Exploratory Analytics and Time Series Modeliing on Petroleum product price Challenge

Abhinav Pathak

```
suppressPackageStartupMessages({
  library(tidyr)
  library(dplyr)
  library(ggplot2)
  library(readxl)
  library(Quandl)
  library(gridExtra)
  library(TSA)
  library(forecast)
  library(sqldf)
})
```

Loading data from excel file

```
crude_oil_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 1", skip = 2)
colnames(crude_oil_data) <- c("Date", "crude_cushing_WTI", "crude_brent_eu")

conv_gasoln_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 2", skip = 2)
colnames(conv_gasoln_data) <- c("Date", "conv_gaso_ny", "conv_gaso_USgulf")

reg_gasoln_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 3", skip = 2)
colnames(reg_gasoln_data) <- c("Date", "reg_gas_la")

heat_oil_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 4", skip = 2)
colnames(heat_oil_data) <- c("Date", "heat_oil_ny")

sulfr_dsl_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 5", skip = 2)
colnames(sulfr_dsl_data) <- c("Date", "sulfr_dsl_ny","sulfr_dsl_USgulf","sulfr_dsl_la")

kersn_jet_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 6", skip = 2)
colnames(kersn_jet_data) <- c("Date", "kersn_jet_USgulf")

propn_data <- read_excel("Jr._Data_Analyst_Project_File.xls", sheet = "Data 7", skip = 2)
colnames(propn_data) <- c("Date", "propn_montBel")
```

Putting all the data in one table

```
crude_oil_data <- crude_oil_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))
heat_oil_data <- heat_oil_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))
propn_data <- propn_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))
reg_gasoln_data <- reg_gasoln_data %>% mutate(year = format(Date, "%Y"),
```

```
                          month = format(Date, '%m'))
conv_gasoln_data <- conv_gasoln_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))
sulfr_dsl_data <- sulfr_dsl_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))
kersn_jet_data <- kersn_jet_data %>% mutate(year = format(Date, "%Y"),
                          month = format(Date, '%m'))

petroleum_product_data <- crude_oil_data %>%
  full_join(heat_oil_data, by = c("month" = "month", "year" = "year")) %>%
  full_join(propn_data, by = c("month" = "month", "year" = "year")) %>%
  full_join(reg_gasoln_data, by = c("month" = "month", "year" = "year")) %>%
  full_join(conv_gasoln_data, by = c("month" = "month", "year" = "year")) %>%
  full_join(sulfr_dsl_data, by = c("month" = "month", "year" = "year")) %>%
  full_join(kersn_jet_data, by = c("month" = "month", "year" = "year")) %>%
  select(-Date.y, -Date.x.x, -Date.y.y, -Date.x.x.x, -Date.y.y.y,-Date)
```

Question 1

Visualizing the change in prices across time

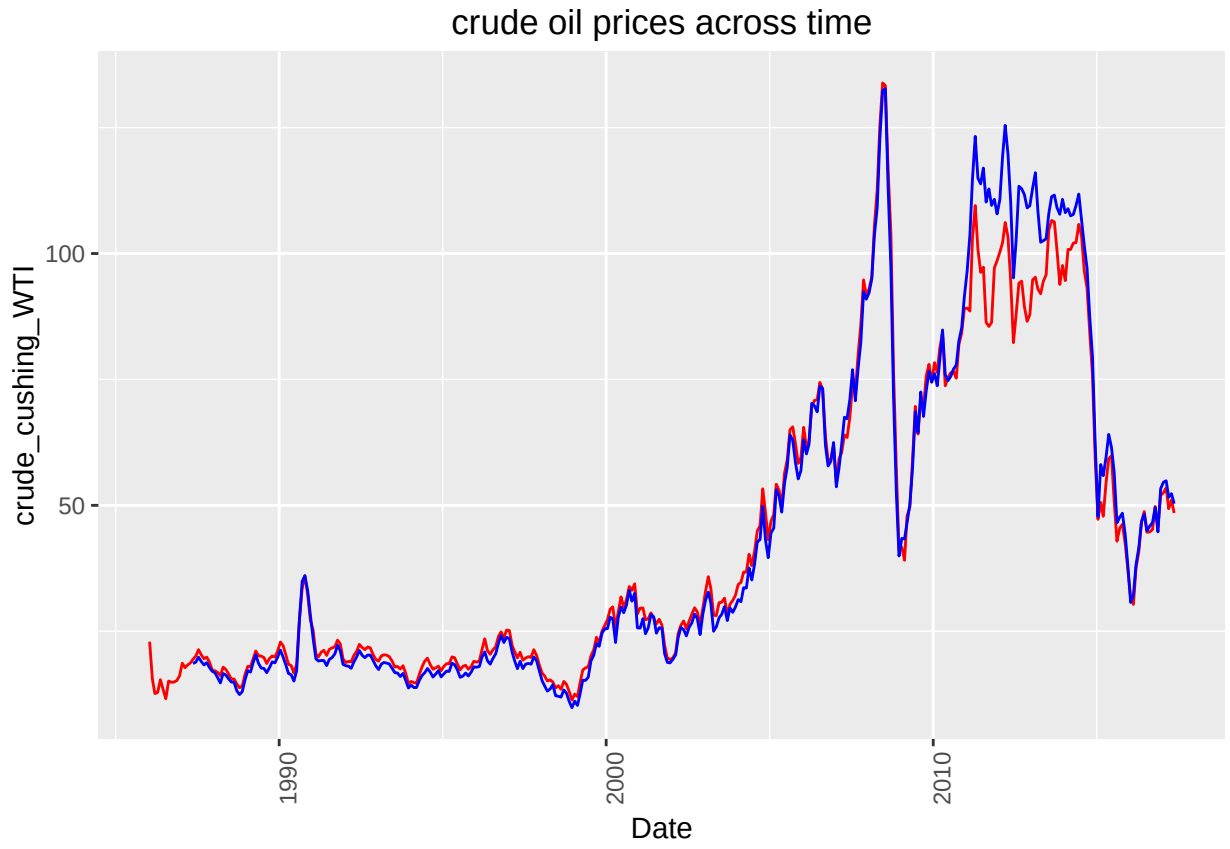
```
g1 <- ggplot(crude_oil_data) +
  geom_line(aes(x= Date, y= crude_cushing_WTI, group =1 ), color = "red") +
  geom_line(aes(x= Date, y= crude_brent_eu ), color = "blue") +
  ggtitle("crude oil prices across time") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

g2 <- ggplot(conv_gasoln_data) +
  geom_line(aes(x= Date, y= conv_gaso_ny , group =1 ), color = "red") +
  geom_line(aes(x= Date, y= conv_gaso_USgulf ), color = "blue") +
  ggtitle("Conventional gasoline prices across time")

g3 <- ggplot(reg_gasoln_data) +
  geom_line(aes(x= Date, y= reg_gas_la , group =1 ), color = "red") +
  ggtitle("Regular gasoline prices across time")

g4 <- ggplot(heat_oil_data) +
  geom_line(aes(x= Date, y= heat_oil_ny , group =1 ), color = "red") +
  ggtitle("Heat oil prices across time")

g5 <- ggplot(sulfr_dsl_data) +
  geom_line(aes(x= Date, y= sulfr_dsl_ny , group =1 ), color = "red") +
    geom_line(aes(x= Date, y= sulfr_dsl_USgulf), color = "blue") +
    geom_line(aes(x= Date, y= sulfr_dsl_la), color = "green") +
  ggtitle("Sulfur Diesel prices across time")

g6 <- ggplot(kersn_jet_data) +
  geom_line(aes(x= Date, y= kersn_jet_USgulf , group =1 ), color = "red") +
  ggtitle("Kerosene Jet fuel prices across time")

g7 <- ggplot(propn_data) +
```

```
  geom_line(aes(x= Date, y= propn_montBel , group =1 ), color = "red") +
  ggtitle("Propane fuel prices across time")
```

```
g1
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

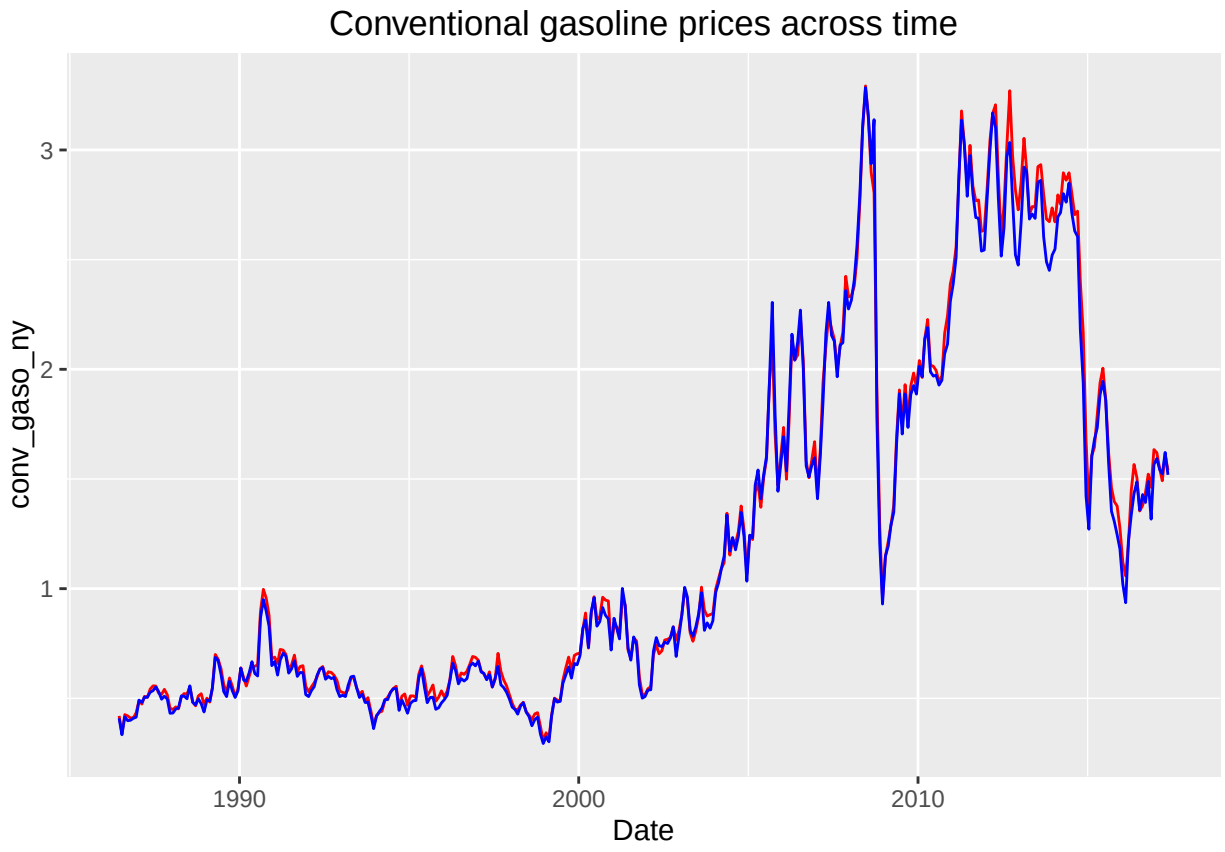
```
## Warning: Removed 17 rows containing missing values (geom_path).
```




```
g2
```

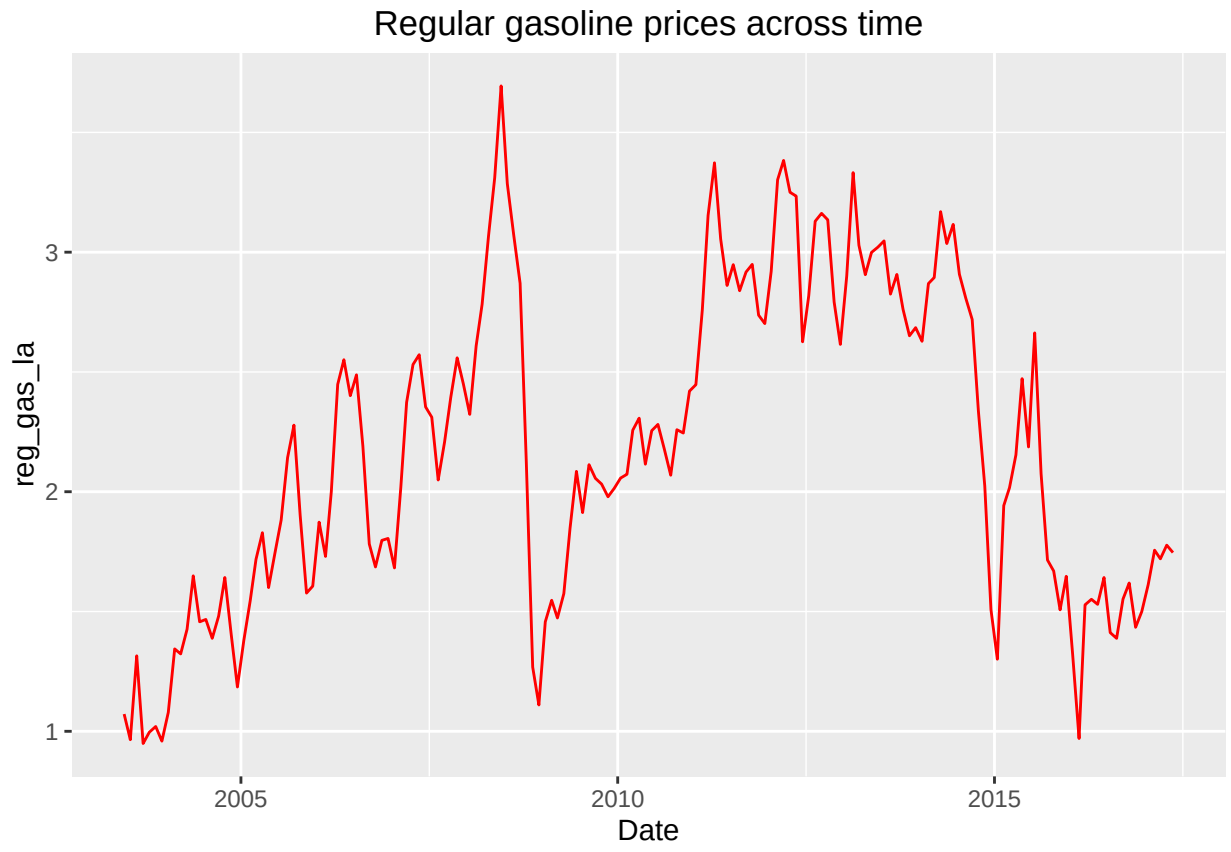
```
## Warning: Removed 1 rows containing missing values (geom_path).
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



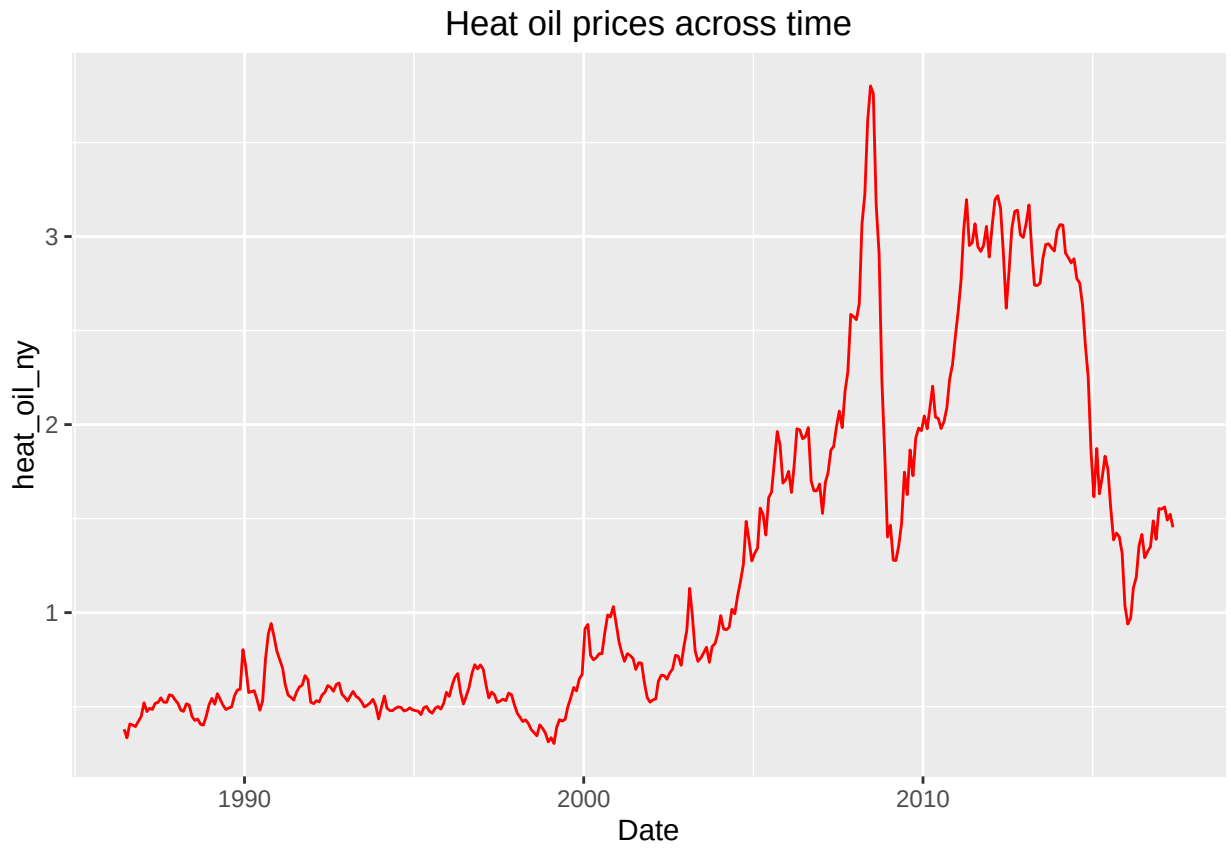

```
g3
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
g4
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

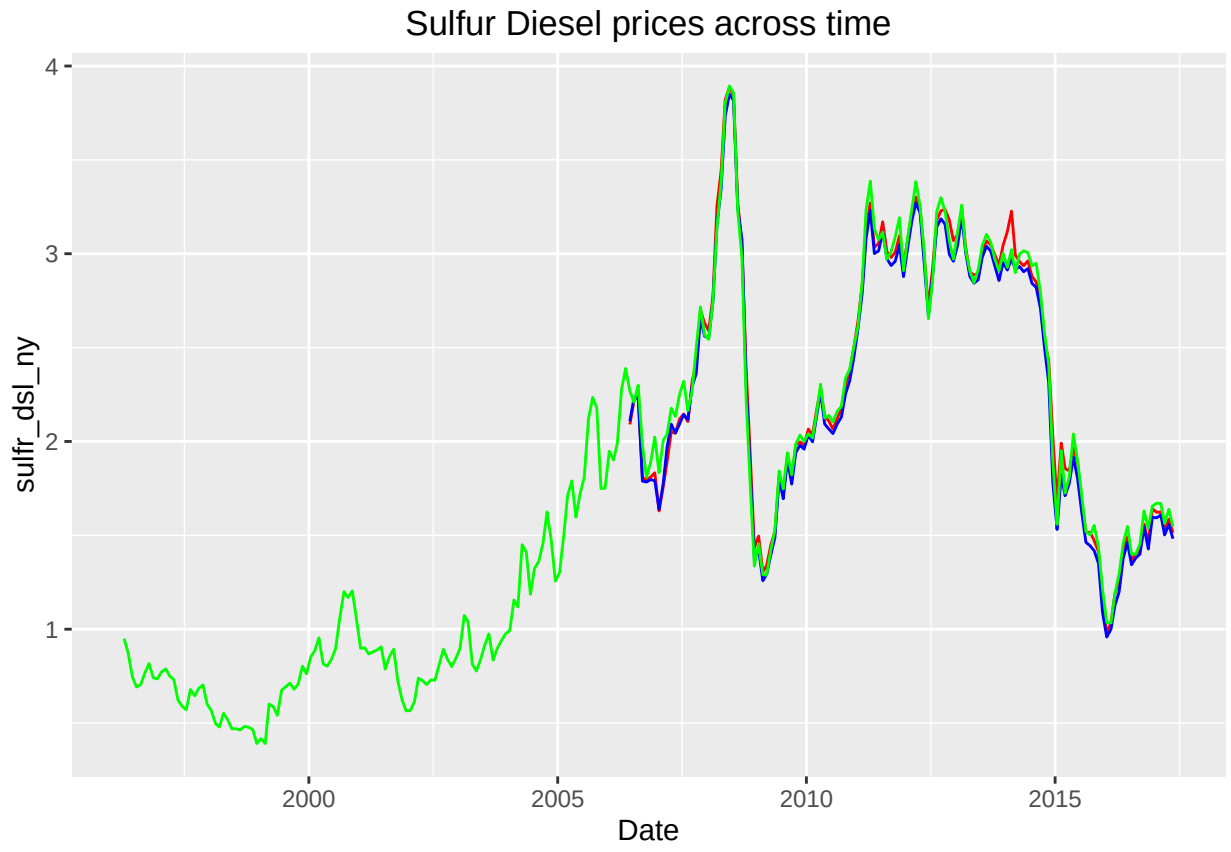



