Project Assignment 3

Isaac Coffie

November 11, 2015

**Does parent involvement in children's education an explanation for any observed improvement in student's academic performances?**

## 1. Setup workspace and call in the data set(s)

#set working directory for my project  
setwd("C:/StatsProject")  
  
#clear all variables from the previous environment  
rm(list = ls())  
  
#load dataset from the file directory  
load("C:/StatsProject/addhealth\_W1.RDATA")  
load("C:/StatsProject/addhealth\_W4.RDATA")

## 2. Create a data subset for my project

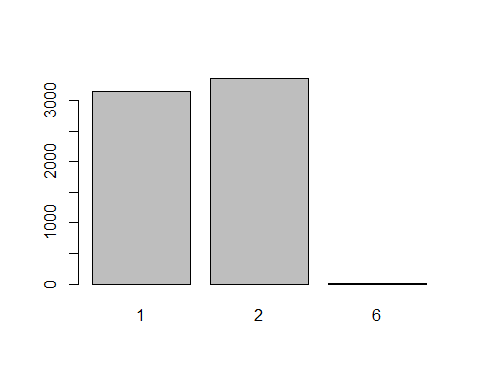
#Selecting variables of interest to my project  
addhealth\_orig<-AddHealth  
my\_vars<-c("AID","BIO\_SEX","age","H1SE4","H1ED11","H1ED12","H1WP5","H1RM1","H1WP17H")  
my\_addhealth<-addhealth\_orig[my\_vars]  
  
#variables from addHealth wave 4   
wave4\_vars<-c("aid","h4da1")  
  
#create a subset with these two variables  
addhealth4\_subset<-addhealth\_public4[wave4\_vars]  
  
#rename AID in addHealth wave 1 to "aid"  
names(my\_addhealth)[1]<-"aid"  
  
#merge the two wave  
my\_addhealth<-merge(my\_addhealth, addhealth4\_subset, by="aid", all = TRUE)

## 3. Label variables and run frequency tables

#form a frequency table   
library(descr)

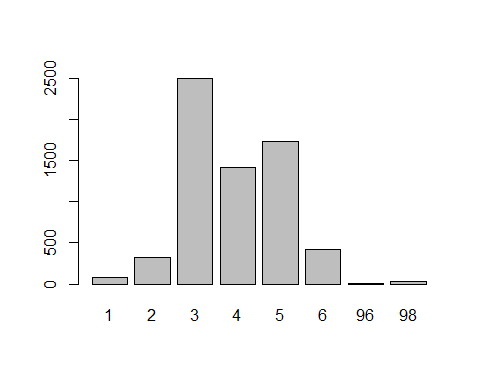
## Warning: package 'descr' was built under R version 3.2.2

freq(as.ordered(my\_addhealth$BIO\_SEX))



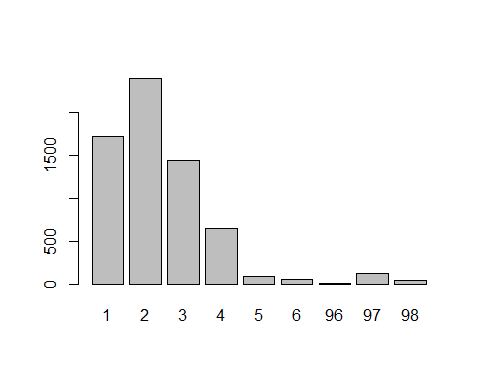
## as.ordered(my\_addhealth$BIO\_SEX)   
## Frequency Percent Cum Percent  
## 1 3147 48.38561 48.39  
## 2 3356 51.59902 99.98  
## 6 1 0.01538 100.00  
## Total 6504 100.00000

freq(as.ordered(my\_addhealth$H1SE4))



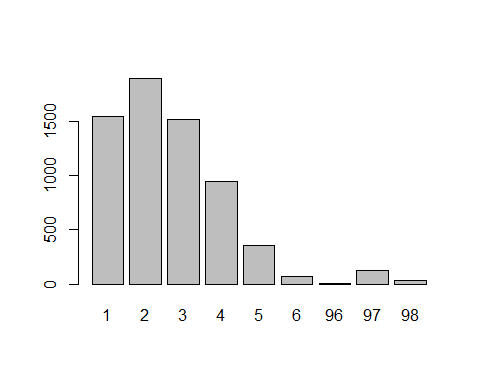
## as.ordered(my\_addhealth$H1SE4)   
## Frequency Percent Cum Percent  
## 1 77 1.18389 1.184  
## 2 322 4.95080 6.135  
## 3 2502 38.46863 44.603  
## 4 1420 21.83272 66.436  
## 5 1735 26.67589 93.112  
## 6 418 6.42681 99.539  
## 96 5 0.07688 99.616  
## 98 25 0.38438 100.000  
## Total 6504 100.00000

freq(as.ordered(my\_addhealth$H1ED11))



