

# Assignment 2

**Loading in the required packages and data to be used to answer the problems.**

```
# Loading required packages.
library(tidyverse) # loads ggplot2, dplyr, readr

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.2      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.4.3      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
  explicit

library(lubridate) # for date/time manipulation

# Loading in the data
df <- read_csv("elmarket_13_19.csv")

Rows: 2556 Columns: 21
-- Column specification -----
Delimiter: ","
dbl  (20): volume, wind_production, price, temperature, precip, windspeed, m...
date  (1): date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Assignment 2

### Task A. Plotting monthly averages of temperature, volume, price and wind production.

```
# Creating a function that plots the average monthly data of a specified variable.
plot_avg_monthly <- function(data, variable) {

  # Check if the specified variable is valid
  if (variable %in% names(data)) {

    # Extract year and month, calculate the average per month
    avg_data <- data %>%
      mutate(year_month = floor_date(date, "month")) %>%
      group_by(year_month) %>%
      summarise(avg_var = mean(get(variable), na.rm = TRUE))

    # Plot
    p <- ggplot(avg_data, aes(x = year_month, y = avg_var)) +
      geom_line() +
      labs(y = paste("Average", variable),
           x = "Month",
           title = paste("Average Monthly", variable)) +
      theme_classic()
    return(p)

    # If the specified variable is invalid
  } else {
    stop("Invalid variable name. Choose a variable present in the data frame.")
  }
}

# Create a loop that creates plots of specified variables

# List of variables that we want to visualize monthly data
variable_list <- c("temperature", "volume", "price", "wind_production")

# Initialized list to store the plot objects
plot_list <- list()

# Looping through the variable list and applying the plot function for each variable
```

```

for (i in variable_list) {
  p <- plot_avg_monthly(data = df, variable = i) # calling the monthly plot function
  plot_list[[i]] <- p                          # storing the plot in a list making it a
}

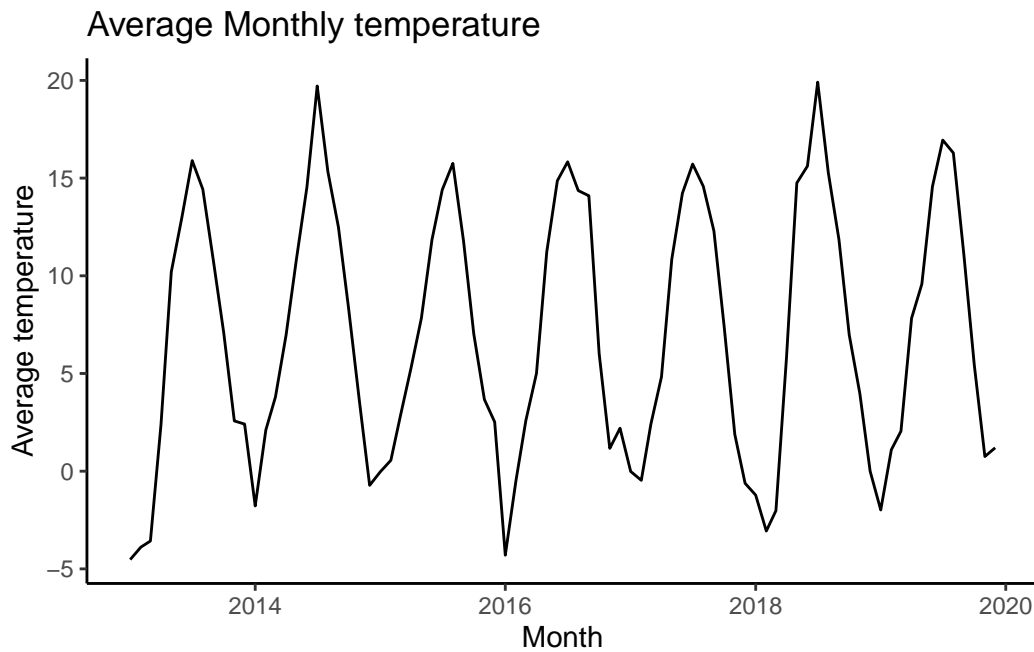
```

We start by commenting the plot of monthly averages of temperature.

