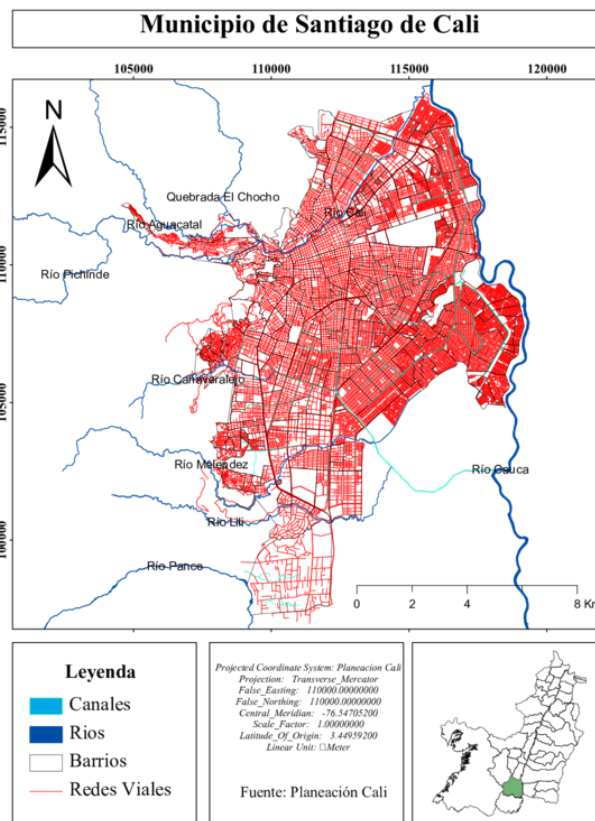




Análisis Espacio-Temporal del COVID-19 y Construcción de Redes de Contacto en Santiago de Cali

**Manuel Enrique Luna Alegría
 Christian Camilo Cadena Lemos**



**Análisis Espacio-Temporal del COVID-19
y Construcción de Redes de Contacto
en Santiago de Cali**

**Manuel Enrique Luna Alegría
Christian Camilo Cadena Lemos**

**Trabajo de grado para optar al título de
Estadístico**

Director

MSc. David Arango Londoño

Co-director

PhD. Jaime Mosquera Restrepo

**UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE ESTADÍSTICA
SANTIAGO DE CALI**

2021

DEDICATORIA

A nuestros padres, a quienes debemos todo.

A nuestros hermanos, un gran apoyo emocional.

A la memoria de Danna Gonzales.

AGRADECIMIENTOS

A Dios, por permitirnos alcanzar esta meta.

Este trabajo no hubiera sido posible sin el constante e incondicional apoyo que nos brindó el profesor David Arango Londoño. También, agradecemos la confianza que la secretaria municipal de salud de Santiago de Cali depositó en nosotros.

Gracias a todo el cuerpo docente de la escuela de estadística, por aportar su grano de arena en nuestra formación como profesionales íntegros.

Gracias a todos los compañeros y amigos, que de alguna u otra forma fueron cruciales para que llegásemos a este punto.

CALIFICACIÓN

Jurado 1:_____

Jurado 2:_____

Índice general

Índice general	XI
Índice de tablas	XIII
Índice de figuras	XV
Resumen	XVIII
Abstract	XIX
1 Introducción	23
1.1 Descripción del problema	23
1.2 Justificación	24
1.3 Antecedentes	26
1.4 Pregunta de investigación	29
1.5 Objetivos	29
1.5.1 Objetivo general	29
1.5.2 Objetivos específicos	29
2 Marco Teórico	30
2.1 Marco Conceptual	30
2.1.1 COVID-19	30
2.1.2 Transmisión	31
2.1.3 Ritmo Reproductivo R_0	31
2.1.4 Duración de la Enfermedad	32
2.1.5 Semana Epidemiológica	32
2.2 Marco Teórico para el Análisis de Redes	32
2.2.1 Ciencia de las Redes	32
2.2.2 Teoría de Grafos	32
2.2.3 Teoría Algebraica de Grafos	34
2.2.4 Grado de un Vértice	34
2.2.5 Índices de Centralidad	35

2.3 Marco Teórico para el Análisis Temporal	36
2.3.1 Suavización Spline	37
2.4 Marco Teórico para el Análisis Espacial	37
2.4.1 Estadística Espacial	37
2.4.2 Patrones Puntuales Espaciales	38
2.4.3 Análisis de la aleatoriedad espacial (AEC - Proceso homogéneo de Poisson)	39
2.4.4 Estimación Kernel de la función de intensidad	43
3 Metodología	45
3.1 Población de Estudio	45
3.2 Depuración de los datos	46
3.3 Análisis Estadístico	47
3.3.1 Análisis Espacio-Temporal	47
3.3.2 Análisis de redes de contactos	50
4 Resultados y discusión	52
4.1 Análisis estadístico descriptivo	52
4.2 Análisis estadístico espacio-temporal	56
4.2.1 Análisis Patrones Puntuales	58
4.3 Análisis de redes de contacto	84
5 Conclusiones, limitaciones y recomendaciones	91
5.1 Conclusiones	91
5.2 Limitaciones	92
5.3 Recomendaciones	93
A Figuras	94
B Tablas	96

Índice de tablas

3.1	Cantidad de casos positivos por Periodo.	48
4.1	Frecuencia de casos positivos entre enero y septiembre del 2020, discriminados por sexo.	53
4.2	Frecuencia de casos positivos COVID-19 según la Nacionalidad	53
4.3	Ranking de casos positivos COVID-19 por barrios	54
4.4	Casos COVID-19 en Santiago de Cali por estrato	55
4.5	Estadísticas de la edad de los casos positivos por COVID-19, discriminado por estrato	56
4.6	Estadísticas de las semanas epidemiológicas discriminado por estrato .	56
4.7	Prueba de Aleatoriedad Espacial por el Método de Cuadrantes con los criterios Test χ^2 y MonteCarlo para los 16 Periodos	58
4.8	Anchos de banda para la estimación de intensidad y función K,G y L por medio de Validación Cruzada	59
4.9	Resumen del análisis espacio temporal	84
4.10	Frecuencia de los tipos de enlaces generados por los individuos positivos de COVID-19	85
4.11	Visualización de resultados de pruebas realizadas para detección de COVID-19	86
4.12	Ranking de los individuos positivos con COVID-19 con mayor contactos registrados	87
B.1	Lista de variables utilizadas en el análisis estadístico	96
B.2	Lista de variables utilizadas para la construcción de las redes de contacto	97
B.3	Paquetes en R utilizados en el estudio	98

Índice de figuras

2.1	Baddeley et al. (2015).Función-k empírica (línea sólida) y función-k teórica para un proceso poisson (línea punteada)	41
2.2	Baddeley et al. (2015).Función-L empírica (línea sólida) y función-L teórica para un proceso poisson (línea punteada)	42
2.3	Bivand et al. (2013). Contraste entre la función G teórica y la función G empírica, para tres patrones de puntos	43
4.1	Frecuencia de casos positivos con nacionalidades diferentes a Colombia	53
4.2	Casos positivos de COVID-19 por barrios y estratos	54
4.3	Visualización de la edad y las semanas epidemiológicas por estrato . . .	55
4.4	Visualización global del comportamiento de la enfermedad entre Enero 18 y Septiembre 26 del 2020	56
4.5	Visualización de la tendencia del fenómeno mediante suavización Spline	57
4.6	Patrón de Puntos - Periodo 1	59
4.7	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 1	60
4.8	Función G, K y L - Periodo 1	60
4.9	Patrón de Puntos - Periodo 2	61
4.10	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 2	62
4.11	Función G, K y L - Periodo 2	62
4.12	Patrón de Puntos - Periodo 3	63
4.13	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 3	64
4.14	Función G, K y L - Periodo 3	64
4.15	Patrón de Puntos - Periodo 4	65
4.16	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 4	65
4.17	Función G, K y L - Periodo 4	66
4.18	Patrón de Puntos - Periodo 5	66
4.19	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 5	67
4.20	Función G, K y L - Periodo 5	67
4.21	Patrón de Puntos - Periodo 6	68
4.22	Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 6	68
4.23	Función G, K y L - Periodo 6	69
4.24	Patrón de Puntos - Periodo 7	69

4.25 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 7	70
4.26 Función G, K y L - Periodo 7	70
4.27 Patrón de Puntos - Periodo 8	71
4.28 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 8	71
4.29 Función G, K y L - Periodo 8	72
4.30 Patrón de Puntos - Periodo 9	72
4.31 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 9	73
4.32 Función G, K y L - Periodo 9	73
4.33 Patrón de puntos - Periodo 10	74
4.34 Estimación no paramétrica del intensidad temporal y espacial - Periodo 10	75
4.35 Función G, K y L - Periodo 10	75
4.36 Patrón de puntos - Periodo 11	76
4.37 Estimación no paramétrica del intensidad temporal y espacial - Periodo 11	76
4.38 Función G, K y L - Periodo 11	76
4.39 Patrón de puntos - Periodo 12	77
4.40 Estimación no paramétrica del intensidad temporal y espacial - Periodo 12	77
4.41 Función G, K y L - Periodo 12	78
4.42 Patrón de puntos - Periodo 13	78
4.43 Estimación no paramétrica del intensidad temporal y espacial - Periodo 13	79
4.44 Función G, K y L - Periodo 13	79
4.45 Patrón de puntos - Periodo 14	80
4.46 Estimación no paramétrica del intensidad temporal y espacial - Periodo 14	80
4.47 Función G, K y L en Periodo 14	81
4.48 Patrón de puntos - Periodo 15	81
4.49 Estimación no paramétrica del intensidad temporal y espacial - Periodo 15	82
4.50 Función G, K y L - Periodo 15	82
4.51 Patrón de puntos - Periodo 16	83
4.52 Estimación no paramétrica del intensidad temporal y espacial - Periodo 16	83
4.53 Función G, K y L - Periodo 16	84
4.54 Ranking de los individuos positivos con COVID-19 con mayor contactos registrados	86
4.55 Fragmento de la matriz de adyacencia	87
4.56 Visualización de algunos arboles estrella del primer bosque	88
4.57 Visualización de algunos arboles estrella del primer bosque	88
4.58 Visualización de algunos arboles estrella del primer bosque	88
4.59 Visualización de algunos arboles estrella del segundo bosque	89
4.60 Visualización de algunos arboles estrella del segundo bosque	89
4.61 Visualización de algunos arboles estrella del segundo bosque	89
A.1 Kolaczyk and Csárdi (2015). Se muestran las cuatro centralidades de vértices para una red de clubes de Karate. El vértice profesor es de color amarillo y el de instructor el azul	94

A.2	https://healthdataminer.com/data-mining/ . Metodología CRISP-DM	95
-----	--	----

Resumen

É STAR preparado es importante, saber esperar lo es aún más, pero aprovechar el momento adecuado es la clave de la vida.

Arthur Schnitzler.
1862–1931

La enfermedad COVID-19 desde sus inicios en China, se ha convertido en una problemática mundial por su alta tasa de contagio y la fácil propagación del virus SARS-coV-2. En este trabajo, se propone un análisis estadístico espacio-temporal de los casos positivos acumulados del COVID-19, desde enero 2020 hasta el 26 de septiembre del 2020 en Santiago de Cali (Colombia). También, una construcción de redes de contactos de los individuos notificados como positivos.

Por las características de la pandemia, el análisis espacio-temporal ayudó a entender que zonas de la ciudad se habían visto mayormente afectadas por la enfermedad, además, mediante pruebas de aleatoriedad espacial se logró determinar los tipos de patrones que se presentaban en diferentes periodos de tiempo. Las redes de contacto, fueron una ayuda adicional de prevención. Es decir, si un individuo era positivo por COVID-19, la IPS donde este se encontraba realizaba otra encuesta para recopilar información importante, como, sus contactos más cercanos, tipo de contacto, etc. Luego, con esta información se construyeron la red de contactos de cada individuo, visualizando de forma previa posibles casos positivos de COVID-19. Estas metodologías podrían ser herramientas útiles para la implementación de planes de mitigación más eficientes contra la COVID-19 u otras enfermedades con características similares de propagación.

Palabras clave: COVID-19, estadística espacial, pandemia, teoría de grafos, análisis temporal, aleatoriedad espacial, epidemiología, Santiago de Cali.

Abstract

$\mathcal{L}$ ^{IFE} is not
a problem to be solved, but a reality to be experienced.

Soren Kierkegaard.
1813–1855

Since its inception in China, COVID-19 disease has become a worldwide problem due to its high infection rate and the easy spread of the SARS-coV-2 virus. In this paper, we propose a spatial-temporal statistical analysis of the accumulated positive cases of COVID-19, from January, 2020 until September 26, 2020 in Santiago de Cali (Colombia). Also, a construction of contact networks of the individuals notified as positive.

Due to the characteristics of the pandemic, the spatio-temporal analysis helped to understand which areas of the city had been most affected by the disease, in addition, through spatial randomization tests it was possible to determine the types of patterns that were present in different periods of time. The contact networks were an additional preventive aid. That is, if an individual was positive for COVID-19, the IPS where he/she was located conducted another survey to collect important information, as, his/her closest contacts, type of contact, etc. Then, with this information, the contact network of each individual was constructed, visualizing in advance possible positive cases of COVID-19. These methodologies could be useful tools for the implementation of more efficient mitigation plans against COVID-19 or other diseases with similar propagation characteristics.

Key words: COVID-19, spatial statistic, pandemic, graph theory, temporal analysis, spatial randomness, epidemiology, Santiago de Cali.

CAPÍTULO 1

Introducción

La vida es muy peligrosa. No por las personas que hacen el mal, sino por las que se sientan a ver lo que pasa.

Albert Einstein.
1879–1955

1.1. Descripción del problema

El COVID-19 es una enfermedad causada por el coronavirus tipo 2 (SARS-CoV-2)¹, se ha propagado con rapidez por el mundo, es tal el caso que, el 11 de marzo del 2020 fue declarado pandemia por la Organización Mundial de la Salud OMS. Este virus tiene sus inicios en diciembre del 2019 en China, exactamente en Wuhan, provincia de Hubei (Huang et al., 2020). Según datos proporcionados por la extensión creada por Google para el COVID-19², hasta el 30 de septiembre del 2020 en el mundo se registró cerca de treinta y tres millones quinientos sesenta y un mil setenta y siete casos confirmados (33,561,077), entre los cuales, un millón cinco mil cuatro (1,005,004) son casos fallecidos.

Colombia no es la excepción, desde el primer caso registrado en el país (06 de marzo del 2020) al 30 de septiembre del 2020, registró ochocientos veintinueve mil seiscientos setenta y nueve (829,679) casos confirmados de los cuales, veinticinco mil novecientos noventa y ocho (25,998) son casos fatales. De los casos confirmados, el Valle del Cauca registró sesenta y dos mil setecientos cuatro (62,704), donde su capital Santiago de

¹El Comité Internacional de Taxonomía de los Virus (ICTV, por sus siglas en inglés), eligió este nombre porque el virus está genéticamente relacionado con el coronavirus responsable del brote de SARS de 2003, aunque se trata de dos virus diferentes.

²<https://www.google.com/covid19-map/?hl=es-419>

Cali registró cuarenta y siete mil cuatrocientos ochenta y un casos (47,481)³.

Para comprender las características epidemiológicas, clínicas y virológicas del COVID-19, la OMS en conjunto con unos expertos crearon protocolos o formularios para investigaciones tempranas de la enfermedad (Grupo et al., 2019). Estos, están a disposición de los países en la página web www.who.int/godata, algunos son: Protocolo de transmisión para los primeros casos COVID-19 y sus contactos cercanos (FFX (FFX), Protocolo de estudio para la transmisión de COVID-19 en hogares familiares (HH), Protocolo para la evaluación de factores de riesgo de COVID-19 para personal sanitario (HW), etc. Se evidencia la importancia de tener acceso a la georreferenciación de los casos activos, así como conocer las redes de contagios para identificar posibles casos aislados⁴ y controlar la propagación del virus de una forma más efectiva.

Por otra parte, la secretaría de salud pública de Santiago Cali cuenta con una base de datos más completa, donde caracterizan a los casos positivos usando los formularios estandarizados por la Organización Mundial de la Salud para el control de los casos. Algunas variables que se utilizan para la caracterización son: dirección, redes de contacto, entre otras variables importantes que se podrían utilizar para mitigar este y futuros virus.

Controlar la propagación del COVID-19 ha sido un desafío para los entes gubernamentales de todos los países afectados, las características del virus hace que esta tarea sea casi imposible. Debido a lo anterior, se presenta la importancia del análisis espacio-temporal en la investigación de la propagación de una pandemia, pues se puede explicar mejor como es la difusión del virus en el espacio-tiempo (Giuliani et al., 2020). Adicionalmente, se propone la construcción de las redes de contacto mediante teoría de grafos de los casos positivos por COVID-19, con el fin de proporcionar más herramientas en la toma de decisiones de salud pública.

1.2. Justificación

Al rededor del mundo muchos académicos e investigadores tratan de contribuir al problema que se presenta por la pandemia del COVID-19, donde las principales contribuciones hacen uso de modelos epidemiológicos deterministas (Kucharski et al., 2020; Liu et al., 2020; Sun et al., 2020). Estas investigaciones se centran en la evolución

³La fecha de revisión de los datos es del 30 de septiembre del 2020, sin embargo debido a las características de la pandemia es susceptible a cambios. Los datos son suministrados por la página web <https://coronaviruscolombia.gov.co/Covid19/index.html>

⁴Casos aislados: Casos positivos COVID-19, pero no se logra identificar la fuente de contagio.

temporal de los fenómenos, con el fin de predecir nuevos casos de infectados, fallecidos y curados, obteniendo información valiosa para la implementación de políticas de salud pública que permitan un manejo adecuado de la situación.

En los últimos 20 años en el área de la salud, la epidemiología espacial ha sido importante para evaluar la incidencia de enfermedades epidémicas, permitiendo desarrollar sistemas de investigación médicas aplicadas (Kirby et al., 2017).

Estas metodologías tuvieron que evolucionar de forma muy rápida debido a la pandemia que se está viviendo actualmente, ratificando que aun hay un largo camino sin explorar sobre técnicas espaciales y temporales aplicados en la salud pública.

El principal vector de contagio de la enfermedad es el ser humano y la forma de contagio es mediante micro partículas expulsadas por la boca y nariz. Por lo cual, el contacto estrecho entre personas causa una celeridad en la propagación del virus. Por lo anterior, se propone analizar los casos positivos por COVID-19 a través del espacio y tiempo, con el fin de identificar patrones poblacionales y facilite la segmentación de zonas donde ocurren mayor o menor contagio. Finalmente, se propone una construcción de las redes de contacto de los individuos positivos, el cual podría facilitar la identificación de los individuos que tengan una mayor red de contactos y esta forma identificar a los posibles nuevos casos positivos.

La ciudad de Santiago de Cali desde el primer caso positivo hasta la fecha del 30 de septiembre del 2020 fue una de las ciudades con mayor *tasa de incidencia diaria* del país, según información reportada por el Instituto Nacional de Salud (INS). Este estudio, se centra en la Ciudad de Santiago de Cali, con el fin de estudiar y analizar la evolución espacio-temporal del virus y construir las redes de contacto de los individuos positivos por COVID-19.

Conocer la estructura de la red permite a los modelos calcular la dinámica epidémica a escala poblacional a partir del comportamiento individual de las infecciones (Keeling and Eames, 2005). Por lo tanto, las características de las redes de contacto se han convertido en importantes temas aplicados que puedan mejorar la comprensión y la predicción de los patrones de epidemias y medidas de intervención. Estas redes, están conformadas por individuos catalogados como casos positivos del COVID-19, los cuales, tienen una red de contacto que podrían ser susceptibles a contagiarse. Analizar las redes de contacto se hace necesario para encontrar las principales causas de contagio del virus, es decir, el tipo de contacto o enlace que tuvieron los individuos positivos.

1.3. Antecedentes

Con el fin de conocer lo que se ha investigado acerca del estudio espacio temporal de casos del COVID-19, así como el análisis de la evolución de redes de contagio de estos casos, se realizó una revisión bibliográfica. A continuación, se presentan algunas de las investigaciones internacionales más relevantes del tema.

[Ranjan and Misra \(2015\)](#), en esta investigación titulada “ **Epidemic Disease Propagation Detection Algorithm using MapReduce for Realistic Social Contact Networks**”, exponen la importancia que tienen las redes de contactos para el control y prevención de epidemias como la influenza. Los autores proponen un algoritmo llamado MapReduce ejecutado en la framework Hadoop, basados en información de teléfonos inteligentes y con modelos matemáticos epidemiológicos, lograron crear las redes de contacto y detectar el estado de salud de cada individuo. Obteniendo redes de contacto muy amplias, donde algunas alcanzaban un número de nodos de 500,000.

Encontraron que, la velocidad de propagación de las epidemias como la influenza es alta entre las redes de contacto de un individuo infectado, además, lograron detectar algunos casos positivos donde no tenían ninguna información de ubicación. Se probó, la eficiencia del algoritmo y concluyeron que la mejor forma para trabajar estas grandes cantidades de datos es mediante cluster.

[Zhao et al. \(2019\)](#), proponen una red neuronal artificial semi-automática llamada Social Media Nested Epidemic Simulation (SimNest) para analizar las redes de contacto por medio de redes sociales e identificar posibles casos de influenza. En el trabajo titulado “**Online Flu Epidemiological Deep Modeling on Disease Contact Network**”, por medio de epidemiología computacional presenta la posibilidad de trabajar con datos en tiempo real proporcionados por Twitter, de esta forma, lograr crear redes de contacto más eficientes para poder analizar, identificar y mitigar la epidemia de la influenza. SimNest, abarca dos componentes muy importantes, el primero, es la capacidad de hacer un análisis espacio-temporal de la epidemia, y la segunda es implementar esto en tiempo real mediante información de las redes sociales. Así, la capacidad de respuesta de un país en cuestión de salud pública podrá ser más eficiente.

Además, los autores realizaron una serie de comparaciones con otros métodos, algunos son: *Linear Autoregressive Exogenous model (LinARX)*, *Multi-variable linear regression model (multiLinReg)*. Donde concluyen, sobre las ventajas de combinar el análisis espacio-temporal y la minería de redes sociales en el estudio de epidemias como la influenza. Los datos con los que probaron a SimNest, fueron obtenidos de tweets en

diferentes estados de E.E.U.U. entre el 1 enero del 2011 y el 15 de abril del 2015.

Huang et al. (2020), en el trabajo titulado **“Spatial-Temporal Distribution of COVID-19 in China and its Prediction: A Data-Driven Modeling Analysis”**, realiza un análisis espacial con datos de la ciudad de Wuhan, en la provincia de Hubei y propone un modelo logístico para la predicción de futuros casos positivos por COVID-19. Los datos utilizados para esta investigación corresponden a los casos de la provincia de Hubei y zonas cercanas, estos, fueron recolectados entre el 13 de enero del 2020 y el 9 de marzo del 2020.

En primera medida, los autores estudian si existe correlación espacial del virus en zonas adyacentes y en distancias geográficas. Mediante la aplicación del índice de Moran⁵, encuentran que no hay correlación significativa en distancias espaciales geográficas, pero para zonas más cercanas a la provincia de Hubei si presentan correlación espacial. Esto indica que, a medida que los casos en la provincia de Hubei aumenten, las zonas adyacentes se verán más afectas y los vecinos de estas zonas van a empezar a presentar casos positivos de la enfermedad. Por otra parte, se estudió el modelo logístico con las características de los datos recolectados, con la finalidad de lograr predecir futuros casos. Se concluyo que, la evolución del virus se puede reducir gradualmente mediante medidas de prevención y aislamiento, porque la correlación espacial de la enfermedad es significativa solo en las zonas vecinas donde existe el virus.

Giuliani et al. (2020), esta investigación tiene como título **“Modelling and Predicting the Spatio Temporal Spread of Coronavirus Disease 2019 (COVID-19) in Italy”**. Los autores en este estudio tienen como objetivo modelar y predecir la propagación espacial y temporal de la enfermedad COVID-19 en Italia. Los datos fueron extraídos por medio de técnicas de raspado de datos en la página web del Departamento de Protección Civil de Italia junto con el Ministerio de Salud Italiano, durante el 26 de febrero del 2020 y el 3 de marzo del 2020.

El análisis estadístico en este estudio se basa principalmente en la evolución del número de infecciones diarias, las cuales consideran y estiman la dinámica temporal de las infecciones, así como su mecanismo geográfico de propagación. El número esperado de infecciones en una provincia en el transcurso del día se descomponen de acuerdo a tres componentes aditivos que representan tres canales diferentes, a través de estos se considera que pueden aumentar los contagios. El primer componente de contagio (catalogado como componente endémico) lo representan por el mecanismo de

⁵**Índice de Moran:** es una medida de autocorrelación espacial desarrollada por Patrick Alfred Pierce Moran.

propagación del COVID-19 dentro de cada provincia, el cual determina la dinámica temporal del contagio dentro de esta. El segundo canal (catalogado como componente epidémico) lo caracterizan por ser atribuible a contagios originados en otras provincias, y en particular de provincias que están geográficamente cercanas entre sí. El tercer canal (catalogado como componente endémico) lo atribuyen a las condiciones específicas de la provincia que determinaron los primeros centros de infecciones y la exposición inicial al riesgo de contagio.

Como resultados encontraron diferencias en la dinámica de la epidemia en diferentes provincias de Italia, lo que demuestran la importancia crucial de un fuerte nivel de coordinación nacional para la aplicación homogénea de medidas de control. Encontrando importantes predicciones a nivel local por medio del análisis espacio-tiempo, ofreciendo una perspectiva decisiva sobre cuando y qué áreas pueden verse más afectadas, permitiendo que los tomadores de decisiones intervengan en las políticas locales.

Wang et al. (2020), en su trabajo titulado como **“Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in the United States”** estudian la dinámica espacio-temporal de la enfermedad, con datos del 20 de enero del 2020 al 25 de abril del 2020 de 48 estados continentales de Estados Unidos. Teniendo en cuenta las medidas de control, los efectos ambientales, los factores socioeconómicos, recursos de los servicios de salud y las condiciones demográficas, los autores proponen un nuevo modelo de transmisión de la enfermedad espacio temporal no paramétrico para los datos que tenían de la epidemia, con el fin de predecir el comportamiento del virus en el espacio y tiempo. El objetivo de este estudio está dividido en tres partes, la primera enfocada en desarrollar un nuevo marco dinámico de modelado epidémico para los datos de vigilancia de salud pública, con el fin de estudiar el patrón espacio-temporal de la propagación del COVID-19. En segundo lugar, comprender los factores que contribuyen a la propagación, modelando los casos de infección diaria considerando factores demográficos, ambientales y socioeconómicos de los Estados Unidos. En tercer lugar, proyectaron el patrón espacio-temporal de la propagación, para predecir el recuento diario de infecciones y el recuento de muertes.

En cuanto a la estimación del modelo, se propone un enfoque de cuasiverosimilitud por medio de la aproximación de spline penalizada y la técnica de mínimos cuadrados ponderados iterativamente. Encontrando como resultado, que este método proporciona resultados de predicción de la incidencia del virus a corto plazo notablemente precisos. Además, los autores aseguran que el modelo puede implementar fácilmente un mapeo de enfermedades para ilustrar áreas de alto riesgo y comprender la dinámica de la propagación de la enfermedad, para así evaluar cómo el virus puede desarrollarse a través del tiempo y el espacio, ayudando a la formulación de políticas que permitan controlar la situación.

1.4. Pregunta de investigación

¿Como ha sido la evolución espacio-temporal de la enfermedad COVID-19 y en que medida las redes de contactos podrían aportar información preventiva de nuevos casos a partir del vértice central⁶ de la red?