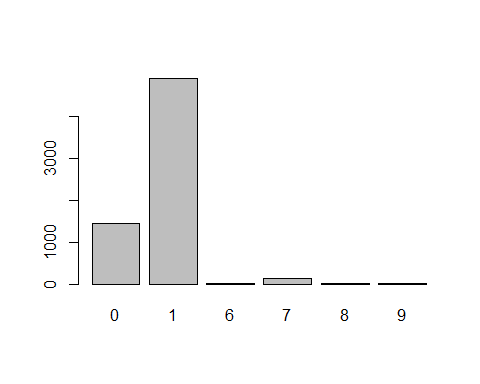
## as.ordered(my\_addhealth$H1ED11)   
## Frequency Percent Cum Percent  
## 1 1712 26.32226 26.32  
## 2 2389 36.73124 63.05  
## 3 1435 22.06335 85.12  
## 4 647 9.94772 95.06  
## 5 87 1.33764 96.40  
## 6 58 0.89176 97.29  
## 96 5 0.07688 97.37  
## 97 128 1.96802 99.34  
## 98 43 0.66113 100.00  
## Total 6504 100.00000

freq(as.ordered(my\_addhealth$H1ED12))



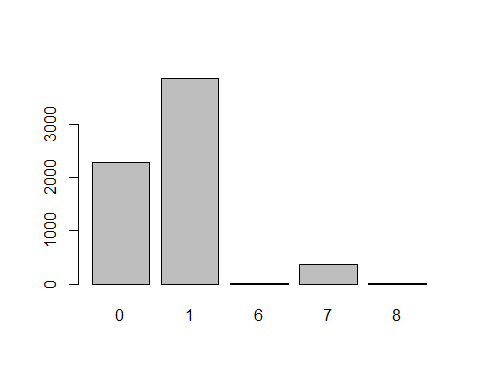
## as.ordered(my\_addhealth$H1ED12)   
## Frequency Percent Cum Percent  
## 1 1552 23.8622 23.86  
## 2 1899 29.1974 53.06  
## 3 1521 23.3856 76.45  
## 4 950 14.6064 91.05  
## 5 353 5.4274 96.48  
## 6 69 1.0609 97.54  
## 96 4 0.0615 97.60  
## 97 128 1.9680 99.57  
## 98 28 0.4305 100.00  
## Total 6504 100.0000

freq(as.ordered(my\_addhealth$H1WP5))



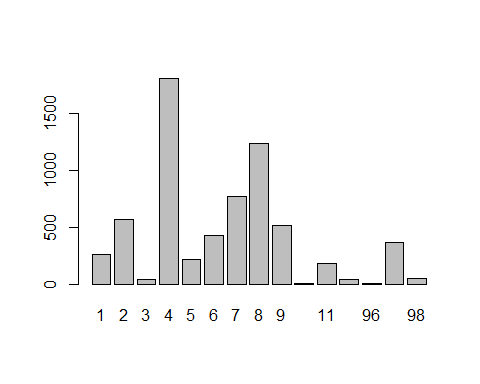
## as.ordered(my\_addhealth$H1WP5)   
## Frequency Percent Cum Percent  
## 0 1451 22.30935 22.31  
## 1 4911 75.50738 97.82  
## 6 2 0.03075 97.85  
## 7 131 2.01415 99.86  
## 8 8 0.12300 99.98  
## 9 1 0.01538 100.00  
## Total 6504 100.00000

freq(as.ordered(my\_addhealth$H1WP17H))



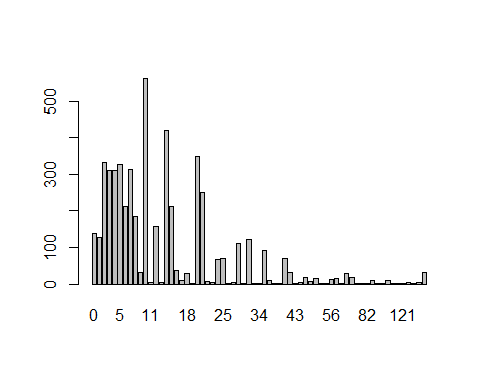
## as.ordered(my\_addhealth$H1WP17H)   
## Frequency Percent Cum Percent  
## 0 2272 34.93235 34.93  
## 1 3851 59.20972 94.14  
## 6 6 0.09225 94.23  
## 7 370 5.68881 99.92  
## 8 5 0.07688 100.00  
## Total 6504 100.00000

freq(as.ordered(my\_addhealth$H1RM1))



## as.ordered(my\_addhealth$H1RM1)   
## Frequency Percent Cum Percent  
## 1 263 4.0437 4.044  
## 2 568 8.7331 12.777  
## 3 41 0.6304 13.407  
## 4 1811 27.8444 41.252  
## 5 217 3.3364 44.588  
## 6 426 6.5498 51.138  
## 7 770 11.8389 62.977  
## 8 1241 19.0806 82.057  
## 9 512 7.8721 89.929  
## 10 7 0.1076 90.037  
## 11 179 2.7522 92.789  
## 12 42 0.6458 93.435  
## 96 4 0.0615 93.496  
## 97 370 5.6888 99.185  
## 98 53 0.8149 100.000  
## Total 6504 100.0000

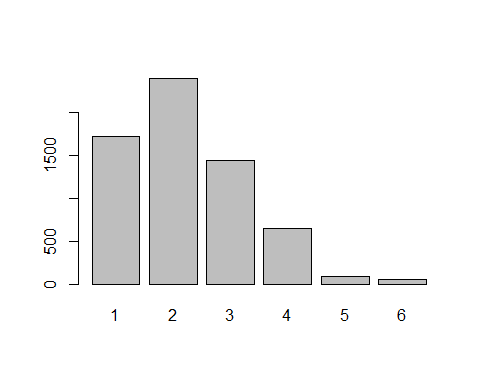
freq(as.ordered(my\_addhealth$h4da1))



