Data Wrangling Project

The Impact of War Devastation and Economic Factors on International Legal Immigration (1990-2020)

Abstract **-** War and economic instability have historically influenced migration patterns worldwide. This study explores and evaluates possible connections between numbers war worldwide, economic factors and immigration between 1990 and 2020.Our analysis integrates data on the number of wars per country, extreme poverty rates, GDP per capita, and migrant stock to determine key drivers of migration. We employ extensive data wrangling, statistical analysis, and visualizations to examine correlations and trends. While our findings attempt to demonstrate that both war and economic conditions significantly impact migration, this study represents an initial stage of a broader investigation.

Introduction **-** Immigration has always shaped human history since the early days [1]. While in the early ages of our species, pre-Neolithic hunter-gatherer humans needed to constantly move for their sustenance [2], often following migratory patterns of other animals or the dynamic structure of resources in their environment [3], as societies settled, shifting from foraging to farming [4], migrations persisted throughout history [5]. At the root of the first modern human migrations were factors such as resource scarcity [6], population pressure [7], and climate change [8]. What drives our curiosity, then, is to see how immigration, one of the most significant socio-political phenomena of the 21st century, is evolving in our contemporary time, evaluating factors like war and economy in order to answer our research questions.

Our ***Research Question*** is how did war devastation influence international legal immigration compared to economic factors between 1990 and 2020? To be able to answer this precisely we have two sub-research questions: How did war devastation influence international legal immigration between 1990 and 2020? How did economic factors influence international legal immigration between 1990 and 2020? To be able to answer these questions we have to precisely define what we mean by these terminologies. **War Devastation** is measured by the number of wars and extreme poverty rate. This definition is based on the logical premise that when a country is engaged in war, a significant portion of its population is directly or indirectly involved in the conflict, reducing workforce participation in economic activities. Additionally, war leads to widespread destruction of infrastructure, further exacerbating economic instability and poverty. **Economic Factors** is measured by GDP per capita and extreme poverty rate. The idea is that economic stability is a fundamental determinant of migration. A high GDP per capita generally indicates better opportunities, while extreme poverty highlights economic distress. By considering both GDP per capita and extreme poverty rates, we aim to capture the dual nature of economic conditions—richness as a pull factor and poverty as a push factor—influencing migration trends.

Data Source **-** For this research, we consulted different datasets from platforms like World Population Review, Our World in Data(Our World in Data), and the Uppsala Conflict Data Program (UCDP - Uppsala Conflict Data Program). These platforms provide extensive, well-documented datasets that allow for large-scale statistical analysis and are widely recognized in academic and policy research. Our World in Data offered the most comprehensive and well-structured dataset, making it an ideal choice for our study due to its meticulous data curation and transparency. The UCDP, maintained by Uppsala University, is an authoritative source for conflict data and further complements our research by providing detailed

records essential for assessing war-related migration trends. These sources provide high-quality data with global coverage, ensuring a reliable foundation for analysis.

We utilize four datasets; **War Data**: Number of ongoing wars per country per year, sourced from the Uppsala Conflict Data Program (UCDP) and the Global Conflict Tracker to ensure continuity. This dataset compiles conflict records to provide insights into war dynamics from 1946 to 2023. **Extreme Poverty Data**: The percentage of the population living below $2.15 per country, obtained from the Our World in Data This dataset ranges from 1963 to 2023. **GDP Data**: GDP per capita per country per year, retrieved from the Maddison Project included in the Our World in Data website. This dataset helps analyze economic conditions from 1990 to 2022. **Migration Data**: The total number of international migrants per country per year, sourced from the Our World in Data. This dataset captures migration trends globally, from 1990 to 2020.

Data Wrangling Methods - The selected datasets were generally well-structured, but some manipulation was necessary to ensure consistency and usability. Before proceeding with the data wrangling steps, we initially visualized the datasets to identify any major challenges such as missing values, inconsistencies, and potential biases. This early exploration allowed us to assess the overall quality of the data and make informed decisions about the best-fit datasets for our research. By evaluating different sources, we ensured that we selected datasets with reliable coverage, minimizing gaps and ensuring the feasibility of our analysis. Eventually, for each of our datasets, we applied different techniques tailored to their specific characteristics. These preparatory steps were crucial in structuring the datasets to facilitate meaningful statistical analysis and visualization.

