Food Inflation Study

Jessica Woods

2023-12-20

```
knitr::opts_chunk$set(echo = TRUE)

library(glmnet)
```

```
## Loading required package: Matrix

## Loaded glmnet 4.1-8
```

```
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

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

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

```
library(ggplot2)
# library(emergency_back_up_brain)
# library(coffee)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

In this R document I will be specifically just cleaning data and correlating data points that are relevant to modeling, looking at the distribution of the data and saving those back to a new CSV file to be visually assessed in Python and modeled in R.

Below is the loaded CSV files needed to start cleaning for modeling and analysis work. The link to get the food inflation information is https://microdata.worldbank.org/index.php/catalog/4483 from December 2019 to December 2023 to complete 4 years of data on internation food inflation. "Monthly food price estimates by product and market 25 countries, 1353 markets, 2007/01/01-2023/12/01, version 2023/12/11" -The World Bank of Microdata website. There are two other factors in the analysis work, that is conflict (gun violence and war) and the US stock Market prices from December 2019 to December 2023.

The current HO: "The occurrence of conflict in key food-producing regions, coupled with fluctuations in the US stock market, significantly influences international food inflation rates. Higher instances of conflict and volatility in the US stock market are expected to correlate with increased food inflation on a global scale." The current Alternative HO : "The impact of conflict and US stock market fluctuations on international food inflation rates may not exhibit a significant correlation."

```
# Food inflation data
food_data <- read.csv("Food_inflation_2019_2023.csv", header=TRUE, sep=",")
# time to clean this up
clean_food_data <- subset(food_data, select = -X)
clean_food_data <-na.omit(clean_food_data)
clean_food_data <- subset(clean_food_data, select = -Market)
clean_food_data <-subset(clean_food_data, select = -Currency)
#fix the date
clean_food_data$Date <-as.Date(clean_food_data$Date, format = "%Y-%m-%d")
```

Narrowing down the top most conflicted countries in the past 4 years according to Wikipedia https://en.wikipedia.org/wiki/List_of_ongoing_armed_conflicts sited. Mexico, Ukraine, Afghanistan, Syria (Syrian Arab Republic), Ethiopia, Yemen Out of those listed I found 3 in the data set to work with.

```
conflicted_countries_food_data <- clean_food_data %>%
  filter(Country %in% c("Afghanistan", "Syrian Arab Republic", "Yemen, Rep."))

conflicted_countries_food_data <- conflicted_countries_food_data %>%
  filter(Open != 0 & Close != 0 & High != 0)
confliced_countries_food_data <- na.omit(conflicted_countries_food_data[c("Open", "Close")])

#Isolate Food Price Index
conflicted_countries_food_price_index<-conflicted_countries_food_data[conflicted_countries_food_data$Pr
conflicted_countries_food_data <- conflicted_countries_food_data %>%
  filter(Product != "food_price_index")
str(conflicted_countries_food_price_index)
```

```
## 'data.frame':    11123 obs. of  8 variables:
##  $ Country: chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ Region : chr  "Badakhshan" "Badakhshan" "Badakhshan" "Badakhshan" ...
##  $ Product: chr  "food_price_index" "food_price_index" "food_price_index" "food_price_index" ...
##  $ Date   : Date, format: "2019-12-01" "2019-12-01" ...
##  $ Open   : num  1.12 1.12 1.13 1.13 1.13 1.13 1.13 1.13 1.23 1.23 ...
##  $ High   : num  1.13 1.13 1.14 1.14 1.14 1.14 1.2 1.2 1.25 1.25 ...
##  $ Low    : num  1.11 1.11 1.12 1.12 1.12 1.12 1.12 1.12 1.21 1.21 ...
##  $ Close  : num  1.13 1.13 1.13 1.13 1.13 1.13 1.2 1.2 1.25 1.25 ...
```

```
#  factor  products, regions, countries
conflicted_countries_food_data$Product <- as.factor(conflicted_countries_food_data$Product)
conflicted_countries_food_data$Country <- as.factor(conflicted_countries_food_data$Country)
```

```r
# Convert 'Region' to a factor within each country
conflicted_countries_food_data$Region <- as.factor(conflicted_countries_food_data$Region)
conflicted_countries_food_data$Region <- factor(
  conflicted_countries_food_data$Region,
  levels = unique(conflicted_countries_food_data$Region)
)

str(conflicted_countries_food_data)
```

```
## 'data.frame':    154938 obs. of  8 variables:
##  $ Country: Factor w/ 3 levels "Afghanistan",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Region : Factor w/ 71 levels "Badakhshan","Badghis",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Product: Factor w/ 30 levels "apples","bananas",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Date   : Date, format: "2019-12-01" "2019-12-01" ...
##  $ Open   : num  49.8 49.8 49.8 49.8 49.9 ...
##  $ High   : num  49.9 49.9 49.9 49.9 50 ...
##  $ Low    : num  49.7 49.7 49.7 49.7 49.8 ...
##  $ Close  : num  49.8 49.8 49.9 49.9 49.9 ...
```

