MSIT 423

Project One

Ray Liu, Jessica Qin, YangHong, Yifan Chen, Yunzi Zhang

Features and model description

We picked some of original variables considering their casual relationships, and finally get the fowling model based on the linear assumption with an adjusted R square of 75.48%.

$$trips = 8.1 - 0.38557ASSAULT - 0.34593BURGLARY + 0.67074THEFT + 0.04846CAPACITY - 1.36232MINORITY^2 + \in_i$$

Predictors	**Definition**	**Transformation**
ASSAULT	A crime ability of threat, excluding bodily harm	
BURGLARY	Unlawful entry into a building for th purpose of committing an offence	
THEFT	Stealing, it increases the demand to our surprise. May be because the theft makes people less likely to ride their own bikes.	
CAPACITY	The maximum carrying amount	
MINORITY	The group type of people in certain location	square

Introduction & Analysis

The goal of this project is to build a predictive model to predict bike demands based on major crime types and city features information. 9 major crime types, assault, battery, burglary, criminal

trespass, deceptive practice, homicide, narcotics, robbery, theft, might affect demands of bikes. Increasing the number of crimes that cause loss of personal possession and direct harm of body could lead the decreasing demands of bikes because people will feel unsafe riding a bike on the road. On the other hand, crimes that happen less frequently and cause major social affect such as homicide will have less effect on the demand of bikes due to the fact that there is nothing people can do to prevent it from happening.

City's city features information can also have profound effect on the demands of bikes. High number of bike routes means there are more demands on bikes. High number of retails stores and restaurant indicate this area is prosperous, therefore having more people, which leads to high demand of bikes.

1.1 Features Initial Analysis and Grouping:

The bike data has 45 predictor variables and 1 response variables. According to the business objective, we only select 9 crime types and city features information. Our initial thought is to composite 9 crimes as one variable since they are highly correlated. Also, we want to group some city features variables that are highly related. However, after grouping, R square has decreased. Also, we want to analyze bike demands based on each type of crime, so we decide not to include composite variables in our model.

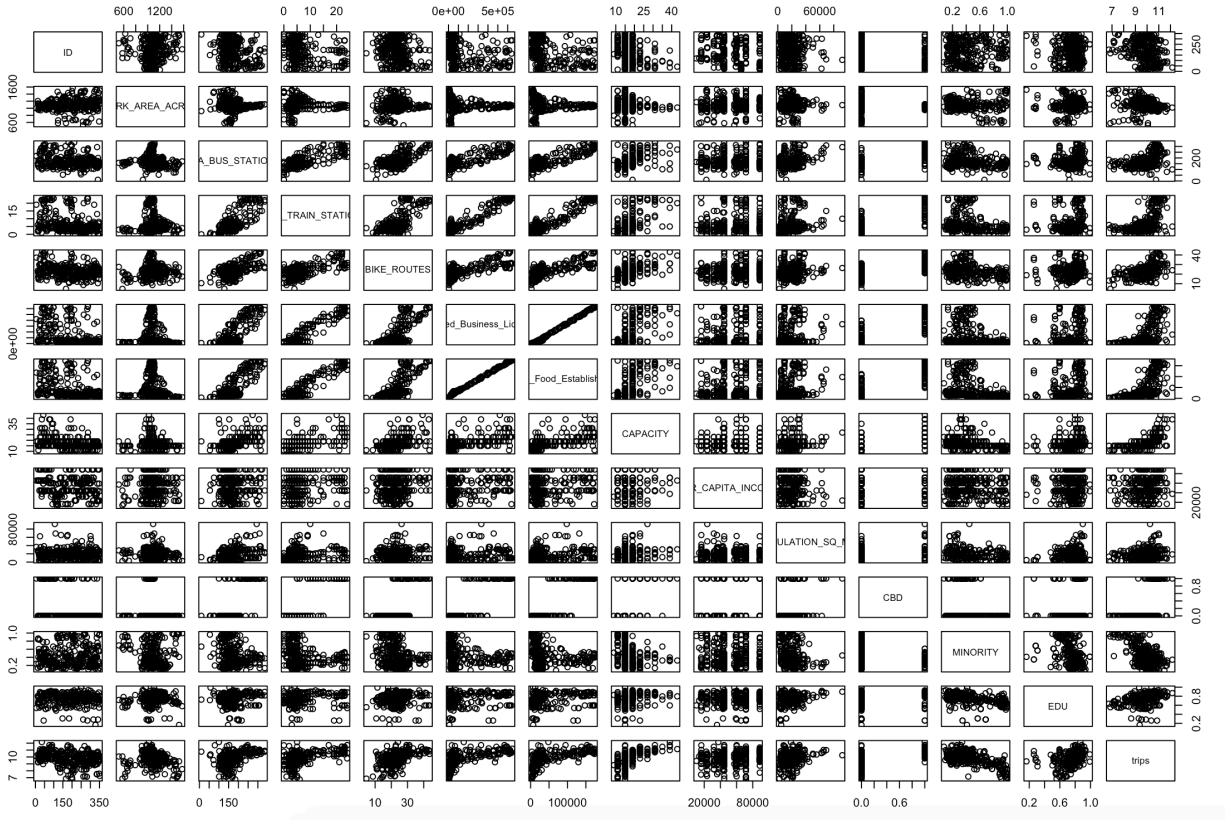
Our model analysis will separate into two parts. First part is to analyze city features variables and build the first model. Second part is to analyze crime variables and build the second model. Finally, we combine the two models together to get final model. The correlations between city features and crime are not strong, so the coefficient of our final model will not be affected.

