

# Lab 1 FRE 501

2024-08-20

## Imports

```
library(lubridate)
```

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

```
library(readr)  
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr   1.1.4      v stringr 1.5.1  
## v forcats 1.0.0      v tibble  3.2.1  
## v ggplot2 3.5.1      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 errors
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.3
```

```
## here() starts at C:/Users/hp/Desktop/FRE 501/Labs/FRE-501-Lab-1
```

```
library(janitor)
```

```
##  
## Attaching package: 'janitor'  
##  
## The following objects are masked from 'package:stats':  
##  
##   chisq.test, fisher.test
```

```
library(tidyverse)  
library(lubridate)  
library(readxl)  
library(gridExtra)
```

```
##  
## Attaching package: 'gridExtra'  
##  
## The following object is masked from 'package:dplyr':  
##  
##   combine
```

```
library(scales)
```

```
## Warning: package 'scales' was built under R version 4.3.3
```

```
##  
## Attaching package: 'scales'  
##  
## The following object is masked from 'package:purrr':  
##  
##   discard  
##  
## The following object is masked from 'package:readr':  
##  
##   col_factor
```

```
library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 4.3.3
```

```
##  
## Attaching package: 'cowplot'  
##  
## The following object is masked from 'package:lubridate':  
##  
##   stamp
```

```
library(stargazer)
```

```
##
## Please cite as:
##
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(tsibble)
```

```
## Warning: package 'tsibble' was built under R version 4.3.3
```

```
## Registered S3 method overwritten by 'tsibble':
##   method          from
##   as_tibble.grouped_df dplyr
##
## Attaching package: 'tsibble'
##
## The following object is masked from 'package:lubridate':
##
##   interval
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, union
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.3.2
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##   smiths
```

## Read data

```
wheat <- read_csv("data/wheat_historical.csv")
```

```
## Rows: 528 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (2): Time, %Chg
## dbl (7): Open, High, Low, Last, Change, Volume, Open Int
##
## 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.
```

```
head(wheat)
```

