Problem 2.8

Sai Ananthula - emw832

```
college <- read.csv("College.csv")
View(college)
head(college)
```

```
##                              X Private Apps Accept Enroll Top10perc Top25perc
## 1 Abilene Christian University     Yes 1660   1232    721        23        52
## 2           Adelphi University     Yes 2186   1924    512        16        29
## 3               Adrian College     Yes 1428   1097    336        22        50
## 4          Agnes Scott College     Yes  417    349    137        60        89
## 5    Alaska Pacific University     Yes  193    146     55        16        44
## 6            Albertson College     Yes  587    479    158        38        62
##   F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD Terminal
## 1        2885         537     7440       3300   450     2200  70       78
## 2        2683        1227    12280       6450   750     1500  29       30
## 3        1036          99    11250       3750   400     1165  53       66
## 4         510          63    12960       5450   450      875  92       97
## 5         249         869     7560       4120   800     1500  76       72
## 6         678          41    13500       3335   500      675  67       73
##   S.F.Ratio perc.alumni Expend Grad.Rate
## 1      18.1          12   7041        60
## 2      12.2          16  10527        56
## 3      12.9          30   8735        54
## 4       7.7          37  19016        59
## 5      11.9           2  10922        15
## 6       9.4          11   9727        55
```

```
rownames(college) <- college[, 1]
View(college)
```

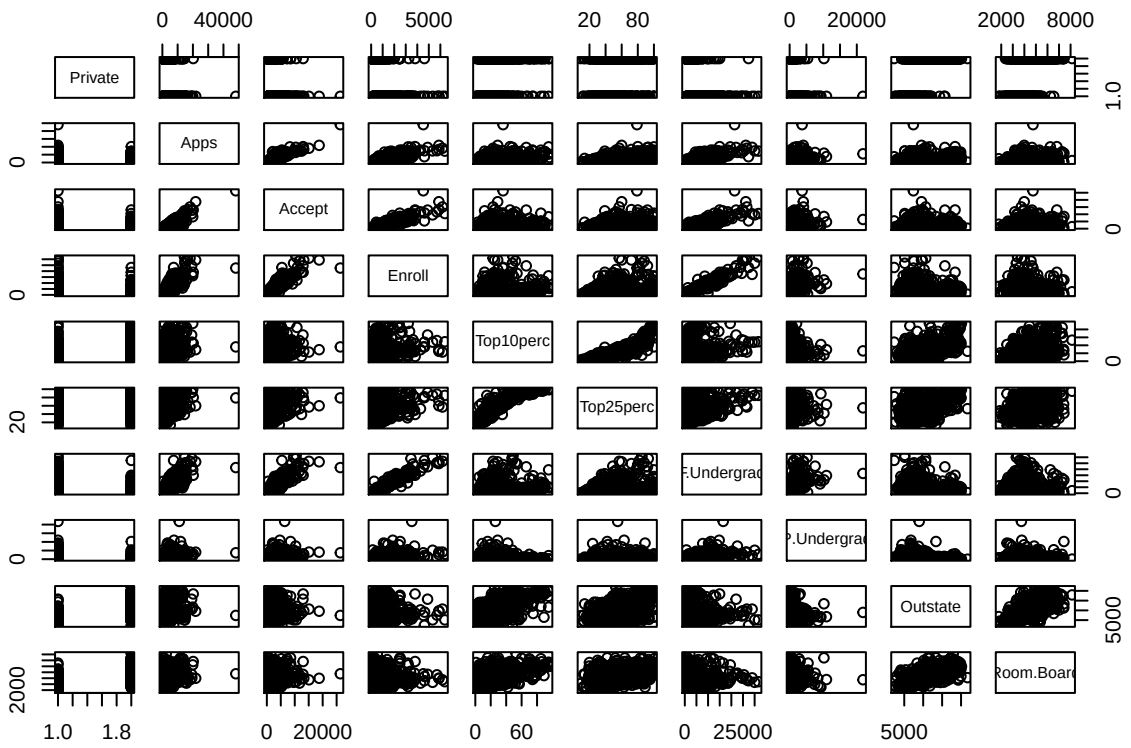
```
college <- college[, -1]
View(college)
```

```
summary(college)
```

