

Visualización del Impacto de las Medidas Gubernamentales en la Evolución del COVID-19

Mónica Mounier^{1,2}, Diana Salazar², y Eduardo Zamudio¹

¹Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones, Félix de Azara 1552, Posadas, Misiones, Argentina

²Facultad de Informática, Universidad Nacional de la Plata, Calle 50 y 120, La Plata, Buenos Aires, Argentina

monicamounier@fceqyn.unam.edu.ar, dianasalazarplata@gmail.com,
eduardo.zamudio@fceqyn.unam.edu.ar

Resumen. En los últimos meses, el mundo ha cambiado drásticamente debido a la aparición y rápida expansión de una nueva enfermedad denominada COVID 19. Esto es debido a la facilidad de contagio de dicha enfermedad, y además, a la existencia de un periodo asintomático el cual aumenta la velocidad de transmisión. Ante ello, cada gobierno ha implementado diferentes medidas, priorizando algunos sectores sobre otros, a fin de disminuir el impacto de la enfermedad en los mismos. El objetivo del trabajo fue diseñar un conjunto de visualizaciones de datos a partir del *clustering* de los países según la tasa de infectados cada 10.000 habitantes, a fin de posibilitar a los expertos la evaluación del impacto de las medidas gubernamentales implementadas por los distintos países en un período de tiempo determinado, con respecto a la evolución del COVID-19. Los *datasets* fueron almacenados en PostgreSQL, por otra parte, el *clustering* y las distintas visualizaciones de datos fueron desarrolladas en Tableau. Los resultados permiten observar que algunas medidas como ser el distanciamiento social, aislamiento y medidas específicas de salud pública, implementadas en los países a tiempo, han impactado positivamente en la disminución de la velocidad de crecimiento de la curva de contagio.

Palabras Claves: COVID-19, Medidas Gubernamentales, *Clustering*, Visualización de datos.

1 Introducción

En los últimos meses, el mundo ha cambiado drásticamente debido a la aparición y rápida expansión de una nueva enfermedad denominada COVID 19. Esto es debido a la facilidad de contagio de dicha enfermedad, y además, a la existencia de un periodo asintomático en las personas afectadas, lo cual aumenta la velocidad de transmisión [1]. Ante ello, cada gobierno ha implementado diferentes medidas, priorizando algunos sectores sobre otros, a fin de disminuir el impacto de la enfermedad en los mismos, según su idiosincrasia e intereses. La fecha de detección del primer caso de la enfermedad en cada país fue diferente en su mayoría. Esto permitió que algunos países tomaran decisiones tempranas en cuanto a la implementación inmediata de medi-

das gubernamentales fundamentales a fin de evitar las devastadoras consecuencias observadas en otros países [2].

Existen estudios que han evaluado la efectividad de algunas medidas gubernamentales referentes al distanciamiento social en determinados países, sin embargo, no hay datos sobre su contribución relativa respecto al control de la pandemia [3,4].

Actualmente, existe la necesidad por parte de los expertos de identificar las medidas gubernamentales adecuadas para disminuir el impacto del COVID 19, a partir de la información disponible en diferentes fuentes de datos [5,6].

El objetivo del presente trabajo fue diseñar un conjunto de visualizaciones de datos a partir del *clustering* de los países según la tasa de infectados cada 10.000 habitantes, a fin de posibilitar a los expertos la evaluación del impacto de las medidas gubernamentales implementadas por los distintos países en un período de tiempo determinado, con respecto a la evolución del COVID-19.

2 Marco Teórico

En esta sección se explican brevemente conceptos sobre COVID-19, y las medidas gubernamentales implementadas para disminuir el impacto de la misma. Además, se explica en que consiste la visualización de datos, así como el método *K-means*.

2.1 COVID-19

En diciembre de 2019, se produjo la aparición de una enfermedad respiratoria atípica aguda en Wuhan, China. Esto se extendió rápidamente de esta ciudad a otras partes de China y al resto del mundo. Pronto se descubrió que un nuevo coronavirus era el responsable, denominado síndrome respiratorio agudo severo coronavirus-2 (SARS-CoV-2, 2019-nCoV) debido a su alta homología (~ 80%) con el SARS-CoV, que causó el síndrome de dificultad respiratoria aguda y alta mortalidad durante 2002-2003. La enfermedad causada por este virus se denominó Coronavirus 19 (COVID-19), y la Organización Mundial de la Salud (OMS) declaró una pandemia [7].

