R Notebook

Code

Hide

```
data <- read.csv("2017_KOTRA_FINAL.csv")
```

```
Warning messages:
1: In readChar(file, size, TRUE) : truncating string with embedded nuls
2: In readChar(file, size, TRUE) : truncating string with embedded nuls
```

Hide

```
data$amount <- gsub(",","",data$amount)
data$amount <- as.numeric(data$amount)
data$weight <- gsub(",","",data$weight)
data$weight <- as.numeric(data$weight)
data <- na.omit(data)
```

Hide

```
data[,"GDP1Y_Diff"] <- log(data$NY_GDP_MKTP_CD_1Y) - log(data$NY_GDP_MKTP_CD)
data[,"Distance_Diff"] <- log(data$KMDIST) - log(data$SNDIST)
data[,"I(TRADE_COUNTRYCD-TRADE_HSCD)"] <- (data$TRADE_COUNTRYCD - data$TRADE_HSCD)/1000000000
data[,"HSCDPercent"] <- (data$TRADE_HSCD_COUNTRYCD/data$TRADE_COUNTRYCD)*100
data[,"Inflation+Unemp_Rate"] <- data$Unemp_Rate + data$Inflation_rate
data <- subset(data,select=-c(COUNTRYNM,UNC_YEAR,HSCD,X171Qex,X172Qex,X173Qex,X174Qex,X171Qim,X172Qim,X173Qim,X174Qim,NY_GDP_MKTP_CD_1Y,NY_GDP_MKTP_CD,KMDIST,SNDIST,Unemp_Rate,Inflation_rate))
quantile(data$TARIFF_AVG)
```

```
  0%  25%  50%  75% 100% 
   0    0    0    5  515 
```

Hide

```
data$TARIFF_AVG  <- ifelse(data$TARIFF_AVG == 0,5,data$TARIFF_AVG)
quantile(data$amount)
```

```
          0%          25%          50%          75%         100% 
0.000000e+00 3.949050e+04 3.508950e+05 2.676368e+06 2.801843e+10 
```

Hide

```
data$amount <- ifelse(data$amount ==0,median(data$amount),data$amount)
quantile(data$weight)
```

```
          0%          25%          50%          75%         100% 
0.000000e+00 1.276156e+03 1.592313e+04 1.692065e+05 7.187643e+09 
```

Hide

```
data$weight <- ifelse(data$weight ==0,median(data$weight),data$weight)
data$KR_TRADE_HSCD_COUNTRYCD  <- ifelse(data$KR_TRADE_HSCD_COUNTRYCD == 0,median(data$KR_TRAD
E_HSCD_COUNTRYCD),data$KR_TRADE_HSCD_COUNTRYCD)
#############################################################################
data$TRADE_COUNTRYCD <- log(data$TRADE_COUNTRYCD)
data$TRADE_HSCD <- log(data$TRADE_HSCD)
data$TARIFF_AVG <- log1p(data$TARIFF_AVG)
data$SP_POP_TOTL <- log(data$SP_POP_TOTL)
data$PA_NUS_FCRF <- log1p(data$PA_NUS_FCRF)
data$TRADE_HSCD_COUNTRYCD <- log(data$TRADE_HSCD_COUNTRYCD)
data$ContainerPortTraffic <- log(data$ContainerPortTraffic)
data$AirTransportFreight <- log1p(data$AirTransportFreight)
data$KR_TRADE_HSCD_COUNTRYCD <- log(data$KR_TRADE_HSCD_COUNTRYCD)
data$amount <- log(data$amount)
data$weight <- log(data$weight)
data$KR_TRADE_HSCD_COUNTRYCD  <- ifelse(data$KR_TRADE_HSCD_COUNTRYCD == 0,mean(data$KR_TRADE_
HSCD_COUNTRYCD),data$KR_TRADE_HSCD_COUNTRYCD)
data$amount<- ifelse(data$amount ==0,mean(data$amount),data$amount)
data$weight <- ifelse(data$weight <=0,mean(data$weight),data$weight)
```

