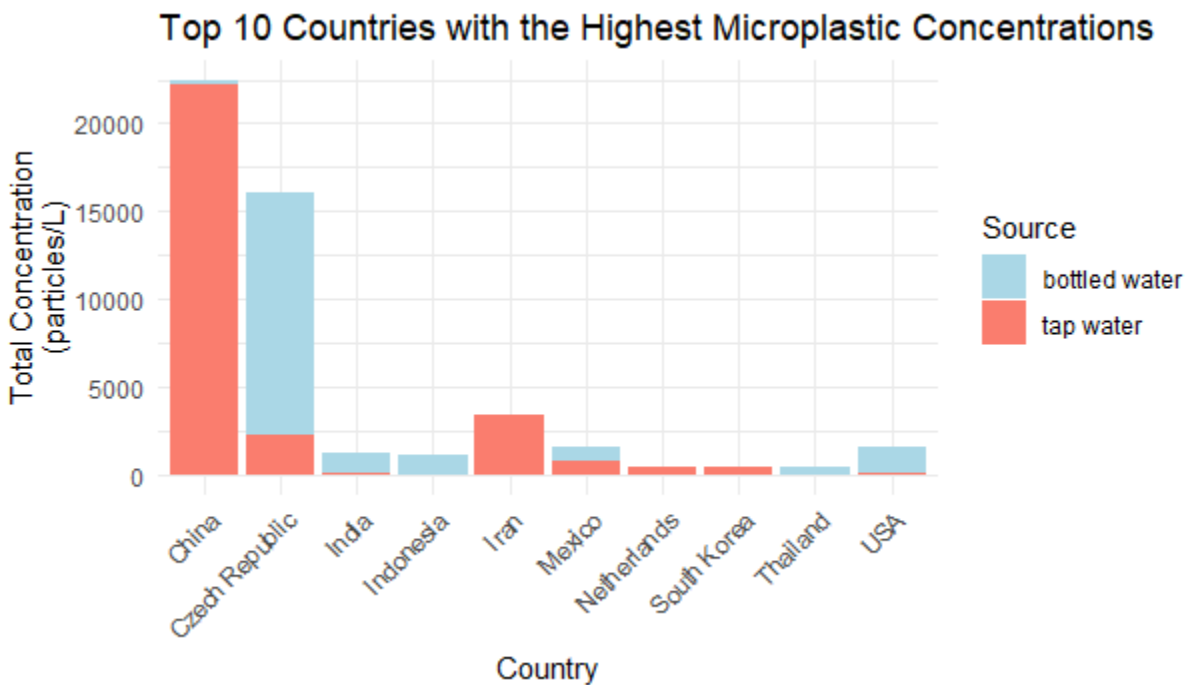


Student: Matthew Terndrup

Date: 5/10/25

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



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 %>%
```

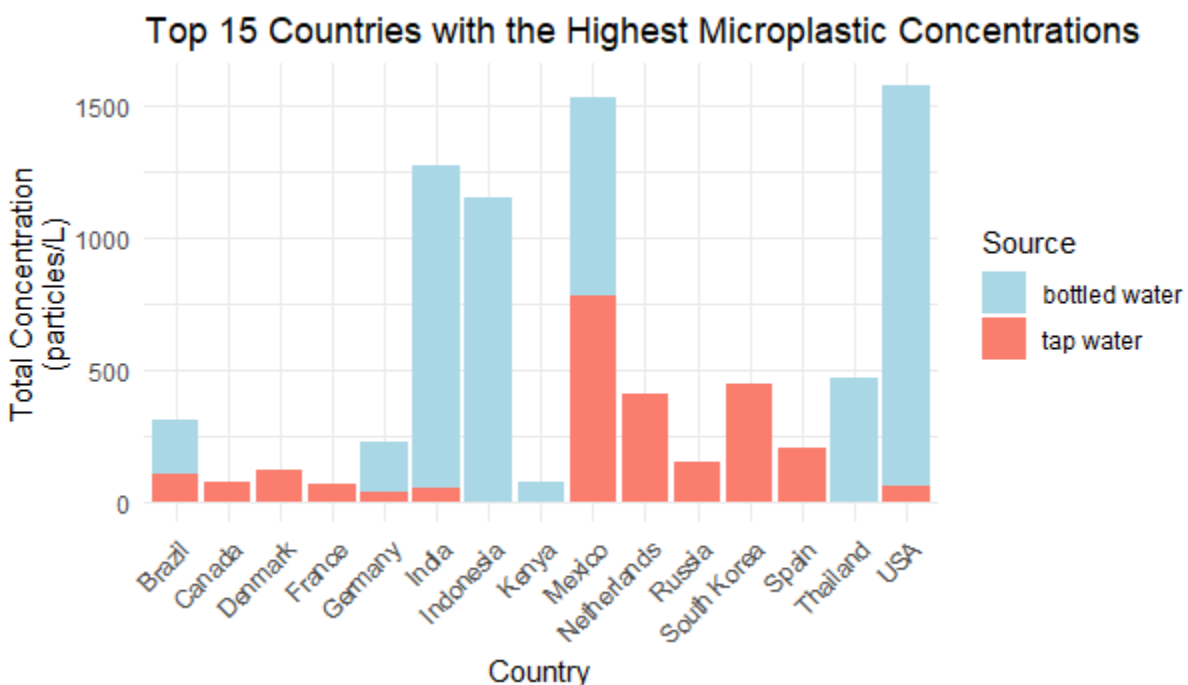
```
  group_by(Countries, Source) %>%
```

```
  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) %>%
```



```

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)")

```