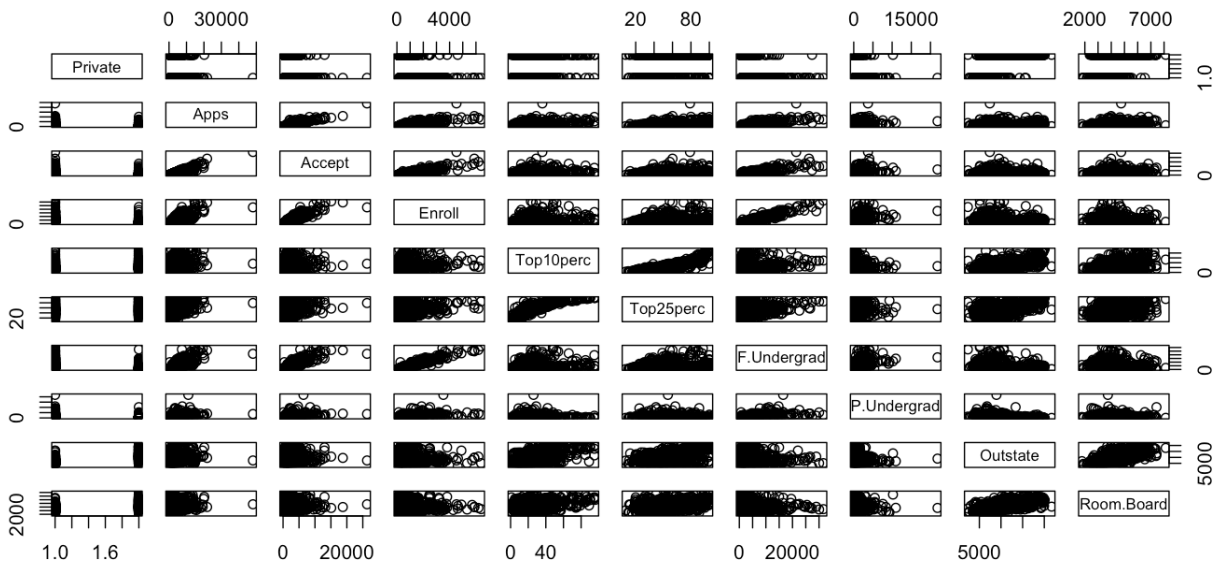
8.a
college<-read.csv("College.csv",header=T,na.strings =“?”)

b.
rownames(college)<-college[,1]
fix(college)
college=college[,-1]
fix(college)
head(college)

8.i

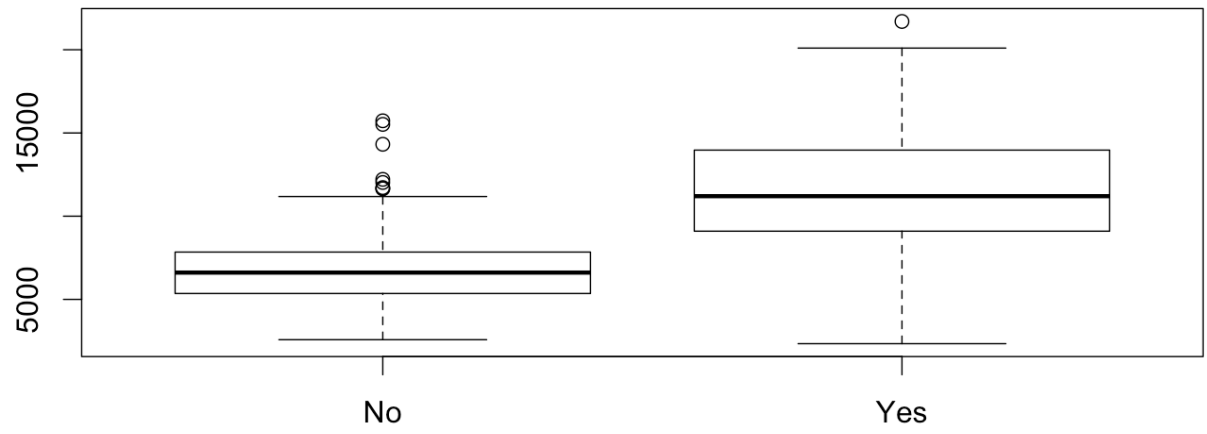
```
> summary(college)
 Private        Apps           Accept          Enroll       Top10perc       Top25perc
 No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00   Min.   :  9.0
 Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00   1st Qu.: 41.0
           Median : 1558   Median : 1110   Median : 434   Median :23.00   Median : 54.0
           Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56   Mean   : 55.8
           3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00   3rd Qu.: 69.0
           Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00   Max.   :100.0
  F.Undergrad     P.Undergrad         Outstate       Room.Board       Books
 Min.   :  139   Min.   :    1.0   Min.   : 2340   Min.   :1780   Min.   :  96.0
 1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0
 Median : 1707   Median :  353.0   Median : 9990   Median :4200   Median : 500.0
 Mean   : 3700   Mean   :  855.3   Mean   :10441   Mean   :4358   Mean   : 549.4
 3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925   3rd Qu.:5050   3rd Qu.: 600.0
 Max.   :31643   Max.   :21836.0   Max.   :21700   Max.   :8124   Max.   :2340.0
    Personal         PhD            Terminal       S.F.Ratio      perc.alumni
 Min.   : 250   Min.   :  8.00   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00
 1st Qu.: 850   1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00
 Median :1200   Median : 75.00   Median : 82.0   Median :13.60   Median :21.00
 Mean   :1341   Mean   : 72.66   Mean   : 79.7   Mean   :14.09   Mean   :22.74
 3rd Qu.:1700   3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
 Max.   :6800   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
```

