DSCI_5340_HW1_Group10

Use the last five years of the Gas production data from aus_production data available from the tsibbledata package and perform the following tasks. Include detailed explanations, wherever appropriate.

```
pacman::p_load(dplyr, fpp3, GGally, gridExtra,lubridate, tidyverse)
theme_set(theme_classic())
search()
```

```
##  [1] ".GlobalEnv"         "package:forcats"     "package:stringr"
##  [4] "package:purrr"      "package:readr"       "package:tidyverse"
##  [7] "package:gridExtra"  "package:GGally"      "package:fable"
## [10] "package:feasts"     "package:fabletools"  "package:tsibbledata"
## [13] "package:tsibble"    "package:ggplot2"     "package:lubridate"
## [16] "package:tidyr"      "package:tibble"      "package:fpp3"
## [19] "package:dplyr"      "package:stats"       "package:graphics"
## [22] "package:grDevices"  "package:utils"       "package:datasets"
## [25] "package:methods"    "Autoloads"           "package:base"
```

Use the last five years of the Gas production data from aus_production data available from the tsibbledata package and perform the following tasks. Include detailed explanations, wherever appropriate.

#For selecting last five years, we have used tail function. #As aus_production is on the Quaterly basis, we have used 4*5 (No of quarters x No. of years)

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

```
 tail(aus_production, 4*5)
```

Quarter <qtr>	Beer <dbl>	Tobacco <dbl>	Bricks <dbl>	Cement <dbl>	Electricity <dbl>	Gas <dbl>
2005 Q3	408	NA	NA	2340	56043	221
2005 Q4	482	NA	NA	2265	54992	180
2006 Q1	438	NA	NA	2027	57112	171
2006 Q2	386	NA	NA	2278	57157	224
2006 Q3	405	NA	NA	2427	58400	233
2006 Q4	491	NA	NA	2451	56249	192
2007 Q1	427	NA	NA	2140	56244	187
2007 Q2	383	NA	NA	2362	55036	234
2007 Q3	394	NA	NA	2536	59806	245
2007 Q4	473	NA	NA	2562	56411	205

1-10 of 20 rows Previous 1 2 Next

#Assigning the last 5 years data to gas_prod_data

```
gas_production_data <- tail(aus_production, 4*5) %>%
  select(Gas)
```

Q1. Plot the above time series. Do these data show a trend-cycle and/or seasonal fluctuations? Elaborate on your answer.

