Data Cleaning

FRE530: Lab 1

- Create a `FRE530 labs` folder in your computer
- Create a `Data` subfolder in your "FRE530 labs" folder
- Download the Weekly U.S. No 2 Diesel Retail Prices (XLS) from the US EIA website and save the file as `diesel_weekly.xls`
- Create a new script called `Lab1.R` and save it in your `FRE 530 labs` folder
- Use `pacman::p_load()` to load the packages you will need for your analysis. Here are some suggested packages: `pacman::p_load(here, readxl, dplyr, janitor, Quandl, xts, lubridate)`.

```
pacman::p_load(here, readxl, dplyr, janitor, Quandl, xts, lubridate)
```

- Use `read_excel()` and `here()` functions to open the diesel data. Call this object `diesel_raw`. *Hint: use `sheet = "Data 1" and skip = 2` to read the Data 1 sheet and skip the first two rows of data. You can also use the `clean_names()` function to read the data in.

```
diesel_raw <- read_excel(here("Data", "diesel_weekly.xls"), sheet = "Data 1", skip = 2) %>%
  clean_names()
```

- Create a new object called `diesel` that cleans the `diesel_raw` object:
  - use `rename()` to rename `weekly_u_s_no_2_diesel_retail_prices_dollars_per_gallon` to `p_diesel`
  - use `filter()` to remove missing `p_diesel` data and to keep observations from June 1, 2012 onwards.
  - use `mutate()` and `ymd()` to convert the date into a certain date structure

```
diesel <- diesel_raw %>%
  rename(p_diesel = weekly_u_s_no_2_diesel_retail_prices_dollars_per_gallon) %>%
  filter(!is.na(p_diesel) & date > "2012-06-01") %>%
  mutate(date = ymd(date))
```

- Use the `Quandl()` command to download weekly soybean data and call it `soybean_raw`. Specifically the code is as follows: `soybean_raw <- Quandl("CHRIS/ICE_IBO2", type = c("raw"), collapse = "weekly")`. If you get a message that you have reached your limit, you will have to create an account here to get an API key. If you were, add this line to your code `Quandl.api_key("add_code_here")`

```
# Quandl.api_key("add_code_here")

soybean_raw <- Quandl("CHRIS/ICE_IBO2", type = c("raw"), collapse = "weekly")
```

- Create a new object called `soybean` that cleans the `soybean_raw` object:
  - Use `dplyr::select()` to select `Date` and `Settle` columns only

    * The reason why we add `dplyr::` before `select()` is because some of you may get an unused argument error. There is another package loaded in R's system that also has the `selecT()` function, and R is confused which package to use. The code `dplyr::` before `select()` tells R to use the `select()` function from the `{dplyr}` package.
  - Use `rename()` to rename `Settle` to `p_soy`
  - Use `mutate()` and `ymd()` to add one day to the dates provided. The reason we are adding one day to the date is because we need to have a common column to merge two dataframes together. We want to merge by the `date` column, but the dates of `diesel` and `soybean` data are not aligned. Diesel prices were reported on Mondays (i.e., June 4, 2012, June 11, 2012, June 18, 2012, etc.), whereas soybean oil prices were reported on Sundays (i.e., June 3, 2012, June 10, 2012, June 17, 2012, etc.)
  - Use `dplyr::select()` to only keep `date` and `p_soy` columns

```
soybean <- soybean_raw %>%
  dplyr::select(Date, Settle)  %>%
  rename(p_soy = Settle) %>%
  mutate(date = ymd(Date) + 1) %>%
  dplyr::select(date, p_soy)
```

- Create a new object called `merge` that merges `diesel` and `soybean` by `date`

```
merge <- merge(soybean, diesel, by = c("date"))
```

- Use `xts()` to convert the `p_diesel` and `p_soy` columns to an xts object, and order by the `date` column. Call this object `soydiesel`

```
soydiesel <- xts(merge[c("p_soy", "p_diesel")], order.by = merge$date)
# soydiesel <- xts(merge[,2:3], order.by = merge$date)
```

- The first five rows and the class of `soydiesel` is provided below.

```
head(soydiesel)
```

```
##            p_soy p_diesel
## 2012-06-04 48.81    3.846
## 2012-06-11 49.67    3.781
## 2012-06-18 48.62    3.729
## 2012-06-25 49.84    3.678
## 2012-07-02 52.59    3.648
## 2012-07-09 53.64    3.683
```

```
class(soydiesel)
```

```
## [1] "xts" "zoo"
```

Codes only

```
pacman::p_load(here, readxl, dplyr, janitor, Quandl, xts, lubridate)

diesel_raw <- read_excel(here("Data", "diesel_weekly.xls"), sheet = "Data 1", skip = 2) %>%
  clean_names()

diesel <- diesel_raw %>%
  rename(p_diesel = weekly_u_s_no_2_diesel_retail_prices_dollars_per_gallon) %>%
  filter(!is.na(p_diesel) & date > "2012-06-01") %>%
  mutate(date = ymd(date))

# Quandl.api_key("add_code_here")
soybean_raw <- Quandl("CHRIS/ICE_IBO2", type = c("raw"), collapse = "weekly")

soybean <- soybean_raw %>%
  dplyr::select(Date, Settle)  %>%
  rename(p_soy = Settle) %>%
  mutate(date = ymd(Date) + 1) %>%
  dplyr::select(date, p_soy)

merge <- merge(soybean, diesel, by = c("date"))

soydiesel <- xts(merge[c("p_soy", "p_diesel")], order.by = merge$date)
# soydiesel <- xts(merge[,2:3], order.by = merge$date)

head(soydiesel)
class(soydiesel)
```