

Research report SARS-CoV-2

Political regime and covid-19 impact

Thibault Herpe

19/05/2020

Abstract

The first part of this study focuses on the analysis of potential correlations between the type of political regime in a country and the impact of the SARS-CoV-2 on that country. A second part explores the similarities and differences in terms of impacts of the crisis between countries with relatively identical medical characteristics.

Contents

1	Collecting Data	2
	Sherbrook University	3
	European Union Open Data Portal	4
	World Bank	4
	World Health Organisation (WHO)	5
	Lockdown Date	7
2	Data Wrangling	8
3	Data Analysis	9
	Correlation tests and p-value analysis	10
	Country segmentation by k-mean algorithm	18
	Covid-19 impact statistics by cluster	22
	Covid-19 impact statistics for cluster 1	23
4	Conclusion	31

Notice : Figures currently reported are probably underestimated and probably manipulated by some governments. However, the figures reported by the covid-19 of each country are subject to change and to be corrected by the scientific community. This study will therefore be updated in order to follow the evolution of the results.

Chapter 1

Collecting Data

Country information is collected from a variety of reliable sources.

Sherbrook University

The Université de Sherbrooke is a large public French-language university in Quebec, Canada with campuses located in Sherbrooke and Longueuil, a suburb of Montreal approximately 130 km west of Sherbrooke. It is one of two universities in the Estrie region of Quebec, and the only French-language university for the region.

Democracy Index (EIU.DEMO.GLOBAL)

“The Democracy Index is an index compiled by the Economist Intelligence Unit (EIU), a UK-based company. It intends to measure the state of democracy in 167 countries, of which 166 are sovereign states and 164 are UN member states.”

find out more information about this index by following this link: https://en.wikipedia.org/wiki/Democracy_Index

Press Freedom Index (PF.LIB.PRESS.RSF.IN)

“The Press Freedom Index is an annual ranking of countries compiled and published by Reporters Without Borders based upon the organisation’s own assessment of the countries’ press freedom records in the previous year. It intends to reflect the degree of freedom that journalists, news organisations, and netizens have in each country, and the efforts made by authorities to respect this freedom. Reporters Without Borders is careful to note that the index only deals with press freedom and does not measure the quality of journalism nor does it look at human rights violations in general.”

find out more information about this index by following this link: https://en.wikipedia.org/wiki/Press_Freedom_Index

European Union Open Data Portal

“The EU Open Data Portal is the point of access to public data published by the EU institutions, agencies and other bodies. Information can be used and reused for commercial or non-commercial purposes. The portal is a key instrument of the EU open data strategy.”

find out more information about this portal by following this link: https://en.wikipedia.org/wiki/EU_Open_Data_Portal

Access to the portal: <https://data.europa.eu/euodp/en/home>

Covid 19 Cases

Data concerning covid-19 collected by the European Union for each country of the world are accessible through this link: <https://opendata.ecdc.europa.eu/covid19/casedistribution/csv>

World Bank

“The World Bank is an international financial institution that provides loans and grants to the governments of poorer countries for the purpose of pursuing capital projects. It comprises two institutions: the International Bank for Reconstruction and Development (IBRD), and the International Development Association (IDA). The World Bank is a component of the World Bank Group. The World Bank’s most recent stated goal is the reduction of poverty.”

find out more information about the World Bank by following this link: https://en.wikipedia.org/wiki/World_Bank

Per Capita Income Index (NY.GNP.PCAP.CD)

“Per capita income (PCI) or average income measures the average income earned per person in a given area (city, region, country, etc.) in a specified year. It is calculated by dividing the area’s total income by its total population.”

find out more information about the Per capita income index by following this link: https://en.wikipedia.org/wiki/Per_capita_income

World Governance Indicators

“Based on a long-standing research program of the World Bank, the Worldwide Governance Indicators capture six key dimensions of governance (Voice & Accountability, Political Stability and Lack of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption) between 1996 and present. They measure the quality of governance in over 200 countries, based on close to 40 data sources produced by over 30 organizations worldwide and are updated annually since 2002.”

find out more information about World Governance Indicators by following this link: https://en.wikipedia.org/wiki/Worldwide_Governance_Indicators

Access to the data by following this link : <https://info.worldbank.org/governance/wgi/>

World Health Organisation (WHO)

“The World Health Organization (WHO) is a specialized agency of the United Nations responsible for international public health.[1] The WHO Constitution, which establishes the agency’s governing structure and principles, states its main objective as”the attainment by all peoples of the highest possible level of health.”

find out more information about the World Health Organisation by following this link: https://en.wikipedia.org/wiki/World_Health_Organization

Universal Health Coverage Index

Universal Health Coverage refers to access to comprehensive health care and public health for all residents of a country or geographical or political region without regard to their economic capacity or personal situation.

"UHC means that all individuals and communities receive the health services they need without suffering financial hardship. It includes the full spectrum of essential, quality health services, from health promotion to prevention, treatment, rehabilitation, and palliative care.

UHC enables everyone to access the services that address the most significant causes of disease and death, and ensures that the quality of those services is good enough to improve the health of the people who receive them."

Find out more information about this index by following this link: [https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-\(uhc\)](https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc))

Acute car hospital beds per 100 000 inhabitants (E992713.T)

"Acute care beds in hospitals are hospital beds that are available for curative care :

Inclusion - Beds accommodating patients where the principal clinical intent is to do one or more of the following: manage labour (obstetrics), cure non-mental illness or provide definitive treatment of injury, perform surgery, relieve symptoms of non-mental illness or injury (excluding palliative care), reduce severity of non-mental illness or injury, protect against exacerbation and/or complication of non-mental illness and/or injury which could threaten life or normal functions, perform diagnostic or therapeutic procedures

Exclusion - Beds allocated for other functions of care (such as psychiatric care, rehabilitation, long-term care and palliative care) - Beds in mental health and substance abuse hospitals (HP.1.2) - Beds for rehabilitation (HC.2) - Beds for palliative care"

Find out more information about this indicator by following this link: https://gateway.euro.who.int/en/indicators/hfa_478-5060-acute-care-hospital-beds-per-100-000/

Lockdown Date

Lockdown Date indicator

“A lockdown is a requirement for people to stay where they are, usually due to specific risks to themselves or to others if they can move freely. The term “stay-at-home” is often used for lockdowns that affect an area, rather than specific locations.” “A stay-at-home order or a movement control order (more common in Southeast Asia) is an order from a government authority to restrict movements of a population as a mass quarantine strategy for suppressing, or mitigating, an epidemic, or pandemic, by ordering residents to stay home except for essential tasks or to work in essential businesses.”

