Crude Oil Price Predictor 2020

Christopher Page

8/12/2020

Introduction/Overview/Executive Summary

Oil is an important commodity. Some of our planet's inhabitants produce it. Many more consume it. All are affected by it. Because of its wide-ranging effects, oil commands collective attention enabled by continuous analysis.

One of the key metrics to take into account in the continuous analysis of the oil indutry is the daily closing price of crude oil futures traded on a global basis in such venues as the New York Mercantile Exchange.

Methods for predicting that price may entail studying the quantitative and qualitative factors at play inside the oil industry as well as in such adjacent industries as transportation and manufacturing.

They may also entail studying the often complex diplomatic, military, economic, cultural, and environmental developments that help explain why crude oil closing prices rise and fall as they do over time.

This project proposes a simpler method, one that predicts crude oil closing prices based on time and the closing prices of complementary (e.g., gasoline) and competing (e.g., platinum) commodities.

The premise is that there is much to be learned by studying the behaviors of commodity traders engaged in making daily investment decisions based on running evaluations of risks and rewards.

That premise informed the development of algorithms with the potential to predict crude oil closing prices with a viable level of accuracy, quantified as the Root Means Square Error (RMSE).

Using a 25,190-row data set constructed with publicly available information downloaded for free from https://www.nasdaq.com/, the author developed and then evaluated fifteen such algorithms.

Evaluation led to the selection of one algorithm that generated the lowest RMSE when applied to a test set. That algorithm used a Ranger model which took into account six key predictors:

(1) Year
(2) Month
(3) Closing Price of Heating Oil
(4) Closing Price of Gasoline
(5) Closing Price of Platinum
(6) Closing Price of Soybeans

The RMSE in question was 2.92, a figure equating to 12% of the 23.51 generated by an algorithm that only took into account the average closing price across the period of observation.

Method/Analysis

```
## Loading required package: caret
```

```
## Loading required package: lattice

## Loading required package: ggplot2

## Loading required package: caTools

## Warning: package 'caTools' was built under R version 4.0.2

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: dslabs

## Loading required package: forcats

## Loading required package: foreach

## Loading required package: gam

## Warning: package 'gam' was built under R version 4.0.2

## Loading required package: splines

## Loaded gam 1.20

## Loading required package: ggrepel

## Loading required package: ggthemes

## Loading required package: lubridate

##
## Attaching package: 'lubridate'

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

## Loading required package: purrr
```

```
##
## Attaching package: 'purrr'

## The following objects are masked from 'package:foreach':
##
##     accumulate, when

## The following object is masked from 'package:caret':
##
##     lift

## Loading required package: randomForest

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

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

## The following object is masked from 'package:ggplot2':
##
##     margin

## Loading required package: tidyverse

## -- Attaching packages ---------------------------------------------------------------------------

## v tibble  3.0.1     v readr   1.3.1
## v tidyr   1.1.0     v stringr 1.4.0

## -- Conflicts ------------------------------------------------------------------------------------
## x purrr::accumulate()      masks foreach::accumulate()
## x lubridate::as.difftime() masks base::as.difftime()
## x randomForest::combine()  masks dplyr::combine()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::lift()            masks caret::lift()
## x randomForest::margin()   masks ggplot2::margin()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
## x purrr::when()            masks foreach::when()
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   Commodity = col_character(),
##   Sector = col_character(),
##   `Closing Price` = col_double()
## )
```

```
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
## Joining, by = "date"
```

```
print("Crude Oil Commodity Prices began the August 2010 to July 2020 period at")
```

```
## [1] "Crude Oil Commodity Prices began the August 2010 to July 2020 period at"
```

```
crude_oil_price_at_beginning_of_period
```

```
## [1] 78.95
```

```
print("They ended that ten-year period at")
```

```
## [1] "They ended that ten-year period at"
```

```
crude_oil_price_at_end_of_period
```

```
## [1] 40.34
```

```
print("During that time, they averaged")
```

```
## [1] "During that time, they averaged"
```

```
mean_crude_oil_price_during_period
```

```
## [1] 70.13307
```

```
print("Running from a minimum of")
```

```
## [1] "Running from a minimum of"
```

```
minimum_crude_oil_price_during_period
```

```
## [1] -37.25
```

```
print("To a maximum of")
```

```
## [1] "To a maximum of"
```

```
maximum_crude_oil_price_during_period
```

```
## [1] 113.45
```

```
print("With a standard deviation of")
```

```
## [1] "With a standard deviation of"
```

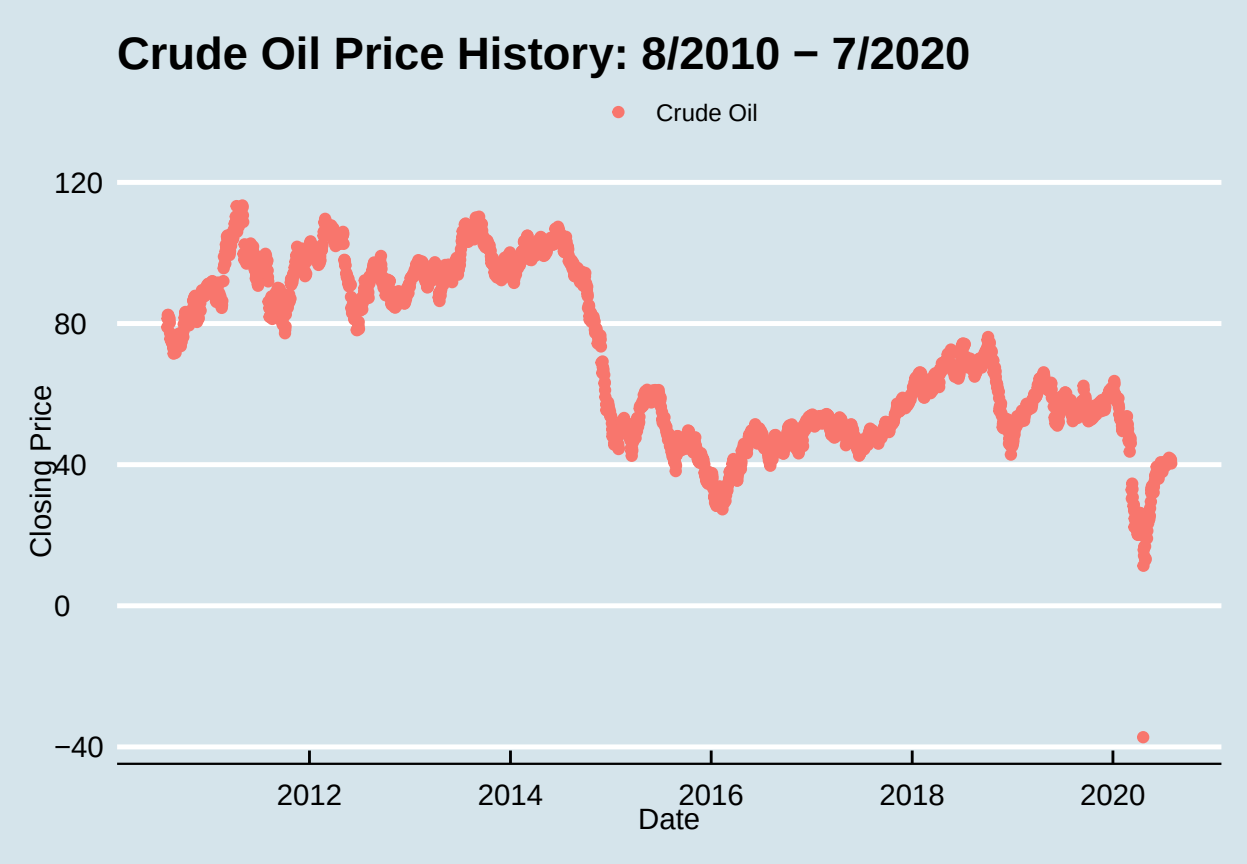
```
standard_deviation_during_period
```

```
## [1] 23.52209
```

```
### Second, this script plots the 2010-2020 timeline of crude oil daily closing prices
### Then, it saves the plot in a .png file entitled "crude_oil_price_history"

crude_oil_plot <- crude_oil %>%
  ggplot(aes(date, closing_price)) +
  geom_point(aes(color = commodity)) +
  xlab("Date") +
  ylab("Closing Price") +
  ggtitle("Crude Oil Price History: 8/2010 - 7/2020") +
  theme_economist() +
    theme(legend.position="top",
          legend.title = element_blank(),
          legend.box = "horizontal" ,
          legend.text=element_text(size=8.5)) +
    guides(col = guide_legend(nrow = 1))

crude_oil_plot
```