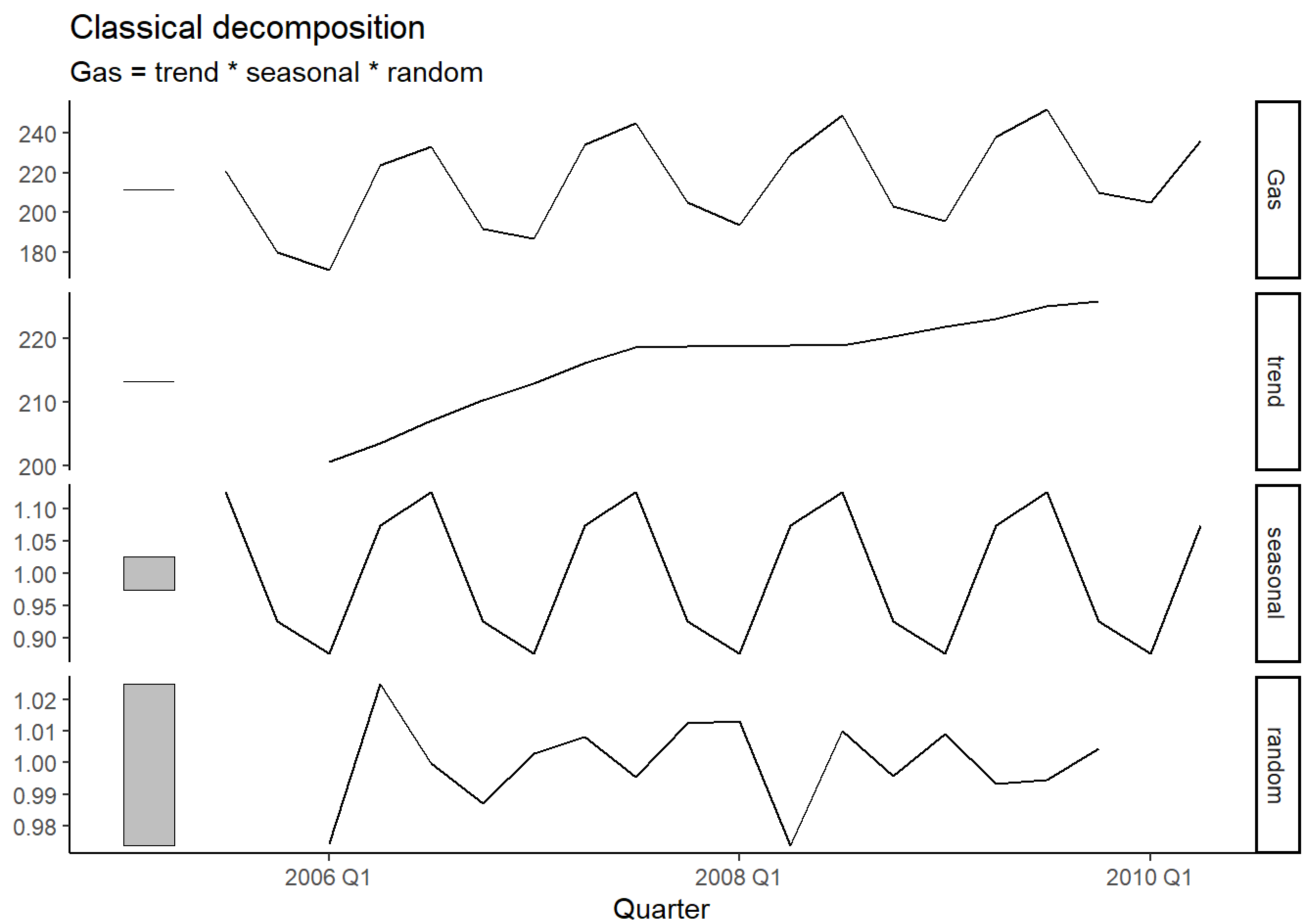
```
gas_production_data <- tail(aus_production, 4*5) %>%
  select(Gas)
gas_production_data %>%
  autoplot(Gas,color = "red") +
  labs(y = "Petajoules")
```

#from the below plotted graph, we can see an ascending pattern or upward trend. Gas production is lowest in Quarter 1's for the every respective year and highest between Quarter 2 and 3, for every respective year. #positive trend is observed from 2006-2010 Q1. #We can observe repeating patterns of peaks and troughs from the data so it would be more helpful to predic specific seasonal pattern.

Q2. Use `classical_decomposition` with `type=multiplicative` to calculate the trend-cycle and seasonal indices. Do the results support the graphical interpretation in the question above?

```
decomp <- gas_production_data %>%
  model(decomp = classical_decomposition(Gas, type = "multiplicative")) %>%
  components()
decomp %>%
  autoplot()
```

