R Notebook

Code

Hide

```
library(quantmod)
library(tidyverse)
```

```
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
── Attaching packages ─────────────────────────────── tidyverse 1.3.1 ──
✓ ggplot2 3.3.5     ✓ purrr   0.3.4
✓ tibble  3.1.2     ✓ dplyr   1.0.7
✓ tidyr   1.1.3     ✓ stringr 1.4.0
✓ readr   1.4.0     ✓ forcats 0.5.1
── Conflicts ────────────────────────────────── tidyverse_conflicts() ──
x dplyr::filter() masks stats::filter()
x dplyr::first()  masks xts::first()
x dplyr::lag()    masks stats::lag()
x dplyr::last()   masks xts::last()
```

Hide

```
library(ggplot2)
```

Hide

```
google_data <- getSymbols(Symbols = "GOOG", src = "yahoo", from = Sys.Date() - 295
3, to = Sys.Date(), auto.assign = FALSE)
google_data <- Cl(google_data)
head(google_data)
```

```
           GOOG.Close
2013-06-10   443.4478
2013-06-11   438.2623
2013-06-12   434.3619
2013-06-13   436.8625
2013-06-14   435.8862
2013-06-17   441.4703
```

Hide

```
chart_Series(google_data, col = "black")
```

Hide

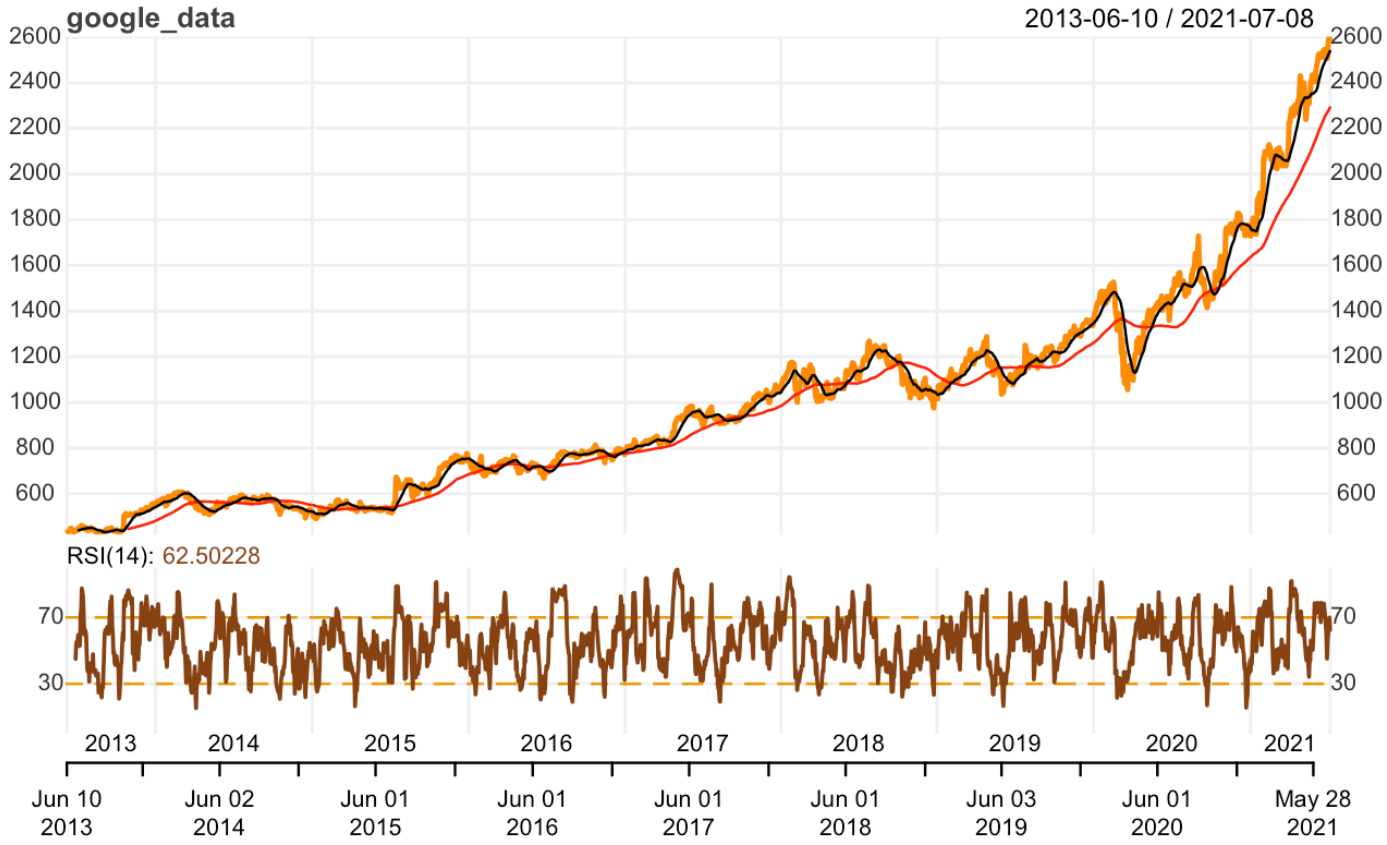
```
add_SMA(n = 100, on = 1, col = "red")
```