```
g5
```

```
## Warning: Removed 123 rows containing missing values (geom_path).
```

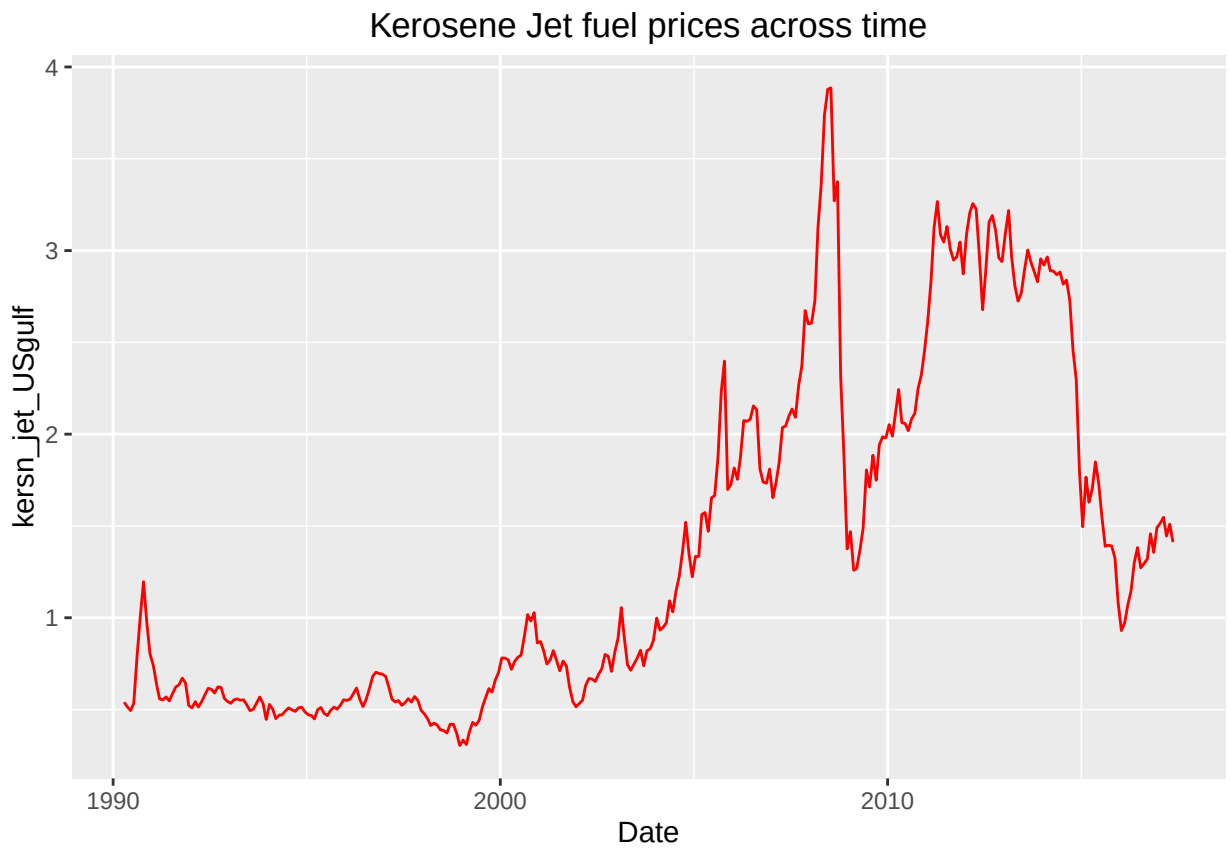
```
## Warning: Removed 123 rows containing missing values (geom_path).
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



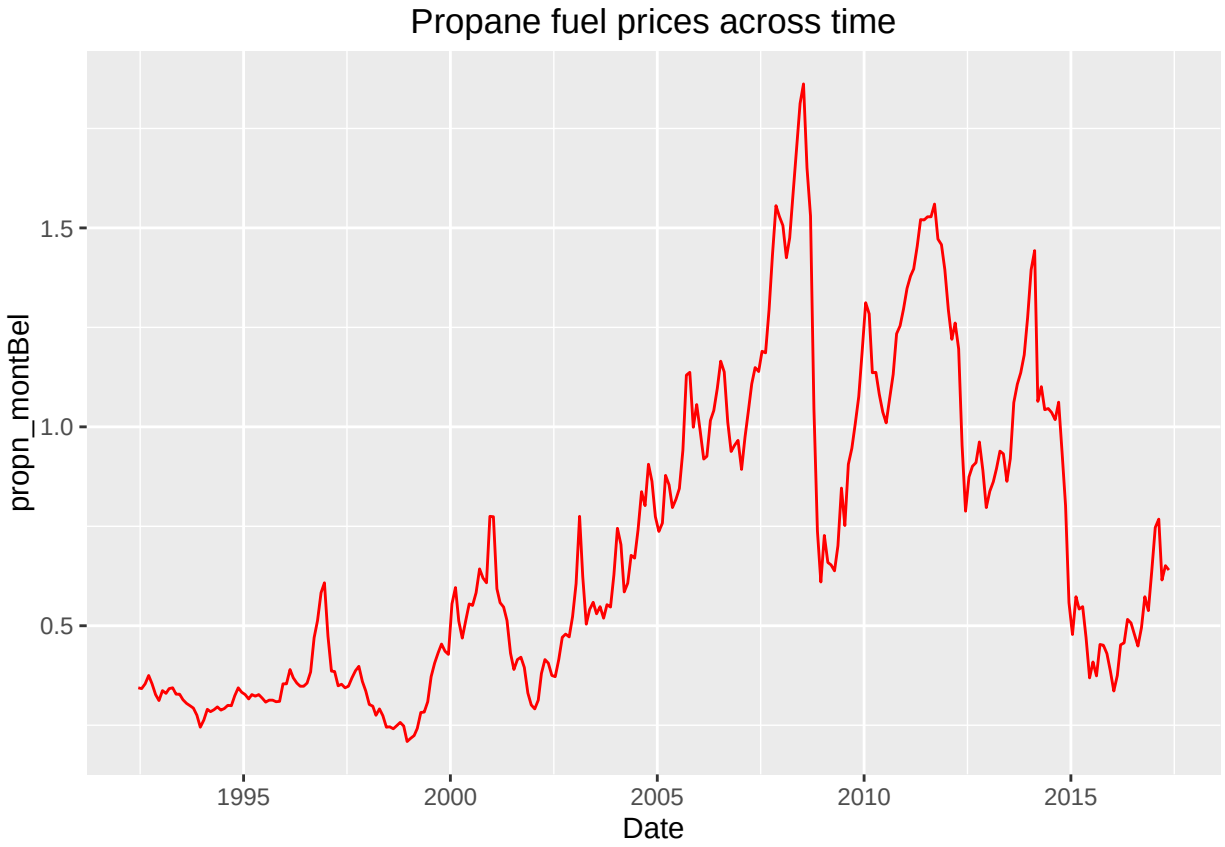

```
g6
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
g7
```

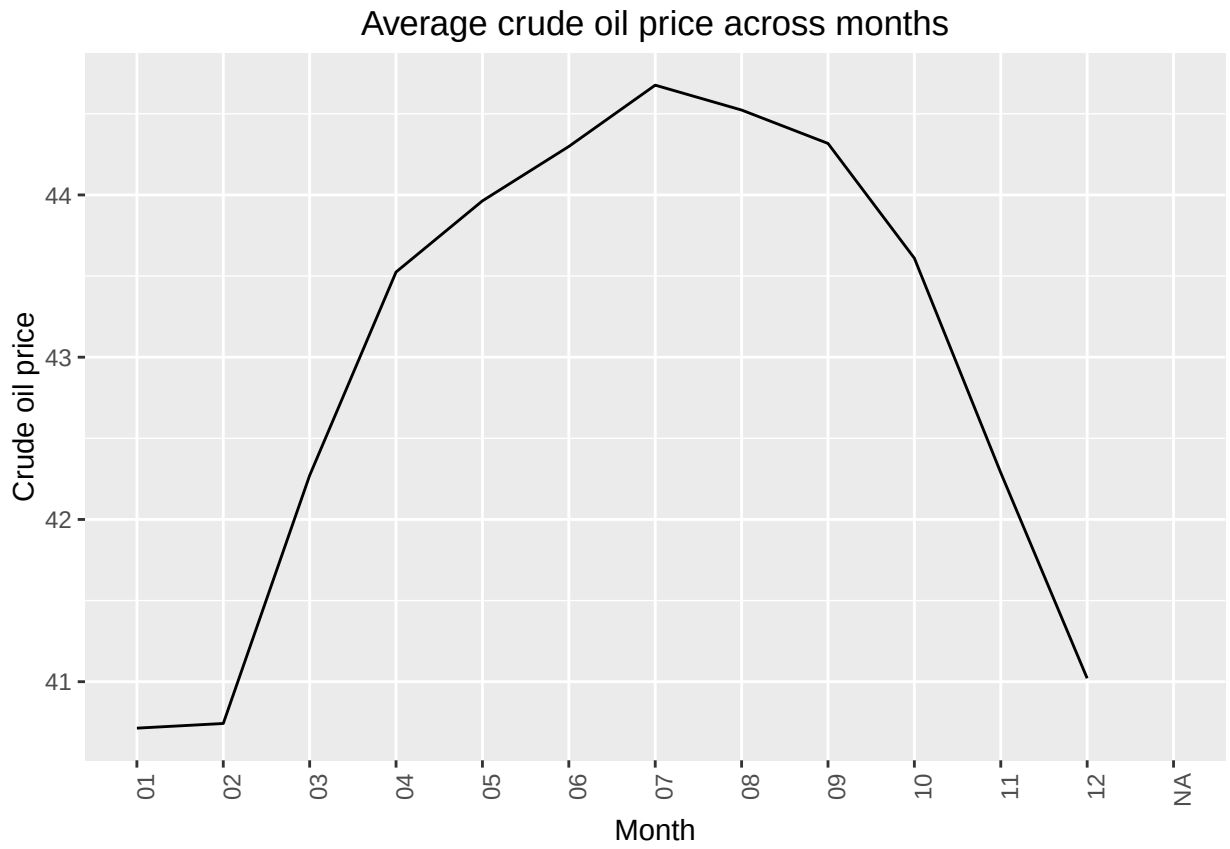
```
## Warning: Removed 1 rows containing missing values (geom_path).
```



- Insight1: There is a similar trend acoss all types of oil prices. Since 2002, oil prices had been rising consistently.
- Insight2: Oil Prices reached its peak in July 2008 and prices were very high on average between 2008 and 2014
- Insight3: Prices have not reduced significantly compared to the average price of last 5-6 year

```
month_wise_crude <- crude_oil_data %>% group_by(Date) %>%
  mutate(year = format(Date,'%Y'),
         month = format(Date, '%m')) %>%
  group_by(month) %>%
  summarise(avg_crude_cushing_price = mean(crude_cushing_WTI))

ggplot(month_wise_crude) +
  geom_line(aes(x= factor(month), y = avg_crude_cushing_price, group = 1 )) +
  labs(x = "Month", y = "Crude oil price", title = "Average crude oil price across months") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

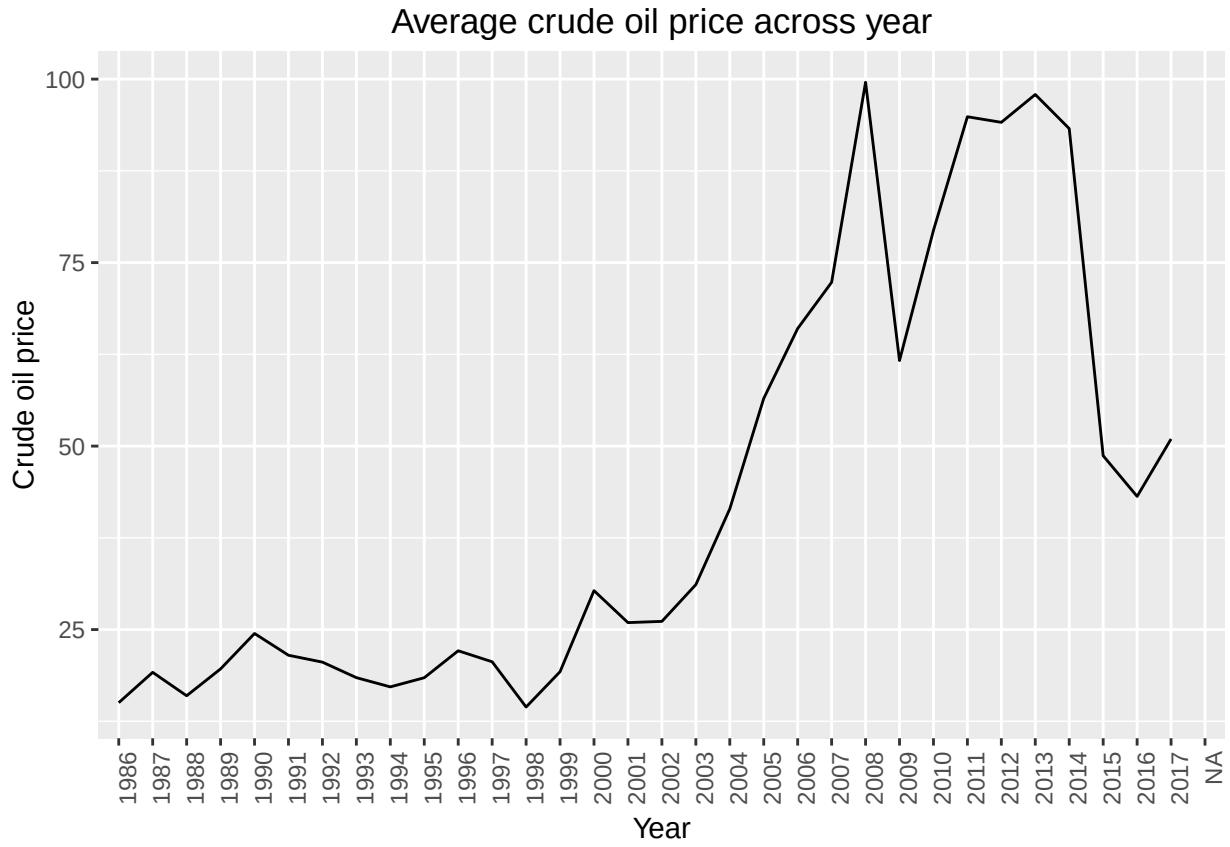


- Insight4: Price of the crude oil is generally higher in summer season (June-July-August) and low in winters (Dec -Jan)

Price across years

```
year_wise_crude <- crude_oil_data %>% group_by(Date) %>%
  mutate(year = format(Date,'%Y'),
         month = format(Date, '%m')) %>%
  group_by(year) %>%
  summarise(avg_crude_cushing_price = mean(crude_cushing_WTI))

ggplot(year_wise_crude) +
    geom_line(aes(x= factor(year), y = avg_crude_cushing_price, group = 1 )) +
    labs(x = "Year", y = "Crude oil price", title = "Average crude oil price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



- Insight5: Price of petroleumn product was at its peak between 2008 and 2014

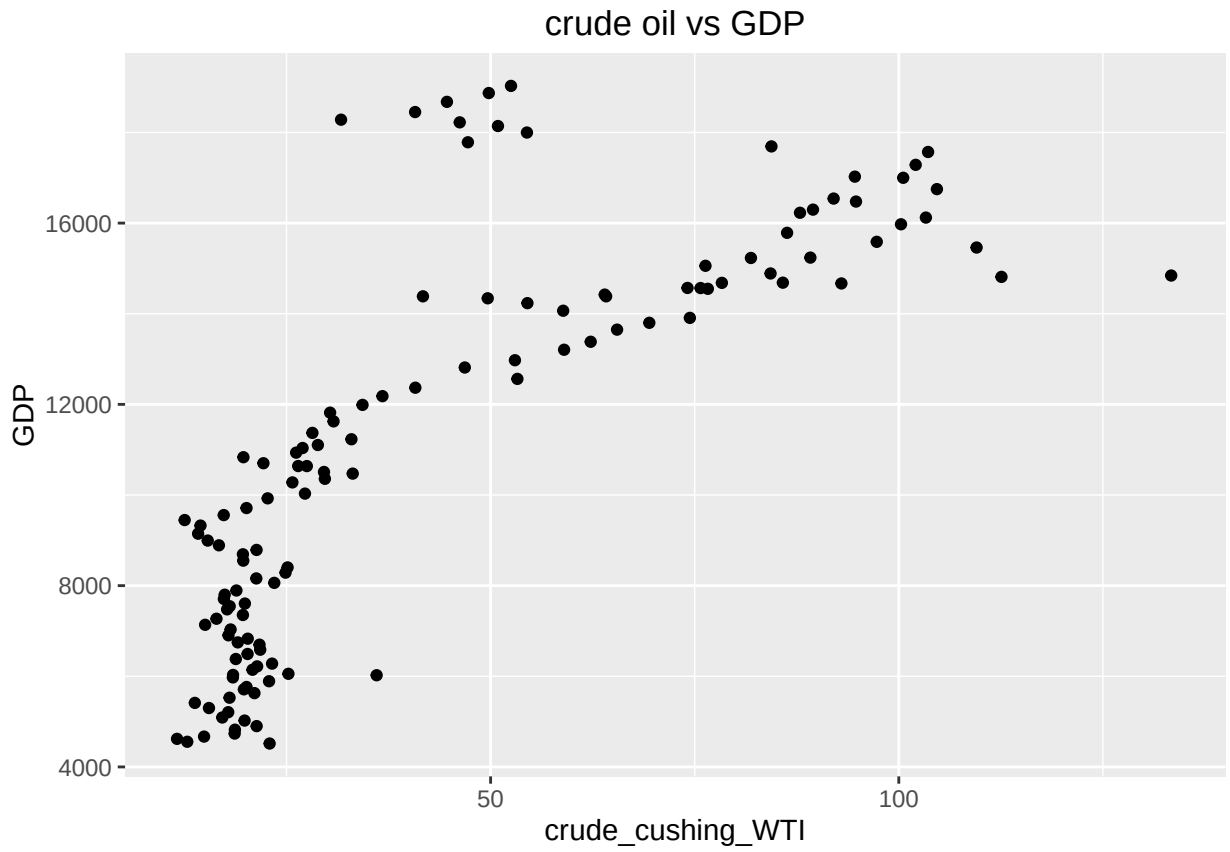
Question 2

Loading external data using Quandl package

```
GDP_data <- Quandl("FRED/GDP", collapse = "monthly" ,start_date="1986-01-01", end_date="2017-01-01") %>%
CPI_data = Quandl("FRED/CPIAUCSL",  collapse = "monthly" ,start_date="1986-01-01", end_date="2017-01-01
Inflation_data = Quandl("WORLDBANK/USA_FP_CPI_TOTL_ZG",  collapse = "monthly" ,start_date="1986-01-01",


External_data <- GDP_data %>%
  full_join(CPI_data, by = c("month" = "month", "year"="year")) %>%
  full_join(Inflation_data, by = c("month" = "month", "year"="year"))

petroleum_product_data.external <- petroleum_product_data %>%
  full_join(External_data,  by = c("month" = "month", "year"="year"))
```

```
ggplot(petroleum_product_data.external) +
  geom_point(aes(x =crude_cushing_WTI, y= GDP )) +
  ggtitle("crude oil vs GDP")
```