```

# Plot monthly average temperature
plot_list$temperature

```



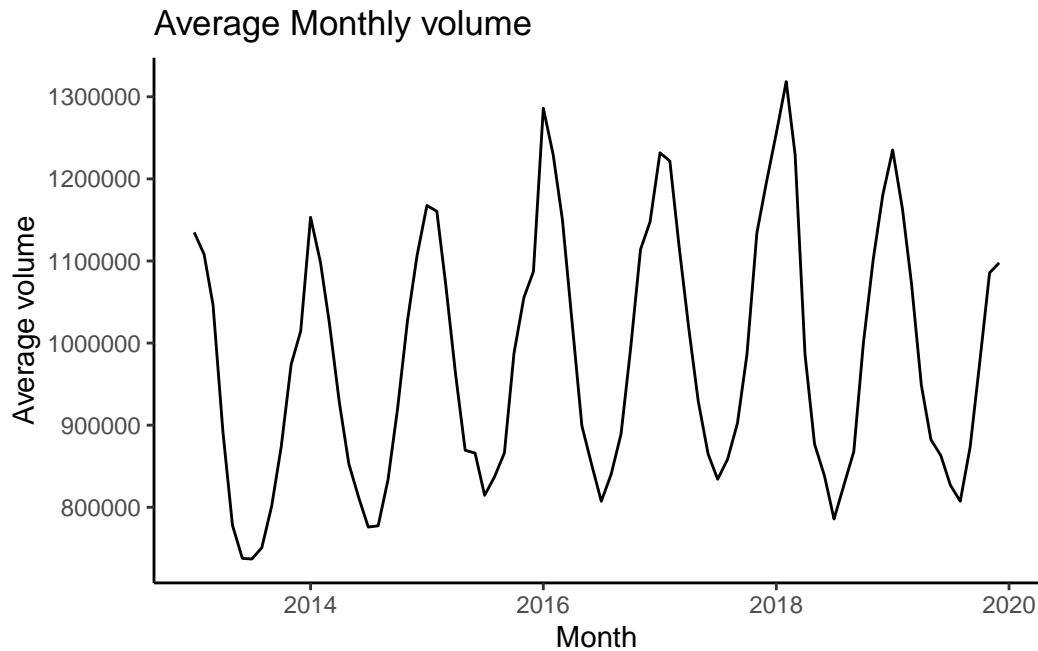
As expected we see variations in average temperatures between the seasons. We notice that the lowest temperatures are recorded during winter and the highest during summers. Also, we notice that the min. and max. temperatures are relatively equal each year.

We can also comment on the plot of monthly averages in volume.

