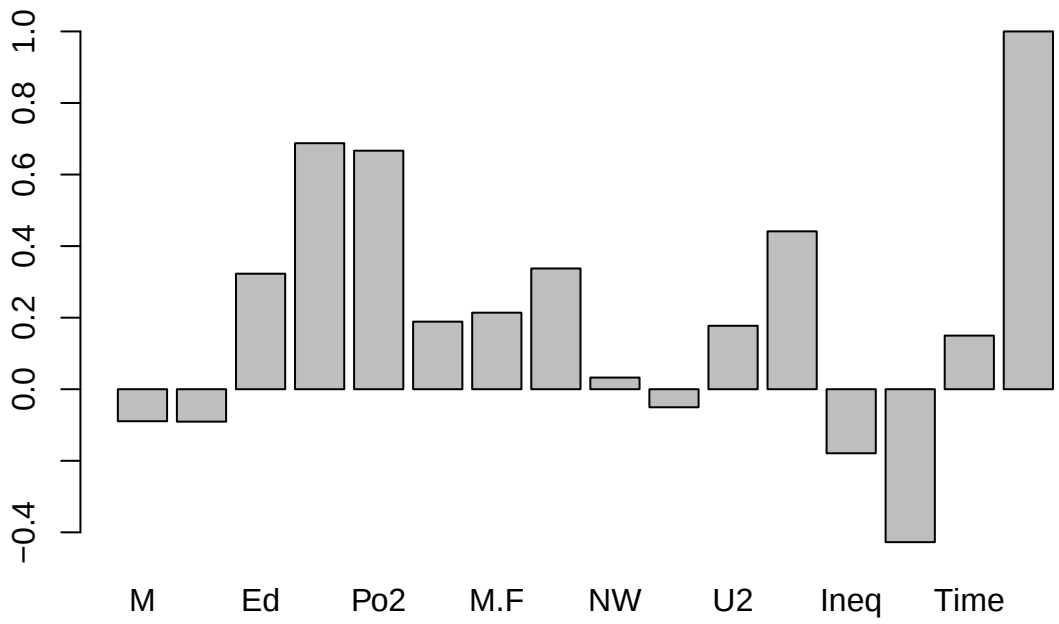
Compiled-Project_Step4.R

Shruti

2020-04-21

```
rm(list=ls())

setwd("~/Desktop/UMASS/Classes/Junior Year Semester 2 Classes/Statistics 525/R_CodeFile_Stats525/IE_Pro
crimerate <- read.csv(file="crime.csv")

#############################################################################################
#To figure out which X works better for the chosen Y (Crime) for this paper,
#We do Pearsons Correlation for the whole data set:

Header_names <- names(crimerate)
increase <- 1
totalFinal_1 <- c()
#pearson:
for (val in Header_names)
{
        mean_y <- mean(crimerate[[Header_names[16]]])
        mean_x1 <- mean(crimerate[[Header_names[increase]]])
        x1_difference <- crimerate[[Header_names[increase]]] - mean_x1
        y_difference <- crimerate[[Header_names[16]]] - mean_y
        upperFinal_1 <- sum(x1_difference * y_difference)
        lowerFinal_1 <- sqrt(sum(x1_difference^2)*sum(y_difference^2))
        totalFinal_1[increase] <- upperFinal_1/lowerFinal_1
        increase <- increase + 1
}

barplot(totalFinal_1, names.arg =  Header_names)
```

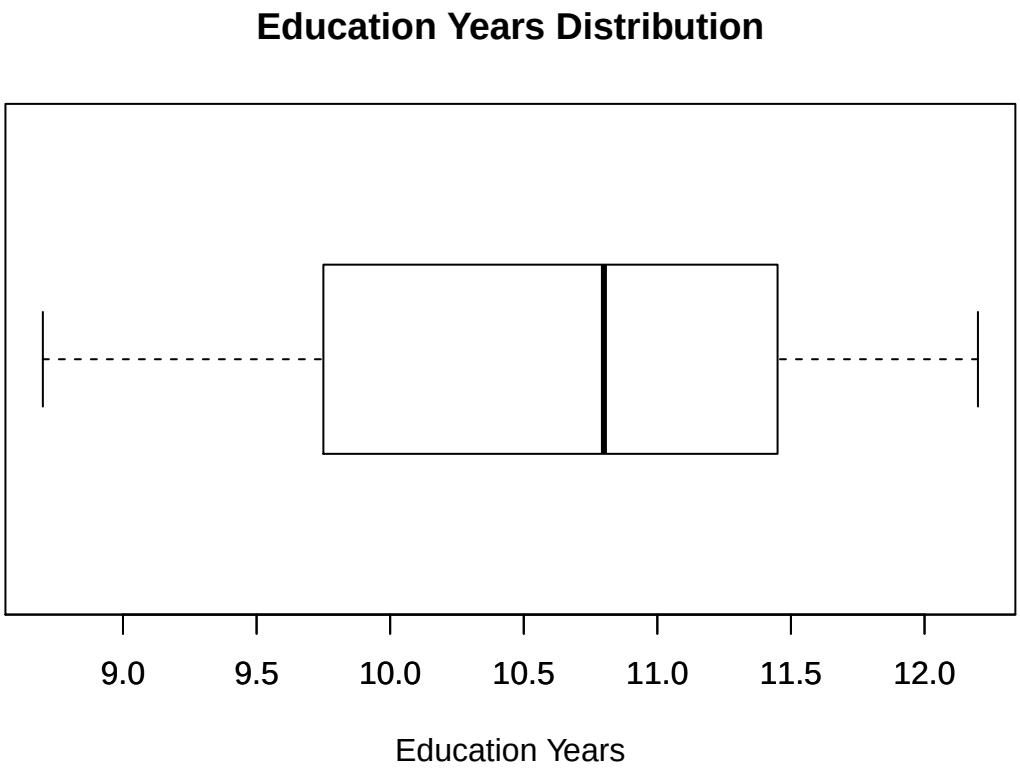



```
##############################################################################
##Original Data Frame

summary(crimerate)
```

```
##        M               So              Ed             Po1
##  Min.   :11.90   Min.   :0.0000   Min.   : 8.70   Min.   : 4.50
##  1st Qu.:13.00   1st Qu.:0.0000   1st Qu.: 9.75   1st Qu.: 6.25
##  Median :13.60   Median :0.0000   Median :10.80   Median : 7.80
##  Mean   :13.86   Mean   :0.3404   Mean   :10.56   Mean   : 8.50
##  3rd Qu.:14.60   3rd Qu.:1.0000   3rd Qu.:11.45   3rd Qu.:10.45
##  Max.   :17.70   Max.   :1.0000   Max.   :12.20   Max.   :16.60
##       Po2              LF              M.F              Pop
##  Min.   : 4.100   Min.   :0.4800   Min.   : 93.40   Min.   :  3.00
##  1st Qu.: 5.850   1st Qu.:0.5305   1st Qu.: 96.45   1st Qu.: 10.00
##  Median : 7.300   Median :0.5600   Median : 97.70   Median : 25.00
##  Mean   : 8.023   Mean   :0.5612   Mean   : 98.30   Mean   : 36.62
##  3rd Qu.: 9.700   3rd Qu.:0.5930   3rd Qu.: 99.20   3rd Qu.: 41.50
##  Max.   :15.700   Max.   :0.6410   Max.   :107.10   Max.   :168.00
##        NW               U1               U2            Wealth
##  Min.   : 0.20   Min.   :0.07000   Min.   :2.000   Min.   :2880
##  1st Qu.: 2.40   1st Qu.:0.08050   1st Qu.:2.750   1st Qu.:4595
##  Median : 7.60   Median :0.09200   Median :3.400   Median :5370
##  Mean   :10.11   Mean   :0.09547   Mean   :3.398   Mean   :5254
##  3rd Qu.:13.25   3rd Qu.:0.10400   3rd Qu.:3.850   3rd Qu.:5915
##  Max.   :42.30   Max.   :0.14200   Max.   :5.800   Max.   :6890
```

```
##       Ineq            Prob              Time          Crime      
##  Min.   :12.60   Min.   :0.00690   Min.   :12.20   Min.   : 342.0
##  1st Qu.:16.55   1st Qu.:0.03270   1st Qu.:21.60   1st Qu.: 658.5
##  Median :17.60   Median :0.04210   Median :25.80   Median : 831.0
##  Mean   :19.40   Mean   :0.04709   Mean   :26.60   Mean   : 905.1
##  3rd Qu.:22.75   3rd Qu.:0.05445   3rd Qu.:30.45   3rd Qu.:1057.5
##  Max.   :27.60   Max.   :0.11980   Max.   :44.00   Max.   :1993.0
```

```r
Y <- crimerate$Crime
X1 <- crimerate$Ed
X2 <- crimerate$NW
X3 <- crimerate$Po1
X4 <- crimerate$Prob
X2t <- log(X2)



################################################################################################
##Boxplot, Outliers and Descriptive Statistics With Original Data Frame

#X1 Boxplot, Outliers and Descriptive Statistics
boxplot(X1, horizontal=TRUE, xlab="Education Years",
        main="Education Years Distribution")
axis(1, at=seq(8,13, by=0.5))
```



```r
summary(X1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.70    9.75   10.80   10.56   11.45   12.20
```

```r
sd(X1)
```

```
## [1] 1.1187
```

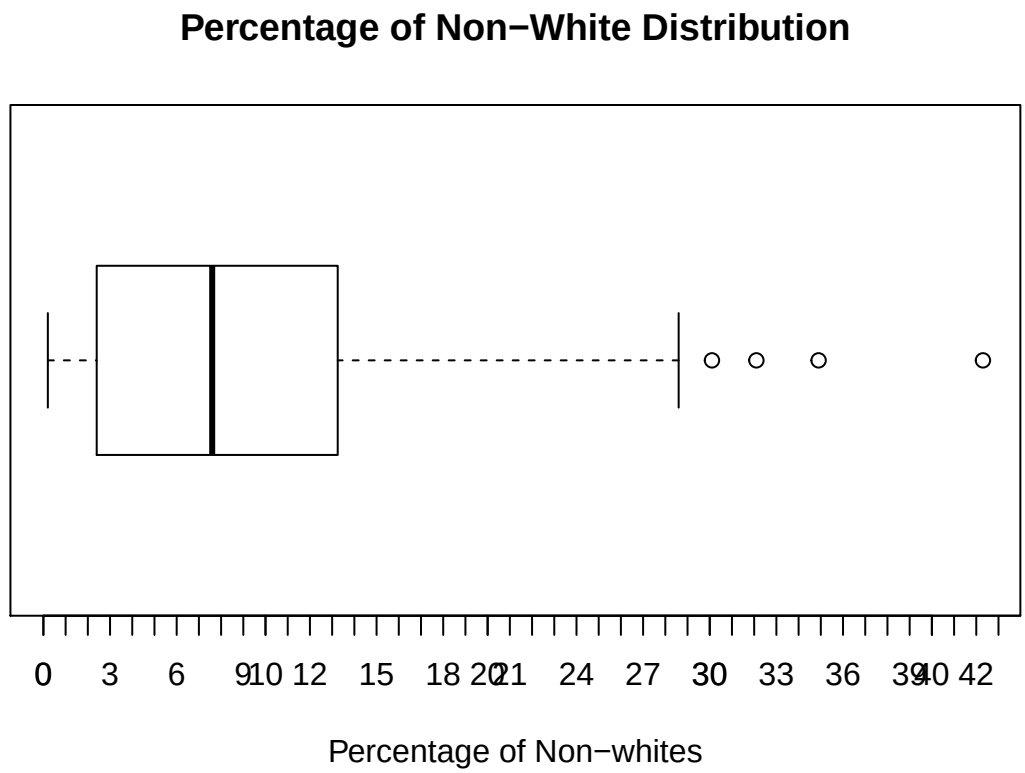
```r
var(X1)
```

```
## [1] 1.251489
```

```r
range(X1)
```

```
## [1]  8.7 12.2
```

```r
#X2 Boxplot, Outliers and Descriptive Statistics
boxplot(X2, horizontal=TRUE, xlab="Percentage of Non-whites",
        main="Percentage of Non-White Distribution")
axis(1, at=seq(0,50, by=1))
```



```r
summary(X2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.20    2.40    7.60   10.11   13.25   42.30
```

```r
sd(X2)
```

```
## [1] 10.28288
```

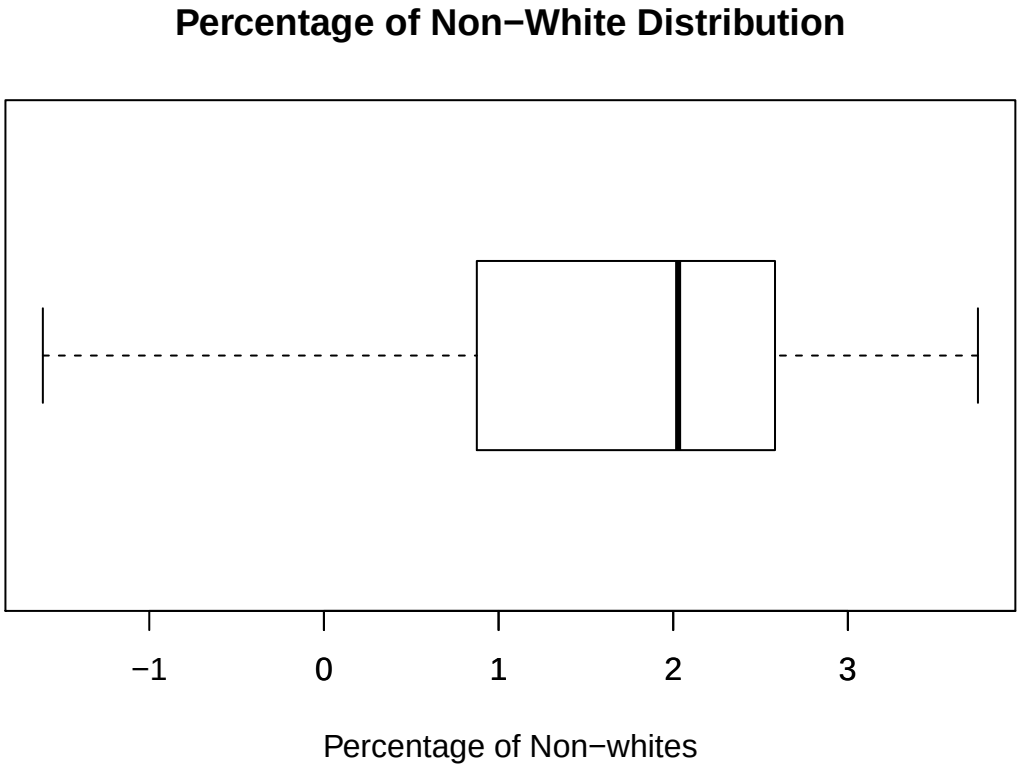
```r
var(X2)
```

```
## [1] 105.7377
```

```r
range(X2)
```

```
## [1]  0.2 42.3
```

```r
#Transformed X2 (X2t) Boxplot, Outliers and Descriptive Statistics
boxplot(X2t, horizontal=TRUE, xlab="Percentage of Non-whites",
        main="Percentage of Non-White Distribution")
axis(1, at=seq(0,50, by=1))
```



```r
summary(X2t)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.6094  0.8755  2.0281  1.7417  2.5828  3.7448
```

```r
sd(X2t)
```

```
## [1] 1.210755
```

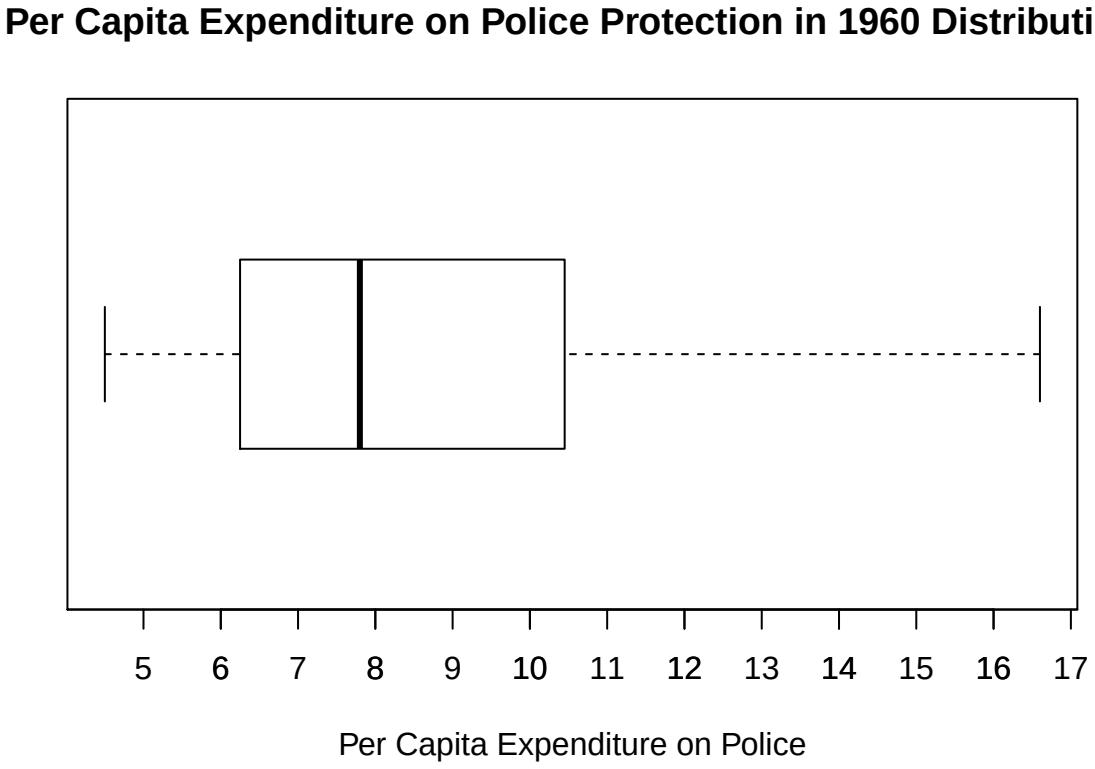
```r
var(X2t)
```

```
## [1] 1.465929
```

```r
range(X2t)
```

```
## [1] -1.609438  3.744787
```

```r
#no outliers in transformed X2t data

#X3 Boxplot, Outliers and Descriptive Statistics
boxplot(X3, horizontal=TRUE, xlab="Per Capita Expenditure on Police",
        main="Per Capita Expenditure on Police Protection in 1960 Distribution")
axis(1, at=seq(4,18, by=1))
```



```r
summary(X3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.50    6.25    7.80    8.50   10.45   16.60
```

```r
sd(X3)
```

```
## [1] 2.971897
```

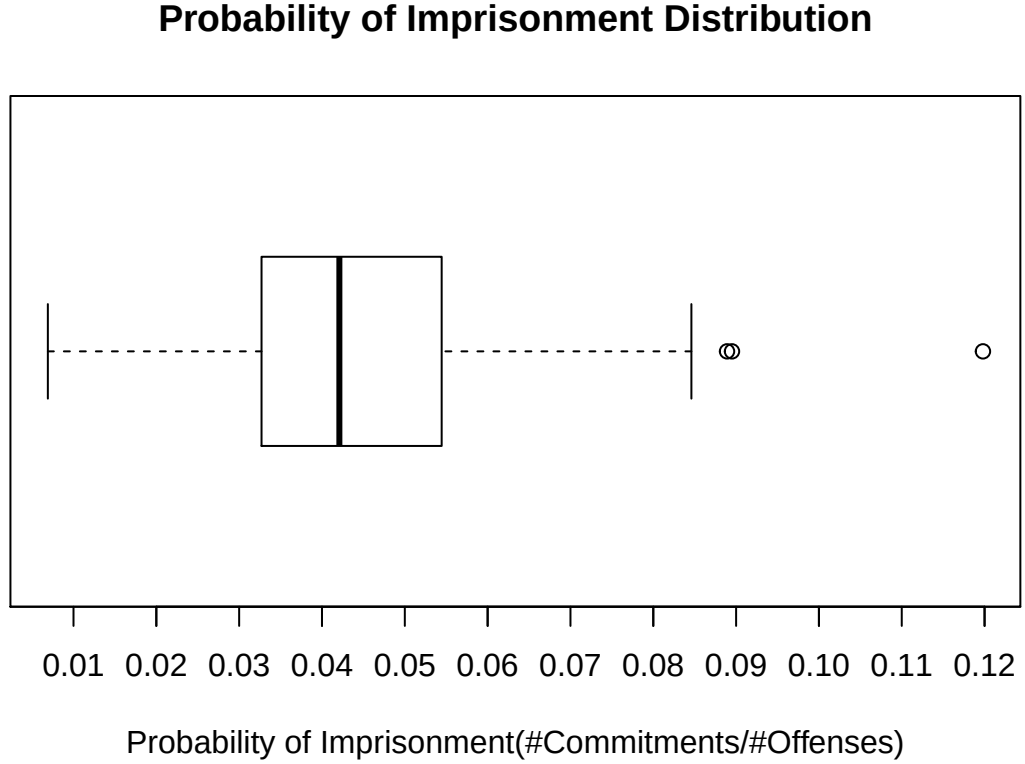
```r
var(X3)
```

```
## [1] 8.832174
```

```r
range(X3)
```

```
## [1]  4.5 16.6
```

```r
#X4 Boxplot, Outliers and Descriptive Statistics
boxplot(X4, horizontal=TRUE, xlab="Probability of Imprisonment(#Commitments/#Offenses)",
        main="Probability of Imprisonment Distribution")
axis(1, at=seq(0,0.13, by=0.01))
```



```r
summary(X4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00690 0.03270 0.04210 0.04709 0.05445 0.11980
```

```r
sd(X4)
```

```
## [1] 0.02273697
```

```r
var(X4)
```

```
## [1] 0.0005169699
```

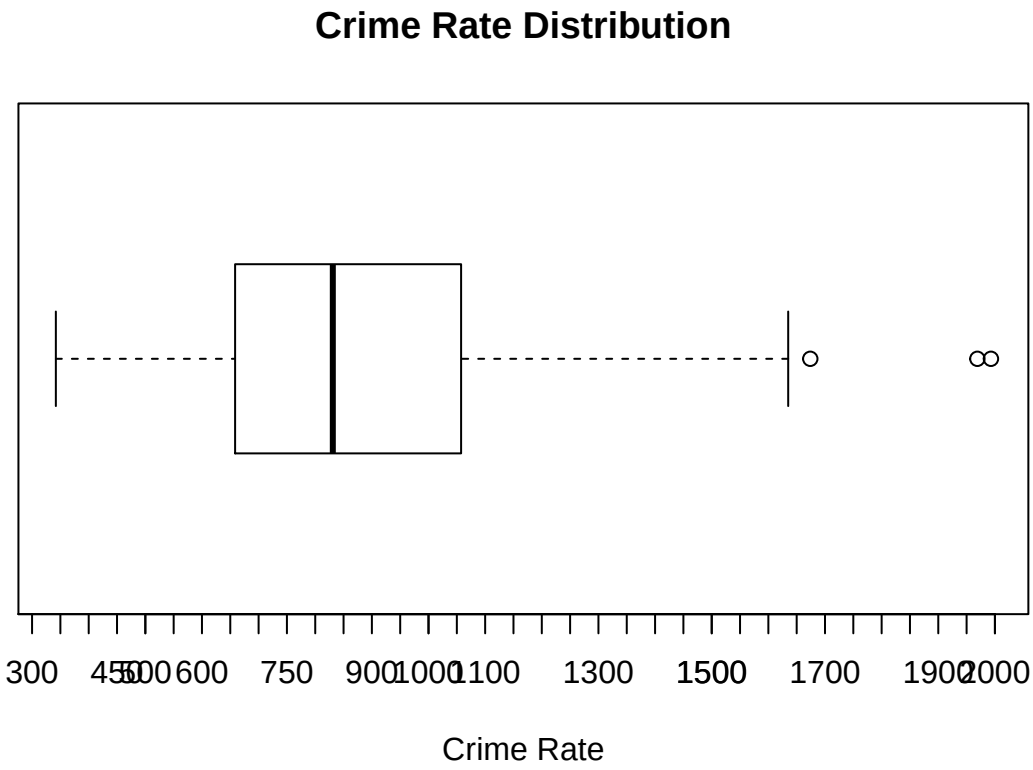
```r
range(X4)
```

```
## [1] 0.006900 0.119804
```

```r
outlierX4 <- which(X4 > 0.085 & Y[18] & Y[22] & Y[42])
outlierX4
```

```
## [1] 18 22 42
```

```r
#Y Boxplot, Outliers and Descriptive Statistics
boxplot(Y, horizontal=TRUE, xlab="Crime Rate",
        main="Crime Rate Distribution")
axis(1, at=seq(0,2000, by=50))
```



```r
summary(Y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   342.0   658.5   831.0   905.1  1057.5  1993.0
```

```r
sd(Y)
```

```
## [1] 386.7627
```

```r
var(Y)
```

```
## [1] 149585.4
```

