Machine Learning

Data Types Conversion

```
data_types=function(data){
  data=data %>%
  mutate(across(c(
                  "MorningClasses",
                  "Transfer",
                  "BUSMAJ",
                  "BUSJMA",
                  "MINOR",
                  "FINCON",
                  "MKTGCON",
                  "BUSCOP",
                  "STANCON",
                  "ACCTCON",
                  "MISCON",
                  "ENTINNCON",
                  "OPERMGTCON",
                  "HRMCON",
                  "INBUCON",
                  "DeclaredConcentration",
                  "At_least_one_concentration_course_per_semester_enrolled"),
                as.factor))
  return(data)
}
```

Course Feature Selection

```
courseFeatureSelection=function(data1){
  data1=data1%>%dplyr::select(-c(STD_INDEX,Mon_classes,
                                 Tues_classes,Wed_classes,
                           Thurs_classes,Fri_classes,
                           MorningClassCount,AfternoonClassCount,HON,
                           TotalClass,OLC_courses_count,
                           Credits.Taken.Annualy,Concentration.Credits,
                           Concentration.Courses.Taken,Semesters_enrolled,
                           Concentration.Credits.Prop.of.Total.Credits,
                           Credits.Towards.a.Certificate,Business.Credits,
                           Certificate.Course,Bus.Course,
                           Certificate.Credits.Prop,
                           Bus.Credit.Prop,Avg.Courses.Taken.per.semester,
                           OLC_courses_prop,LAWPHILCON,ECONDACON,
                           APPCON,Elective_Credits))
  return(data1)
}
```

Random Forest

```
rf=function(data,y){
  set.seed(456)
  partition=createDataPartition(data[[y]],p=0.75,list=FALSE)
```

```
  data_train=data[partition,]
  data_test=data[-partition,]
  gridRF=expand.grid(mtry=c(3,5,10),splitrule="variance",
                     min.node.size=c(1,3,5,10))
  train_control= trainControl(
    method = "cv",
    number = 10,
    verboseIter = FALSE
  )
  rg=train(as.formula(paste0("`", y, "` ~ .")),
           data=data_train,method="ranger",importance = "permutation",
           num.trees=1000,trControl = train_control,tuneGrid=gridRF)

  importance=varImp(rg, scale = FALSE)
  rfImportance=varImp(rg)
  Top5RfImportance=rfImportance$importance%>%as.data.frame()%>%
    rownames_to_column("Feature") %>% arrange(desc(Overall))%>%head(5)
  print(ggplot(data=Top5RfImportance,
               mapping=aes(x=Overall,y= reorder(Feature, Overall)))+
          geom_bar(stat="identity",fill="#A6192E")+xlab("Importance")+
          ylab("Variables")+
          ggtitle("Most Important Variables from the Random Forest Model"))


  predictions=predict(rg, newdata = data_test)
  print("Random Forest RMSE")
  print(sqrt(mean((predictions - data_test[[y]])^2)))
  return(list(Top5RfImportance,rg))
}
```

Linear Regression

```
linearModel=function(data,y){
  set.seed(456)
  partition=createDataPartition(data[[y]],p=0.75,list=FALSE)
  data_train=data[partition,]
  data_test=data[-partition,]
  lg=glm(as.formula(paste0("`", y, "` ~ .")),
         data = data_train, family = gaussian(link = "identity"))
  lgPreds=predict(lg,newdata=data_test)
  print("Linear Regression RMSE")
  print(sqrt(mean((data_test[[y]] - lgPreds)^2)))
  print(summary(lg))
  return(lg)
}
```

Stepwise Regression

```
stepwise=function(model,y,data){
  set.seed(456)
  partition=createDataPartition(data[[y]],p=0.75,list=FALSE)
  data_train=data[partition,]
```

```
  data_test=data[-partition,]
  stepwise=stepAIC(model,direction = "both",trace=FALSE)
  print(summary(stepwise))
  lgPreds=predict(stepwise,newdata=data_test)
  print("Stepwise Regression RMSE")
  print(sqrt(mean((data_test[[y]] - lgPreds)^2)))
}
```

XGboost

```
xg=function(data,y){
  set.seed(456)
  partition=createDataPartition(data[[y]],p=0.75,list=FALSE)
  data_train=data[partition,]
  data_test=data[-partition,]
  gridRF=expand.grid(mtry=c(3,5,10),splitrule="variance",
                     min.node.size=c(1,3,5,10))
  train_control= trainControl(
    method = "cv",
    number = 10,
    verboseIter = FALSE
  )
  rg=train(as.formula(paste0("`", y, "` ~ .")),
           data=data_train,method="ranger",importance = "permutation",
           num.trees=1000,trControl = train_control,tuneGrid=gridRF)

  importance=varImp(rg, scale = FALSE)
  rfImportance=varImp(rg)
  Top5RfImportance=rfImportance$importance%>%as.data.frame()%>%
    rownames_to_column("Feature") %>% arrange(desc(Overall))%>%head(5)
  print(ggplot(data=Top5RfImportance,
               mapping=aes(x=Overall,y= reorder(Feature, Overall)))+
          geom_bar(stat="identity",fill="#A6192E")+xlab("Importance")+
          ylab("Variables")+
          ggtitle("Most Important Variables from the Xgboost Model"))

  predictions=predict(rg, newdata = data_test)
  print("Xgboost RMSE")
  print(sqrt(mean((predictions - data_test[[y]])^2)))
  return(list(Top5RfImportance,rg))
}
```

Credit Feature Selection

```
creditFeatures=function(data){
  data=data%>%dplyr::select(-c(STD_INDEX,Mon_classes,Tues_classes,
                               Wed_classes,Thurs_classes,Fri_classes
                          ,MorningClassCount,AfternoonClassCount,TotalClass,
                          OLC_courses_count,Total.Courses
                          ,Concentration.Courses.Taken,Semesters_enrolled,
                          LAWPHILCON,ECONDACON,APPCON,
                          Credits.Towards.a.Certificate
```

```
                        ,Business.Credits,Certificate.Course,
                        Bus.Course,
                        Concentration.Courses.Prop.of.Total.Courses,Concentration.Credits,
                        Certificate.Courses.Prop.of.Total.Courses,
                        Business.Course.Prop.of.Total.Courses,
                        Avg.Courses.Taken.per.semester,HON,Elective_Credits))
  return(data)
}
```

Partial Dependence Plots Function

```
pdp=function(data,model,variable,target){
  Partial=partial(object=model,pred.var=variable,train = data)
  colnames(Partial)=c("Variable","yhat")
  print(ggplot(Partial,aes(x = Variable, y = yhat))+
    geom_point(color = "#A6192E", size = 3) +
    ggtitle(paste0("Partial Dependence Plot of ", variable, " vs ",target))+
      xlab(variable) +
      ylab(target)
    )
}
```

Reading in Data

```
data=read.csv("EnrollmentsCleaned.csv")
data=data_types(data)
data1=data
```

Annual Credits Taken

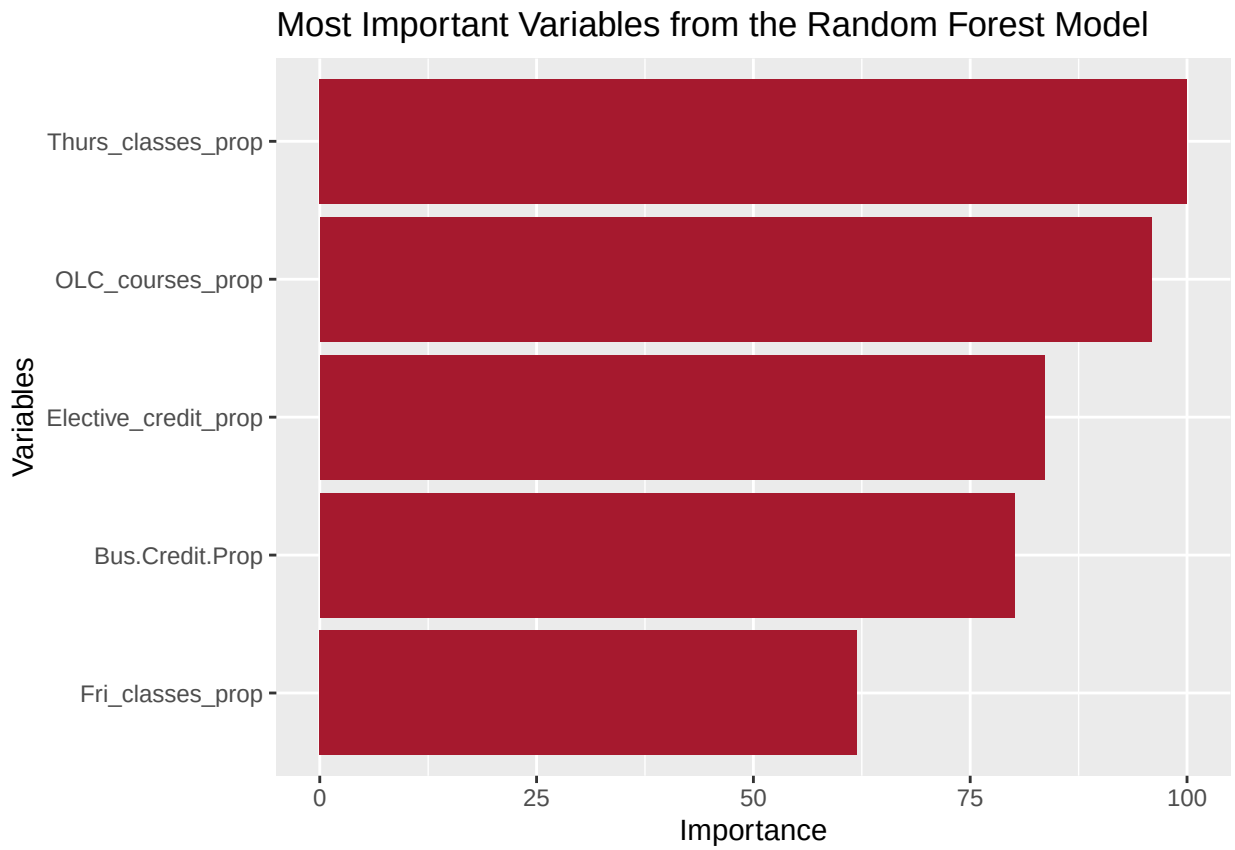
Random Forest and Important Variables

```
data=creditFeatures(data)
colnames(data)
```

```
##  [1] "HourStartMode"
##  [2] "MorningClasses"
##  [3] "Credits.Taken.Annualy"
##  [4] "semesters_with_at_least_one_concentration_course"
##  [5] "Transfer"
##  [6] "BUSMAJ"
##  [7] "BUSJMA"
##  [8] "MINOR"
##  [9] "FINCON"
## [10] "MKTGCON"
## [11] "BUSCOP"
## [12] "STANCON"
## [13] "ACCTCON"
## [14] "MISCON"
## [15] "ENTINNCON"
## [16] "OPERMGTCON"
## [17] "HRMCON"
## [18] "INBUCON"
```

```
## [19] "DeclaredConcentration"
## [20] "At_least_one_concentration_course_per_semester_enrolled"
## [21] "Mon_classes_prop"
## [22] "Tues_classes_prop"
## [23] "Wed_classes_prop"
## [24] "Thurs_classes_prop"
## [25] "Fri_classes_prop"
## [26] "MorningClassProp"
## [27] "AfternoonClassProp"
## [28] "OLC_courses_prop"
## [29] "Concentration.Credits.Prop.of.Total.Credits"
## [30] "Certificate.Credits.Prop"
## [31] "Bus.Credit.Prop"
## [32] "Elective_credit_prop"
```