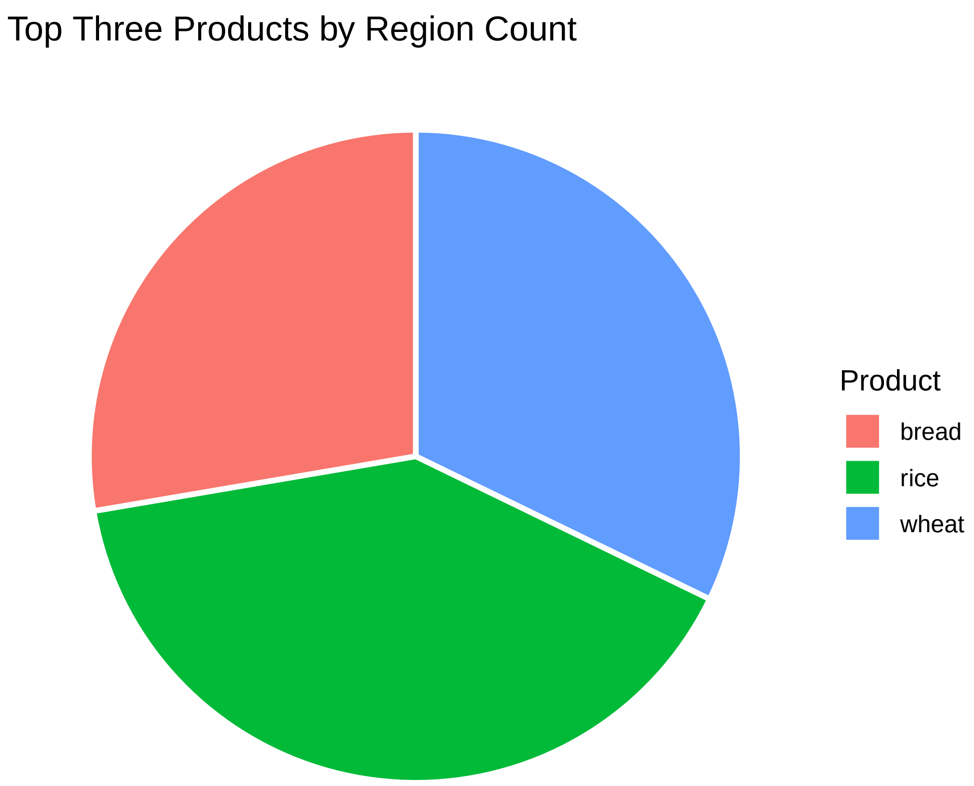
```r
# I want to know which regions have the highest fluctuation in their markets
conflicted_countries_food_data$Fluctuation <- conflicted_countries_food_data$Close - conflicted_countri
# Find observations with the highest fluctuations
top_fluctuations <- conflicted_countries_food_data[order(conflicted_countries_food_data$Fluctuation, de
head(top_fluctuations)
```

```
##                    Country      Region                          Product
## 69874 Syrian Arab Republic Deir-ez-Zor livestock_sheep_two_year_old_male
## 98045 Syrian Arab Republic       Idleb livestock_sheep_two_year_old_male
## 74774 Syrian Arab Republic Deir-ez-Zor livestock_sheep_two_year_old_male
## 73549 Syrian Arab Republic Deir-ez-Zor livestock_sheep_two_year_old_male
## 74770 Syrian Arab Republic Deir-ez-Zor livestock_sheep_two_year_old_male
## 73545 Syrian Arab Republic Deir-ez-Zor livestock_sheep_two_year_old_male
##             Date    Open    High     Low   Close Fluctuation
## 69874 2023-12-01 2449085 3261065 2060071 3261065    811980.1
## 98045 2023-08-01 2843081 3522899 2699200 3522899    679817.8
## 74774 2023-12-01 2591094 3243195 2139313 3243195    652101.2
## 73549 2023-12-01 2613006 3256648 2141895 3256648    643642.5
## 74770 2023-08-01 2217816 2779033 1977785 2779033    561217.1
## 73545 2023-08-01 2266728 2820968 2082072 2820968    554239.3
```

```r
product_region_count <- conflicted_countries_food_data %>%
  group_by(Product) %>%
  summarise(Region_Count = n_distinct(Region)) %>%
  arrange(desc(Region_Count))

# Product with the most regions
product_with_most_regions <- product_region_count[which.max(product_region_count$Region_Count), ]
top_three_products <- product_region_count %>%
  slice_head(n = 3)
# Extract the top three products
top_three_products_list <- top_three_products$Product
```

```r
# Filter the original data to get regions and countries for the top three products
regions_countries_top_three <- conflicted_countries_food_data %>%
  filter(Product %in% top_three_products_list) %>%
  distinct(Product, Region, Country)



# Create a pie chart for the top three regions
ggplot(top_three_products, aes(x = "", y = Region_Count, fill = Product)) +
  geom_bar(stat = "identity", width = 1, color = "white", size = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Top Three Products by Region Count", fill = "Product", x = NULL, y = NULL) +
  theme_void() +
  theme(legend.position = "right") +
  scale_fill_discrete(name = "Product")
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

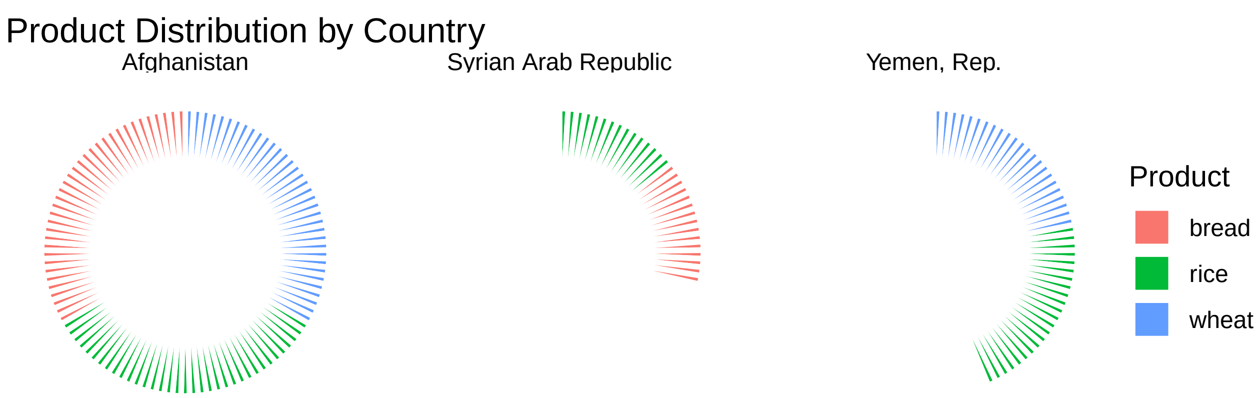



```r
top_three_products
```

```
## # A tibble: 3 x 2
```

```
##   Product Region_Count
##   <fct>          <int>
## 1 rice              71
## 2 wheat             57
## 3 bread             49
```

```r
unique_countries <- unique(regions_countries_top_three$Country)

product_counts <- regions_countries_top_three %>%
  group_by(Country, Product, Region) %>%
  summarise(Count = n(), .groups = "drop")

