

COVID-19 Data Analysis of Tolima-Colombia

Cristian F. Rubio A.

Julio 31, 2020

A. Introduction

A.1. Description & Discussion of the Background

Tolima is one of the 32 departments that make up the Colombian territory with a population of more than 1.33 million inhabitants and a surface density of 23,562 square kilometers. The department is divided by 47 municipalities in total, including its respective capital. As a resident of one of its municipalities, I decided to use this department to develop my project and promote regional development.

Several municipalities in this department, like many others in Colombian territory and throughout the planet, have been affected by the presence of what we call COVID-19 and, although it is a hot topic, there is not enough information to support and / or show The status of the data. We have official sources of information such as the INS (National Institute of Health) but without spatial or positional samples that clearly communicate the data at least in the smallest branches of the Colombian state (territorially speaking), and at the same time impartially share representation of the data.

When these problems are considered, we can think of creating contagion maps, distribution charts, even regressions, and all kinds of visually clear and descriptive information that exposes the data.

B. Data Description

To consider the problem, we can list the data as follows:

- The **INS** (National Institute of Health) is in charge of providing the necessary data on COVID-19 for Colombia.

- With **Forsquare API** the geographic positions of the Tolima municipalities will be obtained.
- Libraries such as **Geopandas** and **Folium** will be used to obtain the maps.
- Each municipality has a related code divipola that will be used to mediate between the data of the INS and the information of the **DANE** (National Administrative Department of Statistics) to collect information and geographic maps to complement or replace the previous points.

C. Methodology

In this project, efforts are directed to the first attempt to expose the information effectively:

ID de caso	Fecha de notificación	Código DIVIPOLA	Ciudad de ubicación	Departamento o Distrito	atención	Edad	Sexo	Tipo	Estado	...	FIS	Fecha de muerte	Fecha diagnostico	
0	1	2020-03-02T00:00:00.000	11001	Bogotá D.C.	Bogotá D.C.	Recuperado	19	F	Importado	Leve	...	2020-02-27T00:00:00.000	NaN	2020-03-06T00:00:00.000
1	2	2020-03-06T00:00:00.000	76111	Guadalajara de Buga	Valle del Cauca	Recuperado	34	M	Importado	Leve	...	2020-03-04T00:00:00.000	NaN	2020-03-09T00:00:00.000
2	3	2020-03-07T00:00:00.000	5001	Medellín	Antioquia	Recuperado	50	F	Importado	Leve	...	2020-02-29T00:00:00.000	NaN	2020-03-09T00:00:00.000
3	4	2020-03-09T00:00:00.000	5001	Medellín	Antioquia	Recuperado	55	M	Relacionado	Leve	...	2020-03-06T00:00:00.000	NaN	2020-03-11T00:00:00.000
4	5	2020-03-09T00:00:00.000	5001	Medellín	Antioquia	Recuperado	25	M	Relacionado	Leve	...	2020-03-08T00:00:00.000	NaN	2020-03-11T00:00:00.000

In the first step, the data is collected, loaded and cleaned for further processing of the information.

