DATA481Gasoline

Mihir Harshe

2026-01-15

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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(lubridate)

gas <- readr::read_csv(
  "/Users/Mihir/Downloads/weekly_gas_prices.csv"
)
```

```
## Rows: 22360 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): fuel, grade, formulation
## dbl  (1): price
## 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.
```

```
glimpse(gas)
```

```
## Rows: 22,360
## Columns: 5
## $ date        <date> 1990-08-20, 1990-08-20, 1990-08-27, 1990-08-27, 1990-09-0~
## $ fuel        <chr> "gasoline", "gasoline", "gasoline", "gasoline", "gasoline"~
## $ grade       <chr> "regular", "regular", "regular", "regular", "regular", "re~
## $ formulation <chr> "all", "conventional", "all", "conventional", "all", "conv~
## $ price       <dbl> 1.191, 1.191, 1.245, 1.245, 1.242, 1.242, 1.252, 1.252, 1.~
```

```r
dim(gas)
```

```
## [1] 22360     5
```

```r
names(gas)
```

```
## [1] "date"        "fuel"        "grade"       "formulation" "price"
```

```r
summary(gas)
```

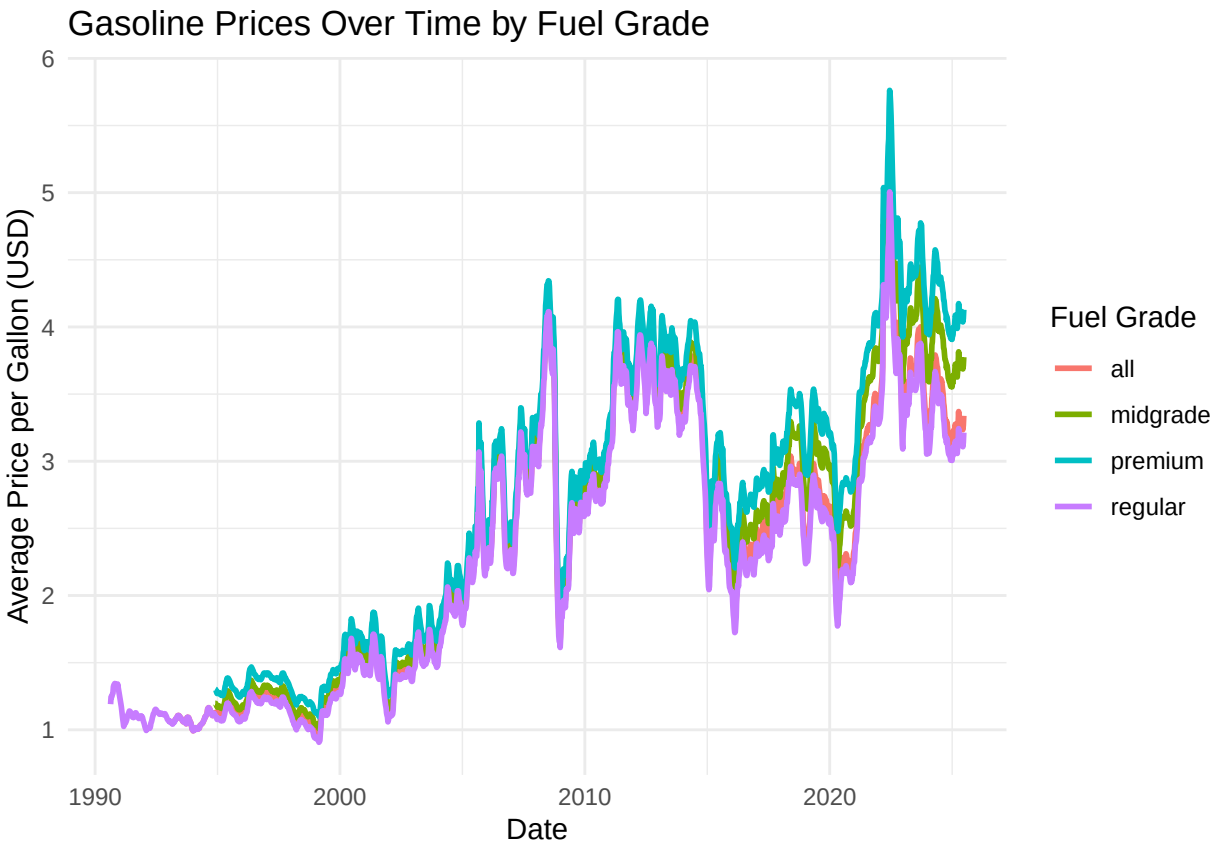
```
##       date                fuel              grade           formulation
##  Min.   :1990-08-20   Length:22360       Length:22360       Length:22360
##  1st Qu.:2002-04-29   Class :character   Class :character   Class :character
##  Median :2010-03-08   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2010-01-07
##  3rd Qu.:2017-10-30
##  Max.   :2025-06-23
##      price
##  Min.   :0.885
##  1st Qu.:1.575
##  Median :2.646
##  Mean   :2.582
##  3rd Qu.:3.389
##  Max.   :6.064
```

```r
if (!dir.exists("RData")) {
  dir.create("RData")
}

save(gas, file = "RData/gasoline_raw.RData")

gas_clean <- gas %>%
  mutate(date = as.Date(date, format = "%m/%d/%y")) %>%
  filter(fuel == "gasoline", formulation == "all") %>%
  group_by(date, grade) %>%
  summarise(
    avg_price = mean(price, na.rm = TRUE),
    .groups = "drop"
  )

ggplot(gas_clean, aes(x = date, y = avg_price, color = grade)) +
  geom_line(linewidth = 1) +
  labs(
    title = "Gasoline Prices Over Time by Fuel Grade",
    x = "Date",
    y = "Average Price per Gallon (USD)",
    color = "Fuel Grade"
  ) +
  theme_minimal()
```


Gasoline Prices Over Time by Fuel Grade
Average Price per Gallon (USD)
6
5
4
3
2
1
1990
2000
2010
2020
Date
Fuel Grade
all
midgrade
premium
regular