Homework 1

Cally Thompson, Diane Keeton, Evan Porter, Jace Crist, Kayla Uhing
University of Nebraska at Omaha

BSAD 8700 - Business Analytics
Due: January 19, 2015

ANSWER FOR 8:

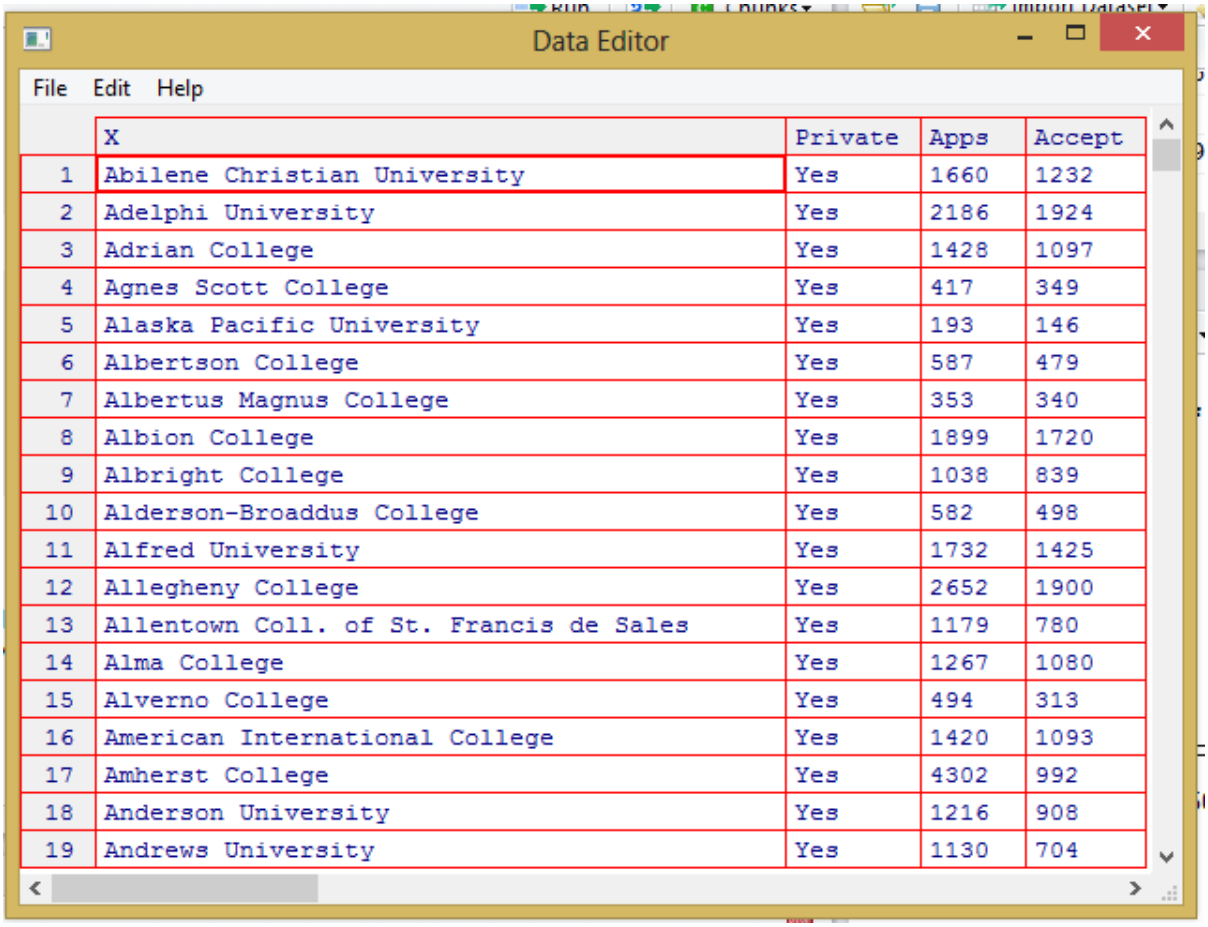
(a)
```
college<-read.csv("college.csv")
tail(college)
```

```
##                                  X Private  Apps Accept Enroll Top10perc
## 772 Worcester Polytechnic Institute     Yes  2768   2314    682        49
## 773         Worcester State College      No  2197   1515    543         4
## 774               Xavier University     Yes  1959   1805    695        24
## 775  Xavier University of Louisiana     Yes  2097   1915    695        34
## 776                 Yale University     Yes 10705   2453   1317        95
## 777    York College of Pennsylvania     Yes  2989   1855    691        28
##     Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal
## 772        86        2802          86    15884       5370   530      730
## 773        26        3089        2029     6797       3900   500     1200
## 774        47        2849        1107    11520       4960   600     1250
## 775        61        2793         166     6900       4200   617      781
## 776        99        5217          83    19840       6510   630     2115
## 777        63        2988        1726     4990       3560   500     1250
##     PhD Terminal S.F.Ratio perc.alumni Expend Grad.Rate
## 772  92       94      15.2          34  10774        82
## 773  60       60      21.0          14   4469        40
## 774  73       75      13.3          31   9189        83
## 775  67       75      14.4          20   8323        49
## 776  96       96       5.8          49  40386        99
## 777  75       75      18.1          28   4509        99
```

(b)
```
#fix(college)
rownames(college)=college[,1]
#fix(college)
college=college[,-1]
#fix(college)
```

Fix 1:

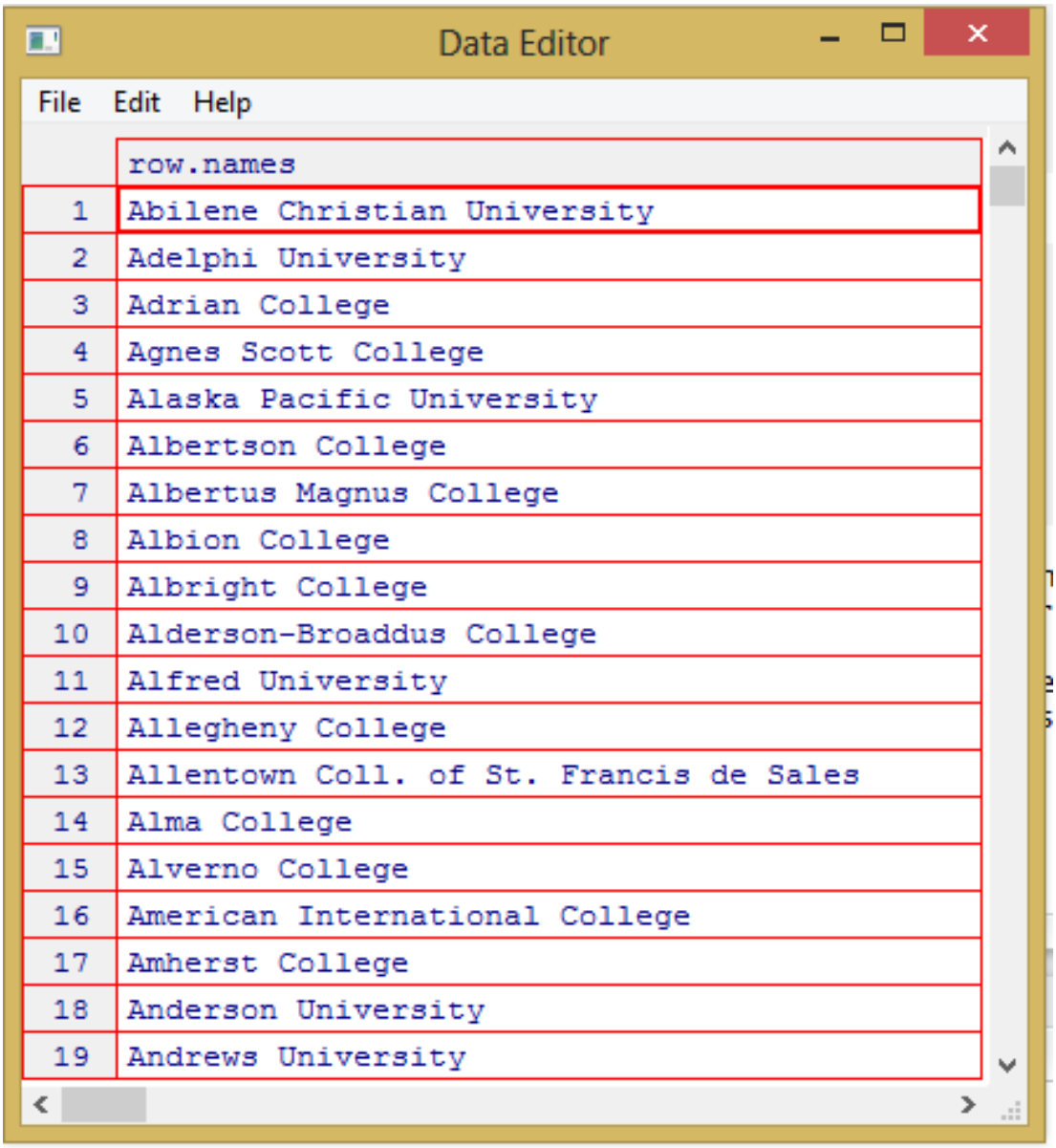
```
a<-readPNG("PNG1.png")
aRaster <- as.raster(a)
grid.raster(aRaster)
```