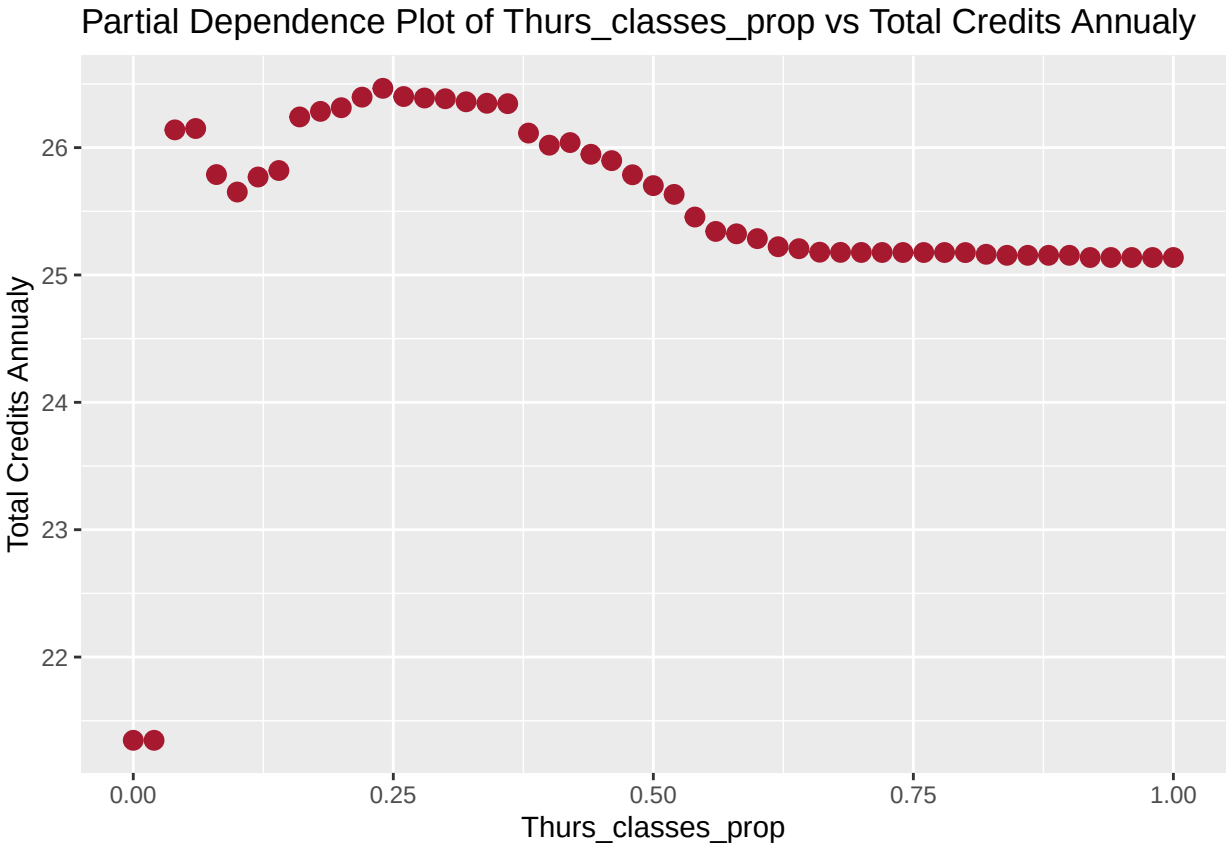
```
RF=rf(data,"Credits.Taken.Annualy")
```

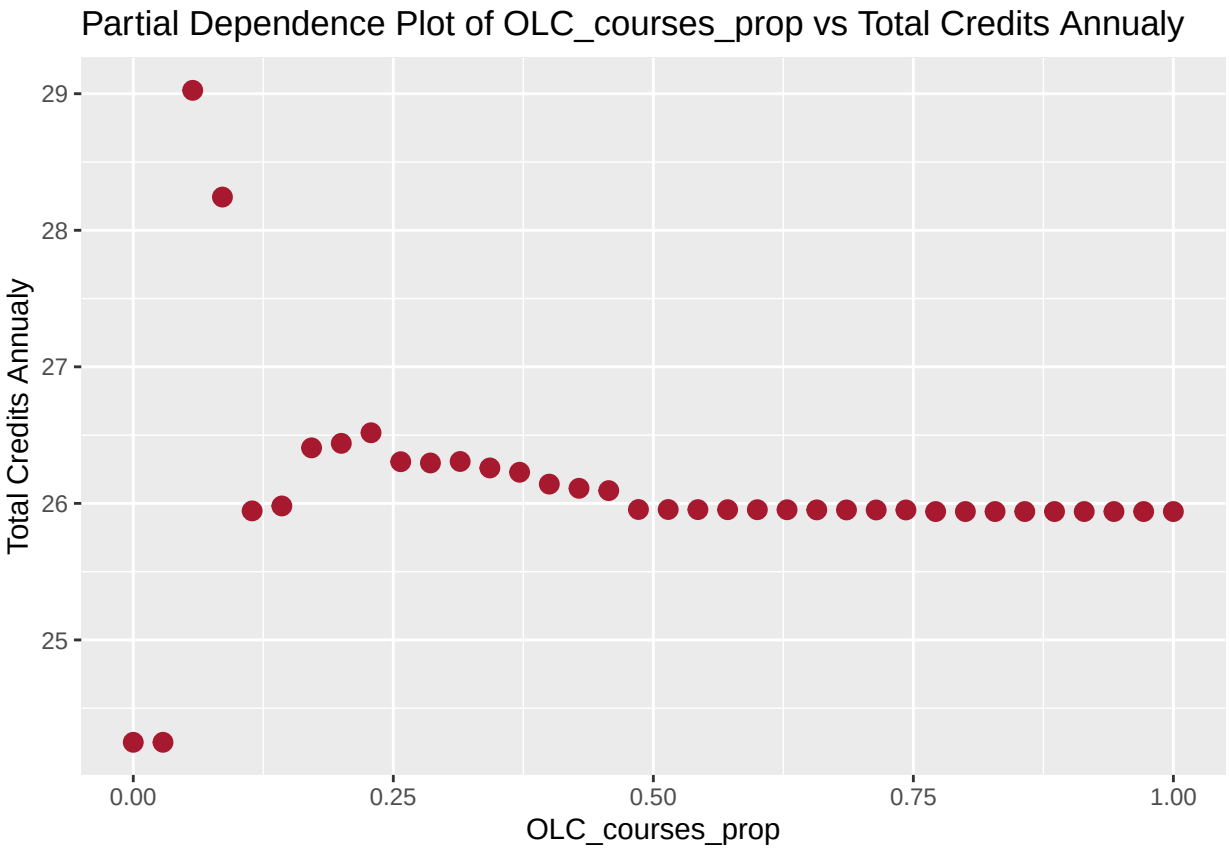


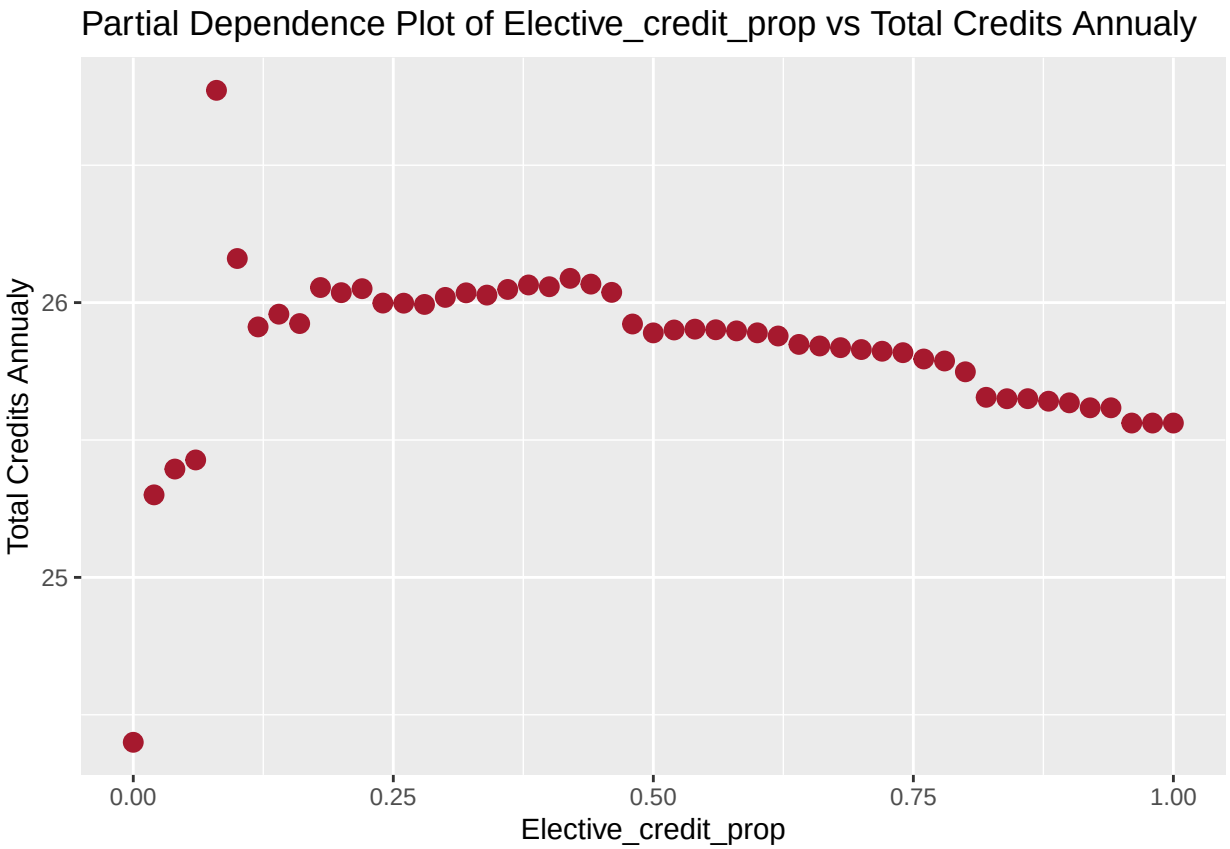

```
## [1] "Random Forest RMSE"
## [1] 5.555304
```

