A Traditional Approach to a Novel Coronavirus

An Analytical Study

by

Souparno Bhattacharya

Objective:

To analyse the correlations between the spread of SARS-COVID-2 in various states/provinces of the world and their respective socioeconomic Key Performance Indicators.

KPIs considered in the analyse:

- Human Development Index
- Population Density (per sq. Km)
- Allocated Finances for Healthcare (in dollars)
- Life expectancy at birth
- Infant mortality rate (per 1000 births)

Introduction:

The novel coronavirus, formally referred to as SARS-COVID-2 or COVID19 has turned out to be one of the most adverse pandemics in recent times. At the time of creation of this report, around 9 million people have been affected by the pandemic since its discovery in late December 2019, with more than 450,000 deaths worldwide. Due to the high contagiousness of the virus, most healthcare systems in the world are overwhelmed by the burden of providing adequate facilities to the sick. Also, as of now, Europe and the Americas have been most gravely affected and even though the spread of the disease is supposedly receding in the European countries, the contagion is very much on the rise in the Americas. On an average, out of the nearly 470,000 closed cases worldwide, 9% have resulted in fatalities with the rest recovering back to normalcy.

The WHO has established that the virus acutely affects the respiratory system and inhibits the immunity system of the host, thus causing various types of respiratory organ failures which are especially lethal for children and senior citizen. Since there is no formal medical remedy available at the moment to treat the infected, the patients are quarantined and provided with ventilator support for atleast two weeks, during which the infected host attempts to get rid of the virus. In most cases the host has 80 to 90 percent chance of survival, but then the chances are subjected to various other factors like previous medical history and age.

What makes the virus especially savage is the ease with which it can infect a host and its extremely high multiplicity, which makes it extremely contagious. The virus usually enters a host through the nose, mouth or eyes. It can remain dormant on a still surface from 3 to 12

hours and can be active for a fortnight in a live host. Thus, anybody coming in contact with any infect surface or entity has an extremely high chance of catching the infection. Also, due to its novel nature, the human immunity system is slow to react to the virus, which combined with its high doubling rate makes it fatal for the patient.

This study attempts to understand the spread of the virus in various geographical areas around the world and tries to establish meaningful correlations (if any) between the number of people affected by the virus and certain traditional socio-economic indicators of the places they live in, such as standard of living, general immunity, population density and the expanse of healthcare infrastructure.

DISCLAIMER: This study has been conducted in a non-institutional capacity for informal exploratory research purposes and does NOT endorse or advice any of the conclusions drawn from the same. The author is NOT accountable for the authenticity of the sources of data used in this analysis.

Data Collection :

The provinces and states of the following countries were taken into consideration:

- Australia
- China
- Germany
- Iceland
- Italy
- Spain

The population, total area and human development index of the states/provinces of each of these countries were parsed from their respective Wikipedia pages (links mentioned in References section). The country profiles were obtained partly from an existing dataset on Github and partly from Wikipedia.

To translate the country wise socioeconomic KPI data to provinces, the weighted HDI of each of the provinces were calculated and then multiplied with the respective factors. Thus, the values of fields like Infant mortality, health expense and life expectancy are relative to the weighted HDI of the corresponding provinces and are not absolute measures.

Please refer to the following datasets in the repository for a better understanding of the datasets used.

- COVID_19_country_data.csv:
- COVID_19_province_data.csv
- Country_profile_variables.csv
- COVID_19_analysis_data.csv
- COVID_19_analysis.xlsx

NOTE: Due to the small size and population of Iceland, it has been considered as a single province

Data Understanding:

- **HDI:**

  Human Development Index is a composite index of a particular civilized geographical area calculated using factors such as life expectancy, education and per capita income indicators. It is an abstract reflection of the standard of living in that area

- **Weighted HDI:**

  To create a relative measure for translating the country specific socio-economic KPIs into province specific KPIs in a graded manner.

  Calculation:

  Weighted_hdi= (province_hdi)/(country_hdi)

- **Population Density:**

  It is the measure of how densely a certain geographical area is civilised.

