Week 9 R Assignment

```
library(RSQLite)
```

```
## Loading required package: DBI
```

```
setwd('C:/Users/Brett/Downloads')
sqlite <- dbDriver("SQLite")
workingdb <- dbConnect(sqlite, "cunyweek9.sqlite")
dbListTables(workingdb)
```

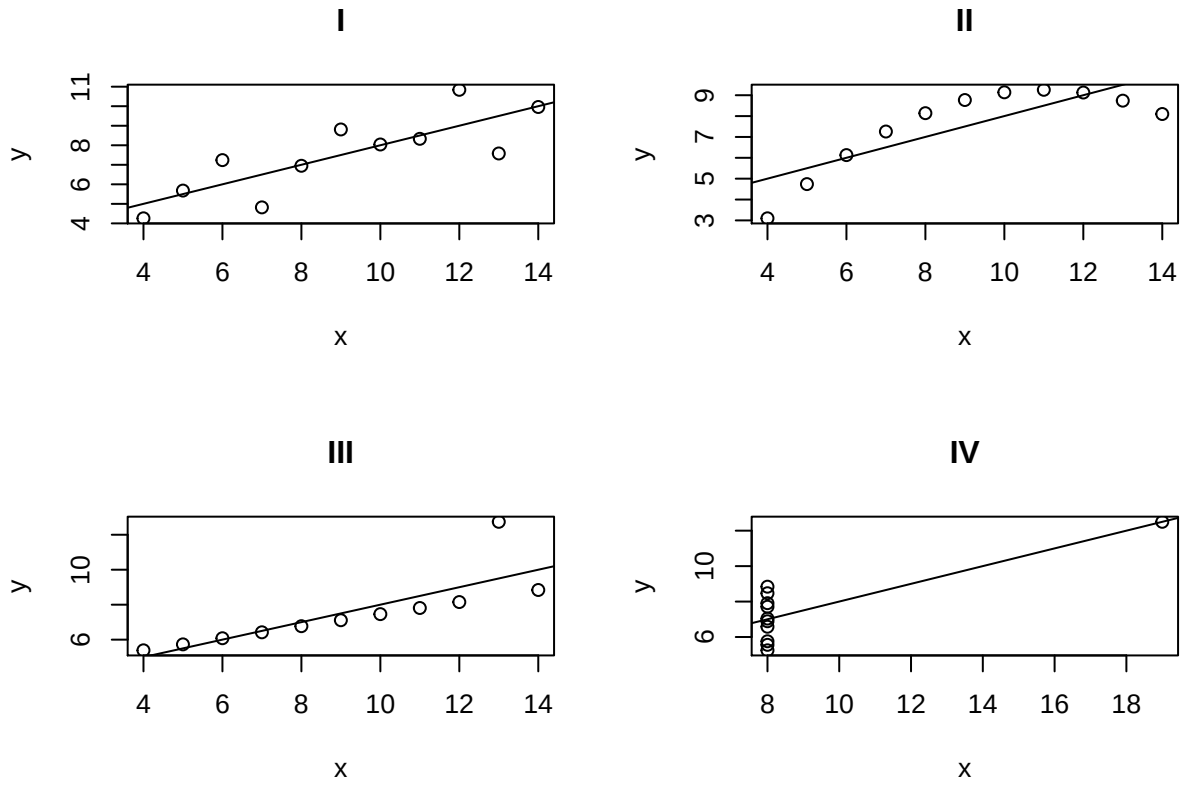
```
## [1] "I"   "II"  "III" "IV"
```

```
I <- dbSendQuery(workingdb, "SELECT * FROM I")
Idf <- data.frame(fetch(I, -1))
II <- dbSendQuery(workingdb, "SELECT * FROM II")
IIdf <- data.frame(fetch(II, -1))
III <- dbSendQuery(workingdb, "SELECT * FROM III")
IIIdf <- data.frame(fetch(III, -1))
IV <- dbSendQuery(workingdb, "SELECT * FROM IV")
IVdf <- data.frame(fetch(IV, -1))
dbDisconnect(workingdb)
```

```
## Warning: RS-DBI driver warning: (closing pending result sets before
## closing this connection)
```

```
## [1] TRUE
```

```
remove(sqlite)
remove(workingdb)
remove(I, II, III, IV)
par(mfrow=c(2,2))
plot(Idf, main="I")
abline(lm(Idf$y ~ Idf$x))
plot(IIdf, main="II")
abline(lm(IIdf$y ~ IIdf$x))
plot(IIIdf, main="III")
abline(lm(IIIdf$y ~ IIIdf$x))
plot(IVdf, main="IV")
abline(lm(IVdf$y ~ IVdf$x))
```