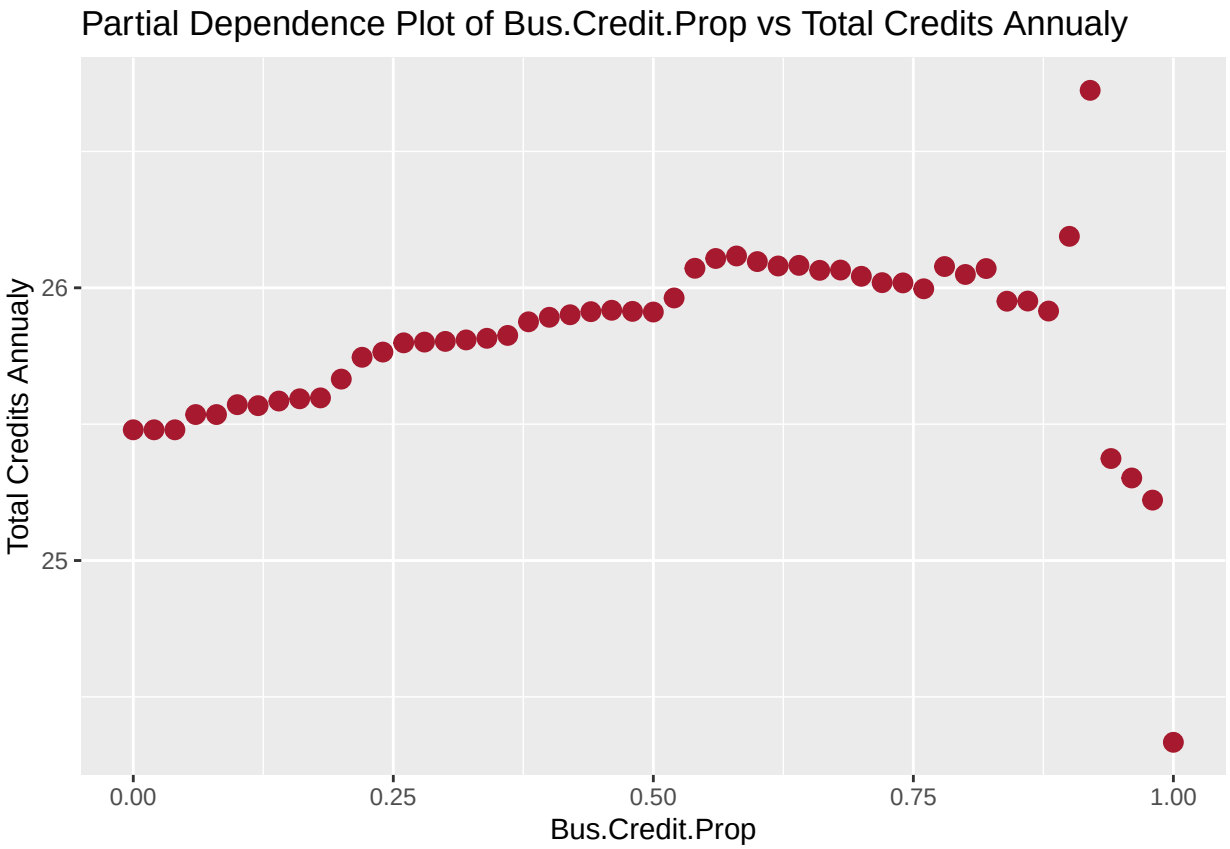
Partial Dependence Plot

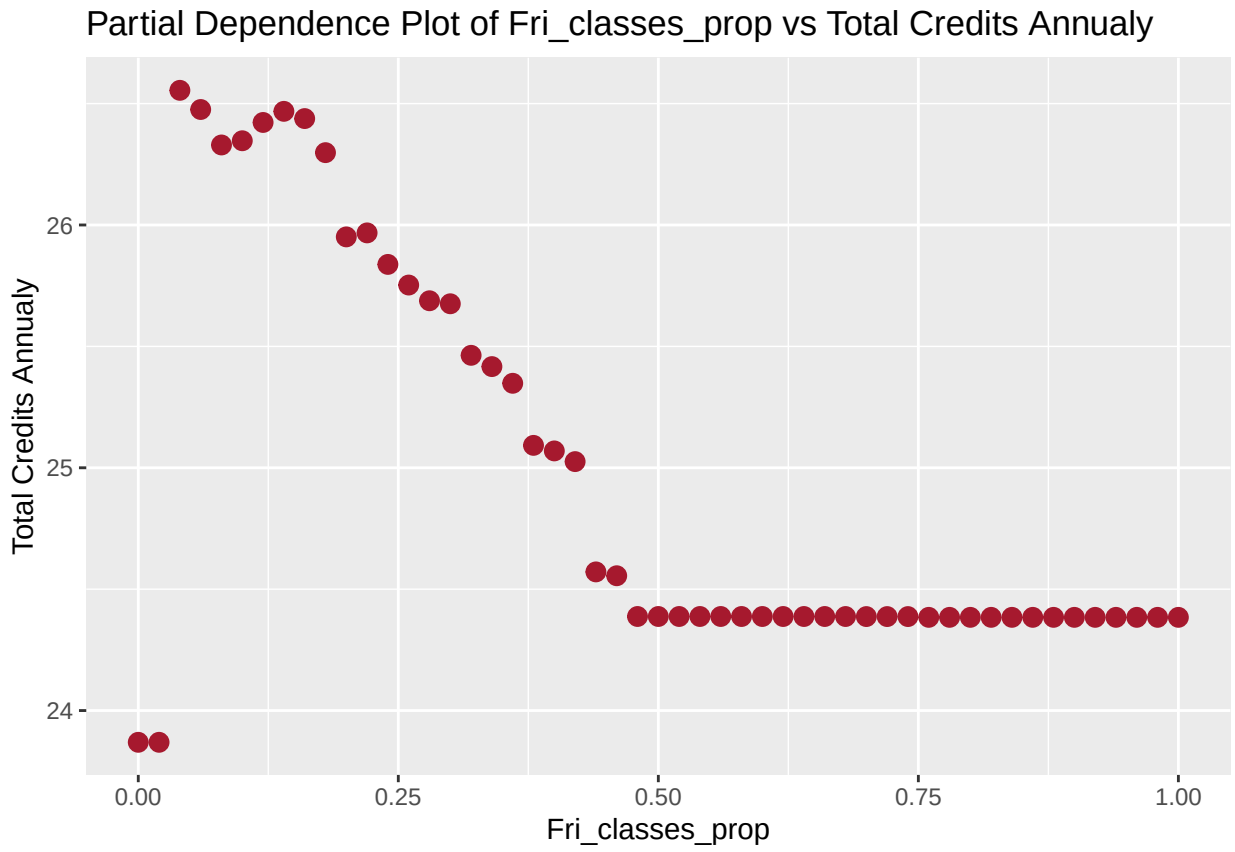
```
invisible(sapply(RF[[1]]$Feature, function(variable) {
  pdp(data = data, model = RF[[2]],
      variable = variable, target = "Total Credits Annualy")
}))
```


Partial Dependence Plot of Thurs_classes_prop vs Total Credits Annualy
Total Credits Annualy
26
25
24
23
22
0.00
0.25
0.50
0.75
1.00
Thurs_classes_prop


Partial Dependence Plot of OLC_courses_prop vs Total Credits Annualy
Total Credits Annualy
29
28
27
26
25
0.00
0.25
0.50
0.75
1.00
OLC_courses_prop


Partial Dependence Plot of Elective_credit_prop vs Total Credits Annualy
Total Credits Annualy
26
25
0.00
0.25
0.50
0.75
1.00
Elective_credit_prop


Partial Dependence Plot of Bus.Credit.Prop vs Total Credits Annualy
Total Credits Annualy
26
25
0.00
0.25
0.50
0.75
1.00
Bus.Credit.Prop



Simple Linear Regression

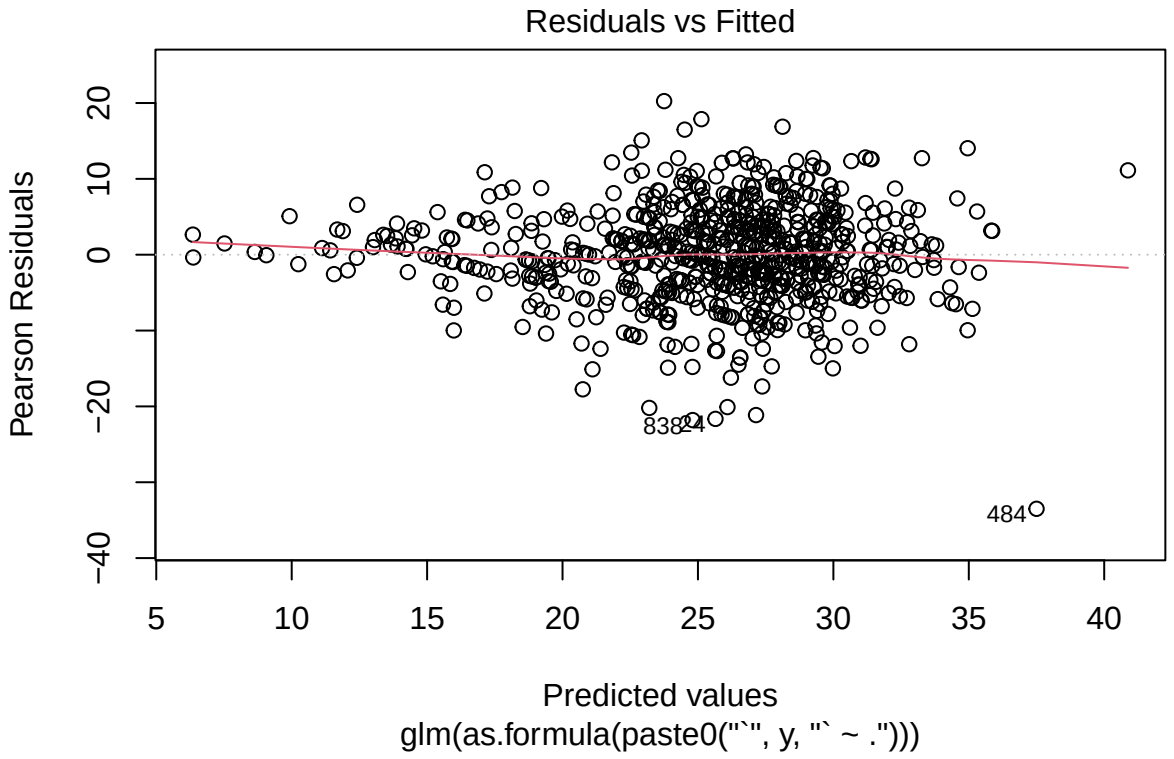
```
Credits_lm=linearModel(data,"Credits.Taken.Annualy")
```

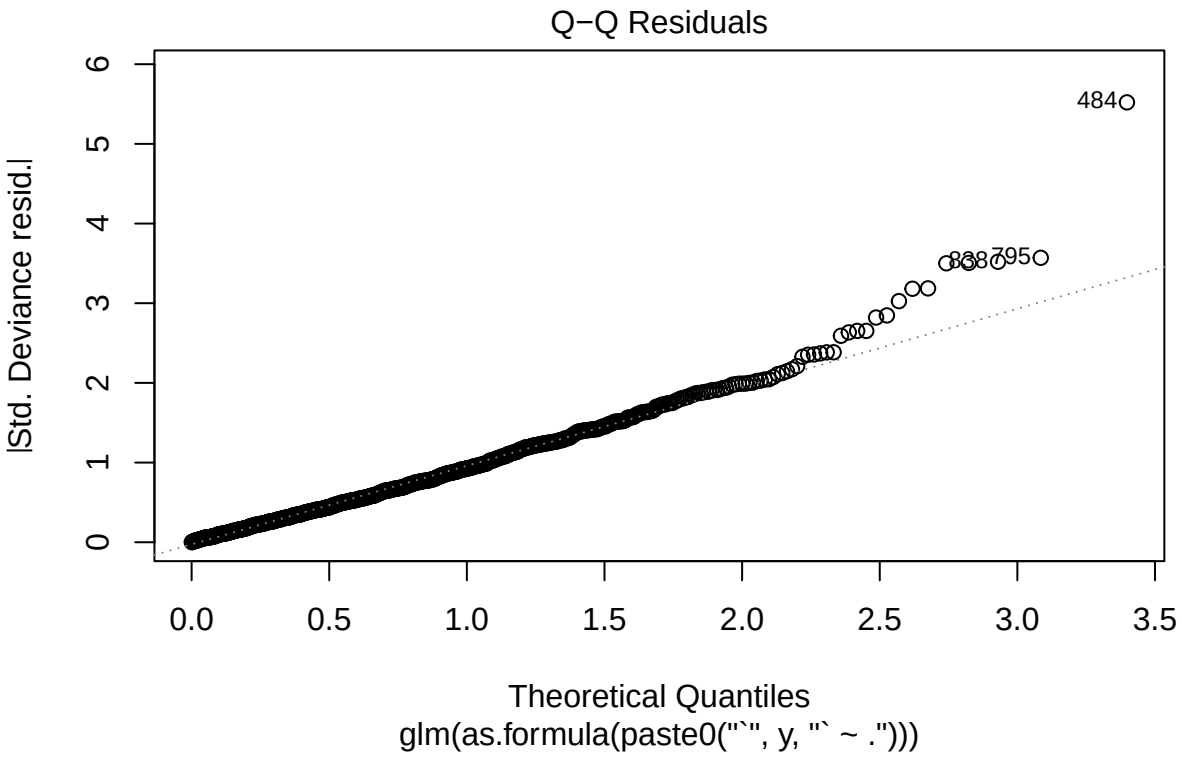
```
## [1] "Linear Regression RMSE"
## [1] 6.381795
##
## Call:
## glm(formula = as.formula(paste0("`", y, "` ~ .")), family = gaussian(link = "identity"),
##     data = data_train)
##
## Coefficients: (1 not defined because of singularities)
##                                                    Estimate Std. Error
## (Intercept)                                        -4.36127    3.94393
## HourStartMode                                       0.33259    0.09974
## MorningClassesTRUE                                 -2.29751    0.91254
## semesters_with_at_least_one_concentration_course    7.14169    0.62401
## TransferTransfer                                   -0.03494    0.52369
## TransferUnknown                                    -7.47584    4.74577
## BUSMAJ1                                            -0.71352    1.68238
## BUSJMA1                                            -0.41371    2.25085
## MINOR1                                             -1.50172    1.26102
## FINCONTRUE                                          0.48850    0.70071
## MKTGCONTRUE                                         1.49581    0.87934
## BUSCOPTRUE                                         -2.93492    0.71095
## STANCONTRUE                                         2.18046    1.14966
```

