Data_fest

Mei Ye Bao

2020-06-13

Data Import

```r
library(PerformanceAnalytics)
```

```
## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend
```

```r
library(quantmod)
```

```
## Warning: package 'quantmod' was built under R version 3.6.2

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```r
library(fpp2)
```

```
## Loading required package: ggplot2

## Loading required package: forecast

## Warning: package 'forecast' was built under R version 3.6.2

## Loading required package: fma

## Loading required package: expsmooth
```

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- t
```

```
## v tibble  2.1.3      v purrr   0.3.2
## v tidyr   1.0.0      v dplyr   0.8.3
## v readr   1.3.1      v stringr 1.4.0
## v tibble  2.1.3      v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------- tidyvers
## x dplyr::filter() masks stats::filter()
## x dplyr::first()  masks xts::first()
## x dplyr::lag()    masks stats::lag()
## x dplyr::last()   masks xts::last()
```

```r
library(ggplot2)
library(stringr)
library(xlsx)
library(nlme)
```

```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse

## The following object is masked from 'package:forecast':
##
##     getResponse
```

```r
library(lme4)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

##
## Attaching package: 'lme4'

## The following object is masked from 'package:nlme':
##
##     lmList
```

```r
covid19_url <- 'https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/ecdc/full_data.
covid19_raw <- read.csv(covid19_url)
#covid19_raw$location == "China"


#China
China <- covid19_raw  %>% filter(location == "China") %>% select(date, new_cases)

#Canada
Canada <- covid19_raw  %>% filter(location == "Canada")  %>% select(date, new_cases)

#USA
USA <- covid19_raw  %>% filter(location == "United States")  %>% select(date, new_cases)
```

```
#JingDong daily
dowjones <- new.env()
JD <- getSymbols("JD", env = dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRUE)

## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/JD?
## period1=-2208988800&period2=1592092800&interval=1d&events=div&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/JD?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/JD?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'
JD_dailyData <- apply.daily(dowjones$JD$JD.Close,last)
#glimpse(JD_dailyData)

#Amazon
AMZN <- getSymbols("AMZN", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRU

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/AMZN?
## period1=-2208988800&period2=1592092800&interval=1d&events=div&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/AMZN?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/AMZN?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

AMZN_dailyData <- apply.daily(dowjones$AMZN$AMZN.Close,last)

#Shopify
SHOP <- getSymbols("SHOP", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRU

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/SHOP?
```

```
## period1=-2208988800&period2=1592092800&interval=1d&events=div&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/SHOP?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/SHOP?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'
```

```r
SHOP_dailyData <- apply.daily(dowjones$SHOP$SHOP.Close,last)

#Alibaba
BABA <- getSymbols("BABA", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRUE
```

```
## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/BABA?
## period1=-2208988800&period2=1592092800&interval=1d&events=div&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/BABA?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query1.finance.yahoo.com/v7/finance/download/BABA?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'
```

```r
BABA_dailyData <- apply.daily(dowjones$BABA$BABA.Close,last)

#PDD
PDD <- getSymbols("PDD", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRUE)
```

```
## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/PDD?
## period1=-2208988800&period2=1592092800&interval=1d&events=div&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/PDD?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'

## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader
## on 'https://query2.finance.yahoo.com/v7/finance/download/PDD?
## period1=-2208988800&period2=1592092800&interval=1d&events=split&crumb=ErI4aPxGb2o'
```

```r
PDD_dailyData <- apply.daily(dowjones$PDD$PDD.Close,last)

#Ebay
EBAY <- getSymbols("EBAY", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRUE
EBAY_dailyData <- apply.daily(dowjones$EBAY$EBAY.Close,last)
```

```r
#WMT
WMT <- getSymbols("WMT", env=dowjones, src="yahoo", from="2019-12-31", to ="2020-06-01", adjust = TRUE)
WMT_dailyData <- apply.daily(dowjones$WMT$WMT.Close,last)
```

Clean Data

```r
dailyData <- data.frame(JD_dailyData, AMZN_dailyData, SHOP_dailyData, BABA_dailyData, PDD_dailyData, EB
#dailyData


