

Problem 1: James 2.4, Exercise 8

Part a

```
> college <- read.csv(file="datasets/College.csv")
> head(college)
```

| | | X | Private | Apps | Accept | Enroll | Top10perc | Top25perc | F.Undergrad | P.Undergrad | Outstate | Room.Board | Books |
|---|------------------------------|-----|---------|------|--------|--------|-----------|-----------|-------------|-------------|----------|------------|-------|
| 1 | Abilene Christian University | Yes | 1660 | 1232 | 721 | 23 | 52 | 2885 | 537 | 7440 | 3300 | 450 | |
| 2 | Adelphi University | Yes | 2186 | 1924 | 512 | 16 | 29 | 2683 | 1227 | 12280 | 6450 | 750 | |
| 3 | Adrian College | Yes | 1428 | 1097 | 336 | 22 | 50 | 1036 | 99 | 11250 | 3750 | 400 | |
| 4 | Agnes Scott College | Yes | 417 | 349 | 137 | 60 | 89 | 510 | 63 | 12960 | 5450 | 450 | |
| 5 | Alaska Pacific University | Yes | 193 | 146 | 55 | 16 | 44 | 249 | 869 | 7560 | 4120 | 800 | |
| 6 | Albertson College | Yes | 587 | 479 | 158 | 38 | 62 | 678 | 41 | 13500 | 3335 | 500 | |

| | Personal | PhD | Terminal | S.F.Ratio | perc.alumni | Expend | Grad.Rate |
|---|----------|-----|----------|-----------|-------------|--------|-----------|
| 1 | 2200 | 70 | 78 | 18.1 | 12 | 7041 | 60 |
| 2 | 1500 | 29 | 30 | 12.2 | 16 | 10527 | 56 |
| 3 | 1165 | 53 | 66 | 12.9 | 30 | 8735 | 54 |
| 4 | 875 | 92 | 97 | 7.7 | 37 | 19016 | 59 |
| 5 | 1500 | 76 | 72 | 11.9 | 2 | 10922 | 15 |
| 6 | 675 | 67 | 73 | 9.4 | 11 | 9727 | 55 |

Part b

| | row.names | 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 |
| 7 | Albertus Magnus College | Yes | 353 | 340 | 103 | 17 | 45 |
| 8 | Albion College | Yes | 1899 | 1720 | 489 | 37 | 68 |
| 9 | Albright College | Yes | 1038 | 839 | 227 | 30 | 63 |
| 10 | Alderson-Broadbudd College | Yes | 582 | 498 | 172 | 21 | 44 |

