Final Project

Introduction

The Data

We will be using data that has extensive information on secondary school students in their math class.

```
data <- read.csv("data/student-mat.csv")
```

Creation of New Variables

In order to provide more insight, we saw room to create informative variables based upon the given data.

The given variables Medu and Fedu give information about the student's parents education history. Using this, we created a new variable "first_gen_college" that indicates if the student would be a first generation college student if they decided to pursue higher education. This will give more tangible and clear insight to how parental education impacts student's performance.

```
data <- data %>%
  mutate(first_gen_college = case_when(
    Medu < 4 & Fedu < 4 ~ "yes",
    TRUE ~"no"
  ))
data[["first_gen_college"]] <- as.factor(data[["first_gen_college"]])
```

Additionally, many variables are self reported ratings from the students on a scale of 1-5. We decided that instead factoring these variables so that scores of 1-3 would be "low" and scores of 4-5 would be "high" would be beneficial to our analysis as it would be more interpretable in context.

```
data <- data %>%
  mutate(famrel = case_when(
    famrel == 1 ~ "low",
    famrel == 2 ~ "low",
    famrel == 3 ~ "low",
    famrel == 4 ~ "high",
    famrel == 5 ~"high"
  ))

data <- data %>%
  mutate(freetime = case_when(
    freetime == 1 ~ "low",
    freetime == 2 ~ "low",
    freetime == 3 ~ "low",
    freetime == 4 ~ "high",
    freetime == 5 ~"high"
  ))

data <- data %>%
  mutate(goout = case_when(
    goout == 1 ~ "low",
    goout == 2 ~ "low",
```

```
    goout == 3 ~ "low",
    goout == 4 ~ "high",
    goout == 5 ~"high"
  ))

data <- data %>%
  mutate(Dalc = case_when(
    Dalc == 1 ~ "low",
    Dalc == 2 ~ "low",
    Dalc == 3 ~ "low",
    Dalc == 4 ~ "high",
    Dalc == 5 ~"high"
  ))

data <- data %>%
  mutate(Walc = case_when(
    Walc == 1 ~ "low",
    Walc == 2 ~ "low",
    Walc == 3 ~ "low",
    Walc == 4 ~ "high",
    Walc == 5 ~"high"
  ))

data <- data %>%
  mutate(health = case_when(
    health == 1 ~ "low",
    health == 2 ~ "low",
    health == 3 ~ "low",
    health == 4 ~ "high",
    health == 5 ~"high"
  ))

data[["sex"]] <- as.factor(data[["sex"]])
data[["address"]] <- as.factor(data[["address"]])
data[["famsize"]] <- as.factor(data[["famsize"]])
data[["Pstatus"]] <- as.factor(data[["Pstatus"]])
data[["Mjob"]] <- as.factor(data[["Mjob"]])
data[["Fjob"]] <- as.factor(data[["Fjob"]])
data[["reason"]] <- as.factor(data[["reason"]])
data[["guardian"]] <- as.factor(data[["guardian"]])
data[["schoolsup"]] <- as.factor(data[["schoolsup"]])
data[["famsup"]] <- as.factor(data[["famsup"]])
data[["paid"]] <- as.factor(data[["paid"]])
data[["activities"]] <- as.factor(data[["activities"]])
data[["nursery"]] <- as.factor(data[["nursery"]])
data[["higher"]] <- as.factor(data[["higher"]])
data[["internet"]] <- as.factor(data[["internet"]])
data[["romantic"]] <- as.factor(data[["romantic"]])
data[["famrel"]] <- as.factor(data[["famrel"]])
data[["freetime"]] <- as.factor(data[["freetime"]])
data[["goout"]] <- as.factor(data[["goout"]])
data[["Dalc"]] <- as.factor(data[["Dalc"]])
data[["Walc"]] <- as.factor(data[["Walc"]])
```

```
data[["health"]] <- as.factor(data[["health"]])
```

Additionally, using information from the famsup and internet variables, we created a variable called "stable_learning_env". If famsup is "yes" and internet is "yes", then stable_learning_env is "yes", otherwise "no".

```
data <- data %>%
  mutate(stable_learning_env = case_when(
    internet =="yes" & famsup =="yes" ~"yes",
    TRUE ~"no"
  ))
data[["stable_learning_env"]] <- as.factor(data[["stable_learning_env"]])
```

Also, we created a new variable "high_freq_absent", which if absences >= 10 for a student, we considered them a highly frequent student.

```
data <- data %>%
  mutate(high_freq_absent = case_when(
    absences >= 10 ~"yes",
    TRUE ~"no"
  ))
data[["high_freq_absent"]] <- as.factor(data[["high_freq_absent"]])
```

We also created a "failed" variable, which was "yes" if failures > 0, and "no" otherwise.

```
data <- data %>%
  mutate(failed = case_when(
    failures > 0 ~"yes",
    TRUE ~"no"
  ))
data[["failed"]] <- as.factor(data[["failed"]])
```

Exploratory Data Analysis

```
summary(data)
```

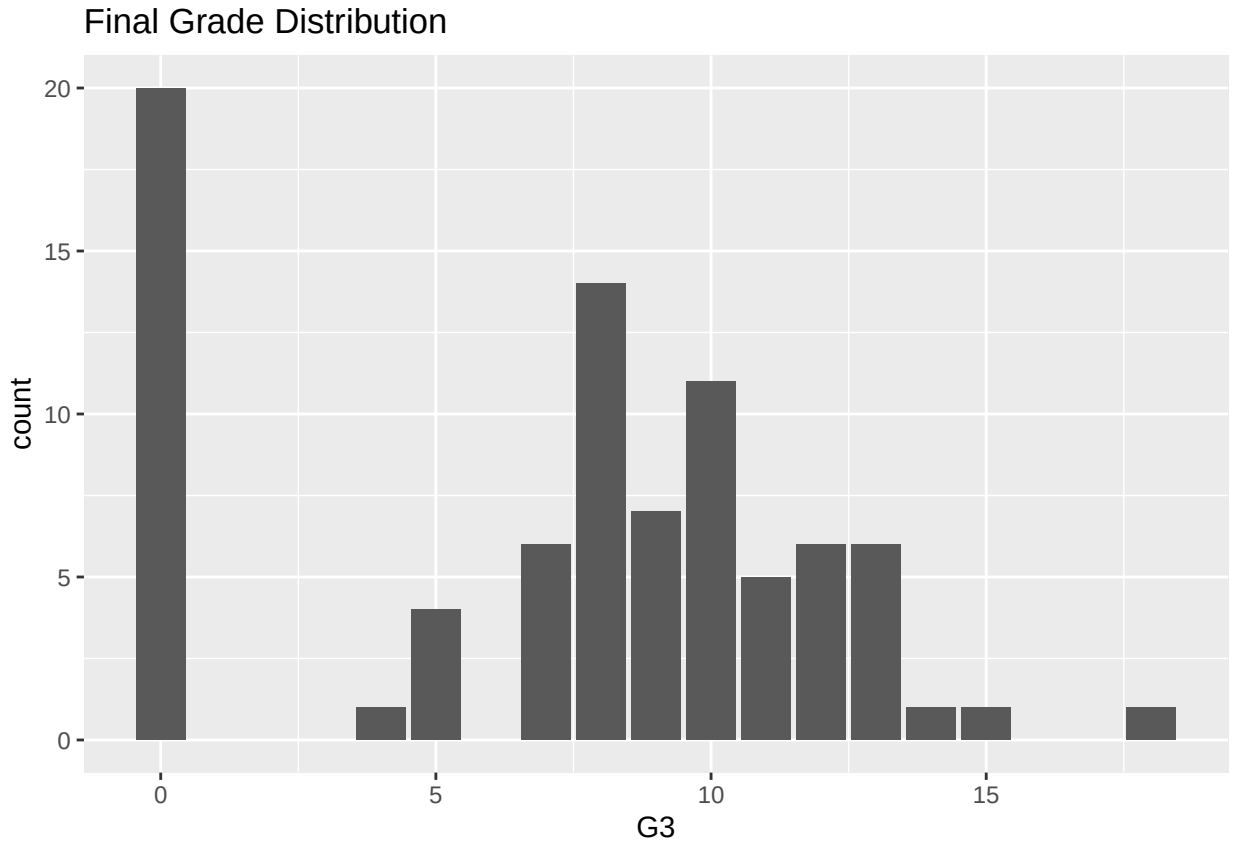
```
##     school           sex          age       address famsize   Pstatus
##  Length:395         F:208   Min.   :15.0   R: 88   GT3:281   A: 41
##  Class :character   M:187   1st Qu.:16.0   U:307   LE3:114   T:354
##  Mode  :character           Median :17.0
##                             Mean   :16.7
##                             3rd Qu.:18.0
##                             Max.   :22.0
##       Medu            Fedu             Mjob           Fjob            reason
##  Min.   :0.000   Min.   :0.000   at_home : 59   at_home : 20   course    :145
##  1st Qu.:2.000   1st Qu.:2.000   health  : 34   health  : 18   home      :109
##  Median :3.000   Median :2.000   other   :141   other   :217   other     : 36
##  Mean   :2.749   Mean   :2.522   services:103   services:111   reputation:105
##  3rd Qu.:4.000   3rd Qu.:3.000   teacher : 58   teacher : 29
##  Max.   :4.000   Max.   :4.000
##    guardian     traveltime      studytime        failures      schoolsup
##  father: 90   Min.   :1.000   Min.   :1.000   Min.   :0.0000   no :344
##  mother:273   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.0000   yes: 51
##  other : 32   Median :1.000   Median :2.000   Median :0.0000
##               Mean   :1.448   Mean   :2.035   Mean   :0.3342
##               3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:0.0000
```

```
##                Max.   :4.000   Max.   :4.000   Max.   :3.0000
##  famsup     paid     activities nursery   higher    internet  romantic
##  no :153   no :214   no :194    no : 81   no : 20   no : 66   no :263
##  yes:242   yes:181   yes:201    yes:314   yes:375   yes:329   yes:132
##
##
##
##
##   famrel    freetime    goout       Dalc       Walc      health
##  high:301   high:155   high:139   high: 18   high: 79   high:212
##  low : 94   low :240   low :256   low :377   low :316   low :183
##
##
##
##
##     absences            G1              G2              G3
##  Min.   : 0.000   Min.   : 3.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 0.000   1st Qu.: 8.00   1st Qu.: 9.00   1st Qu.: 8.00
##  Median : 4.000   Median :11.00   Median :11.00   Median :11.00
##  Mean   : 5.709   Mean   :10.91   Mean   :10.71   Mean   :10.42
##  3rd Qu.: 8.000   3rd Qu.:13.00   3rd Qu.:13.00   3rd Qu.:14.00
##  Max.   :75.000   Max.   :19.00   Max.   :19.00   Max.   :20.00
##  first_gen_college stable_learning_env high_freq_absent failed
##  no :157           no :186             no :312          no :312
##  yes:238           yes:209             yes: 83          yes: 83
##
##
##
##
```

First, I will start off with univariate and bivariate plots of the response variable and key predictors I see being important.

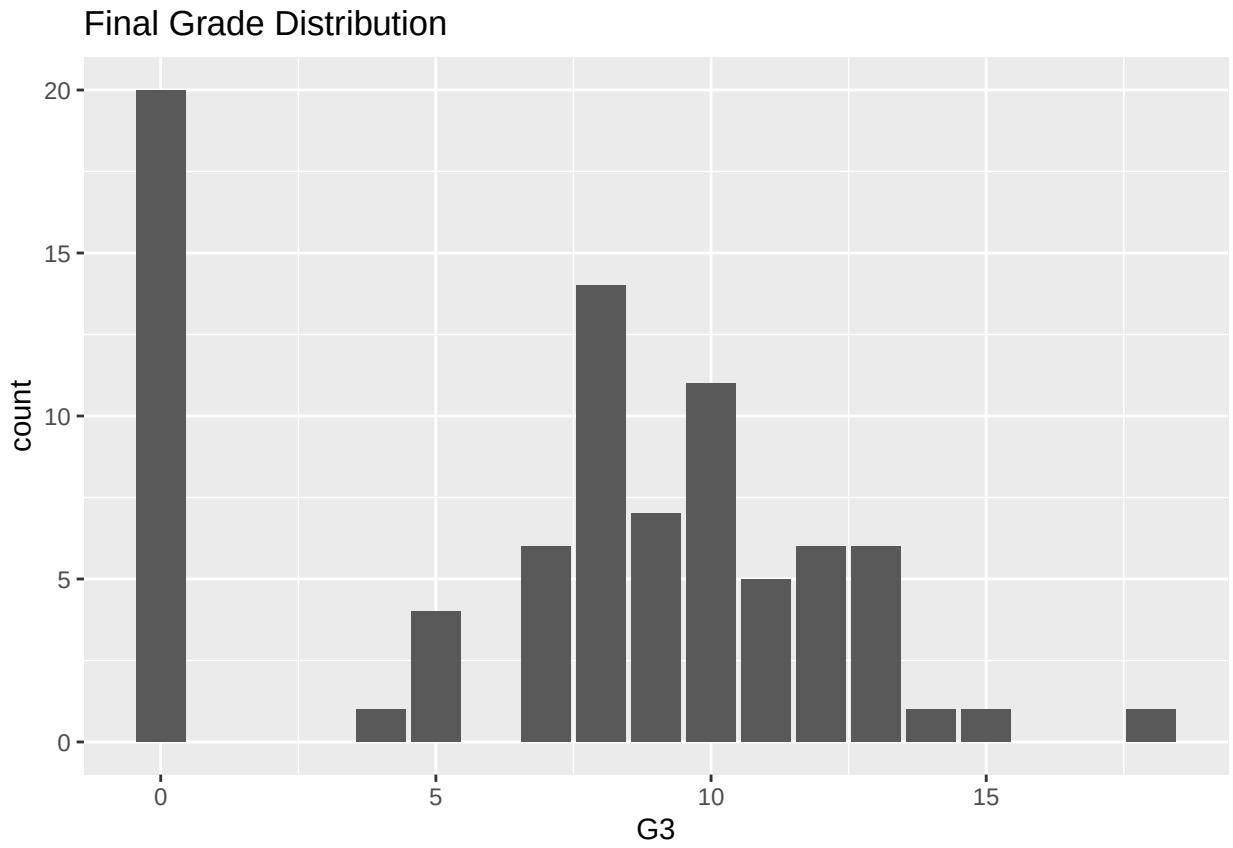
```
data %>%
  filter(failed =="yes") %>%
  ggplot(aes(G3)) +
  geom_histogram(stat = "count") +
  labs(title="Final Grade Distribution")
```

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



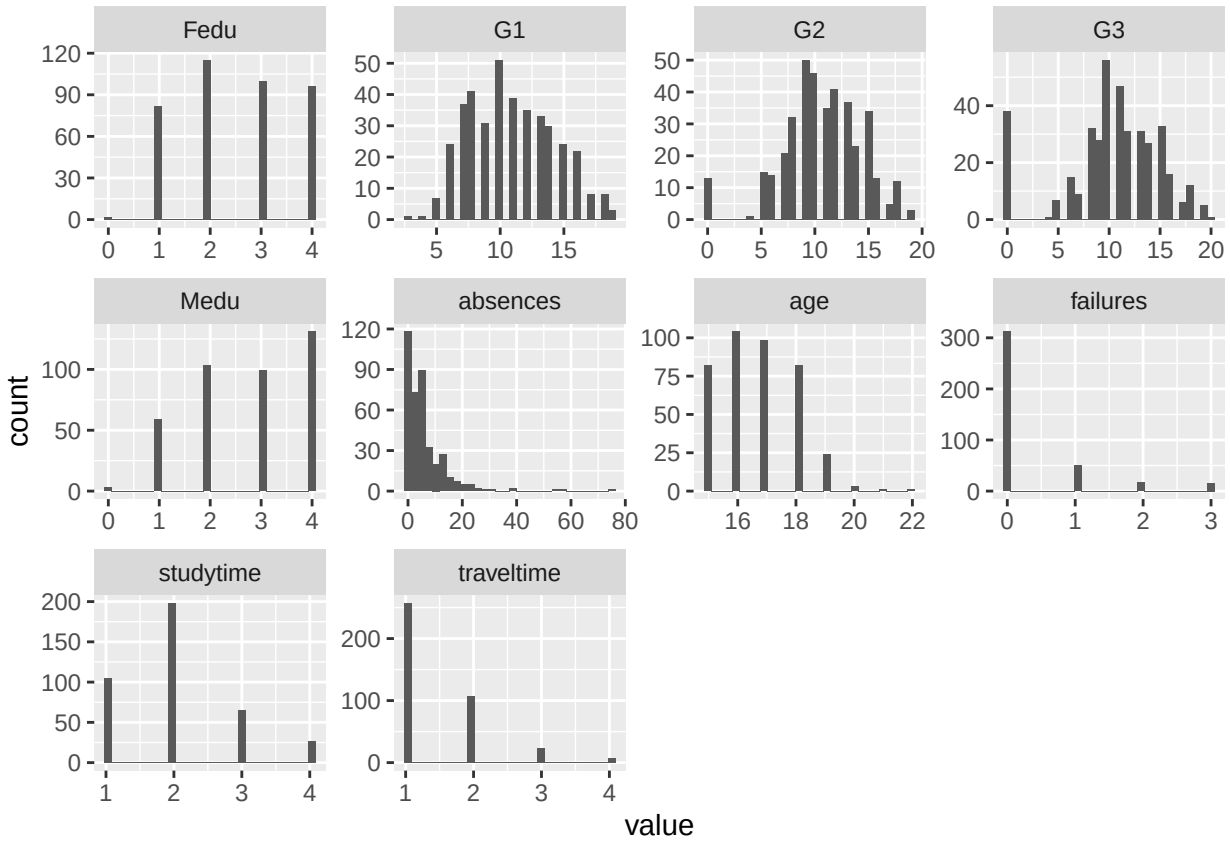

```
data %>%
  filter(failed =="yes") %>%
  ggplot(aes(G3)) +
  geom_histogram(stat = "count") +
  labs(title="Final Grade Distribution")
```

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



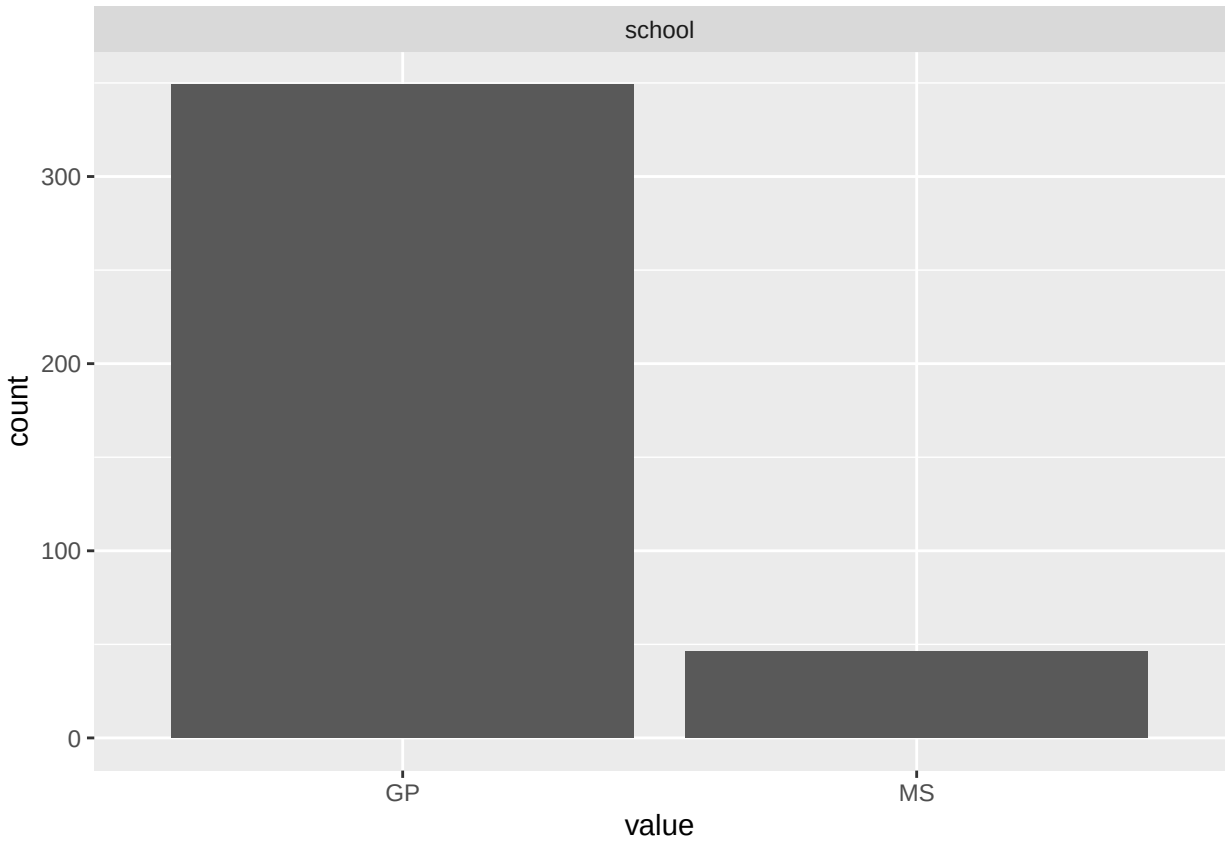

```
data %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
    facet_wrap(~ key, scales = "free") +
    geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



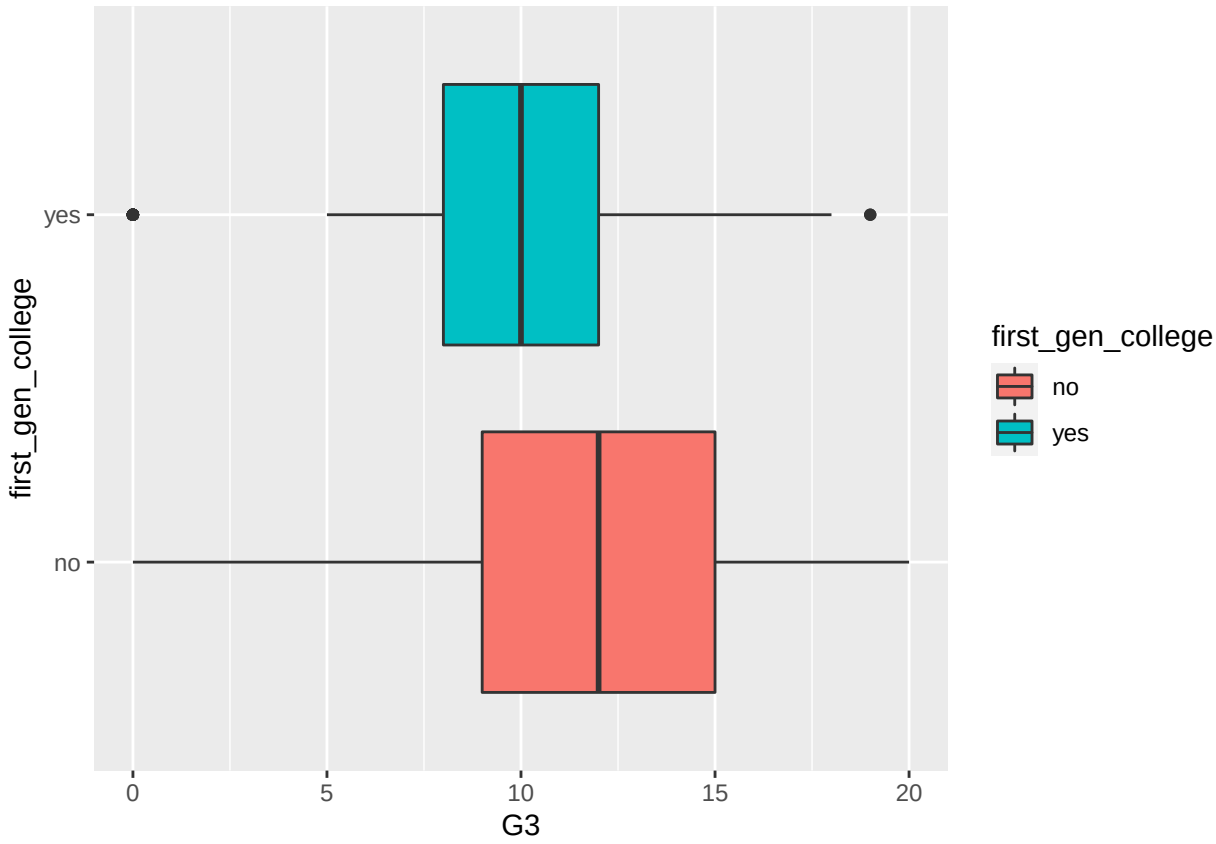

```
data %>%
  keep(is.character) %>%
  gather() %>%
  ggplot(aes(value)) +
    facet_wrap(~ key, scales = "free") +
    geom_histogram(stat="count")
```

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

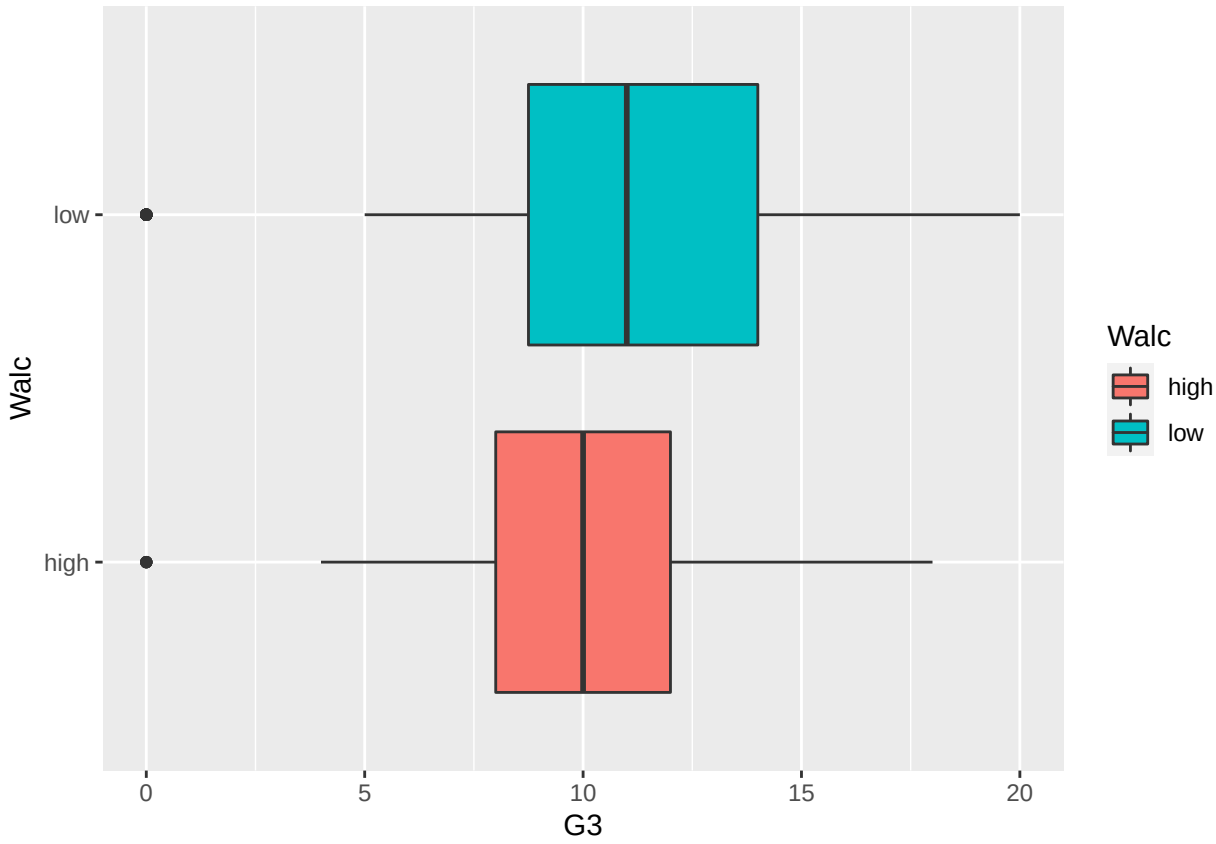


Above we see that the response variable, G3, is pretty normally distributed, thus no transformation is necessary,

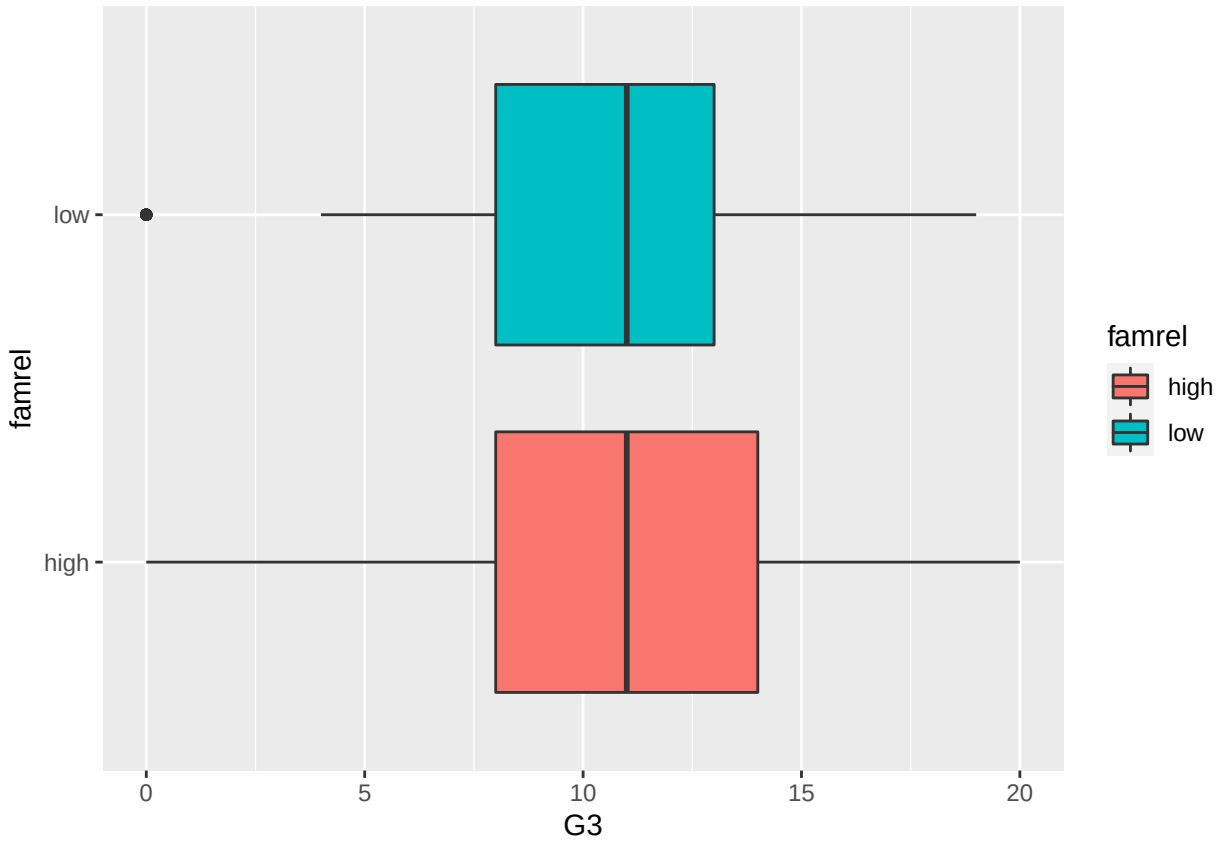
```
ggplot(data = data, aes(x = G3, y = first_gen_college, fill=first_gen_college)) +
  geom_boxplot()
```



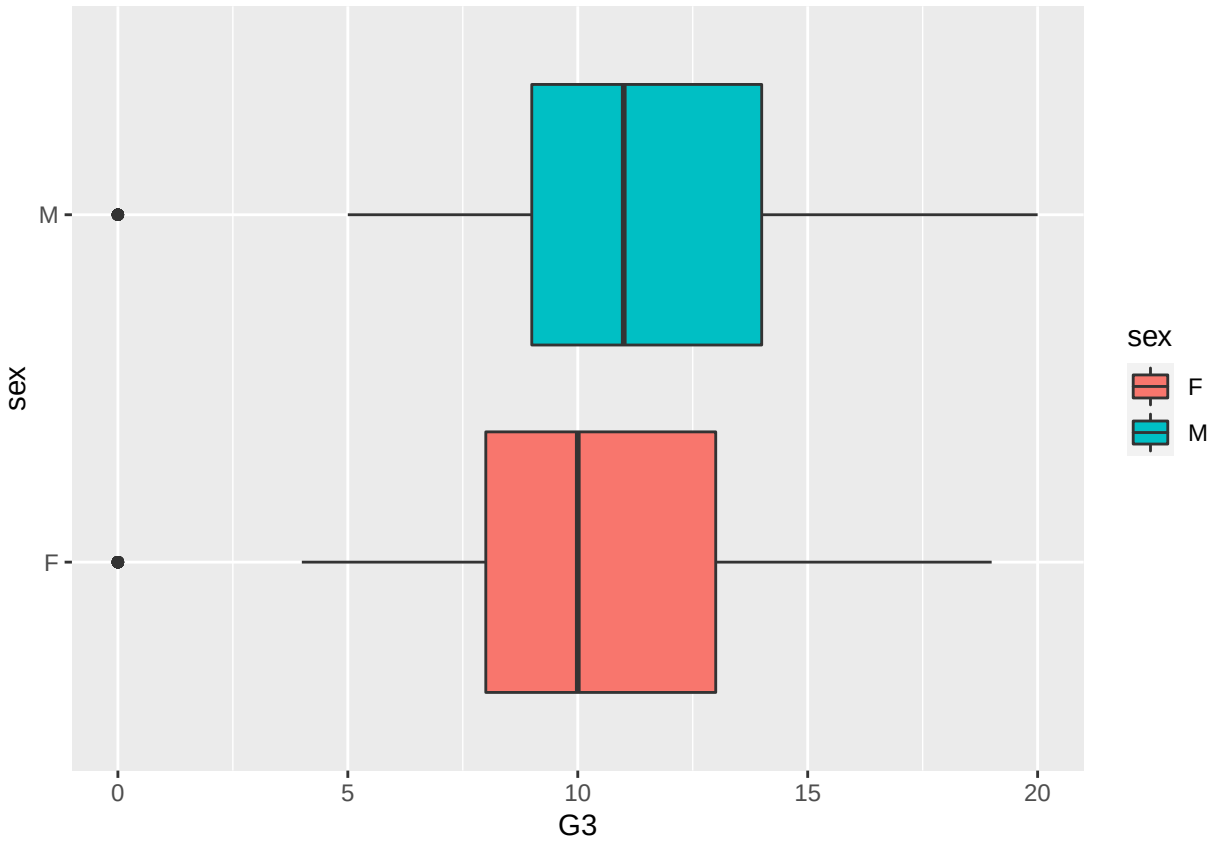