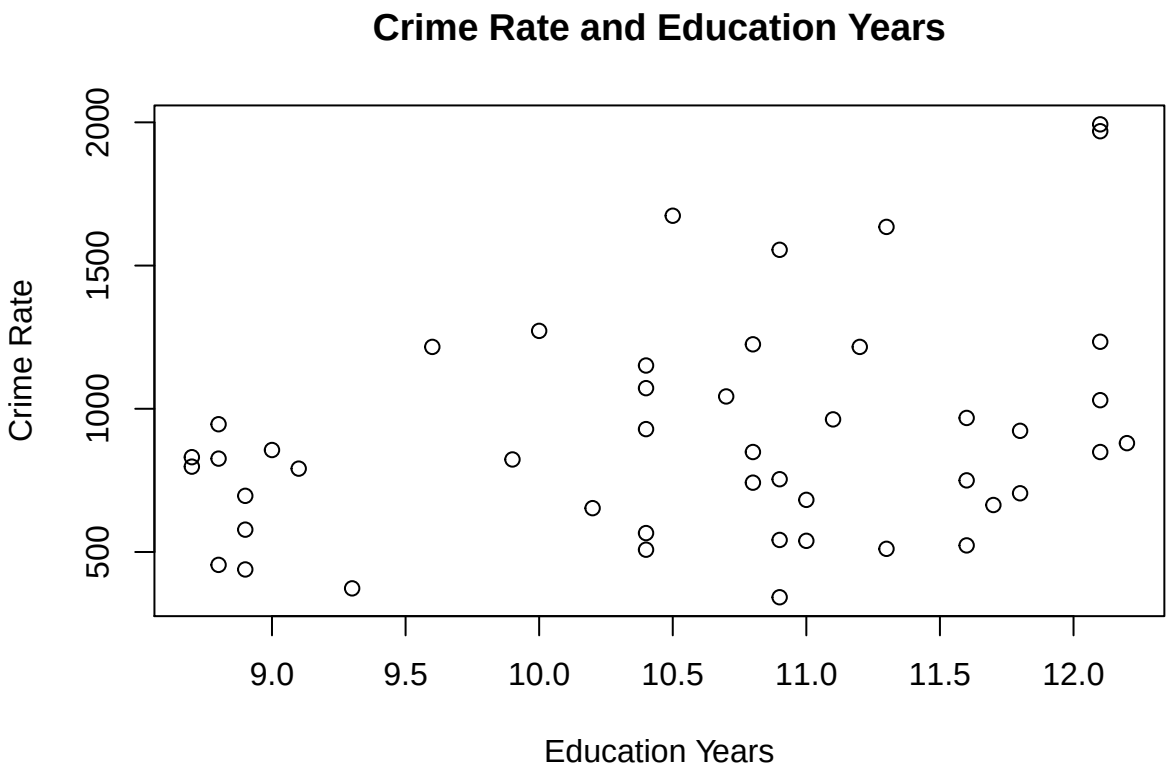
```r
range(Y)
```

```
## [1]  342 1993
```

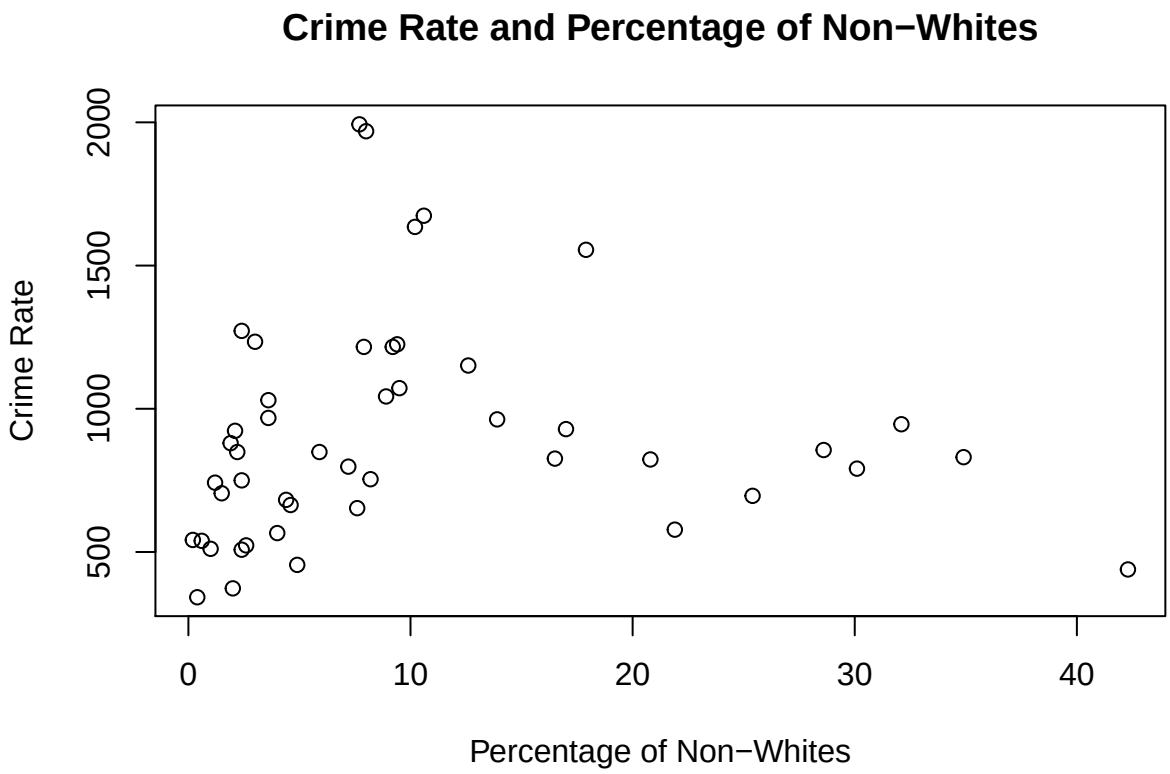
```r
outlierY <- which(
                  Y > 1650 & X1[4] & X2[4] & X3[4] & X4[4]
                  & X1[11] & X2[11] & X3[11] & X4[11]
                  & X1[26] & X2[26] & X3[26] & X4[26]
                  )
outlierY
```

```
## [1]  4 11 26
```

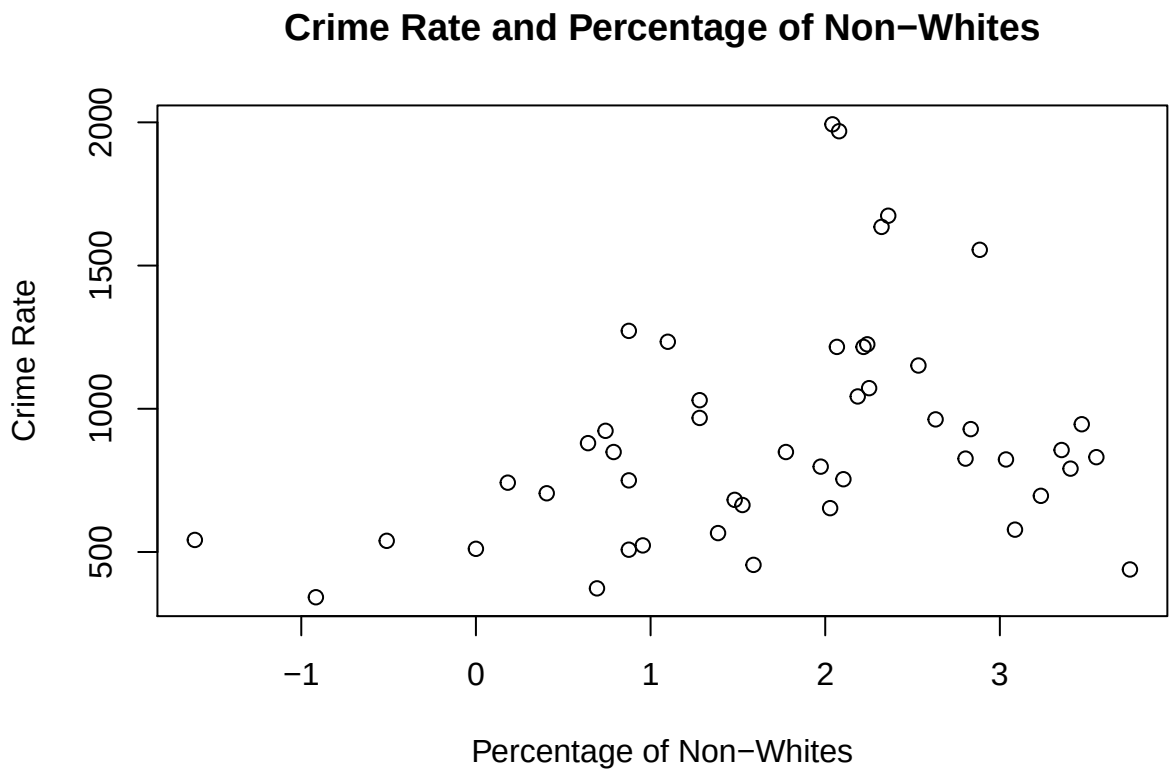
```r
##############################################################################################
##Scatterplots With Original Data Frame & Matrices

#X1 Scatter Plot
plot(X1, Y, xlab="Education Years", ylab="Crime Rate",
     main="Crime Rate and Education Years")
```



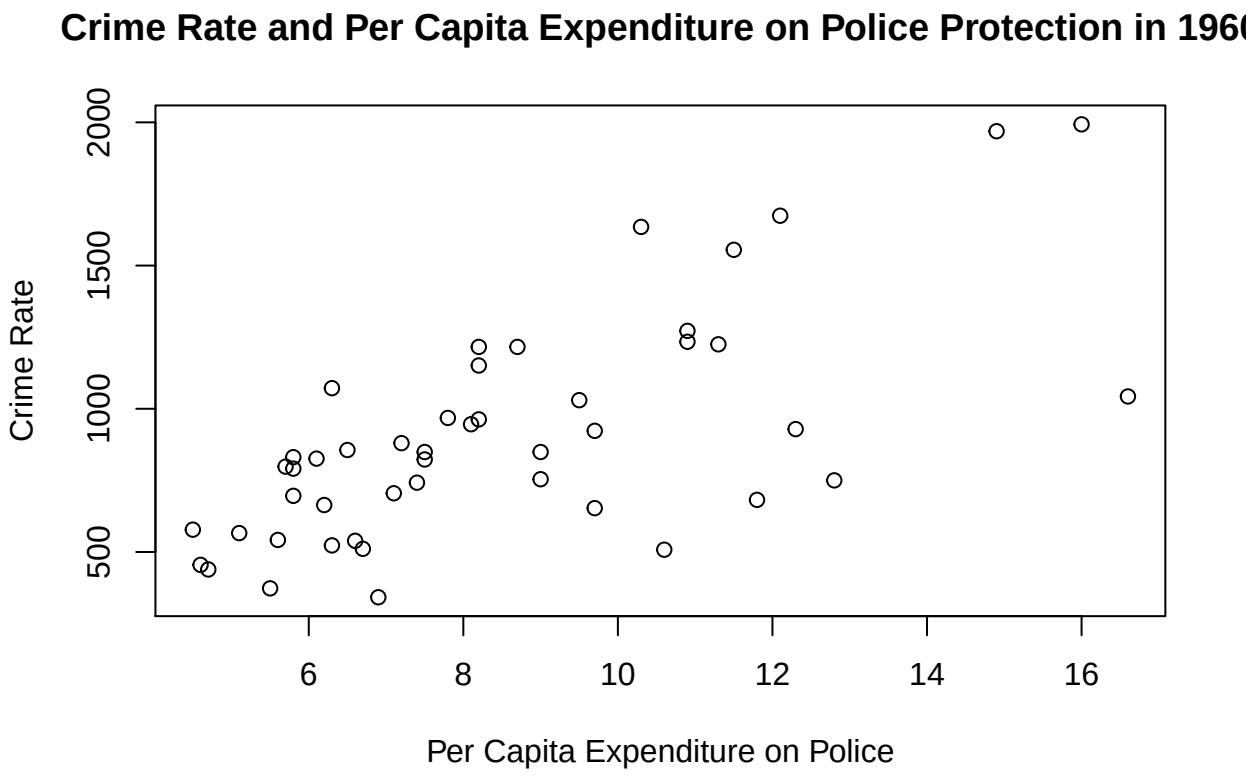

```
#X2 Scatter Plot
plot(X2, Y, xlab="Percentage of Non-Whites", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites")
```



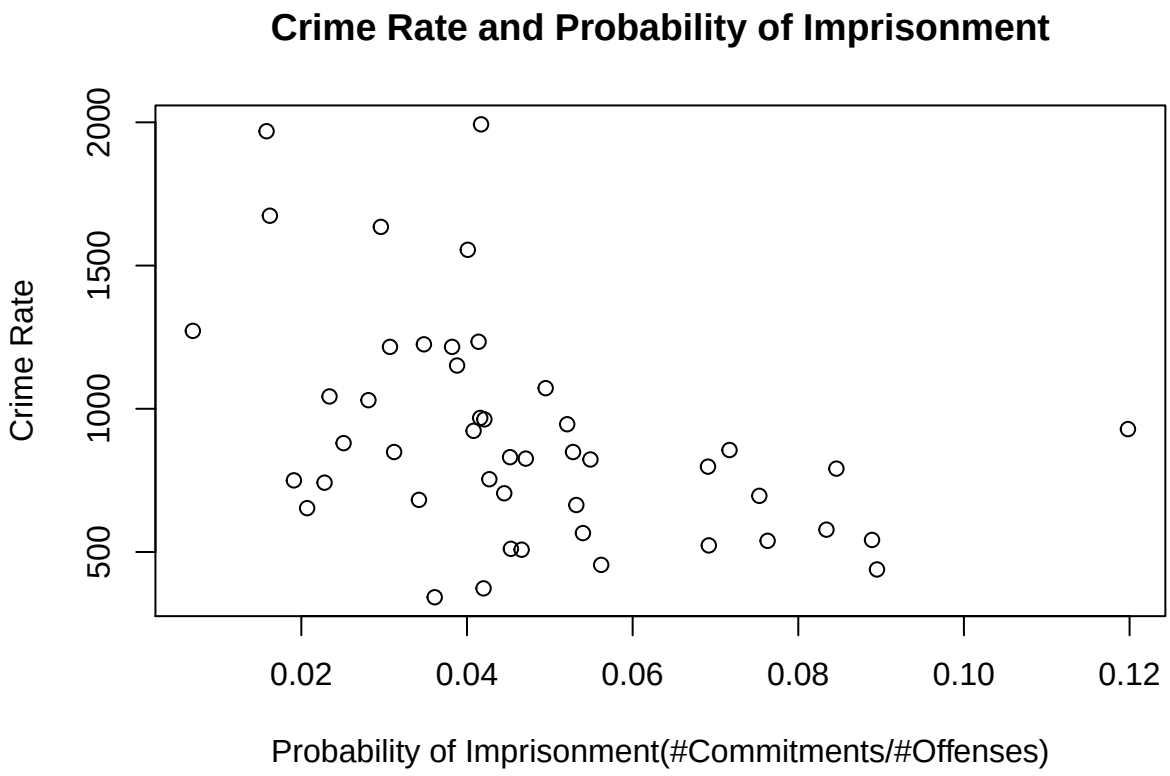

```
#Transformed X2 (X2t) Scatter Plot
plot(X2t, Y, xlab="Percentage of Non-Whites", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites")
```




```
#X3 Scatter Plot
plot(X3, Y, xlab="Per Capita Expenditure on Police", ylab="Crime Rate",
     main="Crime Rate and Per Capita Expenditure on Police Protection in 1960")
```

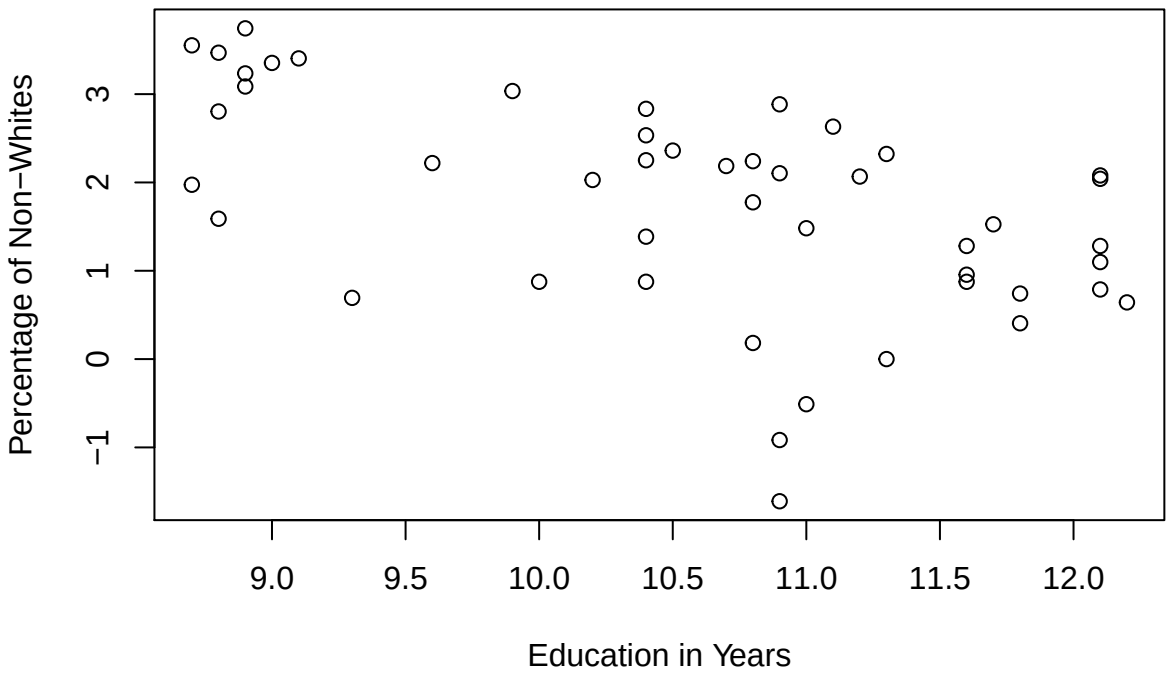



```
#X4 Scatter Plot
plot(X4, Y, xlab="Probability of Imprisonment(#Commitments/#Offenses)", ylab="Crime Rate",
     main="Crime Rate and Probability of Imprisonment")
```




```
#X1 and X2 transformed
plot(X1, X2t, xlab="Education in Years", ylab="Percentage of Non-Whites",
     main="Figure 9-Percentage of Non-Whites and Education in Years")
```

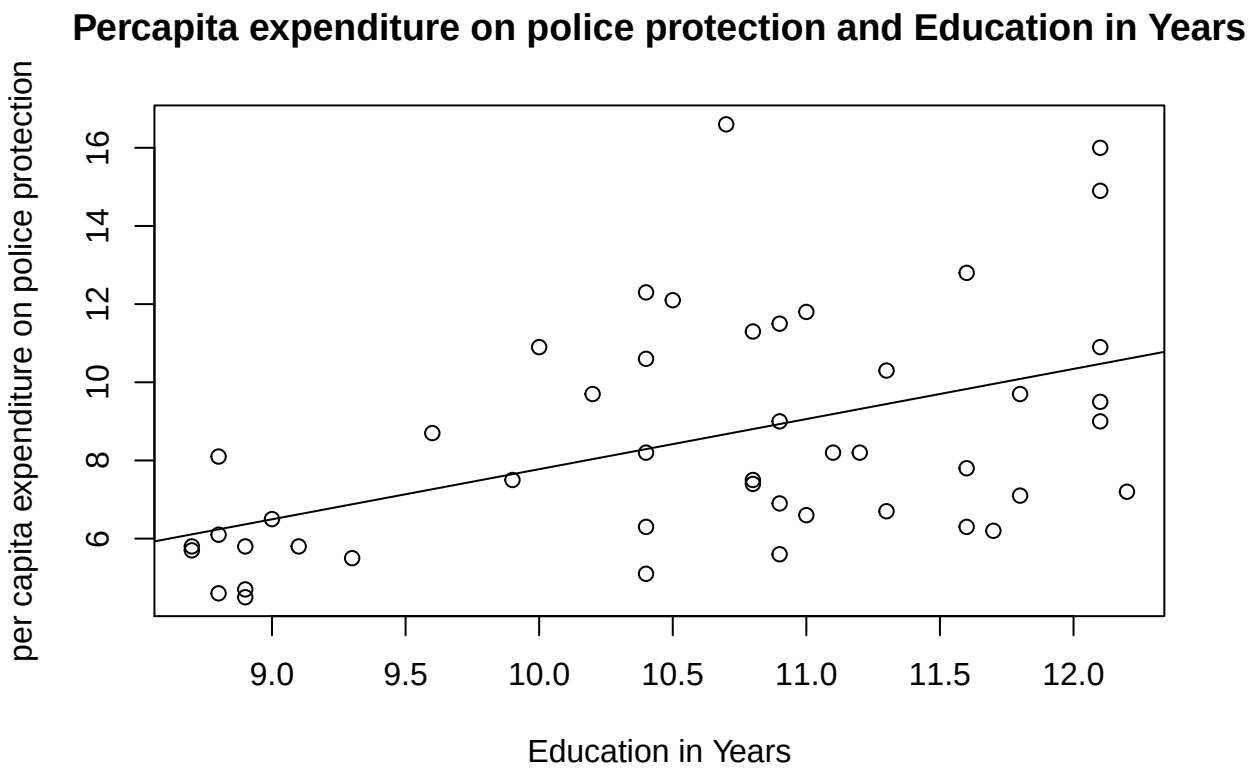
Figure 9−Percentage of Non−Whites and Education in Years




```
cor(X1, X2t, use="all.obs", method="pearson")
```

```
## [1] -0.5388506
```

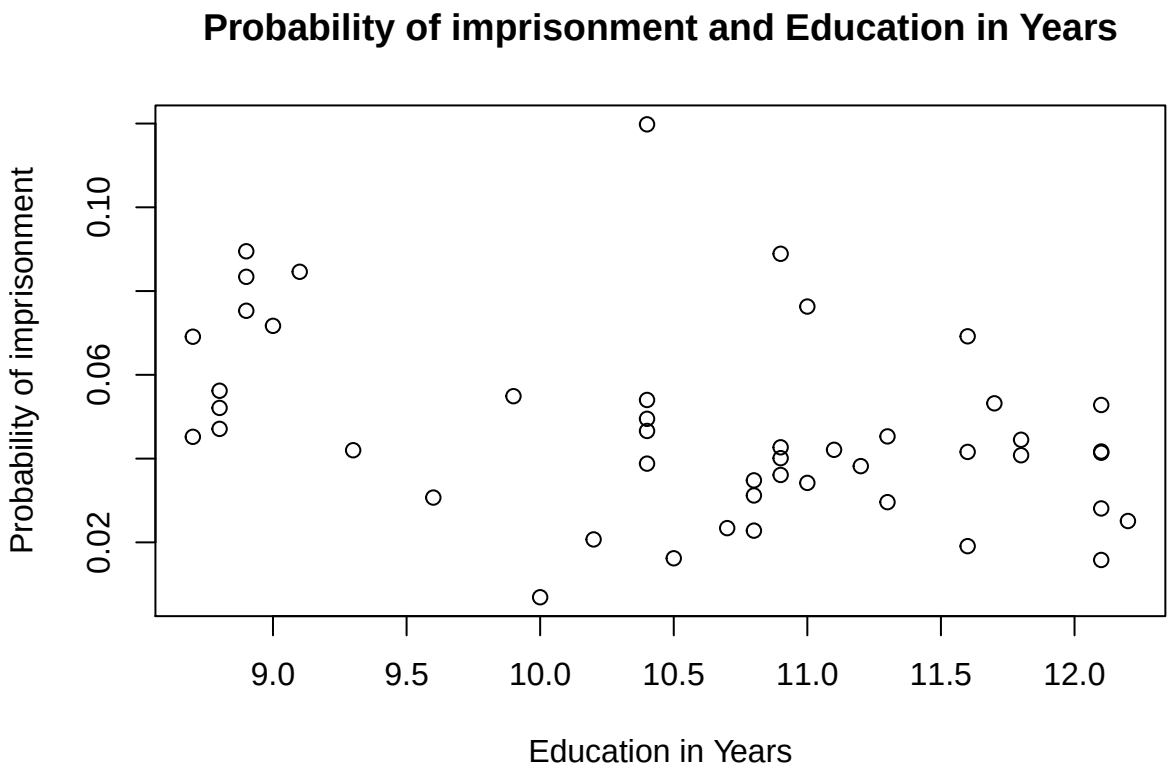
```
#there is a slight downward slope but no significant relationship betwee the two variables

#X1 and X3
plot(X1, X3, xlab="Education in Years", ylab="per capita expenditure on police protection",
     main="Percapita expenditure on police protection and Education in Years")
abline(lm(crimerate$Po1~crimerate$Ed))
```




```
cor(X1, X3, use="all.obs", method="pearson")
```

```
## [1] 0.4829521
```