Hide

```
add_SMA(n = 20, on = 1, col = "black")
```

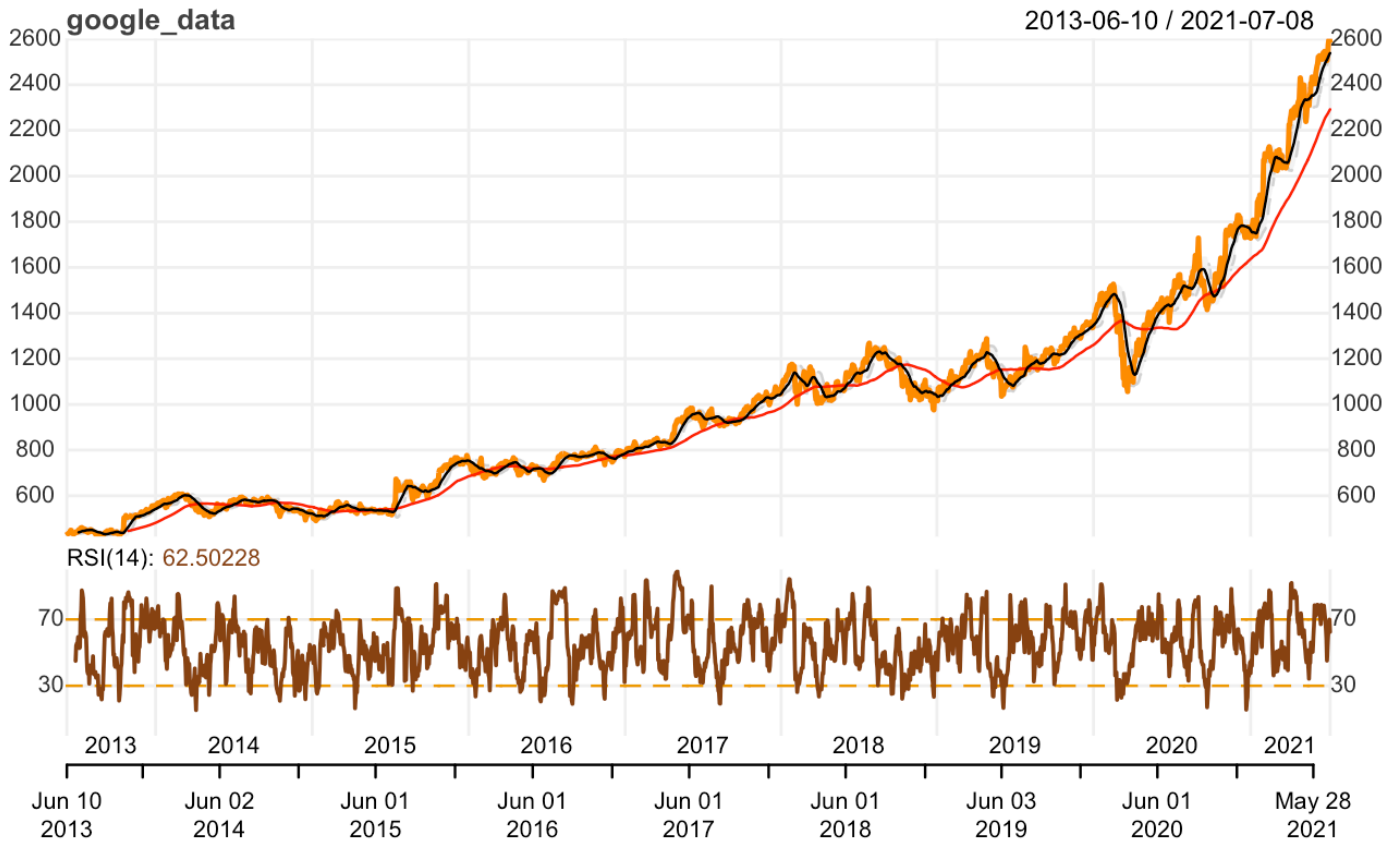
Hide

```
add_RSI(n = 14, maType = "SMA")
```



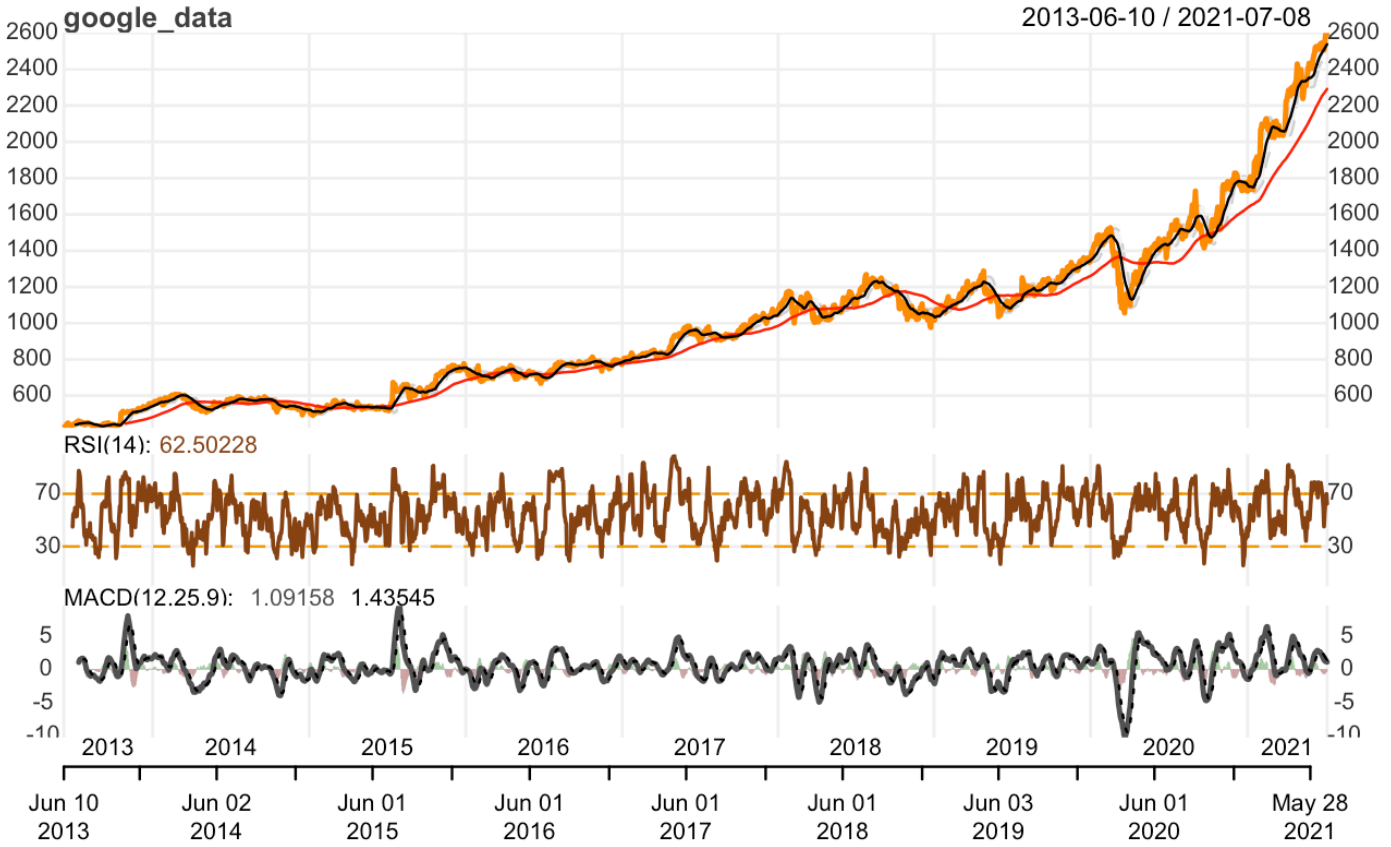
Hide

```
add_BBands(n = 20, maType = "SMA", sd = 1, on = -1)
```