War Dataset Processing - To get a clear understanding of war trends, we started by filtering the UCDP dataset for the years **1990-2020** the period relevant to our research question.One of the challenges we faced was determining how to represent wars across multiple years. Since wars don't necessarily start and end in the same year, we considered them as ongoing from their recorded start year to end year to ensure a more accurate representation of prolonged conflicts.To maintain consistency and facilitate merging with other datasets, we specifically selected the 'year' column from the UCDP API for analysis. Additionally, we cleaned the data by ensuring all year values were integers and removed any missing or inconsistent entries before conducting further statistical analysis and visualization.

Extreme Poverty Data Processing - Before diving into the analysis, we first visualized the dataset to get a sense of its structure and detect any missing or inconsistent data. This allowed us to spot key challenges, such as duplicate country entries and data gaps.To ensure consistency with our research period, we filtered the dataset for the years 1990-2020. This helped maintain alignment with the other datasets used in the study.Given that some countries had missing values in certain years, we addressed this by applying a rolling average interpolation technique. This method helped smooth out inconsistencies while preserving overall trends.We also grouped and standardized regional classifications (e.g., consolidating sub-regions within Africa and Asia) to allow for meaningful regional comparisons. This restructuring step made the dataset more robust for statistical analysis and visualization.

GDP Data Processing - To ensure that our economic data aligns with the rest of the analysis, we filtered the dataset to cover the years 1990-2020. This allows for consistent cross-dataset comparisons. One of the major challenges in working with GDP data is dealing with missing values, as some countries do not report their economic data consistently. To tackle this, we used the k-nearest neighbor (KNN) imputation algorithm, which estimates missing values based on the preceding and succeeding economic data patterns from earlier and later years. This method provides a more informed and accurate estimate than simple mean imputation, preserving economic trends effectively. After filling in missing values, we

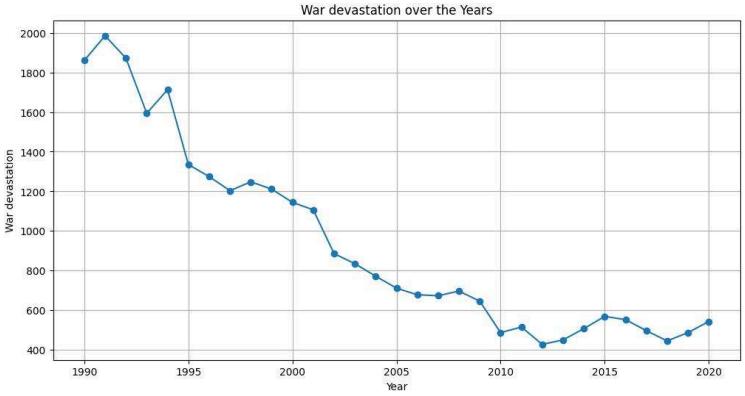
structured the dataset by pivoting it to focus on GDP per capita per year, ensuring that our analysis could effectively capture global economic trends. This preprocessing step made it easier to compute yearly averages and identify meaningful trends over time.

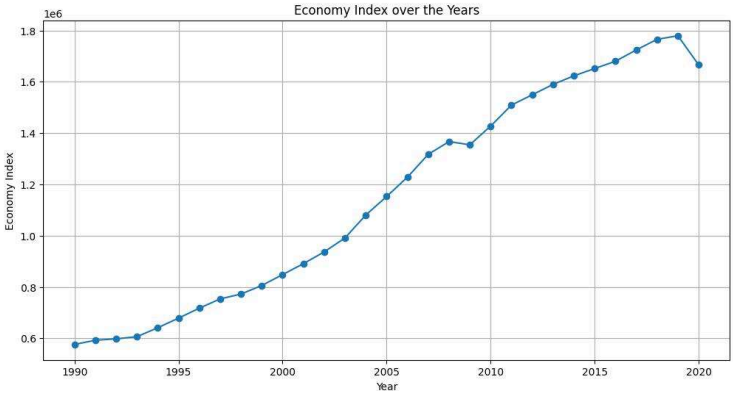
Migration Data Processing - To align with our research period, we filtered the dataset to include only the years 1990-2020, ensuring comparability with other datasets. We then grouped the data by year and aggregated migration numbers to observe global trends over time. This step provided a clearer picture of migration patterns without the noise of country-specific variations. Finally, we standardized the values to ensure consistency across different datasets, making it easier to perform meaningful statistical comparisons. This preparation was essential for uncovering key insights into global migration trends and their relationship with economic and war-related factors.

