

# Gravity of COVID-19 cases in regions of Brazil

Author: Marcello Nery

## Introduction

The worldwide spread of Coronavirus disease (COVID-19) brought a revolution in the lifestyle of everyone. Political decisions have fundamental influence on the situation each country faces. In Brazil, the political scenario already was tense, with the controversial president Jair Bolsonaro in command, making ideological decisions without caring about the well-being of the population. The greatest example the country faces now is his denial about the gravity of the Coronavirus pandemic, refusing to adopt political measures to hold down the increase in the number of cases in Brazil, encouraging non-essential economic activities to stay occurring normally and speaking against governors of states and cities who are doing what is possible for avoiding the worst possible case to happen in their cities and states. The situation of Brazil has rapidly become one of the gravest in the world relative to the pandemic, either economically or in the public health sector.

The objective of this work is to analyze the gravity of the situation for different cities in Brazil in each region of the country, independently and present it with maps of different regions and marking cities with points of different colors, according to the gravity of its situation. For that, I use the K-means clustering algorithm, forming three clusters for each region, each one corresponding to one level of intensity of the number of cases or number of deaths. This study can be useful for politicians to see how the disease is spreading, where it is graver and what politics may be adopted to slow down the number of cases in cities where it is already grave, and to keep the low numbers for cities where the disease rate is still low.

## Data

The data used for this work was extracted from Brazil's official informative website about Coronavirus:

<https://covid.saude.gov.br/>

The data file is updated every working day and for this work I used the one corresponding to 07/01/2020, which means that the data contains numbers of the cases in Brazil from 03/27/2020 until 07/01/2020.

The data file contains the following columns:

- **regiao**: Brazilian region;
- **estado**: The state of the corresponding data;
- **municipio**: City name;
- **coduf**: Code of the state;
- **codmun**: Code of the city;
- **codRegiaoSaude**: Code of the Health region;
- **nomeRegiaoSaude**: Name of the Health region;
- **data**: Date;
- **semanaEpi**: Number of the week;

- [populacaoTCU2019](#): Population number, according to IBGE (available to TCU), in 2019
- [casosAcumulado](#): Number of accumulated cases;
- [casosNovos](#): Number of new cases;
- [obitosAcumulado](#): Number of accumulated deaths;
- [obitosNovos](#): Number of new deaths;
- [RecuperadosNovos](#): Number of recoveries;
- [emAcompanhamentoNovos](#): Number of new followed-up cases;
- [interior/metropolitana](#): Whether the city is from metropolitan or rural area

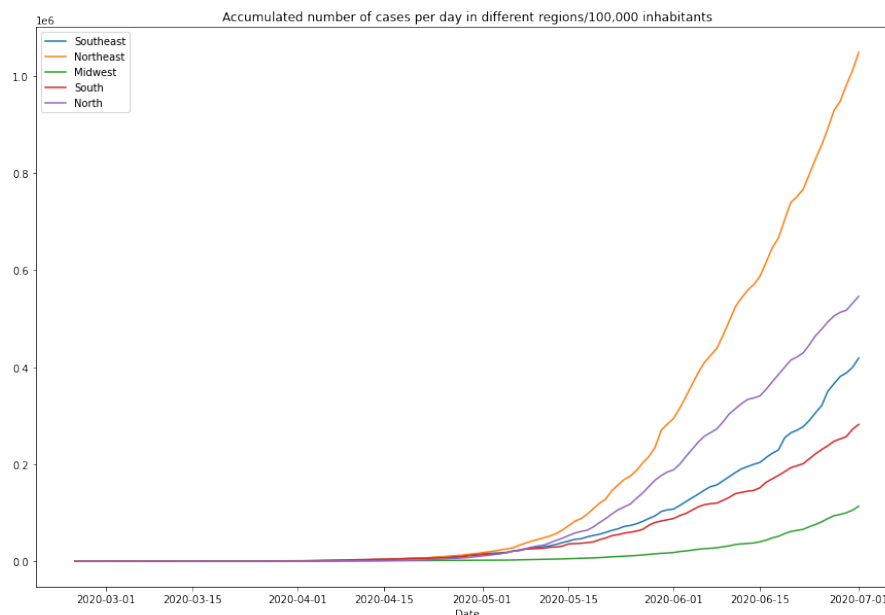
## Methodology

In this work, I make an analysis for each region in Brazil and make classifications of the cities according to their historical accumulated numbers of cases and accumulated numbers of deaths. Before applying the method, I do a small exploratory analysis of the data.

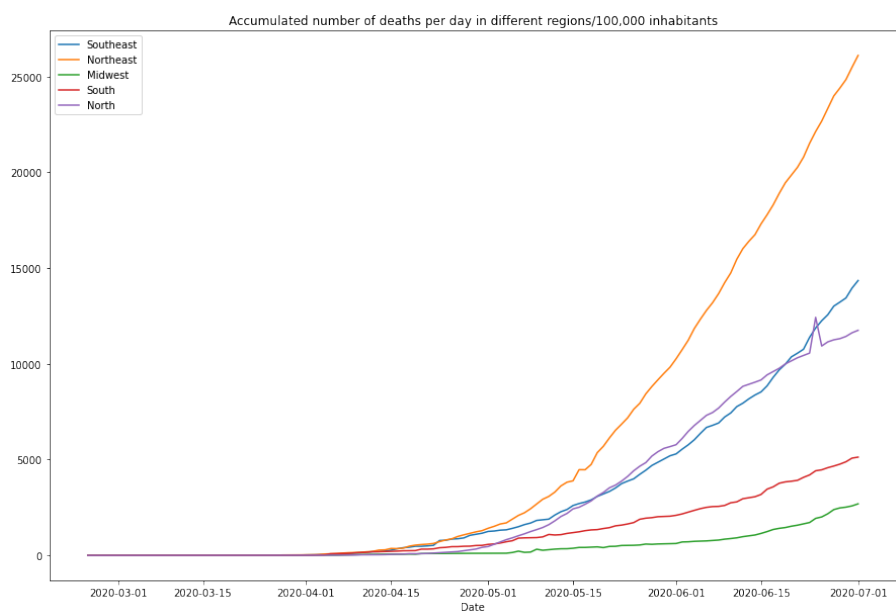
The original data contains accumulated numbers of cases, deaths and daily number of cases and deaths, in absolute magnitude. The first relevant thing I do is to replace these numbers by their corresponding rates/100,000 inhabitants. This puts the number of different cities in the same measure, independently of their population.

Next, I analyze what cities the data contains. The original data contains more city codes than city names, which could suggest that there were cities with wrong names. What actually happens is that there are many cities with the same name in the country, but located in different states. I then create one feature with the corresponding name of the city and its state, which is useful for later searching of geographical coordinates.

