This project started with the discovery of a dataset on Kaggle that showcased the world's population in each country since 1970. There are datapoints in the set every 10 years from 1970 to 2010, then additional datapoints at 2015, 2020 and 2022.Our initial exploration came from a desire to see if we could find any corresponding socioeconomic data that might show correlation with growth rate. After some brainstorming and online sleuthing, we found three datasets we thought could provide a specific hypothesis.

1. Safety
2. Education
3. Poverty

**Base Dataset**  
But before we got to analyzing the influential factors seeking relationships in the data, we needed to make sure we were working with cleaned data. In looking at the different sets we discovered, we realized not every producer of the data used the same syntax or spelling for countries. Everything from "United States" versus "USA" to "Bahamas" versus "The Bahamas" had potential to ruin our data later down the line when we attempted to merge and analyze it. By using GeoApify to search and return identical syntax for country names on each dataset, we knew we could then confidently merge the data later in the process. This clean and cull, as we called it, also removed data reported in the sets that didn't match what we were looking for. This included some territories or regions that were considered countries by some datasets, but not others, but also included some datasets that had regional or continental rows of reporting.

**TALK ABOUT VISUALIZATIONS AND FINDINGS FOR BASE DATA HERE**

**Safety**  
We found a Global Peace Index (GPI) produced online by the Institute for Economics and Peace and hosted online by [VisionOfHumanity.org](http://visionofhumanity.org/). The GPI derives a score for each country based on the level of societal safety and security, the extent of ongoing conflict and the degree of militarisation. The dataset available had the safety index tracked back to 2010 for most countries. This has alignment with our base dataset for the years 2010 to 2022 to allow for comparison.

**TALK ABOUT VISUALIZATIONS AND FINDINGS FOR SAFETY HERE**

**Education**  
To quantify education, we found an initial dataset created by Robert Barro from Harvard and Jong Wha Lee from Korea University that has been dubbed the Barro-Lee dataset online and is considered to be the most comprehensive measure of global education. The initial Barro-Lee study was conducted with data ending in 2010, but we found a GitHub on which Barro continued collaboration to push the data to 2015.This dataset tracks a number of datapoints, so we boiled it down to some key measures that could reflect overall education in a country: average years of education attained and per cent of population with no education. Datapoints dated back to 1950 for many countries, but our base dataset only tracked population to 1970, so we settled on a window of 1970 to 2015 to track the changes.

**TALK ABOUT VISUALIZATIONS AND FINDINGS FOR EDUCATION HERE**

**Poverty**  
To look at poverty, we found the repository for the World Bank's Poverty and Inequality Platform database. This data had thousands of rows measuring different types of poverty across most nations and a mix of years, as well as a mix of measurements. Despite being the largest dataset of the three, it yielded the fewest comparable results by the time it was narrowed down to matching variables and years that we could compare. Using several coding techniques to analyze the data, we found the largest possible dataset to track change based on the type of poverty reporting and version of calculations, while also making sure there were reportable years that matched our base dataset. The result was a dataset of 27 countries that reported poverty the same way between 2000 and 2015, which both are years in our base dataset.

**TALK ABOUT VISUALIZATIONS AND FINDINGS FOR POVERTY HERE**