

Group4

Ophelia & Joanna

Table of contents

1 Data Preprocessing	1
1.1 Data Import and cleaning	1
2 Question 1	3
2.1 Has global coffee production increased over time? In which year did it reach its peak?	3
3 Question 2	4
3.1 Which countries have shown the fastest growth in coffee production over the past 20 years?	4
4 Question 3	6
4.1 Is production and consumption balanced over time? Are there periods of over-supply?	6
5 Github repo URL	7

1 Data Preprocessing

1.1 Data Import and cleaning

```
coffee <- read_csv("C:/Users/USER/Desktop/PBAwork/Final/psd_coffee.csv")

coffee_clean <- coffee |>
  filter(!is.na(Country), !is.na(Year))

coffee_clean <- coffee_clean |>
  mutate(across(where(is.numeric), ~ replace_na(.x, 0)))

coffee_clean <- coffee_clean |>
  mutate(across(where(is.numeric), ~ if_else(.x < 0, 0, .x)))

coffee_clean <- coffee_clean |>
  mutate(
    Net_Exports = Exports - Imports,
    Self_Sufficiency = Production / `Domestic Consumption`
```

)

summary(coffee_clean)

Country	Year	Arabica Production	Bean Exports
Length:6016	Min. :1960	Min. : 0.0	Min. : 0.0
Class :character	1st Qu.:1976	1st Qu.: 0.0	1st Qu.: 0.0
Mode :character	Median :1992	Median : 0.0	Median : 4.0
	Mean :1992	Mean : 744.3	Mean : 814.6
	3rd Qu.:2007	3rd Qu.: 200.0	3rd Qu.: 325.0
	Max. :2023	Max. :49700.0	Max. :41689.0

Bean Imports	Beginning Stocks	Domestic Consumption	Ending Stocks
Min. : 0.0	Min. : 0.0	Min. : 0	Min. : 0.0
1st Qu.: 0.0	1st Qu.: 0.0	1st Qu.: 0	1st Qu.: 0.0
Median : 0.0	Median : 0.0	Median : 14	Median : 0.0
Mean : 372.7	Mean : 457.2	Mean : 673	Mean : 449.4
3rd Qu.: 0.0	3rd Qu.: 83.0	3rd Qu.: 227	3rd Qu.: 81.0
Max. :47000.0	Max. :72461.0	Max. :49070	Max. :72461.0

Exports	Imports	Other Production	Production
Min. : 0.0	Min. : 0.0	Min. : 0.000	Min. : 0
1st Qu.: 0.0	1st Qu.: 0.0	1st Qu.: 0.000	1st Qu.: 0
Median : 9.0	Median : 0.0	Median : 0.000	Median : 21
Mean : 895.6	Mean : 430.3	Mean : 2.211	Mean : 1131
3rd Qu.: 439.0	3rd Qu.: 2.0	3rd Qu.: 0.000	3rd Qu.: 575
Max. :45675.0	Max. :47000.0	Max. :375.000	Max. :69900

Roast & Ground Exports	Roast & Ground Imports	Robusta Production
Min. : 0.00	Min. : 0.00	Min. : 0.0
1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 0.0
Median : 0.00	Median : 0.00	Median : 0.0
Mean : 13.16	Mean : 10.65	Mean : 383.9
3rd Qu.: 0.00	3rd Qu.: 0.00	3rd Qu.: 27.0
Max. :2975.00	Max. :1060.00	Max. :30480.0

Rst,Ground Dom. Consum	Soluble Dom. Cons.	Soluble Exports	Soluble Imports
Min. : 0.0	Min. : 0.00	Min. : 0.0	Min. : 0.00
1st Qu.: 0.0	1st Qu.: 0.00	1st Qu.: 0.0	1st Qu.: 0.00
Median : 11.0	Median : 0.00	Median : 0.0	Median : 0.00
Mean : 588.5	Mean : 84.51	Mean : 67.9	Mean : 46.05
3rd Qu.: 188.2	3rd Qu.: 1.00	3rd Qu.: 0.0	3rd Qu.: 0.00
Max. :47010.0	Max. :6745.00	Max. :4300.0	Max. :6000.00

