EDA

Robert Ford

Table 1: Sample of the Data

region	buyer	supplier	amount	award
Federal	Canadian Centre for Occupational Health and Safety	IPSS Inc	39958	2024-
Federal	Canadian Centre for Occupational Health and Safety	Articulate Global LLC	22785	2024-
Federal	Canadian Centre for Occupational Health and Safety	Colony Labs, dba Scribe	12672	2024-
Federal	Canadian Centre for Occupational Health and Safety	Sheraton Suites Calgary Eau Claire	100000	2024-
Federal	Canadian Centre for Occupational Health and Safety	ROSS & McBRIDE LLP	20000	2024-

Figure 1: template

This graph shows how much was awarded by month

#TODO Color code by buyer

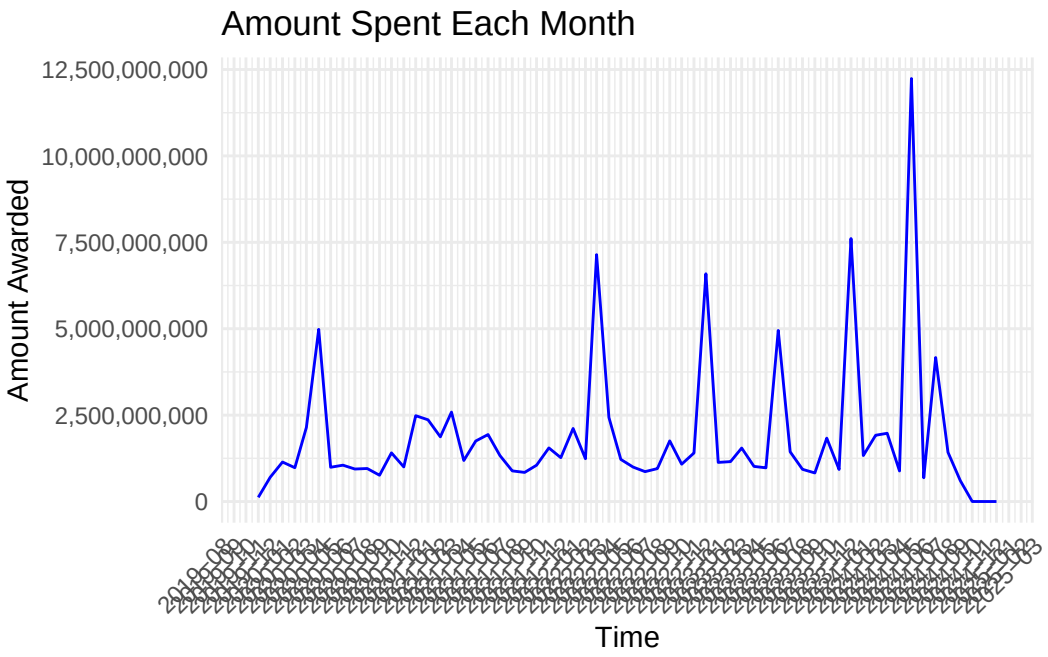


