

# Visual Analytics de Datos Epidemiológicos: Análisis de la Propagación y Olas Pandémicas del COVID-19

CARAZAS QUISPE, Alessandro Jesus

4 de junio de 2025

## Contexto, Motivación y Justificación

La pandemia de COVID-19 generó una disrupción global sin precedentes, evidenciando la necesidad de herramientas analíticas más robustas para interpretar la evolución de la enfermedad. Actualmente, existe una gran disponibilidad de datos epidemiológicos abiertos que permiten analizar dimensiones como el tiempo, el espacio y variables clínicas. Esta abundancia de datos representa una oportunidad única para estudiar patrones de propagación y la aparición de olas pandémicas, con el fin de comprender mejor la dinámica del brote y su comportamiento heterogéneo entre distintas regiones del mundo.

Sin embargo, interpretar estos datos de manera efectiva sigue siendo un desafío. La complejidad inherente a su naturaleza multivariada y espacio-temporal exige herramientas avanzadas que vayan más allá de la visualización tradicional. En este contexto, el enfoque de *visual analytics*, que combina análisis computacional y visualización interactiva, se posiciona como una solución prometedora para facilitar la exploración, descubrimiento de patrones y generación de conocimiento útil para epidemiólogos y autoridades sanitarias [1, 2].

Este proyecto se justifica por la necesidad de contar con un marco analítico capaz de identificar y caracterizar la propagación y las olas epidémicas a lo largo del tiempo y entre regiones. A través de una visualización dinámica e interactiva, se busca no solo representar los datos, sino también generar valor para la toma de decisiones basadas en evidencia científica, especialmente en situaciones críticas como emergencias sanitarias.

## Problema

Durante la pandemia, muchas de las herramientas empleadas para visualizar datos epidemiológicos mostraron importantes limitaciones. Entre ellas destacan la dificultad para integrar múltiples escalas espaciales (país, región, continente), la falta de interactividad en las visualizaciones, la incapacidad de representar la evolución temporal de forma clara y la ausencia de mecanismos para tratar con la incertidumbre en los datos [3, 4].

Estas deficiencias dificultaron la comprensión de fenómenos clave como la propagación del virus entre regiones y la identificación de nuevas olas pandémicas, lo cual afectó negativamente la capacidad de respuesta en salud pública. A pesar de contar con conjuntos de datos detallados y actualizados, persiste una brecha entre la disponibilidad de la información y su análisis efectivo para la detección de tendencias y patrones emergentes [5, 6].

Por tanto, el problema central de este proyecto es la falta de herramientas visuales e interactivas que permitan integrar, analizar y comunicar de manera efectiva los patrones de propagación y olas epidémicas del COVID-19 desde una perspectiva espacio-temporal y multivariada.

## Objetivo

Analizar datos epidemiológicos del COVID-19 mediante técnicas de *visual analytics*, con el propósito de identificar y caracterizar patrones de propagación y olas pandémicas, evaluando su utilidad para apoyar la comprensión del fenómeno y la toma de decisiones en salud pública.

## Descripción del Dataset

El conjunto de datos utilizado proviene del repositorio abierto de *Our World in Data* (OWID), el cual recopila información diaria sobre la evolución de la pandemia de COVID-19 en países y regiones de todo el mundo. La base de datos incluye indicadores sobre contagios, muertes, hospitalizaciones, pruebas, vacunación y respuesta gubernamental. Esta información se actualiza constantemente y permite realizar análisis espacio-temporales detallados.

## Contexto

Este dataset resulta adecuado para estudiar la propagación del COVID-19 y la detección de olas pandémicas debido a su estructura multivariada, cobertura global y periodicidad diaria. Cada fila representa un registro asociado a un país o territorio en una fecha específica. Los datos han sido procesados por OWID a partir de fuentes oficiales como la OMS y autoridades sanitarias nacionales.

