

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE POSGRADOS**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN ESTADÍSTICA APLICADA”

TEMA:

MODELIZACIÓN ESTADÍSTICA MULTIVARIANTE PARA IDENTIFICAR
VARIABLES ATMOSFÉRICAS Y SOCIODEMOGRÁFICAS, ASOCIADOS A LA
COVID-19 EN LA CIUDAD DE GUAYAQUIL

AUTOR:

JHON JAIRO PEÑA BENÍTEZ

Guayaquil - Ecuador

2022

RESUMEN

En el año 2019 apareció un nuevo virus que se lo conocería como COVID 19, este virus se propago por todo el mundo, llegando a casi todos sus rincones, una ciudad se vio muy afectada por este virus la cual es Guayaquil. Es por lo que se decidió analizar de manera multivariante las variables que posiblemente puedan tener incidencia en la propagación del virus, en estas están: atmosféricas (precipitación, temperatura), sociodemográficas (nivel socio económico, densidad poblacional, nivel de instrucción y servicios básicos). Además de estas se utilizó varias bases de datos para espacializar los casos de COVID 19. Se hizo una depuración, limpieza, unión, espacialización de estas bases, para poder hacer los análisis multivariantes respectivos. En el modelamiento se usó análisis de componentes principales, regresión lineal múltiple, inferencia bayesiana con el paquete R INLA. Dando como resultado mapas de zonas (en este caso circuitos) las cuales tuvieron un SIR y un RR alto en la pandemia. Se pudo observar que las variables atmosféricas tenían menos aportación a los modelos que las variables sociodemográficas, las cuales nivel de instrucción y nivel socio económico son las que más tienen una posible influencia como covariables en el contagio del COVID 19.

ABSTRACT

In the year 2019 a new virus appeared, known as COVID 19, this virus spread throughout the world, reaching almost every corner of it, one city was greatly affected by this virus which is Guayaquil. That is why it was decided to analyse in a multivariate way the variables that could possibly have an impact on the transmission of the virus; these are: atmospheric (precipitation, temperature), socio-demographic (socioeconomic level, population density, level of education and basic services). In addition to these, several databases were used to spatialize the COVID 19 cases. These databases were debugged, cleaned, merged and spatialized in order to perform the respective multivariate analyses. Principal component analysis, multiple linear regression and Bayesian inference were used in the modeling with the R INLA package. Resulting in: maps of areas (in this case circuits) that had a high SIR and a high RR in the pandemic. It could be observed that the atmospheric variables had less contribution to the models than the sociodemographic variables, which education level and socioeconomic level are the ones that have more possible influence as covariates in the contagion of COVID 19.

DEDICATORIA

A mi Padre, Madre y Hermana. Julio, Mary y Stefania por el apoyo en todo.

AGRADECIMIENTO

A mi Padre, Madre y Hermana. Julio, Mary y Stefania por el apoyo en todo.

A Sergio, Jhony por guiarme en este camino.

A ESPOL por crear este programa a distancia y pude cumplir mi sueño de estudiar estadística.

A mis compas por colaborar en todo lo que podían durante este periodo de estudio.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Jhon Jairo Peña Benítez

TRIBUNAL DE GRADUACIÓN

OMAR RUIZ, Ph. D.
PRESIDENTE

SERGIO BAUZ, Ph. D.
TUTOR

FRANCISCO MOREIRA, Msc
DOCENTE EVALUADOR

ABREVIATURAS O SIGLAS

Siglas	Abreviatura
CV19	Covid 19
DP	Densidad poblacional
GADMG	Gobierno Autonomo Descentralizado Municipal de Guayaquil
IGM	Instituto geografico Militar
INLA	integrated Nested Laplace Approximation
INOCAR	Instituto Oceanografico y Antartico de la Armada
NI	Nivel de instrucción
OMS	Organizacion Mundial de la salud
PCA	Analisis de Componentes Principales
RR	Riesgos Relativos
SB	Nivel de servicios basicos
SE	Nivel socio economico
SENPLADES	Secretaria Nacional de Planificación y Desarrollo
SIR	Standarized Incidence Ratio

TABLA DE CONTENIDO

CAPÍTULO 1.....	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes	1
1.2. Descripción del problema.....	3
1.3. Objetivos	7
General	7
Específicos.....	7
1.4. Hipótesis	7
1.5. Alcance.....	7
CAPÍTULO 2.....	8
2. MARCO TEÓRICO.....	8
2.1. ANÁLISIS MULTIVARIANTE.....	8
2.2. GEO CODIFICACIÓN Y GEORREFERENCIACIÓN	9
2.3. DATOS ESPACIALES	10
2.4. R - INLA.....	13
2.5. SISTEMAS DE REFERENCIA DE COORDENADAS	14
2.6. ESTADÍSTICA ESPACIAL	16
2.7. CONCEPTOS DE EPIDEMIOLOGIA	17
2.8. MODELOS ESPACIO TEMPORALES EJEMPLO	17
CAPITULO 3.....	20
3. METODOLOGÍA.....	20
3.1. RECOLECCIÓN DE INFORMACIÓN, (EXTRACCIÓN)	20
3.2. TRATAMIENTO DE INFORMACIÓN (TRANSFORMACIÓN).....	22
3.3. MODELADO ESPACIAL (CARGAR).....	24
CAPITULO 4.....	26
4. RESULTADOS.....	26
4.1. BASE DE DATOS DEPURADA	26
4.2. ESTADÍSTICA DESCRIPTIVA	29
4.3. VARIABLE ATMOSFÉRICA.....	31
4.4. VARIABLE SOCIODEMOGRÁFICAS	33
4.4.1. Comparación 2020 y 2021	46
4.5. DASHBOARD	47
CAPÍTULO 5.....	49

5.1. CONCLUSIONES.....	49
5. Referencias	1

LISTADO DE FIGURAS

Figura 1	3
Figura 2	11
Figura 3	12
Figura 4	13
Figura 5	15
Figura 6	16
Figura 7	18
Figura 8	19
Figura 9	22
Figura 10	23
Figura 11	24
Figura 12	27
Figura 13	28
Figura 14	36
Figura 15	39
Figura 16	40
Figura 17	43
Figura 18	45
Figura 19	46
Figura 20	48
Figura 21	48

LISTADO DE TABLAS

Tabla 1	14
Tabla 2	29
Tabla 3	29
Tabla 4	30
Tabla 5	30
Tabla 6	31
Tabla 7	32
Tabla 8	32
Tabla 9	33
Tabla 10	34
Tabla 11	37
Tabla 12	38
Tabla 13	41
Tabla 14	42
Tabla 15	44
Tabla 16	47
Tabla 17	47

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes

La pandemia por Covid-19 es un hito histórico que ha marcado un antes y un después en los sistemas de salud de todo el mundo y ha generado una crisis económica, política y social sin precedente (Takashi, et al, 2020). La enfermedad por Covid-19, se trata principalmente de una neumonía e infección respiratoria producida por una cepa mutante de coronavirus conocida como SARS-CoV-2, de contagio masivo y de rápida transmisión de una persona a otra mediante las secreciones respiratorias, tos, estornudo o contacto común de las superficies contaminadas seguido del contacto de nariz, boca y ojos (Fernández, 2020). Su primera aparición fue en China, específicamente en la provincia de Hubei en diciembre del 2019, en donde solo se registró un aumento de pacientes con enfermedad respiratoria, pero hasta entonces se desconocía su etiología y los signos y síntomas asociados y más comunes a esta enfermedad (Žižek, 2020).

Posterior a ello, el gobierno de China invirtió innumerables esfuerzos para frenar la propagación, pero fueron insuficientes ya que el virus se esparció rápidamente por todo el mundo (Maguiña, Gastelo, & Tequen, 2020). Por esta razón, la Organización Mundial de la Salud (OMS) la declaró pandemia mundial el 11 de marzo del 2020, aludiendo a que se trata de un virus que tiene tendencia a afectar el sistema respiratorio y una vez que ingresa al organismo de la persona provoca una respuesta inmune anormal que afecta a otros órganos del cuerpo y que es letal en la mayoría de los casos (OMS, 2020)

Por tal motivo, los países más afectados por el virus como España, Italia y Estados Unidos y otros países del mundo aplicaron una serie de medidas de bioseguridad para evitar la rápida expansión de la enfermedad (Cuero, 2020), como fue el uso obligatorio de mascarillas, la desinfección y adecuado lavado de manos de forma recurrente, el uso de trajes de bioseguridad en todos los

hospitales y lugares de mayor concurrencia, el confinamiento obligatorio para actividades de índole laboral y escolar y el distanciamiento social (National Library of Medicine, 2020). Esto permitió que los ciudadanos tengan mayor control sobre las posibles formas de evitar el contagio.

Con respecto a la evolución del virus, al momento se conoce que el periodo de incubación puede durar de entre 4 a 7 días, pero en la mayoría de los casos con un 95% se reportó que la duración puede extenderse hasta los 12 o 14 días (Maguiña, Gastelo & Tequen, 2020). También se conoce que los casos pueden ser asintomáticos con un 80%, es decir que presentan un malestar general con tos ligera, mientras que el 15% constituye un cuadro moderado, con tos seca persistente, fiebre y agotamiento, y por último el 5% forman parte de los casos severos caracterizados por fiebre persistente, fallas en otros órganos del cuerpo, afectación cardiovascular y tos, por lo que entre el 3 y 4% de los afectados fallece (Maguiña, Gastelo, & Tequen, 2020).

Aun así, existen aspectos que todavía no se conocen a profundidad sobre el virus, y su gran tasa de morbilidad y mortalidad obligaron a los países a iniciar la investigación sobre las vacunas de una forma nunca vista, de modo que se pueda acelerar su disponibilidad a todos los países afectados (Cuero, 2020). Actualmente, se cuenta con varias marcas de vacunas autorizadas por la Organización Mundial de la Salud. Entre las cuales están: Pfizer-BioNTech con un 95% de eficacia, AstraZeneca-Oxford con 79%, Janssen/Johnson & Johnson 66%, Sputnik V 92% y Sinovac entre el 50-80% de eficacia (Gaus, 2021). En este sentido, los países se han enfrentado al gran reto de vacunar a la mayor parte de su población para alcanzar una transmisión comunitaria, así como también, evitar el riesgo de que el virus se propague a mayor velocidad o mute a otras variaciones para las cuales las vacunas actuales ya no sean efectivas (Gaus, 2021).

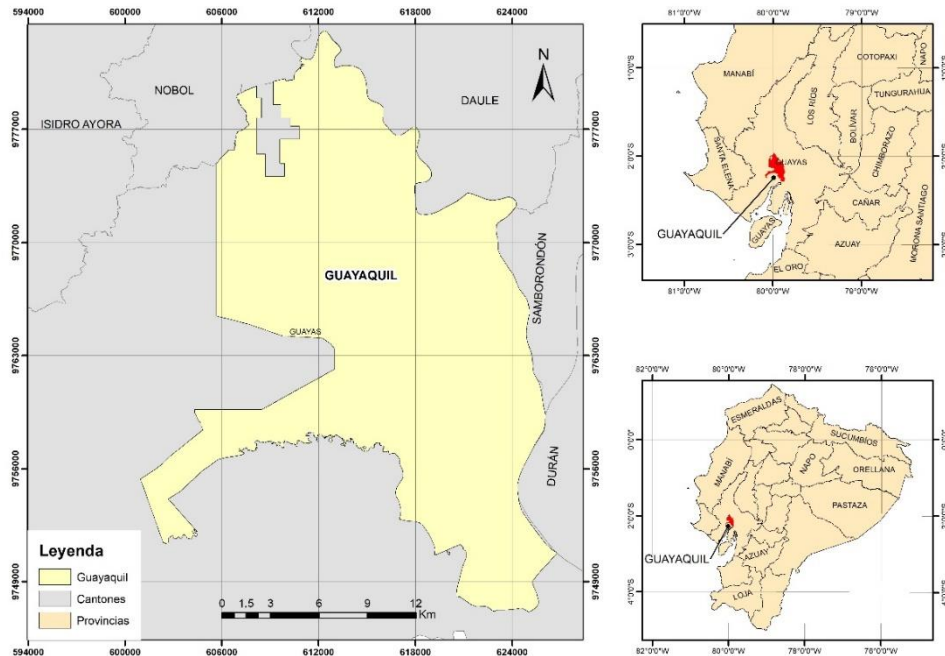
Por otro lado, para el sistema de salud las vacunas solo representan una parte dentro de la batalla contra la pandemia, pues ahora se trata de encaminar la investigación hacia la búsqueda de otras variables que pudiesen estar

asociadas a la rápida propagación del virus, y que no necesariamente están relacionadas al contacto entre las personas (Cantos, Camacho, & Talavera, 2020)

1.2. Descripción del problema

En el transcurso de esta pandemia, hubo una ciudad que resalto por encima de muchas la cual es Guayaquil (Ver Figura 1), a causa de los niveles de contagios que se produjeron en esta y el pésimo manejo que pudo haberse dado por varias autoridades. Medios de comunicación como la BBC news MUNDO ponía en sus titulares "...el drama de Guayaquil, que tiene más muertos por covid-19 que países enteros y lucha a contrarreloj para darles un entierro digno" (BBC, 2020), "Los números se quedan cortos: Lenin Moreno admite que la crisis del covid-19 es más grave de lo informado" (BBC, 2020)

Figura 1
Mapa de ubicación de Guayaquil



Fuente: (INEC, 2014)

Esto crea la necesidad de analizar que ocurrió en los años: 2020 y 2021, dentro de la ciudad que se convirtió en un epicentro en la pandemia y verificar

o constatar las variables que más afecten a la propagación del virus, para que se pueda prevenir de mejor forma a la ciudadanía en un futuro.

Durante todo el tiempo de la pandemia se han divulgado varias teorías: de la forma de transmisión del virus, de los cuidados, de las variables que afecten su propagación, de nuevas amenazas que pueda tener el mundo, por esto es necesario llevar a cabo los estudios para comprobar dichas teorías.

Para entender y estar mejor preparados para el futuro, es importante saber en qué fallamos anteriormente, cuáles son las debilidades que podemos corregir, porque habrá otras variables que no dependen de los ciudadanos ni de los diferentes gobiernos. Es por esto que se planteó esta investigación con variables que son controlables y que pueden ser mejoradas ya sea por el gobierno central o los gobiernos autónomos descentralizados y otras variables que están fuera de nuestro control como lo es la precipitación y temperatura.

