explanation

Qiannan(Nancy) Shen

10/7/2021

Introduction

The dietary health issues is a popular problem nowadays. Excess sugar consumption is one of the serious health problems that appeared in many countries. According to the American Heart Association (AHA), the maximum amount of added sugars you should eat in a day are 37.5 grams per day for male and 25 grams per day for female. However, in most countries, people would take in an amount of sugar which is far from the regular basis. As modern grocery shoppers, we try to be engaged and knowledgeable about nutrition. So, I did this analysis between The quantity of food consumption of sugar and sweeteners (grams per person and day) and calories measures the energy content of the food. And I try to find that the change in the sugar intake from 1960 to 2012 and test whether people prefer to eat sweeter food in a different country. So I choose two-country Solomon Islands and Thailand. Make a comparison of sugar, food supply, and the percentage.

source

```
##Read R Code from the file "wrangling_code.R" and "visualization_function.R"
source(file = "wrangling_code.R", echo = TRUE)
```

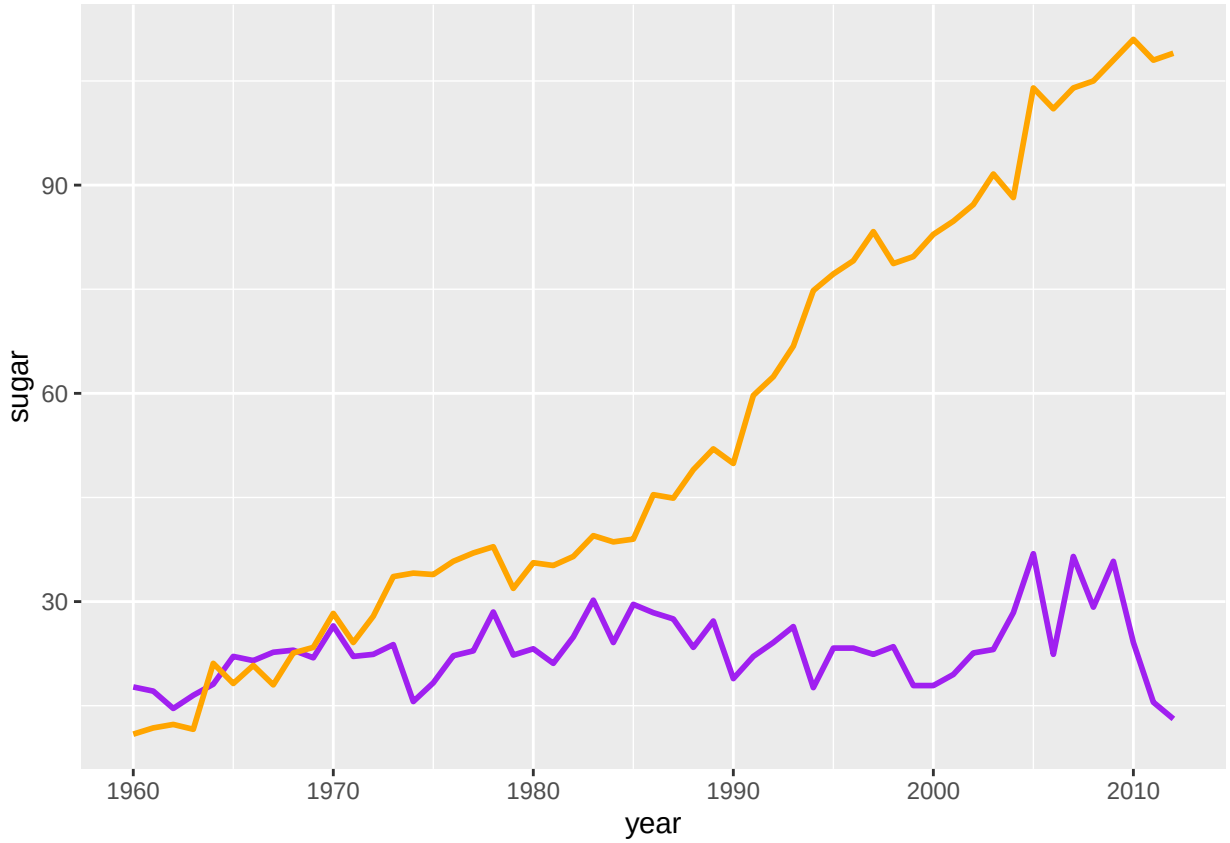
```
##
## > library(readxl)
##
## > library(tidyr)
##
## > library(dplyr)

##
## Attaching package: 'dplyr'

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

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