Part c

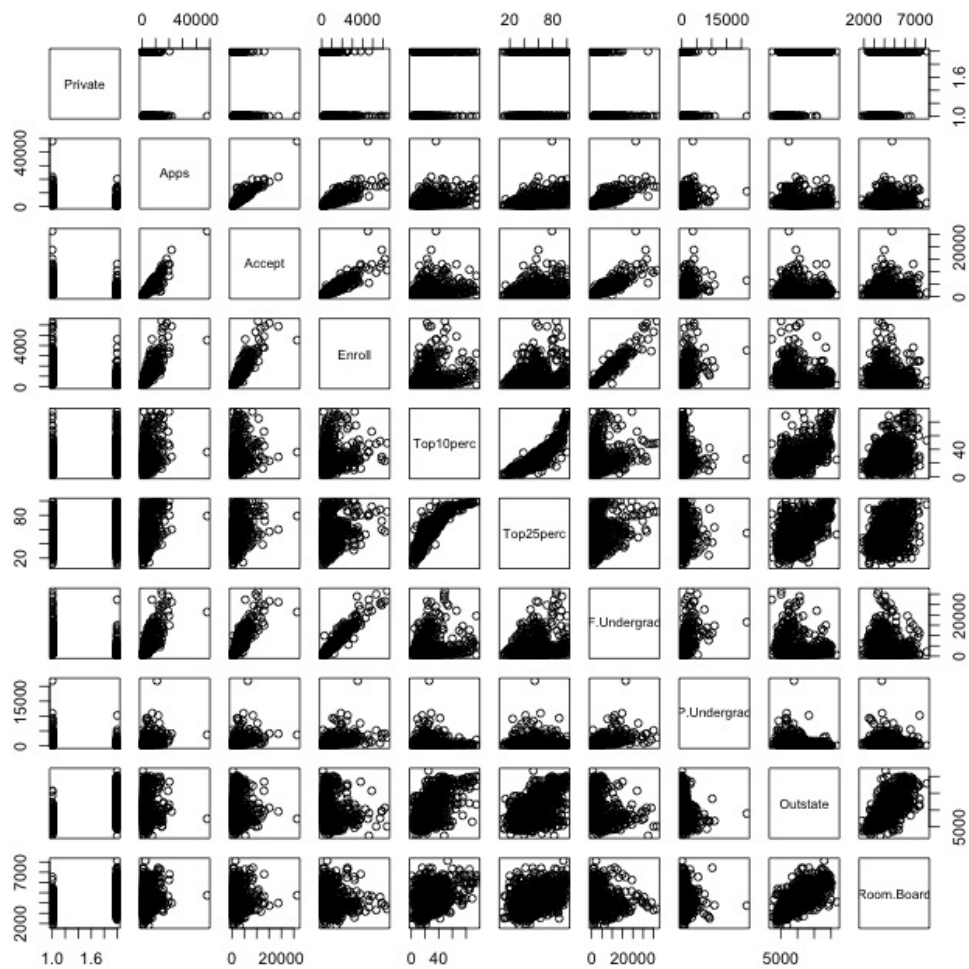
Part i

```
> summary(college)
```

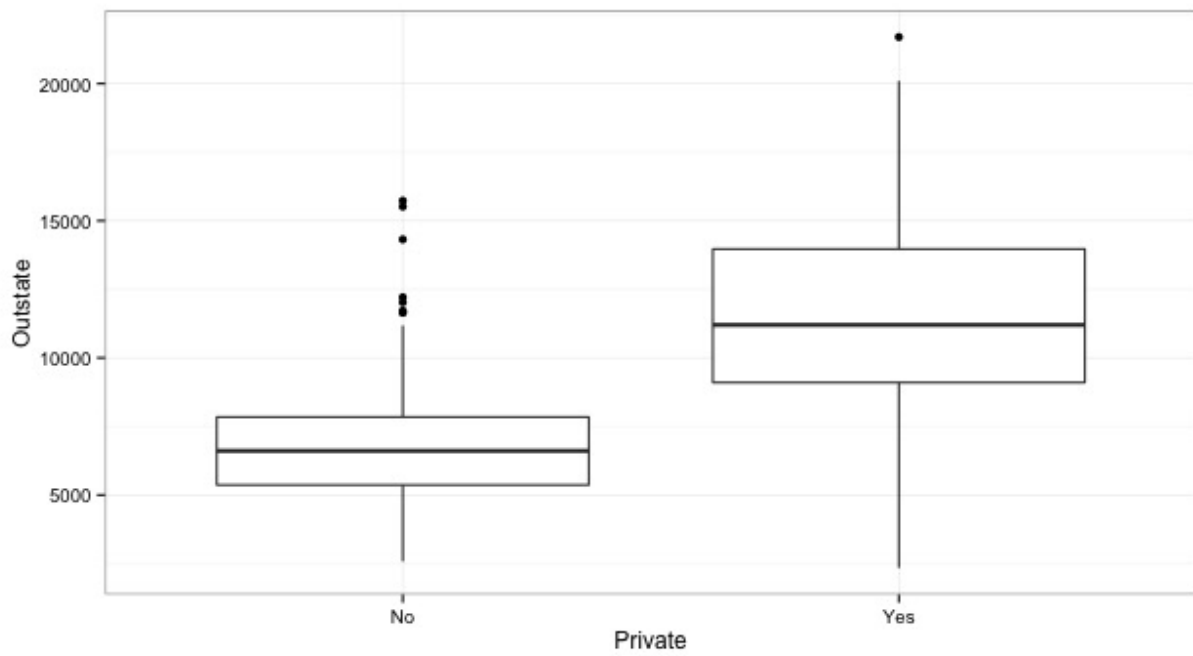
| Private | Apps | Accept | Enroll | Top10perc | Top25perc | F.Undergrad | P.Undergrad | Outstate |
|---------|---------------|---------------|--------------|---------------|---------------|---------------|----------------|---------------|
| No :212 | Min. : 81 | Min. : 72 | Min. : 35 | Min. : 1.00 | Min. : 9.0 | Min. : 139 | Min. : 1.0 | Min. : 2340 |
| Yes:565 | 1st Qu.: 776 | 1st Qu.: 684 | 1st Qu.: 242 | 1st Qu.:15.00 | 1st Qu.: 41.0 | 1st Qu.: 992 | 1st Qu.: 95.0 | 1st Qu.: 7320 |
| | Median : 1558 | Median : 1110 | Median : 434 | Median :23.00 | Median : 54.0 | Median : 1707 | Median : 353.0 | Median : 9990 |
| | Mean : 3002 | Mean : 2019 | Mean : 780 | Mean :27.56 | Mean : 55.8 | Mean : 3700 | Mean : 855.3 | Mean :10441 |
| | 3rd Qu.: 3624 | 3rd Qu.: 2424 | 3rd Qu.: 902 | 3rd Qu.:35.00 | 3rd Qu.: 69.0 | 3rd Qu.: 4005 | 3rd Qu.: 967.0 | 3rd Qu.:12925 |
| | Max. :48094 | Max. :26330 | Max. :6392 | Max. :96.00 | Max. :100.0 | Max. :31643 | Max. :21836.0 | Max. :21700 |

| Room.Board | Books | Personal | PhD | Terminal | S.F.Ratio | perc.alumni | Expend | Grad.Rate |
|--------------|----------------|--------------|----------------|---------------|---------------|---------------|---------------|----------------|
| Min. :1780 | Min. : 96.0 | Min. : 250 | Min. : 8.00 | Min. : 24.0 | Min. : 2.50 | Min. : 0.00 | Min. : 3186 | Min. : 10.00 |
| 1st Qu.:3597 | 1st Qu.: 470.0 | 1st Qu.: 850 | 1st Qu.: 62.00 | 1st Qu.: 71.0 | 1st Qu.:11.50 | 1st Qu.:13.00 | 1st Qu.: 6751 | 1st Qu.: 53.00 |
| Median :4200 | Median : 500.0 | Median :1200 | Median : 75.00 | Median : 82.0 | Median :13.60 | Median :21.00 | Median : 8377 | Median : 65.00 |
| Mean :4358 | Mean : 549.4 | Mean :1341 | Mean : 72.66 | Mean : 79.7 | Mean :14.09 | Mean :22.74 | Mean : 9660 | Mean : 65.46 |
| 3rd Qu.:5050 | 3rd Qu.: 600.0 | 3rd Qu.:1700 | 3rd Qu.: 85.00 | 3rd Qu.: 92.0 | 3rd Qu.:16.50 | 3rd Qu.:31.00 | 3rd Qu.:10830 | 3rd Qu.: 78.00 |
| Max. :8124 | Max. :2340.0 | Max. :6800 | Max. :103.00 | Max. :100.0 | Max. :39.80 | Max. :64.00 | Max. :56233 | Max. :118.00 |