Así, los estudios se han enfocado en evaluar cómo variables atmosféricas, meteorológicas y sociodemográficas pueden influir sobre la gravedad de la enfermedad, la capacidad de propagación y la facilidad de transmisión (Ferrer, 2020; Olcina, Biener, & Marti, 2020)

En este sentido, se afirma que los factores ambientales como la temperatura, la humedad relativa, la velocidad del viento, la radiación solar y la precipitación, pueden influir en el exceso de la mortalidad de una enfermedad (Losa, 2021). Para el presente trabajo, el interés está enfocado al virus por COVID-19. Es así como en el estudio realizado por Waang, et al, (2020) en 429 ciudades, encontró que las temperaturas bajas favorecen la transmisión del virus. Así también, entre los hallazgos más importantes del estudio realizado por Huang, Zhongwei, et al. (2020) se encuentra que el 60% de casos confirmados de COVID-19 ocurrieron en lugares en los cuales la temperatura osciló entre los 5°C a los 15°C, lo que explicaría que el virus pudiese estar propagándose hacia niveles de altitud más altos y que la pandemia por COVID-19 se puede extender de manera más rápida en las grandes ciudades.

Otro aporte importante a escala global sobre la influencia de las variables climáticas y la propagación del virus es el realizado por Falcão, et al. (2020), en el cual se encontró una asociación negativa entre países con temperatura media y las infecciones por COVID-19, así como también se halló que, en los países afectados por el virus, mientras más precipitaciones presentaban estos países, mayores eran las infecciones por COVID-19. Para el caso ecuatoriano, el estudio realizado por Lara & Unda, (2021) en el que se relacionan las variables contaminantes atmosféricas y la propagación de COVID-19, se encontró que en las zonas de la capital en las que se reportó mayor contaminación con dióxido de nitrógeno y dióxido de azufre hubo un mayor índice de personas contagiadas.

Por otro lado, las variables sociodemográficas también son un foco de interés al momento de comprender la velocidad y alcance de contagios por COVID-19 (Medeiros, et al, 2020). Ya que se trata de aportar información sobre la propagación del virus en condiciones sociodemográficas determinadas. En este caso, el estudio realizado por Acosta et al. (2020) en el que se evaluaron variables sociodemográficas, antecedentes y manifestaciones clínicas y radiológicas de los pacientes por COVID-19, se encontró que quienes tenían mayor riesgo de experimentar los niveles graves de la enfermedad eran los adultos mayores y también quienes reportaron otras enfermedades asociadas como hipertensión arterial y obesidad.

Así mismo, Medeiros et al. (2022), encontraron que las tasas de incidencia y de mortalidad en el momento del decreto de estado de alarma del virus se asoció positivamente con las tasas de incidencia, mortalidad del virus y demanda de hospitales actual. Y quienes tenían mayor tasa de mortalidad por el virus eran personas mayores en residencia. En cuanto a los servicios sanitarios, se encontró que la mortalidad y la letalidad del virus no están asociados a las variables demográficas, de gasto sanitario ni de los servicios básicos.

Finalmente, un gran aporte relacionado a las variables sociodemográficas y la incidencia de COVID-19 es el realizado por Yupari-Azabache (2021) en el que se halló que en una muestra de 64 pacientes entre marzo y mayo del 2020, el 85,71% del total de fallecidos fueron del sexo masculino, con una ocupación predominante de jubilados con un 28,57% y la edad promedio de 64 a 67 años. Y entre los síntomas más comunes en pacientes fallecidos, fueron la dificultad respiratoria con un 90,48% con malestar general, tos y auscultación pulmonar anormal. Estos pacientes también presentaron comorbilidad con otras enfermedades, como diabetes y problemas cardiovasculares.

Existen varias correlaciones que contrastan que las altas temperaturas medias y las precipitaciones si tienen efecto para que el contagio del COVID-19 sea más rápido, este estudio se lo realizó en varias ciudades de Brasil en las más afectadas por el virus. (Auler, et al, 2020)

Mohan et al, (2022), analizaron las componentes atmosféricas en Malaysia donde encontró que la temperatura en la que se puede desarrollar mejor el virus, por ende un mayor contagio es de 23 – 25 °C y con una humedad relativa de 70 – 80 %. Estos datos también tienen que ver mucho con el distanciamiento social.

Hay variables como el nivel socio económico que puede afectar de forma directa a la transmisión del virus, por los costos que implica el cuidado para no infectarse (comprar: alcohol, guantes, sanitizantes, mascarillas, entre otros) en la pandemia y la necesidad de salir a trabajar (aumenta la exposición).

Sin embargo, con una verificación y constatación con métodos estadísticos, de la influencia de las diferentes variables obtendremos una idea clara de cómo afectan estas a la proliferación del virus y esto ayudara a estar mejor preparados para una buena planificación en el futuro.

1.3. Objetivos

General

Analizar variables atmosféricas y sociodemográficas que estén asociadas a la propagación de la COVID-19 en la ciudad de Guayaquil, mediante el uso de modelos estadísticos multivariantes, identificando las variables que más influyen.

Específicos

- Obtener un conjunto de datos depurado y apropiado para el análisis estadístico planteado.
- Analizar de manera multivariante las variables atmosféricas y sociodemográficas, asociadas a COVID-19 con datos reales.
- Visualizar los resultados del análisis estadístico, utilizando gráficos dinámicos e interactivos mediante un dashboard.

1.4. Hipótesis

Las variables sociodemográficas y atmosféricas afectan o son relevantes en la propagación del virus COVID – 19.

1.5. Alcance

Esta investigación está centrada en el análisis de las variables atmosféricas y sociodemográficas principales que influyan en la supuesta transmisión y aumento de los contagios de Covid – 19, en la ciudad de Guayaquil en los años 2020 y 2021.

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. ANÁLISIS MULTIVARIANTE

Hace referencia al conjunto de técnicas estadísticas que se propone analizar e interpretar las relaciones entre algunas variables cualitativas o cuantitativas de forma simultánea, a través de la construcción de modelos estadísticos complejos que permitan visualizar el aporte que cada variable brinda a todo el sistema de relaciones, y de esta manera que se pueda explicar, describir e incluso predecir los fenómenos que constituyen el objeto de interés de la investigación (Meneses, 2019). En términos básicos se puede comprender al análisis multivariante como una extensión del análisis bivariante.

Entre las técnicas más utilizadas en este tipo de análisis, se encuentran:

- **El análisis de componentes principales**
- Análisis factorial
- Análisis de correspondencias
- Análisis de conglomerados
- Escalamiento multidimensional
- **Regresión múltiple**
- Correlación canónica
- ANOVA de dos o más factores
- MANOVA,
- Análisis discriminante
- Regresión logística
- Ecuaciones estructurales.

Si bien hay una variedad de técnicas para análisis multivariante, es importante tomar en cuenta que la elección de la técnica depende de dos factores fundamentales en el diseño de la investigación, como son: la pregunta o el objetivo general que impulsa el desarrollo de la investigación, y las características de los datos que se utilizan. Por lo que, el uso de técnicas multivariantes es apropiado cuando el investigador tiene como propósito dar respuesta a preguntas que

involucran el estudio de varias relaciones ya sean éstas de dependencia o independencia con las variables estudiadas de manera conjunta y simultánea (Meneses, 2019).

Análisis de componentes principales

Un problema muy común en el análisis multivariante es el reducir la dimensionalidad, es posible describir con una gran precisión los valores de las variables con un pequeño conjunto de estas, con una pequeña pérdida de información. (Peña, 2002)

Este es el objetivo del análisis de componentes principales (PCA) representar la información de un conjunto con un subconjunto, es decir con menos variables seleccionadas, que son combinaciones lineales del conjunto original. (Peña, 2002)

Esta técnica permite (Peña, 2002):

- Representa de una forma eficiente en un espacio dimensión pequeña, las observaciones en el espacio general p -dimensional.
- Transformar variables correladas en nuevas incorreladas.

Regresión Múltiple

En la regresión lineal múltiple (RLM) se usa más de una variable explicativa a diferencia de la regresión lineal simple que tiene la variable independiente y la dependiente, Esto ofrece una mayor comprensión al momento de construir los modelos requeridos y llegar a estimaciones precisas. (Abuin, 2007)

Este es un modelo muy versátil que puede evaluar una relación de los predictores y el destino continuo. Estos predictores pueden ser derivados, continuos o categóricos, se dice que es un modelo lineal ya que consiste en términos aditivos de los cuales cada uno de estos puede ser un predictor que tiene un coeficiente. (IBM, 2021)

2.2. GEO CODIFICACIÓN Y GEORREFERENCIACIÓN

La georreferenciación es la técnica o proceso en el cual un objeto o entidad se determina una posición dentro de un sistema de coordenadas (x,y) espaciales,

proceso que se usa frecuentemente en los Sistemas de Información Geográfica (SIG) (Davila & Camacho, 2012)

Según Troncoso & Fernández, (2021) para poder realizar una georreferenciación de un número masivo de direcciones de un archivo es necesario una geo codificación previa, sin embargo, esto no suele ser exacto ya que las personas llenan las direcciones de forma no estandarizada.

Es decir, la geo codificación es la transformación de la ubicación, esta puede ser una dirección, nombre de un lugar o entidad, que se encuentre en la superficie de la tierra. Estas se pueden introducir la dirección o varias de estas para así obtener una entidad con coordenadas y con atributos de cada una de ellas, que pueden ser representadas en un plano cartográfico para su respectivo análisis espacial. (ESRI, 2016)

Existen diferentes usos en los que se puede aplicar la geocodificación como: una herramienta para la mejora de una Infraestructura de Datos espaciales a través de un Sistema de Información geográfica para la toma de decisiones en el ámbito urbano o ambiental (Duarte, 2020), para la generación primaria de información georreferenciada para los distintos procesos de desarrollo territorial y gobernabilidad junto a la planeación del espacio público con sus recursos. (Vargas & Álvarez, 2013)

2.3. DATOS ESPACIALES

Según Cressie, (1993) nos dice que existen 3 tipos básicos de datos espaciales los cuales son: datos de área, datos geoestadísticos y los patrones de puntos.

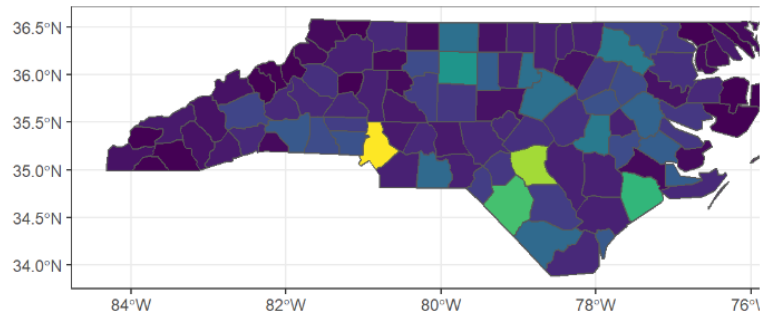
Datos de área

Los datos de área son de forma regular o irregular estos se van dividiendo en un número finito de unidades de área, para una mejor comprensión es visualizarlo con Ecuador y sus provincias, el área total es ecuador y sus subdivisiones son las provincias, que pueden subdividirse en cantones y a su vez en parroquias. En la siguiente Figura podemos observar el número de muertes infantiles repentinas en

los condados de Carolina del Norte de EEUU DEL año 1974. (Pebesma, 2021). Ver Figura 2.

Figura 2

Muertes infantiles súbitas en Carolina del Norte



Fuente: (Moraga, 2019; Pebesma, 2021)

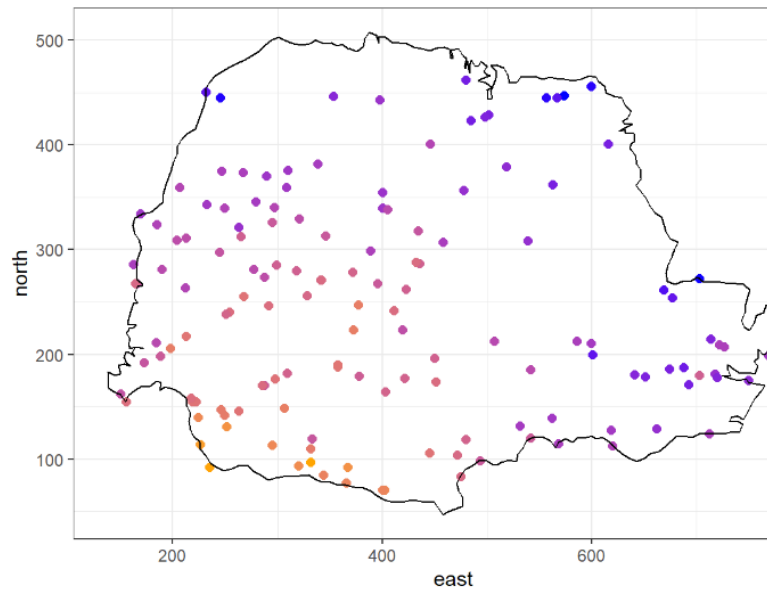
Datos geoestadísticos

Los datos geoestadísticos es el dominio D, es un conjunto fijo continuo. Esto quiere decir que el atributo S puede variar continuamente mientras los puntos del dominio D son fijos, para un mejor entendimiento imagine la contaminación del aire medida por varias estaciones de monitoreo continuo que midan la precipitación estas estaciones. (Moraga, 2019)

En el siguiente ejemplo podemos observar el promedio de mayo a junio de diferentes años de 143 estaciones del estado de Parana, Brasil. (Ribeiro, et al, 2020), Ver Figura 3.

Figura 3

Precipitación media medida en 143 estaciones de registro en el estado de Paraná, Brasil.



Fuente: (Ribeiro, et al, 2020; Moraga, 2019)

Patrones de puntos

A diferencia de los dos anteriores casos en los patrones de puntos el dominio D es aleatorio. Son eventos aleatorios que indican la ocurrencia de los eventos o brindan información adicional, es decir como ejemplos podrían ser personas que tienen una determinada enfermedad en un ciudad o puntos donde sea ha suscitado asaltos a personas. (Moraga, 2019)

En la siguiente Figura 4 podemos observar el brote de colera que afecto a Londres en el año de 1854 estos representan un patrón de puntos. (Li, 2019). Este fue una de las primeras investigaciones de la cartografía combinada con un análisis espacial y salud.

Figura 4

Mapa de John Snow del brote de cólera de Londres de 1854.