The following graphic shows the historical accumulated number of cases for each region.

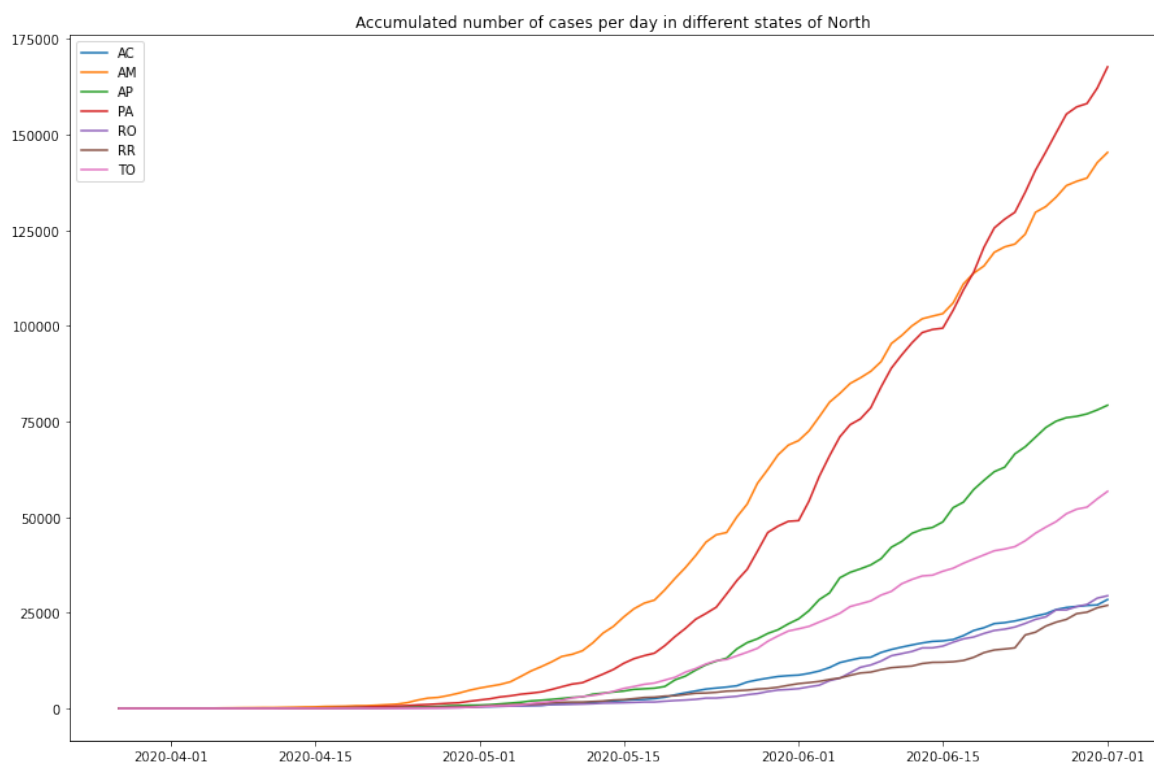


The northeast and north regions are the ones with higher accumulated number of cases, followed by southeast, south and Midwest. The following graphic shows the historical accumulated number of deaths.



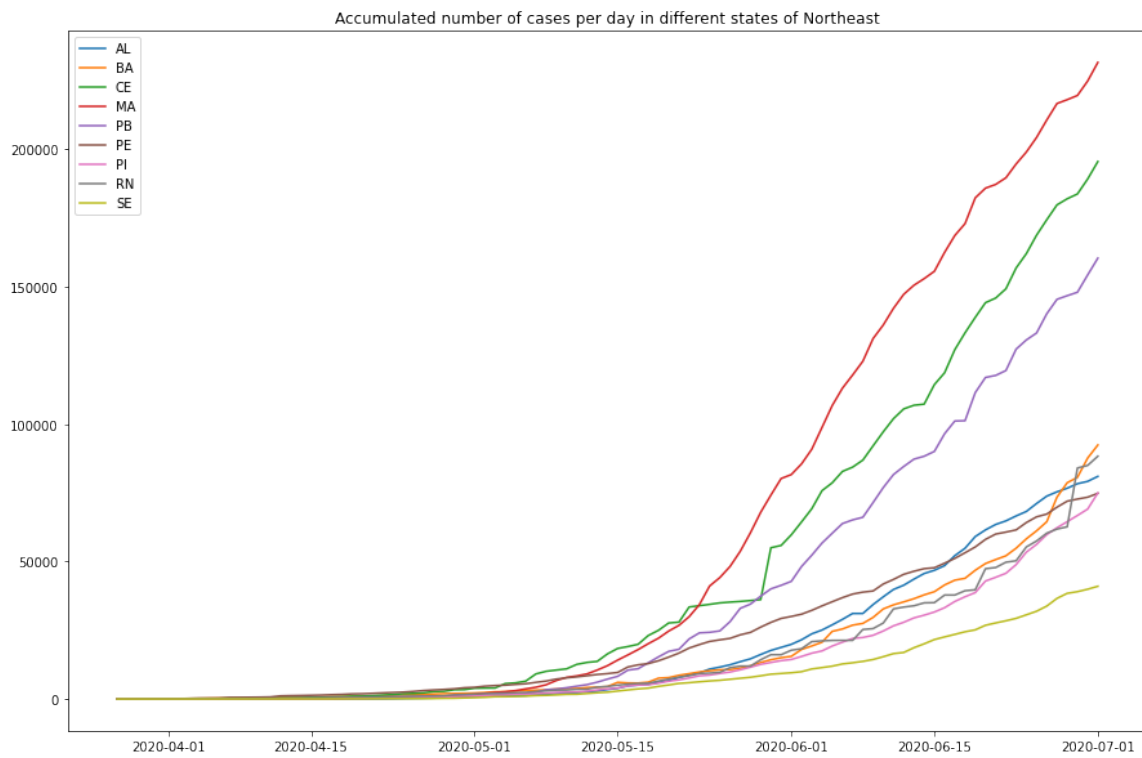
The small tip in the curve of accumulated number of deaths of the north region indicates that there is some wrong data. This is later corrected. The following graphics show how the accumulated cases for each state in each region.

