

```

#loading libraries
library(dplyr)

#loading the dataset
trade <- read.csv("Raw Data.csv")

#View dataset
View(trade) #there are zeros in Weight and Quantity columns

#make a copy and work on the copy
business <- trade

#Modify names of columns for clarity
names(business) <- c("Country", "Year", "Commoditycode", "Commodity", "Flow",
                    "Dollars", "Weight", "Quantityname", "Quantity", "Category")

#structure of data
str(business)

#summary of data
summary(business) #Missing values present in Weight and Quantity
#There is a "No Quantity" value in Quantityname
#No Quantity indicates Quantity is either 0 or NA

#Approach 1: Eliminate all missing values
trader <- na.omit(trade) #if missing values are eliminated, we have 58058 obs.

#Approach 2: Understand why there are missing values

#create new columns for missing weights and quantities
#When Weight is zero and Quantity is also zero, output is 1 else 0
business$zeroWQ <- ifelse(business$Weight < 1 & business$Quantity < 1, "1", "0")
table(business$zeroWQ) # there are 847 rows that account to this scenario
business$zeroWQ <- NULL

#When Weight is zero and Quantity is has NA, output is 1 else 0
#we can discard this condition
business$zeroW_naQ <- ifelse(business$Weight < 1 & business$Quantity == NA, "1", "0")

table(business$zeroW_naQ) #56590 observations says "0"
#From analysis of the dataset there are no such scenario.
#However, the presence of NAs in Quantity accounts for the absence of remaining observations.
#We can discard this scenario.

```

```

business$zeroW_naQ <- NULL

#combine multiple scenarios:

#When Weight is zero, or NAs in Weight, or Quantity is zero, or NAs in Quantity - assign 1
business$Missing <- ifelse(business$Weight < 1 | is.na(business$Weight)
                           | business$Quantity < 1 | is.na(business$Quantity), "1", "0")
table(business$Missing)

#there are 2819 cases, where Weight and Quantity have either zero value or NA
#This column will help us identify the Trade value(in USD) incurred for missing info.

#understand the consequences of missing info by country
missedinfo_au <- filter(business, Country == "Australia", Missing == 1)
nrow(missedinfo_au) # 446 rows have missing info on Weight and Quantity columns
sum(missedinfo_au$Dollars) # 13.5 billion US Dollars (13586563312)

missedinfo_can <- filter(business, Country == "Canada", Missing == 1)
nrow(missedinfo_can) # 1429 rows have missing info on Weight and Quantity
sum(missedinfo_can$Dollars) #32.5 billion US Dollars (32510198277)

missedinfo_usa <- filter(business, Country == "USA", Missing == 1)
nrow(missedinfo_usa) # 994 rows have missing info on Weight and Quantity
sum(missedinfo_usa$Dollars) #36.9 billion US Dollars (36952700815)

#understand the consequences of missing info by Flow
#AUSTRALIA
aus_export <- filter(missedinfo_au, Flow == "Export")
nrow(aus_export) # 52 rows have missing info on Weight and Quantity
sum(aus_export$Dollars) #(12032593547) 12 Billion US Dollars from rows containing missing info

aus_import <- filter(missedinfo_au, Flow == "Import")
nrow(aus_import) # 383 rows have missing info on Weight and Quantity
sum(aus_import$Dollars) #(1552825617) 1.5 Billion US Dollars from rows containing missing info

aus_reimport <- filter(missedinfo_au, Flow == "Re-Import")
nrow(aus_reimport) # 8 rows have missing info on Weight and Quantity
sum(aus_reimport$Dollars) #(938823) 938,823 US Dollars from rows containing missing info

aus_reexport <- filter(missedinfo_au, Flow == "Re-Export")
nrow(aus_reexport) # 3 rows have missing info on Weight and Quantity
sum(aus_reexport$Dollars) #(205325) 205,325 US Dollars from rows containing missing info

#CANADA

```

```

can_export <- filter(missedinfo_can, Flow == "Export")
nrow(can_export) # 637 rows have missing info on Weight and Quantity
sum(can_export$Dollars) #(26578595902) 26.5 Billion US Dollars from rows containing missing info

can_import <- filter(missedinfo_can, Flow == "Import")
nrow(can_import) # 582 rows have missing info on Weight and Quantity
sum(can_import$Dollars) #(5838350510) 5.8 Billion US Dollars from rows containing missing info

can_reimport <- filter(missedinfo_can, Flow == "Re-Import")
nrow(can_reimport) # 100 rows have missing info on Weight and Quantity
sum(can_reimport$Dollars) #(3676340) 3.6 Million US Dollars from rows containing missing info

can_reexport <- filter(missedinfo_can, Flow == "Re-Export")
