SCSData

Yestin Arvin Gochuico

2025-02-06

China’s Growing Presence in the Indo-Pacific Region - Data Exploration and Analysis

Data

Incidents related to Chinese and ASEAN vessels will be isolated since I’m only interested in the conflict with the involved ASEAN countries in disputed territorial waters.

First, we import necessary libraries.

```
library(tidyverse)
library(treemapify)
library(networkD3)
library(htmlwidgets)
```

```
df <- read.csv("SCSIncidents.csv", header=FALSE)
head(df)
```

To prepare the data for data exploration, column names are added.

```
colnames(df) = c('Country', 'Year', 'Month', 'Day', 'Latitude', 'Longitude', 'Level', 'N
otes')
head(df)
```

Here, all incidents related to China are isolated by filtering the `Country` column with `CN`.

```
df_china <- df %>%
  filter(grepl("CN", Country))
head(df_china)
```

In this part, I’m only interested in the China-related incidents that is associated with another country. We can verify this by getting all the observations that has multiple country codes in the `Country` column.

```
df_multi <- df %>%
  filter(grepl("/", Country)) %>%
  filter(grepl("CN", Country))
head(df_multi)
```

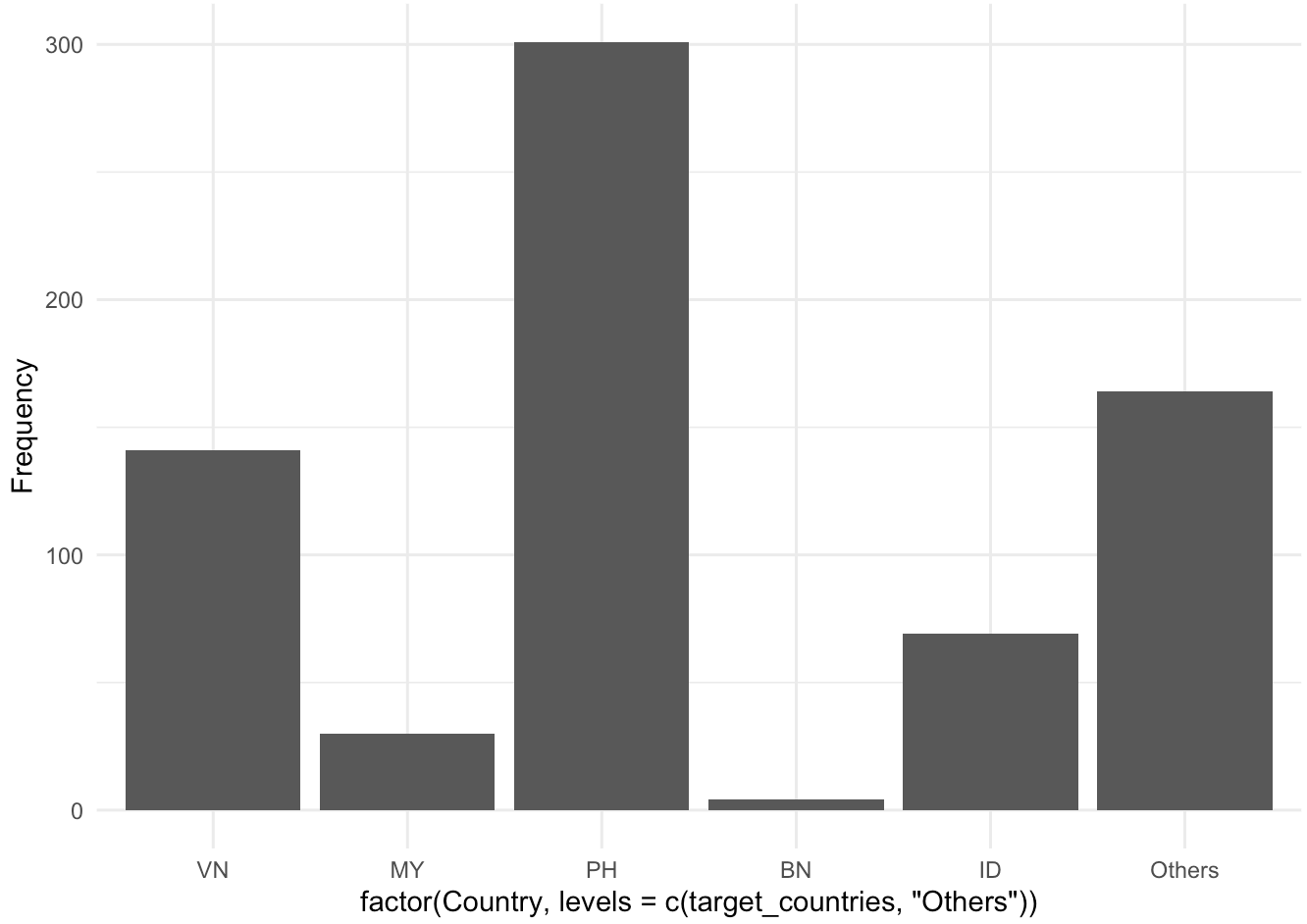
All the countries associated with the incidents will be tallied. For this, we need to remove the `CN` part of observations.

```
Event_TallyPerCountry <- df_multi %>%
  separate_rows(Country, sep = "/") %>%
  filter(Country != "CN")
head(Event_TallyPerCountry)
```

