

CAPÍTULO

7

NUEVAS HERRAMIENTAS

Patrones de la movilidad en tiempos de pandemia: una aproximación con técnicas del “big data”

HALLAZGOS RELEVANTES

- En la pandemia, entre marzo y setiembre del 2020 hubo semanas en las cuales los flujos de tránsito se redujeron hasta en un 90% con respecto a los meses previos. Este comportamiento se constata en la congestión desagregada por cantones, en los cambios de intensidad por día de la semana, hora del día, y en términos agregados en los valores mensuales del 2020 comparado con lo ocurrido en 2019.
- Por primera vez se calcularon conglomerados espaciales de la congestión, analizados con el índice de Morán, que reflejan zonas de alto tránsito coincidentes con áreas densamente pobladas dentro de la GAM. Son cantones donde un 53% de la población ocupada, en promedio, se moviliza fuera de sus cantones para trabajar. De esta relación se pueden inferir importantes flujos intercantonales en esas zonas, que concuerdan con los nudos viales del país.
- Las restricciones a la movilidad impuestas por el Poder Ejecutivo para el control de la pandemia redujeron la intensidad de la congestión al grado que rompieron la relación territorial de esa movilidad a partir de abril, de acuerdo con los análisis estadísticos ejecutados.
- Las mediciones de Waze sobre congestión vial tienen muy alta correlación con los registros de Google que han servido de insumo en la literatura comparada para evaluar la movilidad de personas durante la pandemia. Como Waze tiene microdatos con desagregación territorial y por unidades de tiempo, fue posible efectuar un análisis en profundidad para el caso de Costa Rica.
- La exploración inicial entre la congestión vial, como *proxy* de movilidad de personas, y la cantidad de contagios por covid-19 según cantón muestra una relación positiva en los modelos de regresión evaluados. Es decir, en zonas con alta congestión vial se presentan altos niveles de contagios.
- Otras variables sociodemográficas también indicaron relaciones positivas con los contagios del virus. Destacan la densidad de población, la cantidad de patentes en el cantón y la proporción de personas que trabajan en cantones distintos de donde viven, entre las principales. Estas dos últimas se relacionan con la movilidad de personas por el empleo y las actividades productivas. Estas variables pueden generar riesgos potenciales en el cumplimiento estricto de las medidas de distanciamiento físico para el control del virus.

NUEVOS APORTES PARA LA TOMA DE DECISIONES

- Se exploran los registros de congestión vial de Waze para monitorear la movilización de las personas en medio de la pandemia generada por el covid-19. Esta fuente de datos masivos, con microdatos desagregados territorialmente y en el tiempo, puede ser un insumo útil para entender los cambios en el comportamiento social en zonas con características específicas, en una época atípica como la vivida por la emergencia sanitaria en Costa Rica.
- La información acerca de la magnitud de los cambios en la movilidad de las personas debido a las restricciones establecidas por el Gobierno ante la emergencia por la pandemia, así como por la posible autocontención individual por los riesgos de contagio, tiene potencial para convertirse en un sistema de monitoreo de la cambiante situación, con el fin de ir ajustando medidas sanitarias relacionadas con la movilidad.
- Brinda elementos que les sirvan a quienes toman las decisiones, para identificar los efectos de la aplicación de medidas focalizadas con restricciones donde hay focos de contagio, y de mayor apertura donde no los hay, con el propósito de evitar mayores afectaciones en la actividad económica del país.
- Es también un insumo para valorar la factibilidad de aplicar medidas sobre el tránsito en condiciones posteriores a la pandemia, con el fin de reducir los tiempos perdidos por la congestión, la contaminación y otras externalidades negativas que produce el fuerte embotellamiento en la GAM.

CAPÍTULO

ARMONÍA CON LA NATURALEZA

7

/ Patrones de la movilidad en tiempos de pandemia: una aproximación con técnicas del “big data”

INDICE	Introducción
Hallazgos relevantes 231 Nuevos aportes para la toma de decisiones 231 Introducción 233 Metodología para evaluar la movilidad de personas durante la pandemia 234	<p>El presente capítulo analiza los efectos de la pandemia sobre el país con base en tres factores específicos: las medidas de restricción vehicular adoptadas por el Gobierno para combatir la pandemia, los flujos de movilidad de la población, y la asociación de estos con respecto a los nuevos contagios locales de covid-19.</p>
Proceso de construcción y actualización de la información con técnicas de “big data” 234 Waze como aproximación a la movilidad de personas 235 Modelos de análisis 238 Hipótesis de trabajo 238	<p>Se trata de una investigación aplicada a la toma de decisiones, cuyo objetivo central es crear una herramienta que alimente decisiones informadas en materia de transporte y movilidad durante la atención de la pandemia, en especial el control de la expansión del virus. No obstante, los hallazgos de este capítulo también pueden constituirse en insumos para discusiones posteriores sobre la incidencia que pueden tener políticas concretas para minimizar el grado de congestión vial en el país.</p>
La pandemia modifica las zonas de mayor congestión vial 239	<p>los últimos cinco años. Se partió inicialmente de análisis generales sobre factores globales que condicionan estos patrones, como la falta de ordenamiento territorial de la Gran Área Metropolitana (GAM) y el crecimiento de las ciudades intermedias. Se indagó la relación entre estos factores y los problemas de la población para movilizarse debido a un estilo urbano poco denso, distante y desconectado, con una gran apuesta por el vehículo particular, un sistema de transporte público poco eficiente y una red vial que no da abasto.</p>
Fuerte reducción de la congestión vial en todo el territorio durante la pandemia 239	<p>En los primeros acercamientos se estimaron algunas externalidades negativas de esta situación, como mayores tiempos de viaje, costos económicos y contaminación ambiental, por ejemplo. Se demostró la necesidad de discutir el sistema en su conjunto, y superar la idea de que los problemas encontrados se puedan resolver exclusivamente por la vía de construcción de infraestructura. Después de arribar a estas conclusiones, se hizo necesario avanzar hacia el uso de datos más precisos, con el fin de aportar evidencia más robusta para la toma de decisiones y la deliberación pública en este campo.</p>
Medidas de restricción borraron los patrones territoriales 245	<p>La principal fuente de información empleada son los microdatos de la aplicación móvil Waze, registrados entre enero de 2019 y setiembre de 2020, los cuales permiten una aproximación a los movimientos poblacionales en los ámbitos nacional y subnacional, con perspectiva comparada en el tiempo, y su relación con distintas variables sociales y territoriales. Esta fuente permite alta precisión temporal y espacial en el análisis de los flujos de movilidad diaria, en el marco de las restricciones y políticas establecidas para controlar la expansión del virus.</p>
Factores asociados al comportamiento social en pandemia 247	<p>El tema de los patrones de transporte y movilidad en el país ha sido objeto de estudios detallados en este Informe en</p>
Conclusiones 251	

La conformación en el PEN de una Mesa de Ciencia de Datos y Visualización posibilitó dar un paso adelante en esa dirección, al emplear microdatos y fuentes masivas de información en las investigaciones que el Programa realiza (“*big data*”). Una de las iniciativas impulsadas dentro de esta mesa de trabajo fueron los estudios con los registros de la aplicación Waze, a los cuales se tuvo acceso gracias a un acuerdo de cooperación con el MOPT. Con esta aplicación sobre flujos de tránsito, se midió el nivel de saturación vial desde el plano cantonal y el distrital, hasta un grado detallado de calle por calle en todo el país. De este modo, se logró describir la situación de rutas concretas con precisión de cien metros aproximadamente. Se publicó, en 2019, un mapa para cada uno de los cantones del país con el nivel de congestión promedio de cada calle, en una aplicación web que puede ser accedida en www.dcifra.cr.

El presente capítulo emplea estos microdatos para analizar el comportamiento de la congestión vial en la época de la pandemia, utilizando la información de Waze como una aproximación a la movilidad de personas en el país. Aunque la fuente registra datos con base en los usuarios de la aplicación, diversas pruebas estadísticas demuestran que, para Costa Rica, es una fuente robusta para estimar la movilidad, tal como se comenta en la sección metodológica.

La expectativa teórica de esta investigación se centra en medir la siguiente relación de variables: las medidas sanitarias para el control de la pandemia tienen impactos sobre la movilidad de las personas. Los cambios en esa movilidad, junto a una serie de variables sociodemográficas, además tienen impactos sobre la cantidad de contagios a nivel local en el país.

Se comparan en el tiempo y con precisión territorial las maneras como se modificaron los niveles de congestión, los patrones espaciales de movilidad, y la relación de ese comportamiento, junto a variables sociales, demográficas y económicas, con la cantidad de contagios por cantón. Los resultados pueden ser un insumo de las autoridades para incidir en el control sobre la expansión del virus.

Metodología para evaluar la movilidad de personas durante la pandemia

Este capítulo forma parte de los análisis de la Mesa de Ciencia de Datos y Visualización, del Programa Estado de la Nación (PEN), con el apoyo del Colaboratorio Nacional de Computación Avanzada (CNCA) del Cenat-Conare, y del Ministerio de Obras Públicas y Transportes (MOPT), para el acceso a los datos de Waze.

Analiza –como se indicó en la introducción– los datos de congestión vial registrados por la aplicación de tránsito Waze, junto a otras variables sociodemográficas, como herramienta para aproximar (*proxy*) la movilización de personas antes y durante la pandemia y su relación con los casos de contagio cantonal por covid-19. Se cubre el período entre enero de 2019 y setiembre de 2020 (21 meses) en la mayoría de los procesamientos. Los análisis de regresión incluyen información hasta el mes de julio de 2020.

Con base en este objetivo general, la investigación plantea despejar cuatro interrogantes centrales, a partir de las cuales se organiza el índice de contenidos del presente capítulo:

- ¿Cuáles cantones reportaron mayor congestión durante el 2019 y cómo varió esa congestión durante el 2020?
- ¿Cuánto varió la congestión durante la pandemia –de marzo a setiembre de 2020– con respecto a los patrones anteriores, dadas las medidas de restricción impuestas por el Gobierno y la autocontención de las personas a salir de sus casas?
- ¿Modificaron los cambios registrados en la intensidad de la congestión vial también los patrones territoriales de movilidad?
- ¿Están los cambios en la congestión vial, y otras variables sociodemográficas, relacionados con la cantidad de contagios por covid-19 en el ámbito cantonal?

Para responder la primera interrogante, se construyó un *ranking* de congestión cantonal con los datos del 2019 y se comparó con lo ocurrido durante el 2020, con la finalidad de determinar la variación. El trabajo con la segunda interrogante implicó evaluar múltiples indicadores, con distintos niveles de desagregación, para identificar la magnitud de los cambios en la congestión durante la pandemia. La tercera interrogante se trabajó con modelos de estadística espacial para determinar si la congestión vial cantonal, como medida de aproximación a la movilización de personas, mantiene (o alteró) los patrones territoriales de años anteriores durante los primeros siete meses del año 2020.

Finalmente, para la cuarta interrogante se exploró la relación entre congestión vial y contagios, con datos de marzo a julio de 2020. El objetivo fue producir información para evaluar los posibles efectos de las medidas de restricción vehicular y de comercios sobre la movilidad de personas dirigidas a atender la situación de salud del país. También, se generó nueva información para explorar los riesgos potenciales del comportamiento social en zonas con características sociodemográficas específicas, a partir de las recomendaciones de distanciamiento comunicadas por las autoridades de salud.

Proceso de construcción y actualización de la información con técnicas de “big data”

La base de datos de Waze se utilizó con datos actualizados cada diez minutos en todo el territorio nacional. El PEN-CONARE tuvo acceso a esta información gracias a un convenio de trabajo conjunto con el MOPT (recuadro 7.1).

Para extraer, transformar y cargar los datos (*pipeline de Extract Transform and Load*, ETL por sus siglas en inglés), se aplicaron técnicas de “*big data*” y luego se requirió procesamiento en paralelo para acceder, resumir y analizar la información, etapa denominada como el *pipeline* de analítica. El *pipeline* del ETL se realizó con el lenguaje de programación Python; y el de analítica, con el lenguaje de programación R y el paquete

Recuadro 7.1

Base de datos de Waze para Costa Rica

El Programa Estado de la Nación (PEN) tiene acceso a la base de datos de Waze gracias a un acuerdo de investigación conjunto con el Ministerio de Obras Públicas y Transportes (MOPT). La información es una base de datos que permite la desagregación en distintos niveles: en el nivel territorial hasta segmentos de carretera de cien metros, aproximadamente, y temporalmente cada diez minutos.

La base que se usó para el análisis de este capítulo cubre el período de enero de 2019 a setiembre de 2020. La información de jams por segmento de carretera cada diez minutos contiene 255.831.318 de registros. Entre las variables disponibles, se emplearon las siguientes:

- » IdT: identificador del segmento de carretera
- » Nom_cant: nombre del cantón donde se encuentra el segmento de carretera
- » Cod_canton: identificador del cantón donde se encuentra el segmento de carretera
- » Nom_prov: nombre de la provincia donde

- se encuentra el segmento de carretera
- » Nom_distr: nombre del distrito donde se encuentra el segmento de carretera
- » Cod_dta: código de distrito
- » Id: identificador del jam
- » Delay: atraso promedio generado por el jam en segundos
- » díaSemana: día de la semana en el cual ocurre el jam
- » Año: año en el que ocurre el jam
- » Mes: mes en el que ocurre el jam
- » Día: día del mes en el que ocurre el jam
- » Hora: hora del día en el que ocurre el jam en formato de 24 horas
- » Semanames: número de semana del año en la cual ocurre el jam

Estas variables son producto de la transmisión de información desde teléfonos celulares con la aplicación instalada, que para el caso de Costa Rica se estimaba en más de 500.000 usuarios activos por mes en el 2015 (Cordero, 2015).

Fuente: Elaboración propia con base en Gómez Campos et al., 2020.

sparklyr para análisis con procesamiento en paralelo. Se usaron múltiples paquetes estadísticos y de visualización para los modelos efectuados, todos de código fuente abierta (*Open Source*).