```
##
## > food_supply <- read_excel("food_supply_kilocalories_per_person_and_day.xlsx")
##
## > sugar <- read_excel("sugar_per_person_g_per_day.xlsx")
##
## > tidy_food_supply <- pivot_longer(food_supply, 2:54,
## +     names_to = "year", values_to = "food_supply")
##
## > tidy_sugar <- pivot_longer(sugar, "1960":"2012", names_to = "year",
## +     values_to = "sugar")
##
## > tidy_data <- left_join(tidy_food_supply, tidy_sugar)

## Joining, by = c("country", "year")

##
## > str(tidy_data)
## tibble [9,487 x 4] (S3: tbl_df/tbl/data.frame)
##  $ country    : chr [1:9487] "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ year       : chr [1:9487] "1960" "1961" "1962" "1963" ...
##  $ food_supply: num [1:9487] 3000 2920 2700 2950 2960 2740 2970 2920 2940 2530 ...
##  $ sugar      : num [1:9487] 14.4 12.8 13.2 15.5 16.3 16.7 17.4 17.8 14.5 18.8 ...
##
## > tidy_data <- drop_na(tidy_data)
##
## > tidy_data <- mutate(tidy_data, percentage = sugar/food_supply)
##
## > tidy_Solomon_Islands <- filter(tidy_data, country ==
## +     "Solomon Islands")
##
## > tidy_Thailand <- filter(tidy_data, country == "Thailand")
```

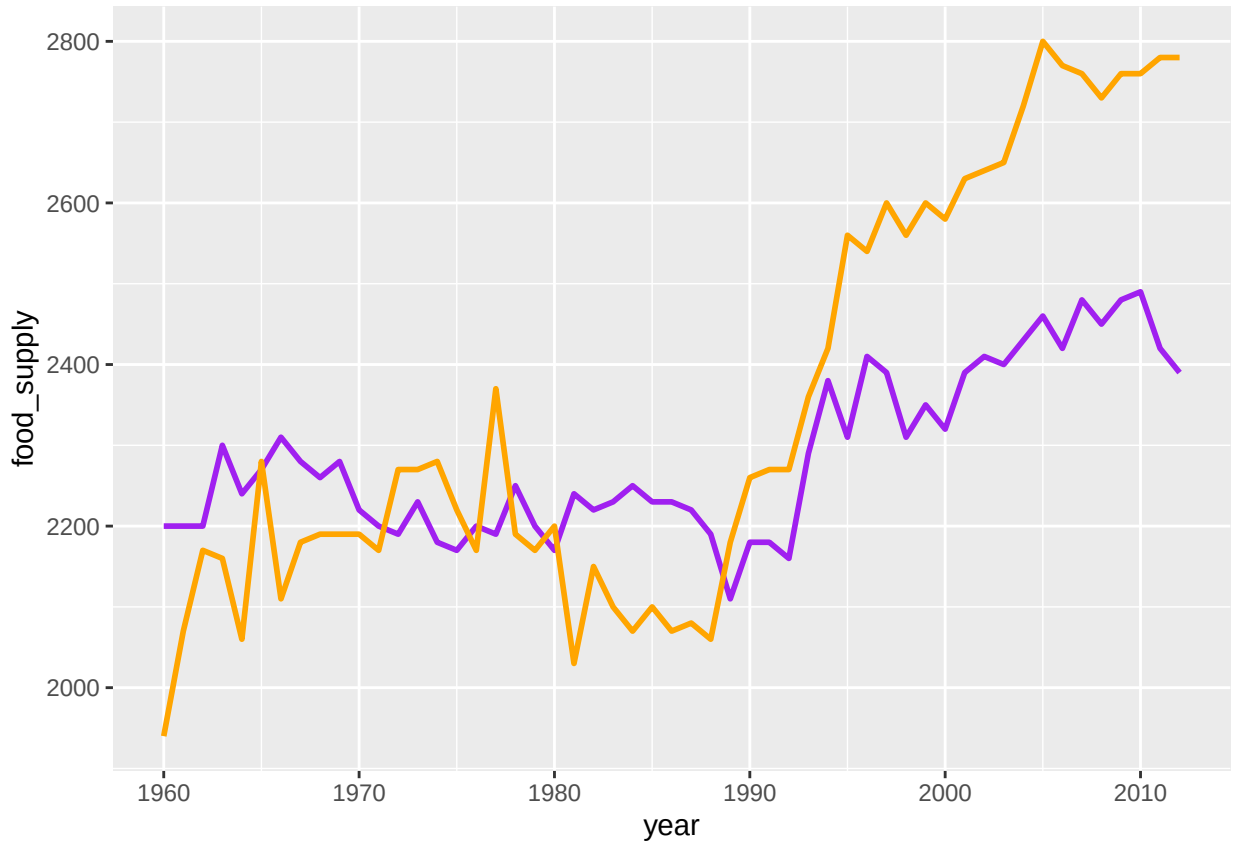
```
source(file = "visualization_function.R", echo = TRUE)
```