We can see that all four have a very similar regression line, but obviously those regression lines don't tell us very much about II, III, or IV

```
summary(Idf)
```

```
##        x               y         
##  Min.   : 4.0   Min.   : 4.26  
##  1st Qu.: 6.5   1st Qu.: 6.32  
##  Median : 9.0   Median : 7.58  
##  Mean   : 9.0   Mean   : 7.50  
##  3rd Qu.:11.5   3rd Qu.: 8.57  
##  Max.   :14.0   Max.   :10.84
```

```
summary(IIdf)
```

```
##        x               y        
##  Min.   : 4.0   Min.   :3.10  
##  1st Qu.: 6.5   1st Qu.:6.70  
##  Median : 9.0   Median :8.14  
##  Mean   : 9.0   Mean   :7.50  
##  3rd Qu.:11.5   3rd Qu.:8.95  
##  Max.   :14.0   Max.   :9.26
```

```r
summary(IIIdf)
```

```
##        x               y         
##  Min.   : 4.0   Min.   : 5.39  
##  1st Qu.: 6.5   1st Qu.: 6.25  
##  Median : 9.0   Median : 7.11  
##  Mean   : 9.0   Mean   : 7.50  
##  3rd Qu.:11.5   3rd Qu.: 7.98  
##  Max.   :14.0   Max.   :12.74
```

```r
summary(IVdf)
```

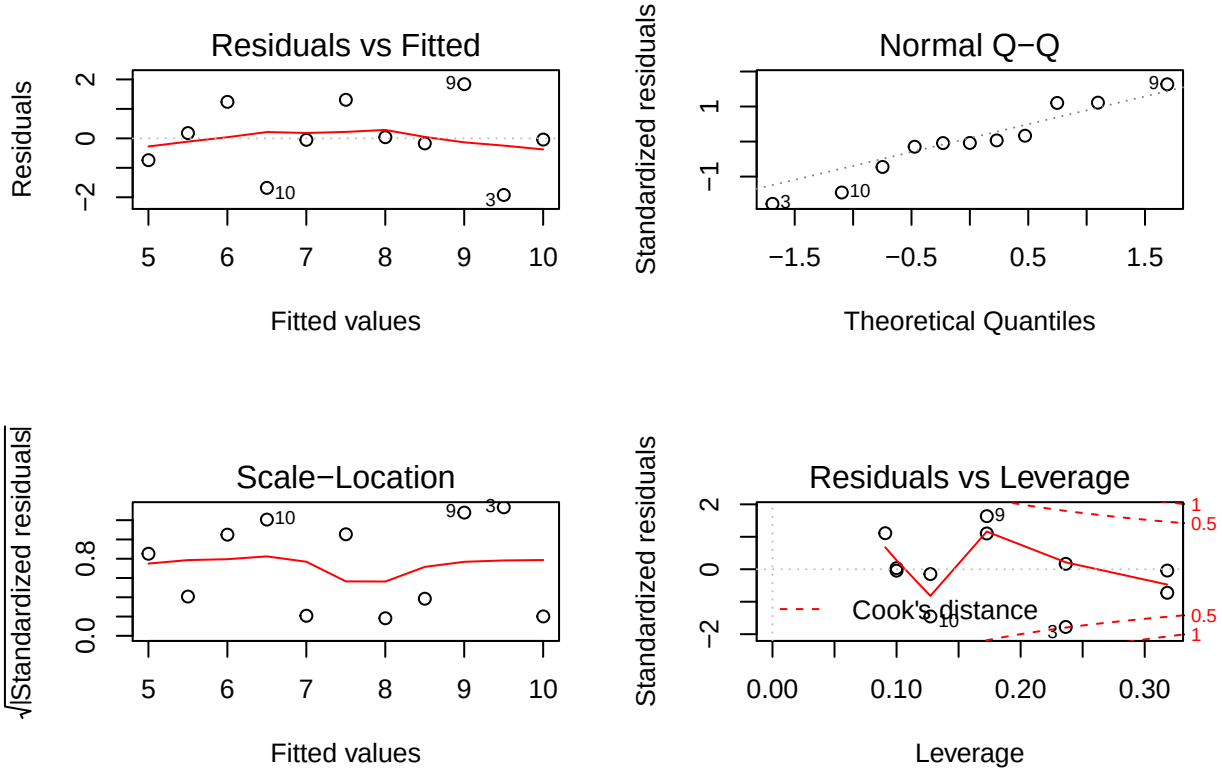
```
##        x            y         
##  Min.   : 8   Min.   : 5.25  
##  1st Qu.: 8   1st Qu.: 6.17  
##  Median : 8   Median : 7.04  
##  Mean   : 9   Mean   : 7.50  
##  3rd Qu.: 8   3rd Qu.: 8.19  
##  Max.   :19   Max.   :12.50
```

