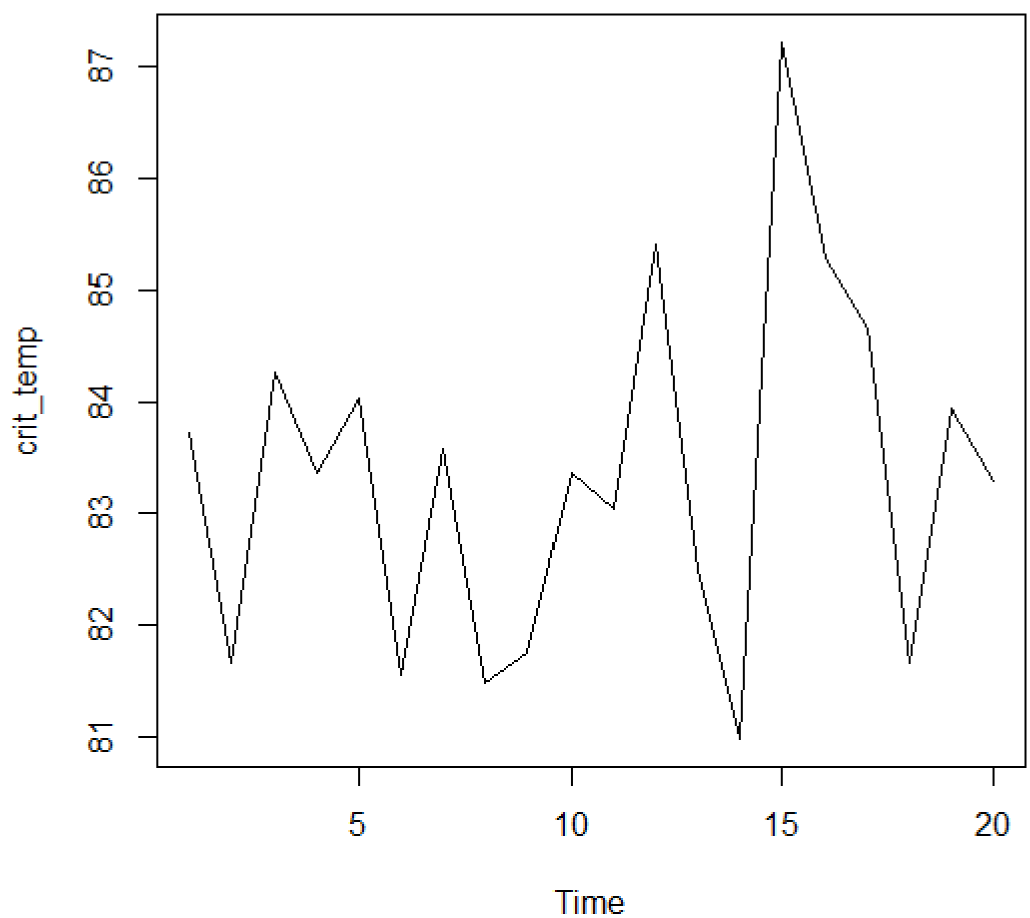
Question1:-

Weekly consumption of soda in my family. Every week we buy cases of Coco cola. We can use the data to do exponential smoothing. I would choose alpha value close to 1. Even though we get only few guest in a month I expect the consumption to be steady. Very low randomness

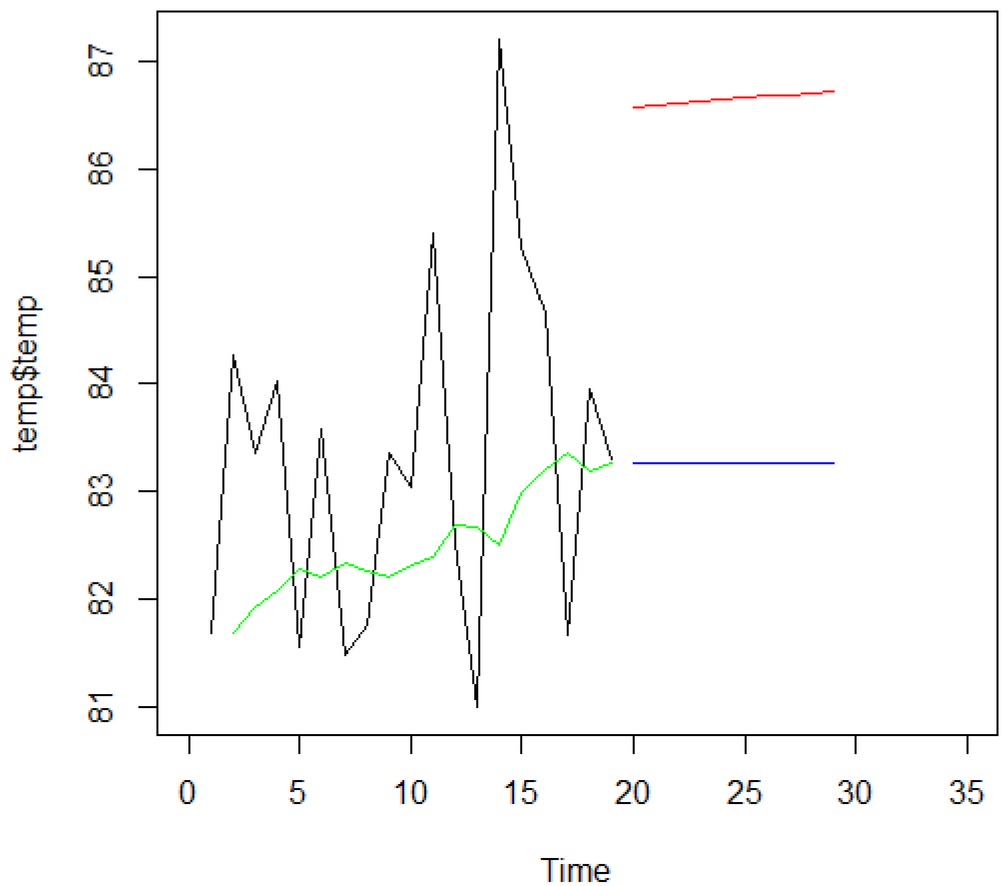
Question2:-

Plot the data to understand the data

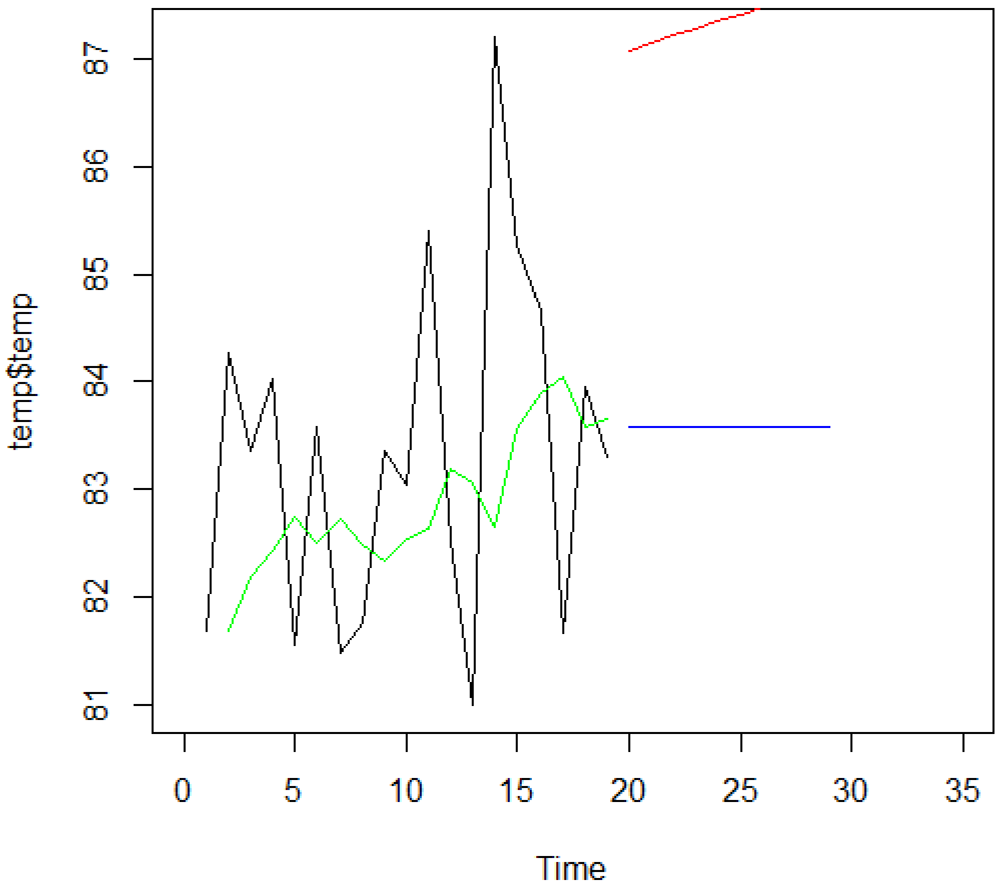
plot.ts(weatherts)



