Exploration of How Trading with the US in Aerospace and Defense Affects Country Factors

STAT 184 - Final Project

Ishan Agrahar

2025-06-30

Table of contents

Introduction

The United States is the world leader in Defense Spending and it is therefore natural that they have a lot of activity with regards to exporting Defensive Technology to other Countries. This report is going to analyze the exports of Aerospace Defense Technology by the US on a year-by-year basis, and attempt to find if a correlation exists between some factors as specified by the below research questions.

To guide analysis, this report will aim to answer the four following research questions.

1. If a country spends more money importing US Aerospace Defense, does that correlate to an increase in Happiness?
2. How much of a countries GDP is allocated towards importing US Aerospace Defense?
3. How does population impact how much a country spends on importing US Aerospace Defense?
4. Is a country's air pollution score impacted by importing US Aerospace Defense?

These questions are interesting because in modern times we are seeing a push for defense spending to increase (i.e the US pushing for higher NATO defense spending). As a world leader and influential country it is interesting to see if there is a correlation between US Aerospace Defense and metrics of daily life that speaks to that influence.

Note: Correlation does not imply Causation. Therefore if by chance there is a correlation found we can only speculate causation and further research would be needed to find if Causation exists which is outside the scope of this report.

Background Information

From the research questions, it is clear that we are comparing multiple metrics to a country's spending on importing US Aerospace Defense. Therefore the best way to answer the research questions is to find data which (1) shows the importing figures for US Aerospace Defense by Country and Year, and (2) represents the metric as specified by each question. Therefore we can further summarize the research questions according to a primary dataset and a corresponding metric dataset.

The metric of the primary dataset shall be: US Aerospace Defense Exports per Year

The corresponding metric of the metric datasets correspond to the above research questions and are as follows:

1. Country Happiness Score per Year
2. Country GDP per Year
3. Country Population per Year
4. Air Pollution per Year

Note that analysis must be done for the years 2015-2024 because those are the years available in the primary dataset.

To answer the research questions, this report will use the powerful tool of visualization where I essentially will take the primary dataset metric of US Aerospace Defense Exports by Country and do a year-by-year comparison against the corresponding metric for some countries and years using scatterplots and line graphs.

Note that I will further specify the countries that are going to be used in analysis because there is a total of fourteen countries reported in the primary dataset.

Data Summary

Primary Dataset Provenance and Attributes - US Aerospace Defense Exports

Where did the data come from: The data comes from the Official Website of the United States International Trade Administration[1]

What purpose did they collect the data: According to the International Trade Administration, they collect data for the purpose of promoting trade and investment, strengthening the competitiveness of U.S. industry, and ensuring fair trade and compliance with trade laws and agreements.

What are the cases: The cases are the amount of money a specific country imported for an item of US Aerospace Defense Technology in a specific year.

Interesting Attributes: We will use the Country, Year, and Value in USD.

The below table is the top three entries of the cleaned dataset to signify a sample data entry.

Year	valueInUSD	country	item
2015	2630	Czech Republic	8805210000–AIR COMBAT SIMULATORS AND PARTS THEREOF
2015	3500	Brazil	8411114050–TURBOJET A/C TURBINES EXC CIVIL, THRUST LE 25 KN
2015	4000	Mexico	8805210000–AIR COMBAT SIMULATORS AND PARTS THEREOF

Metric Dataset Provenance and Attributes - Country Happiness Score

Where did the data come from: The data comes from the World Happiness Report where the files are downloaded from Kaggle for 2015 to 2023[2] and directly from the World Happiness Report for 2024[3].

What purpose did they collect the data: According to the World Happiness Report: By making the essential insights from wellbeing science accessible to all, we give everyone the knowledge to create more happiness for themselves and others.

What are the cases: Each case is the happiness of a country in a year.

[1]Data collected by the US International Trade Administration located at https://www.trade.gov/data-visualization/aerospace-and-defense-data

[2]From Kaggle via Sazidul Islam: https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors

[3]World Happiness Report: https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world

Interesting Attributes: We are interested in country, happiness_score, and year.

The below table is the top three entries of the cleaned dataset to signify a sample data entry.

Table 2: Sample Data Entries

country	happiness_score	Year
Switzerland	7.587	2015
Iceland	7.561	2015
Denmark	7.527	2015

Metric Dataset Provenance and Attributes - Country GDP

Where did the data come from: The data comes from the World Bank Group[4].

What purpose did they collect the data: According to the World Bank Group: Timely and reliable statistics are key inputs to the broad development strategy. Improvements in the quality and quantity of data on all aspects of development are essential if we are to achieve the goal of a world without poverty.