Un modo de transmisión importante para el SARS CoV-2 es la transmisión de persona a persona como el SARS CoV y el MERS CoV. Las vías de transmisión identificadas de COVID-19 son las gotas y la transmisión de contacto. Las personas infectadas pueden contagiar sin tener síntomas. La gravedad de la infección puede variar de infección asintomática a enfermedad crítica. La gravedad clínica de COVID-19 se definió en 5 grupos [1]:

1. **Infección asintomática:** sin ningún signo clínico de síntomas con la prueba de PCR SARSCoV-2 positiva;
2. **Leve:** síntomas de infección aguda del tracto respiratorio superior, incluyen fiebre, fatiga, mialgia, tos, dolor de garganta, secreción nasal y estornudos sin neumonía;
3. **Moderado:** con neumonía, fiebre frecuente y tos; algunos pueden tener sibilancias, pero no hipoxemia obvia, como falta de aliento;
4. **Grave:** progresión rápida alrededor de 1 semana, disnea, con cianosis central, saturación de oxígeno inferior al 92%, con otras manifestaciones de hipoxemia;

5. **Crítico:** pacientes con síndrome de dificultad respiratoria aguda (SDRA) o insuficiencia respiratoria, shock, disfunción orgánica múltiple.

2.2 Medidas gubernamentales

Las medidas implementadas por los gobiernos de los diferentes países en respuesta a la pandemia de coronavirus se dividen en las siguientes categorías principales [6]:

- Distanciamiento social;
- Cierres de emergencia;
- Medidas de salud pública;
- Medidas gubernamentales y socioeconómicas;
- Restricciones de movimiento.

A su vez, cada categoría se divide en varios tipos de medidas las cuales se muestran en la **Tabla 1** [6].

Tabla 1. Medidas gubernamentales

Categorías	Medidas
Medidas gubernamentales y socioeconómicas	Medidas económicas, activación o establecimiento de estructuras administrativas de emergencia, limitación de las importaciones/exportaciones, despliegue militar, y declaración del estado de emergencia
Cierre de emergencia	Cierre total, cierre de campamentos de refugiados u otras minorías, y cierre parcial
Restricciones de movimiento	Documentación adicional de salud requerida a la llegada, cierre de fronteras, chequeos en las fronteras, cierre de fronteras completo, toque de queda, puntos de chequeo en el país, restricciones de viajes nacionales, suspensión de vuelos internacionales, vigilancia y monitoreo, y restricciones de visa
Medidas de salud pública	Enmiendas al reglamento de funerales y entierros, campañas de conciencia, recomendaciones generales, exámenes de salud en aeropuertos y cruces de fronteras, políticas de aislamiento y cuarentena, testeos masivos, fortalecimiento de otras medidas de salud pública, asistencia psicológica y trabajo social médico, requerimiento de uso de equipo protector en público, fortalecimiento del sistema de salud público, y políticas de testeo
Distanciamiento social	Cambios en las políticas relacionadas con las prisiones, cierre de negocios y servicios públicos, limitación de las reuniones públicas, y cierre de colegios

2.3 Visualización de datos

La visualización de datos es el proceso de representación de datos, en formato gráfico, de una manera clara y eficaz. Se convierte en una herramienta poderosa para el análisis e interpretación de datos grandes y complejos, volviéndose un medio eficiente en la transmisión de conceptos en un formato universal. A partir de la generación de visualizaciones adecuadas se logra una mejor comprensión de los datos [8]. Entre las herramientas más utilizadas actualmente para el diseño de visualizaciones de datos se encuentra el Tableau [9].

El proceso de visualización de datos incluye cuatro etapas básicas, combinadas en varios bucles de retroalimentación, siendo las mismas [10]:

1. Recopilación y almacenamiento de datos;
2. Etapa de pre-procesamiento, necesaria para transformar los datos en algo que sea más fácil de manipular. Por lo general, hay alguna forma de reducción de datos para revelar aspectos seleccionados. La exploración de datos es el proceso de cambiar el subconjunto que se está viendo actualmente;
3. Mapeo de los datos seleccionados a una representación visual, que se logra a través de algoritmos informáticos que producen una imagen en la pantalla. La entrada del usuario puede transformar las asignaciones, resaltar subconjuntos o transformar la vista. En general, esto se hace en la computadora del usuario;
4. Sistema perceptivo y cognitivo humano (el perceptor).

