STUDENT PERFORMANCE PREDICTION

The Dataset Variables

School, Sex, Age, Address, FamilySize, ParentsMaritalStatus, MothersEducation, FathersEducation, MothersJob, FathersJob, ReasonToChooseSchool, Guardian, TravelTime, StudyTime, Failures, SchoolSupport, FamilySupport, PaidClasses, Activities, Nursery, HigherEducationInterest, InternetAcess, RomanticRelationship, FamilyRelationship, FreeTime, GoingOut, WorkdayAlcohol, WeekendAlcohol, Health, Absences, FirstPeriodGrade, SecondPeriodGrade, FinalGrade

PACKAgeS AND LIBRARIES

```
#install.packAges("class")
#install.packAges("gmodels")
#install.packAges("caret")
#install.packAges("readr")
#install.packAges("ggplot2")
#install.packAges("GGally")


library(gmodels)

## Warning: package 'gmodels' was built under R version 3.5.1

library(caret)

## Warning: package 'caret' was built under R version 3.5.1

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.5.1

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.1

library(class)

## Warning: package 'class' was built under R version 3.5.1

library(readr)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(GGally)

## Warning: package 'GGally' was built under R version 3.5.1

##
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':
##
##     nasa

library(scales)

##
## Attaching package: 'scales'

## The following object is masked from 'package:readr':
##
##     col_factor

library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

library(nnet)
library(caret)
```

DATA LOAD

```
getwd()

## [1] "C:/Study/500-PrinciplesAndApplications/Project"

setwd('C:/MyRWork/Data/student-alcohol-consumption')
StudentData <- read.csv('studentData.csv')

StudentDataBackup <- read.csv('studentData.csv')

StudentData[duplicated(StudentData),] # No duplicates
```

```
##  [1] school     sex        age        address    famsize    Pstatus
##  [7] Medu       Fedu       Mjob       Fjob       reason     guardian
## [13] traveltime studytime  failures   schoolsup  famsup     paid
## [19] activities nursery    higher     internet   romantic   famrel
## [25] freetime   goout      Dalc       Walc       health     absences
## [31] G1         G2         G3
## <0 rows> (or 0-length row.names)
```

```r
colnames(StudentData) <- c('School','Sex','Age','Address','FamilySize','Paren
tsMaritalStatus','MothersEducation','FathersEducation','MothersJob','FathersJ
ob','ReasonToChooseSchool','Guardian','TravelTime','StudyTime','Failures','Sc
hoolSupport','FamilySupport','PaidClasses','Activities','Nursery','HigherEduc
ationInterest','InternetAcess','RomanticRelationship','FamilyRelationship','F
reeTime','GoingOut','WorkdayAlcohol','WeekendAlcohol','Health','Absences','Fi
rstPeriodGrade','SecondPeriodGrade','FinalGrade') #,'Subject','AverAgeGrade')

colnames(StudentDataBackup) <- c('School','Sex','Age','Address','FamilySize',
'ParentsMaritalStatus','MothersEducation','FathersEducation','MothersJob','Fa
thersJob','ReasonToChooseSchool','Guardian','TravelTime','StudyTime','Failure
s','SchoolSupport','FamilySupport','PaidClasses','Activities','Nursery','High
erEducationInterest','InternetAcess','RomanticRelationship','FamilyRelationsh
ip','FreeTime','GoingOut','WorkdayAlcohol','WeekendAlcohol','Health','Absence
s','FirstPeriodGrade','SecondPeriodGrade','FinalGrade') #,'Subject','AverAgeG
rade')


#Adding a new column -"GradeClassified" that includes the final grades classi
fied into three categories:
#Above averAge - classification for final grades between 15 and 20
#Below averAge - classification for final grades between 0 and 9
#AverAge - classification for final grades between 10 and 14


StudentData$GradeClassified <- ifelse(StudentData$FinalGrade<10,"BelowAverAge
",
                                      ifelse((StudentData$FinalGrade>=10 & St
udentData$FinalGrade<=15),"AverAge","AboveAverAge"))

#Convert GradeClassified as a factored variable with only 3 levels - Above Av
erAge, averAge and below averAge

StudentData$GradeClassified <- as.factor(StudentData$GradeClassified)
```

DATA EXPLORATION

Here we are slicing, dicing and factorising the data in various level for better analysis and also plotting them with each other and the response variable to understand the data in a better way

```r
# AverAge Grades
StudentData$AverAgeGrades <- round(rowMeans(cbind(StudentData$FirstPeriodGrade, StudentData$SecondPeriodGrade, StudentData$FinalGrade)),2)

# No. of students drinking on Weekday and weekend
StudentData$WorkdayAlcohol<- as.factor(StudentData$WorkdayAlcohol)
StudentData$WorkdayAlcohol <- factor(StudentData$WorkdayAlcohol,labels=c("Very Low", "Low", "Medium", "High", "Very High"))

plot(StudentData$WorkdayAlcohol)
```
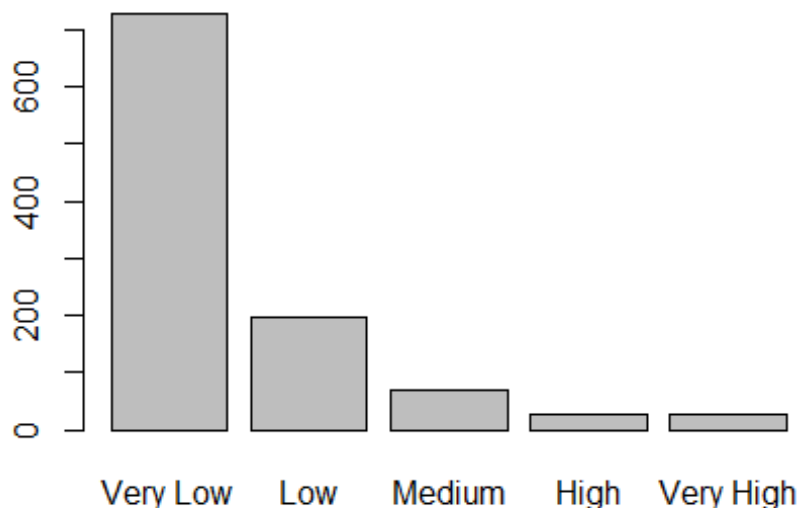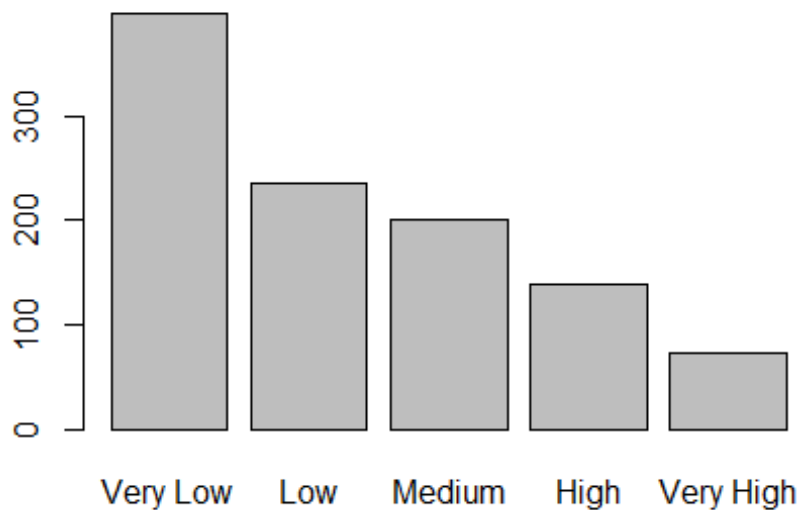


```r
StudentData$WeekendAlcohol <- as.factor(StudentData$WeekendAlcohol)
StudentData$WeekendAlcohol <- factor(StudentData$WeekendAlcohol,labels=c("Very Low", "Low", "Medium", "High", "Very High"))

plot(StudentData$WeekendAlcohol)
```

```
ggcorr(StudentData,label = TRUE,label_alpha = TRUE, hjust = 0.75, size = 2)

## Warning in ggcorr(StudentData, label = TRUE, label_alpha = TRUE, hjust
## = 0.75, : data in column(s) 'School', 'Sex', 'Address', 'FamilySize',
## 'ParentsMaritalStatus', 'MothersJob', 'FathersJob', 'ReasonToChooseSchool'
,
## 'Guardian', 'SchoolSupport', 'FamilySupport', 'PaidClasses',
## 'Activities', 'Nursery', 'HigherEducationInterest', 'InternetAcess',
## 'RomanticRelationship', 'WorkdayAlcohol', 'WeekendAlcohol',
## 'GradeClassified' are not numeric and were ignored
```

The correlation matrix shows relationships between variables including AverAgeGrades, FinalGrade, SecondPeriodGrade, FirstPeriodGrade, Absences, Health, GoingOut, FreeTime, FamilyRelationship, Failures, StudyTime, TravelTime, FathersEducation, othersEducation, and Age. Color scale ranges from -1.0 to 1.0.

