Team members: Pedro Huet
GitHub ID: Pedro-Huet
Email: pedroh@uchicago.edu

Final Paper Write-Up: The state of the State of Mexico during a state of emergency

This is a project I made pertaining the covid-19 data available in my home state: Estado de México (State of Mexico). The project basically consists in finding out if the population size or metropolitan nature of the municipalities were the variables that affected these political units performance during the COVID-19 Pandemic. My research hypothesis was: does the population of a municipality have a negative effect in these local governments' performance tackling COVID-19? Or was it the interconnectivity of the municipalities to the capital (Mexico City) the factor explaining their detrimental performance?

To be able to answer the question, I needed a way to measure the municipalities performance to covid-19. For this, I chose to measure performance by the total amount of confirmed 1) COVID-19 cases and 2) deaths. In this situation, it is reasonable to think that the higher amount of COVID-19 cases (continuous) as well as deaths (continuous) variables indicate that authorities could have offered a better response to halt the spread of the disease among their residents. The independent variables that were used were continuous and dichotomous iterations of the population size of the municipality (raw population size, population size per 1,000 inhabitants and if municipalities had more than 50,000 inhabitants), as well as if the municipality was considered among the Metropolitan Zone of the Valley of Mexico Municipalities or not (dichotomous).

In order to make this analysis possible, I first needed to find a database that contained sufficiently precise and trustworthy information of COVID-19 cases and deaths in the State of México. To do so, I employed the Official Mexican Government Health Department Data (Mexican Government Open Data, 2022), which tracks every reported Covid-19 case in the country, as well as the residence information from the patient during the height of the pandemic (2020 - 2021). Second, I needed to find reliable and updated data regarding the amount of population in the municipalities. The best database for this task is data from the National Institute of Geography and Statistics (2020), whose most recent updated population estimations per municipality in the State of Mexico were in 2020, a relatively short period of time back. To find the precise information regarding the municipalities which were part of the Metropolitan Zone of the Valley of Mexico, I accessed a state government website (State System of Urban, Metropolitan and Housing Information, 2023) with the specific information, whose written content I completely scrapped. Finally, for the data visualization component of my graph, I was able to also find a database from the Institute of Geographical, Statistical and Cadastral Information and Research of the State of Mexico (2018). Since the state's municipality layout have not changed during, at least, the last 5 years, this information should be useful for our purposes.

The approach I took to coding this project was to prioritize the creation of the general database needed to perform the analysis. Therefore, the order in which I performed the projects coding tasks was likely not the same as that of many other students. My first step was to be able to access the information of the total COVID-19 cases in Mexico. Since the database weighs almost 1.4 gigabytes and my computer has a limited amount of VRAM (it was not able to

Team members: Pedro Huet
GitHub ID: Pedro-Huet
Email: pedroh@uchicago.edu

open the data via Python on the onset), I chose to manually extract the column information by chunks, only keep the columns of interest (confirmed COVID-19 cases and deaths among individuals who resided in a municipality of a given state) and manually obtain the data of only for State of México. When talking about confirmed cases, the database defined a few categories in an uncertain manner (suspected, lost results, non-clinical diagnosis). Since the Mexican Government was internationally criticized for not processing COVID-19 data correctly and even trying to hide these for political purposes (Luhnow and Montes, 2021), I have chosen to err in the side of caution and consider all the cases which were not straight out rejected by a COVID-19 test as a confirmed case. In the case of COVID-19 deaths, these are counted as deaths of individuals with confirmed cases.

The second step was to pinpoint which cases corresponded to each municipality (in this database, they were labeled as numbers from 1 to 125) and collapse it (instead of having thousands of confirmed observations, just have a number of cases and deaths for the 125 municipalities), so I used a "dictionary" database that was included in the download to label the municipalities based on their corresponding number (a crosswalk). Next, I merged this database with the population data, which was relatively simple. The only thing I had to do throughout the data is standardize the name of the municipalities, as that was the main tool I was employing to perform the merge. Therefore, I converted the municipalities names in caps throughout all the document.

