

# How to Access Commodity Data from UN Comtrade in R?

Prithviraj Lakkakula

## Objective

In this repository, I will demonstrate how to gather commodity (Maize/Corn) data from a popular trade data source (UN Comtrade, <https://comtrade.un.org/>) in R using an excellent R package, `comtradr`. This R package essentially interfaces with the United Nations Comtrade's API to collect the data.

See the citation below.

To cite package 'comtradr' in publications use:

```
Muir C (2022). _comtradr: Interface with the United Nations Comtrade
API_. R package version 0.3.0,
<https://CRAN.R-project.org/package=comtradr>.
```

A BibTeX entry for LaTeX users is

```
@Manual{,
  title = {comtradr: Interface with the United Nations Comtrade API},
  author = {Chris Muir},
  year = {2022},
  note = {R package version 0.3.0},
  url = {https://CRAN.R-project.org/package=comtradr},
}
```

## Basics

### Limit

First of all, there is limit for accessing the number of calls to the API per hour. If you have a token, then you have some room to increase the number of calls. For more details, you can refer to <https://comtrade.un.org/data/doc/api/#Authentication>

- Without user token: 100 requests per hour.
- With valid user token: 10,000 requests per hour.
- There is a limit set for exporter and importer combination for a maximum of 5.

### Calling UN Comtrade API

If both `start_date` and `end_date` takes input as the same year and `freq` takes `monthly` then it will return all monthly trade data for that specific year.

```
ct_search(reporters = "exporting_country_goes_here",
          partners = c("importing_countries_goes_here"),
          trade_direction = "imports_or_exports_goes_here",
          commod_codes = "enter_your_commodity_code_here",
          start_date = 2012,
          end_date = 2012,
          freq = "monthly")
```

### Country look up

Sometimes the countries were labelled different in the context of exporting and importing scenarios. Since, the countries list is huge, it is good to know how specific countries are labelled. This is easy to look at with `ct_country_lookup()`

```
ct_country_lookup("eu", "reporter")
```

```
[1] "EU-28"
```

```
ct_country_lookup("eu", "partner")
```

```
[1] "Africa CAMEU region, nes" "Eastern Europe, nes"
[3] "Europe EFTA, nes"        "Europe EU, nes"
[5] "Neutral Zone"            "Other Europe, nes"
```

If you look at the outputs of each of the code chunks above, it is clear that the European Union is not labelled same from the point of exporter and importer perspective. Therefore, one needs to make a decision how to approach in the data collection process.

## Commodity lookup

If you want a details of a specific commodity, you can use `ct_commodity_lookup()` and set `verbose = TRUE` as an argument. As the output shows, you will get details about Harmonized System (HS) code. Harmonized System is an international nomenclature for the classification of products.

```
ct_commodity_lookup("maize", verbose = TRUE)
```

```
$maize
```

```
[1] "1005 - Maize (corn)"
[2] "100510 - Cereals; maize (corn), seed"
[3] "100590 - Cereals; maize (corn), other than seed"
[4] "110220 - Cereal flour; of maize (corn)"
[5] "110290 - Cereal flours; other than wheat, meslin, and maize (corn)"
[6] "110313 - Cereal groats and meal; of maize (corn)"
[7] "110423 - Cereal grains; worked (egg hulled, pearled, sliced or kibbled) of maize (corn)"
[8] "110429 - Cereal grains; worked, other than rolled or flaked (egg. Hulled, pearled, sliced)"
[9] "110812 - Starch; maize (corn) starch"
[10] "151521 - Vegetable oils; maize (corn) oil and its fractions, crude, not chemically modified"
[11] "151529 - Vegetable oils; maize (corn) oil and its fractions, other than crude, whether or not refined"
[12] "1904 - Prepared foods obtained by swelling, roasting of cereals or cereal products (except of cereals or cereal products)"
[13] "190430 - Prepared foods obtained by the swelling or roasting of cereals or cereal products"
[14] "190490 - Food preparations; cereal or cereal products (excluding maize), in grain form"
[15] "230210 - Bran, sharps and other residues; of maize (corn), whether or not in the form of pellets"
[16] "230670 - Oil cake and other solid residues; whether or not ground or in the form of pellets"
```