What are the cases: The cases are a country's GDP in a certain year.

Interesting Attributes: We are interested in Country, Year, and GDP in US Dollar.

The below table is the top three entries of the cleaned dataset to signify a sample data entry.

Table 3: Sample Data Entries

country	Year	GDP(USD)
Aruba	1986	405586592
Aruba	1987	487709497
Aruba	1988	596648045

Metric Dataset Provenance and Attributes - Country Population

Where did the data come from: The data comes from the World Bank Group[5].

What purpose did they collect the data: According to the World Bank Group: Timely and reliable statistics are key inputs to the broad development strategy. Improvements in the

[4]World Bank Group: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
[5]World Bank Group: https://data.worldbank.org/indicator/SP.POP.TOTL

quality and quantity of data on all aspects of development are essential if we are to achieve the goal of a world without poverty.

What are the cases: The cases are a countries population in a certain year.

Interesting Attributes: We are interested in country, year, and population.

The below table is the top three entries of the cleaned dataset to signify a sample data entry.

Table 4: Sample Data Entries

country	Year	Population
Aruba	1960	54922
Aruba	1961	55578
Aruba	1962	56320

Metric Dataset Provenance and Attributes - Air Pollution

Where did the data come from: The data is scraped from Wikipedia[6] where they got the table from IQAir 2023 World Air Quality Ranking.

What purpose did they collect the data: According to IQAir, they collect data to empower communities to monitor air quality and support clean air initiatives worldwide[7]

What are the cases: The cases are a countries particulate matter amount in a year.

Interesting Attributes: We are interested in country, particulate matter, and year.

The below table is the top three entries of the cleaned dataset to signify a sample data entry.

Table 5: Sample Data Entries

country	particulateMatter	Year
India	97.1	2018
India	83.3	2019
India	77.1	2020

[6]Air Pollution Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_air_pollution
[7]IQAir: https://www.iqair.com/about-iqair

EDA

Primary Analysis - Countries

The first step I will take to answer the research questions is to explore the primary dataset in attempt to gain more insight of what I can state.

Therefore the first thing I will do is attempt to see what countries are represented by the data. This is exemplified by the below table.

Table 6: Table of a Country vs. Amount they spent importing US Aerospace Defense Technology

country	TotalSpending
Saudi Arabia	8251973477
United Kingdom	5529518157
Israel	4170510025
Singapore	3495911993
Turkey	3071005228
Germany	2982643668
Mexico	2867115003
United Arab Emirates	2765597504
Canada	2668668014
Poland	1606459607
Brazil	1135396633
Czech Republic	579069813
Finland	320628567
Ukraine	10968279

Figure 1: Table of a Country vs. Amount they spent importing US Aerospace Defense Technology

Looking at Figure 1 it is therefore found that only fourteen countries are represented in the primary dataset and therefore in moving forth all following datasets must adhere to analyzing those countries only.

Primary Analysis - Yearly Distribution

The next thing I will do is take a look at the summarized spending of the data to try and get an idea of how the data is distributed yearly.

Table 7: Aerospace Defense Exported by US each Year 2015-2024

Year	totalSpent
2015	3468811283
2016	3788748338
2017	3706836767
2018	4232750903
2019	4078145853
2020	3789740910
2021	3340363113
2022	3973916278
2023	4699210905
2024	4376941618

