Hypothesis: We hypothesize that imposing US tariffs on Chinese goods in 2019 reduced demand for Chinese products. Due to this reduced demand, there was an increase in demand for European Union exports of the same products to the United States.

Theory: Our Theory is that due to tariffs put in place in 2019 against Chinese exports to the United States, European Union exports to the United States --of the same products-- will increase due to the increase in price for Chinese Exports from tariffs.

The data which included trade between China and the U.S in this report was given to us by Professor Buzzard and data between the EU and U.S. was supplied by Professor Khan. The data used consists of two Excel datasets. One included trade item category, trade volume, money, year, and other variables for trade between the United States and China. The other included very specific products traded, the trade volume in both Euros and in product scale, the period of transactions, and if the item was an import or an export between all European Union countries and the United States.

For the purpose of getting visualizations, I had to use the exact variable or products when setting rules for my code. The steps I used to create the histograms for all 8 variables are the same for all graphs. Only small changes such as changing the traded item's name are necessary. The steps in order are as follows:

- Open R studio
- Create a new project and R script.
- Open Excel files being used and save them to a cloud folder.
- Set the working directory to open Excel files using the set working directory function located under session on the menu bar. Pick the cloud folder where the excel files are stored.
- Insert the name of the Excel you want to use into the function read.csv('') and name the function df <- read.csv('excel file.csv')

- Install ggplot2, dplyr, tidyverse using the install.packages(“”) function

- Filter down the data frames used.

    - First, choose which dataset you want to use. This action can be referenced on line 25-26.
    - This is done using the filter() command in lines 12-30, then in lined 105,120, and 135.
    - I call the first new dataset “filtererd_df”, then filtered_df2, etc.
- Then, we must select the time range using our filtered funtcions
    - From the filtered data frame created, the data already stored as a data frame, use the select(x:y) function
    - This will filter years down to only those that are in the function. Reference lines 35-39.
- Continue the previous steps for all 4 variables used in U.S to China dataset (capital goods, food, textiles, chemicals)
- Take the “x” out of the year names by sub setting the vector and make data acceptable for plotting
        - long_df2 <- pivot_longer(time_range_df2,
        - cols = X2017:X2021,
        - names_to = "Year",
        - values_to = "Value")
    - long_df2$Year <- sub("x", "", long_df$Year) ## this will only show the number part of the year
- Now, create the graphs using ggplot
- This will require you to use the filtered data created previously.
- The code for creating this can be referenced in lines 72-78.

- You must use the ggplot() function and insert variables within the parentheses.
- This includes the filtered data contains years and relevant topic information.
- You also must use aes() so that the graph will correctly plot and connect and certain points.
- Once that's complete, use the labs() function to add a main title, as well as an x and y-axis title.
- Set the theme to minimal using the theme_minimal() function to take away any unneeded lines in the background.
- Complete these steps for all 8 variables used between the U.S and EU data and the U.S. and China data.
- For each of the different variables or types of exports used, I readjusted the code.
- This required me to create new data frames for each of the new products used by substituting the names of each item in PERIOD_LAB quotations. I left everything but the title constant to maintain consistency.
- This includes keeping the time range the same, partner, and reporter lab all the same.

Caveats:

We discovered a significant caveat: COVID-19. In all three product bases for EU-US exports, we saw a decrease in aggregate US dollars for 2020 compared to 2019; however, the following year, 2021, the aggregate US dollar amount was more extensive than that of 2019 and followed the sharp trajectory we had hypothesized. In 2020, COVID-19 spread throughout the world, taking everybody by storm. Economies were shut down, people lost their jobs, companies collapsed, and lapses in supply and demand led to a lower aggregate amount of money people spent on said product groups.

Problems Aggregating Code

During the code section of this project, our group opted to use R Studio because of its simplicity to create the visualizations we needed. For the most part, we were successful with this approach, however, we were unable to aggregate multiple items into a new, filtered dataset, which forced us to use individual products and create more graphs to display our findings. The reason aggregation was not possible was because the data in the Excel file changed formats once it was imported into R. Specifically, there were spaces that shouldn't exist, which made putting some variables in quotations nearly impossible. Trying to aggregate the data, even by creating a vector containing multiple variables, was unsuccessful because not all the data that was needed to be used was formatted the same way. This left combined data incompatible with R. It's possible that Stata would have been a better option because it's not as picky when it comes to commas or spaces.