Math111A

Jia Hui Seow

12/11/2020

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------- tidyverse
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------ tidyverse_confli
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Multiple Linear Regression Example
#import data
f1path <- paste0(getwd(), "/000001.SS-6 China Stock prize.csv")
Data1 <- read.csv(f1path)
f2path <- paste0(getwd(), "/^IXIC-2 USA Stock prize Data.csv")
Data2 <- read.csv(f2path)
f3path <- paste0(getwd(), "/^N225-2  Japan Stocks Data Daily .csv")
Data3 <- read.csv(f3path)
head(Data1)
```

```
##         Date     Open     High      Low    Close Adj.Close Volume
## 1 2019-12-31 3036.386 3051.677 3030.512 3050.124  3050.124 217400
## 2 2020-01-02 3066.336 3098.100 3066.336 3085.198  3085.198 292500
## 3 2020-01-03 3089.022 3093.819 3074.518 3083.786  3083.786 261500
## 4 2020-01-06 3070.909 3107.203 3065.309 3083.408  3083.408 312600
## 5 2020-01-07 3085.488 3105.451 3084.329 3104.802  3104.802 276600
## 6 2020-01-08 3094.239 3094.239 3059.131 3066.893  3066.893 297900
```

```
colnames(Data1)
```

```
## [1] "Date"      "Open"      "High"      "Low"       "Close"     "Adj.Close"
## [7] "Volume"
```

```
colnames(Data2)
```

```
## [1] "Date"      "Open"      "High"      "Low"       "Close"     "Adj.Close"
## [7] "Volume"
```

```
colnames(Data3)
```

```
## [1] "Date"      "Open"      "High"      "Low"       "Close"     "Adj.Close"
## [7] "Volume"
```

```
Data1nData2<-merge(Data1,Data2, by = "Date")
dim(Data1nData2)
```

```
## [1] 209  13
```

```
head(Data1nData2)
```

```
##         Date   Open.x   High.x    Low.x  Close.x Adj.Close.x Volume.x  Open.y
## 1 2019-12-31 3036.386 3051.677 3030.512 3050.124    3050.124   217400 8918.74
## 2 2020-01-02 3066.336 3098.100 3066.336 3085.198    3085.198   292500 9039.46
## 3 2020-01-03 3089.022 3093.819 3074.518 3083.786    3083.786   261500 8976.43
## 4 2020-01-06 3070.909 3107.203 3065.309 3083.408    3083.408   312600 8943.50
## 5 2020-01-07 3085.488 3105.451 3084.329 3104.802    3104.802   276600 9076.64
## 6 2020-01-08 3094.239 3094.239 3059.131 3066.893    3066.893   297900 9068.03
##    High.y   Low.y Close.y Adj.Close.y   Volume.y
## 1 8975.36 8912.77 8972.60     8972.60 2182800000
## 2 9093.43 9010.89 9092.19     9092.19 2848370000
## 3 9065.76 8976.43 9020.77     9020.77 2567400000
## 4 9072.41 8943.50 9071.47     9071.47 2788120000
## 5 9091.93 9042.55 9068.58     9068.58 2352850000
## 6 9168.89 9059.38 9129.24     9129.24 2464090000
```

```
Data1nData2nData3<-merge(Data1nData2,Data3, by = "Date")
dim(Data1nData2nData3)
```

```
## [1] 194  19
```

```
head(Data1nData2nData3)
```

```
##         Date   Open.x   High.x    Low.x  Close.x Adj.Close.x Volume.x  Open.y
## 1 2020-01-06 3070.909 3107.203 3065.309 3083.408    3083.408   312600 8943.50
## 2 2020-01-07 3085.488 3105.451 3084.329 3104.802    3104.802   276600 9076.64
## 3 2020-01-08 3094.239 3094.239 3059.131 3066.893    3066.893   297900 9068.03
## 4 2020-01-09 3082.640 3097.329 3080.131 3094.882    3094.882   243400 9202.27
## 5 2020-01-10 3102.294 3105.225 3081.396 3092.291    3092.291   210400 9232.95
## 6 2020-01-14 3120.667 3127.169 3105.605 3106.820    3106.820   230000 9270.61
##    High.y   Low.y Close.y Adj.Close.y   Volume.y     Open     High      Low
## 1 9072.41 8943.50 9071.47     9071.47 2788120000 23319.76 23365.36 23148.53
## 2 9091.93 9042.55 9068.58     9068.58 2352850000 23320.12 23577.44 23299.92
## 3 9168.89 9059.38 9129.24     9129.24 2464090000 23217.49 23303.21 22951.18
## 4 9215.95 9158.50 9203.43     9203.43 2534700000 23530.29 23767.09 23506.15
## 5 9235.20 9164.66 9178.86     9178.86 2378990000 23813.28 23903.29 23761.08
## 6 9298.33 9226.49 9251.33     9251.33 2542170000 23969.04 24059.86 23951.66
##      Close Adj.Close Volume
## 1 23204.86  23204.86  72800
## 2 23575.72  23575.72  64300
## 3 23204.76  23204.76  79400
## 4 23739.87  23739.87  62200
## 5 23850.57  23850.57  55900
## 6 24025.17  24025.17  64200
```

```
tail(Data1nData2nData3)
```

```
##           Date   Open.x   High.x    Low.x  Close.x Adj.Close.x Volume.x
## 189 2020-11-13 3327.229 3327.229 3291.643 3310.105    3310.105   227900
## 190 2020-11-16 3325.621 3346.969 3313.647 3346.969    3346.969   279600
## 191 2020-11-17 3347.150 3347.700 3323.950 3339.900    3339.900   283600
```

```
## 192 2020-11-18 3337.332 3358.894 3333.992 3347.303    3347.303   287500
## 193 2020-11-19 3339.086 3367.333 3330.450 3363.088    3363.088   253300
## 194 2020-11-20 3359.600 3380.150 3356.310 3377.730    3377.730   261300
##       Open.y   High.y    Low.y  Close.y Adj.Close.y   Volume.y      Open
## 189 11794.94 11849.79 11715.52 11829.29    11829.29 3640680000 25405.64
## 190 11847.11 11937.72 11814.89 11924.13    11924.13 4138920000 25652.69
## 191 11913.35 11950.18 11852.41 11899.34    11899.34 4122770000 26043.45
## 192 11896.06 11942.49 11799.96 11801.60    11801.60 4699160000 25860.55
## 193 11779.04 11912.63 11760.98 11904.71    11904.71 5322580000 25628.73
## 194 11892.70 11935.47 11852.51 11854.97    11854.97 5319340000 25486.83
##         High      Low    Close Adj.Close Volume
## 189 25456.18 25215.31 25385.87  25385.87  77600
## 190 25928.18 25640.29 25906.93  25906.93  84900
## 191 26057.30 25851.54 26014.62  26014.62  84800
## 192 25882.14 25656.70 25728.14  25728.14  69200
## 193 25650.86 25474.94 25634.34  25634.34  82000
## 194 25555.37 25425.59 25527.37  25527.37  63400
```

```
all<- Data1nData2nData3
col<-c(1,2,8,14)
all_sub<-all[,col]
head(all_sub)
```

```
##         Date   Open.x  Open.y     Open
## 1 2020-01-06 3070.909 8943.50 23319.76
## 2 2020-01-07 3085.488 9076.64 23320.12
## 3 2020-01-08 3094.239 9068.03 23217.49
## 4 2020-01-09 3082.640 9202.27 23530.29
## 5 2020-01-10 3102.294 9232.95 23813.28
## 6 2020-01-14 3120.667 9270.61 23969.04
```

```
f4path <- paste0(getwd(), "/China-Covid19data.csv")
Data4 <- read.csv(f4path)
f5path <- paste0(getwd(), "/USA-Covid19Data.csv")
Data5 <- read.csv(f5path)
f6path <- paste0(getwd(), "/Japan_Covid19Data.csv")
Data6 <- read.csv(f6path)
head(Data4)
```

```
##         Date    dateRep day month year cases deaths countriesAndTerritories
## 1 2019-12-31 31/12/2019  31    12 2019    27      0                   China
## 2 2020-01-01   1/1/2020   1     1 2020     0      0                   China
## 3 2020-01-02   2/1/2020   2     1 2020     0      0                   China
## 4 2020-01-03   3/1/2020   3     1 2020    17      0                   China
## 5 2020-01-04   4/1/2020   4     1 2020     0      0                   China
## 6 2020-01-05   5/1/2020   5     1 2020    15      0                   China
##   countriesCode
## 1            CN
## 2            CN
## 3            CN
## 4            CN
## 5            CN
## 6            CN
```