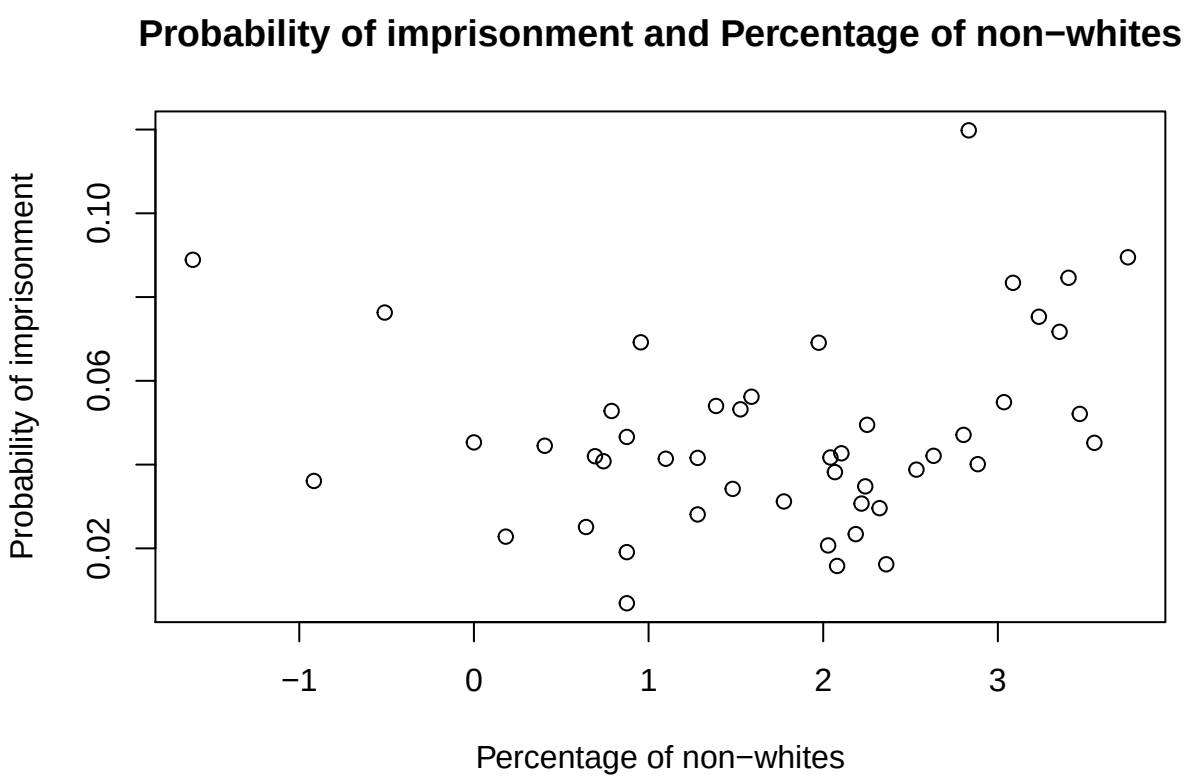
```
#there is a weak positive corelation.

#X1 and X4
plot(X1, X4, xlab="Education in Years", ylab="Probability of imprisonment",
     main="Probability of imprisonment and Education in Years")
```




```
cor(X1, X4, use="all.obs", method="pearson")
```

```
## [1] -0.3899229
```

```
#NO significant relationship

#X2 transformed and X4
plot(X2t, X4, xlab="Percentage of non-whites", ylab="Probability of imprisonment",
     main="Probability of imprisonment and Percentage of non-whites")
```

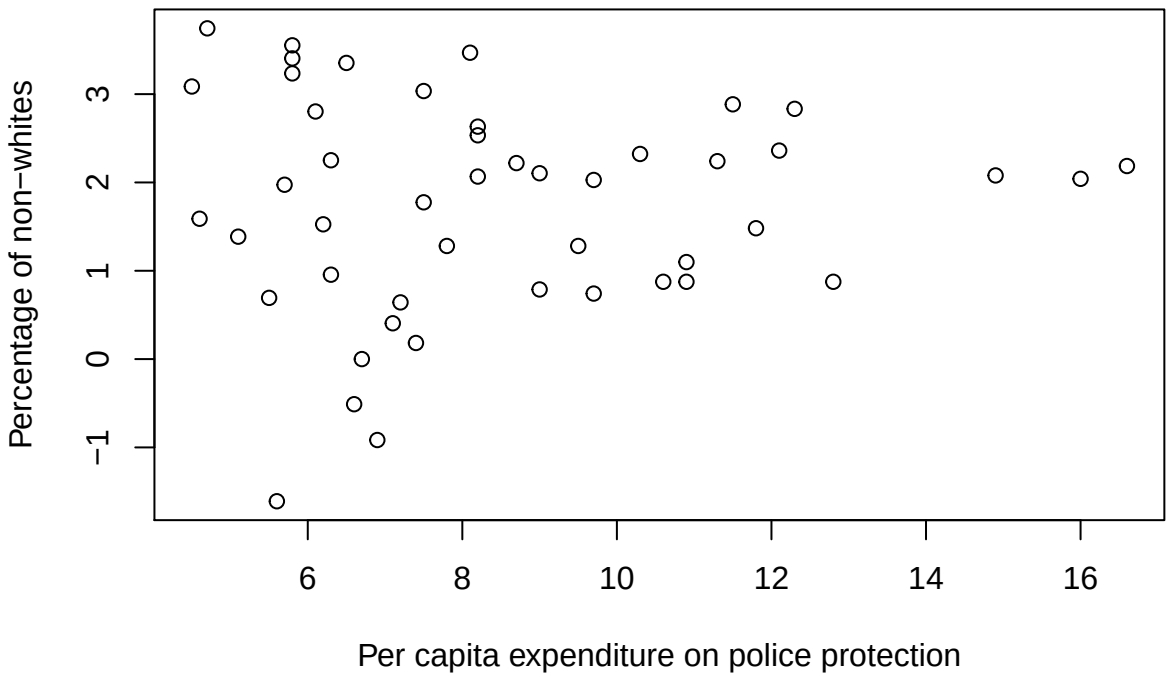



```
cor(X2t, X4, use="all.obs", method="pearson")
```

```
## [1] 0.1854358
```

```
#no significant relationship

#X3 and transformed X2
plot(X3, X2t, ylab="Percentage of non-whites", xlab="Per capita expenditure on police protection",
     main="Figure 9-Per capita expenditure on police protection and Percentage of non-whites")
```

re 9−Per capita expenditure on police protection and Percentage of no

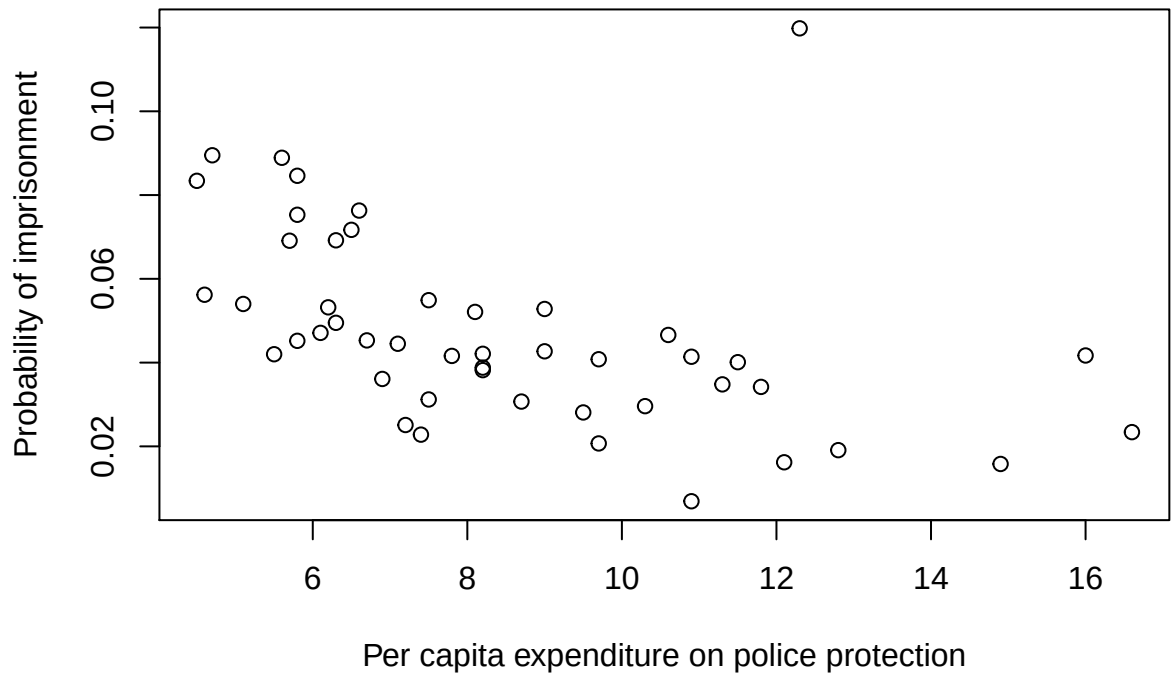



```
cor(X3, X2t, use="all.obs", method="pearson")
```

```
## [1] 0.03654822
```

```
#no significant relationship

#X3 and X4
plot(crimerate$Po1, crimerate$Prob, xlab="Per capita expenditure on police protection", ylab="Probabili
     main="Figure 11-Probability of imprisonment and Per capita expenditure on police protection")
```

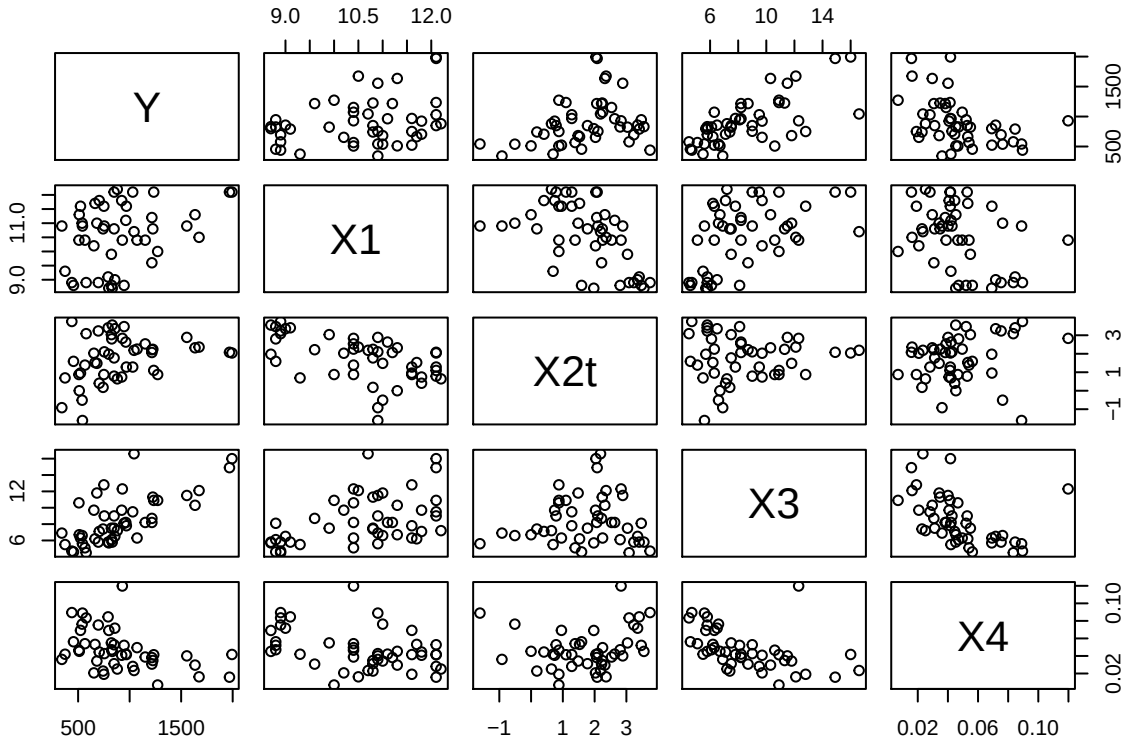
e 11−Probability of imprisonment and Per capita expenditure on police




```
cor(X3, X4, use="all.obs", method="pearson")
```

```
## [1] -0.473247
```

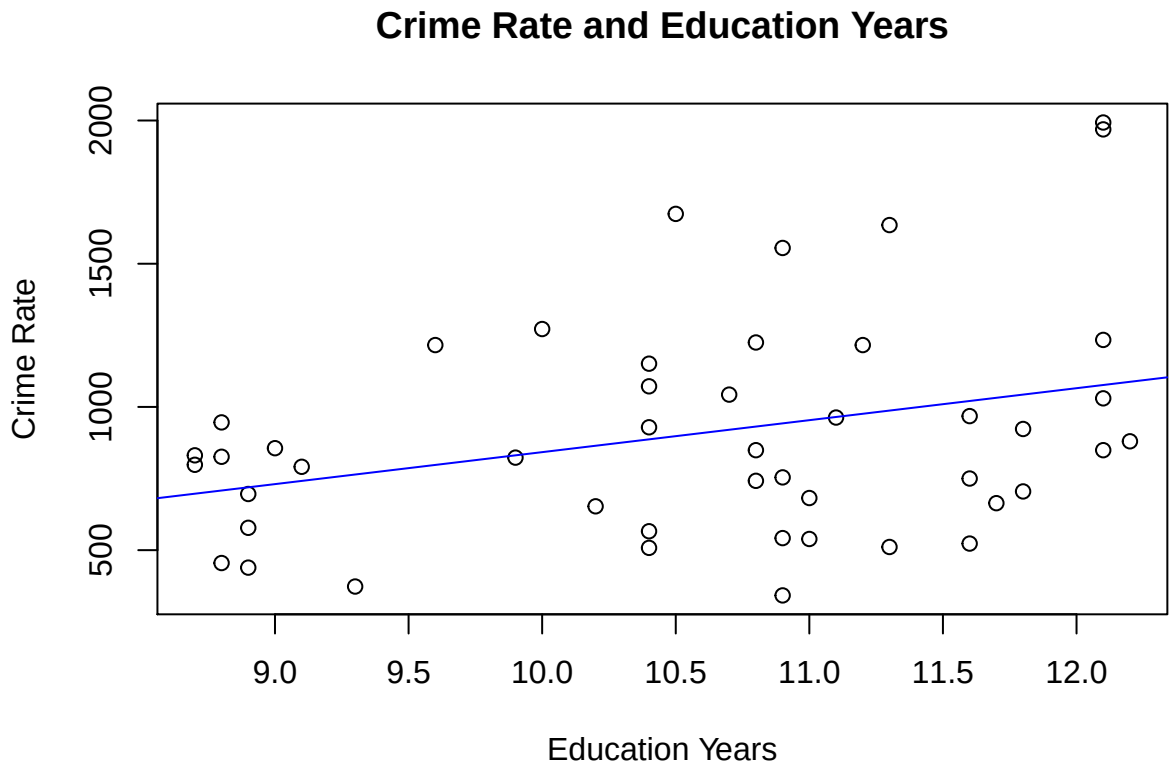
```
#as the percapita expenditure goes up the probability of imprisonment goes down

#Scatterplot Matrix with transformed X2
pairs(~Y+X1+X2t+X3+X4, data=crimerate)
```




```
##########################################################################################
##Linear Regressions & Pearson Correlations For X1, X2, X2 Transformed,
#X3 and x4 With Original Data Frame

#X1 Linear Regression
Y_X1 <- lm(Y ~ X1)
summary(Y_X1)
```

```
##
## Call:
## lm(formula = Y ~ X1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -600.61 -271.25  -46.54  171.33  916.46
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -273.97     518.10  -0.529   0.5996
## X1            111.61      48.78   2.288   0.0269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 370.1 on 45 degrees of freedom
## Multiple R-squared:  0.1042, Adjusted R-squared:  0.08432
## F-statistic: 5.236 on 1 and 45 DF,  p-value: 0.02688
```

```r
plot(Y~X1, xlab="Education Years", ylab="Crime Rate",
     main="Crime Rate and Education Years")
abline(Y_X1, col="blue")
```




```r
cor(X1, Y, use="all.obs", method="pearson")
```

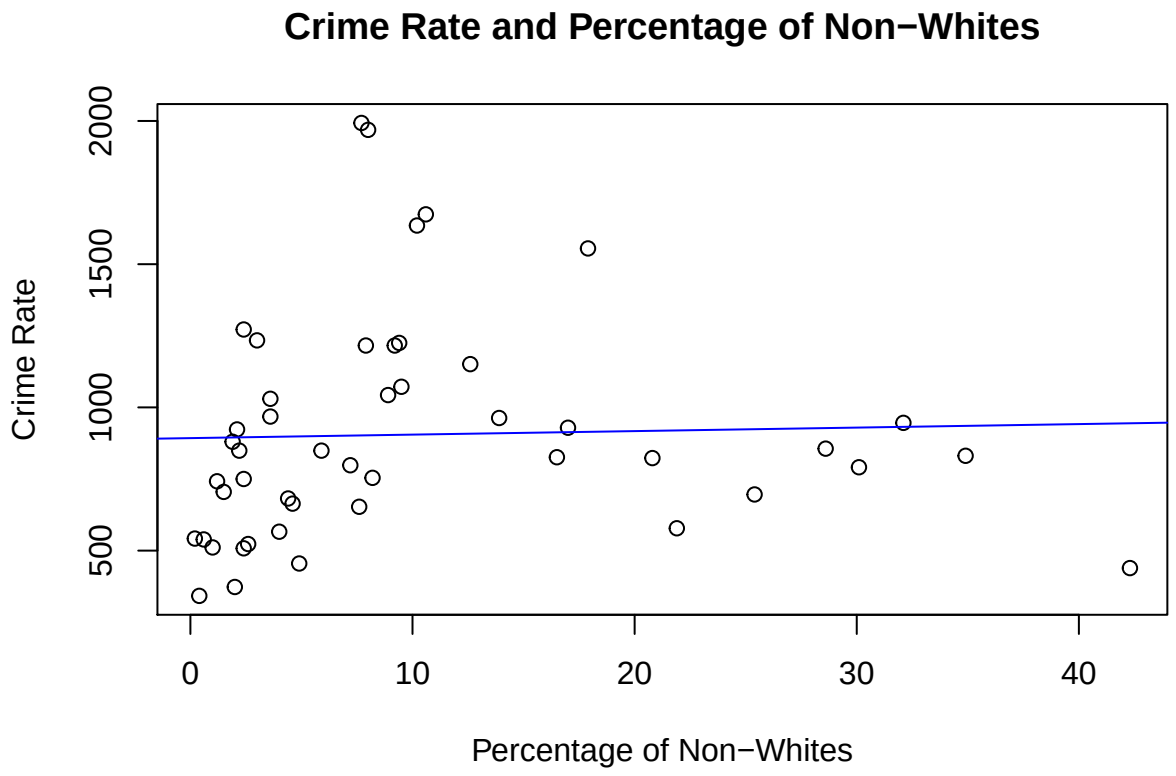
```
## [1] 0.3228349
```

```r
#X2 Linear Regression
Y_X2 <- lm(Y ~ X2)
summary(Y_X2)
```

```
##
## Call:
## lm(formula = Y ~ X2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -551.18 -241.66  -86.92  153.53 1090.87
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  892.686     80.384  11.105 1.76e-14 ***
## X2             1.226      5.604   0.219    0.828
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 390.8 on 45 degrees of freedom
## Multiple R-squared:  0.001063,   Adjusted R-squared:  -0.02114
## F-statistic: 0.04787 on 1 and 45 DF,  p-value: 0.8278
```

```r
plot(Y~X2, xlab="Percentage of Non-Whites", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites")
abline(Y_X2, col="blue")
```




```r
cor(X2, Y, use="all.obs", method="pearson")
```

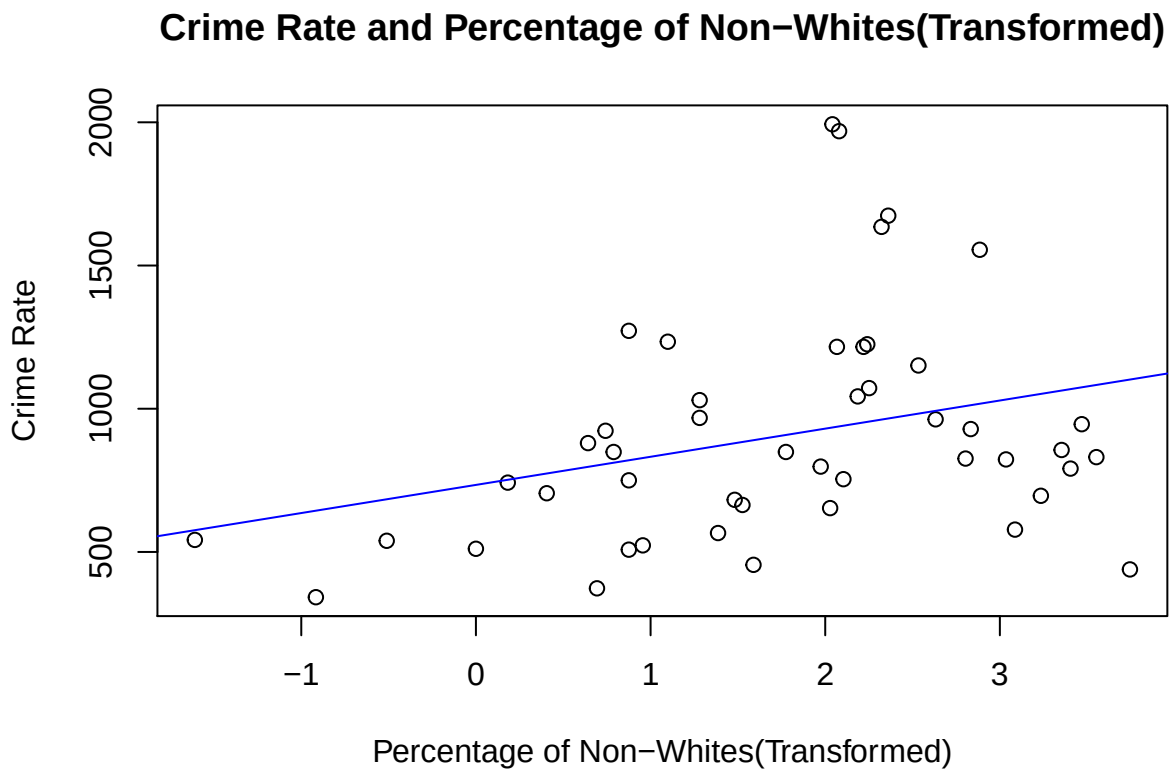
```
## [1] 0.03259884
```

```r
#Transformed X2 (X2t) Linear Regression
X2t <- log(X2)
Y_X2t <- lm(Y ~ X2t)
summary(Y_X2t)
```

```
##
## Call:
## lm(formula = Y ~ X2t)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -662.83 -237.48  -70.01  142.49 1058.49
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   734.02      95.78   7.663 1.07e-09 ***
## X2t            98.22      45.31   2.168   0.0355 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 372.1 on 45 degrees of freedom
## Multiple R-squared:  0.09454,    Adjusted R-squared:  0.07442
## F-statistic: 4.699 on 1 and 45 DF,  p-value: 0.03551
```

```
plot(Y~X2t, xlab="Percentage of Non-Whites(Transformed)", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites(Transformed)")
abline(Y_X2t, col="blue")
```



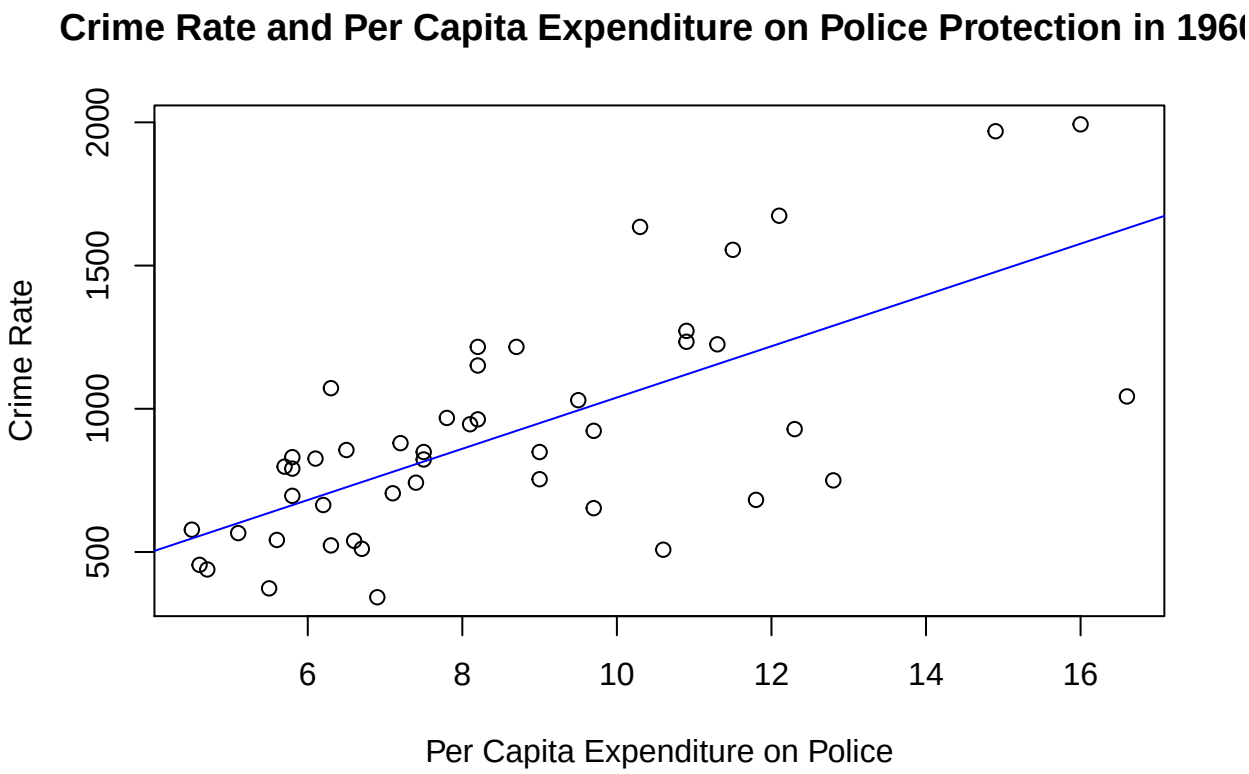

```
cor(X2t, Y, use="all.obs", method="pearson")
```

```
## [1] 0.3074796
```

```r
#X3 Linear Regression
Y_X3 <- lm(Y ~ X3)
summary(Y_X3)
```

```
##
## Call:
## lm(formula = Y ~ X3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -586.91 -155.63   32.52  139.58  568.84
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   144.46     126.69   1.140     0.26
## X3             89.48      14.09   6.353 9.34e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 283.9 on 45 degrees of freedom
## Multiple R-squared:  0.4728, Adjusted R-squared:  0.4611
## F-statistic: 40.36 on 1 and 45 DF,  p-value: 9.338e-08
```

```r
plot(Y~X3, xlab="Per Capita Expenditure on Police", ylab="Crime Rate",
     main="Crime Rate and Per Capita Expenditure on Police Protection in 1960")
abline(Y_X3, col="blue")
```




```
cor(X3, Y, use="all.obs", method="pearson")
```

