ETO5513 Assessment 1 - Data Mini Project

Gun Ho Jang

2024-11-14

Research Question

In recent years, the rising cost of living crisis has placed significant strain on economies worldwide with inflationary pressures on the price of food and energy pushing households and businesses to the brink (Whiting (2022); Shipman et al. (2023)). The increase in food prices has consequentially impacted many people's financial capacity to afford groceries. This project will explore the issue of food insecurity and address the question:

> How much of the global population is unable to afford a healthy diet?

Dataset Introduction

Data for this project was retrieved from Our World in Data (https://ourworldindata.org/food-prices) within the Food Prices Data Explorer. These datasets were originally sourced from the Food and Agriculture Organization of the United Nations ('FAO') and the World Bank. Three different datasets were utilised for this project:

1. Number of people that cannot afford a healthy diet (https://ourworldindata.org/explorers/food-prices?
   Diet=Healthy+diet&Cost+or+Affordability=Affordability&Affordability+metric=Number+that+cannot+afford&Cost+metric=%25+of+average+food+expenditure&

2. Share of population that cannot afford a healthy diet (https://ourworldindata.org/explorers/food-prices?
   Diet=Healthy+diet&Cost+or+Affordability=Affordability&Affordability+metric=Share+that+cannot+afford&Cost+metric=%25+of+average+food+expenditure&co

3. Cost of a healthy diet as a share of average expenditure on food (https://ourworldindata.org/explorers/food-prices?
   Diet=Healthy+diet&Cost+or+Affordability=Cost+&Affordability+metric=Share+that+cannot+afford&Cost+metric=%25+of+average+food+expenditure&country=

We will observe the most recent year with complete data for all variables of interest, which is 2021.

The first two variables of interest *Number of people that cannot afford a healthy diet* and *Share of population that cannot afford a healthy diet* capture the number and percentage share of a country's population that cannot afford a healthy diet. The cost of a healthy diet is classified as "[…] the lowest-cost set of foods available that would meet requirements in dietary guidelines from governments and public health agencies. A diet is defined as unaffordable when its cost, plus basic non-food needs like housing and transportation, exceeds daily income" (Ritchie, Rosado, and Roser (2023)).

The third variable of interest, *Cost of a healthy diet as a share of average expenditure on food*, represents the cost of a healthy diet as a ratio against a person's average food expenditure in a particular country.

```r
# LIsting the variable names
variable_names <- data.frame(
  Variable_Names = c(
    "Country",
    "Continent",
    "Number of people who cannot afford a healthy diet",
    "Share of the population who cannot afford a healthy diet",
    "Affordability of a healthy diet: ratio of cost to food expenditures"
  ))

kable(variable_names, caption = "**Variable Names of Interest**")
```

Variable Names of Interest

Variable_Names
Country
Continent
Number of people who cannot afford a healthy diet
Share of the population who cannot afford a healthy diet
Affordability of a healthy diet: ratio of cost to food expenditures

