Homework 2

Julia Van Dyke

January 11, 2015

7

a

observation1

3

observation2

2

observation3

```
sqrt(1^2+3^2)
```

```
## [1] 3.162278
```

observation4

```
sqrt(1^2+2^2)
```

```
## [1] 2.236068
```

observation5

```
sqrt((-1)^2+1^2)
```

```
## [1] 1.414214
```

observation6

```
sqrt(1^2+1^2+1^2)
```

```
## [1] 1.732051
```

b

Green, the nearest data point is green

c

Red, two out of the nearest three points are red

d

Large, because then the data will better fit the curve

9

a

quantitative: mpg, cylinders, displacement, horsepower, weight, acceleration, year

qualitative: origin, name

b

```
library(ISLR)
range(mpg,data=Auto)
```

```
## [1]  9.0 46.6
```

```
range(cylinders,data=Auto)
```

```
## [1] 3 8
```

```
range(displacement,data=Auto)
```

```
## [1]  68 455
```

```
range(horsepower,data=Auto)
```

```
## [1]  46 230
```

```
range(weight,data=Auto)
```

```
## [1] 1613 5140
```

```
range(acceleration,data=Auto)
```

```
## [1]  8.0 24.8
```

```
range(year,data=Auto)
```

```
## [1] 70 82
```

c

```
mean(mpg,data=Auto)
```

```
## [1] 23.44592
```

```
sd(mpg,data=Auto)
```

```
## [1] 7.805007
```

```
mean(cylinders,data=Auto)
```

```
## [1] 5.471939
```

```
sd(cylinders,data=Auto)
```

```
## [1] 1.705783
```

```
mean(displacement,data=Auto)
```

```
## [1] 194.412
```

```
sd(displacement,data=Auto)
```

```
## [1] 104.644
```

```
mean(horsepower,data=Auto)
```

```
## [1] 104.4694
```

```
sd(horsepower,data=Auto)
```

```
## [1] 38.49116
```

```
mean(weight,data=Auto)
```

```
## [1] 2977.584
```

```
sd(weight,data=Auto)
```

```
## [1] 849.4026
```

```
mean(acceleration,data=Auto)
```

```
## [1] 15.54133
```

```
sd(acceleration,data=Auto)
```

```
## [1] 2.758864
```

```
mean(year,data=Auto)
```

```
## [1] 75.97959
```

```
sd(year,data=Auto)
```

```
## [1] 3.683737
```

d

```
Autominus <- Auto[-c(10:85),]
range(mpg,data=Autominus)
```

```
## [1] 11.0 46.6
```

```
mean(mpg,data=Autominus)
```

```
## [1] 24.40443
```

```r
sd(mpg,data=Autominus)
```

```
## [1] 7.867283
```

```r
range(cylinders,data=Autominus)
```

```
## [1] 3 8
```

```r
mean(cylinders,data=Autominus)
```

```
## [1] 5.373418
```

```r
sd(cylinders,data=Autominus)
```

```
## [1] 1.654179
```

```r
range(displacement,data=Autominus)
```

```
## [1]  68 455
```

```r
mean(displacement,data=Autominus)
```

```
## [1] 187.2405
```

```r
sd(displacement,data=Autominus)
```

```
## [1] 99.67837
```

```r
range(horsepower,data=Autominus)
```

```
## [1]  46 230
```

```r
mean(horsepower,data=Autominus)
```

```
## [1] 100.7215
```

```r
sd(horsepower,data=Autominus)
```

```
## [1] 35.70885
```

```r
range(weight,data=Autominus)
```

```
## [1] 1649 4997
```

```
mean(weight,data=Autominus)
```

```
## [1] 2935.972
```

```
sd(weight,data=Autominus)
```

```
## [1] 811.3002
```

```
range(acceleration,data=Autominus)
```

```
## [1]  8.5 24.8
```

```
mean(acceleration,data=Autominus)
```

```
## [1] 15.7269
```

```
sd(acceleration,data=Autominus)
```

```
## [1] 2.693721
```

```
range(year,data=Autominus)
```

```
## [1] 70 82
```

```
mean(year,data=Autominus)
```

```
## [1] 77.14557
```

```
sd(year,data=Autominus)
```