## A Demo

In this demo,

- I will collect export data of Maize (also called Corn) with HS code 100590 from the United States (as an exporter) to various importers, including China, Japan, South Korea, Indonesia, Vietnam, North African region, the Middle East region, and the European Union region between 2012 and 2022. The frequency of the data is monthly.
- transform and arrange these data
- Export these data to Excel

## Load Required Libraries

```
library(comtradr)
library(tidyverse)
library(janitor)
library(openxlsx)
```

## Write Functions

- Write Functions
- Since, there is a limit for 5 importers/partners at a time, I created a separate data frames and later combined all of them by rows for convenience.

```
# FUNCTIONS
get_comtrade_data <- function(reporter, year){
  df <- ct_search(reporters = reporter,
                  partners = c("China", "Indonesia", "Japan",
                              "Viet Nam", "Rep. of Korea"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%
  select(period, period_desc, trade_flow,
         reporter, partner, netweight_kg,
         trade_value_usd) %>%
  arrange(period)
  return(df)
}

get_comtrade_nafrica <- function(reporter, year){
  df1 <- ct_search(reporters = reporter,
                  partners = c("Algeria", "Libya", "Morocco",
```

```

        "Tunisia", "Mauritania"),
        trade_direction = "exports",
        commod_codes = "100590",
        start_date = year,
        end_date = year,
        freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
  arrange(period)
df2 <- ct_search(reporters = reporter,
                 partners = c("Egypt", "Sudan"),
                 trade_direction = "exports",
                 commod_codes = "100590",
                 start_date = year,
                 end_date = year,
                 freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
  arrange(period)
df <- rbind(
  df1,
  df2
)
return(df)
}

get_comtrade_meast <- function(reporter, year){
  df1 <- ct_search(reporters = reporter,
                   partners = c("Iraq", "Syria", "Israel",
                                "Jordan", "Kuwait"),
                   trade_direction = "exports",
                   commod_codes = "100590",
                   start_date = year,
                   end_date = year,
                   freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
  arrange(period)

```

```

df2 <- ct_search(reporters = reporter,
                  partners = c("Lebanon", "Oman", "Qatar",
                              "Saudi Arabia",
                              "United Arab Emirates"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%
  select(period, period_desc, trade_flow,
         reporter, partner, netweight_kg,
         trade_value_usd) %>%
  arrange(period)
df3 <- ct_search(reporters = reporter,
                  partners = c("Yemen", "Bahrain"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%
  select(period, period_desc, trade_flow,
         reporter, partner, netweight_kg,
         trade_value_usd) %>%
  arrange(period)

df <- rbind(
  df1,
  df2,
  df3
)

return(df)
}

get_comtrade_eu <- function(reporter, year){
  df1 <- ct_search(reporters = reporter,
                  partners = c("Austria", "Belgium", "Bulgaria",
                              "Croatia", "Cyprus"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,

```

```

        end_date = year,
        freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
arrange(period)

df2 <- ct_search(reporters = reporter,
                  partners = c("Czechia", "Denmark", "Estonia",
                              "Finland", "France"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
arrange(period)

df3 <- ct_search(reporters = reporter,
                  partners = c("Germany", "Greece", "Hungary",
                              "Ireland", "Italy"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%
select(period, period_desc, trade_flow,
        reporter, partner, netweight_kg,
        trade_value_usd) %>%
arrange(period)

df4 <- ct_search(reporters = reporter,
                  partners = c("Latvia", "Lithuania", "Luxembourg",
                              "Malta", "Netherlands"),
                  trade_direction = "exports",
                  commod_codes = "100590",
                  start_date = year,
                  end_date = year,
                  freq = "monthly") %>%

