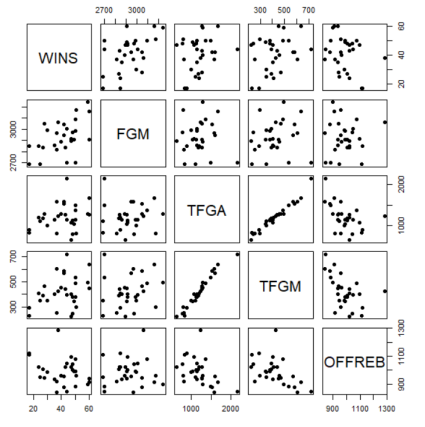
Problem #1

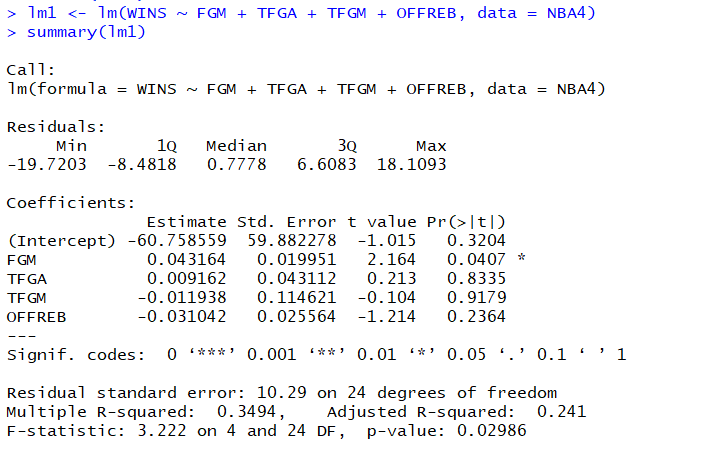
(a)

(1) The matrix scatterplot of WINS (as the response variable), field goals made (FGM), three-point field goals attempted (TFGA), three-point field goals made (TFGM), and the offensive rebounds (OFFREB)



(2) The matrix of correlations among these variables

WINS = Beta\_0 + Beta\_1FGM + Beta\_2TFGA + Beta\_3TFGM + Beta\_4OFFREB



WINS = -60.75 + 0.043 FGM + 0.0092 TFGA – 0.012 TFGM – 0.031 OFFREB

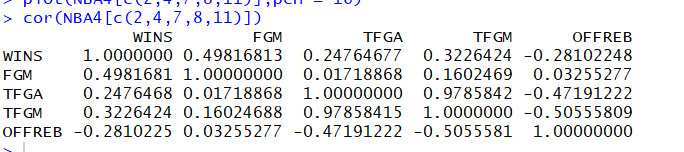
(3)

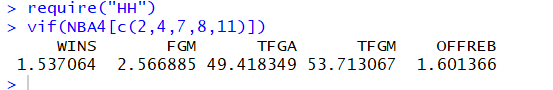
Yes, there have some factors might not significant with WINS, because the P-values are kind of big.

need to show the VIF after the variable is changed.

(b)

WINS = -60.75 + 0.043 FGM + 0.0092 TFGA – 0.012 TFGM – 0.031 OFFREB





As we can read, TFGA and TFGM are MC problems, they are larger than 5.

(c)

We can delete one of them. Or create a new factor which is a ratio of them two like, three-point successful rate (TSR). It is equal with TFGM/TFGA.

Problem #2

should conduct transformation to NUMBER as NUMBER and TIME has a nonlinear relation

(The sentences have gray background are the first time I try the problems, because there have some sentences might confuse you, and I have not time to delete them, so I leave them with gray background)

1. (I do not know the meaning of the question, what does it mean including every variables? Should I image a new possible factors and insert it in? Then test does it significant to the formula? Here seems have a curvilinear relationship, I tried it, but it seems too much and complex (you know, here has two different variances, but what we learned in class only has one basic variance), so I delete them)

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 NUMBER

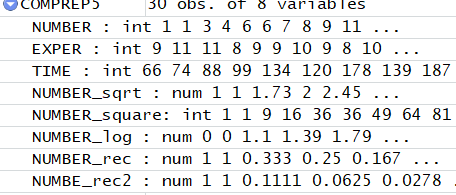
TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 NUMBER + Beta\_3 NUMBER^2

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 sqrt(COMPREP5$NUMBER)

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 log(COMPREP5$NUMBER)

