

# Proj\_Term

Alejandro Casillas

2023-11-06

```
#Libraries
rm(list=ls())
library(readr)
library(tidyverse)

## — Attaching core tidyverse packages — tidyverse
2.0.0 —
## ✓ dplyr      1.1.3      ✓ purrr      1.0.2
## ✓ forcats   1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.3      ✓ tibble     3.2.1
## ✓ lubridate 1.9.2      ✓ tidyr      1.3.0
## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(tidyr)
library(stringr)
library(dplyr)
library(ggplot2)
library(summarytools)

##
## Attaching package: 'summarytools'
##
## The following object is masked from 'package:tibble':
##
##     view

#Data Cleaning Process:
#Data was converted from .xls to .csv.
#####
#Data Read Ins:
CAN_exp = read_csv("CAN_exp.csv")

## Rows: 21 Columns: 33
## — Column specification
##
## Delimiter: ","
## chr (5): Reporter Name, Partner Name, Trade Flow, Product Group,
```

```

Indicator
## dbl (28): 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,
2004, ...
##
## 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.

CAN_imp = read_csv("CAN_imp.csv")

## Rows: 21 Columns: 33
## — Column specification

```

---

```

## Delimiter: ","
## chr (5): Reporter Name, Partner Name, Trade Flow, Product Group,
Indicator
## dbl (28): 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,
2004, ...
##
## 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.

MEX_exp = read_csv("MEX_exp.csv")

## Rows: 21 Columns: 33
## — Column specification

```

---

```

## Delimiter: ","
## chr (5): Reporter Name, Partner Name, Trade Flow, Product Group,
Indicator
## dbl (28): 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,
2004, ...
##
## 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.
```

```
#####
```

```
#CANADA wrangle:
```

```
#Turning my data sets Long and setting the Primary Key as year being the columns, which will now be rows instead so we also create a value column.
```

```
CAN_exp_long = gather(CAN_exp, key = "Year", value = "Value", -`Reporter Name`, -`Partner Name`, -`Trade Flow`, -`Product Group`, -`Indicator`)
```

```
CAN_imp_long = gather(CAN_imp, key = "Year", value = "Value", -`Reporter Name`, -`Partner Name`, -`Trade Flow`, -`Product Group`, -`Indicator`)
```

```
#merging my Long export and import data sets for Canada using bind_rows instead of column_bind since this is long. We turned our years 1994-2021 as our obs.
```

```
CAN_data = bind_rows(CAN_exp_long, CAN_imp_long)
```

```
final_CAN_data = CAN_data %>%
```

```
  mutate(Flow_Product = paste(`Trade Flow`, `Product Group`, sep = "_")) %>%  
#paste export or import next to corresponding Prod Group obs.
```

```
  select(-`Trade Flow`, -`Product Group`, -`Indicator`, -`Reporter Name`, -  
  `Partner Name`) %>% #Keep wanted columns
```

```
  pivot_wider(names_from = Flow_Product, #Now Product Group and Exp or Imp become its own column
```

```
    values_from = `Value`) %>%
```

```
  select(Year, everything()) #Data in thousands of USD
```

```
#####
```

```
#####
```

```
#MEXICO wrangle: Exact same process as above.
```

```
MEX_exp_long = gather(MEX_exp, key = "Year", value = "Value", -`Reporter Name`, -`Partner Name`, -`Trade Flow`, -`Product Group`, -`Indicator`)
```

```
MEX_imp_long = gather(MEX_imp, key = "Year", value = "Value", -`Reporter Name`, -`Partner Name`, -`Trade Flow`, -`Product Group`, -`Indicator`)
```

```
MEX_data = bind_rows(MEX_exp_long, MEX_imp_long)
```

```
final_MEX_data = MEX_data %>%
```

```
  mutate(Flow_Product = paste(`Trade Flow`, `Product Group`, sep = "_")) %>%  
  select(-`Trade Flow`, -`Product Group`, -`Indicator`, -`Reporter Name`, -  
  `Partner Name`) %>%
```

```
  pivot_wider(names_from = Flow_Product,  
    values_from = `Value`) %>%
```

```
  select(Year, everything()) #Data in thousands of USD
```

```
#####
```

```
#####
```

```
#Final merging of data sets as on tidy, In order to merge them with full_join, so that i can distinguish Product Group, flow, and Country I will
```

