Students' Academic Performance

Load Libraries:

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(dplyr)

##
## Attaching package: 'dplyr'

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

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

library(randomForest)

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

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

## The following object is masked from 'package:ggplot2':
##
##     margin

library(class)
library(rpart)
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 3.3.2

library(e1071)
library(caret)

## Warning: package 'caret' was built under R version 3.3.2

## Loading required package: lattice
```

```
library(caTools)
library(party)

## Warning: package 'party' was built under R version 3.3.2

## Loading required package: grid

## Loading required package: mvtnorm

## Warning: package 'mvtnorm' was built under R version 3.3.2

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 3.3.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.3.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.3.2
```

Reading the Data

```
edu <- read.csv('./Edudata.csv')
str(edu)

## 'data.frame':    480 obs. of  17 variables:
##  $ gender                  : Factor w/ 2 levels "F","M": 2 2 2 2 2 1
2 2 1 1 ...
##  $ NationalITy             : Factor w/ 14 levels "Egypt","Iran",..:
5 5 5 5 5 5 5 5 5 5 ...
##  $ PlaceofBirth            : Factor w/ 14 levels "Egypt","Iran",..:
5 5 5 5 5 5 5 5 5 5 ...
##  $ StageID                 : Factor w/ 3 levels
"HighSchool","lowerlevel",..: 2 2 2 2 2 2 3 3 3 3 ...
##  $ GradeID                 : Factor w/ 10 levels "G-02","G-04",..: 2
2 2 2 2 2 5 5 5 5 ...
##  $ SectionID               : Factor w/ 3 levels "A","B","C": 1 1 1 1
1 1 1 1 1 2 ...
##  $ Topic                   : Factor w/ 12 levels
```

```
"Arabic","Biology",..: 8 8 8 8 8 8 9 9 9 8 ...
##  $ Semester                : Factor w/ 2 levels "F","S": 1 1 1 1 1 1
1 1 1 1 ...
##  $ Relation                : Factor w/ 2 levels "Father","Mum": 1 1
1 1 1 1 1 1 1 1 ...
##  $ raisedhands             : int  15 20 10 30 40 42 35 50 12 70 ...
##  $ VisITedResources        : int  16 20 7 25 50 30 12 10 21 80 ...
##  $ AnnouncementsView       : int  2 3 0 5 12 13 0 15 16 25 ...
##  $ Discussion              : int  20 25 30 35 50 70 17 22 50 70 ...
##  $ ParentAnsweringSurvey   : Factor w/ 2 levels "No","Yes": 2 2 1 1
1 2 1 2 2 2 ...
##  $ ParentschoolSatisfaction: Factor w/ 2 levels "Bad","Good": 2 2 1
1 1 1 1 2 2 2 ...
##  $ StudentAbsenceDays      : Factor w/ 2 levels "Above-7","Under-7":
2 2 1 1 1 1 1 2 2 2 ...
##  $ Class                   : Factor w/ 3 levels "H","L","M": 3 3 2 2
3 3 2 3 3 3 ...
```

```
summary(edu)
```

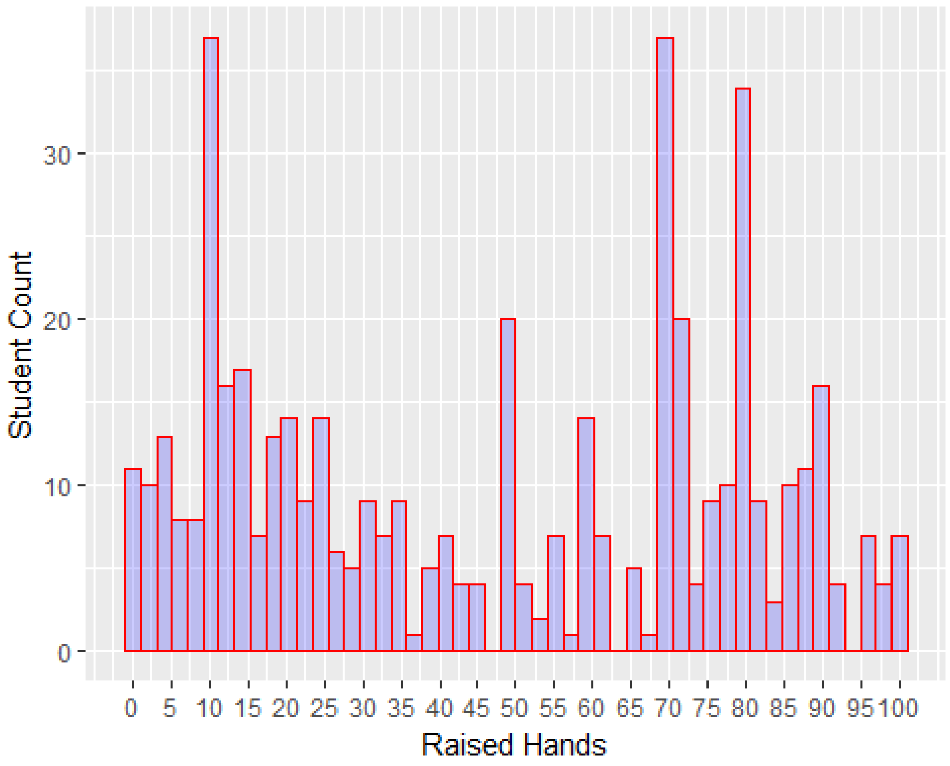
```
##  gender     NationalITy       PlaceofBirth         StageID   
##  F:175   KW        :179   KuwaIT     :180   HighSchool  : 33  
##  M:305   Jordan    :172   Jordan     :176   lowerlevel  :199  
##          Palestine: 28   Iraq       : 22   MiddleSchool:248  
##          Iraq      : 22   lebanon    : 19                     
##          lebanon   : 17   SaudiArabia: 16                     
##          Tunis     : 12   USA        : 16                     
##          (Other)   : 50   (Other)    : 51                     
##     GradeID    SectionID     Topic     Semester   Relation  
##  G-02   :147   A:283     IT     : 95   F:245    Father:283  
##  G-08   :116   B:167     French : 65   S:235    Mum   :197  
##  G-07   :101   C: 30     Arabic : 59                        
##  G-04   : 48             Science: 51                        
##  G-06   : 32             English: 45                        
##  G-11   : 13             Biology: 30                        
##  (Other): 23             (Other):135                        
##   raisedhands     VisITedResources AnnouncementsView   Discussion   
##  Min.   :  0.00   Min.   : 0.0     Min.   : 0.00     Min.   : 1.00  
##  1st Qu.: 15.75   1st Qu.:20.0     1st Qu.:14.00     1st Qu.:20.00  
##  Median : 50.00   Median :65.0     Median :33.00     Median :39.00  
##  Mean   : 46.77   Mean   :54.8     Mean   :37.92     Mean   :43.28  
##  3rd Qu.: 75.00   3rd Qu.:84.0     3rd Qu.:58.00     3rd Qu.:70.00  
##  Max.   :100.00   Max.   :99.0     Max.   :98.00     Max.   :99.00  
##                                                                     
##  ParentAnsweringSurvey ParentschoolSatisfaction StudentAbsenceDays
Class 
##  No :210               Bad :188                 Above-7:191       
H:142  
##  Yes:270               Good:292                 Under-7:289       
L:127  
```

```
##
M:211
##

##

##

##
```

Exploratory Data Analysis

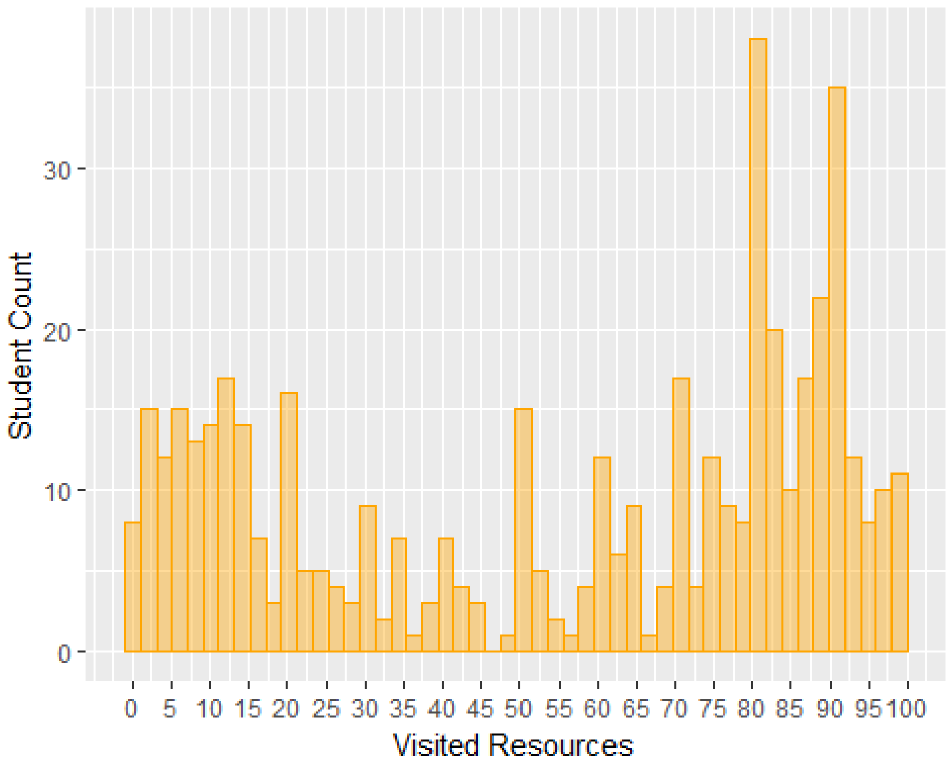
Raised Hands

```
ggplot(edu, aes(x = raisedhands)) + geom_histogram(bins=50, color =
"red",fill="blue",alpha=0.2) +
  scale_x_continuous(breaks = seq(0,100,5)) +
  labs(x = "Raised Hands", y = "Student Count")
```



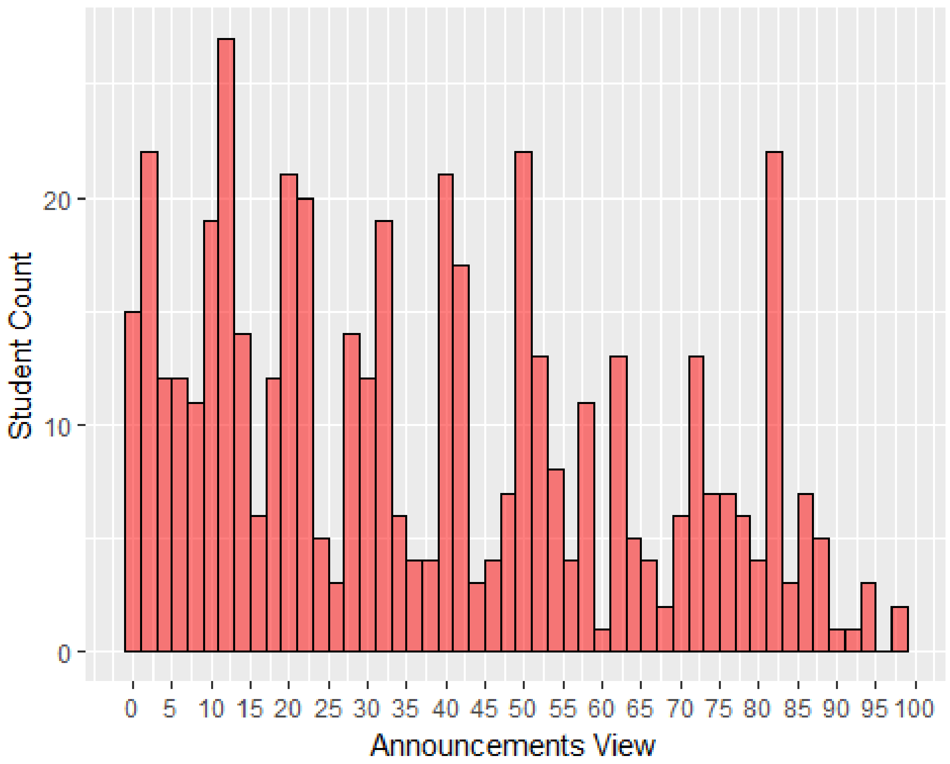
Visited Resources

```
ggplot(edu, aes(x = VisITedResources)) + geom_histogram(bins=50, color
= "orange",fill="orange",alpha=0.4) +
  scale_x_continuous(breaks = seq(0,100,5)) +
  labs(x = "Visited Resources", y = "Student Count")
```



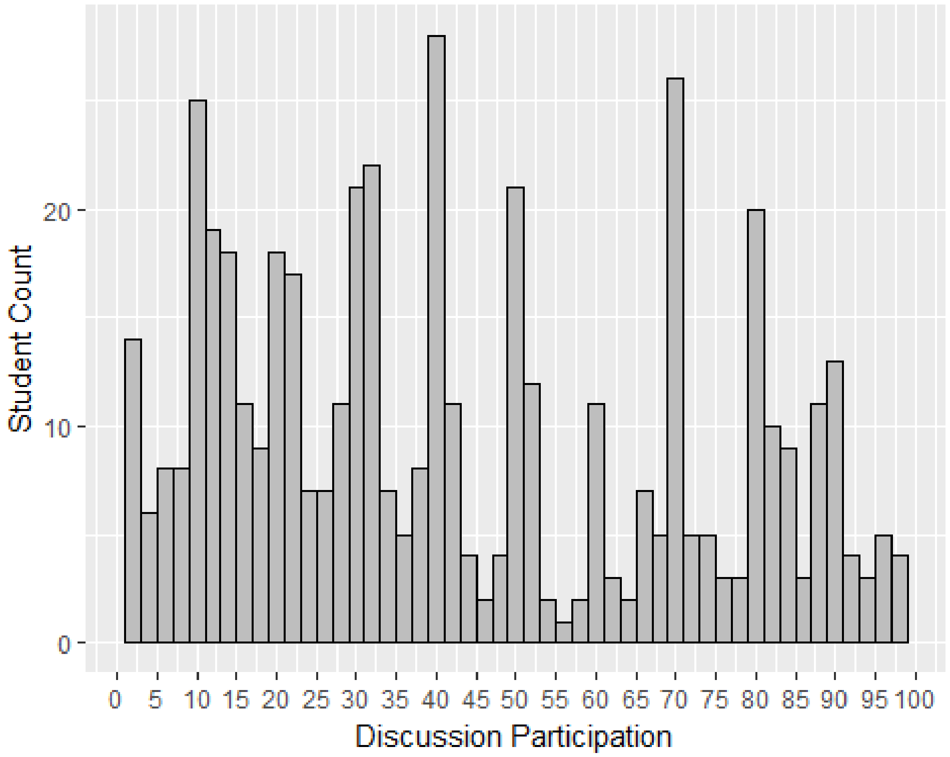
Announcements

```
ggplot(edu, aes(x = AnnouncementsView)) + geom_histogram(bins =
50,color = "black",fill="red",alpha=0.5) +
  scale_x_continuous(breaks = seq(0,100,5)) +
  labs(x = "Announcements View", y = "Student Count")
```



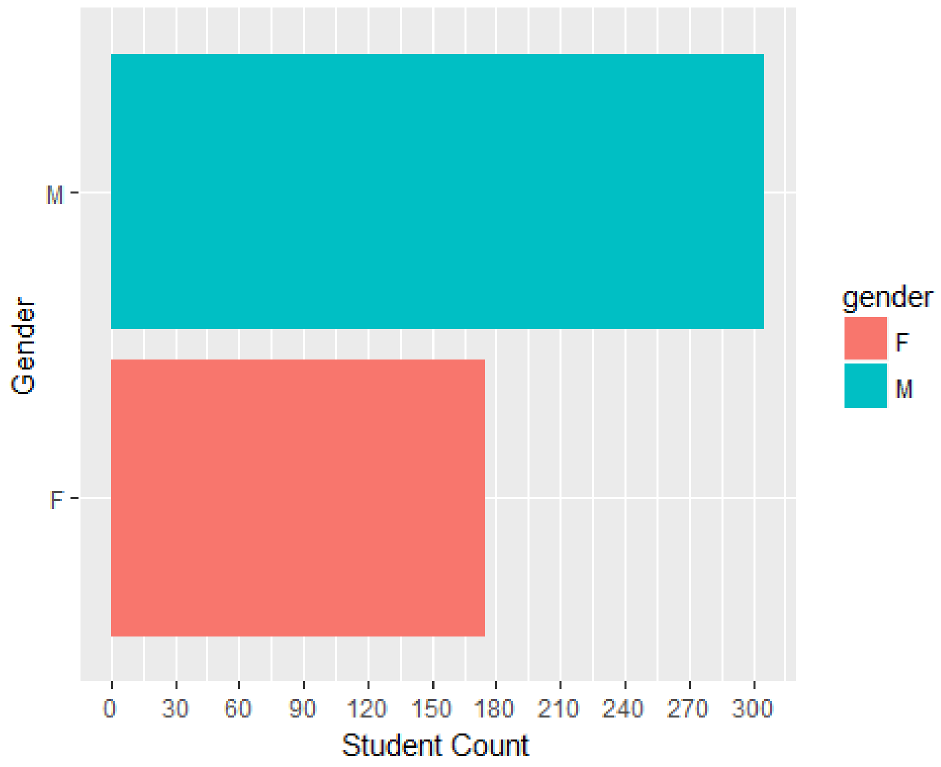
Discussion