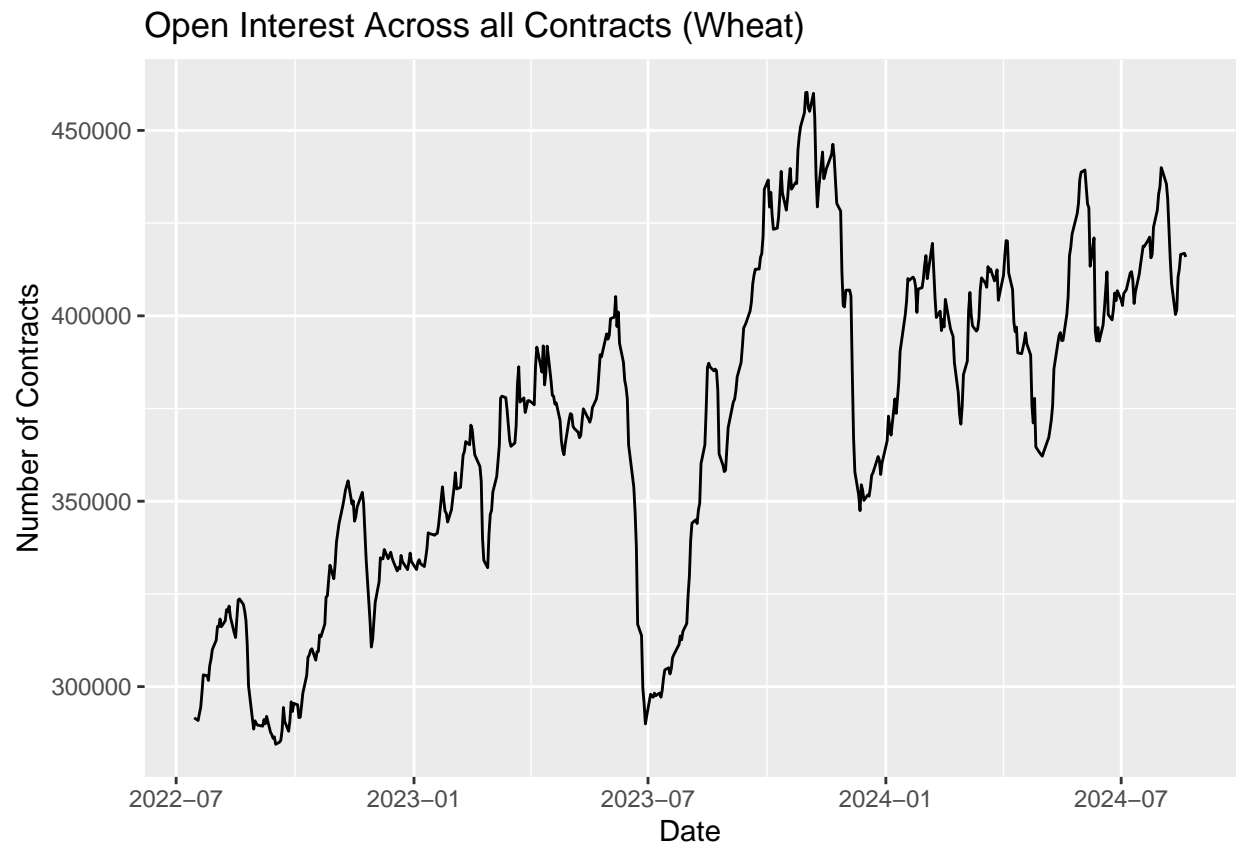
```
## # A tibble: 6 x 9
##   Time      Open High   Low Last Change '%Chg' Volume 'Open Int'
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   <dbl>   <dbl>
## 1 08/20/2024 551   559. 550. 556.   4.25 0.77% 112067 415877
## 2 08/19/2024 554. 554. 545 552.  -0.25 -0.05% 125305 416853
## 3 08/16/2024 549   556. 546. 552.   2.25 0.41% 123773 416551
## 4 08/15/2024 558. 568   548. 550.  -6    -1.08% 161435 412639
## 5 08/14/2024 553   559. 548. 556.   4.5   0.82% 201597 410420
## 6 08/13/2024 560. 562. 549. 552.  -8    -1.43% 163421 401604
```

```
wheat_small <- wheat |>
  mutate(Time = mdy(Time)) |>
  select(Time, Volume, `Open Int`) |>
  mutate(Year = year(Time), Month = month(Time))
head(wheat_small)
```

```
## # A tibble: 6 x 5
##   Time      Volume 'Open Int' Year Month
##   <date>      <dbl>   <dbl> <dbl> <dbl>
## 1 2024-08-20 112067     415877 2024     8
## 2 2024-08-19 125305     416853 2024     8
## 3 2024-08-16 123773     416551 2024     8
## 4 2024-08-15 161435     412639 2024     8
## 5 2024-08-14 201597     410420 2024     8
## 6 2024-08-13 163421     401604 2024     8
```

## Plot OI across all years

```
plt_overall <- ggplot(wheat_small, aes(x=Time, y=`Open Int`)) +
  geom_line() +
  labs(title = "Open Interest Across all Contracts (Wheat)",
       x = "Date",
       y = "Number of Contracts")
plt_overall
```

