Group 01 Landa

Landa Gang

2025-05-05

```
setwd("C:/Users/chemi/Desktop")
datao2 = read.csv("project_data.csv", header = TRUE)
datao = datao2[14:650,]
rownames(datao) <- NULL
datao3 = datao

library(caret)
```

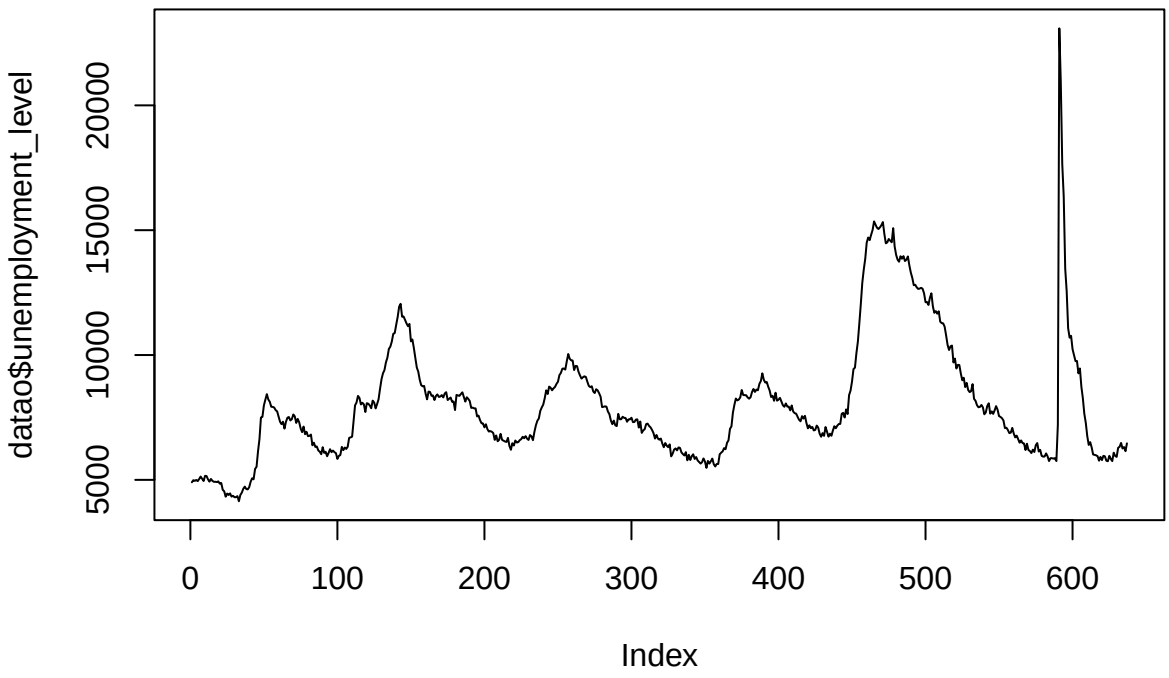
```
## Warning: package 'caret' was built under R version 4.3.3
```

```
## Loading required package: ggplot2
```

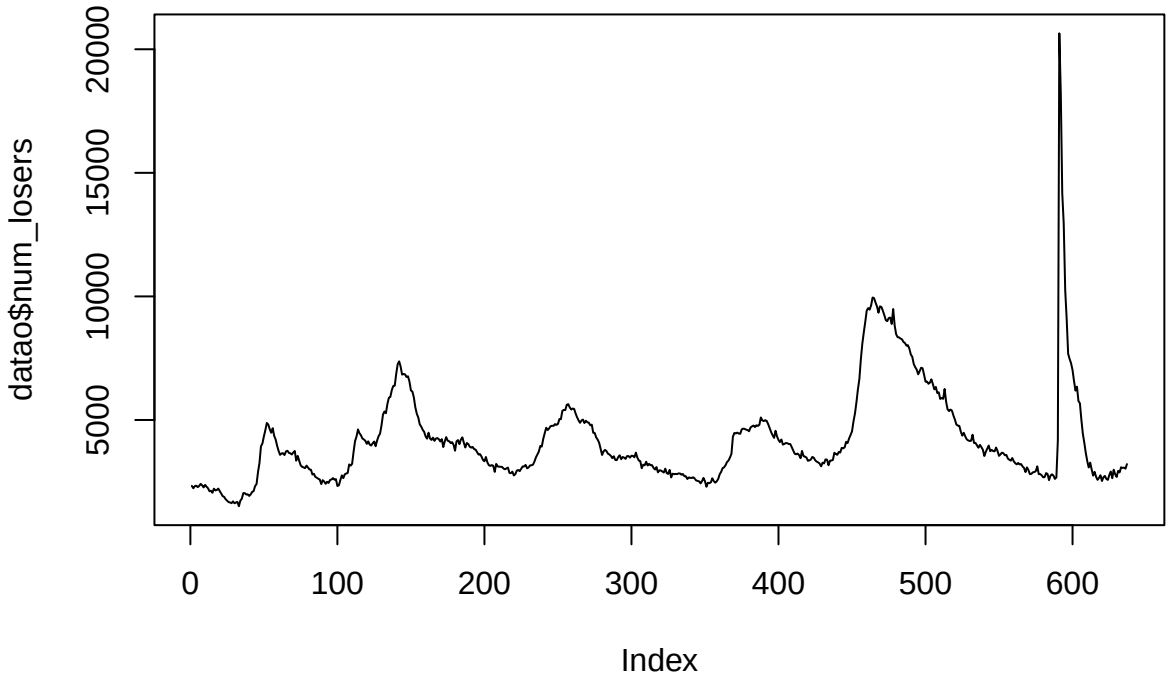
```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Loading required package: lattice
```

```
#description and preprocess
#unemployment
plot(datao$unemployment_level, type = "l")
```



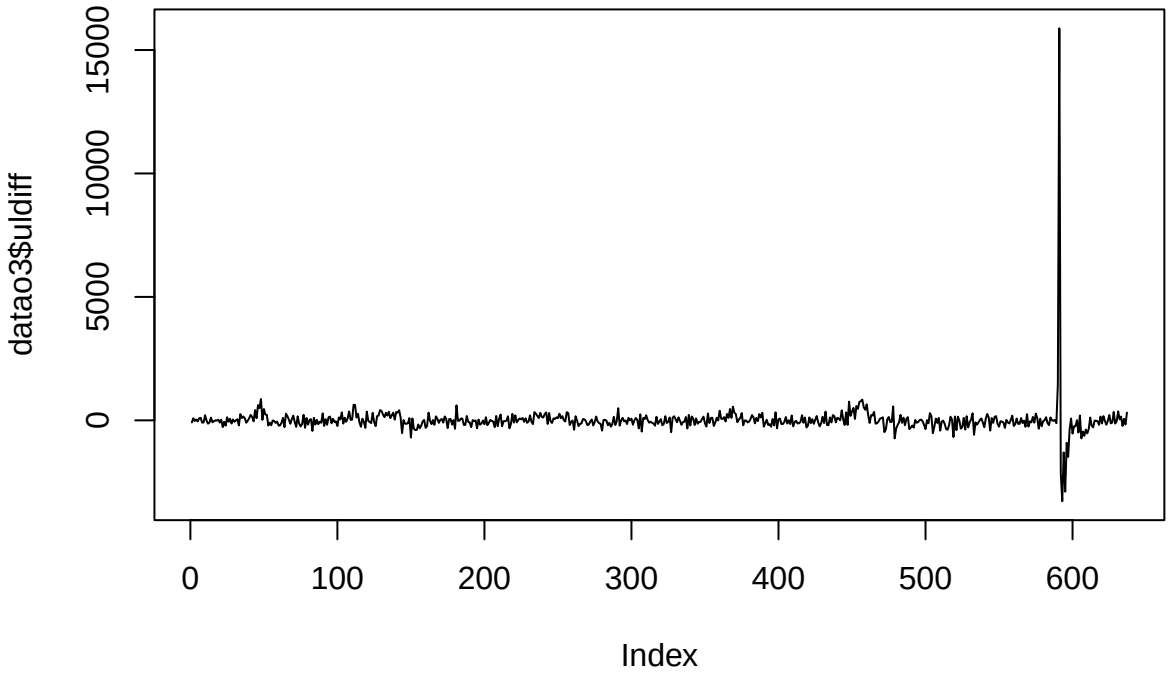

```
plot(datao$num_losers, type = "l")
```



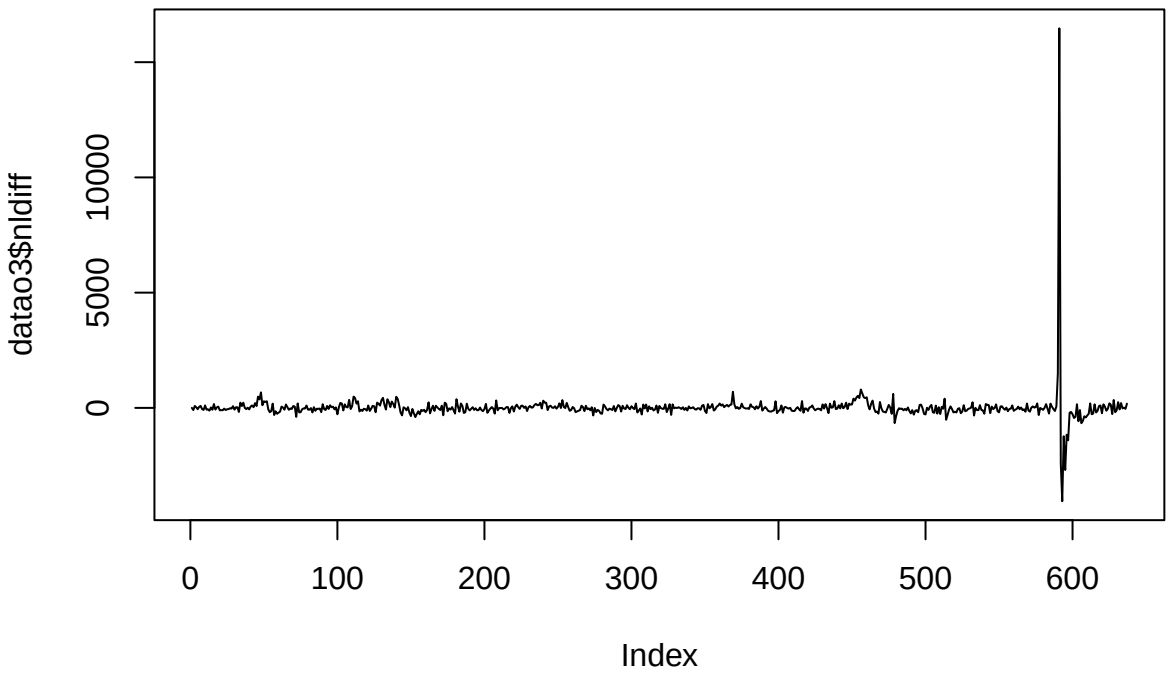

```
cor(datao$unemployment_level, datao$num_losers)
```

```
## [1] 0.9770221
```

```
datao3$uldiff = datao2$unemployment_level[14:650]-datao2$unemployment_level[13:649]
datao3$nldiff = datao2$num_losers[14:650]-datao2$num_losers[13:649]
plot(datao3$uldiff, type = "l")
```



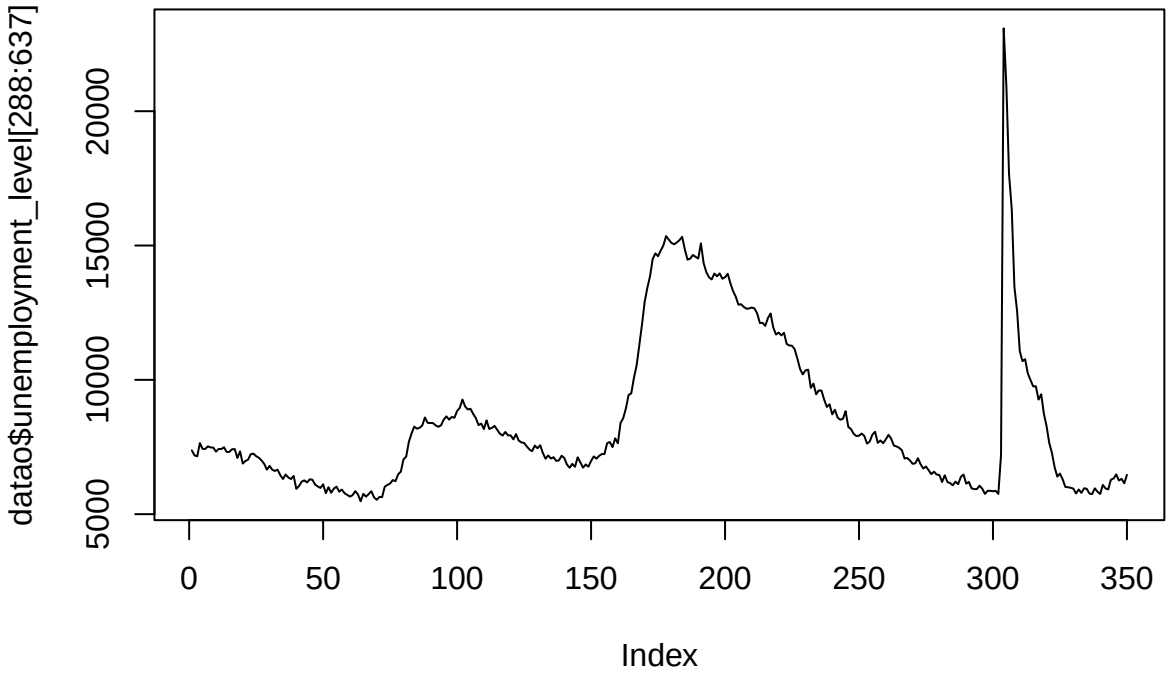

```
plot(datao3$nldiff, type = "l")
```



