Load the Package

```
library(readr)
library(MASS)
library(class)
library(caret)
```

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

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

```
library(e1071)
library(ISLR)
library(leaps)
library(glmnet)
```

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

```
## Loaded glmnet 4.1-1
```

```
library(readr)
```

Read the file

```
covid = read.csv('national2.csv')
covid = covid[,-c(1:3)]
head(covid)
```

```
##   deathDirection deathDirection1 deathDirection2 deathDirection3
## 1           Down            Down              Up            Down
## 2           Down              Up            Down              Up
## 3             Up            Down              Up              Up
## 4           Down              Up              Up              Up
## 5             Up              Up              Up            Down
## 6             Up              Up            Down            Down
##   deathDirection4 inIcuCumulative inIcuCurrently hospitalizedIncrease
## 1              Up           45475           8134                  726
## 2              Up           45453           8409                  503
## 3              Up           45373           8634                 2781
## 4            Down           45293           8970                 1530
## 5            Down           45214           9359                 2172
## 6            Down           45084           9465                 1871
##   hospitalizedCurrently hospitalizedCumulative negative negativeIncrease
## 1                 40199                 878613 74582825           131835
## 2                 41401                 877887 74450990           143835
## 3                 42541                 877384 74307155           271917
## 4                 44172                 874603 74035238           177957
## 5                 45462                 873073 73857281           267001
## 6                 46388                 870901 73590280           255779
##   onVentilatorCumulative onVentilatorCurrently positive positiveIncrease states
## 1                   4281                  2802 28756489            41835     56
## 2                   4280                  2811 28714654            60015     56
## 3                   4275                  2889 28654639            68787     56
## 4                   4267                  2973 28585852            65487     56
## 5                   4260                  3094 28520365            66836     56
## 6                   4257                  3169 28453529            54248     56
##   totalTestResults totalTestResultsIncrease
## 1        363825123                 1170059
## 2        362655064                 1430992
## 3        361224072                 1744417
## 4        359479655                 1590984
## 5        357888671                 1406795
## 6        356481876                 1343519
```

```
# Split the Train and Test
train.size = dim(covid)[1] / 2
train = sample(1:dim(covid)[1], train.size)
test = -train
C.train = covid[train, ]
C.test = covid[test, ]

# Convert to Matrix
ytrain=C.train$deathDirection
ytest=C.test$deathDirection
```

BIC

```
regfit=regsubsets(deathDirection~.,data=covid[train,],nvmax=15)

summary(regfit)
```

```
## Subset selection object
## Call: regsubsets.formula(deathDirection ~ ., data = covid[train, ],
##     nvmax = 15)
## 18 Variables  (and intercept)
##                           Forced in Forced out
## deathDirection1Up             FALSE      FALSE
## deathDirection2Up             FALSE      FALSE
## deathDirection3Up             FALSE      FALSE
## deathDirection4Up             FALSE      FALSE
## inIcuCumulative               FALSE      FALSE
## inIcuCurrently                FALSE      FALSE
## hospitalizedIncrease          FALSE      FALSE
## hospitalizedCurrently         FALSE      FALSE
## hospitalizedCumulative        FALSE      FALSE
## negative                      FALSE      FALSE
## negativeIncrease              FALSE      FALSE
## onVentilatorCumulative        FALSE      FALSE
## onVentilatorCurrently         FALSE      FALSE
## positive                      FALSE      FALSE
## positiveIncrease              FALSE      FALSE
## states                        FALSE      FALSE
## totalTestResults              FALSE      FALSE
## totalTestResultsIncrease      FALSE      FALSE
## 1 subsets of each size up to 15
## Selection Algorithm: exhaustive
##           deathDirection1Up deathDirection2Up deathDirection3Up
## 1  ( 1 )  " "               " "               " "
## 2  ( 1 )  " "               " "               "*"
## 3  ( 1 )  " "               " "               "*"
## 4  ( 1 )  " "               " "               "*"
## 5  ( 1 )  " "               " "               "*"
## 6  ( 1 )  " "               "*"               "*"
## 7  ( 1 )  " "               "*"               "*"
## 8  ( 1 )  " "               "*"               "*"
## 9  ( 1 )  " "               "*"               "*"
## 10  ( 1 ) " "               "*"               "*"
## 11  ( 1 ) " "               "*"               "*"
## 12  ( 1 ) "*"               "*"               "*"
## 13  ( 1 ) "*"               "*"               "*"
## 14  ( 1 ) "*"               "*"               "*"
## 15  ( 1 ) "*"               "*"               "*"
##           deathDirection4Up inIcuCumulative inIcuCurrently hospitalizedIncrease
## 1  ( 1 )  " "               " "             " "            "*"
## 2  ( 1 )  " "               " "             " "            "*"
## 3  ( 1 )  "*"               " "             " "            " "
## 4  ( 1 )  " "               " "             " "            "*"
## 5  ( 1 )  "*"               " "             " "            "*"
## 6  ( 1 )  "*"               " "             " "            "*"
## 7  ( 1 )  "*"               " "             " "            "*"
## 8  ( 1 )  "*"               " "             " "            "*"
## 9  ( 1 )  "*"               "*"             " "            "*"
## 10  ( 1 ) "*"               "*"             " "            "*"
## 11  ( 1 ) "*"               "*"             " "            "*"
```

```
## 12  ( 1 ) "*"                 " "                 "*"                 "*"
## 13  ( 1 ) "*"                 " "                 "*"                 "*"
## 14  ( 1 ) "*"                 "*"                 "*"                 "*"
## 15  ( 1 ) "*"                 "*"                 "*"                 "*"
##           hospitalizedCurrently hospitalizedCumulative negative
## 1  ( 1 )  " "                   " "                     " "
## 2  ( 1 )  " "                   " "                     " "
## 3  ( 1 )  " "                   " "                     " "
## 4  ( 1 )  " "                   " "                     " "
## 5  ( 1 )  " "                   " "                     " "
## 6  ( 1 )  " "                   " "                     " "
## 7  ( 1 )  " "                   " "                     " "
## 8  ( 1 )  " "                   "*"                     " "
## 9  ( 1 )  " "                   " "                     " "
## 10  ( 1 ) "*"                   " "                     " "
## 11  ( 1 ) " "                   " "                     "*"
## 12  ( 1 ) "*"                   " "                     "*"
## 13  ( 1 ) "*"                   " "                     "*"
## 14  ( 1 ) "*"                   " "                     "*"
## 15  ( 1 ) "*"                   " "                     "*"
##           negativeIncrease onVentilatorCumulative onVentilatorCurrently
## 1  ( 1 )  " "              " "                    " "
## 2  ( 1 )  " "              " "                    " "
## 3  ( 1 )  " "              " "                    " "
## 4  ( 1 )  " "              " "                    "*"
## 5  ( 1 )  " "              " "                    "*"
## 6  ( 1 )  " "              " "                    "*"
## 7  ( 1 )  " "              " "                    "*"
## 8  ( 1 )  " "              " "                    "*"
## 9  ( 1 )  " "              " "                    "*"
## 10  ( 1 ) " "              " "                    "*"
## 11  ( 1 ) " "              "*"                    "*"
## 12  ( 1 ) " "              " "                    "*"
## 13  ( 1 ) "*"              " "                    "*"
## 14  ( 1 ) " "              "*"                    "*"
## 15  ( 1 ) "*"              "*"                    "*"
##           positive positiveIncrease states totalTestResults
## 1  ( 1 )  " "      " "              " "    " "
## 2  ( 1 )  " "      " "              " "    " "
## 3  ( 1 )  " "      " "              "*"    " "
## 4  ( 1 )  " "      " "              "*"    " "
## 5  ( 1 )  " "      " "              "*"    " "
## 6  ( 1 )  " "      " "              "*"    " "
## 7  ( 1 )  " "      "*"              "*"    " "
## 8  ( 1 )  " "      " "              "*"    "*"
## 9  ( 1 )  " "      "*"              "*"    "*"
## 10  ( 1 ) " "      "*"              "*"    "*"
## 11  ( 1 ) " "      "*"              "*"    "*"
## 12  ( 1 ) " "      "*"              "*"    "*"
## 13  ( 1 ) " "      "*"              "*"    "*"
## 14  ( 1 ) " "      "*"              "*"    "*"
## 15  ( 1 ) " "      "*"              "*"    "*"
##           totalTestResultsIncrease
## 1  ( 1 )  " "
```

```
## 2  ( 1 )  " "
## 3  ( 1 )  " "
## 4  ( 1 )  " "
## 5  ( 1 )  " "
## 6  ( 1 )  " "
## 7  ( 1 )  " "
## 8  ( 1 )  " "
## 9  ( 1 )  " "
## 10  ( 1 ) " "
## 11  ( 1 ) " "
## 12  ( 1 ) " "
## 13  ( 1 ) " "
## 14  ( 1 ) " "
## 15  ( 1 ) " "
```

