

TEAM 9

industry:

DEMOGRAPHIC ANALYSIS

team:

KALYN AGUIAR

SEAN ALLEN

JUAN MONCALEANO

SUSHMA MYLAVARAPU

A look at how
socioeconomic factors
influence population growth

SOURAV BANERJEE · UPDATED 8 MONTHS AGO

520 New Notebook Download (16 kB)

World Population Dataset

This Dataset contains Population data of every Country/Territory in the world

Data Card Code (78) Discussion (5)

About Dataset

Context

The current US Census Bureau world population estimate in June 2019 shows that the current global population is 7,577,130,400 people on earth, which far exceeds the world population of 7.2 billion in 2015. Our own estimate based on UN data shows the world's population surpassing 7.7 billion.

China is the most populous country in the world with a population exceeding 1.4 billion. It is one of just two countries with a population of more than 1 billion, with India being the second. As of 2018, India has a population of over 1.355 billion people, and its population growth is expected to continue through at least 2050. By the year 2030, the country of India is expected to become the most populous country in the world. This is because India's population will grow, while China is projected to see a loss in population.

The following 11 countries that are the most populous in the world each have populations exceeding 100 million. These include the United States, Indonesia, Brazil, Pakistan, Nigeria, Bangladesh, Russia, Mexico, Japan, Ethiopia, and the Philippines. Of these nations, all are expected to continue to grow except Russia and Japan, which will see their populations drop by 2030 before falling again significantly by 2050.

Many other nations have populations of at least one million, while there are also countries that have just thousands. The smallest population in the world can be found in Vatican City, where only 801 people reside.

Usability 10.00

License Other (specified in description)

Expected update frequency Annually

Tags Tabular Social Science Beginner Exploratory Data Analysis People and Society

Found on Kaggle.com, this dataset showcases the world population by country since 1970 at varying datapoints.

BASE DATASET

Which countries show significant shifts in population growth rates?
What socioeconomic factors may have influenced these changes?

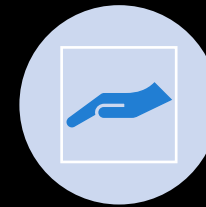
Which factors to test?



