```r
# Importing Dataset
edudata <- read.csv("dataset.csv")
```

```r
# Imporitng the libraries
library(ggplot2)
library(reshape2)
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(MASS)
library(nnet)
```

```r
# Renaming the Columns
colnames(edudata)[colnames(edudata)=="gender"] <- "Gender"
colnames(edudata)[colnames(edudata)=="NationalITy"] <- "Nationality"
colnames(edudata)[colnames(edudata)=="raisedhands"] <- "RaisedHands"
colnames(edudata)[colnames(edudata)=="VisITedResources"] <- "VisitedResources"
edudata$Class <- factor(edudata$Class, levels = c("L","M","H"))
```

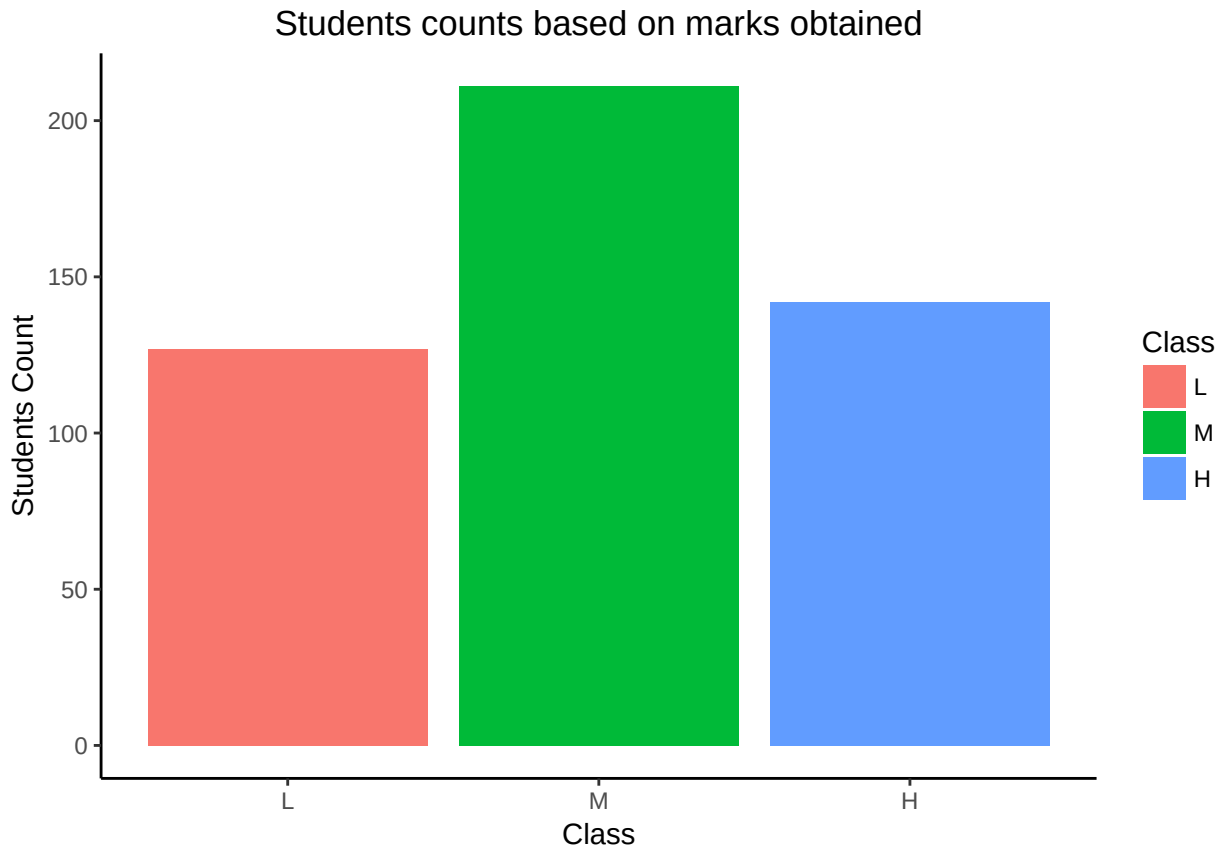
```r
# Exploratory Data Analysis
summary(edudata)
```

```
##  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        L:127
##  Yes:270               Good:292                 Under-7:289        M:211
##                                                                    H:142
##
##
##
##
```

```r
# Finding any missing values
sapply(edudata, function(x) { sum(is.na(x))})
```