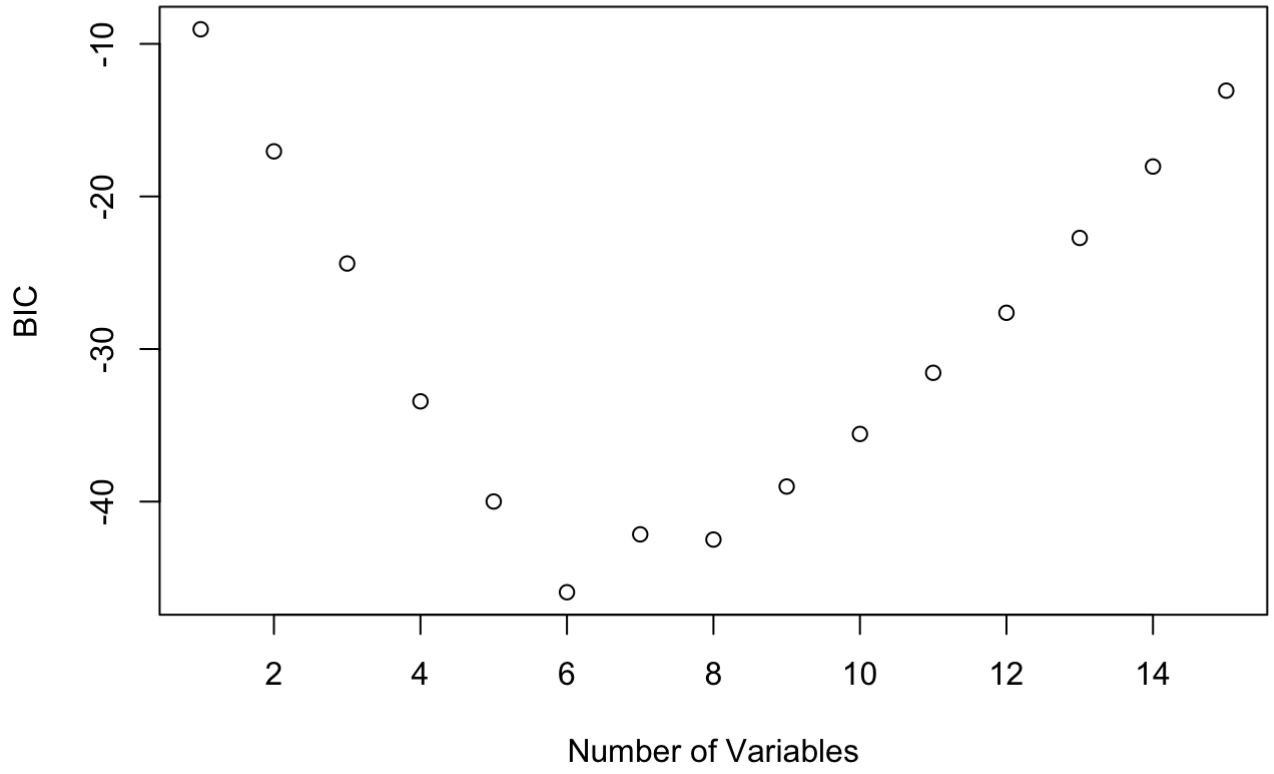
```
reg.summary=summary(regfit)
names(reg.summary)
```

```
## [1] "which"  "rsq"    "rss"    "adjr2"  "cp"     "bic"    "outmat" "obj"
```

```
reg.summary$bic
```

```
##  [1]  -9.043942 -17.043670 -24.393995 -33.428956 -39.993858 -45.939112
##  [7] -42.145885 -42.487193 -39.011632 -35.566080 -31.557814 -27.623418
## [13] -22.718699 -18.038414 -13.065490
```

```
plot(reg.summary$bic,xlab="Number of Variables",ylab="BIC") ### BIC indicates the best m
odel has 9 variables
```




```
best_n=which.min(reg.summary$bic) #find the model with the lowest BIC
best_n                            # Model with 9 regressors is the best
```

```
## [1] 6
```

```
coef(regfit,id=best_n)
```

```
##           (Intercept)     deathDirection2Up     deathDirection3Up
##          9.849361e-01         -2.146228e-01         -2.286028e-01
##     deathDirection4Up  hospitalizedIncrease onVentilatorCurrently
##         -2.700828e-01          1.477690e-04         -8.410839e-05
##                states
##          1.397689e-02
```