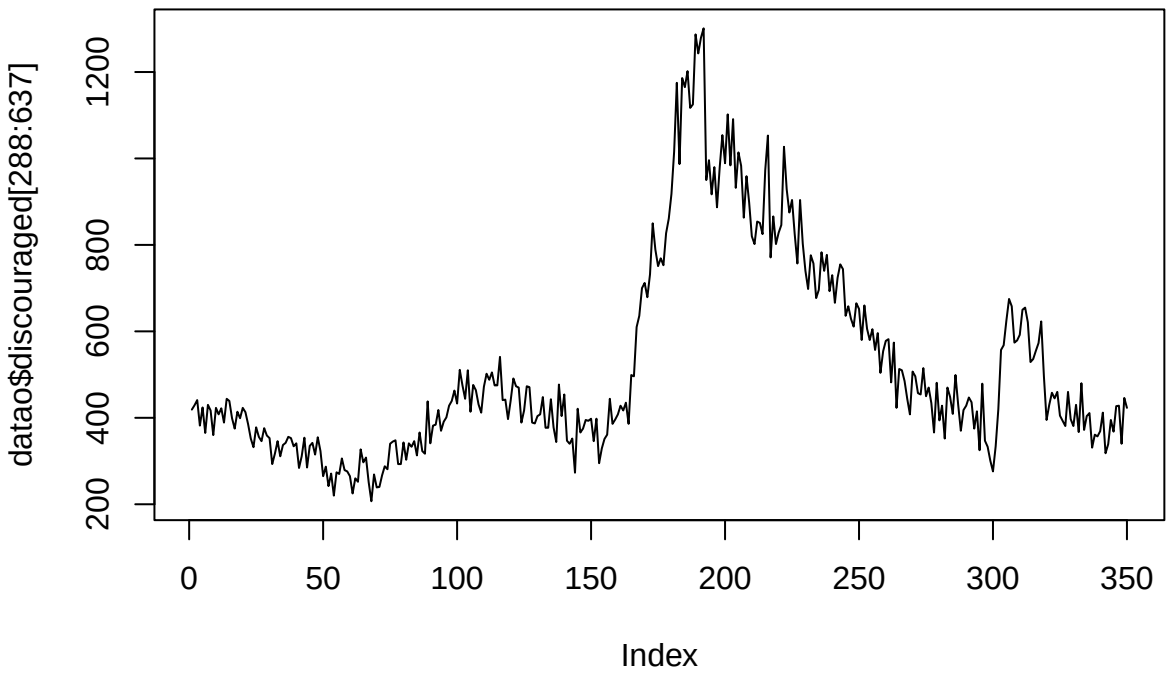

```
cor(datao3$uldiff, datao3$nldiff)
```

```
## [1] 0.9796417
```

```
plot(datao$unemployment_level[288:637], type = "l")
```




```
plot(datao$discouraged[288:637], type = "l")
```



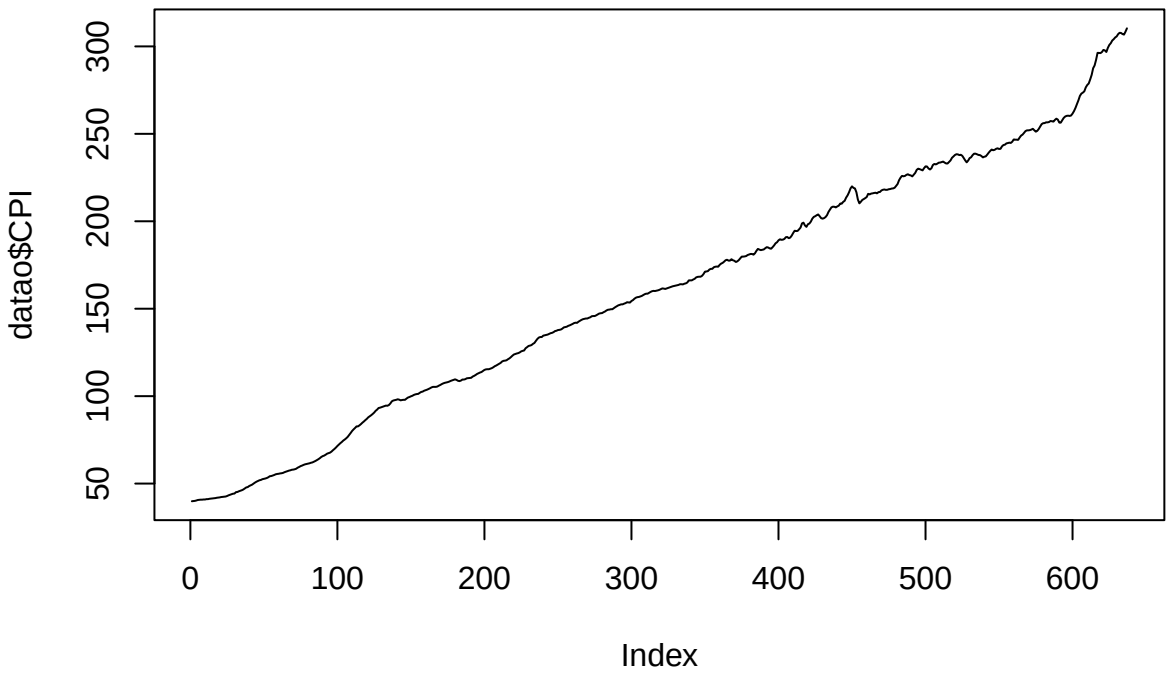

```
cor(datao$unemployment_level[288:637],
    datao$discouraged[288:637])
```

```
## [1] 0.8313918
```

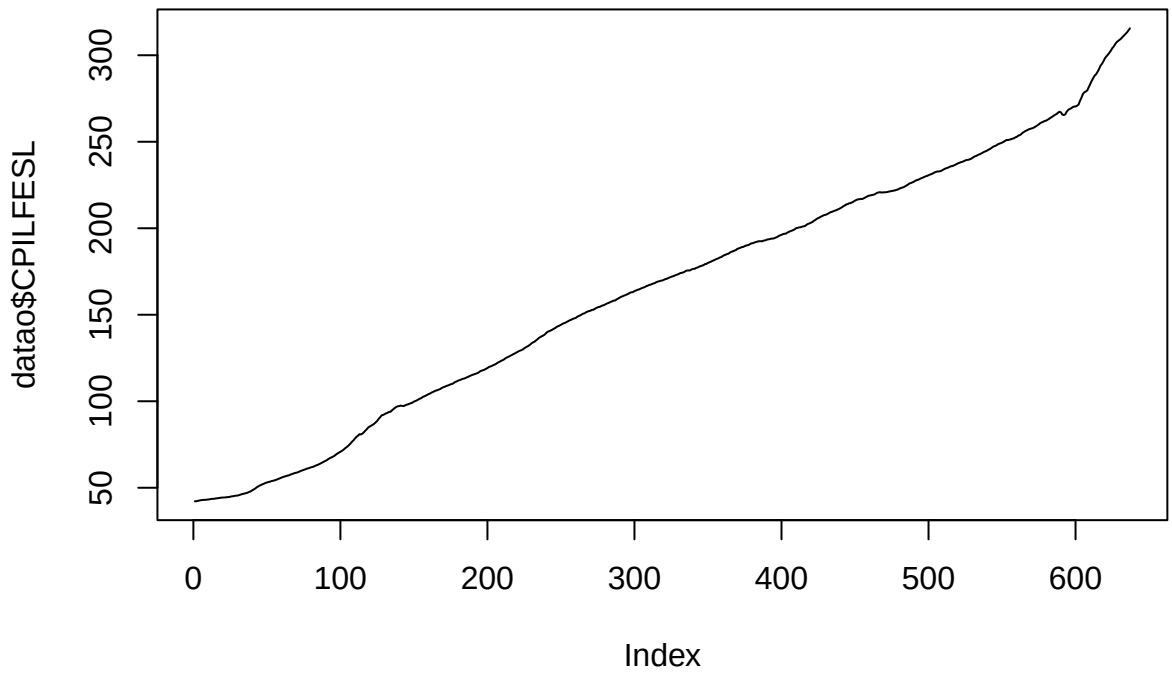
```
datao3$disdiff = datao2$discouraged[14:650]-datao2$discouraged[13:649]

#inflation
cor(datao$CPI, datao$CPILFESL)
```

```
## [1] 0.9988891
```

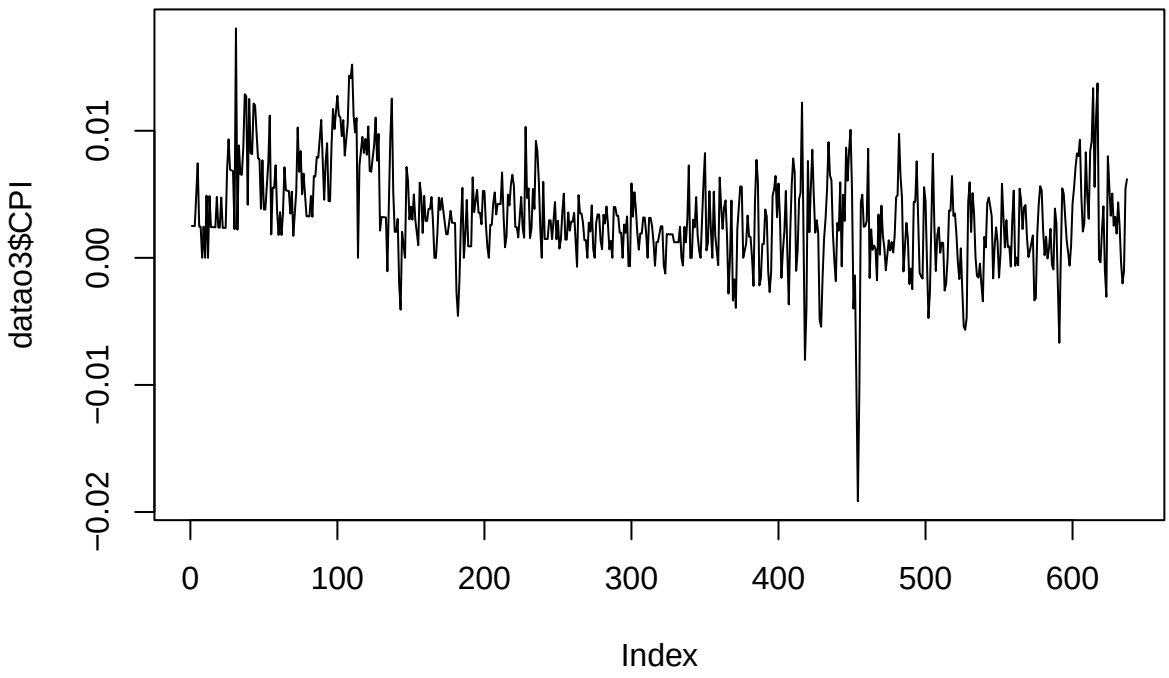
```
plot(datao$CPI, type = "l")
```



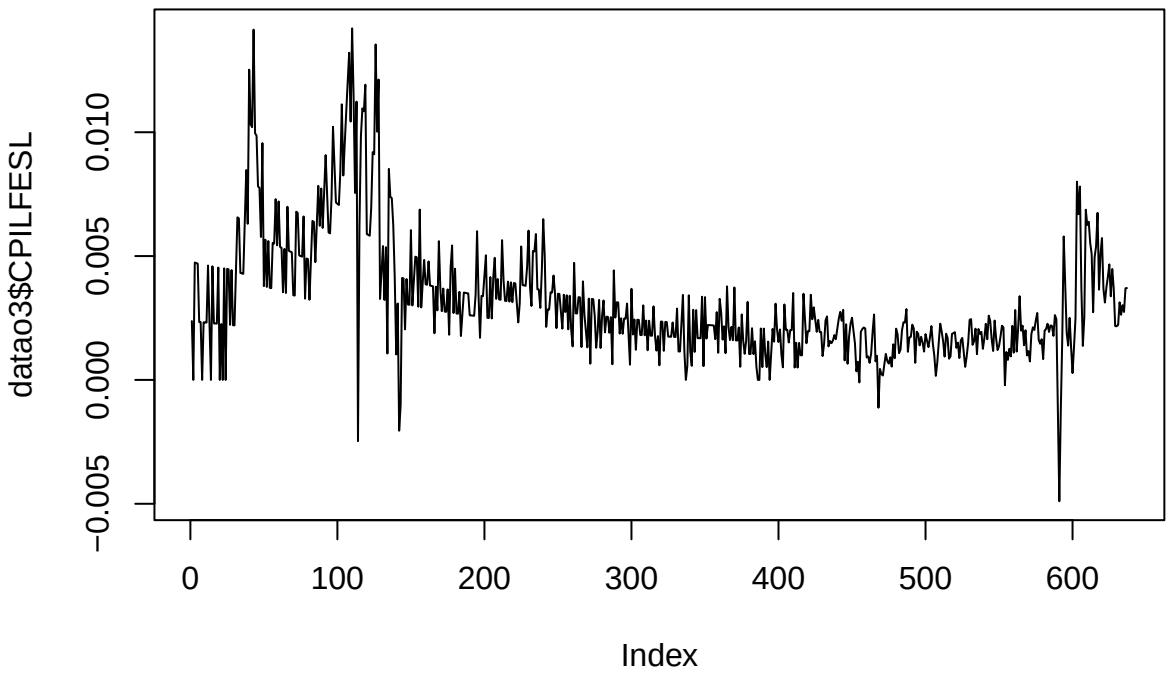

```
plot(datao$CPILFESL, type = "l")
```




```
datao3$CPI = datao2$CPI[14:650]/datao2$CPI[13:649] - 1
datao3$CPILFESL = datao2$CPILFESL[14:650]/datao2$CPILFESL[13:649] - 1
plot(datao3$CPI, type = "l")
```