1.5. Objetivos

1.5.1. Objetivo general

Examinar mediante un análisis Espacio Temporal y construcción de Redes de Contacto, la evolución de los casos positivos de la enfermedad COVID-19 entre los meses de enero y septiembre del 2020 en Santiago de Cali.

1.5.2. Objetivos específicos

- ❶ Desarrollar un análisis descriptivo temporal de la evolución de los casos positivos por COVID-19 en Cali.
- ❷ Desarrollar un análisis espacial de la evolución de los casos positivos por COVID-19 en Cali.
- ❸ Analizar la estructura de las redes de contacto por COVID-19 a partir de los individuos reportados como caso positivo activo.

⁶vértice central: Caso positivo activo inicial para cada red de contacto.

CAPÍTULO 2

Marco Teórico

U^{NA vez terminado}
el juego el rey y el peón vuelven a la misma caja.

Proverbio italiano.

Después de haber revisado estudios previos acerca del tema, en esta sección se presenta el marco teórico de la investigación. Está dividido en cuatro partes.

La primera hace referencia al marco conceptual, por lo que la compone definiciones epidemiológicas sobre las características del COVID-19. La segunda parte está la definición, características y análisis de la ciencia de redes. La tercera parte abarca conceptos correspondientes al análisis exploratorio de series temporales, así como métodos de suavizado. La cuarta parte, habla sobre la teoría de la estadística espacial, haciendo hincapié en los métodos que serán empleados en este estudio.

2.1. Marco Conceptual

A continuación, se presenta algunos conceptos epidemiológicos referentes a la enfermedad COVID-19 que son relevantes para la investigación.

2.1.1. COVID-19

Según la definición de la OMS, el COVID-19 es una enfermedad infecciosa causada por un virus llamado SARS-CoV-2, que en promedio el periodo de incubación es entre 7 y 14 días. Los síntomas asociados a esta enfermedad van desde tos seca, fiebre, dolor muscular, dificultad para respirar y en algunos casos puede presentar un cuadro clínico de deficiencia respiratoria aguda, choque séptico y sepsis ([Xiaoyi et al., 2020](#)).

Según estudios previos de otros brotes causados por coronavirus, se cree que la fuente de infección es de origen animal. En primera instancia, se cree que el reservorio del virus es el murciélago, sin embargo, aún se mantiene investigaciones para aclarar un posible segundo huésped como el pangolín ([Aylward, Bruce \(WHO\); Liang, 2020](#)). Además, como la prevalencia y distribución de coronavirus en diferentes especies de animales silvestres es alta, es posible encontrar casos infecciosos futuros en humanos, en especial en zonas donde hay interacción con estos animales ([Ministerio de Sanidad, 2020](#)).

2.1.2. Transmisión

Los mecanismos conocidos de transmisión son de animal-humano y de humano - humano. En algunas investigaciones se encontró que la forma de transmisión de animal-humano es únicamente mediante secreciones o contacto directo con el animal infectado, se cree que la transmisión se debe primordialmente por tener contacto con el aparato digestivo del animal. Por otra parte, se evidencio tres casos (dos perros en Hong-Kong y un gato en Bélgica) contagiados por el virus, la raíz del contagio puede haber sido sus dueños, pues estos convivían con sus mascotas estando infectados ([Scientifique, 2020](#); [Stacey Knobler, Adel Mahmoud, Stanley Lemon, Alison Mack, Laura Sivitz and Oberholtzer, 2004](#)). El mecanismo de transmisión humano-humano, es el más preocupante para la comunidad mundial, pues se puede contagiar por secreciones o microgotas infectadas que miden no más de 5 micras liberadas a través de la tos, estornudos, respiración o habla y pueden alcanzar una distancia de hasta 2 metros ([Chin et al., 2020](#)). Las manos es otro factor clave para la propagación del virus, pues si están contaminadas y hacen contacto con la mucosa de la boca, ojos o nariz, inmediatamente entra el virus al cuerpo del individuo. Además, algunos estudios sugieren que el virus puede durar en la superficie desde 30 minutos hasta dos días dependiendo de las características del material, las condiciones de humedad y temperatura.

2.1.3. Ritmo Reproductivo R_0

El R_0 es el número promedio de contagios nuevos que una persona infectada puede generar, si el número es alto, mayor será el potencial de propagación de la enfermedad. En el caso del COVID-19 se ha estimado un R_0 de 2.5, esto quiere decir que una persona contagiada tiene la capacidad de infectar al mes 2.5 personas de forma directa y 244 de forma indirecta ([González-jaramillo et al., 2020](#)). Sin embargo, este valor es susceptible a cambios y se disminuye mediante medidas de aislamiento social e higiene.

2.1.4. Duración de la Enfermedad

Cuando los síntomas de la enfermedad son leves, el tiempo promedio de duración es de dos semanas. Si los síntomas de la enfermedad son graves, la duración media pasa de tres a ocho semanas hasta la recuperación o la muerte ([Aylward, Bruce \(WHO\); Liang, 2020](#)). También, el tiempo medio que tarda en aparecer síntomas graves es alrededor de una semana.

2.1.5. Semana Epidemiológica

Es un periodo de tiempo que comprende los días de la semana, pero con la característica de que se empieza a partir del día domingo y termina el sábado. Esta medición estandarizada del tiempo, se compone de 52 semanas en el año y a este conjunto se le denomina calendario epidemiológico ([Vidaurre-Arenas et al.](#)).

2.2. Marco Teórico para el Análisis de Redes

En esta sección se presenta la teoría relacionada con las redes de contacto, como se construyen, que son, análisis, tipos, entre otros conceptos de interés para la investigación.

2.2.1. Ciencia de las Redes

Las redes son interacciones entre diferentes actores⁷ que de una u otra forma tienen alguna relación, esta puede ser cercana o lejana. El mundo está hecho por redes, desde las infraestructuras viales hasta las conexiones celulares de un individuo, cada interacción tiene un fin y un porque ([Luke, 2015](#)). Un ejemplo claro es el internet, donde hay millones de nodos conectados entre si, estos se conectan entre diferentes elementos por medio de links (bordes).

2.2.2. Teoría de Grafos

Matemáticamente un grafo se denota por $G=(V,E)$, el cual es un objeto abstracto conformado por vértices (nodos) V y bordes o lazos (links) E unidos en pares de vértices. El conjunto de vértices se denota por $V(G)$ y el conjunto de bordes se denota por $E(G)$, la cardinalidad de V es denotada por n y la de E esta denotada por m . Dos vértices unidos por un borde se denominan *endvertices*, estos dos vértices unidos son llamados subyacentes o vecinos. Los grafos pueden ser dirigidos o no dirigidos, en los grafos no dirigidos, el orden de los endvertices de un borde es inmaterial, este borde no dirigido que une los vértices $u,v \in V$ es denotado por u,v . En los grafos dirigidos, los bordes

⁷En la ciencia de redes los actores son nodos y estos no necesariamente implican ser seres vivos

dirigidos tienen un comienzo (cola) y un destino (cabeza). Un borde con origen $u \in V$ y destino $v \in V$, está representado por un par ordenado u, v . (Brandes and Erlebach, 2005, p.7)

Los grafos dirigidos y no dirigidos se dividen en tres tipos: Grafos múltiples, son aquellos con bordes paralelos, es decir que el mismo borde ocurre varias veces en E . Grafos simples, son aquellos que los vértices están unidos por un solo borde o mejor dicho no existen bordes paralelos. Grafos Ponderados, son aquellos que tienen un valor numérico en los vértices o en los bordes, en este caso solo se hablará de los pesos de bordes. Dependiendo del contexto, los pesos de borde pueden describir varias propiedades como el costo (por ejemplo, tiempo de viaje o distancia), capacidad, resistencia de interacción o similitud. El peso se denota por w para un borde $e \in E$, así será denotado como $w(e)$ (Brandes and Erlebach, 2005).

De acuerdo a la forma del grafo y las diferentes cualidades de este, así mismo se clasifican. Para efectos de este estudio se trabajarán con arboles y bosques, donde sus bordes son ponderados.

Un grafo toma el nombre de **árbol** cuando cada vértice esta conectado por un único borde y además no existe un circuito y Un bosque es la unión disjunta entre arboles (Aguila, 2013). Según el **teorema 5.5** del libro *Una introducción a las matemáticas discretas y teoría de grafos*, un árbol debe cumplir las siguientes cualidades:

- Si al grafo de árbol se le añade un nuevo borde, este creará un circuito en el grafo.
- Todo árbol tiene n vértices y $m=n-1$ bordes.
- Todo árbol no tiene circuito y esta conectado.

Además, estos tipos de grafos se subdividen de acuerdo a sus características y componentes que poseen. Si son dirigidos o no, se les denomina arboles dirigidos y arboles unidireccionales respectivamente. Los arboles etiquetados, son aquellos que tienen etiquetas en los vértices o bordes, entre otros. Para esta investigación se utilizará un caso particular de los grafos de árbol, denominado *árbol de estrella*. La diferencia entre los otros, es que posee una raíz o un único vértice, el cual esta conectado con otros vértices y a su vez estos vértices tienen un único enlace con la raíz. De esta forma un árbol de estrella con k vértices tendrá $k-1$ hijos (Aguila, 2013).

2.2.3. Teoría Algebraica de Grafos

Anteriormente, se menciona que un grafo dirigido es quien la conexión entre dos vértices solo es posible en una dirección y los grafos no dirigidos el enlace es bidireccional, es decir, que no es significativo la dirección en que se realiza la conexión entre par de vértices. De acuerdo, a lo anterior las matrices de incidencia y adyacencia son diferentes en cada uno de los casos.

Dos grafos denotados $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$, son isomórficos si existe una biyección $\phi = V_1 \rightarrow V_2$ cumpliendo con:

$$\forall u, v \in V : (u, v) \in E_1 \iff (\phi(u), \phi(v)) \in E_2 \quad (2.2.1)$$

De este modo, se consideran dos grafos que son iguales si son isomorfos. Además, si un isomorfismo mapea un grafo sobre si mismo se le denomina automorfismo. Para los grafos no dirigidos el isomorfismo y el automorfismo es análogo ([Brandes and Erlebach, 2005](#)).

Matriz de Incidencias, para un grafo dirigido $G=(V,E)$, la matriz de incidencias corresponderá a $B_{n \times m}$ con n filas y m columnas, donde $V = (v_1, \dots, v_n)$ y $E = (e_1, \dots, e_m)$. Las entradas de $b_{i,j}$ debe satisfacer con:

$$b_{i,j} = \begin{cases} -1, & \text{si } v_i \text{ es el origen de } e_j \\ 1, & \text{si } v_i \text{ es el destino de } e_j \\ 0 & \text{en otro caso} \end{cases} \quad (2.2.2)$$

Matriz de Adyacencia, en un grafo dirigido simple $G=(V,E)$ con $V = (v_1, \dots, v_n)$, es una matriz $n \times n$ representada por $A_{n \times n}$ donde $A(G) = (a_{i,j})_{1 \leq i,j \leq n}$ con:

$$a_{i,j} = \begin{cases} 1, & \text{si } (v_i, v_j) \in E \\ 0 & \text{en otro caso} \end{cases} \quad (2.2.3)$$

Para los grafos no dirigidos simples la matriz A es simétrica con $a_{i,j}=1$, si v_i y v_j son adyacentes. Sin embargo si el grafo no dirigido es ponderado, los valores de 1 cambian por la ponderación correspondiente a cada borde ([Brandes and Erlebach, 2005](#)).

2.2.4. Grado de un Vértice

El grado de un vértice se refiere a la suma de los bordes que contiene cada vértice, es denotado como $deg(v)$. La suma de los grados de los vértices se obtiene de la suma total de los bordes multiplicado por dos ([Aguila, 2013](#)).

$$\sum_{u \in V(G)} deg(u) = 2 * card(E(G)) \quad (2.2.4)$$

Donde, G denota al grafo y u es un vértice perteneciente al conjunto de vértices V del grafo G .

Un vértice es par o impar, si el grado de este es par o impar correspondientemente. Por otra parte, el número de vértices de grado impar en un grafo, es siempre par (Aguila, 2013). El grado promedio del grafo se halla de la siguiente forma:

$$d(G) = \frac{1}{\text{card}(V(G))} * \sum_{u \in V(G)} \text{deg}(u) \quad (2.2.5)$$

2.2.5. Índices de Centralidad

Los índices de centralidad sirven para cuantificar la importancia de un vértice o borde, estos se usan para dar respuestas a preguntas como: ¿Que actores en una red social influyen más? ¿Qué vértices son los principales de una red?, entre otras.

Existen diferentes tipos de medidas de centralidad, la más común es el grado de centralidad de un vértice. También, se definirán los tres índices clásicos para medir la centralidad de los vértices, estos son: cercanía, intermediación y centralidad del vector propio (Kolaczyk and Csárdi, 2015).

Grado de Centralidad de un Vértice, es la medida más simple de centralidad se denota como el grado $d(v)$ de v si el grafo no está dirigido. En grafos dirigidos, se debe tener en cuenta dos variaciones de centralidad: grado de centralidad entrando (in-Degree) $ciD(v) = d^-(v)$ y grado de centralidad saliendo (out-Degree) $coD(v) = d^+(v)$ (Kolaczyk and Csárdi, 2015). El grado de centralidad es una medida local, porque el valor de centralidad de un vértice solo está determinado por el número de sus vecinos. Esta medida es aplicable en redes que representen una situación estática, por ejemplo, se quiere estudiar los resultados de una votación y es de interés analizar el vértice que obtuvo más votos directos. (Brandes and Erlebach, 2005).

Medida de Centralidad de Proximidad (Closeness Centrality), Este enfoque fue introducido por Sabidussi (1966), consiste en medir que tan central es un vértice dependiendo del número de nodos que tiene más cercanos (Kolaczyk and Csárdi, 2015). La ecuación planteada es,

$$c_{Cl}(v) = \frac{1}{\sum_{u \in V} \text{dist}(v, u)} \quad (2.2.6)$$

Donde $\text{dist}(v, u)$ es la mínima longitud o distancia geodésica que une los dos vértices $(v, u) \in V$. Por lo regular para realizar comparaciones entre grafos, esta medida suele

estar normalizada (Kolaczyk, 2009).

Medida de Centralidad de Intermediación (Betweenness Centrality), la medida de centralidad de intermediación más común fue introducida por Freeman (1977), consiste en resumir la extensión en que un vértice esta entre otros pares de vértices (Kolaczyk and Csárdi, 2015). Esta medida se expresa en la siguiente ecuación:

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)} \quad (2.2.7)$$

Donde $\sigma(s, t|v)$ es el número total de caminos más cortos entre s y t que pasan por v , $\sigma(s, t)$ es el número total de caminos más cortos entre s y t independientemente de si pasan o no por v (Kolaczyk and Csárdi, 2015).

Medida de Centralidad del Vector Propio (Eigenvector Centrality), Al igual que las medidas de centralidad anteriormente mencionadas, hay mucha variedad para medir centralidad del vector propio. Sin embargo, la más usada es la propuesta por Bonacich (1972) siguiendo el trabajo de Katz (1953), la medida se basa en nociones de “estado” o “prestigio”, es decir que tratan de capturar la idea de los vecinos más centrales al vértice (Kolaczyk and Csárdi, 2015). Esta medida se puede expresar en términos de vector propio, a continuación, se presenta la ecuación correspondiente:

$$C_{Ei}(v) = \alpha \sum_{(u,v) \in E} C_{Ei}(u) \quad (2.2.8)$$

El vector $C_{Ei} = [C_{Ei}(1), \dots, C_{Ei}(Nv)]^T$ es la solución al problema del valor propios $A_{C_{Ei}} = \alpha^{-1}C_{Ei}$. Donde A es la matriz de adyacencia para el grafo G . Bonacich argumentan que, la selección optima de α^{-1} es el valor propio mayor de A y por lo tanto el vector C_{Ei} será el correspondiente (Kolaczyk, 2009).

Para entender mejor estas medidas de centralidad ver figura A.1.

2.3. Marco Teórico para el Análisis Temporal

El objetivo de analizar el tiempo en este estudio es observar la tendencia del número de contagios. Para este análisis se consideran un conjunto de puntos que representen a través del tiempo el comportamiento del número de contagios, los cuales son caracterizados a través de una regresión no paramétrica.

Según [Durbán Reguera \(2009\)](#) existen diversas formas de realizar una regresión no-paramétrica, las cuales se basan explícitamente en los datos para especificar la forma que tiene el modelo: la curva en un punto dado sólo depende de las observaciones en ese punto y de las observaciones vecinas. Algunas de las técnicas de regresión no-paramétrica existentes son:

- Regresión polinomial local con pesos
- Kernels
- Splines
- Splines con penalizaciones (Pspines)

Las regresiones polinomiales se basan en el modelo lineal al añadir predictores, los cuales se obtienen al elevar cada uno de los predictores originales a una potencia. Por ejemplo, una regresión cúbica utiliza tres variables, X , X^2 y X^3 , como predictores. Esto es entonces, una manera simple de proporcionar un ajuste no lineal a los datos ([James, G., Witten, D., Hastie, T., Tibshirani, 2013](#)).

2.3.1. Suavización Spline

Un suavizador es una herramienta que representa la tendencia de la variable respuesta como función de uno o más predictores lineales X . Formalmente, [Durbán Reguera \(2009\)](#) define los Splines como funciones polinómicas a trozos sobre las que se imponen restricciones en los puntos de unión llamados nodos, estos puntos dividen el rango de x en regiones. Además, los splines dependen del grado del polinomio, número de nodos y la localización de los nodos.

La suavización Spline consiste entonces, en construir una curva que se ajuste a los datos por medio de una regresión polinómica de grado 3. El principal objetivo es minimizar la *suma de residuos al cuadrado penalizados*, por medio del ajuste de dos términos, donde el primero mida la proximidad a los datos y el segundo término penalice la curvatura de la función a través de un parámetro de suavizado ([Durbán Reguera, 2009](#)).

2.4. Marco Teórico para el Análisis Espacial

2.4.1. Estadística Espacial

[Giraldo h \(2002\)](#) propone que la estadística espacial es la reunión de un conjunto de metodologías apropiadas para el análisis de datos, que corresponden a la medición de variables aleatorias en diversos sitios (puntos del espacio o agregaciones espaciales)

de una región. De manera más formal, la estadística espacial trata con el análisis de realizaciones de un proceso estocástico $\{Z(s): s \in D\}$, donde $s \in R^d$ y representa una ubicación en el espacio euclidiano d-dimensional, $Z(s)$ es una variable aleatoria en la ubicación s el cual varía sobre un conjunto de índices $D \subset R^d$.

Además, Giraldo afirma que la estadística espacial se subdivide en tres grandes áreas de estudio que son: Geoestadística, Enmallados y Patrones Espaciales.

Dentro del caso de estudio, se hará referencia a los *Patrones Espaciales*, los cuales a diferencia de la geoestadística y enmallados, las ubicaciones pertenecen a un conjunto D que puede ser *discreto* o *continuo* y su selección no depende del investigador (D aleatorio). El propósito de esta área es determinar si la distribución de los individuos dentro de la región es aleatoria, agregada o uniforme.

2.4.2. Patrones Puntuales Espaciales

Para el desarrollo de los métodos estadísticos en el análisis de patrones de puntos espaciales, es necesario definir el concepto de proceso puntual, que se define como un mecanismo aleatorio cuyo resultado es un patrón de punto. (Rodríguez-cortés, 2018). Otra definición, la realiza (Schabenberger and Gotway, 2017), afirma que un proceso puntual es un arreglo (patrón) de puntos de un conjunto aleatorio D . Formalmente, define que un **patrón espacial** de puntos es un conjunto finito de eventos $\{Z(s) : s \in D\}$ para alguna región plana $D \subset \mathbb{R}^2$. Además de esto, cuando se tienen solo parte de los eventos, el patrón se llama patrón muestreado y cuando se registran todos los eventos de la realización, entonces el patrón es puntual.

Un concepto importante dentro del análisis de patrones puntuales es la **intensidad** λ , Baddeley et al. (2008) lo define como el valor esperado del número de puntos por unidad de área. La intensidad puede ser constante a lo largo de la ventana de observación (homogénea) o puede variar de una localización a otra (no homogénea).

A continuación, se definen tres propiedades importantes en el análisis de patrones puntuales.

- **Estacionariedad** Un proceso puntual $\mathbf{X}$ en $\mathbb{R}^2$, es llamado estacionario (homogéneo) si las propiedades estadísticas de $\mathbf{X}$ y $\mathbf{X} + \nu$ son idénticas para cualquier vector ν , donde se puede afirmar que estos tienen la misma distribución. (Jani-llian and Stoyan, 2008).
- **Isotropía** Un proceso puntual es isotrópico cuando es invariante a la rotación con respecto al origen (Waller and Gotway, 2004).

- **Ergodicidad** un proceso puntual ergódico supone analizar una muestra (un patrón puntual) de un tamaño apropiado que permita obtener resultados estadísticamente significativos. La ergodicidad esta expresada en términos de λ y representa la intensidad, que denota la media de puntos por unidad de área o volumen (Janine Illian and Stoyan, 2008).

En el análisis de los patrones espaciales de puntos, uno de los procedimientos estadísticos fundamentales consiste en el test de aleatoriedad espacial completa (AEC) (Rodríguez-cortés, 2018).

2.4.3. Análisis de la aleatoriedad espacial (AEC - Proceso homogéneo de Poisson)

Prueba de Monte Carlo

Es un caso especial de pruebas por simulación utilizado para evaluar la completa aleatoriedad espacial. La hipótesis es que, un patrón de puntos $Z(s)$ es una realización de un modelo de proceso ψ . Consiste en tener un patrón observado y simular otro bajo el ψ , luego, se construye un estadístico Q de prueba para evaluar si los datos observados y simulados pertenecen al mismo proceso (Schabenberger and Gotway, 2017).

Dependiendo de la hipótesis nula, el valor esperado del estadístico Q es diferente. Por ejemplo, el caso de patrones agregados, se espera un valor de Q pequeño cuando ψ es un proceso homogéneo de Poisson. Este método tiene ventajas al ser sus valores p exactos, por lo cual, no se requiere ninguna aproximación de la distribución teórica del estadístico de prueba. Sin embargo, en casos infinitos el valor-p es inexacto (Schabenberger and Gotway, 2017).

La desventaja que presenta este método, es porque se basa en simulaciones. Por lo tanto, factores como el número de iteraciones, tamaño de simulación, podrían afectar en la precisión de la prueba.

Prueba de conteo por cuadrantes - prueba X^2 para aleatoriedad espacial

Consiste en dividir el patrón de puntos en m cuadrantes de área, $a_1, a_2, \dots, a_m$. Si los puntos quedan distribuidos por su correspondiente área, se obtiene una estimación simple de la intensidad (Quispe Quispe, 2016). De este modo, la hipótesis nula es que los datos parten de un proceso homogéneo y la hipótesis alternativa se refiere que los datos son un proceso Poisson no homogéneo. Para evaluar estas hipótesis se utiliza el estadístico Chi-cuadrado de la siguiente forma:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \bar{n})^2}{\bar{n}}$$

Donde, n_{ij} es el número de eventos en un cuadrante ij , $\bar{n}$ es número esperado de eventos en un cuadrante. r es el número de filas y c el número de columnas, por lo cual el estadístico x^2 se distribuye chi-cuadrado con $r*c-1$ grados de libertad ([Schabenberger and Gotway, 2017](#)).

La K-Función

La K-Función es una de las técnicas que se usa para analizar la correlación de patrones puntuales en procesos estacionarios e isotrópicos, fue propuesta por el profesor británico Bryan Ripley en 1981.

Se define como $\lambda * K(d) = E(M)$, donde λ es la intensidad del proceso, es decir el número esperado de puntos d por unidad de área. En los análisis exploratorios, la estimación de K es una estadística útil para resumir los aspectos de inter-punto, es decir, la distancia de cada punto a todos los registros ([Caballero, 2017](#)). Esta función, Ripley la define como:

$$K(d) = \frac{E(M)}{\lambda} \quad (2.4.1)$$

Donde $E(M)$ es la esperanza del número de eventos adicionales a una distancia d de un evento elegido aleatoriamente y λ es la intensidad. (ecuación 2.4.1)

$$\hat{K}(d) = \frac{\sum \sum_{i \neq j} I_d(d_{ij})}{\hat{\lambda}_n} \quad (2.4.2)$$

En la ecuación 2.4.2, $\hat{\lambda}_n = \frac{n}{A}$, A es el área de la región R y d_{ij} es la distancia entre el i -ésimo y el j -ésimo evento.

$$\hat{K}(d) = \frac{\lambda \pi d^2}{\hat{\lambda}} = \pi d^2 \quad (2.4.3)$$

El patrón es completamente aleatorio si $K(d) = \pi d^2$ (valor teórico), si $K(d) < \pi d^2$ se dice que hay regularidad, es decir que hay muy pocas distancias pequeñas, y si $K(d) > \pi d^2$ el patrón es agregado, es decir que la mayoría de distancias son pequeñas. ([Rodríguez-cortés, 2018](#)). (ecuación 2.4.3)

El análisis consiste en contrastar la función-k teórica de un proceso Poisson homogéneo y la función-k empírica obtenida de las observaciones, de cada uno de los tres periodos temporales. Así, es posible determinar las dependencias espaciales que existen en diferentes etapas de la enfermedad. Cuando la función-k empírica esta por

debajo de la teórica, indica un patrón regular. Así mismo, cuando la función-k empírica esta por encima de la teórica, indica un patrón de datos agregados (ver figura 2.1) (Baddeley et al., 2015).

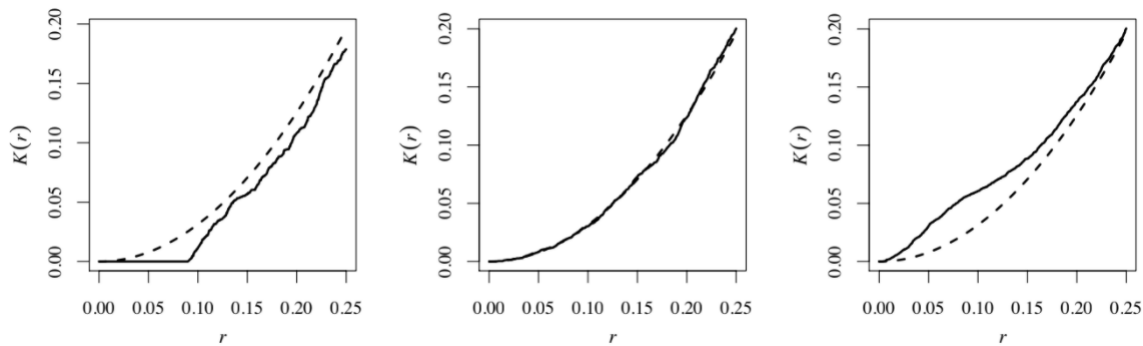


Figura 2.1 Baddeley et al. (2015). Función-k empírica (línea sólida) y función-k teórica para un proceso poisson (línea punteada)

La función L

La función L es otra de las funciones usadas para el estudio de dependencia en patrones puntuales, la cual esta basada en la función K(h), donde tiene la siguiente expresión:

$$L(d) = \sqrt{\frac{K(d)}{\pi}} \quad (2.4.4)$$

Donde la ecuación 2.4.4 es la expresión general, pero bajo un proceso aleatorio la expresión es:

$$L(d) = \sqrt{\frac{\pi d^2}{\pi}} = d \quad (2.4.5)$$

Se dice que si el patrón es aleatorio no debe haber desviación de la recta. (Giraldo, 2011). Para determinar una de las tres características con esta función, es igual al de la función-K de Ripley. Es decir, si la función-L empírica esta por debajo de la función-L teórica indica un patrón regular, si esta por encima indica agregación de datos y si comparten características similares las dos funciones, significa un patrón espacial completamente aleatorio entre los datos, es decir, cumple se distribuyen aleatoriamente al azar y uniformemente.