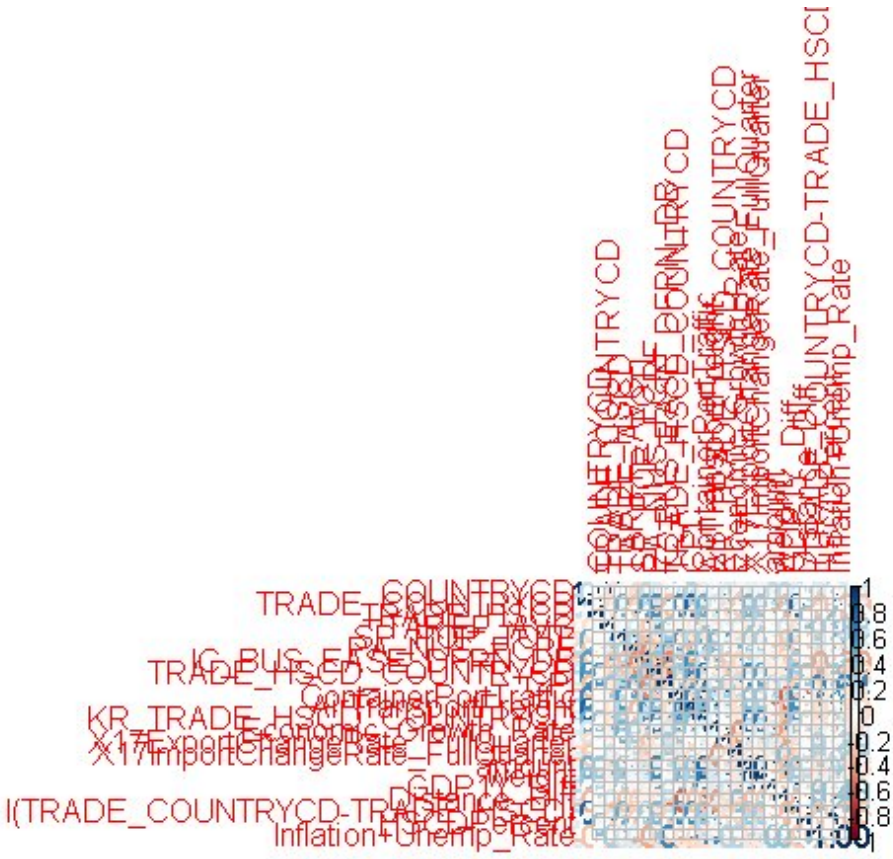
Hide

```
summary(data)
```

```
   COUNTRYCD     TRADE_COUNTRYCD   TRADE_HSCD      TARIFF_AVG       SP_POP_TOTL
 Min.   : 12.0   Min.   :23.63   Min.   :19.31   Min.   :0.09531   Min.   :15.22
 1st Qu.:320.0   1st Qu.:24.90   1st Qu.:21.65   1st Qu.:1.79176   1st Qu.:17.02
 Median :484.0   Median :26.08   Median :22.32   Median :1.79176   Median :17.86
 Mean   :479.7   Mean   :25.86   Mean   :22.43   Mean   :1.91122   Mean   :17.83
 3rd Qu.:704.0   3rd Qu.:26.76   3rd Qu.:23.21   3rd Qu.:1.79176   3rd Qu.:18.64
 Max.   :842.0   Max.   :28.51   Max.   :26.57   Max.   :6.24611   Max.   :21.05
  PA_NUS_FCRF      IC_BUS_EASE_DFRN_DB TRADE_HSCD_COUNTRYCD      CPI         ContainerPortTra
ffic
 Min.   : 0.2649   Min.   :42.67       Min.   : 8.156       Min.   : 98.27   Min.   :11.55
 1st Qu.: 1.4330   1st Qu.:59.82       1st Qu.:16.159       1st Qu.:114.94   1st Qu.:14.98
 Median : 2.1220   Median :69.38       Median :17.437       Median :125.42   Median :15.89
 Mean   : 3.2135   Mean   :68.82       Mean   :17.435       Mean   :138.80   Mean   :15.71
 3rd Qu.: 4.1915   3rd Qu.:78.02       3rd Qu.:18.696       3rd Qu.:147.09   3rd Qu.:16.37
 Max.   :10.4111   Max.   :87.17       Max.   :25.451       Max.   :333.67   Max.   :19.22
 AirTransportFreight KR_TRADE_HSCD_COUNTRYCD Economic_Growth_Rate X17ExportChangeRate_FullQua
rter
 Min.   : 0.4359     Min.   : 0.6931         Min.   :-5.000       Min.   :-0.1734
 1st Qu.: 6.0035     1st Qu.:11.3220         1st Qu.: 1.680       1st Qu.:-0.1648
 Median : 7.1981     Median :13.1060         Median : 3.000       Median : 0.2867
 Mean   : 7.1006     Mean   :13.0577         Mean   : 3.228       Mean   : 0.2078
 3rd Qu.: 8.4766     3rd Qu.:14.9603         3rd Qu.: 5.070       3rd Qu.: 0.5107
 Max.   :10.6357     Max.   :24.8774         Max.   : 7.000       Max.   : 0.6012
 X17ImportChangeRate_FullQuarter     amount           weight           GDP1Y_Diff
 Min.   :-0.04353                Min.   : 0.6931   Min.   : 0.05354   Min.   :-0.20422
 1st Qu.:-0.01590                1st Qu.:10.7877   1st Qu.: 7.47036   1st Qu.:-0.09456
 Median : 0.22629                Median :12.7690   Median : 9.67553   Median :-0.07551
 Mean   : 0.21024                Mean   :12.6680   Mean   : 9.70952   Mean   :-0.06314
 3rd Qu.: 0.33739                3rd Qu.:14.8000   3rd Qu.:12.03888   3rd Qu.:-0.04189
 Max.   : 0.60812                Max.   :24.0561   Max.   :22.69563   Max.   : 0.34130
 Distance_Diff       I(TRADE_COUNTRYCD-TRADE_HSCD)  HSCDPercent        Inflation+Unemp_Rate
 Min.   :-2.229211   Min.   :-32.862               Min.   : 0.000013   Min.   : 2.79
 1st Qu.:-0.396907   1st Qu.:  5.268               1st Qu.: 0.008364   1st Qu.: 6.00
 Median : 0.300014   Median : 19.287               Median : 0.022272   Median : 8.20
 Mean   : 0.001196   Mean   : 33.613               Mean   : 0.083619   Mean   :10.24
 3rd Qu.: 0.558652   3rd Qu.: 41.366               3rd Qu.: 0.058719   3rd Qu.: 9.70
 Max.   : 1.509294   Max.   :240.976               Max.   :10.782907   Max.   :35.54
```