## Entidad de estudio y atributos seleccionados

La entidad de estudio son los países o territorios observados a lo largo del tiempo. A continuación se presentan los atributos relevantes para este proyecto:

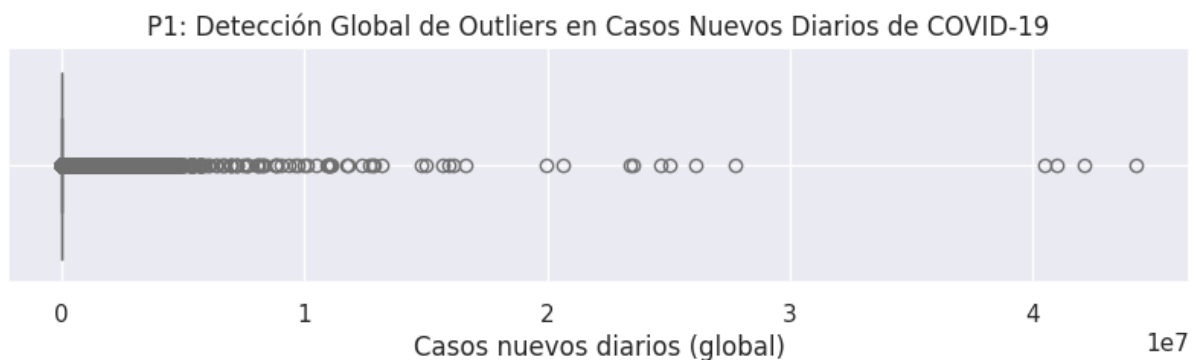
Atributo	Descripción	Tipo	Rango
location	País o región del registro	Categorico	Aproximadamente 200 países o regiones (ISO 3166-1)
date	Fecha del registro	Fecha	Desde 2020-01-01 hasta la fecha actual
new_cases	Nuevos casos confirmados diarios	Entero	0 a más de 100,000 casos diarios, variable según brote
new_cases_smoothed	Promedio móvil 7 días nuevos casos	Decimal	Similar a new_cases, suavizado para reducir variabilidad
total_cases	Total acumulado de casos confirmados	Entero	Desde 0 en adelante, puede superar varios millones
new_deaths	Nuevas muertes diarias reportadas	Entero	0 a varios miles, variable según gravedad del brote
new_deaths_smoothed	Promedio móvil 7 días nuevas muertes	Decimal	Similar a new_deaths, suavizado para observar tendencias
reproduction_rate	Tasa estimada de reproducción (Rt)	Decimal	Generalmente entre 0.5 y 2.5; ¿1 indica expansión
icu_patients	Pacientes en UCI por día	Entero	Variable, desde 0 hasta varios miles según país y brote
hosp_patients	Pacientes hospitalizados totales	Entero	Desde 0 hasta decenas de miles, según demanda hospitalaria
total_tests	Total de pruebas realizadas	Entero	Desde 0 hasta varios millones acumulados
positive_rate	Proporción de pruebas positivas	Decimal	De 0.0 (0 %) a 1.0 (100 %), proporción de positivos
stringency_index	Índice de medidas restrictivas	Decimal	De 0 (sin restricciones) a 100 (máxima restricción)

Cuadro 1: Atributos relevantes para el análisis epidemiológico con sus características y rangos.

La tabla anterior presenta los atributos fundamentales seleccionados para el análisis epidemiológico de los datos. Cada atributo incluye una breve descripción, su tipo de dato y un rango típico de valores que se observan en los registros. Estos rangos reflejan la variabilidad natural de cada variable según el contexto y la evolución de la pandemia en diferentes países o regiones. Por ejemplo, los casos nuevos diarios pueden variar desde cero hasta más de cien mil, dependiendo del brote, mientras que el índice de restricciones varía de cero (sin medidas) a cien (máxima restricción). Este detalle es clave para realizar análisis precisos y detectar posibles anomalías o patrones relevantes en los datos estudiados.