The third step is where I deviated from the average user, as I performed the major web scarping and text processing step at this point when trying to retrieve the municipality data. In the process, I was able to find the ZMVM municipality names throughout the whole text and convert my list into a new column depending if the municipalities matched other elements in the list (ZMVM municipality), marking this as a 1 or else 0. I also had to take care of making the regional pronunciation and writing of some municipalities compatible with the names I had. After having completed this task, I finished merging my databases and imported the Geopandas shapefile into my database, which I promptly merged and set the relevant column as merged data. At this point, I was able to perform data analysis and visualization.

In performing descriptive statistic tests for both the relevant variables, using the most effective charts for each variable, I realized that there was a disproportionally high amount of covid-19 cases and covid-19 deaths in municipalities with large populations and that are considered part of the Metropolitan Zone of the State of Mexico. When displaying the data in the shape of maps, I was also able to visualize that the municipalities with most covid-19 cases were also the ones that had large populations and that are considered part of the Metropolitan Zone of the State of Mexico, although the tendencies regarding deaths were not as clear.

It is then that I performed a series of simple linear regression models to be able to observe the statistical indicators to know if there was an actual relationship between the data. By performing 4 linear regression models (one for each pair of variables), I was able to various results. In the case of population size, the results indicated that highly populated

Team members: Pedro Huet
GitHub ID: Pedro-Huet
Email: pedroh@uchicago.edu

municipalities seemed to have had a noticeable impact on both state cases and death rates, as, on average, for every 10,000 people that inhabited a highly populated municipality there were an additional 218 cases and 15 deaths, at a high statistical significance level (significance level of above 1%). In the case of metropolitan zone municipalities, the results indicated that these metropolitan municipalities seem to have had noticeable impact on state case and deaths rates, as, on average, these had 3,500 more cases and 221 deaths, at a high statistical significance level (significance level of above 1%).

These results seemed enticing, but I wanted to know if there was any overlap between the variables; after all, metropolitan zones could be the most populated. Turning my 4 simple linear models into 2 multivariate regressions with both variables, I was able to see a different picture: while population size retained its magnitude and statistical significance on covid-19 cases and deaths when accounting for metropolitan status (215 more cases and 15 deaths per 10,000 inhabitants), in both regressions, the metropolitan status of the municipality still retained a noticeable amount of significance and magnitude (541 cases at a significance level of above 10%) when explaining the amount of cases in the unit, its magnitude and statistical significance plummeted when explaining deaths (accounted for 5 more deaths at a significance level of even below 70%).

From this, I was able to draw some conclusions that could be useful for future studies. **While both highly populated and metropolitan municipalities tend to perform when mitigating the spread of covid-19, population size clearly had a clearly a larger and more statistically significant effect. When it came to death counts, only highly populated municipalities seem to perform worse. The explanation to this might be that, because metropolitan municipalities have better connectivity to the capital and other relatively developed zones, although people fall ill to covid-19, they are able to better access better medical facilities and services that avoids them from perishing from the illness.**

Sources:

Institute of Geographical, Statistical and Cadastral Information and Research of the State of Mexico. (2018). Government of the state of Mexico. http://acvisor.edomex.gob.mx/AtlasCibernetico/portal/visorAtlas.do

Luhnow, D. and Montes, D. (2021, January 14). 29 Family Members Fell Ill With Covid. Mexico Didn't Count Them. *The Wall Street Journal*. https://www.wsj.com/articles/29-family-members-fell-ill-with-covid-mexico-didnt-count-them-11610642019

Mexican Government Open Data. (2022). Government of Mexico. https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia

National Institute of Geography and Statistics. (2020). Government of Mexico. https://en.www.inegi.org.mx/programas/ccpv/2020/#open_data

State System of Urban, Metropolitan and Housing Information. (2023). Government of the State of Mexico. http://plataforma.seduym.edomex.gob.mx/SIGZonasMetropolitanas/PEIM/descriptiva.do