  Calculation:

  Population_density=(province_population)/(province_area_in_sq_kms)

- **Urban Population:**

  Number of people living in urban areas in a particular province. It is often considered to be a measure of development.

  Calculation:

  Urban_pop=(weighted_hdi)(per_cent_of_urban_population_in_country)*(population_of_province)*

- **Access to Clean Water:**

  Number of people having access to clean drinkable water in rural as well as urban areas. A measure of access of basic hygiene to the population of the province.

  Calculation:

  Clean_water_access=(weighted_hdi)[(per_cent_pop_using_drinkable_water_urban)*(Urban_pop)+(per_cent_pop_using_drinkable_water_rural)*[(population)-(urban_pop)]]*

- **Health Expense:**

  Net amount of funds (in dollars) allocated for the development of healthcare infrastructure and improvement of health standards of the population in a province.

  Calculation:

  Health_expense=(weighted_hdi)[(population_of_province)*[(per_cent_gdp_allocated_for_healthcare)*(GDP_of_country)/(population_of_country)]]*

- **Life Expectancy:**

  An average estimation of the number of years a person is expected to be alive in a specific state/province. An indirect measure of the immunity of the people of the place.

  Calculation:

  Life_expectancy=(weighted_hdi)(life_expectancy_in_the_country)*

- **Infant Mortality Rate:**

  Number of children dying within 12 months of birth per thousand live births. A measure of immunity and nutrition in a community.

  Calculation:

  Infant_mortality=(weighted_hdi)(infant_mortality_in_the_country)*

- **Confirmed Cases:**

  Number of confirmed COVID-19 cases in a given province/state.

Data Wrangling:

- First, **the demographic data parsed** from the webpages were wrangled and compiled using Python (Jupyter Notebooks) resulting in the **'COVID_19_province_data.csv'** Dataset.
- Next **the data regarding the given KPIs of the concerned countries were compiled** and obtained from the 'country_profile_variables.csv' dataset and stored in **'COVID_19_country_data.csv'** using R.
- Finally, the above mentioned **datasets were compiled** via outer-join using Python and **the no. of confirmed cases** obtained from the Jon Hopkins University open source **dataset was** added to this and the result was stored as **'COVID19_analysis_data.csv'**.
- The Final dataset was analysed using Python and visualisations were obtained using R and Excel.

P.T.O

Observations:

Correlations between the KPIs and the No. of Confirmed Cases:

	hdi	population_density	urban_pop	clean_water_access	health_expense	life_expectancy	infant_mortality	confirmed
hdi	1.000000	0.134897	-0.585306	-0.586858	-0.047214	0.840657	-0.706297	0.228712
population_density	0.134897	1.000000	-0.045409	-0.042726	-0.089021	0.347288	0.285398	-0.052090
urban_pop	-0.585306	-0.045409	1.000000	0.999794	0.590572	-0.396119	0.692251	-0.036217
clean_water_access	-0.586858	-0.042726	0.999794	1.000000	0.582696	-0.392871	0.695597	-0.038036
health_expense	-0.047214	-0.089021	0.590572	0.582696	1.000000	-0.130107	0.180100	0.362084
life_expectancy	0.840657	0.347288	-0.396119	-0.392871	-0.130107	1.000000	-0.369076	0.215114
infant_mortality	-0.706297	0.285398	0.692251	0.695597	0.180100	-0.369076	1.000000	-0.234310
confirmed	0.228712	-0.052090	-0.036217	-0.038036	0.362084	0.215114	-0.234310	1.000000

Exb1.1: Pearson Coefficients of the Key Performance Indicators and the No. of Confirmed COVID-19 Cases.

From Exhibit 1.1 it can be observed that No. of Confirmed cases have the strongest positive correlations with the Health expenditure of a Province followed by the Human Development Index and Life Expectancy and strongest negative correlation with the Infant Mortality Rate.