First Model - City Features

2.1 Preliminary analysis:

2.1.1 Omitted unrelated predictors and correlation analysis

We diagnose data by generating correlation matrix and scatter plots. First, we observe that some predictors are to be removed since they are not correlated with demands (trips) as well as not cause of dependent variables.



PIC(3.1)

According to the scatterplot (3.1), predictors to be omitted are:

ID, PARK_AREA_ACRES, PER_CAPITA_INCOME

2.1.2 Correlation analysis

According to the correlation matrix, the multicollinearity exists.

	CTA_TRAIN_STATIONS	Limited_Business License	Retail_Food_Establishment
Limited_Business_License	0.9400	1	0.997
CTA_BUS_STATIONS	0.76	0.79	0.80
Retail_Food_Establishment	0.9396	0.997	1
CBD	0.85	0.882	0.882

If we were to use these correlated predictors in the model, we can end up with wrong conclusions from the model, e.g., concluding a variable is not significant.

2.2 Regression Model:

2.2.1 Predictor selection

According to the correlation matrix and the scatterplot, there is a strong multicollinearity problem. So we use Lasso to select the predictors. As the output in Pic(3.2.3), after considering the correlation, we decide follow predictors into our first model.

CTA_TRAIN_STATIONS, BIKE_ROUTES, Retail_Food_Establishment, CBD, CAPACITY, MINORITY, EDU

```
11 x 1 sparse Matrix of class "dgCMatrix"
                                      1
(Intercept)                8.931377e+00
CTA_BUS_STATIONS           .
CTA_TRAIN_STATIONS        -7.289214e-03
BIKE_ROUTES                6.799076e-03
Limited_Business_License   .
Retail_Food_Establishment  5.348648e-06
CAPACITY                   5.298382e-02
POPULATION_SQ_MILE         .
CBD                        2.480839e-02
MINORITY                  -1.736929e+00
EDU                        4.754218e-01
```

Pic(3.2.1)

2.2.2 Decide independent variable transformation

While the predictors should be transformed may be:

Predictors	Transform assumption1	Transform assumption2
BIKE_ROUTES	N/A	square
CTA_TRAIN_STATIONS,		
RETAIL_FOOD_ESTABLISHMENT		
CAPACITY		
MINORITY		
EDU	N/A	N/A
CBD	Category	

According to the summary of our linear model, the p-value of BIKE_ROUTES, EDU, and CBD are 0.285, 0.078, 0.443 respectively, indicating those three predictors are not significant. After Drop those three predictors, we get our first model(with out crimes) below, with an adjusted R^2 of 70%:

$$trips = 9.125 - 5.744 \times 10^{-2} CTA\_TRAIN\_STATIONS + 1.166 \times 10^{-5} Retail\_Food\_Establishment + 5.872 \times 10^{-2} CAPACITY - 1.81 MINORITY^2 + \in_i$$

With a numerical summary:

```
Call:
lm(formula = trips ~ CTA_TRAIN_STATIONS + Retail_Food_Establishment +
    CAPACITY + I(MINORITY^2), data = com1.2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.24371 -0.30726  0.01372  0.33852  2.06842

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                9.125e+00  1.203e-01  75.870  < 2e-16 ***
CTA_TRAIN_STATIONS        -5.744e-02  1.380e-02  -4.162 4.15e-05 ***
Retail_Food_Establishment  1.166e-05  1.817e-06   6.416 5.56e-10 ***
CAPACITY                   5.872e-02  7.195e-03   8.162 9.67e-15 ***
I(MINORITY^2)             -1.810e+00  1.135e-01 -15.945  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5235 on 295 degrees of freedom
Multiple R-squared:  0.7063,    Adjusted R-squared:  0.7024
F-statistic: 177.4 on 4 and 295 DF,  p-value: < 2.2e-16
```

After calculating the VIF, there is no significant multicollinearity in this model.