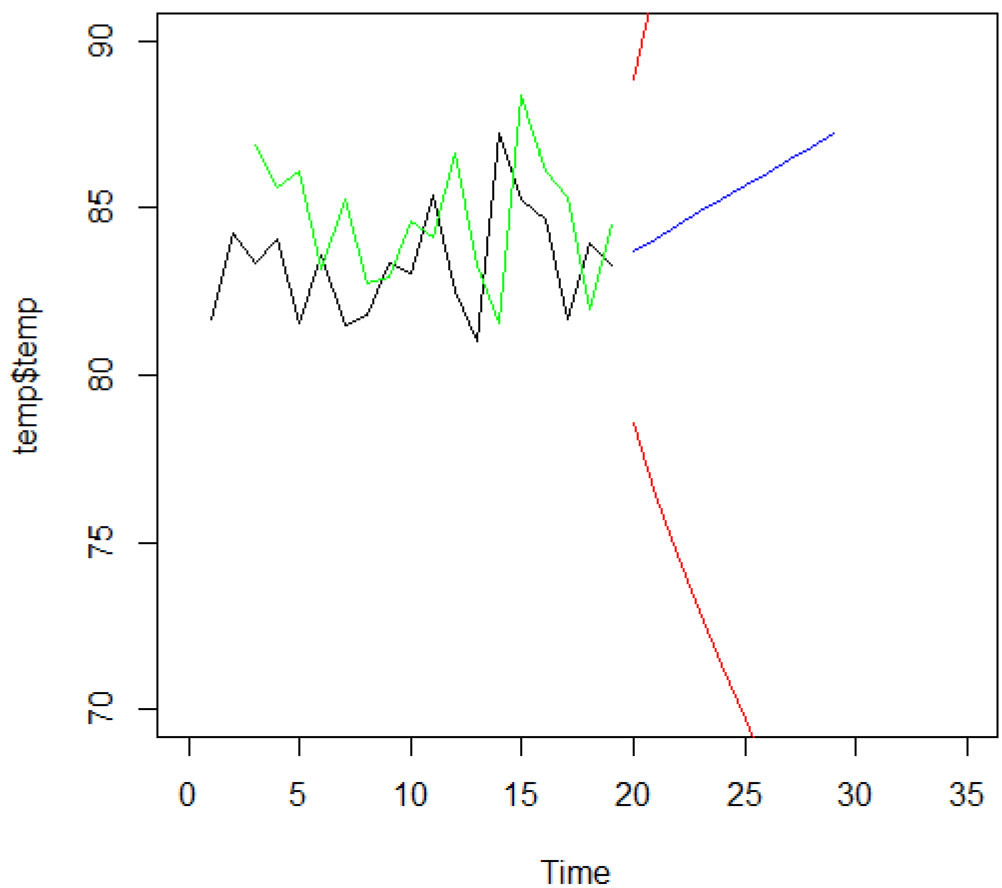
Try Alpha as .1 (exponential value) and beta as FALSE (trending) and gamma (seasonal values) as FALSE.



Try Alpha as .2 (exponential value) and beta as FALSE (trending) and gamma (seasonal values) as FALSE.

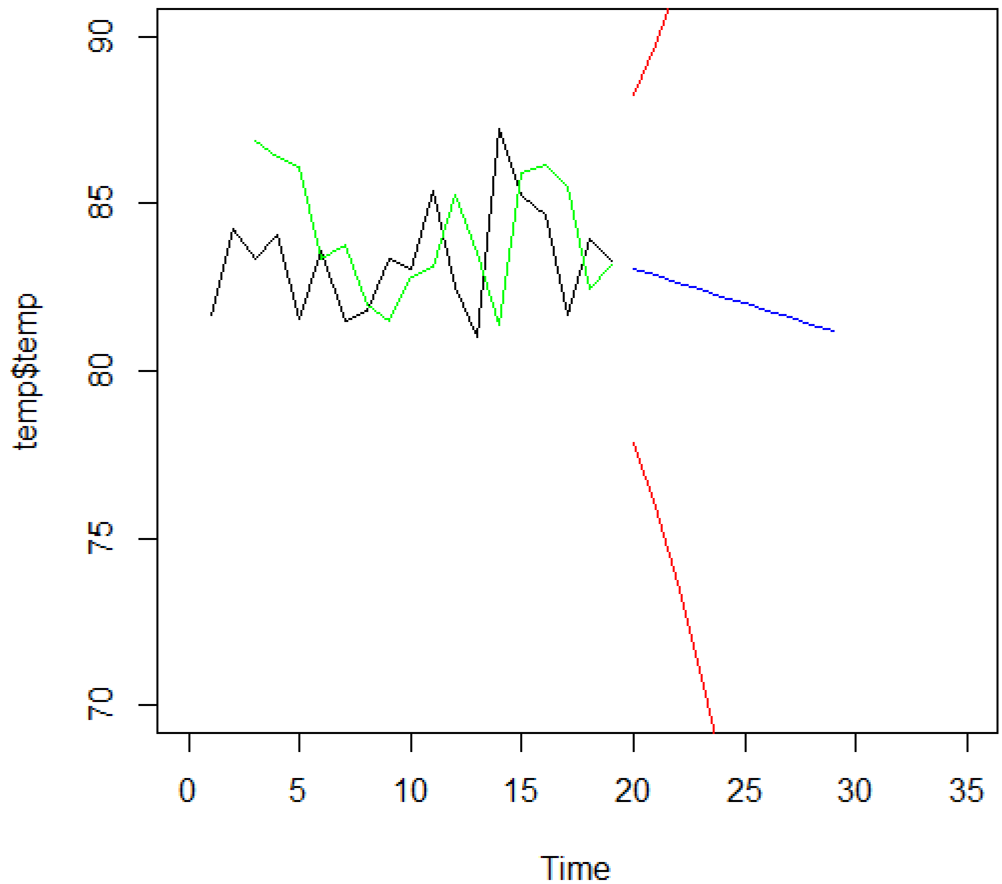


Try Alpha as 1 (exponential value) and beta as .1 (trending)   and gamma (seasonal values) as FALSE.



Letting R predict the value

```
> weather_mean
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = temp$temp, gamma = FALSE)

Smoothing parameters:
 alpha: 0.6422043
 beta : 0.3611714
 gamma: FALSE

Coefficients:
        [,1]
a 83.2556026
b -0.2077225
```