	X	Private	Apps	Accept
1	Abilene Christian University	Yes	1660	1232
2	Adelphi University	Yes	2186	1924
3	Adrian College	Yes	1428	1097
4	Agnes Scott College	Yes	417	349
5	Alaska Pacific University	Yes	193	146
6	Albertson College	Yes	587	479
7	Albertus Magnus College	Yes	353	340
8	Albion College	Yes	1899	1720
9	Albright College	Yes	1038	839
10	Alderson-Broaddus College	Yes	582	498
11	Alfred University	Yes	1732	1425
12	Allegheny College	Yes	2652	1900
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780
14	Alma College	Yes	1267	1080
15	Alverno College	Yes	494	313
16	American International College	Yes	1420	1093
17	Amherst College	Yes	4302	992
18	Anderson University	Yes	1216	908
19	Andrews University	Yes	1130	704

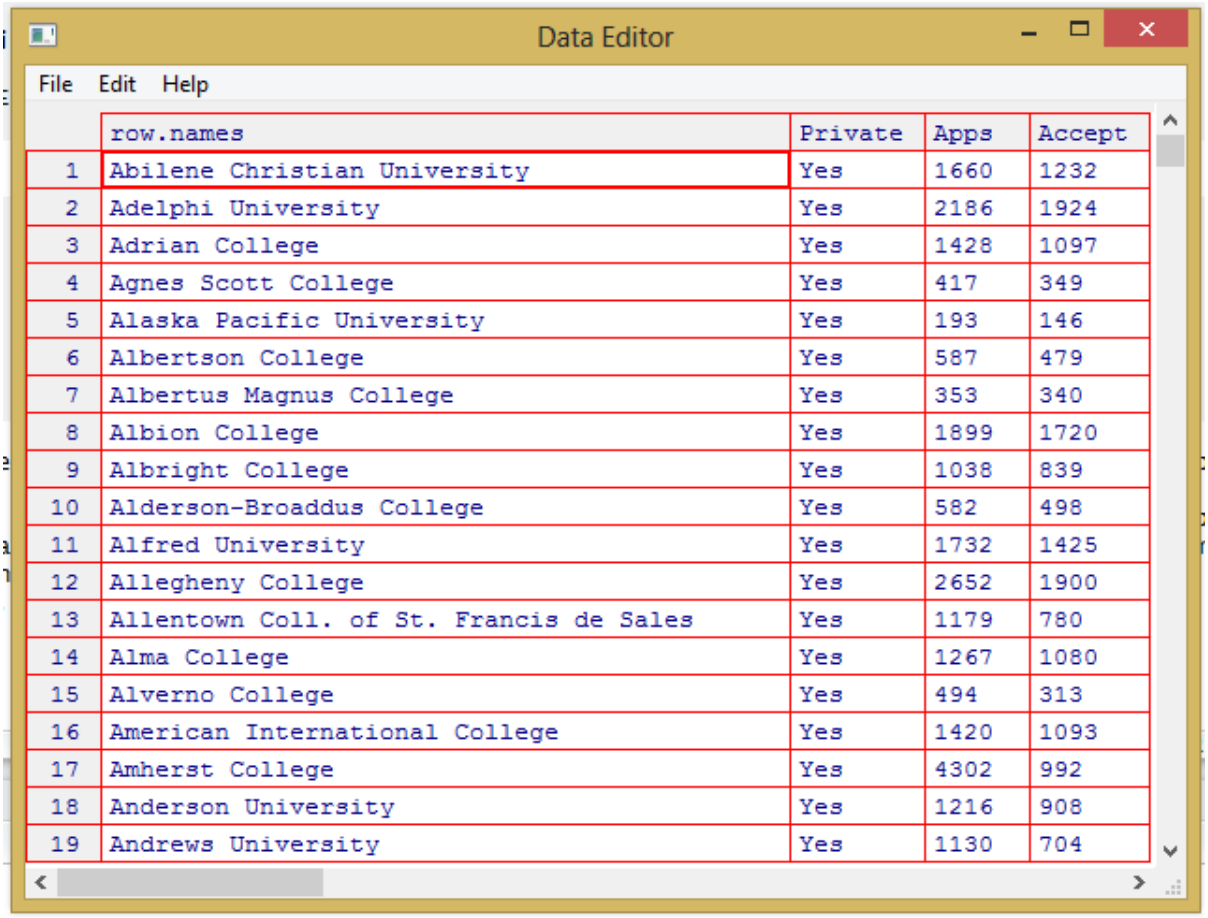
Fix 2:

```
b<-readPNG("PNG2.png")
bRaster <- as.raster(b)
grid.raster(bRaster)
```



Fix 3:

```
c<-readPNG("PNG3.png")
cRaster <- as.raster(c)
grid.raster(cRaster)
```



	row.names	Private	Apps	Accept
1	Abilene Christian University	Yes	1660	1232
2	Adelphi University	Yes	2186	1924
3	Adrian College	Yes	1428	1097
4	Agnes Scott College	Yes	417	349
5	Alaska Pacific University	Yes	193	146
6	Albertson College	Yes	587	479
7	Albertus Magnus College	Yes	353	340
8	Albion College	Yes	1899	1720
9	Albright College	Yes	1038	839
10	Alderson-Broaddus College	Yes	582	498
11	Alfred University	Yes	1732	1425
12	Allegheny College	Yes	2652	1900
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780
14	Alma College	Yes	1267	1080
15	Alverno College	Yes	494	313
16	American International College	Yes	1420	1093
17	Amherst College	Yes	4302	992
18	Anderson University	Yes	1216	908
19	Andrews University	Yes	1130	704