Find out more information about this term by following this link: <https://en.wikipedia.org/wiki/Lockdown>

Chapter 2

Data Wrangling

this part is dedicated to data manipulation. In order to keep this report readable, the programming code is hidden. Access the code on the programming version of this report.

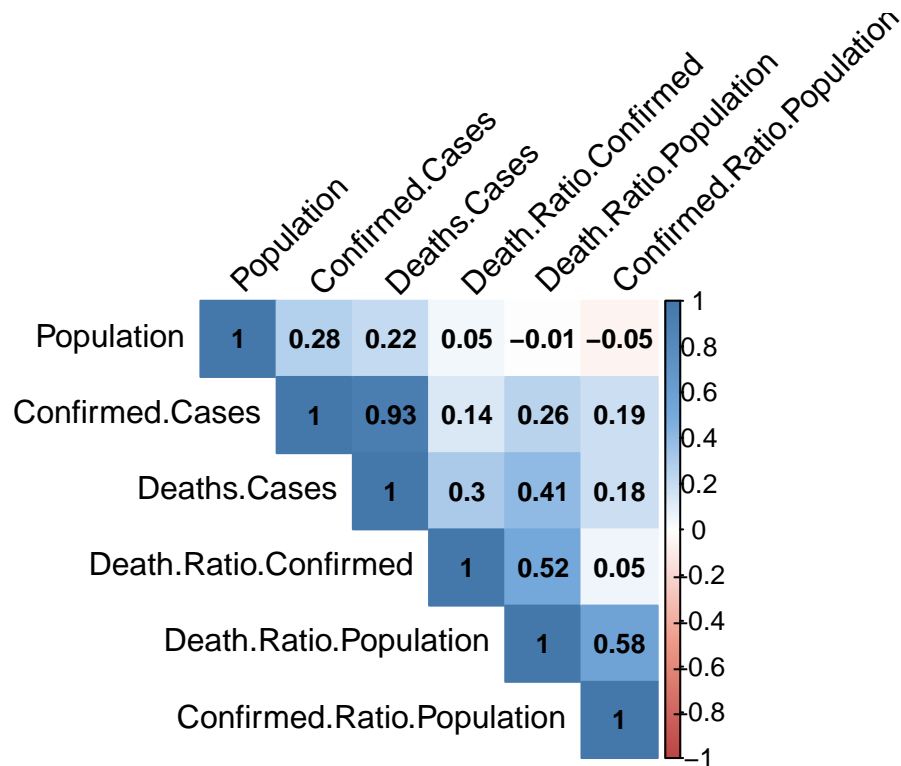
Chapter 3

Data Analysis

First stage of data collection having been completed, an analysis of the data will be carried out in this second part. The dataset contains information for 196 countries on covid-19 such as : confirmed cases number, deaths cases number and ratio like dead person among confirmed person, confirmed cases among total population, Death cases among total population but also the total population of the country and their ISO code and names. A first analysis of correlation between indicators will be carried out, completed by a correlation analysis under p-value test in order to confirm, or not, the statistical validity of this correlation. The objective of the correlation analysis under p-value test is to determine whether the correlation obtained is significant, and therefore exists statistically.

Correlation tests and p-value analysis

correlation test on dataset covid-19



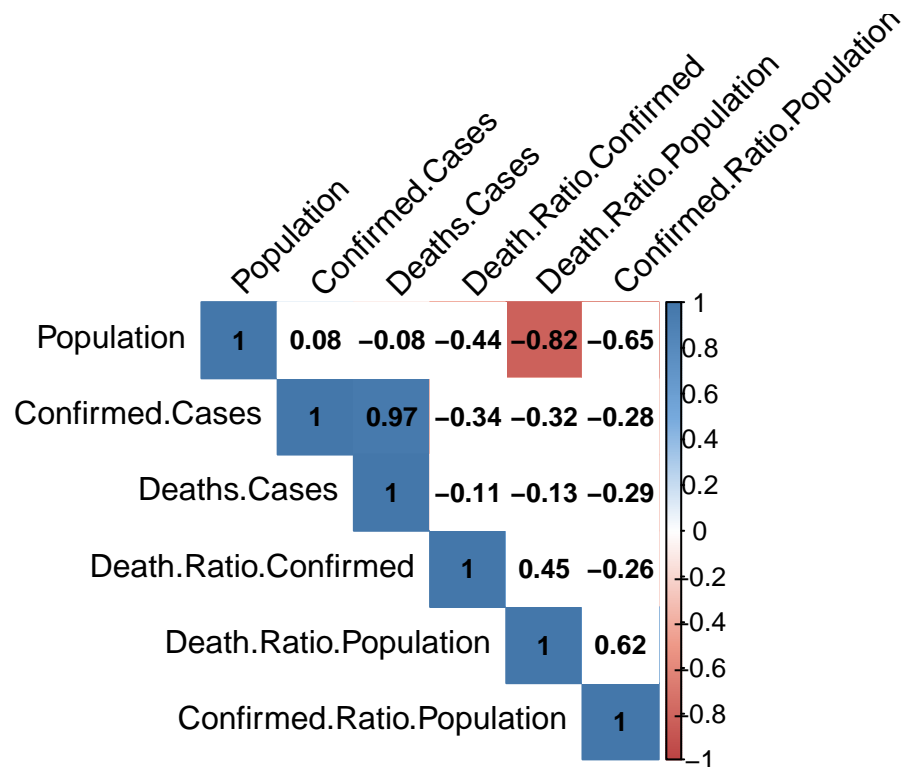
This first correlation matrix shows us correlations between variables. Every intersection between a line and a columns represent the correlation between those two variables. The more the blue is intense, the more the correlation is strong. Numbers in colored squares represent the correlation. If, and only if, this number is above 0.5 then the correlation can be interpreted as strong.

By performing this correlation test on the dataset of covid-19 data, three main correlations appear to be strong. These are the correlation linking confirmed cases to the number of deaths, the correlation linking the number of deaths among confirmed cases to the number of deaths in the population, and the correlation linking the number of deaths in the population to the number of confirmed cases in the population.