POVERTY



EDUCATION

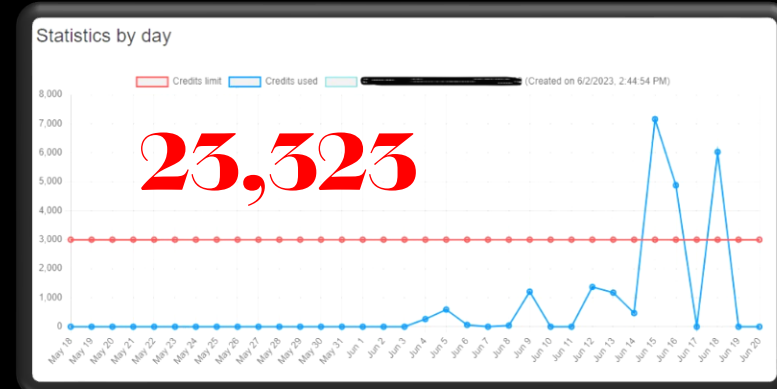


SAFETY

Before we began

- Country iso code is a universal identifier.
- Cleaned data through GeoApify to make sure sets would merge.

```
Indexing: Nauru as nr and confidence: 1
Indexing: Nepal as np and confidence: 1
Indexing: Netherlands as nl and confidence: 1
Indexing: France as fr and confidence: 0
Indexing: New Zealand as nz and confidence: 1
Indexing: Nicaragua as ni and confidence: 1
Indexing: Niger as ne and confidence: 1
Indexing: Nigeria as ng and confidence: 1
Indexing: Niue as nu and confidence: 1
Indexing: North Korea as kp and confidence: 1
Indexing: North Macedonia as mk and confidence: 1
Dropping: {'results': [], 'query': {'text': '', 'parsed': {'country': 'Northern Mariana Islands',
n'}}}}
Indexing: Norway as no and confidence: 1
Indexing: Oman as om and confidence: 1
Indexing: Pakistan as pk and confidence: 1
Indexing: Palau as pw and confidence: 1
Indexing: Panama as pa and confidence: 1
Indexing: Papua New Guinea as pg and confidence: 1
Indexing: Paraguay as py and confidence: 1
Indexing: Peru as pe and confidence: 1
Indexing: Philippines as ph and confidence: 1
Indexing: Poland as pl and confidence: 1
Indexing: Portugal as pt and confidence: 1
Indexing: Romania as ro and confidence: 1
Indexing: Russia as ru and confidence: 1
Indexing: Rwanda as rw and confidence: 1
Indexing: Saudi Arabia as sa and confidence: 1
Indexing: Senegal as sn and confidence: 1
Indexing: Serbia as rs and confidence: 1
Indexing: Singapore as sg and confidence: 1
Indexing: Slovakia as sk and confidence: 1
Indexing: Slovenia as si and confidence: 1
Indexing: South Africa as za and confidence: 1
Indexing: South Korea as kr and confidence: 1
Indexing: Spain as es and confidence: 1
Indexing: Sri Lanka as lk and confidence: 1
Indexing: Sweden as se and confidence: 1
Indexing: Switzerland as ch and confidence: 1
Indexing: Taiwan as tw and confidence: 1
Indexing: Thailand as th and confidence: 1
Indexing: Timor as tl and confidence: 1
Indexing: Togo as tg and confidence: 1
Indexing: Tonga as to and confidence: 1
Indexing: Trinidad and Tobago as tt and confidence: 1
Indexing: Tunisia as tn and confidence: 1
Indexing: Turkey as tr and confidence: 1
Indexing: Uganda as ug and confidence: 1
Indexing: Ukraine as ua and confidence: 1
Indexing: United Kingdom as gb and confidence: 1
Indexing: United States as us and confidence: 1
Indexing: Uruguay as uy and confidence: 1
Indexing: Uzbekistan as uz and confidence: 1
Indexing: Vanuatu as vu and confidence: 1
Indexing: Venezuela as ve and confidence: 1
Indexing: Vietnam as vn and confidence: 1
Indexing: Wales as wa and confidence: 1
Indexing: West Bank as wb and confidence: 1
Indexing: Yemen as ye and confidence: 1
Indexing: Zambia as zw and confidence: 1
Indexing: Zimbabwe as zw and confidence: 1
```



