Explanatory Data Analysis of Trade Statistics of Turkey

MEF University - BDA 503 Essentials of Data Analytics - Fall 2018 - Group Fou*R* Term Project

December 2018

Contents

Abstract

In this term project, trade statistics data of Turkey against the G7 countries and 3 more countries (Russia, India, China) is analysed. The aim of the project is to analyze Turkey's export and import characteristics by product groups considering the trade value and quantities. Within this aim, we take trade values and quantities by product groups in MS Excel file format from trademap.org website for each trade partner for the period between 2013 and 2017. In this study, We followed the steps of explanation of raw data, preparation of raw data through merging the raw data files, cleaning, and manipulating, analysing the data to explore the trade characteristics along with visulations and lastly concluding the comments from analysis.

Explanation of Raw Data

Following the extraction from source (Trademap.org), our raw data consist of 10 countries' trade data against Turkey. This data split with trade direction as 'Export' and 'Import', in other words each country has 2 files as export and import. The other dimension is trade data; one is 'Quantity' of Export/Import, the other is 'Value' in thousand USD of Export/Import, which means each country has one file for 'Quantity' and one file for 'Trade Value'. Therefore we have four main data table for each country (total of 40 excel files).

Data Frame

DF Definition

ExportValues

Export from Turkey to target countries and their values in thousand USD

ExportQuantity

Export from Turkey to target countries and their quantities in units(tons,gallons etc)

ImportValues

Import to Turkey from target counties and their values in thousand USD

ImportQuantity

Import to Turkey from target counties and their quantities in units(tons,gallons etc)

Variables in “Quantity” and “Value” files are as follows:

Country

Product code

Product label

Quantity in 2013

Unit

Quantity in 2014

Unit

Quantity in 2015

Unit

Quantity in 2016

Unit

Quantity in 2017

Unit

France

’0306

Crustaceans, whether in shell or not, live, fresh, chilled, frozen, dried, salted or in brine, …

121

Tons

139

Tons

22

Tons

24

Tons

Country

Product code

Product label

Value in 2013

Value in 2014

Value in 2015

Value in 2016

Value in 2017

France

’0306

Crustaceans, whether in shell or not, live, fresh, chilled, frozen, dried, salted or in brine, . . .

0

1155

1182

270

0

In addition to main raw data, we also have a look-up table for product codes. This lookup table registers product labels in 2 digits code that is less categorized than main data.

Product code

Product label

’01

Live animals

’02

Meat and edible meat offal

Regarding the variables included in these tables; definitions are as follows:

ExportQuantity & ImportQuantity

Country

Trade Partner of Turkey

Product_Code

Traded Product Code in 4 digit

Product_Label

Product Definition

Unit

Metric of Quantity

Quantity_2013

Export or Import quantity in 2013

Quantity_2014

Export or Import quantity in 2014

Quantity_2015

Export or Import quantity in 2015

Quantity_2016

Export or Import quantity in 2016

Quantity_2017

Export or Import quantity in 2017

ExportValues & ImportValues

Country

Trade Partner of Turkey

Product_Code

Traded Product Code in 4 digit

Product_Label

Product Definition

Value_2013

Export or Import value in 2013

Value_2014

Export or Import value in 2014

Value_2015

Export or Import value in 2015

Value_2016

Export or Import value in 2016

Value_2017

Export or Import value in 2017

Data Preparation, Cleaning and Manipulation

We started with loading the needed libraries for preparation, manipulation, analysis and visualisation.

```
library(shiny)
library(tidyverse)
library(dplyr)
library(readxl)
library(leaflet)
library(reshape2)
library(gridExtra)
library(lubridate)
library(plotrix)
library(treemap)
```

First of all, we need to read the excel files constituting our raw data. For this aim, we wrote a function named "getDataSet".

```
PATH <-"C:/Users/emrek/Google Drive/BDA/503-EssentialsOfDataAnalytics/GitHub/gpj18-group_four/Raw Data"
setwd(PATH)

getDataSet <- function(pPattern){
  nm <- list.files(path = PATH, pattern = pPattern)
  if (exists("resultDataset")){
    rm("resultDataset")
  }
  for (file in nm){
    if (!exists("resultDataset")){
      resultDataset <- read_excel(file, col_names=FALSE, skip=1)
    }
```

```r
    else{
      temp_dataset <-read_excel(file, col_names=FALSE, skip=1)
      resultDataset<-rbind(resultDataset, temp_dataset)
      rm(temp_dataset)
    }

  }
  return(resultDataset)

}


import_quantity_dataset <- getDataSet('*.mport.*uant.*')
import_value_dataset <- getDataSet('*.mport.*alu.*')
export_quantity_dataset <- getDataSet('*.xport.*uant.*')
export_value_dataset <- getDataSet('*.xport.*alu.*')
```

This function reads the excel files in the defined path according to given pattern in the filename and bind them in one file named "resultDataset". By using the getDataSet function, we merged the Turkey's export-quantity, export-value, import-quantity and import-value tables for 10 selected countries.

Then, we named the columns and removed the single quotation marks in product code column in each file.

```r
colnames(import_quantity_dataset) <- c("Country","Product_Code","Product_Label","Quantity_2013","Unit1"
colnames(import_value_dataset) <- c("Country","Product_Code","Product_Label","Value_2013","Value_2014",
colnames(export_quantity_dataset) <- c("Country","Product_Code","Product_Label","Quantity_2013","Unit1"
colnames(export_value_dataset) <- c("Country","Product_Code","Product_Label","Value_2013","Value_2014",

import_quantity_dataset$Product_Code<- gsub("'", "", import_quantity_dataset$Product_Code)
import_value_dataset$Product_Code<- gsub("'", "",import_value_dataset$Product_Code)
export_quantity_dataset$Product_Code<- gsub("'", "",export_quantity_dataset$Product_Code)
export_value_dataset$Product_Code<- gsub("'", "",export_value_dataset$Product_Code)
```

After the first step, the view of export & import quantity and export& import value datasets are below in respective order:

```r
glimpse(export_quantity_dataset)
```

```
## Observations: 12,223
## Variables: 13
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "TOTAL", "7214", "6907", "2523", "7208", "6802",...
## $ Product_Label <chr> "All products", "Bars and rods, of iron or non-a...
## $ Quantity_2013 <dbl> NA, 131711, 1795, 22000, 220, 46950, 19782, 2821...
## $ Unit1         <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
## $ Quantity_2014 <dbl> NA, 133195, 2560, 54210, 157114, 49668, 38396, 1...
## $ Unit2         <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
## $ Quantity_2015 <dbl> NA, 1085, 2291, 78195, 10756, 38561, 29677, 4219...
## $ Unit3         <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
## $ Quantity_2016 <dbl> NA, 1759, 1935, 62500, 38354, 33166, 19759, 2239...
## $ Unit4         <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
## $ Quantity_2017 <dbl> 0, 220831, 140214, 129930, 80459, 35413, 34646, ...
## $ Unit5         <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
```

```r
glimpse(import_quantity_dataset)
```

```
## Observations: 12,222
## Variables: 13
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "2601", "7204", "2701", "4401", "0713", "1001", ...
## $ Product_Label <chr> "Iron ores and concentrates, incl. roasted iron ...
## $ Quantity_2013 <dbl> 465332, 373154, 338707, 313714, 194731, 230375, ...
## $ Unit1         <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity_2014 <dbl> 162004, 277774, 491793, 126757, 298731, 89640, N...
## $ Unit2         <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity_2015 <dbl> 494577, 317117, 504464, 297655, 304681, 161211, ...
## $ Unit3         <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity_2016 <dbl> 504892, 372166, 1344338, 308932, 322942, 196538,...
## $ Unit4         <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity_2017 <dbl> 675057, 526312, 485643, 241778, 235982, 145438, ...
## $ Unit5         <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
```

```
glimpse(export_value_dataset)
```

```
## Observations: 12,222
## Variables: 8
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "7214", "8703", "8901", "0802", "8902", "7208", ...
## $ Product_Label <chr> "Bars and rods, of iron or non-alloy steel, not ...
## $ Value_2013    <dbl> 73607, 13976, 0, 57640, 0, 211, 895, 40420, 1052...
## $ Value_2014    <dbl> 73814, 2485, 0, 65460, 0, 96558, 1227, 46362, 11...
## $ Value_2015    <dbl> 616, 2477, 62, 105122, 0, 5033, 1052, 37154, 129...
## $ Value_2016    <dbl> 863, 94, 66747, 94623, 0, 18777, 832, 30918, 171...
## $ Value_2017    <dbl> 106068, 87274, 84922, 75731, 45504, 44261, 40668...
```

```
glimpse(import_value_dataset)
```

```
## Observations: 12,223
## Variables: 8
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "7108", "0713", "7204", "2601", "2701", "2710", ...
## $ Product_Label <chr> "Gold, incl. gold plated with platinum, unwrough...
## $ Value_2013    <dbl> 230689, 132022, 147798, 77901, 62688, 6565, 1065...
## $ Value_2014    <dbl> 8259, 204972, 104914, 25982, 70544, 6244, 169636...
## $ Value_2015    <dbl> 760, 232255, 77607, 43341, 55067, 6183, 47389, 2...
## $ Value_2016    <dbl> 48488, 265707, 84380, 38810, 135545, 7990, 24360...
## $ Value_2017    <dbl> 1030428, 164565, 157861, 78121, 73275, 70084, 63...
```

We have 13 and 8 variables in quantity and value tables respectively for export and import.

Now, it is time to join quantity and value data for both export and import files. While joining the tables, we should also clear the repetitive variable "Unit" for each year.

Also, we need to join product labels in 2 digit product codes with these table.

```
#read the ProductCodes excel
file = "C:/Users/emrek/Google Drive/BDA/503-EssentialsOfDataAnalytics/GitHub/gpj18-group_four/2digitpro
prdCode <- read_excel(file, col_names=FALSE, skip=1)
prdCode$X__1 <- gsub("'", "", prdCode$X__1)

# join the import values and quantities
import_jTable <- import_quantity_dataset %>%
  mutate( Unit= coalesce(Unit1,Unit2 ,Unit3 , Unit4,Unit5)) %>%
```

```r
  select( Country, Product_Code , Unit, Quantity_2013, Quantity_2014, Quantity_2015, Quantity_2016, Quan
  full_join(import_value_dataset,by = c("Country","Product_Code"))
# seperate the first 2 digits of 4 digit product code and join the 2digit product code excel
import_jTable <- import_jTable %>% mutate(PcodeTwo=substr(Product_Code,1,2))
import_jTable <- left_join(import_jTable,prdCode,by =c("PcodeTwo"="X__1") )

# join the export values and quantities
export_jTable <- export_quantity_dataset %>%
  mutate( Unit= coalesce(Unit1,Unit2 ,Unit3 , Unit4,Unit5)) %>%
  select( Country, Product_Code , Unit, Quantity_2013, Quantity_2014,Quantity_2015,Quantity_2016,Quantit
  full_join(export_value_dataset,by = c("Country","Product_Code"))
# seperate the first 2 digits of 4 digit product code and join the 2digit product code excel
export_jTable <- export_jTable %>% mutate(PcodeTwo=substr(Product_Code,1,2))
export_jTable <- left_join(export_jTable,prdCode,by =c("PcodeTwo"="X__1") )
```