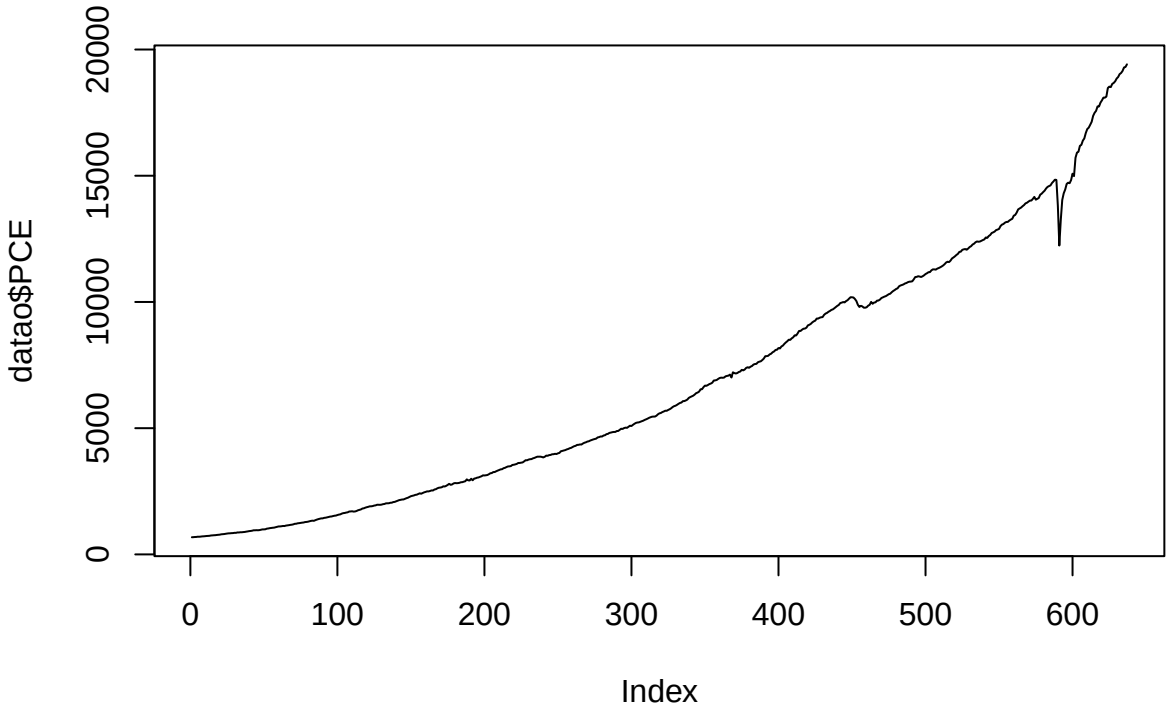

```
plot(datao3$CPILFESL, type = "l")
```



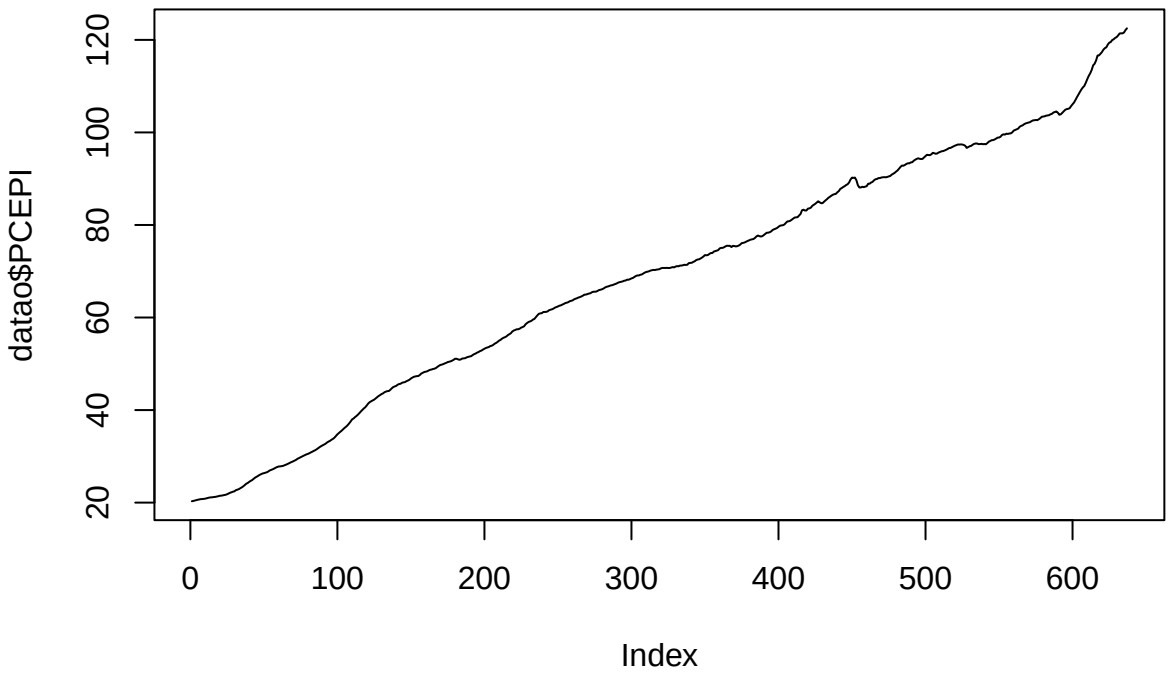

```
cor(datao3$CPI, datao3$CPILFESL)
```

```
## [1] 0.6312404
```

```
#consumption
plot(datao$PCE, type = "l")
```




```
plot(datao$PCEPI, type = "l")
```



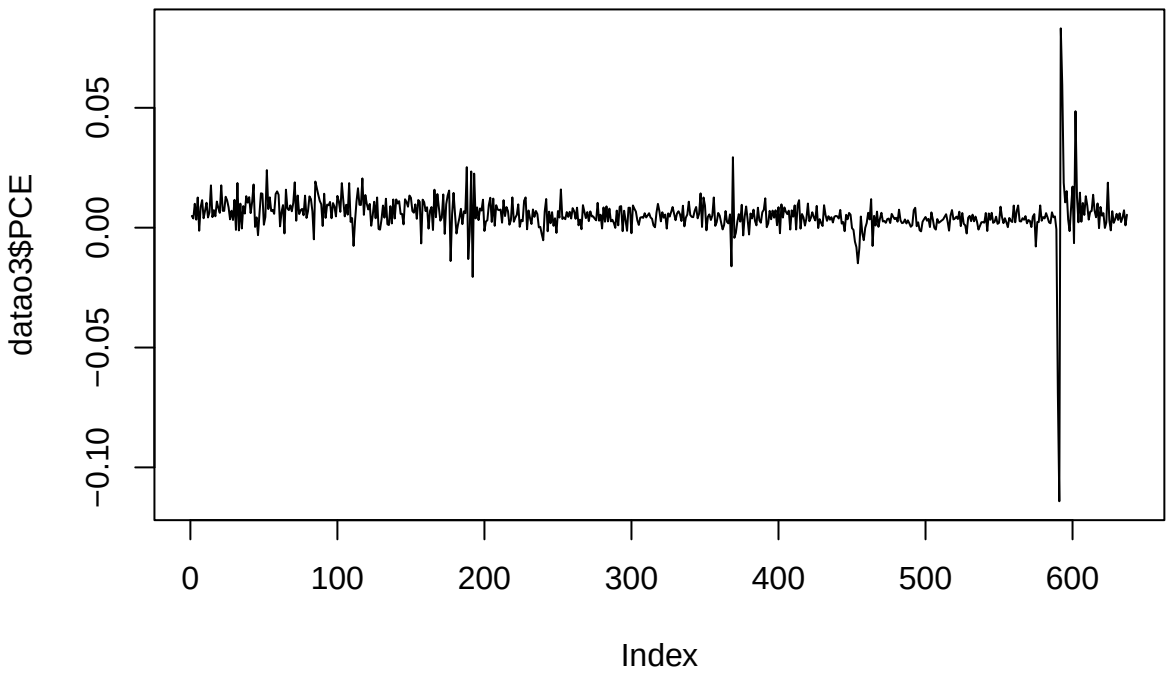

```
cor(datao$PCE, datao$PCEPI)
```

```
## [1] 0.9648003
```

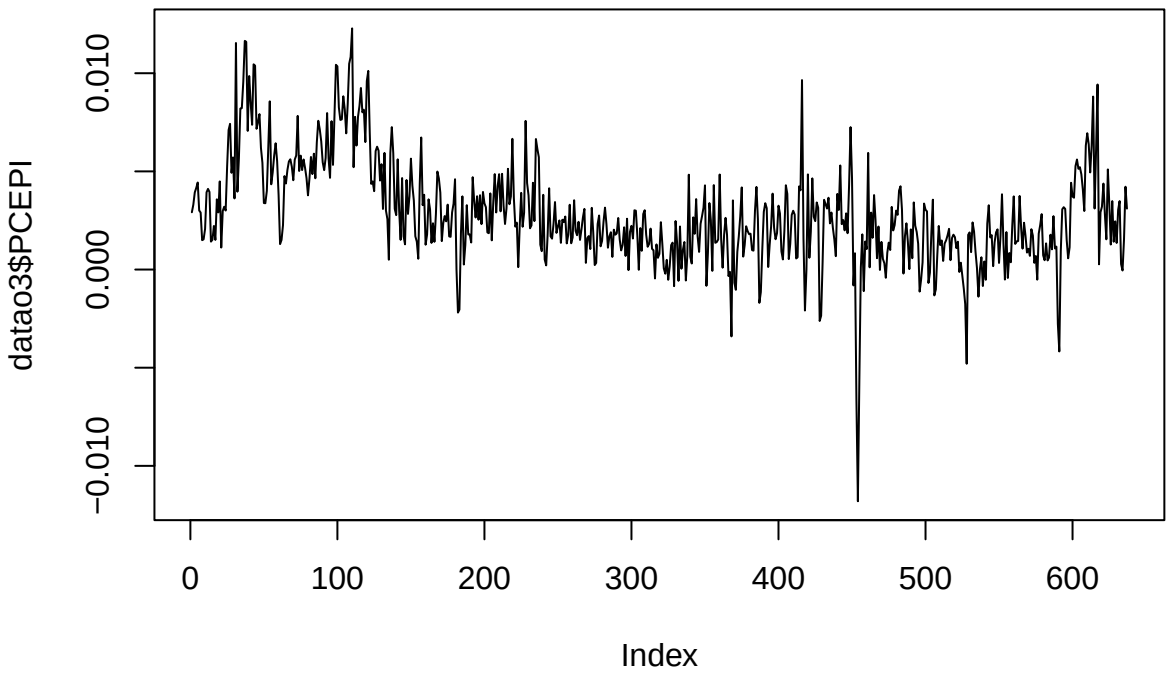
```
cor(datao$PCE, datao$CPI)
```

```
## [1] 0.9766213
```

```
datao3$PCE = datao2$PCE[14:650]/datao2$PCE[13:649] -1
datao3$PCEPI = datao2$PCEPI[14:650]/datao2$PCEPI[13:649] -1
plot(datao3$PCE, type = "l")
```