Selected correlation values:
- FinalGrade / AverAgeGrades: 1
- SecondPeriodGrade: 0.9, 1
- FirstPeriodGrade: 0.9, 0.8, 0.9
- Failures: -0.4, -0.4, -0.4, -0.4
- othersEducation / AverAgeGrades: 0.6

```r
# Question - Does number of Failures affect your grades?
ggplot(StudentData, aes(x=Failures, y=AverAgeGrades, color=Sex))+
geom_jitter(alpha=0.7) +
theme_bw() +
  xlab("Number of Failures") +
  ylab("AverAge grades")
```
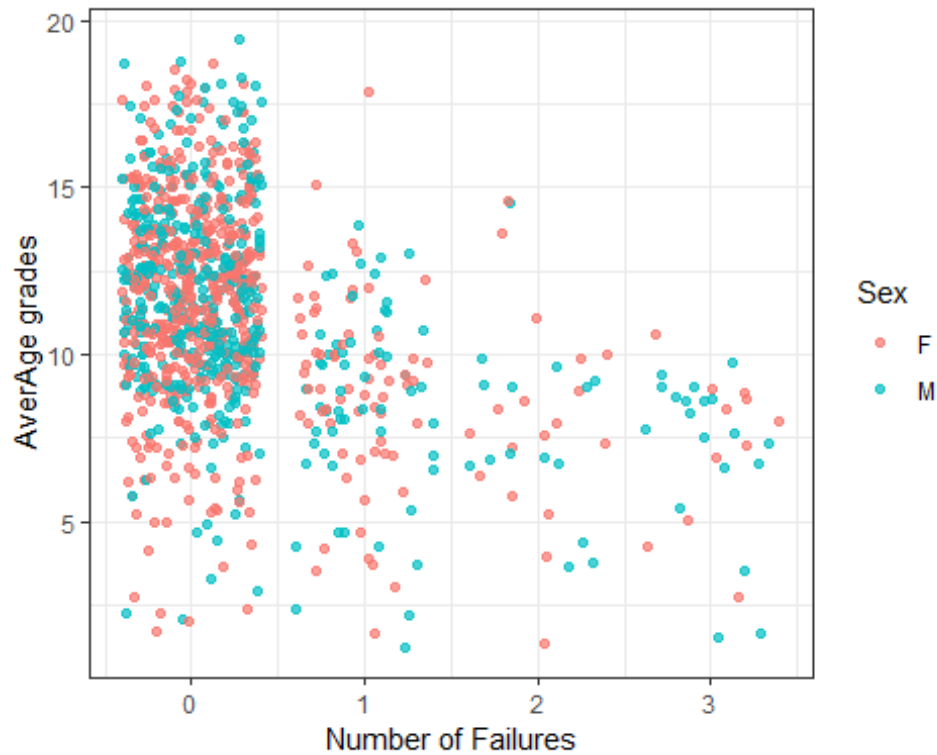
```r
# Observation -  As number of Failures increases, grades decrease


# Question - Does number of Failures affect your grades?
ggplot(StudentData, aes(x=as.factor(HigherEducationInterest),y=AverAgeGrades,
color=Sex))+
  geom_jitter(alpha=0.7) +
  theme_bw() +
  xlab("Number of Failures") +
  ylab("AverAge grades")
```
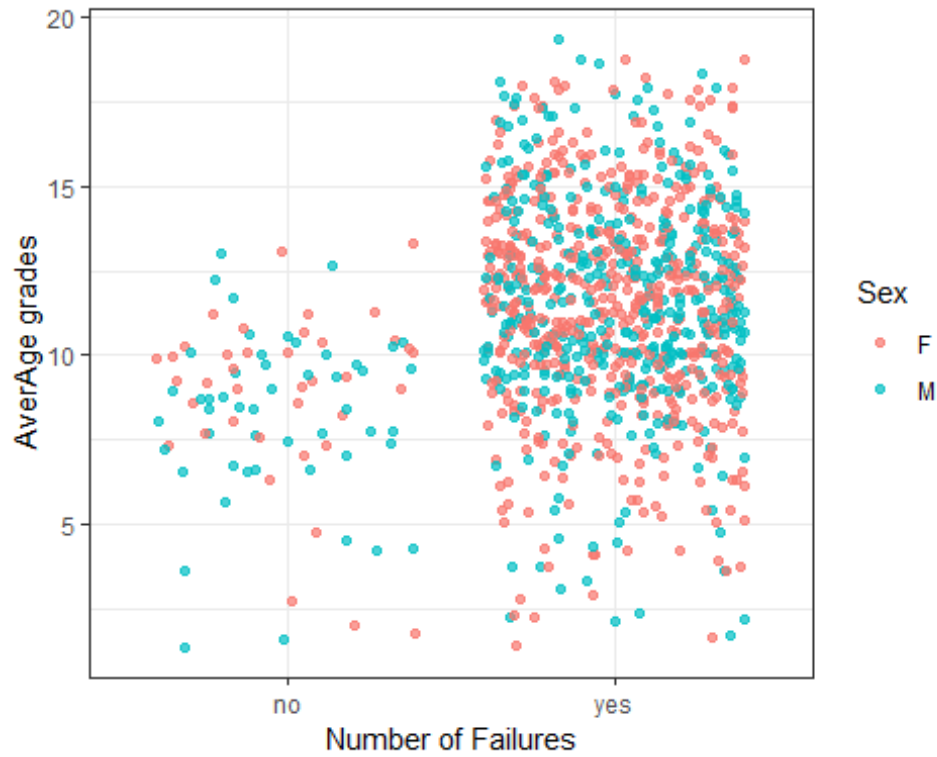
```
# Observation -  As number of Failures increases, grades decrease


# Question - How does mother's education affect the grades?
ggplot(StudentData, aes(x=as.factor(MothersEducation), y=as.numeric(AverAgeGr
ades), fill=MothersEducation))+
  geom_boxplot()
```
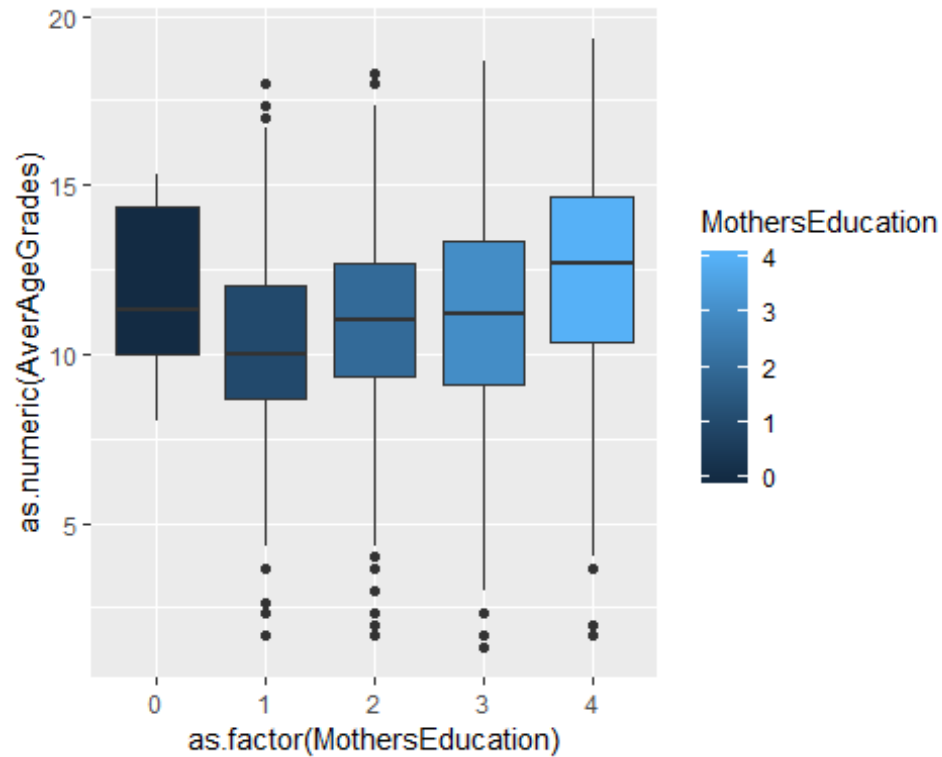
```
# Observation - Grades are better if mother is educated


# Question - How does travel time affect the grades?
ggplot(StudentData, aes(x=as.factor(TravelTime), y=as.numeric(AverAgeGrades),
fill=TravelTime))+
  geom_boxplot()
```
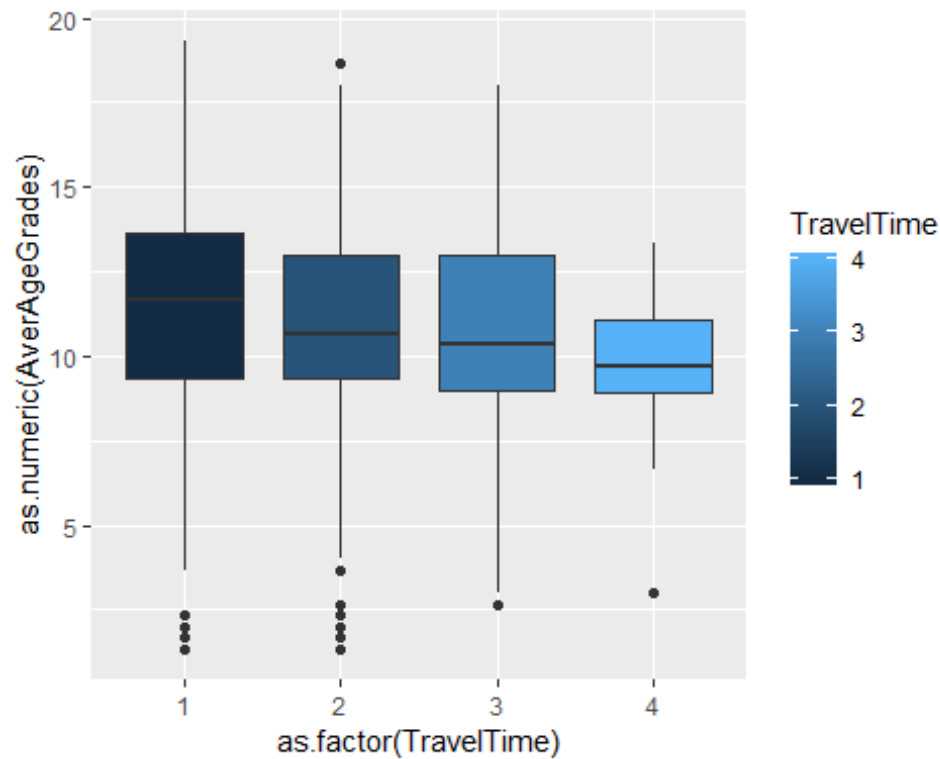
```r
# Observation - Grades gets worse with increase in travel time



# Question - How does going out affect the study time?
ggplot(aes(as.numeric(GoingOut), as.numeric(StudyTime)), data = StudentData)
+
  geom_jitter() +
  geom_smooth(se = FALSE, method = 'lm') +
  xlab("Frequency of Going Out with Friends") +
  ylab("Time Spent Studying")
```
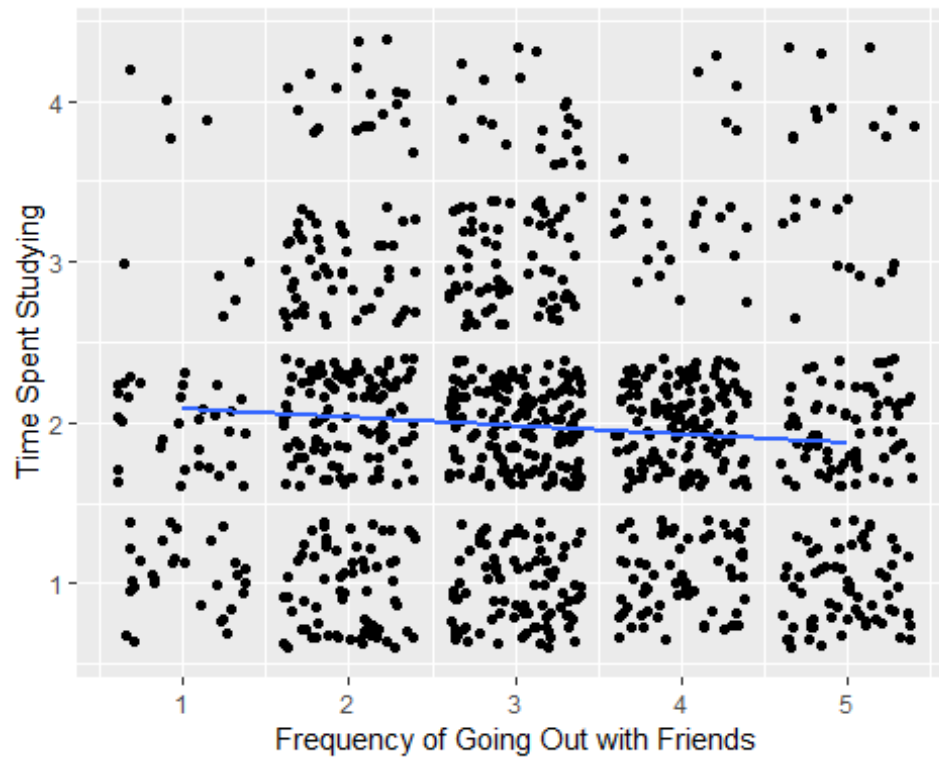
```
# Observation - Going out with friends will lead to lessening of study time



# Some other random observations
ggplot(StudentData, aes(x=as.factor(ParentsMaritalStatus), y=AverAgeGrades))
+
  geom_boxplot(fill="slateblue", alpha=0.2) +
  xlab("parent's cohabitation status ")
```
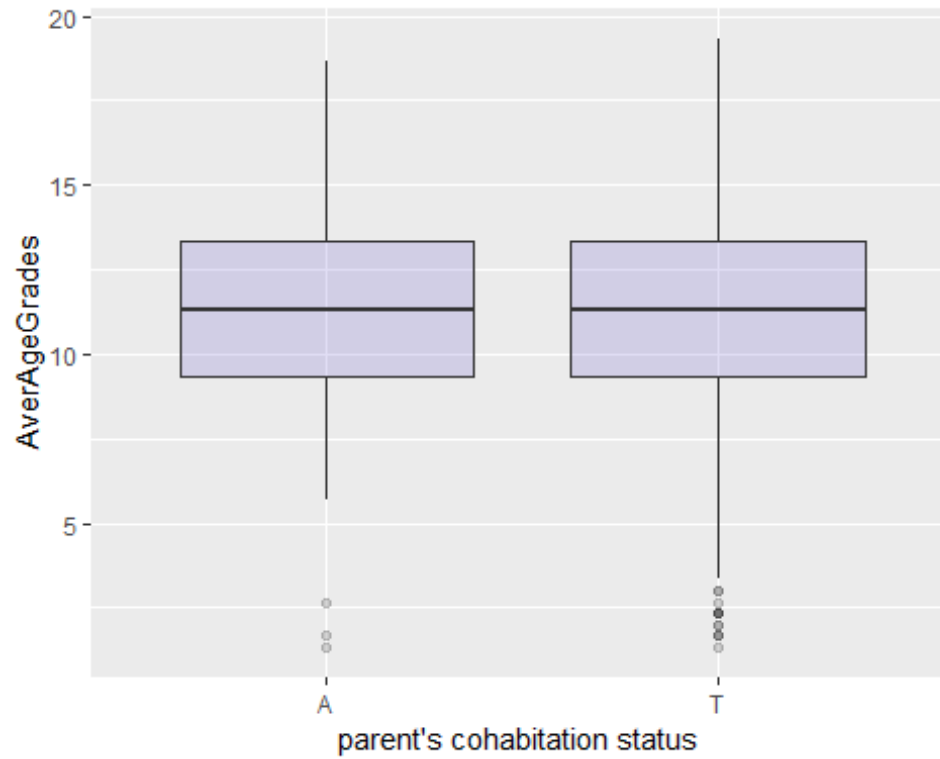
```
ggplot(StudentData, aes(x=FreeTime, y=AverAgeGrades)) +
  geom_point()
```
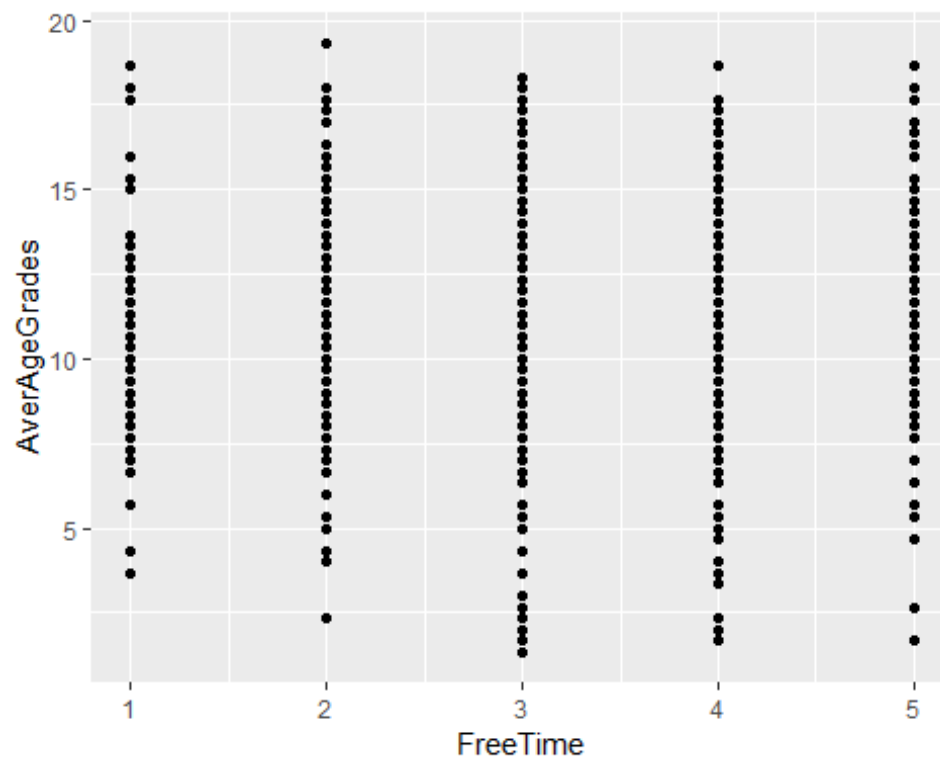
```
ggplot(StudentData, aes(x=WorkdayAlcohol, y=AverAgeGrades)) + geom_point()
```
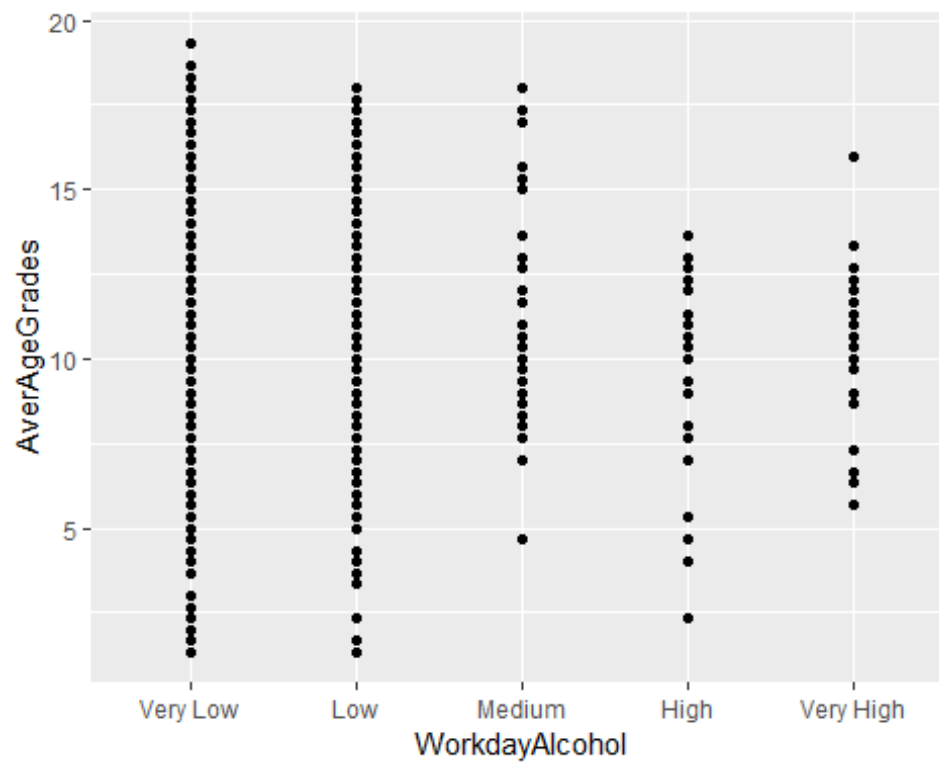


```
ggplot(StudentData, aes(x=as.factor(WorkdayAlcohol), y=AverAgeGrades)) +
  geom_boxplot(fill="slateblue", alpha=0.2)
```
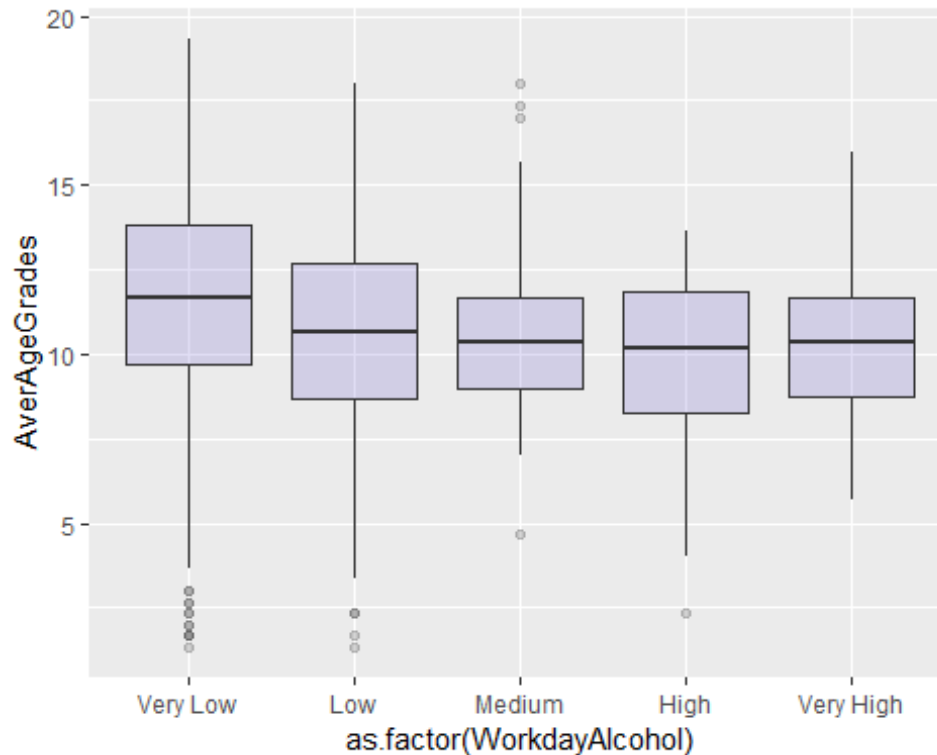
```r
#Question - What is the overall grade distribution of the students?

NumberOfStudents <- as.data.frame(sort(table(StudentData$GradeClassified, dnn
= 'GradeClassified'), decreasing = T),responseName = 'NumberOfStudents')

ggplot(aes(x= reorder(GradeClassified,NumberOfStudents), y = (NumberOfStudent
s/sum(NumberOfStudents)), fill = GradeClassified), data = NumberOfStudents) +
  geom_bar(stat = 'identity') +
  geom_text(stat='identity', aes(label = percent((NumberOfStudents)/sum(Numbe
rOfStudents)))),
          data = NumberOfStudents, hjust = 0.5,vjust=-0.5) +
  theme(axis.text = element_text(size = 12,face = "bold")) +
  xlab("GradeClassified") +
  ylab("Percentage of Students") +
  ggtitle("Percentage of Students in the Grade classified categories")
```
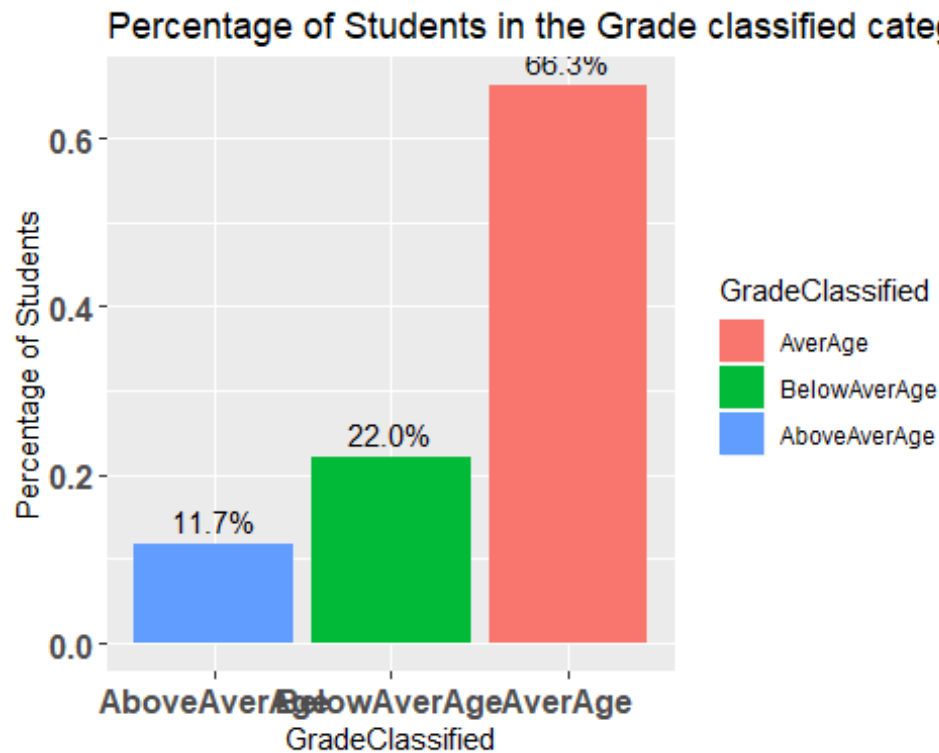
Percentage of Students in the Grade classified catego

```
#Question - How does a romantic relationship affects the grade ?
ggplot(aes(x=GradeClassified, fill = RomanticRelationship),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
 xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with a Student's Romantic Relatio
nship")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with a Student's Roma
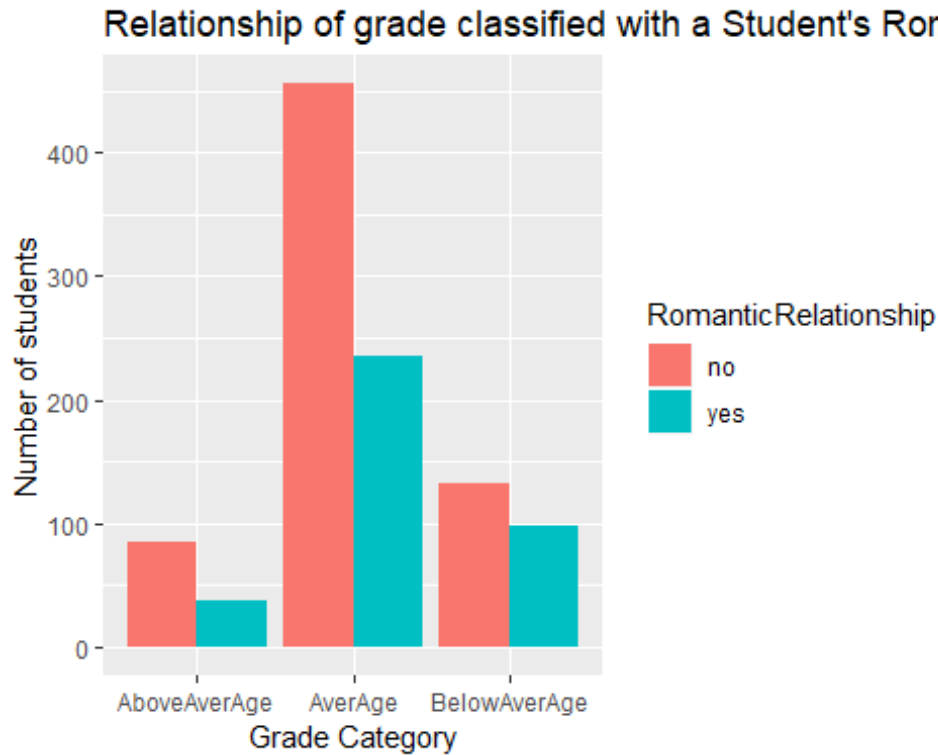

