

Análisis visual multifactor de la propagación espacio-temporal de los casos de Covid-19 y su relación con factores sociodemográficos

John Edson Sanchez Chilo

Universidad Nacional de San Agustín

Keywords:

Abstract

This study presents a comprehensive visual framework to analyze the spatiotemporal evolution of COVID-19 cases in relation to key sociodemographic factors across different regions. The proposed approach integrates multiple data visualizations, enabling simultaneous spatial and analytical interpretation, which enhances evidence-based decision-making. The methodology involves segmenting case data into temporal periods such as quarters and weeks, extracting various statistical metrics, and developing interconnected modules that facilitate personalized analyses for different objectives. Additionally, a socio-demographic data matrix is constructed to identify relevant variables impacting disease spread. The findings demonstrate that combining spatial, temporal, and sociodemographic data through interactive visualizations offers valuable insights into patterns of disease dissemination and socio-economic vulnerabilities. This tool supports public health strategies by enabling stakeholders to explore complex relationships in a user-friendly and dynamic environment.

1 Introducción

La enfermedad por coronavirus 2019 (COVID-19), causada por el virus SARS-CoV-2, emergió a finales de 2019 en Wuhan, China, y rápidamente se convirtió en una pandemia global que afectó todos los aspectos de la vida humana. Su rápida propagación [1], y la falta de preparación para afrontar la pandemia por parte de los gobiernos y las autoridades [2], pusieron a prueba los sistemas de salud, economía y la sociedad en su conjunto, afectando especialmente a ciertos sectores de la población.

La pandemia, evidenció las profundas desigualdades estructurales presentes en muchas regiones. Las poblaciones que ya enfrentaban condiciones sociales y económicas precarias—como acceso limitado a servicios de salud, empleo, mayor densidad poblacional, o mayor número de personas de edad avanzada— fueron particularmente vulnerables a una mayor propagación del virus [3]. En muchos casos, la ausencia de información clara sobre qué factores sociodemográficos influían con mayor peso en la dinámica de contagio dificultó la toma de decisiones efectivas por parte de las autoridades. Esta falta de conocimiento impidió aplicar medidas focalizadas que protegieran a los sectores más expuestos, limitando así la capacidad

de respuesta y aumentando el impacto del virus en comunidades ya desfavorecidas.

Por ello, la integración de datos sociodemográficos es fundamental para comprender mejor la dinámica de propagación de una epidemia o pandemia, ya que factores como la densidad poblacional, la edad, el nivel educativo, el ingreso económico y el acceso a servicios de salud influyen significativamente en la vulnerabilidad y exposición de las personas al virus como el caso de [4]. La relación entre estas variables y la incidencia de casos varía según el contexto local y las desigualdades estructurales presentes [5], por lo que es esencial desarrollar estrategias de salud pública basadas en una comprensión integral de estas dinámicas sociales complejas.

La disponibilidad de datos sobre casos de epidemias por fecha y localidad ha permitido una mayor comprensión del desarrollo de la pandemia en diferentes territorios [6]. Gracias a los sistemas de vigilancia epidemiológica y al registro continuo de casos, es posible realizar análisis temporales y espaciales que revelan cómo se propaga un virus en función de distintos contextos sociales. Sin embargo, esta abundancia de información también presenta importantes desafíos, ya que el volumen y la variedad de los datos dificultan su organización, procesamiento e interpretación efectiva.

Además, el análisis de la propagación de los casos de una epidemia o pandemia se ve notablemente dificultado por la amplia variedad de variables sociodemográficas involucradas, las cuales interactúan de manera dinámica y no lineal. Esta alta dimensionalidad de datos complica la detección de tendencias claras y la interpretación visual de los resultados usando métodos convencionales [7], especialmente cuando se busca mantener la coherencia temporal y geográfica de los datos. A pesar de su importancia, son aún limitados los estudios que integran de forma efectiva datos sociodemográficos con registros epidemiológicos mediante herramientas computacionales y visuales robustas [8], lo que limita la capacidad de los expertos en la toma de decisiones para generar respuestas informadas y focalizadas.

Para abordar este problema, se propone desarrollar una herramienta visual que permita analizar la propagación de los casos de una enfermedad a lo largo del tiempo y el espacio, en relación con los factores sociodemográficos de cada localidad. Esta herramienta facilitará la exploración y visualización de patrones espaciotemporales y sociodemográficos por localidad, además de permitir una interacción dinámica y focalizada con el usuario.

2 Trabajos Relacionados

En los últimos años, el análisis visual de datos se ha consolidado como una herramienta fundamental para comprender fenómenos complejos en el ámbito de la salud pública, especialmente en escenarios de propagación de enfermedades infecciosas. Diversos estudios han explorado la integración de datos espaciotemporales, movilidad humana y variables socio-demográficas para identificar patrones de transmisión y vulnerabilidad en distintas comunidades. Esta sección revisa investigaciones previas que han propuesto métodos de visualización aplicados a epidemias, y mas específicamente a conjuntos de datos de COVID-19, algunas como herramientas interactivas y otras con asitencia inteligente, todo ello con el objetivo de identificar sus aportes, limitaciones y oportunidades de mejora que fundamentan la presente propuesta.

A inicios del año 2020, la aparición del COVID-19 y su rápida propagación a nivel mundial, como lo evidencia el estudio de Park et al. [1], impulsaron a la comunidad científica a enfocar con mayor intensidad sus esfuerzos en el análisis epidemiológico. Riccaboni y Verginer [8] concluyeron que los términos médicos relacionados con COVID-19 experimentaron un incremento promedio de 6.5 veces en la producción científica desde el inicio de la pandemia, lo cual refleja un notable aumento en el interés por investigar enfermedades infecciosas y sus dinámicas de propagación.

En este contexto, el estudio de Rydow et al. [7] demostró que el análisis visual desempeña un papel clave en el estudio epidemiológico, al facilitar la interpretación de modelos complejos y apoyar la toma de decisiones basada en datos.