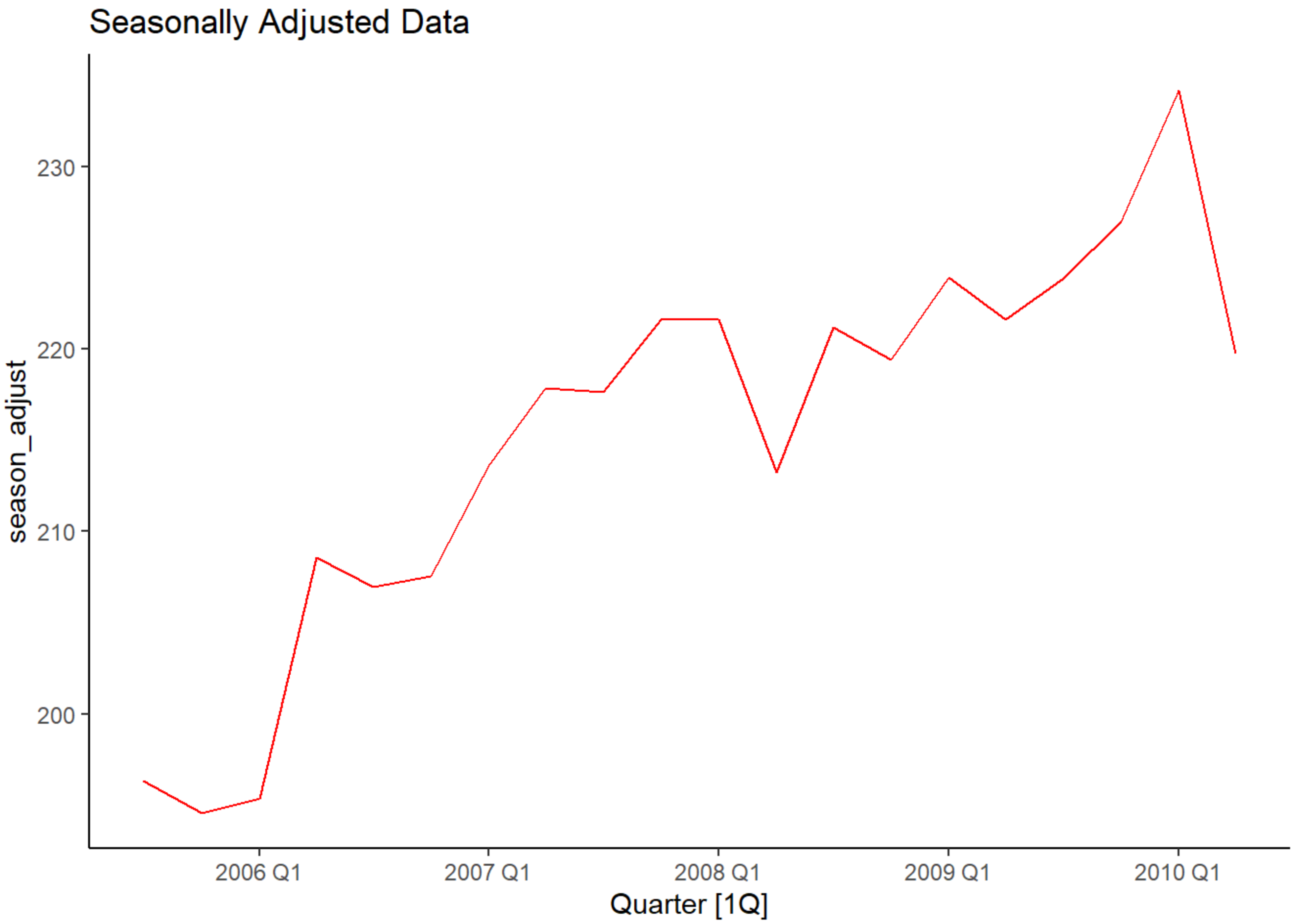
```
## Warning: Removed 2 rows containing missing values (`geom_line()`).
```



#Trend appears to be non-linear between 2007 and 2008.Seasonal trend appears to be sideways. Overall it is observed a positive trend.

Q3. Compute and plot the seasonally-adjusted data.

```
as_tsibble(decomp) %>%
autoplot(season_adjust,color ="red") +
ggtitle("Seasonally Adjusted Data")
```




```
gas_production_data
```

Gas <dbl>	Quarter <qtr>
221	2005 Q3
180	2005 Q4
171	2006 Q1
224	2006 Q2
233	2006 Q3
192	2006 Q4
187	2007 Q1
234	2007 Q2
245	2007 Q3
205	2007 Q4