Hide

```
library(corrplot)
Cor <- cor(data)
corrplot(Cor, method="number")
```



Hide

```
a <- lm(KR_TRADE_HSCD_COUNTRYCD ~ .,data=data)
summary(a)
```

```
Call:
lm(formula = KR_TRADE_HSCD_COUNTRYCD ~ ., data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.5876  -0.7513   0.1307   0.8767   7.0311

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -4.636e+00  8.227e-01  -5.635 1.78e-08 ***
COUNTRYCD                       3.633e-04  6.833e-05   5.317 1.07e-07 ***
TRADE_COUNTRYCD                 1.440e-02  3.356e-02   0.429 0.667868
TRADE_HSCD                     -1.366e-02  1.597e-02  -0.856 0.392255
TARIFF_AVG                     -2.209e-01  2.999e-02  -7.366 1.85e-13 ***
SP_POP_TOTL                     1.016e-01  2.264e-02   4.489 7.21e-06 ***
PA_NUS_FCRF                     5.770e-03  7.443e-03   0.775 0.438223
IC_BUS_EASE_DFRN_DB             4.052e-04  2.654e-03   0.153 0.878653
TRADE_HSCD_COUNTRYCD            4.975e-01  1.462e-02  34.030  < 2e-16 ***
CPI                            -2.461e-05  5.629e-04  -0.044 0.965128
ContainerPortTraffic            1.124e-01  2.016e-02   5.577 2.48e-08 ***
AirTransportFreight            -1.333e-02  1.524e-02  -0.875 0.381853
Economic_Growth_Rate           -2.879e-02  9.320e-03  -3.089 0.002010 **
X17ExportChangeRate_FullQuarter  3.048e-01  8.531e-02   3.573 0.000354 ***
X17ImportChangeRate_FullQuarter  2.858e-01  9.282e-02   3.079 0.002079 **
amount                          3.770e-01  9.749e-03  38.667  < 2e-16 ***
weight                          1.026e-01  7.961e-03  12.893  < 2e-16 ***
GDP1Y_Diff                     -6.515e-01  2.004e-01  -3.251 0.001152 **
Distance_Diff                  -1.389e-01  3.205e-02  -4.335 1.47e-05 ***
`I(TRADE_COUNTRYCD-TRADE_HSCD)` -1.944e-03  4.787e-04  -4.061 4.92e-05 ***
HSCDPercent                     1.130e-01  4.462e-02   2.532 0.011353 *
`Inflation+Unemp_Rate`         -1.094e-02  3.223e-03  -3.395 0.000688 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.58 on 14458 degrees of freedom
Multiple R-squared:  0.6921,    Adjusted R-squared:  0.6917
F-statistic:  1548 on 21 and 14458 DF,  p-value: < 2.2e-16
```