Others relations can be more surprising as the population and confirmed cases, or deaths cases, that are not related at all. This don't come as a matter of facts as it can be pictured that the more your population is big as a country, the more cases you should have.

correlation test under p-value on dataset covid-19

The objective of the correlation analysis under p-value test is to determine whether the correlation obtained is significant, and therefore exists statistically. To run this final test, the p-value chosen was 0.05.

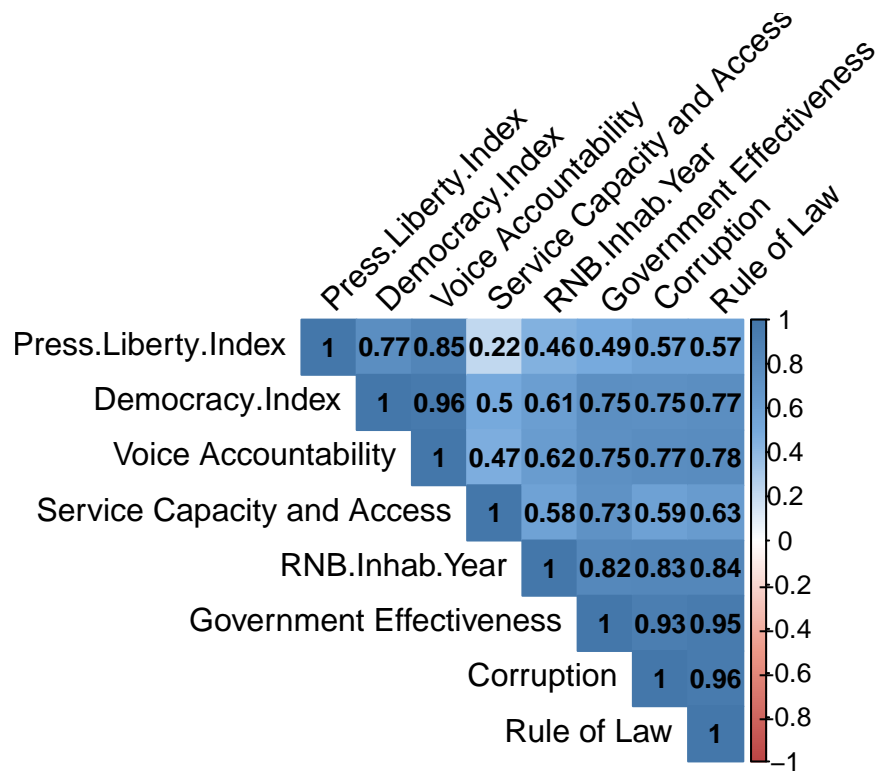


After performing a correlation test with p-value to determine whether the correlations obtained are significant or not, the correlations that still exist appear different. Indeed, only the correlation between the number of confirmed cases and the number of deaths appears to be significant. A surprising negative correlation appears between the size of the country's population and its number of deaths relative to the population. Indeed, this correlation did not appear on the simple correlation test.

correlation test on dataset gathering country's political system characteristics

the same correlation tests are carried out on the dataset gathering, for 147 countries, the data concerning their respective wealth as well as their style of governance, such as for example : Press Liberty index, Democracy index, Per Capita Income index, Corruption index, Rule of Law index, Government Effectiveness index, Voice and Accountability index and Service Capacity and Acces index.

A first analysis of correlation between indicators will be carried out, completed by a correlation analysis under p-value test in order to confirm, or not, the statistical validity of this correlation.



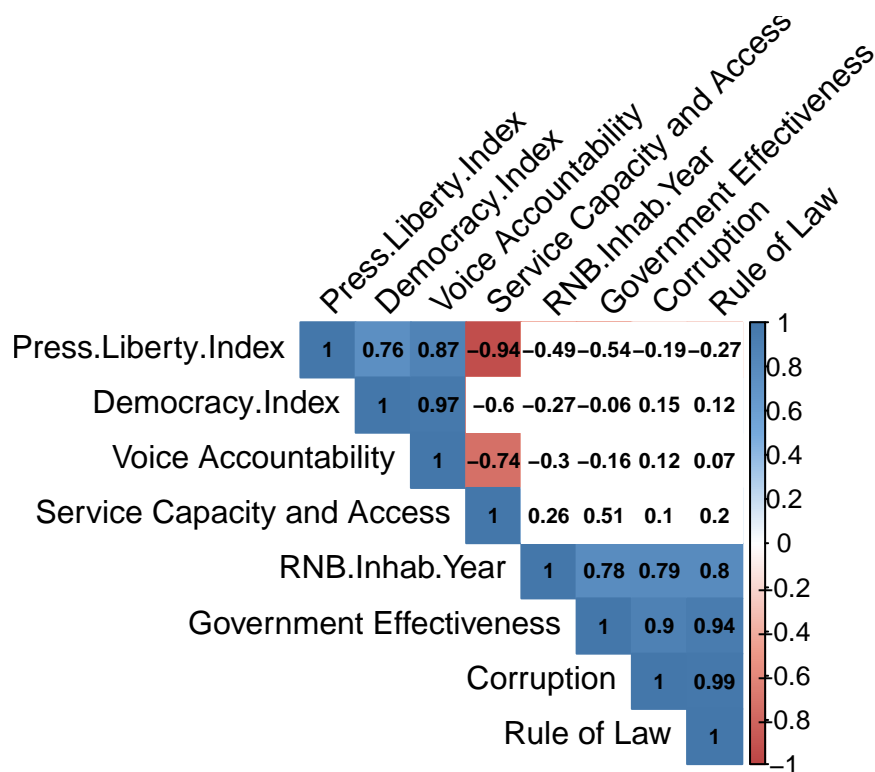
For instance, the correlation between the Press.Liberty.Index and the Service Capacity and Access indicator is the lowest in this matrix, which makes it impossible to establish links between these two indicators at this stage.

On the other hand, the correlation between the Corruption and Rule of Law indicator is the strongest in this matrix, which means that the variation of one of the indicators is highly correlated with the other one. Even if strong correlations can be found, this first matrix correlation can't tell if those correlations are statistically confirmed, in other word that those correlations still exist after performing a p-value test on it.

correlation test under p-value on dataset gathering country's political system characteristics

A p-value test is therefore carried out to analyse the statistical validity of these correlations.

To run this final test, the p-value chosen was 0.05.