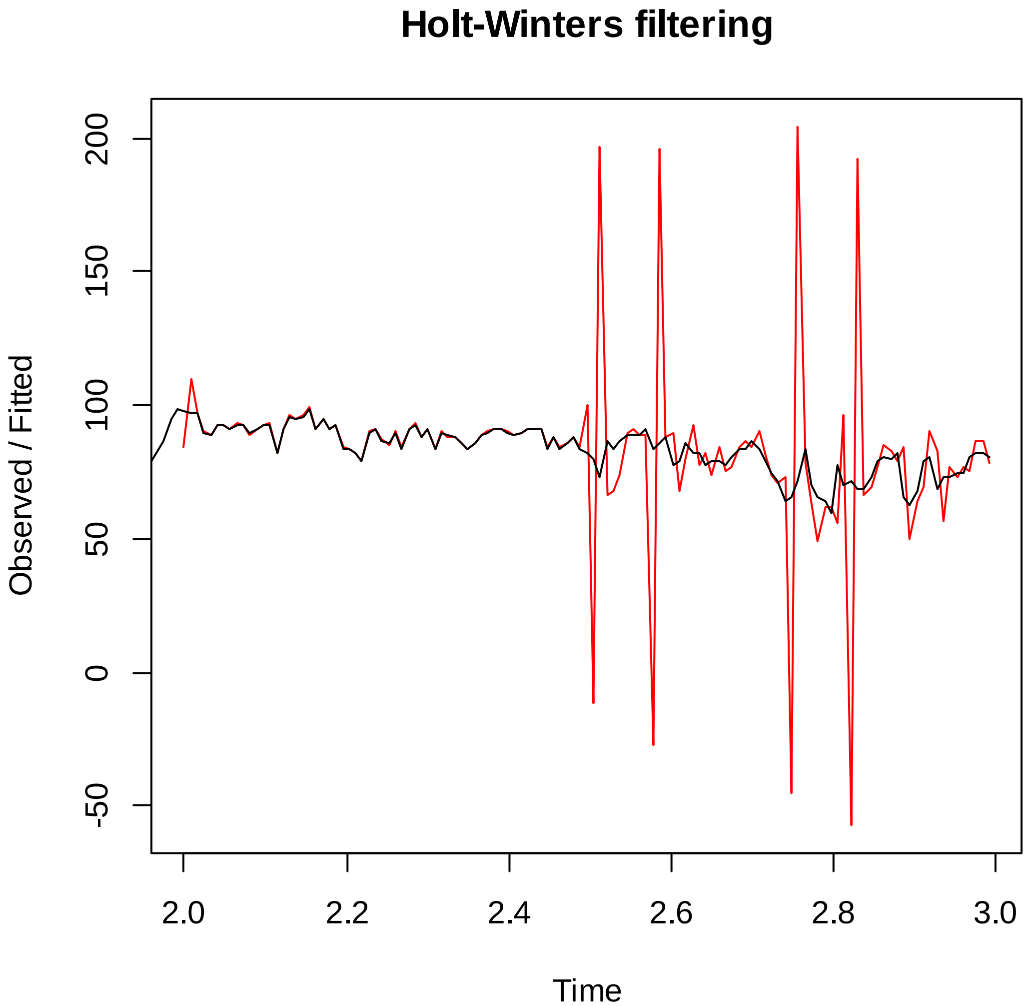
Rsource code:-

```
weather <- read.table("https://d37djvu3ytnwxt.cloudfront.net/assets/courseware
print(head(weather))
crit_temp <- data.frame(sapply(weather[,2:21],mean,header=TRUE))
temp <-crit_temp[2:21,]
temp <- temp[!is.na(temp)]
temp=data.frame(temp)
weather_year <- substr(names(weather)[2:21],2,5)
weather_mean <- HoltWinters(temp$temp,
                              #alpha=1,
                              #beta= 0.1,
                              gamma = FALSE)
weather.predict <- predict(weather_mean,n.ahead=10,prediction.interval=TRUE)
plot.ts(temp$temp,xlim=c(0,35),ylim=c(70,90))
lines(weather_mean$fitted[,1],col="green")
lines(weather.predict[,1],col="blue")
lines(weather.predict[,2],col="red")
lines(weather.predict[,3],col="red")
```

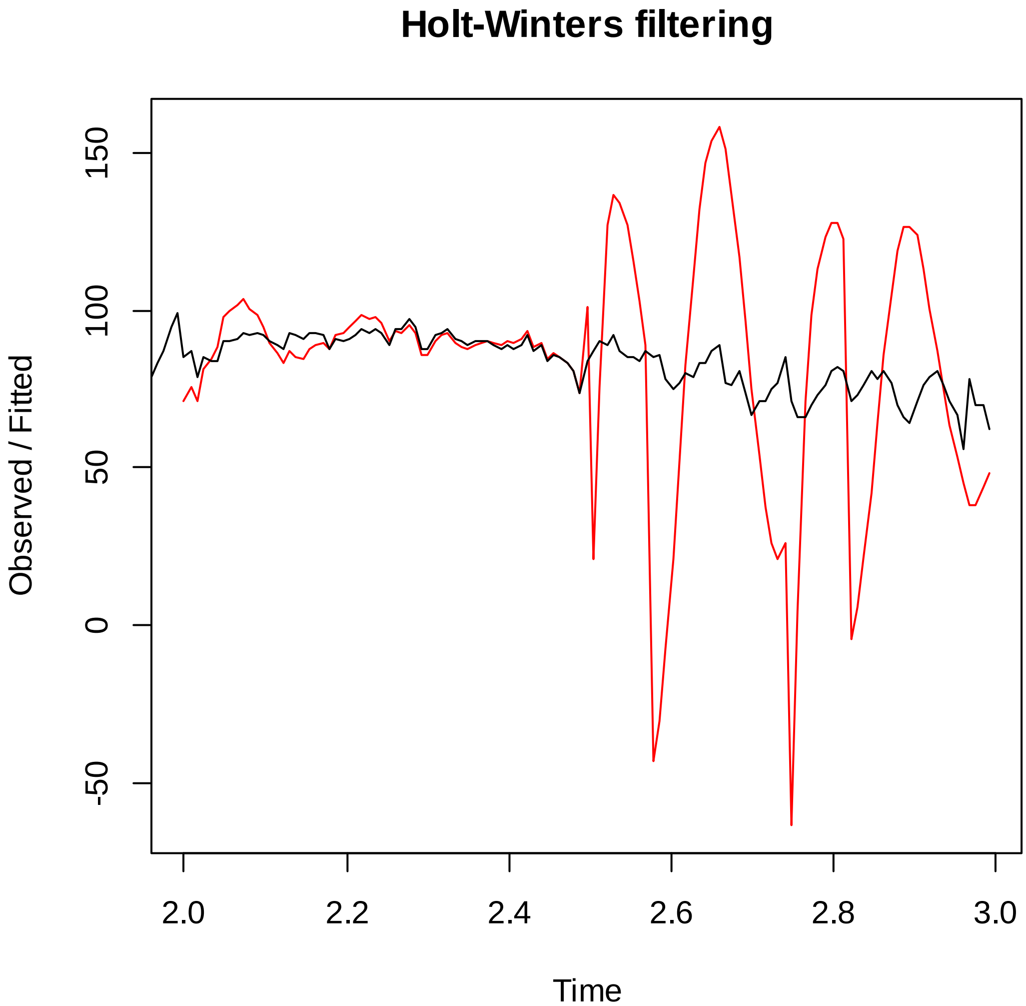
Conclusion:- The trend is upward. So it is difficult to say. However if we forecast with current data then the graph is showing the cool temperatures by end of August and temp. seems to be going to after August

Comparing 1996 data with 2015 data to see warming pattern:

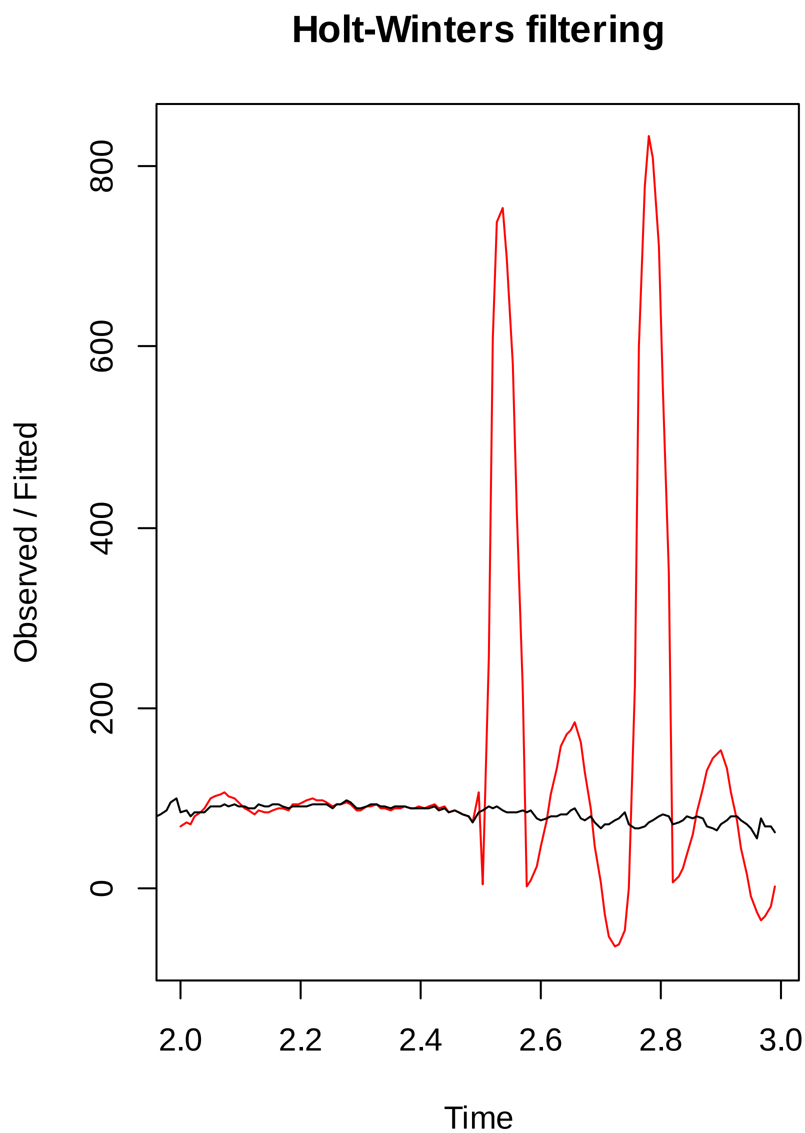
See attached 1996 and 2015 sheet. Based on this I can’t tell if the global warming is happening or not.2015 vs 1996