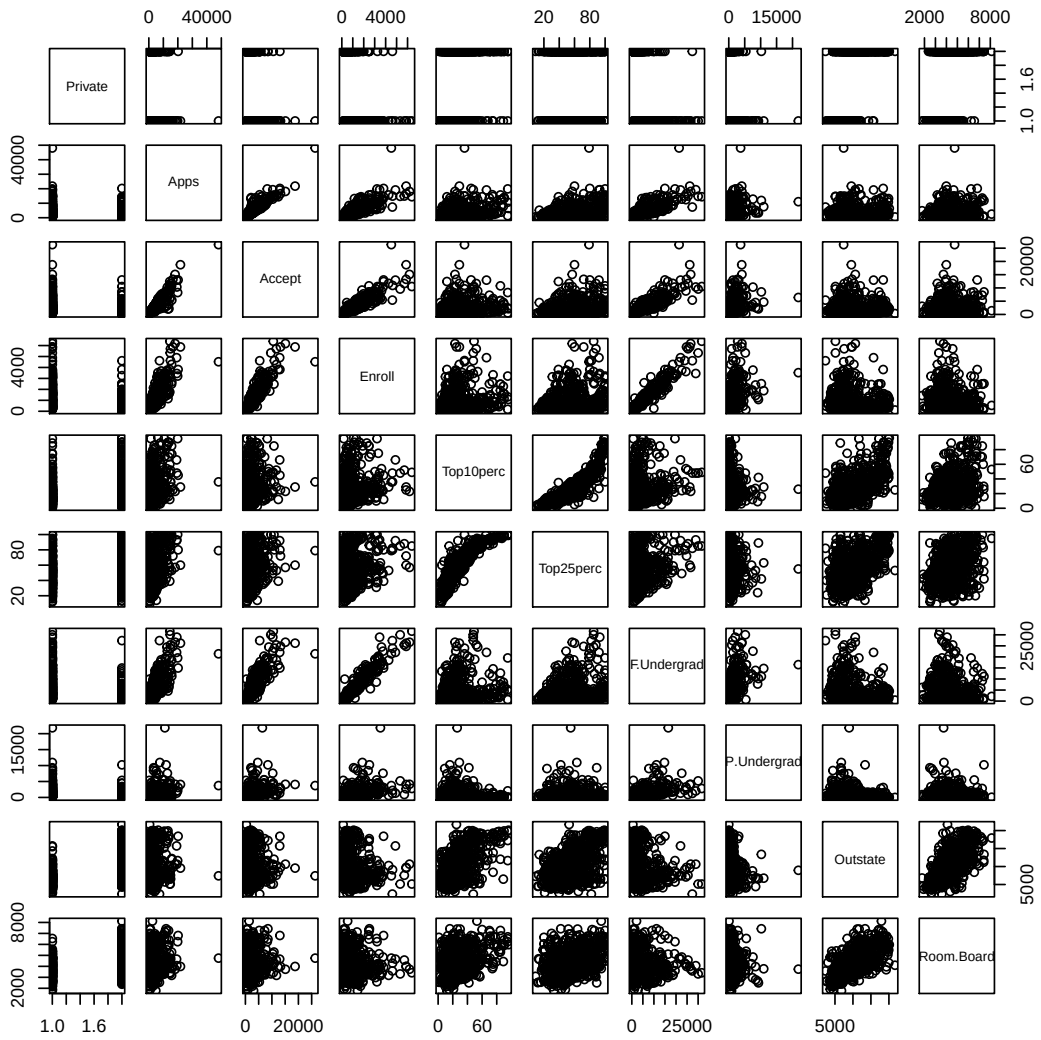
(c) (i)

```
summary(college)
```

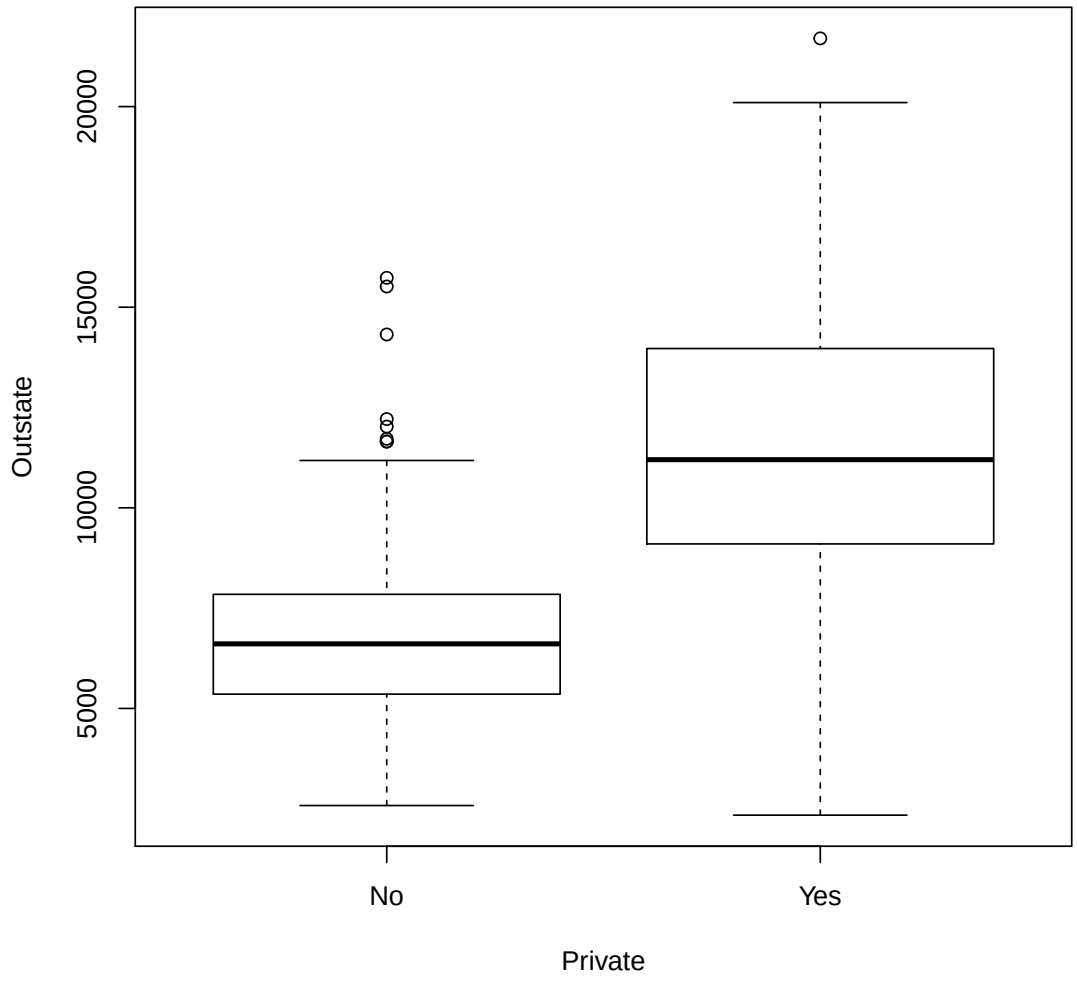
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
##            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)
```
attach(college)
plot(Outstate~Private)
```



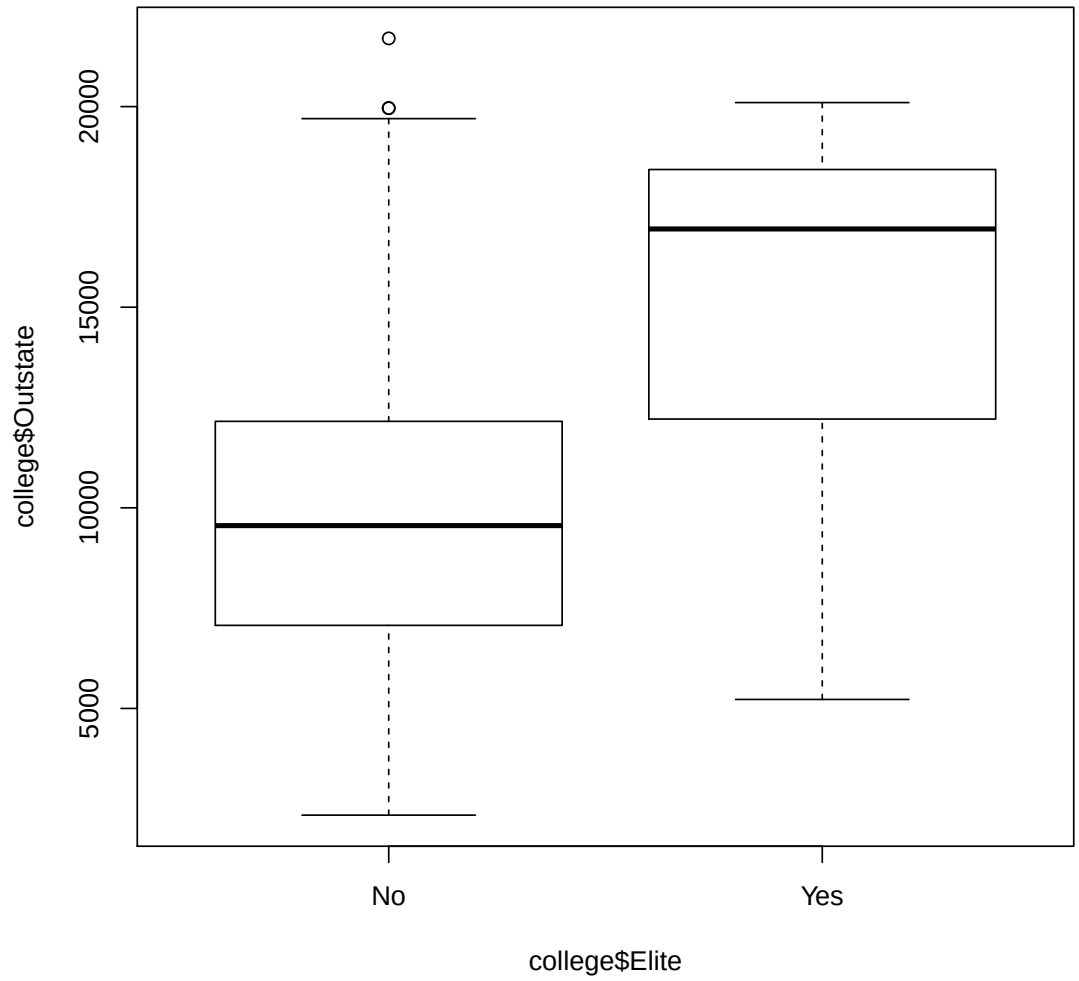
(iv) There are 78 elite Universities.