2.4 Método *K-means*

K-means es una técnica de aprendizaje no supervisado utilizada para particionar automáticamente un conjunto de datos en k grupos, denominados *clusters*. El algoritmo continúa seleccionando k centros de los *clusters* iniciales y luego refinándolos iterativamente de la siguiente manera [11]:

1. Cada instancia d_i se asigna a su centro de *cluster* más cercano;
2. Cada centro de agrupación C_j se actualiza para ser la media de sus instancias.

El algoritmo converge cuando no hay más cambios en la asignación de instancias a los distintos *clusters*.

3 Descripción de la solución implementada

En la siguiente sección se presenta la composición de los *datasets* utilizados, así también como los datos calculados para el desarrollo de las visualizaciones de datos propuestas. Finalmente, se indican las tecnologías utilizadas para su implementación.

3.1 Composición de *COVID-19 Cases dataset*

Este *dataset* presenta una serie temporal de los casos confirmados y las muertes de todos los países del mundo entero que han tenido al menos un caso del COVID-19. El mismo ha sido elaborado por el *Johns Hopkins University Center for Systems Science*

and Engineering (JHU CSSE), y actualizado diariamente desde el 22 de Enero de 2020. A la fecha de la última actualización, 19 de Mayo de 2020, el mismo contaba con un total de 837.998 registros.

En la **Tabla 2** se presentan los campos utilizados para el desarrollo de las visualizaciones [5]:

Tabla 2. Campos utilizados del *COVID-19 Cases Dataset*

Columna	Tipo de Dato	Valores Nulos	Descripción
case_type	Texto	No	Casos confirmados y muertes reportadas
cases	Entero	No	Cantidad de personas que dieron positivo
date	Fecha	No	Fecha en la cual se reportaron los casos
country_region	Texto	No	Nombre del país
iso3	Texto	Si	Código ISO3 del país
lat	Decimal	Si	Latitud del país (georeferencia)
long	Decimal	Si	Longitud del país (georeferencia)
population_count	Entero	Si	Población del país

A fin de realizar las visualizaciones de datos correctamente, se realizaron las siguientes tareas de pre-procesamiento y reducción de datos:

1. **Eliminación de cruceros:** *Cruise Ship* (238 registros con un total de 735 casos);
2. **Normalización de fechas:** Cálculo del día relativo al inicio de la pandemia en cada país, siendo el día 1 en el que se detectaron los primeros casos en dicho país;
3. **Unificación de la población en países con subregiones:** Austria, Canadá, China, USA, y *United Kindom*;
4. **Dataset definitivo:** Fueron considerados para el análisis únicamente los países con una población mayor o igual a 3 millones de habitantes y una duración mínima de la pandemia de 67 días desde el primer caso detectado en el país. A su vez se tuvieron en cuenta únicamente los primeros 78 días de la pandemia de todos los países a fin de comparar la evolución de los distintos países con Argentina, es decir, el período comprendido entre el 03 de Marzo al 19 de Mayo del 2020. Así también, fueron considerados únicamente los países de los cuales se disponía de información referente a las medidas gubernamentales implementadas, quedando fuera de la muestra los siguientes países: *Nigeria, Ethiopia, Taiwan, Ghana, Greece, Egypt, Bosnia and Herzegovina*.

3.2 Composición de ACAPS COVID-19: Government Measures Dataset

Este *dataset* presenta una serie temporal de las medidas gubernamentales, descritas en la Sección 2.2, implementadas por la mayoría de los países que han tenido al menos un caso del COVID-19. El mismo ha sido elaborado por el *Humanitarian Data Exchange*, y actualizado frecuentemente desde el 04 de Marzo de 2020. A la fecha de la última actualización del *dataset*, 19 de Mayo de 2020, el mismo contaba con un total de 11.200 registros.

En la **Tabla 3** se presentan los campos utilizados para el desarrollo de las visualizaciones [6].