# China
glimpse(China)
```

```
## Observations: 167
## Variables: 2
## $ date      <fct> 2019-12-31, 2020-01-01, 2020-01-02, 2020-01-03, 2020...
## $ new_cases <int> 27, 0, 0, 17, 0, 15, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
```

```r
PrepStockChina <- data.frame(JD_dailyData, BABA_dailyData, PDD_dailyData)
Date3 <- rownames(PrepStockChina)
PrepStockChina <- cbind(Date3,PrepStockChina)
StockChina <- left_join(China, PrepStockChina, by = c("date" = "Date3"))
```

```
## Warning: Column `date`/`Date3` joining factors with different levels,
## coercing to character vector
```

```r
# StockChina <- StockChina %>%
#   mutate(Date = as.Date(Date))%>% complete(Date = seq.Date(min(Date), as.Date("2020-06-12"), by="day"
# StockChina <- cbind(China, StockChina)

StockChina <- na.omit(StockChina)
glimpse(StockChina)
```

```
## Observations: 104
## Variables: 5
## $ date       <chr> "2019-12-31", "2020-01-02", "2020-01-03", "2020-01-...
## $ new_cases  <int> 27, 0, 17, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 151, 140, ...
## $ JD.Close   <dbl> 35.23, 37.73, 37.99, 38.00, 38.32, 38.30, 38.90, 39...
## $ BABA.Close <dbl> 212.10, 219.77, 217.00, 216.64, 217.63, 218.00, 221...
## $ PDD.Close  <dbl> 37.82, 41.23, 40.89, 40.09, 41.21, 40.50, 39.66, 38...
```

```r
#Canada
PrepStockCanada <- data.frame(SHOP_dailyData)
Date <- rownames(PrepStockCanada)
PrepStockCanada <- cbind(Date,PrepStockCanada)
StockCanada <- left_join(Canada, PrepStockCanada, by = c("date" = "Date"))
```

```
## Warning: Column `date`/`Date` joining factors with different levels,
## coercing to character vector
```

```r
StockCanada <- na.omit(StockCanada)

#US
PrepStockUS <- data.frame(AMZN_dailyData,EBAY_dailyData,WMT_dailyData)
```

```
Date2 <- rownames(PrepStockUS)
PrepStockUS <- cbind(Date2,PrepStockUS)
StockUS <- left_join(USA, PrepStockUS, by = c("date" = "Date2"))
```