Fuente: (Li, 2019; Moraga, 2019)

2.4. R - INLA

R-INLA es un paquete del software R, el cuál utiliza la aproximación integrada de Laplace INLA, que es el método de la inferencia bayesiana. En estos años se ha ido consolidando como un método alternativo para la cadena de Markov Monte Carlo (MCMC), esto se ha creado por la rapidez de su uso. La metodología INLA se basa en los modelos que se podría decir que son los campos aleatorios latentes de Gaussian Markov (GMRF), que contempla varios modelos que se utilizan. (R – INLA Project, 2022).

Como un ejemplo se ilustra los tiempos que se demora en realizar el mismo procedimiento en R con el paquete rjags (MCMC) vs r-inla (INLA), para *Poisson GLM con un iid random effect*. Ver Tabla 1.

Tabla 1

Tiempo transcurrido en segundos

n	rjags	r-inla
100	30.394	0.383
500	142.532	1.243
5000	1714.468	5.768
25000	8610.32	30.077
100000	got bored after 6 hours	166.819

Fuente: (Morrison, 2017)

2.5. SISTEMAS DE REFERENCIA DE COORDENADAS

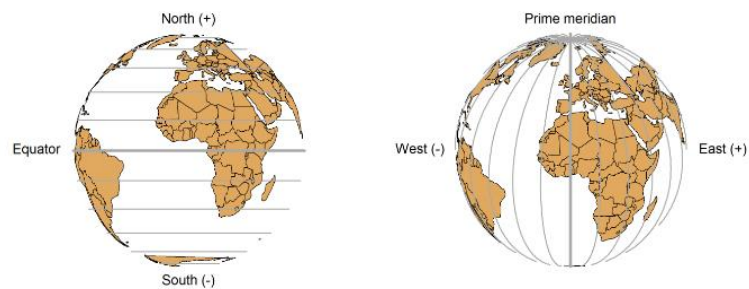
De los aspectos más importantes de los datos espaciales son los sistemas de referencia de coordenadas quienes dan el punto exacto en la superficie de la tierra existen 2 tipos de sistemas de referencia:

Sistema de referencia geográfica

Que es el que mide la posición de una entidad en la superficie de la tierra en un elipsoide o geoide utilizan la latitud y la longitud. Estos son ángulos que pueden medirse en grados o grados, minutos y segundos. En el caso de la latitud su origen se da en el Ecuador que es una línea imaginaria que se encuentra en la mitad de la tierra, equidistante a los polos, con esto se divide a la tierra en norte y sur. Y para el caso de la longitud su punto de partida es el meridiano de grenwich que es una línea imaginaria que va desde el polo norte hasta el polo sur, este determina los hemisferios este y oeste. La latitud en la superficie de la tierra es el ángulo que va desde el Ecuador (0°) hasta el punto mencionado puede ir desde los -90° y 90° . La longitud es el ángulo q se mide desde el meridiano de grenwich hacia el oeste o este pueden ir desde -180° a 180° . (Moraga, 2019) (Ver Figura 5)

Figura 5

Paralelos (izquierda) y meridianos (derecha) de la Tierra.



Fuente: (Moraga, 2019)

Sistema de referencia proyectado

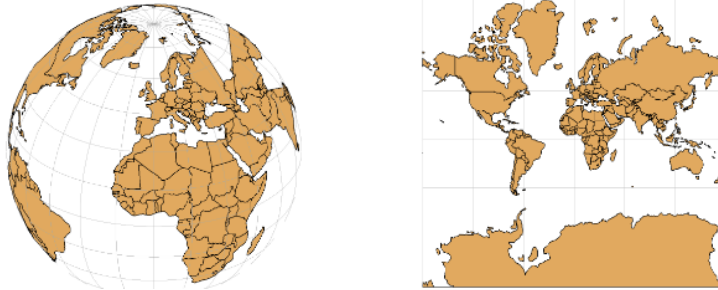
Este sistema utiliza una proyección cartográfica como una transformación de la superficie terrestre a un sistema bidimensional, sin embargo, esto crea distorsiones ya sea en todo el plano o en partes específicas ya que no se puede conservar varias propiedades como forma, distancia, dirección y área. Una de las proyecciones más conocidas es la de Universal Transversa de Mercator (UTM), este trata de conservar ángulos y formas locales divide a la superficie de la tierra en 60 zonas cada una utiliza una proyección de Mercator que mapea toda la tierra en norte y sur. (Moraga, 2019) (Ver Figura 6)

Cada zona viene dada por un número y se la identifica con norte o sur, cada zona se divide por un meridiano central que tiene un valor de 500000m para evitar coordenadas negativas y así mismo en las zonas sur el norte empieza con un valor de 10 000 000m para evitar coordenadas negativas.

En el Ecuador existen varias zonas las cuales son: 17 Sur, 17 Norte, 18 Sur. 18 Norte y en la zona insular son las zonas 15 Norte, 15 Sur, 16 Norte y 16 Sur. La más común es la 17 Sur en la cual se encuentra Guayaquil que es la zona de estudio.

Figura 6

Superficie tridimensional de la Tierra (izquierda) y representación bidimensional de la Tierra (derecha).



Fuente: (Moraga, 2019)

2.6. ESTADÍSTICA ESPACIAL

La geografía automatizada estaba en un desarrollo continuo, mientras transcurría esto, se encontraba necesidades de incorporar el análisis cuantitativo para la dimensión espacial. Cuando tenemos un mapa temático el cual contiene diferentes áreas y como información tiene la densidad poblacional esto es igual o equivale a un histograma en la estadística convencional (Buzai & Montes, 2021)

El análisis cuantitativo evoluciono en el análisis espacial cuantitativo, esto se destaca ya que consolidan las bases conceptuales y las sistematizan con una evolución que se puede aplicar. (Buzai & Santana, 2019)

La integración de los sistemas de información geográfica (SIG), tiene 2 componentes: el grafico el cual podemos observar como resultado en el mapa y el alfanumérico que son las bases de datos que contienen los atributos de cada una de las entidades espaciales. Con esto podemos realizar operaciones estadísticas en las 3 zonas: la gráfica, la alfanumérica y la unión de estas dos.

Ahora si nos vamos más a fondo con los aplicativos podemos llegar a realizar modelos espacio temporales.

2.7. CONCEPTOS DE EPIDEMIOLOGIA

Incidencia

La incidencia en epidemiología es el lapso donde las personas que no tienen o padecen la enfermedad estudiada están en riesgo de contagiarse. (Wasserman, 2013)

Riesgo

Este término se refiere a la ocurrencia o probabilidad de que exista algún resultado adverso, puede ser morbilidad, contagio o mortalidad, es una medida de probabilidad al futuro, es por esto que este mide la probabilidad que se desarrolle la enfermedad en un lapso de tiempo determinado. (Wasserman, 2013)

Riesgo Relativo

Este es la razón de la población con un riesgo absoluto de contagio o muerte con la exposición estudiada contra el riesgo absoluto de contagio o muerte sin la exposición deseada. (Castillo, et al, 2011)

SIR (Radio de incidencia estandarizado)

Es el índice de los casos observados (Y) sobre los casos esperados (E), estos generalmente se usan cuando las tasas de enfermedad se usan comparando las tasas de enfermedad de una población de referencia. (Kelsey & Gold, 2017)

2.8. MODELOS ESPACIO TEMPORALES EJEMPLO

Un ejemplo práctico es el que realiza Moraga, (2019) en el modelado espacio temporal de datos de área para el cáncer de pulmón en Ohio, siguiendo los siguientes pasos:

Datos y mapa

Para el ejemplo utiliza los datos de cáncer de pulmón y el mapa de Ohio que se puede encontrar en el paquete de R, SpatialepiApp. Leemos el archivo alfanumérico que se encuentra en formato .csv y el archivo grafico que se encuentra en formato shape.

Preparación de datos

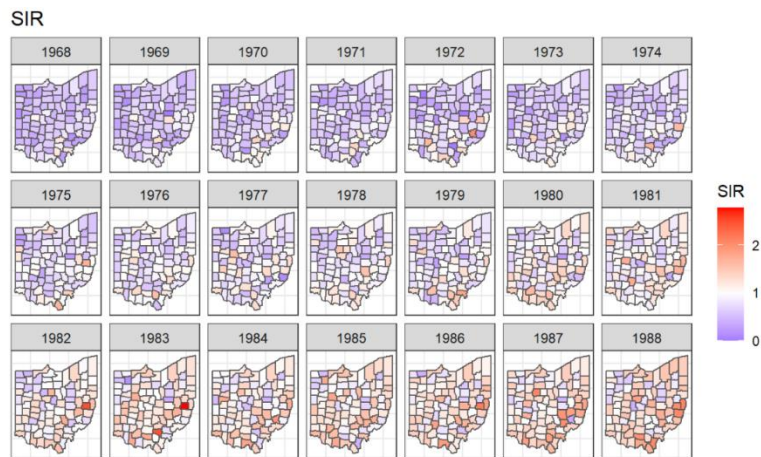
En los datos se encuentran datos desde el año de 1968 a 1988, con los casos de cáncer y la población. Con esto se puede calcular los casos observados, esperados y SIR para los diferentes años. Aumentando nuestra base de datos. Y a su vez agregamos todos esta base de datos a la parte grafica.

Mapeo de SIR

Una vez con los valores de SIR (Razón de incidencia estandarizada), por cada año en el que se estableció en la base de datos, se procede a mapear para una mejor visualización del entendimiento de la epidemia. Ver Figura 7

Figura 7

Mapas de SIR de cáncer de pulmón en los condados de Ohio de 1968 a 1988 creados con una escala de colores divergente.



Fuente: (Moraga, 2019)

Modelado

En este apartado se puede observar un modelo en el cual asumimos que tiene una distribución de poisson. El cual es:

$$Y_i \sim \text{Poisson}(E_i \theta_i), i = 1, \dots, n,$$

E_i = número esperado de casos

θ_i = riesgo relativo

Partimos creando una matriz de vecindades para definir los efectos aleatorios espaciales. Esta matriz está basada en los límites de los vecinos de cada polígono. Además, se debe crear los vectores de índice de los condados y el vector de tiempo que es de los años que se tiene en la base. El primer vector lo debemos replicar ya que tenemos 2 efectos aleatorios distintos.

Seguido de esto agregamos las estimaciones de los riesgos relativos con los respectivos límites tanto superior como inferior con un 95%.

Con esto podemos crear los mapas de riesgo relativo con las diferentes funciones de R-INLA Y ggplot. Como podemos ver a continuación. Ver Figura 8

Figura 8
Riesgos relativos en diferentes años



Fuente: (Moraga, 2019)

CAPITULO 3

3. METODOLOGÍA

3.1. RECOLECCIÓN DE INFORMACIÓN, (EXTRACCIÓN)

Para esta investigación como primer paso se recopiló la información desde varias fuentes de datos oficiales las cuales son:

La primera base de datos recopilada fue por parte del Instituto Oceanográfico y Antártico de la Armada (INOCAR) con variables atmosféricas (**BDAT – Base de datos atmosférica**), esta información es de cada día de los años 2020 y 2021, la información es de:

- Temperatura
- Precipitación

La segunda base de datos es de las variables socio demográficas esta base es del Instituto Geográfico Militar (IGM) (**BDSD- Base de datos sociodemográfica**), fue obtenida desde su geoportal de libre acceso, las variables son:

- Densidad poblacional
- Nivel de instrucción
- Nivel socio económico
- Servicios básicos

Estas variables están representadas gráficamente en sector censales en formato shapefile.

La tercera base es de información geográfica la cual se encuentra en el geo portal del Gobierno Autónomo Descentralizado Municipal de Guayaquil (GADMG) (**BDMG – Base de datos municipio de Guayaquil**) la información es:

- Manzanas
- Lotes
- Vías
- Parroquias urbanas

La cuarta base es de información geográfica extraída de Google Maps (**BDGM – Base de datos Google Maps**) en esta se encuentran las siguientes variables:

- Vías
- Barrios
- Centros de salud
- Sectores
- Parroquias

La quinta base es de la información de los casos covid – 19 (**BDC19 – Base de datos covid 19**), en esta se encuentran las siguientes variables:

- Código CIE 10
- Dirección
- Edad
- Sexo
- Fecha de entrada

La sexta base es de la información de planificación su autor es la ex secretaria nacional de Planificación y desarrollo (SENPLADES) (**BDSP – Base de datos SENPLADES**), en esta se encuentran las siguientes variables:

- Distritos
- Circuitos

3.2. TRATAMIENTO DE INFORMACIÓN (TRANSFORMACIÓN)

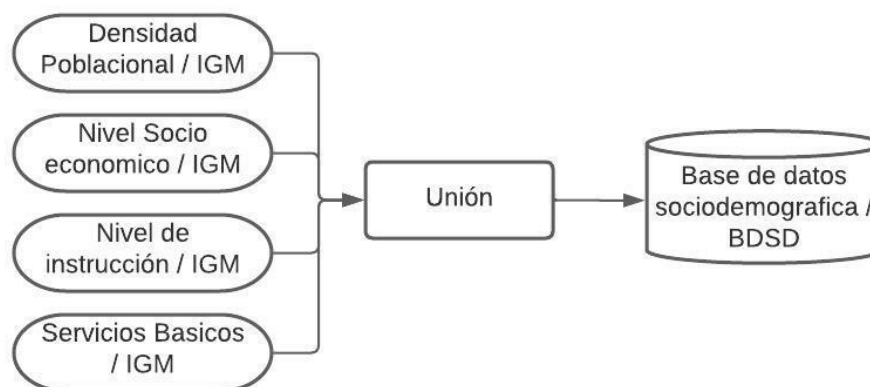
Para el uso e integración de estas distintas bases de datos es necesario tratarlas para poder realizar la investigación.

BDAT. En esta base se hizo la revisión y estandarización de las fechas en un solo formato y la creación de un código por cada fecha para luego realizar un join.

BDS. En esta base se hizo la revisión y unión de los diferentes shapes ya que eran documentos distintos y la misma forma geométrica se pudo unirlos en un mismo archivo con su respectiva estandarización. Ver Figura 9.

Figura 9

Diagrama del proceso para la elaboración de la base de datos sociodemográfica



BDC19. En esta base se hizo una depuración seleccionando las variables del código CIE 10 referentes al covid 19, recordando que al principio de la pandemia no existía un código como tal para este virus por lo que los doctores usaban otros códigos.

Para una mejor geolocalización y/o georreferenciación de las direcciones de esta base se utiliza la ayuda de las bases **BDMG, BDGM** cargadas en un SIG, en bases de datos alfanumérica y Google Earth Pro, todos estos complementos se necesitan ya que existen diferentes formas de decir las direcciones por parte de los usuarios, ellos hacen referencia a su barrio, a sus calles, a sitios conocidos, número de vías, numero de solar o manzana dado esto los diferentes códigos para hacer

geolocalización no tienen una buena precisión por lo que para una georreferenciación segura se la hizo una por una.