Now that we have the dataframe needed, we can now make a bar chart to compare the numbers of China-incidents per ASEAN country involved.

```
target_countries <- c("CN", "VN", "MY", "PH", "BN", "ID")

Event_TallyPerCountry <- Event_TallyPerCountry %>%
  mutate(Country = ifelse(Country %in% target_countries, Country, "Others")) %>%
  group_by(Country) %>%
  count(Country, name = "Frequency")
```

```
ggplot(Event_TallyPerCountry, aes(x = factor(Country, levels = c(target_countries, "Othe
rs")), y = Frequency)) +
  geom_bar(stat = "identity") +
    theme_minimal()
```




```
ggsave("NumIncidentsPerCountry.svg", device = "svg")
```

```
## Saving 7 x 5 in image
```

Next, I'm interested in seeing the trend of total China-related incidents per ASEAN country.

```
Event_TrendPerCountry <- df_multi %>%
  separate_rows(Country, sep = "/") %>%
  filter(Country != "CN")
head(Event_TrendPerCountry)
```

The `Year`, `Month`, and `Day` of the data is separated by columns. It has to be combined to be used for a time series line chart.

```
Event_TrendPerCountry <- Event_TrendPerCountry %>%
  mutate(Country = ifelse(Country %in% target_countries, Country, "Others")) %>%
  group_by(Country) %>%
  mutate(Date = as.Date(paste(Year, Month, Day, sep = "-"), format = "%Y-%m-%d"))
head(Event_TrendPerCountry)
```

Next, we have to add a cumulative frequency column, so the total number of incidents is recorded for each date and for each country.