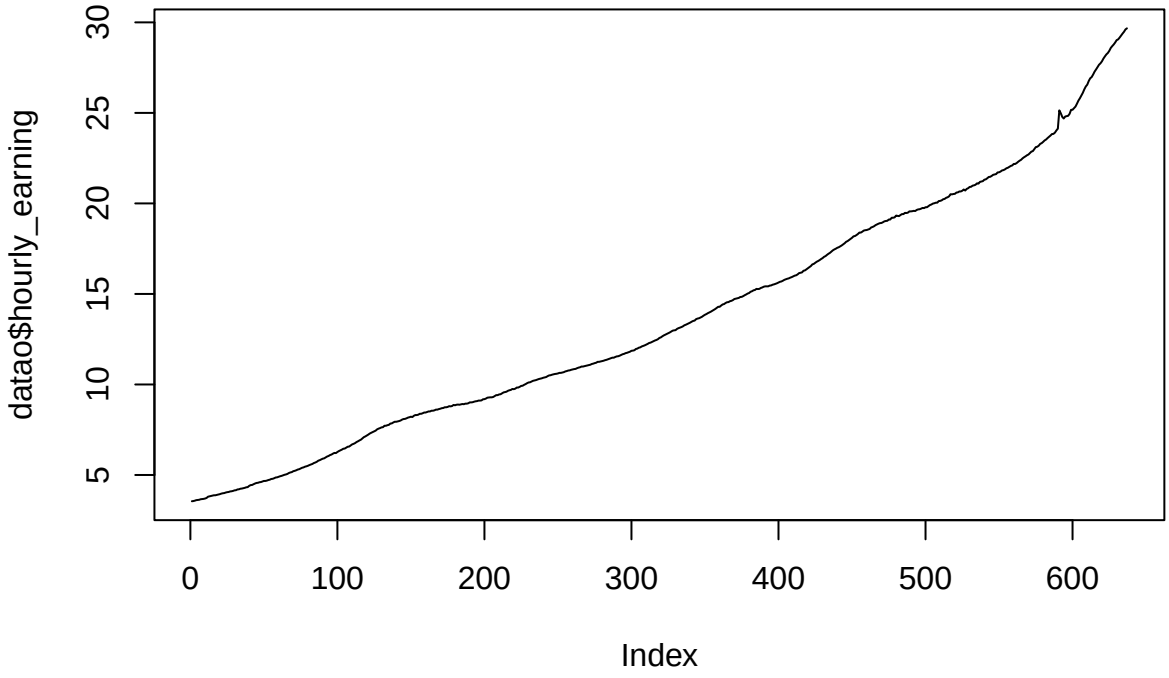

```
plot(datao3$PCEPI, type = "l")
```



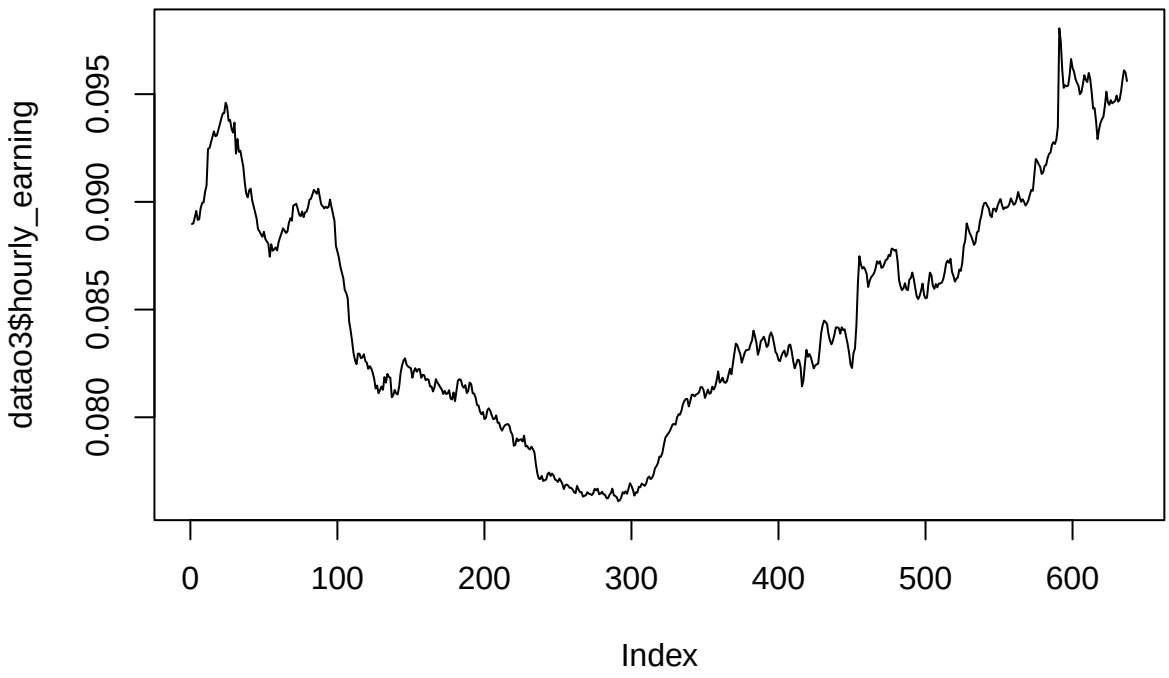

```
cor(datao3$PCE, datao$CPI)
```

```
## [1] -0.172639
```

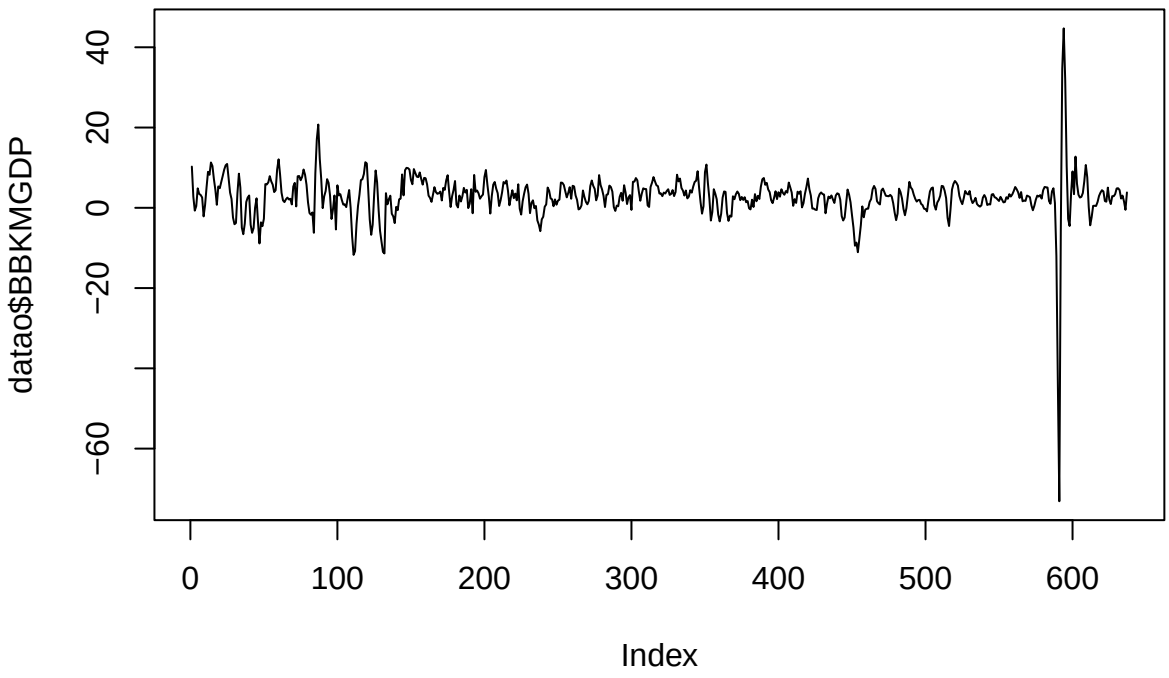
```
#income
plot(datao$hourly_earning, type = "l")
```



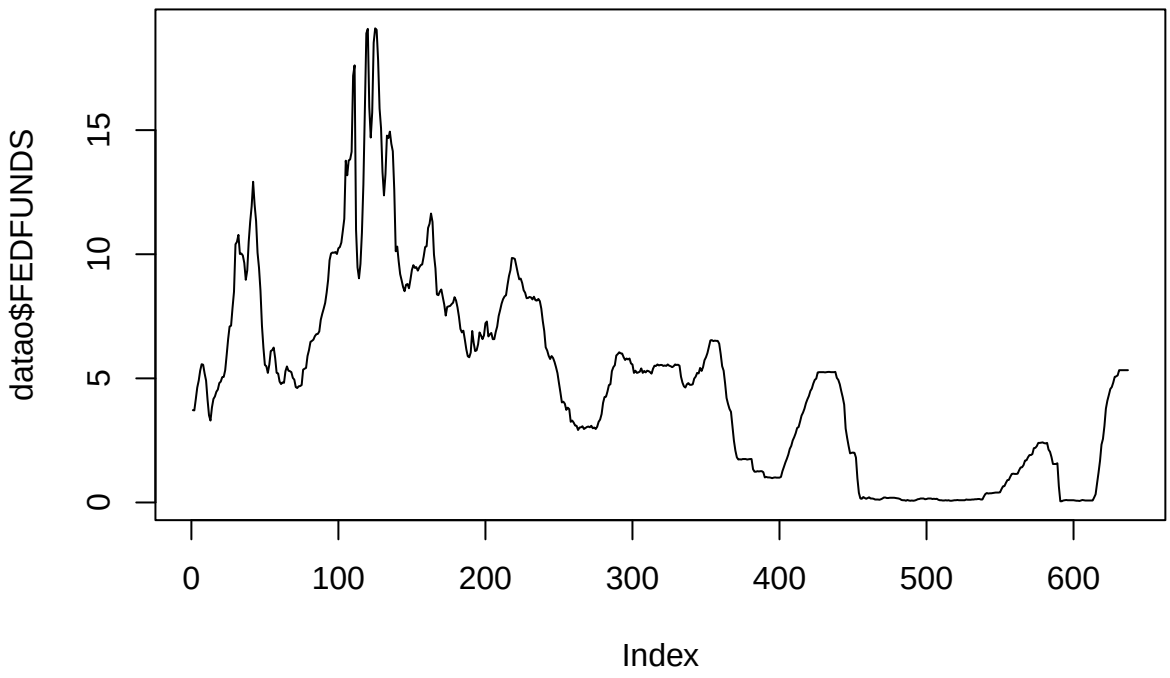

```
datao3$hourly_earning = datao$hourly_earning/datao$CPI
plot(datao3$hourly_earning, type = "l")
```



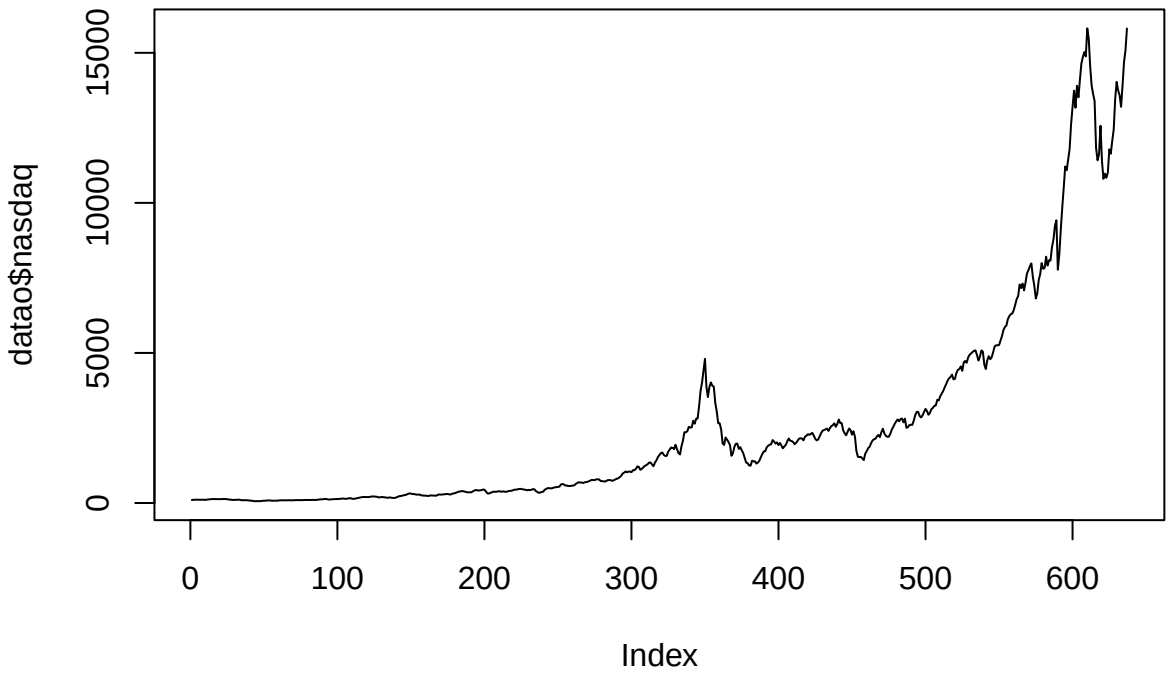

```
#GDP
plot(datao$BBKMGDP, type = "l")
```



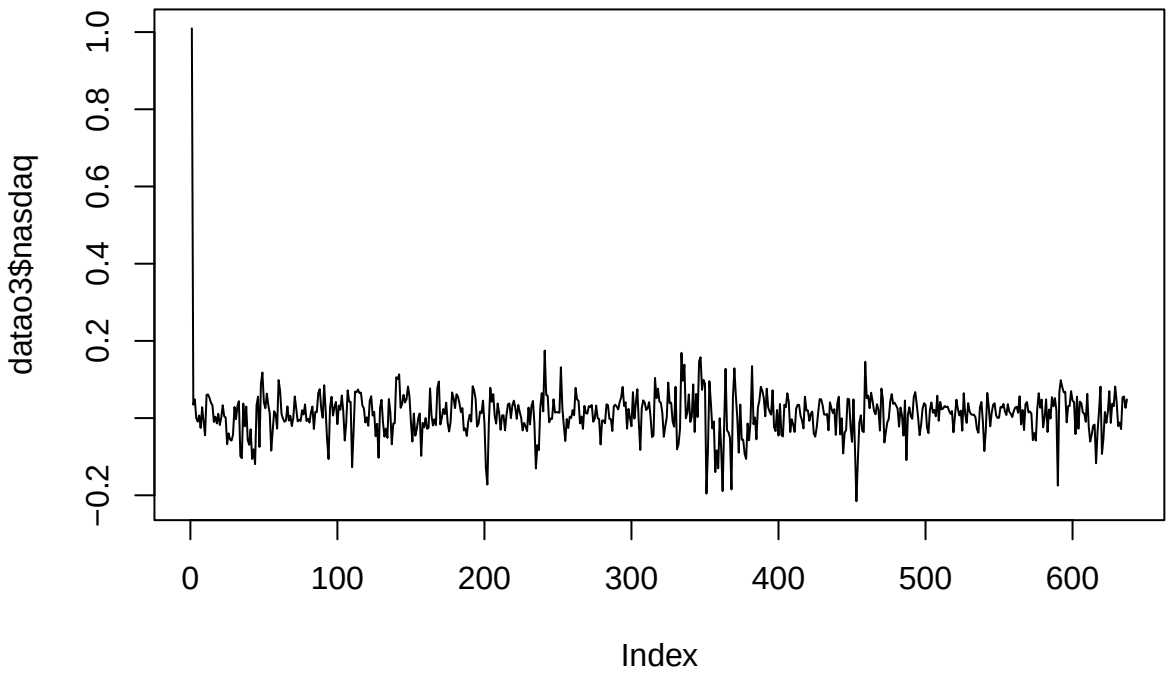

```
#market
plot(datao$FEDFUNDS, type = "l")
```