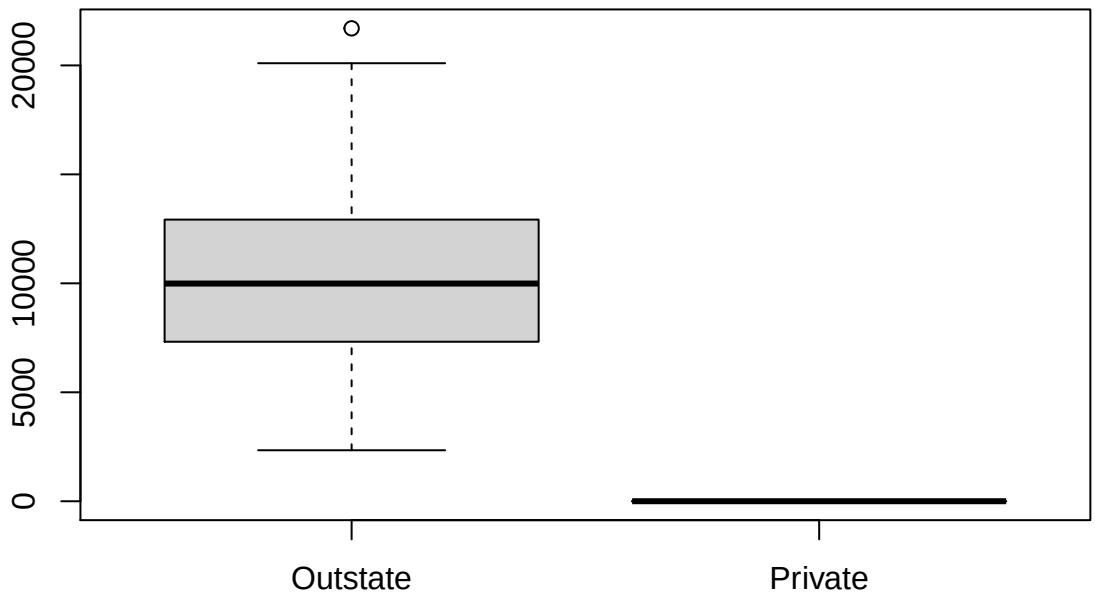
```
##    Private              Apps           Accept          Enroll
##  Length:777         Min.   :   81   Min.   :   72   Min.   :  35
##  Class :character   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242
##  Mode  :character   Median : 1558   Median : 1110   Median : 434
##                     Mean   : 3002   Mean   : 2019   Mean   : 780
##                     3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902
##                     Max.   :48094   Max.   :26330   Max.   :6392
##    Top10perc       Top25perc      F.Undergrad     P.Undergrad
```

```
##  Min.   : 1.00   Min.   :  9.0   Min.   :  139   Min.   :    1.0
##  1st Qu.:15.00   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0
##  Median :23.00   Median : 54.0   Median : 1707   Median :  353.0
##  Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   :  855.3
##  3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0
##  Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0
##     Outstate       Room.Board       Books           Personal
##  Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
##  1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850
##  Median : 9990   Median :4200   Median : 500.0   Median :1200
##  Mean   :10441   Mean   :4358   Mean   : 549.4   Mean   :1341
##  3rd Qu.:12925   3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700
##  Max.   :21700   Max.   :8124   Max.   :2340.0   Max.   :6800
##       PhD            Terminal       S.F.Ratio      perc.alumni
##  Min.   :  8.00   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00
##  1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00
##  Median : 75.00   Median : 82.0   Median :13.60   Median :21.00
##  Mean   : 72.66   Mean   : 79.7   Mean   :14.09   Mean   :22.74
##  3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
##  Max.   :103.00   Max.   :100.0   Max.   :39.80   Max.   :64.00
##      Expend        Grad.Rate
##  Min.   : 3186   Min.   : 10.00
##  1st Qu.: 6751   1st Qu.: 53.00
##  Median : 8377   Median : 65.00
##  Mean   : 9660   Mean   : 65.46
##  3rd Qu.:10830   3rd Qu.: 78.00
##  Max.   :56233   Max.   :118.00
```

```
college[,1] <- as.numeric(factor(college[,1]))
pairs(college[,1:10])
```




```
boxplot(college[,"Outstate"], college[,"Private"],
        names = c("Outstate", "Private"))
```