Hide

```
add_MACD(fast = 12, slow = 25, signal = 9, maType = "SMA", histogram = TRUE)
```



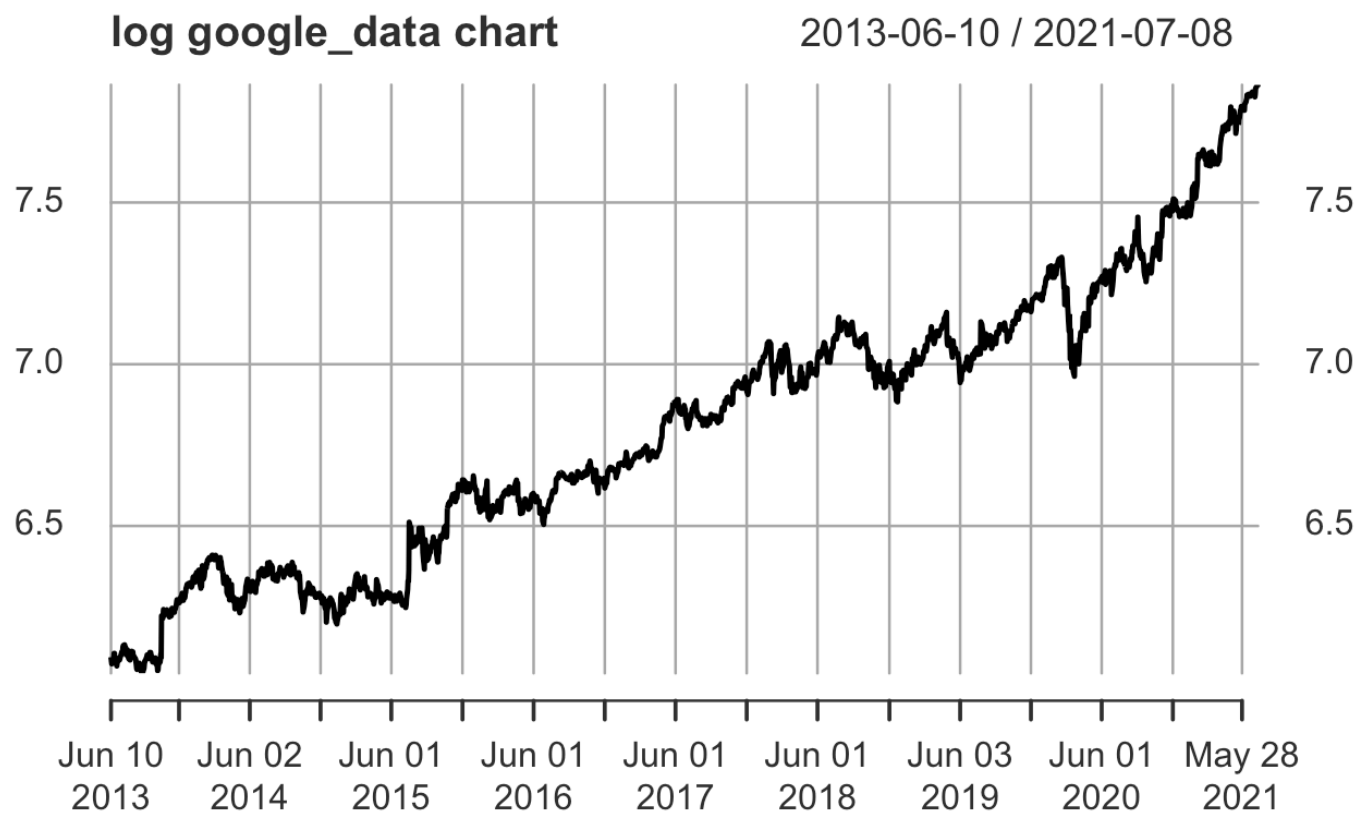
Hide

```
### Log tranformation stock data
google_log <- log(google_data)
head(google_log, n = 10)
```

```
           GOOG.Close
2013-06-10   6.094580
2013-06-11   6.082818
2013-06-12   6.073878
2013-06-13   6.079619
2013-06-14   6.077381
2013-06-17   6.090111
2013-06-18   6.106195
2013-06-19   6.106262
2013-06-20   6.088405
2013-06-21   6.084090
```