BIC-Logistic

```
# Fit a logistic regression
glm.fit=glm(C.train$deathDirection~C.train$deathDirection2+C.train$deathDirection3+C.tra
in$deathDirection4+C.train$hospitalizedIncrease+C.train$hospitalizedCumulative+C.train$o
nVentilatorCurrently+C.train$positiveIncrease+C.train$states+C.train$totalTestResults,fa
mily=binomial,data=C.train)

summary(glm.fit) #print regression output
```

```
##
## Call:
## glm(formula = C.train$deathDirection ~ C.train$deathDirection2 +
##     C.train$deathDirection3 + C.train$deathDirection4 + C.train$hospitalizedIncrease
+
##     C.train$hospitalizedCumulative + C.train$onVentilatorCurrently +
##     C.train$positiveIncrease + C.train$states + C.train$totalTestResults,
##     family = binomial, data = C.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1049  -0.7654  -0.2884   0.7413   2.2529
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -3.787e+00  1.268e+00  -2.987 0.002820 **
## C.train$deathDirection2Up     -1.607e+00  4.468e-01  -3.597 0.000322 ***
## C.train$deathDirection3Up     -1.338e+00  3.901e-01  -3.430 0.000604 ***
## C.train$deathDirection4Up     -1.626e+00  4.096e-01  -3.971 7.16e-05 ***
## C.train$hospitalizedIncrease   7.422e-04  2.016e-04   3.681 0.000232 ***
## C.train$hospitalizedCumulative -1.118e-05  4.219e-06  -2.650 0.008051 **
## C.train$onVentilatorCurrently  -5.889e-04  1.372e-04  -4.293 1.76e-05 ***
## C.train$positiveIncrease        6.567e-06  5.468e-06   1.201 0.229735
## C.train$states                  1.256e-01  2.893e-02   4.341 1.42e-05 ***
## C.train$totalTestResults        2.504e-08  9.437e-09   2.653 0.007971 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.92  on 209  degrees of freedom
## Residual deviance: 196.42  on 200  degrees of freedom
## AIC: 216.42
##
## Number of Fisher Scoring iterations: 5
```

```
# Prediction
glm.prob= predict(glm.fit, C.train, type = "response") #Probability of death direction g
oes up
glm.pred = rep("Down", length(glm.prob)) #Vector of length 316 and set default value for
each element as "Down"
glm.pred[glm.prob > 0.5] = "Up" #down>>up,if the predicted P(up) > 0.5

table(glm.pred, ytest) #Print the Confusion Matrix
```

```
##         ytest
## glm.pred Down Up
##     Down   73 55
##     Up     51 31
```

```
mean(glm.pred==ytest) #Calculate accuracy
```

```
## [1] 0.4952381
```

BIC-LDA

```
library(MASS)

lda.fit=lda(C.train$deathDirection~C.train$deathDirection2+C.train$deathDirection3+C.tra
in$deathDirection4+C.train$hospitalizedIncrease+C.train$hospitalizedCumulative+C.train$o
nVentilatorCurrently+C.train$positiveIncrease+C.train$states+C.train$totalTestResults,fa
mily=binomial,data=C.train) #fit linear discriminant analysis

lda.fit
```

```
## Call:
## lda(C.train$deathDirection ~ C.train$deathDirection2 + C.train$deathDirection3 +
##     C.train$deathDirection4 + C.train$hospitalizedIncrease +
##     C.train$hospitalizedCumulative + C.train$onVentilatorCurrently +
##     C.train$positiveIncrease + C.train$states + C.train$totalTestResults,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$deathDirection2Up C.train$deathDirection3Up
## Down                 0.4390244                 0.4959350
## Up                   0.4022989                 0.2873563
##      C.train$deathDirection4Up C.train$hospitalizedIncrease
## Down                 0.5203252                     1665.626
## Up                   0.3103448                     2609.897
##      C.train$hospitalizedCumulative C.train$onVentilatorCurrently
## Down                       325311.5                     3266.252
## Up                         381103.6                     3585.977
##      C.train$positiveIncrease C.train$states C.train$totalTestResults
## Down                 62607.56       47.95935                 97054780
## Up                   79366.01       55.31034                118704012
##
## Coefficients of linear discriminants:
##                                         LD1
## C.train$deathDirection2Up      -1.079242e+00
## C.train$deathDirection3Up      -1.028441e+00
## C.train$deathDirection4Up      -1.243213e+00
## C.train$hospitalizedIncrease    5.093514e-04
## C.train$hospitalizedCumulative -7.985851e-06
## C.train$onVentilatorCurrently  -4.068617e-04
## C.train$positiveIncrease        4.883638e-06
## C.train$states                  8.442079e-02
## C.train$totalTestResults        1.751097e-08
```

```
lda.pred=predict(lda.fit,C.train) #make predictions for trading days in 2005
lda.class=lda.pred$class #access prediction label
table(lda.class,ytest) #print confusion matrix
```

```
##          ytest
## lda.class Down Up
##      Down   71 53
##      Up     53 33
```

```
mean(lda.class==ytest) #calculate accuracy
```

```
## [1] 0.4952381
```