Summary statistics! Now we're getting somewhere.

Now let's create linear models and see how they look when graphed:

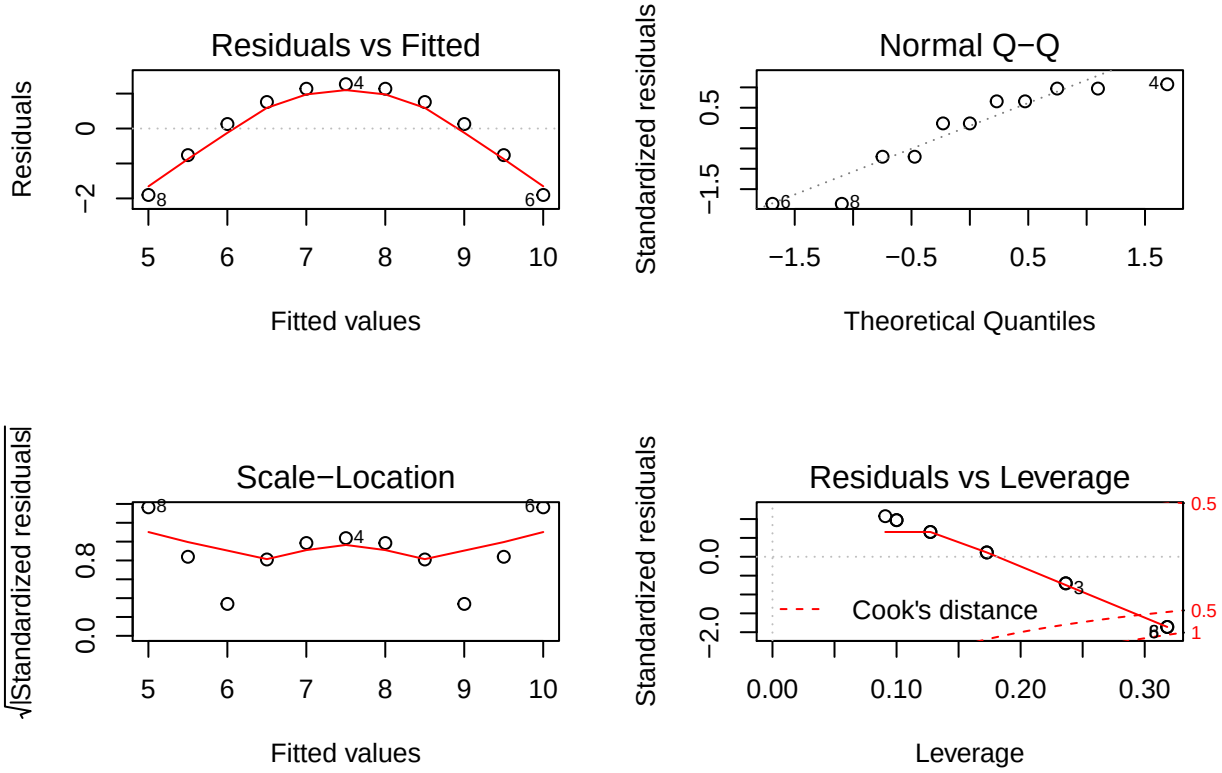
I:

```
par(mfrow=c(2,2))
Ilm <- lm(formula = y ~ x, data = Idf)
plot(Ilm)
```



