

COVID-19 Associative Plots

untitled

03/04/2021

Contents

Introduction	2
Data Structure	2
Exploratory Data Analysis	4
A. Highly correlated variables to Covid 19 cases	4
B. Relationship between GDP per capita and life expectancy	5
C. Relationship between Handwashing Facilities and Extreme Poverty	6

Introduction

Our World in Data COVID-19 data set contains up-to-date data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. The data set was derived from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and can be found here: <https://github.com/CSSEGISandData/COVID-19>. Analyzing this data set will give us some incredible insights into how the covid-19 virus has affected several countries across the globe. We will explore such insights through exploratory data analysis right after we clean and preprocess the data.

Data Structure

In this data set we have a mix of character and numeric variables. It has 59 variables and 70,645 observations.

```
## 'data.frame': 78872 obs. of 59 variables:
## $ iso_code : chr "AFG" "AFG" "AFG" "AFG" ...
## $ continent : chr "Asia" "Asia" "Asia" "Asia" ...
## $ location : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ date : chr "2020-02-24" "2020-02-25" "2020-02-26" "2020-02-27" ...
## $ total_cases : num 1 1 1 1 1 1 1 1 2 4 ...
## $ new_cases : num 1 0 0 0 0 0 0 0 1 2 ...
## $ new_cases_smoothed : num NA NA NA NA NA 0.143 0.143 0 0.143 0.429 ...
## $ total_deaths : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths_smoothed : num NA NA NA NA NA 0 0 0 0 0 ...
## $ total_cases_per_million : num 0.026 0.026 0.026 0.026 0.026 0.026 0.026 0.026 0.051 ...
## $ new_cases_per_million : num 0.026 0 0 0 0 0 0 0 0.026 0.051 ...
## $ new_cases_smoothed_per_million : num NA NA NA NA NA 0.004 0.004 0 0.004 0.011 ...
## $ total_deaths_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths_smoothed_per_million : num NA NA NA NA NA 0 0 0 0 0 ...
## $ reproduction_rate : num NA NA NA NA NA NA NA NA NA NA ...
## $ icu_patients : num NA NA NA NA NA NA NA NA NA NA ...
## $ icu_patients_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ hosp_patients : num NA NA NA NA NA NA NA NA NA NA ...
## $ hosp_patients_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_icu_admissions : num NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_icu_admissions_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_hosp_admissions : num NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_hosp_admissions_per_million : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests : num NA NA NA NA NA NA NA NA NA NA ...
## $ total_tests : num NA NA NA NA NA NA NA NA NA NA ...
## $ total_tests_per_thousand : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_per_thousand : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_smoothed : num NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_smoothed_per_thousand : num NA NA NA NA NA NA NA NA NA NA ...
## $ positive_rate : num NA NA NA NA NA NA NA NA NA NA ...
## $ tests_per_case : num NA NA NA NA NA NA NA NA NA NA ...
## $ tests_units : chr "" "" "" "" ...
## $ total_vaccinations : num NA NA NA NA NA NA NA NA NA NA ...
## $ people_vaccinated : num NA NA NA NA NA NA NA NA NA NA ...
## $ people_fully_vaccinated : num NA NA NA NA NA NA NA NA NA NA ...
```

```

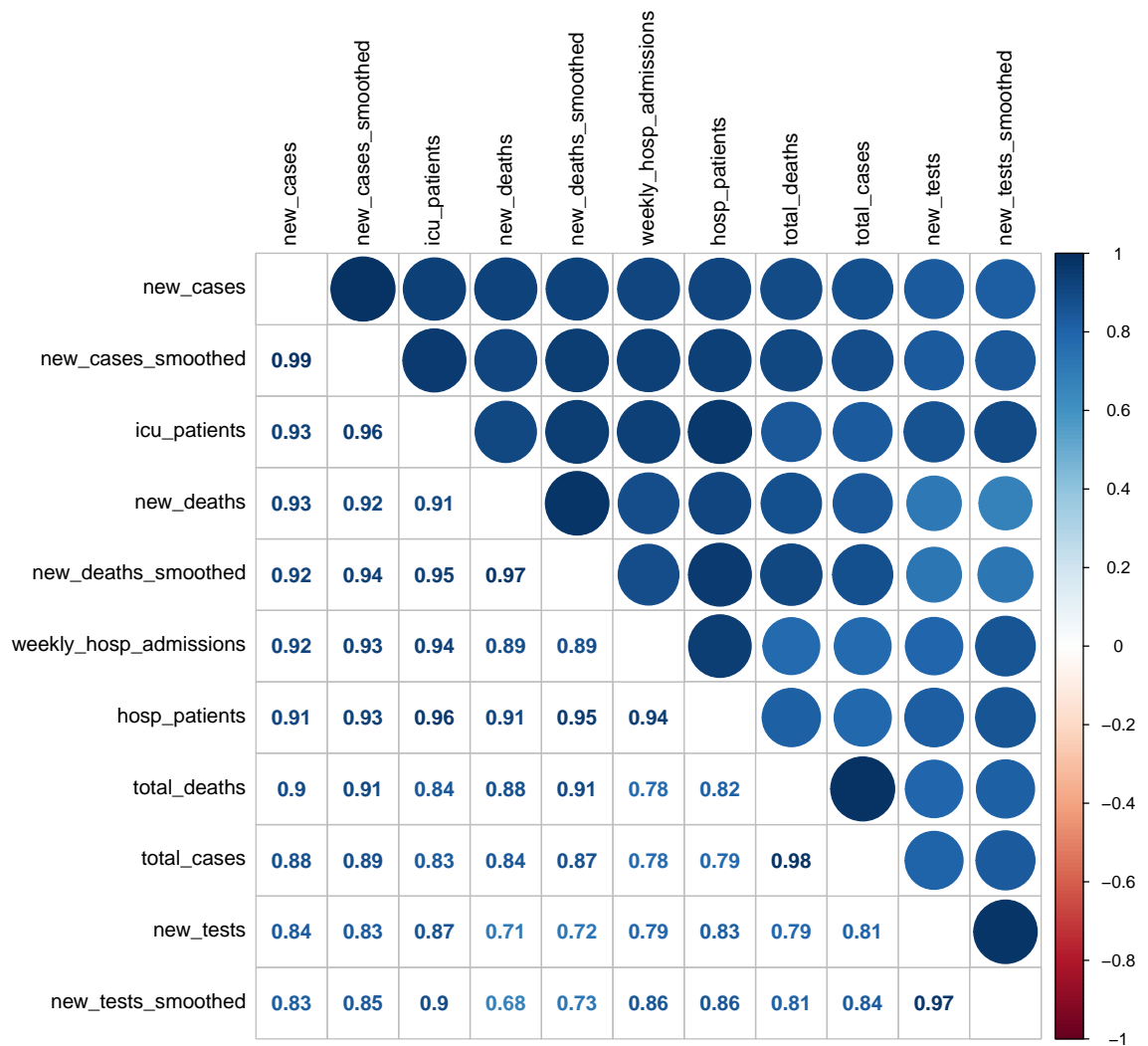
## $ new_vaccinations                : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_vaccinations_smoothed       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_vaccinations_per_hundred  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ people_vaccinated_per_hundred   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ people_fully_vaccinated_per_hundred : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_vaccinations_smoothed_per_million: num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stringency_index                : num  8.33 8.33 8.33 8.33 8.33 ...
## $ population                      : num  38928341 38928341 38928341 38928341 38928341 ...
## $ population_density              : num  54.4 54.4 54.4 54.4 54.4 ...
## $ median_age                     : num  18.6 18.6 18.6 18.6 18.6 18.6 18.6 18.6 18.6 ...
## $ aged_65_older                  : num  2.58 2.58 2.58 2.58 2.58 ...
## $ aged_70_older                  : num  1.34 1.34 1.34 1.34 1.34 ...
## $ gdp_per_capita                  : num  1804 1804 1804 1804 1804 ...
## $ extreme_poverty                 : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ cardiovasc_death_rate           : num  597 597 597 597 597 ...
## $ diabetes_prevalence              : num  9.59 9.59 9.59 9.59 9.59 9.59 9.59 9.59 9.59 ...
## $ female_smokers                   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ male_smokers                     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ handwashing_facilities           : num  37.7 37.7 37.7 37.7 37.7 ...
## $ hospital_beds_per_thousand       : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ life_expectancy                  : num  64.8 64.8 64.8 64.8 64.8 ...
## $ human_development_index          : num  0.511 0.511 0.511 0.511 0.511 0.511 0.511 0.511 0.511