Part ii

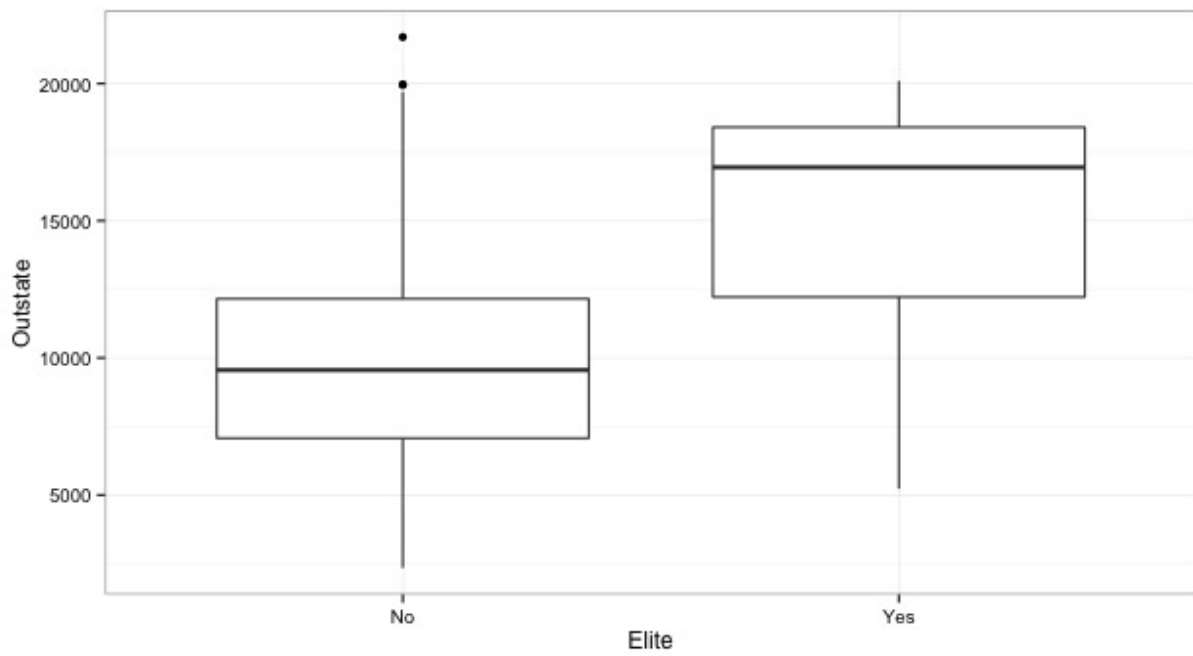


Part iii

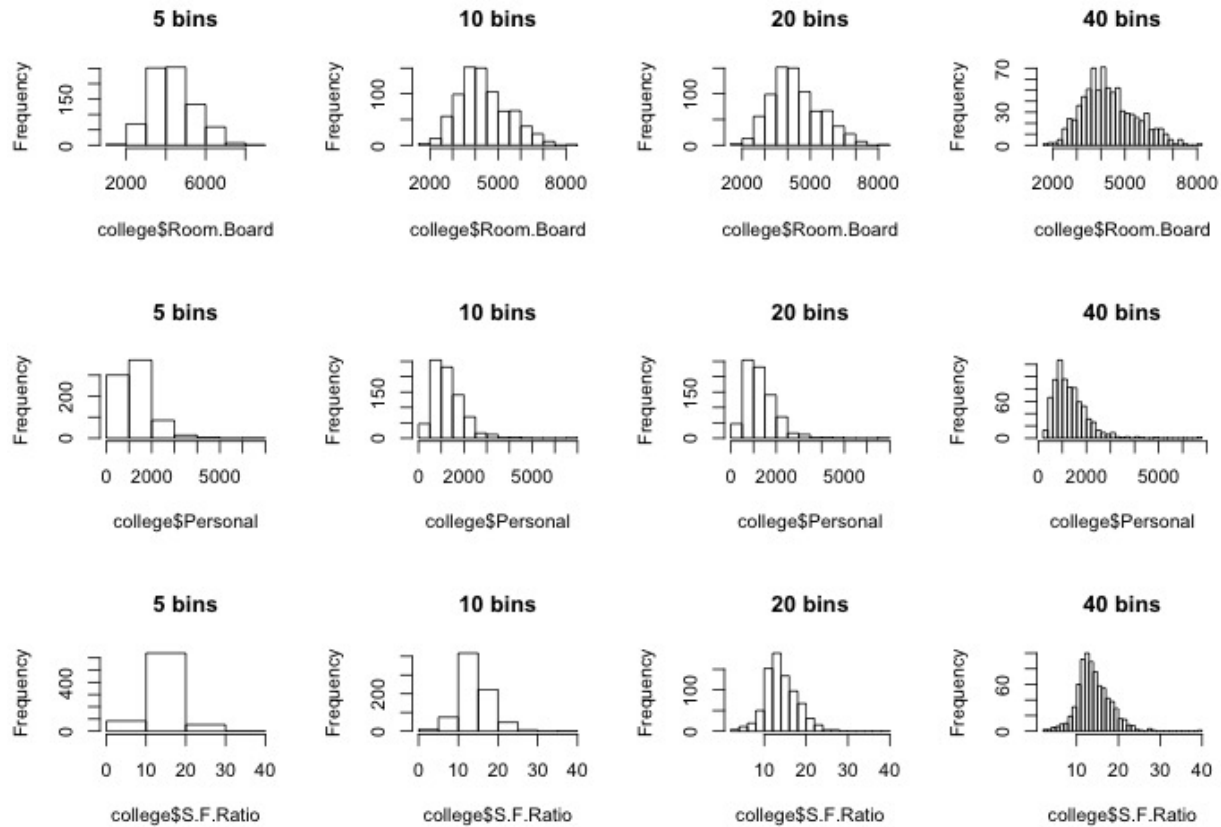


Part iv

There are 78 colleges categorized as "Elite".

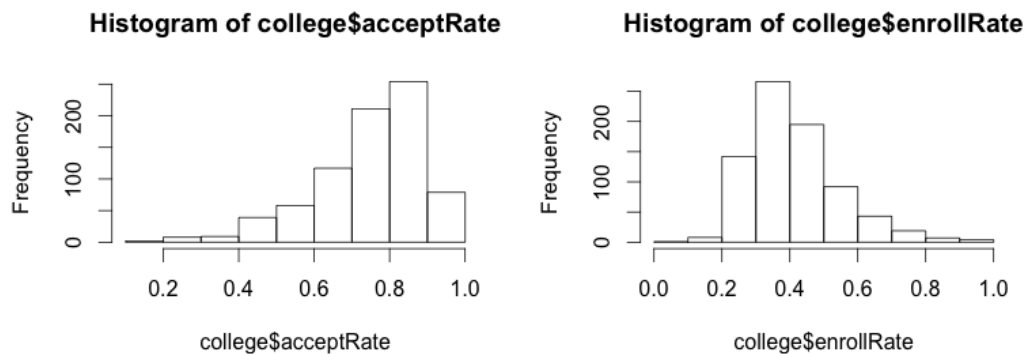


Part v



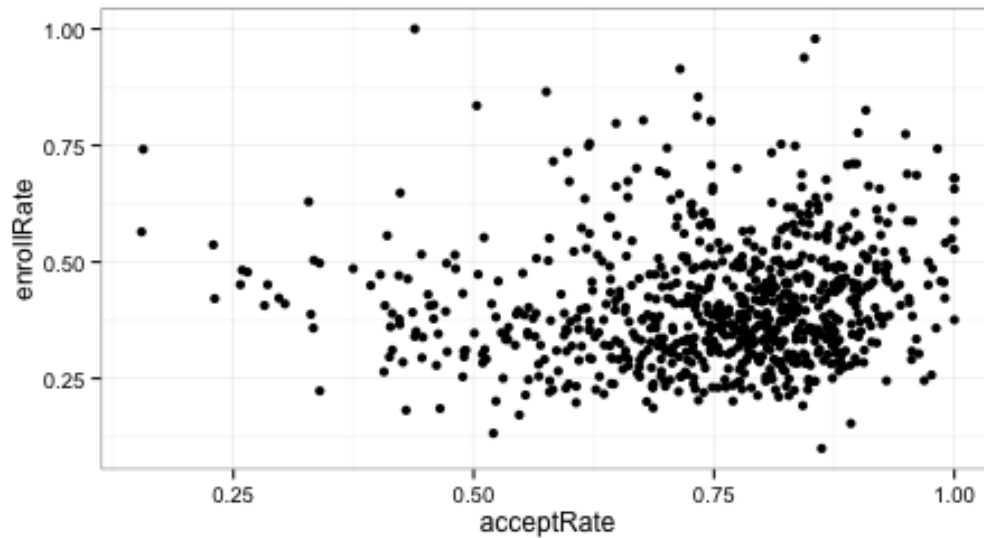
Part vi

I chose to explore acceptance and enrollment rates, with acceptance rate defined as the number of students accepted per application received (`college$Accept/college$Apps`), and enrollment rate defined as the number of students enrolled per applicant accepted (`college$Enroll/college$Accept`).