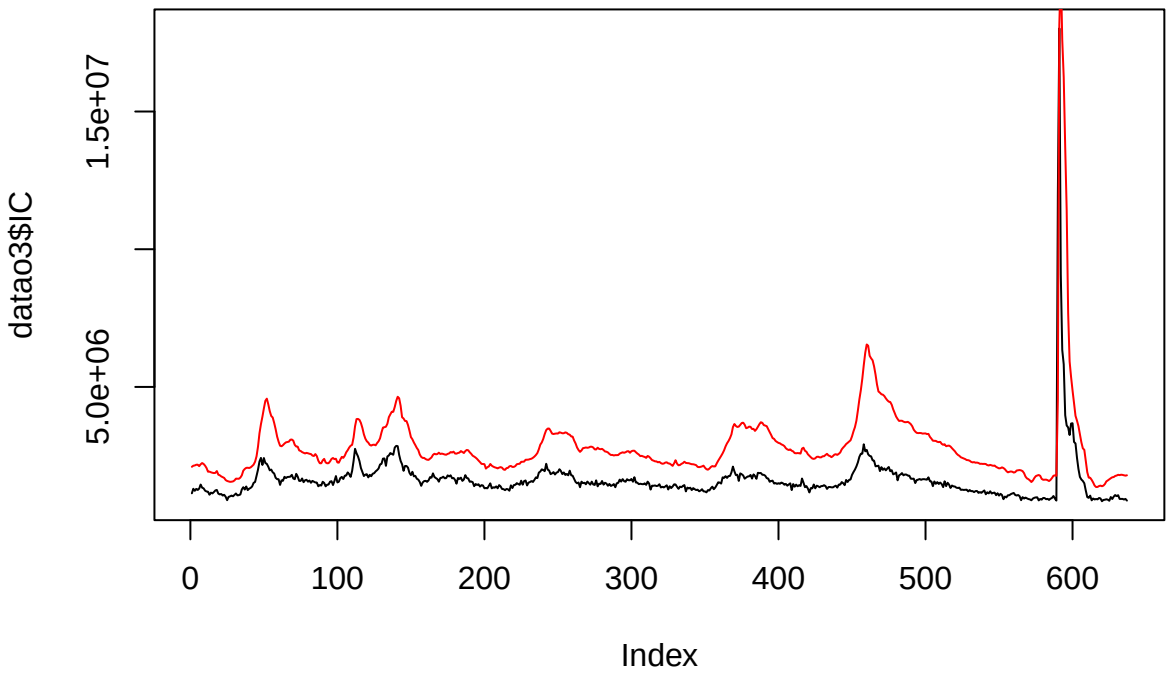

```
datao3$idiff = datao2$FEDFUNDS[14:650]-datao2$FEDFUNDS[13:649]
plot(datao$nasdaq, type = "l")
```




```
datao3$nasdaq = datao2$nasdaq[14:650]/datao2$nasdaq[13:649] -1
datao3$nasdaq[1] = datao2$nasdaq[14]/100
plot(datao3$nasdaq, type = "l")
```




```
#check the correlation
correlx = cor(datao3[,2:ncol(datao3)])
correlx2 = cor(datao3[289:637,2:ncol(datao3)])

#descriptive analysis
plot(datao3$IC, type = "l")
lines(datao3$CC, col = "red")
```




```
cor(datao3$IC, datao3$CC)
```

```
## [1] 0.7896165
```

```
mean(datao3$IC)
```

```
## [1] 1633120
```

```
mean(datao3$CC)
```

```
## [1] 2871793
```

```
sqrt(var(datao3$IC))
```

```
## [1] 965775.9
```

```
sqrt(var(datao3$CC))
```

```
## [1] 1583808
```

```r
datao4=datao3
datao4[,!(names(datao4) %in% c("observation_date"))] = scale(datao3[,!(names(datao3) %in% c("observatio
summary(lm(IC~unemployment_level+discouraged+
     CPI+hourly_earning+UMCSENT_interp+
     BBKMGDP+FEDFUNDS+nasdaq+uldiff+disdiff+idiff,data = datao4[288:637,]))
```

```
##
## Call:
## lm(formula = IC ~ unemployment_level + discouraged + CPI + hourly_earning +
##     UMCSENT_interp + BBKMGDP + FEDFUNDS + nasdaq + uldiff + disdiff +
##     idiff, data = datao4[288:637, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4306 -0.3127 -0.0143  0.2480  7.2228
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.14921    0.06937  -2.151 0.032194 *
## unemployment_level  1.40301    0.06304  22.257  < 2e-16 ***
## discouraged        -0.95516    0.07332 -13.027  < 2e-16 ***
## CPI                 0.02742    0.03752   0.731 0.465477
## hourly_earning      0.43012    0.05006   8.593 3.22e-16 ***
## UMCSENT_interp      0.27911    0.04302   6.487 3.12e-10 ***
## BBKMGDP            -0.29745    0.04830  -6.159 2.09e-09 ***
## FEDFUNDS            0.41402    0.10582   3.912 0.000111 ***
## nasdaq             -0.06197    0.04351  -1.424 0.155281
## uldiff              0.13135    0.04162   3.156 0.001743 **
## disdiff             0.13959    0.03558   3.923 0.000106 ***
## idiff              -0.45704    0.11300  -4.045 6.50e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6351 on 337 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7722, Adjusted R-squared:  0.7648
## F-statistic: 103.9 on 11 and 337 DF,  p-value: < 2.2e-16
```

```r
summary(lm(CC~unemployment_level+discouraged+
             CPI+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+disdiff+idiff,data = datao4[288:637,]))
```

```
##
## Call:
## lm(formula = CC ~ unemployment_level + discouraged + CPI + hourly_earning +
##     UMCSENT_interp + BBKMGDP + FEDFUNDS + nasdaq + uldiff + disdiff +
##     idiff, data = datao4[288:637, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5674 -0.2450  0.0045  0.1982  2.3974
##
```

```
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.16485    0.04991  -3.303 0.001059 **
## unemployment_level  1.89264    0.04535  41.732  < 2e-16 ***
## discouraged        -1.30340    0.05275 -24.708  < 2e-16 ***
## CPI                -0.01527    0.02700  -0.566 0.571936
## hourly_earning      0.40625    0.03601  11.280  < 2e-16 ***
## UMCSENT_interp      0.21309    0.03095   6.884 2.85e-11 ***
## BBKMGDP             0.08093    0.03475   2.329 0.020451 *
## FEDFUNDS            0.33853    0.07614   4.446 1.19e-05 ***
## nasdaq             -0.03063    0.03131  -0.978 0.328598
## uldiff             -0.10104    0.02994  -3.374 0.000826 ***
## disdiff             0.17687    0.02560   6.909 2.45e-11 ***
## idiff              -0.19773    0.08130  -2.432 0.015529 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.457 on 337 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.8794, Adjusted R-squared:  0.8754
## F-statistic: 223.3 on 11 and 337 DF,  p-value: < 2.2e-16
```

```
summary(lm(IC~unemployment_level+
             CPI+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = IC ~ unemployment_level + CPI + hourly_earning +
##     UMCSENT_interp + BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff +
##     idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5174 -0.3033  0.0028  0.2228  8.5924
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3.848e-16  2.579e-02   0.000   1.0000
## unemployment_level  6.079e-01  3.303e-02  18.406  < 2e-16 ***
## CPI                 2.362e-02  3.055e-02   0.773   0.4397
## hourly_earning      2.099e-01  3.172e-02   6.618 7.83e-11 ***
## UMCSENT_interp      1.488e-01  3.363e-02   4.426 1.13e-05 ***
## BBKMGDP            -1.704e-01  3.640e-02  -4.681 3.50e-06 ***
## FEDFUNDS            2.581e-01  3.469e-02   7.440 3.33e-13 ***
## nasdaq              1.126e-02  2.642e-02   0.426   0.6700
## uldiff              1.954e-01  1.304e-01   1.499   0.1344
## nldiff              1.693e-01  1.302e-01   1.300   0.1941
## idiff              -7.566e-02  2.679e-02  -2.824   0.0049 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.651 on 626 degrees of freedom
```

```
## Multiple R-squared:  0.5828, Adjusted R-squared:  0.5762
## F-statistic: 87.46 on 10 and 626 DF,  p-value: < 2.2e-16
```

```
summary(lm(IC~num_losers+
             CPI+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = IC ~ num_losers + CPI + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5719 -0.2691  0.0222  0.2180  8.3838
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.494e-16  2.261e-02   0.000  1.00000
## num_losers      7.105e-01  2.831e-02  25.101  < 2e-16 ***
## CPI             3.567e-02  2.675e-02   1.333  0.18293
## hourly_earning  1.732e-01  2.694e-02   6.429 2.55e-10 ***
## UMCSENT_interp  1.826e-01  2.898e-02   6.300 5.63e-10 ***
## BBKMGDP        -1.944e-01  3.194e-02  -6.087 2.01e-09 ***
## FEDFUNDS        2.745e-01  2.977e-02   9.220  < 2e-16 ***
## nasdaq         -2.427e-03  2.317e-02  -0.105  0.91662
## uldiff          1.787e-01  1.142e-01   1.565  0.11810
## nldiff          1.207e-01  1.141e-01   1.058  0.29041
## idiff          -6.797e-02  2.348e-02  -2.894  0.00393 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5706 on 626 degrees of freedom
## Multiple R-squared:  0.6796, Adjusted R-squared:  0.6745
## F-statistic: 132.8 on 10 and 626 DF,  p-value: < 2.2e-16
```

```
summary(lm(IC~num_losers+
             PCEPI+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = IC ~ num_losers + PCEPI + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5900 -0.2718  0.0194  0.2191  8.4029
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    -1.382e-16  2.261e-02   0.000  1.00000
## num_losers      7.115e-01  2.840e-02  25.050  < 2e-16 ***
## PCEPI           4.109e-02  3.034e-02   1.354  0.17622
## hourly_earning  1.692e-01  2.749e-02   6.155 1.34e-09 ***
## UMCSENT_interp  1.863e-01  2.967e-02   6.278 6.41e-10 ***
## BBKMGDP        -1.958e-01  3.205e-02  -6.110 1.75e-09 ***
## FEDFUNDS        2.678e-01  3.175e-02   8.435 2.28e-16 ***
## nasdaq         -3.905e-03  2.319e-02  -0.168  0.86631
## uldiff          1.780e-01  1.142e-01   1.558  0.11962
## nldiff          1.213e-01  1.141e-01   1.064  0.28792
## idiff          -6.750e-02  2.347e-02  -2.876  0.00417 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5705 on 626 degrees of freedom
## Multiple R-squared:  0.6796, Adjusted R-squared:  0.6745
## F-statistic: 132.8 on 10 and 626 DF,  p-value: < 2.2e-16
```

```
summary(lm(CC~unemployment_level+
             CPILFESL+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = CC ~ unemployment_level + CPILFESL + hourly_earning +
##     UMCSENT_interp + BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff +
##     idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3301 -0.3063  0.0290  0.2351  6.9289
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -2.734e-16  2.621e-02   0.000   1.0000
## unemployment_level  7.979e-01  3.366e-02  23.707  < 2e-16 ***
## CPILFESL            1.733e-03  4.049e-02   0.043   0.9659
## hourly_earning      2.114e-01  3.260e-02   6.485 1.81e-10 ***
## UMCSENT_interp      7.130e-02  3.559e-02   2.004   0.0456 *
## BBKMGDP             8.672e-02  3.704e-02   2.341   0.0195 *
## FEDFUNDS            1.782e-01  4.167e-02   4.276 2.20e-05 ***
## nasdaq              3.785e-02  2.685e-02   1.410   0.1592
## uldiff              2.893e-01  1.326e-01   2.181   0.0295 *
## nldiff             -1.857e-01  1.324e-01  -1.403   0.1613
## idiff              -5.354e-02  2.726e-02  -1.964   0.0499 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6615 on 626 degrees of freedom
## Multiple R-squared:  0.5694, Adjusted R-squared:  0.5625
## F-statistic: 82.77 on 10 and 626 DF,  p-value: < 2.2e-16
```

```r
summary(lm(CC~num_losers+
             CPI+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = CC ~ num_losers + CPI + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3823 -0.2441  0.0495  0.2224  4.3811
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.177e-17  2.095e-02   0.000   1.0000
## num_losers      9.254e-01  2.624e-02  35.271  < 2e-16 ***
## CPI             3.648e-02  2.480e-02   1.471   0.1417
## hourly_earning  1.581e-01  2.497e-02   6.331 4.65e-10 ***
## UMCSENT_interp  1.153e-01  2.686e-02   4.295 2.03e-05 ***
## BBKMGDP         5.433e-02  2.960e-02   1.835   0.0670 .
## FEDFUNDS        1.879e-01  2.759e-02   6.812 2.27e-11 ***
## nasdaq          2.062e-02  2.148e-02   0.960   0.3373
## uldiff          2.682e-01  1.058e-01   2.534   0.0115 *
## nldiff         -2.472e-01  1.057e-01  -2.338   0.0197 *
## idiff          -4.408e-02  2.176e-02  -2.025   0.0432 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5288 on 626 degrees of freedom
## Multiple R-squared:  0.7247, Adjusted R-squared:  0.7203
## F-statistic: 164.8 on 10 and 626 DF,  p-value: < 2.2e-16
```

```r
summary(lm(CC~num_losers+
             PCE+hourly_earning+UMCSENT_interp+
             BBKMGDP+FEDFUNDS+nasdaq+uldiff+
             nldiff+idiff,data = datao4))
```

```
##
## Call:
## lm(formula = CC ~ num_losers + PCE + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4211 -0.2459  0.0556  0.2241  4.2244
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.373e-17  2.076e-02   0.000 1.000000
## num_losers      9.087e-01  2.593e-02  35.039  < 2e-16 ***
## PCE             1.136e-01  3.029e-02   3.749 0.000194 ***
```

```
## hourly_earning  1.501e-01  2.473e-02   6.071 2.20e-09 ***
## UMCSENT_interp  1.070e-01  2.597e-02   4.121 4.28e-05 ***
## BBKMGDP         2.444e-02  3.057e-02   0.800 0.424292
## FEDFUNDS        1.768e-01  2.649e-02   6.676 5.42e-11 ***
## nasdaq          1.050e-02  2.144e-02   0.490 0.624561
## uldiff          2.435e-01  1.051e-01   2.317 0.020818 *
## nldiff         -1.700e-01  1.069e-01  -1.590 0.112310
## idiff          -4.363e-02  2.155e-02  -2.025 0.043311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5239 on 626 degrees of freedom
## Multiple R-squared:  0.7299, Adjusted R-squared:  0.7255
## F-statistic: 169.1 on 10 and 626 DF,  p-value: < 2.2e-16
```

```
lmfullic = lm(IC~num_losers+
              CPI+hourly_earning+UMCSENT_interp+
              BBKMGDP+FEDFUNDS+nasdaq+uldiff+
              nldiff+idiff,data = datao4)
lmfullcc = lm(CC~num_losers+
              PCE+hourly_earning+UMCSENT_interp+
              BBKMGDP+FEDFUNDS+nasdaq+uldiff+
              nldiff+idiff,data = datao4)
summary(lmfullic)
```

```
##
## Call:
## lm(formula = IC ~ num_losers + CPI + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5719 -0.2691  0.0222  0.2180  8.3838
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.494e-16  2.261e-02   0.000  1.00000
## num_losers      7.105e-01  2.831e-02  25.101  < 2e-16 ***
## CPI             3.567e-02  2.675e-02   1.333  0.18293
## hourly_earning  1.732e-01  2.694e-02   6.429 2.55e-10 ***
## UMCSENT_interp  1.826e-01  2.898e-02   6.300 5.63e-10 ***
## BBKMGDP        -1.944e-01  3.194e-02  -6.087 2.01e-09 ***
## FEDFUNDS        2.745e-01  2.977e-02   9.220  < 2e-16 ***
## nasdaq         -2.427e-03  2.317e-02  -0.105  0.91662
## uldiff          1.787e-01  1.142e-01   1.565  0.11810
## nldiff          1.207e-01  1.141e-01   1.058  0.29041
## idiff          -6.797e-02  2.348e-02  -2.894  0.00393 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5706 on 626 degrees of freedom
## Multiple R-squared:  0.6796, Adjusted R-squared:  0.6745
## F-statistic: 132.8 on 10 and 626 DF,  p-value: < 2.2e-16
```

```r
summary(lmfullcc)
```

```
##
## Call:
## lm(formula = CC ~ num_losers + PCE + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + nasdaq + uldiff + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4211 -0.2459  0.0556  0.2241  4.2244
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.373e-17  2.076e-02   0.000 1.000000
## num_losers      9.087e-01  2.593e-02  35.039  < 2e-16 ***
## PCE             1.136e-01  3.029e-02   3.749 0.000194 ***
## hourly_earning  1.501e-01  2.473e-02   6.071 2.20e-09 ***
## UMCSENT_interp  1.070e-01  2.597e-02   4.121 4.28e-05 ***
## BBKMGDP         2.444e-02  3.057e-02   0.800 0.424292
## FEDFUNDS        1.768e-01  2.649e-02   6.676 5.42e-11 ***
## nasdaq          1.050e-02  2.144e-02   0.490 0.624561
## uldiff          2.435e-01  1.051e-01   2.317 0.020818 *
## nldiff         -1.700e-01  1.069e-01  -1.590 0.112310
## idiff          -4.363e-02  2.155e-02  -2.025 0.043311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5239 on 626 degrees of freedom
## Multiple R-squared:  0.7299, Adjusted R-squared:  0.7255
## F-statistic: 169.1 on 10 and 626 DF,  p-value: < 2.2e-16
```

```r
fmic = lm(IC~num_losers+
            hourly_earning+UMCSENT_interp+
            BBKMGDP+FEDFUNDS+idiff,data = datao4)
summary(fmic)
```

```
##
## Call:
## lm(formula = IC ~ num_losers + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5159 -0.3344  0.0200  0.2674  7.3862
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.454e-16  2.415e-02   0.000   1.0000
## num_losers      7.688e-01  2.900e-02  26.507  < 2e-16 ***
## hourly_earning  2.009e-01  2.835e-02   7.086 3.72e-12 ***
## UMCSENT_interp  2.206e-01  2.966e-02   7.439 3.33e-13 ***
## BBKMGDP        -3.889e-01  2.504e-02 -15.529  < 2e-16 ***
```

```
## FEDFUNDS        3.278e-01  2.944e-02  11.134  < 2e-16 ***
## idiff          -6.194e-02  2.493e-02  -2.485   0.0132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6094 on 630 degrees of freedom
## Multiple R-squared:  0.6321, Adjusted R-squared:  0.6286
## F-statistic: 180.4 on 6 and 630 DF,  p-value: < 2.2e-16
```

```
anova(fmic, lmfullic)
```

```
## Analysis of Variance Table
##
## Model 1: IC ~ num_losers + hourly_earning + UMCSENT_interp + BBKMGDP +
##     FEDFUNDS + idiff
## Model 2: IC ~ num_losers + CPI + hourly_earning + UMCSENT_interp + BBKMGDP +
##     FEDFUNDS + nasdaq + uldiff + nldiff + idiff
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    630 234.00
## 2    626 203.79  4    30.211 23.201 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fmcc = lm(CC~num_losers+PCE+
            hourly_earning+UMCSENT_interp+
            FEDFUNDS+nldiff+idiff,data = datao4)
summary(fmcc)
```

```
##
## Call:
## lm(formula = CC ~ num_losers + PCE + hourly_earning + UMCSENT_interp +
##     FEDFUNDS + nldiff + idiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4334 -0.2544  0.0644  0.2322  4.3136
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.463e-17  2.081e-02   0.000   1.0000
## num_losers      9.146e-01  2.582e-02  35.419  < 2e-16 ***
## PCE             1.265e-01  2.855e-02   4.429 1.12e-05 ***
## hourly_earning  1.500e-01  2.478e-02   6.054 2.43e-09 ***
## UMCSENT_interp  1.086e-01  2.534e-02   4.287 2.10e-05 ***
## FEDFUNDS        1.786e-01  2.651e-02   6.740 3.60e-11 ***
## nldiff          5.966e-02  2.879e-02   2.072   0.0387 *
## idiff          -4.467e-02  2.130e-02  -2.098   0.0363 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5251 on 629 degrees of freedom
## Multiple R-squared:  0.7273, Adjusted R-squared:  0.7242
## F-statistic: 239.6 on 7 and 629 DF,  p-value: < 2.2e-16
```

```r
anova(fmcc, lmfullcc)
```

```
## Analysis of Variance Table
##
## Model 1: CC ~ num_losers + PCE + hourly_earning + UMCSENT_interp + FEDFUNDS +
##     nldiff + idiff
## Model 2: CC ~ num_losers + PCE + hourly_earning + UMCSENT_interp + BBKMGDP +
##     FEDFUNDS + nasdaq + uldiff + nldiff + idiff
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    629 173.47
## 2    626 171.81  3    1.6544 2.0093 0.1114
```

```r
fmic2 = lm(IC~num_losers+
            hourly_earning+UMCSENT_interp+
            BBKMGDP+FEDFUNDS+idiff+nldiff,data = datao4)
summary(fmic2)
```

```
##
## Call:
## lm(formula = IC ~ num_losers + hourly_earning + UMCSENT_interp +
##     BBKMGDP + FEDFUNDS + idiff + nldiff, data = datao4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5880 -0.2847  0.0213  0.2123  8.3644
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.621e-16  2.263e-02   0.000  1.00000
## num_losers      7.085e-01  2.793e-02  25.367  < 2e-16 ***
## hourly_earning  1.783e-01  2.668e-02   6.682 5.20e-11 ***
## UMCSENT_interp  1.721e-01  2.827e-02   6.087 2.00e-09 ***
## BBKMGDP        -1.947e-01  3.127e-02  -6.227 8.68e-10 ***
## FEDFUNDS        2.910e-01  2.787e-02  10.441  < 2e-16 ***
## idiff          -6.806e-02  2.337e-02  -2.912  0.00372 **
## nldiff          2.910e-01  3.097e-02   9.396  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5712 on 629 degrees of freedom
## Multiple R-squared:  0.6774, Adjusted R-squared:  0.6738
## F-statistic: 188.7 on 7 and 629 DF,  p-value: < 2.2e-16
```

```r
anova(fmic2, lmfullic)
```

```
## Analysis of Variance Table
##
## Model 1: IC ~ num_losers + hourly_earning + UMCSENT_interp + BBKMGDP +
##     FEDFUNDS + idiff + nldiff
## Model 2: IC ~ num_losers + CPI + hourly_earning + UMCSENT_interp + BBKMGDP +
##     FEDFUNDS + nasdaq + uldiff + nldiff + idiff
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
```

```
## 1    629 205.20
## 2    626 203.79  3    1.4092 1.443 0.2292
```