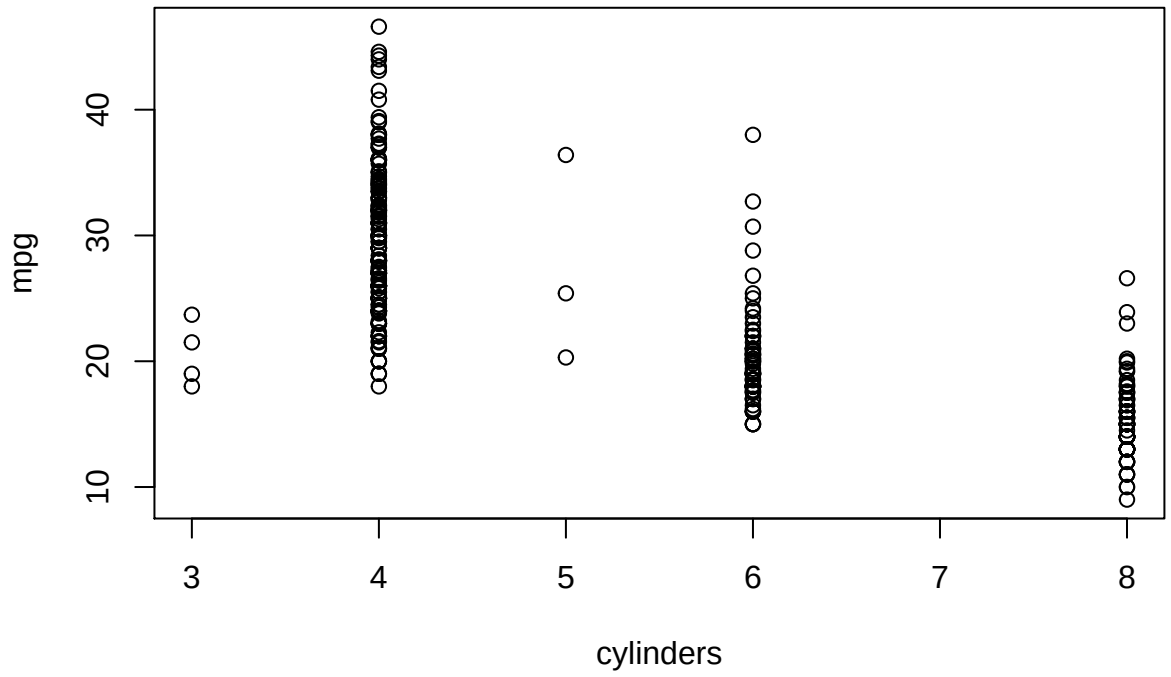
```
## [1] 3.106217
```

e

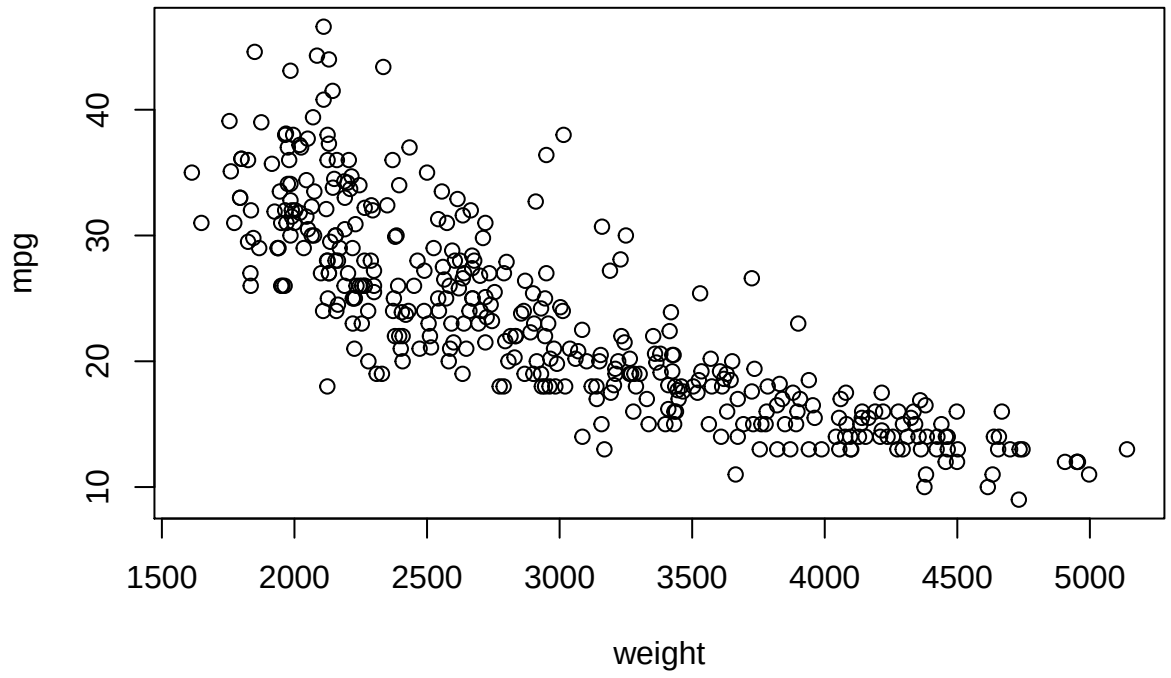
```
attach(Auto)
```

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

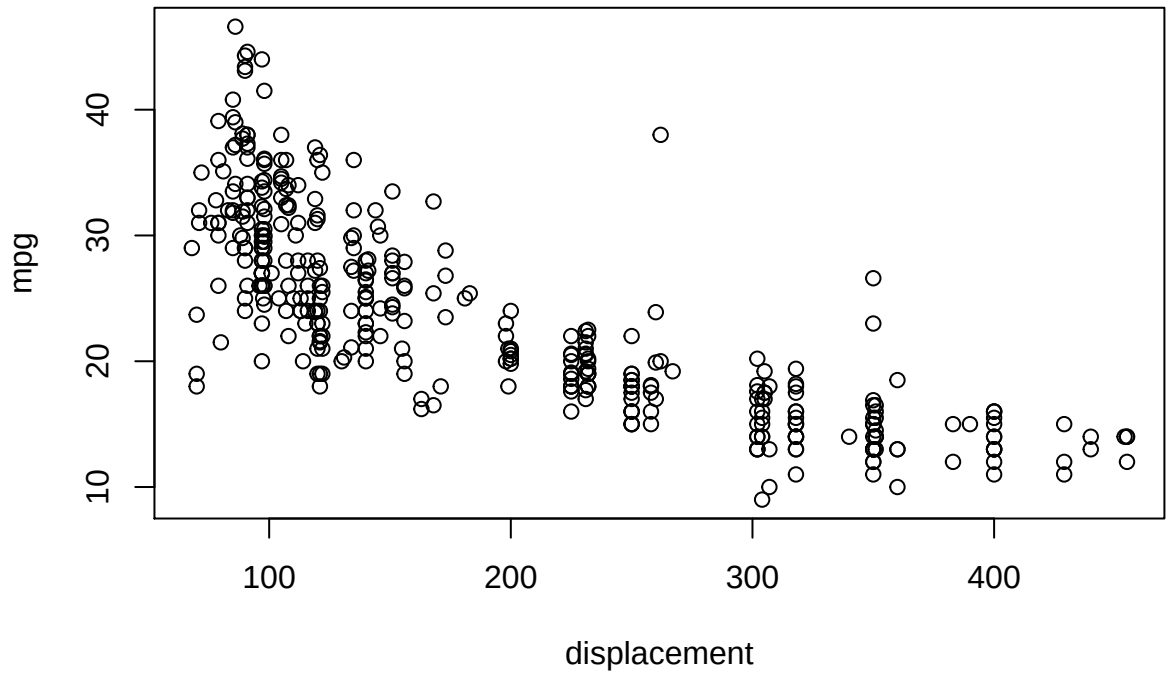
```
plot(cylinders,mpg)
```



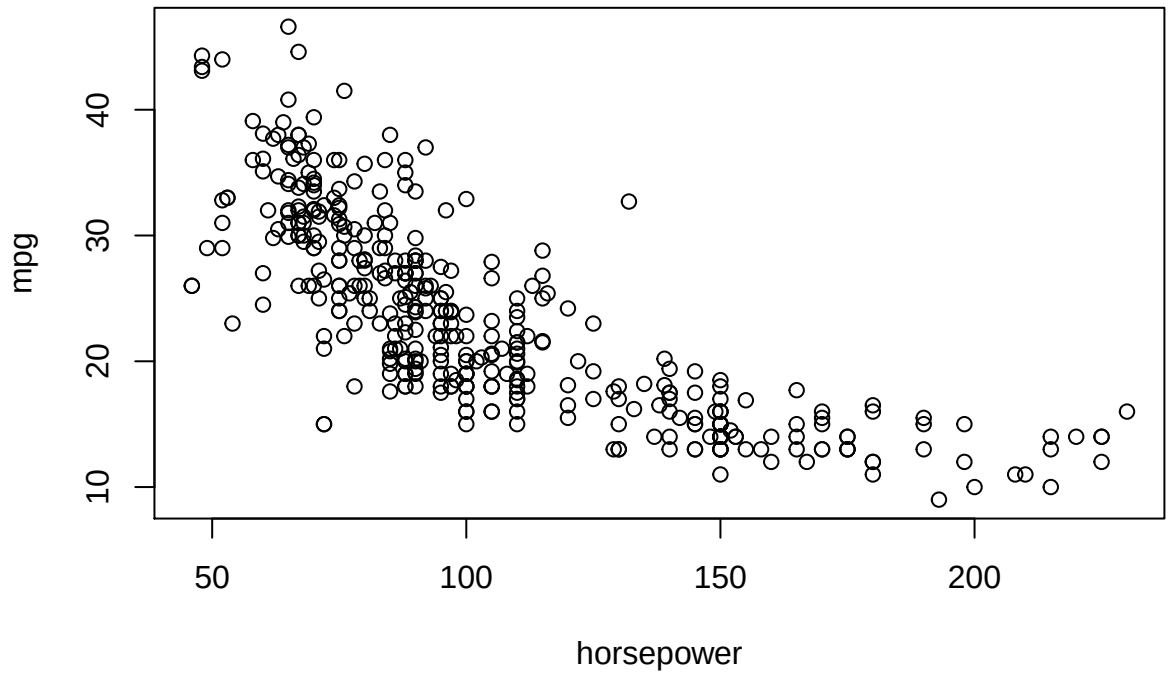