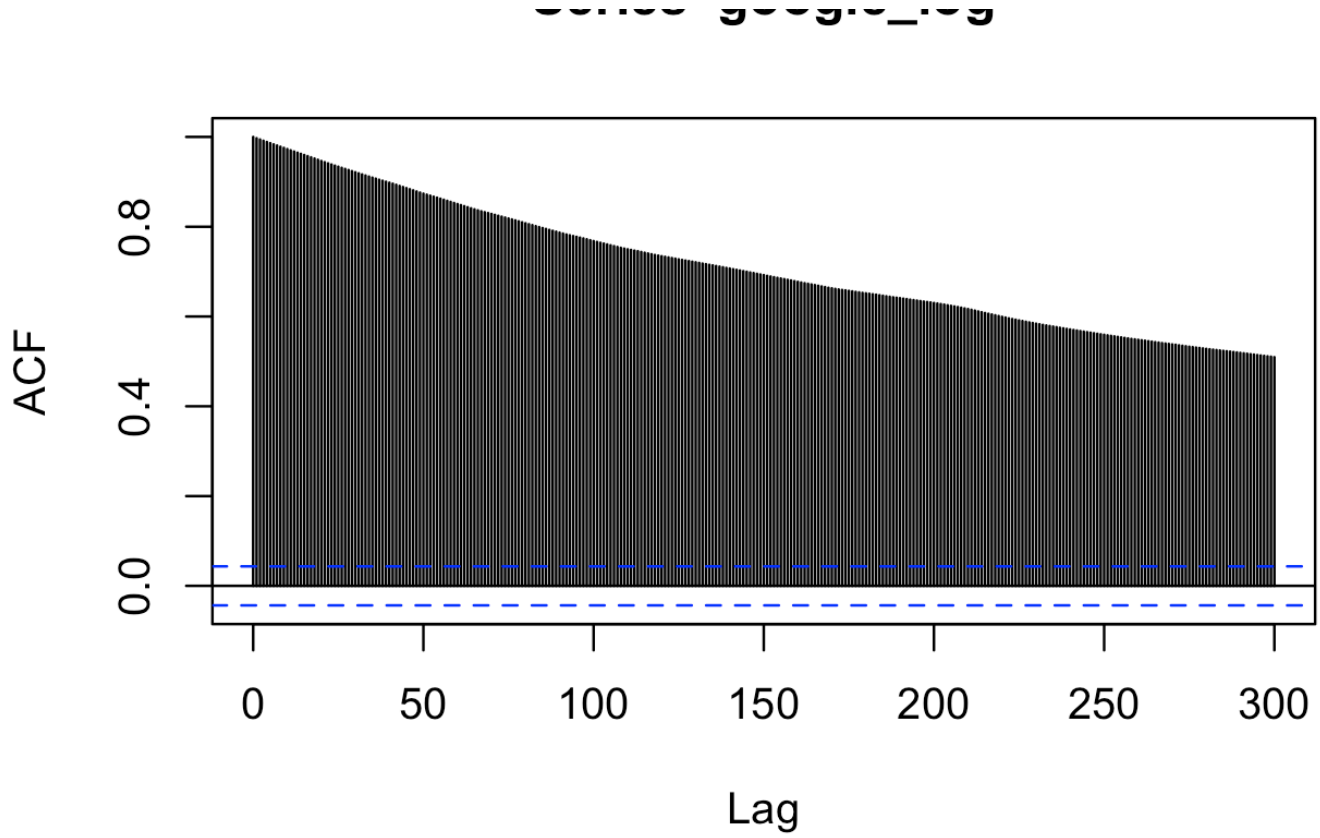
Hide

```
plot(google_log, main = "log google_data chart")
```



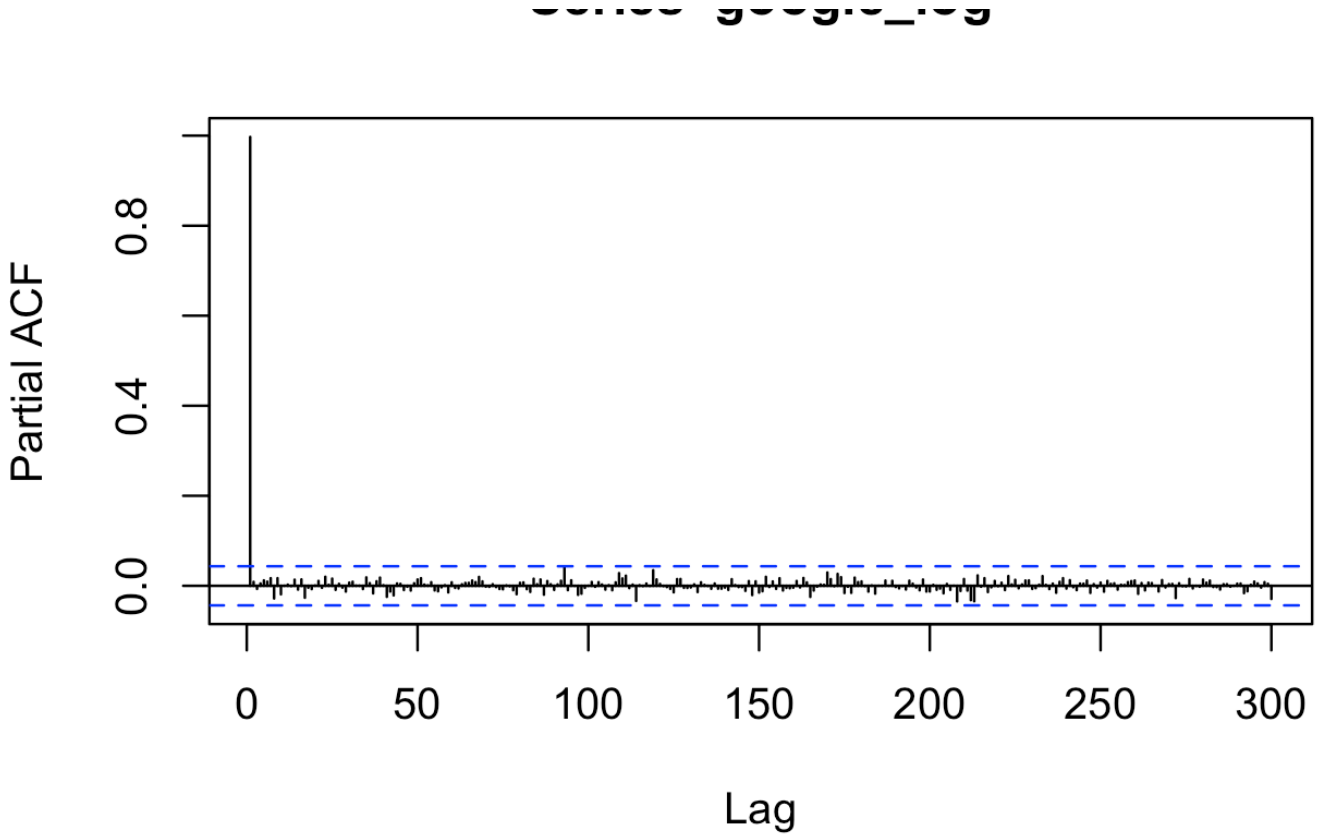
Hide

```
#autocorrelation function
acf_log <- acf(google_log, lag.max = 300)
```



