



## **COVID-19 in Mexico: Exploratory Data Analysis (2020-2022)**

Gerardo Reyes Guzmán

*Universidad LaSalle Bajío Campus Salamanca*

Marco Antonio Escobar Acevedo

*Universidad LaSalle Bajío Campus Salamanca*

Perla Esperanza Rostro Hernández

*Universidad LaSalle Bajío Campus Salamanca*

**Abstract:** The present essay aims at analyzing the official data related with Covid-19 pandemic in Mexico from February 2020 to April 2022. To do that we conducted an Exploratory Data Analysis (EDA) that allowed us to find insightful results. There was a mismatch between the number of Covid-19 deaths reported by the authorities and the total number of deaths of all causes published for the year 2020; we spotted a record velocity of 10,000 Covid-19 deaths in just 8 days during the second wave of the pandemic and found that 70 percent of Covid-19 deaths took place in two regions Mexico's that turn to be strategic in socioeconomic terms. This concentration phenomenon was reinforced through the Price Square Root Law since 47% and 51% of all Covid-19 deaths in 2020 and 2021 respectively, were registered in 6 states.

**Keywords:** Covid-19; Exploratory Data Analysis; Mexico; Cases; Deaths

### **Introduction**

Christopher Murray published at the beginning of 2022 that the omicron variant of SARS-Cov-2 was striking with a huge wave of infections estimated to be around 125 million people a day worldwide, ten times the Delta variant in April 2021. Although large proportions of the population were thought to be infected by the omicron variant in the first quarter of 2022, the effects on health were expected to be modest because the delta variant created some level of immunity among people. However, the global infection-detection of omicron declined from 20% to 5% because a vast proportion of cases were asymptomatic. According to Murray, more than 90% of infections in South Africa were asymptomatic. (Murray, 2022). The ratio of Covid-19 hospitalizations to detected cases hospitalized had declined by about 50% in most states in the USA.

This is the second essay on the Covid-19 pandemic in Mexico in which we resort to EDA to

analyze the official numbers concerning this phenomenon (Gerardo Reyes Guzman, 2021). In the first one, we concluded that Mexican authorities failed at applying the required PCR-tests and that there were shortcomings by the authorities' capacity to determine the real number of infections and victims. We also found important correlations between Covid-19 deaths and income labor per capita, health facilities and violent homicides nationwide. This time we resort once more to an Exploratory Data Analysis (EDA) being aware of its usefulness in science (Tukey, 1962). The objective of this study is to shed some light at the whole pandemic period in hindsight, since by April 2022, authorities allowed the normalization of activities indirectly declaring the end of the pandemic in Mexico (Relaciones Exteriores, 2022).

By mid-April 2022, there were 502 million positive cases and 6.19 million deaths in the world. Regarding the number of deaths, the USA appeared in the first place with 986,000; second, Brazil: 662,000; third, India: 552,000;

fourth, Russia: 365,000 and fifth, Mexico with 324,000 (Google, 2022). The pandemic drove the Mexican economy into a severe recession in 2020. In terms of economic growth e.g., the most affected federal states were Quintana Roo (-24.3%), Baja California Sur (-23.5%), Tlaxcala (-12.1%) and Nayarit (-11.4%). The less affected were Tabasco (3.4%), Baja California Norte (-3.8%) and Chiapas (-3.8%). The national average was -7.9%.

Our analysis aims at enhancing the discussion about the pandemic from the beginning of February 2020 to the first quarter of 2022. We posed the following questions to guide this inquiry: were there excess deaths during the pandemic in Mexico beyond the official number of deaths officially attributed to Covid-19? What insights can be drawn using a different scale, for instance, number of days per 10,000 deaths in relation to the “waves”? what does the number of cases and deaths look like if we use monthly and yearly data in terms of basic statistics parameters? Were Covid-19 deaths concentrated in a certain number of states as the Paretian 80/20 rule dictates, say by filtering the top 13 or applying the Price Square Root Law? We have divided this paper into three sections: methodology, results and discussion, and conclusions. In the first section, we explain the methodology and try to answer the questions mentioned above. We begin by measuring the main changes in deaths from 1994 to 2020 in Mexico outstanding the pandemic’s effects. Second, we describe the four pandemic waves using a scale of number of days per 10,000 deaths to identify the four pandemic waves registered from 27 February 2020 to 20 April 2022. Third, we use graphs to describe the monthly trends of cases and deaths and calculate basic statistic parameters. Here we estimate the slope deaths/cases for each year to clarify the level of fatality along the period. Fourth, we put into practice the Paretian 80/20 rule by filtering the top 13 states and its geographical location to distinguish two regions where Covid-19 struck more intensively. We also apply the Price Square Root Law and proved its validity as well. In section two, we make some comments regarding the analysis and delve into some details related with changes in cases and deaths from 2020 to 2021 focusing on the weirdest cases. Finally, in section three, we synthesize the main findings and emphasize the challenging research questions drawn from this analysis.