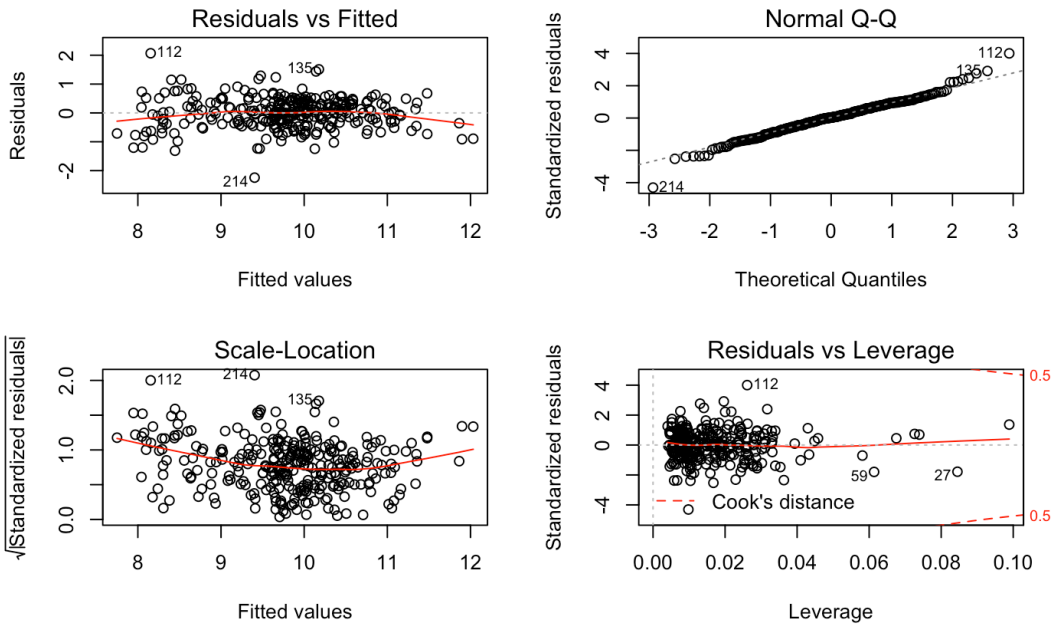
```
CTA_TRAIN_STATIONS Retail_Food_Establishment                    CAPACITY
          8.646809                   8.783578                    1.633999
     I(MINORITY^2)
          1.103478
```

2.3 Model validation

In this section, we validate the linear assumptions of the regression model after we apply the square transformation. Pic(3.3) shows the diagnostic plots of the model. From left to right, up and down, the plot are for linearity, constant variance, uncorrelated errors, and error normality.

The residual plot does not indicate a particular pattern, indicating the model after partly square transformation is linear. Meanwhile, the residuals are like snowstorm, showing a pretty good match. The Q-Q plot shows the normality of errors, and there is no outlier with a Cook’s Distance more than 0.5.

According to present diagnostic, this model obeys the assumption and has no outliers.



Final Model – Crime Added

3.1 Preliminary analysis:

3.1.1 Omitted unrelated predictors and correlation analysis

First, we perform preliminary analysis by looking at the correlations between each crime variable and the bike demand ('trips'). The crime variables with moderate correlations are selected.

Take all the eight crimes into consideration, only Theft, Deceptive_practice, Assault, Burglary, and Robbery seems to be significant. According to the VIF analysis, there is multicollinearity between theft, battery, deceptive, and assault, making battery insignificant.

The multicollinearity comes from the behavior of these crimes, we will explain all the multicollinearity in Section 5.

3.2 Regression Model:

3.2.1 Predictor Selection

From the correlation analysis, we found out that there is multicollinearity between variables, so selection variables simply based on scatterplot and correlation matrix will not be efficient. The whole model might be significant, but individual variable will not be significant. THEFT, ASSAULT, BURGLARY can have impact on demand of bike.
According to the Lasso selection below, we decide to take ASSAULT, BURGLARY, DECEPTIVE_PRACTICE, HOMICIDE, and THEFT into our first model.

```
10 x 1 sparse Matrix of class "dgCMatrix"
                             1
(Intercept)         7.0764378
(Intercept)         .
ASSAULT            -0.5276035
BATTERY             .
BURGLARY           -0.2122191
DECEPTIVE_PRACTICE  0.2161434
HOMICIDE           -0.1798393
NARCOTICS           .
ROBBERY             .
THEFT               0.7456659
```

3.2.2 Decide independent variable transformation

According to the scatter plot, we assume the predictors have the linear relationship with trips, so no transformation needed.

When combine the demographic predictors and the crime predictors, we get the summary below. As the DECEPTIVE_PRACTICE has the VIF of 14.55, and the homicide has the p-value of 0.788, we decide to drop them.

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.100143   0.319115  25.383  < 2e-16 ***
ASSAULT       -0.375555   0.080412  -4.670 4.59e-06 ***
BURGLARY      -0.345342   0.053935  -6.403 6.04e-10 ***
THEFT          0.665550   0.071191   9.349  < 2e-16 ***
HOMICIDE      -0.013415   0.049921  -0.269    0.788
CAPACITY       0.048374   0.006308   7.669 2.57e-13 ***
I(MINORITY^2) -1.354541   0.151145  -8.962  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finally we get the model:

$$trips = 8.1 - 0.38557ASSAULT - 0.34593BURGLARY + 0.67074THEFT + 0.04846CAPACITY - 1.36232MINORITY^2 + \in_i$$

With a numerical summary:

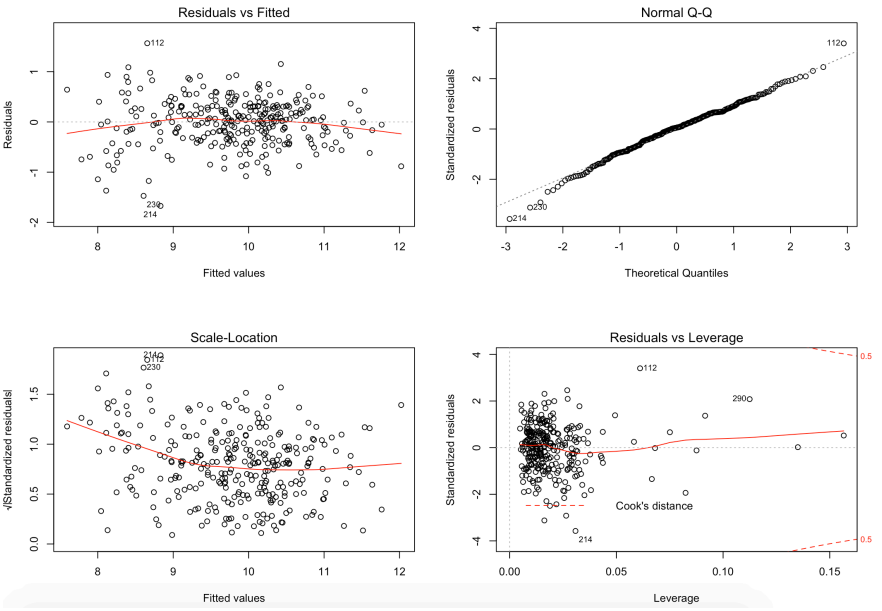
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.10356    0.31836  25.454  < 2e-16 ***
ASSAULT       -0.38557    0.07115  -5.419 1.25e-07 ***
BURGLARY      -0.34593    0.05380  -6.429 5.17e-10 ***
THEFT          0.67074    0.06842   9.804  < 2e-16 ***
CAPACITY       0.04846    0.00629   7.703 2.04e-13 ***
I(MINORITY^2) -1.36232    0.14811  -9.198  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4752 on 294 degrees of freedom
Multiple R-squared:  0.7589,    Adjusted R-squared:  0.7548
F-statistic:   185 on 5 and 294 DF,  p-value: < 2.2e-16
```

3.3 Model validation

In this section, we validate the linear assumptions of the regression model after we apply the square transformation. shows the diagnostic plots of the model. From left to right, up and down, the plot are for linearity, constant variance, uncorrelated errors, and error normality.

The residual plot does not indicate a particular pattern, indicating the model after partly square transformation is linear. Meanwhile, the residuals are like snowstorm, showing a pretty good match. The Q-Q plot shows the normality of errors, and there is no outlier with a Cook's Distance more than 0.5.



While the VIF are all small.

```
> vif(fitcrime2)
     ASSAULT      BURGLARY         THEFT      CAPACITY I(MINORITY^2)
    4.230879      1.556740      4.715884      1.515768      2.279985
```

Interpretation

4.1 Predictor interpretation

CTA_TRAIN_STATION: It is the predictor represent the transportation of the location. For people who transport to place far away would like to ride to the train station. On the other hand, people get way from the train may prefer to ride home.

Retail_Food_Establishment: It is the predictor represent the pedestrian volume of the location. Assume the probability of the people who would like to ride the bike is constant, the big pedestrian volume helps the volume for trips.

MINORITY: The further information is needed to analyze these two predictors. Certain groups of people may less likely to bike.

CAPACITY: A higher capacity may attract more customers, thought the capacity influence the trips only in a slight level.

ASSAULT&BURGLARY: These two predictors are all the indicates of happening of crimes. The crimes will decrease the demand of the trips because of the safety.

THEFT : We distinguish this crime activity from the ASSAULT and BURGLARY, because it has the contradict effect to the other crimes. We make some assumption:

1. The model is not correct, or need to be improved.
2. The less income per person making the frequent theft crimes, meanwhile, the less income people does not have their own bike.
3. A location with many theft crimes making people less like to ride their own bike, because they do not want their own bike to be stolen.

We try to build another models by adding other variables, transformation, and interaction, but all of them give us the positive coefficient of theft. Although we cannot reject the first assumption, we have to find another explanation.
According to the correlation between INCOMRE and THEFT, we reject second assumption.
Finally, we think the third assumption may be the best explanation.

4.2 Correlation interpretation

CTA_TRAIN_STATION, CTA_BUS_STATION: In normal life, the train stations always near to the bus stations for transportation convenient.

Retail_Food_Establishment, CTA_TRAIN_STATION, LIMITED LICENSE: The big volume of pedestrian produced by retail food establishment contributes to the station setting. Meanwhile, the retailers always related to business license.

CBD, Retail_Food_Establishment, CTA_TRAIN_STATION, LIMITED LICENSE: The central of the business district always have more train stations and retailers than other places. Business licenses also different in active business places.

ASSAULT, BATTWERY: As defined, assault involves a threat, but not bodily harm, while battery implies harm; however, the harm usually comes along with threat, making a high correlation between assault and battery.(0.95 correlation between battery and assault.

NARCOTICS,ASSAULT,BATTERY: The correlation between Narcotics and Assault, Battery are both bigger than 0.85. Narcotics can be a part of reasons of Assault and Battery. Combine the analysis in correlation of Assault and Battery, Narcotics has the high probability to be the "w" cause both Assault and Battery, making a high correlation between them.

THEFT,ASSAULT,ROBBERY, DECEPTIVE: relation between Theft and Assault, Robbery, Deceptive are all bigger than 0.8. Assumption 1: Income is the causality of these four crimes, however, cor(bike$PER_CAPITA_INCOME, bike$THEFT) is only 0.12. Assumption 2: Assault, Robbery, and Deceptive always come after the Theft.

Conclusion

From our model, ASSAULT, BURGLARY, THEFT, CAPACITY, and MINORITY can explain 75% percent of the demands, however, the further information is needed to improve the demand like the

MINORITY. What kind of groups people less likely to use the bike? Why they do not like the bike? What can we improve?

There are some places we have to improve:

1. Hope we can do the cross validation or k-fold after finishing the future class.
2. Lasso and Ridge cannot delete all the multicollinearity although they delete most of correlation.

Project 1

MSIT 423, Spring 2019
Due: April 27, 2:00pm

Jessica Qin, Yang Hong, Yunzi Zhang, Yifan Chen, Ray Liu