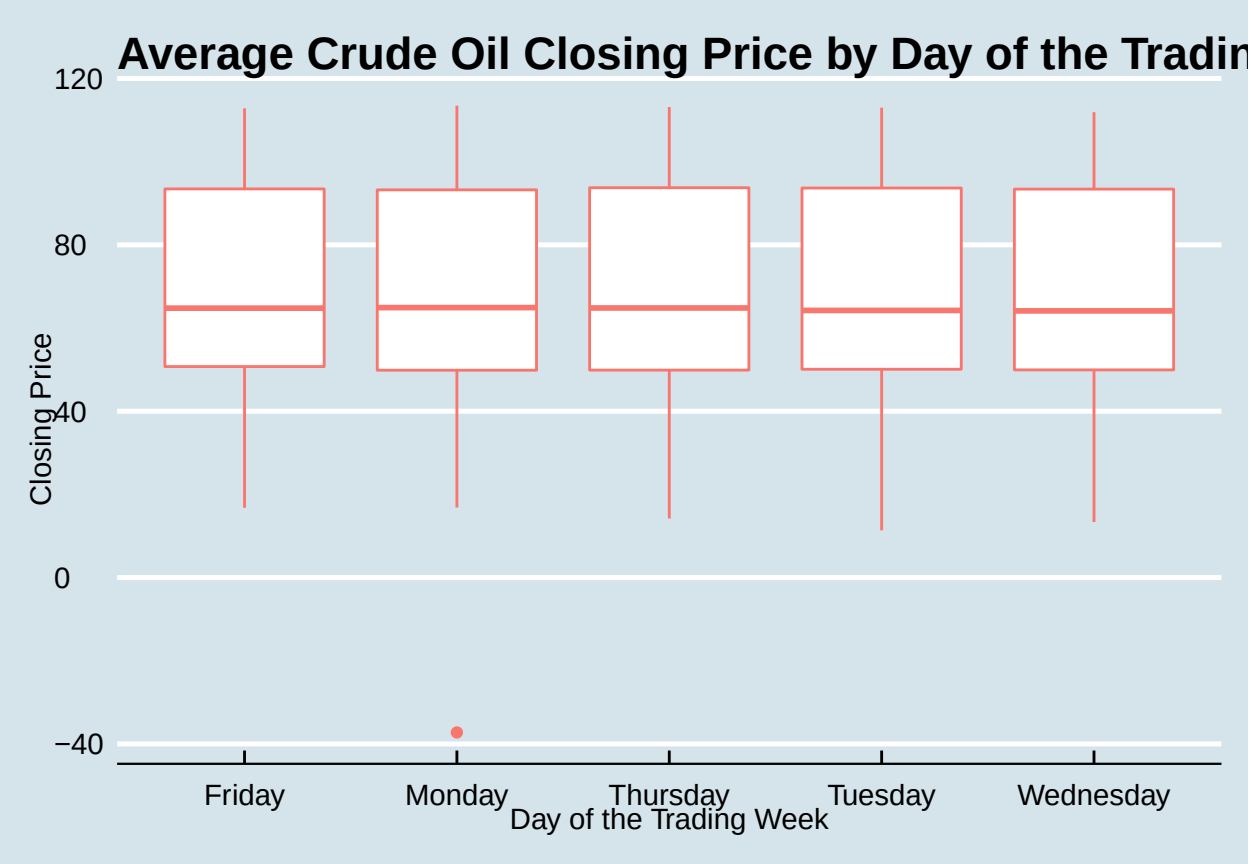

```
ggsave("fig/crude_oil_price_history.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
### Third, this script explores crude oil closing prices in terms of the cyclical components of time
#### It plots the average crude oil closing prices by day of the trading week for 2010-2020
#### Then, it saves the plot in a .png file entitled "crude_oil_average_by_day_of_the_week"

crude_oil_average_by_day_of_the_week_plot <- crude_oil_in_commodities_market %>%
  group_by(date_weekday) %>%
  ggplot(aes(as.factor(date_weekday), closing_price)) +
  geom_boxplot(aes(color = commodity), size = 0.5, show.legend = FALSE) +
  xlab("Day of the Trading Week") +
  ylab("Closing Price") +
  ggtitle("Average Crude Oil Closing Price by Day of the Trading Week: 2010 - 2020") +
  theme_economist()

crude_oil_average_by_day_of_the_week_plot
```



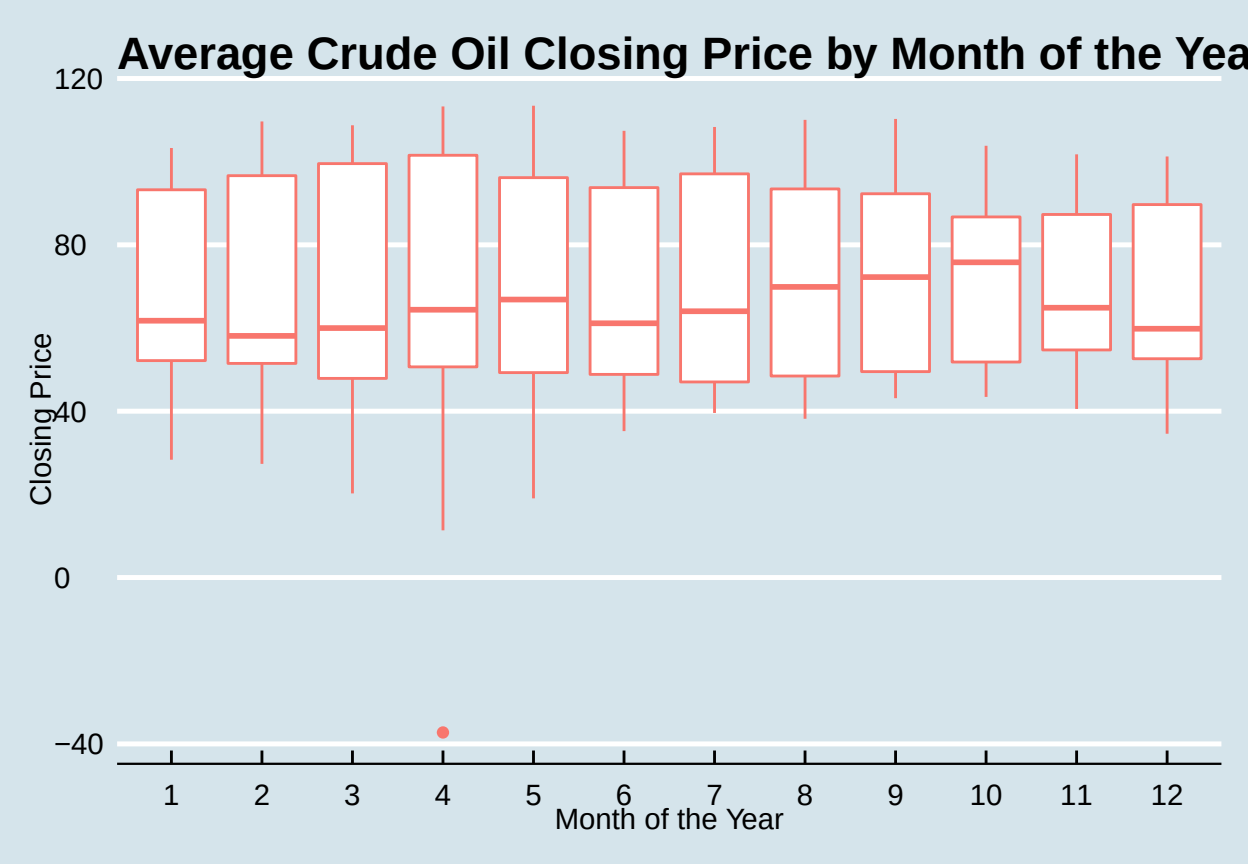

```
ggsave("fig/crude_oil_average_by_day_of_the_week.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#### It plots the average crude oil closing prices by month of the year for 2010-2020
#### Then, it saves the plot in a .png file entitled "crude_oil_average_by_month_of_the_year

crude_oil_average_by_month_of_the_year_plot <- crude_oil_in_commodities_market %>%
  group_by(date_month_of_the_year) %>%
  ggplot(aes(as.factor(date_month_of_the_year), closing_price)) +
  geom_boxplot(aes(color = commodity), size = 0.5, show.legend = FALSE) +
  xlab("Month of the Year") +
  ylab("Closing Price") +
  ggtitle("Average Crude Oil Closing Price by Month of the Year: 2010 - 2020") +
  theme_economist()

crude_oil_average_by_month_of_the_year_plot
```



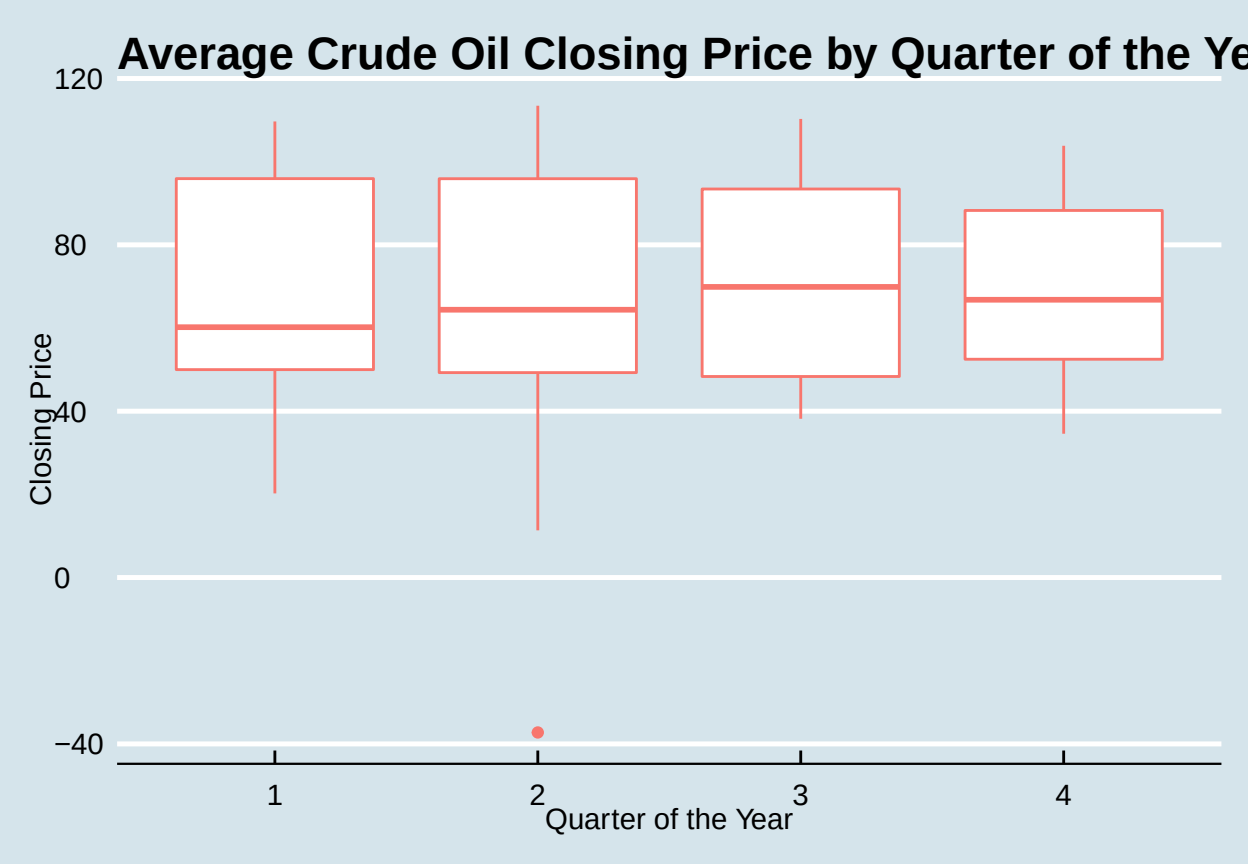

```
ggsave("fig/crude_oil_average_by_month_of_the_year.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#### It plots the average crude oil closing prices by quarter of the year for 2010-2020
#### Then, it saves the plot in a .png file entitled "crude_oil_average_by_month_of_the_year

crude_oil_average_by_quarter_of_the_year_plot <- crude_oil_in_commodities_market %>%
  group_by(date_quarter_of_the_year) %>%
  ggplot(aes(as.factor(date_quarter_of_the_year), closing_price)) +
  geom_boxplot(aes(color = commodity), size = 0.5, show.legend = FALSE) +
  xlab("Quarter of the Year") +
  ylab("Closing Price") +
  ggtitle("Average Crude Oil Closing Price by Quarter of the Year: 2010 - 2020") +
  theme_economist()

crude_oil_average_by_quarter_of_the_year_plot
```




```
ggsave("fig/crude_oil_average_by_quarter_of_the_year.png")
```

