FinalProject

Read in data

```
data_orgin = read.table('Ghosal2020.txt',sep = "\t",header = 1)
data_orgin
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths   CFR
## 1        China      74185       57805      65112.00        2004 2.701
## 2        Italy      21157       17750      12207.00        1441 6.811
## 3        Spain       5232        4906       3097.00         133 2.542
## 4         Iran      11364        7321       9919.00         514 4.523
## 5       France       3661        3570        482.00          79 2.158
## 6           UK        798         769        495.00          11 1.378
## 7  Netherlands        804         792        134.00          10 1.244
## 8      Germany       3675        3621       3130.00           8 0.218
## 9      Belgium        559         555        139.00           3 0.537
## 10 Switzerland       1139        1124        303.00          11 0.966
## 11 South Korea       7979        7198       7294.42          67 0.840
## 12     Austria        504         497        431.00           1 0.198
## 13      Brazil        151         150        151.00           0 0.000
## 14   Indonesia         69          60         38.00           4 5.797
## 15         USA       2183        2126       1117.00          48 2.199
## 16       India        606         554         42.00          10 1.650
##    Week5deaths
## 1         2715
## 2         4825
## 3         1093
## 4         1433
## 5          450
## 6          177
## 7          106
## 8           68
## 9           37
## 10          56
## 11          94
## 12           6
## 13          11
## 14          32
## 15         255
## 16          NA
```

```
india = data_orgin[16,2:6]
india2 = data_orgin[16,3:5]
india3 = data_orgin[16,c(3,6)]
data = data_orgin[1:15,]
result = data[,7]
result
```

```
##  [1] 2715 4825 1093 1433  450  177  106   68   37   56   94    6   11   32  255
```

```
library(MASS)
library(car)
```

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

```
fit <- lm(Week5deaths ~ Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = data)
summary(fit)
```

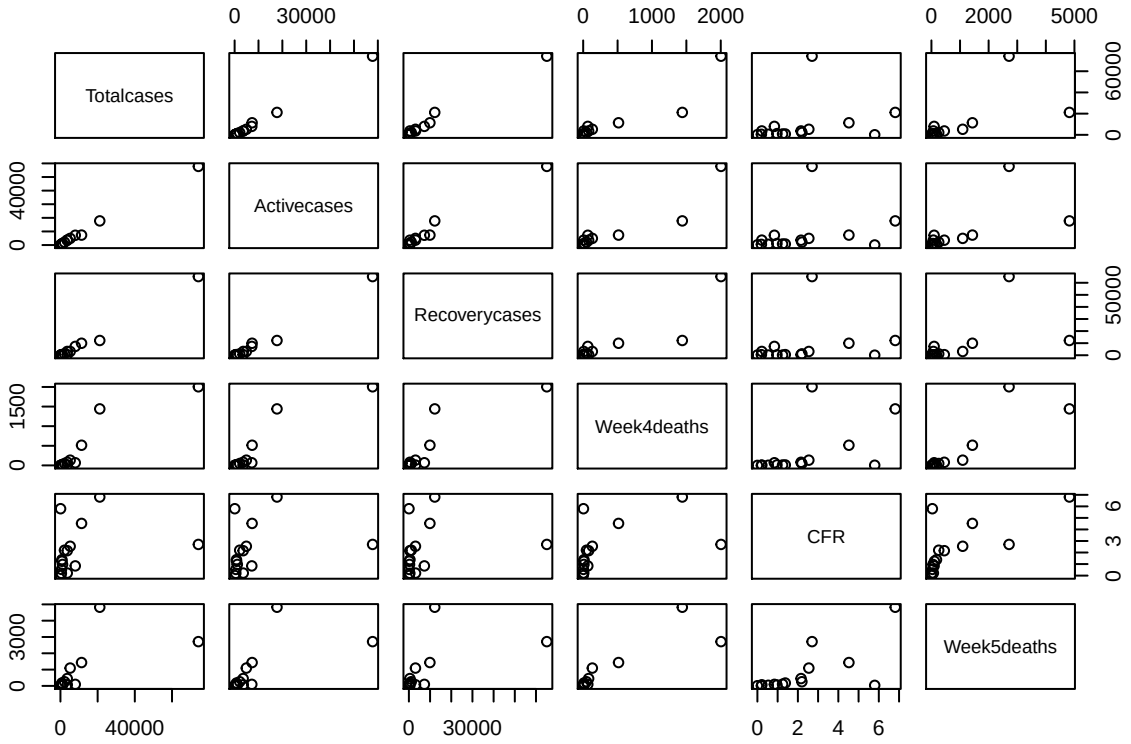
```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths + CFR, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -259.52  -86.60  -69.72   41.31  524.50
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    84.42474  115.00886   0.734 0.481590
## Totalcases     -0.06999    0.21816  -0.321 0.755657
## Activecases     0.12155    0.15538   0.782 0.454134
## Recoverycases  -0.09571    0.10966  -0.873 0.405463
## Week4deaths     3.49750    0.70392   4.969 0.000771 ***
## CFR            33.51329   46.33829   0.723 0.487907
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 234.1 on 9 degrees of freedom
## Multiple R-squared:  0.9807, Adjusted R-squared:  0.9701
## F-statistic:  91.7 on 5 and 9 DF,  p-value: 1.925e-07
```

```
vif(fit)
```

```
##    Totalcases   Activecases Recoverycases   Week4deaths           CFR
##   4356.999162   1338.142312    842.392401     46.419063      2.366481
```

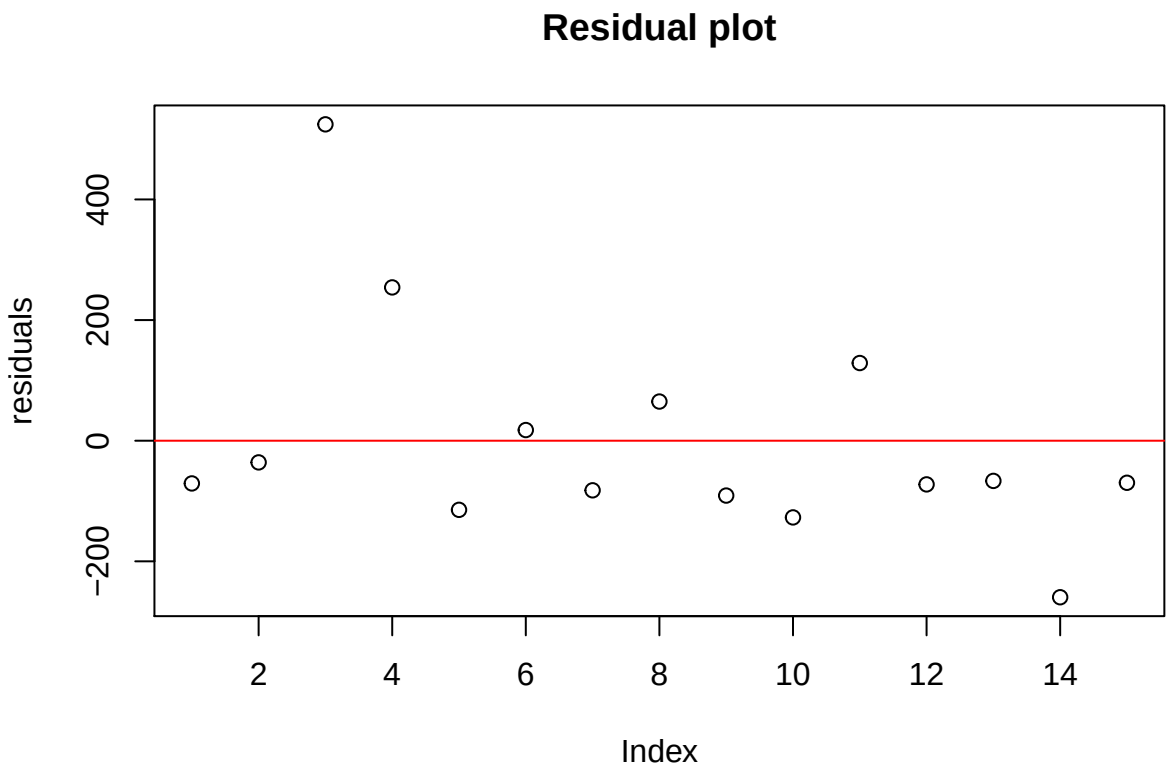
From the coefficient, we can see that the week 5 death is largely determined by CFC. Also, we found out that the week 5 death is negatively accosiated with recovery cases, which is reasonable.

```
pairs(data[,2:7])
```

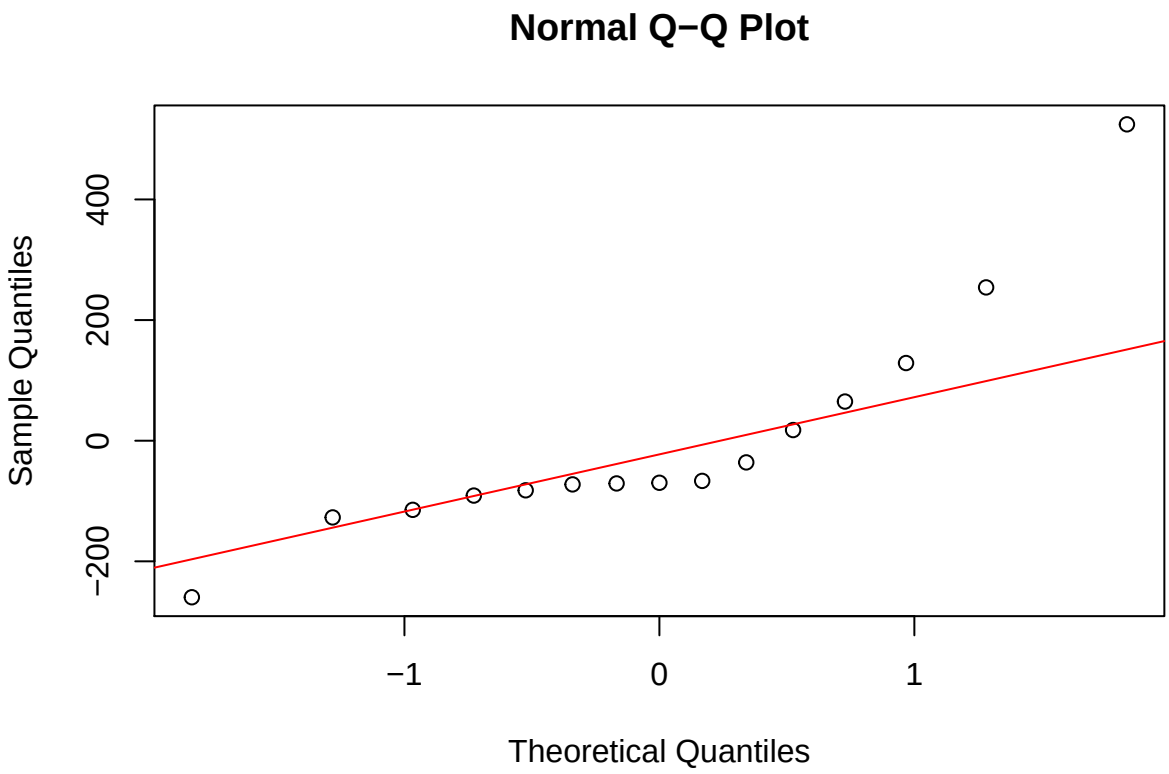


From the pairplot, we also can see that week5 death is slightly positive associated with total case, active case, whereas the week5 death is positively associated with CFC. (CFC is calculated based on other info)

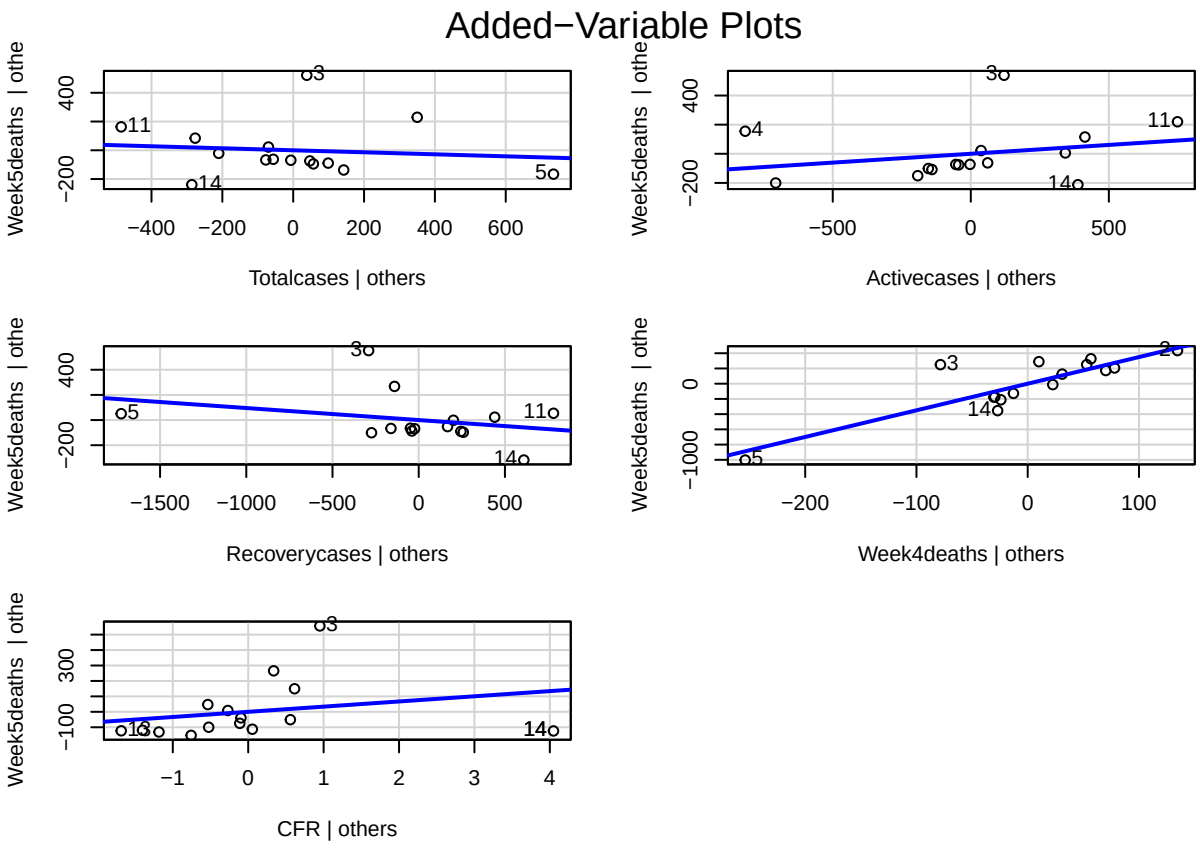
```
plot(fit$residuals, main = 'Residual plot', ylab = 'residuals')
abline(0,0,col = 'red')
```




```
qqnorm(fit$residuals)
qqline(fit$residuals, col="red")
```


Normal Q−Q Plot
Sample Quantiles
−200
0
200
400
−1
0
1
Theoretical Quantiles

```
avPlots(fit)
```



Prediction

We want to predict week 5 death for india. prediction is 195.6021 confidence interval for mean of predictions is [34.3962,356.808]

```
predict(fit, newdata = data[,2:6])
```

```
##           1           2           3           4           5           6
## 2785.827554 4860.947094  568.499800 1178.831255  564.619977  159.321168
##           7           8           9          10          11          12
##  188.261423    3.050843  127.946334  183.175034  -34.794012   78.441317
##          13          14          15
##   77.636182  291.517893  324.718138
```

```
predict(fit, newdata = india)
```

```
##       16
## 195.6021
```

```
predict(fit, newdata = india,interval = "confidence")
```

```
##         fit     lwr     upr
## 16 195.6021 34.3962 356.808
```

```r
predict(fit, newdata = india,interval = "prediction")
```

```
##         fit       lwr     upr
## 16 195.6021 -358.0397 749.244
```