Dataset Description

The R code chunk below demonstrates the data cleaning process for this project.

```r
# Reading individual CSV files into respective dataframes
number_healthy_diet <- read.csv("Data/number-healthy-diet-unaffordable.csv")

share_healthy_diet <- read.csv("Data/share-healthy-diet-unaffordable.csv")

cost_healthy_diet <- read.csv("Data/cost-healthy-diet-share-food-expenditure.csv")

# Filtering the data for most recent year available and selecting the relevant columns
number_healthy_diet_2021 <- number_healthy_diet %>%
  filter(Year == 2021) %>%
  select(Entity, 4)

share_healthy_diet_2021 <- share_healthy_diet %>%
  filter(Year == 2021) %>%
  select(Entity, 4)

cost_healthy_diet_2021 <- cost_healthy_diet %>%
  filter(Year == 2021) %>%
  select(Entity, 4)

# Merging dataframes based on Entity variable (i.e. Name of Country)
food_affordability_data <- number_healthy_diet_2021 %>%
  left_join(share_healthy_diet_2021, by = "Entity") %>%
  left_join(cost_healthy_diet_2021, by = "Entity")

# Renaming Entity variable to 'Country' and cleaning header names to introduce spacing

food_affordability_data <- food_affordability_data %>%
  rename(Country = Entity,
         `Number of people who cannot afford a healthy diet` = Number.of.people.who.cannot.afford.a.healthy.diet,
         `Share of the population who cannot afford a healthy diet` = Share.of.the.population.who.cannot.afford.a.healthy.di
et,
         `Affordability of a healthy diet: ratio of cost to food expenditures` = Affordability.of.a.healthy.diet..ratio.of.c
ost.to.food.expenditures
  )

# Removing invalid entries under 'Country' variable
invalid_country_entries <-  c("East Asia & Pacific (WB)",
                              "Europe & Central Asia (WB)",
                              "High-income countries",
                              "Latin America & Caribbean (WB)",
                              "Low-income countries",
                              "Lower-middle-income countries",
                              "Middle East & North Africa (WB)",
                              "North America (WB)",
                              "South Asia (WB)",
                              "Sub-Saharan Africa (WB)",
                              "Upper-middle-income countries",
                              "World")

food_affordability_data <- food_affordability_data %>%
  filter(!Country %in% invalid_country_entries)

# Removing invalid N/A entries under all variable columns
food_affordability_data <- na.omit(food_affordability_data)

# Re-grouping countries by their Continents using the 'countrycode' package
food_affordability_data <- food_affordability_data %>%
  mutate(
    Continent = countrycode(Country, origin = "country.name", destination = "continent")
  )

# Re-arranging the variables columns
food_affordability_data <- food_affordability_data %>%
  select(Country, Continent, everything())
```

The dataset on t he affordability of healthy diets contains 144 observations and 5 variables. Please note that the code below was used to calculate these values:

```r
```{r number of observations and variables, echo = FALSE}
# Calculating the number of observations and variables

num_observations <- nrow(food_affordability_data)
num_variables <- ncol(food_affordability_data)
```
```

Figure 1. Screenshot of Code

The first two rows of the dataset can be displayed with `str()` in order to view the type of variables within the dataset.

```
# Displaying the first two rows of the dataset
str(food_affordability_data[1:2,])
```

```
## 'data.frame':    2 obs. of  5 variables:
##  $ Country                                               : chr  "Albania" "Algeria"
##  $ Continent                                             : chr  "Europe" "Africa"
##  $ Number of people who cannot afford a healthy diet     : num  400000 8300000
##  $ Share of the population who cannot afford a healthy diet : num  12.6 18.7
##  $ Affordability of a healthy diet: ratio of cost to food expenditures: num  32 65
##  - attr(*, "na.action")= 'omit' Named int [1:4] 58 64 96 114
##   ..- attr(*, "names")= chr [1:4] "58" "64" "96" "114"
```

Data Summary

```
# Summary statistics for 'Number of people who cannot afford a healthy diet' and 'Affordability of a healthy diet: ratio of
cost to food expenditures'
data_summary <- food_affordability_data %>%
  group_by(Continent) %>%
  summarise(
    Mean_Number = format(mean(`Number of people who cannot afford a healthy diet`, na.rm = TRUE), scientific = FALSE),
    Sum_Number = format(sum(`Number of people who cannot afford a healthy diet`, na.rm = TRUE), scientific = FALSE),
    Mean_Share = format(mean(`Share of the population who cannot afford a healthy diet`, na.rm = TRUE), scientific =  FALS
E),
    Median_Share = format(median(`Share of the population who cannot afford a healthy diet`, na.rm = TRUE), scientific = FAL
SE)
  ) %>%
  head(10)

kable(data_summary, caption = "**Summary statistics of the 'Number of people who cannot afford a healthy diet' and 'Share of
the population who cannot afford a healthy diet' variables**")
```

Summary statistics of the 'Number of people who cannot afford a healthy diet' and 'Share of the population who cannot afford a healthy diet' variables