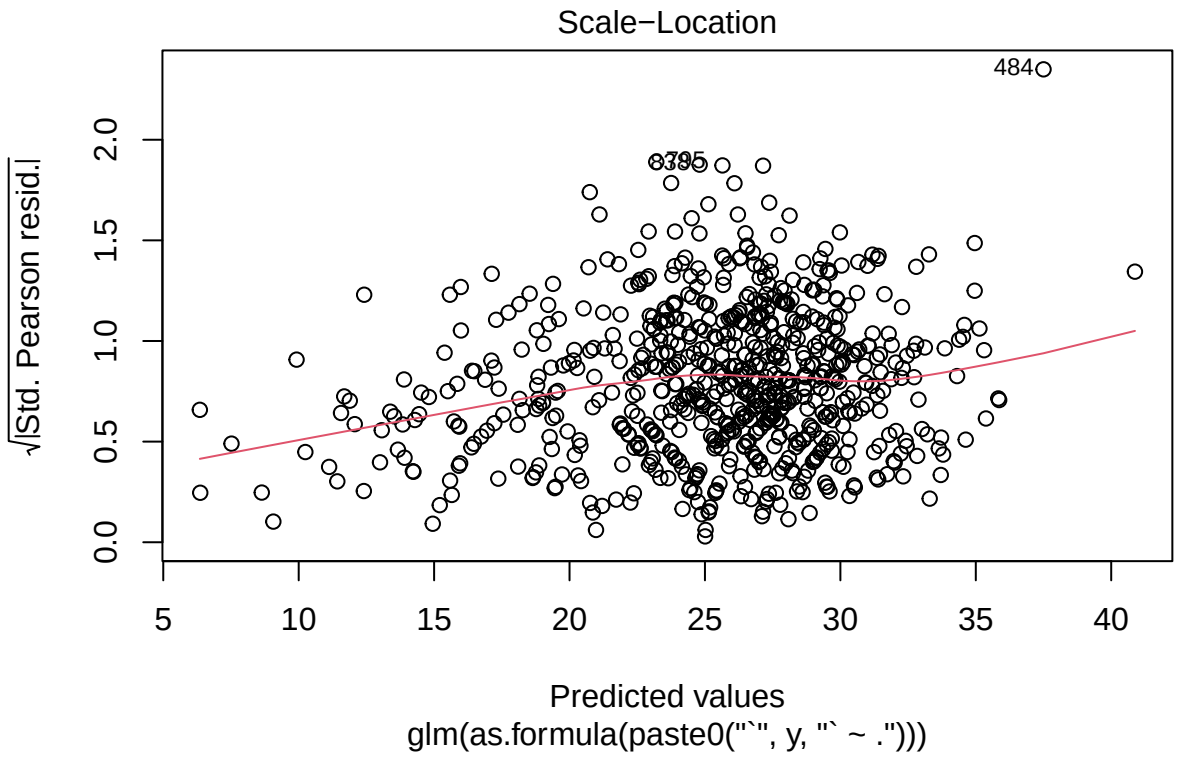
```
## ACCTCONTRUE                                              1.38070    0.79017
## MISCONTRUE                                               1.85851    0.78612
## ENTINNCONTRUE                                           -1.07380    1.51187
## OPERMGTCONTRUE                                           0.22631    0.90621
## HRMCONTRUE                                               0.87250    1.05788
## INBUCONTRUE                                              1.66985    1.16516
## DeclaredConcentration1                                   1.40803    0.92570
## At_least_one_concentration_course_per_semester_enrolled1 -6.13677    1.26469
## Mon_classes_prop                                        -8.05420    3.33124
## Tues_classes_prop                                       -5.43255    3.49089
## Wed_classes_prop                                        -4.12846    3.39385
## Thurs_classes_prop                                      -5.51387    3.66263
## Fri_classes_prop                                        -5.19538    3.43373
## MorningClassProp                                        38.86904    3.98165
## AfternoonClassProp                                      32.12271    3.88313
## OLC_courses_prop                                        25.04350    3.36114
## Concentration.Credits.Prop.of.Total.Credits            -20.05151    2.22680
## Certificate.Credits.Prop                                -3.29555    1.62122
## Bus.Credit.Prop                                         -1.38520    2.12741
## Elective_credit_prop                                        NA         NA
##                                                         t value Pr(>|t|)
## (Intercept)                                              -1.106 0.269183
## HourStartMode                                             3.335 0.000899 ***
## MorningClassesTRUE                                       -2.518 0.012032 *
## semesters_with_at_least_one_concentration_course         11.445  < 2e-16 ***
## TransferTransfer                                         -0.067 0.946831
## TransferUnknown                                          -1.575 0.115644
## BUSMAJ1                                                  -0.424 0.671613
## BUSJMA1                                                  -0.184 0.854223
## MINOR1                                                   -1.191 0.234103
## FINCONTRUE                                                0.697 0.485934
## MKTGCONTRUE                                               1.701 0.089372 .
## BUSCOPTRUE                                               -4.128 4.09e-05 ***
## STANCONTRUE                                               1.897 0.058287 .
## ACCTCONTRUE                                               1.747 0.081015 .
## MISCONTRUE                                                2.364 0.018341 *
## ENTINNCONTRUE                                            -0.710 0.477786
## OPERMGTCONTRUE                                            0.250 0.802865
## HRMCONTRUE                                                0.825 0.409782
## INBUCONTRUE                                               1.433 0.152256
## DeclaredConcentration1                                    1.521 0.128698
## At_least_one_concentration_course_per_semester_enrolled1 -4.852 1.50e-06 ***
## Mon_classes_prop                                         -2.418 0.015868 *
## Tues_classes_prop                                        -1.556 0.120108
## Wed_classes_prop                                         -1.216 0.224220
## Thurs_classes_prop                                       -1.505 0.132659
## Fri_classes_prop                                         -1.513 0.130717
## MorningClassProp                                          9.762  < 2e-16 ***
## AfternoonClassProp                                        8.272 6.55e-16 ***
## OLC_courses_prop                                          7.451 2.72e-13 ***
## Concentration.Credits.Prop.of.Total.Credits              -9.005  < 2e-16 ***
## Certificate.Credits.Prop                                 -2.033 0.042451 *
## Bus.Credit.Prop                                          -0.651 0.515181
## Elective_credit_prop                                        NA       NA
```

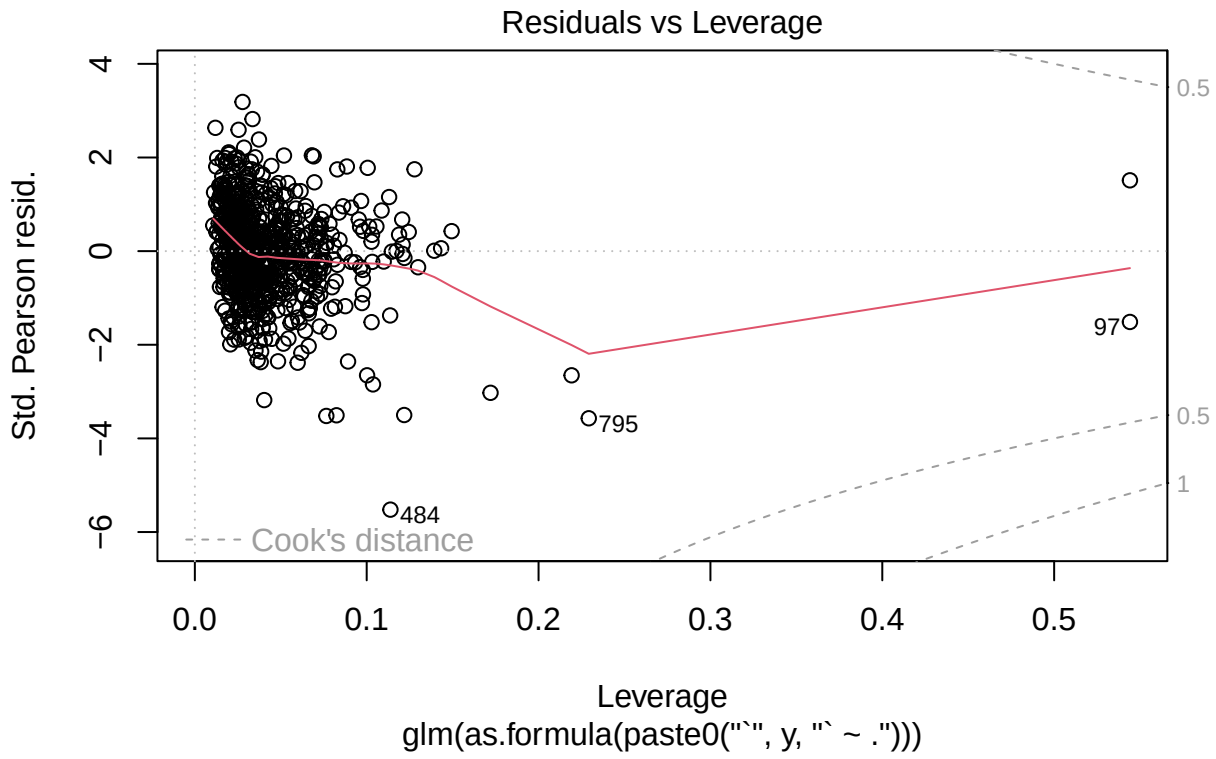
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 41.54494)
##
##     Null deviance: 46778  on 736  degrees of freedom
## Residual deviance: 29289  on 705  degrees of freedom
## AIC: 4871.4
##
## Number of Fisher Scoring iterations: 2
```

```
plot(Credits_lm)
```




Q−Q Residuals
|Std. Deviance resid.|
0
1
2
3
4
5
6
484
795
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Theoretical Quantiles
glm(as.formula(paste0("`", y, "` ~ .")))


Scale−Location
484
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
5
10
15
20
25
30
35
40
Predicted values
glm(as.formula(paste0("`", y, "` ~ .")))

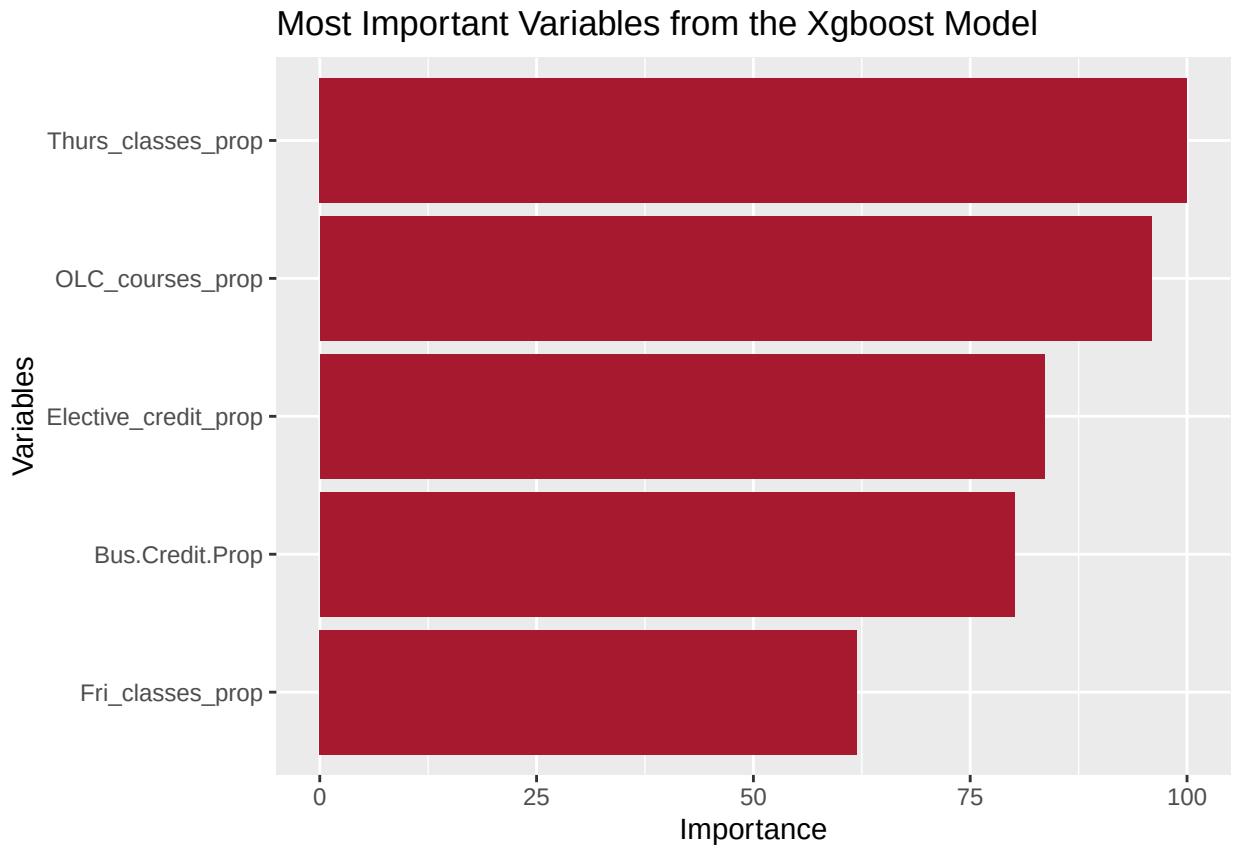