## Methodology

### *a) Deaths and births*

Based on official data, a total of 125,807 people died in Mexico from Covid-19, and 1,426,104 contracted the disease in 2020. In 2021, 173,621 people died and 2,553,619 contracted Covid-19. Until March 2022, 323,016 Mexicans had died of Covid-19, and 5,659,535 had been infected.

INEGI, an official source of statistics in Mexico (Instituto Nacional de Geografía y Estadística, 2022), publishes the total number of deaths - of all causes- countrywide on a yearly basis. As shown in Figure 1, the yearly total deaths increased from 419,074 in 1994 to 747,784 in 2019, at a yearly average rate of 2.16%. It also depicts the balance births minus deaths. This balance has been positive but shows an important downward trend. It went in the same period from 2,485,315 to 1,344,430, a drop of 46%. In 2020, there were 1,629,211 births and 1,086,743 deaths making a record low positive difference of 542,468. That constitutes a severe setback in Mexican demographics. The number of births has been sinking at an average of 49,000 a year during this interval. If this continues, by 2034 the balance will be zero and henceforth the number of deaths will surpass the number of births accelerating the Mexican demographic winter (Larrumbe, 2018).

On the other hand, if Covid-19 had never occurred, the expected number of deaths in 2020 would have been 763,936. INEGI reported 1,086,743 with which we theoretically obtain 322,807 deaths in excess. According to the authorities 125,807 died of Covid-19 and 36,597 more were victims of violence; that makes a total of 162,404. Thus, the rest (160,403 deaths) is still to be clarify but presumably can be attributable to Covid-19 side effects as well.

### *b) The four pandemic waves*

Figure 2 displays the number of days per 10,000 Covid-19 deaths. The pandemic in Mexico began with the first case officially on February 27<sup>th</sup>, 2020, and the first death was registered on March 18<sup>th</sup>, 2020. Following figure 2, we can observe that it took 75 days to accumulate the first 10,000 deaths. During the first wave it took only 15 days on average to fill out an interval of

**The 15th De La Salle University Arts Congress  
March 29-31, 2022**

10,000 deaths; the second wave was the sharpest since it took only 8 days on average to accrue 10,000 deaths; 13 days in the third wave and 27 days in the fourth wave.

Furthermore, if we segment the total number of deaths in milestones of 100,000, we have that on November 19<sup>th</sup>, 2020, the tally marked the first 100,000 deaths, 8 months and 1 day or 35 weeks after the first death. The second milestone, 200,000 deaths, occurred on March 25<sup>th</sup>, 2021, and it only took 4 months and 6 days or 18 weeks. The third milestone, 300,0000 deaths, was scored on January 12<sup>th</sup>, 2022, and took 9 months, 18 days, or 41 weeks. The main peaks in Figure 2 were marked by gathering 10,000 deaths in 8 days as the cumulative deaths went from 150,000 and 170,000. The lowest interval to gather 10,000 deaths took 58 days on average to make a total of 300,000 deaths on January the 12<sup>th</sup> 2022.

### *c) Basis Statistics*

For this EDA, we used 666 observations of death and cases collected daily from February 27<sup>th</sup>, 2020 to April 20<sup>th</sup>, 2022. Figure 3 plots the pandemic's four spikes. In 2020, the first peak was detected in July 2020, with 18,919 deaths and 197,980 cases, a ratio of 10%. The second peak reached the ceiling in January 2021 when 32,729 people died, and 438,156 contracted the disease, a ratio of 7.4%. The third pandemic's peak took place in August 2021, with 18,420 deaths and 496,986 cases, a rate of 3.7%. The number of cases soared to 962,867 in January 2022, but the number of deaths was modestly 6,663, making a ratio of barely 1%. This ratio, also known as the fatality rate, reached its record in June 2020 at 13%; it then went down to 6% in December 2020 and took off again to reach 13% in April 2021; henceforth, it kept constantly falling, indicating that the fatality of the pandemic was leveling off. Since the beginning of 2022, the fatality rate decreased, and it was believed that Covid-19 had become a sort of regular flu.