P-values of the KPIs w.r.t No. of Confirmed Cases:

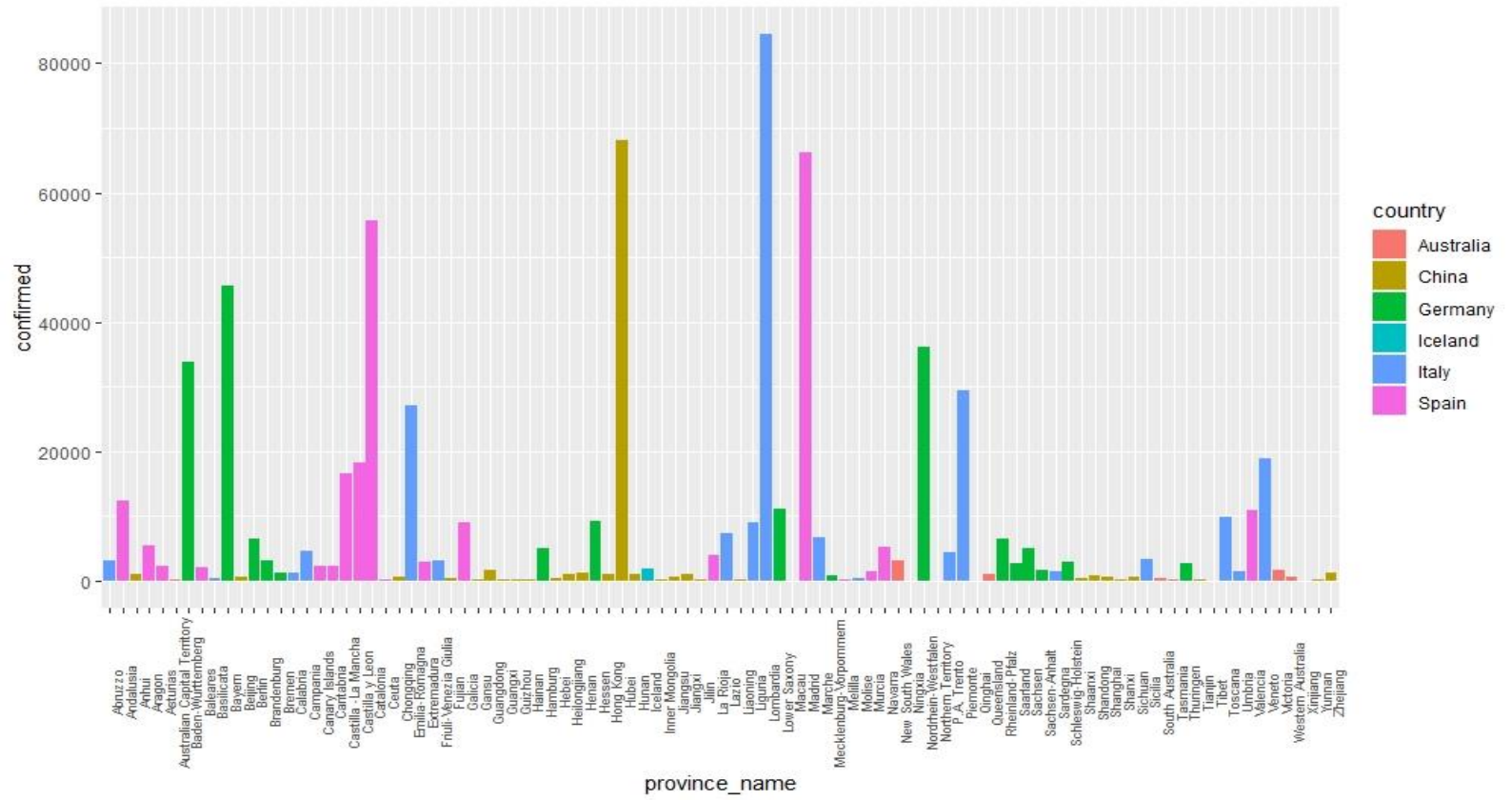
	Feature	P value
0	hdi	0.026606
1	population_density	0.618050
2	urban_pop	0.728931
3	clean_water_access	0.715878
4	health_expense	0.000336
5	life_expectancy	0.037329
6	infant_mortality	0.023022

Exb1.2: P-values of the Key Performance Indicators of a Province with their respective no. of Confirmed Cases.