```
stepwise(Credits_lm,"Credits.Taken.Annualy",data)
```

```
##
## Call:
## glm(formula = Credits.Taken.Annualy ~ HourStartMode + MorningClasses +
##     semesters_with_at_least_one_concentration_course + MKTGCON +
##     BUSCOP + STANCON + ACCTCON + MISCON + DeclaredConcentration +
##     At_least_one_concentration_course_per_semester_enrolled +
##     Mon_classes_prop + MorningClassProp + AfternoonClassProp +
##     OLC_courses_prop + Concentration.Credits.Prop.of.Total.Credits +
##     Certificate.Credits.Prop, family = gaussian(link = "identity"),
##     data = data_train)
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       -6.20395    2.70330
## HourStartMode                                      0.32347    0.09789
## MorningClassesTRUE                                -2.42939    0.88771
## semesters_with_at_least_one_concentration_course   7.33103    0.61618
## MKTGCONTRUE                                        1.26198    0.75938
## BUSCOPTRUE                                        -3.14120    0.68385
## STANCONTRUE                                        2.28382    1.13469
## ACCTCONTRUE                                        1.18452    0.66351
## MISCONTRUE                                         1.60650    0.71641
## DeclaredConcentration1                             1.79489    0.73533
## At_least_one_concentration_course_per_semester_enrolled1  -5.98004    1.23209
```

```
## Mon_classes_prop                                          -4.31840    2.25674
## MorningClassProp                                          33.83828    2.87439
## AfternoonClassProp                                        27.43720    2.77937
## OLC_courses_prop                                          26.30353    2.57022
## Concentration.Credits.Prop.of.Total.Credits              -20.25913    2.12004
## Certificate.Credits.Prop                                  -3.48893    1.47826
##                                                          t value Pr(>|t|)
## (Intercept)                                               -2.295 0.022023 *
## HourStartMode                                              3.304 0.000999 ***
## MorningClassesTRUE                                        -2.737 0.006359 **
## semesters_with_at_least_one_concentration_course          11.898  < 2e-16 ***
## MKTGCONTRUE                                                1.662 0.096975 .
## BUSCOPTRUE                                                -4.593 5.15e-06 ***
## STANCONTRUE                                                2.013 0.044515 *
## ACCTCONTRUE                                                1.785 0.074646 .
## MISCONTRUE                                                 2.242 0.025237 *
## DeclaredConcentration1                                     2.441 0.014889 *
## At_least_one_concentration_course_per_semester_enrolled1  -4.854 1.49e-06 ***
## Mon_classes_prop                                          -1.914 0.056073 .
## MorningClassProp                                          11.772  < 2e-16 ***
## AfternoonClassProp                                         9.872  < 2e-16 ***
## OLC_courses_prop                                          10.234  < 2e-16 ***
## Concentration.Credits.Prop.of.Total.Credits               -9.556  < 2e-16 ***
## Certificate.Credits.Prop                                  -2.360 0.018533 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 41.32015)
##
##     Null deviance: 46778  on 736  degrees of freedom
## Residual deviance: 29751  on 720  degrees of freedom
## AIC: 4853
##
## Number of Fisher Scoring iterations: 2
##
## [1] "Stepwise Regression RMSE"
## [1] 6.320161
```

```
xg(data,"Credits.Taken.Annualy")
```




```
## [1] "Xgboost RMSE"
## [1] 5.555304
```

```
## [[1]]
##                Feature   Overall
## 1   Thurs_classes_prop 100.00000
## 2     OLC_courses_prop  95.88293
## 3 Elective_credit_prop  83.53808
## 4      Bus.Credit.Prop  80.09916
## 5     Fri_classes_prop  61.94993
##
## [[2]]
## Random Forest
##
## 737 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 664, 665, 663, 662, 662, 663, ...
## Resampling results across tuning parameters:
##
##   mtry  min.node.size  RMSE      Rsquared   MAE
##    3     1             5.599787  0.5822358  4.472177
##    3     3             5.611616  0.5809456  4.482014
##    3     5             5.616260  0.5809240  4.500782
##    3    10             5.651520  0.5763800  4.526742
```

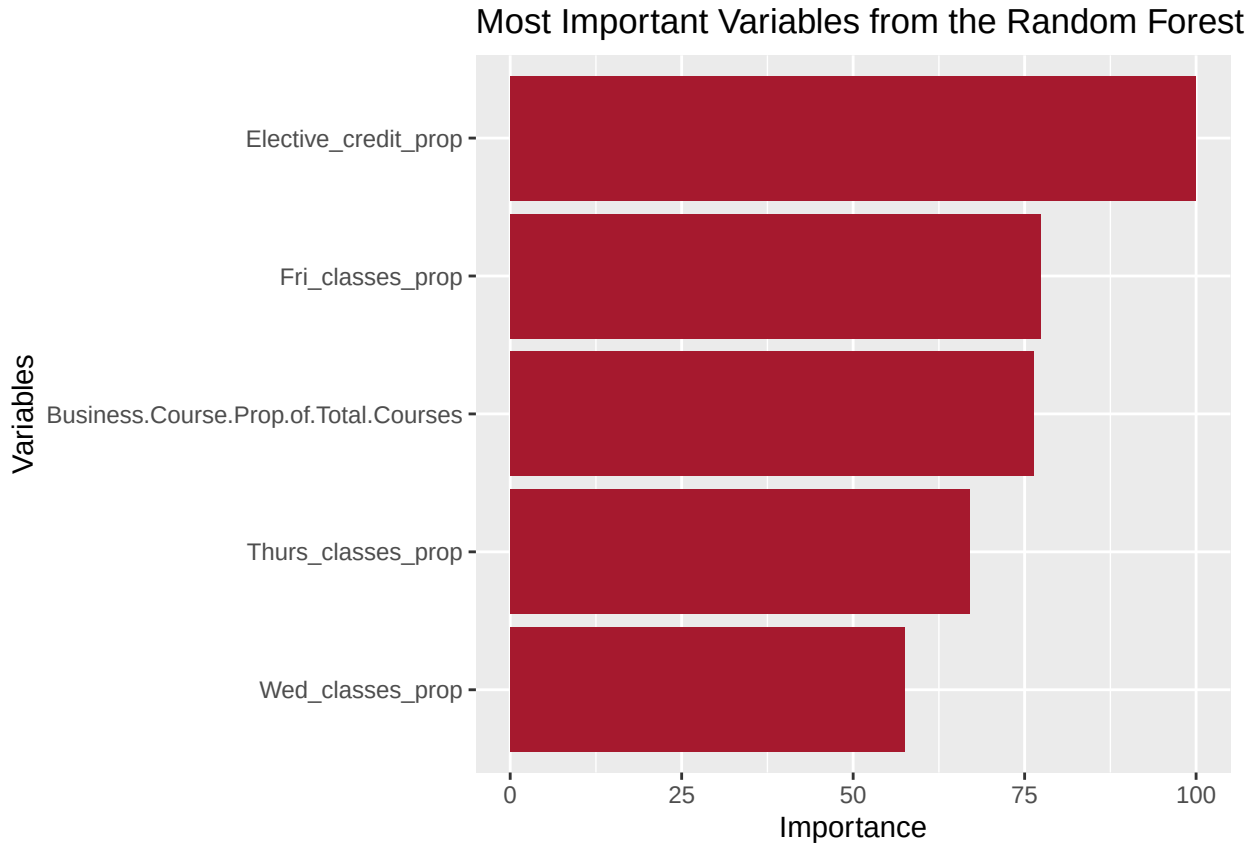
```
##     5     1             5.456211  0.5868285  4.327361
##     5     3             5.469548  0.5840046  4.338944
##     5     5             5.478162  0.5824807  4.352465
##     5    10             5.496671  0.5807983  4.373881
##    10     1             5.395329  0.5786380  4.238665
##    10     3             5.398346  0.5783643  4.234895
##    10     5             5.388420  0.5799455  4.230667
##    10    10             5.409778  0.5769342  4.254852
##
## Tuning parameter 'splitrule' was held constant at a value of variance
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were mtry = 10, splitrule = variance
##  and min.node.size = 5.
```

Annual Courses Taken

Random Forest and Important Variables