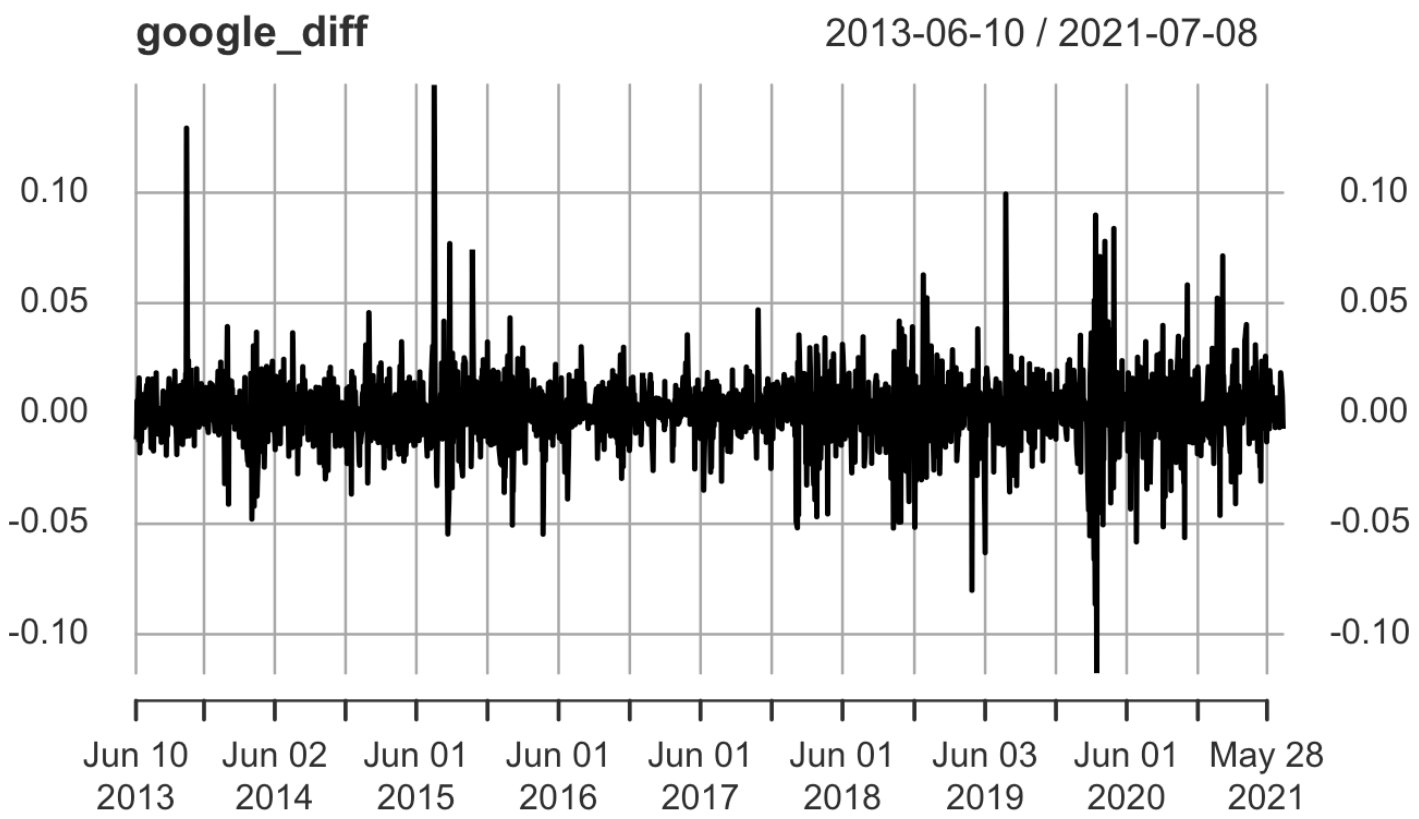
Hide

```
#partial autocorrelation
pacf_log <- pacf(google_log, lag.max = 300)
```



Hide

```
# difference logged data
google_diff <- diff(google_log, lag = 1)

google_diff <- na.locf(google_diff, na.rm = TRUE,
                       fromLast = TRUE)
plot(google_diff)
```



Hide

```
library(urca)
library(tseries)
adf <- adf.test(google_log, alternative = c("stationary", "explosive"),
                k = 0)
adf
```

```

    Augmented Dickey-Fuller Test

data:  google_log
Dickey-Fuller = -2.9438, Lag order = 0, p-value = 0.1787
alternative hypothesis: stationary
```

Hide

```
adf_diff <- adf.test(google_diff, alternative = c("stationary", "explosive"),
                k = 0)
```

```
Warning in adf.test(google_diff, alternative = c("stationary", "explosive"),  :
  p-value smaller than printed p-value
```

