



Reporte Analítico de Indicadores

Indicadores de Telecomunicaciones de economías emergentes.
Estado actual de las telecomunicaciones en México y otras economías emergentes

CENTRO DE ESTUDIOS

Responsable

Mtro. José Alberto Candelaria Barrera

Indicadores de telecomunicaciones de economías emergentes

Estado actual de las telecomunicaciones en México y otras economías emergentes.

Introducción

La experiencia internacional sugiere que las telecomunicaciones juegan un rol esencial en el desarrollo económico de los países, mediante el incremento del empleo, la movilidad del capital, la redistribución del valor agregado entre factores como el propio capital y el trabajo, o, incluso, en la reducción del gasto de materiales y en contaminantes en la economía. Según un estudio de Katz¹ para países de América Latina, un incremento del índice de digitalización de 1% resulta en un aumento de 0.32% en el producto interno bruto, de 0.26% en la productividad laboral, de 0.23% en la productividad multifactorial y de 0.09% en la contribución de las TIC a la productividad laboral.

El progreso tecnológico de las telecomunicaciones es un fenómeno multifactorial que requiere de la consideración de distintos indicadores para ser capturado, por lo que su medición integral es compleja. La existencia de diferentes variables que señalan el avance de los distintos aspectos que conforman el desarrollo tecnológico dificulta establecer conclusiones y definir una estrategia de conectividad acorde con las necesidades del país. Asimismo, la penetración de los servicios de telecomunicaciones no ha sido homogénea a nivel mundial durante los últimos años.

En este trabajo se realiza un análisis estadístico del contraste que existe en el desarrollo de las telecomunicaciones en un conjunto integrado por 20 países con economías emergentes de todo el mundo, incluyendo a México. El objetivo es proveer al personal del Instituto Federal de Telecomunicaciones con un insumo útil que permita identificar brechas con respecto a otras naciones lo que constituye un primer paso en el diseño e implementación de políticas públicas para acortar esas diferencias.

Es así, que se utilizan técnicas estadísticas de agrupamiento de datos para aglutinar a subconjuntos de países según el grado de similitud en el avance que registran en materia de telecomunicaciones, utilizando para ello nueve de las principales variables del sector. Cada agrupamiento incluye a los países que exhiben niveles similares de acceso a infraestructura y a servicios de telecomunicaciones, mientras que entre los diferentes agrupamientos hay divergencias significativas al respecto.

¹ Katz, R. (2018) *La digitalización: una clave para el futuro crecimiento de la productividad en América Latina*.

Este ejercicio permite así establecer qué países han alcanzado un nivel similar en cuanto a la dotación integral de los servicios de telecomunicaciones, incluyendo indicadores de banda ancha móvil, la banda ancha fija, la voz fija, la televisión restringida y la telefonía móvil. Así como la accesibilidad tanto a la banda ancha fija como a móvil.

Los resultados de este análisis multifactorial sobre el avance del sector de las telecomunicaciones en 20 países, permite establecer seis grupos con niveles similares de equipamiento de servicios de telecomunicaciones.

Información estadística

Para el ejercicio que se presenta en este reporte, se utilizan datos actualizados al cuarto trimestre del año 2020, correspondientes a las nueve variables que se enlistan a continuación.

1. Penetración de banda ancha fija.
2. Penetración de TV restringida.
3. Penetración de voz fija.
4. Penetración de banda ancha móvil.
5. Penetración de telefonía móvil.
6. Accesibilidad a la banda ancha fija (Inverso del precio de 5GB de banda ancha fija como porcentaje del GNI per cápita).
7. Accesibilidad a la banda ancha móvil (Inverso del precio de 1.5GB de datos de banda ancha móvil como porcentaje del GNI per cápita).
8. Porcentaje de población con cobertura 4G.
9. Ancho de banda internacional de internet por usuario (bit/seg)

La muestra estudiada se integra por 20 países con economías que pueden catalogarse como emergentes en función de unos ingresos per cápita de al menos 9 mil dólares estadounidenses hasta 28 mil dólares estadounidenses en paridad de poder adquisitivo (PPP, por sus siglas en inglés; ver Cuadro 3). Entre éstos se encuentran naciones de África, Asia, América Latina y Europa, a saber: Armenia, Botswana, Brasil, Bulgaria, China, Colombia, Ecuador, Georgia, Indonesia, Jordania, Líbano, Malasia, México, Macedonia, Namibia, Rusia, Serbia, Sudáfrica, Tailandia, y Turquía.