Following table 1, we see that 125,807 people died in 2020; 173,621 in 2021 and 24,576 until the 1<sup>st</sup> of April 2022. The minimum number of deaths in 2020 was recorded in Baja California Sur, 775 while the maximum in Mexico (officially known as Estado de Mexico), 18,878. The mean was 3,931 and the standard deviation, 4,017. The data

shows a right skewed histogram. During 2021 the minimum was registered in Chiapas: 1096 deaths and the maximum in Mexico: 25,957. The mean was 5,426 and the standard deviation, 5,755. Regarding cases, the same parameters were 7,259 in Campeche and 330,000 in Mexico City and 15,357 in Chiapas and 670,000 in Mexico City during 2020 and 2021 respectively. Total Covid-19 deaths increased by 38%, whereas Covid-19 cases by 79%. These parameters prove that 2021 was worse than 2020, the year in which the pandemic emerged.

Following figure 4, a simple regression model shows that for every 100 cases there were 9.3 deaths in 2020; 6 deaths for every 100 cases in 2021, and 1.3 deaths for every 100 cases in 2022. That means that a during 2020 positive Covid-19 cases were highly lethal and probably had to do with the poor performance of PCR-tests reflected in an imprecise estimation of cases. In 2021, the ratio diminished slightly as the number of cases increased more rapidly than the number of deaths. In 2022, we can see a vigorous spike in the number of cases, but not in the number of deaths. That places us in the fourth wave of Covid-19 with the omicron variant which was highly contagious but not so lethal. The maximum number of deaths reported in a single day in 2020 was 2,789 on October 5<sup>th</sup> with 28,115 cases (ratio: 10%); 4,272 and 6,917 (ratio: 62%) on June 1<sup>st</sup> 2021 and 2,070 and 91,844 (ratio: 2%) February 13<sup>th</sup> 2022 (Data until April 20<sup>th</sup> 2022 under scrutiny). The authorities emphasized that those reports did not mean that so many people had died in a single day. They were the results of adjustments due to pending reports not registered on time. However, data shows 15 times that authorities reported deaths over 1,500 in a single day: 1 time in October 2020; 6 times in January 2021; 3 times in February 2021; 1 time in April 2021; 1 time in June 2021; 1 time in July 2021 and 2 times in February 2022.

### *d) The top 13 and the Price Square Root Law*

Figures 5a (absolute numbers) and 5b (share) depict that Mexico's top 13 federal states concentrated in 2020 and 2021, 72% and 73% of total Covid-19 deaths, respectively. The states that integrated the top 13 in 2020 were the same in 2021, except for Coahuila and Michoacán. Michoacán was not in the top 13 in 2020 but replaced Coahuila in 2021. Coahuila reported a

The 15th De La Salle University Arts Congress  
March 29-31, 2022

decrease in deaths from 4,181 to 3,626 (-13%), whereas, in Michoacán, deaths soared 100%, going from 2,757 to 5,517. Baja California's share fell from 0.04 in 2020 to 0.03 in 2021 because deaths diminished from 5,524 in 2020 to 5,503 in 2021, whereas Veracruz kept its share by an increase of 38%, going from 6,339 to 8,760 deaths (see figure 5b).

The top 13 states appearing in figure 5a form two regions where most of the economic activity occurs (Batalla, 1980). These 13 states comprise 64% of the population; 53% of the national territory; 70% of the GDP; 59% of the violent homicides in 2020; 54% of the migrant remittances, and 74% of the Foreign Direct Investments. The upper region is located along the border with the USA and is well known for its dynamism in trade and investment. The middle region embraces five emblematic production poles as well: Jalisco, Mexico, Mexico City, Veracruz, and Puebla. Thus, we are facing a typical concentration case that puts us face to face with a 20/80 Pareto Rule. In this context, the Price Square Root Law would affirm that *half of the total Covid-19 deaths should have happened in the square root of the total number of states* (Nicolls, 1988). Using this Law, the square root of 32 (states) is 5.6. Rounding that number, 6 federal states should have embraced 50% of Covid-19 deaths. Doing the math, we found that out of 125,807 deaths in 2020, 6 states accounted for 47%: Estado de Mexico (15%), Mexico City (13%), Veracruz (5%), Jalisco (5%), Puebla (5%) and California North (4%). In 2021, with 173,621 deaths, 6 states were responsible for 51% of the deaths: Estado de Mexico (15%), Mexico City (14%), Jalisco (7%), Puebla (6%), Veracruz (5%) and Nuevo León (5%). Therefore, the Price Square Root Law proves to be valid in this case.

## Results and Discussion

To begin with, this concentration derived from the 80/20 Paretian Rule and Price Square Root Law could be explained by resorting to the fundamentals of localization theory, where agglomeration and centripetal forces play an important role in forming regions (Quintana-Romero, 2013; Gutiérrez Casas, 2006). The classical authors of such theory point out that most of the market interactions take place in the center of a region. This exerts an attraction force among the people due to advantages in production's costs (Thunen, 2009; Christaler,

1980). Secondly, it seems intuitive to have found a correlation between Covid-19 deaths 2020 vs 2021. In this context, the states with the highest increase in deaths from 2020 to 2021 in relative terms were Morelos (142%), Baja California Sur (113%), and Michoacán (100%). In absolute terms, the states with the most significant increment in deaths during the same period were Estado de México (17,079), CDMX (6,820), and Jalisco (5,740).

