Homework 3

Fechar Ourotcha

3/5/2020

1. Load the airquality dataset Regress Ozone against each of Solar.R, Wind and Temp (you can ignore Month and Day for now).

a. What are the results of the regression

```
data(airquality)
library(sandwich)
library(AER)
library(MASS)
head(airquality)



lm(airquality$Ozone ~ airquality$Solar.R)
lm(airquality$Ozone ~ airquality$Wind)
lm(airquality$Ozone ~ airquality$Temp)
```

When I regress Ozone with solar.R, I got 18.6 as the intercept and 0.13 as beta 1

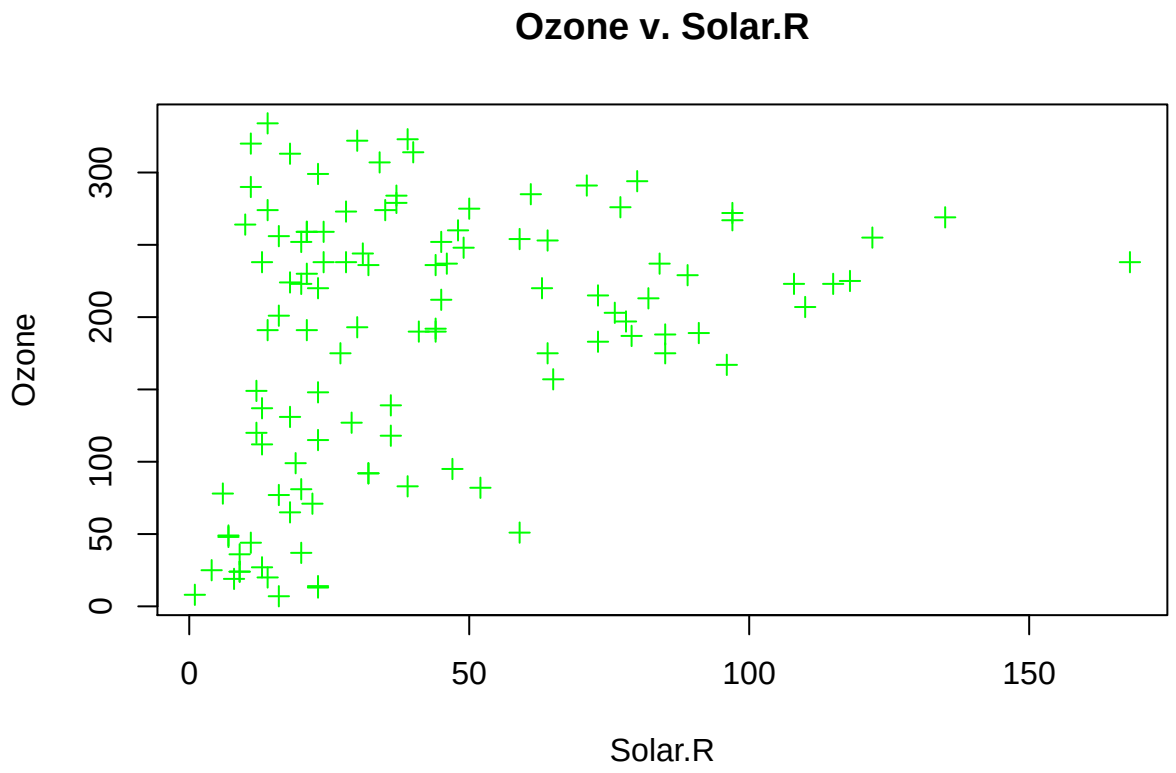
When I regress Ozone with wind, I got 96.87 as the intercept and -5.55 as beta 1