Figure 2: template

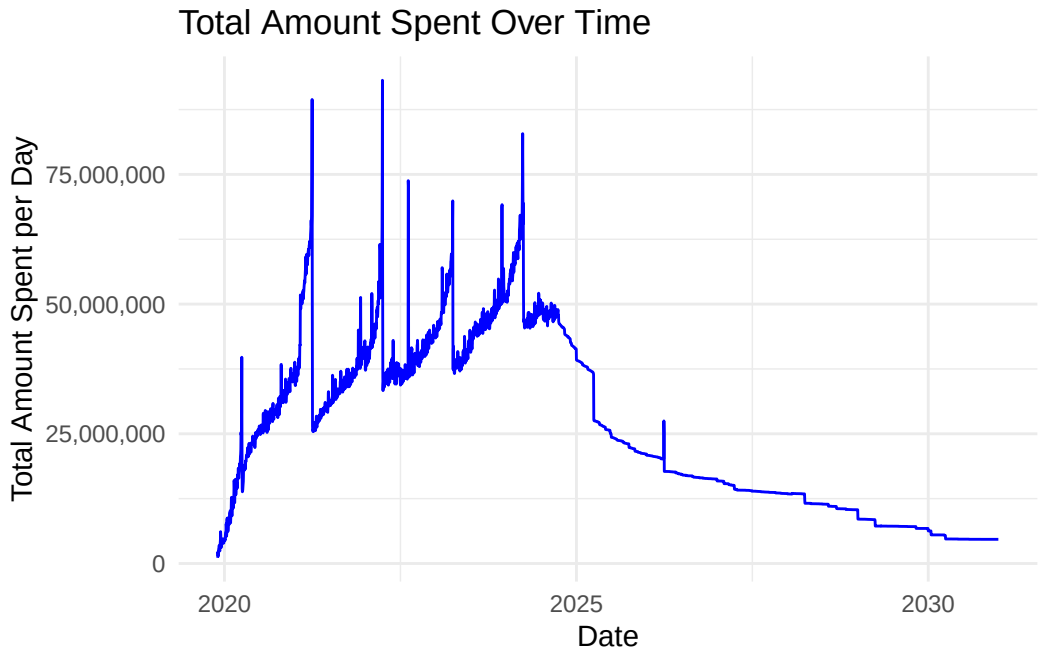


Figure 3: template

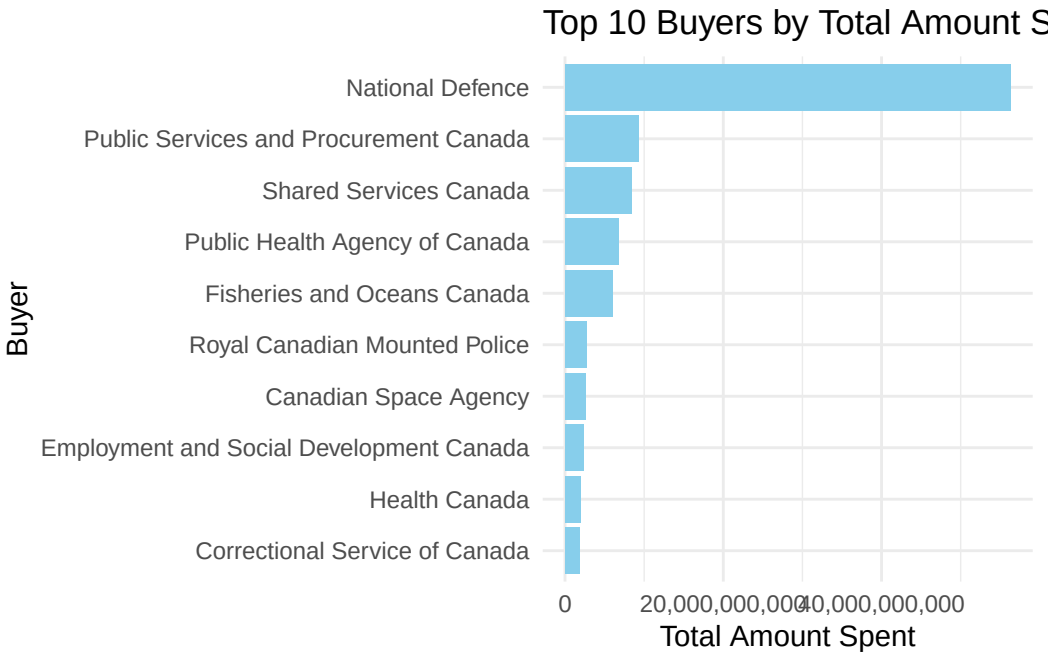


Figure 4: template

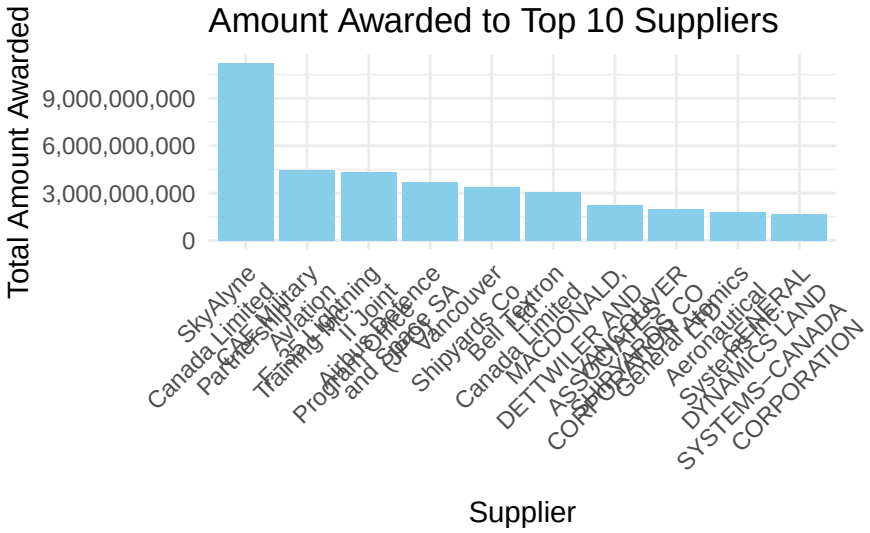


Figure 5: template

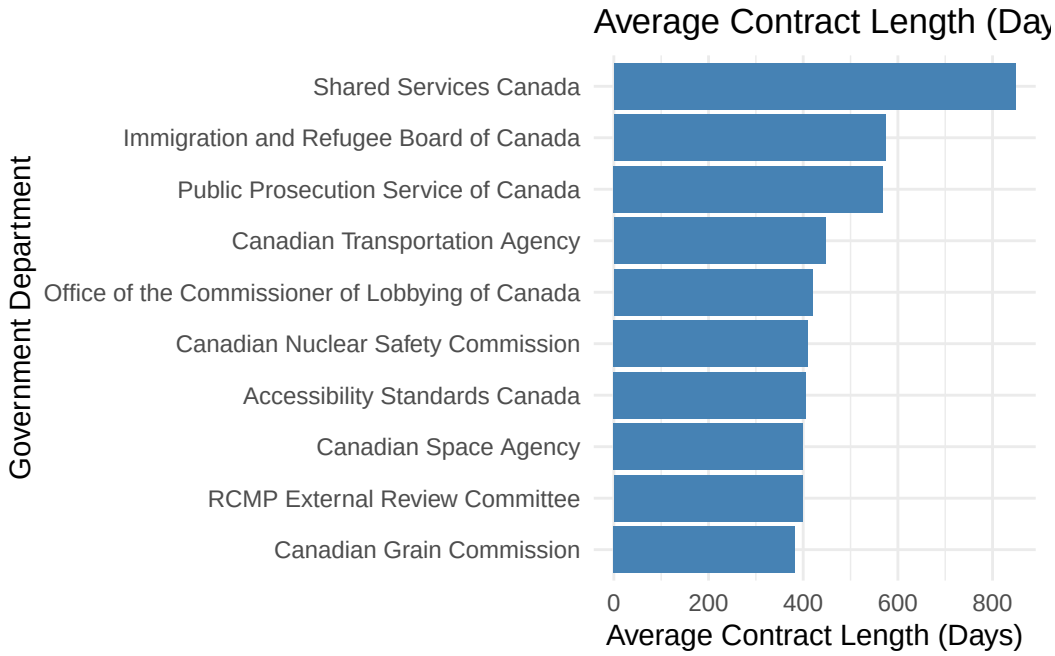


Figure 6: Average Contract Length (Days) by Government Department (Top 10)

