Is there any relationship between GDP and life expectancy?

Mengyang Yi

2024-04-27

Introduction

Understanding the relationship between a country's Gross Domestic Product (GDP) and life expectancy is crucial for policymakers and researchers alike. This project aims to explore the potential links between these two key indicators. While economic development is often associated with improved healthcare and living standards, the nuances of how GDP influences life expectancy require deeper investigation. By analyzing data across various countries and time periods, this study seeks to provide insights into the complex interplay between economic factors and population health outcomes.For this project, I choose 5 countries ranging from 1960 to 2013 to analyze–China, Bangladesh, Mexico, Peru, and Poland.

Let's look at the data at first...

The data

To begin our analysis on the relationship between GDP and life expectancy, the first step is to gather and preprocess the necessary datasets. We will focus on two primary datasets: one containing life expectancy data and another containing GDP data for various countries over a specified time period.

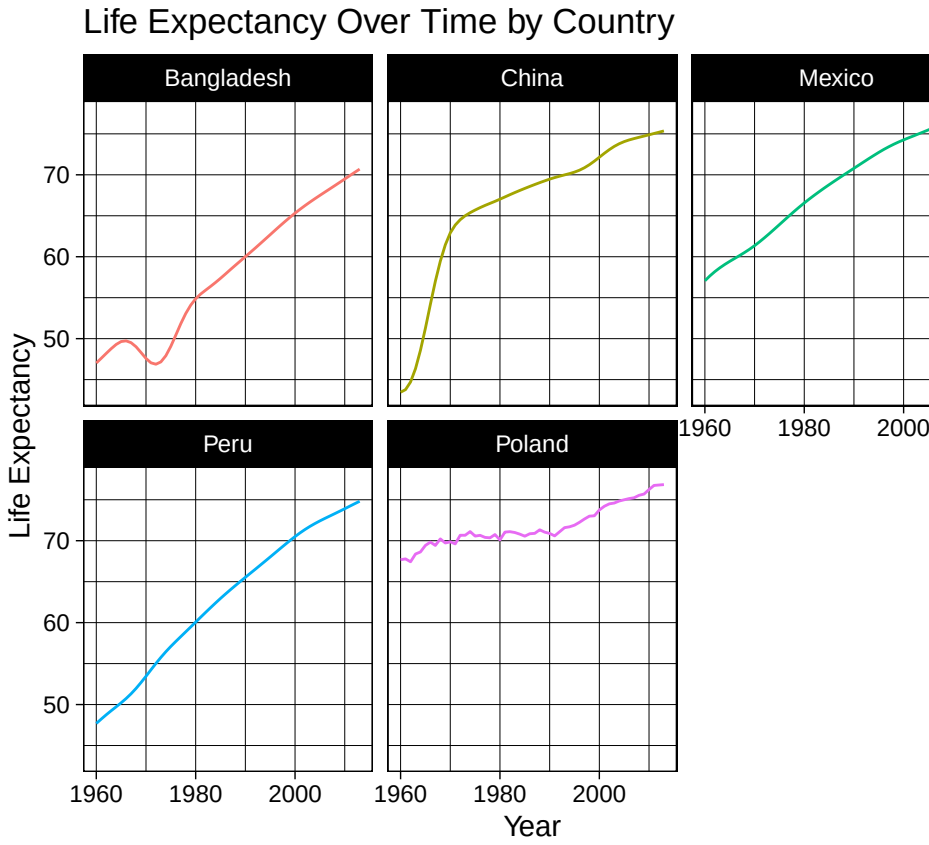
For the Life expectancy dataset, I will use pivot longer to consolidate all the years into a single column. Then I will filter it out and only maintain the information that i want.For the GDP dataset, I will only maintain the mean of the GDP for each selected countries. In the end, I will combine these two datasets and establish a new dataset called "GDP_Life_Expectancy".

```
GDP <- read_csv("/cloud/project/GDP.CSV")
LifeExpectancy <- read_xls("/cloud/project/LifeExpectancy.xls")

LifeExpectancy <- LifeExpectancy |>
  pivot_longer(cols = `1960`:`2013`, names_to = "year", values_to = "Life_Expectancy") |>
  mutate(year = as.integer(year)) |>
  filter(`Country Name` %in% c("China", "Bangladesh", "Mexico", "Peru", "Poland"),
         year >= 1960 & year <= 2013) |>
  rename(country = `Country Name`) |>
  select(country, year, Life_Expectancy)

GDP <- GDP |>
  filter(location_name %in% c("China", "Bangladesh", "Mexico", "Peru", "Poland"),
         year >= 1960 & year <= 2013) |>
  select(location_name, year, gdp_ppp_mean, gdp_usd_mean) |>
  rename(country = location_name)

GDP_Life_Expectancy <- full_join(LifeExpectancy, GDP, by = c("country", "year"))

file_path <- "/cloud/project/GDP_Life_Expectancy.csv"
write.csv(GDP_Life_Expectancy, file = file_path, row.names = FALSE)
```