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

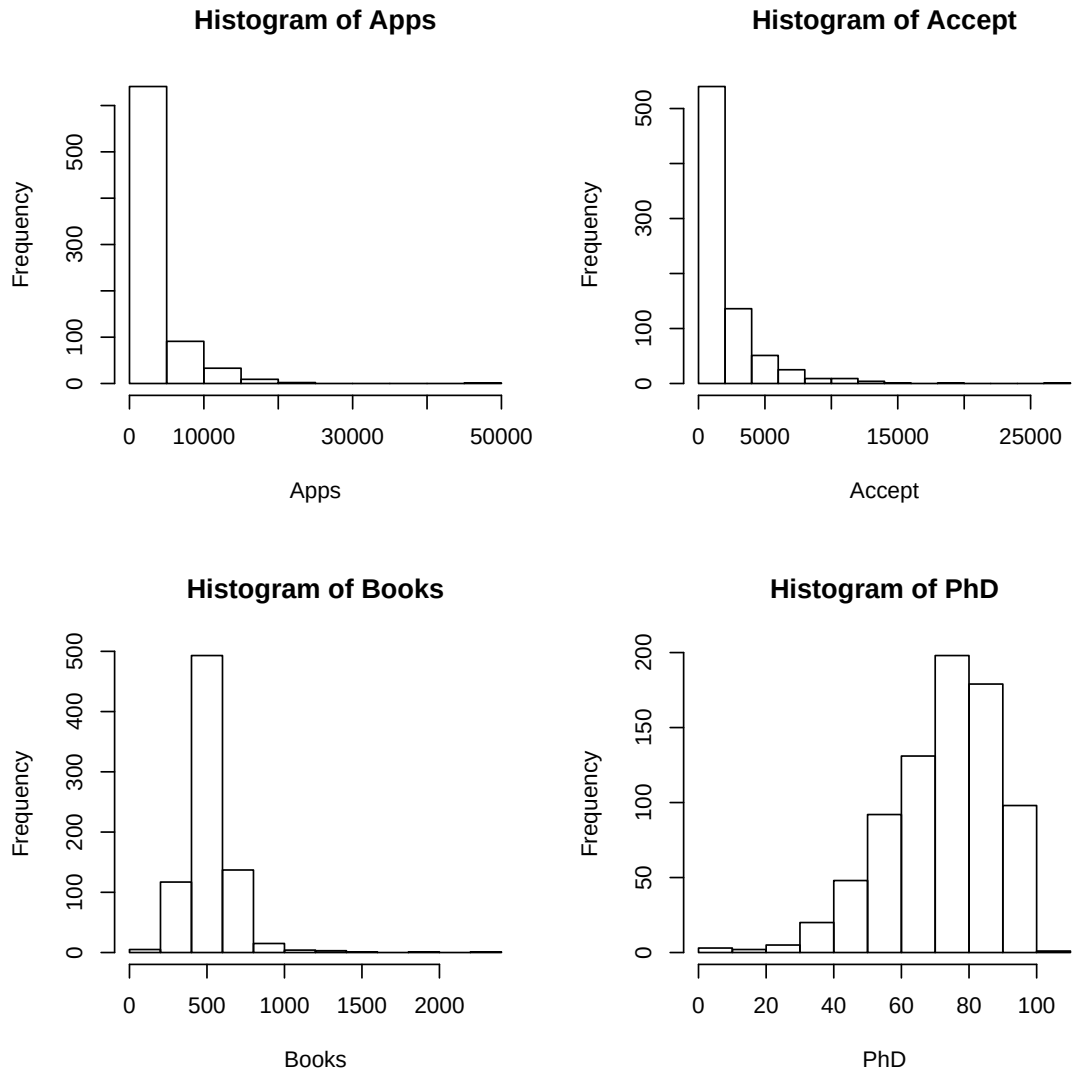
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
##            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$Outstate~college$Elite)
```



(v)
```
hist(Apps)
par(mfrow=c(2,2))
hist(Apps)
hist(Accept)
hist(Books)
hist(PhD)
```



(vi) After reviewing the data and performing the basic functions of R, we have found that there are numerous ways to extract information. There are also multiple different ways to find this data, there are no "right" or "wrong" ways of doing so. Through analyzing qualitative variables, R allows you to get the data you will need, although proper coding must be enforced. The basic commands were a bit difficult to figure out, but the more we did the assignment, the more we started learning what we were actually doing. The visuals of the Histograms and Plot graphs were helpful in dissecting the data.

ANSWER FOR 10:

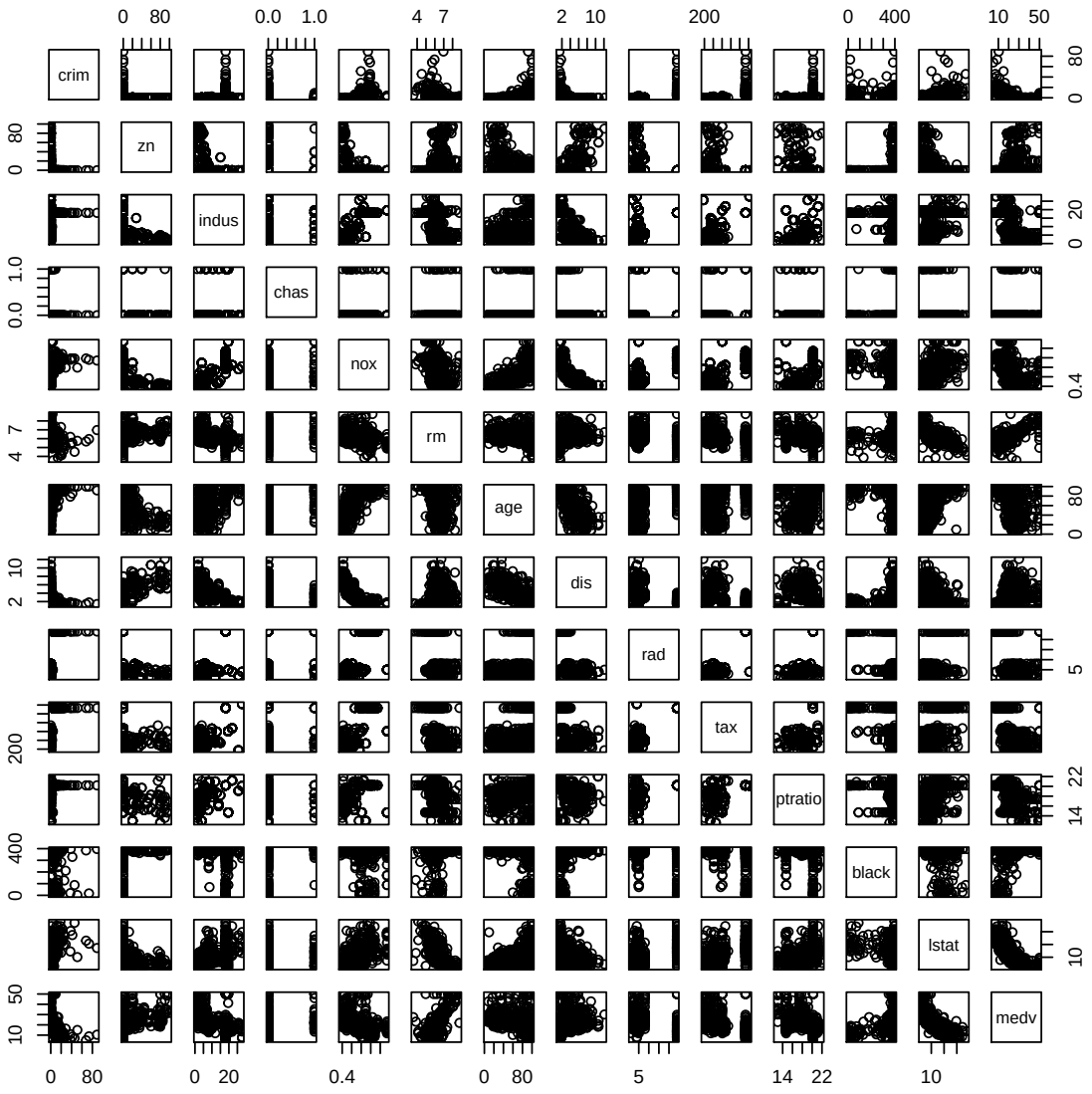
(a)
```
tail(Boston)
```

```
##        crim zn indus chas   nox    rm  age    dis rad tax ptratio  black
## 501 0.22438  0  9.69    0 0.585 6.027 79.7 2.4982   6 391    19.2 396.90
## 502 0.06263  0 11.93    0 0.573 6.593 69.1 2.4786   1 273    21.0 391.99
## 503 0.04527  0 11.93    0 0.573 6.120 76.7 2.2875   1 273    21.0 396.90
```

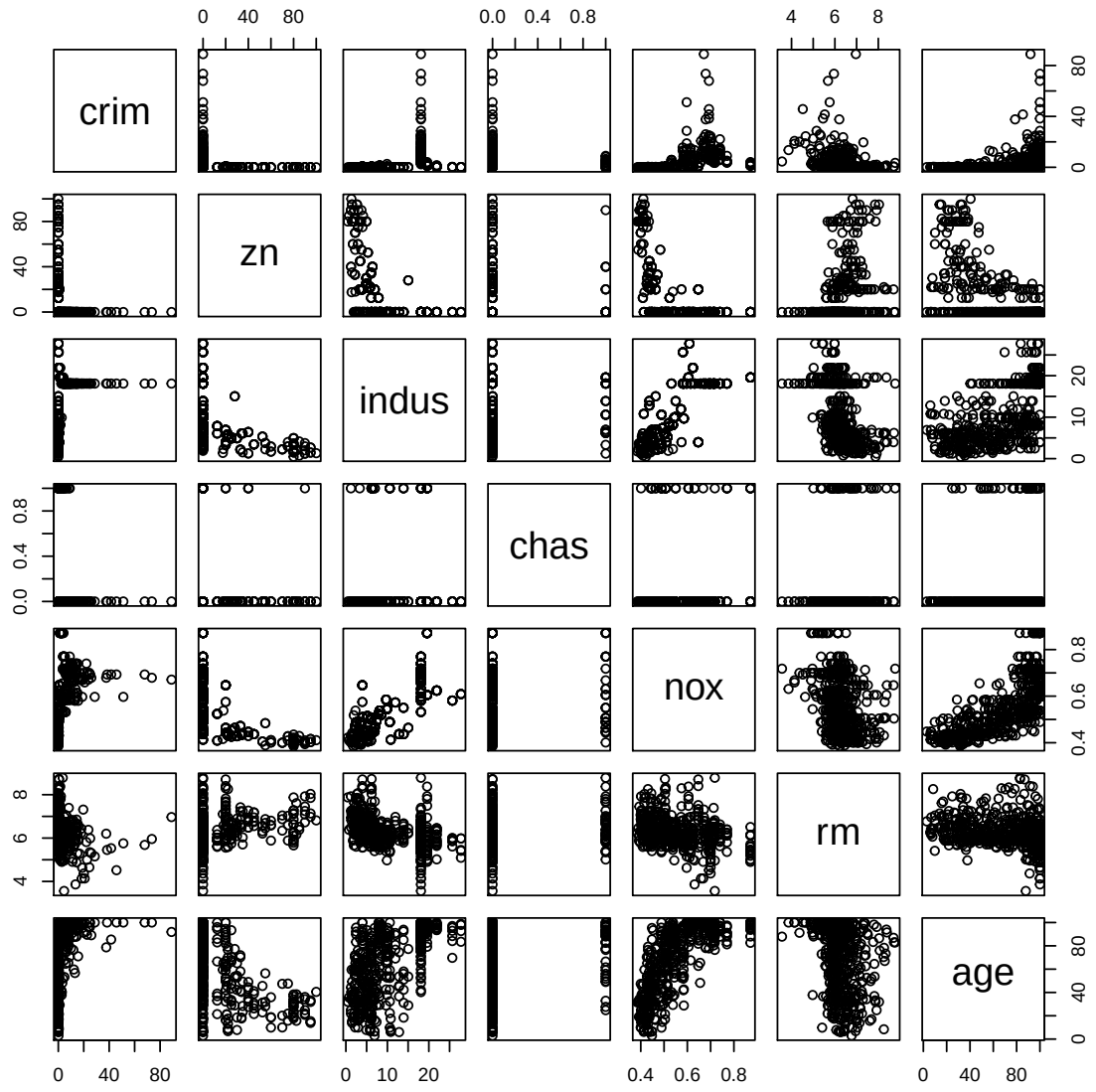
```
## 504 0.06076  0 11.93    0 0.573 6.976 91.0 2.1675   1 273    21.0 396.90
## 505 0.10959  0 11.93    0 0.573 6.794 89.3 2.3889   1 273    21.0 393.45
## 506 0.04741  0 11.93    0 0.573 6.030 80.8 2.5050   1 273    21.0 396.90
##     lstat medv
## 501 14.33 16.8
## 502  9.67 22.4
## 503  9.08 20.6
## 504  5.64 23.9
## 505  6.48 22.0
## 506  7.88 11.9

dim(Boston)