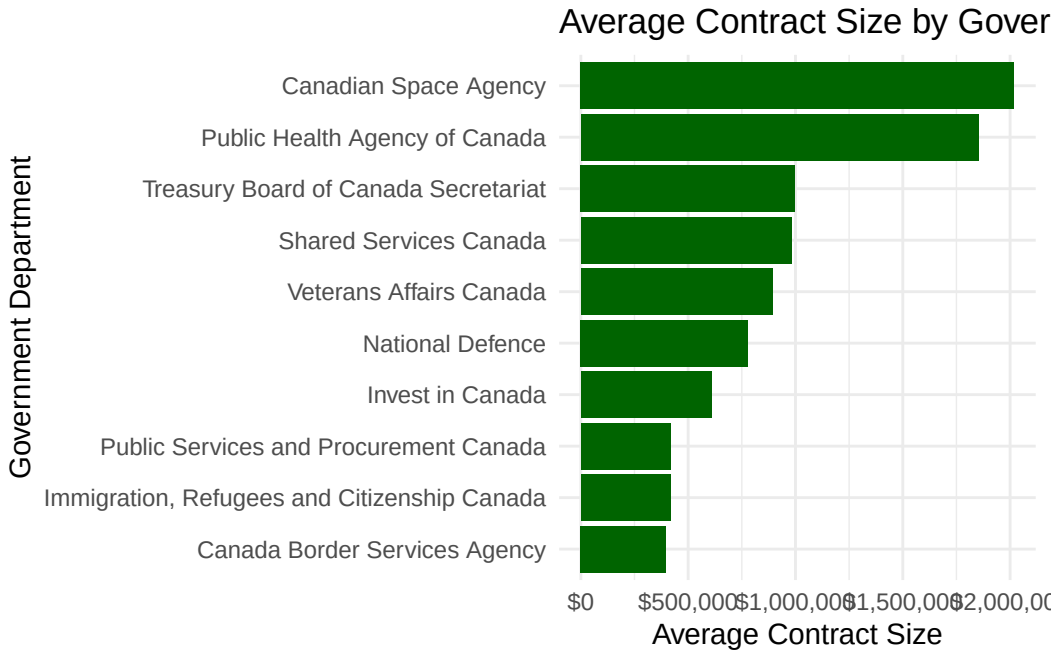
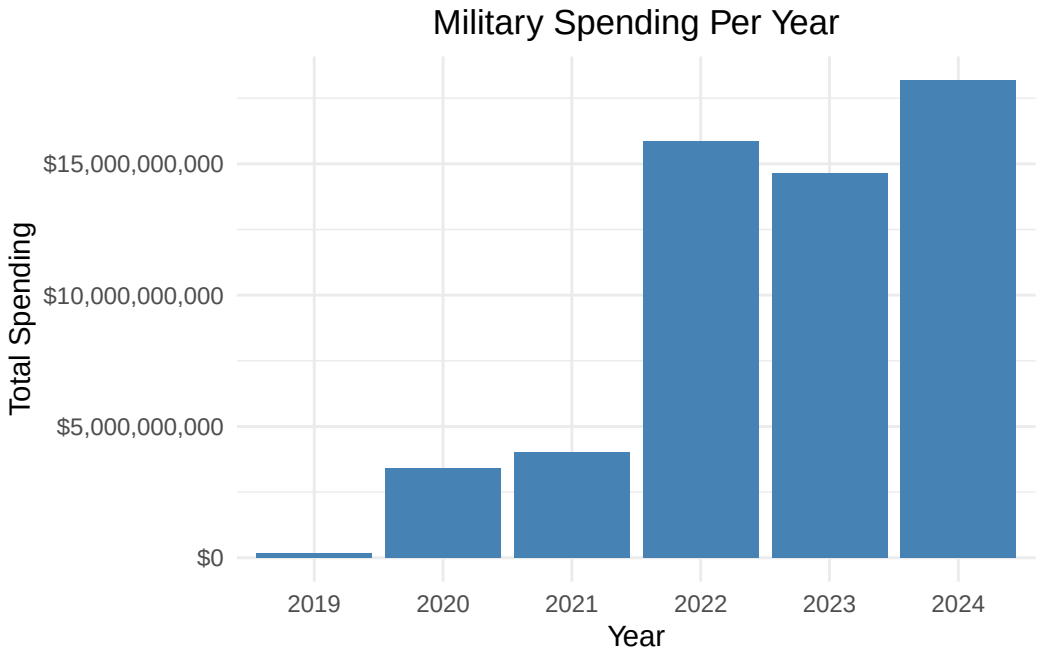