```
ggplot(edu, aes(x = Discussion)) + geom_histogram(bins=50,color =
"black",fill="grey") +
  scale_x_continuous(breaks = seq(0,100,5)) +
  labs(x = "Discussion Participation", y = "Student Count")
```

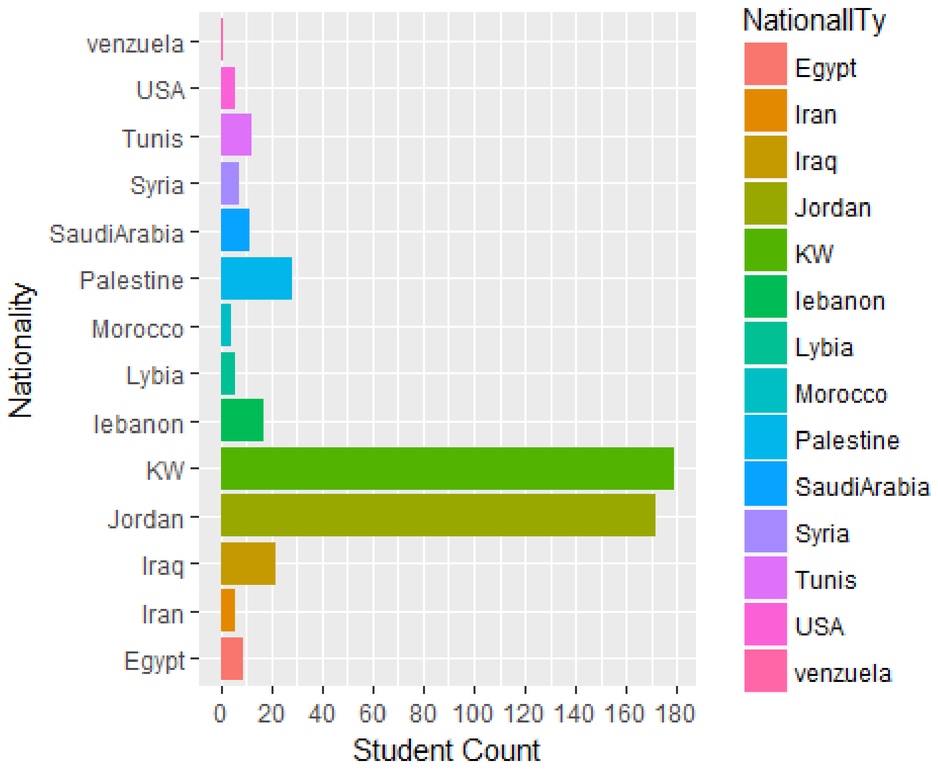


Barplots

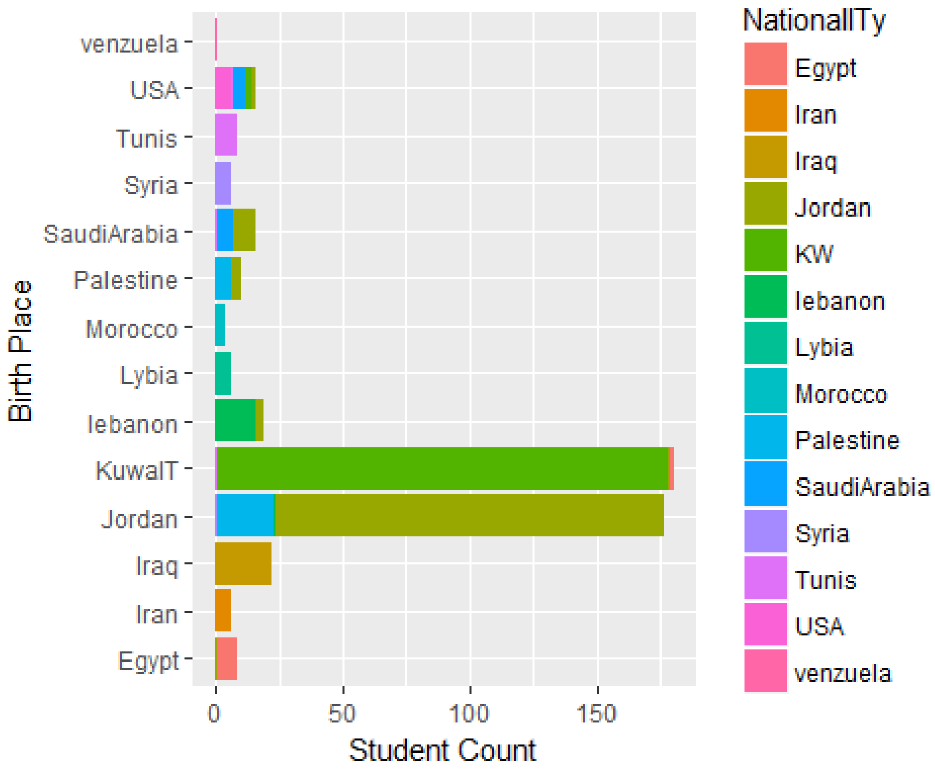
```
ggplot(edu, aes(x = gender)) + geom_bar(aes(fill=gender)) +
  labs(x = "Gender", y = "Student Count") +
  scale_y_continuous(breaks = seq(0,300,30)) + coord_flip()
```



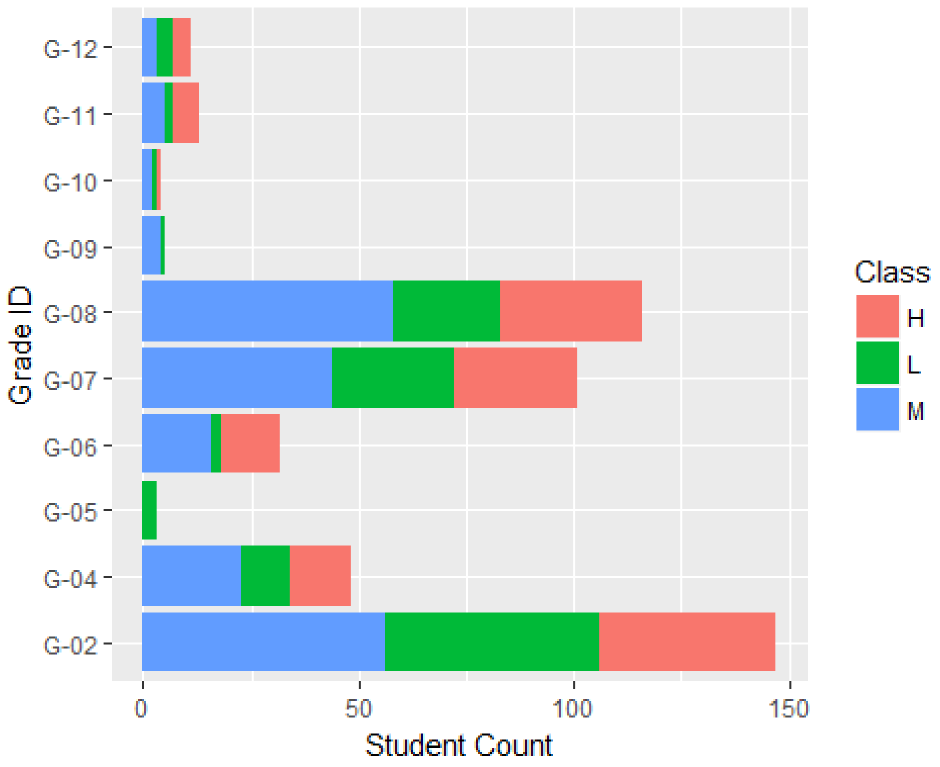
```
ggplot(edu, aes(x = NationalITy)) + geom_bar(aes(fill=NationalITy)) +
  labs(x = "Nationality", y = "Student Count") +
  scale_y_continuous(breaks = seq(0,200,20)) + coord_flip()
```



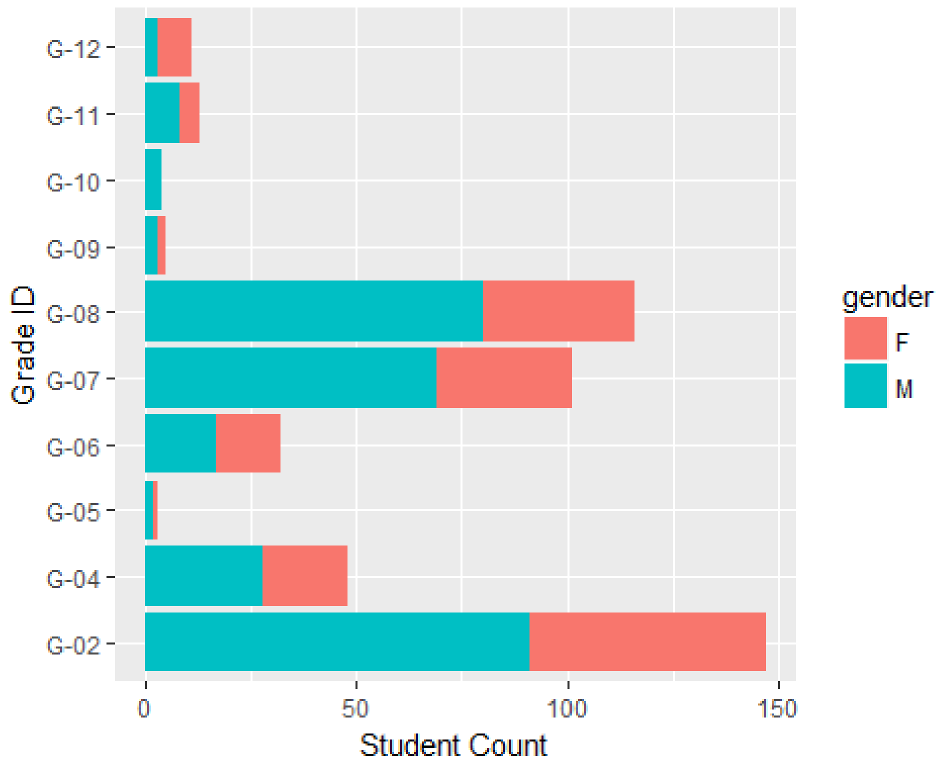
```r
ggplot(edu, aes(x = PlaceofBirth)) + geom_bar(aes(fill = NationalITy))
+
  labs(x = "Birth Place", y = "Student Count") + coord_flip() # usa is
a mix of nationalities
```



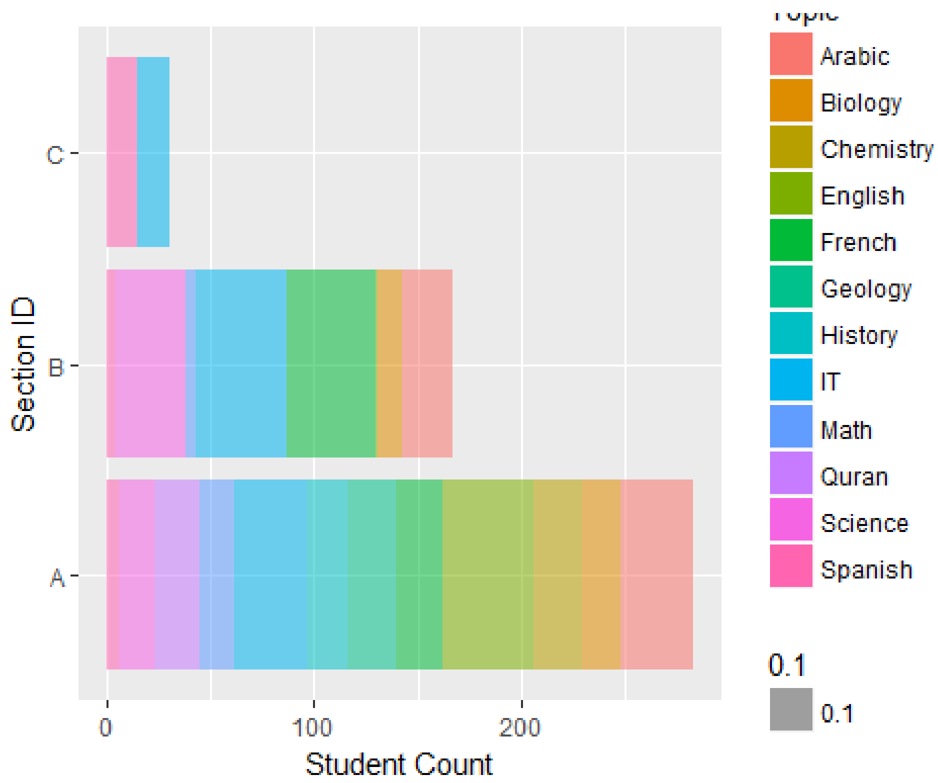
```
ggplot(edu, aes(x = GradeID, fill = Class)) + geom_bar() +
  labs(x = "Grade ID", y = "Student Count") + coord_flip() # g-06 has
students with only low grades
```



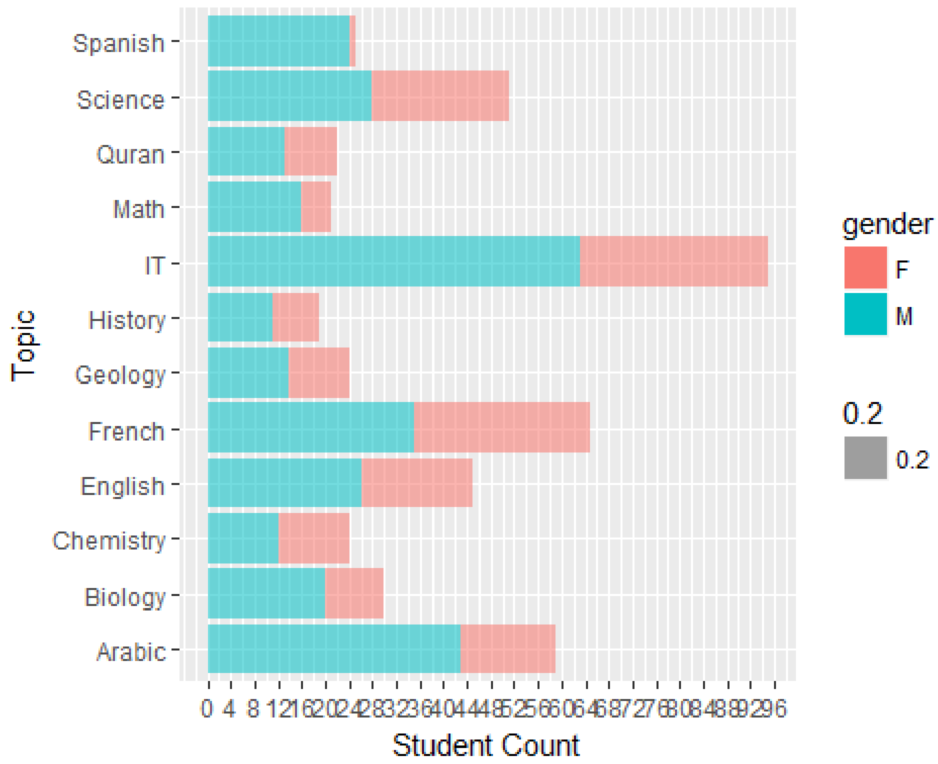
```r
ggplot(edu, aes(x = GradeID, fill = gender)) + geom_bar() +
  labs(x = "Grade ID", y = "Student Count") + coord_flip() # g-10 has
no females
```



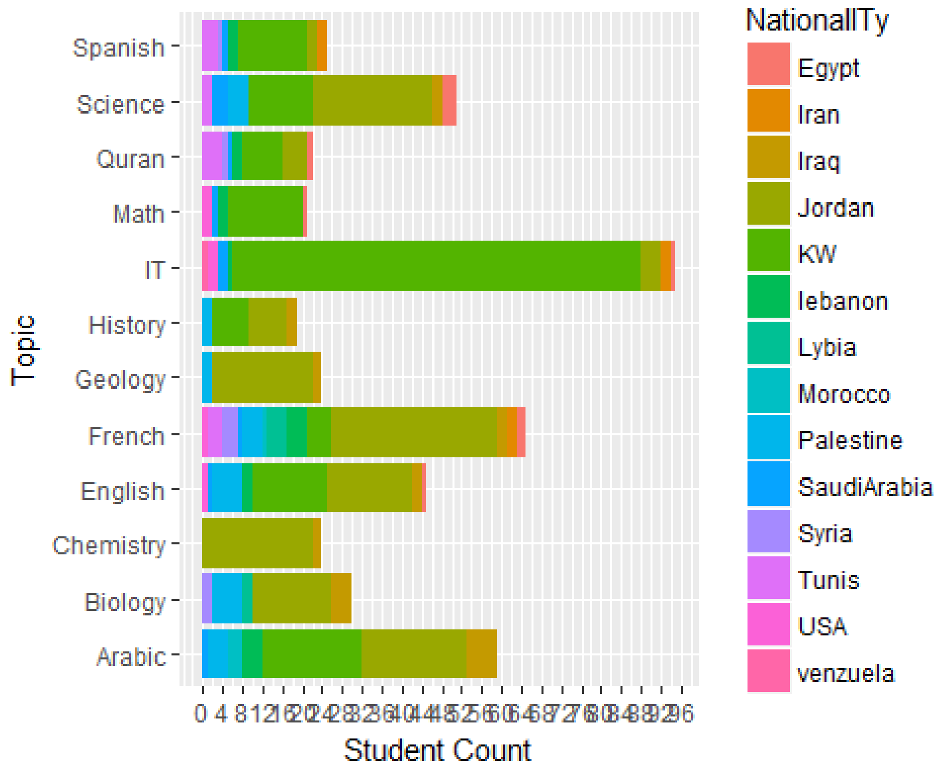
```
ggplot(edu, aes(x = SectionID, fill = Topic,alpha=0.1)) + geom_bar() +
  labs(x = "Section ID", y = "Student Count") +
  coord_flip()
```



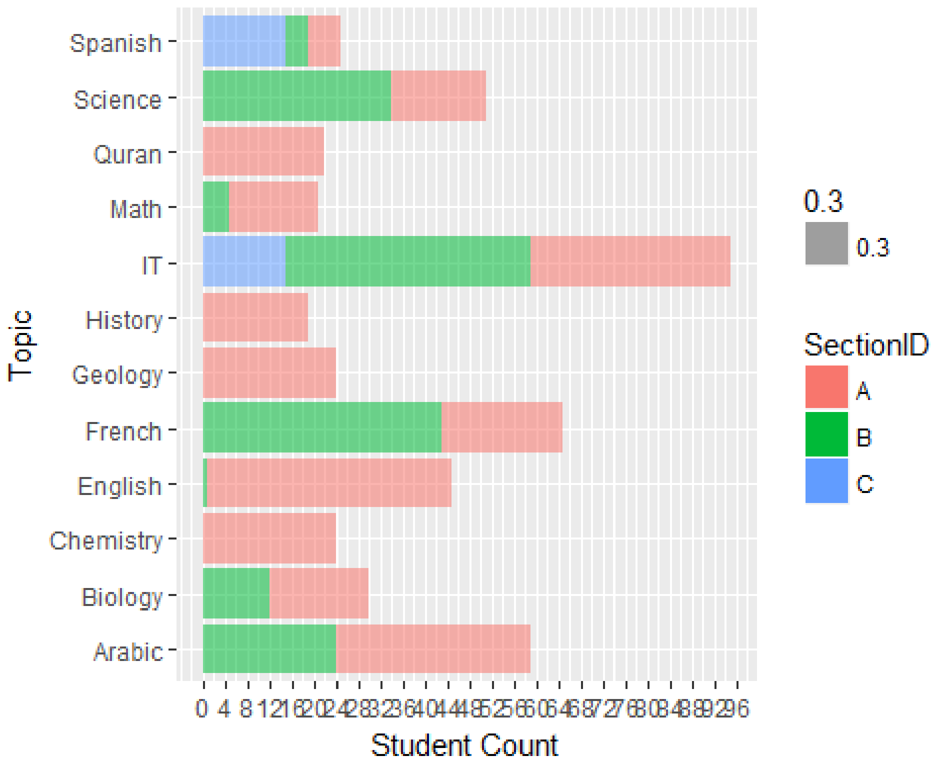
```
ggplot(edu, aes(x = Topic, fill = gender,alpha=0.2)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") +
  scale_y_continuous(breaks = seq(0,100,4)) + coord_flip()
```



```
ggplot(edu, aes(x = Topic, fill = NationalITy)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```

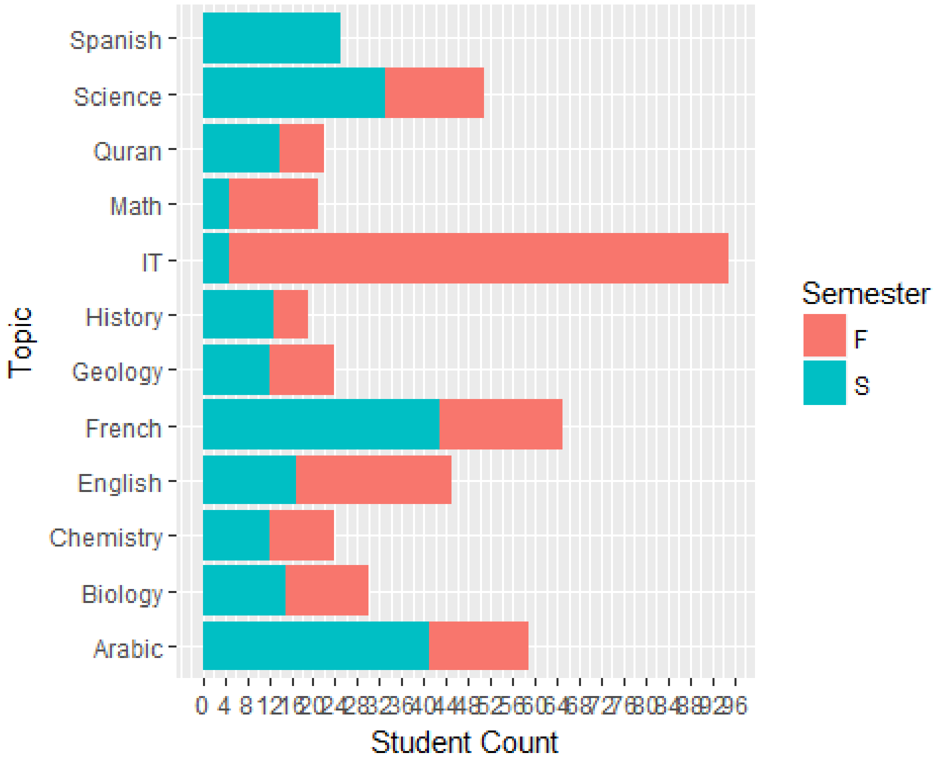


```
ggplot(edu, aes(x = Topic, fill = SectionID,alpha=0.3)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```



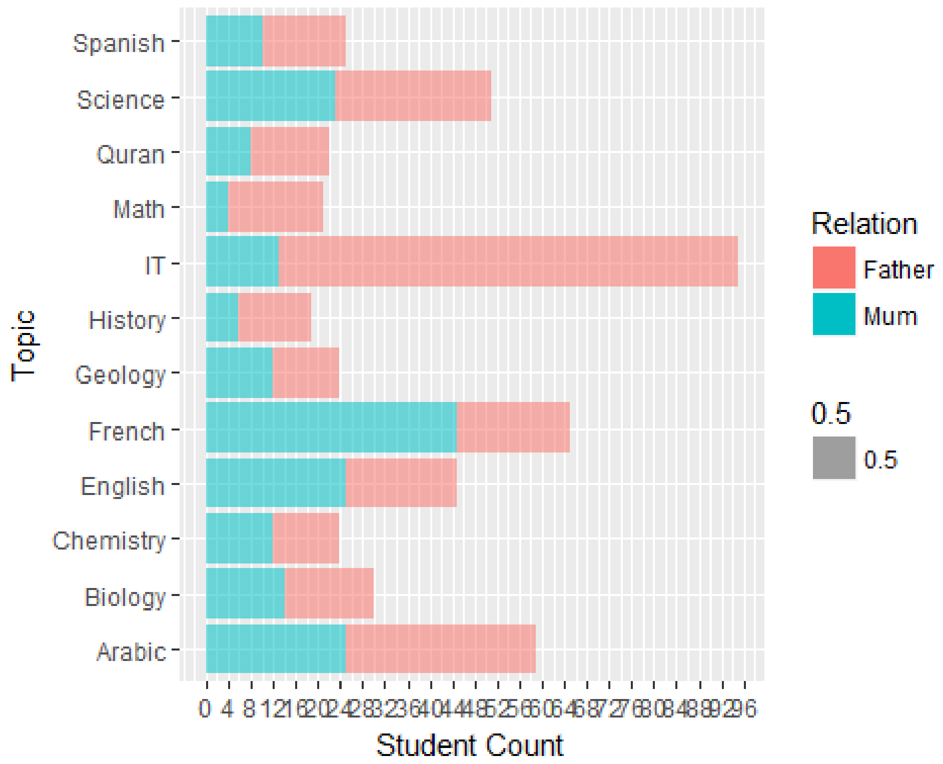
Section C for Mostly Spanish Students

```
ggplot(edu, aes(x = Topic, fill = Semester)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```



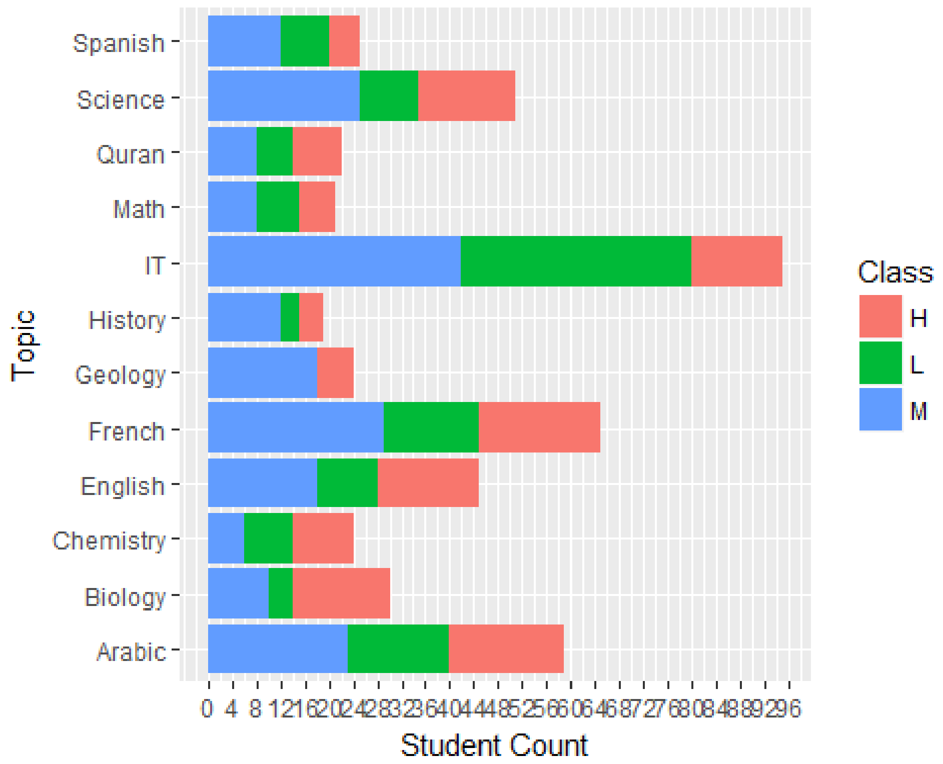
IT Students Are Mostly in 1st Semester

```
ggplot(edu, aes(x = Topic, fill = Relation,alpha=0.5)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```

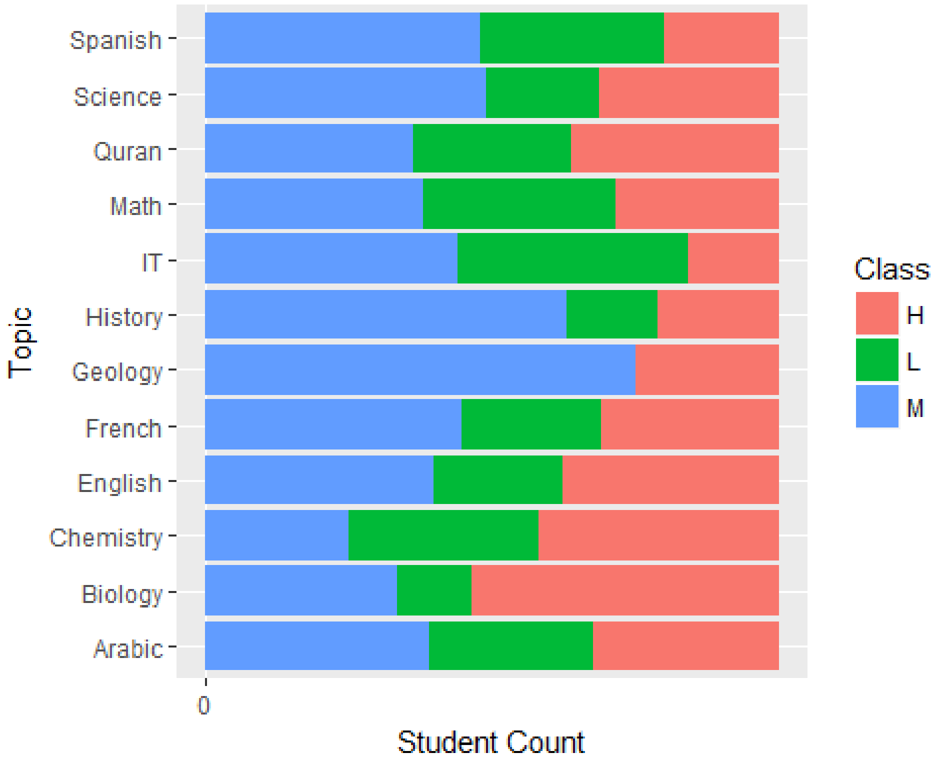


Most French Students have Mom as Guardian in Comparison to Father

```
ggplot(edu, aes(x = Topic, fill = Class)) + geom_bar() +
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```

