Anormalies

Emmanuel Titi

2022-06-14

```
#Loading the libraries
library(Rcpp)
library(AnomalyDetection)
```

```
df=read.csv("http://bit.ly/CarreFourSalesDataset")
head(df)
```

```
##        Date    Sales
## 1  1/5/2019 548.9715
## 2  3/8/2019  80.2200
## 3  3/3/2019 340.5255
## 4 1/27/2019 489.0480
## 5  2/8/2019 634.3785
## 6 3/25/2019 627.6165
```

```
str(df)
```

```
## 'data.frame':    1000 obs. of  2 variables:
##  $ Date : chr  "1/5/2019" "3/8/2019" "3/3/2019" "1/27/2019" ...
##  $ Sales: num  549 80.2 340.5 489 634.4 ...
```

lets convert the date to a datetime data type

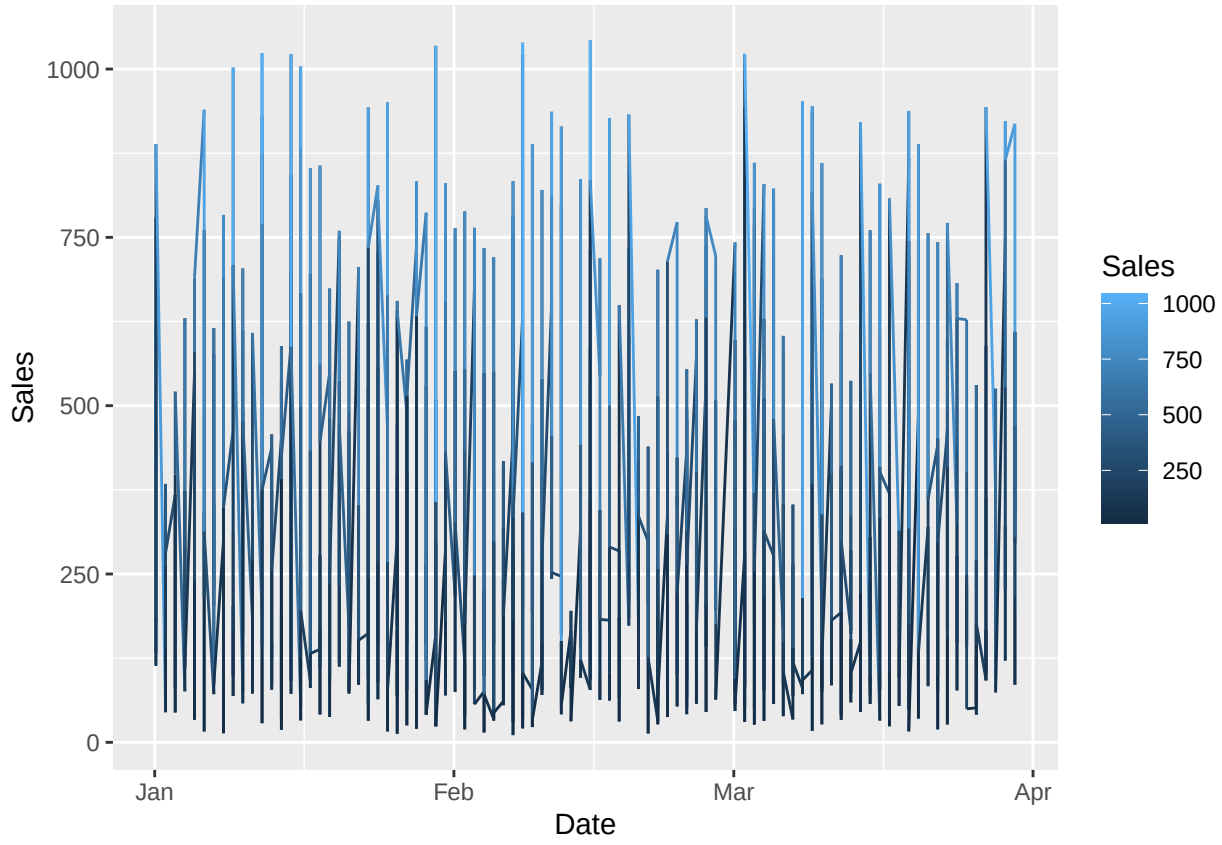
```
df$Date<-as.Date(df$Date,format="%m/%d/%y")
df$Date<-as.POSIXct(df$Date)
head(df)
```

```
##                  Date    Sales
## 1 2020-01-05 03:00:00 548.9715
## 2 2020-03-08 03:00:00  80.2200
## 3 2020-03-03 03:00:00 340.5255
## 4 2020-01-27 03:00:00 489.0480
## 5 2020-02-08 03:00:00 634.3785
## 6 2020-03-25 03:00:00 627.6165
```

```
df=df[order(as.Date(df$Date,format="%m/%d/%y")),]
str(df)
```

```
## 'data.frame':    1000 obs. of  2 variables:
##  $ Date : POSIXct, format: "2020-01-01 03:00:00" "2020-01-01 03:00:00" ...
##  $ Sales: num  457 400 471 388 133 ...
```

```
#Plotting data
library(ggplot2)
ggplot(df, aes(x=Date, y=Sales, color=Sales)) + geom_line()
```