```
ggplot(data = data, aes(x = G3, y = Walc, fill = Walc)) +
  geom_boxplot()
```



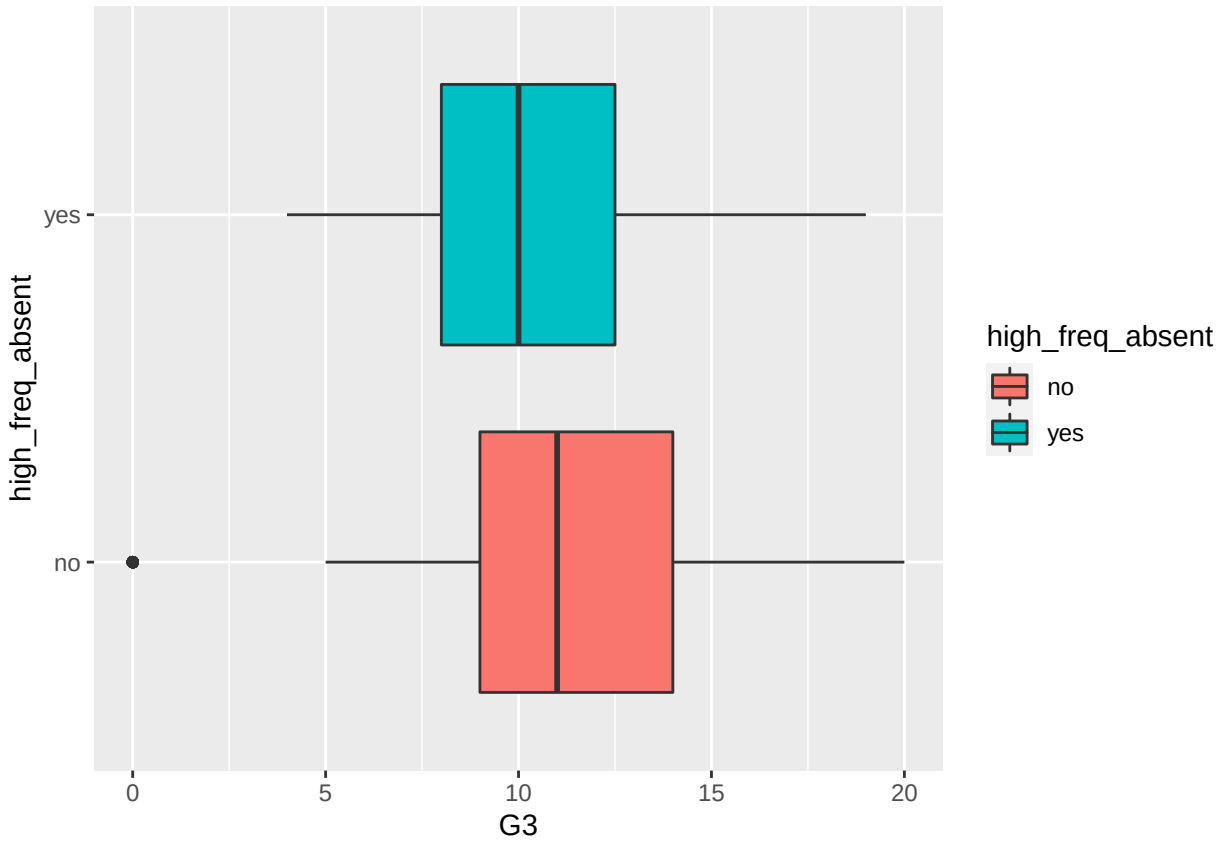

```
ggplot(data = data, aes(x = G3, y = famrel, fill = famrel)) +
  geom_boxplot()
```



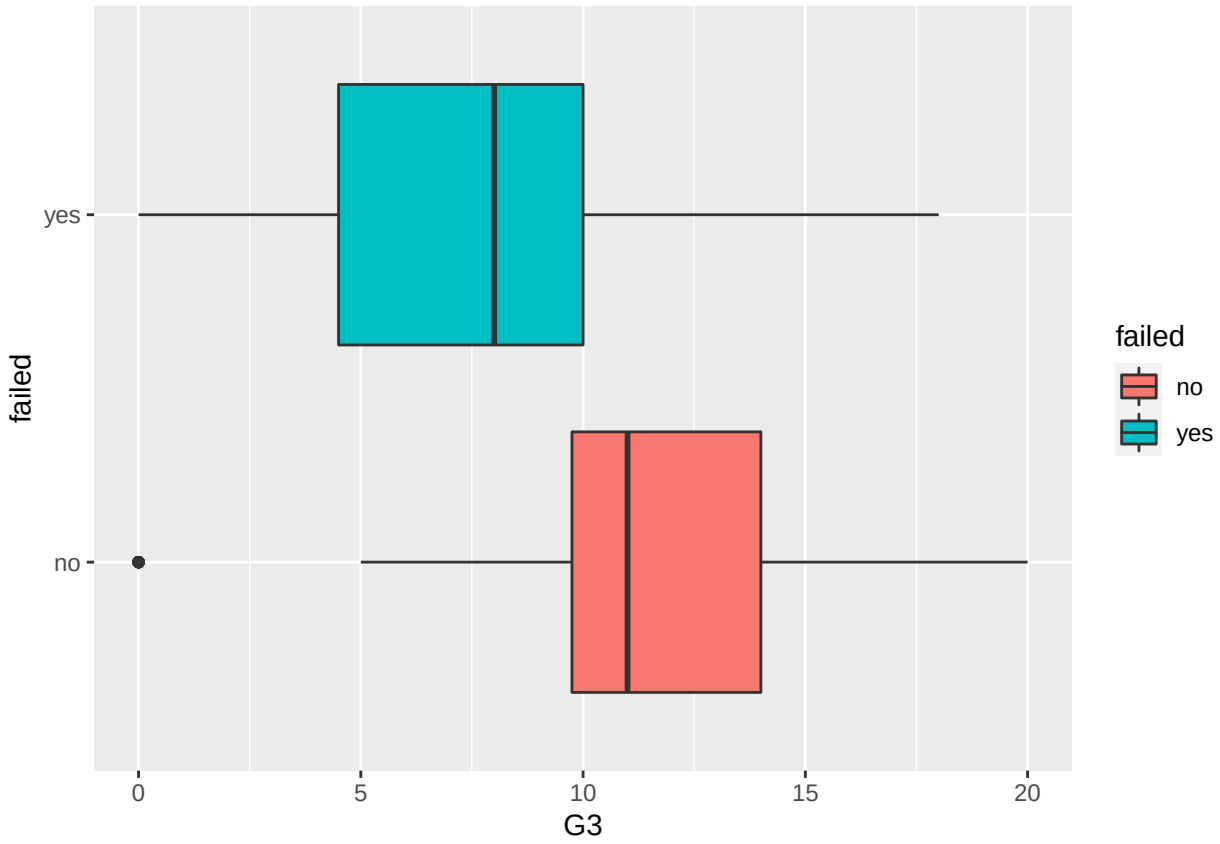

```
ggplot(data = data, aes(x = G3, y= sex, fill = sex)) +
  geom_boxplot()
```



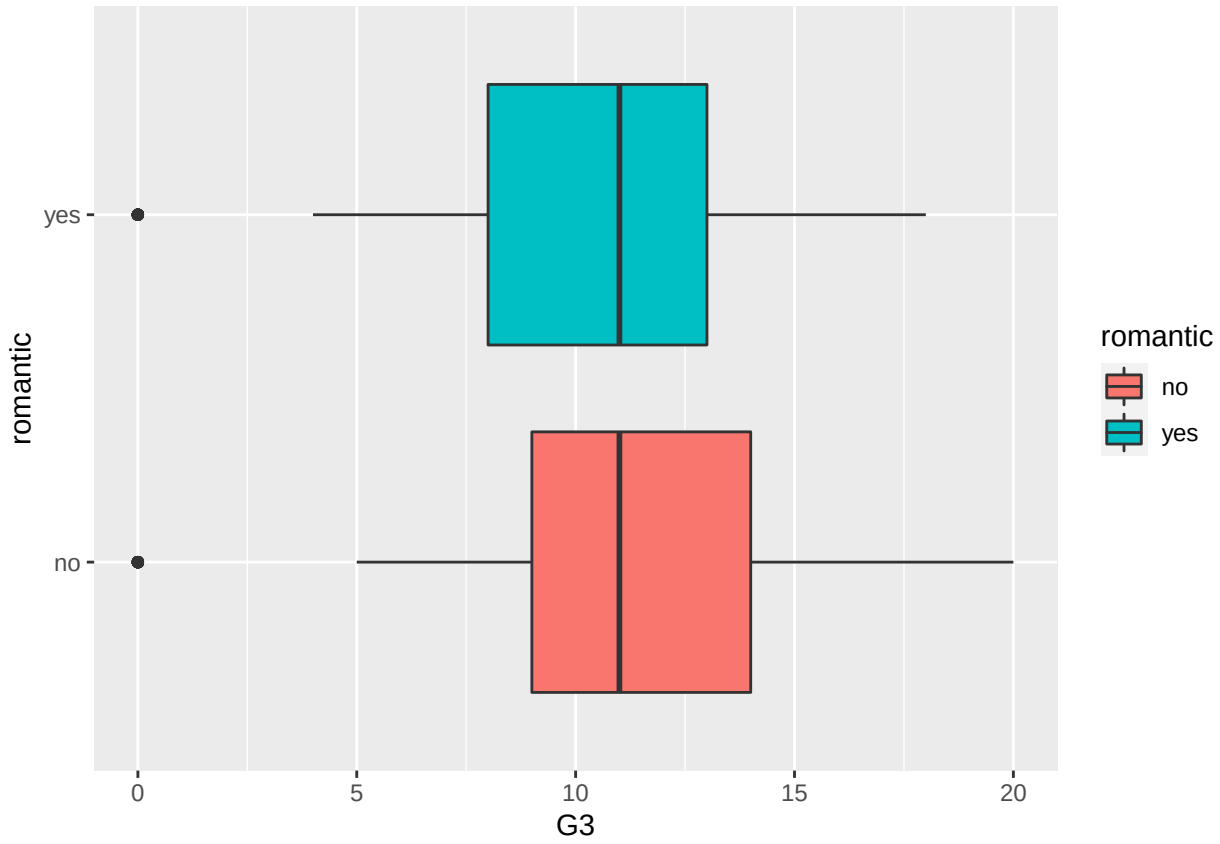

```
ggplot(data = data, aes(x = G3, y = high_freq_absent, fill = high_freq_absent)) +
  geom_boxplot()
```



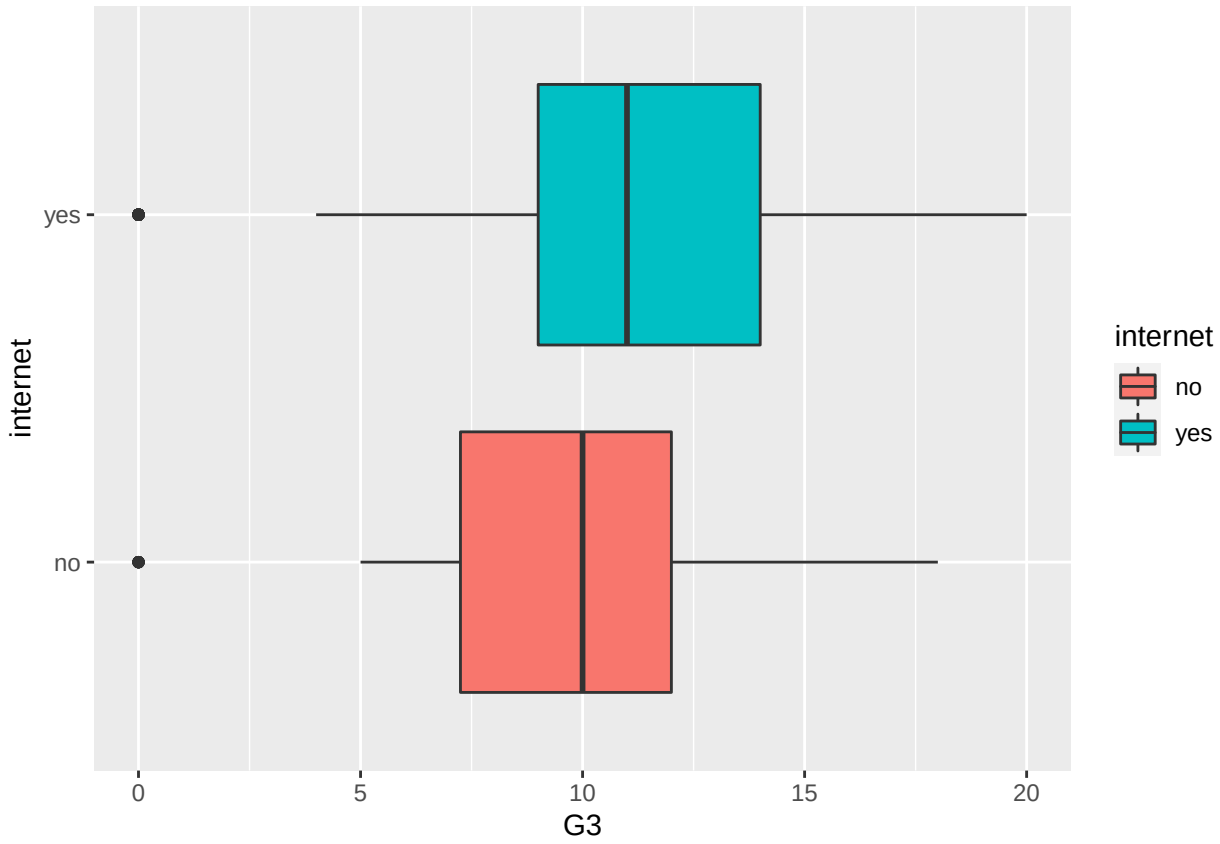

```
ggplot(data = data, aes(x = G3, y=failed, fill = failed)) +
  geom_boxplot()
```



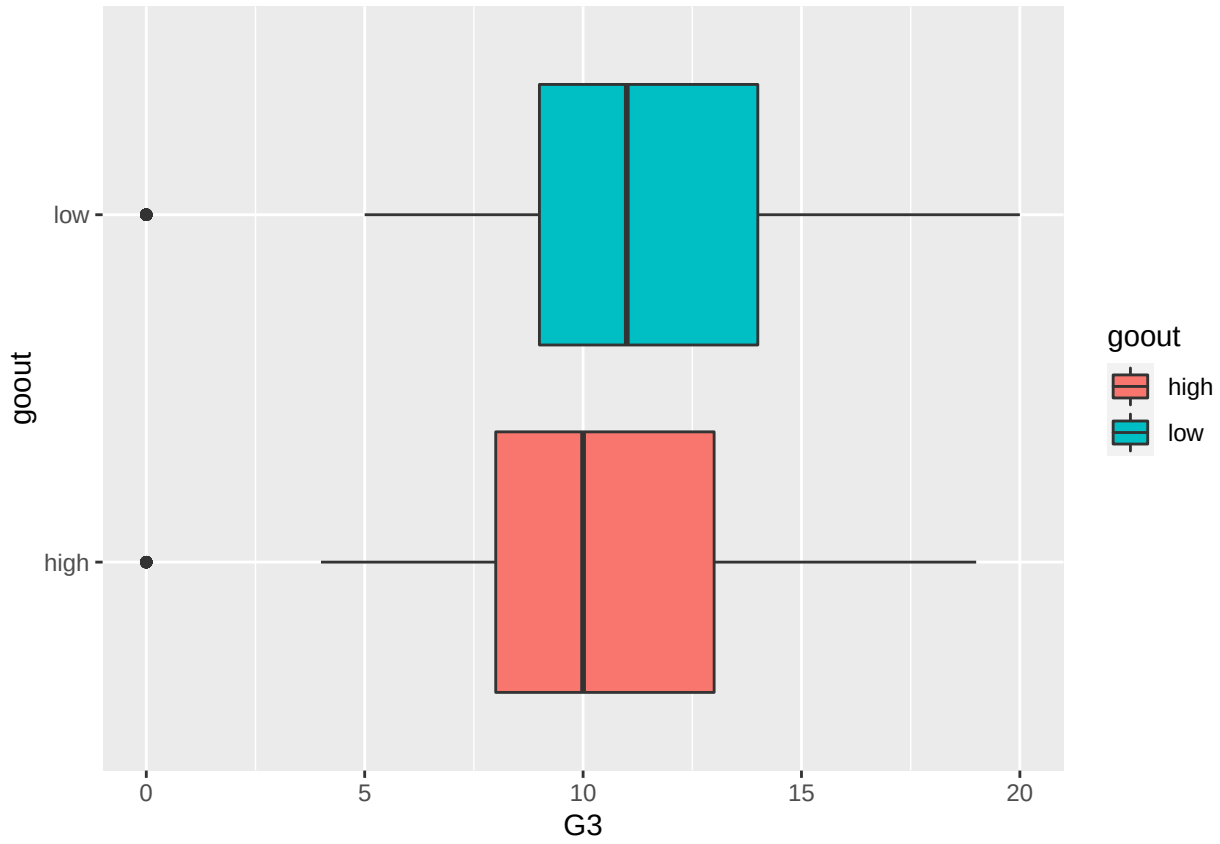

```
ggplot(data = data, aes(x = G3, y=romantic, fill = romantic)) +
  geom_boxplot()
```




```
ggplot(data = data, aes(x = G3, y=internet, fill = internet)) +
  geom_boxplot()
```




```
ggplot(data = data, aes(x = G3, y=goout, fill = goout)) +
  geom_boxplot()
```



From the initial explorations above, we can see a few possible trends. Students who had at least one of the following traits: failed a class previously, were a highly frequent absent student, frequently went out, without internet, were frequent drinkers on the weekend, were in romantic relationships, and were first generation students, on average had lower final grades than their counterparts.

```
names(data)
```

```
##  [1] "school"              "sex"                 "age"
##  [4] "address"             "famsize"             "Pstatus"
##  [7] "Medu"                "Fedu"                "Mjob"
## [10] "Fjob"                "reason"              "guardian"
## [13] "traveltime"          "studytime"           "failures"
## [16] "schoolsup"           "famsup"              "paid"
## [19] "activities"          "nursery"             "higher"
## [22] "internet"            "romantic"            "famrel"
## [25] "freetime"            "goout"               "Dalc"
## [28] "Walc"                "health"              "absences"
## [31] "G1"                  "G2"                  "G3"
## [34] "first_gen_college"   "stable_learning_env" "high_freq_absent"
## [37] "failed"
```

```
num_cols <- unlist(lapply(data, is.numeric))
quant_vars <- data[,num_cols]
cor(quant_vars)
```

```
##                    age        Medu         Fedu  traveltime    studytime
## age         1.000000000 -0.16365842 -0.163438069  0.07064072 -0.004140037
## Medu       -0.163658419  1.00000000  0.623455112 -0.17163930  0.064944137
## Fedu       -0.163438069  0.62345511  1.000000000 -0.15819405 -0.009174639
```

```
## traveltime  0.070640721 -0.17163930 -0.158194054  1.00000000 -0.100909119
## studytime  -0.004140037  0.06494414 -0.009174639 -0.10090912  1.000000000
## failures    0.243665377 -0.23667996 -0.250408444  0.09223875 -0.173563031
## absences    0.175230079  0.10028482  0.024472887 -0.01294378 -0.062700175
## G1         -0.064081497  0.20534100  0.190269936 -0.09303999  0.160611915
## G2         -0.143474049  0.21552717  0.164893393 -0.15319796  0.135879999
## G3         -0.161579438  0.21714750  0.152456939 -0.11714205  0.097819690
##               failures    absences          G1         G2          G3
## age         0.24366538  0.17523008 -0.06408150 -0.1434740 -0.16157944
## Medu       -0.23667996  0.10028482  0.20534100  0.2155272  0.21714750
## Fedu       -0.25040844  0.02447289  0.19026994  0.1648934  0.15245694
## traveltime  0.09223875 -0.01294378 -0.09303999 -0.1531980 -0.11714205
## studytime  -0.17356303 -0.06270018  0.16061192  0.1358800  0.09781969
## failures    1.00000000  0.06372583 -0.35471761 -0.3558956 -0.36041494
## absences    0.06372583  1.00000000 -0.03100290 -0.0317767  0.03424732
## G1         -0.35471761 -0.03100290  1.00000000  0.8521181  0.80146793
## G2         -0.35589563 -0.03177670  0.85211807  1.0000000  0.90486799
## G3         -0.36041494  0.03424732  0.80146793  0.9048680  1.00000000
```

```
#library(corr)
#quant_vars %>% correlate() %>% network_plot(min_cor=0.2)
```

Creating variables for an ordinal final grade, 6-category final grade, and binary final grade

We'd like to examine final grades in multiple ways. The first is as a continuous numerical variable as G3 is.

The second is final grades as an ordered factor variable in order to perform multicategory ordinal logit modeling to see if we could improve fit and predictive power. However, this was unsuccessful.

```
data <- data %>%
  mutate(ord_g3 = factor(G3, ordered=T)
  )
```

The third is final grades as a 6-category ordered factor variable according to the Portuguese education system's classifications. We believe this could address some of the outliers and abnormality in the data (for example, many students received 0's, but no one received a 1, 2, or 3).

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
## The following object is masked from 'package:openintro':
##
##     densityPlot
```

```
data <- data %>%
  mutate(cat_g3 = case_when(
```

```
    G3 == 0 ~ "Poor",
    G3 <= 9 ~ "Weak",
    G3 <= 13 ~ "Sufficient",
    G3 <= 15 ~ "Good",
    G3 <= 17 ~"Very Good",
    G3 <= 20 ~ "Excellent"
  ))
data <- data %>%
  mutate(cat_g3 = factor(cat_g3, levels=c("Poor", "Weak", "Sufficient", "Good", "Very Good", "Excellent
```

The fourth is final grades as a binary factor variable. This is done based on the previous categories in the Portuguese classification system. If the student receives a "poor" or "weak" grade, or G3 < 10, this is considered a "low" grade. If the student received a "sufficient" "good" "very good" or "excellent" grade, this is a high grade.

```
data <- data %>%
  mutate(pf = case_when(
    G3 >= 10 ~ "high",
    G3 < 10 ~ "low"
  ))
data <- data %>%
  mutate(pf = factor(pf, levels=c("high", "low"), ordered = FALSE))
```

Splitting data into training and testing sets

```
attach(data)
set.seed(3)
train_ind <- sample(x = nrow(data), size = 0.8 * nrow(data))
test_ind_neg <- -train_ind
training <- data[train_ind, ]
testing <- data[test_ind_neg, ]
```