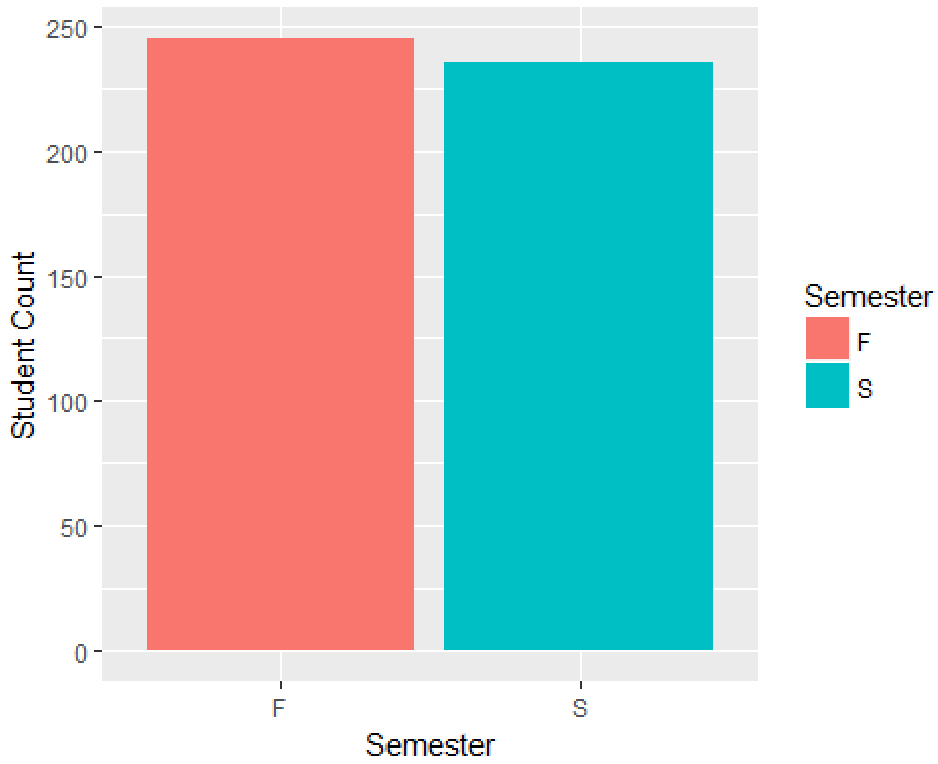


```
ggplot(edu, aes(x = Topic, fill = Class)) + geom_bar(position = "fill")
+
  labs(x = "Topic", y = "Student Count") + coord_flip() +
  scale_y_continuous(breaks = seq(0,100,4))
```

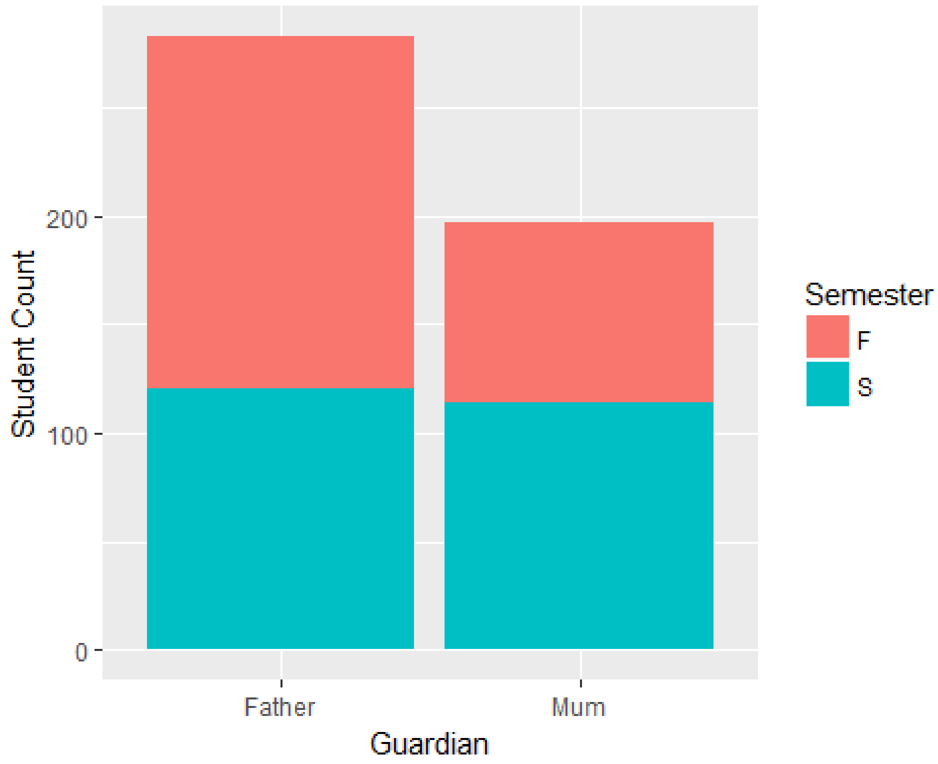


Geology has no low class students

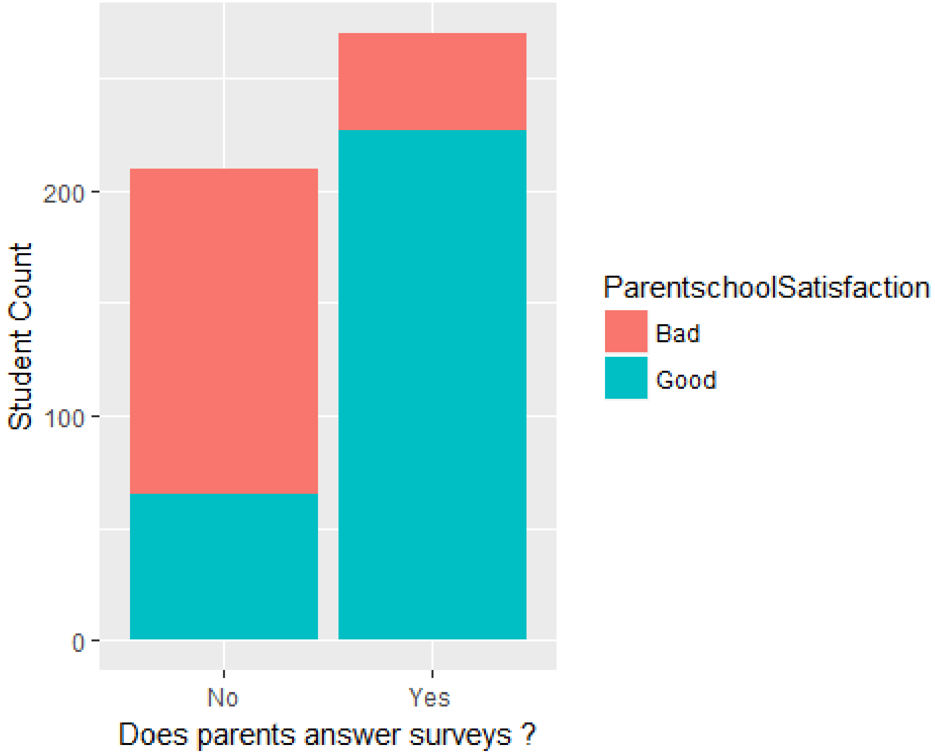
```
ggplot(edu, aes(x = Semester)) + geom_bar(aes(fill=Semester)) +
  labs(x = "Semester", y = "Student Count")
```



```r
ggplot(edu, aes(x = Relation, fill = Semester)) + geom_bar() +
  labs(x = "Guardian", y = "Student Count")
```

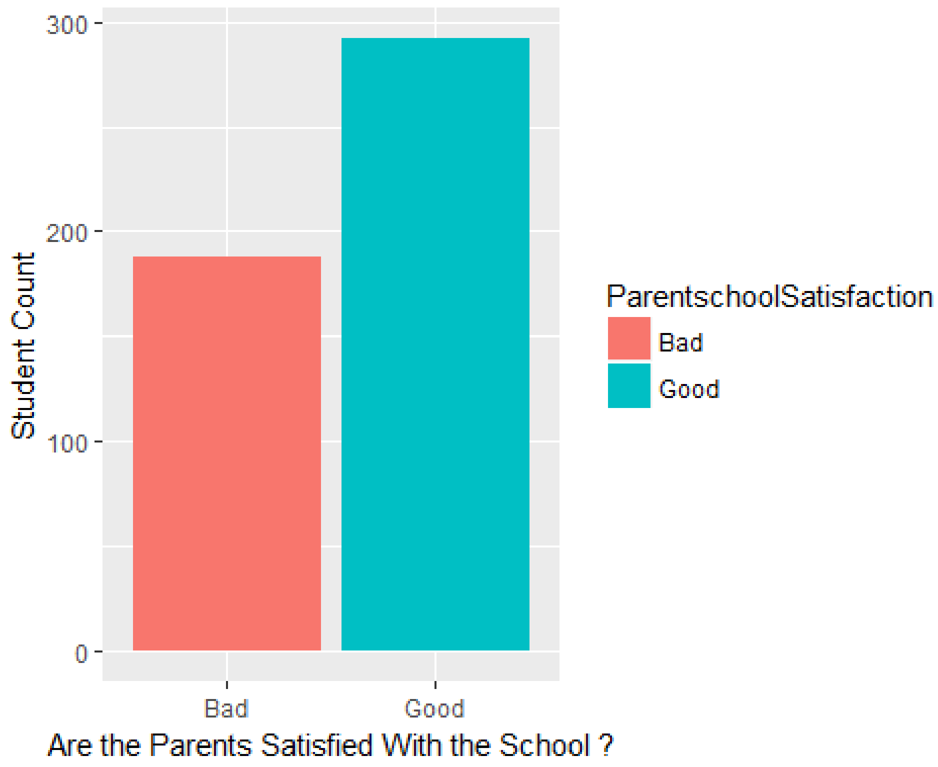


```
ggplot(edu, aes(x = ParentAnsweringSurvey, fill =
ParentschoolSatisfaction)) +
  geom_bar() +
  labs(x = "Does parents answer surveys ?", y = "Student Count")
```

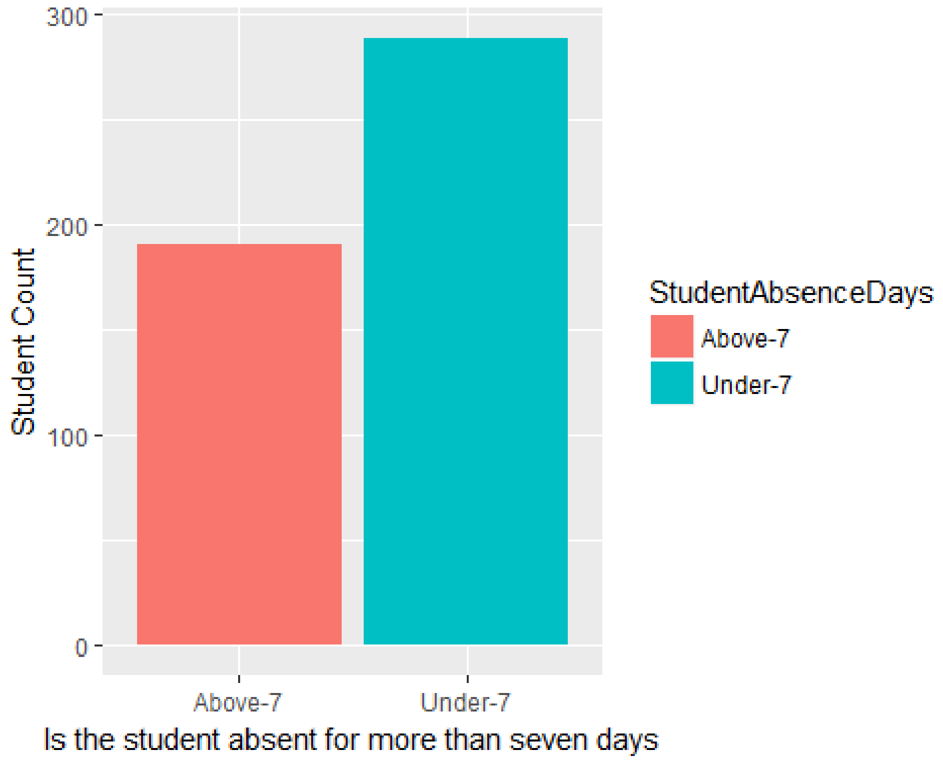


Parent Satisfaction

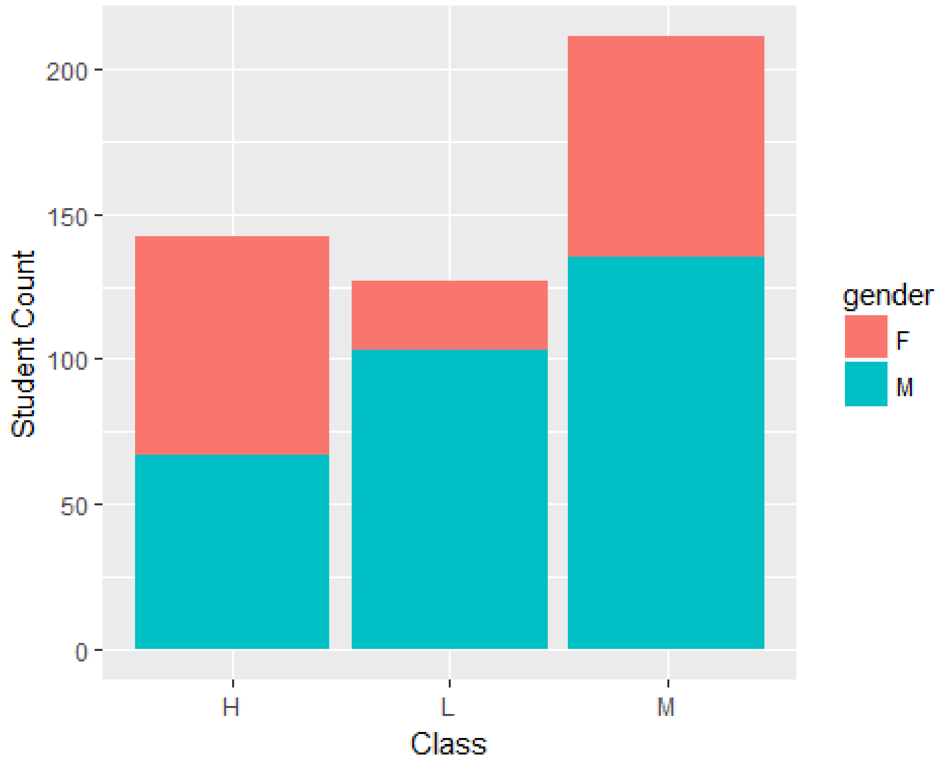
```
ggplot(edu, aes(x = ParentschoolSatisfaction)) +
  geom_bar(aes(fill=ParentschoolSatisfaction)) +
  labs(x = "Are the Parents Satisfied With the School ?", y = "Student
Count")
```



```
ggplot(edu, aes(x = StudentAbsenceDays)) +
geom_bar(aes(fill=StudentAbsenceDays)) +
  labs(x = "Is the student absent for more than seven days", y =
"Student Count")
```

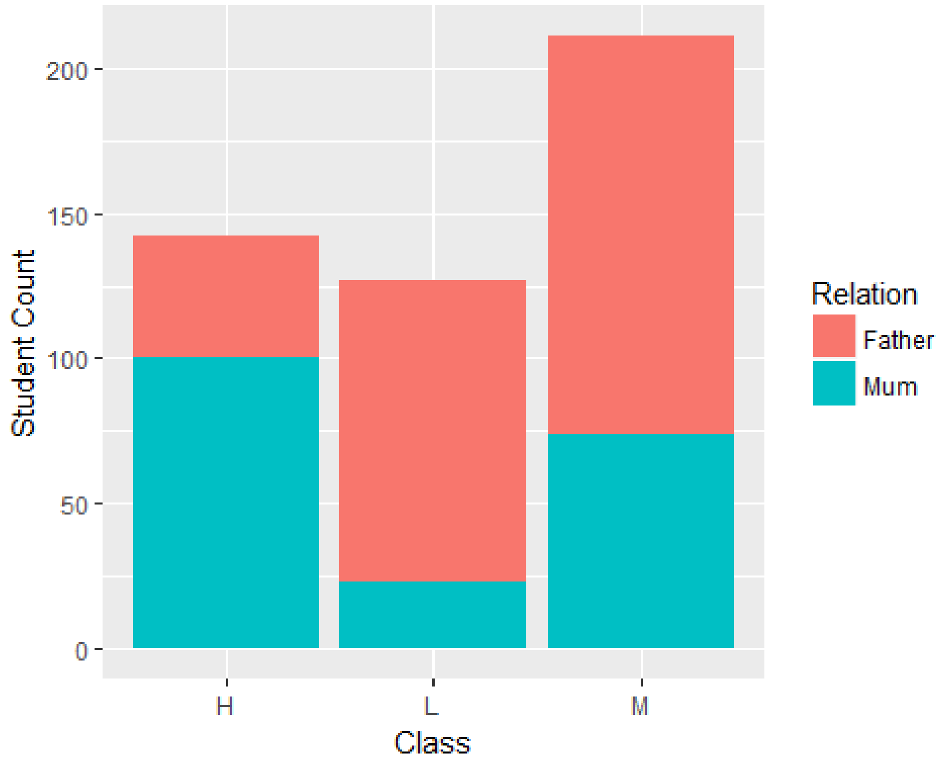


```r
ggplot(edu, aes(x = Class, fill = gender)) + geom_bar() +
  labs(x = "Class", y = "Student Count")
```

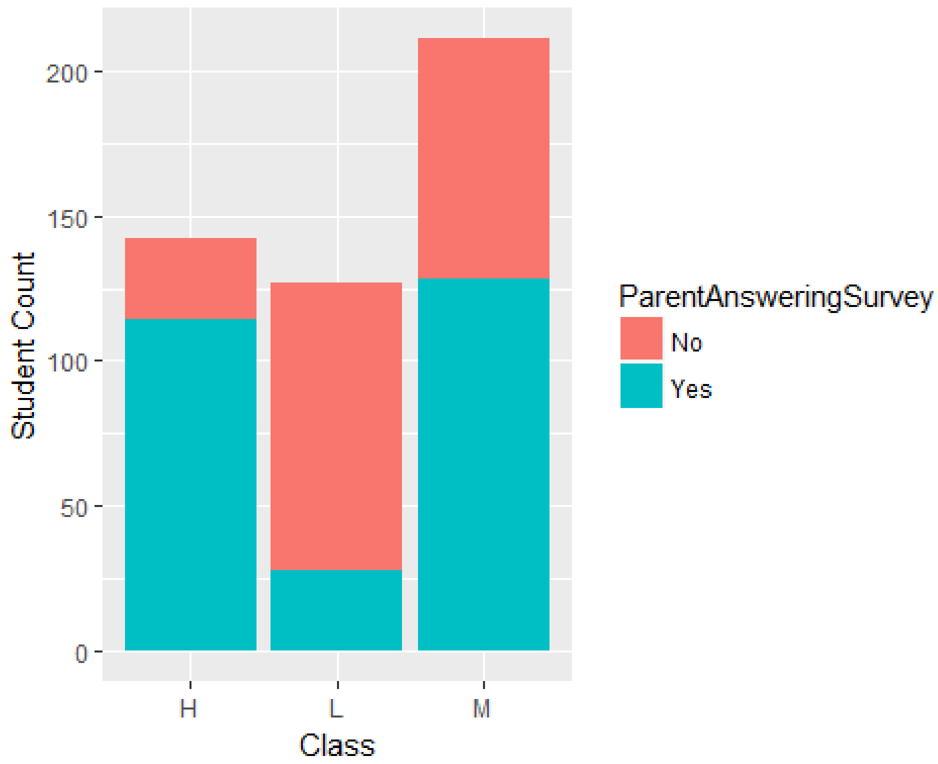


Few Girls in the Low Class

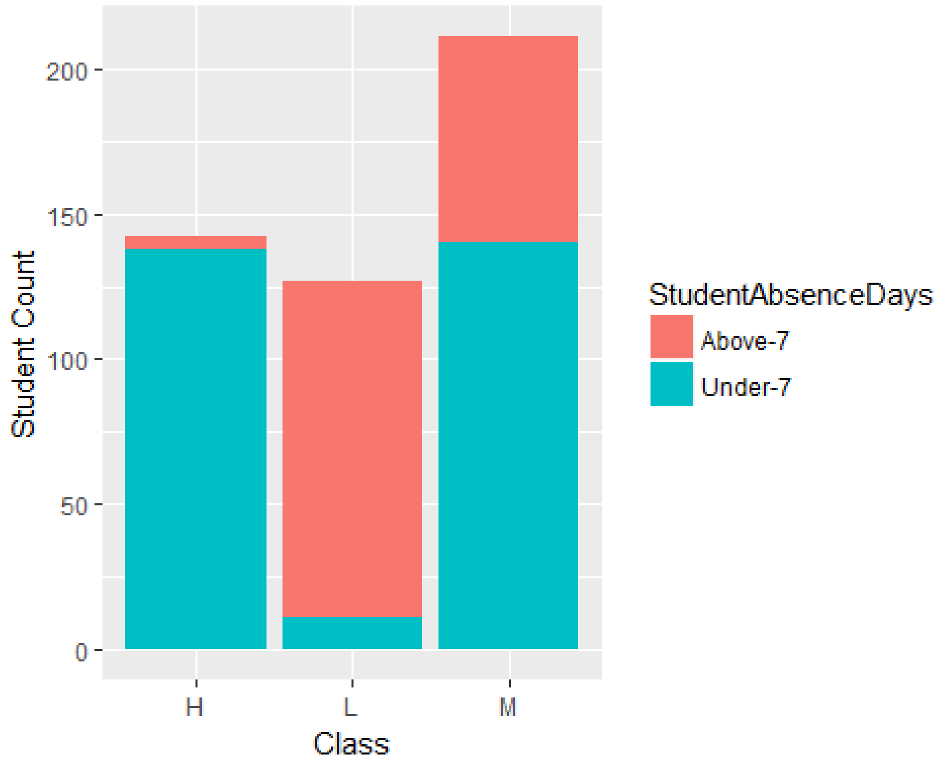
```
ggplot(edu, aes(x = Class, fill = Relation)) + geom_bar() +
  labs(x = "Class", y = "Student Count")
```



```
ggplot(edu, aes(x = Class, fill = ParentAnsweringSurvey)) + geom_bar()
+
  labs(x = "Class", y = "Student Count")
```

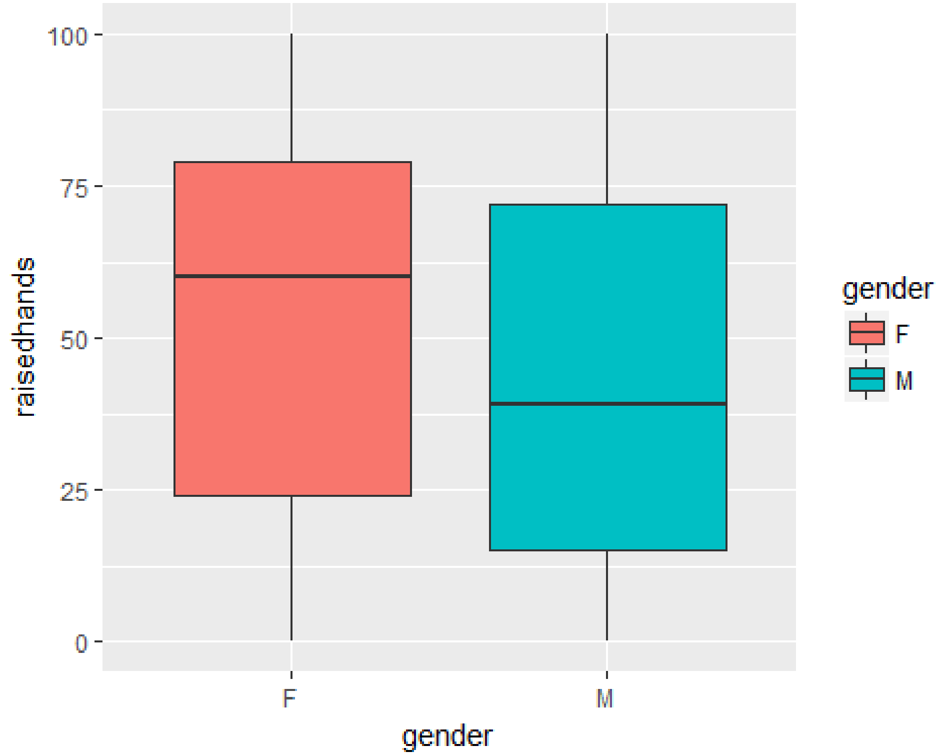


```
ggplot(edu, aes(x = Class, fill = StudentAbsenceDays)) + geom_bar() +
  labs(x = "Class", y = "Student Count")
```