1-10 of 20 rows Previous **1** 2 Next

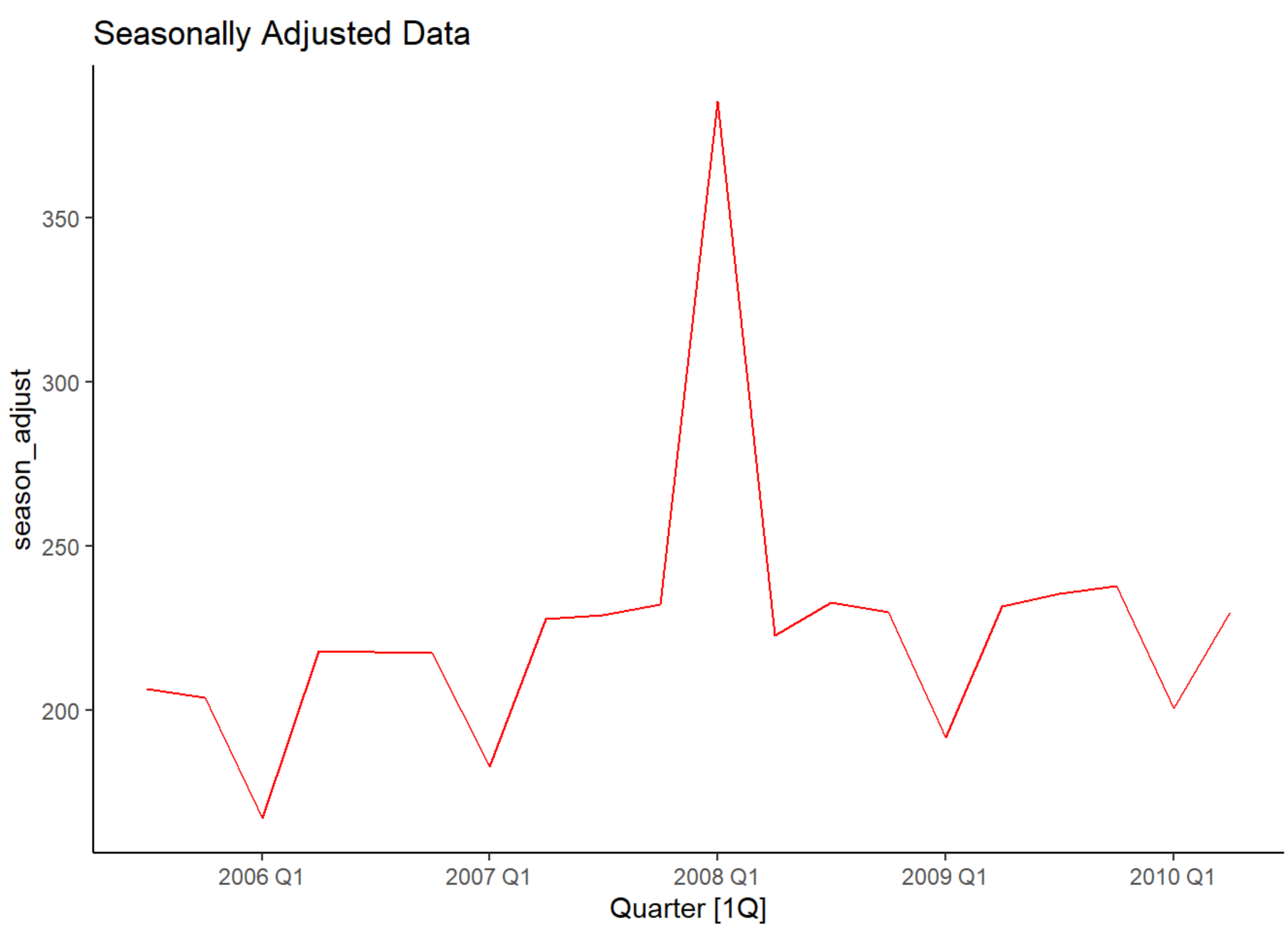
The line chart shows an upward trend over time, with some fluctuations.

Notably, the seasonally adjusted data increased during the observed period.

Q4. 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.

```
gas_production_data[11,1] <- gas_production_data[11,1]+200
gas_production_data %>%
  model(
    classical_decomposition(Gas, type = "multiplicative")
  ) %>%
  components() %>% select(season_adjust) %>%
  autoplot(color = "red") +
  ggtitle("Seasonally Adjusted Data")
```

```
## Plot variable not specified, automatically selected `.vars = season_adjust`
```