# Create a pie chart for each country showing product distribution within regions
ggplot(product_counts, aes(x = "", y = Count, fill = Product)) +
  geom_bar(stat = "identity", width = 1, color = "white", size = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Product Distribution by Country", fill = "Product", x = NULL, y = NULL) +
  theme_void() +
  theme(legend.position = "right") +
  facet_wrap(~ Country)
```

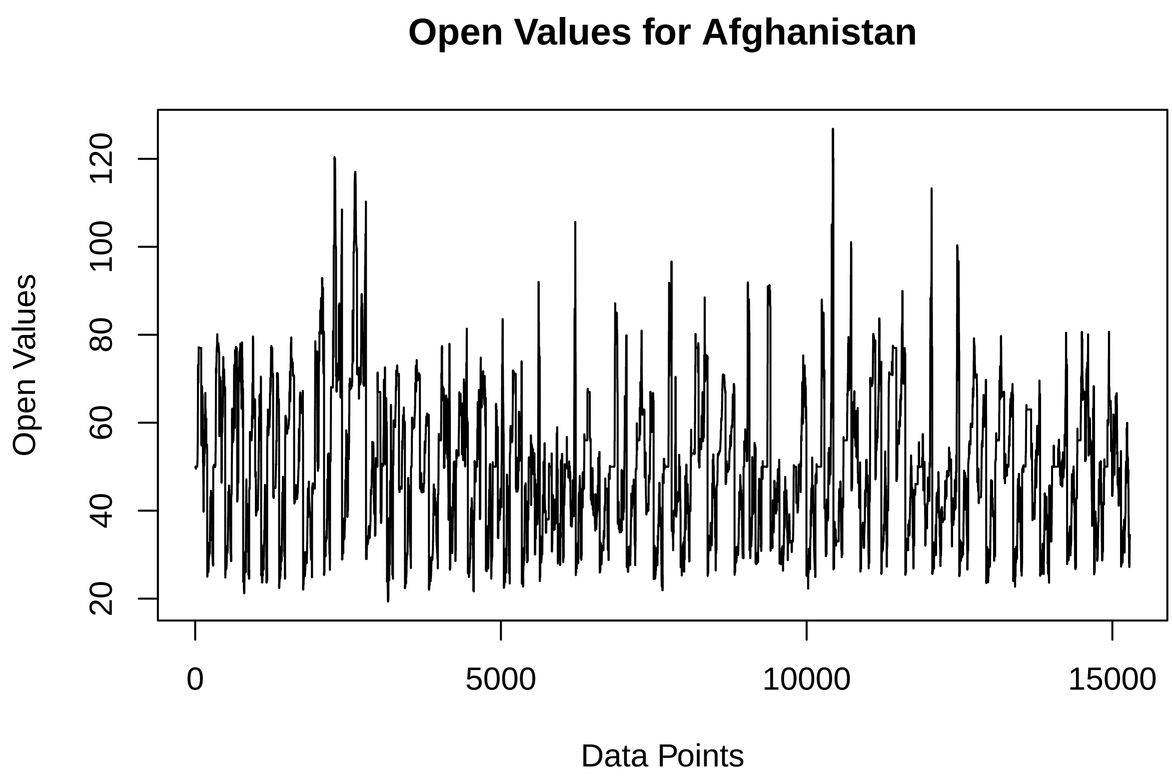



```r
num_unique_regions <- conflicted_countries_food_data %>%
  summarise(Num_Regions = n_distinct(Region))

num_unique_regions
```

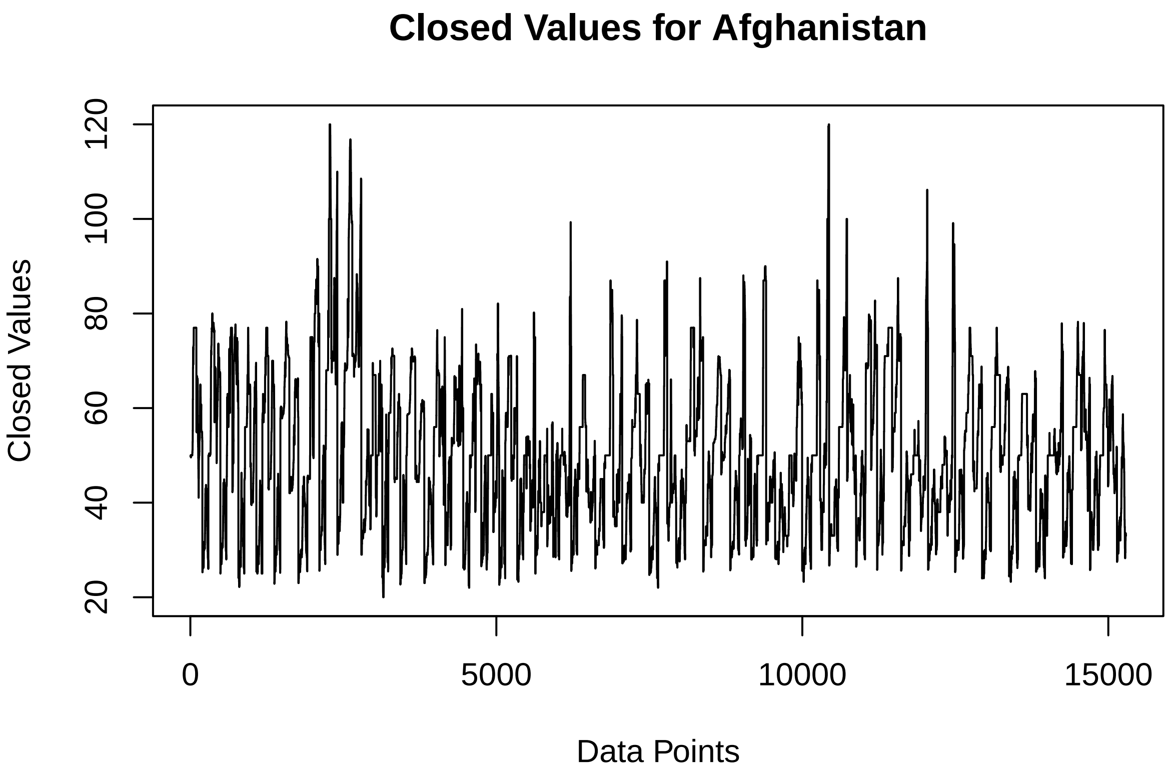
```
##   Num_Regions
## 1          71
```

Afghanistan data on Food Inflation open and close, the difference between the two. Its important to look at the distribution of the data to do further analysis and cleaning of the data. If the differences between open and close prices themselves follow a normal distribution, it could suggest a certain level of regularity and randomness in price movements.

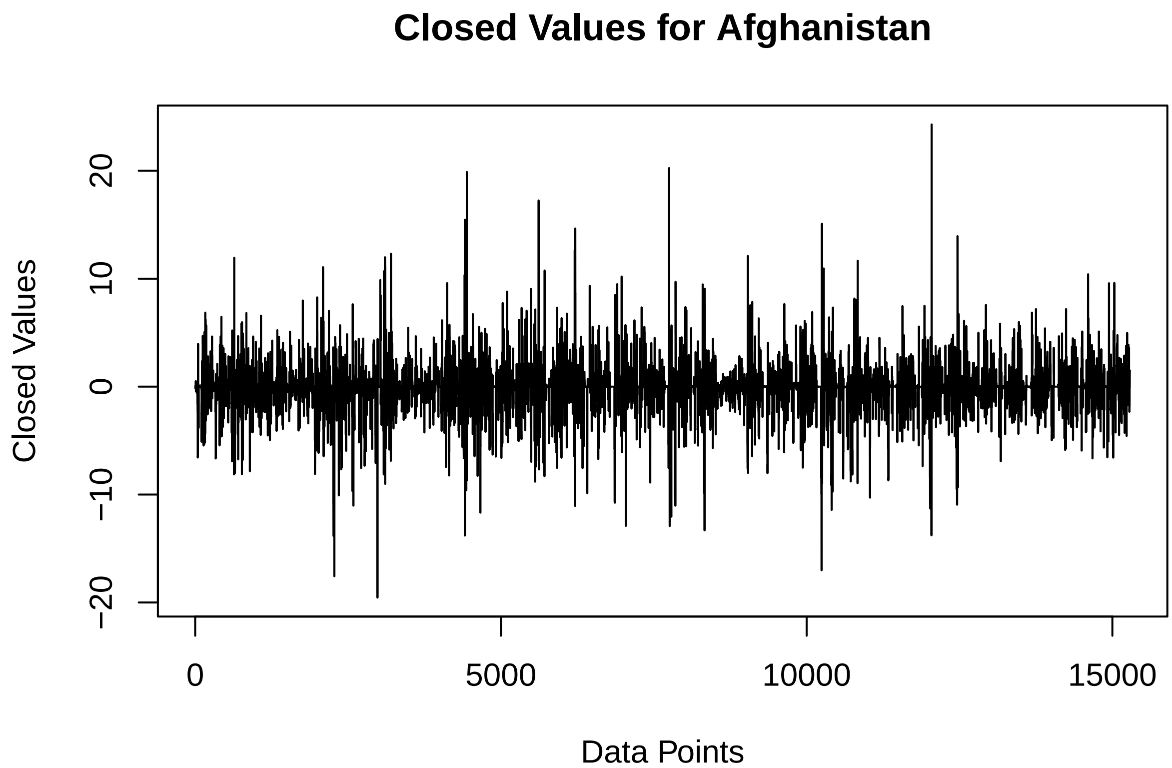
```
data_afghanistan <- conflicted_countries_food_data[conflicted_countries_food_data$Country == "Afghanista