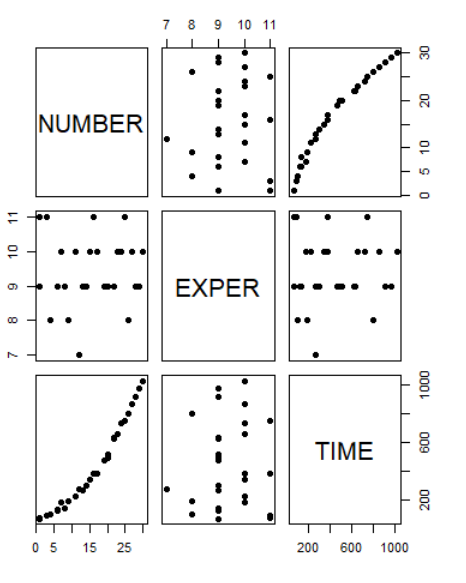
TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 1/COMPREP5$NUMBER

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 1/(COMPREP5$NUMBER\*COMPREP5$NUMBER)

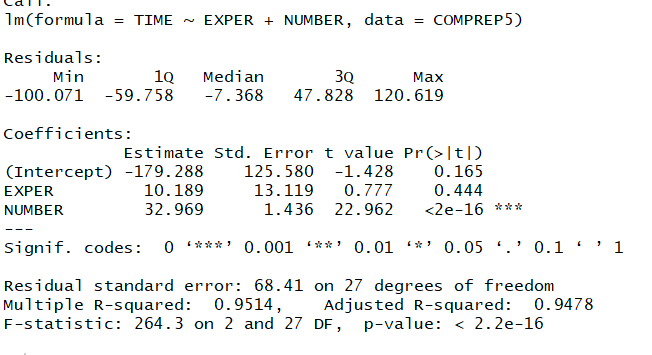


So, I just assume

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 NUMBER



(2)

~~~~

TIME = -179.3 + 10.2 EXPER + 33 NUMBER

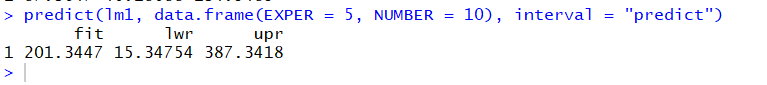
H0: we should not believe this formula

Ha: we should believe this formula

Since P-Value < 0.01, so we reject H0.

I believe this formula, so the time Joe will spend with this client is 201.7 minutes.

Because of (3), I change the answer of (2)

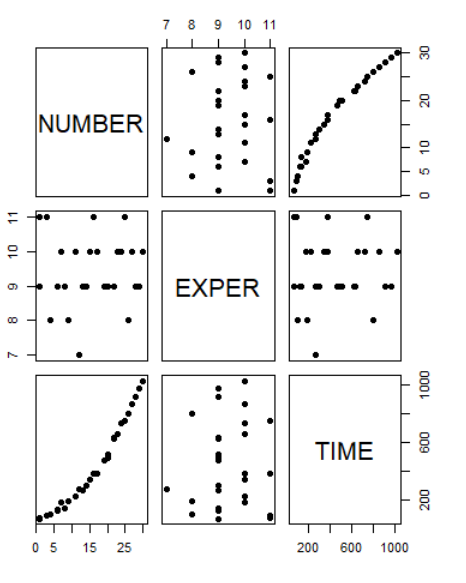


So, the period is 201.34 to 387.34 minutes.

(3)

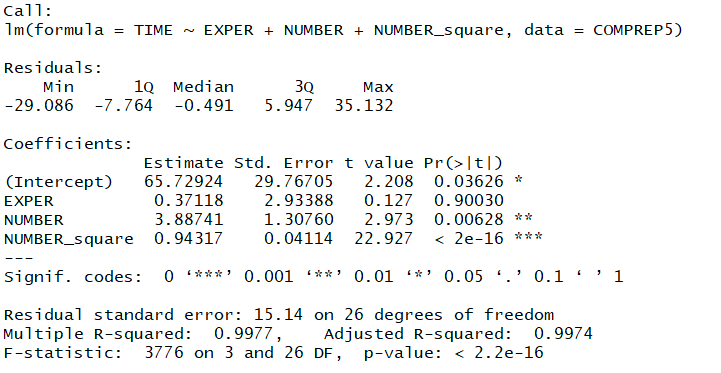
If it is same principle, I guess the time is 10.2 \* 4 + 33 \* 3 – 179.3 = -39.5