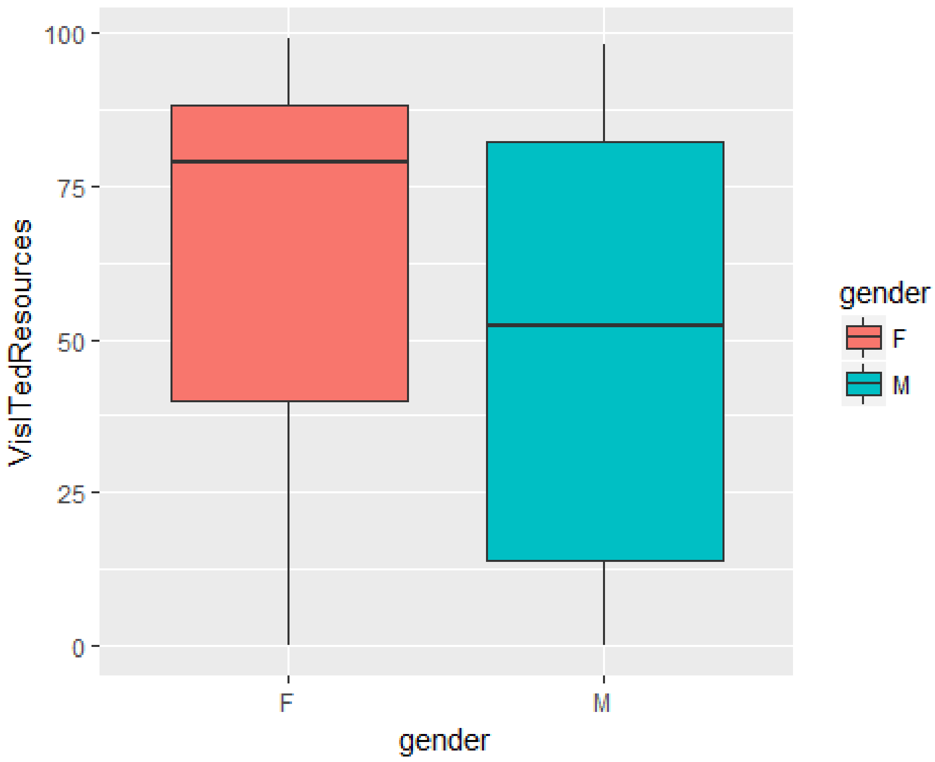


Boxplots

```
ggplot(edu, aes(x = gender, y = raisedhands)) +
geom_boxplot(aes(fill=gender))
```

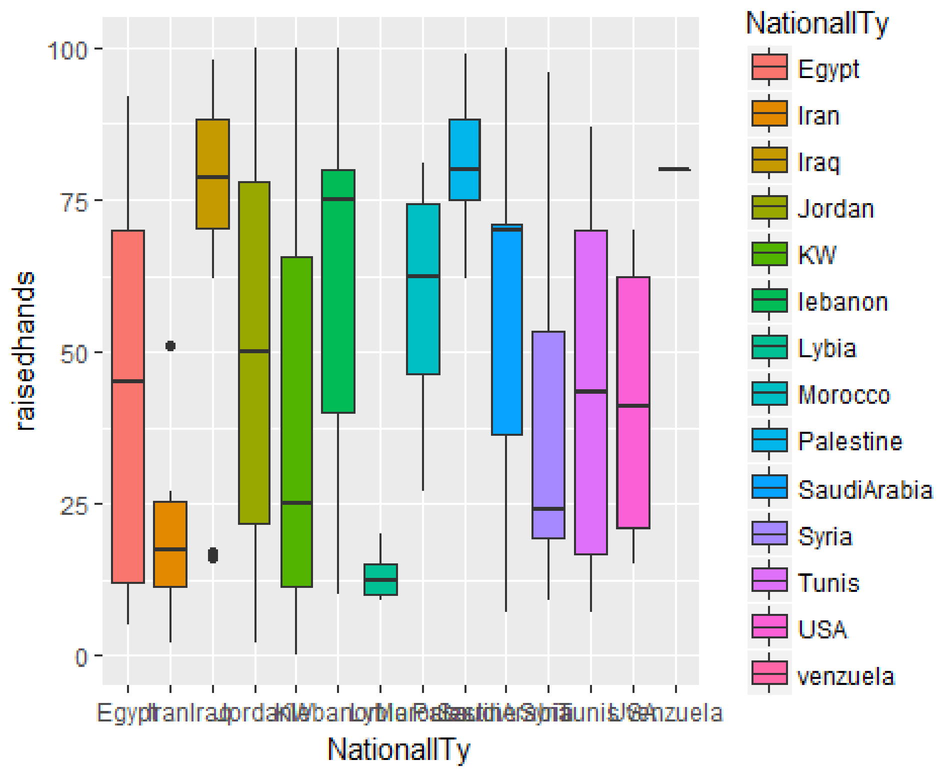


```
ggplot(edu, aes(x = gender, y = VisITedResources)) +
geom_boxplot(aes(fill=gender))
```

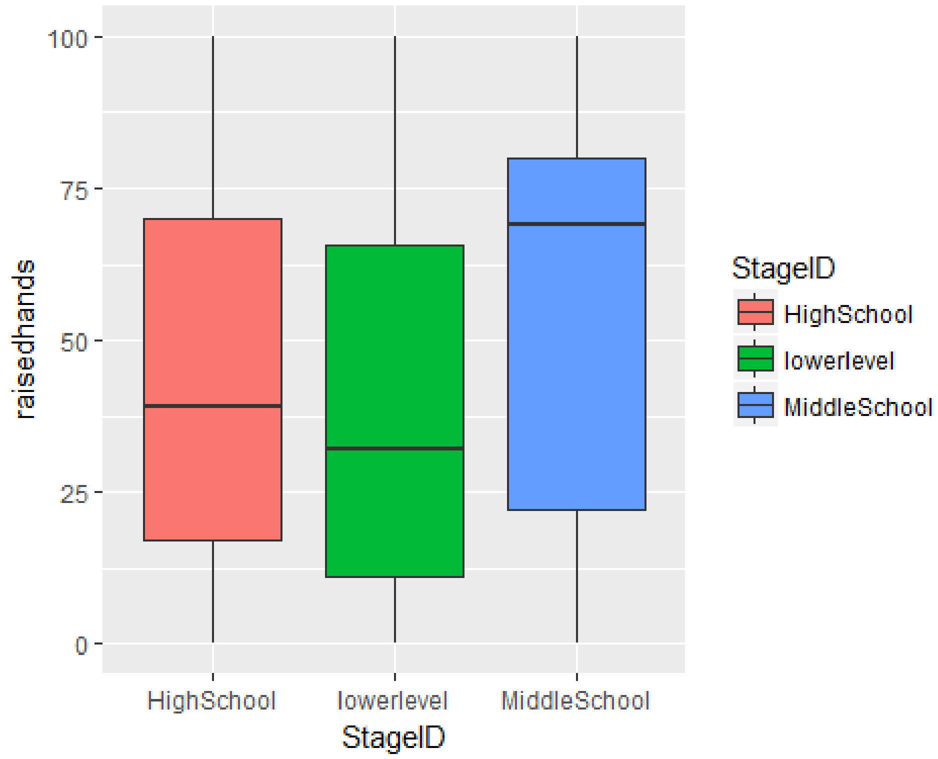


Girls Use More Resources

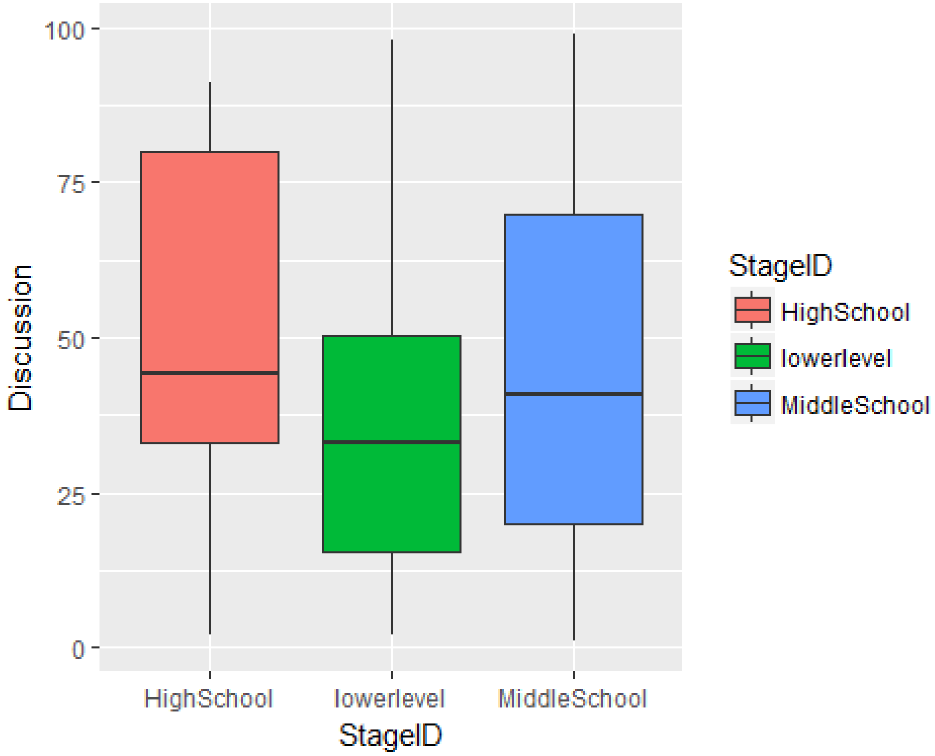
```
ggplot(edu, aes(x = NationalITy, y = raisedhands)) +
geom_boxplot(aes(fill=NationalITy))
```



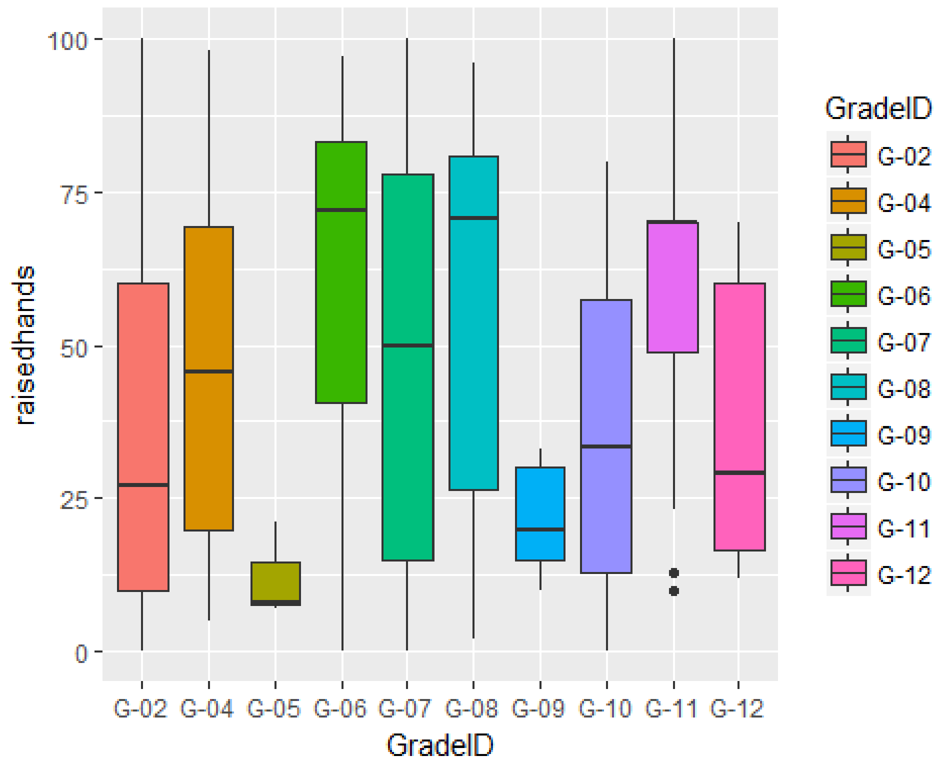
```
ggplot(edu, aes(x = StageID, y = raisedhands)) +
geom_boxplot(aes(fill=StageID))
```



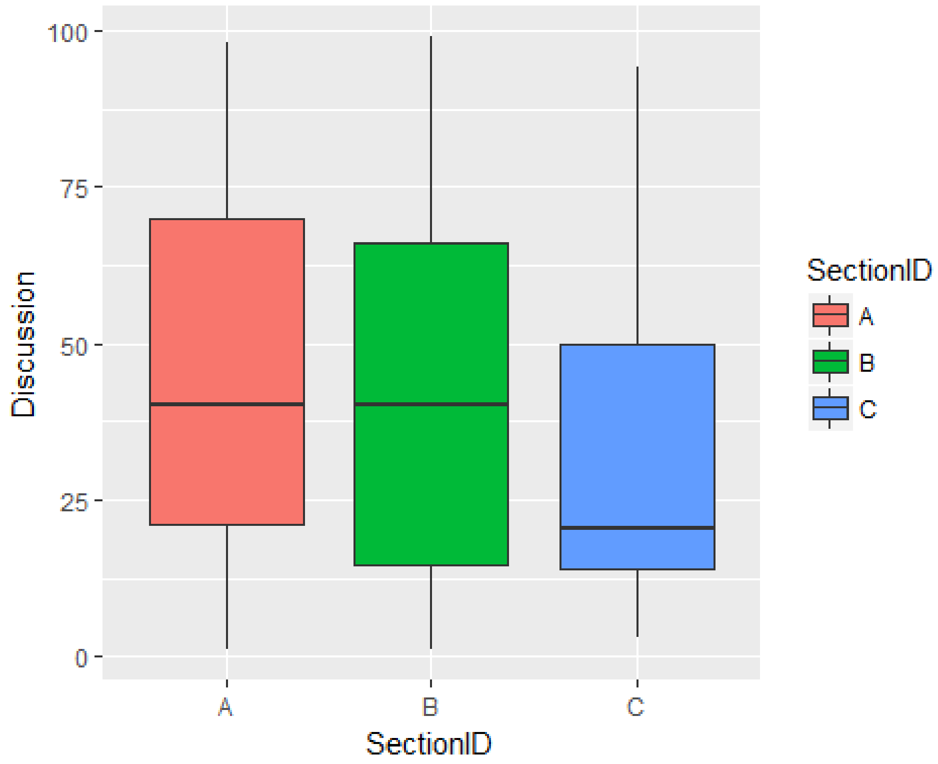
```
ggplot(edu, aes(x = StageID, y = Discussion)) +
geom_boxplot(aes(fill=StageID))
```



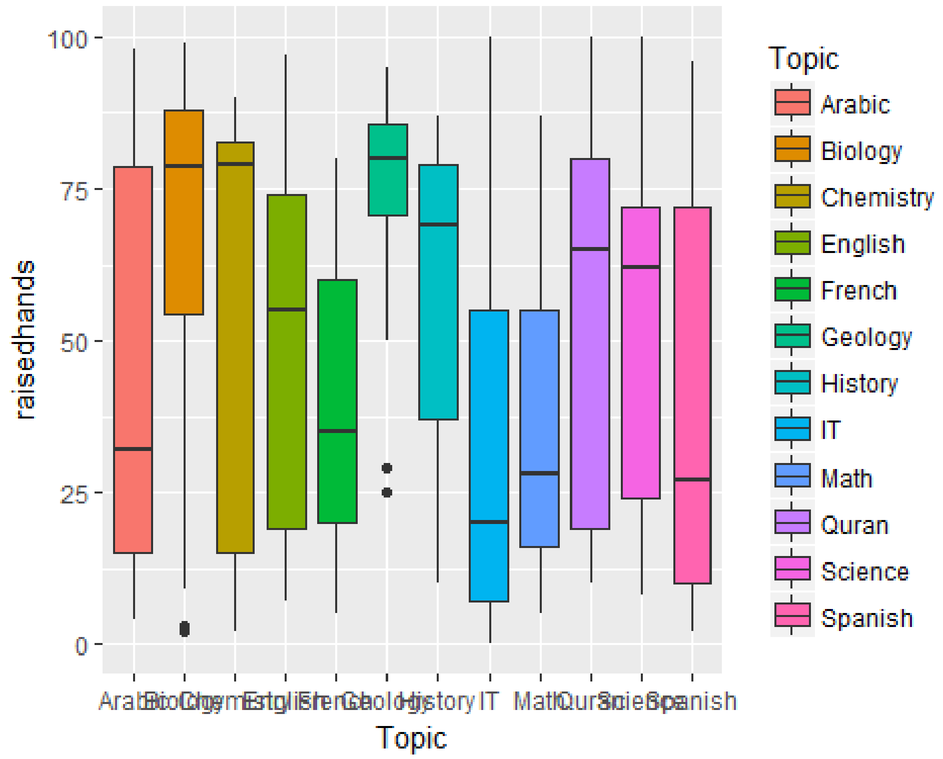
```
ggplot(edu, aes(x = GradeID, y = raisedhands)) +
geom_boxplot(aes(fill=GradeID))
```



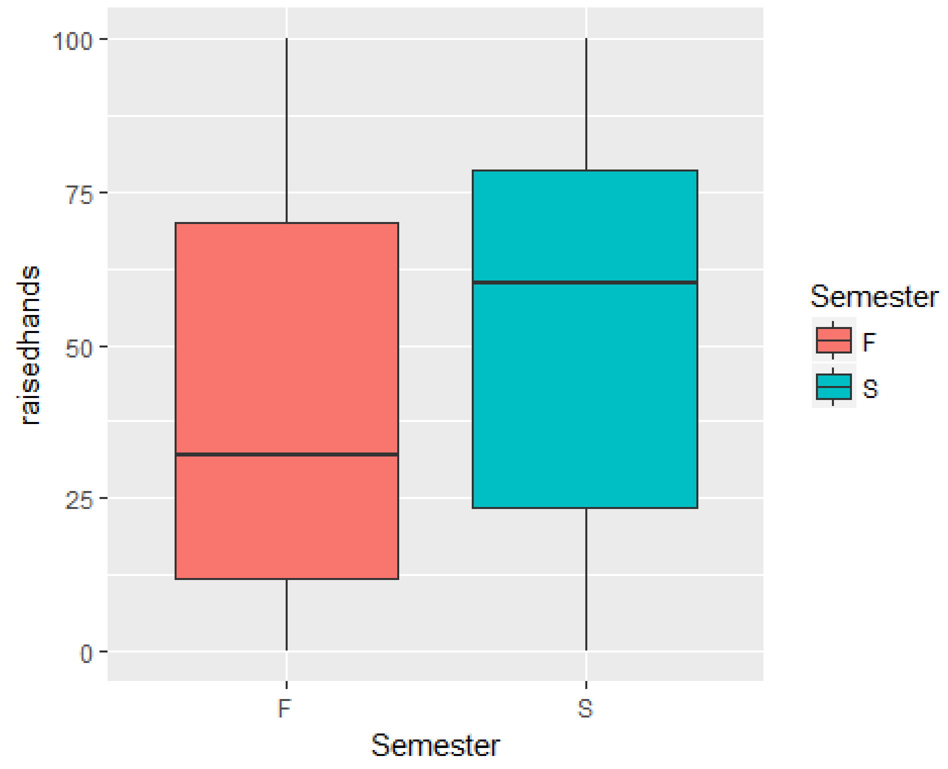
```
ggplot(edu, aes(x = SectionID, y = Discussion)) +
geom_boxplot(aes(fill=SectionID))
```



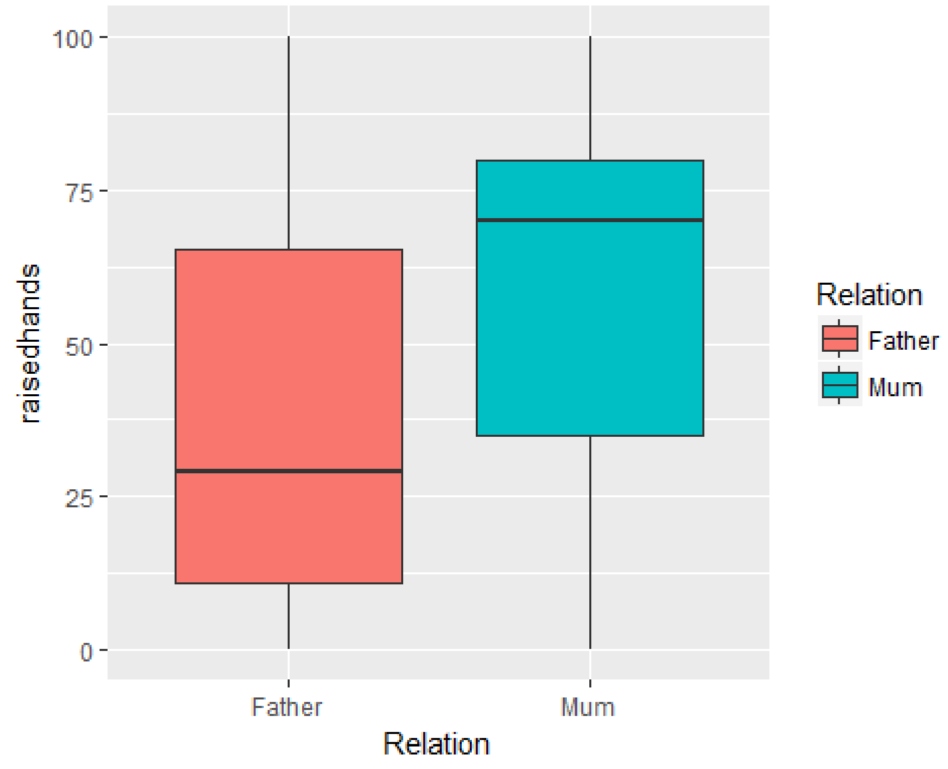
```
ggplot(edu, aes(x = Topic, y = raisedhands)) +
geom_boxplot(aes(fill=Topic))
```