Now we have joined tables for import and export stats of Turkey. Lets glimpse them:

```r
glimpse(import_jTable)
```

```
## Observations: 12,223
## Variables: 16
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "2601", "7204", "2701", "4401", "0713", "1001", ...
## $ Unit          <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity_2013 <dbl> 465332, 373154, 338707, 313714, 194731, 230375, ...
## $ Quantity_2014 <dbl> 162004, 277774, 491793, 126757, 298731, 89640, N...
## $ Quantity_2015 <dbl> 494577, 317117, 504464, 297655, 304681, 161211, ...
## $ Quantity_2016 <dbl> 504892, 372166, 1344338, 308932, 322942, 196538,...
## $ Quantity_2017 <dbl> 675057, 526312, 485643, 241778, 235982, 145438, ...
## $ Product_Label <chr> "Iron ores and concentrates, incl. roasted iron ...
## $ Value_2013    <dbl> 77901, 147798, 62688, 33992, 132022, 93218, 0, 6...
## $ Value_2014    <dbl> 25982, 104914, 70544, 15532, 204972, 31688, 0, 6...
## $ Value_2015    <dbl> 43341, 77607, 55067, 32940, 232255, 47770, 0, 61...
## $ Value_2016    <dbl> 38810, 84380, 135545, 31159, 265707, 54832, 0, 7...
## $ Value_2017    <dbl> 78121, 157861, 73275, 23684, 164565, 34618, 2364...
## $ PcodeTwo      <chr> "26", "72", "27", "44", "07", "10", "27", "27", ...
## $ X__2          <chr> "Ores, slag and ash", "Iron and steel", "Mineral...
```

```r
glimpse(export_jTable)
```

```
## Observations: 12,223
## Variables: 16
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ Product_Code  <chr> "TOTAL", "7214", "6907", "2523", "7208", "6802",...
## $ Unit          <chr> "No quantity", "Tons", "Tons", "Tons", "Tons", "...
## $ Quantity_2013 <dbl> NA, 131711, 1795, 22000, 220, 46950, 19782, 2821...
## $ Quantity_2014 <dbl> NA, 133195, 2560, 54210, 157114, 49668, 38396, 1...
## $ Quantity_2015 <dbl> NA, 1085, 2291, 78195, 10756, 38561, 29677, 4219...
## $ Quantity_2016 <dbl> NA, 1759, 1935, 62500, 38354, 33166, 19759, 2239...
## $ Quantity_2017 <dbl> 0, 220831, 140214, 129930, 80459, 35413, 34646, ...
## $ Product_Label <chr> NA, "Bars and rods, of iron or non-alloy steel, ...
## $ Value_2013    <dbl> NA, 73607, 895, 1531, 211, 40420, 16625, 37956, ...
## $ Value_2014    <dbl> NA, 73814, 1227, 3961, 96558, 46362, 32160, 1484...
## $ Value_2015    <dbl> NA, 616, 1052, 5810, 5033, 37154, 23569, 6140, 0...
## $ Value_2016    <dbl> NA, 863, 832, 4590, 18777, 30918, 12319, 24849, ...
```

```
## $ Value_2017    <dbl> NA, 106068, 40668, 8801, 44261, 30766, 27136, 27...
## $ PcodeTwo      <chr> "TO", "72", "69", "25", "72", "68", "73", "73", ...
## $ X__2          <chr> NA, "Iron and steel", "Ceramic products", "Salt;...
```

Both tables now have 16 variables and 12,223 rows.

But there is a row in Product_Code column as "TOTAL", we cleared it as well.

```
export_jTable <- export_jTable %>% filter(!grepl("TOTAL", Product_Code))
import_jTable <- import_jTable %>% filter(!grepl("TOTAL", Product_Code))
```

We have 16 variables in these tables however quantity and value variables repeat for each year. So actually, it is better to organize the table that it will have 3 variables as "Year", "Quantity" and "Value" instead of 10 columns.

```
#create new objects for each years' quantity
q2013<-import_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2014<-import_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2015<-import_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2016<-import_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2017<-import_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20

#bind the objects
importQuant <- bind_rows(q2013,q2014,q2015,q2016,q2017)

#create new objects for each years' values
v2013<-import_jTable %>% select(Country,Product_Code, Value_2013) %>% rename("Values"=Value_2013) %>% mu
v2014<-import_jTable %>% select(Country,Product_Code, Value_2014) %>% rename("Values"=Value_2014) %>% mu
v2015<-import_jTable %>% select(Country,Product_Code, Value_2015) %>% rename("Values"=Value_2015) %>% mu
v2016<-import_jTable %>% select(Country,Product_Code, Value_2016) %>% rename("Values"=Value_2016) %>% mu
v2017<-import_jTable %>% select(Country,Product_Code, Value_2017) %>% rename("Values"=Value_2017) %>% mu

#bind the objects
importValues <- bind_rows(v2013,v2014,v2015,v2016,v2017)

#join binded quantity and value
import_jTableMELTED <- importQuant %>% left_join(importValues, by=c("Country","Product_Code","year"))


#Repeat with export table

q2013<-export_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2014<-export_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2015<-export_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2016<-export_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20
q2017<-export_jTable %>% select(Country, PcodeTwo, X__2 ,Product_Code, Product_Label, Unit, Quantity_20

exportQuant <- bind_rows(q2013,q2014,q2015,q2016,q2017)

v2013<-export_jTable %>% select(Country,Product_Code, Value_2013) %>% rename("Values"=Value_2013) %>% mu
v2014<-export_jTable %>% select(Country,Product_Code, Value_2014) %>% rename("Values"=Value_2014) %>% mu
v2015<-export_jTable %>% select(Country,Product_Code, Value_2015) %>% rename("Values"=Value_2015) %>% mu
v2016<-export_jTable %>% select(Country,Product_Code, Value_2016) %>% rename("Values"=Value_2016) %>% mu
v2017<-export_jTable %>% select(Country,Product_Code, Value_2017) %>% rename("Values"=Value_2017) %>% mu

exportValues <- bind_rows(v2013,v2014,v2015,v2016,v2017)
```

```r
export_jTableMELTED <- exportQuant %>% left_join(exportValues, by=c("Country","Product_Code","year"))
```

After above coding, now we have more useful tables, with 9 variables and 61100 rows.

```r
glimpse(export_jTableMELTED)
```

```
## Observations: 61,100
## Variables: 9
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ PcodeTwo      <chr> "72", "69", "25", "72", "68", "73", "73", "26", ...
## $ X__2          <chr> "Iron and steel", "Ceramic products", "Salt; sul...
## $ Product_Code  <chr> "7214", "6907", "2523", "7208", "6802", "7306", ...
## $ Product_Label <chr> "Bars and rods, of iron or non-alloy steel, not ...
## $ Unit          <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity      <dbl> 131711, 1795, 22000, 220, 46950, 19782, 28212, N...
## $ year          <chr> "2013", "2013", "2013", "2013", "2013", "2013", ...
## $ Values        <dbl> 73607, 895, 1531, 211, 40420, 16625, 37956, 0, 0...
```

```r
glimpse(import_jTableMELTED)
```

```
## Observations: 61,100
## Variables: 9
## $ Country       <chr> "Canada", "Canada", "Canada", "Canada", "Canada"...
## $ PcodeTwo      <chr> "26", "72", "27", "44", "07", "10", "27", "27", ...
## $ X__2          <chr> "Ores, slag and ash", "Iron and steel", "Mineral...
## $ Product_Code  <chr> "2601", "7204", "2701", "4401", "0713", "1001", ...
## $ Product_Label <chr> "Iron ores and concentrates, incl. roasted iron ...
## $ Unit          <chr> "Tons", "Tons", "Tons", "Tons", "Tons", "Tons", ...
## $ Quantity      <dbl> 465332, 373154, 338707, 313714, 194731, 230375, ...
## $ year          <chr> "2013", "2013", "2013", "2013", "2013", "2013", ...
## $ Values        <dbl> 77901, 147798, 62688, 33992, 132022, 93218, 0, 6...
```

These tables are ready for analysis that will be conducted in the next section. However, before proceeding first lets check the summary of export and import tables.

```r
summary(export_jTableMELTED)
```

```
##    Country            PcodeTwo              X__2          
##  Length:61100       Length:61100       Length:61100      
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##  Product_Code       Product_Label          Unit          
##  Length:61100       Length:61100       Length:61100      
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##     Quantity            year               Values       
##  Min.   :      0   Length:61100       Min.   :      0  
##  1st Qu.:      2   Class :character   1st Qu.:      0  
```

```
##  Median :     35   Mode  :character   Median :      8
##  Mean   :   4067                      Mean   :   4391
##  3rd Qu.:    404                      3rd Qu.:    439
##  Max.   :4879841                      Max.   :3122678
##  NA's   :24869
```

```r
sd(export_jTableMELTED$Values)
```

```
## [1] 36438.36
```

```r
summary(import_jTableMELTED)
```

```
##    Country            PcodeTwo              X__2
##  Length:61100       Length:61100       Length:61100
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Product_Code       Product_Label          Unit
##  Length:61100       Length:61100       Length:61100
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     Quantity             year              Values
##  Min.   :       0   Length:61100       Min.   :       0
##  1st Qu.:       2   Class :character   1st Qu.:       0
##  Median :      41   Mode  :character   Median :      64
##  Mean   :    8872                      Mean   :    9495
##  3rd Qu.:     528                      3rd Qu.:    1611
##  Max.   :14390025                      Max.   :12227836
##  NA's   :18672
```

```r
sd(import_jTableMELTED$Values)
```

```
## [1] 110904.1
```

Since the quantities have different respective units, it is not logical to comment on descriptive statistics for quantity but for the values column we can say that the values are so dispersed with 4391 mean value and 36438 standart deviation for exports and 9495 mean and 110904 standart deviation.