ii.

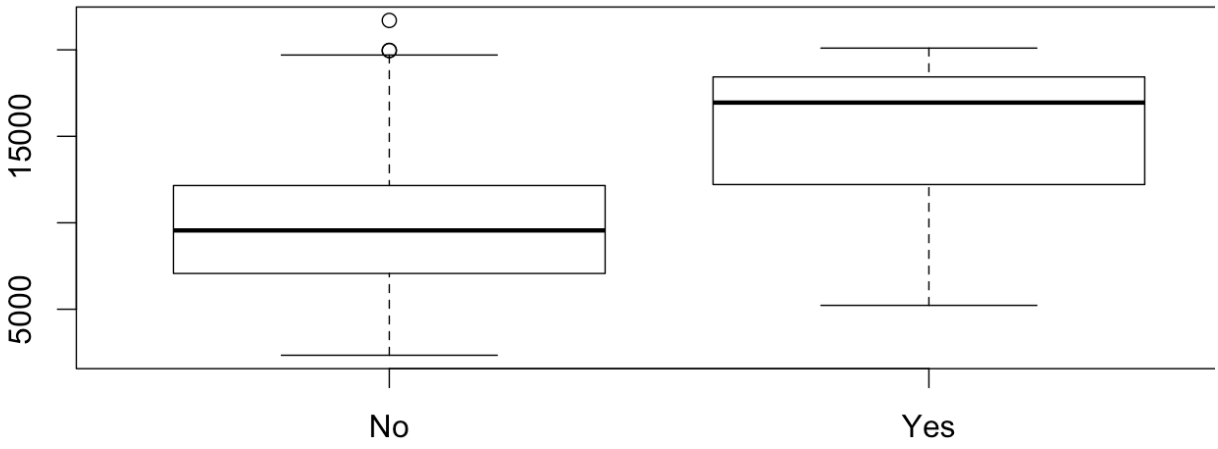


iii.

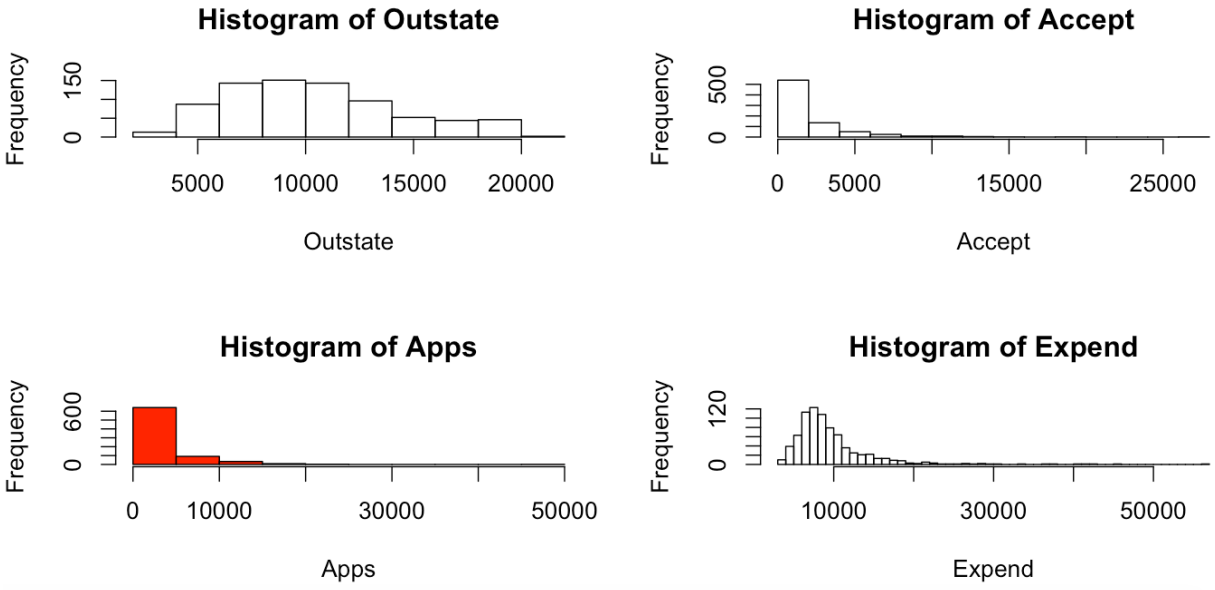
plot(college$Private, college$Outstate)



