DSCI5340_HW1_Group13

Sai Vamshi Palakurthi, Kiran Kumar Gontyala, Pavan Kumar Reddy Naramreddy, Sai Hari Vignesh G

2024-02-11

Let us install the required packages, which are used to get dataset and to plot graph, we need tsibbledata to get aus_production dataset, ggplot2 to plot graphs

```
library(tsibbledata)
library(tsibble)
```

```
##
## Attaching package: 'tsibble'

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

```
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
```

```
library(ggplot2)
library(fpp2)
```

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

## -- Attaching packages --------------------------------------------- fpp2 2.5 --

## v forecast  8.21.1     v expsmooth 2.3
## v fma       2.5

##
```

```
library(forecast)
library(lubridate)
```

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

```
## The following object is masked from 'package:tsibble':
##
##     interval
```

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

#Now we need to use the dataset -> aus_production from tsibbledata

```
data("aus_production")
names(aus_production)
```

```
## [1] "Quarter"     "Beer"        "Tobacco"     "Bricks"      "Cement"
## [6] "Electricity" "Gas"
```

Now let us assign quarter column to date object

```
last_five_years <- aus_production %>% dplyr::select(Quarter,Gas)
```

#Now let us extract last five years information as given in the question

```
last_5_years <- last_five_years %>% filter(year(Quarter) >= 2006)
```

#now we need to get data of 2006, 2007, 2008, 2009,2010, using the below line

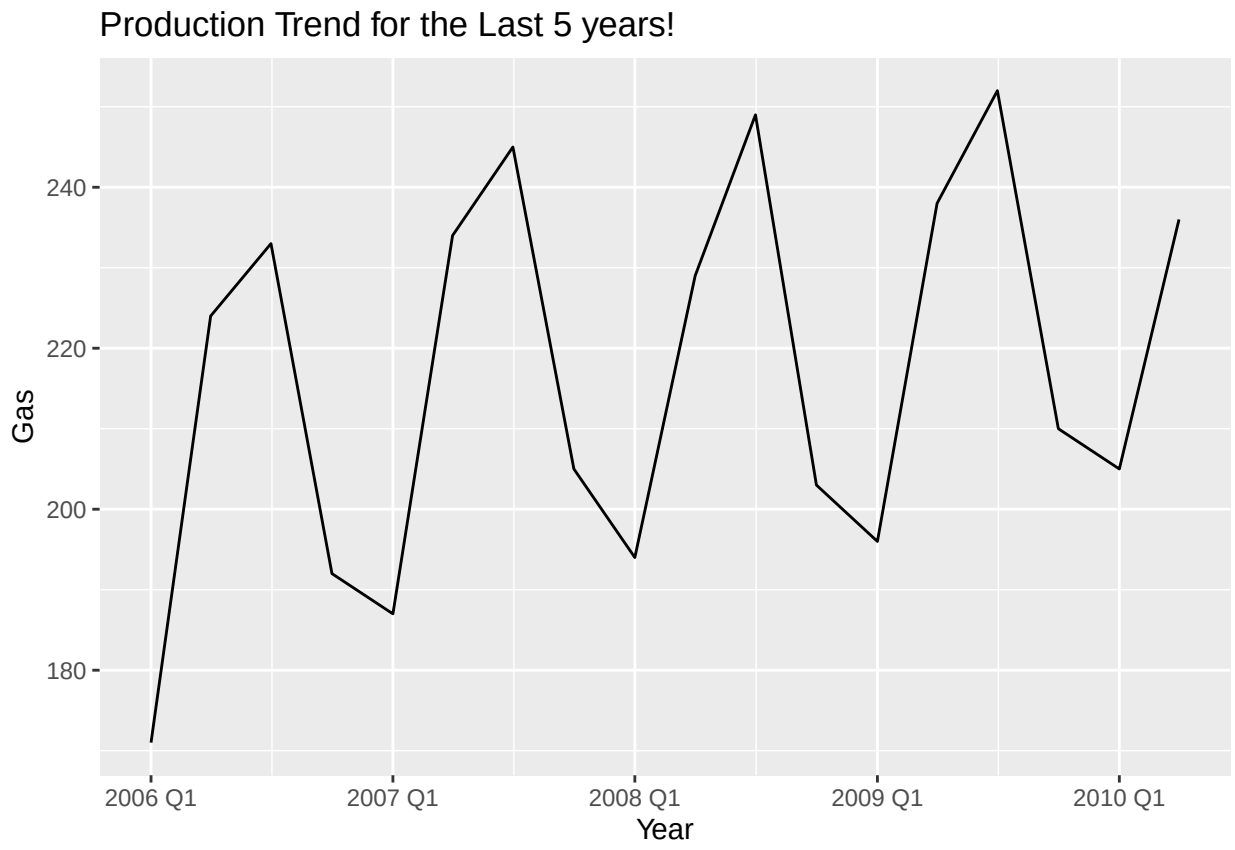
```
last_5_years
```

```
## # A tsibble: 18 x 2 [1Q]
##    Quarter   Gas
##      <qtr> <dbl>
##  1 2006 Q1   171
##  2 2006 Q2   224
##  3 2006 Q3   233
##  4 2006 Q4   192
##  5 2007 Q1   187
##  6 2007 Q2   234
##  7 2007 Q3   245
##  8 2007 Q4   205
##  9 2008 Q1   194
## 10 2008 Q2   229
## 11 2008 Q3   249
## 12 2008 Q4   203
## 13 2009 Q1   196
```

```
## 14 2009 Q2   238
## 15 2009 Q3   252
## 16 2009 Q4   210
## 17 2010 Q1   205
## 18 2010 Q2   236
```