```
# Observation - Students not involved in romantic relationships tends to scor
e than the ones who are in it.


# How does the intake of alcohol during the weekend affect the grade
StudentData$WeekendAlcohol <- as.factor(StudentData$WeekendAlcohol)
ggplot(aes(x=GradeClassified, fill = WeekendAlcohol),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
  xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with a Student's weekend alcohol
consumption ")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with a Student's week
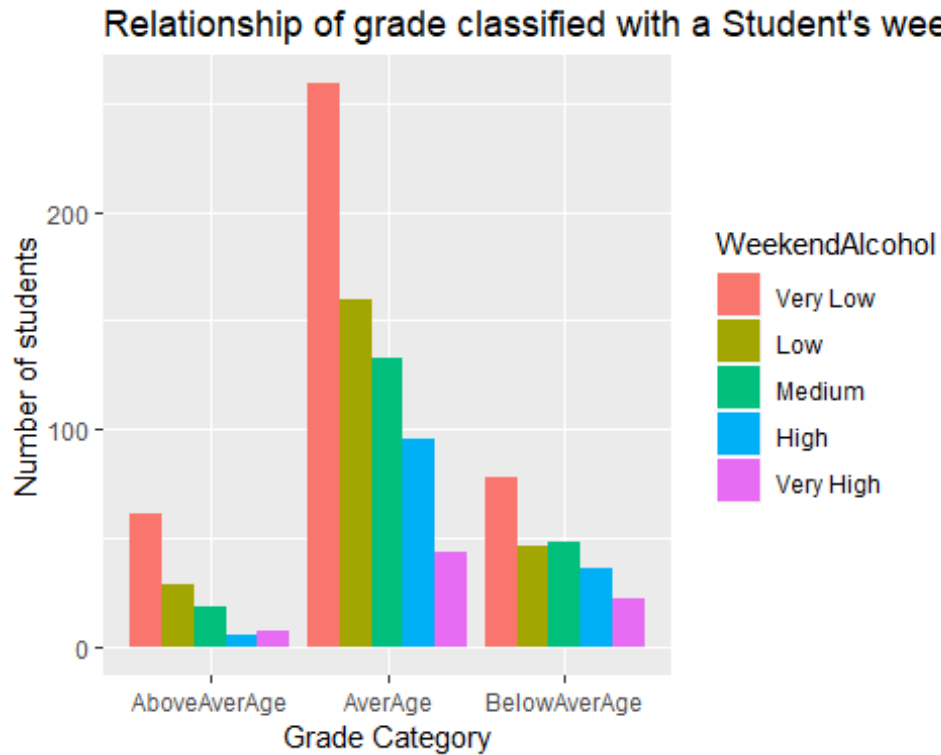

```
# Observation - Students into high level of weekend alcohol intake tends to s
core much less


# Better visualization of the above problem with geom point
 ggplot(aes(x=WeekendAlcohol, y = GradeClassified, color = GradeClassified),
       data = StudentData) +
  geom_point(alpha = 0.5, position="jitter") +
  scale_colour_brewer(palette = 'Set1') +
   theme_bw() + theme(legend.key = element_blank()) +
  xlab("Weekend alcohol") +
  ylab("Grade Category") +
  ggtitle("Relationship of Final grade with weekend alcohol consumption ")
```