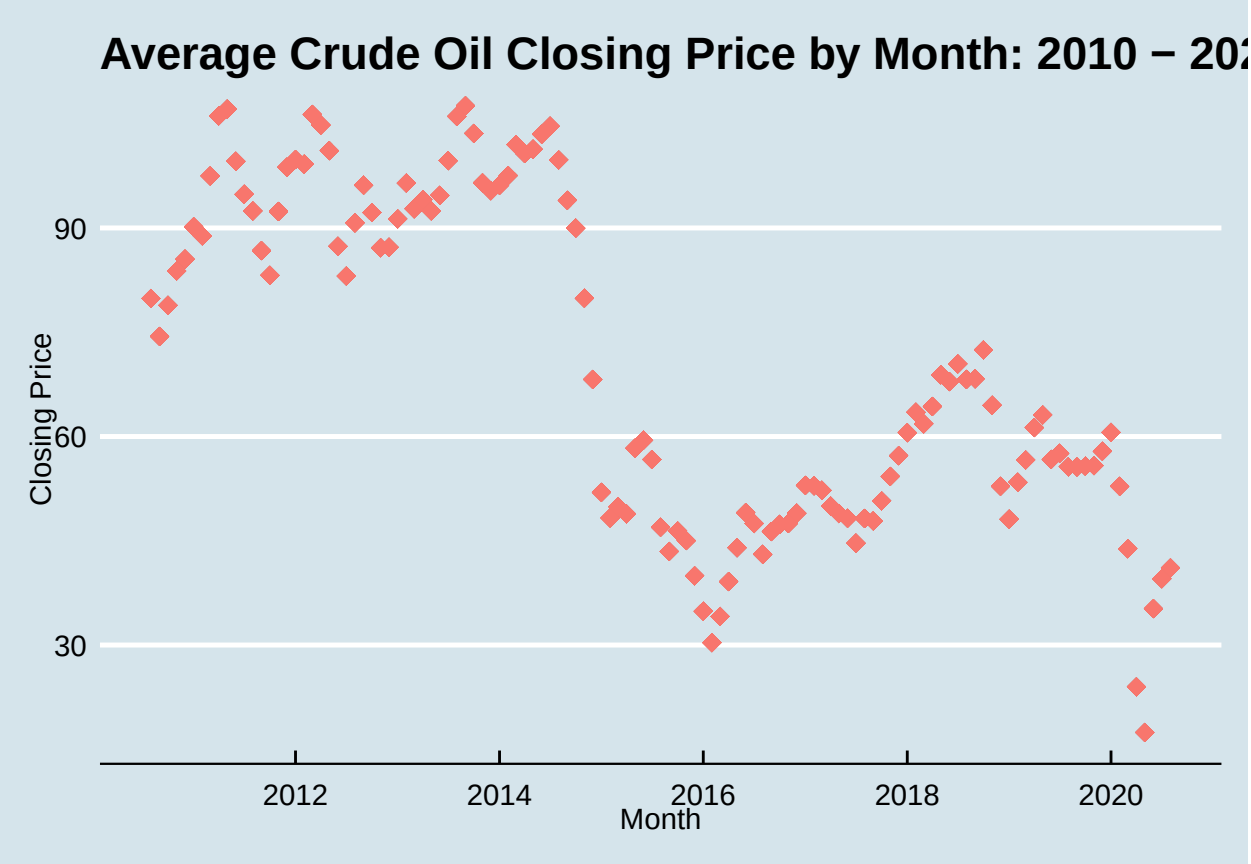
```
## Saving 6.5 x 4.5 in image
```

```
### Fourth, this script explores crude oil closing prices in terms of the linear components of time
#### It plots the average crude oil closing prices by month for each month in 2010-2020
#### Then, it saves the plot in a .png file entitled "crude_oil_average_by_month"

crude_oil_average_by_month <- crude_oil_in_commodities_market %>%
  group_by(date_month) %>%
  mutate(month_avg = mean(closing_price))

crude_oil_average_by_month_plot <- crude_oil_average_by_month %>%
  ggplot(aes(date_month, month_avg)) +
  geom_point(aes(color = commodity), shape = 18, size = 3, show.legend = FALSE) +
  xlab("Month") +
  ylab("Closing Price") +
  ggtitle("Average Crude Oil Closing Price by Month: 2010 - 2020") +
  theme_economist()

crude_oil_average_by_month_plot
```




```
ggsave("fig/crude_oil_average_by_month.png")
```

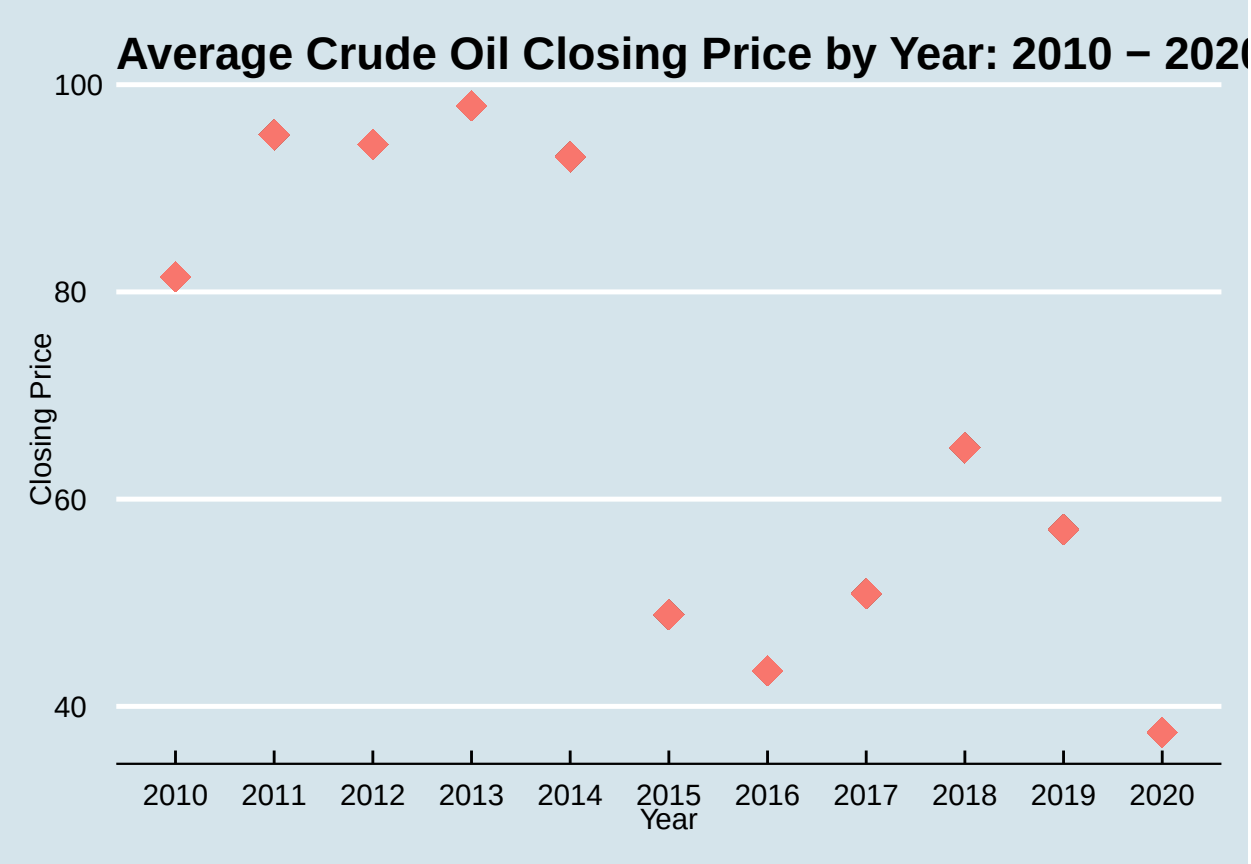
```
## Saving 6.5 x 4.5 in image
```

```
#### It plots the average crude oil closing prices by year for each year in 2010-2020
#### Then, it saves the plot in a .png file entitled "crude_oil_average_by_year"

crude_oil_average_by_year <- crude_oil_in_commodities_market %>%
  group_by(date_year) %>%
  mutate(year_avg = mean(closing_price))

crude_oil_average_by_year_plot <- crude_oil_average_by_year %>%
  ggplot(aes(as.factor(date_year), year_avg)) +
  geom_point(aes(color = commodity), shape = 18, size = 5, show.legend = FALSE) +
  xlab("Year") +
  ylab("Closing Price") +
  ggtitle("Average Crude Oil Closing Price by Year: 2010 - 2020") +
  theme_economist()

crude_oil_average_by_year_plot
```