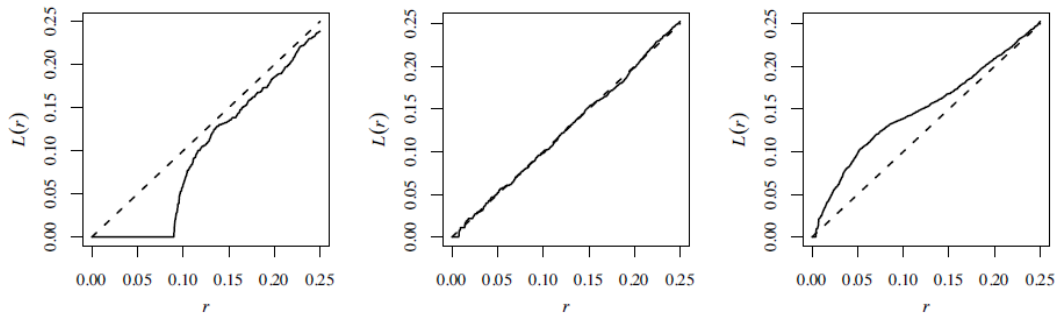


Figura 2.2 [Baddeley et al. \(2015\)](#). Función-L empírica (línea sólida) y función-L teórica para un proceso poisson (línea punteada)

La función $G(D)$

La función $G(D)$ usa como base la mínima distancia entre eventos (lo que se conoce como distancia al vecino más cercano). Donde:

d_i = Mínima distancia entre un evento y sus vecinos

D = Variable aleatoria que define la distancia al sitio del evento más cercano.

$G(D)$ = Función de distribución de D .

Estimación de $G(D)$

Función de distribución empírica:

$$G(\hat{D}) = \frac{Nd_i \leq d}{n} \quad (2.4.6)$$

Distribución teórica de $G(D)$

$$G(d) = P(D \leq d) = 1 - P(D > d) = 1 - P(\text{No hay eventos en } A = \pi d^2)$$

Si $N(A) \sim \text{Poisson}(\lambda \pi d^2)$; entonces

$$P(N(A) = 0) = e^{-\lambda \pi d^2}$$

Por lo tanto:

$$G(d) = 1 - e^{-\lambda \pi d^2} \quad (2.4.7)$$

Donde:

- Si $G(\hat{d})$ crece rápidamente en distancias cortas, los eventos son agregados.
- Si los eventos son regularmente espaciados, $G(\hat{d})$ crece lentamente hacia cierta distancia (espacio eventos) y después crece rápidamente ([Giraldo, 2011](#)).

El análisis, es igual que en los métodos anteriormente mencionados (ver figura 2.3).

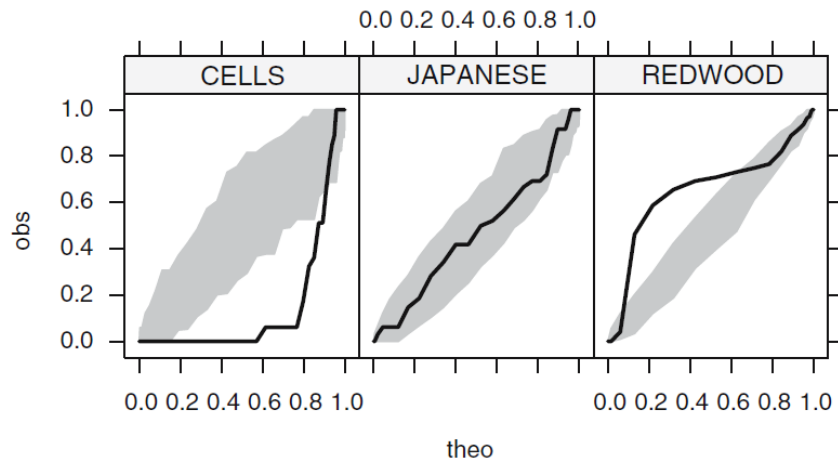


Figura 2.3 Bivand et al. (2013). Contraste entre la función G teórica y la función G empírica, para tres patrones de puntos

Correcciones de Borde

Las función K, G y L presentan 3 correcciones de borde que corrige los pequeños errores cometidos al contar cerca del borde de las ventanas de observación. El corrector de borde más usado propuesto por el profesor Ripley *isotrópico* (iso) en la imagen color negro el cual asume que el proceso es invariante frente a rotaciones, otra corrección es *bord* que es menos eficiente para datos menores a 3000 pero es la más rápida para calcular computacionalmente. La última corrección de borde es la *trans* o corrección por traslación. (Ohser (1983), Ripley (1988)).

2.4.4. Estimación Kernel de la función de intensidad

Es la estimación que se realiza para definir el esperado del número de puntos por unidad de área, donde esta estará delimitada por el perímetro urbano de la ciudad. Investigar la intensidad de un proceso puntual es una de las fases más importantes en el análisis de datos espaciales (Baddeley et al., 2015). Un patrón espacial completamente aleatorio tiene la siguiente función de intensidad:

$$\hat{\lambda} = \frac{n}{|A|} = \frac{\# \text{ de eventos}}{\text{area de } s} \quad (2.4.8)$$

Definiendo a x_i, y_i como las coordenadas de localización de un sitio $s_i(x_i, y_i)$, la estimación de intensidad sería:

$$\hat{\lambda}(s) = \frac{1}{|A|h_x h_y} \sum_{i=1}^n K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right) = \frac{1}{|A|} \sum_{i=1}^n K\left(\frac{s - s_i}{n}\right) \quad (2.4.9)$$

Donde h_x y h_y son los anchos de banda en las direcciones del sistema de coordenadas. Teniendo en cuenta el efecto de borde denominado como $\delta(S)$, el estimador quedaría de la siguiente manera:

$$\lambda(\hat{s}) = \frac{1}{\delta(s)} \sum_{i=1}^n K\left(\frac{s-s_i}{n}\right) \text{ Donde } \delta(s) = \int_A \frac{1}{h^2} K\left(\frac{s-u}{n}\right) du \quad (2.4.10)$$

Donde K representa la función kernel, en estadística no paramétrica, un kernel es una función de ponderación utilizada en las técnicas de estimación de densidad de variables aleatorias $f(x)$. (Guidoum, 2020), existen muchas funciones kernels que permiten la estimación no paramétrica de los datos, en este trabajo se empleará la función Epanechnikov.

Kernel Epanechnikov

$$K(x) = \frac{3}{4}(1-x^2); |x| \leq 1 \quad (2.4.11)$$

En el contexto de la estimación no paramétrica de la intensidad, es mucho mas relevante la elección del ancho de banda que la del núcleo (Kernel)(Guidoum, 2020). Para este estudio, con el fin de escoger el ancho de banda más óptimo se realizará el método de validación cruzada.

Selección del ancho de banda (h) de la densidad Kernel con validación Cruzada

Dentro del problema de elección del ancho de banda no existe un método correcto, entre los más usuales están, el de Berman y Diggle donde proponen un criterio basados en el Cuadrado Medio del Error (MSE *por sus siglas en ingles*) (Berman and Diggle, 1989), es decir, que en este caso el mejor ancho de banda será el que logre minimizar al máximo posible el MSE (Bivand et al., 2013). Otros autores proponen realizar un análisis exploratorio de diferentes métodos de escogencia de anchos de banda, también existe la metodología de selección mediante validación cruzada y hasta técnicas de selección automática. Para este estudio, se combinaron dos criterios. Mediante validación cruzada se selecciona el ancho de banda óptimo que minimice al máximo posible el MSE.

$$M(h) = \frac{MSE(h)}{\lambda^2} - g(0)$$

donde $MSE(h)$ es el error cuadrático medio en el ancho de banda h , $g(0)$ es la función de correlación de pares y λ es la intensidad media.

CAPÍTULO 3

Metodología

Es curioso que la vida, cuanto más vacía, más pesa.

León Daudí.
1905–1985

En esta sección se presenta el proceso de elaboración de la investigación, este componente metodológico está conformado por tres partes. La primera hace referencia a la descripción de la población de estudio, las características más relevantes, entre otros. La segunda parte, es el tratamiento inicial de los datos, como la depuración, organización y presentación de los datos. Por último, está la parte de análisis estadístico el cual se subdivide en: Análisis exploratorio, con el fin de observar comportamientos previos de los datos. Análisis temporal, para evaluar la evolución de los contagios por COVID-19 a través del tiempo. Análisis espacial, donde el objetivo es observar el comportamiento espacial del virus, zonas con más incidencia de casos, entre otros. Análisis de redes de contacto, donde se hará una representación visual y descriptiva de los contactos mencionados por cada uno de los individuos diagnosticados positivos COVID-19.

3.1. Población de Estudio

En esta investigación, la población de estudio serán los casos positivos de la enfermedad COVID-19 en el periodo de enero y septiembre del 2020. Los datos son suministrados por la Secretaria de Salud municipal de la ciudad de Santiago de Cali, donde hasta el 26 de septiembre del 2020 se tenían al rededor de 39643 casos positivos acumulados. Cali es la capital del departamento del Valle del Cauca, el cual, se subdivide en 15 corregimientos, 22 comunas y 249 barrios con 1.822.869 habitantes⁸.

⁸La información de número de viviendas, hogares y personas, a nivel municipal se actualizó el 12 de Noviembre de 2019, de acuerdo con la revisión de límites oficiales municipales dispuestos por el

Se debe mencionar en este apartado, el sesgo de información de casos que ocurrió al principio de la pandemia, donde aproximadamente la información tenía un rezago de 3 a 5 semanas, debido al desconocimiento del comportamiento del virus y los protocolos pertinentes para hacerle frente a la enfermedad. Los casos confirmados en los primeros días del mes de marzo, no eran en realidad los casos de ese mes, sino, de tres o cinco semanas atrás. Luego, como ya se conocía mejor la enfermedad y los protocolos, los resultados fueron más recientes, con una o dos semanas de atraso aproximadamente. La hoja electrónica en formato excel, cuenta con información pertinente de estos individuos, como lo es su dirección de vivienda, coordenadas de ubicación, lugar de trabajo, edad, entre otros.(Ver tabla de variables [B.1](#))

Para la construcción de la redes de contagio, la secretaría de salud publica municipal suministró otros registros, correspondiente a una encuesta que los individuos notificados positivos por COVID-19, realizaban voluntariamente. La encuesta consistía en obtener información de los círculos sociales, familiares y laborales con los que tuvo contacto el individuo antes de ser notificado positivo. Los registro que se lograron recolectar entre los periodos de marzo y octubre del 2020, fueron de 1588 casos de contactos. (Ver tabla de variables [B.2](#)).

3.2. Depuración de los datos

Antes de procesar la información, se reestructuró la base de datos con el fin de dejarla técnicamente correcta para el análisis, en este paso se hizo una revisión y clasificación de cada una de las variables contenidas en la base de datos.

Inicialmente, la estructura de los datos proporcionados por la secretaria de salud de Cali no era la ideal para lograr cumplir los objetivos propuestos en este trabajo. Por lo tanto, se realizó un análisis exploratorio previo con el fin de evaluar las variables de interés y como estaban conformadas cada una de ellas. Se encontró, con variables cualitativas cuyas categorías estaban repetidas, es decir, habían registros que presentaban la misma categoría, pero por temas de escritura se aparentaban ser diferentes. El tratamiento de datos perdidos, fue otro problema critico en la depuración de la base de datos suministrada. Algunas variables presentaban mas del 50 % de datos faltantes, por lo cual, fue necesario descartarlas en la depuración y dejar estrictamente las variables relevantes para este estudio.

Mediante la metodología CRISP-DM (ver figura [A.2](#)), se logró hacer una limpieza eficiente de los datos para lograr enfatizar en los objetivos propuestos ([Britos, 2005](#); [De Jonge and Van Der Loo, 2013](#)). Esta metodología consiste en evaluar seis pasos:

1. Comprender el problema.

2. Entender los datos.
3. Preparación de los datos.
4. Modelado.
5. Evaluación.
6. Despliegue.

Para efectos del estudio, el paso cuatro es igual al análisis espacio-temporal y la construcción de redes, donde luego se evalúa la metodología utilizada en el análisis para futuros modelados estadísticos y despliegues.

Por otra parte, los datos faltantes no fue posible eliminarlos o imputarlos, debido a las características de los datos, una eliminación de un registro podría ser un individuo que no quiso dar información al respecto y una imputación implicaría estimar una condición que puede no ser la verdadera.

3.3. Análisis Estadístico

3.3.1. Análisis Espacio-Temporal

Previo al análisis espacial, se realizaron dos estudios. El primero, un análisis exploratorio de los datos con el fin de conocer el comportamiento de la enfermedad y la forma como se estaba propagando, logrando identificar algunos sectores críticos de la ciudad. El segundo, un estudio descriptivo temporal, con el objetivo de encontrar tendencias de la enfermedad en diferentes periodos del tiempo estudiado. De este modo, mediante suavizadores Spline se trató de ajustar una curva donde tuviera un equilibrio entre sesgo y variabilidad.

Para esto se ajustó a la nube de puntos un suavizado tipo spline con un parámetro de suavizado $\lambda=0.4$, con el fin de no sub-ajustar o sobre-ajustar la curva de tendencia ([Durbán Reguera, 2009](#)). Los tres periodos se dividen de la siguiente forma:

Periodo 1: Corresponde a los primeros casos registrados de la pandemia, es decir, los meses entre marzo y mayo.

Periodo 2: Este periodo son los meses intermedios y donde ocurrió el primer pico de la enfermedad en la ciudad de Cali, junio y julio.

Periodo 3: Son los dos últimos meses del intervalo de tiempo estudiado en esta investigación, estos son, agosto y septiembre.

Luego, con los datos proporcionados por la secretaría de salud municipal, el cual, contenían las direcciones de los casi 40000 casos positivos por COVID-19. Se realiza la estandarización de estas direcciones y posteriormente una geocodificación a un sistema de coordenadas de longitud y latitud (Elías-Cuertas et al., 2020). Donde, con la ayuda del software estadístico R y algunas librerías de visualización y análisis de datos espaciales (ver tabla B.3), se construye el polígono de la ciudad de Santiago de Cali dividido por barrios (Wikle et al., 2019).

Inmediatamente, se realiza un análisis de patrones puntuales en el espacio y tiempo. Esto con el fin de evaluar el comportamiento del virus según las semanas epidemiológicas comprendidas entre enero y septiembre del 2020.

Los periodos se dividieron en 16, el cual se conforman de la siguiente manera:

Nº PERIODOS	SEMANAS	REGISTRO DE CASOS POSITIVOS COVID-19
1	3 hasta la 22	2825
2	23 y 24	2196
3	25 y 26	2808
4	27	2019
5	28	2981
6	29	3367
7	30	3640
8	31	3360
9	32	2802
10	33	2811
11	34	2098
12	35	1839
13	36	1656
14	37	1939
15	38	1915
16	39	1310

Tabla 3.1 Cantidad de casos positivos por Período.

Principalmente, el análisis se centra en un conjunto de puntos finitos (individuos positivos por COVID-19) en el espacio y tiempo. Además, dentro del análisis espacial se tuvieron en cuenta las pruebas de aleatoriedad espacial y la de estimación de la intensidad, con el fin de observar las correlaciones existentes entre los patrones puntuales.

Evaluación de aleatoriedad espacial

Teniendo en cuenta los datos georeferenciados (coordenadas de Latitud y Longitud) dentro del perímetro urbano de Santiago de Cali se realiza en la evaluación de la aleatoriedad espacial donde las hipótesis estadísticas son:

H_0 : El patrón de puntos es aleatorio vs H_a : El patrón de puntos no es aleatorio

Test de completa aleatoriedad espacial - Basada en Cuadrantes

Antes de comenzar el estudio del patrón de puntos, es necesario realizar el test de completa aleatoriedad espacial para caracterizar dicho patrón de punto, el rechazo de la hipótesis nula conlleva a un patrón agregado ó regular. En principio, con los datos georeferenciados en el perímetro de Santiago de Cali se realiza el test por el método de cuadrantes y distancias (Funciones K,L y G).

La prueba Chi-cuadrado tiene como requisito de que todas las casillas observadas tengan una frecuencia mayor a 5 eventos, en algunos casos puede que este problema conlleve a sesgo en la estimación por el no cumplimiento de este requisito, es por eso que se decidió realizar esta misma prueba con el criterio de Monte carlo mediante 1000 simulaciones. Según [Hatfield \(2018\)](#), la prueba basada en cuadrantes presenta una limitación y es debido a que la división en cuadrantes puede que no se cumpla de manera homogénea en el espacio geográfico, en nuestro caso el perímetro de Santiago de Cali no es un espacio geográfico completamente cuadrado, por lo tanto, para esta limitación, se realiza también el análisis de la aleatoriedad espacial por el método basado en distancias.

Test de completa aleatoriedad espacial - Basada en Distancias

Para evaluar las hipótesis dentro de cada Periodo basada en el método de distancias, se empleó las funciones K de Ripley, L y G, donde para cada una, si la curva representada por el patrón de puntos está por encima de la curva del proceso Poisson (completa aleatoriedad espacial) se considera que el proceso es agregado, si ocurre lo contrario (por debajo) se considera que el patrón de puntos es regular.

Dentro de las curvas dadas por las funciones G, K y L se encuentran las correcciones de borde la cual corrige los errores cometidos al contar cerca del borde de las ventanas de observación, es decir puntos que puedan estar por fuera de la región caracterizada. El corrector de borde que se utilizó para la determinación de la naturaleza del patrón de puntos (Patrón Agregado o Regular) mediante análisis gráfico, si la función expresada por la corrección de borde (*bord(r)*) estaba por encima de la curva del proceso poisson (*pois(r)*) el patrón es catalogado como agregado para ese periodo de tiempo, si ocurre lo contrario (por debajo de la curva (*pois(r)*) es catalogado como un patrón regular.

Estimación de la intensidad.

Después de que se define si el proceso es aleatorio, agregado o regular (rechazo de la hipótesis nula), se realiza la estimación de la intensidad del patrón de puntos, la cual es relevante para determinar la distribución espacial y temporal del periodo.

Para realizar dicha estimación se utilizó el método no paramétrico Kernel cuya función de núcleo fue la Epanechnikov y para la elección del ancho de banda (λ) se utilizó el criterio de validación cruzada (con la función del *bw.diggle* y *bw.ppl* de la librería *spatsat* en el software R con la función *density*) el cual se basa en escoger el mínimo Cuadrado Medio del Error (MSE). [Baddeley et al. \(2008\)](#) sugiere el uso del algoritmo *bw.ppl()* porque, en su experiencia, tiende a producir los valores más apropiados cuando el patrón se caracteriza por ser agregado, por lo tanto los anchos de banda para este trabajo son calculados con dicho criterio.

3.3.2. Análisis de redes de contactos

Luego de hacer el análisis descriptivo, temporal y espacial de los datos, se construye la red de contactos de cada uno de los individuos de estudio. Estas redes serán estáticas y no dirigidas, el objetivo de este análisis es evaluar como son las evoluciones de los contagios a través de la red de contactos de los individuos, así la secretaría de salud municipal tendrá un mejor conocimiento de la distribución del virus.

Es importante aclarar que, por cuestiones de tratamiento de datos los individuos fueron codificados de tal manera que fuera evidente el nombre real de cada uno de ellos. De esta forma, se respeta el derecho a la privacidad de los individuos mostrados en este análisis de redes.

El contexto teórico en la construcción de las redes tiene diferentes etapas, la primera y fundamental es la depuración de los datos, luego viene la construcción de la matriz de adyacencia, descripción de vértices y bordes, visualización de la red y análisis. Esta construcción estará fundamentada de acuerdo a diferentes autores ([Brandes and Erlebach, 2005](#); [Kolaczyk, 2009](#); [Kolaczyk and Csárdi, 2015](#); [Luke, 2015](#)).

De acuerdo a lo anterior, los grafos contruidos para este estudio, tienen el nombre de arboles, el cual se caracterizan por tener un único camino simple entre dos vértices y no existe circuito. Adicional, el tipo de árbol construido en cada uno de los grupos de contacto se le denomina árbol estrella, porque existe un único vértice que se conecta con otros vértices. Ahora, la construcción teórica de estos arboles dado su código de Prüfer⁹, es ([Aguila, 2013](#)):

- Se inicia con un árbol vacío¹⁰T y se obtiene dos conjuntos de datos L_1 y L_2 .
- $L_1=[1,2,...,n]$ ($n \in \mathbb{Z}$) y $L_2 = [p_1, p_2, ..., p_{n-2}]$.
- Sea $w_1=\min[i:i \in L_1 \wedge i \in L_2]$.
- Se agrega la arista o borde (w_1, p_1) al árbol T y se actualiza:

⁹**Código de Prüfer:** Es una manera de codificar la estructura de un árbol con n vértices ([Aguila, 2013](#))

- $V(T)=V(T) \cup (w_1, p_1)$
 - $E(T)=E(T) \cup (w_1, p_1)$
 - $L_1=L_1/w_1$
 - $L_2=L_2/p_1$
- El proceso de forma general es para cada $w_i, i=1,2,\dots,n-2$:
- $w_i=\min[i:i \in L_1 \wedge i \in L_2]$
 - $V(T)=V(T) \cup (w_i, p_i)$
 - $E(T)=E(T) \cup (w_i, p_i)$
 - $L_1=L_1/w_i$
 - $L_2=L_2/p_i$

El total de los registros trabajados en este tópico fueron de 1849, de los cuales, se logró construir 1587 enlaces con 262 individuos diagnosticados con COVID-19 encuestados en el periodo de marzo-octubre. La construcción de los arboles, se realizó con la ayuda del software estadístico R y el proceso se divide en 4 fases:

Fase 1: Las librerías utilizadas para la construcción de los grafos (ver tabla B.3), funcionan con dos hojas electrónicas de datos, pero al final se unifican para crear el grafo. La primera hoja, corresponde a la información de los enlaces o links, pesos y tipos de enlaces. La segunda hoja, contiene todos los vértices sin enlace, una ID previamente creada para cada individuo, identificación del genero (cualitativo y binario), numero de comuna, entre otras variables relevantes para el análisis.

Fase 2: Una vez teniendo los datos debidamente organizados, se construyen los grafos.

Fase 3: Construcción de matriz de adyacencia, el cual, por las características de los grafos no dirigidos, sera una matriz cuadrada simétrica donde tendrá los valores de los pesos en los enlaces existentes entre cada unos de los individuos. La dimensión de esta matriz es de 1849 x 1849.

Fase 4: Finalmente, se agrupan los arboles de estrella contruidos. Esta unión entre arboles, es denominada bosque.

Construir primero los grafos en vez de la matriz técnicamente es posibles, en cualquiera de los dos caminos se llega al mismo resultado. Pero, el coste computacional es más alto al construir primero la matriz y luego los grafos, además, por la cantidad de enlaces que se tenían este coste era un factor a tener en cuenta en el tratamiento de los datos.

¹⁰**Árbol Vacío:** $V(T)=E(T)=\emptyset$

CAPÍTULO 4

Resultados y discusión

D^{IME} y lo olvido,
enséñame y lo recuerdo, involúcrame y lo aprendo.

Benjamin Franklin.
1706–1790

4.1. Análisis estadístico descriptivo

En esta sección, se presentan algunos estadísticos descriptivos del fenómeno estudiado. Estos datos van desde enero 18 y septiembre 26 del 2020, es evidente que la fecha trabajada en este estudio no concuerda con el primer reporte de caso positivo por COVID-19 realizado en la ciudad de Cali, el motivo es -según la secretaria municipal- porque esta fecha se refiere a los registros que inicialmente tuvieron sospechas y luego si fueron confirmados. Lo anterior pudo haber sido porque, la enfermedad se dispersó por el mundo muy rápido, tanto fue su rapidez que los países al principio no sabían como manejar esta situación.

La nacionalidad que destaca en los reportes de la ciudad de Cali es colombiana, con un 67.8% sobre el total de los datos. Los datos perdidos (N/D), fueron del 31.61% para esta variable. Es probable que el alto índice de no respuesta en esta variable se debe a una recolección de datos mal estructurada o simplemente los individuos se negaron a brindar información.

Se evidencia en la tabla 4.1, un total de casos positivos acumulados de 39643 donde, el 50.23% son hombres y el 49.1% mujeres. Este primer resumen es interesante porque, refleja que la enfermedad COVID-19 afecta en igual proporciones a hombre y mujeres.

	Sexo	Número de casos	Porcentaje
1	M	19916	50.23 %
2	F	19462	49.1 %
3	N/D	265	0.67 %

Tabla 4.1 Frecuencia de casos positivos entre enero y septiembre del 2020, discriminados por sexo.

Se registraron 231 individuos de 18 nacionalidades diferentes a Colombiana, equivalentes al 0.583 % del total de los datos. Se observa en la ciudad de Cali una mayor proporción de casos con nacionalidad venezolana y congolés, correspondientes al 73.1 % y 6.5 % (respectivamente) del total de extranjeros registrados (ver figura 4.1 y tabla 4.2).

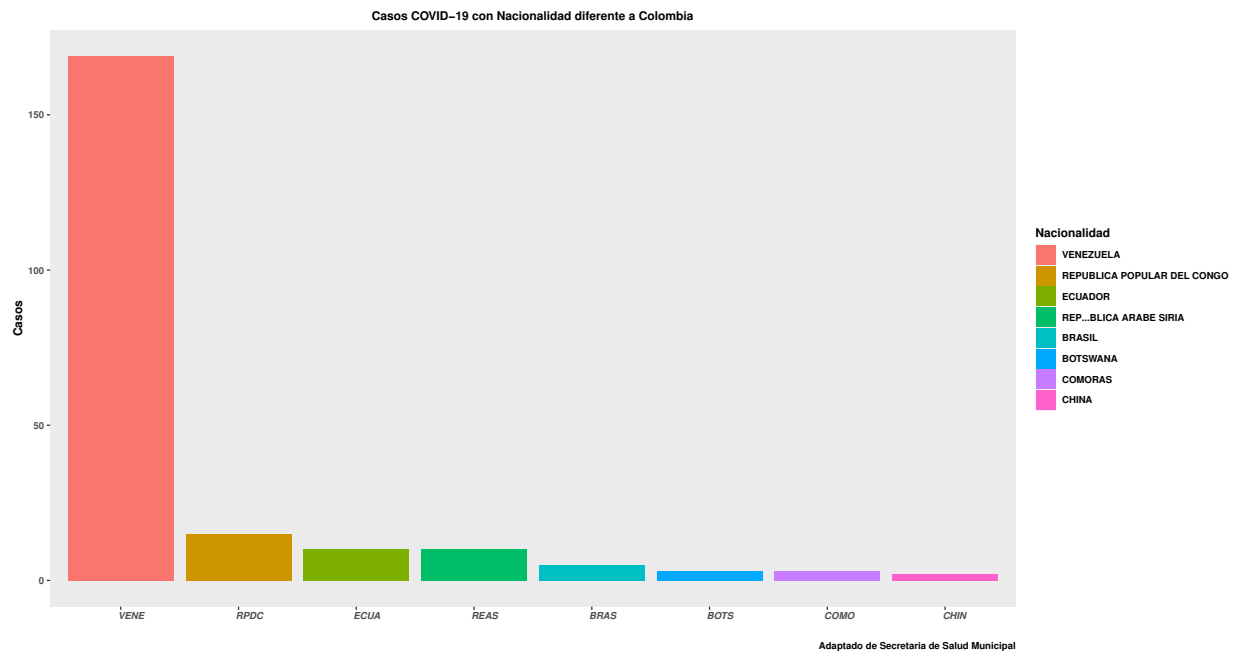


Figura 4.1 Frecuencia de casos positivos con nacionalidades diferentes a Colombia

	Nacionalidad	Frecuencia Absoluta	Frecuencia Relativa
1	VENEZUELA	169	73.1 %
2	REPÚBLICA POPULAR DEL CONGO	15	6.5 %
3	ECUADOR	10	4.33 %
4	REPÚBLICA ÁRABE SIRIA	10	4.33 %
5	BRASIL	5	2.16 %
6	BOTSWANA	3	1.3 %
7	COMORAS	3	1.3 %
8	CHINA	2	0.87 %
9	GAMBIA	2	0.87 %
10	ITALIA	2	0.87 %
11	MÉXICO	2	0.87 %
12	PERÚ	2	0.87 %
13	ANTILLAS NEERLANDESAS	1	0.433 %
14	ARGENTINA	1	0.433 %
15	BOSNIA	1	0.433 %
16	CUBA	1	0.433 %
17	ESTADOS UNIDOS DE AMÉRICA	1	0.433 %
18	SUDÁFRICA	1	0.433 %

Tabla 4.2 Frecuencia de casos positivos COVID-19 según la Nacionalidad

Como el objetivo principal de este proyecto, es analizar a nivel espacio temporal los contagios, se hace indispensable observar patrones de la enfermedad discriminados por barrios y estratos. De esta forma, se logra tener un panorama inicial descriptivo de las mayores incidencias del COVID-19.