Tabla 3. Campos utilizados del *CAPS COVID-19: Government Measures dataset*

Columna	Tipo de Dato	Valores Nulos	Descripción
id	Entero	No	Identificador
country	Texto	Si	Nombre del País
iso	Texto	Si	Código ISO3 del país
region	Texto	No	Región del país
category	Texto	No	Categoría de la medida
measure	Texto	Si	Medidas gubernamentales
targeted_pop_group	Texto	No	Medida dirigida a grupos de personas
comments	Texto	Si	Detalles sobre las medidas
date_implemented	Fecha	Si	Fecha de implementación de la medida
source	Texto	No	Fuente de la información
link	Texto	Si	Enlace de la fuente
entry_date	Fecha	No	Fecha de actualización

A fin de realizar las visualizaciones de datos correctamente, se realizaron las siguientes tareas de pre-procesamiento y reducción de datos, siendo las mismas:

1. **Datos faltantes:** Fueron cargadas las fechas faltantes en el campo *date_implemented*, algunas a partir de la revisión exhaustiva de las fuentes indicadas en el campo *source*, y otras, a partir de la fecha informada en el campo *entry_date*;
2. **Unificación de datos:** Fueron unificadas las descripciones tanto de las categorías, como de las medidas;
3. **Correcciones de datos:** Se realizaron algunas correcciones sobre los datos en los campos *iso* y *region* de los registros correspondientes a algunos países;
4. **Dataset definitivo:** Fueron considerados para el análisis únicamente los países con una población mayor o igual a 3 millones de habitantes y una duración mínima de la pandemia de 67 días desde el primer caso detectado en el país. Finalmente, para la muestra se incluyeron 95 países en total.

3.3 Campos calculados

A continuación se describen los campos calculados creados en Tableau, necesarios para posibilitar la comparación en la evolución de la pandemia en los distintos países:

- **Infected Rate:** Para posibilitar la visualización de la evolución del COVID-19 relativa a la población de los distintos países cada 10.000 habitantes, fue creado un campo denominado *Infected Rate*, disponible en fórmula (1);

$$Infected\ Rate = \frac{cases}{population_count} \times 10.000 \quad (1)$$

donde *cases* incluye únicamente los casos confirmados

- **Day:** A fin de comparar la evolución de la pandemia en los distintos países, fue creado el campo calculado en Tableau denominado *day*. Este se refiere al día relativo a la fecha de inicio de la pandemia en cada país, donde en el día 1 se detectaron los primeros casos. Por ejemplo, en el caso de Argentina, el día 1 fue el 3 de Marzo de 2020, y el día 78 se correspondió con el 19 de Mayo de 2020, la cuál fue la última fecha de actualización del *dataset*;
- **Period:** A fin de visualizar la distribución semanal de las primeras medidas gubernamentales implementadas por los distintos países, fue creado el campo calculado *Period*. En este sentido, se han dividido los 78 días en los siguientes grupos: *Pre-pandemic Week* (medidas gubernamentales implementadas días anteriores al inicio de la pandemia en el país), *1st Week* a *11th Week*.

3.4 Configuraciones del Tableau

A continuación se indica las distintas configuraciones realizadas en el Tableau:

- **Base de Datos:** Fue creada una conexión a la base de datos PostgreSQL, a partir de los datos seleccionados de los *datasets* mencionados en las secciones 3.1 y 3.2, almacenados en dos tablas denominadas *cases* y *government_measures*, respectivamente;
- **Clustering:** Para la realización del *clustering* se utilizó *K-means*, siendo este el único método disponible actualmente en el Tableau. La cantidad de *clusters* seleccionada fue automática (determinada por Tableau).

3.5 Visualizaciones de datos

A continuación se indican las distintas visualizaciones diseñadas para abordar el problema planteado:

- **Clustering of the Countries:** Visualización del resultado del *clustering* de las series de tiempo de los países según evolución de la pandemia teniendo en cuenta la tasa de infectados relativa correspondiente al último día de la muestra, y visualización geográfica de los mismos;
- **Weekly Distribution of the First Government Measures:** Visualización de la distribución semanal de los países según las primeras medidas gubernamentales implementadas. Se utilizó una tabla de resaltado que muestra la distribución porcentual y cuantitativa de países según cada *cluster*;
- **Government Measures and Infected Rate:** Se pueden visualizar las primeras medidas implementadas en contraste a la serie de tiempos según el *Infected Rate* para el período de análisis.