Estado de México and CDMX have the largest number of inhabitants per square kilometer, 796 and 6,085 respectively. Ergo, these two places represent a potential risk for future pandemics. There were also federal states which reported a significant increase in positive cases but not in deaths and vice versa, a significant increase in deaths but not cases. For instance, in Quintana Roo the cases increased from 15,855 in 2020 to 46,221 in 2021 (191%), but deaths only augmented by 2% going from 2,028 to 2,073. The same was observed in Tabasco, where cases soared from 44,248 to 100,752 (128%), but deaths fell from 3200 to 2559 (-20%). Conversely, in Michoacán, cases increased from 33,104 to 40,849 (23%), whereas the number of deaths soared from 2,757 to 5,517 (100%). These peculiarities may have reflected flaws in the official statistics reports or were possibly the results of public policies put into practice by the local authorities.

## Conclusions

We found the following answers to the questions posed at the beginning. Firstly, there were mismatches in the way authorities presented the data regarding Covid-19. More people died of all causes than the number officially attributed to Covid-19 and homicides. We believe that the excess of deaths has to do with illnesses related to the pandemic. The demographic statistics for 2021 have not yet been published as this essay goes to print. In this context, the constant reduction of the yearly balance of births minus deaths should be addressed by the coming public policies, since it represents the beginning of a demographic winter, the way West Europe is already experimenting it (Larrumbe, 2018). Secondly, we could observe the intensity of the pandemic, and the waves by using a scale of number of days per 10,000 deaths. In this context, we could identify two episodes of a death toll at a record velocity of



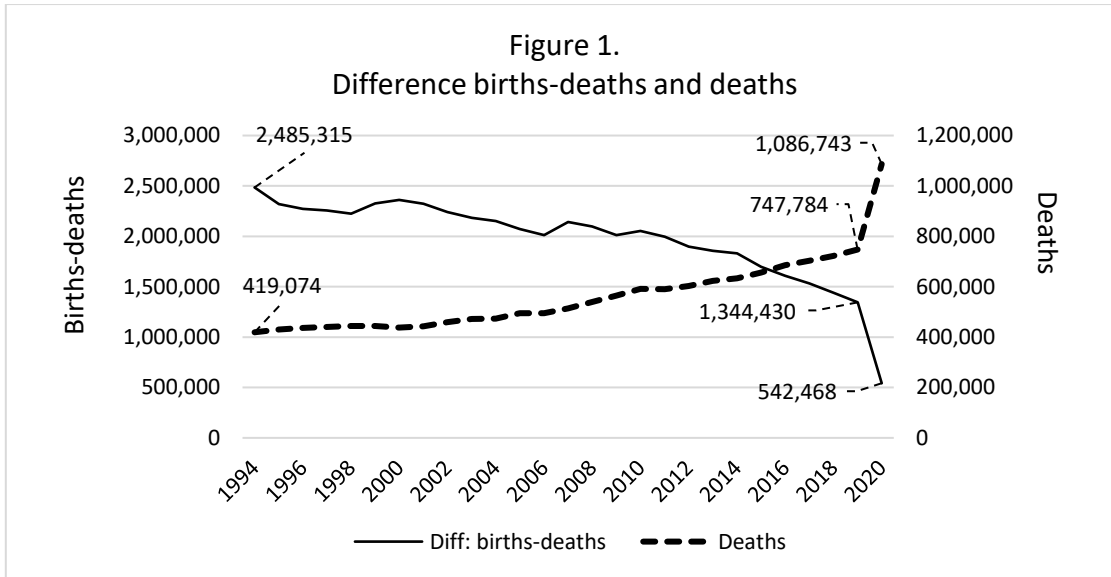
8 days per 10,000 people as figure 2 indicates. Thirdly, by tracing the monthly cases and deaths over the two years, we realized that most of the fatalities happened in 2021, whereas the fatality rate reached a ceiling of 13% in June 2020 and April 2021, respectively. The fatality rate kept falling and touched the bottom at the beginning of 2022, sealing the end of the pandemic. Fourthly, 13 Federal States accounted for more than 70% of the deaths forming two regions with key socioeconomic indicators. The Price Square Root Law also proved to work as we found that 47% of the deaths in 2020 and 51% of the deaths in 2021 occurred in 6 states. This concentration phenomenon matches the fundamentals of the Localization Theory, which is based on the existence of agglomeration and centripetal forces defining regions. This can be used as a guide to distribute rationally scarce resources to attend pandemics like Covid-19, then had these resources been conveyed to such regions, the number of deaths could have been diminished. We must also emphasize that the two states with the worst economic performance in 2020 were neither in the top 13 nor in the top 6 but constitute the backbone of Mexican tourism; they were Baja California Sur (-23.5) and Quintana Roo (-24.1). In other words, the economic sector most affected by the pandemic was Services, which mainly comprises restaurants and hotels. Finally, two questions can be posed to continue this research, why did the authorities report fewer deaths of Covid-19 than the total mortality in 2020, and why did deaths fall in 2021 in states like Coahuila and California but increase strongly in others like Michoacán and Morelos?