La unidad de observación es la congestión –en adelante *jams*– por segmentos de carretera para el período comprendido entre enero de 2019 y setiembre de 2020. La base de datos contiene poco más de 255 millones de registros. Para llegar a este nivel de desagregación, se siguieron tres pasos:

- Extracción de todos los jams y sus atributos en formato JSON, a través de conexión con el API de Waze.
- Segmentación de la red vial nacional y cantonal del país en tramos de cien metros, aproximadamente, y asigna-

ción de un identificador único a cada segmento. Se aplicó la metodología desarrollada en la investigación de Cubero et al. (2019), y para esta ocasión se actualizó la red vial.

- Intersección de la base de *jams* –punto #1– con la de segmentos de carretera –punto #2–, para obtener un conteo de *jams* en cada segmento de carretera. Los segmentos que no presentan información se mantienen en cero. Esto genera un polígono de toda la red vial, con indicadores de la congestión a lo largo del año –total de *jams*, velocidad promedio, tamaño y coordenadas espaciales, entre los principales–.

Esta metodología evita que la extensión de un atasco afecte el conteo de la presencia o no de esta en un tramo de carretera.

En lenguaje técnico eso significa que se “controla” la variable “congestión” por la variable “extensión”. Por ejemplo, un atasco A, que se extiende por 500 metros, se contabiliza como cinco unidades de congestión. Otro atasco B, con extensión de un kilómetro, se registra en la base de datos como diez unidades. Esta es una medida más precisa de la congestión para aproximar la movilidad agregada.

Como la toma de datos ocurre cada diez minutos desde el API de Waze, algunos casos de congestión pueden estar duplicados en la base de datos. Este sobreregistro no genera mayores alteraciones, por dos motivos: el análisis no se centra en el valor absoluto, sino en la tendencia y cambios comparativos en el tiempo y territorio para todos los casos; además, la temporalidad en la captura de datos es la misma durante todo el período de análisis, por lo cual el posible sobreregistro no altera la serie de tiempo por partes.

Waze como aproximación a la movilidad de personas

En este análisis, se usa la congestión vial como una variable para aproximar la movilidad de personas (variable *proxy*). Una noción sencilla conceptualmente está detrás de esta aproximación: alta congestión en una zona de carretera denota fuerte aglomeración de vehículos. Eso incluye autos privados y transporte público (autobuses comerciales y privados, taxis, busetas de estudiantes, etc.). Desde esa perspectiva, esta aglomeración de vehículos representa una aglomeración de personas que están transitando de un lugar a otro. Identificar en espacio y tiempo el tránsito vial permite construir un rastro de movilización de personas en el territorio nacional a lo largo del día y mes. Por lo anterior, en este trabajo se entiende la congestión vial como una forma indirecta de medir flujos de personas.

¿Cuán robusta puede ser esa medición? El caso de Costa Rica presenta un escenario propicio para estimar la movilidad de personas, con base en datos de congestión vial y contagios por covid-19, por varias razones. Primero, se cuenta con información de la aplicación Waze para

un largo período de tiempo –enero de 2019 a setiembre de 2020–. Estos datos, además, están altamente correlacionados con otras medidas que han probado ser robustas para aproximar la movilidad de personas en medio de la pandemia, en especial los registros agregados por país que publica Google como parte de su proyecto “*Community Mobility Reports*”. A diferencia de Google, los datos de Waze posibilitan, para el caso de Costa Rica, profundizar el análisis local para determinar comportamientos específicos.

Vale indicar que esas mediciones no son una representación perfecta de la realidad; es decir, no representan un censo de la movilización de las personas. No obstante, dado el uso masivo de tecnología celular en el país, constituyen una de las herramientas más robustas disponibles hasta el momento para medir los fenómenos en estudio en este capítulo.

La atención a la pandemia del covid-19 trajo consigo medidas como el teletrabajo y las restricciones viales, que abren la oportunidad de estudio para determinar su impacto sobre la movilidad de las personas. Investigaciones recientes sobre las redes viales y de tránsito señalan las ventajas de recurrir a fuentes de información generadas por la telefonía celular u otras tecnologías con transmisión de datos en tiempo real.

Las aplicaciones móviles tienen antecedentes de uso con éxito para planificar obras viales e integrar los efectos que los datos móviles generan en la toma de decisiones públicas (Townsend, 2014). La plataforma Waze incluye en su sitio web casos de estudio alrededor del mundo en torno de la implementación de obras viales, intervenciones específicas frente a escenarios de emergencia u otras necesidades con base en la información recolectada por esta aplicación móvil¹. Como ejemplo, Sánchez et al. (2019) usan Waze como fuente de datos para validar el uso de semáforos dinámicos, también conocidos como “semáforos inteligentes”, porque cambian su temporización de luces según las condiciones actuales de tráfico. De esta manera, las nuevas fuentes de información con alta penetración en la sociedad, como la telefonía celular, se pueden integrar a esquemas de planificación de obras para el manejo del tránsito.

Estas nuevas formas de generar datos sobre el tránsito en carretera, o mediante señales de geolocalización enviadas por la tecnología celular, también han sido un insumo en medio de la pandemia. La premisa es clara: frente a un virus con alto nivel de propagación, una de las medidas para contener los contagios por covid-19 es reducir la movilidad de la población. De ahí que múltiples gobiernos en el mundo estén implantando restricciones de tránsito y cierres de lugares públicos y comercios en general, para reducir los desplazamientos y aglomeraciones de las personas.

Los efectos de esas medidas se confirman con los datos disponibles de distintas fuentes. Con información de redes sociales, en específico, de localización por medio de la aplicación Twitter, se aproximó el movimiento de las personas en medio de la pandemia. Con un estudio sobre 587 millones de tuits, Huang et al. (2020) concluyen que la información de Twitter es una fuente que permite aproximar la dinámica de movilidad de manera confiable, aunque plantea limitaciones sobre la representatividad de la población. Además, detectan que los patrones de desplazamiento han variado considerablemente durante las fases de la pandemia. Para el caso de Estados Unidos, los datos muestran un período de contención de la movilidad que luego se rompe, y ello coincide con las protestas contra el racismo y la violencia policial de los últimos meses.

Estos hallazgos concuerdan con otras investigaciones que utilizan datos agregados y anonimizados de teléfonos celulares. Un caso reciente logró medir la efectividad del distanciamiento físico, definido a partir de los cambios en la movilidad por día, para reducir el contagio en veinticinco condados de los Estados Unidos. Se encontró una alta correlación –con coeficientes de Pearson² mayores de 0,7– en veinte de los veinticinco condados del estudio. Estos resultados agregan evidencia empírica sobre los efectos de las medidas de restricción para aminorar la propagación del virus (Badr et al., 2020). Conclusiones similares se obtuvieron en una investigación en Italia en la cual se usaron sensores automáticos de conteo de autos en carretera a lo largo del todo el país (Cartení et al.,

2020), metodología utilizada en Costa Rica con el fin de estimar costos del tránsito en zona de alto flujo (recuadro 7.2).

Otro estudio en China demostró que las medidas de prevención impuestas sobre la movilidad de las personas en Wuhan y las zonas circundantes influyeron sobre los patrones de esparcimiento del covid-19 en el territorio. No obstante, también detectaron que esas medidas son perceptiblemente más efectivas cuando los contagios son más localizados. Cuando una zona amplia es contagiada, de manera más o menos homogénea, los protocolos presentan reducción en su efectividad (Kraemer et al., 2020).

Otros estudios (Fang et al., 2020; Alvarez et al., 2020; Satyakti, 2020) se suman a la creciente literatura sobre las herramientas para medir la movilidad de personas y su relación con los contagios por el virus. Como lo indica el Imperial College de Londres en uno de sus reportes, los “datos de movilidad representan un importante proxy para medir el distanciamiento social”. Con datos de Apple y de Google para 53 países encontraron, con ambas fuentes de información, una relación entre la movilidad y el contagio. La movilidad explicó más del 85% de la variabilidad en la transmisión del virus (Nouvellet et al., 2020).

Los análisis con datos de Google han probado ser un insumo central para los intentos públicos de controlar la pandemia. La empresa creó un sitio web específico para ese fin, donde actualiza sus reportes de movilidad por país. Sin embargo, tienen limitaciones: son datos agregados, los cuales no permiten ver el comportamiento local, lo que con Waze sí se puede realizar para el caso de Costa Rica.

¿Cómo se generan los datos de localización de usuarios de Google y Waze? En ambos casos, se utiliza una combinación de las señales de la aplicación y las del GPS del teléfono celular donde está instalada la aplicación. Cuando se abre el Google Maps, por ejemplo, se envían señales de la localización de los usuarios, pero, si el usuario tiene habilitado el GPS de su teléfono, también se mandan señales, aunque la aplicación no esté en uso. De manera similar sucede con la aplicación de Waze. Según los términos

Recuadro 7.2

Predicción y costos del flujo vehicular en la ruta 27

En el marco de una investigación para optar por el grado de ingeniería en producción industrial del Instituto Tecnológico de Costa Rica, se utilizó la tecnología de los sistemas de transporte inteligente (STI), conocidos como contadores de tránsito, en la ruta 27 para recolectar datos sobre flujo de tráfico por tipo de vehículo.

La investigación tuvo como objetivo aplicar modelos de predicción del flujo de tránsito con base en los registros de los STI disponibles sobre la ruta 27. Estos medidores sirven para mejorar la eficiencia del tráfico, la seguridad vial, el control de accidentes, la planificación de rutas y al uso eficiente de la infraestructura vial, entre otros.

Entre San José y Caldera existen doce puntos de conteo que registran el flujo en ambos sentidos. Entre los principales hallazgos se encontró que un 94% del tránsito en esa ruta son vehículos livianos. El tramo con mayor concentración vial es de aproximadamente 10 kilómetros. En este trayecto destacan los puntos ubicados en el kilómetro 2+900, que se encuentra a dos kilómetros más novecientos metros a partir del Gimnasio Nacional en La Sabana (antes del peaje de Escazú). En esta zona pasan hasta 2.269 vehículos en una hora en promedio. El segundo punto en importancia se ubica en el kilómetro 10+540, eso es a diez kilómetros más quinientos cuarenta metros a partir del Gimnasio Nacional (punto localizado por el puente de río Corrogres).

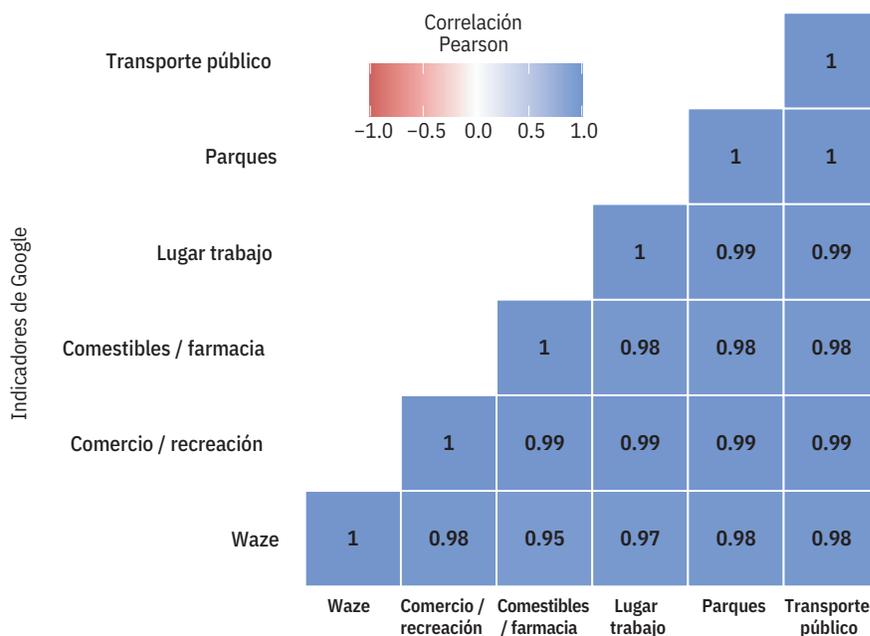
Se hizo, además, una estimación del costo social por demora considerando variables como: la longitud del tramo, el valor del tiempo en congestión, el número promedio de vehículos que pasan por el tramo, el factor de ocupación media de los vehículos y la velocidad promedio en el tramo, entre los principales. En promedio se estableció que el costo social de transitar por la ruta es de aproximadamente 109.760 colones por hora. Esta estimación de costos permite aproximar los efectos de la congestión en función de las horas perdidas por congestionamiento, el deterioro de la salud pública y la contaminación.

Fuente: Elaboración propia con base en Rivera, 2020.

Gráfico 7.1

Correlación entre las mediciones de Waze sobre congestión vial e indicadores de Google sobre movilidad para Costa Rica. 2020

(valores diarios entre el 17 de febrero y el 31 de julio)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT y Google.

de privacidad de la aplicación, para mejorar la experiencia de usuario se puede recolectar información sobre el viaje y la localización, incluso cuando no está empleando la aplicación. Esto significa que ambas aplicaciones pueden enviar señales de localización que alimentan sus bases de datos sin necesidad de iniciar un viaje desde las aplicaciones. Solo basta tenerla instalada en el teléfono celular.

Un experimento efectuado en el Instituto de Seguridad en Aplicaciones Distribuidas (SVA, por su sigla en alemán), de la Universidad Técnica de Hamburgo, comprobó la posibilidad de geolocalización con alto nivel de precisión de los usuarios a partir de sus movimientos en carretera en ambas aplicaciones (Jeske, 2013).

¿Es Waze una medida tan confiable, como Google, para aproximar el desplazamiento de personas en Costa Rica? Por la forma de recolectar los datos, existe evidencia de que ambas son un “proxy” satisfactorio, por cuanto no se requiere calcular una ruta de viaje en automóvil

para que se envíen los datos. Cada aparato celular se convierte en un sensor potencial de movimiento.