```
##
## > library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## > library(ggplot2)
##
## > source(file = "wrangling_code.R", echo = TRUE)
##
```

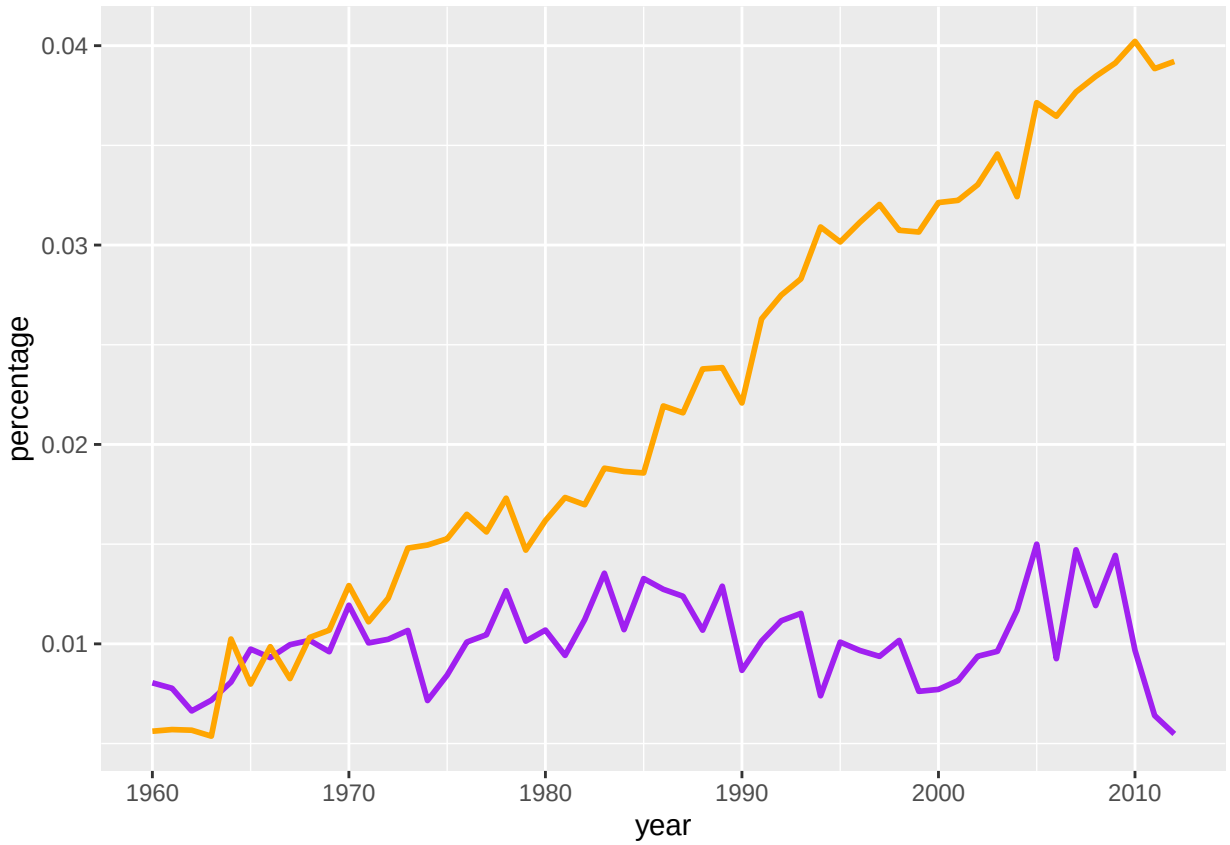
```
## > library(readxl)
##
## > library(tidyr)
##
## > library(dplyr)
##
## > food_supply <- read_excel("food_supply_kilocalories_per_person_and_day.xlsx")
##
## > sugar <- read_excel("sugar_per_person_g_per_day.xlsx")
##
## > tidy_food_supply <- pivot_longer(food_supply, 2:54,
## +     names_to = "year", values_to = "food_supply")
##
## > tidy_sugar <- pivot_longer(sugar, "1960":"2012", names_to = "year",
## +     values_to = "sugar")
##
## > tidy_data <- left_join(tidy_food_supply, tidy_sugar)