Hide

```
library(car)
vif(a)
```

```
                     COUNTRYCD                TRADE_COUNTRYCD                     TRADE_HS
CD
                      1.860457                       9.616653                      2.3933
86
                    TARIFF_AVG                    SP_POP_TOTL                     PA_NUS_FC
RF
                      1.324358                       5.913737                      2.5255
09
           IC_BUS_EASE_DFRN_DB            TRADE_HSCD_COUNTRYCD                          C
PI
                      5.137724                       4.719072                      3.3679
88
          ContainerPortTraffic             AirTransportFreight          Economic_Growth_Ra
te
                      4.362944                       5.642329                      3.3072
13
X17ExportChangeRate_FullQuarter X17ImportChangeRate_FullQuarter                       amou
nt
                      2.932478                       2.633597                      5.3676
13
                        weight                     GDP1Y_Diff                  Distance_Di
ff
                      4.335120                       1.741559                      3.4717
09
`I(TRADE_COUNTRYCD-TRADE_HSCD)`                    HSCDPercent         `Inflation+Unemp_Rat
e`
                      3.620062                       1.296843                      3.4994
58
```

Hide

```
outlierTest(a)
```

```
       rstudent unadjusted p-value Bonferroni p
16924 -6.716149         1.9356e-11   2.8027e-07
4064  -6.502790         8.1432e-11   1.1791e-06
5012  -6.249862         4.2234e-10   6.1155e-06
3243  -6.063447         1.3655e-09   1.9773e-05
20966 -6.024054         1.7423e-09   2.5228e-05
3624  -5.946576         2.8012e-09   4.0562e-05
8945  -5.625698         1.8819e-08   2.7249e-04
11228 -5.610901         2.0497e-08   2.9680e-04
11215 -5.603491         2.1391e-08   3.0975e-04
9585  -5.466515         4.6658e-08   6.7560e-04
```

Hide

```
data2 <- data[-c(16924,4064,5012,3243,20966,3624,8945,11228,11215,9585),]
```

