Analyzing Potential Economic Dependencies of the COVID-19 Mortality Rate

STA303H1 Final Project Part 1

Chenxu (Robin) Mao

Mar 2024

Table of Contents

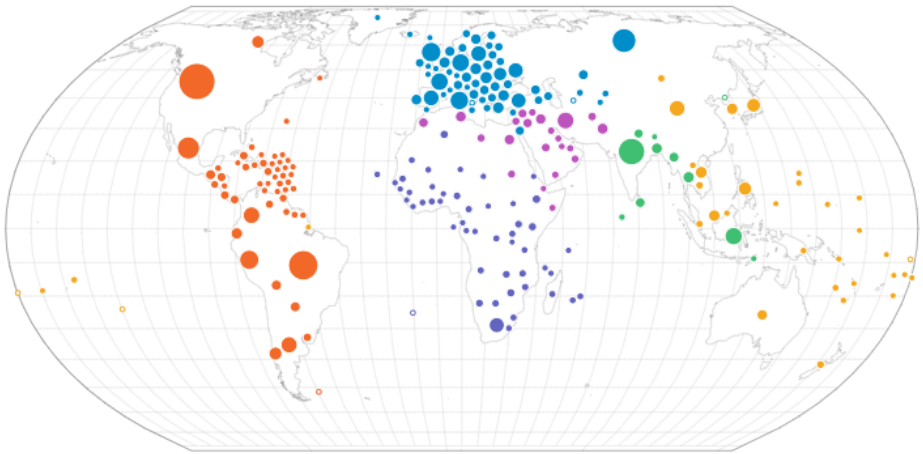
Introduction: Importance

Last 7 days | Last 28 days | Total cumulative

Count | Rate per 100 000

Number of COVID-19 deaths reported to WHO (cumulative total)

World



WHO Regions Africa Americas Eastern Mediterranean Europe South-East Asia Western Pacific

7,035,337

Reported COVID-19 deaths

25 February 2024

Number of COVID-19 deaths reported to WHO (cumulative total)

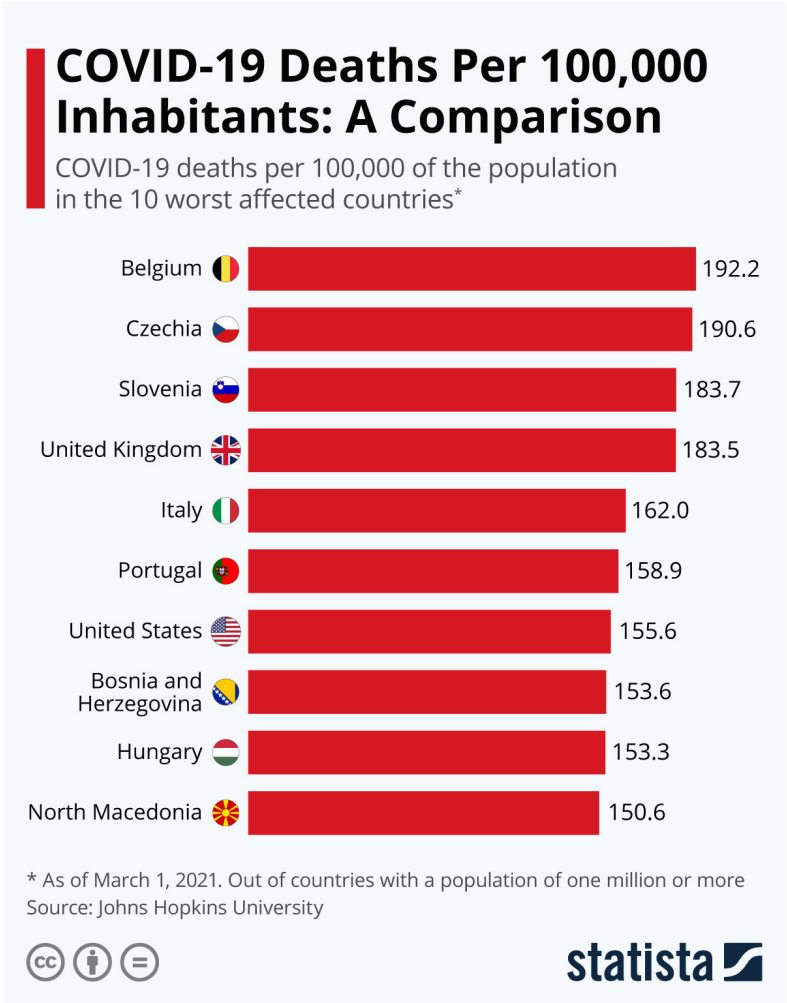
World

Country	Deaths
United States of America	1.2m
Brazil	702.1k
India	533.5k
Show 228 more	
Saint Helena	0
Tokelau	0
Turkmenistan	0

