Problem 1:

(a) better (b) worse (c) better (d) worse

Problem 2:

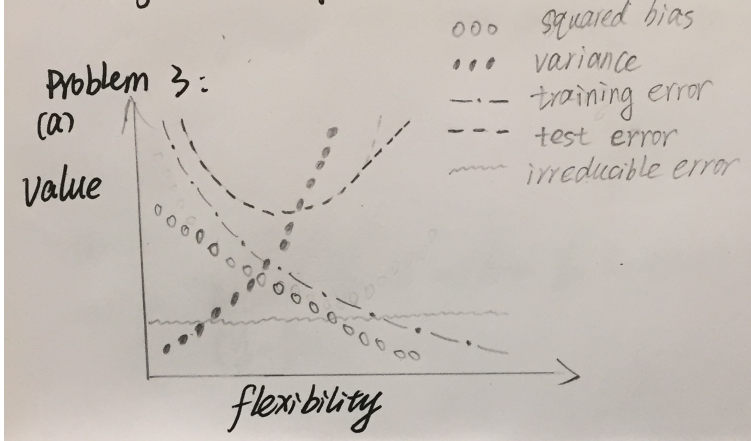
(a) regression, inference, $n=500$, $p=3$

(b) classification, prediction, $n=20$, $p=13$

(c) regression, inference, $n=52$, $p=3$

Problem 3:

(a)



(b) squared bias: This will keep decreasing because as the model becomes more flexible, it can fit the training data better, hence lower squared bias.

variance: With more flexibility, the model tends to be more unstable, hence a higher variance.

training error: With more flexibility, the model can always fit the training data better, eventually leads to even 0 traing error.

test error: Firstly, with more flexibility, the model can get more trend of the data and such ~~trend~~ trend generalizes well to the test data, hence a lower test error. But later on with even more flexibility, the model is fitting noise in the training data and such noise does not generalize well to the test data, therefore the test error goes up again.

irreducible error: The is the part of error that can never be reduced and it will remain constant as the variance of ε.

Problem 4:

(a) No. Because we do not know y_0. So there's no way for us to compute $(y_0 - \hat{f}(x_0))^2$

(b) No. bias = $(E\hat{f}(x_0) - f(x_0))^2$. Though we can estimate $E\hat{f}(x_0)$, we don't know f, therefore we can not estimate bias.

(C). Yes. Variance = $E(\hat{f}(x_0) - E\hat{f}(x_0))^2$. we can sample multiple time to get multiple $\hat{f}$, then we average all $\hat{f}(x_0)$ to estimate $E\hat{f}(x_0)$. then we calculate all of the $(\hat{f}(x_0) - E\hat{f}(x_0))^2$ (with $E\hat{f}(x_0)$ replaced by its estimator) and average them, this gives us a estimator of variance.

(b) No. Because MSE = bias + Variance + irreducible error. We can only estimate Variance. The bias and irreducible error sum up and we can not tell how much each part contribute to the total MSE (Not to mention that we cannot estimate MSE)

Problem 5: a)

```
college=read.csv("College.csv")
```

b)

```
rownames(college)=college[,1]
fix(college)

college=college[,-1]
fix(college)
```

c)

i.