Una vez completada las coordenadas de las direcciones se transforma en un archivo espacial con la ayuda de las coordenadas, estas se las transforma a coordenadas proyectadas.

Se realiza un join con la BDAT para complementar la base así se tiene coordenada del paciente, temperatura y precipitación del día en que entro al hospital u de los días atrás que puedo haberse contagiado. Seguido de esto se hace una unión espacial para obtener las variables sociodemográficas, completando la base de datos.

Por otro lado, se hace un join espacial para obtener cuantos casos se obtuvo por barrios, distritos y circuitos. Ver Figura 10 y 11.

Sumado a esto se utilizó la base de datos de las infografías publicadas para alimentar la base de datos atmosféricas, con las fechas de los pacientes que entraron al hospital esto con 12 y 14 días posteriores de temperatura y precipitación.

Figura 10
Diagrama del proceso de elaboración de las bases de datos 1

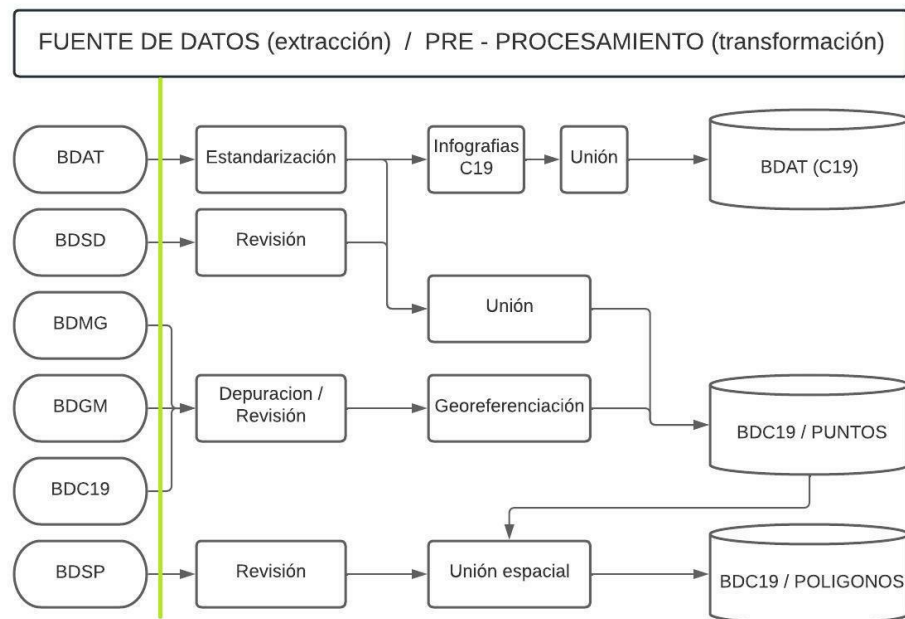
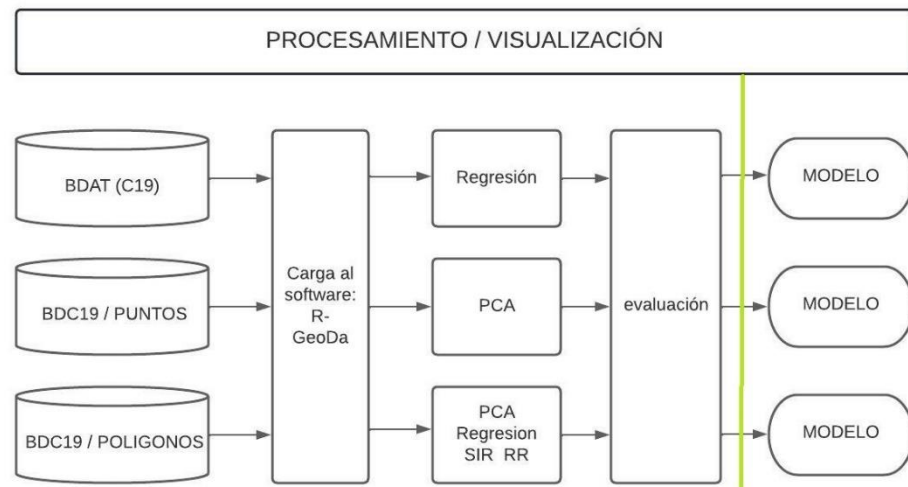


Figura 11

Diagrama del proceso de elaboración de las bases de datos 2



3.3. MODELADO ESPACIAL (CARGAR)

Una vez preparadas las bases de datos se procede a llamarlas en el software R, y se realiza un análisis descriptivo de las mismas.

Se realiza con la BDC19 (puntos) que contiene las variables atmosféricas y sociodemográficas se realiza un PCA, con el software GeoDa.

Con la BDAT(C19) se realiza una regresión lineal, con el software R, implementando las variables del tiempo de 12 y 14 días anteriores al ingreso al hospital.

Con la BDC19 (polígonos) se prepara la base de datos siguiendo los pasos usados por Moraga, (2019).

Para el análisis de los datos cargamos la BDC19 a R, tenemos los casos observados sin embargo necesitamos saber los casos esperados, estos se calculan con la población de las zonas y la población total. Con los datos observados (Y) y esperados (E) podemos calcular el SIR

$$SIR_i = \frac{Y_i}{E_i}$$

Con esto se mapea el SIR, se calcula la matriz de vecindades y se hace inferencia con INLA para obtener los RR. En la inferencia se encuentran dos efectos aleatorios uno para modelar la variación residual espacial y el otro el ruido no estructurado, para la formula utilizamos las covariables y los efectos aleatorios para estos utilizamos los modelos besag e iid. Ajustamos el modelo con INLA, y la formula.

Con esto se calcula los RR y se puede mapear cuales son los circuitos con más riesgo relativo de la ciudad de Guayaquil.

Obtenidos estos datos se creó el Dashboard para crear un mejor entorno de presentación de los resultados obtenidos.

Los datos fueron procesados en los softwares: Qgis versión 3.24.3 y Rstudio versión Desktop 2022.07.0+548, Geoda.

CAPITULO 4

4. RESULTADOS

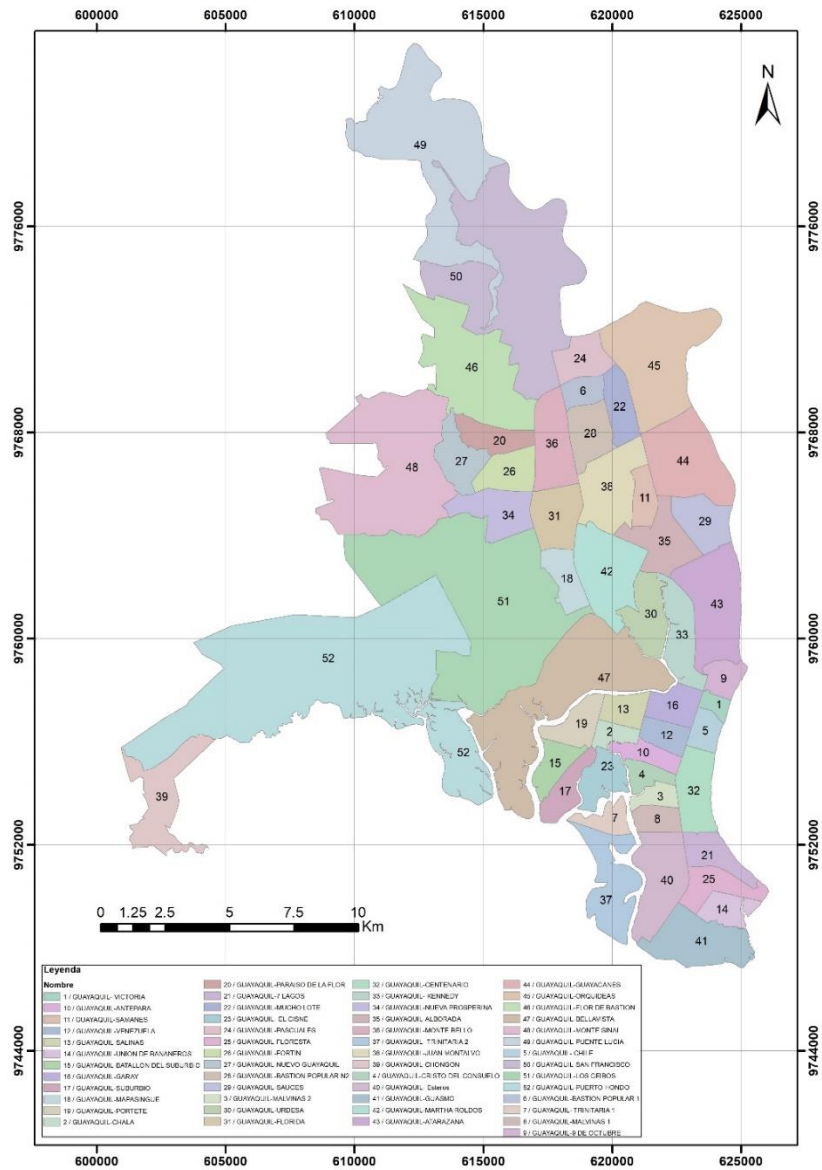
4.1. BASE DE DATOS DEPURADA

Con la base de datos obtenidas de los diferentes entes que provén información, se depuraron y se siguió la metodología antes mostrada para conseguir las bases finales, las cuales se derivaron en bases sociodemográficas, atmosféricas y mixtas.

BASE COVID 19 DE POLÍGONOS

Estas bases para un mejor análisis se las hizo con la de circuitos de la ciudad de Guayaquil (Ver Figura 12) que es cubierta por 52 polígonos y se las dividió por años 2020 y 2021. En estas bases solo se encuentran las variables sociales por cada circuito, en base a los sectores censales se hizo un promedio de cada variable por circuito.

Figura 12
Mapa de circuitos de Guayaquil

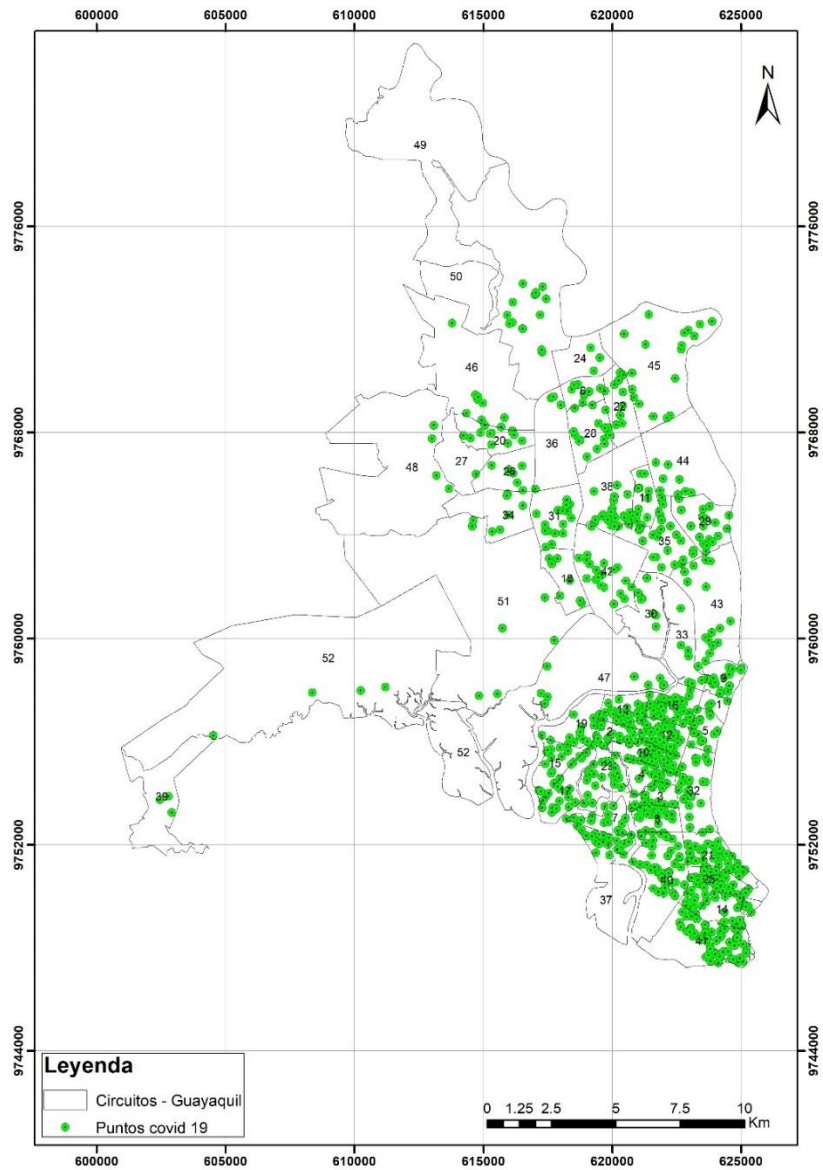


BASE COVID 19 DE PUNTOS

En esta base se encuentra las variables atmosféricas y sociodemográficas y la ubicación de cada paciente que se georreferenció (Ver Figura 13). Las variables atmosféricas se encuentran por día y las sociodemográficas son las calculadas por el IGM en base al INEC.

Figura 13

Mapa de casos de COVID georreferenciados en la ciudad de Guayaquil



BASE DE VARIABLES ATMOSFÉRICAS

En esta base se encuentra la precipitación y temperatura diaria de cada paciente que entro al hospital, los datos son de 12 y 14 días antes de su entrada, esta decisión se la tomo en base a Maguiña Vargas, Gastelo Acosta, & Tequen Bernilla, (2020). Más otras variables creadas para un mejor análisis que incluyen la información con las infografías oficiales del gobierno de como avanzaba el COVID 19 en Guayaquil.

4.2. ESTADÍSTICA DESCRIPTIVA

Para explicar mejor los datos obtenidos de las bases depuradas se hizo una pequeña exploración en sus datos con fines descriptivos por cada año. Ver Tabla 2 y 3.