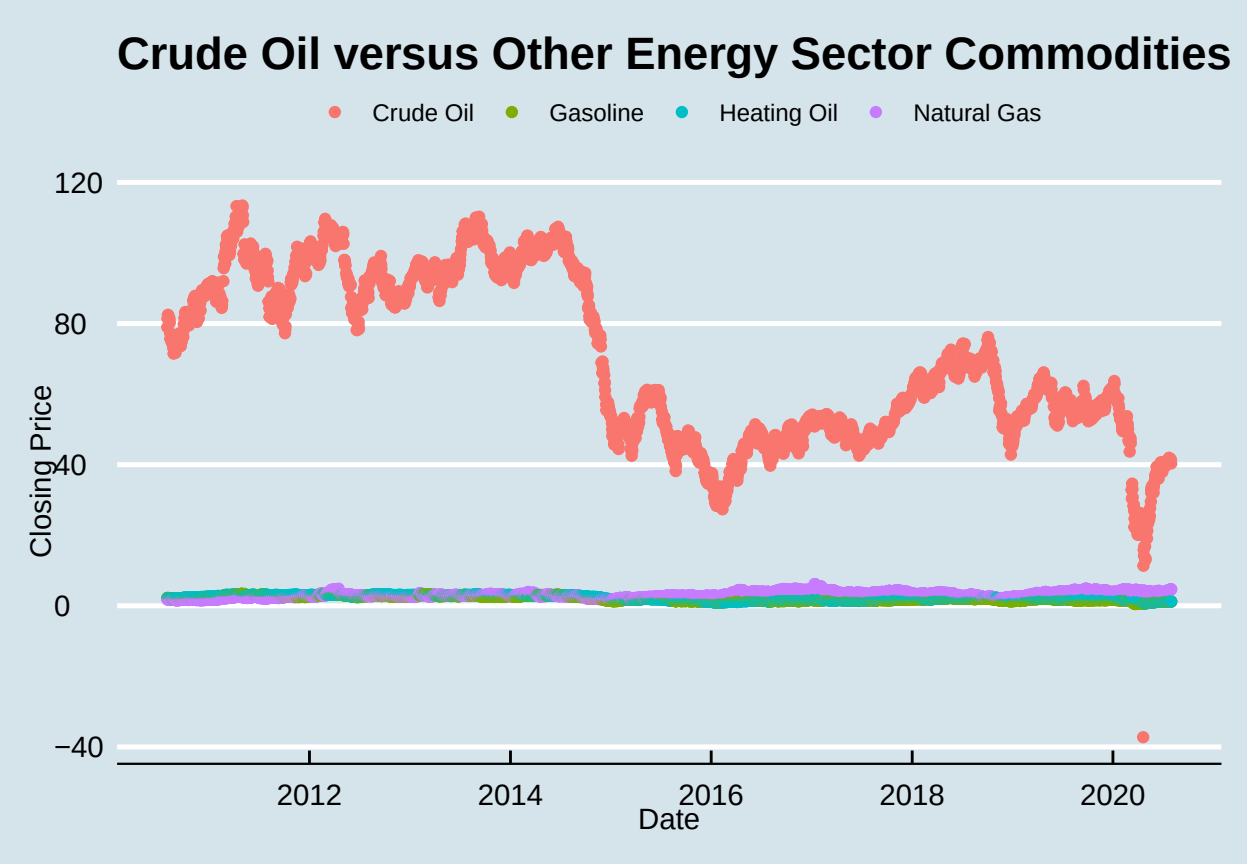

```
ggsave("fig/crude_oil_average_by_year.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
### Fourth, this script explores crude oil closing prices in comparison to those of other complementary
#### It plots the 2010-2020 timeline of crude oil daily closing prices compared to those of the other e
#### Then, it saves the plot in a .png file entitled "crude_oil_price_history_in_comparison_to_three_ot

crude_oil_vs_other_energy_plot <- energy %>%
  ggplot(aes(date, closing_price)) +
  geom_point(aes(color = commodity)) +
  xlab("Date") +
  ylab("Closing Price") +
  ggtitle("Crude Oil versus Other Energy Sector Commodities") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_vs_other_energy_plot
```



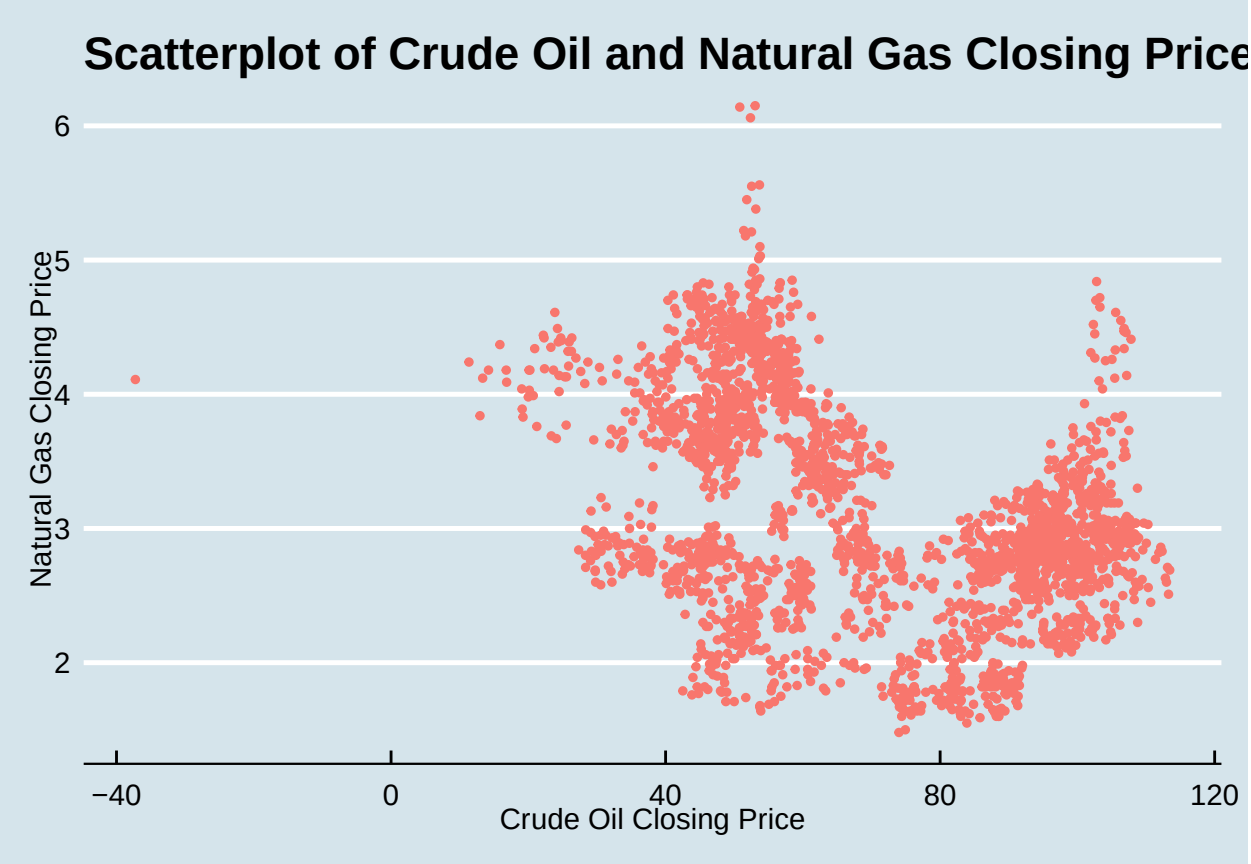

```
ggsave("fig/crude_oil_in_comparison_to_three_other_energy_sector_commodities.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#### As a next step, it generates scatterplots portraying the relationships between crude oil closing p

crude_oil_natural_gas_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, natural_gas_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Natural Gas Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Natural Gas Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_natural_gas_scatterplot
```