Linear model – note this is incorrect, should be on training set

```
base_lm <- lm(G3 ~ . -G2 -G1 -ord_g3 -cat_g3 -pf, data=training)
summary(base_lm)
```

```
##
## Call:
## lm(formula = G3 ~ . - G2 - G1 - ord_g3 - cat_g3 - pf, data = training)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.521  -2.044   0.376   2.489   9.741
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           18.203829   5.365987   3.392 0.000796 ***
## schoolMS               1.150014   0.867133   1.326 0.185878
## sexM                   1.322519   0.556764   2.375 0.018225 *
## age                   -0.484580   0.241157  -2.009 0.045483 *
## addressU               0.469861   0.656448   0.716 0.474752
## famsizeLE3             1.084061   0.538517   2.013 0.045095 *
```

```
## PstatusT               -0.938572   0.804552  -1.167 0.244403
## Medu                    0.109100   0.415204   0.263 0.792932
## Fedu                   -0.198417   0.350209  -0.567 0.571476
## Mjobhealth              0.867372   1.261872   0.687 0.492436
## Mjobother              -0.267653   0.802816  -0.333 0.739094
## Mjobservices            0.851853   0.890079   0.957 0.339390
## Mjobteacher            -1.408800   1.187437  -1.186 0.236491
## Fjobhealth             -0.610959   1.658086  -0.368 0.712808
## Fjobother              -0.391708   1.059598  -0.370 0.711911
## Fjobservices           -0.165284   1.115764  -0.148 0.882346
## Fjobteacher             0.801340   1.452904   0.552 0.581714
## reasonhome              0.449578   0.618690   0.727 0.468058
## reasonother             0.908713   0.889464   1.022 0.307859
## reasonreputation        1.042423   0.663933   1.570 0.117561
## guardianmother         -0.077074   0.625039  -0.123 0.901953
## guardianother           0.375031   1.092420   0.343 0.731635
## traveltime             -0.395034   0.395929  -0.998 0.319292
## studytime               0.605942   0.320779   1.889 0.059959 .
## failures               -0.718930   0.657655  -1.093 0.275285
## schoolsupyes           -1.484089   0.750939  -1.976 0.049130 *
## famsupyes              -0.465877   1.229810  -0.379 0.705117
## paidyes                 0.765679   0.536344   1.428 0.154557
## activitiesyes          -0.308944   0.491852  -0.628 0.530450
## nurseryyes             -0.363035   0.626732  -0.579 0.562899
## higheryes               0.008231   1.180199   0.007 0.994440
## internetyes             1.106572   1.016492   1.089 0.277286
## romanticyes            -1.349381   0.529977  -2.546 0.011445 *
## famrellow              -0.107383   0.600760  -0.179 0.858270
## freetimelow            -1.525774   0.537121  -2.841 0.004842 **
## gooutlow                1.589621   0.578545   2.748 0.006404 **
## Dalclow                 0.241061   1.376739   0.175 0.861135
## Walclow                 0.106693   0.763599   0.140 0.888981
## healthlow               0.626273   0.501042   1.250 0.212395
## absences                0.092750   0.043125   2.151 0.032380 *
## first_gen_collegeyes   -1.342709   0.900456  -1.491 0.137083
## stable_learning_envyes -0.901463   1.351184  -0.667 0.505232
## high_freq_absentyes    -0.563752   0.902594  -0.625 0.532763
## failedyes              -2.151741   1.233152  -1.745 0.082130 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.08 on 272 degrees of freedom
## Multiple R-squared:  0.3484, Adjusted R-squared:  0.2453
## F-statistic: 3.382 on 43 and 272 DF,  p-value: 6.586e-10
```

```
step.model <- stepAIC(base_lm, direction="both")
```

```
## Start:  AIC=929.2
## G3 ~ (school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + reason + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + G1 + G2 + first_gen_college +
##     stable_learning_env + high_freq_absent + failed + ord_g3 +
##     cat_g3 + pf) - G2 - G1 - ord_g3 - cat_g3 - pf
```

```
##
##                       Df Sum of Sq    RSS    AIC
## - Fjob                 4    22.451 4549.3 922.77
## - guardian             2     3.588 4530.4 925.45
## - reason               3    46.928 4573.8 926.46
## - higher               1     0.001 4526.8 927.20
## - Walc                 1     0.325 4527.2 927.23
## - Dalc                 1     0.510 4527.4 927.24
## - famrel               1     0.532 4527.4 927.24
## - Medu                 1     1.149 4528.0 927.28
## - famsup               1     2.388 4529.2 927.37
## - Fedu                 1     5.342 4532.2 927.58
## - nursery              1     5.584 4532.4 927.59
## - high_freq_absent     1     6.493 4533.3 927.66
## - activities           1     6.566 4533.4 927.66
## - stable_learning_env  1     7.408 4534.3 927.72
## - address              1     8.526 4535.4 927.80
## - traveltime           1    16.568 4543.4 928.36
## - internet             1    19.723 4546.6 928.58
## - failures             1    19.889 4546.7 928.59
## - Pstatus              1    22.649 4549.5 928.78
## - health               1    26.002 4552.8 929.01
## <none>                             4526.8 929.20
## - school               1    29.273 4556.1 929.24
## - paid                 1    33.918 4560.8 929.56
## - first_gen_college    1    37.005 4563.8 929.78
## - failed               1    50.673 4577.5 930.72
## - Mjob                 4   146.645 4673.5 931.28
## - studytime            1    59.385 4586.2 931.32
## - schoolsup            1    65.004 4591.8 931.71
## - age                  1    67.198 4594.0 931.86
## - famsize              1    67.443 4594.3 931.88
## - absences             1    76.982 4603.8 932.53
## - sex                  1    93.905 4620.7 933.69
## - romantic             1   107.890 4634.7 934.65
## - goout                1   125.644 4652.5 935.86
## - freetime             1   134.296 4661.1 936.44
##
## Step:  AIC=922.77
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + reason + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + first_gen_college + stable_learning_env +
##     high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - guardian             2     2.657 4552.0 918.95
## - reason               3    50.687 4600.0 920.27
## - Dalc                 1     0.015 4549.3 920.77
## - higher               1     0.064 4549.4 920.77
## - famrel               1     0.177 4549.5 920.78
## - famsup               1     1.619 4550.9 920.88
## - Medu                 1     1.665 4551.0 920.88
```

```
## - Fedu                 1     2.497 4551.8 920.94
## - Walc                 1     2.546 4551.8 920.94
## - nursery              1     6.318 4555.6 921.21
## - activities           1     6.794 4556.1 921.24
## - high_freq_absent     1     8.555 4557.8 921.36
## - address              1     9.586 4558.9 921.43
## - stable_learning_env  1    10.009 4559.3 921.46
## - traveltime           1    13.469 4562.8 921.70
## - failures             1    19.270 4568.6 922.10
## - internet             1    21.770 4571.1 922.28
## - Pstatus              1    25.698 4575.0 922.55
## - paid                 1    26.206 4575.5 922.58
## - health               1    26.464 4575.8 922.60
## <none>                             4549.3 922.77
## - school               1    29.812 4579.1 922.83
## - first_gen_college    1    40.522 4589.8 923.57
## - failed               1    48.982 4598.3 924.15
## - Mjob                 4   144.439 4693.7 924.64
## - studytime            1    59.801 4609.1 924.89
## - famsize              1    61.530 4610.8 925.01
## - schoolsup            1    62.702 4612.0 925.09
## - age                  1    64.436 4613.7 925.21
## - absences             1    83.540 4632.8 926.52
## - sex                  1    97.313 4646.6 927.46
## - romantic             1   105.515 4654.8 928.01
## - goout                1   121.823 4671.1 929.12
## + Fjob                 4    22.451 4526.8 929.20
## - freetime             1   138.896 4688.2 930.27
##
## Step:  AIC=918.95
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + reason + traveltime + studytime + failures +
##     schoolsup + famsup + paid + activities + nursery + higher +
##     internet + romantic + famrel + freetime + goout + Dalc +
##     Walc + health + absences + first_gen_college + stable_learning_env +
##     high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - reason               3    52.356 4604.3 916.57
## - Dalc                 1     0.012 4552.0 916.95
## - higher               1     0.181 4552.1 916.96
## - famrel               1     0.188 4552.1 916.96
## - Medu                 1     1.234 4553.2 917.04
## - famsup               1     1.629 4553.6 917.06
## - Fedu                 1     1.788 4553.7 917.08
## - Walc                 1     2.812 4554.8 917.15
## - activities           1     6.997 4558.9 917.44
## - nursery              1     8.052 4560.0 917.51
## - high_freq_absent     1     8.288 4560.2 917.53
## - stable_learning_env  1     9.808 4561.8 917.63
## - traveltime           1    11.700 4563.6 917.76
## - address              1    11.755 4563.7 917.77
## - failures             1    18.962 4570.9 918.27
## - internet             1    21.221 4573.2 918.42
```

```
## - paid                 1    26.936 4578.9 918.82
## - Pstatus              1    26.973 4578.9 918.82
## - school               1    28.739 4580.7 918.94
## - health               1    28.776 4580.7 918.94
## <none>                             4552.0 918.95
## - first_gen_college    1    40.067 4592.0 919.72
## - failed               1    47.162 4599.1 920.21
## - Mjob                 4   146.092 4698.0 920.93
## - famsize              1    61.085 4613.0 921.16
## - schoolsup            1    61.905 4613.9 921.22
## - studytime            1    62.135 4614.1 921.24
## - age                  1    65.132 4617.1 921.44
## - absences             1    83.207 4635.2 922.68
## + guardian             2     2.657 4549.3 922.77
## - sex                  1    97.826 4649.8 923.67
## - romantic             1   106.205 4658.2 924.24
## + Fjob                 4    21.520 4530.4 925.45
## - goout                1   131.437 4683.4 925.95
## - freetime             1   148.342 4700.3 927.09
##
## Step:  AIC=916.57
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + traveltime + studytime + failures + schoolsup +
##     famsup + paid + activities + nursery + higher + internet +
##     romantic + famrel + freetime + goout + Dalc + Walc + health +
##     absences + first_gen_college + stable_learning_env + high_freq_absent +
##     failed
##
##                       Df Sum of Sq    RSS    AIC
## - Dalc                 1     0.052 4604.4 914.57
## - higher               1     0.070 4604.4 914.57
## - famsup               1     0.559 4604.9 914.60
## - famrel               1     0.754 4605.1 914.62
## - Fedu                 1     2.686 4607.0 914.75
## - activities           1     3.055 4607.4 914.78
## - Walc                 1     3.533 4607.8 914.81
## - Medu                 1     3.838 4608.1 914.83
## - high_freq_absent     1     6.069 4610.4 914.98
## - nursery              1     6.181 4610.5 914.99
## - address              1     6.724 4611.0 915.03
## - stable_learning_env  1    13.913 4618.2 915.52
## - traveltime           1    13.959 4618.3 915.52
## - internet             1    24.904 4629.2 916.27
## - failures             1    25.319 4629.6 916.30
## - school               1    25.928 4630.2 916.34
## - Pstatus              1    26.882 4631.2 916.41
## <none>                             4604.3 916.57
## - first_gen_college    1    36.492 4640.8 917.06
## - paid                 1    36.508 4640.8 917.06
## - failed               1    39.346 4643.7 917.25
## - health               1    40.226 4644.5 917.31
## - schoolsup            1    61.327 4665.6 918.75
## - studytime            1    63.615 4667.9 918.90
## - age                  1    63.645 4668.0 918.90
```

```
## - famsize              1    63.740 4668.0 918.91
## + reason               3    52.356 4552.0 918.95
## - Mjob                 4   161.512 4765.8 919.46
## + guardian             2     4.326 4600.0 920.27
## - absences             1    87.084 4691.4 920.49
## - sex                  1    90.937 4695.2 920.75
## - romantic             1   100.812 4705.1 921.41
## + Fjob                 4    25.589 4578.7 922.80
## - goout                1   135.607 4739.9 923.74
## - freetime             1   146.195 4750.5 924.44
##
## Step:  AIC=914.57
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + traveltime + studytime + failures + schoolsup +
##     famsup + paid + activities + nursery + higher + internet +
##     romantic + famrel + freetime + goout + Walc + health + absences +
##     first_gen_college + stable_learning_env + high_freq_absent +
##     failed
##
##                       Df Sum of Sq    RSS    AIC
## - higher               1     0.072 4604.4 912.57
## - famsup               1     0.543 4604.9 912.61
## - famrel               1     0.775 4605.1 912.62
## - Fedu                 1     2.667 4607.0 912.75
## - activities           1     3.035 4607.4 912.78
## - Walc                 1     3.592 4608.0 912.82
## - Medu                 1     3.858 4608.2 912.83
## - high_freq_absent     1     6.057 4610.4 912.98
## - nursery              1     6.579 4610.9 913.02
## - address              1     6.837 4611.2 913.04
## - traveltime           1    13.907 4618.3 913.52
## - stable_learning_env  1    14.057 4618.4 913.53
## - internet             1    25.117 4629.5 914.29
## - failures             1    25.268 4629.6 914.30
## - school               1    26.243 4630.6 914.37
## - Pstatus              1    27.298 4631.7 914.44
## <none>                             4604.4 914.57
## - first_gen_college    1    36.450 4640.8 915.06
## - paid                 1    37.459 4641.8 915.13
## - failed               1    39.650 4644.0 915.28
## - health               1    40.189 4644.5 915.32
## + Dalc                 1     0.052 4604.3 916.57
## - schoolsup            1    61.468 4665.8 916.76
## - age                  1    63.625 4668.0 916.91
## - studytime            1    63.723 4668.1 916.91
## - famsize              1    63.834 4668.2 916.92
## + reason               3    52.397 4552.0 916.95
## - Mjob                 4   161.615 4766.0 917.47
## + guardian             2     4.340 4600.0 918.27
## - absences             1    87.486 4691.8 918.52
## - sex                  1    91.454 4695.8 918.78
## - romantic             1   101.206 4705.6 919.44
## + Fjob                 4    25.373 4579.0 920.82
## - goout                1   135.654 4740.0 921.74
```

```
## - freetime             1   147.561 4751.9 922.54
##
## Step:  AIC=912.57
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + traveltime + studytime + failures + schoolsup +
##     famsup + paid + activities + nursery + internet + romantic +
##     famrel + freetime + goout + Walc + health + absences + first_gen_college +
##     stable_learning_env + high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - famsup               1     0.568 4605.0 910.61
## - famrel               1     0.775 4605.2 910.63
## - Fedu                 1     2.608 4607.0 910.75
## - activities           1     2.997 4607.4 910.78
## - Walc                 1     3.640 4608.1 910.82
## - Medu                 1     3.842 4608.3 910.84
## - high_freq_absent     1     5.985 4610.4 910.98
## - nursery              1     6.603 4611.0 911.03
## - address              1     6.856 4611.3 911.04
## - stable_learning_env  1    13.985 4618.4 911.53
## - traveltime           1    14.027 4618.5 911.54
## - internet             1    25.076 4629.5 912.29
## - failures             1    26.533 4631.0 912.39
## - school               1    26.680 4631.1 912.40
## - Pstatus              1    27.355 4631.8 912.45
## <none>                             4604.4 912.57
## - first_gen_college    1    36.472 4640.9 913.07
## - paid                 1    38.301 4642.7 913.19
## - failed               1    39.622 4644.1 913.28
## - health               1    40.231 4644.7 913.32
## + higher               1     0.072 4604.4 914.57
## + Dalc                 1     0.054 4604.4 914.57
## - schoolsup            1    61.580 4666.0 914.77
## - famsize              1    63.825 4668.3 914.92
## + reason               3    52.291 4552.1 914.96
## - studytime            1    64.699 4669.1 914.98
## - age                  1    65.805 4670.2 915.06
## - Mjob                 4   161.555 4766.0 915.47
## + guardian             2     4.407 4600.0 916.27
## - absences             1    88.034 4692.5 916.56
## - sex                  1    92.455 4696.9 916.86
## - romantic             1   103.467 4707.9 917.60
## + Fjob                 4    25.392 4579.0 918.83
## - goout                1   135.631 4740.1 919.75
## - freetime             1   147.699 4752.1 920.55
##
## Step:  AIC=910.61
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + traveltime + studytime + failures + schoolsup +
##     paid + activities + nursery + internet + romantic + famrel +
##     freetime + goout + Walc + health + absences + first_gen_college +
##     stable_learning_env + high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
```

```
## - famrel               1     0.830 4605.8 908.67
## - Fedu                 1     2.797 4607.8 908.81
## - activities           1     2.876 4607.9 908.81
## - Walc                 1     3.666 4608.7 908.86
## - Medu                 1     3.812 4608.8 908.87
## - high_freq_absent     1     5.966 4611.0 909.02
## - address              1     6.534 4611.5 909.06
## - nursery              1     6.559 4611.6 909.06
## - traveltime           1    14.064 4619.1 909.58
## - school               1    27.105 4632.1 910.47
## - failures             1    27.727 4632.7 910.51
## - Pstatus              1    28.520 4633.5 910.56
## <none>                             4605.0 910.61
## - first_gen_college    1    36.695 4641.7 911.12
## - paid                 1    37.733 4642.7 911.19
## - failed               1    39.057 4644.1 911.28
## - health               1    40.099 4645.1 911.35
## - internet             1    51.288 4656.3 912.11
## + famsup               1     0.568 4604.4 912.57
## + higher               1     0.097 4604.9 912.61
## + Dalc                 1     0.038 4605.0 912.61
## - schoolsup            1    61.235 4666.2 912.79
## - famsize              1    64.114 4669.1 912.98
## - age                  1    65.241 4670.2 913.06
## + reason               3    51.133 4553.9 913.08
## - studytime            1    65.634 4670.6 913.09
## - Mjob                 4   161.786 4766.8 913.52
## + guardian             2     4.384 4600.6 914.31
## - absences             1    87.488 4692.5 914.56
## - sex                  1    95.379 4700.4 915.09
## - stable_learning_env  1    98.444 4703.4 915.30
## - romantic             1   103.086 4708.1 915.61
## + Fjob                 4    24.872 4580.1 916.90
## - goout                1   135.297 4740.3 917.76
## - freetime             1   147.293 4752.3 918.56
##
## Step:  AIC=908.67
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + traveltime + studytime + failures + schoolsup +
##     paid + activities + nursery + internet + romantic + freetime +
##     goout + Walc + health + absences + first_gen_college + stable_learning_env +
##     high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - Fedu                 1     2.624 4608.5 906.85
## - activities           1     2.893 4608.7 906.87
## - Medu                 1     3.936 4609.8 906.94
## - Walc                 1     4.401 4610.2 906.97
## - nursery              1     6.369 4612.2 907.11
## - high_freq_absent     1     6.397 4612.2 907.11
## - address              1     6.710 4612.5 907.13
## - traveltime           1    13.790 4619.6 907.61
## - school               1    26.389 4632.2 908.48
## - Pstatus              1    28.121 4633.9 908.59
```

```
## - failures             1    28.339 4634.2 908.61
## <none>                             4605.8 908.67
## - first_gen_college    1    36.610 4642.4 909.17
## - paid                 1    38.222 4644.1 909.28
## - failed               1    38.998 4644.8 909.33
## - health               1    39.269 4645.1 909.35
## - internet             1    51.344 4657.2 910.17
## + famrel               1     0.830 4605.0 910.61
## + famsup               1     0.623 4605.2 910.63
## + higher               1     0.098 4605.7 910.66
## + Dalc                 1     0.056 4605.8 910.67
## - schoolsup            1    61.018 4666.8 910.83
## - famsize              1    64.288 4670.1 911.05
## - age                  1    64.415 4670.2 911.06
## + reason               3    51.723 4554.1 911.10
## - studytime            1    65.938 4671.8 911.16
## - Mjob                 4   162.394 4768.2 911.62
## + guardian             2     4.337 4601.5 912.37
## - absences             1    88.472 4694.3 912.68
## - sex                  1    97.356 4703.2 913.28
## - stable_learning_env  1    97.881 4703.7 913.32
## - romantic             1   102.853 4708.7 913.65
## + Fjob                 4    24.261 4581.6 915.00
## - goout                1   134.730 4740.6 915.78
## - freetime             1   152.588 4758.4 916.97
##
## Step:  AIC=906.85
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Mjob + traveltime + studytime + failures + schoolsup + paid +
##     activities + nursery + internet + romantic + freetime + goout +
##     Walc + health + absences + first_gen_college + stable_learning_env +
##     high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - Medu                 1     2.786 4611.2 905.04
## - activities           1     3.168 4611.6 905.07
## - Walc                 1     4.633 4613.1 905.17
## - high_freq_absent     1     5.751 4614.2 905.24
## - nursery              1     6.964 4615.4 905.33
## - address              1     7.031 4615.5 905.33
## - traveltime           1    12.660 4621.1 905.72
## - school               1    25.585 4634.0 906.60
## - failures             1    26.390 4634.8 906.65
## - Pstatus              1    27.375 4635.8 906.72
## <none>                             4608.5 906.85
## - first_gen_college    1    35.702 4644.2 907.29
## - paid                 1    38.723 4647.2 907.49
## - health               1    39.562 4648.0 907.55
## - failed               1    40.756 4649.2 907.63
## - internet             1    52.513 4661.0 908.43
## + Fedu                 1     2.624 4605.8 908.67
## + famsup               1     0.806 4607.6 908.79
## + famrel               1     0.656 4607.8 908.81
## + Dalc                 1     0.032 4608.4 908.85
```

```
## + higher               1     0.028 4608.4 908.85
## - schoolsup            1    63.257 4671.7 909.16
## + reason               3    52.446 4556.0 909.23
## - famsize              1    65.203 4673.7 909.29
## - age                  1    65.395 4673.8 909.30
## - Mjob                 4   160.168 4768.6 909.65
## - studytime            1    70.570 4679.0 909.65
## + guardian             2     3.030 4605.4 910.64
## - absences             1    86.716 4695.2 910.74
## - sex                  1    96.713 4705.2 911.41
## - stable_learning_env  1   102.554 4711.0 911.81
## - romantic             1   103.215 4711.7 911.85
## + Fjob                 4    20.078 4588.4 913.47
## - goout                1   136.489 4744.9 914.07
## - freetime             1   157.762 4766.2 915.49
##
## Step:  AIC=905.04
## G3 ~ school + sex + age + address + famsize + Pstatus + Mjob +
##     traveltime + studytime + failures + schoolsup + paid + activities +
##     nursery + internet + romantic + freetime + goout + Walc +
##     health + absences + first_gen_college + stable_learning_env +
##     high_freq_absent + failed
##
##                       Df Sum of Sq    RSS    AIC
## - activities           1     3.365 4614.6 903.27
## - Walc                 1     4.818 4616.1 903.37
## - high_freq_absent     1     5.218 4616.5 903.40
## - nursery              1     6.626 4617.9 903.49
## - address              1     7.387 4618.6 903.55
## - traveltime           1    13.152 4624.4 903.94
## - school               1    24.848 4636.1 904.74
## - failures             1    26.223 4637.5 904.83
## - Pstatus              1    28.620 4639.9 905.00
## <none>                             4611.2 905.04
## - paid                 1    39.037 4650.3 905.70
## - health               1    42.629 4653.9 905.95
## - failed               1    42.685 4653.9 905.95
## - internet             1    52.013 4663.3 906.59
## + Medu                 1     2.786 4608.5 906.85
## + Fedu                 1     1.474 4609.8 906.94
## + famrel               1     0.791 4610.4 906.99
## + famsup               1     0.734 4610.5 906.99
## + Dalc                 1     0.053 4611.2 907.04
## + higher               1     0.030 4611.2 907.04
## + reason               3    54.539 4556.7 907.28
## - famsize              1    62.951 4674.2 907.33
## - schoolsup            1    64.412 4675.7 907.42
## - age                  1    65.820 4677.1 907.52
## - Mjob                 4   164.228 4775.5 908.10
## - studytime            1    74.553 4685.8 908.11
## - first_gen_college    1    79.652 4690.9 908.45
## + guardian             2     2.634 4608.6 908.86
## - absences             1    90.700 4701.9 909.20
## - sex                  1   100.420 4711.7 909.85
```

```
## - stable_learning_env  1   100.575 4711.8 909.86
## - romantic             1   102.155 4713.4 909.97
## + Fjob                 4    21.251 4590.0 911.58
## - goout                1   134.594 4745.8 912.13
## - freetime             1   156.601 4767.8 913.59
##
## Step:  AIC=903.27
## G3 ~ school + sex + age + address + famsize + Pstatus + Mjob +
##     traveltime + studytime + failures + schoolsup + paid + nursery +
##     internet + romantic + freetime + goout + Walc + health +
##     absences + first_gen_college + stable_learning_env + high_freq_absent +
##     failed
##
##                       Df Sum of Sq    RSS    AIC
## - Walc                 1     4.868 4619.5 901.60
## - high_freq_absent     1     4.880 4619.5 901.61
## - nursery              1     6.698 4621.3 901.73
## - address              1     8.135 4622.7 901.83
## - traveltime           1    12.891 4627.5 902.15
## - failures             1    25.150 4639.8 902.99
## - school               1    26.162 4640.8 903.06
## <none>                             4614.6 903.27
## - Pstatus              1    30.752 4645.4 903.37
## - paid                 1    40.143 4654.7 904.01
## - health               1    42.979 4657.6 904.20
## - failed               1    44.128 4658.7 904.28
## - internet             1    51.734 4666.3 904.79
## + activities           1     3.365 4611.2 905.04
## + Medu                 1     2.983 4611.6 905.07
## + Fedu                 1     1.653 4613.0 905.16
## + famrel               1     0.804 4613.8 905.22
## + famsup               1     0.593 4614.0 905.23
## + Dalc                 1     0.032 4614.6 905.27
## + higher               1     0.005 4614.6 905.27
## - famsize              1    63.284 4677.9 905.58
## - age                  1    64.243 4678.8 905.64
## - schoolsup            1    64.342 4678.9 905.65
## + reason               3    50.738 4563.9 905.78
## - studytime            1    72.147 4686.8 906.17
## - Mjob                 4   165.955 4780.6 906.44
## - first_gen_college    1    77.768 4692.4 906.55
## + guardian             2     2.630 4612.0 907.09
## - absences             1    89.997 4704.6 907.37
## - sex                  1    97.922 4712.5 907.91
## - stable_learning_env  1   100.748 4715.4 908.10
## - romantic             1   103.906 4718.5 908.31
## + Fjob                 4    21.318 4593.3 909.81
## - goout                1   136.309 4750.9 910.47
## - freetime             1   154.642 4769.2 911.69
##
## Step:  AIC=901.6
## G3 ~ school + sex + age + address + famsize + Pstatus + Mjob +
##     traveltime + studytime + failures + schoolsup + paid + nursery +
##     internet + romantic + freetime + goout + health + absences +
```

```
##     first_gen_college + stable_learning_env + high_freq_absent +
##     failed
##
##                        Df Sum of Sq    RSS    AIC
## - nursery               1     5.268 4624.7 899.96
## - high_freq_absent      1     6.309 4625.8 900.04
## - address               1     8.398 4627.9 900.18
## - traveltime            1    14.141 4633.6 900.57
## - school                1    26.891 4646.4 901.44
## - failures              1    26.947 4646.4 901.44
## <none>                               4619.5 901.60
## - Pstatus               1    29.540 4649.0 901.62
## - paid                  1    39.240 4658.7 902.28
## - failed                1    42.935 4662.4 902.53
## - health                1    46.283 4665.8 902.76
## - internet              1    51.566 4671.0 903.11
## + Walc                  1     4.868 4614.6 903.27
## + activities            1     3.415 4616.1 903.37
## + Medu                  1     3.176 4616.3 903.39
## + Fedu                  1     1.807 4617.7 903.48
## + famrel                1     1.570 4617.9 903.50
## + famsup                1     0.651 4618.8 903.56
## + Dalc                  1     0.215 4619.3 903.59
## + higher                1     0.025 4619.4 903.60
## - age                   1    61.719 4681.2 903.80
## - famsize               1    62.196 4681.7 903.83
## - schoolsup             1    63.775 4683.2 903.94
## + reason                3    51.505 4568.0 904.06
## - studytime             1    73.235 4692.7 904.58
## - first_gen_college     1    76.131 4695.6 904.77
## - Mjob                  4   167.753 4787.2 904.88
## + guardian              2     2.997 4616.5 905.40
## - absences              1    88.701 4708.2 905.61
## - sex                   1    93.167 4712.6 905.91
## - stable_learning_env   1    97.503 4717.0 906.21
## - romantic              1   103.781 4723.3 906.63
## + Fjob                  4    23.788 4595.7 907.97
## - freetime              1   159.147 4778.6 910.31
## - goout                 1   192.048 4811.5 912.48
##
## Step:  AIC=899.96
## G3 ~ school + sex + age + address + famsize + Pstatus + Mjob +
##     traveltime + studytime + failures + schoolsup + paid + internet +
##     romantic + freetime + goout + health + absences + first_gen_college +
##     stable_learning_env + high_freq_absent + failed
##
##                        Df Sum of Sq    RSS    AIC
## - high_freq_absent      1     6.963 4631.7 898.44
## - address               1     8.034 4632.8 898.51
## - traveltime            1    14.290 4639.0 898.94
## - failures              1    27.191 4651.9 899.82
## - Pstatus               1    28.623 4653.4 899.91
## - school                1    28.697 4653.4 899.92
## <none>                               4624.7 899.96
```

```
## - paid                 1    37.447 4662.2 900.51
## - failed               1    42.458 4667.2 900.85
## - health               1    46.355 4671.1 901.12
## - internet             1    52.921 4677.7 901.56
## + nursery              1     5.268 4619.5 901.60
## + activities           1     3.472 4621.3 901.73
## + Walc                 1     3.438 4621.3 901.73
## + Medu                 1     2.822 4621.9 901.77
## + Fedu                 1     2.266 4622.5 901.81
## + famrel               1     1.167 4623.6 901.89
## + famsup               1     0.613 4624.1 901.92
## + higher               1     0.027 4624.7 901.96
## + Dalc                 1     0.000 4624.7 901.96
## - famsize              1    59.246 4684.0 901.99
## - age                  1    60.691 4685.4 902.08
## - schoolsup            1    65.344 4690.1 902.40
## + reason               3    49.762 4575.0 902.55
## - studytime            1    70.102 4694.8 902.72
## - first_gen_college    1    72.970 4697.7 902.91
## - Mjob                 4   165.869 4790.6 903.10
## + guardian             2     4.296 4620.4 903.67
## - absences             1    90.215 4715.0 904.07
## - sex                  1    92.011 4716.8 904.19
## - stable_learning_env  1    95.337 4720.1 904.41
## - romantic             1   103.737 4728.5 904.97
## + Fjob                 4    23.943 4600.8 906.32
## - freetime             1   163.248 4788.0 908.93
## - goout                1   193.818 4818.6 910.94
##
## Step:  AIC=898.44
## G3 ~ school + sex + age + address + famsize + Pstatus + Mjob +
##     traveltime + studytime + failures + schoolsup + paid + internet +
##     romantic + freetime + goout + health + absences + first_gen_college +
##     stable_learning_env + failed
##
##                       Df Sum of Sq    RSS    AIC
## - address              1     7.266 4639.0 896.94
## - traveltime           1    12.535 4644.2 897.29
## - failures             1    26.837 4658.5 898.27
## - Pstatus              1    27.719 4659.4 898.33
## - school               1    28.778 4660.5 898.40
## <none>                             4631.7 898.44
## - paid                 1    37.157 4668.9 898.97
## - failed               1    44.182 4675.9 899.44
## - health               1    44.411 4676.1 899.46
## + high_freq_absent     1     6.963 4624.7 899.96
## - internet             1    52.830 4684.5 900.02
## + nursery              1     5.921 4625.8 900.04
## + Walc                 1     4.647 4627.1 900.12
## + activities           1     3.072 4628.6 900.23
## - famsize              1    56.687 4688.4 900.28
## + Medu                 1     2.201 4629.5 900.29
## + famrel               1     1.819 4629.9 900.32
## + Fedu                 1     1.779 4629.9 900.32
```

```
## + famsup               1     0.603 4631.1 900.40
## + Dalc                 1     0.227 4631.5 900.42
## + higher               1     0.010 4631.7 900.44
## - age                  1    62.584 4694.3 900.68
## - schoolsup            1    63.205 4694.9 900.72
## - first_gen_college    1    69.920 4701.6 901.17
## + reason               3    47.128 4584.6 901.21
## - studytime            1    73.006 4704.7 901.38
## - Mjob                 4   163.204 4794.9 901.38
## + guardian             2     4.080 4627.6 902.16
## - sex                  1    91.838 4723.5 902.64
## - stable_learning_env  1    98.774 4730.5 903.11
## - romantic             1   103.936 4735.6 903.45
## - absences             1   114.267 4746.0 904.14
## + Fjob                 4    25.543 4606.2 904.69
## - freetime             1   159.728 4791.4 907.15
## - goout                1   197.640 4829.3 909.64
##
## Step:  AIC=896.94
## G3 ~ school + sex + age + famsize + Pstatus + Mjob + traveltime +
##     studytime + failures + schoolsup + paid + internet + romantic +
##     freetime + goout + health + absences + first_gen_college +
##     stable_learning_env + failed
##
##                       Df Sum of Sq    RSS    AIC
## - traveltime           1    19.405 4658.4 896.25
## - school               1    24.034 4663.0 896.57
## - failures             1    27.449 4666.4 896.80
## - Pstatus              1    28.802 4667.8 896.89
## <none>                             4639.0 896.94
## - paid                 1    37.373 4676.3 897.47
## - failed               1    44.429 4683.4 897.95
## - health               1    47.008 4686.0 898.12
## + address              1     7.266 4631.7 898.44
## + high_freq_absent     1     6.195 4632.8 898.51
## + nursery              1     5.515 4633.5 898.56
## + Walc                 1     4.853 4634.1 898.60
## + activities           1     3.770 4635.2 898.68
## + Medu                 1     2.582 4636.4 898.76
## + famrel               1     2.077 4636.9 898.79
## + Fedu                 1     2.015 4637.0 898.80
## + famsup               1     0.256 4638.7 898.92
## + Dalc                 1     0.088 4638.9 898.93
## + higher               1     0.007 4639.0 898.94
## - famsize              1    59.107 4698.1 898.94
## - age                  1    62.979 4701.9 899.20
## - internet             1    63.691 4702.7 899.24
## - schoolsup            1    64.147 4703.1 899.28
## - first_gen_college    1    68.916 4707.9 899.60
## - studytime            1    69.102 4708.1 899.61
## + reason               3    42.757 4596.2 900.01
## - Mjob                 4   166.990 4806.0 900.11
## + guardian             2     5.880 4633.1 900.53
## - sex                  1    88.924 4727.9 900.94
```

```
## - romantic             1   103.242 4742.2 901.89
## - stable_learning_env  1   106.317 4745.3 902.10
## - absences             1   110.550 4749.5 902.38
## + Fjob                 4    26.656 4612.3 903.11
## - freetime             1   160.290 4799.3 905.67
## - goout                1   195.612 4834.6 907.99
##
## Step:  AIC=896.25
## G3 ~ school + sex + age + famsize + Pstatus + Mjob + studytime +
##     failures + schoolsup + paid + internet + romantic + freetime +
##     goout + health + absences + first_gen_college + stable_learning_env +
##     failed
##
##                       Df Sum of Sq    RSS    AIC
## - school               1    15.148 4673.5 895.28
## <none>                             4658.4 896.25
## - failures             1    29.802 4688.2 896.27
## - Pstatus              1    32.616 4691.0 896.46
## + traveltime           1    19.405 4639.0 896.94
## - paid                 1    41.404 4699.8 897.05
## - failed               1    42.976 4701.3 897.16
## + address              1    14.136 4644.2 897.29
## - health               1    48.064 4706.4 897.50
## + Walc                 1     6.191 4652.2 897.83
## + nursery              1     5.412 4653.0 897.89
## - famsize              1    55.299 4713.7 897.98
## + high_freq_absent     1     3.864 4654.5 897.99
## + activities           1     3.828 4654.5 897.99
## + Medu                 1     3.517 4654.9 898.02
## - age                  1    57.122 4715.5 898.11
## + famrel               1     1.760 4656.6 898.14
## + Fedu                 1     0.922 4657.5 898.19
## + Dalc                 1     0.199 4658.2 898.24
## + famsup               1     0.145 4658.2 898.24
## + higher               1     0.059 4658.3 898.25
## - schoolsup            1    60.937 4719.3 898.36
## - first_gen_college    1    68.126 4726.5 898.84
## - internet             1    69.706 4728.1 898.95
## - studytime            1    72.231 4730.6 899.12
## + reason               3    44.326 4614.0 899.23
## + guardian             2     3.635 4654.7 900.01
## - sex                  1    85.887 4744.3 900.03
## - Mjob                 4   184.698 4843.1 900.54
## - romantic             1   109.682 4768.1 901.61
## - absences             1   110.083 4768.5 901.64
## - stable_learning_env  1   115.969 4774.3 902.03
## + Fjob                 4    24.247 4634.1 902.61
## - freetime             1   167.854 4826.2 905.44
## - goout                1   206.149 4864.5 907.94
##
## Step:  AIC=895.28
## G3 ~ sex + age + famsize + Pstatus + Mjob + studytime + failures +
##     schoolsup + paid + internet + romantic + freetime + goout +
##     health + absences + first_gen_college + stable_learning_env +
```

```
##     failed
##
##                         Df Sum of Sq    RSS    AIC
## <none>                               4673.5 895.28
## - Pstatus                1    30.564 4704.1 895.34
## - failures               1    32.554 4706.1 895.47
## - failed                 1    40.951 4714.5 896.04
## - paid                   1    42.024 4715.5 896.11
## - age                    1    43.819 4717.3 896.23
## + school                 1    15.148 4658.4 896.25
## + traveltime             1    10.519 4663.0 896.57
## - health                 1    51.313 4724.8 896.73
## + nursery                1     6.947 4666.6 896.81
## + address                1     6.412 4667.1 896.85
## + Walc                   1     6.200 4667.3 896.86
## + activities             1     4.656 4668.9 896.97
## + high_freq_absent       1     4.592 4668.9 896.97
## + Medu                   1     2.462 4671.1 897.11
## - famsize                1    57.629 4731.1 897.15
## + Fedu                   1     0.815 4672.7 897.23
## + famrel                 1     0.763 4672.8 897.23
## + famsup                 1     0.420 4673.1 897.25
## + higher                 1     0.286 4673.2 897.26
## + Dalc                   1     0.037 4673.5 897.28
## - schoolsup              1    61.967 4735.5 897.44
## - studytime              1    66.222 4739.7 897.73
## - first_gen_college      1    67.388 4740.9 897.80
## - internet               1    68.381 4741.9 897.87
## + reason                 3    41.716 4631.8 898.45
## - sex                    1    82.718 4756.2 898.82
## + guardian               2     2.785 4670.7 899.09
## - Mjob                   4   180.840 4854.4 899.28
## - absences               1    99.398 4772.9 899.93
## - romantic               1   105.360 4778.9 900.33
## + Fjob                   4    27.128 4646.4 901.44
## - stable_learning_env    1   122.336 4795.9 901.45
## - freetime               1   173.037 4846.6 904.77
## - goout                  1   211.170 4884.7 907.25
```

```
summary(step.model)
```

```
##
## Call:
## lm(formula = G3 ~ sex + age + famsize + Pstatus + Mjob + studytime +
##     failures + schoolsup + paid + internet + romantic + freetime +
##     goout + health + absences + first_gen_college + stable_learning_env +
##     failed, data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8210  -2.0556   0.4938   2.6140   9.9229
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             14.75149    3.54759   4.158 4.21e-05 ***
```

```
## sexM                    1.14701    0.50283   2.281 0.023256 *
## age                    -0.32975    0.19861  -1.660 0.097924 .
## famsizeLE3              0.97273    0.51088   1.904 0.057884 .
## PstatusT               -1.04089    0.75067  -1.387 0.166610
## Mjobhealth              1.45938    1.05318   1.386 0.166893
## Mjobother               0.09317    0.70128   0.133 0.894392
## Mjobservices            1.24145    0.75471   1.645 0.101053
## Mjobteacher            -1.05269    1.01161  -1.041 0.298916
## studytime               0.60166    0.29478   2.041 0.042138 *
## failures               -0.87145    0.60896  -1.431 0.153478
## schoolsupyes           -1.41766    0.71803  -1.974 0.049274 *
## paidyes                 0.80462    0.49487   1.626 0.105038
## internetyes             1.51136    0.72870   2.074 0.038945 *
## romanticyes            -1.29428    0.50274  -2.574 0.010528 *
## freetimelow            -1.63431    0.49535  -3.299 0.001089 **
## gooutlow                1.81405    0.49772   3.645 0.000316 ***
## healthlow               0.83218    0.46318   1.797 0.073416 .
## absences                0.07111    0.02844   2.501 0.012944 *
## first_gen_collegeyes   -1.22804    0.59645  -2.059 0.040381 *
## stable_learning_envyes -1.54991    0.55870  -2.774 0.005889 **
## failedyes              -1.86537    1.16221  -1.605 0.109562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.987 on 294 degrees of freedom
## Multiple R-squared:  0.3272, Adjusted R-squared:  0.2792
## F-statistic:  6.81 on 21 and 294 DF,  p-value: 6.694e-16
```

Based on the backwards stepwise regression model, we can see that the variables sex, famsize, Mjob, studytime, schoolsup, romantic, freetime, goout, absences, first_gen_college, stable_learning_environment failed are active. Multiple R-squared: 0.3105, Adjusted R-squared: 0.2687

Based on these active variables, some interactions that we think could be significant are: schoolsup*failed, famsup*first_gen_college, higher*first_gen_college.

```
activelm <- lm(G3 ~ (sex + famsize + Mjob + studytime + schoolsup + romantic + freetime + goout + absen

summary(activelm)
```

```
##
## Call:
## lm(formula = G3 ~ (sex + famsize + Mjob + studytime + schoolsup +
##     romantic + freetime + goout + absences + first_gen_college +
##     stable_learning_env)^2, data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7310  -1.8779   0.1507   2.2384   7.7562
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             16.978563   4.945398   3.433
## sexM                                    -6.279275   2.863416  -2.193
## famsizeLE3                               7.139516   3.122936   2.286
## Mjobhealth                              -3.697491   5.992618  -0.617
## Mjobother                               -1.180328   3.939477  -0.300
```

```
## Mjobservices                          -1.673599   4.163254  -0.402
## Mjobteacher                            3.298140   5.515226   0.598
## studytime                             -3.263700   1.649762  -1.978
## schoolsupyes                           1.459600   3.921646   0.372
## romanticyes                           -9.857215   3.690357  -2.671
## freetimelow                           -0.687540   3.089813  -0.223
## gooutlow                               1.605623   3.093781   0.519
## absences                              -0.139293   0.252030  -0.553
## first_gen_collegeyes                  -7.680166   3.425045  -2.242
## stable_learning_envyes                 1.373828   2.834812   0.485
## sexM:famsizeLE3                       -2.907723   1.417452  -2.051
## sexM:Mjobhealth                        8.135787   3.200323   2.542
## sexM:Mjobother                         1.559082   1.870911   0.833
## sexM:Mjobservices                      3.591237   2.047598   1.754
## sexM:Mjobteacher                       1.581279   2.687976   0.588
## sexM:studytime                         0.522376   0.716722   0.729
## sexM:schoolsupyes                     -3.227237   2.039394  -1.582
## sexM:romanticyes                       2.314159   1.314770   1.760
## sexM:freetimelow                       0.744900   1.228326   0.606
## sexM:gooutlow                          1.691006   1.289970   1.311
## sexM:absences                          0.092127   0.129626   0.711
## sexM:first_gen_collegeyes              4.088623   1.517734   2.694
## sexM:stable_learning_envyes            0.351908   1.291994   0.272
## famsizeLE3:Mjobhealth                 -6.315772   3.015589  -2.094
## famsizeLE3:Mjobother                  -3.273923   1.909237  -1.715
## famsizeLE3:Mjobservices               -2.393133   2.004824  -1.194
## famsizeLE3:Mjobteacher                -3.363680   2.974405  -1.131
## famsizeLE3:studytime                   0.232114   0.902695   0.257
## famsizeLE3:schoolsupyes               -3.350568   2.264998  -1.479
## famsizeLE3:romanticyes                -1.529419   1.423288  -1.075
## famsizeLE3:freetimelow                -0.403799   1.362434  -0.296
## famsizeLE3:gooutlow                    0.206195   1.381607   0.149
## famsizeLE3:absences                   -0.250150   0.087877  -2.847
## famsizeLE3:first_gen_collegeyes       -1.960437   1.661981  -1.180
## famsizeLE3:stable_learning_envyes      2.453701   1.308341   1.875
## Mjobhealth:studytime                   3.030828   2.022031   1.499
## Mjobother:studytime                    1.729165   1.129140   1.531
## Mjobservices:studytime                 2.290233   1.225339   1.869
## Mjobteacher:studytime                  1.490936   1.880559   0.793
## Mjobhealth:schoolsupyes                5.013922   6.571863   0.763
## Mjobother:schoolsupyes                 0.012477   2.327656   0.005
## Mjobservices:schoolsupyes              2.247615   2.664760   0.843
## Mjobteacher:schoolsupyes               6.255315   4.541613   1.377
## Mjobhealth:romanticyes                10.227248   3.469559   2.948
## Mjobother:romanticyes                  2.833341   1.889034   1.500
## Mjobservices:romanticyes               3.740122   2.088750   1.791
## Mjobteacher:romanticyes                6.314009   2.776216   2.274
## Mjobhealth:freetimelow                -0.922184   2.849252  -0.324
## Mjobother:freetimelow                  0.516055   1.929145   0.268
## Mjobservices:freetimelow              -0.819100   2.032401  -0.403
## Mjobteacher:freetimelow               -2.795190   2.675217  -1.045
## Mjobhealth:gooutlow                   -2.534428   2.922361  -0.867
## Mjobother:gooutlow                    -1.962467   1.881660  -1.043
## Mjobservices:gooutlow                 -0.823569   2.030101  -0.406
```

```
## Mjobteacher:gooutlow                          -0.478557   2.816652  -0.170
## Mjobhealth:absences                           -0.102298   0.301855  -0.339
## Mjobother:absences                            -0.069680   0.164307  -0.424
## Mjobservices:absences                         -0.003119   0.178437  -0.017
## Mjobteacher:absences                          -0.193009   0.245773  -0.785
## Mjobhealth:first_gen_collegeyes               -0.073575   3.326435  -0.022
## Mjobother:first_gen_collegeyes                 0.101772   2.200334   0.046
## Mjobservices:first_gen_collegeyes             -1.271246   2.336154  -0.544
## Mjobteacher:first_gen_collegeyes               9.759505   6.511814   1.499
## Mjobhealth:stable_learning_envyes             -5.566339   3.061283  -1.818
## Mjobother:stable_learning_envyes              -3.127598   1.858491  -1.683
## Mjobservices:stable_learning_envyes           -4.792954   1.971321  -2.431
## Mjobteacher:stable_learning_envyes            -9.459456   2.893356  -3.269
## studytime:schoolsupyes                        -2.587018   1.169829  -2.211
## studytime:romanticyes                          1.323596   0.925304   1.430
## studytime:freetimelow                          0.708880   0.817306   0.867
## studytime:gooutlow                             0.019852   0.846497   0.023
## studytime:absences                            -0.011006   0.066291  -0.166
## studytime:first_gen_collegeyes                 1.925774   1.106861   1.740
## studytime:stable_learning_envyes               1.014599   0.737860   1.375
## schoolsupyes:romanticyes                       4.297657   2.466701   1.742
## schoolsupyes:freetimelow                       1.142835   2.032938   0.562
## schoolsupyes:gooutlow                         -0.251228   2.226135  -0.113
## schoolsupyes:absences                         -0.240036   0.122725  -1.956
## schoolsupyes:first_gen_collegeyes              6.682693   2.275772   2.936
## schoolsupyes:stable_learning_envyes           -1.286722   1.931616  -0.666
## romanticyes:freetimelow                       -1.694420   1.306745  -1.297
## romanticyes:gooutlow                           1.544419   1.340978   1.152
## romanticyes:absences                           0.072144   0.094856   0.761
## romanticyes:first_gen_collegeyes               2.271525   1.593368   1.426
## romanticyes:stable_learning_envyes            -1.107765   1.277031  -0.867
## freetimelow:gooutlow                          -0.913480   1.244635  -0.734
## freetimelow:absences                           0.145371   0.109207   1.331
## freetimelow:first_gen_collegeyes              -0.797808   1.528367  -0.522
## freetimelow:stable_learning_envyes            -0.985060   1.319413  -0.747
## gooutlow:absences                             -0.035154   0.102970  -0.341
## gooutlow:first_gen_collegeyes                  1.678947   1.588931   1.057
## gooutlow:stable_learning_envyes                1.147120   1.299395   0.883
## absences:first_gen_collegeyes                  0.115368   0.116775   0.988
## absences:stable_learning_envyes                0.234354   0.078485   2.986
## first_gen_collegeyes:stable_learning_envyes -3.030900   1.563712  -1.938
##                                               Pr(>|t|)
## (Intercept)                                   0.000715 ***
## sexM                                          0.029380 *
## famsizeLE3                                    0.023215 *
## Mjobhealth                                    0.537879
## Mjobother                                     0.764759
## Mjobservices                                  0.688087
## Mjobteacher                                   0.550462
## studytime                                     0.049167 *
## schoolsupyes                                  0.710116
## romanticyes                                   0.008137 **
## freetimelow                                   0.824121
## gooutlow                                      0.604303
```

```
## absences                                  0.581052
## first_gen_collegeyes                      0.025954 *
## stable_learning_envyes                    0.628432
## sexM:famsizeLE3                           0.041437 *
## sexM:Mjobhealth                           0.011717 *
## sexM:Mjobother                            0.405581
## sexM:Mjobservices                         0.080870 .
## sexM:Mjobteacher                          0.556960
## sexM:studytime                            0.466889
## sexM:schoolsupyes                         0.115010
## sexM:romanticyes                          0.079802 .
## sexM:freetimelow                          0.544863
## sexM:gooutlow                             0.191288
## sexM:absences                             0.478027
## sexM:first_gen_collegeyes                 0.007617 **
## sexM:stable_learning_envyes               0.785593
## famsizeLE3:Mjobhealth                     0.037393 *
## famsizeLE3:Mjobother                      0.087820 .
## famsizeLE3:Mjobservices                   0.233910
## famsizeLE3:Mjobteacher                    0.259362
## famsizeLE3:studytime                      0.797320
## famsizeLE3:schoolsupyes                   0.140522
## famsizeLE3:romanticyes                    0.283767
## famsizeLE3:freetimelow                    0.767224
## famsizeLE3:gooutlow                       0.881501
## famsizeLE3:absences                       0.004845 **
## famsizeLE3:first_gen_collegeyes           0.239465
## famsizeLE3:stable_learning_envyes         0.062083 .
## Mjobhealth:studytime                      0.135359
## Mjobother:studytime                       0.127134
## Mjobservices:studytime                    0.062968 .
## Mjobteacher:studytime                     0.428755
## Mjobhealth:schoolsupyes                   0.446333
## Mjobother:schoolsupyes                    0.995728
## Mjobservices:schoolsupyes                 0.399905
## Mjobteacher:schoolsupyes                  0.169835
## Mjobhealth:romanticyes                    0.003553 **
## Mjobother:romanticyes                     0.135103
## Mjobservices:romanticyes                  0.074757 .
## Mjobteacher:romanticyes                   0.023929 *
## Mjobhealth:freetimelow                    0.746510
## Mjobother:freetimelow                     0.789336
## Mjobservices:freetimelow                  0.687332
## Mjobteacher:freetimelow                   0.297262
## Mjobhealth:gooutlow                       0.386765
## Mjobother:gooutlow                        0.298140
## Mjobservices:gooutlow                     0.685380
## Mjobteacher:gooutlow                      0.865246
## Mjobhealth:absences                       0.735017
## Mjobother:absences                        0.671927
## Mjobservices:absences                     0.986071
## Mjobteacher:absences                      0.433131
## Mjobhealth:first_gen_collegeyes           0.982374
## Mjobother:first_gen_collegeyes            0.963151
```

```
## Mjobservices:first_gen_collegeyes           0.586891
## Mjobteacher:first_gen_collegeyes            0.135402
## Mjobhealth:stable_learning_envyes           0.070403 .
## Mjobother:stable_learning_envyes            0.093845 .
## Mjobservices:stable_learning_envyes         0.015858 *
## Mjobteacher:stable_learning_envyes          0.001254 **
## studytime:schoolsupyes                      0.028053 *
## studytime:romanticyes                       0.154034
## studytime:freetimelow                       0.386720
## studytime:gooutlow                          0.981312
## studytime:absences                          0.868291
## studytime:first_gen_collegeyes              0.083309 .
## studytime:stable_learning_envyes            0.170539
## schoolsupyes:romanticyes                    0.082885 .
## schoolsupyes:freetimelow                    0.574590
## schoolsupyes:gooutlow                       0.910251
## schoolsupyes:absences                       0.051767 .
## schoolsupyes:first_gen_collegeyes           0.003679 **
## schoolsupyes:stable_learning_envyes         0.506034
## romanticyes:freetimelow                     0.196127
## romanticyes:gooutlow                        0.250713
## romanticyes:absences                        0.447749
## romanticyes:first_gen_collegeyes            0.155424
## romanticyes:stable_learning_envyes          0.386656
## freetimelow:gooutlow                        0.463785
## freetimelow:absences                        0.184546
## freetimelow:first_gen_collegeyes            0.602205
## freetimelow:stable_learning_envyes          0.456123
## gooutlow:absences                           0.733133
## gooutlow:first_gen_collegeyes               0.291851
## gooutlow:stable_learning_envyes             0.378320
## absences:first_gen_collegeyes               0.324285
## absences:stable_learning_envyes             0.003153 **
## first_gen_collegeyes:stable_learning_envyes 0.053893 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.143 on 216 degrees of freedom
## Multiple R-squared:  0.4664, Adjusted R-squared:  0.2219
## F-statistic: 1.907 on 99 and 216 DF,  p-value: 4.802e-05
```

From this, we can see that there seem to be significant interaction effects between sex and schoolsup, sex and first_gen_college, sex and Mjob, schoolsup and absences, schoolsup and studytime, schoolsup and first_gen_college, absences and failed, failed and first_gen_college, Mjob and studytime, Mjob and first_gen_college, studytime and famsup.