## Joining, by = c("country", "year")

##
## > str(tidy_data)
## tibble [9,487 x 4] (S3: tbl_df/tbl/data.frame)
##  $ country    : chr [1:9487] "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ year       : chr [1:9487] "1960" "1961" "1962" "1963" ...
##  $ food_supply: num [1:9487] 3000 2920 2700 2950 2960 2740 2970 2920 2940 2530 ...
##  $ sugar      : num [1:9487] 14.4 12.8 13.2 15.5 16.3 16.7 17.4 17.8 14.5 18.8 ...
##
## > tidy_data <- drop_na(tidy_data)
##
## > tidy_data <- mutate(tidy_data, percentage = sugar/food_supply)
##
## > tidy_Solomon_Islands <- filter(tidy_data, country ==
## +     "Solomon Islands")
##
## > tidy_Thailand <- filter(tidy_data, country == "Thailand")
##
## > compare_plot <- function(country1, country2, indicator) {
## +     if (indicator == "food_supply") {
## +         tidy_country1 <- filter(tidy_data, count .... [TRUNCATED]
##
## > compare_plot("Solomon Islands", "Thailand", "sugar")
```




```
##
## > compare_plot("Solomon Islands", "Thailand", "food_supply")
```




```
##
## > compare_plot("Solomon Islands", "Thailand", "percentage")
```



Wrangling_code

```
#Load the packages: tidyr, dplyr
library(readxl)
library(tidyr)
library(dplyr)
```

```
#Read the excel file and get the dataset of food supply and sugar
food_supply <- read_excel("food_supply_kilocalories_per_person_and_day.xlsx")
sugar <- read_excel("sugar_per_person_g_per_day.xlsx")
```