nrow(can_reexport) # 110 rows have missing info on Weight and Quantity
sum(can_reexport$Dollars) #(89575525) 89 Million US Dollars from rows containing missing info

#USA
usa_export <- filter(missedinfo_usa, Flow == "Export")
nrow(usa_export) # 385 rows have missing info on Weight and Quantity
sum(usa_export$Dollars) #(17946793912) 17.9 Billion US Dollars from rows containing missing info

usa_import <- filter(missedinfo_usa, Flow == "Import")
nrow(usa_import) # 305 rows have missing info on Weight and Quantity
sum(usa_import$Dollars) #(18418691692) 18.4 Billion US Dollars from rows containing missing info

usa_reimport <- filter(missedinfo_usa, Flow == "Re-Import")
nrow(usa_reimport) # 0 rows have missing info on Weight and Quantity
sum(usa_reimport$Dollars) #0 US Dollars from rows containing missing info

usa_reexport <- filter(missedinfo_usa, Flow == "Re-Export")
nrow(usa_reexport) # 254 rows have missing info on Weight and Quantity
sum(usa_reexport$Dollars) #(587215211) 5.8 Billion US Dollars from rows containing missing info

#UNDERSTAND THE TOTAL USD for each country from business dataset
#AUSTRALIA
totalusd_au <- filter(business, Country == "Australia", Dollars >= 1)
nrow(totalusd_au) #Australia occupies 24,921 rows in the dataset
sum(totalusd_au$Dollars) #Total US dollars account to 458 billion

#excluding rows which have missing info
usd_au <- filter(totalusd_au, Missing == 0)
nrow(usd_au) #24475 rows with complete info
sum(usd_au$Dollars)# 445 Billion US Dollars (Difference)
options(scipen = 999)

```

```
#filter above by Flow
```

```
expo_au <- filter(usd_au, Flow == "Export")  
nrow(expo_au) #11133 rows with complete info  
sum(expo_au$Dollars)# 382 Billion US Dollars
```

```
imp_au <- filter(usd_au, Flow == "Import")  
nrow(imp_au) #9964 rows with complete info  
sum(imp_au$Dollars)# 62.6 Billion US Dollars
```

```
reimp_au <- filter(usd_au, Flow == "Re-Import")  
nrow(reimp_au) #1335 rows with complete info  
sum(reimp_au$Dollars)# 0.27 Billion US Dollars
```

```
reexpo_au <- filter(usd_au, Flow == "Re-Export")  
nrow(reexpo_au) #2043 rows with complete info  
sum(reexpo_au$Dollars)# 0.19 Billion US Dollars
```

```
#CANADA
```

```
totalusd_can <- filter(business, Country == "Canada", Dollars >= 1)  
nrow(totalusd_can) #Canada occupies 29,932 rows in the dataset  
sum(totalusd_can$Dollars) #Total US dollars account to 775 billion
```

```
#excluding rows which have missing info
```

```
usd_can <- filter(totalusd_can, Missing == 0)  
nrow(usd_can) #28,503 rows with complete info  
sum(usd_can$Dollars)# 742 Billion US Dollars (Difference)
```

```
#filter above by Flow
```

```
expo_can <- filter(usd_can, Flow == "Export")  
nrow(expo_can) # 10592 rows with complete info  
sum(expo_can$Dollars)# 488 Billion US Dollars
```

```
imp_can <- filter(usd_can, Flow == "Import")  
nrow(imp_can) #11248 rows with complete info  
sum(imp_can$Dollars)# 249 Billion US Dollars
```

```
reimp_can <- filter(usd_can, Flow == "Re-Import")  
nrow(reimp_can) #3307 rows with complete info  
sum(reimp_can$Dollars)# 0.5 Billion US Dollars
```

```
reexpo_can <- filter(usd_can, Flow == "Re-Export")  
nrow(reexpo_can) #3356 rows with complete info  
sum(reexpo_can$Dollars)# 4 Billion US Dollars
```

```
#USA
```

```
totalusd_usa <- filter(business, Country == "USA", Dollars >= 1)
nrow(totalusd_usa) #USA occupies 4237 rows in the dataset
sum(totalusd_usa$Dollars) #Total US dollars account to 601 billion
```

```
#excluding rows which have missing info
```

```
usd_usa <- filter(totalusd_usa, Missing == 0)
nrow(usd_usa) #3293 rows with complete info
sum(usd_usa$Dollars) # 564 Billion US Dollars (Difference)
```

```
#filter above by Flow
```

```
expo_usa <- filter(usd_usa, Flow == "Export")
nrow(expo_usa) #1144 rows with complete info
sum(expo_usa$Dollars) # 429 Billion US Dollars
```

```
imp_usa <- filter(usd_usa, Flow == "Import")
nrow(imp_usa) #1210 rows with complete info
sum(imp_usa$Dollars) # 133 Billion US Dollars
```

```
reimp_usa <- filter(usd_usa, Flow == "Re-Import")
nrow(reimp_usa) #0 rows with complete info
sum(reimp_usa$Dollars) # 0 US Dollars
```

```
reexpo_usa <- filter(usd_usa, Flow == "Re-Export")
nrow(reexpo_usa) #939 rows with complete info
sum(reexpo_usa$Dollars) # 1.9 Billion US Dollars
```

```
#INSIGHTS SO FAR
```

```
#USA has not re-imported any commodity from 1988 - 2016
```

```
#There have been import bills and export payouts for data having no info of commodities weight and quantity.