```
## Warning: Column `date`/`Date2` joining factors with different levels,
## coercing to character vector
```
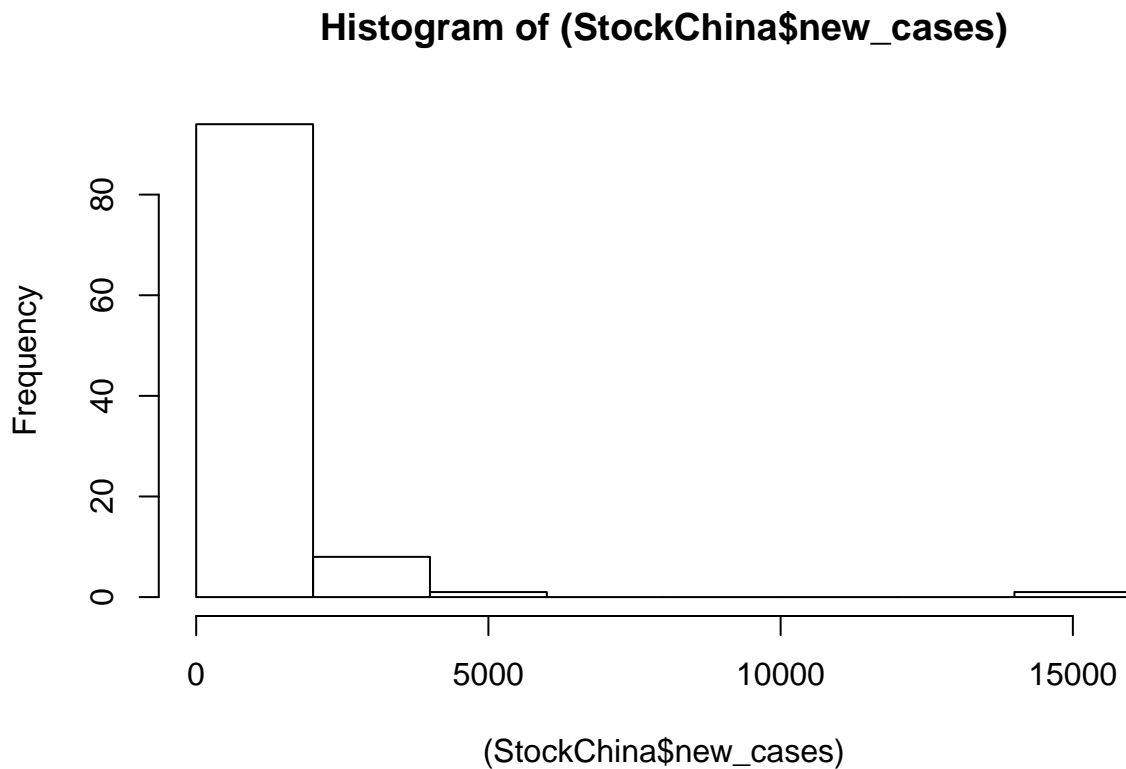
```
StockUS <- na.omit(StockUS)
#StockUS
```
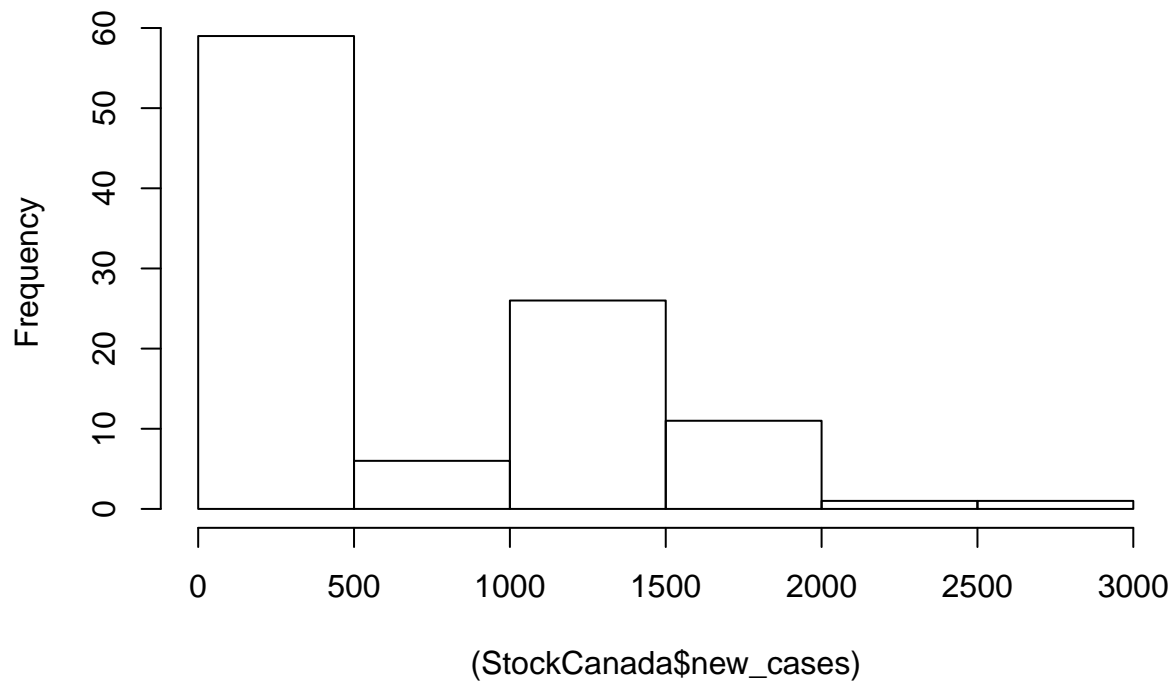
Analysis

Visualizing the data

Before considering models, a histogram of our data is plotted to see the distribution of the data.
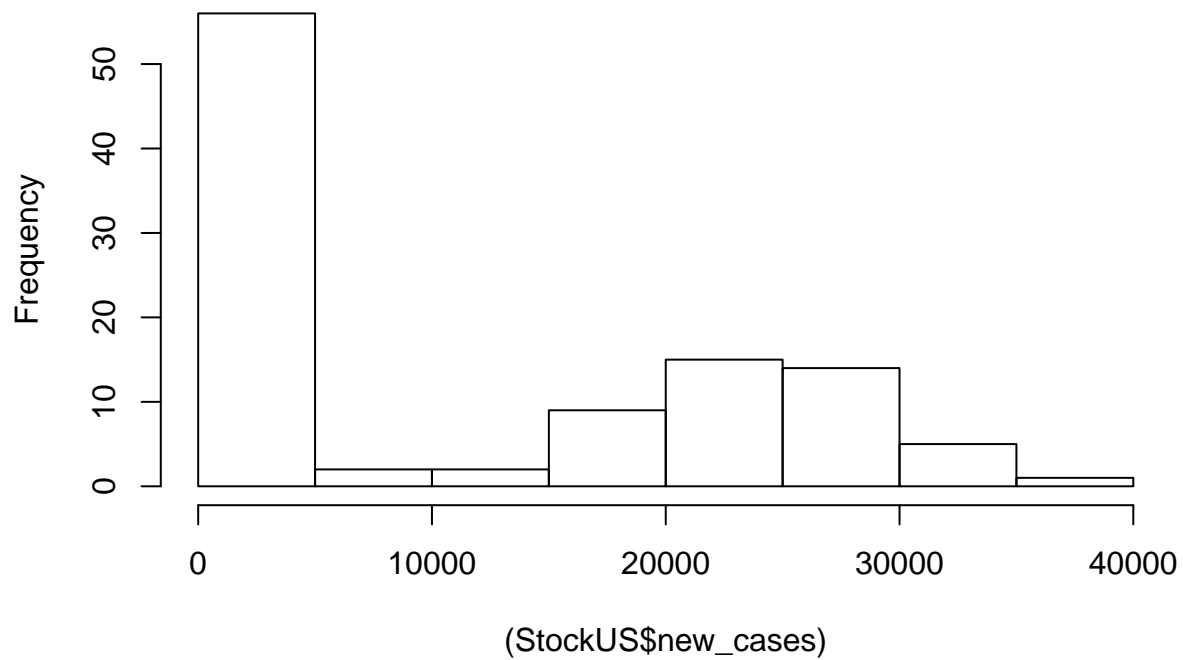
```
hist((StockChina$new_cases))
```


Histogram of (StockChina$new_cases)