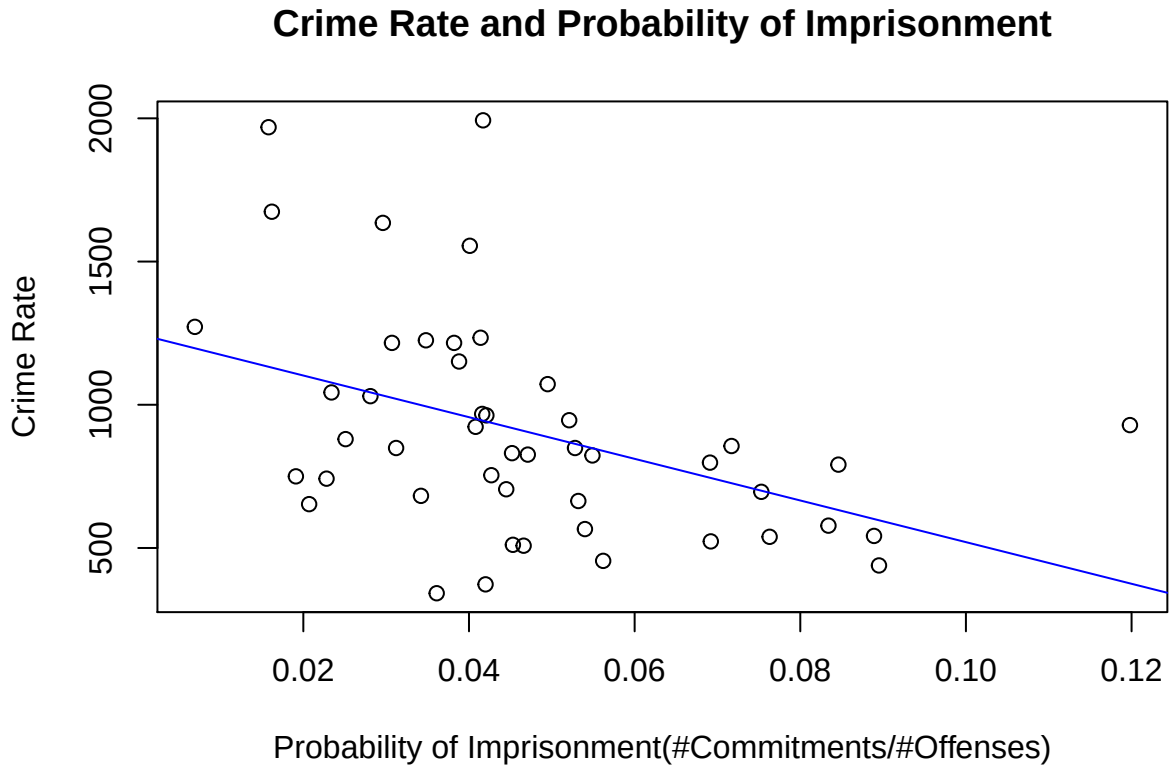
```
## [1] 0.6876045
```

```
#X4 Linear Regression
Y_X4 <- lm(Y ~ X4)
summary(Y_X4)
```

```
##
## Call:
## lm(formula = Y ~ X4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -643.01 -207.81  -27.83  171.53 1048.70
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1247.5      119.6  10.427 1.38e-13 ***
## X4           -7270.6     2292.4  -3.172  0.00273 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 353.5 on 45 degrees of freedom
## Multiple R-squared:  0.1827, Adjusted R-squared:  0.1645
```

```
## F-statistic: 10.06 on 1 and 45 DF,  p-value: 0.00273
```

```
plot(Y~X4, xlab="Probability of Imprisonment(#Commitments/#Offenses)", ylab="Crime Rate",
     main="Crime Rate and Probability of Imprisonment")
abline(Y_X4, col="blue")
```




```
cor(X4, Y, use="all.obs", method="pearson")
```

```
## [1] -0.4274222
```

```
#Multiple Regression-Model 1
model1<-data.frame(Y,X1,X2t,X3,X4)
model1reg <- lm(Y~X1+X2t+X3+X4)
summary(model1reg)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2t + X3 + X4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -478.71 -115.54  -18.42  136.91  442.01
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -794.67     518.74  -1.532 0.133039
## X1            104.91      48.26   2.174 0.035398 *
## X2t           157.09      39.26   4.001 0.000251 ***
## X3             55.95      16.44   3.403 0.001474 **
## X4          -3348.16    1871.67  -1.789 0.080851 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 246.6 on 42 degrees of freedom
## Multiple R-squared:  0.6289, Adjusted R-squared:  0.5936
## F-statistic:  17.8 on 4 and 42 DF,  p-value: 1.292e-08
```

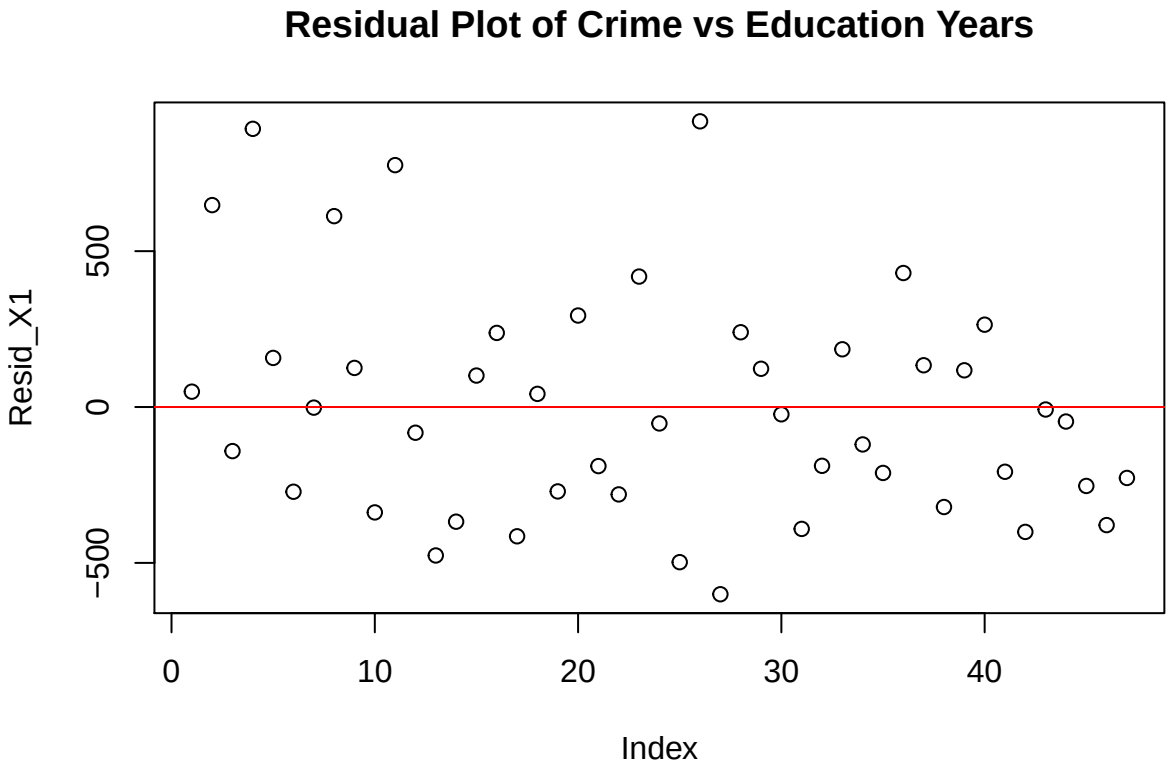
```
#R-squared
summary(model1reg)$r.squared
```

```
## [1] 0.6289268
```

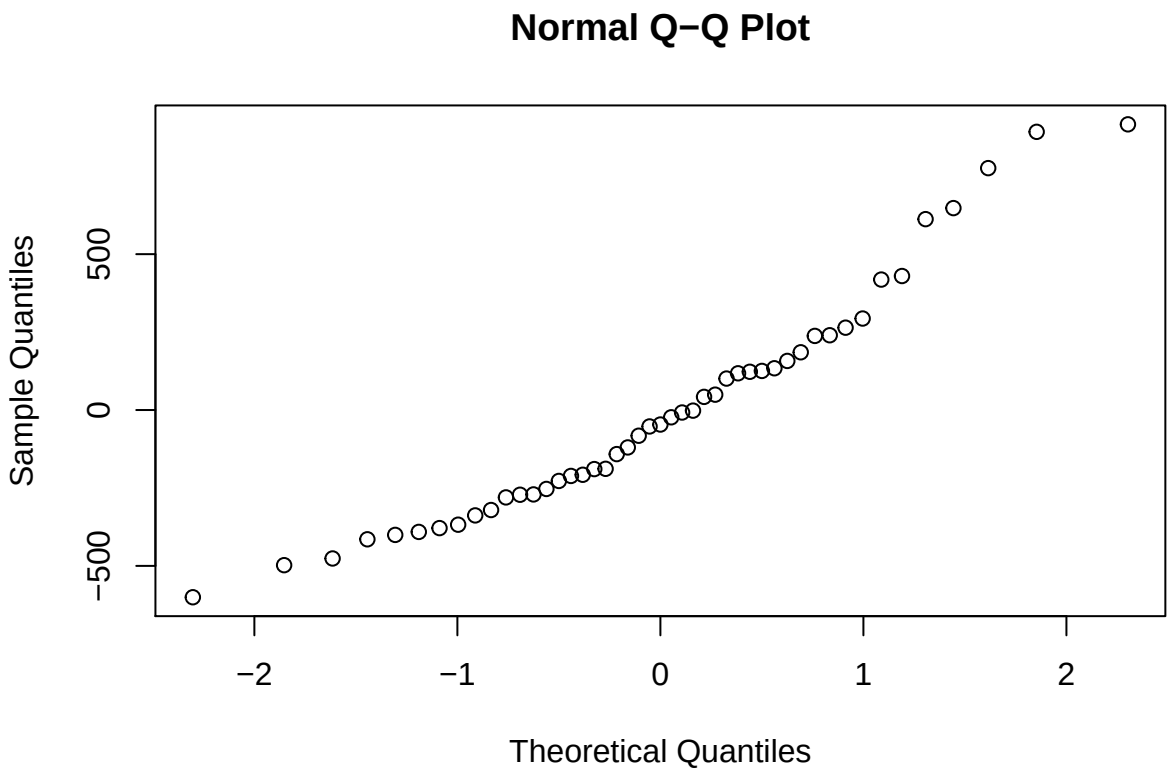
```
#Correlation Matrix
cor(model1)
```

```
##              Y         X1         X2t          X3         X4
## Y    1.0000000  0.3228349  0.30747955  0.68760446 -0.4274222
## X1   0.3228349  1.0000000 -0.53885060  0.48295213 -0.3899229
## X2t  0.3074796 -0.5388506  1.00000000  0.03654822  0.1854358
## X3   0.6876045  0.4829521  0.03654822  1.00000000 -0.4732470
## X4  -0.4274222 -0.3899229  0.18543580 -0.47324704  1.0000000
```

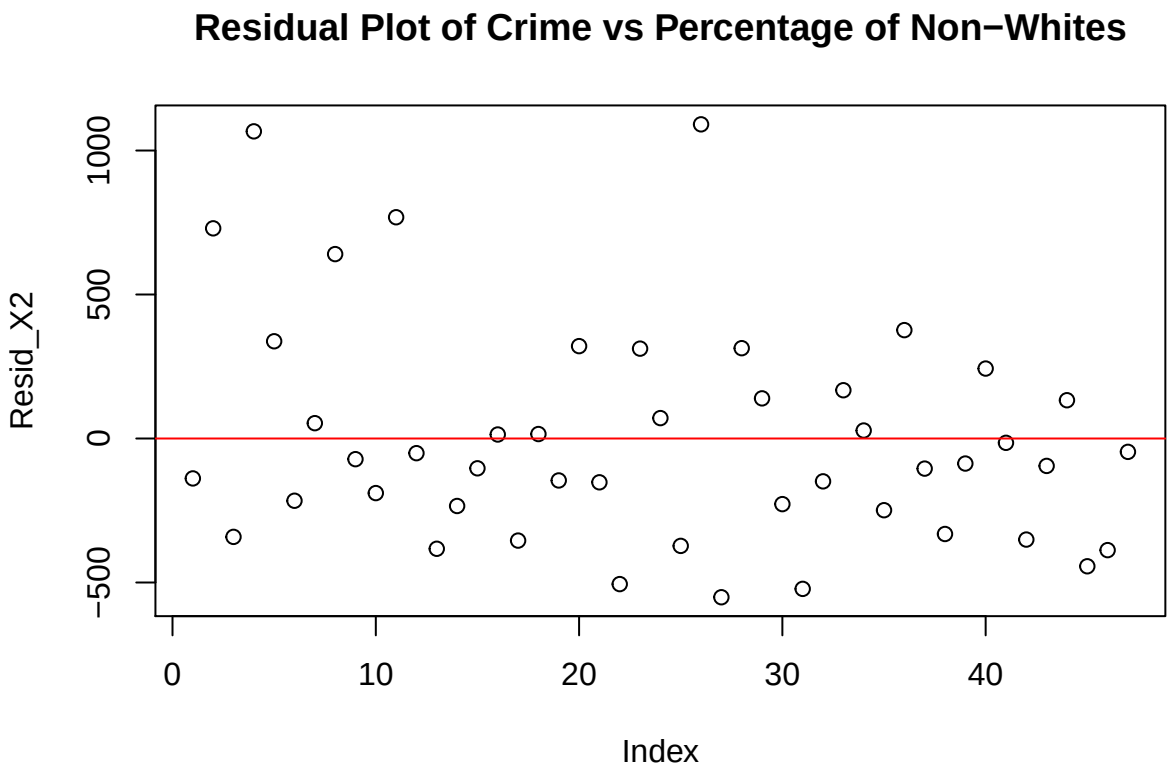
```
##############################################################################################
##Residual, Residuals vs Fitted Values and QQ Plots With Original Data Frame

#X1 Residual & QQ Plots to Check
Resid_X1 <- Y_X1$residuals
plot(Resid_X1, main="Residual Plot of Crime vs Education Years")
abline(h=0, col="red")
```



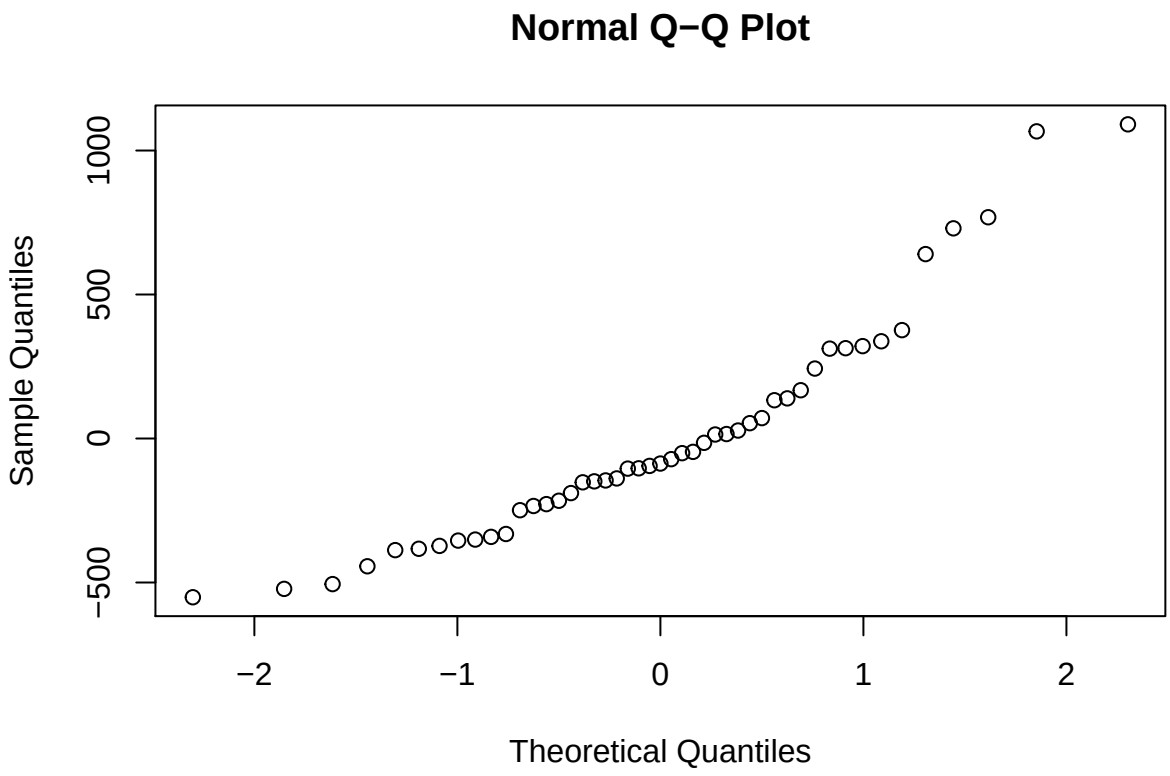

```
qqnorm(Resid_X1)
```



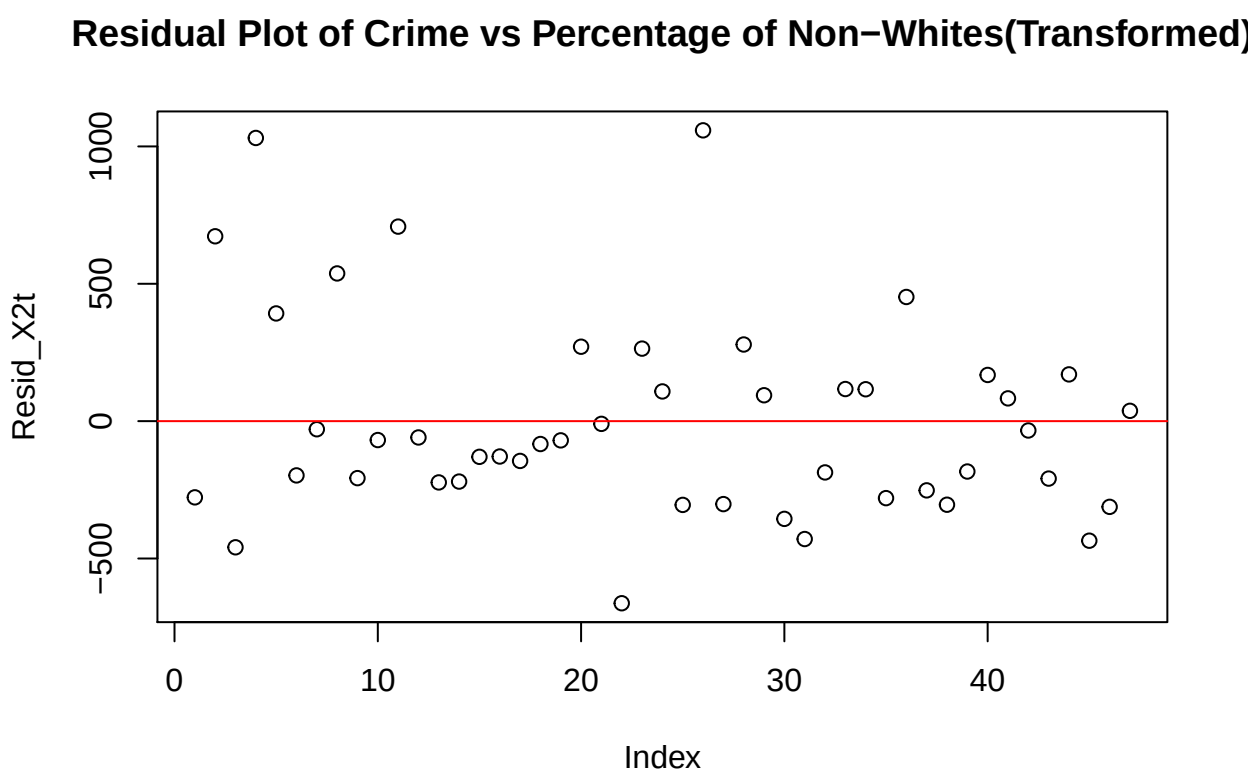

```
#X2 Residual & QQ Plots to Check
Resid_X2 <- Y_X2$residuals
plot(Resid_X2, main="Residual Plot of Crime vs Percentage of Non-Whites")
abline(h=0, col="red")
```



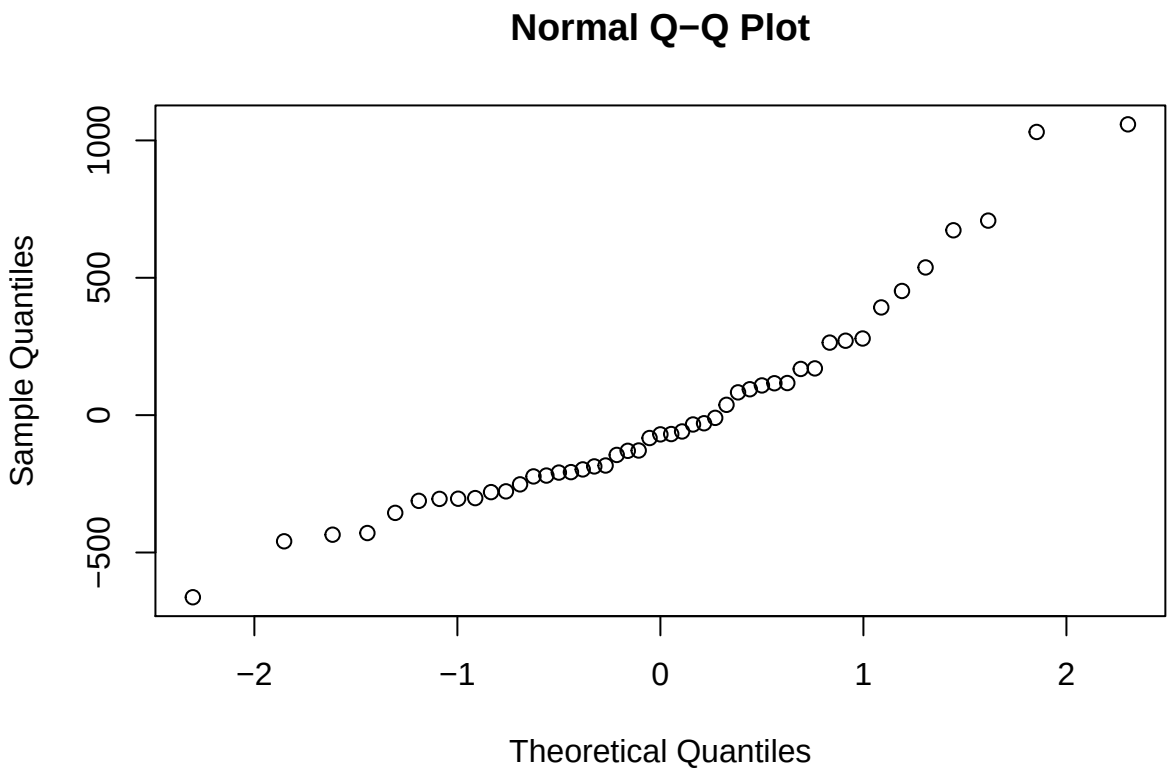

```
qqnorm(Resid_X2)
```




```r
#Transformed X2 (X2t) Residual Plot to Check
Resid_X2t <- Y_X2t$residuals
plot(Resid_X2t, main="Residual Plot of Crime vs Percentage of Non-Whites(Transformed)")
abline(h=0, col="red")
```