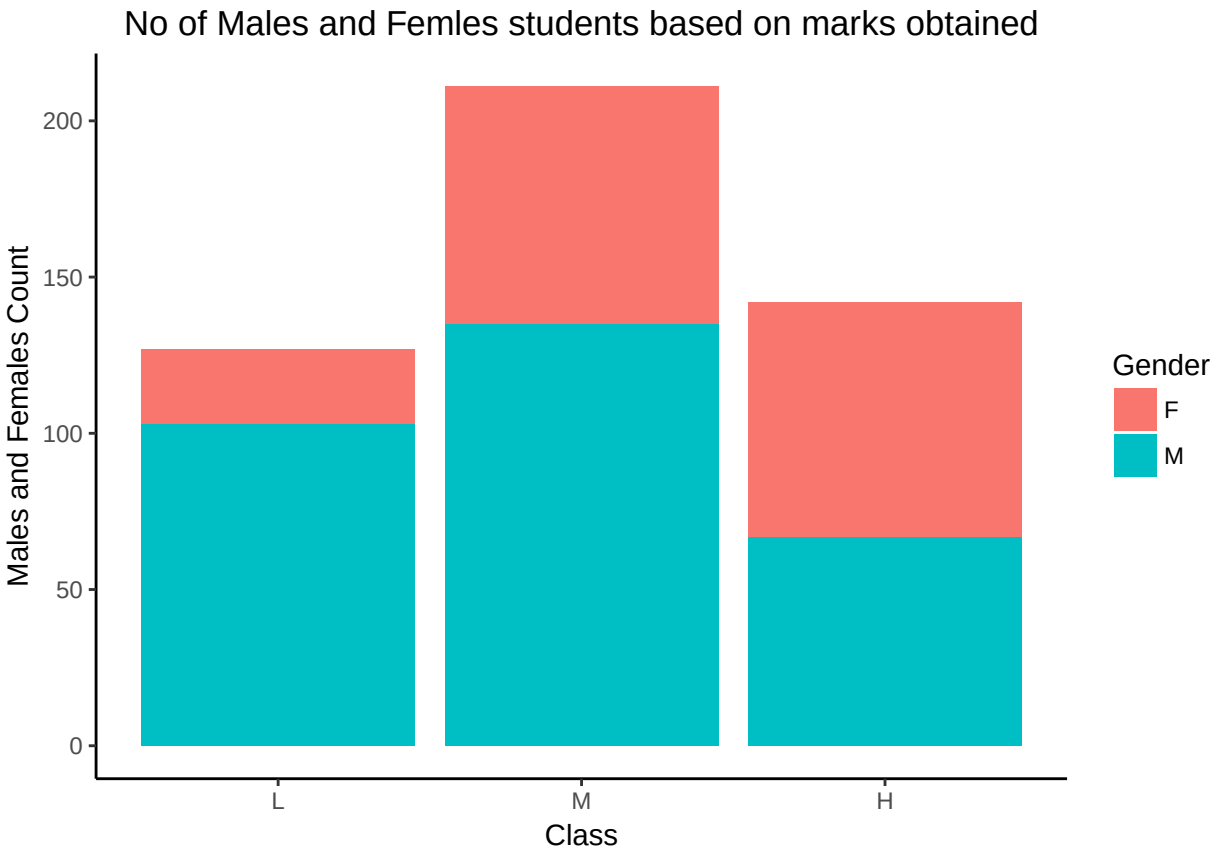
```
##                    Gender             Nationality             PlaceofBirth
##                         0                       0                        0
##                 StageID                  GradeID               SectionID
##                         0                       0                        0
##                   Topic                 Semester                Relation
##                         0                       0                        0
##             RaisedHands          VisitedResources        AnnouncementsView
##                         0                       0                        0
##              Discussion    ParentAnsweringSurvey ParentschoolSatisfaction
##                         0                       0                        0
##       StudentAbsenceDays                   Class
##                         0                       0
```

```r
# Student Class Performance
ggplot(data = edudata, aes(x = Class, fill = Class)) + geom_bar()+ggtitle("Students counts based on mark
  theme(plot.title = element_text(hjust = 0.5))+ylab("Students Count")
```