```
hist((StockCanada$new_cases))
```

Histogram of (StockCanada$new_cases)

(StockCanada$new_cases)

```
hist((StockUS$new_cases))
```

Histogram of (StockUS$new_cases)

(StockUS$new_cases)

```
shapiro.test(StockChina$new_cases)
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  StockChina$new_cases
## W = 0.36197, p-value < 2.2e-16
```

```r
shapiro.test(StockCanada$new_cases)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  StockCanada$new_cases
## W = 0.78361, p-value = 4.528e-11
```

```r
shapiro.test(StockUS$new_cases)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  StockUS$new_cases
## W = 0.77321, p-value = 2.255e-11
```

```r
# The data is normal if the p-value is above 0.05. So we now know our variable is normally distributed.
# Thus, not normal
```

Stocks and COVID-19 cases in China

```r
summary(StockChina)
```

```
##      date             new_cases          JD.Close        BABA.Close
##  Length:104         Min.   :    0.0   Min.   :35.23   Min.   :176.3
##  Class :character   1st Qu.:    6.0   1st Qu.:39.56   1st Qu.:197.9
##  Mode  :character   Median :   49.5   Median :41.26   Median :206.6
##                     Mean   :  596.5   Mean   :42.35   Mean   :206.2
##                     3rd Qu.:  208.2   3rd Qu.:44.02   3rd Qu.:216.7
##                     Max.   :15141.0   Max.   :55.53   Max.   :230.5
##    PDD.Close
##  Min.   :31.77
##  1st Qu.:36.02
##  Median :38.20
##  Mean   :42.27
##  3rd Qu.:47.59
##  Max.   :68.70
```

