Class Activity 411

Rickey Huang

4/11/2022

```
data <- read.csv(file = "houses.csv")
head(data)
```

```
##    Price Living.Area Baths Bedrooms Fireplace Acres Age
## 1 142212        1982   1.0        3         N  2.00 133
## 2 134865        1676   1.5        3         Y  0.38  14
## 3 118007        1694   2.0        3         Y  0.96  15
## 4 138297        1800   1.0        2         Y  0.48  49
## 5 129470        2088   1.0        3         Y  1.84  29
## 6 206512        1456   2.0        3         N  0.98  10
```

Question 1

Part a

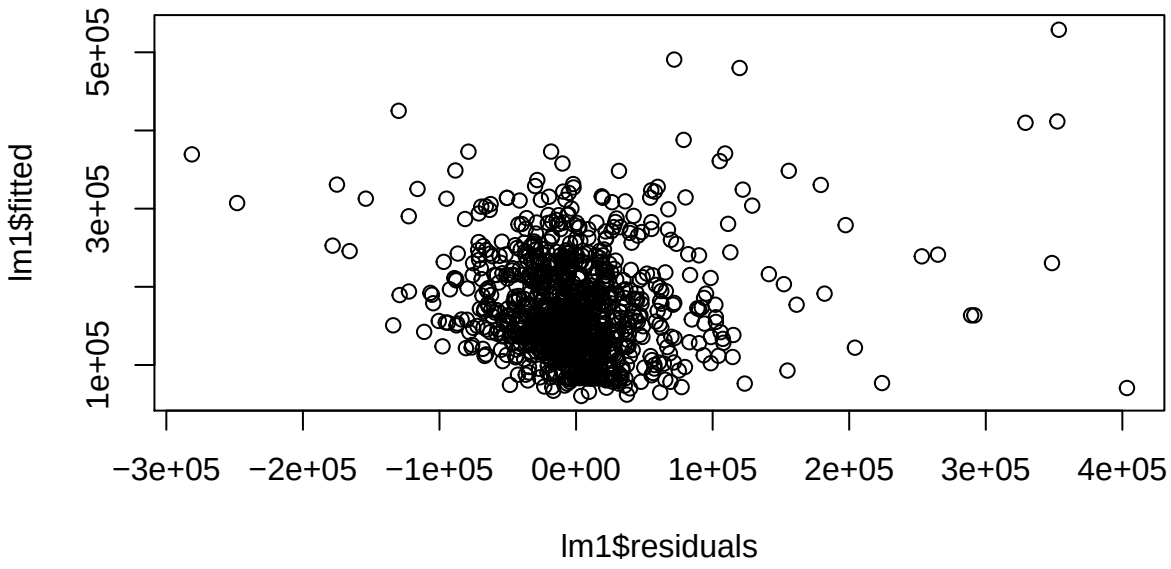
Fit a model Price ~ Living.Area

```
lm1 <- lm(Price~Living.Area, data = data)
```

Part b

Plot ehat vs. fitted

```
plot(lm1$residuals,lm1$fitted)
```



Part c

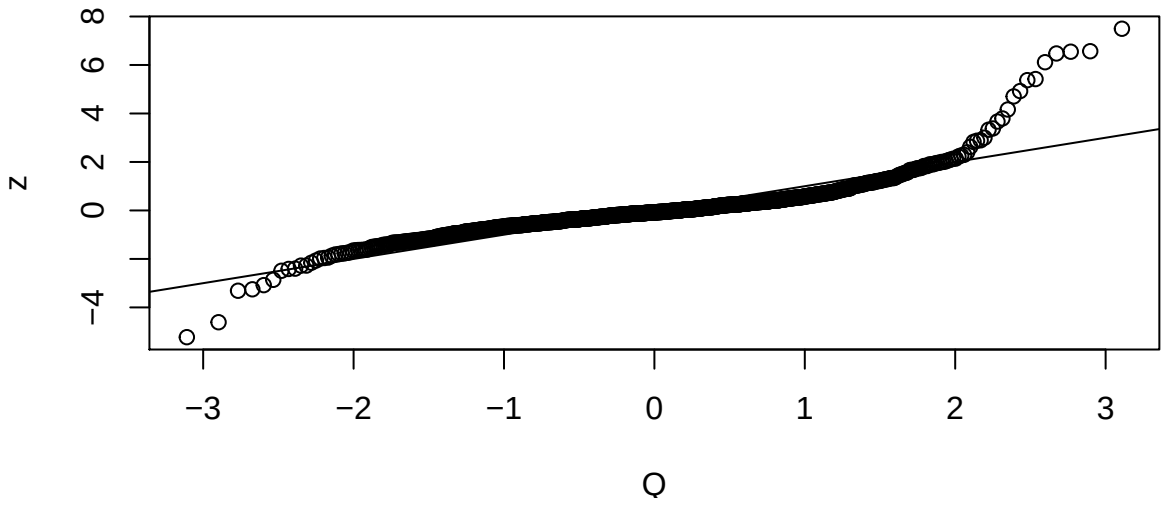
Make a residual histogram

```
hist(lm1$residuals)
```

Part d

Make a QQ-plot

```
e1 <- lm1$residuals
e1 <- sort(e1)
n <- dim(model.matrix(lm1))[1]
Q <- qnorm(seq(1:n)/(n+1))
sigmahat <- sqrt(sum(e1^2)/(n-2))
z <- e1/sigmahat
plot(Q,z)
abline(0,1)
```



Part e

Can you see any bad problem individuals?

Part f

What happens if you remove them?

Question 2

Part a

Create a vector of internally studentized residuals

Part b

Creaate a vector of externally studentized residuals

Part c

Create a vector of Cook's Distance

Part d

Do the large values correspond to the data points you thought were bad from last time?

Part e

What happends if you remove them?