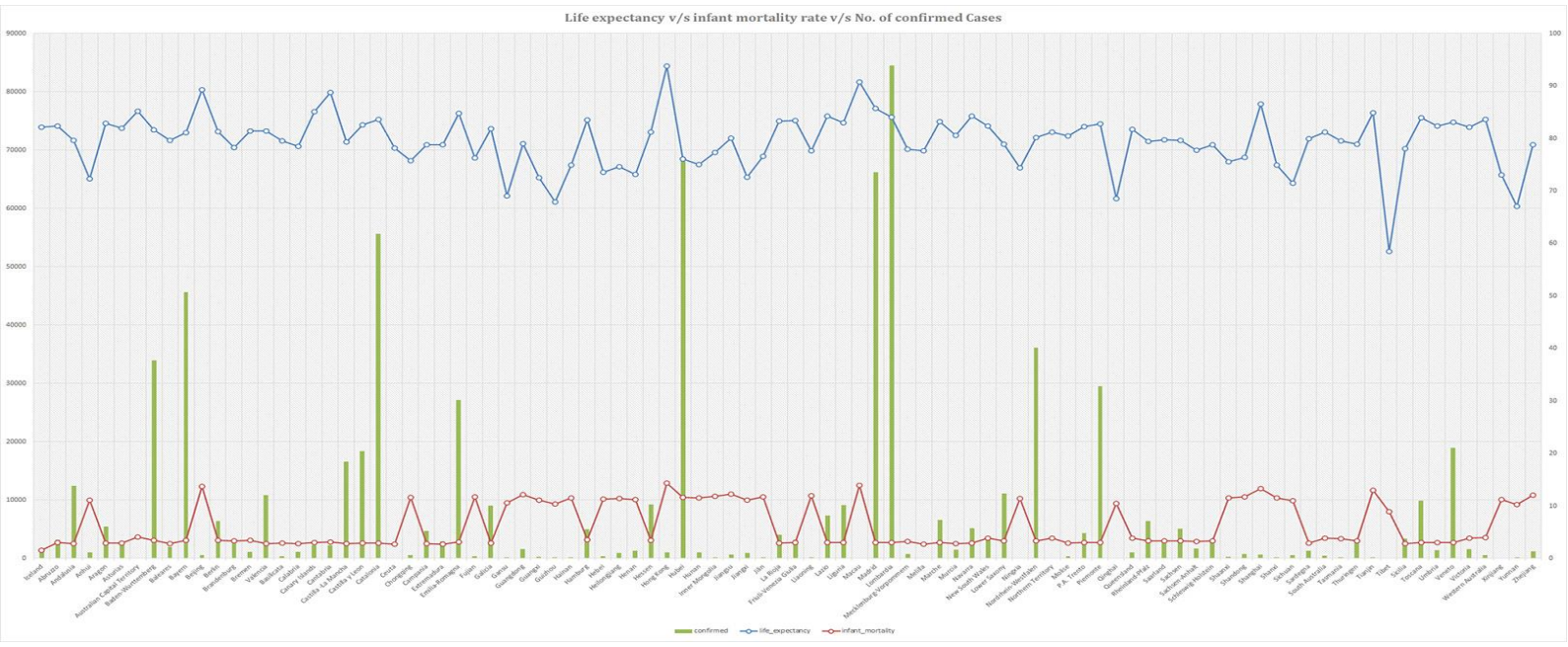
Exhibit 1.2 observes high levels of significance (p-values) for factors like Population Density, Urban Population and Access to Clean Water (here, taken as a measure of basic hygiene).

NOTE: The above observations can be inferred in various ways and the conclusions have been kept to the discretion of the reader. This study does not endorse or advise any particular conclusion derived from the given observations.