```

```

select(period, period_desc, trade_flow,
       reporter, partner, netweight_kg,
       trade_value_usd) %>%
arrange(period)

df5 <- ct_search(reporters = reporter,
                 partners = c("Poland", "Portugal", "Romania",
                              "Slovakia", "Slovenia"),
                 trade_direction = "exports",
                 commod_codes = "100590",
                 start_date = year,
                 end_date = year,
                 freq = "monthly") %>%
select(period, period_desc, trade_flow,
       reporter, partner, netweight_kg,
       trade_value_usd) %>%
arrange(period)

df6 <- ct_search(reporters = reporter,
                 partners = c("Spain", "Sweden"),
                 trade_direction = "exports",
                 commod_codes = "100590",
                 start_date = year,
                 end_date = year,
                 freq = "monthly") %>%
select(period, period_desc, trade_flow,
       reporter, partner, netweight_kg,
       trade_value_usd) %>%
arrange(period)

df <- rbind(
  df1,
  df2,
  df3,
  df4,
  df5,
  df6
)

return(df)

```



```
}
```

## Data Collection

```
usa_df_2012 <- get_comtrade_data("USA", "2012")
usa_df_2013 <- get_comtrade_data("USA", "2013")
usa_df_2014 <- get_comtrade_data("USA", "2014")
usa_df_2015 <- get_comtrade_data("USA", "2015")
usa_df_2016 <- get_comtrade_data("USA", "2016")
usa_df_2017 <- get_comtrade_data("USA", "2017")
usa_df_2018 <- get_comtrade_data("USA", "2018")
usa_df_2019 <- get_comtrade_data("USA", "2019")
usa_df_2020 <- get_comtrade_data("USA", "2020")
usa_df_2021 <- get_comtrade_data("USA", "2021")
usa_df_2022 <- get_comtrade_data("USA", "2022")

usa_nafrica_2012 <- get_comtrade_nafrica("USA", 2012)
usa_nafrica_2013 <- get_comtrade_nafrica("USA", 2013)
usa_nafrica_2014 <- get_comtrade_nafrica("USA", 2014)
usa_nafrica_2015 <- get_comtrade_nafrica("USA", 2015)
usa_nafrica_2016 <- get_comtrade_nafrica("USA", 2016)
usa_nafrica_2017 <- get_comtrade_nafrica("USA", 2017)
usa_nafrica_2018 <- get_comtrade_nafrica("USA", 2018)
usa_nafrica_2019 <- get_comtrade_nafrica("USA", 2019)
usa_nafrica_2020 <- get_comtrade_nafrica("USA", 2020)
usa_nafrica_2021 <- get_comtrade_nafrica("USA", 2021)
usa_nafrica_2022 <- get_comtrade_nafrica("USA", 2022)

usa_meast_2012 <- get_comtrade_meast("USA", 2012)
usa_meast_2013 <- get_comtrade_meast("USA", 2013)
usa_meast_2014 <- get_comtrade_meast("USA", 2014)
usa_meast_2015 <- get_comtrade_meast("USA", 2015)
usa_meast_2016 <- get_comtrade_meast("USA", 2016)
usa_meast_2017 <- get_comtrade_meast("USA", 2017)
usa_meast_2018 <- get_comtrade_meast("USA", 2018)
usa_meast_2019 <- get_comtrade_meast("USA", 2019)
usa_meast_2020 <- get_comtrade_meast("USA", 2020)
usa_meast_2021 <- get_comtrade_meast("USA", 2021)
usa_meast_2022 <- get_comtrade_meast("USA", 2022)
```

```

usa_eu_2012 <- get_comtrade_eu("USA", 2012)
usa_eu_2013 <- get_comtrade_eu("USA", 2013)
usa_eu_2014 <- get_comtrade_eu("USA", 2014)
usa_eu_2015 <- get_comtrade_eu("USA", 2015)
usa_eu_2016 <- get_comtrade_eu("USA", 2016)
usa_eu_2017 <- get_comtrade_eu("USA", 2017)
usa_eu_2018 <- get_comtrade_eu("USA", 2018)
usa_eu_2019 <- get_comtrade_eu("USA", 2019)
usa_eu_2020 <- get_comtrade_eu("USA", 2020)
usa_eu_2021 <- get_comtrade_eu("USA", 2021)
usa_eu_2022 <- get_comtrade_eu("USA", 2022)

usa_df <-
  rbind(
    usa_df_2012,
    usa_nafrica_2012,
    usa_meast_2012,
    usa_eu_2012,
    usa_df_2013,
    usa_nafrica_2013,
    usa_meast_2013,
    usa_eu_2013,
    usa_df_2014,
    usa_nafrica_2014,
    usa_meast_2014,
    usa_eu_2014,
    usa_df_2015,
    usa_nafrica_2015,
    usa_meast_2015,
    usa_eu_2015,
    usa_df_2016,
    usa_nafrica_2016,
    usa_meast_2016,
    usa_eu_2016,
    usa_df_2017,
    usa_nafrica_2017,
    usa_meast_2017,
    usa_eu_2017,
    usa_df_2018,
    usa_nafrica_2018,
    usa_meast_2018,