Outlier

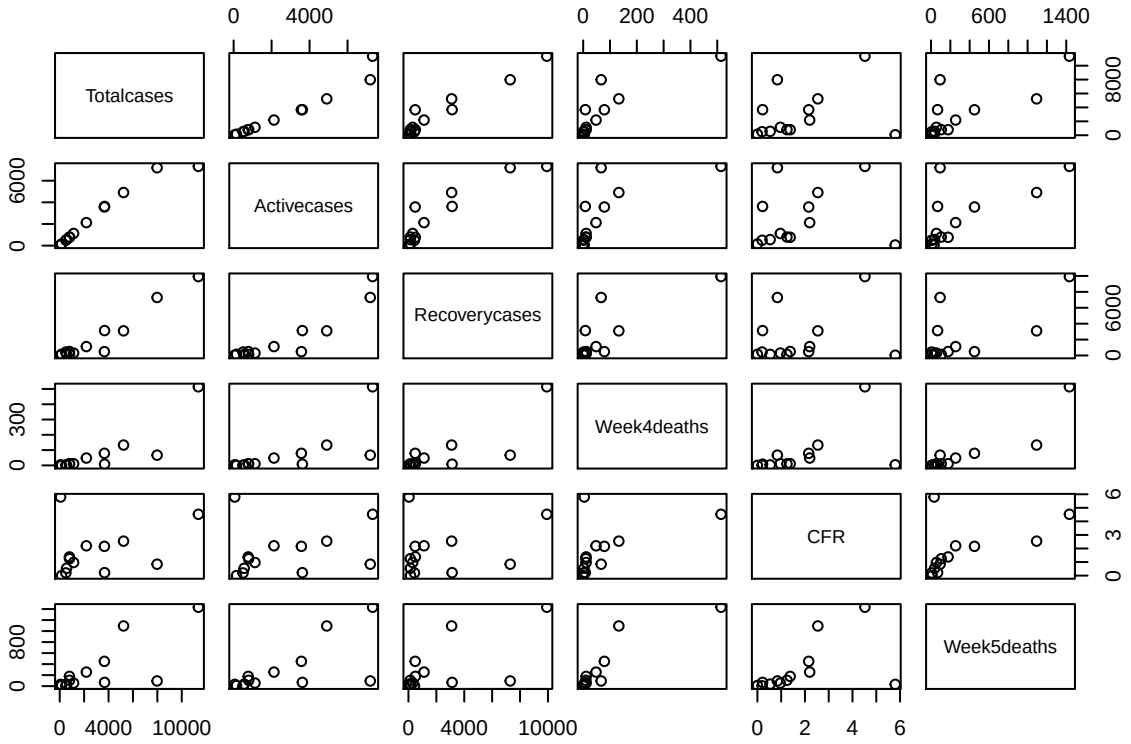
We did found outliers inside our dataset, which is China and Italy. We will analyze our fitting model by dropping info from china and italy, and then seperate

```r
dropped= data[3:15,]
drop_china = data[-1,]
drop_italy= data[-2,]
```

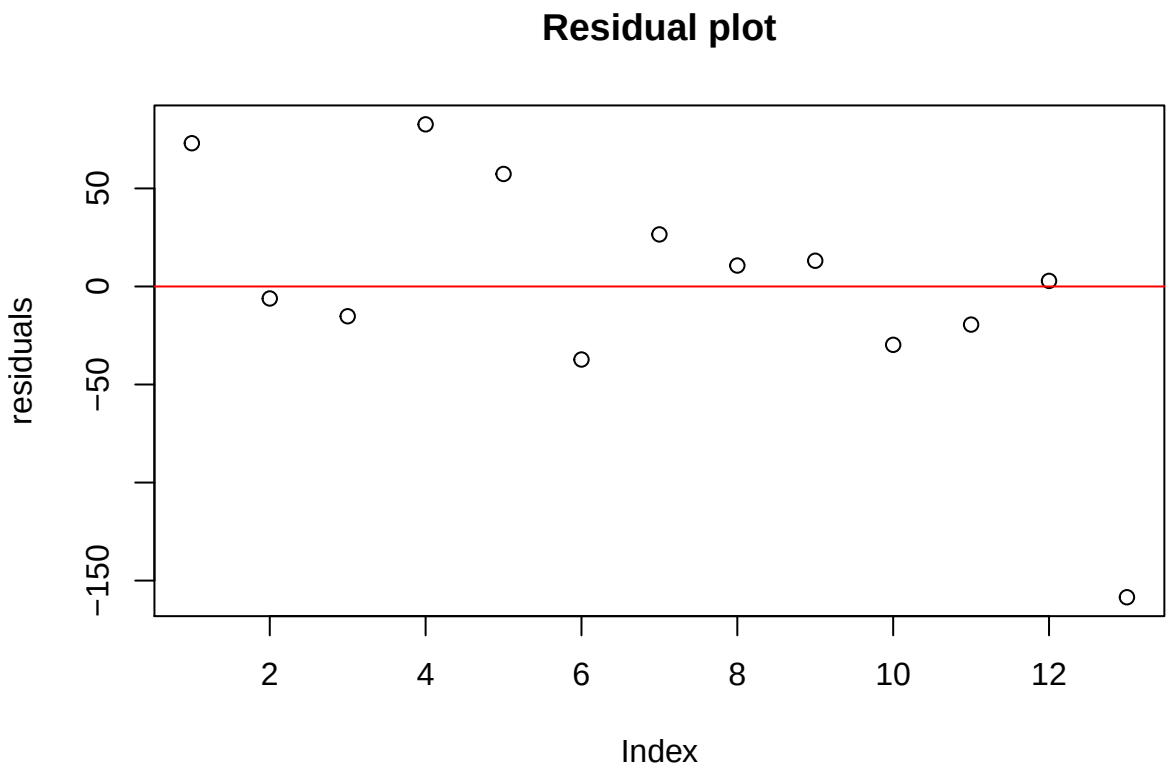
```r
fitnew <- lm(Week5deaths ~ Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = dropped)
summary(fitnew)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths + CFR, data = dropped)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -158.490  -19.439    2.842   26.544   82.676
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   26.15720   42.91299   0.610 0.561422
## Totalcases    -1.24408    0.22462  -5.539 0.000870 ***
## Activecases    1.13137    0.18198   6.217 0.000438 ***
## Recoverycases  0.14856    0.05649   2.630 0.033933 *
## Week4deaths   11.31865    1.43362   7.895 9.91e-05 ***
## CFR           -5.16809   16.48295  -0.314 0.763006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.73 on 7 degrees of freedom
## Multiple R-squared:  0.9819, Adjusted R-squared:  0.9689
## F-statistic: 75.87 on 5 and 7 DF,  p-value: 6.083e-06
```

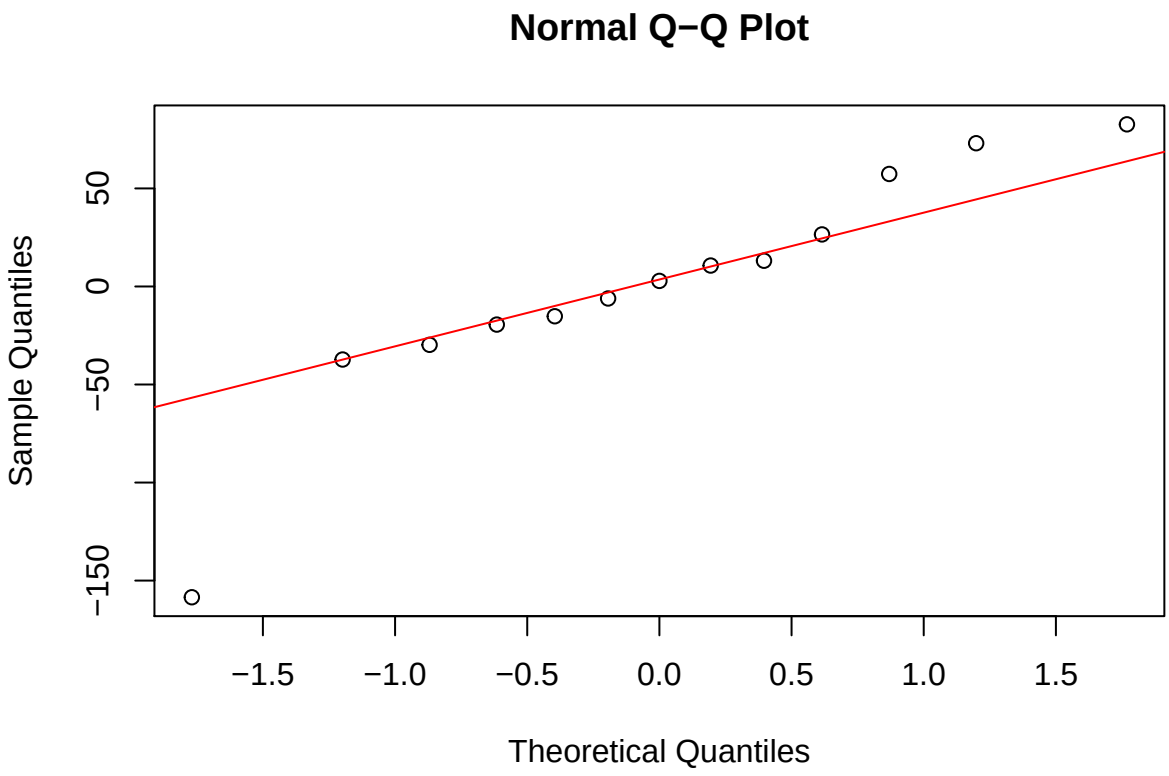
```r
pairs(dropped[,2:7])
```




```
plot(fitnew$residuals, main = 'Residual plot', ylab = 'residuals')
abline(0,0,col = 'red')
```



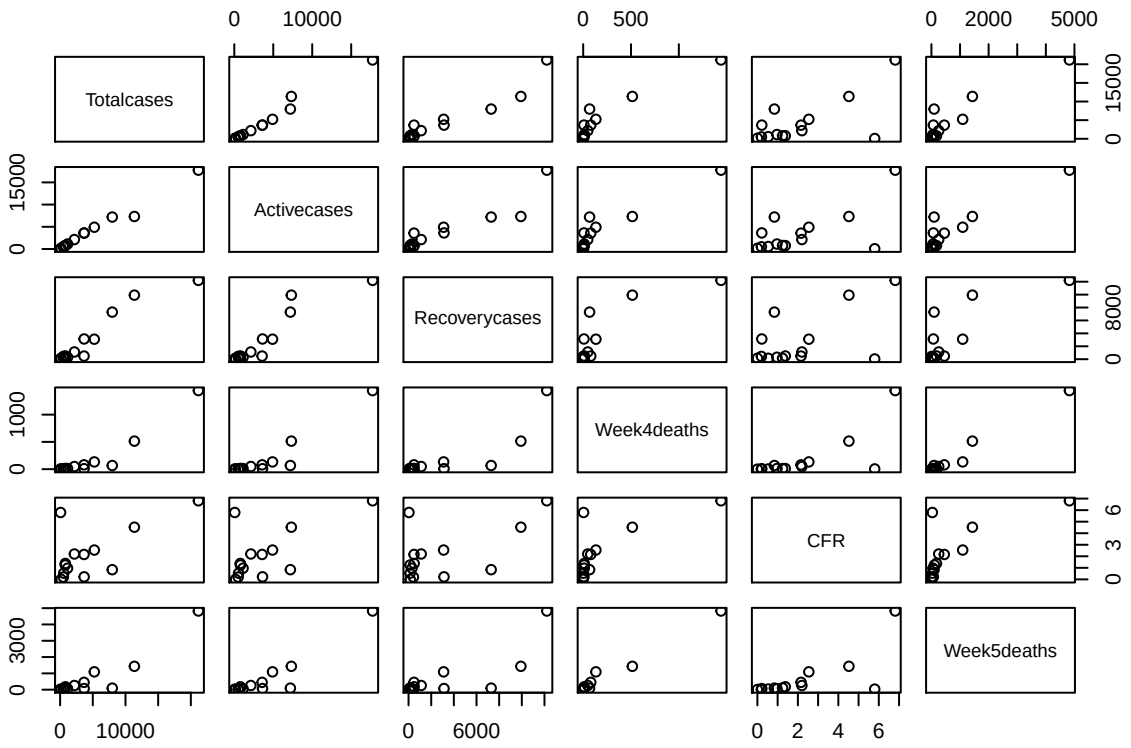

```
qqnorm(fitnew$residuals)
qqline(fitnew$residuals, col="red")
```



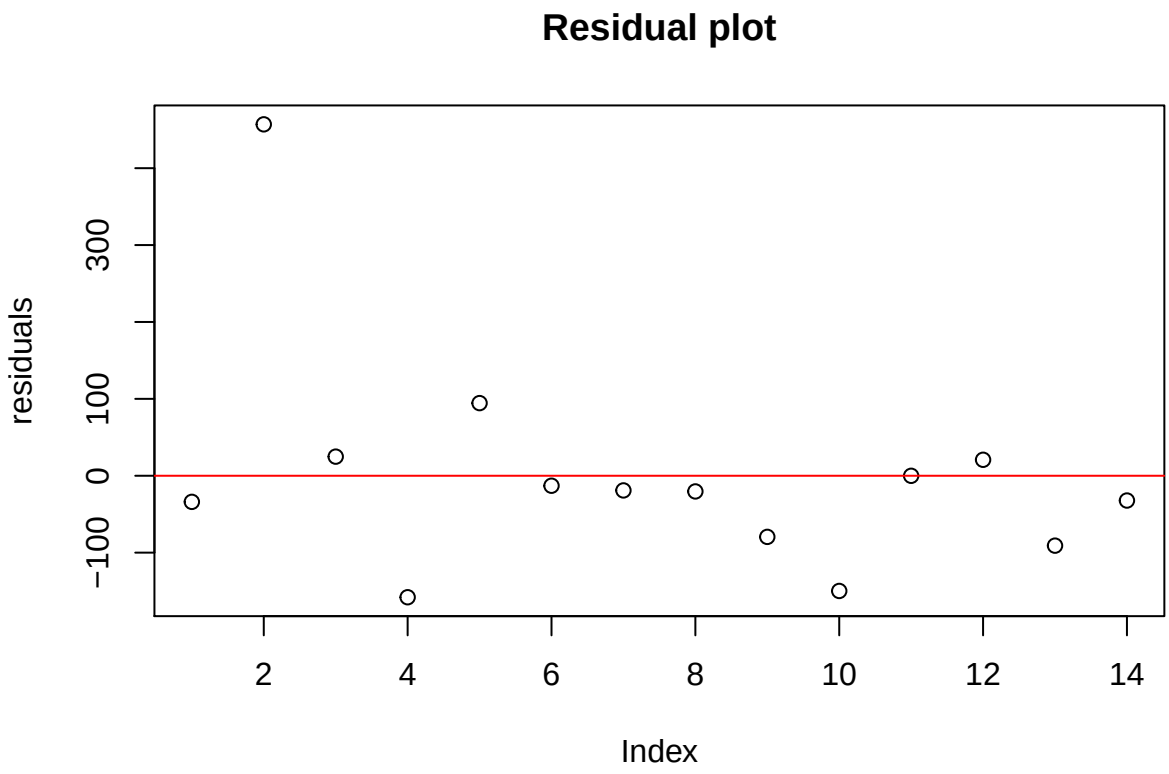

```
fitchina <- lm(Week5deaths ~ Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = drop_c
summary(fitchina)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths + CFR, data = drop_china)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -158.01  -68.12  -19.75   15.55  457.01
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -10.29184  100.42915  -0.102   0.9209
## Totalcases      0.05055    0.18243   0.277   0.7887
## Activecases     0.06445    0.12723   0.507   0.6261
## Recoverycases  -0.11111    0.08847  -1.256   0.2446
## Week4deaths     2.68683    0.65747   4.087   0.0035 **
## CFR            20.58698   37.66645   0.547   0.5996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 188.4 on 8 degrees of freedom
## Multiple R-squared:  0.9868, Adjusted R-squared:  0.9786
## F-statistic: 119.7 on 5 and 8 DF,  p-value: 2.692e-07
```

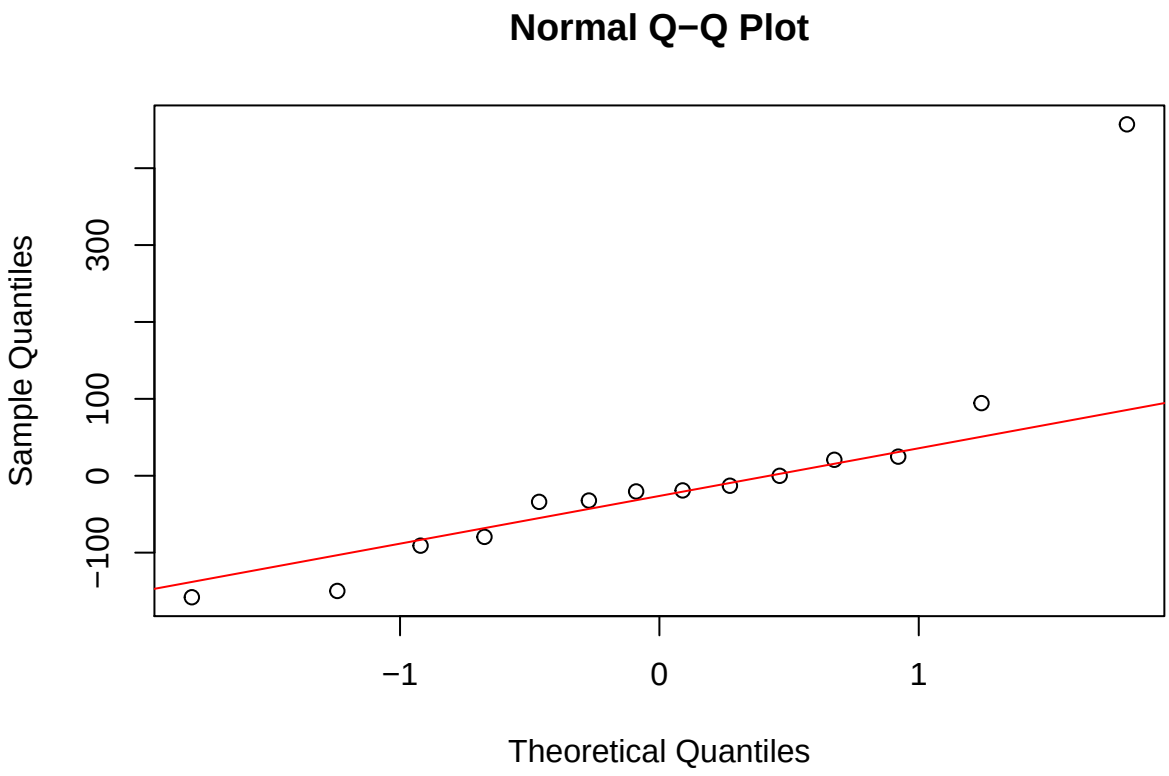
```r
pairs(drop_china[,2:7])
```




```r
plot(fitchina$residuals, main = 'Residual plot', ylab = 'residuals')
abline(0,0,col = 'red')
```