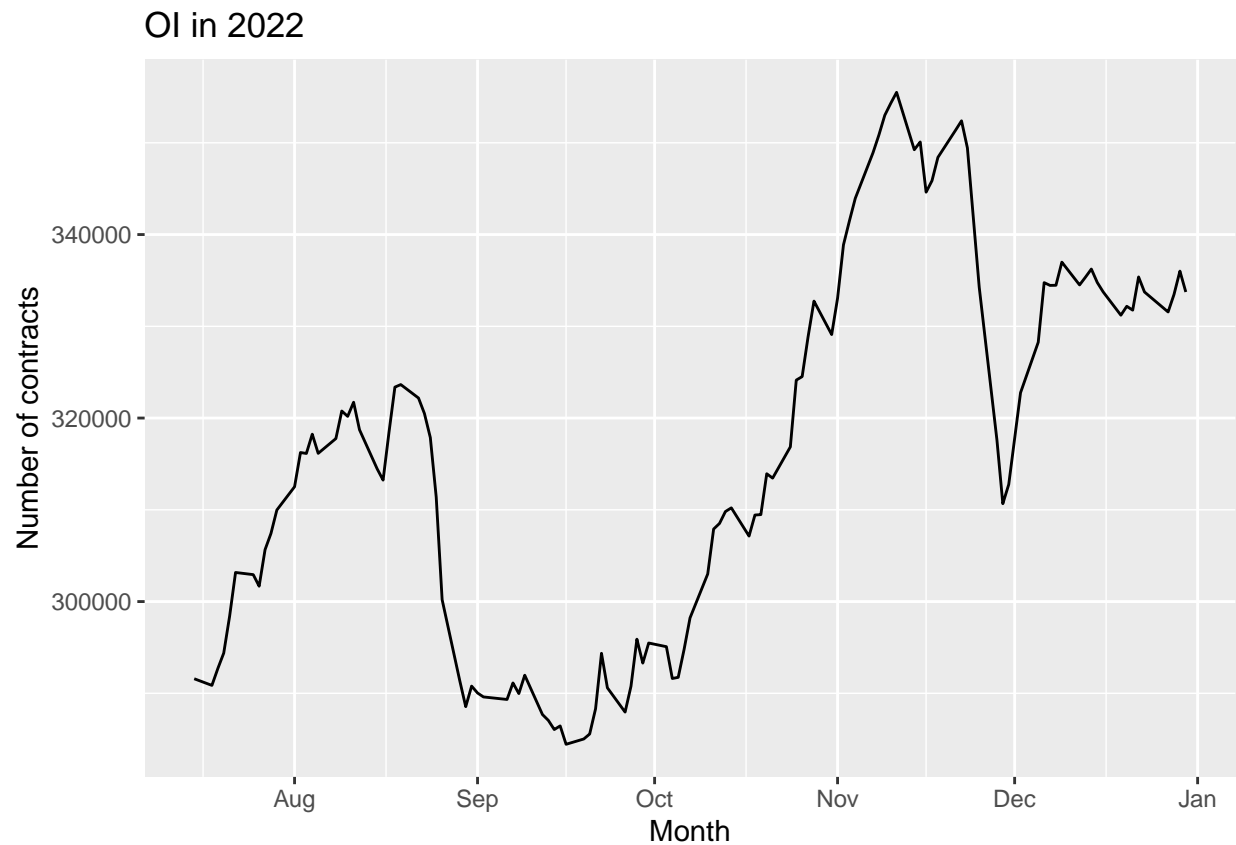


## Plot OI for one particular year

```
wheat_2022 <- wheat_small |> filter(Year==2022)

plt_2022 <- ggplot(wheat_2022, aes(x=Time,y=`Open Int`)) +
  geom_line() +
  labs(title = "OI in 2022",x = "Month",
        y = "Number of contracts")

plt_2022
```