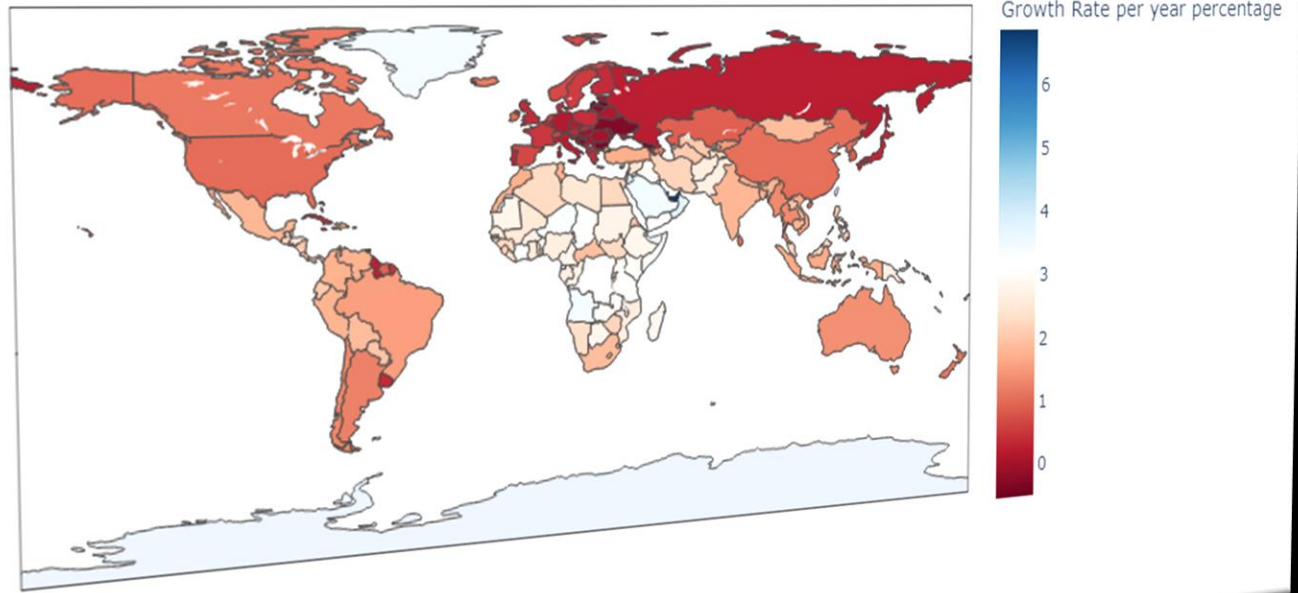
- ~~“U.S.A.”~~
- ~~“USA”~~
- ~~“United States of America”~~

Use API to fetch common syntax and iso code

- us
- United States

Cleaned base dataset

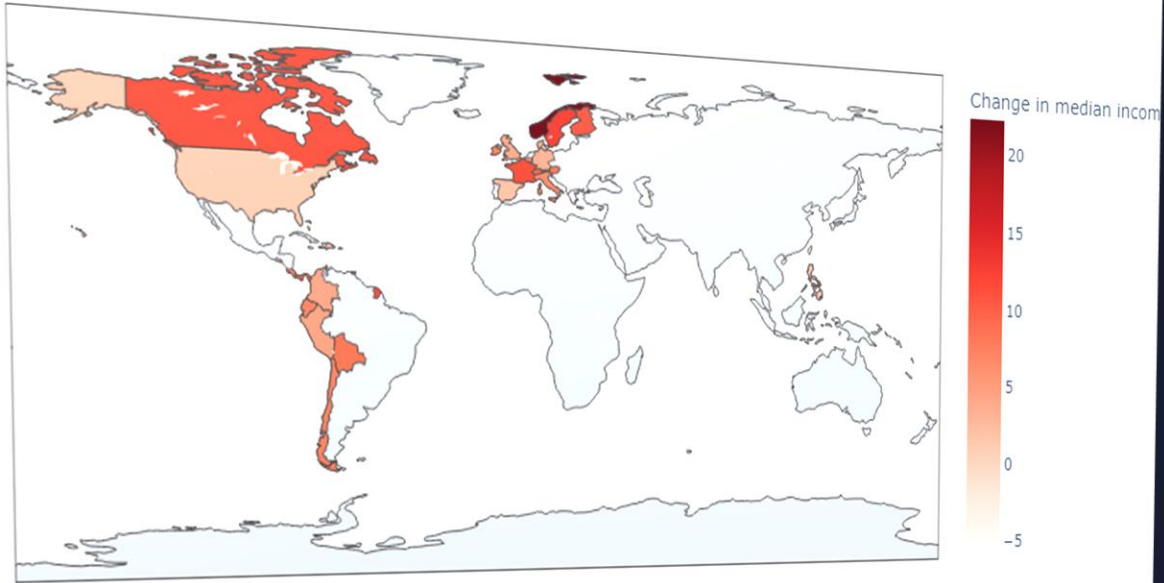
Growth Rate from 1970 to 2022



$$\text{Growth rate} = \left(\frac{\text{new value}}{\text{original value}} \right)^{\frac{1}{N}} - 1$$