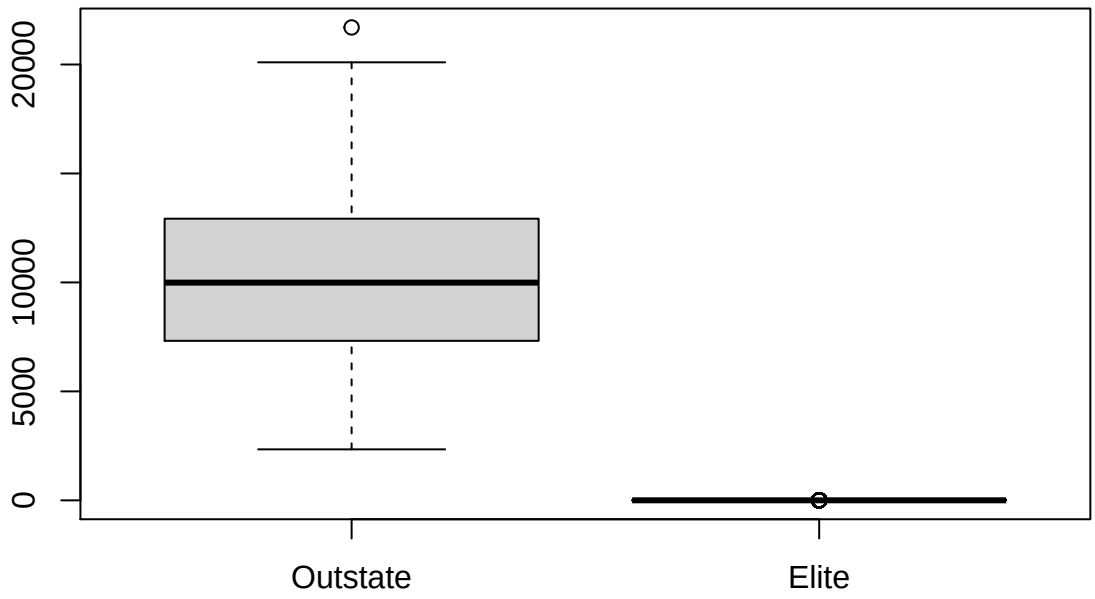



```
Elite <- rep ("No", nrow (college))
Elite[college$Top10perc > 50] <- " Yes "
Elite <- as.factor (Elite)
college <- data.frame (college , Elite)
summary(college)
```