La fuente de los indicadores estadísticos son el World Cellular Information Service y el World Broadband Information Service de la consultora OMDIA² (ver Cuadro 1) para las variables de penetración, excepto para el caso de México, los cuales provienen del Banco de Información de Telecomunicaciones del IFT³. En tanto que los datos de las variables de accesibilidad a la banda ancha fija y móvil provienen de la ITU⁴. En lo referente a éstas dos últimas variables se toman los inversos de los precios de cada servicio como porcentaje del ingreso per cápita; esto en virtud de que a un mayor porcentaje del precio con respecto al ingreso per cápita, menor ingreso disponible para cubrir otras necesidades. Es así que al tomar el inverso de dichas variables se le está otorgando un sentido positivo a cada una de ellas. Por otra parte, el ancho de banda internacional de internet por usuario (en bits por segundo) se refiere a la capacidad utilizada de ancho de banda internacional;

² <https://omdia.tech.informa.com/>

³ <https://bit.ift.org.mx/BitWebApp/>

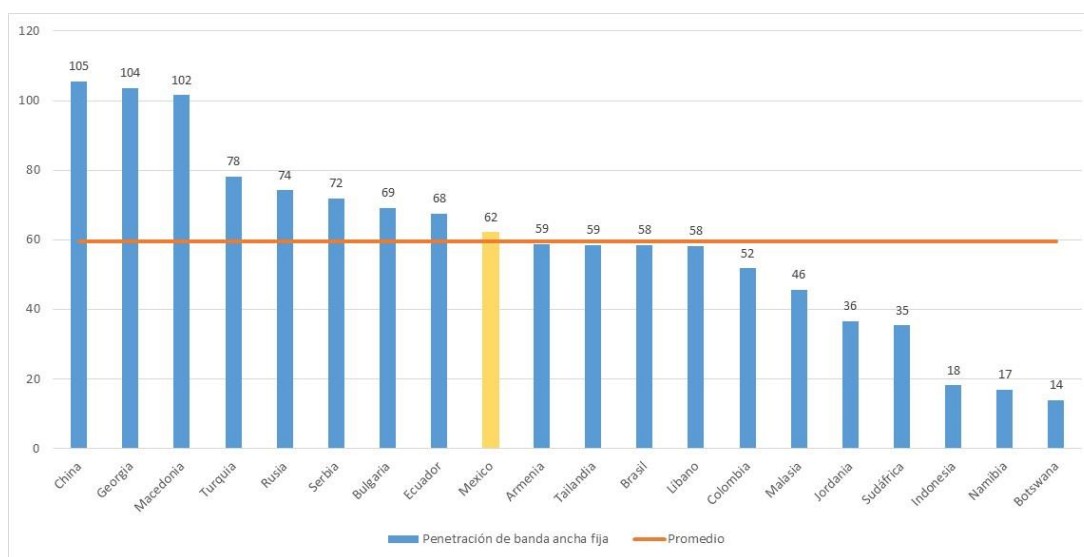
⁴ <https://www.itu.int/itu-d/sites/statistics/>

es decir, a la suma de la capacidad utilizada de todos los puntos de intercambio de tráfico dividida entre el número de usuarios de internet en cada país.

Posteriormente, se procede a analizar de forma esquemática cada una de las nueve variables.