En las primeras etapas del desarrollo de herramientas de visualización aplicadas a estudios epidemiológicos, era común el uso de representaciones gráficas simples, como mapas de calor [9], series temporales [10], o gráficos compuestos de múltiples vistas [11]. Si bien estas visualizaciones permitían comunicar ciertos aspectos generales de la evolución de la enfermedad, su capacidad para representar relaciones complejas entre variables espaciales, temporales y socio-demográficas era limitada.

Los estudios sobre enfermedades infecciones se apoyan principalmente en dos tipos de datos, los datos espacio-temporales y los datos multidimensionales. Los primeros permiten capturar la evolución y distribución geográfica de las infecciones a lo largo del tiempo, es decir la visualización geográfica [12, 13, 14] puede representar de forma efectiva, datos multivariantes en un espacio y tiempo específico.

Los segundos integran múltiples variables clínicas, epidemiológicas y ambientales para ofrecer una perspectiva más compleja y completa del fenómeno. Podemos incluir datos sobre movilidad urbana [15, 16], factores económicos [17], sociodemográficos [18, 19] y otros más.

En el contexto específico de la visualización de datos relacionados con la pandemia de COVID-19, debido a la rápida y extensa propagación del virus a nivel global, surgió la necesidad de desarrollar múltiples tipos de gráficos y representaciones visuales que permitan analizar y comunicar eficazmente su impacto en distintas naciones [20] [21] [22].

Además de los esfuerzos gubernamentales, instituciones

académicas han aportado significativamente al desarrollo de herramientas visuales para el seguimiento de la pandemia. Por ejemplo, la Universidad Johns Hopkins creó un dashboard interactivo ampliamente reconocido para visualizar en tiempo real la evolución global del COVID-19 [23]. Paralelamente, para facilitar el acceso y análisis de datos, también pusieron a disposición un repositorio público en GitHub que recopila y actualiza constantemente la información epidemiológica [24].

En la investigación relacionada con datos de COVID-19, comúnmente se manejan dos tipos principales de fuentes de datos, el primer tipo corresponde a datos directamente vinculados con la enfermedad, que incluyen información objetiva y cuantificable como el número de casos confirmados, las tasas de recuperación y las tasas de mortalidad [25, 14]. Estos datos son fundamentales para el seguimiento epidemiológico y permiten evaluar la evolución y gravedad del brote en diferentes regiones. Por otro lado, existe un segundo tipo de datos, denominados indirectamente vinculados, que no reflejan factores objetivos directamente relacionados con la infección, pero que capturan aspectos comunitarios y sociales derivados de la pandemia. Entre estos se encuentran indicadores relacionados con los impactos sociales, como cambios en el comportamiento poblacional o en la movilidad [15, 16], así como los efectos económicos y financieros ocasionados por la crisis sanitaria [17] o también los datos sociodemográficos de la población [18]. Aunque estos datos no están directamente conectados con la transmisión o la incidencia del virus, resultan cruciales para comprender el alcance total de la pandemia y para diseñar estrategias integrales de respuesta que consideren tanto la salud pública como las dimensiones sociales y económicas.

3 Análisis de Requerimientos

Después de revisar estudios previos para identificar los principales desafíos en el análisis de los datos. Se destacó la necesidad de herramientas que simplifiquen la representación, selección y análisis de la información, permitiendo estudiar las diferentes tendencias, estacionalidades y variaciones entre periodos. Asimismo, las tareas fueron complementadas con referencias a trabajos previos, resaltando la importancia de acelerar el estudio de la propagación multifactor de casos Covid-19 y su relación con factores sociodemográficos, de las cuales se seleccionaron cuatro tareas que se pueden realizar en la herramienta de análisis propuesta:

- **R1. Análisis de Comportamiento Regional:** El sistema de visualización debe ser capaz de identificar y agrupar regiones con patrones de comportamiento epidemiológico similares. Para cada grupo, se debe calcular y mostrar de manera visual las medidas de tendencia central (media, mediana) y de dispersión (mínimo, máximo, total) de los casos epidemiológicos. Además, se deben correlacionar y mostrar los factores sociodemográficos promedio que caracterizan a cada grupo de regiones.
- **R2. Identificación y Comparación de Factores Sociodemográficos Clave:** El sistema debe identificar los factores

sociodemográficos más influyentes en la dinámica epidemiológica de cada grupo de regiones. Para facilitar el análisis, el usuario podrá comparar visualmente estos factores clave de un grupo con los de otros grupos, permitiendo entender las diferencias y similitudes en sus perfiles sociodemográficos.

- **R3. Visualización de la Propagación Espacio-Temporal de Casos:** El sistema debe visualizar la evolución de la propagación epidemiológica en un mapa a lo largo del tiempo. Las regiones se deben colorear según su pertenencia a un grupo (clúster) definido previamente. La densidad de casos en cada región se representará visualmente, identificando de manera automática y destacada los puntos de brote, zonas de riesgo y alertas en diferentes periodos.
- **R4. Control de Parámetros y Exploración Temporal:** El sistema debe proveer un panel de control interactivo que permita al usuario seleccionar y configurar el conjunto de datos de trabajo y los factores sociodemográficos para el análisis. Adicionalmente, el panel debe ofrecer herramientas para la exploración temporal, permitiendo la reproducción automática de la evolución de los casos o el control manual a través de una línea de tiempo.