We see some huge spikes at different intervals. There are a lot of anomalies in this data. *Decomposing*

```
library(tibble)
df1<-as_tibble(df)
class(df1)
```

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

```
library(anomalize)
```

```
## == Use anomalize to improve your Forecasts by 50%! ==============================
## Business Science offers a 1-hour course - Lab #18: Time Series Anomaly Detection!
## </> Learn more at: https://university.business-science.io/p/learning-labs-pro </>
```

```
dfz<-df1%>%
  time_decompose(Sales,method = "stl",frequency = "24 hours",trend="1 day")%>%
  anomalize(remainder,method="gesd",alpha=0.05,max_anoms = 0.1)%>%
  plot_anomaly_decomposition()
```

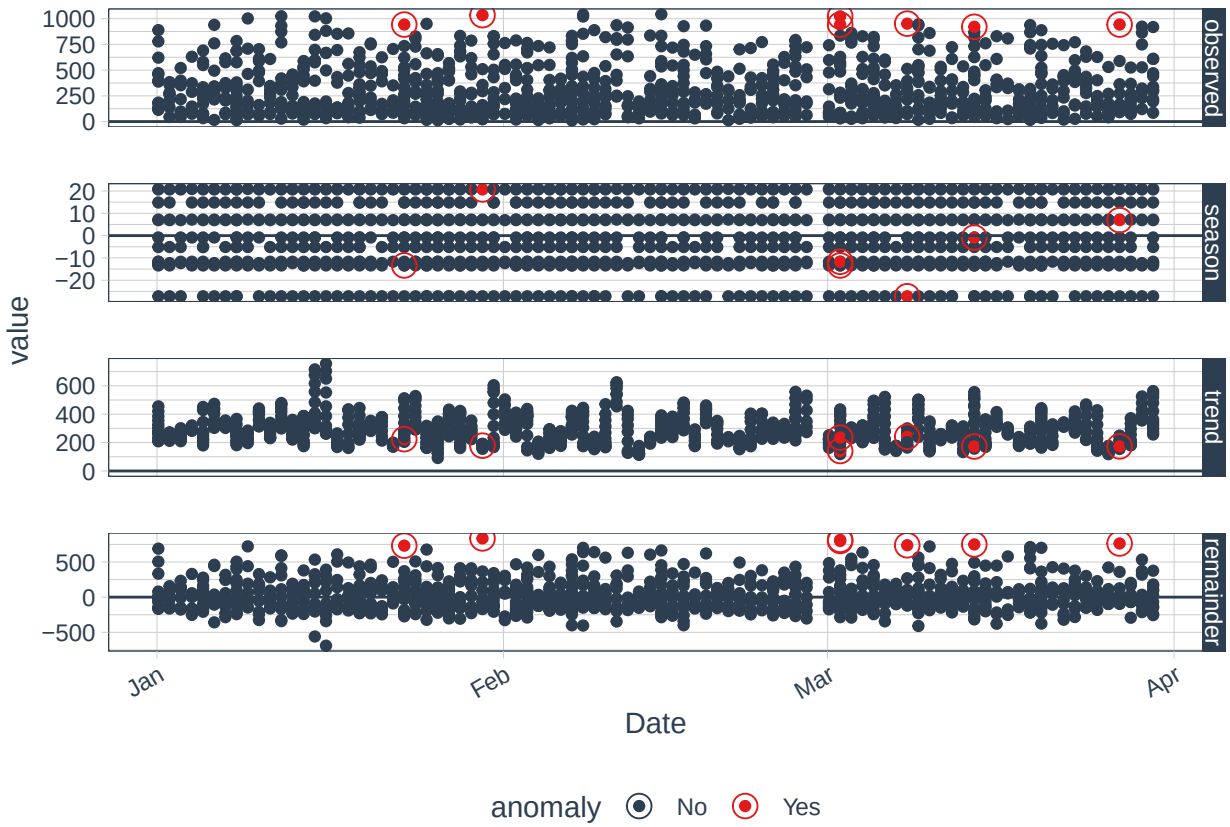
```
## Converting from tbl_df to tbl_time.
## Auto-index message: index = Date
```

```
## frequency = 11 seconds

## trend = 11 seconds

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