Hide

```
reg <- lm(KR_TRADE_HSCD_COUNTRYCD ~ .,data=data2)
summary(reg)
```

```
Call:
lm(formula = KR_TRADE_HSCD_COUNTRYCD ~ ., data = data2)

Residuals:
     Min       1Q   Median       3Q      Max
-10.5864  -0.7522   0.1310   0.8761   7.0322

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -4.645e+00  8.230e-01  -5.644 1.69e-08 ***
COUNTRYCD                       3.666e-04  6.835e-05   5.364 8.29e-08 ***
TRADE_COUNTRYCD                 1.441e-02  3.357e-02   0.429 0.667616
TRADE_HSCD                     -1.334e-02  1.597e-02  -0.835 0.403817
TARIFF_AVG                     -2.200e-01  3.000e-02  -7.332 2.38e-13 ***
SP_POP_TOTL                     1.017e-01  2.264e-02   4.493 7.06e-06 ***
PA_NUS_FCRF                     5.728e-03  7.447e-03   0.769 0.441796
IC_BUS_EASE_DFRN_DB             4.270e-04  2.655e-03   0.161 0.872211
TRADE_HSCD_COUNTRYCD            4.970e-01  1.462e-02  33.990  < 2e-16 ***
CPI                            -2.721e-05  5.631e-04  -0.048 0.961454
ContainerPortTraffic            1.124e-01  2.016e-02   5.576 2.51e-08 ***
AirTransportFreight            -1.340e-02  1.524e-02  -0.879 0.379204
Economic_Growth_Rate           -2.876e-02  9.324e-03  -3.084 0.002044 **
X17ExportChangeRate_FullQuarter  3.045e-01  8.532e-02   3.569 0.000360 ***
X17ImportChangeRate_FullQuarter  2.883e-01  9.284e-02   3.105 0.001906 **
amount                          3.769e-01  9.750e-03  38.655  < 2e-16 ***
weight                          1.030e-01  7.964e-03  12.928  < 2e-16 ***
GDP1Y_Diff                     -6.523e-01  2.004e-01  -3.255 0.001137 **
Distance_Diff                  -1.395e-01  3.205e-02  -4.351 1.36e-05 ***
`I(TRADE_COUNTRYCD-TRADE_HSCD)` -1.943e-03  4.788e-04  -4.058 4.98e-05 ***
HSCDPercent                     1.132e-01  4.463e-02   2.536 0.011221 *
`Inflation+Unemp_Rate`         -1.094e-02  3.223e-03  -3.395 0.000687 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.58 on 14450 degrees of freedom
Multiple R-squared:  0.6922,    Adjusted R-squared:  0.6918
F-statistic:  1547 on 21 and 14450 DF,  p-value: < 2.2e-16
```

Hide

```
step.reg3 <- step(reg,direction = 'both',trace='FALSE')
summary(step.reg3)
```

```
Call:
lm(formula = KR_TRADE_HSCD_COUNTRYCD ~ COUNTRYCD + TARIFF_AVG +
    SP_POP_TOTL + TRADE_HSCD_COUNTRYCD + ContainerPortTraffic +
    Economic_Growth_Rate + X17ExportChangeRate_FullQuarter +
    X17ImportChangeRate_FullQuarter + amount + weight + GDP1Y_Diff +
    Distance_Diff + `I(TRADE_COUNTRYCD-TRADE_HSCD)` + HSCDPercent +
    `Inflation+Unemp_Rate`, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max
-10.6445  -0.7479   0.1313   0.8747   7.0607

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -4.519e+00  3.100e-01 -14.576  < 2e-16 ***
COUNTRYCD                        3.538e-04  6.078e-05   5.821 5.96e-09 ***
TARIFF_AVG                      -2.187e-01  2.893e-02  -7.561 4.24e-14 ***
SP_POP_TOTL                      1.100e-01  1.418e-02   7.758 9.22e-15 ***
TRADE_HSCD_COUNTRYCD             4.898e-01  1.013e-02  48.331  < 2e-16 ***
ContainerPortTraffic             1.048e-01  1.685e-02   6.219 5.13e-10 ***
Economic_Growth_Rate            -2.582e-02  8.490e-03  -3.041 0.002361 **
X17ExportChangeRate_FullQuarter  3.046e-01  7.267e-02   4.191 2.79e-05 ***
X17ImportChangeRate_FullQuarter  2.883e-01  8.465e-02   3.406 0.000661 ***
amount                           3.766e-01  9.628e-03  39.115  < 2e-16 ***
weight                           1.035e-01  7.910e-03  13.083  < 2e-16 ***
GDP1Y_Diff                      -6.387e-01  1.865e-01  -3.425 0.000617 ***
Distance_Diff                   -1.387e-01  2.997e-02  -4.628 3.73e-06 ***
`I(TRADE_COUNTRYCD-TRADE_HSCD)` -1.925e-03  4.267e-04  -4.512 6.47e-06 ***
HSCDPercent                      1.089e-01  4.371e-02   2.491 0.012745 *
`Inflation+Unemp_Rate`          -1.191e-02  2.984e-03  -3.992 6.59e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.58 on 14456 degrees of freedom
Multiple R-squared:  0.6922,    Adjusted R-squared:  0.6918
F-statistic:  2167 on 15 and 14456 DF,  p-value: < 2.2e-16
```