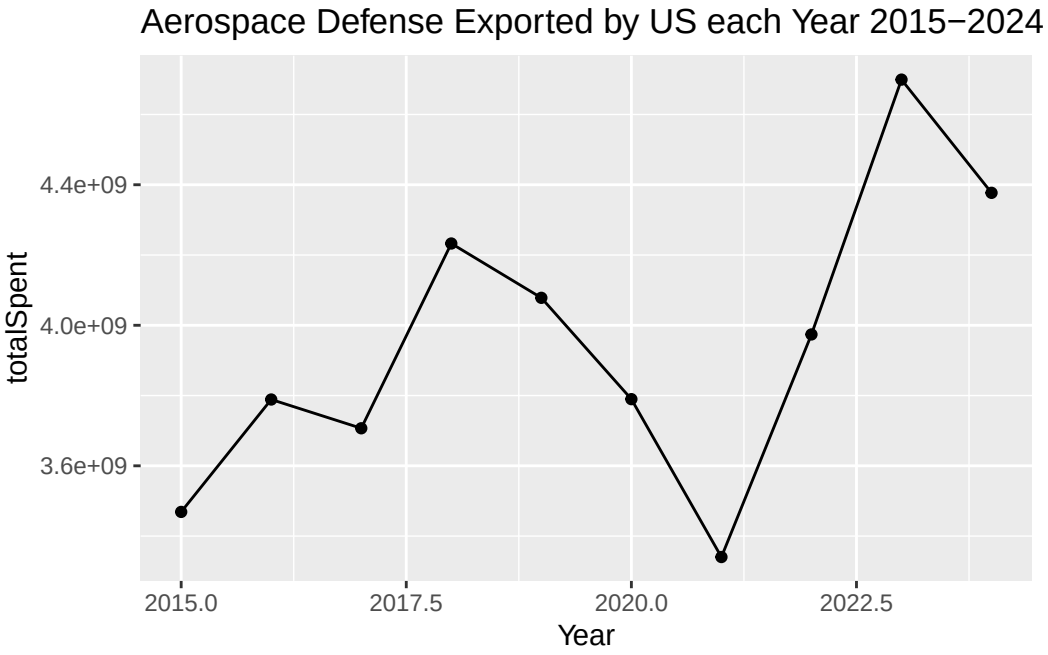

Figure 2: Yearly Distribution of Trade

As can be seen by Figure 2 it seemingly increases till 2018 and then trade goes down likely due to COVID, and then picks back up heavily after 2021. This trend may be useful when we try to answer the research questions.

Lastly I will look at how different countries spent money over the years.

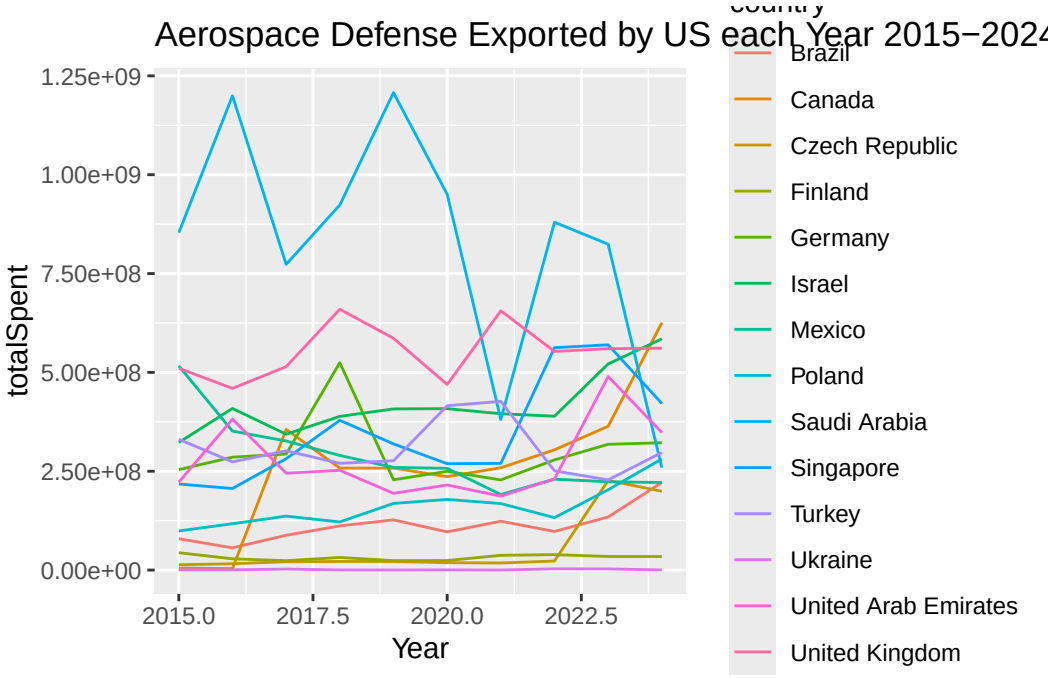

Figure 3: Yearly Distribution of Trade by Country

As can be seen in Figure 3 there is a variety of spending across the years where Saudi Arabia seems to be the highest spending across most years with Ukraine being the lowest.

The next part of data analysis will focus on the primary dataset against the metric datasets for the research questions.

Research Question 1

This section aims to answer the research question: If a country spends more money importing US Aerospace Defense, does that correlate to an increase in Happiness?

To do this I will make a visualization of happiness score versus value exported from the US Aerospace Defense to that country to see if there is a visible correlation. At this point it is more convient to look at the sum of spending per year so I will now remove the specific items from each year/country and just have the cases for the primary dataset be the amount a country spent in importing US Aerospace Defense Technology in a year.