```r
plot(JD_dailyData)
```

JD_dailyData   2019−12−31 / 2020−05−29
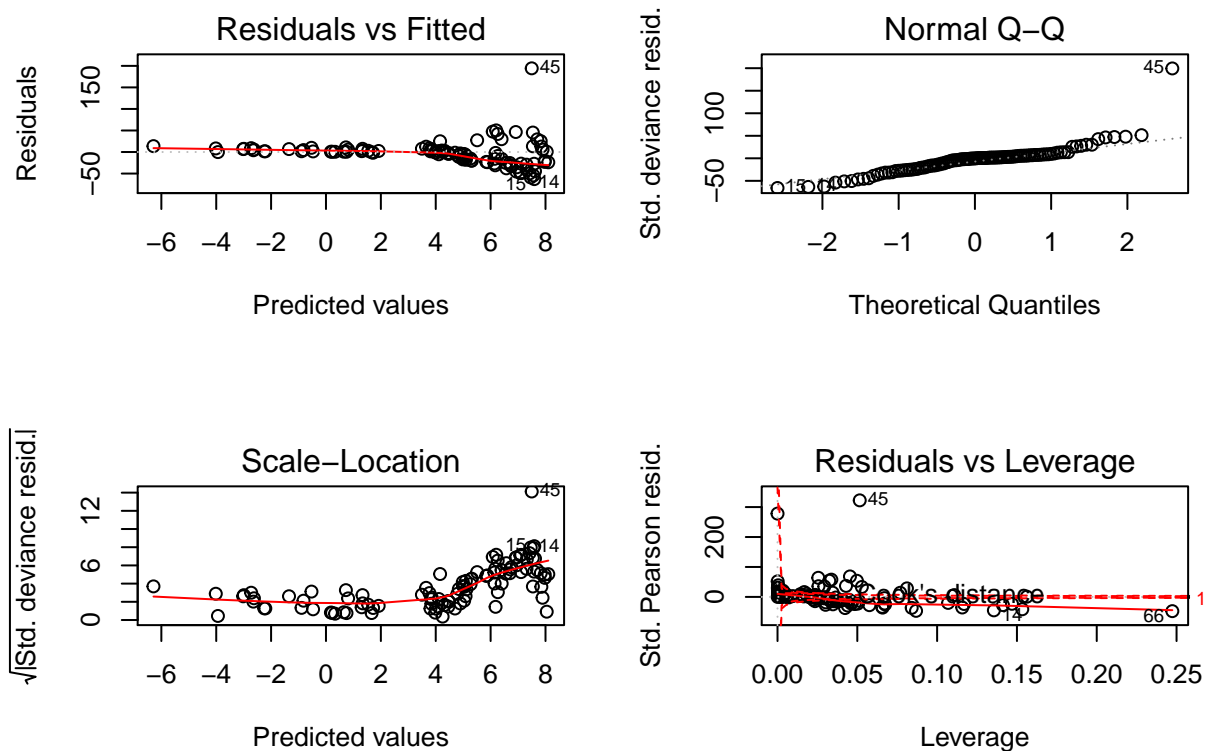
```
#Stocks in China with COVID-19 cases in China

model_China <- glm(new_cases ~ JD.Close + BABA.Close + PDD.Close, family = poisson, data = StockChina)


# Checking Model Assumptions
par(mfrow = c(2, 2))
plot(model_China)
```

Residuals vs Fitted

Residuals

Normal Q–Q

Std. deviance resid.

Scale–Location

√|Std. deviance resid.|

Residuals vs Leverage

Std. Pearson resid.

Cook's distance

```r
qqnorm(residuals(model_China))

