DATA624 - Homework 1

Esteban Aramayo

2022-06-05

# Week 1, Getting Started and Time Series Graphics : May 31 - June 05

## Exercise 2.10.1

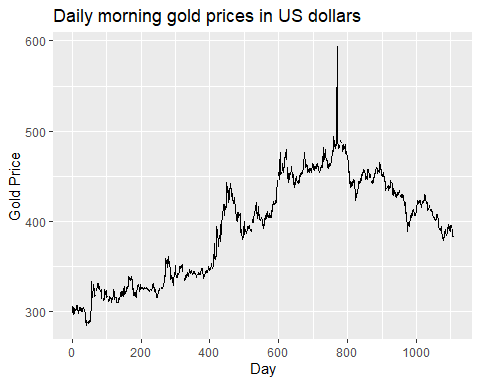
Use the help function to explore what the series gold, woolyrnq and gas represent.

1. Use autoplot() to plot each of these in separate plots.

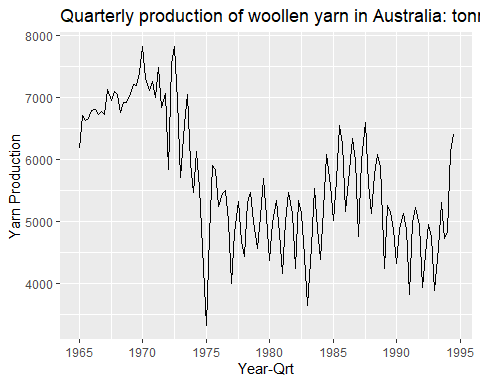
### Solution

library(fpp2)

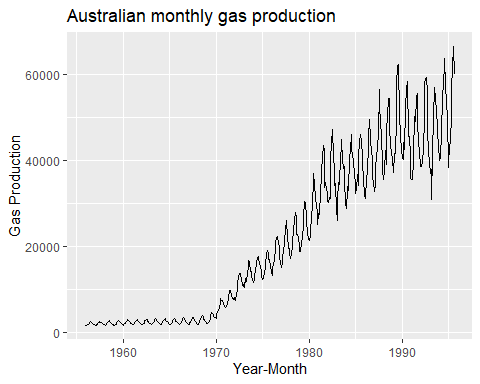
autoplot(gold) +  
 ggtitle("Daily morning gold prices in US dollars") +  
 ylab("Gold Price") +  
 xlab("Day")



autoplot(woolyrnq) +  
 ggtitle("Quarterly production of woollen yarn in Australia: tonnes") +  
 ylab("Yarn Production") +  
 xlab("Year-Qrt")



autoplot(gas) +  
 ggtitle("Australian monthly gas production") +  
 ylab("Gas Production") +  
 xlab("Year-Month")



1. What is the frequency of each series? Hint: apply the frequency() function.

### Solution

The frequency of the series is annual because its frequency value is 1.

frequency(gold)

#> [1] 1

The frequency of the series is quarterly because its frequency value is 4.

frequency(woolyrnq)

#> [1] 4

The frequency of the series is monthly because its frequency value is 12.

frequency(gas)

#> [1] 12

1. Use which.max() to spot the outlier in the gold series. Which observation was it?

### Solution

The outlier observation in the series is 593.7, which is located at the position 770.

gold[which.max(gold)]

#> [1] 593.7

## Exercise 2.10.3

Download some monthly Australian retail data from the book website. These represent retail sales in various categories for different Australian states, and are stored in a MS-Excel file.

1. You can read the data into R with the following script:

retaildata <- readxl::read\_excel(“retail.xlsx”, skip=1)

The second argument (skip=1) is required because the Excel sheet has two header rows.

### Solution

# library(rio)  
# retail\_url <- "https://otexts.com/fpp2/extrafiles/retail.xlsx"  
# retaildata <- rio::import(file = retail\_url, which = 1)  
# head(retaildata)  
  
library(readxl)  
retaildata <- readxl::read\_xlsx("retail.xlsx", skip = 1)  
head(retaildata)

#> # A tibble: 6 x 190  
#> `Series ID` A3349335T A3349627V A3349338X A3349398A A3349468W  
#> <dttm> <dbl> <dbl> <dbl> <dbl> <dbl>  
#> 1 1982-04-01 00:00:00 303. 41.7 63.9 409. 65.8  
#> 2 1982-05-01 00:00:00 298. 43.1 64 405. 65.8  
#> 3 1982-06-01 00:00:00 298 40.3 62.7 401 62.3  
#> 4 1982-07-01 00:00:00 308. 40.9 65.6 414. 68.2  
#> 5 1982-08-01 00:00:00 299. 42.1 62.6 404. 66   
#> 6 1982-09-01 00:00:00 305. 42 64.4 412. 62.3  
#> # ... with 184 more variables: A3349336V <dbl>, A3349337W <dbl>,  
#> # A3349397X <dbl>, A3349399C <dbl>, A3349874C <dbl>, A3349871W <dbl>,  
#> # A3349790V <dbl>, A3349556W <dbl>, A3349791W <dbl>, A3349401C <dbl>,  
#> # A3349873A <dbl>, A3349872X <dbl>, A3349709X <dbl>, A3349792X <dbl>,  
#> # A3349789K <dbl>, A3349555V <dbl>, A3349565X <dbl>, A3349414R <dbl>,  
#> # A3349799R <dbl>, A3349642T <dbl>, A3349413L <dbl>, A3349564W <dbl>,  
#> # A3349416V <dbl>, A3349643V <dbl>, A3349483V <dbl>, A3349722T <dbl>, ...

1. Select one of the time series as follows (but replace the column name with your own chosen column):

myts <- ts(retaildata[,“A3349873A”], frequency=12, start=c(1982,4))

1. Explore your chosen retail time series using the following functions:

autoplot(), ggseasonplot(), ggsubseriesplot(), gglagplot(), ggAcf()

Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

### Solution