open_values_afghanistan <- data_afghanistan$Open
closed_values_afghanistan <- data_afghanistan$Close
difference_afghanistan <- open_values_afghanistan - closed_values_afghanistan


# Create a sequence for x-axis (assuming you want a sequence of numbers as x-axis)
x <- seq(length(open_values_afghanistan))
x2 <-seq(length(closed_values_afghanistan))
x3 <- seq(length(difference_afghanistan))
# Plotting Open values for Afghanistan
plot(x, open_values_afghanistan, type = "l", xlab = "Data Points", ylab = "Open Values", main = "Open Va
```



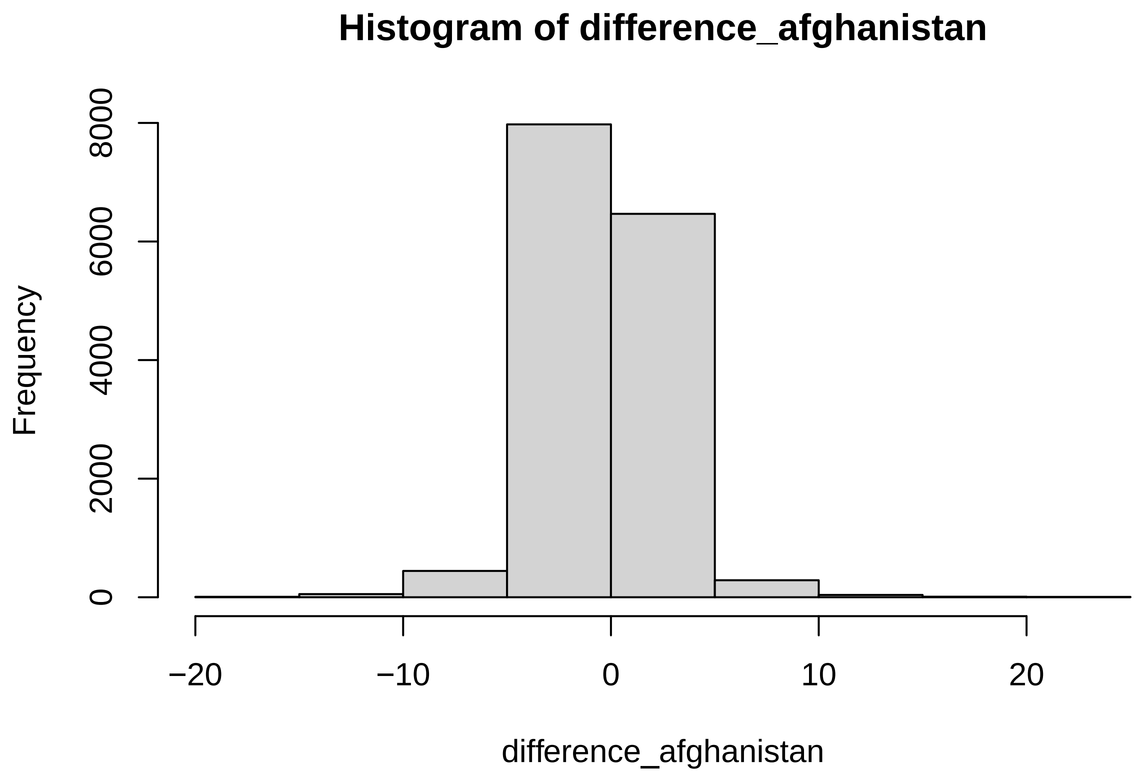