Relationship of Final grade with weekend alcohol
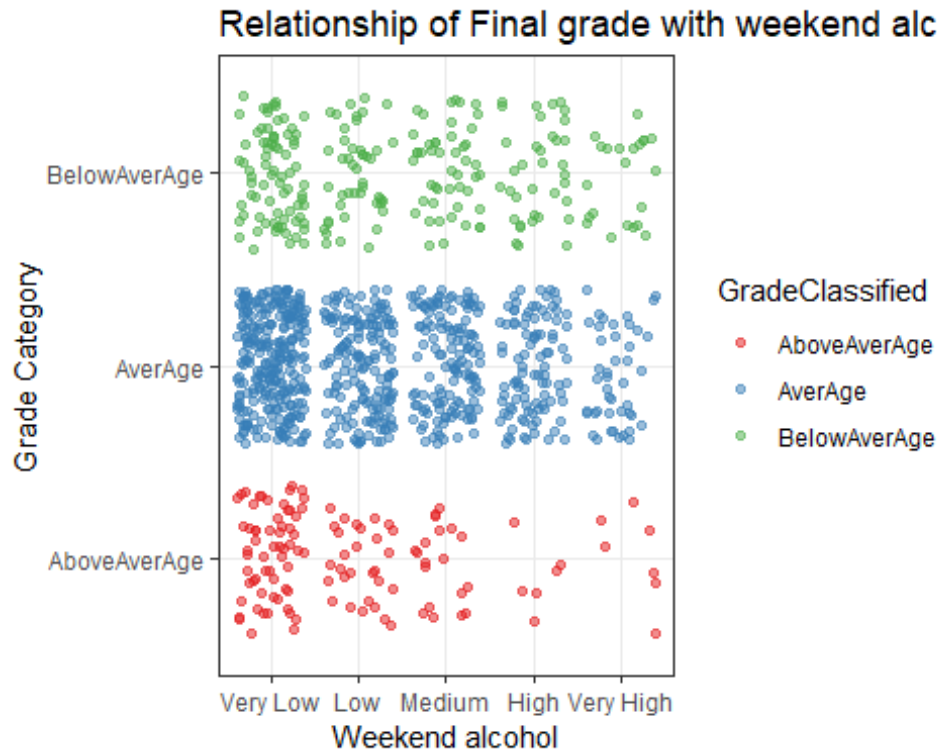
```r
# Checking the dependency of Grade classified with Weekeend Alcohol consumpti
on using Chi-Square Test
chisq.test(StudentData$GradeClassified,StudentData$WeekendAlcohol)

##
##  Pearson's Chi-squared test
##
## data:  StudentData$GradeClassified and StudentData$WeekendAlcohol
## X-squared = 18.151, df = 8, p-value = 0.02012

# P-Value is less than the significant value so the null hypothesis should be
rejected and conclude that grades are affected by weekend alcohol pattern


# Are Grade classified and Mother's Education related?
StudentData$MothersEducation <- as.factor(StudentData$MothersEducation)

ggplot(aes(x=GradeClassified, fill = MothersEducation),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
  xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with a Student's Mother's Educati
on")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with a Student's Moth
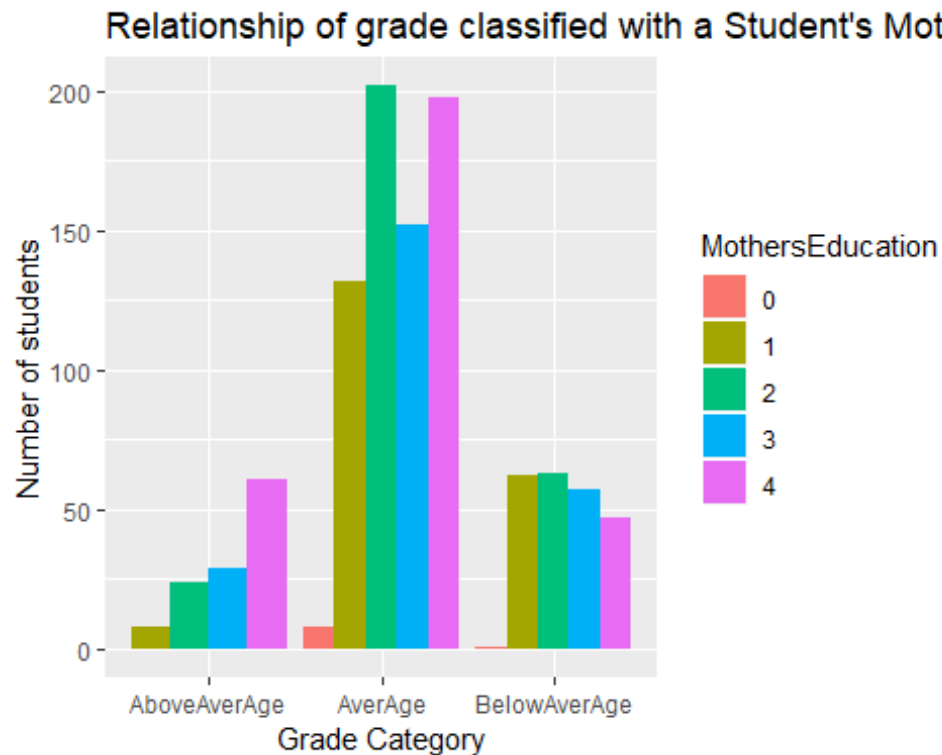

```r
# Not a clear pattern here for all the level of grades, though mothers of stu
dents with above average grade are more educated

# Above problem in different visualization
ggplot(aes(y=MothersEducation, x = GradeClassified, color = GradeClassified),
       data = StudentData) +
  geom_point( alpha = 1, position="jitter") +
  scale_colour_brewer(palette = 'Set2') +
  theme_bw() + theme(legend.key = element_blank()) +
  ylab("Mothers Education") +
  xlab("Grade Category") +
  ggtitle("Relationship of Final grade with a Student's Mother's Education ")
```

Relationship of Final grade with a Student's Mother's Ec
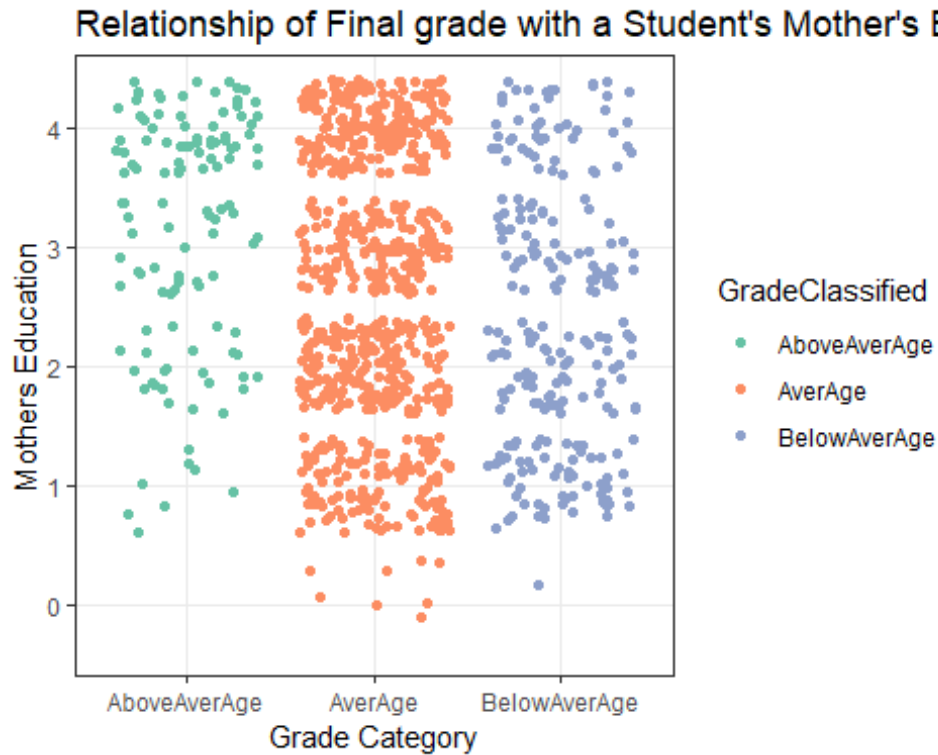

```
# Question - What is the relationship of number of failures and grade score by a student?
StudentData$Failures <- as.factor(StudentData$Failures)
ggplot(aes(x=GradeClassified, fill = Failures),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
  xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with a Student's Failures ")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with a Student's Failure
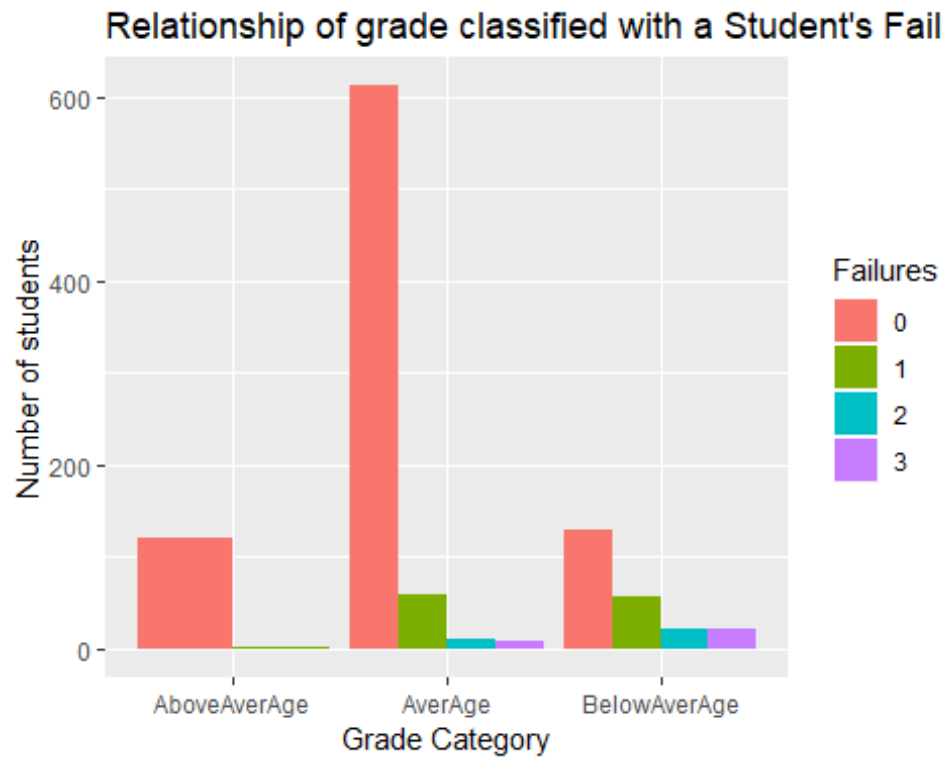

```
# Observation - students with no or less failure tends to score better grades

# Another visualization for above problem
ggplot(aes(y=Failures, x = GradeClassified, color = GradeClassified),
       data = StudentData) +
  geom_point(alpha = 0.5, position="jitter") +
  theme_bw() + theme(legend.key = element_blank()) +
  scale_colour_brewer(palette = "Set1") +
  xlab("Grade categoryl") +
  ylab("Failures") +
  ggtitle("Relationship of Final grade with a Student's Failure ")
```

Relationship of Final grade with a Student's Failure
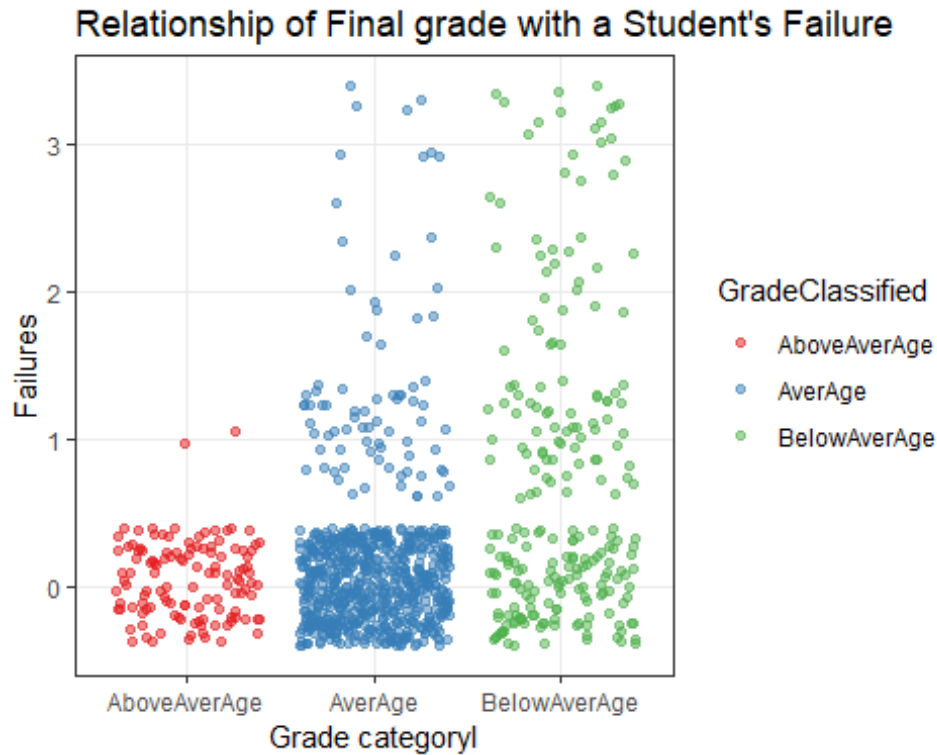

```
# Question - How does going out pattern affect the grades?
StudentData$GoingOut <- as.factor(StudentData$GoingOut)

ggplot(aes(x=GradeClassified, fill = GoingOut),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
 xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with a Student Going Out ")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with a Student Going
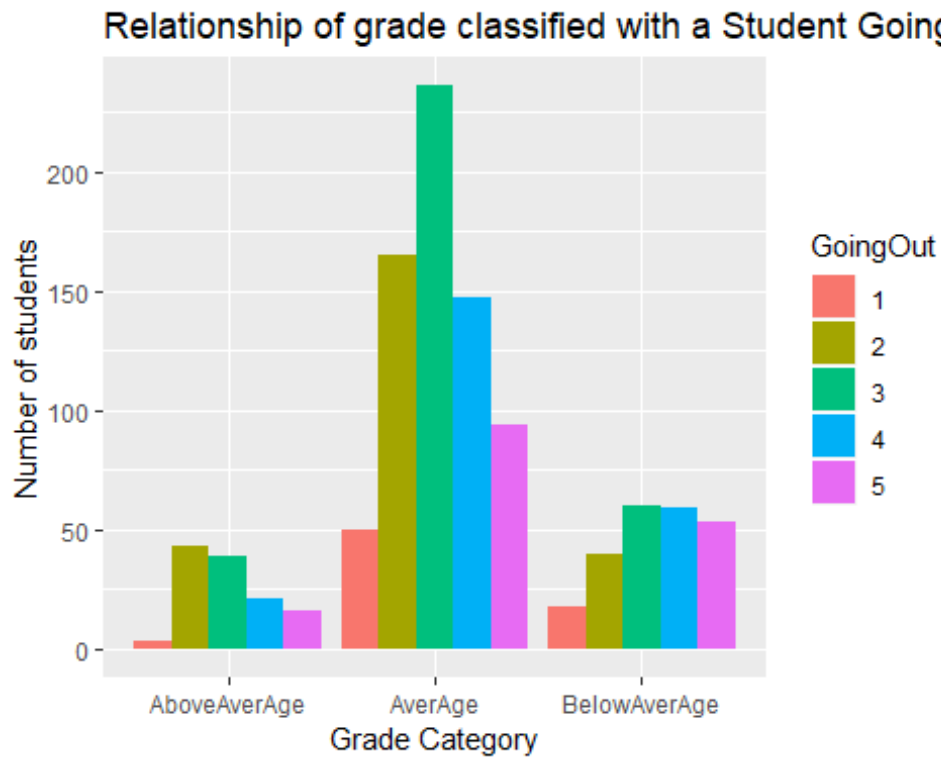