Para que los datos sean robustos, se necesita alta penetración de tecnología celular, un fenómeno que también ocurre en Costa Rica. De acuerdo con un reciente informe de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el país es uno de los líderes de la región en penetración móvil, con suscripciones por cada cien habitantes mayores que el promedio de la OCDE y de América Latina (OCDE, 2020).

La evidencia empírica respalda el grado de similitud entre los datos de movilidad de Google y los de congestión vial de Waze. Como ejercicio de validación, para este capítulo se unieron las dos series de datos con registros diarios entre el 17 de febrero y el 31 de julio de 2020 (166 días) y se correlacionaron para determinar cuán semejante son.

En efecto, ambas mediciones están altamente correlacionadas. En el gráfico 7.1, se observa la relación entre los

datos agregados por país de Waze y cinco de los indicadores que usa Google para reportar los cambios en la movilidad de personas. La correlación de Waze con las mediciones de Google es de 0,95 o superior. Es decir, ambas aplicaciones están midiendo el mismo fenómeno en niveles muy similares, en este caso los cambios en los patrones de movilización de las personas, ya sea por sus viajes en auto o por su desplazamiento con otros medios. Cuando Google señala altos niveles de movilización de personas, Waze registra altos niveles de congestión vial.

Datos agregados de Waze muestran un descenso del 85%, en promedio, en el congestionamiento del tráfico en San José durante ciertas fases de la pandemia (Galindo, J, 2020)⁵. Los datos de Google (2020), que toman en cuenta la cantidad de visitas y duración de estadía en zonas de aglomeración típicas como parques, comercios de suministros como supermercados o zonas de transporte público, indican que la reducción en la movilidad ha variado entre un 40% y un 80%, dependiendo del lugar.

En suma, existe suficiente evidencia empírica para Costa Rica, y se respalda con la literatura comparada más reciente, sobre el uso de aplicaciones como Waze o Google para aproximar mediciones sobre el desplazamiento de las personas en medio de la pandemia. Por consiguiente, en este Informe se usa el concepto de congestión vial como una forma equivalente a la de movilidad de personas.

Modelos de análisis

Los modelos de análisis de este capítulo consideran como premisa que las medidas del Poder Ejecutivo comprenden fases de restricción y apertura a lo largo del período analizado, y su cumplimiento incluye diversas medidas con alcance nacional, cantonal o distrital. Por consiguiente, los patrones de movilización de las personas se han visto afectados, entre otros factores, por esas fases y por su aplicación territorial.

Se utilizaron modelos de estadística espacial y técnicas del “big data”, con el fin de identificar las variaciones de intensidad de esa movilidad por cantón para Costa Rica. Se definen tres momentos

para la comparación: lo ocurrido durante todo el 2019, luego entre enero y febrero de 2020 antes de la pandemia, y en los meses durante la pandemia, específicamente de marzo a setiembre.

La siguiente etapa fue la aplicación de modelos espaciales con el índice de Morán⁶, para identificar la existencia de conglomerados de cantones con alta intensidad de congestión vial durante los meses del año 2020. Finalmente, se exploró la relación entre Waze, como proxy de movilización, y la cantidad de contagios por cantón. Se realizaron, calibraron y compararon modelos de regresión lineal y de Poisson con ajustes (modelo de regresión binomial negativo) que contemplaron los datos del Ministerio de Salud (Minsa) sobre contagios por covid-19, los de congestión de Waze e indicadores sociodemográficos para toda la serie de tiempo con base en los registros del Instituto Nacional de Estadística y Censos de Costa Rica (INEC) para la mayoría de los casos. Los modelos incluyeron pruebas con distintos rezagos entre la medición de Waze y los datos de contagio del Minsa, por cuanto el virus tiene un período de incubación de entre ocho y quince días, aproximadamente.

Hipótesis de trabajo

Los análisis efectuados en este capítulo tienen una lógica secuencial de acuerdo con una serie de hipótesis de trabajo asociadas a las cuatro preguntas de investigación formuladas al inicio de esta sección. Los análisis empíricos se orientaron a aportar evidencia para rechazar, o no, estas hipótesis.

Como primera hipótesis se señala que la base de datos de Waze es una medición indirecta que permite, con ciertos grados de confiabilidad, aproximar el flujo de transeúntes en el territorio nacional. Con base en ello se plantea que la pandemia y las medidas tomadas por el Gobierno modificaron los flujos de tránsito de forma diferenciada localmente, según las variaciones en el *ranking* de congestión vial por cantón.

La segunda hipótesis es que las restricciones de tránsito y comercio por la pandemia redujeron significativamente la intensidad de la congestión vial entre

marzo y setiembre en todo el país, comparada con lo ocurrido entre enero y febrero de 2020, y lo registrado durante todo el 2019. Las variaciones en la intensidad no fueron constantes, pues las medidas del Gobierno estuvieron sujetas a fases de cierre y apertura que variaron a lo largo de período en estudio.

La tercera hipótesis sugiere que la pandemia alteró la relación territorial de la congestión vial en ciertos cantones de la GAM. En efecto, hay conglomerados de alta congestión como producto del flujo de personas que, por trabajo y educación entre los principales motivos, transitan por la GAM. Ese comportamiento estaría asociado, parcialmente, a la movilidad entre cantones. No obstante, se postula que, durante los meses en pandemia analizados, la movilización se redujo al punto que desaparecieron los flujos de personas más considerables. Por lo tanto, se desvaneció la relación territorial de la congestión vial. Desde ese punto de vista, las medidas estarían cumpliendo su objetivo de limitar el movimiento de personas para reducir la incidencia de contagio por covid-19.

Una cuarta hipótesis es que las mediciones de Waze, junto a otras variables sociodemográficas, están relacionadas con la cantidad de contagios por covid-19 según cantón. Mucha congestión vial hoy en un cantón estaría asociada con una mayor incidencia de casos en dieciséis días. Lo mismo ocurre con otras variables relacionadas con comportamientos sociales en zonas de la GAM con características sociodemográficas específicas, que podrían atentar contra las medidas de distanciamiento físico para el control de la pandemia.

Tanto la comprobación como el rechazo de estas hipótesis brinda evidencia que puede ser útil para definir medidas de política pública orientada a disminuir la incidencia de contagios en zonas específicas. Además, con alguna certidumbre sobre el comportamiento por cantón, la información también podría dar pistas sobre cómo realizar cierres donde hay focos de contagio, y mayor apertura donde no los hay, con el fin de afectar lo menos posible la actividad económica del país.

La pandemia modifica las zonas de mayor congestión vial

A finales del año 2019, el PEN presentó públicamente una aplicación con indicadores cantonales llamada “Dcifra tu cantón”, la cual puede ser accedida en la dirección www.dcifra.cr. Incluyó, entre otros análisis, una evaluación exhaustiva de la congestión vial de cada uno de los 82 cantones del país. A partir de ese estudio, en esta edición del Informe se actualizaron los datos con lo sucedido en 2019 y 2020, y se hizo un *ranking* comparado de los municipios con mayores problemas de tránsito. El objetivo fue determinar los cambios relativos en la congestión durante la pandemia, como una primera medición general de lo ocurrido.

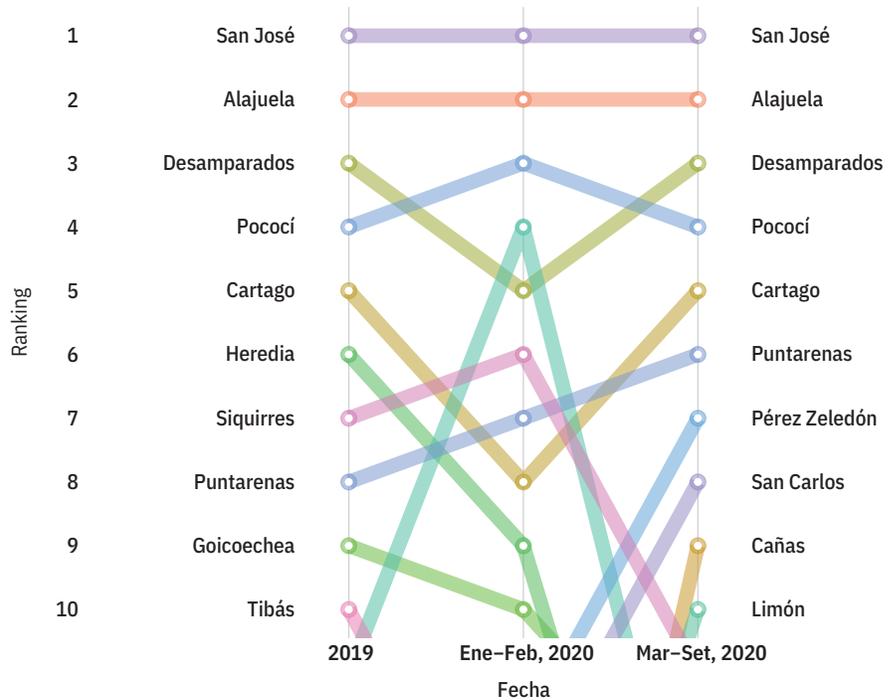
Según los resultados, un grupo de cantones presentaron la mayor cantidad de flujos de tránsito en el 2019 y se mantuvieron en las primeras posiciones durante el 2020, incluso en los meses de pandemia. En este grupo destacan San José, Alajuela y Desamparados. Son cantones con alta densidad de población, cabeceras de provincia o con alta concentración de conexiones de transporte público y flujo de trabajadores, como en el caso de San José. También sobresale en los primeros lugares de congestión el cantón de Pococí en Limón, lo cual puede estar vinculado con la insuficiencia de la ruta 32 para abarcar el flujo vehicular por esa zona, sumado a los trabajos de ampliación que han generado cierres y trabajos en la vía durante los últimos meses.

Un segundo grupo de cantones tuvieron flujos regularmente altos en 2019, pero redujeron su importancia relativa durante el 2020, en particular con las medidas de restricción. Entre estos están el cantón central de Heredia, Goicoechea y Tibás. Los tres se ubicaron entre los diez con más congestión en 2019; sin embargo, durante el 2020, el flujo vehicular disminuyó de manera significativa al compararlo con otras zonas del país.

Por último, el tercer grupo es el de municipios que no estuvieron en las primeras posiciones durante el año 2019, pero sí aparecen con altos flujos en el 2020, en los meses de pandemia. En este caso se encuentran Pérez Zeledón, San

Gráfico 7.2

Cambios en el ranking de cantones con mayor congestión vial (diez primeros lugares)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

Carlos, Cañas y Limón. El gráfico 7.2 representa los movimientos en el *ranking* de mayor congestión en tres momentos: el 2019, entre enero y febrero de 2020 antes de la pandemia, y entre marzo y setiembre durante la pandemia.

Lo ocurrido en estos cantones da una primera evidencia de las variaciones en la movilidad en período de pandemia. Para precisar este elemento, considérese el caso de San José como ejemplo. Se mantiene como el cantón con mayor congestión hasta el momento, pero se observa una reducción importante de la actividad cuando se comparan las condiciones de todo el 2019 con lo registrado entre marzo y setiembre de 2020, en período de pandemia. En promedio, la congestión se redujo un 66% en este cantón.

Por segmento de horario destacan los cambios en la hora pico de la tarde. En 2019 el segmento horario de mayor congestión sucedió entre 4 p. m. y 6 p. m., que coincide con el fin de la jornada laboral y de clases en el sistema educativo

público. Durante marzo y setiembre se mantuvo la hora pico de la tarde, pero se movió al segmento entre 3 p. m. y 5 p. m., que concuerda con el límite de hora permitido para circular cuando se aplicaron las restricciones más fuertes.

El mapa 7.1 muestra, de acuerdo con la intensidad de los colores, la reducción del tránsito en los dos momentos para el cantón central de San José. Nótese la disminución en el casco central y en la zona este. Se mantienen alto en las rutas de la General Cañas que lo conectan con Alajuela y Heredia, el entronque con la vía de circunvalación a la altura del monumento del Agua en la Uruca y en general en toda la ruta 39 conocida como “la circunvalación”.

Fuerte reducción de la congestión vial en todo el territorio durante la pandemia

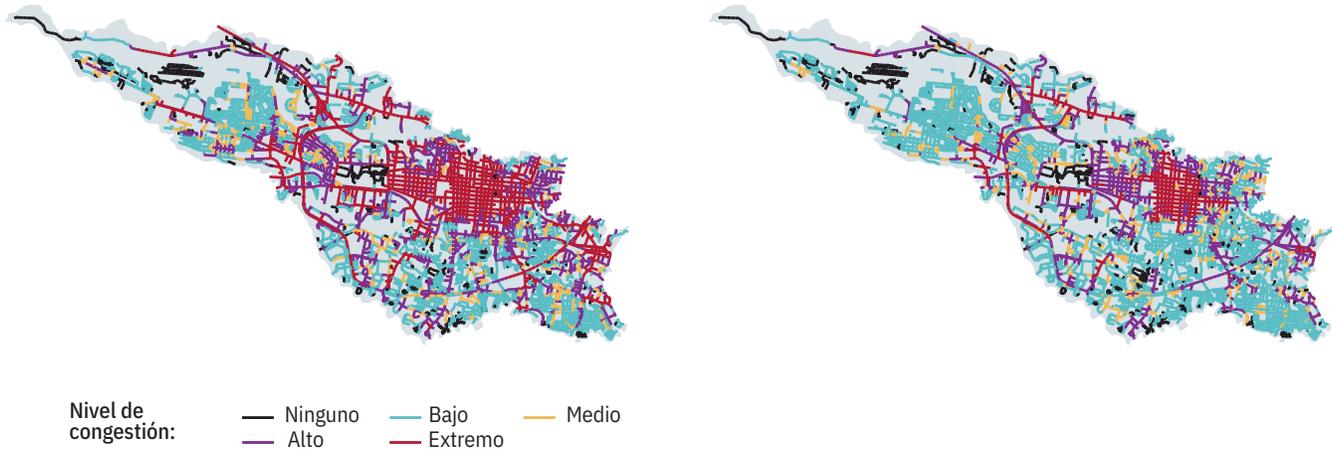
El ejemplo de San José da pie para estimar la reducción de la congestión en todo el país y valorar el impacto de las medidas

Mapa 7.1

Cantidad de jams promedio por segmento^{a/} en el cantón central de San José, según nivel de intensidad

San José, 2019

San José, marzo–setiembre, 2020



a/ Se entiende por jams el conteo de congestión por segmento de carretera (cien metros aproximadamente) reportado por la aplicación Waze para un lapso de tiempo definido.

Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

restrictivas de tránsito y de operación de comercios como efecto agregado (recuadro 7.3). El comportamiento general de los datos muestra que, en el período de la pandemia, se redujo significativamente la congestión vial, debido, entre otros aspectos, a esas medidas impuestas por el Gobierno desde el 12 de marzo. No obstante, al observar los indicadores que miden la temporalidad y ubicación de los flujos de tránsito, el comportamiento no es uniforme.

Entre las tendencias que se mantienen antes y durante la pandemia, en todo el territorio nacional, está la mayor congestión relativa durante los viernes, y también en la hora "pico" de la tarde, entre 4 p. m. y 7 p. m. Por supuesto, los niveles de intensidad son distintos entre enero-febrero y los meses posteriores, pero los datos evidencian que en general la pandemia, aunque redujo sustancialmente esa intensidad, no anuló por completo la movilidad de personas en esos dos momentos. En el gráfico 7.3 se aprecia ese comportamiento de manera comparada.

Si bien los niveles de cada mes se

redujeron después de febrero, la estacionalidad de los viernes se mantiene durante la pandemia. Como se observa en el gráfico 7.4, el único mes en el cual desaparece ese patrón por día de la semana es abril, que coincide con la Semana Santa, cuando se aplicó una fuerte restricción vehicular en todo el territorio nacional. Se requiere un mayor análisis para detectar la dirección de esos flujos los viernes, los puntos de concentración y hacia dónde se dirigen, con el fin de determinar posibles impactos.

Sí se observan diferencias entre lo ocurrido antes y durante la pandemia en otros indicadores. Por ejemplo, la hora "pico" de la mañana, entre 6 a. m. y 7 a. m. aproximadamente, es evidente en febrero, situación coincidente con el inicio del calendario escolar del Ministerio de Educación Pública (MEP). Se mantiene en marzo y desaparece en abril de 2020. Por orden del Poder Ejecutivo, a partir del 17 de marzo se suspendieron las clases presenciales en centro educativos públicos y privados. En ambos casos se continuó con métodos de enseñanza a distancia. En mayo, junio, julio, agosto

y setiembre no se registran flujos importantes de congestión en el segmento de hora de las mañanas.

El tiempo de atraso promedio es una manera de medir momentos de alta congestión. En los puntos de mayor atasco se espera un incremento significativo de los tiempos promedio de viaje. Al obtener este indicador por mes y hora, se refleja con mayor detalle el efecto de la hora "pico" en la mañana durante febrero y marzo y luego cómo desaparece a partir de abril (gráfico 7.5).

Este breve repaso, con los datos de congestión de Waze, revela el efecto de las restricciones impuestas por el Gobierno para reducir la movilidad de las personas durante la pandemia, junto a otros aspectos no medidos aquí. Esas medidas no han sido constantes en el tiempo y, por consiguiente, sus efectos son diferenciados. Las primeras acciones se tomaron a mediados de marzo y se endurecieron en abril, durante la Semana Santa, para evitar fuertes flujos de personas, comportamiento normal en esas fechas.

En términos comparados, la serie de datos evidencia que el inicio del año

Recuadro 7.3

Restricciones emitidas por el Ejecutivo ante la pandemia

El 6 de marzo de 2020 se confirmó el primer caso de covid-19 en Costa Rica. A partir de ese momento comenzó una estrategia de medidas decretadas por el Gobierno para controlar los efectos de la pandemia en el país. Desde el inicio se plantearon diversas fases de restricción y apertura para el tránsito vial, comercios y actividades sociales y masivas. Con el tiempo, esas fases iniciales fueron modificadas por las autoridades de acuerdo con los efectos del virus en distintos momentos y lugares del territorio. El objetivo central era desincentivar las aglomeraciones y la movilidad de personas para reducir el contacto entre ellas.

Las medidas, en general, han tenido una vigencia promedio de dieciséis días. Algunas de las que permanecieron por más tiempo fueron el cierre obligatorio de templos y cultos religiosos, seguido por el cierre total de playas en todo el país y las restricciones vehiculares que han estado sujetas a diferentes cambios a lo largo de los meses.

Con la detección del primer caso de covid-19 empezó la primera fase de restricción. La campaña “Quédate en casa” empezó el 9 de marzo, cuando se declaró alerta amarilla. Un día después se instruyó el teletrabajo al sector público. Los primeros efectos sobre la congestión fueron evidentes: a partir de la segunda semana de marzo el cambio porcentual de la congestión es negativo y decreciente para todo el mes. En abril continuaron las medidas restrictivas con un endurecimiento en las disposiciones para el tránsito de vehículos y la actividad comercial durante la Semana Santa, entre el 6 y el 12 de ese mes. Según las mediciones de Waze, hasta setiembre de 2020, esta ha sido la semana de menor movilidad de personas durante la pandemia.

En mayo se moderó la estrategia y se pasó a la primera fase de apertura de actividades económicas, pero con capacidad restringida. Durante este mes predominaron las medidas de intensidad media, las cuales generaron un leve incremento en la congestión vial.

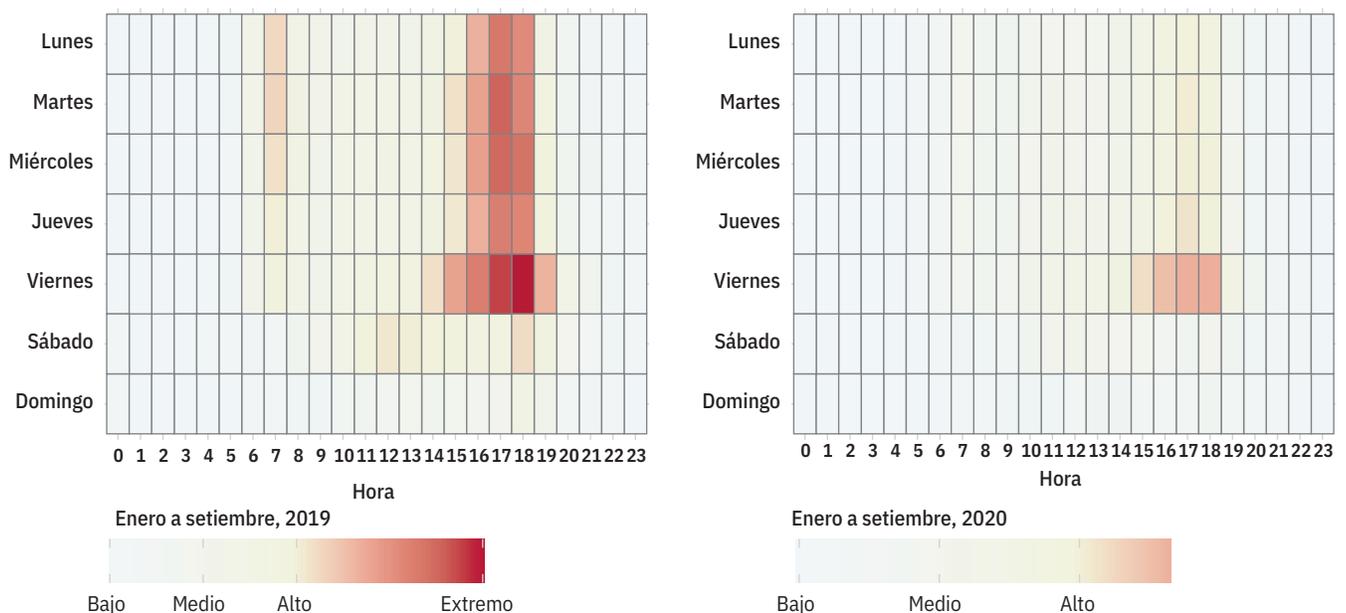
Luego vino otro período de endurecimiento relativo durante junio y se tomaron medidas diferenciadas para los cantones en alerta naranja. En el transcurso del mes, y conforme aumentaron los niveles de contagio, se sumaron múltiples cantones a la lista de zonas en alerta naranja. Cantones en esta condición tuvieron medidas de restricción más severas con respecto a otras zonas del país, debido al aumento de los casos de contagio o por estar ubicados en áreas de alta vulnerabilidad de contagio. Del 20 al 22 de junio se instruyó un esquema más restrictivo sobre actividades económicas.

En julio, se destacó la declaratoria precautoria de transmisión comunitaria para la GAM y la puesta en práctica del cerco epidemiológico entre el 12 y el 19. Esta fase fue tan restrictiva como la Semana Santa. Durante el resto del mes se mantuvieron importantes restricciones viales para los cantones en alerta naranja.

Fuente: Gómez Campos et al., 2020, con datos del Ministerio de la Presidencia, 2020.

Gráfico 7.3

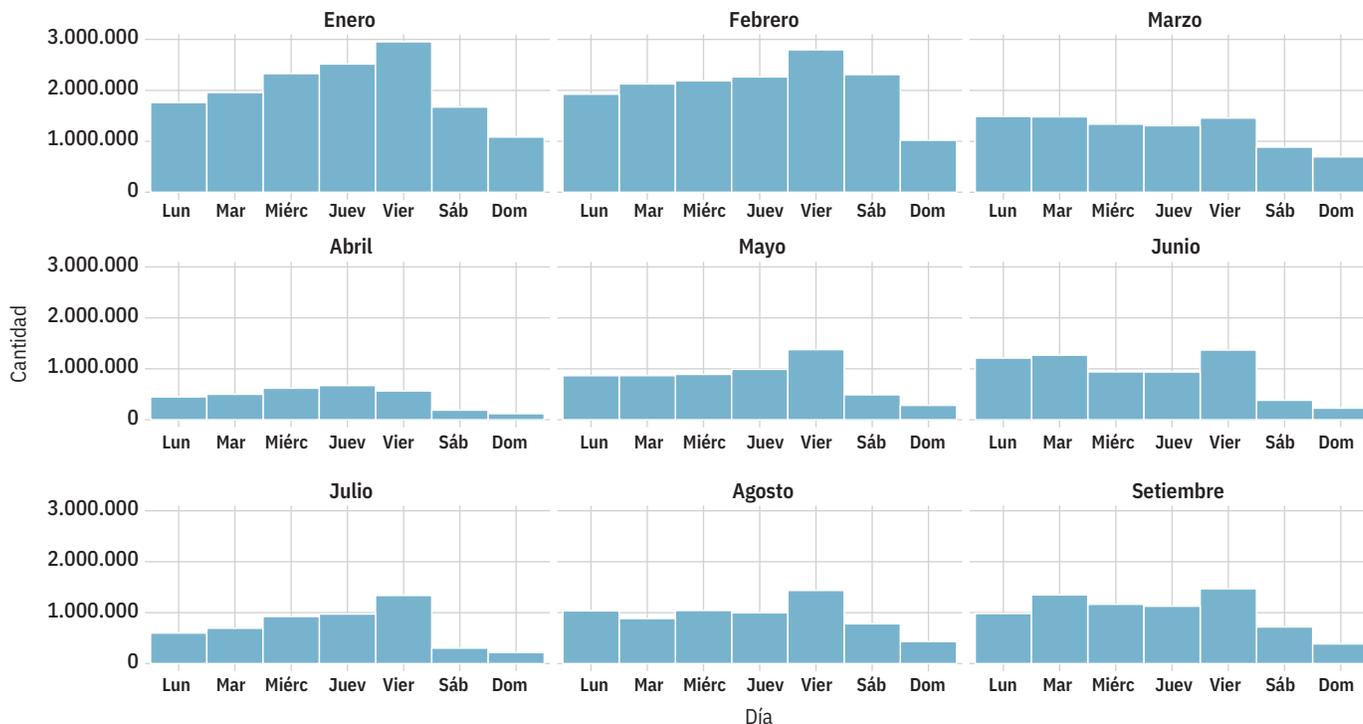
Nivel de congestión vial por día de la semana y hora