Holt-Winters filtering
Observed / Fitted
200
150
100
50
0
-50
2.0
2.2
2.4
2.6
2.8
3.0
Time


Holt-Winters filtering
Observed / Fitted
Time
150
100
50
0
-50
2.0
2.2
2.4
2.6
2.8
3.0

Using seasonal = multiplicative. Was not helpful in this case see chart below



Question3::-

Problem: - Reaching Work on time

Predictors:-

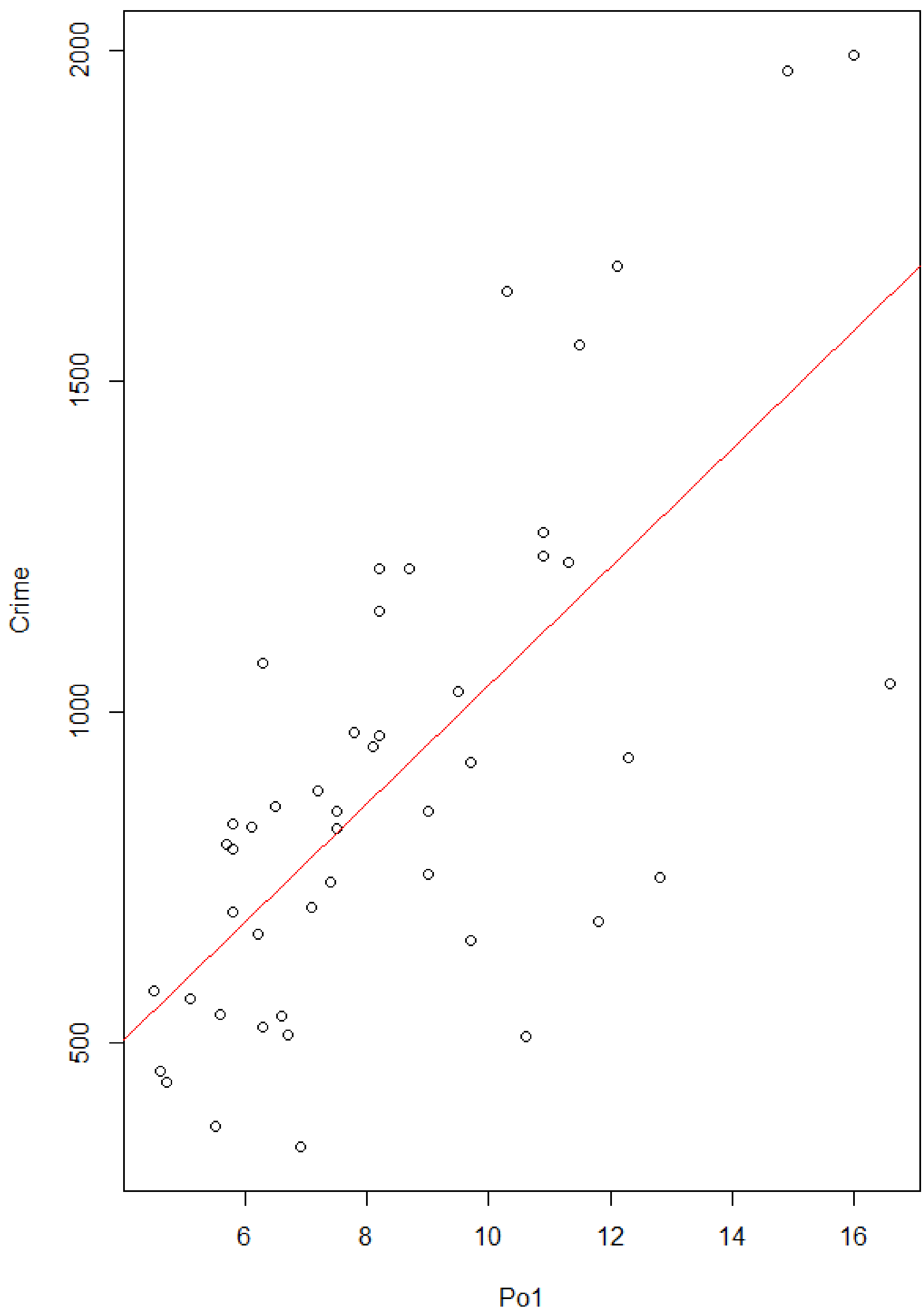
1. Start time
2. Traffic in 101.
3. School traffic.
4. Elevator wait time
5. Speed

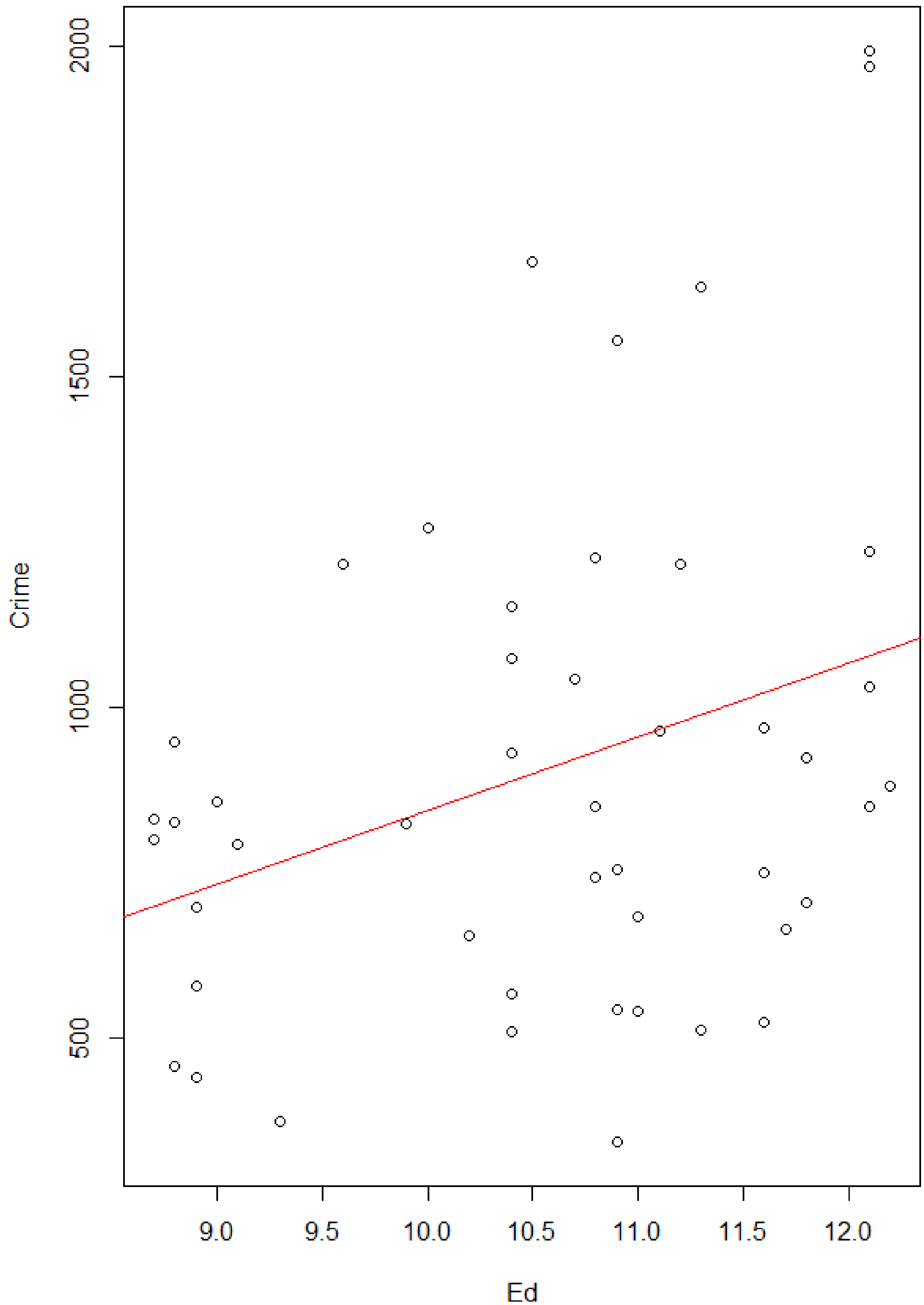
Question 4:-

```
> model = lm(Crime~.,data=crime)
> model

Call:
lm(formula = Crime ~ ., data = crime)

Coefficients:
(Intercept)            M           So
 -5.984e+03    8.783e+01   -3.803e+00
         Ed          Po1          Po2
  1.883e+02    1.928e+02   -1.094e+02
         LF          M.F          Pop
 -6.638e+02    1.741e+01   -7.330e-01
         NW           U1           U2
  4.204e+00   -5.827e+03    1.678e+02
     Wealth         Ineq         Prob
  9.617e-02    7.067e+01   -4.855e+03
       Time
 -3.479e+00
```