4 Descripción general del sistema

La figura 1 presenta el flujo del presente trabajo. Es una aplicación web la cual esta dividida en seis módulos: Datos(Análisis Exploratorio de los mismos), Preprocesamiento, Modelo para Generar Datos de Valor, Análisis de Datos adicional, Visualización y las Interacciones que cumplen con los requerimientos del trabajo,

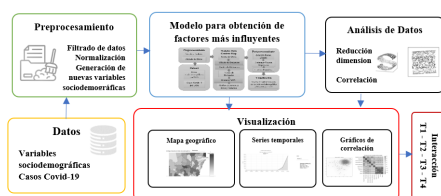


Figure 1. Flujo de Trabajo para el sistema

4.1 Dataset

El conjunto de datos utilizado en este estudio contiene los casos de Covid-19 encontrados dentro de la región de Nueva Gales del Sur, en Australia, y abarca el período desde el 25 de enero de 2020 hasta el 07 de febrero de 2022, siendo un total de 973412 casos. Los datos fueron recopilados por el Ministerio de Salud de Nueva Gales del Sur y están disponibles en la página de Salud del Gobierno [26]. Este conjunto de datos incluye casos Covid-19 contienen la fecha, el código del distrito local de salud (LHD) y el código del Área de Gobierno Local (LGA) siendo un total de 128 en Nueva Gales del Sur.

El segundo conjunto contiene datos sociodemográficos y económicos a nivel de Local Government Areas (LGAs) como Edad Media, Ingreso Medio por Persona, Ingreso Medio por Familia, Costo Medio de Hipoteca, Costo Medio de Renta, Población masculina o femenida por edades, Habitaciones por Familia, Medios de Transporte usados, Porcentaje de Personas que viven solas, y otros más, siendo un total de 128 datos sociodemográficos por LGA, extraídos del censo de 2016, las cuales serán usadas con la intención de analizar características comunitarias y posiblemente correlacionarlas con los patrones de contagio.

4.2 Preprocesamiento

El preprocesamiento es una fase crítica en el análisis de datos, ya que prepara los datos brutos para su análisis posterior. Esta etapa asegura que los datos sean limpios, coherentes y adecuados para los métodos analíticos que se aplicarán.

4.3 Limpieza de Datos

El conjunto de datos de casos Covid-19 presenta registros de casos sin la ubicación específica del LGA donde fue encontrado, en este caso datos sin ubicación espacial no representan ningún aporte al estudio, por lo cual serán eliminados completamente, ya que al ser ubicaciones tampoco no se pueden aplicar técnicas de imputación u otras, además que solo representan el 4% del total de registros por lo que no tienen mayor relevancia.

El conjunto de datos de datos sociodemográficos abarca los 128 LGAs presentes en Nueva Gales del Sur, adicionalmente con una área no incorporada en ninguna jurisdicción, por lo cual será ignorada y no utilizada en el estudio.

Tampoco se trabajaron con diferentes áreas de gobierno local las cuales presentan datos sociodemográficos incoherentes, o que no los tengan completos, esto con el fin evitar problemas con los modelos al usar datos incompletos, además que en este caso tampoco se puede realizar ningún método de imputación al tratarse de áreas de gobierno local totalmente independientes, por lo cual un acercamiento geográfico por los vecinos no sería de gran ayuda y se prefiere mantener solo los datos completos.

4.3.1 Agrupación de datos

Se tiene como datos casi un millón de casos de Covid-19 encontrados en las diferentes áreas de gobierno local en Nueva Gales del Sur, esta cantidad inmensa de datos, no puede ser visualizada directamente, por lo cual se procederá agrupar mediante un conteo de casos por mes y área de gobierno local, de esta manera se busca reducir datos innecesarios y solo mantener un resumen de ellos, el cual podrá ser usado a posterior por el modelo y la visualización.

Se hicieron agrupaciones de casos por trimestre, de manera que los tres años de datos se puedan dividir y abarcar un total de 12 periodos, de los cuales se halló métricas de máximo y mínimo número de casos, media y mediana. Aparte también se hicieron subdivisiones de los trimestres en semanas, y de igual forma de hallaron las métricas anteriormente mencionadas para

cada trimestre. Todo ello se hizo con el fin de obtener los datos listos para ser representados mas adelante en la serie temporal.

4.4 Modelo

El esquema de la propuesta, se estructura en módulos interconectados diseñados para abordar de manera integral el análisis evolutivo espacio-temporal de los casos de Covid-19 y su relación con factores sociodemograficos. Cada módulo incluye visualizaciones específicas que se adaptan a diferentes objetivos analíticos, permitiendo un enfoque personalizado según la tarea a realizar.

4.5 Data Context Map: Modelo para datos sociomográficos mas relevantes de locaciones similares

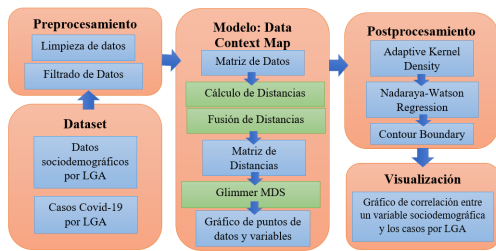


Figure 2. Modelo para la obtención de factores mas relevantes

4.5.1 Matriz de Datos

A partir del conjunto de datos sociodemograficos, se crea una matriz de datos $X \in R^{m \times n}$, donde:

- m es el número de registros, en este caso cada uno de los LGAs y sus respectivos datos, representan un registro.
- n es el número de variables sociodemográficas (edad, ingreso, movilidad, etc.).

La matriz de datos debe de tomar solo valores numéricos por lo que se ignoraran los campos de nombre o código relacionados al LGA, y solo se tomara en cuenta los 10 atributos sociodemográficos del LGA mas relevantes.

Adicionalmente se realizara una normalización de la matriz de datos con el fin de obtener valores entre 0 y 1, esto debido a que si mantenemos los valores como estan, las variables con datos numericos mas grandes como poblacion o area geografica, tendrian mayor influencia en el modelo, por sobre variables como costo de renta u otros, lo cual no es lo que buscamos.