Somewhat surprisingly, there's little correlation between acceptance and enrollment rate.

```
> cor(college$acceptRate, college$enrollRate)
[1] 0.0824304
```

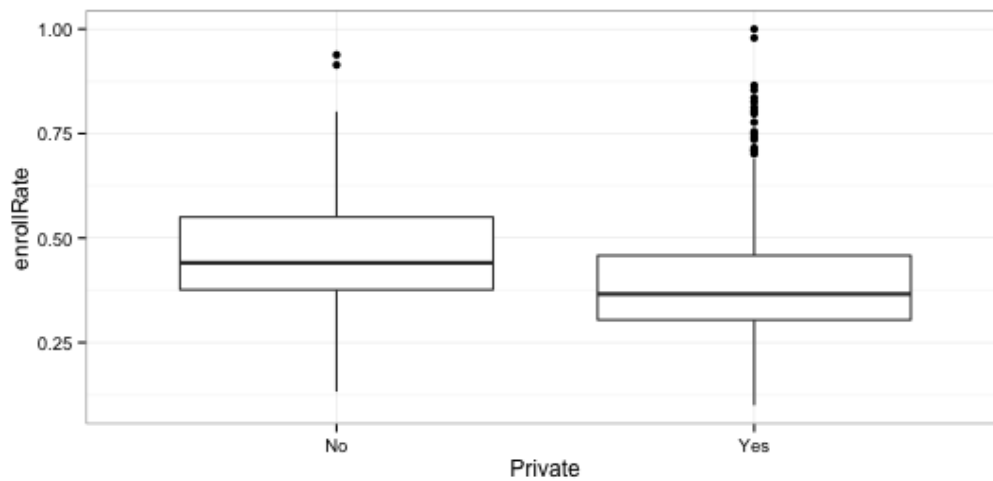
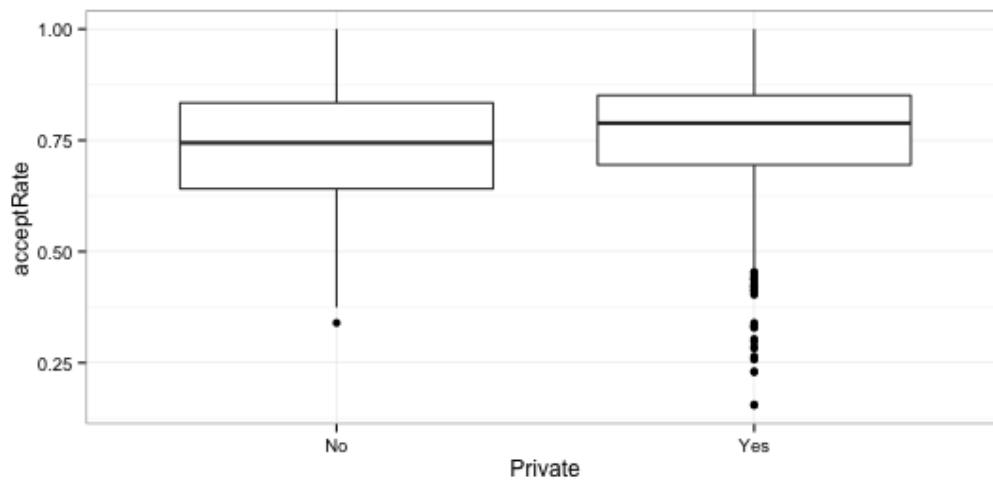


I also looked at summary statistics by public vs private schools.

```
> college %>%
+   group_by(Private) %>%
+   summarize(min.acceptRate=min(acceptRate),
+             median.acceptRate=median(acceptRate),
+             mean.acceptRate=mean(acceptRate),
+             max.acceptRate=max(acceptRate),
+             min.enrollRate=min(enrollRate),
+             median.enrollRate=median(enrollRate),
+             mean.enrollRate=mean(enrollRate),
+             max.enrollRate=max(enrollRate)) %>%
+   as.data.frame()
```

| | Private | min.acceptRate | median.acceptRate | mean.acceptRate | max.acceptRate | min.enrollRate |
|---|---------|----------------|-------------------|-----------------|----------------|----------------|
| 1 | No | 0.3397060 | 0.7443387 | 0.7265305 | 1 | 0.13242009 |
| 2 | Yes | 0.1544863 | 0.7885653 | 0.7545812 | 1 | 0.09975397 |

| | median.enrollRate | mean.enrollRate | max.enrollRate |
|---|-------------------|-----------------|----------------|
| 1 | 0.4405908 | 0.4620216 | 0.9382716 |
| 2 | 0.3660934 | 0.3932510 | 1.0000000 |



Problem 2: [James](#) 2.4, Exercise 9

Part a

Quantitative predictors: mpg, cylinders, displacement, horsepower, weight, acceleration, year

Qualitative predictors: origin, name

Part b

| | statistic | mpg | cylinders | displacement | horsepower | weight | acceleration | year | origin | name |
|---|-----------|------|-----------|--------------|------------|--------|--------------|------|--------|------|
| 1 | min | 9.0 | 3 | 68 | 46 | 1613 | 8.0 | 70 | NA | NA |
| 2 | max | 46.6 | 8 | 455 | 230 | 5140 | 24.8 | 82 | NA | NA |

Part c

| | statistic | mpg | cylinders | displacement | horsepower | weight | acceleration | year | origin |
|---|-----------|-----------|-----------|--------------|------------|-----------|--------------|-----------|--------|
| 1 | mean | 23.445918 | 5.471939 | 194.412 | 104.46939 | 2977.5842 | 15.541327 | 75.979592 | NA |
| 2 | sd | 7.805007 | 1.705783 | 104.644 | 38.49116 | 849.4026 | 2.758864 | 3.683737 | NA |

Part d

| | statistic | mpg | cylinders | displacement | horsepower | weight | acceleration | year | origin |
|---|-----------|-----------|-----------|--------------|------------|-----------|--------------|-----------|--------|
| 1 | min | 11.000000 | 3.000000 | 68.00000 | 46.00000 | 1649.0000 | 8.500000 | 70.000000 | NA |
| 2 | max | 46.600000 | 8.000000 | 455.00000 | 230.00000 | 4997.0000 | 24.800000 | 82.000000 | NA |
| 3 | mean | 24.404430 | 5.373418 | 187.24051 | 100.72152 | 2935.9715 | 15.726899 | 77.145570 | NA |
| 4 | sd | 7.867283 | 1.654179 | 99.67837 | 35.70885 | 811.3002 | 2.693721 | 3.106217 | NA |

Part e

MPG generally increased between 1970 and 1982.

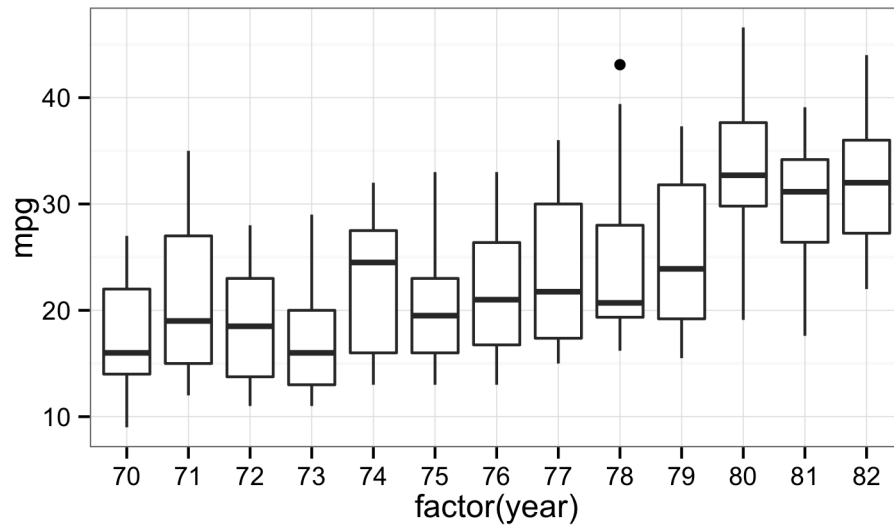


Figure 1: MPG vs year

As the number of cylinders increases, MPG generally decreases.