```
plot(x2, closed_values_afghanistan, type = "l", xlab = "Data Points", ylab = "Closed Values", main = "Cl
```



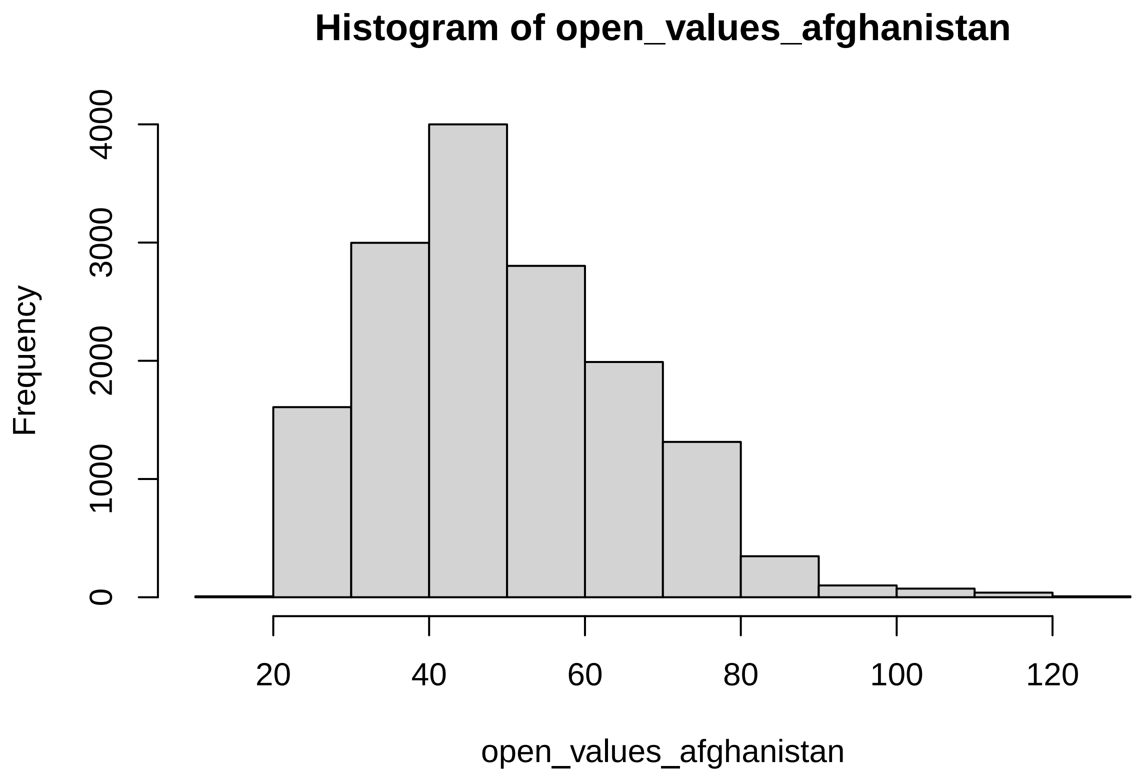

```
plot(x3, difference_afghanistan, type = "l", xlab = "Data Points", ylab = "Closed Values", main = "Clos
```



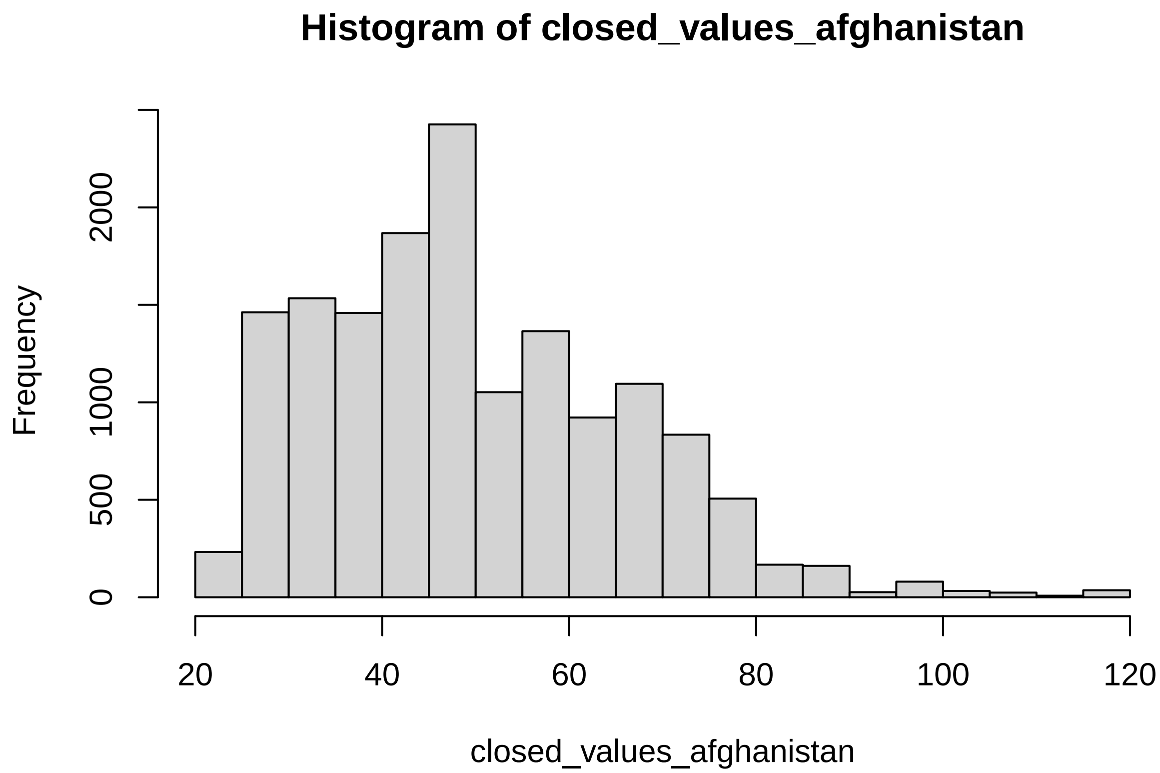

```
#seems to be normally distributed, this helps narrow down appropriate models to examine Afghanistan spe
hist(difference_afghanistan)
```