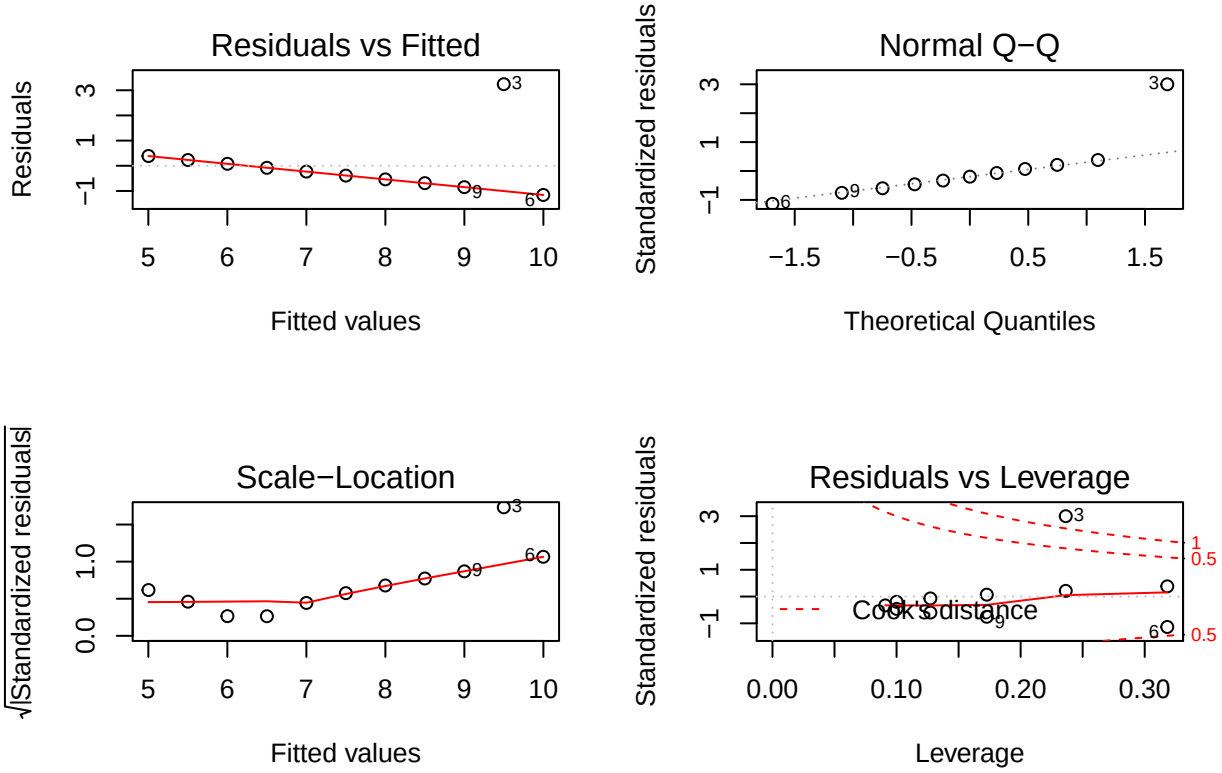
II:

```
par(mfrow=c(2,2))
IIlm <- lm(formula = y ~ x, data = IIdf)
plot(IIlm)
```