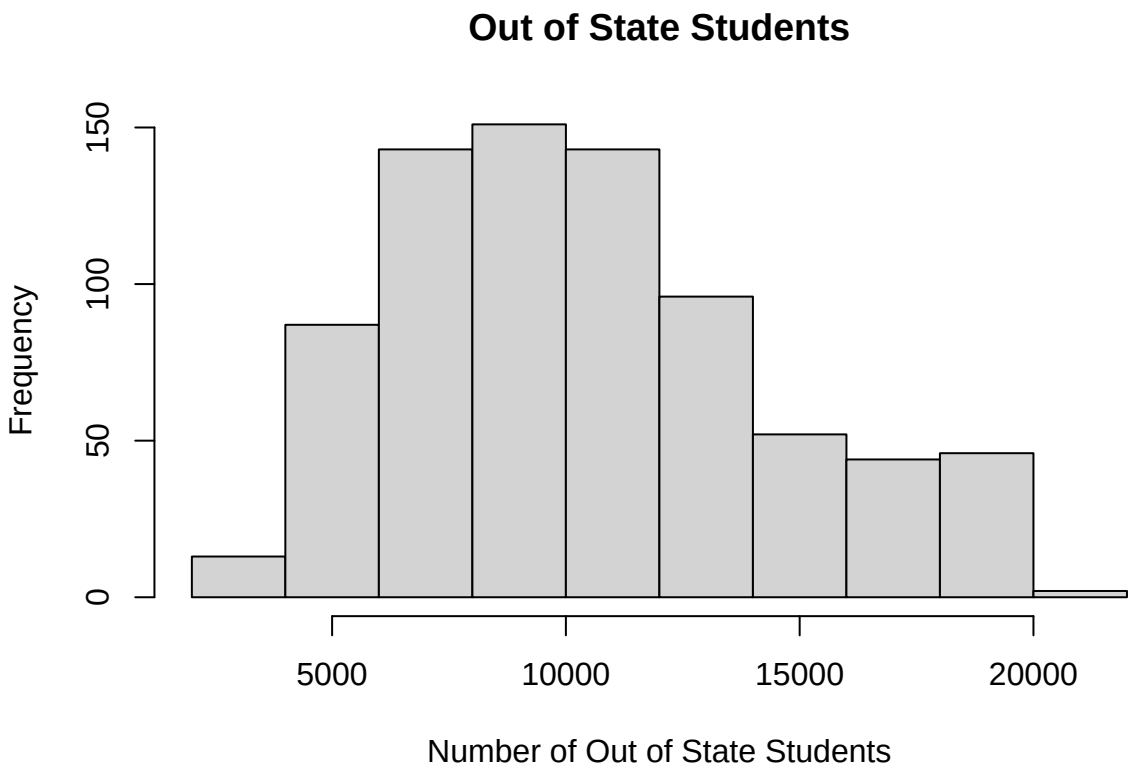
```
##     Private           Apps           Accept          Enroll       Top10perc
##  Min.   :1.000   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  1st Qu.:1.000   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##  Median :2.000   Median : 1558   Median : 1110   Median : 434   Median :23.00
##  Mean   :1.727   Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56
##  3rd Qu.:2.000   3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00
##  Max.   :2.000   Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books           Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
```