BIC-QDA

```
qda.fit=qda(C.train$deathDirection~C.train$deathDirection2+C.train$deathDirection3+C.tra
in$deathDirection4+C.train$hospitalizedIncrease+C.train$hospitalizedCumulative+C.train$o
nVentilatorCurrently+C.train$positiveIncrease+C.train$states+C.train$totalTestResults,fa
mily=binomial,data=C.train) #fit a quadratic discriminant analysis

qda.fit
```

```
## Call:
## qda(C.train$deathDirection ~ C.train$deathDirection2 + C.train$deathDirection3 +
##     C.train$deathDirection4 + C.train$hospitalizedIncrease +
##     C.train$hospitalizedCumulative + C.train$onVentilatorCurrently +
##     C.train$positiveIncrease + C.train$states + C.train$totalTestResults,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$deathDirection2Up C.train$deathDirection3Up
## Down                 0.4390244                 0.4959350
## Up                   0.4022989                 0.2873563
##      C.train$deathDirection4Up C.train$hospitalizedIncrease
## Down                 0.5203252                     1665.626
## Up                   0.3103448                     2609.897
##      C.train$hospitalizedCumulative C.train$onVentilatorCurrently
## Down                       325311.5                     3266.252
## Up                         381103.6                     3585.977
##      C.train$positiveIncrease C.train$states C.train$totalTestResults
## Down                 62607.56       47.95935                 97054780
## Up                   79366.01       55.31034                118704012
```

```
qda.pred=predict(qda.fit,C.train) #make predictions for trading days in 2005
qda.class=qda.pred$class #access prediction label
table(qda.class,ytest) #print confusion matrix
```

```
##          ytest
## qda.class Down Up
##      Down   56 42
##      Up     68 44
```

```
mean(qda.class==ytest) #calculate accuracy
```

```
## [1] 0.4761905
```

Stepwise

```
model_stepwise <- train(deathDirection ~ ., data =C.train,
                    method = 'glmStepAIC', direction = 'both',
                    trace = 0,
               trControl = trainControl(method = 'none', verboseIter = FALSE))

summary(model_stepwise) # This method gives 8 variables
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1946  -0.7402  -0.2681   0.7160   2.1994
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -3.868e+00  1.271e+00  -3.044 0.002338 **
## deathDirection1Up     -7.731e-01  4.541e-01  -1.703 0.088629 .
## deathDirection2Up     -1.736e+00  4.706e-01  -3.688 0.000226 ***
## deathDirection3Up     -1.615e+00  4.207e-01  -3.837 0.000124 ***
## deathDirection4Up     -1.873e+00  4.445e-01  -4.214 2.51e-05 ***
## inIcuCurrently         3.267e-04  2.082e-04   1.569 0.116553
## hospitalizedIncrease   9.159e-04  2.312e-04   3.961 7.45e-05 ***
## hospitalizedCurrently -7.739e-05  3.547e-05  -2.182 0.029121 *
## negative              -1.956e-07  7.041e-08  -2.778 0.005472 **
## onVentilatorCurrently -9.952e-04  3.527e-04  -2.822 0.004779 **
## positiveIncrease       2.545e-05  8.802e-06   2.892 0.003833 **
## states                 1.402e-01  2.995e-02   4.680 2.87e-06 ***
## totalTestResults       4.010e-08  1.408e-08   2.848 0.004402 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.92  on 209  degrees of freedom
## Residual deviance: 190.97  on 197  degrees of freedom
## AIC: 216.97
##
## Number of Fisher Scoring iterations: 5
```

Stepwise-Logistic

```
glm.fit=glm(C.train$deathDirection ~ C.train$deathDirection2+C.train$deathDirection3+C.t
rain$deathDirection4+C.train$inIcuCumulative+C.train$inIcuCurrently+C.train$positiveIncr
ease+C.train$states+C.train$totalTestResults,family=binomial,data=C.train)

summary(glm.fit) #print regression output
```

```
##
## Call:
## glm(formula = C.train$deathDirection ~ C.train$deathDirection2 +
##     C.train$deathDirection3 + C.train$deathDirection4 + C.train$inIcuCumulative +
##     C.train$inIcuCurrently + C.train$positiveIncrease + C.train$states +
##     C.train$totalTestResults, family = binomial, data = C.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0804  -0.8226  -0.2976   0.7879   2.4579
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -4.092e+00  1.380e+00  -2.965 0.003023 **
## C.train$deathDirection2Up -1.236e+00  4.004e-01  -3.088 0.002015 **
## C.train$deathDirection3Up -1.284e+00  3.682e-01  -3.486 0.000490 ***
## C.train$deathDirection4Up -1.673e+00  3.843e-01  -4.353 1.34e-05 ***
## C.train$inIcuCumulative   -3.148e-04  9.001e-05  -3.497 0.000470 ***
## C.train$inIcuCurrently    -1.996e-04  5.753e-05  -3.469 0.000523 ***
## C.train$positiveIncrease   2.292e-05  5.971e-06   3.839 0.000123 ***
## C.train$states             1.441e-01  2.995e-02   4.812 1.50e-06 ***
## C.train$totalTestResults   3.622e-08  1.082e-08   3.349 0.000811 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.92  on 209  degrees of freedom
## Residual deviance: 213.62  on 201  degrees of freedom
## AIC: 231.62
##
## Number of Fisher Scoring iterations: 6
```

```
# Prediction
glm.prob= predict(glm.fit, C.train, type = "response") #Probability of death direction g
oes up
glm.pred = rep("Down", length(glm.prob)) #Vector of length 316 and set default value for
each element as "Down"
glm.pred[glm.prob > 0.5] = "Up" #down>>up,if the predicted P(up) > 0.5

table(glm.pred, ytest) #Print the Confusion Matrix
```

```
##         ytest
## glm.pred Down Up
##     Down   75 56
##     Up     49 30
```

```
mean(glm.pred==ytest) #Calculate accuracy
```

```
## [1] 0.5
```