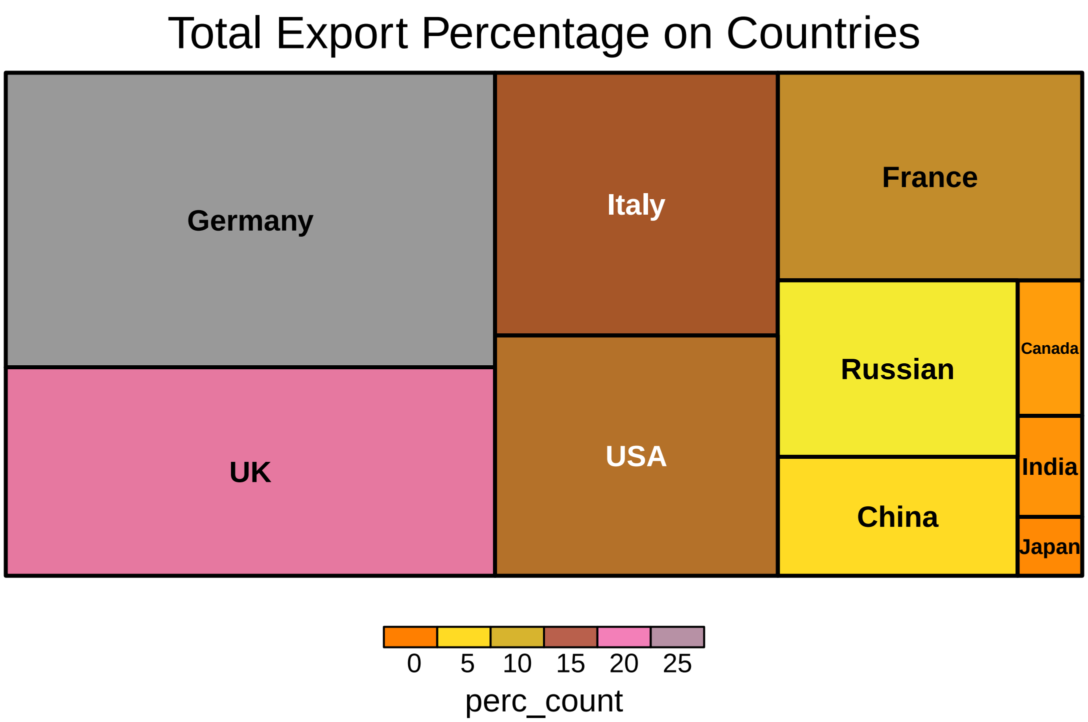
Analysis

Breakdown of Total Export and Import of Turkey Over Countries

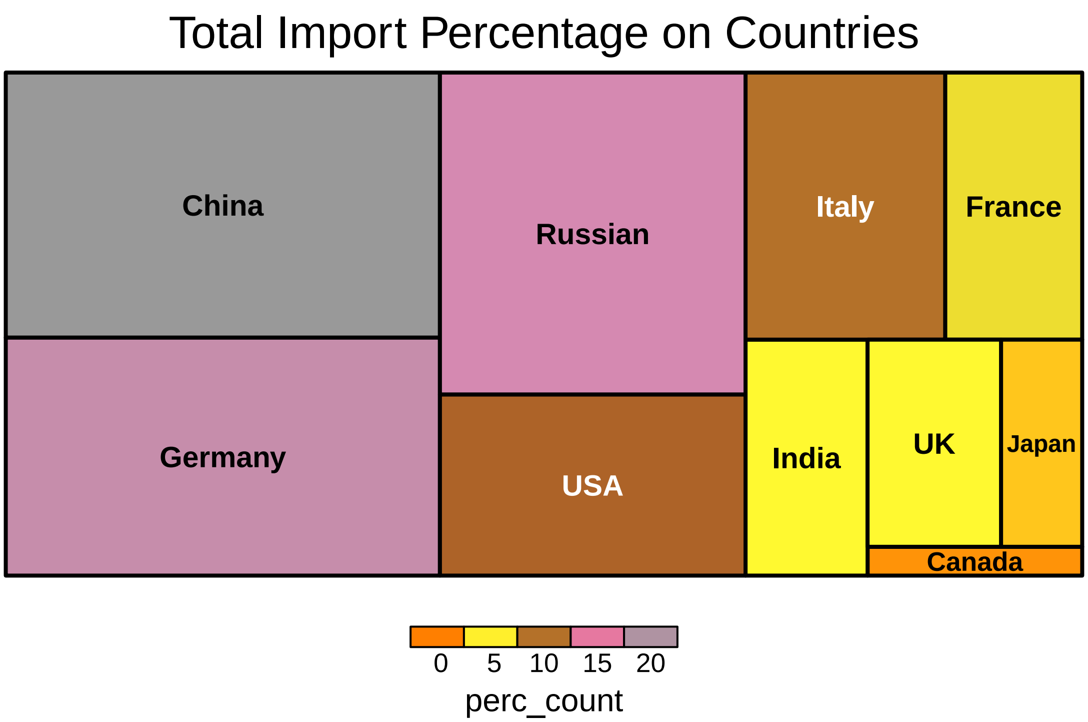
Our analysis starts with to understandthe shares of trade volume of the selected 10 countries on Turkey's import and export.

```r
# TREEMAP FOR EXPORT

exptable1=export_jTableMELTED%>%group_by(Country)%>%
  summarise(co_total=sum(Values,na.rm = TRUE))%>%
```

```
  arrange(desc(co_total,year))

pex1_table<-exptable1%>%
  mutate(perc_count=(co_total/sum(exptable1$co_total))*100)%>%
  select(Country,co_total,perc_count)

treemap(pex1_table,index = c("Country"),vSize = "perc_count",vColor = "perc_count",title="Total Export P
        fontsize.title = 17,type="value",palette = "Set1")
```



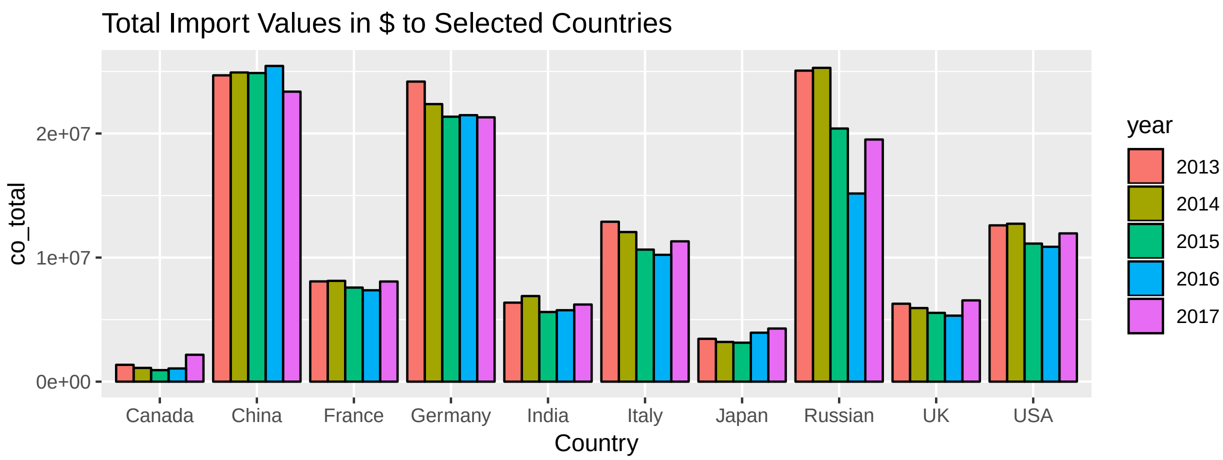

```
# TREEMAP FOR IMPORT

imptable1=import_jTableMELTED%>%group_by(Country)%>%
  summarise(co_total=sum(Values,na.rm = TRUE))%>%
  arrange(desc(co_total,year))

pim1_table<-imptable1%>%
  mutate(perc_count=(co_total/sum(imptable1$co_total))*100)%>%
  select(Country,co_total,perc_count)

treemap(pim1_table,index = c("Country"),vSize = "perc_count",vColor = "perc_count",title="Total Import P
        fontsize.title = 17,type="value",palette = "Set1")
```

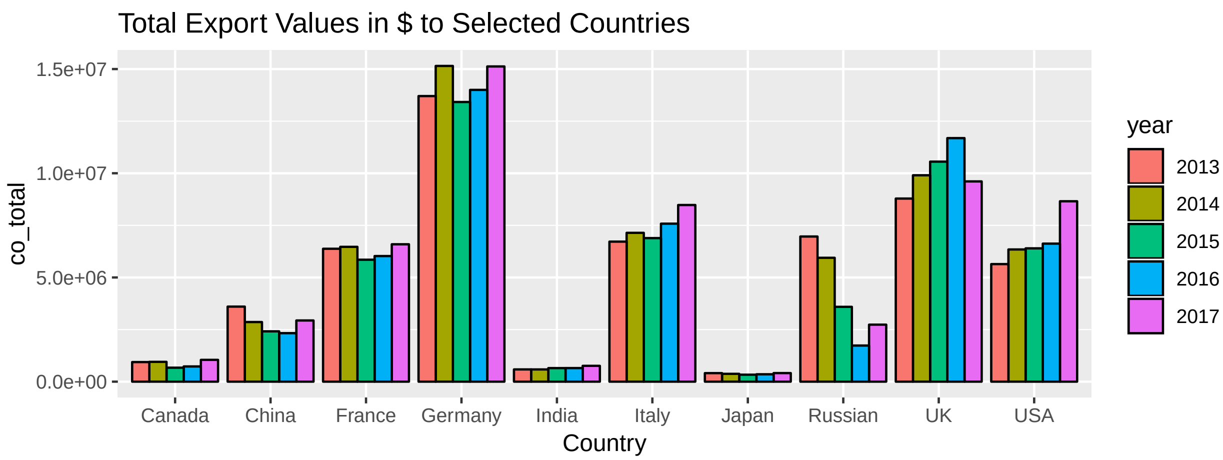


As seen above, Turkeys' biggest trade partners are China, Germany, Russia and US at import side while they are Germany, UK, Italy and USA at export side. Germany and USA are the 2 top countries seen in both directions of trade.

When we breakdown this info to years, it is seen in below graphs that except Russia, fluctuations over years are not seem so significant but it is worth to deep dive Year over Year (YoY) changes.

```
# Export Values in $
ttable=export_jTableMELTED%>%group_by(Country,year)%>%
  summarise(co_total=sum(Values,na.rm = TRUE))%>%
  arrange(desc(co_total,year))
#Import Values in $
vtable=import_jTableMELTED%>%group_by(Country,year)%>%summarise(co_total=sum(Values,na.rm = TRUE))%>%ar

#Import & Export percentage of Countries

a<-ggplot(data=vtable, aes(x=Country, y=co_total,fill=year)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black")+
  ggtitle("Total Import Values in $ to Selected Countries")


b<-ggplot(data=ttable, aes(x=Country, y=co_total,fill=year)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black")+ggtitle("Total Export Values in $



grid.arrange(a,b,  nrow = 2)
```



