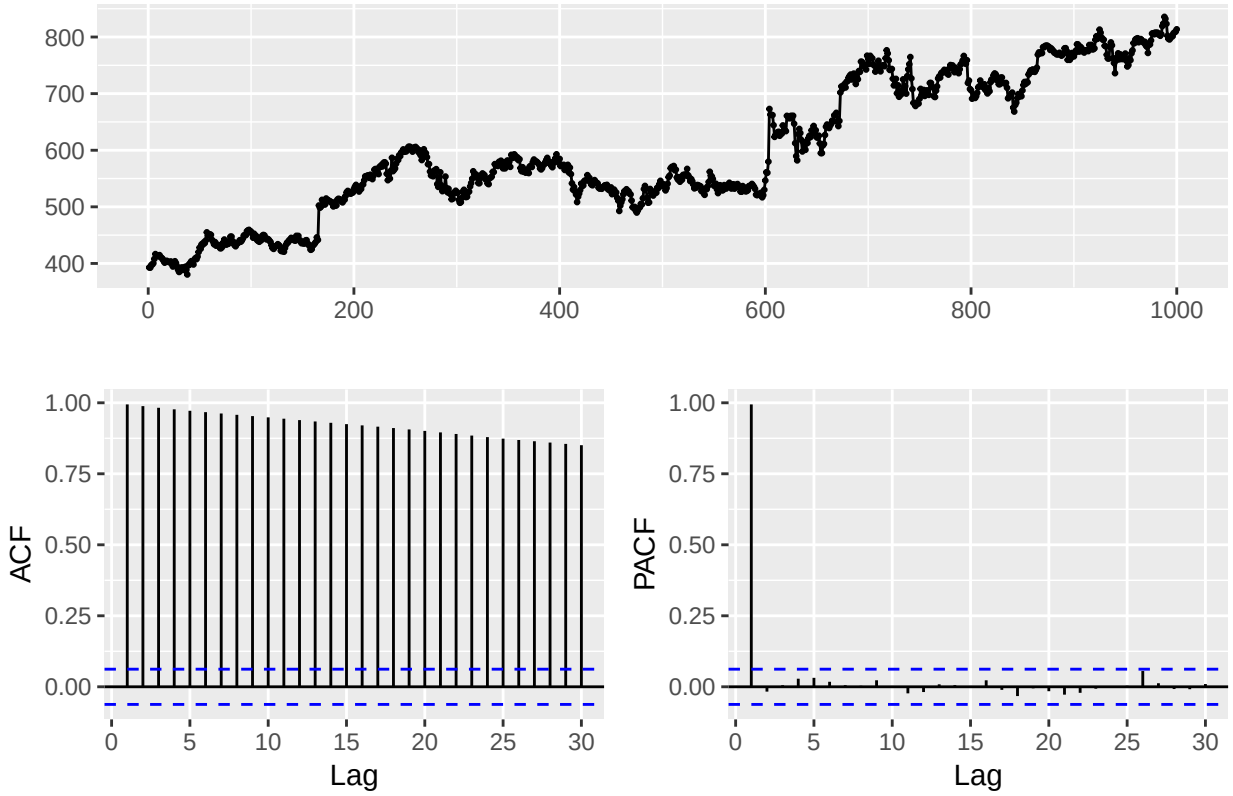
Google

Julien JACQUES

2/27/2020

Let start by looking at the series

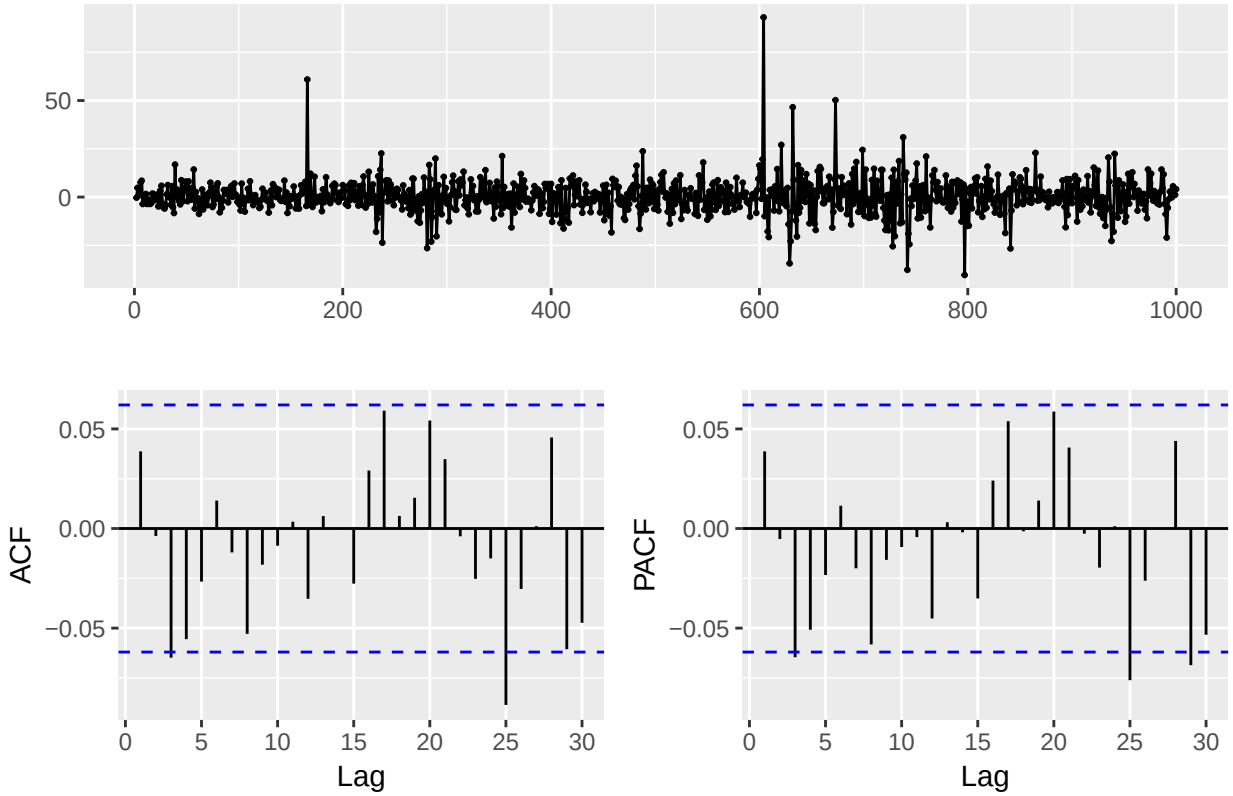
```
library(fpp2)
ggtsdisplay(goog)
```



There is a long-term (linear) trend. There is no reason to have seasonal pattern, and we do not see it in the data.

We can try to remove the linear trend

```
ggtsdisplay(diff(goog))
```



The remaining look like a white noise: that means that we will have nothing to modelize apart the trend.

```
Box.test(diff(goog),lag = 25)
```

```
##
##  Box-Pierce test
##
## data:  diff(goog)
## X-squared = 32.569, df = 25, p-value = 0.1422
```