4.5.2 Cálculo de Distancias

A partir de la matriz de datos, se debe conseguir la matriz de distancias, la cual esta compuesta por tres tipos de distancias:

1. **Distancia entre datos (Data-to-Data, DD):** se calcula la distancia euclidiana entre pares de registros, representa la

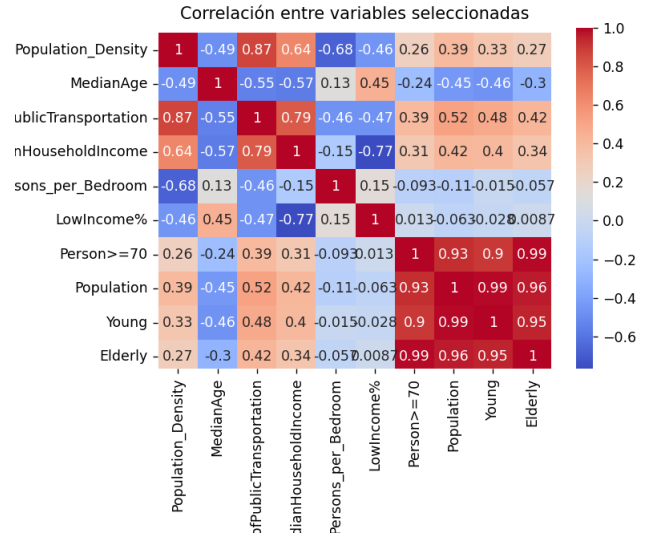


Figure 3. Grafico de correlación entre variables

similitud que hay entre LGAs.

$$DD_{ij} = \|X_i - X_j\|$$

2. **Distancia entre variables (Variable-to-Variable, VV):** se mide la disimilitud entre variables como:

$$VV_{ij} = 1 - |\text{correlación}(V_i, V_j)|$$

Esta métrica refleja el grado de independencia entre atributos.

3. **Distancia entre datos y variables (Data-to-Variable, DV):** se representa cada variable V_j como un vector unitario $e_j \in R^n$, y se calcula la distancia euclidiana desde cada registro D_i a dicho vector:

$$DV_{ij} = \|D_i - e_j\|$$

donde $e_j = [0, \dots, 1, \dots, 0]$ con 1 en la posición j .

4.5.3 Fusión de Distancias

Se integra toda la información en una única matriz de distancias compuesta $D \in R^{(m+n) \times (m+n)}$:

$$D = DDDVDV^T VV^T$$

4.5.4 Reducción de Dimensionalidad con MDS

Se aplica *Multidimensional Scaling* (MDS) sobre la matriz D para obtener una proyección bidimensional que preserve las distancias originales en la medida de lo posible. El resultado es un conjunto de coordenadas en R^2 para cada registro y variable.

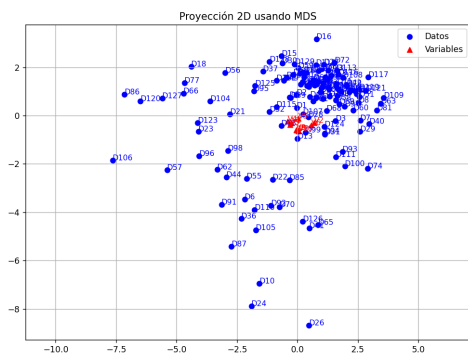


Figure 4. Grafico de puntos del GLimmer MDS

4.6 Técnicas de Visualización y Análisis Post-Proyección

Una vez obtenida la proyección bidimensional de los datos y variables mediante MDS, se aplicaron técnicas complementarias con el fin de mejorar la interpretación visual y apoyar la identificación de patrones locales en el espacio proyectado.

4.6.1 Adaptive Kernel Density Estimation (AKDE)

La *Adaptive Kernel Density Estimation* es una técnica no paramétrica que permite estimar la función de densidad de probabilidad subyacente a una distribución de puntos en un espacio 2D. En este trabajo, se utilizó AKDE para identificar regiones con alta concentración de registros, lo que sugiere la presencia de clústeres naturales.

A diferencia de un KDE tradicional con ancho de banda fijo, AKDE adapta localmente el ancho del kernel según la densidad del entorno: regiones densas utilizan kernels más estrechos, mientras que regiones dispersas emplean kernels más amplios. Esto permite una representación más precisa de la estructura espacial de los datos proyectados.

4.6.2 Regresión Kernel de Nadaraya-Watson

La regresión de Nadaraya-Watson es un estimador no paramétrico que aproxima una variable de interés en función de las coordenadas espaciales, utilizando una ponderación basada en funciones kernel. Para cada punto en la grilla 2D, el valor estimado se calcula como un promedio ponderado de los valores de los datos vecinos, donde la ponderación decrece con la distancia.

En este estudio, se aplicó esta técnica para generar mapas suavizados de variables sociodemográficas (por ejemplo, ingreso o educación) sobre el espacio proyectado, lo cual facilita la identificación de regiones con valores altos o bajos y permite interpretar cómo ciertas variables varían a lo largo del mapa de contexto.

4.6.3 Contornos de Densidad (Contour Boundary)

Los contornos de densidad son líneas que delimitan regiones de igual densidad dentro del espacio proyectado. Se obtienen

aplicando niveles de umbral sobre la superficie de densidad estimada (por ejemplo, mediante AKDE), generando fronteras visuales entre zonas de distinta concentración.

En el *Data Context Map*, estos contornos permiten visualizar con mayor claridad los límites aproximados de clústeres, así como zonas transicionales entre grupos. Esta técnica complementa el análisis al ofrecer una representación topográfica de la estructura local de los datos.

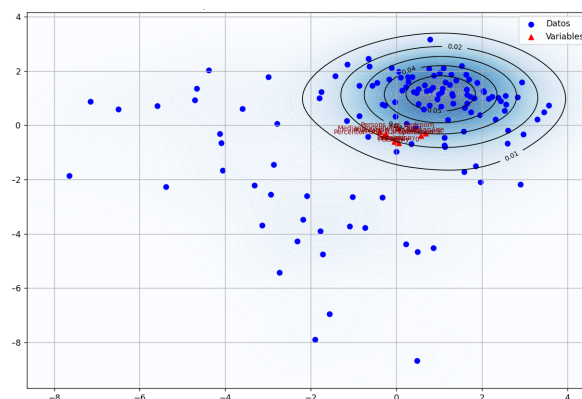


Figure 5. Grafico con AKDE + Contorno de Puntos

El *Context Map* resultante 5 permite interpretar:

- Grupos de registros similares, representados por puntos cercanos (clústeres).
- Variables cercanas a dichos grupos, interpretadas como atributos representativos.
- Correlaciones entre variables, visibles a través de su proximidad espacial.