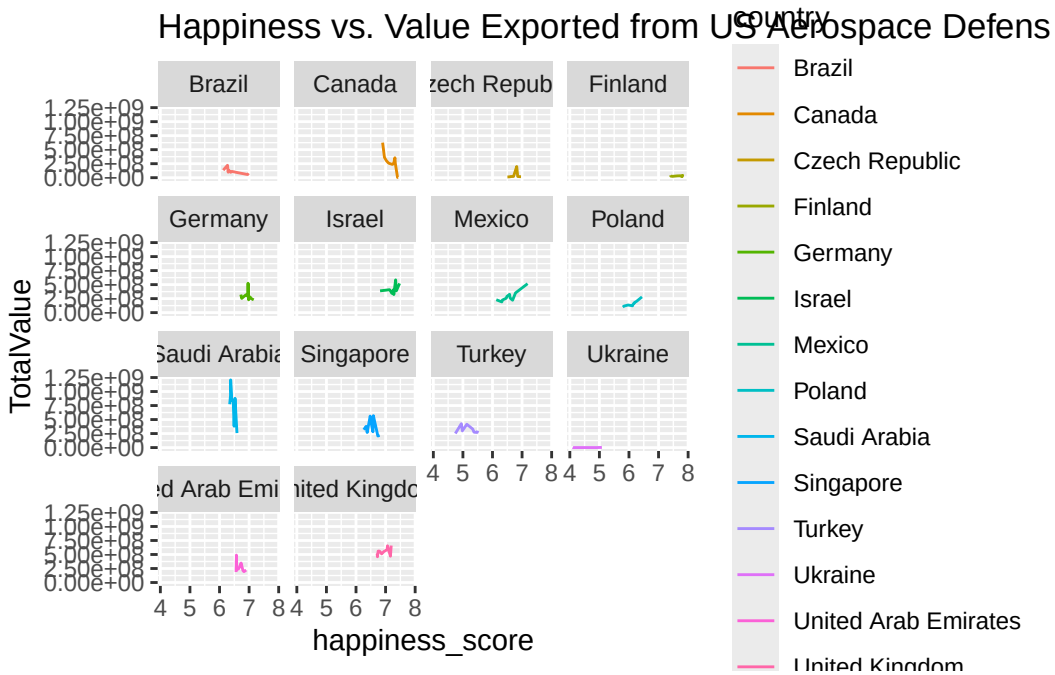

Figure 4: Happiness vs. Value

As seen in Figure 4 there unfortunately does not seem to be a general correlation to trading with the US and happiness, note that some countries like Saudi Arabia and Canada seem to be be happier when there is less trade and opposite for Mexico, but there is mixed results elsewhere.

Research Question 2

This section aims to answer the research question: How much of a countries GDP is allocated towards importing US Aerospace Defense?

To do this I will display a ratio where I take the countries value spent on aerospace trade and divide that by the countries total GDP.