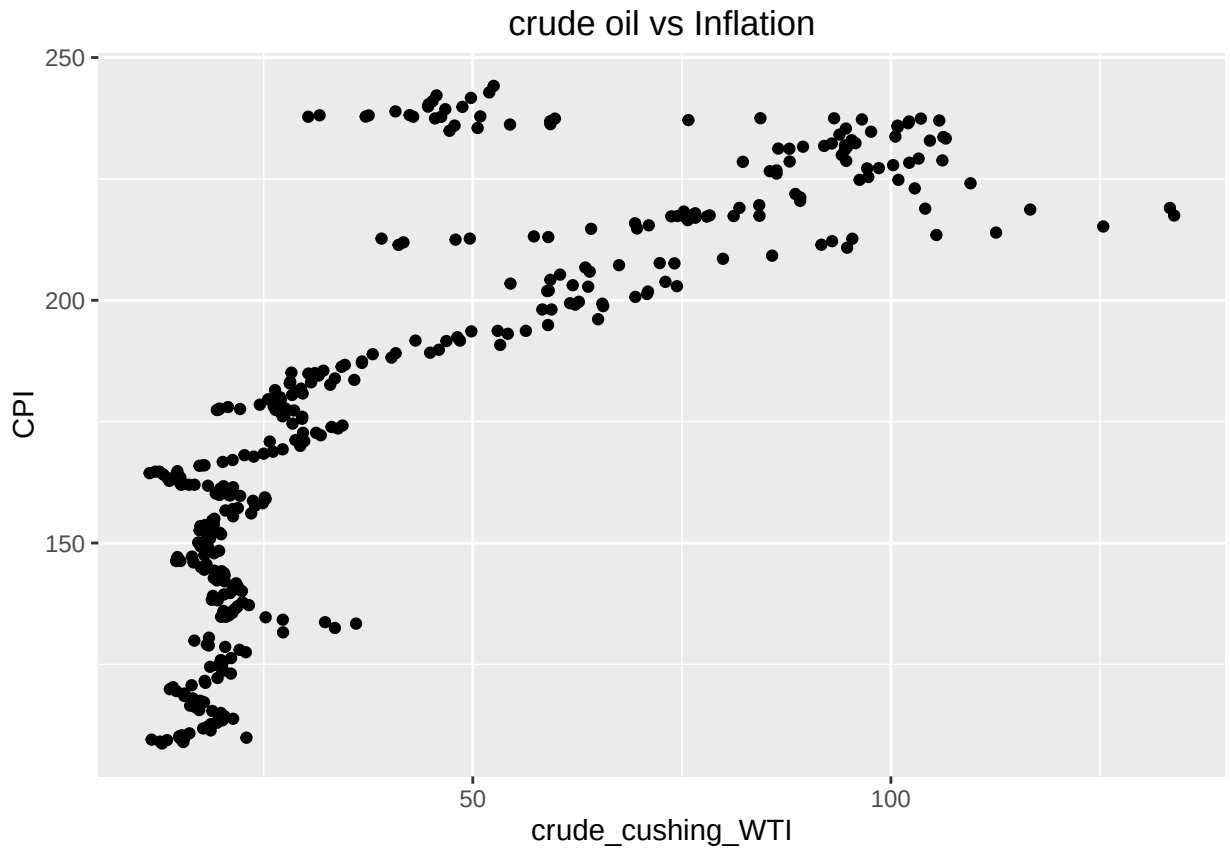
```
## Warning: Removed 281 rows containing missing values (geom_point).
```



- Oil prices are correlated with the growth of economy

```
ggplot(petroleum_product_data.external) +
  geom_point(aes(x =crude_cushing_WTI, y= CPI )) +
  ggtitle("crude oil vs Inflation")
```

```
## Warning: Removed 33 rows containing missing values (geom_point).
```



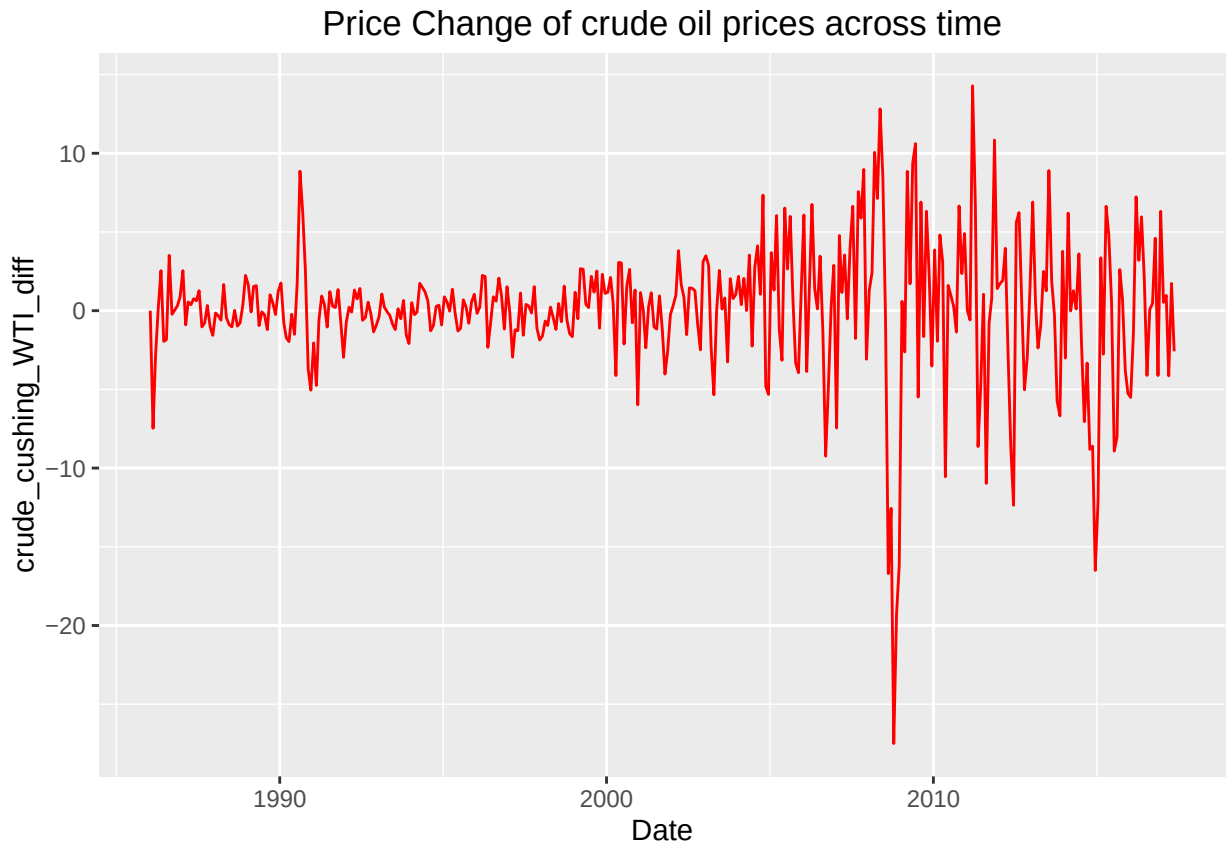
- Oil prices are also increasin with consumer price index. Therefore it is clear that external factors are also dependent in a way on oil prices although nothing can be said about causal relationships and it is just a correlation

Question 3

Price changes month over month

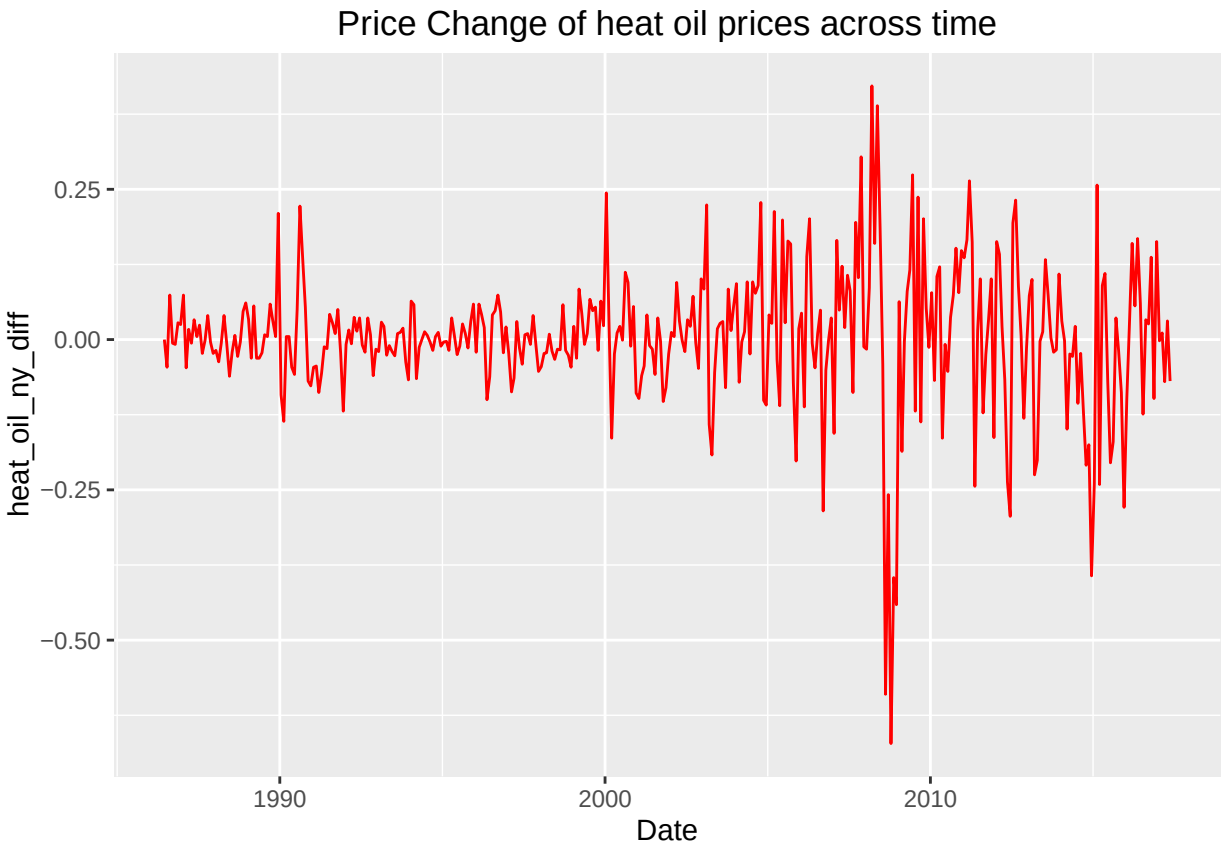
```
crude_oil_data$crude_cushing_WTI_diff <- c(0, diff(crude_oil_data$crude_cushing_WTI, lag = 1))
ggplot(crude_oil_data) +
  geom_line(aes(x= Date, y= crude_cushing_WTI_diff), color = "red") +
  ggtitle("Price Change of crude oil prices across time")
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
heat_oil_data$heat_oil_ny_diff <- c(0, diff(heat_oil_data$heat_oil_ny, lag = 1))
ggplot(heat_oil_data) +
  geom_line(aes(x= Date, y= heat_oil_ny_diff), color = "red") +
  ggtitle("Price Change of heat oil prices across time")
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



- Price has changed more abruptly from 2008 to 2015, whereas prior to 2008 the prices were a little stable

Question 4

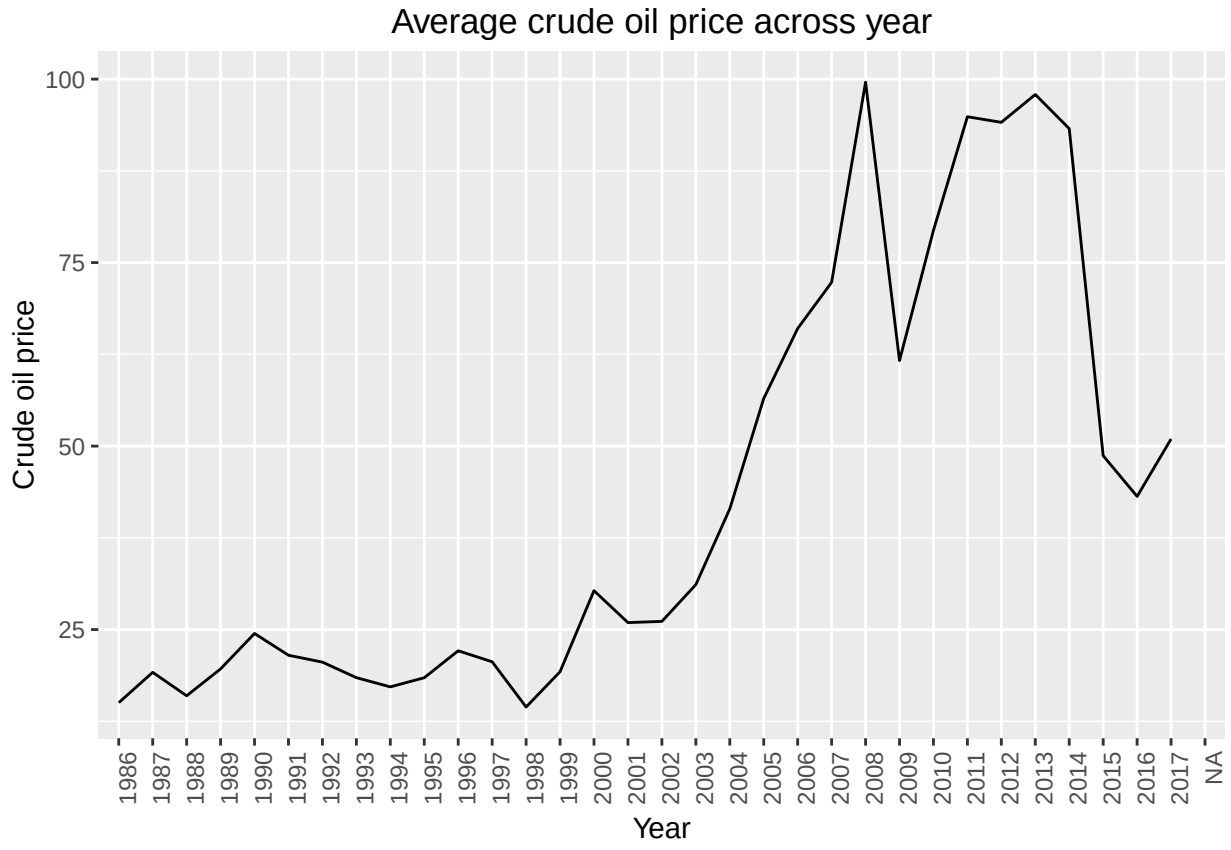
Price changes and movement across year

```
petroleum_product_data_byYear <- petroleum_product_data %>%
  group_by(year)  %>%
  summarise(crude_cushing_WTI = mean(crude_cushing_WTI, na.rm = TRUE),
            heat_oil_ny = mean(heat_oil_ny, na.rm = TRUE),
            propn_montBel =mean(propn_montBel, na.rm = TRUE),
            reg_gas_la = mean(reg_gas_la, na.rm = TRUE),
            conv_gaso_ny = mean(conv_gaso_ny, na.rm = TRUE),
            sulfr_dsl_ny = mean(sulfr_dsl_ny, na.rm = TRUE),
            kersn_jet_USgulf = mean(kersn_jet_USgulf, na.rm = TRUE))

h1 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = crude_cushing_WTI, group = 1 )) +
    labs(x = "Year", y = "Crude oil price", title = "Average crude oil price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

h2 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = heat_oil_ny, group = 1 )) +
    labs(x = "Year", y = "heat oil ny", title = "Average heat oil price across year") +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

h3 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = propn_montBel, group = 1 )) +
    labs(x = "Year", y = "price", title = "Average propane price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

h4 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = reg_gas_la, group = 1 )) +
    labs(x = "Year", y = "price", title = "Average regular gasoline price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

h5 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = conv_gaso_ny, group = 1 )) +
    labs(x = "Year", y = "price", title = "Average conventional gasoline price across year") +   theme(a

h6 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = sulfr_dsl_ny, group = 1 )) +
    labs(x = "Year", y = "price", title = "Average sulfur diesel price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

h7 <- ggplot(petroleum_product_data_byYear) +
    geom_line(aes(x= factor(year), y = kersn_jet_USgulf, group = 1 )) +
    labs(x = "Year", y = "price", title = "Average kerosene jet price across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

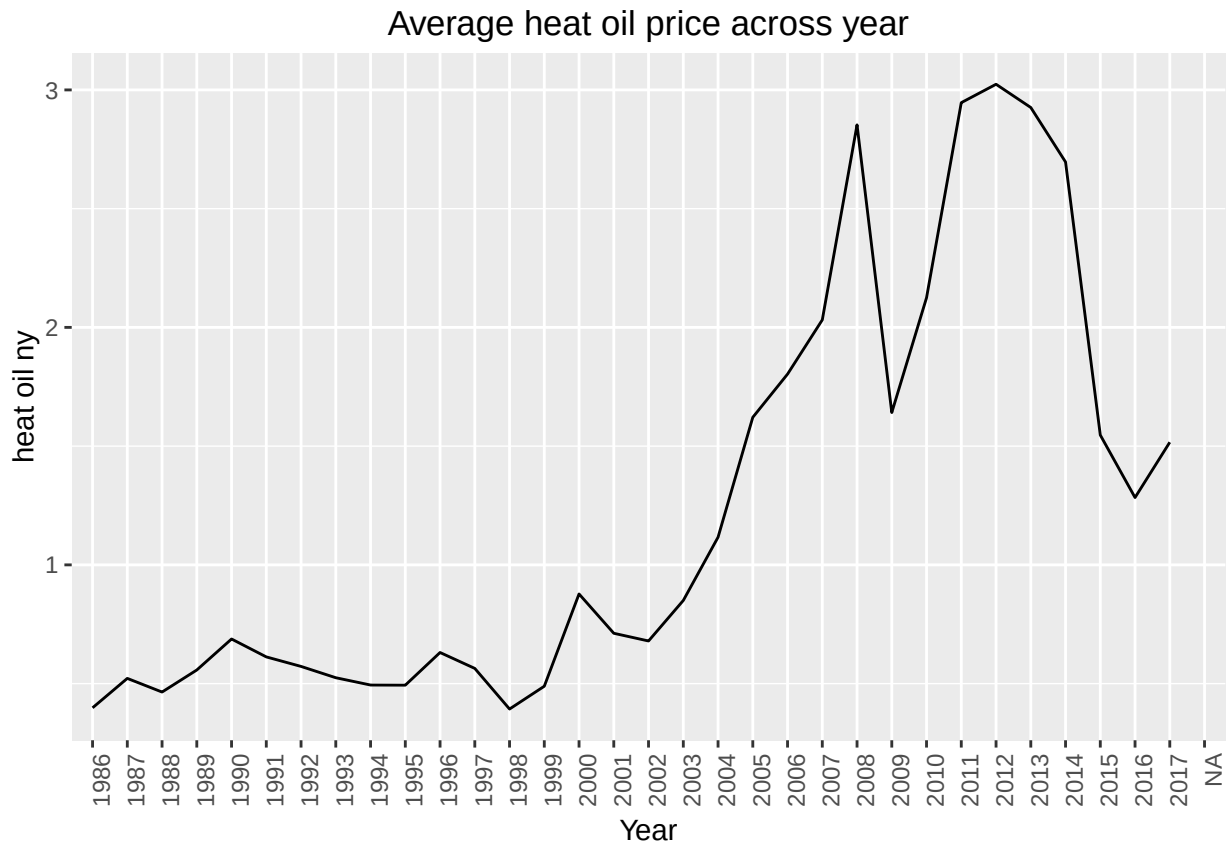
```
h1
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



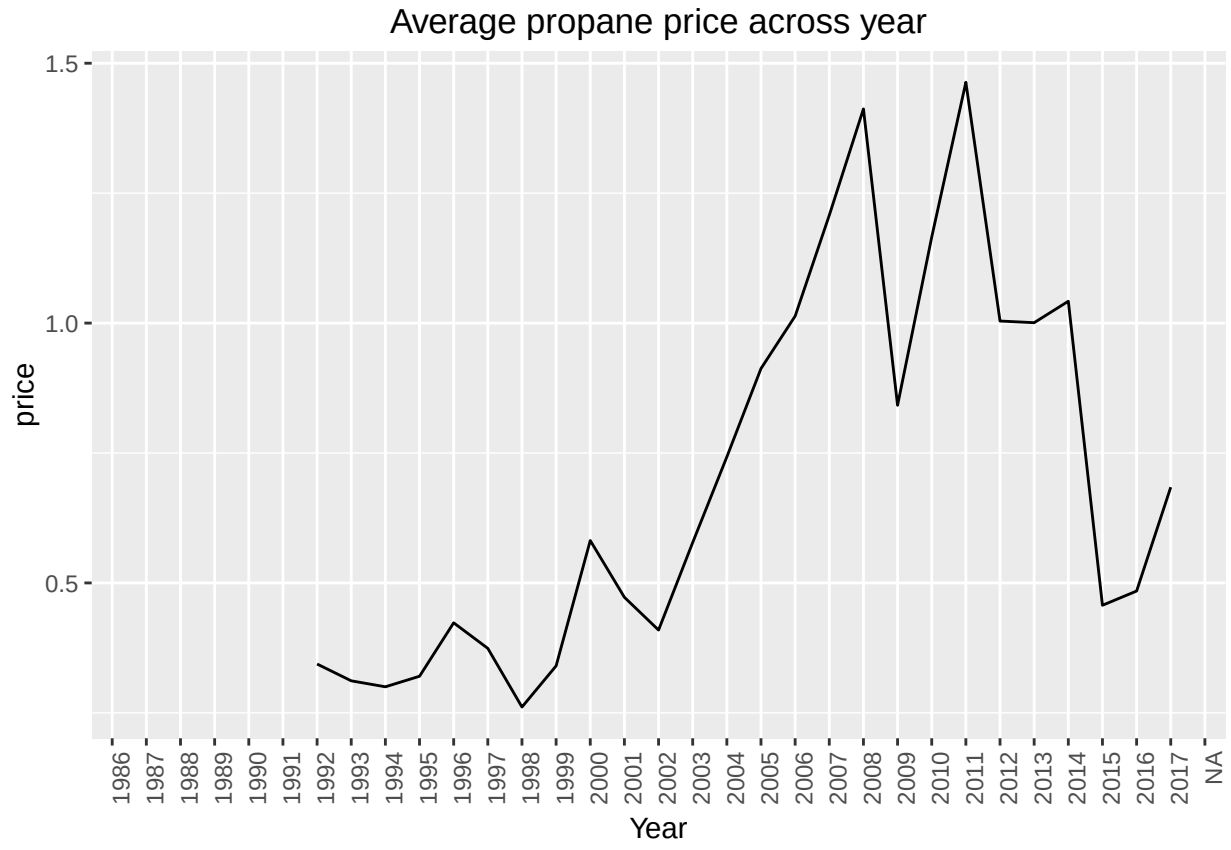