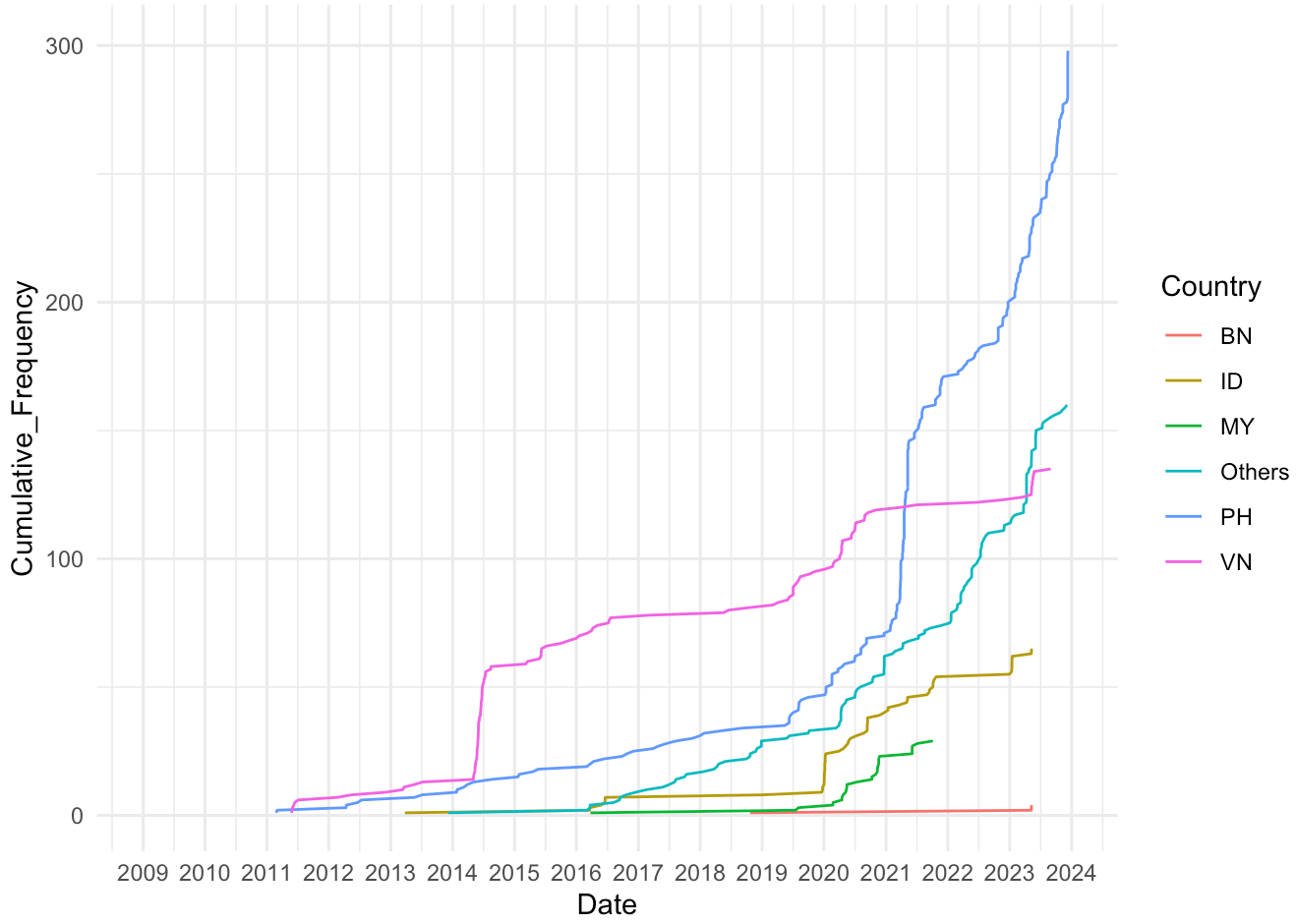
```
#sort data by Country and Date
Event_TrendPerCountry <- Event_TrendPerCountry %>%
  arrange(Country, Date)

#add cumulative frequency column
Event_TrendPerCountry <- Event_TrendPerCountry %>%
  group_by(Country) %>%
  mutate(Cumulative_Frequency = row_number())
head(Event_TrendPerCountry)
```

```
Event_TrendPerCountry %>%
  filter(Country == "ID")
```

Now that the dataframe is done, I can now plot the numerical data as a line chart with countries as the categories with a legend.

```
ggplot(Event_TrendPerCountry, aes(x = Date, y = Cumulative_Frequency, color = Country))
+
  geom_line() +
  theme_minimal() +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  expand_limits(x = as.Date(c("2009-01-01", "2023-12-31")))
```

```
## Warning: Removed 18 rows containing missing values or values outside the scale range
## (`geom_line()`).
```




```
ggsave("PerCountry.svg", device = "svg")
```

```
## Saving 7 x 5 in image
```

```
## Warning: Removed 18 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

Difference in biggest vessel size