Hide

```
set.seed(156)
library(forecast)
n = nrow(data2)
i = 1:n
train_list <- sample(i,n*0.7)
valid_list <- setdiff(i,train_list)
train_df <- data2[train_list,]
valid_df <- data2[valid_list,]
reg <- lm(KR_TRADE_HSCD_COUNTRYCD ~ .,data=train_df)
step.reg3 <- step(reg,direction = 'both', trace = FALSE)
step.pred <- predict(step.reg3,valid_df)
accuracy(step.pred,valid_df$KR_TRADE_HSCD_COUNTRYCD)
```

```
                ME     RMSE     MAE       MPE    MAPE
Test set 0.03101841 1.561186 1.11875 -2.397746 10.7332
```

Hide

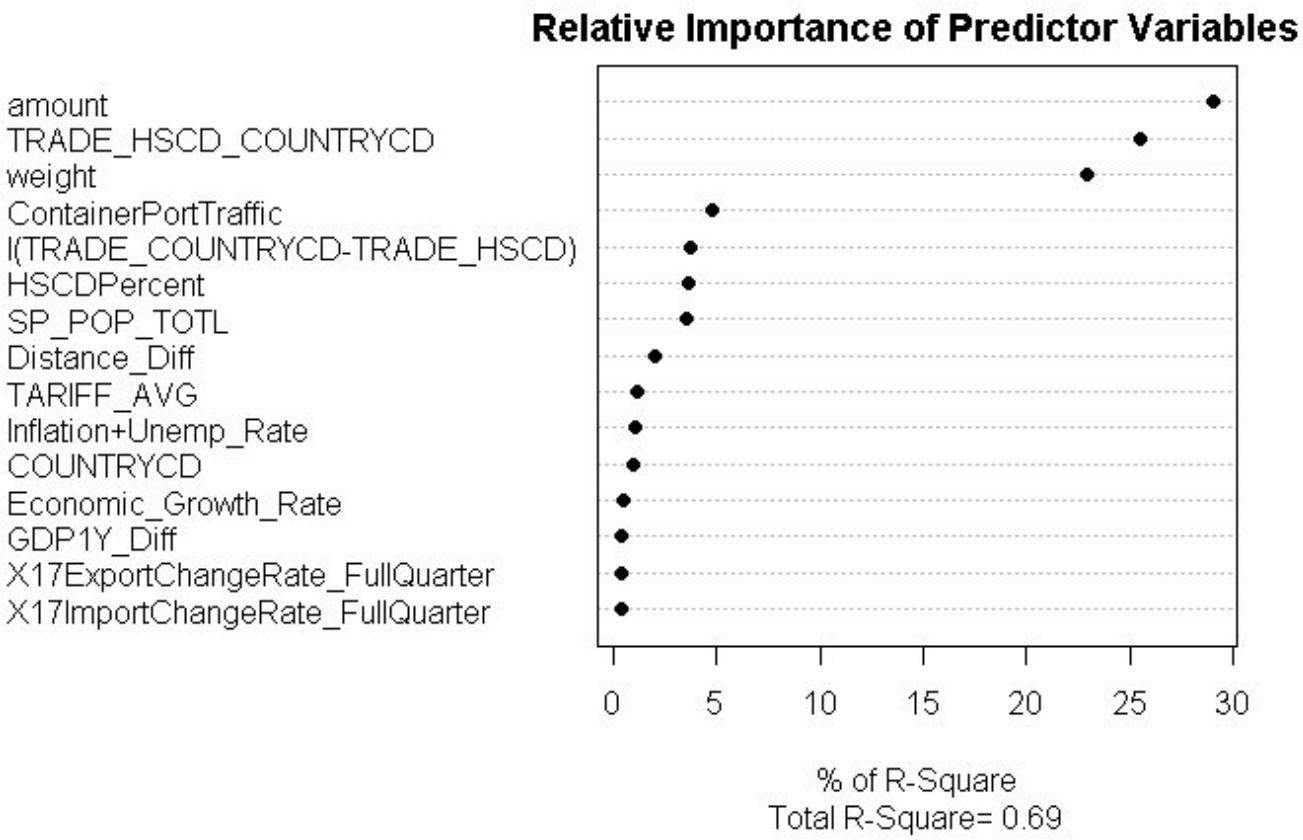
```
residuals <- step.pred[1:20] - valid_df$KR_TRADE_HSCD_COUNTRYCD[1:20]
data.frame("Predicted"=step.pred[1:20], "Actual"=valid_df$KR_TRADE_HSCD_COUNTRYCD[1:20], "Res
iduals" = residuals)
```

```
   Predicted    Actual   Residuals
1   9.176407  7.972121  1.20428620
2  14.560593 15.336205 -0.77561236
5  12.693466 12.519722  0.17374368
9  11.696177 10.748540  0.94763647
10 14.431929 14.444482 -0.01255269
22 13.481881 14.030317 -0.54843602
29 13.897581 13.341562  0.55601960
31 13.832337 12.390504  1.44183242
32 13.575845 12.130848  1.44499735
39 13.376975 14.308534 -0.93155930
50 12.663504 13.875044 -1.21153987
51 15.494209 15.752319 -0.25811008
53 12.155724  8.495765  3.65995891
70 11.074789 10.578751  0.49603782
71 15.178778 15.022027  0.15675104
80 12.829105 13.947579 -1.11847401
82 10.458744  9.539284  0.91945939
90 13.294349 11.718719  1.57563005
91 12.571263 13.355191 -0.78392805
95 16.853453 15.896400  0.95705293
```