Linear model on the training data set

```
base_lm_tr <- lm(G3 ~ . -G2 -G1 -ord_g3 -stable_learning_env, data = training)
step.model_tr <- stepAIC(base_lm_tr, direction="both")
```

```
## Start:  AIC=25.96
## G3 ~ (school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + reason + guardian + traveltime + studytime +
```

```
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + G1 + G2 + first_gen_college +
##     stable_learning_env + high_freq_absent + failed + ord_g3 +
##     cat_g3 + pf) - G2 - G1 - ord_g3 - stable_learning_env
##
##
## Step:  AIC=25.96
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + reason + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + first_gen_college + high_freq_absent +
##     failed + cat_g3
##
##                      Df Sum of Sq    RSS    AIC
## - reason              3       1.2  254.3  21.41
## - guardian            2       1.6  254.8  23.97
## - activities          1       0.0  253.2  23.97
## - internet            1       0.0  253.2  23.98
## - first_gen_college   1       0.0  253.2  23.98
## - age                 1       0.0  253.2  24.00
## - absences            1       0.1  253.2  24.03
## - Fjob                4       5.0  258.2  24.16
## - freetime            1       0.2  253.3  24.17
## - school              1       0.2  253.4  24.18
## - Pstatus             1       0.4  253.6  24.44
## - nursery             1       0.5  253.7  24.56
## - romantic            1       0.6  253.8  24.73
## - Mjob                4       5.5  258.7  24.79
## - traveltime          1       0.7  253.9  24.86
## - high_freq_absent    1       0.7  253.9  24.88
## - sex                 1       0.8  253.9  24.92
## - goout               1       0.9  254.1  25.05
## - paid                1       0.9  254.1  25.13
## - health              1       0.9  254.1  25.13
## - famsize             1       1.0  254.2  25.21
## - Walc                1       1.4  254.5  25.65
## - address             1       1.4  254.6  25.68
## - schoolsup           1       1.6  254.8  25.96
## <none>                            253.2  25.96
## - famsup              1       2.0  255.1  26.39
## - Medu                1       2.0  255.1  26.40
## - failed              1       2.1  255.3  26.58
## - studytime           1       2.1  255.3  26.61
## - higher              1       2.3  255.5  26.87
## - famrel              1       2.8  256.0  27.44
## - Fedu                1       3.5  256.7  28.36
## - Dalc                1       3.9  257.1  28.82
## - failures            1       4.0  257.2  28.95
## - cat_g3              5    4281.1 4534.3 927.72
##
## Step:  AIC=21.41
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
```

```
##     Fedu + Mjob + Fjob + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + first_gen_college + high_freq_absent +
##     failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - guardian           2       1.6  255.9  19.37
## - first_gen_college  1       0.0  254.3  19.41
## - internet           1       0.0  254.4  19.43
## - age                1       0.0  254.4  19.44
## - activities         1       0.0  254.4  19.48
## - absences           1       0.1  254.4  19.53
## - school             1       0.1  254.5  19.57
## - freetime           1       0.2  254.6  19.69
## - Fjob               4       5.2  259.6  19.86
## - nursery            1       0.4  254.7  19.86
## - Mjob               4       5.3  259.6  19.88
## - Pstatus            1       0.4  254.8  19.92
## - romantic           1       0.5  254.9  20.06
## - high_freq_absent   1       0.7  255.0  20.23
## - paid               1       0.7  255.1  20.31
## - traveltime         1       0.9  255.2  20.47
## - sex                1       0.9  255.2  20.50
## - goout              1       0.9  255.3  20.55
## - famsize            1       1.0  255.3  20.61
## - health             1       1.2  255.6  20.91
## - Walc               1       1.3  255.7  21.06
## <none>                           254.3  21.41
## - schoolsup          1       1.6  256.0  21.45
## - address            1       1.7  256.0  21.48
## - failed             1       1.9  256.2  21.74
## - famsup             1       2.1  256.4  21.98
## - studytime          1       2.2  256.6  22.15
## - Medu               1       2.3  256.7  22.30
## - higher             1       2.5  256.8  22.47
## - famrel             1       2.9  257.3  23.03
## - Fedu               1       3.7  258.1  24.01
## - failures           1       3.8  258.1  24.08
## - Dalc               1       4.0  258.3  24.32
## + reason             3       1.2  253.2  25.96
## - cat_g3             5    4330.2 4584.6 925.21
##
## Step:  AIC=19.37
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + activities + nursery + higher +
##     internet + romantic + famrel + freetime + goout + Dalc +
##     Walc + health + absences + first_gen_college + high_freq_absent +
##     failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - internet           1       0.0  255.9  17.38
## - first_gen_college  1       0.0  255.9  17.38
```

```
## - activities         1       0.0  256.0  17.42
## - absences           1       0.1  256.0  17.45
## - nursery            1       0.2  256.1  17.58
## - school             1       0.2  256.2  17.66
## - age                1       0.3  256.2  17.77
## - Fjob               4       5.3  261.2  17.81
## - freetime           1       0.4  256.3  17.87
## - Mjob               4       5.6  261.5  18.19
## - romantic           1       0.7  256.6  18.20
## - Pstatus            1       0.7  256.7  18.27
## - high_freq_absent   1       0.8  256.7  18.35
## - sex                1       0.8  256.8  18.40
## - goout              1       0.9  256.8  18.47
## - famsize            1       0.9  256.8  18.48
## - paid               1       0.9  256.9  18.54
## - health             1       1.0  256.9  18.58
## - traveltime         1       1.1  257.0  18.75
## - schoolsup          1       1.5  257.4  19.24
## <none>                           255.9  19.37
## - Walc               1       1.7  257.7  19.52
## - address            1       2.0  257.9  19.80
## - higher             1       2.1  258.0  19.93
## - famsup             1       2.1  258.1  20.00
## - failed             1       2.2  258.1  20.05
## - studytime          1       2.3  258.2  20.19
## - Medu               1       2.5  258.4  20.39
## - famrel             1       2.5  258.5  20.50
## + guardian           2       1.6  254.3  21.41
## - failures           1       3.5  259.4  21.67
## - Dalc               1       4.0  259.9  22.27
## - Fedu               1       4.0  259.9  22.27
## + reason             3       1.1  254.8  23.97
## - cat_g3             5    4333.4 4589.4 921.54
##
## Step:  AIC=17.38
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + activities + nursery + higher +
##     romantic + famrel + freetime + goout + Dalc + Walc + health +
##     absences + first_gen_college + high_freq_absent + failed +
##     cat_g3
##
##                      Df Sum of Sq    RSS    AIC
## - first_gen_college  1       0.0  255.9  15.38
## - activities         1       0.0  256.0  15.42
## - absences           1       0.1  256.0  15.45
## - nursery            1       0.2  256.1  15.58
## - school             1       0.2  256.2  15.66
## - age                1       0.3  256.2  15.77
## - freetime           1       0.4  256.3  15.87
## - Fjob               4       5.4  261.3  15.93
## - romantic           1       0.7  256.6  16.22
## - Pstatus            1       0.7  256.7  16.28
## - Mjob               4       5.7  261.6  16.33
```

```
## - high_freq_absent   1       0.8  256.7  16.35
## - sex                1       0.8  256.8  16.40
## - goout              1       0.9  256.8  16.47
## - famsize            1       0.9  256.8  16.49
## - paid               1       1.0  256.9  16.56
## - health             1       1.0  256.9  16.58
## - traveltime         1       1.1  257.1  16.76
## - schoolsup          1       1.5  257.4  17.24
## <none>                            255.9  17.38
## - Walc               1       1.8  257.7  17.53
## - address            1       2.0  258.0  17.89
## - higher             1       2.1  258.0  17.95
## - famsup             1       2.1  258.1  18.00
## - failed             1       2.2  258.1  18.05
## - studytime          1       2.3  258.2  18.21
## - Medu               1       2.5  258.4  18.39
## - famrel             1       2.5  258.5  18.50
## + internet           1       0.0  255.9  19.37
## + guardian           2       1.6  254.4  19.43
## - failures           1       3.5  259.4  19.67
## - Dalc               1       4.0  259.9  20.28
## - Fedu               1       4.0  259.9  20.28
## + reason             3       1.1  254.8  21.97
## - cat_g3             5    4346.0 4601.9 920.40
##
## Step:  AIC=15.38
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + activities + nursery + higher +
##     romantic + famrel + freetime + goout + Dalc + Walc + health +
##     absences + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - activities         1       0.0  256.0  13.42
## - absences           1       0.1  256.0  13.45
## - nursery            1       0.2  256.1  13.59
## - school             1       0.2  256.2  13.67
## - age                1       0.3  256.3  13.77
## - freetime           1       0.4  256.3  13.87
## - Fjob               4       5.4  261.3  13.98
## - romantic           1       0.7  256.6  14.22
## - Pstatus            1       0.7  256.7  14.28
## - high_freq_absent   1       0.8  256.7  14.35
## - Mjob               4       5.7  261.7  14.40
## - sex                1       0.8  256.8  14.42
## - goout              1       0.9  256.8  14.48
## - famsize            1       0.9  256.8  14.49
## - paid               1       1.0  256.9  14.56
## - health             1       1.0  256.9  14.58
## - traveltime         1       1.1  257.1  14.76
## - schoolsup          1       1.5  257.4  15.25
## <none>                            255.9  15.38
## - Walc               1       1.8  257.7  15.55
## - address            1       2.0  258.0  15.89
```

```
## - higher             1       2.1  258.0  15.95
## - famsup             1       2.1  258.1  16.01
## - failed             1       2.2  258.1  16.06
## - studytime          1       2.3  258.2  16.21
## - famrel             1       2.5  258.5  16.51
## - Medu               1       3.0  258.9  17.05
## + first_gen_college  1       0.0  255.9  17.38
## + internet           1       0.0  255.9  17.38
## + guardian           2       1.6  254.4  17.43
## - failures           1       3.5  259.4  17.69
## - Dalc               1       4.0  259.9  18.28
## - Fedu               1       4.8  260.8  19.29
## + reason             3       1.1  254.8  20.00
## - cat_g3             5    4380.2 4636.1 920.74
##
## Step:  AIC=13.42
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + nursery + higher + romantic +
##     famrel + freetime + goout + Dalc + Walc + health + absences +
##     high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - absences           1       0.1  256.0  11.50
## - nursery            1       0.2  256.1  11.63
## - school             1       0.2  256.2  11.69
## - age                1       0.3  256.3  11.84
## - freetime           1       0.4  256.3  11.89
## - Fjob               4       5.4  261.3  11.98
## - romantic           1       0.7  256.6  12.25
## - Pstatus            1       0.8  256.7  12.36
## - high_freq_absent   1       0.8  256.8  12.41
## - Mjob               4       5.7  261.7  12.42
## - sex                1       0.8  256.8  12.43
## - famsize            1       0.9  256.9  12.53
## - goout              1       0.9  256.9  12.54
## - health             1       1.0  256.9  12.61
## - paid               1       1.0  256.9  12.62
## - traveltime         1       1.1  257.1  12.81
## - schoolsup          1       1.5  257.5  13.30
## <none>                           256.0  13.42
## - Walc               1       1.8  257.7  13.59
## - address            1       2.1  258.1  14.00
## - higher             1       2.1  258.1  14.04
## - famsup             1       2.1  258.1  14.04
## - failed             1       2.2  258.1  14.08
## - studytime          1       2.3  258.2  14.21
## - famrel             1       2.5  258.5  14.54
## - Medu               1       3.0  258.9  15.07
## + activities         1       0.0  255.9  15.38
## + first_gen_college  1       0.0  256.0  15.42
## + internet           1       0.0  256.0  15.42
## + guardian           2       1.6  254.4  15.49
## - failures           1       3.5  259.5  15.71
```

```
## - Dalc               1       4.0  260.0  16.30
## - Fedu               1       4.8  260.8  17.29
## + reason             3       1.1  254.8  18.00
## - cat_g3             5    4382.2 4638.2 918.88
##
## Step:  AIC=11.5
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + nursery + higher + romantic +
##     famrel + freetime + goout + Dalc + Walc + health + high_freq_absent +
##     failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - nursery            1       0.2  256.2   9.72
## - school             1       0.2  256.2   9.74
## - age                1       0.3  256.3   9.88
## - Fjob               4       5.3  261.3   9.99
## - freetime           1       0.4  256.4  10.00
## - romantic           1       0.6  256.7  10.28
## - Pstatus            1       0.7  256.8  10.41
## - Mjob               4       5.7  261.7  10.45
## - goout              1       0.9  256.9  10.59
## - sex                1       0.9  256.9  10.59
## - famsize            1       0.9  256.9  10.63
## - high_freq_absent   1       1.0  257.0  10.67
## - health             1       1.0  257.0  10.68
## - paid               1       1.0  257.0  10.69
## - traveltime         1       1.1  257.1  10.85
## - schoolsup          1       1.6  257.6  11.42
## <none>                           256.0  11.50
## - Walc               1       1.8  257.8  11.72
## - higher             1       2.1  258.1  12.05
## - failed             1       2.1  258.1  12.09
## - famsup             1       2.1  258.2  12.12
## - address            1       2.2  258.2  12.16
## - studytime          1       2.3  258.3  12.33
## - famrel             1       2.6  258.6  12.66
## - Medu               1       3.0  259.0  13.17
## + absences           1       0.1  256.0  13.42
## + activities         1       0.0  256.0  13.45
## + internet           1       0.0  256.0  13.50
## + first_gen_college  1       0.0  256.0  13.50
## + guardian           2       1.5  254.5  13.61
## - failures           1       3.4  259.5  13.71
## - Dalc               1       4.1  260.2  14.57
## - Fedu               1       4.8  260.8  15.33
## + reason             3       1.2  254.8  16.04
## - cat_g3             5    4459.9 4715.9 922.14
##
## Step:  AIC=9.72
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + traveltime + studytime + failures +
##     schoolsup + famsup + paid + higher + romantic + famrel +
##     freetime + goout + Dalc + Walc + health + high_freq_absent +
```

```
##     failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - school             1       0.2  256.4   7.99
## - age                1       0.3  256.5   8.07
## - freetime           1       0.4  256.6   8.19
## - Fjob               4       5.5  261.7   8.41
## - romantic           1       0.6  256.8   8.50
## - Pstatus            1       0.8  257.0   8.68
## - goout              1       0.8  257.1   8.76
## - Mjob               4       5.9  262.1   8.87
## - sex                1       1.0  257.2   8.90
## - health             1       1.0  257.2   8.92
## - famsize            1       1.0  257.2   8.97
## - high_freq_absent   1       1.0  257.2   8.98
## - paid               1       1.1  257.3   9.02
## - traveltime         1       1.1  257.4   9.13
## - schoolsup          1       1.5  257.7   9.60
## <none>                           256.2   9.72
## - Walc               1       1.9  258.1  10.04
## - failed             1       2.1  258.3  10.24
## - higher             1       2.1  258.3  10.28
## - address            1       2.2  258.4  10.43
## - famsup             1       2.3  258.5  10.50
## - studytime          1       2.5  258.7  10.74
## - famrel             1       2.5  258.7  10.80
## - Medu               1       3.0  259.2  11.34
## + nursery            1       0.2  256.0  11.50
## + absences           1       0.1  256.1  11.63
## + activities         1       0.0  256.2  11.68
## + first_gen_college  1       0.0  256.2  11.72
## + internet           1       0.0  256.2  11.72
## - failures           1       3.4  259.6  11.86
## + guardian           2       1.3  254.9  12.07
## - Dalc               1       4.0  260.2  12.58
## - Fedu               1       5.0  261.2  13.81
## + reason             3       1.1  255.1  14.35
## - cat_g3             5    4468.6 4724.9 920.73
##
## Step:  AIC=7.99
## G3 ~ sex + age + address + famsize + Pstatus + Medu + Fedu +
##     Mjob + Fjob + traveltime + studytime + failures + schoolsup +
##     famsup + paid + higher + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - age                1       0.2  256.6   6.18
## - freetime           1       0.3  256.8   6.41
## - Fjob               4       5.3  261.7   6.47
## - romantic           1       0.6  257.0   6.73
## - Pstatus            1       0.8  257.2   6.98
## - goout              1       0.8  257.3   7.01
## - Mjob               4       5.8  262.2   7.06
## - sex                1       1.0  257.4   7.20
```

```
## - traveltime         1       1.0  257.4   7.21
## - famsize            1       1.0  257.4   7.21
## - health             1       1.0  257.5   7.25
## - paid               1       1.0  257.5   7.27
## - high_freq_absent   1       1.2  257.6   7.45
## - schoolsup          1       1.5  257.9   7.81
## <none>                           256.4   7.99
## - Walc               1       1.8  258.2   8.19
## - failed             1       2.0  258.4   8.46
## - famsup             1       2.1  258.6   8.62
## - higher             1       2.2  258.7   8.75
## - famrel             1       2.4  258.8   8.88
## - address            1       2.6  259.0   9.18
## - studytime          1       2.7  259.1   9.25
## - Medu               1       2.9  259.3   9.53
## + school             1       0.2  256.2   9.72
## + nursery            1       0.2  256.2   9.74
## + absences           1       0.0  256.4   9.93
## + activities         1       0.0  256.4   9.96
## + first_gen_college  1       0.0  256.4   9.99
## + internet           1       0.0  256.4   9.99
## - failures           1       3.3  259.7  10.05
## + guardian           2       1.4  255.0  10.22
## - Dalc               1       3.8  260.3  10.68
## - Fedu               1       4.9  261.4  12.00
## + reason             3       1.0  255.4  12.74
## - cat_g3             5    4486.9 4743.3 919.96
##
## Step:  AIC=6.18
## G3 ~ sex + address + famsize + Pstatus + Medu + Fedu + Mjob +
##     Fjob + traveltime + studytime + failures + schoolsup + famsup +
##     paid + higher + romantic + famrel + freetime + goout + Dalc +
##     Walc + health + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - freetime           1       0.3  256.9   4.61
## - Fjob               4       5.4  262.0   4.82
## - romantic           1       0.6  257.2   4.94
## - goout              1       0.8  257.4   5.13
## - Pstatus            1       0.8  257.4   5.19
## - Mjob               4       5.8  262.4   5.25
## - health             1       0.9  257.5   5.35
## - traveltime         1       1.0  257.6   5.39
## - sex                1       1.0  257.6   5.41
## - famsize            1       1.0  257.6   5.42
## - paid               1       1.1  257.7   5.49
## - high_freq_absent   1       1.3  257.9   5.81
## <none>                           256.6   6.18
## - Walc               1       1.9  258.5   6.53
## - schoolsup          1       2.1  258.6   6.71
## - failed             1       2.2  258.8   6.85
## - famrel             1       2.3  258.9   6.96
## - famsup             1       2.4  259.0   7.09
## - higher             1       2.5  259.1   7.20
```

```
## - address            1       2.5  259.1   7.24
## - studytime          1       2.8  259.4   7.62
## - Medu               1       3.0  259.6   7.89
## + nursery            1       0.2  256.4   7.97
## + age                1       0.2  256.4   7.99
## + school             1       0.1  256.5   8.07
## + absences           1       0.0  256.5   8.14
## + activities         1       0.0  256.5   8.14
## + first_gen_college  1       0.0  256.6   8.18
## + internet           1       0.0  256.6   8.18
## + guardian           2       1.6  255.0   8.22
## - failures           1       3.3  259.9   8.26
## - Dalc               1       4.0  260.6   9.06
## - Fedu               1       5.0  261.6  10.30
## + reason             3       1.0  255.6  10.95
## - cat_g3             5    4532.6 4789.2 921.01
##
## Step:  AIC=4.61
## G3 ~ sex + address + famsize + Pstatus + Medu + Fedu + Mjob +
##     Fjob + traveltime + studytime + failures + schoolsup + famsup +
##     paid + higher + romantic + famrel + goout + Dalc + Walc +
##     health + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - Fjob               4       5.4  262.3   3.12
## - goout              1       0.6  257.5   3.30
## - romantic           1       0.7  257.6   3.42
## - Pstatus            1       0.7  257.6   3.48
## - paid               1       0.9  257.9   3.75
## - famsize            1       0.9  257.9   3.77
## - traveltime         1       1.0  257.9   3.77
## - health             1       1.0  258.0   3.86
## - sex                1       1.2  258.1   4.07
## - Mjob               4       6.2  263.1   4.12
## - high_freq_absent   1       1.3  258.3   4.24
## <none>                           256.9   4.61
## - Walc               1       2.0  258.9   5.08
## - schoolsup          1       2.1  259.0   5.14
## - famrel             1       2.1  259.0   5.17
## - failed             1       2.3  259.2   5.38
## - famsup             1       2.3  259.2   5.44
## - higher             1       2.4  259.4   5.58
## - address            1       2.4  259.4   5.59
## - studytime          1       2.8  259.7   6.02
## + freetime           1       0.3  256.6   6.18
## - Medu               1       3.1  260.0   6.37
## + age                1       0.2  256.8   6.41
## + nursery            1       0.1  256.8   6.43
## + guardian           2       1.7  255.2   6.46
## + school             1       0.1  256.9   6.52
## + absences           1       0.1  256.9   6.54
## + activities         1       0.0  256.9   6.58
## + internet           1       0.0  256.9   6.60
## + first_gen_college  1       0.0  256.9   6.61
```

```
## - failures           1       3.3  260.2   6.65
## - Dalc               1       4.3  261.3   7.90
## - Fedu               1       4.9  261.8   8.56
## + reason             3       1.1  255.9   9.29
## - cat_g3             5    4689.6 4946.5 929.22
##
## Step:  AIC=3.12
## G3 ~ sex + address + famsize + Pstatus + Medu + Fedu + Mjob +
##     traveltime + studytime + failures + schoolsup + famsup +
##     paid + higher + romantic + famrel + goout + Dalc + Walc +
##     health + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - Pstatus            1       0.5  262.8   1.72
## - traveltime         1       0.5  262.8   1.76
## - health             1       0.7  263.0   1.94
## - goout              1       0.7  263.0   1.99
## - romantic           1       0.8  263.1   2.12
## - paid               1       1.0  263.3   2.32
## - sex                1       1.0  263.3   2.37
## - high_freq_absent   1       1.1  263.4   2.49
## - Mjob               4       6.3  268.5   2.58
## - famsize            1       1.2  263.5   2.58
## <none>                           262.3   3.12
## - Walc               1       2.2  264.5   3.75
## - famrel             1       2.3  264.5   3.84
## - failed             1       2.4  264.7   4.02
## - higher             1       2.6  264.9   4.23
## - address            1       2.6  264.9   4.26
## - famsup             1       2.6  264.9   4.27
## + Fjob               4       5.4  256.9   4.61
## - schoolsup          1       3.0  265.3   4.72
## + age                1       0.3  262.0   4.77
## + nursery            1       0.3  262.0   4.77
## - studytime          1       3.1  265.3   4.79
## + freetime           1       0.3  262.0   4.82
## + guardian           2       1.7  260.5   5.02
## + internet           1       0.0  262.2   5.07
## + first_gen_college  1       0.0  262.3   5.09
## + absences           1       0.0  262.3   5.11
## + activities         1       0.0  262.3   5.12
## + school             1       0.0  262.3   5.12
## - failures           1       3.4  265.7   5.16
## - Medu               1       3.7  266.0   5.58
## - Fedu               1       4.4  266.6   6.32
## - Dalc               1       4.7  267.0   6.74
## + reason             3       1.2  261.1   7.68
## - cat_g3             5    4719.6 4981.9 923.47
##
## Step:  AIC=1.72
## G3 ~ sex + address + famsize + Medu + Fedu + Mjob + traveltime +
##     studytime + failures + schoolsup + famsup + paid + higher +
##     romantic + famrel + goout + Dalc + Walc + health + high_freq_absent +
##     failed + cat_g3
```

```
##
##                     Df Sum of Sq    RSS    AIC
## - traveltime         1       0.5  263.3   0.29
## - health             1       0.7  263.5   0.61
## - goout              1       0.8  263.6   0.65
## - romantic           1       0.9  263.6   0.75
## - paid               1       0.9  263.7   0.78
## - sex                1       1.0  263.7   0.88
## - Mjob               4       6.1  268.9   0.95
## - high_freq_absent   1       1.2  264.0   1.19
## - famsize            1       1.4  264.2   1.43
## <none>                           262.8   1.72
## - Walc               1       2.1  264.9   2.25
## - failed             1       2.3  265.1   2.50
## - famrel             1       2.4  265.2   2.58
## - higher             1       2.6  265.3   2.79
## - famsup             1       2.6  265.4   2.84
## - address            1       2.6  265.4   2.88
## + Pstatus            1       0.5  262.3   3.12
## - schoolsup          1       2.9  265.7   3.21
## + nursery            1       0.3  262.4   3.32
## + age                1       0.3  262.5   3.35
## + guardian           2       1.9  260.8   3.38
## - studytime          1       3.1  265.9   3.40
## + Fjob               4       5.1  257.6   3.48
## + freetime           1       0.2  262.6   3.50
## - failures           1       3.3  266.1   3.63
## + first_gen_college  1       0.0  262.8   3.69
## + internet           1       0.0  262.8   3.71
## + activities         1       0.0  262.8   3.72
## + absences           1       0.0  262.8   3.72
## + school             1       0.0  262.8   3.72
## - Medu               1       3.5  266.3   3.95
## - Fedu               1       4.6  267.4   5.25
## - Dalc               1       4.9  267.7   5.53
## + reason             3       1.2  261.6   6.24
## - cat_g3             5    4732.6 4995.3 922.32
##
## Step:  AIC=0.29
## G3 ~ sex + address + famsize + Medu + Fedu + Mjob + studytime +
##     failures + schoolsup + famsup + paid + higher + romantic +
##     famrel + goout + Dalc + Walc + health + high_freq_absent +
##     failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - health             1       0.7  264.0  -0.87
## - goout              1       0.7  264.0  -0.83
## - paid               1       0.8  264.1  -0.73
## - Mjob               4       5.9  269.2  -0.69
## - sex                1       0.9  264.2  -0.58
## - romantic           1       0.9  264.2  -0.57
## - high_freq_absent   1       1.1  264.3  -0.43
## - famsize            1       1.6  264.8   0.17
## <none>                           263.3   0.29
```

```
## - Walc               1       2.0  265.3   0.69
## - address            1       2.2  265.5   0.92
## - failed             1       2.3  265.6   1.04
## - famrel             1       2.3  265.6   1.08
## - famsup             1       2.5  265.8   1.28
## - higher             1       2.6  265.9   1.45
## + traveltime         1       0.5  262.8   1.72
## - studytime          1       2.9  266.1   1.74
## + Pstatus            1       0.4  262.8   1.76
## + guardian           2       2.0  261.2   1.83
## + nursery            1       0.4  262.9   1.87
## - schoolsup          1       3.0  266.3   1.91
## + age                1       0.3  263.0   1.95
## + freetime           1       0.2  263.1   2.10
## - failures           1       3.3  266.5   2.19
## + internet           1       0.0  263.2   2.28
## + first_gen_college  1       0.0  263.2   2.28
## + school             1       0.0  263.2   2.28
## + activities         1       0.0  263.3   2.29
## + absences           1       0.0  263.3   2.29
## - Medu               1       3.5  266.8   2.48
## + Fjob               4       4.8  258.5   2.54
## - Fedu               1       4.4  267.7   3.57
## - Dalc               1       5.0  268.3   4.29
## + reason             3       1.3  261.9   4.71
## - cat_g3             5    4740.1 5003.3 920.83
##
## Step:  AIC=-0.87
## G3 ~ sex + address + famsize + Medu + Fedu + Mjob + studytime +
##     failures + schoolsup + famsup + paid + higher + romantic +
##     famrel + goout + Dalc + Walc + high_freq_absent + failed +
##     cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - goout              1       0.8  264.7  -1.94
## - paid               1       0.8  264.7  -1.93
## - romantic           1       0.9  264.9  -1.79
## - high_freq_absent   1       0.9  264.9  -1.75
## - sex                1       1.0  264.9  -1.72
## - Mjob               4       6.6  270.6  -1.07
## - famsize            1       1.5  265.5  -1.02
## <none>                           264.0  -0.87
## - Walc               1       1.8  265.7  -0.73
## - famrel             1       2.0  266.0  -0.48
## - address            1       2.1  266.1  -0.37
## - failed             1       2.4  266.4   0.03
## - famsup             1       2.5  266.4   0.06
## - higher             1       2.6  266.6   0.23
## + health             1       0.7  263.3   0.29
## - studytime          1       2.8  266.8   0.51
## + Pstatus            1       0.5  263.4   0.52
## + traveltime         1       0.4  263.5   0.61
## + nursery            1       0.4  263.6   0.65
## + freetime           1       0.2  263.7   0.87
```

```
## + age                1       0.2  263.8   0.90
## - schoolsup          1       3.3  267.2   1.00
## + guardian           2       1.7  262.2   1.06
## + internet           1       0.0  263.9   1.12
## + activities         1       0.0  263.9   1.12
## + first_gen_college  1       0.0  263.9   1.12
## + absences           1       0.0  263.9   1.12
## + school             1       0.0  264.0   1.13
## - failures           1       3.4  267.3   1.14
## - Medu               1       3.8  267.8   1.64
## + Fjob               4       4.4  259.5   1.76
## - Fedu               1       4.5  268.5   2.47
## - Dalc               1       5.2  269.2   3.29
## + reason             3       1.5  262.4   3.31
## - cat_g3             5    4765.1 5029.1 920.45
##
## Step:  AIC=-1.94
## G3 ~ sex + address + famsize + Medu + Fedu + Mjob + studytime +
##     failures + schoolsup + famsup + paid + higher + romantic +
##     famrel + Dalc + Walc + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - paid               1       0.7  265.4  -3.12
## - romantic           1       1.0  265.7  -2.79
## - high_freq_absent   1       1.0  265.7  -2.74
## - sex                1       1.1  265.8  -2.68
## - Mjob               4       6.5  271.2  -2.32
## - famsize            1       1.6  266.3  -2.09
## <none>                           264.7  -1.94
## - address            1       2.1  266.8  -1.48
## - famrel             1       2.2  267.0  -1.28
## - famsup             1       2.3  267.1  -1.15
## - failed             1       2.4  267.2  -1.05
## + goout              1       0.8  264.0  -0.87
## - higher             1       2.6  267.3  -0.84
## + health             1       0.7  264.0  -0.83
## + Pstatus            1       0.6  264.2  -0.61
## - studytime          1       3.0  267.7  -0.43
## + traveltime         1       0.4  264.3  -0.40
## + nursery            1       0.4  264.4  -0.37
## - Walc               1       3.3  268.0  -0.07
## + age                1       0.1  264.6  -0.06
## - schoolsup          1       3.3  268.0  -0.01
## + freetime           1       0.0  264.7   0.01
## + activities         1       0.0  264.7   0.05
## + internet           1       0.0  264.7   0.05
## + first_gen_college  1       0.0  264.7   0.06
## + school             1       0.0  264.7   0.06
## + absences           1       0.0  264.7   0.06
## - failures           1       3.4  268.1   0.11
## + guardian           2       1.6  263.1   0.14
## + Fjob               4       4.6  260.1   0.51
## - Medu               1       4.0  268.7   0.79
## - Fedu               1       4.4  269.1   1.27
```

```
## - Dalc               1       5.1  269.9   2.14
## + reason             3       1.5  263.2   2.27
## - cat_g3             5    4889.7 5154.4 926.23
##
## Step:  AIC=-3.12
## G3 ~ sex + address + famsize + Medu + Fedu + Mjob + studytime +
##     failures + schoolsup + famsup + higher + romantic + famrel +
##     Dalc + Walc + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - sex                1       0.9  266.4  -4.01
## - high_freq_absent   1       1.0  266.4  -3.99
## - romantic           1       1.0  266.4  -3.93
## - famsize            1       1.5  266.9  -3.31
## <none>                           265.4  -3.12
## - Mjob               4       6.8  272.3  -3.07
## - famsup             1       1.9  267.3  -2.88
## - address            1       2.1  267.5  -2.63
## - famrel             1       2.1  267.5  -2.60
## - failed             1       2.2  267.6  -2.48
## - higher             1       2.4  267.8  -2.32
## + health             1       0.7  264.7  -1.98
## + paid               1       0.7  264.7  -1.94
## + goout              1       0.7  264.7  -1.93
## + nursery            1       0.5  265.0  -1.68
## + Pstatus            1       0.4  265.0  -1.65
## + traveltime         1       0.3  265.1  -1.52
## - studytime          1       3.2  268.6  -1.38
## - Walc               1       3.2  268.6  -1.34
## + age                1       0.1  265.3  -1.26
## + internet           1       0.0  265.4  -1.16
## - failures           1       3.4  268.8  -1.16
## + guardian           2       1.7  263.7  -1.15
## + freetime           1       0.0  265.4  -1.14
## + activities         1       0.0  265.4  -1.14
## + first_gen_college  1       0.0  265.4  -1.12
## + school             1       0.0  265.4  -1.12
## + absences           1       0.0  265.4  -1.12
## - schoolsup          1       3.5  268.9  -1.01
## + Fjob               4       4.7  260.7  -0.75
## - Medu               1       3.9  269.3  -0.53
## - Fedu               1       4.3  269.7  -0.07
## - Dalc               1       5.5  270.9   1.39
## + reason             3       1.2  264.2   1.41
## - cat_g3             5    4910.2 5175.6 925.53
##
## Step:  AIC=-4.01
## G3 ~ address + famsize + Medu + Fedu + Mjob + studytime + failures +
##     schoolsup + famsup + higher + romantic + famrel + Dalc +
##     Walc + high_freq_absent + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - high_freq_absent   1       0.8  267.1  -5.09
## - romantic           1       0.8  267.2  -5.01
```

```
## <none>                           266.4  -4.01
## - famsize           1       1.7  268.0  -4.01
## - famrel            1       1.9  268.2  -3.79
## - address           1       1.9  268.3  -3.71
## - famsup            1       2.1  268.5  -3.48
## - Mjob              4       7.3  273.7  -3.42
## - failed            1       2.4  268.7  -3.22
## + sex               1       0.9  265.4  -3.12
## - studytime         1       2.6  268.9  -2.97
## + goout             1       0.8  265.6  -2.93
## - Walc              1       2.6  269.0  -2.92
## + health            1       0.7  265.6  -2.90
## + paid              1       0.6  265.8  -2.68
## - higher            1       2.8  269.2  -2.67
## + nursery           1       0.5  265.8  -2.65
## + Pstatus           1       0.4  266.0  -2.48
## + traveltime        1       0.3  266.0  -2.39
## + age               1       0.1  266.2  -2.16
## + freetime          1       0.1  266.3  -2.10
## + internet          1       0.0  266.3  -2.05
## + first_gen_college 1       0.0  266.3  -2.03
## + absences          1       0.0  266.3  -2.02
## + guardian          2       1.7  264.7  -2.02
## + school            1       0.0  266.3  -2.02
## + activities        1       0.0  266.4  -2.01
## - failures          1       3.4  269.8  -1.99
## - Medu              1       3.8  270.2  -1.51
## - schoolsup         1       3.8  270.2  -1.49
## + Fjob              4       4.5  261.8  -1.44
## - Fedu              1       4.2  270.6  -1.01
## + reason            3       1.4  265.0   0.36
## - Dalc              1       5.8  272.2   0.83
## - cat_g3            5    5011.2 5277.5 929.69
##
## Step:  AIC=-5.09
## G3 ~ address + famsize + Medu + Fedu + Mjob + studytime + failures +
##     schoolsup + famsup + higher + romantic + famrel + Dalc +
##     Walc + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - romantic           1       1.0  268.1  -5.89
## - Mjob               4       6.8  273.9  -5.17
## <none>                           267.1  -5.09
## - famsize            1       1.8  268.9  -4.96
## - famsup             1       2.0  269.1  -4.76
## - famrel             1       2.1  269.2  -4.60
## - address            1       2.1  269.3  -4.56
## - studytime          1       2.4  269.5  -4.28
## - Walc               1       2.5  269.6  -4.15
## + goout              1       0.8  266.3  -4.06
## + high_freq_absent   1       0.8  266.4  -4.01
## + sex                1       0.8  266.4  -3.99
## - higher             1       2.7  269.9  -3.88
## + health             1       0.6  266.5  -3.83
```

```
## + nursery            1       0.6  266.5  -3.82
## + paid               1       0.5  266.6  -3.72
## - failed             1       2.9  270.0  -3.66
## + Pstatus            1       0.5  266.7  -3.65
## + guardian           2       2.0  265.1  -3.45
## + absences           1       0.3  266.9  -3.42
## + age                1       0.2  266.9  -3.35
## + traveltime         1       0.2  266.9  -3.34
## - Medu               1       3.2  270.4  -3.27
## + internet           1       0.1  267.1  -3.17
## + freetime           1       0.1  267.1  -3.16
## + first_gen_college  1       0.0  267.1  -3.12
## + school             1       0.0  267.1  -3.12
## + activities         1       0.0  267.1  -3.09
## - failures           1       3.8  270.9  -2.65
## + Fjob               4       4.5  262.7  -2.42
## - Fedu               1       4.0  271.1  -2.42
## - schoolsup          1       4.2  271.4  -2.13
## + reason             3       1.1  266.1  -0.35
## - Dalc               1       6.4  273.6   0.42
## - cat_g3             5    5030.3 5297.4 928.88
##
## Step:  AIC=-5.89
## G3 ~ address + famsize + Medu + Fedu + Mjob + studytime + failures +
##     schoolsup + famsup + higher + famrel + Dalc + Walc + failed +
##     cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - Mjob               4       6.4  274.6  -6.39
## <none>                            268.1  -5.89
## - famsize            1       1.9  270.1  -5.63
## - famsup             1       2.0  270.1  -5.58
## - famrel             1       2.1  270.3  -5.40
## - address            1       2.1  270.3  -5.39
## + romantic           1       1.0  267.1  -5.09
## + high_freq_absent   1       0.9  267.2  -5.01
## + goout              1       0.9  267.3  -4.92
## - studytime          1       2.8  270.9  -4.65
## - Walc               1       2.8  270.9  -4.62
## + nursery            1       0.6  267.5  -4.61
## + sex                1       0.6  267.6  -4.59
## + health             1       0.6  267.6  -4.57
## + paid               1       0.6  267.6  -4.56
## + guardian           2       2.2  265.9  -4.52
## - Medu               1       2.9  271.0  -4.49
## + Pstatus            1       0.5  267.6  -4.49
## + absences           1       0.4  267.7  -4.41
## - failed             1       3.1  271.3  -4.23
## + traveltime         1       0.3  267.9  -4.21
## + age                1       0.3  267.9  -4.20
## - higher             1       3.2  271.3  -4.14
## + internet           1       0.1  268.0  -4.05
## + freetime           1       0.1  268.1  -3.97
## + first_gen_college  1       0.0  268.1  -3.90
```

```
## + school             1       0.0  268.1  -3.90
## + activities         1       0.0  268.1  -3.89
## + Fjob               4       4.6  263.5  -3.41
## - failures           1       3.9  272.0  -3.38
## - Fedu               1       4.2  272.3  -3.01
## - schoolsup          1       4.9  273.1  -2.13
## + reason             3       0.9  267.2  -0.98
## - Dalc               1       7.1  275.2   0.36
## - cat_g3             5    5122.8 5391.0 932.41
##
## Step:  AIC=-6.39
## G3 ~ address + famsize + Medu + Fedu + studytime + failures +
##     schoolsup + famsup + higher + famrel + Dalc + Walc + failed +
##     cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - Medu               1       0.2  274.7  -8.21
## - famsup             1       1.6  276.2  -6.51
## <none>                           274.6  -6.39
## - studytime          1       1.8  276.4  -6.32
## + health             1       1.3  273.3  -5.93
## + Mjob               4       6.4  268.1  -5.89
## - higher             1       2.4  277.0  -5.65
## + sex                1       1.1  273.5  -5.63
## + paid               1       0.9  273.7  -5.41
## - famrel             1       2.6  277.2  -5.38
## + nursery            1       0.8  273.8  -5.32
## + goout              1       0.7  273.9  -5.21
## + romantic           1       0.7  273.9  -5.17
## - famsize            1       2.9  277.4  -5.12
## - Walc               1       2.9  277.5  -5.03
## - address            1       3.0  277.6  -4.99
## + internet           1       0.4  274.2  -4.86
## + guardian           2       2.1  272.5  -4.83
## - Fedu               1       3.1  277.7  -4.78
## + freetime           1       0.3  274.3  -4.78
## + high_freq_absent   1       0.3  274.3  -4.73
## + Pstatus            1       0.3  274.3  -4.67
## + traveltime         1       0.2  274.4  -4.57
## - failed             1       3.3  277.9  -4.56
## + age                1       0.2  274.4  -4.56
## + absences           1       0.1  274.5  -4.52
## + first_gen_college  1       0.1  274.5  -4.47
## + activities         1       0.0  274.6  -4.42
## + school             1       0.0  274.6  -4.39
## - failures           1       3.7  278.3  -4.16
## + Fjob               4       4.9  269.7  -4.05
## - schoolsup          1       6.0  280.6  -1.53
## + reason             3       1.0  273.6  -1.48
## - Dalc               1       6.4  281.0  -1.08
## - cat_g3             5    5220.6 5495.2 930.46
##
## Step:  AIC=-8.21
## G3 ~ address + famsize + Fedu + studytime + failures + schoolsup +
```

```
##     famsup + higher + famrel + Dalc + Walc + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - studytime          1       1.7  276.5  -8.24
## <none>                            274.7  -8.21
## - famsup             1       1.8  276.5  -8.18
## + health             1       1.3  273.4  -7.76
## - higher             1       2.4  277.2  -7.41
## + sex                1       1.0  273.7  -7.36
## - famrel             1       2.6  277.4  -7.22
## + paid               1       0.8  273.9  -7.15
## + goout              1       0.7  274.0  -7.07
## + nursery            1       0.7  274.0  -7.06
## - address            1       2.8  277.6  -6.96
## + romantic           1       0.6  274.1  -6.94
## - famsize            1       2.9  277.6  -6.93
## - Walc               1       3.0  277.7  -6.78
## + guardian           2       2.1  272.6  -6.69
## + internet           1       0.3  274.4  -6.60
## + freetime           1       0.3  274.4  -6.56
## + Pstatus            1       0.2  274.5  -6.48
## + high_freq_absent   1       0.2  274.5  -6.46
## + traveltime         1       0.2  274.6  -6.41
## + age                1       0.2  274.6  -6.40
## + Medu               1       0.2  274.6  -6.39
## + absences           1       0.1  274.7  -6.31
## + activities         1       0.0  274.7  -6.24
## - failed             1       3.5  278.2  -6.23
## + first_gen_college  1       0.0  274.7  -6.21
## + school             1       0.0  274.7  -6.21
## - Fedu               1       3.7  278.4  -6.04
## + Fjob               4       5.0  269.7  -6.01
## - failures           1       3.8  278.5  -5.87
## + Mjob               4       3.7  271.0  -4.49
## - schoolsup          1       5.9  280.7  -3.49
## + reason             3       1.0  273.7  -3.40
## - Dalc               1       6.4  281.1  -2.97
## - cat_g3             5    5322.2 5597.0 934.26
##
## Step:  AIC=-8.24
## G3 ~ address + famsize + Fedu + failures + schoolsup + famsup +
##     higher + famrel + Dalc + Walc + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## - famsup             1       1.3  277.7  -8.77
## <none>                            276.5  -8.24
## + studytime          1       1.7  274.7  -8.21
## - higher             1       2.0  278.5  -7.92
## + health             1       1.2  275.2  -7.64
## + paid               1       1.0  275.4  -7.43
## - famrel             1       2.5  279.0  -7.41
## - address            1       2.5  279.0  -7.39
## + nursery            1       1.0  275.5  -7.33
## + romantic           1       0.9  275.6  -7.27
```

```
## - famsize            1       2.7  279.1  -7.20
## + goout              1       0.8  275.7  -7.16
## + guardian           2       2.3  274.1  -6.93
## + sex                1       0.5  276.0  -6.77
## + internet           1       0.4  276.0  -6.72
## + age                1       0.3  276.2  -6.55
## + Pstatus            1       0.3  276.2  -6.55
## + Fjob               4       5.4  271.0  -6.50
## + freetime           1       0.2  276.2  -6.49
## - Fedu               1       3.3  279.8  -6.48
## + high_freq_absent   1       0.2  276.3  -6.46
## + traveltime         1       0.1  276.4  -6.36
## + absences           1       0.1  276.4  -6.33
## + Medu               1       0.1  276.4  -6.32
## + activities         1       0.1  276.4  -6.32
## - failed             1       3.5  280.0  -6.27
## + first_gen_college  1       0.0  276.5  -6.25
## + school             1       0.0  276.5  -6.25
## - Walc               1       3.6  280.1  -6.15
## - failures           1       4.2  280.7  -5.46
## + Mjob               4       3.4  273.1  -4.11
## - schoolsup          1       5.5  282.0  -4.02
## + reason             3       0.9  275.6  -3.24
## - Dalc               1       6.4  282.9  -2.98
## - cat_g3             5    5341.4 5617.9 933.44
##
## Step:  AIC=-8.77
## G3 ~ address + famsize + Fedu + failures + schoolsup + higher +
##     famrel + Dalc + Walc + failed + cat_g3
##
##                     Df Sum of Sq    RSS    AIC
## <none>                           277.7  -8.77
## + famsup             1       1.3  276.5  -8.24
## + studytime          1       1.2  276.5  -8.18
## + health             1       1.2  276.5  -8.14
## - higher             1       2.4  280.1  -8.10
## + nursery            1       1.0  276.8  -7.90
## - address            1       2.5  280.3  -7.89
## - famrel             1       2.6  280.3  -7.88
## + romantic           1       0.9  276.9  -7.74
## - Fedu               1       2.7  280.5  -7.71
## + goout              1       0.8  277.0  -7.65
## + sex                1       0.7  277.1  -7.54
## + guardian           2       2.3  275.5  -7.38
## + age                1       0.4  277.3  -7.29
## + paid               1       0.4  277.3  -7.27
## - famsize            1       3.1  280.8  -7.26
## - Walc               1       3.2  281.0  -7.11
## + Fjob               4       5.5  272.2  -7.10
## + Pstatus            1       0.3  277.5  -7.08
## + internet           1       0.3  277.5  -7.08
## + freetime           1       0.2  277.5  -7.02
## + Medu               1       0.2  277.6  -6.96
## + high_freq_absent   1       0.1  277.6  -6.91
```

```
## + traveltime         1       0.1  277.7  -6.88
## + activities         1       0.1  277.7  -6.85
## + absences           1       0.0  277.7  -6.83
## + school             1       0.0  277.7  -6.79
## + first_gen_college  1       0.0  277.7  -6.77
## - failed             1       3.7  281.5  -6.58
## - failures           1       4.5  282.3  -5.68
## - schoolsup          1       5.7  283.5  -4.31
## + Mjob               4       2.9  274.8  -4.11
## + reason             3       1.1  276.7  -3.98
## - Dalc               1       6.3  284.0  -3.69
## - cat_g3             5    5383.9 5661.7 933.89
```

```
AIC(step.model_tr)
```

```
## [1] 889.9955
```

```
summary(step.model_tr)
```

```
##
## Call:
## lm(formula = G3 ~ address + famsize + Fedu + failures + schoolsup +
##     higher + famrel + Dalc + Walc + failed + cat_g3, data = training)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4957 -0.4924 -0.0280  0.5656  2.3831
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.78188    0.41240  28.569  < 2e-16 ***
## addressU      0.21727    0.13092   1.660  0.09806 .
## famsizeLE3    0.22315    0.12191   1.830  0.06817 .
## Fedu          0.09029    0.05282   1.709  0.08843 .
## failures     -0.33333    0.15097  -2.208  0.02800 *
## schoolsupyes -0.41118    0.16514  -2.490  0.01332 *
## higheryes    -0.40362    0.25262  -1.598  0.11115
## famrellow     0.21856    0.13162   1.661  0.09784 .
## Dalclow      -0.77438    0.29695  -2.608  0.00957 **
## Walclow       0.28854    0.15421   1.871  0.06231 .
## failedyes     0.54916    0.27424   2.003  0.04613 *
## cat_g3.L     14.42916    0.21651  66.643  < 2e-16 ***
## cat_g3.Q     -3.84056    0.18469 -20.795  < 2e-16 ***
## cat_g3.C      1.36622    0.18446   7.407 1.32e-12 ***
## cat_g3^4     -0.37757    0.17256  -2.188  0.02944 *
## cat_g3^5      0.41451    0.13520   3.066  0.00237 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9622 on 300 degrees of freedom
## Multiple R-squared:   0.96,  Adjusted R-squared:  0.958
## F-statistic: 480.2 on 15 and 300 DF,  p-value: < 2.2e-16
```