```
plot(weight,mpg)
```



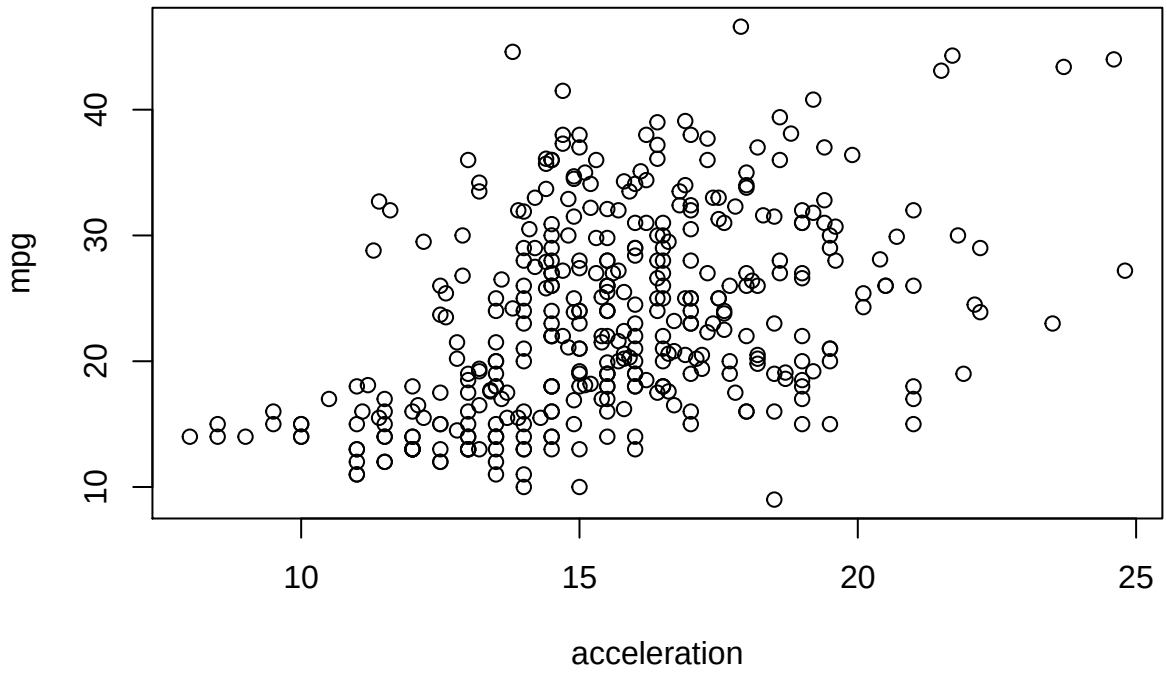

```
plot(displacement,mpg)
```



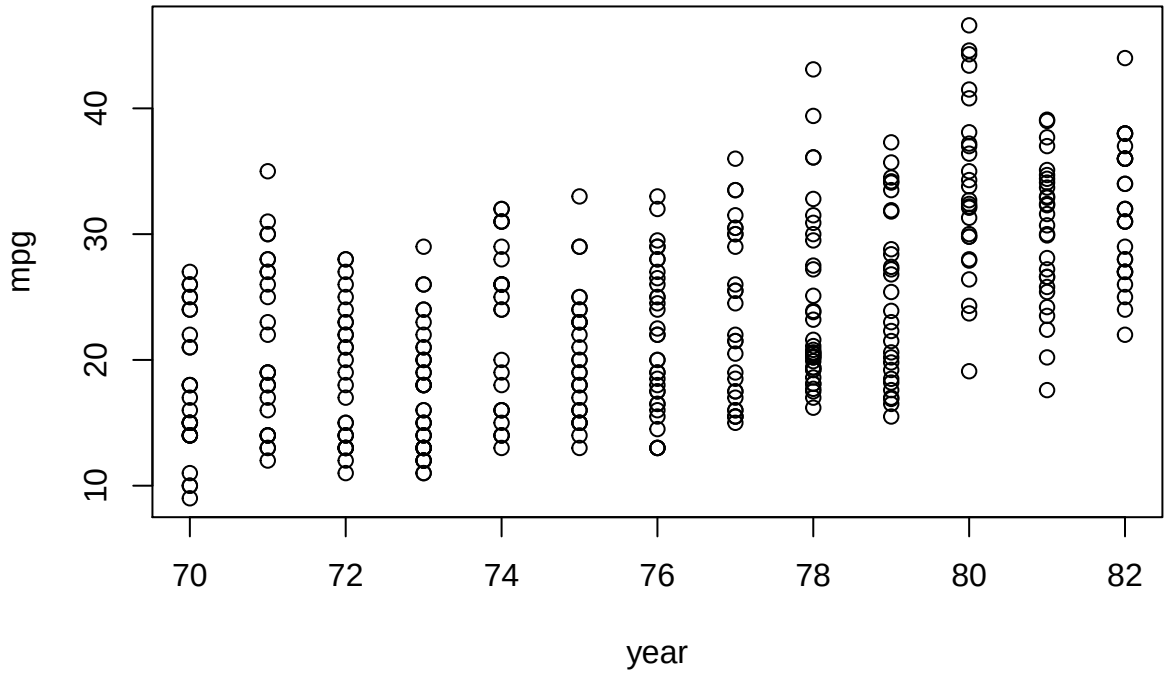

```
plot(horsepower,mpg)
```




```
plot(acceleration,mpg)
```




```
plot(year,mpg)
```



#There is no relationship between mpg and