3.6 Tecnologías utilizadas

Para la implementación las visualizaciones se utilizaron las siguientes tecnologías:

- **Tableau Desktop:** Es un software de visualización de datos que permite ver y comprender datos en cuestión de minutos. Los *dashboards* interactivos ayudan a descubrir información oculta al instante. Tableau aprovecha la capacidad natural de las personas para detectar patrones visuales rápidamente [9];
- **PostgreSQL:** Es un sistema de gestión de bases de datos relacional de código abierto orientado a objetos [12];
- **pgAdmin:** Es un cliente de código abierto para la administración de bases de datos PostgreSQL [13].

4 Resultados obtenidos

El resultado obtenido fueron tres *clusters*, los cuales fueron renombrados respecto al valor del *Infected Rate* promedio, siendo los nombres *Low*, *Medium*, y *High*, como se muestra en la **Tabla 3**.

Tabla 3. Resultados obtenidos en el *clustering* de países según el *Infected Rate*

<i>Cluster</i>	Total de Países	Porcentaje
<i>Low</i>	64	67,37%
<i>Medium</i>	21	22,10%
<i>High</i>	10	10,53%

En la **Tabla 3** se puede observar que la mayor cantidad de países se encuentra en el *cluster Low*, debido a que el valor de *Infected Rate* correspondiente al último día del período analizado fue inferior al de otros países agrupados por *K-means*.

En la visualización denominada *Clustering of the Countries* se muestran las series de tiempo de países agrupados en los distintos *clusters*, diferenciados por color (**Fig. 1**). En la parte superior de la figura se muestra la serie de tiempo de los países a partir del *Infected Rate* según el *Day*, y en la parte inferior se observa la ubicación geográfica de los mismos.

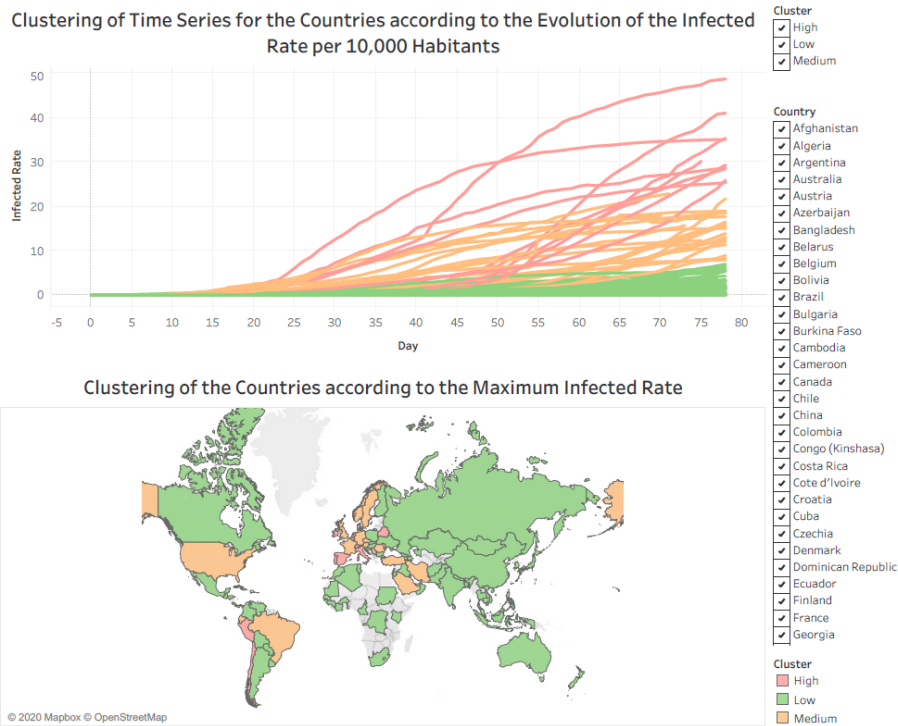


Fig. 1. Clustering of the Countries, generado con Tableau (<https://public.tableau.com>)

En la visualización denominada *Weekly Distribution of the First Government Measures* puede observarse la distribución semanal de los países según las primeras medidas gubernamentales implementadas y el *cluster* de pertenencia. Se muestra tanto la distribución porcentual relativa a la cantidad de países contenidos en el *cluster* (parte superior de la figura), como la cantidad de países que tomaron la medida gubernamental por primera vez (**Fig. 2** y **Fig. 3**).