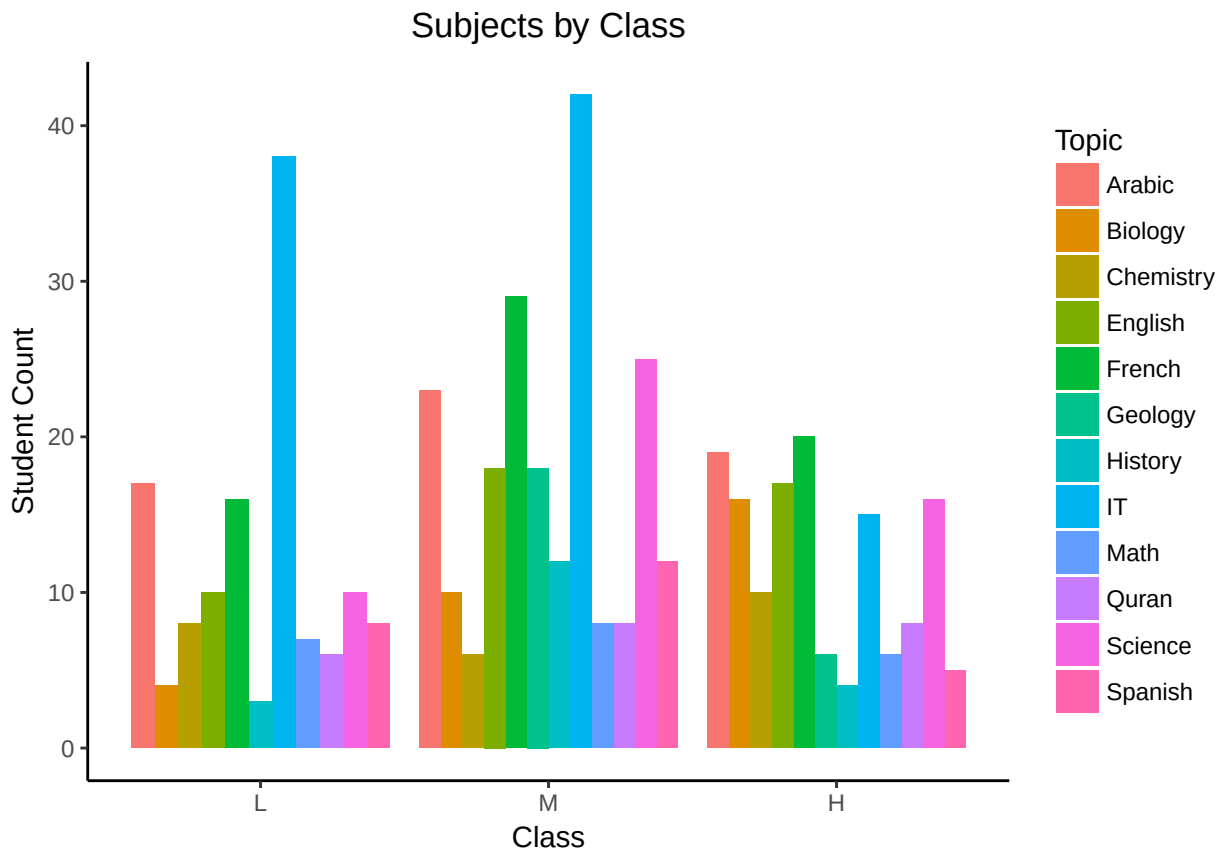
We can conculde that Maximum students performance is in Middle Class. That means most of the student marks/grades are in interval between 70 to 89,

```r
# Gender based Class performance Calculation
ggplot(data = edudata, aes(x = Class, fill = Gender)) + geom_bar() +ggtitle("No of Males and Femles stu
  theme(plot.title = element_text(hjust = 0.5))+ylab("Males and Females Count")
```