```

```

    usa_eu_2018,
    usa_df_2019,
    usa_nafrica_2019,
    usa_meast_2019,
    usa_eu_2019,
    usa_df_2020,
    usa_nafrica_2020,
    usa_meast_2020,
    usa_eu_2020,
    usa_df_2021,
    usa_nafrica_2021,
    usa_meast_2021,
    usa_eu_2021,
    usa_df_2022,
    usa_nafrica_2022,
    usa_meast_2022,
    usa_eu_2022
  )

head(usa_df)
summary(usa_df)
usa_df %>%
  distinct(partner)

```

## Data Transformation

I transformed the data as shown below in a step-by-step fashion

- Replace NAs to zeros
- Clean names of the columns
- Pivot the data into wider format
- Sum appropriate columns of countries in certain cases such as groups/regions
- Transforming the data from Kilograms to Metric Tons by multiplying each value with a conversion factor of 0.001
- Retained specific columns with `transmute()` and deleted the rest.

```

main_u <- usa_df %>%
  select(period, reporter, partner, netweight_kg) %>%
  filter(partner %in% c("China", "Indonesia", "Japan",
                       "Rep. of Korea", "Viet Nam")) %>%
  spread(key = "partner", value = "netweight_kg") %>%

```

```

replace(is.na(.), 0) %>%
clean_names() %>%
transmute(period, reporter,
           China_MT = china*0.001,
           Indonesia_MT = indonesia*0.001,
           Japan_MT = japan*0.001,
           Rep_of_Korea_MT = rep_of_korea*0.001,
           Vietnam_MT = viet_nam*0.001)

n_africa_u <- usa_df %>%
  select(period, reporter, partner, netweight_kg) %>%
  filter(partner %in% c("Algeria", "Libya", "Egypt",
                       "Morocco", "Tunisia", "Sudan",
                       "Mauritania")) %>%
  spread(key = "partner", value = "netweight_kg") %>%
  replace(is.na(.), 0) %>%
  mutate(n_africa = rowSums(., 3:8)) %>%
  clean_names() %>%
  transmute(period, reporter, Algeria_MT = algeria*0.001,
            Egypt_MT = egypt*0.001,
            Mauritania_MT = mauritania*0.001,
            Morocco_MT = morocco*0.001,
            Sudan_MT = sudan*0.001,
            Tunisia_MT = tunisia*0.001,
            N_Africa_MT = n_africa*0.001)

middle_east_u <- usa_df %>%
  select(period, reporter, partner, netweight_kg) %>%
  filter(partner %in% c("Iraq", "Syria", "Israel",
                       "Jordan", "Kuwait",
                       "Lebanon",
                       "Oman", "Qatar",
                       "Saudi Arabia",
                       "United Arab Emirates",
                       "Yemen", "Bahrain")) %>%
  spread(key = "partner", value = "netweight_kg") %>%
  replace(is.na(.), 0) %>%
  mutate(middle_east = rowSums(., 3:14)) %>%
  clean_names() %>%
  transmute(period, reporter, Bahrain_MT = bahrain*0.001,