En la **Fig. 2** se muestran los resultados obtenidos para los países pertenecientes al *cluster Low*, donde se seleccionaron dos categorías: medidas de salud pública y medidas de distanciamiento social. Con respecto a las primeras, el 21,88% de países en la primera semana comenzó a realizar exámenes de salud en aeropuertos y pasos fronterizos, y en la segunda semana el 7,81% implementó la misma medida, siendo un total de 19 países. Así también, en la primera semana, el 18,75% implementó políticas de aislamiento y cuarentena, y la segunda semana lo hizo el 15,63%, 22 países en total. En cuanto a las medidas de distanciamiento social, se observó que en la primera semana el 15,63% procedió al cierre de negocios y servicios públicos, y la segunda semana lo hizo también el 7,81%, siendo un total de 15 países. Además, en la primera semana el 17,19% limitó las reuniones públicas, y 28,13% lo hizo la segunda semana, alcanzando un 45,32% acumulado (29 países). Finalmente, en la primera semana el

26,56% de los países procedió al cierre de colegios, y en la segunda semana el 23,44% implementó la misma medida, lo cual representa un total del 50% de los países del cluster (32 países).

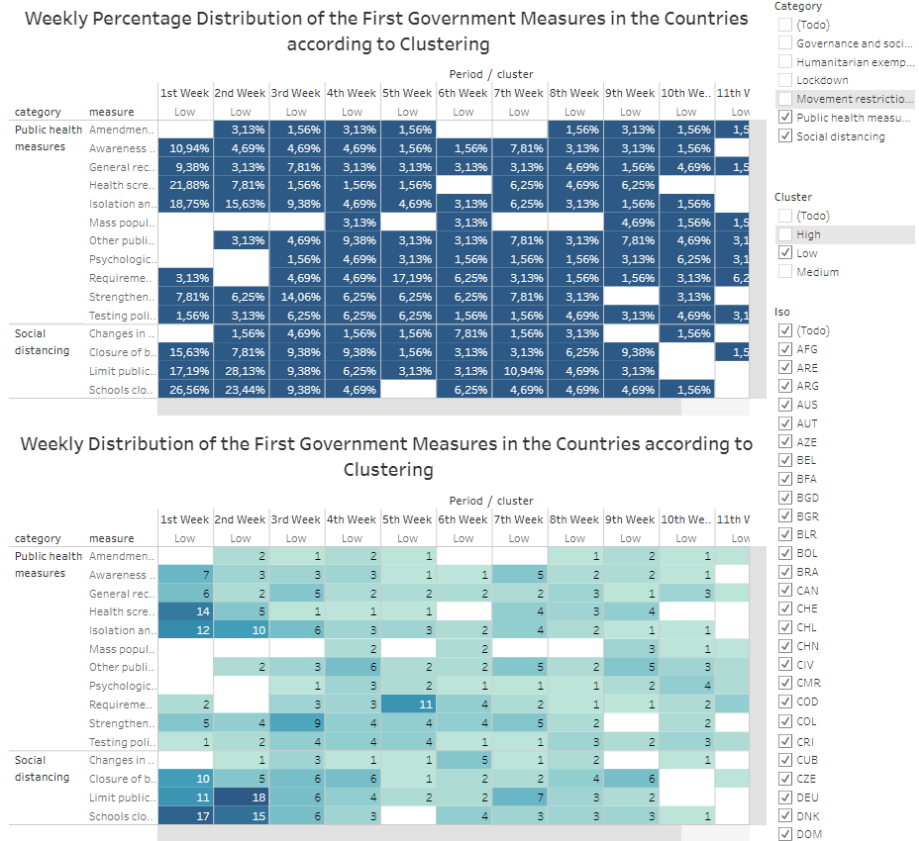


Fig. 2. Weekly Distribution of the First Government Measures – Cluster Low, generado con Tableau (<https://public.tableau.com>)