Source: World Health Organization

World Health Organization, 2024[(1)]

Introduction: Importance



Infographic: How Coronavirus Deaths Vary per Million Inhabitants, 2021 [2]

Introduction: Research Question

Research Question

What is the correlation between the COVID-19 mortality rate and economic factors such as Population Count, GDP, Individual Wealth, Literacy Rate, Governmental Type, and Omicron Variant Proportion, between 2019 and 2023?

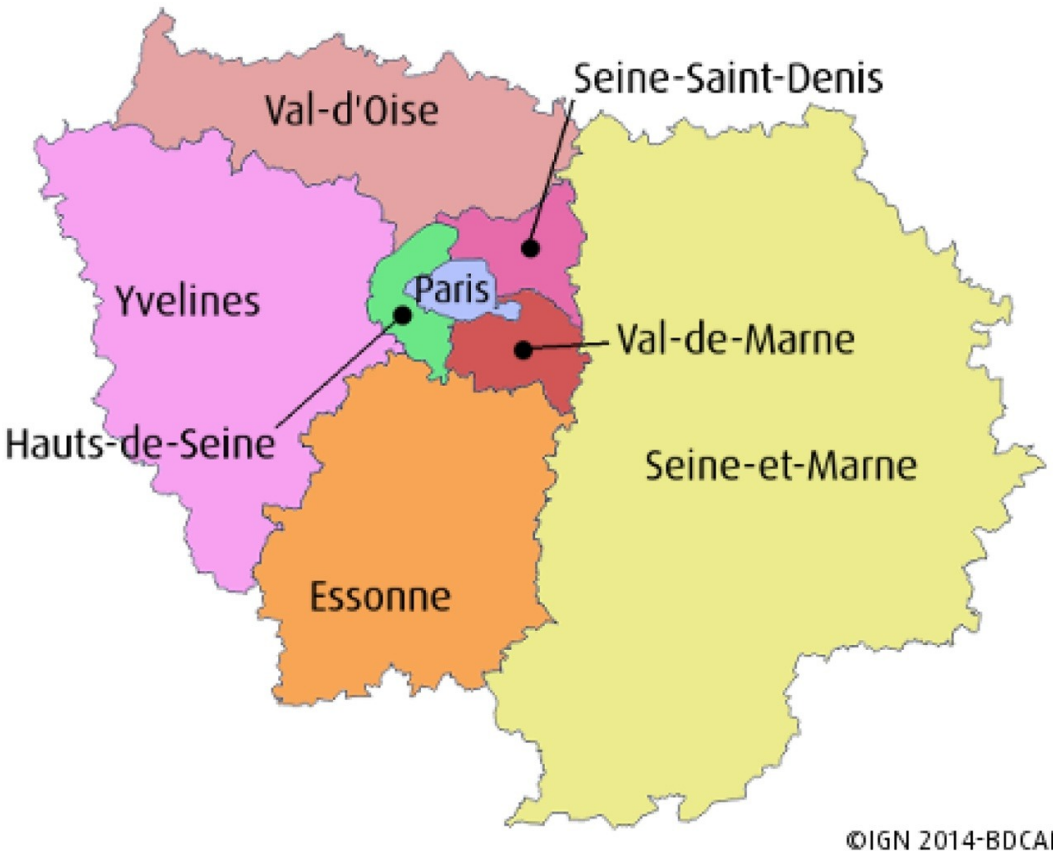
Literature Review 1

- The role of economic structural factors in determining pandemic mortality rates. Goutte et al. (2020) [10]
- **Research Question**: How the role of economic structural factors in the COVID-19 mortality rate in France?
- **Data**: from a governmental agency, INSEE
- **Process**:
  - Collected data from INSEE, conduct PCA analysis
  - Divide data into region-specific groups based on economic and structural characteristics
  - Analyze and model the region-specific data groups
  - Conclude a final model with 30 socioeconomic variables, such as unemployment rate, aging index, population density, etc. (cont')