Este enfoque permite analizar relaciones complejas entre datos y variables desde una perspectiva visual integrada, facilitando la detección de patrones relevantes en contextos sociodemográficos e identificando regiones con comportamiento similar (clúster).

4.7 Diseño visual

La herramienta visual 6 fue implementada para resolver los requerimientos previamente mencionados, los cuales ofrecen una exploración visual e interactiva a través de las diferentes secciones.

Para una mejor ilustración, aplicamos nuestro dataset pre-procesado sobre data de casos de Covid-19 en la región de Nueva Gales del Sur, además de añadir el conjunto de datos de sus variables sociodemográficas. La herramienta visual ha seguido cierto flujo de trabajo para generar cada uno de los gráficos, los cuales han sido detallados en la sección de modelo, y por lo mismo también cumplen con ofrecer al usuario la información que este necesita de la forma más sencilla y entendible posible.

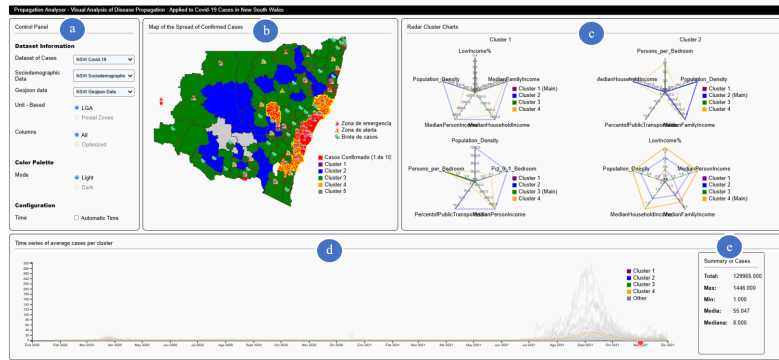


Figure 6. Herramienta de visualización para la propagación de casos epidemiológicos

4.7.1 Panel de Control

El Panel de Control, ubicado en la parte izquierda superior del dashboard, permite configurar los parámetros fundamentales del análisis de manera flexible e intuitiva. Su diseño facilita la selección y combinación de distintos conjuntos de datos, niveles geográficos y modos de visualización, lo que permite adaptar la exploración según las necesidades del usuario o del contexto epidemiológico.

A través de un conjunto de desplegables, es posible elegir el conjunto de datos epidemiológicos a utilizar, en este caso “NSW Covid-19”, junto con el dataset sociodemográfico correspondiente que incluye variables como ingresos, densidad poblacional o número de personas por habitación. Asimismo, se especifica el archivo de datos geoespaciales en formato GeoJSON, el cual contiene los límites administrativos utilizados para la representación geográfica. Esta integración de capas de datos permite enriquecer el análisis espacial con información de contexto social y económico.

El panel también incluye la opción de definir el nivel de agregación territorial sobre el que se visualizarán los datos, permitiendo seleccionar entre unidades basadas en LGA (Local Government Area) o zonas postales. En la configuración actual, se ha priorizado el uso de LGAs debido a su relevancia administrativa y disponibilidad de datos integrados. Adicionalmente, el usuario puede optar por incluir todas las columnas sociodemográficas o un subconjunto optimizado que reduce el número de variables sin sacrificar la riqueza del análisis multivariado.

Otro elemento configurable es la paleta de colores, que permite alternar entre un modo claro y un modo oscuro. Esto mejora la legibilidad según el entorno de trabajo o preferencia del analista. En la versión actual, se emplea el modo claro por defecto para facilitar la visualización.

Finalmente, se ofrece una opción de configuración temporal a través de la función “Automatic Time”. Esta permite activar una animación automática que recorre el comportamiento temporal de los casos confirmados en el mapa y la serie temporal, brindando una visión dinámica de la evolución del brote. Esta funcionalidad temporal es especialmente útil para identificar momentos críticos, observar la progresión del virus y evaluar retrospectivamente la efectividad de las medidas adoptadas en distintos periodos.

4.7.2 Mapa de propagación de Casos Confirmados

El mapa de propagación de casos Fig. 6(b), realiza una representación visual completa de el esparcimiento de casos por las diferentes regiones en los diferentes periodos de tiempo, es por ello que este mapa se complementa con ayuda del manejo de tiempo en la serie temporal Fig 1(d) con la cual uno puede avanzar o retroceder en el tiempo de forma manual, o dejarlo a libertad con el modo automático.

El mapa esta coloreado de manera que regiones con comportamiento similar de casos (las cuales representan un mismo cluster) tengan el mismo color, y con ello en la leyenda se especifique a que grupo pertenecen.

Ademas del color de fondo, se añade una capa visual de iconos que representan los casos confirmados y el nivel de riesgo que representa la cantidad de casos. Primero tenemos los puntos rojos, los cuales van apareciendo a través del tiempo en las diferentes regiones, estos puntos siguen una escala de 1 punto por cada cierta cantidad de casos confirmados para capturar mejor la magnitud relativa en zonas con diferencias marcadas de casos. Para el conjunto de datos de Covid se ha seguido una escala de 1 a 10. Segundo, se tiene iconos que representan el nivel de gravedad de los casos confirmados en una región específica, las zonas con brotes incipientes se marcan con icono de “virus verde”, las zonas de alerta con un “símbolo de alerta” de color amarillo, y las zonas de emergencia con un “símbolo de sirena” de color rojo.

4.7.3 Gráficos Radiales por Cluster

A la derecha del mapa se encuentran los gráficos de radar Fig. 6(c), una visualización multivariada que permite analizar en detalle los perfiles sociodemográficos con las variables más influyentes de cada grupo de regiones con comportamiento similar (clúster) identificado.