### North



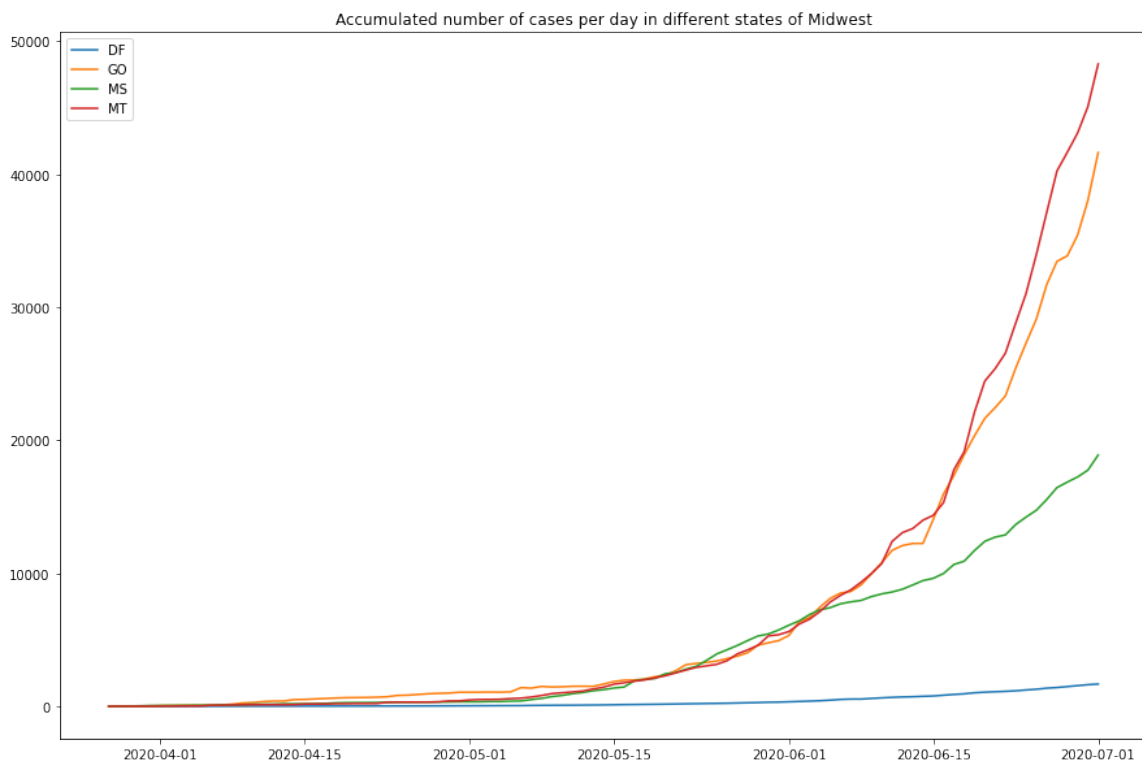
North states: Acre (AC), Amazonas (AM), Amapá (AP), Pará (PA), Rondônia (RO), Roraima (RR), Tocantins (TO).

## Northeast



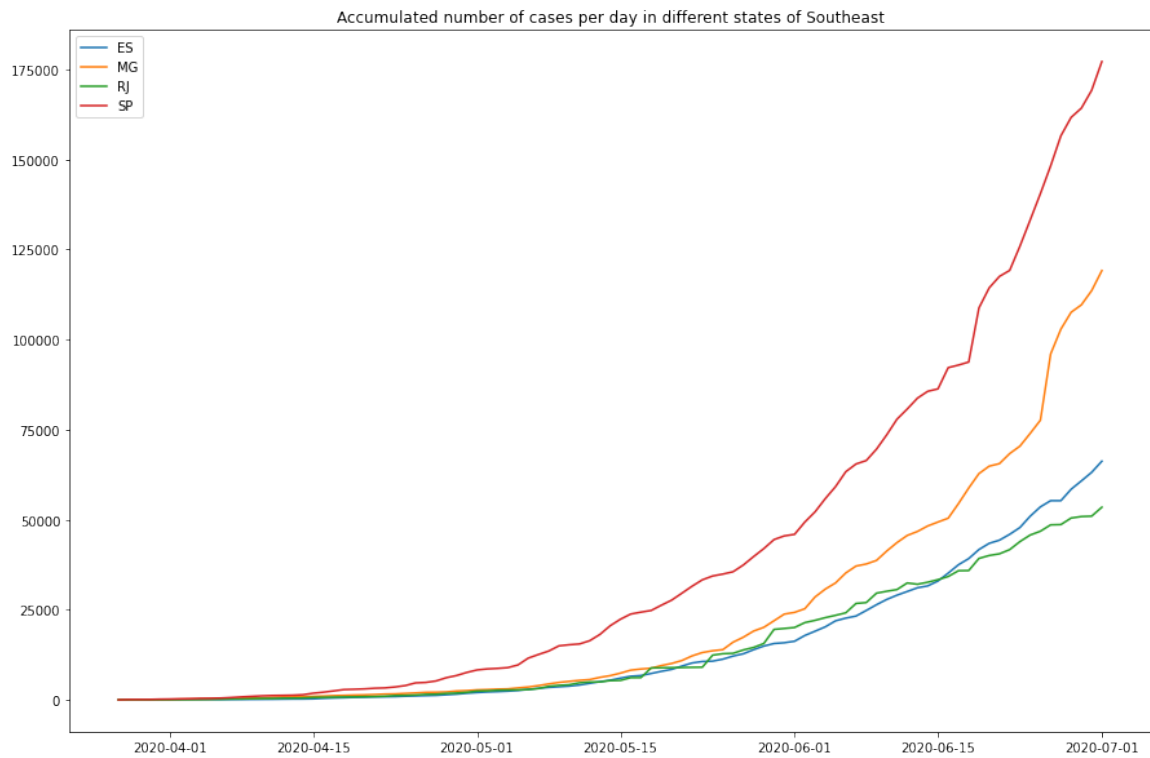
Northeast states: Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN), Sergipe (SE).