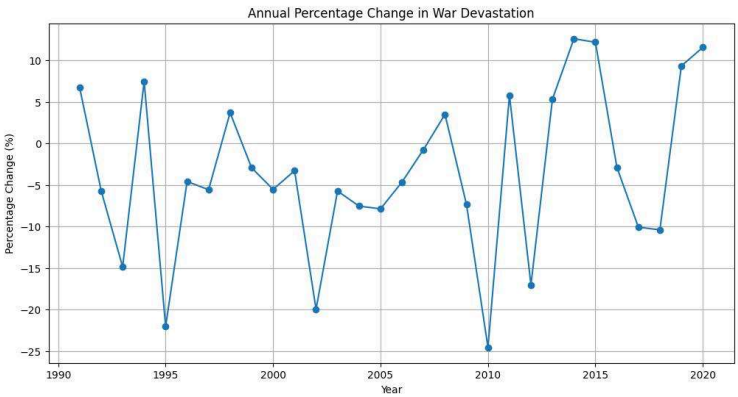
The variables concerning war devastation and economic index, which were previously discussed, were calculated through the merger of tables based on annual data. For the calculation of war devastation, the product of poverty and wars per year was determined. In contrast, for the economic index, the percentage of individuals not living in poverty was derived by subtracting the poverty percentage from one hundred, which was then multiplied by the GDP to compute the economic index. Subsequently, owing to the lack of standardization in the data (normalization is the more appropriate term, as it refers to adjusting data to have common scales), the conversion to percentages was achieved using the "pct_change()" function from pandas, and the resultant figures were multiplied by 100 to transform them from decimal form to percentages. Additionally, the change in the economic index and war devastation from 1990 to 2020 was calculated as a percentage to ascertain the presence of any discernible trends. This analysis was aimed at evaluating the variations over the specified period, thereby facilitating a clearer understanding of the temporal dynamics affecting these variables.

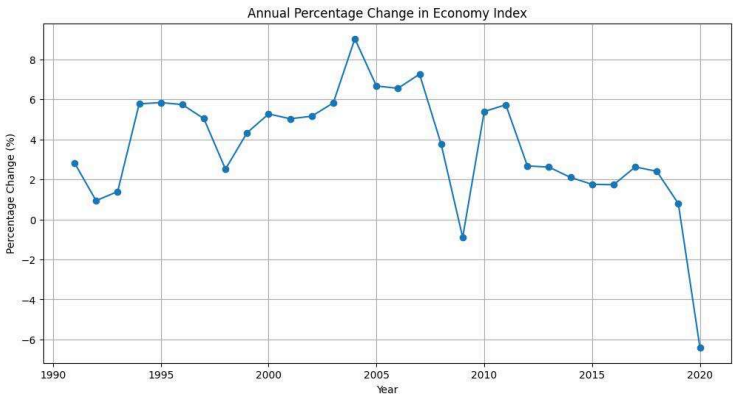
For the purpose of statistical analysis, the dataset was filtered to reflect five-year intervals. The variable of immigration was then subjected to analysis using the Pearson Correlation Coefficient, employing the "pearsonr(x, y)" function from the "scipy.stats" library.

Findings









The analysis involving the Pearson correlation coefficient provided significant insights into the relationships between different variables. Specifically, the correlation between war devastation and the number of legal international immigrants per year yielded a correlation coefficient of -0.8013163492661401 with a p-value of 0.030268861919301934. This p-value suggests that the correlation is not statistically significant. However, this negative correlation suggests a strong inverse relationship; as war devastation increases, the number of legal international immigrants tends to decrease.

Conversely, the economic index's correlation with the number of legal international immigrants per year was determined to be 0.9536600509457536, with a highly significant p-value of 0.0008659149869167201. This is not a statistically significant p value. However, this positive correlation indicates a very strong relationship where improvements in the economic index, which suggest better economic conditions, are associated with an increase in the number of legal international immigrants. This pattern is expected as better economic conditions are required for legal international immigration.

These findings are crucial for understanding migration trends and the factors influencing them. Policymakers and researchers might use this data to forecast immigration trends or to plan economic and social strategies to manage the impacts of war and economic changes on migration.

References

[1] The Migration History of Humans: DNA Study Traces Human Origins Across the Continents | Scientific American

[2]Hunter–gatherer migrations, mobility and social relations: A case study from the Early Bronze Age Baikal region, Siberia - ScienceDirect

[3] Hunter-Gatherer Mobility Strategies | Journal of Anthropological Research: Vol 39, No 3

[4] From Foraging To Farming: Explaining The Neolithic Revolution - Weisdorf - 2005 - Journal of Economic Surveys - Wiley Online Library

[5] Patterns of Neolithic migration | BEAN Project | Results in brief | FP7 | CORDIS | European Commission

[6] Early Human Migration - World History Encyclopedia

[7] The Ancient Migration: Unveiling the Global Journey of Early Humans – A Brief History of the World To 1500

[8] Early human migrations - Wikipedia