Cada gráfico radial representa un grupo (clúster) de regiones con comportamientos epidemiológicos similares. Para facilitar la comparación visual, el grupo principal se muestra con líneas y un área resaltada, utilizando el color distintivo de su clúster. Los perfiles de los demás grupos se superponen de manera opaca con sus respectivos colores.

Esta técnica de visualización permite una comparación di-

recta e intuitiva de los factores sociodemográficos clave, destacando las variables que más diferencian a un grupo de los otros. Así, se pueden identificar patrones de comportamiento que están fuertemente correlacionados con ciertas características sociodemográficas.

4.7.4 Serie Temporal de Casos Confirmados

Debajo del mapa principal se encuentra la serie temporal de casos promedio por clúster Fig. 6(d), una visualización que representa la evolución mensual del número de casos confirmados desde el inicio del brote. Cada línea de esta gráfica corresponde a un clúster previamente identificado en el análisis multivariado, manteniendo los mismos colores usados en el mapa de propagación y los gráficos radiales.

Esta visualización permite observar cómo varió la intensidad de los contagios a lo largo del tiempo en cada grupo. A diferencia de mostrar series individuales por región, esta gráfica consolida los datos a nivel de clúster, lo que facilita identificar patrones agregados y diferencias estructurales entre grupos. Por ejemplo, algunos clústeres pueden haber tenido picos tempranos de contagio mientras que otros experimentaron crecimientos más tardíos o prolongados.

Además de ello, al hacer hover sobre una de las líneas, el sistema despliega en color gris las líneas individuales correspondientes a cada región que conforma ese clúster, permitiendo visualizar cómo se distribuyeron internamente los casos dentro del grupo. De manera sincronizada, el mapa principal también resalta geográficamente las regiones pertenecientes a dicho clúster, generando así una vista dual y coordinada entre el comportamiento temporal y su localización espacial.

También al lado derecho se encuentra el resumen estadístico mensual Fig. 6(e), una pequeña sección numérica ubicada junto al gráfico que presenta indicadores clave como el total de casos, la media, la mediana y el valor máximo registrado durante el mes seleccionado. Esta síntesis numérica actúa como un soporte contextual que ayuda a interpretar cuantitativamente lo observado en el gráfico y el mapa

En la parte inferior de la gráfica se encuentra un indicador móvil en forma de punto rojo, el cual marca el mes actualmente seleccionado en el análisis. Este punto puede ser arrastrado manualmente por el usuario, siempre y cuando la opción de tiempo automático esté desactivada en el panel de control. Esta interacción permite al analista detener la animación automática y focalizarse en un mes específico de interés. Al mover el punto rojo, el sistema actualiza automáticamente tanto el mapa de propagación de casos como el resumen estadístico mensual, de modo que todo el entorno visual del dashboard se sincroniza con la temporalidad seleccionada.

Gráficos/Requerimientos	R1	R2	R3	R4
Panel de Control				X
Mapa de Casos	X		X	
Gráfico Radiales por cluster	X	X		
Serie Temporal por Cluster			X	X
Resumen de Casos				X

Figure 7. Graficos vs Requerimientos

5 Casos de Estudio

5.1 Caso de Estudio 1: Análisis de la Evolución de Regiones de Alto Riesgo

Este caso de estudio tiene como objetivo demostrar la capacidad de la herramienta para identificar y analizar la evolución de regiones que experimentaron brotes iniciales y cuya propagación se intensificó de manera alarmante a lo largo del tiempo. Utilizando un conjunto de datos de COVID-19, se busca comprender cómo la herramienta permite vincular el comportamiento epidemiológico con los factores sociodemográficos subyacentes.

Para llevar a cabo el análisis, se sigue un proceso de exploración visual integrado en la herramienta. En primer lugar, se utiliza el mapa de propagación para obtener una visión general de la situación en los primeros periodos de la pandemia. Al avanzar la línea de tiempo en el panel de control, se observa la aparición de los primeros puntos de brote en diversas regiones. La herramienta ha agrupado estas regiones con comportamientos de propagación inicial en el Clúster 1, representándolas con el color morado.

A continuación, la atención se dirige al gráfico de series temporales para examinar la evolución de este grupo. Se observa que, a pesar de haber manifestado los brotes iniciales, la curva de casos del Clúster 1 muestra un crecimiento exponencial sostenido, superando al resto de los grupos en el tiempo. Esta observación confirma que el Clúster 1 representa a las regiones con una progresión de la enfermedad particularmente alarmante, convirtiéndolas en un foco de alto riesgo.

Finalmente, para identificar los factores que podrían explicar esta evolución, se analiza el gráfico radial del Clúster 1. La visualización comparativa revela un perfil sociodemográfico distintivo: los valores más elevados en este clúster corresponden a la densidad de población y a las personas por dormitorio. La superposición opaca de los perfiles de los demás clústeres confirma que estas variables son significativamente más altas en las regiones del Clúster 1.

Este hallazgo sugiere una fuerte correlación entre las condiciones de alta densidad de población y el riesgo de una propagación acelerada de la enfermedad, demostrando cómo la herramienta permite vincular de manera intuitiva el comportamiento epidemiológico con sus posibles causas sociodemográficas.

5.2 Caso de Estudio 2: Comparación de Perfiles de Propagación Sostenida vs. Controlada

El objetivo de este caso de estudio es contrastar el comportamiento epidemiológico de dos grupos de regiones para identificar las diferencias en sus patrones de propagación y los factores sociodemográficos asociados. Para ello, se comparará el Clúster 2 (regiones en azul), que exhibió un crecimiento sostenido pero menos abrupto, con el Clúster 4 (regiones en verde), que logró controlar la propagación de la enfermedad en una etapa temprana.