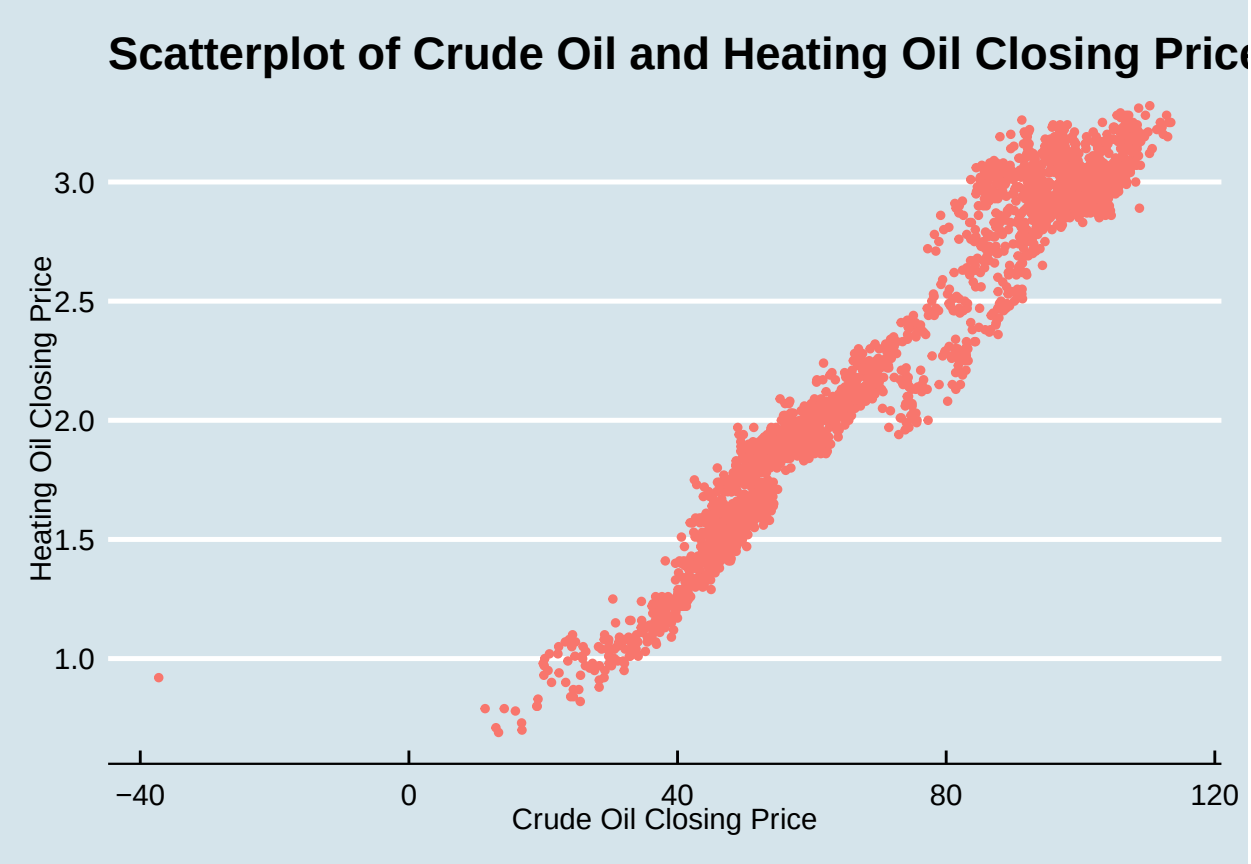



```
ggsave("fig/scatterplot_crude_oil_versus_natural_gas.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_heating_oil_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, heating_oil_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Heating Oil Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Heating Oil Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_heating_oil_scatterplot
```

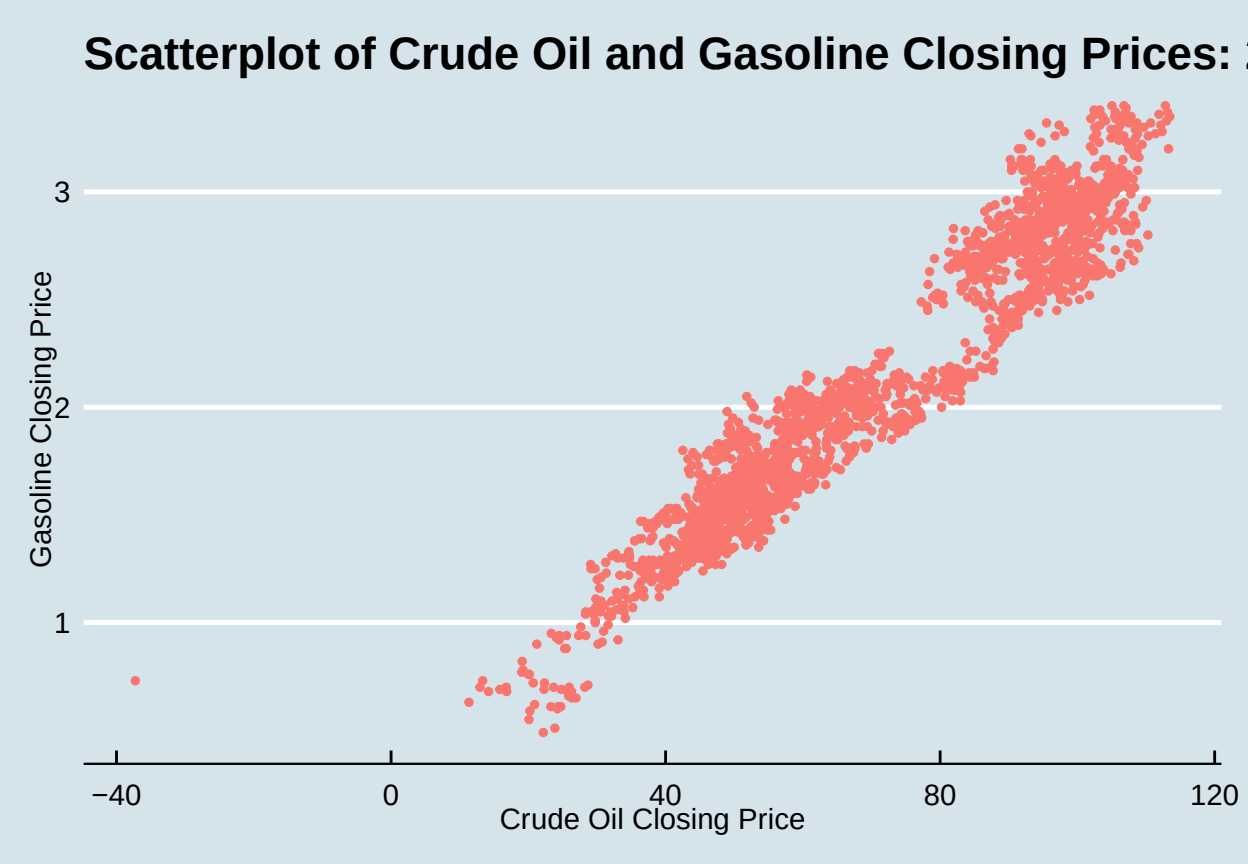



```
ggsave("fig/scatterplot_crude_oil_versus_heating_oil.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_gasoline_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, gasoline_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Gasoline Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Gasoline Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_gasoline_scatterplot
```




```
ggsave("fig/scatterplot_crude_oil_versus_gasoline.png")
```

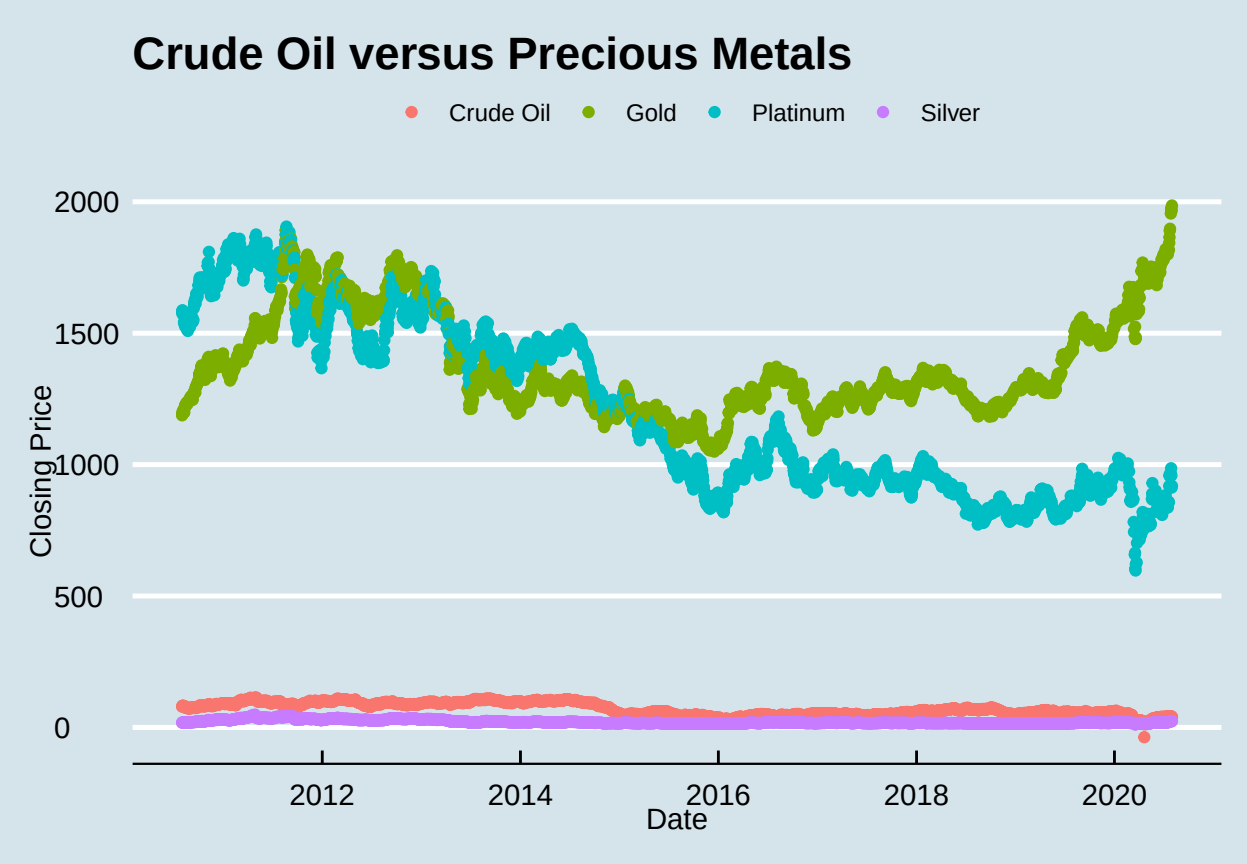
```
## Saving 6.5 x 4.5 in image
```

```
### It plots the 2010-2020 timeline of crude oil daily closing prices compared to those of the three pr
### Then, it saves the plot in a .png file entitled "crude_oil_price_history_in_comparison_to_three_com

crude_oil_vs_precious_metals <- bind_rows(crude_oil, precious_metals)

crude_oil_vs_precious_metals_plot <- crude_oil_vs_precious_metals %>%
  ggplot(aes(date, closing_price)) +
  geom_point(aes(color = commodity)) +
  xlab("Date") +
  ylab("Closing Price") +
  ggtitle("Crude Oil versus Precious Metals") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_vs_precious_metals_plot
```