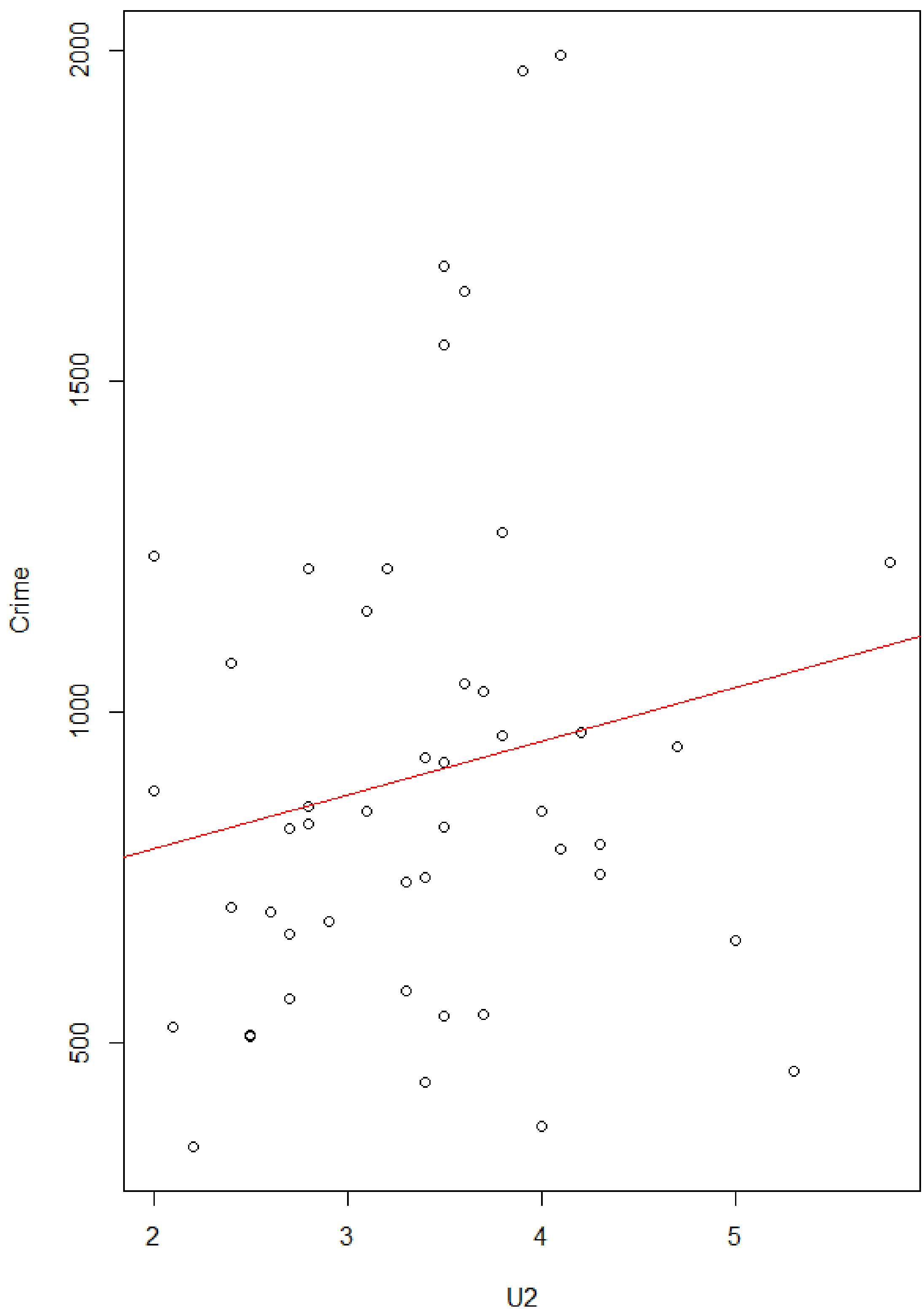
Picking the low Coefficients

Ed	mean years of schooling of the population aged 25 years or over
Po1	per capita expenditure on police protection in 1960
U2	unemployment rate of urban males 35–39