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

Gráfico 7.4

Congestión por día de la semana y mes. 2020

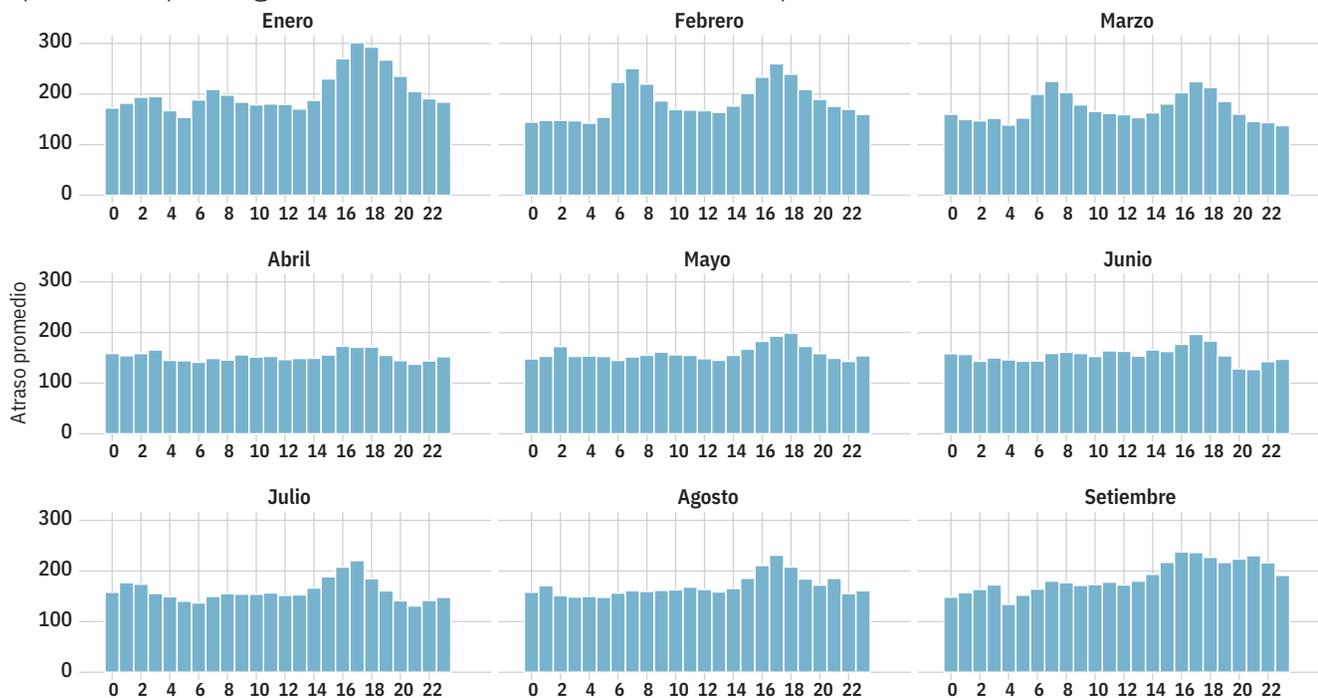


Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

Gráfico 7.5

Atraso promedio en segundos, por hora del día y mes. 2020

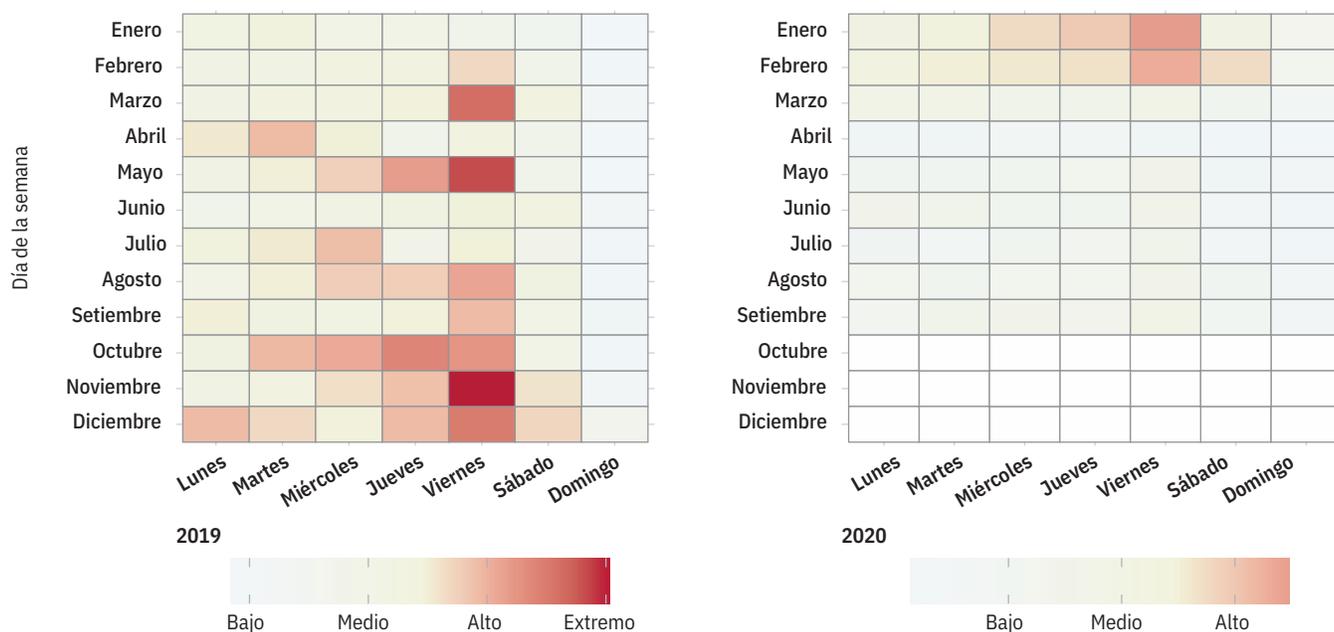
(promedio por segmento de carretera de 100 metros aproximadamente)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

Gráfico 7.6

Nivel de la congestión vial por día de la semana y mes. 2019 y 2020



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

2020 tuvo mayor congestión en enero y febrero que en el 2019, pero la pandemia y las medidas de restricción modificaron el comportamiento siguiente. El gráfico 7.6 muestra las diferencias entre ambos años por mes y día de la semana, y como se indicó anteriormente, es clara la estacionalidad de la congestión: mayor movilidad los viernes. Nótese el cambio a partir de marzo del 2020 con respecto a lo sucedido en el 2019.

¿Cuánto se redujo la congestión de marzo a setiembre, en medio de la pandemia, con respecto a período previo? Para medirlo, se estableció un promedio de congestión diaria considerando los datos de enero y febrero de 2020, para usarlo como base comparativa con lo ocurrido después.

De acuerdo con los resultados, hasta Semana Santa la reducción de la movilidad de personas alcanzó un 90% de los niveles registrados en enero y febrero de 2020. Posteriormente se flexibilizaron las medidas y la movilidad creció hasta un 40%. No obstante, conforme crecieron los casos de contagio por covid-19,

el Poder Ejecutivo tomó nuevas acciones más estrictas. La más importante, hasta setiembre de 2020 que cubre este análisis, sucedió entre el 11 y el 19 de julio. En esas fechas se decretó un “cerco epidemiológico de la GAM”. Eso significó que todos los cantones de esta zona se catalogaron en alerta de emergencia naranja y entraron a regir medidas más estrictas para la movilidad y la operación de comercios.

El gráfico 7.7 ilustra los efectos de esas medidas tomadas por el Gobierno, así como la reducción del tránsito en el período de análisis. La conclusión es clara: las medidas restrictivas, junto a otros comportamientos de autocontención, redujeron la movilidad de personas en medio de la pandemia. Valga decir que este es un elemento central para contener epidemias.

Los niveles medidos por Waze concuerdan, en la tendencia, con los datos de Google sobre movilidad de personas en Costa Rica durante la pandemia. Esto confirma una de las hipótesis de este capítulo: los datos de congestión vial son

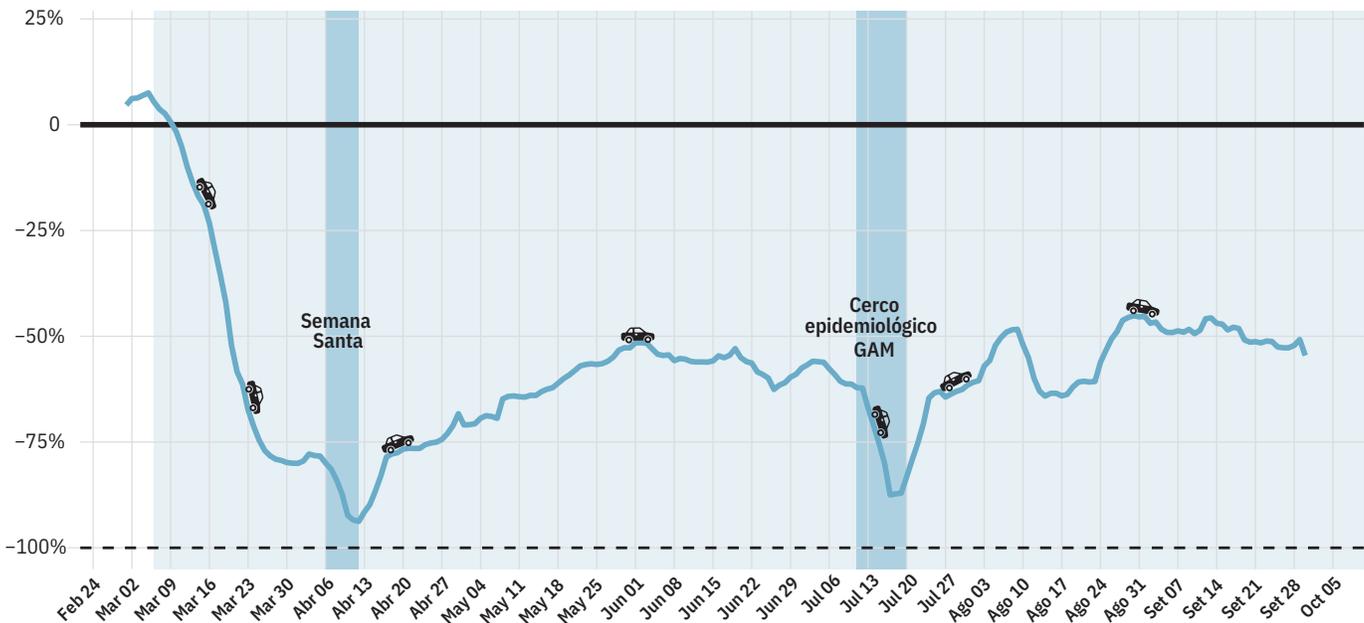
un *proxy* para medir la movilización de las personas. Las razones están relacionadas con la forma automática de generación de datos con tecnología celular de Google y Waze, explicadas al inicio de este documento.

Google registra datos agregados por país para lugares específicos. Por ejemplo, zonas de comercio y recreación, como teatros, cines, restaurantes y bares. También, tiene datos para medir cómo ha variado el desplazamiento de las personas en comercios de abastecimiento, como supermercados y farmacias. Además, cuenta con registros de las zonas de paradas de transporte público, entre otros.

Esos datos indican reducciones hasta de un 80% en Semana Santa y casi un 70% en los días del cerco epidemiológico en la GAM durante julio. Estos dos momentos de mayor restricción o autocontención de las personas a salir coinciden con las mediciones de Waze y corroboran las semejanzas entre esta y Google (gráfico 7.8).

Gráfico 7.7

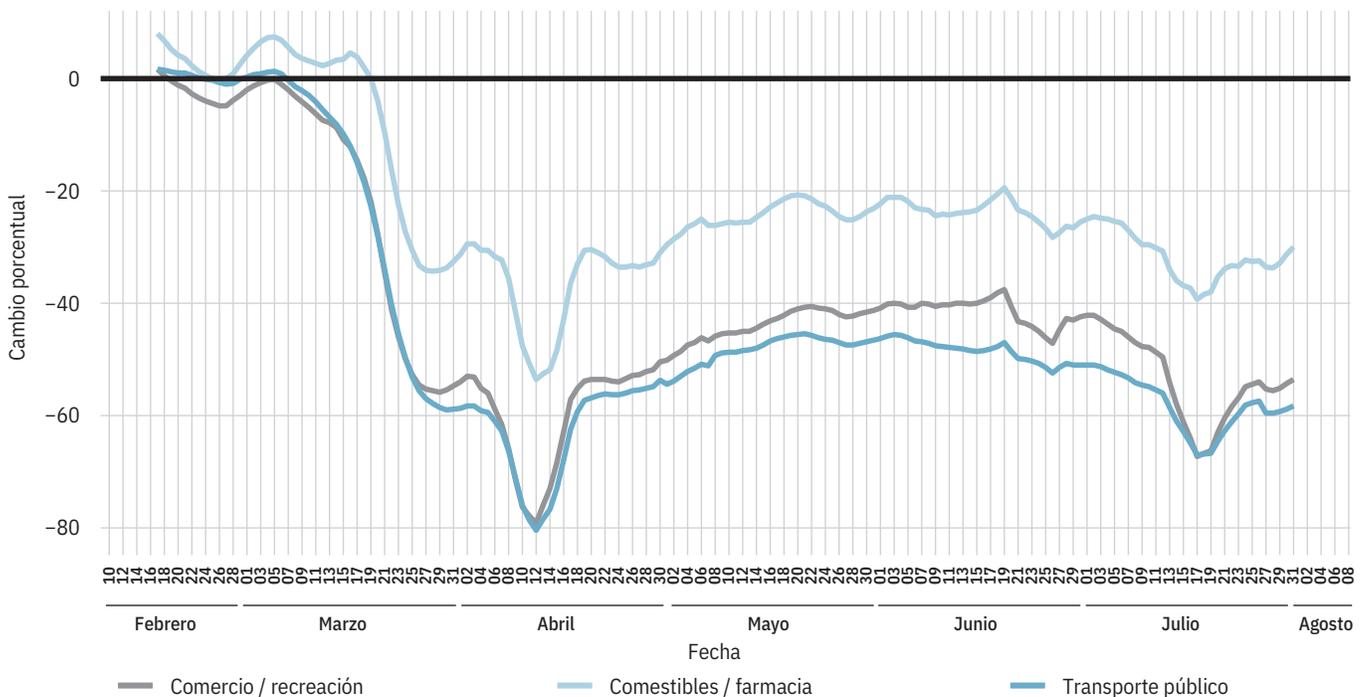
Cambio porcentual diario en la congestión vial con respecto a enero-febrero de 2020
(promedio móvil de siete días)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

Gráfico 7.8

Cambios en la movilidad para Costa Rica con base en Google Mobility Data



Fuente: Gómez Campos et al., 2020, con datos de Google.

Medidas de restricción borraron los patrones territoriales

De los datos se desprende que tanto la intensidad de la congestión como la movilidad de las personas se redujeron considerablemente a partir de marzo de 2020, con períodos de restablecimiento relativo. La pregunta relevante por contestar ahora es si eso generó un cambio en los patrones territoriales de desplazamiento de la población. En el análisis subyace la hipótesis de que las restricciones dirigidas a reducir el tránsito de manera importante procuran evitar la diseminación del virus a lo largo del país.

Tomando como referencia los datos de 2019, es posible hallar patrones espaciales de la congestión ubicados, mayoritariamente, en zonas dentro de la GAM. Son focos de fuerte congestión que, además, tienen alta proporción de desplazamiento de personas del cantón para ir a trabajar, de acuerdo con la Encuesta Nacional de Hogares (Enaho) y los datos del último censo de población del INEC. Al combinarse tales factores, muy probablemente ese conglomerado de cantones de la GAM con alta congestión se deba a flujos de tránsito intercantonales con afectación recíproca.