En este orden de ideas, se observa en la gráfica 4.2 dos representaciones de los casos positivos acumulados en el periodo de tiempo estudiado, donde, a la izquierda se aprecia el ranking de los barrios de Cali con mayor incidencia acumulada y a la derecha la frecuencia de los casos por estrato. Se hace evidente que, los estratos dos y tres fueron los mayormente afectados por esta enfermedad (ver tablas 4.3 y 4.4). El porcentaje de N/D fue de 32.97 % para la variable *Barrios* y 39.2 % para la variable *Estrato*.

Con este primer acercamiento, el panorama del comportamiento de la enfermedad se empieza a ver un poco mejor y se evidencia las zonas donde toca hacer hincapié por parte de las entidades sanitarias. Sin embargo, en el análisis espacio temporal se podrá apreciar de una forma más clara como se comportó esta enfermedad entre los meses de enero y septiembre del 2020.

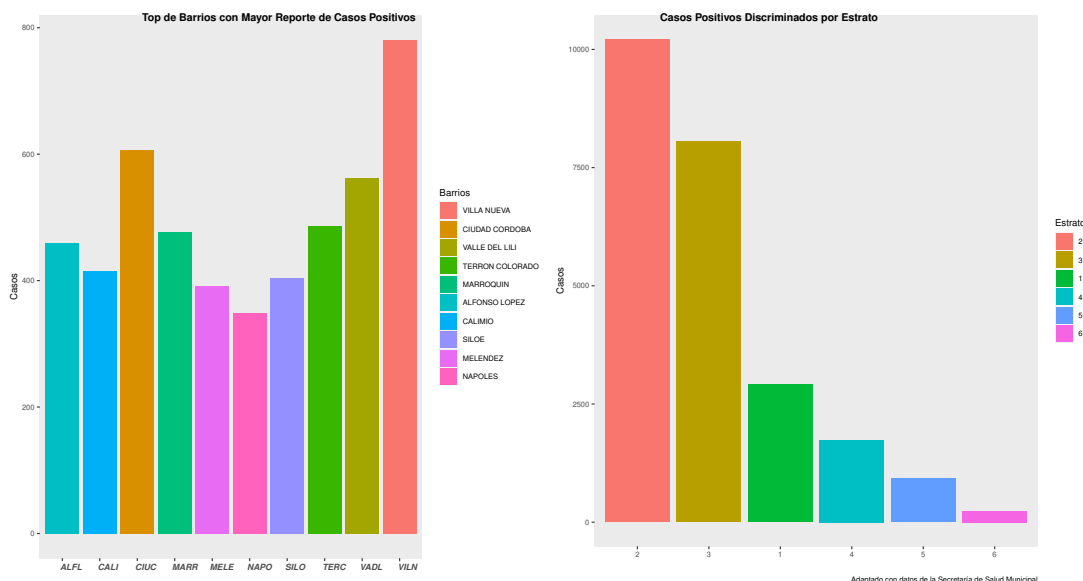


Figura 4.2 Casos positivos de COVID-19 por barrios y estratos

Barrios	Frecuencia Absoluta	Frecuencia Relativa
1 VILLA NUEVA	781	2.94 %
2 CIUDAD CORDOBA	607	2.29 %
3 VALLE DEL LILI	563	2.121 %
4 TERRÓN COLORADO	486	1.83 %
5 MARROQUÍN	477	1.8 %
6 ALFONSO LOPEZ	460	1.73 %
7 CALIMIO	415	1.563 %
8 SILOE	404	1.52 %
9 MELENDEZ	392	1.48 %
10 NAPOLES	349	1.32 %

Tabla 4.3 Ranking de casos positivos COVID-19 por barrios

Estrato	Frecuencia Absoluta	Frecuencia Relativa
1	2915	12.10 %
2	10216	42.42 %
3	8054	33.44 %
4	1739	7.22 %
5	925	3.84 %
6	234	0.97 %

Tabla 4.4 Casos COVID-19 en Santiago de Cali por estrato

El análisis con los diagramas de cajas para la variable edad (ver figura 4.3), se centra en encontrar diferencias entre edades y semanas epidemiológicas discriminando los individuos por los estratos socio-económicos. Se encontró que, las medianas de edades para cada uno de los estratos son muy similares, por lo cual es posible intuir de acuerdo al gráfico que la mayor densidad de individuos contagiados están entre los 39 y 41 años (ver tabla 4.5).

Ahora, con las semanas epidemiológicas se encuentran leves diferencias en la longitud de los bigotes de cada uno de los diagramas por estrato, esto evidencia que los primeros casos registrados se encontraban muy posiblemente en los estratos 1 y 2. Diferencias grandes entre la medianas no se encontraron, estas oscilan entre la semana 27 y 29, por lo cual, la mitad de los registro casi coincide con el primer reporte de pico de la enfermedad (ver tabla 4.6).

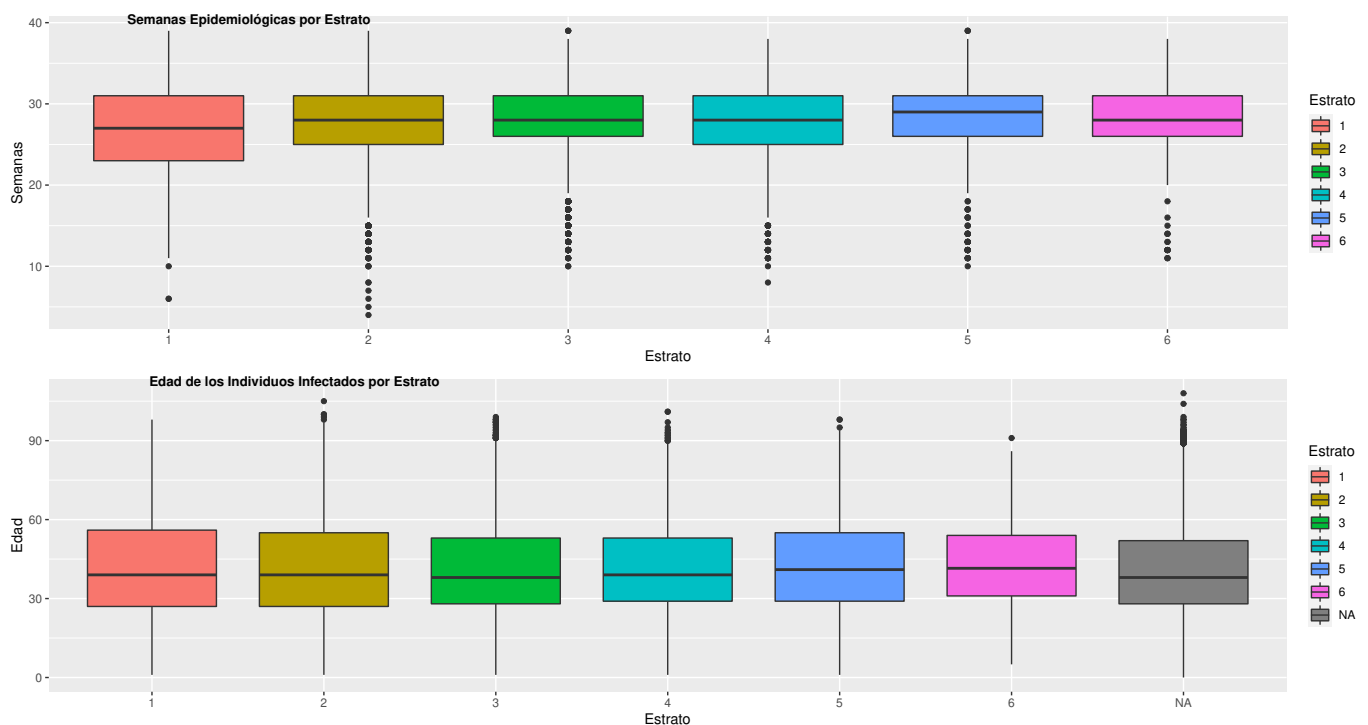


Figura 4.3 Visualización de la edad y las semanas epidemiológicas por estrato

Estrato	Edad media	D.E.	CV	Mediana
1	41.78	19.61	46.92	39.00
2	41.68	18.85	45.22	39.00
3	40.89	18.06	44.16	38.00
4	41.82	17.84	42.66	39.00
5	42.66	19.03	44.60	41.00
6	43.75	17.02	38.91	41.50
N/D	40.17	17.77	44.24	38.00

Tabla 4.5 Estadísticas de la edad de los casos positivos por COVID-19, discriminado por estrato

Estrato	Semana media	D.E.	CV	Mediana
1	26.81	5.33	19.89	27.00
2	27.52	4.87	17.68	28.00
3	27.90	5.23	18.76	28.00
4	27.66	5.55	20.05	28.00
5	27.64	5.99	21.66	29.00
6	27.10	6.37	23.52	28.00

Tabla 4.6 Estadísticas de las semanas epidemiológicas discriminado por estrato

Finalmente, se presentan una visión global del comportamiento de la enfermedad en la ciudad de Cali. En la figura 4.4, se muestra en rojo las incidencias diarias que ocurrieron en el periodo de tiempo estudiado y la línea azul los casos positivos acumulados. De acuerdo a esta información, el primer pico ocurrió el 27 de julio del 2020 y luego se observa un leve descenso de la enfermedad.

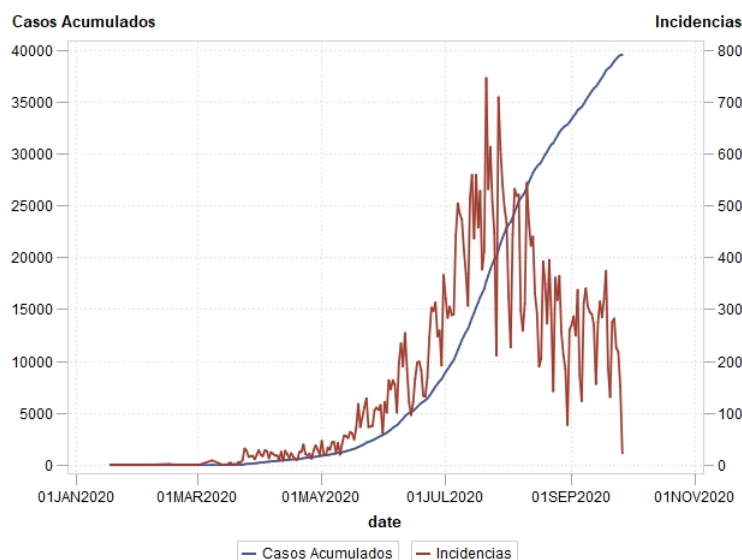


Figura 4.4 Visualización global del comportamiento de la enfermedad entre Enero 18 y Septiembre 26 del 2020

4.2. Análisis estadístico espacio-temporal

Es importante realizar un análisis previo temporal, donde se conozca de antemano tendencias y comportamientos de la enfermedad COVID-19 a través del tiempo. Así,

en la figura 4.5 se muestra una aplicación del método de la suavización Spline como herramienta para visualizar descriptivamente el comportamiento de la pandemia en Cali.

Este primer análisis descriptivo temporal, se dividió en tres periodos con el fin de representar diferentes etapas de la enfermedad a través del tiempo. De esta forma, en el periodo 1 (*Inicio de la pandemia*) se evidencia que los casos positivos entre los meses enero y principios de marzo no superaban más de dos incidencias diarias. También, en este primer periodo por políticas de salud pública, para todo el territorio nacional se decreta la cuarentena obligatoria (línea vertical negra), esta medida se toma al ver un repunte de la enfermedad en las últimas dos semanas de marzo. En este periodo, es observable que después de la segunda semana de marzo la tendencia de la enfermedad fue creciente, sin embargo, en algunos casos hubieron periodos descendentes. Pero, hay que tener en cuenta que hubo un periodo aproximadamente de tres semanas, entre marzo y abril, donde la tendencia decreciente fue a causa de daños en equipos indispensables para la detección de la enfermedad.

El segundo periodo (*Primer pico*), tiene una particularidad creciente. Aunque aun estaba vigente la cuarentena obligatoria en estas fechas, se evidencia que el virus alcanzó su máximo de incidencias diarias superior a los 750 casos positivos. Adicionalmente, se observa que la curva presenta mayor variabilidad que en el primer periodo.

El 1 de septiembre del 2020 se levanta la cuarentena obligatoria en toda Colombia, esto se representa en el periodo tres (*Leve descenso*) con la línea negra vertical. Aquí se muestra que, la enfermedad empezó a tener una tendencia decreciente, por lo cual, motivó a los entes gubernamentales a implementar medidas menos drásticas. Sin embargo, aunque es evidente la tendencia decreciente comparándolo con los otros dos periodos, se observa que esta tendencia no es muy marcada y presenta una alta variabilidad.

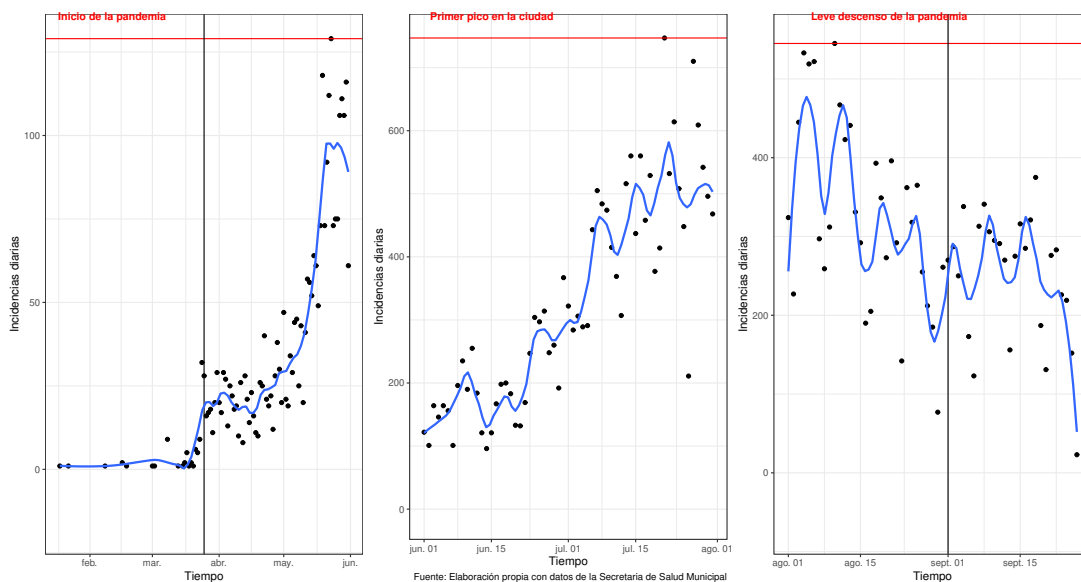


Figura 4.5 Visualización de la tendencia del fenómeno mediante suavización Spline

4.2.1. Análisis Patrones Puntuales

Para este análisis se propuso dividir en periodos de acuerdo a semanas epidemiológicas, dado que la función de correlación espacial (K y L) funcionan computacionalmente bien si el número de datos es menor a 3000. De manera que se dividió el análisis para que cada periodo cumpla esta condición. Sin embargo, es necesario aclarar que sólo en 3 Periodos (6,7 y 8) hubo más de 3000 casos en la semana epidemiológica.

La tabla 4.7 presenta los test de completa aleatoriedad espacial basada en el método de Cuadrantes y con los criterios χ^2 y Montecarlo, se tiene que para los 16 periodos se rechaza la hipótesis nula, es decir que existe suficiente evidencia estadística para decir que la distribución espacial del patrón de puntos para cada uno de los periodos no es aleatorio.

	Método	Prueba	Resultado	Valor-p
Periodo 1	En cuadrantes	Test χ^2	Agregado	0,000009
		Test de Montecarlo	Agregado	0,000050
Periodo 2	En cuadrantes	Test χ^2	Agregado	0,000002
		Test de Montecarlo	Agregado	0,000400
Periodo 3	En cuadrantes	Test χ^2	Agregado	0,000010
		Test de Montecarlo	Agregado	0,000041
Periodo 4	En cuadrantes	Test χ^2	Agregado	0,000007
		Test de Montecarlo	Agregado	0,000046
Periodo 5	En cuadrantes	Test χ^2	Agregado	0,000001
		Test de Montecarlo	Agregado	0,000012
Periodo 6	En cuadrantes	Test χ^2	Agregado	0,000001
		Test de Montecarlo	Agregado	0,000055
Periodo 7	En cuadrantes	Test χ^2	Agregado	0,000002
		Test de Montecarlo	Agregado	0,000008
Periodo 8	En cuadrantes	Test χ^2	Agregado	0,000009
		Test de Montecarlo	Agregado	0,000400
Periodo 9	En cuadrantes	Test χ^2	Agregado	0,000001
		Test de Montecarlo	Agregado	0,000049
Periodo 10	En cuadrantes	Test χ^2	Agregado	0,000002
		Test de Montecarlo	Agregado	0,000040
Periodo 11	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000440
Periodo 12	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000087
Periodo 13	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000046
Periodo 14	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000070
Periodo 15	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000090
Periodo 16	En cuadrantes	Test χ^2	Agregado	0,000000
		Test de Montecarlo	Agregado	0,000065

Tabla 4.7 Prueba de Aleatoriedad Espacial por el Método de Cuadrantes con los criterios Test χ^2 y Montecarlo para los 16 Periodos

En la tabla 4.8 se presentan los anchos de banda calculados para cada periodo el cual permite estimar las funciones K, G, L y la función de intensidad, esto se realiza mediante el criterio de validación cruzada.

Periodo	Bandwidth
Periodo 1	9.0250
Periodo 2	26.6443
Periodo 3	37.8974
Periodo 4	19.1147
Periodo 5	39.5947
Periodo 6	44.1458
Periodo 7	64.6259
Periodo 8	81.6926
Periodo 9	82.2647
Periodo 10	56.7060
Periodo 11	51.1349
Periodo 12	53.7565
Periodo 13	30.4925
Periodo 14	13.1982
Periodo 15	39.8223
Periodo 16	49.8347

Tabla 4.8 Anchos de banda para la estimación de intensidad y función K,G y L por medio de Validación Cruzada

Periodo 1

El periodo 1 está comprendido entre la fecha 18/01/2020 y 30/05/2020, (Semana 3 a la 20) con un acumulado de casos de 2825.

Patrón de Puntos Periodo 1



Figura 4.6 Patrón de Puntos - Periodo 1

En la figura 4.6 se presenta el patrón puntual del periodo 1, se observa una concentración (varios puntos de casos positivos) en el Centro y hacia el Oriente de la Ciudad. Además en este periodo existen varios puntos por fuera de la perímetro de Santiago de Cali.

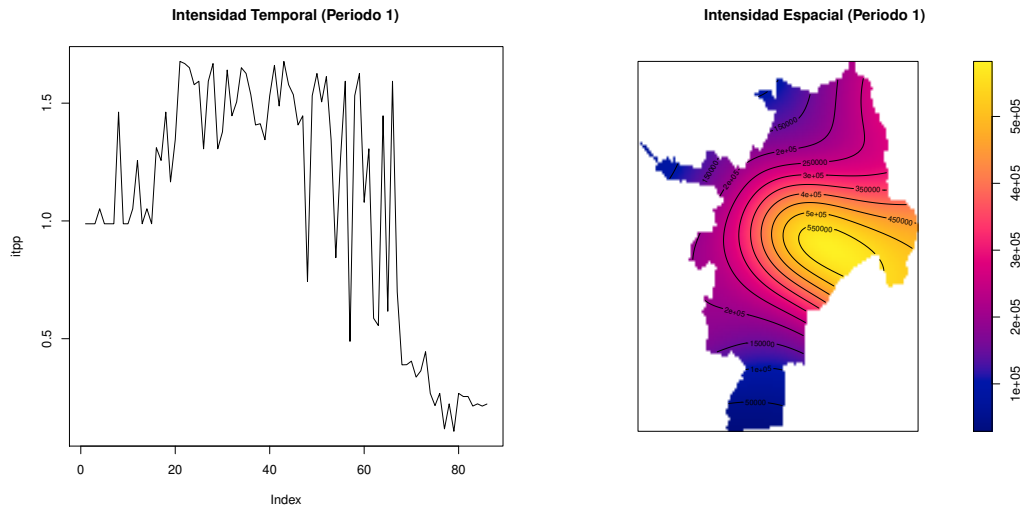


Figura 4.7 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 1

Para observar el comportamiento espacio-temporal de los periodos se tiene la figura 4.7 que representa a la izquierda la intensidad temporal y al lado derecho la espacial, se observa descriptivamente que el patrón es agregado con concentración en el Oriente y Distrito de Aguablanca de Cali para este Periodo 1.

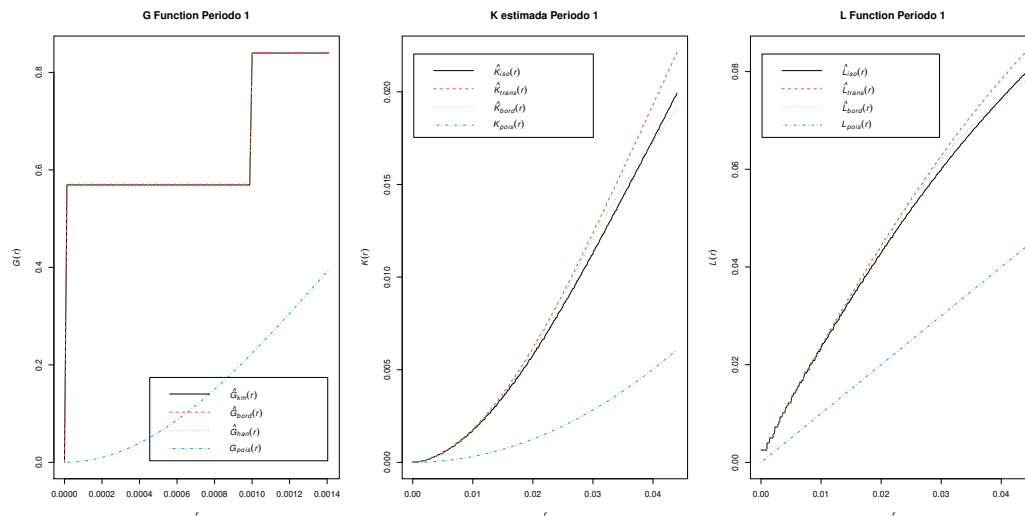


Figura 4.8 Función G, K y L - Periodo 1

En la figura 4.8 se presenta las funciones G, K y L respectivamente, para analizar la correlación espacial del periodo.

Para cada una de estas funciones se presentan con la correcciones de borde. En la función G se tiene que el proceso es totalmente agregado ya que las 3 correcciones de borde están por encima de la línea azul clara (proceso es aleatorio ó Proceso Poisson). Se tiene en este Periodo que para la función K y L, cada una de las correcciones de borde está por encima del proceso homogéneo de Poisson, observando entonces

un proceso agregado. Para la función L se tienen las mismas correcciones de borde, presentando los mismos resultados.(Ver Figura 4.8).

Periodo 2

En el periodo 2 se tienen las semanas epidemiológicas 23 y 24 las cuales están comprendidas entre la fecha 31/05/2020 y 13/06/2020, con un acumulado de casos de 2196.

En la figura 4.9 se presenta el patrón de puntos para este periodo, donde se observa una concentración de puntos en el Oriente de la Ciudad. Para observar mejor esta concentración se tiene el patrón de intensidad temporal y espacial en la figura 4.10, por el lado izquierdo el patrón presenta varios saltos ó picos a través de estas dos semanas, en el lado derecho se observa la intensidad espacial en la que al parecer existe un patrón agregado, concentrado la mayor cantidad de casos (Color amarillo) en el Centro la Ciudad.

Patrón de Puntos Periodo 2



Figura 4.9 Patrón de Puntos - Periodo 2

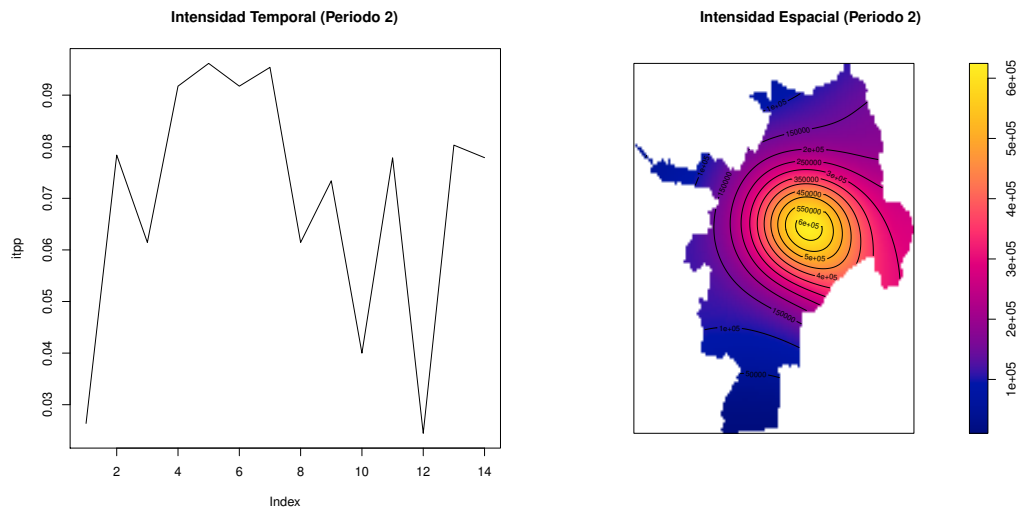


Figura 4.10 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 2

En la figura 4.11 se presenta las funciones G,K,L respectivamente. Donde la función G al tener las curvas escalonadas por encima de la patrón aleatorio (curva azul clara) nos indica que el patrón es agregado. En mayor medida esto se nota para las funciones K y L en las que se destaca que el patrón es agregado al estar por encima de la barra azul que representa un patrón completamente aleatorio o proceso Poisson.