## [1] 506  14

#?(Boston)
```

The Boston data frame has 506 rows and 14 columns.
The columns and rows represent data including crime, age and median value of owner-occupied homes in $ 1000s

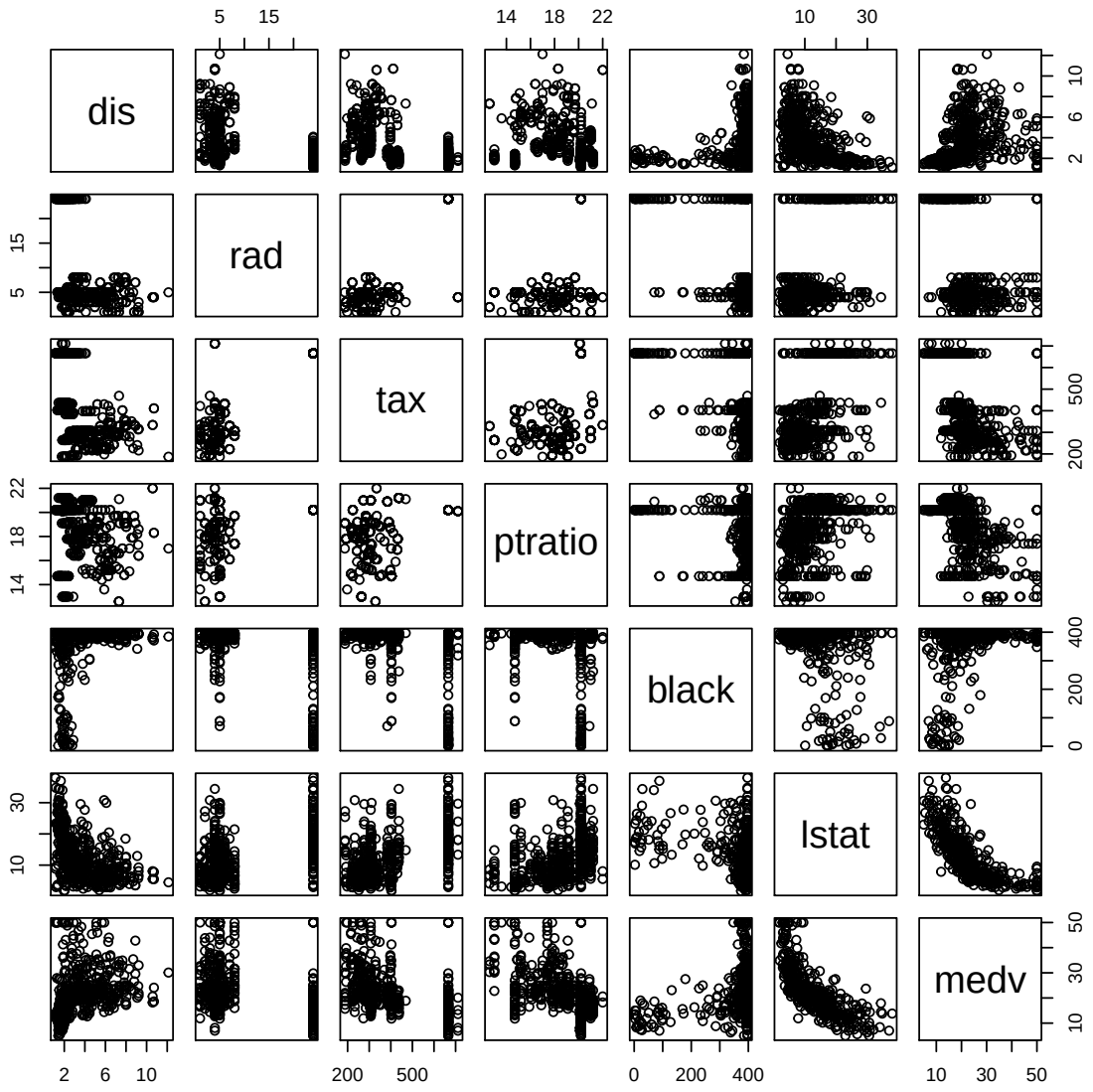
(b)
```
pairs(Boston)
```