Total Distribution	Total Supply	Net_Exports	Self_Sufficiency
Min. : 0	Min. : 0	Min. : -43970.0	Min. : 0.00
1st Qu.: 0	1st Qu.: 0	1st Qu.: 0.0	1st Qu.: 1.00
Median : 112	Median : 112	Median : 1.0	Median : 3.20
Mean : 2018	Mean : 2018	Mean : 465.3	Mean : Inf
3rd Qu.: 1105	3rd Qu.: 1105	3rd Qu.: 341.5	3rd Qu.:12.29

```
Max.      :97806      Max.      :97806      Max.      : 45603.0      Max.      : Inf
NA's      :1948
```

2 Question 1

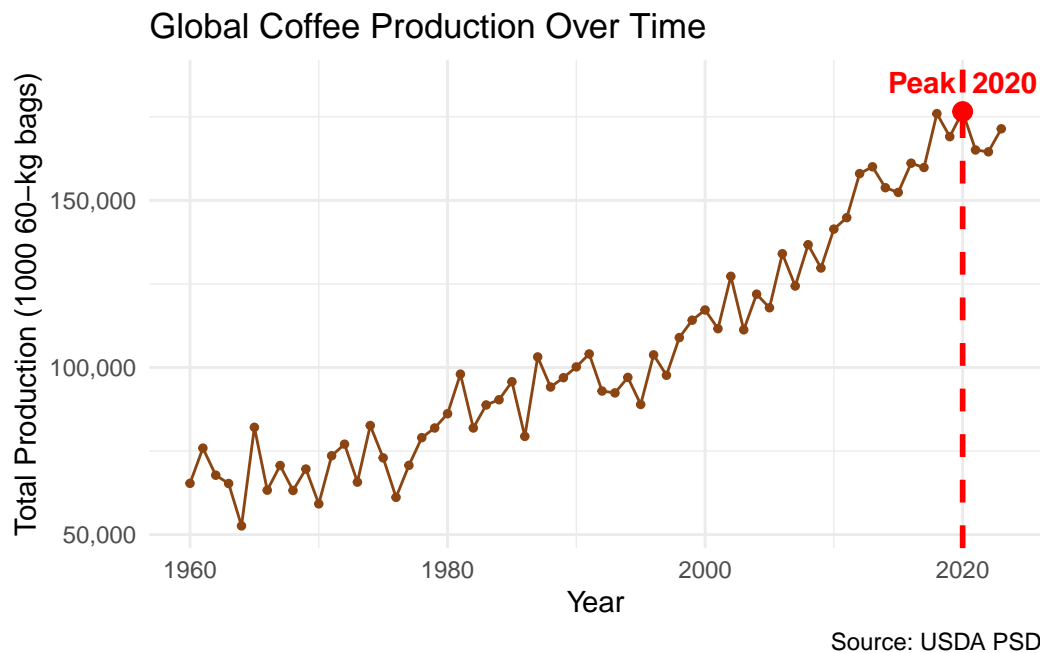
2.1 Has global coffee production increased over time? In which year did it reach its peak?

```
global_production <- coffee_clean |>
  group_by(Year) |>
  summarise(Total_Production = sum(Production, na.rm = TRUE))

peak_year <- global_production |>
  filter(Total_Production == max(Total_Production))

plot1 <- ggplot(data = global_production,
               mapping = aes(x = Year, y = Total_Production)) +
  geom_line(color = "#8B4513", linewidth = 0.5) +
  geom_point(color = "#8B4513", size = 1) +
  geom_vline(xintercept = peak_year$Year,
            linetype = "dashed",
            color = "red",
            linewidth = 1) +
  geom_point(data = peak_year,
            aes(x = Year, y = Total_Production),
            color = "red",
            size = 3) +
  annotate("text",
         x = peak_year$Year,
         y = peak_year$Total_Production * 1.05,
         label = paste0("Peak ", peak_year$Year),
         color = "red",
         fontface = "bold") +
  scale_y_continuous(labels = comma) +
  labs(
    title = "Global Coffee Production Over Time",
    x = "Year",
    y = "Total Production (1000 60-kg bags)",
    caption = "Source: USDA PSD"
  ) +
  theme_minimal()

print(plot1)
```



Global coffee production has shown a steady upward trend from 1960 to 2023, reflecting continuous expansion in global supply and cultivation capacity. The production reached its historical peak in 2020, at approximately 170 million 60-kg bags.

