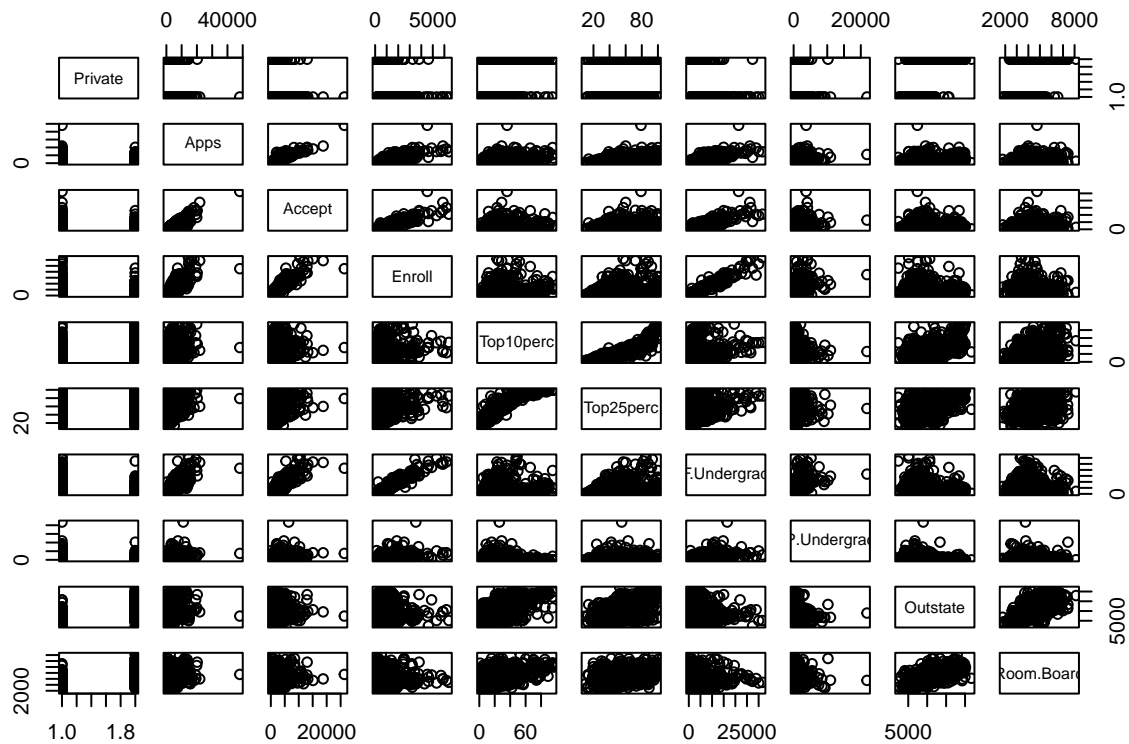


## Lab01 Brandon Habschied

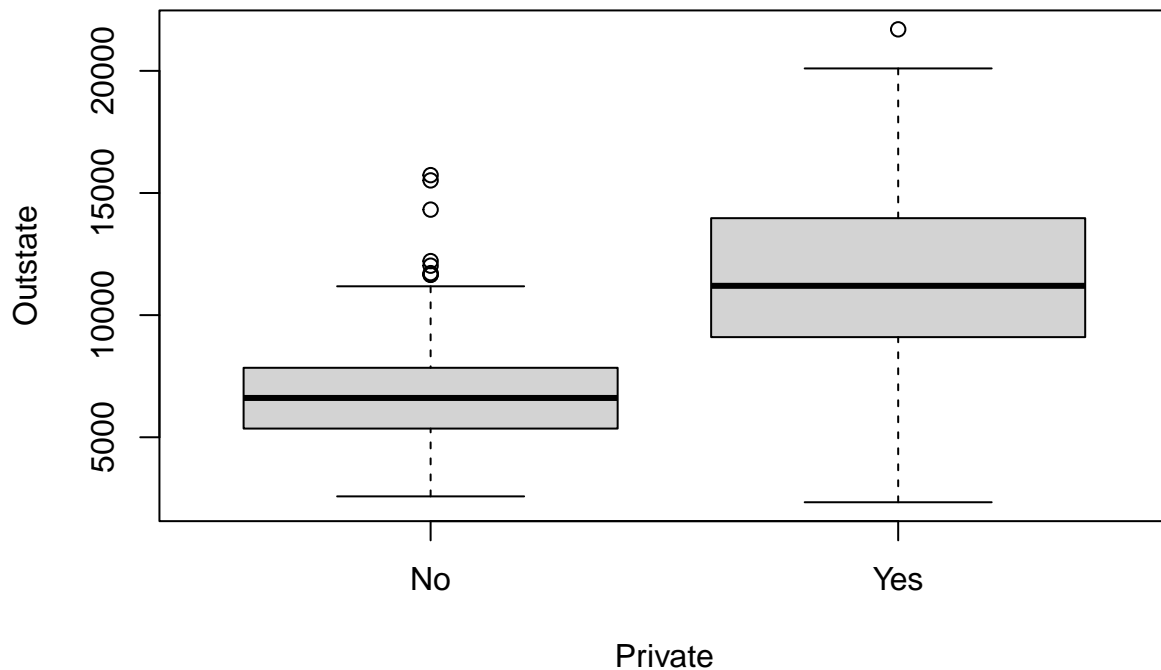
```
college <- read.csv("C:/Users/brand/OneDrive/Desktop/School/DataDrivenDiscovery/Lab1/College.csv")
fix(college)
rownames(college) = college[,1]
fix(college)
college = college[,-1]
fix(college)
summary(college)
```

```
##      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$Private = as.factor(college$Private)
pairs(college[,1:10])
```



```
plot(college$Private, college$Outstate, xlab="Private", ylab="Outstate")
```

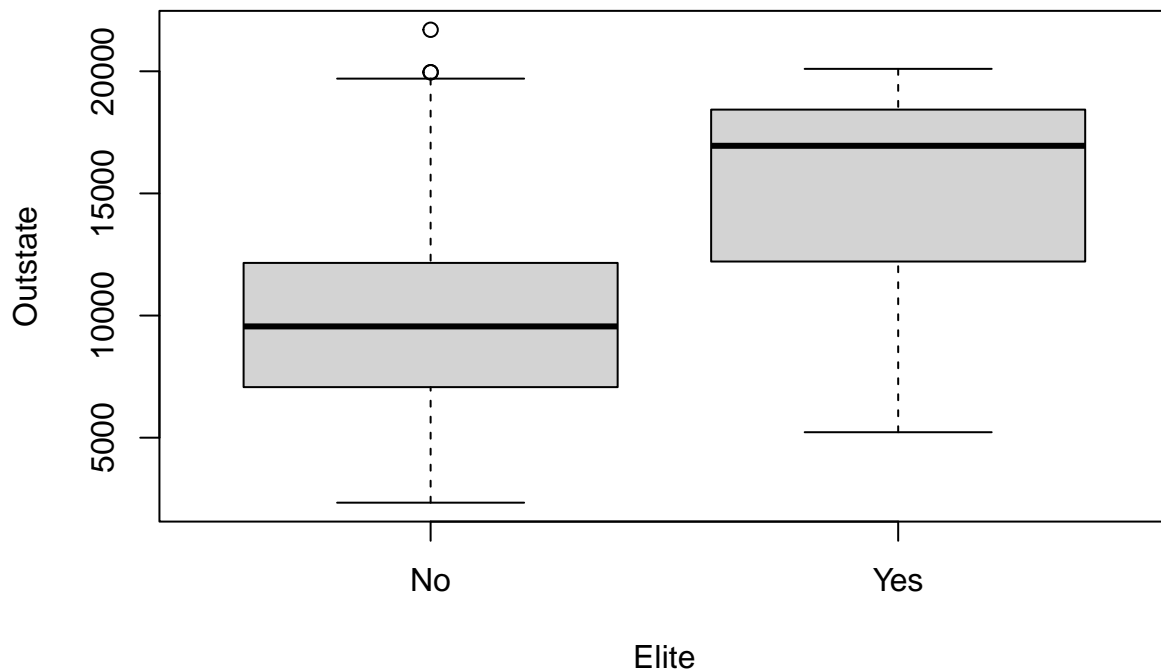


```
Elite = rep("No",nrow(college))
Elite[college$Top10perc>50]="Yes"
Elite=as.factor(Elite)
college=data.frame(college,Elite)
summary(college)
```

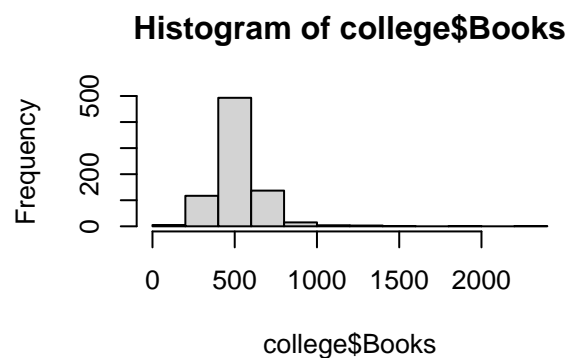
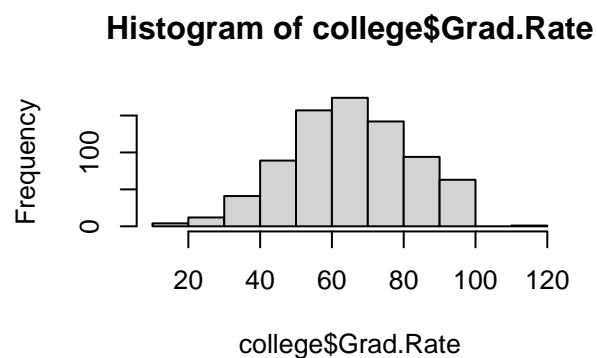
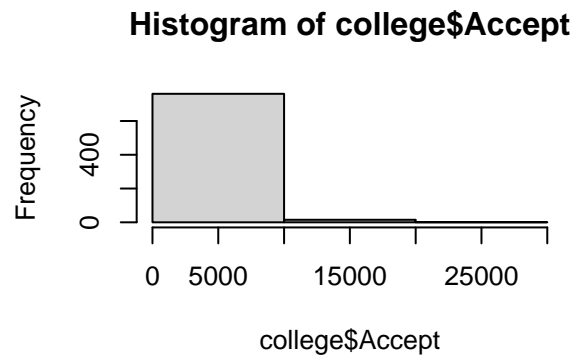
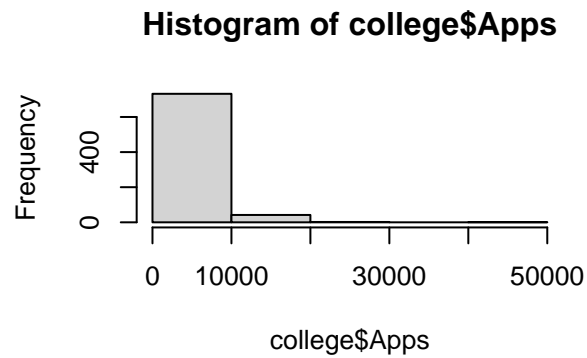
```