Continent	Mean_Number	Sum_Number	Mean_Share	Median_Share
Africa	17862500	857400004	59.45	60.5
Americas	7258333	174199996	28.00833	26.45
Asia	48845455	1611900030	25.17273	18
Europe	1110811	41100000	8.202703	5
Oceania	700000	1400000	35	35

From the summary statistics, the continents of Asia holds the highest number of people unable to afford a healthy diet (at 1.612 billion) which is then followed by Africa (at 857 million), although Africa's population had a higher mean and median percentage share of the population being impacted compared to Asia's percentage share. **Please note that certain continents may not be represented accurately due to unavailable data and smaller pools of countries, such as Oceania (i.e. Australia and Fiji) where Fiji has skewed the represented data for the entire region.**

Given that Asia has the largest population unable to afford a healthy diet, we can examine this by further breaking down the numbers by country:

```
# Break down of 'Number of people who cannot afford a healthy diet' by countries in Asia (sorted by largest population numbe
r)
asia_data <- food_affordability_data %>%
  filter(Continent == "Asia") %>%
  select(Country, `Number of people who cannot afford a healthy diet`) %>%
  arrange(desc(`Number of people who cannot afford a healthy diet`))

kable (asia_data, caption = "**Number of People Unable to Afford a Healthy Diet in Asia**")
```

Number of People Unable to Afford a Healthy Diet in Asia

Country	Number of people who cannot afford a healthy diet
India	830900030
China	251300000
Pakistan	135800000
Indonesia	127600000
Bangladesh	87600000
Philippines	58500000

Country	Number of people who cannot afford a healthy diet
Syria	18400000
Nepal	13100000
Thailand	13100000
Iraq	12400000
Japan	9500000
Vietnam	9400000
Sri Lanka	7400000
Turkey	7400000
Uzbekistan	5300000
South Korea	5200000
Laos	4300000
Tajikistan	3200000
Kyrgyzstan	2400000
Israel	1600000
Armenia	1500000
Jordan	1500000
Taiwan	1100000
Kazakhstan	1000000
Malaysia	900000
Mongolia	800000
Palestine	300000
Bhutan	200000
Azerbaijan	100000
Lebanon	100000
Cyprus	0
Maldives	0
United Arab Emirates	0

The data reveals that India, China and Pakista are the three most affected countries, with a combined total of 1.218 billion people who are unable to afford a healthy diet. This accounts for nearly 76% of the reported affected population in Asia. In contrast, Cyprus, Maldives, and the United Arab Emirates report none being impacted. This could indicate that either negligible numbers are affected, or potential under-reporting or data unavailability for these countries.

Visualisations

```
# Creating the scatter plot for Affordability of a Healthy Diet across Continents
ggplot(food_affordability_data, aes(x = Continent, y = `Affordability of a healthy diet: ratio of cost to food expenditures
`)) +
  geom_point() +
  labs(
    title = "Affordability of a Healthy Diet across Continents",
    x = "Continent",
    y =  "Ratio of Cost to Food Expenditures"
  ) +
  theme(plot.title = element_text(face = "bold"))
```

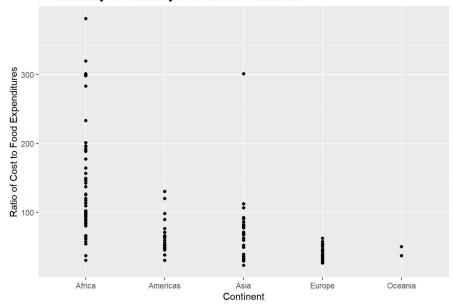**Affordability of a Healthy Diet across Continents**