Stepwise-LDA

```
lda.fit=lda(C.train$deathDirection ~ C.train$deathDirection2+C.train$deathDirection3+C.t
rain$deathDirection4+C.train$inIcuCumulative+C.train$inIcuCurrently+C.train$positiveIncr
ease+C.train$states+C.train$totalTestResults,family=binomial,data=C.train) #fit a quadra
tic discriminant analysis

lda.fit
```

```
## Call:
## lda(C.train$deathDirection ~ C.train$deathDirection2 + C.train$deathDirection3 +
##     C.train$deathDirection4 + C.train$inIcuCumulative + C.train$inIcuCurrently +
##     C.train$positiveIncrease + C.train$states + C.train$totalTestResults,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$deathDirection2Up C.train$deathDirection3Up
## Down                 0.4390244                 0.4959350
## Up                   0.4022989                 0.2873563
##      C.train$deathDirection4Up C.train$inIcuCumulative C.train$inIcuCurrently
## Down                 0.5203252               16218.13               9279.561
## Up                   0.3103448               19167.03              10750.069
##      C.train$positiveIncrease C.train$states C.train$totalTestResults
## Down                 62607.56       47.95935                 97054780
## Up                   79366.01       55.31034                118704012
##
## Coefficients of linear discriminants:
##                                      LD1
## C.train$deathDirection2Up -1.024365e+00
## C.train$deathDirection3Up -1.138688e+00
## C.train$deathDirection4Up -1.440401e+00
## C.train$inIcuCumulative   -2.642730e-04
## C.train$inIcuCurrently    -1.724692e-04
## C.train$positiveIncrease   1.970696e-05
## C.train$states             1.065214e-01
## C.train$totalTestResults   3.036542e-08
```

```
lda.pred=predict(lda.fit,C.train) #make predictions for trading days in 2005
lda.class=lda.pred$class #access prediction label
table(lda.class,ytest) #print confusion matrix
```

```
##          ytest
## lda.class Down Up
##      Down   75 55
##      Up     49 31
```

```
mean(lda.class==ytest) #calculate accuracy
```

```
## [1] 0.5047619
```

Stepwise-QDA

```
qda.fit=qda(C.train$deathDirection ~ C.train$deathDirection2+C.train$deathDirection3+C.t
rain$deathDirection4+C.train$inIcuCumulative+C.train$inIcuCurrently+C.train$positiveIncr
ease+C.train$states+C.train$totalTestResults,family=binomial,data=C.train) #fit a quadra
tic discriminant analysis

qda.fit
```

```
## Call:
## qda(C.train$deathDirection ~ C.train$deathDirection2 + C.train$deathDirection3 +
##     C.train$deathDirection4 + C.train$inIcuCumulative + C.train$inIcuCurrently +
##     C.train$positiveIncrease + C.train$states + C.train$totalTestResults,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$deathDirection2Up C.train$deathDirection3Up
## Down                 0.4390244                 0.4959350
## Up                   0.4022989                 0.2873563
##      C.train$deathDirection4Up C.train$inIcuCumulative C.train$inIcuCurrently
## Down                 0.5203252                16218.13               9279.561
## Up                   0.3103448                19167.03              10750.069
##      C.train$positiveIncrease C.train$states C.train$totalTestResults
## Down                 62607.56       47.95935                 97054780
## Up                   79366.01       55.31034                118704012
```

```
qda.pred=predict(qda.fit,C.train) #make predictions for trading days in 2005
qda.class=qda.pred$class #access prediction label
table(qda.class,ytest) #print confusion matrix
```

```
##          ytest
## qda.class Down Up
##      Down   52 35
##      Up     72 51
```

```
mean(qda.class==ytest) #calculate accuracy
```

```
## [1] 0.4904762
```