```

Exploratory Data Analysis

A. Highly correlated variables to Covid 19 cases

In this correlogram, we see the most correlated variables to new covid cases. We maintained variables that have a strong correlation with the new covid cases variable, that is, we only kept variables with a correlation greater than the absolute of 0.8. These variables were new_cases, new_cases_smoothed, icu_patients, new_deaths, new_deaths_smoothed, weekly_hosp_admissions and hosp_patients, total_deaths, total_cases, new_tests and new_tests_smoothed.



B. Relationship between GDP per capita and life expectancy

In the diagram below we show the relationship between GDP per capita and life expectancy across the world. Countries with lower GDP per capita, mostly found in Africa, have a lower life expectancy, while countries with a higher GDP per capita, like those found in Europe, have a higher life expectancy. However, it is interesting to note that more prosperous countries with higher GDP per capita and life expectancy suffered more covid cases than less advanced countries with lower GDP per capita and life expectancy.



C. Relationship between Handwashing Facilities and Extreme Poverty

In this diagram we will explore the relationship between handwashing facilities and extreme poverty. Countries that are extremely poor like those in Africa, have the least amount of handwashing facilities, while those in richer countries have more hand washing facilities. There is an inverse relationship between extreme poverty and handwashing facilities as illustrated by the regression line. Paradoxically, countries that are well of and have a lot of hand washing facilities seem to suffer higher positive cases of the coronavirus, countries in Asia and South America stand out in this regard. On the other hand, countries in African, with fewer hand washing facilities, had a lower positive case count.