Q1 #Now let us plot the time series plot

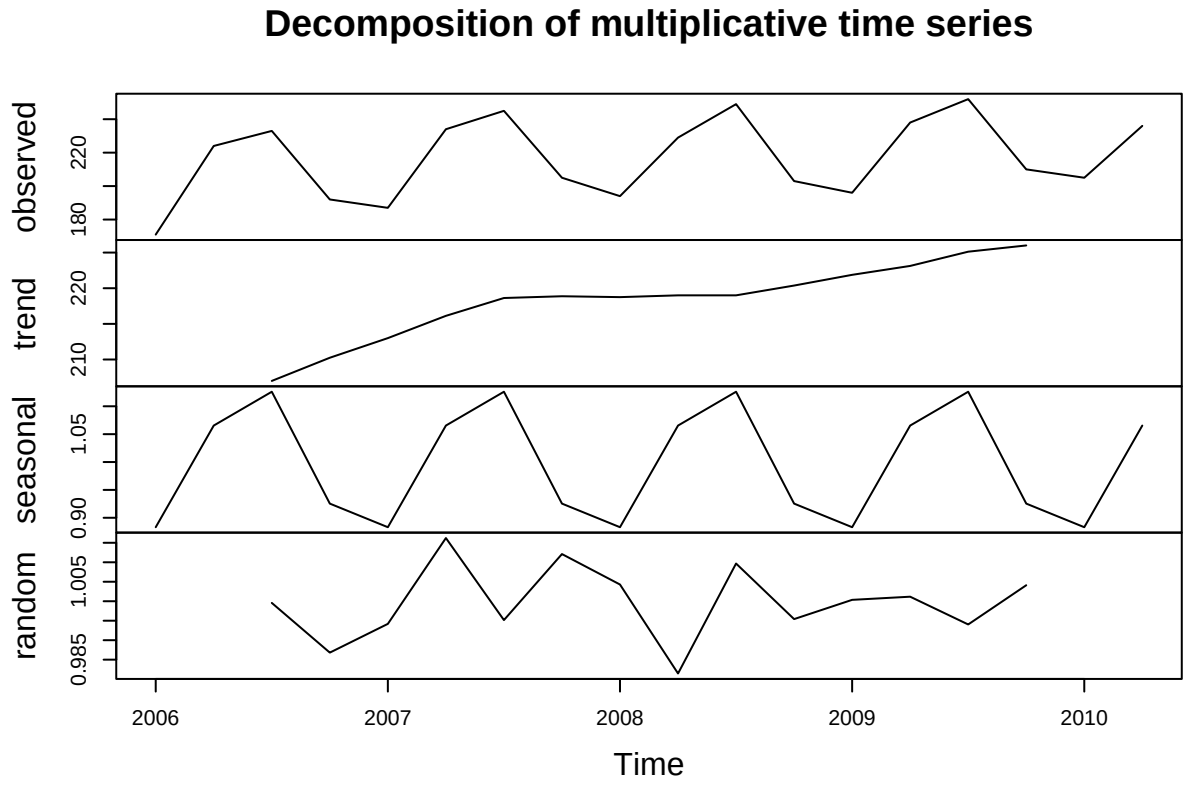
```
ggplot(last_5_years, aes(x = Quarter, y = Gas)) +
  geom_line() +
  labs(title = "Production Trend for the Last 5 years!",
  x = "Year",
  y = "Gas")
```