Random Forest

```
### Random Forest with p/3 = 5 variables for regression model

### Note!!! : sqrt(p) for classification

library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

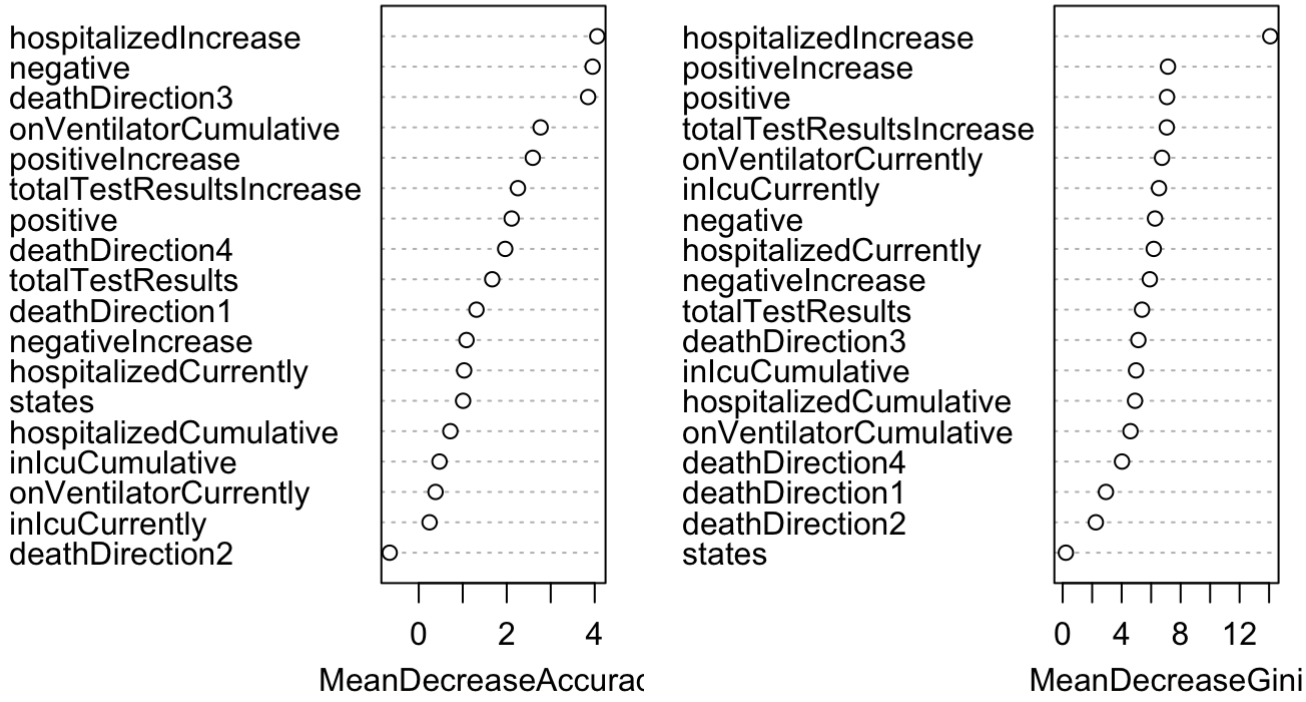
```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
rf.covid=randomForest(deathDirection~.,data=covid, subset=train,mtry=3,importance=TRUE,n
tree=25)
importance(rf.covid)
```

```
##                              Down           Up MeanDecreaseAccuracy
## deathDirection1           1.0702962  0.644928299             1.3119393
## deathDirection2          -0.4762113 -0.680507505            -0.6607643
## deathDirection3           2.3059043  3.228043682             3.8497285
## deathDirection4           2.2991566  0.622685101             1.9648802
## inIcuCumulative           0.9299590 -0.415780754             0.4719734
## inIcuCurrently            1.1002918 -0.774559811             0.2457038
## hospitalizedIncrease      2.6055179  2.974913062             4.0575441
## hospitalizedCurrently     1.6540920  0.052733152             1.0322264
## hospitalizedCumulative    2.0052773 -1.678958441             0.7182989
## negative                  3.6078538  0.654154540             3.9517759
## negativeIncrease         -0.0957177  1.389893385             1.0856522
## onVentilatorCumulative    1.8606764  1.687257285             2.7696237
## onVentilatorCurrently     0.4815791 -0.007325406             0.3819068
## positive                  3.0550911 -0.717081608             2.1120346
## positiveIncrease          2.6961794 -0.098475724             2.5924938
## states                    1.0206207  0.777585805             1.0083808
## totalTestResults          2.3747123 -1.877034634             1.6702190
## totalTestResultsIncrease  2.7476665 -0.952301593             2.2544159
##                          MeanDecreaseGini
## deathDirection1                 2.9345002
## deathDirection2                 2.2481404
## deathDirection3                 5.1334229
## deathDirection4                 4.0239005
## inIcuCumulative                 4.9758217
## inIcuCurrently                  6.5316956
## hospitalizedIncrease           14.0787532
## hospitalizedCurrently           6.1777055
## hospitalizedCumulative          4.9079845
## negative                        6.2626482
## negativeIncrease                5.9099438
## onVentilatorCumulative          4.5994572
## onVentilatorCurrently           6.7380125
## positive                        7.0780364
## positiveIncrease                7.1415265
## states                          0.2171258
## totalTestResults                5.3808919
## totalTestResultsIncrease        7.0631950
```

```
varImpPlot(rf.covid)
```

rf.covid




```
# Choose the first 7 variables in Ginis
# onVentilatorCurrently+totalTestResultsIncrease+positiveIncrease+negativeIncrease+hospi
talizedIncrease+totalTestResults+hospitalizedCumulative
```

Random Forest-Logistic

```
glm.fit=glm(C.train$deathDirection ~ C.train$onVentilatorCurrently+C.train$totalTestResu
ltsIncrease+C.train$positiveIncrease+C.train$negativeIncrease+C.train$hospitalizedIncrea
se+C.train$totalTestResults+C.train$hospitalizedCumulative,family=binomial,data=C.train)

summary(glm.fit) #print regression output
```

```
##
## Call:
## glm(formula = C.train$deathDirection ~ C.train$onVentilatorCurrently +
##     C.train$totalTestResultsIncrease + C.train$positiveIncrease +
##     C.train$negativeIncrease + C.train$hospitalizedIncrease +
##     C.train$totalTestResults + C.train$hospitalizedCumulative,
##     family = binomial, data = C.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2502  -0.9026  -0.7026   1.0937   2.0485
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -1.273e+00  4.118e-01  -3.091  0.00200 **
## C.train$onVentilatorCurrently   -3.400e-04  1.247e-04  -2.726  0.00642 **
## C.train$totalTestResultsIncrease -1.215e-06  1.264e-06  -0.961  0.33666
## C.train$positiveIncrease         1.385e-06  8.009e-06   0.173  0.86273
## C.train$negativeIncrease        -9.881e-08  1.909e-06  -0.052  0.95872
## C.train$hospitalizedIncrease     8.213e-04  1.819e-04   4.516 6.29e-06 ***
## C.train$totalTestResults        -6.641e-09  8.617e-09  -0.771  0.44090
## C.train$hospitalizedCumulative   5.874e-06  4.719e-06   1.245  0.21328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.92  on 209  degrees of freedom
## Residual deviance: 252.39  on 202  degrees of freedom
## AIC: 268.39
##
## Number of Fisher Scoring iterations: 4
```