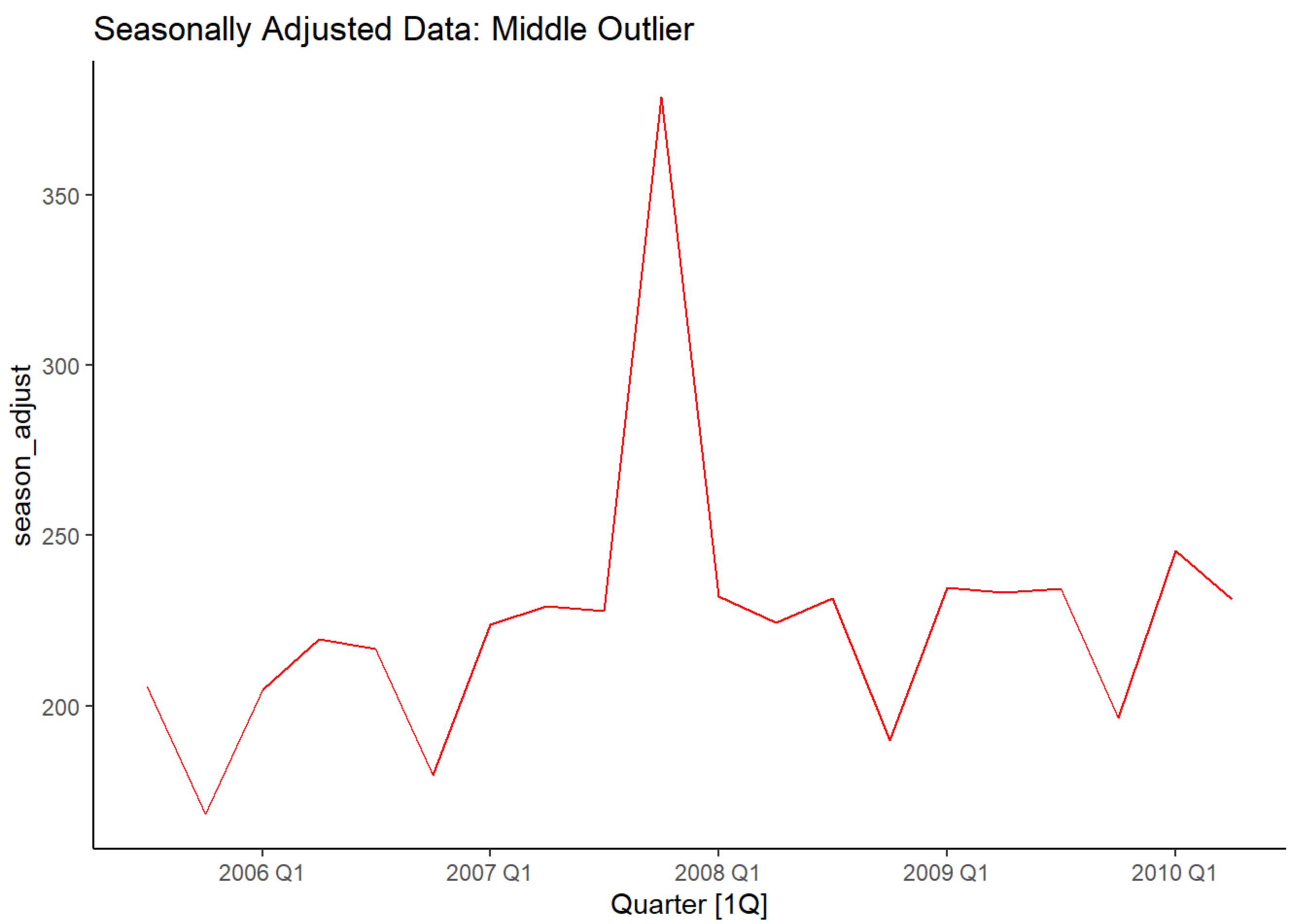
#The addition of 200 in 2008 Q1, which is considered an outlier, causes a sharp increase in production, elevating it from approximately 220 petajoules to over 350 petajoules for that particular period. This outlier disrupts the previously established seasonal data estimates.

Q5. 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?

Outlier in the middle

```
middle <- tail(aus_production, 5*4) %>% select(Gas)
middle[10,1] = middle[10,1] + 200
middle %>%
  model(
    classical_decomposition(Gas, type = "multiplicative")
  ) %>%
  components() %>% select(season_adjust) %>%
  autoplot(color ="red") +
  ggtitle("Seasonally Adjusted Data: Middle Outlier")
```

```
## Plot variable not specified, automatically selected `.vars = season_adjust`
```