Use `classical_decomposition` with `type=multiplicative` to calculate the trend-cycle and seasonal indices. Do the results support the graphical interpretation in the question

Q2 #Now let us decompose the time series using type = "multiplicative"

```
last_5_years <- last_5_years$Gas
last_5_years <- ts(last_5_years, start = c(2006,1), end = c(2010, 2), frequency = 4)
last_5_years_decompse <- decompose(last_5_years, type = "multiplicative")
plot(last_5_years_decompse)
```



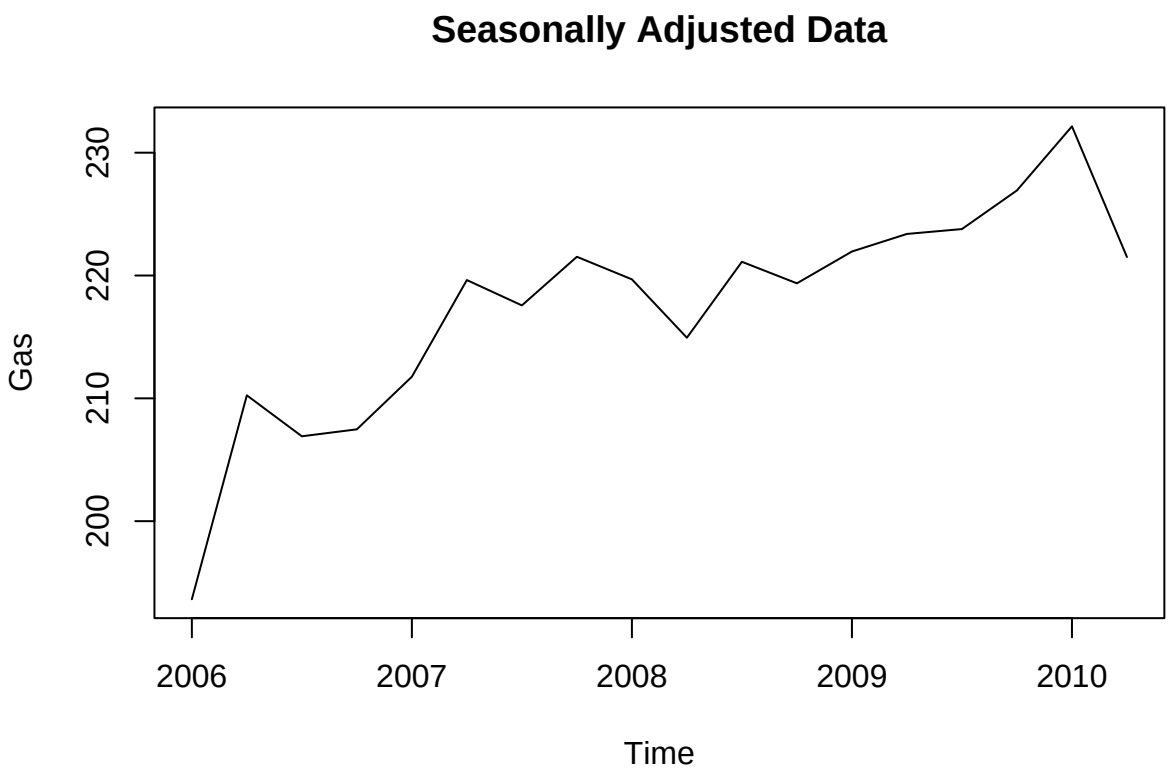
As can be seen from the plot above, there is an increasing trend-cycle with low production at the beginning and end of each year. Production increases from the first quarter until reaching its peak in the second and third quarters, at which point it starts to drop once more.