```
colnames(Data4)
```

```
## [1] "Date"                    "dateRep"
## [3] "day"                     "month"
## [5] "year"                    "cases"
## [7] "deaths"                  "countriesAndTerritories"
## [9] "countriesCode"
```

```
colnames(Data5)
```

```
## [1] "Date"                    "dateRep"
## [3] "day"                     "month"
## [5] "year"                    "cases"
## [7] "deaths"                  "countriesAndTerritories"
## [9] "countriesCode"
```

```
colnames(Data6)
```

```
## [1] "Date"                    "dateRep"
## [3] "day"                     "month"
## [5] "year"                    "cases"
## [7] "deaths"                  "countriesAndTerritories"
## [9] "countriesCode"
```

```
Data4nData5<-merge(Data4,Data5, by = "Date")
dim(Data4nData5)
```

```
## [1] 326  17
```

```
head(Data4nData5)
```

```
##         Date  dateRep.x day.x month.x year.x cases.x deaths.x
## 1 2019-12-31 31/12/2019    31      12   2019      27        0
## 2 2020-01-01   1/1/2020     1       1   2020       0        0
## 3 2020-01-02   2/1/2020     2       1   2020       0        0
## 4 2020-01-03   3/1/2020     3       1   2020      17        0
## 5 2020-01-04   4/1/2020     4       1   2020       0        0
## 6 2020-01-05   5/1/2020     5       1   2020      15        0
##   countriesAndTerritories.x countriesCode.x  dateRep.y day.y month.y year.y
## 1                     China              CN 31/12/2019    31      12   2019
## 2                     China              CN   1/1/2020     1       1   2020
## 3                     China              CN   2/1/2020     2       1   2020
## 4                     China              CN   3/1/2020     3       1   2020
## 5                     China              CN   4/1/2020     4       1   2020
## 6                     China              CN   5/1/2020     5       1   2020
##   cases.y deaths.y countriesAndTerritories.y countriesCode.y
## 1       0        0  United_States_of_America             US
## 2       0        0  United_States_of_America             US
## 3       0        0  United_States_of_America             US
## 4       0        0  United_States_of_America             US
## 5       0        0  United_States_of_America             US
## 6       0        0  United_States_of_America             US
```

```
Data4nData5nData6<-merge(Data4nData5,Data6, by = "Date")
dim(Data4nData5nData6)
```

```
## [1] 326  25
```

```
head(Data4nData5nData6)
```

```
##         Date  dateRep.x day.x month.x year.x cases.x deaths.x
## 1 2019-12-31 31/12/2019    31      12   2019      27        0
## 2 2020-01-01   1/1/2020     1       1   2020       0        0
## 3 2020-01-02   2/1/2020     2       1   2020       0        0
## 4 2020-01-03   3/1/2020     3       1   2020      17        0
## 5 2020-01-04   4/1/2020     4       1   2020       0        0
## 6 2020-01-05   5/1/2020     5       1   2020      15        0
##   countriesAndTerritories.x countriesCode.x  dateRep.y day.y month.y year.y
## 1                     China              CN 31/12/2019    31      12   2019
## 2                     China              CN   1/1/2020     1       1   2020
## 3                     China              CN   2/1/2020     2       1   2020
## 4                     China              CN   3/1/2020     3       1   2020
## 5                     China              CN   4/1/2020     4       1   2020
## 6                     China              CN   5/1/2020     5       1   2020
##   cases.y deaths.y countriesAndTerritories.y countriesCode.y    dateRep day
## 1       0        0  United_States_of_America              US 31/12/2019  31
## 2       0        0  United_States_of_America              US   1/1/2020   1
## 3       0        0  United_States_of_America              US   2/1/2020   2
## 4       0        0  United_States_of_America              US   3/1/2020   3
## 5       0        0  United_States_of_America              US   4/1/2020   4
## 6       0        0  United_States_of_America              US   5/1/2020   5
##   month year cases deaths countriesAndTerritories countriesCode
## 1    12 2019     0      0                   Japan            JP
## 2     1 2020     0      0                   Japan            JP
## 3     1 2020     0      0                   Japan            JP
## 4     1 2020     0      0                   Japan            JP
## 5     1 2020     0      0                   Japan            JP
## 6     1 2020     0      0                   Japan            JP
```

```
all2<- Data4nData5nData6
all2_sub <- all2[7:326,]
tail(all2_sub)
```

```
##           Date  dateRep.x day.x month.x year.x cases.x deaths.x
## 321 2020-11-15 15/11/2020    15      11   2020      11        0
## 322 2020-11-16 16/11/2020    16      11   2020       7        0
## 323 2020-11-17 17/11/2020    17      11   2020      15        0
## 324 2020-11-18 18/11/2020    18      11   2020       8        0
## 325 2020-11-19 19/11/2020    19      11   2020      12        0
## 326 2020-11-20 10/11/2020    20      11   2020      17        0
##     countriesAndTerritories.x countriesCode.x  dateRep.y day.y month.y year.y
## 321                     China              CN 15/11/2020    15      11   2020
## 322                     China              CN 16/11/2020    16      11   2020
## 323                     China              CN 17/11/2020    17      11   2020
## 324                     China              CN 18/11/2020    18      11   2020
## 325                     China              CN 19/11/2020    19      11   2020
## 326                     China              CN 20/11/2020    20      11   2020
##     cases.y deaths.y countriesAndTerritories.y countriesCode.y    dateRep day
## 321  165277     1255  United_States_of_America              US 15/11/2020  15
## 322  132170      614  United_States_of_America              US 16/11/2020  16
## 323  168425     1006  United_States_of_America              US 17/11/2020  17
## 324  154316     1467  United_States_of_America              US 18/11/2020  18
## 325  170005     1850  United_States_of_America              US 19/11/2020  19
```

```
## 326  188020     2018  United_States_of_America             US 20/11/2020  20
##     month year cases deaths countriesAndTerritories countriesCode
## 321    11 2020  1694      3                   Japan            JP
## 322    11 2020     0      0                   Japan            JP
## 323    11 2020  2649     20                   Japan            JP
## 324    11 2020  1489     10                   Japan            JP
## 325    11 2020  2151      9                   Japan            JP
## 326    11 2020  2301     21                   Japan            JP
```

```
col_covidcases<-c(1,6,14,22)
col_coviddeath<-c(1,7,15,23)
covidcases<-all2_sub[,col_covidcases]
coviddeath<-all2_sub[,col_coviddeath]
head(covidcases)
```

```
##          Date cases.x cases.y cases
## 7  2020-01-06       0       0     0
## 8  2020-01-07       0       0     0
## 9  2020-01-08       0       0     0
## 10 2020-01-09       0       0     0
## 11 2020-01-10       0       0     0
## 12 2020-01-11       0       0     0
```

```
head(coviddeath)
```

```
##          Date deaths.x deaths.y deaths
## 7  2020-01-06        0        0      0
## 8  2020-01-07        0        0      0
## 9  2020-01-08        0        0      0
## 10 2020-01-09        0        0      0
## 11 2020-01-10        0        0      0
## 12 2020-01-11        1        0      0
```

```
f7path <- paste0(getwd(), "/Math111AUnemploymentRate.csv")
Data7 <- read.csv(f7path)
dim(Data7)
```

```
## [1] 320   4
```

```
head(Data7)
```

```
##         Date ChinaUnemply USAUnemply JapanUnemply
## 1 2020-01-06          5.2        3.6          2.2
## 2 2020-01-07          5.2        3.6          2.2
## 3 2020-01-08          5.2        3.6          2.2
## 4 2020-01-09          5.2        3.6          2.2
## 5 2020-01-10          5.2        3.6          2.2
## 6 2020-01-11          5.2        3.6          2.2
```