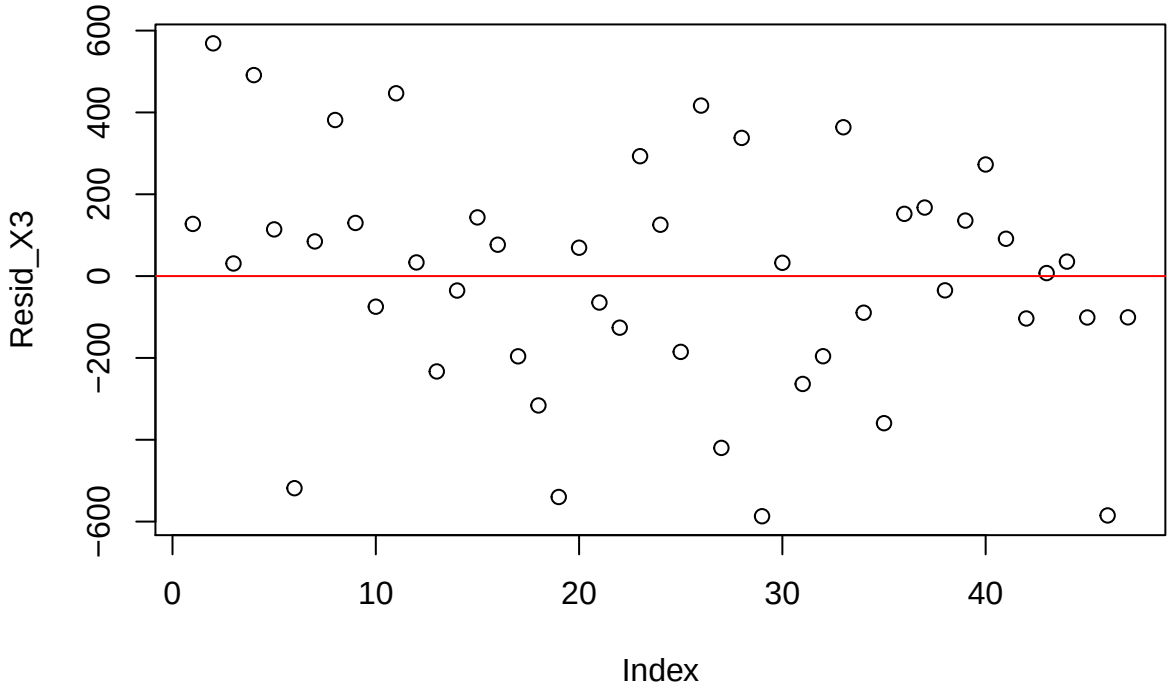



```
qqnorm(Resid_X2t)
```

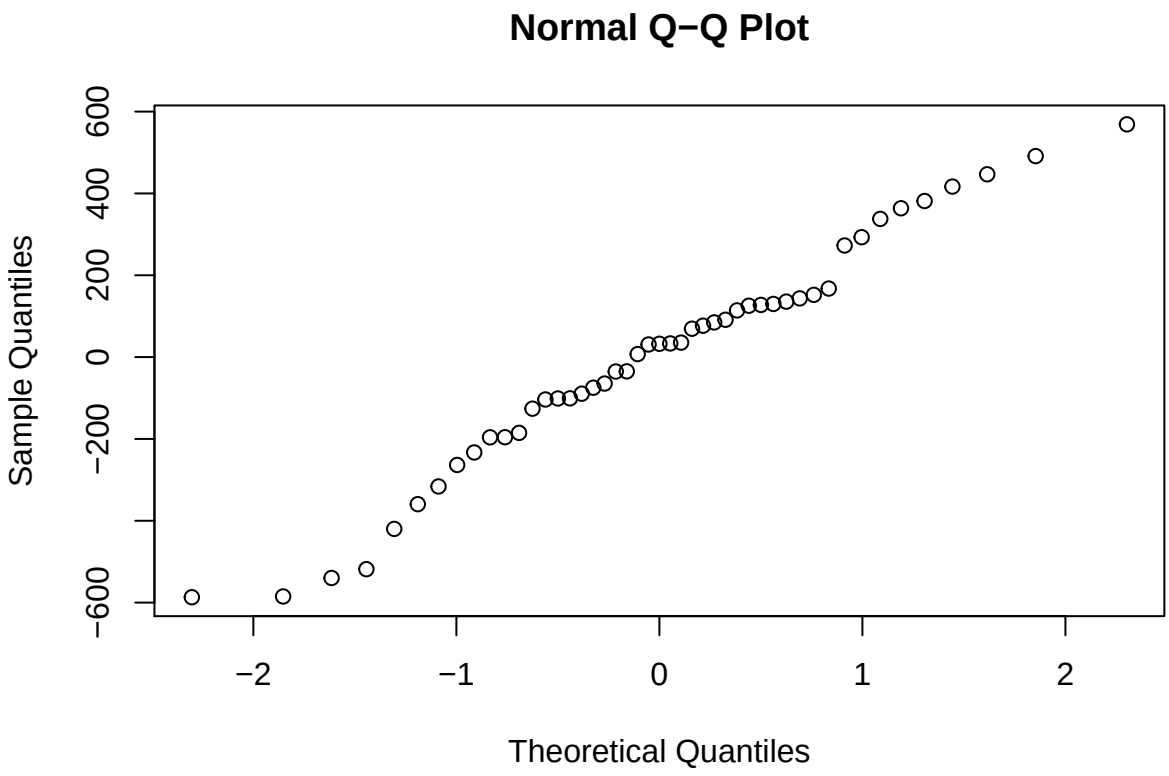



```
#X3 Residual & QQ Plots to Check
Resid_X3 <- Y_X3$residuals
plot(Resid_X3, main="Residual Plot of Crime vs Per Capita Expenditure on Police Protection in 1960")
abline(h=0, col="red")
```

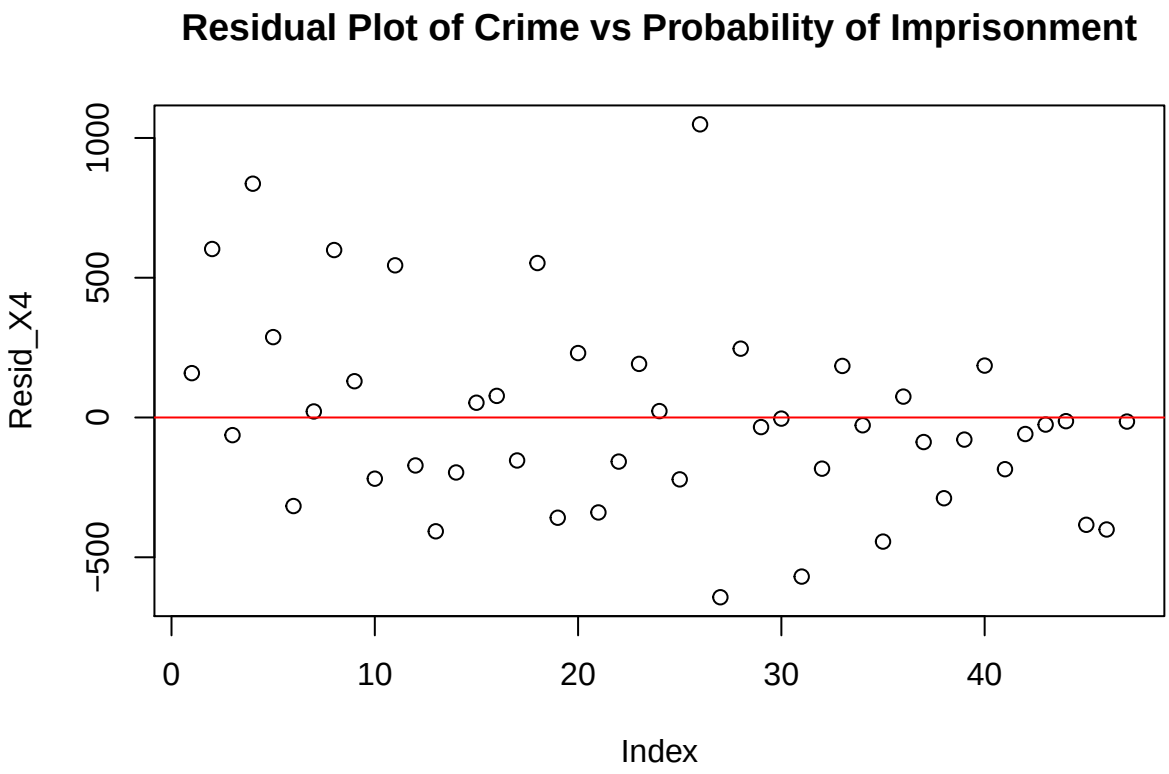
esidual Plot of Crime vs Per Capita Expenditure on Police Protection ir



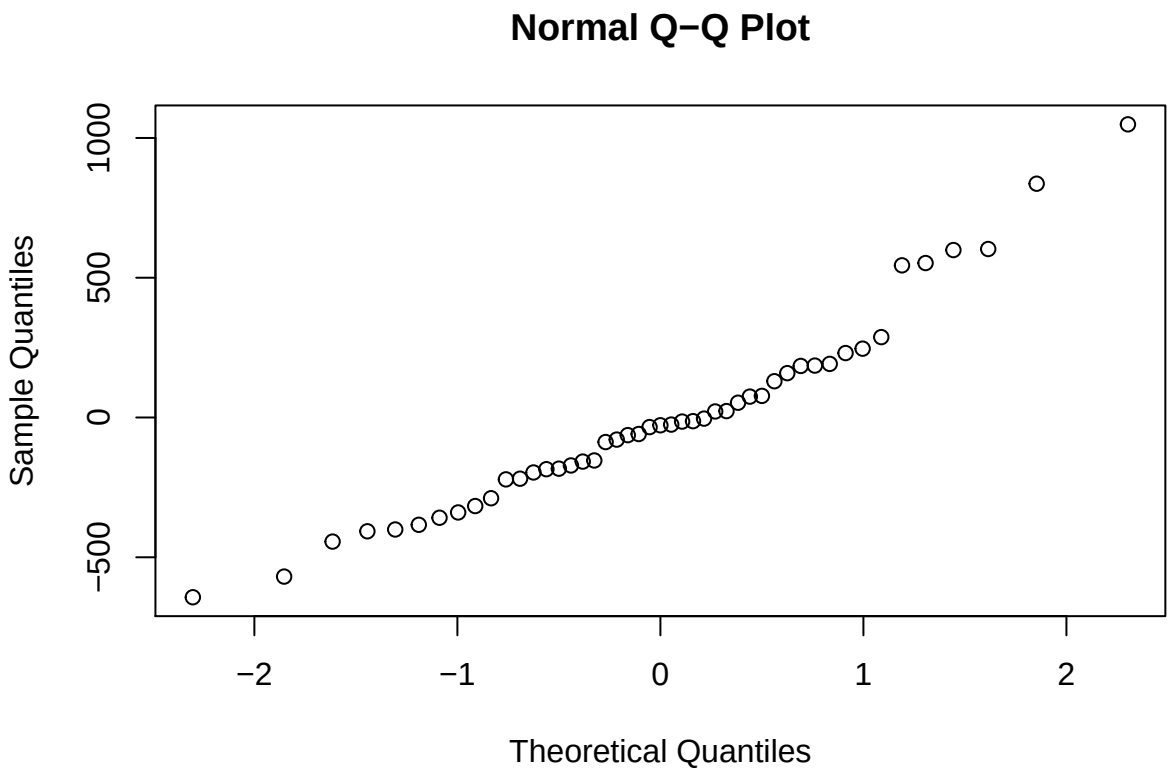

```
qqnorm(Resid_X3)
```



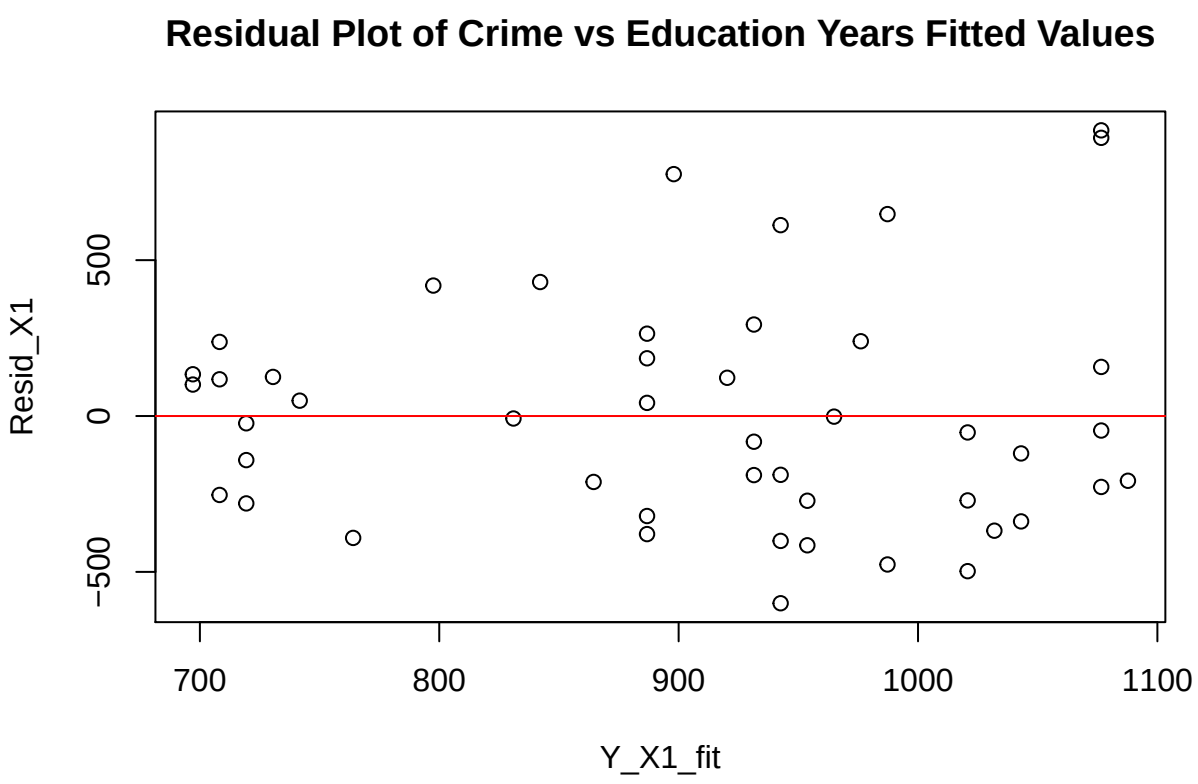

```r
#X4 Residual & QQ Plots to Check
Resid_X4 <- Y_X4$residuals
plot(Resid_X4, main="Residual Plot of Crime vs Probability of Imprisonment")
abline(h=0, col="red")
```



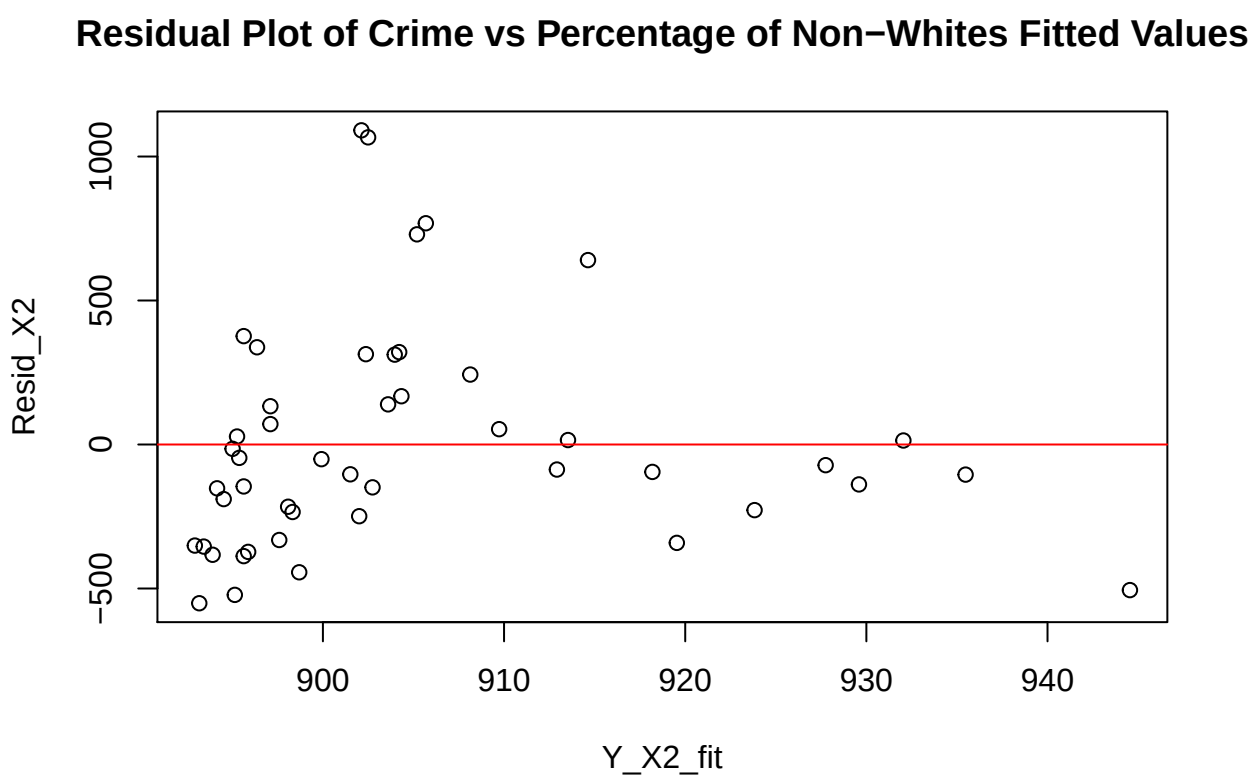

```
qqnorm(Resid_X4)
```



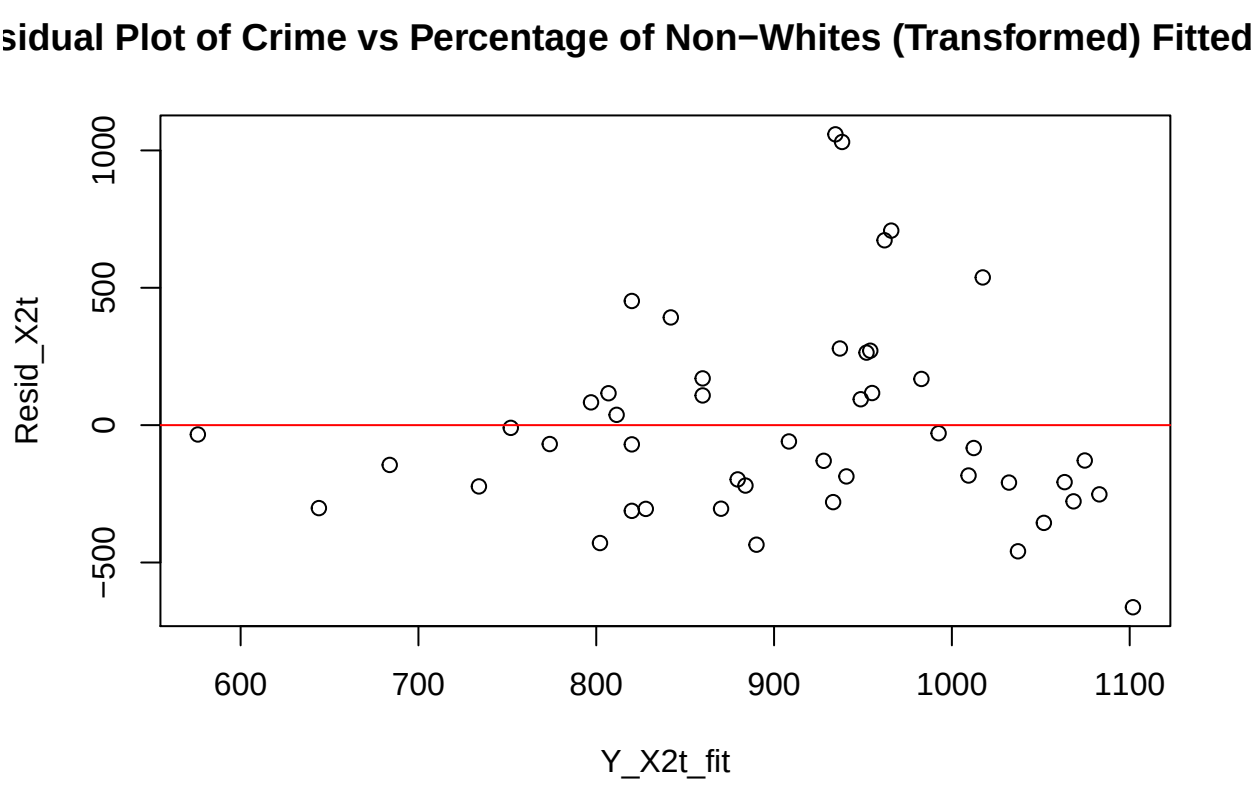

```
#X1 Residuals against Fitted values Plot
Y_X1_fit <- Y_X1$fitted.values
plot(Y_X1_fit, Resid_X1,
     main="Residual Plot of Crime vs Education Years Fitted Values")
abline(h=0, col="red")
```




```
#X2 Residuals against Fitted values Plot
Y_X2_fit <- Y_X2$fitted.values
plot(Y_X2_fit, Resid_X2,
     main="Residual Plot of Crime vs Percentage of Non-Whites Fitted Values")
abline(h=0, col="red")
```

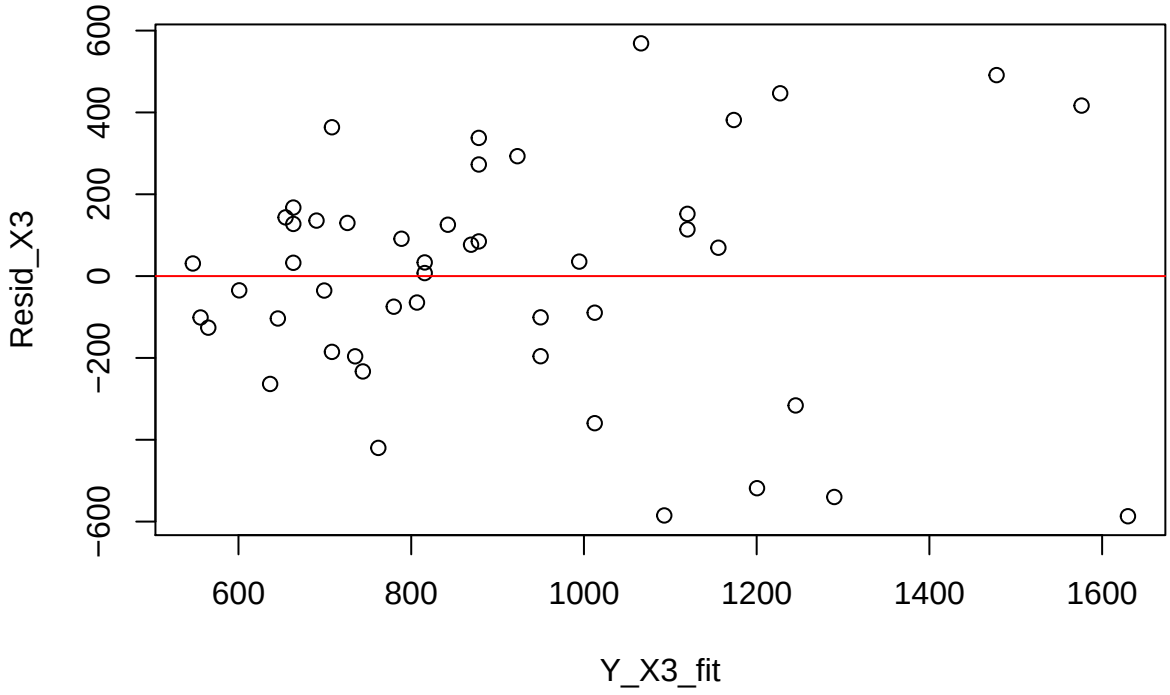



```
#X2t Residuals against Fitted values Plot
Y_X2t_fit <- Y_X2t$fitted.values
plot(Y_X2t_fit, Resid_X2t,
     main="Residual Plot of Crime vs Percentage of Non-Whites (Transformed) Fitted Values")
abline(h=0, col="red")
```