$(N = \text{number of years})$

Change in median income from 2000 to 2015



Income below which half the population lives
(median income)

POVERTY

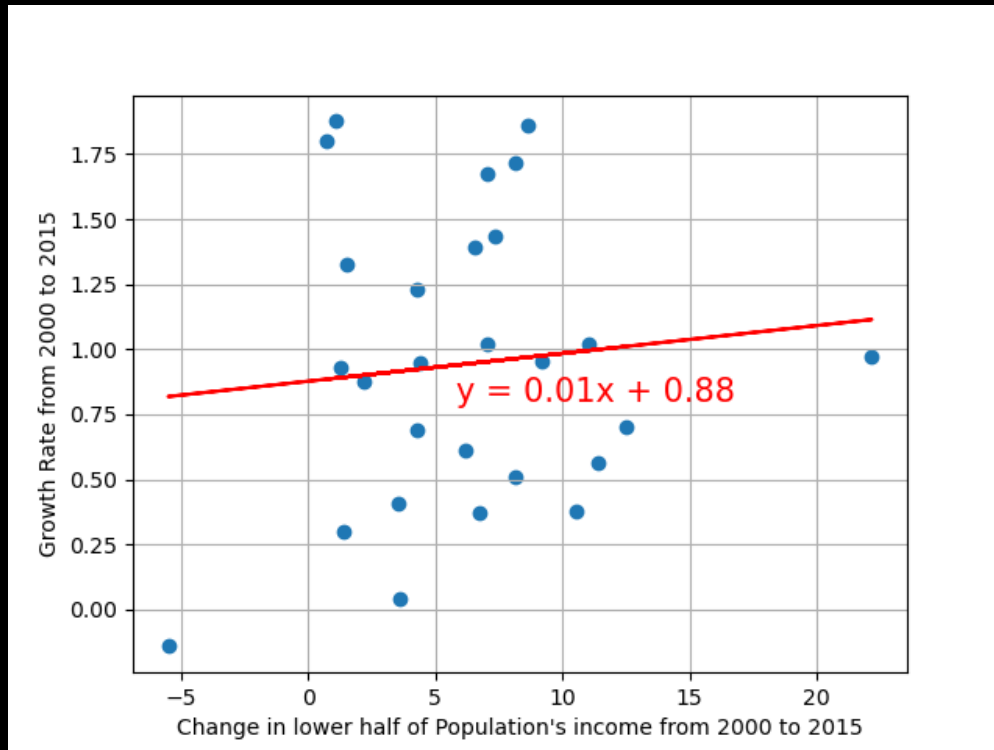
Null Hypothesis:

An improvement in a country's median income over time will have no significant impact on the growth rate of its population.

Alternative Hypothesis:

An improvement in a country's median income over time will have a significant impact on the growth rate of its population.