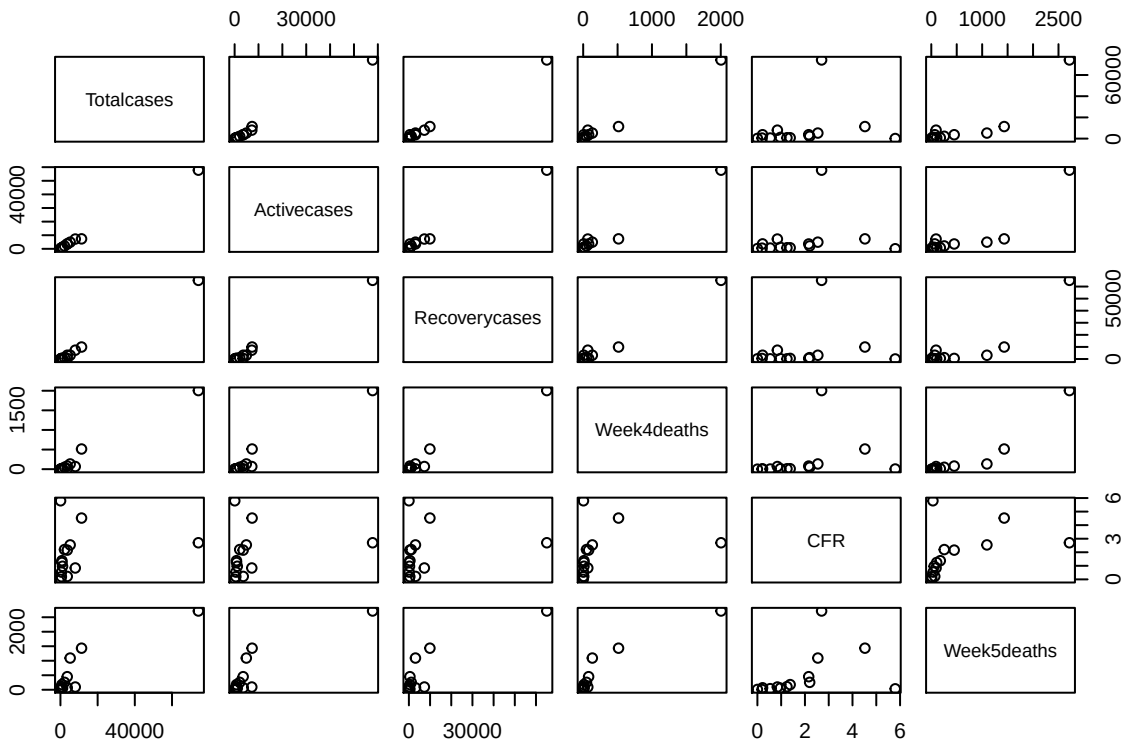

```
qqnorm(fitchina$residuals)
qqline(fitchina$residuals, col="red")
```



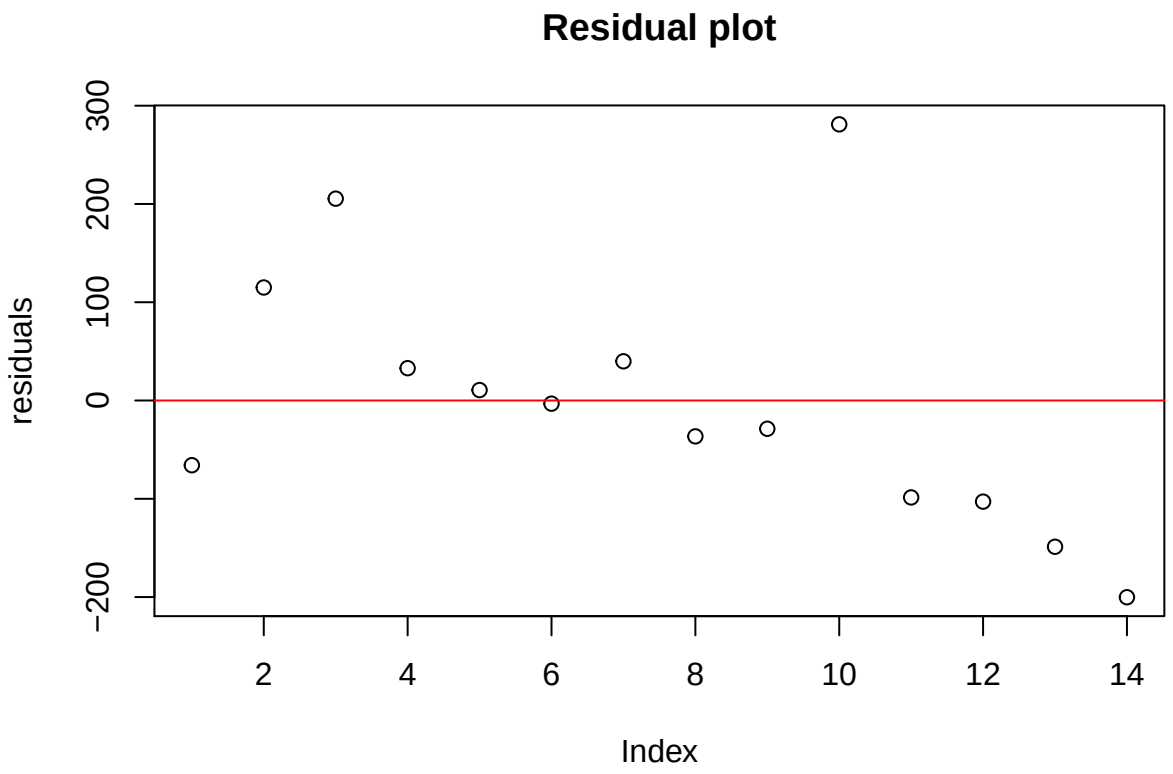

```
fitita <- lm(Week5deaths ~ Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = drop_ita
summary(fitnew)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths + CFR, data = dropped)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -158.490  -19.439    2.842   26.544   82.676
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   26.15720   42.91299   0.610 0.561422
## Totalcases    -1.24408    0.22462  -5.539 0.000870 ***
## Activecases    1.13137    0.18198   6.217 0.000438 ***
## Recoverycases  0.14856    0.05649   2.630 0.033933 *
## Week4deaths   11.31865    1.43362   7.895 9.91e-05 ***
## CFR           -5.16809   16.48295  -0.314 0.763006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.73 on 7 degrees of freedom
## Multiple R-squared:  0.9819, Adjusted R-squared:  0.9689
## F-statistic: 75.87 on 5 and 7 DF,  p-value: 6.083e-06
```

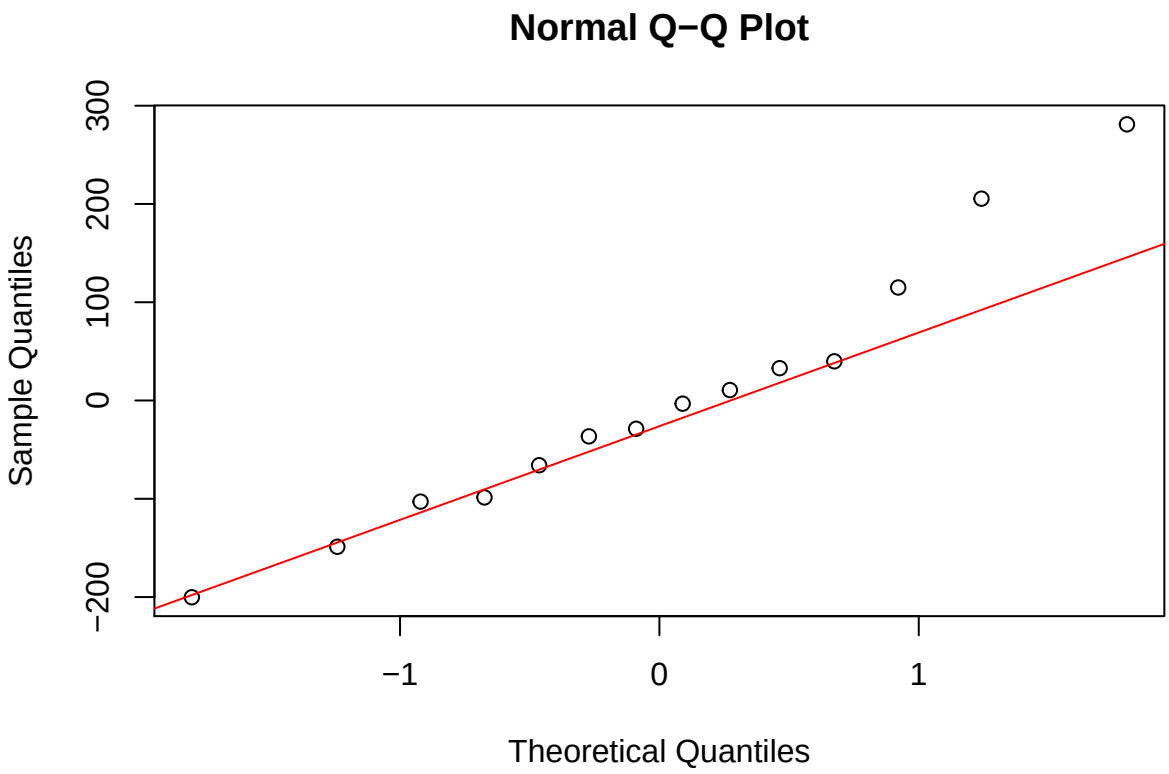
```
pairs(drop_italy[,2:7])
```




```
plot(fitita$residuals, main = 'Residual plot', ylab = 'residuals')
abline(0,0,col = 'red')
```




```
qqnorm(fitita$residuals)
qqline(fitita$residuals, col="red")
```



Although r-square did not incrase much, the pair plot did correlate better. However, we found out the slope all some variables changed a lot, especially CFR, it turned from positive 30 to -5 for both counries dropped, 20 for china dropped and 5 for only italy dropped. We doubt our original linear regression.

Scenario2

As we can see in the first part, the patterns between each pair of variables are not linear. Also, the residual plot shows extreme outliers, which means the linear model is very sensitive to outliers. These extreme outliers are influential and may cause errors to our linear model. Therefore, we revise the model by taking a logarithmic transformation of the entire dataset to reduce the effect of some influential skewed data(outliers).

But before we apply the logarithmic transformation of the dataset, we need first eliminate Brazil's data. That is because 'Week4deaths' and 'CFR' are both 0. If we take log of them, the entries will become '-inf' and cause errors.

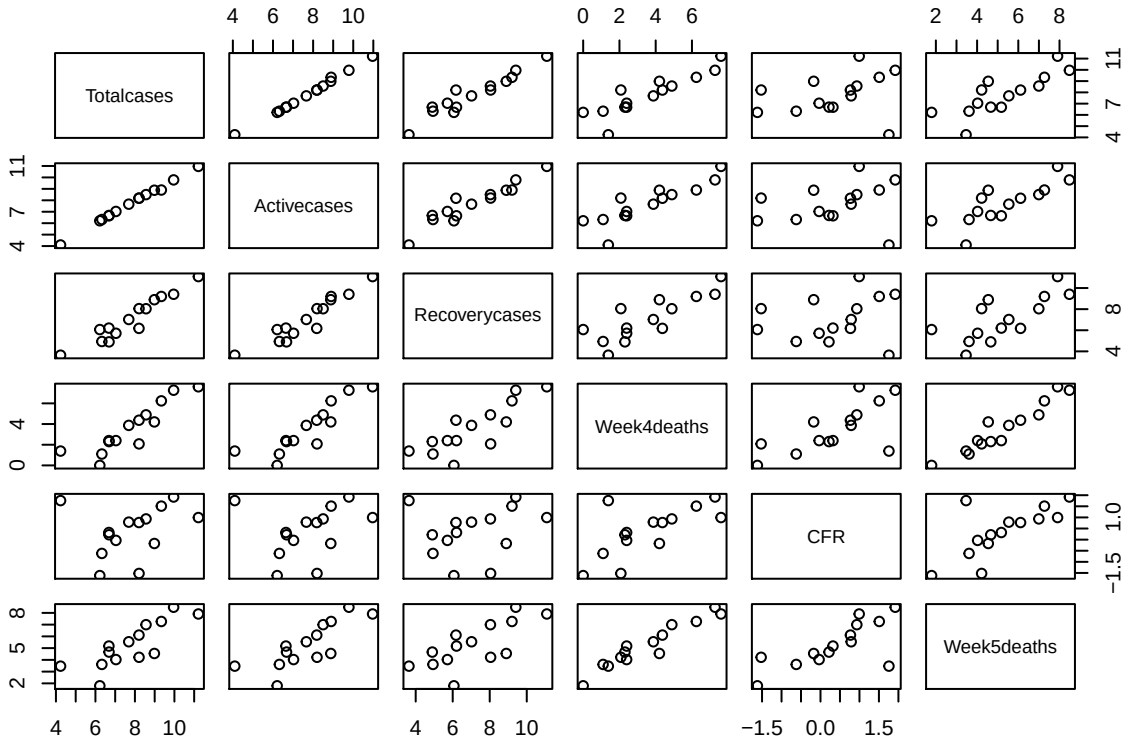
```
logdata<-data[-13,]
logdata<-data.frame(logdata['Countries'],log(logdata['Totalcases']),log(logdata['Activecases']),log(log
logdata
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths         CFR
## 1        China  11.214317   10.964831     11.083864    7.602900  0.99362207
## 2        Italy   9.959726    9.784141      9.409765    7.273093  1.91853895
## 3        Spain   8.562549    8.498214      8.038189    4.890349  0.93295117
## 4         Iran   9.338206    8.898502      9.202207    6.242223  1.50917549
## 5       France   8.205492    8.180321      6.177944    4.369448  0.76918187
```

```
## 6           UK   6.682109    6.645091      6.204558    2.397895  0.32063317
## 7  Netherlands   6.689599    6.674561      4.897840    2.302585  0.21833199
## 8      Germany   8.209308    8.194506      8.048788    2.079442 -1.52326022
## 9      Belgium   6.326149    6.318968      4.934474    1.098612 -0.62175718
## 10 Switzerland   7.037906    7.024649      5.713733    2.397895 -0.03459144
## 11 South Korea   8.984568    8.881558      8.894865    4.204693 -0.17435339
## 12     Austria   6.222576    6.208590      6.066108    0.000000 -1.61948825
## 14   Indonesia   4.234107    4.094345      3.637586    1.386294  1.75734054
## 15         USA   7.688455    7.661998      7.018402    3.871201  0.78800271
##    Week5deaths
## 1     7.906547
## 2     8.481566
## 3     6.996681
## 4     7.267525
## 5     6.109248
## 6     5.176150
## 7     4.663439
## 8     4.219508
## 9     3.610918
## 10    4.025352
## 11    4.543295
## 12    1.791759
## 14    3.465736
## 15    5.541264
```

We draw the pair graph to see the pattern between each variable. Compare to the original pair graph, our new graph shows a noticeable linear relationship between each pair of variables.

```
pairs(logdata[c("Totalcases", "Activecases", "Recoverycases","Week4deaths","CFR","Week5deaths")])
```



fit We use our new dataset to fit a new linear model. Although R-square decreases a little bit, but it still shows good fit.

```
fit2 <- lm(Week5deaths~Totalcases+Activecases+Recoverycases+Week4deaths+CFR, data = logdata)
summary(fit2)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths + CFR, data = logdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0140 -0.2195  0.0440  0.2784  0.7883
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -581.1014  1067.4115  -0.544    0.601
## Totalcases     125.2276   231.8770   0.540    0.604
## Activecases      1.6159     2.1969   0.736    0.483
## Recoverycases   -0.1072     0.3223  -0.332    0.748
## Week4deaths   -125.9814   231.8386  -0.543    0.602
## CFR            127.0035   231.8375   0.548    0.599
##
## Residual standard error: 0.5877 on 8 degrees of freedom
## Multiple R-squared:  0.9408, Adjusted R-squared:  0.9038
## F-statistic: 25.43 on 5 and 8 DF,  p-value: 0.000103
```