#Question 3 - Compute and plot the seasonally-adjusted data #Compute and plot the seasonally adjusted data

```
last_5_years_seasonally_adjusted <- last_5_years/last_5_years_decompse$seasonal
```

#now we need to plot the last_5_years_seasonally_adjusted

```
plot(last_5_years_seasonally_adjusted, main = "Seasonally Adjusted Data", ylab = "Gas")
```



#Question 4 - Change one observation to be an outlier (e.g., add 200 to the 2008 Q1 observation), and #recompute the seasonally adjusted data. What is the effect of the outlier? Explain? #> #As there is an change in one observation to be an outliner #> #here we have added 200 to the 2008 Q1 observation

```
last_5_years[9] <- last_5_years[9] + 200
```

#Now we need to recompute the seasonally adjusted data

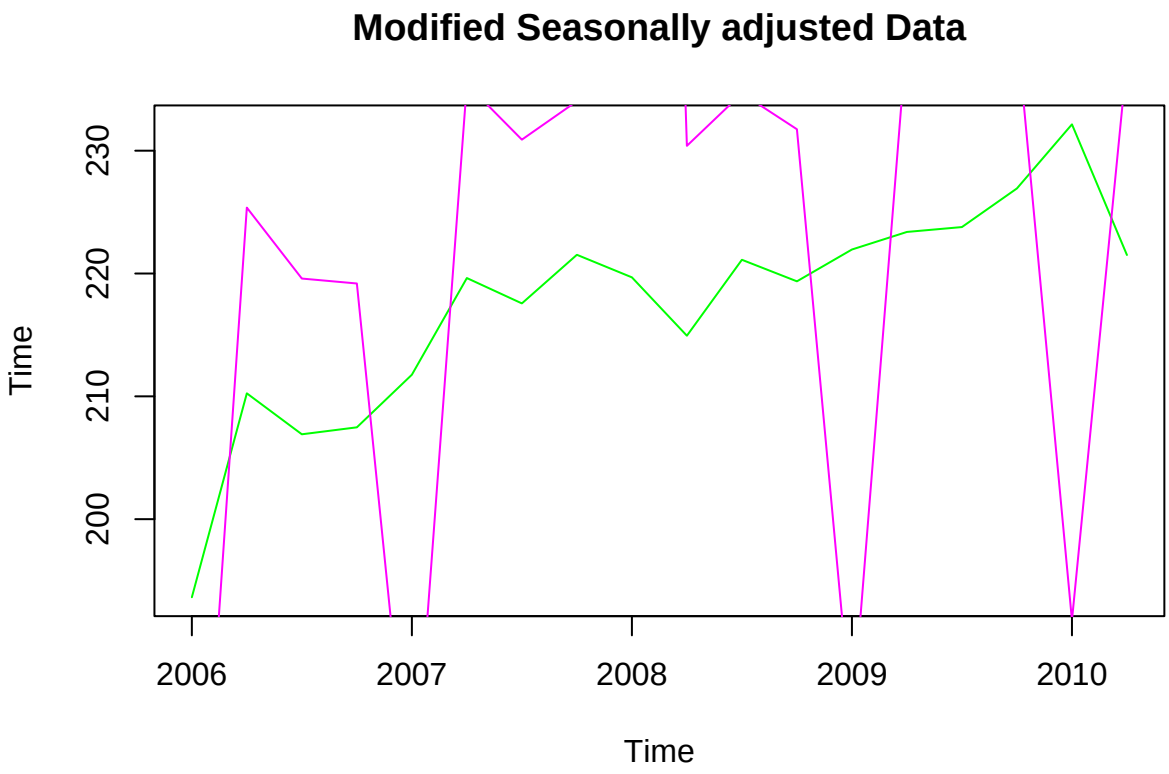
```
last_5_years_recompute <- decompose(last_5_years, type = "multiplicative")
```

#Now we need to obtain seasonally adjusted data, as we have done in previous steps

```
last_5_years_seasonally_adjusted_recompute <- last_5_years/ last_5_years_recompute$seasonal
```