## as.ordered(my\_addhealth$h4da1)   
## Frequency Percent Valid Percent Cum Percent  
## 0 137 2.10640 2.67944 2.679  
## 1 128 1.96802 2.50342 5.183  
## 2 332 5.10455 6.49325 11.676  
## 3 311 4.78167 6.08253 17.759  
## 4 311 4.78167 6.08253 23.841  
## 5 327 5.02768 6.39546 30.237  
## 6 213 3.27491 4.16585 34.403  
## 7 314 4.82780 6.14121 40.544  
## 8 186 2.85978 3.63779 44.181  
## 9 31 0.47663 0.60630 44.788  
## 10 562 8.64084 10.99159 55.779  
## 11 4 0.06150 0.07823 55.858  
## 12 158 2.42927 3.09016 58.948  
## 13 4 0.06150 0.07823 59.026  
## 14 421 6.47294 8.23391 67.260  
## 15 211 3.24416 4.12674 71.387  
## 16 36 0.55351 0.70409 72.091  
## 17 9 0.13838 0.17602 72.267  
## 18 29 0.44588 0.56718 72.834  
## 19 1 0.01538 0.01956 72.854  
## 20 350 5.38130 6.84530 79.699  
## 21 251 3.85916 4.90906 84.608  
## 22 7 0.10763 0.13691 84.745  
## 23 3 0.04613 0.05867 84.803  
## 24 68 1.04551 1.32994 86.133  
## 25 71 1.09164 1.38862 87.522  
## 26 2 0.03075 0.03912 87.561  
## 27 4 0.06150 0.07823 87.639  
## 28 110 1.69127 2.15138 89.791  
## 29 1 0.01538 0.01956 89.810  
## 30 123 1.89114 2.40563 92.216  
## 32 1 0.01538 0.01956 92.235  
## 34 2 0.03075 0.03912 92.275  
## 35 91 1.39914 1.77978 94.054  
## 36 11 0.16913 0.21514 94.270  
## 37 1 0.01538 0.01956 94.289  
## 38 1 0.01538 0.01956 94.309  
## 40 70 1.07626 1.36906 95.678  
## 42 31 0.47663 0.60630 96.284  
## 43 1 0.01538 0.01956 96.304  
## 45 5 0.07688 0.09779 96.401  
## 48 18 0.27675 0.35204 96.753  
## 49 6 0.09225 0.11735 96.871  
## 50 16 0.24600 0.31293 97.184  
## 52 2 0.03075 0.03912 97.223  
## 54 1 0.01538 0.01956 97.242  
## 56 13 0.19988 0.25425 97.497  
## 60 15 0.23063 0.29337 97.790  
## 63 2 0.03075 0.03912 97.829  
## 70 28 0.43050 0.54762 98.377  
## 72 17 0.26138 0.33249 98.709  
## 75 1 0.01538 0.01956 98.729  
## 80 1 0.01538 0.01956 98.748  
## 82 1 0.01538 0.01956 98.768  
## 84 9 0.13838 0.17602 98.944  
## 85 1 0.01538 0.01956 98.963  
## 98 1 0.01538 0.01956 98.983  
## 100 10 0.15375 0.19558 99.179  
## 105 1 0.01538 0.01956 99.198  
## 110 1 0.01538 0.01956 99.218  
## 121 1 0.01538 0.01956 99.237  
## 140 4 0.06150 0.07823 99.315  
## 150 1 0.01538 0.01956 99.335  
## 996 3 0.04613 0.05867 99.394  
## 998 31 0.47663 0.60630 100.000  
## NA's 1391 21.38684   
## Total 6504 100.00000 100.00000

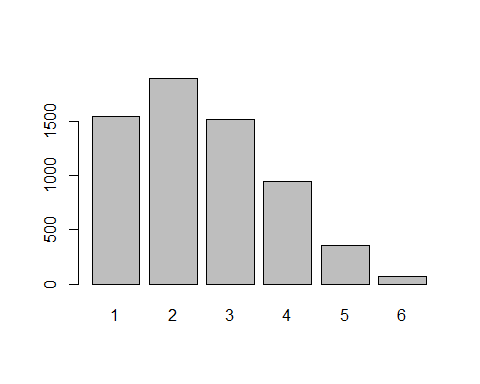
## 4. Data management

#(C)Replace "don't know", "Legitimate skip" and "refused" with NA  
  
#Grade in English and Art  
my\_addhealth$H1ED11[my\_addhealth$H1ED11==96 | my\_addhealth$H1ED11==97 |my\_addhealth$H1ED11==98]<-NA  
freq(as.ordered(my\_addhealth$H1ED11))



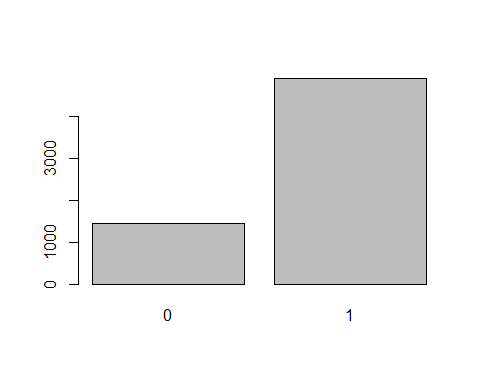
## as.ordered(my\_addhealth$H1ED11)   
## Frequency Percent Valid Percent Cum Percent  
## 1 1712 26.3223 27.0544 27.05  
## 2 2389 36.7312 37.7528 64.81  
## 3 1435 22.0633 22.6770 87.48  
## 4 647 9.9477 10.2244 97.71  
## 5 87 1.3376 1.3748 99.08  
## 6 58 0.8918 0.9166 100.00  
## NA's 176 2.7060   
## Total 6504 100.0000 100.0000