When I regress Ozone with , I got -147.0 as the intercept and 2.43 as beta 1

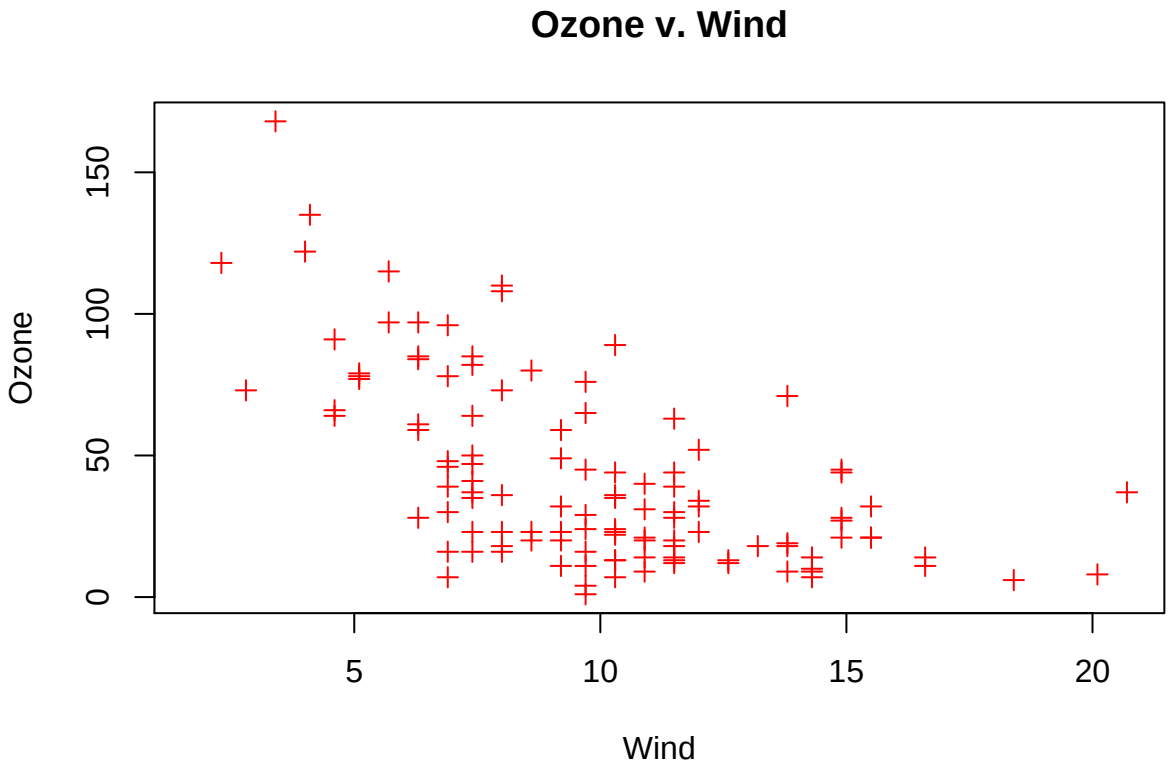
b. Make a scatterplot of each independent variable against the dependent variable.

```
#par(mfcol= c(3,1))
plot(airquality$Ozone, airquality$Solar.R,
     main = "Ozone v. Solar.R",
     xlab = "Solar.R",
```

