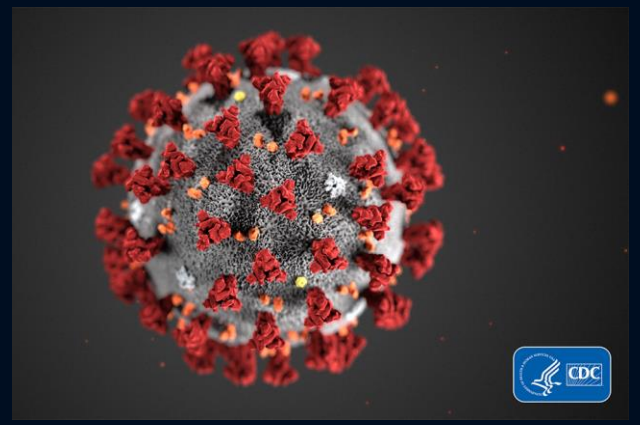
IS590PR Final Project

Analysis of the relationship between the extent of COVID-19 and the developed level of countries

JINGTAO YU
PENGXIN BIAN
YILIN TAN

Background



- The Covid-19 outbreak has caused significant global panic.
- More than 253, 000 deaths and 3.61 million confirmed cases.
- The are some data suggest that the covid-19 pandemic is hitting some countries particularly hard.
- It seems like it hit some high-income countries even harder.
- Our target is to find out the relationship between the extent of COVID-19 and the developed level of countries.

Hypotheses

- 1. The GDP per capita has positive linear relationship with COVID-19 confirmed rate and death rate
- 2. The expected year of schooling has positive linear relationship with COVID-19 confirmed rate and death rate
- 3. The life expectancy at birth has positive linear relationship with COVID-19 confirmed rate and death rate
- 4. The internet user percentage has positive linear relationship with COVID-19 confirmed rate and death rate

Dataset

- **GDP per capita (current US):** API_NY.GDP.PCAP.CD_DS2_en_csv_v2_988471.csv
- **Life expectancy at birth, total (years):** API_SP.DYN.LE00.IN_DS2_en_csv_v2_988752.csv
- **Expected year of schooling:** Expected years of schooling (years).csv
- **Internet user percentage**: Internet users, total (% of population).csv
- **Global Covid-19 confirmed cases:** time_series_covid19_confirmed_global.csv
- **Global Covid-19 death cases:** time_series_covid19_deaths_global.csv
- **Population by Country :**API_SP.POP.TOTL_DS2_en_csv_v2_988606.csv

Code Explanation

- 13 functions totally
- 7 functions for data preprocessing **Pandas.read_csv()**
  - read_covid_csv(): Read datasets of COVID-19 confirmed cases and death cases and store the results in dictionaries
  - read_covid_data(): Transfer the data in dictionaries into DataFrame type
  - read_life_expectancy(): Read the dataset of life expectancy
  - read_population(): Read the dataset of population
  - read_GDP(): Read the dataset of GDP
  - read_Education(): Read the dataset of education
  - read_Internet(): Read the dataset of Internet

Code Explanation(Cont.)

- 3 functions for calculating the statistics for analysis
  - **calculate_covariance():** calculate the covariance of confirmed rate and death rate for life expectancy, GDP, education and Internet separately

    Numpy.cov()
  - calculate_correlation_coefficient(): calculate the correlation coefficient of confirmed rate and death rate for life expectancy, GDP, education and Internet separately

    Numpy.corr()
  - calculate_significance_of_coefficient(): calculate the significance of coefficient of confirmed rate and death rate for life expectancy, GDP, education and Internet separately

    Scipy.stats.pearsonr()

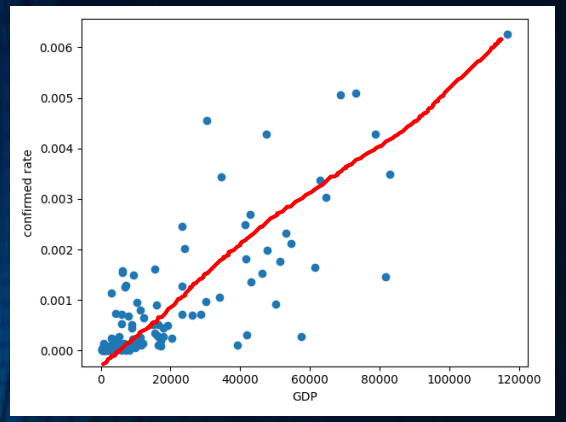
Code Explanation(Cont.)