```

# Plot monthly average volume
plot_list$volume

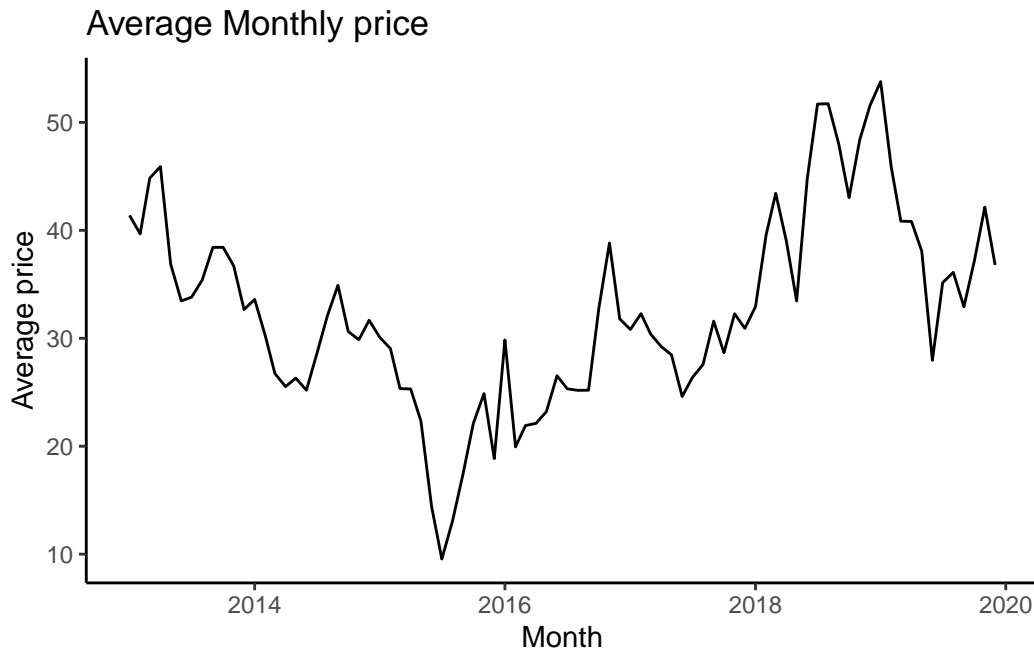
```



In the plot we see that the energy consumption is at max during winter and is lowest during winters. There are also differences between different years, but no great differences.

The next plot to examine is the plot of monthly averages of price.

```
# Plot monthly average price  
plot_list$price
```



We notice some seasonal trends in the prices of energy, however, the biggest variations is between years. We notice a substantial dip in energy prices during the summer of 2015, however the price continued to rise up until the winter of 2019. After that it has returned to 2018-levels.

The last single plot to examine is the plot of monthly wind production.

```
# Plot monthly average wind production
plot_list$wind_production.
```

NULL

Also in this plot we notice seasonal variations with highs during autumn/winter and lows during summer.

Lastly, we can examine the relationship between the different variables.

```
# Create a function that handles multiple variables and creates single plot
plot_avg_monthly_multi <- function(data, variables) {

  # Check if the specified variable is valid
  if (all(variables %in% names(data))) {
```

```

# Extract year and month, calculate the average per month of both variables
avg_data <- data %>%
  mutate(year_month = floor_date(date, "month")) %>%
  group_by(year_month) %>%
  summarise(avg_var1 = mean(get(variables[1]), na.rm = TRUE),
            avg_var2 = mean(get(variables[2]), na.rm = TRUE),
            .groups = "drop")

# Compute a scaling factor to map both variables onto a similar scale
scale_factor <- mean(avg_data$avg_var1, na.rm = TRUE) /
  mean(avg_data$avg_var2, na.rm = TRUE)

# Plot
p <- ggplot(avg_data, aes(x = year_month)) +
  geom_line(aes(y = avg_var1, color = "red")) +
  geom_line(aes(y = avg_var2 * scale_factor, color = "blue", linetype = "dashed")) +
  scale_y_continuous(
    name = paste("Average", variables[1]),
    sec.axis = sec_axis(~. / scale_factor, name = paste("Average", variables[2])),
    labels = scales::number
  ) +
  labs(
    x = "Month",
    title = paste("Monthly averages of", paste(variables, collapse = " & "))
  ) +
  theme_classic()
return(p)

# If the specified variable is invalid
} else {
  stop("Invalid variable name. Choose variables present in the data frame.")
}
}