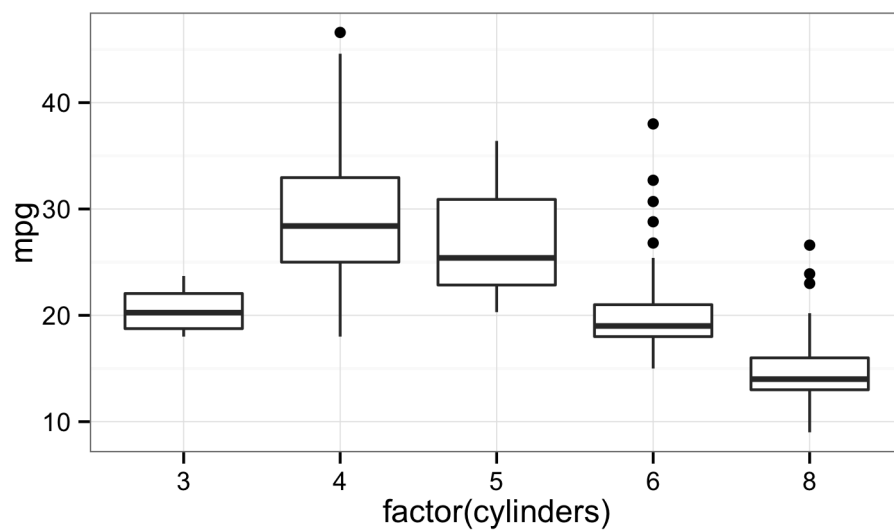
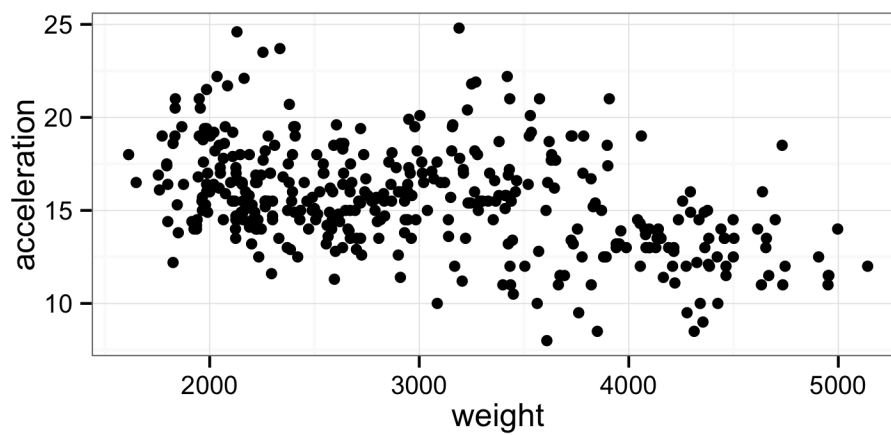
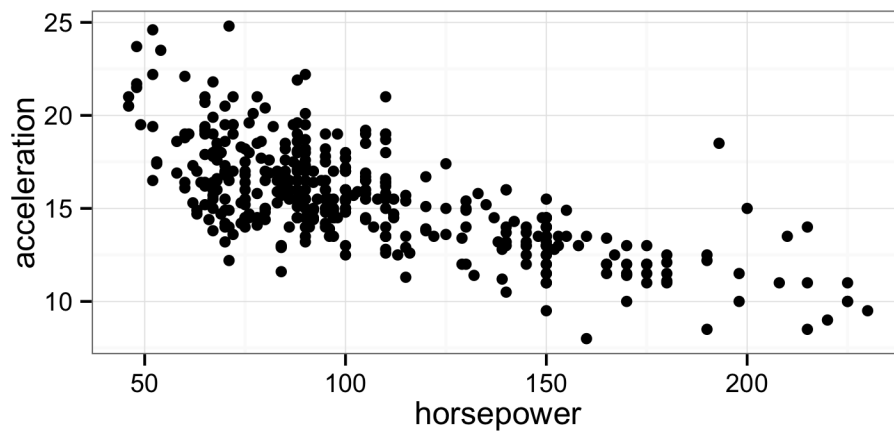


Figure 2: MPG vs number of cylinders

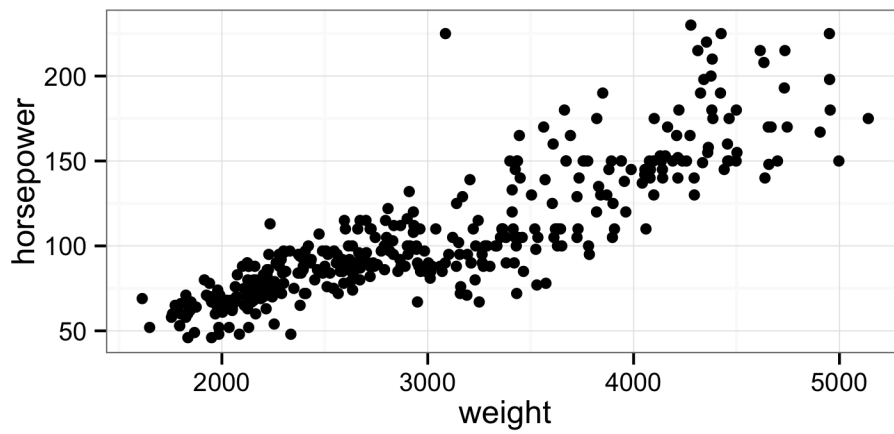
There isn't a particularly strong relationship between weight and acceleration (at least not one that is easily seen graphically).



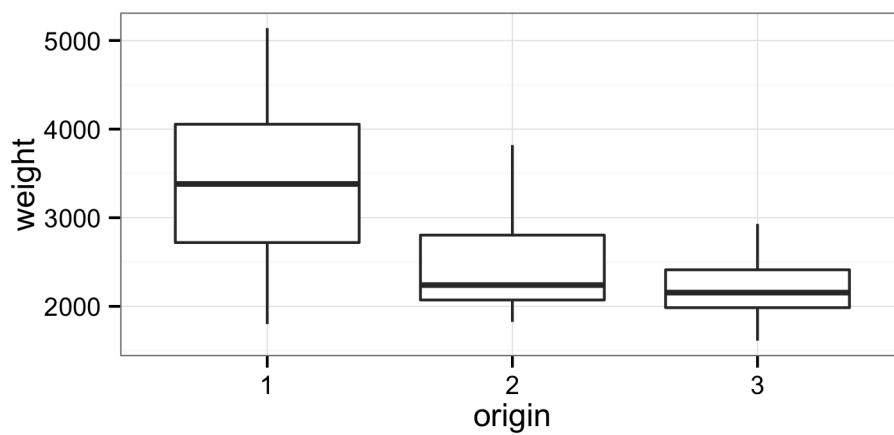
But there is a strong negative relationship between horsepower and acceleration.