#Grade in Mathematics  
my\_addhealth$H1ED12[my\_addhealth$H1ED12==96 | my\_addhealth$H1ED12==97 |my\_addhealth$H1ED12==98]<-NA  
freq(as.ordered(my\_addhealth$H1ED12))



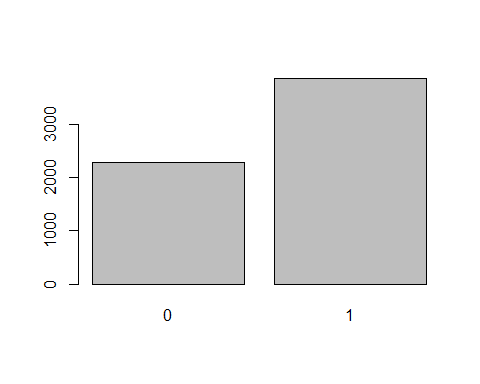
## as.ordered(my\_addhealth$H1ED12)   
## Frequency Percent Valid Percent Cum Percent  
## 1 1552 23.862 24.464 24.46  
## 2 1899 29.197 29.934 54.40  
## 3 1521 23.386 23.975 78.37  
## 4 950 14.606 14.975 93.35  
## 5 353 5.427 5.564 98.91  
## 6 69 1.061 1.088 100.00  
## NA's 160 2.460   
## Total 6504 100.000 100.000

#Parent decision on when to watch tv programs  
my\_addhealth$H1WP5[my\_addhealth$H1WP5==6 | my\_addhealth$H1WP5==7 |my\_addhealth$H1WP5==8 |my\_addhealth$H1WP5==9]<-NA  
freq(as.ordered(my\_addhealth$H1WP5))



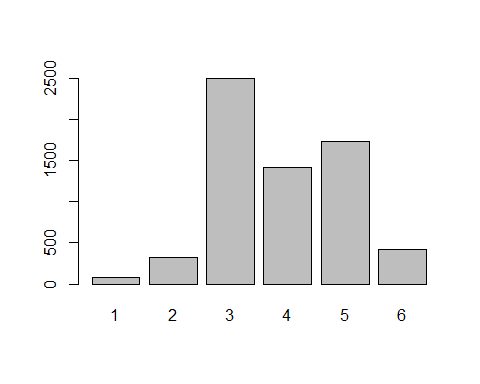
## as.ordered(my\_addhealth$H1WP5)   
## Frequency Percent Valid Percent Cum Percent  
## 0 1451 22.309 22.81 22.81  
## 1 4911 75.507 77.19 100.00  
## NA's 142 2.183   
## Total 6504 100.000 100.00

#conversation with parent on school work or grade  
my\_addhealth$H1WP17H[my\_addhealth$H1WP17H==6 | my\_addhealth$H1WP17H==7 | my\_addhealth$H1WP17H==8]<-NA  
freq(as.ordered(my\_addhealth$H1WP17H))



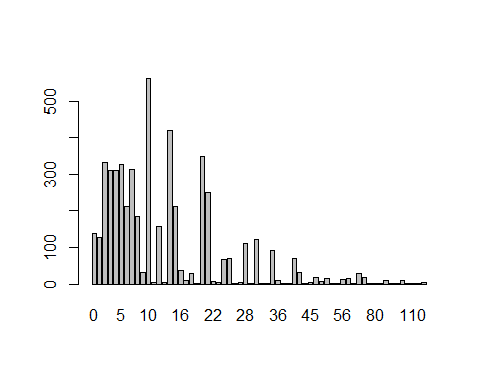
## as.ordered(my\_addhealth$H1WP17H)   
## Frequency Percent Valid Percent Cum Percent  
## 0 2272 34.932 37.11 37.11  
## 1 3851 59.210 62.89 100.00  
## NA's 381 5.858   
## Total 6504 100.000 100.00

#how intelligent are you compared to your peers  
my\_addhealth$H1SE4[my\_addhealth$H1SE4==96 | my\_addhealth$H1SE4==98]<-NA  
freq(as.ordered(my\_addhealth$H1SE4))



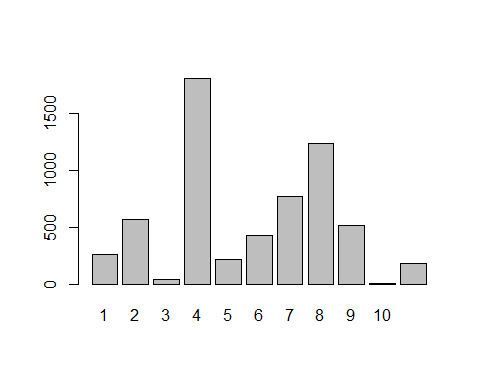
## as.ordered(my\_addhealth$H1SE4)   
## Frequency Percent Valid Percent Cum Percent  
## 1 77 1.1839 1.189 1.189  
## 2 322 4.9508 4.974 6.163  
## 3 2502 38.4686 38.647 44.810  
## 4 1420 21.8327 21.934 66.744  
## 5 1735 26.6759 26.800 93.543  
## 6 418 6.4268 6.457 100.000  
## NA's 30 0.4613   
## Total 6504 100.0000 100.000