## Midwest



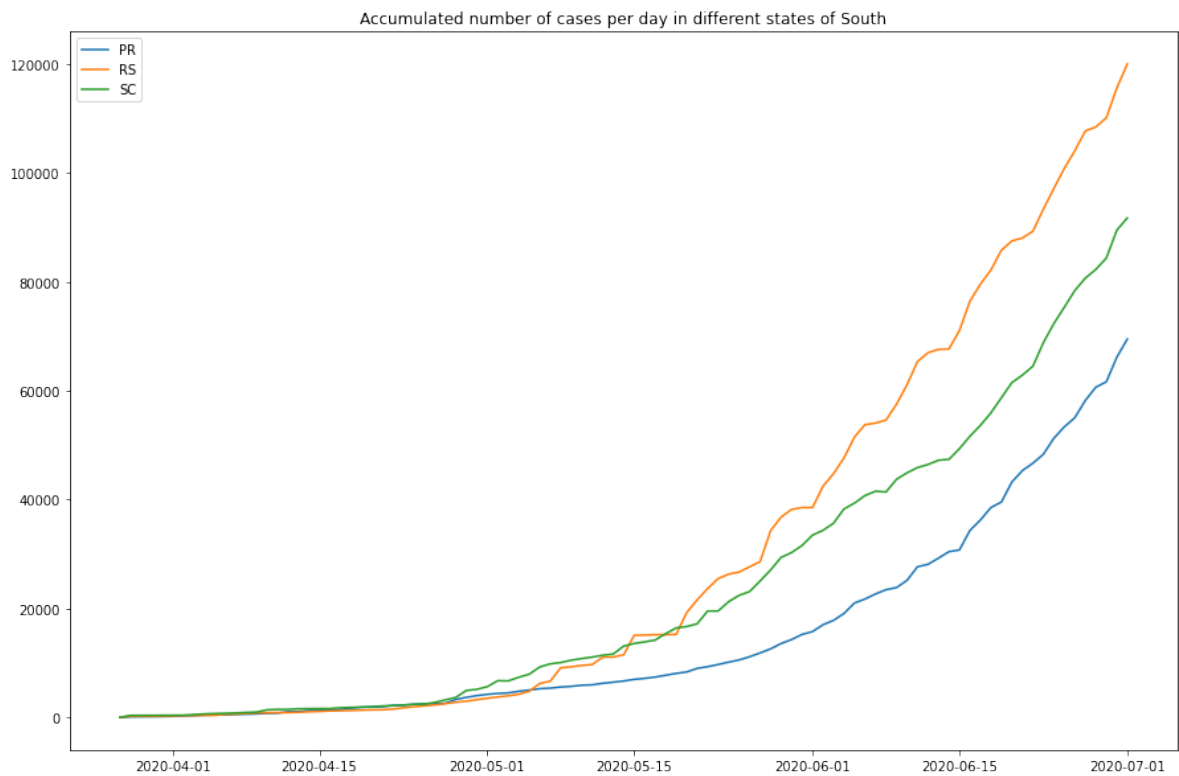
Midwest states: Distrito Federal (DF), Goiás (GO), Mato Grosso do Sul (MS), Mato Grosso (MT)

## Southeast



Southeast states: Espírito Santo (ES), Minas Gerais (MG), Rio de Janeiro (RJ), São Paulo (SP).

## South



South states: Paraná (PR), Rio Grande do Sul (RS), Santa Catarina (SC).

## Results

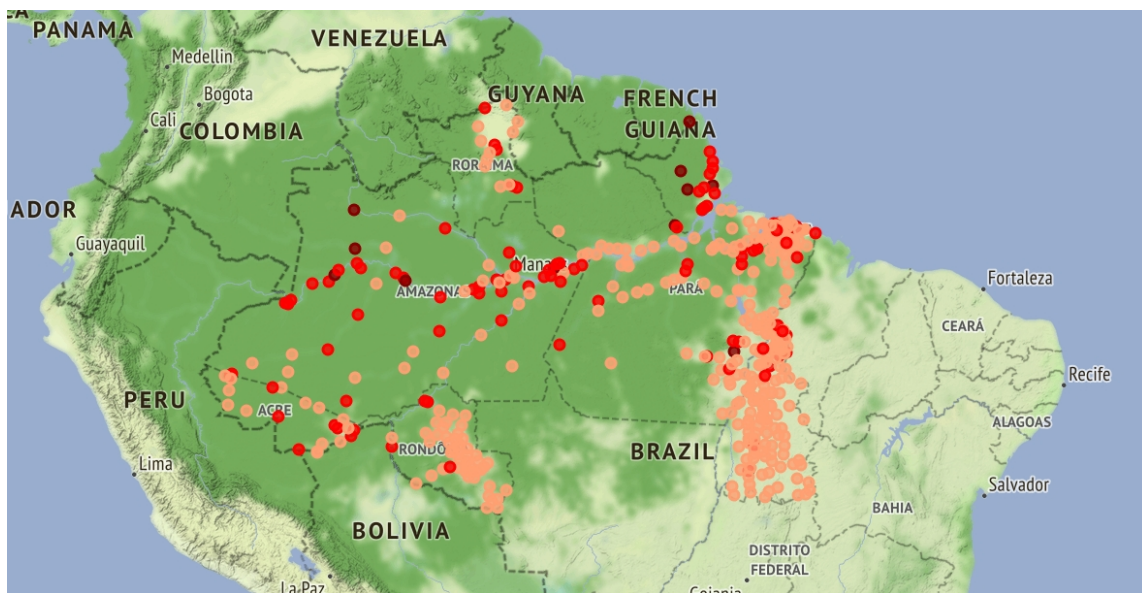
For applying the K-means clustering algorithm, I construct 10 dataframes: one dataframe for accumulated number of cases for each region and one dataframe for accumulated number of deaths for each region as well. The index of each dataframe is the city name and its corresponding state. I first re-scaled all the data so that every column has numbers in the same measure for applying the algorithm, but the result was not satisfactory. The algorithm created one main cluster with almost all of the cities in the region and other clusters with less than ten cities. The best thing is to keep the numbers as they were so the algorithm consider recent days more relevant than early ones, as the number of cases and deaths increases with time.

I set the K-means clustering algorithm using three clusters for every possible situation, to separate cities with low intensity, intermediate intensity and high intensity in their number of cases and number of deaths. Beginning with the number of cases, I find the average number for each cluster in each region, to set an hierarchy between the clusters. The following numbers correspond to the accumulated number of cases/100,000 inhabitants for each cluster in each region in the last day in the data, July 1<sup>st</sup> of 2020.

	Low intensity cluster	Intermediate intensity cluster	High intensity cluster
North	623.6	2,625	6,884
Northeast	283.2	1,162	2,710
Midwest	180.4	883.2	2,526
Southeast	112.6	516.2	1,220
South	181.5	982.4	7,069

For the following maps, I choose tones of red that vary from light to dark depending depending on the intensity of the cluster.