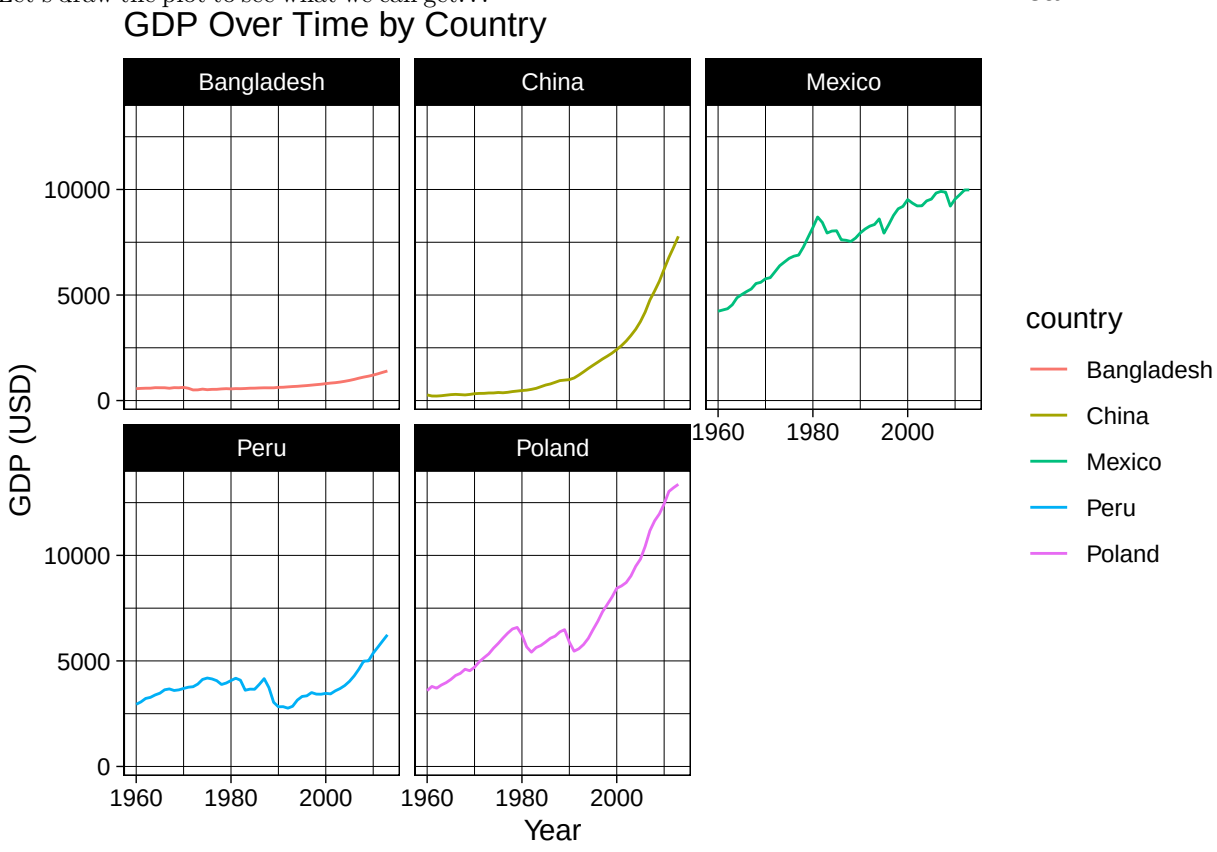
```r
print(GDP_Life_Expectancy)
```

```
## # A tibble: 270 x 5
##    country     year Life_Expectancy gdp_ppp_mean gdp_usd_mean
##    <chr>      <dbl>           <dbl>        <dbl>        <dbl>
##  1 Bangladesh  1960            47.0         1238          563
##  2 Bangladesh  1961            47.6         1239          576
##  3 Bangladesh  1962            48.2         1248          585
##  4 Bangladesh  1963            48.8         1301          586
##  5 Bangladesh  1964            49.3         1286          610
##  6 Bangladesh  1965            49.6         1282          608
##  7 Bangladesh  1966            49.7         1278          606
##  8 Bangladesh  1967            49.5         1223          582
##  9 Bangladesh  1968            49.0         1260          611
## 10 Bangladesh  1969            48.3         1272          607
## # i 260 more rows
```

The result



Let's draw the plot to see what we can get. . .