#Number of hours spent on watching TV and videos  
my\_addhealth$h4da1[my\_addhealth$h4da1==996 | my\_addhealth$h4da1==998 |my\_addhealth$h4da1==144 | my\_addhealth$h4da1==150]<-NA  
freq(as.ordered(my\_addhealth$h4da1))



## as.ordered(my\_addhealth$h4da1)   
## Frequency Percent Valid Percent Cum Percent  
## 0 137 2.10640 2.69791 2.698  
## 1 128 1.96802 2.52068 5.219  
## 2 332 5.10455 6.53801 11.757  
## 3 311 4.78167 6.12446 17.881  
## 4 311 4.78167 6.12446 24.006  
## 5 327 5.02768 6.43954 30.445  
## 6 213 3.27491 4.19456 34.640  
## 7 314 4.82780 6.18354 40.823  
## 8 186 2.85978 3.66286 44.486  
## 9 31 0.47663 0.61048 45.096  
## 10 562 8.64084 11.06735 56.164  
## 11 4 0.06150 0.07877 56.243  
## 12 158 2.42927 3.11146 59.354  
## 13 4 0.06150 0.07877 59.433  
## 14 421 6.47294 8.29067 67.724  
## 15 211 3.24416 4.15518 71.879  
## 16 36 0.55351 0.70894 72.588  
## 17 9 0.13838 0.17724 72.765  
## 18 29 0.44588 0.57109 73.336  
## 19 1 0.01538 0.01969 73.356  
## 20 350 5.38130 6.89248 80.248  
## 21 251 3.85916 4.94289 85.191  
## 22 7 0.10763 0.13785 85.329  
## 23 3 0.04613 0.05908 85.388  
## 24 68 1.04551 1.33911 86.727  
## 25 71 1.09164 1.39819 88.125  
## 26 2 0.03075 0.03939 88.165  
## 27 4 0.06150 0.07877 88.243  
## 28 110 1.69127 2.16621 90.410  
## 29 1 0.01538 0.01969 90.429  
## 30 123 1.89114 2.42221 92.852  
## 32 1 0.01538 0.01969 92.871  
## 34 2 0.03075 0.03939 92.911  
## 35 91 1.39914 1.79204 94.703  
## 36 11 0.16913 0.21662 94.919  
## 37 1 0.01538 0.01969 94.939  
## 38 1 0.01538 0.01969 94.959  
## 40 70 1.07626 1.37850 96.337  
## 42 31 0.47663 0.61048 96.948  
## 43 1 0.01538 0.01969 96.967  
## 45 5 0.07688 0.09846 97.066  
## 48 18 0.27675 0.35447 97.420  
## 49 6 0.09225 0.11816 97.538  
## 50 16 0.24600 0.31508 97.853  
## 52 2 0.03075 0.03939 97.893  
## 54 1 0.01538 0.01969 97.913  
## 56 13 0.19988 0.25601 98.169  
## 60 15 0.23063 0.29539 98.464  
## 63 2 0.03075 0.03939 98.503  
## 70 28 0.43050 0.55140 99.055  
## 72 17 0.26138 0.33478 99.390  
## 75 1 0.01538 0.01969 99.409  
## 80 1 0.01538 0.01969 99.429  
## 82 1 0.01538 0.01969 99.449  
## 84 9 0.13838 0.17724 99.626  
## 85 1 0.01538 0.01969 99.646  
## 98 1 0.01538 0.01969 99.665  
## 100 10 0.15375 0.19693 99.862  
## 105 1 0.01538 0.01969 99.882  
## 110 1 0.01538 0.01969 99.902  
## 121 1 0.01538 0.01969 99.921  
## 140 4 0.06150 0.07877 100.000  
## NA's 1426 21.92497   
## Total 6504 100.00000 100.00000

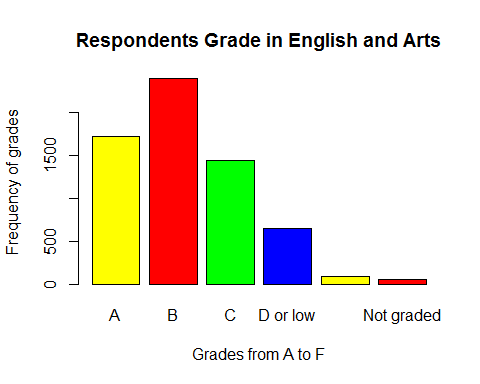
#Educational background of parent  
my\_addhealth$H1RM1[my\_addhealth$H1RM1==12 | my\_addhealth$H1RM1==96 | my\_addhealth$H1RM1==97 |my\_addhealth$H1RM1==98]<-NA  
freq(as.ordered(my\_addhealth$H1RM1))



## as.ordered(my\_addhealth$H1RM1)   
## Frequency Percent Valid Percent Cum Percent  
## 1 263 4.0437 4.3579 4.358  
## 2 568 8.7331 9.4118 13.770  
## 3 41 0.6304 0.6794 14.449  
## 4 1811 27.8444 30.0083 44.457  
## 5 217 3.3364 3.5957 48.053  
## 6 426 6.5498 7.0588 55.112  
## 7 770 11.8389 12.7589 67.871  
## 8 1241 19.0806 20.5634 88.434  
## 9 512 7.8721 8.4838 96.918  
## 10 7 0.1076 0.1160 97.034  
## 11 179 2.7522 2.9660 100.000  
## NA's 469 7.2109   
## Total 6504 100.0000 100.0000