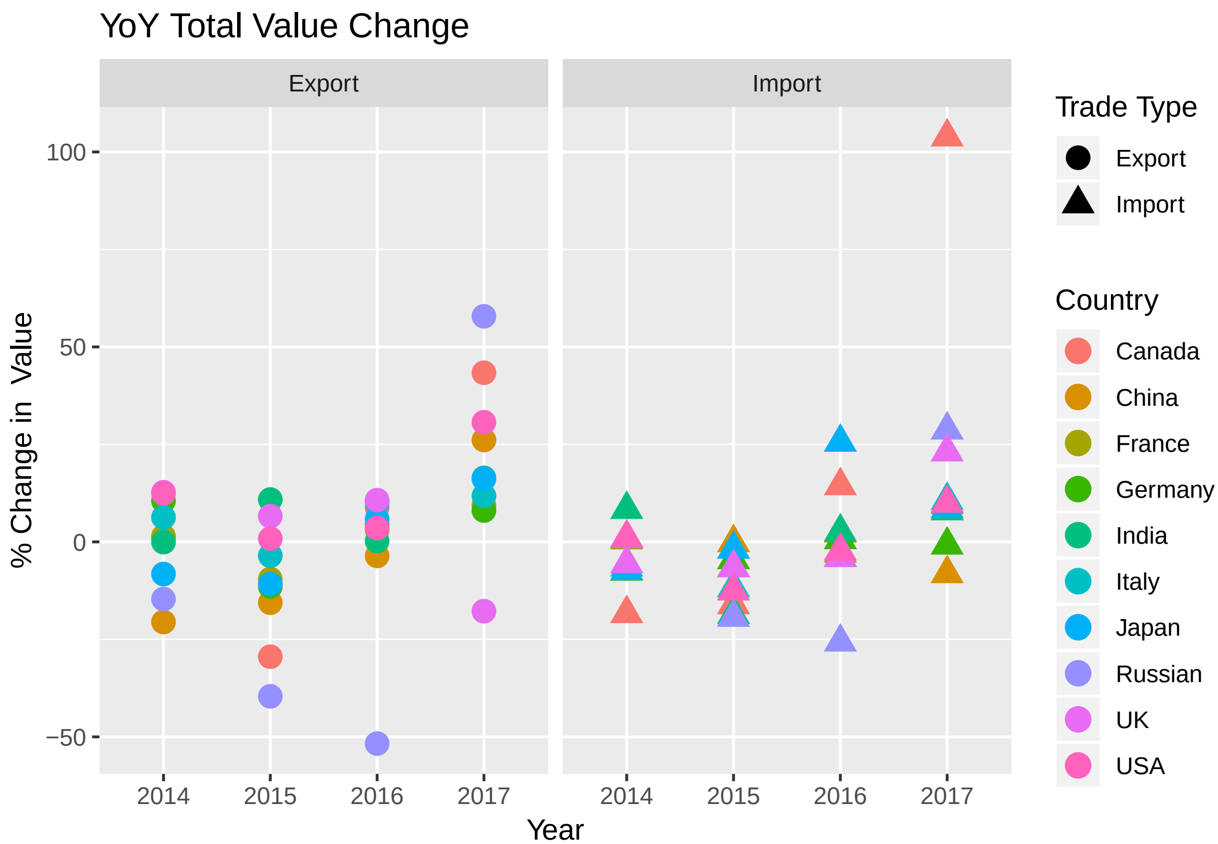


Analysis of Year over Year Changes in Trade Values

In order to review YoY changes, we calculated delta of trade value for consecutive years and plotted as below seperating export and import.

```
tableExport <- export_jTableMELTED %>%
  group_by(Country, year) %>%
  summarise(totalExpV=sum(Values,na.rm=TRUE)) %>%
  arrange(Country, year) %>%
  mutate(ExpValChng = (totalExpV - lag(totalExpV))/lag(totalExpV)*100)


tableImport <- import_jTableMELTED %>%
  group_by(Country, year) %>%
  summarise(totalImpV=sum(Values,na.rm=TRUE)) %>%
  arrange(Country, year) %>%
  mutate(ImpValChng = (totalImpV - lag(totalImpV))/lag(totalImpV)*100)

table <- full_join(tableExport,tableImport) %>%
  select(Country, year, Export="ExpValChng", Import = "ImpValChng") %>%
  melt(id=c("Country", "year")) %>%
  filter(year!=2013)
```

```r
ggplot(table, aes(x=year)) +
  geom_point(aes(y=value, color=Country, shape=variable), size=4) +
  facet_grid(~variable) +
   labs(title="YoY Total Value Change",
    x="Year",
    y="% Change in  Value",
    shape=c("Trade Type"))
```

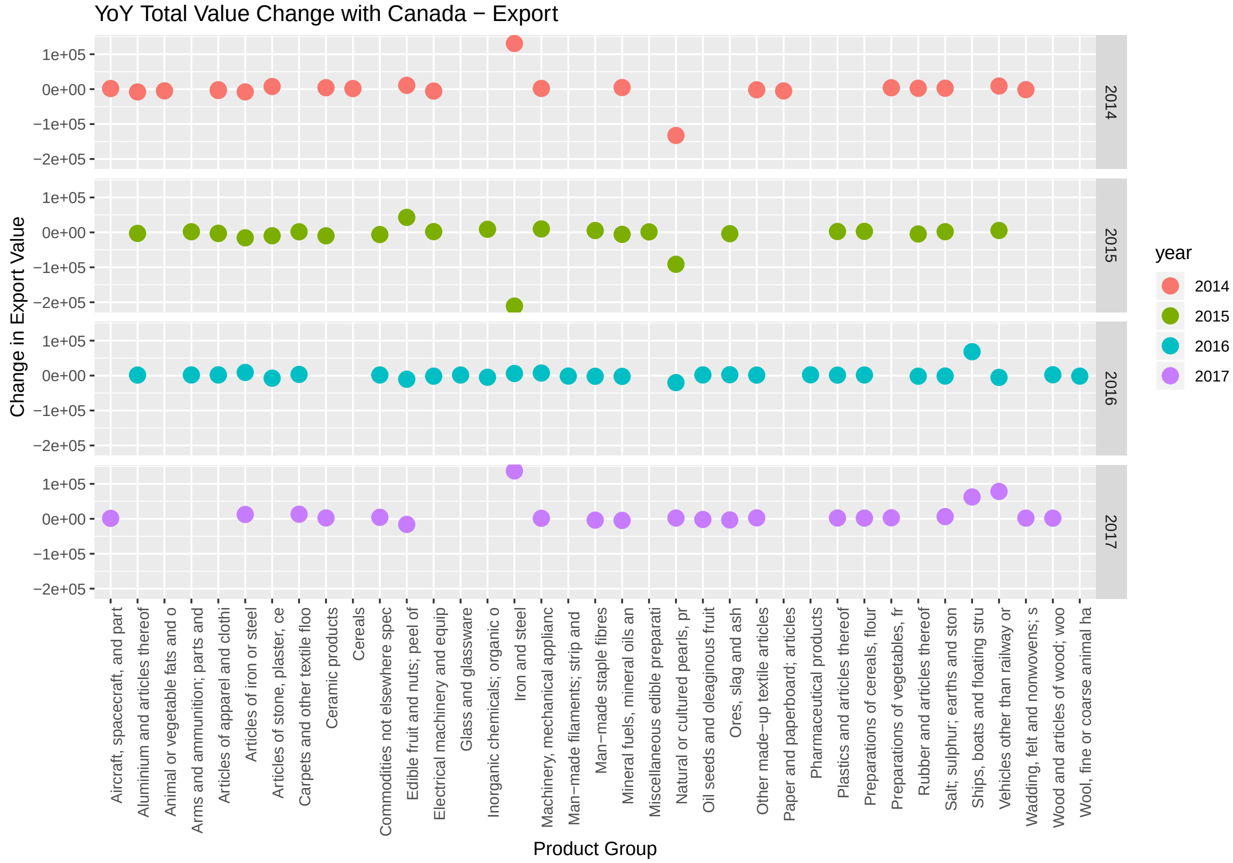


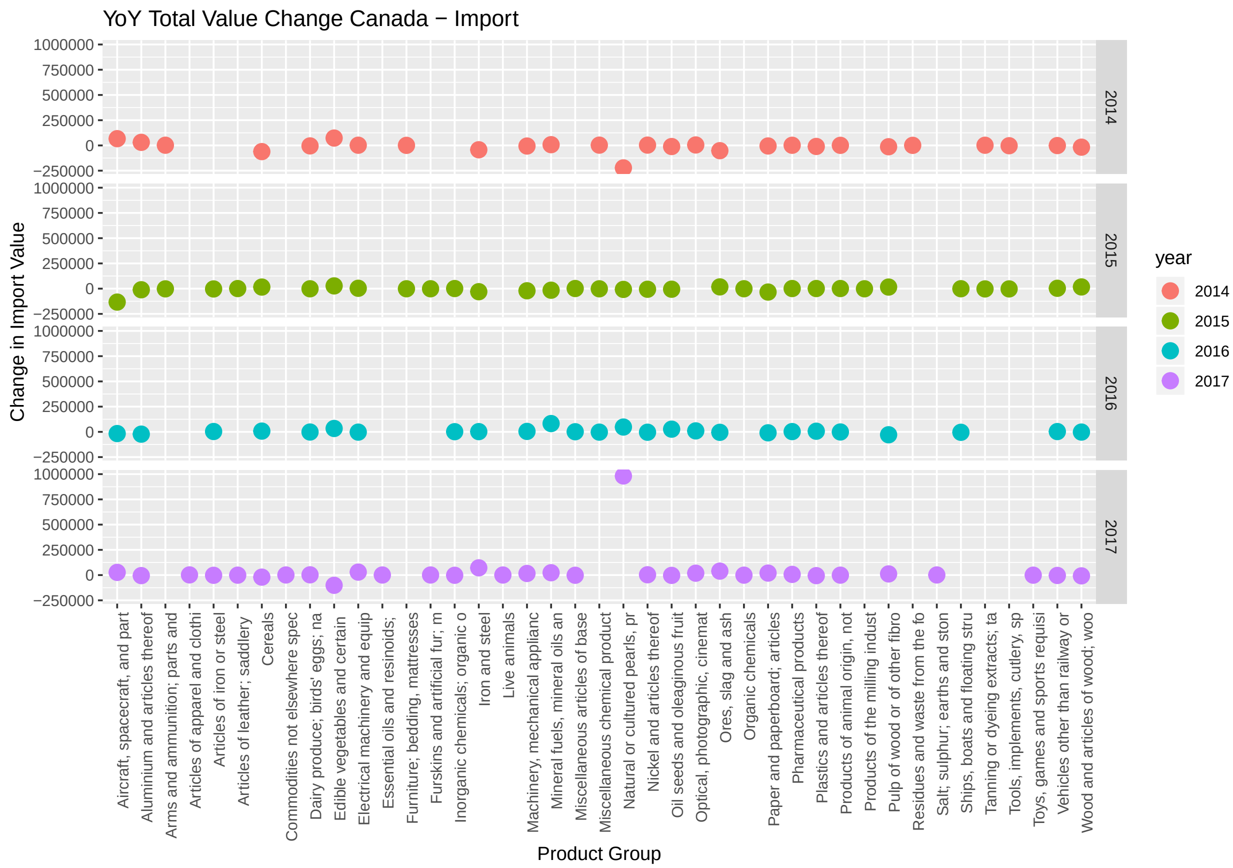
From the graph, it is observable that especially in the years of 2015, 2016 and 2017 exports from Russia and Canada deviates strongly, up to above %50. And for the import side, again in 2016 and 2017 seems that Russia and Canada are top deviators and suprisingly in 2017 imports from Canada rises more than 100%.

So, we should check which products drives these fluctuations. In order to visualize it, we found the highest delta products through filtering the ones exceeding 2 interquantile range (IQR - the difference between third and first quantile).