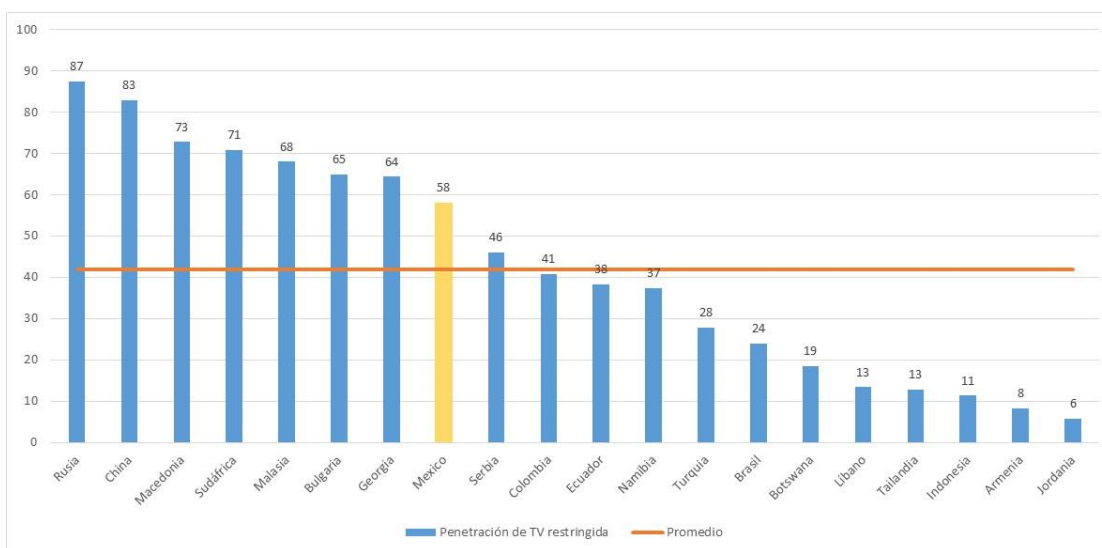
La Gráfica 1 muestra a la variable de penetración de banda ancha fija por hogares para cada uno de los 24 países. En este caso China alcanza una penetración de 105 accesos de banda ancha fija por cada 100 hogares; en tanto que Georgia llega a los 104 accesos de banda ancha fija por cada 100 hogares. México ocupa la novena posición con una penetración de 62 accesos por cada 100 hogares. Por su parte, Namibia y Botswana con 17 y 14 accesos, respectivamente, acusan el mayor rezago. El promedio para los veinticuatro países es de 59 accesos por cada 100 hogares.

Gráfica 1. Penetración de banda ancha fija en hogares



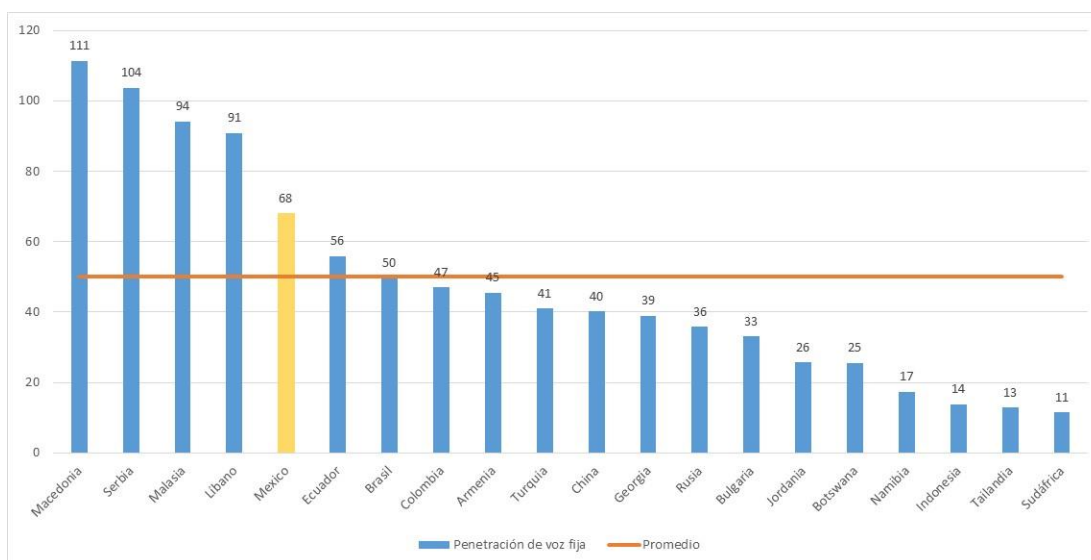
Por otra parte, la variable de penetración de TV restringida por hogares se muestra en la Gráfica 2. La primera posición la ocupa Rusia con una penetración de 87 suscripciones por cada 100 hogares, seguida de China con 83 suscripciones por cada 100 hogares. Los últimos puestos los ocupan Armenia y Jordania, con 8 y 6 suscripciones por cada 100 hogares. México con 58 suscripciones por cada 100 hogares se encuentra por encima del promedio de la región el cual es de 42 suscripciones por cada 100 hogares.

Gráfica 2. Penetración de TV restringida



La Gráfica 3 muestra la penetración de voz fija por hogares. Se observa que Macedonia alcanza una penetración de 111 suscriptores por cada 100 hogares; en tanto que la de Serbia es de 104 suscriptores por cada 100 hogares. Estos países se encuentran en las posiciones uno y dos, respectivamente. En cuanto a los puestos finales, Sudáfrica con 11 y Tailandia con 13 suscriptores por cada 100 hogares, son los países con mayor rezago. México con 68 suscriptores por cada 100 hogares se encuentra por encima del promedio, el cual asciende a 50 suscripciones por cada 100 hogares.