iv.
summary(college$Elite)
plot(college$Elite, college$Outstate)
There are 78 yes.



v.
par(mfrow=c(2,2))
hist(Expend,breaks=50)
hist(Apps,col=2)
hist(Accept)
hist(Outstate)



vi.
mean(Accept/Apps)
The Accept rate is 0.74.
mean(Enroll/Accept)
The Enroll rate is 0.41.
plot(Top25perc,Grad.Rate)
The rate of Top25perc have a little influence of Grad.Rage.

9.
Auto=read.csv("Auto.csv",header=T,na.strings ="?")
fix(Auto)
dim(Auto)
Auto=na.omit(Auto)

a.
quantitative: mpg, cylinders, displacement, horsepower, weight,acceleration, year,origin
qualitative: name

b.
> sapply(Auto[, 1:8], range)

	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
[1,]	9.0	3	68	46	1613	8.0	70	1
[2,]	46.6	8	455	230	5140	24.8	82	3

c.
> sapply(Auto[, 1:8], mean)

mpg	cylinders	displacement	horsepower	weight	acceleration
23.45	5.47	194.41	104.47	2977.58	15.54
year	origin				
75.98	1.58				

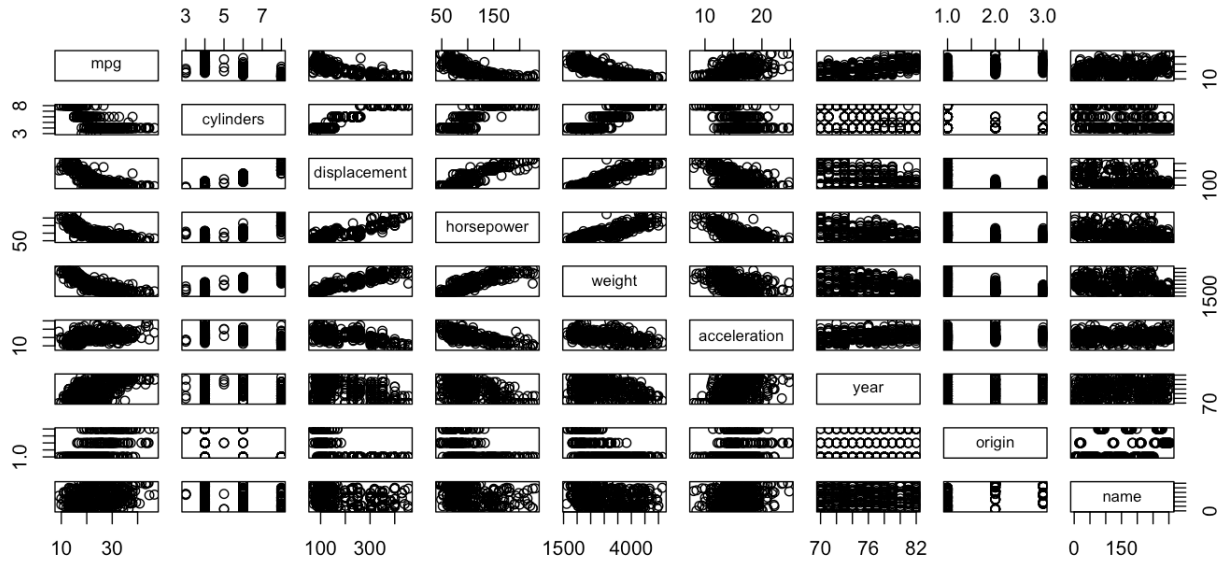
> sapply(Auto[, 1:8], sd)

mpg	cylinders	displacement	horsepower	weight	acceleration
7.8050075	1.7057832	104.6440039	38.4911599	849.4025600	2.7588641
year	origin				
3.6837365	0.8055182				