Tabla 2

VARIABLES SOCIALES Y ATMOSFÉRICAS DEL 2020 EN FORMATO DE PUNTOS

	P10	T10	P12	T12	P14	T14	SB	SE	DP	NI
Min. :	0.000	23.000	0.000	23.000	0.000	23.000	0.200	0.200	0.200	0.100
1st Qu.:	0.000	26.000	0.000	26.000	0.000	26.000	1.000	0.400	0.400	0.500
Median :	0.000	27.000	0.000	27.000	0.000	27.000	1.000	0.600	0.600	0.700
Mean :	1.545	27.020	0.988	27.040	1.837	27.200	0.978	0.619	0.588	0.520
3rd Qu.:	0.100	28.050	0.100	28.120	0.000	28.330	1.000	0.800	0.800	0.700
Max. :	75.300	29.870	75.300	31.000	82.800	31.000	1.000	1.000	1.000	0.700

Nota: P10 = Precipitación 10 días antes de entrar al hospital, T10 = Temperatura 10 días antes de entrar al hospital, P12 = Precipitación 12 días antes de entrar al hospital, T12 = Temperatura 12 días antes de entrar al hospital, P14 = Precipitación 14 días antes de entrar al hospital, T14 = Temperatura 14 días antes de entrar al hospital, SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Tabla 3

VARIABLES SOCIALES Y ATMOSFÉRICAS DEL 2021 EN FORMATO DE PUNTOS

	P10	T10	P12	T12	P14	T14	SB	SE	DP	NI
Min. :	0.000	23.330	0.000	23.330	0.000	23.000	0.200	0.200	0.200	0.100
1st Qu.:	0.000	25.500	0.000	25.520	0.000	25.530	1.000	0.600	0.400	0.500
Median :	0.000	26.700	0.000	26.930	0.000	26.630	1.000	0.600	0.600	0.700
Mean :	4.752	26.720	4.684	26.760	4.148	26.700	0.974	0.633	0.618	0.515
3rd Qu.:	0.600	27.600	0.800	27.820	1.200	27.650	1.000	0.800	0.800	0.700
Max. :	104.600	30.400	81.500	30.400	104.600	30.400	1.000	1.000	1.000	0.700

Nota: P10 = Precipitación 10 días antes de entrar al hospital, T10 = Temperatura 10 días antes de entrar al hospital, P12 = Precipitación 12 días antes de entrar al hospital, T12 = Temperatura 12 días antes de entrar al hospital, P14 = Precipitación 14 días antes de entrar al hospital, T14 = Temperatura 14 días antes de entrar al hospital, SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Con las 2 tablas mostradas se puede observar las pequeñas diferencias de los casos puntuales en el año 2020 y 2021.

Los niveles de servicios básicos son bastante uniformes con una media de 0.97, en el caso del nivel socio económico va de 0.61 a 0.63 un nivel medio, en la densidad poblacional va de 0.58 a 0.61 un nivel medio y en el nivel de instrucción el valor es de 0.51 un valor muy central.

Las siguientes medidas son de las variables sociales en formato polígono del año 2020 y 2021, Ver Tabla 4 y 5.

Tabla 4

Medidas descriptivas de las variables sociodemográficas del 2020 en formato de área.

	SB	SE	DP	NI	Y	E	SIR
Min. :	0.381	0.200	0.419	0.100	0.000	1.368	0.000
1st Qu.:	0.854	0.446	0.641	0.397	3.000	6.556	0.486
Median :	0.986	0.596	0.782	0.507	6.000	8.690	0.838
Mean :	0.893	0.592	0.769	0.466	9.019	9.019	0.963
3rd Qu.:	1.000	0.721	0.941	0.590	11.500	11.838	1.392
Max. :	1.000	0.985	1.000	0.659	38.000	19.299	2.926

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Tabla 5

Medidas descriptivas de las variables sociodemográficas del 2021 en formato de área.

	SB	SE	DP	NI	Y	E	SIR
Min. :	0.381	0.200	0.419	0.100	0.000	1.202	0.000
1st Qu.:	0.854	0.446	0.641	0.397	3.000	5.759	0.428
Median :	0.986	0.596	0.782	0.507	5.000	7.634	0.878
Mean :	0.893	0.592	0.769	0.466	7.923	7.923	1.006
3rd Qu.:	1.000	0.721	0.941	0.590	9.500	10.400	1.221
Max. :	1.000	0.985	1.000	0.659	32.000	16.953	4.210

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Se observo que en el nivel de servicios básicos existe una igualdad en los años 2020 y 2021 con un 0.89 de media, en el nivel socio económico ocurre el mismo fenómeno los dos años presentan el mismo valor de 0.59, la densidad poblacional tiene el mismo valor de 0.76 y el nivel de instrucción de la cabeza familiar tienen un valor de 0.4658.

Esto sucede ya que estas variables sociodemográficas son fijas y las variables atmosféricas son móviles van cambiando cada día. Sin embargo, en las bases de área (polígonos) se puede contabilizar los casos que han ocurrido en el año de una manera mejor para el análisis espacial.

4.3. VARIABLE ATMOSFÉRICA

Se trabajo con la base de puntos la cual contiene variables atmosféricas y sociodemográficas, se realizó un análisis de componentes principales, este análisis se lo hizo con el software Geoda especialista en estadística espacial. Ver Tabla 6.

Tabla 6

Análisis de componentes principales de las variables sociodemográficas y atmosféricas

PCA						
method: eigen						
Standard deviation:						
1.3566	1.1345	1.0070	0.9875	0.7770	0.5289	
Proportion of variance:						
0.3067	0.2145	0.1690	0.1625	0.1006	0.0466	
Cumulative proportion:						
0.3067	0.5212	0.6902	0.8528	0.9534	1.0000	
Kaiser criterion: 3						
95% threshold criterion: 4						
Eigenvalues:						
1.8403	1.2871	1.0141	0.9751	0.6037	0.2798	
Variable Loadings:						
	PC1	PC2	PC3	PC4	PC5	PC6
P12	-0.076	-0.102	-0.681	-0.719	-0.053	-0.022
T12	-0.064	0.109	-0.725	0.672	0.085	-0.012
SB	0.173	-0.801	-0.031	0.090	0.255	0.505
SE	0.635	-0.296	-0.048	0.041	0.025	-0.710
DP	0.461	0.459	-0.029	-0.146	0.706	0.239
NI	-0.587	-0.195	0.084	-0.025	0.653	-0.428
Squared correlations:						
	PC1	PC2	PC3	PC4	PC5	PC6
P12	0.011	0.013	0.470	0.504	0.002	0.000
T12	0.008	0.015	0.533	0.440	0.004	0.000
SB	0.055	0.826	0.001	0.008	0.039	0.071
SE	0.741	0.113	0.002	0.002	0.000	0.141
DP	0.391	0.271	0.001	0.021	0.301	0.016
NI	0.635	0.049	0.007	0.001	0.257	0.051

Nota: P12 = Precipitación 12 días antes de entrar al hospital, T12 = Temperatura 12 días antes de entrar al hospital, SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Se puede observar en los primeros dos componentes principales, que las variables que más contribuyen son las sociodemográficas, las atmosféricas en el componente

1 tienen una contribución de 0.011 para la precipitación y de 0.008 para la de temperatura a 12 días de haber entrado al hospital. Por lo que se decidió analizarlo de otro punto.

Con la base de datos se hizo una clasificación a las lluvias (Ver Tabla 7), del día 12 y 14 anterior a partir de su entrada al centro de salud.

Según Maguiña Vargas, Gastelo Acosta, & Tequen Bernilla, 2020, el 95 % del tiempo en el periodo de incubación es de 12 días y otros determinan que puede ser de 2 a 14, por lo que para el análisis se conserva el día 12 y 14 antes de entrar al centro de salud.

Tabla 7
Categoría de lluvias respecto a mm/h

PRECIPITACIÓN mm/h	CATEGORÍA
0.1 - 2	débil
Feb-15	moderada
15 - 30	fuerte
30 - 60	muy fuerte
> 60	torrencial

Una vez hecho la clasificación se procedió a realizar una regresión lineal, con un pequeño detalle se creó otra variable de lluvias consecutivas del día 12 y 14, los resultados fueron favorables. Ver Tabla 8.

Tabla 8
Modelo lineal de la variable atmosférica

Residuals:				
Min	1Q	Median	3Q	Max
-29432	-12639	-4410	15169	34097
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	92302.8	12173.9	7.582	1.16E-13 ***
ES_lluvia_fuerte_12_141	2695.5	8187.8	0.329	0.742
temperat_12	-2619.6	459.3	-5.704	1.77E-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16330 on 657 degrees of freedom

Multiple R-squared: 0.0473, Adjusted R-squared: 0.0444

F-statistic: 16.31 on 2 and 657 DF, p-value: 1.223e-07

Nota: ES_lluvia_fuerte_12_141 es la nueva variable de lluvias consecutivas de 12 y 14 días antes de entrar al hospital.

En este modelo la variable independiente son los casos informados por el gobierno en infografías de Guayaquil, las dependientes son las lluvias consecutivas fuertes del día 12 y 14 anterior y la temperatura del día 12 anterior.

Analizando los coeficientes se puede observar que la temperatura es significativa, no obstante, en el caso de la precipitación se ve que cuando existe lluvias consecutivas los casos tienen tendencia a subir.

Esto podría tener varias razones por el comportamiento de las personas que se da cuando llueve.

4.4. VARIABLE SOCIODEMOGRÁFICAS

En esta sección se realizó procesamiento y análisis de las variables sociodemográficas en los circuitos de la ciudad de Guayaquil en los años 2020 y 2021 respectivamente.

Para el año 2020 en un inicio se realizó un análisis exploratorio, con el análisis de componentes principales. Ver Tabla 9.

Tabla 9

PCA de las variables sociodemográficas del año 2020

PCA method: eigen				
Standard deviation:				
	1.410621	1.31652	0.465371	0.245671
Proportion of variance:				
	0.497463	0.433306	0.054142	0.015089
Cumulative proportion:				
	0.497463	0.930769	0.984911	1
Kaiser criterion: 2				
95% threshold criterion: 2				
Eigenvalues:				
	1.98985	1.73323	0.21657	0.0603542
Variable Loadings:				
	PC1	PC2	PC3	PC4
SB	0.5459	0.4462	-0.4493	0.5486
SE	0.6976	-0.0167	-0.0421	-0.7151

DP	-0.0213	-0.7184	-0.6943	0.0369
NI	-0.4636	0.5334	-0.5606	-0.4316

Squared correlations:

	PC1	PC2	PC3	PC4
SB	0.5931	0.3451	0.0437	0.0182
SE	0.9683	0.0005	0.0004	0.0309
DP	0.0009	0.8946	0.1044	0.0001
NI	0.4276	0.4931	0.0681	0.0112

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Se pudo observar que en el componente principal 1 las variables que más aportan es el nivel socioeconómico, los servicios básicos, el nivel de instrucción y por último la densidad poblacional.

Seguido de esto se hizo una regresión lineal. Ver Tabla 10.

Tabla 10

Regresión de las variables sociodemográficas del año 2020

REGRESSION				
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set :	dis_c_2020_a			
Dependent Variable :	Y	Number of Observations:	52	
Mean dependent var :	9.01923	Number of Variables :	5	
S.D. dependent var :	8.44095	Degrees of Freedom :	47	
R-squared :	0.380826	F-statistic :	7.22688	
Adjusted R-squared :	0.32813	Prob(F-statistic):	0.00012747	
Sum squared residual:	2294.03	Log likelihood :	-172.242	
Sigma-square:	48.8091	Akaike info criterion :	354.484	
S.E. of regression :	6.98635	Schwarz criterion :	364.241	
Sigma-square ML : ML	44.1159			
S.E of regression ML:	6.64198			
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.0587577	15.379	0.00382063	0.99703
SB	-13.7366	14.1992	-0.96742	0.33829
SE	33.2014	14.0282	2.36677	0.02212
DP	-16.1137	9.09955	-1.77082	0.08308
NI	29.9612	13.2886	2.25465	0.02886
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 40.542180				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	69.057	0	
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	

Breusch-Pagan test	4	8.4224	0.07727
Koenker-Bassett test	4	2.5451	0.63659
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	14	11.6839	0.63167

COEFFICIENTS VARIANCE MATRIX

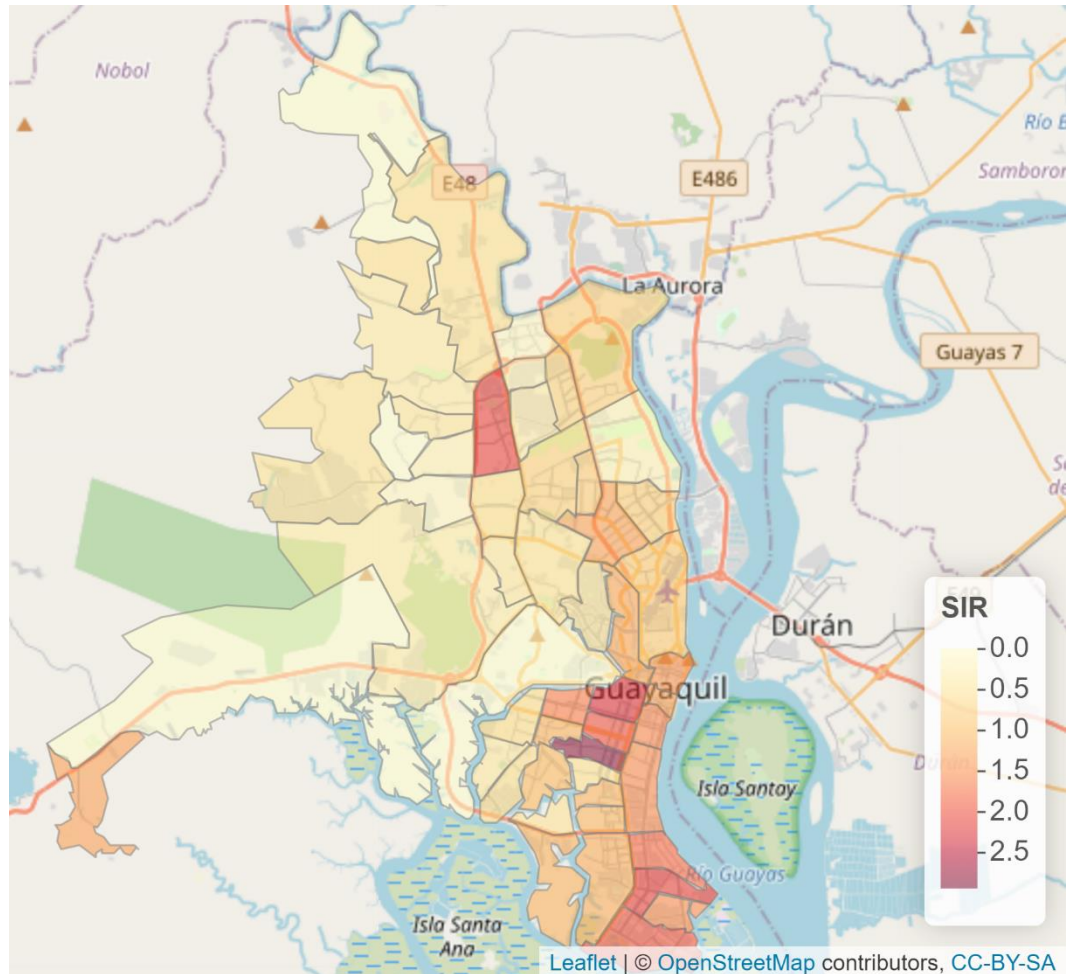
CONSTANT	SB	SE	DP	NI
236.515	-75.374	-27.280	-126.092	-118.455
-75.374	201.618	-168.824	51.607	-95.149
-27.280	-168.824	196.790	-8.194	145.465
-126.092	51.607	-8.194	82.802	45.517
-118.455	-95.149	145.465	45.517	176.588

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

En la regresión lineal los coeficientes predominantes son de las variables de nivel socio económico, el nivel de instrucción, los servicios básicos y densidad poblacional respectivamente, se pudo observar que coincide la primera y última variable con el PCA.