En la Fig. 3 se muestran los resultados obtenidos para los países pertenecientes al cluster High, donde se seleccionaron las mismas categorías analizadas para el cluster Low a fin de comparar los resultados. Con respecto a las medidas de salud pública, en las dos primeras semanas ninguno de los países realizó exámenes de salud en aeropuertos y pasos fronterizos, fue en la tercera semana que algunos de los países comenzaron a implementar dicha medida. Por otro lado, en la primera semana, solamente el 10 % implementó políticas de aislamiento y cuarentena, y la segunda semana lo hizo el 20%. En cuanto a las medidas de distanciamiento social, se observó que en las dos primeras semanas ninguno de los países procedió al cierre de negocios y servicios públicos. Además, en la primera semana únicamente el 10% limitó las reuniones públicas, y el 20% lo hizo una semana más tarde. Finalmente, en la primera semana solamente el 10% de los países procedió al cierre de colegios, y en la segunda semana

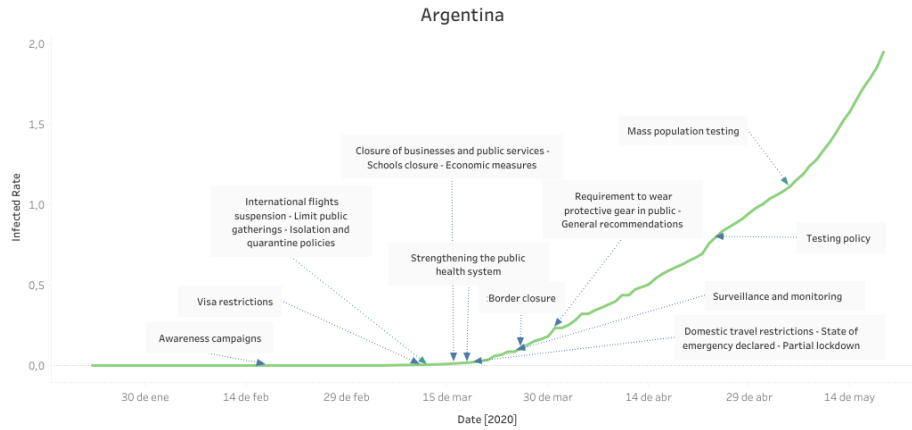


Fig. 4. Government Measures and Infected Rate - Argentina, generado con Tableau (<https://public.tableau.com>)

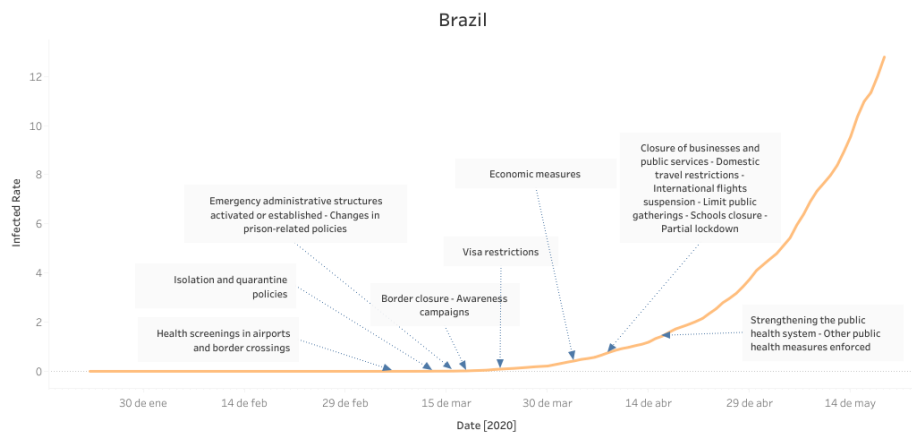


Fig. 5. Government Measures and Infected Rate - Brazil, generado con Tableau (<https://public.tableau.com>)