```
# open and close are distributed to the right
hist(open_values_afghanistan)
```

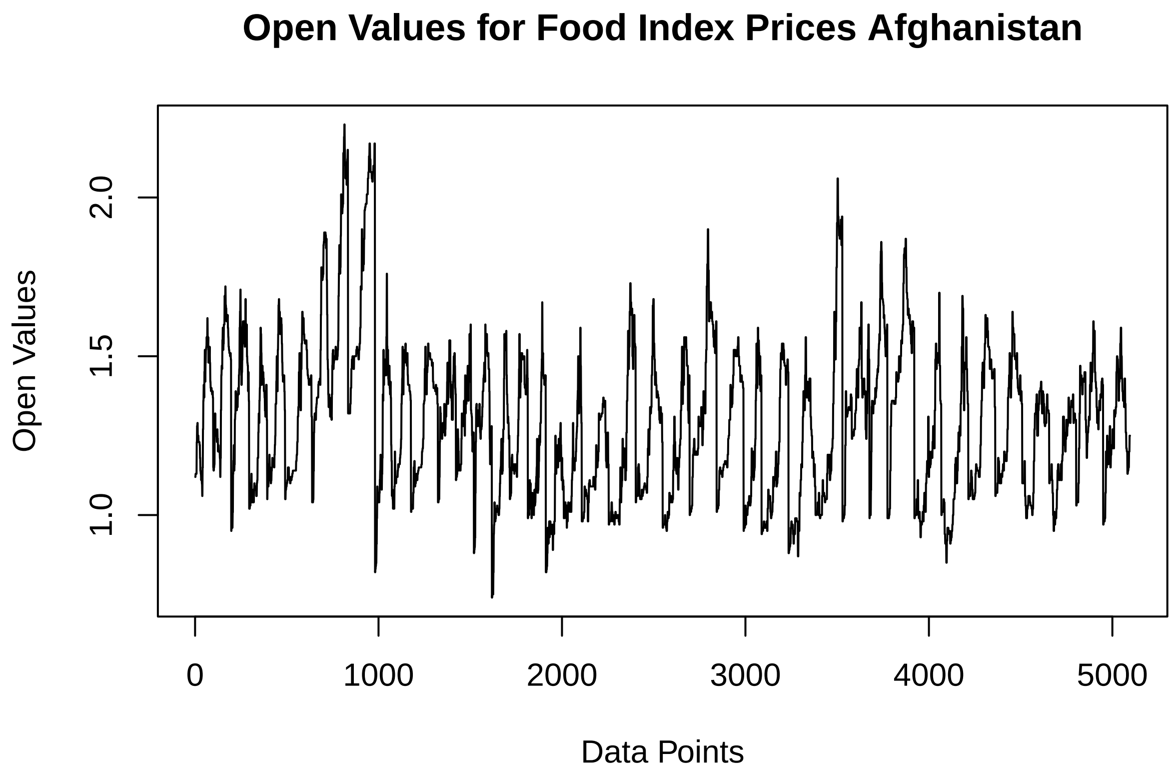



```
hist(closed_values_afghanistan)
```

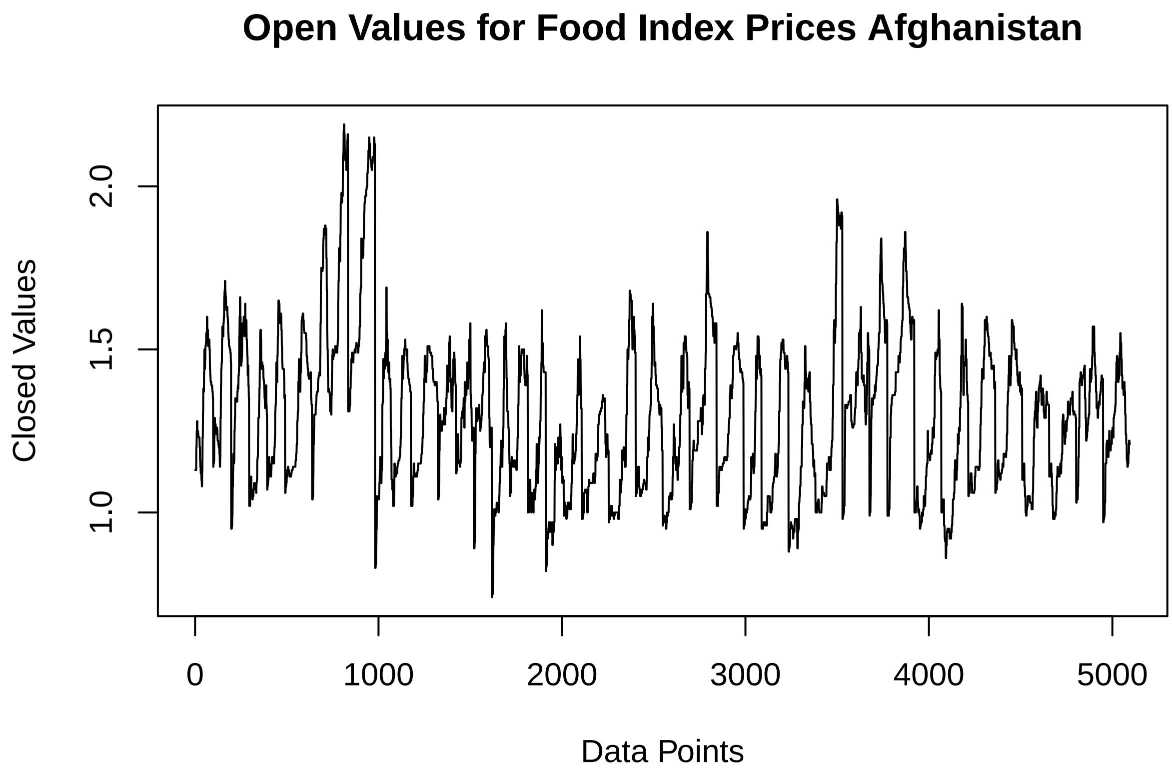


In financial markets, a normal distribution of price index differences might support the idea of market efficiency, suggesting that market prices reflect all available information, and arbitrage opportunities might be limited. Furthermore, the questioning factor still remains that the open and close histograms are right skewed, which indicate otherwise. Its considered a bullish market indication.

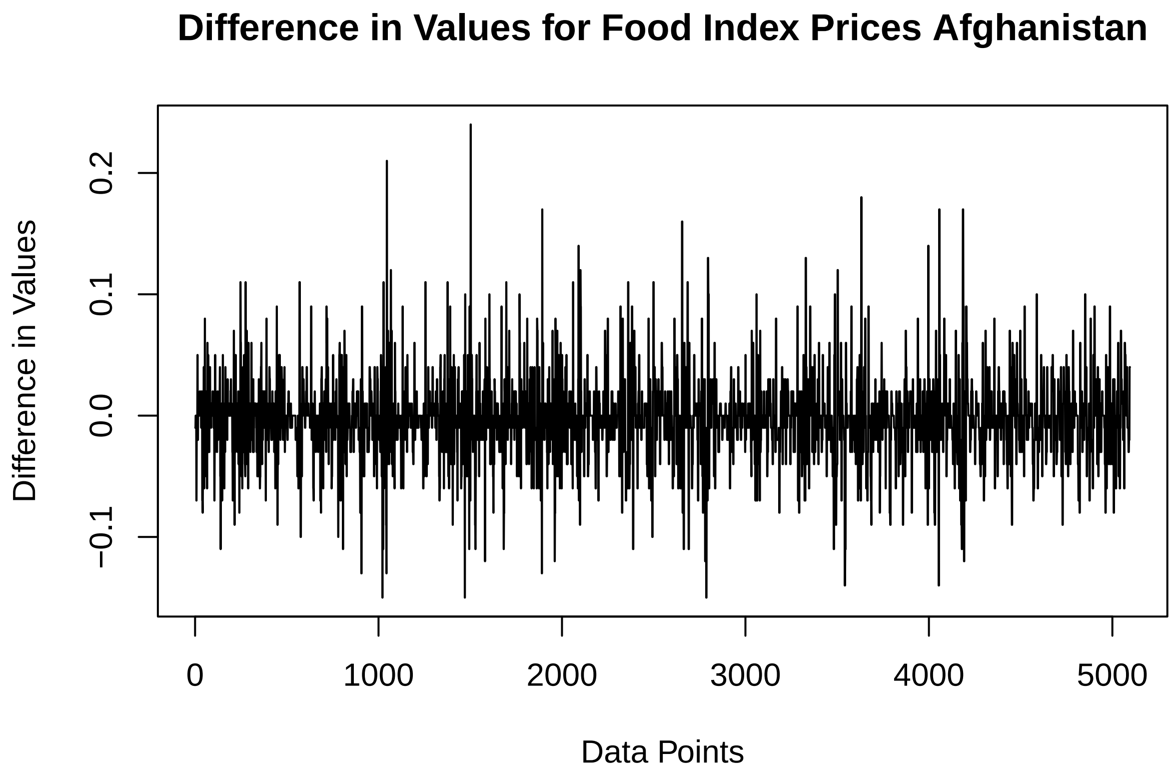
```
# check to values of the Food Index Prices and see if they reflect a normal distribution or not
# Food Price Index
index_afghanistan_data <- conflicted_countries_food_price_index[conflicted_countries_food_price_index$Co
index_afghanistan_open <- index_afghanistan_data$Open
index_afghanistan_close <- index_afghanistan_data$Close
index_afghanistan_difference <- index_afghanistan_open - index_afghanistan_close