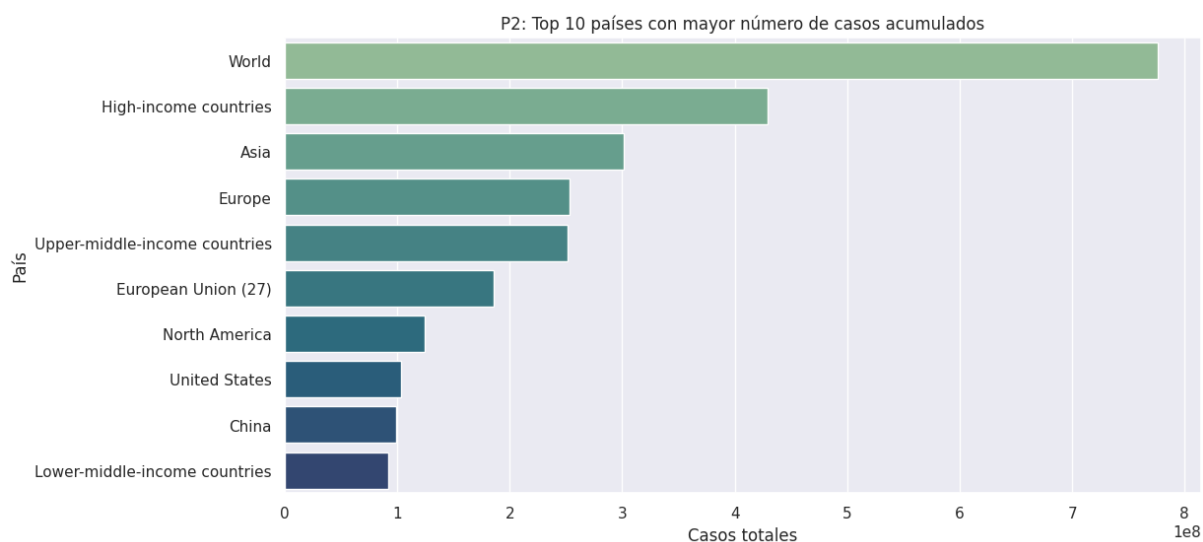
## Preguntas

Pregunta 1: ¿Qué problemas hay en el dataset?



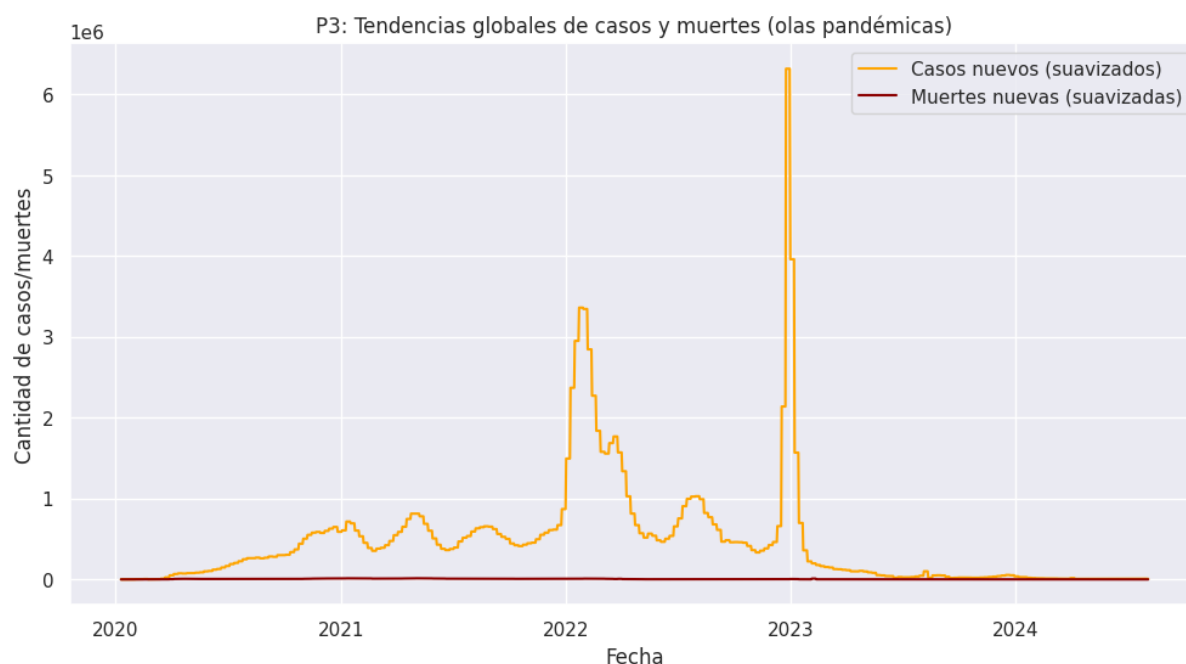
Se detectaron datos anómalos como casos negativos, lo cual es inválido. Esto puede ser causado por errores de carga, correcciones retroactivas o fallas en los sistemas nacionales.