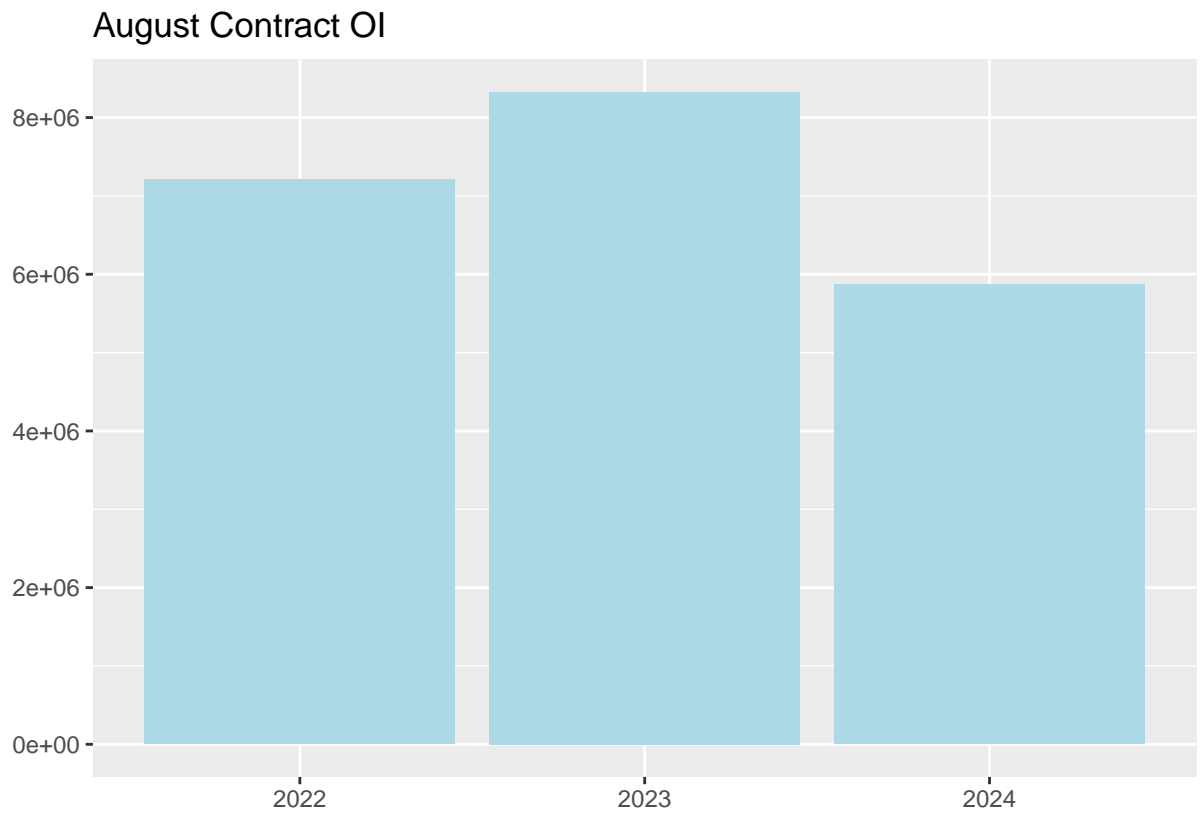


## Plot OI for one specific month (all years)

```
month_data <- wheat_small |> filter(Month == 8)

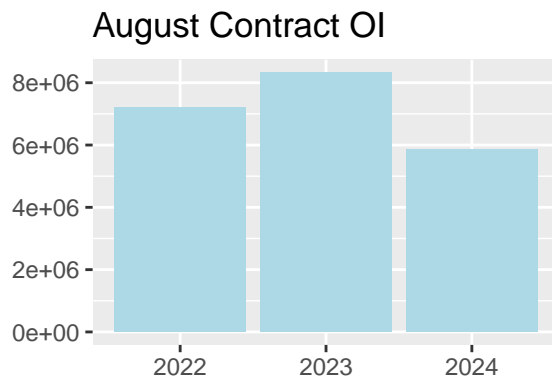
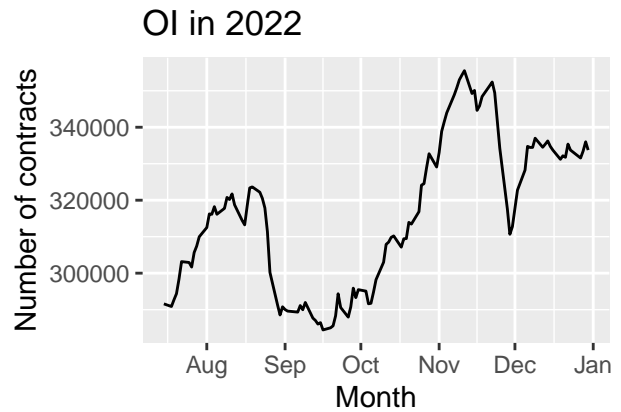
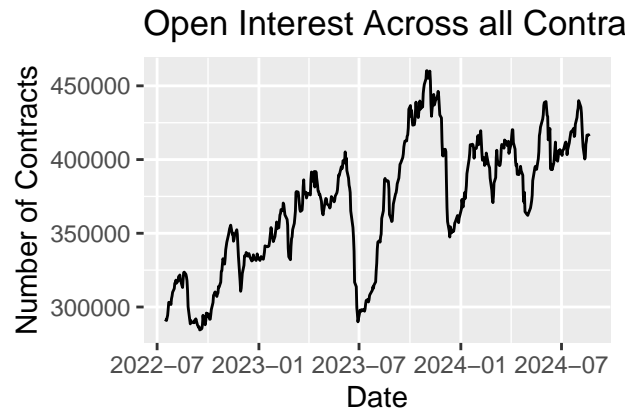
plt_month <- ggplot(month_data,
  aes(x=factor(Year), y=`Open Int`)) +
  geom_col(fill = "lightblue") +
  labs(title = "August Contract OI", x = "", y = "")

plt_month
```



## Plotting in a grid

```
plot_grid(plt_overall, plt_2022, plt_month,  
          ncol = 2,  
          align = 'v')
```



## Create month and year dummies

month	year	d1	d2	d3	d4	d5	d6	d7	d8
11	2020	0	0	0	0	0	0	0	0
11	2020	0	0	0	0	0	0	0	0
11	2020	0	0	0	0	0	0	0	0
11	2020	0	0	0	0	0	0	0	0
11	2020	0	0	0	0	0	0	0	0
11	2020	0	0	0	0	0	0	0	0