Hide

```
relweights <- function(fit,...){
  R <- cor(fit$model)
  nvar <- ncol(R)
  rxx <- R[2:nvar, 2:nvar]
  rxy <- R[2:nvar, 1]
  svd <- eigen(rxx)
  evec <- svd$vectors
  ev <- svd$values
  delta <- diag(sqrt(ev))
  lambda <- evec %*% delta %*% t(evec)
  lambdasq <- lambda ^ 2
  beta <- solve(lambda) %*% rxy
  rsquare <- colSums(beta ^ 2)
  rawwgt <- lambdasq %*% beta ^ 2
  import <- (rawwgt / rsquare) * 100
  import <- as.data.frame(import)
  row.names(import) <- names(fit$model[2:nvar])
  names(import) <- "Weights"
  import <- import[order(import),1, drop=FALSE]
  dotchart(import$Weights, labels=row.names(import),
           xlab="% of R-Square", pch=19,
           main="Relative Importance of Predictor Variables",
           sub=paste("Total R-Square=", round(rsquare, digits=3)),
           ...)
  return(import)
}
relweights(step.reg3)
```

```
                                  Weights
X17ImportChangeRate_FullQuarter  0.3801897
X17ExportChangeRate_FullQuarter  0.3854453
GDP1Y_Diff                       0.4155515
Economic_Growth_Rate             0.4866492
COUNTRYCD                        0.9577733
Inflation+Unemp_Rate             1.0279721
TARIFF_AVG                       1.1513395
Distance_Diff                    2.0159682
SP_POP_TOTL                      3.5339865
HSCDPercent                      3.6531960
I(TRADE_COUNTRYCD-TRADE_HSCD)    3.7531349
ContainerPortTraffic             4.7538742
weight                          22.9143464
TRADE_HSCD_COUNTRYCD            25.5109195
amount                          29.0596536
```


Relative Importance of Predictor Variables
amount
TRADE_HSCD_COUNTRYCD
weight
ContainerPortTraffic
I(TRADE_COUNTRYCD-TRADE_HSCD)
HSCDPercent
SP_POP_TOTL
Distance_Diff
TARIFF_AVG
Inflation+Unemp_Rate
COUNTRYCD
Economic_Growth_Rate
GDP1Y_Diff
X17ExportChangeRate_FullQuarter
X17ImportChangeRate_FullQuarter
0
5
10
15
20
25
30
% of R-Square
Total R-Square= 0.69