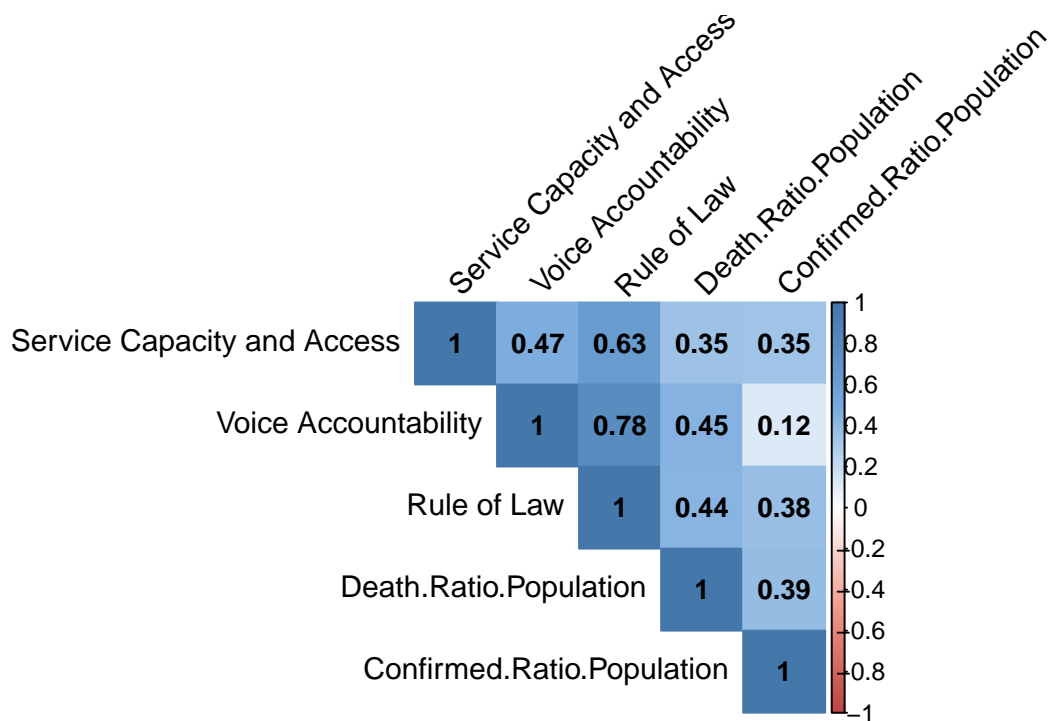
By doing this test, it clearly appears that a lot of correlations precedently found, does not existe anymore. Every square still colored shows significant correlations. The blue ones a positive correlation, and red ones a negative one.

By reading the matrix, it can be seen that some indicators best describe others, like the Rule of Law index which is strongly, and significantly, correlated to the RNB per Inhabitant per year, the Government effectiveness, the Corruption index. Therefore, if it's only taken to perform test, it will well represents effectively the evolution of other ones. An interesting negative correlation is the one linking the Press Liberty Index to the Service Capacity and Access indicator. It is difficult to interpret this result, but several possibilities emerge. The first and most desirable would be that countries with the least free press would have better hospital capacity and easier access to care. The second possibility could be that the official figures for available places and access to healthcare systems are not representative of reality. The Voice Accountability indicator confirms the observation made earlier, without being able to provide any further details.

Taking these observations into account, only three indicators are selected for further analysis. The Rule of Law, Service Capacity and Access indicator and the Press.Liberty.Index will be selected. Their selection is based on the analysis of their statistical correlations with the others. Indeed, in order to simplify the analyses as much as possible, a reduction in the number of dimensions of the indicator matrices must be made.

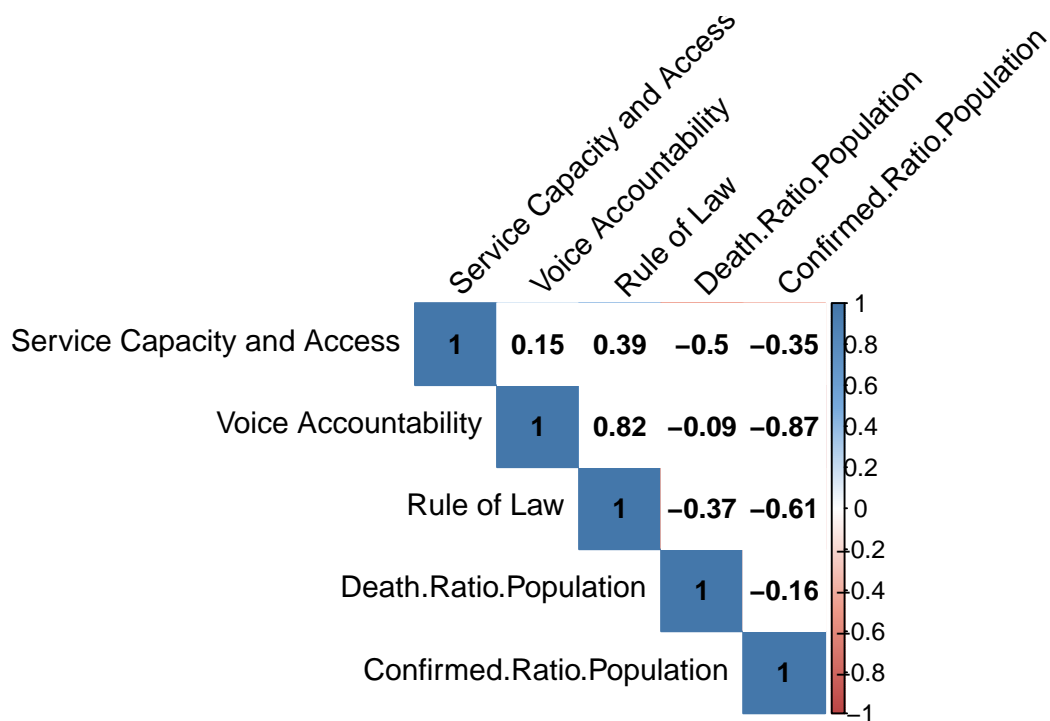
correlation test on dataset regrouping covid-19 figures and country's political system

For this last correlation test, a new dataset is created in order to group together the tree main indicators of the political regime of the previously selected countries, and the two main indicators of the impact of covid-19 on the populations of the affected countries.