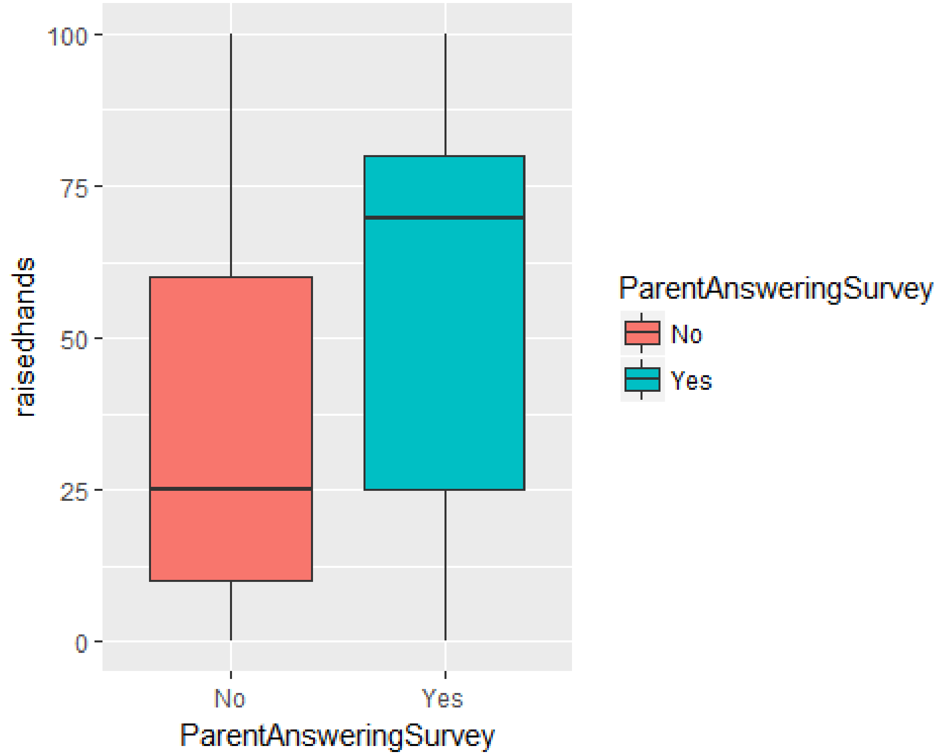
```
ggplot(edu, aes(x = Semester, y = raisedhands)) +
geom_boxplot(aes(fill=Semester))
```



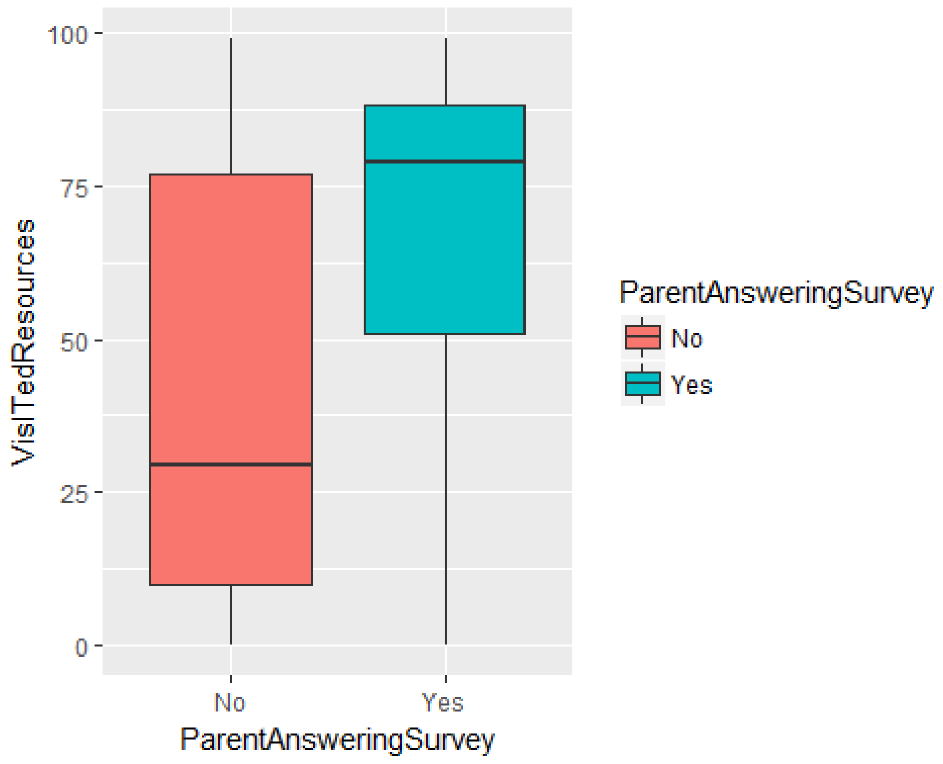
```
ggplot(edu, aes(x = Relation, y = raisedhands)) +
geom_boxplot(aes(fill=Relation))
```



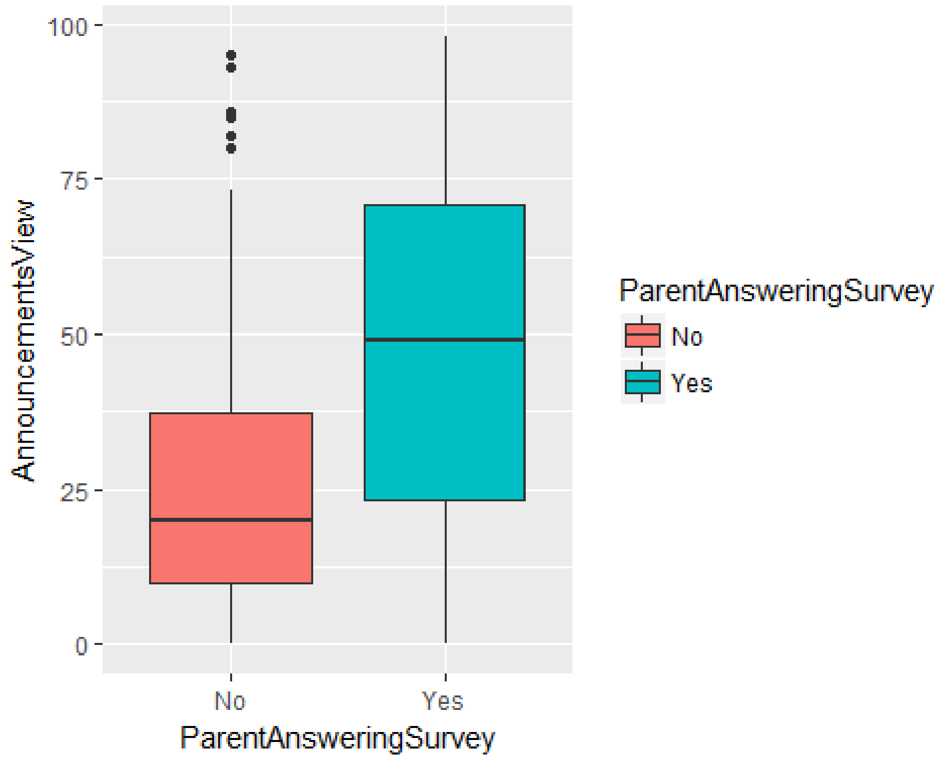
```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = raisedhands)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



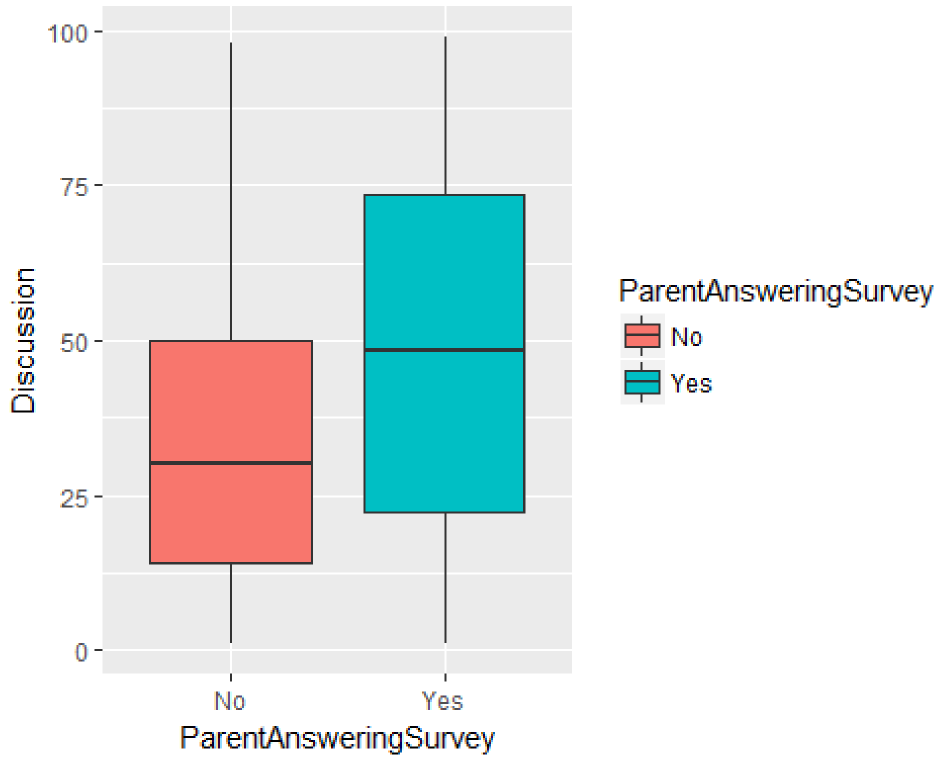
```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = VisITedResources)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



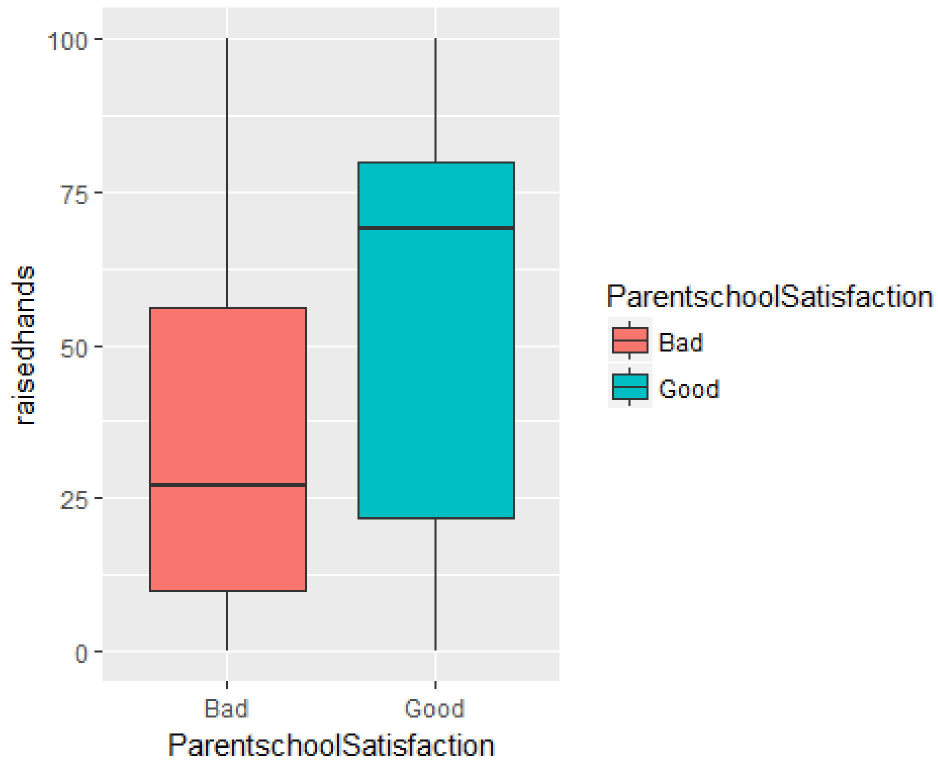
```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = AnnouncementsView)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



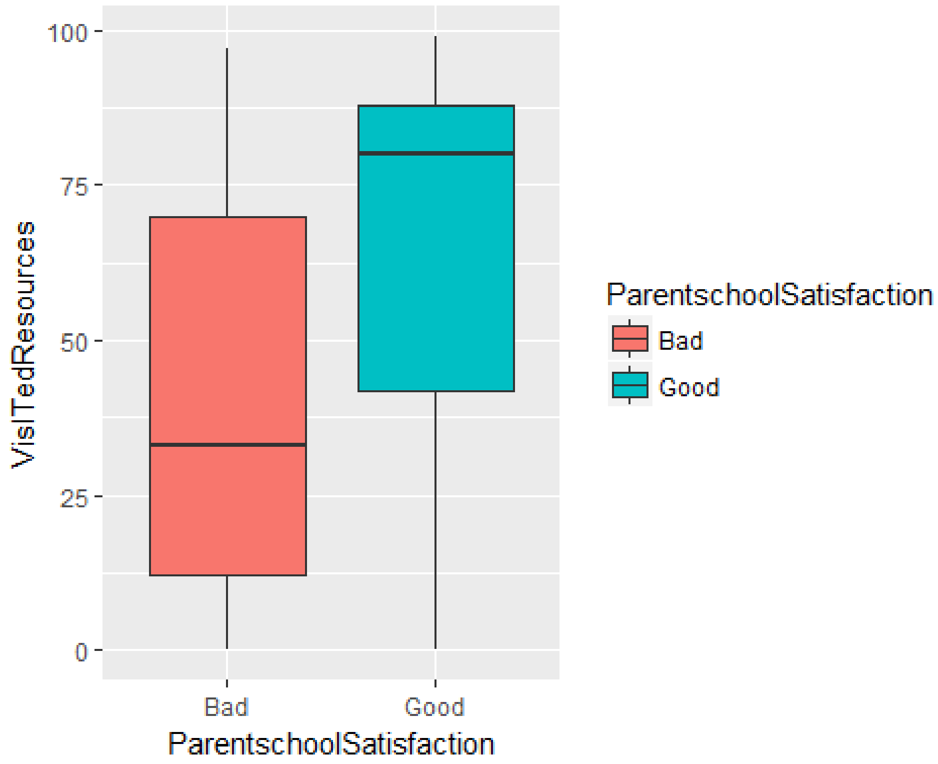
```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = Discussion)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



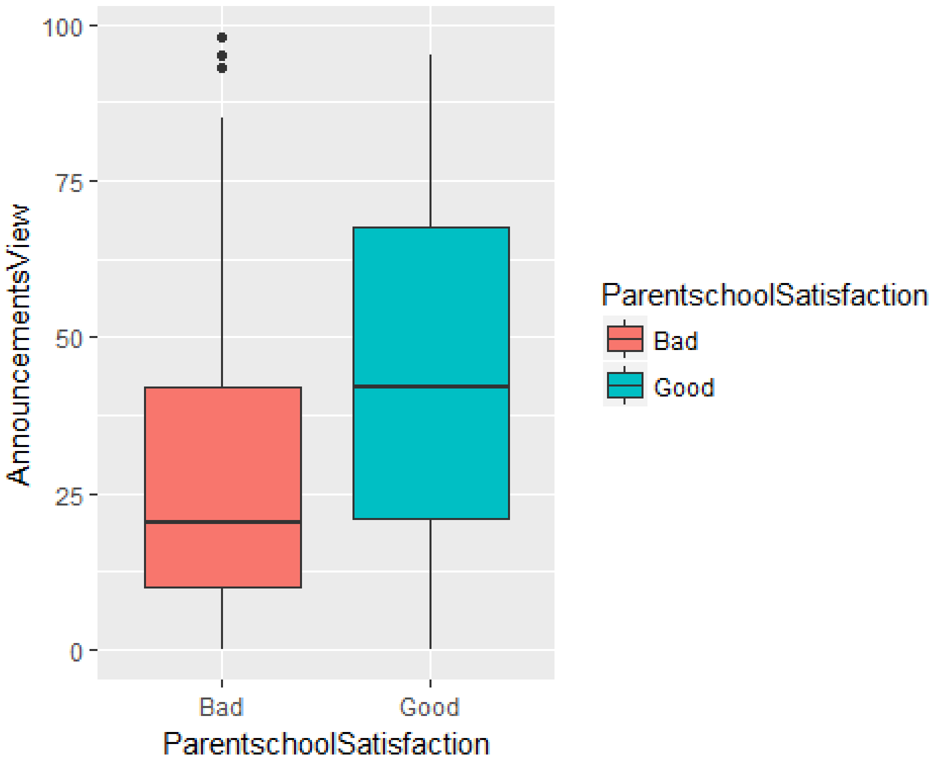
```
ggplot(edu, aes(x = ParentschoolSatisfaction, y = raisedhands)) +
geom_boxplot(aes(fill=ParentschoolSatisfaction))
```



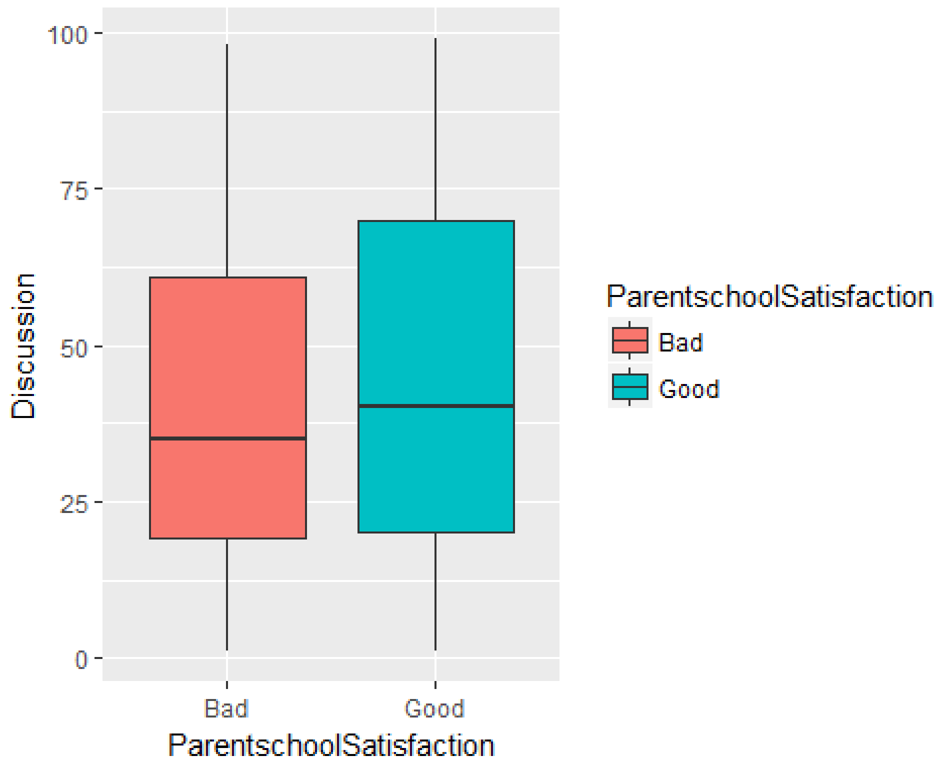
```
ggplot(edu, aes(x = ParentschoolSatisfaction, y = VisITedResources)) +
geom_boxplot(aes(fill=ParentschoolSatisfaction))
```



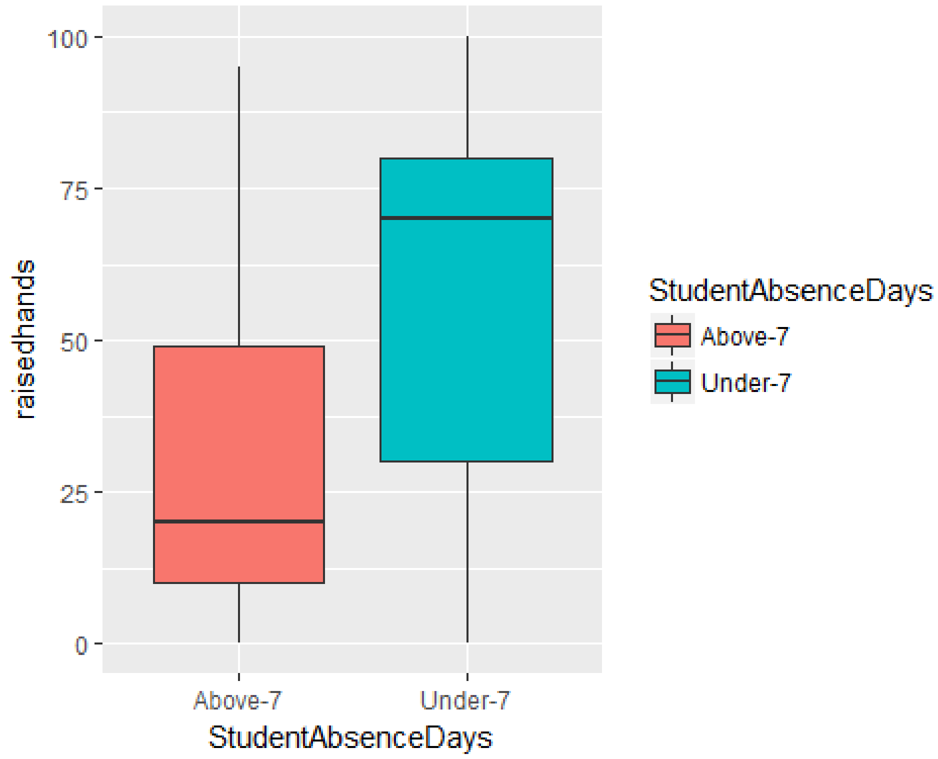
```
ggplot(edu, aes(x = ParentschoolSatisfaction, y = AnnouncementsView)) +
geom_boxplot(aes(fill=ParentschoolSatisfaction))
```



```
ggplot(edu, aes(x = ParentschoolSatisfaction, y = Discussion)) +
geom_boxplot(aes(fill=ParentschoolSatisfaction))
```

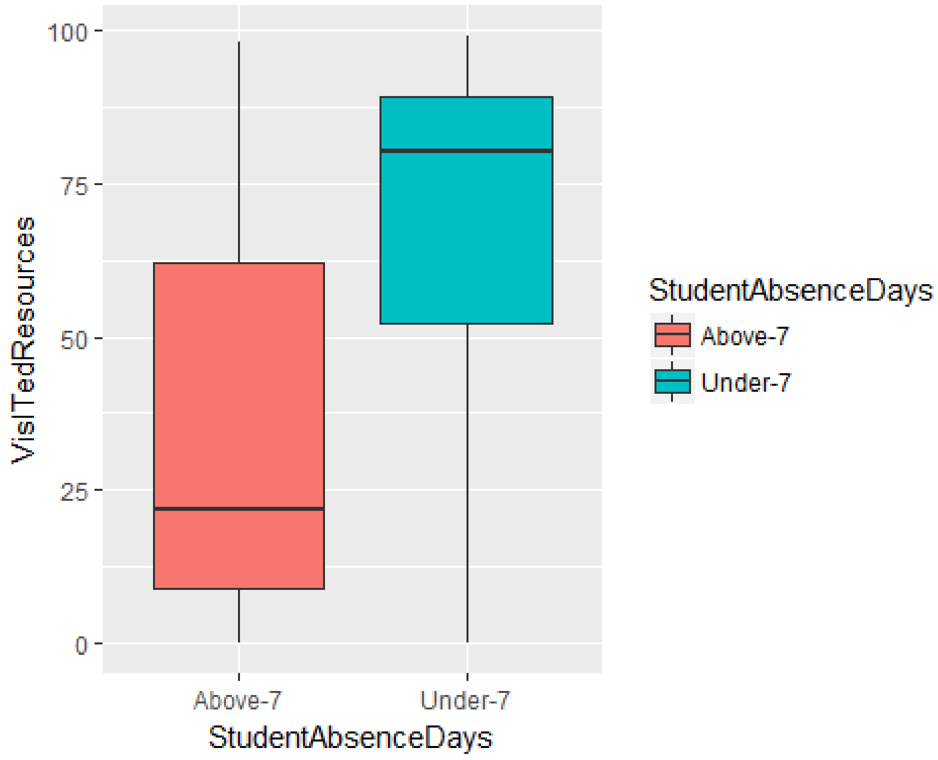


```
ggplot(edu, aes(x = StudentAbsenceDays, y = raisedhands)) +
geom_boxplot(aes(fill=StudentAbsenceDays))
```