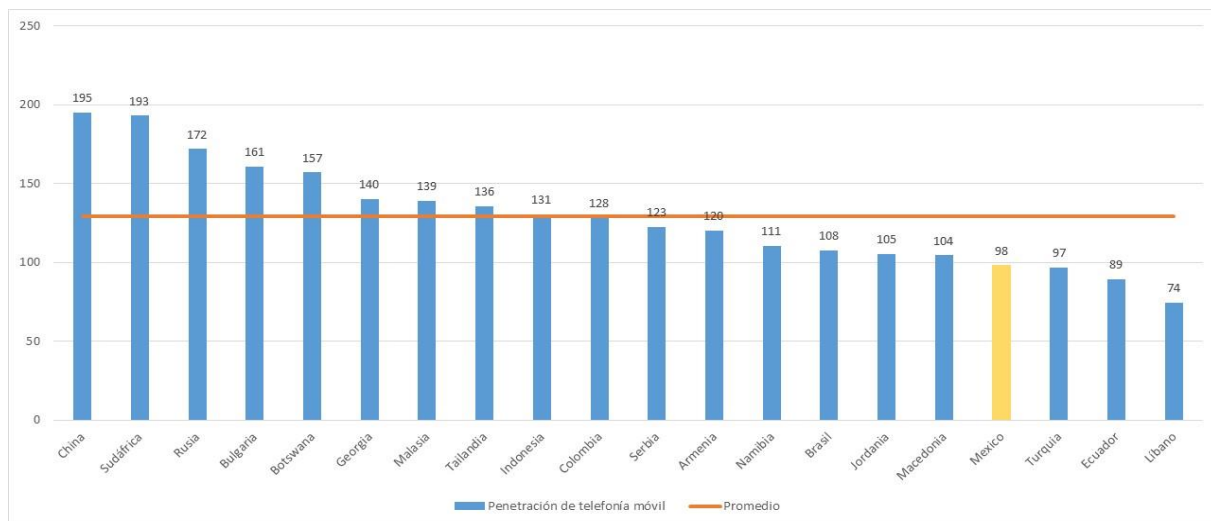
Gráfica 3. Penetración de voz fija



Por otra parte, China cuenta con la tasa de penetración de líneas de telefonía móvil por cada 100 habitantes más elevada para los veinte países de la muestra, con 195 líneas por cada 100 habitantes,

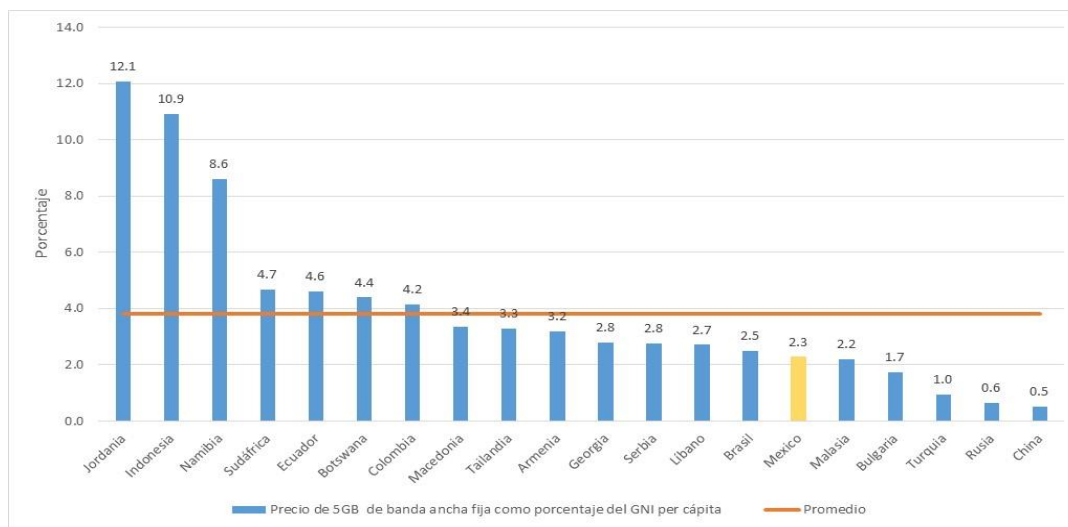
seguido de Sudáfrica con 193 líneas por cada 100 habitantes (Gráfica 4). México reporta 98 líneas por cada 100 habitantes. Líbano con 74 líneas por cada 100 habitantes reporta el mayor rezago.