So, I guess this formula would be wrong within some area of NUMBER and EXPER,



I am not sure, here would be too much if I paste them, so I just use this here. Plus here only leave 50 mins to 0:00 o’clock even this is my fault, so sorry.

TIME = Beta\_0 + Beta\_1 EXPER + Beta\_2 NUMBER + Beta\_3 NUMBER^2

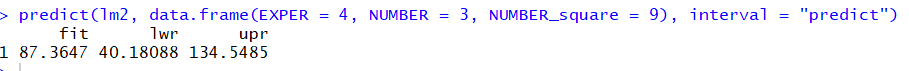


H0: we should not believe this formula

Ha: we should believe this formula

Since P-Value < 0.05, so we reject H0.

I can believe this formula. Compare with the screen shot in (2), this is more fitter.



So, the period is 87.36 to 134.55 minutes.

try to transform y

Problem#3

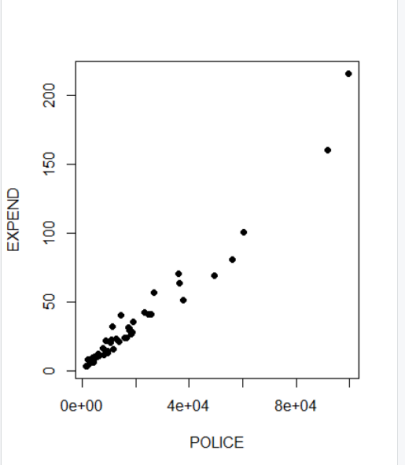
(a)

Yes, we need to let the data in EXPEND be divided by 1,000,000.



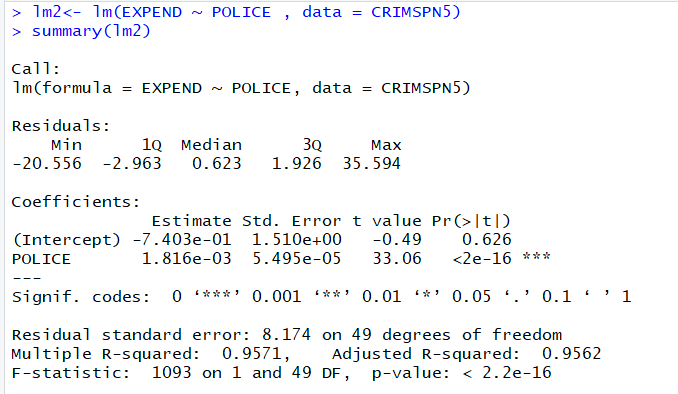
The spending has relationship with state, but not with their name, so I would not give them a ID number. I do not know what I could add in, so I only use the POLICE.

(b)