## 5. Univariate graphing

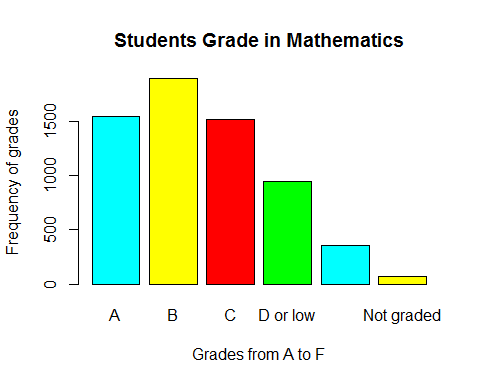
#A frequency distribution table and a bar graph for categorical variables  
  
#Grade in English  
freq(as.ordered(my\_addhealth$H1ED11), names=c("A","B","C","D or low","No Exams","Not graded"), main="Respondents Grade in English and Arts", xlab="Grades from A to F", ylab="Frequency of grades", col = c("yellow","red","green","blue"))



## as.ordered(my\_addhealth$H1ED11)   
## Frequency Percent Valid Percent Cum Percent  
## 1 1712 26.3223 27.0544 27.05  
## 2 2389 36.7312 37.7528 64.81  
## 3 1435 22.0633 22.6770 87.48  
## 4 647 9.9477 10.2244 97.71  
## 5 87 1.3376 1.3748 99.08  
## 6 58 0.8918 0.9166 100.00  
## NA's 176 2.7060   
## Total 6504 100.0000 100.0000

Judging by the bar plot shown above, we can clearly conclude that the category of gradepoint award which had the highest frequency is Grade "B" followed by "A" then to "C". However, the grade point with labels 'D or below', 'Not graded' and 'No exams' had the least frequency was.

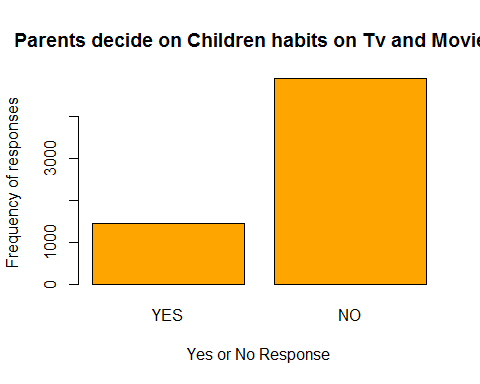
#Grade in Mathematics  
freq(as.ordered(my\_addhealth$H1ED12), name=c("A","B","C", "D or low", "No Exams", "Not graded"),main="Students Grade in Mathematics", xlab="Grades from A to F", ylab="Frequency of grades", col = c("cyan","yellow","red","green"))



## as.ordered(my\_addhealth$H1ED12)   
## Frequency Percent Valid Percent Cum Percent  
## 1 1552 23.862 24.464 24.46  
## 2 1899 29.197 29.934 54.40  
## 3 1521 23.386 23.975 78.37  
## 4 950 14.606 14.975 93.35  
## 5 353 5.427 5.564 98.91  
## 6 69 1.061 1.088 100.00  
## NA's 160 2.460   
## Total 6504 100.000 100.000

The bar plot above reveals that the category of grade awards which had the highest frequency is Grade "B" followed by "A" then "C". However, the grade point with the least frequency was D or below, Not graded and others.

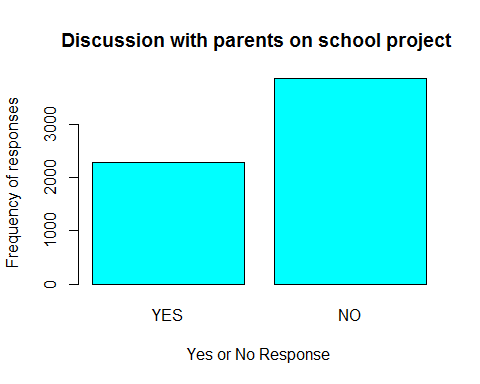
#parents controlling their children on what tv programs to watch  
freq(as.ordered(my\_addhealth$H1WP5), name=c("YES","NO"), main="Parents decide on Children habits on Tv and Movies ", xlab="Yes or No Response", ylab="Frequency of responses",col=c("orange"))



## as.ordered(my\_addhealth$H1WP5)   
## Frequency Percent Valid Percent Cum Percent  
## 0 1451 22.309 22.81 22.81  
## 1 4911 75.507 77.19 100.00  
## NA's 142 2.183   
## Total 6504 100.000 100.00

using the graph above, majority (more than 4000 people) said their parents do not set rules for them on what television programs and movies they should watch.On the other hand, less than 2000 responded that their parents control their watching of Televisoin programs.