```
# Observation - Not a specific pattern here



#Question - What is the grade of the students who aspire for higher education
?
ggplot(aes(x=GradeClassified, fill =HigherEducationInterest),
       data = StudentData) +
  geom_histogram(stat = "count", position=position_dodge()) +
 xlab("Grade Category") +
  ylab("Number of students") +
  ggtitle("Relationship of grade classified with Higher Education interest ")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Relationship of grade classified with Higher Education
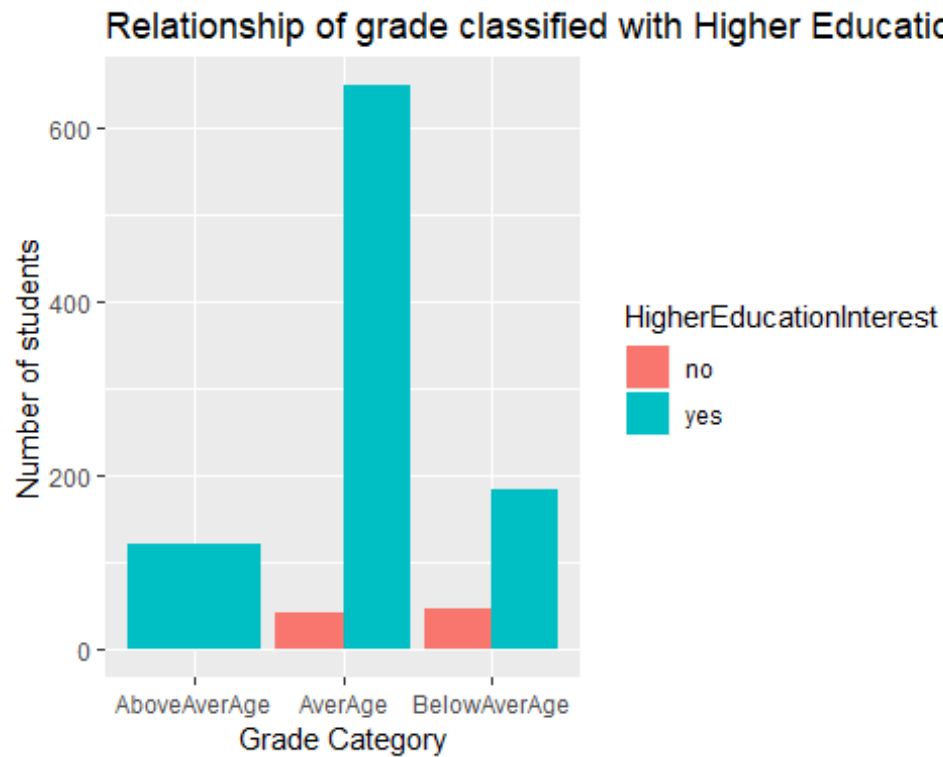

```r
# Observation - Almost all the students with good grades aspire for higher st
udies
```

```r
# Another visualization for above problem
ggplot(aes(x=HigherEducationInterest, y = GradeClassified, color = GradeClass
ified),
       data = StudentData) +
  geom_point(alpha = 0.5, position="jitter") +
  theme_bw() + theme(legend.key = element_blank()) +
  scale_colour_brewer(palette = 'Set2') +
  xlab("HigherEducation Interest") +
  ylab("Grade Category") +
  ggtitle("Relationship of grade classified with Higher Education Interest ")
```

Relationship of grade classified with Higher Ed
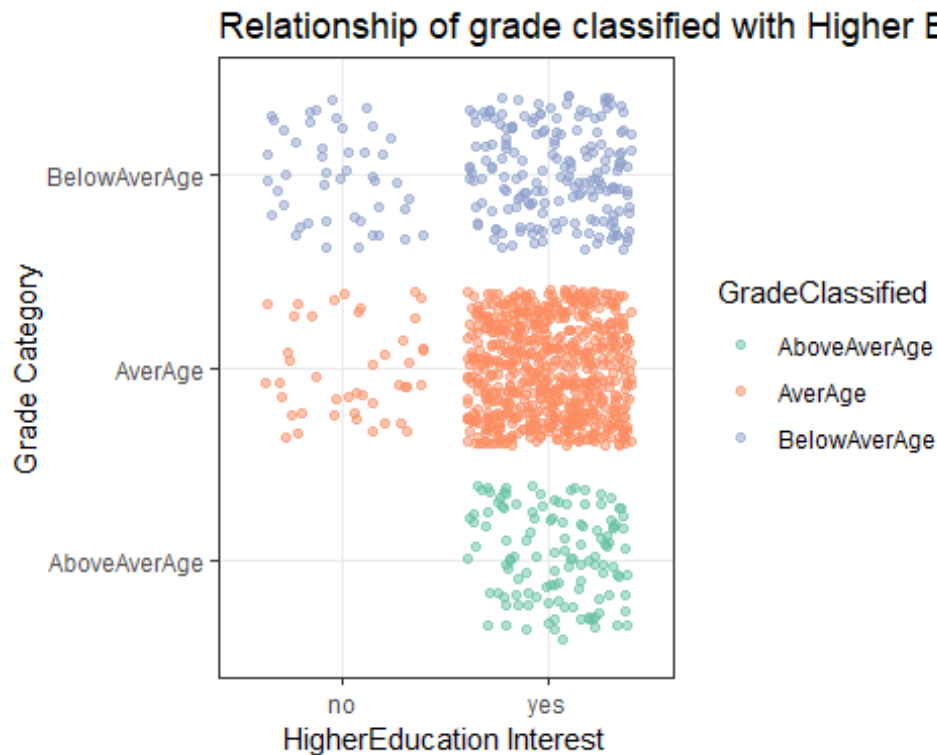
FIRST MODEL - MULTIPLE LINEAR REGRESSION METHOD IMPLEMENTATION

Influential attibutes seems to be - **NUMERICAL** - Age, MothersEducation, FathersEducation, TravelTime, StudyTime, Failures, FamilyRelationship, WeekendAlcohol, WorkdayAlcohol, GoingOut, FreeTime