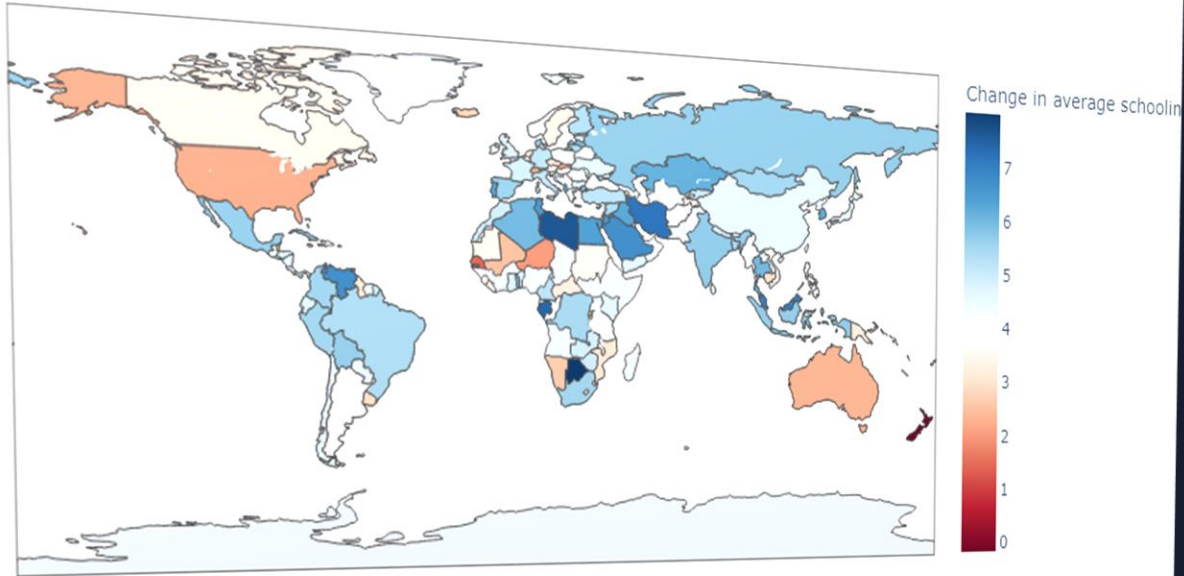
Change in median income versus growth rate



- Scatter plot has small sample size
- Correlation coefficient = 0.1
- P-value = 0.62476

Conclusion: We fail to reject the null hypothesis

Change in average schooling from 1970 to 2015



Barro-Lee average years of schooling attained
(average schooling)

EDUCATION

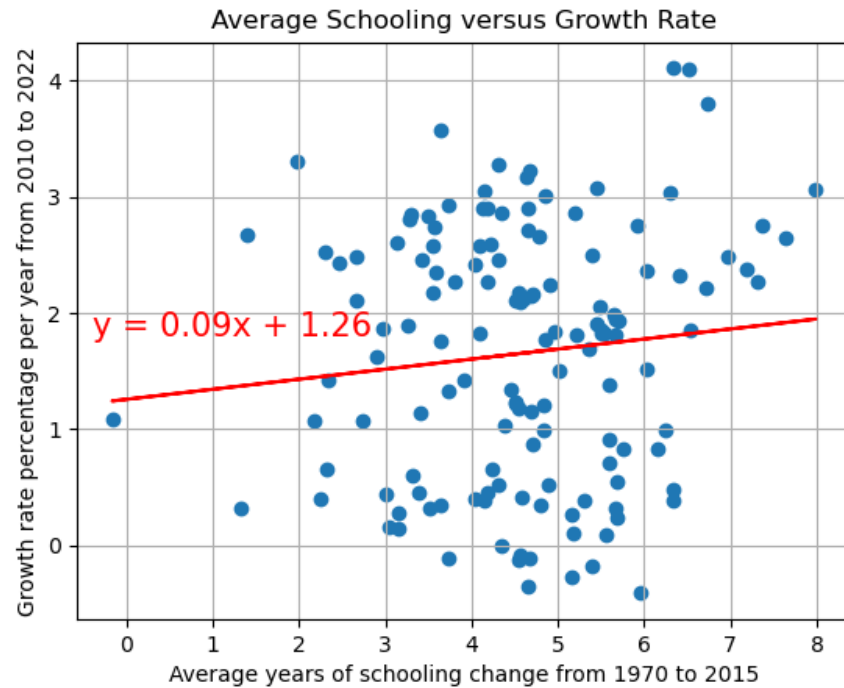
Null Hypothesis:

An improvement in a country's average schooling over time will have no significant impact on the growth rate of its population.

Alternative Hypothesis:

An improvement in a country's average schooling over time will have a significant impact on the growth rate of its population.