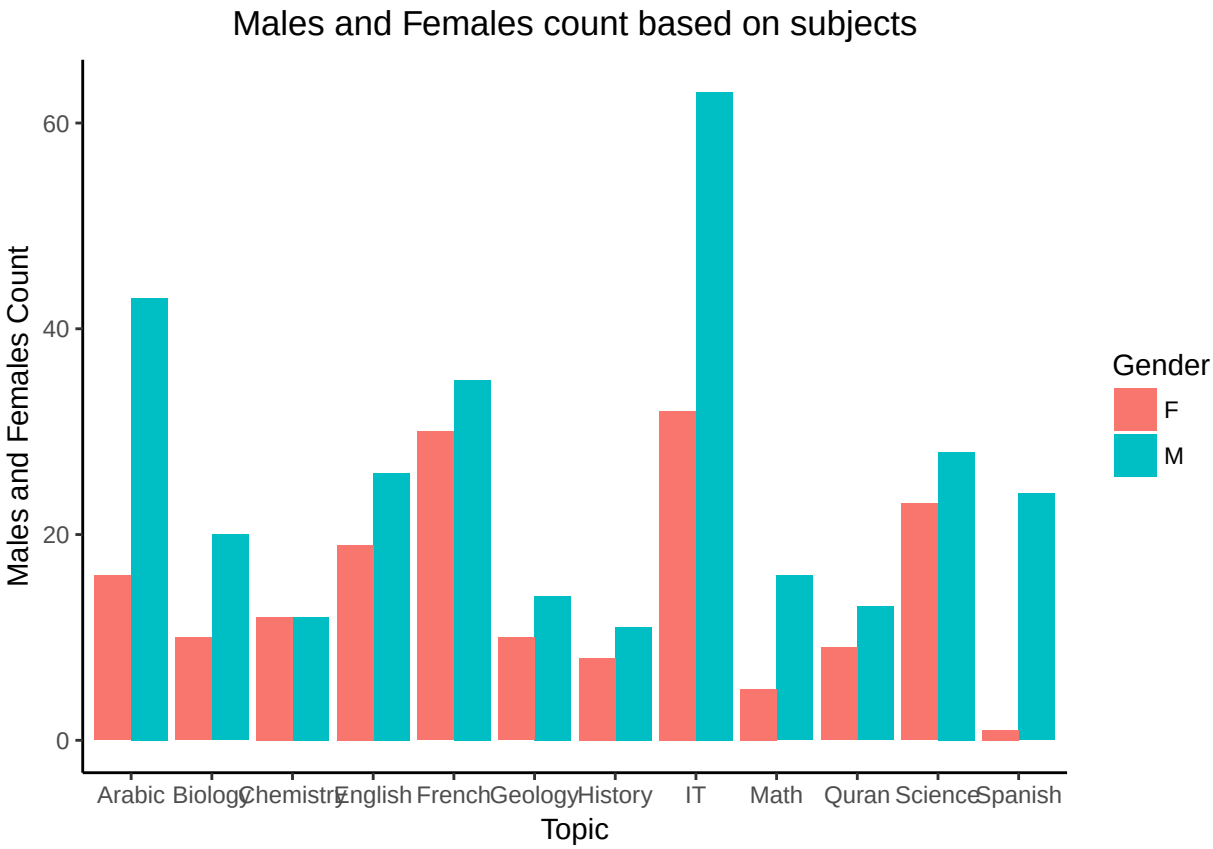
From the plot we can conculde that very few female students are there in low class grade whereas males have higher ratio of lower grades. Therefore, they are categorize under low class.

```
# Student Topics baseed on Class
ggplot(data = edudata, aes(x = Class, fill = Topic)) + geom_bar(position = "dodge")+ggtitle("Subjects by
  theme(plot.title = element_text(hjust = 0.5))+ylab("Student Count")
```



Most students are studying IT in M and L class. In H class most students are studying French.

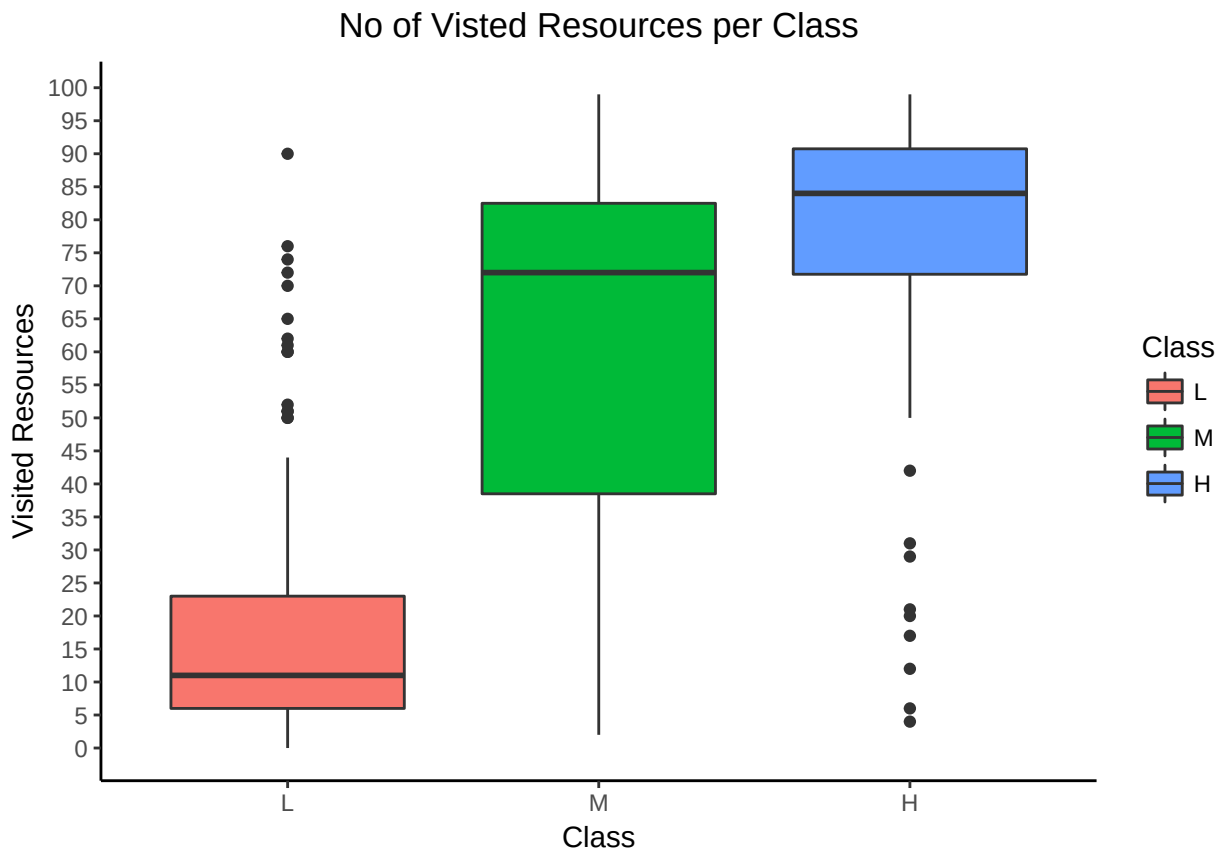
```
# Student count based on Gender for Topics
ggplot(data = edudata, aes(x = Topic, fill = Gender)) + geom_bar(position = "dodge")+ggtitle("Males and
  theme(plot.title = element_text(hjust = 0.5))+ylab("Males and Females Count")
```