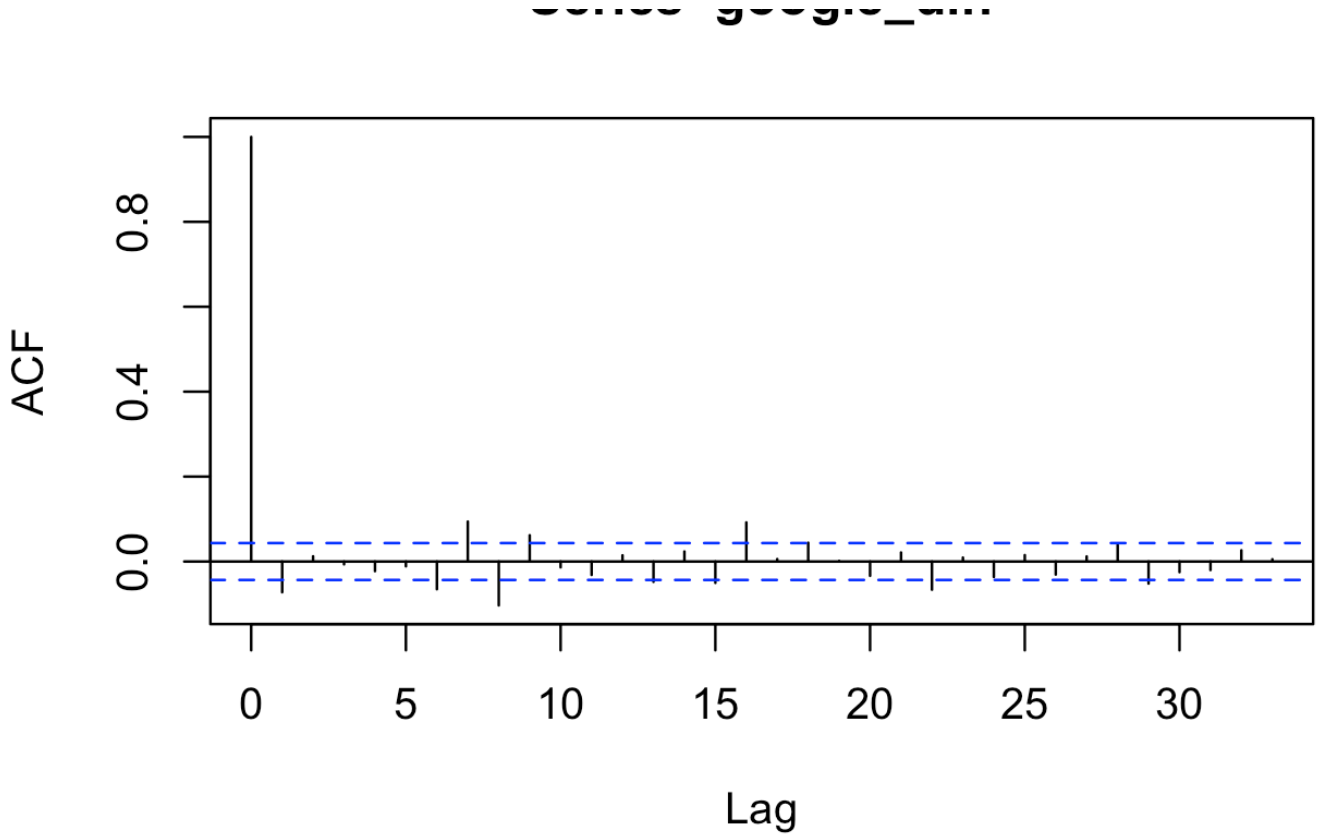
Hide

```
adf_diff
```

```
    Augmented Dickey-Fuller Test

data:  google_diff
Dickey-Fuller = -48.478, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

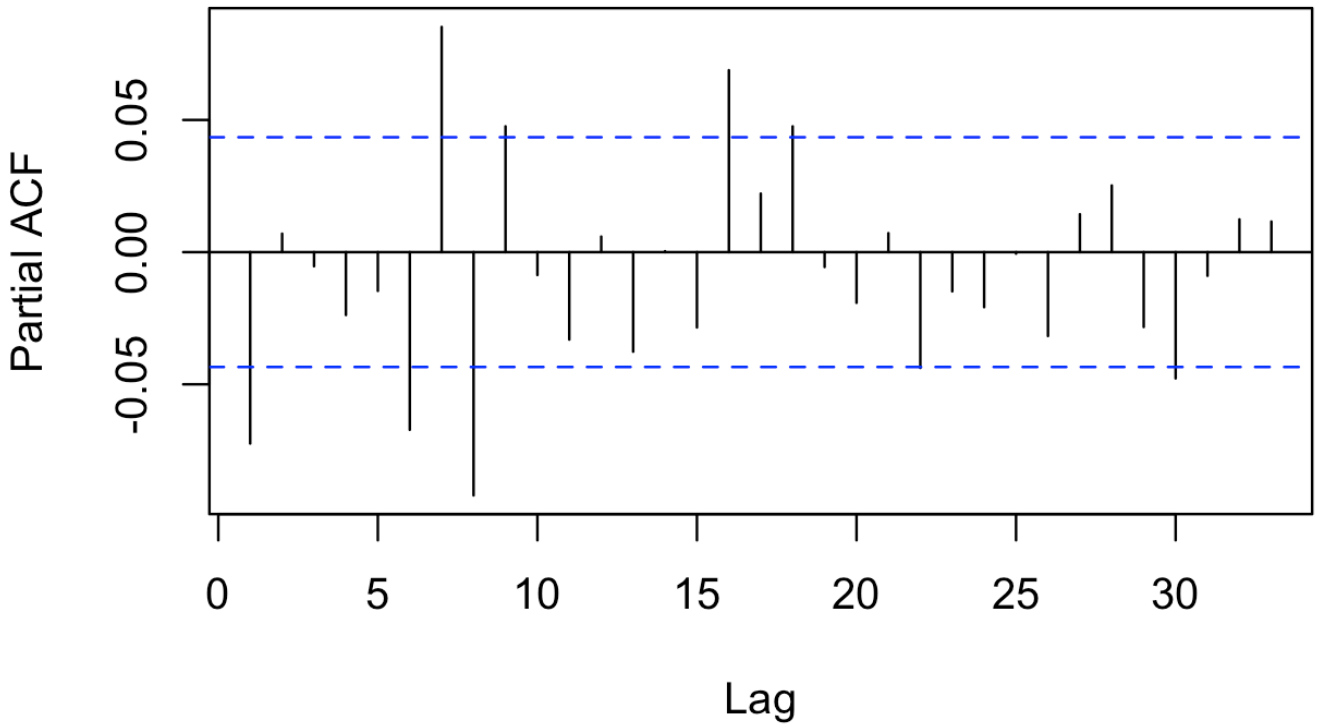
Hide

```
diff.acf <- acf(google_diff)
```



Hide

```
diff.pacf <- pacf(google_diff)
```



Hide

```
# splitting into train and test data
library(caTools)
train_data <- google_diff[1:1270]
```

Hide

```
library(forecast)
set.seed(123)
arima_model <- auto.arima(train_data, stationary = TRUE, ic = c("aicc", "aic", "bi
c"),
                          trace = TRUE)
```

```
 Fitting models using approximations to speed things up...

 ARIMA(2,0,2) with non-zero mean : -7194.721
 ARIMA(0,0,0) with non-zero mean : -7200.796
 ARIMA(1,0,0) with non-zero mean : -7200.642
 ARIMA(0,0,1) with non-zero mean : -7200.911
 ARIMA(0,0,0) with zero mean     : -7199.307
 ARIMA(1,0,1) with non-zero mean : -7198.623
 ARIMA(0,0,2) with non-zero mean : -7199.054
 ARIMA(1,0,2) with non-zero mean : -7201.028
 ARIMA(1,0,3) with non-zero mean : -7199.62
 ARIMA(0,0,3) with non-zero mean : -7197.133
 ARIMA(2,0,1) with non-zero mean : -7196.575
 ARIMA(2,0,3) with non-zero mean : Inf
 ARIMA(1,0,2) with zero mean     : -7197.717

 Now re-fitting the best model(s) without approximations...