```
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##     Terminal       S.F.Ratio      perc.alumni        Expend
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
##    Grad.Rate        Elite
##  Min.   : 10.00    Yes : 78
##  1st Qu.: 53.00   No   :699
##  Median : 65.00
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

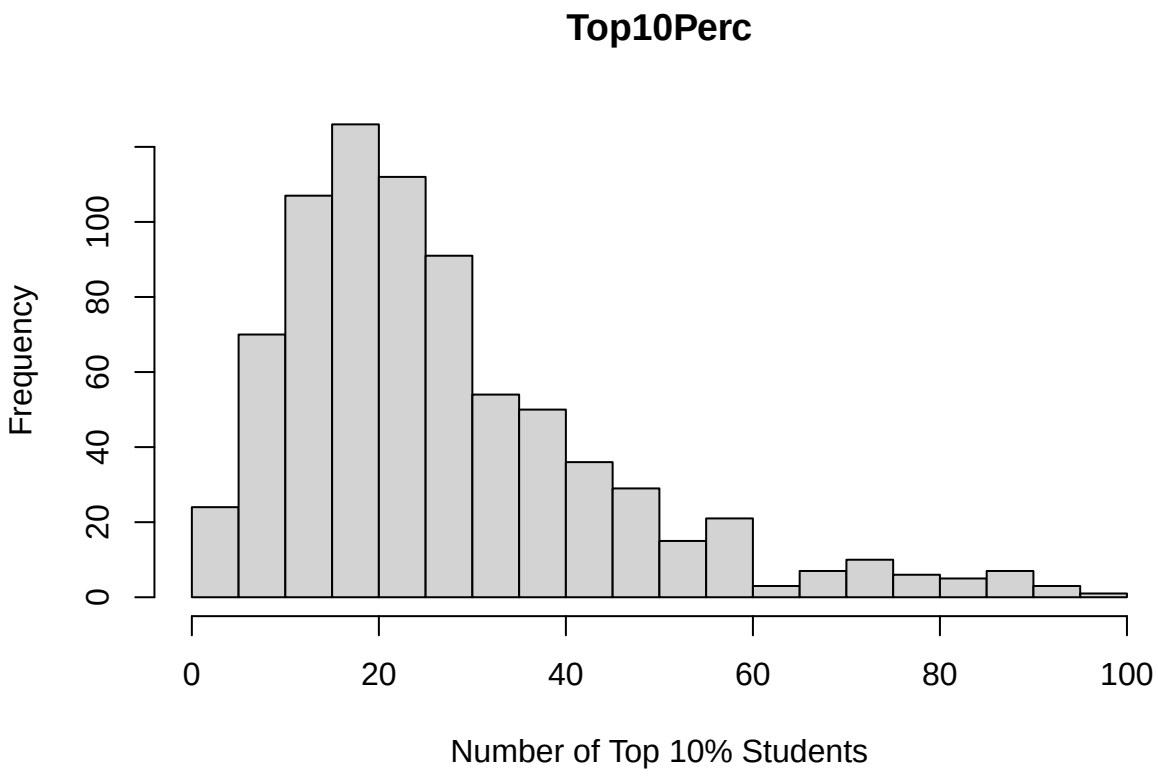
```
boxplot(college[,"Outstate"], college[,"Elite"],
        names = c("Outstate", "Elite"))
```



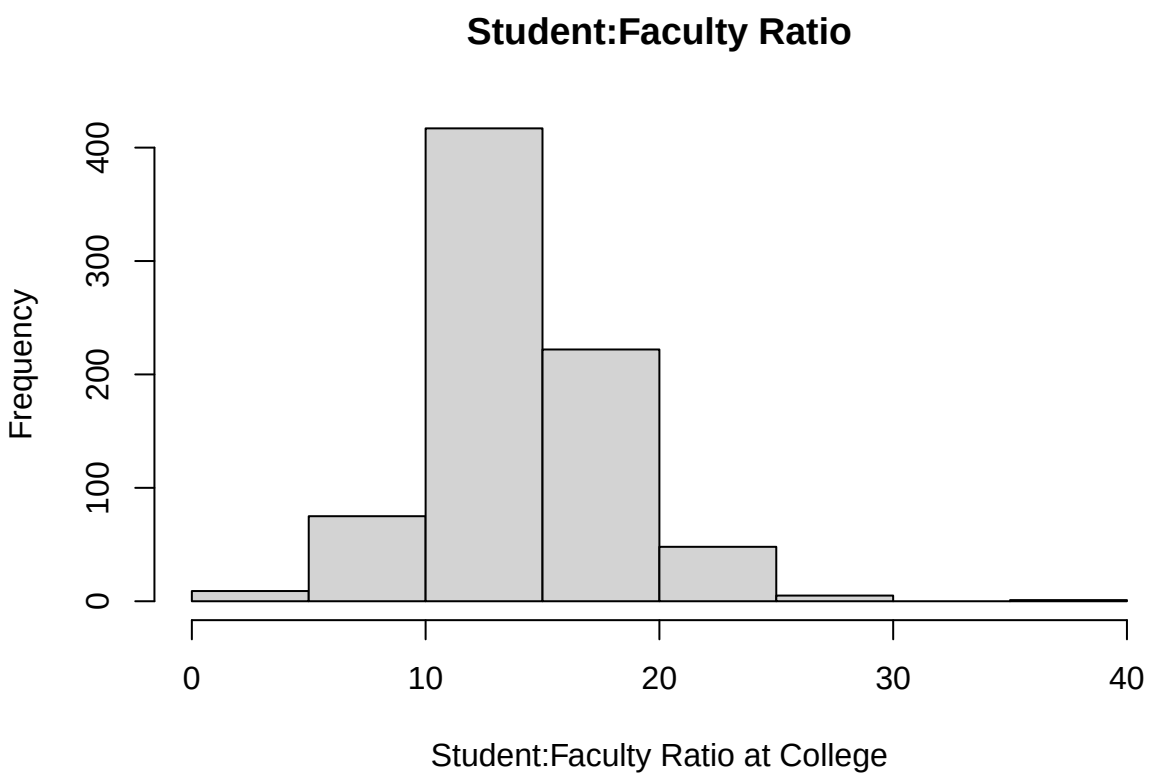