Gráfica 4. Penetración de telefonía móvil



La Gráfica 5 muestra el precio de 5GB de banda ancha fija como porcentaje del GNI per cápita⁵. Jordania es la nación que presenta una mayor proporción del precio del servicio con respecto a los ingresos con un 12.1%; a ésta le sigue Indonesia con 10.9%. México con un 2.3% se encuentra por debajo del promedio general correspondiente a 4%. En tanto que Rusia (0.6%) y China (0.5%) ocupan las últimas posiciones.

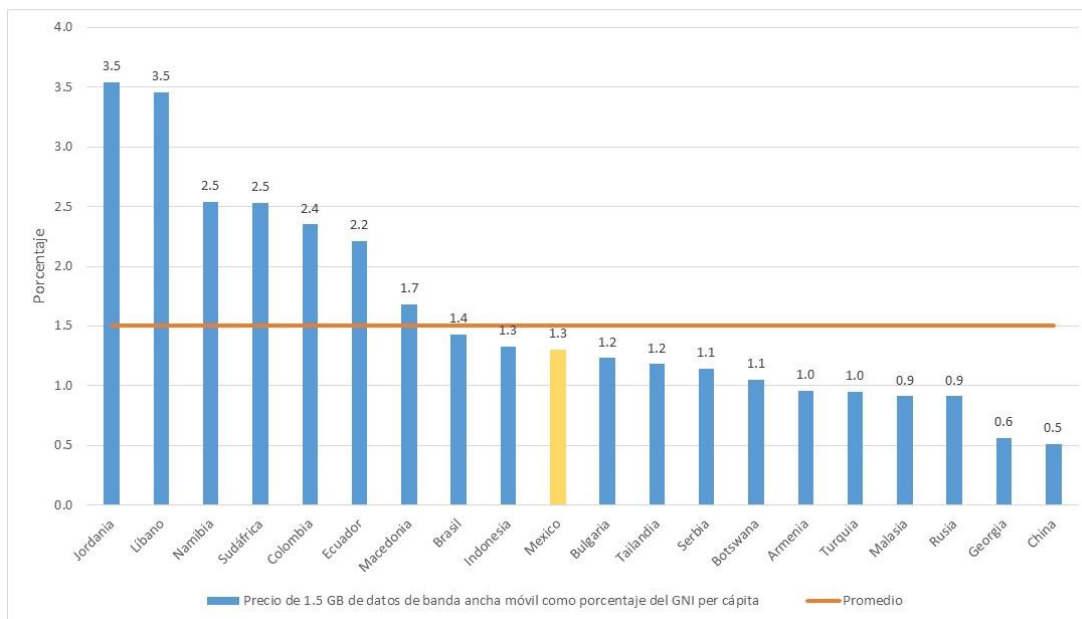
Gráfica 5. Precio de 5GB de banda ancha fija como porcentaje del PIB per cápita



⁵ Para fines del ejercicio estadístico se utilizaron los inversos de los precios como porcentaje del PIB per cápita para cada país de la muestra, en virtud de que permite darle un sentido positivo a la variable.

En la gráfica 6 se presenta el precio de 1.5 GB de datos de banda ancha móvil como porcentaje del GNI per cápita⁶. El primer sitio lo ocupa Jordania con un 3.5%, seguido de Líbano con 3.4%. Georgia (0.6%) y China (0.5%) ocupan los últimos puestos. México con un 1.3% del precio del servicio como porcentaje de sus ingresos se encuentra por debajo del promedio general para la muestra de veinte países que es de 2%.