x1_index <- seq(length(index_afghanistan_open))
x2_index <- seq(length(index_afghanistan_close))
x3_index <- seq(length(index_afghanistan_difference))

plot(x1_index, index_afghanistan_open, type = "l", xlab = "Data Points", ylab = "Open Values", main = "(
```



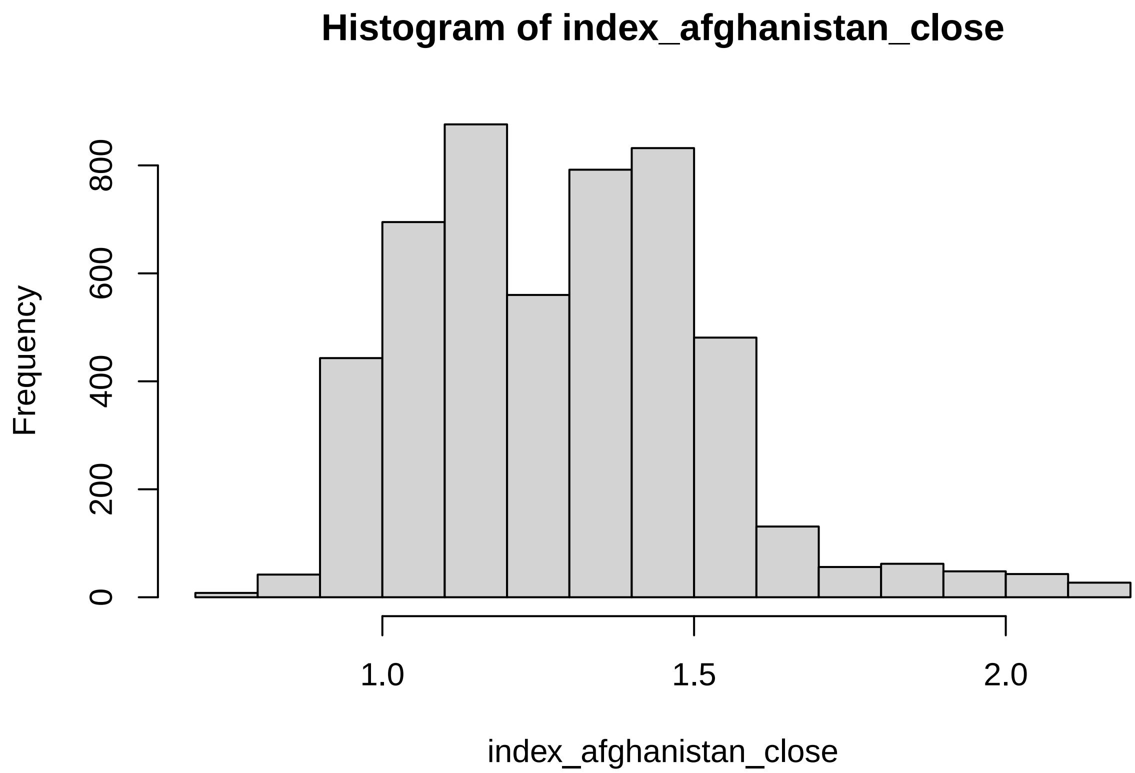