## Private      Apps      Accept      Enroll      Top10perc
## No :212  Min.   : 81   Min.   : 72   Min.   : 35   Min.   : 1.00
## Yes:565  1st Qu.: 776  1st Qu.: 604  1st Qu.: 242  1st Qu.:15.00
##          Median : 1558 Median : 1110 Median : 434 Median :23.00
##          Mean   : 3002 Mean   : 2019 Mean   : 780 Mean  :27.56
##          3rd Qu.: 3624 3rd Qu.: 2424 3rd Qu.: 902 3rd Qu.:35.00
##          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    No :699
## 1st Qu.: 53.00    Yes: 78
## Median : 65.00
## Mean   : 65.46
## 3rd Qu.: 78.00
## Max.   :118.00
```

```
plot(college$Elite, college$Outstate, xlab="Elite", ylab="Outstate")
```



```
par(mfrow=c(2,2))
hist(college$Apps,breaks = 5)
hist(college$Accept, breaks = 3)
hist(college$Grad.Rate, breaks = 10)
hist(college$Books, breaks = 10)
```



```
par(mfrow=c(1,1))
print(sd(college$Enroll))
```

```
## [1] 929.1762
```

```
# D.i) The predictors Private and Elite are qualitative. ALL
# the other predictors are Quantitative
print("D.i) The predictors Private and Elite are qualitative.
      ALL the other predictors are Quantitative")
```

```
## [1] "D.i) The predictors Private and Elite are qualitative. \n      ALL the other predictors are Quantitative"
```

```
# D.ii) The enroll predictor has a range of 35-6392. Its mean is 780
# and its standard deviation is 929.1762
cat("D.ii) The enroll predictor has a range of", min(college$Enroll), "-",
    max(college$Enroll), "a mean of", mean(college$Enroll), "and a std dev of",
    sd(college$Enroll))
```

```
## D.ii) The enroll predictor has a range of 35 - 6392 a mean of 779.973 and a std dev of 929.1762
```

```
# D.iii) The new range for enroll is 35-6392. Its new mean is 823.1
# and its new standard deviation is 972.2258
```

```
college = college[-(100:200),]  
cat("D.iii) The enroll predictor without 100-200 has a range of",  
    min(college$Enroll), "-", max(college$Enroll), "a mean of \n",  
    mean(college$Enroll), "and a std dev of", sd(college$Enroll))
```

```
## D.iii) The enroll predictor without 100-200 has a range of 35 - 6392 a mean of  
## 823.0695 and a std dev of 972.2258
```