# Checking three companies' stock in relationship to COVID-19 new cases in China
summary(model_China)
```

```
## 
## Call:
## glm(formula = new_cases ~ JD.Close + BABA.Close + PDD.Close,
##     family = poisson, data = StockChina)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -62.924  -19.992   -0.168    5.776  194.269
## 
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.5237428  0.1381013  -54.48   <2e-16 ***
## JD.Close     0.1760270  0.0025987   67.74   <2e-16 ***
## BABA.Close   0.1051885  0.0005653  186.08   <2e-16 ***
## PDD.Close   -0.4156188  0.0026185 -158.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for poisson family taken to be 1)
## 
##     Null deviance: 202508  on 103  degrees of freedom
## Residual deviance:  92962  on 100  degrees of freedom
## AIC: 93537
## 
## Number of Fisher Scoring iterations: 8
```

All three companies showed significant interation with COVID-19

```
library(pbkrtest)
library(sjPlot)
tab_model(model_China)
```

new_cases

Predictors

Incidence Rate Ratios

CI

p

(Intercept)

0.00

$0.00 - 0.00$

<0.001

JD.Close

1.19

$1.19 - 1.20$

<0.001

BABA.Close

1.11

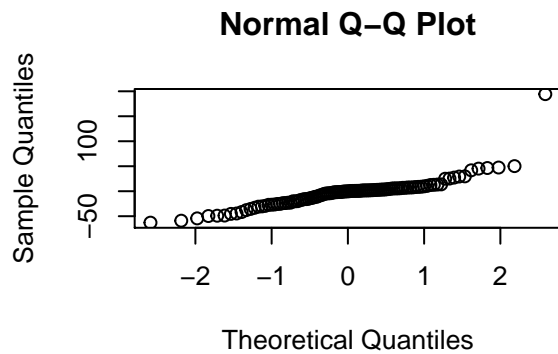$1.11 - 1.11$

<0.001

PDD.Close

0.66

$0.66 - 0.66$

<0.001

Observations

104

R2 Nagelkerke

1.000

Normal Q–Q Plot

Stocks and COVID-19 cases in Canada

```
#Stocks in Canada with COVID-19 cases in Canada

model_Canada <- glm(new_cases ~ SHOP_dailyData, data = StockCanada)
model2_Canada <- glm(new_cases ~ SHOP_dailyData, family = poisson,data = StockCanada)
```