```
library(car)
```

```
## Loading required package: carData
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

0.1 Loading the data from the csv file

```
setwd("~/Desktop/Ray Liu")
bike<-read.csv("bike.csv")
com <- bike[,c(1:13,45)]
com1.2<- com[,c(3:8,10:14)]
cor(com1.2)
```

```
##                           CTA_BUS_STATIONS CTA_TRAIN_STATIONS BIKE_ROUTES
## CTA_BUS_STATIONS                 1.0000000         0.76391087  0.57678715
## CTA_TRAIN_STATIONS               0.7639109         1.00000000  0.68212020
## BIKE_ROUTES                      0.5767872         0.68212020  1.00000000
## Limited_Business_License         0.7930528         0.94001627  0.72209437
## Retail_Food_Establishment        0.8020726         0.93956091  0.74978709
## CAPACITY                         0.4670364         0.60617220  0.46837336
## POPULATION_SQ_MILE               0.3038109         0.08970660  0.09750206
## CBD                              0.7233554         0.85270533  0.63539809
## MINORITY                        -0.1816221        -0.23396665 -0.25718405
## EDU                              0.1119116         0.09380398  0.12139609
## trips                            0.4300325         0.52622037  0.51112087
##                           Limited_Business_License
## CTA_BUS_STATIONS                        0.79305280
## CTA_TRAIN_STATIONS                      0.94001627
## BIKE_ROUTES                             0.72209437
## Limited_Business_License                1.00000000
## Retail_Food_Establishment               0.99706184
## CAPACITY                                0.60607181
## POPULATION_SQ_MILE                      0.07680321
## CBD                                     0.88216517
## MINORITY                               -0.19282372
## EDU                                     0.14236166
## trips                                   0.56061304
```