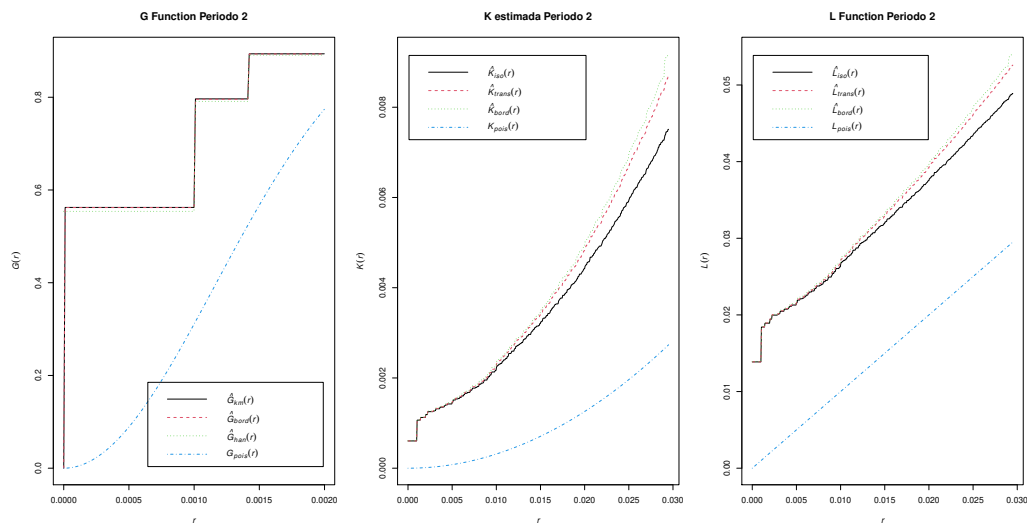


Figura 4.11 Función G, K y L - Periodo 2

Periodo 3

En el periodo 3 se tienen las semanas epidemiológicas 25 y 26 las cuales están comprendidas entre la fecha 14/06/2020 y 27/06/2020, con un acumulado de casos de 2808. Para este periodo se observa en el patrón puntual una presencia de puntos más expandida y no tan concentrada en una sola zona de Cali, a pesar de que en este periodo se presentan casos en varias zonas de la Ciudad la concentración de puntos

sigue siendo en la zona Oriente de la Ciudad. (Ver figura 4.12).

Patrón de Puntos Periodo 3

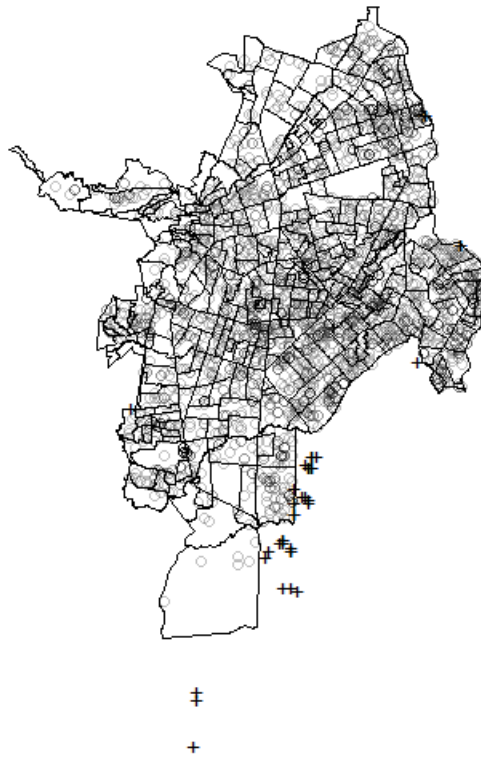


Figura 4.12 Patrón de Puntos - Periodo 3

En la intensidad espacio temporal se tiene que la intensidad temporal presenta en su mayoría un comportamiento regular ya que solo presenta una caída.^{al} final de periodo, así mismo para la intensidad espacial se observa que ésta intensidad está repartida por todo el espacio geográfico de Santiago de Cali pero enfocándose aún en el Oriente de Cali, aunque cabe notar la diferencia con el anterior Periodo es que en este la concentración no es tan marcada y parece un poco más aleatoria. (Ver figura 4.13) Para clasificar si el patrón es regular, agregado o aleatorio se presenta entonces la figura 4.14 en observa para cada una de las funciones G,K y L que las curvas están por encima de la curva de aleatoriedad espacial, cabe resaltar que a pesar de que están por encima se nota una diferencia con el anterior Periodo 2 que evidentemente era un proceso agregado, en este tenemos la diferencia entre las curvas con el de la aleatoriedad espacial es un poco menor al anterior periodo.

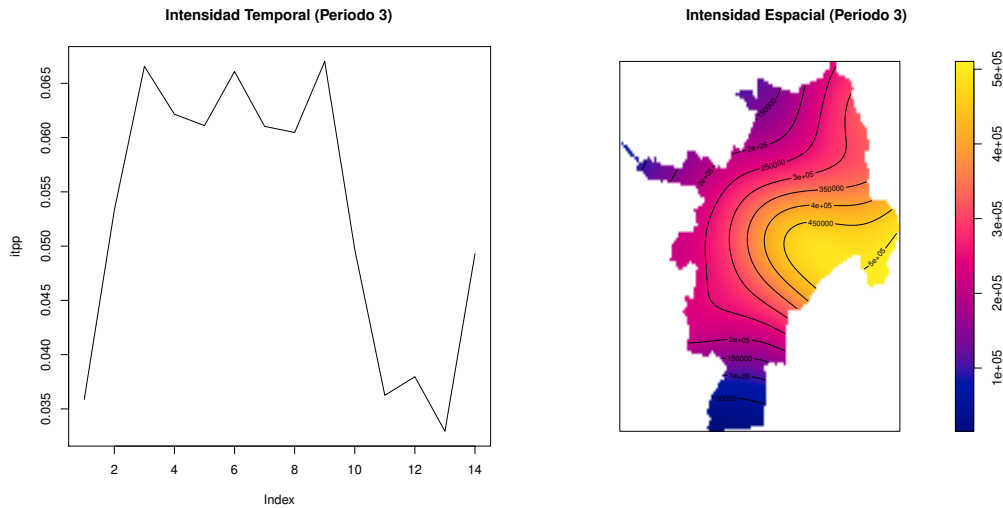


Figura 4.13 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 3

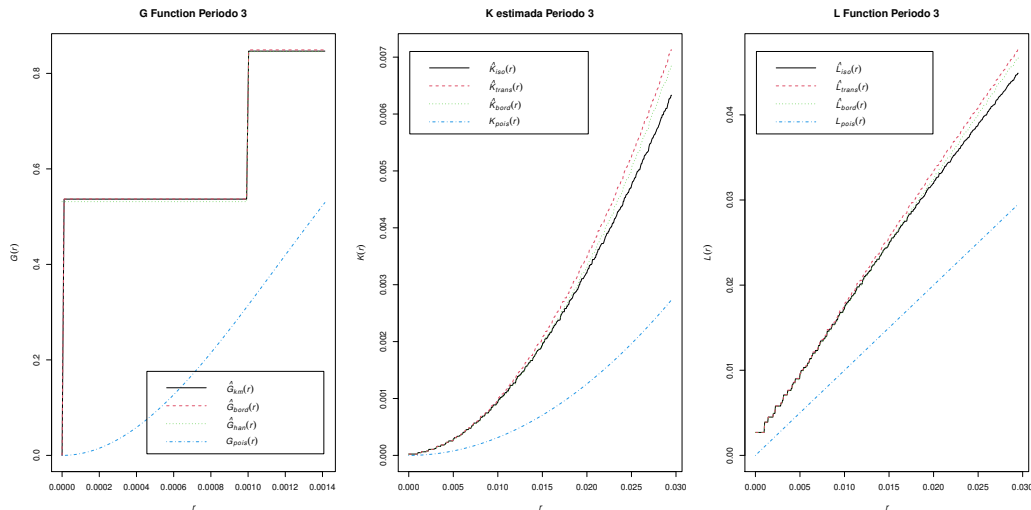


Figura 4.14 Función G, K y L - Periodo 3

Periodo 4

En el periodo 4 se tiene la semana epidemiológica 27 la cual está comprendida entre la fecha 28/06/2020 y 04/07/2020, con un acumulado de casos de 2020. Para esta semana epidemiológica se presenta el patrón de puntos en la figura 4.15, donde en primera medida no se presenta concentración de puntos que destaquen un patrón agregado, se nota presencia de casos al rededor de toda la Ciudad y por fuera del perimetro de la Ciudad.

Para analizar la intensidad espacio temporal en este periodo, la figura 4.16 se observa en la parte izquierda que hay un salto de casos en la intensidad temporal siendo indicio de patrón agregado, similarmente en la parte derecha se observa una concentración en el Oriente de la Ciudad.

Para caracterizar entonces si el patrón es agregado se presenta las gráficas G,K y L donde confirman el comportamiento agregado que existe en este periodo (semana 27) donde la mayor parte de los casos se produjo en el Oriente de la Ciudad.(Ver figura 4.17

Patrón de Puntos Periodo 4

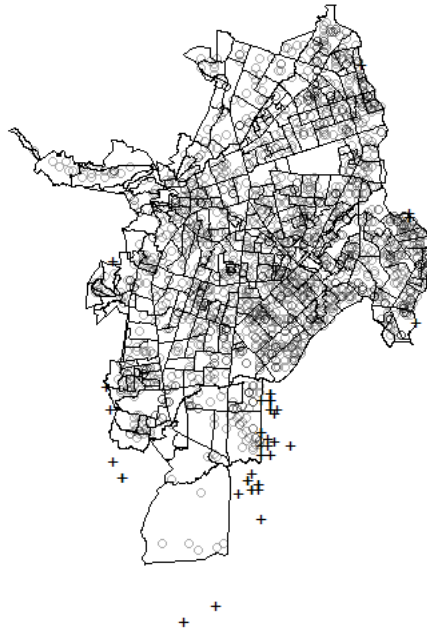


Figura 4.15 Patrón de Puntos - Periodo 4

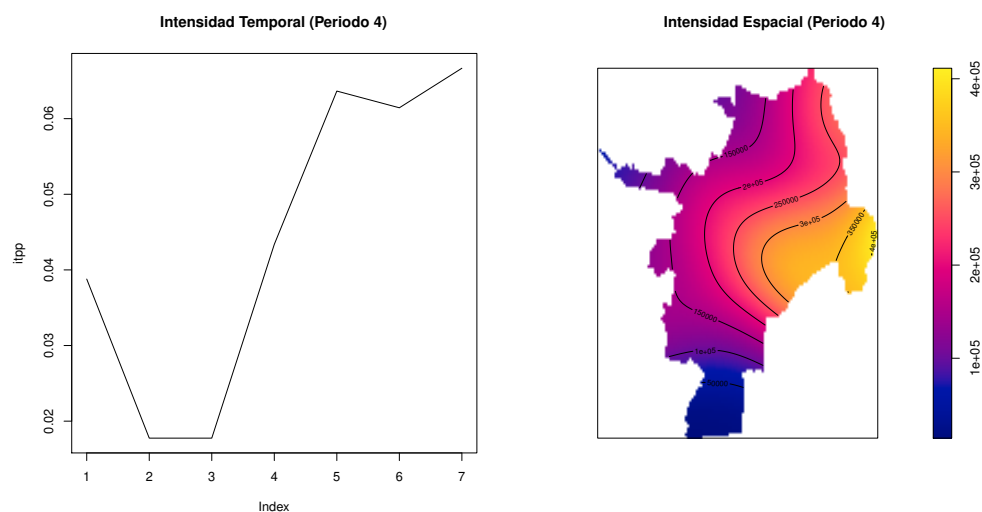


Figura 4.16 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 4

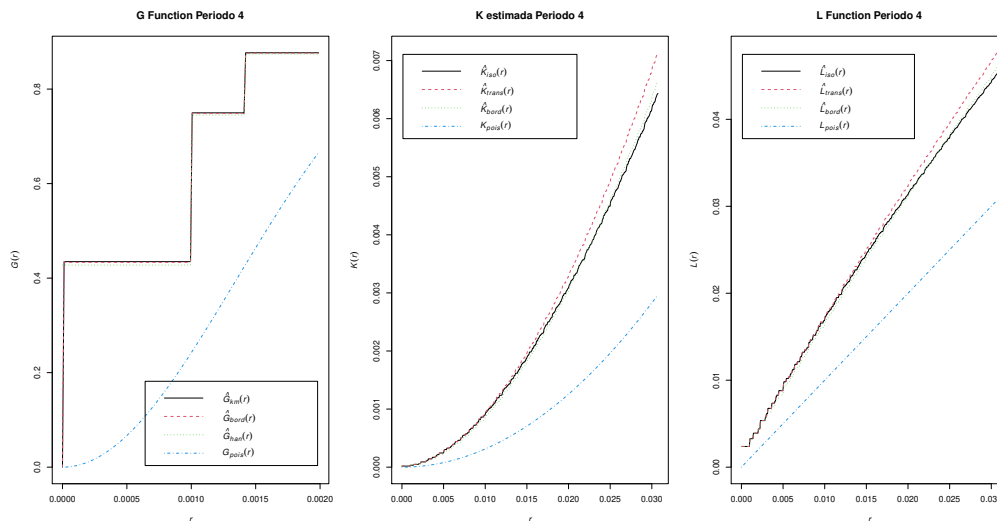


Figura 4.17 Función G, K y L - Periodo 4

Periodo 5

En el periodo 5 se tiene la semana epidemiológica 28 la cual está comprendida entre la fecha 05/07/2020 y 11/07/2020, con un acumulado de casos de 2981. En este periodo se sigue presentando incidencia de COVID-19 en el perímetro urbano y por fuera de este.

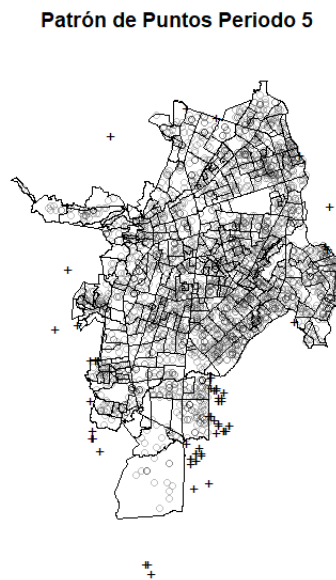


Figura 4.18 Patrón de Puntos - Periodo 5

Para el estudio de la intensidad en la figura 4.19 se presenta que la intensidad espacial y temporal, donde se observa en color amarillo la zona con más incidencia de COVID-19 y esta es la zona oriente del Santiago de Cali.

La prueba formal para caracterizar el patrón de puntos en este periodo se tiene que, la función G, K y L están por encima de la curva de la aleatoriedad espacial, por lo tanto se dice que el patrón de puntos en este periodo es agregado cuya concentración principalmente se encuentra en el Oriente y Nor-Oriente de la Ciudad.

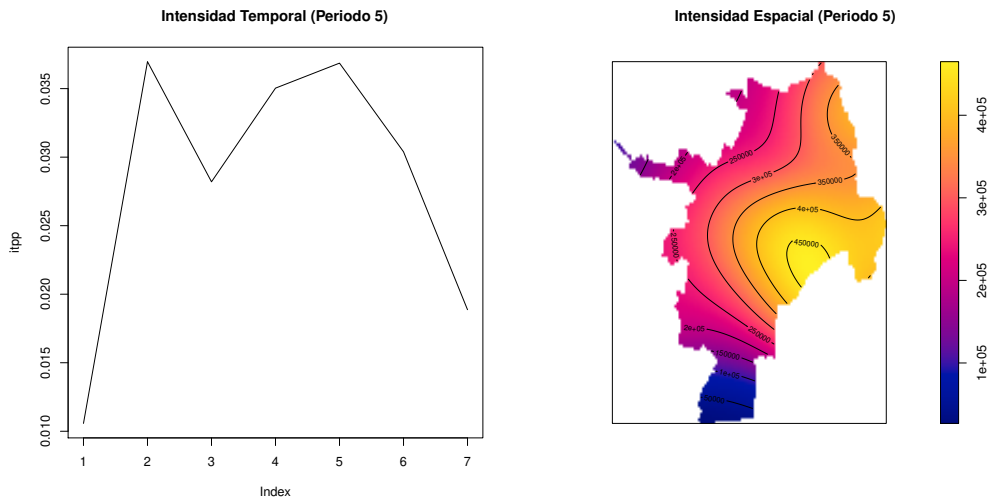


Figura 4.19 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 5

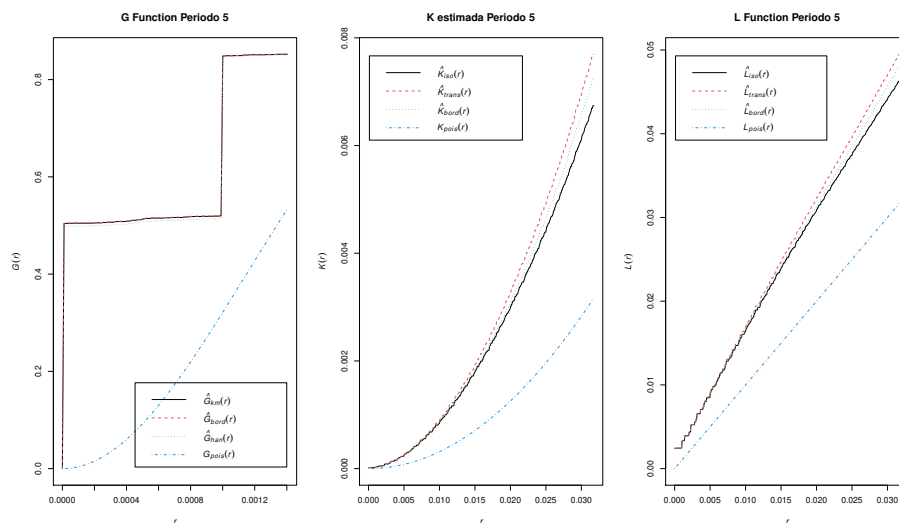


Figura 4.20 Función G, K y L - Periodo 5

Periodo 6

En el periodo 6 se tiene la semana epidemiológica 29 la cual está comprendida entre la fecha 12/07/2020 y 18/07/2020, con un acumulado de casos de 3367.

Patrón de Puntos Periodo 6



Figura 4.21 Patrón de Puntos - Periodo 6

Este es uno de los periodos (semana epidemiológica) con más casos de Covid-19, se observa el patrón de puntos en la figura 4.21, la intensidad espacio temporal se presenta en la figura 4.22 donde la concentración de de puntos se consagran en la zona Oriente y Norte de la ciudad (color amarillo) y se presenta poca incidencia en el sur de la Ciudad. Para identificar la aleatoriedad de los casos de COVID-19 en la figura 4.23 donde las curvas de la función G se presentan por encima de la curva de aleatoriedad espacial, por el lado de la función K y L sólo se presenta que la curva de corrección de borde está por encima de la curva del proceso poisson indicando que el proceso es agregado. Cabe resaltar que debido a que para este periodo hubo más de 3000 casos en las funciones K y L sólo aparece la corrección de borde.

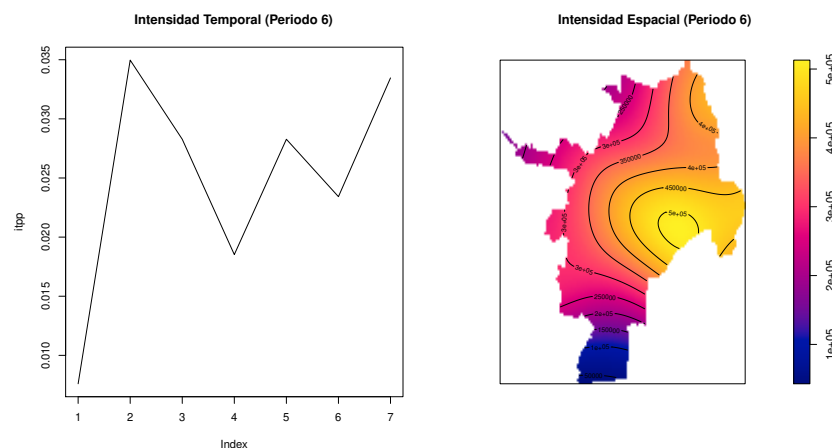


Figura 4.22 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 6

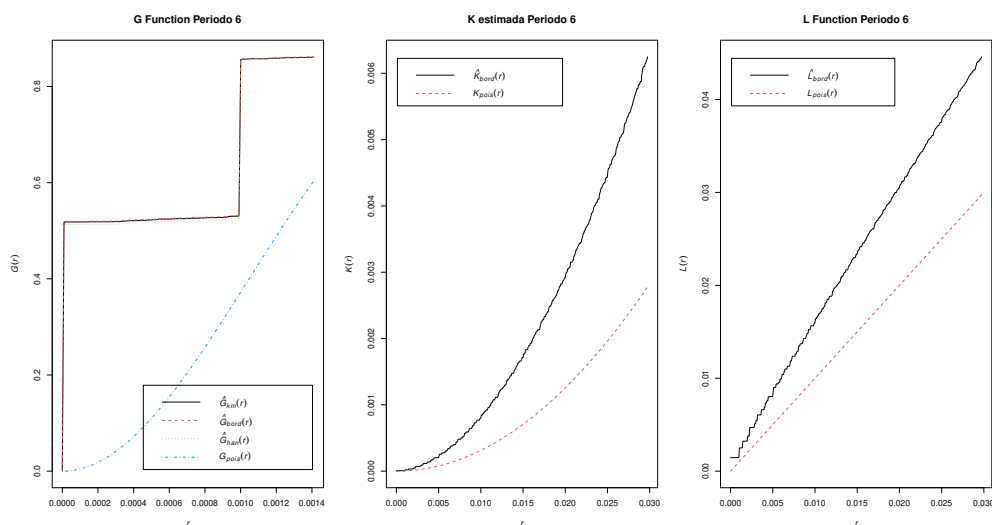


Figura 4.23 Función G, K y L - Periodo 6

Periodo 7

En el periodo 7 se tiene la semana epidemiológica 30 la cual está comprendida entre la fecha 19/07/2020 y 25/07/2020, con un acumulado de casos de 3640.

Patrón de Puntos Periodo 7

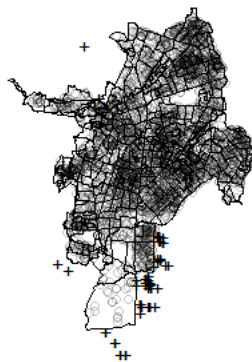


Figura 4.24 Patrón de Puntos - Periodo 7

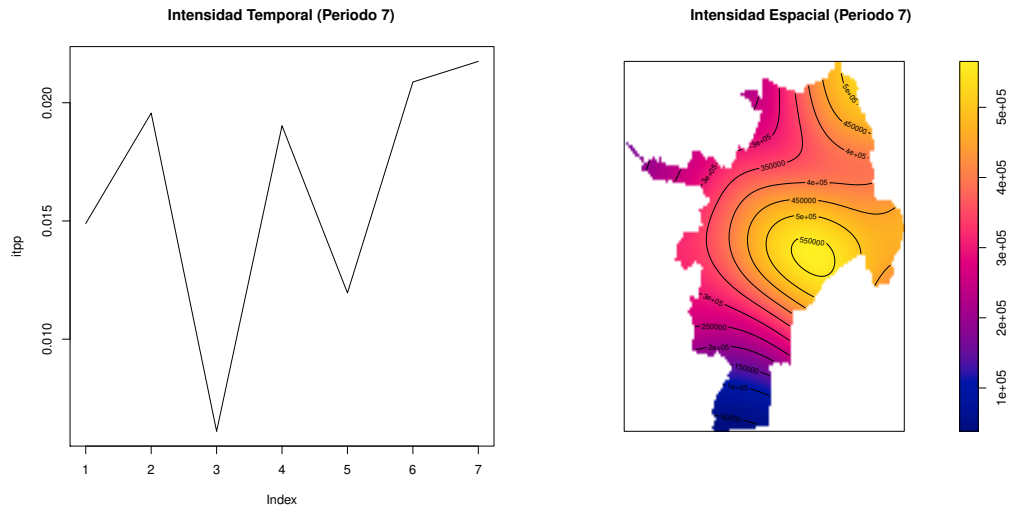


Figura 4.25 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 7

Esta es la semana epidemiológica con mayor incidencia de COVID-19 en la Ciudad caracterizado por el patrón de puntos en la figura 4.24, se observa que se presentan más casos a las afueras de la Ciudad. La intensidad de estos casos para este periodo al igual que los anteriores se presenta en su mayoría en la zona oriente de la Ciudad, la mayor parte de los casos se encuentra en las comunas 11,12,13,14, 15 y 16, cuya intensidad espacio temporal se encuentra en la figura 4.25. Debido a que en este periodo al igual que en el periodo 6 y 8 hay mas de 3000 casos las funciones K y L solo presentan la corrección de borde (Ver figura 4.26). Con respecto a la función G, K y L se observa que las curvas están por encima de la aleatoriedad espacial, se dice entonces que para este periodo el patrón de puntos es agregado cuyas zonas de concentración del virus es en la zona Oriente de Santiago de Cali.

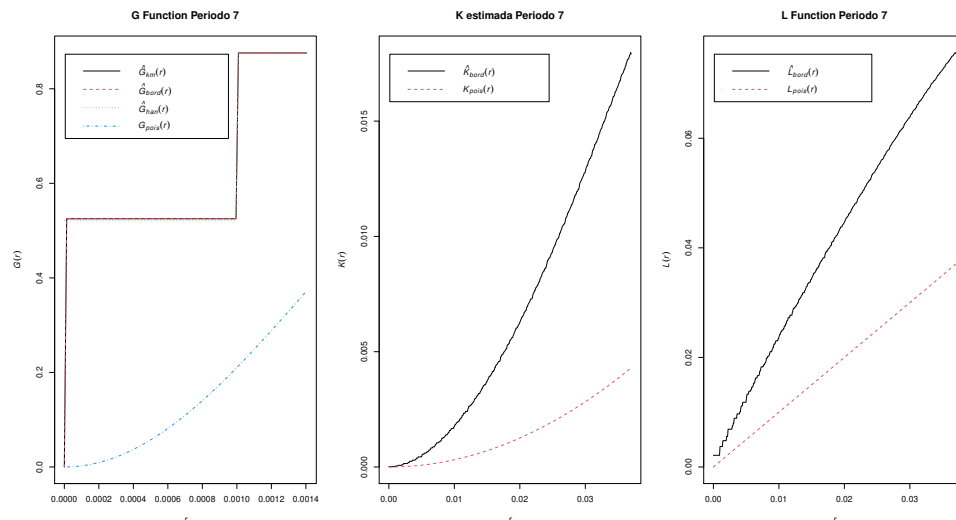


Figura 4.26 Función G, K y L - Periodo 7

Periodo 8

En el periodo 8 se tiene la semana epidemiológica 31 la cual está comprendida entre la fecha 26/07/2020 y 01/08/2020, con un acumulado de casos de 3360. Para esta semana los casos se mantienen por encima de los 3000 casos de Covid-19, la figura 4.28 se presenta la intensidad del patrón de puntos para este periodo, se observa que a pesar de que la concentración de casos de Covid-19 sigue siendo en la zona norte y oriente de la Ciudad, a diferencia del anterior periodo hay una menor concentración en estas zonas. Sin embargo, el patrón de puntos para este periodo sigue siendo un agregado, (Ya que las curvas de las funciones G, K y L están por encima del proceso aleatorio de Poisson) siendo la zona oriente y norte las más afectadas. (Ver Figura 4.29)

Patrón de Puntos Periodo 8

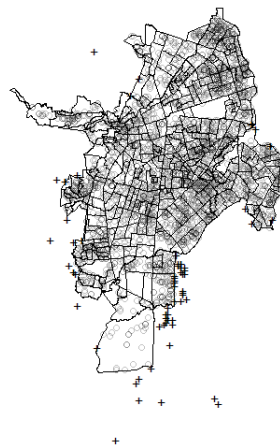


Figura 4.27 Patrón de Puntos - Periodo 8

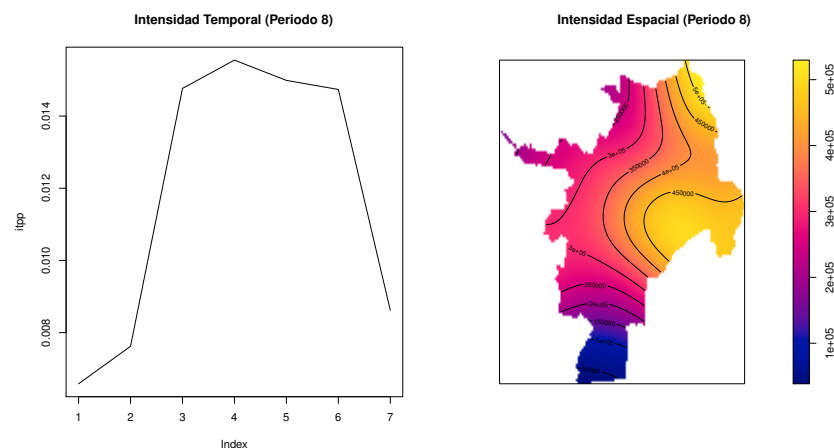


Figura 4.28 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 8

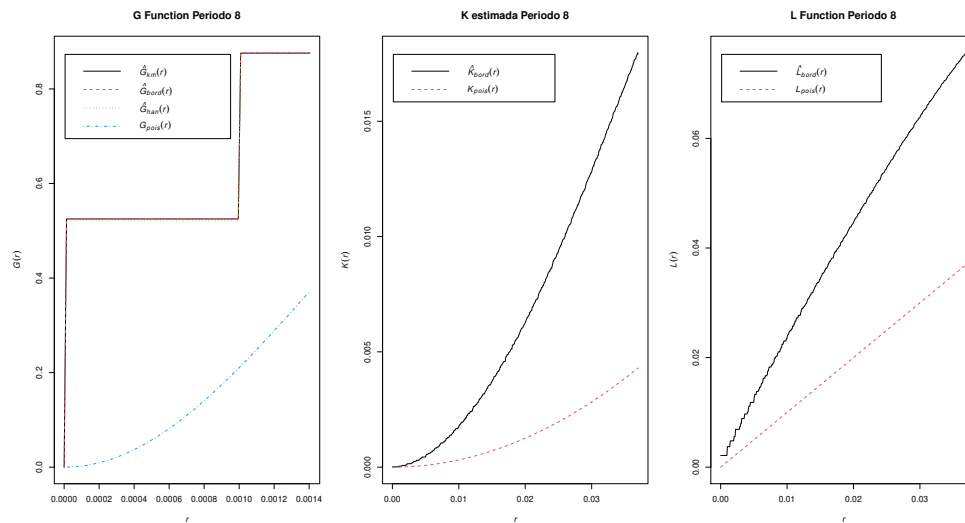


Figura 4.29 Función G, K y L - Periodo 8

Periodo 9

Patrón de Puntos Periodo 9



Figura 4.30 Patrón de Puntos - Periodo 9

En el periodo 9 se tiene la semana epidemiológica 32 la cual está comprendida entre la fecha 02/08/2020 y 08/08/2020, con un acumulado de casos de 2802. En esta

semana epidemiológica los casos vuelven a caer por debajo de los 3000, presentándose varios casos en la periferia de la Ciudad. Con respecto a la intensidad se observa que las zonas Oriente y Norte de Cali siguen siendo las más comunes para la concentración del virus donde estadísticamente se observa que si existe una diferencia de casos con respecto a otras zonas, ya que el proceso es agregado según las funciones G, K y L cuyas curvas se encuentran por encima del proceso aleatorio de Poisson. (Ver figura 4.32).