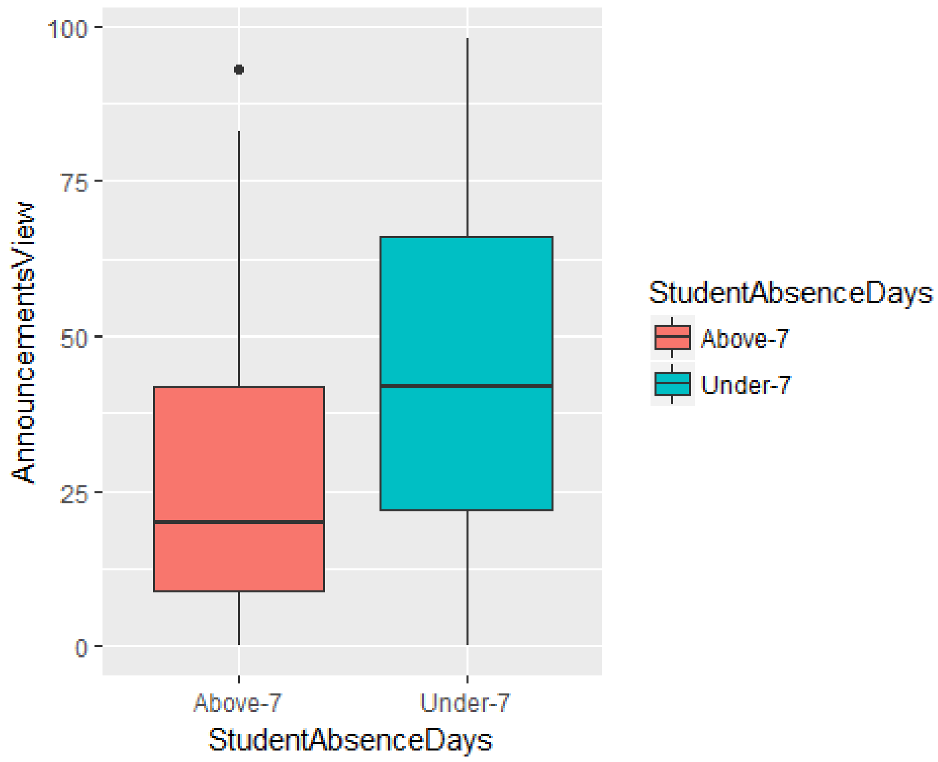


More Students Leave = Less Hand Raises

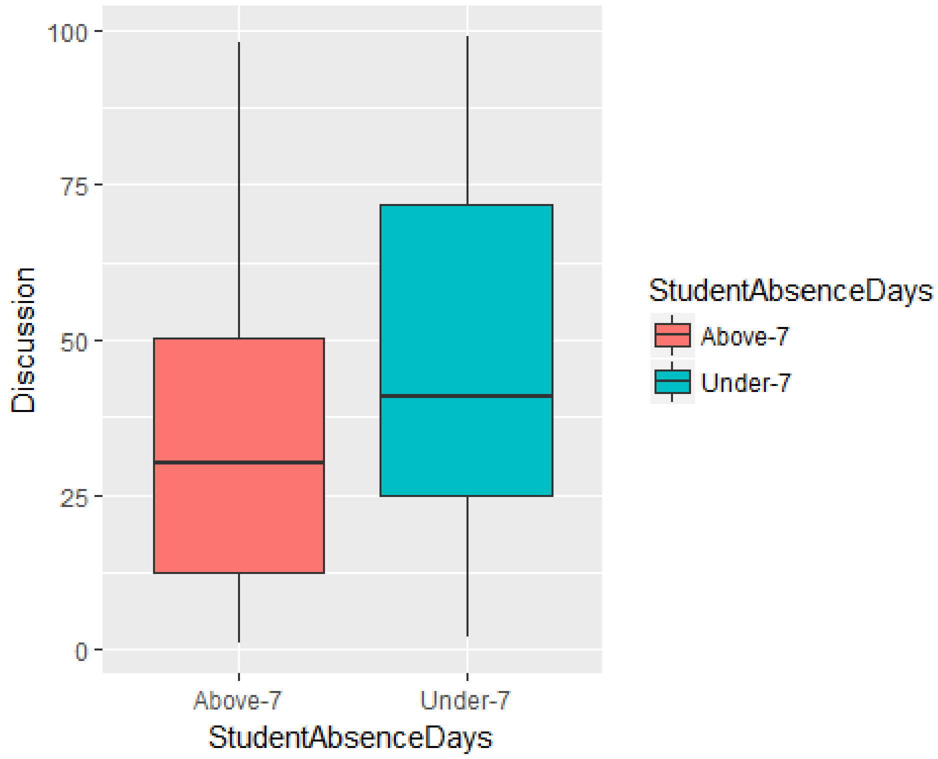
```
ggplot(edu, aes(x = StudentAbsenceDays, y = VisITedResources)) +
geom_boxplot(aes(fill=StudentAbsenceDays))
```



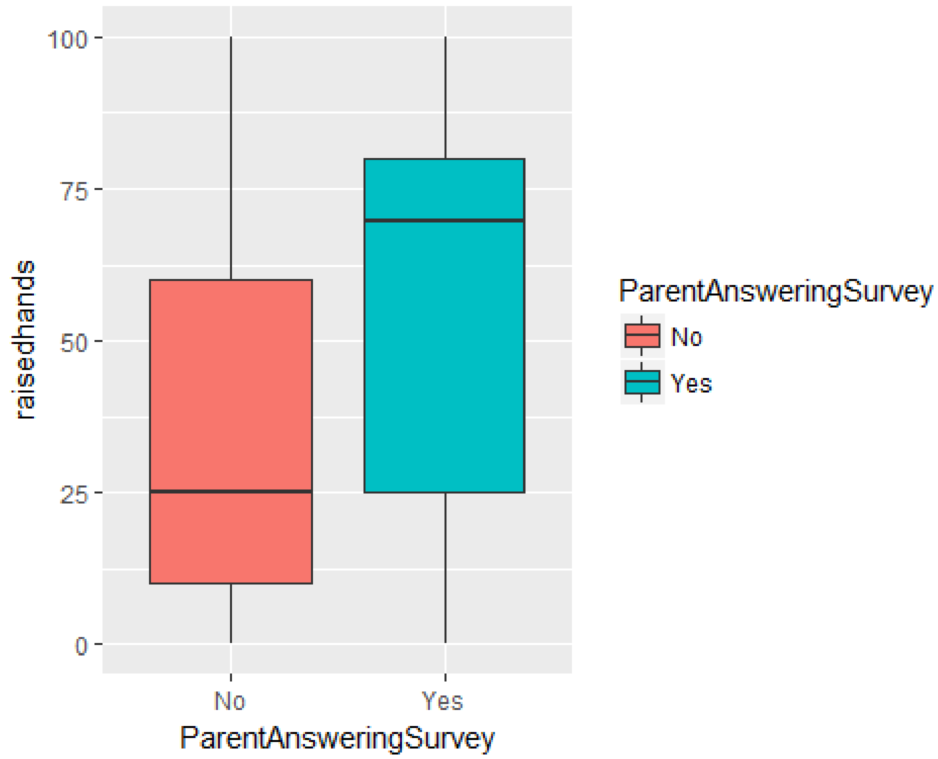
```
ggplot(edu, aes(x = StudentAbsenceDays, y = AnnouncementsView)) +
geom_boxplot(aes(fill=StudentAbsenceDays))
```



```
ggplot(edu, aes(x = StudentAbsenceDays, y = Discussion)) +
geom_boxplot(aes(fill=StudentAbsenceDays))
```

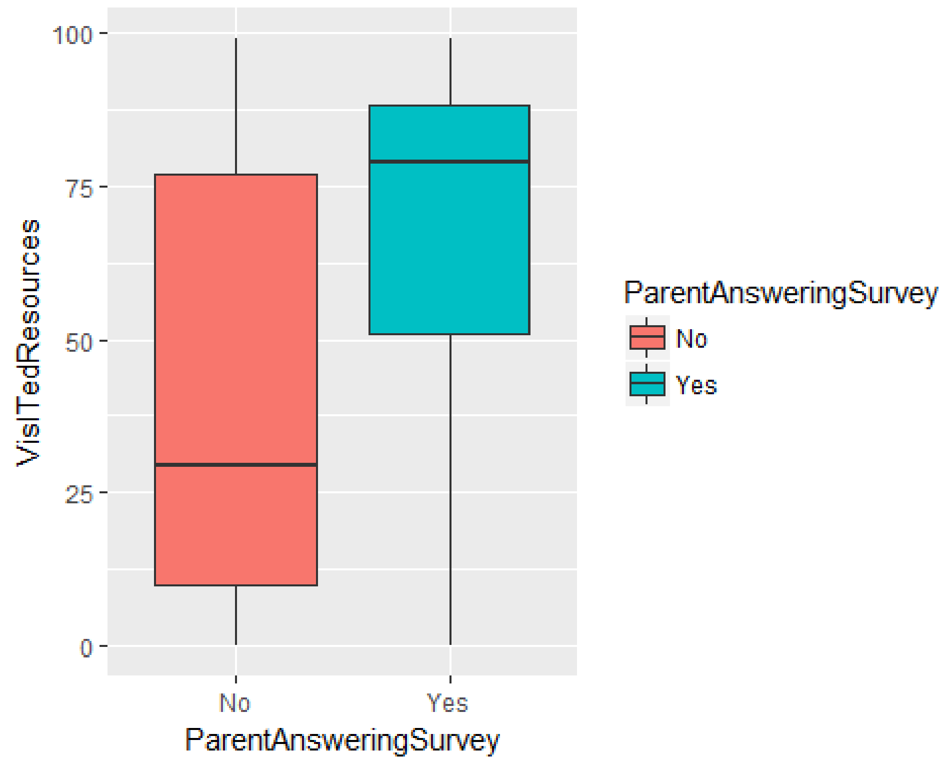


```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = raisedhands)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```

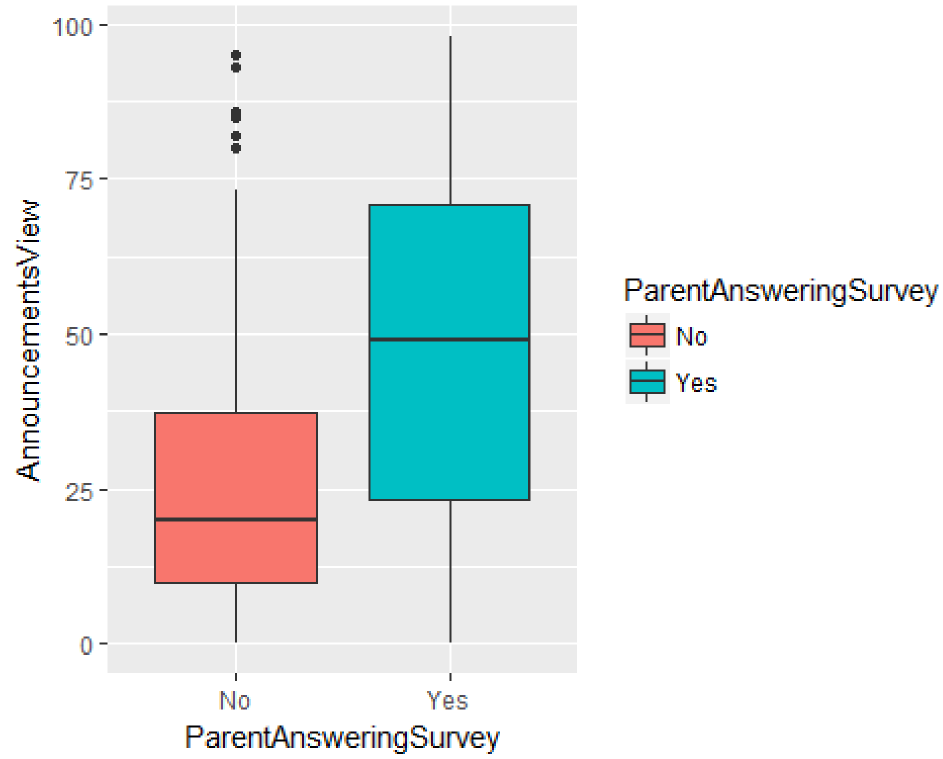


Yes Answers to Surveys = More Raised hands

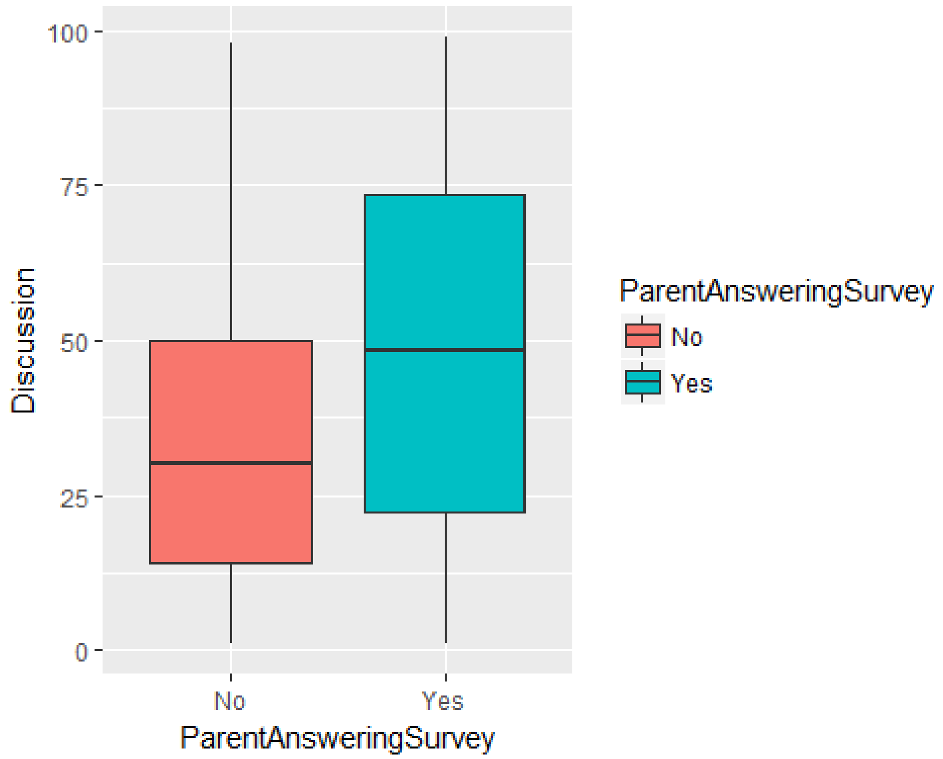
```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = VisITedResources)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = AnnouncementsView)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```

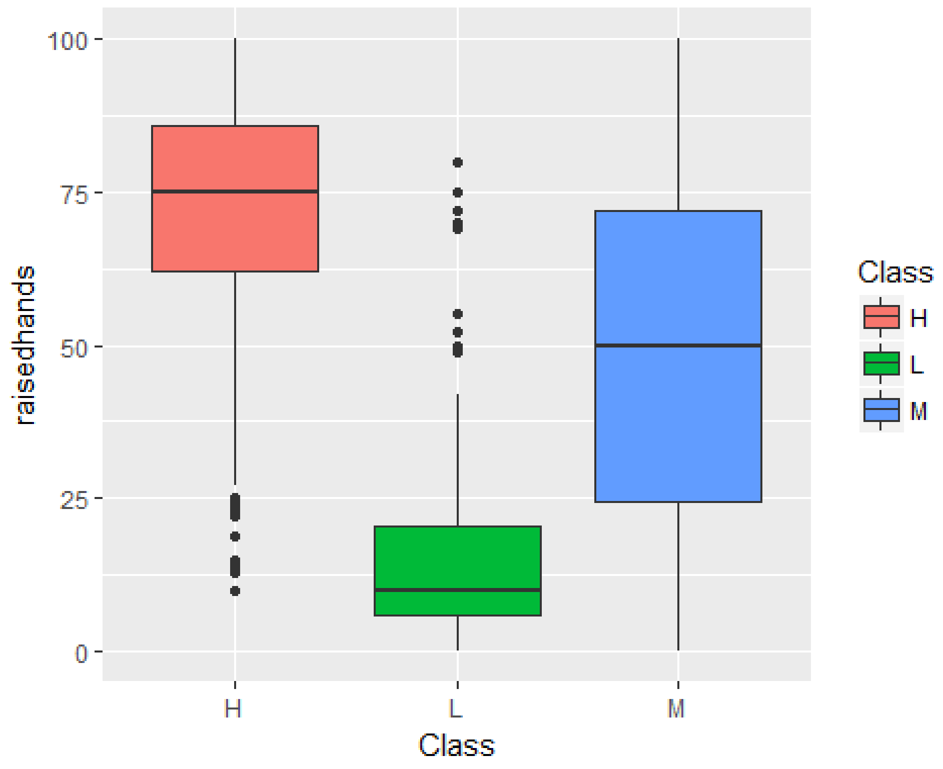


```
ggplot(edu, aes(x = ParentAnsweringSurvey, y = Discussion)) +
geom_boxplot(aes(fill=ParentAnsweringSurvey))
```



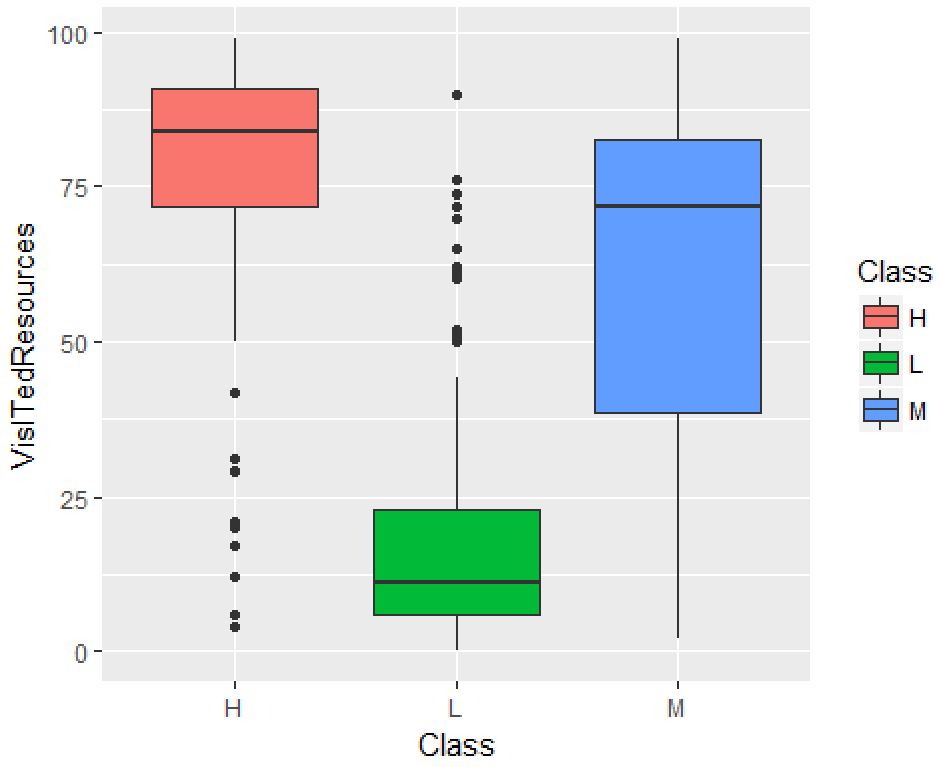
Class-Wise Boxplots

```
ggplot(edu, aes(x = Class, y = raisedhands)) +
geom_boxplot(aes(fill=Class))
```



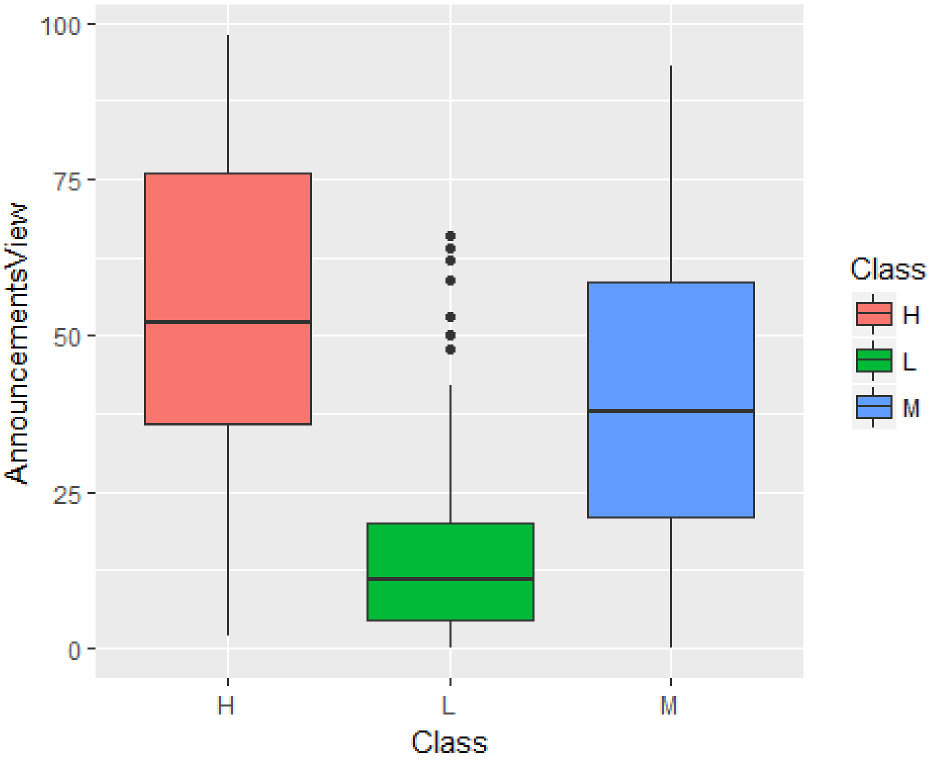
High Marks = Active Participation

```
ggplot(edu, aes(x = Class, y = VisITedResources)) +
geom_boxplot(aes(fill=Class))
```



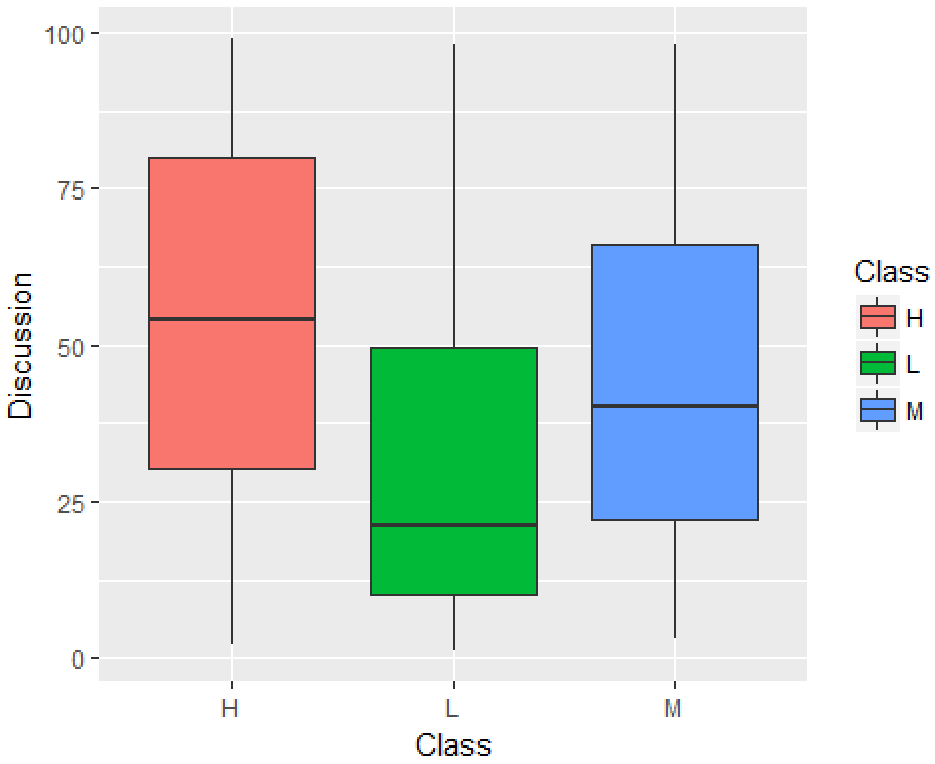
High Marks by visited Resources

```
ggplot(edu, aes(x = Class, y = AnnouncementsView)) +
geom_boxplot(aes(fill=Class))
```



