ANOMALY DETECTION

Anomalies of a time series data can be defined as outliers of the **remainders** once the linear trend and seasonal periodicity are taken into account.

We use library "anomalize" in R to identify anomalies. We call the function `time_decompose()` with the option `method="stl"` (factoring the linear and seasonal components), and the function `anomalize()` with the option `method="iqr"`. By default, this method defines outliers as observations lying below $Q1-3\cdot IQR$ or above $Q3+3\cdot IQR$, where $Q1$ is the first quartile (25th percentile), $Q3$ is the third quartile (75th percentile), and the interquartile range is $IQR = Q3 - Q1$.

The default setting can be changed by specifying a value for alpha other than 0.05 (option `"alpha="`). Outliers are defined as values that lie $0.15/\alpha \cdot IQR$ distance away from the quartiles. The default value of $\alpha = 0.05$, thus resulting in the multiplicative constant of 3. If alpha is increased, more observations become outliers. If alpha is decreased, fewer observations are labeled as outliers.

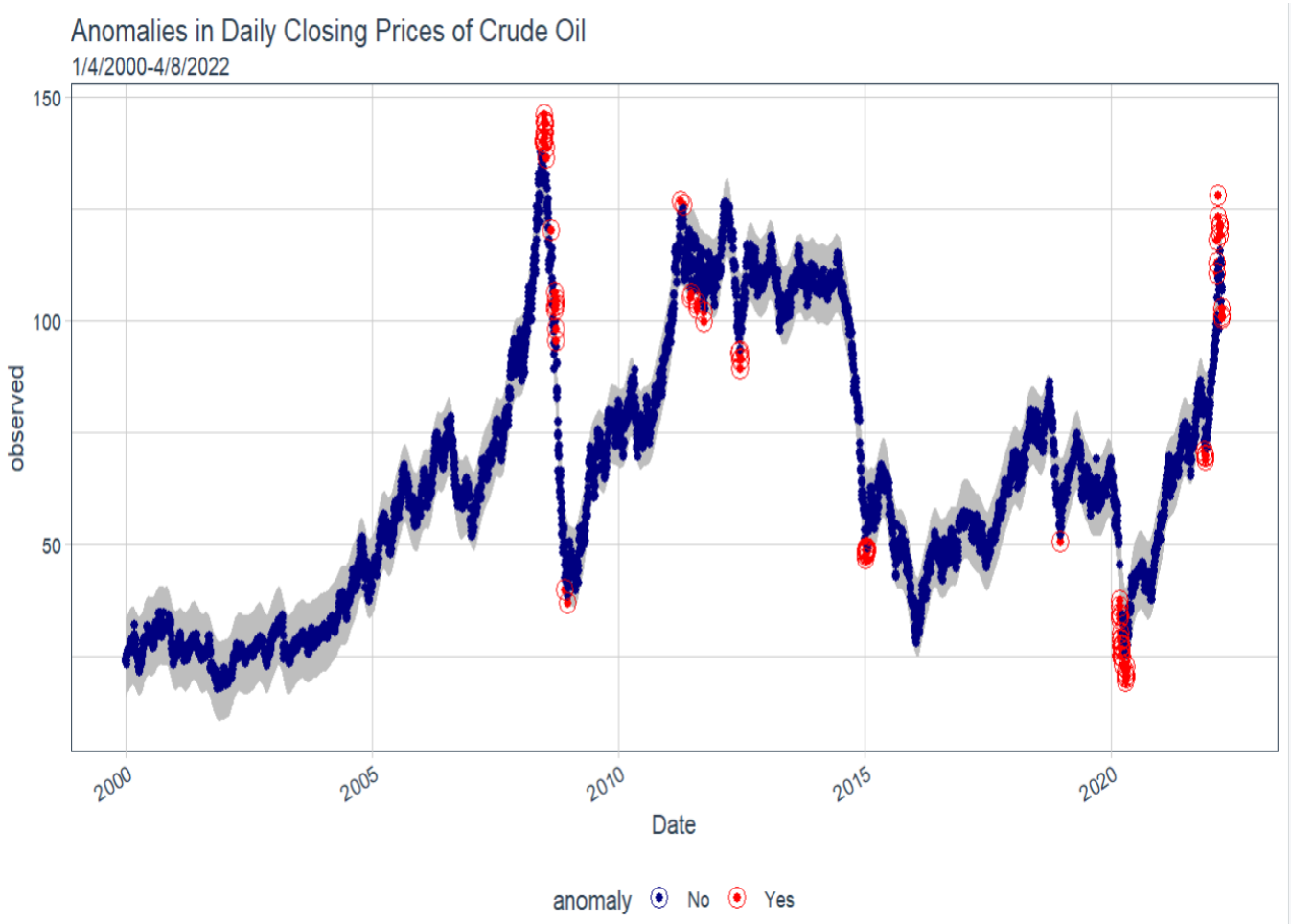
Example. We use the crude oil data from the previous example to detect and plot anomalies in the daily closing prices.

```
crudeoil.data<- read.csv(file="./crudeoil_data.csv", header=TRUE, sep=",")

crudeoil.data$Date<- as.Date(crudeoil.data$Date, format="%Y-%m-%d")

library(tibbletime) #creates indices for date in time series data
crudeoil.data_tbl <- as_tbl_time(crudeoil.data, Date)

library(anomalize)
library(tidyverse)
crudeoil.data_tbl %>% time_decompose(Close, method="stl") %>% anomalize(remainder, method="iqr")
%>% time_recompose() %>% plot_anomalies(time_recomposed=TRUE, color_no='navy', color_yes='red',
fill_ribbon='gray', size_circles=4) + labs(title="Anomalies in Daily Closing Prices of Crude Oil",
subtitle="1/4/2000-4/8/2022")
```


Anomalies in Daily Closing Prices of Crude Oil
1/4/2000-4/8/2022
150
100
50
observed
2000
2005
2010
2015
2020
Date
anomaly
No
Yes