And unsurprisingly, as the weight of a car increases, so does its horsepower.



Japanese cars (3) are usually lighter than American (1) and European (2) cars.



Japanese cars also have higher MPG.

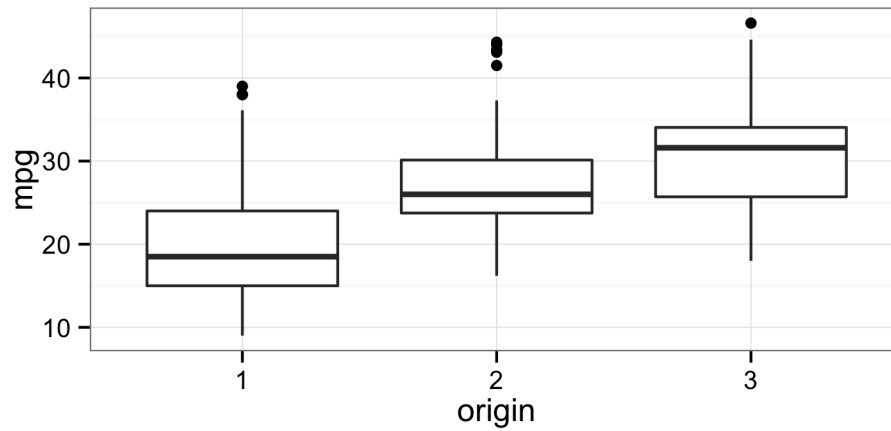


Figure 3: MPG by origin

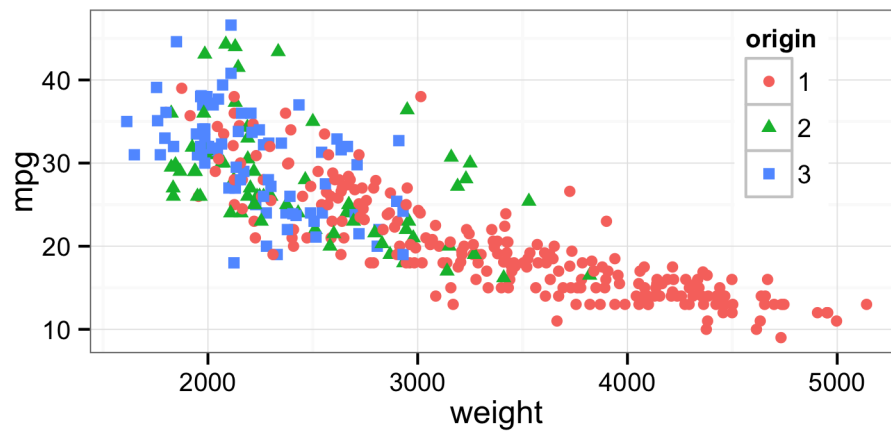


Figure 4: MPG vs weight, by origin

Part f

As shown in **Part e**, the year (Figure 1), number of cylinders (Figure 2), origin (Figure 3), and weight (Figure 4) of a car are all useful in predicting MPG. Displacement is also a useful predictor, but acceleration is not (see Figures 5 and 6 below).

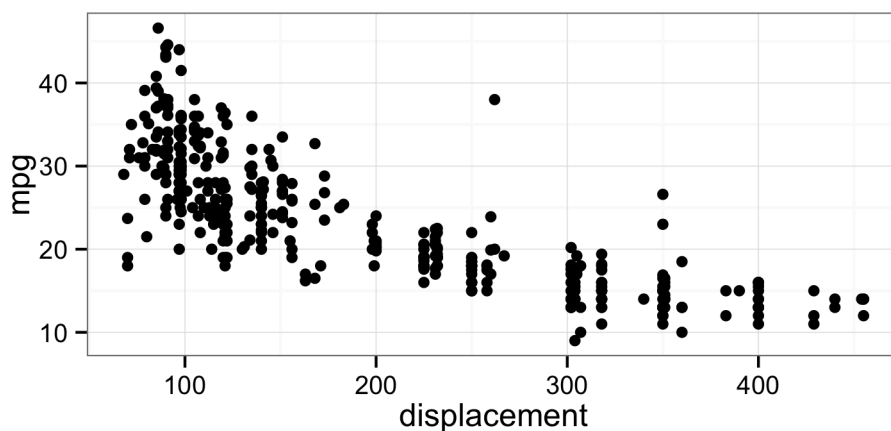


Figure 5: MPG vs displacemet

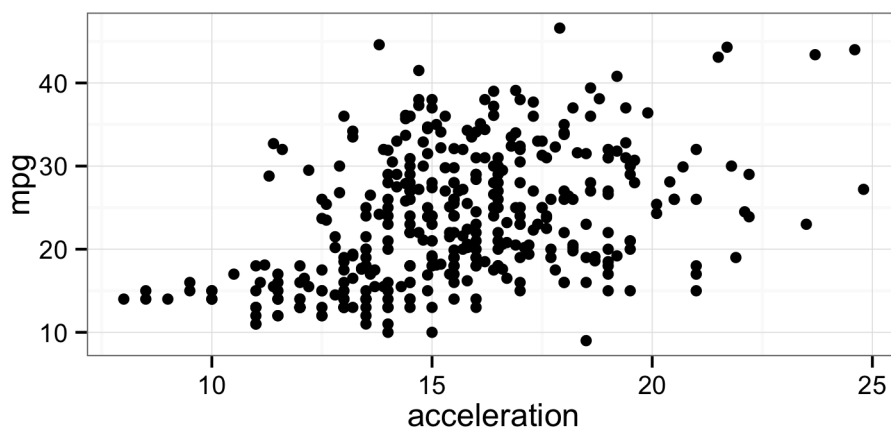


Figure 6: MPG vs acceleration

Problem 3: [James 2.4](#), Exercise 10

Part a

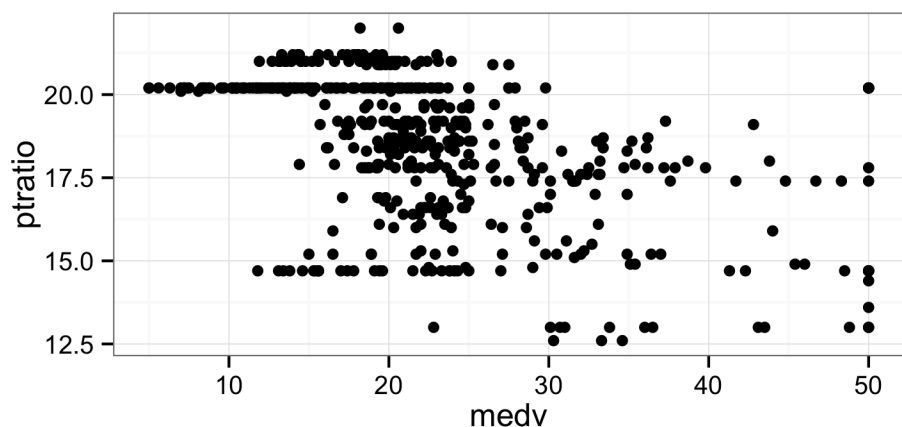
In the `Boston` dataset there are 506 rows and 14 columns. Each row represents a town in Boston. The column definitions, as written in the help file, are:

- `crim`: per capita crime rate by town.
- `zn`: proportion of residential land zoned for lots over 25,000 sq.ft.
- `indus`: proportion of non-retail business acres per town.
- `chas`: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
- `nox`: nitrogen oxides concentration (parts per 10 million).
- `rm`: average number of rooms per dwelling.