Literary Review 1: Result

- **Model**: PCA: Principal Component Analysis
- Reduces dimension, transforms variables into a new set of uncorrelated variables (linear combination of the original predictors)
- **Result**: 30 Socioeconomic variables, such as unemployment rate, aging index, population density, and education level, career, etc.
- **Critical Variables**:
  - Unemployment Rate: A higher unemployment rate was associated with higher mortality rates.
  - Poverty Rate: Departments with higher poverty rates experienced higher mortality rates.
  - Share of Social Minima in Income: This variable indicates the proportion of income coming from social minimum benefits, which was higher in departments with higher mortality rates.
  - Share of Population with Little or No Diploma: A higher proportion of the population with low educational attainment was associated with higher mortality rates.

Literary Review 1: Graph



Data groups based on the regional economic and structural status. Goutte et al. (2020) [10]

Literary Review 1: Relevance

Research Question: How the role of economic structural factors in the COVID-19 mortality rate in France?
Dataset: non-public datasets from INSEE
Variables: Unemployment Rate, Poverty Rate, Share of Social Minima in Income, Share of Population with Little or No Diploma, and 26 more.
Similarities:

- data segmentation, use region-specific traits as confounding variables.
- similar approach in economic factors.

Differences:

- no PCA (Principal Component Analysis).
- more mathematical and assumption-focused.
- a broader baseline region.

Literature Review 2

Social, Economic, and Regional Determinants of Mortality in Hospitalized Patients With COVID-19 in Brazil. Rodrigues et al. (2022). [11]

Research Question: What are the social, economic, and regional determinants of mortality in hospitalized patients with COVID-19 in Brazil, and how do these factors influence the odds of mortality based on the patients' economic, social, and epidemiological characteristics?

Data: SIVEP-Gripe by the Brazilian Ministry of Health

Process:

- Collected data from SIVEP-Gripe and PNAD 2020
- Categorized Brazilian states into two macro-regions (Northern and Central-South)
- Conduct logistic regressions
- Conclude

Literature Review 2: Result

- **Model**: Logistic Regression
- Useful in binary outcome, adjustments for confounding variables & non-linearity, interoperability.
- **Result**: a total of 18 socioeconomic & health factors, combined with some random effect adjustments.
- **Critical Variables**:
  - Age: categorical & binary, indicates whether the individual age is above 50.
  - Ethnicity: categorical & binary, 1 if the individual is white.
  - Gender: categorial & binary, 1 if the individual is male.
  - Comorbidities: including cardiovascular disease, diabetes, hematological disease, Down syndrome, obesity, pulmonary disease, liver disease, immunosuppression, renal disease, and neurological disease.
  - Region: categorical & binary. 1 if the individual is in a Northern province.
  - Inverse of relative income

Literature Review 2: Parameters

Parameter		Estimate	Wald qui-square	Pr > ChiSq
Intercept		−1.5534	4143.2854	<0.0001
Age_less_50_years		−0.8672	2678.2748	<0.0001
Ethnicity	Non-white	0.0603	128.3516	<0.0001
Gender	Male	0.1614	1127.7744	<0.0001
Cardiovascular_disease		0.1729	1114.5035	<0.0001
Diabetes		0.1875	1137.8771	<0.0001
Hematological_disease		0.2217	59.5684	<0.0001
Down_syndrome		0.2888	37.9069	<0.0001
Region	Northern	0.3127	2061.7896	<0.0001
Age_50_years_higher		0.4106	652.1627	<0.0001
Obesity		0.4556	2596.4568	<0.0001
Illiterate_patient		0.4566	1096.2094	<0.0001
Pulmonary_disease		0.4858	1513.6012	<0.0001
Liver_disease		0.5021	376.4814	<0.0001
Immunosuppresions		0.5278	1157.7189	<0.0001
Renal_disease		0.6774	2906.2803	<0.0001
Neurological_disease		0.6880	3007.4068	<0.0001
I_relative_income*		0.8959	1029.9471	<0.0001
Number of observations used			840,201	

**I relative income means the inverse of the relative income. The higher that indicator, the worse the individual's economic condition.*