```
dfz
```



Our data has more than 1 record per day ,we hence set our frequency of 24hrs and 1 day.

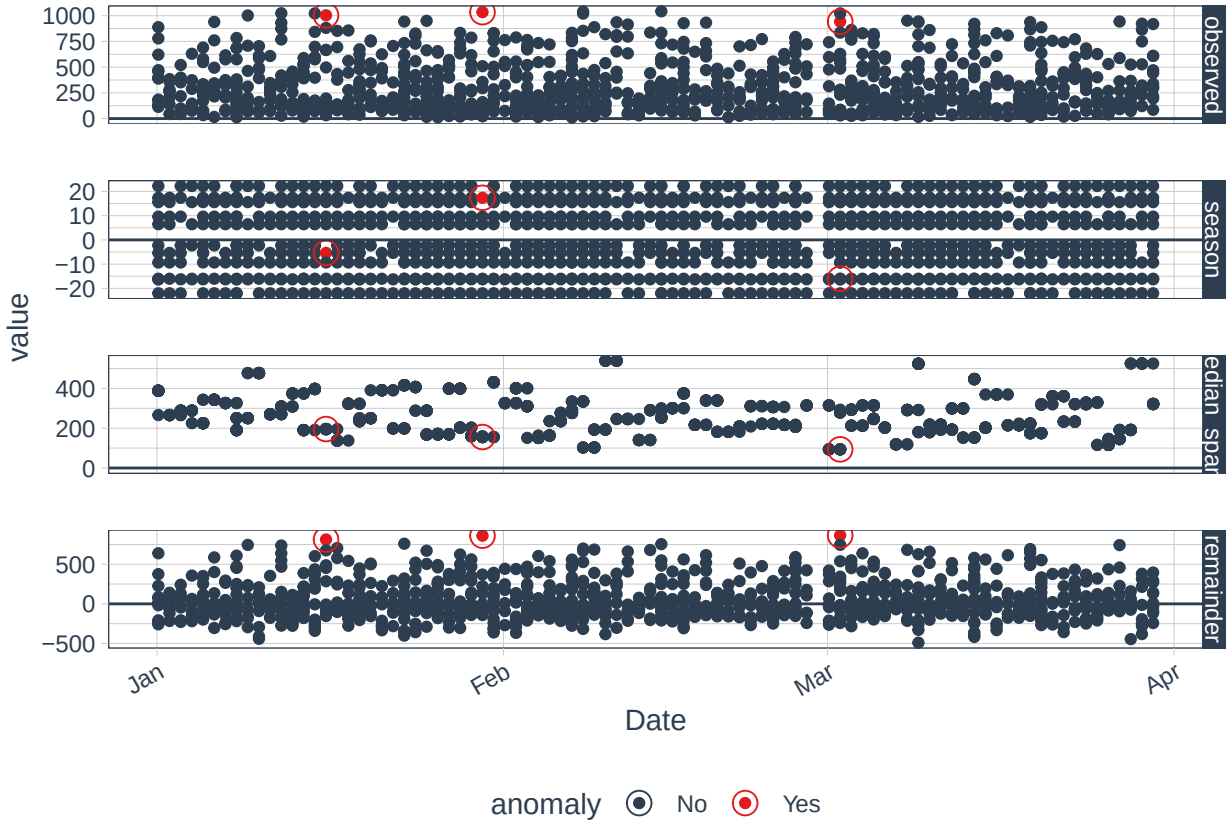
```
dft<-df1%>%
  time_decompose(Sales,method = "twitter",frequency = "24 hours",trend="1 day")%>%
  anomalize(remainder,method="gesd",alpha=0.05,max_anoms = 0.1)%>%
  plot_anomaly_decomposition()
```