AIC of 1793.426 and adjusted R-squared of 0.2742.

Active variables: goout, absences, first_gen_college, failed, freetime, romantic, famsup, schoolsup, studytime,

famsize, sex. Let's explore these with interaction terms.

```
activelm_tr <- lm(G3 ~ (sex + schoolsup + romantic + freetime + goout + absences + failed + studytime +
summary(activelm_tr)
```

```
##
## Call:
## lm(formula = G3 ~ (sex + schoolsup + romantic + freetime + goout +
##     absences + failed + studytime + first_gen_college + famsup +
##     famsize), data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.5751  -2.1790   0.4765   2.7999  11.0657
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          10.60588    0.99491  10.660  < 2e-16 ***
## sexM                  0.84292    0.50654   1.664 0.097129 .
## schoolsupyes         -0.92949    0.68924  -1.349 0.178478
## romanticyes          -1.35942    0.50766  -2.678 0.007813 **
## freetimelow          -1.42836    0.50007  -2.856 0.004581 **
## gooutlow              1.70615    0.50120   3.404 0.000753 ***
## absences              0.07599    0.02813   2.701 0.007297 **
## failedyes            -3.75325    0.59064  -6.355 7.63e-10 ***
## studytime             0.56316    0.29587   1.903 0.057930 .
## first_gen_collegeyes -1.34272    0.49968  -2.687 0.007603 **
## famsupyes            -0.87684    0.49917  -1.757 0.079996 .
## famsizeLE3            1.14706    0.51719   2.218 0.027304 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.097 on 304 degrees of freedom
## Multiple R-squared:  0.2655, Adjusted R-squared:  0.2389
## F-statistic: 9.987 on 11 and 304 DF,  p-value: 1.635e-15
```

```
AIC(activelm_tr)
```

```
## [1] 1801.82
```

AIC of 1801.82, Adjusted R-squared of 0.2389.

Adding all interaction terms:

```
interlm_tr <- lm(G3 ~ (sex + schoolsup + romantic + freetime + goout + absences + failed + studytime + 
summary(interlm_tr)
```

```
##
## Call:
## lm(formula = G3 ~ (sex + schoolsup + romantic + freetime + goout +
##     absences + failed + studytime + first_gen_college + famsup +
##     famsize)^2, data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.1133  -1.9546   0.3402   2.2558   8.0897
##
```

```
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       17.828089   2.724408   6.544 3.39e-10 ***
## sexM                              -2.236645   1.989617  -1.124 0.262028
## schoolsupyes                       2.491733   2.900969   0.859 0.391205
## romanticyes                       -2.924267   2.589279  -1.129 0.259826
## freetimelow                       -2.019097   2.075748  -0.973 0.331642
## gooutlow                           1.696368   2.184918   0.776 0.438251
## absences                          -0.123817   0.172425  -0.718 0.473376
## failedyes                         -5.471058   2.886823  -1.895 0.059226 .
## studytime                         -2.406763   1.037396  -2.320 0.021150 *
## first_gen_collegeyes              -7.000153   1.957673  -3.576 0.000419 ***
## famsupyes                         -3.708275   1.989413  -1.864 0.063498 .
## famsizeLE3                         1.443291   2.318042   0.623 0.534095
## sexM:schoolsupyes                 -2.993395   1.677001  -1.785 0.075483 .
## sexM:romanticyes                   1.287244   1.071327   1.202 0.230682
## sexM:freetimelow                   1.086732   1.060899   1.024 0.306664
## sexM:gooutlow                     -0.667762   1.122071  -0.595 0.552307
## sexM:absences                     -0.009894   0.099224  -0.100 0.920654
## sexM:failedyes                    -3.771541   1.544833  -2.441 0.015328 *
## sexM:studytime                     0.167683   0.614759   0.273 0.785263
## sexM:first_gen_collegeyes          4.029339   1.103393   3.652 0.000317 ***
## sexM:famsupyes                     1.168833   1.082460   1.080 0.281279
## sexM:famsizeLE3                   -0.415000   1.199880  -0.346 0.729734
## schoolsupyes:romanticyes           2.301138   1.966379   1.170 0.243023
## schoolsupyes:freetimelow           0.798298   1.701002   0.469 0.639258
## schoolsupyes:gooutlow              0.312639   1.893847   0.165 0.869014
## schoolsupyes:absences             -0.244795   0.104775  -2.336 0.020265 *
## schoolsupyes:failedyes             2.951505   1.786648   1.652 0.099799 .
## schoolsupyes:studytime            -2.227824   0.914124  -2.437 0.015506 *
## schoolsupyes:first_gen_collegeyes  3.809002   1.644812   2.316 0.021384 *
## schoolsupyes:famsupyes            -0.150740   1.589188  -0.095 0.924508
## schoolsupyes:famsizeLE3           -3.380800   1.893210  -1.786 0.075357 .
## romanticyes:freetimelow           -1.775277   1.097661  -1.617 0.107074
## romanticyes:gooutlow              -0.260153   1.149870  -0.226 0.821196
## romanticyes:absences              -0.018206   0.079839  -0.228 0.819808
## romanticyes:failedyes             -0.209183   1.298112  -0.161 0.872111
## romanticyes:studytime              1.104782   0.738687   1.496 0.136023
## romanticyes:first_gen_collegeyes   0.487793   1.144907   0.426 0.670436
## romanticyes:famsupyes             -0.478762   1.171035  -0.409 0.683011
## romanticyes:famsizeLE3             0.419434   1.265768   0.331 0.740646
## freetimelow:gooutlow              -1.544626   0.993952  -1.554 0.121449
## freetimelow:absences               0.133237   0.079625   1.673 0.095522 .
## freetimelow:failedyes             -2.576119   1.337191  -1.927 0.055178 .
## freetimelow:studytime              0.592657   0.655277   0.904 0.366638
## freetimelow:first_gen_collegeyes   0.974338   1.060793   0.918 0.359246
## freetimelow:famsupyes             -0.383726   1.107122  -0.347 0.729187
## freetimelow:famsizeLE3            -0.642029   1.131025  -0.568 0.570783
## gooutlow:absences                 -0.067710   0.086300  -0.785 0.433442
## gooutlow:failedyes                 0.526401   1.319646   0.399 0.690312
## gooutlow:studytime                 1.208907   0.697376   1.734 0.084243 .
## gooutlow:first_gen_collegeyes      0.114464   1.128380   0.101 0.919282
## gooutlow:famsupyes                -0.982709   1.089033  -0.902 0.367733
## gooutlow:famsizeLE3               -0.573774   1.175861  -0.488 0.626007
```

```
## absences:failedyes                 0.215436   0.070183   3.070 0.002381 **
## absences:studytime                -0.006741   0.056951  -0.118 0.905869
## absences:first_gen_collegeyes      0.116287   0.081528   1.426 0.155018
## absences:famsupyes                 0.171607   0.072160   2.378 0.018156 *
## absences:famsizeLE3               -0.152720   0.081077  -1.884 0.060779 .
## failedyes:studytime               -0.340481   0.950143  -0.358 0.720387
## failedyes:first_gen_collegeyes     2.933576   1.635308   1.794 0.074043 .
## failedyes:famsupyes                0.694334   1.308619   0.531 0.596179
## failedyes:famsizeLE3               1.667873   1.465060   1.138 0.256034
## studytime:first_gen_collegeyes     0.989623   0.642846   1.539 0.124967
## studytime:famsupyes                1.177579   0.656217   1.794 0.073947 .
## studytime:famsizeLE3               0.913739   0.752968   1.214 0.226083
## first_gen_collegeyes:famsupyes     0.202676   1.124631   0.180 0.857130
## first_gen_collegeyes:famsizeLE3   -0.920305   1.184496  -0.777 0.437920
## famsupyes:famsizeLE3               0.116438   1.163129   0.100 0.920339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.79 on 249 degrees of freedom
## Multiple R-squared:  0.4852, Adjusted R-squared:  0.3488
## F-statistic: 3.556 on 66 and 249 DF,  p-value: 3.206e-13
```

```
AIC(interlm_tr)
```

```
## [1] 1799.471
```

Significant interactions: absences*famsup, absences*failed schoolsup*first_gen_college, schoolsup*absences, schoolsup*studytime, sex*failed, sex*first_gen_college

```
interlm_tr1 <- lm(G3 ~ (sex + schoolsup + romantic + freetime + goout + absences + failed + studytime +
summary(interlm_tr1)
```

```
##
## Call:
## lm(formula = G3 ~ (sex + schoolsup + romantic + freetime + goout +
##     absences + failed + studytime + first_gen_college + famsup +
##     famsize + absences * famsup + absences * failed + schoolsup *
##     first_gen_college + schoolsup * absences + schoolsup * studytime +
##     sex * failed + sex * first_gen_college), data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4178  -2.2623   0.5875   2.6372   8.9101
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     11.693293   1.020762  11.455  < 2e-16 ***
## sexM                            -0.887111   0.727008  -1.220 0.223349
## schoolsupyes                     0.352447   1.999087   0.176 0.860175
## romanticyes                     -1.643156   0.483126  -3.401 0.000763 ***
## freetimelow                     -1.265993   0.472756  -2.678 0.007821 **
## gooutlow                         1.556515   0.482777   3.224 0.001405 **
## absences                         0.006135   0.044127   0.139 0.889517
## failedyes                       -4.414005   0.906859  -4.867 1.84e-06 ***
## studytime                        0.805798   0.304397   2.647 0.008549 **
## first_gen_collegeyes            -3.405193   0.672109  -5.066 7.13e-07 ***
```

```
## famsupyes                          -1.267101   0.578374  -2.191 0.029243 *
## famsizeLE3                          1.136219   0.492371   2.308 0.021706 *
## absences:famsupyes                  0.066071   0.056944   1.160 0.246872
## absences:failedyes                  0.202897   0.061362   3.307 0.001061 **
## schoolsupyes:first_gen_collegeyes   4.459123   1.373647   3.246 0.001303 **
## schoolsupyes:absences              -0.099735   0.077493  -1.287 0.199090
## schoolsupyes:studytime             -1.634406   0.749457  -2.181 0.029984 *
## sexM:failedyes                     -2.147021   1.123717  -1.911 0.057013 .
## sexM:first_gen_collegeyes           3.729817   0.948321   3.933 0.000104 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.84 on 297 degrees of freedom
## Multiple R-squared:  0.3696, Adjusted R-squared:  0.3314
## F-statistic: 9.673 on 18 and 297 DF,  p-value: < 2.2e-16
```

```
AIC(interlm_tr1)
```

```
## [1] 1767.509
```

Schoolsup seems to have a high correlation with first_gen_college and studytime. Absences hasa high correlation with failed. Both of these make sense intuitively. Sex and first_Gen_college seems to have a high correlation, which does not make much sense intuitively and should be explored further.

Including only active variables and paring down the model:

```
interlm_tr2 <- lm(G3 ~ (romantic + freetime + goout + failed + studytime + first_gen_college + famsup +
summary(interlm_tr2)
```

```
##
## Call:
## lm(formula = G3 ~ (romantic + freetime + goout + failed + studytime +
##     first_gen_college + famsup + famsize + absences * failed +
##     schoolsup * first_gen_college + schoolsup * studytime + sex *
##     first_gen_college), data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.1510  -2.1029   0.4616   2.4331   8.6819
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        11.56650    1.00954  11.457  < 2e-16 ***
## romanticyes                        -1.57486    0.48280  -3.262 0.001234 **
## freetimelow                        -1.20937    0.47412  -2.551 0.011246 *
## gooutlow                            1.67484    0.47822   3.502 0.000532 ***
## failedyes                          -5.57431    0.70675  -7.887 5.80e-14 ***
## studytime                           0.82586    0.30416   2.715 0.007008 **
## first_gen_collegeyes               -3.32647    0.67292  -4.943 1.28e-06 ***
## famsupyes                          -0.96589    0.47204  -2.046 0.041608 *
## famsizeLE3                          1.19057    0.48812   2.439 0.015304 *
## absences                            0.02018    0.03122   0.646 0.518453
## schoolsupyes                        0.03990    1.96359   0.020 0.983804
## sexM                               -1.06718    0.72433  -1.473 0.141707
## failedyes:absences                  0.23837    0.05954   4.004 7.87e-05 ***
## first_gen_collegeyes:schoolsupyes   4.18746    1.35602   3.088 0.002203 **
```

```
## studytime:schoolsupyes            -1.62659    0.75000  -2.169 0.030884 *
## first_gen_collegeyes:sexM          3.34855    0.92366   3.625 0.000339 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.857 on 300 degrees of freedom
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.3255
## F-statistic: 11.14 on 15 and 300 DF,  p-value: < 2.2e-16
```

```
AIC(interlm_tr2)
```

```
## [1] 1767.442
```

AIC of 1767.442, Adjusted R-squared of 0.3255.

Using the model on the testing set:

```
pred.lm <- predict(interlm_tr2, testing)
pred.lm
```

```
##        11        17        18        21        31        32        34        43
## 12.717789 13.664748 10.446235 13.825867 12.859976 11.650605 13.829146 11.690972
##        46        58        59        60        72        76        78        80
##  9.281686 11.731340 13.877800 12.758157 14.268208 11.771707 10.434069 14.169365
##        89        94        95        97        98       115       116       119
## 11.376307 12.717789 12.961230  9.278423  6.181993 12.434563 11.225503  9.594648
##       120       130       133       149       150       152       155       160
## 13.121112 11.486639  7.957911  9.249891  6.963512  6.939987  8.730383  5.092277
##       174       175       178       180       187       194       196       200
##  1.533208 10.197155 11.044474 10.156482  9.312343 11.008302 13.299388 15.117727
##       224       226       227       229       231       235       239       240
## 11.672689  7.344933  8.884210 12.545147 13.581960 14.435933 10.397581  6.598803
##       243       244       250       254       260       264       269       273
## 12.515736 11.790639 11.812723 11.812723 12.834007 11.263805  9.812116 11.831628
##       279       281       282       283       284       288       292       300
##  5.768204 11.546389 11.876063 12.253786 10.946656 10.338282 13.543646 11.750747
##       304       315       322       328       329       336       340       341
## 11.043036  7.653679  8.058670 10.542833 13.219463 14.367006  9.472058  3.393286
##       347       358       369       378       385       386       389
## 12.049147 10.013290  6.990610 16.164352 11.045477 11.223437  8.907049
```

```
mse_test <- mean((pred.lm - testing$G3)^2)
testing$G3
```

```
##  [1]  9 14 10 15 12 17 12 18  6 15  9 16 10 10 11  5 10 10 14 15 10  9 16  8 13
## [26] 18 12  0 10 14 12 12  0  9  6 11 11 10 15 10 13  7 15  9 14  6 11  0  0 12
## [51] 15  8  0  9 10 11  8  8 10 12 10 12 15 16 18 13  9 10  9 15 10 11 16 11 10
## [76] 10  5 10  8
```

```
mse_test
```

```
## [1] 18.86165
```

Regression random forest

```
library(randomForest)
reg.rf <- randomForest(G3 ~ . -G1 -G2 -G3 -ord_g3 -famsup -internet -Medu -Fedu, data=training, mtry=3,
```

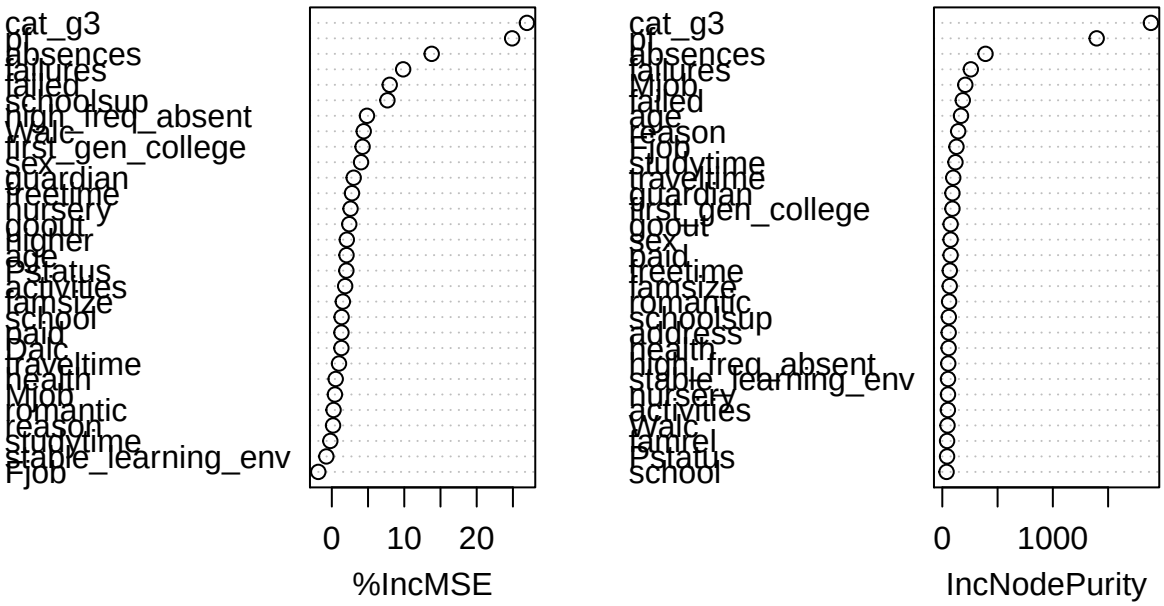
```
                         importance=TRUE, na.action=na.omit)
print(reg.rf)
```

```
##
## Call:
##  randomForest(formula = G3 ~ . - G1 - G2 - G3 - ord_g3 - famsup -      internet - Medu - Fedu, data =
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 4.915282
##                     % Var explained: 77.64
```

```
importance(reg.rf)
```

```
##                        %IncMSE IncNodePurity
## school               1.3487120     38.593271
## sex                  4.0322388     75.162968
## age                  2.0273684    169.083527
## address             -2.0817327     57.675152
## famsize              1.5210162     67.162107
## Pstatus              2.0065689     43.155041
## Mjob                 0.4269056    207.942562
## Fjob                -1.8683843    128.268967
## reason               0.1685197    143.725490
## guardian             3.0175058     91.762950
## traveltime           0.9894995     98.950533
## studytime           -0.2098575    118.945636
## failures             9.8451942    257.973987
## schoolsup            7.6754381     58.881257
## paid                 1.3263437     75.146152
## activities           1.8405176     50.425898
## nursery              2.5847390     51.530905
## higher               2.0682247     28.662453
## romantic             0.2376836     61.569808
## famrel              -1.9370858     43.205851
## freetime             2.7809252     67.945159
## goout                2.4048622     75.264740
## Dalc                 1.3197823      8.379323
## Walc                 4.4021639     47.143460
## health               0.5343537     57.289239
## absences            13.7857279    390.949107
## first_gen_college    4.2468703     91.744769
## stable_learning_env -0.7563498     51.705492
## high_freq_absent     4.8500305     53.864402
## failed               7.9919510    185.257969
## cat_g3              26.9317342   1887.141422
## pf                  24.8918809   1396.530977
```

```
varImpPlot(reg.rf)
```




```
yhat_rf <- predict(reg.rf, newdata = testing)
mse_test.rf <- mean((yhat_rf - testing$G3)^2)

mse_test.rf
```

```
## [1] 4.116767
```

Improved test MSE compared to the linear model. test MSE = 14.085. 28.5% variation explained; mean of squared residuals is 15.7.

```
reg.rf1 <- randomForest(G3 ~ failed + absences + schoolsup + first_gen_college + age + studytime + famsi
                        importance=TRUE, na.action=na.omit)
print(reg.rf1)
```

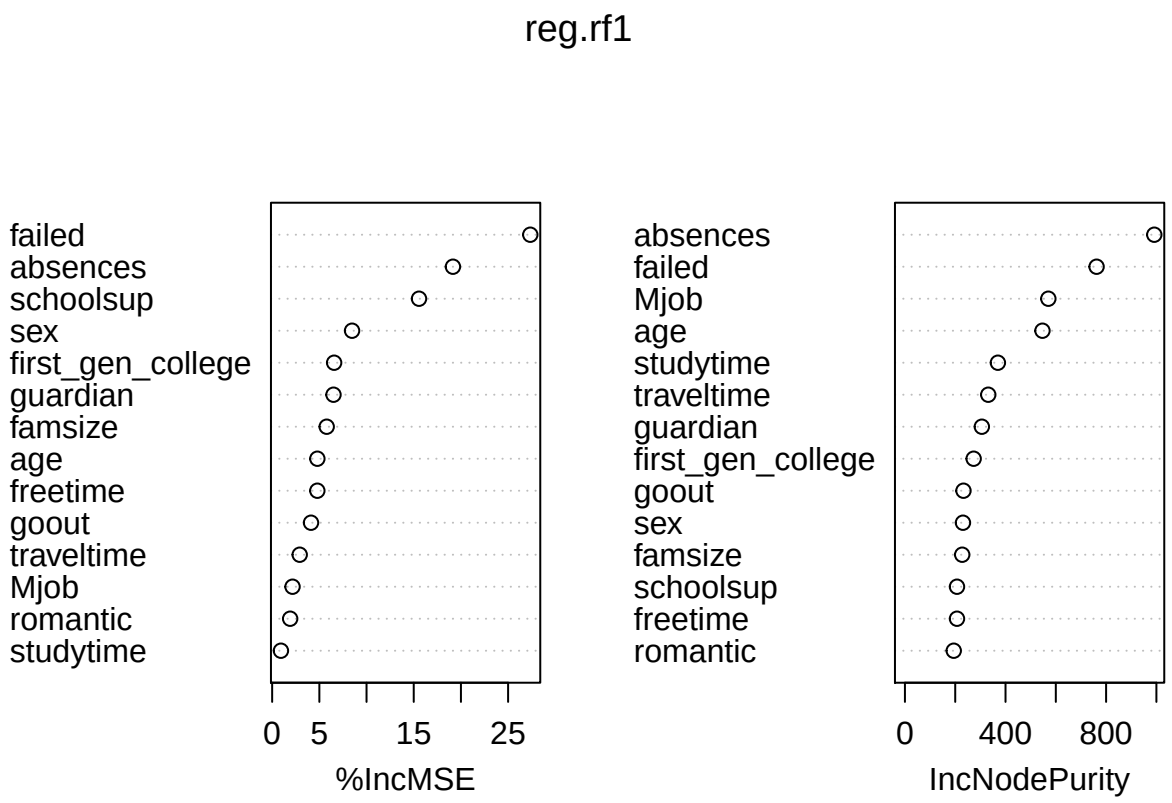
```
##
## Call:
##  randomForest(formula = G3 ~ failed + absences + schoolsup + first_gen_college +      age + studytime
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 15.12239
##                     % Var explained: 31.21
```

```
importance(reg.rf1)
```

```
##                      %IncMSE IncNodePurity
## failed            27.3499827      762.1617
## absences          19.1545094      991.6197
## schoolsup         15.5597996      207.0295
## first_gen_college  6.5657481      273.2462
```

```
## age                4.7892973      546.8774
## studytime          0.9291226      369.6630
## famsize            5.7792007      227.7688
## guardian           6.4997815      305.8171
## freetime           4.7869360      207.0203
## Mjob               2.1598038      570.3071
## romantic           1.9072324      194.0506
## traveltime         2.9179682      331.3511
## sex                8.4653943      230.7749
## goout              4.1164839      232.7351
```

```
varImpPlot(reg.rf1)
```




```
yhat_rf1 <- predict(reg.rf1, newdata = testing)
mse_test.rf1 <- mean((yhat_rf1 - testing$G3)^2)

mse_test.rf1
```

```
## [1] 14.65102
```

Test MSE of 14.973; 32.4% of var explained by model; mean of squared residuals: 14.86

Multicategory ordinal logit model

Due to the way grades are assigned as values between 0 and 20, we would like to consider G3 as an ordered categorical variable with 21 levels. This would allow us to fit a multicategory ordinal logistic model to the data.