```
pairs(Boston[,1:7])
```




```
pairs(Boston[,8:14])
```



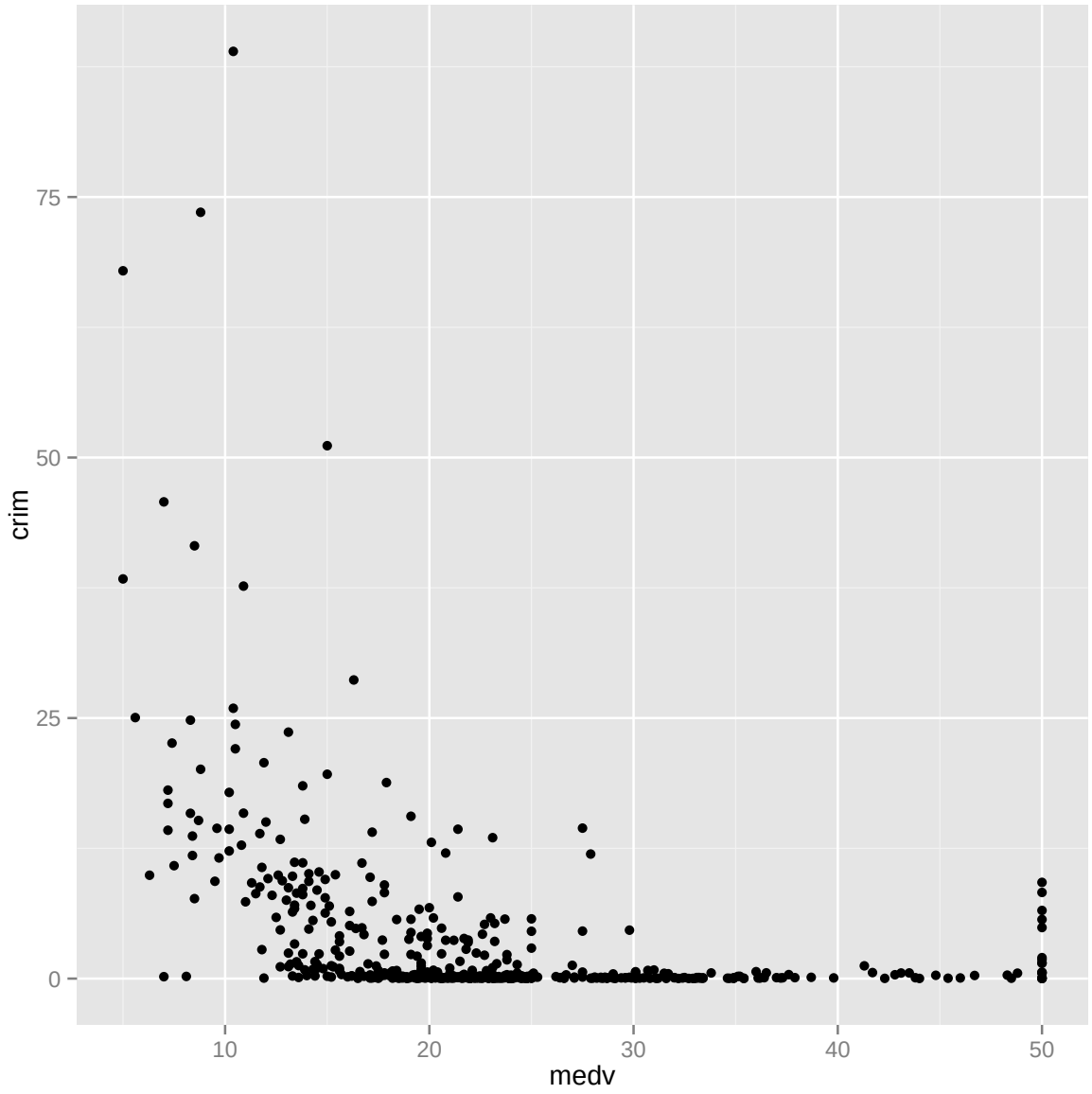

```
round(cor(Boston),2)
```

```
##           crim    zn indus  chas   nox    rm   age   dis   rad   tax
## crim      1.00 -0.20  0.41 -0.06  0.42 -0.22  0.35 -0.38  0.63  0.58
## zn       -0.20  1.00 -0.53 -0.04 -0.52  0.31 -0.57  0.66 -0.31 -0.31
## indus     0.41 -0.53  1.00  0.06  0.76 -0.39  0.64 -0.71  0.60  0.72
## chas     -0.06 -0.04  0.06  1.00  0.09  0.09  0.09 -0.10 -0.01 -0.04
## nox       0.42 -0.52  0.76  0.09  1.00 -0.30  0.73 -0.77  0.61  0.67
## rm       -0.22  0.31 -0.39  0.09 -0.30  1.00 -0.24  0.21 -0.21 -0.29
## age       0.35 -0.57  0.64  0.09  0.73 -0.24  1.00 -0.75  0.46  0.51
## dis      -0.38  0.66 -0.71 -0.10 -0.77  0.21 -0.75  1.00 -0.49 -0.53
## rad       0.63 -0.31  0.60 -0.01  0.61 -0.21  0.46 -0.49  1.00  0.91
## tax       0.58 -0.31  0.72 -0.04  0.67 -0.29  0.51 -0.53  0.91  1.00
## ptratio   0.29 -0.39  0.38 -0.12  0.19 -0.36  0.26 -0.23  0.46  0.46
## black    -0.39  0.18 -0.36  0.05 -0.38  0.13 -0.27  0.29 -0.44 -0.44
```