```
vessel <- read.csv("VesselSize.csv", header=FALSE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## incomplete final line found by readTableHeader on 'VesselSize.csv'
```

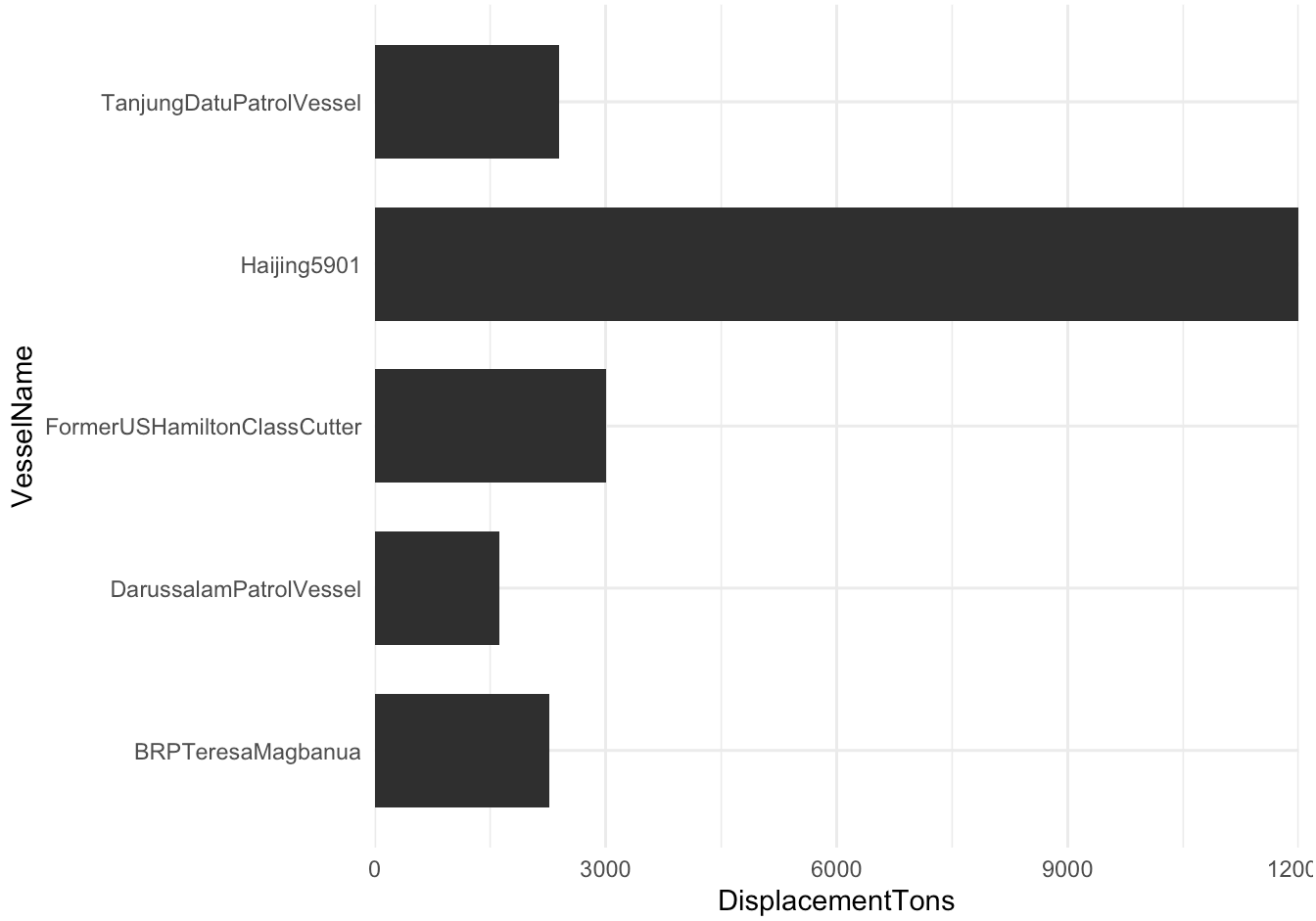
```
head(vessel)
```