We examine the EDA and active variables in the linear model to choose the predictors in our base model.

Fitting the base model:

```r
mod <-polr(ord_g3 ~ . -G1 -G2 -G3 -cat_g3 -pf, data = training)
summary(mod)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = ord_g3 ~ . - G1 - G2 - G3 - cat_g3 - pf, data = training)
##
## Coefficients:
##                             Value Std. Error t value
## schoolMS                  0.32605    0.38236  0.8527
## sexM                      0.56589    0.24600  2.3004
## age                      -0.28037    0.10765 -2.6045
## addressU                  0.18757    0.28015  0.6695
## famsizeLE3                0.50097    0.23647  2.1185
## PstatusT                 -0.32472    0.34653 -0.9371
## Medu                      0.06932    0.18259  0.3796
## Fedu                     -0.08955    0.15325 -0.5844
## Mjobhealth                0.60023    0.54994  1.0915
## Mjobother                -0.05752    0.35774 -0.1608
## Mjobservices              0.54672    0.39896  1.3704
## Mjobteacher              -0.61096    0.52074 -1.1732
## Fjobhealth               -0.40733    0.69420 -0.5868
## Fjobother                -0.17855    0.48037 -0.3717
## Fjobservices             -0.05255    0.49757 -0.1056
## Fjobteacher               0.46618    0.66467  0.7014
## reasonhome                0.23177    0.26518  0.8740
## reasonother               0.25966    0.38080  0.6819
## reasonreputation          0.39833    0.28845  1.3809
## guardianmother            0.04580    0.26486  0.1729
## guardianother             0.46437    0.48249  0.9624
## traveltime               -0.14252    0.17642 -0.8079
## studytime                 0.37461    0.14709  2.5467
## failures                 -0.40709    0.28580 -1.4244
## schoolsupyes             -1.07299    0.33629 -3.1907
## famsupyes                -0.49298    0.54604 -0.9028
## paidyes                   0.20052    0.23526  0.8523
## activitiesyes            -0.11540    0.21736 -0.5309
## nurseryyes               -0.17096    0.26950 -0.6343
## higheryes                -0.15919    0.49945 -0.3187
## internetyes               0.39537    0.44970  0.8792
## romanticyes              -0.50355    0.23510 -2.1419
## famrellow                 0.05791    0.26647  0.2173
## freetimelow              -0.67456    0.23496 -2.8709
## gooutlow                  0.77564    0.24852  3.1210
## Dalclow                   0.07991    0.59539  0.1342
## Walclow                   0.25638    0.32121  0.7982
## healthlow                 0.29549    0.21720  1.3604
## absences                  0.02002    0.02009  0.9966
## first_gen_collegeyes     -0.70199    0.38869 -1.8060
## stable_learning_envyes   -0.16982    0.59977 -0.2831
## high_freq_absentyes      -0.16074    0.38422 -0.4184
## failedyes                -0.91443    0.54086 -1.6907
```

```
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -7.1685  2.3463    -3.0552
## 4|5   -7.1297  2.3461    -3.0390
## 5|6   -6.9426  2.3448    -2.9609
## 6|7   -6.5614  2.3424    -2.8012
## 7|8   -6.3439  2.3413    -2.7096
## 8|9   -5.7195  2.3376    -2.4467
## 9|10  -5.3221  2.3345    -2.2798
## 10|11 -4.6194  2.3278    -1.9844
## 11|12 -3.9471  2.3222    -1.6997
## 12|13 -3.5499  2.3203    -1.5299
## 13|14 -3.0007  2.3201    -1.2933
## 14|15 -2.4425  2.3202    -1.0527
## 15|16 -1.6383  2.3195    -0.7063
## 16|17 -1.0749  2.3222    -0.4629
## 17|18 -0.7489  2.3251    -0.3221
## 18|19  0.2523  2.3411     0.1078
## 19|20  2.1350  2.5097     0.8507
##
## Residual Deviance: 1513.116
## AIC: 1633.116
```

```
acc.ord <- predict(mod, training)
ctable <- table(training$G3, acc.ord)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 20.89
```

```
ctable
```

```
##     acc.ord
##       0  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   0  18  0  0  0  0  0  0 10  5  0  0  0  0  0  0  0  0  0
##   4   0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
##   5   3  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0
##   6   1  0  0  0  0  0  0  9  1  0  0  0  1  0  0  0  0  0
##   7   4  0  0  0  0  0  0  1  3  0  0  0  0  0  0  0  0  0
##   8  11  0  0  0  0  0  0  8  7  0  0  0  1  0  0  0  0  0
##   9   6  0  0  0  0  0  0  8  6  0  0  0  0  0  0  0  0  0
##   10  8  0  0  0  0  0  0 14 12  0  0  0  5  0  0  0  0  0
##   11  2  0  0  0  0  0  0 12 20  0  1  0  5  0  0  0  0  0
##   12  3  0  0  0  0  0  0  3 12  0  0  0  5  0  0  0  0  0
##   13  5  0  0  0  0  0  0  2 17  0  0  0  4  0  0  0  0  0
##   14  1  0  0  0  0  0  0  2 12  0  0  0  8  0  0  0  0  0
##   15  0  0  0  0  0  0  0  3  7  0  1  0 14  0  0  0  0  0
##   16  0  0  0  0  0  0  0  1  4  0  0  0  7  0  0  0  0  0
##   17  0  0  0  0  0  0  0  0  3  0  0  0  2  0  0  0  0  0
##   18  0  0  0  0  0  0  0  1  4  0  0  0  4  0  0  0  0  0
##   19  0  0  0  0  0  0  0  0  1  0  0  0  4  0  0  0  0  0
##   20  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
```

```
mod1 <- polr(ord_g3 ~ failed + high_freq_absent + romantic + internet + goout + first_gen_college + Wal
summary(mod1)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = ord_g3 ~ failed + high_freq_absent + romantic +
##     internet + goout + first_gen_college + Walc + sex + schoolsup +
##     famsup + absences + studytime + higher, data = training)
##
## Coefficients:
##                          Value Std. Error t value
## failedyes             -1.34171    0.27018 -4.9659
## high_freq_absentyes    0.03565    0.34970  0.1019
## romanticyes           -0.45766    0.22322 -2.0502
## internetyes            0.36012    0.27651  1.3024
## gooutlow               0.58461    0.22879  2.5552
## first_gen_collegeyes -0.66042    0.22335 -2.9568
## Walclow                0.33622    0.27525  1.2215
## sexM                   0.55216    0.22206  2.4865
## schoolsupyes          -0.69882    0.28905 -2.4177
## famsupyes             -0.48594    0.21594 -2.2503
## absences               0.01048    0.01726  0.6071
## studytime              0.25616    0.13427  1.9078
## higheryes              0.37412    0.45201  0.8277
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -1.6658  0.6572    -2.5347
## 4|5   -1.6293  0.6566    -2.4816
## 5|6   -1.4556  0.6534    -2.2276
## 6|7   -1.1027  0.6485    -1.7004
## 7|8   -0.9014  0.6467    -1.3939
## 8|9   -0.3219  0.6437    -0.5000
## 9|10   0.0449  0.6427     0.0698
## 10|11  0.6847  0.6438     1.0635
## 11|12  1.2954  0.6464     2.0040
## 12|13  1.6591  0.6479     2.5610
## 13|14  2.1640  0.6516     3.3212
## 14|15  2.6641  0.6579     4.0496
## 15|16  3.3940  0.6711     5.0577
## 16|17  3.9280  0.6858     5.7276
## 17|18  4.2419  0.6982     6.0754
## 18|19  5.2132  0.7670     6.7971
## 19|20  7.0544  1.1934     5.9112
##
## Residual Deviance: 1562.392
## AIC: 1622.392
```

```
(ctable <- coef(summary(mod1)))
```

```
##
## Re-fitting to get Hessian
```

```
##                            Value Std. Error     t value
## failedyes             -1.34171397 0.27018479 -4.96591237
## high_freq_absentyes    0.03565031 0.34969709  0.10194625
## romanticyes           -0.45766071 0.22322374 -2.05023314
```

```
## internetyes           0.36011512 0.27651008  1.30235803
## gooutlow              0.58461188 0.22879147  2.55521711
## first_gen_collegeyes -0.66042320 0.22335452 -2.95683831
## Walclow               0.33622440 0.27525303  1.22151026
## sexM                  0.55215988 0.22206496  2.48647905
## schoolsupyes         -0.69881781 0.28904608 -2.41766920
## famsupyes            -0.48593653 0.21594151 -2.25031553
## absences              0.01047719 0.01725748  0.60710999
## studytime             0.25615867 0.13426713  1.90782854
## higheryes             0.37411910 0.45201352  0.82767238
## 0|4                  -1.66581605 0.65721302 -2.53466683
## 4|5                  -1.62930927 0.65655494 -2.48160386
## 5|6                  -1.45558081 0.65342826 -2.22760613
## 6|7                  -1.10268571 0.64846971 -1.70044290
## 7|8                  -0.90144383 0.64670148 -1.39391027
## 8|9                  -0.32186448 0.64369588 -0.50002569
## 9|10                  0.04485165 0.64272846  0.06978319
## 10|11                 0.68472806 0.64384467  1.06349884
## 11|12                 1.29538659 0.64638677  2.00404255
## 12|13                 1.65914607 0.64785012  2.56100295
## 13|14                 2.16402771 0.65158608  3.32116935
## 14|15                 2.66414639 0.65787988  4.04959395
## 15|16                 3.39401724 0.67106437  5.05766269
## 16|17                 3.92801161 0.68580175  5.72761970
## 17|18                 4.24194365 0.69821069  6.07544927
## 18|19                 5.21317802 0.76696900  6.79711702
## 19|20                 7.05442354 1.19339185  5.91123825
```

Calculate and store p-values:

```
p1 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p1))
```

```
##                           Value Std. Error     t value      p value
## failedyes            -1.34171397 0.27018479 -4.96591237 6.837883e-07
## high_freq_absentyes   0.03565031 0.34969709  0.10194625 9.187993e-01
## romanticyes          -0.45766071 0.22322374 -2.05023314 4.034169e-02
## internetyes           0.36011512 0.27651008  1.30235803 1.927940e-01
## gooutlow              0.58461188 0.22879147  2.55521711 1.061216e-02
## first_gen_collegeyes -0.66042320 0.22335452 -2.95683831 3.108111e-03
## Walclow               0.33622440 0.27525303  1.22151026 2.218929e-01
## sexM                  0.55215988 0.22206496  2.48647905 1.290142e-02
## schoolsupyes         -0.69881781 0.28904608 -2.41766920 1.562027e-02
## famsupyes            -0.48593653 0.21594151 -2.25031553 2.442892e-02
## absences              0.01047719 0.01725748  0.60710999 5.437779e-01
## studytime             0.25615867 0.13426713  1.90782854 5.641338e-02
## higheryes             0.37411910 0.45201352  0.82767238 4.078561e-01
## 0|4                  -1.66581605 0.65721302 -2.53466683 1.125543e-02
## 4|5                  -1.62930927 0.65655494 -2.48160386 1.307926e-02
## 5|6                  -1.45558081 0.65342826 -2.22760613 2.590679e-02
## 6|7                  -1.10268571 0.64846971 -1.70044290 8.904765e-02
## 7|8                  -0.90144383 0.64670148 -1.39391027 1.633447e-01
## 8|9                  -0.32186448 0.64369588 -0.50002569 6.170570e-01
## 9|10                  0.04485165 0.64272846  0.06978319 9.443662e-01
## 10|11                 0.68472806 0.64384467  1.06349884 2.875558e-01
```

```
## 11|12                1.29538659 0.64638677  2.00404255 4.506550e-02
## 12|13                1.65914607 0.64785012  2.56100295 1.043705e-02
## 13|14                2.16402771 0.65158608  3.32116935 8.964113e-04
## 14|15                2.66414639 0.65787988  4.04959395 5.130658e-05
## 15|16                3.39401724 0.67106437  5.05766269 4.244263e-07
## 16|17                3.92801161 0.68580175  5.72761970 1.018496e-08
## 17|18                4.24194365 0.69821069  6.07544927 1.236411e-09
## 18|19                5.21317802 0.76696900  6.79711702 1.067334e-11
## 19|20                7.05442354 1.19339185  5.91123825 3.395455e-09
```

Confidence intervals for parameter estimates:

```
(ci1 <- confint(mod1))
```

```
## Waiting for profiling to be done...
```

```
##
## Re-fitting to get Hessian
```

```
##                              2.5 %      97.5 %
## failedyes            -1.876197497 -0.81556618
## high_freq_absentyes  -0.656411452  0.71764477
## romanticyes          -0.897663611 -0.02188466
## internetyes          -0.180649012  0.90497042
## gooutlow              0.137337761  1.03514595
## first_gen_collegeyes -1.100400976 -0.22416437
## Walclow              -0.203641426  0.87718770
## sexM                  0.118264255  0.98945878
## schoolsupyes         -1.268171885 -0.13267874
## famsupyes            -0.910849856 -0.06369438
## absences             -0.021851056  0.04656485
## studytime            -0.006577242  0.52033547
## higheryes            -0.505211121  1.27490555
```

Analyzing the p-values and confidence intervals allows us to determine whether the coefficient estimates are significant. Based on these, failed, romantic, goout, first_gen_college, sex, schoolsup, famsup, studytime seem to be active. (Studytime is dubious, but we will include it in the next model)

Refitting a model with these predictors:

```
mod2 <- polr(ord_g3 ~ failed + romantic + goout + first_gen_college + studytime + sex + schoolsup + fams
summary(mod2)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = ord_g3 ~ failed + romantic + goout + first_gen_college +
##     studytime + sex + schoolsup + famsup, data = training)
##
## Coefficients:
##                       Value Std. Error t value
## failedyes            -1.3882     0.2606  -5.327
## romanticyes          -0.4157     0.2160  -1.924
## gooutlow              0.6874     0.2119   3.245
## first_gen_collegeyes -0.6953     0.2189  -3.177
## studytime             0.2749     0.1333   2.062
## sexM                  0.4689     0.2127   2.205
```

```
## schoolsupyes         -0.6642     0.2881  -2.305
## famsupyes            -0.4203     0.2133  -1.970
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -2.5429  0.4453    -5.7101
## 4|5   -2.5068  0.4443    -5.6425
## 5|6   -2.3344  0.4396    -5.3103
## 6|7   -1.9834  0.4324    -4.5866
## 7|8   -1.7842  0.4294    -4.1546
## 8|9   -1.2081  0.4228    -2.8573
## 9|10  -0.8412  0.4200    -2.0027
## 10|11 -0.2072  0.4175    -0.4963
## 11|12  0.3981  0.4182     0.9519
## 12|13  0.7604  0.4192     1.8142
## 13|14  1.2610  0.4225     2.9847
## 14|15  1.7560  0.4295     4.0886
## 15|16  2.4801  0.4464     5.5558
## 16|17  3.0130  0.4678     6.4414
## 17|18  3.3261  0.4857     6.8484
## 18|19  4.2924  0.5793     7.4098
## 19|20  6.1276  1.0817     5.6650
##
## Residual Deviance: 1567.543
## AIC: 1617.543
```

```
(ctable <- coef(summary(mod2)))
```

```
##
## Re-fitting to get Hessian
```

```
##                           Value Std. Error    t value
## failedyes            -1.3881639  0.2605997 -5.3268047
## romanticyes          -0.4156752  0.2160358 -1.9241039
## gooutlow              0.6873737  0.2118547  3.2445526
## first_gen_collegeyes -0.6953188  0.2188570 -3.1770460
## studytime             0.2748992  0.1332997  2.0622646
## sexM                  0.4689357  0.2126871  2.2048153
## schoolsupyes         -0.6642084  0.2881418 -2.3051443
## famsupyes            -0.4203058  0.2133383 -1.9701374
## 0|4                  -2.5429074  0.4453387 -5.7100521
## 4|5                  -2.5067847  0.4442706 -5.6424726
## 5|6                  -2.3344286  0.4396061 -5.3102731
## 6|7                  -1.9834475  0.4324423 -4.5866175
## 7|8                  -1.7841536  0.4294407 -4.1545983
## 8|9                  -1.2081500  0.4228275 -2.8573116
## 9|10                 -0.8412191  0.4200487 -2.0026706
## 10|11                -0.2072168  0.4175067 -0.4963196
## 11|12                 0.3980517  0.4181620  0.9519078
## 12|13                 0.7604341  0.4191666  1.8141570
## 13|14                 1.2609616  0.4224815  2.9846550
## 14|15                 1.7559528  0.4294754  4.0885992
## 15|16                 2.4801341  0.4464006  5.5558481
## 16|17                 3.0129643  0.4677514  6.4413788
## 17|18                 3.3260960  0.4856779  6.8483581
```

```
## 18|19                 4.2923566  0.5792787  7.4098294
## 19|20                 6.1275584  1.0816527  5.6649962
```

```
p2 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p2))
```

```
##                           Value Std. Error    t value      p value
## failedyes            -1.3881639  0.2605997 -5.3268047 9.995554e-08
## romanticyes          -0.4156752  0.2160358 -1.9241039 5.434156e-02
## gooutlow              0.6873737  0.2118547  3.2445526 1.176353e-03
## first_gen_collegeyes -0.6953188  0.2188570 -3.1770460 1.487835e-03
## studytime             0.2748992  0.1332997  2.0622646 3.918255e-02
## sexM                  0.4689357  0.2126871  2.2048153 2.746706e-02
## schoolsupyes         -0.6642084  0.2881418 -2.3051443 2.115849e-02
## famsupyes            -0.4203058  0.2133383 -1.9701374 4.882263e-02
## 0|4                  -2.5429074  0.4453387 -5.7100521 1.129416e-08
## 4|5                  -2.5067847  0.4442706 -5.6424726 1.676252e-08
## 5|6                  -2.3344286  0.4396061 -5.3102731 1.094611e-07
## 6|7                  -1.9834475  0.4324423 -4.5866175 4.504849e-06
## 7|8                  -1.7841536  0.4294407 -4.1545983 3.258595e-05
## 8|9                  -1.2081500  0.4228275 -2.8573116 4.272462e-03
## 9|10                 -0.8412191  0.4200487 -2.0026706 4.521266e-02
## 10|11                -0.2072168  0.4175067 -0.4963196 6.196689e-01
## 11|12                 0.3980517  0.4181620  0.9519078 3.411438e-01
## 12|13                 0.7604341  0.4191666  1.8141570 6.965355e-02
## 13|14                 1.2609616  0.4224815  2.9846550 2.838983e-03
## 14|15                 1.7559528  0.4294754  4.0885992 4.339860e-05
## 15|16                 2.4801341  0.4464006  5.5558481 2.762670e-08
## 16|17                 3.0129643  0.4677514  6.4413788 1.183930e-10
## 17|18                 3.3260960  0.4856779  6.8483581 7.470236e-12
## 18|19                 4.2923566  0.5792787  7.4098294 1.264620e-13
## 19|20                 6.1275584  1.0816527  5.6649962 1.470278e-08
```

```
(ci2 <- confint(mod2))
```

```
## Waiting for profiling to be done...
##
## Re-fitting to get Hessian
```

```
##                             2.5 %       97.5 %
## failedyes            -1.90462685 -0.881708189
## romanticyes          -0.84097272  0.006566793
## gooutlow              0.27365841  1.104747592
## first_gen_collegeyes -1.12660028 -0.268012697
## studytime             0.01411035  0.537164395
## sexM                  0.05297154  0.887306542
## schoolsupyes         -1.23199282 -0.100060789
## famsupyes            -0.83978194 -0.002875816
```

AIC has decreased.

Based on the p-values and confidence intervals, romantic does not seem to be significant. Let's try excluding it.

Pared-down model again:

```
mod3 <- polr(ord_g3 ~ failed + goout + first_gen_college + sex + schoolsup + studytime, data = training
summary(mod3)
```

```
## Call:
## polr(formula = ord_g3 ~ failed + goout + first_gen_college +
##     sex + schoolsup + studytime, data = training, Hess = TRUE)
##
## Coefficients:
##                         Value Std. Error t value
## failedyes             -1.4470     0.2594  -5.577
## gooutlow               0.6862     0.2115   3.244
## first_gen_collegeyes -0.5623     0.2119  -2.654
## sexM                   0.5365     0.2106   2.547
## schoolsupyes          -0.6138     0.2822  -2.175
## studytime              0.2189     0.1311   1.670
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -2.1140  0.4116    -5.1354
## 4|5   -2.0782  0.4106    -5.0620
## 5|6   -1.9083  0.4060    -4.7005
## 6|7   -1.5631  0.3995    -3.9125
## 7|8   -1.3668  0.3969    -3.4435
## 8|9   -0.7982  0.3913    -2.0401
## 9|10  -0.4358  0.3892    -1.1199
## 10|11  0.1921  0.3879     0.4952
## 11|12  0.7943  0.3898     2.0379
## 12|13  1.1538  0.3918     2.9448
## 13|14  1.6485  0.3967     4.1551
## 14|15  2.1387  0.4054     5.2759
## 15|16  2.8588  0.4247     6.7305
## 16|17  3.3897  0.4478     7.5696
## 17|18  3.7015  0.4667     7.9310
## 18|19  4.6606  0.5642     8.2598
## 19|20  6.4828  1.0738     6.0371
##
## Residual Deviance: 1574.549
## AIC: 1620.549
```

```
(ctable <- coef(summary(mod3)))
```

```
##                           Value Std. Error    t value
## failedyes             -1.4470044  0.2594410 -5.5773921
## gooutlow               0.6862095  0.2115274  3.2440692
## first_gen_collegeyes -0.5623425  0.2119141 -2.6536341
## sexM                   0.5364682  0.2106266  2.5470106
## schoolsupyes          -0.6138372  0.2821635 -2.1754659
## studytime              0.2188984  0.1310851  1.6698957
## 0|4                   -2.1139669  0.4116495 -5.1353568
## 4|5                   -2.0782289  0.4105540 -5.0620109
## 5|6                   -1.9083008  0.4059759 -4.7005269
## 6|7                   -1.5630672  0.3995011 -3.9125480
## 7|8                   -1.3667896  0.3969204 -3.4434854
## 8|9                   -0.7982483  0.3912773 -2.0401089
## 9|10                  -0.4358495  0.3891718 -1.1199410
## 10|11                  0.1920664  0.3878675  0.4951856
## 11|12                  0.7942916  0.3897682  2.0378563
## 12|13                  1.1537589  0.3917941  2.9448095
```

```
## 13|14                 1.6485090  0.3967448  4.1550867
## 14|15                 2.1387107  0.4053722  5.2759186
## 15|16                 2.8587993  0.4247500  6.7305461
## 16|17                 3.3897018  0.4478022  7.5696401
## 17|18                 3.7014831  0.4667133  7.9309568
## 18|19                 4.6605957  0.5642498  8.2598092
## 19|20                 6.4827907  1.0738176  6.0371431
```

```
p3 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p3))
```

```
##                            Value Std. Error    t value      p value
## failedyes             -1.4470044  0.2594410 -5.5773921 2.441512e-08
## gooutlow               0.6862095  0.2115274  3.2440692 1.178351e-03
## first_gen_collegeyes -0.5623425  0.2119141 -2.6536341 7.963013e-03
## sexM                   0.5364682  0.2106266  2.5470106 1.086501e-02
## schoolsupyes          -0.6138372  0.2821635 -2.1754659 2.959522e-02
## studytime              0.2188984  0.1310851  1.6698957 9.494000e-02
## 0|4                   -2.1139669  0.4116495 -5.1353568 2.816093e-07
## 4|5                   -2.0782289  0.4105540 -5.0620109 4.148574e-07
## 5|6                   -1.9083008  0.4059759 -4.7005269 2.594911e-06
## 6|7                   -1.5630672  0.3995011 -3.9125480 9.132736e-05
## 7|8                   -1.3667896  0.3969204 -3.4434854 5.742675e-04
## 8|9                   -0.7982483  0.3912773 -2.0401089 4.133948e-02
## 9|10                  -0.4358495  0.3891718 -1.1199410 2.627389e-01
## 10|11                  0.1920664  0.3878675  0.4951856 6.204691e-01
## 11|12                  0.7942916  0.3897682  2.0378563 4.156431e-02
## 12|13                  1.1537589  0.3917941  2.9448095 3.231536e-03
## 13|14                  1.6485090  0.3967448  4.1550867 3.251642e-05
## 14|15                  2.1387107  0.4053722  5.2759186 1.320927e-07
## 15|16                  2.8587993  0.4247500  6.7305461 1.690275e-11
## 16|17                  3.3897018  0.4478022  7.5696401 3.742596e-14
## 17|18                  3.7014831  0.4667133  7.9309568 2.174638e-15
## 18|19                  4.6605957  0.5642498  8.2598092 1.459055e-16
## 19|20                  6.4827907  1.0738176  6.0371431 1.568666e-09
```

```
(ci3 <- confint(mod3))
```

```
## Waiting for profiling to be done...
```

```
##                              2.5 %      97.5 %
## failedyes             -1.96148845 -0.94318819
## gooutlow               0.27307924  1.10288470
## first_gen_collegeyes -0.97950820 -0.14817187
## sexM                   0.12485281  0.95095558
## schoolsupyes          -1.16913473 -0.06070744
## studytime             -0.03766773  0.47670960
```

All predictors are significant, but AIC has increased compared to mod2.

Evaluating accuracy of the model for the training set:

```
acc.ord3 <- predict(mod3, training)
ctable <- table(training$G3, acc.ord3)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 17.09
```

```
ctable
```

```
##     acc.ord3
##       0  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   0  16  0  0  0  0  0  0  7  7  0  0  0  3  0  0  0  0  0
##   4   1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   5   2  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0
##   6   1  0  0  0  0  0  0  6  5  0  0  0  0  0  0  0  0  0
##   7   5  0  0  0  0  0  0  0  1  0  0  0  2  0  0  0  0  0
##   8  12  0  0  0  0  0  0  4  9  0  0  0  2  0  0  0  0  0
##   9   6  0  0  0  0  0  0  4  9  0  0  0  1  0  0  0  0  0
##   10  8  0  0  0  0  0  0  6 20  0  0  0  5  0  0  0  0  0
##   11  4  0  0  0  0  0  0  8 23  0  0  0  5  0  0  0  0  0
##   12  3  0  0  0  0  0  0  5 11  0  0  0  4  0  0  0  0  0
##   13  5  0  0  0  0  0  0  3 17  0  0  0  3  0  0  0  0  0
##   14  1  0  0  0  0  0  0  3  8  0  0  0 11  0  0  0  0  0
##   15  0  0  0  0  0  0  0  2 14  0  0  0  9  0  0  0  0  0
##   16  0  0  0  0  0  0  0  2  6  0  0  0  4  0  0  0  0  0
##   17  0  0  0  0  0  0  0  1  3  0  0  0  1  0  0  0  0  0
##   18  1  0  0  0  0  0  0  1  5  0  0  0  2  0  0  0  0  0
##   19  0  0  0  0  0  0  0  0  3  0  0  0  2  0  0  0  0  0
##   20  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
```

Very terrible accuracy even for the training set.

What if we add interaction terms?