## References

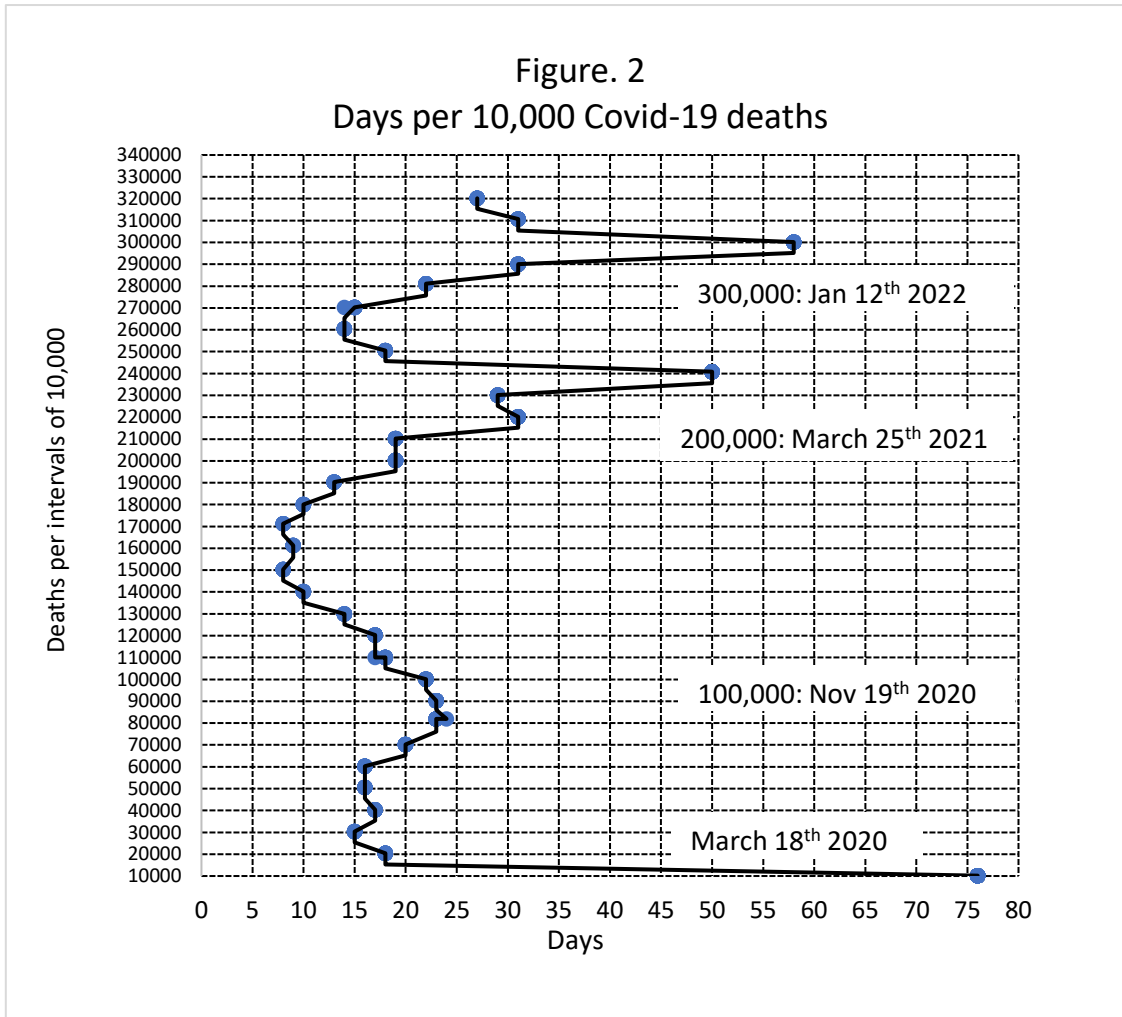
- Batalla, A. B. (1980). *Geografía Económica de México, teoría, fenómenos generales, análisis regional*. México: Trillas.
- BBVA Research. (2018). *Situación Regional Sectorial Mexico*. CDMX: BBVA Research.
- Christaler, W. (1980). *Die zentralen Orte in Süddeutschland : Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Berlin: WBG Academic.
- Gerardo Reyes Guzman, M. A. (2021). Covid-19 Deaths in Mexico: Exploratory Data Analysis. *DLSU Congress, Vol. 5*.
- Google. (10 de abril de 2022). *Enfermedad por el nuevo coronavirus*. Obtenido de Estadísticas: <https://www.google.com/search?q=muer+tes+por+coronavirus+en+el+mundo&oq=muer+tes+por+coronavirus+en+el&aqs=chrome.0.0i512j69i57j0i512l8.55317j1j15&sourceid=chrome&ie=UTF-8>
- Gutiérrez Casas, L. E. (2006). Teorías del crecimiento regional y el desarrollo divergente. Propuesta de un marco de referencia. *Revista de Ciencias Sociales y Humanidades*, 185-227.
- Instituto Nacional de Geografía y Estadística. (5 de February de 2022). *Demografía y Sociedad*. Obtenido de Mortalidad: <https://www.inegi.org.mx/temas/mortalidad/>
- Larrumbe, A. M. (2018). *Demographic suicide in the western hemisphere and half the world*. Columbia: SC.
- Murray, C. J. (2022). Covid-19 will continue but the end of the pandemic is near. *The Lancet*, [https://doi.org/10.1016/S0140-6736\(22\)00100-3](https://doi.org/10.1016/S0140-6736(22)00100-3).
- Nicolls, P. T. (1988). Price's square root law: Empirical validity and relation to Lotka's law. *Information Processing & Management*, 469-477.
- Quintana-Romero, L. A.-S. (2013). Growth, Concentration, Inequality and Regional Policy in Mexico. En J. A. Cuadrado-Roura, *Regional Problems and Policies in Latin America. Advances in Spatial Science* (págs. [https://doi.org/10.1007/978-3-642-39674-8\\_8](https://doi.org/10.1007/978-3-642-39674-8_8)). Berlin: Springer.
- Relaciones Exteriores. (10 de April de 2022). *Secretaría de Relaciones Exteriores*. Obtenido de Mexico's Covid-19 Monitoring System: [embamex.sre.gob.mx](http://embamex.sre.gob.mx)
- Thunen, J. V. (2009). *The Isolated State in Relation to Agriculture and Political Economy*. Basingstoke, United Kingdom: Palgrave Macmillan.
- Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 1-67.