Para establecer estas relaciones, se elaboraron modelos estadísticos espaciales con el índice de Morán global, y una prueba para confirmar que la relación espacial es significativa, denominada Test de Monte Carlo, con base en la metodología de Medina y Solymosi (2019). En los períodos donde la relación espacial es significativa, se ejecuta el índice local de Morán como medida local de asociación espacial (*local indicators of spatial association*, LISA, por su sigla en inglés) para determinar cuáles cantones conforman estos conglomerados.

El primer paso fue construir un modelo con el promedio mensual de congestiones durante el 2019, para determinar el estado “normal” del tránsito y cuáles zonas constituyen el conglomerado de alta congestión. Luego se ejecutó el mismo modelo para cada uno de los meses del 2020. De esta forma, se corroboró el comportamiento en el período previo al inicio de la pandemia –enero y

febrero–, así como la comparación posterior cuando se aplicaron las medidas de restricción. Se usan los datos mensuales como criterio de corte para tratar de aislar los efectos de las distintas medidas en este tiempo.

Cabe recordar que las regulaciones impuestas por el Poder Ejecutivo han variado en el tiempo, pero también en el nivel de afectación para cada cantón, con restricciones diferenciadas por distritos y zonas específicas como la GAM y la zona norte del país, por ejemplo. La evaluación por mes de estas proporciones proporciona un panorama más preciso de lo que está ocurriendo. Es importante señalar que esta es una aproximación inicial, por lo cual deben explorarse otros modelos complementarios para contrastar los hallazgos de este estudio.

Según los resultados del análisis de los datos del 2019, existe un único conglomerado significativo con ocho cantones que registran alta congestión. Estas zonas presentan una relación espacial significativa. En otras palabras, es ahí donde están los flujos viales más relevantes del país. Son seis casos de la pro-

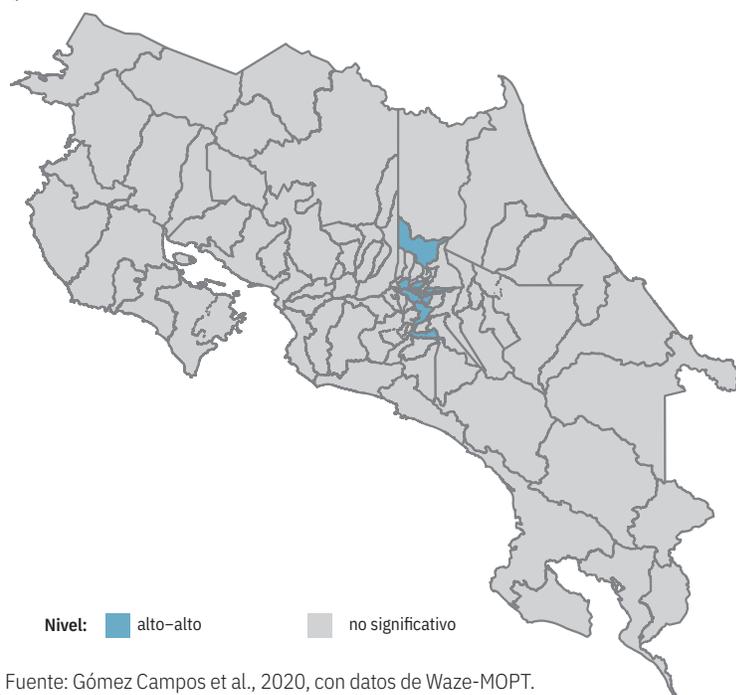
vincia de San José: cantón central de San José, Desamparados, Goicoechea, Tibás, Montes de Oca y Curridabat, y otros dos de la provincia de Heredia: el cantón central y Santo Domingo. Todos estos municipios están ubicados en la GAM.

De acuerdo con los datos del Censo de Población del 2011, con actualización de proyecciones de población al 2020, en esos ocho cantones, en promedio, un 53% de la población ocupada residente en ellos sale regularmente del cantón, pues su lugar de trabajo se ubica en otras zonas del país⁷. Estos datos dan sustento a la tesis de que los municipios con mayor congestión comparten características demográficas y presiones viales que generan afectaciones mutuas. El problema radica en que esos flujos intercantonales pueden provocar riesgos crecientes, en medio de la pandemia, de esparcimiento del virus, sobre todo si eso ocurre infringiendo las medidas de distanciamiento físico. En el mapa 7.2, se observa la zona de afectación según el promedio mensual de congestión vial de Waze considerando los datos de todo el 2019.

Mapa 7.2

Conglomerado de alta congestión vial con base en el índice de Morán, 2019

(promedio mensual)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

¿Se ha modificado esta situación durante el 2020, específicamente entre marzo y setiembre, a raíz de las restricciones impuestas por el Poder Ejecutivo y la atención a la pandemia? La respuesta es afirmativa: la restricción alteró de manera importante el comportamiento del tránsito en todo el país, al punto de borrar los flujos antes descritos. Durante enero, febrero y marzo se mantuvo el conglomerado de cantones con alto nivel de congestión, aunque con variaciones en la composición de ese grupo. A partir de abril desaparece el patrón territorial de la congestión. En ninguno de los meses siguientes la relación espacial fue significativa, de acuerdo con los análisis realizados. El cuadro 7.1 muestra los resultados del test de Monte Carlo para determinar si la relación espacial es significativa en cada uno de los meses del 2020.

Nótese que en junio de 2020 la relación espacial no es significativa por muy poco margen, de acuerdo con el “p-value” de la prueba estadística⁸. Este período coincide con un período de apertura moderada y de aumento relativo de la movilidad de las personas que, por poco, vuelven a generar conglomerados de congestión

Cuadro 7.1

Nivel de significancia^{a/} del test de Monte Carlo sobre la congestión vial, por mes

(relación significativa en conglomerados espaciales)

Período	p-value	Significativo
2019	0,01283	Sí
Enero de 2020	0,03736	Sí
Febrero de 2020	0,01135	Sí
Marzo de 2020	0,03555	Sí
Abril de 2020	0,17882	No
Mayo de 2020	0,07974	No
Junio de 2020	0,05204	No
Julio de 2020	0,09841	No
Agosto de 2020	0,28530	No
Setiembre de 2020	0,22863	No

a/ La prueba estadística es significativa si el p-value es menor de 0,05.

Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

en el territorio, según los análisis realizados. Concuera, temporalmente, con el aumento relativo de casos de contagios en el país que, posteriormente, llevaron a las autoridades a decretar medidas de restricción más fuertes con el cerco epi-

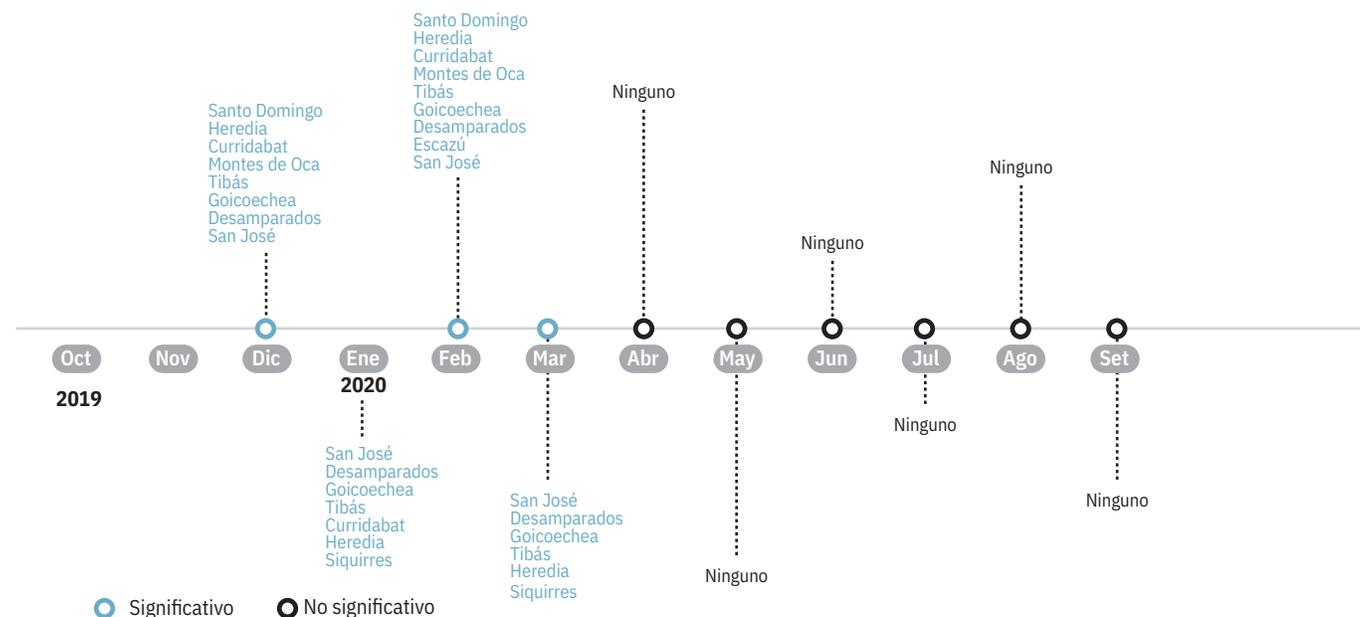
demiológico en la GAM, a mediados de julio, para tratar de controlar las tasas crecientes de contagio.

Como se aprecia en el gráfico 7.9, durante el 2019 y los primeros tres meses del 2020 los cantones donde hubo con-

Gráfico 7.9

Cantones que forman parte del conglomerado de alta congestión, por mes

(índice de Morán global y local)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT.

glomerados de alta congestión son, en su mayoría, los mismos: casi todos dentro de la GAM, con excepción de Siquirres. Son lugares con altas densidades de población, en su mayoría cantones dormitorio con una alta proporción de personas que trabajan en zonas vecinas y requieren desplazarse para ello.

Estos resultados evidencian que, en términos generales, la restricción disminuyó la intensidad de la congestión, pero también modificó los patrones de movilización de personas en todo el país. En ese sentido, al parecer las restricciones cumplieron su objetivo: reducir el desplazamiento de la gente en el territorio.

Factores asociados al comportamiento social en pandemia

Este apartado explora la asociación entre la congestión vial, como indicador proxy de la movilidad de personas, y otras variables sociodemográficas cantonales, con la cantidad de contagios por covid-19 en Costa Rica. Se usan dos modelos de regresión para el análisis, luego de calibrar parámetros (recuadro 7.4).

El principal hallazgo indica que variables como la congestión vial, la densidad de población, y la proporción de personas que salen a trabajar a cantones distintos de donde residen, entre otras, están asociadas positivamente con el contagio del virus. La hipótesis es que estas variables están vinculadas a conductas y a condiciones sociales que ponen en riesgo el cumplimiento de las normas de distancia física recomendadas por las autoridades de salud para prevenir las infecciones.

Los dos modelos seleccionados por su mejor ajuste, entre las doce combinaciones de parámetros utilizados, fueron el modelo lineal con el promedio móvil de casos nuevos de covid-19 por día con dieciséis días de rezago, y en el modelo binomial negativo con el promedio móvil de casos nuevos de covid-19 por día con dieciséis días de rezago. Es decir, en ambos procedimientos se validaron los mismos parámetros de medición para la variable dependiente y los días de rezago.

El cuadro 7.2 tiene los resultados estadísticos comparativos de los modelos. Se indican con asteriscos los efectos

Cuadro 7.2

Coefficientes de los modelos de regresión lineal y binomial negativo
(variable dependiente: casos nuevos de COVID-19 por cantón por día)

	Lineal	Binomial negativo
(Intercept)	-1573,12 *** (50,19)	-802,35 *** (9,91)
Jamstotal	0,84 *** (0,16)	0,12 *** (0,02)
Fecha	0,09 *** (0,00)	0,04 *** (0,00)
DensidadPoblacion	1,67 *** (0,14)	0,14 *** (0,02)
SectorPrimario	-1,40 *** (0,19)	-0,22 *** (0,03)
Escolaridad2011	-3,78 *** (0,23)	-0,02 (0,03)
PorcPob_NoAsegurada	0,41 *** (0,11)	0,49 *** (0,02)
PorcPob_Urbana	-0,23 (0,18)	0,17 *** (0,03)
Patentes_canton	1,11 *** (0,20)	0,53 *** (0,03)
Trabajadores_canton	4,36 *** (0,19)	-0,11 *** (0,02)
Mov_intercantonal	2,09 *** (0,17)	0,4 *** (0,03)
nobs	9021	9021
r.squared	0,41	
adj.r.squared	0,41	
sigma	8,31	
statistic	627,34	
p.value	0,00	
df	11,00	
logLik	-31896,69	-11852,45
AIC	63817,39	23728,91
BIC	63902,67	23814,19
deviance	622243,64	8042,41
df.residual	9010	9010
null.deviance		30215,55
df.null		9020
pseudo.r.squared		0,69
pseudo.r.squared.mcfadden		0,30

Todos los predictores continuos son centrados con el promedio y escalados con una desviación estándar.

*** p < 0,001, ** p < 0,01, * p < 0,05

Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT y Minsa.

Recuadro 7.4

Regresiones para el análisis de movilidad y contagios por covid-19

El análisis de regresión se realiza con información para el período de marzo a julio de 2020. Se utilizan dos modelos de regresión para este análisis, con doce combinaciones distintas en sus parámetros para la calibración. El primer modelo es uno lineal, que sirvió para identificar las relaciones principales entre las variables. En primera instancia, los modelos lineales indicaron que la variable *jams* y algunas otras socio-demográficas resultaban significativas usando casos nuevos y casos activos como variable respuesta.