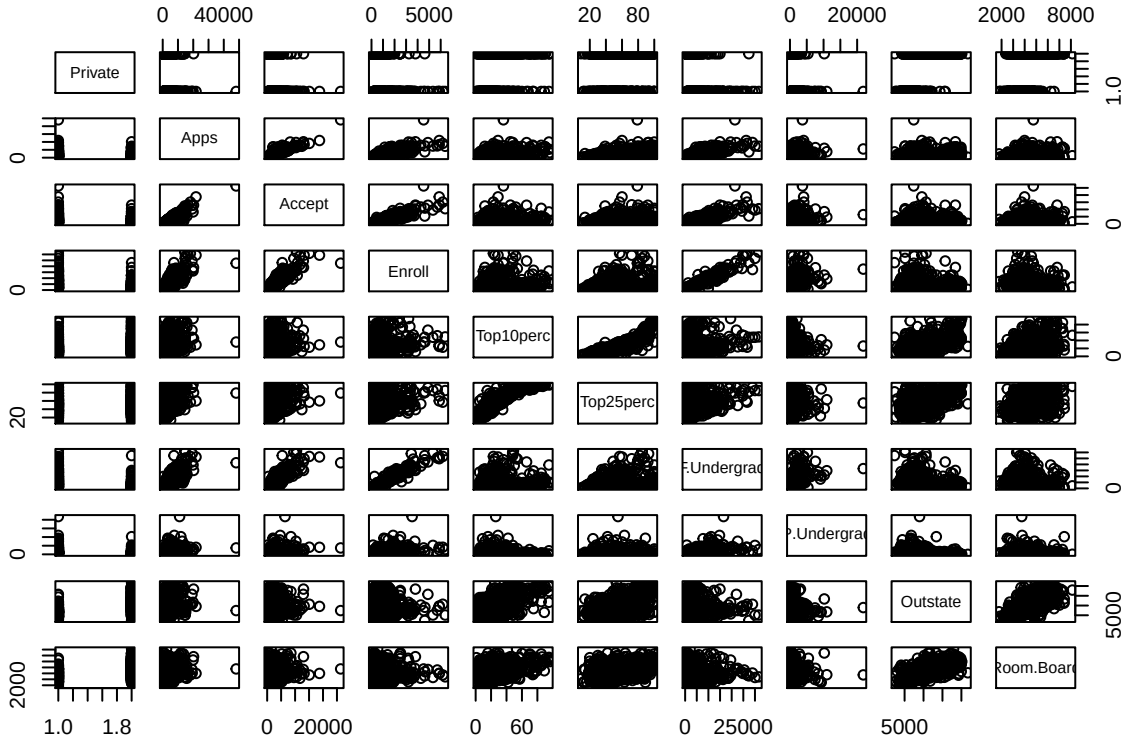
```
summary(college)
```

```
##  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
##  Min.   : 10.00
##  1st Qu.: 53.00
##  Median : 65.00
```

```
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

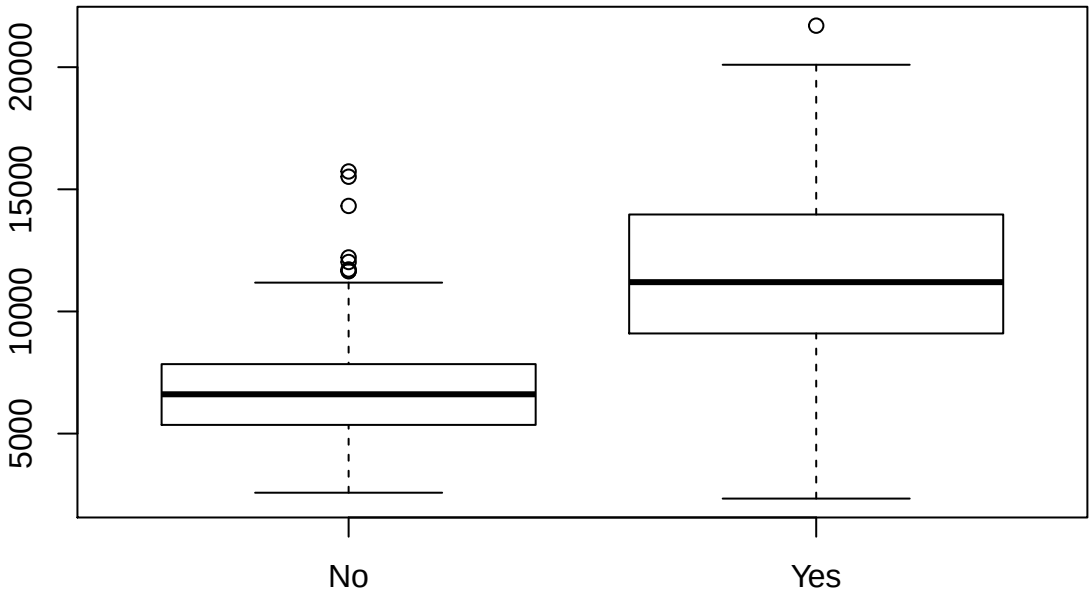
ii.

```
pairs(college[,1:10])
```



iii.