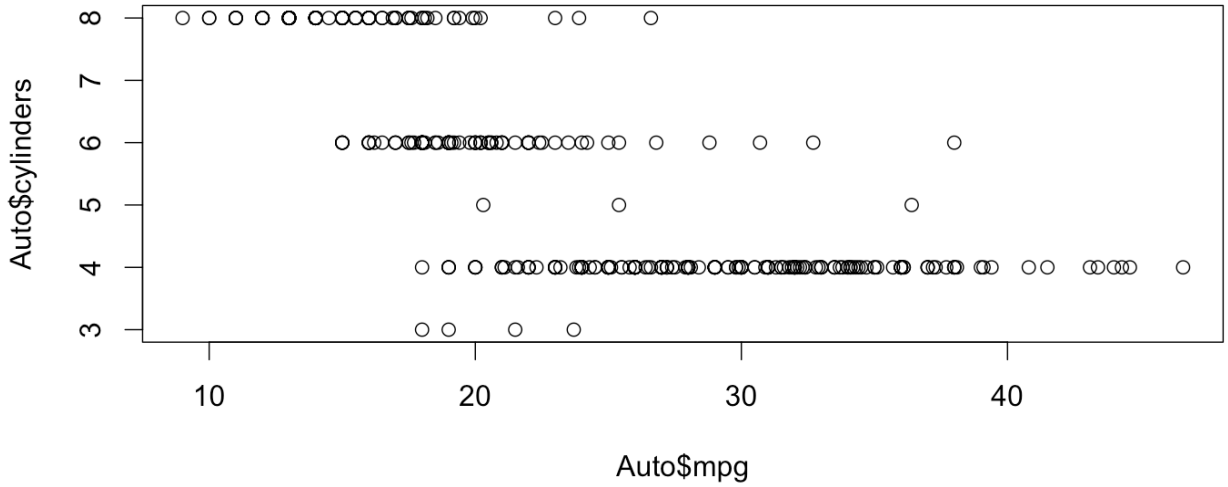
d.
Auto2<-Auto[-(10:85),]

```
> sapply(Auto2[, 1:8], range)
      mpg cylinders displacement horsepower weight acceleration year origin
[1,] 11.0         3           68         46   1649          8.5   70      1
[2,] 46.6         8          455        230   4997         24.8   82      3
> sapply(Auto2[, 1:8], mean)
         mpg    cylinders displacement   horsepower       weight acceleration
   24.404430     5.373418   187.240506   100.721519  2935.971519    15.726899
        year       origin
   77.145570     1.601266
> sapply(Auto2[, 1:8], sd)
         mpg    cylinders displacement   horsepower       weight acceleration
    7.867283     1.654179    99.678367    35.708853   811.300208     2.693721
        year       origin
    3.106217     0.819910
```

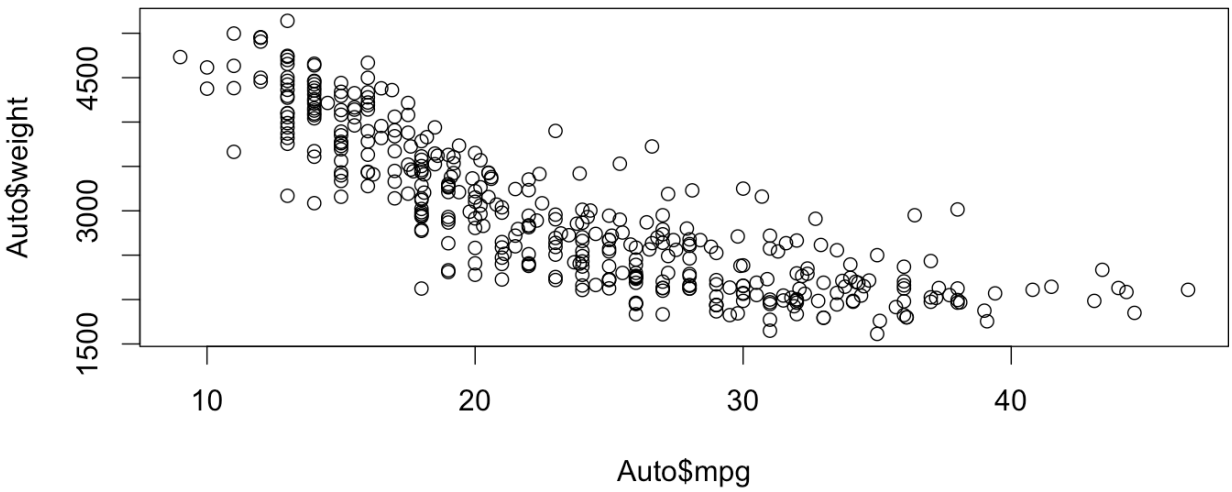
e.
pairs(Auto)