Se siguió con un análisis observacional epidemiológico se creó la variable de casos esperados y de SIR (standardized incidence ratio) (Ver Figura 14). Con esto podemos observar donde son los circuitos que tiene un mayor SIR y se pudo observar que los valores más predominantes son en el sur.

Figura 14
Modelo SIR año 2020, por circuitos



El SIR puede dar una idea de riesgo, sin embargo, tiene bastantes limitaciones sobre todo en áreas que están divididas en áreas pequeñas y grandes es por esto por lo que se decidió utilizar los circuitos, tanto como las parroquias y distritos se observa bastante desigualdad en las áreas que conformaban Guayaquil, los circuitos se encontraban de una forma más homogénea. En vista de esto se hace modelos que puedan tomar en cuenta la información de sus vecinos y de otras covariables para dar un suavizado a los SIR, siguiendo la metodología que usa Moraga, 2019.

Es por esto, por lo que asumimos que el modelo tiene una distribución de Poisson.

$$Y_i \sim \text{Poisson}(E_i \theta_i), i = 1, \dots, n,$$

E_i = número esperado de casos

θ_i = riesgo relativo

El logaritmo de θ_i se lo expresa:

$$\log\theta_i = \beta_0 + \beta_1 * SB_i + \dots + u_i + v_i$$

Donde:

β_0 : es el intercepto que representa el riesgo total

β_1 : el coeficiente de la covariable

u_i : componente estructural espacial

v_i : efecto espacial no estructurado

Se procede a crear la matriz de vecindades, la cual crea una lista de vecinos de cada polígono

Ver Tabla 11.

Tabla 11
Matriz de vecindades

Poligono	Numero de vecinos							
	1	2	3	4	5	6	7	8
V1	5	9	16	33				
V2	10	12	13	17	19	23		
V3	4	8	32	4	8	32		
V4	3	10	32	3	10	32		
V5	1	12	16	32				
V6	22	24	28	36	50			
V7	37							
V8	3	32	40	3	32	40		
V9	1	16	33	43				
V10	2	4	12	23	32			
V11	35	38	44					
V12	2	5	10	13	16	32		
V13	2	12	16	19				
V14	25	41	25	41				
V15	17	19	17	19				
V16	1	5	9	12	13	33		
V17	2	15	19	23				
V18	31	42	51					
V19	2	13	15	17				
V20	26	27	36	46				
V21	25	32	40					
V22	6	24	28	38	44	45		
V23	2	10	17					
V24	6	22	45	50				
V25	14	21	40	41				
V26	20	27	31	34	36			
V27	20	26	34	46	48			
V28	6	22	36	38				

V29	35	43	44						
V30	33	35	42	47					
V31	18	26	34	36	38	42	51		
V32	3	4	5	8	10	12	21	40	
V33	1	9	16	30	35	43	47		
V34	26	27	31	48	51				
V35	11	29	30	33	38	42	43	44	
V36	6	20	26	28	31	38	46	50	
V37	7								
V38	11	22	28	31	35	36	42	44	
V39	52								
V40	8	21	25	32	41				
V41	14	25	40	14	25	40			
V42	18	30	31	35	38	47	51		
V43	9	29	33	35	9	29	33	35	
V44	11	22	29	35	38	45			
V45	22	24	44						
V46	20	27	36	48	50				
V47	30	33	42	51					
V48	27	34	46	51					
V49	50								
V50	6	24	36	46	49				
V51	18	31	34	42	47	48	52		
V52	39	51							

Este modelo contiene 2 efectos aleatorios, uno es
vi: es para el ruido que no está estructurado
ui: la variación residual espacial

Con esto podemos especificar la fórmula que se incluye la respuesta, covariable y los efectos, en la inferencia bayesiana existe el modelo besag el cual usaremos en esta fórmula junto con la matriz de vecindades. Usamos INLA con la formula expuesta y la distribución poisson. En la Tabla 12 podemos ver los resultados.

Tabla 12
Coeficientes de la inferencia de circuitos 2020

	Fixed effects:						
	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	-2.741	1.272	-5.281	-2.728	-0.275	-2.703	0
SB	0.725	1.324	-1.837	0.711	3.367	0.684	0
SE	1.791	1.097	-0.37	1.792	3.947	1.794	0
DP	0.198	0.781	-1.329	0.193	1.749	0.185	0
NI	1.277	0.877	-0.464	1.283	2.986	1.295	0

Random effects:						
Name	Model					
idareau	Besags ICAR model					
idareav	IID model					

Model hyperparameters:						
	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for idareau	6.22	3.04	2.34	5.55	13.99	4.47
Precision for idareav	19149.58	19077.7	1305.23	13403.72	69968.73	3583.73

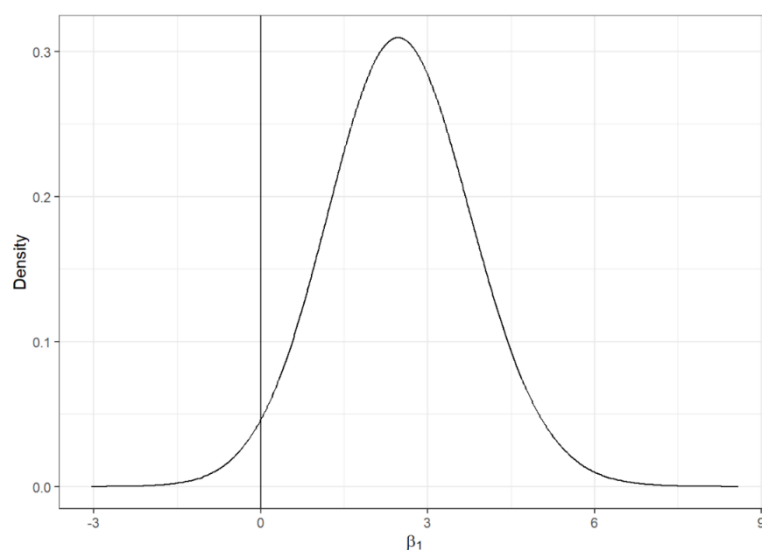
Marginal log-Likelihood:	-178.83
--------------------------	---------

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Se puede observar que las variables de nivel socio económico y el nivel de instrucción son los de más alto valor esto quiere decir que estos inciden más a la hora de contagio. Cabe recalcar que las mismas variables aparecen primeras en el PCA y LM.

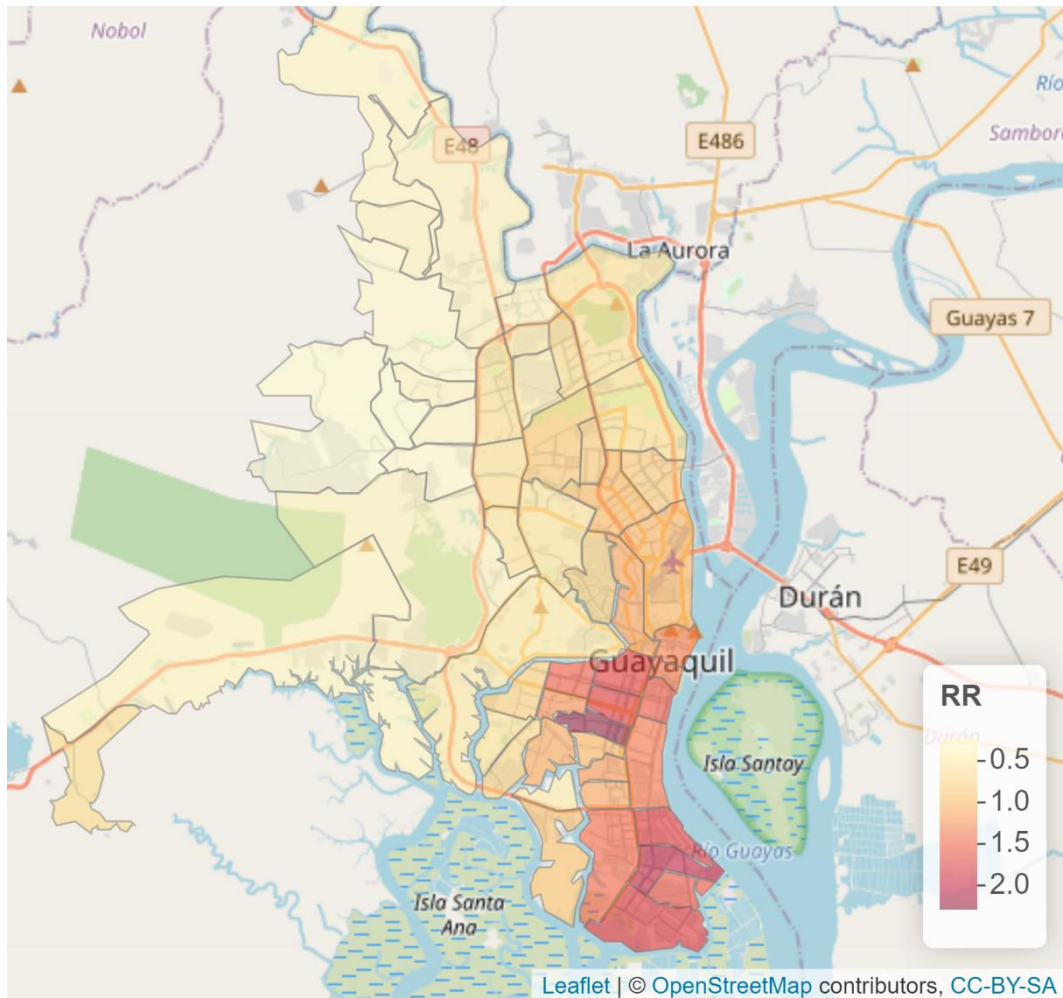
Para una apreciación visual se dibujó la distribución posterior, antes se calculó una suavizado de la distribución marginal de la variable más alta en los resultados el cual es el nivel socio económico. Ver Figura 15.

Figura 15
Distribución posterior marginal SE 2020



Una vez con estos resultados podemos mapear los riesgos relativos que están dados por la media posterior y su credibilidad están contenidos entre los cuartiles 0.25 y 0.75. Ver Figura 16.

Figura 16
Mapa de riesgos relativos en Guayaquil en el 2020



Con esto podemos observar las zonas con mayor riesgo relativo para el contagio de COVID 19.

Este mismo procedimiento se lo realizo para el año 2021. A continuación, se presenta las diferentes Tablas y Figuras del 2021.

Al igual que en 2020 se comenzó realizando un PCA (Ver Tabla 13), en la cual las variables que más aportan al componente 1 respectivamente son: nivel socio

económico, servicios básicos, nivel de instrucción, y por último densidad poblacional.

Tabla 13

PCA de las variables sociodemográficas del año 2021

PCA method: eigen				
Standard deviation:				
	1.410621	1.31652	0.465371	0.245671
Proportion of variance:				
	0.497463	0.433306	0.054142	0.015089
Cumulative proportion:				
	0.497463	0.930769	0.984911	1
Kaiser criterion: 2				
95% threshold criterion: 2				
Eigenvalues:				
	1.98985	1.73323	0.21657	0.0603542
Variable Loadings:				
	PC1	PC2	PC3	PC4
SB	0.5459	0.4462	-0.4493	0.5486
SE	0.6976	-0.0167	-0.0421	0.7151
DP	-0.0213	-0.7184	-0.6943	0.0369
NI	-0.4636	0.5334	-0.5606	0.4316
Squared correlations:				
	PC1	PC2	PC3	PC4
SB	0.5931	0.3451	0.0437	0.0182
SE	0.9683	0.0005	0.0004	0.0309
DP	0.0009	0.8946	0.1044	0.0001
NI	0.4276	0.4931	0.0681	0.0112

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Seguido de esto se realizó una regresión lineal la cual podemos ver en la Tabla 14, en esta las variables de mayor coeficiente son en el siguiente orden: nivel socio económico, nivel de instrucción, densidad poblacional y servicios básicos.

Tabla 14

Regresión de las variables sociodemográficas del año 2021

REGRESSION				
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set :	circuito 2021			
Dependent Variable :	Y	Number of Observations:	52	
Mean dependent var :	7.92308	Number of Variables :	5	
S.D. dependent var :	7.28762	Degrees of Freedom :	47	
R-squared :	0.393481	F-statistic :	7.62283	
Adjusted R-squared :	0.341862	Prob(F-statistic):	8.08E-05	
Sum squared residual:	1675.02	Log likelihood :	-164.066	
Sigma-square:	35.6387	Akaike info criterion :	338.131	
S.E. of regression :	5.96982	Schwarz criterion :	347.887	
Sigma-square ML : ML	32.2119			
S.E of regression ML:	5.67555			
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-12.1267	13.1413	-0.922786	0.36083
SB	-12.5424	12.1332	-1.03372	0.30656
SE	33.5788	11.987	2.80125	0.00737
DP	-6.88939	7.77554	-0.886034	0.38011
NI	35.7707	11.3551	3.15019	0.00284
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 40.542180				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	55.0396	0	
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	4	7.4024	0.11609	
Koenker-Bassett test	4	2.4337	0.65655	
SPECIFICATION ROBUST TEST				
TEST	DF	VALUE	PROB	
White	14	9.0848	0.82558	
COEFFICIENTS VARIANCE MATRIX				
CONSTANT	SB	SE	DP	NI
172.695	-55.035	-19.919	-92.068	-86.492
-55.035	147.214	-123.269	37.682	-69.474
-19.919	-123.269	143.689	-5.983	106.214
-92.068	37.682	-5.983	60.459	33.235
-86.492	-69.474	106.214	33.235	128.938

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Se procedió a realizar el modelamiento con inferencia bayesiana, primero se calculó los casos esperados y luego el SIR (Ver Figura 17), seguido del cálculo de matriz de vecindades que por ser el mismo polígono es la matriz que en 2020.