As a result of the test, only two correlations appear to be strong. These correlations are not new, as they correspond to those previously obtained. The main result of this test is that no correlation can be established between the indicators showing the impact of covid-19 on populations and the type of political regime in place within each country. However, a p-value test should be performed to confirm, or not, this result.

correlation test under p-value on dataset regrouping covid-19 figures and country's political system



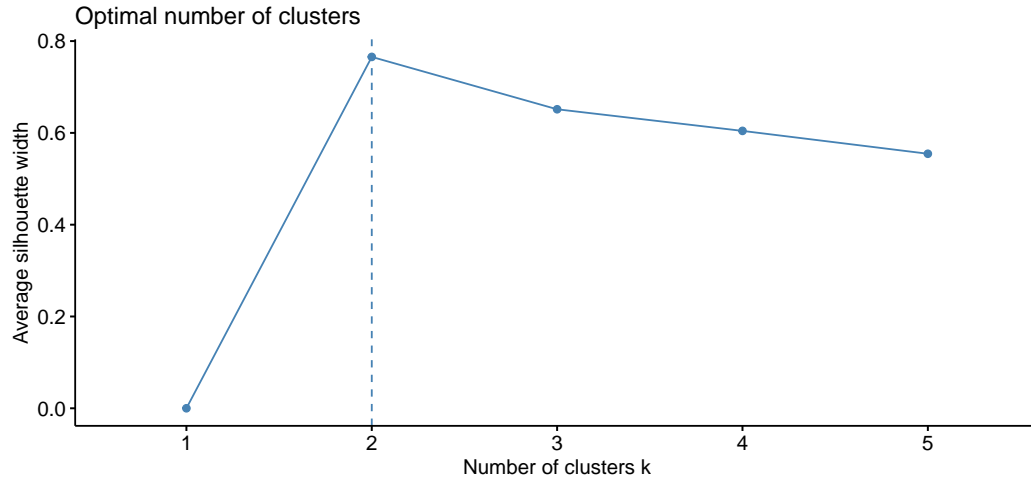
Unsurprisingly, no significant correlations appear between the type of political regime in each country and the impact of covid-19. Currently, no link between political regime and the impact of covid-19 can be established, in fact, the analysis of the correlations of the different indicators does not allow us to affirm such a hypothesis. However, it should be kept in mind that countries with a high level of press repression are likely to correct, if not manipulate, the official figures of the impact of covid-19 on their population.

As the covid-19 figures are likely to evolve, it will be interesting to follow the evolution of these correlations over time.

Country segmentation by k-mean algorithm

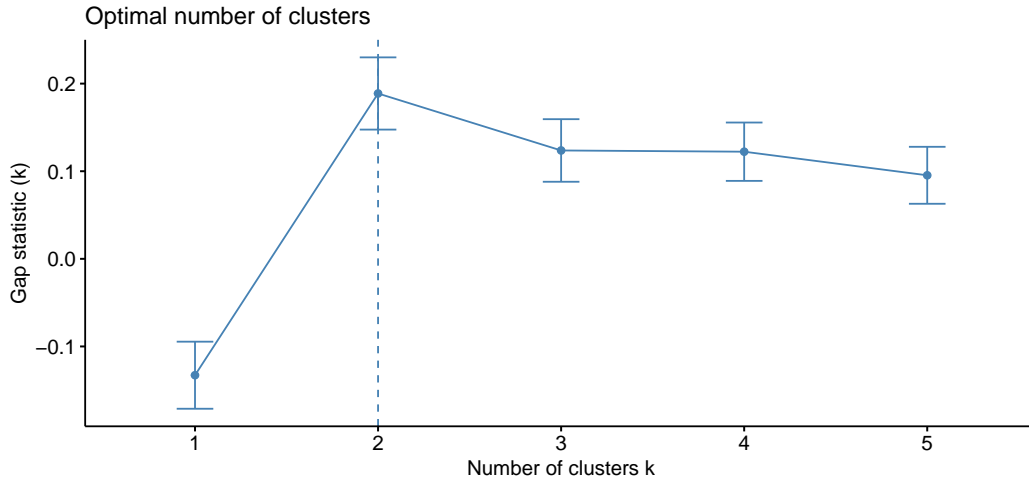
After performing these correlation tests, a K-means algorithm is performed in order to determine groups of countries that are relatively identical in terms of hospital facilities, and thus to evaluate to which extent the covid-19 impacts these countries in the same manner. the goal is to create groups of country based on the evaluation of their : Service Capacity and Access index, Rule of law index and Voice Accountability index. Before performing the kmeans algorithm, it is needed to by determine the optimal number of clusters. in order to avoid too vague clustering, it will be important not to fragment the dataset into more than 3 groups. By doing so, the subgroups should contain more than the 30 individuals needed to perform statistically reliable tests. Having 148 countries, the ideal groups would contain in the 45 to 50 countries.

The optimal number of clusters is determined prior to the execution of the kmeans agorithm. In order to determine this optimal number, the silhouette clustering test is performed. For each observation, the silhouette coefficient corresponds to the difference between the average distance with observations from the same group (cohesion) and the average distance with observations from other neighbouring groups (separation). If this difference is negative, the observation is on average closer to the neighbouring group than to its own: it is therefore misclassified. Conversely, if this difference is positive, the point is on average closer to its group than to the neighbouring group: it is therefore well ranked.



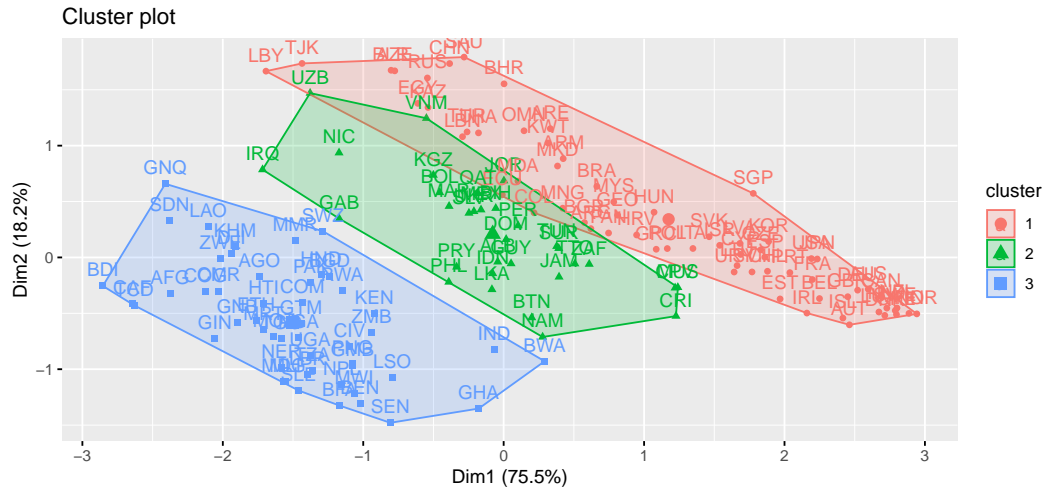
This silhouette clustering algorithm shows that 2 cluster is the optimal number. However, in order to avoid a too imprecise partitioning, the dataset will be subdivided into 3 groups which corresponds to the second optimal partitioning.