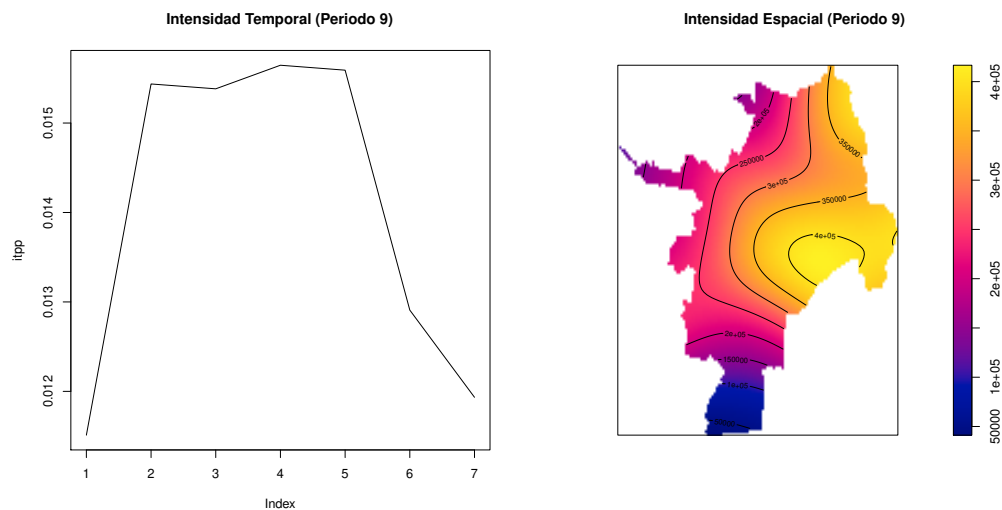


Figura 4.31 Estimación no-paramétrica de la intensidad temporal y espacial - Periodo 9

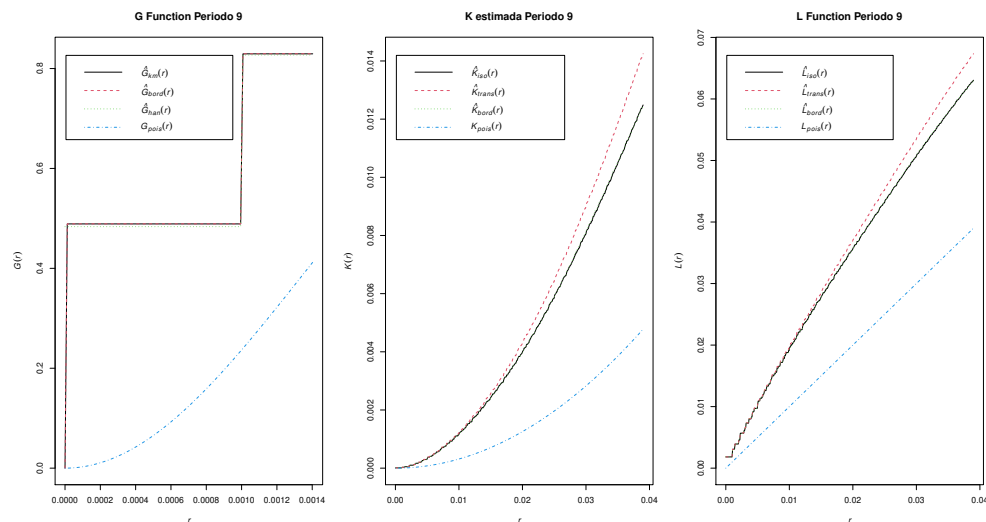


Figura 4.32 Función G, K y L - Periodo 9

Periodo 10

En el periodo 10 se tiene la semana epidemiológica 33 la cual está comprendida entre la fecha 02/08/2020 y 08/08/2020, con un acumulado de casos de 2811. En la figura 4.33 correspondiente al patrón de puntos, es evidente que la mayor densidad

de estos se presenta en la zona oriental y centro de la ciudad de Cali. Estos están comprendidos en su mayoría por las comunas 11,12,13,14,15,16 y 21. Sin embargo, en comparación con los periodos pasados se observa en este, que la enfermedad a estado avanzando a los largo de la ciudad a medida que pasan los días.

Esta intensidad de casos es posible verla más clara con un gráfico suavizado, mediante la metodología kernel en el gráfico derecho de la figura 4.34, se hacen evidentes las zonas más afectadas por la enfermedad. Donde, en este caso se confirma el análisis anterior por medio de patrones puntuales. Por otra parte en la gráfica izquierda (intensidad temporal), se observa una semana epidemiológica muy variante en los primeros tres días, porque, hay un declive de la enfermedad y luego un repunte que se mantiene en el tiempo.

Con las funciones K,L y G, se busca encontrar si en esta semana epidemiológica existe un patrón agregado o una dispersión estadísticamente significativa. De este modo, si se observa la figura 4.35, en los tres estadísticos la función empírica esta por encima de la teórica. Es decir, que este periodo de tiempo presenta un patrón de datos agregados.

Patrón de Puntos Periodo 10

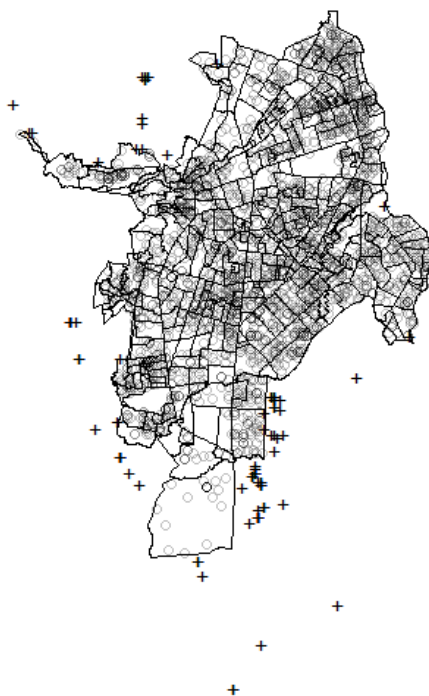


Figura 4.33 Patrón de puntos - Periodo 10

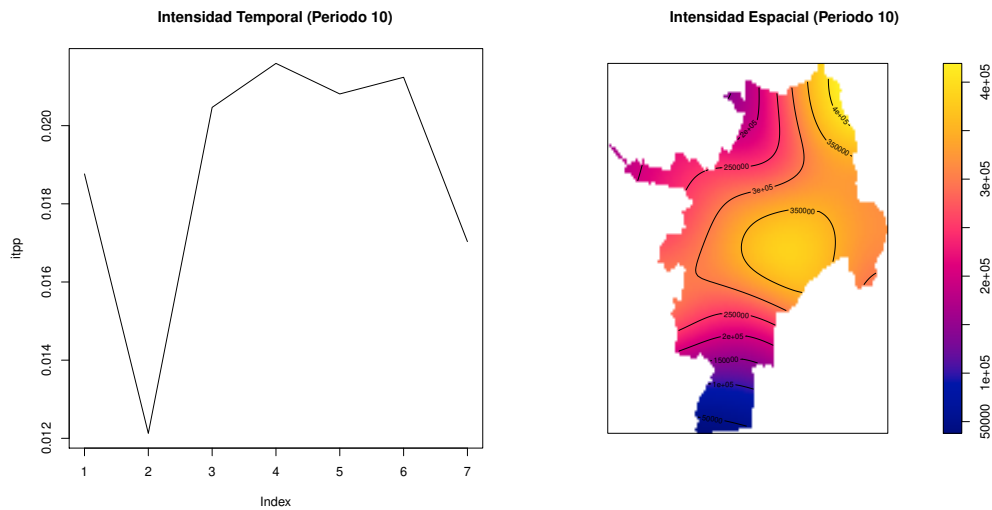


Figura 4.34 Estimación no paramétrica del intensidad temporal y espacial - Periodo 10

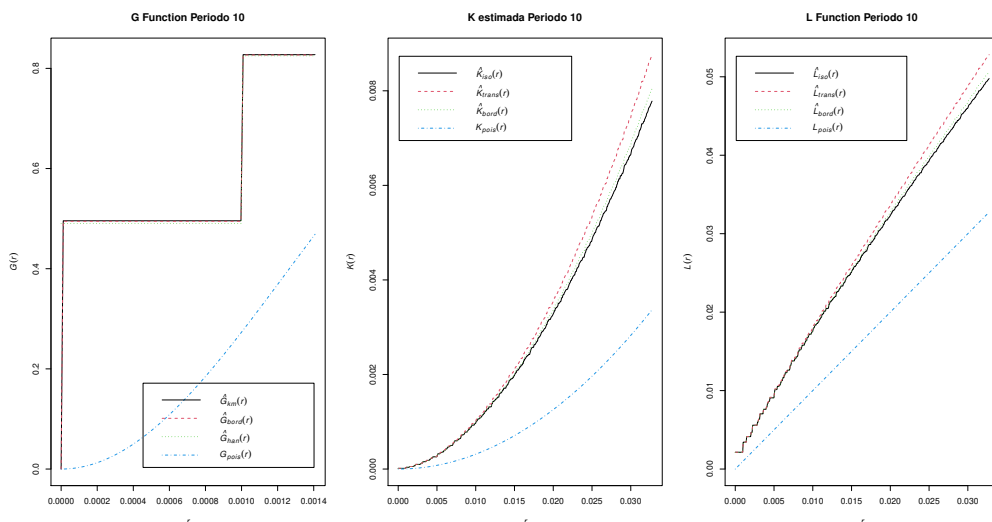


Figura 4.35 Función G, K y L - Periodo 10

Periodo 11

En el periodo 11 se tiene la semana epidemiológica 34 la cual está comprendida entre la fecha 16/08/2020 y 22/08/2020, con un acumulado de casos de 2098. Nuevamente se observa mayor incidencia de casos en las zonas nor-oriente, centro y oriente de la ciudad (ver figura 4.36). Sin embargo, en la gráfica de intensidad espacial (figura 4.36) se observa un aumento de casos en la zona nor-occidente. Adicionalmente, en el gráfico de intensidad temporal, se hace evidente un aumento de casos en esta semana epidemiológica.



Figura 4.36 Patrón de puntos - Periodo 11

Al igual que los periodos anteriores, se presenta un patrón de puntos agregados en este periodo. Esto debido a que, la mayor incidencia de casos se observan en áreas específicas de la ciudad mencionadas anteriormente (ver figura 4.38).

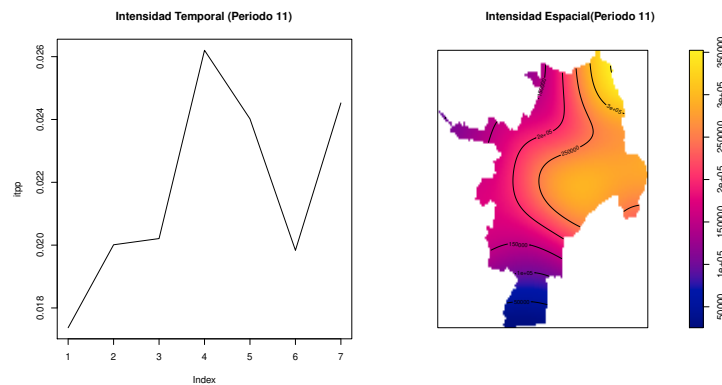


Figura 4.37 Estimación no paramétrica del intensidad temporal y espacial - Periodo 11

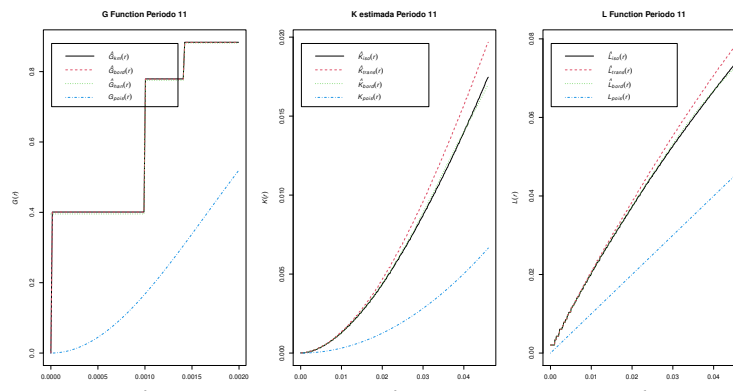


Figura 4.38 Función G, K y L - Periodo 11

Periodo 12



Figura 4.39 Patrón de puntos - Periodo 12

En el periodo 12 se tiene la semana epidemiológica 35 la cual está comprendida entre la fecha 23/08/2020 y 29/08/2020, con un acumulado de casos de 1839. Aunque la zona oriente y nor-oriente sigue siendo el epicentro de la enfermedad en Cali, se observa en este periodo mas incidencias en las zonas nor-occidente y parte del sur (ver figuras 4.39 y 4.40).

Adicionalmente, en la figura 4.41 se aprecia que se presenta un patrón de datos agregados. Esto, podría indicar que la enfermedad se dispersa con mayor facilidad en zonas donde hay mayor densidad de individuos.

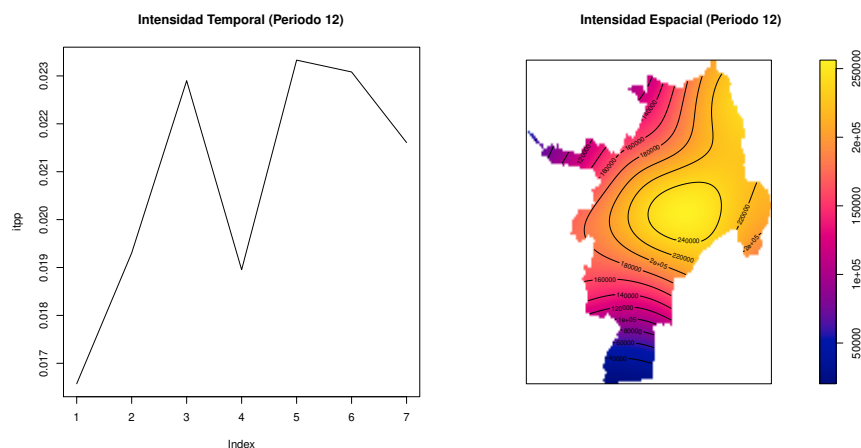


Figura 4.40 Estimación no paramétrica del intensidad temporal y espacial - Periodo 12

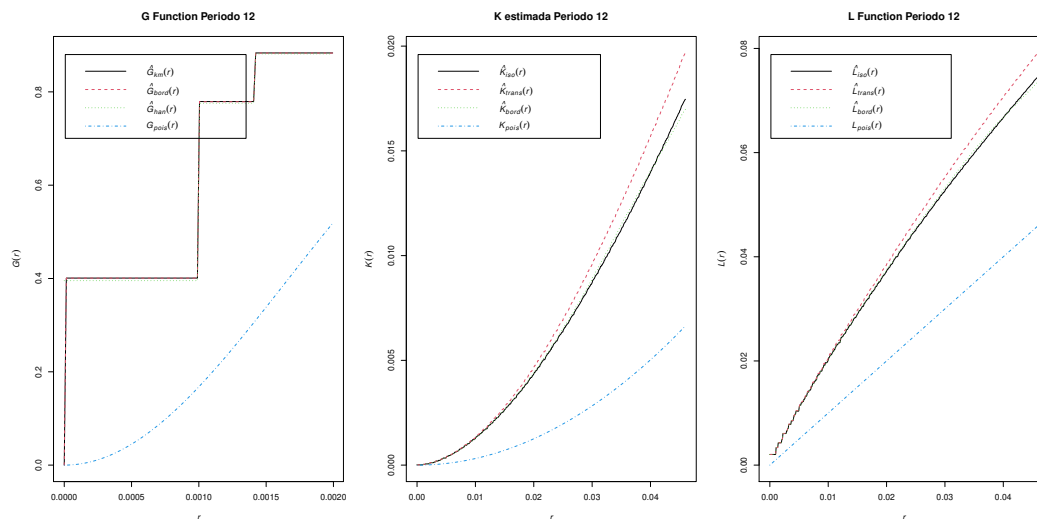


Figura 4.41 Función G, K y L - Periodo 12

Periodo 13

En el periodo 13 se tiene la semana epidemiológica 36 la cual está comprendida entre la fecha 30/08/2020 y 05/09/2020, con un acumulado de casos de 1656. Es apreciable tanto en la figura 4.42 y la gráfica de intensidad espacial en la figura 4.43, el aumento de casos en zonas que en periodos anteriores no presentaban grandes incidencias. Se observa en este periodo, una dispersión de la enfermedad por casi el 90 % del territorio de la ciudad. Sin embargo, la parte comprendida por las comunas 17 y 22, se observan incidencias bajas de la enfermedad.



Figura 4.42 Patrón de puntos - Periodo 13

Por otra parte, nuevamente se observa un patrón agregado por parte de los datos espaciales. Esto se hace evidente en la figura 4.44, donde la función empírica esta por encima de la función teóricas en los tres estadísticos planteados para evaluar la aleatoriedad espacial.

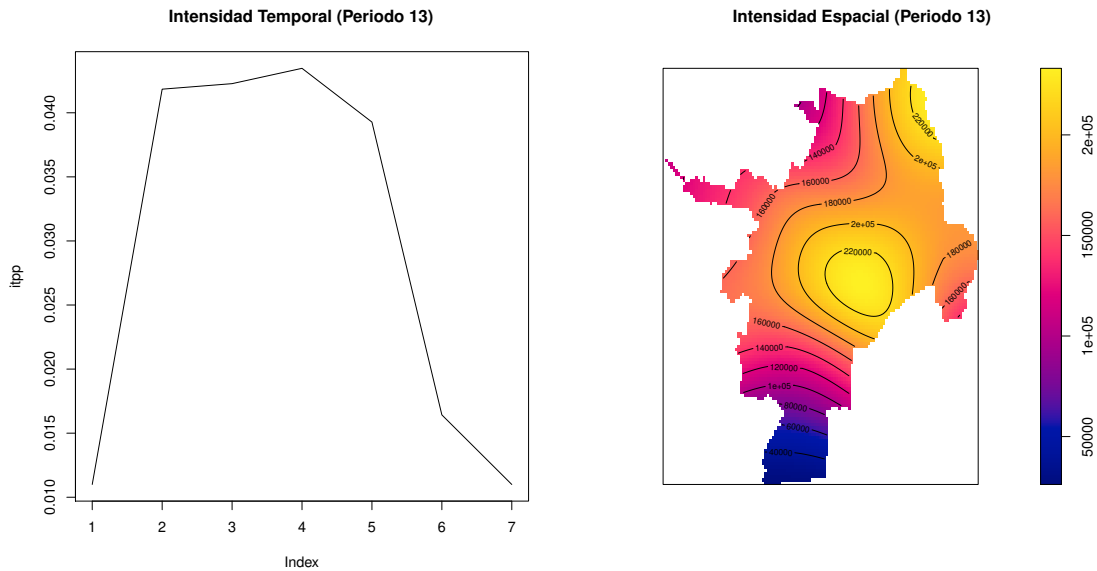


Figura 4.43 Estimación no paramétrica del intensidad temporal y espacial - Periodo 13

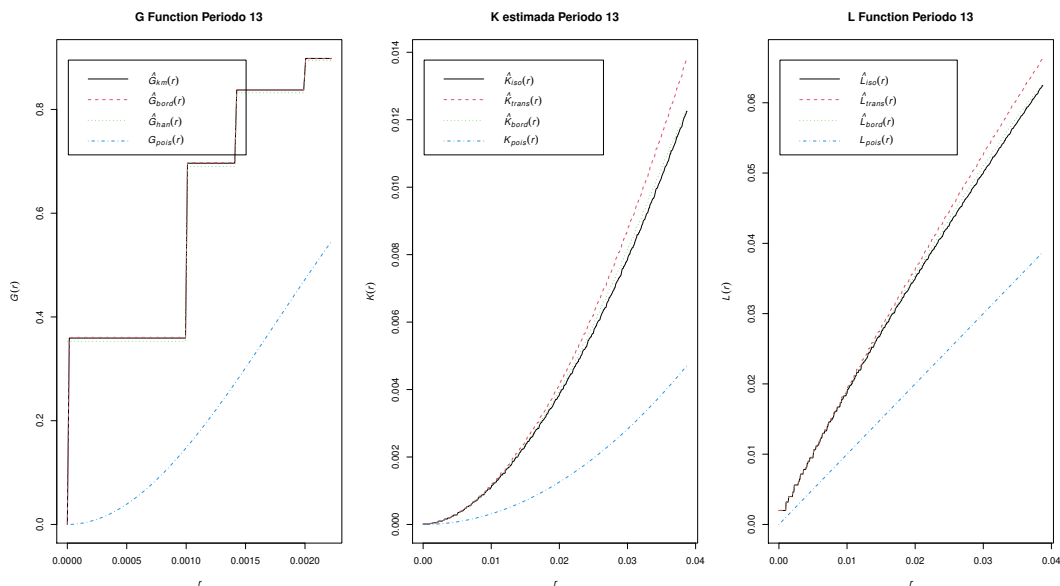


Figura 4.44 Función G, K y L - Periodo 13

Periodo 14

Patrón de Puntos Periodo 14



Figura 4.45 Patrón de puntos - Periodo 14

En el periodo 14 se tiene la semana epidemiológica 37 la cual está comprendida entre la fecha 06/09/2020 y 12/09/2020, con un acumulado de casos de 1939. En este periodo se observa que la mayor incidencia de casos se presenta en el nor-oriente de la ciudad, afectando en su mayoría a las comunas 5,6 y 7 (ver figura 4.45 Y 4.46). En el gráfico de intensidad temporal (4.46), se empieza a observar una intensidad que cambia a medida que pasan los días y al final se observa un decrecimiento importante de esta intensidad de casos.

También, según las funciones G,K y L, el patrón observado de los puntos sigue siendo agregado al igual que los periodos anteriores. La diferencia de este periodo radica que, el mayor número de incidencias de la enfermedad ya no las presenta el oriente-centro, sino, la zona nor-oriental de la ciudad.

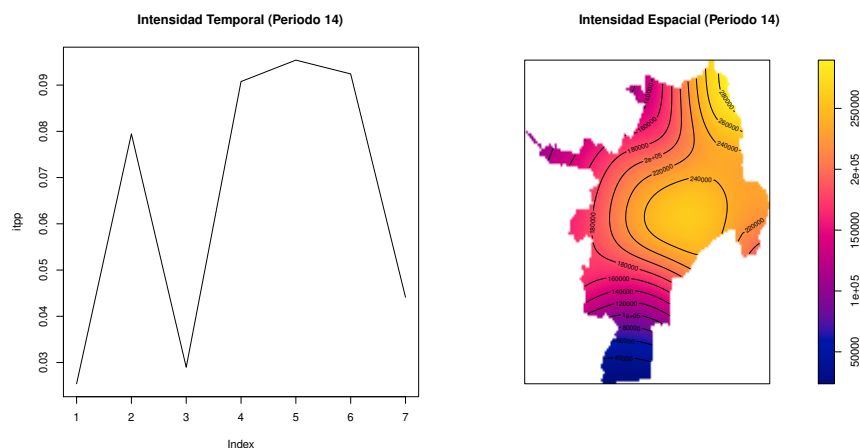


Figura 4.46 Estimación no paramétrica del intensidad temporal y espacial - Periodo 14

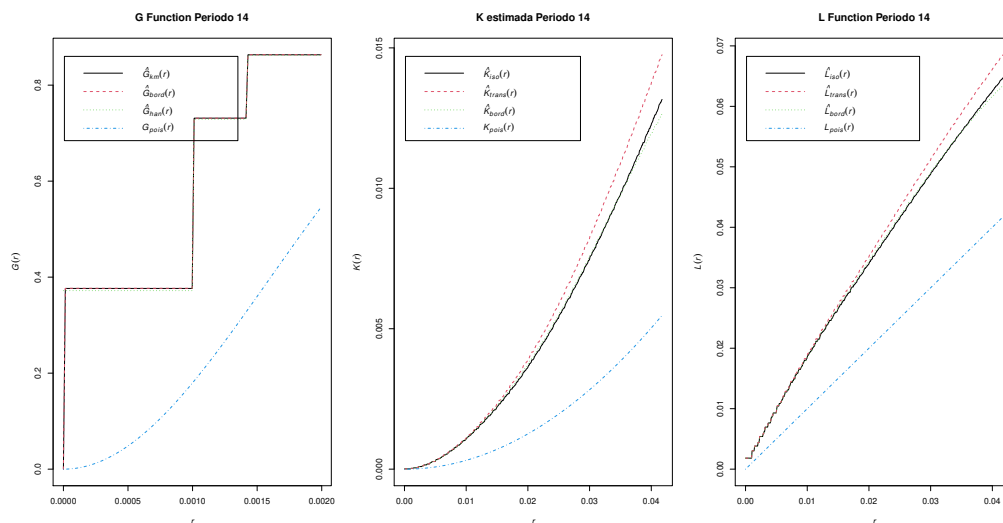


Figura 4.47 Función G, K y L en Periodo 14

Periodo 15

En el periodo 15 se tiene la semana epidemiológica 38 la cual está comprendida entre la fecha 13/09/2020 y 19/09/2020, con un acumulado de casos de 1915. Se observan dos puntos amarillos en la ciudad y los patrones puntuales están siendo mas densos en zonas específicas de Cali. Los puntos amarillos indican la mayor intensidad de la enfermedad y esto corrobora el gráfico de patrones puntuales (ver figura 4.48 Y 4.49).

