correlation plot

Stephanie Yang jy2777

12/6/2017

```
library("forecast")
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone/tz/2017c.
## 1.0/zoneinfo/America/New_York'
```

```
setwd("~/Desktop/Homework/Statistical Methods/Project/datasets")
#transpose t()
starbucks.metrics <- t(read.csv("starbucks_metrics2.csv", header=FALSE))

#labels columns
colnames(starbucks.metrics)<- c("Date","Likes (Total) FB","Comments (Total) FB","Shares (Total) FB","Rea

#removes duplicate row
starbucks.metrics1 <- starbucks.metrics[-1,]

##Cleaning the metrics sheet
#1) removes space in column titles
colnames(starbucks.metrics1) <- gsub(" ","",colnames(starbucks.metrics1))

#2) removes % symbol of column 10
starbucks.metrics1[,c(11,15,19,20,21,26,27,28,31,33,38,39,43,44,52,59,60,61,69,71,72,73)] <- as.numeric
starbucks.metrics1 <- as.data.frame(starbucks.metrics1)
class(starbucks.metrics1)
```

```
## [1] "data.frame"
```

```
#3) removes comma separator for thousands, except for date column which is type character not numeric
#gsub to replace "," with "", and then convert the string to numeric using as.numeric
starbucks.metrics1[,2:73] <- lapply(starbucks.metrics1[,2:73], function(x) as.numeric(gsub(",","", as.c
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
##Transforms Monthly to Quarterly Data:
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

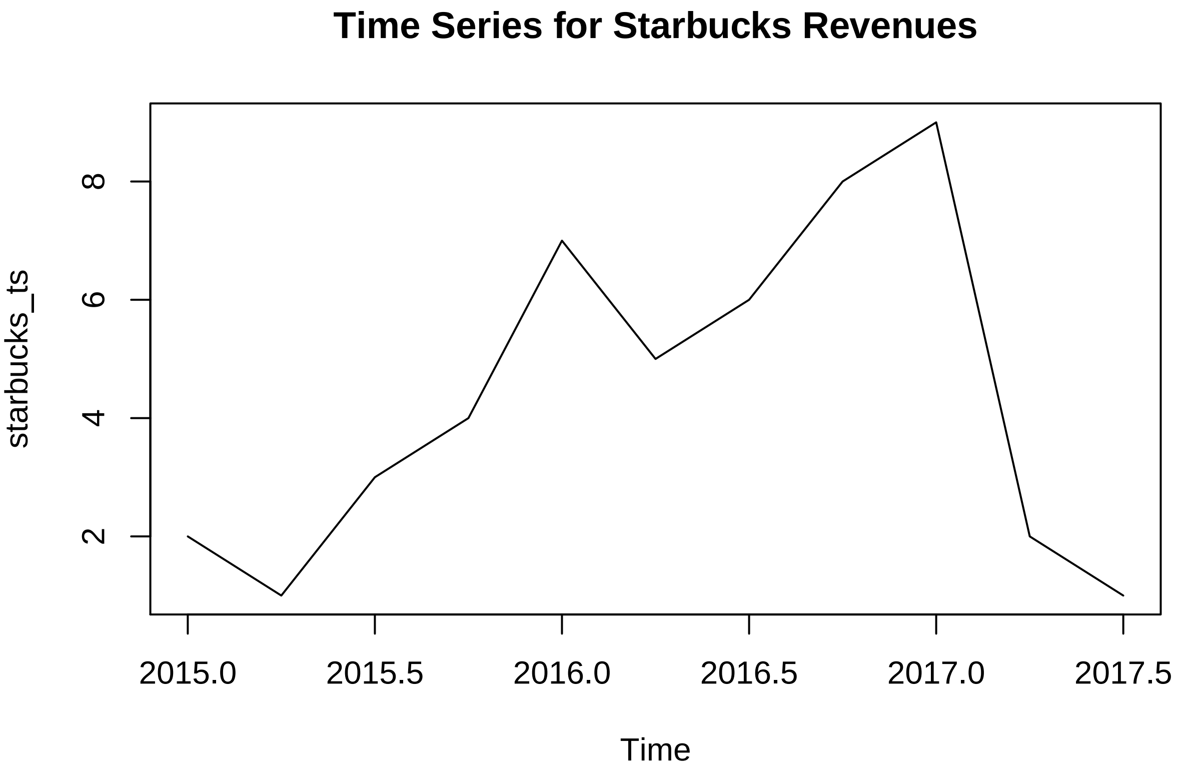
```
## The following object is masked from 'package:base':
##
##     date
```