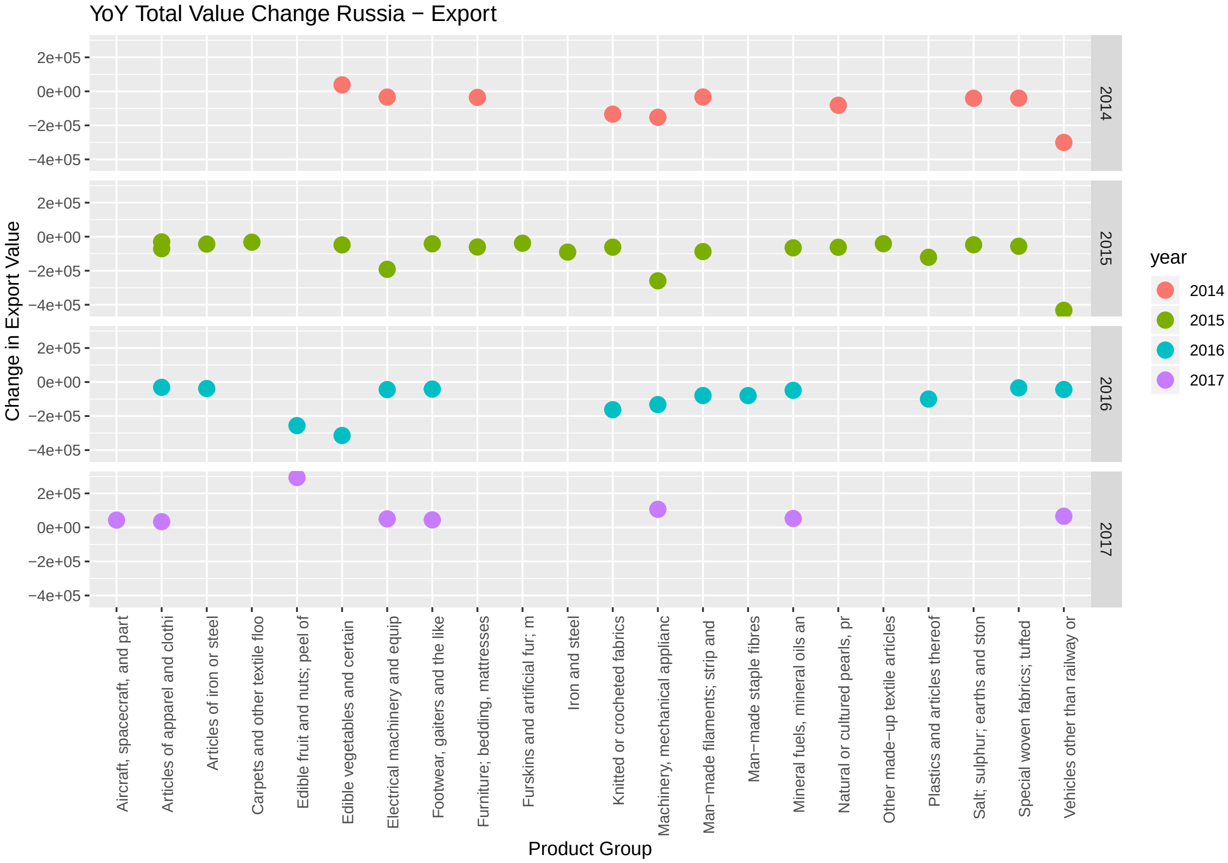
```r
# Canada and Russia significantly deviates. Investigate these countries in detail with traded products

# First Canada

# Export Analysis

summarytableCanadaEx <- export_jTableMELTED %>%
  filter(Country=="Canada") %>%
  group_by(X__2, year) %>%
  summarise(ExportValue=sum(Values)) %>%
  arrange(X__2, year) %>%
```

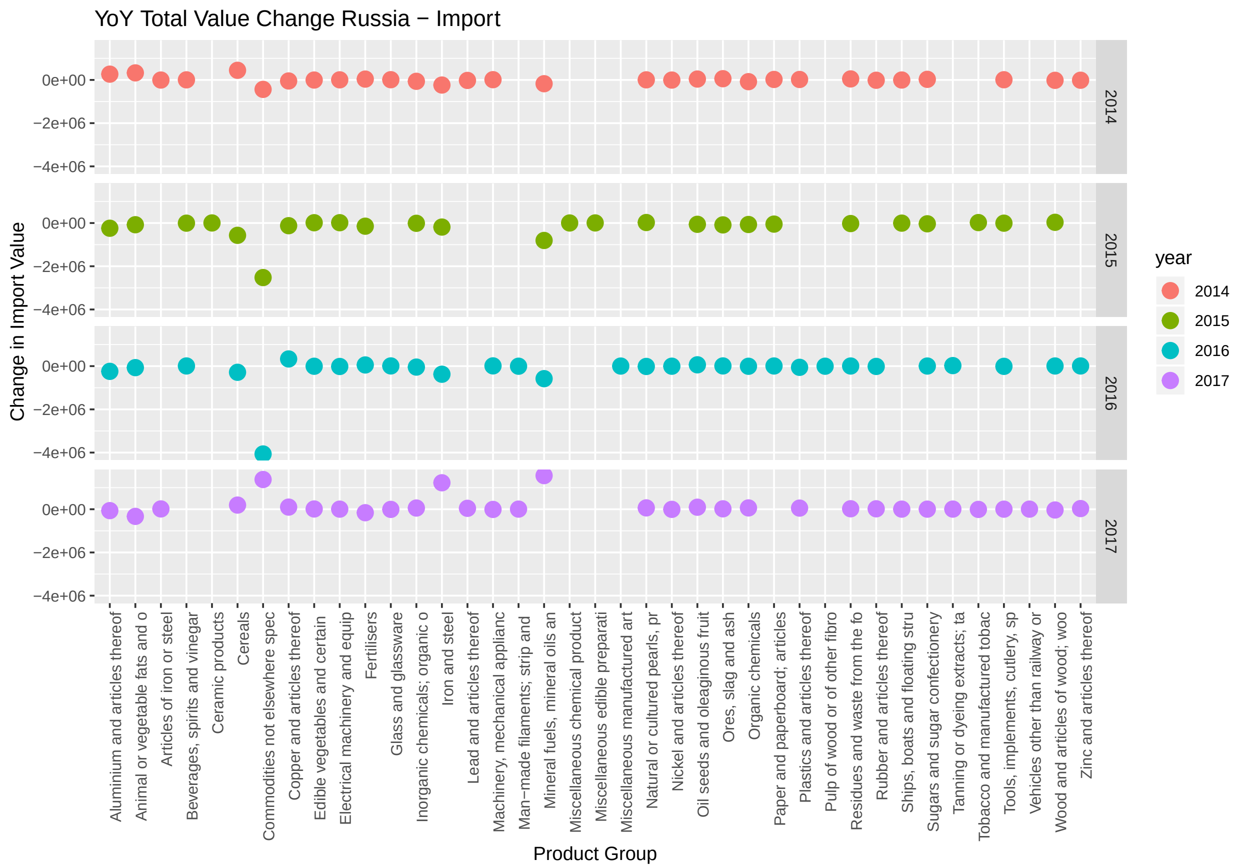
```
  mutate(ExpPercChngYoY = (ExportValue - lag(ExportValue))/lag(ExportValue)*100) %>%
  mutate(ExpChngYoy=((ExportValue - lag(ExportValue)))) %>%
  filter(year!=2013)


#mean(summarytableCanadaEx$ExpChngYoy)
#median(summarytableCanadaEx$ExpChngYoy)
#sd(summarytableCanadaEx$ExpChngYoy)
#quantile(summarytableCanadaEx$ExpChngYoy, probs = c(0.25, 0.50, 0.75, 1))
IQR= quantile(summarytableCanadaEx$ExpChngYoy, probs = c(0.75)) - quantile(summarytableCanadaEx$ExpChng
altsinir=quantile(summarytableCanadaEx$ExpChngYoy, probs = c(0.25)) - (2*IQR)
ustsinir=quantile(summarytableCanadaEx$ExpChngYoy, probs = c(0.75)) + (2*IQR)


summarytableCanadaEx <- summarytableCanadaEx %>%
  filter(abs(ExpChngYoy)>max(abs(altsinir),ustsinir))


plotCanEx <- ggplot(summarytableCanadaEx, aes(x=substr(X__2, start = 1, stop = 30))) +
  geom_point(aes(y=ExpChngYoy, color=year), size=4) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  facet_grid (~year) +
  facet_grid(rows = vars(year)) +
  labs(title="YoY Total Value Change with Canada - Export",
       x="Product Group",
       y="Change in Export Value",
       Color=c("Year"))
```


YoY Total Value Change with Canada − Export
Change in Export Value
Product Group
year
2014
2015
2016
2017
1e+05
0e+00
−1e+05
−2e+05
Aircraft, spacecraft, and part
Aluminium and articles thereof
Animal or vegetable fats and o
Arms and ammunition; parts and
Articles of apparel and clothi
Articles of iron or steel
Articles of stone, plaster, ce
Carpets and other textile floo
Ceramic products
Cereals
Commodities not elsewhere spec
Edible fruit and nuts; peel of
Electrical machinery and equip
Glass and glassware
Inorganic chemicals; organic o
Iron and steel
Machinery, mechanical applianc
Man−made filaments; strip and
Man−made staple fibres
Mineral fuels, mineral oils an
Miscellaneous edible preparati
Natural or cultured pearls, pr
Oil seeds and oleaginous fruit
Ores, slag and ash
Other made−up textile articles
Paper and paperboard; articles
Pharmaceutical products
Plastics and articles thereof
Preparations of cereals, flour
Preparations of vegetables, fr
Rubber and articles thereof
Salt; sulphur; earths and ston
Ships, boats and floating stru
Vehicles other than railway or
Wadding, felt and nonwovens; s
Wood and articles of wood; woo
Wool, fine or coarse animal ha



YoY Total Value Change Canada − Import
Change in Import Value
1000000
750000
500000
250000
0
−250000
2014
2015
2016
2017
year
Aircraft, spacecraft, and part
Aluminium and articles thereof
Arms and ammunition; parts and
Articles of apparel and clothi
Articles of iron or steel
Articles of leather; saddlery
Cereals
Commodities not elsewhere spec
Dairy produce; birds' eggs; na
Edible vegetables and certain
Electrical machinery and equip
Essential oils and resinoids;
Furniture; bedding, mattresses
Furskins and artificial fur; m
Inorganic chemicals; organic o
Iron and steel
Live animals
Machinery, mechanical applianc
Mineral fuels, mineral oils an
Miscellaneous articles of base
Miscellaneous chemical product
Natural or cultured pearls, pr
Nickel and articles thereof
Oil seeds and oleaginous fruit
Optical, photographic, cinemat
Ores, slag and ash
Organic chemicals
Paper and paperboard; articles
Pharmaceutical products
Plastics and articles thereof
Products of animal origin, not
Products of the milling indust
Pulp of wood or of other fibro
Residues and waste from the fo
Salt; sulphur; earths and ston
Ships, boats and floating stru
Tanning or dyeing extracts; ta
Tools, implements, cutlery, sp
Toys, games and sports requisi
Vehicles other than railway or
Wood and articles of wood; woo
Product Group


YoY Total Value Change Russia − Export
Change in Export Value
2e+05
0e+00
−2e+05
−4e+05
2014
2015
2016
2017
year
Aircraft, spacecraft, and part
Articles of apparel and clothi
Articles of iron or steel
Carpets and other textile floo
Edible fruit and nuts; peel of
Edible vegetables and certain
Electrical machinery and equip
Footwear, gaiters and the like
Furniture; bedding, mattresses
Furskins and artificial fur; m
Iron and steel
Knitted or crocheted fabrics
Machinery, mechanical applianc
Man−made filaments; strip and
Man−made staple fibres
Mineral fuels, mineral oils an
Natural or cultured pearls, pr
Other made−up textile articles
Plastics and articles thereof
Salt; sulphur; earths and ston
Special woven fabrics; tufted
Vehicles other than railway or
Product Group



It is concluded in above graphs that;

- The observations of change of export to Canada in 2015 and 2017 are due to export of "Iron and Steel".
- The observation of change of import from Canada in 2017 is due to import of "Natural or cultured pearls…".
- The decrease in export to Russia in 2015 drives by "Vehicles other than railway..".
- The decrease in export to Russia in 2016 and increase in 2017 is the effect of "Edible fruits and vegetables".
- Imports from Russia deviates due to "Commodities not elsewhere speecified".

Besides this analysis, we can discover more on this topic with different perspectives. To do this, we decided to build up shiny app with dependents of Country, Product and Year. Please follow the link to access to Shiny app. ***Link***

This image taken from Shiny app and presents the trade values of edible vegetables over years. It is observable that, in 2016 Turkey's edible vegetable export to Russia dramatically downs.

Analysis of Trade Deficit by Countries and Years

To discover the country's effects to trade deficit, we calculated net_value which formulated as 'Export_value - Import_value'. And visualized that value on the basis of year.