Sin embargo, debido a la naturaleza discreta de las variables casos nuevos y casos activos, así como a la falta de normalidad y homoscedasticidad, se considera también el modelo de Poisson, por cuanto se ajusta mejor a los datos. Este último se recomienda cuando la variable respuesta (casos nuevos y casos activos, en este estudio) son un conteo, es decir, son igual o mayor a cero y son números enteros. Además, es un conteo a través del tiempo: cada día y semana tiene su propio valor. Al efectuar el análisis con el modelo de Poisson, se encontró un problema en los datos: presentó sobredispersión. En otras palabras, la media no es igual a la variancia, sino que esta es bastante mayor a la media, por un factor no necesariamente proporcional a la media. “Una forma de relajar esta restricción de igualdad media-varianza del modelo de regresión de Poisson es especificar una distribución que permita un modelado más flexible. En este sentido, el modelo paramétrico estándar para datos de recuento con presencia de sobredispersión es el modelo de Regresión Binomial Negativa” (Alcaide, 2015). Con base en esto se decidió ajustar el modelo utilizando una distribución binomial negativa. Este cambio ayuda a mejorar la estimación de los coeficientes ajustando el error de la estimación al comportamiento real de los datos.

La investigación se realizó de manera secuencial. Primero se probaron las variables (dependientes e independientes) y los parámetros para calibrar doce combinaciones en los modelos de regresión. Luego se escogieron los dos de mejor ajuste para ejecutar los análisis finales con la mejor combinación. Los parámetros estimados fueron los siguientes:

- **Temporalidad:** los datos se agruparon por día, por semana y con un promedio móvil a siete días.
- **Indicador covid-19:** se usaron casos nuevos y también casos activos por día.
- **Rezago:** utilizando el coeficiente de correlación de Pearson, se obtuvo la correlación entre los casos nuevos y los casos activos con los *jams* de Waze, con el fin de aproximar el nivel óptimo de días de rezago. Para casos nuevos, se obtuvo una correlación máxima a los dieciséis días y para casos activos a los dieciocho días.

Los modelos incluyeron las siguientes variables independientes:

- **Jamstotal:** promedio móvil de siete días de la cantidad de *jams* por día por cantón.
- **Fecha:** día del año.
- **DensidadPoblacion:** habitantes por km² de cada cantón con datos del INEC.
- **SectorPrimario:** proporción de empleos formales en el sector primario (o agropecuario) por cantón con datos del INEC.
- **Escolaridad2011:** porcentaje de la población con secundaria completa o más por cantón con datos del INEC.
- **PorcPob_NoAsegurada:** porcentaje de la población sin seguro social de la CCSS con datos del INEC.

- **PorcPob_Urbana:** porcentaje de la población que vive en zonas urbanas en el cantón con datos del INEC.
- **Patentes_canton:** total de patentes registradas en cada cantón con datos del Omipyme-UNED.
- **Trabajadores_canton:** cantidad de personas que trabajan en empresas formales de ese cantón con datos de la Revec-BCCR.
- **Mov_intercantonal:** porcentaje de personas ocupadas mayores de 15 años que trabajan en un cantón distinto de donde residen con datos del INEC.

Para seleccionar los modelos con el mejor ajuste, se usó el índice de información de Akaike (AIC). AIC se define como menos dos veces el logaritmo de la máxima verosimilitud más dos veces el número de parámetros (Akaike, 1974). Para el modelo lineal, se utilizó el R² como medida de bondad de ajuste. R² ofrece una aproximación al porcentaje de variancia de la variable respuesta, que es explicado por las variables predictoras en el modelo. En el caso del modelo binomial negativo, no se emplea este coeficiente, pero existen alternativas como el pseudo R² de McFadden. Este último utiliza el logaritmo de la verosimilitud. El cuadro 7.3 resume los resultados con todas las combinaciones realizadas para calibrar el análisis.

Cuadro 7.3

Parámetros de los modelos de regresión, según nivel de ajuste^{a/}
(se marcan en color los modelos con mejor ajuste)

Modelo	Temporalidad de datos	Indicador covid-19	Días de rezago	R ² -ajustado	Pseudo-R ² (McFadden)
Lineal (lm)	Diario	Casos nuevos	16	0,36	-
Lineal (lm)	Diario	Casos activos	18	0,34	-
Lineal (lm)	Semanal	Casos nuevos	16	0,39	-
Lineal (lm)	Semanal	Casos activos	18	0,32	-
Lineal (lm)	Promedio móvil: 7 días	Casos nuevos	16	0,41	-
Lineal (lm)	Promedio móvil: 7 días	Casos activos	18	0,34	-
Binomial negativo (glm)	Diario	Casos nuevos	16	-	0,21
Binomial negativo (glm)	Diario	Casos activos	18	-	0,17
Binomial negativo (glm)	Semanal	Casos nuevos	16	-	0,20
Binomial negativo (glm)	Semanal	Casos activos	18	-	0,11
Binomial negativo (glm)	Promedio móvil: 7 días	Casos nuevos	16	-	0,30
Binomial negativo (glm)	Promedio móvil: 7 días	Casos activos	18	-	0,18

a/ La selección se hizo con base en los coeficientes de ajuste. En el lineal se usó el parámetro R²-ajustado, y para el binomial negativo se utilizó el Pseudo-R² (McFadden) y el criterio de información de Akaike (AIC).

Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT y Minsa.

estadísticamente significativos para las variables independientes. Si el valor es positivo, significa que la variable respectiva tiene una asociación en la misma dirección que la variable de casos nuevos de covid-19 (dependiente). Por ejemplo, en zonas donde hay una alta congestión (JamsTotal_pm7), también hay un alto nivel de contagio del virus. Si el valor es negativo, ocurren relaciones inversas entre la variable independiente y la dependiente.

Los resultados generan evidencia empírica suficiente para aceptar la hipótesis de investigación: la asociación positiva entre un indicador adelantado, la congestión vial, medida por Waze para aproximar la movilidad de personas, y los casos de covid-19 por cantón. Se registran altos niveles de congestión vial en zonas que también presentan altos niveles de contagio del virus en términos generales. La relación es significativa en los dos modelos de regresión: el lineal y el binomial negativo. Es importante recordar que los indicadores de congestión de Waze tienen una alta correlación con los indicadores de movilidad que registra Google. Estos últimos se han utilizado, en dis-

tintos países, para estudiar la relación entre la movilidad y los contagios en los últimos meses, tema desarrollado en una sección anterior de este capítulo.

Asimismo, se encuentran asociaciones positivas entre los casos nuevos de covid-19 y la densidad de población, el porcentaje de población no asegurada por la CCSS, el total de patentes de empresas registradas en el cantón, el porcentaje de personas que laboran en cantones diferentes de donde residen y la fecha.

La hipótesis es que varias de esas variables tienden a dificultar el cumplimiento estricto de las medidas de distanciamiento. Mayor movilidad de personas, desde zonas con muchos trabajadores que deben trasladarse a otros cantones a laborar, así como municipios receptores que además tienen alta densidad de población y de parque empresarial instalado, pueden potenciar el acercamiento físico de las personas y crear condiciones más propicias para el contagio masivo. A eso se suman personas sin seguro de la CCSS que, generalmente, se desempeñan en el sector informal de la economía con condiciones más vulnerables, para quienes el distanciamiento social es difícil de cumplir.

Es importante subrayar que las relaciones arriba descritas no expresan causalidad, sino asociación. Dan información sobre fenómenos que pueden ser relevantes para entender comportamientos y condiciones sociales en medio de la pandemia y tomar decisiones basadas en ello. ¿Ocurren esas asociaciones en todos los cantones y de manera constante en el tiempo? Probablemente no: los resultados no son concluyentes; son una aproximación, validada en términos estadísticos, pero no determinante, en buena medida porque aún se desconocen muchas de las características del virus, la manera en que actúa y sus variadas formas de contagio.

Otras asociaciones entre variables consideradas en los modelos son negativas. Tal es el caso del sector primario: localidades donde hay más producción agropecuaria tienden a ser zonas con menor nivel de contagios por covid-19. De nuevo, esa relación no es de causalidad: solamente muestra que ambos hechos están presentes en un mismo territorio. Lo central es entender las dinámicas sociales en estas regiones, para comprender las razones por las cuales esos

fenómenos se presentan en conjunto. Una hipótesis es que son lugares con menor densidad de población y, por ende, donde es más fácil cumplir las recomendaciones de distanciamiento físico recomendadas por los expertos en salud para aminorar el contagio.

En el gráfico 7.10 se observan los coeficientes de los modelos de regresión. Si están a la derecha de la línea punteada, la relación es positiva con la variable de contagios por covid-19. A la inversa, si se ubican hacia la izquierda, la asociación es negativa.

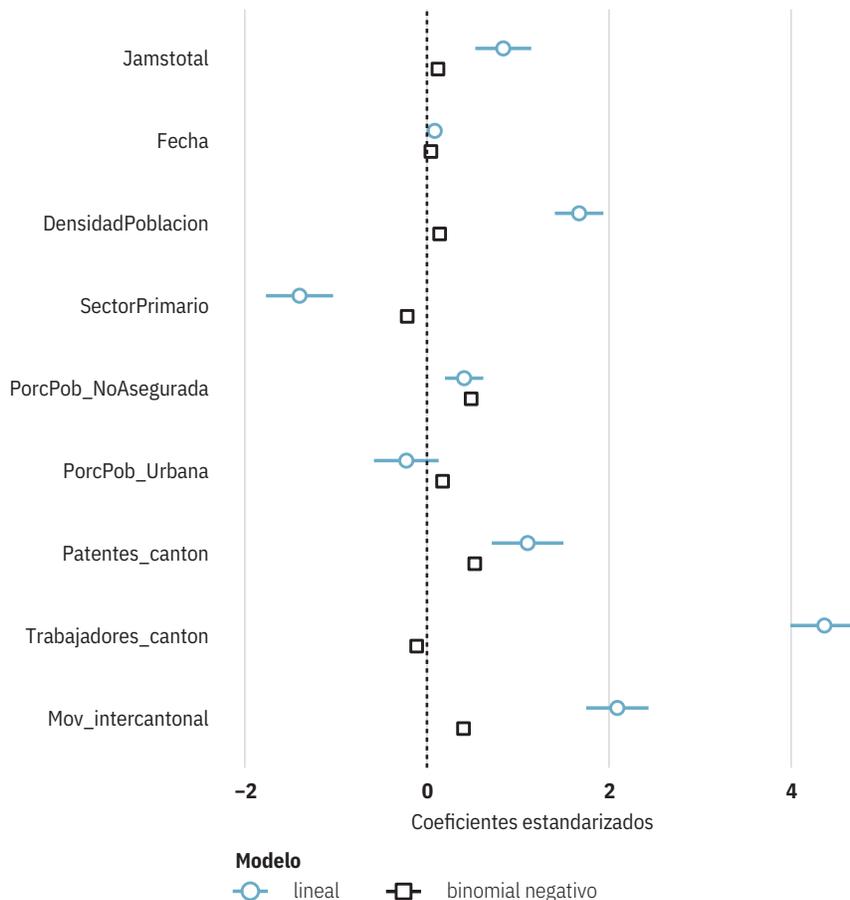
Por último, se exploraron territorialmente los cantones donde estas relaciones son más consistentes. Los resultados del modelo binomial, el de mejor ajuste a los datos disponibles, indican que la relación entre congestión vial y casos de covid-19 es estadísticamente significativa en 44 cantones (mapa 7.3), con un nivel de confianza del 95%. Esos cantones, organizados por provincia, son los siguientes:

- San José (8): Escazú, Aserrí, Mora, Santa Ana, Alajuelita, Tibás, Montes de Oca y Curridabat.
- Alajuela (9): cantón central, Atenas, Naranjo, Orotina, Zarcero, Valverde Vega, Los Chiles, Guatuso y Río Cuarto.
- Cartago (5): cantón central, La Unión, Turrialba, Oreamuno y El Guarco.
- Heredia (6): Santo Domingo, San Rafael, San Isidro, Belén, San Pablo y Sarapiquí.
- Guanacaste (5): Liberia, Nicoya, Bagaces, Carrillo y Cañas.
- Puntarenas (7): cantón central, Esparza, Montes de Oro, Osa, Quepos, Parrita y Corredores.
- Limón (4): Siquirres, Talamanca, Matina y Guácimo.

En los cantones donde esa relación no fue significativa, se necesitan otras estrategias para calibrar los modelos y las

Gráfico 7.10

Variables asociadas con casos nuevos de covid-19 por cantón, según modelo de regresión



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT y Minsa.

variables. Como parte de los hallazgos, se detectaron anomalías en las colas de la distribución de los datos que requieren mayor estudio. Tal es el caso del cantón central de San José. Tiene momentos de pocos contagios con mucha congestión y, viceversa, muchos contagios con poca congestión. No son la regla, pero constituyen casos extremos que impiden la precisión del modelo. Además, este cantón se ve influenciado por otras variables difíciles de captar en el análisis sin un preprocesamiento específico. Por ejemplo: es la capital del país, es una zona más de paso y menos de residencia de las personas. Parte de su alta congestión se explica por los flujos viales desde o hacia zonas aledañas que concentran parques empresariales y zonas residen-

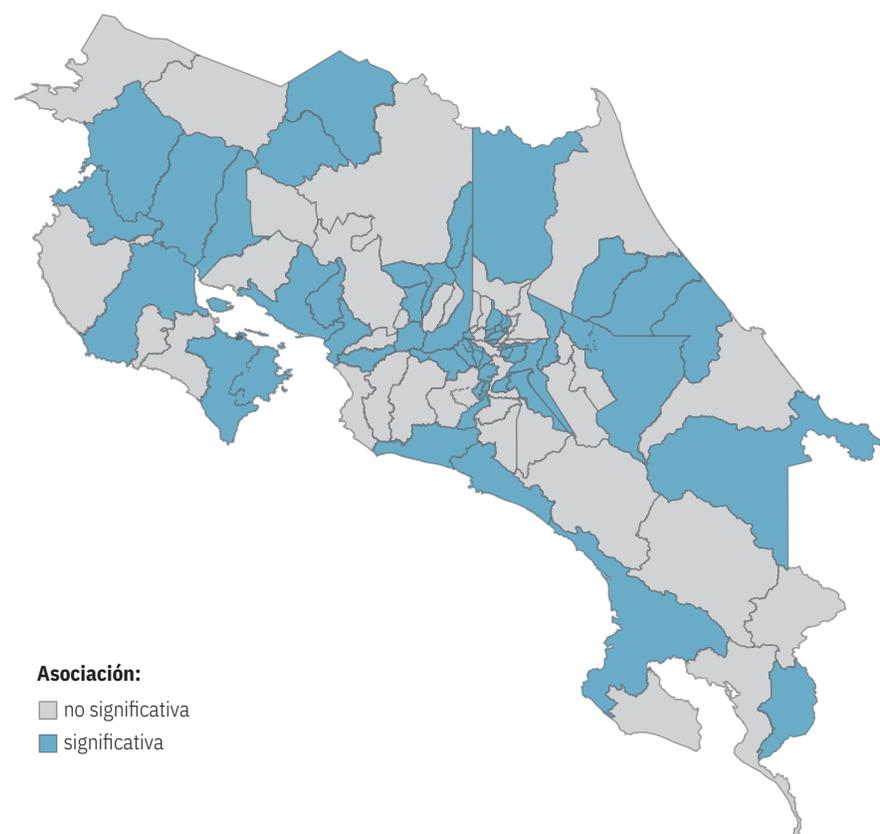
ciales. También, concentra una buena parte de las instituciones públicas, que en pandemia adoptaron modalidades de teletrabajo. En estos casos, los valores de congestión vial exigen un tratamiento más exhaustivo.