```
boxplot(Outstate~Private,data=college)
```

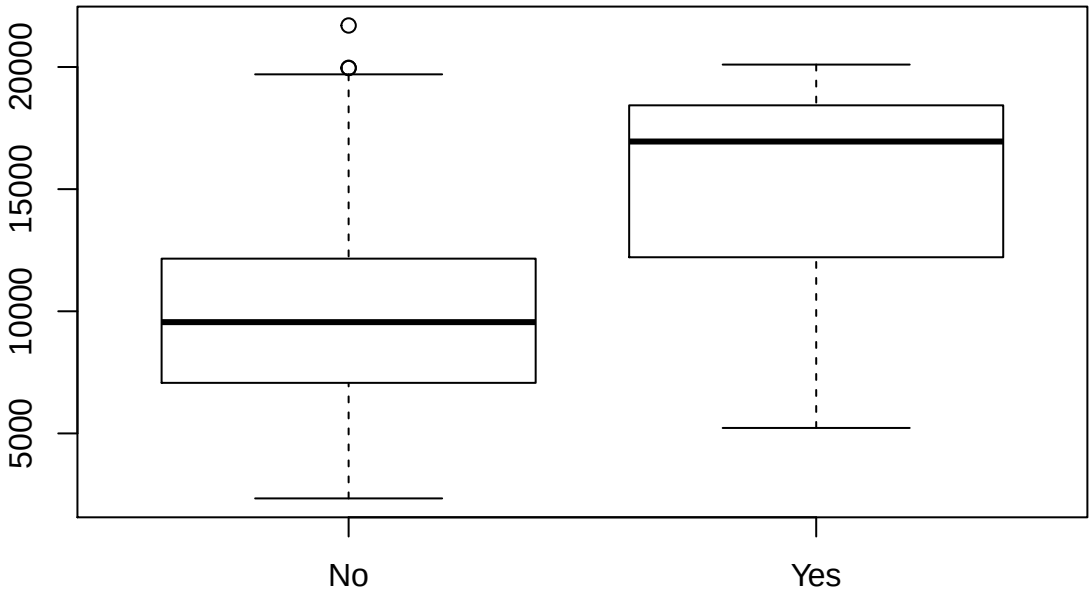


iv.

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

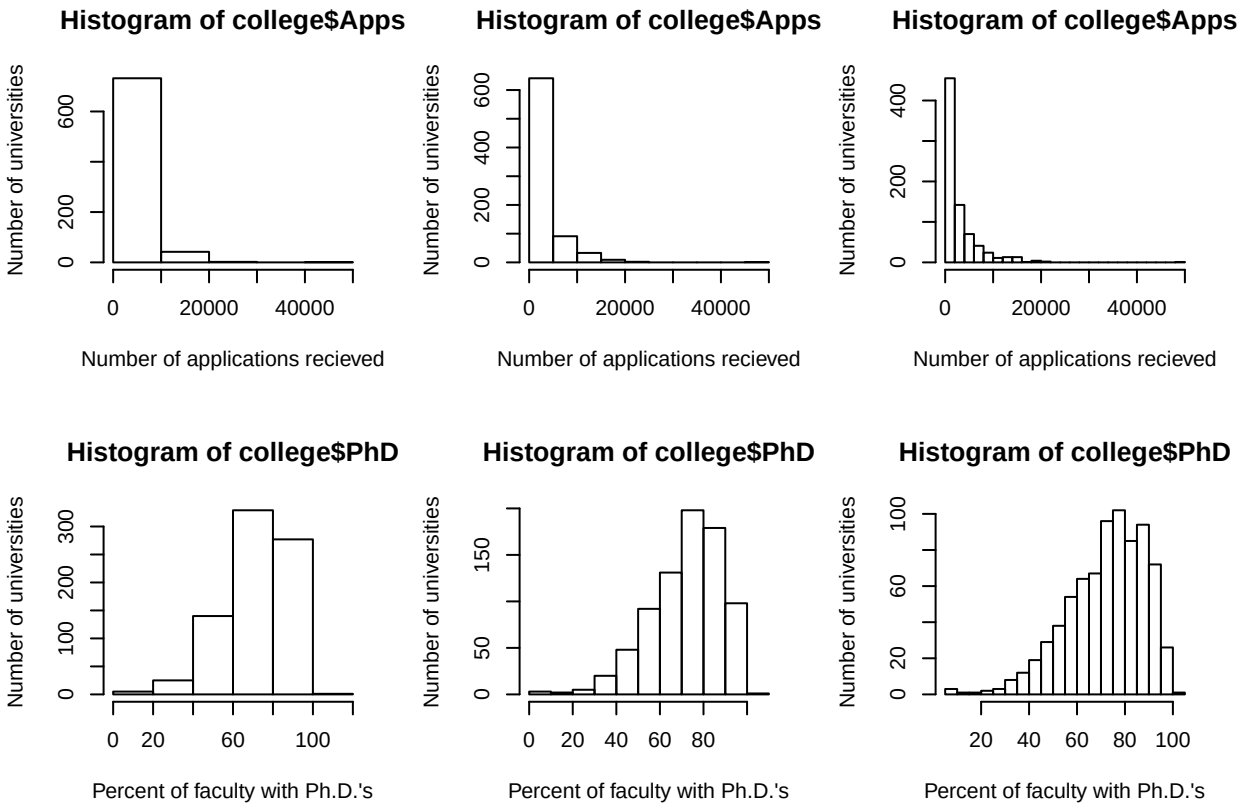
```
##  No Yes
## 699  78
```