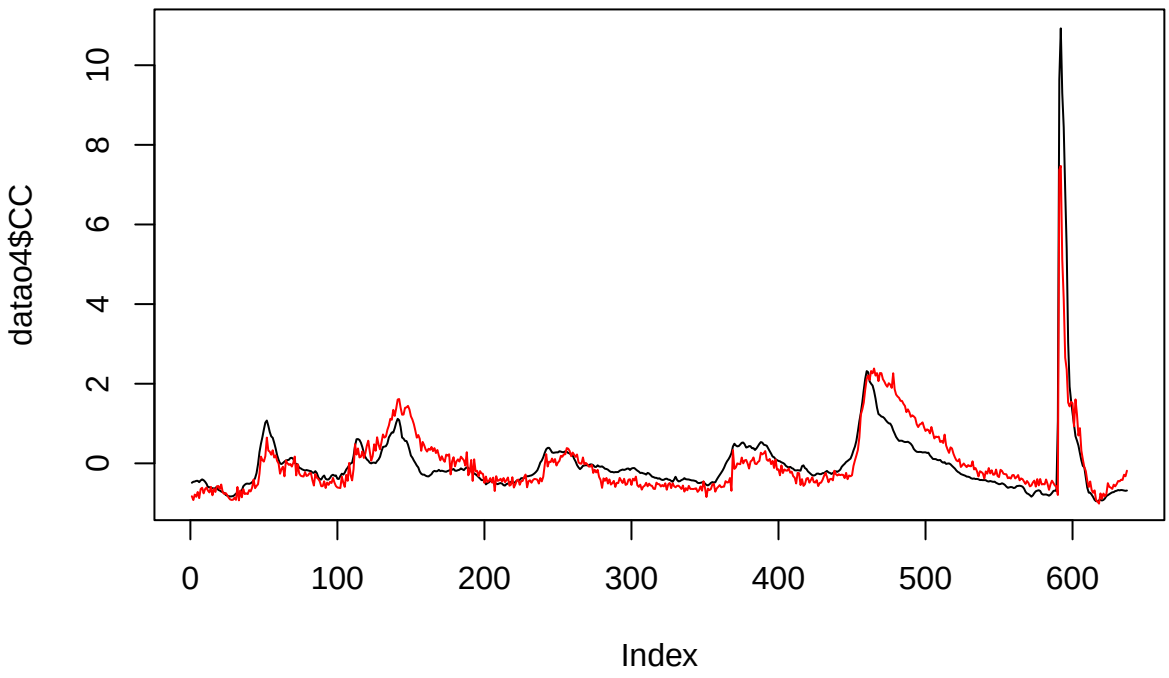
```
fmic = fmic2

y = datao4$IC
x = model.matrix(fmic)[,2:ncol(model.matrix(fmic))]
set.seed(2)
train_control <- trainControl(method = "cv", number = 5)
lmcvic <- train(
  x = x,
  y = y,
  method = "lm",
  trControl = train_control
)
(lmcvic$results$RMSE)^2
```

```
## [1] 0.4874865
```

```
y = datao4$CC
x = model.matrix(fmcc)[,2:ncol(model.matrix(fmcc))]
lmcvcc <- train(
  x = x,
  y = y,
  method = "lm",
  trControl = train_control
)
(lmcvcc$results$RMSE)^2
```

```
## [1] 0.7503188
```

```
plot(datao4$CC, type = "l")
lines(predict(lmcvcc), col = "red")
```




```
#regularization

library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.3.3
```

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

```
## Loaded glmnet 4.1-8
```

```
y = datao4$IC
x = datao4[, (names(datao4) %in% c("num_losers", "CPI", "PCE",
                                   "hourly_earning", "UMCSENT_interp",
                                   "BBKMGDP",
                                   "FEDFUNDS","nasdaq",
                                   "nldiff",
                                   "idiff"))]
grid <- expand.grid(
  alpha = seq(0, 1, length = 11),
  lambda = 10^seq(-5, 10, length = 100)
)

set.seed(1)
model = train(
```

```
  x = x,
  y = y,
  method = "glmnet",
  trControl = train_control,
  tuneGrid = grid
)
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo,
## : There were missing values in resampled performance measures.
```

```
best_alpha = model$bestTune$alpha
best_alpha
```

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

```
best_lambda = model$bestTune$lambda
best_lambda
```

```
## [1] 0.0869749
```

```
model$results[model$results$alpha == best_alpha & model$results$lambda == best_lambda, ]$RMSE^2
```

```
## [1] 0.5274346
```

```
model <- glmnet(x, y, alpha = best_alpha, lambda = best_lambda)
coef(model)
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                              s0
## (Intercept)    -6.580066e-17
## num_losers      5.043076e-01
## CPI             .
## PCE             .
## hourly_earning  .
## UMCSENT_interp  .
## BBKMGDP        -1.093611e-01
## FEDFUNDS        5.814048e-02
## nasdaq          .
## nldiff          2.986317e-01
## idiff           .
```

```
y = datao4$CC
set.seed(1)
model2 = train(
  x = x,
  y = y,
  method = "glmnet",
  trControl = train_control,
  tuneGrid = grid
)
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo,
## : There were missing values in resampled performance measures.
```

```
best_alpha2 = model2$bestTune$alpha
best_alpha2
```

```
## [1] 0.9
```

```
best_lambda2 = model2$bestTune$lambda
best_lambda2
```

```
## [1] 0.1747528
```