# Plot relationship between variables

# Initialize empty list
plot_list_multi <- list()

# Loop over variables and store comparison plots in list
for (i in 1:(length(variable_list) - 1)) {

```

```

for (j in (i + 1):length(variable_list)) {
  current_vars <- c(variable_list[i], variable_list[j])
  p <- plot_avg_monthly_multi(df, current_vars)
  plot_name <- paste(current_vars, collapse = "_vs_")
  plot_list_multi[[plot_name]] <- p
}
}

```

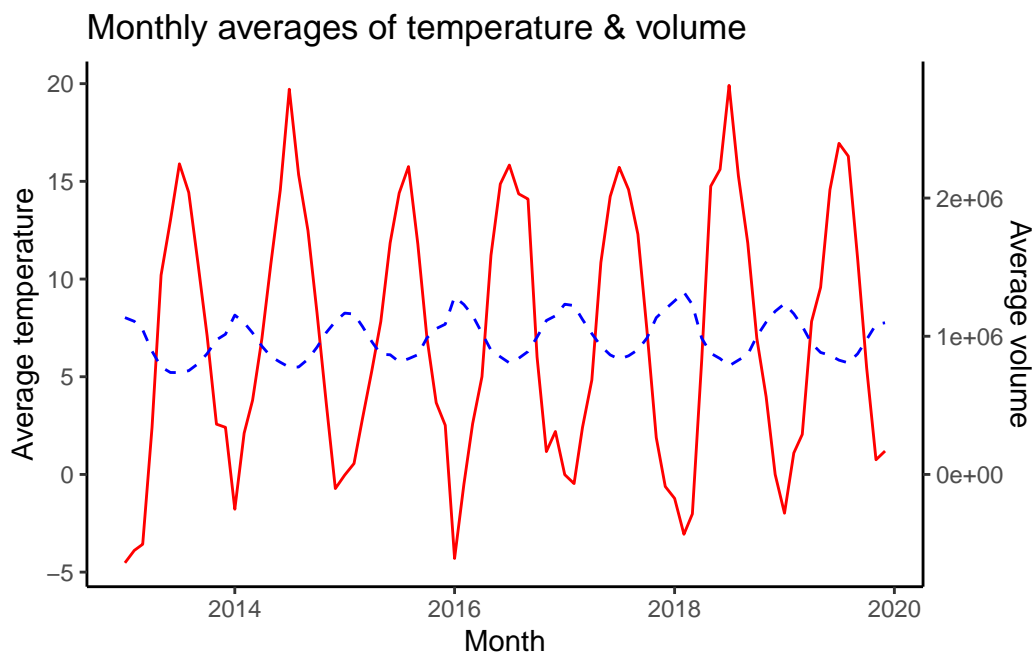
Comparing the four specified variables adds up to six comparison plots. We chose only to discuss the three most relevant, temperature vs. volume and price vs. volume. Notice that to compare the variables, we have added a second axis to each plot since the variables have different scales. In some cases, like the temperature vs. volume plot, this has lead to some errors. Errors like this is easy to spot if you compare with the single variable plot. Despite the scaling errors we chose to include them since they visualize relationships between the variables.

The first plot we examine is the temperature vs. volume plot.

```

# Compare temperature vs. volume
plot_list_multi$temperature_vs_volume

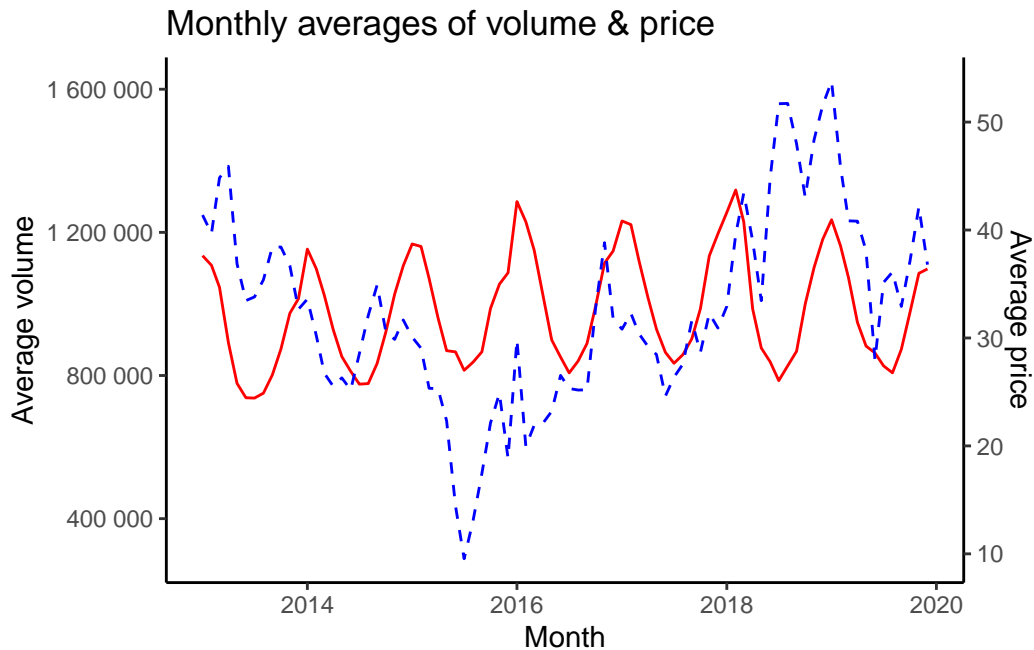
```



Despite the errors in scaling, we notice that when temperatures increases, the volume decrease and vice versa.

The second plot of interest is the volume vs. price plot.