```
boxplot(Outstate~Elite,data=college)
```



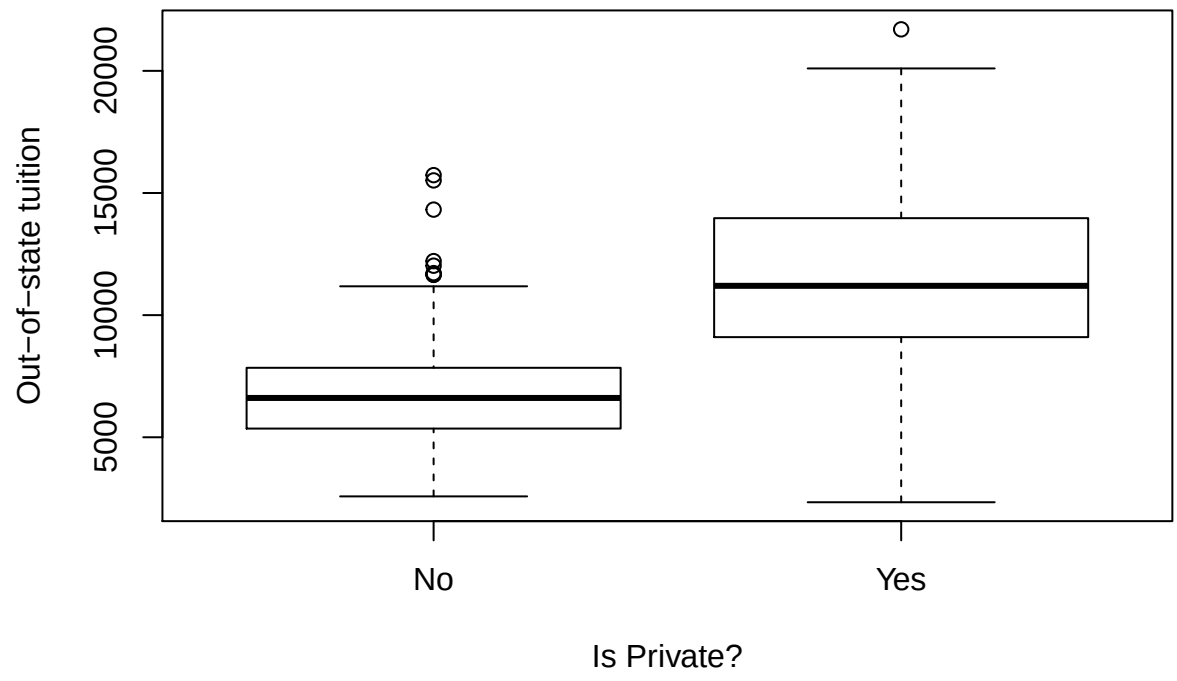
v.