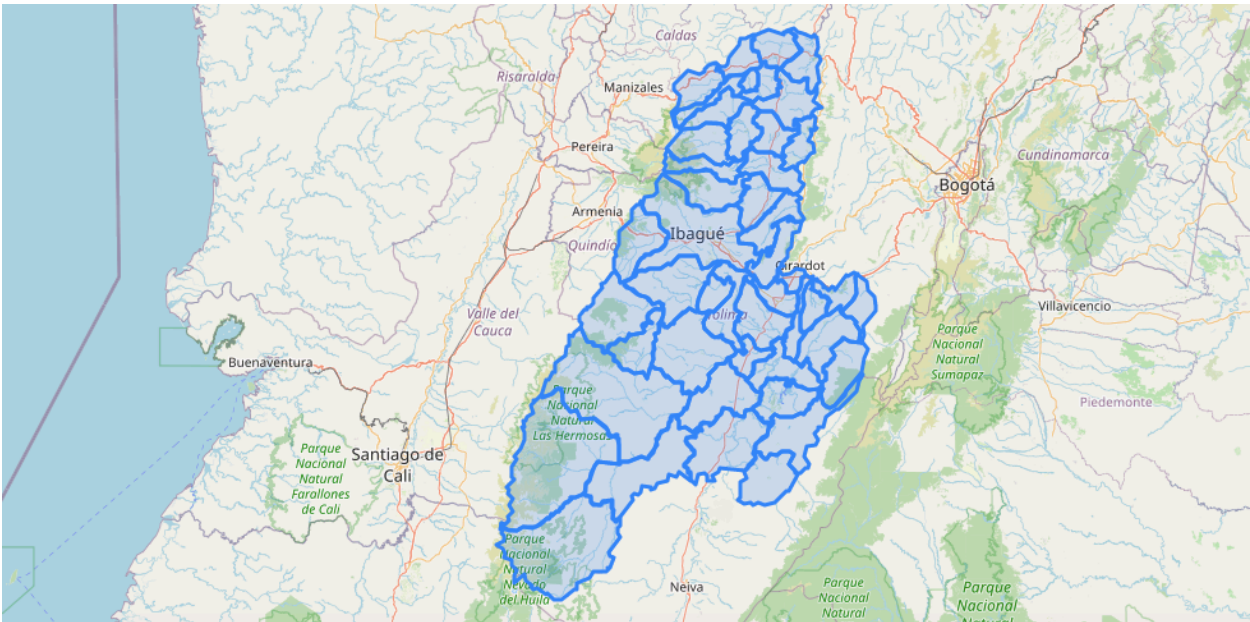
	muncodigo	departamentos	municipios		muncodigo	departamentos	municipios	counts
0	73001	TOLIMA	IBAGUÉ	0	73001	TOLIMA	IBAGUÉ	966
1	73001	TOLIMA	IBAGUÉ	1	73268	TOLIMA	ESPINAL	1144
2	73001	TOLIMA	IBAGUÉ	2	73449	TOLIMA	MELGAR	58
3	73001	TOLIMA	IBAGUÉ	3	73319	TOLIMA	GUAMO	21
4	73001	TOLIMA	IBAGUÉ	4	73624	TOLIMA	ROVIRA	8

We pre-process the data and obtain how many cases exist for each municipality.

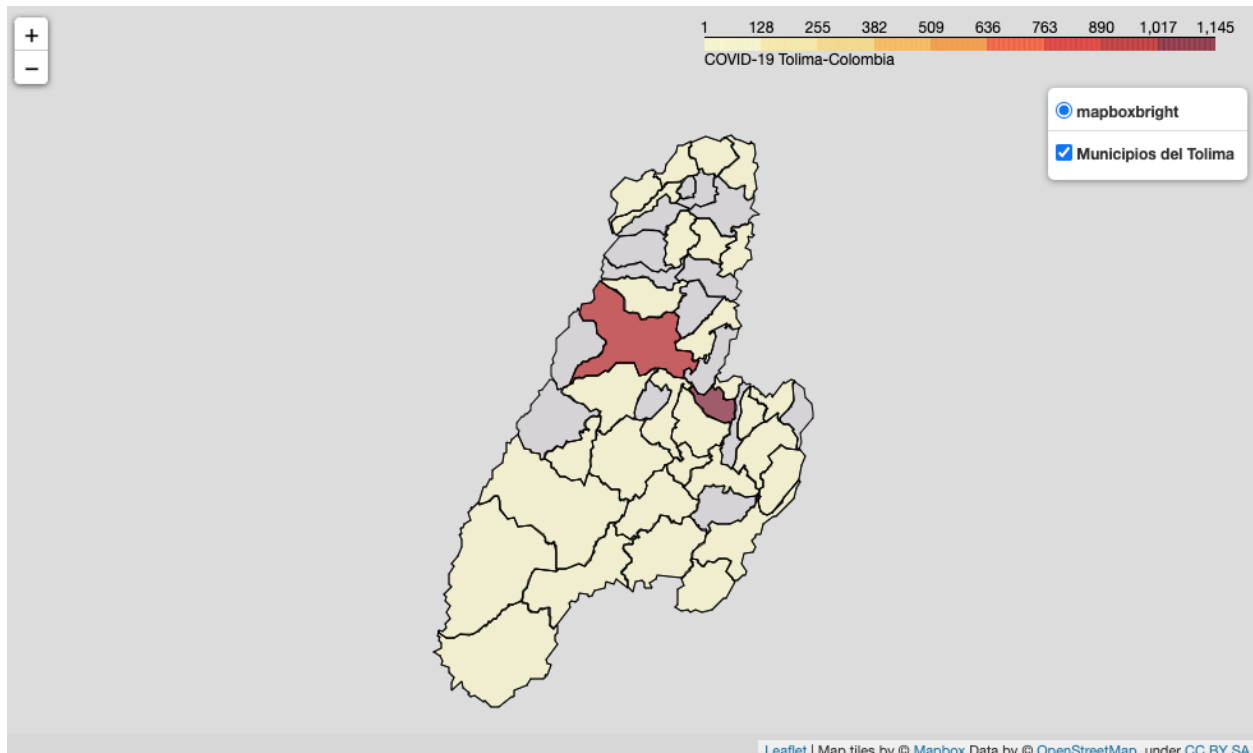
In the second step, it will seek to generate the contagion map of the department of Tolima-Colombia to achieve visual ease and clarity in the information, as well as allowing interactivity between the user and the exposed data.