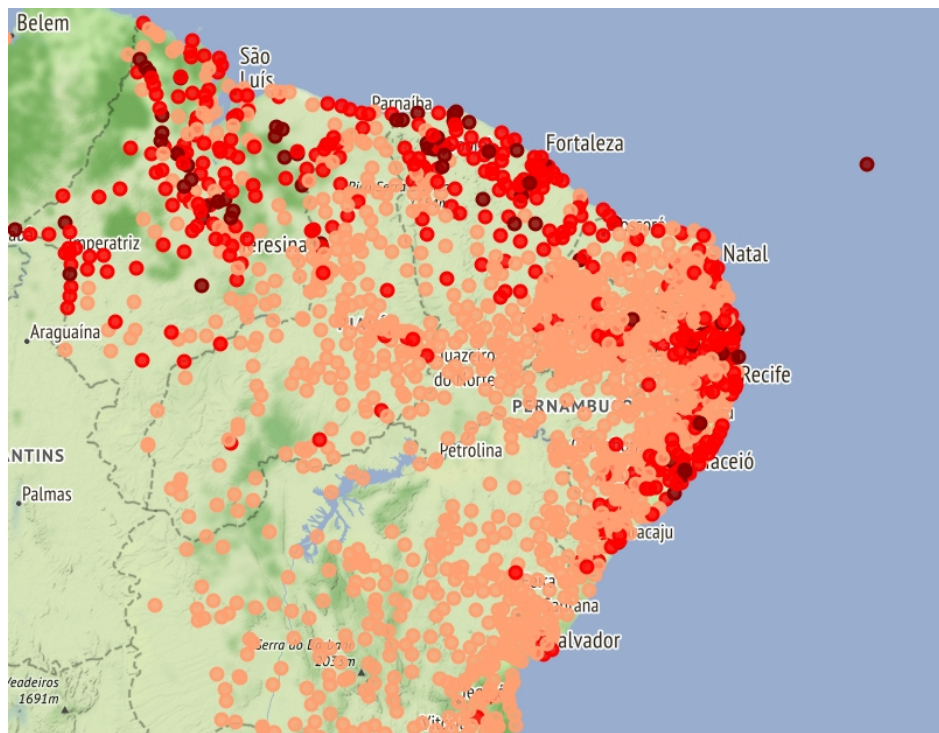
North



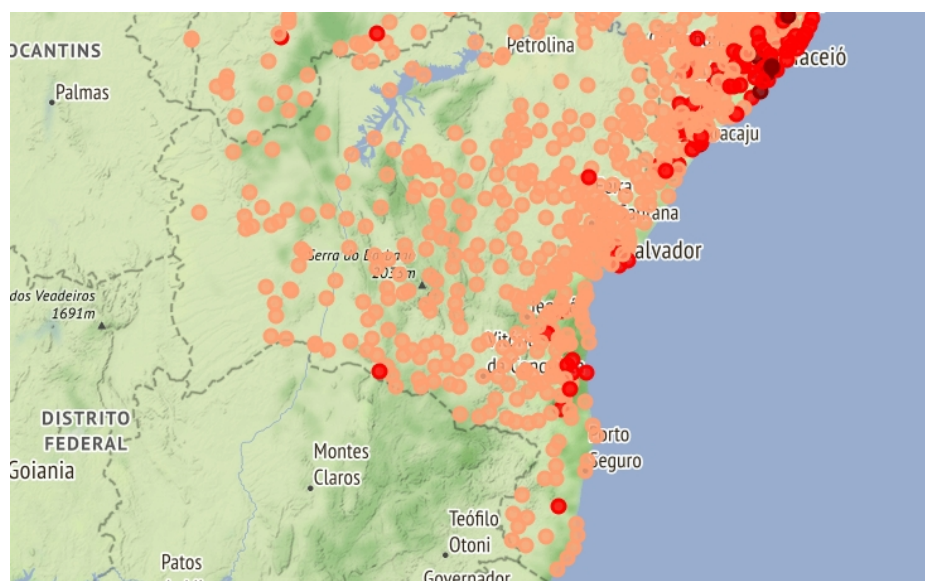
The north region has 450 cities officially, and the data contains 434 of them. The markers' colors are darker as high is the intensity of the accumulated number of cases. Only the state of Pará and Amazonas have cities with high

intensity of number of cases. The great majority of cities in the north region is in the group of cities with low intensity of cases.

#### Northeast



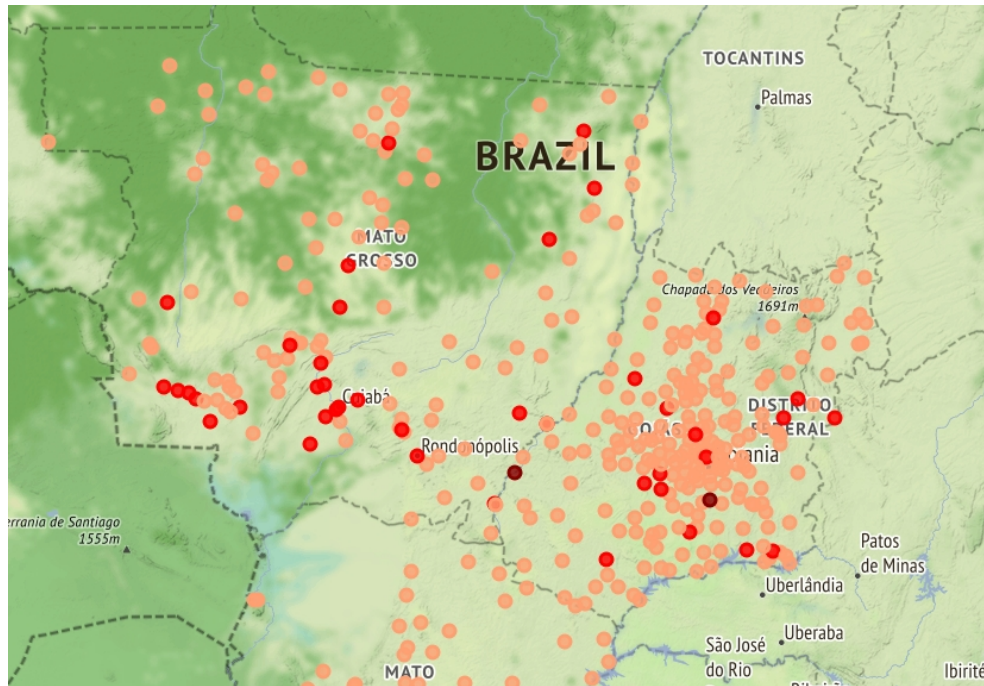
The northeast region has 1,794 cities officially, and the data contains 1,759 of them. As it has much more cities than the north, there are much more occurrences of cities with high intensity of cases, more concentrated in the litoral, as the main cities are there. The exception is the state of Maranhão (top left), with more even distribution of cities with high intensity of Covid-19 cases.