3 Question 2

3.1 Which countries have shown the fastest growth in coffee production over the past 20 years?

```
max_year <- max(coffee_clean$Year, na.rm = TRUE)
start_year <- max_year - 20

country_growth <- coffee_clean |>
  filter(Year >= start_year, Year <= max_year) |>
  group_by(Country) |>
  summarise(
    First_Production = Production[Year == min(Year)][1],
    Last_Production = Production[Year == max(Year)][1],
    .groups = "drop"
  ) |>
  filter(First_Production > 0) |>
  mutate(
    Growth_Rate = ((Last_Production - First_Production) / First_Production) * 100
  ) |>
  filter(Growth_Rate > 0) |>
  arrange(desc(Growth_Rate)) |>
  head(15)

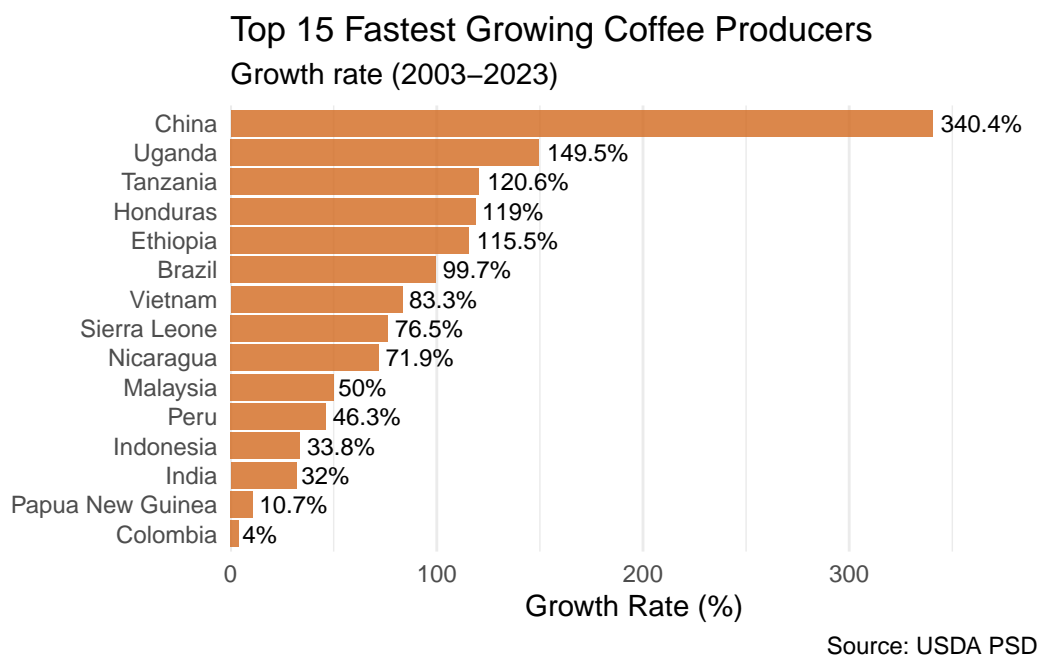
plot2 <- ggplot(data = country_growth,
```

```

        mapping = aes(x = reorder(Country, Growth_Rate),
                        y = Growth_Rate)) +
geom_col(fill = "#D2691E", alpha = 0.8) +
geom_text(aes(label = paste0(round(Growth_Rate, 1), "%")),
          hjust = -0.1,
          size = 3) +
coord_flip() +
scale_y_continuous(expand = expansion(mult = c(0, 0.15))) +
labs(
  title = "Top 15 Fastest Growing Coffee Producers",
  subtitle = paste0("Growth rate (", start_year, "-", max_year, ")"),
  x = NULL,
  y = "Growth Rate (%)",
  caption = "Source: USDA PSD"
) +
theme_minimal() +
theme(
  panel.grid.major.y = element_blank()
)

print(plot2)