```
colnames(vessel) = c('Country', 'VesselName', 'DisplacementTons')
head(vessel)
```

```
vessel %>%
  arrange(DisplacementTons)
```

```
##       Country                 VesselName DisplacementTons
## 1      Brunei      DarussalamPatrolVessel             1625
## 2 Philippines           BRPTeresaMagbanua             2265
## 3   Indonesia     TanjungDatuPatrolVessel             2400
## 4     Vietnam FormerUSHamiltonClassCutter             3000
## 5       China                 Haijing5901            12000
```

```
ggplot(vessel, aes(x = VesselName, y = DisplacementTons)) +
  geom_tile(width = 0.7, aes(y = DisplacementTons / 2, height = DisplacementTons)) +
  coord_flip() +
  scale_y_continuous(expand = c(0, 0)) +
  theme_minimal()
```



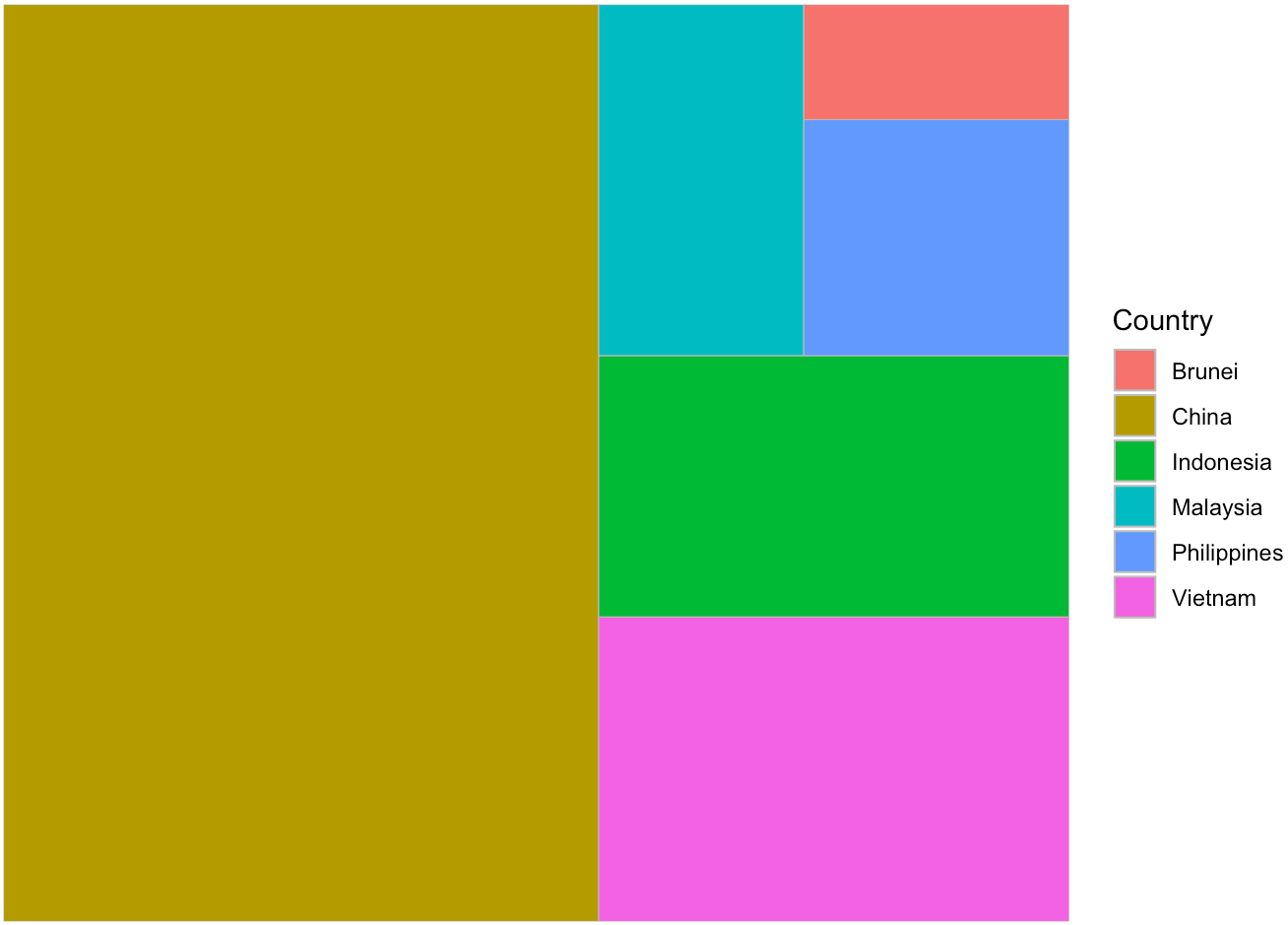