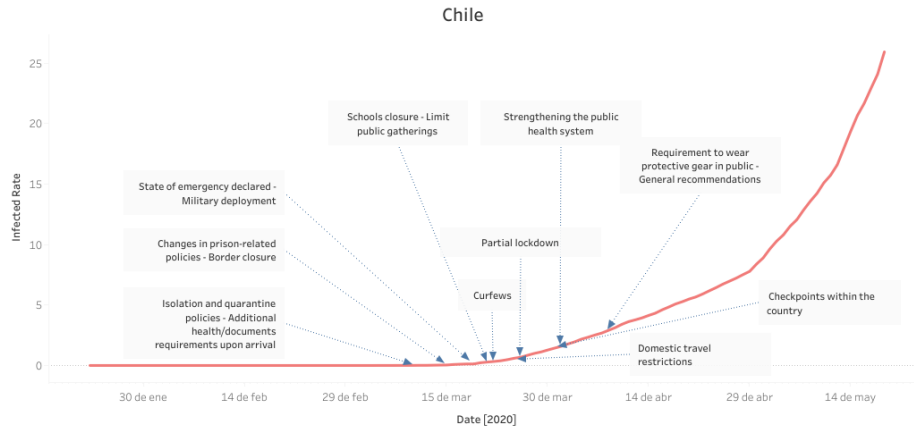


Fig. 6. Government Measures and Infected Rate - Chile, generado con Tableau (<https://public.tableau.com>)

5 Conclusiones

En el presente trabajo se concluye que algunas medidas específicas de salud pública y de distanciamiento social implementadas a tiempo han impactado positivamente en la disminución de la velocidad de crecimiento de la curva de contagio, ya que por ejemplo, en los países del *cluster Low* 19 países habían comenzado a realizar exámenes de salud en aeropuertos y pasos fronterizos, mientras que los del *cluster High* no habían implementado la medida en ese periodo. A su vez, en cuanto a las medidas de distanciamiento social, 32 países del *cluster Low* realizaron el cierre de escuelas a las dos semanas desde el primer caso diagnosticado de la enfermedad, mientras que únicamente 2 países del *cluster High* lo habían realizado. A su vez, 15 países del *cluster Low* procedieron al cierre de negocios y servicios públicos dentro de las dos primeras semanas, mientras que ninguno de los países del *cluster High* lo había implementado.

Como trabajo futuro pueden incluirse otras variables las cuales no se encontraban en los *datasets* utilizados, tales como, la densidad demográfica (factor importante que aumenta la velocidad de contagio en ciudades muy conglomeradas), así también como, las comorbilidades de la población de los diferentes países, las cuales provocan un mayor riesgo de muerte en caso de contagio en la población.

6 Bibliografía

1. Bulut, C., Kato, Y.: Epidemiology of COVID-19. *Turkish journal of medical sciences*, 50 (SI-1), 563–570 (2020)
2. Saglietto, A. *et al.*: COVID-19 in Europe: the Italian lesson. *Lancet*, 395 (10230), 1110-1111 (2020)
3. Banerjee, T., Arnab, N.: US county level analysis to determine If social distancing

- slowed the spread of COVID-19. *Revista Panamericana de Salud Pública*, 44, e90 (2020)
4. Viner, R. *et al.*: School closure and management practices during coronavirus. *The Lancet Child & Adolescent Health*, 4 (5), 397-404 (2020)
 5. COVID-19 Cases. <https://data.world/covid-19-data-resource-hub/covid-19-case-counts/workspace/file?filename=COVID-19+Cases.csv>, último acceso 20/05/2020
 6. ACAPS COVID-19: Government Measures Dataset. <https://data.humdata.org/dataset/acaps-covid19-government-measures-dataset>, último acceso 20/05/2020
 7. Yuki, K., Fujiogi, M., Koutsogiannaki, S.: COVID-19 pathophysiology: A review. *Clinical Immunology*, 108427 (2020)
 8. Rivera, E. *et al.*: XX Workshop de Investigadores en Ciencias de la Computación - WICC 2018. *Minería de datos y visualización de información*, Resistencia (2018)
 9. Tableau. <https://www.tableau.com/>, último acceso 30/05/2020
 10. Ware, C.: Information visualization: Perception for design. *Information visualization: Perception for design*. Massachusetts: Elsevier, 4 (2013)
 11. MacQueen, J.: Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 281-297 (1967)
 12. PostgreSQL. <https://www.postgresql.org/>, último acceso 30/05/2020
 13. pgAdmin. <https://www.pgadmin.org/>, último acceso 20/05/2020