Se calcula la formula con las variables sociales que por ser de un mismo polígono son las mismas del 2020, sin embargo, en este modelo cambian los casos esperados, los casos activos y el SIR. Se realizó el mismo procedimiento para: inferencia (Ver tabla 15), distribución posterior (Ver Figura 18) RR (Ver Figura 19).

Figura 17
Modelo SIR año 2021, por circuitos

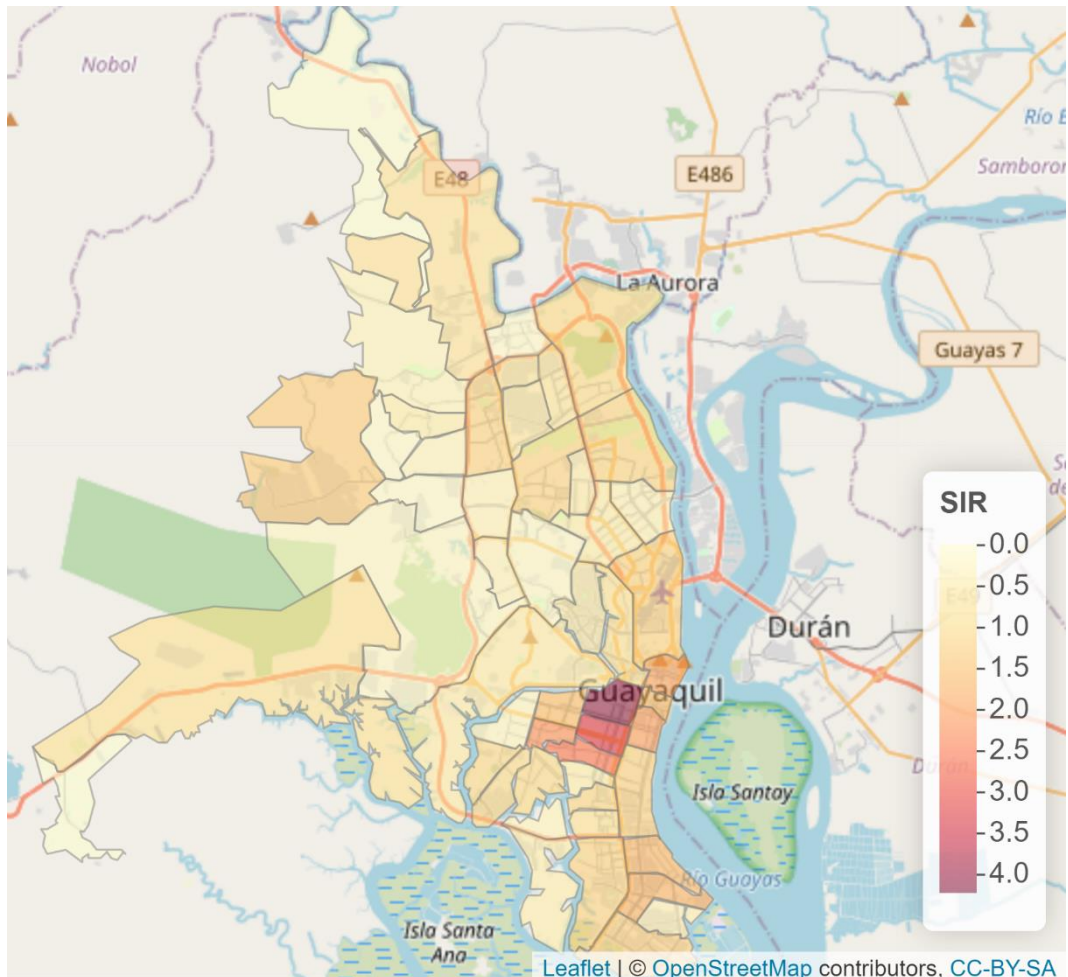


Tabla 15

Coeficientes de la inferencia de distritos 2021

Fixed effects:							
	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	-4.342	1.513	-7.336	-4.336	-1.384	-4.322	0
SB	-0.312	1.515	-3.309	-0.304	2.644	-0.291	0
SE	2.514	1.306	-0.008	2.498	5.129	2.468	0
DP	1.909	0.974	0.003	1.906	3.836	1.898	0
NI	3.004	1.033	0.991	2.998	5.054	2.985	0

Random effects:	
Name	Model
idareau	Besags ICAR model
idareav	IID model

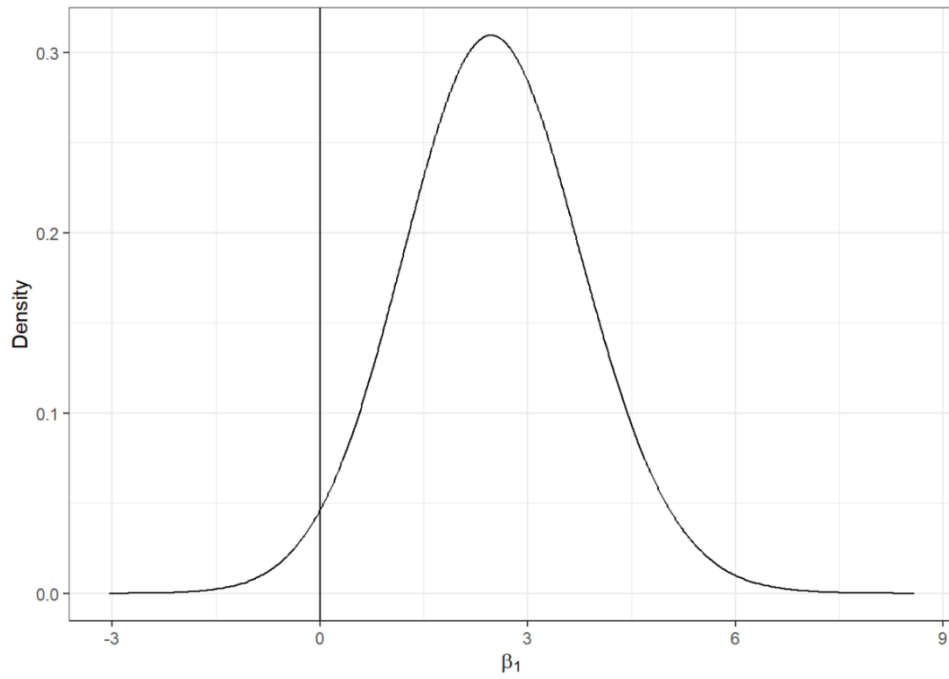
Model hyperparameters:							
	mean	sd	0.025quant	0.5quant	0.975quant	mode	
Precision for idareau	2.98	1.1	1.4	2.8	5.66	2.46	
Precision for idareav	18796.5	18853.71	1297.72	13108.39	69088.09	3574.66	

Marginal log-Likelihood: -177.73

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

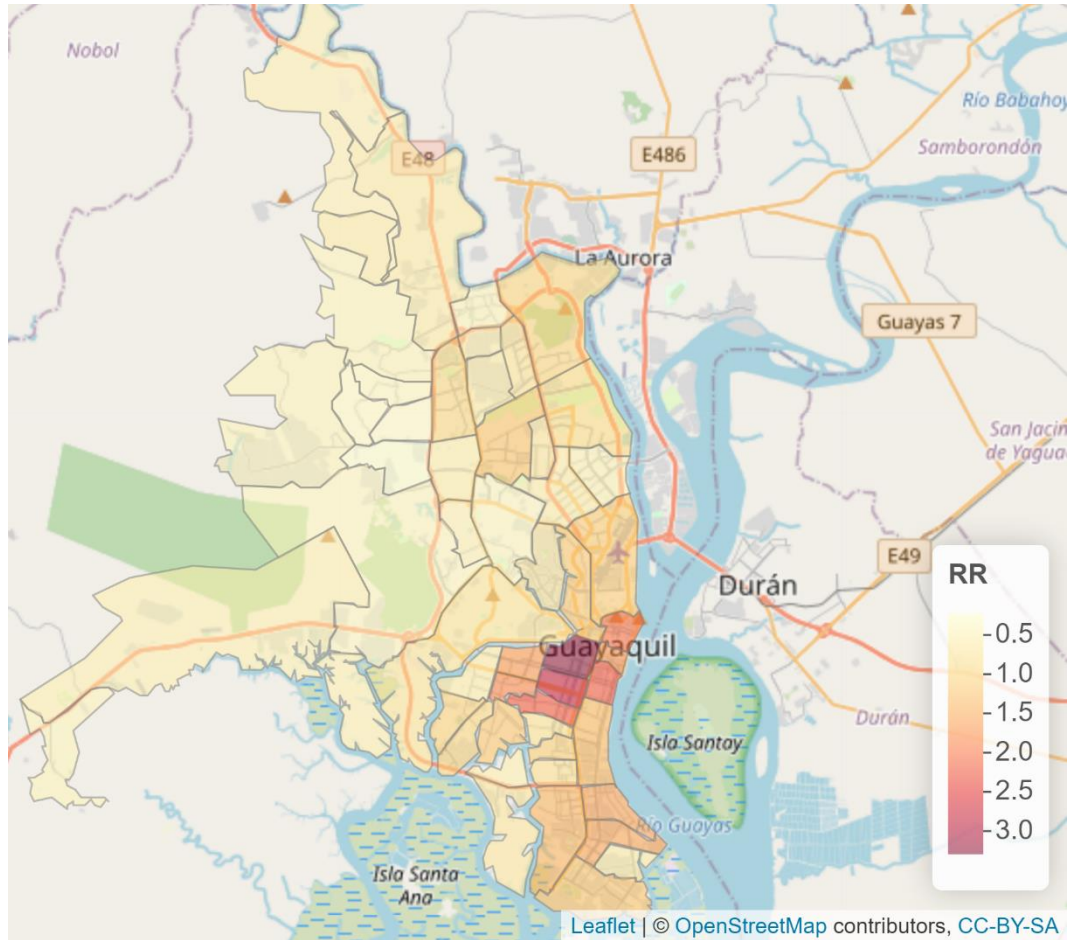
Ya con esto se puede ver la distribución marginal, para poder comparar con la del 2020 se realizó de la variable de nivel socio económico. Ver Figura 18

Figura 18
Distribución posterior SE 2021



Con estos resultados podemos calcular los RR de cada circuito podemos verlos en la siguiente Figura 19

Figura 19
Mapa de riesgos relativos en Guayaquil en el 2021



4.4.1. Comparación 2020 y 2021

Se hizo la comparación respectiva de los años 2020 y 2021, en el PCA se pudo observar que los niveles de jerarquía se mantienen en los dos años. En la regresión lineal se empieza a ver cambios En el orden de los valores de coeficientes en las dos posiciones de arriba se mantiene el nivel socio económico y el nivel de instrucción y en el caso de los 2 coeficientes de abajo hay un cruce entre servicios básicos y densidad poblacional. En el caso de la inferencia existe este cruce los 2 valores más altos son nivel socio económico y nivel de instrucción y los valores más bajos se mantiene servicios básicos y densidad poblacional al igual que los 2 superiores se cruzan cuando cambian de año. Ver tabla 16

Tabla 16

Comparación de los diferentes análisis de los años 2020 y 2021

Variables	PCA - PC1		Regresión Lineal		Inferencia	
	2020	2021	2020	2021	2020	2021
SB	0.593	0.593	-13.736	-12.542	0.726	-0.312
SE	0.968	0.968	33.201	33.578	1.791	2.515
DP	0.001	0.001	-16.113	-6.889	0.198	1.909
NI	0.427	0.427	29.961	35.770	1.277	3.004

Nota: SB = Servicios Básicos, SE = Nivel socio económico, DP= densidad poblacional, NI= Nivel de instrucción.

Igualmente se comparó los Y, E, SIR Y RR del año 2020 y 2021 en cada circuito (Ver Tabla 17)

Tabla 17

Comparación de los SIR, RR, Y, E de los años 2020 y 2021

CIRCUITOS A 2020	Y (CASOS)	E (CASOS ESPERADOS)	SIR	RR
IGUALES	7	0	1	0
AUMENTARON	27	52	26	22
DISMINUYERON	18	0	25	30

El análisis fue del 2020 al 2021, con respecto al número de casos en 7 circuitos se mantuvieron iguales, en 27 hubo un aumento de casos y en 18 distritos disminuyeron. Con respecto a E los casos esperados en 0 circuitos se mantuvieron iguales en 52 circuitos hubo un aumento y ninguno disminuyó. Con respecto a los SIR en 1 circuito se mantuvo igual, en 26 hubo un aumento y en 25 disminuyó. Con respecto a los RR en los circuitos en ninguno se mantuvo igual, en 22 hubo un alza y en 30 circuitos disminuyeron.

4.5. DASHBOARD

Para una mejor visualización de todos los procesos se decidió hacer un dashboard para mantener una comprensión visual y mejor entendimiento. Este se lo dividió en 3 partes en la primera se encuentra un resumen, objetivos, conclusiones y tablas para comprender mejor el problema, Ver Figura , en la segunda se encuentra el mapa de RR riesgos relativos con el SIR y las distribuciones posteriores del 2020 y en la

tercera hoja se encuentra el mapa de RR riesgos relativos con el SIR y las distribuciones posteriores del 2021.

Los mapas se los hizo interactivos es decir si se pasa con el cursor encima de los circuitos se muestra toda la información de cada circuito.