```
h2
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



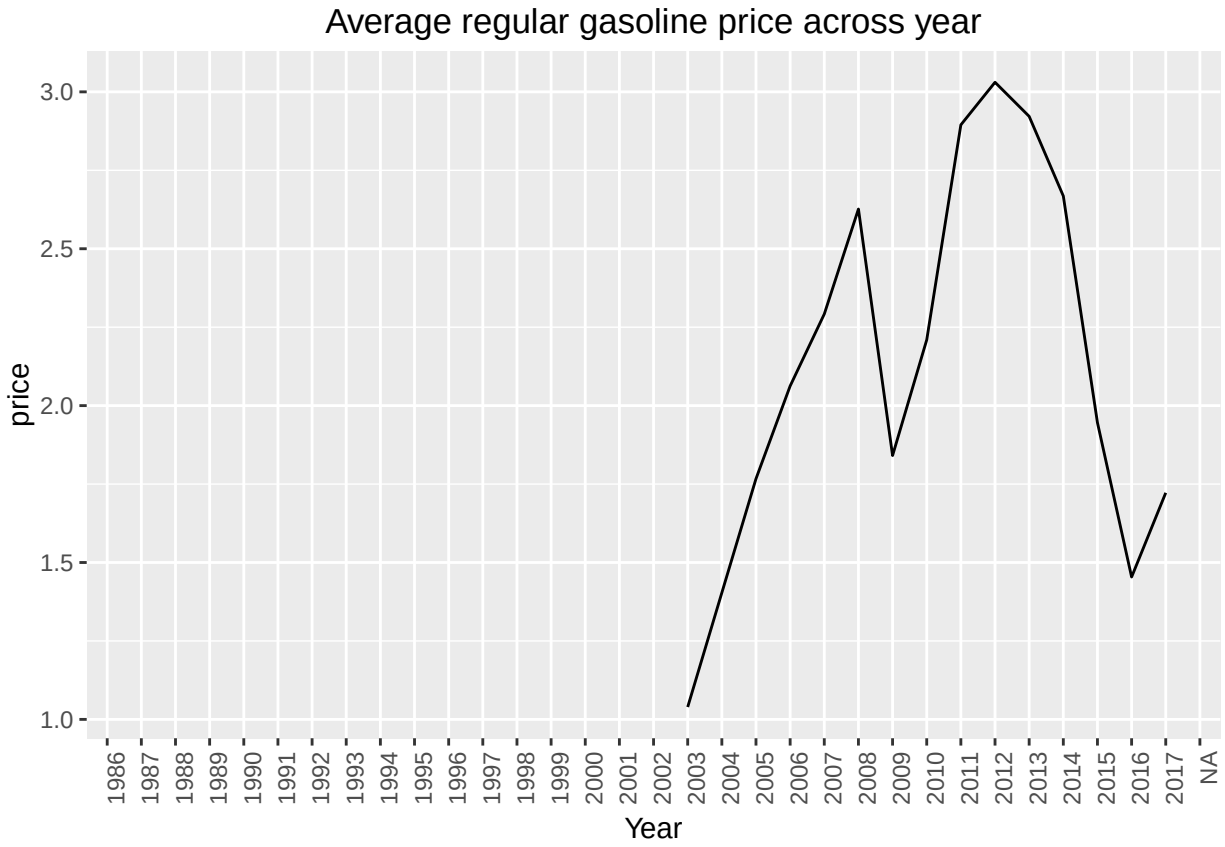

```
h3
```

```
## Warning: Removed 7 rows containing missing values (geom_path).
```



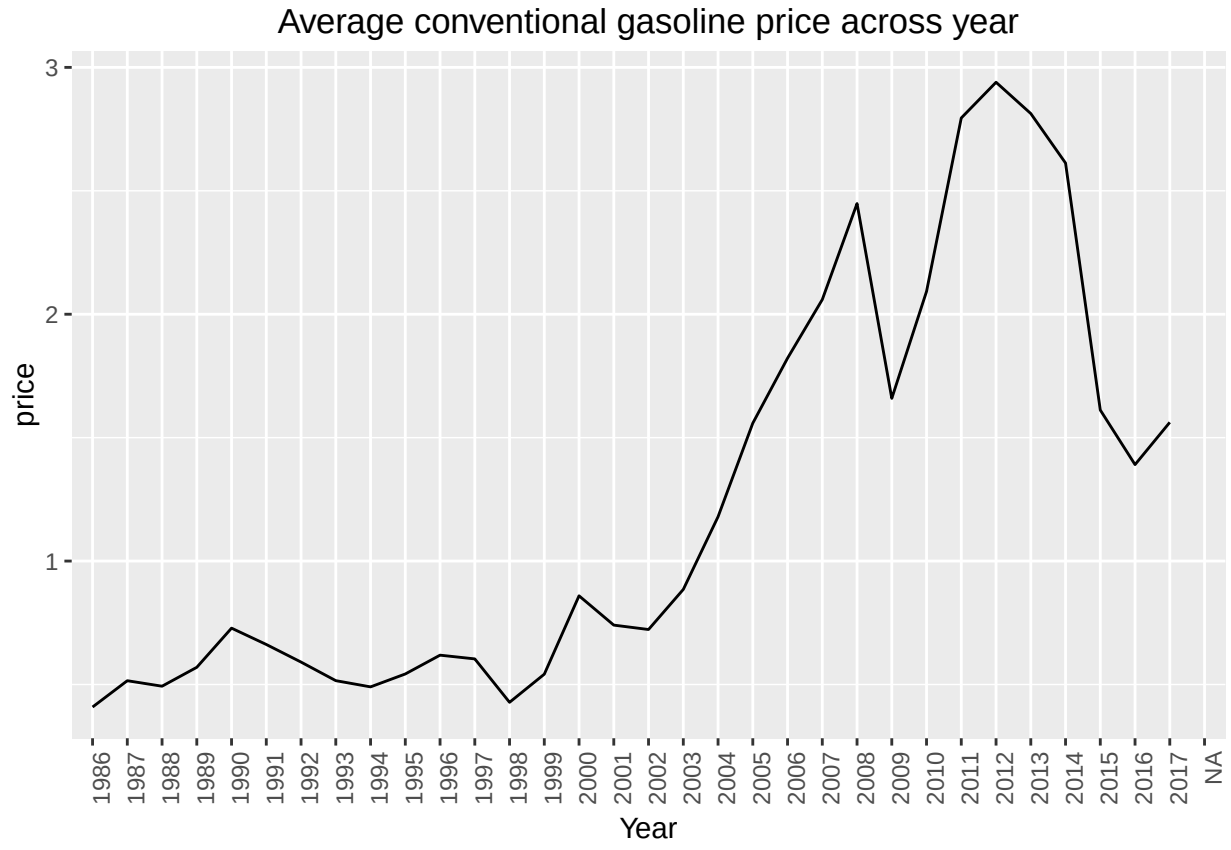

```
h4
```

```
## Warning: Removed 18 rows containing missing values (geom_path).
```



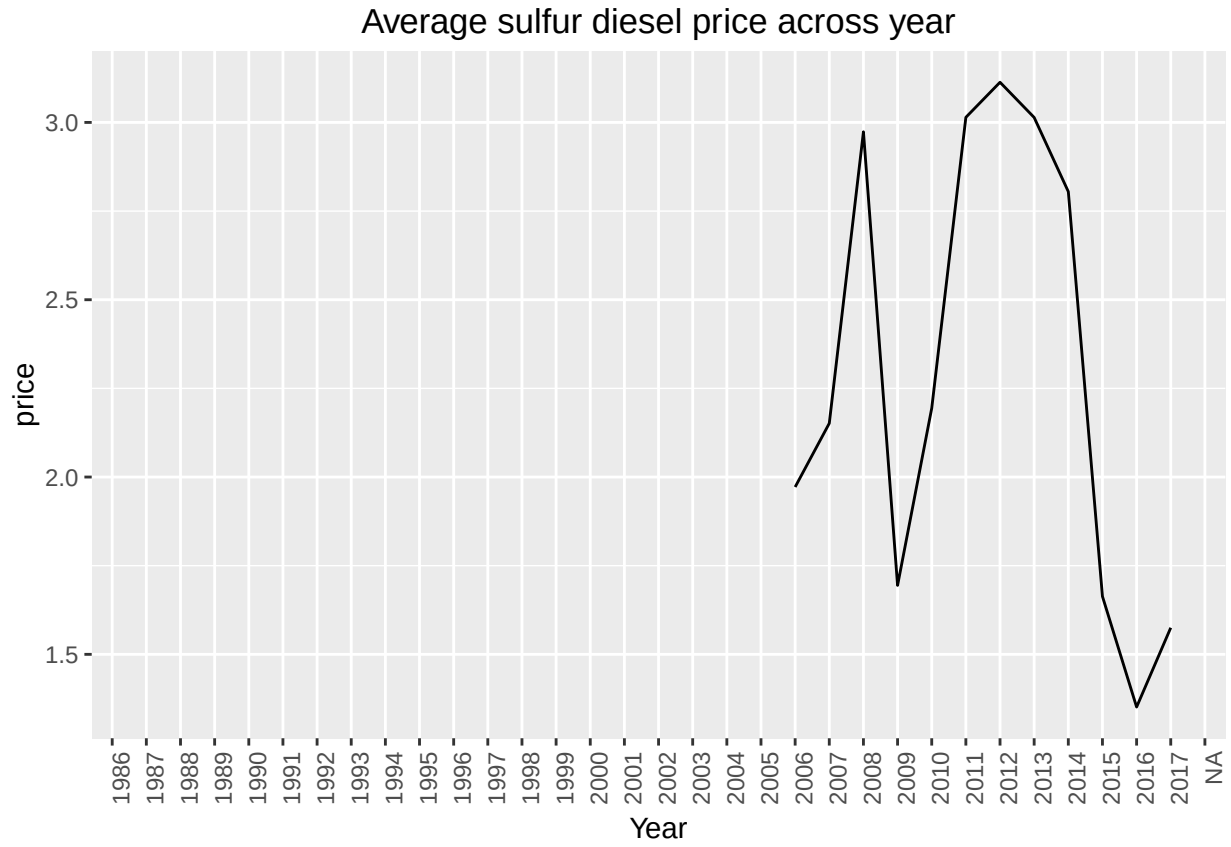

```
h5
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



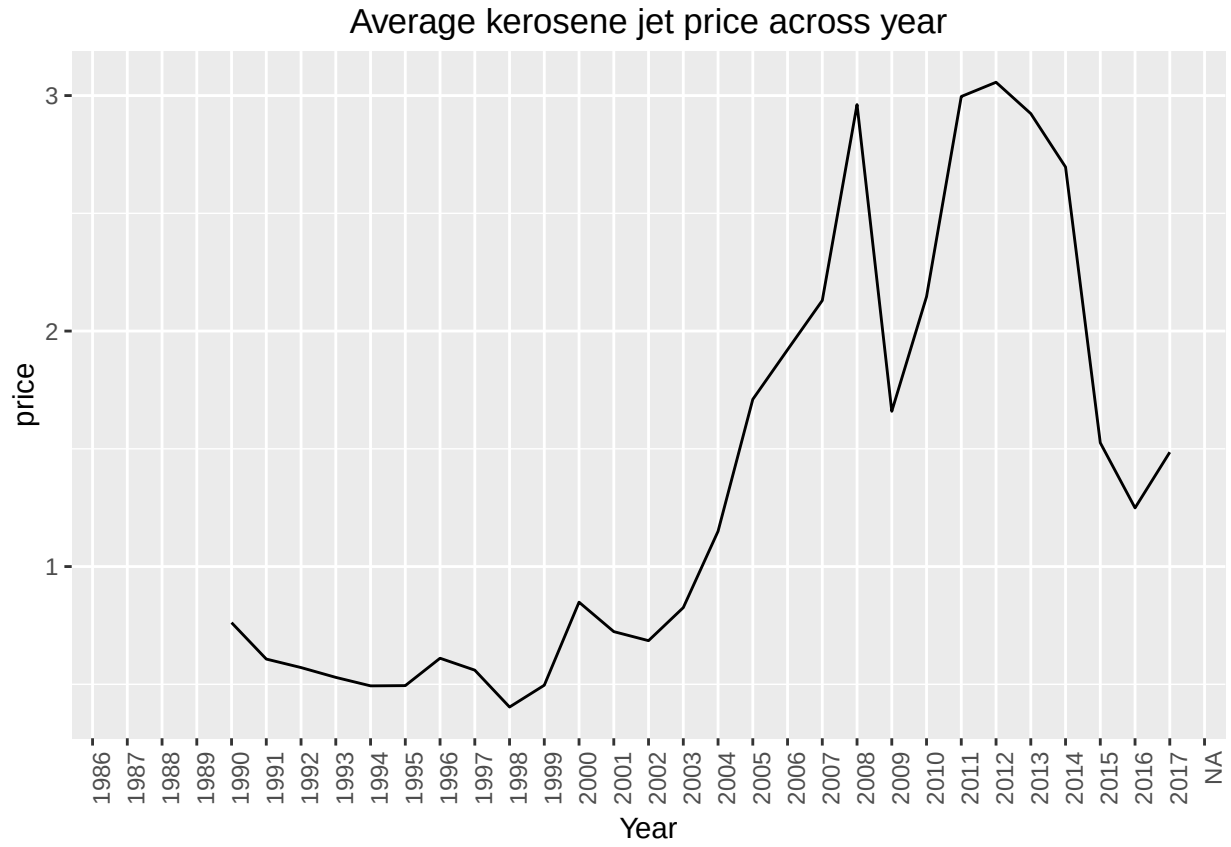

```
h6
```

```
## Warning: Removed 21 rows containing missing values (geom_path).
```




```
h7
```

```
## Warning: Removed 5 rows containing missing values (geom_path).
```



* Year wise trend has been similar for all the petroleum producrs with highers between 2008 and 2014

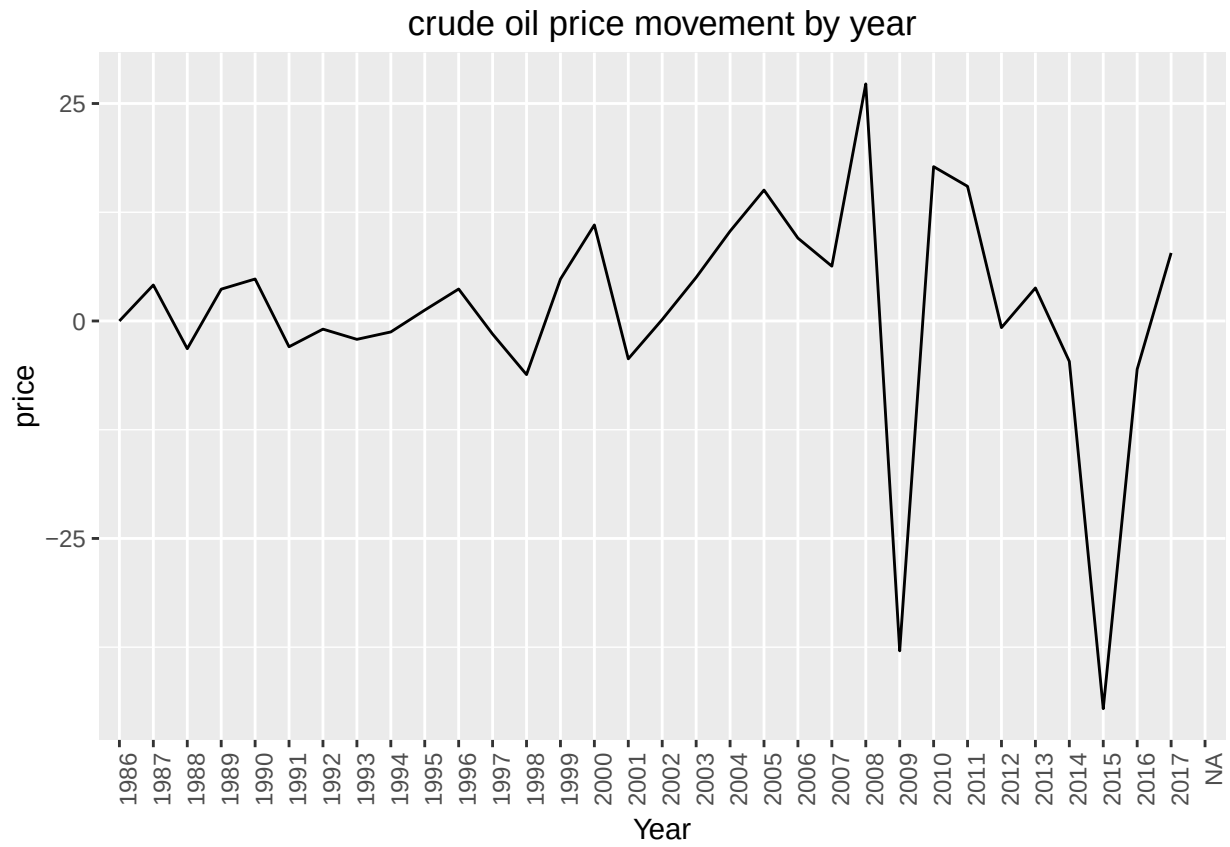
Price movement

```
petroleum_product_data_byYear_diff <- petroleum_product_data_byYear %>%
  mutate(crude_cushing_WTI = c(0, diff(crude_cushing_WTI, lag =1)),
            heat_oil_ny = c(0, diff(heat_oil_ny, lag =1)),
            propn_montBel = c(0, diff(propn_montBel, lag =1)),
            reg_gas_la = c(0, diff(reg_gas_la, lag =1)),
            conv_gaso_ny = c(0, diff(conv_gaso_ny, lag =1)),
            sulfr_dsl_ny = c(0, diff(sulfr_dsl_ny, lag =1)),
            kersn_jet_USgulf = c(0, diff(kersn_jet_USgulf, lag =1)))

d1 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = crude_cushing_WTI, group = 1 )) +
    labs(x = "Year", y = "price", title = "crude oil price movement by year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

d2 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = heat_oil_ny, group = 1 )) +
    labs(x = "Year", y = "price", title = "heat oil price movement by year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

d3 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = propn_montBel, group = 1 )) +
```

```
    labs(x = "Year", y = "price", title = "propane price movement by year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

d4 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = reg_gas_la, group = 1 )) +
    labs(x = "Year", y = "price", title = "regular gasoline price movement across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

d5 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = conv_gaso_ny, group = 1 )) +
    labs(x = "Year", y = "price", title = "conventional gasoline price movement across year") +   theme

d6 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = sulfr_dsl_ny, group = 1 )) +
    labs(x = "Year", y = "price", title = "sulfur diesel price movement across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

d7 <- ggplot(petroleum_product_data_byYear_diff) +
    geom_line(aes(x= factor(year), y = kersn_jet_USgulf, group = 1 )) +
    labs(x = "Year", y = "price", title = "kerosene jet price movement across year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