**CATEGORICAL** - FathersJob, MothersJob, ParentsMaritalStatus, ReasonToChooseSchool, InternetAcess, HigherEducationInterest, Guardian, School, Sex, Address, FamilySize,SchoolSupport, FamilySupport, PaidClasses, Activities, Nursery, RomanticRelationship,

```
StudentData1 <- StudentData[c("Age","MothersEducation", "FathersEducation", "TravelTime", "StudyTime", "Failures",
                              "FamilyRelationship","FreeTime", "WeekendAlcohol", "WorkdayAlcohol", "GoingOut", "Health","Absences")]

unique(sort(StudentData$RomanticRelationship))

## [1] no  yes
## Levels: no yes
```

```r
StudentData1$School.gp <- ifelse(StudentData$School == "GP",1,0) # GP = 1, MS
= 0
StudentData1$HigherEducationInterest.yes <- ifelse(StudentData$HigherEducatio
nInterest == "yes",1,0) # yes = 1, no = 0
StudentData1$ParentsMaritalStatus.T <- ifelse(StudentData$ParentsMaritalStatu
s == "T",1,0) # Together = 1, Apart = 0
StudentData1$InternetAcess.yes <- ifelse(StudentData$InternetAcess == "yes",1
,0) # yes = 1, no = 0
StudentData1$Sex.F <- ifelse(StudentData$Sex == "F",1,0)  # F = 1, M = 0
StudentData1$Address.R <- ifelse(StudentData$Address == "R",1,0)  # R = 1, U
= 0
StudentData1$FamilySize.GT3 <- ifelse(StudentData$FamilySize == "GT3",1,0)  #
GT3 = 1, LE3 = 0
StudentData1$SchoolSupport.yes <- ifelse(StudentData$SchoolSupport == "yes",1
,0) # yes = 1, no = 0
StudentData1$FamilySupport.yes <- ifelse(StudentData$FamilySupport == "yes",1
,0) # yes = 1, no = 0
StudentData1$PaidClasses.yes <- ifelse(StudentData$PaidClasses == "yes",1,0)
# yes = 1, no = 0
StudentData1$Activities.yes <- ifelse(StudentData$Activities == "yes",1,0) #
yes = 1, no = 0
StudentData1$Nursery.yes <- ifelse(StudentData$Nursery == "yes",1,0) # yes =
1, no = 0
StudentData1$RomanticRelationship.yes <- ifelse(StudentData$RomanticRelations
hip == "yes",1,0) # yes = 1, no = 0


StudentData1$Guardian.M <- ifelse(StudentData$Guardian == "mother",1,0) # Mot
her = 1, Other = 0
StudentData1$Guardian.F <- ifelse(StudentData$Guardian == "father",1,0) # Fat
her = 1, Other = 0


StudentData1$ReasonToChooseSchool.course <- ifelse(StudentData$ReasonToChoose
School == "course",1,0)
StudentData1$ReasonToChooseSchool.home <- ifelse(StudentData$ReasonToChooseSc
hool == "home",1,0)
StudentData1$ReasonToChooseSchool.reputation <- ifelse(StudentData$ReasonToCh
ooseSchool == "reputation",1,0)

StudentData1$MothersJob.athome <- ifelse(StudentData$MothersJob == "at_home",
1,0)
StudentData1$MothersJob.Health <- ifelse(StudentData$MothersJob == "Health",1
,0)
StudentData1$MothersJob.services<-ifelse(StudentData$MothersJob == "services"
,1,0)
StudentData1$MothersJob.teacher<- ifelse(StudentData$MothersJob == "teacher",
1,0)
```

```
StudentData1$FathersJob.athome <- ifelse(StudentData$FathersJob == "at_home",
1,0)
StudentData1$FathersJob.Health <- ifelse(StudentData$FathersJob == "Health",1
,0)
StudentData1$FathersJob.services<-ifelse(StudentData$FathersJob == "services"
,1,0)
StudentData1$FathersJob.teacher<- ifelse(StudentData$FathersJob == "teacher",
1,0)


StudentData1$AverAgeGrades <- StudentData$AverAgeGrades



StudentData2 <- StudentData1



##### MODEL 1 ############################
model1 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model1)

##
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -10.4284  -1.5092   0.0983   1.6946   7.3636
##
## Coefficients: (2 not defined because of singularities)
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               9.594454   2.012584   4.767 2.15e-06 ***
## Age                       0.064080   0.083416   0.768 0.442553
## MothersEducation1        -0.950579   0.963349  -0.987 0.324008
## MothersEducation2        -0.919146   0.968225  -0.949 0.342694
## MothersEducation3        -0.775206   0.979450  -0.791 0.428858
## MothersEducation4         0.042035   1.002183   0.042 0.966552
## FathersEducation          0.044725   0.109621   0.408 0.683361
## TravelTime               -0.081093   0.132906  -0.610 0.541897
## StudyTime                 0.364069   0.114593   3.177 0.001534 **
## Failures1                -2.209801   0.295241  -7.485 1.58e-13 ***
## Failures2                -3.401519   0.518641  -6.559 8.73e-11 ***
## Failures3                -3.518427   0.547949  -6.421 2.09e-10 ***
## FamilyRelationship        0.064715   0.096708   0.669 0.503540
## FreeTime                  0.061429   0.093592   0.656 0.511755
## WeekendAlcoholLow        -0.065353   0.244718  -0.267 0.789484
## WeekendAlcoholMedium      0.127219   0.276384   0.460 0.645404
## WeekendAlcoholHigh       -0.199222   0.351275  -0.567 0.570748
## WeekendAlcoholVery High   0.417511   0.520306   0.802 0.422494
## WorkdayAlcoholLow        -0.379839   0.264477  -1.436 0.151264
```

```
## WorkdayAlcoholMedium                  -0.056033    0.415882   -0.135 0.892851
## WorkdayAlcoholHigh                     -1.262847    0.610142   -2.070 0.038733 *
## WorkdayAlcoholVery High                -0.496150    0.706812   -0.702 0.482872
## GoingOut2                               1.119004    0.383504    2.918 0.003604 **
## GoingOut3                               0.506669    0.373408    1.357 0.175129
## GoingOut4                               0.172580    0.396716    0.435 0.663640
## GoingOut5                              -0.067471    0.432127   -0.156 0.875957
## Health                                 -0.140006    0.063279   -2.213 0.027158 *
## Absences                               -0.008728    0.015007   -0.582 0.560993
## School.gp                               0.418064    0.237412    1.761 0.078560 .
## HigherEducationInterest.yes             1.370638    0.342387    4.003 6.72e-05 ***
## ParentsMaritalStatus.T                  0.035015    0.288431    0.121 0.903401
## InternetAcess.yes                       0.379093    0.231761    1.636 0.102218
## Sex.F                                   0.099588    0.203052    0.490 0.623919
## Address.R                              -0.267201    0.220718   -1.211 0.226338
## FamilySize.GT3                         -0.371558    0.199063   -1.867 0.062261 .
## SchoolSupport.yes                      -1.375148    0.286198   -4.805 1.79e-06 ***
## FamilySupport.yes                      -0.321349    0.188068   -1.709 0.087822 .
## PaidClasses.yes                        -0.767621    0.220756   -3.477 0.000529 ***
## Activities.yes                          0.071334    0.181357    0.393 0.694155
## Nursery.yes                             0.105301    0.223094    0.472 0.637030
## RomanticRelationship.yes               -0.422684    0.189123   -2.235 0.025641 *
## Guardian.M                             -0.599121    0.389825   -1.537 0.124637
## Guardian.F                             -0.319715    0.422621   -0.757 0.449526
## ReasonToChooseSchool.course            -0.021954    0.308089   -0.071 0.943207
## ReasonToChooseSchool.home               0.035259    0.332072    0.106 0.915463
## ReasonToChooseSchool.reputation         0.260204    0.343227    0.758 0.448564
## MothersJob.athome                      -0.023913    0.261460   -0.091 0.927147
## MothersJob.Health                             NA          NA       NA       NA
## MothersJob.services                     0.491218    0.235406    2.087 0.037172 *
## MothersJob.teacher                     -0.573079    0.336736   -1.702 0.089095 .
## FathersJob.athome                      -0.011837    0.386633   -0.031 0.975583
## FathersJob.Health                             NA          NA       NA       NA
## FathersJob.services                    -0.150780    0.209022   -0.721 0.470857
## FathersJob.teacher                      1.095987    0.408052    2.686 0.007354 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.739 on 992 degrees of freedom
## Multiple R-squared:  0.3112, Adjusted R-squared:  0.2758
## F-statistic: 8.788 on 51 and 992 DF,  p-value: < 2.2e-16

StudentData2$FathersJob.Health <- NULL
StudentData2$FathersJob.teacher <- NULL
StudentData2$MothersJob.athome <- NULL
StudentData2$Nursery.yes <- NULL


##### MODEL 2 ###########################
model2 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model2)
```

```
## 
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5846  -1.5181   0.0884   1.7402   7.3389
## 
## Coefficients: (1 not defined because of singularities)
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  9.539754   2.012894   4.739 2.46e-06 ***
## Age                          0.070951   0.083545   0.849  0.39594
## MothersEducation1           -0.898952   0.964065  -0.932  0.35133
## MothersEducation2           -0.928771   0.966429  -0.961  0.33677
## MothersEducation3           -0.818759   0.976635  -0.838  0.40204
## MothersEducation4            0.109087   0.995279   0.110  0.91275
## FathersEducation             0.117350   0.105594   1.111  0.26669
## TravelTime                  -0.067907   0.132799  -0.511  0.60922
## StudyTime                    0.353103   0.114658   3.080  0.00213 **
## Failures1                   -2.217325   0.295775  -7.497 1.45e-13 ***
## Failures2                   -3.397493   0.519603  -6.539 9.91e-11 ***
## Failures3                   -3.517597   0.549046  -6.407 2.29e-10 ***
## FamilyRelationship           0.050849   0.096710   0.526  0.59915
## FreeTime                     0.062098   0.093783   0.662  0.50803
## WeekendAlcoholLow           -0.071619   0.245088  -0.292  0.77018
## WeekendAlcoholMedium         0.067118   0.276017   0.243  0.80793
## WeekendAlcoholHigh          -0.325060   0.348641  -0.932  0.35137
## WeekendAlcoholVery High      0.311625   0.518205   0.601  0.54774
## WorkdayAlcoholLow           -0.335925   0.264541  -1.270  0.20444
## WorkdayAlcoholMedium        -0.053777   0.416543  -0.129  0.89730
## WorkdayAlcoholHigh          -1.190248   0.610731  -1.949  0.05159 .
## WorkdayAlcoholVery High     -0.447721   0.706506  -0.634  0.52641
## GoingOut2                    1.059268   0.382057   2.773  0.00567 **
## GoingOut3                    0.477413   0.374023   1.276  0.20210
## GoingOut4                    0.156225   0.397275   0.393  0.69423
## GoingOut5                   -0.097805   0.431239  -0.227  0.82063
## Health                      -0.140727   0.063238  -2.225  0.02628 *
## Absences                    -0.008664   0.015040  -0.576  0.56468
## School.gp                    0.448258   0.237641   1.886  0.05955 .
## HigherEducationInterest.yes  1.364373   0.341787   3.992 7.04e-05 ***
## ParentsMaritalStatus.T       0.003617   0.288490   0.013  0.99000
## InternetAcess.yes            0.336572   0.230410   1.461  0.14440
## Sex.F                        0.075573   0.202431   0.373  0.70898
## Address.R                   -0.260923   0.220927  -1.181  0.23787
## FamilySize.GT3              -0.331456   0.197805  -1.676  0.09412 .
## SchoolSupport.yes           -1.323503   0.286147  -4.625 4.23e-06 ***
## FamilySupport.yes           -0.325634   0.188227  -1.730  0.08394 .
## PaidClasses.yes             -0.805967   0.220470  -3.656  0.00027 ***
## Activities.yes               0.060482   0.181655   0.333  0.73924
## RomanticRelationship.yes    -0.393922   0.189230  -2.082  0.03762 *
```

```
## Guardian.M                          -0.522295    0.387142   -1.349   0.17761
## Guardian.F                          -0.209723    0.420053   -0.499   0.61769
## ReasonToChooseSchool.course         -0.062456    0.308365   -0.203   0.83954
## ReasonToChooseSchool.home            0.001209    0.331887    0.004   0.99709
## ReasonToChooseSchool.reputation      0.215092    0.342469    0.628   0.53011
## MothersJob.Health                          NA          NA       NA         NA
## MothersJob.services                  0.560876    0.228140    2.458   0.01412 *
## MothersJob.teacher                  -0.447866    0.332805   -1.346   0.17870
## FathersJob.athome                   -0.093399    0.383071   -0.244   0.80742
## FathersJob.services                 -0.258196    0.205424   -1.257   0.20909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.745 on 995 degrees of freedom
## Multiple R-squared:  0.3061, Adjusted R-squared:  0.2726
## F-statistic: 9.145 on 48 and 995 DF,  p-value: < 2.2e-16

StudentData2$FathersJob.athome <- NULL
StudentData2$ReasonToChooseSchool.home <- NULL
StudentData2$Sex.F <- NULL
StudentData2$ParentsMaritalStatus.T <- NULL


##### MODEL 3 ############################
model3 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model3)

##
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -10.597   -1.530    0.095    1.725    7.372
##
## Coefficients: (1 not defined because of singularities)
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               9.582760   1.988900    4.818 1.67e-06 ***
## Age                       0.071134   0.083008    0.857 0.391680
## MothersEducation1        -0.905327   0.954247   -0.949 0.342985
## MothersEducation2        -0.933983   0.954303   -0.979 0.327962
## MothersEducation3        -0.827452   0.961968   -0.860 0.389904
## MothersEducation4         0.094250   0.982621    0.096 0.923606
## FathersEducation          0.118600   0.105226    1.127 0.259973
## TravelTime               -0.067588   0.131593   -0.514 0.607635
## StudyTime                 0.359858   0.112763    3.191 0.001461 **
## Failures1                -2.222187   0.294764   -7.539 1.06e-13 ***
## Failures2                -3.398121   0.518140   -6.558 8.72e-11 ***
## Failures3                -3.522775   0.547499   -6.434 1.92e-10 ***
## FamilyRelationship        0.049403   0.096069    0.514 0.607196
## FreeTime                  0.056164   0.092579    0.607 0.544215
```

```
## WeekendAlcoholLow                       -0.072036    0.244416   -0.295 0.768265
## WeekendAlcoholMedium                      0.068463    0.273355    0.250 0.802288
## WeekendAlcoholHigh                       -0.339256    0.344018   -0.986 0.324295
## WeekendAlcoholVery High                   0.289496    0.511694    0.566 0.571684
## WorkdayAlcoholLow                        -0.344160    0.262581   -1.311 0.190267
## WorkdayAlcoholMedium                     -0.072005    0.413389   -0.174 0.861758
## WorkdayAlcoholHigh                       -1.195525    0.607099   -1.969 0.049201 *
## WorkdayAlcoholVery High                  -0.462179    0.702878   -0.658 0.510977
## GoingOut2                                 1.056893    0.380902    2.775 0.005628 **
## GoingOut3                                 0.474989    0.373125    1.273 0.203313
## GoingOut4                                 0.168468    0.395301    0.426 0.670071
## GoingOut5                                -0.087681    0.429565   -0.204 0.838304
## Health                                   -0.141800    0.062811   -2.258 0.024187 *
## Absences                                 -0.008422    0.014892   -0.566 0.571817
## School.gp                                 0.446472    0.232407    1.921 0.055007 .
## HigherEducationInterest.yes               1.374474    0.339375    4.050 5.52e-05 ***
## InternetAcess.yes                         0.336942    0.228835    1.472 0.141222
## Address.R                                -0.258103    0.219565   -1.176 0.240067
## FamilySize.GT3                           -0.324639    0.191645   -1.694 0.090584 .
## SchoolSupport.yes                        -1.313743    0.283622   -4.632 4.10e-06 ***
## FamilySupport.yes                        -0.320330    0.185619   -1.726 0.084704 .
## PaidClasses.yes                          -0.806091    0.219794   -3.667 0.000258 ***
## Activities.yes                            0.050750    0.178937    0.284 0.776759
## RomanticRelationship.yes                 -0.386793    0.187277   -2.065 0.039147 *
## Guardian.M                               -0.520892    0.384131   -1.356 0.175397
## Guardian.F                               -0.213766    0.414856   -0.515 0.606474
## ReasonToChooseSchool.course              -0.061102    0.203508   -0.300 0.764054
## ReasonToChooseSchool.reputation           0.219593    0.240241    0.914 0.360910
## MothersJob.Health                               NA          NA       NA       NA
## MothersJob.services                       0.555930    0.227020    2.449 0.014504 *
## MothersJob.teacher                       -0.451425    0.331225   -1.363 0.173223
## FathersJob.services                      -0.248531    0.201309   -1.235 0.217279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.74 on 999 degrees of freedom
## Multiple R-squared:  0.306,  Adjusted R-squared:  0.2754
## F-statistic: 10.01 on 44 and 999 DF,  p-value: < 2.2e-16

StudentData2$Age <- NULL
StudentData2$FreeTime <- NULL
StudentData2$WeekendAlcohol <- NULL
StudentData2$Activities.yes <- NULL
StudentData2$Guardian.F <- NULL


##### MODEL 4 #############################
model4 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model4)
```

```
## 
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -10.560  -1.529   0.115   1.725   7.434 
## 
## Coefficients: (1 not defined because of singularities)
##                                 Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                     10.827001   1.148133   9.430  < 2e-16 ***
## MothersEducation1               -0.971265   0.949842  -1.023 0.306764    
## MothersEducation2               -0.991123   0.949031  -1.044 0.296573    
## MothersEducation3               -0.908665   0.957325  -0.949 0.342761    
## MothersEducation4                0.014477   0.976065   0.015 0.988169    
## FathersEducation                 0.110517   0.104615   1.056 0.291028    
## TravelTime                      -0.072510   0.130583  -0.555 0.578828    
## StudyTime                        0.384417   0.109992   3.495 0.000495 ***
## Failures1                       -2.160167   0.287731  -7.508 1.32e-13 ***
## Failures2                       -3.269448   0.503001  -6.500 1.26e-10 ***
## Failures3                       -3.338857   0.535584  -6.234 6.68e-10 ***
## FamilyRelationship               0.060928   0.094333   0.646 0.518502    
## WorkdayAlcoholLow               -0.395776   0.233669  -1.694 0.090622 .  
## WorkdayAlcoholMedium            -0.082051   0.370481  -0.221 0.824769    
## WorkdayAlcoholHigh              -1.091332   0.567045  -1.925 0.054562 .  
## WorkdayAlcoholVery High         -0.106582   0.572146  -0.186 0.852259    
## GoingOut2                        1.038550   0.378128   2.747 0.006130 ** 
## GoingOut3                        0.492039   0.367492   1.339 0.180902    
## GoingOut4                        0.207161   0.383986   0.540 0.589660    
## GoingOut5                       -0.037904   0.405319  -0.094 0.925511    
## Health                          -0.145802   0.062100  -2.348 0.019075 *  
## Absences                        -0.007578   0.014624  -0.518 0.604446    
## School.gp                        0.446338   0.227518   1.962 0.050065 .  
## HigherEducationInterest.yes      1.304931   0.334086   3.906 0.000100 ***
## InternetAcess.yes                0.323140   0.227456   1.421 0.155723    
## Address.R                       -0.234358   0.218306  -1.074 0.283291    
## FamilySize.GT3                  -0.330355   0.190918  -1.730 0.083874 .  
## SchoolSupport.yes               -1.355485   0.275381  -4.922 9.99e-07 ***
## FamilySupport.yes               -0.319956   0.183052  -1.748 0.080786 .  
## PaidClasses.yes                 -0.809152   0.218541  -3.703 0.000225 ***
## RomanticRelationship.yes        -0.344186   0.184765  -1.863 0.062777 .  
## Guardian.M                      -0.369720   0.196021  -1.886 0.059566 .  
## ReasonToChooseSchool.course     -0.060272   0.202524  -0.298 0.766067    
## ReasonToChooseSchool.reputation  0.240212   0.236787   1.014 0.310604    
## MothersJob.Health                      NA         NA      NA       NA    
## MothersJob.services              0.567035   0.226039   2.509 0.012278 *  
## MothersJob.teacher              -0.398294   0.326421  -1.220 0.222680    
## FathersJob.services             -0.268345   0.198682  -1.351 0.177117    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 2.735 on 1007 degrees of freedom
## Multiple R-squared:  0.303,  Adjusted R-squared:  0.278
## F-statistic: 12.16 on 36 and 1007 DF,  p-value: < 2.2e-16

StudentData2$TravelTime <- NULL
StudentData2$ReasonToChooseSchool.course <- NULL
StudentData2$MothersJob.teacher <- NULL



##### MODEL 5 ############################
model5 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model5)

## 
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
## 
## Residuals:
##      Min      1Q  Median      3Q     Max
## -10.7757 -1.5277  0.1078  1.7232  7.5543
## 
## Coefficients: (1 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.705014   1.114849   9.602  < 2e-16 ***
## MothersEducation1           -1.003440   0.948754  -1.058 0.290473
## MothersEducation2           -1.029097   0.947945  -1.086 0.277912
## MothersEducation3           -0.944374   0.956305  -0.988 0.323622
## MothersEducation4           -0.156770   0.965768  -0.162 0.871081
## FathersEducation             0.109438   0.104314   1.049 0.294370
## StudyTime                    0.381902   0.109848   3.477 0.000529 ***
## Failures1                   -2.157134   0.287425  -7.505 1.35e-13 ***
## Failures2                   -3.278888   0.500826  -6.547 9.33e-11 ***
## Failures3                   -3.347453   0.535033  -6.257 5.81e-10 ***
## FamilyRelationship           0.060143   0.094279   0.638 0.523669
## WorkdayAlcoholLow           -0.413650   0.232760  -1.777 0.075845 .
## WorkdayAlcoholMedium        -0.115239   0.366956  -0.314 0.753556
## WorkdayAlcoholHigh          -1.120145   0.564144  -1.986 0.047352 *
## WorkdayAlcoholVery High     -0.113575   0.570235  -0.199 0.842168
## GoingOut2                    1.055735   0.376729   2.802 0.005170 **
## GoingOut3                    0.517525   0.366393   1.412 0.158115
## GoingOut4                    0.240691   0.382709   0.629 0.529547
## GoingOut5                   -0.013608   0.404345  -0.034 0.973160
## Health                      -0.148161   0.061822  -2.397 0.016731 *
## Absences                    -0.007451   0.014565  -0.512 0.609071
## School.gp                    0.450994   0.226347   1.992 0.046587 *
## HigherEducationInterest.yes  1.307359   0.333667   3.918 9.52e-05 ***
## InternetAcess.yes            0.310807   0.225974   1.375 0.169310
## Address.R                   -0.275251   0.209982  -1.311 0.190213
## FamilySize.GT3              -0.311986   0.190376  -1.639 0.101569
```

```
## SchoolSupport.yes                  -1.330220    0.274495  -4.846 1.46e-06 ***
## FamilySupport.yes                  -0.325025    0.182764  -1.778 0.075641 .
## PaidClasses.yes                    -0.800844    0.217349  -3.685 0.000241 ***
## RomanticRelationship.yes           -0.322932    0.184036  -1.755 0.079610 .
## Guardian.M                         -0.394563    0.193591  -2.038 0.041797 *
## ReasonToChooseSchool.reputation    0.297320    0.210732   1.411 0.158584
## MothersJob.Health                        NA          NA      NA       NA
## MothersJob.services                 0.634326    0.219000   2.896 0.003855 **
## FathersJob.services                -0.284517    0.197570  -1.440 0.150155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.734 on 1010 degrees of freedom
## Multiple R-squared:  0.3016, Adjusted R-squared:  0.2787
## F-statistic: 13.22 on 33 and 1010 DF,  p-value: < 2.2e-16

StudentData2$WorkdayAlcohol <- NULL
StudentData2$ReasonToChooseSchool.reputation <- NULL
StudentData2$Absences <- NULL
StudentData2$FamilyRelationship <- NULL


##### MODEL 6 ############################
model6 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model6)

##
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5760  -1.5216   0.0819   1.7502   7.1591
##
## Coefficients: (1 not defined because of singularities)
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         10.578809   1.078340   9.810  < 2e-16 ***
## MothersEducation1   -0.828318   0.945612  -0.876  0.38126
## MothersEducation2   -0.833910   0.943730  -0.884  0.37710
## MothersEducation3   -0.804873   0.952284  -0.845  0.39820
## MothersEducation4   -0.003079   0.962807  -0.003  0.99745
## FathersEducation     0.117479   0.104356   1.126  0.26054
## StudyTime            0.434224   0.107676   4.033 5.93e-05 ***
## Failures1           -2.251423   0.282009  -7.984 3.82e-15 ***
## Failures2           -3.335381   0.499624  -6.676 4.04e-11 ***
## Failures3           -3.333812   0.534653  -6.235 6.59e-10 ***
## GoingOut2            1.006916   0.375848   2.679  0.00750 **
## GoingOut3            0.467573   0.363882   1.285  0.19910
## GoingOut4            0.122419   0.377110   0.325  0.74553
## GoingOut5           -0.128202   0.394197  -0.325  0.74508
```

```
## Health                        -0.160402   0.060637  -2.645  0.00829 **
## School.gp                       0.468926   0.221341   2.119  0.03437 *
## HigherEducationInterest.yes     1.385610   0.331951   4.174 3.25e-05 ***
## InternetAcess.yes               0.325658   0.225378   1.445  0.14878
## Address.R                      -0.264765   0.207067  -1.279  0.20131
## FamilySize.GT3                 -0.273774   0.188746  -1.450  0.14723
## SchoolSupport.yes              -1.341649   0.274456  -4.888 1.18e-06 ***
## FamilySupport.yes              -0.307763   0.181701  -1.694  0.09061 .
## PaidClasses.yes                -0.872097   0.215989  -4.038 5.80e-05 ***
## RomanticRelationship.yes       -0.372672   0.181732  -2.051  0.04055 *
## Guardian.M                     -0.357150   0.192082  -1.859  0.06326 .
## MothersJob.Health                    NA         NA      NA       NA
## MothersJob.services             0.633786   0.218625   2.899  0.00382 **
## FathersJob.services            -0.301507   0.196120  -1.537  0.12452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.738 on 1017 degrees of freedom
## Multiple R-squared:  0.2945, Adjusted R-squared:  0.2764
## F-statistic: 16.33 on 26 and 1017 DF,  p-value: < 2.2e-16

StudentData2$FathersEducation <- NULL
StudentData2$FamilySupport.yes <- NULL
StudentData2$Guardian.M <- NULL
StudentData2$Address.R <- NULL


##### MODEL 7 ############################
model7 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model7)

##
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4582  -1.5887   0.0931   1.8128   7.0476
##
## Coefficients: (1 not defined because of singularities)
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              10.402050   1.063234   9.783  < 2e-16 ***
## MothersEducation1        -0.992418   0.946617  -1.048 0.294710
## MothersEducation2        -0.949639   0.943213  -1.007 0.314263
## MothersEducation3        -0.848468   0.946924  -0.896 0.370450
## MothersEducation4         0.005777   0.945165   0.006 0.995124
## StudyTime                 0.413565   0.107067   3.863 0.000119 ***
## Failures1                -2.208558   0.281181  -7.855 1.01e-14 ***
## Failures2                -3.275544   0.498732  -6.568 8.12e-11 ***
## Failures3                -3.428970   0.532287  -6.442 1.81e-10 ***
## GoingOut2                 0.958916   0.376130   2.549 0.010935 *
```

```
## GoingOut3                    0.454247    0.364248    1.247 0.212653
## GoingOut4                    0.111207    0.377700    0.294 0.768487
## GoingOut5                   -0.185734    0.394504   -0.471 0.637881
## Health                      -0.159452    0.060694   -2.627 0.008739 **
## School.gp                    0.550640    0.211922    2.598 0.009503 **
## HigherEducationInterest.yes  1.383503    0.332133    4.166 3.37e-05 ***
## InternetAcess.yes            0.366781    0.224411    1.634 0.102479
## FamilySize.GT3              -0.291778    0.188393   -1.549 0.121747
## SchoolSupport.yes           -1.347727    0.274088   -4.917 1.02e-06 ***
## PaidClasses.yes             -0.930827    0.214701   -4.335 1.60e-05 ***
## RomanticRelationship.yes    -0.371647    0.182193   -2.040 0.041622 *
## MothersJob.Health                  NA          NA       NA       NA
## MothersJob.services          0.610196    0.218501    2.793 0.005326 **
## FathersJob.services         -0.235218    0.195418   -1.204 0.228997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.746 on 1021 degrees of freedom
## Multiple R-squared:  0.2877, Adjusted R-squared:  0.2723
## F-statistic: 18.74 on 22 and 1021 DF,  p-value: < 2.2e-16

StudentData2$InternetAcess.yes <- NULL
StudentData2$FamilySize.GT3 <- NULL
StudentData2$FathersJob.services <- NULL



##### MODEL 8 #############################
model8 <- lm(AverAgeGrades ~ ., data = StudentData2)
summary(model8)

##
## Call:
## lm(formula = AverAgeGrades ~ ., data = StudentData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1957  -1.6055   0.0848   1.7937   6.9750
##
## Coefficients: (1 not defined because of singularities)
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             10.32876    1.05130   9.825  < 2e-16 ***
## MothersEducation1       -1.00978    0.94615  -1.067 0.286109
## MothersEducation2       -0.93596    0.94301  -0.993 0.321177
## MothersEducation3       -0.83865    0.94698  -0.886 0.376038
## MothersEducation4        0.06464    0.94391   0.068 0.945414
## StudyTime                0.40483    0.10717   3.778 0.000167 ***
## Failures1               -2.23201    0.28116  -7.938 5.35e-15 ***
## Failures2               -3.28830    0.49935  -6.585 7.25e-11 ***
## Failures3               -3.52102    0.53188  -6.620 5.79e-11 ***
## GoingOut2                0.98284    0.37640   2.611 0.009155 **
```

```
## GoingOut3                        0.49538     0.36396    1.361 0.173784
## GoingOut4                        0.16508     0.37758    0.437 0.662049
## GoingOut5                       -0.12309     0.39350   -0.313 0.754494
## Health                          -0.16083     0.06054   -2.657 0.008016 **
## School.gp                        0.62226     0.20835    2.987 0.002888 **
## HigherEducationInterest.yes      1.38080     0.33268    4.150 3.59e-05 ***
## SchoolSupport.yes               -1.36109     0.27380   -4.971 7.80e-07 ***
## PaidClasses.yes                 -0.91929     0.21464   -4.283 2.02e-05 ***
## RomanticRelationship.yes        -0.35071     0.18201   -1.927 0.054267 .
## MothersJob.Health                    NA          NA       NA       NA
## MothersJob.services              0.61232     0.21349    2.868 0.004213 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.75 on 1024 degrees of freedom
## Multiple R-squared:  0.2832, Adjusted R-squared:  0.2699
## F-statistic:  21.3 on 19 and 1024 DF,  p-value: < 2.2e-16

# MODEL 8 has R-squared as 0.265 and proves to have significant values in it.
Refer analysis.xlsx
```