IT has most number of males student

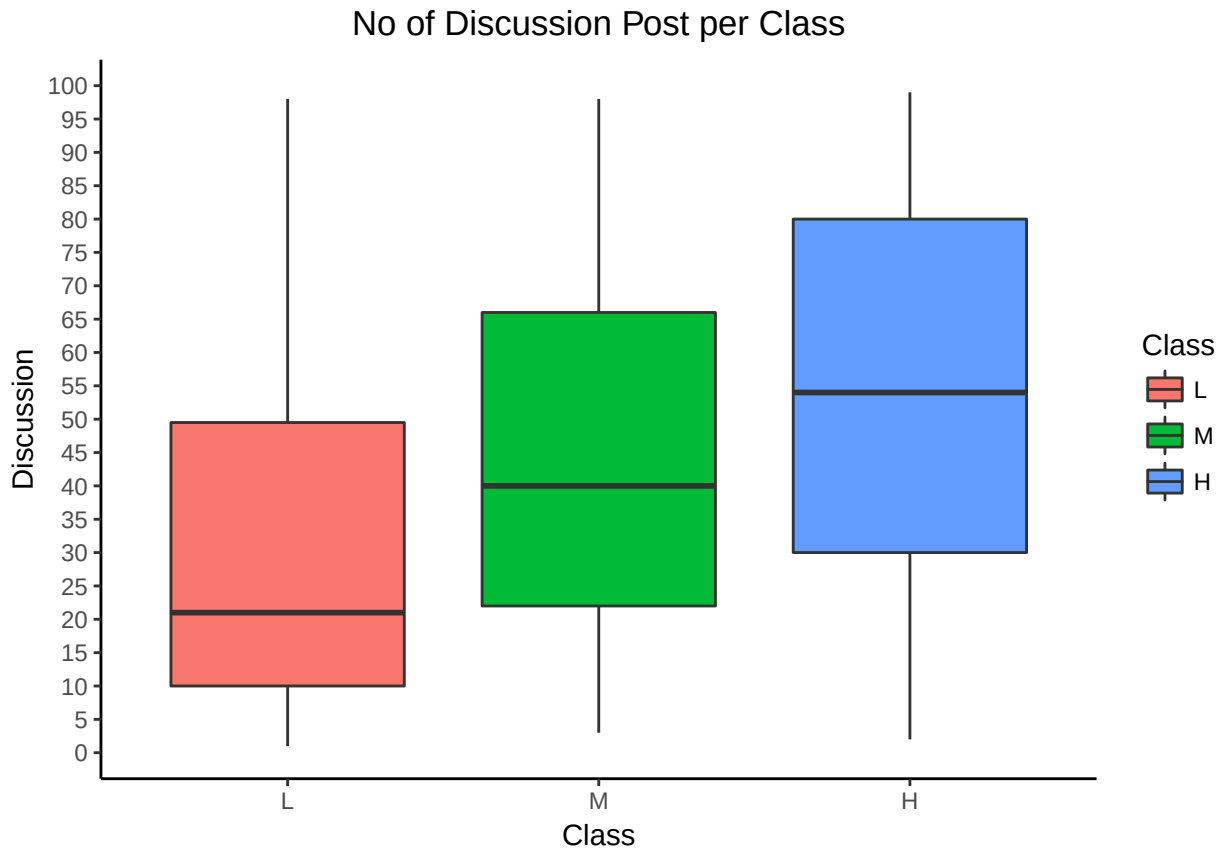
Box Plots

```
# Students that visit more resources posted by instructor gets higher marks
ggplot(data = edudata, aes(x = Class, fill = Class, y = VisitedResources)) + geom_boxplot() +
    labs(x = "Class", y = "Visited Resources") +
    scale_y_continuous(breaks = seq(0,100,5))+ggtitle("No of Visted Resources per Class")+theme_classic
  theme(plot.title = element_text(hjust = 0.5))+ylab("Visited Resources")
```