3 groups of countries will therefore be determined by the K-mean algorithm, but it is necessary to ensure the quality of the resulting clusters, to do a Gap Statistic test is performed. The Gap Statistic is a method developed by R. Tibshirani, G. Walther, and T. Hastie from the Stanford University in 2001. they describe their method in the following way : “The technique uses the output of any clustering algorithm, comparing the change in within-cluster dispersion with that expected under an appropriate reference null distribution”. This means that the more the gap statistic number for each k is high, the better this number of clusters (k) represent a good clustering.



If 2 clusters still represents the optimal number of clusters, as the Silhouette method already indicated, 3 being the second best option, it is possible to choose 3 as the optimal number of clusters. This method is particularly sensitive to cluster overlapping which can explain why 3 is not the optimal number of clusters. Confirmation will be made by seeing groups of countries during the performance of the k-means algorithm. keeping in mind that 30 countries for each clusters is a minimum to perform statistics, 5 in the maximum number of clusters tested. Indeed, if 5 clusters are created they will contain less than 30 countries in each one and thus, statistical tests cannot be performed.

Once all the parameters of the k-mean algorithm have been determined, it should be applied to the dataset containing the variables selected for each country that are : : Voice Accountability index, Rule of Law index and Service Capacity and Access index.



3 groups of 66, 32 and 50 countries sharing common characteristics according to the selected variables are obtained. As previously suspected, overlapping cluster can be observed which means that this clustering cannot be the best one. With that said, this overlapping is due to only 3 countries which does not represent a threat for the overall clustering.

Covid-19 impact statistics by cluster

Once 3 clusters of countries have been obtained, analyses will be carried out to determine the impact of covid-19 on the countries belonging to the same cluster.

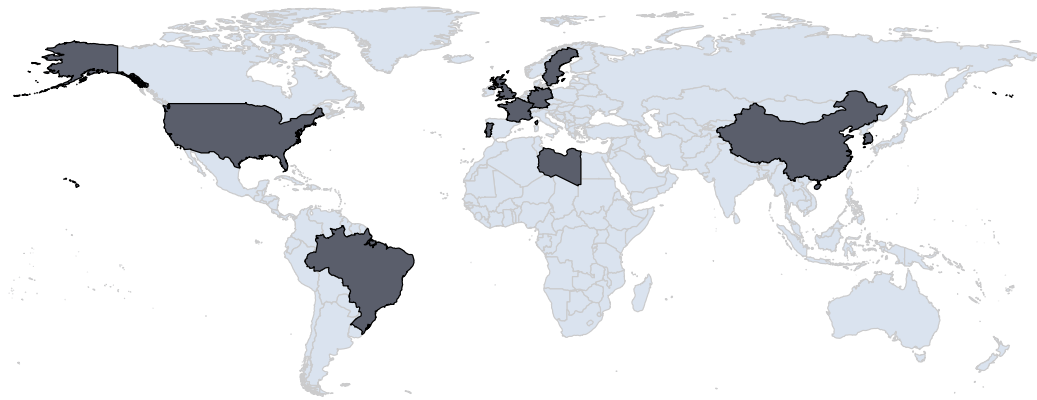
Calculations of the averages of each variable within the same cluster are performed and grouped in the table above.

	Cluster 1	Cluster 2	Cluster 3
Population	51361720	29873995	56331209
Confirmed Cases	78561.00	15702.19	8757.68
Deaths Cases	5123.9091	663.6875	208.9000
Deaths per millions inhabitants	105.810321	16.290250	3.421643
Confirmed cases per millions inhabitants	2168.1808	1191.5840	224.1798
Death ratio per hundred confirmed cases	4.431692	2.798265	2.378104

Some differences are noticeable, such as the number of confirmed cases, which varies greatly, or the number of deaths per million inhabitants. A factor 10 appears between each cluster concerning the number of deaths per million inhabitants, which represents a very important variation. This can have several origins, such as the lack of detection of cases due to lack of testing, a population that is relatively untouched because it is either totally confined or far from global communication channels. The cause can also be very different and take a political turn with figures deliberately modified so as not to appear catastrophic.

Covid-19 impact statistics for cluster 1

Further analysis will be conducted on Cluster 1, which includes the countries of the European Union, but also the largest countries in the world in terms of geopolitics, without forgetting poor countries, sometimes at war, but with a relatively efficient health system. Cluster 1 is composed of Brazil, China, France, Germany, Libya, Portugal, Republic of Korea, Sweden, the United Kingdom and the United States.



As can be seen on this map, this cluster is relatively representative of all the main country typologies, and represents almost all continents. This is very important in order not to distort the analyses. Indeed, if the observations represented only one typology of State or emanated from only one continent, this would lead to generalize specific results to a given situation. For example, selecting only European countries does not allow us to visualize the dynamics of this pandemic on a global scale.

Thus the selection consists of :

- Brazil: located in South America, Brazil is a poor country whose management of the crisis is strongly debated at the international and national level with a resignation of the politicians in charge of the management of the health system. In addition, Brazil has a large number of urban slums, which compromises the effectiveness of lockdown measures. Finally, the country's health system was in serious trouble even before crisis's emergence.