```
data1=courseFeatureSelection(data1)
```

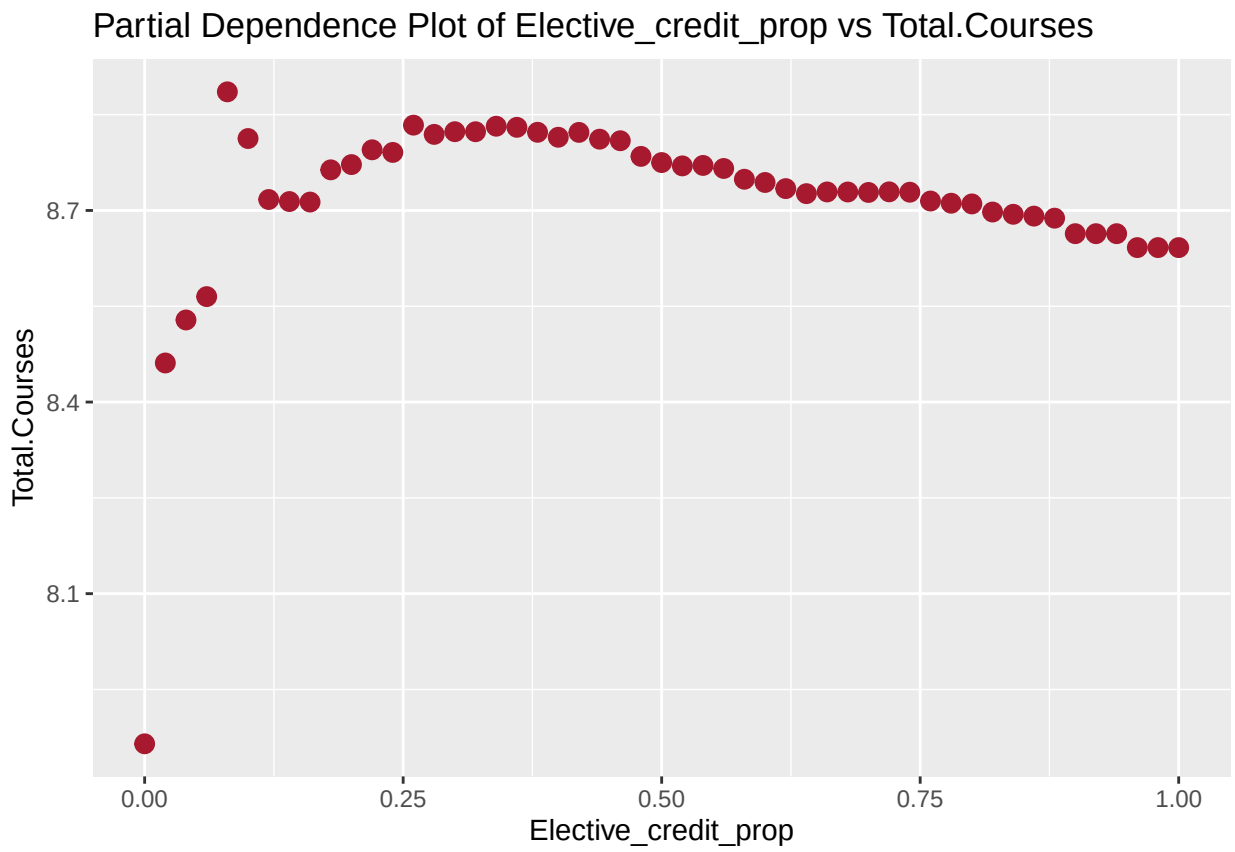
```
RF_Courses=rf(data1,"Total.Courses")
```

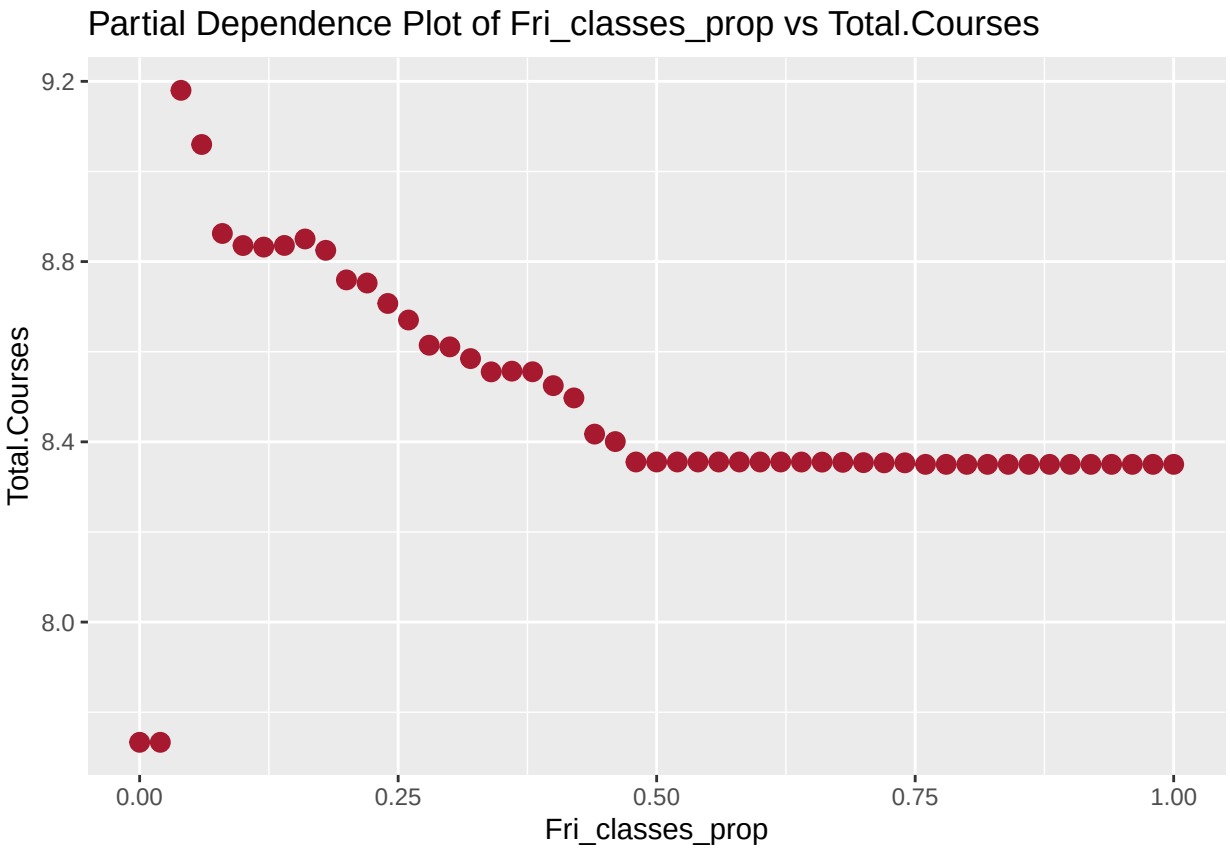


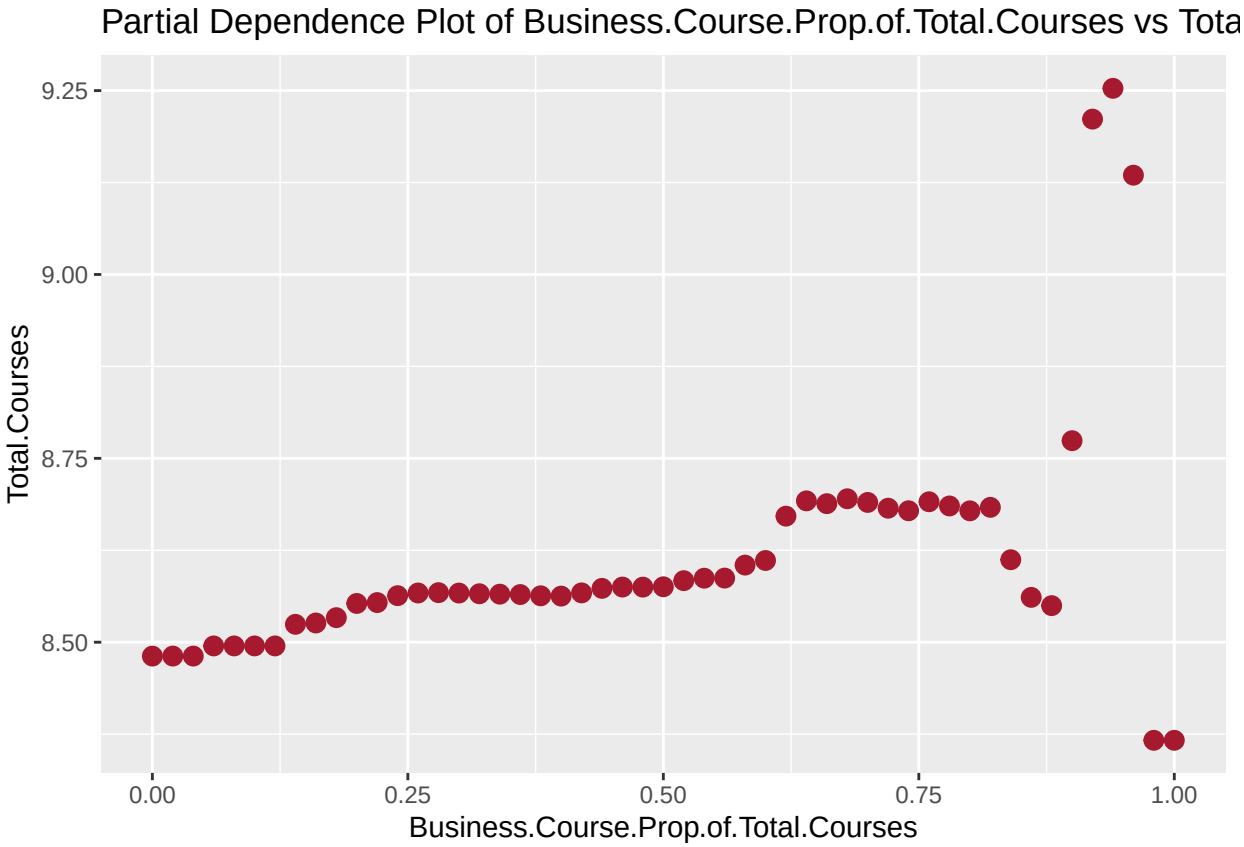