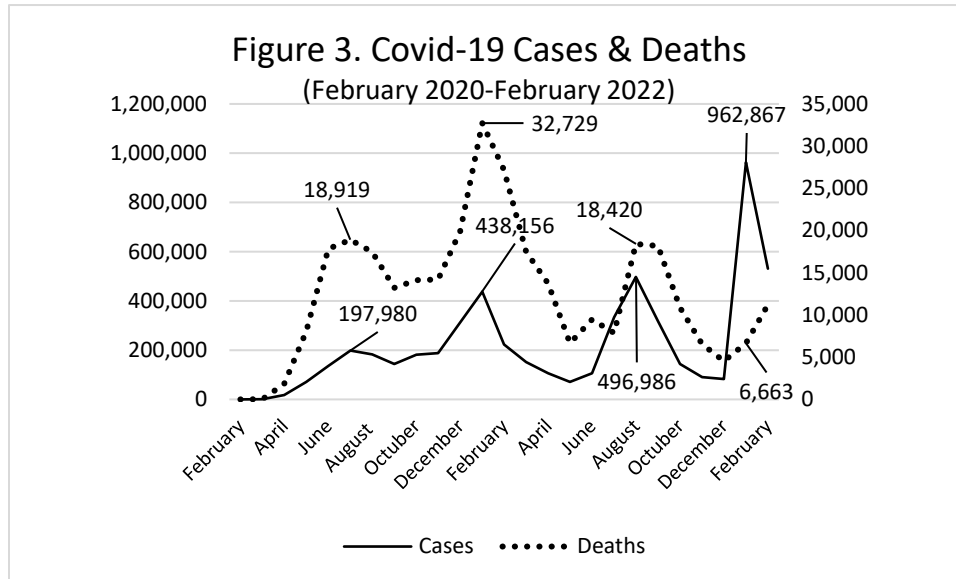
**APPENDIX**



Source: Own calculations with data of INEGI (Mexico's National Institute of Geography and Statistics)



Source: Own calculations with data of (Google, 2022)



Source: Own calculations with data of (Google, 2022)