*need to name the columns as following:*

```
final_CAN_data = final_CAN_data %>%  
  rename_with(~paste0(., "_Canada"), -Year)
```

```
final_MEX_data = final_MEX_data %>%  
  rename_with(~paste0(., "_Mexico"), -Year)
```

```
trade_volume_df = full_join(final_CAN_data, final_MEX_data, by = "Year")
```

```
save(trade_volume_df, file = "trade_volume_data.RData")
```

```
#####  
#####
```

*#Make sure final df tidy is clean and has no corrupt, garbage, nas, or need for transformations*

```
sum(is.na(trade_volume_df)) #No NA Values
```

```
## [1] 0
```

*#From Conversions we should turn year into numer from chr*

```
trade_volume_df$Year = as.numeric(trade_volume_df$Year)
```

```
descr(trade_volume_df$Import_Transportation_Mexico)
```

```
## Descriptive Statistics
```

```
## trade_volume_df$Import_Transportation_Mexico
```

```
## N: 28
```

```
##
```

```
## Import_Transportation_Mexico
```

```
## -----
```

```
## Mean 45246414.72
```

```
## Std.Dev 29423482.51
```

```
## Min 7344559.91
```

```
## Q1 26112153.51
```

```
## Median 33342760.56
```

```
## Q3 73272322.14
```

```
## Max 104234615.20
```

```
## MAD 24952716.61
```

```
## IQR 45225431.17
```

```
## CV 0.65
```

```
## Skewness 0.61
```

```
## SE.Skewness 0.44
```

```
## Kurtosis -1.11
```

```
## N.Valid 28.00
```

```
## Pct.Valid 100.00
```

*#In Thousands of USD, little bit to hectic, so lets put it in Millions of USD*

```
## [1] 104234615
```

```

min(trade_volume_df$Import_Transportation_Mexico)

## [1] 7344560

columns_to_scale = names(trade_volume_df)[names(trade_volume_df) != "Year"]

trade_volume_df = trade_volume_df %>%
  mutate_at(.vars = columns_to_scale, .funs = list(~./1e6)) #Now in Millions
of USD

descr(trade_volume_df$Import_Transportation_Mexico)

## Descriptive Statistics
## trade_volume_df$Import_Transportation_Mexico
## N: 28
##
##              Import_Transportation_Mexico
## -----
##              Mean                45.25
##              Std.Dev             29.42
##              Min                  7.34
##              Q1                   26.11
##              Median               33.34
##              Q3                   73.27
##              Max                 104.23
##              MAD                  24.95
##              IQR                  45.23
##              CV                   0.65
##              Skewness              0.61
##              SE.Skewness           0.44
##              Kurtosis              -1.11
##              N.Valid               28.00
##              Pct.Valid             100.00

max(trade_volume_df$Import_Transportation_Mexico) #In Millions of USD, much
better. Plus, very limited observations so extreme values relevant.

## [1] 104.2346

min(trade_volume_df$Import_Transportation_Mexico)

## [1] 7.34456

#####
#####
#Explore Mexico Imports:
mx_trans_hist = ggplot(trade_volume_df, aes(x =
Import_Transportation_Mexico)) +
  geom_histogram(binwidth = 6, fill = "blue", color = "black") +
  labs(title = "Histogram of Import_Transportation_Mexico (in Millions of
USD)", x = "Millions of USD", y = "Frequency")