```



Over the past two decades, China has emerged as the fastest-growing coffee producer, with an exceptional 340% increase in output. Other countries showing strong expansion include Uganda, Tanzania, and Honduras, each exceeding 100% growth. This reflects the rapid development of coffee cultivation in Asia and Africa, reshaping the global coffee supply landscape.

4 Question 3

4.1 Is production and consumption balanced over time? Are there periods of oversupply?

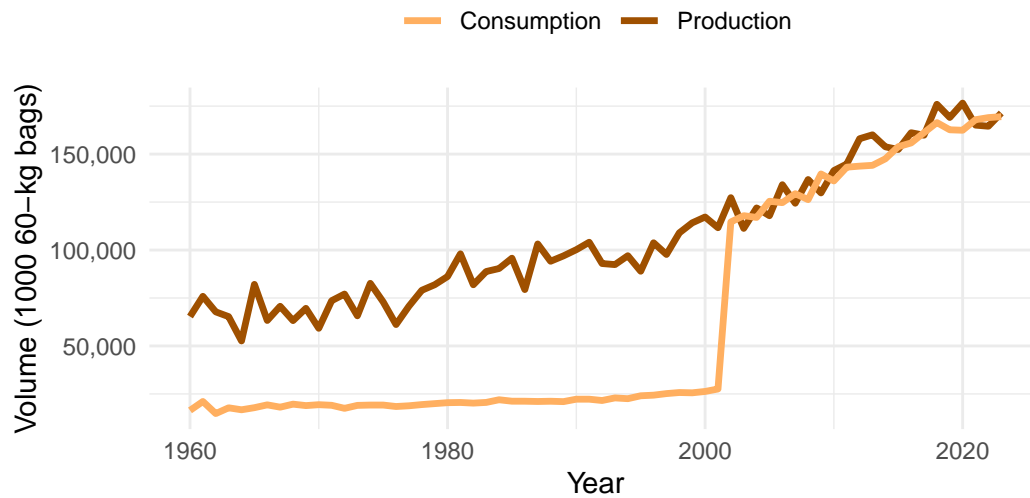
```
global_balance <- coffee_clean |>
  group_by(Year) |>
  summarise(
    Production = sum(Production, na.rm = TRUE),
    Consumption = sum(`Domestic Consumption`, na.rm = TRUE),
    .groups = "drop"
  ) |>
  mutate(
    Gap = Production - Consumption
  )

plot3 <- ggplot(data = global_balance,
               mapping = aes(x = Year)) +
  geom_line(aes(y = Production, color = "Production"), linewidth = 1.2) +
  geom_line(aes(y = Consumption, color = "Consumption"), linewidth = 1.2) +
  scale_color_manual(values = c("Production" = "#9F5000",
                                "Consumption" = "#FFAF60")) +
  scale_y_continuous(labels = comma) +
  labs(
    title = "Global Coffee Production vs Consumption",
    subtitle = "Red areas indicate periods of oversupply",
    x = "Year",
    y = "Volume (1000 60-kg bags)",
    color = NULL,
    caption = "Source: USDA PSD"
  ) +
  theme_minimal() +
  theme(
    legend.position = "top"
  )

print(plot3)
```

Global Coffee Production vs Consumption

Red areas indicate periods of oversupply



Source: USDA PSD

Global coffee production and consumption have both increased steadily over time, with production generally outpacing consumption. Notable periods of oversupply appear after the early 2000s, when production began to grow more rapidly, creating occasional supply surpluses in the global market.

5 Github repo URL

https://github.com/ophelia0207/PBA_DataAnalysisProject