**Table 1. Mexico’s 32 States: Covid-19 Statistical parameters**

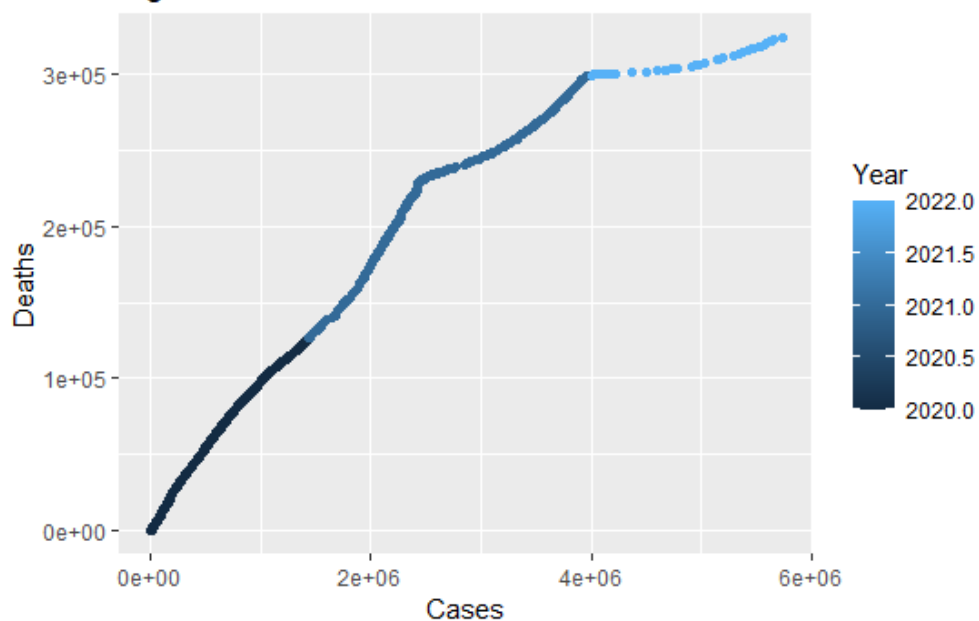
Parameters	2020		2021		Δ%	
	Deaths	Cases	Deaths	Cases	Deaths	Cases
Min.	775	7,259	1,096	15,357	41%	112%
1 <sup>st</sup> Qu.	1,520	17,498	1,818	35,950	20%	105%
Median	2,908	30,347	3,697	50,229	27%	66%
Mean	3,931	44,551	5,426	79,708	38%	79%
3 <sup>rd</sup> Qu.	4,602	45,007	5,506	73,398	20%	63%
Max.	18,878	330,000	25,957	670,000	37%	103%
Std. Dev	4,017	58,997	5,755	116,334	43%	97%
Total	125,807	1,426,124	173,621	2,554,619	38%	79%

Source: Compiled by Gerardo Reyes based on (Google, 2022).





**Figure 4. Covid-19: Correlation Deaths vs Cases**



Source: Compiled by Gerardo Reyes based on (Google, 2022)