```
ggsave("vesselsize.svg", device = "svg")
```

```
## Saving 7 x 5 in image
```

```
MCI <- read.csv("MilitaryCapabilityIndex.csv")
MCI
```

```
##       Country MilitaryCapabilityIndex
## 1       China                    69.7
## 2     Vietnam                    18.3
## 3    Malaysia                     9.2
## 4 Philippines                     8.0
## 5      Brunei                     3.9
## 6   Indonesia                    15.7
```

```
ggplot(MCI, aes(area = MilitaryCapabilityIndex, fill = Country)) +
  geom_treemap() +
  theme_minimal()
```




```
ggsave("MCITreeMap.svg", device = "svg")
```

```
## Saving 7 x 5 in image
```

```
MSP <- read.csv("MaritimeSeaPower.csv")
head(MSP)
```

```
MSP <- MSP %>%
  filter(Country %in% c("China", "Vietnam", "Malaysia", "Philippines", "Brunei", "Indone
sia"))
head(MSP)
```

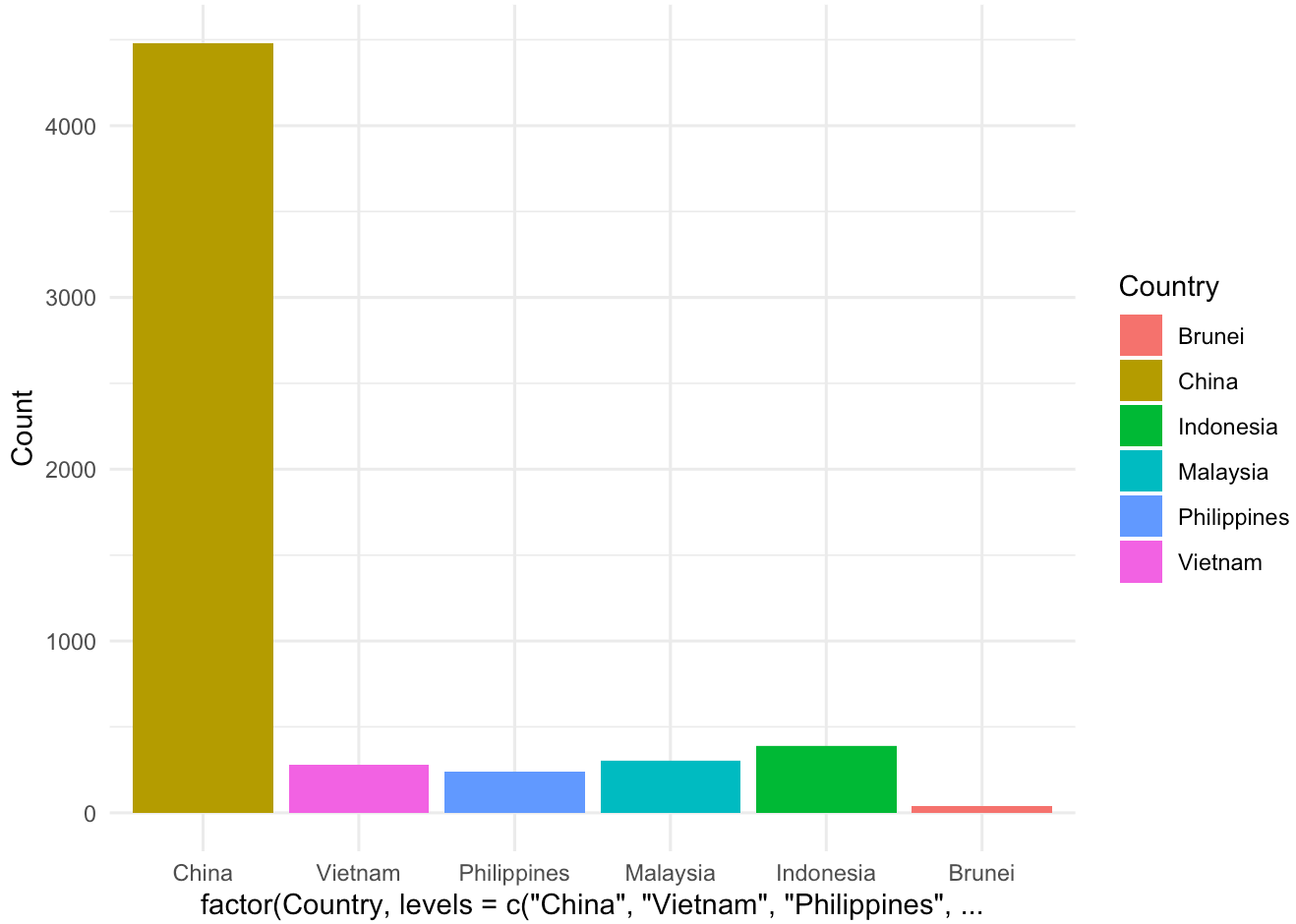
```
MSP_Count <- MSP %>%
  mutate(Total = rowSums(select(., -Country), na.rm = TRUE))
head(MSP_Count)
```

```
MSP_Long <- MSP %>%
  pivot_longer(cols = -Country, names_to = "AssetType", values_to = "Count") %>%