```
par(mfrow=c(2,3))
hist(college$Apps,breaks=5,xlab='Number of applications recieved',ylab='Number of universities')
hist(college$Apps,breaks=10,xlab='Number of applications recieved',ylab='Number of universities')
hist(college$Apps,breaks=20,xlab='Number of applications recieved',ylab='Number of universities')
hist(college$PhD,breaks=5,xlab='Percent of faculty with Ph.D.\'s',ylab='Number of universities')
hist(college$PhD,breaks=10,xlab='Percent of faculty with Ph.D.\'s',ylab='Number of universities')
hist(college$PhD,breaks=20,xlab='Percent of faculty with Ph.D.\'s',ylab='Number of universities')
```

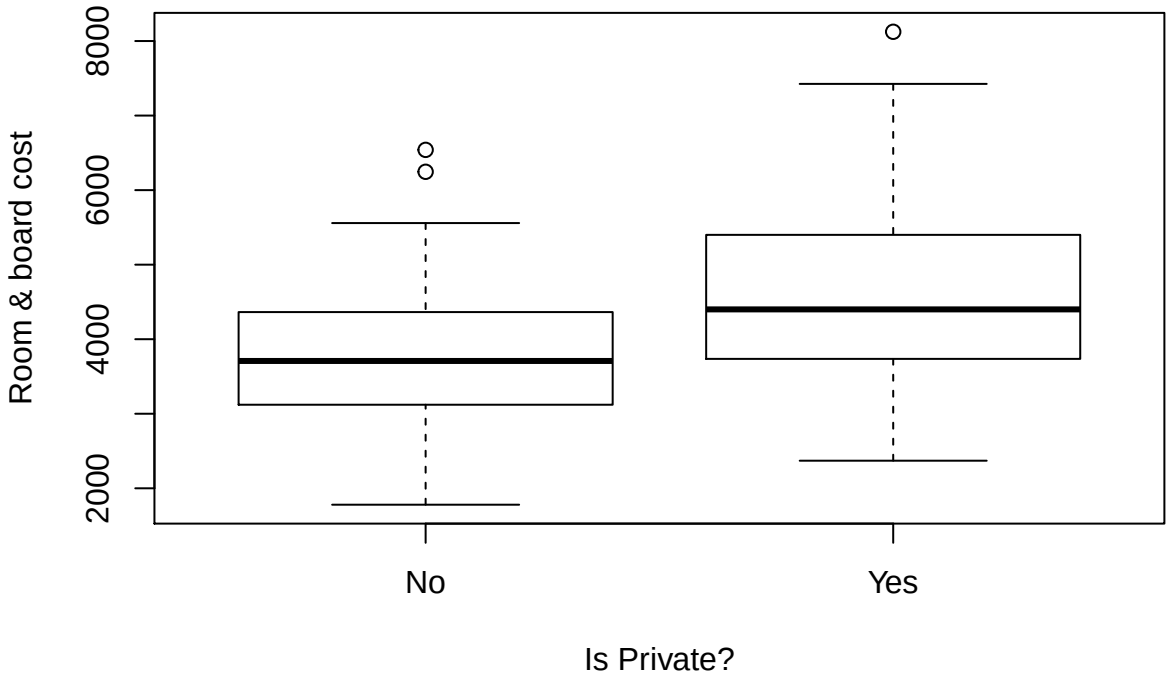


vi. The following 3 boxplots show that private university is generally more expensive to attend(high out-of-state tuition and room & board cost). And at the same time, students in private universities are spending less on there personal spending, probably because that the high expense of attending school has already shrinked their wallet.

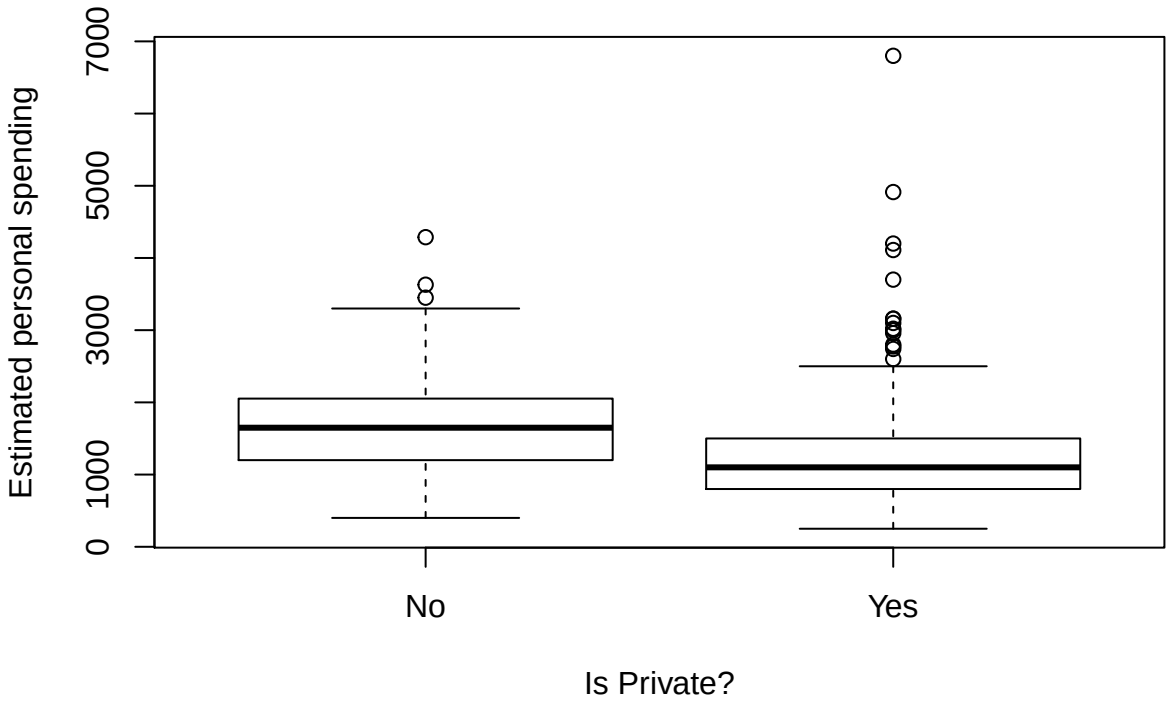
```
boxplot(Outstate~Private,data=college,xlab="Is Private?",ylab="Out-of-state tuition")
```