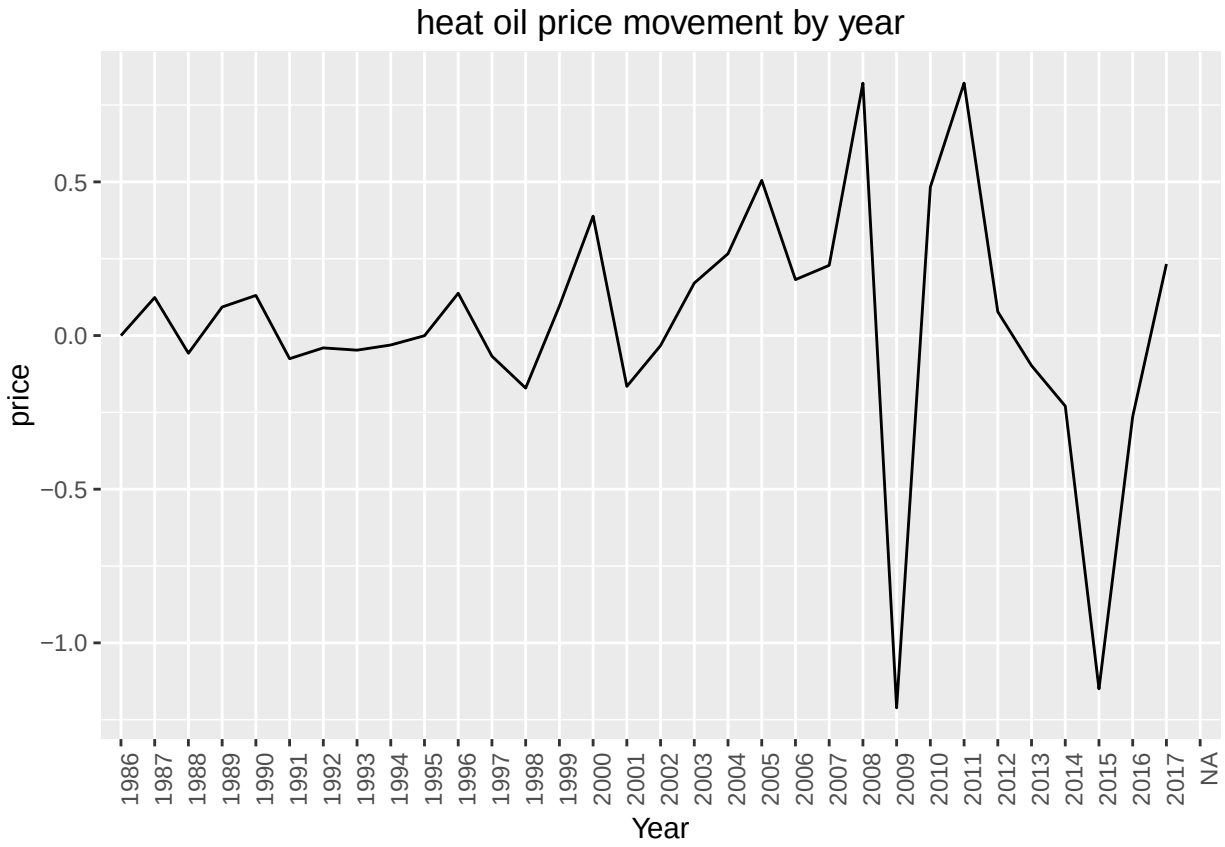
```
d1
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



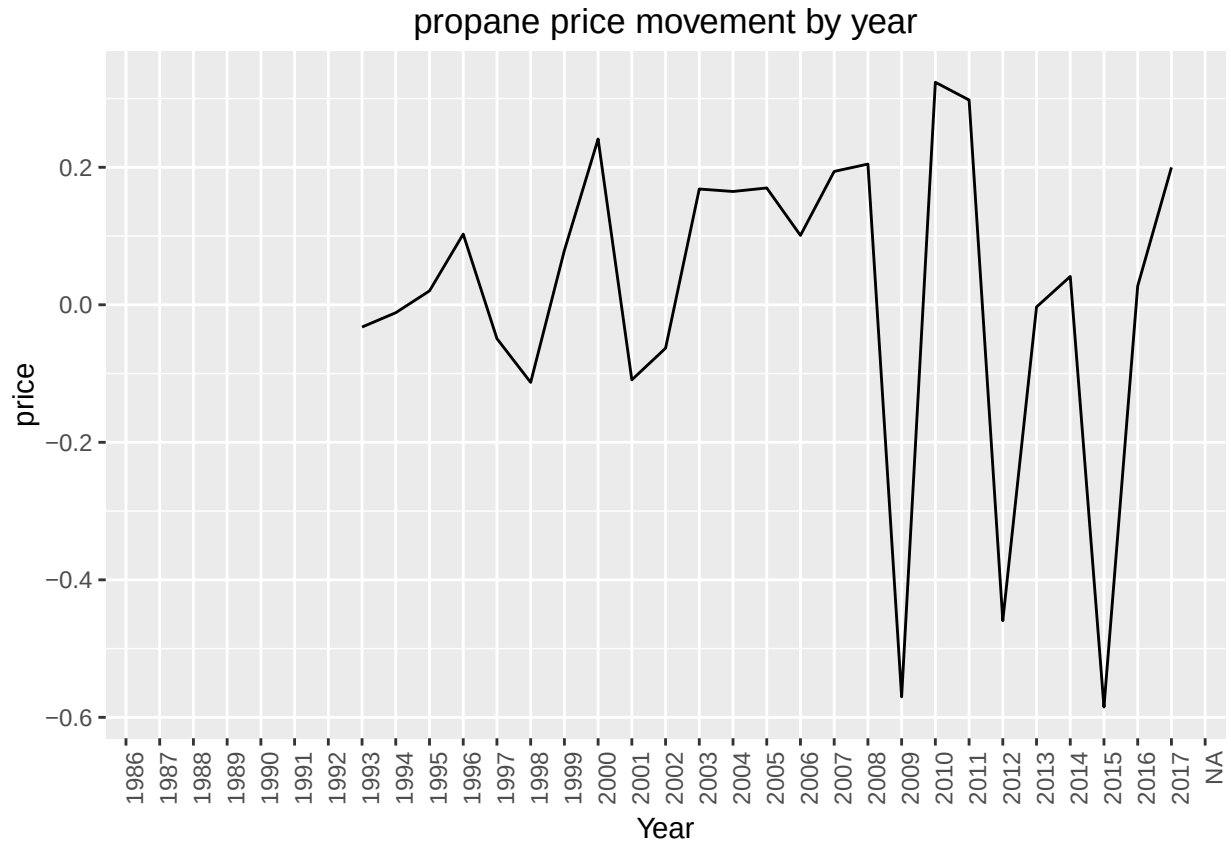
```
d2
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



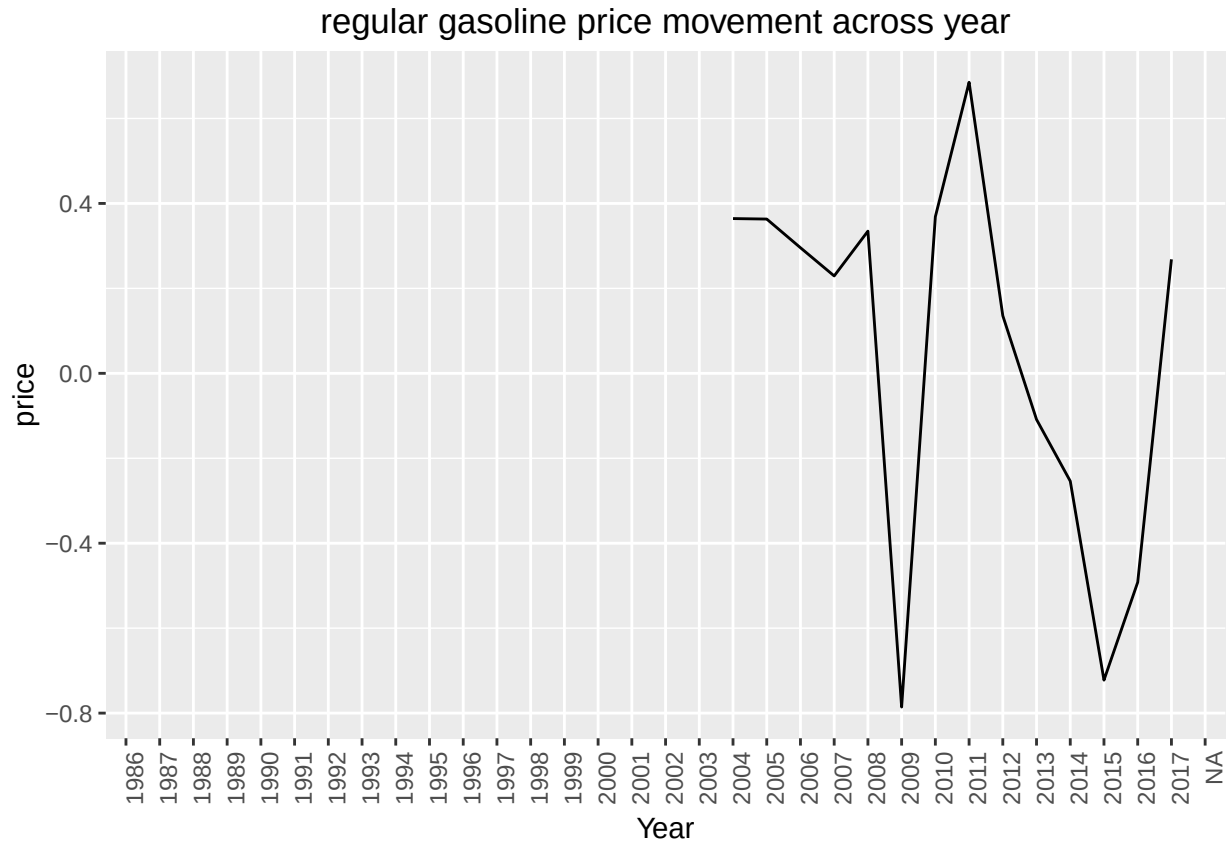

```
d3
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



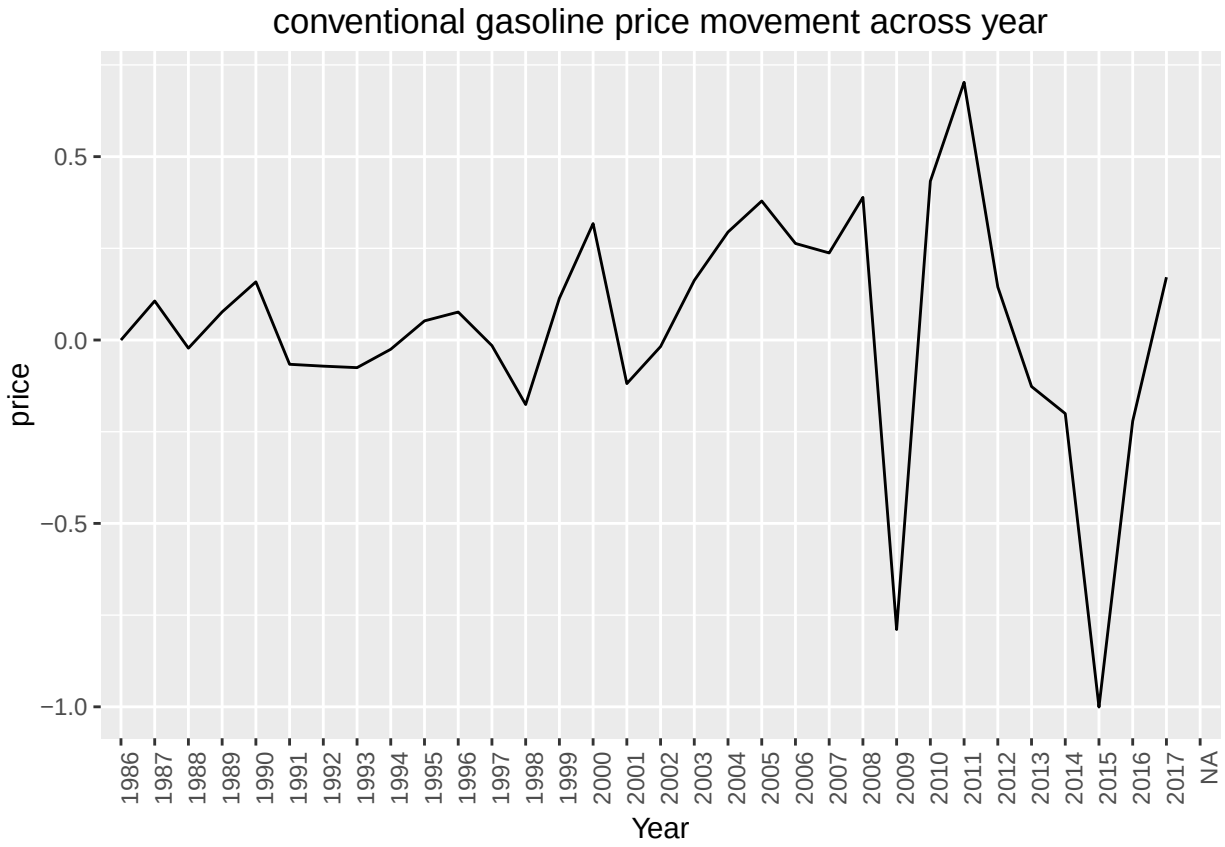

```
d4
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



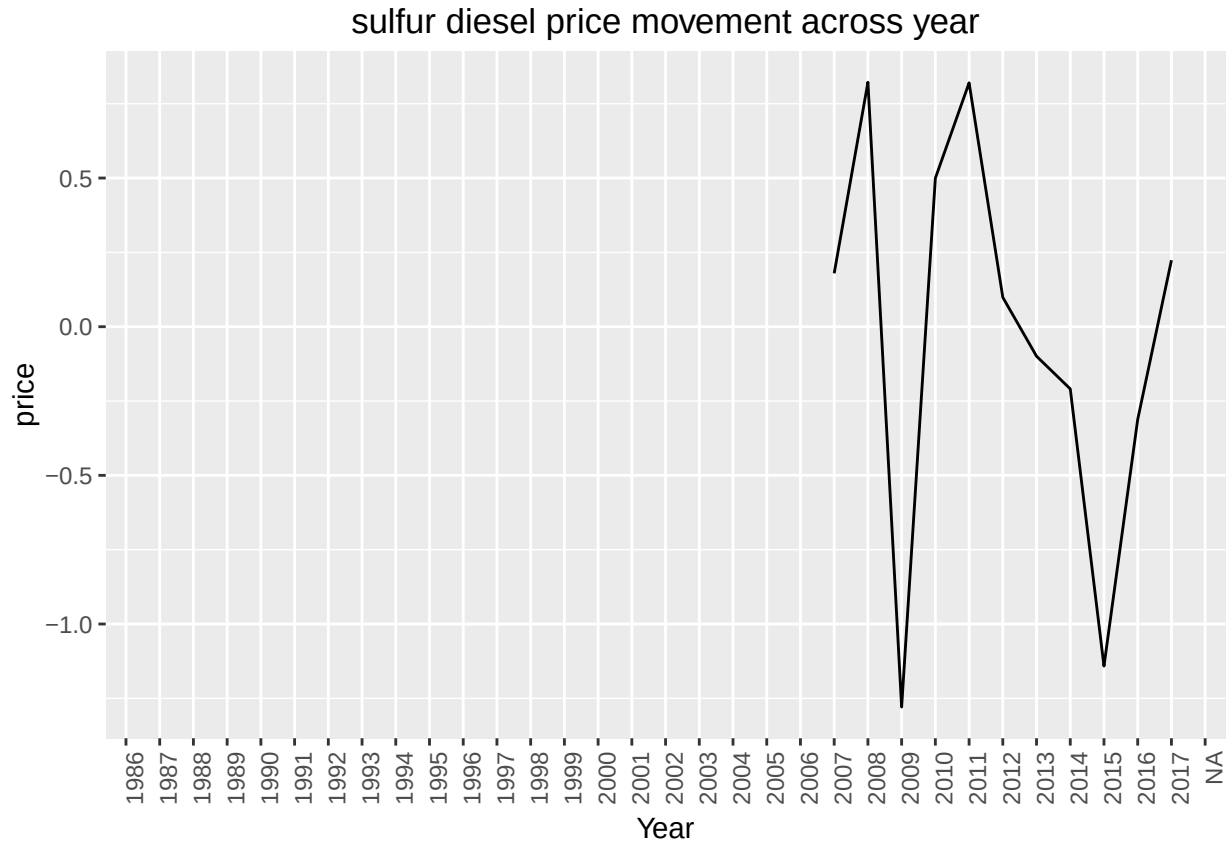

```
d5
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



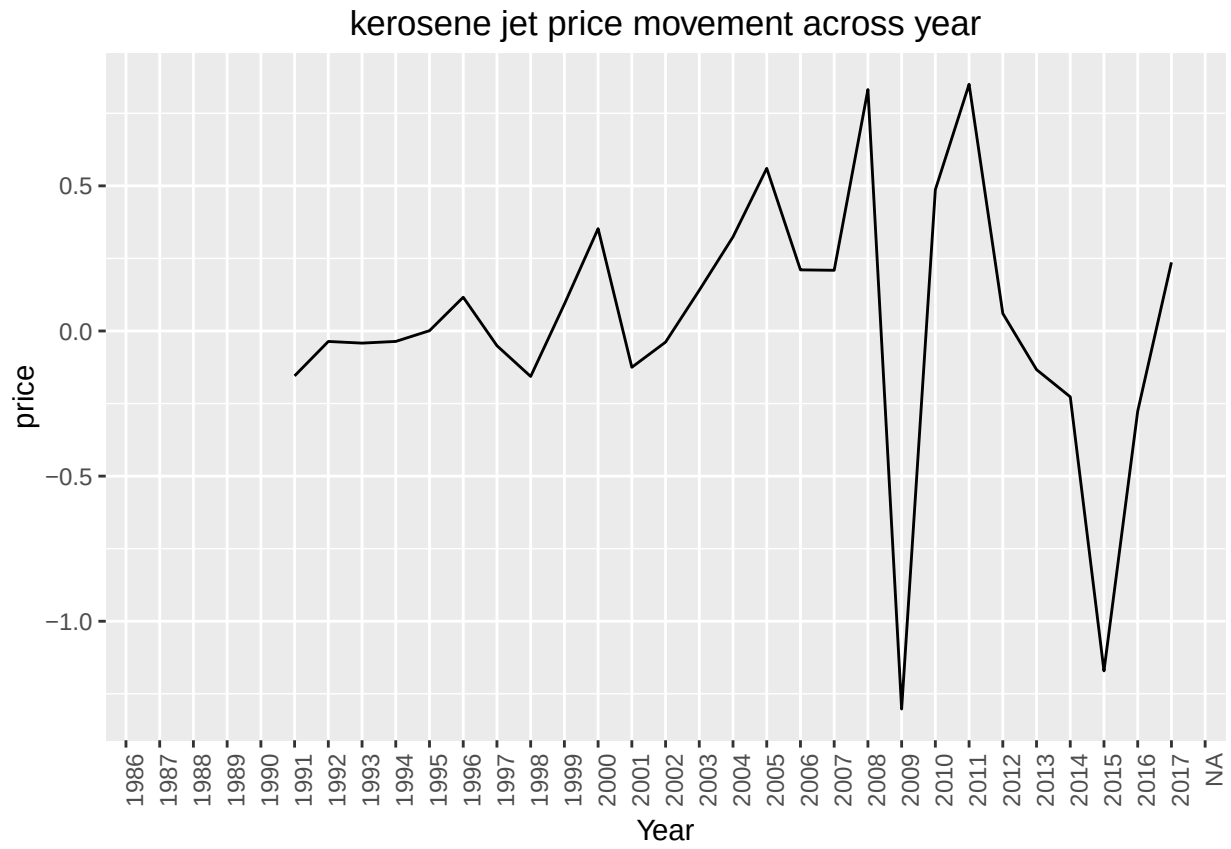

```
d6
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
d7
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



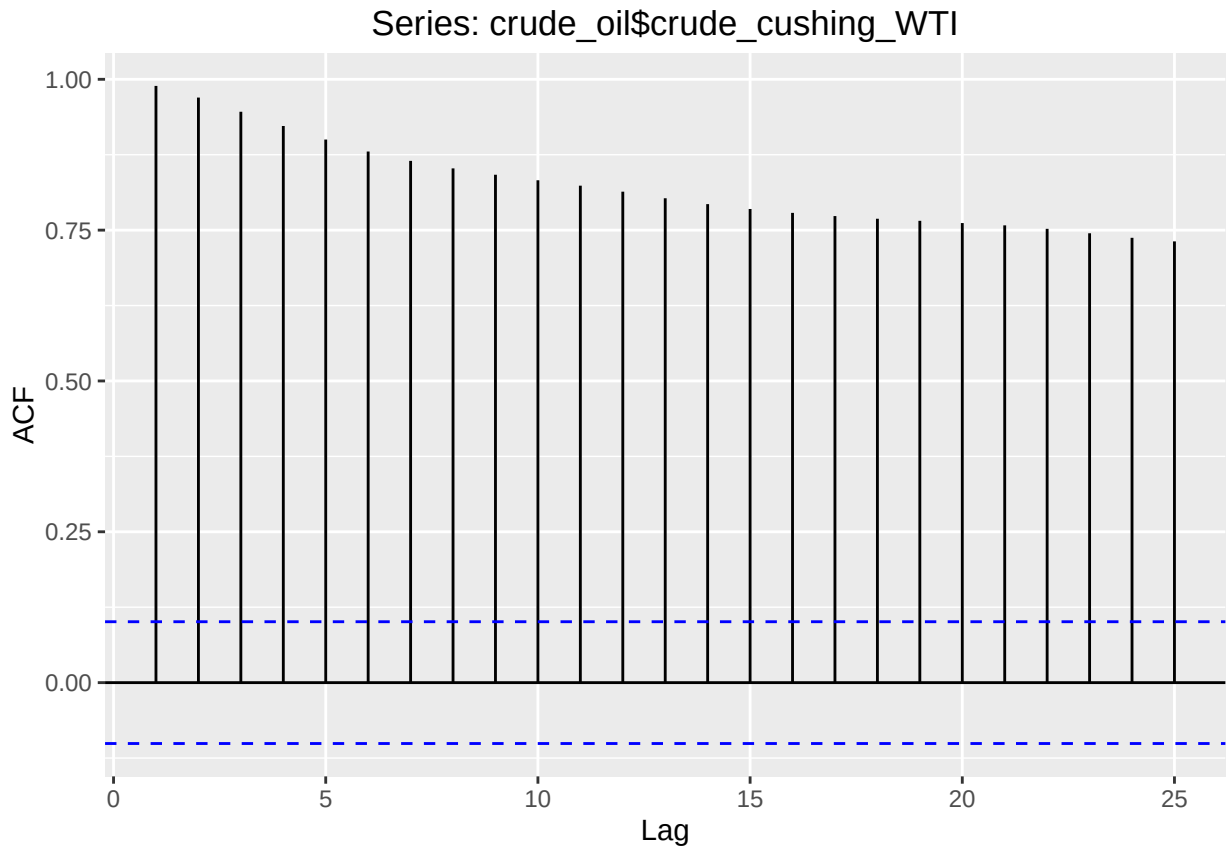
- Price movement have been volatile in recent years fom 2008 to 2014