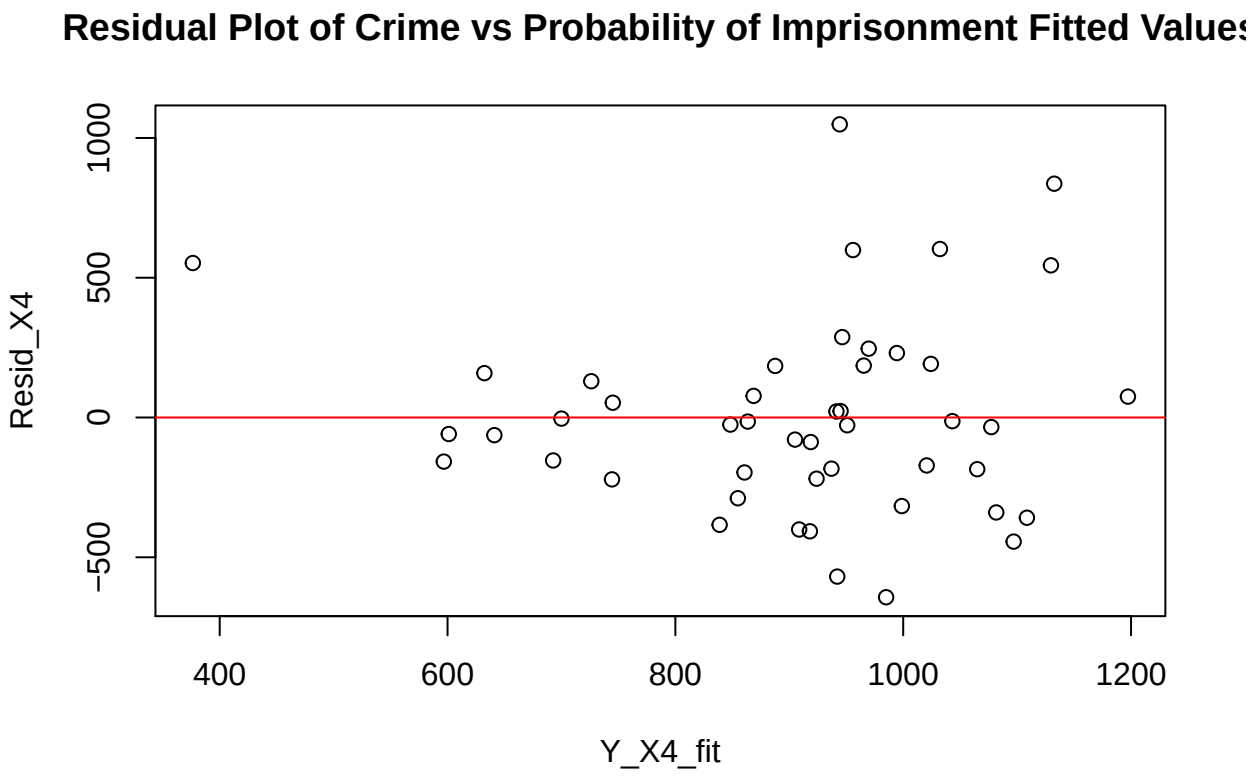



```
#X3 Residuals against Fitted values Plot
Y_X3_fit <- Y_X3$fitted.values
plot(Y_X3_fit, Resid_X3,
     main="Residual Plot of Crime vs Per Capita Expenditure on Police Protection in 1960 Fitted Values"
abline(h=0, col="red")
```

l Plot of Crime vs Per Capita Expenditure on Police Protection in 1960 l

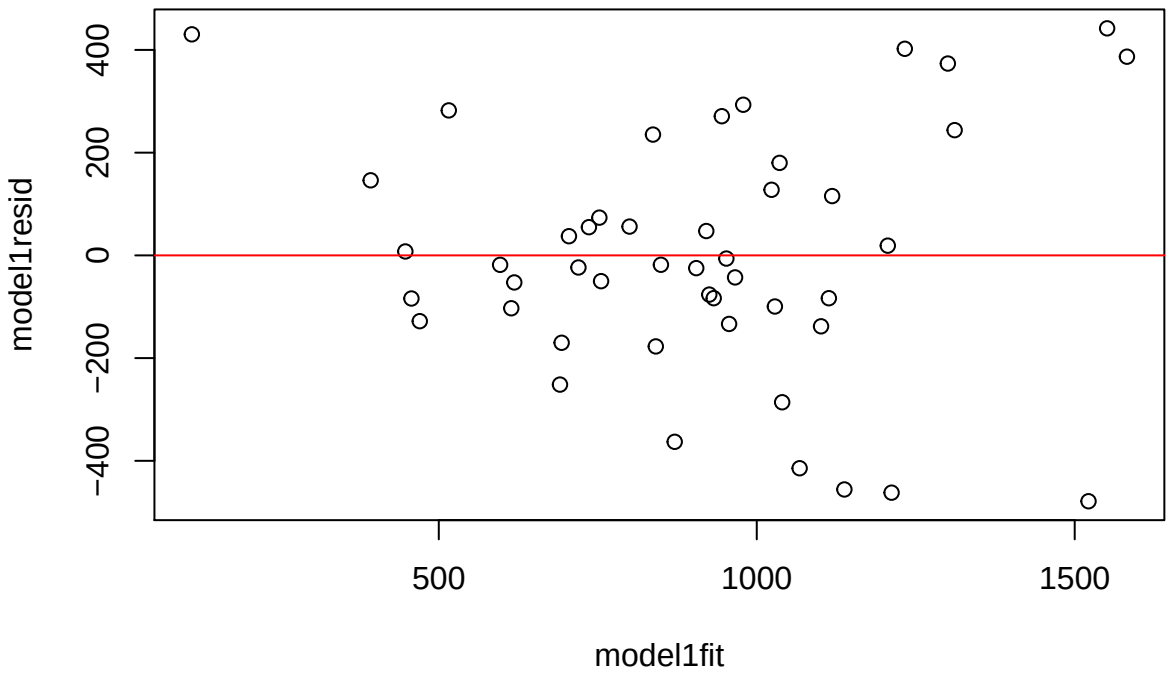



```
#X4 Residuals against Fitted values Plot
Y_X4_fit <- Y_X4$fitted.values
plot(Y_X4_fit, Resid_X4,
     main="Residual Plot of Crime vs Probability of Imprisonment Fitted Values")
abline(h=0, col="red")
```




```
#Multiple Regression Residuals against Fitted values Plot
model1resid <- model1reg$residuals
model1fit <- model1reg$fitted.values
plot(model1fit, model1resid,
     main="Figure 16-Multiple Regression(Model1) Residual vs Fitted Values Plot")
abline(h=0, col="red")
```

Figure 16−Multiple Regression(Model1) Residual vs Fitted Values Pl

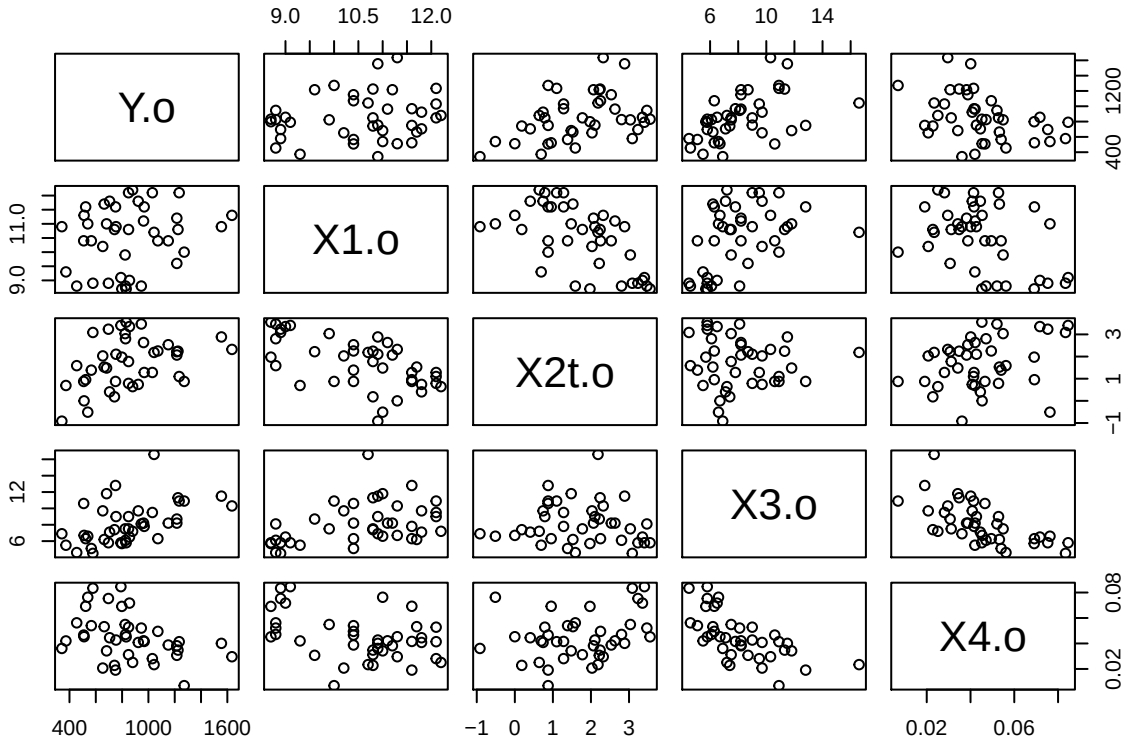



```
##########################################################################################################################
##New Data Frame with Outliers Omitted

#Outliers omitted were X4: i=18, 22, 42, Y: i=4, 11, 26
outlier_crime <- rbind(crimerate[1:3, ], crimerate[5:10, ],
                       crimerate[12:17, ], crimerate[19:21, ],
                       crimerate[23:25, ], crimerate[27:41, ], crimerate[43:47, ])
summary(outlier_crime)
```

```
##        M               So              Ed            Po1
##  Min.   :11.90   Min.   :0.0000   Min.   : 8.70   Min.   : 4.500
##  1st Qu.:13.00   1st Qu.:0.0000   1st Qu.: 9.60   1st Qu.: 6.300
##  Median :13.60   Median :0.0000   Median :10.80   Median : 7.500
##  Mean   :13.88   Mean   :0.3415   Mean   :10.53   Mean   : 8.144
##  3rd Qu.:14.70   3rd Qu.:1.0000   3rd Qu.:11.30   3rd Qu.: 9.700
##  Max.   :17.70   Max.   :1.0000   Max.   :12.20   Max.   :16.600
##       Po2              LF              M.F             Pop
##  Min.   : 4.100   Min.   :0.4800   Min.   : 93.4   Min.   :  3.00
##  1st Qu.: 6.000   1st Qu.:0.5310   1st Qu.: 96.4   1st Qu.: 10.00
##  Median : 7.100   Median :0.5600   Median : 97.7   Median : 25.00
##  Mean   : 7.702   Mean   :0.5614   Mean   : 98.2   Mean   : 34.22
##  3rd Qu.: 9.500   3rd Qu.:0.5950   3rd Qu.: 99.0   3rd Qu.: 40.00
##  Max.   :15.700   Max.   :0.6410   Max.   :104.9   Max.   :168.00
##        NW              U1              U2            Wealth
##  Min.   : 0.4   Min.   :0.07000   Min.   :2.000   Min.   :3180
##  1st Qu.: 2.4   1st Qu.:0.08000   1st Qu.:2.700   1st Qu.:4570
```

```
##  Median : 5.9   Median :0.09200   Median :3.300   Median :5290
##  Mean   : 9.5   Mean   :0.09544   Mean   :3.359   Mean   :5190
##  3rd Qu.:12.6   3rd Qu.:0.10500   3rd Qu.:3.800   3rd Qu.:5880
##  Max.   :34.9   Max.   :0.14200   Max.   :5.800   Max.   :6890
##       Ineq            Prob             Time           Crime
##  Min.   :12.60   Min.   :0.00690   Min.   :16.10   Min.   : 342.0
##  1st Qu.:16.60   1st Qu.:0.03420   1st Qu.:21.70   1st Qu.: 664.0
##  Median :19.40   Median :0.04210   Median :25.80   Median : 826.0
##  Mean   :19.56   Mean   :0.04491   Mean   :26.56   Mean   : 853.5
##  3rd Qu.:22.80   3rd Qu.:0.05320   3rd Qu.:30.00   3rd Qu.:1030.0
##  Max.   :26.40   Max.   :0.08460   Max.   :44.00   Max.   :1635.0
```

```
Y.o <- outlier_crime$Crime
X1.o <- outlier_crime$Ed
X2.o <- outlier_crime$NW
X3.o <- outlier_crime$Po1
X4.o <- outlier_crime$Prob
X2t.o <- log(X2.o)

#Scatterplot Matrix with transformed X2
pairs(~Y.o+X1.o+X2t.o+X3.o+X4.o, data=outlier_crime)
```

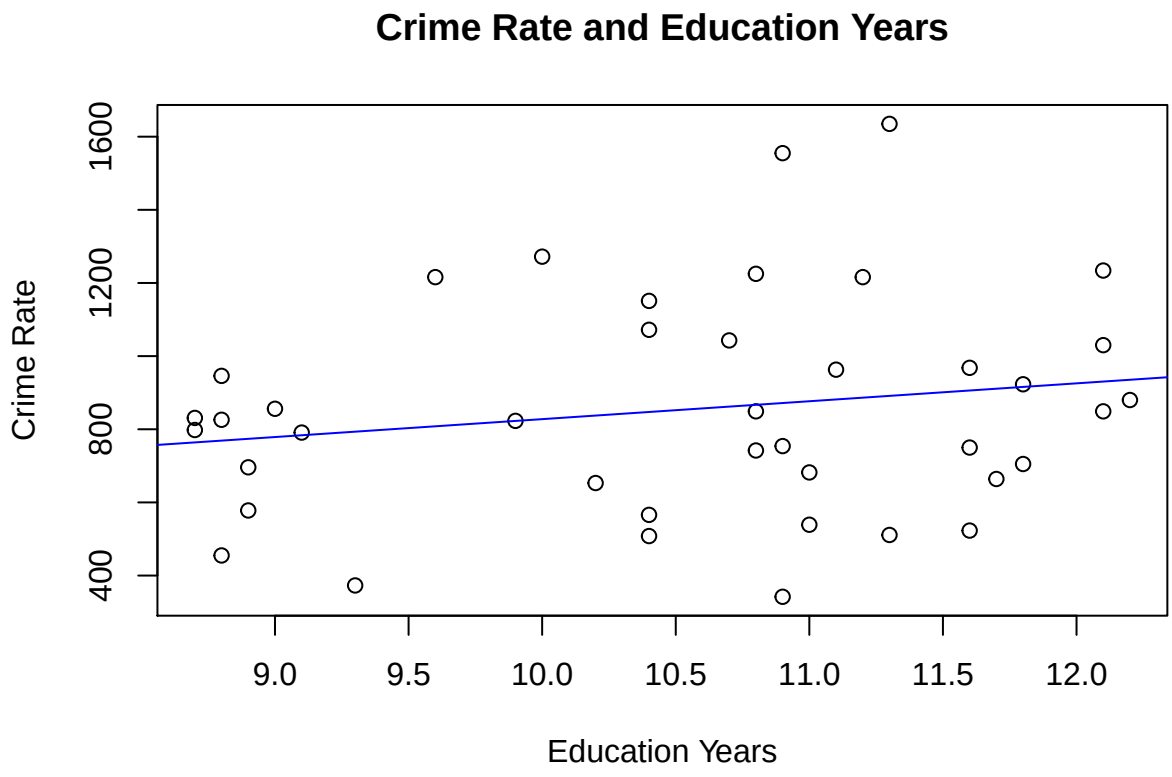



```
#############################################################################################
##Linear Regressions & Pearson Correlations With New Data Frame
```

```r
#X1.o Linear Regression
Y.o_X1.o <- lm(Y.o ~ X1.o)
summary(Y.o_X1.o)
```

```
##
## Call:
## lm(formula = Y.o ~ X1.o)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -529.75 -195.87    0.19  177.02  743.67
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   338.32     442.23   0.765    0.449
## X1.o           48.94      41.78   1.171    0.249
##
## Residual standard error: 295.1 on 39 degrees of freedom
## Multiple R-squared:  0.03398,    Adjusted R-squared:  0.009212
## F-statistic: 1.372 on 1 and 39 DF,  p-value: 0.2486
```

```r
plot(Y.o~X1.o, xlab="Education Years", ylab="Crime Rate",
     main="Crime Rate and Education Years")
abline(Y.o_X1.o, col="blue")
```



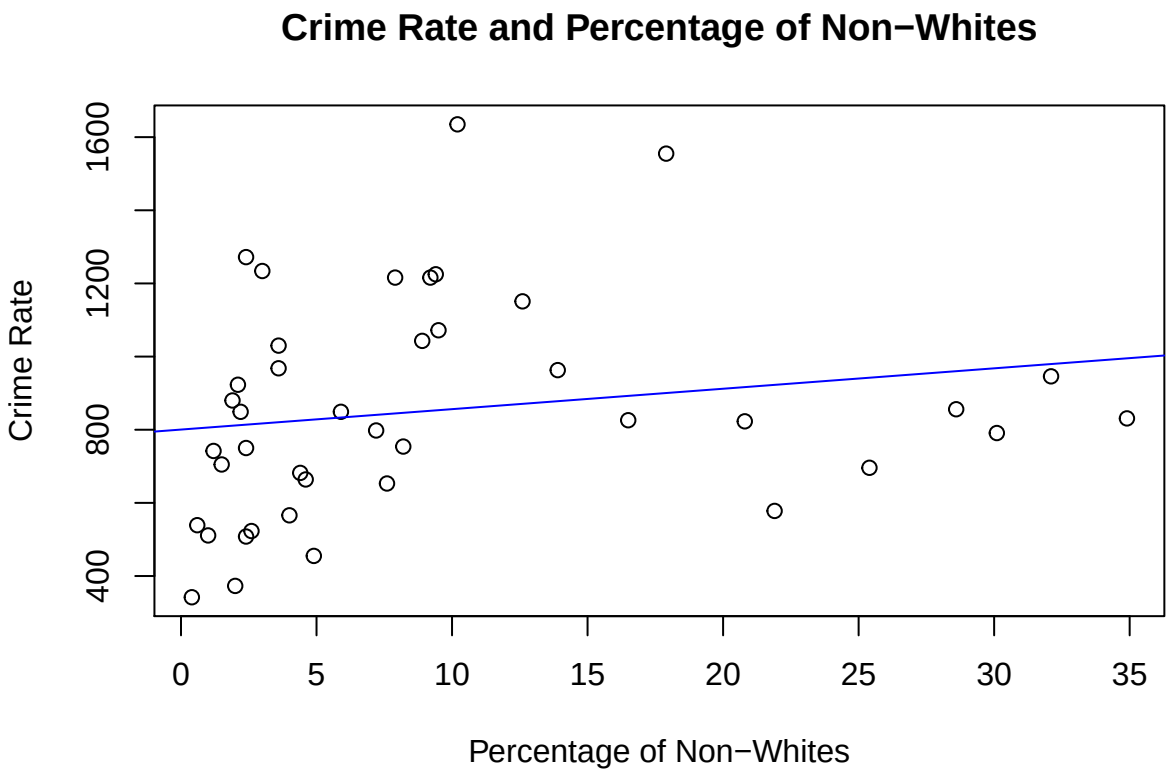
```r
cor(X1.o, Y.o, use="all.obs", method="pearson")
```

```
## [1] 0.184342
```

```r
#X2.o Linear Regression
Y.o_X2.o <- lm(Y.o ~ X2.o)
summary(Y.o_X2.o)
```

```
##
## Call:
## lm(formula = Y.o ~ X2.o)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -460.71 -189.89  -65.17  192.86  777.61
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  800.474     65.423  12.235  6.3e-15 ***
## X2.o           5.580      4.884   1.143     0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 295.3 on 39 degrees of freedom
## Multiple R-squared:  0.03239,    Adjusted R-squared:  0.007578
## F-statistic: 1.305 on 1 and 39 DF,  p-value: 0.2602
```

```r
plot(Y.o~X2.o, xlab="Percentage of Non-Whites", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites")
abline(Y.o_X2.o, col="blue")
```




```
cor(X2.o, Y.o, use="all.obs", method="pearson")
```

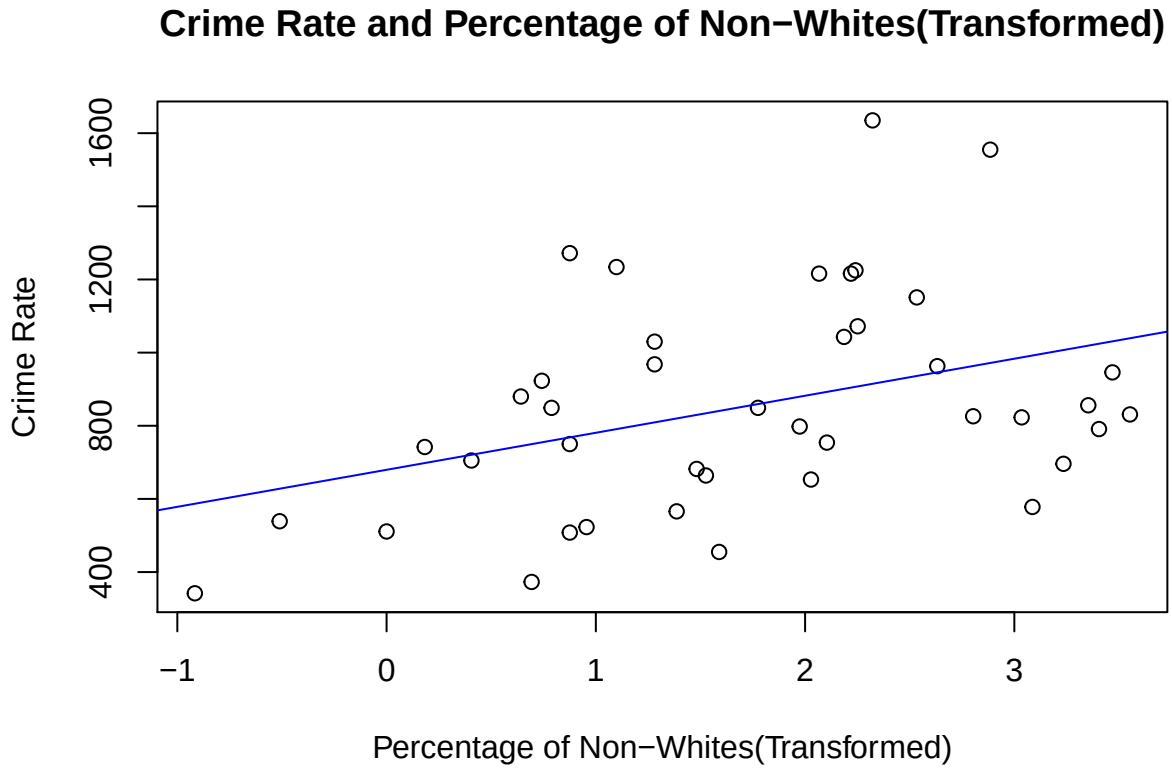
```
## [1] 0.1799685
```

```
#X2t.o Linear Regression *
Y.o_X2t.o <- lm(Y.o ~ X2t.o)
summary(Y.o_X2t.o)
```

```
##
## Call:
## lm(formula = Y.o ~ X2t.o)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -414.10 -208.28  -81.49  164.45  720.25
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   679.63      79.85   8.511    2e-10 ***
## X2t.o         101.24      39.07   2.592   0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 277.3 on 39 degrees of freedom
## Multiple R-squared:  0.1469, Adjusted R-squared:  0.125
```

```
## F-statistic: 6.716 on 1 and 39 DF,  p-value: 0.01338
```

```
plot(Y.o~X2t.o, xlab="Percentage of Non-Whites(Transformed)", ylab="Crime Rate",
     main="Crime Rate and Percentage of Non-Whites(Transformed)")
abline(Y.o_X2t.o, col="blue")
```




```
cor(X2t.o, Y.o, use="all.obs", method="pearson")
```

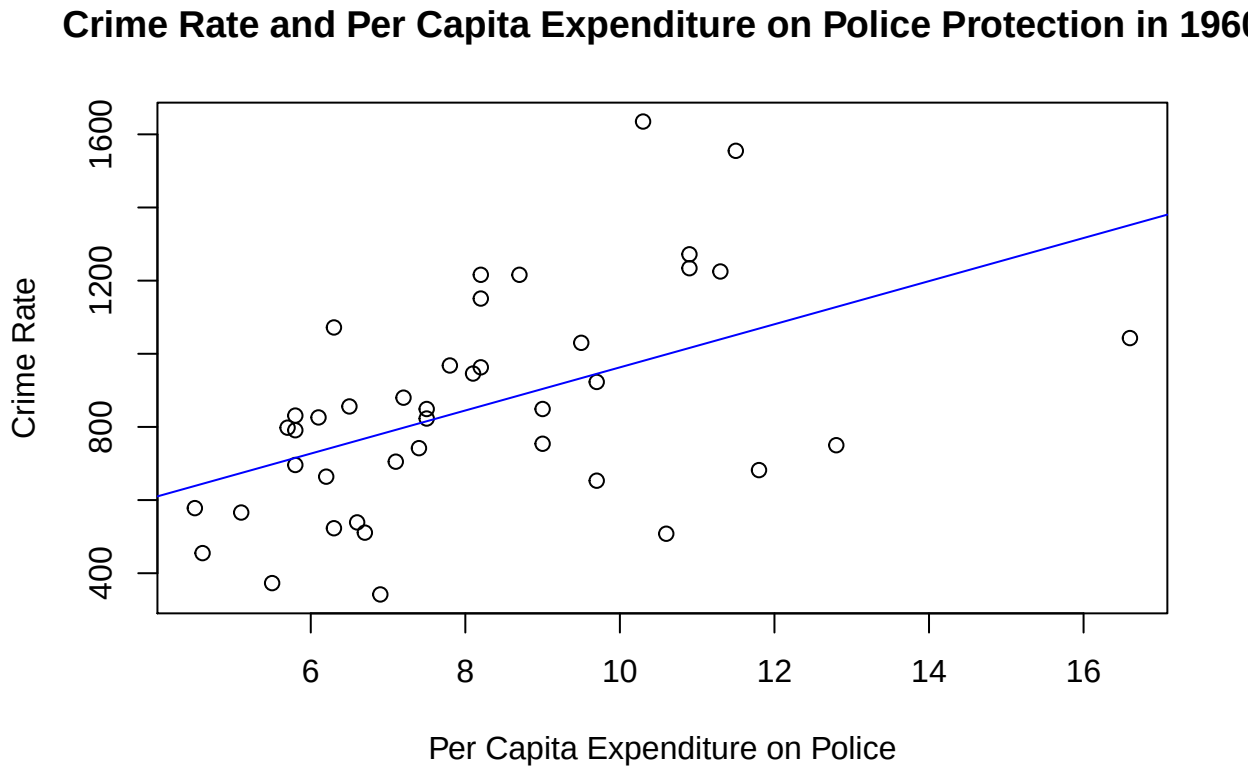
```
## [1] 0.3832827
```

```
#X3.o Linear Regression
Y.o_X3.o <- lm(Y.o ~ X3.o)
summary(Y.o_X3.o)
```

```
##
## Call:
## lm(formula = Y.o ~ X3.o)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -490.28 -189.57    7.47  115.69  654.41
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   373.40     138.96   2.687 0.010538 *
## X3.o           58.95      16.32   3.612 0.000855 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 259.9 on 39 degrees of freedom
## Multiple R-squared:  0.2507, Adjusted R-squared:  0.2315
## F-statistic: 13.05 on 1 and 39 DF,  p-value: 0.0008554
```

```
plot(Y.o~X3.o, xlab="Per Capita Expenditure on Police", ylab="Crime Rate",
     main="Crime Rate and Per Capita Expenditure on Police Protection in 1960")
abline(Y.o_X3.o, col="blue")
```




```
cor(X3.o, Y.o, use="all.obs", method="pearson")
```