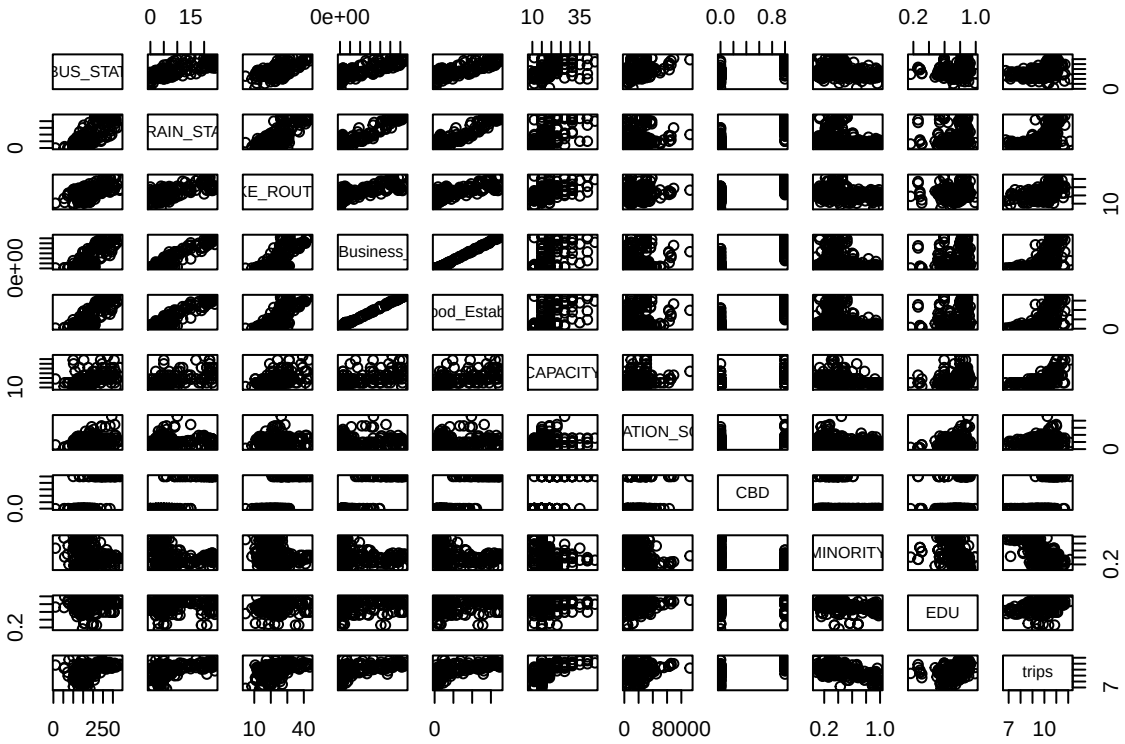
```
##                            Retail_Food_Establishment   CAPACITY
## CTA_BUS_STATIONS                           0.8020726  0.4670364
## CTA_TRAIN_STATIONS                         0.9395609  0.6061722
## BIKE_ROUTES                                0.7497871  0.4683734
## Limited_Business_License                   0.9970618  0.6060718
## Retail_Food_Establishment                  1.0000000  0.6119165
## CAPACITY                                   0.6119165  1.0000000
## POPULATION_SQ_MILE                         0.1009049  0.1121183
## CBD                                        0.8824601  0.6093845
## MINORITY                                  -0.2355995 -0.2152348
## EDU                                        0.1580518  0.1899403
## trips                                      0.5931837  0.5944283
##                            POPULATION_SQ_MILE        CBD   MINORITY
## CTA_BUS_STATIONS                   0.30381094  0.7233554 -0.1816221
## CTA_TRAIN_STATIONS                 0.08970660  0.8527053 -0.2339666
## BIKE_ROUTES                        0.09750206  0.6353981 -0.2571841
## Limited_Business_License           0.07680321  0.8821652 -0.1928237
## Retail_Food_Establishment          0.10090491  0.8824601 -0.2355995
## CAPACITY                           0.11211830  0.6093845 -0.2152348
## POPULATION_SQ_MILE                 1.00000000  0.1817645 -0.2667885
## CBD                                0.18176454  1.0000000 -0.1686462
## MINORITY                          -0.26678850 -0.1686462  1.0000000
## EDU                                0.26077076  0.1216539 -0.3768247
## trips                              0.21655885  0.5269037 -0.6369958
##                                          EDU      trips
## CTA_BUS_STATIONS                  0.11191158  0.4300325
## CTA_TRAIN_STATIONS                0.09380398  0.5262204
## BIKE_ROUTES                       0.12139609  0.5111209
## Limited_Business_License          0.14236166  0.5606130
## Retail_Food_Establishment         0.15805181  0.5931837
## CAPACITY                          0.18994029  0.5944283
## POPULATION_SQ_MILE                0.26077076  0.2165588
## CBD                               0.12165392  0.5269037
## MINORITY                         -0.37682469 -0.6369958
## EDU                               1.00000000  0.3584473
## trips                             0.35844727  1.0000000
```

```
plot(com1.2)
library(MASS)
tran1=cbind(log(com1.2[,c(1:6,9)]),com1.2[,7:8],com1.2[,10:11])
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

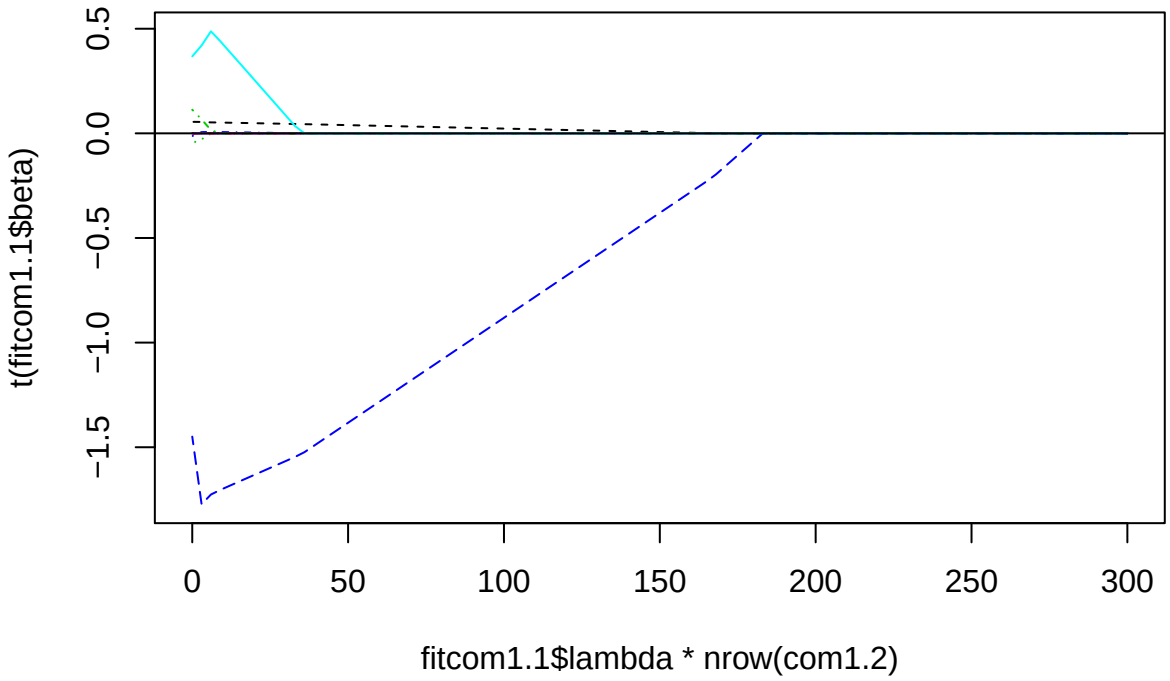
```
## Loaded glmnet 2.0-16
```




```
lam = seq(0,300,length=101)/nrow(com1.2)
x = model.matrix(trips~., com1.2)
fitcom1.1=glmnet(x,com1.2$trips,alpha=1, lambda = lam)
cv.lasso=cv.glmnet(x, com1.2$trips, alpha=1, lambda = lam)
coef(cv.lasso)
```