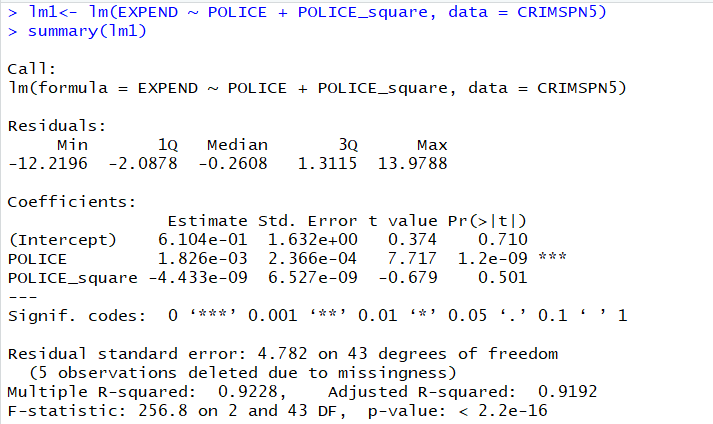


I assume the regression model might be

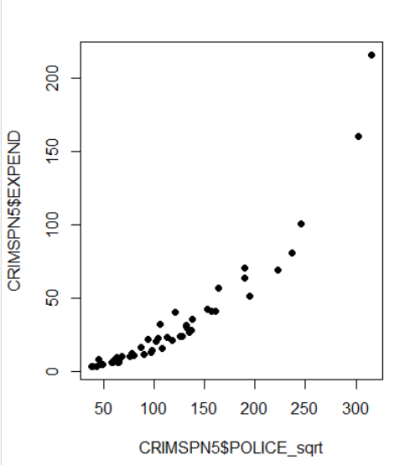
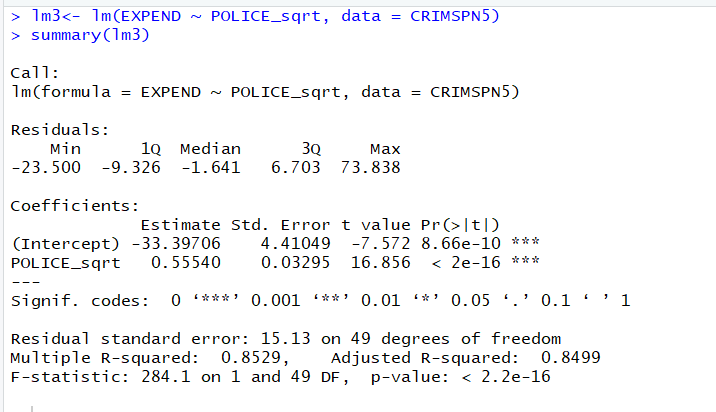
1. EXPEND = Beta\_0 + Beta\_1 POLICE



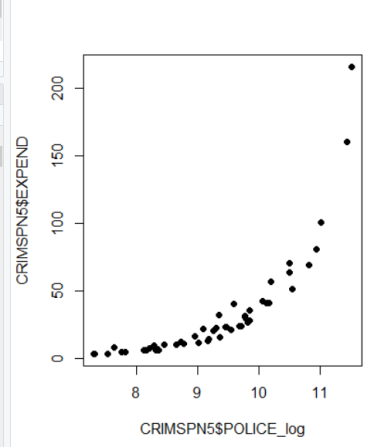
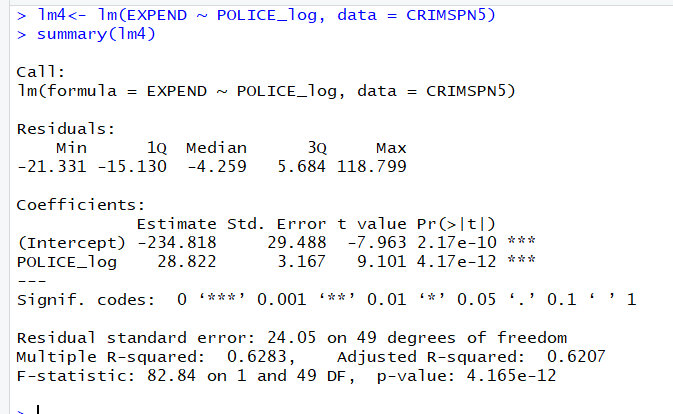
1. EXPEND = Beta\_0 + Beta\_1 POLICE + Beta\_2 POLICE^2



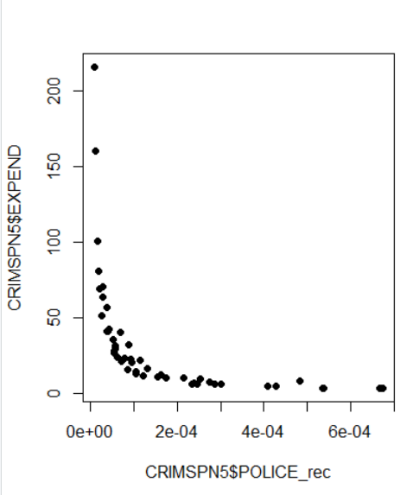
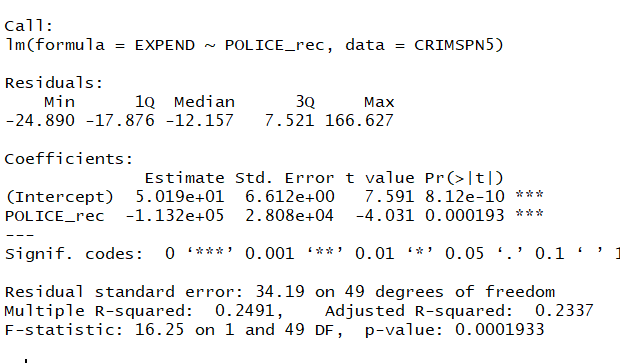
1. EXPEND = Beta\_0 + Beta\_1 sqrt(POLICE)

1. EXPEND = Beta\_0 + Beta\_1 log(POLICE)

1. EXPEND = Beta\_0 + Beta\_1 1/POLICE

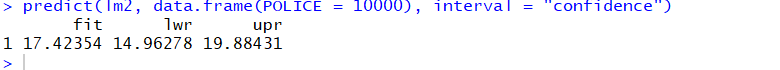
1. EXPEND = Beta\_0 + Beta\_1 1/(POLICE)^2

You can see, the fitted level would be lower than 5., so compare all of them, we could have a conclusion

The 1. would be better.

So, the (multiple) regression model is EXPEND = -0.7403 + 0.001816 POLICE

(c)



So, it is between 17.42 and 19.88 million dollars.

interval = 'predict'(-2pt)