```
plot_list_multi$volume_vs_price
```



We notice the at times close relationship of the two variables. In periods of high volume, the price tends to be relatively high and vice versa. This is likely due to a energy surplus in periods with low demand resulting in low prices, and increased prices in periods of high demand.

### Task B. Why will a OLS regression on quantity not provide an estimate of demand?

An OLS regression of quantity on price alone, even with controls, is not fitted to estimate demand due to many reasons:

1. Endogeneity exists as price and quantity in markets are simultaneously determined
2. Omitted variables like for example consumer income, advertising and seasonality. These can affect and lead to omitted variable bias.
3. Measurement errors in price and quantity can contribute to bias for the estimates for the parameters



4. OLS (or Gauss-Markow theorem) assumes a linear relationship. This may not hold if the demand curves are nonlinear.
5. Heteroskedasticity which is variability in the error term given any explanatory variable. In this case the variability in quantity varies across price levels which in return can affect standard errors.
6. The demand's dynamic behavior and seasonality may not be sufficiently captured. To estimate demand accurately, more advanced econometric methods are often required. These advanced methods usually contain instrumental variable regression and structural equation modelling. Controlling for relevant factors and improve our data quality can also make the estimation precision better.

### **Task C. Considering choice of instrument for price when estimating demand**

When choosing a instrument for price when estimating demand, it's important to consider the validity of the instrument in terms of the requirements for a valid instrument. These requirements can be relevance, exogeneity, and exclusion restrictions:

1. Why is temperature not a valid instrument for price when estimating demand? The key requirements for exogeneity may not be satisfied if one were to use temperature as an instrument for price when estimating demand. (Exogeneity refers to the condition in which an independent variable is unrelated to the error term, indicating that it is not influenced by unobserved factors in the model.) Temperature is usually affected by factors that also may affect demand, like seasonality and consumer behavior. The instrument would not be valid due to endogeneity concerns if temperature is correlated with unobservable demand shocks.
2. Why can magazine levels (or deviations) and wind power production potentially be good instruments for price when estimating demand? Magazine levels and wind power production could potentially be sufficient instruments for price as they can fulfil the requirements for a valid instrument. They should be relevant in other terms they should be affecting the price. In addition, they should be exogenous, meaning that they should not be correlated with unobservable factors. But establishing their exogeneity requires thoughtful consideration and testing.
3. Why is it necessary to control for seasonality (say, calendar month) and temperature? It is necessary to control for seasonality and temperature to avoid omitted variable bias. (Omitted variable bias appear if a relevant variable that affects the dependent variable is excluded in a regression model. This leads to biased and estimates of the coefficients of the variables that are included.) Seasonality captures variation in the systematic demand throughout the year unrelated to price. Meanwhile temperature affects demand independently. When we control for these factors, we ensure precise demand estimates.

4. How could controlling for weekday and year be useful? Incorporating controls for weekday and year is beneficial to justify for demand patterns variation. Weekdays and weekends usually give different demand profiles given the consumers consumption. Meanwhile years would capture trends that are long-term and macroeconomic affects on demand. These controls strengthen the accuracy of demand estimates, while taking into account sources of variation unrelated to price.