plot(Auto$mpg,Auto$cylinders): more mpg, less cylinders



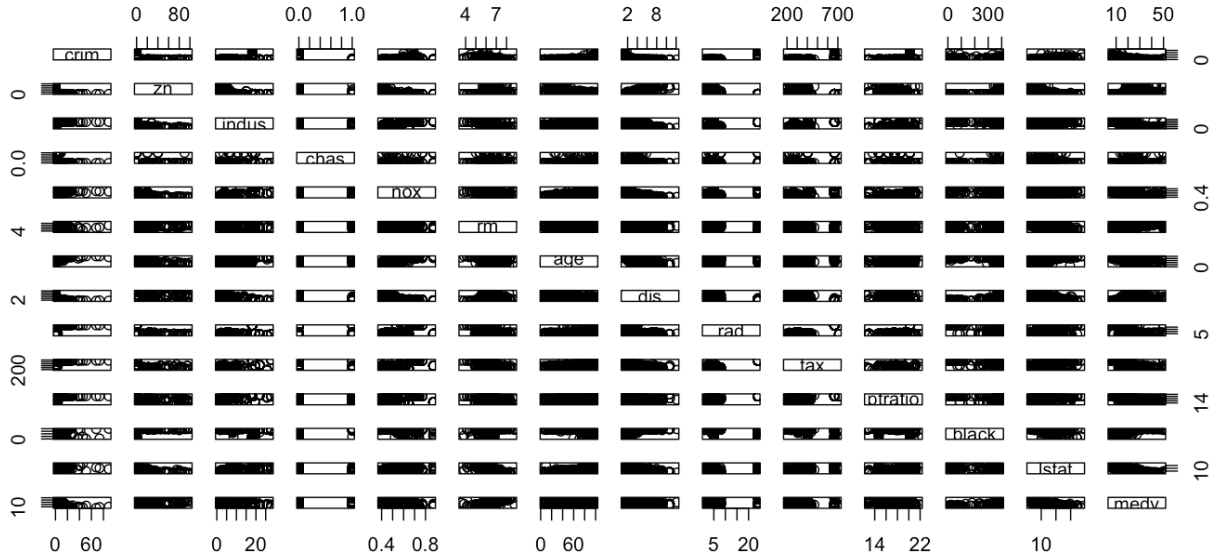
plot(Auto$mpg, Auto$weight): more mpg, less weight



f.
From plot e we can see that all predictors can infect mpg except name. So name is useless as a predictor.

10.a.
Boston data set has 506 rows and 14 columns. The rows represent 506 Housing Values in Suburbs of Boston and the columns represent 14 features.

b.
pairs(Boston)



With age become older, crim increase.
With dis become larger, crim decrease.
There are no infect between has and nox.
…

c.
age: with age become older, crim increase.
dis: with dis become larger, crim decrease.
rad: high rad, high crim
tax: high tax, high crim

d.
Most cities have low crime rates, but there are 18 cities crime rate > 20.
Most tax rates is between 660 - 680.
Partite rate is high around 20.

e.
sum(Boston$chas == 1)
35

f.
median(Boston$ptratio)
19.05

g.
subset(Boston, medv == min(Boston$medv))
399 and 406

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
399	38.3518	0	18.1	0	0.693	5.453	100	1.4896	24	666	20.2	396.90	30.59	5
406	67.9208	0	18.1	0	0.693	5.683	100	1.4254	24	666	20.2	384.97	22.98	5

Some of the features are good and some not. So, its good but not best or worst.

h.
dim(subset(Boston, rm > 7))
64
dim(subset(Boston, rm > 8))
13
They have lower crime.