```
UnemployRate <- Data7

ALL<-merge(all2, all, by = "Date")
dim(ALL)
```

```
## [1] 194  43
```

```
head(ALL)
```

```
##         Date  dateRep.x day.x month.x year.x cases.x deaths.x
```

```
## 1 2020-01-06   6/1/2020     6       1   2020       0        0
## 2 2020-01-07   7/1/2020     7       1   2020       0        0
## 3 2020-01-08   8/1/2020     8       1   2020       0        0
## 4 2020-01-09   9/1/2020     9       1   2020       0        0
## 5 2020-01-10  10/1/2020    10       1   2020       0        0
## 6 2020-01-14 14/01/2020    14       1   2020       0        0
##   countriesAndTerritories.x countriesCode.x  dateRep.y day.y month.y year.y
## 1                     China              CN   6/1/2020     6       1   2020
## 2                     China              CN   7/1/2020     7       1   2020
## 3                     China              CN   8/1/2020     8       1   2020
## 4                     China              CN   9/1/2020     9       1   2020
## 5                     China              CN  10/1/2020    10       1   2020
## 6                     China              CN 14/01/2020    14       1   2020
##   cases.y deaths.y countriesAndTerritories.y countriesCode.y    dateRep day
## 1       0        0  United_States_of_America              US   6/1/2020   6
## 2       0        0  United_States_of_America              US   7/1/2020   7
## 3       0        0  United_States_of_America              US   8/1/2020   8
## 4       0        0  United_States_of_America              US   9/1/2020   9
## 5       0        0  United_States_of_America              US  10/1/2020  10
## 6       0        0  United_States_of_America              US 14/01/2020  14
##   month year cases deaths countriesAndTerritories countriesCode   Open.x
## 1     1 2020     0      0                   Japan            JP 3070.909
## 2     1 2020     0      0                   Japan            JP 3085.488
## 3     1 2020     0      0                   Japan            JP 3094.239
## 4     1 2020     0      0                   Japan            JP 3082.640
## 5     1 2020     0      0                   Japan            JP 3102.294
## 6     1 2020     0      0                   Japan            JP 3120.667
##     High.x    Low.x  Close.x Adj.Close.x Volume.x  Open.y  High.y   Low.y
## 1 3107.203 3065.309 3083.408    3083.408   312600 8943.50 9072.41 8943.50
## 2 3105.451 3084.329 3104.802    3104.802   276600 9076.64 9091.93 9042.55
## 3 3094.239 3059.131 3066.893    3066.893   297900 9068.03 9168.89 9059.38
## 4 3097.329 3080.131 3094.882    3094.882   243400 9202.27 9215.95 9158.50
## 5 3105.225 3081.396 3092.291    3092.291   210400 9232.95 9235.20 9164.66
## 6 3127.169 3105.605 3106.820    3106.820   230000 9270.61 9298.33 9226.49
##   Close.y Adj.Close.y   Volume.y     Open     High      Low    Close Adj.Close
## 1 9071.47     9071.47 2788120000 23319.76 23365.36 23148.53 23204.86  23204.86
## 2 9068.58     9068.58 2352850000 23320.12 23577.44 23299.92 23575.72  23575.72
## 3 9129.24     9129.24 2464090000 23217.49 23303.21 22951.18 23204.76  23204.76
## 4 9203.43     9203.43 2534700000 23530.29 23767.09 23506.15 23739.87  23739.87
## 5 9178.86     9178.86 2378990000 23813.28 23903.29 23761.08 23850.57  23850.57
## 6 9251.33     9251.33 2542170000 23969.04 24059.86 23951.66 24025.17  24025.17
##   Volume
## 1  72800
## 2  64300
## 3  79400
## 4  62200
## 5  55900
## 6  64200
```

```
colnames(ALL)
```

```
##  [1] "Date"                      "dateRep.x"
##  [3] "day.x"                     "month.x"
##  [5] "year.x"                    "cases.x"
##  [7] "deaths.x"                  "countriesAndTerritories.x"
```

```
##  [9] "countriesCode.x"           "dateRep.y"
## [11] "day.y"                     "month.y"
## [13] "year.y"                    "cases.y"
## [15] "deaths.y"                  "countriesAndTerritories.y"
## [17] "countriesCode.y"           "dateRep"
## [19] "day"                       "month"
## [21] "year"                      "cases"
## [23] "deaths"                    "countriesAndTerritories"
## [25] "countriesCode"             "Open.x"
## [27] "High.x"                    "Low.x"
## [29] "Close.x"                   "Adj.Close.x"
## [31] "Volume.x"                  "Open.y"
## [33] "High.y"                    "Low.y"
## [35] "Close.y"                   "Adj.Close.y"
## [37] "Volume.y"                  "Open"
## [39] "High"                      "Low"
## [41] "Close"                     "Adj.Close"
## [43] "Volume"
```

```
col3<-c(1,6,14,22,26,32,38)
all3_sub<-ALL[,col3]
head(all3_sub)
```

```
##         Date cases.x cases.y cases   Open.x  Open.y     Open
## 1 2020-01-06       0       0     0 3070.909 8943.50 23319.76
## 2 2020-01-07       0       0     0 3085.488 9076.64 23320.12
## 3 2020-01-08       0       0     0 3094.239 9068.03 23217.49
## 4 2020-01-09       0       0     0 3082.640 9202.27 23530.29
## 5 2020-01-10       0       0     0 3102.294 9232.95 23813.28
## 6 2020-01-14       0       0     0 3120.667 9270.61 23969.04
```

```
AALL<-merge(all3_sub, Data7, by = "Date")
dim(AALL)
```

```
## [1] 194  10
```

```
head(AALL)
```

```
##         Date cases.x cases.y cases   Open.x  Open.y     Open ChinaUnemply
## 1 2020-01-06       0       0     0 3070.909 8943.50 23319.76          5.2
## 2 2020-01-07       0       0     0 3085.488 9076.64 23320.12          5.2
## 3 2020-01-08       0       0     0 3094.239 9068.03 23217.49          5.2
## 4 2020-01-09       0       0     0 3082.640 9202.27 23530.29          5.2
## 5 2020-01-10       0       0     0 3102.294 9232.95 23813.28          5.2
## 6 2020-01-14       0       0     0 3120.667 9270.61 23969.04          5.2
##   USAUnemply JapanUnemply
## 1        3.6          2.2
## 2        3.6          2.2
## 3        3.6          2.2
## 4        3.6          2.2
## 5        3.6          2.2
## 6        3.6          2.2
```

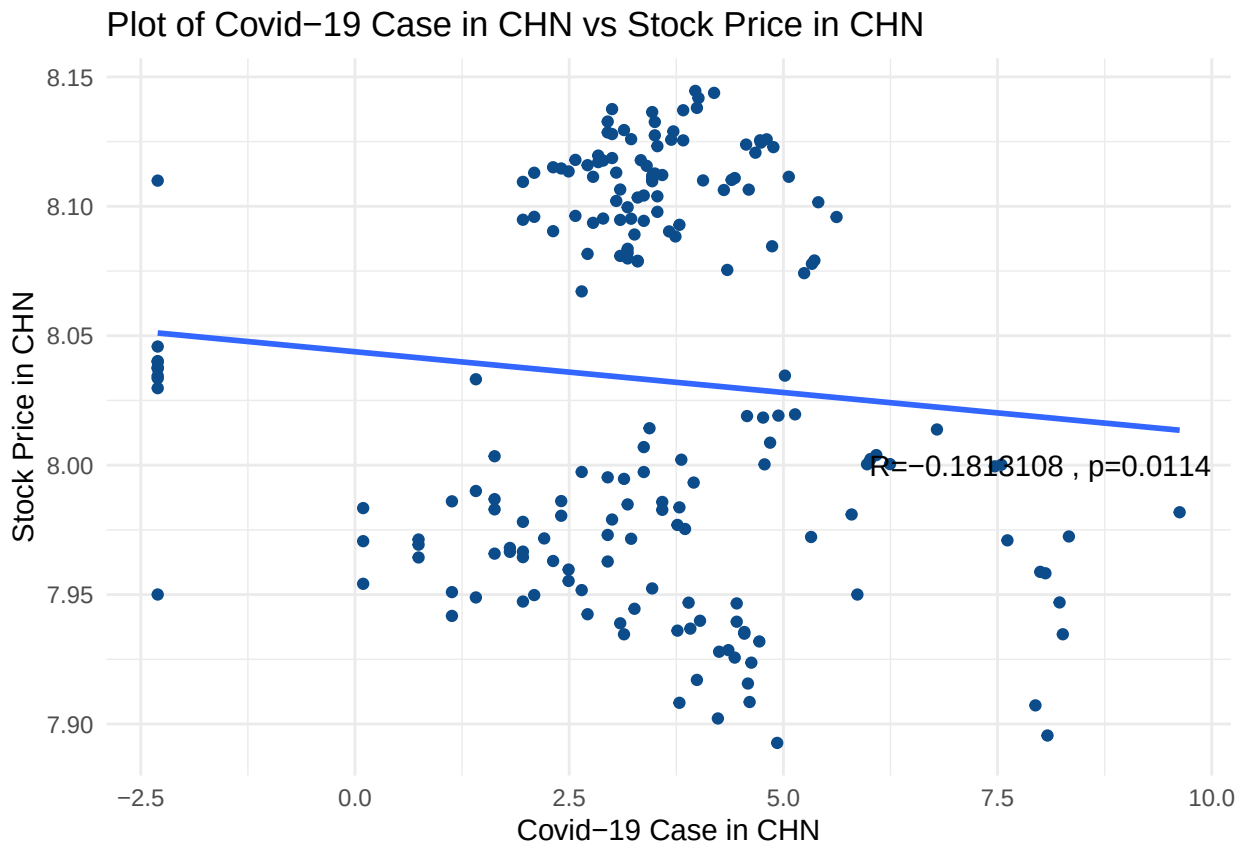
```
colnames(AALL)=c("date","c1","c2","c3","o1","o2","o3","u1","u2","u3")
head(AALL)
```