Let's base our interaction terms on the discussion for the linear model.

```
mod4 <- polr(ord_g3 ~ failed + goout + romantic +  first_gen_college + sex + schoolsup + sex*schoolsup +
summary(mod4)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = ord_g3 ~ failed + goout + romantic + first_gen_college +
##     sex + schoolsup + sex * schoolsup + sex * first_gen_college +
##     schoolsup * failed + schoolsup * studytime + schoolsup *
##     first_gen_college + studytime * famsup, data = training)
##
## Coefficients:
##                                      Value Std. Error  t value
## failedyes                          -1.72945     0.2981 -5.80124
## gooutlow                            0.67401     0.2169  3.10806
## romanticyes                        -0.48910     0.2204 -2.21930
## first_gen_collegeyes               -1.43664     0.3180 -4.51738
## sexM                               -0.01446     0.3391 -0.04265
## schoolsupyes                        0.73242     0.9305  0.78712
## studytime                           0.36175     0.2487  1.45475
## famsupyes                          -0.76421     0.5894 -1.29664
## sexM:schoolsupyes                  -1.06747     0.6217 -1.71701
## first_gen_collegeyes:sexM           1.06194     0.4232  2.50904
## failedyes:schoolsupyes              1.15774     0.6593  1.75607
## schoolsupyes:studytime             -1.11414     0.3374 -3.30178
## first_gen_collegeyes:schoolsupyes   1.59232     0.6093  2.61341
## studytime:famsupyes                 0.14783     0.2850  0.51876
```

```
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -3.0101  0.5949    -5.0599
## 4|5   -2.9717  0.5939    -5.0034
## 5|6   -2.7903  0.5898    -4.7309
## 6|7   -2.4230  0.5833    -4.1537
## 7|8   -2.2159  0.5806    -3.8163
## 8|9   -1.6123  0.5757    -2.8006
## 9|10  -1.2173  0.5741    -2.1202
## 10|11 -0.5423  0.5738    -0.9451
## 11|12  0.0965  0.5744     0.1680
## 12|13  0.4842  0.5739     0.8438
## 13|14  1.0192  0.5741     1.7752
## 14|15  1.5345  0.5779     2.6551
## 15|16  2.2702  0.5892     3.8531
## 16|17  2.8082  0.6045     4.6455
## 17|18  3.1241  0.6180     5.0554
## 18|19  4.0937  0.6937     5.9014
## 19|20  5.9292  1.1498     5.1569
##
## Residual Deviance: 1537.762
## AIC: 1599.762
```

```
(ctable <- coef(summary(mod4)))
```

```
##
## Re-fitting to get Hessian
```

```
##                                       Value Std. Error     t value
## failedyes                        -1.72945368  0.2981178 -5.80124295
## gooutlow                          0.67400726  0.2168580  3.10805786
## romanticyes                      -0.48909643  0.2203829 -2.21930313
## first_gen_collegeyes             -1.43663982  0.3180252 -4.51737778
## sexM                             -0.01446239  0.3390652 -0.04265371
## schoolsupyes                      0.73241720  0.9305008  0.78712151
## studytime                         0.36174519  0.2486653  1.45474731
## famsupyes                        -0.76421191  0.5893770 -1.29664348
## sexM:schoolsupyes                -1.06746616  0.6217004 -1.71701055
## first_gen_collegeyes:sexM         1.06194139  0.4232461  2.50904014
## failedyes:schoolsupyes            1.15773926  0.6592797  1.75606684
## schoolsupyes:studytime           -1.11413949  0.3374358 -3.30178248
## first_gen_collegeyes:schoolsupyes 1.59232262  0.6092889  2.61341153
## studytime:famsupyes               0.14782789  0.2849627  0.51876227
## 0|4                              -3.01008158  0.5948850 -5.05993887
## 4|5                              -2.97166463  0.5939322 -5.00337322
## 5|6                              -2.79029263  0.5898075 -4.73085276
## 6|7                              -2.42295680  0.5833207 -4.15373034
## 7|8                              -2.21585268  0.5806252 -3.81632165
## 8|9                              -1.61234628  0.5757246 -2.80055115
## 9|10                             -1.21732081  0.5741468 -2.12022571
## 10|11                            -0.54231022  0.5738069 -0.94510930
## 11|12                             0.09649365  0.5743725  0.16799839
## 12|13                             0.48423331  0.5738879  0.84377688
## 13|14                             1.01921806  0.5741418  1.77520259
```

```
## 14|15                                1.53449674  0.5779426  2.65510251
## 15|16                                2.27018077  0.5891862  3.85307839
## 16|17                                2.80819116  0.6045010  4.64546978
## 17|18                                3.12405658  0.6179583  5.05544881
## 18|19                                4.09365582  0.6936810  5.90135194
## 19|20                                5.92920227  1.1497705  5.15685736
```

```
p4 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p4))
```

```
##                                       Value Std. Error     t value
## failedyes                        -1.72945368  0.2981178 -5.80124295
## gooutlow                          0.67400726  0.2168580  3.10805786
## romanticyes                      -0.48909643  0.2203829 -2.21930313
## first_gen_collegeyes             -1.43663982  0.3180252 -4.51737778
## sexM                             -0.01446239  0.3390652 -0.04265371
## schoolsupyes                      0.73241720  0.9305008  0.78712151
## studytime                         0.36174519  0.2486653  1.45474731
## famsupyes                        -0.76421191  0.5893770 -1.29664348
## sexM:schoolsupyes                -1.06746616  0.6217004 -1.71701055
## first_gen_collegeyes:sexM         1.06194139  0.4232461  2.50904014
## failedyes:schoolsupyes            1.15773926  0.6592797  1.75606684
## schoolsupyes:studytime           -1.11413949  0.3374358 -3.30178248
## first_gen_collegeyes:schoolsupyes  1.59232262  0.6092889  2.61341153
## studytime:famsupyes               0.14782789  0.2849627  0.51876227
## 0|4                              -3.01008158  0.5948850 -5.05993887
## 4|5                              -2.97166463  0.5939322 -5.00337322
## 5|6                              -2.79029263  0.5898075 -4.73085276
## 6|7                              -2.42295680  0.5833207 -4.15373034
## 7|8                              -2.21585268  0.5806252 -3.81632165
## 8|9                              -1.61234628  0.5757246 -2.80055115
## 9|10                             -1.21732081  0.5741468 -2.12022571
## 10|11                            -0.54231022  0.5738069 -0.94510930
## 11|12                             0.09649365  0.5743725  0.16799839
## 12|13                             0.48423331  0.5738879  0.84377688
## 13|14                             1.01921806  0.5741418  1.77520259
## 14|15                             1.53449674  0.5779426  2.65510251
## 15|16                             2.27018077  0.5891862  3.85307839
## 16|17                             2.80819116  0.6045010  4.64546978
## 17|18                             3.12405658  0.6179583  5.05544881
## 18|19                             4.09365582  0.6936810  5.90135194
## 19|20                             5.92920227  1.1497705  5.15685736
##                                      p value
## failedyes                        6.582515e-09
## gooutlow                         1.883212e-03
## romanticyes                      2.646611e-02
## first_gen_collegeyes             6.261014e-06
## sexM                             9.659776e-01
## schoolsupyes                     4.312107e-01
## studytime                        1.457392e-01
## famsupyes                        1.947539e-01
## sexM:schoolsupyes                8.597724e-02
## first_gen_collegeyes:sexM        1.210597e-02
## failedyes:schoolsupyes           7.907700e-02
## schoolsupyes:studytime           9.607254e-04
```

```
## first_gen_collegeyes:schoolsupyes 8.964329e-03
## studytime:famsupyes               6.039265e-01
## 0|4                               4.193909e-07
## 4|5                               5.633572e-07
## 5|6                               2.235787e-06
## 6|7                               3.270986e-05
## 7|8                               1.354559e-04
## 8|9                               5.101542e-03
## 9|10                              3.398702e-02
## 10|11                             3.446031e-01
## 11|12                             8.665845e-01
## 12|13                             3.987941e-01
## 13|14                             7.586444e-02
## 14|15                             7.928431e-03
## 15|16                             1.166420e-04
## 16|17                             3.393034e-06
## 17|18                             4.293798e-07
## 18|19                             3.605349e-09
## 19|20                             2.511290e-07
```

```
(ci4 <- confint(mod4))
```

```
## Waiting for profiling to be done...
##
## Re-fitting to get Hessian
```

```
##                                       2.5 %      97.5 %
## failedyes                        -2.3214364 -1.15123396
## gooutlow                          0.2503920  1.10114465
## romanticyes                      -0.9230151 -0.05837766
## first_gen_collegeyes             -2.0646870 -0.81670972
## sexM                             -0.6799620  0.65071516
## schoolsupyes                     -1.0998497  2.56413384
## studytime                        -0.1227550  0.85203523
## famsupyes                        -1.9233145  0.38932033
## sexM:schoolsupyes                -2.2962304  0.15377624
## first_gen_collegeyes:sexM         0.2363092  1.89633281
## failedyes:schoolsupyes           -0.1337302  2.46425196
## schoolsupyes:studytime           -1.7820611 -0.45214925
## first_gen_collegeyes:schoolsupyes  0.4009686  2.79629741
## studytime:famsupyes              -0.4093914  0.70866271
```

AIC has decreased significantly compared to the previous models without interaction terms, by nearly 20. However, in this model, sex, its interaction with schoolsup, and its interaction with first_gen_college all seem to be insignificant. The interaction between studytime and famsup and failed and schoolsup do not seem significant either, so let us remove it to pare down the model:

```
mod5 <- polr(ord_g3 ~ failed + goout + romantic + schoolsup + first_gen_college + schoolsup * studytime
summary(mod5)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = ord_g3 ~ failed + goout + romantic + schoolsup +
##     first_gen_college + schoolsup * studytime + schoolsup * first_gen_college,
##     data = training)
```

```
##
## Coefficients:
##                                    Value Std. Error  t value
## failedyes                        -1.39146     0.2622 -5.30667
## gooutlow                          0.59649     0.2134  2.79476
## romanticyes                      -0.50391     0.2163 -2.32943
## schoolsupyes                      0.03364     0.8300  0.04052
## first_gen_collegeyes             -0.85007     0.2317 -3.66947
## studytime                         0.29305     0.1393  2.10388
## schoolsupyes:studytime           -0.85318     0.3244 -2.62985
## schoolsupyes:first_gen_collegeyes  1.52426     0.5720  2.66499
##
## Intercepts:
##       Value   Std. Error t value
## 0|4   -2.6921  0.4042    -6.6607
## 4|5   -2.6558  0.4029    -6.5914
## 5|6   -2.4839  0.3975    -6.2485
## 6|7   -2.1376  0.3891    -5.4942
## 7|8   -1.9409  0.3852    -5.0386
## 8|9   -1.3590  0.3772    -3.6032
## 9|10  -0.9824  0.3739    -2.6277
## 10|11 -0.3366  0.3707    -0.9081
## 11|12  0.2744  0.3708     0.7400
## 12|13  0.6418  0.3719     1.7259
## 13|14  1.1514  0.3752     3.0688
## 14|15  1.6489  0.3824     4.3124
## 15|16  2.3707  0.4009     5.9140
## 16|17  2.9048  0.4245     6.8436
## 17|18  3.2188  0.4440     7.2496
## 18|19  4.1810  0.5450     7.6720
## 19|20  6.0069  1.0636     5.6479
##
## Residual Deviance: 1563.245
## AIC: 1613.245
```

```
(ctable <- coef(summary(mod5)))
```

```
##
## Re-fitting to get Hessian
##                                         Value Std. Error     t value
## failedyes                        -1.39146287  0.2622101 -5.30667226
## gooutlow                          0.59648735  0.2134306  2.79475999
## romanticyes                      -0.50390805  0.2163229 -2.32942552
## schoolsupyes                      0.03363511  0.8300358  0.04052249
## first_gen_collegeyes             -0.85006782  0.2316595 -3.66947101
## studytime                         0.29304592  0.1392885  2.10387705
## schoolsupyes:studytime           -0.85318089  0.3244225 -2.62984500
## schoolsupyes:first_gen_collegeyes  1.52425650  0.5719558  2.66498990
## 0|4                              -2.69208825  0.4041720 -6.66074948
## 4|5                              -2.65581637  0.4029201 -6.59142141
## 5|6                              -2.48392676  0.3975209 -6.24854363
## 6|7                              -2.13758424  0.3890606 -5.49421935
## 7|8                              -1.94093740  0.3852119 -5.03862266
## 8|9                              -1.35895482  0.3771508 -3.60321358
```

```
## 9|10                                -0.98243182  0.3738798 -2.62766733
## 10|11                               -0.33664864  0.3707238 -0.90808488
## 11|12                                0.27441039  0.3708471  0.73995555
## 12|13                                0.64181542  0.3718808  1.72586325
## 13|14                                1.15140976  0.3751961  3.06882090
## 14|15                                1.64894914  0.3823753  4.31238440
## 15|16                                2.37070007  0.4008641  5.91397458
## 16|17                                2.90477643  0.4244524  6.84358578
## 17|18                                3.21877485  0.4439939  7.24959282
## 18|19                                4.18096474  0.5449623  7.67202626
## 19|20                                6.00685744  1.0635640  5.64785684
```

```
p5 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p5))
```

```
##                                       Value Std. Error     t value
## failedyes                        -1.39146287  0.2622101 -5.30667226
## gooutlow                          0.59648735  0.2134306  2.79475999
## romanticyes                      -0.50390805  0.2163229 -2.32942552
## schoolsupyes                      0.03363511  0.8300358  0.04052249
## first_gen_collegeyes             -0.85006782  0.2316595 -3.66947101
## studytime                         0.29304592  0.1392885  2.10387705
## schoolsupyes:studytime           -0.85318089  0.3244225 -2.62984500
## schoolsupyes:first_gen_collegeyes  1.52425650  0.5719558  2.66498990
## 0|4                              -2.69208825  0.4041720 -6.66074948
## 4|5                              -2.65581637  0.4029201 -6.59142141
## 5|6                              -2.48392676  0.3975209 -6.24854363
## 6|7                              -2.13758424  0.3890606 -5.49421935
## 7|8                              -1.94093740  0.3852119 -5.03862266
## 8|9                              -1.35895482  0.3771508 -3.60321358
## 9|10                             -0.98243182  0.3738798 -2.62766733
## 10|11                            -0.33664864  0.3707238 -0.90808488
## 11|12                             0.27441039  0.3708471  0.73995555
## 12|13                             0.64181542  0.3718808  1.72586325
## 13|14                             1.15140976  0.3751961  3.06882090
## 14|15                             1.64894914  0.3823753  4.31238440
## 15|16                             2.37070007  0.4008641  5.91397458
## 16|17                             2.90477643  0.4244524  6.84358578
## 17|18                             3.21877485  0.4439939  7.24959282
## 18|19                             4.18096474  0.5449623  7.67202626
## 19|20                             6.00685744  1.0635640  5.64785684
##                                       p value
## failedyes                        1.116447e-07
## gooutlow                         5.193826e-03
## romanticyes                      1.983653e-02
## schoolsupyes                     9.676766e-01
## first_gen_collegeyes             2.430529e-04
## studytime                        3.538917e-02
## schoolsupyes:studytime           8.542381e-03
## schoolsupyes:first_gen_collegeyes 7.699064e-03
## 0|4                              2.724347e-11
## 4|5                              4.356354e-11
## 5|6                              4.142975e-10
## 6|7                              3.924425e-08
## 7|8                              4.688939e-07
```

```
## 8|9                                  3.143071e-04
## 9|10                                 8.597255e-03
## 10|11                                3.638334e-01
## 11|12                                4.593270e-01
## 12|13                                8.437202e-02
## 13|14                                2.149054e-03
## 14|15                                1.615033e-05
## 15|16                                3.339494e-09
## 16|17                                7.723504e-12
## 17|18                                4.180264e-13
## 18|19                                1.693003e-14
## 19|20                                1.624604e-08
```

```
(ci5 <- confint(mod5))
```

```
## Waiting for profiling to be done...
##
## Re-fitting to get Hessian
```

```
##                                           2.5 %      97.5 %
## failedyes                           -1.91099915 -0.88183866
## gooutlow                             0.17940546  1.01668340
## romanticyes                         -0.92995496 -0.08125812
## schoolsupyes                        -1.60642853  1.65956106
## first_gen_collegeyes                -1.30698473 -0.39813983
## studytime                            0.02039631  0.56697874
## schoolsupyes:studytime              -1.49158352 -0.21490640
## schoolsupyes:first_gen_collegeyes    0.40814640  2.65646321
```

This has resulted in an increase in the AIC, which is still lower than the first three models.

Let's check the accuracy of this model with interaction terms:

```
acc.ord4 <- predict(mod4, training)
ctable <- table(training$G3, acc.ord4)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 19.94
```

```
ctable
```

```
##     acc.ord4
##       0  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   0  18  0  0  0  0  0  0 10  3  0  1  0  1  0  0  0  0  0
##   4   1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   5   4  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
##   6   1  0  0  0  0  0  0 10  1  0  0  0  0  0  0  0  0  0
##   7   4  0  0  0  0  0  0  1  1  0  0  0  2  0  0  0  0  0
##   8  14  0  0  0  0  0  0  7  3  0  0  0  3  0  0  0  0  0
##   9   8  0  0  0  0  0  0  4  6  0  2  0  0  0  0  0  0  0
##   10  5  0  0  0  0  0  0 14 10  0  2  0  8  0  0  0  0  0
##   11  5  0  0  0  0  0  0 13 15  0  1  0  6  0  0  0  0  0
##   12  3  0  0  0  0  0  0  7  9  0  1  0  3  0  0  0  0  0
##   13  4  0  0  0  0  0  0  5 10  0  3  0  6  0  0  0  0  0
##   14  0  0  0  0  0  0  0  2  8  0  4  0  9  0  0  0  0  0
##   15  0  0  0  0  0  0  0  3  8  0  1  0 13  0  0  0  0  0
##   16  0  0  0  0  0  0  0  0  9  0  0  0  3  0  0  0  0  0
##   17  0  0  0  0  0  0  0  0  3  0  0  0  2  0  0  0  0  0
```

```
##   18  1  0  0  0  0  0  0  0  4  0  2  0  2  0  0  0  0  0
##   19  0  0  0  0  0  0  0  0  2  0  0  0  3  0  0  0  0  0
##   20  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
```

The accuracy is even lower than mod3, at only 19.94% for the training set.

Checking on testing set:

```
pred.ord3 <- predict(mod3, testing)
ctable <- table(testing$G3, pred.ord3)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 8.86
```

```
pred.ord4 <- predict(mod4, testing)
ctable <- table(testing$G3, pred.ord4)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 11.39
```

```
pred.ord5 <- predict(mod5, testing)
ctable <- table(testing$G3, pred.ord5)
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
## [1] 10.13
```

Accuracy rates are even lower, at 8.86%, 11.39%, and 10.13%.

Highly inaccurate model, not a good fit for the data.

6-category grades modeling

```
set.seed(3)
train_ind <- sample(x = nrow(data), size = 0.8 * nrow(data))
test_ind_neg <- -train_ind
ftrain <- data[train_ind, ]
ftest <- data[test_ind_neg, ]
```

Trying out a multicat ordinal logit on this:

```
mod6 <- polr(cat_g3 ~ failed + goout + romantic + schoolsup + first_gen_college + schoolsup * studytime
summary(mod6)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = cat_g3 ~ failed + goout + romantic + schoolsup +
##     first_gen_college + schoolsup * studytime + schoolsup * first_gen_college,
##     data = ftrain)
##
## Coefficients:
##                                    Value Std. Error  t value
## failedyes                      -1.55916     0.2803 -5.56248
## gooutlow                        0.58202     0.2260  2.57501
## romanticyes                    -0.68049     0.2310 -2.94618
## schoolsupyes                    0.01527     0.8809  0.01734
## first_gen_collegeyes           -0.87441     0.2449 -3.56977
## studytime                       0.31024     0.1444  2.14898
```

```
## schoolsupyes:studytime            -0.89744     0.3472 -2.58483
## schoolsupyes:first_gen_collegeyes  1.78767     0.6142  2.91072
##
## Intercepts:
##                       Value   Std. Error t value
## Poor|Weak            -2.8407  0.4234    -6.7090
## Weak|Sufficient      -1.0846  0.3904    -2.7781
## Sufficient|Good       1.1195  0.3899     2.8716
## Good|Very Good        2.3538  0.4154     5.6661
## Very Good|Excellent   3.2057  0.4575     7.0075
##
## Residual Deviance: 873.2587
## AIC: 899.2587
```

```
(ctable <- coef(summary(mod6)))
```

```
##
## Re-fitting to get Hessian
```

```
##                                       Value Std. Error     t value
## failedyes                        -1.55915962  0.2802994 -5.56247945
## gooutlow                          0.58201779  0.2260254  2.57501100
## romanticyes                      -0.68048602  0.2309721 -2.94618281
## schoolsupyes                      0.01527404  0.8808503  0.01734011
## first_gen_collegeyes             -0.87441246  0.2449495 -3.56976600
## studytime                         0.31023730  0.1443649  2.14898050
## schoolsupyes:studytime           -0.89744352  0.3471959 -2.58483360
## schoolsupyes:first_gen_collegeyes  1.78767301  0.6141683  2.91072192
## Poor|Weak                        -2.84071658  0.4234218 -6.70895259
## Weak|Sufficient                  -1.08462029  0.3904224 -2.77806907
## Sufficient|Good                   1.11949859  0.3898567  2.87156458
## Good|Very Good                    2.35375515  0.4154121  5.66607292
## Very Good|Excellent               3.20565589  0.4574631  7.00746349
```

```
p6 <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p6))
```

```
##                                       Value Std. Error     t value
## failedyes                        -1.55915962  0.2802994 -5.56247945
## gooutlow                          0.58201779  0.2260254  2.57501100
## romanticyes                      -0.68048602  0.2309721 -2.94618281
## schoolsupyes                      0.01527404  0.8808503  0.01734011
## first_gen_collegeyes             -0.87441246  0.2449495 -3.56976600
## studytime                         0.31023730  0.1443649  2.14898050
## schoolsupyes:studytime           -0.89744352  0.3471959 -2.58483360
## schoolsupyes:first_gen_collegeyes  1.78767301  0.6141683  2.91072192
## Poor|Weak                        -2.84071658  0.4234218 -6.70895259
## Weak|Sufficient                  -1.08462029  0.3904224 -2.77806907
## Sufficient|Good                   1.11949859  0.3898567  2.87156458
## Good|Very Good                    2.35375515  0.4154121  5.66607292
## Very Good|Excellent               3.20565589  0.4574631  7.00746349
##                                       p value
## failedyes                        2.659684e-08
## gooutlow                         1.002369e-02
## romanticyes                      3.217222e-03
## schoolsupyes                     9.861653e-01
```

```
## first_gen_collegeyes              3.573003e-04
## studytime                         3.163595e-02
## schoolsupyes:studytime            9.742600e-03
## schoolsupyes:first_gen_collegeyes 3.605948e-03
## Poor|Weak                         1.960263e-11
## Weak|Sufficient                   5.468299e-03
## Sufficient|Good                   4.084453e-03
## Good|Very Good                    1.461074e-08
## Very Good|Excellent               2.426773e-12
```

```
(ci5 <- confint(mod6))
```

```
## Waiting for profiling to be done...
##
## Re-fitting to get Hessian
```

```
##                                       2.5 %     97.5 %
## failedyes                       -2.11603937 -1.0156926
## gooutlow                         0.14100166  1.0278318
## romanticyes                     -1.13625714 -0.2299976
## schoolsupyes                    -1.71834716  1.7492857
## first_gen_collegeyes            -1.35840403 -0.3972182
## studytime                        0.02754372  0.5941023
## schoolsupyes:studytime          -1.58076549 -0.2136889
## schoolsupyes:first_gen_collegeyes  0.58713044  3.0005349
```

```
acc.ord6 <- predict(mod6, ftrain)
ctable <- table(ftrain$cat_g3, acc.ord6)
ctable
```

```
##             acc.ord6
##              Poor Weak Sufficient Good Very Good Excellent
##   Poor          3   10         20    0         0         0
##   Weak          5   22         46    0         0         0
##   Sufficient    2   13        113    2         0         0
##   Good          0    0         47    1         0         0
##   Very Good     0    0         17    0         0         0
##   Excellent     0    0         14    1         0         0
```

Still not very accurate for the training

Random forest:

```
rf.cat<-randomForest(cat_g3~. -G1 -G2 -G3 -ord_g3 -pf -famsup -internet -Medu -Fedu,data = ftrain, mtry
print(rf.cat)
```

```
##
## Call:
##  randomForest(formula = cat_g3 ~ . - G1 - G2 - G3 - ord_g3 - pf -      famsup - internet - Medu - Fe
##                Type of random forest: classification
##                      Number of trees: 50
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 57.91%
## Confusion matrix:
##            Poor Weak Sufficient Good Very Good Excellent class.error
## Poor         15    7          8    2         1         0   0.5454545
## Weak          5   18         48    2         0         0   0.7534247
```

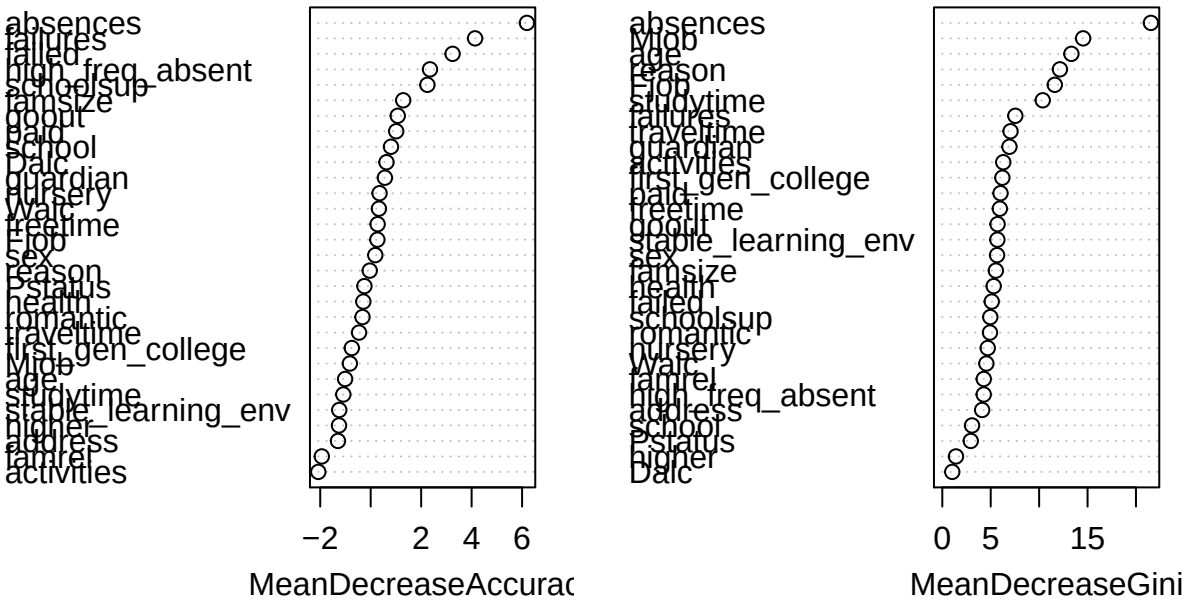
```
## Sufficient    3   29         93    3         2         0   0.2846154
## Good          4    5         32    7         0         0   0.8541667
## Very Good     0    0         11    5         0         1   1.0000000
## Excellent     1    0         12    1         1         0   1.0000000
```

```
importance(rf.cat)
```

```
##                            Poor        Weak  Sufficient        Good
## school                1.418224026  0.40894251 -0.57960533  2.54388488
## sex                  -0.653501187  0.60370568  0.24977571  0.48066823
## age                   0.448344220 -1.17420634 -0.24533484 -0.59042922
## address              -0.693495819 -0.21330911 -1.50717981  0.96654289
## famsize               1.764145807  0.98725385  0.28687545  0.38019026
## Pstatus               0.437039727 -0.53762447  0.07045264 -0.87903414
## Mjob                 -1.109135394 -0.44161141 -1.50681345  0.95931824
## Fjob                  1.297625783  0.40257487  0.05823661  0.09064474
## reason               -0.097894846 -1.71667103  0.18204719  1.70951479
## guardian              0.147158285  0.15798101  1.23071197 -0.92995689
## traveltime           -0.425294519 -0.28006441  0.14534919 -1.43228661
## studytime             0.057173173  0.38568536 -0.50241649 -2.03731931
## failures              2.556769774  3.08729766  0.45793818  2.53571001
## schoolsup             1.753438366  1.25517987  0.78379658  1.70510374
## paid                  1.118627552  0.26189803  0.51657208 -0.63798601
## activities           -1.370551184 -1.20699948 -1.13224555 -0.95825557
## nursery              -1.298410047  1.02781486  0.46410186 -1.32884393
## higher               -2.316940279 -0.04532634 -0.68585552  1.78471688
## romantic             -0.680407593  0.30727455 -0.33609909 -0.40894821
## famrel               -1.078387545 -0.50289296 -1.24361414 -0.90902151
## freetime             -0.001088676  0.02740684  0.56238463 -0.82769673
## goout                -2.022831566  1.07316905  0.87287338  0.79658049
## Dalc                  0.000000000  0.41953213  0.20864014  1.01015254
## Walc                 -0.858439370 -0.04346963 -0.01723483  2.44025218
## health                0.793248289  0.58067311 -0.45732022 -0.18750654
## absences              7.448999273  1.58862031  2.78077467  2.04587569
## first_gen_college     0.416878406  0.25922961 -1.89180829 -0.13036804
## stable_learning_env -0.765481584  0.04294296 -0.09970439 -1.10686914
## high_freq_absent      2.915807853  0.34734604  0.95599262  3.03573666
## failed                2.885781564  1.52474214  0.81895718  2.58167937
##                        Very Good  Excellent MeanDecreaseAccuracy
## school               -1.01015254  1.0101525           0.80498543
## sex                  -1.34359993 -0.1754656           0.18194613
## age                  -1.44845467  1.2963037          -1.01353341
## address              -1.01015254  1.0101525          -1.29671970
## famsize               0.76232872  1.0101525           1.28901079
## Pstatus               0.00000000  0.4678229          -0.24896753
## Mjob                  1.14907792  0.5579040          -0.82550818
## Fjob                 -1.21657276 -0.4272368           0.26508220
## reason               -0.22067177  0.5833694          -0.03728804
## guardian             -0.66011578 -1.0101525           0.56306717
## traveltime            1.00191205 -1.0101525          -0.46143791
## studytime             0.51054941  0.3558777          -1.08499648
## failures              0.00000000  1.0101525           4.13700025
## schoolsup            -1.43177092  1.0101525           2.24716923
## paid                  0.97943770  1.0980312           1.01047467
## activities           -0.51987524  0.0000000          -2.07632814
```

```
## nursery              0.75798367  1.4229360          0.35161402
## higher               0.00000000  0.0000000         -1.25637567
## romantic            -0.31832142  1.0101525         -0.32461487
## famrel               1.01015254 -1.0101525         -1.93774260
## freetime             1.01015254 -0.6024591          0.27578889
## goout               -0.60245906  1.0101525          1.07199306
## Dalc                 0.00000000  0.0000000          0.62635885
## Walc                 0.49601049  1.0101525          0.32577261
## health              -1.56071254  0.1562119         -0.29254478
## absences             0.57155643  0.3742818          6.19172692
## first_gen_college    0.06283591  0.9200461         -0.74819158
## stable_learning_env -1.01015254 -1.4400461         -1.24639512
## high_freq_absent     1.76776695 -0.8904292          2.35017255
## failed               1.01015254  1.4229360          3.24673911
##                     MeanDecreaseGini
## school                      3.072587
## sex                         5.656577
## age                        13.331134
## address                     4.112408
## famsize                     5.529408
## Pstatus                     2.937808
## Mjob                       14.542602
## Fjob                       11.614861
## reason                     12.143577
## guardian                    6.934204
## traveltime                  7.047091
## studytime                  10.372108
## failures                    7.541551
## schoolsup                   4.955187
## paid                        6.012111
## activities                  6.294656
## nursery                     4.692421
## higher                      1.414178
## romantic                    4.924411
## famrel                      4.280805
## freetime                    5.942191
## goout                       5.707473
## Dalc                        1.028423
## Walc                        4.552716
## health                      5.306674
## absences                   21.543826
## first_gen_college           6.205268
## stable_learning_env         5.686258
## high_freq_absent            4.278214
## failed                      5.106471
```

```
varImpPlot(rf.cat)
```




```
rf.acc<- predict(rf.cat, ftrain, type = 'class')
t<-table(predictions=rf.acc, actual=ftrain$cat_g3)
t
```

```
##             actual
## predictions  Poor Weak Sufficient Good Very Good Excellent
##   Poor         33    0          0    0         0         0
##   Weak          0   71          0    0         0         0
##   Sufficient    0    2        130    0         1         0
##   Good          0    0          0   48         0         0
##   Very Good     0    0          0    0        16         0
##   Excellent     0    0          0    0         0        15
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.9905063
```

Very fitted model with accuracy for training data >99%.

Let's see what the accuracy rate for the testing set is:

```
rf.pred<- predict(rf.cat, ftest, type = 'class')
t<-table(predictions=rf.pred, actual=ftest$cat_g3)
t
```

```
##             actual
## predictions  Poor Weak Sufficient Good Very Good Excellent
##   Poor          2    1          2    0         0         0
##   Weak          1    0          4    0         1         1
##   Sufficient    2   17         24    9         2         1
##   Good          0    1          5    3         1         1
##   Very Good     0    0          0    0         1         0
```

```
##   Excellent     0    0          0    0         0         0
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.3797468
```

43.03% accuracy, which is an improvement.

Let's choose the most important variables, as well as interaction effects we believe to be important based on previous exploration:

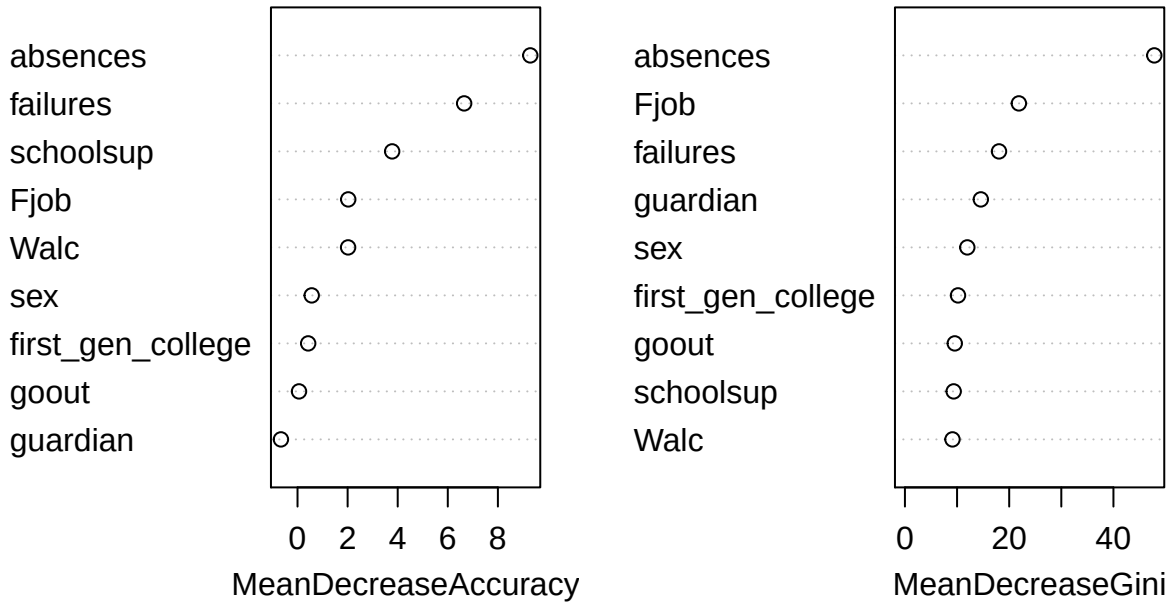
```
rf.cat1<-randomForest(cat_g3~failures + absences + sex + Walc + Fjob +goout + schoolsup + first_gen_coll
print(rf.cat1)
```

```
##
## Call:
##  randomForest(formula = cat_g3 ~ failures + absences + sex + Walc +      Fjob + goout + schoolsup + f
##                Type of random forest: classification
##                      Number of trees: 50
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 52.22%
## Confusion matrix:
##            Poor Weak Sufficient Good Very Good Excellent class.error
## Poor         23    3          4    2         0         1   0.3030303
## Weak          5   27         38    2         0         1   0.6301370
## Sufficient    5   23         87   11         2         2   0.3307692
## Good          2    5         23   13         1         4   0.7291667
## Very Good     1    1         10    3         1         1   0.9411765
## Excellent     0    1         10    3         1         0   1.0000000
```

```
importance(rf.cat1)
```

```
##                        Poor       Weak Sufficient       Good  Very Good
## failures          4.2424935  4.1450188  1.0631210  3.9687210  2.9769975
## absences         14.7974760  1.7270164  3.9337529  1.4591986  1.8980916
## sex               3.3357840  1.5103454 -2.3359910  1.0936902  1.5714323
## Walc              2.4844928 -0.3991997 -0.6085898  3.7842420  3.8025511
## Fjob              0.3852836  2.2506175  1.2083110 -1.8090465  2.1483694
## goout            -1.5310686  1.6782038 -1.3062458  0.8902641  3.6881474
## schoolsup         4.0204822  4.3824578 -1.2463088  3.2099174  1.0610934
## first_gen_college 0.6975070  0.4266225 -2.2457970  4.6074291 -0.9355315
## guardian          1.2666280 -2.5306220 -0.3286237 -1.1966790  3.0171426
##                   Excellent MeanDecreaseAccuracy MeanDecreaseGini
## failures          1.0101525           6.65313915        18.045934
## absences         -1.4509532           9.29408139        47.834972
## sex              -0.9379438           0.56632438        11.968161
## Walc              1.0101525           2.01966213         9.122728
## Fjob              0.1039052           2.02501814        21.861115
## goout             0.8013456           0.05896317         9.559409
## schoolsup         0.0000000           3.77886103         9.351629
## first_gen_college 0.7912918           0.42524516        10.182273
## guardian         -0.8099566          -0.66267503        14.557878
```

```
varImpPlot(rf.cat1)
```




```
rf.acc<- predict(rf.cat1, ftrain, type = 'class')
t<-table(predictions=rf.acc, actual=ftrain$cat_g3)
t
```

```
##             actual
## predictions  Poor Weak Sufficient Good Very Good Excellent
##   Poor         30    4          2    1         0         0
##   Weak          0   55          2    0         0         0
##   Sufficient    2   12        123   13         6         6
##   Good          1    1          1   34         1         2
##   Very Good     0    0          1    0         9         0
##   Excellent     0    1          1    0         1         7
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.8164557
```

54.75% OOB estimate of error rate and 83.5% accuracy rate for the training data.

```
rf.pred1<- predict(rf.cat1, ftest, type = 'class')
t<-table(predictions=rf.pred1, actual=ftest$cat_g3)
t
```

```
##             actual
## predictions  Poor Weak Sufficient Good Very Good Excellent
##   Poor          2    1          2    0         0         0
##   Weak          1    1          6    1         0         0
##   Sufficient    1   15         23    7         3         1
##   Good          0    2          4    3         1         2
##   Very Good     0    0          0    1         0         0
##   Excellent     1    0          0    0         1         0
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.3670886
```

37.97% Accuracy, which is less than the full RF model.