 ARIMA(1,0,2) with non-zero mean : Inf
 ARIMA(0,0,1) with non-zero mean : -7200.91

 Best model: ARIMA(0,0,1) with non-zero mean
```

Hide

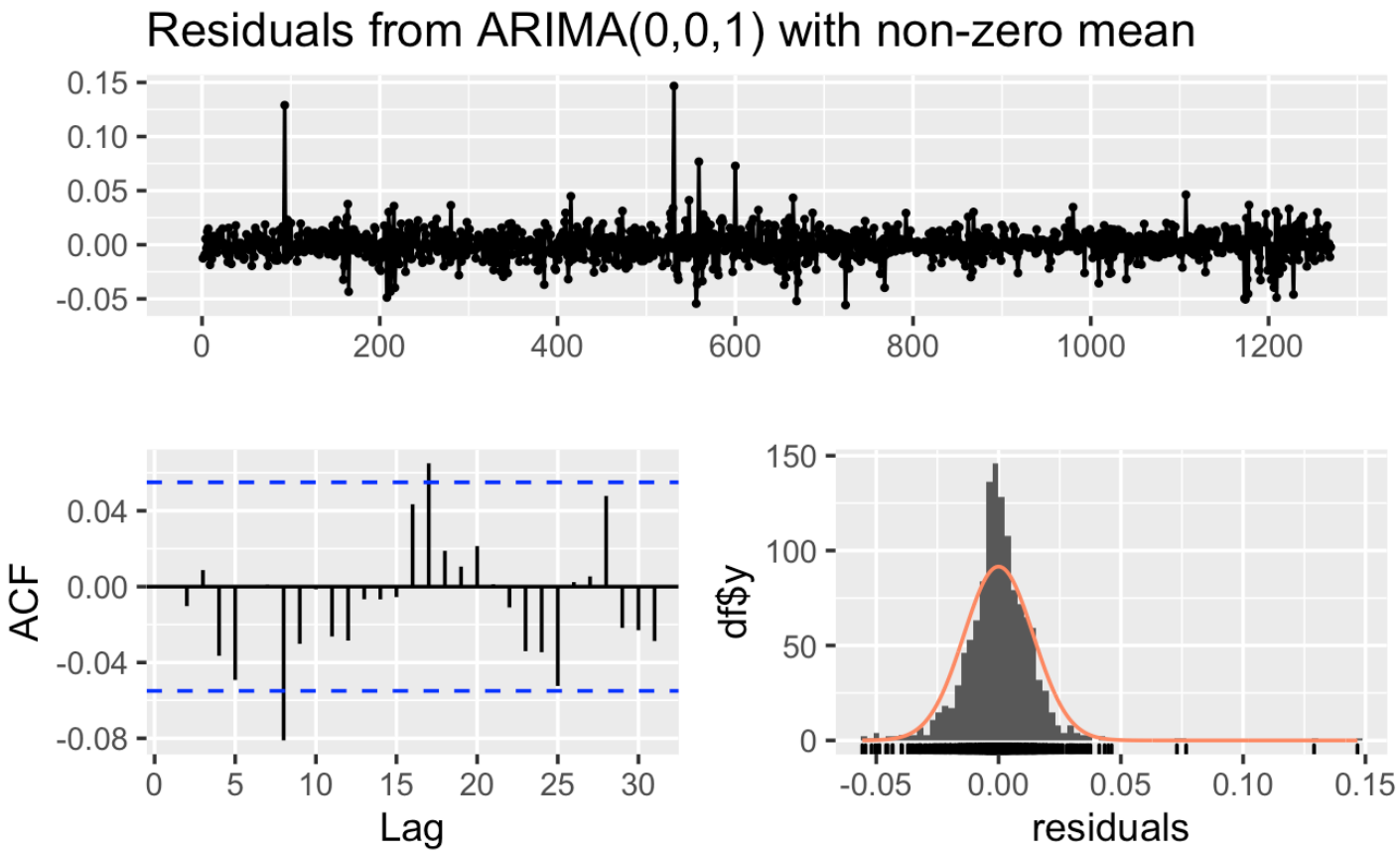
```
#summary for choosen best arima(p,d,q) model
#Autoregressive(p), lag(d), moving average(q)
summary(arima_model)
```

```
Series: train_data
ARIMA(0,0,1) with non-zero mean

Coefficients:
         ma1   mean
      0.0413  7e-04
s.e.  0.0284  4e-04

sigma^2 estimated as 0.0002012:  log likelihood=3603.46
AIC=-7200.93   AICc=-7200.91   BIC=-7185.49

Training set error measures:
                      ME       RMSE         MAE      MPE     MAPE
Training set 6.084837e-08 0.01417447 0.009626458 97.07964 137.4596
                   MASE          ACF1
Training set 0.6927495 -0.0004338891
```

Hide

```
#diagnostic checking
checkresiduals(arima_model)
```

```
    Ljung-Box test

data:  Residuals from ARIMA(0,0,1) with non-zero mean
Q* = 14.624, df = 8, p-value = 0.06688

Model df: 2.   Total lags used: 10
```



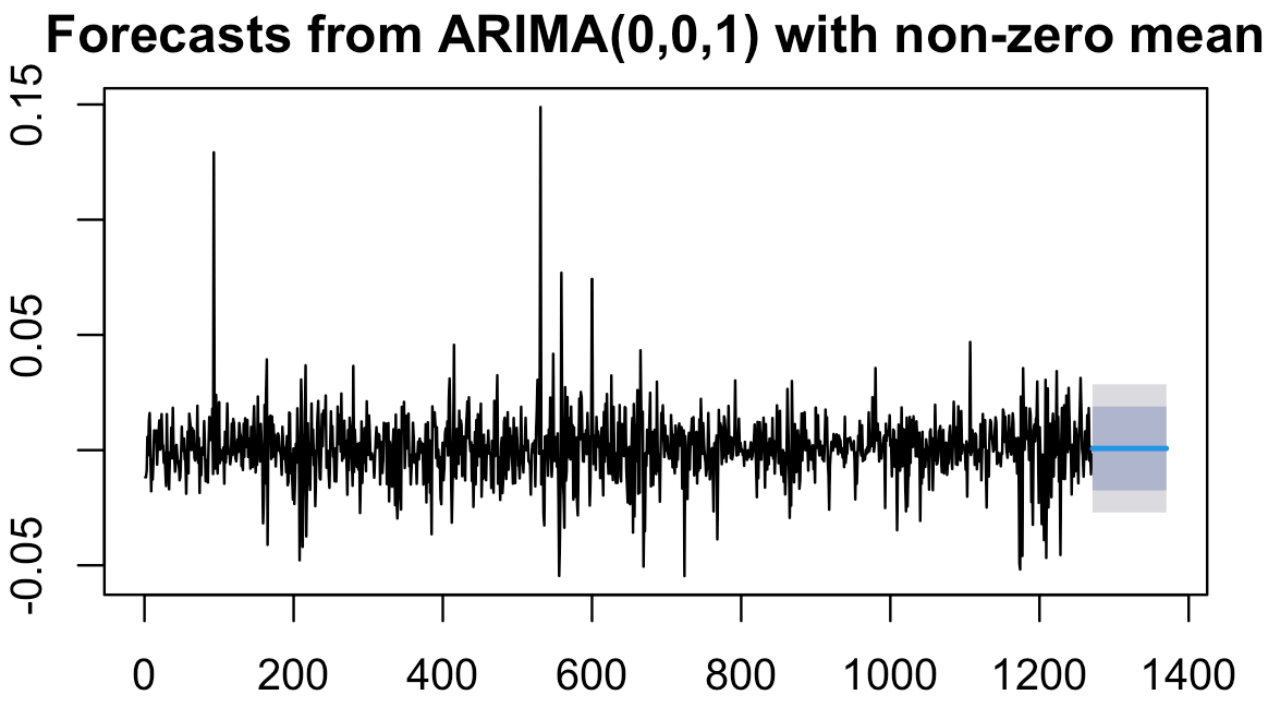
Hide