- China: Located in Asia, China is a relatively rich country is the epicenter of the crisis, the local government of Wuhan tried to hide the emergence of this epidemic and therefore slowed down the establishment of lockdown measures by the central government. China has very good figures which seem to be more related to the strong control of the press by the government than to the real effectiveness of the lockdown measures taken.
- France: Located in Western Europe, France is a wealthy country often cited in the past for its very efficient healthcare system. In 2000, France was ranked first in the WHO's world ranking of health systems, but it has now fallen to 15th place due to the savings made in the health system and its inadequate management. By taking action relatively late, France has avoided the collapse of its health system, but has some of the worst figures in the world. This is partly due to the fairly accurate counting it does compared to other countries.
- Germany: Located in Europe, Germany is a rich country and seems to be going through the crisis well as well as being seen as an example by having a low mortality rate as well as a non-total and regionalized lockdown. Germany quickly conducted testing campaigns that enabled it to contain as far as possible the outbreaks of the pandemic on its territory.
- Libya: located in Africa, Libya is a country at war which had a relatively functional health system before the health crisis. However the figures for the impact of the pandemic on the country seem to be inaccurate and underestimated due to a lack of testing. For the part of the country currently controlled by the central government, however, measures were taken very early on which probably helped to limit the spread of the virus.
- Portugal: Located in Southern Europe, Portugal is a country in economic difficulty following the financial crisis of 2008. This has resulted in massive cost-savings at the expense of the healthcare system, which has suffered heavy staff cuts. However, Portugal, relatively far from the main global communication routes, took advantage of the additional time before the first patient appeared on its territory to take strong and rapid measures. This is now reflected in the excellent figures that the country is reporting.

It is an illustration of the success of a good management of the health crisis and proves that the quality of the health system alone is not enough to get through a pandemic with serenity.

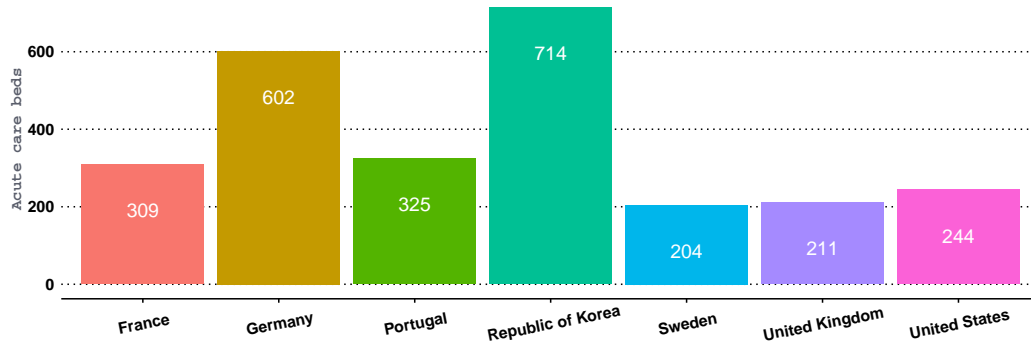
- The Republic of Korea: Located in Asia, the Republic of Korea is a wealthy country accustomed to the management of epidemics. In 2015, South Korea faced the MERS virus which led it to maintain very high preventive measures and response capacities in case of an epidemic or pandemic. The population is also aware of the cleanliness of public spaces and numerous are those who protect themselves on a daily basis by fear of illness. This has led the country to be the world's number one example in the management of this crisis, and has allowed the country to avoid lockdown sparing it the economic consequences it induces.
- Sweden: Located in Northern Europe, Sweden is a very wealthy country that has made the radical choice not to confine itself in any way. While it does not have more catastrophic figures than France or the United Kingdom, it does have more catastrophic figures compared to its Nordic neighbors such as Norway, Denmark, and Finland. Indeed, by being far removed from the world's main trade routes, Sweden could have avoided many of the deaths it recorded. This is directly due to the strategy it has put in place which is based on collective immunity, in total opposition to its neighbours.
- United Kingdom: Located in Western Europe, the United Kingdom is a wealthy country that took action very late and is now feeling the full effects of the savings that have been made in the healthcare system for many years. The Prime Minister of the United Kingdom initially opted for the Swedish strategy based on collective immunity, before changing strategy and opting for population lockdown. This led to confusion and a lack of preparation in a country that was spared at the beginning of the crisis. Today the United Kingdom has the 4th worst figures among developed countries and the worst of this selection.

- United States of America: Located in North America, the United States is a wealthy country with a relatively good healthcare system. The United States has been slow to confine its population and some states have chosen not to confine. The mortality figures from covid-19 classifies the United States as one of the worst countries in the world even though a massive testing campaign is being performed. However, the campaign has been relatively successful in containing the epidemic, despite the total lack of coordination by the federal government, which, like Brazil, is experiencing deep divisions between the specialists in charge of managing the health crisis and the president.

Acute care bed facilities

An initial analysis of the number of acute care beds available, among the countries making public their data within the selection of the ten countries, allows to perceive marked disparities in hospital equipment in spite of the relative homogeneity in terms of hospital infrastructure of this cluster.

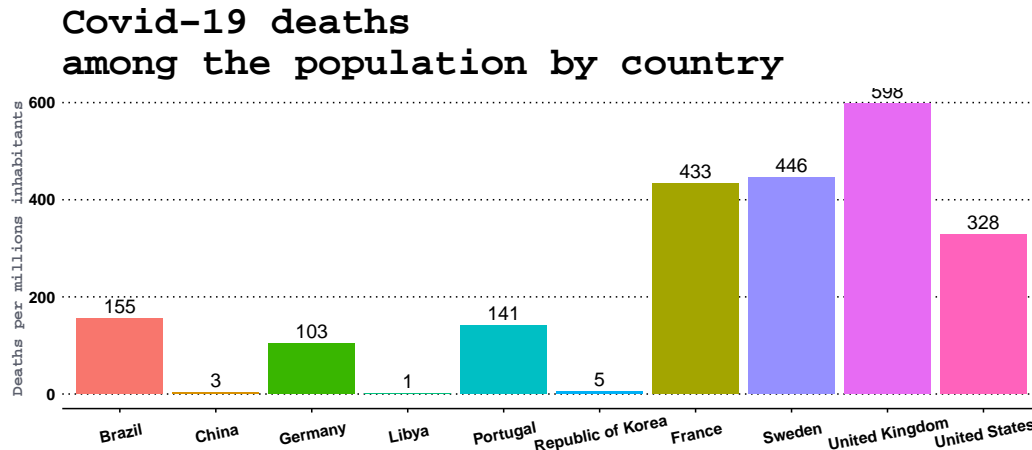
**Acute care beds in 2017
per 100,000 inhabitants**



This chart shows that the number of intensive care beds available per 100,000 population in 2017 varies significantly, from 714 in the Republic of Korea to only 204 in Sweden. Two countries, France and Portugal, appear in the average of this selection with 309 and 325 beds available per 100,000 inhabitants respectively. This variability is interesting because it depicts a very different initial situation depending on the country, and therefore a response strategy that will be impacted by it.

