Project 1 Group 5

Aaron Zalki, Mia Siracusa, Lidiia Tronina, John Suh, Henry Vasquez

6/21/2020

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data <- readxl::read\_excel('./Data Set for Class.xls')%>%  
 mutate(SeriesInd = as.Date(SeriesInd, "1899-12-30"))  
  
data%>%  
 summary()

## SeriesInd group Var01   
## Min. :2011-05-06 Length:10572 Min. : 9.03   
## 1st Qu.:2013-01-29 Class :character 1st Qu.: 23.10   
## Median :2014-11-03 Mode :character Median : 38.44   
## Mean :2014-11-01 Mean : 46.98   
## 3rd Qu.:2016-08-05 3rd Qu.: 66.78   
## Max. :2018-05-01 Max. :195.18   
## NA's :854   
## Var02 Var03 Var05 Var07   
## Min. : 1339900 Min. : 8.82 Min. : 8.99 Min. : 8.92   
## 1st Qu.: 12520675 1st Qu.: 22.59 1st Qu.: 22.91 1st Qu.: 22.88   
## Median : 21086550 Median : 37.66 Median : 38.05 Median : 38.05   
## Mean : 37035741 Mean : 46.12 Mean : 46.55 Mean : 46.56   
## 3rd Qu.: 42486700 3rd Qu.: 65.88 3rd Qu.: 66.38 3rd Qu.: 66.31   
## Max. :480879500 Max. :189.36 Max. :195.00 Max. :189.72   
## NA's :842 NA's :866 NA's :866 NA's :866

data1 <- data %>%  
 filter(between(SeriesInd, as.Date("2011-05-06"),as.Date("2017-10-13")))  
data1%>%  
 summary()

## SeriesInd group Var01   
## Min. :2011-05-06 Length:9732 Min. : 9.03   
## 1st Qu.:2012-12-10 Class :character 1st Qu.: 23.10   
## Median :2014-07-25 Mode :character Median : 38.44   
## Mean :2014-07-23 Mean : 46.98   
## 3rd Qu.:2016-03-01 3rd Qu.: 66.78   
## Max. :2017-10-13 Max. :195.18   
## NA's :14   
## Var02 Var03 Var05 Var07   
## Min. : 1339900 Min. : 8.82 Min. : 8.99 Min. : 8.92   
## 1st Qu.: 12520675 1st Qu.: 22.59 1st Qu.: 22.91 1st Qu.: 22.88   
## Median : 21086550 Median : 37.66 Median : 38.05 Median : 38.05   
## Mean : 37035741 Mean : 46.12 Mean : 46.55 Mean : 46.56   
## 3rd Qu.: 42486700 3rd Qu.: 65.88 3rd Qu.: 66.38 3rd Qu.: 66.31   
## Max. :480879500 Max. :189.36 Max. :195.00 Max. :189.72   
## NA's :2 NA's :26 NA's :26 NA's :26

data1%>%  
 filter(!complete.cases(.))

## # A tibble: 26 x 7  
## SeriesInd group Var01 Var02 Var03 Var05 Var07  
## <date> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2011-06-03 S06 NA NA NA NA NA  
## 2 2014-07-01 S05 NA NA NA NA NA  
## 3 2017-06-11 S03 NA 42343600 NA NA NA  
## 4 2017-06-11 S02 NA 38160300 NA NA NA  
## 5 2017-06-11 S01 NA 7329600 NA NA NA  
## 6 2017-06-11 S06 NA 19885500 NA NA NA  
## 7 2017-06-11 S05 NA 16610900 NA NA NA  
## 8 2017-06-11 S04 NA 9098800 NA NA NA  
## 9 2017-06-12 S03 NA 50074700 NA NA NA  
## 10 2017-06-12 S02 NA 45801300 NA NA NA  
## # … with 16 more rows

data\_cc <- data1%>%  
 filter(complete.cases(.))   
medians <- data\_cc %>%  
 group\_by(group) %>%  
 summarise(med = median(Var01))  
data1[is.na(data1$Var01) & data1$group == 'S01', ]$Var01 <- medians$med[medians$group == 'S01'][1]   
data1[is.na(data1$Var01) & data1$group == 'S02', ]$Var01 <- medians$med[medians$group == 'S02'][1]   
data1[is.na(data1$Var01) & data1$group == 'S03', ]$Var01 <- medians$med[medians$group == 'S03'][1]   
data1[is.na(data1$Var01) & data1$group == 'S04', ]$Var01 <- medians$med[medians$group == 'S04'][1]   
data1[is.na(data1$Var01) & data1$group == 'S05', ]$Var01 <- medians$med[medians$group == 'S05'][1]   
data1[is.na(data1$Var01) & data1$group == 'S06', ]$Var01 <- medians$med[medians$group == 'S06'][1]   
  
medians <- data\_cc %>%  
 group\_by(group) %>%  
 summarise(med = median(Var02))  
data1[is.na(data1$Var02) & data1$group == 'S05', ]$Var02 <- medians$med[medians$group == 'S05'][1]   
data1[is.na(data1$Var02) & data1$group == 'S06', ]$Var02 <- medians$med[medians$group == 'S06'][1]   
  
medians <- data\_cc %>%  
 group\_by(group) %>%  
 summarise(med = median(Var03))  
data1[is.na(data1$Var03) & data1$group == 'S01', ]$Var03 <- medians$med[medians$group == 'S01'][1]   
data1[is.na(data1$Var03) & data1$group == 'S02', ]$Var03 <- medians$med[medians$group == 'S02'][1]   
data1[is.na(data1$Var03) & data1$group == 'S03', ]$Var03 <- medians$med[medians$group == 'S03'][1]   
data1[is.na(data1$Var03) & data1$group == 'S04', ]$Var03 <- medians$med[medians$group == 'S04'][1]   
data1[is.na(data1$Var03) & data1$group == 'S05', ]$Var03 <- medians$med[medians$group == 'S05'][1]   
data1[is.na(data1$Var03) & data1$group == 'S06', ]$Var03 <- medians$med[medians$group == 'S06'][1]   
  