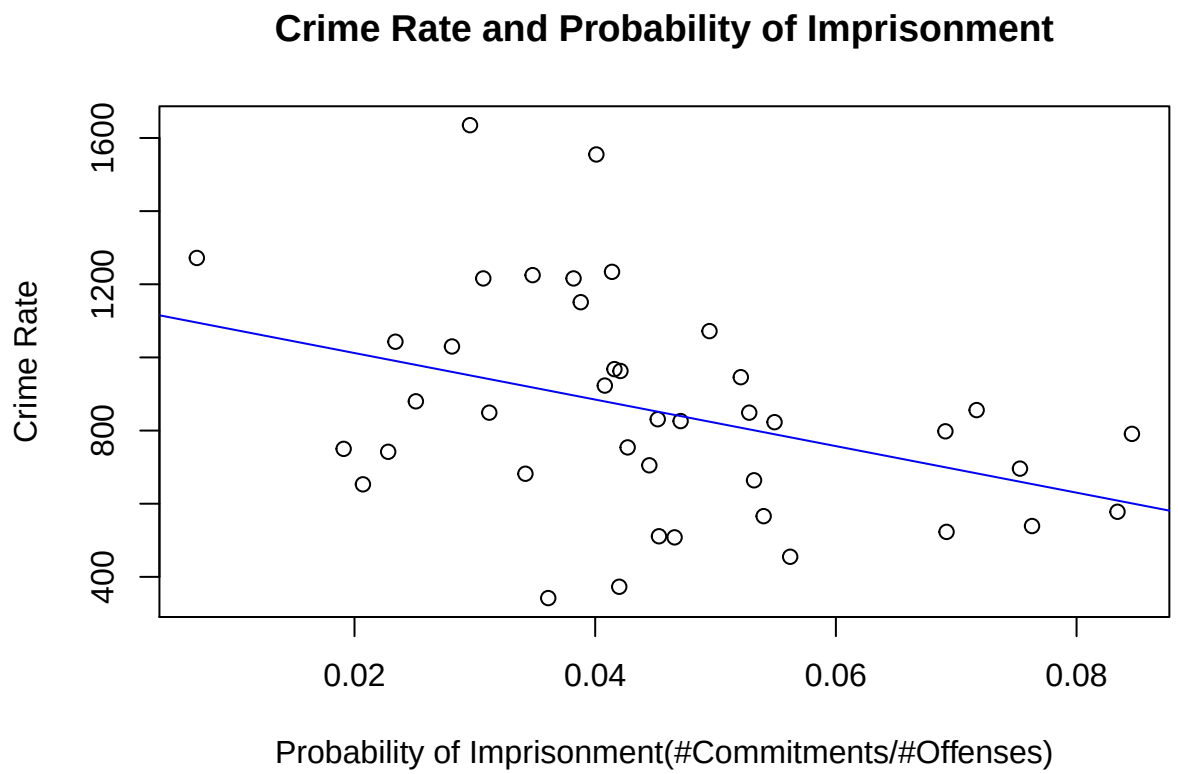
```
## [1] 0.5007164
```

```
#X4.o Linear Regression
Y.o_X4.o <- lm(Y.o ~ X4.o)
summary(Y.o_X4.o)
```

```
##
## Call:
## lm(formula = Y.o ~ X4.o)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -567.57 -175.94   33.09  172.97  684.06
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1139.3      118.3   9.630 7.36e-12 ***
## X4.o         -6363.9     2451.4  -2.596   0.0132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 277.2 on 39 degrees of freedom
## Multiple R-squared:  0.1473, Adjusted R-squared:  0.1255
## F-statistic: 6.739 on 1 and 39 DF,  p-value: 0.01323
```

```
plot(Y.o~X4.o, xlab="Probability of Imprisonment(#Commitments/#Offenses)", ylab="Crime Rate",
     main="Crime Rate and Probability of Imprisonment")
abline(Y.o_X4.o, col="blue")
```




```
cor(X4.o, Y.o, use="all.obs", method="pearson")
```

```
## [1] -0.3838468
```

```r
#Multiple Regression-Model 2
model2 <- data.frame(Y.o, X1.o,X2t.o,X3.o,X4.o)
model2reg <- lm(Y.o~X1.o+X2t.o+X3.o+X4.o)
summary(model2reg)
```

```
##
## Call:
## lm(formula = Y.o ~ X1.o + X2t.o + X3.o + X4.o)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -384.70 -130.67  -30.17  118.01  452.76
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -576.72     474.30  -1.216   0.2319
## X1.o          111.73      41.28   2.706   0.0103 *
## X2t.o         195.82      38.64   5.067 1.22e-05 ***
## X3.o           19.43      18.32   1.061   0.2959
## X4.o        -5353.79    2496.23  -2.145   0.0388 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.3 on 36 degrees of freedom
## Multiple R-squared:  0.5683, Adjusted R-squared:  0.5203
## F-statistic: 11.85 on 4 and 36 DF,  p-value: 3.044e-06
```

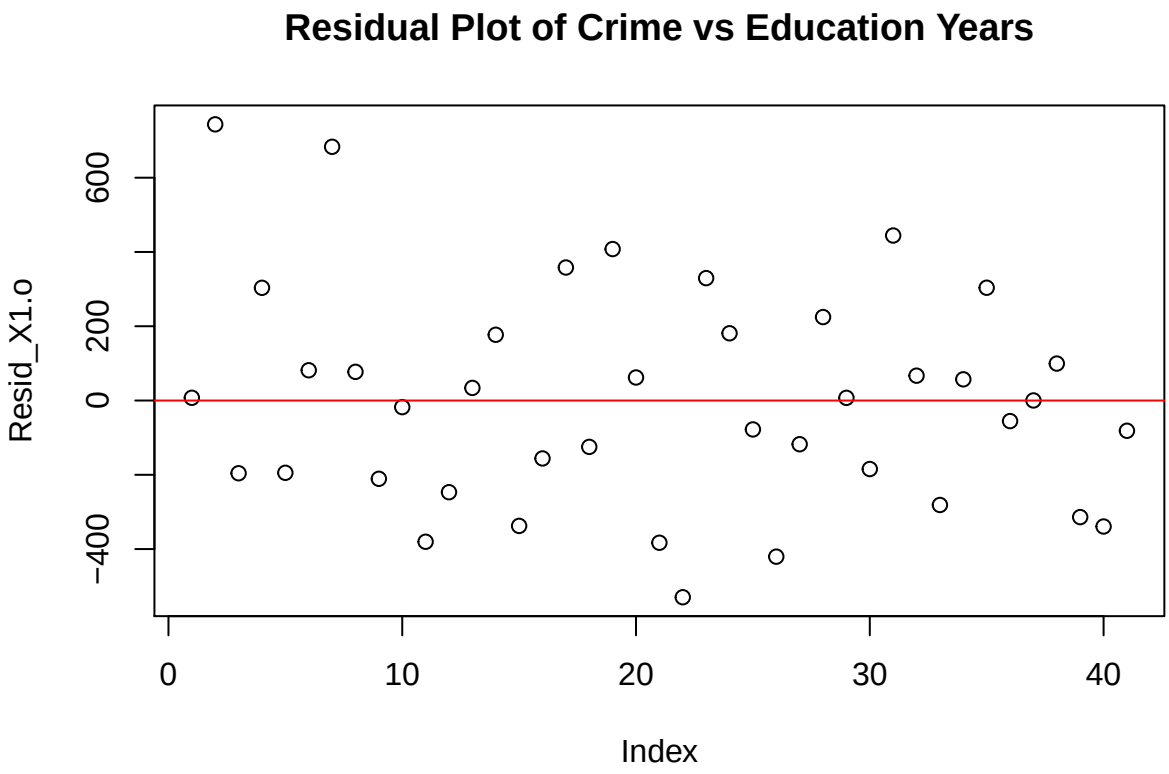
```r
#R-squared
summary(model2reg)$r.squared
```

```
## [1] 0.5683115
```

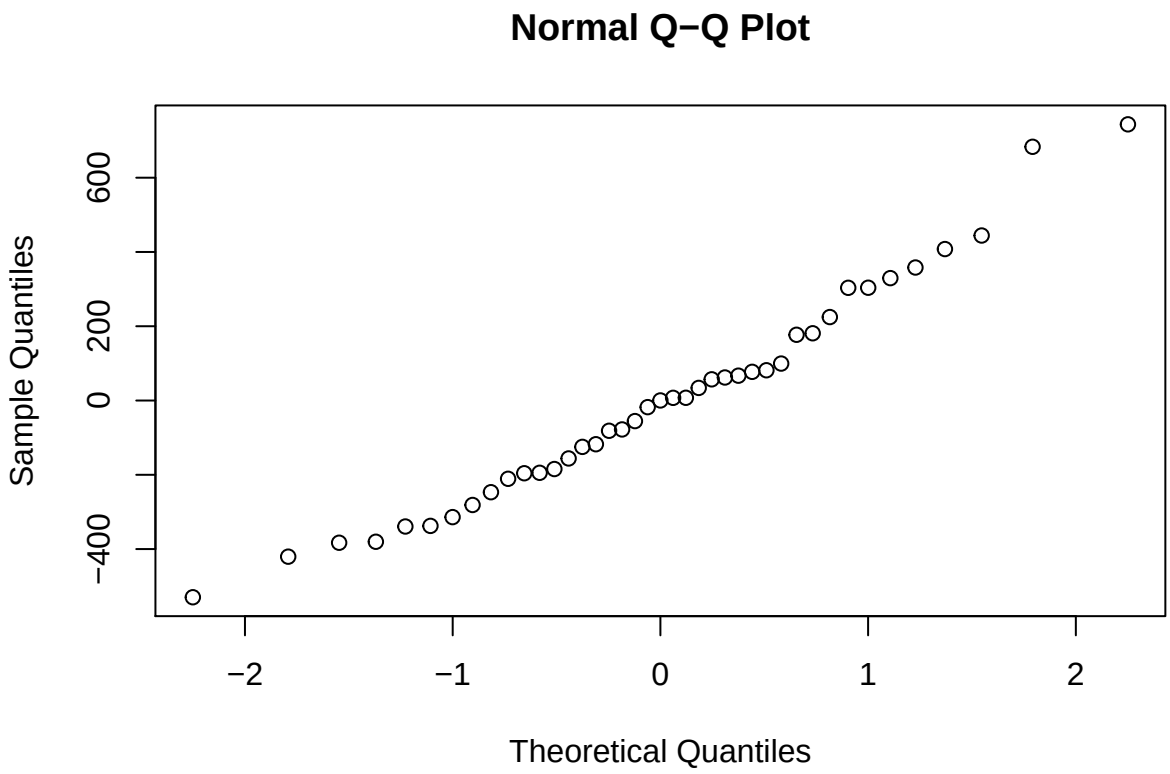
```r
#Correlation Matrix
cor(model2)
```

```
##              Y.o       X1.o       X2t.o        X3.o       X4.o
## Y.o    1.0000000  0.1843420  0.38328269  0.50071642 -0.3838468
## X1.o   0.1843420  1.0000000 -0.59664526  0.42206265 -0.4215966
## X2t.o  0.3832827 -0.5966453  1.00000000 -0.06367576  0.2986594
## X3.o   0.5007164  0.4220626 -0.06367576  1.00000000 -0.6356398
## X4.o  -0.3838468 -0.4215966  0.29865940 -0.63563981  1.0000000
```

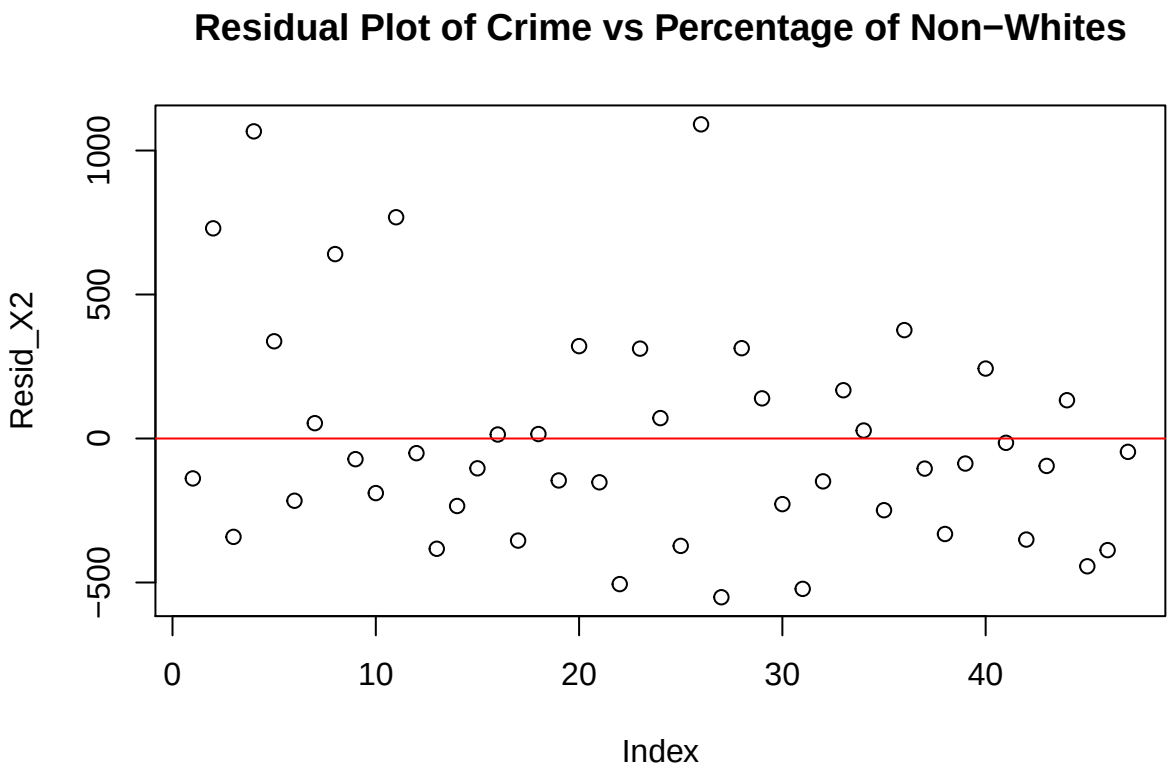
```r
##############################################################################################
##Residual, Residual vs Fitted Values and QQ Plots With New Data Frame

#X1.o Residual & QQ Plots to Check
Resid_X1.o <- Y.o_X1.o$residuals
plot(Resid_X1.o, main="Residual Plot of Crime vs Education Years")
abline(h=0, col="red")
```



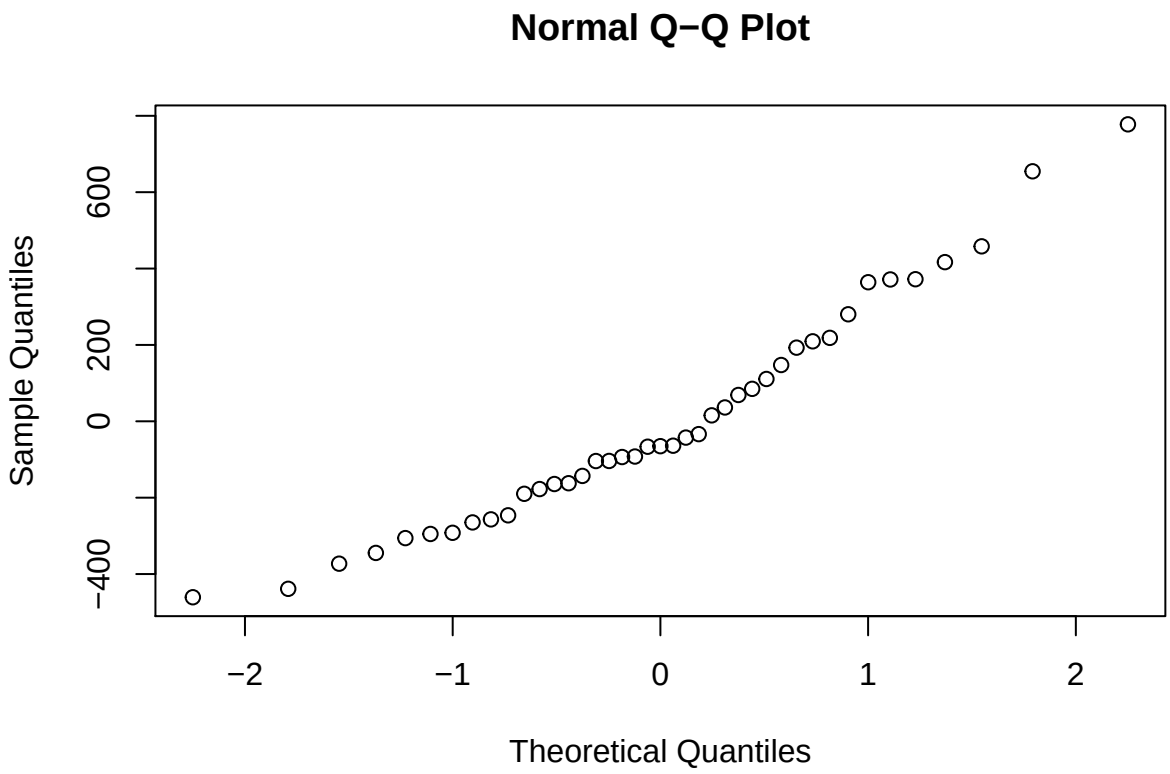

```
qqnorm(Resid_X1.o)
```



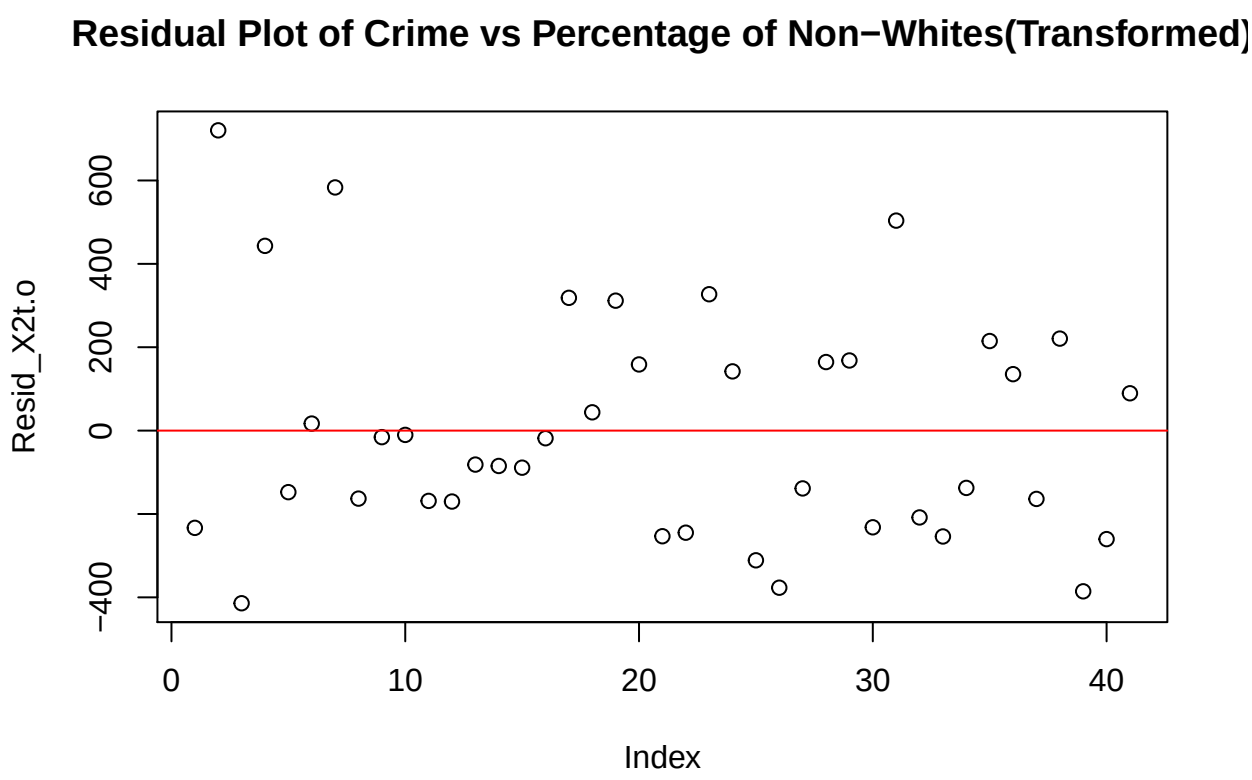

```
#X2.o Residual & QQ Plots to Check
Resid_X2.o <- Y.o_X2.o$residuals
plot(Resid_X2, main="Residual Plot of Crime vs Percentage of Non-Whites")
abline(h=0, col="red")
```



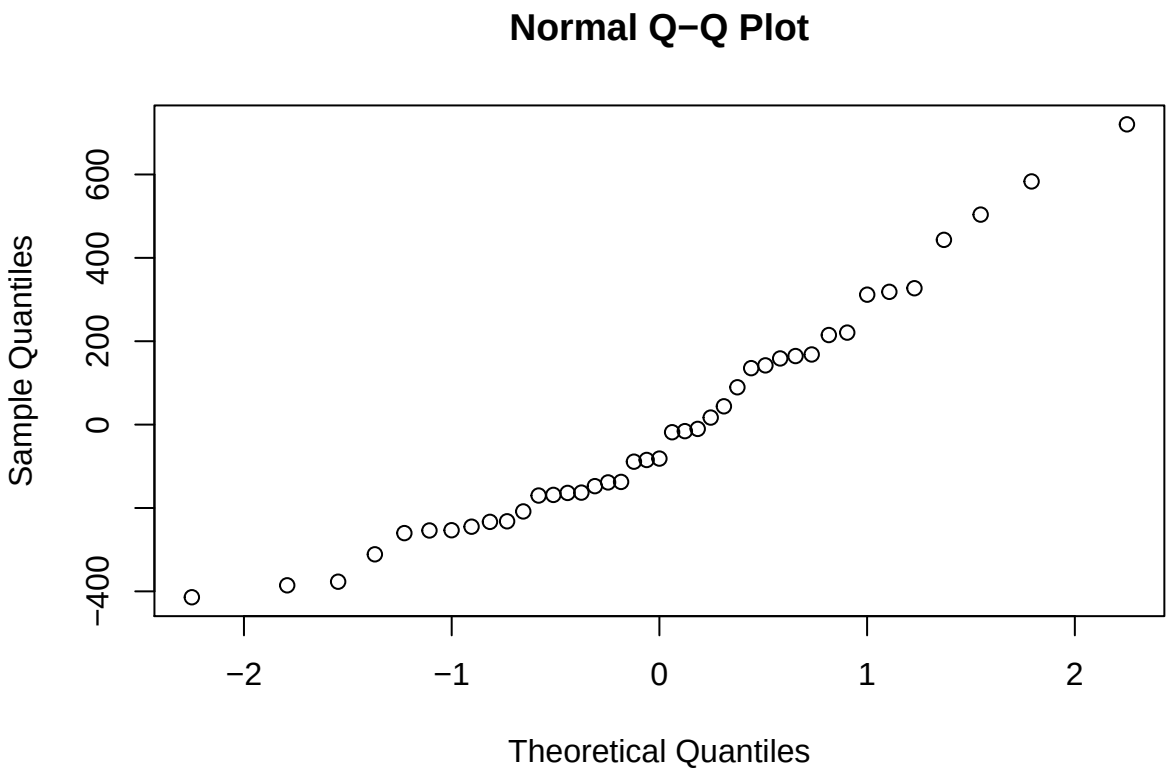

```
qqnorm(Resid_X2.o)
```




```
#X2t.o Residual Plot to Check
Resid_X2t.o <- Y.o_X2t.o$residuals
plot(Resid_X2t.o, main="Residual Plot of Crime vs Percentage of Non-Whites(Transformed)")
abline(h=0, col="red")
```