	geometry	muncodigo	departamentos	municipios	counts
4	POLYGON ((-74.90818 4.24985, -74.90812 4.24985...	73268	TOLIMA	ESPINAL	1144
11	POLYGON ((-75.07451 4.56792, -75.07423 4.56814...	73001	TOLIMA	IBAGUÉ	966
6	POLYGON ((-74.55565 4.26726, -74.55656 4.26664...	73449	TOLIMA	MELGAR	58
8	POLYGON ((-75.03287 4.28047, -75.03259 4.28042...	73678	TOLIMA	SAN LUIS	57
27	POLYGON ((-75.03279 3.81964, -75.03318 3.81932...	73217	TOLIMA	COYAIMA	34

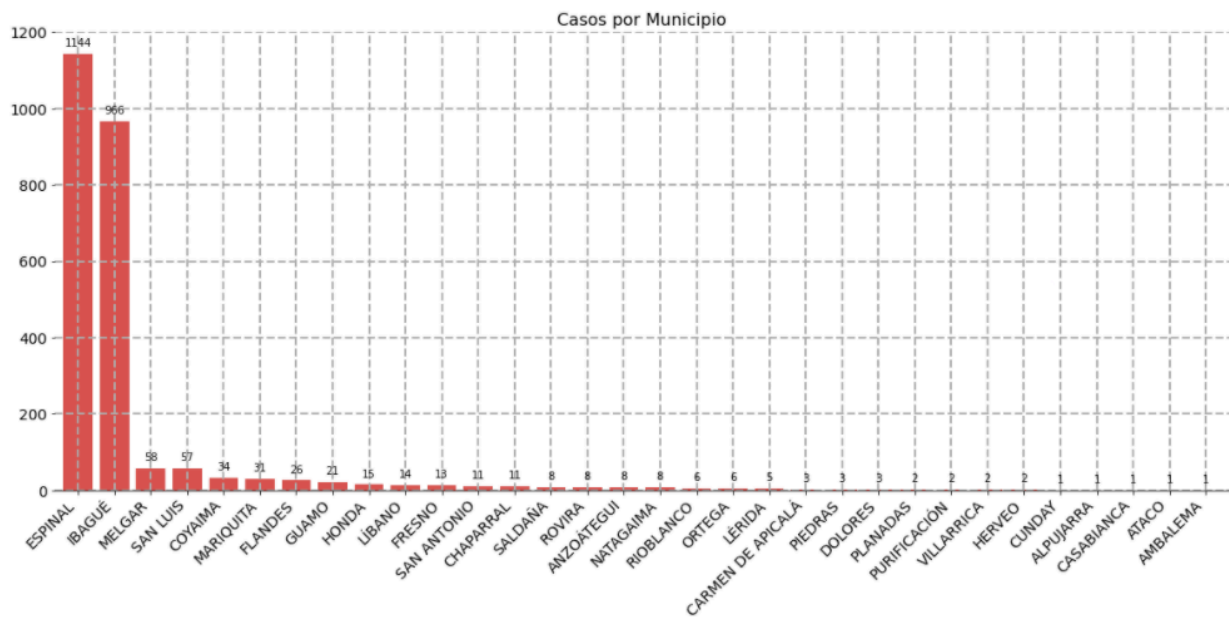
We proceed to obtain the respective location coordinates by asking for the municipality