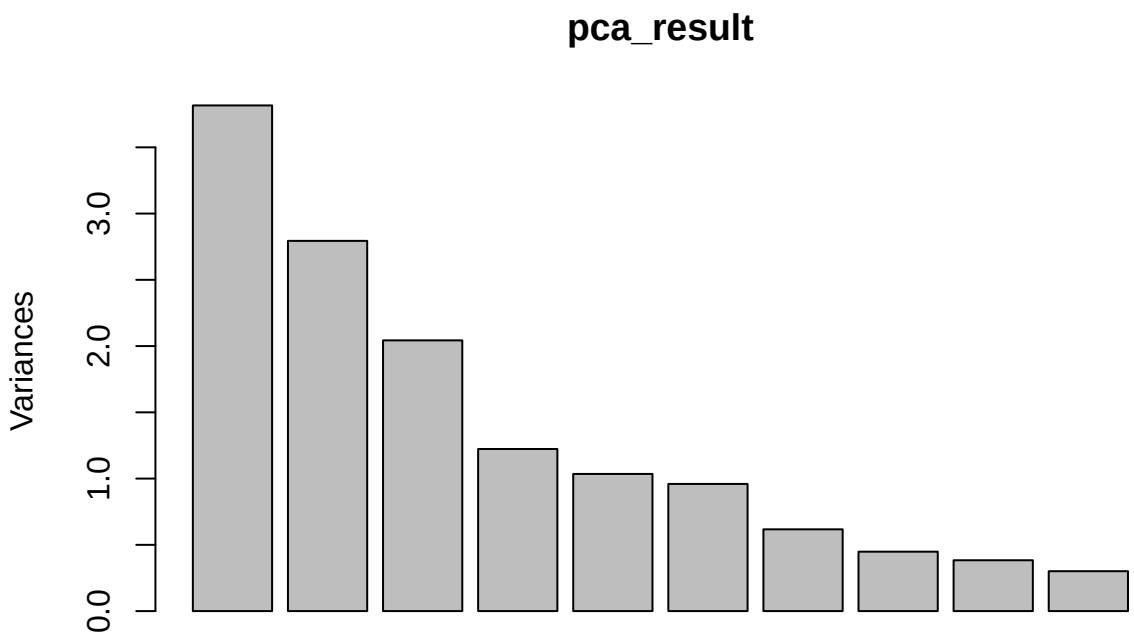
```
model2$results[model2$results$alpha == best_alpha2 & model2$results$lambda == best_lambda2, ]$RMSE^2
```

```
## [1] 0.3288756
```

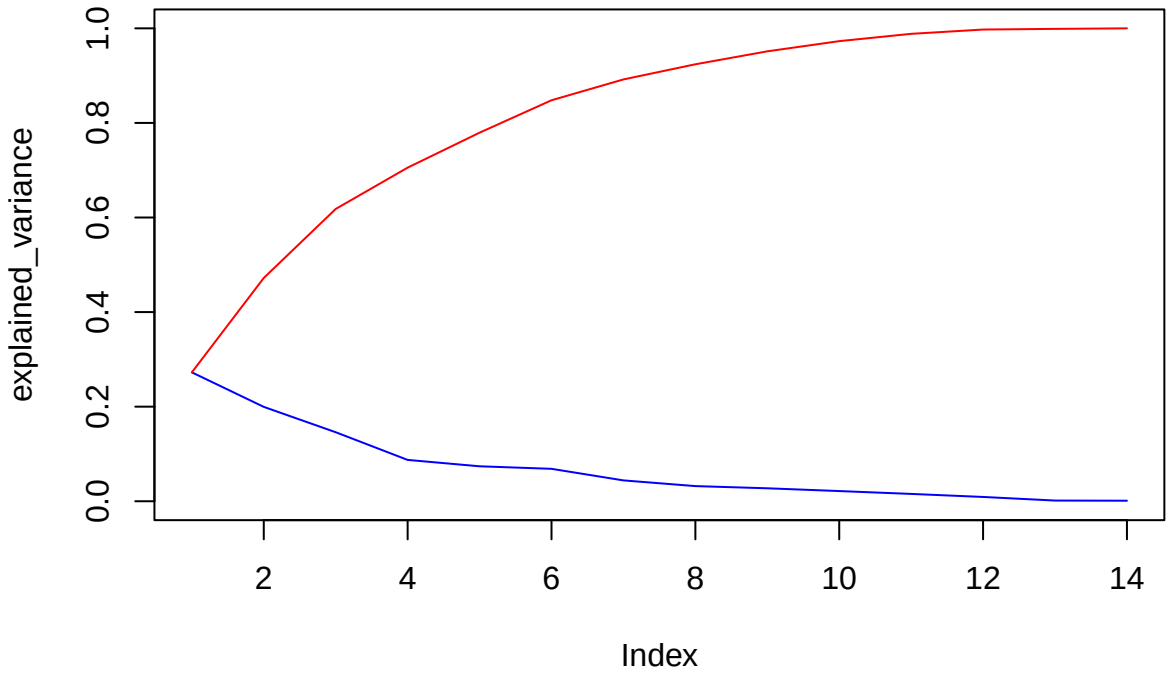
```
model2 <- glmnet(x, y, alpha = best_alpha2, lambda = best_lambda2)
coef(model2)
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                              s0
## (Intercept)    -4.618228e-19
## num_losers      6.577078e-01
## CPI             .
## PCE             .
## hourly_earning  .
## UMCSENT_interp  .
## BBKMGDP         .
## FEDFUNDS        .
## nasdaq          .
## nldiff          .
## idiff           .
```

```
#PCA
xpca = datao4[, (!names(datao4) %in% c("observation_date", "CC",
                                       "IC","disdiff","discouraged"))]
pca_result <- prcomp(xpca)
screeplot(pca_result)
```




```r
explained_variance <- pca_result$sdev^2 / sum(pca_result$sdev^2)
cumulative_variance <- cumsum(explained_variance)
plot(explained_variance,
     type = "l",
     col = "blue",ylim=c(0,1))
lines(cumulative_variance, col = "red", type = "l")
```




```
eigenvalues <- pca_result$sdev^2
sum(eigenvalues > 1)
```

```
## [1] 5
```

```
cumulative_variance[5]
```

```
## [1] 0.7793811
```

```
pca_result$rotation[,1:5]
```

```
##                             PC1         PC2         PC3          PC4
## unemployment_level  0.282563681 -0.11044356  0.50515205 -0.245797567
## num_losers          0.287458372 -0.08250575  0.51706791 -0.201934807
## CPI                -0.338546480  0.28014434  0.18190970 -0.008760172
## CPILFESL           -0.343663302  0.31406765  0.12707088 -0.076186367
## PCEPI              -0.360098351  0.29883243  0.20800065  0.004452191
## PCE                -0.330000021 -0.26996777  0.18716205 -0.079704346
## hourly_earning     -0.019256228  0.04380085  0.23908055  0.772954307
## UMCSENT_interp      0.022035337 -0.22849526 -0.50505723 -0.112019296
## BBKMGDP            -0.273801995 -0.37766807  0.06451377 -0.054614198
## FEDFUNDS           -0.277708210  0.30802936 -0.10604391 -0.403082214
## nasdaq              0.008310115 -0.09014842  0.08593537 -0.180139199
## uldiff              0.315246437  0.41386018 -0.07267987 -0.009084744
```

```
## nldiff              0.317622016  0.41179249 -0.08850579  0.001192076
## idiff              -0.113831328 -0.03268578 -0.04218958  0.280273851
##                             PC5
## unemployment_level  0.16343010
## num_losers          0.12341842
## CPI                -0.04029229
## CPILFESL            0.05406713
## PCEPI              -0.08396384
## PCE                -0.07347395
## hourly_earning     -0.21658220
## UMCSENT_interp     -0.04681026
## BBKMGDP            -0.02864438
## FEDFUNDS            0.10382313
## nasdaq             -0.83320605
## uldiff             -0.09441589
## nldiff             -0.09432015
## idiff               0.41027023
```

```
pca_scores <- pca_result$x[, 1:4]
y = datao4$IC
modelpca <- lm(y ~ pca_scores)
summary(modelpca)
```

```
##
## Call:
## lm(formula = y ~ pca_scores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3550 -0.2768  0.0029  0.1921  8.5397
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.808e-16  2.703e-02   0.000        1
## pca_scoresPC1  2.668e-01  1.385e-02  19.266  < 2e-16 ***
## pca_scoresPC2  2.059e-01  1.619e-02  12.720  < 2e-16 ***
## pca_scoresPC3  2.462e-01  1.893e-02  13.009  < 2e-16 ***
## pca_scoresPC4 -1.386e-01  2.446e-02  -5.668  2.2e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6823 on 632 degrees of freedom
## Multiple R-squared:  0.5374, Adjusted R-squared:  0.5345
## F-statistic: 183.6 on 4 and 632 DF,  p-value: < 2.2e-16
```

```
pca_scores2 <- pca_result$x[, 1:4]
y = datao4$CC
modelpca2 <- lm(y ~ pca_scores2)
summary(modelpca2)
```

```
##
## Call:
## lm(formula = y ~ pca_scores2)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2440 -0.2264  0.0185  0.2064  6.3679
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.209e-17  2.610e-02   0.000   1.0000
## pca_scores2PC1  1.860e-01  1.337e-02  13.910  < 2e-16 ***
## pca_scores2PC2 -3.948e-02  1.563e-02  -2.527   0.0118 *
## pca_scores2PC3  4.420e-01  1.827e-02  24.190  < 2e-16 ***
## pca_scores2PC4 -1.649e-01  2.362e-02  -6.982 7.36e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6587 on 632 degrees of freedom
## Multiple R-squared:  0.5688, Adjusted R-squared:  0.5661
## F-statistic: 208.4 on 4 and 632 DF,  p-value: < 2.2e-16
```

```
pca_scores <- pca_result$x[, 1:4]
pca_scores2 <- pca_result$x[, 1:4]
set.seed(3)
train_control <- trainControl(method = "cv", number = 5)
lmcvic2 <- train(
  x = pca_scores,
  y = y,
  method = "lm",
  trControl = train_control
)
(lmcvic2$results$RMSE)^2
```

```
## [1] 0.7676103
```

```
y = datao4$CC
lmcvcc2 <- train(
  x = pca_scores2,
  y = y,
  method = "lm",
  trControl = train_control
)
(lmcvcc2$results$RMSE)^2
```

```
## [1] 0.7474254
```

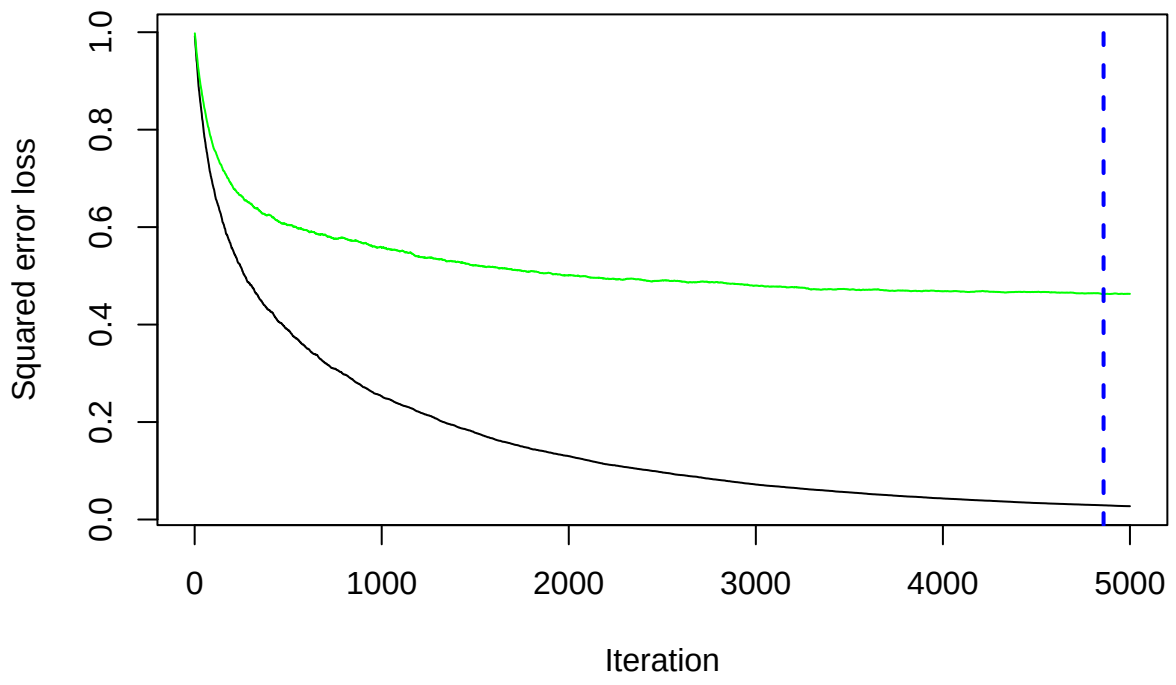
```
#boosting
library(gbm)
```

```
## Warning: package 'gbm' was built under R version 4.3.3
```

```
## Loaded gbm 2.2.2
```

```
## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.co
```

```
y = datao4$IC
mbx = cbind(y,xpca)

set.seed(2)
modelboosting <- gbm(
  formula = y ~ .,
  data = mbx,
  distribution = "gaussian",
  n.trees = 5000,
  interaction.depth = 3,
  shrinkage = 0.01,
  cv.folds = 5,
)
best_iter <- gbm.perf(modelboosting, method = "cv")
```