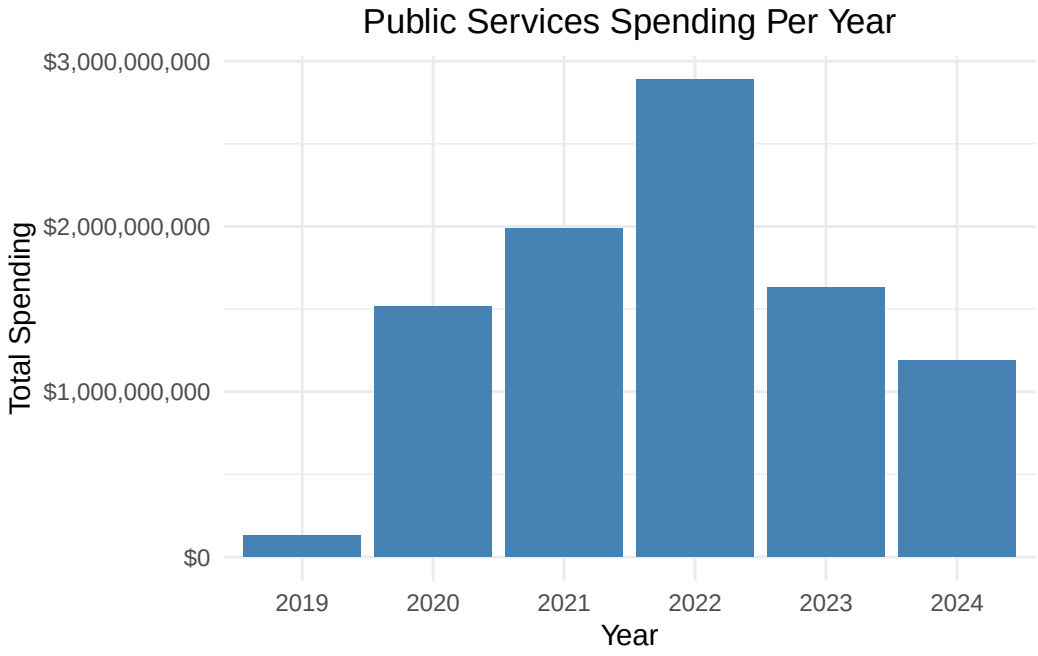
Figure 7: Average Contract Size by Government Department (Top 10)

```r
# Convert award_date to Date type and extract the year
df$award_date <- as.Date(df$award_date, format = "%Y-%m-%d")
df$year <- format(df$award_date, "%Y")

# Filter for military spending (where buyer is "National Defence")
military_df <- df %>%
  filter(buyer == "National Defence")

# Summarize total spending per year
military_spending_per_year <- military_df %>%
  group_by(year) %>%
  summarize(total_spending = sum(amount, na.rm = TRUE))

# Create the bar chart
ggplot(data = military_spending_per_year, aes(x = year, y = total_spending)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Military Spending Per Year", x = "Year", y = "Total Spending") +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$")) +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
df$year <- format(df$award_date, "%Y")
```

```r
# Filter for military spending (where buyer is "National Defence")
public_services_df <- df %>%
  filter(buyer == "Public Services and Procurement Canada")

# Summarize total spending per year
public_spending_per_year <- public_services_df %>%
  group_by(year) %>%
  summarize(total_spending = sum(amount, na.rm = TRUE))

# Create the bar chart
ggplot(data = public_spending_per_year, aes(x = year, y = total_spending)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Public Services Spending Per Year", x = "Year", y = "Total Spending") +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$")) +
  theme(plot.title = element_text(hjust = 0.5))
```