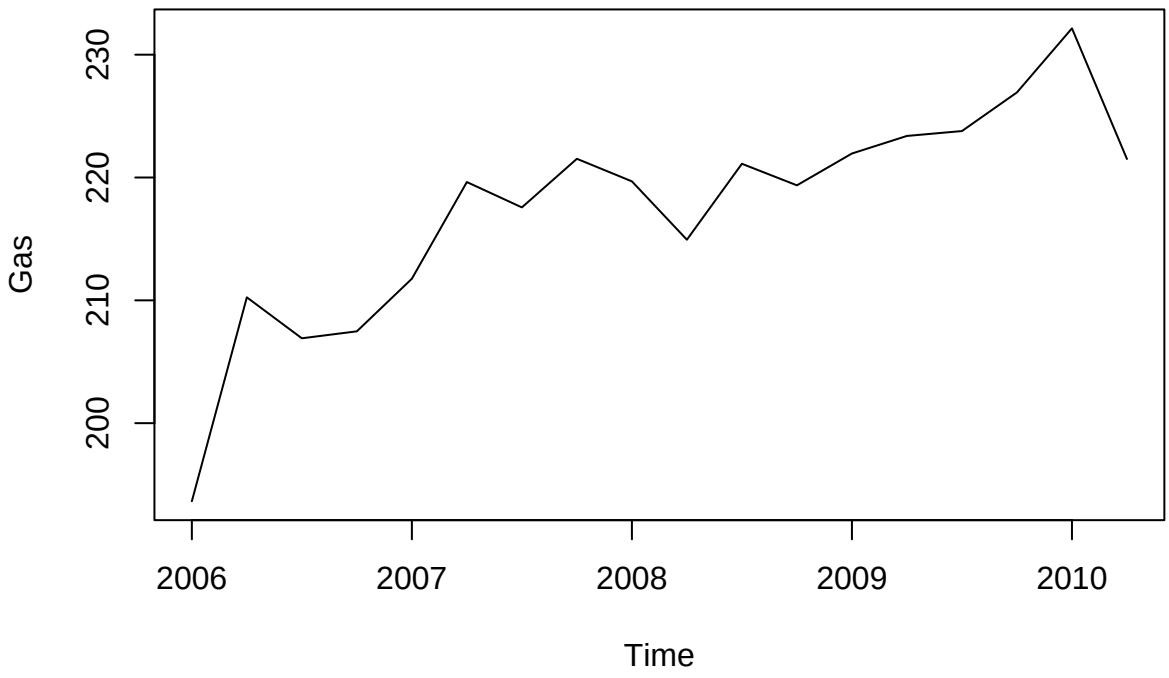
#now plot the modified seasonally adjusted data

```
par(mfrow = c(1,1))
plot(last_5_years_seasonally_adjusted, main = "Modified Seasonally adjusted Data", ylab = "Time",  col =
lines(last_5_years_seasonally_adjusted_recompute, col = "magenta")
```