Bonus questions

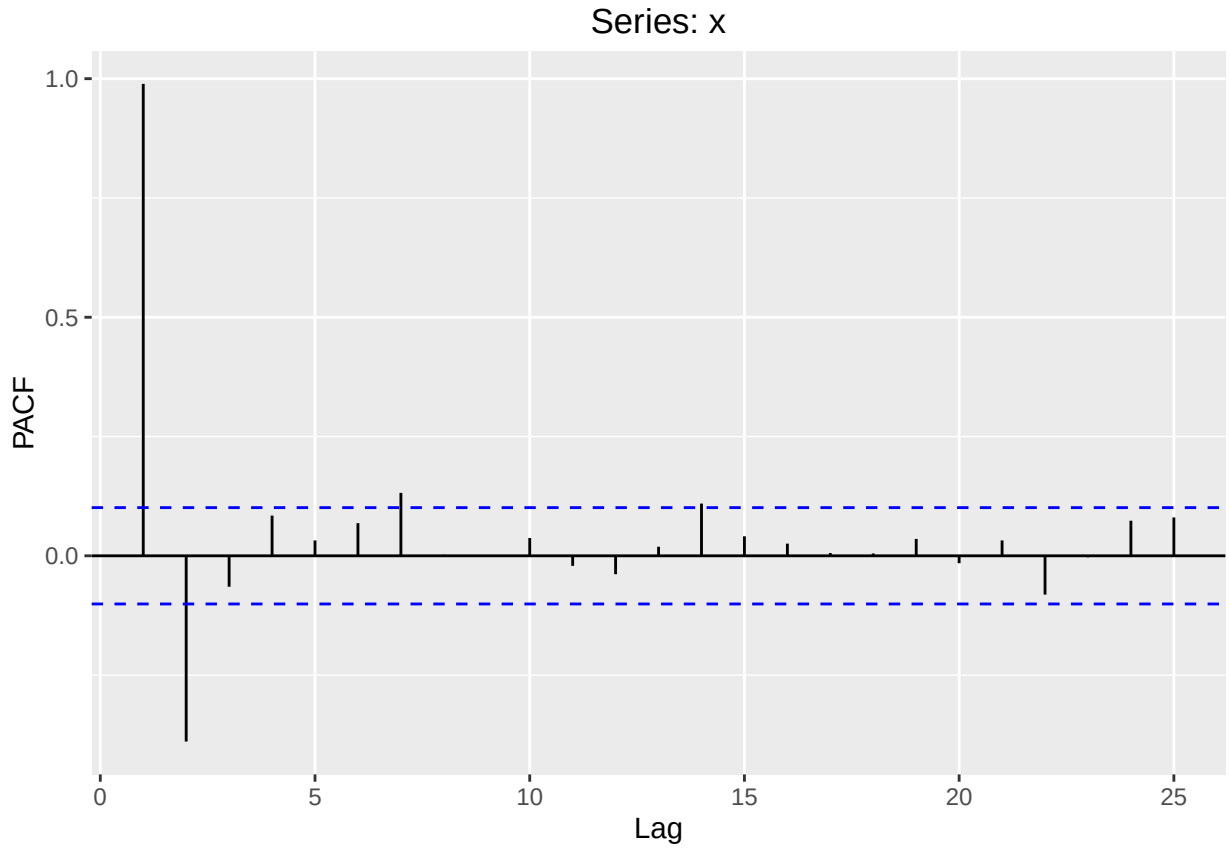
Question 5

Predict 6 months of Crude oil prices

```
crude_oil <- petroleum_product_data %>%
  select(crude_cushing_WTI) %>%
  filter(!is.na(crude_cushing_WTI))

ggAcf(crude_oil$crude_cushing_WTI)
```




```
ggPacf(crude_oil$crude_cushing_WTI)
```



check for stationarity

```
adf.test(crude_oil$crude_cushing_WTI, alternative = "stationary")
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  crude_oil$crude_cushing_WTI
## Dickey-Fuller = -2.2227, Lag order = 7, p-value = 0.483
## alternative hypothesis: stationary
```

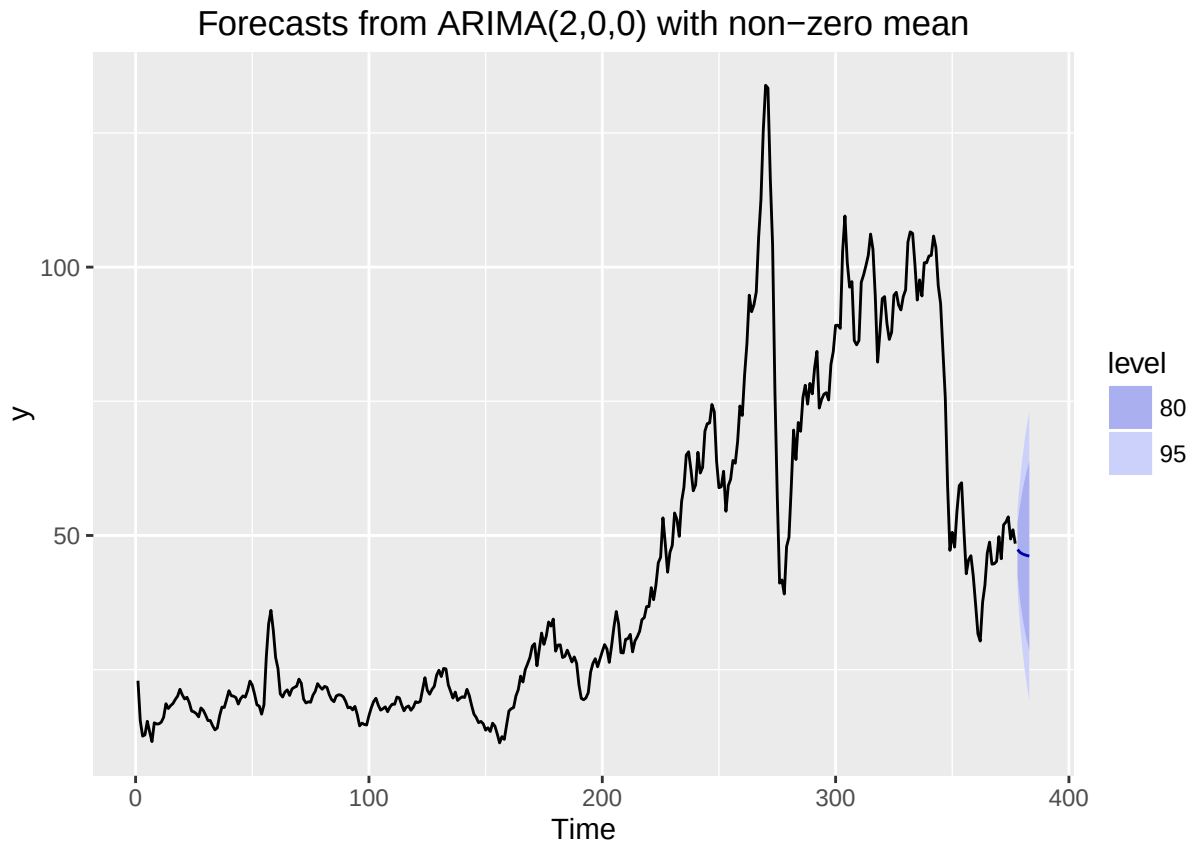
- The series is stationary
- From ACF and PACF plot, it is clear that this is AR2 process

```
crude_oil.model <- Arima(crude_oil$crude_cushing_WTI,order=c(2,0,0))
```

```
crude_oil.forecast <- forecast(crude_oil.model, h = 6)
crude_oil.forecast
```

```
##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 378       47.37248 42.29614 52.44883 39.60888 55.13608
## 379       46.85825 38.22544 55.49105 33.65552 60.06098
## 380       46.58412 35.07006 58.09817 28.97489 64.19334
## 381       46.40819 32.52960 60.28679 25.18271 67.63367
## 382       46.27344 30.41246 62.13441 22.01617 70.53070
## 383       46.15689 28.59899 63.71478 19.30440 73.00937
```

```
autoplot(crude_oil.forecast)
```



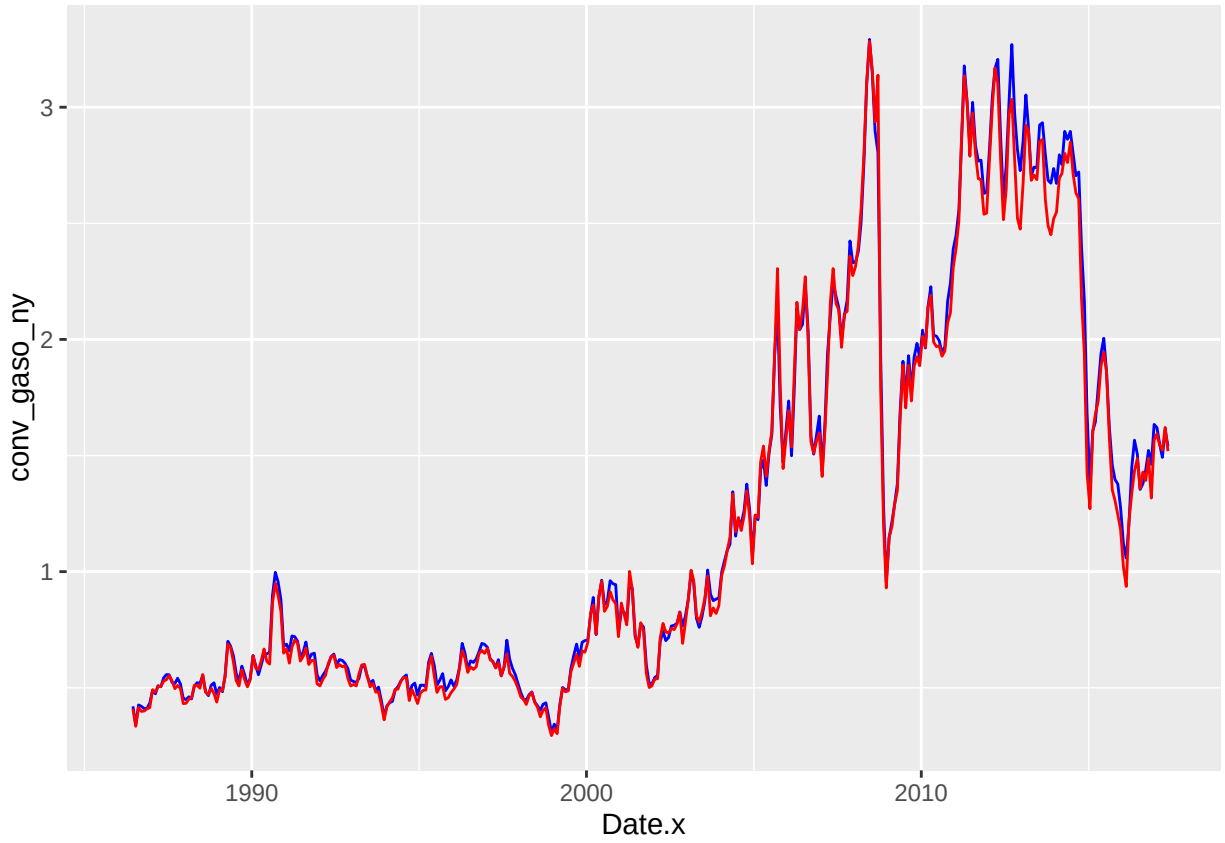
- According to the forecast, for the next 6 months crude oil prices will vary between 45 and 48

Question 6

```
ggplot(petroleum_product_data) +
  geom_line(aes(x = Date.x, y = conv_gaso_ny), color = "blue",show.legend = ) +
    geom_line(aes(x = Date.x, y = conv_gaso_USgulf), color = "red")
```

```
## Warning: Removed 6 rows containing missing values (geom_path).
```

```
## Warning: Removed 6 rows containing missing values (geom_path).
```



find correlation

```
ny <- petroleum_product_data$conv_gaso_ny
ny[is.na(ny)] <- 0
gulf <- petroleum_product_data$conv_gaso_USgulf
gulf[is.na(gulf)] <- 0
cor(ny,gulf)
```

```
## [1] 0.9980751
```

- Correlation is very high between conventional gasoline prices between Ny and us gulf

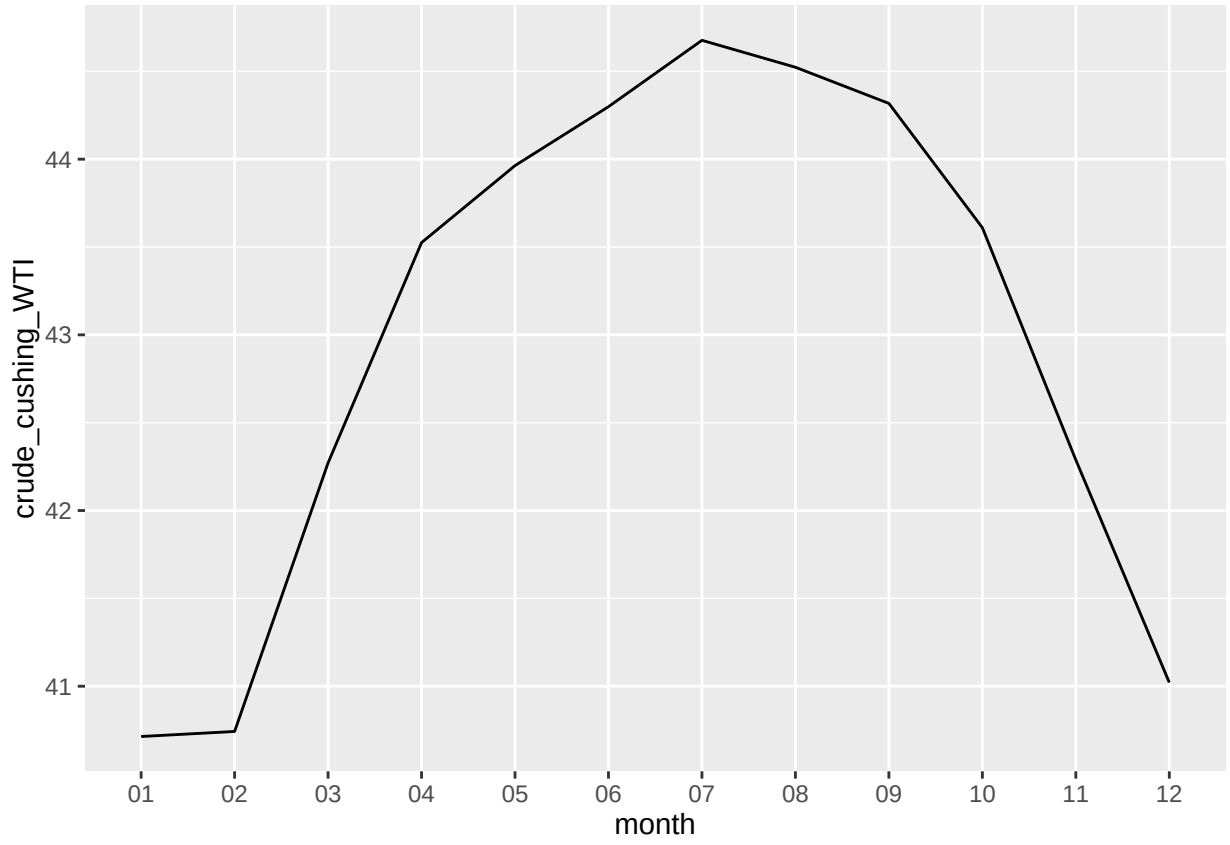
Question 7

```
petroleum_product_data_byMonth <- petroleum_product_data %>%
  group_by(month) %>%
  summarise(crude_cushing_WTI = mean(crude_cushing_WTI, na.rm = TRUE),
            heat_oil_ny = mean(heat_oil_ny, na.rm = TRUE),
            propn_montBel =mean(propn_montBel, na.rm = TRUE),
            reg_gas_la = mean(reg_gas_la, na.rm = TRUE),
            conv_gaso_ny = mean(conv_gaso_ny, na.rm = TRUE),
            sulfr_dsl_ny = mean(sulfr_dsl_ny, na.rm = TRUE),
            kersn_jet_USgulf = mean(kersn_jet_USgulf, na.rm = TRUE))
```

```
m1 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= crude_cushing_WTI, group = 1))
m2 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= heat_oil_ny, group = 1))
m3 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= propn_montBel, group = 1))
m4 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= reg_gas_la, group = 1))
m5 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= conv_gaso_ny, group = 1))
m6<- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= sulfr_dsl_ny, group = 1))
m7 <- ggplot(petroleum_product_data_byMonth) +
  geom_line(aes(x= month, y= kersn_jet_USgulf, group = 1))
```