Patrón de Puntos Periodo 15

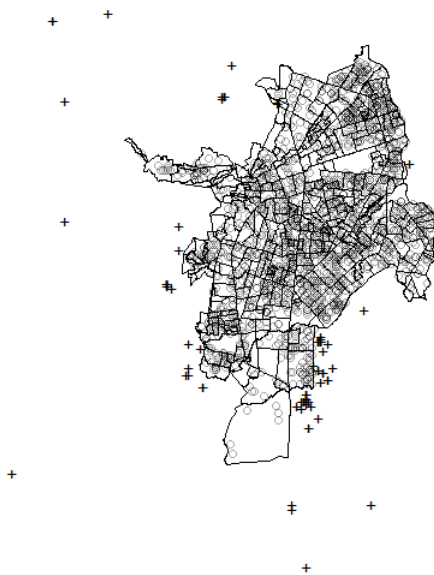


Figura 4.48 Patrón de puntos - Periodo 15

Así, es posible indicar que las dos zonas de Cali que presentan mayor incidencias de la enfermedad, son la parte oriente y nor-oriente de esta. También, si se compara con lo periodos iniciales se observa que la enfermedad en este periodo, logró dispersarse por casi todo el territorio urbano de la ciudad de Cali.

En la figura 4.50, se observa que los datos son agregados, donde sobresalen los dos grupos mencionados anteriormente.

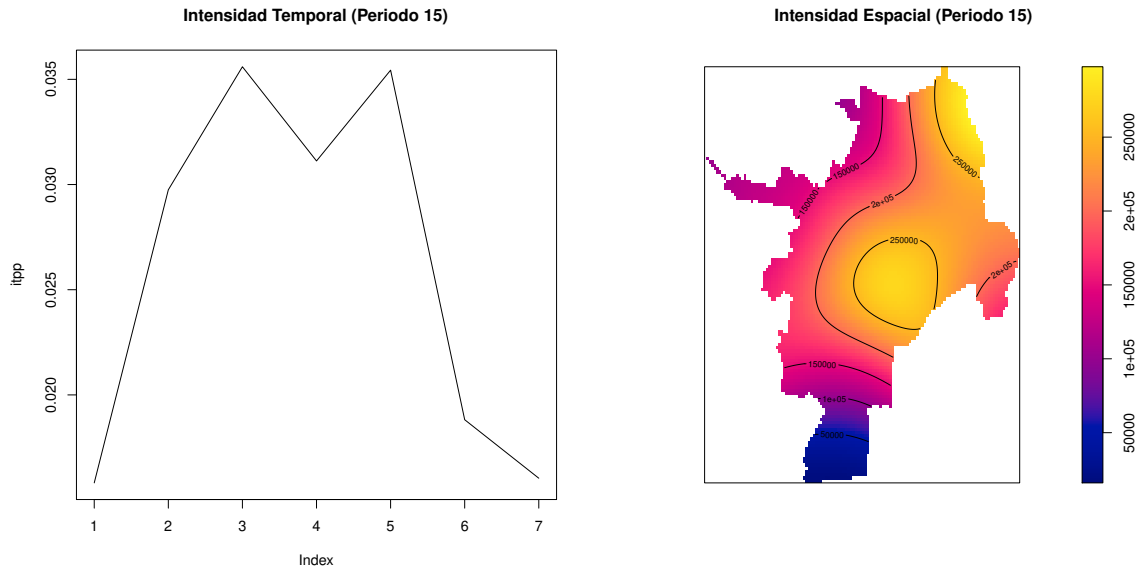


Figura 4.49 Estimación no paramétrica del intensidad temporal y espacial - Periodo 15

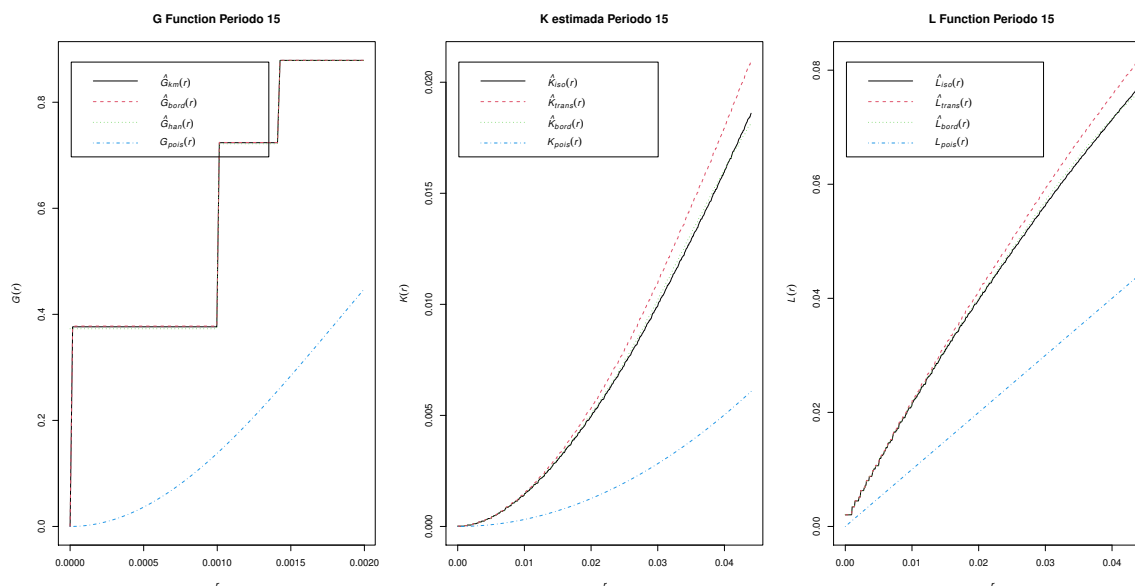


Figura 4.50 Función G, K y L - Periodo 15

Periodo 16

Patrón de Puntos Periodo 16



Figura 4.51 Patrón de puntos - Periodo 16

En el periodo 16 se tiene la semana epidemiológica 39 la cual está comprendida entre la fecha 20/09/2020 y 26/09/2020, con un acumulado de casos de 1310. Esta es la ultima semana epidemiológica de esta investigación, diferencias significativas entre el periodo no se encontraron. Sin embargo, se observa que la incidencias en esta semana son más bajas en comparación con el periodo 15, pero, las zonas más afectadas siguen siendo las mismas (figuras 4.51 Y 4.52).

La gráfica de intensidad temporal (figura 4.52), muestra claramente una tendencia decreciente después del quinto día de la semana estadia, lo que concuerdan con los datos de la pagina oficial del gobierno colombiano. También, en la figura 4.53 se presenta un patrón agregado de punto, lo cual, coincide con los periodos anteriormente analizados.

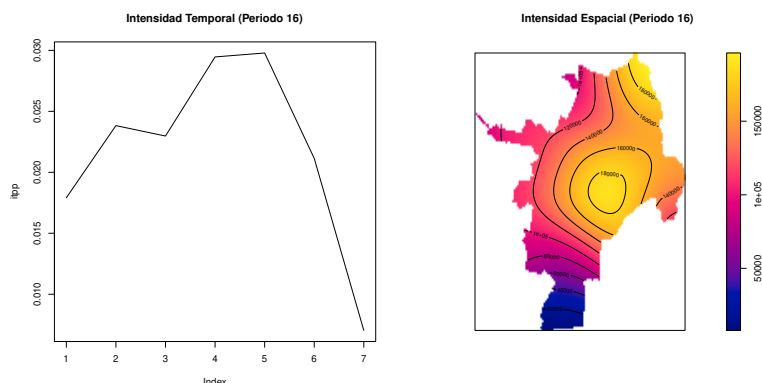


Figura 4.52 Estimación no paramétrica del intensidad temporal y espacial - Periodo 16

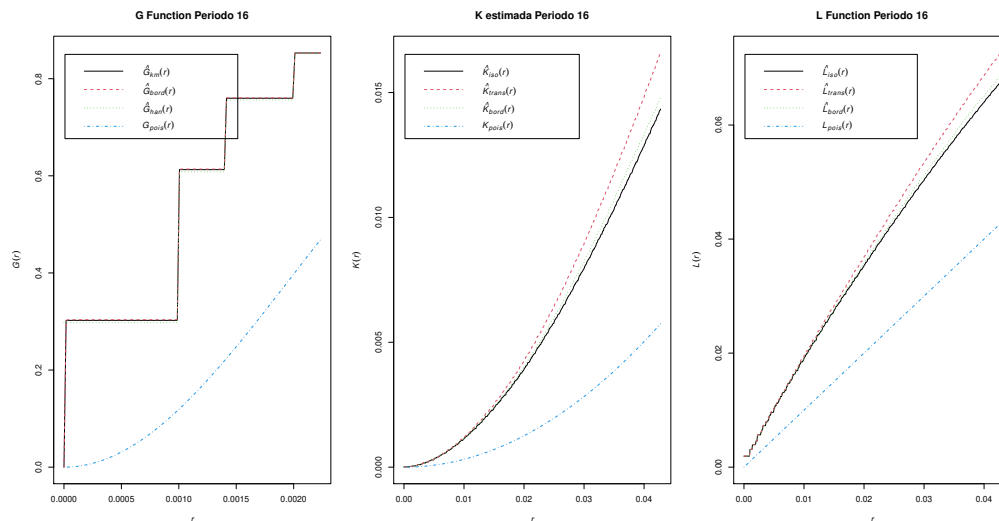


Figura 4.53 Función G, K y L - Periodo 16

En la tabla 4.9, se presenta un resumen ejecutivo donde se evidencia la evolución de la pandemia por todo el territorio municipal de la ciudad de Cali. Se observa, que los primeros caso fueron registrados en el centro de la ciudad, de allí fue creciendo al oriente y oeste de la ciudad, finalmente termina afectando a la zona norte. El sur destaca por ser la zona menos afectada en este periodo de tiempo.

Tabla 4.9 Resumen del análisis espacio temporal

Periodos	Gráficos	Semanas epidemiológicas	Casos positivos	Tipo de patrón	Zonas con mayor densidad de casos
1	4.7 al 4.8	3 hasta la 22	2825	Agregado	centro
2	4.9 al 4.11	23 y 24	2196	Agregado	centro
3	4.12 al 4.14	25 y 26	2808	Agregado	centro-suroriente
4	4.15 al 4.17	27	2019	Agregado	centro-suroriente
5	4.18 al 4.20	28	2981	Agregado	centro-suroriente/nororiente
6	4.21 al 4.23	29	3367	Agregado	centro-suroriente/nororiente
7	4.24 al 4.26	30	3640	Agregado	centro-suroriente/nororiente
8	4.27 al 4.29	31	3360	Agregado	centro-suroriente/nororiente
9	4.30 al 4.32	32	2802	Agregado	centro-suroriente/nororiente
10	4.33 al 4.35	33	2811	Agregado	centro-suroriente/nororiente-oeste
11	4.36 al 4.38	34	2098	Agregado	centro-suroriente/nororiente-oeste
12	4.39 al 4.41	35	1839	Agregado	centro-suroriente/nororiente-oeste
13	4.42 al 4.44	36	1656	Agregado	centro-suroriente/nororiente-oeste
14	4.45 al 4.47	37	1939	Agregado	centro-suroriente/nororiente-oeste
15	4.48 al 4.50	38	1915	Agregado	centro-suroriente/nororiente-oeste-norte
16	4.51 al 4.53	39	1310	Agregado	centro-suroriente/nororiente-oeste-norte

4.3. Análisis de redes de contacto

Como se mencionó en la metodología, se obtuvo un total de 1849 individuos para la construcción de las redes de contacto, estos se dividen entre 262 vértices raíz (correspondientes a los casos positivos encuestados) y 1587 vértices hijos (Enlaces o conexiones totales construidas). En esta parte, se observa una de las importancias de las redes de contacto como metodología aliada para explorar este tipo de enfermedades, es evidente el alcance del COVID-19 porque con tan solo 262 individuos positivos encuestados, se logró encontrar 1587 casos positivos posibles de esta enfermedad.

El peso de los bordes, se realizó teniendo en cuenta el tipo de contacto que el individuo tuvo con otro. Los pesos van de uno hasta cuatro, donde, uno hace referencia a menos probabilidad de contagio y cuatro a mayor probabilidad de contagio. Estos se desglosan en cuatro tipos:

- **Social:** Referente a cualquier contacto estrecho¹¹ que el individuo halla tenido con personas que no hagan parte del círculo familiar y laboral. A este primer tipo de contacto se le asignó un peso de uno, debido a las características del periodo de tiempo estudiado. Es decir, como a finales de marzo y principios de septiembre la cuarentena obligatoria estaba vigente, la probabilidad de contagio por contacto social era baja.
- **Laboral:** Es cualquier contacto estrecho que el individuo (no personal de salud) tiene en su círculo laboral. El peso de este tipo de contacto es de dos, porque en la mayoría de las empresas se tomaron medidas estrictas de bioseguridad.
- **Familiar:** Cualquier contacto estrecho que el individuo tuvo en su círculo familiar. El peso es de tres, porque se hace evidente que en la mayoría de reuniones familiares, visitas etc., las medidas de bioseguridad son ignoradas.
- **Personal de salud:** Individuos que desempeñan labores en el área de la salud, como médicos, personal de aseo, vigilantes, entre otros. Se le asigna un peso de cuatro, por la alta exposición que estos individuos tienen ante el virus.

En la tabla 4.10, se presentan las frecuencias de los tipos de vínculos representados en los grafos. Se evidencia, que en el periodo de tiempo estudiado, el mayor vínculo fue familiar seguido por social. Esta variable presentó un 14 % de N/D.

Tipo de contacto	Frecuencia absoluta	Frecuencia relativa	Peso del enlace
SOCIAL	225	16.5 %	1
LABORAL	331	24.24 %	2
FAMILIAR	778	57 %	3
PERSONAL DE SALUD	31	2.3 %	4

Tabla 4.10 Frecuencia de los tipos de enlaces generados por los individuos positivos de COVID-19

La tabla 4.11 muestra los resultados de la variable “Toma de prueba”, esta variable se midió únicamente con los contactos que los 262 individuos confirmados con COVID-19 reportaron en la encuesta. Se observa una cifra preocupante, porque el porcentaje de datos perdidos fue 75.9 % del total de los datos (1587). Adicionalmente, se reportaron 14 pruebas negativas, 342 en estado pendiente y 26 positivas. Este fenómeno se

¹¹**Contacto estrecho:** Estar con una persona en un lugar cerrado a menos de dos metros de distancia, sin tapabocas y por más de 15 minutos (ver <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/VSP/pspm01-man-lineamientos-formulacion-casos-vsp.pdf>).

produjo por varios factores, entre los que están, no colaboración de las personas, mala recolección de datos, no conocimiento de la situación en ese momento, negación a hacerse la prueba, entre otros.

Resultado	Frecuencia absoluta	Frecuencia relativa
NEGATIVO	14	0.88 %
PENDIENTE	342	21.55 %
POSITIVO	26	1.64 %

Tabla 4.11 Visualización de resultados de pruebas realizadas para detección de COVID-19

En la figura 4.54 y la tabla 4.12, es posible observar los casos positivos por COVID-19 que tuvieron más contactos antes de ser confirmados y aislados según las medidas bioseguridad vigentes. Se hace evidente la facilidad con la que se puede propagar el virus si no se tienen medidas efectivas y eficaces para la detección del mismo. También, gracias a la construcción de las redes de contacto, es mucho más sencillo identificar estos casos sospechosos, con el fin de mantener controlada la propagación de la enfermedad.

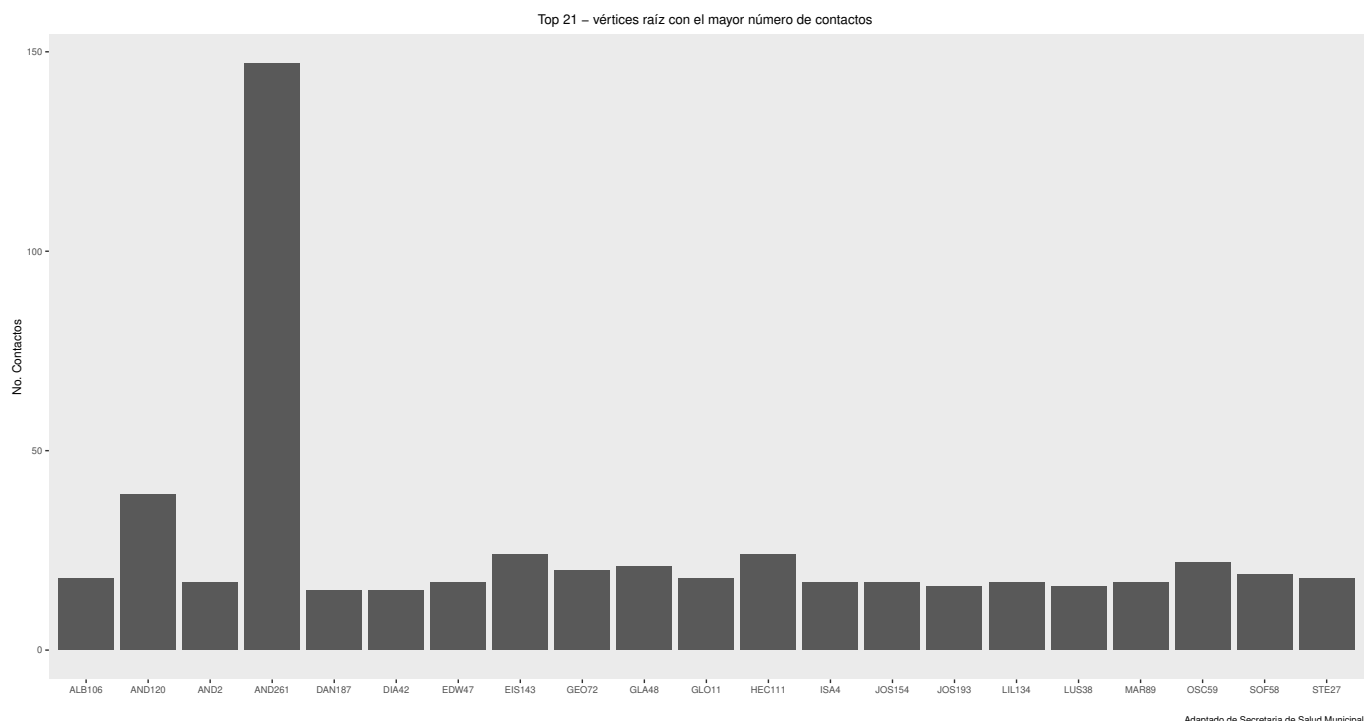


Figura 4.54 Ranking de los individuos positivos con COVID-19 con mayor contactos registrados

	Individuos	Frecuencia absoluta	Frecuencia relativa
1	AND261	147	9.26 %
2	AND120	39	2.5 %
3	EIS143	24	1.5 %
4	HEC111	24	1.5 %
5	OSC59	22	1.39 %
6	GLA48	21	1.32 %
7	GEO72	20	1.26 %
8	SOF58	19	1.2 %
9	ALB106	18	1.13 %
10	GLO11	18	1.13 %
11	STE27	18	1.13 %
12	AND2	17	1.07 %
13	EDW47	17	1.07 %
14	ISA4	17	1.07 %
15	JOS154	17	1.07 %
16	LIL134	17	1.07 %
17	MAR89	17	1.07 %
18	JOS193	16	1 %
19	LUS38	16	1 %
20	DAN187	15	0.94 %
21	DIA42	15	0.94 %

Tabla 4.12 Ranking de los individuos positivos con COVID-19 con mayor contactos registrados

La matriz de adyacencia no fue posible mostrarla completa en este documento, dada las dimensiones que tiene. La matriz es simétrica y cuadrada de tamaño 1849 x 1849 (debido a los tipos de grafos no dirigidos), los cruces entre los vértices tendrán los pesos correspondientes a dicho enlace y ceros en el resto (ver figura 4.55).

	CRI11	MIR12	MIR13	SAR14	WAL15	JHO16	VAN17	JOS28	ALB29	MIR210	JUA211
ROD1	1	3	3	3	3	3	3	3	0	0	0
AND2	0	0	0	0	0	0	0	0	1	1	1
JAI3	0	0	0	0	0	0	0	0	0	0	0
ISA4	0	0	0	0	0	0	0	0	0	0	0
HER5	0	0	0	0	0	0	0	0	0	0	0
ALF7	0	0	0	0	0	0	0	0	0	0	0
MAR8	0	0	0	0	0	0	0	0	0	0	0
JUA9	0	0	0	0	0	0	0	0	0	0	0
MAB10	0	0	0	0	0	0	0	0	0	0	0
GLO11	0	0	0	0	0	0	0	0	0	0	0
YU14	0	0	0	0	0	0	0	0	0	0	0
ANT16	0	0	0	0	0	0	0	0	0	0	0
JHO17	0	0	0	0	0	0	0	0	0	0	0
RUB18	0	0	0	0	0	0	0	0	0	0	0
VIR19	0	0	0	0	0	0	0	0	0	0	0
CA20	0	0	0	0	0	0	0	0	0	0	0
MAR21	0	0	0	0	0	0	0	0	0	0	0
ROS22	0	0	0	0	0	0	0	0	0	0	0
CAR25	0	0	0	0	0	0	0	0	0	0	0
ADR26	0	0	0	0	0	0	0	0	0	0	0
STE27	0	0	0	0	0	0	0	0	0	0	0
DAH28	0	0	0	0	0	0	0	0	0	0	0

Figura 4.55 Fragmento de la matriz de adyacencia

Los grafos ayudan a visualizar mejor el fenómeno a estudiar, a entender de una forma visual e interactiva la evolución del COVID-19. También, son una herramienta útil porque permiten agrupar o segmentar, para identificar a nivel epidemiológico individuos que podrían ser más propensos a contagiar nuevos individuos, además informan nuevos casos posibles de contagios.

A continuación, se presentan algunos de los grafos construidos a partir de la base

de datos suministrada por la secretaria de salud municipal. El análisis se dividió en tres partes por conveniencia visual, por lo cual, en las siguientes figuras se presentan tres bosques y algunos arboles estrella contenido en los tres bosques. Los vértices naranjas son individuos de sexo femenino y los grises de sexo masculino, el grosor del enlace es proporcional al peso asignado (ver figuras de la 4.56 a la 4.61).