Crime
Po1
500
1000
1500
2000
6
8
10
12
14
16

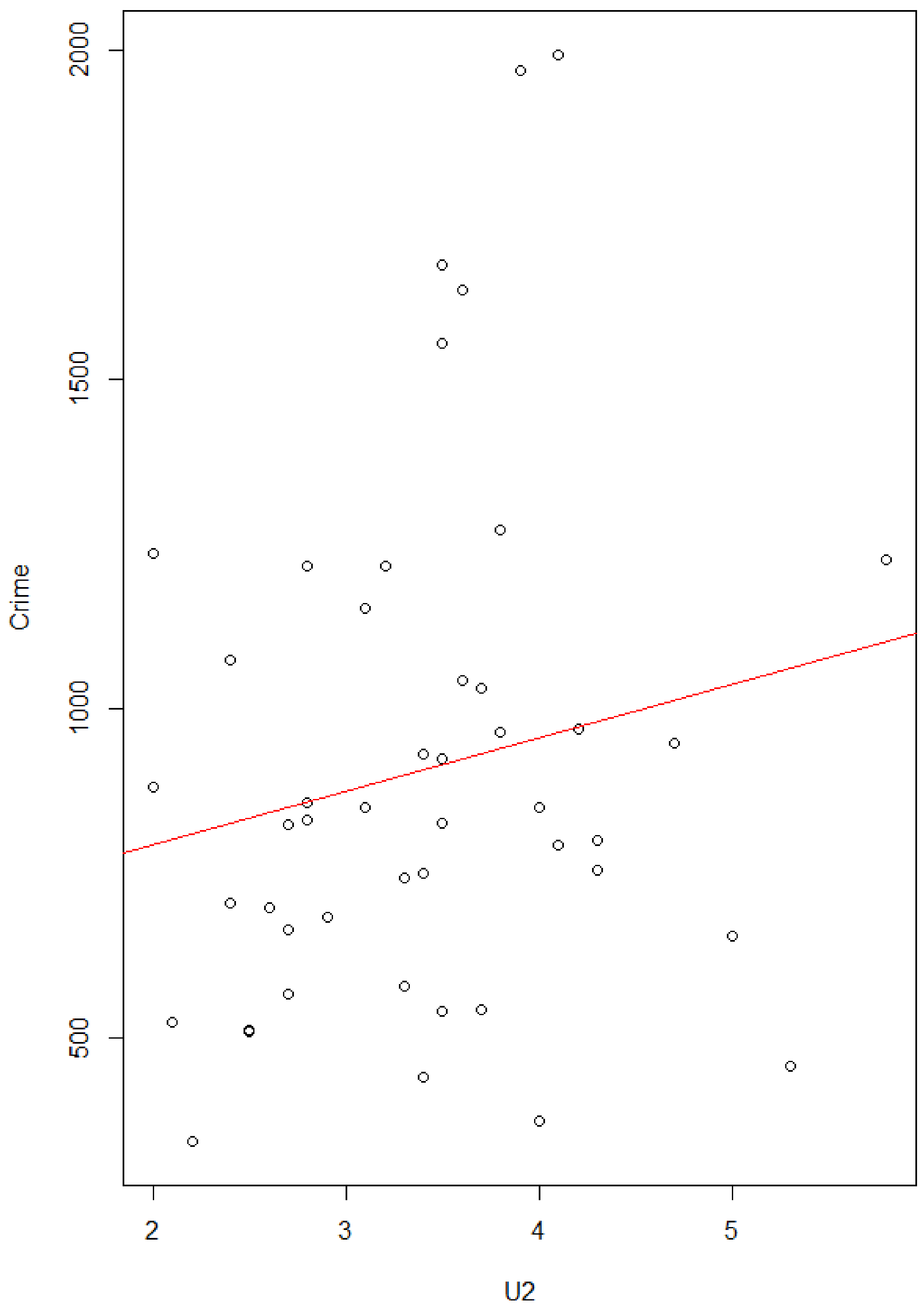


2000
1500
1000
500
Crime
9.0
9.5
10.0
10.5
11.0
11.5
12.0
Ed

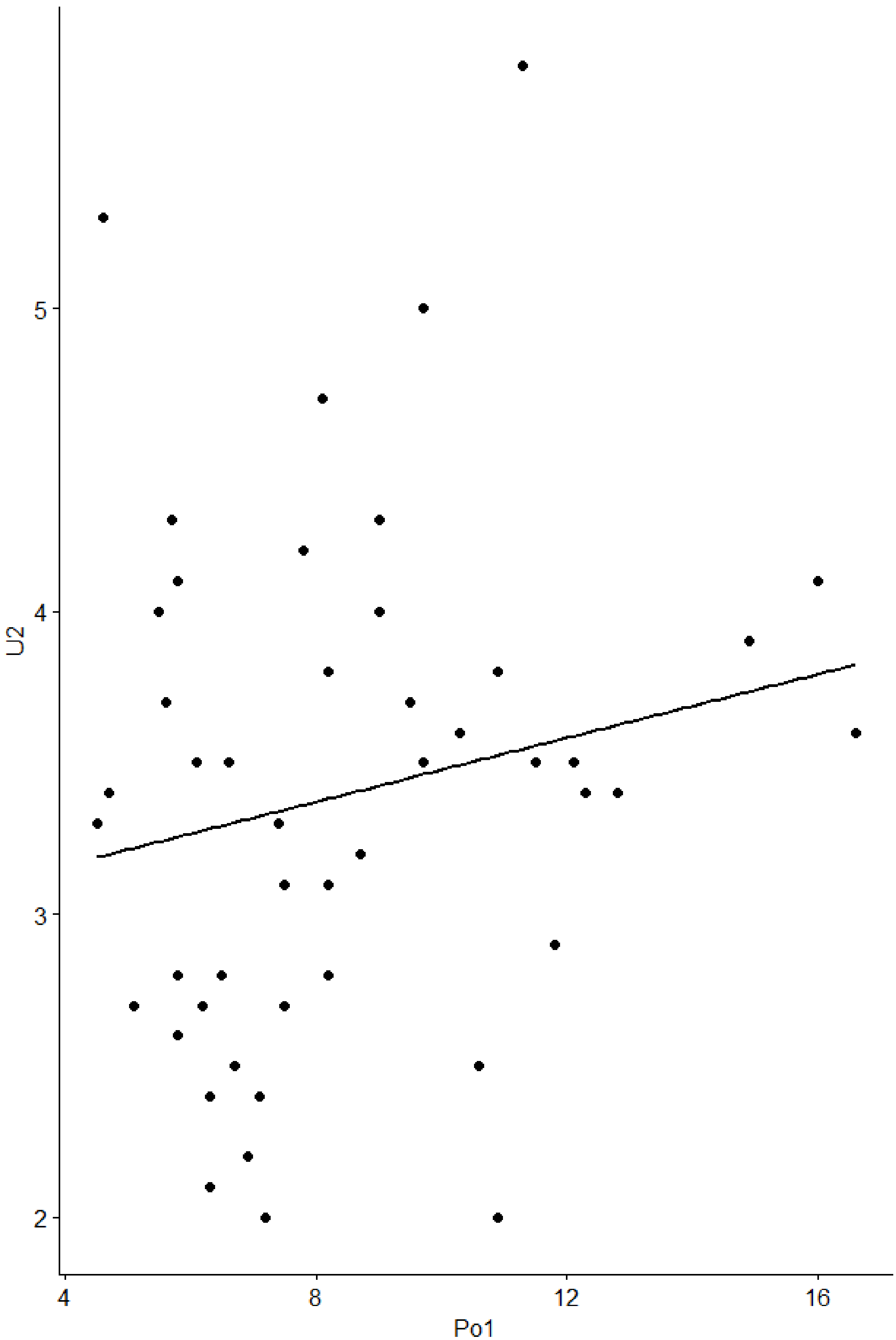


2000
1500
1000
500
Crime
2
3
4
5
U2

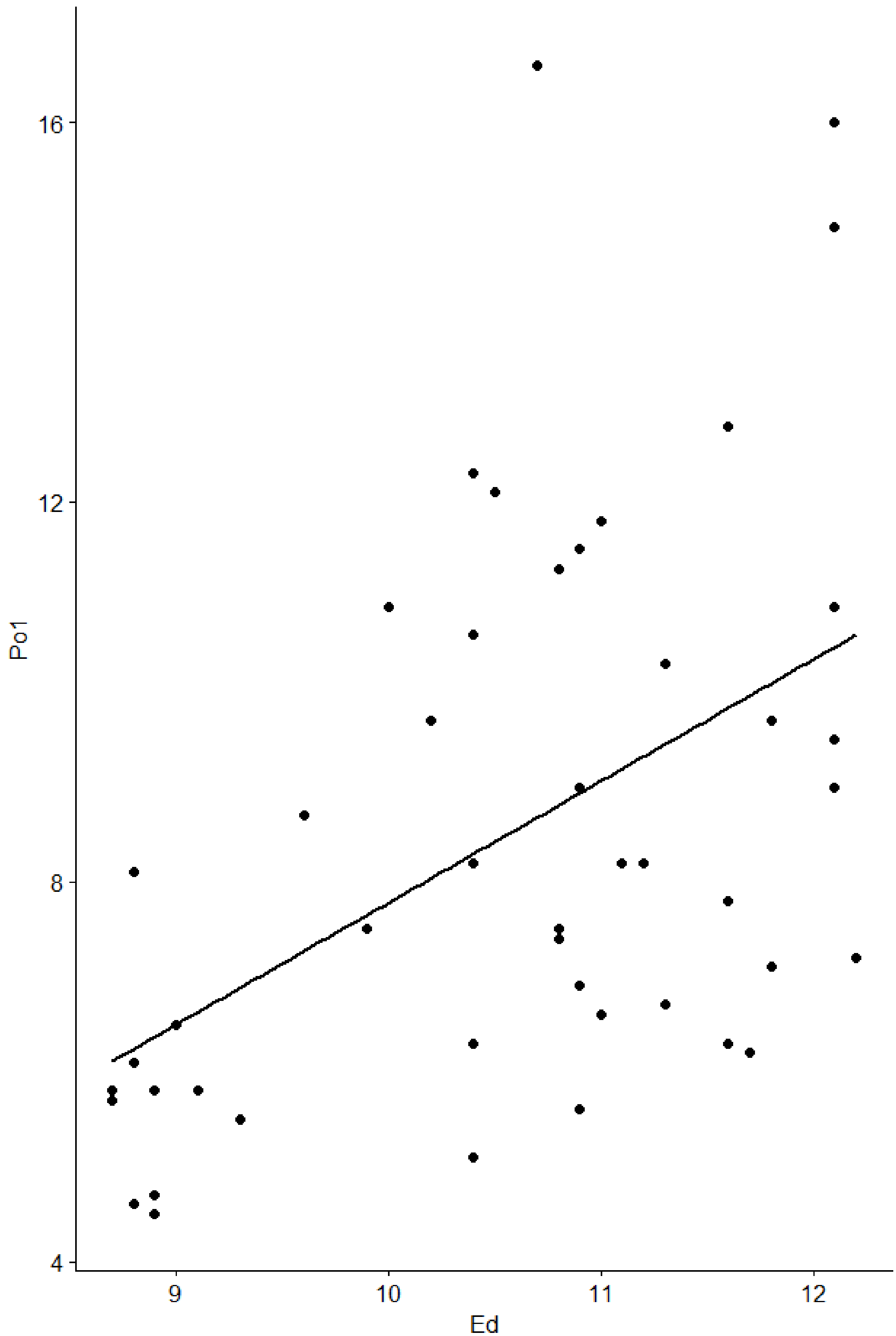
Now Correlation:- between various predictors