```
## Compare family: Normal VS Poisson
summary(model_Canada)
```

```
##
## Call:
## glm(formula = new_cases ~ SHOP_dailyData, data = StockCanada)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -658.1   -432.8   -272.7    377.7   1765.8
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -920.2132    238.8822   -3.852 0.000205 ***
## SHOP_dailyData     2.9055      0.4461    6.513 2.83e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 345258)
##
##     Null deviance: 49860752  on 103  degrees of freedom
## Residual deviance: 35216315  on 102  degrees of freedom
## AIC: 1625.3
##
## Number of Fisher Scoring iterations: 2
```

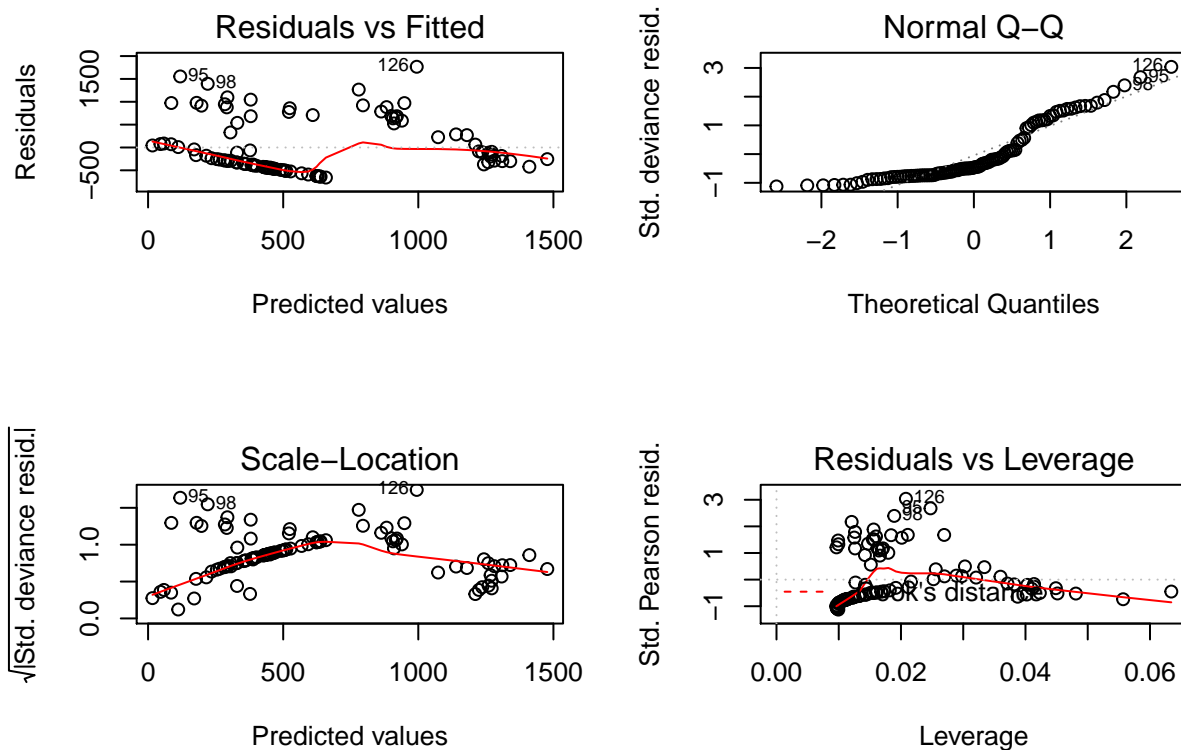```
summary(model2_Canada)
```

```
##
## Call:
## glm(formula = new_cases ~ SHOP_dailyData, family = poisson, data = StockCanada)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -33.18   -28.09   -16.45     16.27     59.30
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.998e+00  1.744e-02   229.2   <2e-16 ***
## SHOP_dailyData 4.259e-03 2.819e-05   151.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 97867  on 103  degrees of freedom
## Residual deviance: 75661  on 102  degrees of freedom
## AIC: 76181
##
## Number of Fisher Scoring iterations: 6
### Comparing AIC, model 1 has a much smaller AIC, so model 1 is selected as the better model
```

```
# Checking Model Assumptions
par(mfrow = c(2, 2))
plot(model_Canada)
```



```
# Checking three companies' stock in relationship to COVID-19 new cases in China
summary(model_Canada)
```

```
##
## Call:
## glm(formula = new_cases ~ SHOP_dailyData, data = StockCanada)
##
## Deviance Residuals:
```

13

```
##    Min      1Q  Median      3Q     Max
## -658.1  -432.8  -272.7   377.7  1765.8
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -920.2132   238.8822  -3.852 0.000205 ***
## SHOP_dailyData   2.9055     0.4461   6.513 2.83e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 345258)
##
##     Null deviance: 49860752  on 103  degrees of freedom
## Residual deviance: 35216315  on 102  degrees of freedom
## AIC: 1625.3
##
## Number of Fisher Scoring iterations: 2
### All three companies showed significant interation with COVID-19
tab_model(model_Canada)
```

new_cases

Predictors

Estimates

CI

p

(Intercept)