Rodrigues et al. (2022). [11]

Literature Review 2: Relevance

- **Research Question**: What are the social, economic, and regional determinants of mortality in hospitalized patients with COVID-19 in Brazil, and how do these factors influence the odds of mortality based on the patients' economic, social, and epidemiological characteristics?
- **Dataset**: non-public datasets from SIVEP-Gripe
- **Variables**: Age, ethnicity, gender, comorbidities, region, social status, and 11 more.
- **Similarities**:
  - generalized linear model.
  - data segmentation, use region-specific traits as confounding variables.
- **Differences**:
  - more assumption-focuses.
  - more focuses on the economic factors.
  - a broader baseline region.

Literature Review 3

- A global analysis of the effect of temperature, socioeconomic and environmental factors on the spread and mortality rate of the COVID-19 pandemic, Mizanur et al.(2020) [12]
- **Research Question**: "Can temperatures and other socioeconomic factors explain the spatial variability of the rate of spread and mortality of COVID-19 at the global scale?"
- **Data**: WHO, JHU, UN database
- **Process**:
  - Collected and cleaned data from open-sourced database
  - Classified countries based on the national income
  - Conduct partial correlation analysis
  - Evaluate model using AIC
  - Conclude

Literature Review 3: Result

- **Model**: Partial Correlation Analysis
- Useful for controlling confounding variables such as humidity and other environmental factors.
- **Result**:
  - Global Scale: Intercept, Population Growth Rate, Percentage of People Over 60, Min Temperature
  - High Income: Intercept, Population Growth, Frosted Area, Age 60+, Min Temperature
  - Low Income: Intercept, Population Density, GDP Growth Rate, Temperature Variation
- **Critical Variables**:
  - Population Growth Rate
  - Percentage of People Over 60: Percentage of the population that is over 60
  - GDP Growth: Growth in Gross Domestic Product
  - Temperature: Different implementation among different groups with different national wealth

Literature Review 3: Relevance

- **Research Question**: "Can temperatures and other socioeconomic factors explain the spatial variability of the rate of spread and mortality of COVID-19 at the global scale?"
- **Dataset**: Open-source dataset from WHO, JHU, etc.
- **Variables**: Population Growth Rate, Percentage of People Over 60, GDP Growth, Temperature
- **Similarities**:
  - same datasets
  - similar usage of generalized linear models
- **Differences**:
  - sightly differences in the variable of interests
  - Poisson regression instead of logistic regression, with more numerical variables
  - a more detailed data segmentation as introduced in literature review 1 & 2

Data Description: Dataset

Main Dataset

COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University
https://github.com/CSSEGISandData/COVID-19 [3]
Note1: All other datasets are cited in references.
Note2: Not all variables from all datasets will be used.

Data Description: Variable Chosen

Y(dependent):
COVID-19 Mortality Rate: mortality proportion of a country's population. Measured in deaths per million.

X(independent):
Population: Population count of a country. We use it as it could be directly related to the COVID-19 spread pattern. [4]
GDP: Gross Domestic Product, the total annual output of a country. Measured in $USD. We use it as it represents the quantitative wealth of a country.[5]
GDPperCapita: the total annual income of a citizen of a country. Measured in $USD. We use it as it represents the quantitative individual wealth of a country.[6]
Unemployment Rate: the proportion of the unemployed labor force proportion within the country. We use it as it represents the labor structure.[7]
Inflation Rate: measured in CPI (Consumer Price Index), the inflation rate of the country. We use it as it represents the monetary status of a country.[8]

Latitude & Longitude: The official latitude and longitude of the country.

Data Description: Exploratory Data Analysis

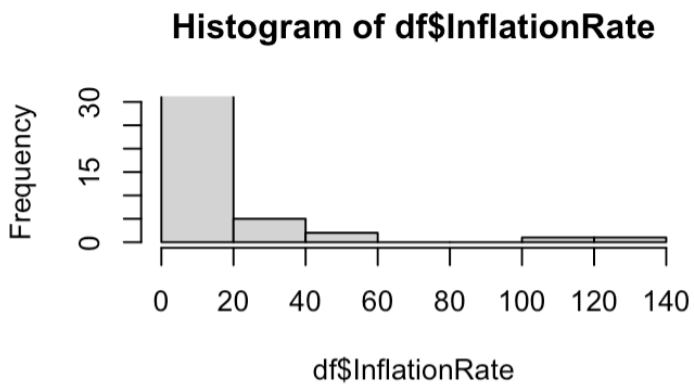
Original dataset: 202 time series
Final dataset: 131 time series, each with a mortality toll time series of size 1152, with daily mortality counts between 2021 and 2023.
Summary Variable Statistics:

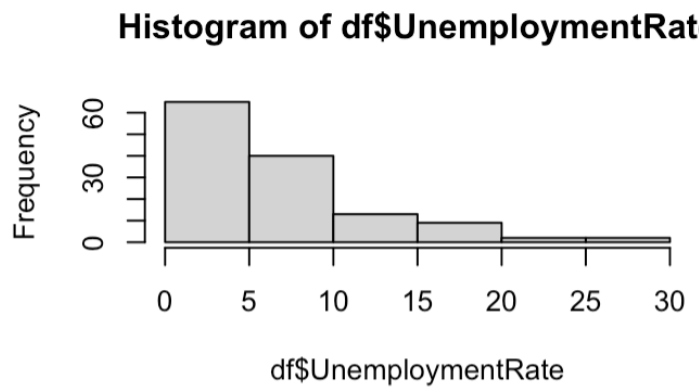
Country	InflationRate	UnemploymentRate	GDPperCapita	GDP	Population
Min	1.350779	259.025	787	542686976	222382
Max	138.8085	125006	5415	2.54397×10^{13}	1417173173
Mean	11.57413	18944.7	7507	692785674001	50890490
S.d.	15.93115	25618.92	7560	2.773837×10^{12}	178855864

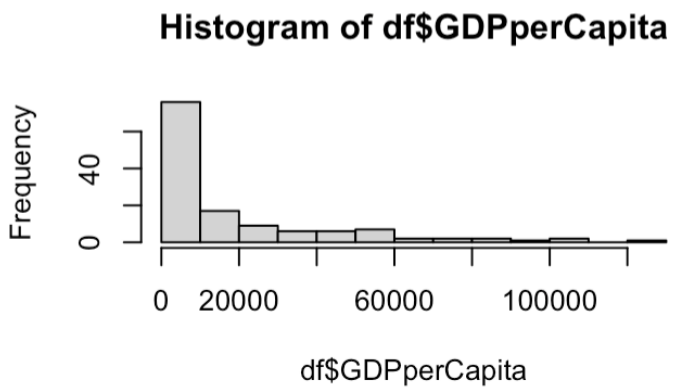
Potential Problem: a decreased sample size

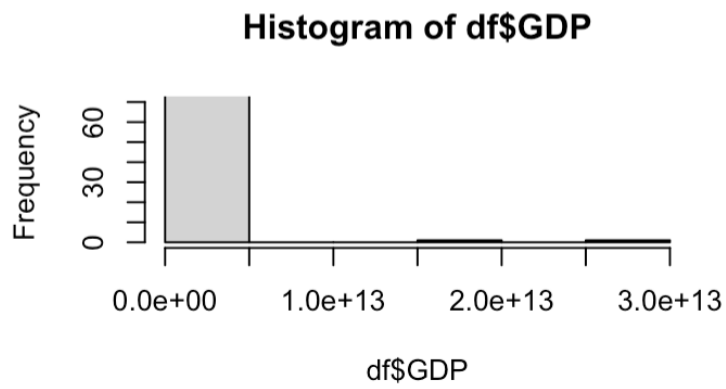
Data Description: Exploratory Data Analysis

Predictor Histogram:





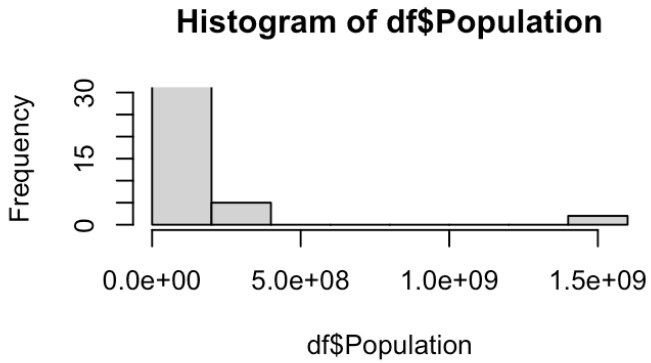




Significantly right-skewed. Needs to be transformed.