Crime
2000
1500
1000
500
2
3
4
5
U2



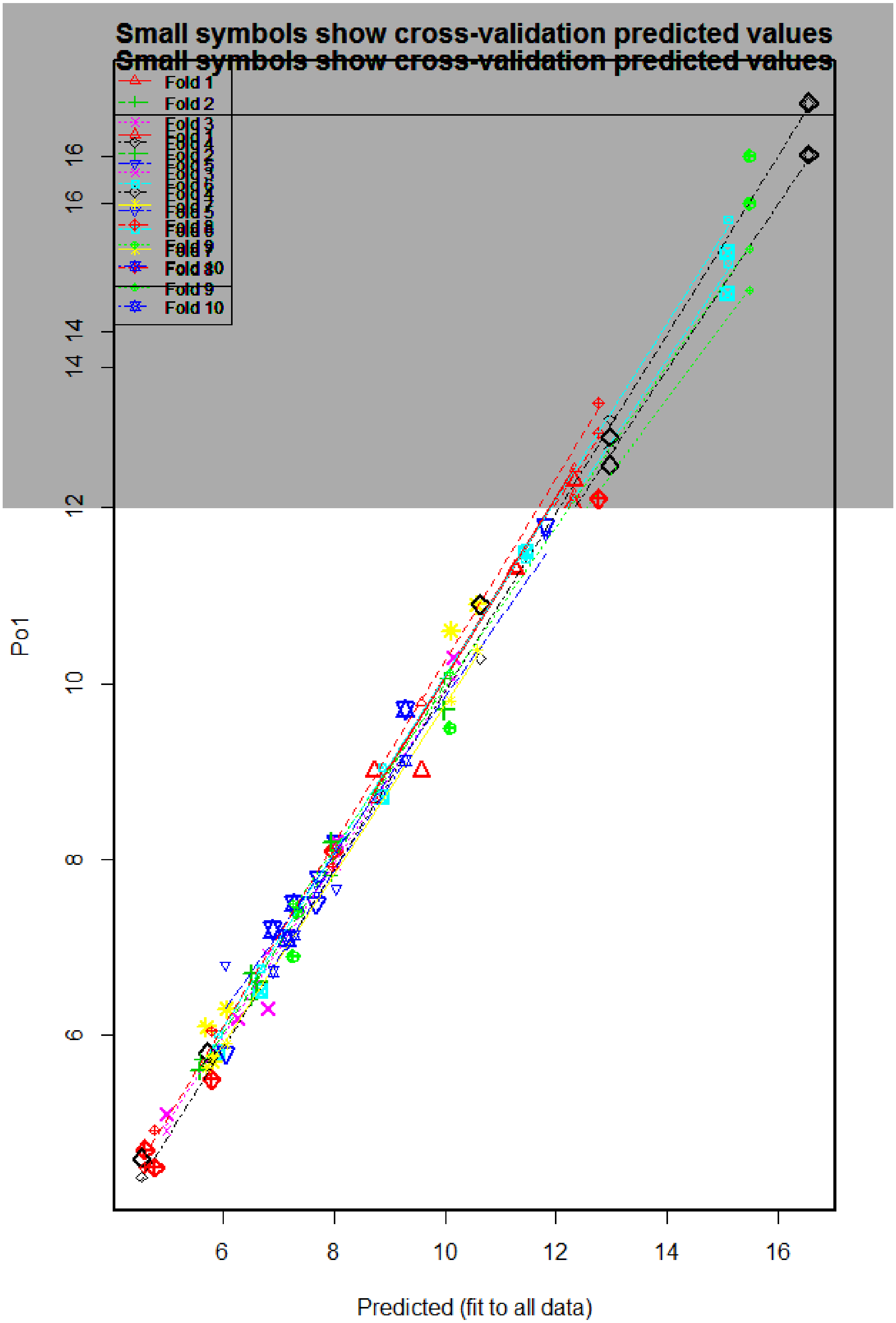
U2
5
4
3
2
4
8
12
16
Po1


Po1
16
12
8
4
9
10
11
12
Ed

Decision time:- Iam not a crime expert. Even though there is a positive correlation I will use all three variable.

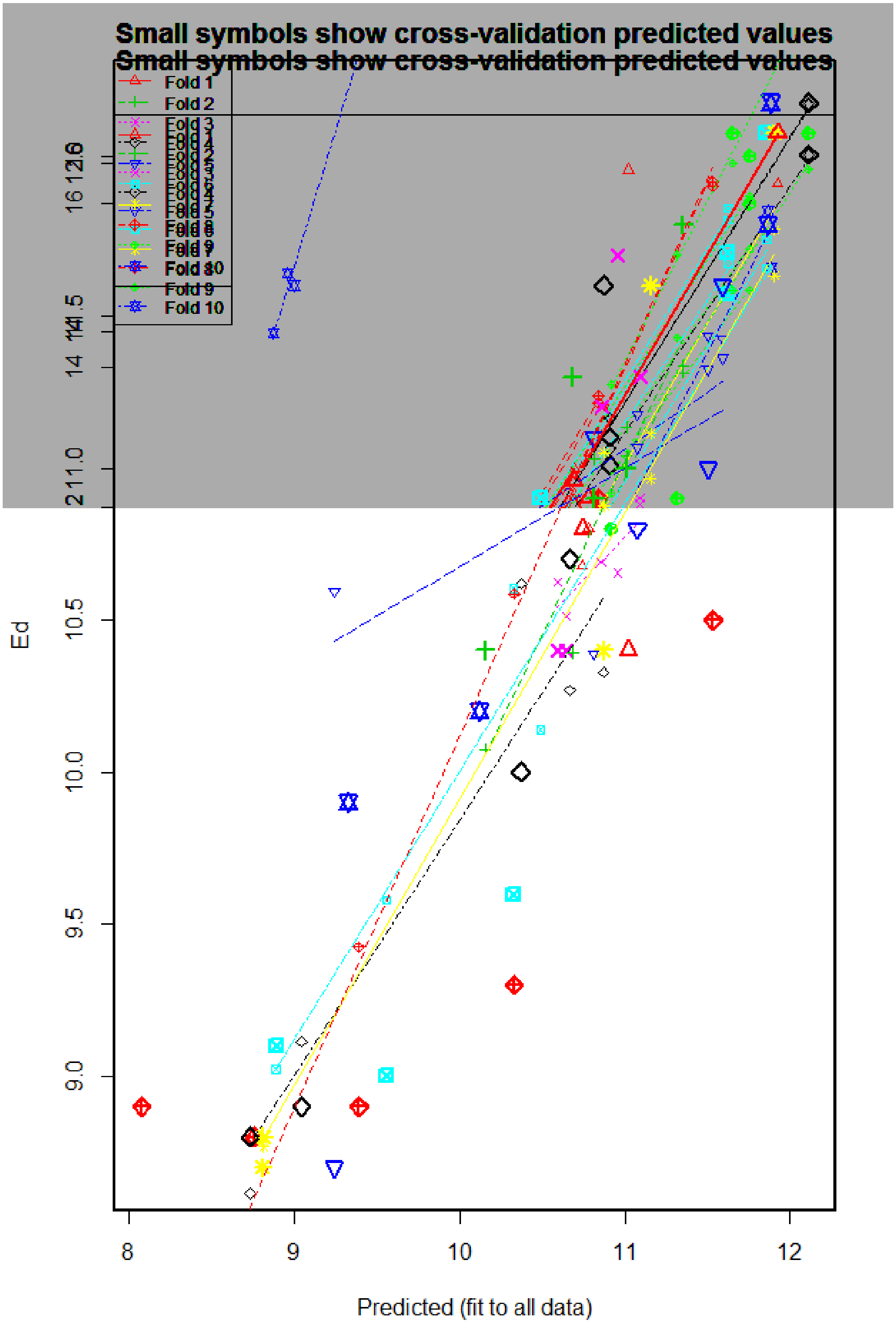
```
> model = glm(Crime~Ed + Po1 +U2,data=crime)

> model


Call:  glm(formula = Crime ~ Ed + Po1 + U2, data = crime)


Coefficients:

(Intercept)          Ed         Po1          U2

     39.716       3.386      87.546      25.150


Degrees of Freedom: 46 Total (i.e. Null);  43 Residual

Null Deviance:      6881000

Residual Deviance: 3609000      AIC: 672.1
```

Correlation continued..
Quality of FIT:-


Small symbols show cross-validation predicted values
Small symbols show cross-validation predicted values
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Fold 6
Fold 7
Fold 8
Fold 9
Fold 10
Po1
Predicted (fit to all data)
6
8
10
12
14
16

```
> sqrt(mean((mod1$cvpred - mod1$Po1)^2))
[1] 0.45
> sqrt(mean((mod2$cvpred - mod2$Po1)^2))
[1] 2.3
> sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
[1] 0.207
> sqrt(mean((mod2$cvpred - mod2$Predicted)^2))
[1] 2.33
>
```