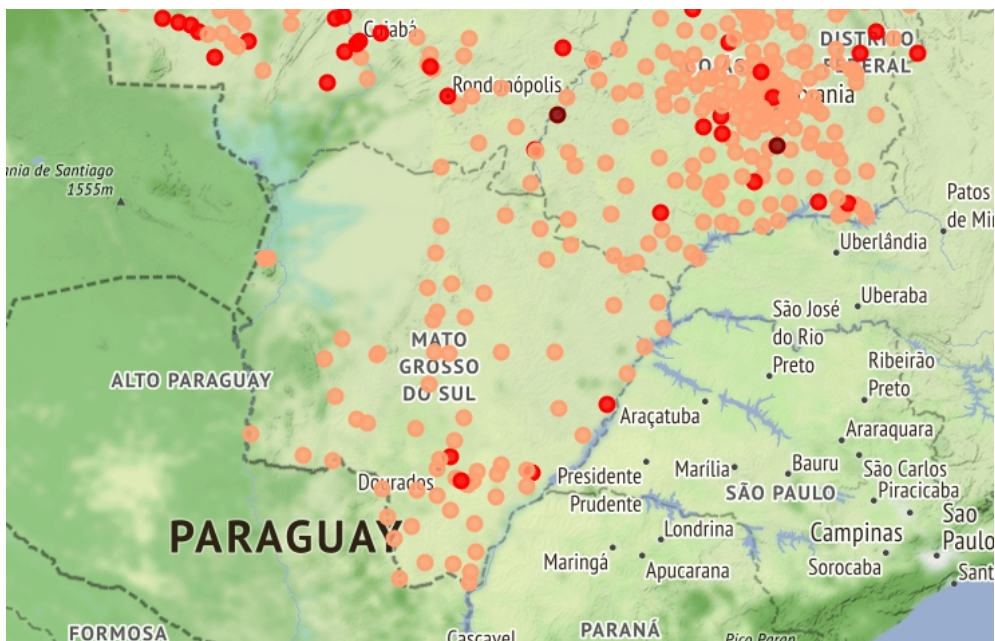
Still in the northeast region, as we move down to the south, to the state of Bahia, the occurrences of cities with intermediate or high intensity of covid-19 cases is much smaller than in the upper part of the region. The state of Bahia adopted a great number of measures for controlling the spread of the disease.



## Midwest



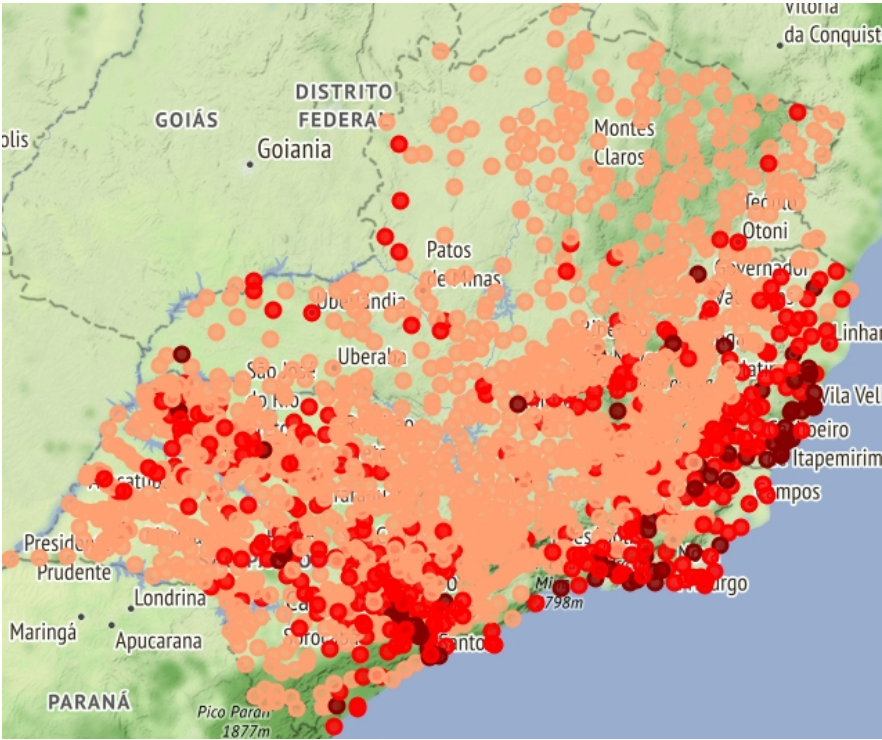
With officially 466 cities and 424 cities in the data, in the midwest, there is only one city with high intensity of number of cases, which is Guia Lopes da Laguna, MS. The cities with intermediate level of occurrence are spread near the states capitals: Brasília, DF; Goiânia, GO; Cuiabá, MT, with some few occurrences in cities far from the capitals.



Still in the midwest, just few cities in the state of Mato Grosso do Sul have intermediate intensity of accumulated number of cases. The capital of the state, Campo Grande have low intensity of number of cases, so are the cities near it, in the geographical center of the state.