The goal is to simulate a one-hot encoded columns for month and year:

```
wheat_small <- wheat_small %>%
  mutate(d1 = ifelse(Month==1,1,0),
         d2 = ifelse(Month==2,1,0),
         d3 = ifelse(Month==3,1,0),
         d4 = ifelse(Month==4,1,0),
         d5 = ifelse(Month==5,1,0),
         d6 = ifelse(Month==6,1,0),
         d7 = ifelse(Month==7,1,0),
         d8 = ifelse(Month==8,1,0),
         d9 = ifelse(Month==9,1,0),
         d10 = ifelse(Month==10,1,0),
         d11 = ifelse(Month==11,1,0),
         D2022 = ifelse(Year==2022,1,0),
         D2023 = ifelse(Year==2023,1,0),
```



```

D2024 = ifelse(Year==2024,1,0))

head(wheat_small)

```

```

## # A tibble: 6 x 19
##   Time      Volume 'Open Int' Year Month   d1    d2    d3    d4    d5    d6
##   <date>      <dbl>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2024-08-20 112067      415877 2024     8     0     0     0     0     0     0
## 2 2024-08-19 125305      416853 2024     8     0     0     0     0     0     0
## 3 2024-08-16 123773      416551 2024     8     0     0     0     0     0     0
## 4 2024-08-15 161435      412639 2024     8     0     0     0     0     0     0
## 5 2024-08-14 201597      410420 2024     8     0     0     0     0     0     0
## 6 2024-08-13 163421      401604 2024     8     0     0     0     0     0     0
## # i 8 more variables: d7 <dbl>, d8 <dbl>, d9 <dbl>, d10 <dbl>, d11 <dbl>,
## #   D2022 <dbl>, D2023 <dbl>, D2024 <dbl>

```

## Regression

```

wheat_lm <- lm(`Open Int` ~ d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8 + d9 + d10 +
               d11 + D2022 + D2023 + D2024,
               data = wheat_small)
summary(wheat_lm)

```

```

##
## Call:
## lm(formula = `Open Int` ~ d1 + d2 + d3 + d4 + d5 + d6 + d7 +
##   d8 + d9 + d10 + d11 + D2022 + D2023 + D2024, data = wheat_small)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -71923 -14976   2285  18254  51768
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   431085     4486   96.107 < 2e-16 ***
## d1            -41511     5361  -7.744 5.17e-14 ***
## d2            -32490     5422  -5.993 3.89e-09 ***
## d3            -22163     5282  -4.196 3.21e-05 ***
## d4            -24959     5373  -4.645 4.33e-06 ***
## d5            -22699     5273  -4.305 2.00e-05 ***
## d6            -25489     5372  -4.744 2.71e-06 ***
## d7            -39938     4916  -8.125 3.35e-15 ***
## d8            -11615     4673  -2.485  0.0133 *
## d9             -4958     5045  -0.983  0.3262
## d10             24896     4986   4.993 8.16e-07 ***
## d11             40768     5015   8.129 3.25e-15 ***
## D2022          -119931     3539 -33.888 < 2e-16 ***
## D2023           -43735     2493 -17.541 < 2e-16 ***
## D2024              NA          NA      NA      NA
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22840 on 514 degrees of freedom
## Multiple R-squared:  0.7352, Adjusted R-squared:  0.7285
## F-statistic: 109.8 on 13 and 514 DF,  p-value: < 2.2e-16
```

## Print in a readable table (stargazer)

```
stargazer(wheat_lm, type = "text", digits = 0,
          title="Results", align=TRUE)
```

```
##
## Results
## =====
##                               Dependent variable:
##                               -----
##                               'Open Int'
##                               -----
## d1                           -41,511***
##                               (5,361)
##
## d2                           -32,490***
##                               (5,421)
##
## d3                           -22,163***
##                               (5,282)
##
## d4                           -24,959***
##                               (5,373)
##
## d5                           -22,699***
##                               (5,273)
##
## d6                           -25,489***
##                               (5,372)
##
## d7                           -39,938***
##                               (4,915)
##
## d8                           -11,615**
##                               (4,673)
##
## d9                           -4,958
##                               (5,045)
##
## d10                          24,896***
##                               (4,986)
##
## d11                          40,768***
##                               (5,015)
##
```

```

## D2022                -119,931***
##                      (3,539)
##
## D2023                -43,735***
##                      (2,493)
##
## D2024
##
##
## Constant             431,085***
##                      (4,485)
##
## -----
## Observations          528
## R2                    1
## Adjusted R2           1
## Residual Std. Error   22,842 (df = 514)
## F Statistic           110*** (df = 13; 514)
## =====
## Note:                 *p<0.1; **p<0.05; ***p<0.01

```