Figura 20
Dashboard ventana 1

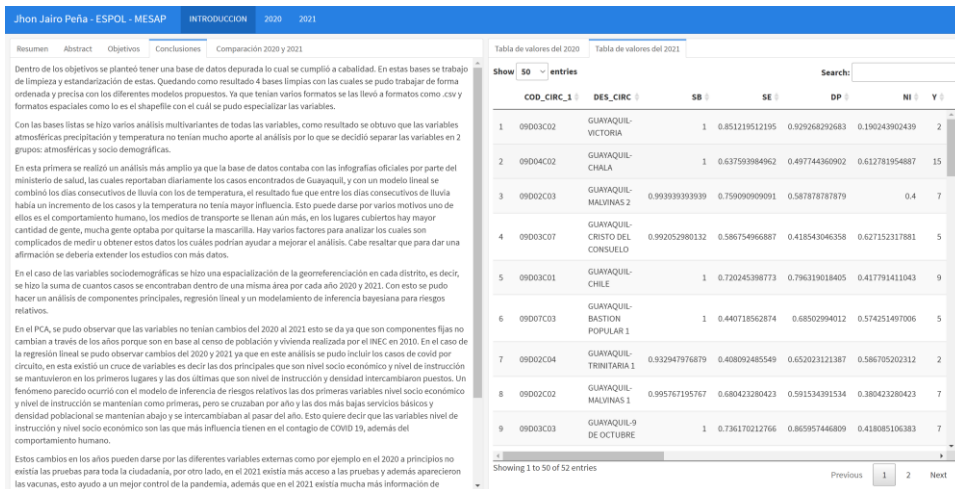
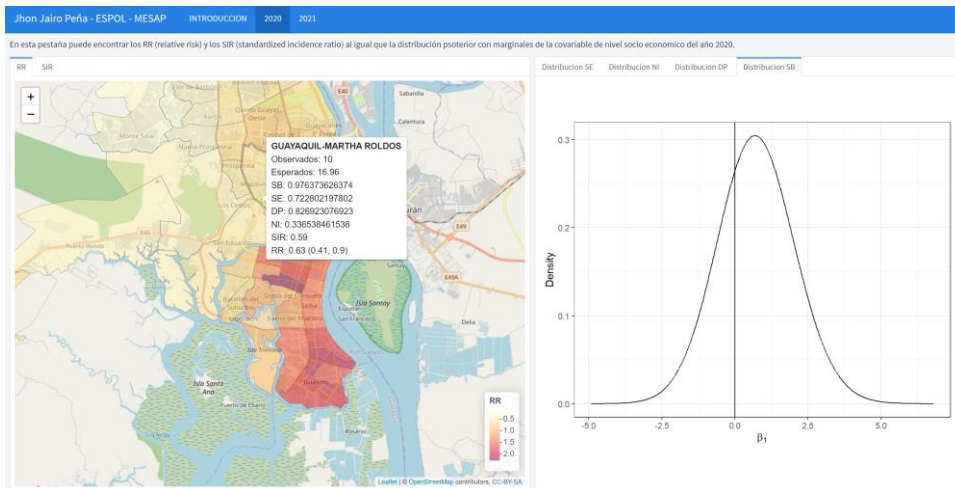


Figura 21
Dashboard ventana 2 con mapas interactivos



CAPÍTULO 5

5.1. CONCLUSIONES

Dentro de los objetivos se planteó tener una base de datos depurada lo cual se cumplió a cabalidad. En estas bases se trabajo de limpieza y estandarización de estas. Quedando como resultado 4 bases limpias con las cuales se pudo trabajar de forma ordenada y precisa con los diferentes modelos propuestos. Ya que tenían varios formatos se las llevó a formatos como .csv y formatos espaciales como lo es el shapefile con el cuál se pudo especializar las variables.

Con las bases listas se hizo varios análisis multivariantes de todas las variables, como resultado se obtuvo que las variables atmosféricas precipitación y temperatura no tenían mucho aporte al análisis por lo que se decidió separar las variables en 2 grupos: atmosféricas y socio demográficas.

En esta primera se realizó un análisis más amplio ya que la base de datos contaba con las infografías oficiales por parte del ministerio de salud, las cuales reportaban diariamente los casos encontrados de Guayaquil, y con un modelo lineal se combinó los días consecutivos de lluvia con los de temperatura, el resultado fue que entre los días consecutivos de lluvia había un incremento de los casos y la temperatura no tenía mayor influencia. Esto puede darse por varios motivos uno de ellos es el comportamiento humano, los medios de transporte se llenan aún más, en los lugares cubiertos hay mayor cantidad de gente, mucha gente optaba por quitarse la mascarilla. Hay varios factores para analizar los cuales son complicados de medir u obtener estos datos los cuáles podrían ayudar a mejorar el análisis. Cabe resaltar que para dar una afirmación se debería extender los estudios con más datos.

En el caso de las variables sociodemográficas se hizo una espacialización de la georreferenciación en cada distrito, es decir, se hizo la suma de cuantos casos se encontraban dentro de una misma área por cada año 2020 y 2021. Con esto se

pudo hacer un análisis de componentes principales, regresión lineal y un modelamiento de inferencia bayesiana para riesgos relativos.

En el PCA, se pudo observar que las variables no tenían cambios del 2020 al 2021 esto se da ya que son componentes fijas no cambian a través de los años porque son en base al censo de población y vivienda realizada por el INEC en 2010. En el caso de la regresión lineal se pudo observar cambios del 2020 y 2021 ya que en este análisis se pudo incluir los casos de covid por circuito, en esta existió un cruce de variables es decir las dos principales que son nivel socio económico y nivel de instrucción se mantuvieron en los primeros lugares y las dos últimas que son nivel de instrucción y densidad intercambiaron puestos. Un fenómeno parecido ocurrió con el modelo de inferencia de riesgos relativos las dos primeras variables nivel socio económico y nivel de instrucción se mantenían como primeras, pero se cruzaban por año y las dos más bajas servicios básicos y densidad poblacional se mantenían abajo y se intercambiaban al pasar del año. Esto quiere decir que las variables nivel de instrucción y nivel socio económico son las que más influencia tienen en el contagio de COVID 19, además del comportamiento humano.

Estos cambios en los años pueden darse por las diferentes variables externas como por ejemplo en el 2020 a principios no existía las pruebas para toda la ciudadanía, por otro lado, en el 2021 existía más acceso a las pruebas y además aparecieron las vacunas, esto ayudo a un mejor control de la pandemia, además que en el 2021 existía mucha más información de técnicas que pueden ayudar a prevenir el contagio.

Se planteo hacer un aplicativa visual en este caso un dashboard el cuál permita tener una comprensión visual de los modelos que se hizo, en este caso se enfocó en los modelos SIR y RR ya que se los puede presentar de una manera atractiva a la ciudadanía en forma de mapas interactivos de los años 2020 y 2021. Estos mapas al pasar el cursor por el circuito se despliega toda la información del mismo, por lo que se cumplió el objetivo.

Para un mejor análisis es necesario la totalidad o la mayoría de datos de los casos covid de los centros de salud de la ciudad de Guayauil con una buena descripción de su ubicación y el día posible de la fecha de contagio.

5. Referencias

- Abuin, J. (2007). Regresión lineal múltiple.
- Acosta, G., Escobar, G., Bernaola, G., Alfaro, J., Taype, W., Marcos, C., & Amado, J. (2020). Caracterización de pacientes con COVID-19 grave atendidos en un hospital de referencia nacional del Perú. *Revista Peruana de Medicina Experimental y Salud Pública*. doi:<http://dx.doi.org/10.17843/rpmesp.2020.372.5437>
- Auler, A., Cássaro, F., da Silva, V., & Pires, L. (2020). Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities. *Science of The Total Environment*, 729. doi:<https://doi.org/10.1016/j.scitotenv.2020.139090>
- BBC. (2020). Coronavirus en Ecuador | "Los números se quedan cortos": Lenín Moreno admite que la crisis del covid-19 es más grave de lo informado. *BBC News Mundo*. Recuperado el 5 de Septiembre de 2021, de <https://www.bbc.com/mundo/noticias-america-latina-52143469>
- BBC. (2020). Coronavirus en Ecuador: el drama de Guayaquil, que tiene más muertos por covid-19 que países enteros y lucha a contrarreloj para darles un entierro digno. *BBC News Mundo*. Recuperado el 5 de Septiembre de 2021, de <https://www.bbc.com/mundo/noticias-america-latina-52116100>
- Buzai, G., & Montes, E. (2021). *Estadística espacial: Fundamentos y aplicación con sistemas de información geográfica*. Buenos Aires, Argentina.
- Buzai, G., & Santana, M. (2019). *Métodos cuantitativos en Geografía Humana*. Buenos Aires: Universidad Nacional de Luján.
- Cantos, J., Camacho, S., & Talavera, J. (2020). Aspectos atmosféricos y climáticos en la expansión de la pandemia (COVID-19) en la provincia de Alicante. *Investigaciones Geográficas*, 275-297. España. doi:<https://doi.org/10.14198/INGEO2020.OCBCMT>
- Castillo, C., Mujica, O. j., Loyola, E., & Canela, J. (2011). Módulos de principios de epidemiología para el control de enfermedades.
- Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.
- Cuero, C. (2020). La pandemia del COVID-19 [The COVID-19 pandemic]. doi:10.37980/im.journal.rmdp.2020872
- Davila Martinez, F., & Camacho Arranz, E. (2012). Georreferenciación de documentos cartográficos para la gestión de archivos y cartotecas: " propuesta metodológica".
- Duarte, A. (2020). Geocodificador de domicilios en el marco de una Infraestructura de Datos Espaciales (IDE). Una herramienta SIG para el proceso de generación de Información Geográfica para la toma de decisiones. *Universidad Nacional de Luján*. Obtenido de <http://ri.unlu.edu.ar/xmlui/handle/rediunlu/991>
- ESRI. (2016). *ArcGIS desktop*. Obtenido de ¿Qué es la geocodificación?: <https://desktop.arcgis.com/es/arcmap/10.4/manage-data/geocoding/what-is-geocoding.htm>
- Fernandez Cordova, G. C. (2020). *El nuevo Coronavirus y la pandemia del Covid-19*.
- Ferrer, R. (2020). Pandemia por Covid-19: el mayor reto de la historia del intensivismo. *Medicina intensiva*,. doi:10.1016/j.medin.2020.04.002
- Gaus, D. (2021). COVID-19: vacunas. (6 (1)). doi: <https://doi.org/10.23936/pfr.v6i1.196>
- Huang, Z., Huang, J., Gu, Q., Du, P., Liang, H., & Dong, Q. (2020). Optimal temperature zone for the dispersal of COVID-19. *Science of The Total Environment*.

- IBM. (2021). Regresión lineal múltiple. Obtenido de <https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=tests-multiple-linear-regression>
- INEC. (2014). Instituto Nacional de Estadísticas y Censos. *Geografía Estadística*. Obtenido de <https://www.ecuadorencifras.gob.ec/geografia-estadistica/>
- Kelsey, J., & Gold, E. B. (2017). *Observational Epidemiology*. 295 - 307. doi:<https://doi.org/10.1016/B978-0-12-803678-5.00310-6>
- Lara, R. E., & Unda, F. B. (2021). Influencia de la contaminación atmosférica en la incidencia de enfermedades cardiovasculares y COVID-19 en parroquias del Distrito Metropolitano de Quito.
- Li, P. (2019). cholera: Amend, Augment and Aid Analysis of John Snow's Cholera Map.
- Losa, L. (2021). Aplicación del análisis de series temporales a variables atmosféricas.
- Maguiña, C., Gastelo, R., & Tequen, A. (2020). *El nuevo Coronavirus y la pandemia del Covid-19*. *Revista Medica Herediana*. doi:<http://dx.doi.org/10.20453/rmh.v31i2.3776>
- Medeiros, A., Daponte, A., Moreira, D., Toledo, R., Costa de Lima, K., & Gil, E. (2022). Factores asociados a la incidencia y la mortalidad por COVID-19 en las comunidades autónomas. *Gaceta Sanitaria*. doi:<https://dx.doi.org/10.1016/j.gaceta.2020.05.004>
- Meneses, J. (2019). Introducción al análisis multivariante.
- Mohan Viswanathan, P., Sabarathinam, C., Karuppanan, S., & Gopalakrishnan, G. (2022). Determination of vulnerable regions of SARS-CoV-2 in Malaysia using meteorology and air quality data.
- Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. *Chapman & Hall/CRC Biostatistics Series*. Obtenido de <https://www.paulamoraga.com/book-geospatial/index.html>
- National Library of Medicine. (2020). COVID-19 Pandemic: the greatest challenge in the history of critical care. Recuperado el 11 de Mayo de 2022, de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7151426/>
- Olcina, J., Biener, S., & Marti, J. (2020). Aspectos atmosféricos y climáticos en la expansión de la pandemia (COVID-19) en la provincia de Alicante. *Investigaciones Geográficas*. doi:<https://doi.org/10.14198/INGEO2020.OCBCMT>
- OMS. (2020). *Organización Mundial de la salud, Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020*. Recuperado el 11 de Junio de 2022, de <https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Pebesma, E. (2021). Sf: Simple Features for r. Obtenido de <https://CRAN.R-project.org/package=sf>.
- Peña, D. (2002). Análisis de datos multivariantes.
- Project, R. -I. (9 de 6 de 2022). *R - INLA Project*. Obtenido de What is INLA?: <https://www.r-inla.org/what-is-inla>
- Ribeiro, J., Paulo, J., Schlather, M., Bivand, R., & Ripley, B. (2020). geoR: Análisis de Datos Geoestadísticos. Obtenido de <http://www.leg.ufpr.br/geoR>
- Takashi, K., Arce, S., Arias, E., Àvalos, G., De La Cruz, G., Dávila, C., . . . Yauri, H. (2020). *Posible influencia de la estacionalidad y las variables atmosféricas en el COVID - 19*.
- Troncoso, F. H., & Fernández, N. E. (2021). Limpieza, corrección y geocodificación de grandes bases de datos de direcciones mediante minería de texto. *Universidad ciencia y tecnología*, 25(109 (2021)). doi:<https://doi.org/10.47460/uct.v25i109.451>

- Vargas, J. A., & Álvarez, D. H. (2013). Proceso de geocodificación de direcciones en la ciudad de Medellín, una técnica determinística de georreferenciación de direcciones. *Ingenierías USBMed*, 4(1). doi:<https://doi.org/10.21500/20275846.278>
- Waang, M., Jiang, A., Gong, L., Guo, W., Li, C., Zheng, J., . . . Li, H. (2020). Temperature Significantly Change COVID-19 Transmission in 429 cities. *MedRxiv*. doi:<https://doi.org/10.1101/2020.02.22.20025791>
- Wasserman, A. (2013). Bases epidemiológicas para la comprensión de los factores de riesgo.
- Yupari, I., Bardales, L., Rodriguez, J., Barros, J., & Rodriguez, Á. (2021). Factores de riesgo de mortalidad por COVID-19 en pacientes hospitalizados: Un modelo de regresión logística. *Revista de la Facultad de Medicina Humana*. doi:<http://dx.doi.org/10.25176/rfmh.v21i1.3264>
- Žižek, S. (2020). *Pandemia: La covid-19 estremece al mundo*. Anagrama.