```
#To tidy a dataset, pivot the offending columns into a new pair of variables
tidy_food_supply <- pivot_longer(food_supply, 2:54, names_to = "year", values_to = "food_supply")
tidy_sugar <- pivot_longer(sugar, '1960':'2012', names_to = "year", values_to = "sugar")

#To combine the tidied versions of tidy_sugar and tidy_food_supply into a single tibble, we need to use
tidy_data <- left_join(tidy_food_supply, tidy_sugar)
```

```
## Joining, by = c("country", "year")
```

```
#Drop tibble's rows containing missing values
tidy_data <- drop_na(tidy_data)

#Mutate adds new variables and preserves existing
#the new variable is percentage which is sugar/food_supply
tidy_data <- mutate(tidy_data,
                    percentage = sugar/food_supply)
```

Visualization Function

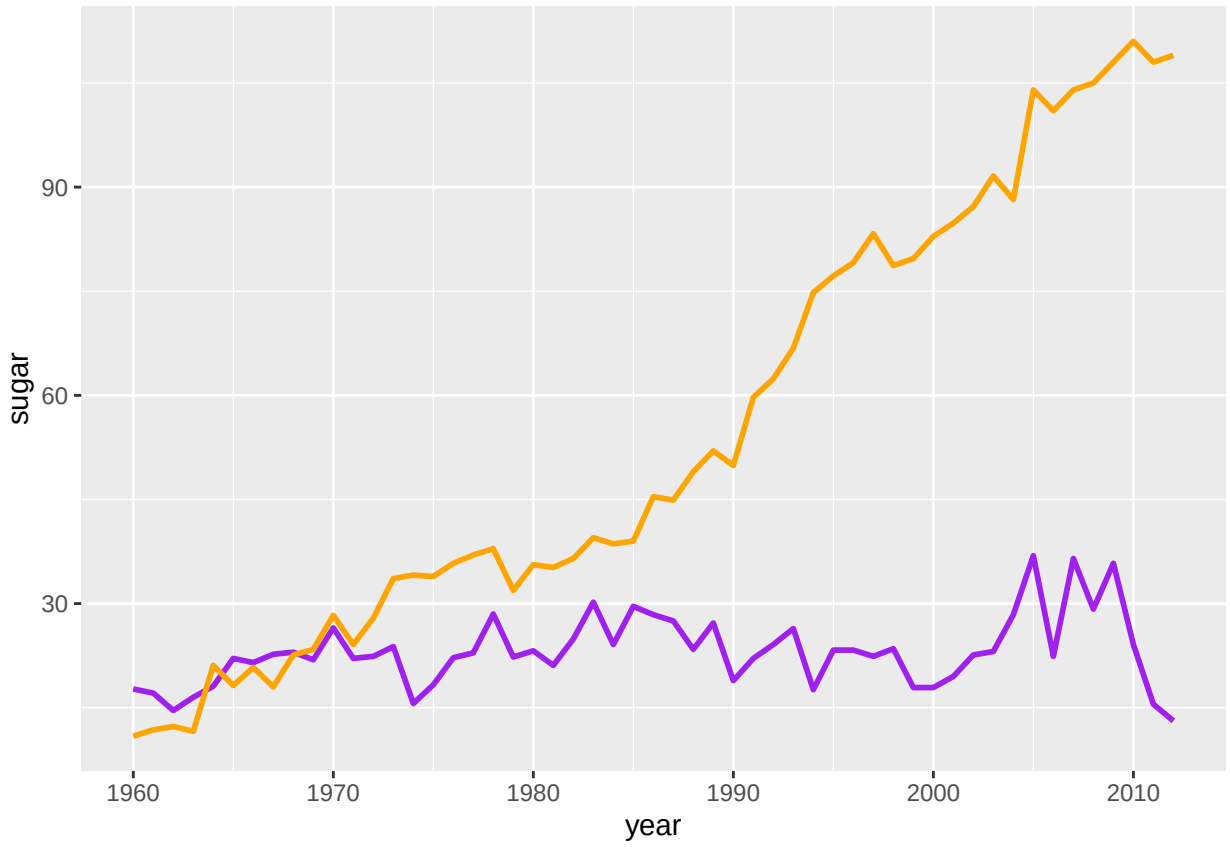
```
#load the package tidyverse and ggplot2
library(tidyverse)
library(ggplot2)