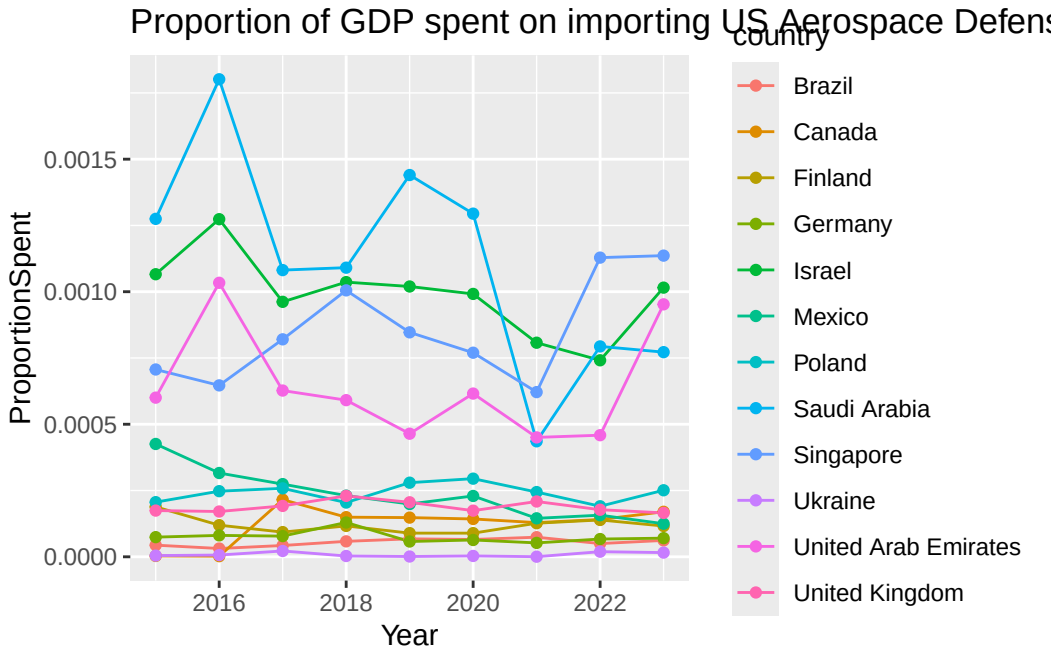

Figure 5: Proportion of GDP spent on importing US Aerospace Defense

Looking at Figure 5 it seems that the trend roughly seems to follow the total spent by country per year analysis that was done in Primary Analysis. As can be seen Saudi Arabia has the top trading proportion of GDP till 2021 where Israel led in 2021 and then Singapore in 2022 and 2023. This is stunning because Saudia Arabia continued to have the highest total spent by country per year as noted in Figure 3. Ukraine, as expected from total spent by country per year, is the lowest in proportion.

Research Question 3

This section aims to answer the research question: How does population impact how much a country spends on importing US Aerospace Defense?

To answer this question I will make a graph across the years of population and then put the value spent on imports in size so we can see if there is a relationship with year too.

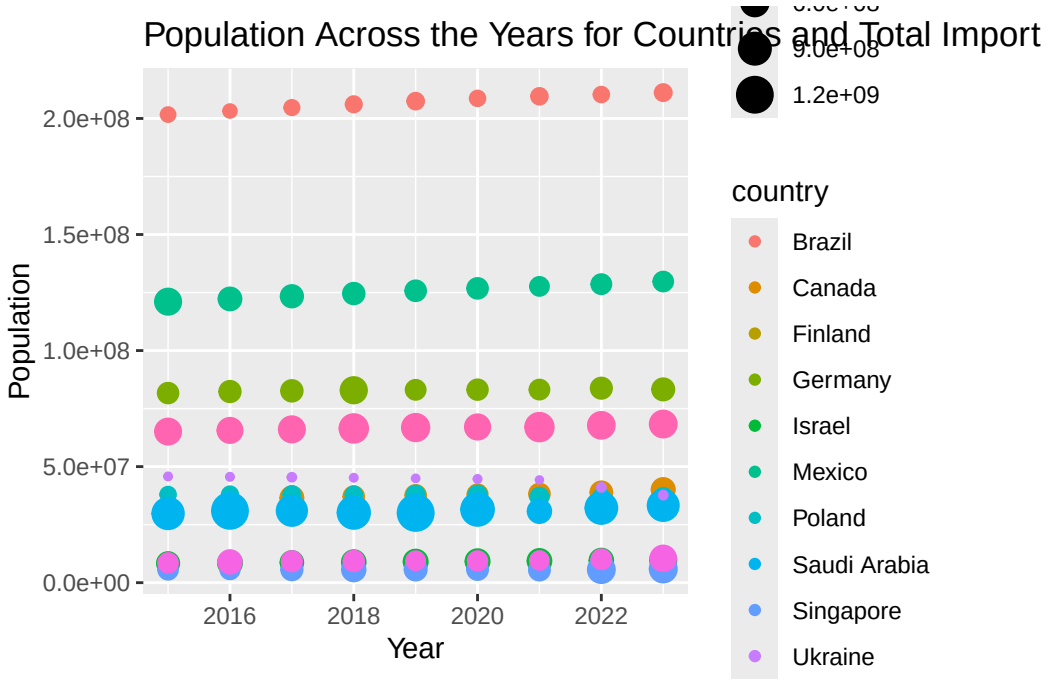

Figure 6: Population Across Years

From Figure 6, there unfortunately is not a general trend and most countries look like there is no correlation but there may be a trend with Israel and Brazil for example where as population increases over the years, defense decreases.

Research Question 4

This section aims to answer the research question: Is a country's air pollution score impacted by importing US Aerospace Defense?

To do this a line graph of the particulate matter (how bad the air is) versus the total value of imports for each country should be sufficient to see if there is a correlation.