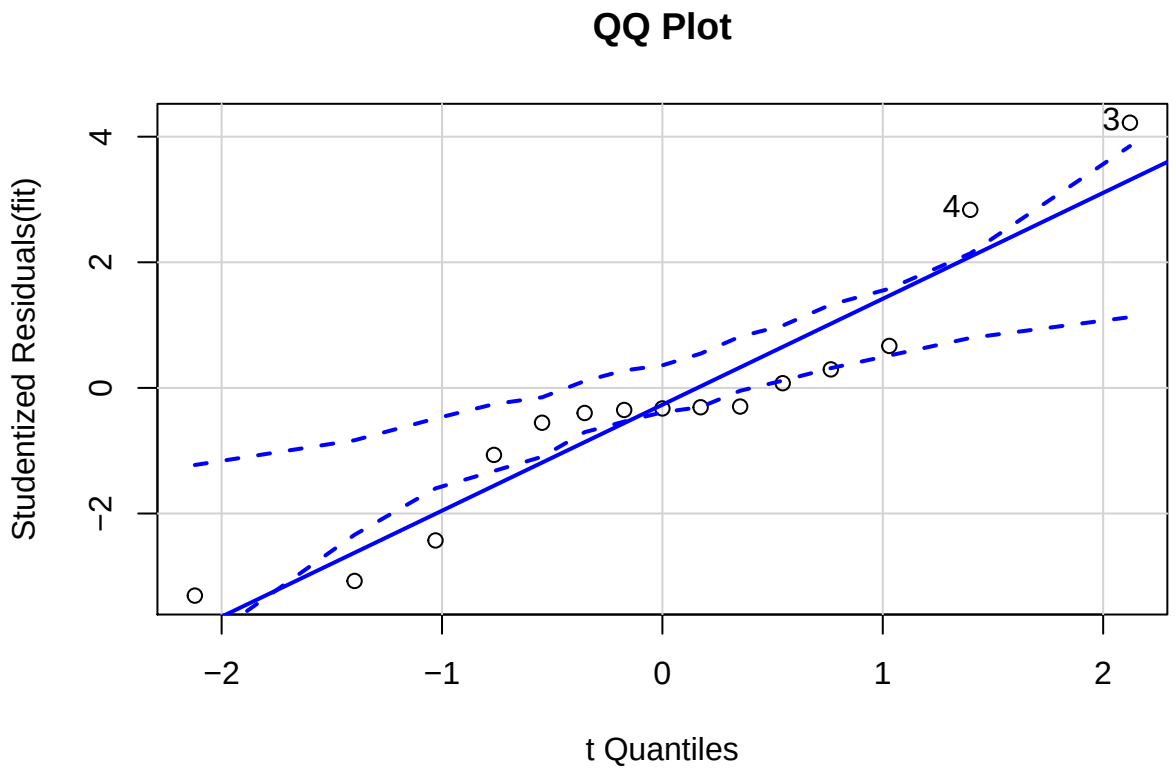
```
vif(fit2)
```

```
##    Totalcases   Activecases Recoverycases   Week4deaths           CFR
##  6.452851e+06  5.407396e+02  1.708497e+01  1.092204e+07  2.438601e+06
```

regression diagnostics

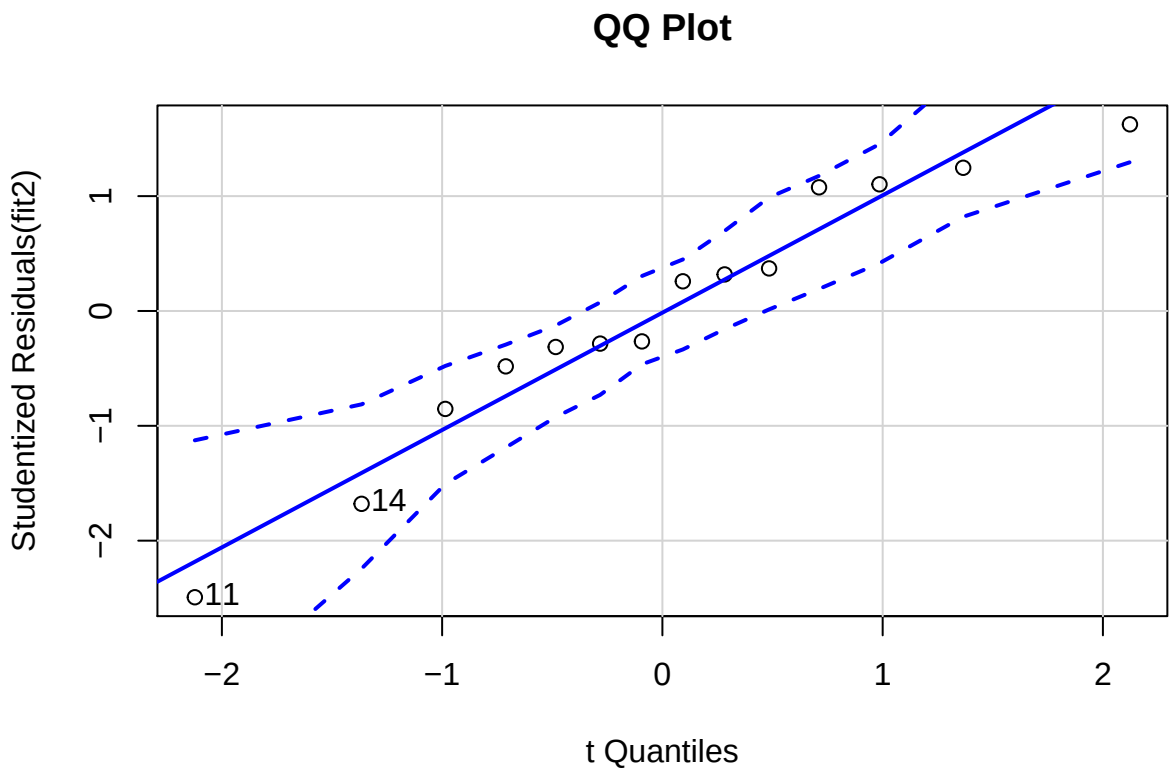
Now we do regression diagnostics of our new model. By comparing the qqplots for the first model and the second model, we can see our new model's residuals lie approximately along the line which indicates a better fit. Our new leverage plot also shows that our new model is less affected by outliers.

```
qqPlot(fit, main = "QQ Plot")
```



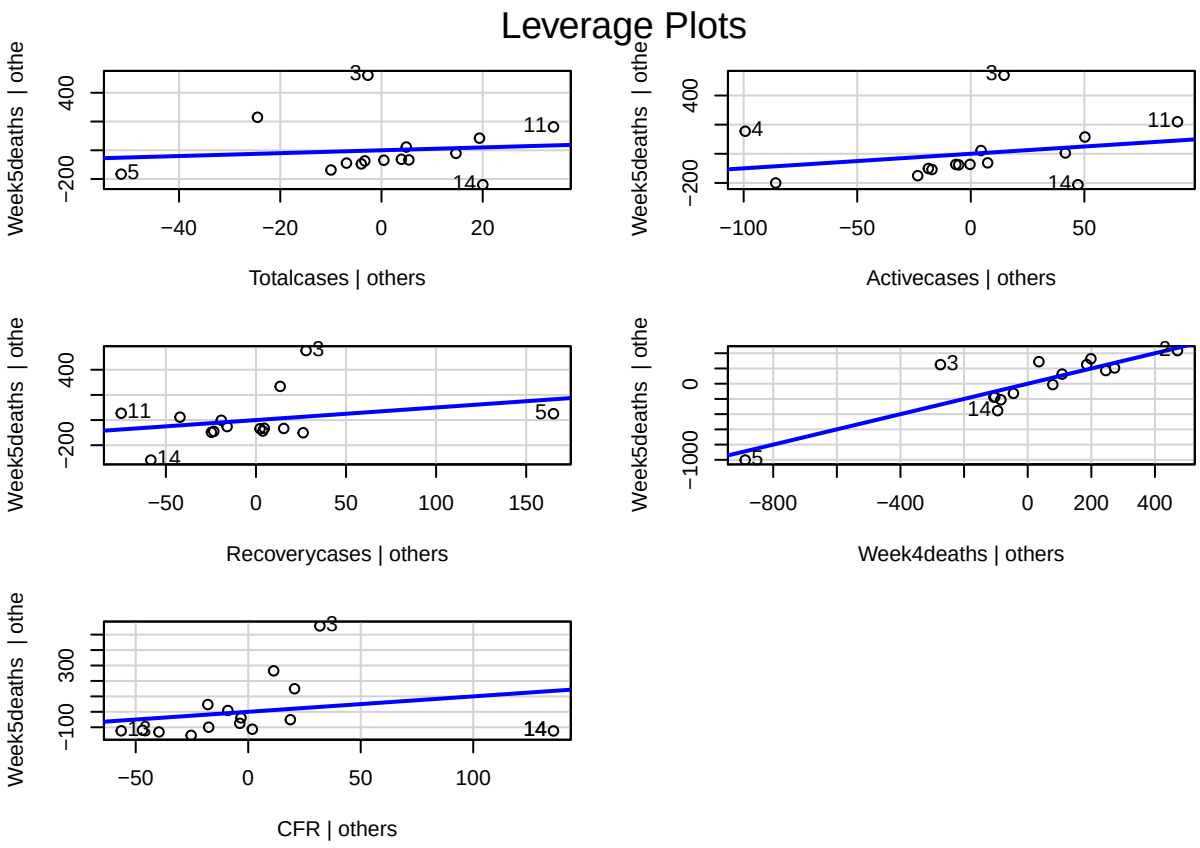

```
## [1] 3 4
```

```
qqPlot(fit2, main="QQ Plot") #qq plot for studentized resid
```

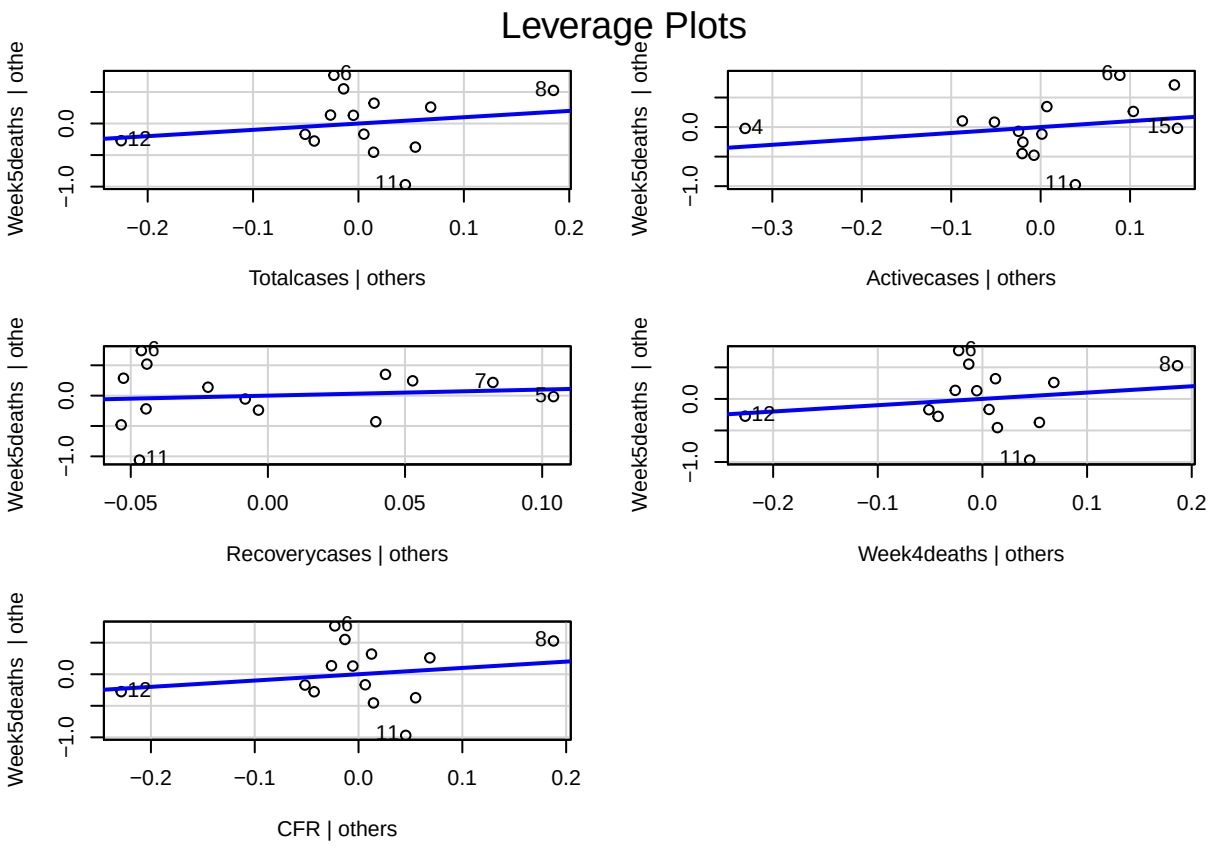



```
## 11 14
## 11 13
```

```
leveragePlots(fit) # leverage plots
```