```
##         date c1 c2 c3       o1      o2       o3  u1  u2  u3
```

```
## 1 2020-01-06  0  0  0 3070.909 8943.50 23319.76 5.2 3.6 2.2
## 2 2020-01-07  0  0  0 3085.488 9076.64 23320.12 5.2 3.6 2.2
## 3 2020-01-08  0  0  0 3094.239 9068.03 23217.49 5.2 3.6 2.2
## 4 2020-01-09  0  0  0 3082.640 9202.27 23530.29 5.2 3.6 2.2
## 5 2020-01-10  0  0  0 3102.294 9232.95 23813.28 5.2 3.6 2.2
## 6 2020-01-14  0  0  0 3120.667 9270.61 23969.04 5.2 3.6 2.2
```

```
# scatterplot
library(ggplot2)
sp1<-ggplot(AALL) +
  aes(x = log(c1+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Covid-19 Case in CHN vs Stock Price in CHN",
       x = "Covid-19 Case in CHN",y ="Stock Price in CHN") +
  annotate("text", x=8, y=8, label= "R=-0.1813108 , p=0.0114")# Add correlation coefficient & P
sp1
```

```
## `geom_smooth()` using formula 'y ~ x'
```



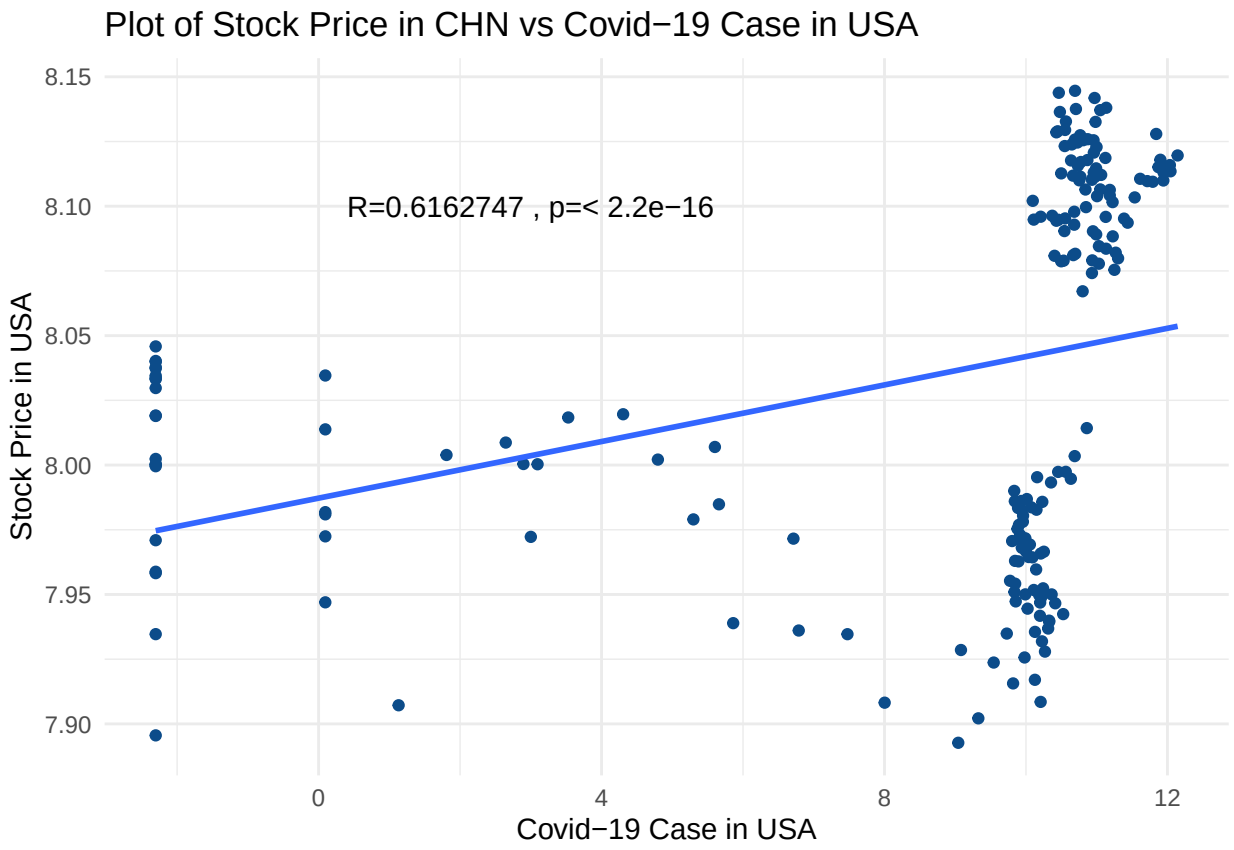

```
cor.test(AALL$o1, AALL$c1) #-0.1813108  almost 0
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$o1 and AALL$c1
## t = -2.5547, df = 192, p-value = 0.0114
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.3141612 -0.0414958
## sample estimates:
##        cor
## -0.1813108
```

```
sp2<-ggplot(AALL) +
  aes(x = log(c2+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Covid-19 Case in USA",
       x = "Covid-19 Case in USA",y ="Stock Price in USA") +
  annotate("text", x=3, y=8.1, label= "R=0.6162747 , p=< 2.2e-16")# Add correlation coefficient
sp2
```

```
## `geom_smooth()` using formula 'y ~ x'
```



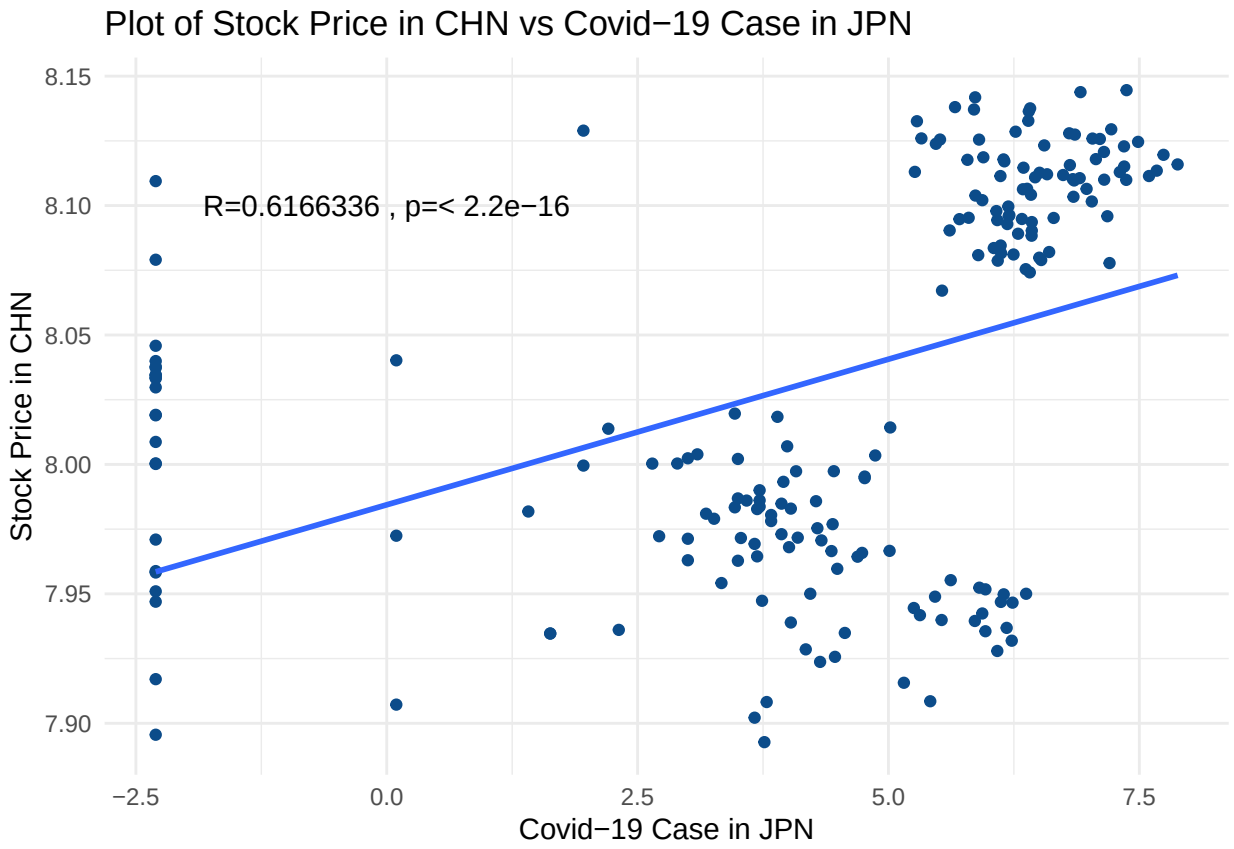