#Read R Code from the file "wrangling_code.R"
source(file = "wrangling_code.R", echo = TRUE)
```

```
##
## > library(readxl)
##
## > library(tidyr)
##
## > library(dplyr)
##
## > food_supply <- read_excel("food_supply_kilocalories_per_person_and_day.xlsx")
##
## > sugar <- read_excel("sugar_per_person_g_per_day.xlsx")
##
## > tidy_food_supply <- pivot_longer(food_supply, 2:54,
## +     names_to = "year", values_to = "food_supply")
##
## > tidy_sugar <- pivot_longer(sugar, "1960":"2012", names_to = "year",
## +     values_to = "sugar")
##
## > tidy_data <- left_join(tidy_food_supply, tidy_sugar)

## Joining, by = c("country", "year")

##
## > str(tidy_data)
## tibble [9,487 x 4] (S3: tbl_df/tbl/data.frame)
##  $ country    : chr [1:9487] "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ year       : chr [1:9487] "1960" "1961" "1962" "1963" ...
##  $ food_supply: num [1:9487] 3000 2920 2700 2950 2960 2740 2970 2920 2940 2530 ...
##  $ sugar      : num [1:9487] 14.4 12.8 13.2 15.5 16.3 16.7 17.4 17.8 14.5 18.8 ...
##
## > tidy_data <- drop_na(tidy_data)
##
## > tidy_data <- mutate(tidy_data, percentage = sugar/food_supply)
##
```

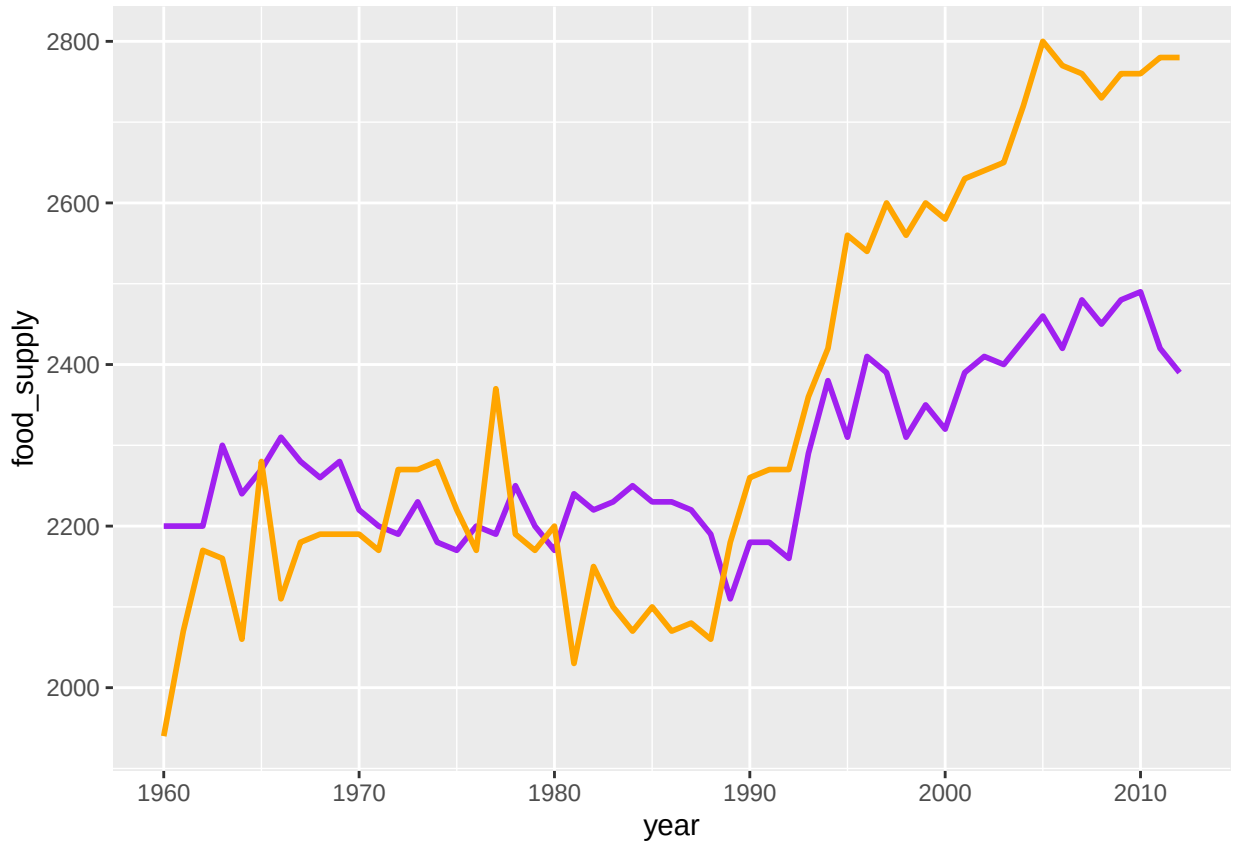
```
## > tidy_Solomon_Islands <- filter(tidy_data, country ==
## +     "Solomon Islands")
##
## > tidy_Thailand <- filter(tidy_data, country == "Thailand")
```