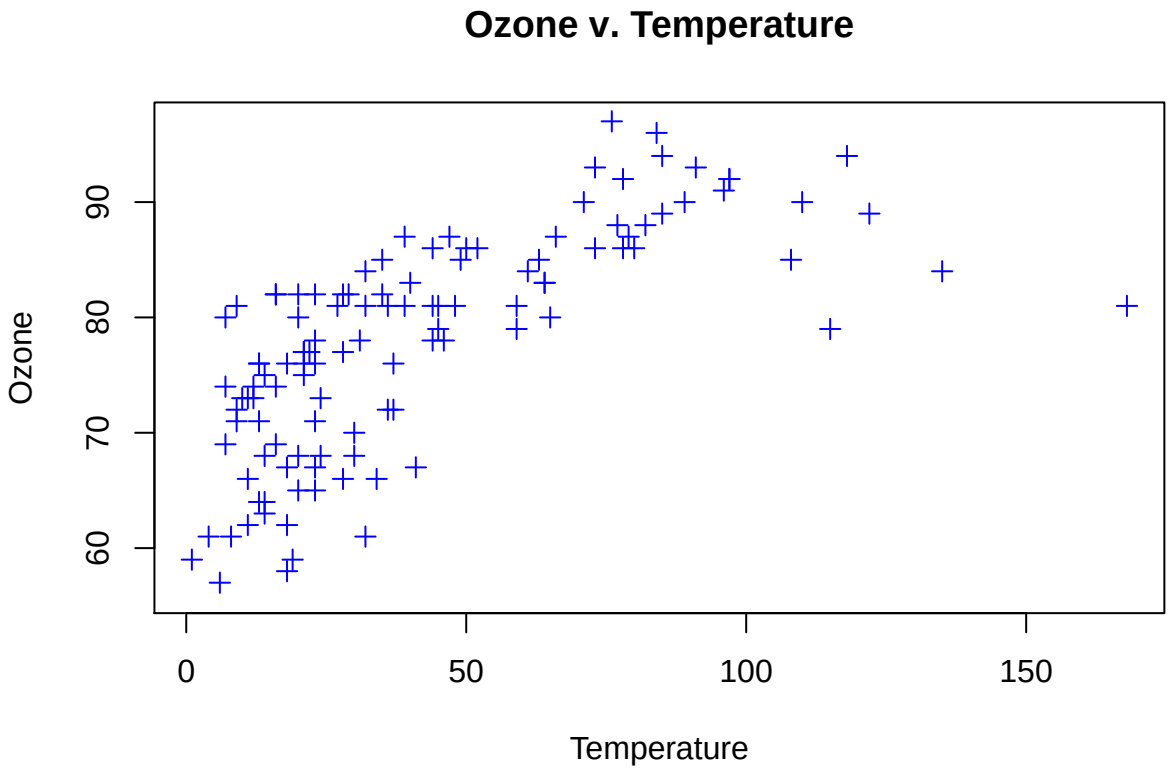
```
     ylab = "Ozone",
     pch = 3,
     col = "green")
```




```
plot(airquality$Ozone ~ airquality$Wind,
     main = "Ozone v. Wind",
     xlab = "Wind",
     ylab = "Ozone",
     pch = 3,
     col = "red")
```




```
plot(airquality$Ozone , airquality$Temp,
     main = "Ozone v. Temperature",
     xlab = "Temperature",
     ylab = "Ozone",
     pch = 3,
     col = "blue")
```



c. Without getting into formal regression results, does each of the models look "good" to you? Why and why not?

The last two seems good to me. Although it's not a strong one, they still show some correlation between the independent variable and the dependent

variable. The points for the first graph are too far apart.

2. Load the anscombe data set. This is a dataset of four pairs of (x,y) variables. For each of the 4 pairs (x1 and y1, x2 and y2, ect) #a. Find the average y value #b. Find the standard deviation of the y value

```
data("anscombe")
#head(anscombe)

mean(anscombe$y1)
```

```
## [1] 7.500909
```

```r
mean(anscombe$y2)
```

```
## [1] 7.500909
```

```r
mean(anscombe$y3)
```

```
## [1] 7.5
```

```r
mean(anscombe$y4)
```

```
## [1] 7.500909
```

```r
sd(anscombe$y1)
```

```
## [1] 2.031568
```

```r
sd(anscombe$y2)
```

```
## [1] 2.031657
```

```r
sd(anscombe$y3)
```

```
## [1] 2.030424
```

```r
sd(anscombe$y4)
```

```
## [1] 2.030579
```

#c. Make a simple regression and report the coefficients.

```r
lm(anscombe$y1 ~ anscombe$x1)
```

```
##
## Call:
## lm(formula = anscombe$y1 ~ anscombe$x1)
##
## Coefficients:
## (Intercept)  anscombe$x1
##      3.0001       0.5001
```

```r
lm(anscombe$y2 ~ anscombe$x2)
```

```
##
## Call:
## lm(formula = anscombe$y2 ~ anscombe$x2)
##
## Coefficients:
## (Intercept)  anscombe$x2
##       3.001        0.500
```

```
lm(anscombe$y3 ~ anscombe$x3)
```

```
##
## Call:
## lm(formula = anscombe$y3 ~ anscombe$x3)
##
## Coefficients:
## (Intercept)  anscombe$x3
##      3.0025       0.4997
```

```
lm(anscombe$y4 ~ anscombe$x4)
```

```
##
## Call:
## lm(formula = anscombe$y4 ~ anscombe$x4)
##
## Coefficients:
## (Intercept)  anscombe$x4
##      3.0017       0.4999
```

#d. make a scatterplot of each of the four pairs

```
lm(anscombe$y3 ~ anscombe$x3)
```

```
##
## Call:
## lm(formula = anscombe$y3 ~ anscombe$x3)
##
## Coefficients:
## (Intercept)  anscombe$x3
##      3.0025       0.4997
```