```
cor.test(AALL$c2, AALL$o1) #0.6162747
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$c2 and AALL$o1
## t = 10.843, df = 192, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5205968 0.6966666
## sample estimates:
```

```
##       cor
## 0.6162747
```

```
sp3<-ggplot(AALL) +
  aes(x = log(c3+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Covid-19 Case in JPN",
       x = "Covid-19 Case in JPN",y ="Stock Price in CHN") +
  annotate("text", x=0, y=8.1, label= "R=0.6166336 , p=< 2.2e-16") # Add correlation coefficient
sp3
```

```
## `geom_smooth()` using formula 'y ~ x'
```

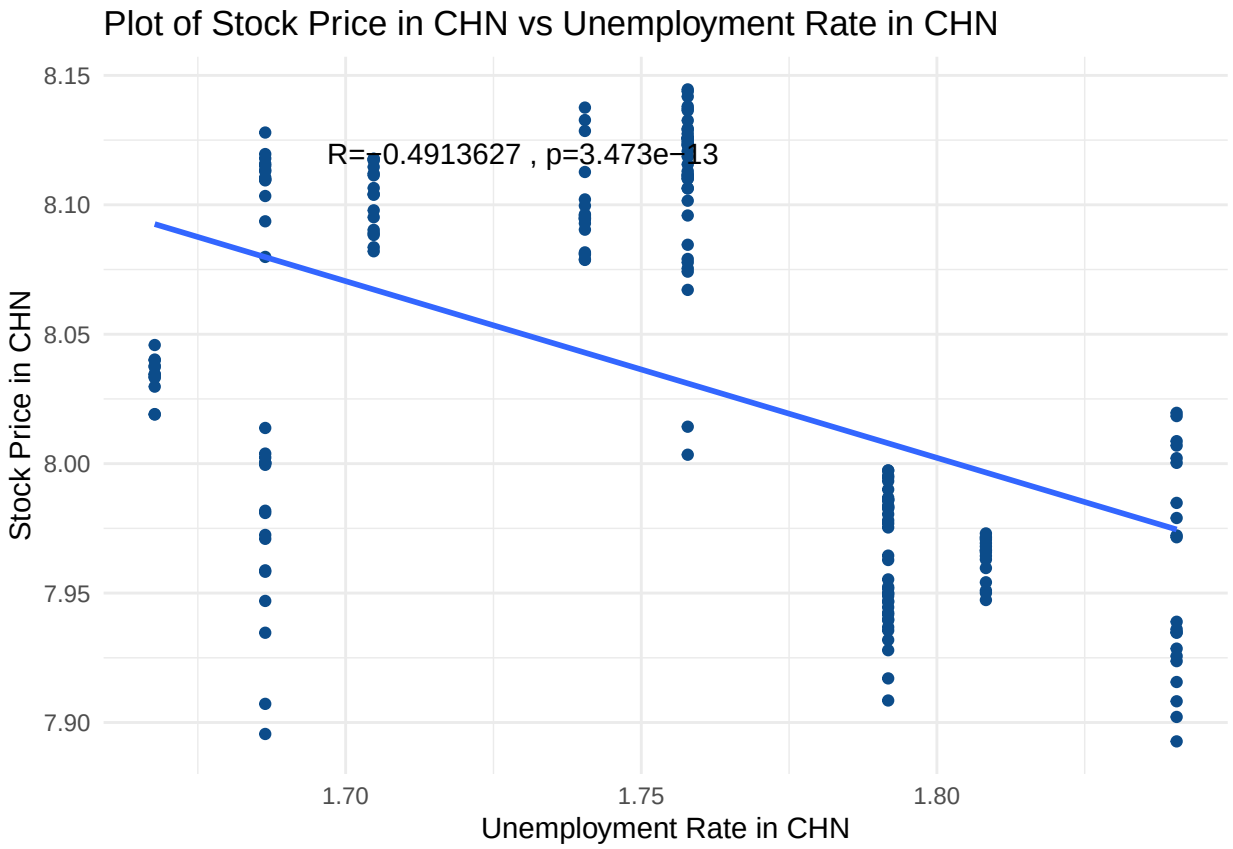



```
cor.test(AALL$c3, AALL$o1) #0.6166336
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$c3 and AALL$o1
## t = 10.853, df = 192, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5210187 0.6969644
## sample estimates:
##       cor
## 0.6166336
```

```
sp4<-ggplot(AALL) +
  aes(x = log(u1+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Unemployment Rate in CHN",
       x = "Unemployment Rate in CHN",y ="Stock Price in CHN") +
  annotate("text", x=1.73, y=8.12, label= "R=-0.4913627 , p=3.473e-13") # Add correlation coefficient
sp4
```

```
## `geom_smooth()` using formula 'y ~ x'
```



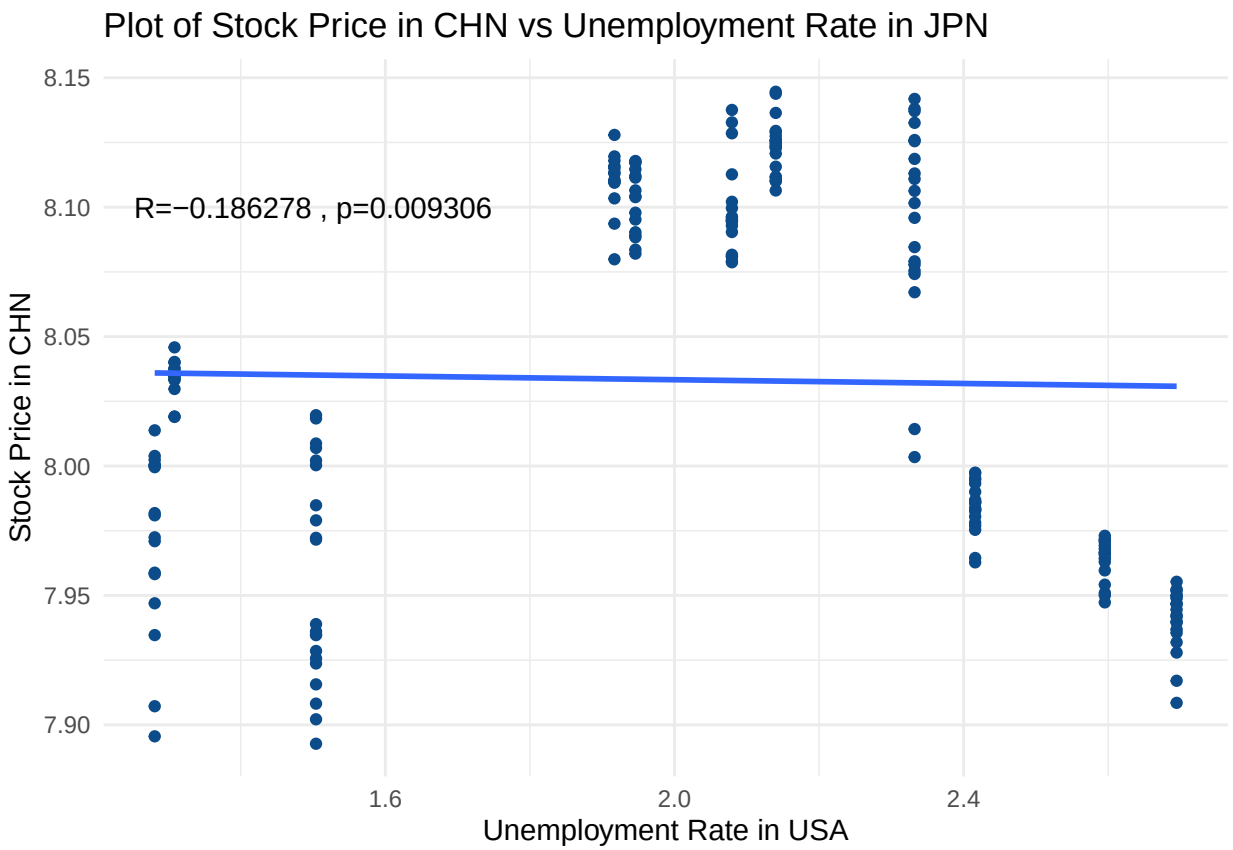