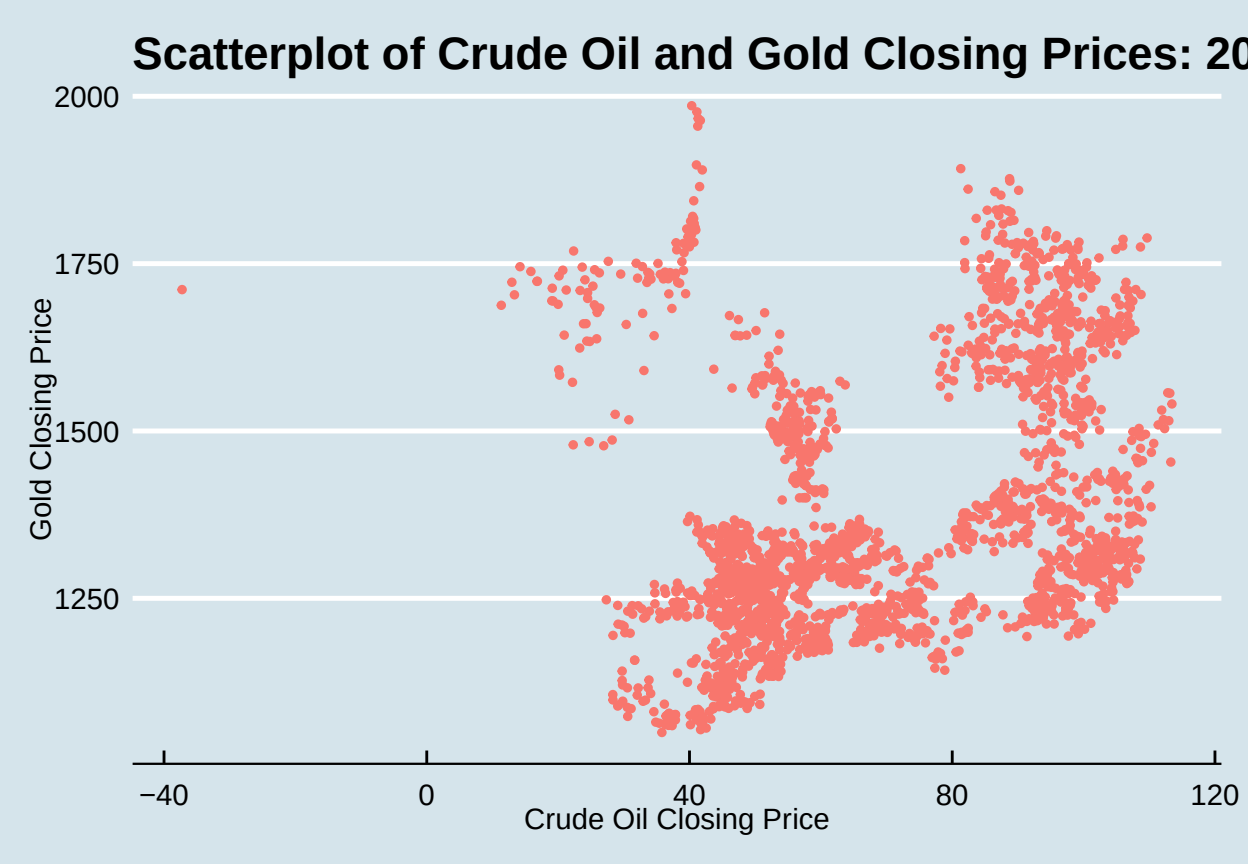

```
ggsave("fig/crude_oil_in_comparison_to_three_precious_metals_sector_commodities.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#### As a next step, it generates scatterplots portraying the relationships between crude oil closing p

crude_oil_gold_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, gold_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Gold Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Gold Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_gold_scatterplot
```

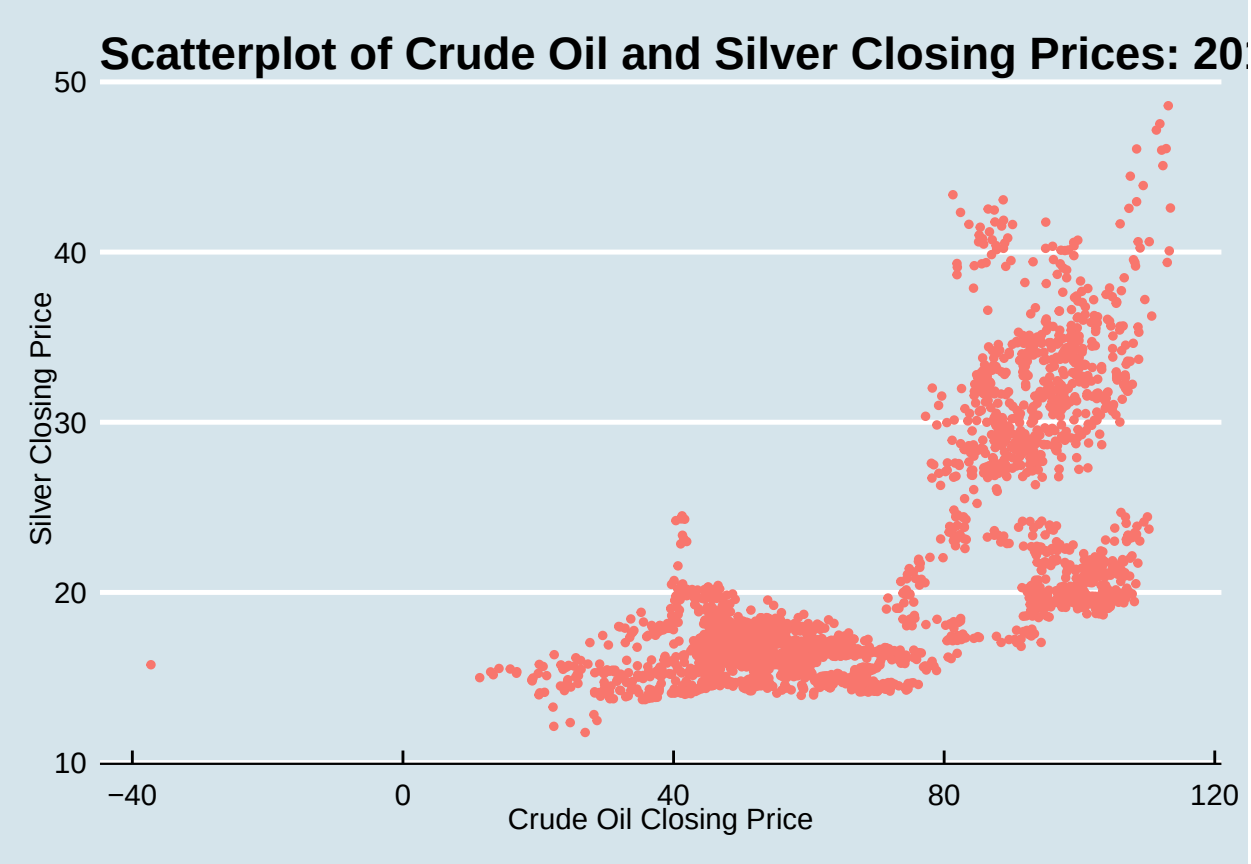



```
ggsave("fig/scatterplot_crude_oil_versus_gold.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_silver_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, silver_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Silver Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Silver Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_silver_scatterplot
```