```
# Prediction
glm.prob= predict(glm.fit, C.train, type = "response") #Probability of death direction g
oes up
glm.pred = rep("Down", length(glm.prob)) #Vector of length 316 and set default value for
each element as "Down"
glm.pred[glm.prob > 0.5] = "Up" #down>>up,if the predicted P(up) > 0.5

table(glm.pred, ytest) #Print the Confusion Matrix
```

```
##         ytest
## glm.pred Down Up
##     Down   87 59
##     Up     37 27
```

```
mean(glm.pred==ytest) #Calculate accuracy
```

```
## [1] 0.5428571
```

Random Forest-LDA

```
lda.fit=lda(C.train$deathDirection ~ C.train$onVentilatorCurrently+C.train$totalTestResu
ltsIncrease+C.train$positiveIncrease+C.train$negativeIncrease+C.train$hospitalizedIncrea
se+C.train$totalTestResults+C.train$hospitalizedCumulative,family=binomial,data=C.train)
#fit a quadratic discriminant analysis

lda.fit
```

```
## Call:
## lda(C.train$deathDirection ~ C.train$onVentilatorCurrently +
##     C.train$totalTestResultsIncrease + C.train$positiveIncrease +
##     C.train$negativeIncrease + C.train$hospitalizedIncrease +
##     C.train$totalTestResults + C.train$hospitalizedCumulative,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$onVentilatorCurrently C.train$totalTestResultsIncrease
## Down                       3266.252                          819656.2
## Up                         3585.977                          956889.9
##      C.train$positiveIncrease C.train$negativeIncrease
## Down                 62607.56                 171559.4
## Up                   79366.01                 188722.6
##      C.train$hospitalizedIncrease C.train$totalTestResults
## Down                     1665.626                 97054780
## Up                       2609.897                118704012
##      C.train$hospitalizedCumulative
## Down                       325311.5
## Up                         381103.6
##
## Coefficients of linear discriminants:
##                                             LD1
## C.train$onVentilatorCurrently    -3.876825e-04
## C.train$totalTestResultsIncrease -1.484660e-06
## C.train$positiveIncrease          2.291505e-06
## C.train$negativeIncrease         -2.299795e-07
## C.train$hospitalizedIncrease      9.397352e-04
## C.train$totalTestResults         -9.198560e-09
## C.train$hospitalizedCumulative    7.626309e-06
```

```
lda.pred=predict(lda.fit,C.train) #make predictions for trading days in 2005
lda.class=lda.pred$class #access prediction label
table(lda.class,ytest) #print confusion matrix
```

```
##          ytest
## lda.class Down Up
##      Down   88 60
##      Up     36 26
```

```
mean(lda.class==ytest) #calculate accuracy
```

```
## [1] 0.5428571
```

Random Forest-QDA

```
qda.fit=qda(C.train$deathDirection ~ C.train$onVentilatorCurrently+C.train$totalTestResu
ltsIncrease+C.train$positiveIncrease+C.train$negativeIncrease+C.train$hospitalizedIncrea
se+C.train$totalTestResults+C.train$hospitalizedCumulative,family=binomial,data=C.train)
#fit a quadratic discriminant analysis

qda.fit
```

```
## Call:
## qda(C.train$deathDirection ~ C.train$onVentilatorCurrently +
##     C.train$totalTestResultsIncrease + C.train$positiveIncrease +
##     C.train$negativeIncrease + C.train$hospitalizedIncrease +
##     C.train$totalTestResults + C.train$hospitalizedCumulative,
##     data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      C.train$onVentilatorCurrently C.train$totalTestResultsIncrease
## Down                       3266.252                          819656.2
## Up                         3585.977                          956889.9
##      C.train$positiveIncrease C.train$negativeIncrease
## Down                 62607.56                 171559.4
## Up                   79366.01                 188722.6
##      C.train$hospitalizedIncrease C.train$totalTestResults
## Down                     1665.626                 97054780
## Up                       2609.897                118704012
##      C.train$hospitalizedCumulative
## Down                       325311.5
## Up                         381103.6
```

```
qda.pred=predict(qda.fit,C.train) #make predictions for trading days in 2005
qda.class=qda.pred$class #access prediction label
table(qda.class,ytest) #print confusion matrix
```

```
##          ytest
## qda.class Down Up
##      Down   95 70
##      Up     29 16
```

```
mean(qda.class==ytest) #calculate accuracy
```

```
## [1] 0.5285714
```

Full data-Logistic

```
glm.fit=glm(C.train$deathDirection ~.,family=binomial,data=C.train)