```
cor.test(AALL$u1, AALL$o1) #-0.4913627
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$u1 and AALL$o1
## t = -7.8173, df = 192, p-value = 3.473e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5913069 -0.3765531
## sample estimates:
##        cor
## -0.4913627
```

```
sp5<-ggplot(AALL) +
  aes(x = log(u2+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
```

```
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Unemployment Rate in JPN",
       x = "Unemployment Rate in USA",y ="Stock Price in CHN") +
  annotate("text", x=1.5, y=8.1, label= "R=-0.186278 , p=0.009306") # Add correlation coefficient
sp5
```

```
## `geom_smooth()` using formula 'y ~ x'
```



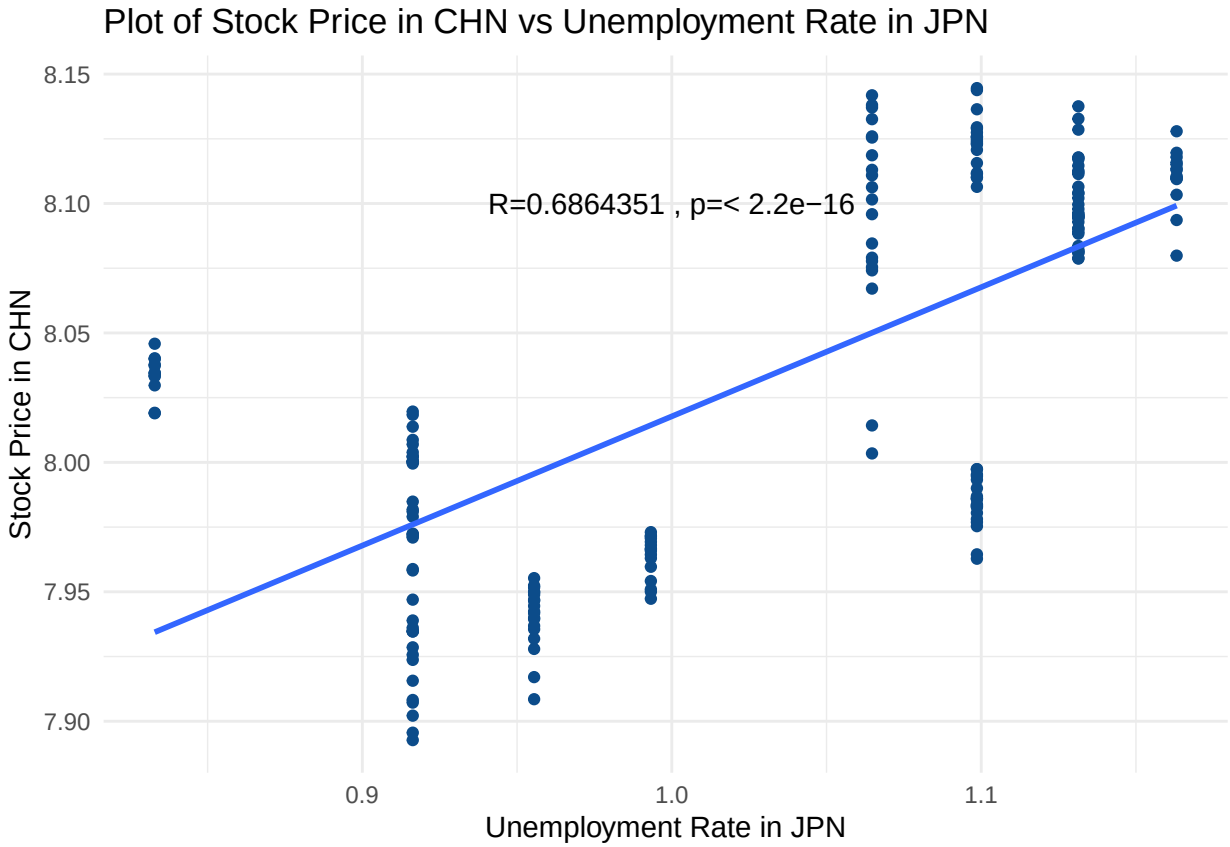

```
cor.test(AALL$u2, AALL$o1) #-0.186278
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$u2 and AALL$o1
## t = -2.6271, df = 192, p-value = 0.009306
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.31878730 -0.04662678
## sample estimates:
##       cor
## -0.186278
```

```
sp6<-ggplot(AALL) +
  aes(x = log(u3+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Unemployment Rate in JPN",
       x = "Unemployment Rate in JPN",y ="Stock Price in CHN") +
```

```
  annotate("text", x=1, y=8.1, label= "R=0.6864351 , p=< 2.2e-16") # Add correlation coefficient
sp6
```

```
## `geom_smooth()` using formula 'y ~ x'
```

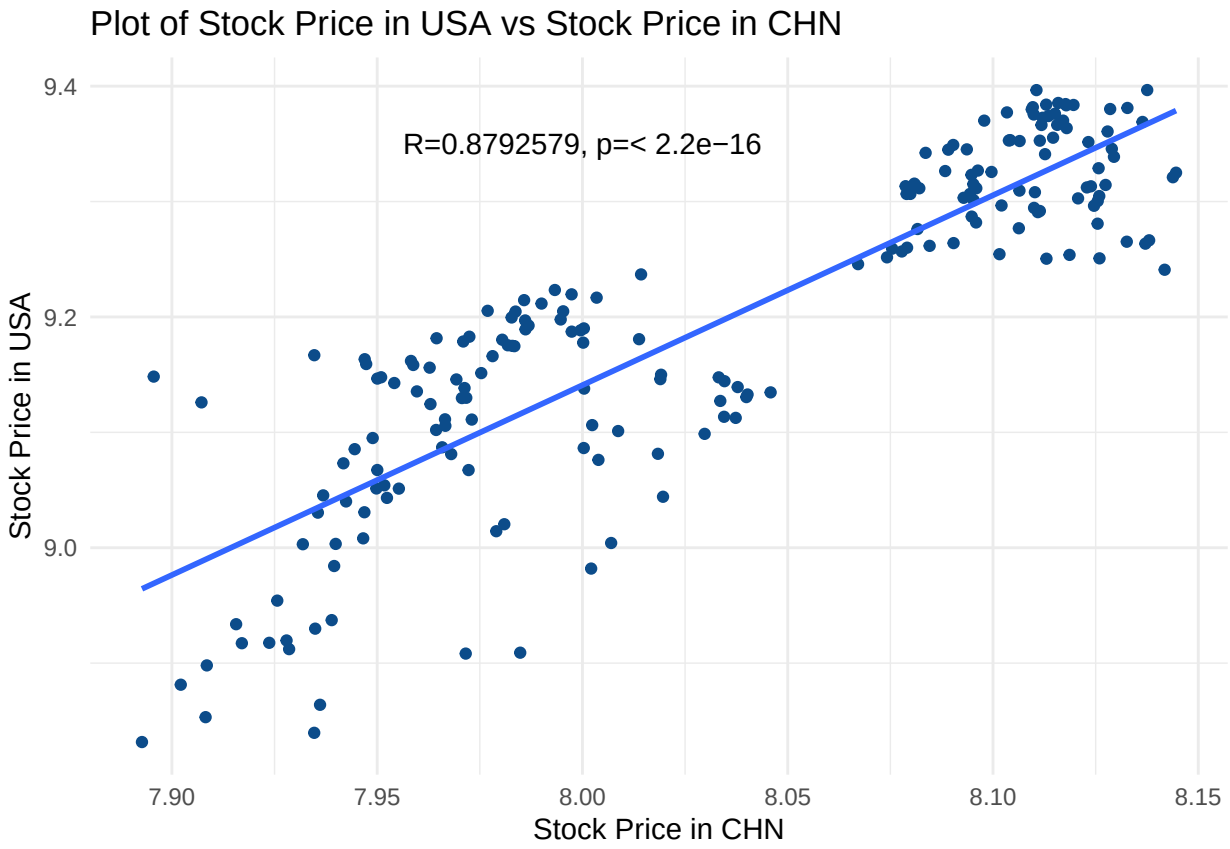



```
cor.test(AALL$u3, AALL$o1) #0.6864351
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$u3 and AALL$o1
## t = 13.08, df = 192, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6039645 0.7543620
## sample estimates:
##       cor
## 0.6864351
```

```
sp7<-ggplot(AALL) +
  aes(x = log(o1+0.1),y = log(o2+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in USA vs Stock Price in CHN",
       x = "Stock Price in CHN",y ="Stock Price in USA") +
  annotate("text", x=8, y=9.35, label= "R=0.8792579, p=< 2.2e-16")
sp7
```