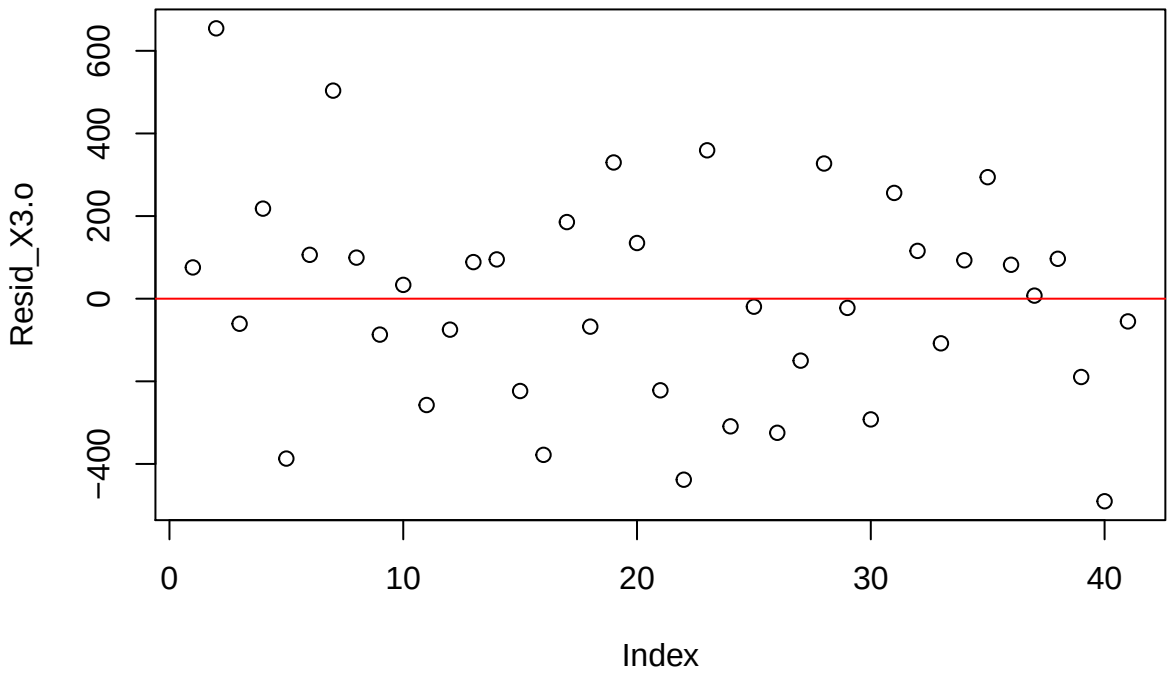



```
qqnorm(Resid_X2t.o)
```

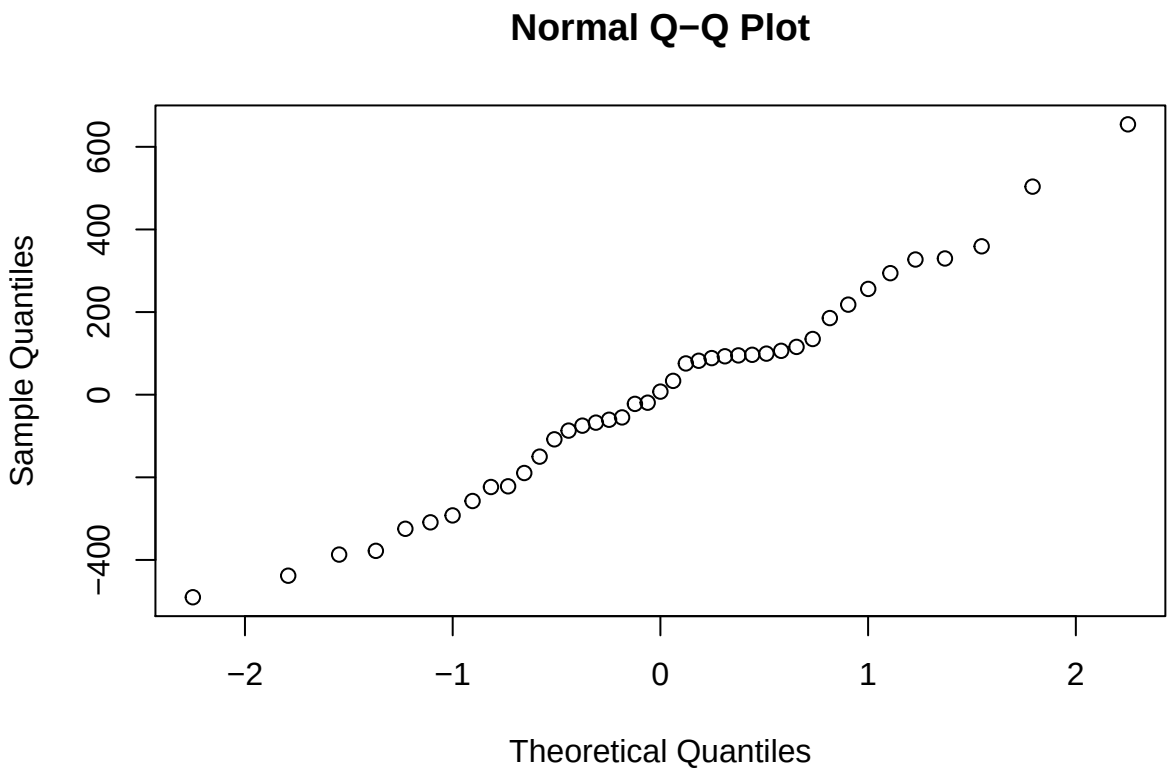



```
#X3.o Residual & QQ Plots to Check
Resid_X3.o <- Y.o_X3.o$residuals
plot(Resid_X3.o, main="Residual Plot of Crime vs Per Capita Expenditure on Police Protection in 1960")
abline(h=0, col="red")
```

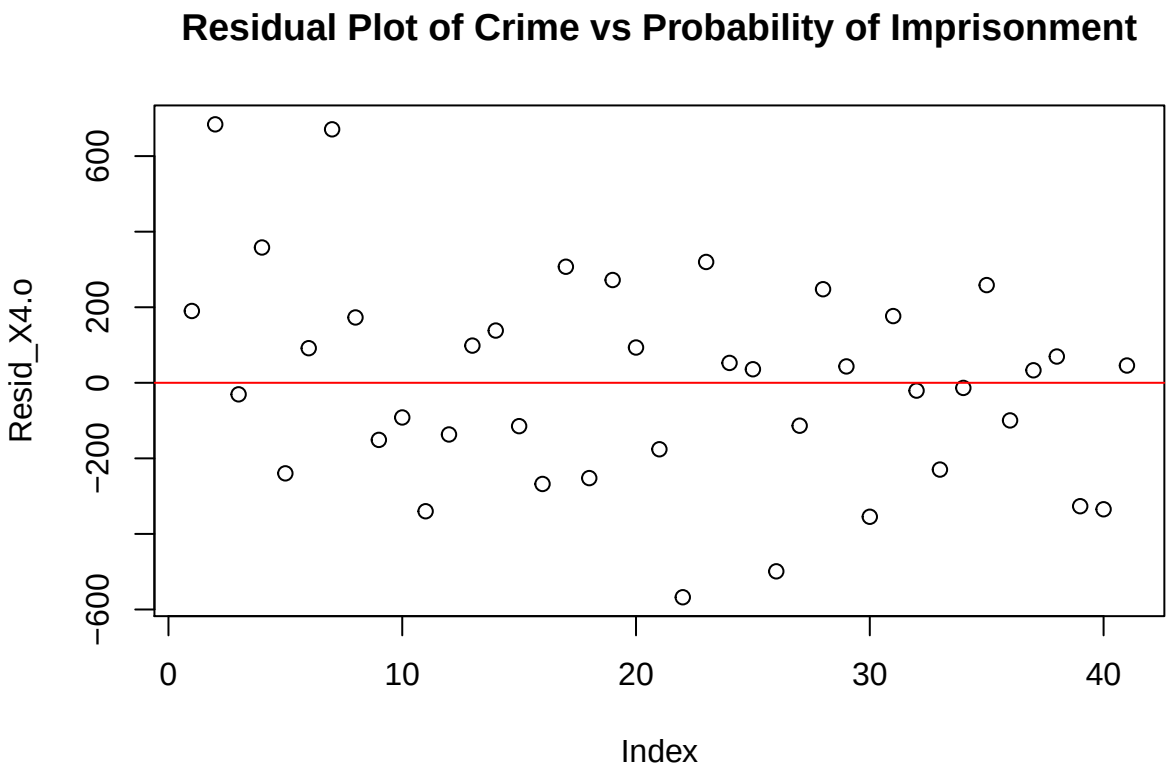
esidual Plot of Crime vs Per Capita Expenditure on Police Protection ir



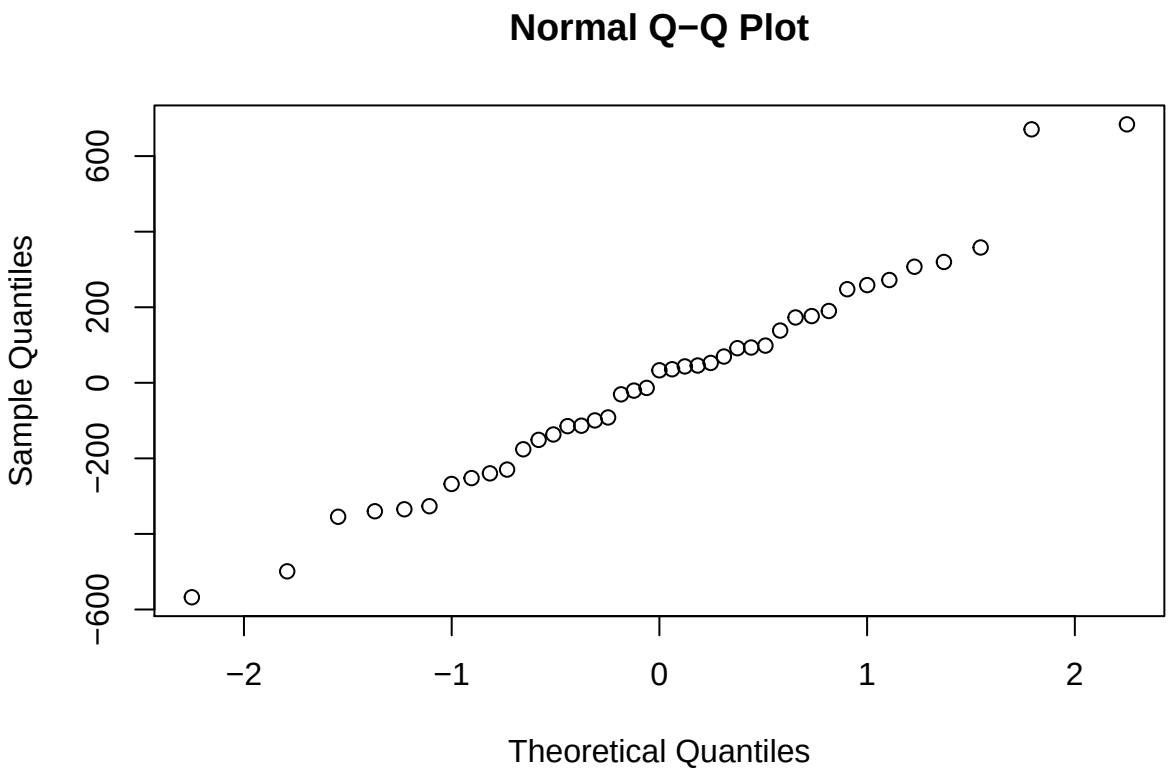

```
qqnorm(Resid_X3.o)
```



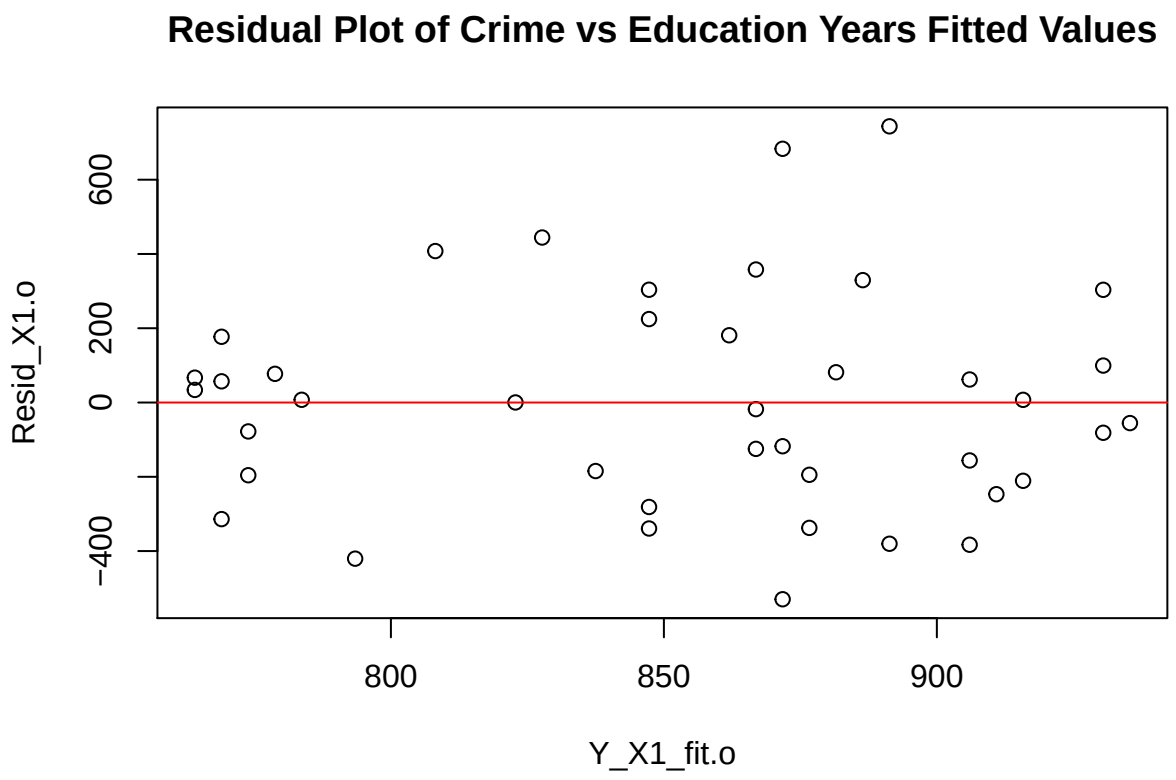

```
#X4.o Residual & QQ Plots to Check
Resid_X4.o <- Y.o_X4.o$residuals
plot(Resid_X4.o, main="Residual Plot of Crime vs Probability of Imprisonment")
abline(h=0, col="red")
```



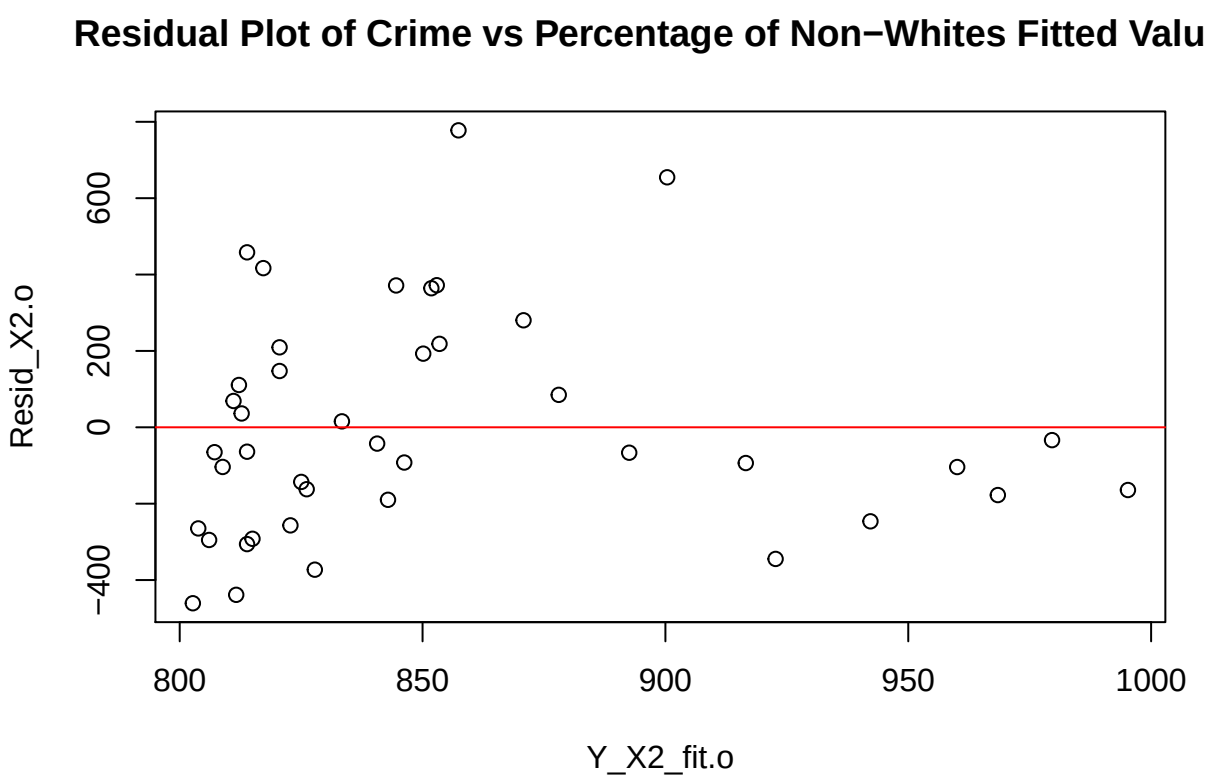

```
qqnorm(Resid_X4.o)
```



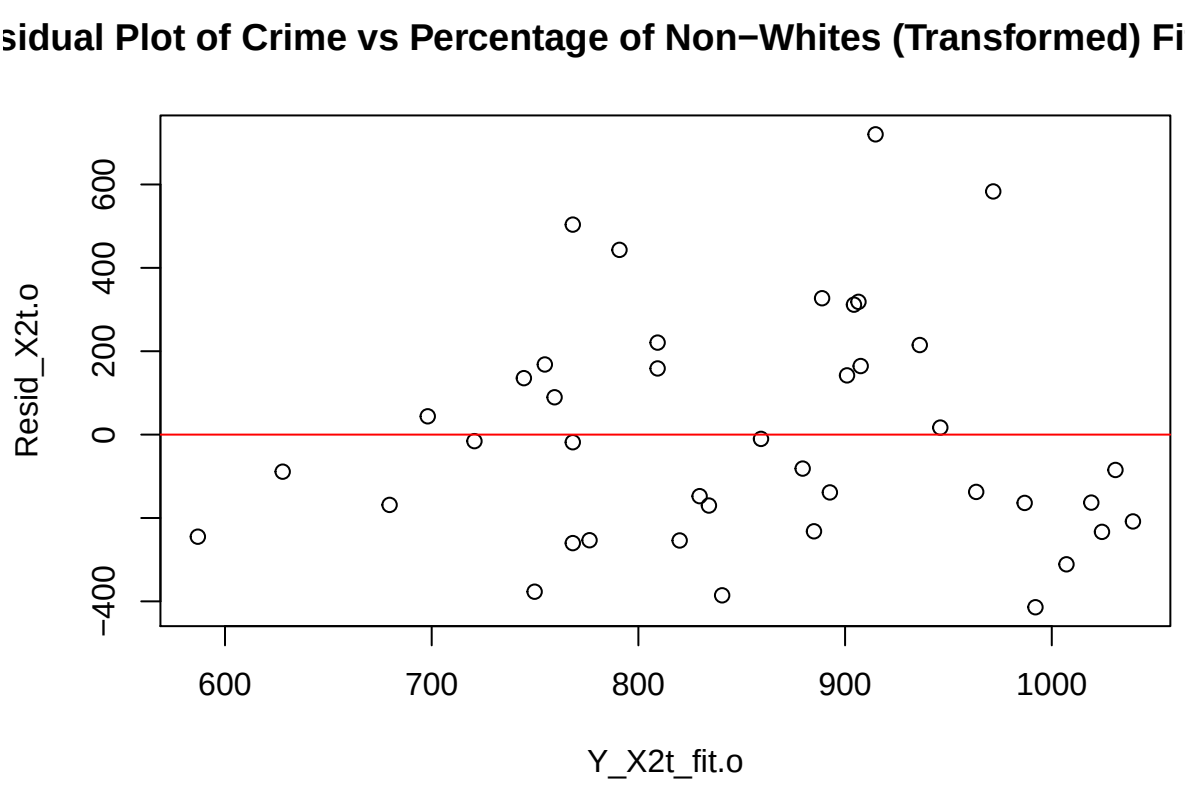

```
#X1.o Residuals against Fitted values Plot
Y_X1_fit.o <- Y.o_X1.o$fitted.values
plot(Y_X1_fit.o, Resid_X1.o,
     main="Residual Plot of Crime vs Education Years Fitted Values")
abline(h=0, col="red")
```




```
#X2.o Residuals against Fitted values Plot
Y_X2_fit.o <- Y.o_X2.o$fitted.values
plot(Y_X2_fit.o, Resid_X2.o,
     main="Residual Plot of Crime vs Percentage of Non-Whites Fitted Values")
abline(h=0, col="red")
```

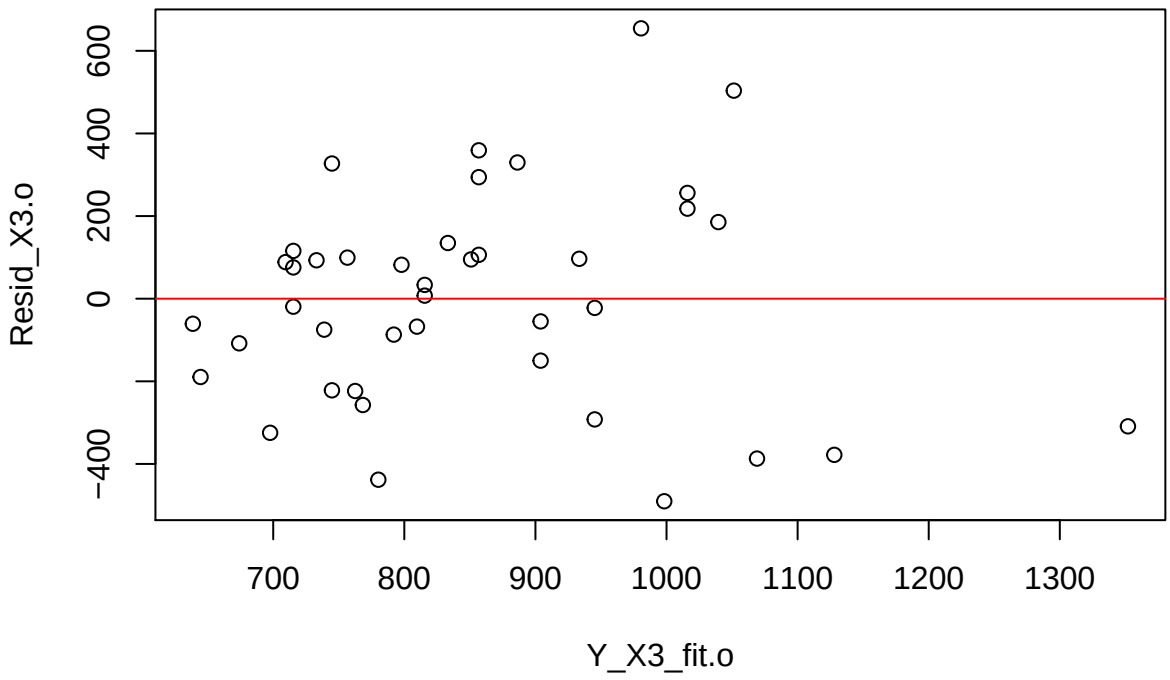



```
#X2t.o Residuals against Fitted values Plot
Y_X2t_fit.o <- Y.o_X2t.o$fitted.values
plot(Y_X2t_fit.o, Resid_X2t.o,
     main="Residual Plot of Crime vs Percentage of Non-Whites (Transformed) Fitted Values")
abline(h=0, col="red")
```

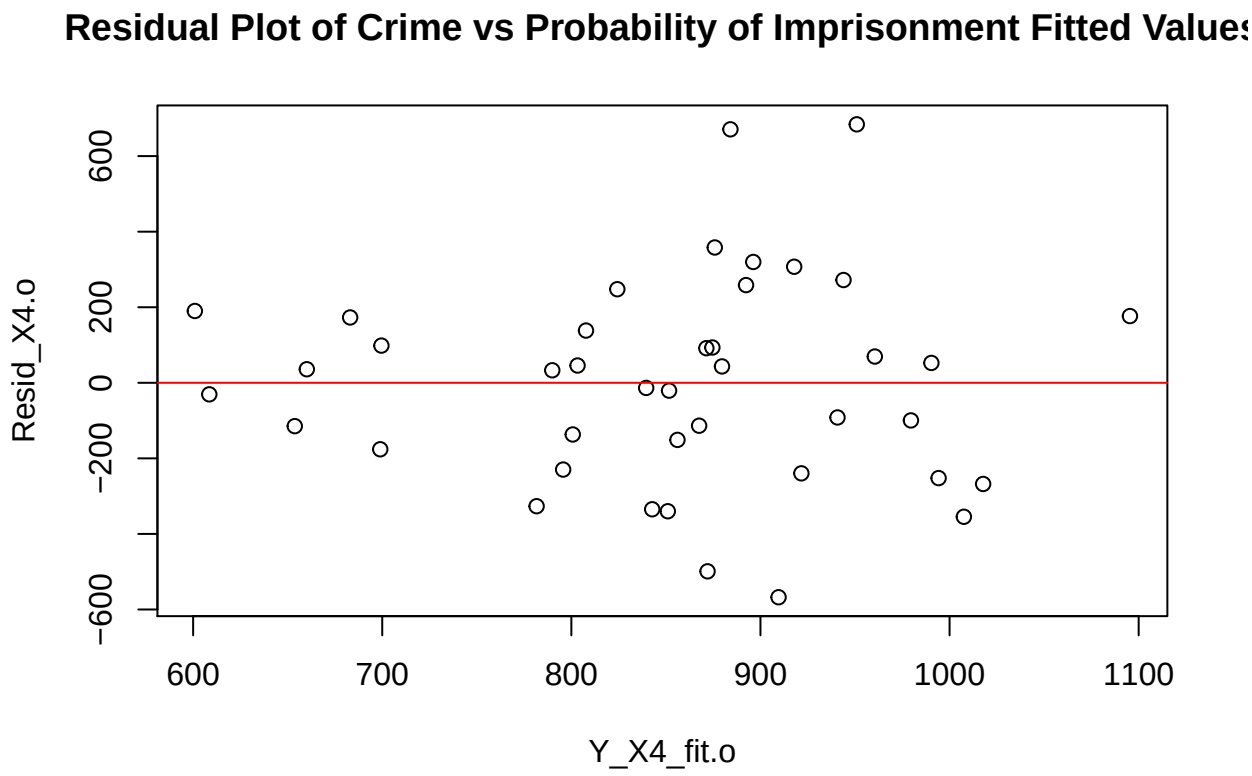



```
#X3.o Residuals against Fitted values Plot
Y_X3_fit.o <- Y.o_X3.o$fitted.values
plot(Y_X3_fit.o, Resid_X3.o,
     main="Residual Plot of Crime vs Per Capita Expenditure on Police Protection in 1960 Fitted Values"
abline(h=0, col="red")
```

l Plot of Crime vs Per Capita Expenditure on Police Protection in 1960 l




```
#X4.o Residuals against Fitted values Plot
Y_X4_fit.o <- Y.o_X4.o$fitted.values
plot(Y_X4_fit.o, Resid_X4.o,
     main="Residual Plot of Crime vs Probability of Imprisonment Fitted Values")
abline(h=0, col="red")
```




```
#Multiple Regression(Model2) Residuals against Fitted values Plot
model2resid <- model2reg$residuals
model2fit <- model2reg$fitted.values
plot(model2fit, model2resid,
     main="Figure 17-Multiple Regression(Model2) Residual vs Fitted Values Plot")
abline(h=0, col="red")
```

Figure 17−Multiple Regression(Model2) Residual vs Fitted Values Pl