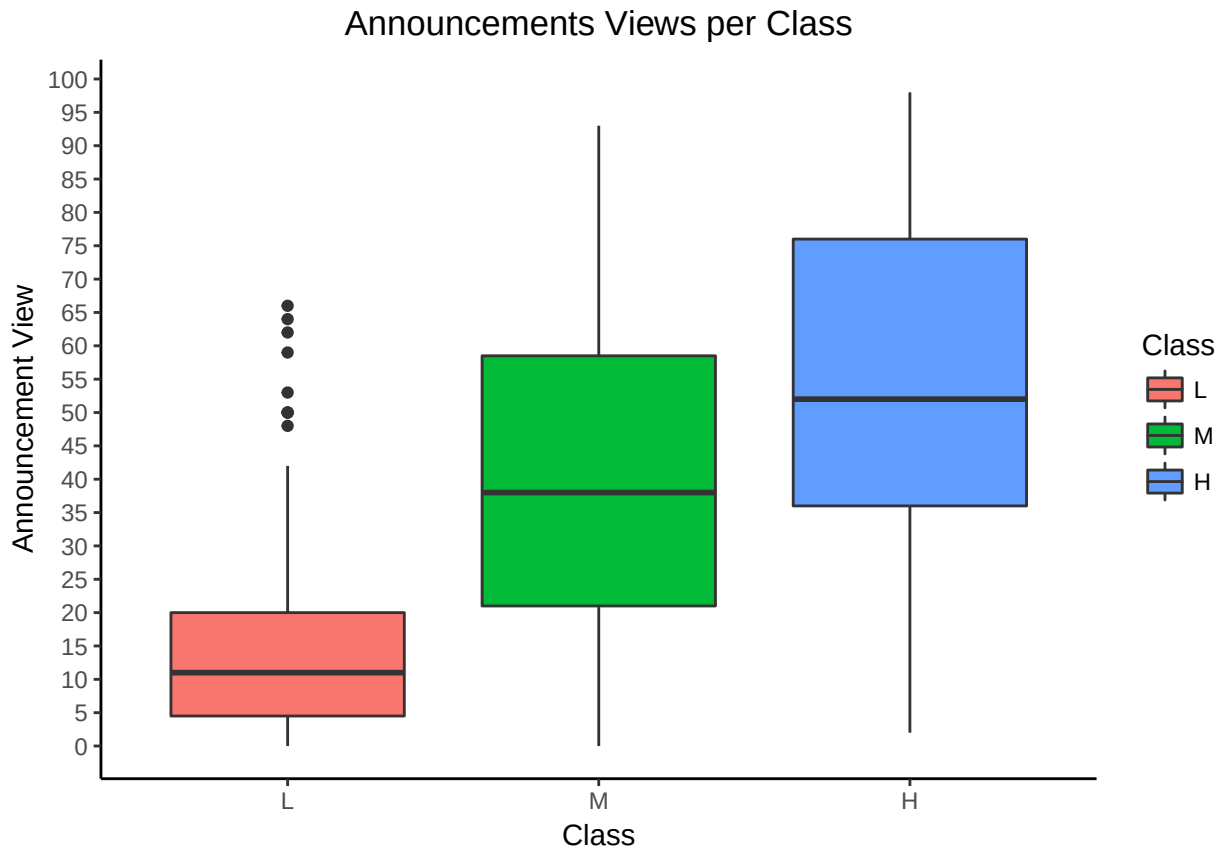
From the box plot we can see that on average student that visited resources online got higher grades.Lower Class perofrmance and Higher class peroformance students have few outliers.

```
# Box plot for Discussion post and students grades
ggplot(data = edudata, aes(x = Class, fill= Class, y = Discussion)) + geom_boxplot() +
    labs(x = "Class", y = "Discussion Post") +
    scale_y_continuous(breaks = seq(0,100,5))+ggtitle("No of Discussion Post per Class")+theme_classic()
  theme(plot.title = element_text(hjust = 0.5))+ylab("Discussion")
```