```
#creates a year and quarter column per row
starbucks.metrics1$Date <- ymd(starbucks.metrics1$Date)
starbucks.metrics1$year = year(starbucks.metrics1$Date)
starbucks.metrics1$quarter = quarter(starbucks.metrics1$Date)

#aggregates quarters of same year and takes their sum (sales are aldo sums) : HOW TO DO IT WITH AGGREGA
library("reshape2")
starbucks.metrics2 <- melt(starbucks.metrics1[,2:75], id=c("quarter", "year"))
starbucks.metrics2 <- dcast(starbucks.metrics2, year + quarter ~ variable, fun.aggregate = sum)
```

```
write.csv(starbucks.metrics2,file="colgate_vizmetrics.csv")
starbucks.metrics2 <- starbucks.metrics2[1:9,]
#quarterly sales data, data points from CapitalIQ over 2 years
chipotle.sales <- read.csv("starbucks_sales.csv")

#===============================================================
#quarterly sales data, data points from CapitalIQ over 2 years
chipotle.sales <- read.csv("starbucks_sales.csv")

#cleaning sales sheet: subsets,transposes and reformats data
chipotle.sales2 <- chipotle.sales[c(10,13),46:54]
chipotle.sales2 <- t(chipotle.sales2)
colnames(chipotle.sales2)<- c("Date","Sales")
typeof(chipotle.sales2[,2])
```

```
## [1] "character"
```

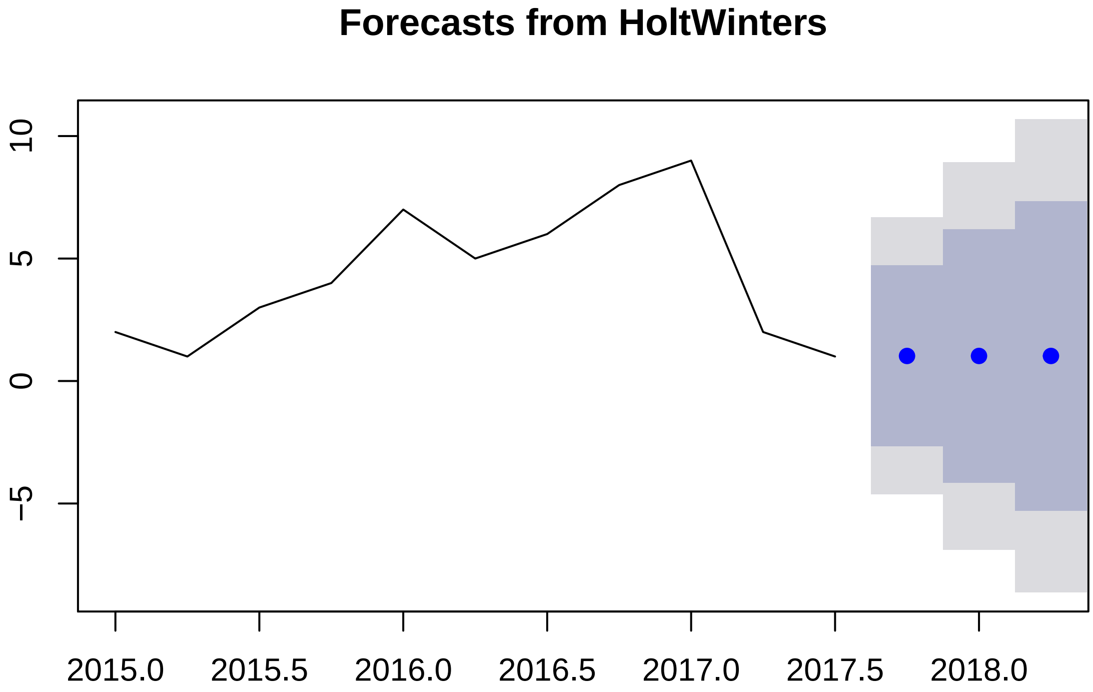
```
chipotle.sales2[,2] <- as.numeric(gsub(",","",chipotle.sales2[,2]))
chipotle.sales2 <- as.data.frame(chipotle.sales2)

starbucks_sale <- chipotle.sales2$Sales


# Creating time series
starbucks_ts <- ts(starbucks_sale, start = c(2015, 1), end=c(2017,3), frequency=4)
plot(starbucks_ts, main="Time Series for Starbucks Revenues")
```