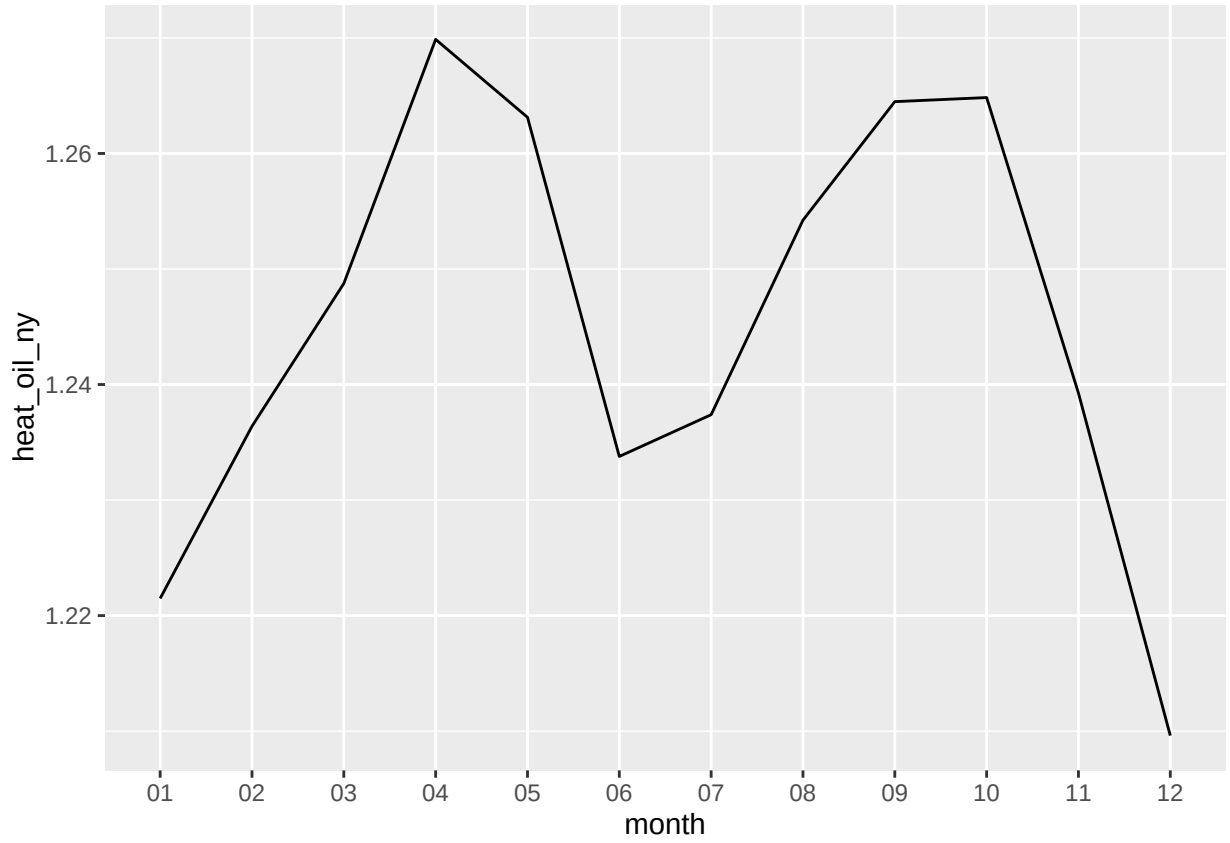
```
m1
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



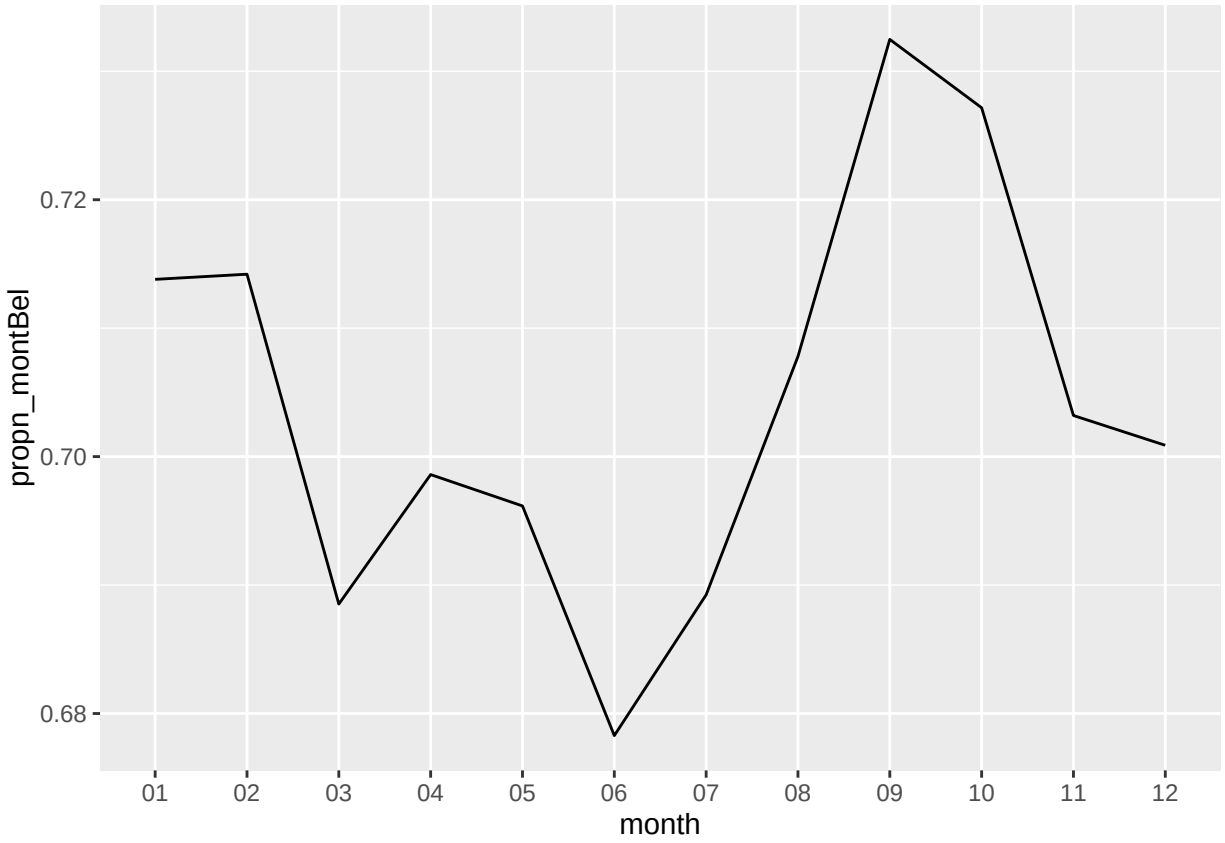

```
m2
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



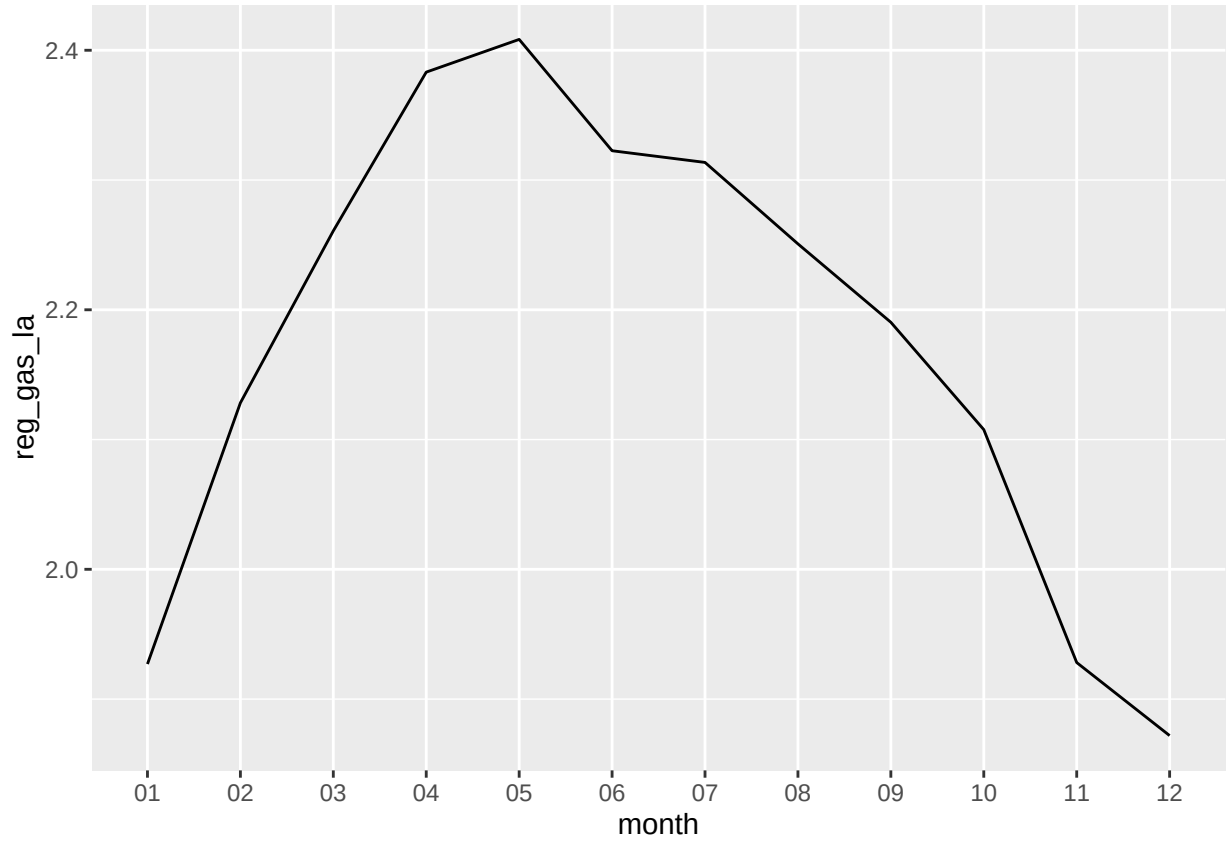

```
m3
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



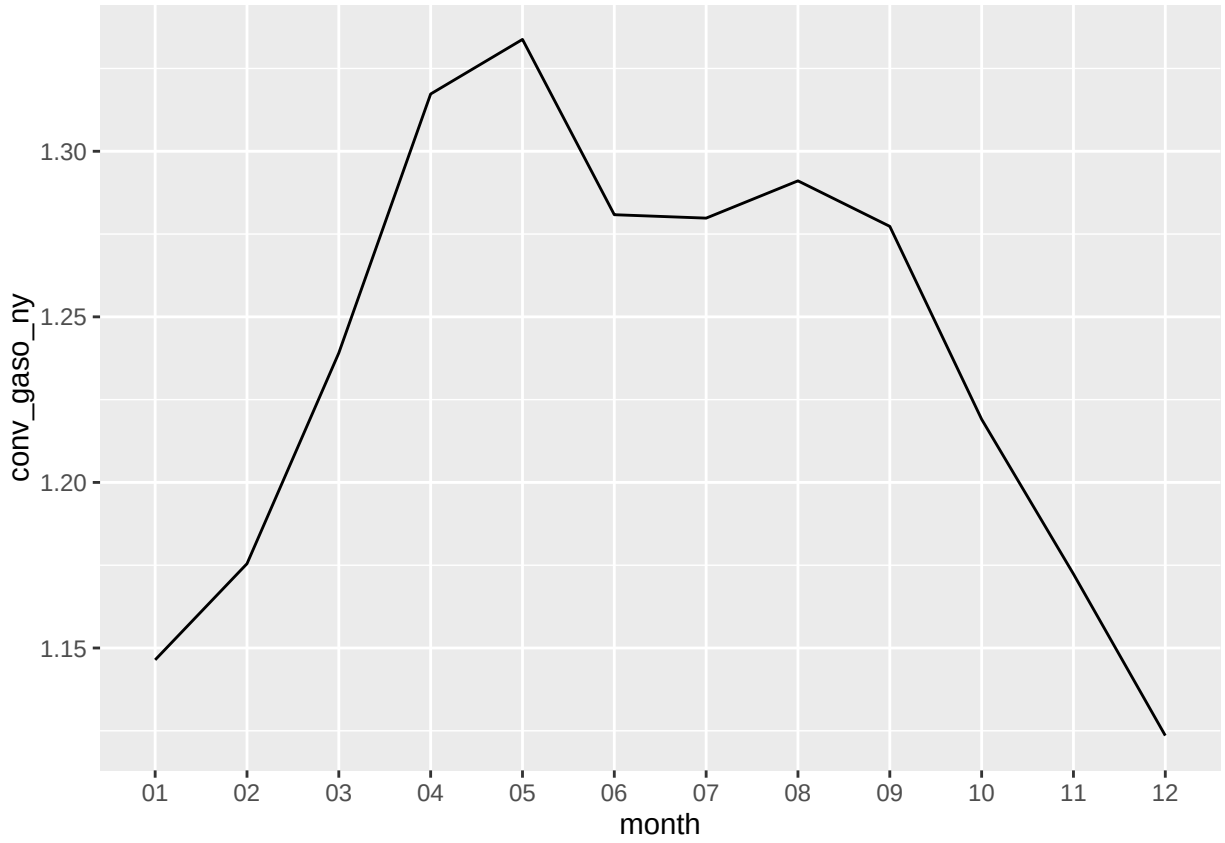

```
m4
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



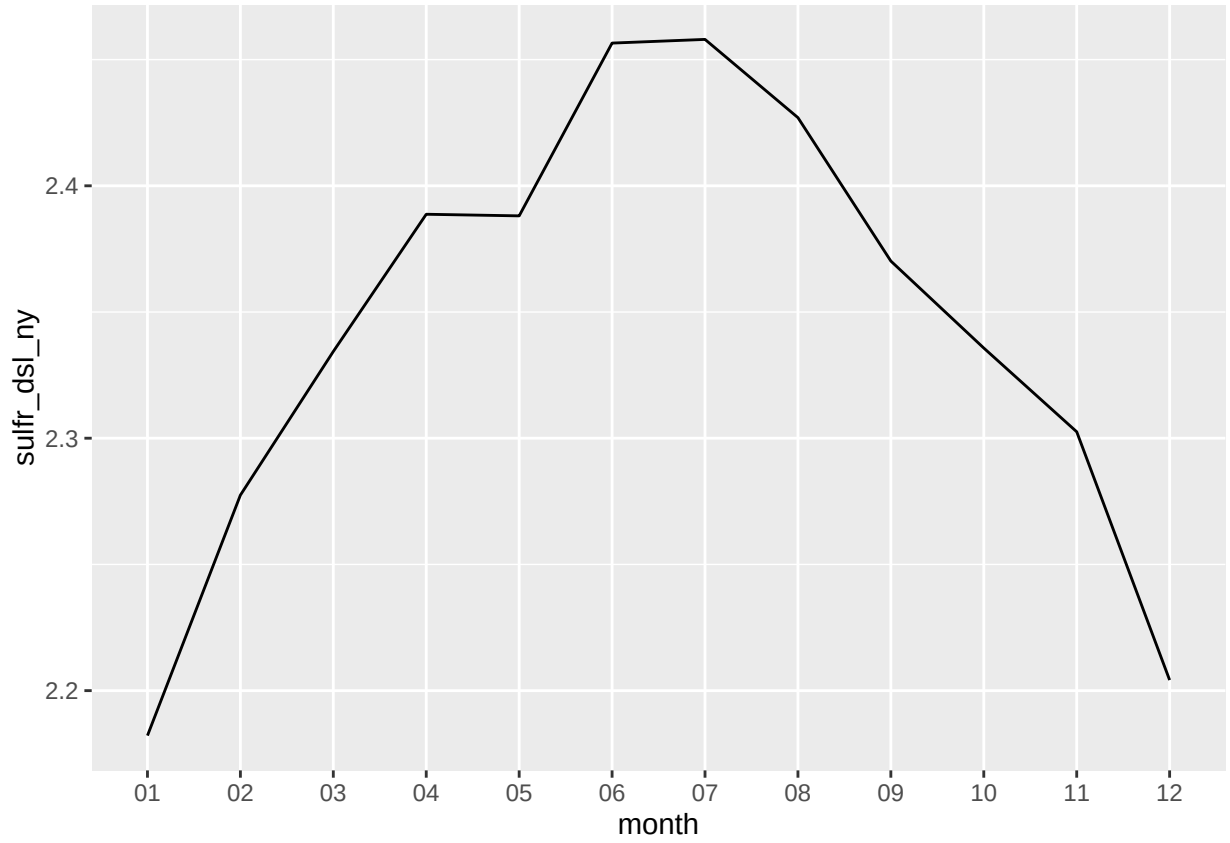

```
m5
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



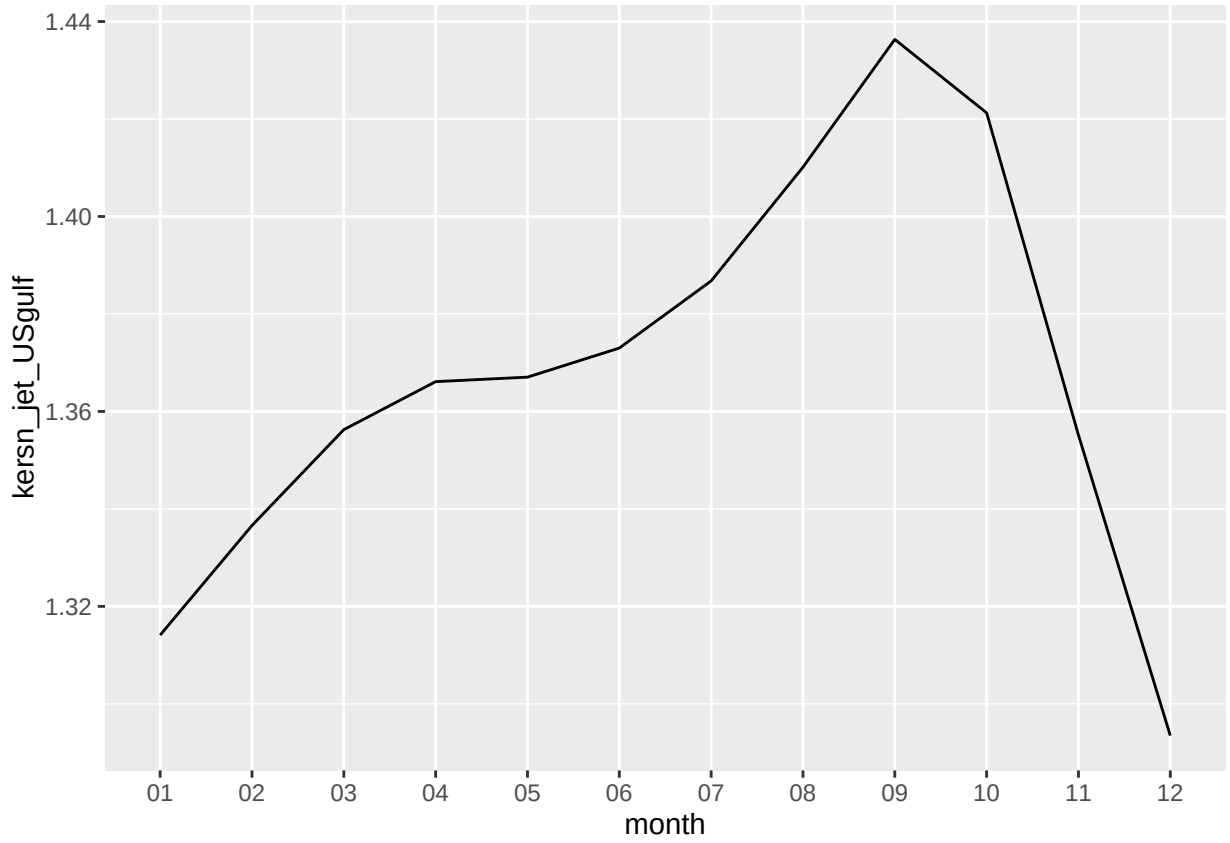

```
m6
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
m7
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



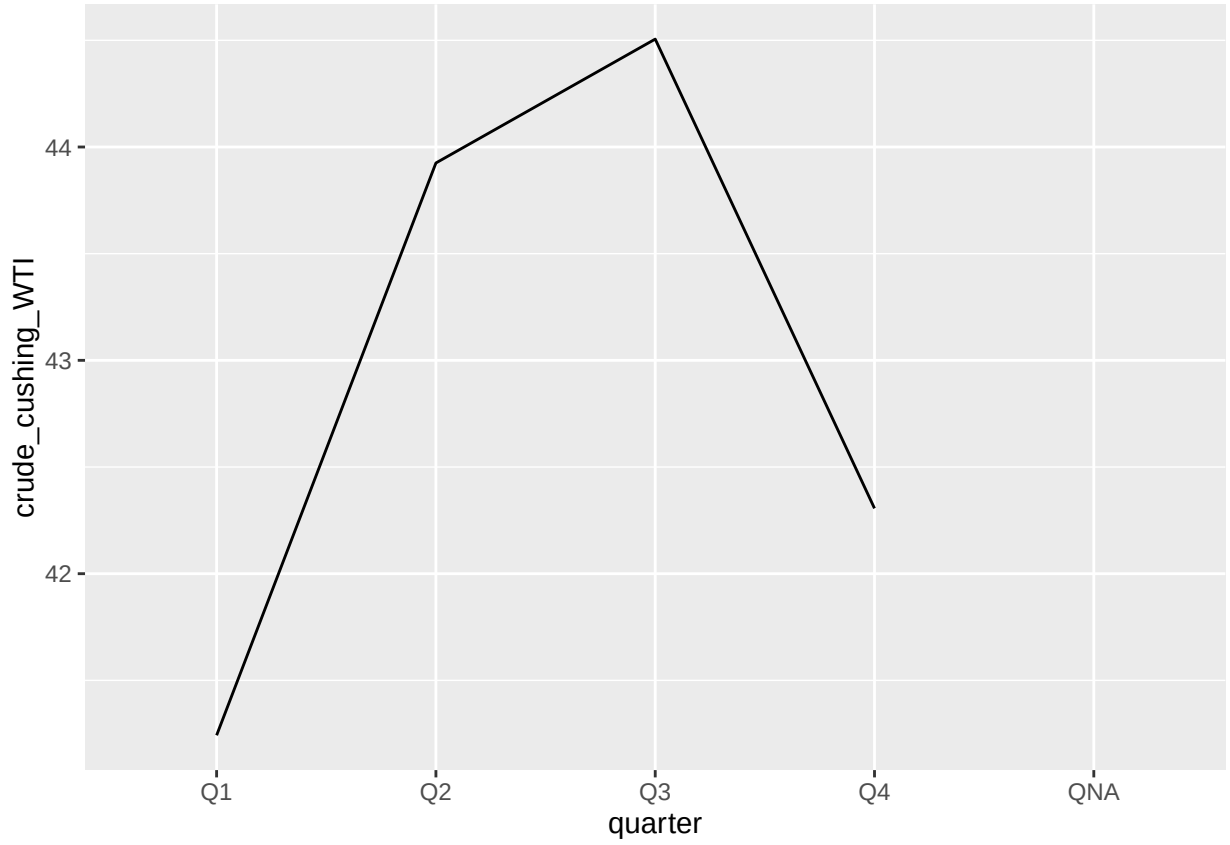
- From all the 7 plots above for different petroleum products, it is clear that prices in December and January are generally lower than other months, Except in the case of propane prices, where is is lowest in the month of June

```
petroleum_product_data_byQtr <- petroleum_product_data %>%
  mutate (quarter = quarters(as.Date(Date.x))) %>%
  group_by(quarter) %>%
  summarise(crude_cushing_WTI = mean(crude_cushing_WTI, na.rm = TRUE),
            heat_oil_ny = mean(heat_oil_ny, na.rm = TRUE),
            propn_montBel =mean(propn_montBel, na.rm = TRUE),
            reg_gas_la = mean(reg_gas_la, na.rm = TRUE),
            conv_gaso_ny = mean(conv_gaso_ny, na.rm = TRUE),
            sulfr_dsl_ny = mean(sulfr_dsl_ny, na.rm = TRUE),
            kersn_jet_USgulf = mean(kersn_jet_USgulf, na.rm = TRUE))


q1 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= crude_cushing_WTI, group = 1))
q2 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= heat_oil_ny, group = 1))
q3 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= propn_montBel, group = 1))
q4 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= reg_gas_la, group = 1))
q5 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= conv_gaso_ny, group = 1))
q6<- ggplot(petroleum_product_data_byQtr) +
```

```
  geom_line(aes(x= quarter, y= sulfr_dsl_ny, group = 1))
q7 <- ggplot(petroleum_product_data_byQtr) +
  geom_line(aes(x= quarter, y= kersn_jet_USgulf, group = 1))
```