```

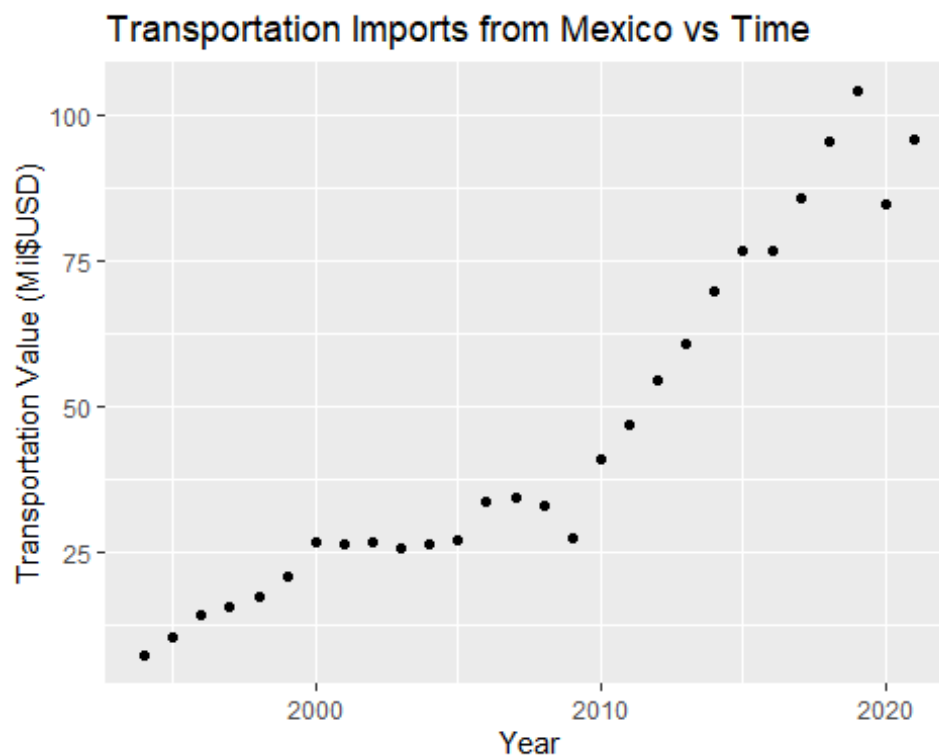
```

#mx_trans_qqnorm = qqnorm(trade_volume_df$Import_Transportation_Mexico)
#mx_trans_qqline = qqline(trade_volume_df$Import_Transportation_Mexico, col =
"red")

#Viz 1: Positive Trend: Transportation Imports from Mexico Over Time
mx_trans_vs_time = ggplot(trade_volume_df, aes(x = Year, y =
Import_Transportation_Mexico)) +
  geom_point() +
  labs(x = "Year", y = "Transportation Value (Mil$USD)") +
  ggtitle("Transportation Imports from Mexico vs Time")

mx_trans_vs_time

```



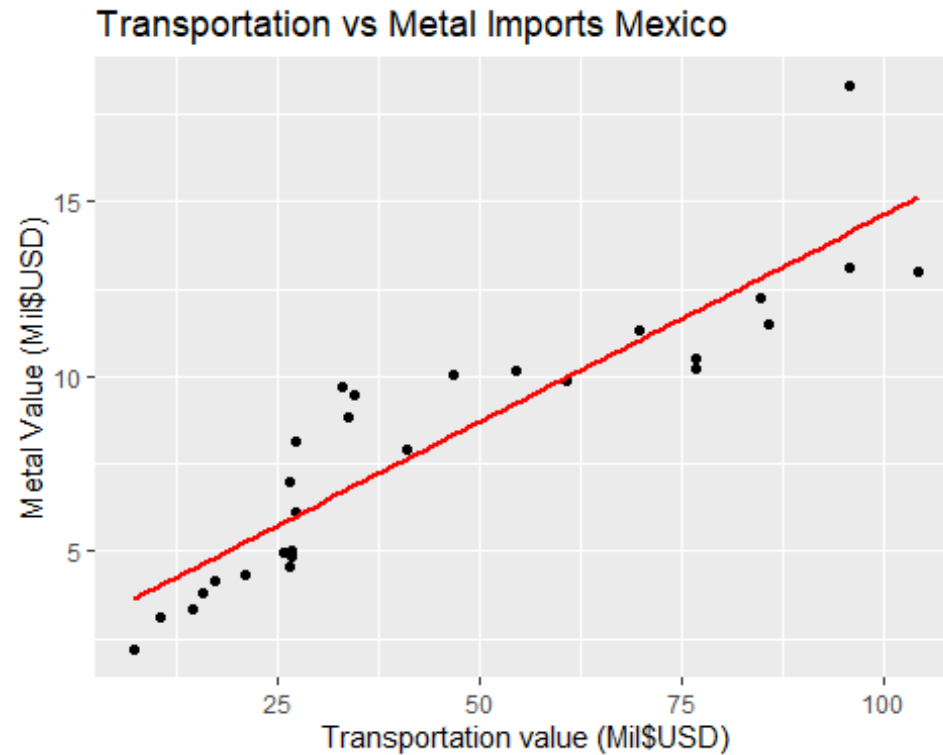
```

#Viz 2: Positive Trend: Transportation Imports from Mexico and Metal Imports
from Mexico
mx_trans_vs_met = ggplot(trade_volume_df, aes(x =
Import_Transportation_Mexico, y = Import_Metals_Mexico)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(x = "Transportation value (Mil$USD)", y = "Metal Value (Mil$USD)") +
  ggtitle("Transportation vs Metal Imports Mexico")

mx_trans_vs_met

## `geom_smooth()` using formula = 'y ~ x'