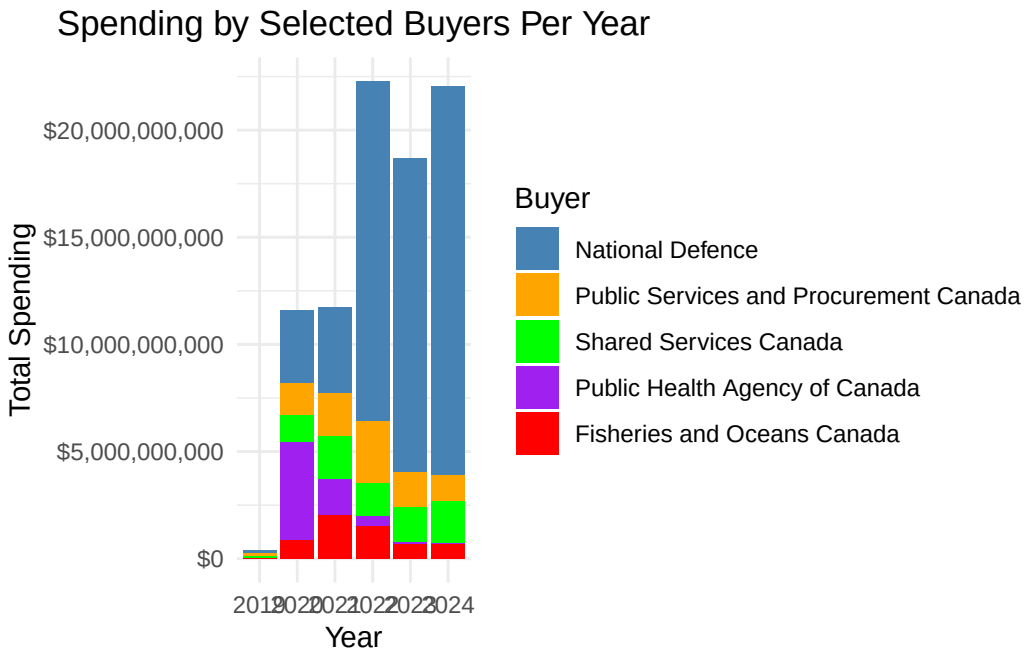



```r
df$year <- format(df$award_date, "%Y")

# Filter data for the specified buyers
selected_buyers <- c(
  "National Defence",
  "Public Services and Procurement Canada",
```

```
  "Shared Services Canada",
  "Public Health Agency of Canada",
  "Fisheries and Oceans Canada"
)

stacked_df <- df %>%
  filter(buyer %in% selected_buyers) %>%
  group_by(year, buyer) %>%
  summarize(total_spending = sum(amount, na.rm = TRUE)) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

```
# Ensure the order of the stacks matches the specified order
stacked_df$buyer <- factor(stacked_df$buyer, levels = selected_buyers)

# Create the stacked bar chart
ggplot(data = stacked_df, aes(x = year, y = total_spending, fill = buyer)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(
    values = c("National Defence" = "steelblue",
               "Public Services and Procurement Canada" = "orange",
               "Shared Services Canada" = "green",
               "Public Health Agency of Canada" = "purple",
               "Fisheries and Oceans Canada" = "red")
  ) +
  labs(
    title = "Spending by Selected Buyers Per Year",
    x = "Year",
    y = "Total Spending",
    fill = "Buyer"
  ) +
  theme_minimal() +
  scale_y_continuous(labels = dollar_format(prefix = "$")) +
  theme(plot.title = element_text(hjust = 0.5))
```

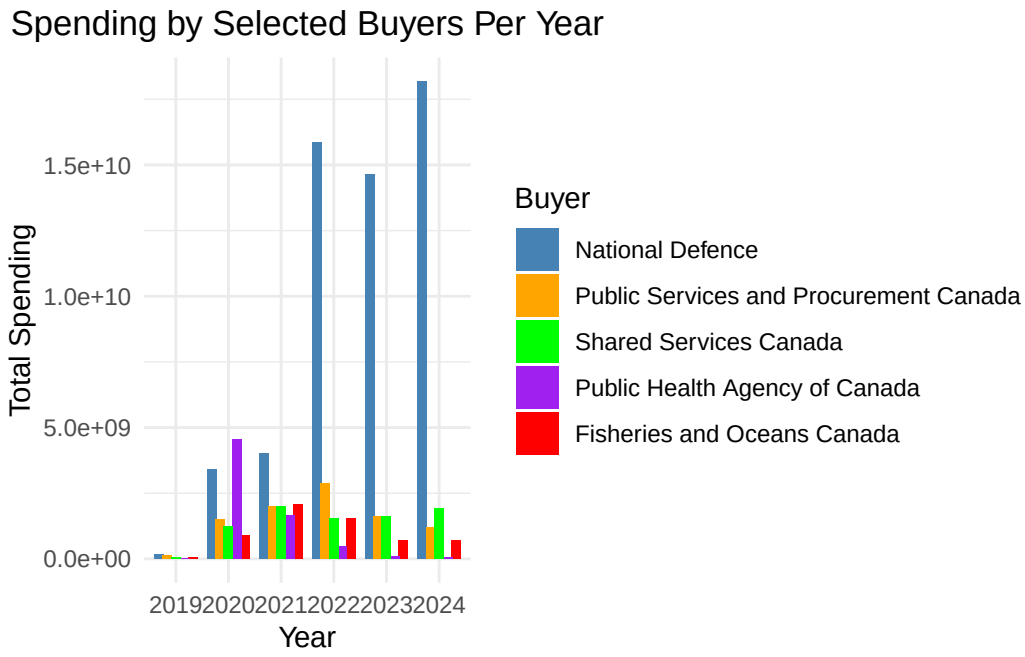