From the box plot we can see that the student who gets higher marks in class participates more in disucssion forums and lower grades student participate less in discussion formums.

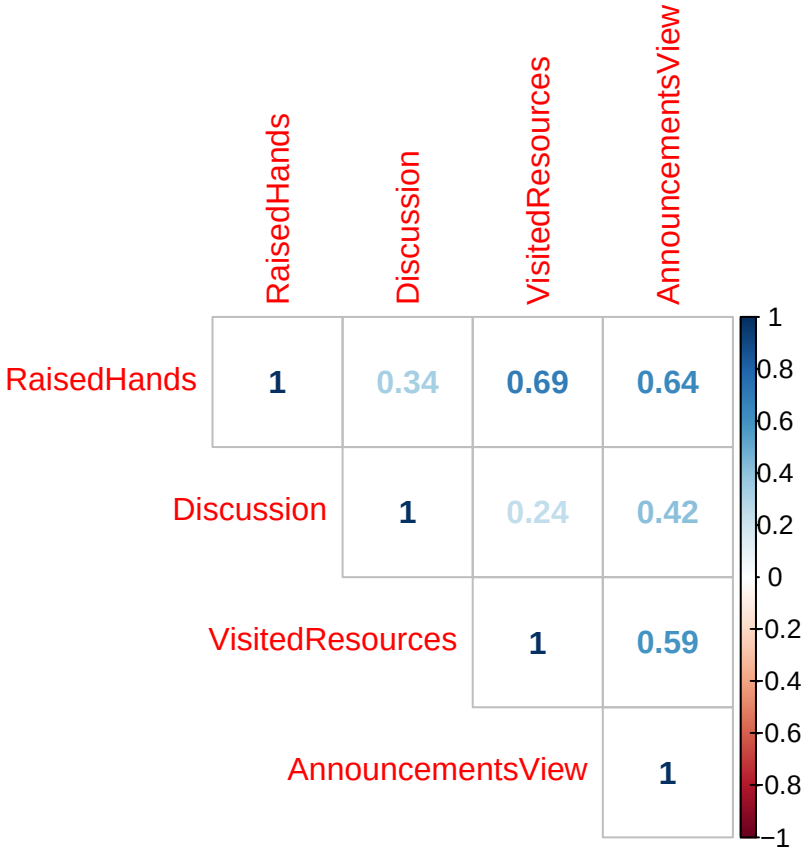
```
# Box plot for Announcement view and students grades
ggplot(data = edudata, aes(x = Class, fill = Class, y = AnnouncementsView)) + geom_boxplot() +
    labs(x = "Class", y = "Announcement View") +
    scale_y_continuous(breaks = seq(0,100,5))+ggtitle("Announcements Views per Class")+theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))+ylab("Announcement View")
```



High marks students visits more announcments views

Corelation Analysis

```
correlation <- cor(edudata[,c("RaisedHands","Discussion","VisitedResources","AnnouncementsView")])
corrplot(correlation,type="upper", method = "number")
```




```
cor(edudata[,c("RaisedHands","Discussion","VisitedResources","AnnouncementsView")])
```

```
##                   RaisedHands Discussion VisitedResources
## RaisedHands         1.0000000  0.3393860        0.6915717
## Discussion          0.3393860  1.0000000        0.2432918
## VisitedResources    0.6915717  0.2432918        1.0000000
## AnnouncementsView   0.6439178  0.4172900        0.5945000
##                   AnnouncementsView
## RaisedHands               0.6439178
## Discussion                0.4172900
## VisitedResources          0.5945000
## AnnouncementsView         1.0000000
```

Statistical Tests

```
attach(edudata)
```

Visited Resources Null hypothesis: There is no difference between mean visited resources hands between males and female. Alternative hypothesis: There is difference between mean visited resources between males and female.

```
# Performing T test to find the meann difference between teh visited resources online for males and fem
t.test(VisitedResources~Gender)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  VisitedResources by Gender
## t = 4.8533, df = 394.3, p-value = 1.753e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   8.615498 20.348436
## sample estimates:
## mean in group F mean in group M
##        64.00000        49.51803
```