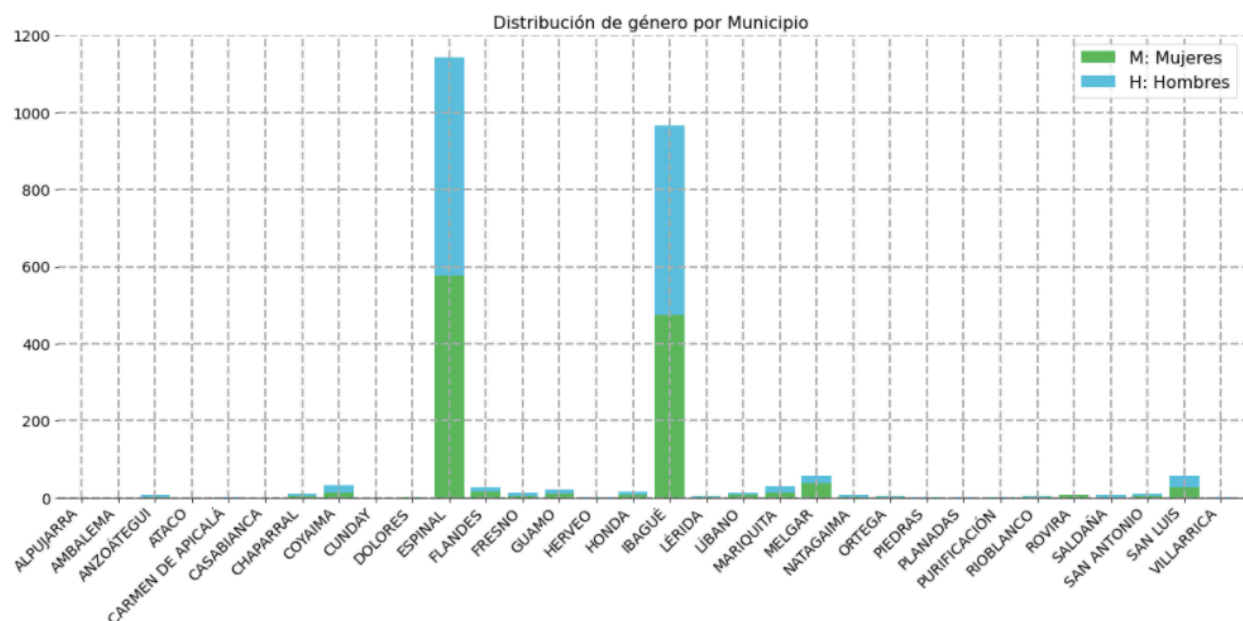


Now we pass the number of people infected by each municipality on an interactive map so that the user can move around the interface provided and feel comfortable in understanding the data.



In the third step, it will focus on showing the number of cases per municipality in an orderly manner, in addition to the gender distribution by infected municipality.





D. Results and Discussion

Our analysis shows that the largest number of confirmed cases of COVID-19 in the department of Tolima-Colombia is mainly concentrated in two municipalities, Espinal and Ibagué, the capital of Tolima. The figures are alarming since Espinal, with approximately 76.000 inhabitants, has already registered 1.144 people to date (July 31, 2020), while Ibagué, with some 570.000 inhabitants, registers 966 cases of contagion.

Of the 47 municipalities of the department, to date 32 of them have at least one confirmed case. The data is grouped by gender to observe the distribution of the contact for both cases, that is, both men and women to try to observe patterns of behavior biased to some gender, but we found that in the end there are no bias marks because the virus does not discriminate.

The portion of the results exposed to barely opens the gap for an in-depth analysis of the data, since COVID-19 is an ongoing issue with small samples of a definitive vaccine and a big road to what it really is.

E. Conclusion

The objective of this project was based on generating and exposing from the information acquired by the INS (National Institute of Health), other official organizations such as the DANE and open source software, for the implementation of the information effectively and clear. There is a long way to go and it is hardly an effort to achieve a definitive result from which one can conclude about the other data, and more than concluding and trying to perhaps arrive at punctual answers, it is more an attempt to be an exquisite point of information in the presentation of the data.

F. Future directions

I plan to further explore and analyze the data and share it with the local community who are primarily the most interested as well as the people interested in data science who want to replicate this work in their city or country.