First, we grouped and manipulated the data and continued with visulation;

```
export_Product_Group <- export_jTableMELTED%>%
group_by(Country,PcodeTwo,X__2,year)%>%
summarize(Quantity = sum(Quantity), Values = sum(Values))%>%
rename(Product_Group_Code =PcodeTwo, Product_Group_Name = X__2,Export_Quantity = Quantity,Export_Value
```

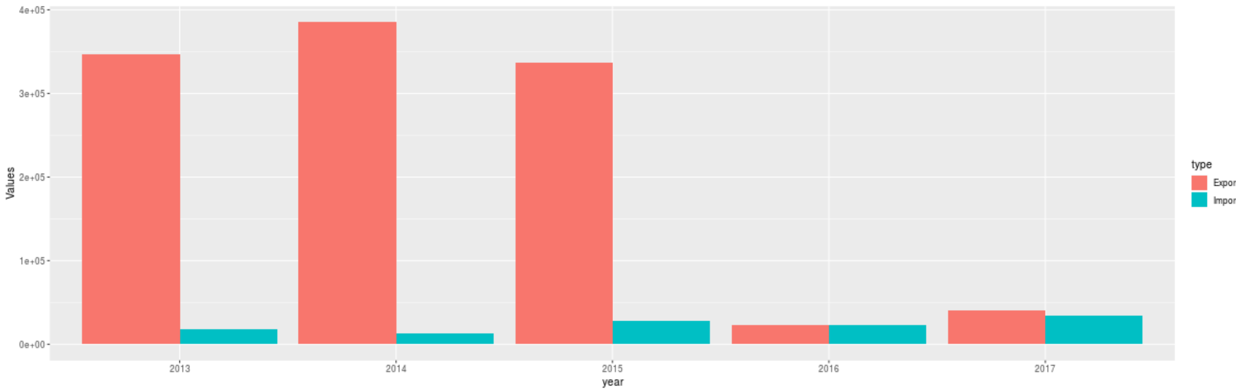


Figure 1: Edible vegetable export to Russia

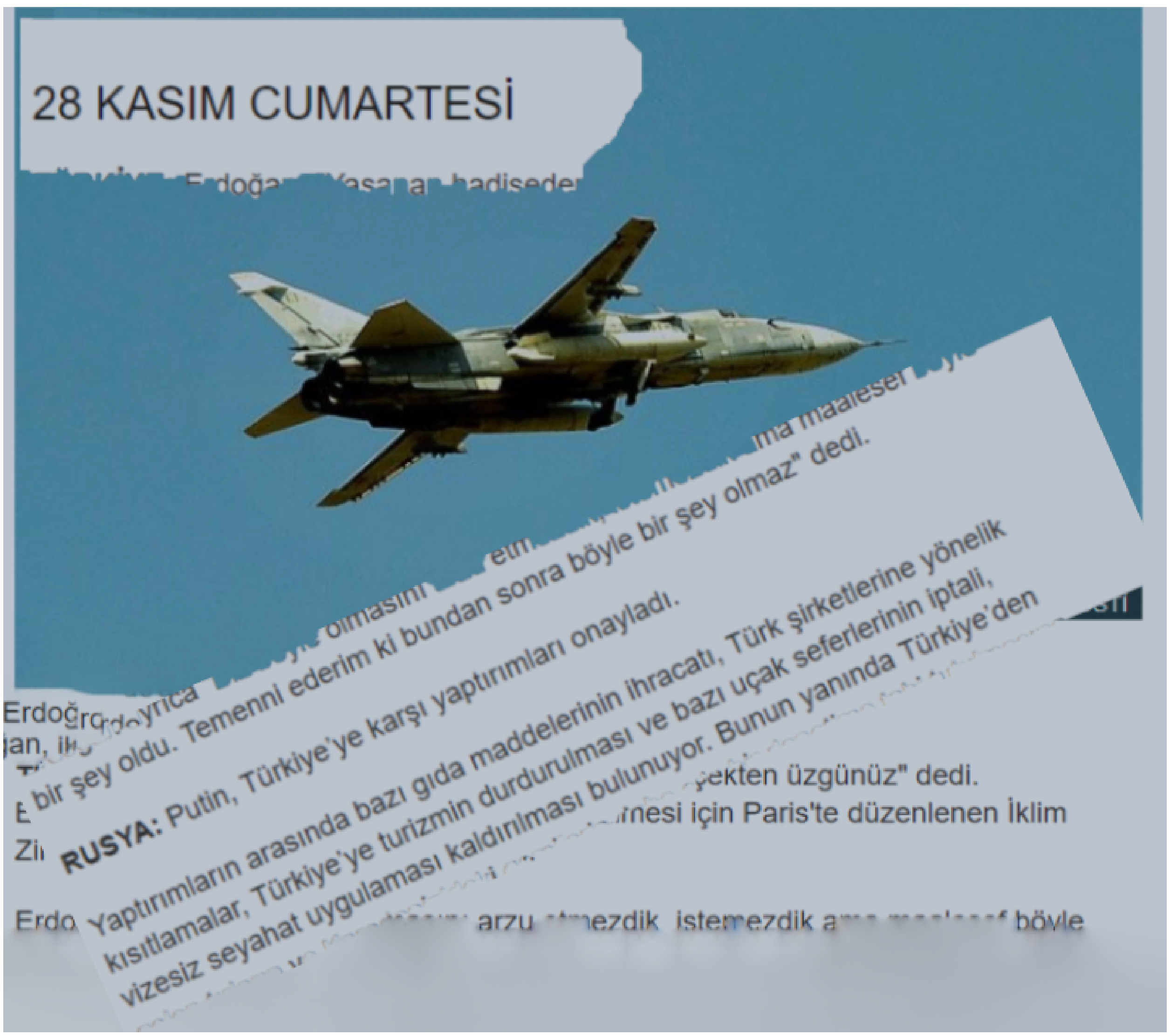
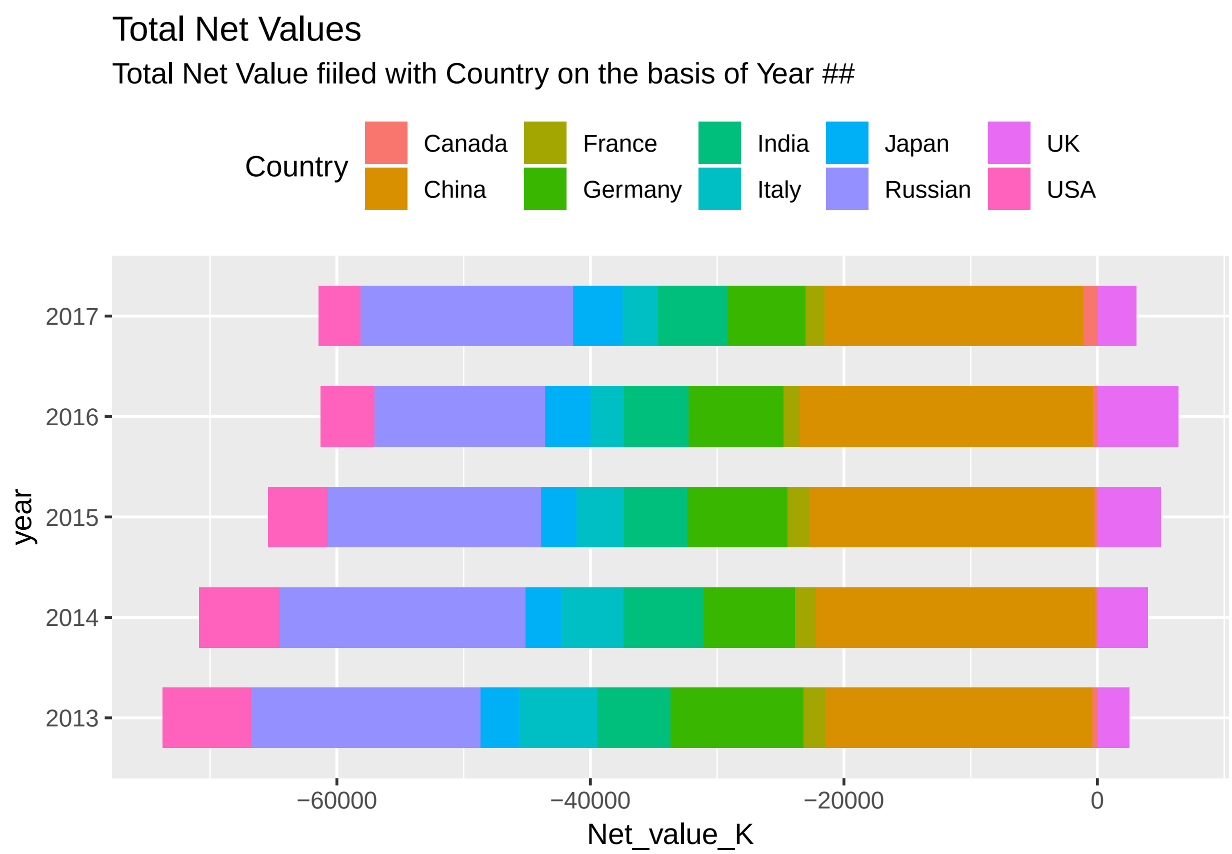


Figure 2: Russian sanctions to Turkey

```
import_Product_Group <- import_jTableMELTED%>%
group_by(Country,PcodeTwo,X__2,year)%>%
summarize(Quantity = sum(Quantity), Values = sum(Values))%>%
rename(Product_Group_Code =PcodeTwo, Product_Group_Name = X__2,Import_Quantity = Quantity,Import_Value =

Export_Import_JTable <- export_Product_Group %>%
  full_join(import_Product_Group,by=c("Country","Product_Group_Code","Product_Group_Name","year")) %>%
mutate(Import_Quantity = replace(Import_Quantity, is.na(Import_Quantity),0),
        Export_Quantity = replace(Export_Quantity, is.na(Export_Quantity),0))%>%
select(Country,Product_Group_Code,Product_Group_Name,year,Export_Quantity,Import_Quantity,Export_Value,I
```

```
####Net Value By Country###
Net_Value_By_Country <- Export_Import_JTable %>%
group_by(Country,year)%>%
summarize (Export_Value = sum(Export_Value), Import_Value = sum(Import_Value),
           Net_Value = sum(Export_Value - Import_Value),Net_value_K = round((sum(Export_Value - Import_V


# Total Net Value fiiled with Country on the basis of Year #
ggplot(Net_Value_By_Country, aes(x=year, y=Net_value_K)) +
  geom_bar(stat='identity', aes(fill=Country), width=.6,position = position_stack(reverse = TRUE))  +
  labs(subtitle="Total Net Value fiiled with Country on the basis of Year ##", title= "Total Net Values
  coord_flip()+
  theme(legend.position = "top")
```