```

```

Iraq_MT = iraq*0.001,
Israel_MT = israel*0.001,
Jordan_MT = jordan*0.001,
Kuwait_MT = kuwait*0.001,
Lebanon_MT = lebanon*0.001,
Oman_MT = oman*0.001,
Qatar_MT = qatar*0.001,
Saudi_Arabia_MT = saudi_arabia*0.001,
Syria_MT = syria*0.001,
UAE_MT = united_arab_emirates*0.001,
Yemen_MT = yemen*0.001,
Middle_East_MT = middle_east*0.001)

european_union_u <- usa_df %>%
  select(period, reporter, partner, netweight_kg) %>%
  filter(partner %in% c("Austria", "Belgium", "Bulgaria",
    "Croatia", "Cyprus",
    "Czechia", "Denmark",
    "Estonia", "Finland", "France",
    "Germany", "Greece", "Hungary",
    "Ireland", "Italy",
    "Latvia", "Lithuania",
    "Luxembourg", "Malta", "Netherlands",
    "Poland", "Portugal",
    "Romania", "Slovakia",
    "Slovenia",
    "Spain", "Sweden")) %>%
  spread(key = "partner", value = "netweight_kg") %>%
  replace(is.na(.), 0) %>%
  mutate(european_union = rowSums(. [, 3:26])) %>%
  clean_names() %>%
  transmute(period, reporter, Austria_MT = austria*0.001,
    Belgium_MT = belgium*0.001,
    Bulgaria_MT = bulgaria*0.001,
    Croatia_MT = croatia*0.001,
    Cyprus_MT = cyprus*0.001,
    Denmark_MT = denmark*0.001,
    Estonia_MT = estonia*0.001,
    Finland_MT = finland*0.001,
    France_MT = france*0.001,
    Germany_MT = germany*0.001,

```

```

Greece_MT = greece*0.001,
Hungary_MT = hungary*0.001,
Ireland_MT = ireland*0.001,
Italy_MT = italy*0.001,
Latvia_MT = latvia*0.001,
Lithuania_MT = lithuania*0.001,
Malta_MT = malta*0.001,
Netherlands_MT = netherlands*0.001,
Poland_MT = poland*0.001,
Portugal_MT = portugal*0.001,
Romania_MT = romania*0.001,
Slovenia_MT = slovenia*0.001,
Spain_MT = spain*0.001,
Sweden_MT = sweden*0.001,
European_Union_MT = european_union*0.001)

```

## Export the data into Excel

```

# Create a workbook (spreadsheet)
USA_Exports <- createWorkbook()

# Add three sheets to the spreadsheet
addWorksheet(USA_Exports, sheetName = "Main_Importers")
addWorksheet(USA_Exports, sheetName = "North_Africa")
addWorksheet(USA_Exports, sheetName = "Middle_East")
addWorksheet(USA_Exports, sheetName = "The_EU")
# Write data to each sheet
writeDataTable(
  USA_Exports,
  sheet = "Main_Importers",
  x = main_u
)

writeDataTable(
  USA_Exports,
  sheet = "North_Africa",
  x = n_africa_u
)

writeDataTable(

```

```
    USA_Exports,  
    sheet = "Middle_East",  
    x = middle_east_u  
)  
  
writeDataTable(  
    USA_Exports,  
    sheet = "The_EU",  
    x = european_union_u  
)  
  
saveWorkbook(USA_Exports, "USA_Exports.xlsx")
```