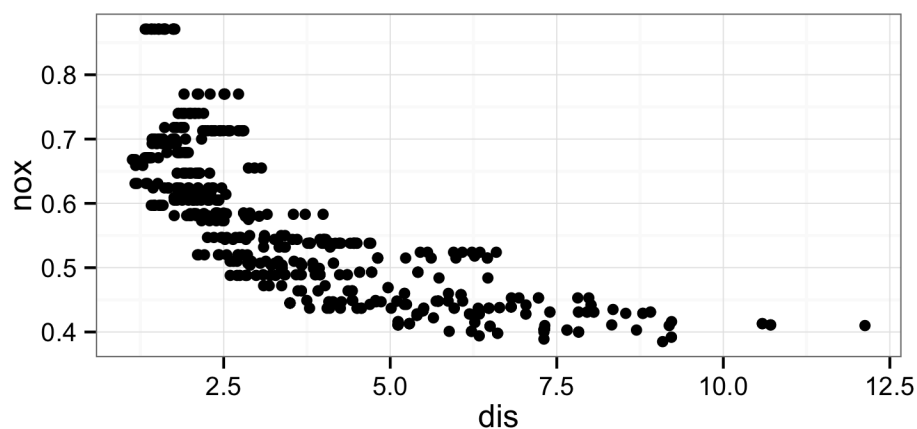
age: proportion of owner-occupied units built prior to 1940.
dis: weighted mean of distances to five Boston employment centres.
rad: index of accessibility to radial highways.
tax: full-value property-tax rate per \$10,000.
ptratio: pupil-teacher ratio by town.
black: $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town.
lstat: lower status of the population (percent).
medv: median value of owner-occupied homes in \$1000s.

Part b

Pupil-teacher ratio decreases slightly as home value increases.



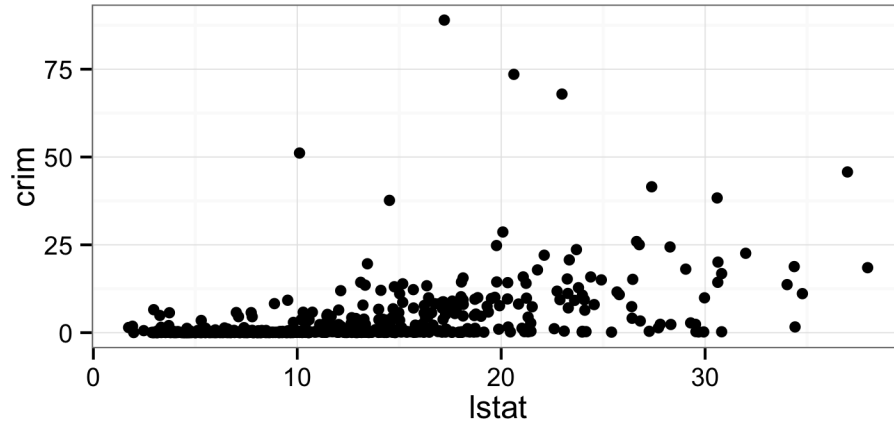
As distance from employment centers increases, nitrogen oxides concentration decreases.



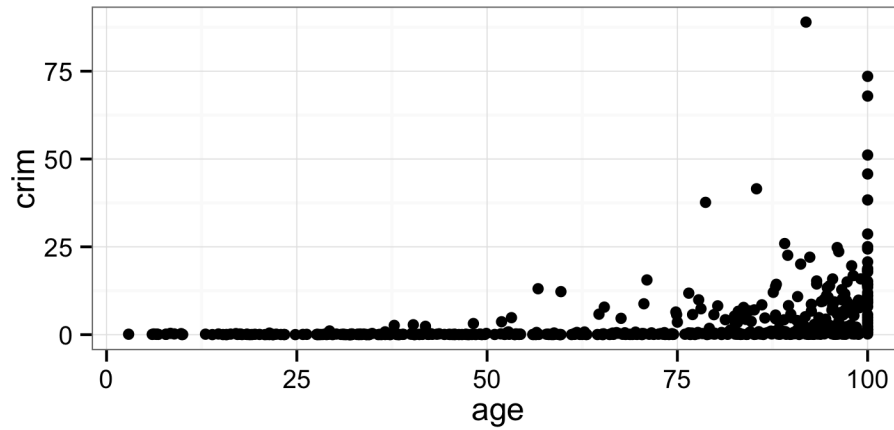
For plots incorporating crime rate (**crim**), see **Part c**.

Part c

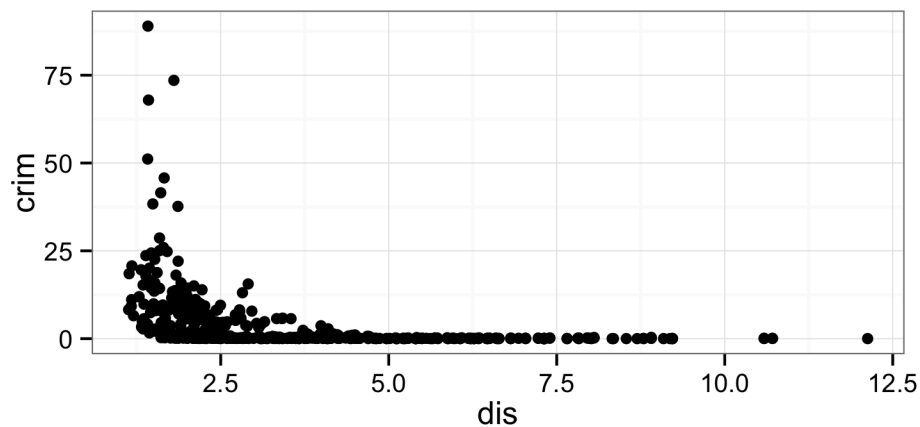
Crime rates are associated with many of the predictors. As "lower status of the population (percent)" increases, crime rates do too.