Data Description: Exploratory Data Analysis

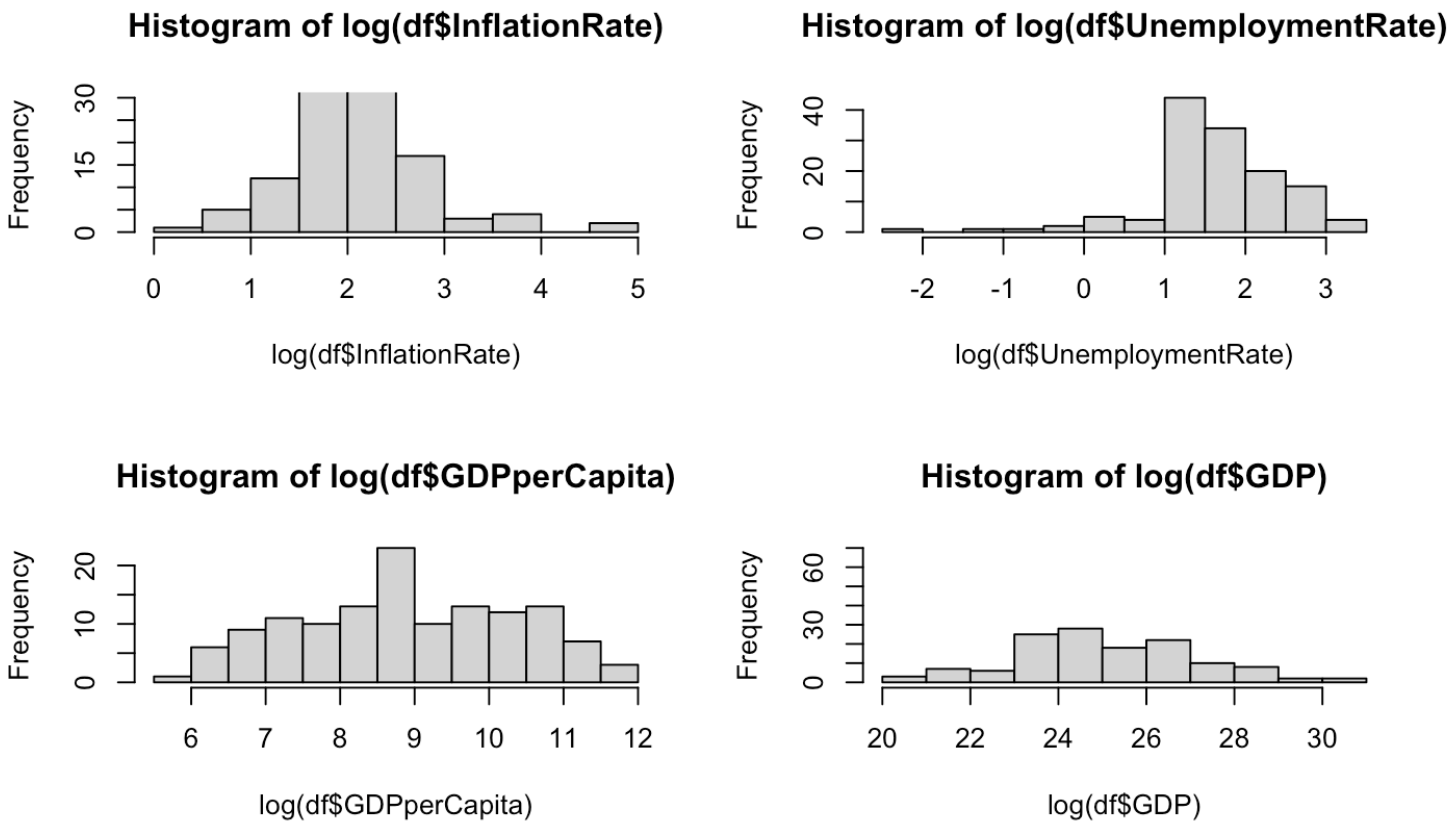
Predictor Histogram:



Significantly right-skewed. Needs transformation.