year, acceleration, or cylinders. There is correlation between mpg and weight, displacement, and horsepower. #***f*** #Weight, displacement, and horsepower could be used to predict mpg.

10

a

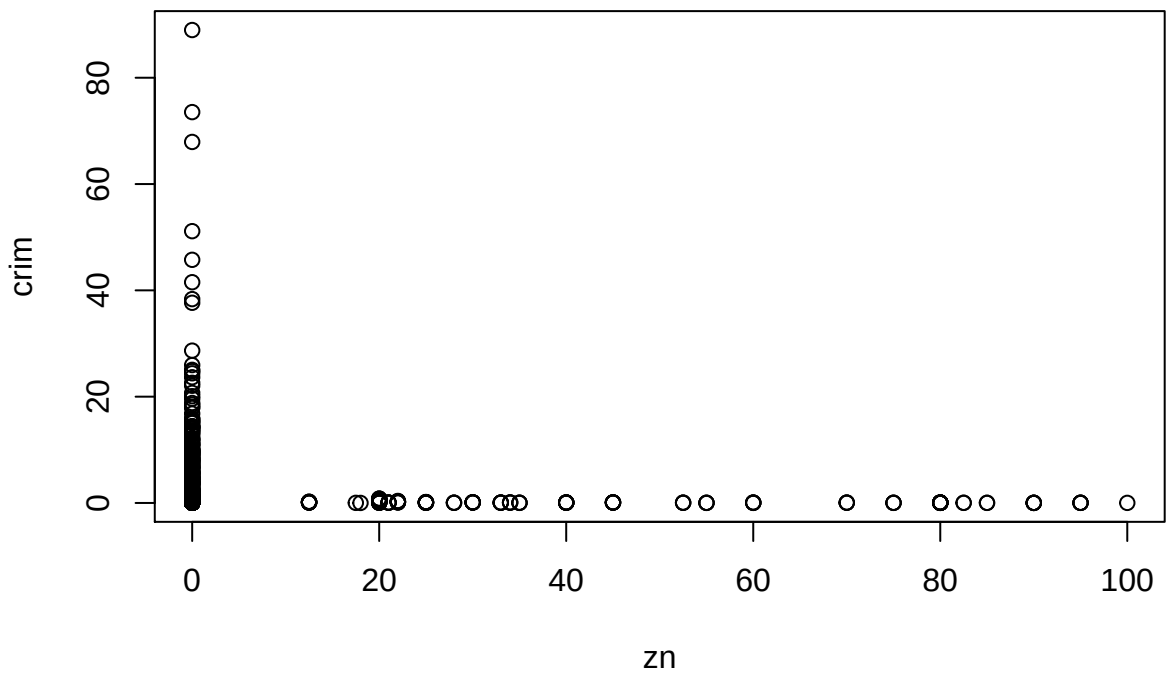
506 rows and 14 columns. rows are observations, columns are variables.