```
## `geom_smooth()` using formula 'y ~ x'
```

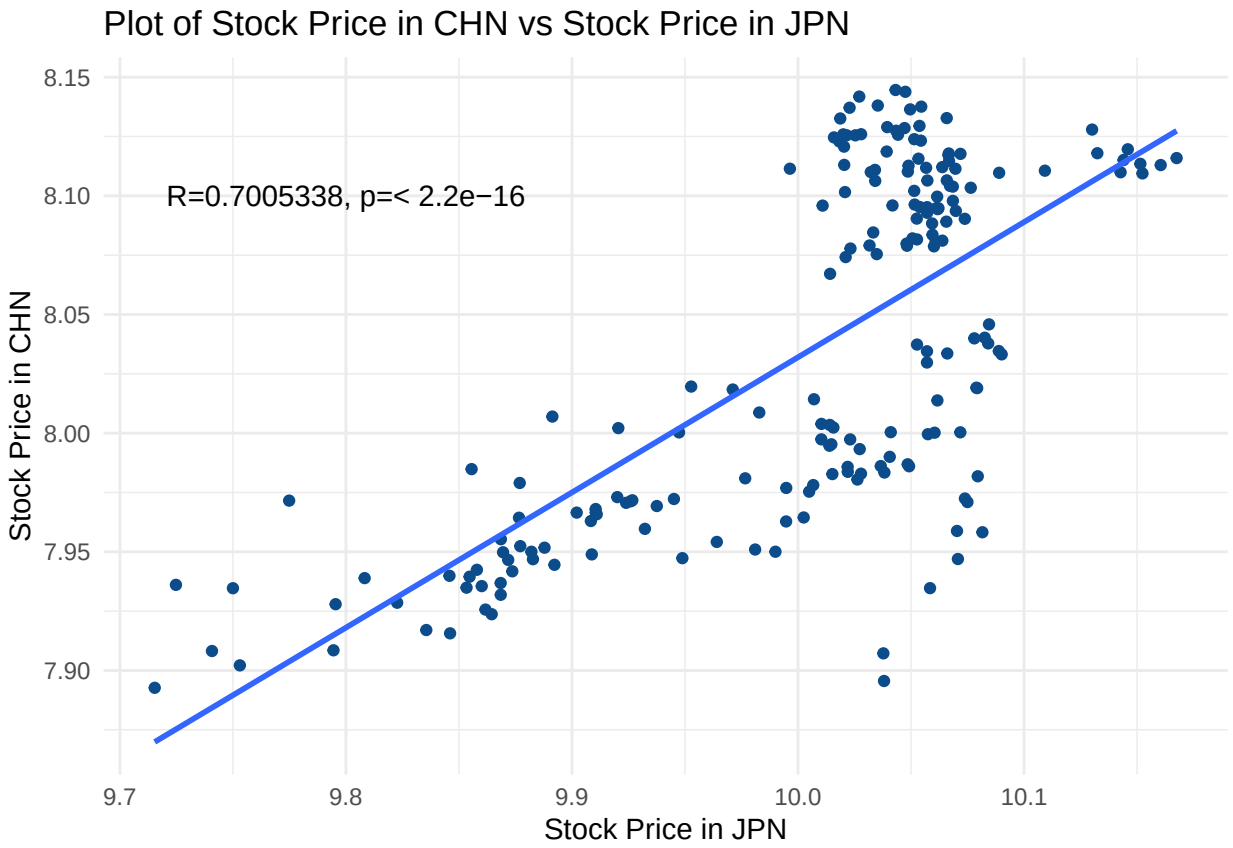



```
cor.test(AALL$o1, AALL$o2) #0.8792579
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$o1 and AALL$o2
## t = 25.577, df = 192, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8427735 0.9077002
## sample estimates:
##       cor
## 0.8792579
```

```
sp8<-ggplot(AALL) +
  aes(x = log(o3+0.1),y = log(o1+0.1)) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal() + geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Stock Price in CHN vs Stock Price in JPN",
       x = "Stock Price in JPN",y ="Stock Price in CHN") +
  annotate("text", x=9.8, y=8.1, label= "R=0.7005338, p=< 2.2e-16")
sp8
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
cor.test(AALL$o3, AALL$o1) #0.7005338
```

```
##
##  Pearson's product-moment correlation
##
## data:  AALL$o3 and AALL$o1
## t = 13.602, df = 192, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6209378 0.7658306
## sample estimates:
##       cor
## 0.7005338
```

```
library(tidyverse)
##Spilt the dataset into train set and test set
set.seed(1)
row.number <- sample(1:nrow(AALL), 0.8*nrow(AALL))
train = AALL[row.number,]
test = AALL[-row.number,]
dim(train)
```

```
## [1] 155  10
```

```
dim(test)
```

```
## [1] 39 10
```

```r
head(train)
```

```
##           date   c1    c2   c3       o1       o2       o3  u1   u2  u3
## 68  2020-04-28   26 22541  191 2819.991  8825.69 19776.18 5.9 14.7 2.5
## 167 2020-10-13   18 41653  326 3353.121 11901.76 23667.90 5.4  6.9 3.0
## 129 2020-08-06  122 52804 1134 3380.760 10989.98 22471.71 5.7  8.4 2.9
## 162 2020-09-28   27 36248  441 3224.977 11084.38 23391.96 5.6  7.9 3.0
## 43  2020-03-19   44  2988   44 2719.406  6996.45 16995.77 6.2  4.4 2.4
## 14  2020-02-04 3237     0    0 2685.269  9398.39 22881.13 5.3  3.5 2.4
```

```r
head(test)
```

```
##         date c1 c2 c3       o1      o2       o3  u1  u2  u3
## 3 2020-01-08  0  0  0 3094.239 9068.03 23217.49 5.2 3.6 2.2
## 4 2020-01-09  0  0  0 3082.640 9202.27 23530.29 5.2 3.6 2.2
## 5 2020-01-10  0  0  0 3102.294 9232.95 23813.28 5.2 3.6 2.2
## 6 2020-01-14  0  0  0 3120.667 9270.61 23969.04 5.2 3.6 2.2
## 8 2020-01-16  0  0  0 3095.734 9313.45 23960.20 5.2 3.6 2.2
## 9 2020-01-17  4  0  0 3081.464 9392.37 24103.45 5.2 3.6 2.2
```

```r
# Multivariate Linear Regression
fit2 <- lm(o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3, data=train)
fit2
```

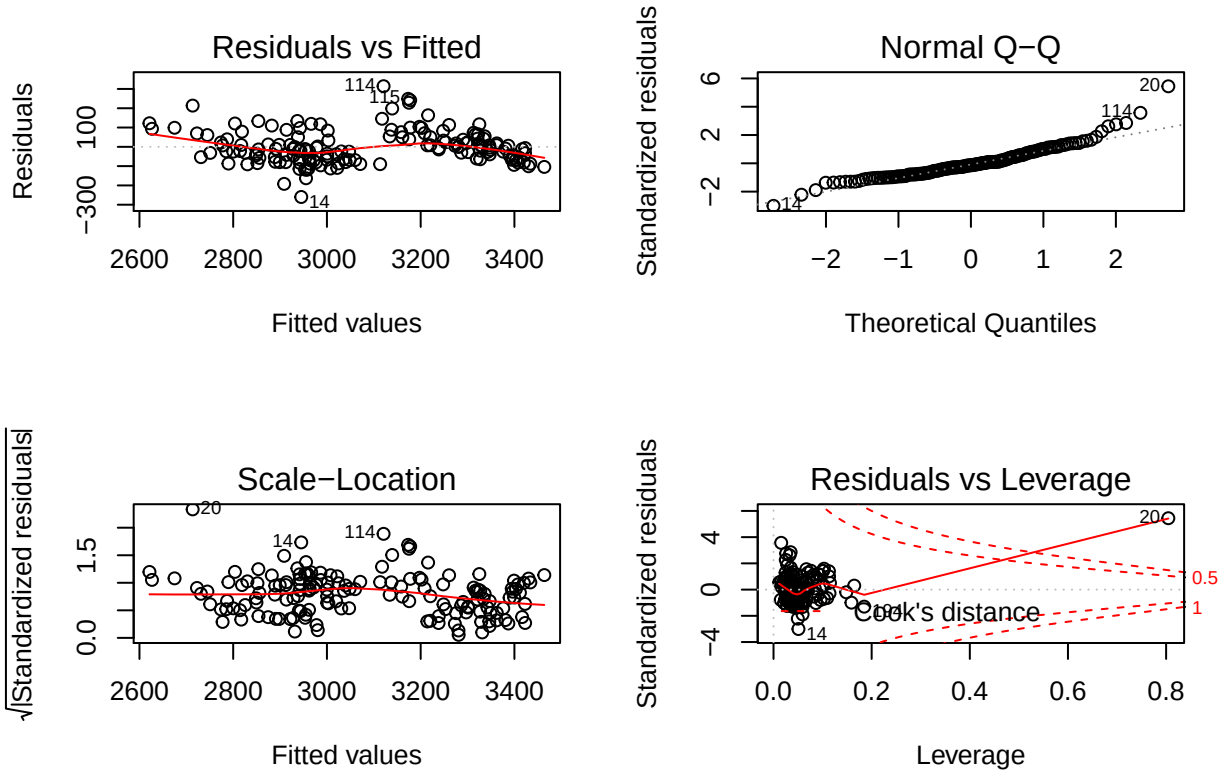
```
##
## Call:
## lm(formula = o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3, data = train)
##
## Coefficients:
## (Intercept)           c1           c2           c3           u1           u2
##   1.785e+03   -2.165e-02    2.988e-04    5.680e-02    9.499e+01   -1.251e+01
##          u3           o2           o3
##  -2.371e+02    2.215e-01   -3.245e-02
```