```
## 12 x 1 sparse Matrix of class "dgCMatrix"
##                                        1
## (Intercept)                9.101224e+00
## (Intercept)                .
## CTA_BUS_STATIONS          -7.149317e-04
## CTA_TRAIN_STATIONS        -2.856942e-02
## BIKE_ROUTES                5.899240e-03
## Limited_Business_License   .
## Retail_Food_Establishment  8.399333e-06
## CAPACITY                   5.443639e-02
## POPULATION_SQ_MILE         5.970144e-07
## CBD                        6.655048e-02
## MINORITY                  -1.772462e+00
## EDU                        4.199945e-01
```

```
matplot(fitcom1.1$lambda*nrow(com1.2), t(fitcom1.1$beta), type="l"); abline(h=0)
```



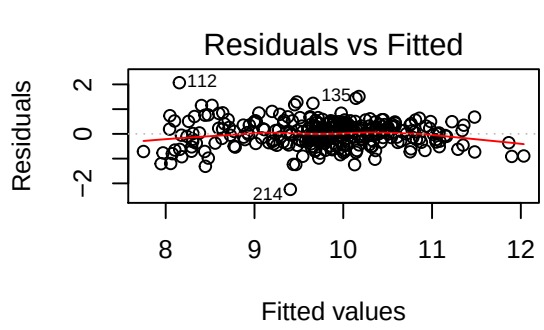

```
fit2.1= lm(trips~ CTA_TRAIN_STATIONS + BIKE_ROUTES + Retail_Food_Establishment + CAPACITY +I
summary(fit2.1)
```

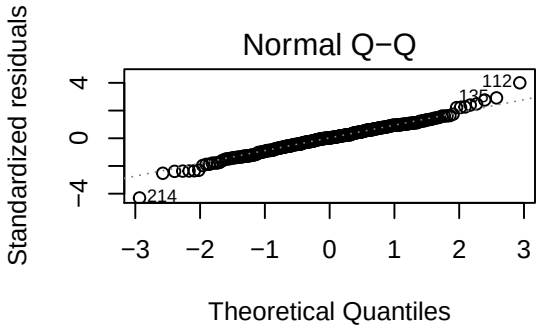
```
##
## Call:
## lm(formula = trips ~ CTA_TRAIN_STATIONS + BIKE_ROUTES + Retail_Food_Establishment +
##     CAPACITY + I(MINORITY^2) + EDU + CBD, data = com1.2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16160 -0.33107  0.01739  0.35924  2.03863
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 8.675e+00  2.718e-01  31.916  < 2e-16 ***
## CTA_TRAIN_STATIONS         -5.252e-02  1.421e-02  -3.695 0.000263 ***
## BIKE_ROUTES                 7.460e-03  6.967e-03   1.071 0.285211
## Retail_Food_Establishment   9.502e-06  2.246e-06   4.230 3.13e-05 ***
## CAPACITY                    5.604e-02  7.354e-03   7.621 3.54e-13 ***
## I(MINORITY^2)              -1.737e+00  1.212e-01 -14.327  < 2e-16 ***
## EDU                         4.627e-01  2.614e-01   1.770 0.077768 .
## CBD                         1.259e-01  1.637e-01   0.769 0.442588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5221 on 292 degrees of freedom
## Multiple R-squared:  0.7108, Adjusted R-squared:  0.7039
## F-statistic: 102.5 on 7 and 292 DF,  p-value: < 2.2e-16
```

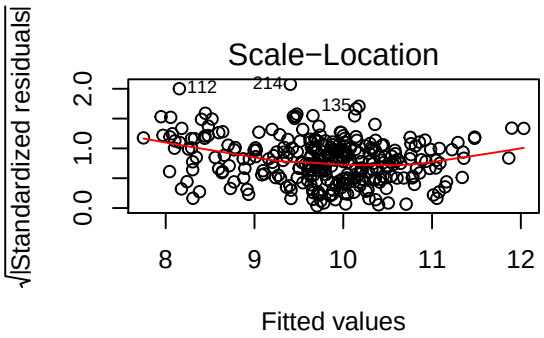
```r
fit2.2= lm(trips~ CTA_TRAIN_STATIONS + Retail_Food_Establishment + CAPACITY +I(MINORITY^2) ,
summary(fit2.2)
```

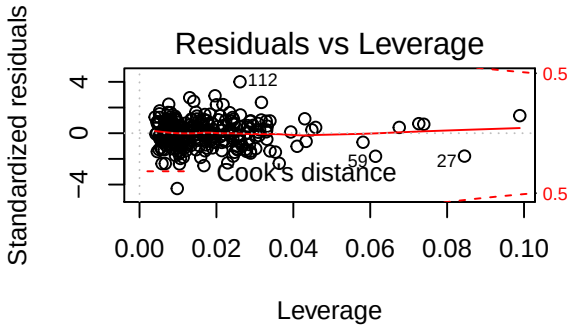
```
##
## Call:
## lm(formula = trips ~ CTA_TRAIN_STATIONS + Retail_Food_Establishment +
##     CAPACITY + I(MINORITY^2), data = com1.2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24371 -0.30726  0.01372  0.33852  2.06842
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 9.125e+00  1.203e-01  75.870  < 2e-16 ***
## CTA_TRAIN_STATIONS         -5.744e-02  1.380e-02  -4.162 4.15e-05 ***
## Retail_Food_Establishment   1.166e-05  1.817e-06   6.416 5.56e-10 ***
## CAPACITY                    5.872e-02  7.195e-03   8.162 9.67e-15 ***
## I(MINORITY^2)              -1.810e+00  1.135e-01 -15.945  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5235 on 295 degrees of freedom
## Multiple R-squared:  0.7063, Adjusted R-squared:  0.7024
## F-statistic: 177.4 on 4 and 295 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(fit2.2)
```





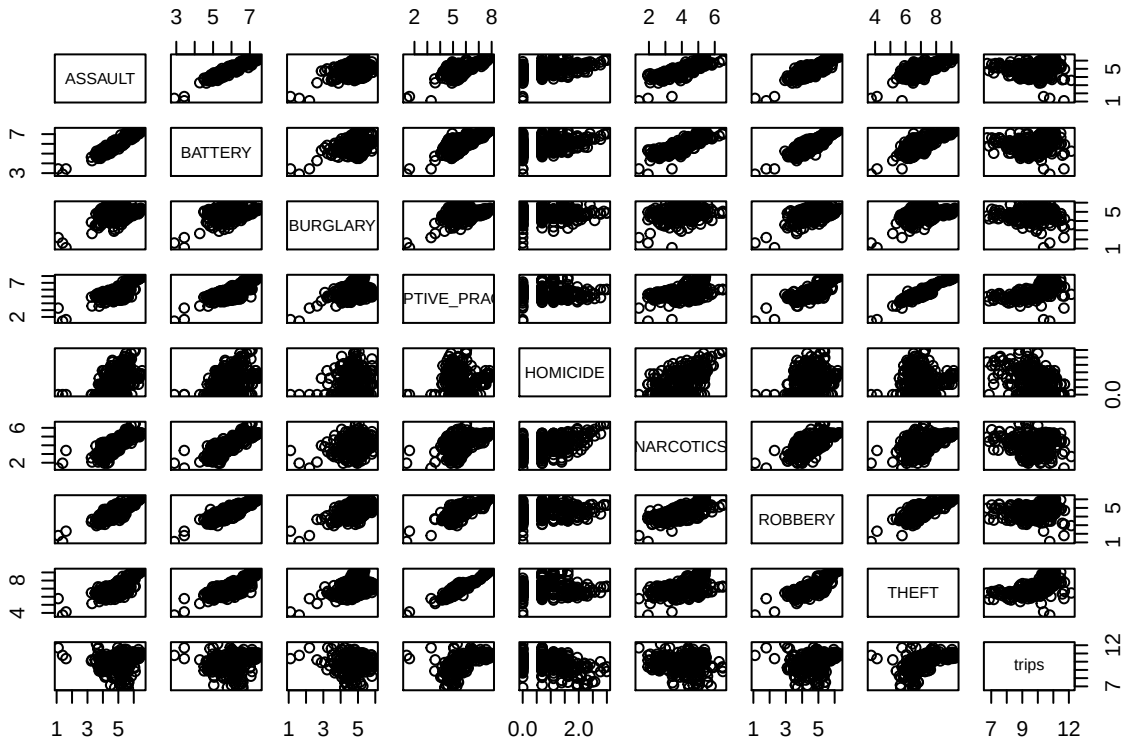