El análisis comienza en el gráfico de series temporales (d), donde se observa que el Clúster 2 muestra un crecimiento gradual y sostenido de casos a lo largo del tiempo, sin picos tan

pronunciados como los del Clúster 1. En contraste, el Clúster 4 presenta una curva de casos significativamente más plana, indicando una propagación contenida. La visualización simultánea de ambos perfiles en el mismo gráfico temporal permite apreciar esta diferencia de manera clara.

A continuación, se utiliza el gráfico radial (c) para comparar los perfiles sociodemográficos de estos dos clústeres. Al seleccionar y resaltar el Clúster 2, se observa que sus valores más altos corresponden a un porcentaje elevado de personas que utilizan transporte público. Por otro lado, al analizar el perfil del Clúster 4, se observa que sus variables más altas están asociadas a un ingreso familiar promedio más elevado y un bajo porcentaje de personas por dormitorio. La superposición de los dos perfiles en el gráfico radial destaca de inmediato estas diferencias.

Este contraste visual sugiere una hipótesis clave: los clústeres con una propagación más controlada, como el Clúster 4, podrían estar asociados a factores de mayor bienestar económico y menor densidad de convivencia. En cambio, los clústeres con una propagación más sostenida, como el Clúster 2, podrían verse influenciados por variables relacionadas con la movilidad y el uso de transporte público, lo que facilitaría una transmisión comunitaria persistente. Este estudio demuestra cómo la herramienta es eficaz no solo para identificar patrones de alto riesgo, sino también para explorar las posibles causas detrás de una gestión exitosa de la crisis.

6 Discusión de Resultados

Durante el desarrollo del análisis visual, se lograron identificar diversos patrones y conocimientos relevantes que responden a las cuatro tareas analíticas planteadas (R1–R4).

Uno de los patrones más significativos hallados fue la presencia de agrupamientos regionales con comportamientos epidemiológicos similares. A través del uso combinado de técnicas de clusterización y visualización geográfica, se identificó que regiones con alta densidad poblacional y mayor número de personas por dormitorio —como las contenidas en el Clúster 1— tendieron a presentar una progresión acelerada y sostenida de casos. Este patrón sugiere una fuerte relación entre condiciones de hacinamiento y mayor riesgo de propagación, reforzando la importancia de factores comunitarios en la evolución epidemiológica. La visualización conjunta de métricas como la media, mediana y valores extremos facilitó una comprensión integral de las dinámicas dentro de cada grupo.

Mediante el análisis de los gráficos radiales por clúster, se identificaron factores sociodemográficos clave diferenciadores entre regiones con comportamientos contrastantes. Por ejemplo, se observó que el Clúster 4, con menor propagación de casos, presentó valores más altos en ingreso familiar promedio y menor densidad habitacional, mientras que el Clúster 2, con propagación sostenida, mostró un mayor uso de transporte público, lo cual puede facilitar la transmisión comunitaria del virus. Este hallazgo respalda la hipótesis de que el bienestar económico y las condiciones de movilidad son determinantes críticos para entender la vulnerabilidad de una región.

Gracias a la integración del mapa animado con la serie temporal, se pudo identificar zonas de brote y evolución sincrónica

entre grupos de regiones. Un patrón relevante fue que los brotes iniciales ocurrieron en ciertas áreas específicas (Clúster 1), que posteriormente extendieron su influencia hacia otras zonas. Asimismo, se evidenció que algunos grupos geográficos experimentaron picos de contagio en momentos diferentes, lo cual sugiere una propagación escalonada y heterogénea en el tiempo. Este análisis espacio-temporal permitió generar alertas visuales claras sobre regiones críticas, apoyando estrategias de intervención focalizada.

El panel de control permitió a los usuarios explorar la evolución dinámica de los casos a través de múltiples configuraciones temporales y territoriales, lo que reveló un patrón interesante: el comportamiento temporal de los casos por clúster tiende a seguir trayectorias coherentes y sincronizadas dentro del grupo, pero diferenciadas entre grupos. Al manipular la línea de tiempo, se observó que ciertas regiones respondieron con mayor rapidez a las medidas de control, mientras que otras mantuvieron un crecimiento constante, lo que evidencia la importancia del seguimiento temporal interactivo para una evaluación retrospectiva más precisa de la eficacia de las políticas implementadas.

7 Conclusiones

Este trabajo presentó una herramienta visual interactiva capaz de integrar información epidemiológica y sociodemográfica para analizar la propagación espacio-temporal del COVID-19. A través de la segmentación por clústeres, el sistema permitió identificar regiones con comportamientos epidemiológicos similares y asociarlos con factores comunitarios clave como la densidad poblacional, el ingreso económico y la movilidad urbana. La implementación de visualizaciones coordinadas —como mapas dinámicos, series temporales y gráficos radiales— facilitó una comprensión profunda y accesible de la evolución de la pandemia en distintos contextos locales.

Los resultados obtenidos evidencian que existen patrones consistentes entre ciertos perfiles sociodemográficos y la intensidad o control de los brotes. Regiones con mayor hacinamiento o dependencia del transporte público tendieron a presentar una propagación más sostenida, mientras que aquellas con mejores condiciones económicas mostraron curvas más contenidas. Esta evidencia refuerza la necesidad de considerar variables estructurales al momento de diseñar estrategias de salud pública, ya que las condiciones sociales tienen un impacto directo en la vulnerabilidad frente a una crisis sanitaria.

Finalmente, la herramienta desarrollada demostró ser útil no solo para el análisis técnico de datos, sino también como una plataforma de exploración que promueve la toma de decisiones informadas. Su flexibilidad para ajustar parámetros, visualizar tendencias en tiempo real y comparar escenarios permite adaptar su uso a diversas etapas de una emergencia sanitaria. Este enfoque multidimensional representa un aporte significativo al campo del análisis visual aplicado a la salud pública y puede ser extendido a otros contextos epidemiológicos o territoriales.