Data Description: Exploratory Data Analysis

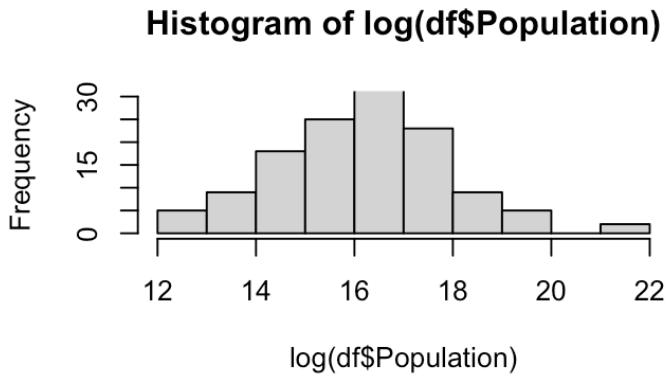
log(Predictor) Histogram:



Log(unemployment rate) needs to be further transformed.

Data Description: Exploratory Data Analysis

log(Predictor) Histogram:



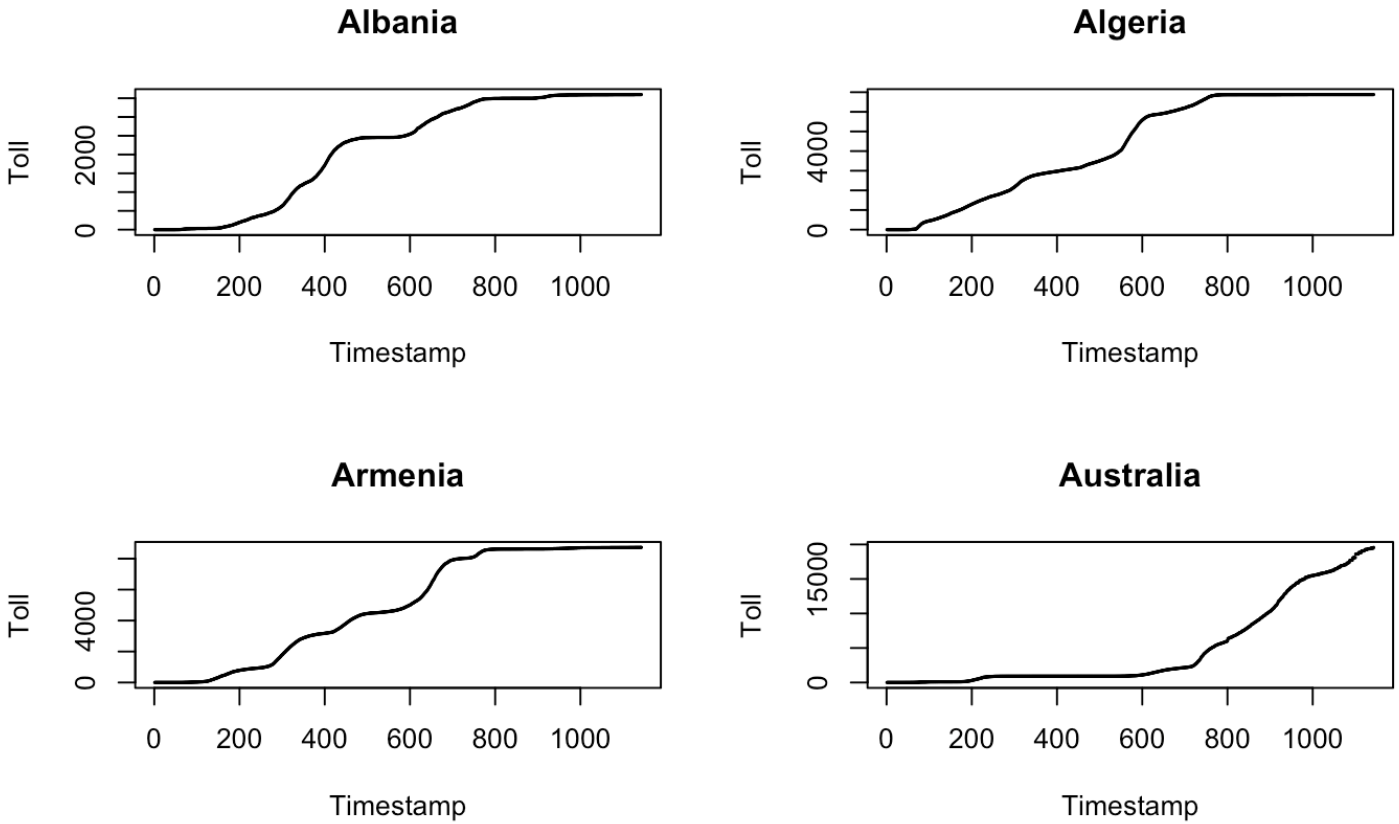
Data Description: Exploratory Data Analysis

Correlation plot among predictors:

Collinearity exists between population and GDP. Needs transformation, and outliers and influential points removal.