  
medians <- data\_cc %>%  
 group\_by(group) %>%  
 summarise(med = median(Var05))  
data1[is.na(data1$Var05) & data1$group == 'S01', ]$Var05 <- medians$med[medians$group == 'S01'][1]   
data1[is.na(data1$Var05) & data1$group == 'S02', ]$Var05 <- medians$med[medians$group == 'S02'][1]   
data1[is.na(data1$Var05) & data1$group == 'S03', ]$Var05 <- medians$med[medians$group == 'S03'][1]   
data1[is.na(data1$Var05) & data1$group == 'S04', ]$Var05 <- medians$med[medians$group == 'S04'][1]   
data1[is.na(data1$Var05) & data1$group == 'S05', ]$Var05 <- medians$med[medians$group == 'S05'][1]   
data1[is.na(data1$Var05) & data1$group == 'S06', ]$Var05 <- medians$med[medians$group == 'S06'][1]   
  
medians <- data\_cc %>%  
 group\_by(group) %>%  
 summarise(med = median(Var07))  
data1[is.na(data1$Var07) & data1$group == 'S01', ]$Var07 <- medians$med[medians$group == 'S01'][1]   
data1[is.na(data1$Var07) & data1$group == 'S02', ]$Var07 <- medians$med[medians$group == 'S02'][1]   
data1[is.na(data1$Var07) & data1$group == 'S03', ]$Var07 <- medians$med[medians$group == 'S03'][1]   
data1[is.na(data1$Var07) & data1$group == 'S04', ]$Var07 <- medians$med[medians$group == 'S04'][1]   
data1[is.na(data1$Var07) & data1$group == 'S05', ]$Var07 <- medians$med[medians$group == 'S05'][1]   
data1[is.na(data1$Var07) & data1$group == 'S06', ]$Var07 <- medians$med[medians$group == 'S06'][1]   
  
data1%>%  
 summary()

## SeriesInd group Var01   
## Min. :2011-05-06 Length:9732 Min. : 9.03   
## 1st Qu.:2012-12-10 Class :character 1st Qu.: 23.16   
## Median :2014-07-25 Mode :character Median : 38.40   
## Mean :2014-07-23 Mean : 46.98   
## 3rd Qu.:2016-03-01 3rd Qu.: 66.80   
## Max. :2017-10-13 Max. :195.18   
## Var02 Var03 Var05 Var07   
## Min. : 1339900 Min. : 8.82 Min. : 8.99 Min. : 8.92   
## 1st Qu.: 12521025 1st Qu.: 22.63 1st Qu.: 22.93 1st Qu.: 22.92   
## Median : 21086550 Median : 37.62 Median : 38.01 Median : 37.98   
## Mean : 37031871 Mean : 46.12 Mean : 46.55 Mean : 46.56   
## 3rd Qu.: 42464900 3rd Qu.: 65.97 3rd Qu.: 66.43 3rd Qu.: 66.39   
## Max. :480879500 Max. :189.36 Max. :195.00 Max. :189.72

S01 – Forecast Var01, Var02 S02 – Forecast Var02, Var03 S03 – Forecast Var05, Var07 S04 – Forecast Var01, Var02 S05 – Forecast Var02, Var03 S06 – Forecast Var05, Var07

## S01 – Forecast Var01

s01\_var01 <- data1%>%  
 filter(group =='S01' ) %>%  
 select(SeriesInd, Var01)

library(fpp2)

## Loading required package: ggplot2

## Registered S3 methods overwritten by 'ggplot2':  
## method from   
## [.quosures rlang  
## c.quosures rlang  
## print.quosures rlang

## Loading required package: forecast

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

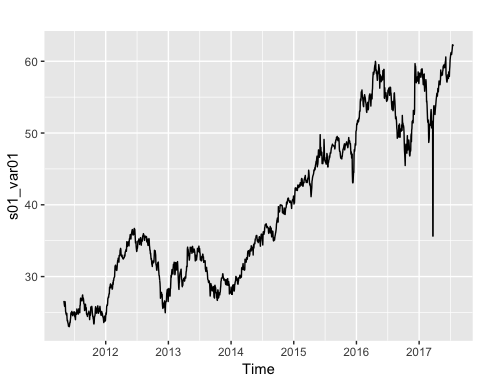
## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

## Registered S3 methods overwritten by 'forecast':  
## method from   
## fitted.fracdiff fracdiff  
## residuals.fracdiff fracdiff

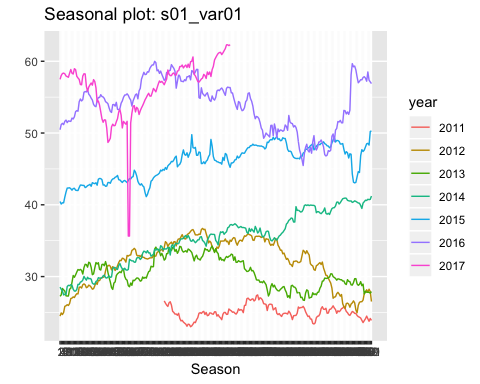
## Loading required package: fma

## Loading required package: expsmooth

s01\_var01<- ts(s01\_var01[,2], start = c(2011,88), frequency = 261)  
autoplot(s01\_var01)



ggseasonplot(s01\_var01)



s01\_var01 %>% decompose(type = "multiplicative") %>%  
 autoplot() + xlab("Year")