b

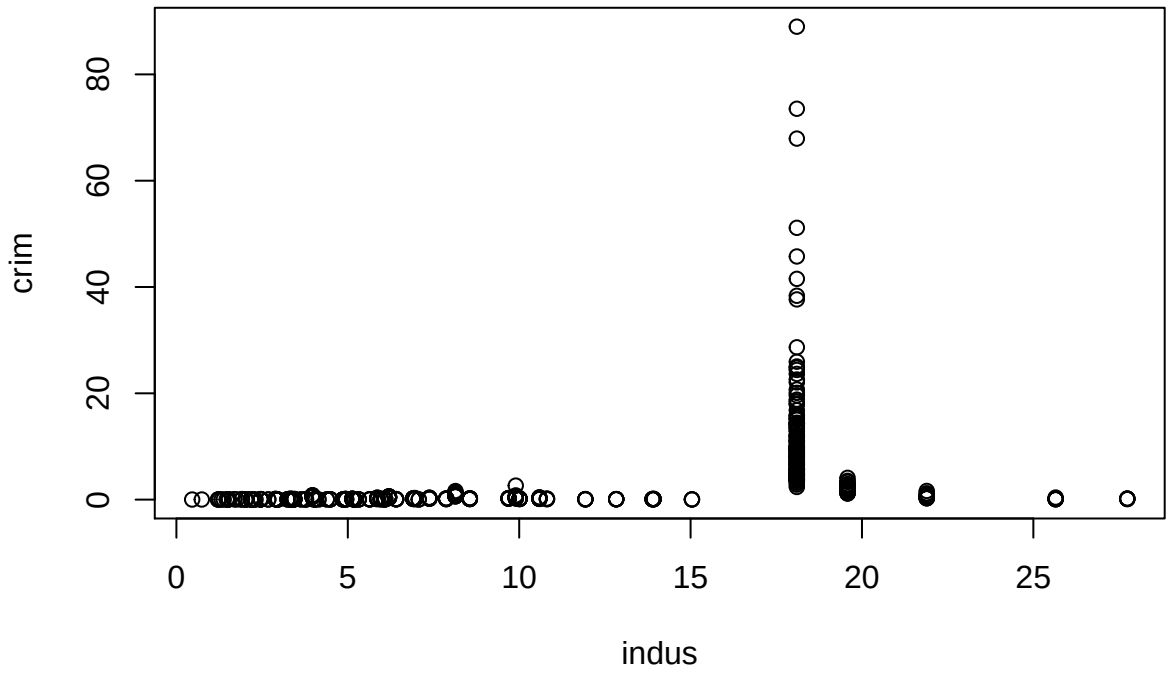
```
library(MASS)
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

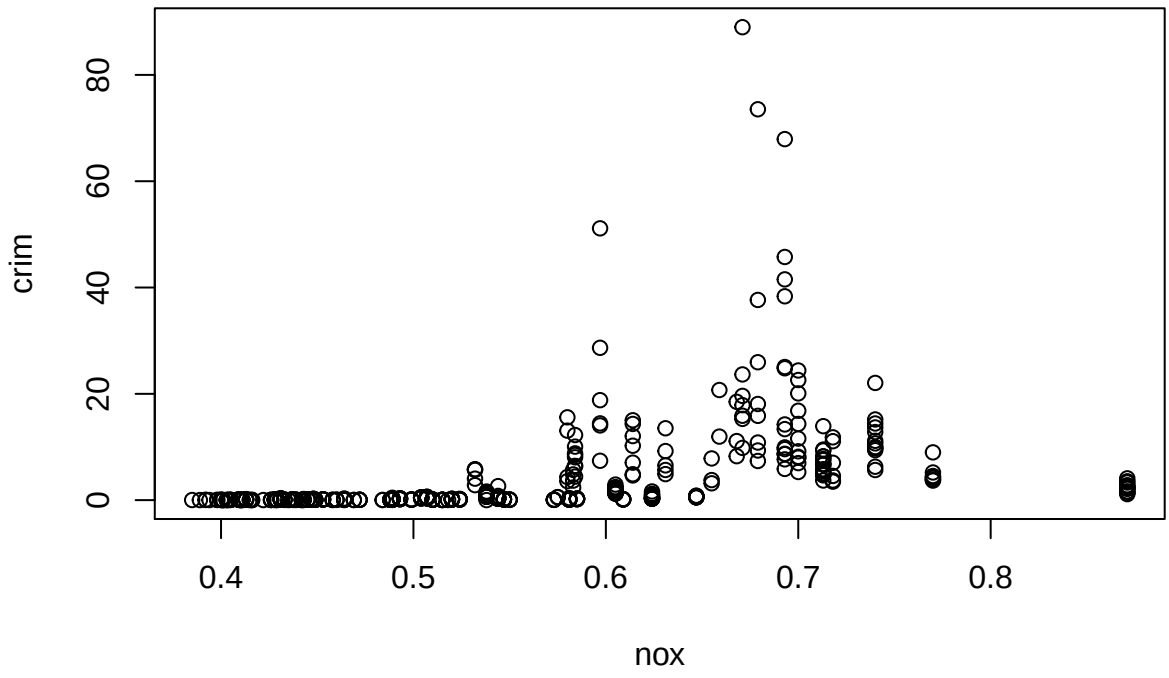
```
attach(Boston)
plot(zn,crim)
```



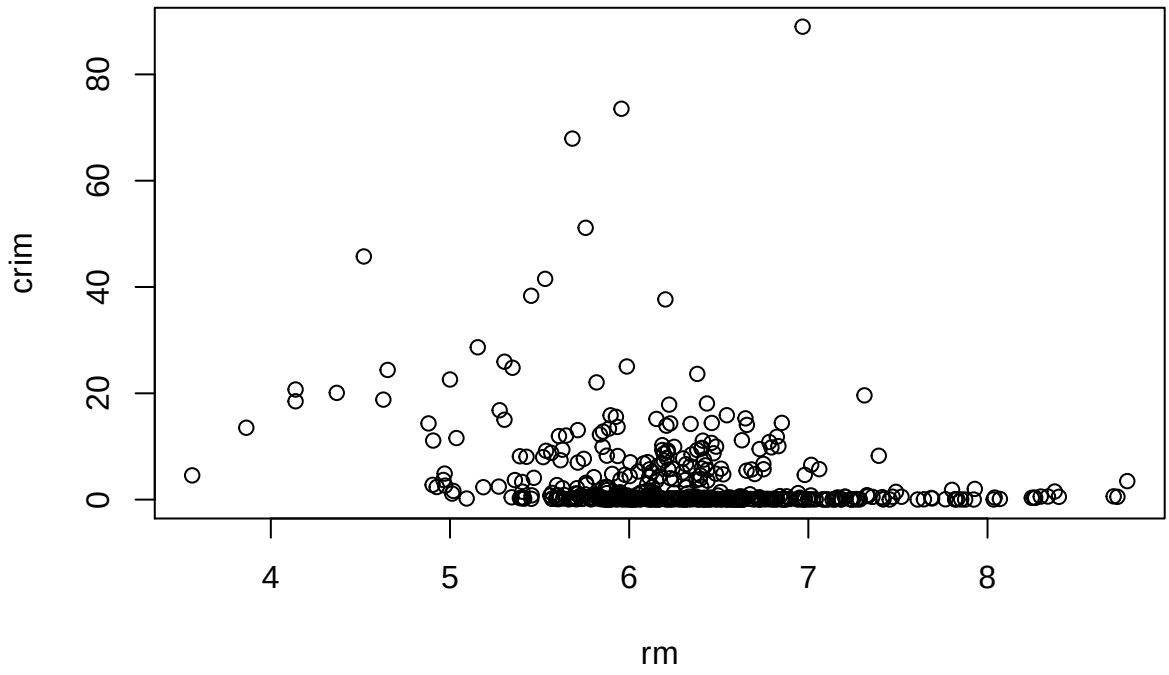