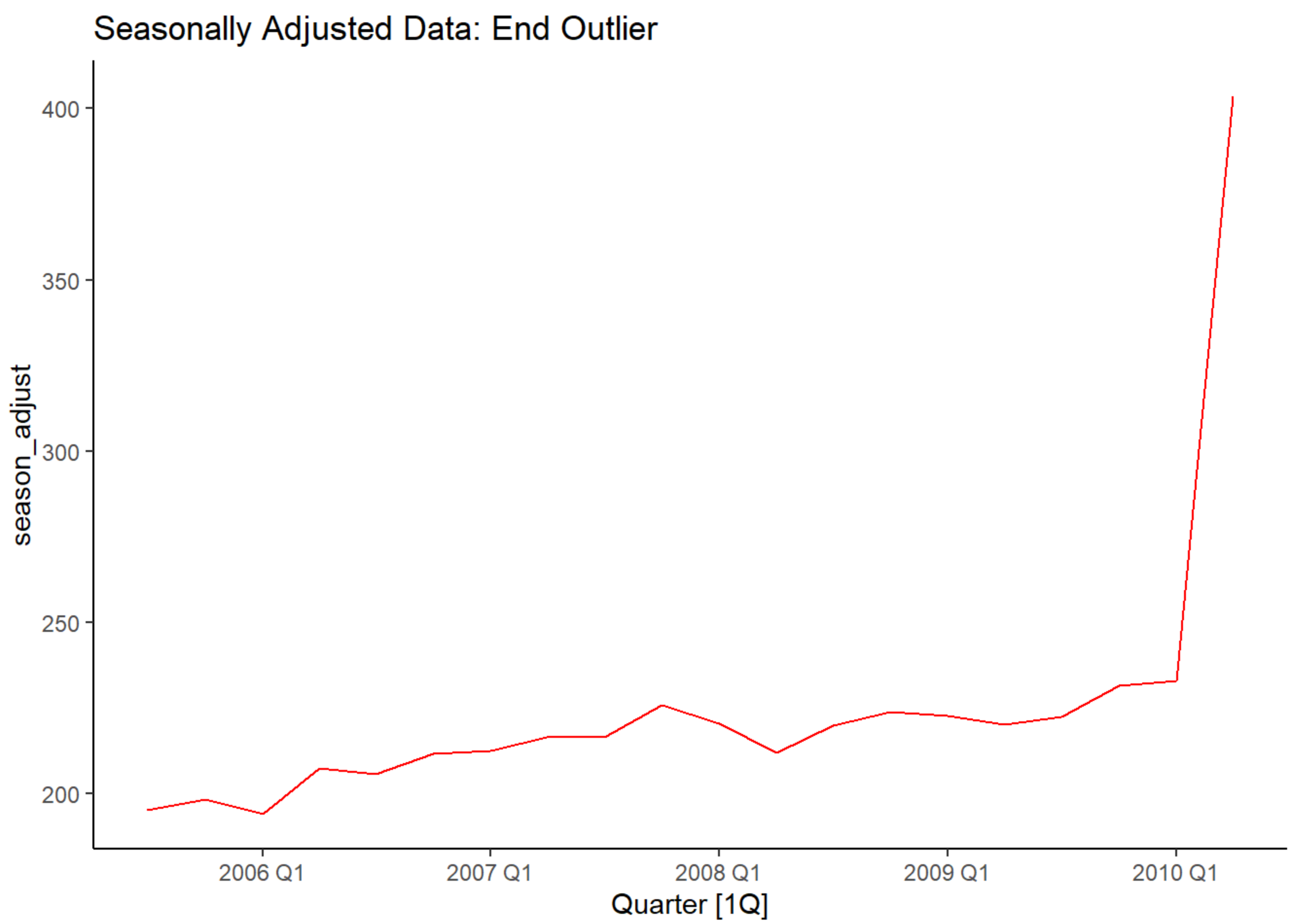


#

Outlier in the end

```
end <- tail(aus_production, 5*4) %>% select(Gas)
end[20,1] = end[20,1] + 200
end %>%
  model(
    classical_decomposition(Gas, type = "multiplicative")
  ) %>%
  components() %>% select(season_adjust) %>%
  autoplot(color = "red") +
  ggtitle("Seasonally Adjusted Data: End Outlier")
```

```
## Plot variable not specified, automatically selected `.vars = season_adjust`
```



#When we add 200 to the last available quarter which is Q2 of 2010, we observe a decrease in seasonal fluctuations compared to the interim periods. This demonstrates that the placement of an outlier significantly influences the overall trend.