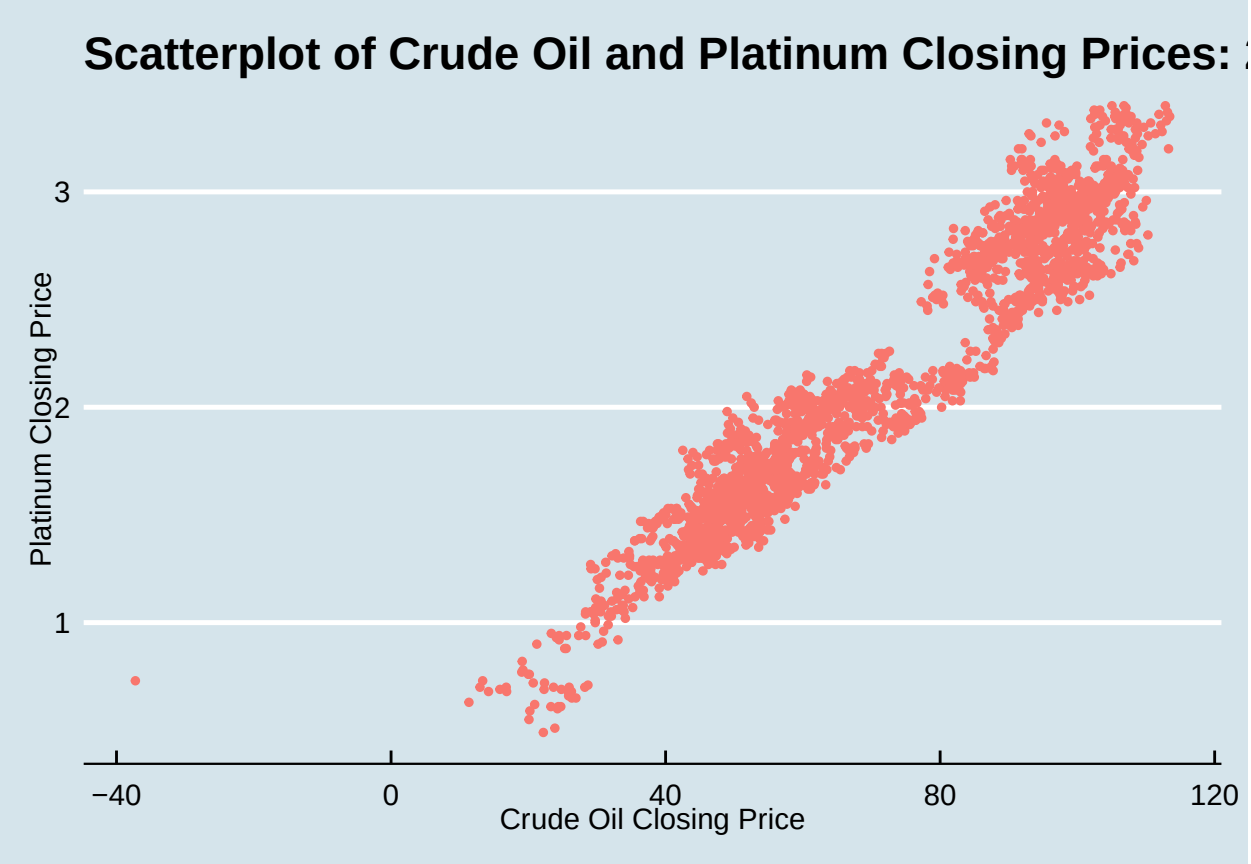



```
ggsave("fig/scatterplot_crude_oil_versus_silver.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_platinum_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, gasoline_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Platinum Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Platinum Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_platinum_scatterplot
```




```
ggsave("fig/scatterplot_crude_oil_versus_platinum.png")
```

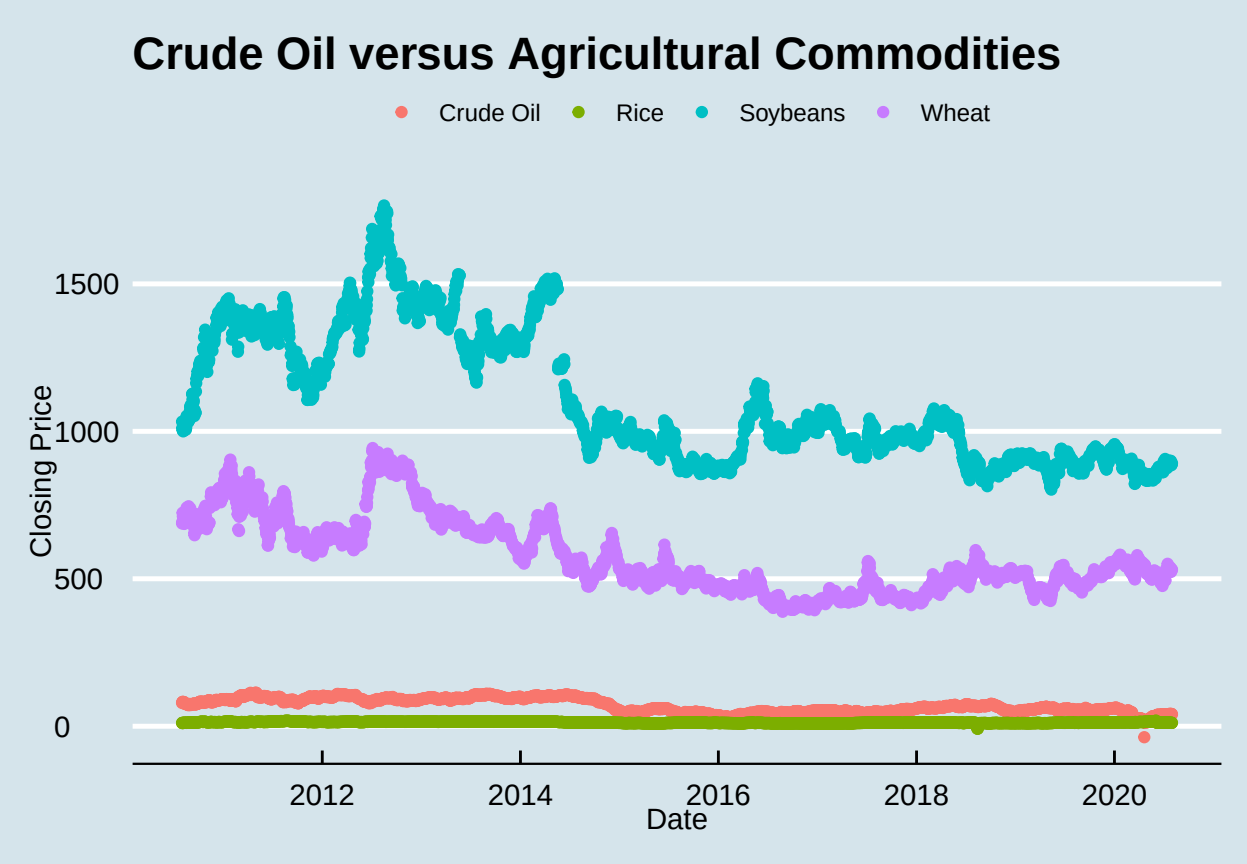
```
## Saving 6.5 x 4.5 in image
```

```
### It plots the 2010-2020 timeline of crude oil daily closing prices compared to those of the three ag
### Then, it saves the plot in a .png file entitled "crude_oil_price_history_in_comparison_to_three_com

crude_oil_vs_agriculture <- bind_rows(crude_oil, agriculture)

crude_oil_vs_agriculture_plot <- crude_oil_vs_agriculture %>%
  ggplot(aes(date, closing_price)) +
  geom_point(aes(color = commodity)) +
  xlab("Date") +
  ylab("Closing Price") +
  ggtitle("Crude Oil versus Agricultural Commodities") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_vs_agriculture_plot
```



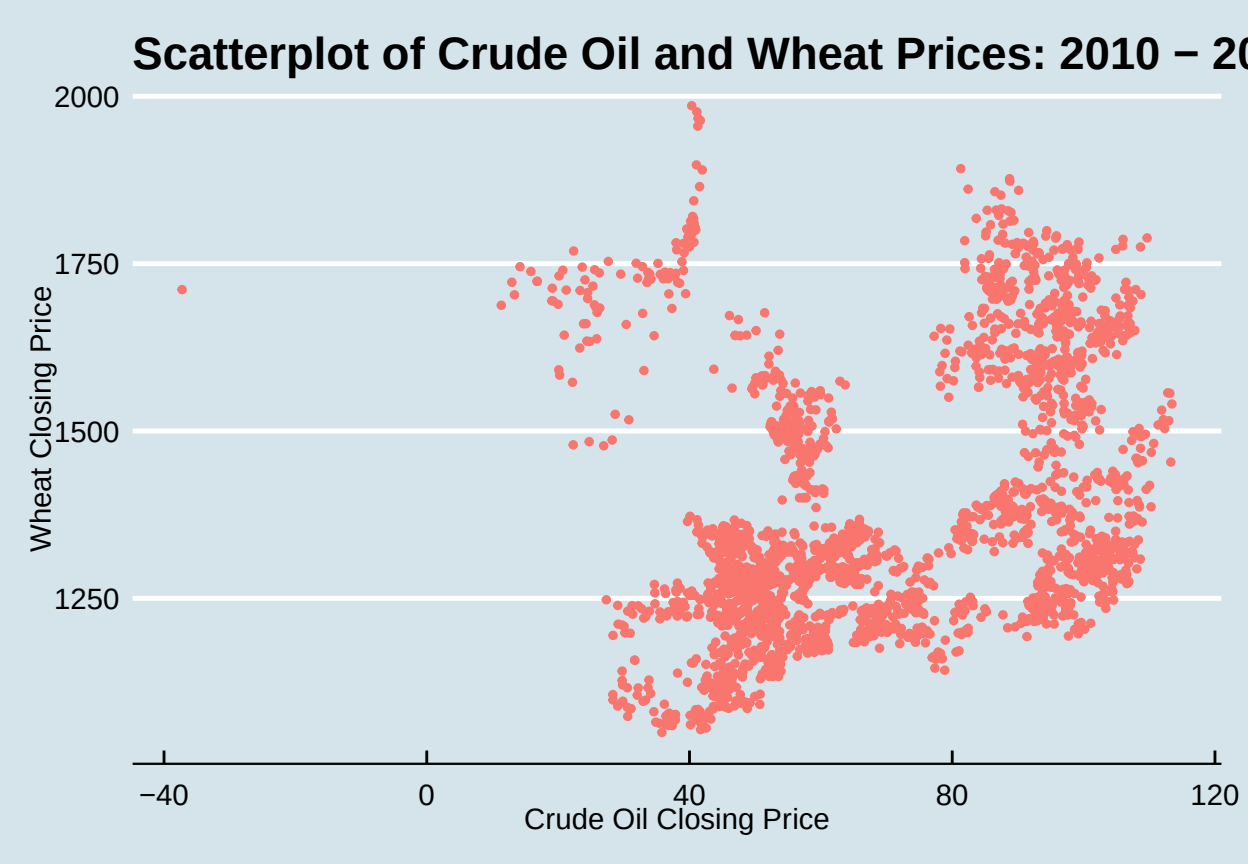

```
ggsave("fig/crude_oil_in_comparison_to_three_agricultural_sector_commodities.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#### As a final step, it generates scatterplots portraying the relationships between crude oil closing
```

```
crude_oil_wheat_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, gold_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Wheat Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Wheat Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_wheat_scatterplot
```