```
plot(indus,crim)
```



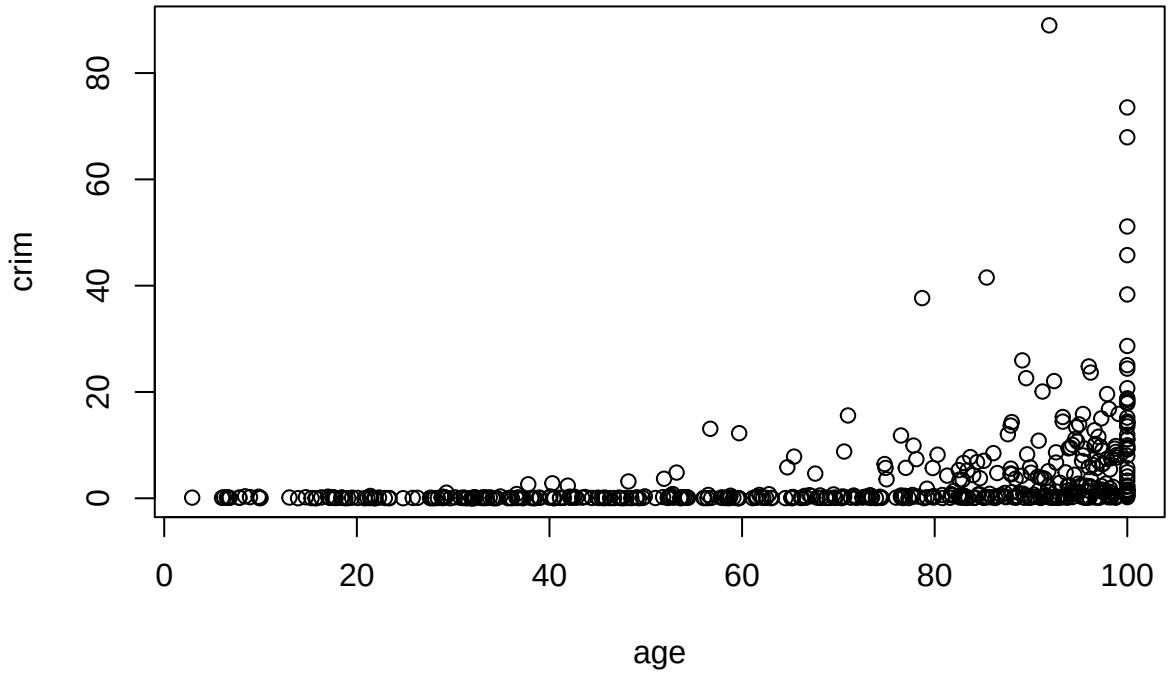

```
plot(nox,crim)
```



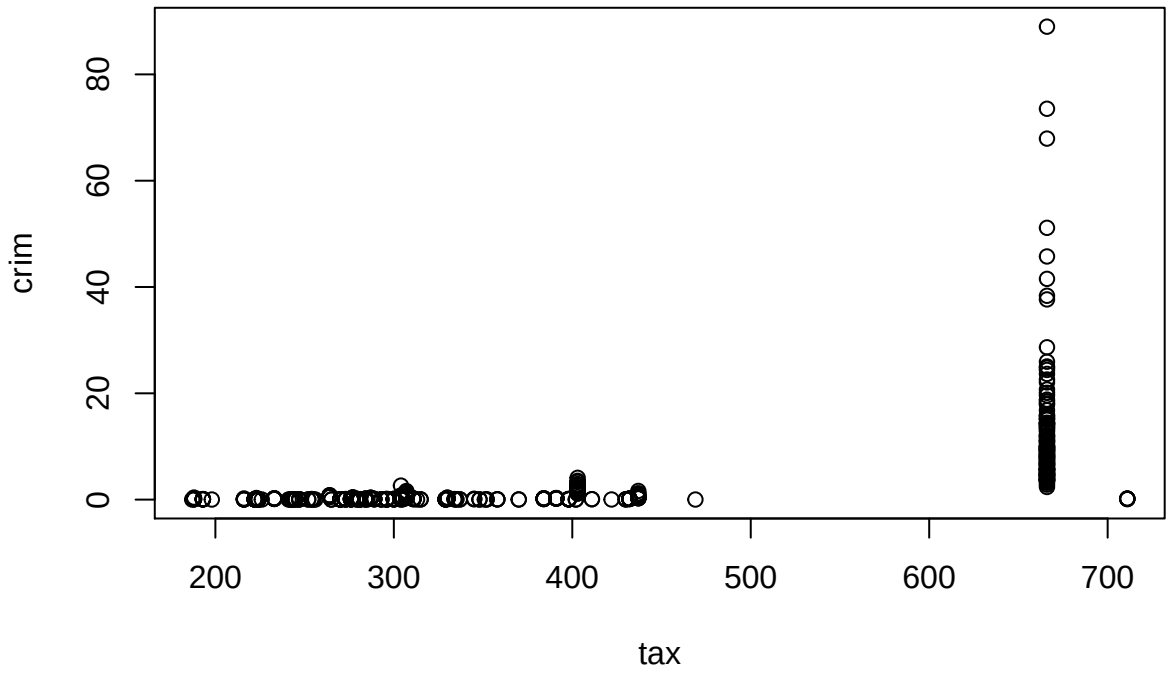

```
plot(rm,crim)
```



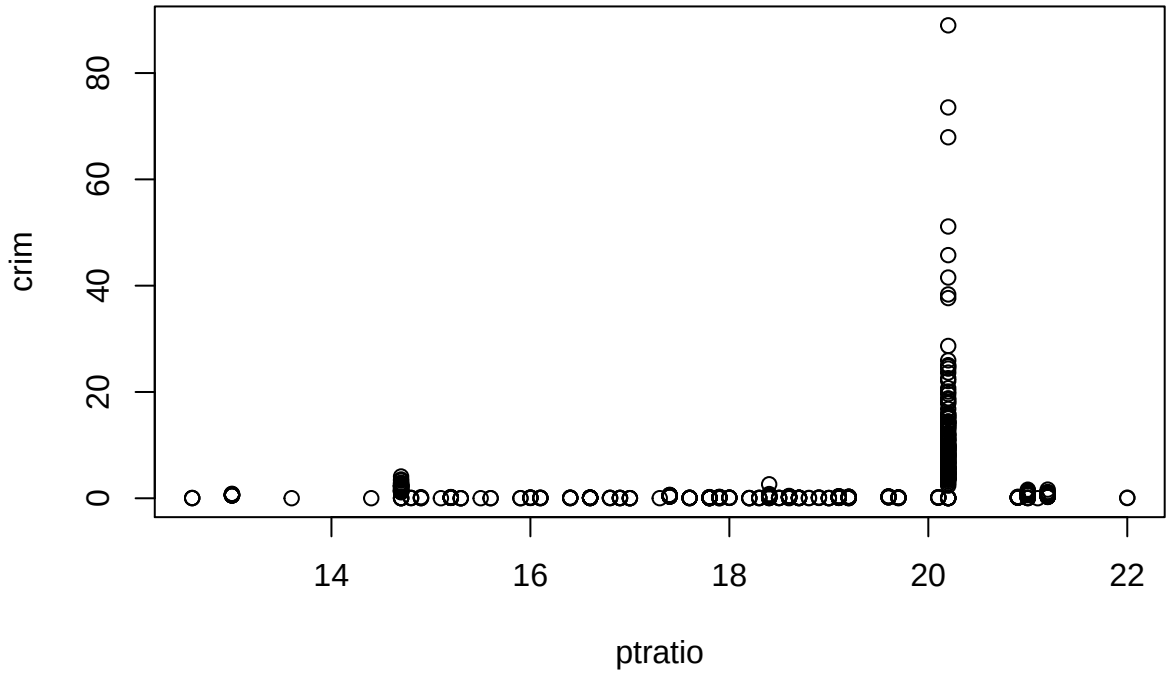

```
plot(age,crim)
```



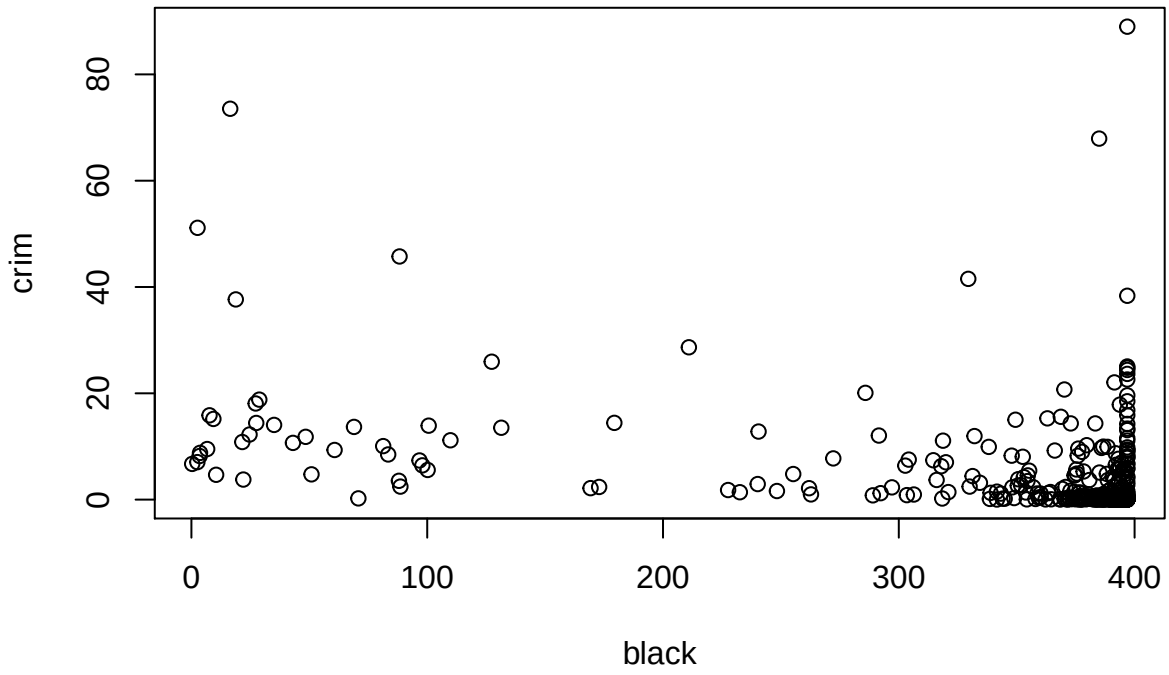

```
plot(tax,crim)
```



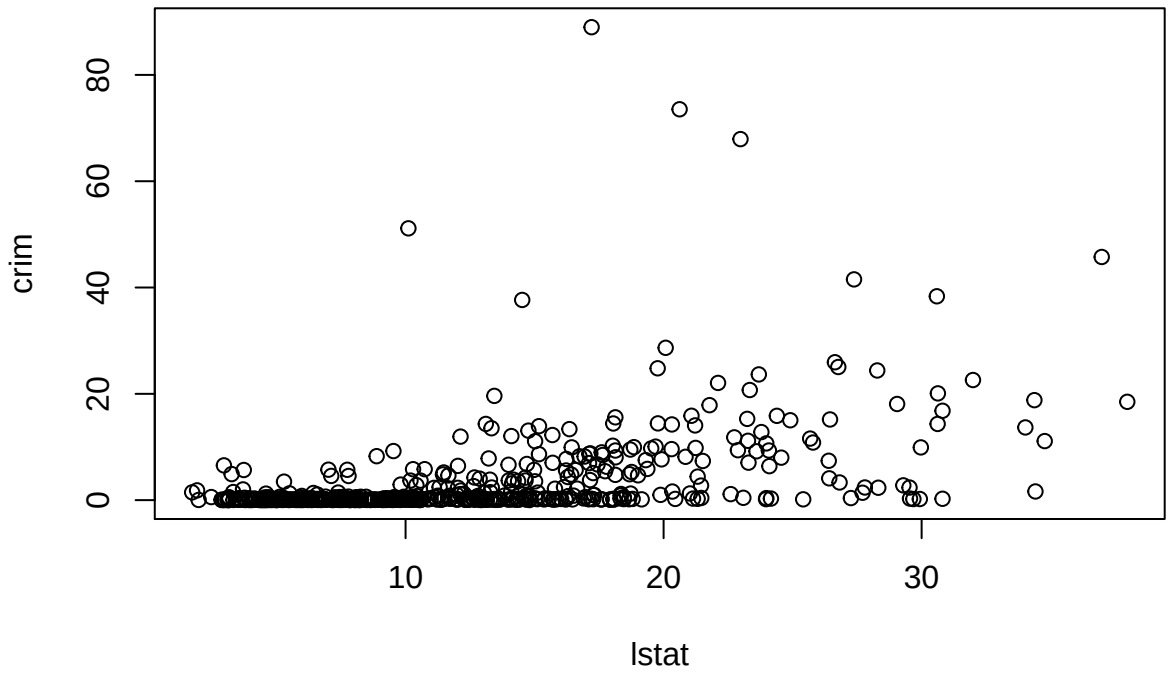

```
plot(ptratio,crim)
```



