Github:https://github.com/kylebradford/CSC442-Project-Econ-vs-Stock-Market

Credits and Context:
- Dataset 1 : Stock market data found on Stooq: https://stooq.com/db/h/

This data was collected to track the trends of various global stock markets around the world. Specifically the U.S., U.K., Japan, Hong Kong, Poland, and Hungary. It includes the Open and Close values for the day as well as volume traded and the ticker that indicates what stock was being used as a representative of the market's performance.

- Dataset 2 : Global economy indicators found on Kaggle:

https://www.kaggle.com/datasets/prasad22/global-economy-indicators

This data was collected to track the change in the economic indicators of over 200 different countries. It was collected through the Economic Statistics Branch of the United Nations Statistics Division, and the national statistical services of the countries included in the dataset. It has data spanning back as far as 1970 up to 2021.

Cleaning and Wrangling:

File Aggregation: Our script iterates through all .txt files in the directory and merges them into a single CSV for easier manipulation. Column names were stripped of leading/trailing spaces to avoid inconsistencies. The Country column values were also stripped to ensure proper matching. Country Filtering: Our economic data contained almost every country whereas our stock data contained less than 10 countries so we filtered our economic data down to only the countries where we had matching stock data. (Germany, Hong Kong, Hungary, Japan, Poland, UK, US). We then removed columns that were not useful to our data analysis including mining and agriculture data.

Data Merging:

To merge the data we had to account for the economic indicators being yearly values and the stock market being daily. We considered two approaches, one where we used quarterly reports of the stock market data, and another where we take average statistics by week. We settled on the former method because our economic data only provides statistics by the year. We wanted to include the nuance of the cyclical nature of the stock market by using quarters without bloating our dataset with weekly reports that would not be indicative of yearly economic indicator movement. By consolidating the stock data into just four records per year, we are able to reduce noise and keep our dataset at a manageable size. We then aligned the daily stock data with the annual economic indicators, ensuring each record had the correct year. This final structure enables a clearer comparison between stock performance and broader economic trends, improving the insights we can draw from the data.

Team Credits:

Kyle Bradford - Cleaning, wrangling, merging, github and file organization

William Mungas - Planning and developing data merge, and coordinating meetings

Logan Blake - Data writeup/description and team planning

Issac Adams - Dataset cleaning/merging