405" "Home Economics"

[486,] "20405" "Visual Arts"

[487,] "20405" "General Arts"

[488,] "20405" ""

[489,] "20406" "Visual Arts"

[490,] "20406" "General Arts"

[491,] "20406" "Home Economics"

[492,] "20406" "Business"

[493,] "20406" ""

[494,] "20407" "Visual Arts"

[495,] "20407" "Home Economics"

[496,] "20407" "Technical"

[497,] "20407" "Agriculture"

[498,] "20407" ""

[499,] "20407" "General Arts"

[500,] "20408" "Visual Arts"

[ reached getOption("max.print") -- omitted 2580 rows ]

> codeprogram<-codeprogram[!duplicated(codeprogram),]

> #Delete the duplicate combinations of schoolcode and program

> dim(codeprogram)

[1] 3080 2

> #3080 2

>

> n1<-nrow(codeprogram)

> for (i in 1:n1){

+

+ if(codeprogram[i,2]==""){

+ codeprogram[i,]=NA

+ }

+ }

> #Subsititute program="" to NA to delete the empty column

>

> codeprogram<-codeprogram[complete.cases(codeprogram),]

> dim(codeprogram)

[1] 2773 2

> # The number of choices is 2773

>

>

> misscore<-matrix(datstu[,"score"],nrow=340823,ncol=1)

> length(misscore[is.na(misscore)])

[1] 179887

> # The number of missing test score is 179887

>

> #apply to the same school

> ppnum<-as.matrix(rep(1:6,340823))

> ppchoice<-as.vector(as.matrix(datstu[,5:10]))

> ppchoice<-as.matrix(ppchoice)

> ppsch<-cbind(ppchoice,ppnum)

> ppsch<-ppsch[complete.cases(ppsch),]

> ppsame<-ppsch[order(ppsch[,1]),]

> a<-as.vector(table(ppsame[,1]))

> a

[1] 5891 1958 8419 2474 1496 4015 4075 4181 9151 3017 2242 2733 6019 4150 7484 7934 15987 11331

[19] 7766 2998 18 118 50 77 80 6282 72 14 1 114 61 14 149 1898 4 11

[37] 26 3 21 31 7937 4349 5948 2918 8164 59 4 6237 14 2585 128 3 1 24

[55] 3355 8494 5092 1666 2712 5479 3476 2655 3541 5 16 73 7 3 6114 2728 5804 7069

[73] 15223 7021 75 15 4388 5716 1654 1248 3511 4944 4540 6740 10352 5423 6967 4129 658 1795

[91] 2553 3268 2442 1475 1855 6969 3869 5243 7820 5064 12875 2919 4495 4011 1862 163 675 3088

[109] 1870 1909 8523 605 3350 2314 1331 2099 661 4548 5899 1798 5746 10086 9841 8111 2295 1370

[127] 1824 5152 5292 1410 1758 7319 4976 1811 3485 6398 4807 1631 6781 1362 1259 6536 2123 223

[145] 38 2425 8016 3318 3890 3445 2099 1881 1837 1865 1738 3136 6050 2061 5065 7162 6018 7046

[163] 2374 2971 5 3295 2094 7237 6229 3330 5010 4150 1298 72 8694 7616 5557 8813 8042 3518

[181] 5660 49 53 5860 5124 2750 9142 4257 5283 7511 9242 4991 9305 33 42 6672 4917 3318

[199] 6302 3723 44 9979 7204 1877 3112 1730 2418 4824 5127 1856 2032 4679 13886 2408 7676 3523

[217] 2666 9400 5129 2428 8333 3773 5142 1832 2017 3850 1321 5265 5949 2951 4407 1965 1874 654

[235] 7454 4879 3791 972 5534 10326 11228 7025 3313 830 1526 1433 2272 1737 10800 2430 3102 4688

[253] 9071 5600 7377 2616 3578 2915 2535 5068 9076 2528 3704 29 48 355 10 36 7 44

[271] 14 60 11 1 13 1 29 6 2383 19 5 21 18 3 2652 13121 3347 13999

[289] 5294 3934 1883 9433 5204 1879 1077 474 6955 2874 935 1676 3796 3418 4191 4650 16215 9731

[307] 3415 2183 7399 3070 5674 10642 5028 6696 4775 573 2320 5341 5248 3780 1427 2273 745 3689

[325] 8372 4218 1589 2947 3718 1403 4145 7696 5202 4180 789 1558 4326 1003 8390 1439 2519 13114

[343] 11629 14833 3517 3048 1115 5584 1202 11419 15834 2319 1700 3164 3319 2425 7800 5041 1727 1900

[361] 25 6889 1250 3781 1254 756 2772 7482 3429 8165 2495 1291 4254 1149 2063 6517 2648 423

[379] 699 1967 5366 694 1932 695 518 7129 6062 1714 1476 382 7 8082 2225 4824 1199 12198

[397] 1109 2304 654 1737 1987 565 2690 4085 5087 813 5090 3490 2476 2506 4788 591 7684 3115

[415] 1806 1456 439 935 2935 587 599 1200 751 84 16 1269 2126 810 1356 522 1 3556

[433] 807 999 7791 1622 709 6295 5477 4949 2334 4725 1291 1881 1138 512 1600 5362 5642 3541

[451] 4666 1648 1610 860 2033 2909 3278 5835 1616 4239 1978 850 1485 896 604 6758 1818 124

[469] 2422 746 322 1702 1106 3831 425 870 855 1028 3200 2431 2379 887 116 3515 877 9338

[487] 7382 9857 10976 7575 7030 1140 6556 3962 1498 4335 1611 7888 3163 3607 5303 2452 7101 902

[505] 1389 5428 2068 3920 556 5106 1577 370 1336 1437 1289 402 1361 1769 4630 2737 7285 3041

[523] 2828 5753 2941 2226 883 6098 4008 4273 7411 626 5776 6482 1282 1226 1054 500 393 192

[541] 4842 7848 2225 5790 27 3320 4911 1811 1521 3479 1946 1729 1836 2019 1612 2088 3061 2071

[559] 3281 3064 5651 1420 667 243 190 67 44 143 29 70 72 98 34 198 92 112

[577] 14 120 74 13 175 1977 101 35 31 868 5 1012 51 52 103 52 72 358

[595] 443 620 44 899 166 67 2386 213 1361 33 2984 43 83 1294 4253 105 130 64

[613] 170 65 6 164 49 33 35 34 37 4 14 68 1320 1921 3 372 780 146

[631] 6 158 66 1355 2450 400 10 12 27 19

>

>

> choices<-as.matrix(datstu[,5:10])

> dim(choices[complete.cases(choices),])

[1] 323089 6

> # The number of students who apply to less than 6 choices is 323089