Contents lists available at ScienceDirect

Data in Brief

Data Article

Spatio-temporal dataset of COVID-19 outbreak in Mexico

Jean-François Mas

Laboratorio de análisis espacial, Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de La Huerta, C.P. 58190 Morelia Michoacán, México

ABSTRACT

Our understanding of how COVID-19 spreads over a territory needs to be improved. For example, the evaluation of disease spatiotemporal distribution and its association with other characteristics can help identify covariates, model the behavior of the epidemic, and provide useful information for decision making. Data were compiled from the National Population Council (CONAPO), Google, the National Institute of Statistics and Geography (INEGI), and the Secretary of Health. The data describe the cases of COVID and characteristics of the population, such as distribution, mobility, and prevalence of chronic diseases such as diabetes, hypertension, and obesity. These data were processed to be compatible and georeferenced to a common geographic framework to facilitate spatial analysis in a geographic information system (GIS).

© 2021 Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

E-mail address: jfmas@ciga.unam.mx

Specifications Table

Subject	Infectious Diseases
Specific subject area	Georeferenced data related to COVID-19 confirmed cases and deaths, community mobility, demographic characteristics, and pre-existing diseases.
Type of data	Tables in comma-separated values (CSV) format
	Digital maps (shapefiles)
How data were acquired	The original datasets were downloaded from the official websites of Google Community Mobility Reports, CONAPO, INEGI, and the Secretary of Health.
Data format	Raw
	Filtered
Parameters for data collection	Relevant data to evaluate the spatiotemporal patterns of the COVID-19 pandemic in Mexico and possible co-variables that may explain these patterns were collected.
Description of data collection	The Mexican Secretary of Health daily collects epidemiological data about COVID-19.
	The 2010 municipal population was obtained through the INEGI population census, while CONAPO estimated the 2020 population through a projection based on the population growth trends in previous years and the population's structure by age groups and sex [1].
	The prevalence of obesity, hypertension, and diabetes diseases per municipality, was estimated by INEGI from the National Health and Nutrition Survey (ENSANUT) [2].
	Google elaborated community Mobility Reports with aggregated and anonymized data from the user's location history information.
Data source location	Institution: Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México (UNAM)
	City/State: Morelia, Michoacán
	Country: Mexico
	The dataset is hosted at Mendeley Data (see Data accessibility)
	Primary data sources:
	Epidemiological data - Mexican Secretary of Health http://datosabiertos.salud.gob.mx/gobmx/salud/datos_abiertos/datos_abiertos_covid19.zip
	Municipality maps INEGI https://www.inegi.org.mx/temas/mg/
	Population census 2010 INEGI https://www.inegi.org.mx/programas/ccpv/2010/default.html#Datos_abiertos
	Population 2020 projection CONAPO https://datos.gob.mx/busca/dataset/proyecciones-de-la-poblacion-de-mexico-y-de-las-entidades-federativas-2016–2050
	Prevalence of obesity, hypertension and diabetes diseases by municipality INEGI https://www.inegi.org.mx/investigacion/pohd/2018/
	Community Mobility Reports https://www.google.com/covid19/mobility/?hl=en
Data accessibility	The dataset is hosted at Mendeley Data
	Repository name: Data_Mexico_COVID19
	Data identification number: 10.17632/mc37xdzw74.1 (DOI)
	Direct URL to data: https://data.mendeley.com/datasets/mc37xdzw74/1
Related research article	Mas, J-F., Stage 1 Registered Report: Spatiotemporal patterns of the COVID-19 epidemic in Mexico at the municipality level, PeerJ, 9:e10622 https://doi.org/10.7717/peerj.10622

Value of the Data

- Pandemic is a spatial process. Thus, uncovering spatial patterns and associations is useful to bring new insights into its spread and how to fight it.
- This dataset can be used by other researchers to apply spatial analyses to assess the pandemic spreading patterns. Decision-makers can benefit from these data to define policies.
- This dataset can be used to assess the pandemic spreading patterns and assess covariates' role as age or chronic diseases. For instance, the distribution of population with chronic diseases can help identify vulnerable regions and the community mobility data in the assessment and tuning of social distancing strategies.
- These data can also be employed with other datasets as sociodemographic data from INEGI to assess COVID-19 impact.

1. Data Description

GIS layers (shapefiles)

- Municipalities_COVID.shp: Map (shapefile) with weekly COVID indices at the municipality level: field new1 to new49 indicates the number of weekly confirmed new cases (week 1 to 49 of year 2020). The other weekly COVID indices are organized in the same manner. Field cumul indicates the number of weekly confirmed cumulative cases; field activ the number of weekly actives cases; field death the number of weekly deaths; field actr the weekly rate of active cases (number per 100,000 inhabitants) and, field deatr the weekly death rate (number of deaths per 100,000 inhabitants). Additional fields are CVEGEO (unique municipality' code), CVE_ENT (State number), CVE_MUN (Municipality'number), NOMGEO (Municipality name, Pop2010 (Municipality 2010 population from INEGI census) and, Pop2020 (Municipality 2020 population from CONAPO projection).
- Municipalities_chronic_diseases.shp: Map of the proportion of inhabitants above 20 years with obesity, hypertension, and diabetes at the municipal level.
- States_mobility.shp: Map of mobility reduction at the state level (weekly average) in various categories of places (shops and leisure spaces, supermarkets and pharmacies, parks, transport stations, workplaces, and residential areas).

The GIS layers are based on the Lambert conformal conic projection, a conic map projection which minimizes deviation from the unit scale within a region comprising the two standard parallels. The first two shapefiles, aggregated at the municipality level, are based on the most recent municipalities' boundaries (2019 with 2465 municipalities).