Pregunta 2: ¿Qué se descubrió al analizar los datos? (comportamiento global)



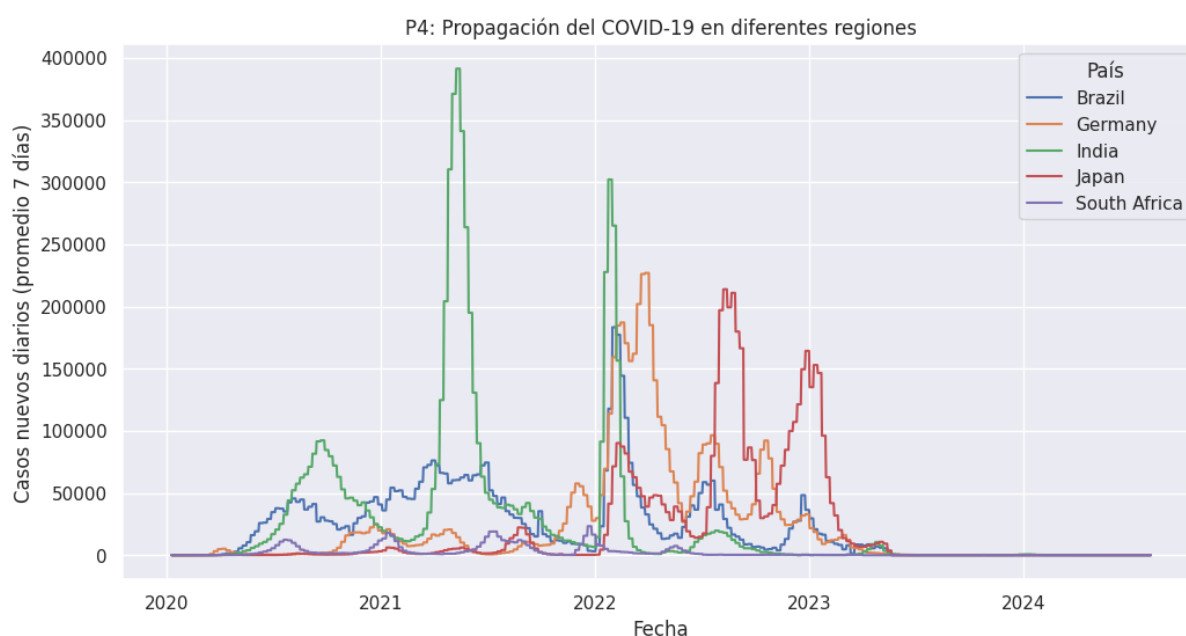
Asia y Europa encabezan la lista, lo cual es consistente con el impacto mediático y los reportes de la OMS. Sirve como base para priorizar visualizaciones regionales.

Pregunta 3: ¿Qué reflejan los patrones de tendencia?



Las curvas evidencian las olas pandémicas: periodos de crecimiento rápido seguidos de descensos, permitiendo estudiar picos y fases de control.

Pregunta 4: **¿Cómo varía la propagación geográficamente?**



Se observan distintos patrones de propagación. India y Germania presentan olas abruptas, mientras Japón y Brasil muestran picos más modulados. Esto permite personalizar la respuesta sanitaria por región.

## Referencias

- [1] Yuxuan Dong, Chun Jie Liang, Yifan Chen, and Jing Hua. A visual modeling method for spatiotemporal and multidimensional features in epidemiological analysis: Applied covid-19 aggregated datasets. *Computational Visual Media*, 10(1):161–186, 2024.
- [2] Donghee Kim, Beatriz Cánovas-Segura, Marta Campos, and Jose M. Juarez. Visualization of spatial-temporal epidemiological data: A scoping review. *Technologies*, 12(3):31, 2024.
- [3] Thierry Joliveau, Robin Cura, and Henri Commenges. Covid-19 geoviz for spatio-temporal structures detection. In *Proceedings of the International Cartographic Association*, volume 4, page 37, 2021.
- [4] Saturnino Luz and Masood Masoodian. Temporal and spatial elements in interactive epidemiological maps. arXiv preprint arXiv:2206.06048, 2022.
- [5] Jiannan Chen, Alan M. MacEachren, and Diansheng Guo. Visual analytics for epidemiologists: Understanding the interactions between age, time, and disease with multi-panel graphs. *PLOS ONE*, 6(2):e14683, 2011.
- [6] University of Memphis School of Public Health. Visual communication of public health data: A scoping review. *Frontiers in Public Health*, 13:1234567, 2025.