- 3 functions for analysing and showing the results visually
  - correlation_analysis(): Main function of the correlation analysis to join the data frames together and call the display_analysis_result() function

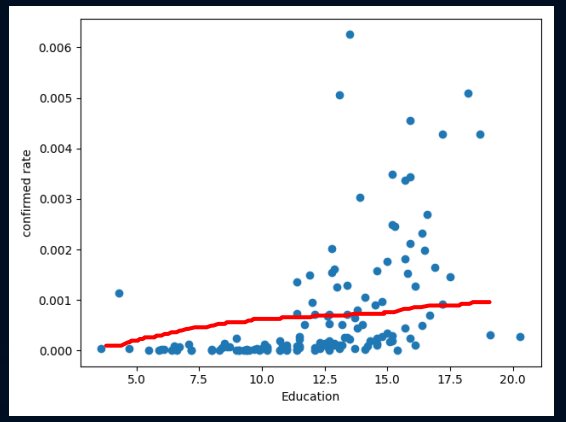
    Pandas.merge()
  - display_analysis_result(): Call the calculation functions and display results in the forms of text and images
  - draw_scatter_plot(): Generate a scatter plot based on two given data frame columns

    Matplotlib.pyplot.scatter()

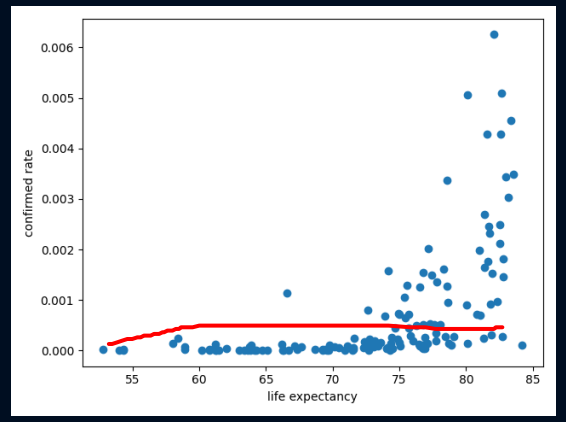
Scatter plots



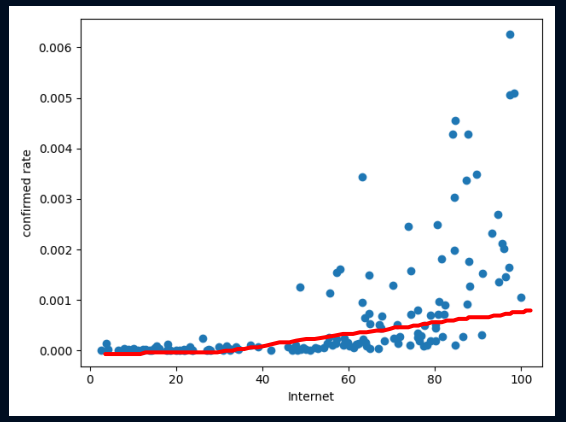
GDP per capita



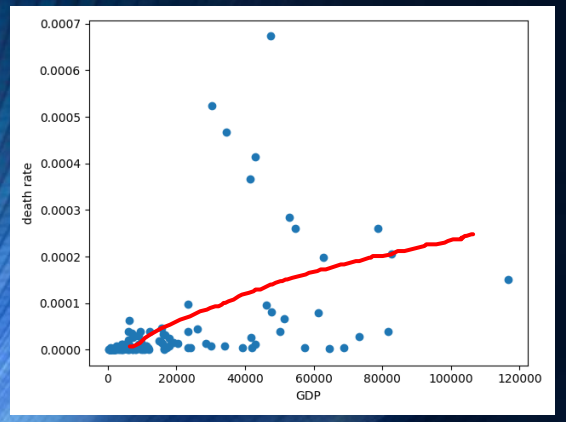
Expected year of schooling

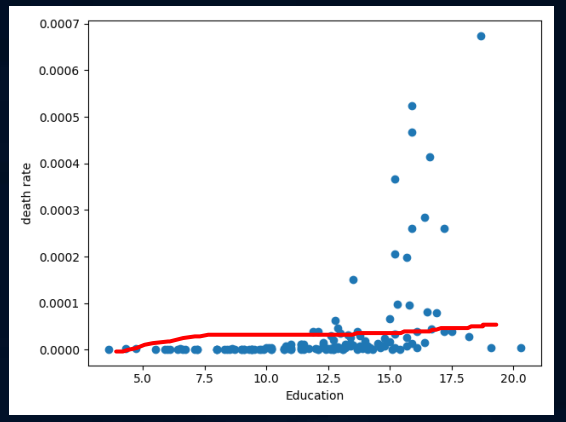


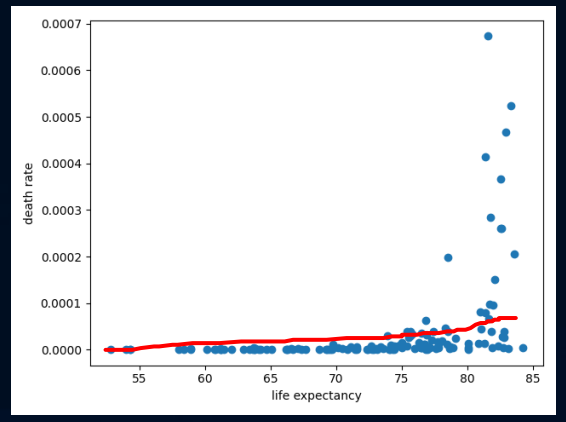
Life expectancy at birth

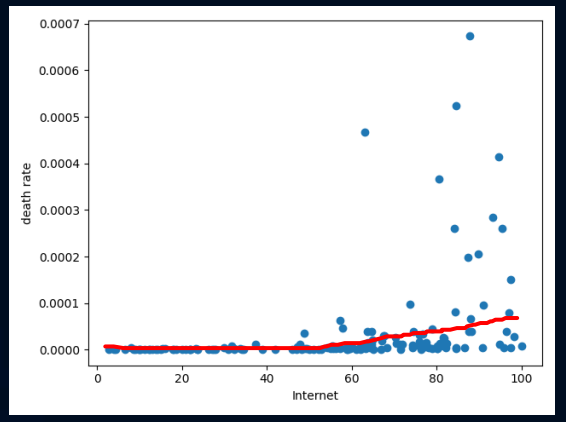


Internet user percentage









Correlation with COVID-19 confirmed rate:

Factor	GDP per capita	Expected year of schooling	Life expectancy at birth	Internet user percentage
Covariance	20.71	0.0019	0.0048	0.019
Correlation coefficient	0.82	0.48	0.55	0.57
Significance of coefficient	3.57e-36	1.08e-09	1.47e-12	5.35e-14

Correlation with COVID-19 death rate:

Factor	GDP per capita	Expected year of schooling	Life expectancy at birth	Internet user percentage
Covariance	1.04	0.00013	0.00030	0.00099
Correlation coefficient	0.49	0.41	0.40	0.36
Significance of coefficient	6.34e-10	3.53e-07	9.37e-07	1.24e-05

Conclusion

All our four hypothesis are supported

That means that countries with higher developed level tend to have more serious extend of COVID-19

We think this may because:

1. Testing problems in less developed countries
2. Low connectivity in less developed countries

Thank You And Stay Safe

Reference

COVID-19 confirmed case and death case:
https://github.com/CSSEGISandData/COVID-19

GDP per capita:
https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Expected year of schooling & Internet user percentage :
http://hdr.undp.org/en/data#

Life expectancy at birth:
https://data.worldbank.org/indicator/SP.DYN.LE00.IN

https://en.wikipedia.org/wiki/COVID-19_pandemic