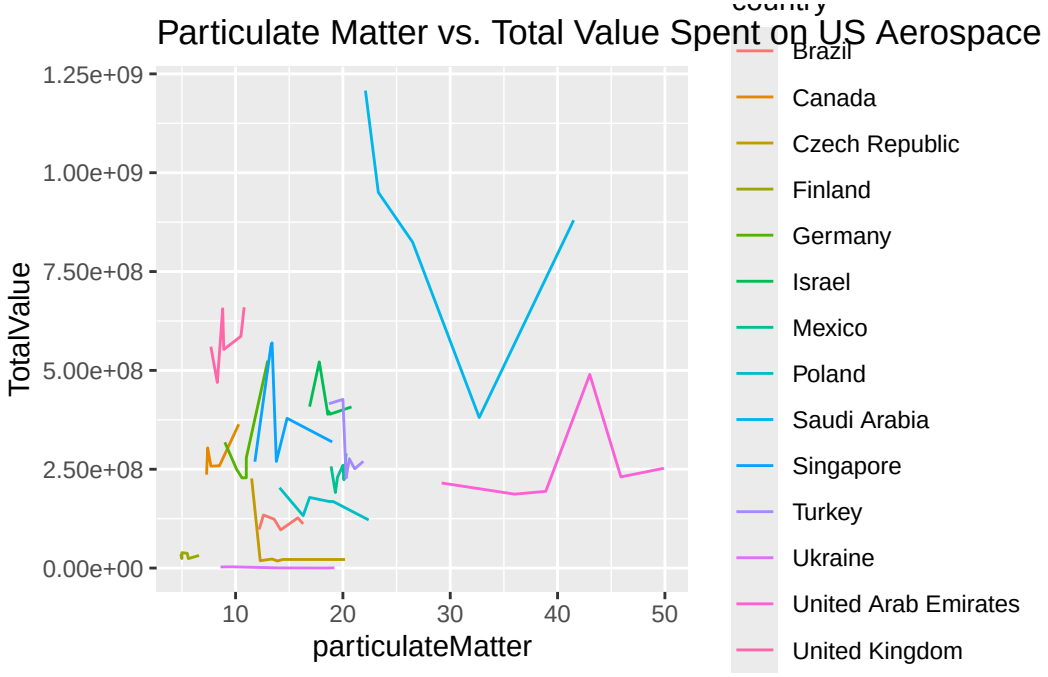

Figure 7: Particulate Matter vs. Total Value Import by Country

Looking at Figure 7 there does not really seem to be a correlation at all between trading US Aerospace Defense and particulate matter so I unfortunately must conclude for this question that there is not a trend at current scale for this metric.

Conclusion

To first remind of what the research questions that guide this report see below

1. If a country spends more money importing US Aerospace Defense, does that correlate to an increase in Happiness?
2. How much of a countries GDP is allocated towards importing US Aerospace Defense?
3. How does population impact how much a country spends on importing US Aerospace Defense?
4. Is a country's air pollution score impacted by importing US Aerospace Defense?

Through the powerful tool of visualization I have explored the trends in the data at a country basis to see if there are trends in the above metrics compared to my primary dataset of importing US Aerospace Defense. To conclude I will summarize what was ascertained in each visualization. For the first question I found that there was not a general trend but for some countries there did seem to be a correlation where if a country spends more money importing US Aerospace Defense it does that correlate to an increase in Happiness; as well as the opposite of that trend. For the second question I found that the proportion of a countries GDP had parallels to the amount that they spent in terms of ranking of countries and suprisingly that Saudi Arabia's proportion went down while they still had the highest total spending indicating that their economy must've greatly increased. In the third question I found that there was not a general trend but for some countries there may be a correlation which could be explored more on a country-by-country basis. Note that a general trend was not discovered for any research question except for the Question 4 where I state with a good amount of confidence that, based on current data, a country's air pollution score is not impacted by importing US Aerospace Defense.

Cumulatively looking at the questions as a whole, it would be at a better benefit to look at the data from a country per country basis as my generalization hurt my ability to find trends and geographically there are going to be differences.

References

"Aerospace and Defense Data." *International Trade Administration | Trade.Gov*, www.trade.gov/data-visualization/aerospace-and-defense-data. Accessed 30 June 2025.

"First in Air Quality." *IQAir*, www.iqair.com/about-iqair. Accessed 30 June 2025.

"GDP (Current US$)." *World Bank Open Data*, data.worldbank.org/indicator/NY.GDP.MKTP.CD. Accessed 30 June 2025.

"Happiest Countries in the World 2025." *Worldpopulationreview.Com*, worldpopulationreview.com/country-rankings/happiest-countries-in-the-world. Accessed 30 June 2025.

Islam, Sazidul. "World Happiness Report (till 2023)." *Kaggle*, 9 Sept. 2023, www.kaggle.com/datasets/sazidthe1/happiness-scores-and-factors.

"List of Countries by Air Pollution." *Wikipedia*, Wikimedia Foundation, 21 May 2025, en.wikipedia.org/wiki/List_of_countries_by_air_pollution.