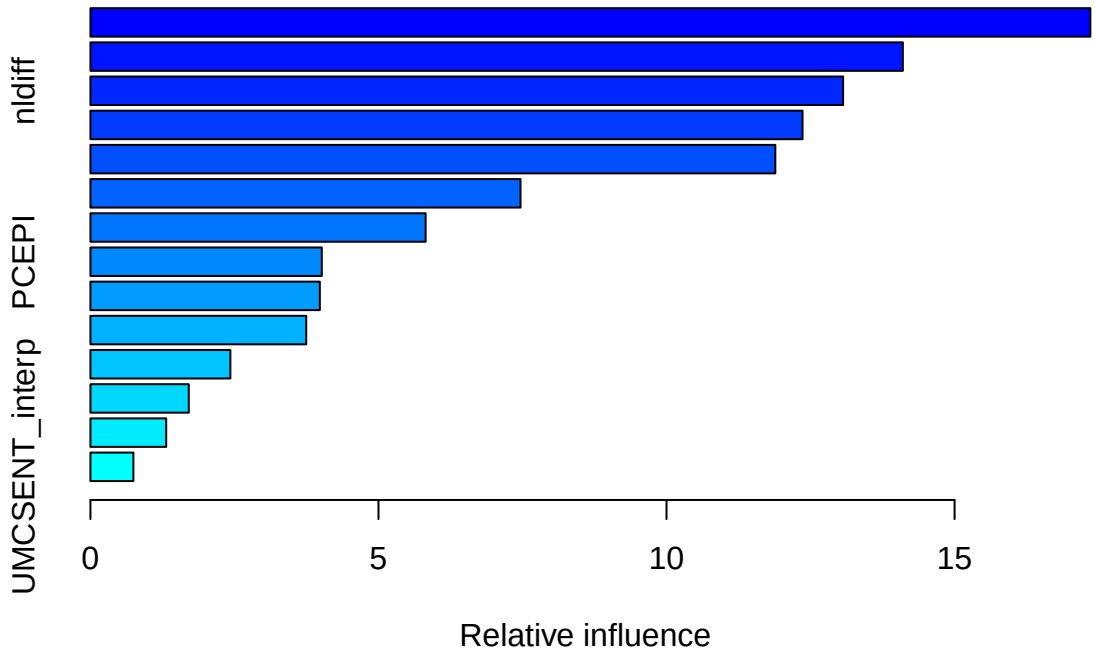

```
best_iter
```

```
## [1] 4859
```

```
modelboosting$cv.error[best_iter]
```

```
## [1] 0.4626691
```

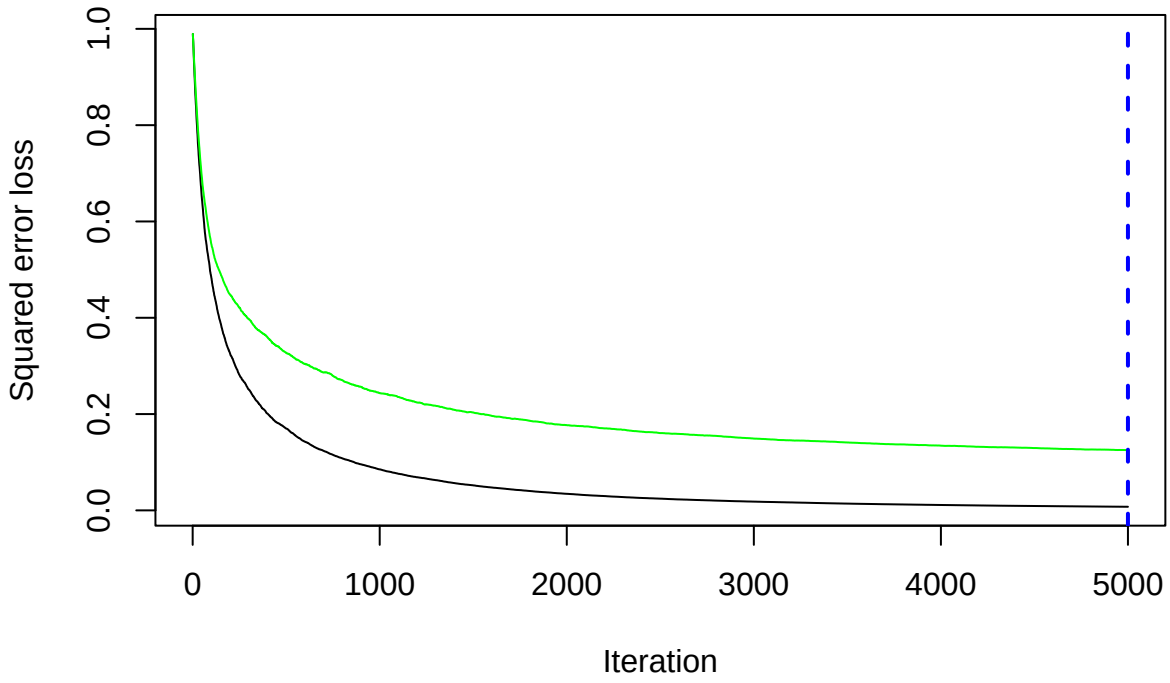
```r
summary(modelboosting)
```




```
##                                   var    rel.inf
## CPILFESL                     CPILFESL 17.3572035
## hourly_earning         hourly_earning 14.1026861
## nldiff                         nldiff 13.0652305
## BBKMGDP                       BBKMGDP 12.3612459
## num_losers                 num_losers 11.8887453
## uldiff                         uldiff  7.4654329
## PCE                               PCE  5.8187405
## PCEPI                           PCEPI  4.0162243
## idiff                           idiff  3.9820179
## unemployment_level unemployment_level  3.7455773
## CPI                               CPI  2.4293026
## FEDFUNDS                     FEDFUNDS  1.7069470
## nasdaq                         nasdaq  1.3149408
## UMCSENT_interp         UMCSENT_interp  0.7457054
```

```r
y = datao4$CC
mbx = cbind(y,xpca)

set.seed(1)
modelboosting2 <- gbm(
  formula = y ~ .,
  data = mbx,
```

```
  distribution = "gaussian",
  n.trees = 5000,
  interaction.depth = 3,
  shrinkage = 0.01,
  cv.folds = 5,
)
best_iter2 <- gbm.perf(modelboosting2, method = "cv")
```



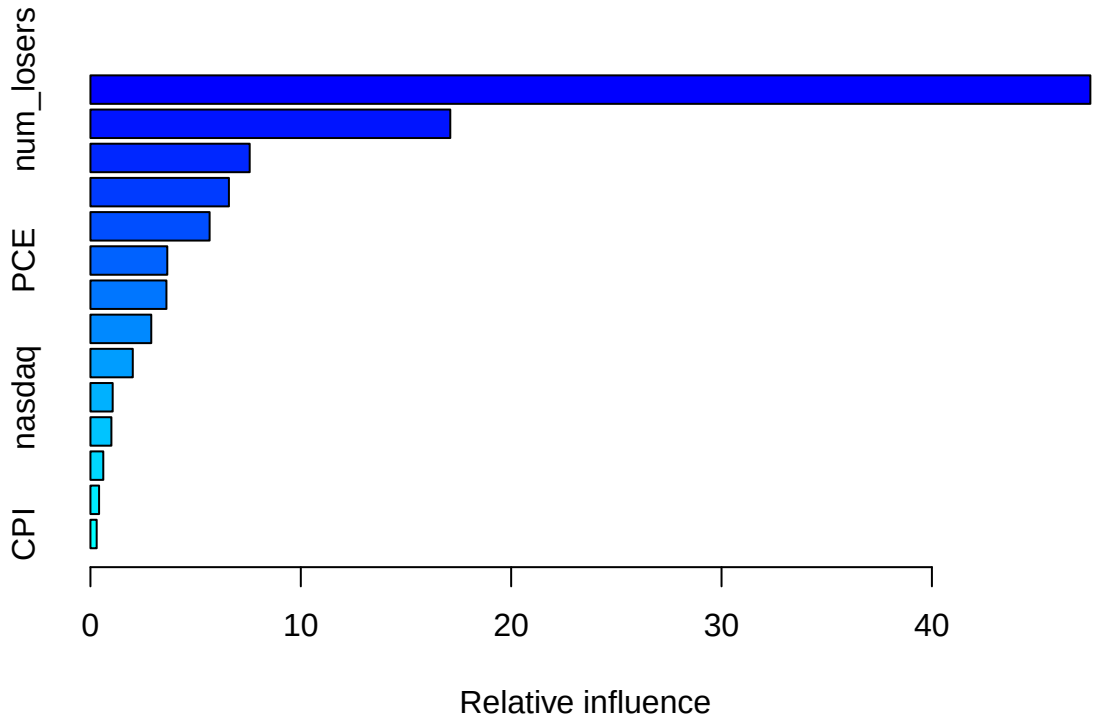

```
best_iter2
```

```
## [1] 5000
```

```
modelboosting2$cv.error[best_iter2]
```

```
## [1] 0.125189
```

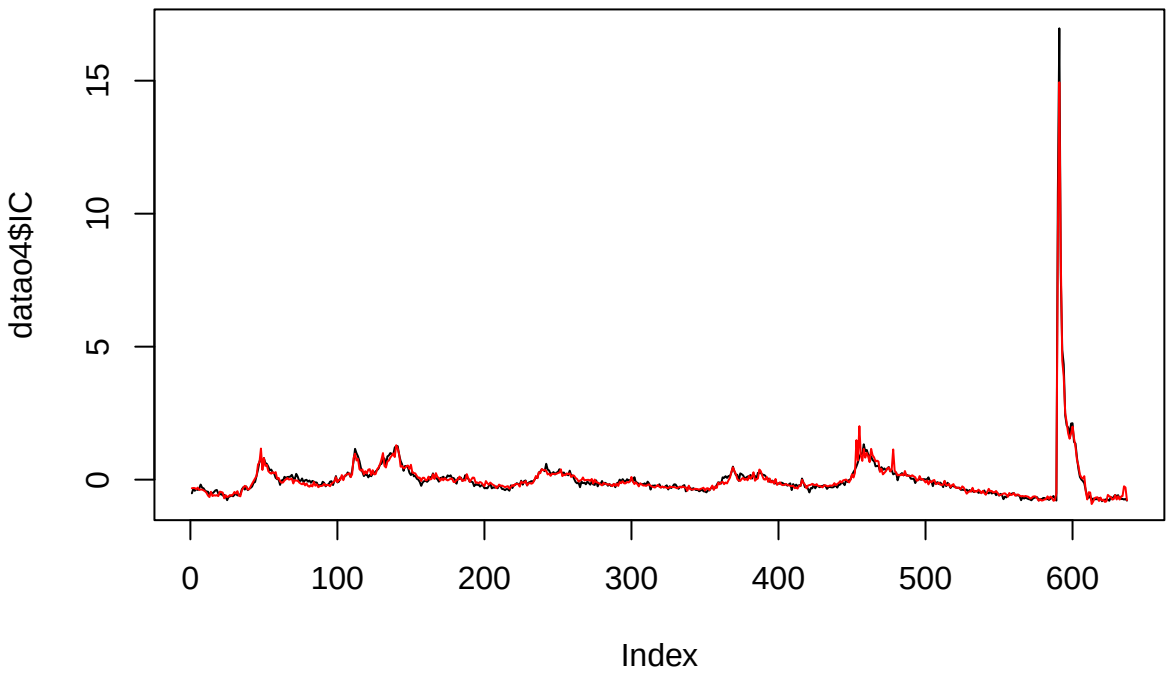
```
summary(modelboosting2)
```




```
##                                   var    rel.inf
## num_losers                 num_losers 47.5326069
## hourly_earning         hourly_earning 17.1052395
## unemployment_level unemployment_level  7.5708496
## nldiff                         nldiff  6.5830574
## uldiff                         uldiff  5.6648931
## PCE                               PCE  3.6534437
## BBKMGDP                       BBKMGDP  3.6131493
## CPILFESL                     CPILFESL  2.8915072
## FEDFUNDS                     FEDFUNDS  2.0130612
## nasdaq                         nasdaq  1.0572219
## UMCSENT_interp         UMCSENT_interp  0.9966262
## PCEPI                           PCEPI  0.6107841
## idiff                           idiff  0.4105221
## CPI                               CPI  0.2970379
```

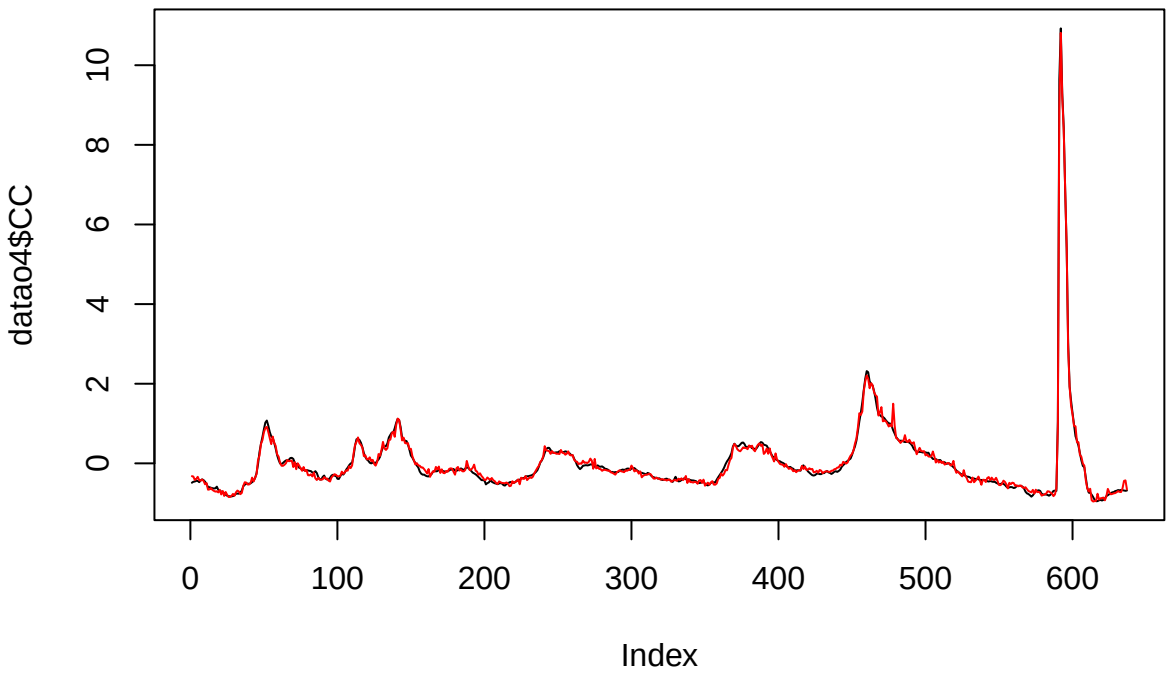
```
plot(datao4$IC, type = "l")
lines(predict(modelboosting), col = "red")
```

```
## Using 4859 trees...
```




```
plot(datao4$CC, type = "l")
lines(predict(modelboosting2), col = "red")
```

```
## Using 5000 trees...
```


data4$CC
0
2
4
6
8
10
0
100
200
300
400
500
600
Index