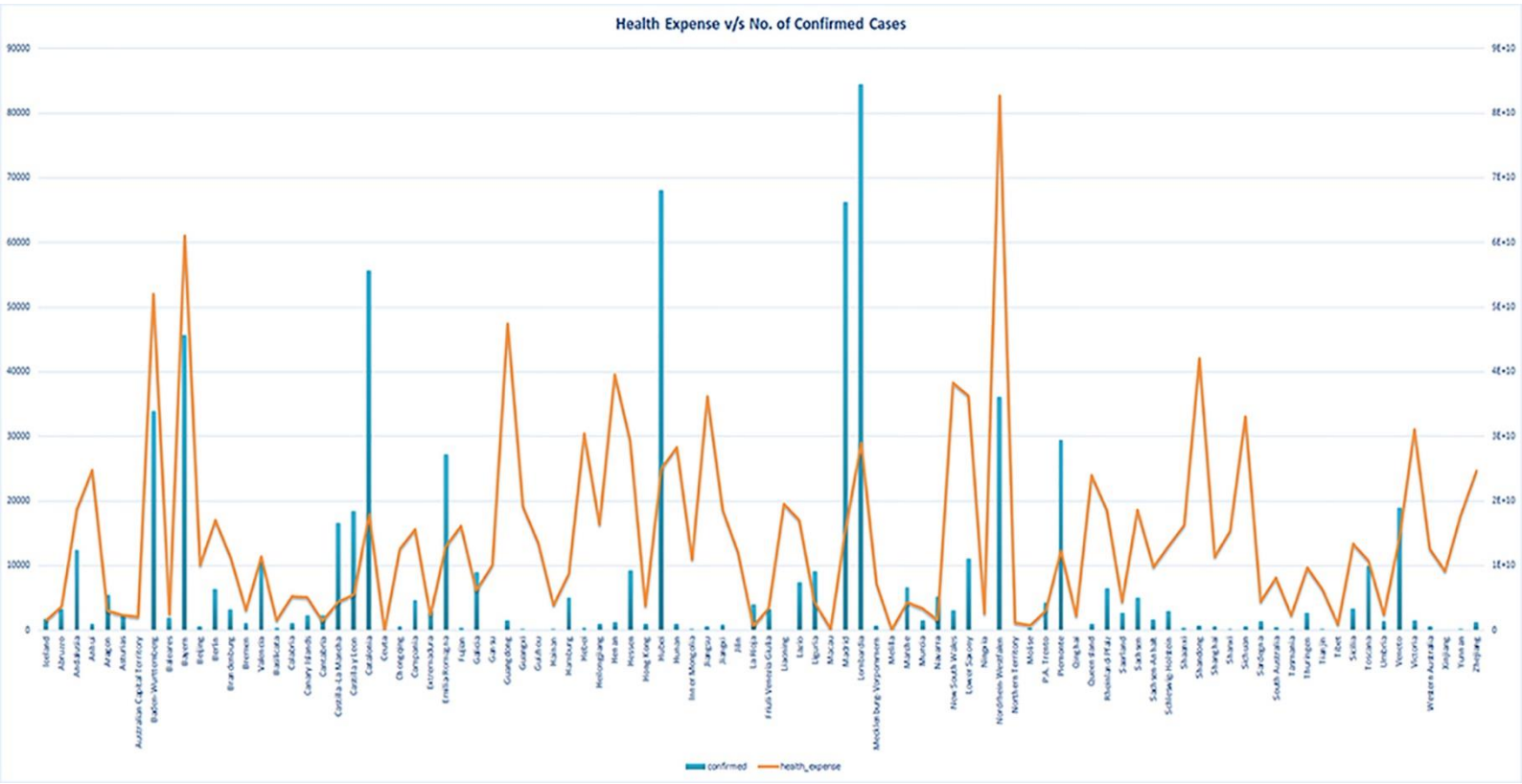
Visualisations:



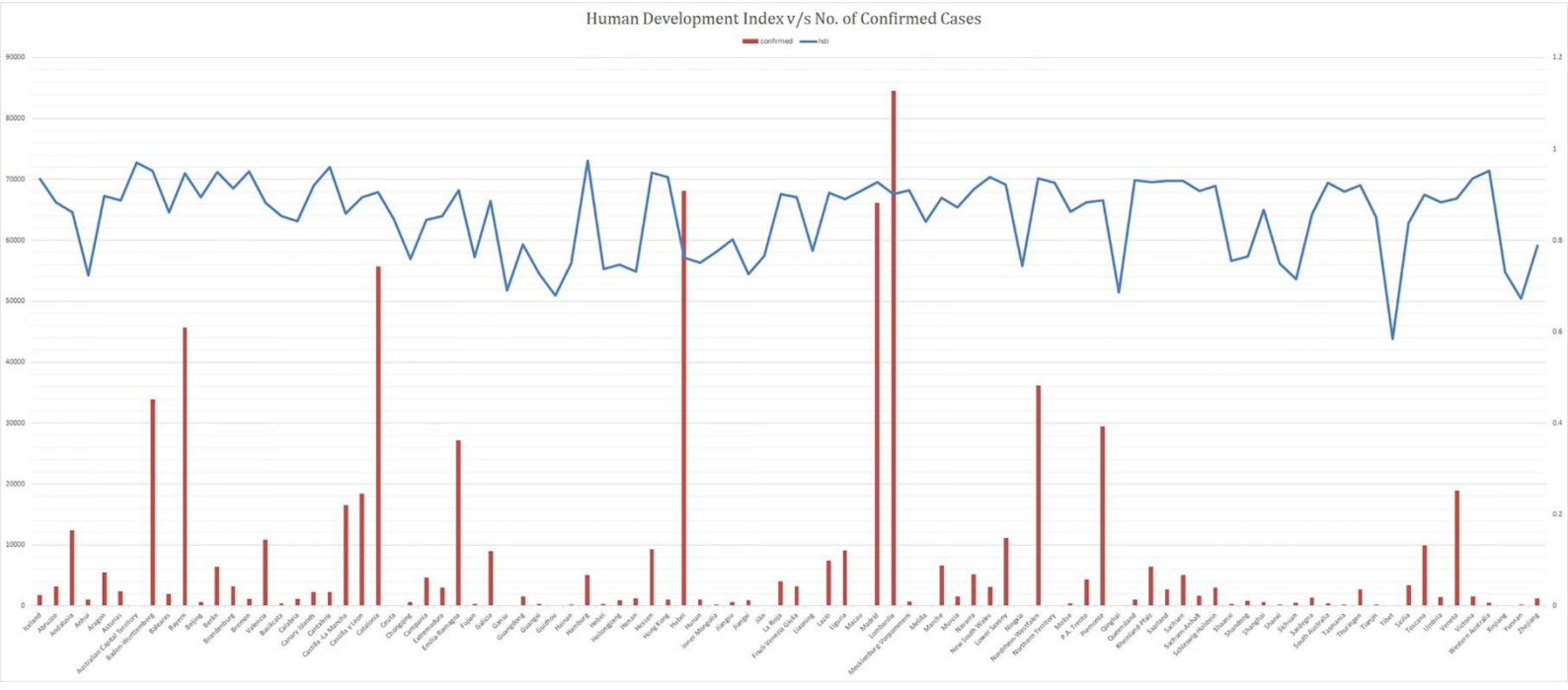
Exb2.1: No. of Confirmed Cases in Various Provinces