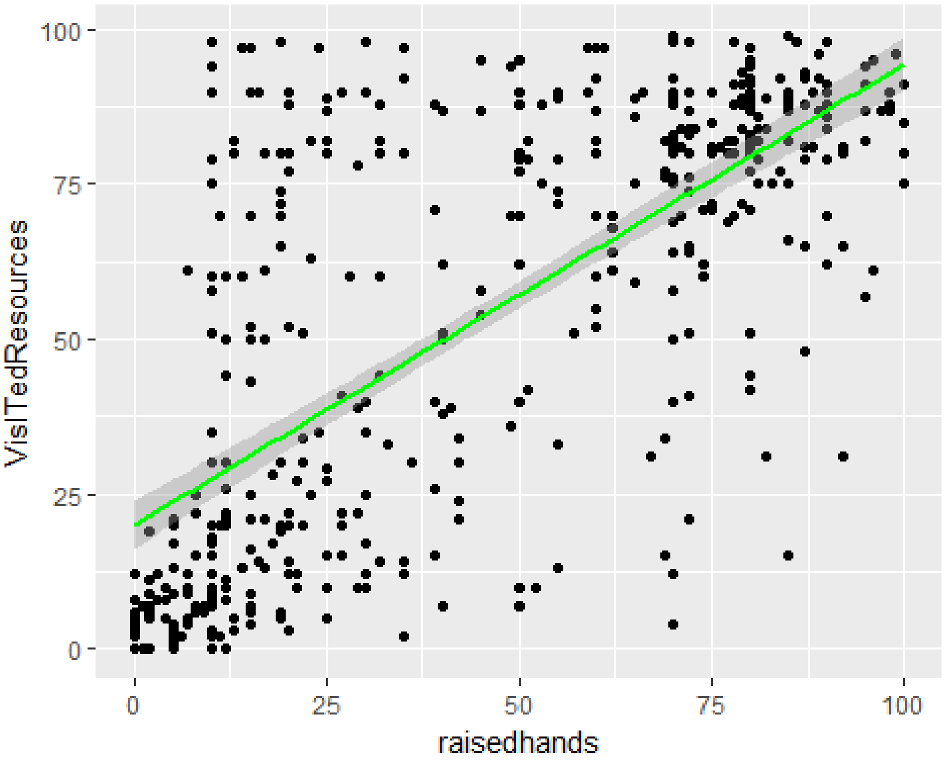
More Marks More Announcements

```
ggplot(edu, aes(x = Class, y = Discussion)) +
geom_boxplot(aes(fill=Class))
```

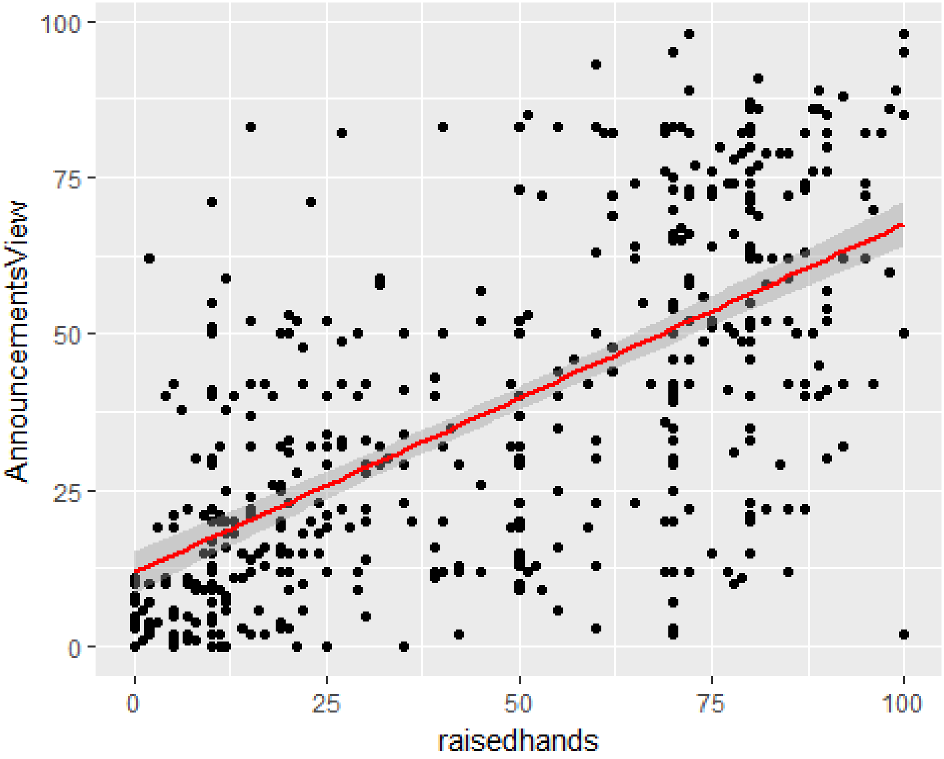


Scatterplots

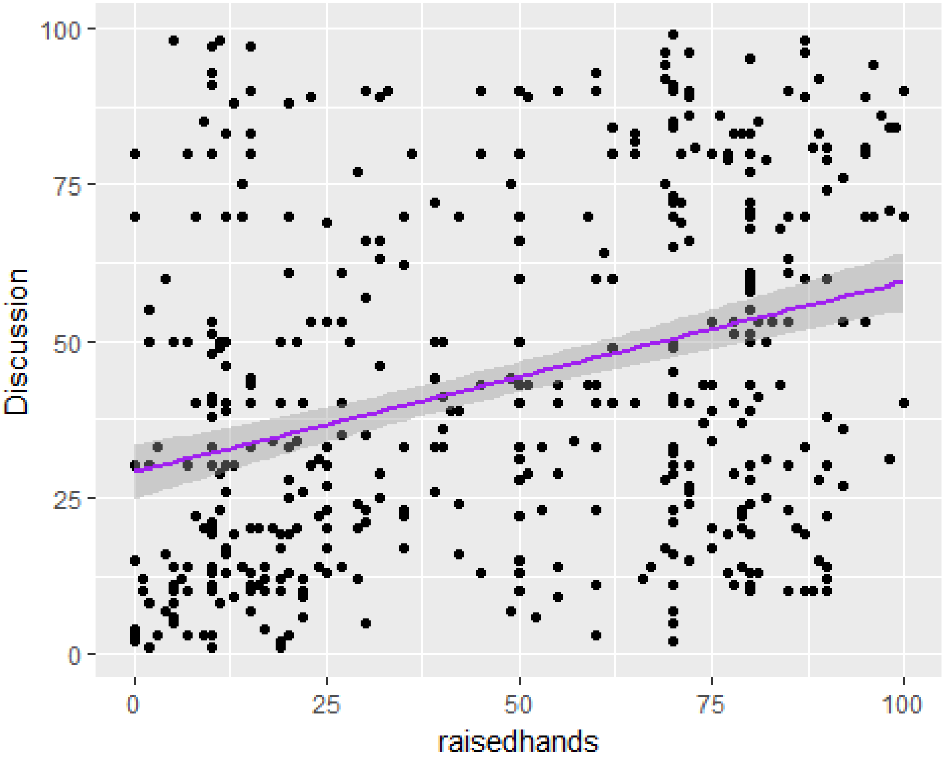
```
ggplot(edu, aes(x = raisedhands, y = VisITedResources)) + geom_point()
+
  geom_smooth(method = "lm",color='green')
```



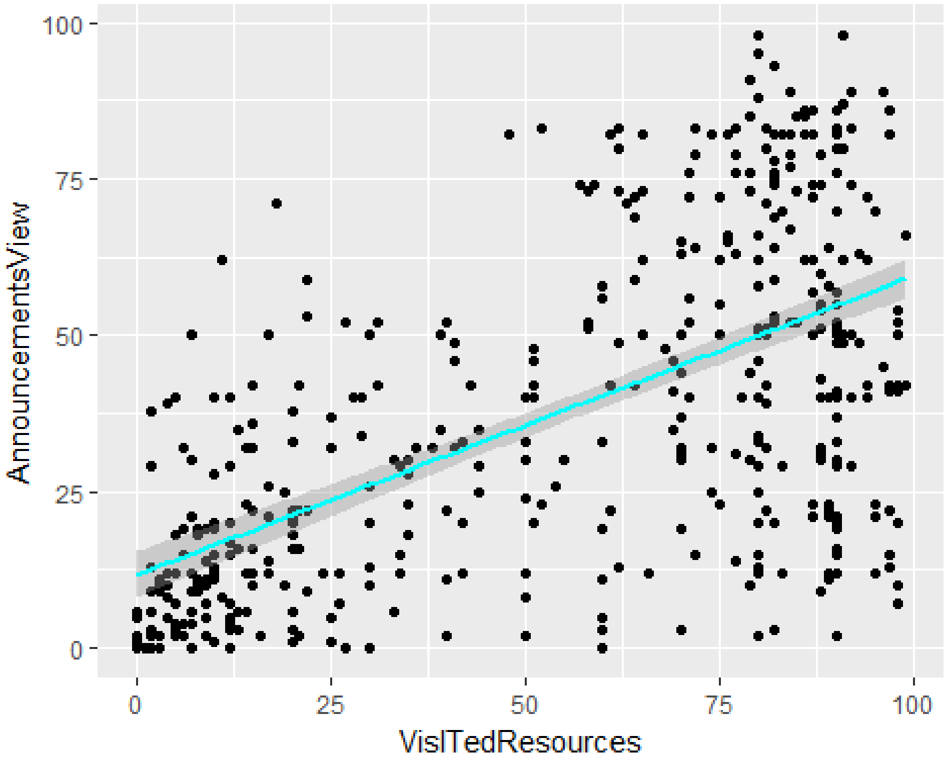
```
ggplot(edu, aes(x = raisedhands, y = AnnouncementsView)) + geom_point()
+
  geom_smooth(method = "lm",color='red')
```



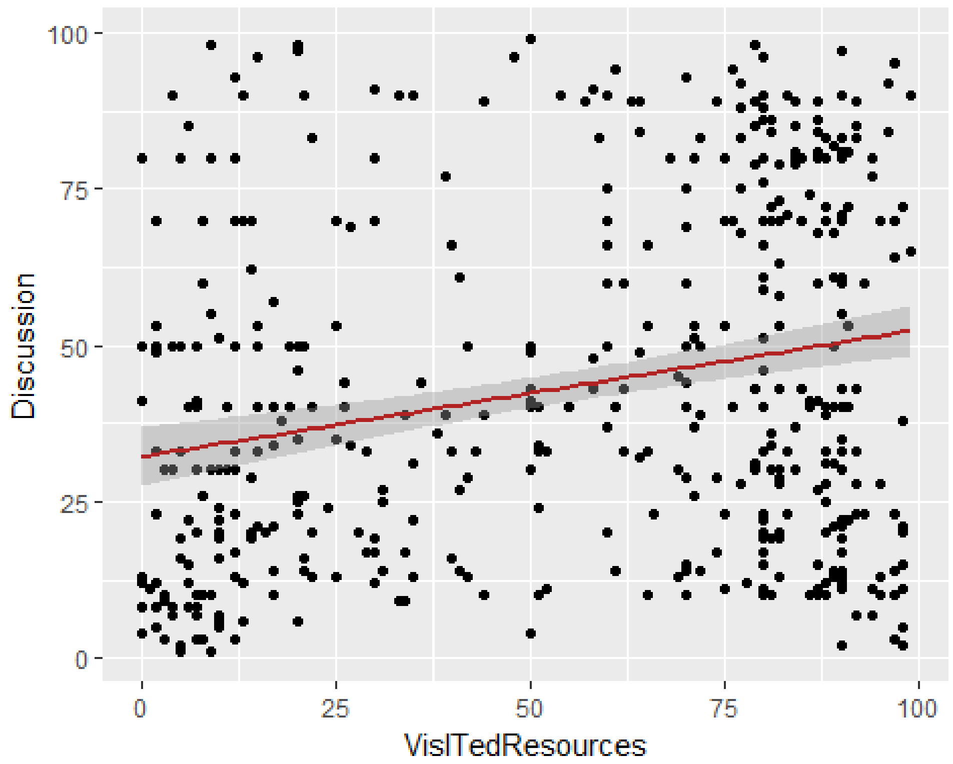
```
ggplot(edu, aes(x = raisedhands, y = Discussion)) + geom_point() +
  geom_smooth(method = "lm",color='purple')
```



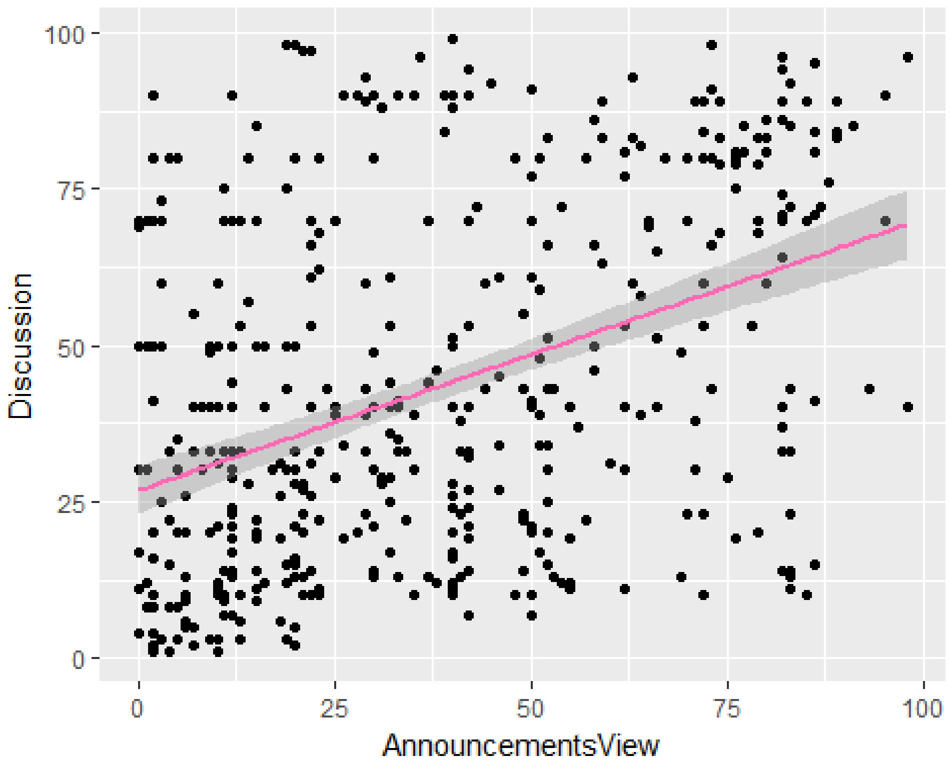
```r
ggplot(edu, aes(x = VisITedResources, y = AnnouncementsView)) +
geom_point() +
  geom_smooth(method = "lm",color='cyan')
```



```r
ggplot(edu, aes(x = VisITedResources, y = Discussion)) + geom_point() +
  geom_smooth(method = "lm",color='firebrick')
```

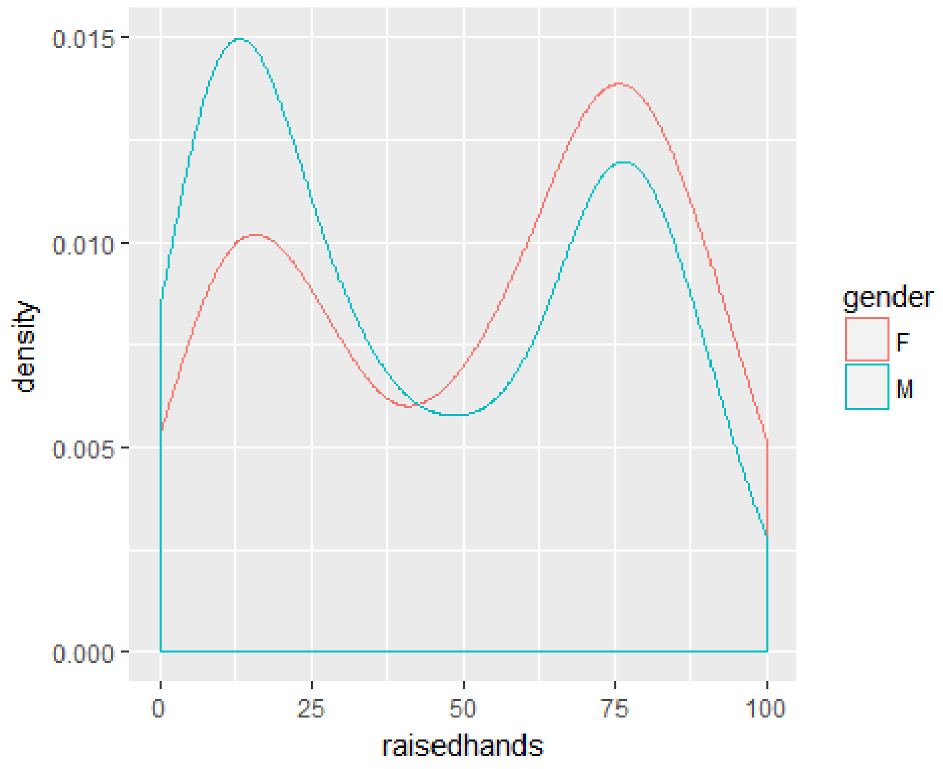


```
ggplot(edu, aes(x = AnnouncementsView, y = Discussion)) + geom_point()
+
  geom_smooth(method = "lm",color='hotpink')
```

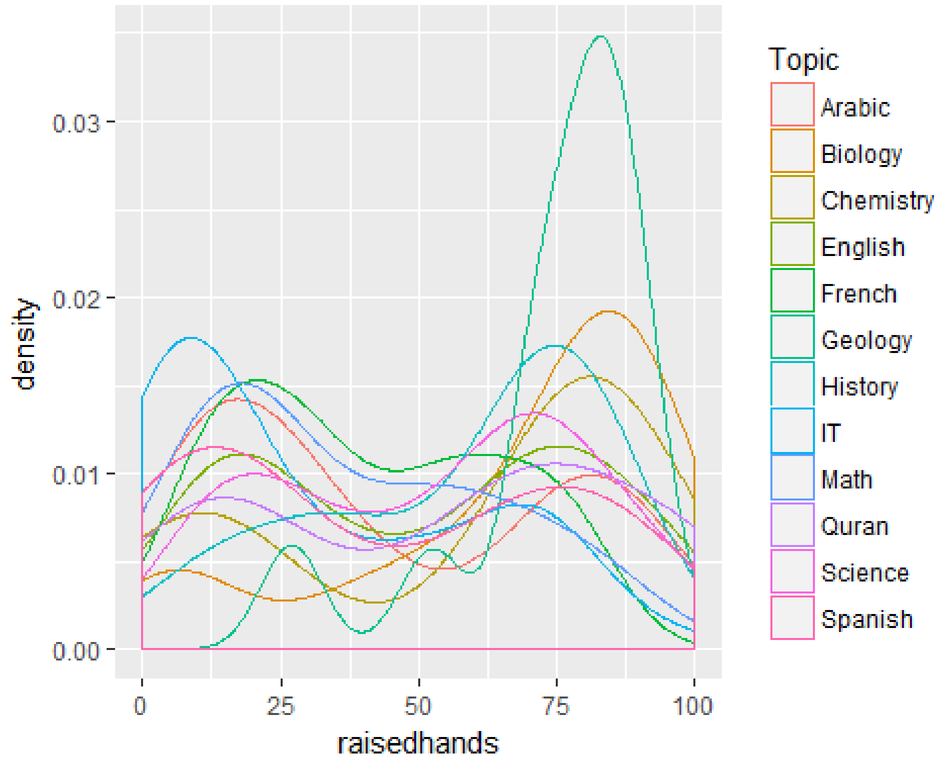


Density Plots

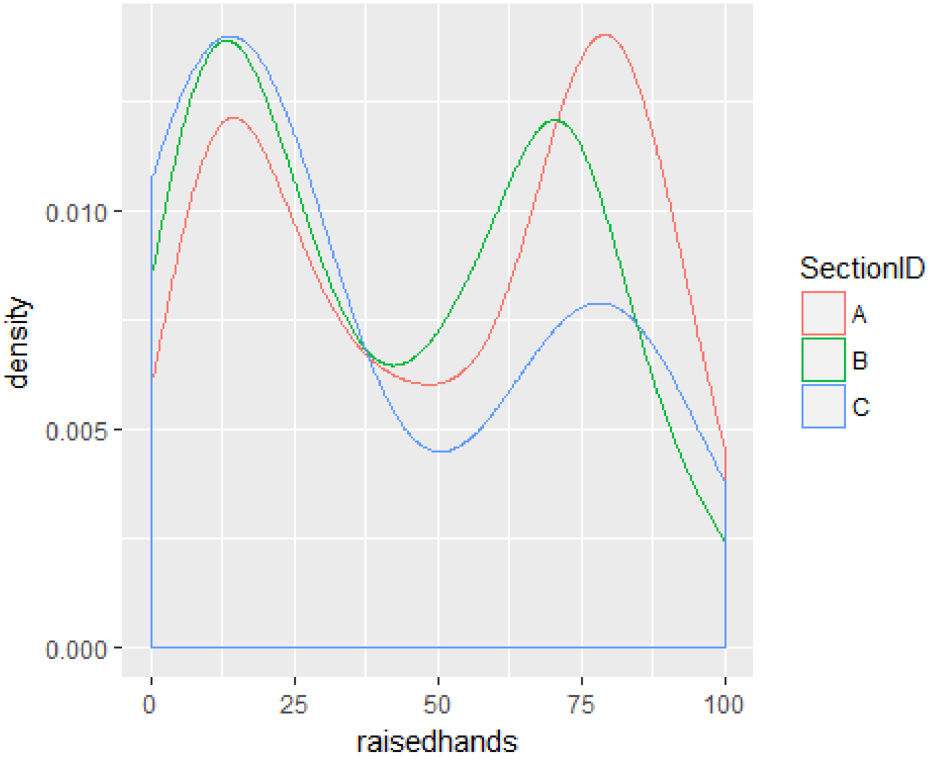
```
ggplot(edu, aes(x = raisedhands, color = gender)) + geom_density()
```