```
#ggplot function to make the visualization, I set there arguments in this function.
#The three arguments are two countries I select to compare and the variable I want to compare.
compare_plot <- function (country1, country2, indicator) {

 #Use if - else to do with different variables
 if(indicator == 'food_supply') {

   #filter() allows you to subset observations based on their values. The first argument is the name of
  #we set two subgroups by the country
  tidy_country1 <- filter(tidy_data, country == country1)
  tidy_country2 <- filter(tidy_data, country == country2)

  #convert the year to numeric so that it could be continuous
  tidy_country1$year <- as.numeric(tidy_country1$year)
  tidy_country2$year <- as.numeric(tidy_country2$year)

  #ggplot the graph between the food_supply and year
  ggplot()+
    geom_line(data = tidy_country1, aes(x = year, y = food_supply), group = 1,
              color = "purple", lwd = 1)+
    geom_line(data = tidy_country2, aes(x = year, y = food_supply), group = 1,
              color = "orange", lwd =1)

} else if(indicator == 'sugar') {
  tidy_country1 <- filter(tidy_data, country == country1)
  tidy_country2 <- filter(tidy_data, country == country2)
  tidy_country1$year <- as.numeric(tidy_country1$year)
  tidy_country2$year <- as.numeric(tidy_country2$year)
  ggplot()+
    geom_line(data = tidy_country1, aes(x = year, y = sugar),  color = "purple", lwd = 1)+
    geom_line(data = tidy_country2, aes(x = year, y = sugar),  color = "orange", lwd =1)