```
df$year <- format(df$award_date, "%Y")

# Filter data for the specified buyers
selected_buyers <- c(
  "National Defence",
  "Public Services and Procurement Canada",
  "Shared Services Canada",
  "Public Health Agency of Canada",
  "Fisheries and Oceans Canada"
)

grouped_df <- df %>%
  filter(buyer %in% selected_buyers) %>%
  group_by(year, buyer) %>%
  summarize(total_spending = sum(amount, na.rm = TRUE)) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

```
# Ensure the order of the groups matches the specified order
grouped_df$buyer <- factor(grouped_df$buyer, levels = selected_buyers)
```

```r
# Create the grouped bar chart
ggplot(data = grouped_df, aes(x = year, y = total_spending, fill = buyer)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  scale_fill_manual(
    values = c("National Defence" = "steelblue",
               "Public Services and Procurement Canada" = "orange",
               "Shared Services Canada" = "green",
               "Public Health Agency of Canada" = "purple",
               "Fisheries and Oceans Canada" = "red")
  ) +
  labs(
    title = "Spending by Selected Buyers Per Year",
    x = "Year",
    y = "Total Spending",
    fill = "Buyer"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
df$year <- format(df$award_date, "%Y")

# Define the selected buyers
selected_buyers <- c(
```

```
  "National Defence",
  "Public Services and Procurement Canada",
  "Shared Services Canada",
  "Public Health Agency of Canada",
  "Fisheries and Oceans Canada"
)

# Add a new column for buyer categories
df <- df %>%
  mutate(buyer_category = ifelse(buyer %in% selected_buyers, buyer, "Other"))

# Summarize total spending per year for each buyer category
grouped_df <- df %>%
  group_by(year, buyer_category) %>%
  summarize(total_spending = sum(amount, na.rm = TRUE)) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

```
# Ensure the order of the groups matches the specified order, with "Other" last
grouped_df$buyer_category <- factor(
  grouped_df$buyer_category,
  levels = c(selected_buyers, "Other")
)

# Create the grouped bar chart
ggplot(data = grouped_df, aes(x = year, y = total_spending, fill = buyer_category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  scale_fill_manual(
    values = c(
      "National Defence" = "steelblue",
      "Public Services and Procurement Canada" = "orange",
      "Shared Services Canada" = "green",
      "Public Health Agency of Canada" = "purple",
      "Fisheries and Oceans Canada" = "red",
      "Other" = "gray"
    )
  ) +
  labs(
    title = "Spending by Selected and Other Buyers Per Year",
```

```
    x = "Year",
    y = "Total Spending",
    fill = "Buyer"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

nding by Selected and Other Buyers Per Year