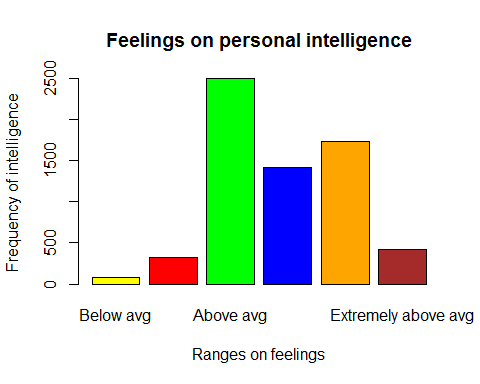
#discussion with parent on school work  
freq(as.ordered(my\_addhealth$H1WP17H), name=c("YES","NO"), main="Discussion with parents on school project ", xlab="Yes or No Response", ylab="Frequency of responses", col=c("cyan"))



## as.ordered(my\_addhealth$H1WP17H)   
## Frequency Percent Valid Percent Cum Percent  
## 0 2272 34.932 37.11 37.11  
## 1 3851 59.210 62.89 100.00  
## NA's 381 5.858   
## Total 6504 100.000 100.00

From the figure above, most of the respondents(more than 3500) said they do not have a regular discussion with their parents on their school projects, grades and assignments.

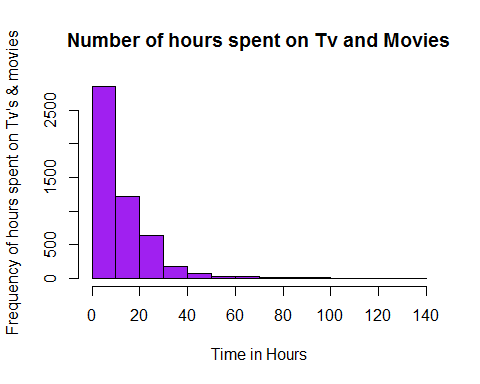
#how intelligent are you compared to your peers  
freq(as.ordered(my\_addhealth$H1SE4), names=c("Below avg","Slightly below avg","Above avg","Slightly above avg","Moderately above avg","Extremely above avg"), main="Feelings on personal intelligence", xlab="Ranges on feelings", ylab="Frequency of intelligence", col = c("yellow","red","green","blue","orange","brown"))



## as.ordered(my\_addhealth$H1SE4)   
## Frequency Percent Valid Percent Cum Percent  
## 1 77 1.1839 1.189 1.189  
## 2 322 4.9508 4.974 6.163  
## 3 2502 38.4686 38.647 44.810  
## 4 1420 21.8327 21.934 66.744  
## 5 1735 26.6759 26.800 93.543  
## 6 418 6.4268 6.457 100.000  
## NA's 30 0.4613   
## Total 6504 100.0000 100.000

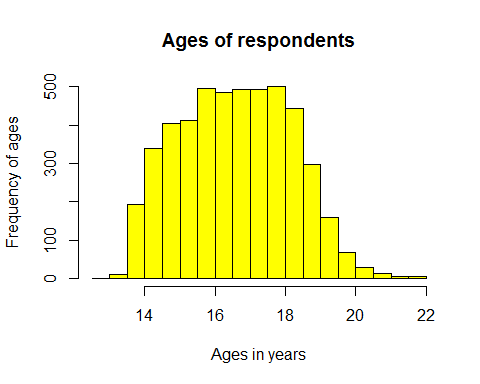
The diagram above shows that the category of intelligence feeling that had the highest frequecny were reported to be the ones above average, followed by moderately above average and slightly above average. However, people who said their intelligence level was below that of their peers was very low.

#Histogram for quantitative variables  
  
#Number of hours spent on watching TV and videos  
hist(my\_addhealth$h4da1, main = "Number of hours spent on Tv and Movies", xlab="Time in Hours", ylab="Frequency of hours spent on Tv's & movies", col=c("purple"))



From the above diagram, the number of hours spent on watching movies and Tv was concentrated between the hours of 0 to 40. Again, the data was heavily skewed to the left after the 60th hours of watching movies and Tv 's was recorded.

#Ages of respondent  
hist(my\_addhealth$age, main = "Ages of respondents", xlab="Ages in years", ylab="Frequency of ages", col = c("yellow"))

 From the figure above, the distribution is almost close to a symmetrical distribution with most of the ages centered at 16.5 years. Also, the variation between the data seems to be very small enough.

## 6. Summary statistics (quantitative variables only)

#Mean and standard deviation of respondent'sage  
mean(my\_addhealth$age,na.rm = TRUE)

## [1] 16.54155

sd(my\_addhealth$age,na.rm = TRUE)

## [1] 1.581993

The mean age of respondents is 16.541 and the distance between the mean age and the individual age is 1.581. Thus, the individual ages of respondents are not too far away from the mean ages.

##Mean and standard deviation of the number of hours spent watching television  
mean(my\_addhealth$h4da1,na.rm = TRUE)

## [1] 13.56833

sd(my\_addhealth$h4da1,na.rm = TRUE)

## [1] 13.68613

The mean number of hours spent on watching television is 13.569 and the variation of the individual data from the mean is 13.686. This means the there is a wider distance between the mean and the actual data of respondents.

#Summary of ages and number of number of hours spent on watching television  
summary(my\_addhealth$h4da1)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 5.00 10.00 13.57 20.00 140.00 1426

summary(my\_addhealth$age)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 12.68 15.26 16.51 16.54 17.76 21.93 1667

The median age is 16.51 years. The IQR of the ages of respondents = 3rd Qu - 1st Qu = (17.76- 15.26) years = 2.5 years.

Also, the IQR of the number of hours spent on watching televison and movies is (20-5) = 15 hours. The median number of hours spent on movies is 10 hours.