```
# Exponential Models
fit <- HoltWinters(starbucks_ts, beta=FALSE, gamma=FALSE)
forecast(fit, 3)
```

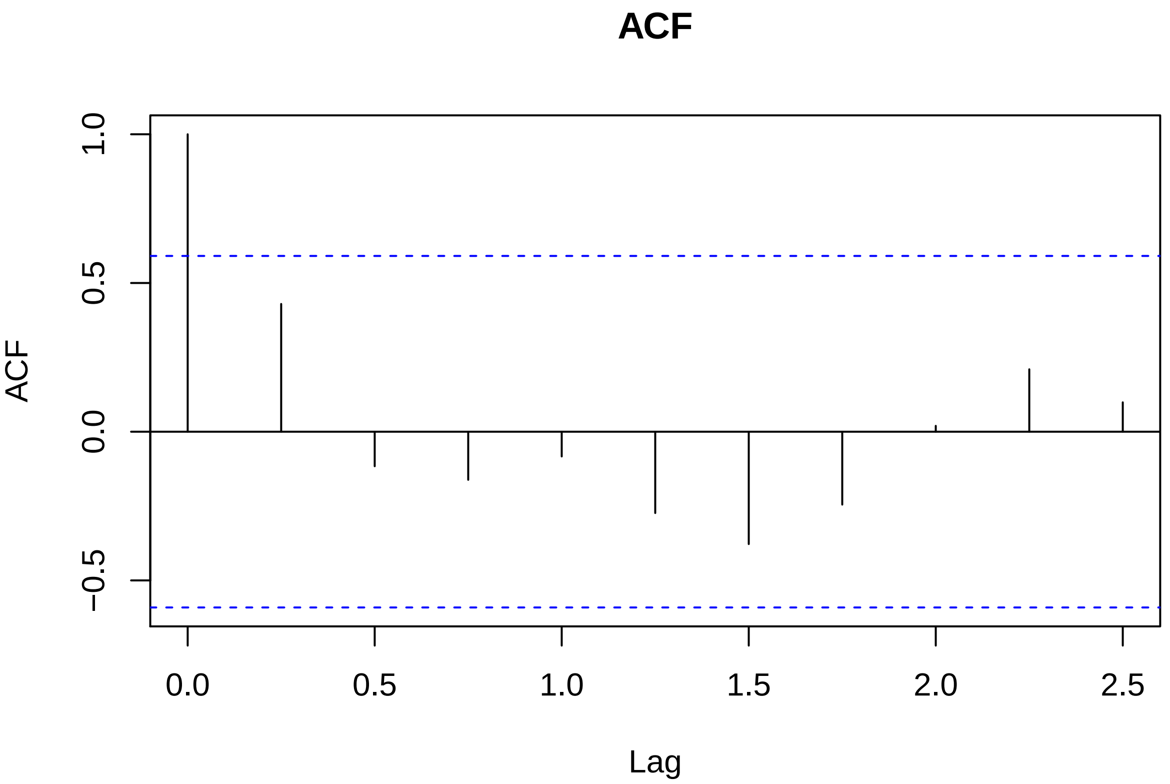
```
##         Point Forecast      Lo 80    Hi 80     Lo 95     Hi 95
## 2017 Q4       1.023382 -2.673185 4.719949 -4.630029  6.676794
## 2018 Q1       1.023382 -4.151148 6.197912 -6.890379  8.937143
## 2018 Q2       1.023382 -5.292227 7.338991 -8.635509 10.682273
```

```
plot(forecast(fit,3))
```

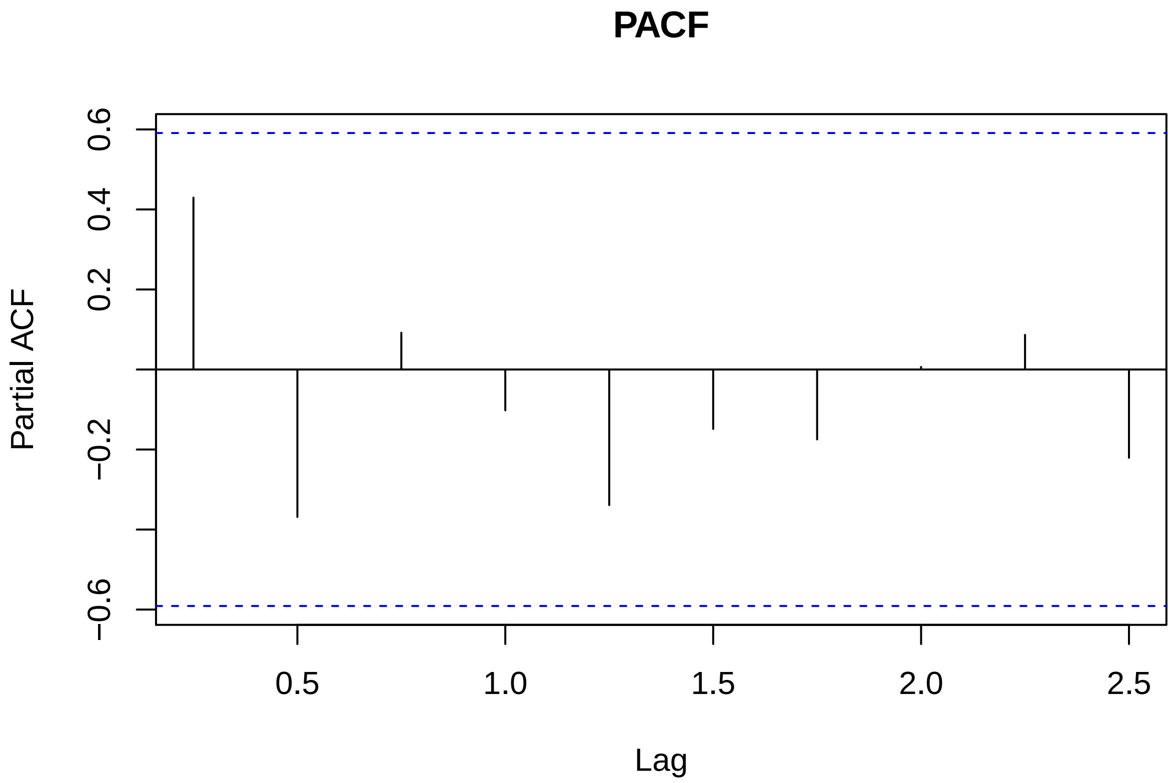


Forecasts from HoltWinters