Figure 2. Scatter Plot Visualisation

The scatter plot provides a comparison of the affordability of a healthy diet across continents. However, a box plot may be more suitable in providing further insights by highlighting the median, quartiles, and any outliers.

```
# Creating the box plot for Affordability of a Healthy Diet across Continents
ggplot(food_affordability_data, aes(x = Continent, y = `Affordability of a healthy diet: ratio of cost to food expenditures
`)) +
  geom_boxplot() +
  geom_hline(yintercept = 100, color = "red", linetype = "dashed") +
  labs(
    title = "Affordability of a Healthy Diet across Continents",
    x = "Continent",
    y =  "Ratio of Cost to Food Expenditures"
  ) +
  theme(plot.title = element_text(face = "bold"))
```

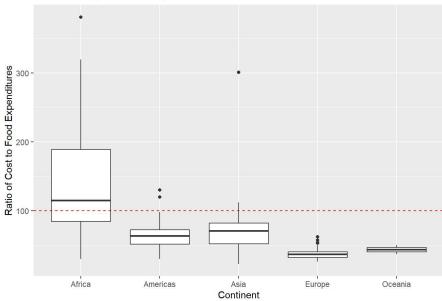**Affordability of a Healthy Diet across Continents**

Figure 3. Box Plot Visualisation

The following observations have been made:

- *Countries within the continents of Africa, the Americas, and Asia display data points that surpass 100% of the Ratio of Cost to Food Expenditures, indicating that people in these regions would have to spend more than their average food expenditure budget would allow in order to afford a healthy diet. A red dashed line has been added to the box plot visualisation to mark the 100% threshold.*

- In Africa, the highest ratio is 318% for the nation of Burundi, and for Asia, this is Syria at 301%. Both nations are characterised by a history of volatility and conflict, and struggle with issues of poverty (BBC (2023a); BBC (2023b)). These countries thus appear as the sole outliers in the box plot for their respective continents.

Conclusion

The project's research question aimed to determine how much of the global population cannot afford a healthy diet. The findings indicate that approximately 2.686 billion people worldwide lack the means to afford a healthy diet, primarily in Africa and Asia. In Asia, nearly 76% of those affected are located in India, China, and Pakistan, where food budgets are unable to match the total cost required to secure adequate nutrition. With one-third of the world impacted, the FAO's Director of the Agrifood Economics and Policy Division has highlighted that this represents a structural issue in our agrifood systems, and it will require significant investment in order to address this problem and to achieve the United Nation's goal of Zero Hunger by 2030 (Emsden (2024)).

References

BBC. 2023a. "Burundi Country Profile." https://www.bbc.com/news/world-africa-13085064 (https://www.bbc.com/news/world-africa-13085064).

———. 2023b. "Syria Country Profile." https://www.bbc.com/news/world-middle-east-14703856 (https://www.bbc.com/news/world-middle-east-14703856).

Emsden, C. 2024. "Healthy Diets Remain Unaffordable for a Third of the World's Population." Food; Agriculture Organization of the United Nations. https://www.fao.org/newsroom/detail/healthy-diets-remain-unaffordable-for-a-third-of-the-world-s-population/en (https://www.fao.org/newsroom/detail/healthy-diets-remain-unaffordable-for-a-third-of-the-world-s-population/en).

Ritchie, H., P. Rosado, and M. Roser. 2023. "Food Prices." Our World in Data. https://ourworldindata.org/food-prices (https://ourworldindata.org/food-prices).

Shipman, A., A. Smondel, B. Y. Adhinegara, J. W. Diamond, L. G. Vega, M. H. Shahrour, P. Martin, and W. Simpson. 2023. "Global Economy 2023: How Countries Around the World Are Tackling the Cost of Living Crisis." The Conversation. https://theconversation.com/global-economy-2023-how-countries-around-the-world-are-tackling-the-cost-of-living-crisis-196740 (https://theconversation.com/global-economy-2023-how-countries-around-the-world-are-tackling-the-cost-of-living-crisis-196740).

Whiting, K. 2022. "The Cost-of-Living Crisis Is Having a Global Impact. Here's What Countries Are Doing to Help." World Economic Forum. https://www.weforum.org/stories/2022/09/cost-of-living-crisis-global-impact/ (https://www.weforum.org/stories/2022/09/cost-of-living-crisis-global-impact/).