Data Description: Exploratory Data Analysis

Example time series



Have to check ACF, PCF, if stationary, etc.

Have to convert toll into a proper rate.

Discussion: Plan for Answering Research Question

- Poisson Regression for Rate
- Potential Confounding Variable:
  - Lat & Long (literature review 1, 2), for data segmentation
  - further binary definition on numerical variables
- Response variable: COVID-19 Mortality Rate
- Predictor variables:
  - Population
  - GDP
  - GDP per Capita
  - Unemployment Rate
  - Inflation Rate
  - Lat & Long

Discussion: Plan for Answering Research Question

- Process:
  1. Clean dataset
  2. Model diagnostics for assumption violations
  3. Variable reduction using AIC or BIC
  4. Model diagnostics on the reduced model
  5. Validations such as cross-validation, Receiver Operating Characteristics
  6. Accuracy & bias test on the reduced model
- If final model high accuracy $\rightarrow$ there exists correlation between COVID-19 Mortality Rate & the predictor variables given ...

Discussion: GLM Assumptions

Note that we are using Poisson Regression.

1. GLM: Independent $y_i's$ (EDA supported)
2. GLM: $g(\mu_i) = \vec{x}_i^T \vec{\beta}$, given function $g(.)$ (EDA supported)
3. Poisson: $y_i \geq 0, \forall y$ (stands by default)
4. Poisson: Mortality distribution (conditioning on the model) follows a Poisson Distribution (to be further determined)
5. Poisson: Mean and Variance are identical (to be further determined)
6. Regression: Absence of multicollinearity (EDA somewhat supported)
7. Regression: Lack of Influential Points (to be further determined)

References

1. World Health Organization. (2024). COVID-19 deaths — WHO COVID-19 dashboard. Datadot. https://data.who.int/dashboards/covid19/deaths?n=c
2. Infographic: How Coronavirus deaths vary per million inhabitants. (2021, March 1). Statista Infographics. https://www.statista.com/chart/21170/coronavirus-death-rate-worldwide/
3. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1.
4. World Bank. (2021). Population, Total — Data. Worldbank.org. https://data.worldbank.org/indicator/SP.POP.TOTL
5. World Bank. (2021b). GDP (current US$) — Data. The World Bank. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
6. World Bank. (2021a). GDP per capita (current US$) — Data. Worldbank.org. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD
7. World Bank. (2023). Unemployment, total (% of total labor force) (modeled ILO estimate). Worldbank.org. https://data.worldbank.org/indicator/sl.uem.totl.zs —Data. Worldbank.org. https://data.worldbank.org/indicator/SE.ADT.LITR.ZS
8. World Bank. (2023a). Inflation, consumer prices (annual %) — Data. Worldbank.org. https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG
9. https://github.com/albertyw/avenews/blob/master/old/data/average-latitude-longitude-countries.csv
10. Goutte, S., Péran, T., & Porcher, T. (2020). The role of economic structural factors in determining pandemic mortality rates: Evidence from the COVID-19 outbreak in France. Research in International Business and Finance, 54, 101281. https://doi.org/10.1016/j.ribaf.2020.101281
11. Rodrigues, W., da Costa Frizzera, H., Trevisan, D. M. de Q., & Prata, D. (2022, March 9). Social, economic, and regional determinants of mortality in hospitalized patients with COVID-19 in Brazil. Frontiers. https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.856137/full
12. Rahman, M., Islam, M., Shimanto, M.H. et al. A global analysis on the effect of temperature, socio-economic and environmental factors on the spread and mortality rate of the COVID-19 pandemic. Environ Dev Sustain 23, 9352–9366 (2021). https://doi.org/10.1007/s10668-020-01028-x