References

- [1] M. Park, A. R. Cook, J. T. Lim, Y. Sun, and B. L. Dickens, "A systematic review of covid-19 epidemiology based on current evidence," *Journal of Clinical Medicine*, vol. 9, no. 4, p. 967, 2020.
- [2] F. Y. Al-Ashwal, M. Kubas, M. Zawiah, A. N. Bitar, R. Mukred Saeed, S. A. S. Sulaiman, A. H. Khan, and S. M. S. Ghadzi, "Healthcare workers' knowledge, preparedness, counselling practices, and perceived barriers to confront covid-19: A cross-sectional study from a war-torn country, yemen," *PloS one*, vol. 15, no. 12, p. e0243962, 2020.
- [3] B. L. Perry, B. Aronson, and B. A. Pescosolido, "Pandemic precarity: Covid-19 is exposing and exacerbating inequalities in the american heartland," *Proceedings of the National Academy of Sciences*, vol. 118, no. 8, p. e2020685118, 2021.
- [4] J. M. Baena-Díez, M. Barroso, S. I. Cordeiro-Coelho, J. L. Díaz, and M. Grau, "Impact of covid-19 outbreak by income: hitting hardest the most deprived," *Journal of Public Health*, vol. 42, pp. 698–703, 08 2020.
- [5] G. A. Millett, A. T. Jones, D. Benkeser, S. Baral, L. Mercer, C. Beyrer, B. Honermann, E. Lankiewicz, L. Mena, J. S. Crowley, J. Sherwood, and P. S. Sullivan, "Assessing differential impacts of covid-19 on black communities," *Annals of Epidemiology*, vol. 47, pp. 37–44, 2020.
- [6] M. A. Billah, M. M. Miah, and M. N. Khan, "Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence," *PloS one*, vol. 15, no. 11, p. e0242128, 2020.
- [7] E. Rydow, R. Borgo, H. Fang, T. Torsney-Weir, B. Swallow, T. Porphyre, C. Turkay, and M. Chen, "Development and evaluation of two approaches of visual sensitivity analysis to support epidemiological modeling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 1255–1265, 2023.
- [8] M. Riccaboni and L. Verginer, "The impact of the covid-19 pandemic on scientific research in the life sciences," *PLOS ONE*, vol. 17, no. 2, p. e0263001, 2022.
- [9] P. Sanz-Leon, L. H. W. Hamilton, S. J. Raison, A. J. X. Pan, N. J. Stevenson, R. M. Stuart, R. G. Abeysuriya, C. C. Kerr, S. B. Lambert, and J. A. Roberts, "Modelling herd immunity requirements in queensland: Impact of vaccination effectiveness, hesitancy and variants of sars-cov-2," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 380, no. 2233, p. 20210311, 2022.
- [10] M. Angelini and G. Cazzetta, "Progressive visualization of epidemiological models for covid-19 visual analysis," in *AVI Workshop on Big Data Applications*, pp. 163–173, Springer, 2020.
- [11] D. Antweiler, D. Sessler, M. Rosknecht, B. Abb, S. Ginzel, and J. Kohlhammer, "Uncovering chains of infections through spatio-temporal and visual analysis of covid-19 contact traces," *Computers & Graphics*, vol. 106, pp. 1–8, 2022.
- [12] M. Thöny, R. Schnürer, R. Sieber, L. Hurni, and R. Pajarola, "Storytelling in interactive 3d geographic visualization systems," *ISPRS International Journal of Geo-Information*, vol. 7, no. 3, p. 123, 2018.
- [13] S. Goodwin, J. Dykes, A. Slingsby, and C. Turkay, "Visualizing multiple variables across scale and geography," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 599–608, 2016.
- [14] L. Mason, B. Hicks, and J. S. Almeida, "Epivecs: exploring spatiotemporal epidemiological data using cluster embedding and interactive visualization," *Scientific Reports*, vol. 13, no. 1, p. 21193, 2023.
- [15] Q. Cao, R. Jiang, C. Yang, Z. Fan, X. Song, and R. Shibasaki, "Metapopulation graph neural networks: Deep metapopulation epidemic modeling with human mobility," *arXiv preprint arXiv:2306.14857*, 2023.
- [16] Y. Kang, S. Gao, Y. Liang, M. Li, J. Rao, and J. Kruse, "Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic," *Scientific data*, vol. 7, no. 1, p. 390, 2020.
- [17] H. Zhang, P. Li, Z. Zhang, W. Li, J. Chen, X. Song, R. Shibasaki, and J. Yan, "Epidemic versus economic performances of the covid-19 lockdown: A big data driven analysis," *Cities*, vol. 120, p. 103502, 2022.
- [18] Y. Dong, C. J. Liang, Y. Chen, and J. Hua, "A visual modeling method for spatiotemporal and multidimensional features in epidemiological analysis: Applied covid-19 aggregated datasets," *Computational Visual Media*, vol. 10, no. 1, pp. 161–186, 2024.
- [19] Y. Yu, Y. Wang, Y. Zhang, H. Qu, and D. Liu, "Inclusiviz: Visual analytics of human mobility data for understanding and mitigating urban segregation," *arXiv preprint arXiv:2501.03594*, 2025.
- [20] Ministerio de Salud del Perú, "Sala situacional covid-19," 2025. Accedido: 23 de mayo de 2025.
- [21] Australian Government Department of Health and Aged Care, "Monitoring and reporting on covid-19," 2025. Accedido: 23 de mayo de 2025.
- [22] California Department of Public Health, "Weekly respiratory virus report," 2025. Accedido: 23 de mayo de 2025.
- [23] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020.

- [24] Johns Hopkins University Center for Systems Science and Engineering (CSSE), "Covid-19 data repository," 2020–2025. Accedido: 23 de mayo de 2025.
- [25] G. Wan, Z. Liu, M. S. Lau, B. A. Prakash, and W. Jin, "Epidemiology-aware neural ode with continuous disease transmission graph," *arXiv preprint arXiv:2410.00049*, 2024.
- [26] Government of New South Wales, "Nsw government data portal," 2025. Accessed: 2025-06-12.