JUAN (education)

Intro   
This project focuses on analyzing the World Population Dataset to identify notable changes in population growth rates across countries. By investigating the socioeconomic factors that contribute to these shifts, we aim to gain valuable insights into the complex interplay between population trends and various indicators of development.”

Base dataset  
This project utilizes a dataset sourced from Kaggle, which presents comprehensive information on the world's population in each country since 1970. The data originates from United Nations population estimates and projections. With data points available at 10-year intervals from 1970 to 2010, as well as additional points for 2015, 2020, and 2022, we analyze population growth rates and identify significant shifts over time. By leveraging this dataset, we gain a deeper understanding of the complex interplay between population dynamics and socioeconomic factors, contributing to sustainable development strategies.

Education 1: We used the Barro-Lee dataset, an extensive global education measure, to quantify education levels. The dataset, curated by Robert Barro of Harvard and Jong Wha Lee of Korea University, extends to 2015, having been initially capped in 2010.From this rich dataset, we focused on two key indicators: the average years of education attained and the percentage of the population without education. Our analysis period is 1970 to 2015, based on the availability of our population data. We then set up our study hypotheses: Null Hypothesis: An increase in a country's average education does not significantly affect population growth. Alternative Hypothesis: An increase in a country's average education significantly affects population growth.

Education 2: Following our study design, we plotted the change in average schooling against the growth rate to visually represent the data spread. This scatter plot is an effective tool to get an immediate understanding of the data trends. Our analysis further quantifies the relationship through the correlation coefficient, which stands at 0.11. This value indicates a weak positive relationship between a country's average schooling and its population growth rate. Moreover, the P-value of 0.20548 surpasses the typical significance level of 0.05. This means the observed data is quite likely under the null hypothesis. Thus, we reach the conclusion given the existing data, we cannot reject the null hypothesis. In other words, the evidence does not support a significant impact of average schooling on the growth rate of a country's population.

KALYN (poverty)

Intro

Talk about first question which countries show significant shift in population. 2nd question which socioeconomic factors effect this change  
Settling on the three factors Poverty, education, and safety and why we landed on those three  
Speak about remote work and not much data as of yet because only became really prevalent a couple years a go  
Finding datasets associated with the factors.

Slide 1  
Worked with the poverty dataset from OWID repository on GitHub  
initially more than 4k rows of data. Brought down to about 30 rows of relevant data. Much smaller sample size but still had a nice variety of countries  
Looked specifically at the income below which half the population lives or the "median income" and if there was any correlation to population growth  
Cleaned down dataset welfare type to just income, reporting level to just national, and PIP version for standardized rate of poverty line in 2017.

Slide 2  
We can see through the scatter plot that most of the points were clustered in the centre  
the correlation coefficient shows not a very high correlation  
And the p value being way above the accepted significance value of 0.05 we would fail to reject the null hypothesis.

**SEAN (base)**

CLEANING SLIDE

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.

Through our exploring, we realized different sources used different syntax, spelling and version for the names of 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.

We realized that 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, such as some territories or regions that were considered countries by some datasets, but not others.

The choropleth maps we wanted to use to visualize the data also did their mapping through what's known as an iso code, which is a unique two- or three-letter identifier for each country.

Some datasets had the two-letter, other the three and some of them had none.

We used a fifth dataset that was a listing of all countries and iso codes to further refine the datasets, as GeoApify also returned the two-letter iso code.

This was an exercise in patience with a couple of the larger datasets, and I'm hopeful we don't have another GeoApify-related project this month, as my API requests to the website eclipsed 23,000 during the process.

GROWTH RATE

We also learned something new about growth rate. In realizing the growth rate listed on the base dataset wasn't matching what we assumed was the proper calculation, we found there are different growth rate formulas for different data.

The usual "final data minus starting data divided by starting data" isn't used as commonly with demographics. There is an average growth rate over time formula that we found online that better expresses population growth and was used in the base dataset.

This formula takes into account how many years you are calculating, so the final percentage in our analysis is better expressed as "X country had a population growth rate of Y.2% over Z years."

**SUSHMA (safety)**

Slide 1

* Hello everyone, I am Sushma from Team 9.
* I would be briefing about Safety dataset.
* Safety is a critical aspect in various domains.
* To monitor and analyze safety trends, a comprehensive safety dataset is essential.
* 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 militarization.
* The dataset available had the safety index tracked back to 2008 for most countries. This has alignment with our base dataset for the years 2010 to 2022 to allow for comparison.

Slide 2

* After cleaning the dataset, I’ve calculated the Change in GPI versus growth rate:
  + Scatter plot shows clustered datapoints.
  + Correlation coefficient = 0.15
  + P-value = 0.05296

And hence, we can conclude that “We can (softly) reject the null hypothesis.”

* + As you can see, it is not a strong positive correlation.
  + But the p-value is very close to the 0.05 threshold for statistical significance.

*\* Remember, the higher the p-value, the less statistically significant the correlation\**

Slide 3

* The correlation coefficient for each of the three factors was not a strong correlation, but it was positive for all three.
* We initially thought the poverty data set would yield the strongest correlations, but the dataset proved to have too many variables and by the time it was cleaned the sample size was too small to make strong conclusions.
* The education dataset was closer to being statistically significant. Perhaps looking at a different datapoint could show a stronger relationship with growth rates.
* In the end, only safety had a p-value that showed that its correlation within the data to growth rate was statistically significant – even if that correlation was weak.
* If a country’s GPI improves, the rate at which its population grows will increase.

*I hope we covered all the required points.*

*We are happy to answer any questions.*