```
library(car)
vif(fit2.2)
```

```
##        CTA_TRAIN_STATIONS Retail_Food_Establishment
##                  8.646809                  8.783578
##                  CAPACITY             I(MINORITY^2)
##                  1.633999                  1.103478
```

```
#crime
bike2<- bike[,c(15,16,17,22,24,31,40,43,45)]
plot(bike2)
```




```
x2 = model.matrix(trips~., bike2)
crime.lasso=cv.glmnet(x2, bike2$trips, alpha=1, lambda = lam)
coef(crime.lasso)
```

```
## 10 x 1 sparse Matrix of class "dgCMatrix"
##                                1
## (Intercept)         7.01991182
## (Intercept)         .
## ASSAULT            -0.67839317
## BATTERY             .
## BURGLARY           -0.24494203
## DECEPTIVE_PRACTICE  0.30282790
## HOMICIDE           -0.12179572
## NARCOTICS           .
## ROBBERY            -0.09351282
## THEFT               0.86462708
```

```
fitcrime1=lm(trips ~ ASSAULT+BURGLARY+DECEPTIVE_PRACTICE+HOMICIDE+THEFT+CAPACITY +I(MINORITY
fitcrime2=lm(trips ~ ASSAULT+BURGLARY+THEFT+CAPACITY +I(MINORITY^2), data=bike)
summary(fitcrime2)
```

```
##
## Call:
## lm(formula = trips ~ ASSAULT + BURGLARY + THEFT + CAPACITY +
##     I(MINORITY^2), data = bike)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.67197 -0.30379  0.02627  0.31190  1.56668
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.10356    0.31836  25.454  < 2e-16 ***
## ASSAULT       -0.38557    0.07115  -5.419 1.25e-07 ***
## BURGLARY      -0.34593    0.05380  -6.429 5.17e-10 ***
## THEFT          0.67074    0.06842   9.804  < 2e-16 ***
## CAPACITY       0.04846    0.00629   7.703 2.04e-13 ***
## I(MINORITY^2) -1.36232    0.14811  -9.198  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4752 on 294 degrees of freedom
## Multiple R-squared:  0.7589, Adjusted R-squared:  0.7548
## F-statistic:   185 on 5 and 294 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(fitcrime2)
```