Announcement Views Null hypothesis: There is no difference between announcement views between male and females. Alternative hypothesis: There is difference between announcement views between male and females.

```
# Performing T test for students viewing announcment with gender
t.test(AnnouncementsView~Gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  AnnouncementsView by Gender
## t = 1.1592, df = 379.79, p-value = 0.2471
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.004752  7.764190
## sample estimates:
## mean in group F mean in group M
##        39.74857        36.86885
```

Discussion Null hypothesis: There is no difference between announcement views between male and females. Alternative hypothesis: There is difference between announcement views between male and females.

```
t.test(Discussion~Gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  Discussion by Gender
## t = 2.6864, df = 338.33, p-value = 0.007579
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.91559 12.39073
## sample estimates:
## mean in group F mean in group M
##        47.82857        40.67541
```

One Way ANOVA

Null hypothesis: There is no difference among mean visited resources by Class. Alternative hypothesis: There is difference between at least two groups mean visited resources by Class.

```
fit <- aov(VisitedResources~Class)
summary(fit)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Class         2 257596  128798   230.5 <2e-16 ***
## Residuals   477 266568     559
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Null hypothesis: There is no difference among mean announcement views by Class. Alternative hypothesis: There is difference between at least two groups mean announcement views by Class.

```
fit <- aov(AnnouncementsView~Class)
summary(fit)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Class         2  99306   49653   98.72 <2e-16 ***
## Residuals   477 239902     503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Null hypothesis: There is no difference among mean discussionby Class. Alternative hypothesis: There is difference between at least two groups mean discussionby Class.

```
fit <- aov(Discussion~Class)
summary(fit)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Class         2  35031   17516   25.25 3.76e-11 ***
## Residuals   477 330850     694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multinomial Logistic Regression

```
# Multinomial Logistic Regression
model <- multinom(Class ~ RaisedHands + VisitedResources + AnnouncementsView + Discussion , data = eduda
```

```
## # weights:  18 (10 variable)
## initial  value 527.333899
## iter  10 value 396.503412
## iter  20 value 339.177653
## final  value 339.177644
## converged
```

```
summary(model)
```

```
## Call:
## multinom(formula = Class ~ RaisedHands + VisitedResources + AnnouncementsView +
##     Discussion, data = edudata)
##
## Coefficients:
##   (Intercept) RaisedHands VisitedResources AnnouncementsView Discussion
## M   -2.705545  0.02742201       0.03481683        0.02901283 0.01060729
## H   -6.537994  0.05085812       0.05853911        0.02656988 0.01947731
##
## Std. Errors:
##   (Intercept) RaisedHands VisitedResources AnnouncementsView  Discussion
## M   0.3750504 0.008492583      0.006578314       0.009821992 0.006244215
## H   0.6401695 0.009758013      0.008672398       0.011032798 0.007524808
##
```

```
## Residual Deviance: 678.3553
## AIC: 698.3553
```

```
# Finding P Value for the model
z <- summary(model)$coefficients/summary(model)$standard.errors
z
```

```
##   (Intercept) RaisedHands VisitedResources AnnouncementsView Discussion
## M   -7.213816    3.228936         5.292668          2.953864   1.698739
## H  -10.212911    5.211934         6.750048          2.408263   2.588413
```

```
#  Performing 2 Tailed Z Test
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
```

```
##    (Intercept)  RaisedHands VisitedResources AnnouncementsView  Discussion
## M 5.440093e-13 1.242518e-03     1.205449e-07       0.003138224 0.089368457
## H 0.000000e+00 1.868822e-07     1.477951e-11       0.016028613 0.009641927
```

```
# Extract the coefficients from the model and exponentiate
exp(coef(model))
```

```
##   (Intercept) RaisedHands VisitedResources AnnouncementsView Discussion
## M 0.066833891    1.027801         1.035430          1.029438   1.010664
## H 0.001447389    1.052174         1.060286          1.026926   1.019668
```

```
# Finding predicted Propbabilites
head(pp <- fitted(model))
```

```
##           L         M           H
## 1 0.8045369 0.1855441 0.009919038
## 2 0.7392168 0.2439696 0.016813530
## 3 0.8615327 0.1328638 0.005603509
## 4 0.5974364 0.3637632 0.038800383
## 5 0.2222089 0.6105213 0.167269791
## 6 0.3093185 0.5694851 0.121196329
```