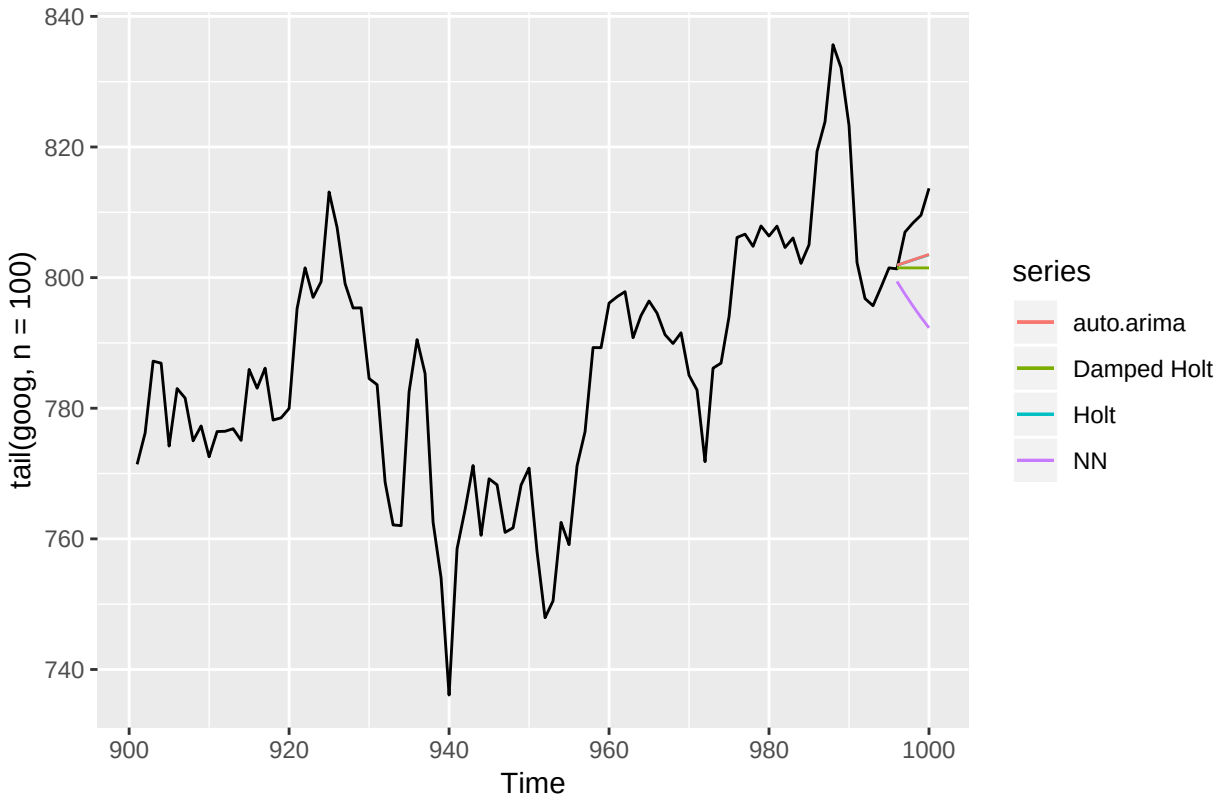
We choose lag=25 since we saw a slightlty significatn ACF at lag 25. In general we observe only the first 10 ACF

We extract training and test set

```
goog_train=head(goog,n=995)
goog_test=tail(goog,n = 5)
```

We test different models

```
fit1=holt(goog_train,h=5, damped=FALSE)
fit2=holt(goog_train,h=5, damped=TRUE)
fit3=auto.arima(goog_train)
prev3=forecast(fit3,h=5)
fit4=nnetar(goog_train)
prev4=forecast(fit4,h=5)
autoplot(tail(goog,n=100))+
  autolayer(fit1$mean,series="Holt")+
  autolayer(fit2$mean,series="Damped Holt")+
  autolayer(prev3$mean,series="auto.arima")+
  autolayer(prev4$mean,series="NN")
```



Forecasting seems to be not very efficient, what is not surprising since the only pattern we observe is a trend which is essentially on a long term.

We can nevertheless comput the RMSE:

```
cat('Holt: ',sqrt(mean((fit1$mean-goog_test)^2)),'\n')
```

```
## Holt:  6.322007
```

```
cat('Damped Holt: ',sqrt(mean((fit2$mean-goog_test)^2)),'\n')
```

```
## Damped Holt:  7.630459
```

```
cat('auto.arima: ',sqrt(mean((prev3$mean-goog_test)^2)),'\n')
```

```
## auto.arima:  6.289618
```

```
cat('NN: ',sqrt(mean((prev4$mean-goog_test)^2)),'\n')
```

```
## NN:  13.83536
```