The graph above illustrates a positive correlation between life expectancy and year for each country. Perhaps this is the result of society changing and individuals being more health-concerned. We also can tell that there is a positive correlation between GDP and year for each country.Therefore, it's reasonable to question if GDP has an effect on life expectancy. Hence, it is crucial to examine the correlation between GDP and life expectancy. This may be accomplished by using the correlation approach to see whether GDP and life expectancy are significantly related. Using this statistical method, we may learn more about the possible relationship between changes in GDP and changes in life expectancy over time. NOTE: 1) A correlation coefficient close to 1 indicates a strong positive linear relationship (as GDP increases, life expectancy tends to increase). 2) A correlation coefficient close to -1 indicates a strong negative linear relationship (as GDP increases, life expectancy tends to decrease). 3) A correlation coefficient near 0 indicates a weak or no linear relationship between the two variables.

```
cor_test <- cor.test(GDP_Life_Expectancy$gdp_usd_mean, GDP_Life_Expectancy$Life_Expectancy, method = "p
print(cor_test)
```

```
##
##  Pearson's product-moment correlation
##
## data:  GDP_Life_Expectancy$gdp_usd_mean and GDP_Life_Expectancy$Life_Expectancy
## t = 12.587, df = 268, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5286081 0.6794584
## sample estimates:
##       cor
## 0.6095213
```

A low p-value (p-value < 2.2e-16) suggests that the correlation coefficient is statistically significant, indicating a reliable linear relationship between GDP and life expectancy.

Let's fit it into linear regression model with GDP (gdp_usd_mean) as the independent variable and life expectancy (Life_Expectancy) as the dependent variable:. . .

```
model <- lm(Life_Expectancy ~ gdp_usd_mean, data = GDP_Life_Expectancy)
summary(model)
```

```
##
## Call:
## lm(formula = Life_Expectancy ~ gdp_usd_mean, data = GDP_Life_Expectancy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.814  -5.015   1.407   5.253   9.760
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.878e+01  6.737e-01   87.26   <2e-16 ***
## gdp_usd_mean 1.605e-03  1.275e-04   12.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.765 on 268 degrees of freedom
## Multiple R-squared:  0.3715, Adjusted R-squared:  0.3692
## F-statistic: 158.4 on 1 and 268 DF,  p-value: < 2.2e-16
```

As all the p-values are less than 0.05 in a linear regression model, it indicates that the associated predictor variable is likely to have a meaningful and statistically significant impact.

So we can get a linear equation as following: Life expectancy = 58.78 + 0.001605 * GDP (USD)

An intriguing issue arises: we thus wonder how long individuals will be able to live in next century...

prediction

Identifying the relationship between year and GDP (USd) and using it as a bridge to forecast life expectancy is one potential approach.

```
model <- lm(gdp_usd_mean ~ year , data = GDP_Life_Expectancy)
summary(model)
```

```
##
## Call:
## lm(formula = gdp_usd_mean ~ year, data = GDP_Life_Expectancy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4913.9 -2720.8    44.6  2575.3  7053.1
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -155706.37   23165.24  -6.722 1.08e-10 ***
## year             80.49      11.66   6.902 3.69e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2986 on 268 degrees of freedom
## Multiple R-squared:  0.1509, Adjusted R-squared:  0.1478
## F-statistic: 47.64 on 1 and 268 DF,  p-value: 3.694e-11
```

By establishing the relationship between GDP and year, we can get a linear equation as : GDP (USD) = -155706.37 + 80.49* year Now, the things become easy, we can plug year from 2024 to 2124 to predict...

```
years <- seq(2024, 2124)
GDP <- -155706.37 + 80.49 * years
life_expectancy <- 58.78 + 0.001605 * GDP
data <- data.frame(year = years, GDP = GDP, life_expectancy = life_expectancy)
ggplot <-
  ggplot(data,aes(x = year, y = life_expectancy)) +
  geom_line() +
  labs(x = "Years", y = "Life Expectancy")
ggplotly(tooltip = c("year", "life_expectancy"))
```

```
## QStandardPaths: XDG_RUNTIME_DIR not set, defaulting to '/tmp/runtime-r1859698'
## TypeError: Attempting to change the setter of an unconfigurable property.
## TypeError: Attempting to change the setter of an unconfigurable property.
```