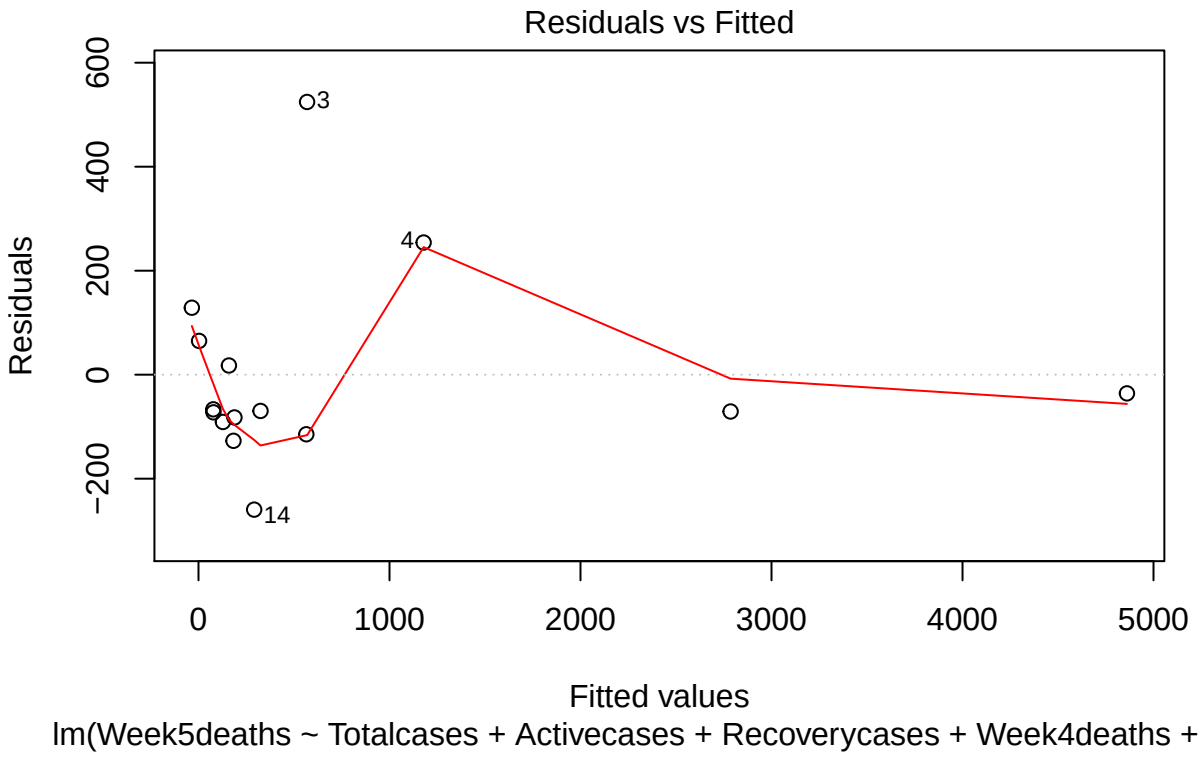


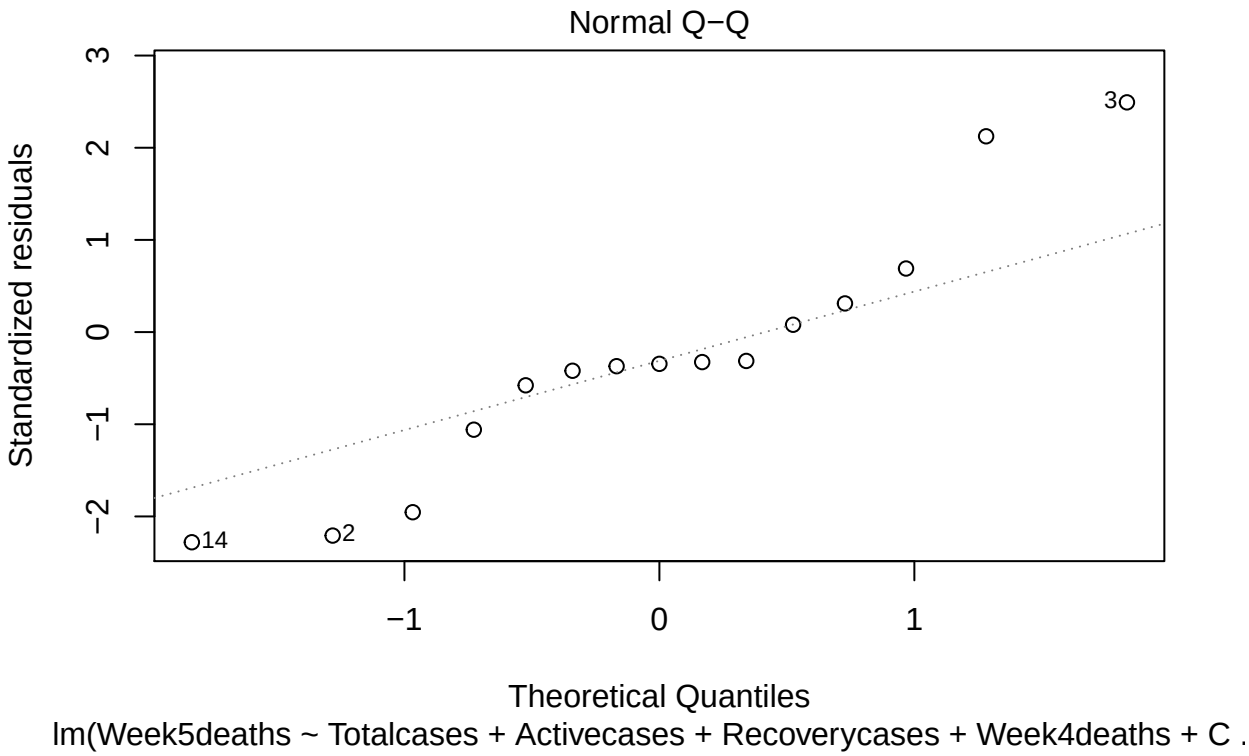

```
leveragePlots(fit2)
```

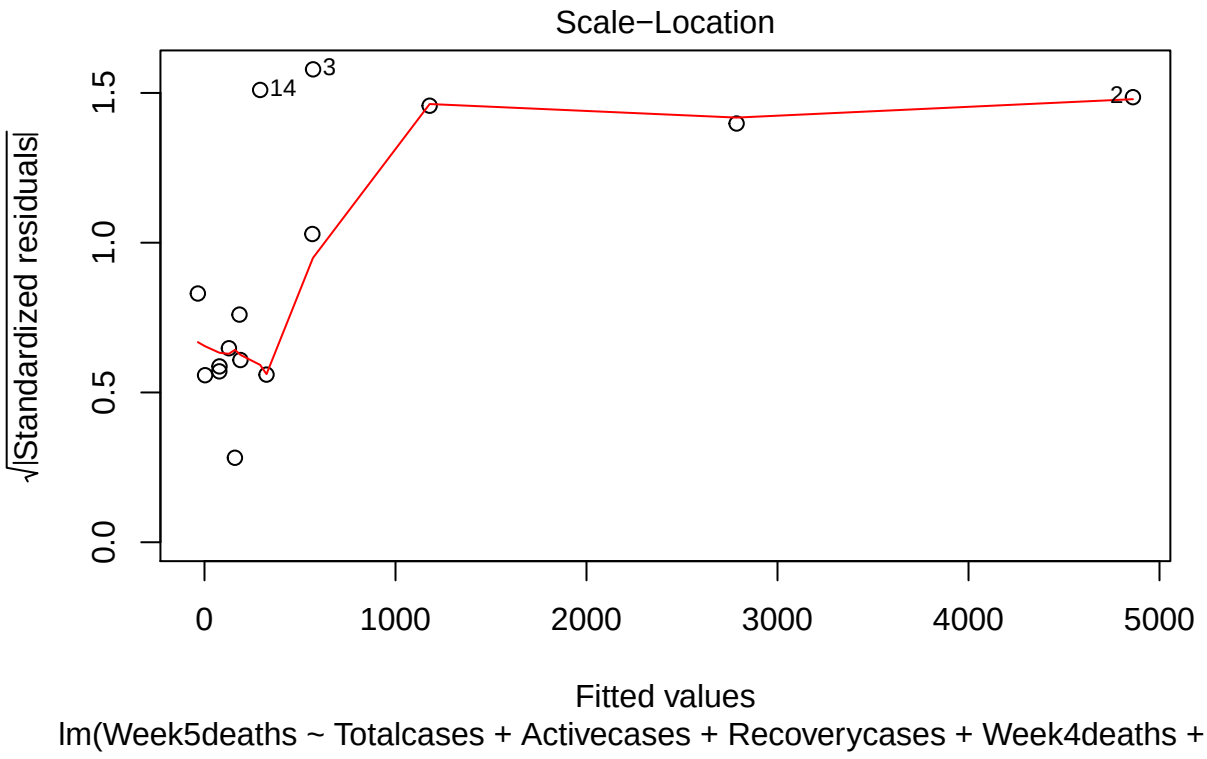


Moreover, we have less residuals that are outside of Cook's distance, which indicate that we have less influential outliers. However, in analysis of normality of residuals. We find that the first model's residuals are more likely to follow a normal distribution.

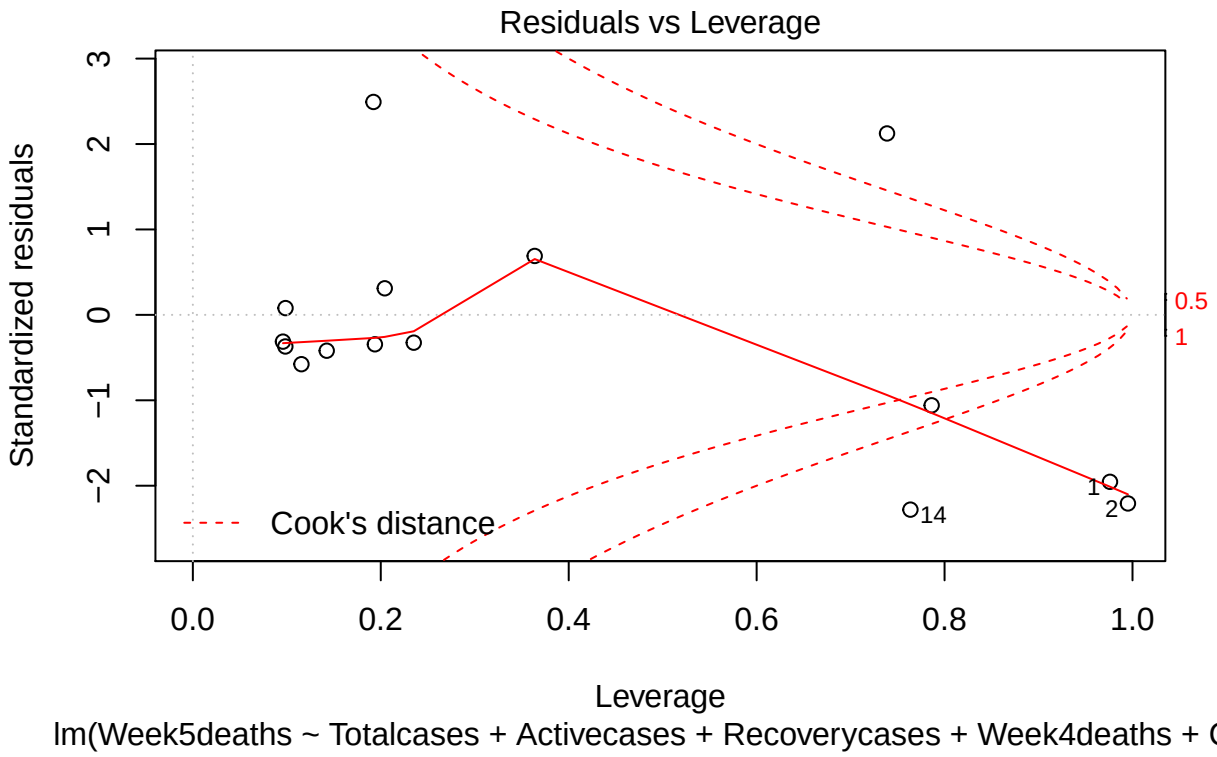
```
plot(fit)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .
3
4
14
600
400
200
0
−200
0
1000
2000
3000
4000
5000

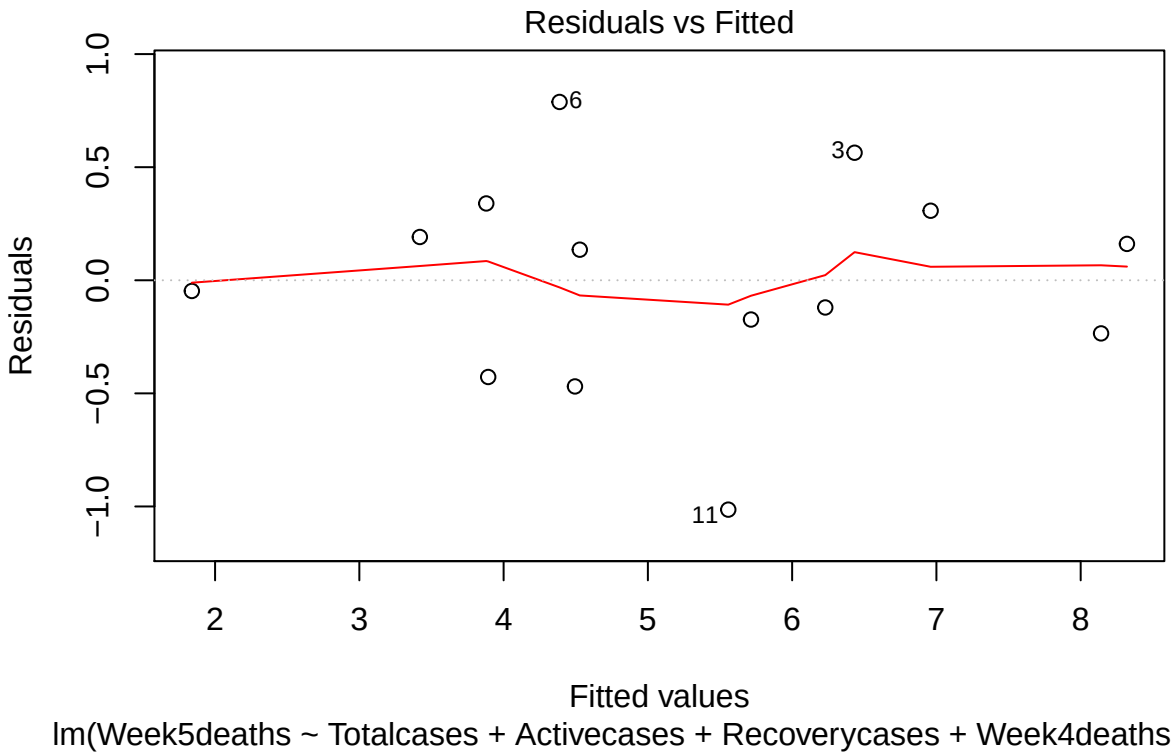

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .
14
2
3
−2
−1
0
1
2
3
−1
0
1

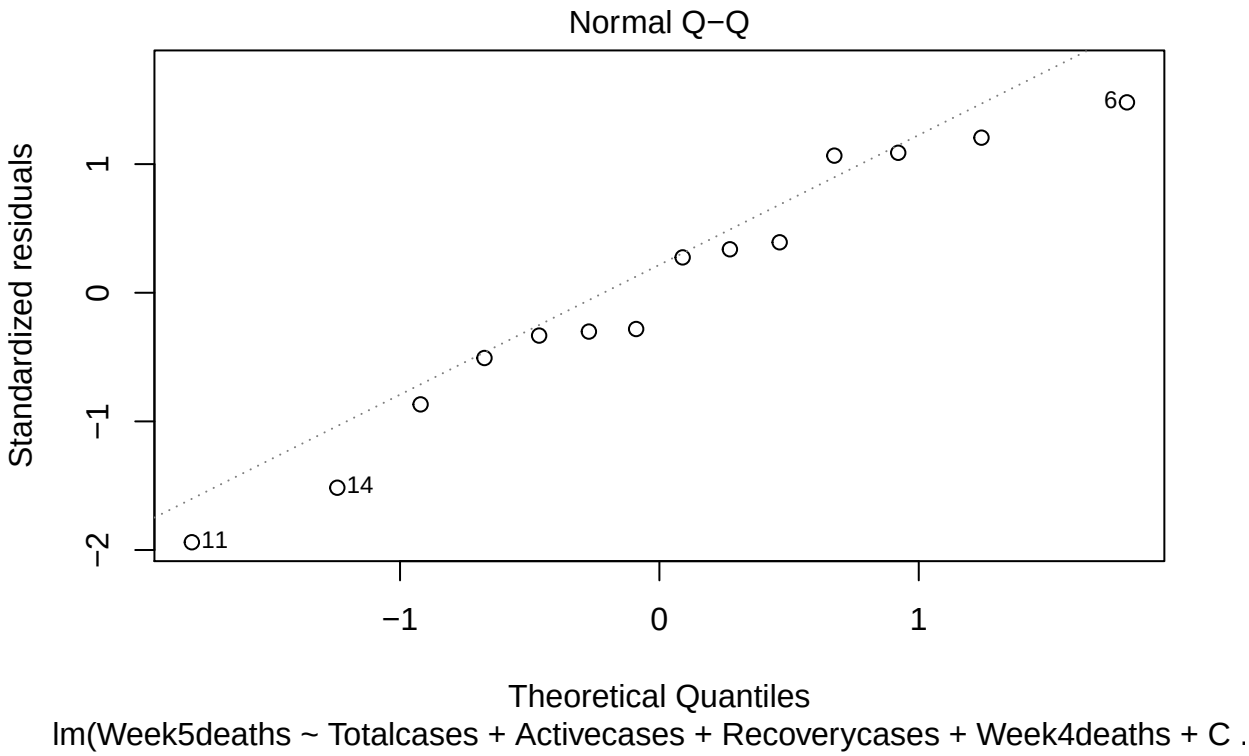


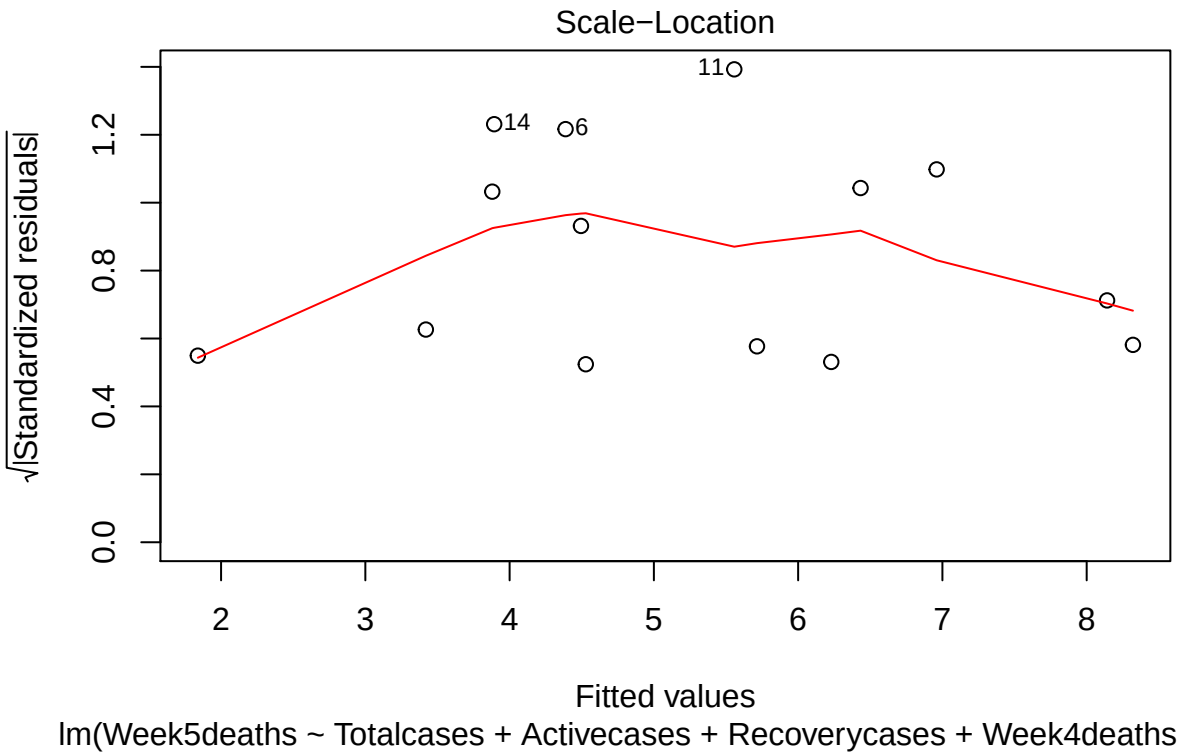

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

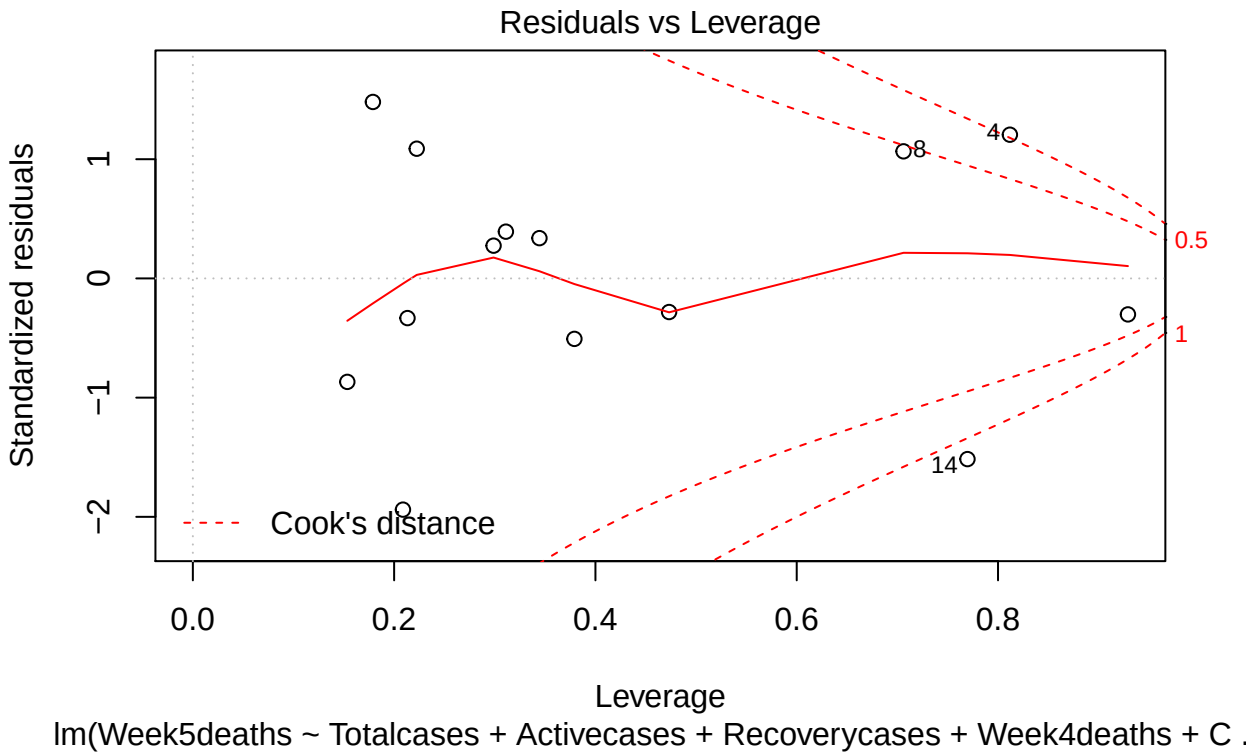

Residuals vs Leverage
Standardized residuals
3
2
1
0
−1
−2
0.0
0.2
0.4
0.6
0.8
1.0
Leverage
Cook's distance
0.5
1
14
1
2
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .

```
plot(fit2)
```


Residuals vs Fitted
Residuals
1.0
0.5
0.0
−0.5
−1.0
6
3
11
2
3
4
5
6
7
8
Fitted values
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .
6
14
11
−2
−1
0
1
−1
0
1


Scale−Location
11
14
6
√|Standardized residuals|
0.0
0.4
0.8
1.2
2
3
4
5
6
7
8
Fitted values
lm(Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + C .



prediction

Now, we use our new model to give a point estimate and a prediction interval for the number of deaths in India in week 5. We estimate the week5 death to be 93.87909 with 95% confidence interval to be [21.7043,406.0616] 95% prediction interval to be [12.76397,690.4813]

```
exp(predict(fit2,log(india),interval = "confidence"))
```

```
##         fit     lwr      upr
## 16 93.87909 21.7043 406.0616
```

```
exp(predict(fit2,log(india),interval = "prediction"))
```

```
##         fit      lwr      upr
## 16 93.87909 12.76397 690.4813
```

The estimation using the second model differs from the first one. Scenerio 3

Scenerio 4 Then we use model selection to check whether

Model Selection

```
set.seed(2020)
k <- 5
n <- 14
val.size <- floor(n/k)
folds_i <- sample(rep(1:k, length = n))
cv.mse <- rep(0, k)
ind.remain=1:n
for (round in 1:k){
  #val.ind <- sample(ind.remain, val.size, replace = FALSE)
  val.ind <- which(folds_i == round)
  fit<- lm(Week5deaths~Totalcases+Activecases+Recoverycases+Week4deaths+CFR, data = logdata[-val.ind,])
  y.hat <- predict(fit, logdata[val.ind,-7])
  cv.mse[round] <- mean((logdata[val.ind,7] - y.hat)^2)
}
mean(cv.mse)
```

```
## [1] 1.916331
```

```
library(MASS)
library(datasets)
library(olsrr)
```

```
##
## Attaching package: 'olsrr'

## The following object is masked from 'package:MASS':
##
##     cement

## The following object is masked from 'package:datasets':
##
##     rivers
```

```
ols_step_backward_aic(fit, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Totalcases
## 2 . Activecases
## 3 . Recoverycases
## 4 . Week4deaths
## 5 . CFR
##
##  Step 0: AIC = 20.14657
##  Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + CFR
```

```
##
## ---------------------------------------------------------------------
## Variable          DF     AIC      Sum Sq     RSS     R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------
## Recoverycases     1     18.147     0.000    1.173    0.974        0.958
## Totalcases        1     20.625     0.269    1.442    0.967        0.949
## Week4deaths       1     20.651     0.272    1.445    0.967        0.949
## CFR               1     20.673     0.275    1.447    0.967        0.949
## Activecases       1     20.775     0.287    1.460    0.967        0.948
## ---------------------------------------------------------------------
##
##
## Variables Removed:
##
## - Recoverycases
##
##
##   Step 1 : AIC = 18.14664
##  Week5deaths ~ Totalcases + Activecases + Week4deaths + CFR
##
## ---------------------------------------------------------------------
## Variable        DF     AIC      Sum Sq     RSS     R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------
## Totalcases      1     18.725     0.281    1.454    0.967        0.955
## Week4deaths     1     18.759     0.285    1.458    0.967        0.955
## CFR             1     18.781     0.288    1.461    0.967        0.955
## Activecases     1     19.617     0.393    1.566    0.965        0.951
## ---------------------------------------------------------------------
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - Recoverycases
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ---------------------------------------------------------------
## R                        0.987       RMSE                0.409
## R-Squared                0.974       Coef. Var           7.529
## Adj. R-Squared           0.958       MSE                 0.168
## Pred R-Squared           0.784       MAE                 0.253
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
```

```
##                Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression      43.099         4         10.775    64.319    0.0000
## Residual         1.173         7          0.168
## Total           44.271        11
## ----------------------------------------------------------------------
##
##                                    Parameter Estimates
## ------------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig         lower      upper
## ------------------------------------------------------------------------------------------------
## (Intercept)    -962.812       737.172                -1.306    0.233    -2705.947    780.323
##  Totalcases     207.643       160.316      196.052     1.295    0.236     -171.444    586.729
## Activecases       2.019         1.318        1.841     1.532    0.169       -1.097      5.134
## Week4deaths    -208.877       160.100     -259.588    -1.305    0.233     -587.453    169.699
##         CFR     209.854       160.084      122.672     1.311    0.231     -168.685    588.393
## ------------------------------------------------------------------------------------------------

##
##
##                      Backward Elimination Summary
## ---------------------------------------------------------------------
## Variable          AIC       RSS      Sum Sq      R-Sq       Adj. R-Sq
## ---------------------------------------------------------------------
## Full Model       20.147    1.173     43.099     0.97351       0.95144
## Recoverycases    18.147    1.173     43.099     0.97351       0.95838
## ---------------------------------------------------------------------
```

```
ols_step_backward_aic(fit2, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Totalcases
## 2 . Activecases
## 3 . Recoverycases
## 4 . Week4deaths
## 5 . CFR
##
##  Step 0: AIC = 31.01351
##  Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths + CFR
##
## ------------------------------------------------------------------------
## Variable         DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------
## Recoverycases     1     29.206      0.038     2.801     0.940         0.913
## Totalcases        1     29.515      0.101     2.864     0.939         0.911
## Week4deaths       1     29.521      0.102     2.865     0.939         0.911
## CFR               1     29.529      0.104     2.867     0.939         0.911
## Activecases       1     29.930      0.187     2.950     0.937         0.909
## ------------------------------------------------------------------------
```

```
##
##
## Variables Removed:
##
## - Recoverycases
##
##
##   Step 1 : AIC = 29.20559
##  Week5deaths ~ Totalcases + Activecases + Week4deaths + CFR
##
## ------------------------------------------------------------------------
## Variable       DF     AIC      Sum Sq     RSS     R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## Totalcases     1     27.838     0.130    2.931    0.937        0.918
## Week4deaths    1     27.851     0.132    2.934    0.937        0.918
## CFR            1     27.861     0.134    2.936    0.937        0.918
## Activecases    1     29.024     0.389    3.190    0.932        0.911
## ------------------------------------------------------------------------
##
## - Totalcases
##
##
##   Step 2 : AIC = 27.83843
##  Week5deaths ~ Activecases + Week4deaths + CFR
##
## ------------------------------------------------------------------------
## Variable       DF     AIC      Sum Sq     RSS     R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## Week4deaths    1     26.866     0.223    3.154    0.932        0.920
## Activecases    1     28.027     0.496    3.427    0.927        0.913
## CFR            1     28.432     0.597    3.528    0.924        0.911
## ------------------------------------------------------------------------
##
## - Week4deaths
##
##
##   Step 3 : AIC = 26.86551
##  Week5deaths ~ Activecases + CFR
##
## ------------------------------------------------------------------------
## Variable       DF     AIC      Sum Sq     RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## CFR            1     48.811    14.292    17.446    0.626        0.595
## Activecases    1     52.535    19.608    22.762    0.512        0.472
## ------------------------------------------------------------------------
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - Recoverycases
## - Totalcases
## - Week4deaths
```

```
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.966       RMSE                 0.535
## R-Squared               0.932       Coef. Var           10.158
## Adj. R-Squared          0.920       MSE                  0.287
## Pred R-Squared          0.861       MAE                  0.395
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression     43.524         2         21.762    75.895    0.0000
## Residual        3.154        11          0.287
## Total          46.679        13
## -------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    -0.729         0.687                -1.062    0.311    -2.241    0.782
## Activecases     0.730         0.088        0.665    8.269    0.000     0.536    0.925
##         CFR     0.980         0.139        0.568    7.060    0.000     0.674    1.285
## ----------------------------------------------------------------------------------------------


##
##
##                      Backward Elimination Summary
## ---------------------------------------------------------------
## Variable          AIC       RSS      Sum Sq      R-Sq      Adj. R-Sq
## ---------------------------------------------------------------
## Full Model       31.014    2.763    43.915    0.94080       0.90380
## Recoverycases    29.206    2.801    43.877    0.93998       0.91331
## Totalcases       27.838    2.931    43.748    0.93721       0.91837
## Week4deaths      26.866    3.154    43.524    0.93243       0.92014
## ---------------------------------------------------------------
```

```r
library(MASS)
step1 <- step(fit, direction="backward")
```

```
## Start:  AIC=-15.91
## Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths +
##     CFR
```

```
##
##                 Df Sum of Sq    RSS     AIC
## - Recoverycases  1  0.000006 1.1726 -17.908
## <none>                       1.1726 -15.908
## - Totalcases     1  0.269003 1.4416 -15.430
## - Week4deaths    1  0.272192 1.4448 -15.403
## - CFR            1  0.274755 1.4474 -15.382
## - Activecases    1  0.287181 1.4598 -15.279
##
## Step:  AIC=-17.91
## Week5deaths ~ Totalcases + Activecases + Week4deaths + CFR
##
##               Df Sum of Sq    RSS     AIC
## <none>                     1.1726 -17.908
## - Totalcases   1   0.28102 1.4537 -17.330
## - Week4deaths  1   0.28514 1.4578 -17.296
## - CFR          1   0.28787 1.4605 -17.273
## - Activecases  1   0.39321 1.5658 -16.438
```

```
step1$anova
```

```
##              Step Df     Deviance Resid. Df Resid. Dev       AIC
## 1                  NA           NA         6   1.172624 -15.90795
## 2 - Recoverycases  1 6.293813e-06         7   1.172630 -17.90789
```

```
library(MASS)
step2 <- step(fit2, direction="backward")
```

```
## Start:  AIC=-10.72
## Week5deaths ~ Totalcases + Activecases + Recoverycases + Week4deaths +
##     CFR
##
##                 Df Sum of Sq    RSS     AIC
## - Recoverycases  1  0.038175 2.8015 -12.525
## - Totalcases     1  0.100745 2.8641 -12.215
## - Week4deaths    1  0.101996 2.8653 -12.209
## - CFR            1  0.103658 2.8670 -12.201
## - Activecases    1  0.186875 2.9502 -11.801
## <none>                       2.7633 -10.717
##
## Step:  AIC=-12.52
## Week5deaths ~ Totalcases + Activecases + Week4deaths + CFR
##
##               Df Sum of Sq    RSS     AIC
## - Totalcases   1   0.12954 2.9310 -13.892
## - Week4deaths  1   0.13223 2.9337 -13.879
## - CFR          1   0.13424 2.9357 -13.869
## - Activecases  1   0.38865 3.1901 -12.706
## <none>                     2.8015 -12.525
##
## Step:  AIC=-13.89
## Week5deaths ~ Activecases + Week4deaths + CFR
##
```

```
##               Df Sum of Sq    RSS     AIC
## - Week4deaths  1   0.22311 3.1541 -14.865
## <none>                     2.9310 -13.892
## - Activecases  1   0.49588 3.4269 -13.704
## - CFR          1   0.59658 3.5276 -13.298
##
## Step:  AIC=-14.86
## Week5deaths ~ Activecases + CFR
##
##               Df Sum of Sq     RSS     AIC
## <none>                      3.1541 -14.865
## - CFR          1    14.292 17.4464   7.081
## - Activecases  1    19.608 22.7617  10.804
```

```
step2$anova
```

```
##              Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                  NA         NA         8   2.763315 -10.71677
## 2 - Recoverycases  1 0.03817462         9   2.801489 -12.52469
## 3    - Totalcases  1 0.12953997        10   2.931029 -13.89185
## 4   - Week4deaths  1 0.22311372        11   3.154143 -14.86477
```

```
fit3 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = data)#0.9701
#first iteration delete totalcases
fit3 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths + CFR, data = data)#0.9727
fit3 <- lm(Week5deaths~Totalcases + Recoverycases + Week4deaths + CFR, data = data)#0.9712
fit3 <- lm(Week5deaths~Totalcases + Activecases+Week4deaths + CFR, data = data)#0.9708
fit3 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + CFR, data = data)#0.8991
fit3 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + Week4deaths, data = data)#0.9715
summary(fit3)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Totalcases + Activecases + Recoverycases +
##     Week4deaths, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -152.54 -121.25  -74.57   27.82  556.34
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   142.60129   80.21370   1.778 0.105810
## Totalcases     -0.03777    0.20840  -0.181 0.859812
## Activecases     0.09066    0.14579   0.622 0.547943
## Recoverycases  -0.10868    0.10558  -1.029 0.327537
## Week4deaths     3.63501    0.66139   5.496 0.000263 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 228.5 on 10 degrees of freedom
## Multiple R-squared:  0.9796, Adjusted R-squared:  0.9715
## F-statistic: 120.2 on 4 and 10 DF,  p-value: 2.068e-08
```