```
hist(college[,"Outstate"],main = "Out of State Students", xlab = "Number of Out of State Students")
```




```
hist(college[,"Top10perc"],main = "Top10Perc", xlab = "Number of Top 10% Students",breaks = 20)
```



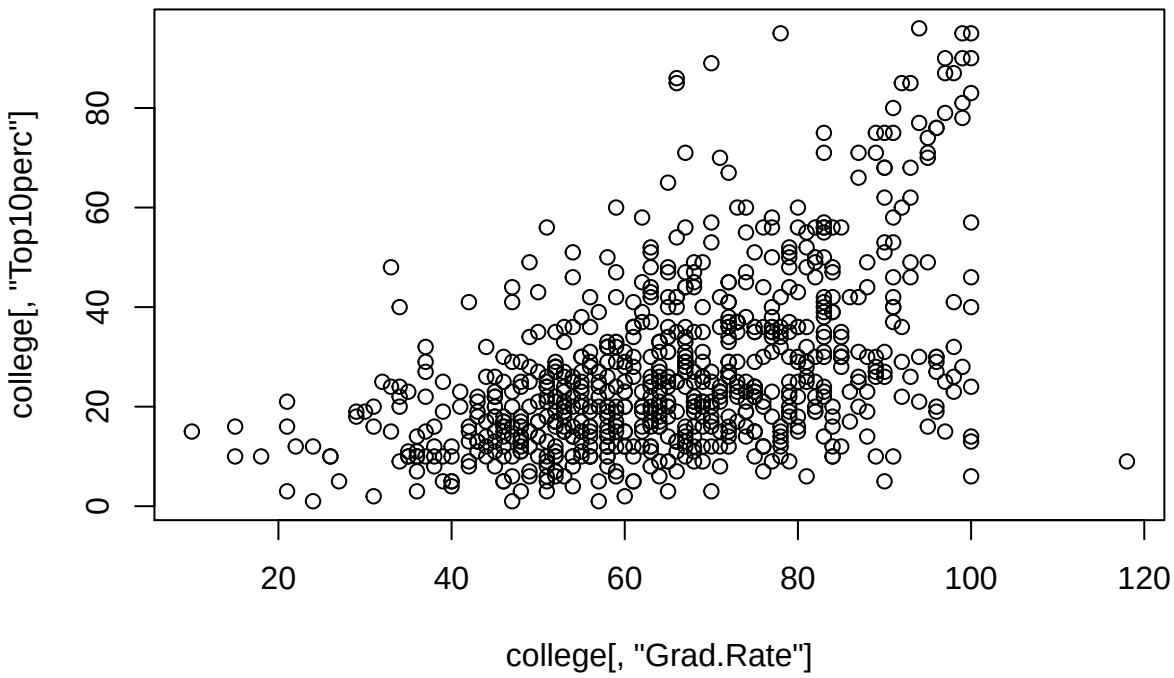

```
hist(college[,"S.F.Ratio"],main = "Student:Faculty Ratio", xlab = "Student:Faculty Ratio at College",
```




```
relation <- lm(college[,"Grad.Rate"]~college[,"Top10perc"])
summary(relation)
```

```
##
## Call:
## lm(formula = college[, "Grad.Rate"] ~ college[, "Top10perc"])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.410  -9.834   0.288   9.080  61.482
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             52.17990    0.99431   52.48   <2e-16 ***
## college[, "Top10perc"]  0.48201    0.03039   15.86   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.94 on 775 degrees of freedom
## Multiple R-squared:  0.245,  Adjusted R-squared:  0.244
## F-statistic: 251.5 on 1 and 775 DF,  p-value: < 2.2e-16
```

```
plot(college[,"Grad.Rate"],college[,"Top10perc"])
```



There seems to be a strong correlation between the number of 10 percent students who go to any given college and the graduation rate. This is justified by the p-value having 16 0's in front of it (calculated using the linear regression function). Using this information from the test data it would be rather easy to predict the graduation rate at a university given the number of students who attend said college that were in the top 10% of their high school's graduating class.