Mortality rate among population

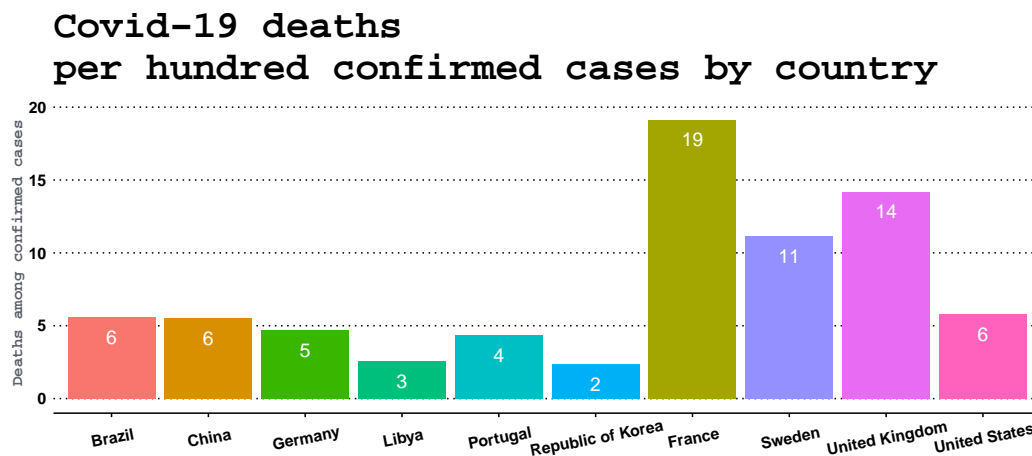
Analysis of mortality rates among the population shows a clear distinction between two groups of countries.



A first group consisting of France, Sweden, the United Kingdom and the United States is easily identifiable. Indeed, the figures obtained by these countries are clearly superior to the others. Thus the United Kingdom records about 550 deaths per million inhabitants, which is more than five times the figure for Germany. As stated in the introduction to this chapter, the Republic of Korea, with the exception of Libya, posted exceptionally low figures, confirming its exemplary handling of the crisis. If for Germany and the Republic of Korea the figures seem to be right, this is not the case for other countries which are in very different situations such as Brazil which is still at the beginning of the epidemic on its territory or China which, although the cradle of the epidemic, seems to have frozen its figures several months ago.

Mortality rate among confirmed cases

the following analysis concerns the number of deaths per hundred confirmed cases. This number is directly impacted by the testing policy of countries. Indeed, if the country can practice only relatively few tests, this will induce high mortality rate among those who test positive. Conversely, if the country has the capacity to perform mass testing like this, it improves its rate.

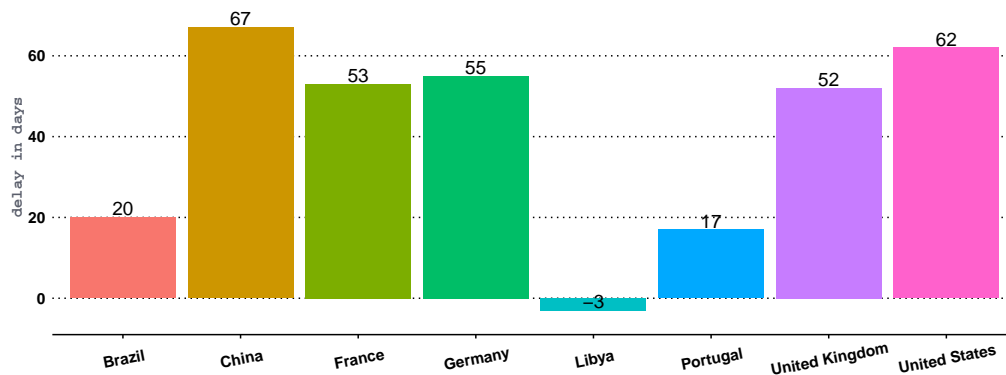


France, which has almost no testing among its population, has a much worse ratio and has the worst selection ratio, with about 20 deaths per hundred tested. In fact, as tests are mainly carried out only on patients admitted to hospital in an already deteriorated health condition, this leads to a higher mortality rate in this ratio. Conversely, the United States, which screens its population on a massive scale, has a very favourable ratio, on a par with China or Brazil and barely higher than that of Germany. However, the United States has the fourth worst ratio in terms of deaths among the population, which shows the caution with which these ratios must be interpreted.

Delay before lockdown

The last but not least graph looks at the time lag between the first confirmed case of covid-19 on the territory of each country and the beginning, or not, of the lockdown measures.

Days between first case and country's lockdown



As shown in this graph, some countries, such as Portugal, have reacted very quickly to the presence of the virus on their territories by establishing a generalized containment of the population in only 17 days. Other countries such as China and the United States took very late measures with 67 and 62 days respectively. France is in the average of this selection by confining its population within 53 days, 3 days faster than its neighbour Germany. However, it should not be forgotten that countries such as Portugal and the United Kingdom were hit much later than others, although the United Kingdom did not take advantage of this to prepare for the pandemic. Finally, some specific cases are to be commented like that of the Republic of Korea and Sweden which chose not to confine their population at all, or Libya which took restrictive measures, more than confinement, 3 days before the first case of covid-19 on its territory.

Chapter 4

Conclusion

This study, conducted in collaboration with the SKEMA Business School's Data Science Research Laboratory, demonstrated that at present, no links can be established between authoritarian or democratic political regimes and the impact of covid-19 on the population of countries. Furthermore, the study of differences in existing hospital infrastructure within the same cluster of countries does not predict the impact of the coronavirus on the population of the countries. In conclusion, if the type of regime is not at the root of the good or bad results obtained in the fight against covid-19, and if the ability to cope with the pandemic in each country's medical system does not play a role within the cluster of countries with the best healthcare resources in the world, only government actions can explain such variability in the figures recorded in the context of this pandemic.

Once again, the figures currently reported are probably underestimated and probably manipulated by some governments. However, the figures reported by the covid-19 of each country are subject to change and to be corrected by the scientific community. This study will therefore be updated in order to follow the evolution of the results.