```



```
#####
#####
can_mach_hist = ggplot(trade_volume_df, aes(x = `Export_Mach and
Elec_Canada`)) +
  geom_histogram(binwidth = 6, fill = "blue", color = "black") +
  labs(title = "Histogram of Export_Mach & Elec_Canada (Mil$USD)", x =
"Millions of USD", y = "Frequency")

#can_machqqnorm = qqnorm(trade_volume_df$`Export_Mach and Elec_Canada`)
#can_machqqline = qqline(trade_volume_df$`Export_Mach and Elec_Canada`, col =
"red")

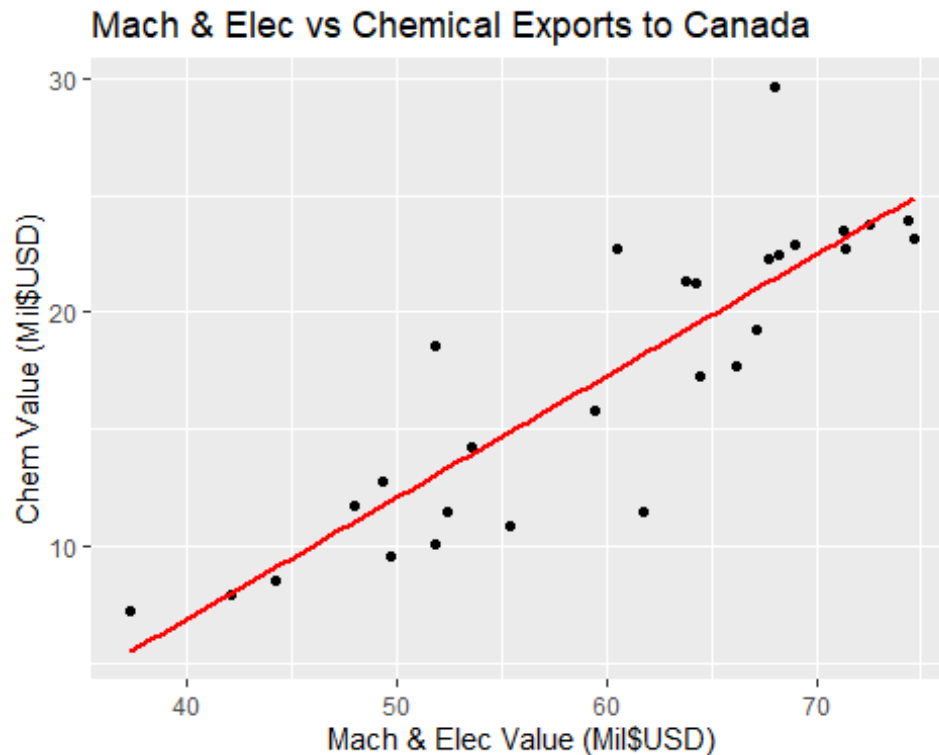
can_mach_vs_time = ggplot(trade_volume_df, aes(x = Year, y = `Export_Mach and
Elec_Canada`)) +
  geom_point() +
  labs(x = "Year", y = "Exports Mach & Elec") +
  ggtitle("Mach & Elec Exports to Canada vs Time")

#Viz 3: Positive Trend: Machine & Electronic Exports to Cananada and Chemical
Exports to Canada
can_mach_vs_chem = ggplot(trade_volume_df, aes(x = `Export_Mach and
Elec_Canada`, y = `Export_Chemicals_Canada`)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(x = "Mach & Elec Value (Mil$USD)", y = "Chem Value (Mil$USD)") +
```

```
ggtitle("Mach & Elec vs Chemical Exports to Canada")
```

```
can_mach_vs_chem
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



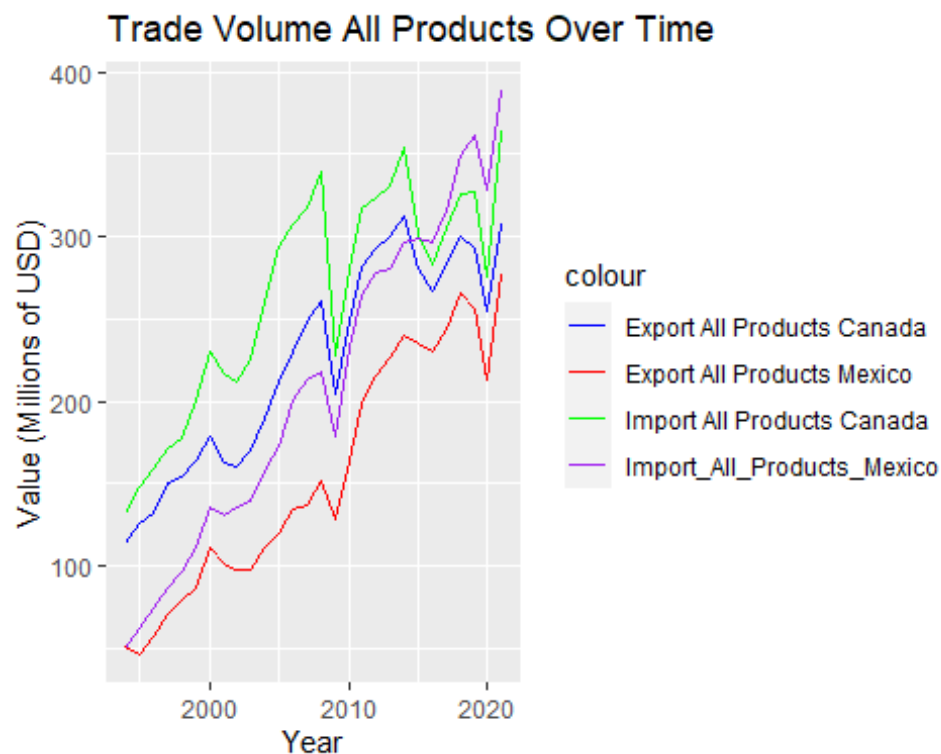
```
#####
#####
#Viz 4: Positive Trend of ALL Products Exports & Imports Over Time