```
arima <- arima(train_data, order = c(0, 0, 1))
summary(arima)
```

```
Call:
arima(x = train_data, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.0413      7e-04
s.e.  0.0284      4e-04

sigma^2 estimated as 0.0002009:  log likelihood = 3603.46,  aic = -7200.93

Training set error measures:
                      ME       RMSE          MAE      MPE     MAPE
Training set 6.084837e-08 0.01417447 0.009626458 97.07964 137.4596
                   MASE          ACF1
Training set 0.6927495 -0.0004338891
```

Hide

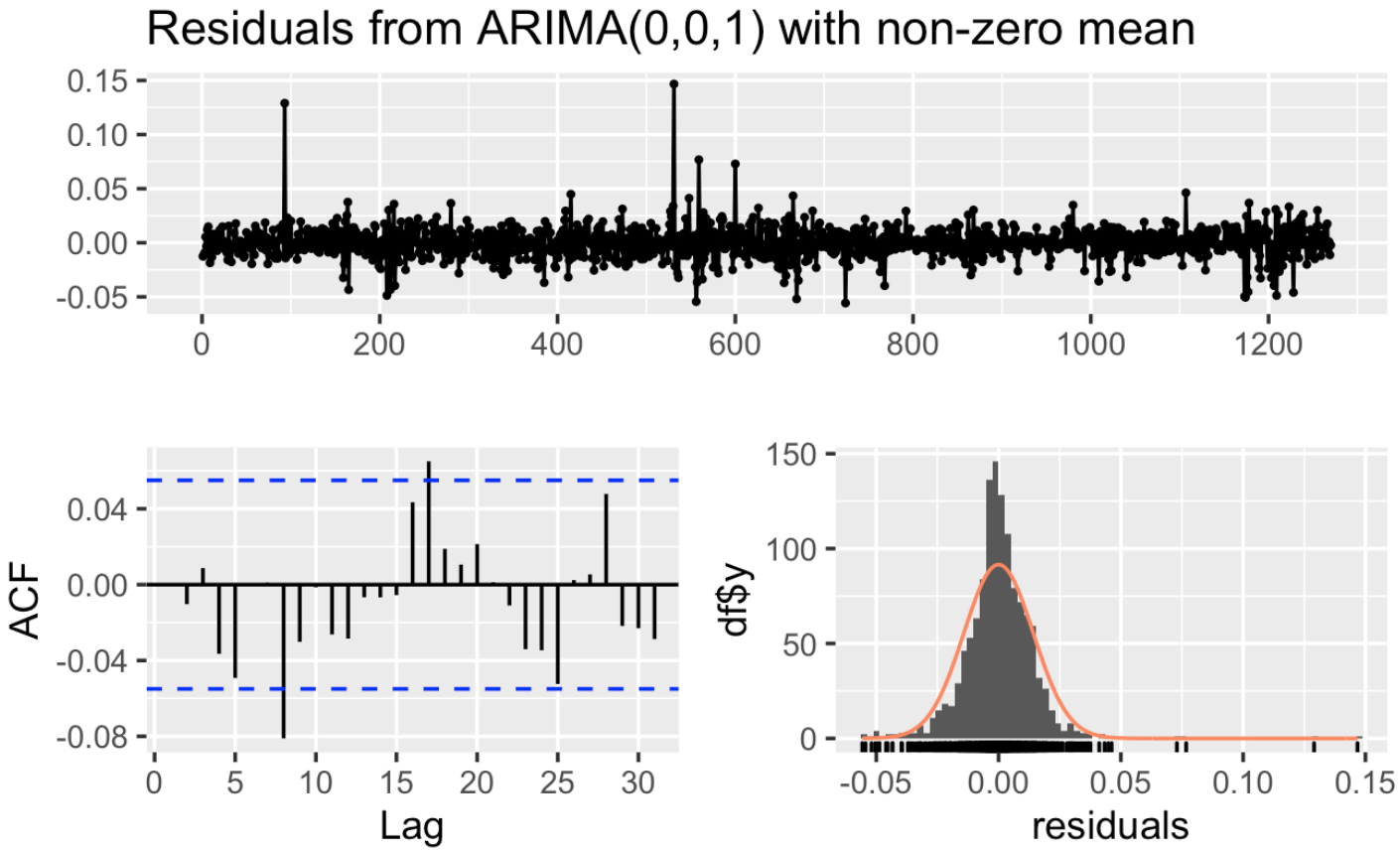
```
forecast1 <- forecast(arima, h = 100)
plot(forecast1)
```



Hide

```
checkresiduals(arima)
```

```
    Ljung-Box test

data:  Residuals from ARIMA(0,0,1) with non-zero mean
Q* = 14.624, df = 8, p-value = 0.06688

Model df: 2.   Total lags used: 10
```



Hide

```
arima <- arima(google_log[1:1270], order = c(0, 0, 1))
summary(arima)
```

```
Call:
arima(x = google_log[1:1270], order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.9672     6.5369
s.e.  0.0054     0.0077

sigma^2 estimated as 0.01972:  log likelihood = 689.74,  aic = -1373.48

Training set error measures:
                      ME      RMSE       MAE          MPE     MAPE
Training set 0.0001205015 0.1404198 0.1210119 -0.08534688 1.850192
                  MASE      ACF1
Training set 12.57917 0.9275646
```

Hide

```
forecast_ori <- forecast(arima, h = 200)
a <- ts(google_log)
forecast_ori %>% autoplot() + autolayer(a)
```