```r
#second iteration (delete CFR
fit3 <- lm(Week5deaths~Recoverycases + Week4deaths + CFR, data = data)#0.9701
fit3 <- lm(Week5deaths~Activecases + Week4deaths + CFR, data = data)#0.945
fit3 <- lm(Week5deaths~Activecases +Recoverycases + CFR, data = data)#0.8389
fit3 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths, data = data)#0.974 use this
#third iteration delete
fit3 <- lm(Week5deaths~Recoverycases + Week4deaths, data = data)#0.9723
fit3 <- lm(Week5deaths~Activecases + Week4deaths, data = data)#0.9491
fit3 <- lm(Week5deaths~Activecases +Recoverycases, data = data)#0.7353
summary(fit3)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + Recoverycases, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -987.57 -324.07  -90.76   80.79 1743.97
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -84.9131   222.9610  -0.381 0.709976
## Activecases     0.4771     0.0969   4.923 0.000352 ***
## Recoverycases  -0.3749     0.0862  -4.349 0.000946 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 696.1 on 12 degrees of freedom
## Multiple R-squared:  0.7731, Adjusted R-squared:  0.7353
## F-statistic: 20.44 on 2 and 12 DF,  p-value: 0.0001364
```

```r
#decide to delete totalcases and CFR
fit.ms1 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths, data = data)
vif(fit.ms1)
```

```
##   Activecases Recoverycases   Week4deaths
##      155.9321      103.6383       12.1180
```

```r
summary(fit.ms1)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + Recoverycases + Week4deaths,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -156.87 -118.74  -76.23   35.33  553.20
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   138.20887   73.02541   1.893  0.08501 .
```

```
## Activecases     0.06596    0.04943   1.334  0.20904
## Recoverycases  -0.12657    0.03585  -3.531  0.00471 **
## Week4deaths     3.53342    0.33519  10.542 4.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 218.2 on 11 degrees of freedom
## Multiple R-squared:  0.9796, Adjusted R-squared:  0.974
## F-statistic: 175.8 on 3 and 11 DF,  p-value: 1.426e-09
```

```
#predict
predict(fit.ms1,india,interval = "confidence")
```

```
##         fit      lwr      upr
## 16 204.7712 63.99058 345.5518
```

```
predict(fit.ms1,india,interval = "prediction")
```

```
##         fit       lwr      upr
## 16 204.7712 -295.7113 705.2537
```

```
fit4 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + Week4deaths + CFR, data = logdata)#0.9

#first delete recoverycases
fit4 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths + CFR, data = logdata)#0.9114
fit4 <- lm(Week5deaths~Totalcases +Recoverycases + Week4deaths + CFR, data = logdata)#0.9078
fit4 <- lm(Week5deaths~Totalcases + Activecases + Week4deaths + CFR, data = logdata)#0.9133
fit4 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + CFR, data = logdata)#0.9113
fit4 <- lm(Week5deaths~Totalcases + Activecases +Recoverycases + Week4deaths,data = logdata)#0.9113
#second delete total
fit4 <- lm(Week5deaths~Activecases + Week4deaths + CFR, data = logdata)#0.9184
fit4 <- lm(Week5deaths~Totalcases + Week4deaths + CFR, data = logdata)#0.9112
fit4 <- lm(Week5deaths~Totalcases + Activecases + CFR, data = logdata)#0.9183
fit4 <- lm(Week5deaths~Totalcases + Activecases + Week4deaths, data = logdata)#0.9182
#Third delete week4deaths
fit4 <- lm(Week5deaths~ Week4deaths + CFR, data = logdata)#0.9132
fit4 <- lm(Week5deaths~Activecases  + CFR, data = logdata)#0.9201 use this
fit4 <- lm(Week5deaths~Activecases + Week4deaths, data = logdata)#0.9107
#fourth
fit4 <- lm(Week5deaths~CFR, data = logdata)#0.4717
fit4 <- lm(Week5deaths~Activecases, data = logdata)#0.5951
summary(fit4)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases, data = logdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1691 -0.5950  0.3074  0.9356  1.4131
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4353     1.5300  -0.938 0.366697
## Activecases   0.8691     0.1938   4.484 0.000747 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.206 on 12 degrees of freedom
## Multiple R-squared:  0.6262, Adjusted R-squared:  0.5951
## F-statistic: 20.11 on 1 and 12 DF,  p-value: 0.0007471
```

```
#decide to delete recovery total week4
fit.ms2 <- lm(Week5deaths~Activecases  + CFR, data = logdata)
summary(fit.ms2)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + CFR, data = logdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04174 -0.34305  0.06642  0.32780  0.73924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.7293     0.6868  -1.062    0.311
## Activecases   0.7302     0.0883   8.269 4.76e-06 ***
## CFR           0.9796     0.1388   7.060 2.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5355 on 11 degrees of freedom
## Multiple R-squared:  0.9324, Adjusted R-squared:  0.9201
## F-statistic:  75.9 on 2 and 11 DF,  p-value: 3.662e-07
```

```
#predict
exp(predict(fit.ms2,log(india),interval = "confidence"))
```

```
##         fit     lwr      upr
## 16 79.35504 51.9564 121.2021
```

```
exp(predict(fit.ms2,log(india),interval = "prediction"))
```

```
##         fit      lwr      upr
## 16 79.35504 22.68168 277.6347
```

#check collinearity and choose model

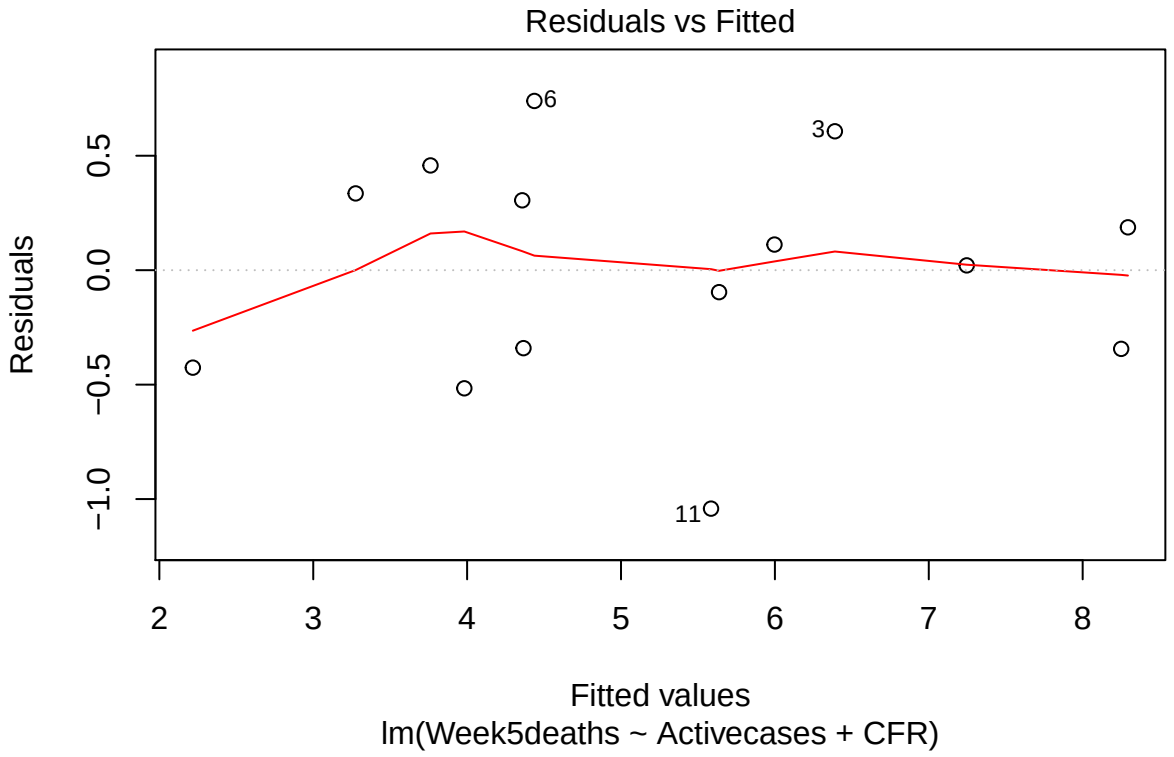
```
vif(fit.ms1) #too large
```

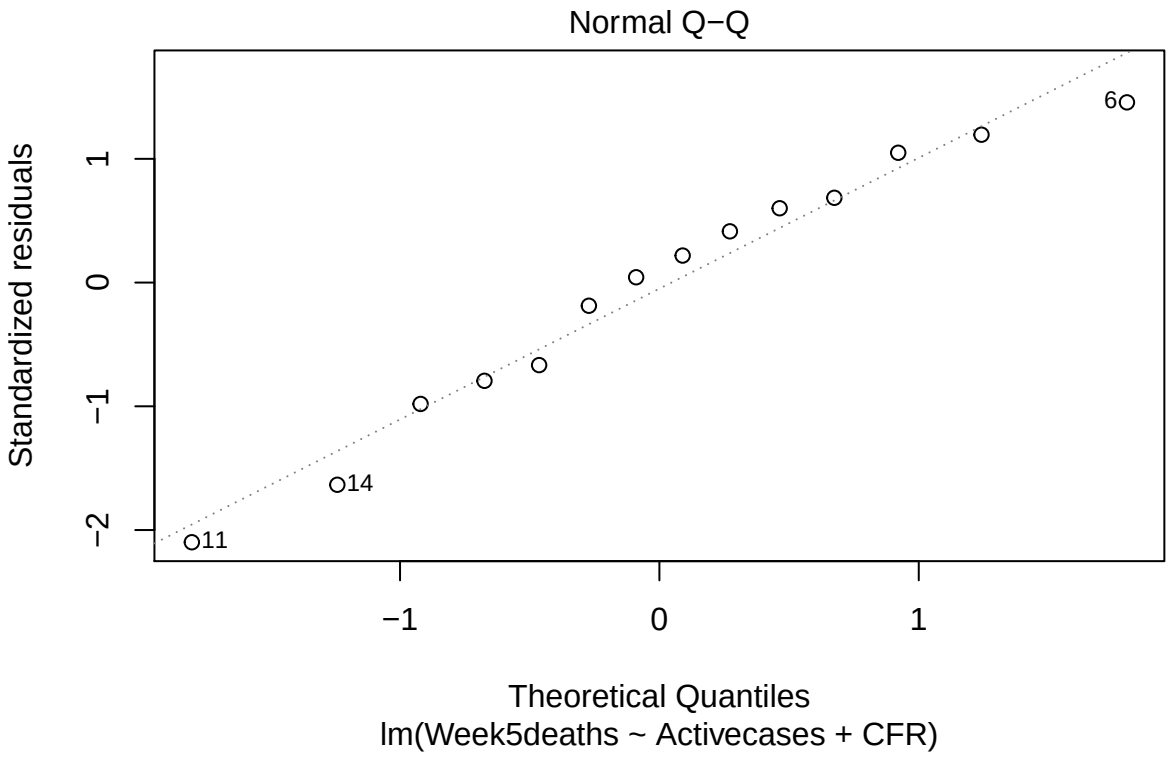
```
##   Activecases Recoverycases   Week4deaths
##      155.9321      103.6383       12.1180
```

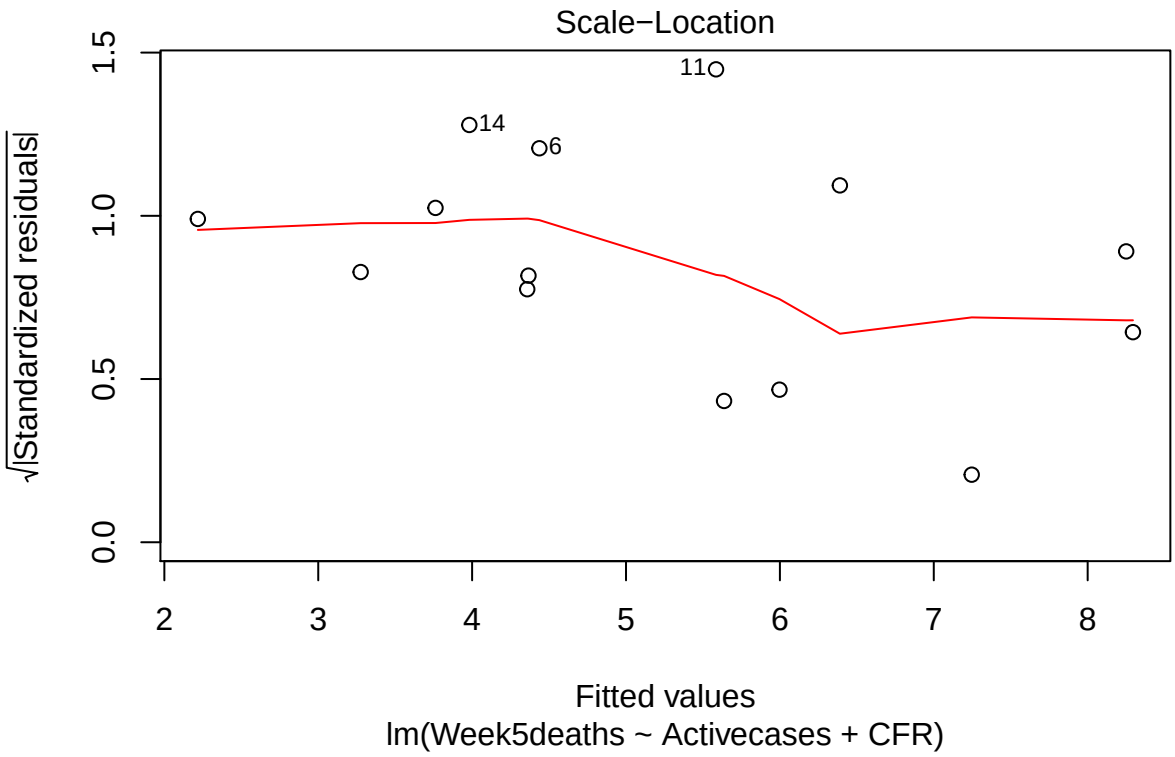
```r
vif(fit.ms2) #use this
```

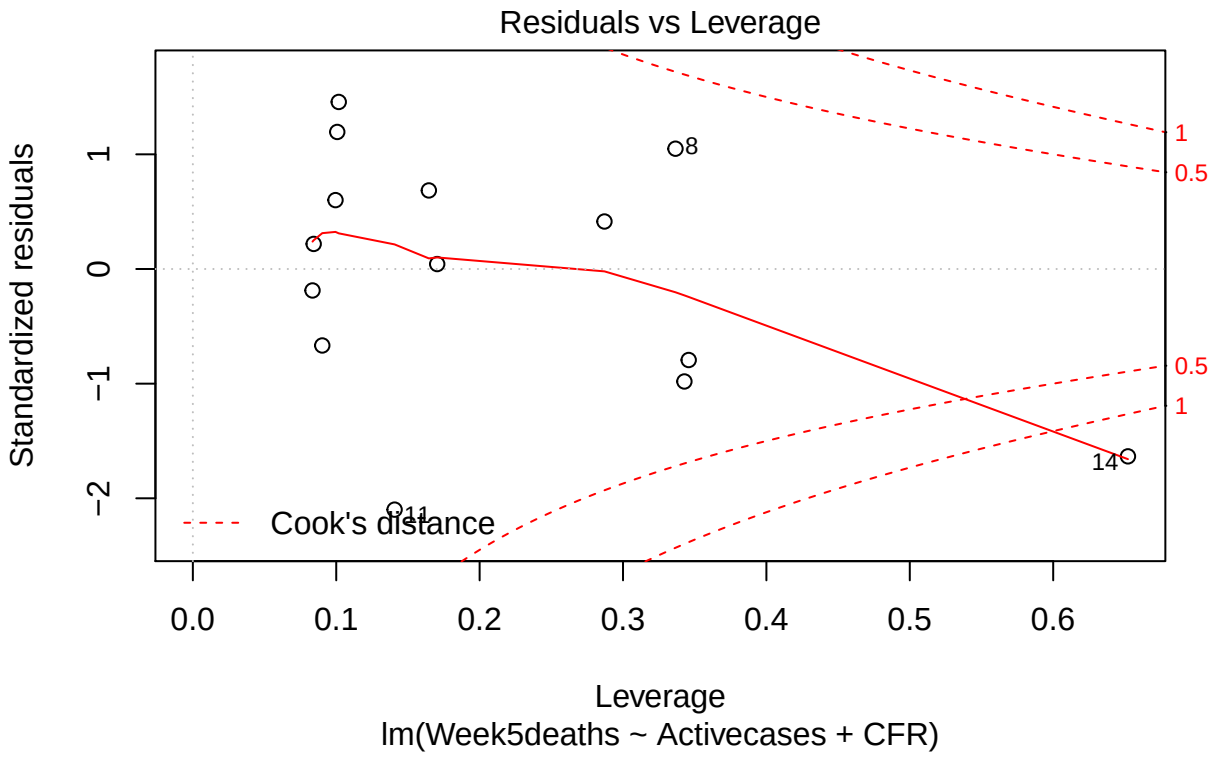
```
## Activecases         CFR
##    1.052277    1.052277
```

```r
plot(fit.ms2)
```




Normal Q−Q
Standardized residuals
−2
−1
0
1
11
14
6
−1
0
1
Theoretical Quantiles
lm(Week5deaths ~ Activecases + CFR)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
11
14
6
2
3
4
5
6
7
8
Fitted values
lm(Week5deaths ~ Activecases + CFR)