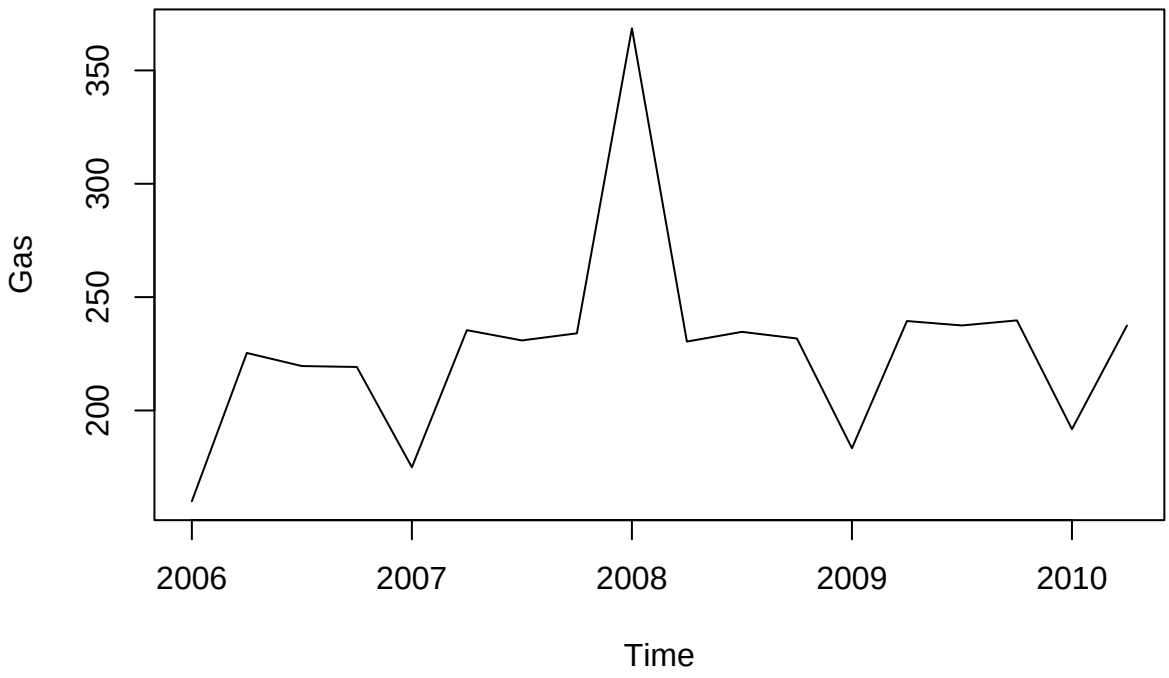


#Now let us plot the original and modified seasonally adjusted data

```
par(mfrow = c(1,1))
plot(last_5_years_seasonally_adjusted, ylab = "Gas")
```




```
plot(last_5_years_seasonally_adjusted_recompute, ylab = "Gas")
```



Does it make any difference if the outlier is near the end rather than in the middle of the time series? Why or why not?

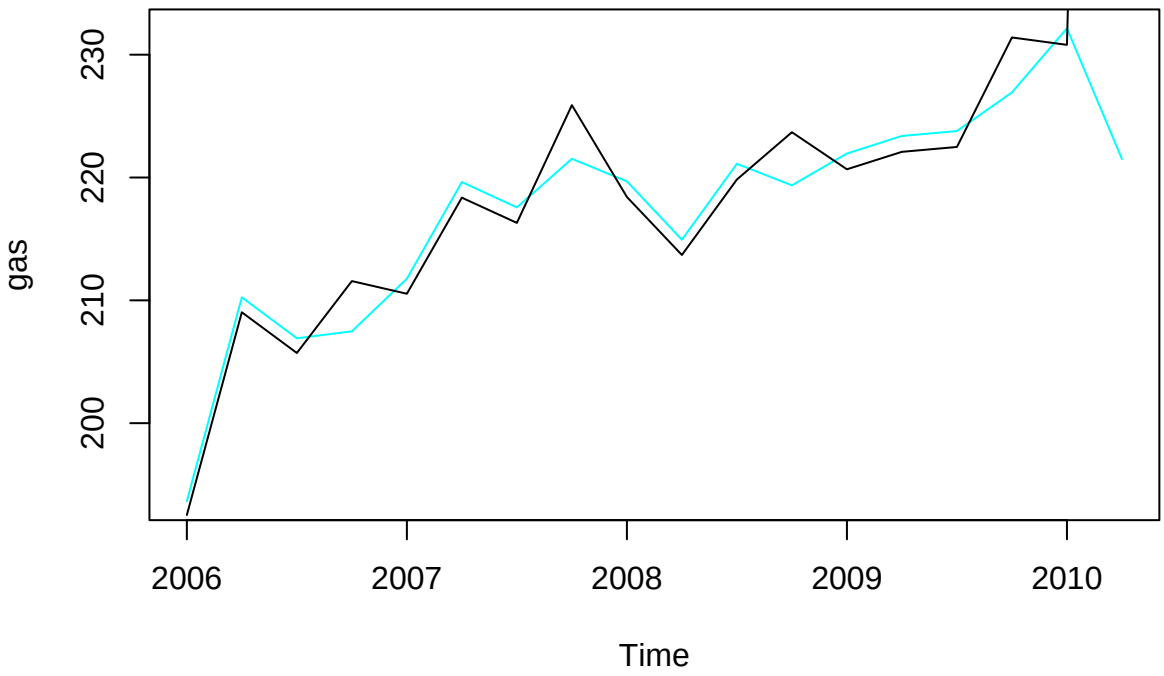
#Now let us check what happens if the outliner is at the end.

```
last_5_years[9] <- last_5_years[9] - 200
last_5_years[18] <- last_5_years[18] + 200
last_5_years
```

```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 2006  171  224  233  192
## 2007  187  234  245  205
## 2008  194  229  249  203
## 2009  196  238  252  210
## 2010  205  436
```

#Now we need to decompose , we need to follow the steps of ques 4

```
outliner_at_the_end <- decompose(last_5_years, type = "multiplicative")
seasonally_outliner_at_the_end <- last_5_years/outliner_at_the_end$seasonal
plot(last_5_years_seasonally_adjusted, ylab = "gas",col = 'cyan')
lines(seasonally_outliner_at_the_end, ylab ="gas", col = 'black')
```



When the outliner is towards the end as opposed to the centre, there is a difference. In the seasonally corrected data above, the outliner at the end of the series has moved towards the right, however, the outliner in the middle of the data caused a greater vertical than a horizontal movement.