SECOND MODEL - MULTINOMIAL LINEAR REGRESSION

```
#----------------------------------------------------------------------
-----------------------------------------------------


# Creating testing and Training Data using random sampling.

set.seed(100)

TrainingDataRows <- sample(1:nrow(StudentData), 0.5*nrow(StudentData))
TrainingData <- StudentData[TrainingDataRows,]
TestingData <- StudentData[-TrainingDataRows,]


library(MASS)


#Working with the multinominal model, Model 1 is a full model that includes a
ll the data columns

library(nnet)

Model1 <- multinom(GradeClassified ~. , data = TrainingData)
```

```
## # weights:  177 (116 variable)
## initial  value 573.475615
## iter  10 value 268.510288
## iter  20 value 161.005700
## iter  30 value 91.326342
## iter  40 value 17.796776
## iter  50 value 0.104183
## iter  60 value 0.000892
## final   value 0.000054
## converged
```

```
summary(Model1)
```

```
## Call:
## multinom(formula = GradeClassified ~ ., data = TrainingData)
##
## Coefficients:
##                 (Intercept)  SchoolMS      SexM      Age AddressU
## AverAge            309.7377 -34.47498 -4.305233 34.0337 26.80762
## BelowAverAge       914.5532 -29.02363 -8.866951 29.0174 34.61877
##                 FamilySizeLE3 ParentsMaritalStatusT MothersEducation1
## AverAge             -44.99166              1.869051          10.39419
## BelowAverAge        -36.65100             40.445180         243.03473
##                 MothersEducation2 MothersEducation3 MothersEducation4
## AverAge                 -30.24506         -34.47009           27.8485
## BelowAverAge            201.12942         189.38651          261.8649
##                 FathersEducation MothersJobhealth MothersJobother
## AverAge                 26.21161       -10.938563       -38.90487
## BelowAverAge            33.19841        -6.306989       -42.24507
##                 MothersJobservices MothersJobteacher FathersJobhealth
## AverAge                  -21.72345         -51.87056        -42.93773
## BelowAverAge             -14.70348         -25.25430        -32.85754
##                 FathersJobother FathersJobservices FathersJobteacher
## AverAge                48.62661           15.42930         -88.07632
## BelowAverAge           69.62778           23.34707         -12.02976
##                 ReasonToChooseSchoolhome ReasonToChooseSchoolother
## AverAge                        -5.700137                  61.03654
## BelowAverAge                  -10.022552                  69.20021
##                 ReasonToChooseSchoolreputation Guardianmother Guardianother
## AverAge                               48.91955      7.1745154      5.995264
## BelowAverAge                          44.78664     -0.8312289     49.322666
##                 TravelTime StudyTime Failures1 Failures2 Failures3
## AverAge           47.11085  2.455486  14.04800 -57.05997 -318.6370
## BelowAverAge      43.96847  4.793192  13.73492 -72.81390 -334.5042
##                 SchoolSupportyes FamilySupportyes PaidClassesyes
## AverAge                 88.37923        -21.99924     -16.091300
## BelowAverAge            86.51983        -27.33472      -6.756314
##                 Activitiesyes Nurseryyes HigherEducationInterestyes
## AverAge              2.152165  -33.00986                   280.6577
## BelowAverAge       -15.580695  -40.92331                   258.9921
```

```
##              InternetAcessyes RomanticRelationshipyes FamilyRelationship
## AverAge             43.13225               -10.76907          -13.09049
## BelowAverAge        36.70913               -12.40041          -13.92751
##              FreeTime GoingOut2 GoingOut3 GoingOut4 GoingOut5
## AverAge      51.02561  50.79704  56.14654  72.33354   9.17286
## BelowAverAge 53.34762  65.17764  59.06718  57.19749  26.22929
##              WorkdayAlcoholLow WorkdayAlcoholMedium WorkdayAlcoholHigh
## AverAge              -9.859824            -100.42709          -90.54408
## BelowAverAge         10.859902             -78.03675          -51.49813
##              WorkdayAlcoholVery High WeekendAlcoholLow
## AverAge                     -170.8211         18.471681
## BelowAverAge                -194.1782          4.610073
##              WeekendAlcoholMedium WeekendAlcoholHigh
## AverAge                  15.04228          -19.22034
## BelowAverAge            -17.58508          -40.11882
##              WeekendAlcoholVery High    Health   Absences FirstPeriodGrade
## AverAge                     60.33238 -5.881429 -1.6353961         3.304309
## BelowAverAge                58.92288 -1.093455 -0.5321485        14.676792
##              SecondPeriodGrade FinalGrade AverAgeGrades
## AverAge               3.485915  -64.63862     -35.15162
## BelowAverAge         15.297677 -156.60111     -51.57762
## 
## Std. Errors:
##              (Intercept) SchoolMS      SexM      Age   AddressU
## AverAge        696.3071 2134.291 3221.9306 2618.346 6701.2081
## BelowAverAge   444.0357 1008.103  806.4928 2765.245  841.0066
##              FamilySizeLE3 ParentsMaritalStatusT MothersEducation1
## AverAge          4261.956              3790.418          342.7478
## BelowAverAge     2697.730              1033.368          357.1664
##              MothersEducation2 MothersEducation3 MothersEducation4
## AverAge               4014.369         3806.7843          1002.631
## BelowAverAge          1073.183          983.3742          1039.001
##              FathersEducation MothersJobhealth MothersJobother
## AverAge              2981.479         305.7892      1392.67965
## BelowAverAge         3000.177         341.3115        18.50987
##              MothersJobservices MothersJobteacher FathersJobhealth
## AverAge                4838.8134          1037.073         951.7798
## BelowAverAge            977.1869          1038.992         951.7798
##              FathersJobother FathersJobservices FathersJobteacher
## AverAge            2277.8203          2138.9642         66.284899
## BelowAverAge        871.2573           352.5881          6.289769
##              ReasonToChooseSchoolhome ReasonToChooseSchoolother
## AverAge                      1076.199                235.4575320
## BelowAverAge                 1076.199                  0.5770737
##              ReasonToChooseSchoolreputation Guardianmother Guardianother
## AverAge                            1680.7755       3012.958      8.588831
## BelowAverAge                        951.9498       2955.322      8.588831
##              TravelTime StudyTime Failures1 Failures2    Failures3
## AverAge        1403.266  4405.941  4.161118  341.2869 2.572130e-07
## BelowAverAge   2290.096  2088.779  4.161118  341.2869 2.469436e-07
```

```
##                 SchoolSupportyes FamilySupportyes PaidClassesyes
## AverAge                2.025745        2616.0983      230.88384
## BelowAverAge           2.025745         936.4812       14.75834
##                 Activitiesyes Nurseryyes HigherEducationInterestyes
## AverAge            1799.1622   6726.024                   990.9797
## BelowAverAge        802.5891   3028.036                   806.5511
##                 InternetAcessyes RomanticRelationshipyes FamilyRelationship
## AverAge                 1552.939               4503.8323           4083.424
## BelowAverAge            1227.800                 27.9558           2326.495
##                 FreeTime    GoingOut2 GoingOut3 GoingOut4 GoingOut5
## AverAge         4811.489 3917.961334  1733.144  4425.174  3601.415
## BelowAverAge    5641.180    2.756997  1050.446  1076.081  3601.407
##                 WorkdayAlcoholLow WorkdayAlcoholMedium WorkdayAlcoholHigh
## AverAge                  5004.388             2846.779           4.183281
## BelowAverAge             3267.734             2846.779           4.183281
##                 WorkdayAlcoholVery High WeekendAlcoholLow
## AverAge                        6.289769       2061.387369
## BelowAverAge                   6.289769          6.204273
##                 WeekendAlcoholMedium WeekendAlcoholHigh
## AverAge                     3238.697           3075.112
## BelowAverAge                3251.888           2935.972
##                 WeekendAlcoholVery High   Health Absences FirstPeriodGrade
## AverAge                       5229.7286 4968.997 7302.138          4118.50
## BelowAverAge                   908.6234 4130.041 5793.737          3644.37
##                 SecondPeriodGrade FinalGrade AverAgeGrades
## AverAge                  6519.761   5978.742      848.6614
## BelowAverAge             6912.746   3722.844      867.1863
##
## Residual Deviance: 0.0001089357
## AIC: 232.0001

library(caret)

# Predicting our model 1

Prediction = predict(Model1, newdata=TestingData)
Accuracy <- table(Prediction, TestingData[,"GradeClassified"])
sum(diag(Accuracy))/sum(Accuracy)

## [1] 0.9214559

confusionMatrix(data=Prediction, TestingData$GradeClassified)

## Confusion Matrix and Statistics
##
##                Reference
## Prediction      AboveAverAge AverAge BelowAverAge
##    AboveAverAge           48      15            0
##    AverAge                12     335            8
##    BelowAverAge            0       6           98
##
```

```
## Overall Statistics
##
##                  Accuracy : 0.9215
##                    95% CI : (0.895, 0.943)
##       No Information Rate : 0.682
##       P-Value [Acc > NIR] : < 2.2e-16
##
##                     Kappa : 0.837
##    Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: AboveAverAge Class: AverAge
## Sensitivity                      0.80000         0.9410
## Specificity                      0.96753         0.8795
## Pos Pred Value                   0.76190         0.9437
## Neg Pred Value                   0.97386         0.8743
## Prevalence                       0.11494         0.6820
## Detection Rate                   0.09195         0.6418
## Detection Prevalence             0.12069         0.6801
## Balanced Accuracy                0.88377         0.9103
##                      Class: BelowAverAge
## Sensitivity                       0.9245
## Specificity                       0.9856
## Pos Pred Value                    0.9423
## Neg Pred Value                    0.9809
## Prevalence                        0.2031
## Detection Rate                    0.1877
## Detection Prevalence              0.1992
## Balanced Accuracy                 0.9551
```