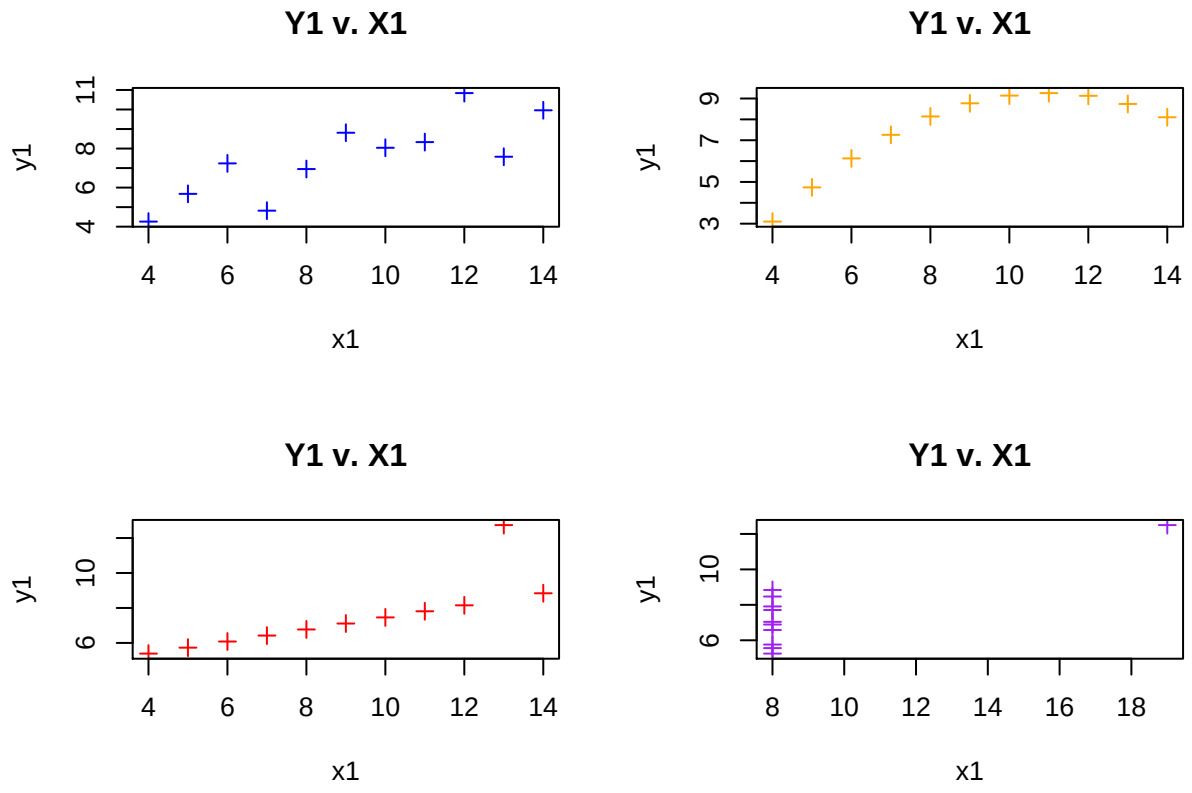
```
par(mfrow=c(2,2))

a <- plot(anscombe$x1, anscombe$y1,
     main = "Y1 v. X1",
     xlab = "x1",
     ylab = "y1",
     pch = 3,
     col = "blue")

b <- plot(anscombe$x2, anscombe$y2,
     main = "Y1 v. X1",
     xlab = "x1",
     ylab = "y1",
     pch = 3,
     col = "orange")
c <- plot(anscombe$x3, anscombe$y3,
     main = "Y1 v. X1",
     xlab = "x1",
     ylab = "y1",
```

```
     pch = 3,
     col = "red")

d <- plot(anscombe$x4, anscombe$y4,
     main = "Y1 v. X1",
     xlab = "x1",
     ylab = "y1",
     pch = 3,
     col = "purple")
```




```
#i.  hint you can use par(mfrow=c(2,2)) to put all 4 on a graph
```

e. What do you notice?

Although the coeffients are very similar, all four graphs are scattered differently