En otros casos, se encontraron indicios de estacionalidad en los datos. Por lo tanto, se optó por "suavizar" la serie de datos con promedios móviles, para reducir la influencia estacional de algunas variables. De ese modo, los modelos mostraron un mejor ajuste. Sin embargo, se deben utilizar otros procedimientos para precisar la magnitud de esas influencias o "ruidos" propios de los datos, y recalibrar los modelos. Eso es parte de la agenda pendiente de investigación en el tema abordado por este capítulo.

Mapa 7.3

Cantones con asociación significativa entre congestión vial y casos de covid-19

(modelo de regresión binomial)



Fuente: Gómez Campos et al., 2020, con datos de Waze-MOPT y Minsa.

Conclusiones

Los análisis efectuados en este capítulo proporcionan un nuevo acervo de información, inédita en el país, para explicar en qué medida la movilidad de personas, estimada de manera indirecta con Waze, y otras variables sociodemográficas están asociadas con los contagios de covid-19 en el ámbito local en Costa Rica. Los resultados obtenidos son aproximaciones a un tema complejo, dado lo reciente e incierto de la propagación y control del virus del covid-19. A pesar de esas dificultades, los hallazgos de este estudio demuestran la utilidad de las herramientas de “big data” e inteligencia de datos para los tomadores de decisión que requieren información actualizada sobre distintas variantes y asociaciones con el virus.

De acuerdo con las autoridades de salud, la principal recomendación para tratar de limitar el contagio del virus es reducir el contacto físico al máximo, mientras se obtiene más y mejor información sobre su forma de actuar y procedimientos médicos para curarlo. En un escenario ideal, sin movilización de personas y aislamiento completo el virus reduciría sustancialmente su capacidad de esparcimiento y letalidad.

En la práctica, ese escenario ideal no es posible, por muchas razones. Por ello, es clave entender cuánta cantidad de personas se movilizan en medio de la pandemia y a dónde. La relación de estas y otras variables sobre los casos de contagios locales también puede ser un insumo esencial para monitorear la situación en el tiempo. Este capítulo proporciona

insumos importantes para comprender el alcance y sentido de la relación entre estos aspectos.

Con poco más de 255 millones de registros de congestión vial por segmentos de carretera de aproximadamente cien metros en todo el país entre enero de 2019 y setiembre de 2020, se pudo constatar la validez de Waze como fuente de información para estimar la movilidad de las personas en tiempo real en el territorio nacional. Los datos de congestión vial de Waze se registran de manera automática por medio de información enviada por aparatos celulares que tienen instalada la aplicación, incluso aunque esta no se use. Esta forma de transmisión de datos también es usada por otras aplicaciones como Google, que han probado ser una buena herramienta para estimar la movilidad de personas en medio de la pandemia.

Además, este capítulo confirma la muy alta correlación entre las mediciones de Waze y Google para todo el territorio, lo cual permite validar los registros y aprovechar los microdatos de Waze para investigación en profundidad sobre territorios específicos, tema que no es posible examinar con los datos de Google.

Los análisis realizados generaron información para sustentar empíricamente las hipótesis de trabajo planteadas al inicio de la investigación. La evidencia para la primera hipótesis corrobora que, durante la pandemia, se modificaron las zonas de mayor congestión en el país. Con excepción de los cantones centrales de San José y Alajuela, que se mantienen como los de mayor tránsito antes y durante la pandemia, las demás localidades sufrieron cambios importantes. Zonas de alto tránsito, como Heredia, Goicoechea y Tibás, ubicadas en el *ranking* de los diez cantones con más tránsito durante el 2019, salieron de esa lista durante la pandemia, pues tuvieron una reducción considerable de sus flujos viales. En cambio otras zonas, como Pérez Zeledón, San Carlos, Cañas y Limón, subieron a los primeros lugares.

En concordancia con lo planteado en la segunda hipótesis, se constató que las restricciones sobre el tránsito y el cierre de comercios en el país ocasionaron reducciones considerables en los

flujos de personas en todo el territorio, junto a otros aspectos, como la posible autocontención individual a salir de las casas de habitación. Estas reducciones no fueron constantes en el tiempo. Se confirma un efecto mayor al inicio de la pandemia, entre marzo y abril, con disminuciones hasta de un 90% en Semana Santa, con respecto a los valores previos. Luego hubo fases de apertura y cierres de manera intermitente. Entre mayo y junio sucedió una recuperación del tránsito, cuando las autoridades permitieron el restablecimiento parcial del comercio y de la movilidad vial. Después se pasó a otra fase de fuerte restricción a la movilidad y comercios, entre el 11 y el 19 de julio, con el cerco epidemiológico en los cantones de la GAM. Y, finalmente, vino una nueva fase de apertura.

Los postulados de la tercera hipótesis de investigación también se validaron: de acuerdo con la evidencia, en medio de la pandemia desaparecieron los patrones territoriales de la congestión prevalecientes en 2019 y los primeros meses del 2020, sobre todo en ciertas zonas de la GAM con una alta proporción de personas que salen de su cantón diariamente para trabajar. Eso significa que las medidas de restricción cumplieron su objetivo de minimizar la movilidad de forma sustancial.

Además, la evidencia confirma la relación planteada en la cuarta hipótesis de investigación de este estudio, que también coincide con estudios similares en otros países, tal y como se reseñó en

el apartado metodológico al inicio de este capítulo: las mediciones de Waze, que estiman la movilidad de personas, junto a otras variables sociodemográficas, están relacionadas con la cantidad de contagios por covid-19 según cantón. En efecto, modelos de regresión con información local muestran factores sociales vinculados a mayores casos de contagio del virus por cantón. Esos factores tienen en común que podrían generar riesgos tangibles sobre el cumplimiento de las medidas de distanciamiento físico para reducir la propagación del virus.

Dos modelos de regresión, uno lineal y otro de Poisson con ajuste a uno binomial negativo, permitieron concluir que variables como la congestión vial (movilidad), la densidad de población, la proporción de personas ocupadas que viajan a trabajar a otros cantones, la cantidad de patentes (actividad económica) y la proporción de población no asegurada (vulnerabilidad laboral) están asociadas positivamente con los nuevos contagios diarios de covid-19 por cantón. Se trata de factores y condiciones sociales que, en medio de la pandemia, podrían afectar el balance de salud del país.

Un resultado interesante, es que las relaciones identificadas en los modelos de regresión no se cumplen en todos los cantones, ni tienen, cuando es posible identificarlas, la misma intensidad. Esto sugiere que los efectos de las restricciones sobre la propagación del virus pueden ser distintos dependiendo de factores contextuales, que aún deben ser mejor

investigados. También permiten concluir que restricciones generales, de afectación a todo el país, no serían igualmente efectivas.

Estos análisis muestran, además, la importancia de incursionar en nuevas fuentes de información masiva, suministradas por las tecnologías inalámbricas características del “internet de las cosas”, capaces de generar millones de datos en tiempo real, como en el caso de Waze o Google. Esas fuentes requieren, además, análisis sofisticados que combinan técnicas del “*big data*” y mucho poder computacional para la investigación. El resultado es la posibilidad de medir fenómenos complejos y de alta variabilidad en el tiempo, que de otra forma sería casi imposible estudiar en profundidad.

Finalmente, la agenda futura de investigación requiere seguir actualizando, durante el transcurso de la pandemia, la información que se presenta en este Informe. Ello permitirá no solo reevaluar estas primeras conclusiones conforme se disponga de más información sobre el virus, y otros factores sociales, sino también plantear recomendaciones más específicas sobre la manera de optimizar el efecto de las restricciones a los flujos poblacionales sobre el esparcimiento del covid-19. Asimismo, es necesario explorar otros métodos estadísticos, para contrastar los hallazgos con diversos enfoques y continuar con la precisión de las mediciones utilizadas, con el propósito de mejorar el ajuste de los modelos.

Investigadores principales: Steffan Gómez Campos y Mariana Cubero Corella.

Insumos: *Patrones de la congestión vial en tiempos de pandemia: una aproximación a la movilidad de personas con técnicas del “big data”*, de Steffan Gómez-Campos, Mariana Cubero Corella, María Fernanda Salas García, Kenneth Obando Rodríguez y Víctor Yeom Song.

Borrador del capítulo: Steffan Gómez Campos.

Coordinación: Leonardo Merino y Karen Chacón.

Edición técnica: Jorge Vargas Cullell, Leonardo Merino, Natalia Morales Aguilar, Karen Chacón y Steffan Gómez Campos.

Asesoría metodológica: Rafael Segura Carmona, Jorge Vargas Cullell, Natalia Morales Aguilar, Marcela Alfaro Córdoba y Fabio Sánchez Peña.

Actualización y creación de bases de datos: Karlissa Calderón Zúñiga, Rafael Segura Carmona, Steffan Gómez-Campos, Mariana Cubero Corella, María Fernanda Salas García, Kenneth Obando Rodríguez y Víctor Yeom Song.

Procesamiento de datos: Steffan Gómez Campos, Mariana Cubero Corella, María Fernanda Salas García.

Visualización de datos: Steffan Gómez Campos y Mariana Cubero Corella.

Lectores críticos: Marcela Alfaro, Fabio Sánchez, María Estelí Jarquín, Jorge Vargas Cullell, Leonardo Merino, Natalia Morales Aguilar.

Revisión y corrección de cifras: Rafael Segura Carmona.

Corrección de estilo y edición de textos: Mireya González Núñez.

Diseño y diagramación: Erick Valdelomar/Insignia Ng.

Un agradecimiento especial a Fabio Sánchez Peña y María Estelí Jarquín, quienes fungieron como lectores críticos de este capítulo, así como a Marcela Alfaro Córdoba, y al equipo del EpiMEC del Centro de Investigación en Matemática Pura y Aplicada (CIMPA) de la Universidad de Costa Rica (UCR) por sus comentarios metodológicos. Al Departamento de Informática del MOPT por el acceso a los datos de Waze para Costa Rica. Al Laboratorio Nacional de Computación Avanzada (CNCA) del Cenat-Conare por la asesoría técnica.

El taller de consulta se realizó el 3 de setiembre de 2020, con la participación de Rosario Alfaro, Manuel Alfaro, Ronald Alfaro, Jaime Allen, Raquel Arriola, Guido Barrientos, María Laura Brenes, Vera Brenes, Diana Camacho, Federico Cartín, Karen Chacón, Lenin Corrales, Mariana Cubero, Felipe De León, Esteban Durán, Marianela Espinoza, Diego Fernández, Steffan Gómez-Campos, Vladimir González, Miguel Gutiérrez, María Estelí Jarquín, Diana Jiménez, Arlene Méndez, Laura Mora, Ana Lucía Moya, Kenneth Obando, Andrea San Gil, Fabio Sánchez, Rafael Segura, Arturo Steinvorth, Leonardo Merino y Jorge Vargas Cullell.

Notas

1 Para más detalle, ver <https://www.waze.com/ccp/casestudies>.

2 El coeficiente de correlación de Pearson es una medida estadística del grado de asociación entre dos variables aleatorias cuantitativas. Varía entre -1 y 1. Si el indicador se acerca a 1 existe una correlación positiva, es decir, cuando una de las variables aumenta, la otra también lo hace. Si se acerca a -1 hay una correlación negativa, es decir, cuando una variable crece, la otra disminuye. Conforme el valor se acerca a 0, se reduce la posibilidad de que exista una relación lineal entre ellas.

3 Para más detalle, ver <https://www.google.com/covid19/mobility/>.

4 Para más detalle sobre los términos de privacidad y recolección de datos de la aplicación Waze, ver <https://www.waze.com/legal/privacy#information-that-is-being-collected>.

5 Para actualizaciones diarias se puede visitar la página <https://www.iadb.org/en/topics-effectiveness-improving-lives/coronavirus-impact-dashboard>.

6 El índice de Morán es un indicador para medir la autocorrelación espacial. Es un indicador multidireccional y multidimensional, que permite establecer relaciones en conjuntos de datos de alta complejidad por su estructura. Puede obtener valores entre -1 y 1. Cuando el valor es -1, significa que hay correlación perfecta negativa, es decir, existen conglomerados de valores distintos (también conocido como dispersión perfecta). Cuando se obtiene un valor de 1, entonces hay conglomerados espaciales perfectos de valores similares. Si se obtiene un valor de 0, significa que no hay autocorrelación en los datos, es decir, están distribuidos espacialmente de forma aleatoria.

7 La pregunta concreta es la siguiente: Para las personas con empleo, ¿en qué provincia y cantón se ubica el establecimiento/negocio/ finca/ institución para el que trabaja?

8 La prueba estadística es significativa si el *p-value* es menor de 0,05.