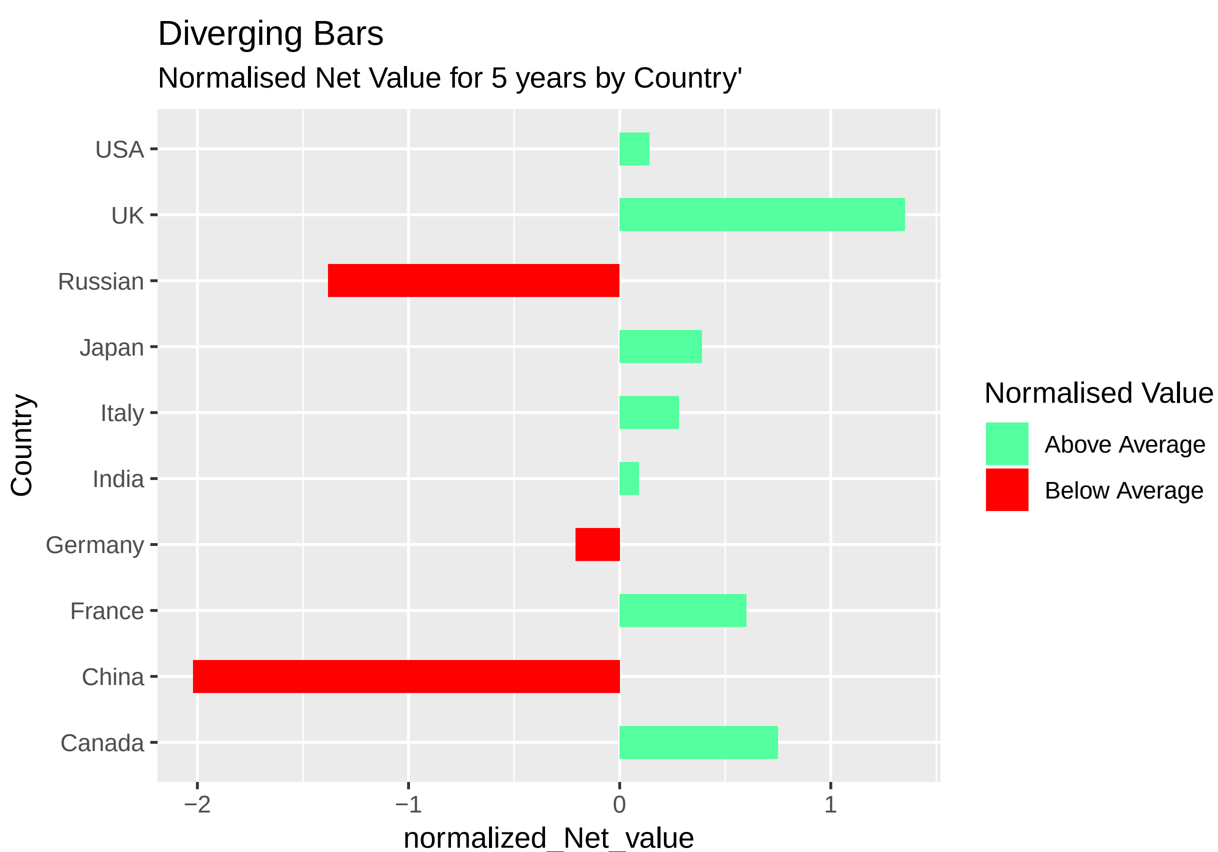


As seen in the graph, only UK gives trade surplus for all years analysed. On the other hand China, Russia, Germany and USA have the most significant effect on trade deficit.

An Alternative perspective to explore effects of countries

The effects of countries are changing year by year. So that, as an alternative perspective we decided to take average net value for all countries in five years and normalized that value to show which country's effects are more than others.

```
#Normalized of Average Values for five years
Normalized_Values_By_Country <- Export_Import_JTable %>%
  #filter (year == 2017)%>%
  group_by(Country)%>%
  summarize (Net_Value = mean(Export_Value - Import_Value))%>%
  mutate(normalized_Net_value =  round((Net_Value  - mean(Net_Value))/sd(Net_Value ),2)) %>%
  mutate(value_Net_type =  ifelse(normalized_Net_value < 0, "below", "above"))

#Diverging Bars for normalized_Export_value
ggplot(Normalized_Values_By_Country, aes(x=Country, y=normalized_Net_value , label=normalized_Net_value)
  geom_bar(stat='identity', aes(fill=value_Net_type), width=.5)  +
  scale_fill_manual(name="Normalised Value",
                    labels = c("Above Average", "Below Average"),
                    values = c("above"="seagreen1", "below"="Red")) +
  labs(subtitle="Normalised Net Value for 5 years by Country'",
       title= "Diverging Bars") +
  coord_flip()
```



Considering the 5 years period, it is clear that China, Germany and Russia has contribution to trade deficit

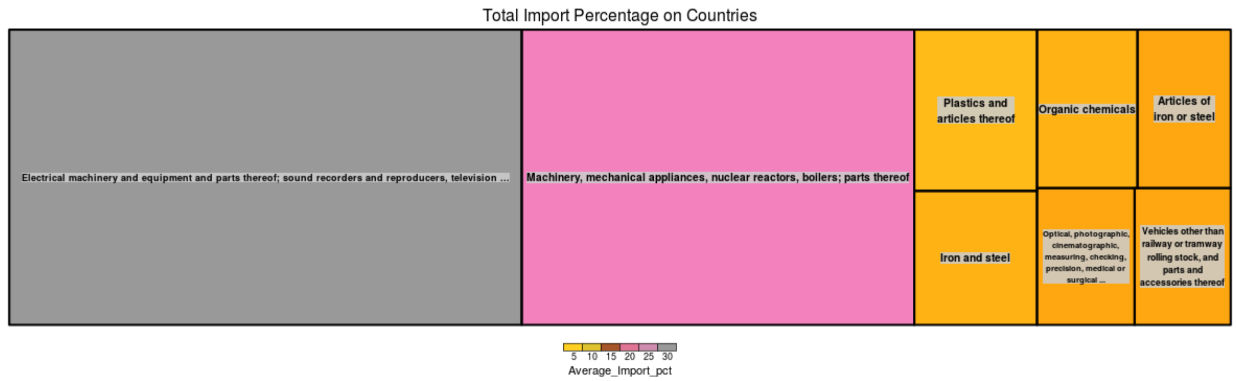


Figure 3: Import from China

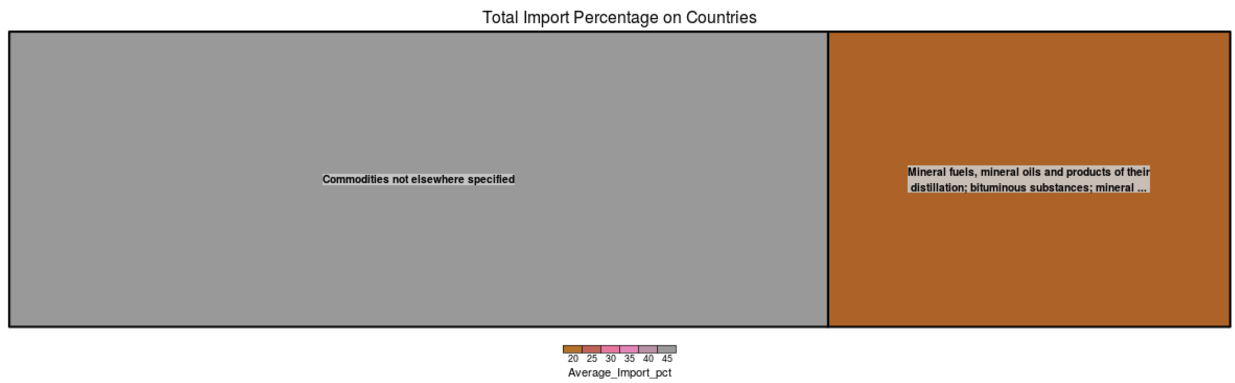


Figure 4: Import from Russia

above average.

We drilled down the analysis for the countries which decreases the average, namely China, Russia and Germany. For this aim, we build up shiny app to review the products by defining the Country and ratio showing cumulative percentage of share of products over total import value of this country.

Please follow the link for interactive graph: ***Link***

Through the shiny app, it is understood that

- 70% of import value from China is coming from 8 products while 50% is from electrical machinery, equipments and machinery, mechanical appliances
- Mineral fuels, mineral oils and products of their distillation constitute %21 of import from Russia, while the bigggest product group is classified as "Other""
- Vehicles other than railway is the top product imported from Germany and mechanical and electronical machinery and equipments are followed it.

Correlation of Trade Deficit with FX Rate and BIST100 index

In order to understand the correlation between trade deficit/surplus and BIST100 index and USD/TL rates, we added this economic indicators to our analysis:

First we read the USD and BIST100 index values and rename the columns; Then proceeded with drawing the graph to represent USD/TL rate, BIST100 index and Trade Deficit in time series.

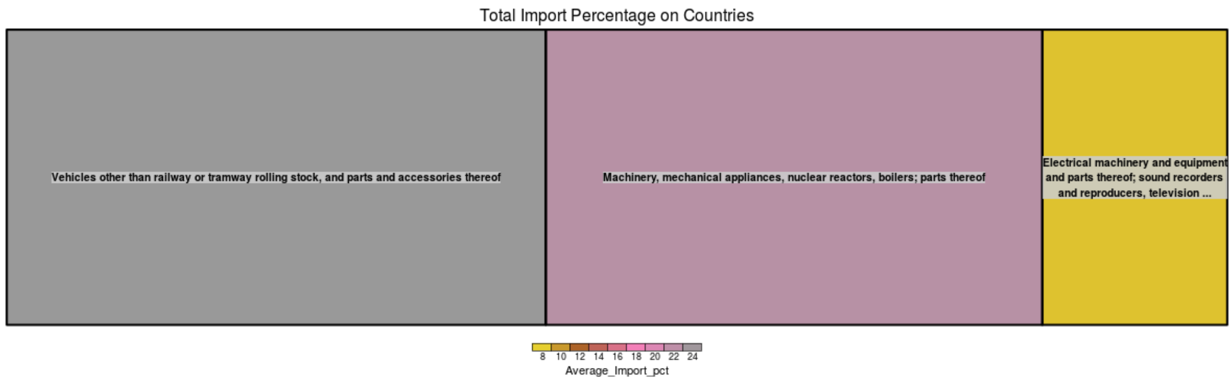
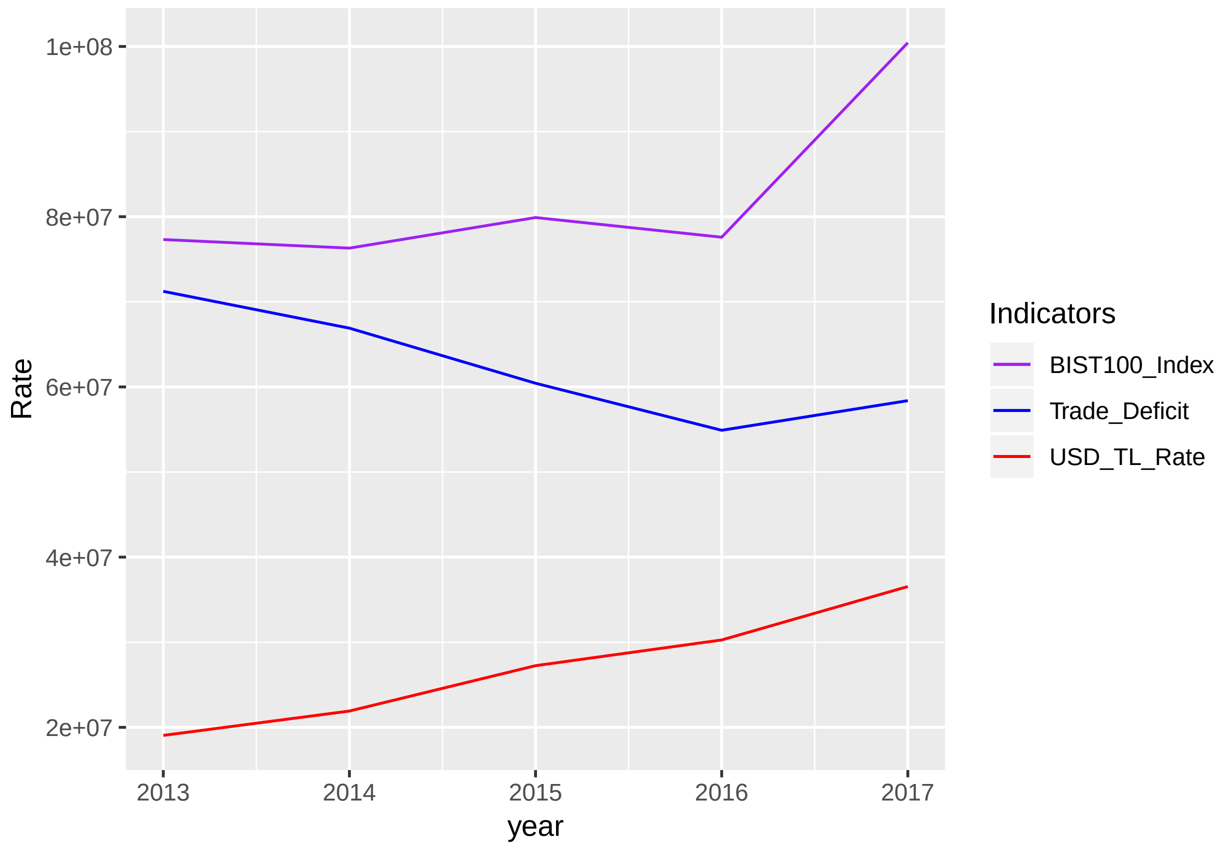