```
## [1] "Random Forest RMSE"
## [1] 1.826375
```

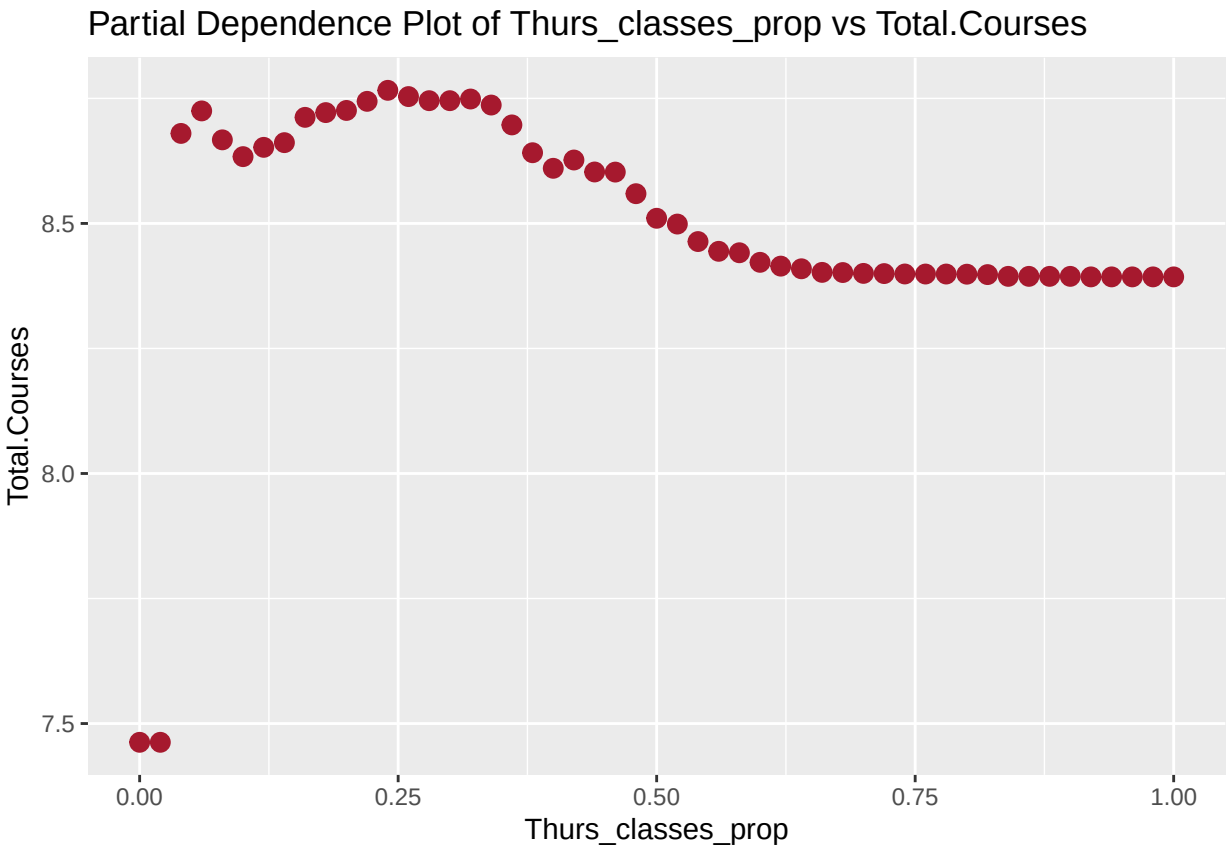
Partial Dependence Plot

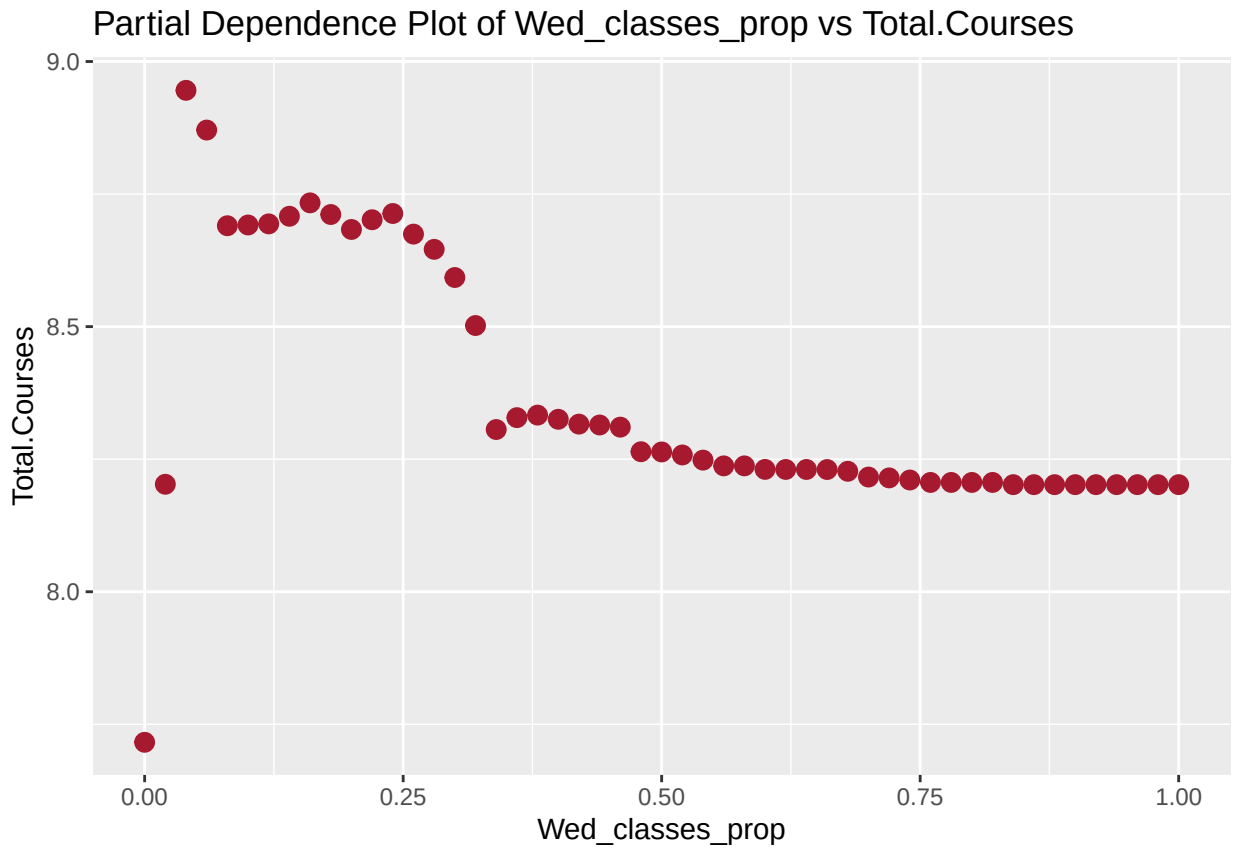
```
invisible(sapply(RF_Courses[[1]]$Feature, function(variable) {
  pdp(data = data1, model = RF_Courses[[2]],
      variable = variable, target = "Total.Courses")
}))
```




Partial Dependence Plot of Fri_classes_prop vs Total.Courses
Total.Courses
9.2
8.8
8.4
8.0
0.00
0.25
0.50
0.75
1.00
Fri_classes_prop


Partial Dependence Plot of Business.Course.Prop.of.Total.Courses vs Tota
Total.Courses
9.25
9.00
8.75
8.50
0.00
0.25
0.50
0.75
1.00
Business.Course.Prop.of.Total.Courses


Partial Dependence Plot of Thurs_classes_prop vs Total.Courses
Total.Courses
8.5
8.0
7.5
0.00
0.25
0.50
0.75
1.00
Thurs_classes_prop



Simple Linear Regression

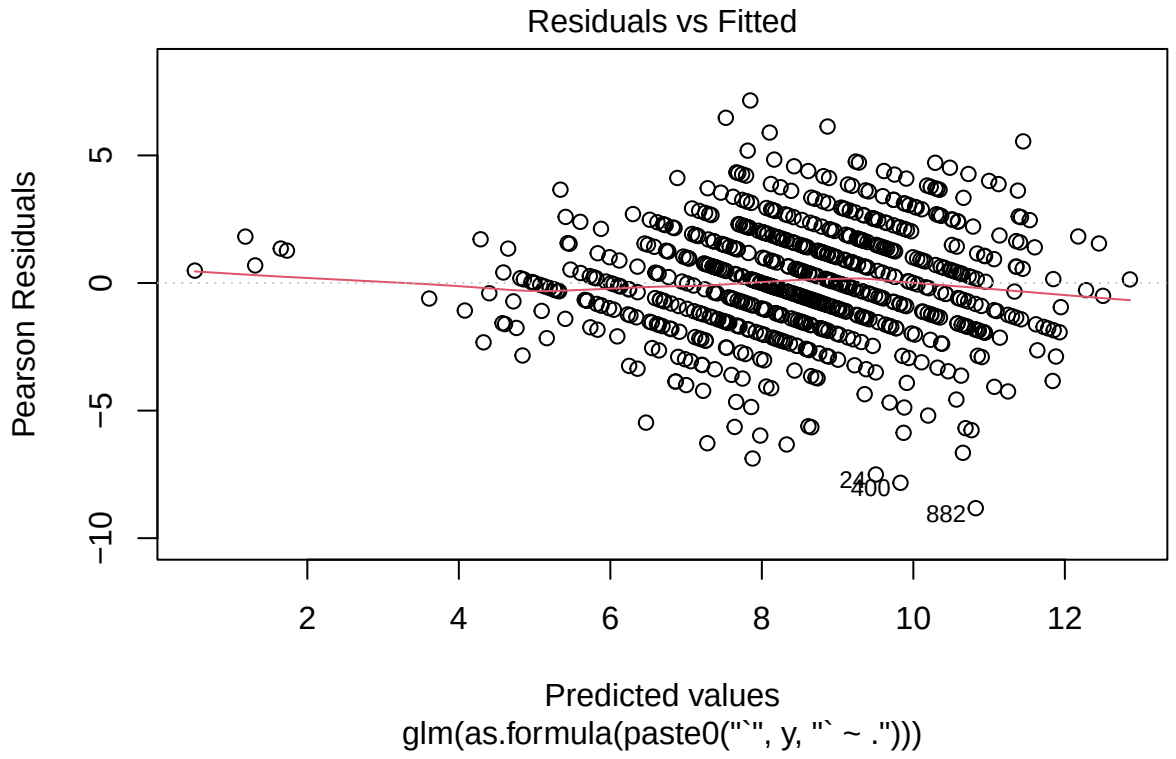
```
Course_lm=linearModel(data1,"Total.Courses")
```

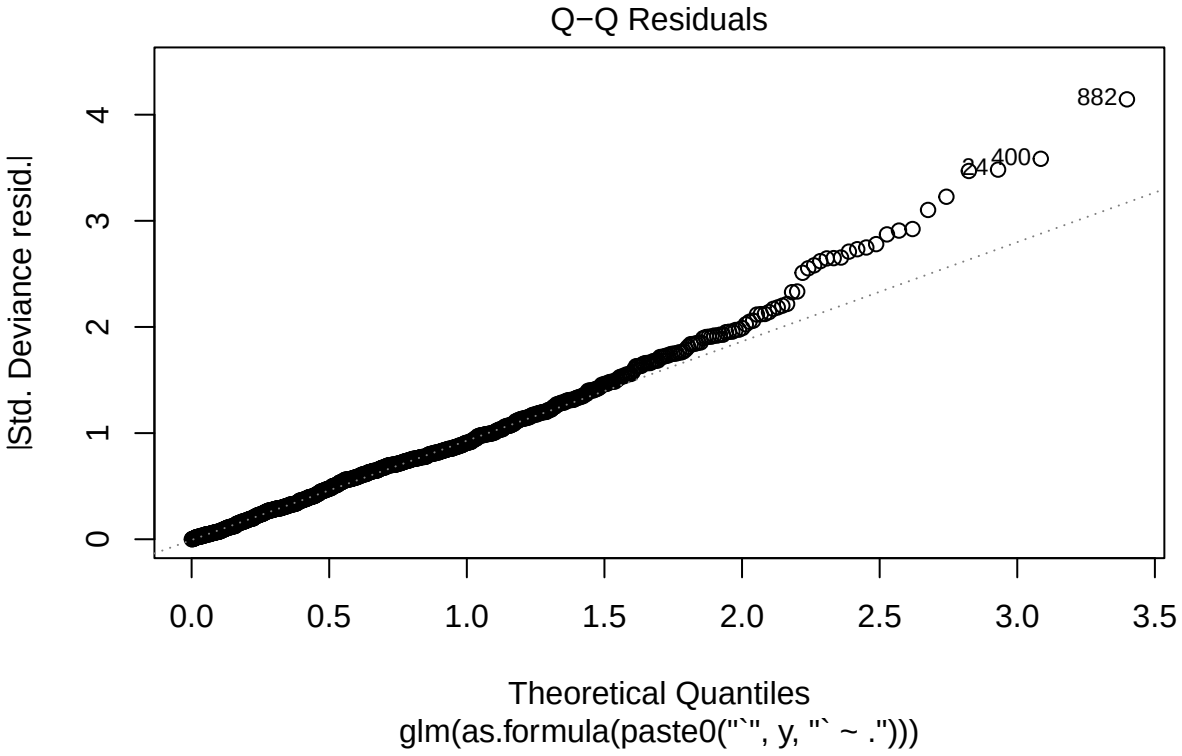
```
## [1] "Linear Regression RMSE"
## [1] 2.392846
##
## Call:
## glm(formula = as.formula(paste0("`", y, "` ~ .")), family = gaussian(link = "identity"),
##     data = data_train)
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                     -0.49103    4.14673
## HourStartMode                                    0.16238    0.03437
## MorningClassesTRUE                              -0.77070    0.30749
## semesters_with_at_least_one_concentration_course 2.55600    0.21391
## TransferTransfer                                 0.00433    0.18190
## TransferUnknown                                 -1.20353    1.21146
## BUSMAJ1                                         -0.07054    0.66115
## BUSJMA1                                         -1.22673    0.85283
## MINOR1                                          -0.96723    0.40409
## FINCONTRUE                                       0.14922    0.24686
## MKTGCONTRUE                                      0.29768    0.28630
## BUSCOPTRUE                                      -1.15087    0.24526
## STANCONTRUE                                      0.71938    0.41538
## ACCTCONTRUE                                      0.45807    0.27722
```

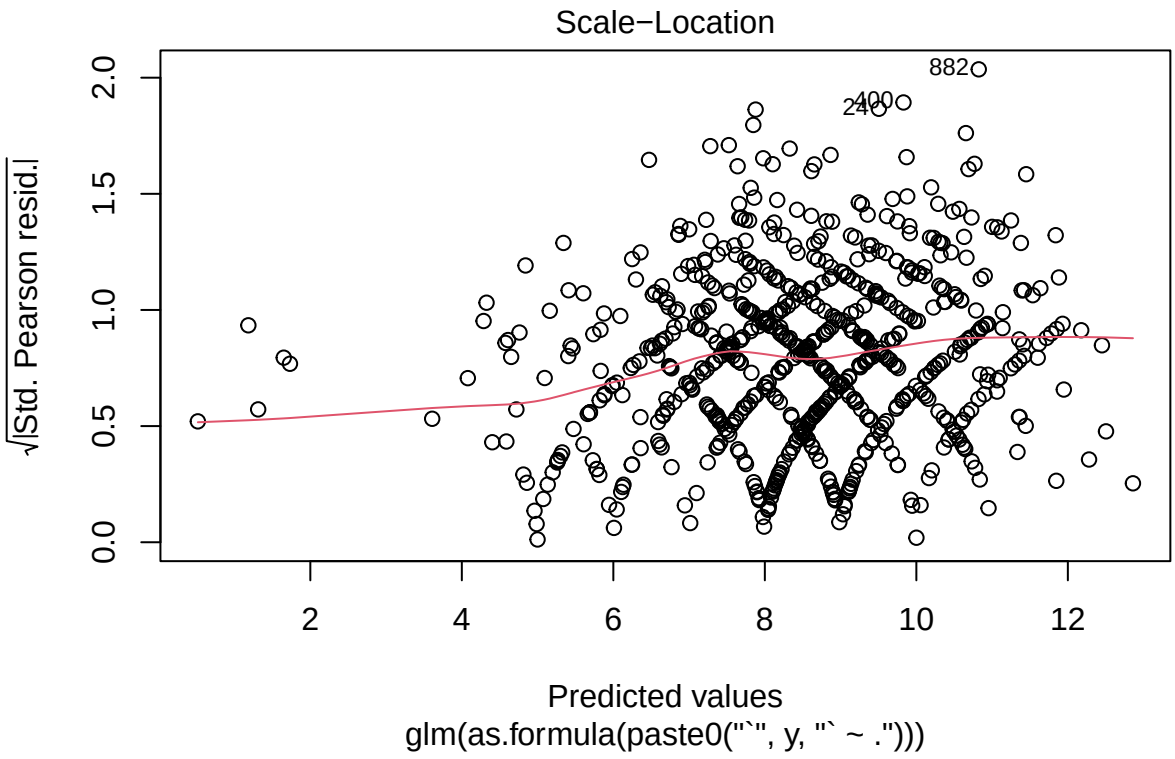
```
## MISCONTRUE                                                 0.42805    0.27810
## ENTINNCONTRUE                                             -0.70707    0.57843
## OPERMGTCONTRUE                                            -0.12944    0.30830
## HRMCONTRUE                                                 0.06714    0.36636
## INBUCONTRUE                                                0.55687    0.41208
## DeclaredConcentration1                                     0.81503    0.32413
## At_least_one_concentration_course_per_semester_enrolled1 -1.61280    0.45371
## Mon_classes_prop                                          -7.26282    1.13058
## Tues_classes_prop                                         -6.81432    1.19657
## Wed_classes_prop                                          -5.65893    1.15288
## Thurs_classes_prop                                        -6.63688    1.25567
## Fri_classes_prop                                          -4.51335    1.13223
## MorningClassProp                                          12.89300    1.35468
## AfternoonClassProp                                        10.06755    1.31568
## Concentration.Courses.Prop.of.Total.Courses               -7.51336    0.74367
## Certificate.Courses.Prop.of.Total.Courses                 -0.89731    0.54517
## Business.Course.Prop.of.Total.Courses                      2.13743    4.07623
## Elective_credit_prop                                       5.90049    3.97968
##                                                           t value Pr(>|t|)
## (Intercept)                                                -0.118 0.905774
## HourStartMode                                               4.724 2.79e-06 ***
## MorningClassesTRUE                                         -2.506 0.012421 *
## semesters_with_at_least_one_concentration_course           11.949  < 2e-16 ***
## TransferTransfer                                            0.024 0.981014
## TransferUnknown                                            -0.993 0.320831
## BUSMAJ1                                                    -0.107 0.915063
## BUSJMA1                                                    -1.438 0.150757
## MINOR1                                                     -2.394 0.016944 *
## FINCONTRUE                                                  0.604 0.545740
## MKTGCONTRUE                                                 1.040 0.298810
## BUSCOPTRUE                                                 -4.692 3.24e-06 ***
## STANCONTRUE                                                 1.732 0.083733 .
## ACCTCONTRUE                                                 1.652 0.098900 .
## MISCONTRUE                                                  1.539 0.124209
## ENTINNCONTRUE                                              -1.222 0.221966
## OPERMGTCONTRUE                                             -0.420 0.674732
## HRMCONTRUE                                                  0.183 0.854638
## INBUCONTRUE                                                 1.351 0.177014
## DeclaredConcentration1                                      2.515 0.012140 *
## At_least_one_concentration_course_per_semester_enrolled1   -3.555 0.000404 ***
## Mon_classes_prop                                           -6.424 2.44e-10 ***
## Tues_classes_prop                                          -5.695 1.81e-08 ***
## Wed_classes_prop                                           -4.909 1.14e-06 ***
## Thurs_classes_prop                                         -5.286 1.67e-07 ***
## Fri_classes_prop                                           -3.986 7.41e-05 ***
## MorningClassProp                                            9.517  < 2e-16 ***
## AfternoonClassProp                                          7.652 6.51e-14 ***
## Concentration.Courses.Prop.of.Total.Courses               -10.103  < 2e-16 ***
## Certificate.Courses.Prop.of.Total.Courses                  -1.646 0.100224
## Business.Course.Prop.of.Total.Courses                       0.524 0.600190
## Elective_credit_prop                                        1.483 0.138613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