#From the accuracy calculation, it has predicted the grade calssified at rate of 92%. The confusion matrix identified the wrongly predicted grades
#We shouldnt probably consider this since this includes the Final grade column with which we derived our Grade classified

 # model 2 multinominal , reduced model

Model2 <- **multinom**(GradeClassified **~** MothersEducation **+**RomanticRelationship **+** WeekendAlcohol **+** GoingOut **+**
                    HigherEducationInterest **+** MothersEducation **+** StudyTime **+** InternetAcess **+** Health , data = TrainingData)

```
## # weights:  57 (36 variable)
## initial  value 573.475615
## iter  10 value 430.070786
## iter  20 value 423.113564
## iter  30 value 420.907412
## iter  40 value 420.355238
## iter  50 value 420.284721
```

```
## final  value 420.284578
## converged
```

```
# Predicting our model 2

Prediction2 = predict(Model2, newdata=TestingData)
Accuracy <- table(Prediction2, TestingData[,"GradeClassified"])
sum(diag(Accuracy))/sum(Accuracy)
```

```
## [1] 0.6743295
```

```
confusionMatrix(data=Prediction2, TestingData$GradeClassified)
```

```
## Confusion Matrix and Statistics
##
##                Reference
## Prediction     AboveAverAge AverAge BelowAverAge
##    AboveAverAge           1       0            0
##    AverAge               59     337           92
##    BelowAverAge           0      19           14
##
## Overall Statistics
##
##                Accuracy : 0.6743
##                  95% CI : (0.6323, 0.7144)
##     No Information Rate : 0.682
##     P-Value [Acc > NIR] : 0.6655
##
##                   Kappa : 0.0678
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: AboveAverAge Class: AverAge
## Sensitivity                     0.016667        0.94663
## Specificity                     1.000000        0.09036
## Pos Pred Value                  1.000000        0.69057
## Neg Pred Value                  0.886756        0.44118
## Prevalence                      0.114943        0.68199
## Detection Rate                  0.001916        0.64559
## Detection Prevalence            0.001916        0.93487
## Balanced Accuracy               0.508333        0.51850
##                      Class: BelowAverAge
## Sensitivity                      0.13208
## Specificity                      0.95433
## Pos Pred Value                   0.42424
## Neg Pred Value                   0.81186
## Prevalence                       0.20307
## Detection Rate                   0.02682
## Detection Prevalence             0.06322
## Balanced Accuracy                0.54320
```

```
#From the accuracy calculation, it has predicted the grade calssified  at rat
e of 67%. The confusion matrix identified the wrongly predicted grades which
is HigherEducationInterest compared to Model 2

#Model 3 - adding  more variables

Model3 <- multinom(GradeClassified ~ Address + FathersEducation + TravelTime
+ Failures + SchoolSupport + MothersEducation +RomanticRelationship + Weekend
Alcohol  +  GoingOut +
                   HigherEducationInterest + MothersEducation  + StudyTime
+ InternetAcess + Health +FreeTime + WorkdayAlcohol  , data = TrainingData)

## # weights:  93 (60 variable)
## initial  value 573.475615
## iter  10 value 431.366933
## iter  20 value 398.255182
## iter  30 value 394.424336
## iter  40 value 392.374278
## iter  50 value 391.626982
## iter  60 value 391.365770
## iter  70 value 391.355588
## final  value 391.355558
## converged
```

IMPLEMENTING KNN METHOD

```
StudDataChunk1 <- StudentDataBackup

# Converting categorical into numerical values

unique(sort(StudentDataBackup$school))

## NULL

StudDataChunk1$Sex <-  as.numeric(factor(StudDataChunk1$Sex,
                                    levels=c("F","M")))


StudDataChunk1$Address <-  as.numeric(factor(StudDataChunk1$Address,
                                    levels=c("R","U")))

StudDataChunk1$ParentsMaritalStatus <-  as.numeric(factor(StudDataChunk1$Pare
ntsMaritalStatus,
                              levels=c("A","T")))

StudDataChunk1$FamilySize <-  as.numeric(factor(StudDataChunk1$FamilySize,
                                    levels=c("GT3","LE3")))
```

```r
StudDataChunk1$School <-  as.numeric(factor(StudDataChunk1$School,
                                       levels=c("GP","MS")))

StudDataChunk1$MothersJob <-  as.numeric(factor(StudDataChunk1$MothersJob,
                              levels=c("at_home","health","other","service
s","teacher")))

StudDataChunk1$FathersJob <-  as.numeric(factor(StudDataChunk1$FathersJob,
                              levels=c("at_home","health","other","service
s","teacher")))

StudDataChunk1$ReasonToChooseSchool <-  as.numeric(factor(StudDataChunk1$Reas
onToChooseSchool,
                              levels=c("course","home","other","reputation
")))

StudDataChunk1$Guardian <-  as.numeric(factor(StudDataChunk1$Guardian,
                              levels=c("father","mother","other" )))

StudDataChunk1$SchoolSupport <-  as.numeric(factor(StudDataChunk1$SchoolSuppo
rt,
                              levels=c("no","yes")))

StudDataChunk1$FamilySupport <-  as.numeric(factor(StudDataChunk1$FamilySuppo
rt,
                              levels=c("no","yes")))

StudDataChunk1$PaidClasses <-  as.numeric(factor(StudDataChunk1$PaidClasses,
                              levels=c("no","yes")))

StudDataChunk1$Activities <-  as.numeric(factor(StudDataChunk1$Activities,
                              levels=c("no","yes")))

StudDataChunk1$Nursery <-  as.numeric(factor(StudDataChunk1$Nursery,
                              levels=c("no","yes")))

StudDataChunk1$HigherEducationInterest <-  as.numeric(factor(StudDataChunk1$H
igherEducationInterest,
                              levels=c("no","yes")))

StudDataChunk1$InternetAcess <-  as.numeric(factor(StudDataChunk1$InternetAce
ss,
                              levels=c("no","yes")))

StudDataChunk1$RomanticRelationship <-  as.numeric(factor(StudDataChunk1$Roma
nticRelationship,
                              levels=c("no","yes")))
```

```r
StudDataChunk1$AverAgeGrade <- round(rowMeans(cbind(StudDataChunk1$FirstPerio
dGrade, StudDataChunk1$SecondPeriodGrade, StudDataChunk1$FinalGrade)),2)


StudDataChunk1$GradeClassified <- ifelse(StudDataChunk1$AverAgeGrade<12,"Fail
","Pass")


StudDataChunk2 <- StudDataChunk1

# Function to Normalize the data


Normalize <- function(x) {
  return((x - min(x))/ (max(x) -  min(x)))
}

# Normalising the student performance data

StudData_Norm <-  as.data.frame(lapply(StudDataChunk2[,c(1:30)], Normalize))
StudData_Norm$GradesClassified <- StudDataChunk2$GradeClassified


# Sampling rows out of it & Training the model - 80% train and 20% testing


set.seed(1234)
index=sample(1:2, length(StudDataChunk2$GradeClassified),replace= T, prob = c
(.8,.2))

StudData_train <- StudData_Norm[index==1,]
StudData_test  <- StudData_Norm[index==2,]


PredictedModel <- knn(train = StudData_train[,1:30], test = StudData_test[,1:
30],
                 cl = StudData_train$GradesClassified, k=21)

# Evaluating the model

CrossTable(x = StudData_test$GradesClassified, y= PredictedModel, prop.chisq
= FALSE)

## 
## 
##     Cell Contents
## |-------------------------|
## |                       N |
```

```
## |            N / Row Total |
## |            N / Col Total |
## |          N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  218
##
##
##                                | PredictedModel
## StudData_test$GradesClassified |      Fail |      Pass | Row Total |
## ------------------------------|-----------|-----------|-----------|
##                         Fail |        76 |        42 |       118 |
##                              |     0.644 |     0.356 |     0.541 |
##                              |     0.724 |     0.372 |           |
##                              |     0.349 |     0.193 |           |
## ------------------------------|-----------|-----------|-----------|
##                         Pass |        29 |        71 |       100 |
##                              |     0.290 |     0.710 |     0.459 |
##                              |     0.276 |     0.628 |           |
##                              |     0.133 |     0.326 |           |
## ------------------------------|-----------|-----------|-----------|
##                 Column Total |       105 |       113 |       218 |
##                              |     0.482 |     0.518 |           |
## ------------------------------|-----------|-----------|-----------|
##
##
##    k  Fail   Pass    Match   TOTAL   ACCURACY
##    20  76    69      145     218     0.7005
##    25  72    68      140     218     0.6763
##    28  65    71      136     218     0.6570
##    21  76    71      147     218     0.7053 ****
##    22  72    69      141     218     0.6812
##    19  74    69      143     218     0.6908
```