Tables (CSV files)

- datosCOVID.csv: pre-processed raw table with daily information with additional fields as the day and week of the year, and the municipality code.
- tab_new.csv, tab_culum.csv, tab_activ.csv, tab_death.csv, tab_activrate.csv, and, tab_deathrate.csv. Tables of the weekly number of confirmed cases, accumulated cases, active cases, deaths, active cases rate, and death rate per municipality with 2010 and 2020 population.
- Population2020_CONAPO.csv: Table that shows population structure (sex, quinquennial age groups) per municipality.
- pop2010–2020.csv: 2010 and 2020 municipality populations based on updated (2019) municipality boundaries (total of 2465 municipalities).
- Edo2CVE_ENT.csv: Table of equivalent names of the Mexican states in INEGI data and Google reports.

R code

- Elaborate_table_covid.R: Script used to download and process raw epidemiological data from the Secretary of Health. It produces the tables datosCOVID.csv, tab_new.csv, tab_culum.csv, tab_activ.csv, tab_death.csv, tab_activrate.csv and, tab_deathrate.csv and the shapefiles municipalities_new.csv, municipalities_culum.csv, municipalities_activ.csv, municipalities_death.csv, municipalities_activrate.csv and, municipalities_deathrate.csv.
- Elaborate_table_preexsisting.R: Script used to process data from INEGI on chronic diseases. It produces the GIS layer municipalities_chronic_diseases.shp.
- Elaborate_map_mobility.R: Script used to process data from Google mobility reports. It filters data by selecting information for Mexico, calculates weekly average mobility change and produces the shapefile states_mobility.shp.

2. Experimental Design, Materials and Methods

The main problem in elaborating compatible spatial data was that the different sources used different municipal references due to the creation of new municipalities during the last decade.

The 2010 population census employs 2456 municipalities, the 2020 projected population (elaborated in 2015) 2457, the chronic diseases database 2458, and the epidemiological data 2465.

The information of the 2010 Census at the settlement level was obtained from INEGI. It is a CSV table that contains the name of the settlement, the code of the State and the municipality, the geographical coordinates (in concatenated format), and a large number of population attributes. The coordinates were used to create a point GIS layer, which was overlaid with the maps of the municipal boundaries. This operation allows calculating the 2010 population in the 2019 municipalities and also to know the proportions of the population 'transferred' from one municipality to another one when using versions with a different number of municipalities. These proportions were used to estimate the values of attributes for newly created municipalities. For instance, the 2020 projected population of newly created municipalities was estimated assuming that the redistribution of the population from existing to new municipalities follows the same proportions as observed for the 2010 population. As this assumption can be inaccurate, a flag was assigned to the municipalities which population was adjusted (newly created municipalities or municipalities which a part of the population was assigned to a new municipality). In the case of the map of chronic diseases, the prevalence of disease was adjusted using the same rationale.

The epidemiological data about COVID-19 were downloaded from the government open data web page as a CSV table. For each patient, identified through an anonymized code, it presents the type and the location (State) of the health institution where she/he was treated, the result of the COVID test, sex, age, nationality, ethnicity, the State and municipality of residence, whether hospitalization was required, whether an intubation was required, the date of admission to the care unit, the date of onset of symptoms and eventually the date of death. The table also indicates a diagnosis of pneumonia, case of pregnancy, and the existence of a pre-existing disease such as diabetes, chronic obstructive pulmonary disease (COPD), asthma, immunosuppression, hypertension, cardiovascular diseases, obesity, chronic kidney failure, as well as a smoking habit. It also indicates whether the patient had contact with any other case diagnosed with SARS CoV-2.

The dates were converted into the day and week of the year to make the selection of specific periods easier. For each municipality and each week, the number of new confirmed cases, cumulative cases, active cases, deaths, rate of active cases, rate of death was computed. Both rates are based on the projected municipal 2020 population and are expressed as the number of cases per 100,000 inhabitants. The number of active cases is based on the assumption that it takes, on average, two weeks to recover from an infection. These different indices were saved as tables and GIS layers (shapefile). Concerning the georeferencing of epidemiological data, it is worth noting that the association between the cases and the municipalities is based on the municipality of residence. However, the State of residence and the State where hospitalization occurred are different in 5.9% of the confirmed cases and 6.7% of the deaths.

COVID-19 community mobility reports were filtered to Mexico. The weekly values of percent change from baseline were averaged for each State. State names were normalized to use them as merging keys when joining the mobility data with the attribute table of the map of State boundaries from INEGI.

The entire processing was carried out using the R program [3].

Ethics Statement

All data presented in this document is anonymized and aggregated at municipality or state level.

CRediT Author Statement

J-F. Mas: Conceptualization, Methodology, Software, Writing, Reviewing and Editing.

Declaration of Competing Interest

I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] CONAPO, 2019. Nota técnica: Proyecciones de La Población De Los Municipios De México 2015-2030. https://www.gob.mx/cms/uploads/attachment/file/487810/Nota_Tecnica_Proyecciones_Municipales_2019_01.pdf (Accessed 8 August 2020).

[2] INEGI, 2020. Prevalencia De Obesidad, Hipertensión y Diabetes Para Los Municipios de México 2018 Estimación para Áreas Pequeñas, Nota metodológica, 51 p. https://www.inegi.org.mx/contenidos/investigacion/pohd/2018/doc/a_peq_2018_nota_met.pdf (Accessed 8 August 2020).

[3] R Core TeamR: A language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020 https://www.R-project.org/.