```

```
#USA leads with 36.9 billion US Dollars for data having missing info
```

```
#Canada follows USA with 32.5 billion US Dollars
```

```
#While Australia generated 13.5 billion US Dollars (Missing info highest in Export and lowest in Re export Flow)
```

```
#Total Trade in US Dollars in descending order (business dataset)
```

```
#CANADA - 775 BILLION
```

```
#USA - 601 BILLION
```

```
#AUS - 458 BILLION
```

```
#Excluding rows with missing info: Ranking in Largest Trade Share
```

```
#CANADA - 742 BILLION
```

```
#USA - 564 BILLION
```

```
#AUS - 445 BILLION
```

```
#Understanding Trade balance (data with no missing info)
```

```
#Trade balance = Total Value of Export - Total Value of Import
```

```
tradebalance_usa <- sum(expo_usa$Dollars) - sum(imp_usa$Dollars)
```

```
tradebalance_usa # 295 Billion Trade surplus
```

```
tradebalance_can <- sum(expo_can$Dollars) - sum(imp_can$Dollars)
```

```
tradebalance_can # 238 Billion Trade surplus
```

```
tradebalance_aus <- sum(expo_aus$Dollars) - sum(imp_aus$Dollars)
```

```
tradebalance_aus # 319 Billion Trade deficit
```

```
#Lets aggregate the data
```

```
tableau <- rbind(usd_aus,usd_can,usd_usa)
```

```
tableau$Missing <- NULL
```

```
#Considering the tidied data for the Tableau visualization
```

```
write.csv(tableau, "tableau.csv", row.names = FALSE)
```

```
#Calculating Trade balance for complete dataset (including missing info)
```

```
#AUSTRALIA
```

```
tb_impoaus <- filter(totalusd_aus, Flow == "Import")
```

```
sum(tb_impoaus$Dollars) #64 billion US dollars
```

```
tb_expoaus <- filter(totalusd_aus, Flow == "Export")
```

```
sum(tb_expoaus$Dollars) #394 billion US dollars
```

```
tb_aus <- sum(tb_expoaus$Dollars) - sum(tb_impoaus$Dollars)
```

```
tb_aus # 329 billion US Dollars Trade surplus
```

```
#CANADA
```

```
tb_impocan <- filter(totalusd_can, Flow == "Import")
```

```
sum(tb_impocan$Dollars) #255 billion US dollars
```

```
tb_expocan <- filter(totalusd_can, Flow == "Export")
```

```
sum(tb_expocan$Dollars) #514 billion US dollars
```

```
tb_can <- sum(tb_expocan$Dollars) - sum(tb_impocan$Dollars)
```

```
tb_can # 259 billion US Dollars Trade surplus
```

```
#USA
```

```
tb_impousa <- filter(totalusd_usa, Flow == "Import")
```

```
sum(tb_impousa$Dollars) #152 billion US dollars
```

```
tb_expousa <- filter(totalusd_usa, Flow == "Export")
sum(tb_expousa$Dollars) #447 billion US dollars

tb_usa <- sum(tb_expousa$Dollars) - sum(tb_impousa$Dollars)
tb_usa # 295 billion US Dollars Trade surplus

#Trade balance Ranking (descending) - complete dataset
#AUSTRALIA = 329 billion US dollars (Trade Surplus)
#USA = 295 billion US dollars (Trade Surplus)
#CANADA = 259 billion US dollars (Trade Surplus)
```