## 7. Bivariate output and graphing

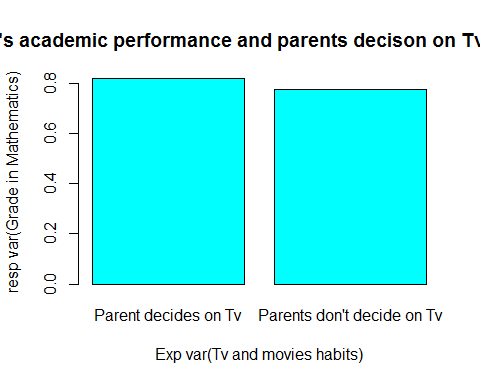
#create a new variable mycoodebook$intel to collapse response variables   
my\_addhealth$intel<-rep(NA, nrow(my\_addhealth))  
   
my\_addhealth$intel[my\_addhealth$H1SE4 == 1 | my\_addhealth$H1SE4 == 2 | my\_addhealth$H1SE4 == 3]<-1  
my\_addhealth$intel[my\_addhealth$H1SE4 == 4 | my\_addhealth$H1SE4 == 5 | my\_addhealth$H1SE4 == 6]<-2  
  
  
#cOLLAPSE VARIABLE H1WP5 into two categorical responses  
my\_addhealth$parent<-rep(NA, nrow(my\_addhealth))  
my\_addhealth$parent[my\_addhealth$H1WP5 == 0]<-1  
my\_addhealth$parent[my\_addhealth$H1WP5 == 1]<-2  
  
#collapse variable H1ED11 (english grades)into a new variable my\_addhealth$english  
my\_addhealth$english<-rep(NA, nrow(my\_addhealth))  
my\_addhealth$english[my\_addhealth$H1ED11 == 1 | my\_addhealth$H1ED11 == 2 | my\_addhealth$H1ED11 == 3]<-1  
my\_addhealth$english[my\_addhealth$H1ED11 == 4 | my\_addhealth$H1ED11 == 5 | my\_addhealth$H1ED11 == 6]<-2

#collapse variable H1ED12 (maths grades)into a new variable my\_addhealth$maths  
my\_addhealth$maths<-rep(NA, nrow(my\_addhealth))  
   
my\_addhealth$maths[my\_addhealth$H1ED12 == 1 | my\_addhealth$H1ED12 == 2 | my\_addhealth$H1ED12 == 3] <-1  
my\_addhealth$maths[my\_addhealth$H1ED12 == 4 | my\_addhealth$H1ED12 == 5 | my\_addhealth$H1ED12 == 6]<-2  
  
  
#Collapse variable H1WP17H (concersation with parent) into two responses  
my\_addhealth$chat<-rep(NA, nrow(my\_addhealth))  
my\_addhealth$chat[my\_addhealth$H1WP17H == 0]<-1  
my\_addhealth$chat[my\_addhealth$H1WP17H == 1]<-2

#create a table to illustrate both the response variable(grades in maths) and explanatory variable (parents decision on what Tv programs kids should watch )  
newtable<-table(my\_addhealth$maths, my\_addhealth$parent)  
tab1\_colProp <-round( prop.table(newtable, 2),3)   
tab1\_colProp

##   
## 1 2  
## 1 0.818 0.777  
## 2 0.182 0.223

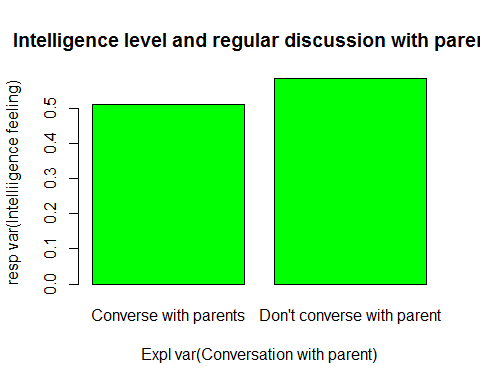
barplot(tab1\_colProp[1,], col=c("cyan"), names=c("Parent decides on Tv","Parents don't decide on Tv"), main="Student's academic performance and parents decison on Tv's & movies", xlab="Exp var(Tv and movies habits)", ylab = "resp var(Grade in Mathematics)")

 Using the barplot above, we can verify that there is a 81.8% for students to pass mathematics with a semester grade of A and B given that their parents set rules for them as to when and what Tv programs to watch. However, there is 77.7% for a student to pass in mathematics with a grade of B or bettwe given that their parents do not set rules for them as to what Tv and movies they should watch. I chose bar chart because it is easy to analyse responses using the height of the response variables.

#A bivariet graph showing ones intelligence level feeling and regular discussion with parents   
  
tab1<-table(my\_addhealth$intel, my\_addhealth$chat)  
tab1\_colProp2 <-round( prop.table(tab1, 2),3)   
tab1\_colProp2

##   
## 1 2  
## 1 0.489 0.415  
## 2 0.511 0.585

barplot(tab1\_colProp2[2,], col=c("green"), names=c("Converse with parents","Don't converse with parent"),main="Intelligence level and regular discussion with parents", xlab="Expl var(Conversation with parent)", ylab = "resp var(Intelliigence feeling)")

 From the figure above, there is a 51.1% chance for respondents who regularly have conversation with their parents on school project and assignments to have an intelligent level above their peers. On the filp side, there is a 58.5% probability for students who do not converse with their parent to have an intelligence level above their peers. I also chose this graph because it makes comparision between two variables more interperatable and eaier.  
## 8. Bivariate analysis

## 9. Save and quit

#SAVE DATA AND VARIABLES  
save(my\_vars,my\_addhealth,file = "dataSubset.RDATA")