trade_volume_plot = ggplot(trade_volume_df, aes(x = Year)) +
  geom_line(aes(y = `Export_All Products_Canada`, color =
"Export_All_Products_Canada")) +
  geom_line(aes(y = `Import_All Products_Canada`, color =
"Import_All_Products_Canada")) +
  geom_line(aes(y = `Export_All Products_Mexico`, color =
"Export_All_Products_Mexico")) +
  geom_line(aes(y = `Import_All Products_Mexico`, color =
"Import_All_Products_Mexico")) +
  labs(
    title = "Trade Volume All Products Over Time",
    x = "Year",
    y = "Value (Millions of USD)"
  ) +
  scale_color_manual(
    values = c(
      "Export_All_Products_Canada" = "blue",
      "Import_All_Products_Canada" = "green",
```



```

    "Export_All_Products_Mexico" = "red",
    "Import_All_Products_Mexico" = "purple"
  ),
  labels = c(
    "Export_All_Products_Canada" = "Export All Products Canada",
    "Import_All_Products_Canada" = "Import All Products Canada",
    "Export_All_Products_Mexico" = "Export All Products Mexico",
    "Import_All_Products_Mexico" = "Import All Products Mexico"
  )
)
trade_volume_plot

```



```

#####
###
# Write Up A Report. Still Missing. Need and Want to speak to professor. Team
member still working on their part so i can do the whole report yet.

```