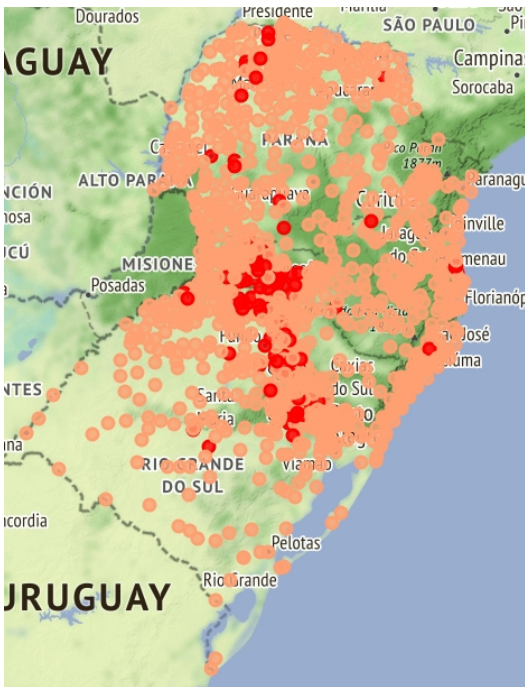


## Southeast



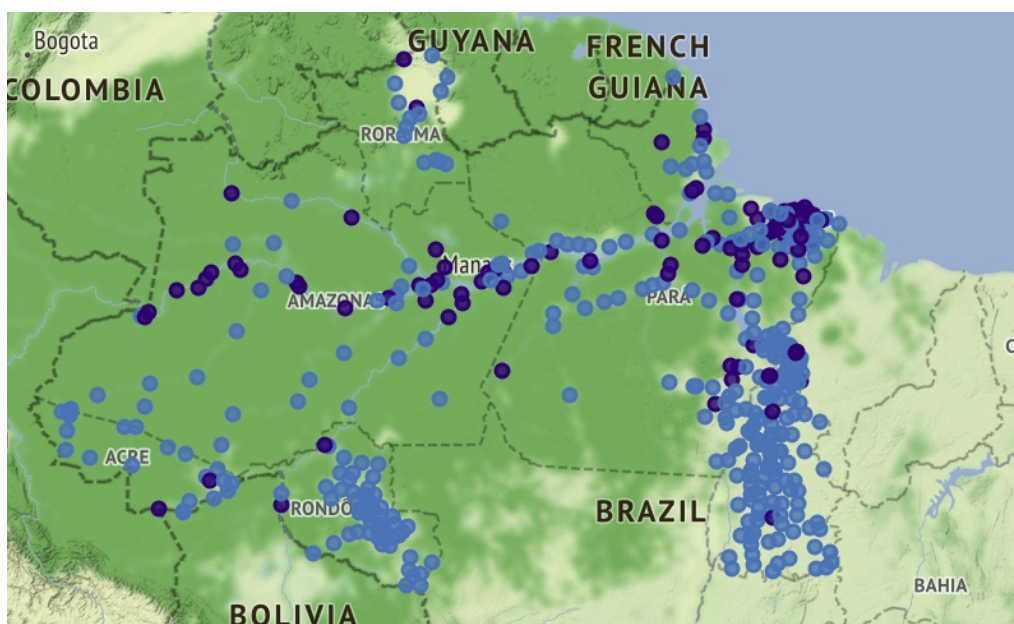
The southeast has officially 1668 cities and 1649 of them are present in the data. The southeast was the first region to present a confirmed case, in February, in the city of São Paulo, SP. The cities with high intensity of accumulated number of cases are more concentrated near the littoral as well as in northeast, but there is a great number of cities with intermediate intensity also in regions far from the littoral. Despite being the second state with higher number of cases, as presented in a previous graphic, the state of Minas Gerais is very huge and has a great number of cities, being the state with fewer cities classified as intermediate or high intensity compared to other states in the southeast.

South

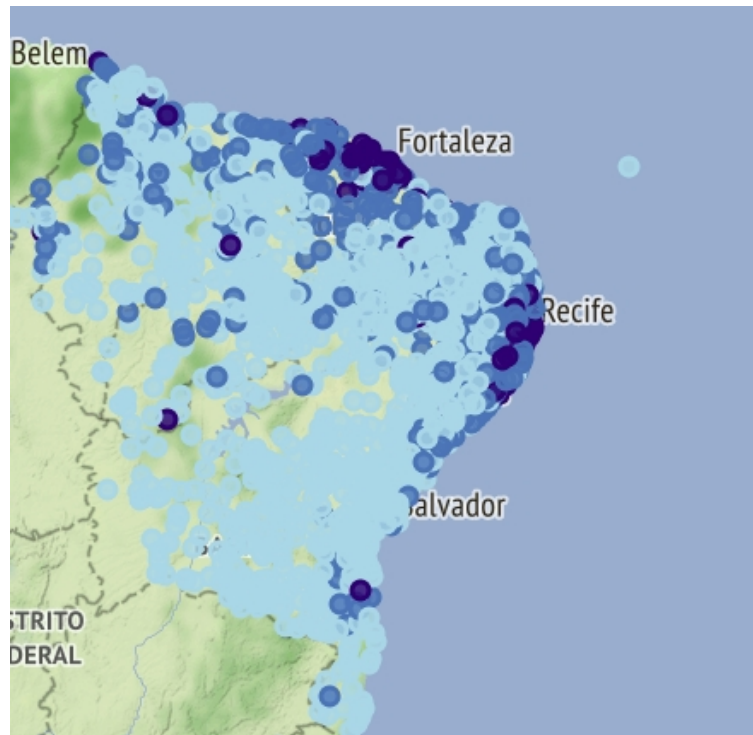


The south region has officially 1,191 cities and 1,067 of them are present in the data. Only two cities belong to the group of cities with high intensity of number of cases, which are Entre Rios, SC and Nova Araçá, RS, both of them are cities from the rural area. The cities with intermediate intensity of number of cases are concentrated more in the center of the region, in contrast with the pattern appearing in northeast and southeast, where cities of littoral are the ones with higher intensity of cases. Also, the cities with intermediate number are not the ones near the capitals of the states.

	Low intensity cluster	Intermediate intensity cluster	High intensity cluster
North	0	13.6	67.4
Northeast	5.08	31.1	68.6
Midwest	3.46	26.4	62.7
Southeast	2.90	25.9	61.7
South	2.81	35.5	151



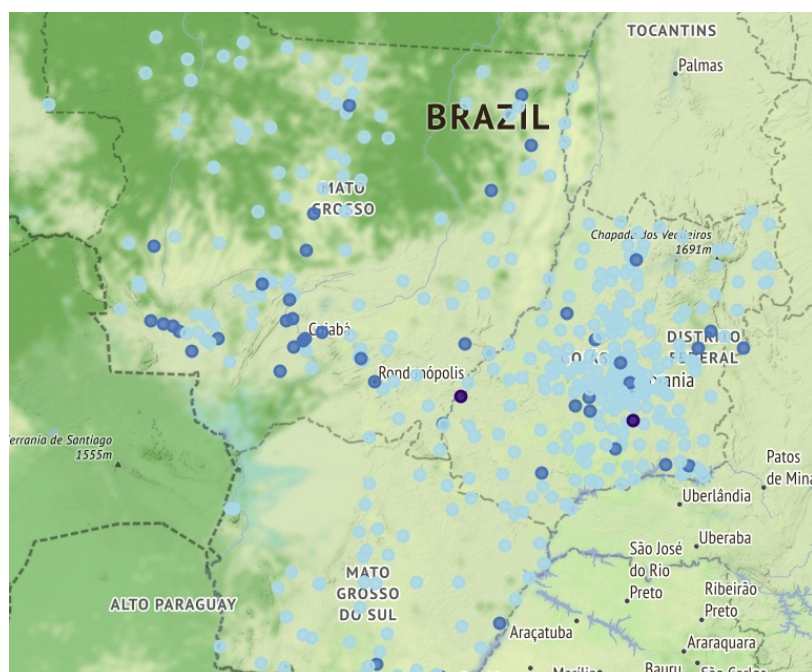
## Northeast



In the northeast region, the distribution pattern of cities is a bit different from the distribution for accumulated number of cases. There are two reasons for this:

- In the state of Maranhão, the cities are more frequently categorized as low or intermediate intensity of number of deaths than for the number of cases;
- The cities with high intensity of number of deaths are less frequent than the ones with high intensity of number of cases, for the whole region.