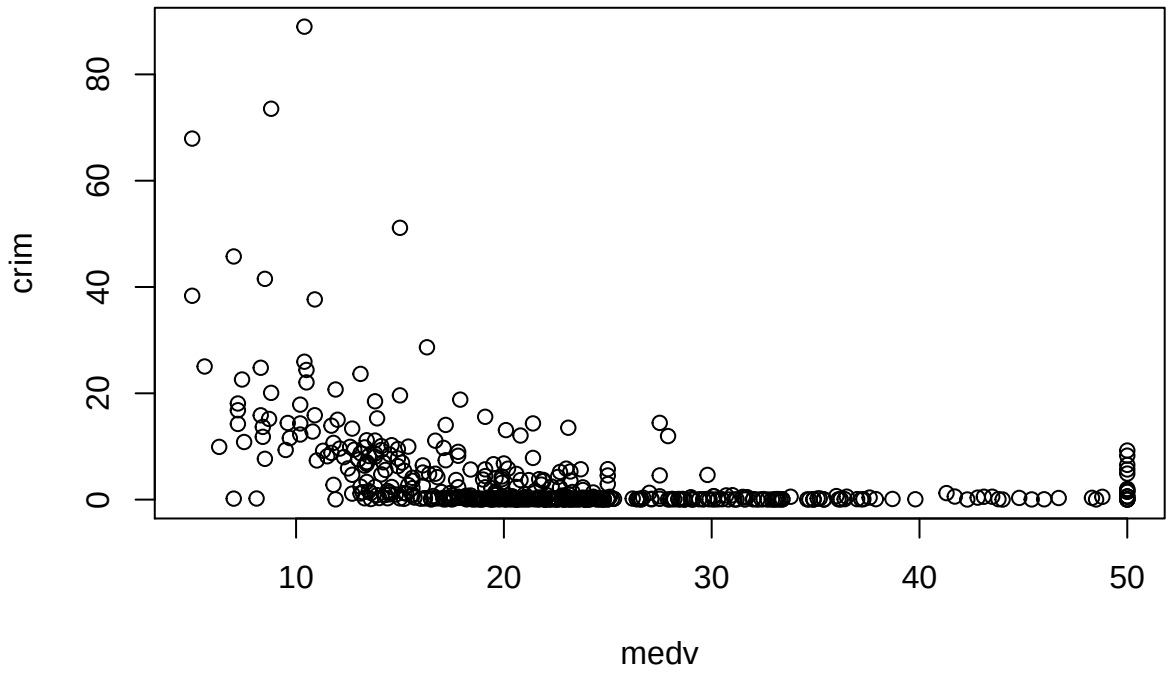

```
plot(black,crim)
```




```
plot(lstat,crim)
```




```
plot(medv,crim)
```



#most of the plots contain several outliers. Nox, rm, age, and lstat seem to have most correlation with crim. #**c** #nox-positive correlation, rm-positive, age-positive, lstat-positive #**d**

```
range(crim)
```

```
## [1]  0.00632 88.97620
```

```
range(tax)
```

```
## [1] 187 711
```

```
range(ptratio)
```

```
## [1] 12.6 22.0
```

crim and tax have very wide ranges, while ptratio has a smaller range

e

```
sum(chas)
```

```
## [1] 35
```

f

```
median(ptratio)
```

```
## [1] 19.05
```

g

```
min(medv)
```

```
## [1] 5
```

h

```
sum(rm>7)
```

```
## [1] 64
```

```
sum(rm>8)
```

```
## [1] 13
```