summary(glm.fit) #print regression output
```

```
##
## Call:
## glm(formula = C.train$deathDirection ~ ., family = binomial,
##     data = C.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2295  -0.7001  -0.2676   0.6861   2.1917
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -3.886e+00  1.276e+00  -3.044 0.002334 **
## deathDirection1Up        -7.981e-01  4.671e-01  -1.709 0.087522 .
## deathDirection2Up        -1.850e+00  4.969e-01  -3.722 0.000197 ***
## deathDirection3Up        -1.657e+00  4.377e-01  -3.786 0.000153 ***
## deathDirection4Up        -1.923e+00  4.524e-01  -4.251 2.13e-05 ***
## inIcuCumulative          -3.283e-04  7.572e-04  -0.434 0.664576
## inIcuCurrently            2.702e-04  2.905e-04   0.930 0.352293
## hospitalizedIncrease      9.194e-04  2.485e-04   3.700 0.000215 ***
## hospitalizedCurrently    -5.668e-05  4.863e-05  -1.165 0.243863
## hospitalizedCumulative    1.819e-06  2.317e-05   0.079 0.937424
## negative                 -4.973e-07  4.068e-07  -1.222 0.221555
## negativeIncrease         -1.802e-06  2.880e-06  -0.626 0.531427
## onVentilatorCumulative    5.117e-03  6.333e-03   0.808 0.419161
## onVentilatorCurrently    -1.121e-03  5.139e-04  -2.182 0.029134 *
## positive                  1.077e-07  5.987e-07   0.180 0.857193
## positiveIncrease          2.168e-05  1.570e-05   1.381 0.167326
## states                    1.434e-01  3.113e-02   4.608 4.06e-06 ***
## totalTestResults          6.645e-08  7.215e-08   0.921 0.357054
## totalTestResultsIncrease  1.047e-06  1.745e-06   0.600 0.548638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.92  on 209  degrees of freedom
## Residual deviance: 189.41  on 191  degrees of freedom
## AIC: 227.41
##
## Number of Fisher Scoring iterations: 5
```

```
# Prediction
glm.prob= predict(glm.fit, C.train, type = "response") #Probability of death direction g
oes up
glm.pred = rep("Down", length(glm.prob)) #Vector of length 316 and set default value for
each element as "Down"
glm.pred[glm.prob > 0.5] = "Up" #down>>up,if the predicted P(up) > 0.5

table(glm.pred, ytest) #Print the Confusion Matrix
```

```
##         ytest
## glm.pred Down Up
##     Down   74 55
##     Up     50 31
```

```
mean(glm.pred==ytest) #Calculate accuracy
```

```
## [1] 0.5
```

Full data-LDA

```
lda.fit=lda(C.train$deathDirection ~.,family=binomial,data=C.train) #fit a quadratic dis
criminant analysis

lda.fit
```

```
## Call:
## lda(C.train$deathDirection ~ ., data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      deathDirection1Up deathDirection2Up deathDirection3Up deathDirection4Up
## Down         0.3089431         0.4390244         0.4959350         0.5203252
## Up           0.5402299         0.4022989         0.2873563         0.3103448
##      inIcuCumulative inIcuCurrently hospitalizedIncrease hospitalizedCurrently
## Down        16218.13       9279.561             1665.626               47061.11
## Up          19167.03      10750.069             2609.897               55060.40
##      hospitalizedCumulative negative negativeIncrease onVentilatorCumulative
## Down               325311.5 24512191         171559.4               1633.593
## Up                 381103.6 29736458         188722.6               1950.908
##      onVentilatorCurrently positive positiveIncrease   states totalTestResults
## Down              3266.252  7351195         62607.56 47.95935         97054780
## Up                3585.977  8898537         79366.01 55.31034        118704012
##      totalTestResultsIncrease
## Down                 819656.2
## Up                   956889.9
##
## Coefficients of linear discriminants:
##                                    LD1
## deathDirection1Up        -3.959585e-01
## deathDirection2Up        -1.146237e+00
## deathDirection3Up        -1.149903e+00
## deathDirection4Up        -1.348052e+00
## inIcuCumulative          -1.335668e-04
## inIcuCurrently            1.580236e-04
## hospitalizedIncrease      5.700596e-04
## hospitalizedCurrently    -3.091732e-05
## hospitalizedCumulative   -5.666372e-07
## negative                 -3.808591e-07
## negativeIncrease         -9.798032e-07
## onVentilatorCumulative    3.583729e-03
## onVentilatorCurrently    -7.282941e-04
## positive                  3.246727e-09
## positiveIncrease          1.468120e-05
## states                    9.060980e-02
## totalTestResults          5.224313e-08
## totalTestResultsIncrease  3.492887e-07
```

```
lda.pred=predict(lda.fit,C.train) #make predictions for trading days in 2005
lda.class=lda.pred$class #access prediction label
table(lda.class,ytest) #print confusion matrix
```

```
##          ytest
## lda.class Down Up
##      Down   72 53
##      Up     52 33
```

```
mean(lda.class==ytest) #calculate accuracy
```

```
## [1] 0.5
```

Full data-QDA

```
qda.fit=qda(C.train$deathDirection ~.,family=binomial,data=C.train) #fit a quadratic dis
criminant analysis

qda.fit
```

```
## Call:
## qda(C.train$deathDirection ~ ., data = C.train, family = binomial)
##
## Prior probabilities of groups:
##      Down        Up
## 0.5857143 0.4142857
##
## Group means:
##      deathDirection1Up deathDirection2Up deathDirection3Up deathDirection4Up
## Down         0.3089431         0.4390244         0.4959350         0.5203252
## Up           0.5402299         0.4022989         0.2873563         0.3103448
##      inIcuCumulative inIcuCurrently hospitalizedIncrease hospitalizedCurrently
## Down        16218.13       9279.561             1665.626              47061.11
## Up          19167.03      10750.069             2609.897              55060.40
##      hospitalizedCumulative negative negativeIncrease onVentilatorCumulative
## Down               325311.5 24512191         171559.4               1633.593
## Up                 381103.6 29736458         188722.6               1950.908
##      onVentilatorCurrently positive positiveIncrease   states totalTestResults
## Down              3266.252  7351195         62607.56 47.95935         97054780
## Up                3585.977  8898537         79366.01 55.31034        118704012
##      totalTestResultsIncrease
## Down                 819656.2
## Up                   956889.9
```

```
qda.pred=predict(qda.fit,C.train) #make predictions for trading days in 2005
qda.class=qda.pred$class #access prediction label
table(qda.class,ytest) #print confusion matrix
```

```
##          ytest
## qda.class Down Up
##      Down   62 44
##      Up     62 42
```

```
mean(qda.class==ytest) #calculate accuracy
```

```
## [1] 0.4952381
```