The RF models indicate that for grade categorization, the most important variables are absences, failed, guardian, studytime, Mjob and Fjob, schoolsup, age, goout, first_gen_college (not in that order).

Modeling for low-high grades

Considering final grades as a continuous variable and ordinal categorical variable gave poor results. Therefore, we'd like to model a binary variable that indicates whether the student has a high grade (grade >= 10) or low grade (<10).

```
set.seed(3)
train_ind1 <- sample(x = nrow(data), size = 0.8 * nrow(data))
test_ind_neg1 <- -train_ind1
ftrain1 <- data[train_ind1, ]
ftest1 <- data[test_ind_neg1, ]
```

Fitting a decision tree on pass-fail

```
data[["pf"]] <- as.factor(data[["pf"]])
training[["pf"]] <- as.factor(training[["pf"]])
testing[["pf"]] <- as.factor(testing[["pf"]])
treepf <- tree(pf ~ . -G1 -G2 -G3 -ord_g3 -failures -reason -health -age -nursery -ord_g3, data=training
```

```
## Warning in tree(pf ~ . - G1 - G2 - G3 - ord_g3 - failures - reason - health - :
## NAs introduced by coercion
```

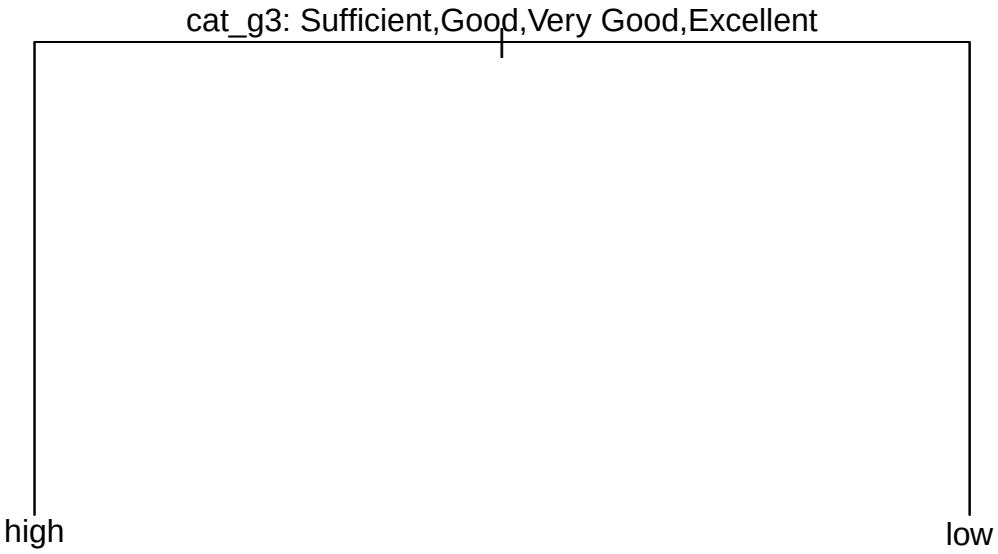
```
treepf
```

```
## node), split, n, deviance, yval, (yprob)
##       * denotes terminal node
##
## 1) root 316 403.2 high ( 0.6646 0.3354 )
##   2) cat_g3: Sufficient,Good,Very Good,Excellent 210   0.0 high ( 1.0000 0.0000 ) *
##   3) cat_g3: Poor,Weak 106   0.0 low ( 0.0000 1.0000 ) *
```

```
summary(treepf)
```

```
##
## Classification tree:
## tree(formula = pf ~ . - G1 - G2 - G3 - ord_g3 - failures - reason -
##     health - age - nursery - ord_g3, data = training)
## Variables actually used in tree construction:
## [1] "cat_g3"
## Number of terminal nodes:  2
## Residual mean deviance:  0 = 0 / 314
## Misclassification error rate: 0 = 0 / 316
```

```
plot(treepf)
text(treepf, pretty = 0)
```



Initial Tree Diagnostic

```
tree.pred <- predict(treepf, testing, type = "class")
```

```
## Warning in pred1.tree(object, tree.matrix(newdata)): NAs introduced by coercion
```

```
table(tree.pred, testing$pf)
```

```
##
## tree.pred high low
##      high   55   0
##      low     0  24
```

```
sum(diag(table(tree.pred, testing$pf)))/79
```

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

Misclassification rate: 0.38. This can likely be decreased with other methods- using all variables likely overfits.

###Pruning

```
set.seed(3)
cv.pf <- cv.tree(treepf, FUN = prune.misclass)
```

```
## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion
```

```
## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion
```

```
## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion
```

```
## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion
```

```
## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion
```

```
## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion
```

```
## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion
```

```
## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion

## Warning in tree(model = m[rand != i, , drop = FALSE]): NAs introduced by
## coercion

## Warning in pred1.tree(tree, tree.matrix(nd)): NAs introduced by coercion
```

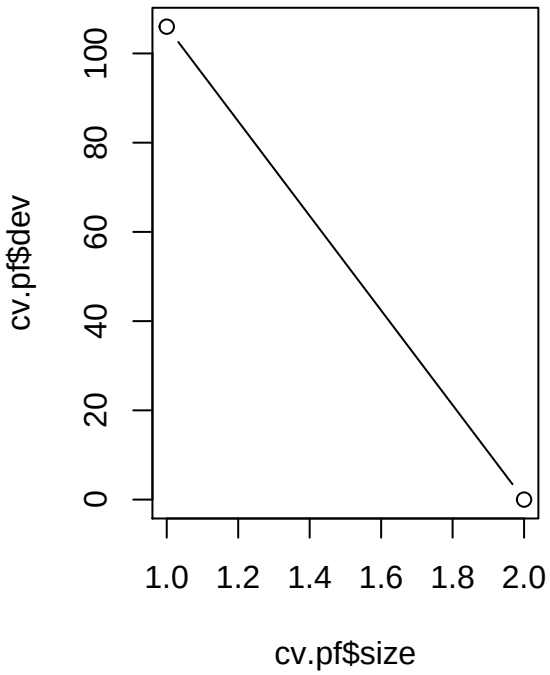
```
names(cv.pf)
```

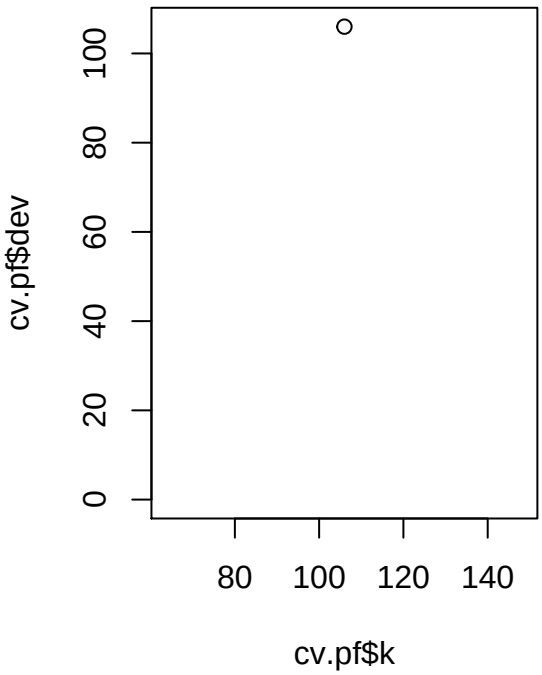
```
## [1] "size"   "dev"    "k"      "method"
```

```
cv.pf
```

```
## $size
## [1] 2 1
##
## $dev
## [1]   0 106
##
## $k
## [1] -Inf  106
##
## $method
## [1] "misclass"
##
## attr(,"class")
## [1] "prune"         "tree.sequence"
```

```
par(mfrow = c(1,2))
plot(cv.pf$size, cv.pf$dev, type = "b")
plot(cv.pf$k, cv.pf$dev, type = "b")
```

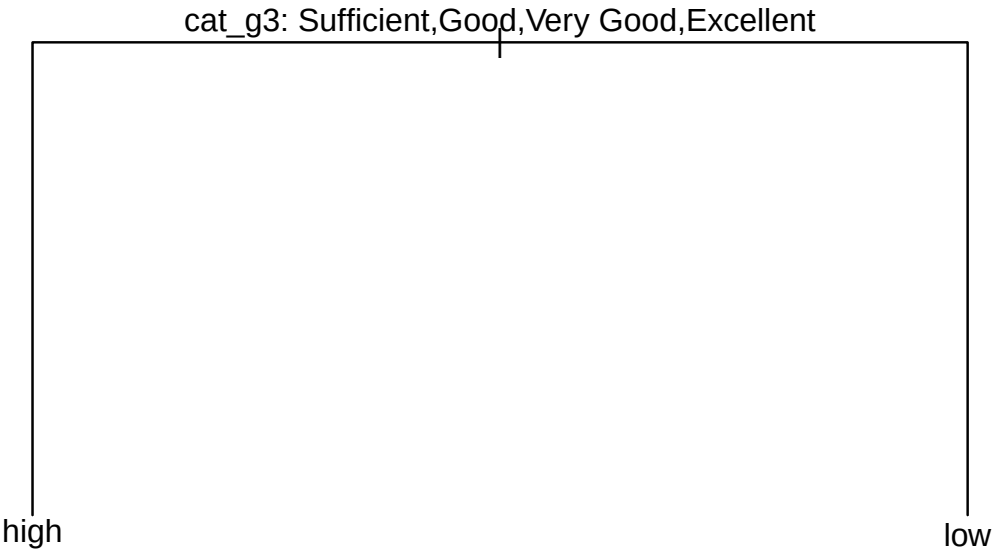




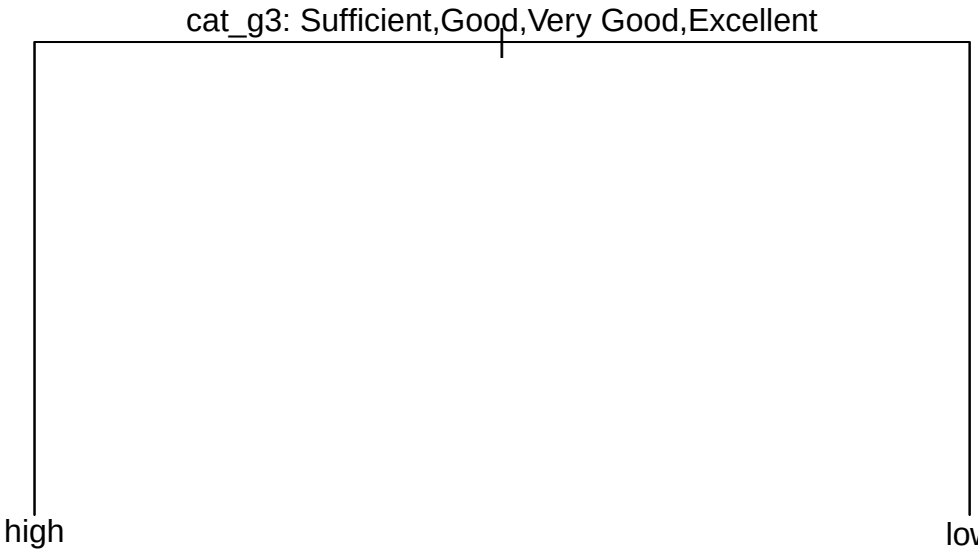

```
prune.pf <- prune.misclass(treepf, best = 3)
```

```
## Warning in prune.tree(tree = treepf, best = 3, method = "misclass"): best is
## bigger than tree size
```

```
plot(prune.pf)
text(prune.pf, pretty = 0)
```




```
prune.short <- prune.misclass(treepf, best = 2)
plot(prune.short)
text(prune.short, pretty = 0)
```




```
treepred2 <- predict(prune.pf, testing, type = "class")
```

```
## Warning in pred1.tree(object, tree.matrix(newdata)): NAs introduced by coercion
```

```
table(treepred2, testing$pf)
```

```
##
## treepred2 high low
##      high   55   0
##      low     0  24
```

```
sum(diag(table(treepred2, testing$pf)))/79
```

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

```
treepred3 <- predict(prune.short, testing, type = "class")
```

```
## Warning in pred1.tree(object, tree.matrix(newdata)): NAs introduced by coercion
```

```
table(treepred3, testing$pf)
```

```
##
## treepred3 high low
##      high   55   0
##      low     0  24
```

```
sum(diag(table(treepred3, testing$pf)))/79
```

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

Misclassification rateL .32.

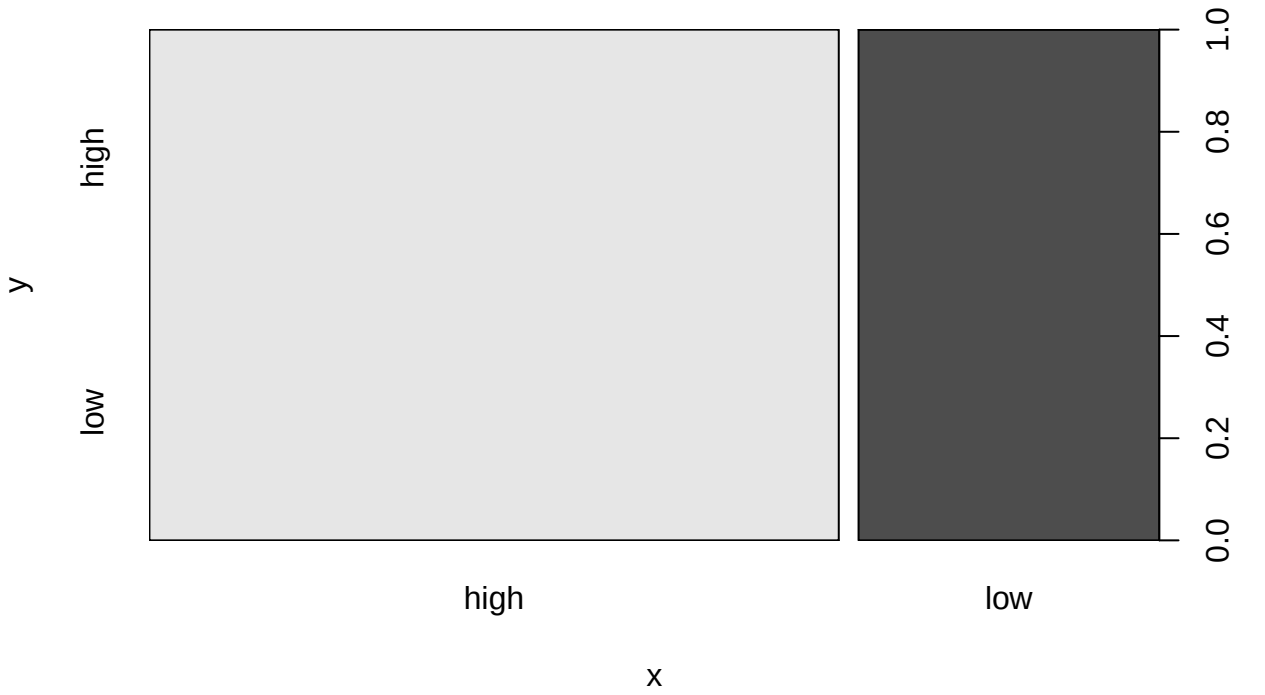
Bagging

```
library(randomForest)
set.seed(1)
bag.pf <- randomForest(pf ~ . -G1 -G2 -G3 -ord_g3 -failures -reason -health -age -nursery -ord_g3, data=
bag.pf
```

```
##
## Call:
```

```
##  randomForest(formula = pf ~ . - G1 - G2 - G3 - ord_g3 - failures -      reason - health - age - nurs
##                Type of random forest: classification
##                      Number of trees: 75
## No. of variables tried at each split: 28
##
##         OOB estimate of  error rate: 0%
## Confusion matrix:
##      high low class.error
## high  210   0           0
## low     0 106           0
```

```
yhat.bag <- predict(bag.pf, testing)
plot(yhat.bag, testing$pf)
```




```
table(yhat.bag, testing$pf)
```

```
##
## yhat.bag high low
##     high   55   0
##     low     0  24
```

```
sum(diag(table(yhat.bag, testing$pf)))/79
```

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

Boosting

```
library(gbm)
```

```
## Loaded gbm 2.1.8
```

```
attach(data)
```

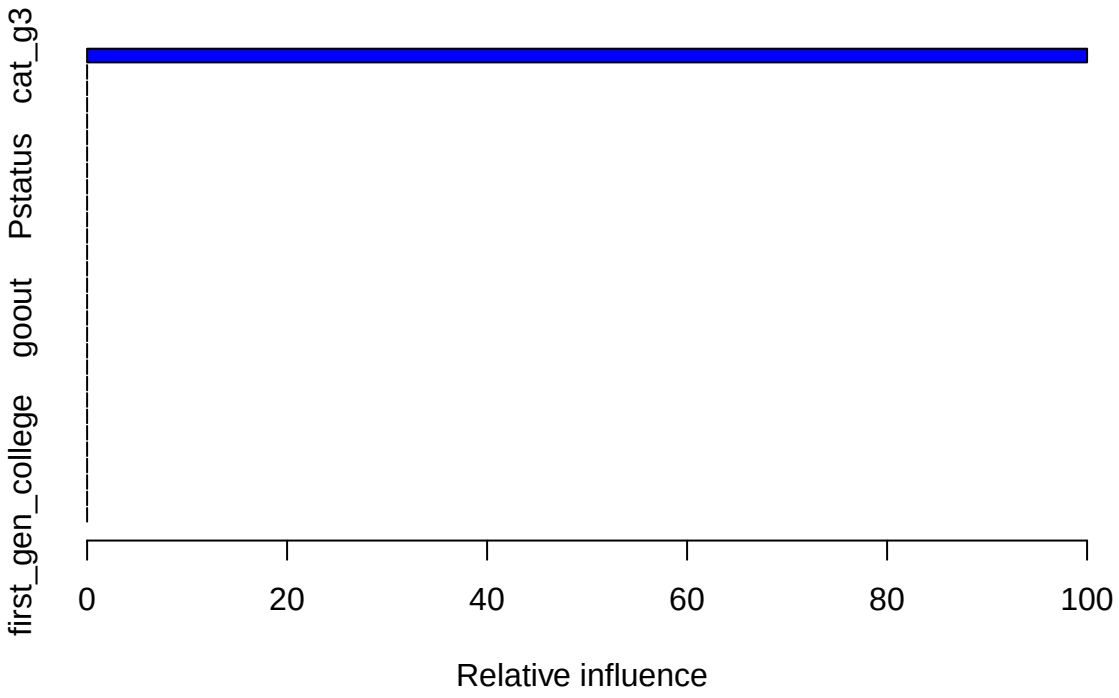
```
## The following objects are masked from data (pos = 4):
##
```

```
##     Dalc, Fedu, Fjob, G1, G2, G3, Medu, Mjob, Pstatus, Walc, absences,
##     activities, address, age, cat_g3, failed, failures, famrel,
##     famsize, famsup, first_gen_college, freetime, goout, guardian,
##     health, high_freq_absent, higher, internet, nursery, ord_g3, paid,
##     pf, reason, romantic, school, schoolsup, sex, stable_learning_env,
##     studytime, traveltime
```

```
data[["pf_factor"]] <- as.factor(data[["pf"]])
data[["pf_bin"]] <- as.numeric(data[["pf_factor"]])-1
training[["pf_factor"]] <- as.factor(training[["pf"]])
training[["pf_bin"]] <- as.numeric(training[["pf_factor"]])-1
testing[["pf_factor"]] <- as.factor(testing[["pf"]])
testing[["pf_bin"]] <- as.numeric(testing[["pf_factor"]])-1

set.seed(1)
boost.pf <- gbm(pf_bin ~ . -pf_factor -pf -school -G1 -G2 -G3 -ord_g3 -failures -reason -health -age -nu
                    distribution = "bernoulli", n.trees = 500,
                    interaction.depth = 2)

summary(boost.pf)
```




```
##                                 var      rel.inf
## cat_g3                       cat_g3 1.000000e+02
## absences                   absences 1.527612e-27
## traveltime               traveltime 7.700951e-29
## Mjob                           Mjob 5.879732e-29
## studytime                 studytime 8.341861e-30
## Medu                           Medu 4.589843e-30
## schoolsup                 schoolsup 3.356194e-30
## Fjob                           Fjob 3.083552e-30
## Pstatus                     Pstatus 1.407552e-31
## famsize                     famsize 1.559680e-32
```

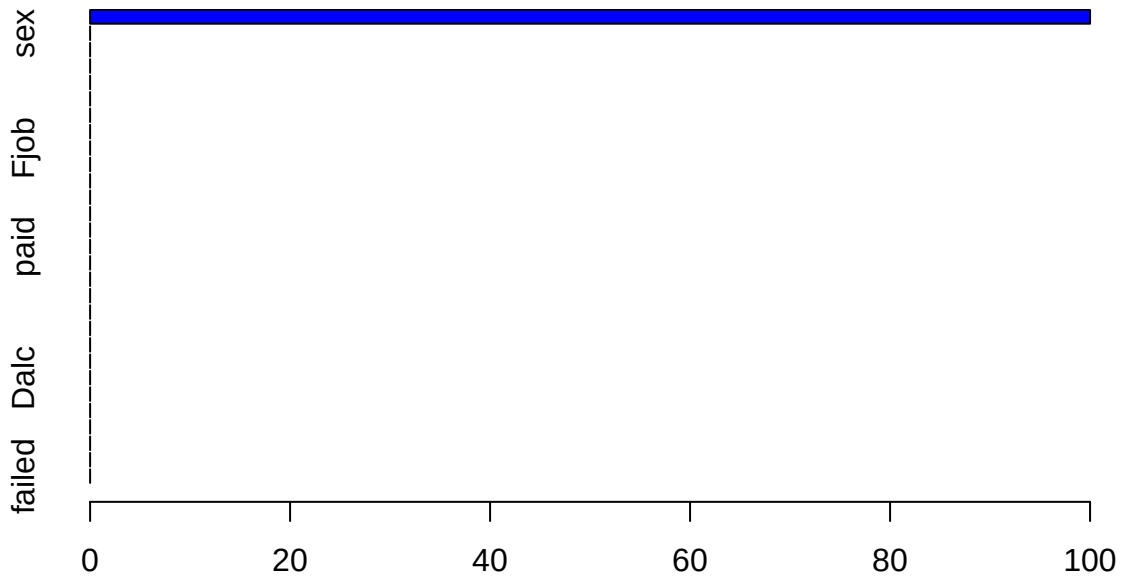
```
## address                         address 8.849969e-34
## internet                       internet 1.157148e-34
## high_freq_absent       high_freq_absent 2.399331e-38
## Walc                               Walc 2.156494e-42
## failed                           failed 4.545526e-46
## Fedu                               Fedu 1.215371e-49
## goout                             goout 4.070774e-61
## guardian                       guardian 2.196326e-62
## paid                               paid 4.185927e-66
## famrel                           famrel 4.076438e-66
## romantic                       romantic 5.379314e-68
## stable_learning_env stable_learning_env 8.809047e-72
## activities                   activities 2.816272e-73
## sex                                 sex 0.000000e+00
## famsup                           famsup 0.000000e+00
## higher                           higher 0.000000e+00
## freetime                       freetime 0.000000e+00
## Dalc                               Dalc 0.000000e+00
## first_gen_college     first_gen_college 0.000000e+00
```

```
predboost1 <- predict(boost.pf, testing,
                      n.trees = 500)
table(predboost1, testing$pf_bin)
```

```
##
## predboost1           0  1
##   -51.1132477456973 55  0
##   36.7658750128531   0 24
```

Lower interaction depth

```
boost.pf1 <- gbm(pf_bin ~ . -pf_factor -pf -school -G1 -G2 -G3 -ord_g3 -failures -reason -health -age -
                   distribution = "bernoulli", n.trees = 500,
                   interaction.depth = 1)
summary(boost.pf1)
```




```
##                                    var rel.inf
## cat_g3                           cat_g3     100
## sex                                 sex       0
## address                         address       0
## famsize                         famsize       0
## Pstatus                         Pstatus       0
## Medu                               Medu       0
## Fedu                               Fedu       0
## Mjob                               Mjob       0
## Fjob                               Fjob       0
## guardian                       guardian       0
## traveltime                   traveltime       0
## studytime                     studytime       0
## schoolsup                     schoolsup       0
## famsup                           famsup       0
## paid                               paid       0
## activities                   activities       0
## higher                           higher       0
## internet                       internet       0
## romantic                       romantic       0
## famrel                           famrel       0
## freetime                       freetime       0
## goout                             goout       0
## Dalc                               Dalc       0
## Walc                               Walc       0
## absences                       absences       0
## first_gen_college     first_gen_college       0
## stable_learning_env stable_learning_env       0
## high_freq_absent       high_freq_absent       0
## failed                           failed       0
```

```
predboost1 <- predict(boost.pf, testing,
                      n.trees = 500)
```

```
table(predboost1, testing$pf_bin)
```

```
##
## predboost1           0  1
##   -51.1132477456973 55  0
##   36.7658750128531   0 24
```

Not much difference.

Fitting random forest on low-high binary

Fitting with ALL predictors:

```
rf.bin<-randomForest(pf~. -G1 -G2 -G3 -ord_g3 - cat_g3 -Medu -Fedu,data = ftrain1,mtry=3, ntree=50, imp
print(rf.bin)
```

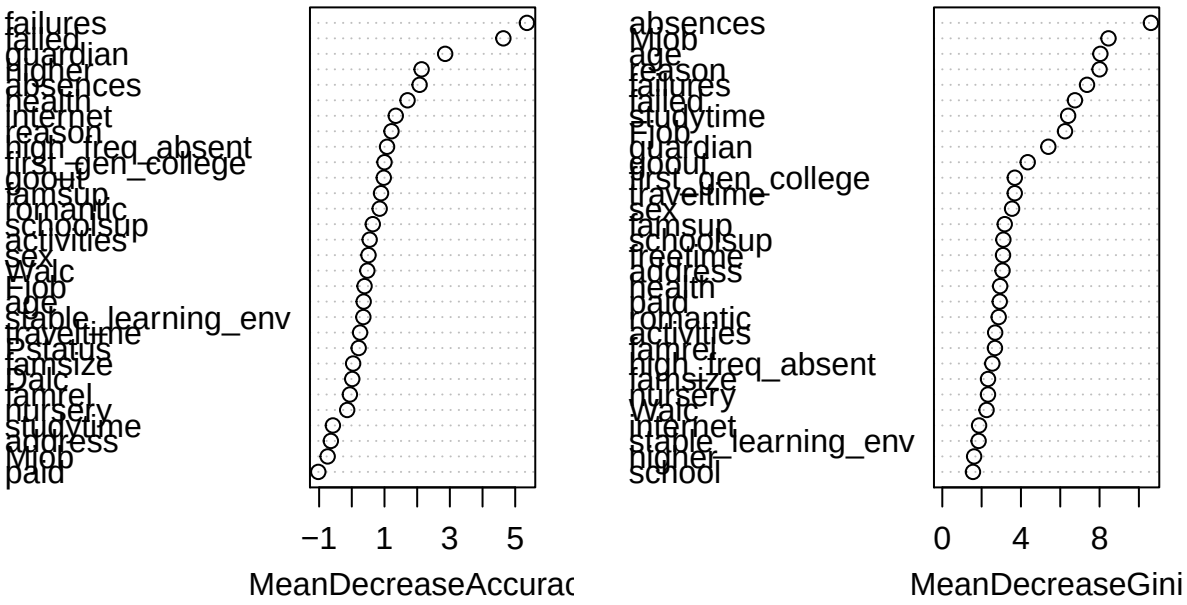
```
##
## Call:
##  randomForest(formula = pf ~ . - G1 - G2 - G3 - ord_g3 - cat_g3 -      Medu - Fedu, data = ftrain1, r
##                Type of random forest: classification
##                      Number of trees: 50
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 27.85%
## Confusion matrix:
##      high low class.error
## high  190  20   0.0952381
## low    68  38   0.6415094
```

```
importance(rf.bin)
```

```
##                          high         low MeanDecreaseAccuracy
## school              -2.485581508 -1.52673306          -2.63710303
## sex                 -0.965165302  2.93414745           0.51049234
## age                  0.139033230  0.40753652           0.36305088
## address             -1.265315512  0.56919163          -0.64142255
## famsize              0.313937542 -0.15509877           0.04017503
## Pstatus              0.300092074 -0.04079301           0.21024630
## Mjob                -0.665033163 -0.73081250          -0.73948831
## Fjob                -0.379109665  1.06891868           0.39015400
## reason               0.714816698  1.24057039           1.21413248
## guardian             2.173626595  1.84588051           2.85619441
## traveltime           0.646069637 -0.43241156           0.25330303
## studytime           -0.342003707 -0.08909346          -0.57883698
## failures             4.346277380  3.55520510           5.35786996
## schoolsup           -0.411547231  1.69806708           0.64211123
## famsup               1.635973757 -0.52964534           0.89440050
## paid                -1.202828598  0.12282011          -1.02521760
## activities          -0.233423760  1.23855081           0.54681514
## nursery              0.380920261 -0.51327766          -0.14170752
## higher               1.297165181  1.66139689           2.13552428
## internet             1.517377344  0.09930989           1.34488620
## romantic             0.326795259  1.39052208           0.85260390
## famrel               0.258084483 -0.34734701          -0.05859069
## freetime            -1.697979099 -1.19984074          -1.89631703
## goout               -0.366115745  2.29795989           0.97810480
```

```
## Dalc                -0.227459308  0.20745016          0.01294779
## Walc                -1.455065051  1.97279834          0.47546418
## health               1.035644277  1.46258333          1.70557566
## absences             2.505092087  0.74901892          2.07446675
## first_gen_college    0.002474154  1.88146528          1.00016786
## stable_learning_env -0.137473745  1.10737207          0.35268623
## high_freq_absent     0.028780867  1.83391235          1.07867783
## failed               4.759518594  3.25610088          4.63730082
##                     MeanDecreaseGini
## school                     1.5600159
## sex                        3.5491078
## age                        8.0402404
## address                    3.0645379
## famsize                    2.3253121
## Pstatus                    1.4532774
## Mjob                       8.4544034
## Fjob                       6.2418644
## reason                     7.9999579
## guardian                   5.3937605
## traveltime                 3.6843505
## studytime                  6.3999760
## failures                   7.3620293
## schoolsup                  3.1061548
## famsup                     3.1764733
## paid                       2.9223962
## activities                 2.6885872
## nursery                    2.3242116
## higher                     1.6192776
## internet                   1.8714043
## romantic                   2.8689542
## famrel                     2.6783574
## freetime                   3.0896971
## goout                      4.3452508
## Dalc                       0.8336794
## Walc                       2.2521656
## health                     2.9419715
## absences                  10.6122048
## first_gen_college          3.6871370
## stable_learning_env        1.8476062
## high_freq_absent           2.5439946
## failed                     6.7471392
```

```
varImpPlot(rf.bin)
```




```
rf.acc<- predict(rf.bin, ftrain1, type = 'class')
t<-table(predictions=rf.acc, actual=ftrain1$pf)
t
```

```
##            actual
## predictions high low
##        high  210   2
##        low     0 104
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.9936709
```

Predictions on testing set:

```
rf.pred2<- predict(rf.bin, ftest1, type = 'class')
t<-table(predictions=rf.pred2, actual=ftest1$pf)
t
```

```
##            actual
## predictions high low
##        high   48  20
##        low     7   4
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.6582278
```

72.15% accuracy rate.

Finding the best random forest model by including important predictors:

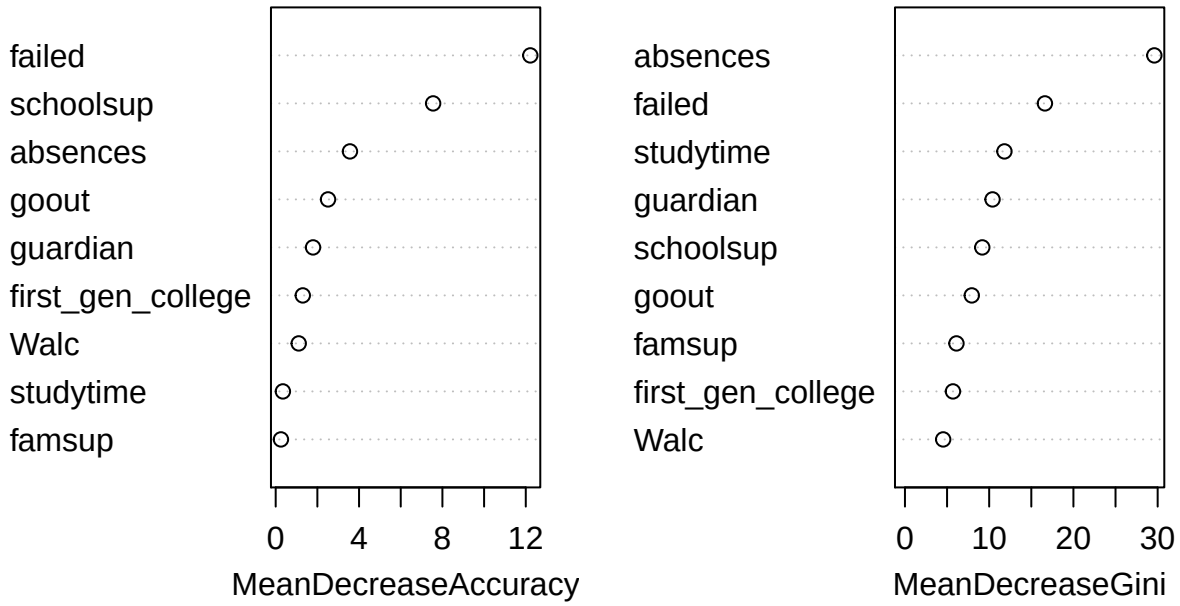
```
rf.bin1<-randomForest(pf~failed + absences+ guardian + studytime + goout + schoolsup + first_gen_colleg
print(rf.bin1)
```

```
##
## Call:
##  randomForest(formula = pf ~ failed + absences + guardian + studytime +      goout + schoolsup + firs
##                Type of random forest: classification
##                      Number of trees: 50
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 27.85%
## Confusion matrix:
##      high low class.error
## high  182  28   0.1333333
## low    60  46   0.5660377
```

```
importance(rf.bin1)
```

```
##                         high        low MeanDecreaseAccuracy MeanDecreaseGini
## failed             8.76476089 10.7432931           12.2176093        16.615519
## absences           3.10547085  2.0738397            3.5634858        29.604750
## guardian           2.58574861 -0.5726153            1.7926224        10.396056
## studytime         -0.04201686  0.5169681            0.3448446        11.809784
## goout              1.57374996  2.8259365            2.5109677         7.934012
## schoolsup          5.17404427  5.9413297            7.5551707         9.177453
## first_gen_college  0.20283606  1.8521200            1.2991851         5.697752
## Walc               0.86179592  0.5995871            1.1049734         4.536099
## famsup             1.32940595 -0.9431626            0.2515182         6.114414
```

```
varImpPlot(rf.bin1)
```




```
rf.acc1<- predict(rf.bin1, ftrain1, type = 'class')
t<-table(predictions=rf.acc1, actual=ftrain1$pf)
t
```

```
##            actual
## predictions high low
##        high  208  25
##        low     2  81
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.914557
```

The pared-down model has an 00B estimate of error rate of 25.95% and a training set prediction accuracy rate of 90.19%.

Predictions on testing set:

```
rf.pred3<- predict(rf.bin1, ftest1, type = 'class')
t<-table(predictions=rf.pred3, actual=ftest1$pf)
t
```

```
##            actual
## predictions high low
##        high   48  18
##        low     7   6
```

```
sum(diag(t))/sum(t)
```

```
## [1] 0.6835443
```

72.15% prediction accuracy rate.

Overall the random-forests for pass-fail indicate that the most important factors affecting whether the student passes/fails are failed, absences, guardian, studytime, goout, schoolsup, first_gen_college.