} else if (indicator == 'percentage') {
  tidy_country1 <- filter(tidy_data, country == country1)
  tidy_country2 <- filter(tidy_data, country == country2)
  tidy_country1$year <- as.numeric(tidy_country1$year)
  tidy_country2$year <- as.numeric(tidy_country2$year)
  ggplot()+
    geom_line(data = tidy_country1, aes(x = year, y = percentage), group = 1,
              color = "purple", lwd = 1)+
    geom_line(data = tidy_country2, aes(x = year, y = percentage), group = 1,
              color = "orange", lwd =1)
} else stop('warning: Check the indicator')
}
```

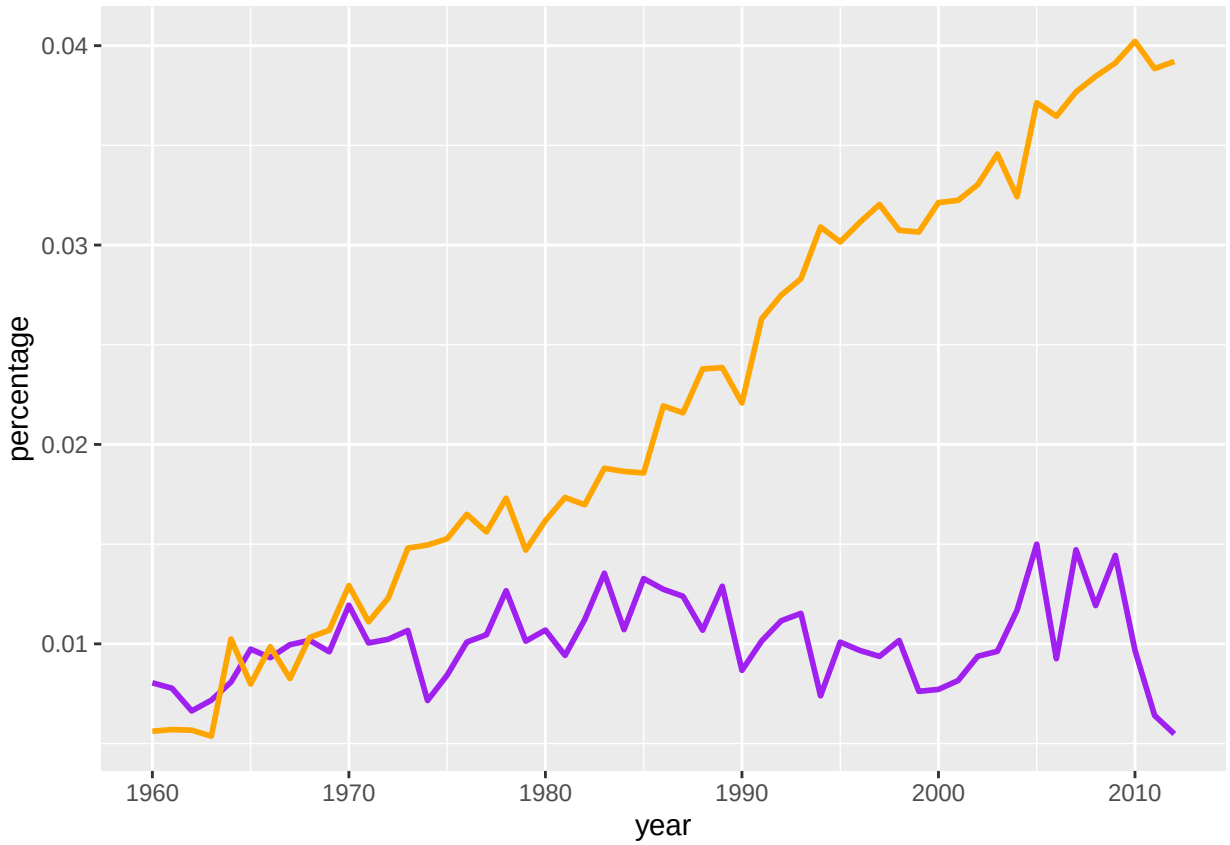
```
#input the variable and use the function to get the plot
compare_plot("Solomon Islands", "Thailand",'sugar')
```




```
compare_plot("Solomon Islands", "Thailand",'food_supply')
```




```
compare_plot("Solomon Islands", "Thailand",'percentage')
```


0.04
0.03
0.02
0.01
percentage
1960
1970
1980
1990
2000
2010
year