```
gdp<-read.csv('GDP.csv',header = TRUE)
gdp<-gdp[,c("Country.Name","X2018") ]
row.names(gdp) <-  1:nrow(gdp)
levels(gdp$Country.Name) <- c(levels(gdp$Country.Name), "UK")
gdp$Country.Name[gdp$Country.Name == 'United Kingdom'] <- 'UK'
levels(gdp$Country.Name) <- c(levels(gdp$Country.Name), "USA")
gdp$Country.Name[gdp$Country.Name == 'United States'] <- 'USA'
levels(gdp$Country.Name) <- c(levels(gdp$Country.Name), "Iran")
gdp$Country.Name[gdp$Country.Name == 'Iran, Islamic Rep.'] <- 'Iran'
levels(gdp$Country.Name) <- c(levels(gdp$Country.Name), "South Korea")
gdp$Country.Name[gdp$Country.Name == 'Korea, Rep.'] <- 'South Korea'
```

```
col=data_orgin$Countries
gdp<-gdp[gdp$Country.Name %in% col,]
library(plyr)
gdp<-rename(gdp, c("Country.Name"="Countries", "X2018"="GDP"))
```

```
total_origin <- merge(data_orgin,gdp,by=c("Countries"))
total_origin
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths   CFR
## 1      Austria        504         497        431.00           1 0.198
## 2      Belgium        559         555        139.00           3 0.537
## 3       Brazil        151         150        151.00           0 0.000
## 4        China      74185       57805      65112.00        2004 2.701
```

```
## 5       France       3661        3570        482.00          79 2.158
## 6      Germany       3675        3621       3130.00           8 0.218
## 7        India        606         554         42.00          10 1.650
## 8    Indonesia         69          60         38.00           4 5.797
## 9         Iran      11364        7321       9919.00         514 4.523
## 10       Italy      21157       17750      12207.00        1441 6.811
## 11 Netherlands        804         792        134.00          10 1.244
## 12 South Korea       7979        7198       7294.42          67 0.840
## 13       Spain       5232        4906       3097.00         133 2.542
## 14 Switzerland       1139        1124        303.00          11 0.966
## 15          UK        798         769        495.00          11 1.378
## 16         USA       2183        2126       1117.00          48 2.199
##    Week5deaths          GDP
## 1            6 4.552858e+11
## 2           37 5.427611e+11
## 3           11 1.885483e+12
## 4         2715 1.360815e+13
## 5          450 2.777535e+12
## 6           68 3.947620e+12
## 7           NA 2.718732e+12
## 8           32 1.042173e+12
## 9         1433           NA
## 10        4825 2.083864e+12
## 11         106 9.136585e+11
## 12          94 1.619424e+12
## 13        1093 1.419042e+12
## 14          56 7.051404e+11
## 15         177 2.855297e+12
## 16         255 2.054434e+13
```

```
india.new = total_origin[16,2:6]
india2.new = total_origin[16,c(3,4,5,8)]
india3.new= total_origin[16,c(3,4,5,6,8)]
total = total_origin[-7,]
total
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths   CFR
## 1      Austria        504         497        431.00           1 0.198
## 2      Belgium        559         555        139.00           3 0.537
## 3       Brazil        151         150        151.00           0 0.000
## 4        China      74185       57805      65112.00        2004 2.701
## 5       France       3661        3570        482.00          79 2.158
## 6      Germany       3675        3621       3130.00           8 0.218
## 8    Indonesia         69          60         38.00           4 5.797
## 9         Iran      11364        7321       9919.00         514 4.523
## 10       Italy      21157       17750      12207.00        1441 6.811
## 11 Netherlands        804         792        134.00          10 1.244
## 12 South Korea       7979        7198       7294.42          67 0.840
## 13       Spain       5232        4906       3097.00         133 2.542
## 14 Switzerland       1139        1124        303.00          11 0.966
## 15          UK        798         769        495.00          11 1.378
## 16         USA       2183        2126       1117.00          48 2.199
##    Week5deaths          GDP
## 1            6 4.552858e+11
```

```
## 2           37 5.427611e+11
## 3           11 1.885483e+12
## 4         2715 1.360815e+13
## 5          450 2.777535e+12
## 6           68 3.947620e+12
## 8           32 1.042173e+12
## 9         1433           NA
## 10        4825 2.083864e+12
## 11         106 9.136585e+11
## 12          94 1.619424e+12
## 13        1093 1.419042e+12
## 14          56 7.051404e+11
## 15         177 2.855297e+12
## 16         255 2.054434e+13
```

```
fit.new1 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths+GDP, data = total)
summary(fit.new1)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + Recoverycases + Week4deaths +
##     GDP, data = total)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -326.84  -55.58   -9.94   22.99  448.18
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.230e+01  9.120e+01   0.354  0.73133
## Activecases    1.756e-01  6.957e-02   2.524  0.03256 *
## Recoverycases -2.014e-01  4.893e-02  -4.116  0.00261 **
## Week4deaths    2.863e+00  4.449e-01   6.435  0.00012 ***
## GDP           -4.090e-12  1.076e-11  -0.380  0.71257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 198.6 on 9 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9796
## F-statistic: 157.2 on 4 and 9 DF,  p-value: 2.562e-08
```

```
predict(fit.new1,india2.new,interval = "confidence")
```

```
##         fit       lwr      upr
## 16 233.9757 -208.1301 676.0814
```

```
predict(fit.new1,india2.new,interval = "prediction")
```

```
##         fit       lwr      upr
## 16 233.9757 -396.2606 864.2119
```

```
elder<-read.csv('Elder Percentage.csv',header = TRUE)
```

```
elder<-elder[,c("Country.Name","X2018") ]
row.names(elder) <-  1:nrow(elder)
levels(elder$Country.Name) <- c(levels(elder$Country.Name), "UK")
elder$Country.Name[elder$Country.Name == 'United Kingdom'] <- 'UK'
levels(elder$Country.Name) <- c(levels(elder$Country.Name), "USA")
elder$Country.Name[elder$Country.Name == 'United States'] <- 'USA'
levels(elder$Country.Name) <- c(levels(elder$Country.Name), "Iran")
elder$Country.Name[elder$Country.Name == 'Iran, Islamic Rep.'] <- 'Iran'
levels(elder$Country.Name) <- c(levels(elder$Country.Name), "South Korea")
elder$Country.Name[elder$Country.Name == 'Korea, Rep.'] <- 'South Korea'
```

```
elder<-elder[elder$Country.Name %in% col,]
library(plyr)
elder<-rename(elder, c("Country.Name"="Countries", "X2018"="elder_percentage"))
elder
```

```
##       Countries elder_percentage
## 13      Austria        19.001566
## 16      Belgium        18.788744
## 28       Brazil         8.922838
## 36  Switzerland        18.623217
## 39        China        10.920884
## 54      Germany        21.461962
## 69        Spain        19.378508
## 76       France        20.034625
## 80           UK        18.395866
## 105   Indonesia         5.857166
## 108       India         6.179956
## 111        Iran         6.184574
## 115       Italy        22.751680
## 125 South Korea        14.418556
## 175 Netherlands        19.196193
## 250         USA        15.807654
```

```
total_origin <- merge(total_origin,elder,by=c("Countries"))
total_origin
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths   CFR
## 1      Austria        504         497        431.00           1 0.198
## 2      Belgium        559         555        139.00           3 0.537
## 3       Brazil        151         150        151.00           0 0.000
## 4        China      74185       57805      65112.00        2004 2.701
## 5       France       3661        3570        482.00          79 2.158
## 6      Germany       3675        3621       3130.00           8 0.218
## 7        India        606         554         42.00          10 1.650
## 8    Indonesia         69          60         38.00           4 5.797
## 9         Iran      11364        7321       9919.00         514 4.523
## 10       Italy      21157       17750      12207.00        1441 6.811
## 11 Netherlands        804         792        134.00          10 1.244
## 12 South Korea       7979        7198       7294.42          67 0.840
```

```
## 13       Spain       5232        4906       3097.00         133 2.542
## 14 Switzerland       1139        1124        303.00          11 0.966
## 15          UK        798         769        495.00          11 1.378
## 16         USA       2183        2126       1117.00          48 2.199
##    Week5deaths          GDP elder_percentage
## 1            6 4.552858e+11        19.001566
## 2           37 5.427611e+11        18.788744
## 3           11 1.885483e+12         8.922838
## 4         2715 1.360815e+13        10.920884
## 5          450 2.777535e+12        20.034625
## 6           68 3.947620e+12        21.461962
## 7           NA 2.718732e+12         6.179956
## 8           32 1.042173e+12         5.857166
## 9         1433           NA         6.184574
## 10        4825 2.083864e+12        22.751680
## 11         106 9.136585e+11        19.196193
## 12          94 1.619424e+12        14.418556
## 13        1093 1.419042e+12        19.378508
## 14          56 7.051404e+11        18.623217
## 15         177 2.855297e+12        18.395866
## 16         255 2.054434e+13        15.807654
```

```
india.new = total_origin[16,2:6]
india2.new = total_origin[16,c(3,4,5,6,8,9)]
india3.new= total_origin[16,c(3,4,5,6,8,9)]
total = total_origin[-7,]
total
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths   CFR
## 1      Austria        504         497        431.00           1 0.198
## 2      Belgium        559         555        139.00           3 0.537
## 3       Brazil        151         150        151.00           0 0.000
## 4        China      74185       57805      65112.00        2004 2.701
## 5       France       3661        3570        482.00          79 2.158
## 6      Germany       3675        3621       3130.00           8 0.218
## 8    Indonesia         69          60         38.00           4 5.797
## 9         Iran      11364        7321       9919.00         514 4.523
## 10       Italy      21157       17750      12207.00        1441 6.811
## 11 Netherlands        804         792        134.00          10 1.244
## 12 South Korea       7979        7198       7294.42          67 0.840
## 13       Spain       5232        4906       3097.00         133 2.542
## 14 Switzerland       1139        1124        303.00          11 0.966
## 15          UK        798         769        495.00          11 1.378
## 16         USA       2183        2126       1117.00          48 2.199
##    Week5deaths          GDP elder_percentage
## 1            6 4.552858e+11        19.001566
## 2           37 5.427611e+11        18.788744
## 3           11 1.885483e+12         8.922838
## 4         2715 1.360815e+13        10.920884
## 5          450 2.777535e+12        20.034625
## 6           68 3.947620e+12        21.461962
## 8           32 1.042173e+12         5.857166
## 9         1433           NA         6.184574
## 10        4825 2.083864e+12        22.751680
```

```
## 11         106 9.136585e+11        19.196193
## 12          94 1.619424e+12        14.418556
## 13        1093 1.419042e+12        19.378508
## 14          56 7.051404e+11        18.623217
## 15         177 2.855297e+12        18.395866
## 16         255 2.054434e+13        15.807654
```

```
india2.new
```

```
##    Activecases Recoverycases Week4deaths   CFR          GDP elder_percentage
## 16        2126          1117          48 2.199 2.054434e+13         15.80765
```

```
fit.new2 <- lm(Week5deaths~Activecases +Recoverycases + Week4deaths+GDP+elder_percentage, data = total)
summary(fit.new2)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + Recoverycases + Week4deaths +
##     GDP + elder_percentage, data = total)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -326.70  -55.72   -9.51   22.80  448.21
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.141e+01  2.424e+02   0.130  0.90007
## Activecases       1.754e-01  8.519e-02   2.059  0.07347 .
## Recoverycases    -2.013e-01  6.199e-02  -3.247  0.01175 *
## Week4deaths       2.863e+00  4.945e-01   5.790  0.00041 ***
## GDP              -4.086e-12  1.146e-11  -0.357  0.73053
## elder_percentage  6.169e-02  1.540e+01   0.004  0.99690
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 210.6 on 8 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9771
## F-statistic: 111.8 on 5 and 8 DF,  p-value: 3.52e-07
```

```
predict(fit.new2,india2.new,interval = "confidence")
```

```
##         fit       lwr      upr
## 16 233.9626 -244.1089 712.0341
```

```
predict(fit.new2,india2.new,interval = "prediction")
```

```
##         fit       lwr      upr
## 16 233.9626 -447.5013 915.4266
```

final model from advance analysis

```
logtot <- data.frame(total['Countries'],log(total['Totalcases']),log(total['Activecases']),
                     log(total['Recoverycases']),log(total['Week4deaths']),log(total['CFR']),
                     log(total['Week5deaths']),log(total['GDP']),log(total['elder_percentage']))
logtot <- logtot[-3,]
logtot
```

```
##      Countries Totalcases Activecases Recoverycases Week4deaths         CFR
## 1      Austria   6.222576    6.208590      6.066108    0.000000 -1.61948825
## 2      Belgium   6.326149    6.318968      4.934474    1.098612 -0.62175718
## 4        China  11.214317   10.964831     11.083864    7.602900  0.99362207
## 5       France   8.205492    8.180321      6.177944    4.369448  0.76918187
## 6      Germany   8.209308    8.194506      8.048788    2.079442 -1.52326022
## 8    Indonesia   4.234107    4.094345      3.637586    1.386294  1.75734054
## 9         Iran   9.338206    8.898502      9.202207    6.242223  1.50917549
## 10       Italy   9.959726    9.784141      9.409765    7.273093  1.91853895
## 11 Netherlands   6.689599    6.674561      4.897840    2.302585  0.21833199
## 12 South Korea   8.984568    8.881558      8.894865    4.204693 -0.17435339
## 13       Spain   8.562549    8.498214      8.038189    4.890349  0.93295117
## 14 Switzerland   7.037906    7.024649      5.713733    2.397895 -0.03459144
## 15          UK   6.682109    6.645091      6.204558    2.397895  0.32063317
## 16         USA   7.688455    7.661998      7.018402    3.871201  0.78800271
##    Week5deaths      GDP elder_percentage
## 1     1.791759 26.84419         2.944521
## 2     3.610918 27.01994         2.933258
## 4     7.906547 30.24169         2.390677
## 5     6.109248 28.65259         2.997462
## 6     4.219508 29.00413         3.066282
## 8     3.465736 27.67233         1.767666
## 9     7.267525       NA         1.822058
## 10    8.481566 28.36525         3.124639
## 11    4.663439 27.54072         2.954712
## 12    4.543295 28.11309         2.668516
## 13    6.996681 27.98100         2.964165
## 14    4.025352 27.28166         2.924409
## 15    5.176150 28.68020         2.912126
## 16    5.541264 30.65361         2.760494
```

```
fitt <- lm(Week5deaths~Activecases  + CFR +GDP+ elder_percentage, data = logtot)
vif(fitt)
```

```
##      Activecases              CFR              GDP elder_percentage
##         1.991565         1.398128         1.726913         1.641337
```

```
summary(fitt)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + CFR + GDP + elder_percentage,
##     data = logtot)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55978 -0.28965  0.08152  0.31162  0.48991
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -7.6243     4.1859  -1.821 0.106018
## Activecases        0.5670     0.1028   5.513 0.000565 ***
## CFR                1.1441     0.1391   8.227 3.57e-05 ***
## GDP                0.1602     0.1461   1.096 0.304905
## elder_percentage   1.2697     0.4523   2.807 0.022935 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4445 on 8 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9627, Adjusted R-squared:  0.9441
## F-statistic: 51.64 on 4 and 8 DF,  p-value: 9.372e-06
```

```
fit.final <- lm(Week5deaths~Activecases  + CFR +elder_percentage, data = logtot)
vif(fit.final)
```

```
##      Activecases              CFR elder_percentage
##         1.150589         1.491733         1.445111
```

```
summary(fit.final)
```

```
##
## Call:
## lm(formula = Week5deaths ~ Activecases + CFR + elder_percentage,
##     data = logtot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86474 -0.22634 -0.08323  0.37005  0.57515
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.30233    1.05728  -2.178   0.0545 .
## Activecases       0.68515    0.08369   8.186 9.62e-06 ***
## CFR               1.12923    0.14975   7.541 1.97e-05 ***
## elder_percentage  0.68279    0.37094   1.841   0.0955 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4854 on 10 degrees of freedom
## Multiple R-squared:  0.9495, Adjusted R-squared:  0.9344
## F-statistic: 62.71 on 3 and 10 DF,  p-value: 8.677e-07
```

```
indiapre <- total_origin[7,c(2:9)]
indiapre
```

```
##   Totalcases Activecases Recoverycases Week4deaths  CFR Week5deaths
## 7        606         554            42          10 1.65          NA
##            GDP elder_percentage
## 7 2.718732e+12         6.179956
```

```
exp(predict(fit.final,log(indiapre),interval = "confidence"))
```

```
##        fit      lwr      upr
## 7 46.29107 21.66149 98.92504
```

```
exp(predict(fit.final,log(indiapre),interval = "prediction"))
```

```
##        fit      lwr      upr
## 7 46.29107 12.34759 173.5451
```

```
#so our final model is fit.final
```