-920.21

-1388.41 – -452.01

<0.001

SHOP_dailyData

2.91

2.03 – 3.78

<0.001

Observations

104
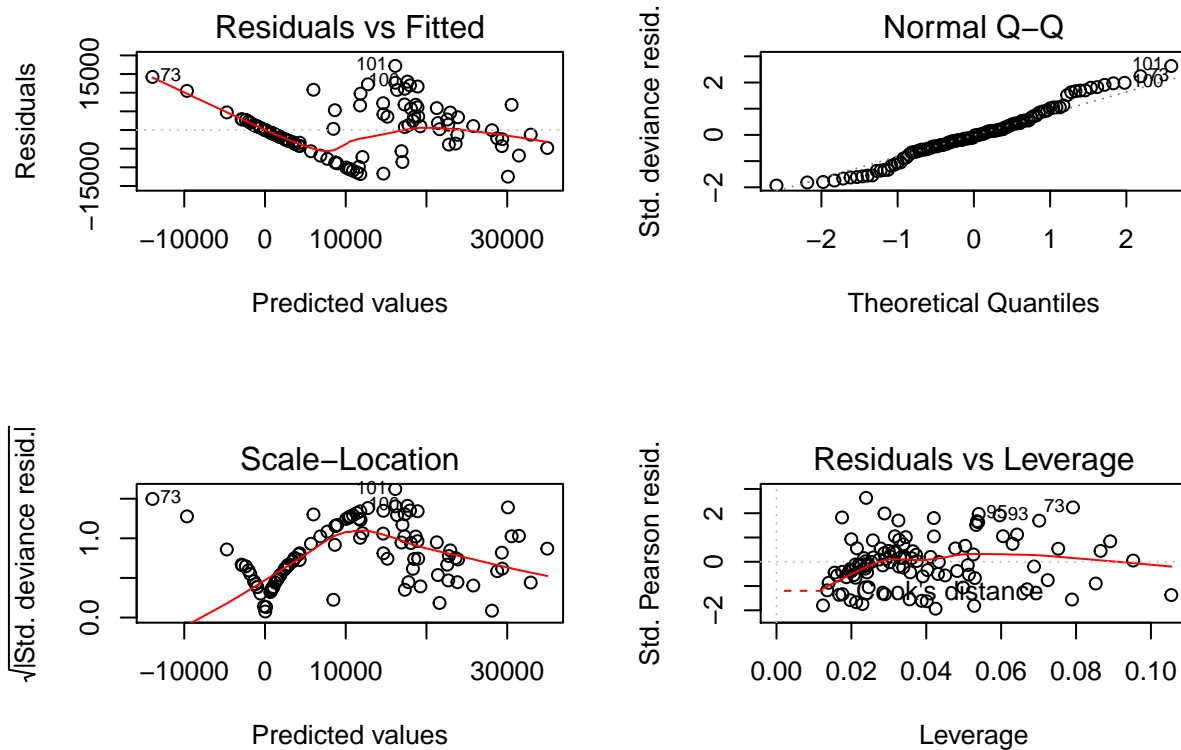
R2 Nagelkerke

1.000

Stocks and COVID-19 cases in US

```
#Stocks in China with COVID-19 cases in US

model_US <- glm(new_cases ~ AMZN_dailyData +EBAY_dailyData + WMT_dailyData, data = StockUS)
```

```
# Checking Model Assumptions
par(mfrow = c(2, 2))
plot(model_US)
```

Residuals vs Fitted

Normal Q–Q

Scale–Location

Residuals vs Leverage

```
# Checking three companies' stock in relationship to COVID-19 new cases in China
summary(model_US)
```

```
##
## Call:
## glm(formula = new_cases ~ AMZN_dailyData + EBAY_dailyData + WMT_dailyData,
##     data = StockUS)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -12503.6   -3657.2   -394.9   3555.9   17185.2
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.164e+05  1.589e+04  -7.325 6.23e-11 ***
## AMZN_dailyData  4.343e+01  6.432e+00   6.752 9.73e-10 ***
## EBAY_dailyData -1.797e+03  2.566e+02  -7.003 2.95e-10 ***
## WMT_dailyData   8.607e+02  1.914e+02   4.498 1.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 43703541)
##
##     Null deviance: 1.5659e+10  on 103   degrees of freedom
## Residual deviance: 4.3704e+09  on 100   degrees of freedom
```

```
## AIC: 2130.7
##
## Number of Fisher Scoring iterations: 2
### All three companies showed significant interation with COVID-19
tab_model(model_US)
```

new_cases

Predictors

Estimates

CI

p

(Intercept)

-116387.11

-147529.16 – -85245.07

<0.001

AMZN_dailyData

43.43

30.82 – 56.04

<0.001

EBAY_dailyData

-1796.90

-2299.84 – -1293.96

<0.001

WMT_dailyData

860.70

485.62 – 1235.78

<0.001

Observations

104

R2 Nagelkerke

1.000