"Population, Total." *World Bank Open Data*, data.worldbank.org/indicator/SP.POP.TOTL. Accessed 30 June 2025.

Code Appendix

```
## Style Guide: The Tidyverse Style Guide using styler for compliance
## Note: Set your working directory to the STAT-184-Expenditure-Analysis folder

# ----------------------------------------------------------------

## Dependencies
library(dplyr)
library(tidyr)
library(readr)
library(rvest)
library(knitr)
library(ggplot2)

# ----------------------------------------------------------------

## Scraped Data

### This will scrape the air pollution wikipedia table and save it
web_page <- "https://en.wikipedia.org/wiki/List_of_countries_by_air_pollution"
SetOfTables <- web_page %>%
  read_html() %>%
  html_nodes(css = "table") %>%
  html_table(fill = TRUE)
# 4th table has what I want
airPollution <- SetOfTables[[4]]
airPollution <- airPollution %>%
  rename(
    rank = Rank,
    country = `Country/Region`,
    "Y2018" = `2018`,
    "Y2019" = `2019`,
    "Y2020" = `2020`,
    "Y2021" = `2021`,
    "Y2022" = `2022`,
    "Y2023" = `2023`)

# Note this needed because 2023 is a number and I don't want to risk going backwards
airPollution$Y2018 <- parse_number(airPollution$Y2018)
airPollution$Y2019 <- parse_number(airPollution$Y2019)
airPollution$Y2020 <- parse_number(airPollution$Y2020)
```

```r
airPollution$Y2021 <- parse_number(airPollution$Y2021)
airPollution$Y2022 <- parse_number(airPollution$Y2022)

airPollution <- airPollution %>%
  pivot_longer(
    cols = starts_with("Y"),
    names_to = "Year",
    values_to = "particulateMatter")
airPollution$Year <- parse_number(airPollution$Year)

# Uncomment below line to save csv
#write_csv(airPollution, "airPollution.csv")

# ------------------------------------------------------------------

## Datasets

### Loading all datasets from csv files
# Primary Dataset
primaryDefenseTrade <- read.csv("data/defense/USDefenseExport.csv") %>%
  rename(
    Year = YEAR,
    valueInUSD = VALUE,
    country = COUNTRY,
    item = HS.CODE
  )
primaryDefenseTrade$valueInUSD <- parse_number(primaryDefenseTrade$valueInUSD)
primaryDefenseTrade = primaryDefenseTrade %>%
  drop_na()
primaryDefenseTrade$Year = as.numeric(primaryDefenseTrade$Year)

# Research Q1
yearAppend <- function(datFrame, numYear) {
  datFrame %>% mutate(Year = numYear)
}

happinessScore <- read.csv("data/happiness/WHR_2015.csv") %>%
  yearAppend(2015) %>%
  select(country, happiness_score, Year)

for (i in 2016:2023) {
  happinessScore <- happinessScore %>%
```

```
    rbind(
      yearAppend(
        read.csv(
          paste("data/happiness/WHR_", i, ".csv", sep = "")
        ),
        numYear = i
      ) %>%
        select(country, happiness_score, Year)
    )
}

happinessScore <- happinessScore %>%
  rbind(
    read.csv("data/happiness/WEB_WHR_2024.csv") %>%
      yearAppend(2024) %>%
      rename(happiness_score = HappiestCountriesWorldHappinessReportScore2024) %>%
      select(country, happiness_score, Year)
  ) %>%
  drop_na()

# Research Q2
gdpPerCountry <- read.csv("data/GDPUSD/GDPinUSD.csv")

gdpPerCountry$X2024 <- NULL # This one and the below col are just NA
gdpPerCountry$X <- NULL
gdpPerCountry$Indicator.Code <- NULL # Below are the same vals or unneeded
gdpPerCountry$Indicator.Name <- NULL
gdpPerCountry$Country.Code <- NULL

gdpPerCountry <- gdpPerCountry %>%
  pivot_longer(
    cols = starts_with("X"),
    names_to = "Year",
    values_to = "GDP(USD)"
  ) %>%
  rename(country=Country.Name) %>%
  drop_na()
gdpPerCountry$Year <- parse_number(gdpPerCountry$Year)

# Research Q3
populationPerCountry <- read.csv("data/Population/PopulationData.csv")
```

```r
populationPerCountry$X2024 <- NULL # This one and the below col are just NA
populationPerCountry$X <- NULL
populationPerCountry$Indicator.Code <- NULL # Below are the same vals or unneeded
populationPerCountry$Indicator.Name <- NULL
populationPerCountry$Country.Code <- NULL

populationPerCountry <- populationPerCountry %>%