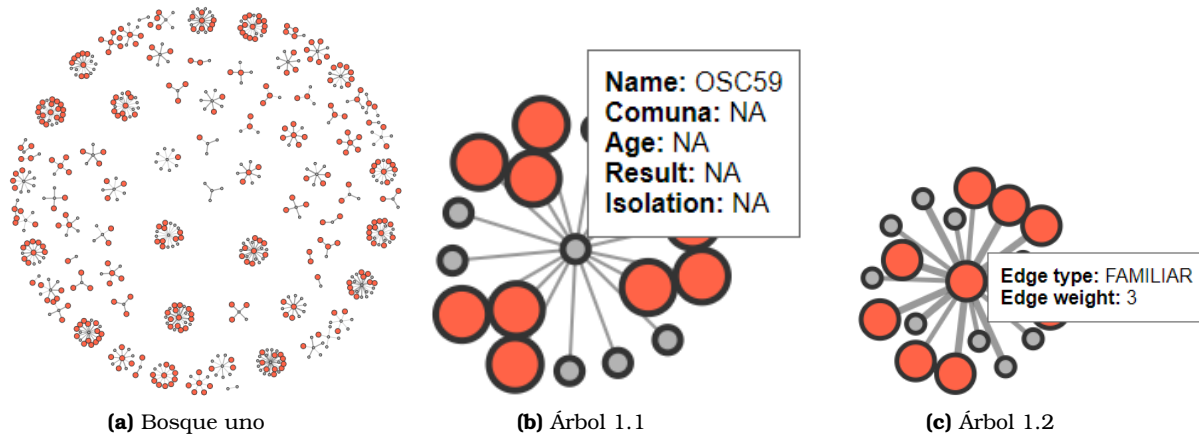


Figura 4.56 Visualización de algunos arboles estrella del primer bosque

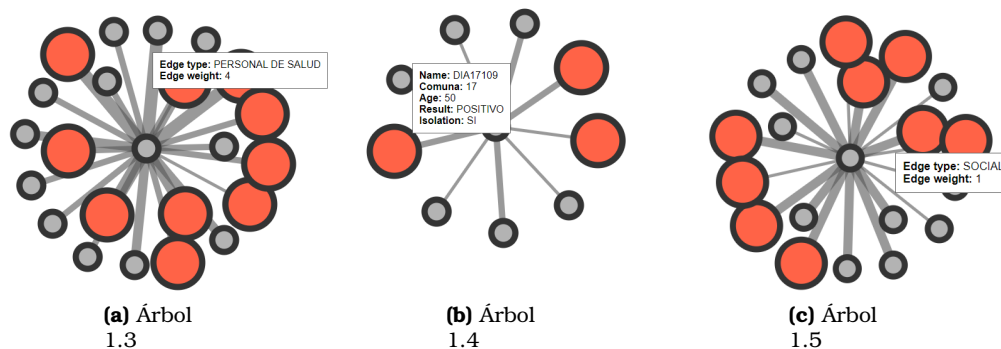


Figura 4.57 Visualización de algunos arboles estrella del primer bosque

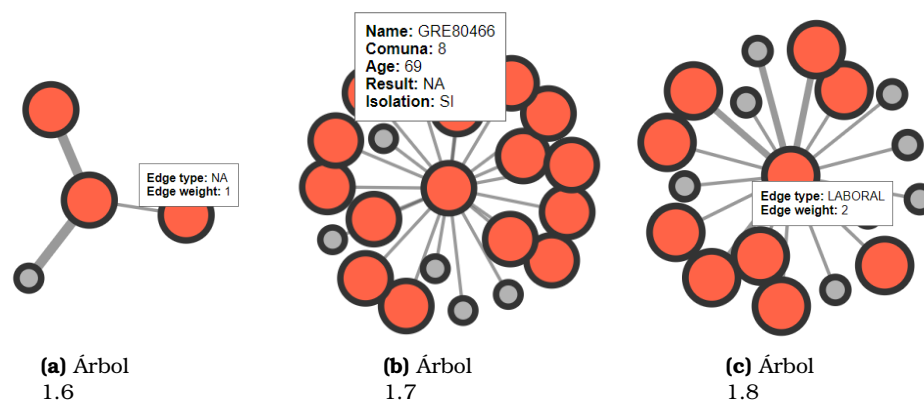


Figura 4.58 Visualización de algunos arboles estrella del primer bosque

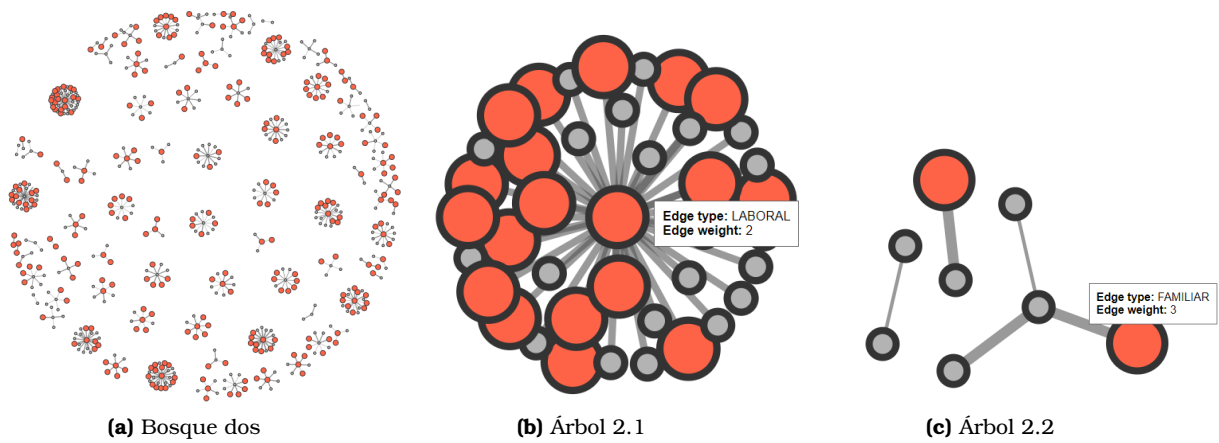


Figura 4.59 Visualización de algunos arboles estrella del segundo bosque

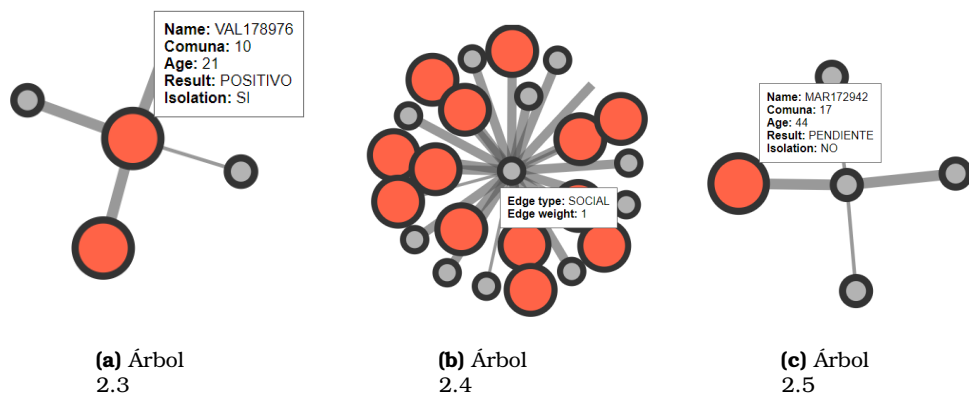


Figura 4.60 Visualización de algunos arboles estrella del segundo bosque

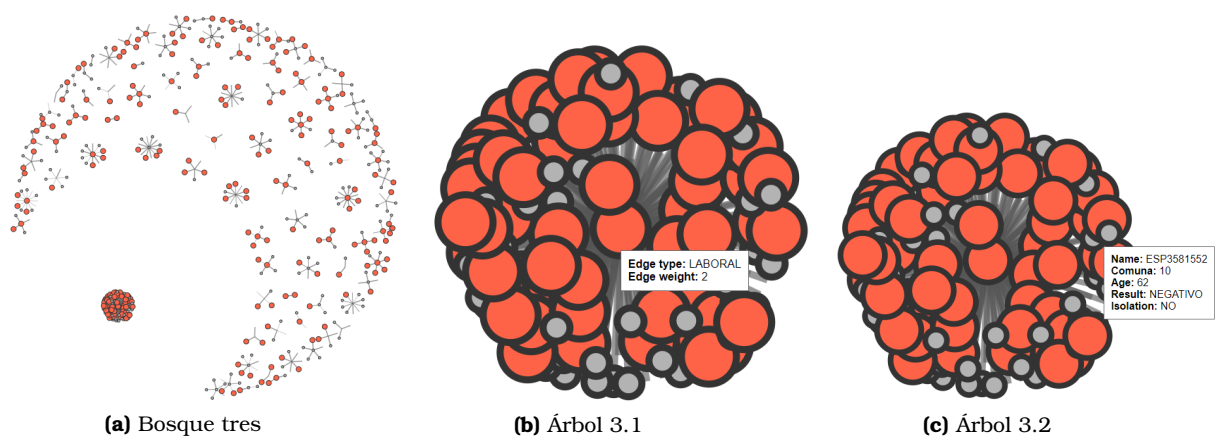


Figura 4.61 Visualización de algunos arboles estrella del segundo bosque

La facilidad de los grafos para encontrar centralidades, importancia de los actores (vértices) y peso de los enlaces. Permite cuantificar la estructura social en el espacio y tiempo, de esta forma es posible encontrar el grado de afección en la sociedad.

En las figuras anteriores (ver figuras de la 4.56 a la 4.61), se observan los tres bosquejos contruidos, el cual contienen los 262 arboles identificados. Algunos presentan mayor intensidad de contactos que otros. Gracias a los grafos, aparte de medir la intensidad o cantidad de contactos, es posible identificar la forma de contacto que existen entre vértices, el peso del enlace y el vértice central u origen. Con esta información los expertos pueden entender como la enfermedad afecta a los círculos sociales, laborales y familiares de los individuos, la evolución y zonas de mayor exposición, identificar aspectos de mejoramiento en la toma de decisiones para evitar la celeridad del contagio, crear un sistema de alerta con los posibles casos positivos (manteniendo el derecho a la privacidad), identificar enlaces con mayor probabilidad de contagio dependiendo de las circunstancias, entre otros aspectos importantes para el control de esta enfermedad y otras de características similares. (Ver los grafos completos en el siguiente link, tener en cuenta que se deben descargar y abrir por medio de Google Chrome. https://drive.google.com/drive/folders/1MD0g29qDl2oVHLPiJSKa6gM_JrQqfEbi?usp=sharing)

CAPÍTULO 5

Conclusiones, limitaciones y recomendaciones

*DORMÍA y soñaba que la vida
era bella; desperté y advertí que la vida era deber.*

Immanuel Kant.
1724–1804

5.1. Conclusiones

Esta enfermedad que ha afectado a casi todos los países del mundo ha sido muy compleja mitigarla o controlarla debido a su forma de contagio y su rápida propagación entre las personas. Es por esto que se decidió aportar un enfoque estadístico basado en algunos estudios realizados en Italia, China y demás países mencionados en los antecedentes de este trabajo. Donde, con ayuda de las técnicas estadísticas espaciales, es posible tener un panorama mejor sobre los patrones que esta presentando la enfermedad en diferentes periodos de tiempo en la ciudad de Cali.

Con los datos suministrados por la secretaria de salud de Santiago de Cali, se logró abordar la problemática que ha presentado la enfermedad COVID-19 desde un punto de vista espacial y temporal. Inicialmente, se pensó dividir en tres periodos temporales de la enfermedad, -Inicio de la pandemia en Cali; Inicio del primer pico y Final del primer pico de la pandemia-. Pero, dada la alta tasa de incidencias se optó por agrupar los periodos de acuerdo a las semanas epidemiológicas registradas, así, se pasó de tener tres periodos de tiempo a 16 periodos.

Adicionalmente, para tener una perspectiva individual de los contagios de acuerdo a la teoría de matemáticas discretas y teoría de grafos, se decidió presentar unos grafos de tipo árbol estrella. Dado que, se centra en el individuo que ya esta categorizado como

positivo y mediante una encuesta realizada en la IPS donde se notificó la condición se genera la red de contactos de cada uno de estos individuos.

Se evidenció, que los 16 periodos presentaron un patrón agregado, teniendo agrupaciones en diferentes zonas de Cali, los grupos que sobresalieron por su mayor incidencia -sobre todo a partir del quinto periodo- fueron las zonas centro, oriente, sur-oriente y nor-oriente de la ciudad. Estos resultados, son coherentes con las características de contagio de la enfermedad y la densidad de población que existe en estas zonas. Es decir, que en zonas donde se presentaban mayor aglomeración de personas o donde las características socio-demográficas inciden en diferentes decisiones de los individuos, la propagación de la enfermedad resultaba ser más rápida.

Las redes de contacto construidas, podrían ser una herramienta practica e informativa para identificar nuevos posibles casos de COVID-19 u otras enfermedades donde las características de contagio sean similares. En estos grafos, resulta fácil identificar el vértice central y las conexiones que este crea a través del tiempo, además, de tener información importante como el tipo de contacto que tuvo con la persona, si ha estado en aislamiento, barrio, si se ha realizado la prueba, etc.

La información en conjunto (tanto la espacio-temporal y la de los grafos), aportaría a nuevos estudios de caracter epidemiológicos. Es decir, la secretaria de salud publica podría definir nuevas variables de estudio como estilo de vida, salud mental, datos genéticos, historial clínico, entre otras variables que los expertos crean relevantes tanto para los casos ya diagnosticados, como para los casos sospechosos encontrados en las redes de contacto. Se podría segmentar las población de acuerdo a estas nuevas variables, encontrar los tipos de enlaces entre cada grupo y lograr identificar previamente las zonas e individuos mas vulnerables a esta enfermedad.

Adicionalmente, se podrían evaluar aspectos sociales y psicológicos en las poblaciones que mayormente fueron afectadas por la enfermedad, causas de contagio, factores claves que influyeron en las zonas con mayor densidad de contagios, el manejo de los casos aislados o asintomáticos, etc. Finalmente, se podrían plantear estudios a nivel genético que facilitaría la comprensión de la evolución del virus antes y después del contagio, mecanismos nuevos de propagación, variantes del virus, afectaciones en el cuerpo humano posteriores al contagio, anticuerpos generados por el cuerpo y otras variables, que podrían servir a la hora de replantear acciones de respuesta en salud pública para gestionar casos y reducir el potencial de propagación.

5.2. Limitaciones

Cuando la enfermedad del COVID-19 llego a Colombia y específicamente a la ciudad de Cali, las entidades gubernamentales no tenían una metodología clara para la obtención de datos de las tasas de incidencia de esta enfermedad, los protocolos de muestras

rápidas no estaban implementadas, entre otros aspectos. Debido a esto, se generó un subregistro de casos posiblemente de más de dos semanas. Este subregistro fue una limitación en los resultados, porque es muy probable que estos se encuentren desfasados por días o semanas.

Para el caso de la construcción de los grafos, la proporción de la tasa de no respuesta de 0.7 aproximadamente, en la variable "Toma de prueba", probablemente limitó el resultado final de la investigación. Algunos individuos, se negaban a dar información sobre los círculos sociales, familiares y laborales.

5.3. Recomendaciones

Es conveniente que en futuras investigaciones, se logren implementar grafos dinámicos, donde se puedan enlazar la evolución espacial y temporal de una determinada enfermedad. Como se observó anteriormente, las dos metodologías utilizadas -tanto la espacio-temporal como la de grafos- se abordaron de una forma individual y aunque fueron valiosos los resultados, una aplicación conjunta con visualizaciones dinámicas podría llegar a ser más informativo.

Adicionalmente, implementar una estandarización de encuestas y la digitalización de las mismas, ayudaría a optimizar tiempos en la generación de la información y los datos recopilados tendrían un margen de error menor. Lo anterior, sería un aporte para la implementación de un sistema de recolección de datos global organizado.

Finalmente, sería interesante en futuros trabajos combinar las metodologías utilizadas en este trabajo con metodologías enfocadas en la estimación del subregistro de casos. Esto ayudaría a minimizar sesgos a la hora de realizar una proyección o predicción de futuros casos de la enfermedad.

APÉNDICE A

Figuras de Tipos de Centralidades y Metodología CRISP-DM

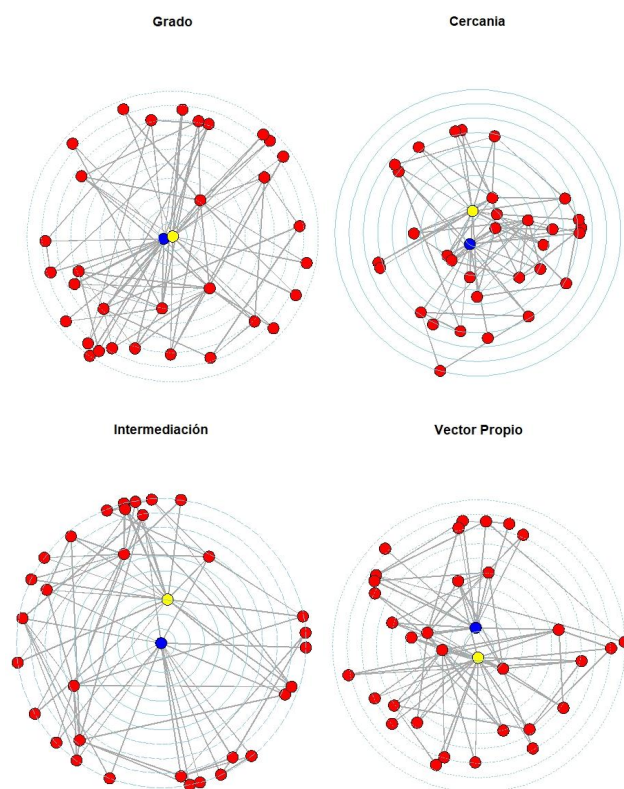


Figura A.1 [Kolaczyk and Csárdi \(2015\)](#). Se muestran las cuatro centralidades de vértices para una red de clubes de Karate. El vértice profesor es de color amarillo y el de instructor el azul

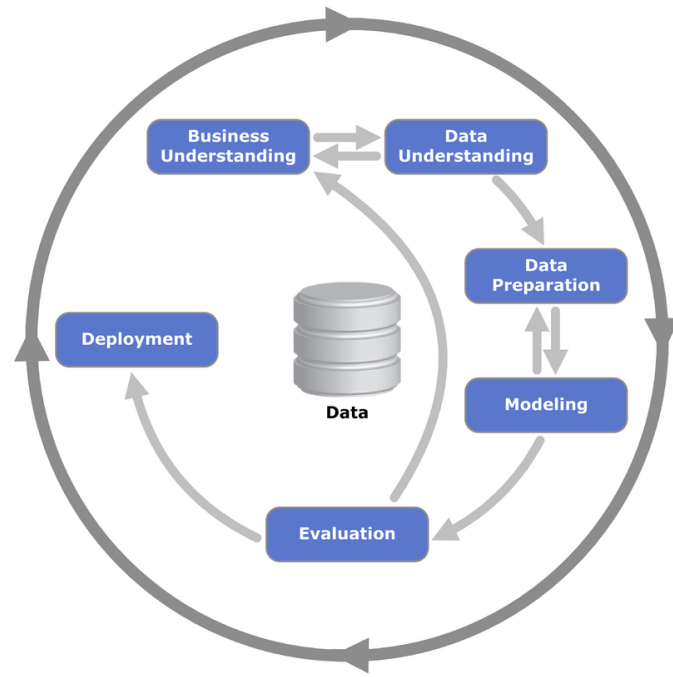


Figura A.2 <https://healthdataminer.com/data-mining/>. Metodología CRISP-DM

APÉNDICE B

Tablas

Tabla B.1 Lista de variables utilizadas en el análisis estadístico

Variable	Tipo	Escala	Descripción
fec_not	Cuantitativa continua	razón	Fecha registrada por el personal de salud del individuo confirmado positivo de COVID-19, el cual, reporta los primeros síntomas.
semanas	Cuantitativa discreta	razón	Semana epidemiológica
Confirmados	Cualitativa	nominal	Individuos positivos por COVID-19
edad	Cuantitativa discreta	razón	Edad del individuo confirmado
nacionalidad	Cualitativa	nominal	País de nacimiento del individuo
sexo	Cualitativa	nominal	Sexo del individuo
dirección	Cualitativa	nominal	Lugar de residencia del individuo positivo
estrato	Cualitativa	ordinal	Nivel socio-económico
País	Cualitativa	nominal	País de procedencia
x	Cuantitativa continua	intervalo	Latitud
y	Cuantitativa continua	intervalo	Longitud
Barrios_orig	Cualitativa	nominal	Barrio de residencia del individuo
Cod_Barrio	Cualitativa	nominal	Código del barrio

Tabla B.2 Lista de variables utilizadas para la construcción de las redes de contacto

Variable	Descripción
from	Vértice inicial, individuo confirmado positivo
to	Contactos que el individuo positivo menciona en la encuesta
type	Tipo de enlace entre cada uno de los individuos infectados y los contactos
weight	Peso de enlace
id	Identidad única de casa uno de los individuos y contactos
sexo	Genero de los individuos
sexo.type	Variable numérica dicotómica de identificación de sexo
type.comuna	Numero de comuna
Resultado	Estado de la prueba de detección del COVID-19 realizado a los contactos
Aislamiento	Variable dicotómica con respuesta S/N, referente a si estuvo asilado o no
Edad	Edad registrada de acuerdo a los datos suministrados por los individuos

Tabla B.3 Paquetes en R utilizados en el estudio

Paquete R	Descripción
Análisis descriptivo	
DMwR2	Paquete especializado en técnicas de minerías de datos.
funModeling	Especializado en la exploración y visualización de datos.
stringr-tidyverse	Muy practico para el tratamiento de datos. El paquete proporciona un conjunto cohesivo de funciones diseñadas para hacer que trabajar con cadenas sea lo más fácil posible
dplyr	Útil para la manipulación de data frame
visdat	Visualización preliminar de datos.
lubridate	Útil para el tratamiento de fechas en R.
Redes de contacto	
igraph	Paquete especializado para visualización y análisis de grafos.
ggraph	Extensión de ggplot2 para la visualización de grafos.
network	Paquete utilizado para la modificación de objetos de red.
ndtv	Paquete para generar grafos dinámicos.
Análisis Espacial	
ggplot2	Es un paquete del lenguaje R, utilizado para la visualización de todo tipo de gráficos.
spatstat	Se enfoca en el análisis espacial de patrones puntuales, principalmente en patrones de puntos de dos dimensiones, se utiliza para el análisis espacio temporal de los patrones puntuales. Se utiliza para el análisis de patrones puntuales.
maptools	Conjunto de herramientas para manipular datos geográficos. El paquete también proporciona una interfaz para intercambiar objetos espaciales con paquetes como 'PBSmapping', 'Spatstat.geom', 'maps' y otros. Se utiliza para leer el shapefile del mapa de Cali.
lubridate	Función para trabajar con 'fechas-horas' e 'intervalos de tiempo' rápido y fácil de usar en datos de fecha y hora, extracción y actualización de componentes de una fecha y hora. Se utiliza para dividir los periodos de tiempo en semanas epidemiológicas con la función epiweek.

Bibliografía

- R. P. Aguila. *Una introducción a las matemáticas discretas y teoría de grafos*. El Cid Editor, 2013.
- W. P. Aylward, Bruce (WHO); Liang. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). *The WHO-China Joint Mission on Coronavirus Disease 2019*, 2019(February):16–24, 2020. URL <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>.
- A. Baddeley, E. Rubak, and R. Turner. *Spatial point patterns: methodology and applications with R*. CRC press, 2015.
- A. Baddeley et al. Analysing spatial point patterns in r. In *Workshop notes version*, volume 3, 2008.
- M. Berman and P. Diggle. Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):81–92, 1989. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1989.tb01750.x.
- R. S. Bivand, E. J. Pebesma, V. Gomez-Rubio, and E. J. Pebesma. *Applied spatial data analysis with R*, volume 2. Springer, 2013.
- U. Brandes and T. Erlebach. *Network Analysis - Methodological Foundations*, volume 3418. Springer Science & Business Media, 2005. ISBN 3540709592.
- P. Britos. Objetivos de negocio y procesos de minería de datos basados en sistemas inteligentes. *Reportes Técnicos en Ingeniería del Software*, 7(1):26–29, 2005.
- Y. B. Caballero. Test de aleatoriedad para procesos puntuales espaciales basado en el cálculo de la dimensión fractal. *Universidad Nacional de Colombia Facultad de Ciencias Departamento de Estadística*, 2017.
- A. W. H. Chin, J. T. S. Chu, M. R. A. Perera, K. P. Y. Hui, H.-L. Yen, M. C. W. Chan, M. Peiris, and L. L. M. Poon. Stability of SARS-CoV-2 in different environmental conditions. *The Lancet Microbe*, pages 0–4, 2020. ISSN 26665247. doi: 10.1016/s2666-5247(20)30003-3.
- E. De Jonge and M. Van Der Loo. *An introduction to data cleaning with R*. Statistics Netherlands Heerlen, 2013.
- M. Durbán Reguera. An introduction to smoothing with penalties: P-splines. *BEIO, Boletín de Estadística e Investigación Operativa*, 25(3):195–205, 2009. ISSN 1889-3805.

- D. Elías-Cuartas, D. Arango-Londoño, G. Guzmán-Escarria, E. Muñoz, D. Caicedo, D. Ortega-Lenis, A. Fandiño-Losada, J. Mena, M. Torres, L. Barrera, et al. Análisis espacio-temporal del sars-cov-2 en cali, colombia. *Revista de Salud Pública*, 22:1–6, 2020.
- R. Giraldo. Estadística Espacial - Notas de Clase. *Universidad Nacional de Colombia*, page 136, 2011.
- R. Giraldo h. Introducción a la geoestadística: Teoría y Aplicación. *Universidad Nacional de Colombia*, page 91, 2002.
- D. Giuliani, M. M. Dickson, G. Espa, and F. Santi. Modelling and Predicting the Spatio-Temporal Spread of Coronavirus Disease 2019 (COVID-19) in Italy. *SSRN Electronic Journal*, 2019:1–17, 2020. doi: 10.2139/ssrn.3559569.
- V. González-jaramillo, N. González-jaramillo, C. Gómez-restrepo, C. A. Palacio-acosta, A. Gómez-lópez, and O. H. Franco. Proyecciones de impacto de la pandemia COVID-19 en la población colombiana, según medidas de mitigación. Datos preliminares de modelos epidemiológicos para el periodo del 18 de marzo al 18 de abril de 2020. *Revista de Salud Pública.*, 22:1–6, 2020. doi: <https://doi.org/10.15446/rsap.V22.85789>.
- E. Grupo, L. Oms, M. Oriente, F. Few, and E. Ffx. FOCO TÉCNICO : Investigaciones epidemiológicas y clínicas precoces sobre el COVID-19 para una respuesta de salud pública. pages 2–3, 2019.
- A. C. Guidoum. Kernel Estimator and Bandwidth Selection for Density and its Derivatives: The kedd Package. pages 1–22, 2020. URL <http://arxiv.org/abs/2012.06102>.
- G. Hatfield. Spatial statistics. *Practical Mathematics for Precision Farming*, (2):75–104, 2018. doi: 10.2134/practicalmath2016.0102.
- R. Huang, M. Liu, and Y. Ding. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *Journal of Infection in Developing Countries*, 14(3):246–253, 2020. ISSN 19722680. doi: 10.3855/jidc.12585.
- R. James, G., Witten, D., Hastie, T., Tibshirani. *An Introduction to Statistical Learning - with Applications in R* | Gareth James | Springer. 2013. ISBN 9781461471370. doi: 10.1007/978-1-4614-7138-7. URL <https://www.springer.com/gp/book/9781461471370> <http://www.springer.com/us/book/9781461471370>.
- H. S. Janine Illian, Antti Penttinen and D. Stoyan. Statistical analysis and modelling of spatial point patterns. *John Wiley Sons*, page vol. 70, 2008.

- M. J. Keeling and K. T. Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005. ISSN 17425662. doi: 10.1098/rsif.2005.0051.
- R. S. Kirby, E. Delmelle, and J. M. Eberth. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*, 2017. ISSN 1047-2797. doi: 10.1016/j.annepidem.2016.12.001. URL <http://dx.doi.org/10.1016/j.annepidem.2016.12.001>.
- E. Kolaczyk. *Statistical analysis of network data - Methods and models*. 2009. ISBN 9780387881454. doi: 10.1007/978-0-387-88146-1. URL <http://link.springer.com/content/pdf/10.1007/b98971.pdf>.
- E. D. Kolaczyk and G. Csárdi. *Statistical Analysis of Network Data with R*, volume 66. 2015. ISBN 9781493909827. doi: 10.18637/jss.v066.b01. URL <http://www.springer.com/series/6991>.
- A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, and S. Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 3099(20):1–7, 2020. ISSN 14744457. doi: 10.1016/S1473-3099(20)30144-4.
- Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2):1–4, 2020. ISSN 17088305. doi: 10.1093/jtm/taaa021.
- D. A. Luke. *A User's Guide to Network Analysis in R*. 2015. ISBN 978-3-642-38950-4. doi: 10.1007/978-3-642-38951-1. URL <http://link.springer.com/10.1007/978-3-642-38951-1>.
- E. Ministerio de Sanidad. Información Científica-Técnica. Enfermedad por Coronavirus, COVID-19. page 37, 2020.
- J. Ohser. On estimators for the reduced second moment measure of point processes. *Series Statistics*, 14(1):63–71, 1983. doi: 10.1080/02331888308801687. URL <https://doi.org/10.1080/02331888308801687>.
- B. Quispe Quispe. Modelos estadísticos en procesos puntuales espaciales poisson para evaluar la distribución espacial de los hechos delictivos en lima, Perú. 2016.
- R. Ranjan and R. Misra. Epidemic disease propagation detection algorithm using MapReduce for realistic social contact networks. *2014 International Conference on High Performance Computing and Applications, ICHPCA 2014*, (ii), 2015. doi: 10.1109/ICHPCA.2014.7045325.

- B. Ripley. *Statistical Inference for Spatial Processes*. Cambridge University Press, 1988. ISBN 9780521352345. URL <https://books.google.com.co/books?id=Ah90QgAACAAJ>.
- F. J. Rodríguez-cortés. Tema 3. ¿qué es la intensidad de un patrón de puntos y porqué usarla? In *Escuela de Estadística, Univeridad Nacional de Colombia, Medellín, Colombia*, 2018.
- O. Schabenberger and C. A. Gotway. *Statistical methods for spatial data analysis*. CRC press, 2017.
- C. Scientifique. Risque zoonotique du SARS-CoV2 (Covid-19) associé aux animaux de compagnie : infection de l ’ animal vers l ’ homme et de l ’ homme vers l ’ animal. pages 1–21, 2020.
- Stacey Knobler, Adel Mahmoud, Stanley Lemon, Alison Mack, Laura Sivitz and K. Oberholtzer. *Learning from SARS: Preparing for the Next Disease Outbreak Workshop Summary*. 2004. ISBN 0309594332. URL <http://www.nap.edu/catalog/10915.html{%}0AVisit>.
- K. Sun, J. Chen, and C. Viboud. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *The Lancet Digital Health*, 2(4):e201–e208, 2020. ISSN 25897500. doi: 10.1016/S2589-7500(20)30026-1.
- M. Vidaurre-Arenas, R. Martínez-Piedra, A. d. S. OPS, et al. Boletín epidemiológico.
- L. A. Waller and C. A. Gotway. Applied spatial statistics for public health data. *John Wiley Sons*, page vol. 368, 2004.
- L. Wang, G. Wang, L. Gao, X. Li, S. Yu, M. Kim, Y. Wang, and Z. Gu. Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in the United States. pages 1–26, 2020. URL <http://arxiv.org/abs/2004.14103>.
- C. K. Wikle, A. Zammit-Mangion, and N. Cressie. *Spatio-temporal Statistics with R*. CRC Press, 2019.
- Xiaoyi, F. Wei, Huang, L. Hu, L. Wen, and K. Chen. Epidemiology and Clinical Characteristics of COVID-19. 23(April):268–271, 2020. doi: 10.34172/aim.2020.09.
- L. Zhao, J. Chen, F. Chen, F. Jin, W. Wang, C. T. Lu, and N. Ramakrishnan. Online flu epidemiological deep modeling on disease contact network. *GeoInformatica*, 2019. ISSN 13846175. doi: 10.1007/s10707-019-00376-9.