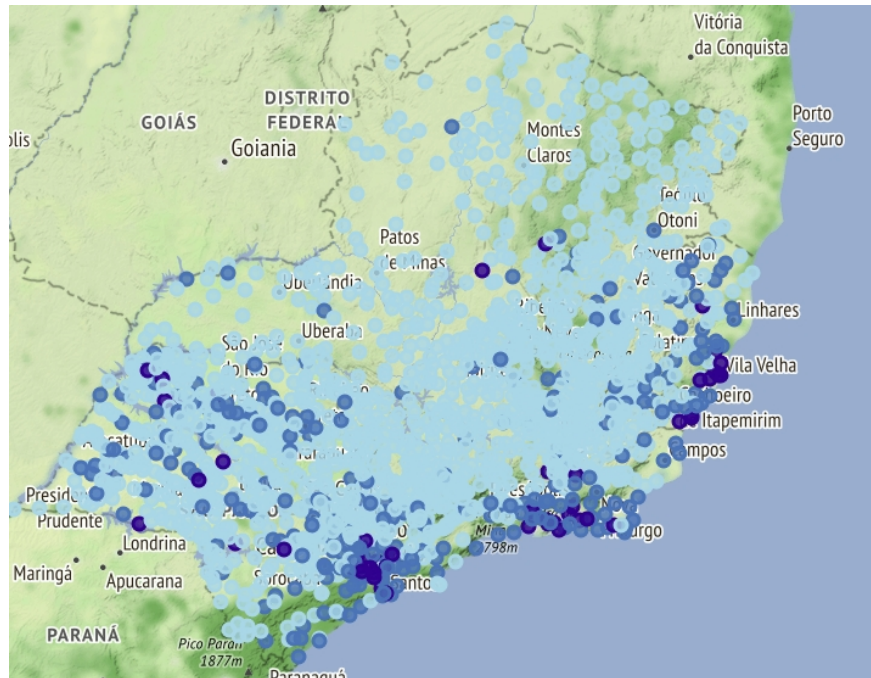
## Midwest





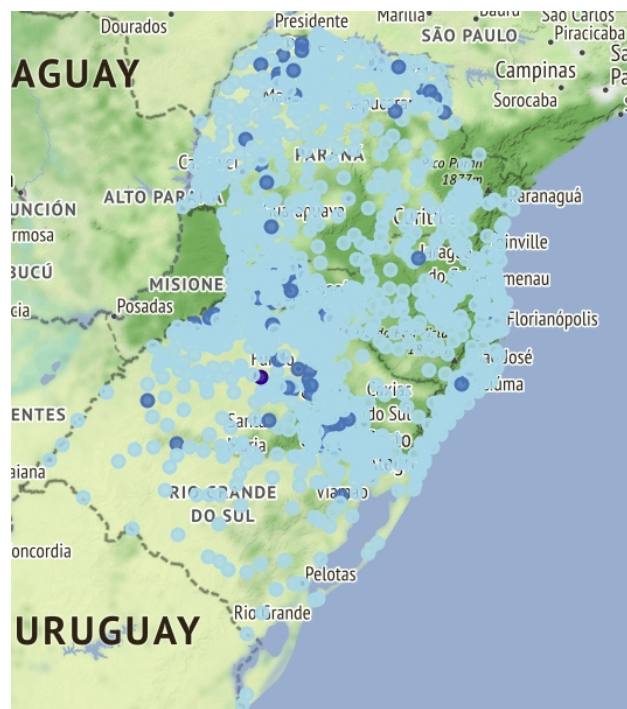
There are only two cities with high intensity in the number of deaths, which are Ponte Branca, MT and Professor Jamil, GO. These cities are different from the city with high intensity in the number of cases. The cities with intermediate intensity of number of deaths are also spread near the states capitals, which also happens for cities with intermediate intensity of number of cases.

#### Southeast



The distribution pattern for the intensity in the number of deaths is less strong than the distribution in the number of cases. The north of the region has few cities with intermediate number of deaths and even fewer with high number, in opposition to what is seen in the map for the accumulated number of cases. Also, the regions with intermediate and high intensity are more concentrated in the state of Espírito Santo, littoral of Rio de Janeiro, but more evenly distributed through the state of São Paulo.

#### South



Only few cities in the south region have intermediate number of deaths, and only one has high number of deaths, which is Saldanha Marinho, RS. We can do the same as we did for the north region and consider that there are only two clusters in this case, one with low intensity of deaths and one with intermediate intensity of deaths.

## Discussion

The accumulated number of cases/100,000 inhabitants shows that the order of the regions by descending accumulated number is northeast, north, southeast, south and midwest. From these regions, the ones with greater number of cities with high intensity of numbers of cases are northeast and southeast, that are the two biggest regions in the country and concentrates a high fraction of the country's population.

We can consider that every region in the country lives a grave situation relative to the spread of the coronavirus. Some states more than others.

In the case of the north region, the gravest situation is concentrated in the states of Pará (PA) and Amazonas (AM), both of them having almost all of the cities with intermediate and high intensity of accumulated cases and deaths.

In the northeast, the gravest situation are found in the states of Maranhão (MA), Ceará (CE) and Paraíba (PB), with respect to the accumulated number of cases, but for the accumulated number of deaths, the state of Maranhão has a less grave situation than Ceará and Paraíba. The pattern of the spread is from the littoral to the interior, indicating directions to reduce the transit of people in the attempt of reducing the spread.

In the midwest, the gravest situations occur in the states of Mato Grosso (MT) and Mato Grosso do Sul (MS), though the distribution pattern of the intensity of accumulated cases and deaths are more homogeneously distributed than in other regions, more concentrated in some parts.

The southeast was the first region to register a case of infection by the virus in the city of São Paulo. After that, the numbers of cases have grown very fast and spread through the whole region. The higher accumulated numbers are concentrated in the states of São Paulo and Minas Gerais, though the state of Rio de Janeiro can be considered in a graver situation than Minas Gerais, as it is much smaller with smaller number of cities compared to Minas Gerais.

In the south region, it can be seen in the numbers of accumulated cases that, in comparison to other regions that the high intensity clusters for number of cases and deaths have very large numbers compared to clusters of same level of intensity in other regions. This is due to the number of cities in the clusters. The situation in the region is less grave than in the north, northeast and south, but graver than midwest.

## Conclusion

The spread of coronavirus in Brazil was very fast and intense. In less than 4 months, most of the metropolitan cities in the country has already a great number of cases and deaths. However, there are many cities that still did not register big number of cases or deaths.

Political measures in a national level could have slown down the spread of the disease in the country, and could also make the pattern of the spread more homogeneous and smooth throughout the country. The maps show some level of regularity in the distribution of cities with high intensity of cases and deaths, but there are many cities that do not follow this pattern throughout the country;

The notion of where the disease is more concentrated makes possible the adoption of measures to block the spread from some regions with more intense numbers to the others less intense numbers.