**Figure 5a. Top 13 States affected by Covid-19 in 2020 and 2021  
(Absolute numbers)**

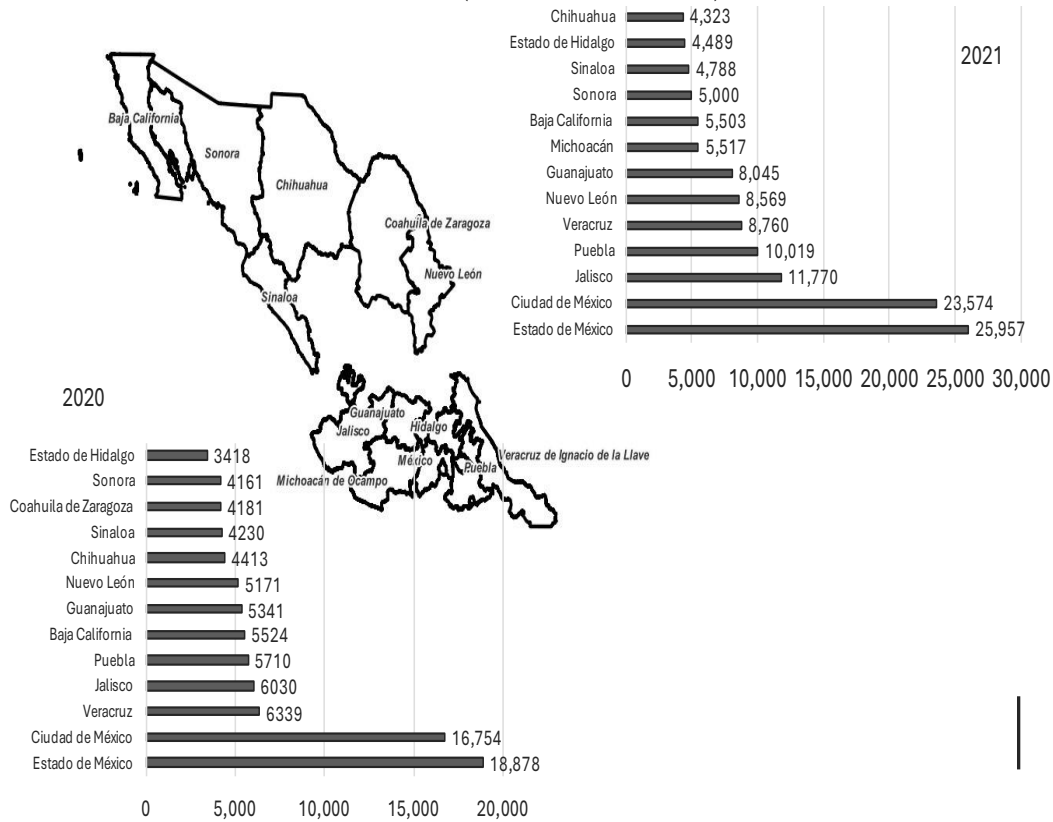
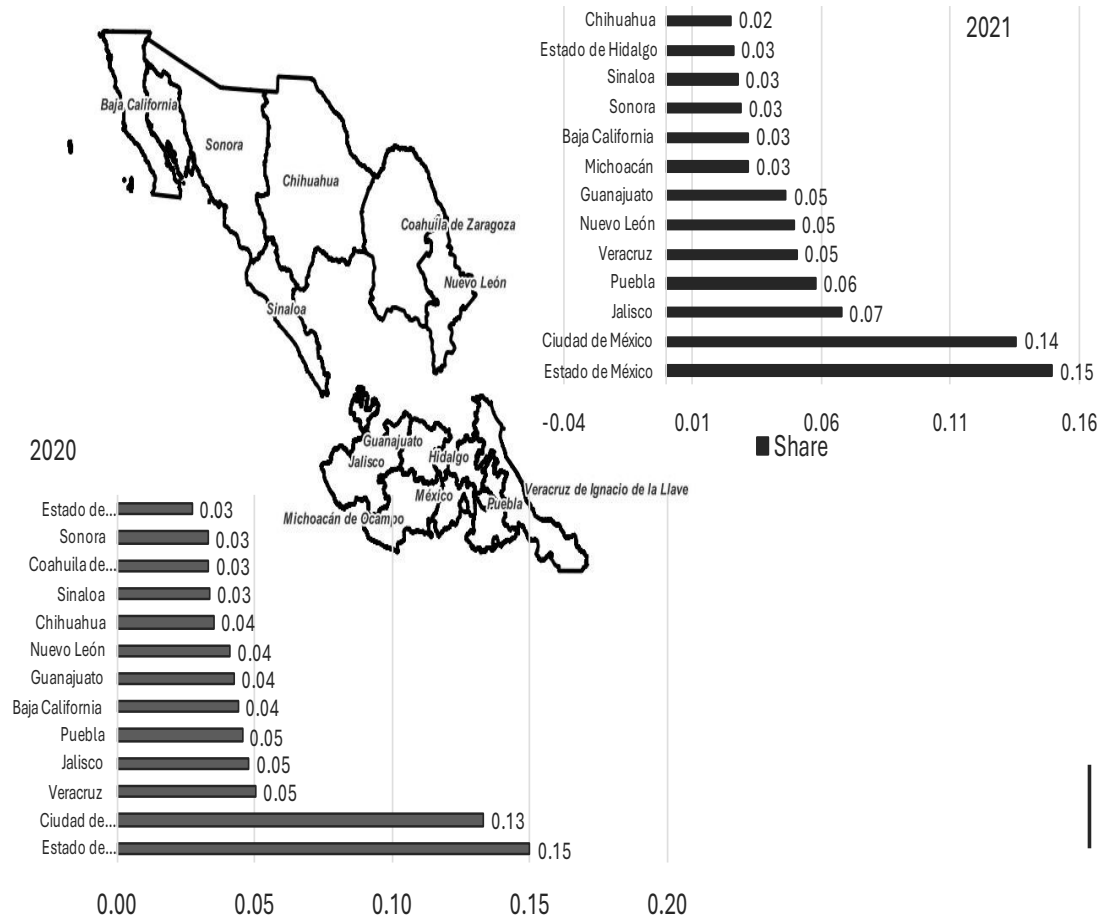


Figure 5b. Top 13 States affected by Covid-19 in 2020 and 2021  
(Share %)



Source: Own calculations with data of INEGI