```
## Automated Forecasting
fit.1 <- ets(starbucks_ts)
fit.2 <- auto.arima(starbucks_ts)


# ARIMA forecasting
acf(starbucks_ts, main="ACF")
```



```
pacf(starbucks_ts, main="PACF")
```

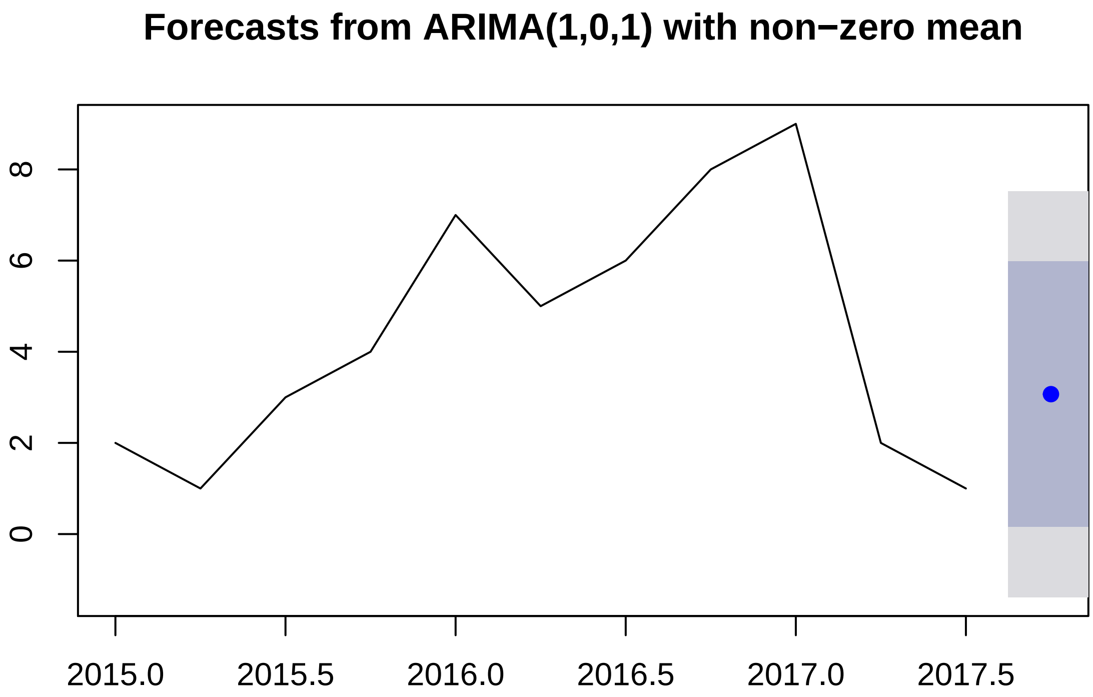


```
fit.3 <- arima(starbucks_ts, order = c(1,0,1))
```

```
accuracy(fit.3)
```

```
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.05819247 2.271644 1.908321 -43.97826 71.41148 0.9087244
##                    ACF1
## Training set 0.05730439
```

```
plot(forecast(fit.3, 1))
```


Forecasts from ARIMA(1,0,1) with non−zero mean