  filter(Count > 0)
head(MSP_Long)
```

```
MSP_Count <- MSP_Long %>%
  group_by(Country) %>%
  summarise(Total = sum(Count, na.rm = TRUE))
MSP_Count
```

```
## # A tibble: 6 × 2
##   Country     Total
##   <chr>       <int>
## 1 Brunei         39
## 2 China        4481
## 3 Indonesia     388
## 4 Malaysia      302
## 5 Philippines   237
## 6 Vietnam       282
```

```
ggplot(MSP_Long, aes(x = factor(Country, levels = c("China", "Vietnam", "Philippines",
"Malaysia", "Indonesia", "Brunei")), y= Count, fill = Country)) +
  geom_bar(stat="identity") +
  theme_minimal()
```




```
ggsave("MSP.svg", device = "svg")
```

```
## Saving 7 x 5 in image
```

```
df_china_subi_reef <- df_china %>%
  filter(Latitude >= 10, Latitude <= 14, Longitude >= 111, Longitude <= 115)
head(df_china_subi_reef)
```

```
df_china <- df_china %>%
  select("Latitude", "Longitude")
head(df_china)
```

```
names(df_china) <- NULL
head(df_china)
```

```
write.csv(df_china,"SCSData_Clean.csv", row.names = FALSE)
```

```
write.csv(df_china_subi_reef,"SCSData_Subi.csv", row.names = FALSE)
```