  pivot_longer(
    cols = starts_with("X"),
    names_to = "Year",
    values_to = "Population"
  ) %>%
  rename(country=Country.Name) %>%
  drop_na()
populationPerCountry$Year <- parse_number(populationPerCountry$Year)

# Research Q4
airPollutionPerCountry <- read.csv("data/Airpollution/airPollution.csv") %>%
  select(country,particulateMatter,Year) %>%
  drop_na()

# ------------------------------------------------------------

## Data Summary

# Primary Dataset Provenance and Attributes - US Aerospace Defense Exports
primaryDefenseTrade %>%
  head(3) %>%
  kable()

# Metric Dataset Provenance and Attributes - Country Happiness Score
happinessScore %>%
  head(3) %>%
  kable(caption = "Sample Data Entries")

# Metric Dataset Provenance and Attributes - Country GDP
gdpPerCountry %>%
  head(3) %>%
  kable(caption = "Sample Data Entries")

# Metric Dataset Provenance and Attributes - Country Population
populationPerCountry %>%
```

```
  head(3) %>%
  kable(caption = "Sample Data Entries")

# Metric Dataset Provenance and Attributes - Air Pollution
airPollutionPerCountry %>%
  head(3) %>%
  kable(caption = "Sample Data Entries")

# ------------------------------------------------------------

## Exploratory Data Analysis
# Country vs. Amount they spent importing US Aerospace Defense Technology
primaryDefenseTrade %>%
  group_by(country) %>%
  summarize(TotalSpending = sum(valueInUSD)) %>%
  arrange(desc(TotalSpending)) %>%
  kable(caption = "Table of a Country vs. Amount they spent importing US Aerospace Defense Te

# Aerospace Defense Exported by US each Year 2015-2024
spreadD <- primaryDefenseTrade %>%
  group_by(Year) %>%
  summarize(totalSpent = sum(valueInUSD))

spreadD %>% kable(caption = "Aerospace Defense Exported by US each Year 2015-2024")
ggplot(data = spreadD, aes(x = Year, y = totalSpent)) +
  geom_point() +
  geom_line() +
  labs(title = "Aerospace Defense Exported by US each Year 2015-2024")

# Yearly Distribution of Trade by Country

spreadD <- primaryDefenseTrade %>%
  group_by(Year, country) %>%
  summarize(totalSpent = sum(valueInUSD), .groups = "drop")

ggplot(data = spreadD, aes(x = Year, y = totalSpent)) +
  geom_line() +
  aes(colour = country) +
  theme(legend.position = "right") +
  labs(title = "Aerospace Defense Exported by US each Year 2015-2024 by Country")

# Research Q1
```

```r
primaryDefenseTrade <- primaryDefenseTrade %>%
  group_by(country, Year) %>%
  summarize(TotalValue = sum(valueInUSD), .groups = "drop")

rq1 <- primaryDefenseTrade %>% left_join(happinessScore, by = join_by(Year, country))

ggplot(data = rq1, aes(x = happiness_score, y = TotalValue)) +
  geom_line() +
  aes(colour = country) +
  facet_wrap(~country, ncol = 4) +
  theme(legend.position = "right") +
  labs(title = "Happiness vs. Value Exported from US Aerospace Defense per Country")

# Research Q2
rq2 <- primaryDefenseTrade %>%
  inner_join(gdpPerCountry, by = join_by(country, Year)) %>%
  mutate(ProportionSpent = TotalValue / `GDP(USD)`)

ggplot(data = rq2, aes(x = Year, y = ProportionSpent)) +
  geom_line() +
  geom_point() +
  aes(colour = country) +
  theme(legend.position = "right") +
  labs(title = "Proportion of GDP spent on importing US Aerospace Defense by Country")

# Research Q3
rq3 <- primaryDefenseTrade %>%
  inner_join(populationPerCountry, by = join_by(country, Year))

ggplot(data = rq3, aes(x = Year, y = Population)) +
  geom_point() +
  aes(colour = country) +
  aes(size = TotalValue) +
  theme(legend.position = "right") +
  labs(title = "Population Across the Years for Countries and Total Import of US Aerospace D

# Research Q4
rq4 <- primaryDefenseTrade %>%
  inner_join(airPollutionPerCountry, by = join_by(country, Year))

ggplot(data = rq4, aes(x = particulateMatter, y = TotalValue)) +
  geom_line() +
```

```
  aes(colour = country) +
  theme(legend.position = "right") +
  labs(title = "Particulate Matter vs. Total Value Spent on US Aerospace Defense Import")
```