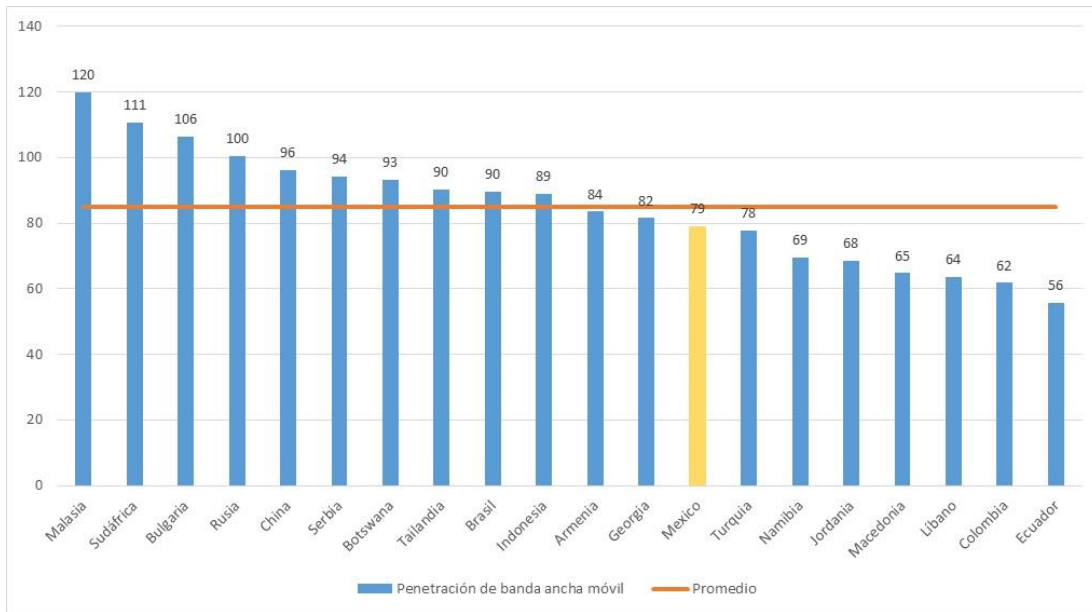
Gráfica 6. Precio de 1.5 GB de datos de banda ancha móvil como porcentaje del PIB per cápita



Aunado a lo anterior, la Gráfica 7 muestra a la variable de penetración de la banda ancha móvil. Malasia ocupa la primera posición con 120 líneas por cada 100 habitantes, seguido de Sudáfrica con 111 líneas por cada 100 habitantes. Colombia y Ecuador ocupan las posiciones más rezagadas con 62 y 56 líneas por cada 100 habitantes, respectivamente. México con 79 líneas por cada 100 habitantes se encuentra por debajo del promedio de la muestra de veinte países el cual equivale a 85 líneas por cada 100 habitantes.

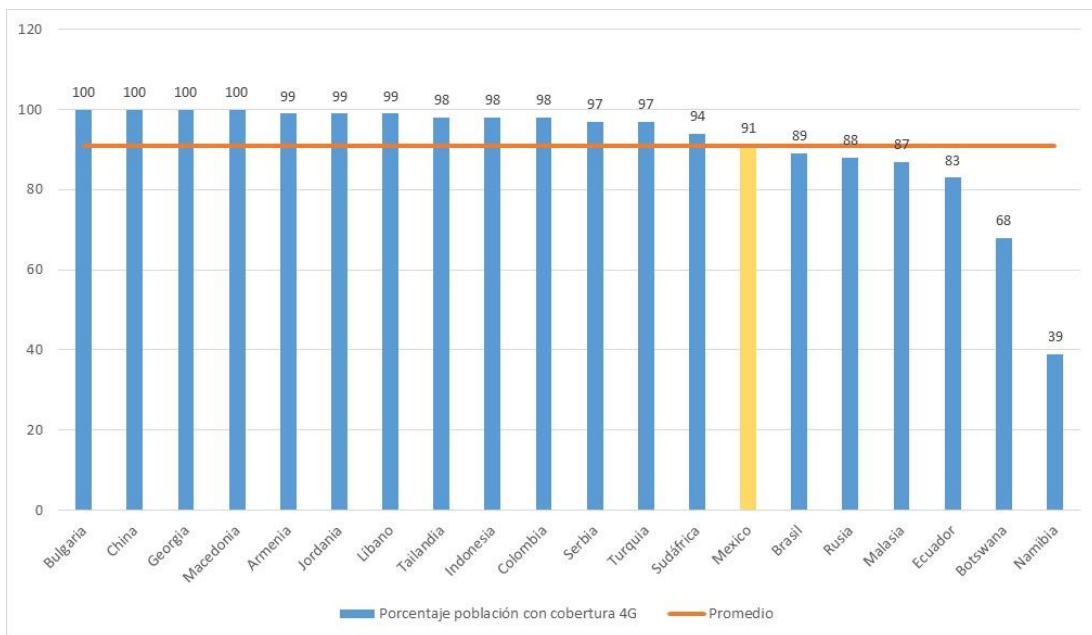
⁶ Idem.

