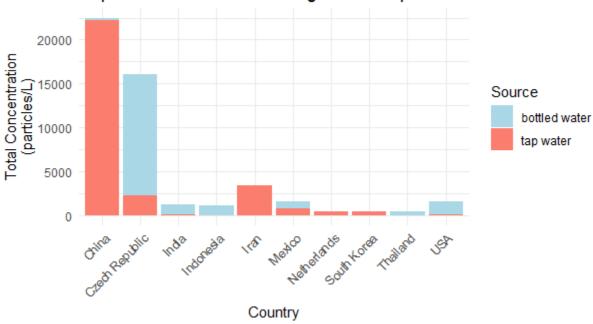
Student: Matthew Terndrup Date: 5/10/25

Data Source: https://catalog.data.gov/dataset/microplastics-in-drinking-water

Top 10 Countries with the Highest Microplastic Concentrations



1. Load data safely

data <- read.csv("samples_geocoded.csv", stringsAsFactors = FALSE)

#2. Filter for:

- Concentration_Units == "particles/L"

- Numeric Concentration values

- Non-NA Countries

library(dplyr)

data_clean <- data %>%

filter(Concentration_Units == "particles/L") %>%

filter(grepl("^[0-9.]+\$", Concentration)) %>%

mutate(Concentration = as.numeric(Concentration)) %>%

filter(!is.na(Countries)) # Remove rows where Countries is NA

3. Summarize total concentration by Country and Source

data_by_country_source <- data_clean %>%

group by(Countries, Source) %>%

summarise(Total_Concentration = sum(Concentration, na.rm = TRUE))

4. Get the total concentration per country, and select the top 10 countries top_10_countries <- data_by_country_source %>%

group by(Countries) %>%

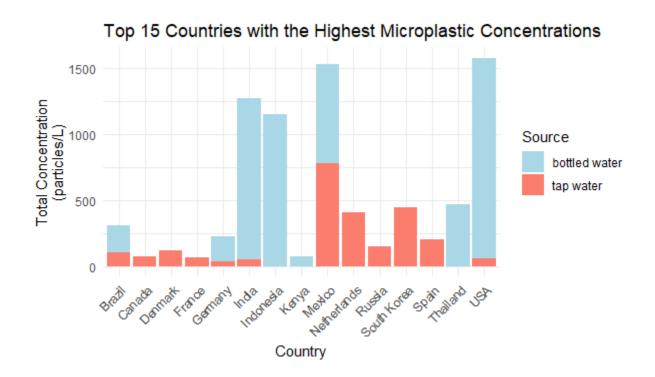
```
summarise(Total_Concentration = sum(Total_Concentration, na.rm = TRUE)) %>% top_n(10, Total_Concentration) %>% arrange(desc(Total_Concentration))
```

5. Filter the original data to include only the top 10 countries data_top_10 <- data_by_country_source %>% filter(Countries %in% top_10_countries\$Countries)

6. Create the stacked bar chart for the top 10 countries with custom y-axis label library(ggplot2)

```
ggplot(data_top_10, aes(x = Countries, y = Total_Concentration, fill = Source)) +
geom_bar(stat = "identity") +
theme_minimal() +
labs(title = "Top 10 Countries with the Highest Microplastic Concentrations",
        x = "Country",
        y = "Total Concentration") +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
scale_fill_manual(values = c("bottled water" = "lightblue", "tap water" = "salmon")) +
ylab("Total Concentration\n(particles/L)") # Add "particles/L" below the y-axis label
```

Adjusted Chart showing the Top 15 Countries - China, Czech Republic and Iran excluded *Note: I still need to figure out how to change labels and such to state that these two countries were removed from the chart *



```
data_clean <- data %>%
 filter(Concentration Units == "particles/L") %>%
 filter(grepl("^[0-9.]+$", Concentration)) %>%
 mutate(Concentration = as.numeric(Concentration)) %>%
 filter(!is.na(Countries)) %>%
 filter(!Countries %in% c("China", "Czech Republic", "Iran"))
data_by_country_source <- data_clean %>%
 group by(Countries, Source) %>%
 summarise(Total Concentration = sum(Concentration, na.rm = TRUE))
top 15 countries <- data by country source %>%
 group_by(Countries) %>%
 summarise(Total Concentration = sum(Total Concentration, na.rm = TRUE)) %>%
 top_n(15, Total_Concentration) %>%
 arrange(desc(Total_Concentration))
data_top_15 <- data_by_country_source %>%
 filter(Countries %in% top 15 countries$Countries)
ggplot(data_top_15, aes(x = Countries, y = Total_Concentration, fill = Source)) +
 geom bar(stat = "identity") +
 theme minimal() +
 labs(title = "Top 15 Countries with the Highest Microplastic Concentrations",
    x = "Country",
    y = "Total Concentration") +
 theme(axis.text.x = element text(angle = 45, hjust = 1)) +
 scale_fill_manual(values = c("bottled water" = "lightblue", "tap water" = "salmon")) +
 ylab("Total Concentration\n(particles/L)")
```