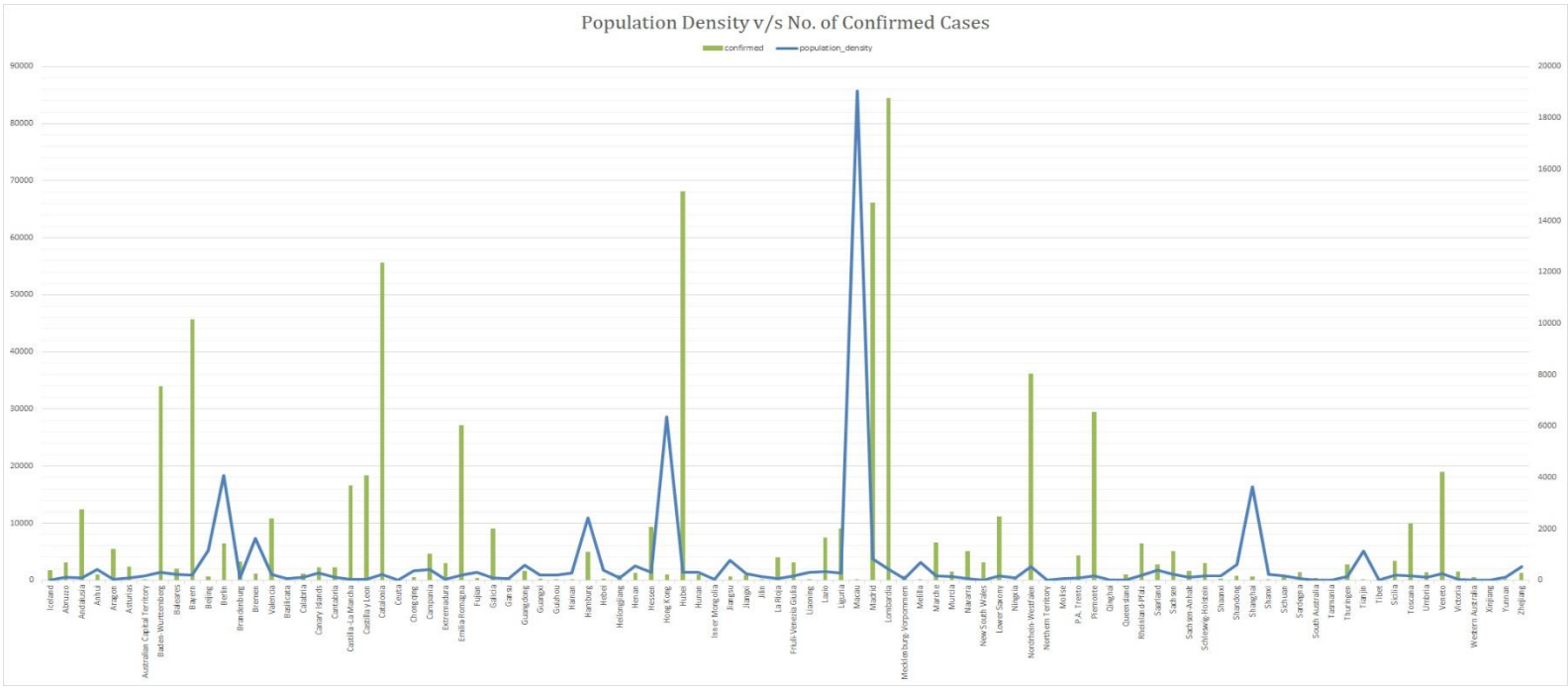
Exb 2.2: Life Expectancy(Blue) v/s Infant Mortality Rate(Red) v/s No. of Confirmed Cases



Exb 2.3: Health Expense(Orange) v/s No. of Confirmed Cases



Exb 2.4: HDI(Blue) v/s No. of Confirmed Cases



Exb: 2.5: Population Density(Blue) v/s No. of Confirmed Cases.

References:

- https://en.wikipedia.org/wiki/Autonomous_communities_of_Spain
- https://en.wikipedia.org/wiki/Provinces_of_China
- https://en.wikipedia.org/wiki/Provinces_of_Italy
- https://en.wikipedia.org/wiki/States_of_Germany
- https://en.wikipedia.org/wiki/States_and_territories_of_Australia
- https://en.wikipedia.org/wiki/List_of_Italian_regions_by_Human_Development_Index
- https://en.wikipedia.org/wiki/List_of_Spanish_autonomous_communities_by_Human_Development_Index
- https://en.wikipedia.org/wiki/List_of_German_states_by_Human_Development_Index
- https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Greater_China_by_Human_Development_Index
- https://en.wikipedia.org/wiki/List_of_Australian_states_and_territories_by_Human_Development_Index
- http://hdr.undp.org/en/countries/profiles/ISL
- https://github.com/CSSEGISandData/COVID-19