Gráfica 7. Penetración de banda ancha móvil



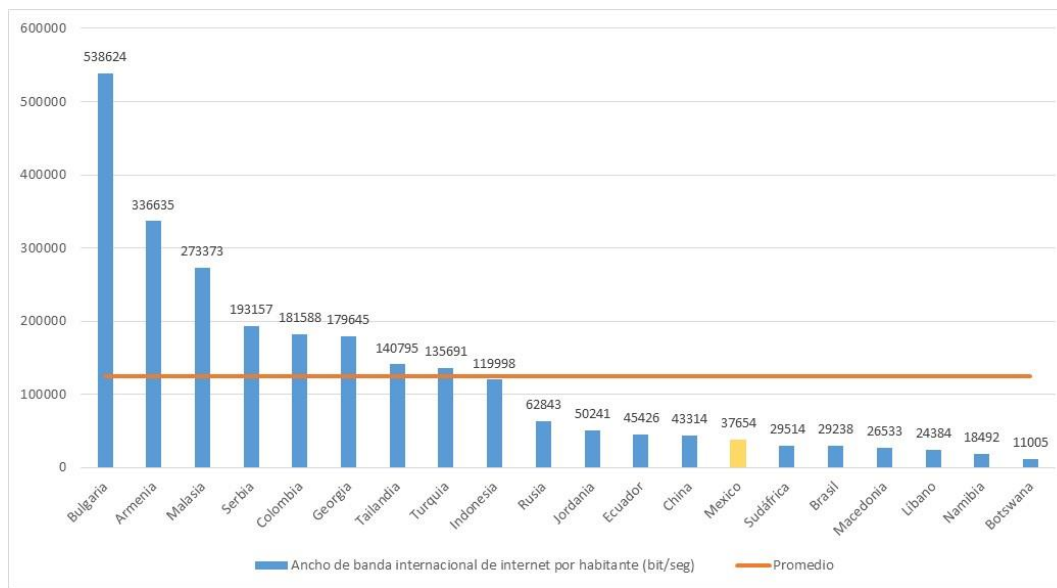
La gráfica 8 presenta el porcentaje de población con cobertura 4G. Se puede observar que Bulgaria, China, Georgia y Macedonia alcanzan el 100 de su población. En tanto que Armenia, Jordania y el Líbano llegan hasta el 99 por ciento. Por su parte, Botswana con 68 por ciento y Namibia con el 39 por cientos se encuentran al final de la gráfica. México presenta un 91 por ciento.

Gráfica 8. Porcentaje de población con cobertura 4G



Finalmente, la gráfica 9 presenta el ancho de banda internacional de internet por usuario en bits por segundo. Bulgaria se encuentra en el primer lugar de los veinte países con 538,624 bits por segundo por usuario de internet, seguida de Armenia con 336,635 bits por segundo. En el caso de México alcanza los 37,654 bits por segundo.

Gráfica 9. Ancho de banda internacional de internet por usuario (bit/seg)



Cuadro 1. Variables estadísticas del sector de las telecomunicaciones por país.

País	PIB per cápita en PPP	Inverso del precio de 5GB de banda ancha fija como porcentaje del GNI per cápita	Inverso del precio de 1.5GB de datos de banda ancha móvil como porcentaje del GNI per cápita	Penetración de banda ancha fija	Penetración de telefonía móvil	Penetración de TV restringida	Penetración de voz fija	Penetración de banda ancha móvil	Porcentaje de población con cobertura 4G	Ancho de banda internacional por usuario de internet (bits/seg)
Armenia	9668	0.3	1.0	58.8	120.0	8.2	45.4	84	99	336635
Botswana	17024	0.2	1.0	13.8	157.0	18.6	25.5	93	68	11005
Brasil	15553	0.4	0.7	58.5	107.6	24.0	49.6	90	89	29238
Bulgaria	20948	0.6	0.8	69.1	160.5	64.8	33.0	106	100	538624
China	16842	2.0	2.0	105.5	194.9	83.0	40.2	96	100	43314
Colombia	14503	0.2	0.4	51.8	128.1	40.8	47.0	62	98	181588
Ecuador	11612	0.2	0.5	67.5	89.3	38.3	55.8	56	83	45426
Georgia	10674	0.4	1.8	103.5	140.0	64.3	38.9	82	100	179645
Indonesia	12310	0.1	0.8	18.1	130.5	11.3	13.8	89	98	119998
Jordania	9173	0.1	0.3	36.5	105.1	5.8	25.6	68	99	5024
Líbano	14513	0.4	0.3	58.2	74.4	13.4	90.9	64	99	24384
Malasia	29511	0.5	1.1	45.5	138.8	68.0	94.2	120	87	273373
México	18656	0.4	0.8	62	98	58	68	79	91	37654