```r
#Show results
summary(fit2)
```

```
##
## Call:
## lm(formula = o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -259.91  -63.90  -11.42   49.56  314.18
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.785e+03  3.647e+02   4.895 2.57e-06 ***
## c1          -2.165e-02  5.910e-03  -3.664 0.000347 ***
## c2           2.988e-04  3.704e-04   0.807 0.421228
## c3           5.680e-02  2.385e-02   2.382 0.018518 *
## u1           9.499e+01  4.781e+01   1.987 0.048808 *
## u2          -1.251e+01  2.460e+00  -5.087 1.10e-06 ***
## u3          -2.371e+02  7.434e+01  -3.189 0.001748 **
## o2           2.215e-01  1.961e-02  11.297  < 2e-16 ***
## o3          -3.245e-02  1.024e-02  -3.169 0.001861 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.76 on 146 degrees of freedom
## Multiple R-squared:  0.8634, Adjusted R-squared:  0.8559
## F-statistic: 115.4 on 8 and 146 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(fit2)
```




```
#Predicted Result for linear
pred2 <- predict(fit2, newdata = test)
pred2
```

```
##        3        4        5        6        8        9       11       12
## 2967.845 2987.430 2985.043 2988.331 2998.107 3010.854 3021.309 3014.030
##       18       30       32       38       54       63       65       66
## 2947.616 2989.856 3027.107 2772.830 2779.549 2839.607 2852.250 2865.581
##       71       72       77       81       82       99      101      109
## 2943.591 2926.838 2959.598 2972.113 3005.808 2993.995 3010.879 3132.832
##      116      118      120      131      133      137      143      147
## 3146.817 3169.362 3215.204 3254.123 3294.182 3374.434 3426.849 3468.394
##      153      154      164      166      177      179      191
## 3245.147 3220.074 3244.095 3371.021 3313.337 3264.347 3464.216
```

```
modelEval2<-cbind(test$o1,pred2)
colnames(modelEval2) <- c('Actual', 'Predicted')
modelEval2
```

```
##       Actual Predicted
## 3   3094.239  2967.845
```

```
## 4   3082.640  2987.430
## 5   3102.294  2985.043
## 6   3120.667  2988.331
## 8   3095.734  2998.107
## 9   3081.464  3010.854
## 11  3038.487  3021.309
## 12  3037.952  3014.030
## 18  2860.499  2947.616
## 30  2899.310  2989.856
## 32  2981.806  3027.107
## 38  2936.016  2772.830
## 54  2806.968  2779.549
## 63  2842.237  2839.607
## 65  2850.510  2852.250
## 66  2834.939  2865.581
## 71  2882.710  2943.591
## 72  2901.570  2926.838
## 77  2872.524  2959.598
## 81  2863.046  2972.113
## 82  2827.897  3005.808
## 99  2929.880  2993.995
## 101 2966.901  3010.879
## 109 3380.954  3132.832
## 116 3356.359  3146.817
## 118 3243.911  3169.362
## 120 3315.182  3215.204
## 131 3379.487  3254.123
## 133 3328.175  3294.182
## 137 3444.565  3374.434
## 143 3333.490  3426.849
## 147 3420.469  3468.394
## 153 3225.784  3245.147
## 154 3275.918  3220.074
## 164 3232.710  3244.095
## 166 3287.328  3371.021
## 177 3240.740  3313.337
## 179 3235.760  3264.347
## 191 3347.150  3464.216
```

```
rmse <- sqrt(sum((pred2 - test$o1)^2)/length(test$o1))
rmse #96.55912
```

```
## [1] 96.55912
```

```
c(RMSE = rmse, R2=summary(fit2)$r.squared)
```

```
##       RMSE         R2
## 96.5591184  0.8634111
```

```
#RMSE          R2
#96.5591184  0.8634111
```

```
#Since the the differrent between the the first row of o2 and the second row of o2(Stock Price in USA)
# from 1 to 155 so and we get the mean of the different=1.339084
Diff_of_CHN_SP<-c()
for (i in 2:155){
```

```
  Diff_of_CHN_SP[i-1]<-AALL[i,5]-AALL[i-1,5]
}
Diff_of_CHN_SP
```

```
##   [1]   14.579101    8.750977  -11.599121   19.654052   18.373047  -17.497070
##   [7]   -7.436035  -14.269776    4.325928  -47.302978   -0.535157 -321.253906
##  [13]  -31.428955  107.102051   34.518799   32.040039    1.569091   35.056885
##  [19]   31.588135  -27.278076   81.543945   -1.887939    2.357910   40.366211
##  [25]  -40.176026    5.858887    4.561035  -67.848877  -25.331054  107.578857
##  [31]  -25.083008   54.349121    3.784912  -52.759033  -68.245849   82.826904
##  [37]  -65.745850 -131.784179   93.069091 -101.017089   -3.961915  -72.916015
##  [43]  -41.815918   25.428955   72.283936  -13.406983   31.087891  -53.264893
##  [49]   27.587891  -23.765869  -23.312989   53.347901   33.392090   -1.051026
##  [55]   19.922119  -41.241211   10.203125   31.861084  -28.232177   37.128173
##  [61]    4.850831    1.828125  -28.171143   36.444092  -15.571045  -22.697022
##  [67]    7.749024   12.393066   44.088867    6.237061   18.860107   -6.953125
##  [73]  -11.661865    4.103027   -6.351074   -8.183105   25.163086   -1.220948
##  [79]   -5.745117  -27.675049  -35.148925   19.423095   -9.113037   -2.623047
##  [85]   36.380127   44.352784   14.074218    1.446778   -8.648926   18.792969
##  [91]   -2.447022   11.743164  -11.487060  -62.988037   31.477050    4.550049
##  [97]   19.839844   -2.790039    8.910156   28.110840   -6.012940   12.094971
## [103]    0.093018   -7.970948   26.075928   32.538086  164.120117  193.114991
## [109]  -43.402100   65.930908   15.452149  -39.548096   55.636963  -12.946045
## [115]  -65.718994 -141.956787   29.508789   86.636230  -15.365234 -104.795899
## [121]   15.747071   -4.145020   77.583984  -18.776123   51.387207   44.256836
## [127]  -13.109863   17.429932  -10.172119    8.899170  -51.994141    0.682129
## [133]  -12.506104   58.233155   68.031982    2.630859  -58.599853   -5.738037
## [139]   10.885986    1.766846  -21.071778  -38.318115   12.799073   70.260986
## [145]  -26.808106   30.727051  -16.437012  -67.624023  -35.190918  -20.219971
## [151]    1.234864  -56.447998   50.134033    1.209961
```

```
mean(Diff_of_CHN_SP) #1.339084
```

```
## [1] 1.339084
```

```
#The RMSE that we get is 96.55912 which is alot larger than the number 1.339084.
#Therefore, this might not be a good model fit.
#Hence, we will start improve the model using the backward stepwise regression.

#Performing a stepwise regression
step(fit2, direction = "forward")
```

```
## Start:  AIC=1399.36
## o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3
```

```
##
## Call:
## lm(formula = o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3, data = train)
##
## Coefficients:
## (Intercept)           c1           c2           c3           u1           u2
##   1.785e+03   -2.165e-02    2.988e-04    5.680e-02    9.499e+01   -1.251e+01
##          u3           o2           o3
##  -2.371e+02    2.215e-01   -3.245e-02
```

```r
#It doesn't give me any additional information

library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
fit2 <- lm(o1 ~ c1 + c2 + c3 + u1 + u2 + u3 + o2 + o3, data=train)
bptest(fit2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fit2
## BP = 27.612, df = 8, p-value = 0.000554
```