```
plot(x2_index, index_afghanistan_close, type = "l", xlab = "Data Points", ylab = "Closed Values", main =
```



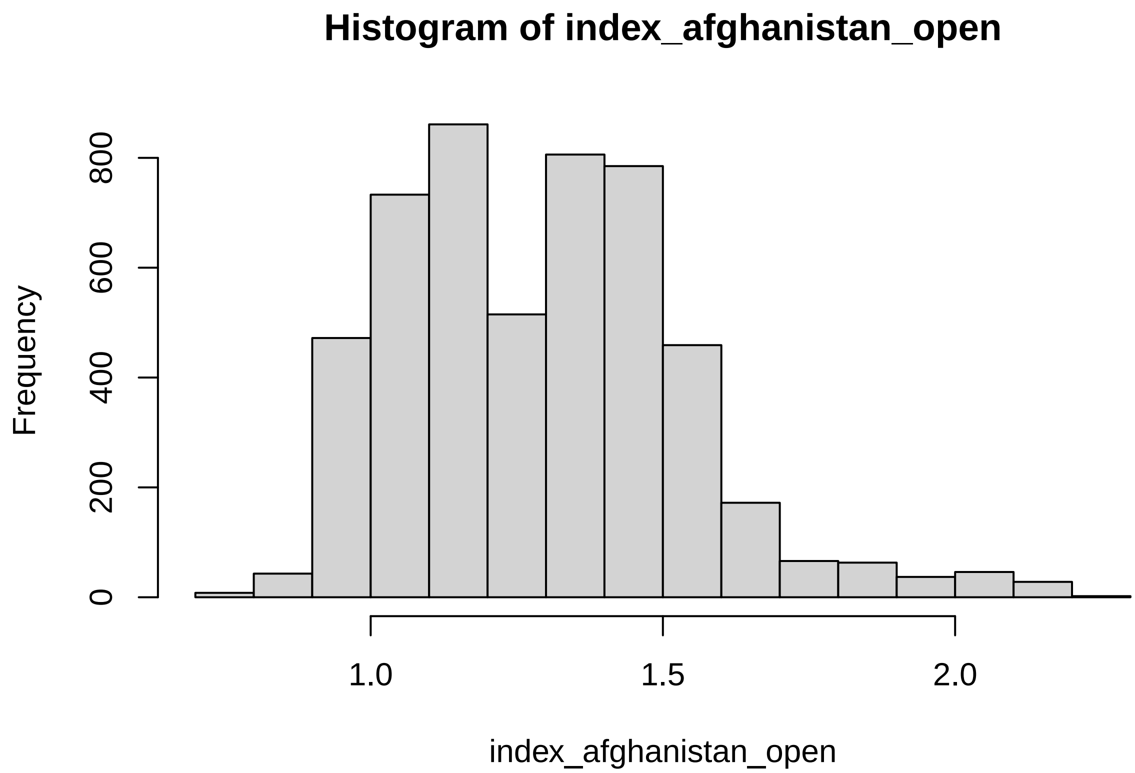

```
plot(x3_index, index_afghanistan_difference, type = "l", xlab = "Data Points", ylab = "Difference in Va
```



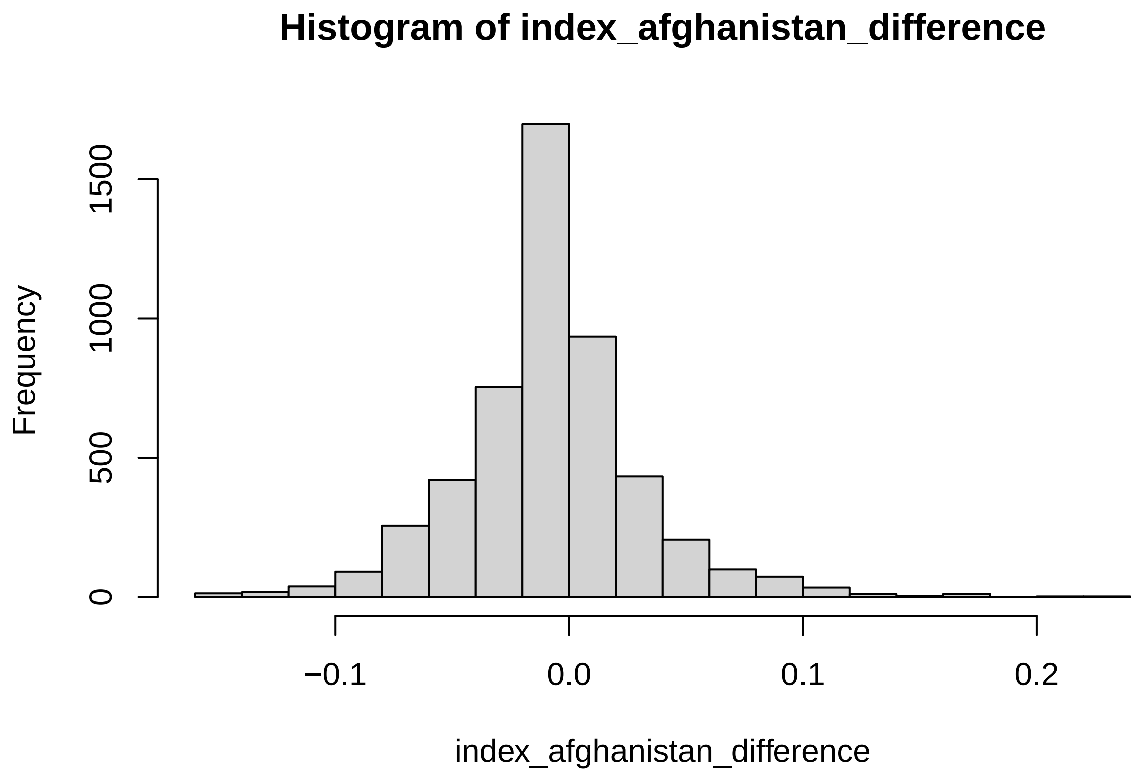

```
# these are sckewed to the right to the higher values
hist(index_afghanistan_close)
```




```
hist(index_afghanistan_open)
```




```
# this seems to hold a normal distribution which is probably best to work with-
hist(index_afghanistan_difference)
```



Below is the conflict data which is from https://acleddata.com/data-export-tool/. This requires a access key to get this data, it seems to be the most up to date data on conflicts. That coincides with the Dates from the Food inflation.

```
# Conflict data
```

Below is the the US Stock Market Data which is from https://www.ers.usda.gov/data-products/wheat-data/ and Yahoo Stocks for rought rice https://finance.yahoo.com/quote/ZR%3DF/history?period1=1575158400&period2=1703030400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true.

```
# US Stock Market Data
```