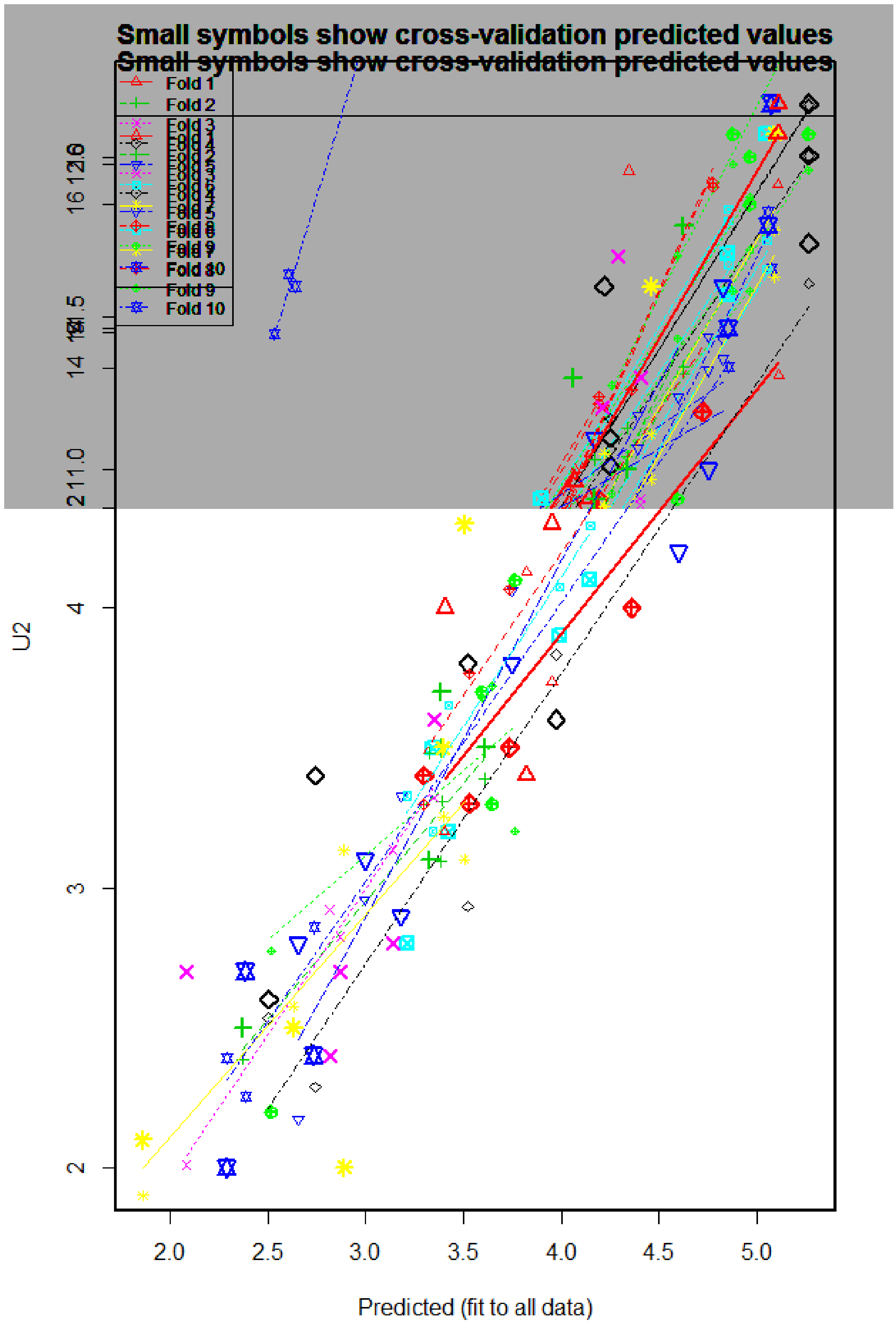
Ed:-



Small symbols show cross-validation predicted values
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Fold 6
Fold 7
Fold 8
Fold 9
Fold 10
Ed
Predicted (fit to all data)

```
> sqrt(mean((mod1$cvpred - mod1$Ed)^2))
[1] 0.685
> sqrt(mean((mod2$cvpred - mod2$Ed)^2))
[1] 3.77
> sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
[1] 0.32
> sqrt(mean((mod2$cvpred - mod2$Predicted)^2))
[1] 3.78
>
```

U2


Small symbols show cross-validation predicted values
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Fold 6
Fold 7
Fold 8
Fold 9
Fold 10
U2
2
3
4
Predicted (fit to all data)
2.0
2.5
3.0
3.5
4.0
4.5
5.0

```
> sqrt(mean((mod1$cvpred - mod1$U2)^2))
[1] 0.536
> sqrt(mean((mod2$cvpred - mod2$U2)^2))
[1] 1.4
> sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
[1] 0.233
> sqrt(mean((mod2$cvpred - mod2$Predicted)^2))
[1] 1.32
>
```

Results:-

```
> test1 <- test[,c("Ed","Po1","U2")]
> predict(model, newdata=test1)
   1
1215
>
```

R Code:-

```
crime <- read.table("http://www.statsci.org/data/general/uscrime.txt",sep="\t", header=TRUE)
print(head(crime))
names(crime)
#[1] "M"      "So"     "Ed"     "Po1"    "Po2"    "LF"     "M.F"    "Pop"
#[9] "NW"     "U1"     "U2"     "Wealth" "Ineq"   "Prob"   "Time"   "Crime"

modelall = glm(Crime~.,data=crime)
modelall

model = glm(Crime~Ed + Po1 +U2,data=crime)
model


plot(Crime~Po1,data=crime,na.rm=TRUE)
#calculate mean of Crime NA removed
mean.Po1 <- mean(crime$Po1)
abline(h=mean.Po1)
#use lm to fit a regression line
modelPo1 = lm(Crime~Po1,data=crime)
modelPo1
abline(modelPo1,col='red')
plot (modelPo1)



plot(Crime~Ed,data=crime,na.rm=TRUE)
#calculate mean of Crime NA removed
mean.Ed <- mean(crime$Ed)
abline(h=mean.Ed)
#use lm to fit a regression line
modelEd = lm(Crime~Ed,data=crime)
modelEd
abline(modelEd,col='red')
plot (modelEd)


plot(Crime~U2,data=crime,na.rm=TRUE)
#calculate mean of Crime NA removed
mean.U2 <- mean(crime$U2)
abline(h=mean.U2)
#use lm to fit a regression line
modelU2 = lm(Crime~U2,data=crime)
modelU2
abline(modelU2,col='red')
plot (modelU2)

ggscatter(crime,x="Ed", y="U2",add="reg.line",cor.method = "pearson")
ggscatter(crime,x="Po1", y="U2",add="reg.line",cor.method = "pearson")
ggscatter(crime,x="Ed", y="Po1",add="reg.line",cor.method = "pearson")

test <- data.frame(M=14,So=0,Ed=10.0,Po1=12.0,Po2=15.5,LF=.64,M.F=94,pop=150,NW=1.1,U1=.12,U2=3.6,Wealth=3200,Ineq=20.1,Prob=.04,Time=39.0)

fit <- lm(Po1 ~ ., data=crime)
mod1 <- cv.lm(crime,fit,m=10)
mod2 <- cv.lm(crime,Po1 ~ .,m=10)
sqrt(mean((mod1$cvpred - mod1$Po1)^2))
sqrt(mean((mod2$cvpred - mod2$Po1)^2))
sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
sqrt(mean((mod2$cvpred - mod2$Predicted)^2))
```

```
fit <- lm(Ed ~ ., data=crime)
mod1 <- cv.lm(crime,fit,m=10)
mod2 <- cv.lm(crime,Ed ~ .,m=10)
sqrt(mean((mod1$cvpred - mod1$Ed)^2))
sqrt(mean((mod2$cvpred - mod2$Ed)^2))
sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
sqrt(mean((mod2$cvpred - mod2$Predicted)^2))


fit <- lm(U2 ~ ., data=crime)
mod1 <- cv.lm(crime,fit,m=10)
mod2 <- cv.lm(crime,U2 ~ .,m=10)
sqrt(mean((mod1$cvpred - mod1$U2)^2))
sqrt(mean((mod2$cvpred - mod2$U2)^2))
sqrt(mean((mod1$cvpred - mod1$Predicted)^2))
sqrt(mean((mod2$cvpred - mod2$Predicted)^2))


predict(model, newdata=test1, interval='confidence')
```