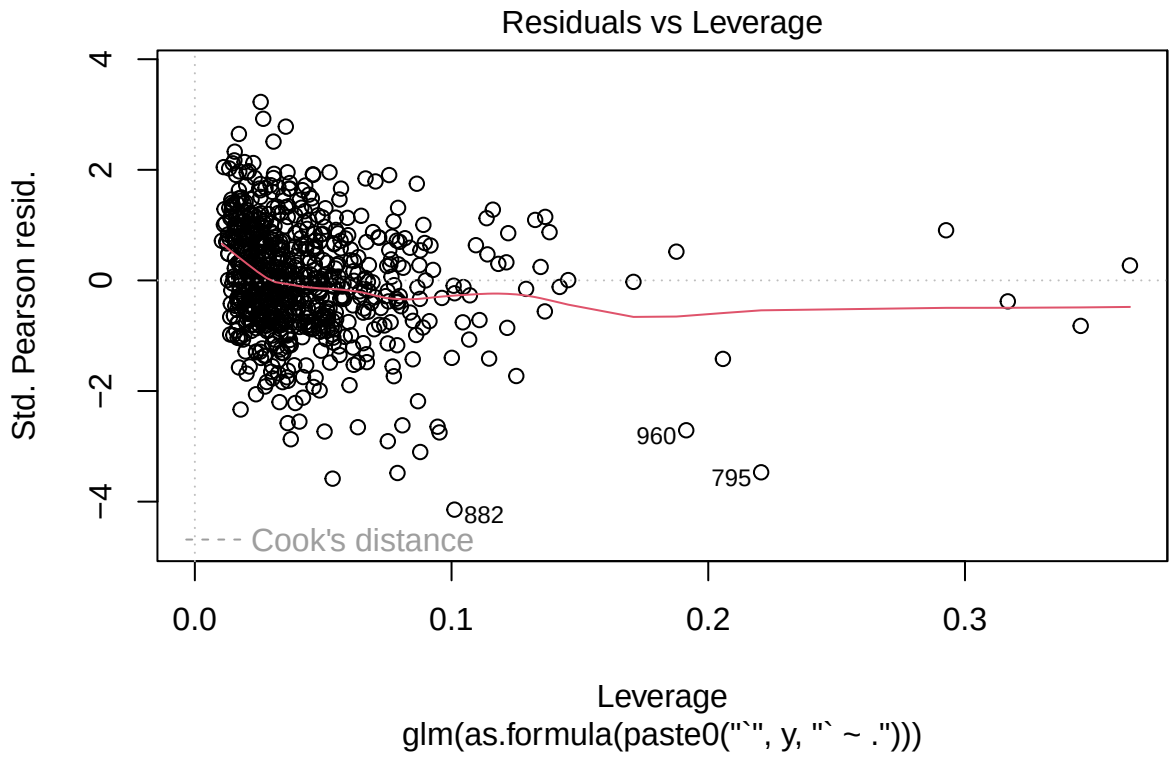
```
## (Dispersion parameter for gaussian family taken to be 5.042521)
##
##     Null deviance: 5695.9  on 737  degrees of freedom
## Residual deviance: 3560.0  on 706  degrees of freedom
## AIC: 3321.7
##
## Number of Fisher Scoring iterations: 2
```

```
plot(Course_lm)
```




Q−Q Residuals
|Std. Deviance resid.|
0
1
2
3
4
882
400
240
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Theoretical Quantiles
glm(as.formula(paste0("`", y, "` ~ .")))


Scale−Location
882
400
240
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2
4
6
8
10
12
Predicted values
glm(as.formula(paste0("`", y, "` ~ .")))




```
stepwise(Course_lm,"Total.Courses",data1)
```

```
##
## Call:
## glm(formula = Total.Courses ~ HourStartMode + MorningClasses +
##     semesters_with_at_least_one_concentration_course + BUSJMA +
##     MINOR + BUSCOP + ENTINNCON + DeclaredConcentration + At_least_one_concentration_course_per_semes
##     Mon_classes_prop + Tues_classes_prop + Wed_classes_prop +
##     Thurs_classes_prop + Fri_classes_prop + MorningClassProp +
##     AfternoonClassProp + Concentration.Courses.Prop.of.Total.Courses +
##     Certificate.Courses.Prop.of.Total.Courses + Elective_credit_prop,
##     family = gaussian(link = "identity"), data = data_train)
##
## Coefficients:
##                                                        Estimate Std. Error
## (Intercept)                                             1.50568    0.80426
## HourStartMode                                           0.16341    0.03414
## MorningClassesTRUE                                     -0.75932    0.30399
## semesters_with_at_least_one_concentration_course        2.58444    0.20482
## BUSJMA1                                                -1.22412    0.54834
## MINOR1                                                 -0.99210    0.39773
## BUSCOPTRUE                                             -1.25904    0.23670
## ENTINNCONTRUE                                          -0.86246    0.56053
## DeclaredConcentration1                                  1.18583    0.22988
## At_least_one_concentration_course_per_semester_enrolled1 -1.54394    0.44541
## Mon_classes_prop                                       -7.36829    1.08169
```

```
## Tues_classes_prop                                          -6.94453    1.13432
## Wed_classes_prop                                           -5.90198    1.10978
## Thurs_classes_prop                                         -6.76610    1.19941
## Fri_classes_prop                                           -4.61041    1.08314
## MorningClassProp                                           13.17815    1.28845
## AfternoonClassProp                                         10.29493    1.25725
## Concentration.Courses.Prop.of.Total.Courses                -7.37495    0.71192
## Certificate.Courses.Prop.of.Total.Courses                  -0.89740    0.49627
## Elective_credit_prop                                        3.68034    0.56398
##                                                           t value Pr(>|t|)
## (Intercept)                                                 1.872 0.061594 .
## HourStartMode                                               4.787 2.06e-06 ***
## MorningClassesTRUE                                         -2.498 0.012718 *
## semesters_with_at_least_one_concentration_course           12.618  < 2e-16 ***
## BUSJMA1                                                    -2.232 0.025896 *
## MINOR1                                                     -2.494 0.012839 *
## BUSCOPTRUE                                                 -5.319 1.39e-07 ***
## ENTINNCONTRUE                                              -1.539 0.124332
## DeclaredConcentration1                                      5.158 3.22e-07 ***
## At_least_one_concentration_course_per_semester_enrolled1   -3.466 0.000559 ***
## Mon_classes_prop                                           -6.812 2.04e-11 ***
## Tues_classes_prop                                          -6.122 1.52e-09 ***
## Wed_classes_prop                                           -5.318 1.40e-07 ***
## Thurs_classes_prop                                         -5.641 2.43e-08 ***
## Fri_classes_prop                                           -4.257 2.35e-05 ***
## MorningClassProp                                           10.228  < 2e-16 ***
## AfternoonClassProp                                          8.188 1.21e-15 ***
## Concentration.Courses.Prop.of.Total.Courses               -10.359  < 2e-16 ***
## Certificate.Courses.Prop.of.Total.Courses                  -1.808 0.070979 .
## Elective_credit_prop                                        6.526 1.28e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 5.019249)
##
##     Null deviance: 5695.9  on 737  degrees of freedom
## Residual deviance: 3603.8  on 718  degrees of freedom
## AIC: 3306.7
##
## Number of Fisher Scoring iterations: 2
##
## [1] "Stepwise Regression RMSE"
## [1] 2.420863
```