```
## lstat    0.46 -0.41  0.60 -0.05  0.59 -0.61  0.60 -0.50  0.49  0.54
## medv    -0.39  0.36 -0.48  0.18 -0.43  0.70 -0.38  0.25 -0.38 -0.47
##         ptratio black lstat  medv
## crim       0.29 -0.39  0.46 -0.39
## zn        -0.39  0.18 -0.41  0.36
## indus      0.38 -0.36  0.60 -0.48
## chas      -0.12  0.05 -0.05  0.18
## nox        0.19 -0.38  0.59 -0.43
## rm        -0.36  0.13 -0.61  0.70
## age        0.26 -0.27  0.60 -0.38
## dis       -0.23  0.29 -0.50  0.25
## rad        0.46 -0.44  0.49 -0.38
## tax        0.46 -0.44  0.54 -0.47
## ptratio    1.00 -0.18  0.37 -0.51
## black     -0.18  1.00 -0.37  0.33
## lstat      0.37 -0.37  1.00 -0.74
## medv      -0.51  0.33 -0.74  1.00
```

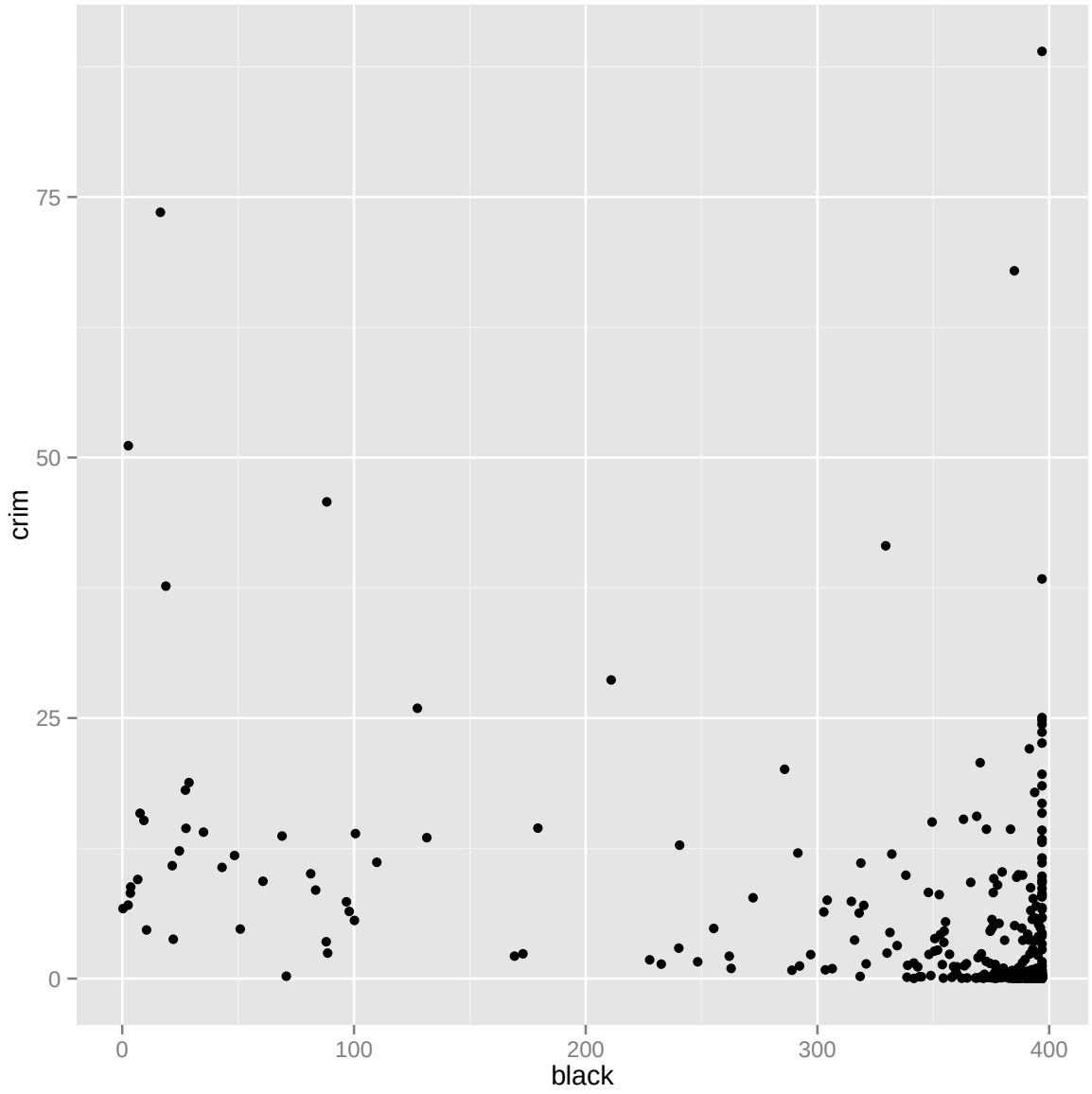
There is a negative correlation between the lower status of the population (lstat) and median value of owner occupied homes (medv). This correlation could be attributed to the fact that the lower the individual status, the lower the income of the home. Another example of negative correlation would be from the nitrogen oxcide concentration (nox) and the weighted mean of distances to Boston employment centers (dis). This correlation is related to the fact that as employment centers are more spread out, the polution decreases. An example of positive correlation would be the relationship between the average number of rooms per dwelling (rm) and median value of owner occupied home (medv). This graph shows that the number of rooms in a dwelling go up as the median value of the home increases.

(c)
```
ggplot(Boston, aes(medv,crim))+geom_point()
```

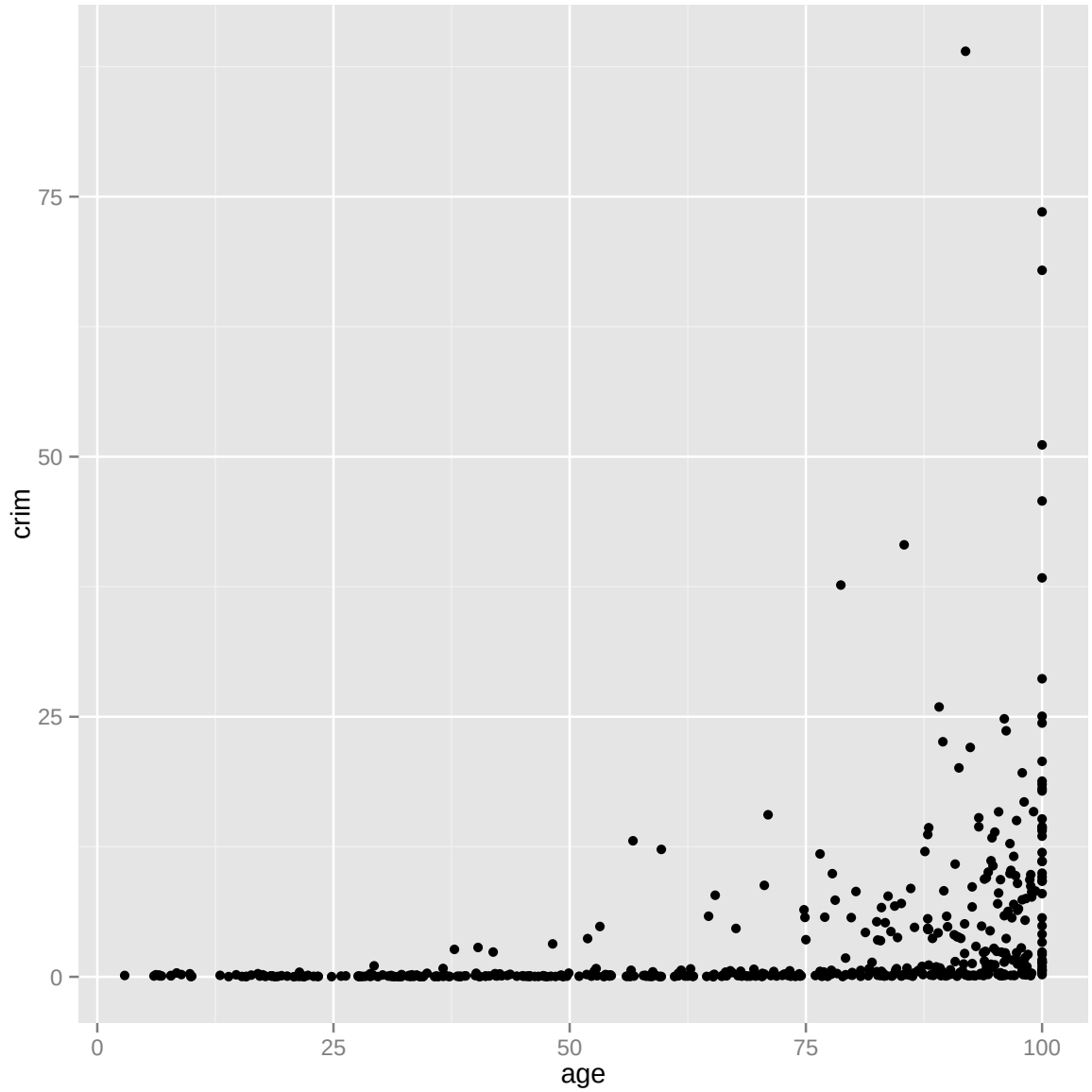


In regard to predictors associated with crime rate, it is shown that there seems to be a relationship such that the higher the median value of the home, the lower the crime rate.

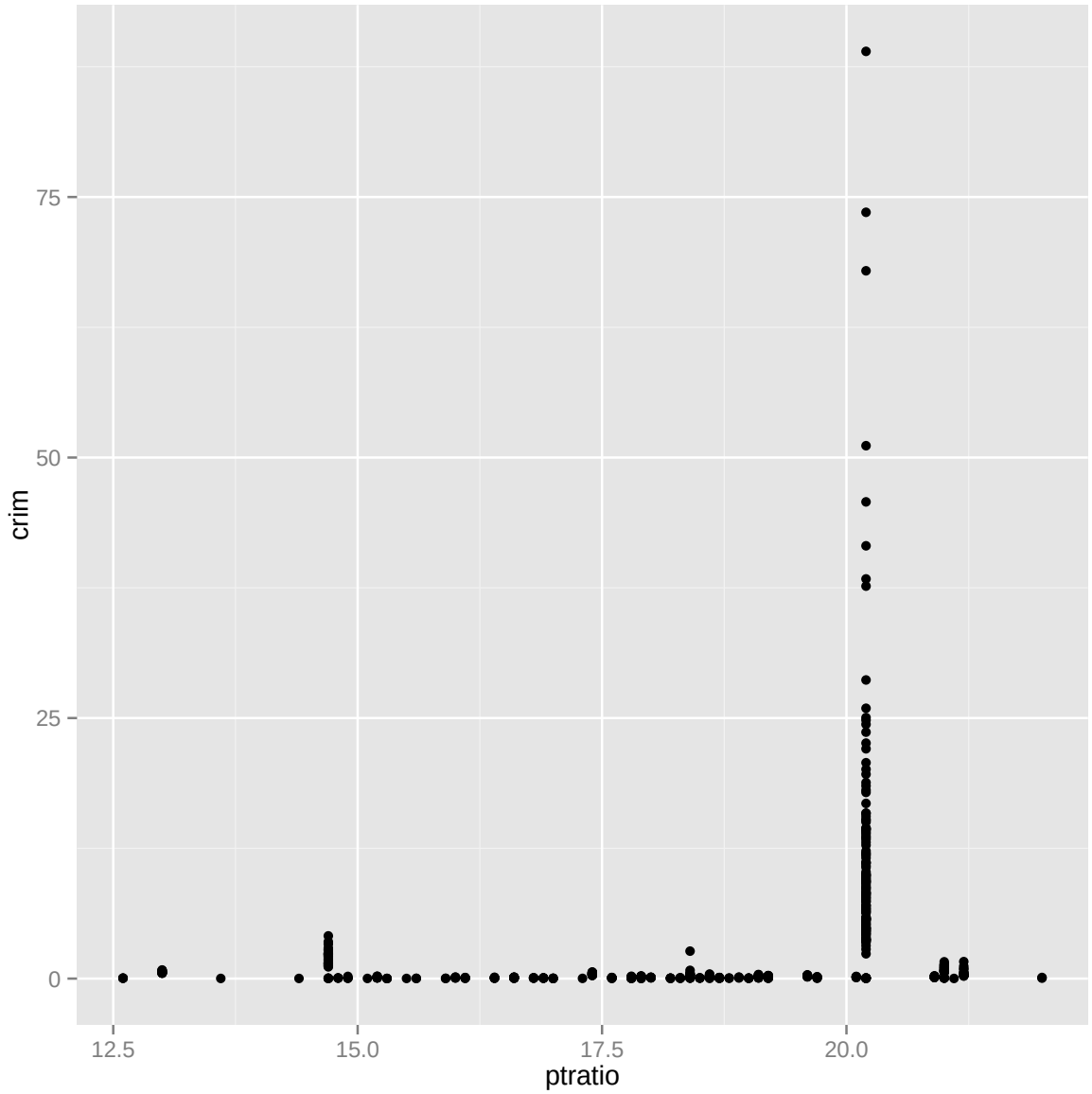
(d)
```
write.csv(Boston, "Boston.csv")
# colnames(Boston)
ggplot(Boston, aes(black,crim))+geom_point()
```



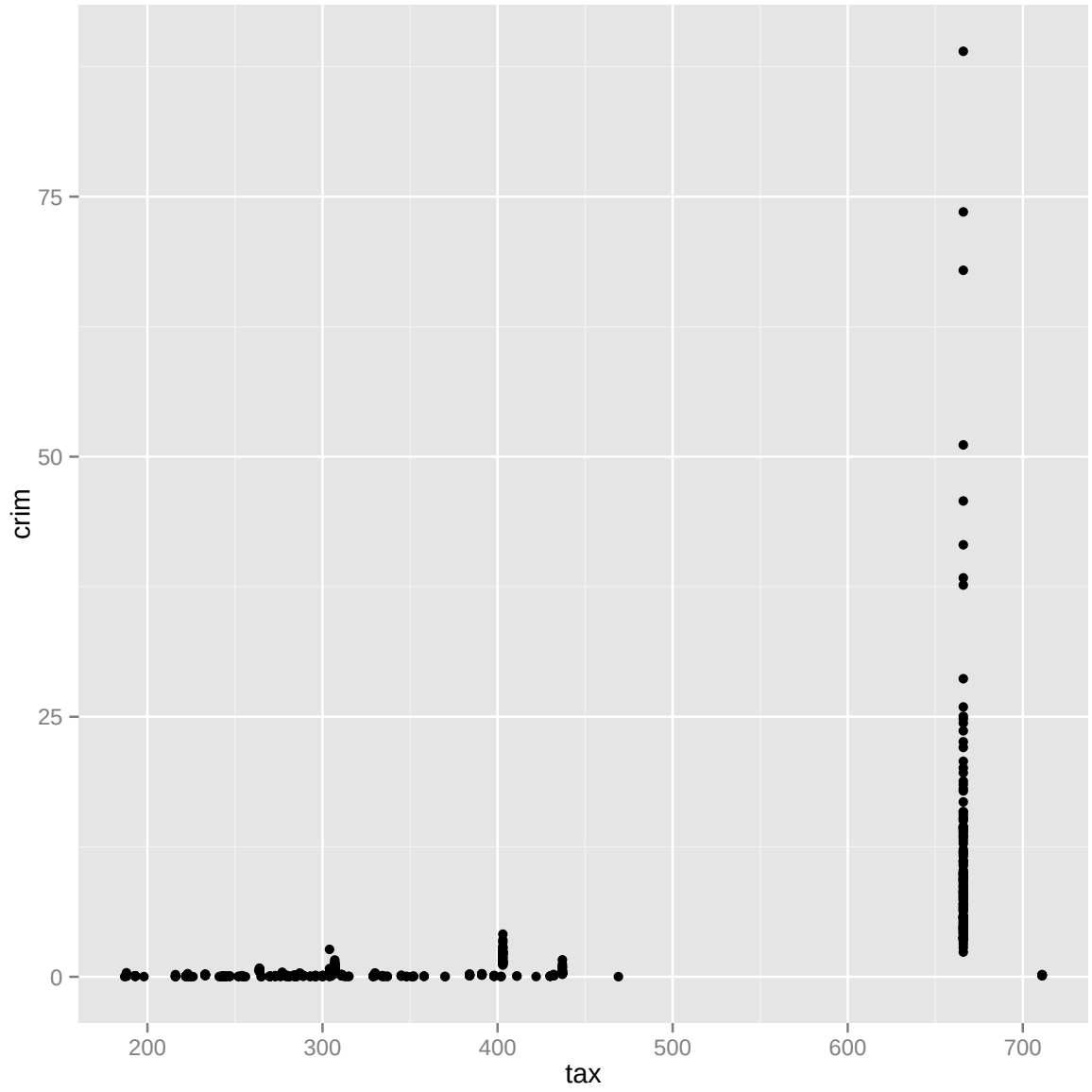