Figure 5: Import from Germany

```
fileUSD= "C:/Users/emrek/Google Drive/BDA/503-EssentialsOfDataAnalytics/GitHub/gpj18-group_four/EVDS.xl
usdR <- read_excel(fileUSD, col_names=FALSE, skip=1)


colnames(usdR) <- c("DateR","ExchangeRate")
usd <- na.omit(usdR)
remove(usdR)

usdN<-usd %>% mutate(Date=dmy(DateR))
usdM<- usdN %>% group_by(year=year(Date)) %>% summarise(Rate=mean(ExchangeRate)*10000000)


fileBIST= "C:/Users/emrek/Google Drive/BDA/503-EssentialsOfDataAnalytics/GitHub/gpj18-group_four/BIST10
bist100R <- read_excel(fileBIST, col_names=FALSE, skip=1)


colnames(bist100R) <- c("Date", "Year","Value", "Açılış",   "Yüksek",   "Düşük",    "Hac.", "Fark %")
bist100 <- bist100R  %>% select(Date,Year,Value) %>%
  group_by(Year) %>%
  summarise(Value = mean(Value)*1000)
```

```
sumImport<- import_jTable %>% group_by(Country) %>% summarise(Value_2013=sum(Value_2013),Value_2014=sum
sumExport<- export_jTable %>% group_by(Country) %>% summarise(Value_2013=sum(Value_2013),Value_2014=sum
joinedTableDeficit<-sumExport %>%  inner_join(sumImport,by="Country") %>%
  mutate(Value_2013=Value_2013.y-Value_2013.x,Value_2014=Value_2014.y-Value_2014.x, Value_2015=Value_20
         Value_2017=Value_2017.y-Value_2017.x) %>%
  select(Country, Value_2013, Value_2014, Value_2015, Value_2016, Value_2017)

joinedTableDeficitMelt <- melt(joinedTableDeficit, id = c("Country"))
deficitPerCountry<-joinedTableDeficitMelt %>% mutate(year = case_when( variable == "Value_2013" ~ 2013,
                                                                       variable == "Value_2014" ~ 2014,
                                                                       variable == "Value_2015" ~ 2015,
                                                                       variable == "Value_2016" ~ 2016,
                                                                       variable == "Value_2017" ~ 2017)
  select(Country,year,value)

deficitPerYear <- deficitPerCountry %>%
  group_by(year) %>% summarise(value=sum(value))
```

```
p<-ggplot() +
  geom_line(data=usdM,aes(x=year, y=Rate,group = 1, colour="USD_TL_Rate"))+
  geom_line(data=deficitPerYear,aes(x=year, y=value,group = 1, colour="Trade_Deficit"))+
  geom_line(data=bist100,aes(x=Year, y=Value,group = 1,colour="BIST100_Index")) +
  scale_colour_manual(name="Indicators",values=c(USD_TL_Rate="red", Trade_Deficit="blue", BIST100_Index=


print(p)
```



It is concluded that trade deficit has negatively correlated with USD/TL rate between 2013 and 2016 as expected and for the year 2017 we saw the effect of stagflation - while the USD/TL rate increasing (accordingly we know that inflation rate is strongly positively correlatied with the inflation rate (inf rate: 2015-12:5.71%, 2016-12:9.94%, 2017-12:15.47%)) and weak or negative economic growth. Besides, there is not a direct correlation between BIST100 index and Trade Deficit Values.

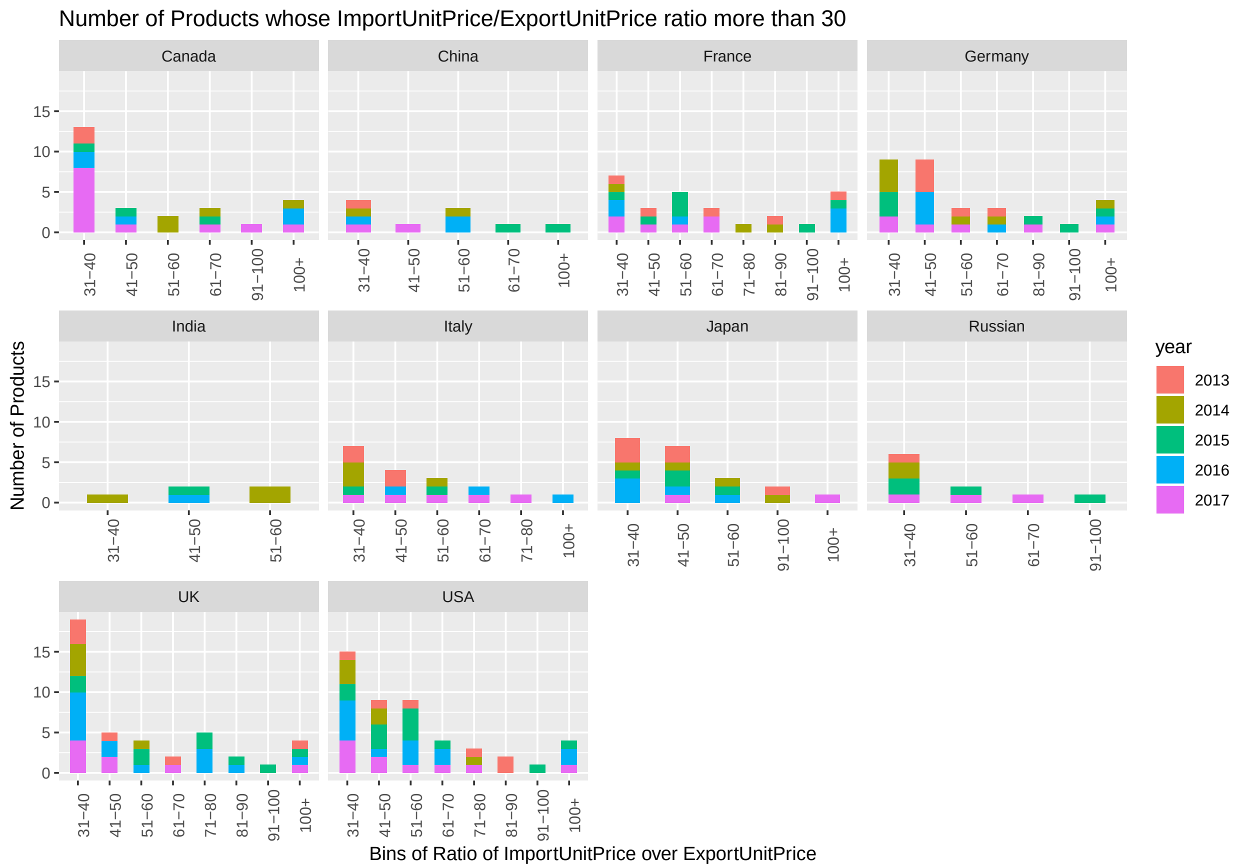
Analysis of the Import Unit Prices and Export Unit Prices

In order to understand the products which Turkey is both exporting and importing but importing with more value addition, we calculated unit prices on each product for both import and export and we followed finding the ratio of Import Unit Price / Export Unit Price and groping them in bins of 10.

```
unit_price_import <- import_jTableMELTED %>% select(Country,Product_Label, X__2 ,Product_Label, Product_
  mutate(Unit_Price = Values/Quantity)
unit_price_export <- export_jTableMELTED %>% select(Country,Product_Label, X__2 ,Product_Label, Product_
```

```r
  mutate(Unit_Price = Values/Quantity)

unit_price_import <- unit_price_import %>% mutate(type="Import")
unit_price_export <- unit_price_export %>% mutate(type="Export")

joint_table <- unit_price_export %>% select(Country,Product_Label,year,Export_Unit_Price=Unit_Price) %>%
  full_join(unit_price_import, by =c("Country","Product_Label","year")) %>% select(Country,Product_Label
  mutate(Ratio=Import_Unit_Price/Export_Unit_Price)
melted_data <- joint_table %>% melt(id=c("Country","Product_Label","year","Ratio")) %>% rename("type"=va
  filter(value  >0 & Ratio >0 & Ratio !=Inf ) #& Ratio <=1000 & value <=10000)

melted_data <- melted_data %>% mutate(Bins=0)
melted_data$Bins <- cut(melted_data$Ratio, breaks=c(1,10,20,30,40,50,60,70,80,90,100,1000), labels=c("1-
melted_data <- melted_data %>% filter((!is.na(Bins) & Bins!="1-10" &  Bins!="11-20" &  Bins!="21-30") )
  select(Country,Product_Label, year, Ratio, Bins)
melted_data <- distinct(melted_data)

 #
```

```r
ggplot(melted_data, aes(Bins)) +
  geom_bar(aes(fill=year), width = 0.5) +

  facet_wrap(~Country, scales = "free_x") +
  labs(title="Number of Products whose ImportUnitPrice/ExportUnitPrice ratio more than 30",
       y="Number of Products",
       x="Bins of Ratio of ImportUnitPrice over ExportUnitPrice") +
  theme(axis.text.x = element_text(angle=90))
```



As seen in the graph;

- France, Germany, UK and US are the strongest countries in terms of giving strong positive trade balance with variety of product groups while India is the weakest.

In order to examine the details, we builded ***Shiny*** app.

- Highest ratio products for India is "Yeasts, active or inactive; other dead single-cell micro-organisms" and "Pharmaceutical preparations".
- "Pharmaceutical preparations" is coming in top list for most of the countries.
- USA and UK have 47 and 42 products that have ratio more than 30 and these countries are top 2.

K-Means Clustering

cvv

Conclusion

cvv

References

TradeMap

TCMB

TUIK

Top 50 ggplot2 Visualizations

R Markdown: The Definitive Guide

R-statistics blog