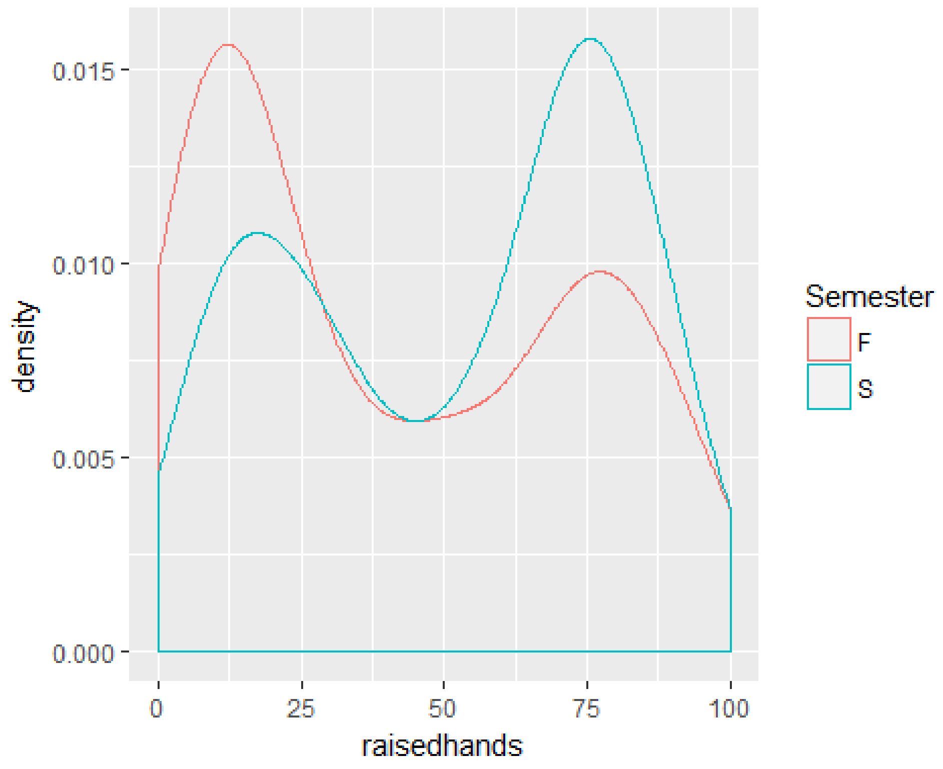
```
ggplot(edu, aes(x = raisedhands, color = Topic)) + geom_density()
```



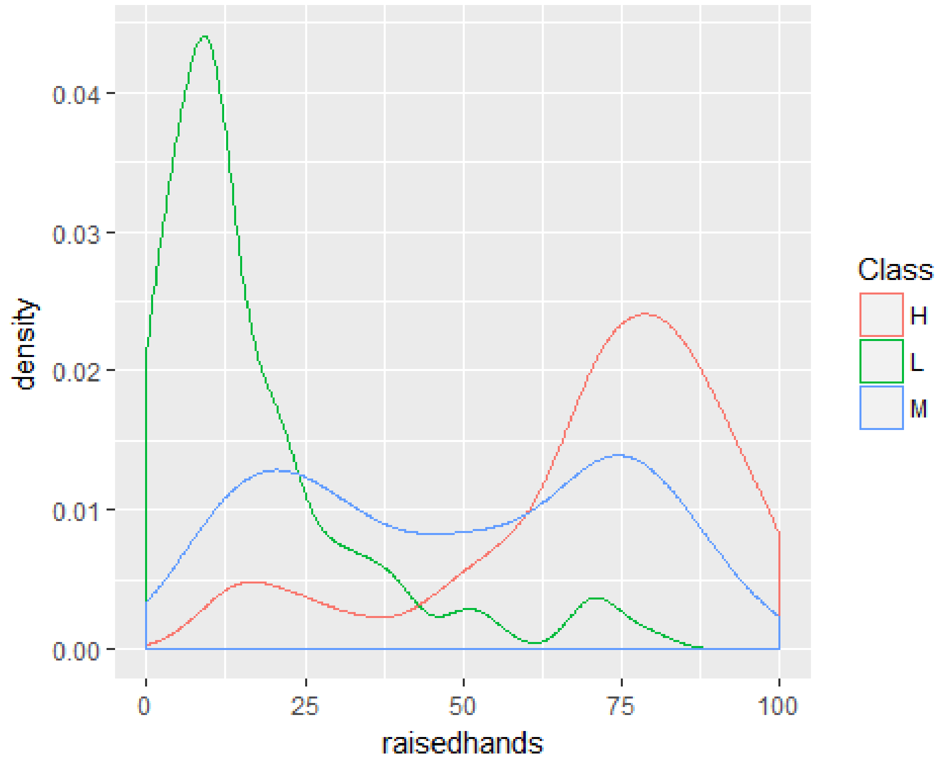
```r
ggplot(edu, aes(x = raisedhands, color = SectionID)) + geom_density()
```



```r
ggplot(edu, aes(x = raisedhands, color = Semester)) + geom_density()
```

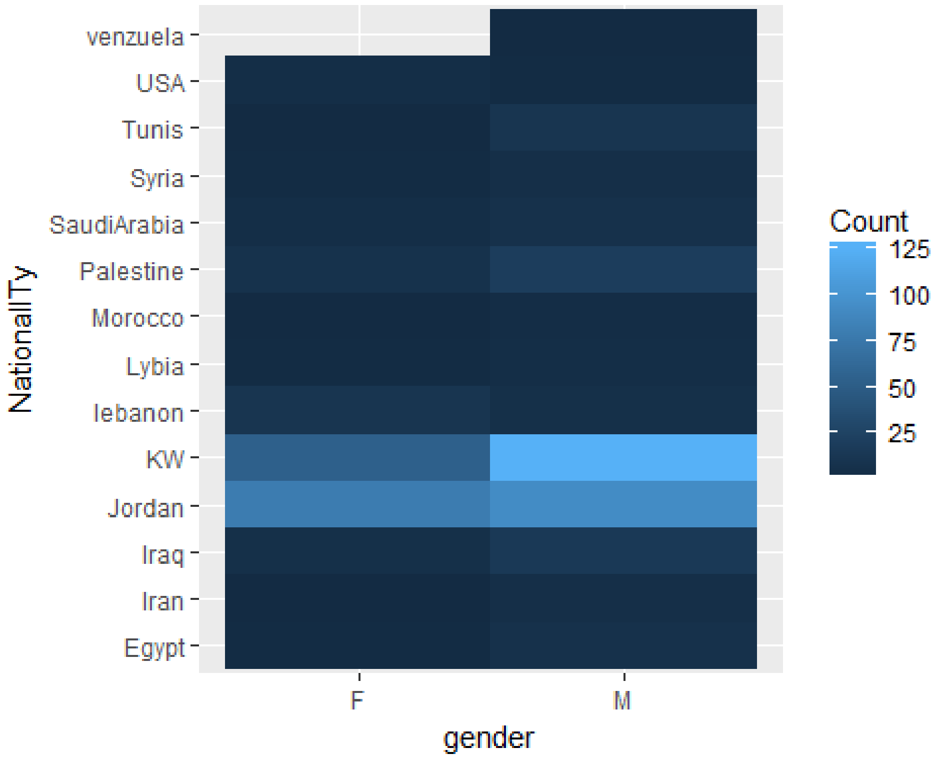


```
ggplot(edu, aes(x = raisedhands, color = Class)) + geom_density()
```



Tile Map

```
tile.map <- edu %>% group_by(gender, NationalITy) %>%
  summarise(Count = n()) %>% arrange(desc(Count))

ggplot(tile.map, aes(x = gender, NationalITy, fill = Count)) +
geom_tile()
```



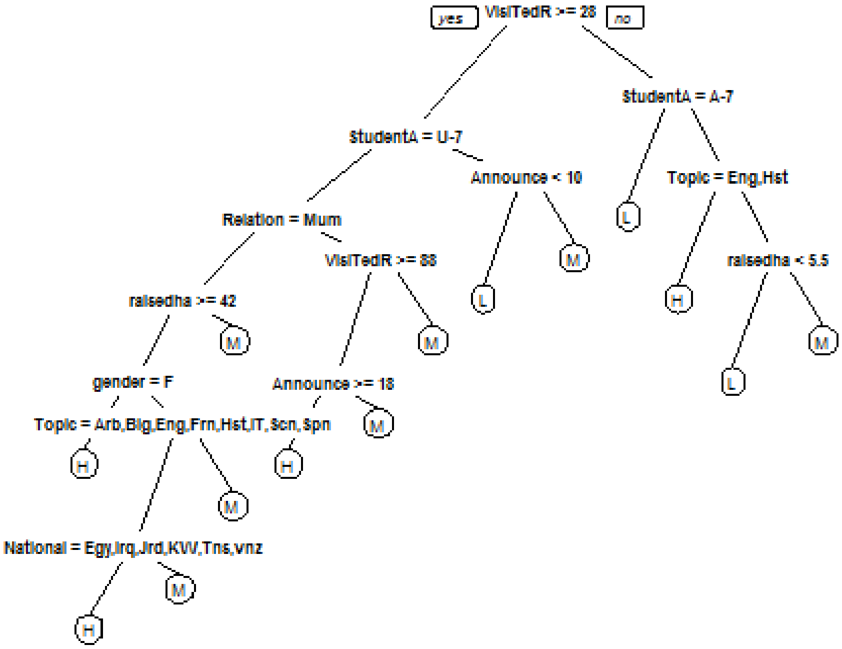
Predictive Modeling

Splitting data into train and cross-validation sets.

```
set.seed(23210)
split <- sample.split(edu$Class, SplitRatio = 0.75)
train <- subset(edu, split == T)
cv <- subset(edu, split == F)
```

Decision Tree

```
tree.model <- rpart(Class ~ ., data = train, method = "class",
minbucket = 1)
prp(tree.model)
```



```
tree.predict <- predict(tree.model, cv, type = "class")
table(cv$Class, tree.predict)

##    tree.predict
##       H  L  M
##    H 26  0 10
##    L  0 22 10
##    M  8  2 43
```

Decision Tree Using Caret Package

```
rpart.control = trainControl(method = "repeatedcv", number = 10,
repeats = 3)
rpart.grid = expand.grid(.cp = seq(0.01, 0.5, 0.02))
rpart.model.caret <-train(Class ~ ., data = train, method = "rpart",
preProcess = "scale",
    trControl = rpart.control, tuneGrid = rpart.grid)

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19,
## uniqueCut = 10, : These variables have zero variances:
NationalITyvenzuela,
## PlaceofBirthvenzuela, GradeIDG-10

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19,
## uniqueCut = 10, : These variables have zero variances:
```

```
NationalITyvenzuela,
## PlaceofBirthvenzuela

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19,
## uniqueCut = 10, : These variables have zero variances:
NationalITyvenzuela,
## PlaceofBirthvenzuela

rpart.predict.caret <- predict.train(rpart.model.caret, cv)
confusionMatrix(rpart.predict.caret, cv$Class)

## Confusion Matrix and Statistics
## 
##           Reference
## Prediction  H  L  M
##          H 26  0 13
##          L  0 25  2
##          M 10  7 38
## 
## Overall Statistics
## 
##                Accuracy : 0.7355
##                  95% CI : (0.6476, 0.8116)
##     No Information Rate : 0.438
##     P-Value [Acc > NIR] : 3.108e-11
## 
##                   Kappa : 0.5906
##  Mcnemar's Test P-Value : NA
## 
## Statistics by Class:
## 
##                      Class: H Class: L Class: M
## Sensitivity            0.7222   0.7812   0.7170
## Specificity            0.8471   0.9775   0.7500
## Pos Pred Value         0.6667   0.9259   0.6909
## Neg Pred Value         0.8780   0.9255   0.7727
## Prevalence             0.2975   0.2645   0.4380
## Detection Rate         0.2149   0.2066   0.3140
## Detection Prevalence   0.3223   0.2231   0.4545
## Balanced Accuracy      0.7846   0.8794   0.7335
```

Accuracy -> 0.7355

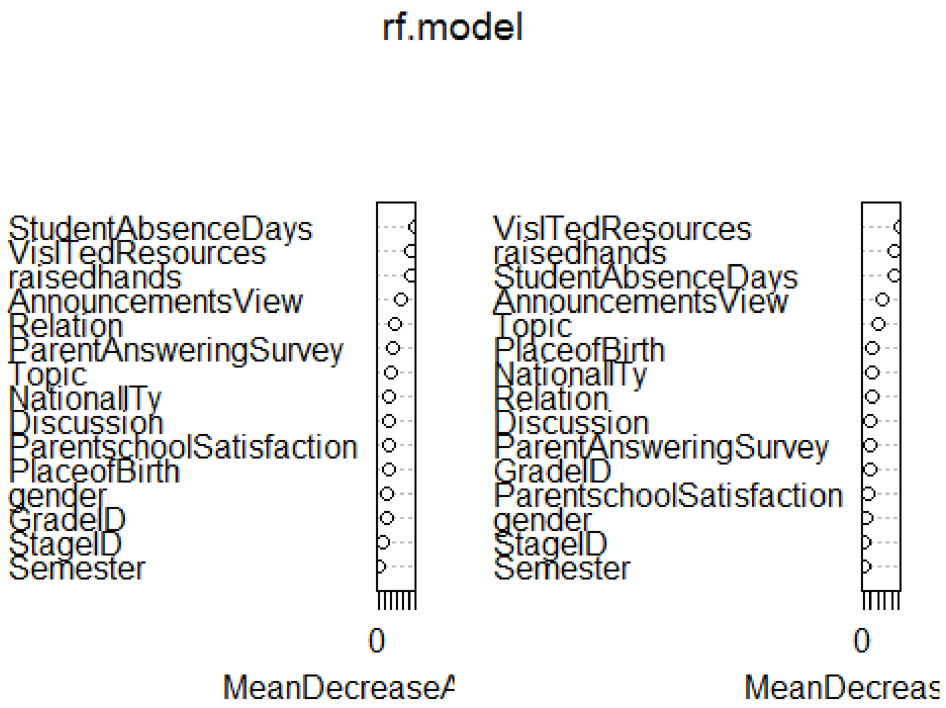
Random Forest

```
set.seed(10005)

rf.model <- randomForest(Class ~ .- SectionID , data = train,
importance = TRUE,
      ntree = 2000, nodesize = 20)
```

```
rf.predict <- predict(rf.model, cv)
confusionMatrix(cv$Class, rf.predict)
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction  H  L  M
##          H 19  0 17
##          L  0 23  9
##          M 11  2 40
## 
## Overall Statistics
## 
##                Accuracy : 0.6777
##                  95% CI : (0.5867, 0.7598)
##     No Information Rate : 0.5455
##     P-Value [Acc > NIR] : 0.002089
## 
##                   Kappa : 0.4906
##  Mcnemar's Test P-Value : NA
## 
## Statistics by Class:
## 
##                      Class: H Class: L Class: M
## Sensitivity            0.6333   0.9200   0.6061
## Specificity            0.8132   0.9062   0.7636
## Pos Pred Value         0.5278   0.7188   0.7547
## Neg Pred Value         0.8706   0.9775   0.6176
## Prevalence             0.2479   0.2066   0.5455
## Detection Rate         0.1570   0.1901   0.3306
## Detection Prevalence   0.2975   0.2645   0.4380
## Balanced Accuracy      0.7233   0.9131   0.6848
```

```
varImpPlot(rf.model)
```



Accuracy -> 0.6777

C-Forest Utilizing Party

```
cforest.model = cforest(Class ~ .-SectionID , data = train,
    controls=cforest_unbiased(ntree=2000, mtry = 3))

cforest.prediction = predict(cforest.model, cv, OOB = TRUE, type =
"response")
confusionMatrix(cv$Class, cforest.prediction)

## Confusion Matrix and Statistics
##
##             Reference
## Prediction   H  L  M
##            H 23  0 13
##            L  0 25  7
##            M  9  2 42
##
## Overall Statistics
##
##                Accuracy : 0.7438
##                  95% CI : (0.6565, 0.8188)
##     No Information Rate : 0.5124
##     P-Value [Acc > NIR] : 1.596e-07
##
```

```
##                    Kappa : 0.5984
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: H Class: L Class: M
## Sensitivity            0.7188   0.9259   0.6774
## Specificity            0.8539   0.9255   0.8136
## Pos Pred Value         0.6389   0.7812   0.7925
## Neg Pred Value         0.8941   0.9775   0.7059
## Prevalence             0.2645   0.2231   0.5124
## Detection Rate         0.1901   0.2066   0.3471
## Detection Prevalence   0.2975   0.2645   0.4380
## Balanced Accuracy      0.7863   0.9257   0.7455
```

Accuracy -> 0.7438

Suppost Vector Machines

```
svm.model <- svm(Class ~ ., data = train, kernel = "radial", cost = 10,
gamma = 0.15)
svm.predict <- predict(svm.model, cv)
confusionMatrix(cv$Class, svm.predict)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  H  L  M
##          H 26  0 10
##          L  1 27  4
##          M  7  5 41
##
## Overall Statistics
##
##                Accuracy : 0.7769
##                  95% CI : (0.6922, 0.8475)
##     No Information Rate : 0.4545
##     P-Value [Acc > NIR] : 4.519e-13
##
##                   Kappa : 0.6553
##  Mcnemar's Test P-Value : 0.6502
##
## Statistics by Class:
##
##                      Class: H Class: L Class: M
## Sensitivity            0.7647   0.8438   0.7455
## Specificity            0.8851   0.9438   0.8182
## Pos Pred Value         0.7222   0.8438   0.7736
## Neg Pred Value         0.9059   0.9438   0.7941
## Prevalence             0.2810   0.2645   0.4545
## Detection Rate          0.2149    0.2231    0.3388
## Detection Prevalence    0.2975    0.2645    0.4380
## Balanced Accuracy       0.8249    0.8938    0.7818
```

Accuracy -> 0.777

Ensemble Model

```
results <- data.frame(tree = tree.predict, rpart = rpart.predict.caret, 
rf = rf.predict,
      cforest = cforest.prediction, svm = svm.predict,
      actual.class = cv$Class, final.prediction = rep("-",nrow(cv)))

results
```

```
##     tree rpart rf cforest svm actual.class final.prediction
## 9      M     M  M       M   M            M                -
## 11     M     M  M       M   M            H                -
## 20     H     H  M       M   M            H                -
## 25     L     L  L       L   L            L                -
## 27     M     M  M       M   M            M                -
## 41     L     L  L       L   L            L                -
## 46     L     L  L       L   L            L                -
## 47     L     L  L       L   L            L                -
## 49     M     M  M       M   M            H                -
## 54     H     H  M       H   H            H                -
## 57     L     L  L       L   M            L                -
## 59     M     M  M       M   M            M                -
## 62     M     M  M       M   M            M                -
## 64     M     M  M       M   L            M                -
## 68     H     H  H       H   H            H                -
## 71     M     L  M       M   M            L                -
## 72     M     M  M       M   M            M                -
## 81     M     M  M       L   L            L                -
## 94     H     M  M       M   H            H                -
## 97     M     M  M       M   M            H                -
## 102    H     H  H       H   H            H                -
## 117    M     M  M       M   M            M                -
## 128    L     L  M       L   L            L                -
## 129    M     M  M       L   L            L                -
## 130    M     M  M       M   H            L                -
## 132    M     M  M       M   M            M                -
## 134    L     M  L       L   L            L                -
## 138    H     H  H       H   H            M                -
## 145    L     L  L       L   L            L                -
## 148    M     M  M       M   M            M                -
## 154    M     M  M       M   M            L                -
## 159    M     L  M       M   L            L                -
## 162    H     H  H       H   M            M                -
## 167    H     H  M       M   M            M                -
## 169    H     H  M       M   M            H                -
## 170    M     M  M       M   L            M                -
## 175    M     M  M       M   M            M                -
## 176    L     L  L       L   L            L                -
## 177    H     H  M       M   M            H                -
## 182    M     M  M       M   L            L                -
## 184    M     M  M       M   M            M                -
## 185    M     L  L       L   L            L                -
## 186    H     H  H       H   M            H                -
## 188    M     M  H       H   H            M                -
## 192    L     L  L       M   L            L                -
## 195    M     H  M       M   M            M                -
## 196    M     M  M       M   M            M                -
## 203    M     M  M       M   L            M                -
## 208    M     M  M       M   M            M                -
## 211    M     M  M       M   M            H                -
## 213    H     H  M       M   M            H                -
## 216    L     L  L       L   L            L                -
## 219    H     H  M       H   H            H                -
## 220    H     H  M       M   M            M                -
## 224    H     H  M       M   H            H                -
## 230    L     L  L       L   L            L                -
## 232    L     L  L       L   L            L                -
## 234    M     M  M       M   M            M                -
## 237    L     L  L       L   L            L                -
## 239    M     M  M       M   M            M                -
## 241    H     H  H       H   H            H                -
## 244    L     L  L       L   L            M                -
## 251    M     M  M       M   H            H                -
## 261    L     L  L       L   L            L                -
## 265    H     H  H       H   H            M                -
## 270    M     M  M       M   M            M                -
## 272    H     H  H       H   H            M                -
## 273    H     M  M       M   H            M                -
## 278    M     H  H       H   M            H                -
## 279    M     H  H       H   M            M                -
## 284    H     H  H       H   H            H                -
## 290    M     H  H       H   M            M                -
## 295    M     M  M       M   M            M                -
## 302    M     L  L       L   L            L                -
## 311    M     M  M       M   M            M                -
## 314    H     H  H       H   H            H                -
## 318    H     H  H       H   H            H                -
## 327    L     L  L       L   L            L                -
## 330    H     H  H       H   M            M                -
## 332    L     L  L       L   L            L                -
## 335    L     L  L       L   L            L                -
## 336    L     L  L       L   L            L                -
## 341    M     M  M       M   M            M                -
## 345    M     M  M       M   H            H                -
## 348    M     M  M       M   H            H                -
## 350    L     L  L       L   L            L                -
## 352    L     L  L       L   L            L                -
## 353    M     M  M       M   H            M                -
## 356    M     M  M       H   H            H                -
## 357    M     M  M       M   H            M                -
## 366    M     M  H       H   H            H                -
## 367    L     L  L       L   M            M                -
## 369    H     H  H       H   H            H                -
## 386    H     H  H       H   H            H                -
## 392    M     M  M       M   M            M                -
## 393    H     H  H       H   H            H                -
## 397    M     M  H       M   M            M                -
## 402    M     M  M       M   M            M                -
## 408    L     L  L       L   L            L                -
## 409    M     M  M       M   M            M                -
## 412    H     H  H       H   H            H                -
## 413    M     M  H       M   M            M                -
## 417    H     H  H       H   H            H                -
## 418    H     H  H       H   H            H                -
## 423    M     M  M       M   M            M                -
## 425    H     H  H       H   H            H                -
## 427    M     M  M       M   L            M                -
## 433    M     M  M       M   M            M                -
## 435    M     M  M       M   M            M                -
## 436    M     M  M       M   M            M                -
## 440    M     H  H       H   M            M                -
## 441    M     H  M       M   M            M                -
## 442    M     H  M       M   M            M                -
## 446    M     M  M       M   M            L                -
## 450    H     H  M       H   H            H                -
## 460    H     H  H       H   H            H                -
## 464    M     M  M       M   M            M                -
## 465    H     H  H       H   H            H                -
## 467    H     H  H       H   H            H                -
## 471    M     M  M       M   M            M                -
## 478    M     M  M       M   M            M                -
```

```
getmode <- function(x) {
  unique.x <- unique(x)
  unique.x[which.max(tabulate(match(x, unique.x)))]
}

results$final.prediction <- apply(results, 1, getmode)
confusionMatrix(results$actual.class, results$final.prediction)
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction  H  L  M
##            H 28  0  8
##            L  0 24  8
##            M  5  2 46
## 
## Overall Statistics
##                                           
##                Accuracy : 0.8099          
##                  95% CI : (0.7286, 0.8755)
##     No Information Rate : 0.5124          
##     P-Value [Acc > NIR] : 1.018e-11       
##                                           
##                   Kappa : 0.7019          
##  Mcnemar's Test P-Value : NA              
## 
## Statistics by Class:
## 
##                      Class: H Class: L Class: M
## Sensitivity            0.8485   0.9231   0.7419
## Specificity            0.9091   0.9158   0.8814
## Pos Pred Value         0.7778   0.7500   0.8679
## Neg Pred Value         0.9412   0.9775   0.7647
## Prevalence             0.2727   0.2149   0.5124
## Detection Rate         0.2314   0.1983   0.3802
## Detection Prevalence   0.2975   0.2645   0.4380
## Balanced Accuracy      0.8788   0.9194   0.8116
```

Accuracy -> 0.810 (best)