```r
boxplot(Room.Board~Private,data=college,xlab="Is Private?",ylab="Room & board cost")
```

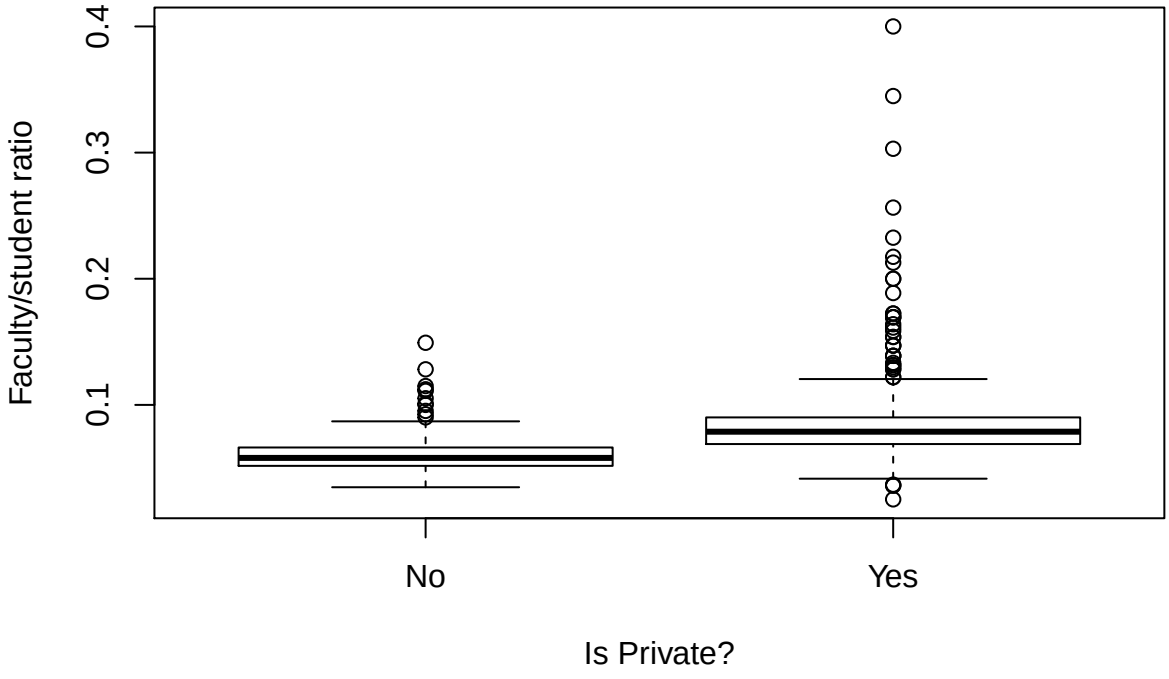



```r
boxplot(Personal~Private,data=college,xlab="Is Private?",ylab="Estimated personal spending")
```



The following 2 boxplots show that private universities are indeed put into good use of their higher tuition to provide the students with better learning environment, including a higher faculty/student ratio and a higher instructional expenditure per student. Therefore, my advice is, if you can afford a private university, then go to one. Because it's probably worth the money.

```
boxplot(1/S.F.Ratio~Private,data=college,xlab="Is Private?",ylab="Faculty/student ratio")
```




```
boxplot(Expend~Private,data=college,xlab="Is Private?",ylab="Instructional expenditure per student")
```