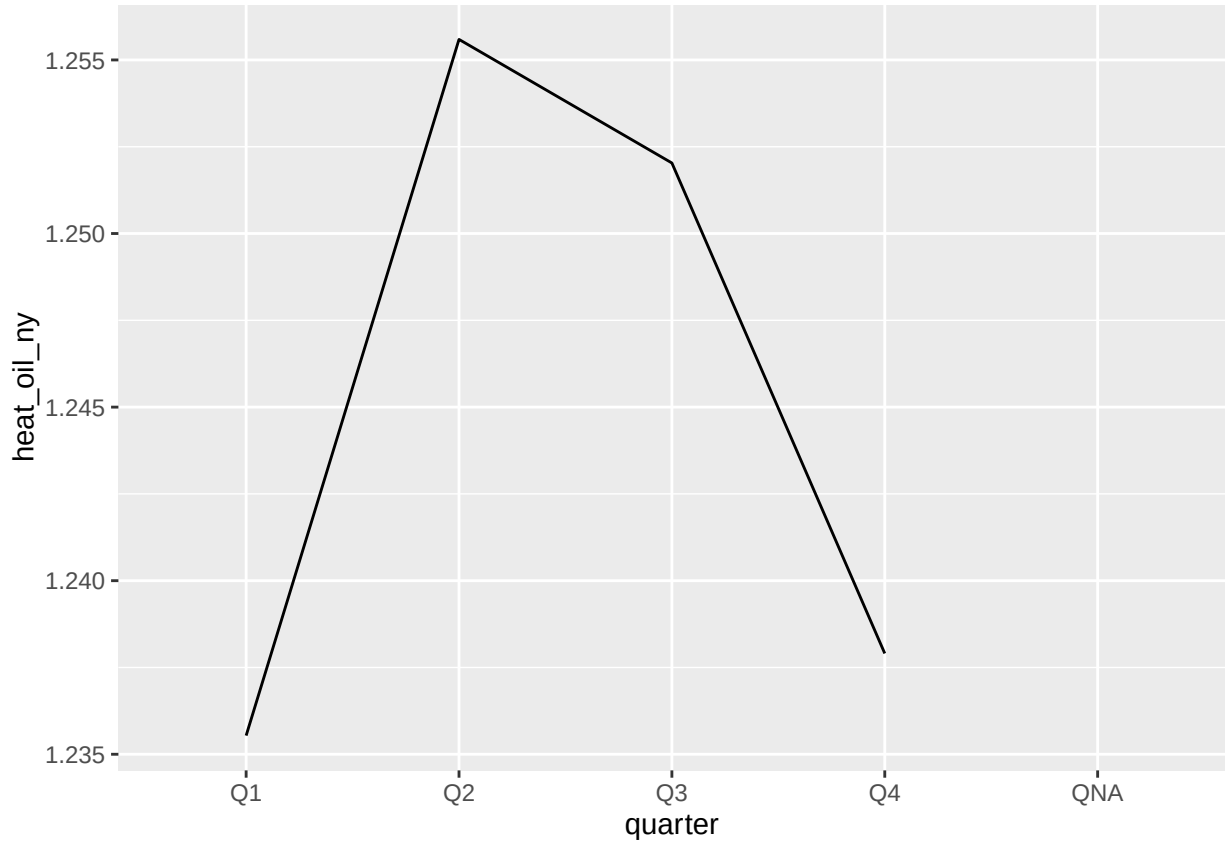
```
q1
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



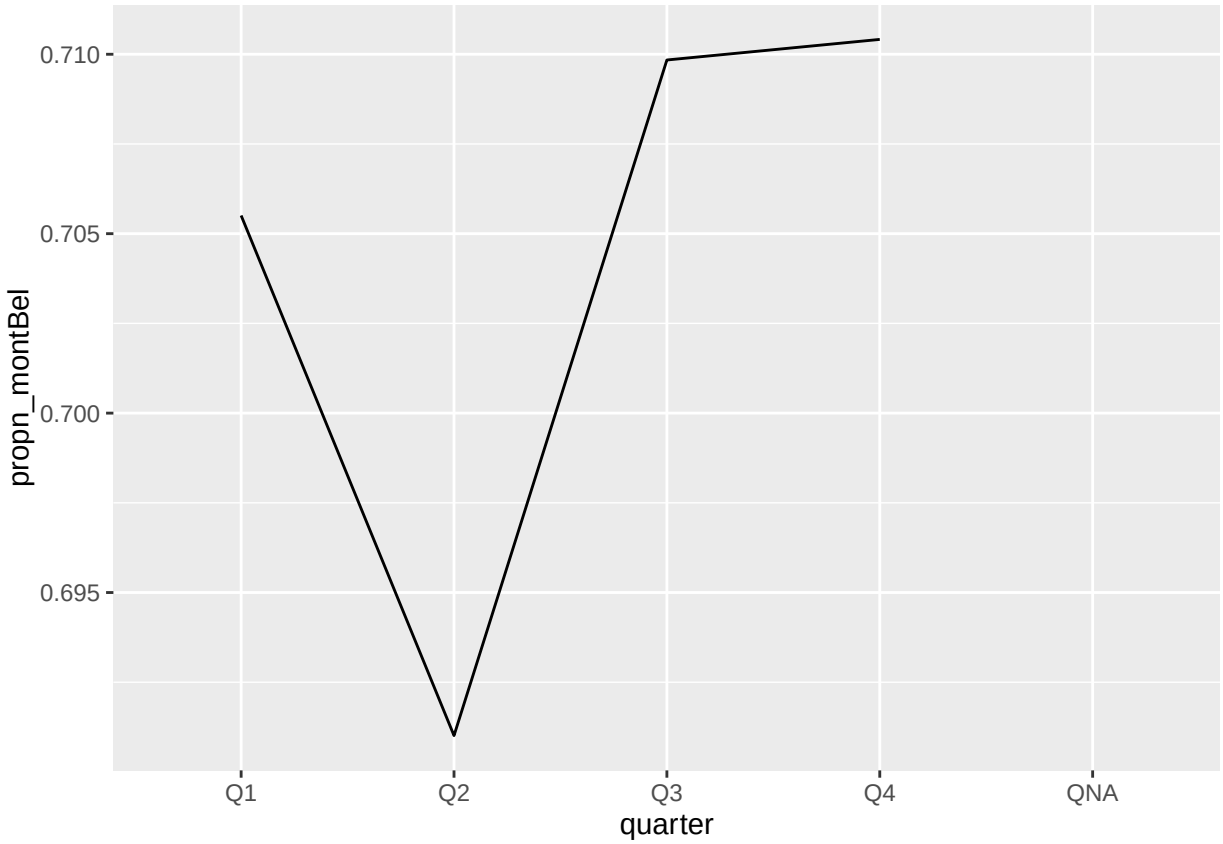

```
q2
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



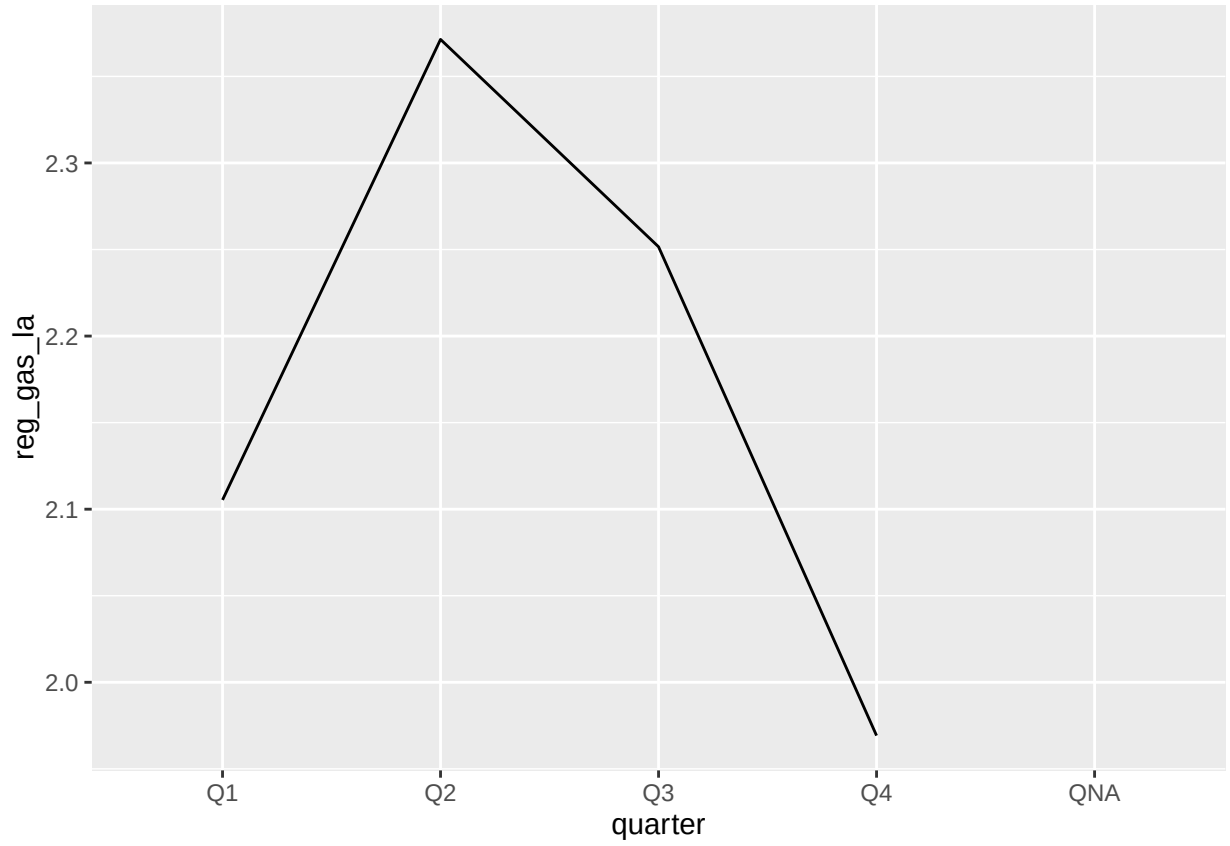

```
q3
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



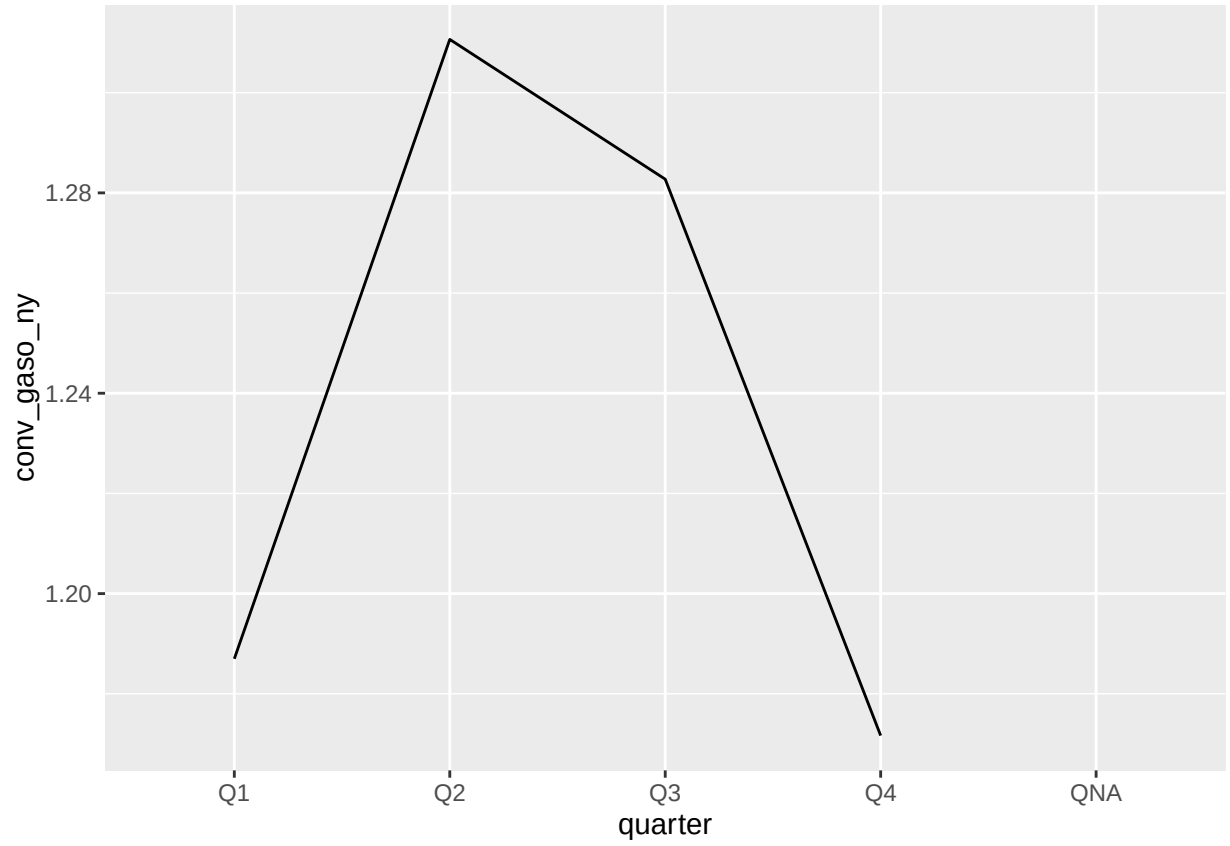

```
q4
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



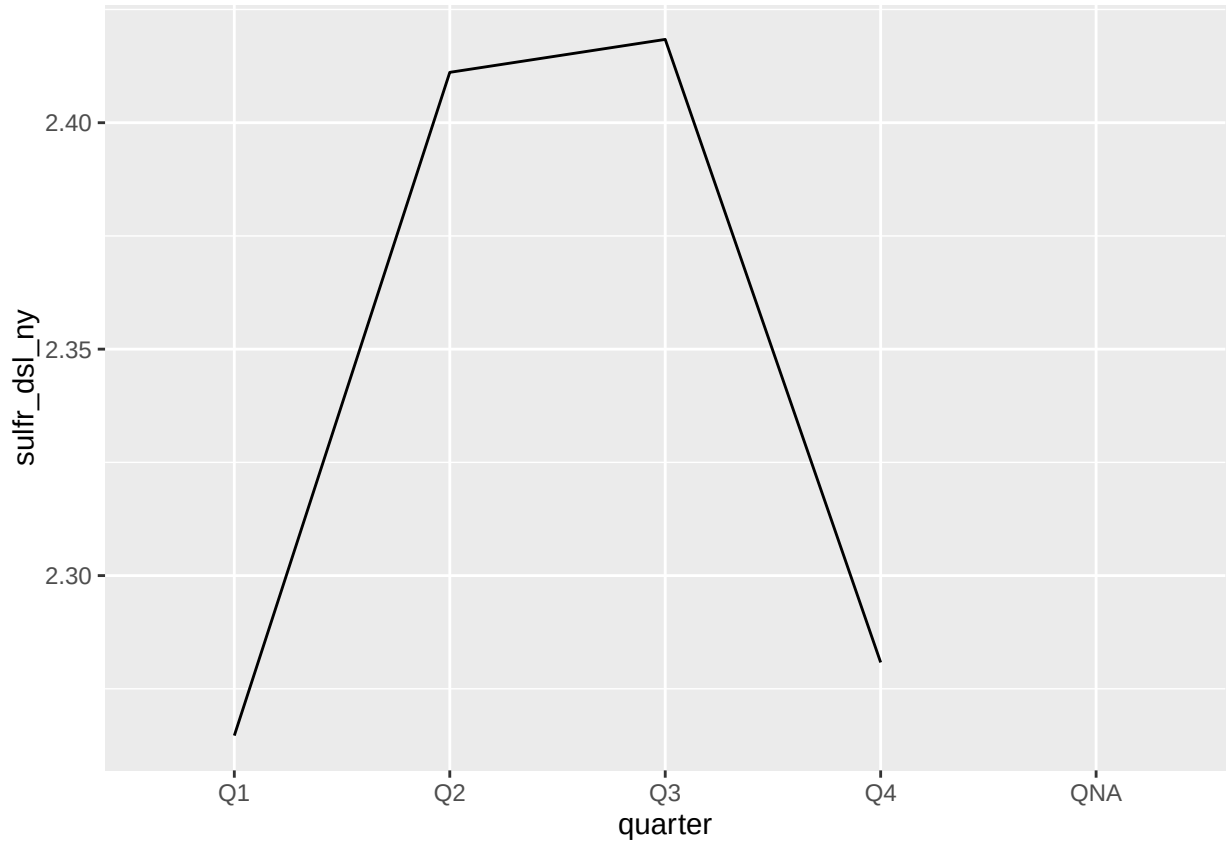

```
q5
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



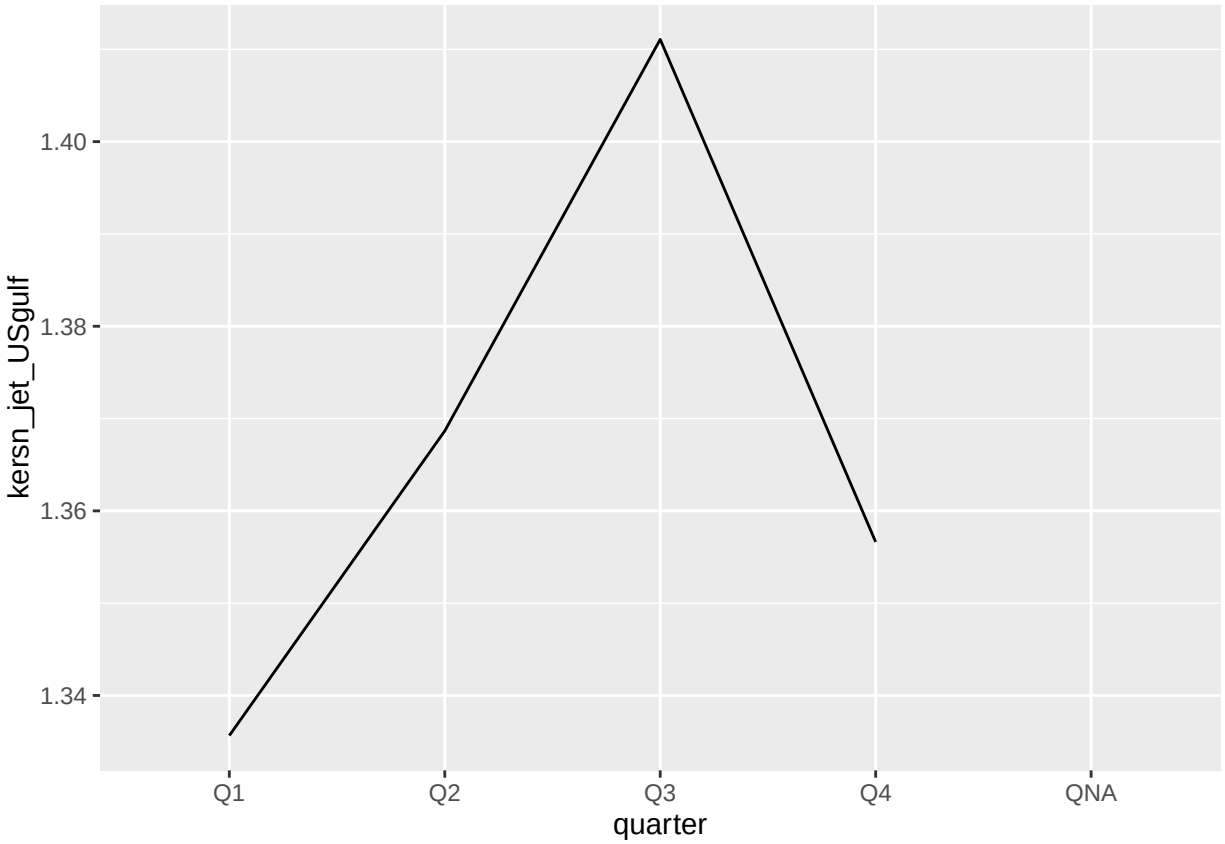

```
q6
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```




```
q7
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



- Quarter 1 and Quarter 4 are generally the ones with the lowest prices, except propane prices where it is quarter 2 with lowest price