```
## Converting from tbl_df to tbl_time.
## Auto-index message: index = Date

## frequency = 11 seconds

## median_span = 11 seconds
```

```
dft
```



Twitter detects less anommalies than stl.

lets recompose the time series

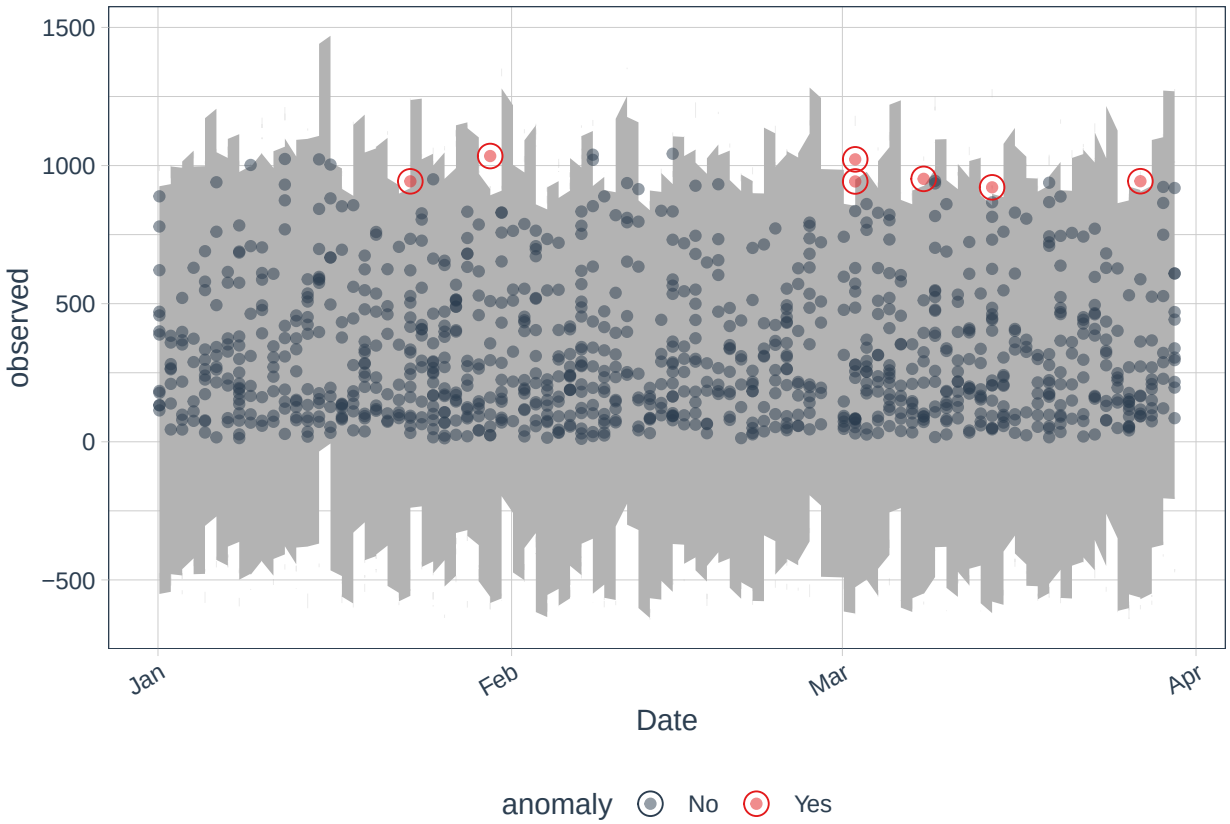
```
dfr<-df1%>%
  time_decompose(Sales,method = "stl",frequency = "24 hours",trend="1 day")%>%
  anomalize(remainder,method="gesd",alpha=0.05,max_anoms = 0.1)%>%
  time_recompose()%>%
  plot_anomalies(time_recompose=TRUE,ncol=3,alpha_dots=0.5)
```

```
## Converting from tbl_df to tbl_time.
## Auto-index message: index = Date

## frequency = 11 seconds

## trend = 11 seconds
```

```
dfr
```


1500
1000
500
0
−500
observed
Jan
Feb
Mar
Apr
Date
anomaly
No
Yes