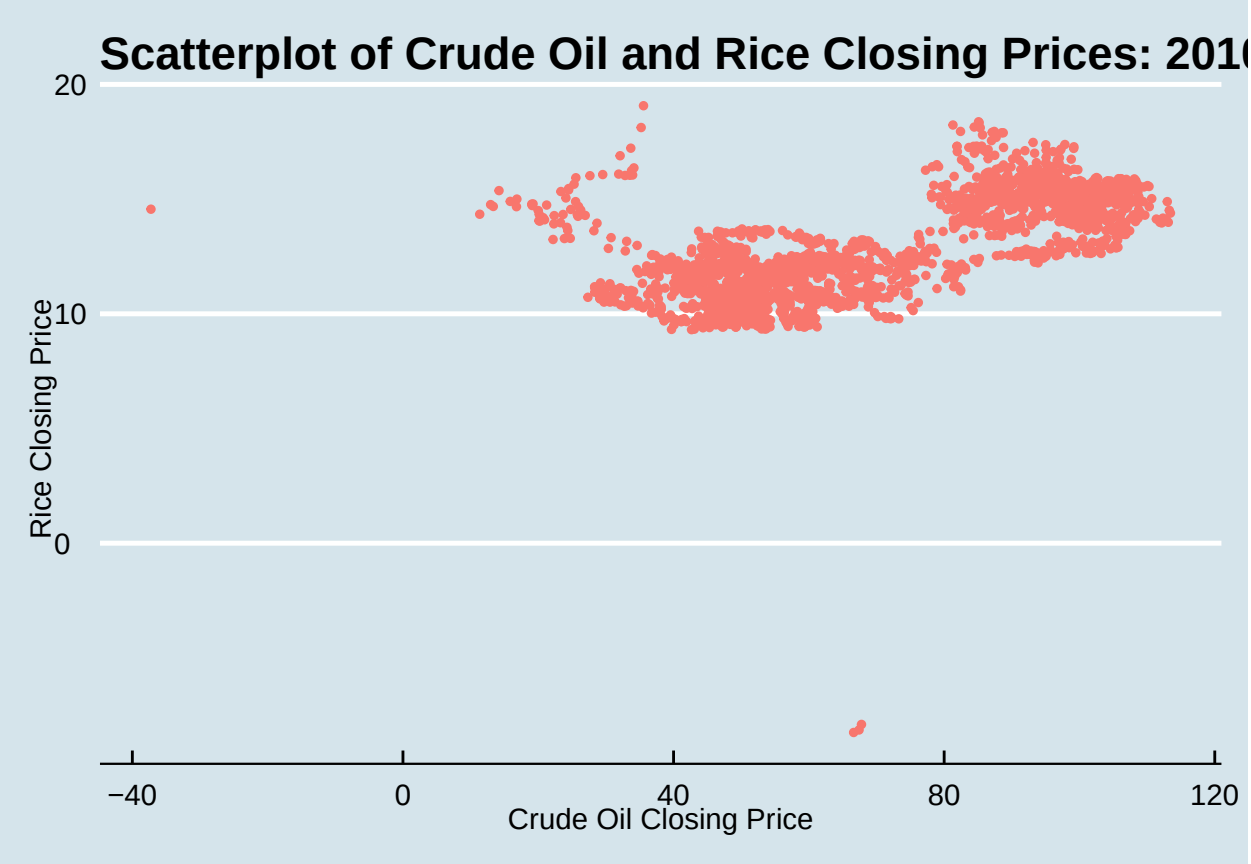



```
ggsave("fig/scatterplot_crude_oil_versus_wheat.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_rice_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, rice_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Rice Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Rice Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_rice_scatterplot
```

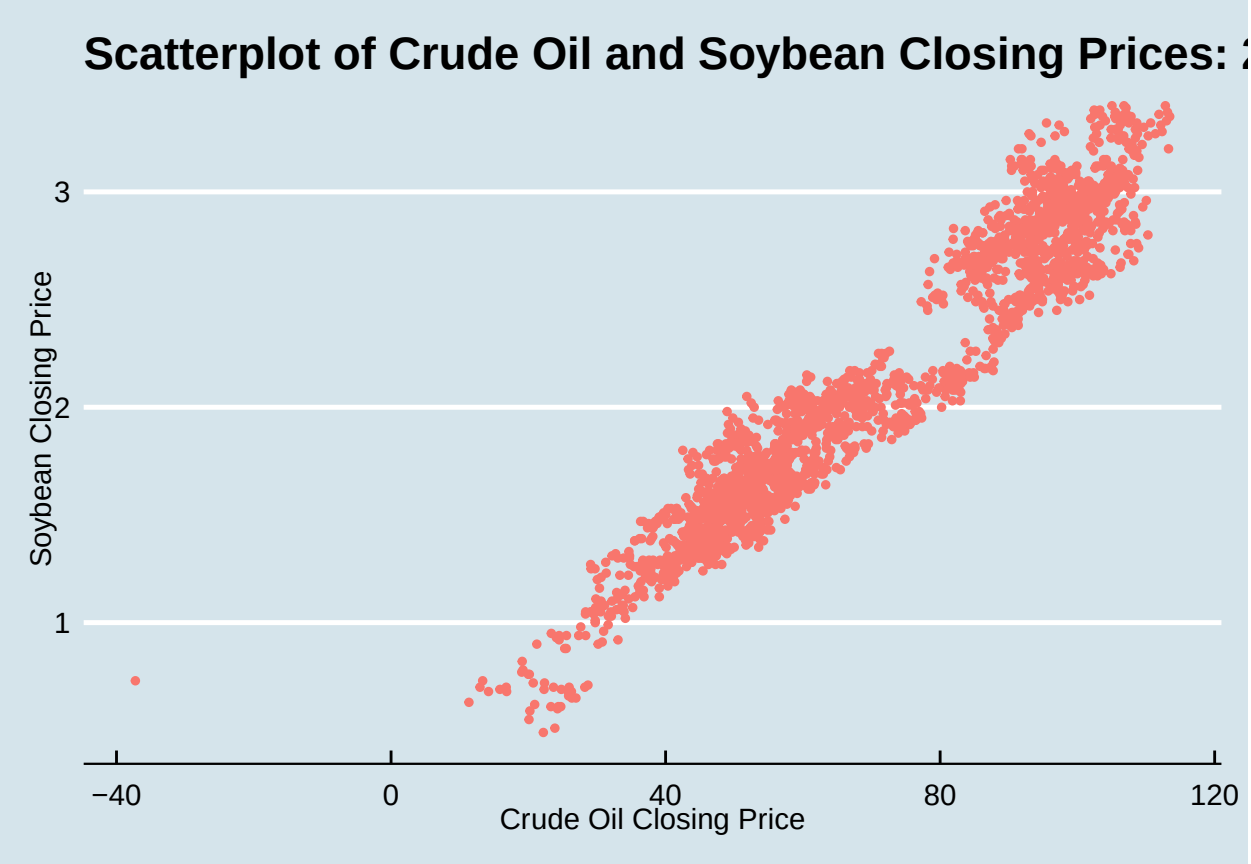



```
ggsave("fig/scatterplot_crude_oil_versus_rice.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
crude_oil_soybeans_scatterplot <- crude_oil_in_commodities_market %>%
  ggplot(aes(closing_price, gasoline_closing_price)) +
  geom_point(aes(color = commodity), size = 1, show.legend = FALSE) +
  xlab("Crude Oil Closing Price") +
  ylab("Soybean Closing Price") +
  ggtitle("Scatterplot of Crude Oil and Soybean Closing Prices: 2010 - 2020") +
  theme_economist() +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box = "horizontal" ,
        legend.text=element_text(size=8.5)) +
  guides(col = guide_legend(nrow = 1))

crude_oil_soybeans_scatterplot
```




```
ggsave("fig/scatterplot_crude_oil_versus_soybean.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: The ‘i‘ argument of ‘‘[‘()‘ can’t be a matrix as of tibble 3.0.0.
## Convert to a vector.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_warnings()‘ to see where this warning was generated.
```

```
### This section evaluates a set of fifteen potentially viable algorithms for predicting cruide oil clo

### Each of the first nine algorithms considers crude oil in isolation from other globally traded commo
### In isolation, the available predictors include both the cyclical and the linear components of time
### One cyclical component is each observation's day of the trading week (Monday, Tuesday,...Friday)
### Others are the month (1,2..12) and the quarter (1,2,..4) of each year on the 2010-2020 timeline
### Liner components are the year (e.g., 2010) and month (e.g., 8/2010) of each observation

#### ALGORITHM 01 (AVERAGE)

##### This algorithm predicts crude oil closing prices based solely on the average crude oil price for
##### It will serve as a baseline from which to compare the analytical viability of other, more sophist

mu <- mean(crude_oil_train$closing_price)

predicted_price_algorithm_01 <- mu
```

```
RMSE01 <- RMSE(predicted_price_algorithm_01, crude_oil_test$closing_price)

RMSE01
```

```
## [1] 23.51292
```