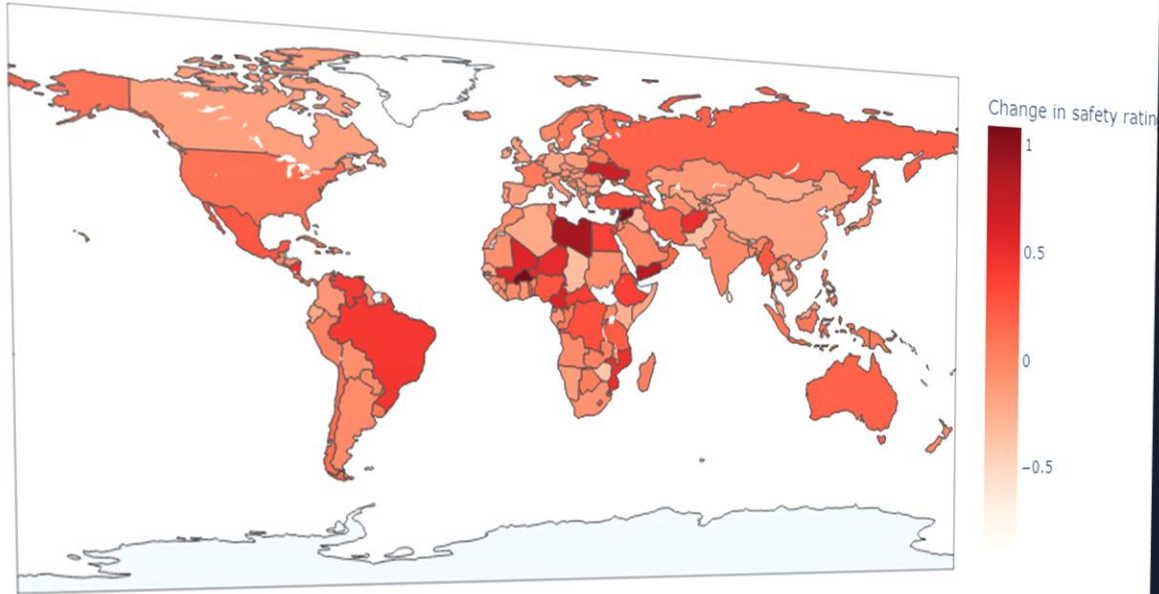
Change in average schooling versus growth rate



- Scatter plot shows spread of data
- Correlation coefficient = 0.11
- P-value = 0.20548

Conclusion: We fail to reject the null hypothesis

Change in GPI from 2010 to 2022



Global Peace Index

SAFETY

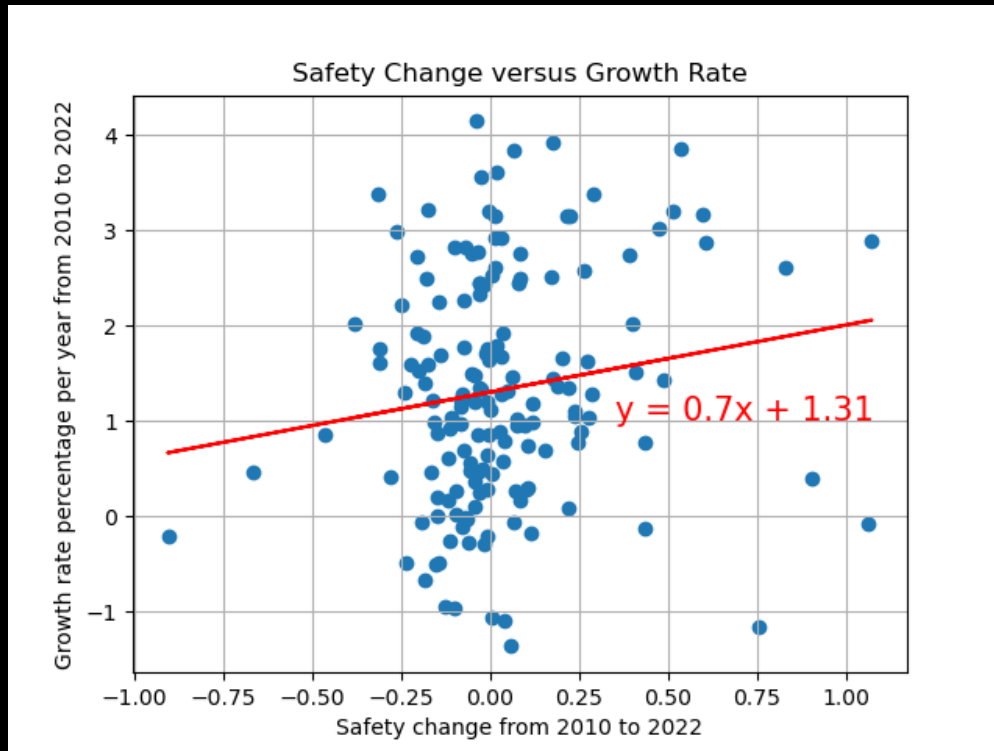
Null Hypothesis:

An improvement in a country's GPI over time will have no significant impact on the growth rate of its population.

Alternative Hypothesis:

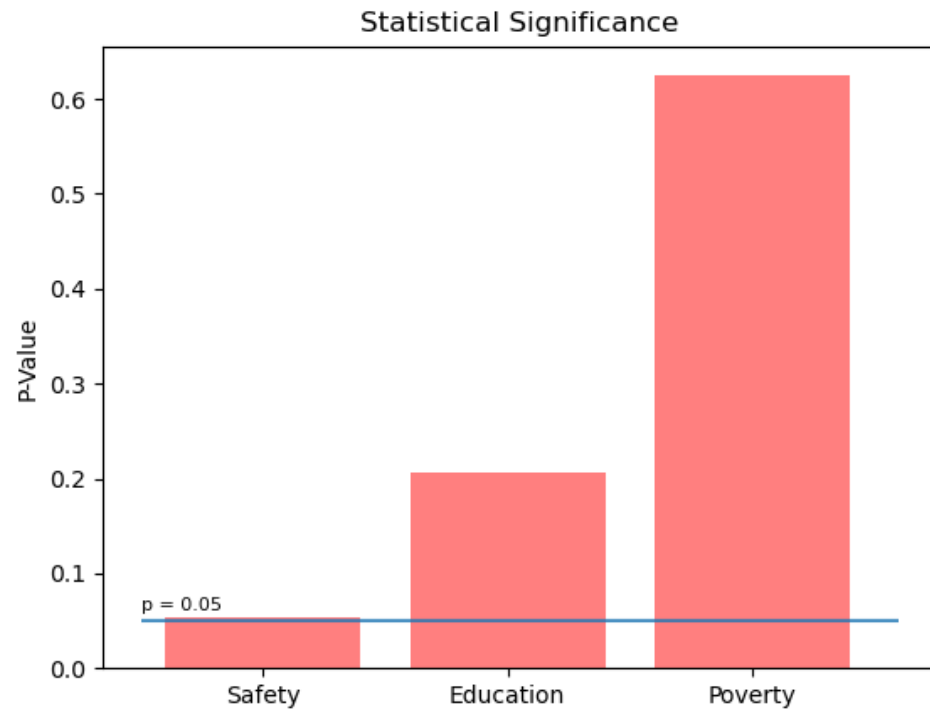
An improvement in a country's GPI over time will have a significant impact on the growth rate of its population.

Change in GPI versus growth rate



- Scatter plot shows clustered datapoints
- Correlation coefficient = 0.15
- P-value = 0.05296

Conclusion: We can (softly) reject the null hypothesis.



** Remember: The higher the p-value, the less statistically significant the correlation.*

Safety showed the most statistically significant correlation to population growth rate.

If a country's GPI improves, the rate at which its population grows will increase.

CONCLUSION