```
ggplot(Boston, aes(age,crim))+geom_point()
```



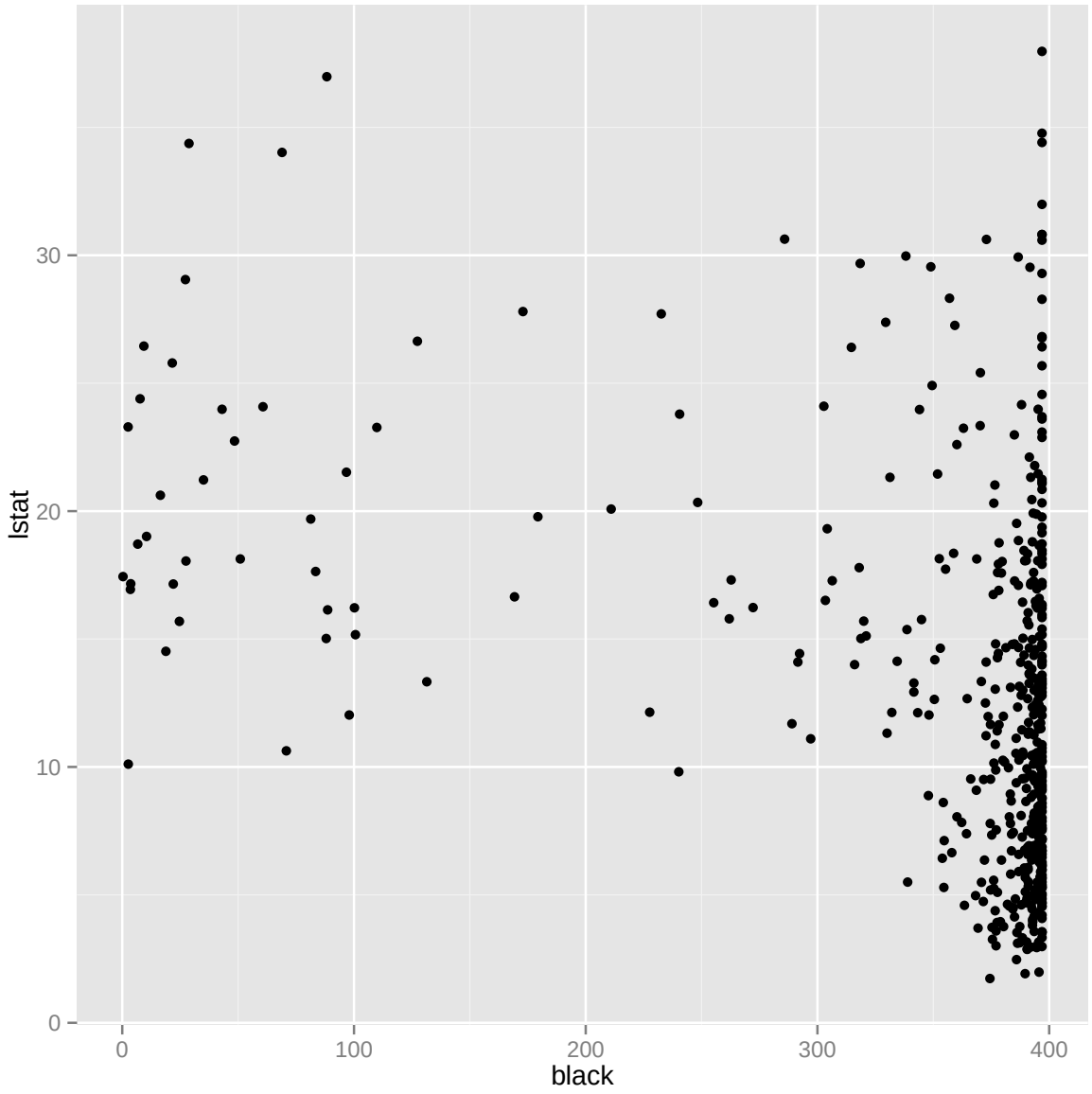

```
ggplot(Boston, aes(ptratio,crim))+geom_point()
```



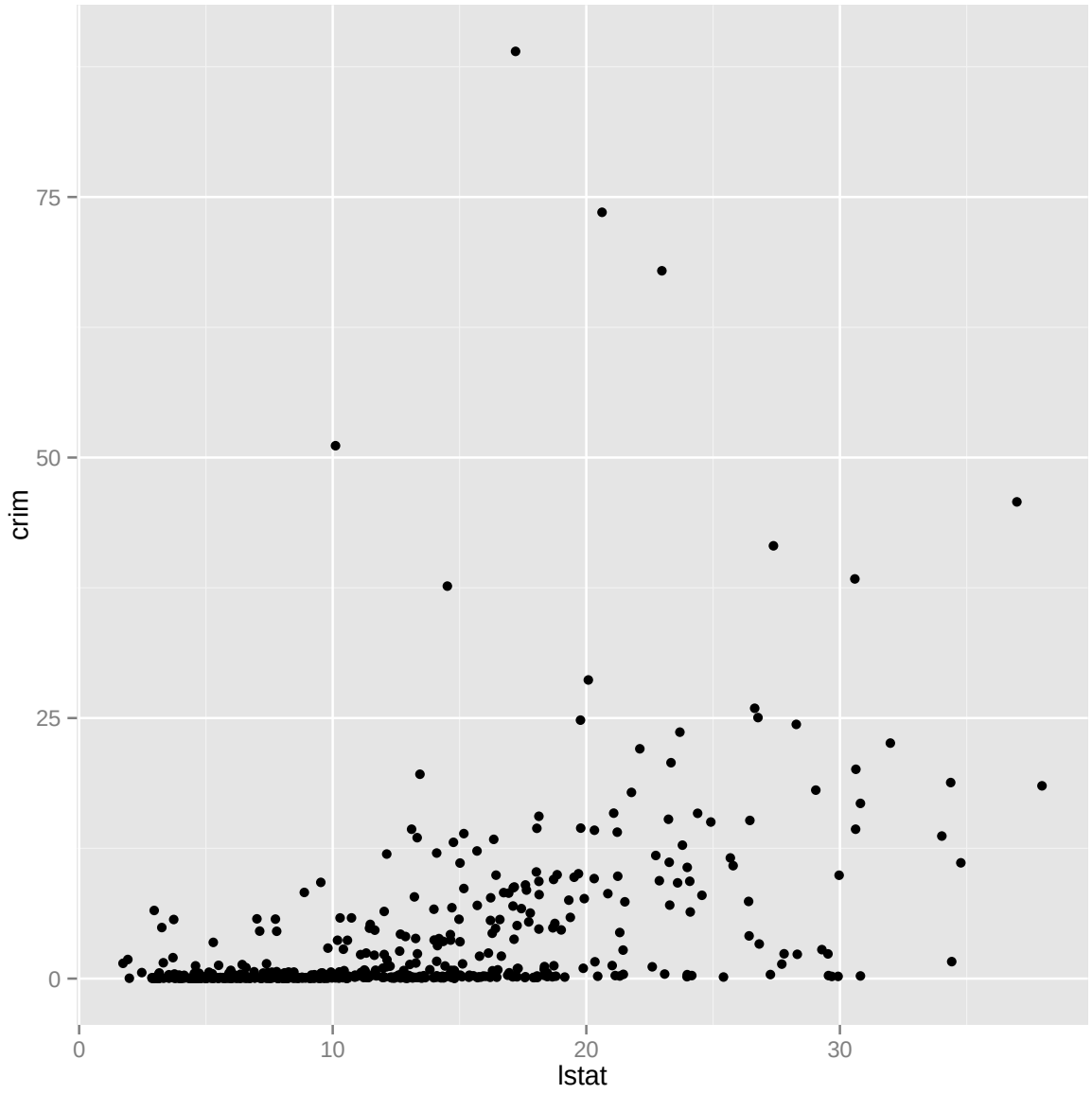

```
ggplot(Boston, aes(tax,crim))+geom_point()
```




```
ggplot(Boston, aes(black,lstat))+geom_point()
```




```
ggplot(Boston, aes(lstat,crim))+geom_point()
```



We were able to determine that crime rates increase in older houses and in areas with higher black populations; by using scatterplots. Also, we determined that suburbs 400-500 have higher tax rates, and 357-488 have above average crime rates. Pupil teacher ratios appear to remain relatively constant within all suburbs.

(e)
```
Boston1<-Boston %>%
  filter(chas>0)
dim(Boston1)

## [1] 35 14
```

We were able to calculate that 35 suburbs in the data set bound the Charles River

(f)
```
median(Boston$ptratio)

## [1] 19.05
```

The median of the pupil teacher ratio is 19.05

(g)
```
Boston2<-Boston %>%
  mutate(suburb=as.numeric(1:nrow(Boston))) %>%
  select(suburb,medv) %>%
  arrange(-medv)
tail(Boston2,3)

##     suburb medv
## 504    401  5.6
## 505    399  5.0
## 506    406  5.0
```

Suburb 399 and 406 have the lowest median value of owner occupied homes.

(h)
```
Boston3<-Boston %>%
  filter(rm>7)
dim(Boston3)

## [1] 64 14

round((64/506)*100)

## [1] 13
```

13% have rooms with more than 7 rooms

```
Boston3<-Boston %>%
  filter(rm>8)
dim(Boston3)

## [1] 13 14

round((13/506)*100,2)

## [1] 2.57
```

3% have rooms with more than 8 rooms