III:

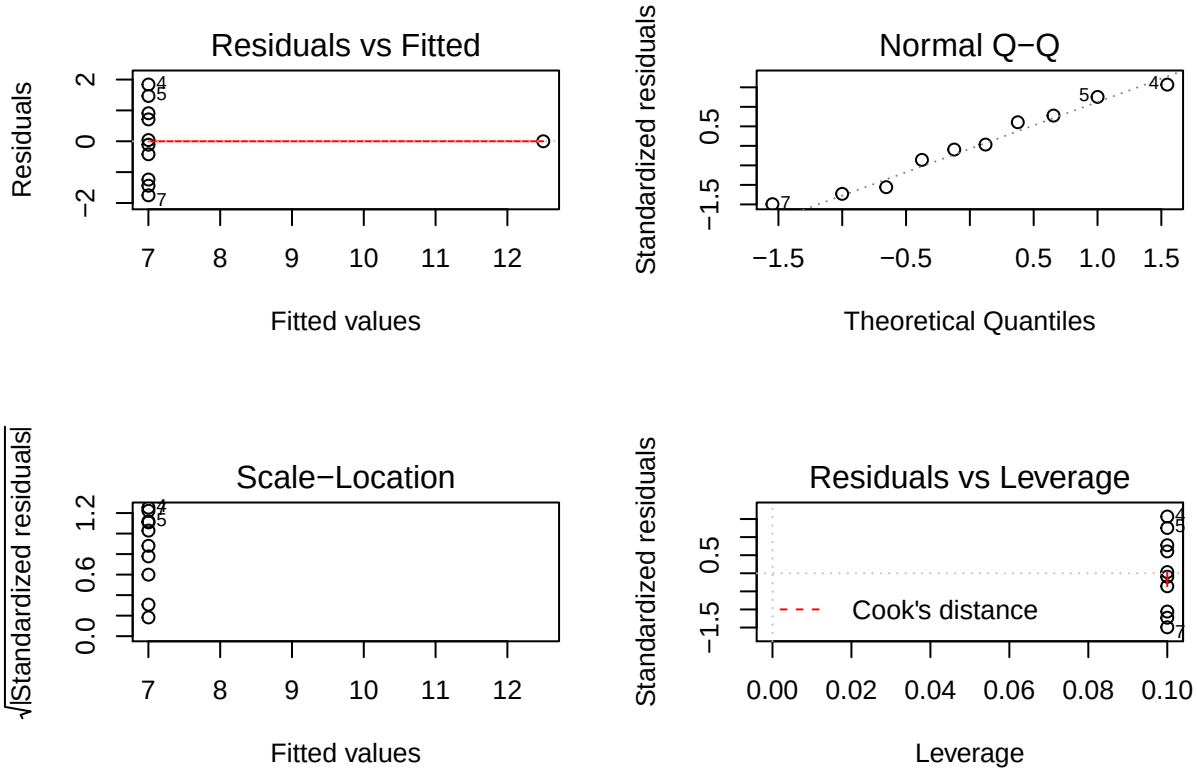
```
par(mfrow=c(2,2))
IIIlm <- lm(formula = y ~ x, data = IIIdf)
plot(IIIlm)
```



IV:

```
par(mfrow=c(2,2))
IVlm <- lm(formula = y ~ x, data = IVdf)
plot(IVlm)
```

```
## Warning: not plotting observations with leverage one:
##   8
## Warning: not plotting observations with leverage one:
##   8
```



Now we have much better predictors for our data. The next step will be working with the data. . . should we drop outliers?