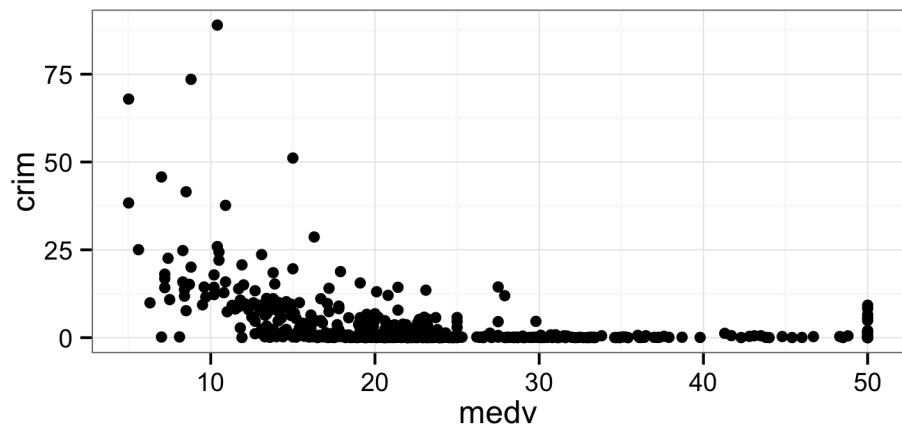
Similarly, towns with a higher proportion of buildings built before 1940 have higher crime rates as well.



Towns further from employment centers have lower crime rates.



And crime rates decrease as the median home value increases.



Part d

Yes, some suburbs of Boston appear to have particularly high crime rates. The suburbs in rows 381, 399, 401, 405, 406, 411, 414, 415, 418, 419, and 428 each have a per capita crime rate of more than 25. Crime rate ranges from 0.00632 to 88.97620.

There also appear to be some towns with incredibly high tax rates. There are 137 towns with full-value property-tax rate per \$10,000 of more than 600. Tax rate ranges from 187 to 711. A subset of the 137 towns are listed below:

```
[1] 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373
...
[121] 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493
```

It doesn't look like any towns have particularly high pupil-teacher ratios. Pupil-teacher ratio ranges from 12.6 to 22.0. 20.2 students/teacher appears to be a very popular ratio.

Range of each predictor:

| | stat | crim | zn | indus | chas | nox | rm | age | dis | rad | tax | ptratio | black | lstat | medv |
|---|------|----------|-----|-------|------|-------|-------|-------|---------|-----|-----|---------|--------|-------|------|
| 1 | min | 0.00632 | 0 | 0.46 | 0 | 0.385 | 3.561 | 2.9 | 1.1296 | 1 | 187 | 12.6 | 0.32 | 1.73 | 5 |
| 2 | max | 88.97620 | 100 | 27.74 | 1 | 0.871 | 8.780 | 100.0 | 12.1265 | 24 | 711 | 22.0 | 396.90 | 37.97 | 50 |

Part e

35 town in the dataset bound the Charles river

Part f

Among towns in the dataset, the median pupil-teacher ratio is 19.05.

Part g

The towns in row numbers 399 and 406 have the lowest median values of owner-occupied homes at \$5,000.

For those towns:

| | rowNum | crim | zn | indus | chas | nox | rm | age | dis | rad | tax | ptratio | black | lstat | medv |
|---|--------|---------|----|-------|------|-------|-------|-----|--------|-----|-----|---------|--------|-------|------|
| 1 | 399 | 38.3518 | 0 | 18.1 | 0 | 0.693 | 5.453 | 100 | 1.4896 | 24 | 666 | 20.2 | 396.90 | 30.59 | 5 |
| 2 | 406 | 67.9208 | 0 | 18.1 | 0 | 0.693 | 5.683 | 100 | 1.4254 | 24 | 666 | 20.2 | 384.97 | 22.98 | 5 |

Crime rates here are on the upper end of the range, and the towns have an average number of rooms per dwelling. Neither town borders the Charles river, and 100% of owner-occupied buildings in both towns were built before 1940.

Part h

There are 64 towns that average more than 7 rooms per dwelling, and 13 towns that average more than 8 rooms per dwelling.

Towns with more than 8 rooms per dwelling:

| | rowNum | crim | zn | indus | chas | nox | rm | age | dis | rad | tax | ptratio | black | lstat | medv |
|----|--------|---------|----|-------|------|--------|-------|------|--------|-----|-----|---------|--------|-------|------|
| 1 | 98 | 0.12083 | 0 | 2.89 | 0 | 0.4450 | 8.069 | 76.0 | 3.4952 | 2 | 276 | 18.0 | 396.90 | 4.21 | 38.7 |
| 2 | 164 | 1.51902 | 0 | 19.58 | 1 | 0.6050 | 8.375 | 93.9 | 2.1620 | 5 | 403 | 14.7 | 388.45 | 3.32 | 50.0 |
| 3 | 205 | 0.02009 | 95 | 2.68 | 0 | 0.4161 | 8.034 | 31.9 | 5.1180 | 4 | 224 | 14.7 | 390.55 | 2.88 | 50.0 |
| 4 | 225 | 0.31533 | 0 | 6.20 | 0 | 0.5040 | 8.266 | 78.3 | 2.8944 | 8 | 307 | 17.4 | 385.05 | 4.14 | 44.8 |
| 5 | 226 | 0.52693 | 0 | 6.20 | 0 | 0.5040 | 8.725 | 83.0 | 2.8944 | 8 | 307 | 17.4 | 382.00 | 4.63 | 50.0 |
| 6 | 227 | 0.38214 | 0 | 6.20 | 0 | 0.5040 | 8.040 | 86.5 | 3.2157 | 8 | 307 | 17.4 | 387.38 | 3.13 | 37.6 |
| 7 | 233 | 0.57529 | 0 | 6.20 | 0 | 0.5070 | 8.337 | 73.3 | 3.8384 | 8 | 307 | 17.4 | 385.91 | 2.47 | 41.7 |
| 8 | 234 | 0.33147 | 0 | 6.20 | 0 | 0.5070 | 8.247 | 70.4 | 3.6519 | 8 | 307 | 17.4 | 378.95 | 3.95 | 48.3 |
| 9 | 254 | 0.36894 | 22 | 5.86 | 0 | 0.4310 | 8.259 | 8.4 | 8.9067 | 7 | 330 | 19.1 | 396.90 | 3.54 | 42.8 |
| 10 | 258 | 0.61154 | 20 | 3.97 | 0 | 0.6470 | 8.704 | 86.9 | 1.8010 | 5 | 264 | 13.0 | 389.70 | 5.12 | 50.0 |
| 11 | 263 | 0.52014 | 20 | 3.97 | 0 | 0.6470 | 8.398 | 91.5 | 2.2885 | 5 | 264 | 13.0 | 386.86 | 5.91 | 48.8 |
| 12 | 268 | 0.57834 | 20 | 3.97 | 0 | 0.5750 | 8.297 | 67.0 | 2.4216 | 5 | 264 | 13.0 | 384.54 | 7.44 | 50.0 |
| 13 | 365 | 3.47428 | 0 | 18.10 | 1 | 0.7180 | 8.780 | 82.9 | 1.9047 | 24 | 666 | 20.2 | 354.55 | 5.29 | 21.9 |

These towns all have very low crime rates and have lower values for `lstat`. They generally have `medv` values towards the higher end of the range.

Problem 4: Yale Faces B

Part a

The *CroppedYale/yaleB01/yaleB01p00A - 005E + 10.pgm* image is of class `pixmapGrey` (part of the `pixmap` package). The original image is made up of 32,256 pixels (192 x 168).