Namibia	10471	0.1	0.4	16.9	110.5	37.4	17.2	69	39	18492
Macedonia	15290	0.3	0.6	101.6	104.4	72.8	111.4	65	100	26533
Rusia	25763	1.6	1.1	74.2	171.8	87.4	35.7	100	88	62843
Serbia	15432	0.4	0.9	71.9	122.6	46.1	103.7	94	97	193157
Sudáfrica	13526	0.2	0.4	35.4	192.9	70.9	11.4	111	94	29514
Tailandia	17910	0.3	0.8	58.5	135.7	12.8	12.8	90	98	140795
Turquía	28002	1.1	1.1	78.2	96.7	27.9	40.9	78	97	135691

Determinación de agrupamientos según la disponibilidad de los servicios de telecomunicaciones en los países.

Para integrar grupos de países con un avance similar en cuanto a su equipamiento tomando en cuenta los nueve indicadores de telecomunicaciones descritos en el Cuadro 1, se usa la técnica de partición conocida como K-Medias⁷. Dicha técnica es aceptada generalmente por la robustez de sus resultados y utiliza la distancia entre los datos para saber si éstos son parecidos o diferentes entre sí. Las técnicas permiten identificar los países que sean relativamente homogéneos entre sí, con base en la disponibilidad actual de los nuevos servicios de telecomunicaciones de manera simultánea, formando grupos integrados por países con la mayor similitud entre sí, pero diferentes respecto a otros grupos definidos.

En el análisis se emplea una segunda técnica de agrupamiento, el método Ward⁸. El empleo de dos técnicas diferentes permite contrastar los resultados y tener mayor solidez en las conclusiones. Para una mayor descripción véase el Anexo Estadístico.

A partir de este ejercicio estadístico, se definieron seis grupos o *clústers* como el número óptimo a considerar. Para llegar a esta conclusión se utiliza una gráfica conocida como *Screeplot* la cual se muestra en el Anexo Metodológico, y en la cual la idea consiste en que en aquella parte de la gráfica en donde se muestre un quiebre será un indicativo de que la calidad del modelo comienza a aumentar de manera más lenta conforme se incrementa el número de *clústers*. Es decir, la calidad del modelo ya no se incrementará de manera sustancial conforme aumenta la complejidad de el mismo (i.e. el número de *clústers*).

Estos agrupamientos se reportan en el Cuadro siguiente:

Cuadro 2. Países según agrupamiento

Agrupamiento	Países
1	China, Georgia, Rusia

⁷ Para una descripción detallada, véase: Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

⁸ El segundo método se trata de una técnica jerárquica conocida como método Ward. La idea de ambas técnicas es que aquellas naciones que compartan características semejantes estarán juntas en un mismo grupo, y a su vez, separadas de los otros grupos definidos. En función de los resultados que generan ambas técnicas, se consideró que el método de K-Medias es el más apropiado para el análisis.

2	Bulgaria, Malasia, Serbia
3	Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México, Turquía
4	Sudáfrica
5	Armenia, Indonesia, Tailandia
6	Botswana, Namibia

La metodología empleada permite establecer que en estos seis *clústers* de países los miembros de cada grupo exhiben niveles similares de equipamiento de servicios de telecomunicaciones fijas y móviles, así como de accesibilidad a la banda ancha fija y móvil. Así, por ejemplo, México comparte un nivel de desarrollo en telecomunicaciones similar a Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia y Turquía, considerando el agregado de los indicadores; pero no así con Armenia o Bulgaria, por citar un ejemplo.

Cabe destacar que las técnicas de agrupamiento permiten integrar en grupos a los países con niveles similares de avance o rezago, pero no dan un orden.

Para ordenar los agrupamientos en cuanto al nivel de desarrollo alcanzado con base en los indicadores elegidos, se construye un índice con base en la suma de las nueve variables estandarizadas⁹ para cada país y grupo. El índice tiene un valor mayor en la medida que el nivel relativo de cada país es más elevado. El signo negativo del índice sugiere un nivel promedio inferior a la media del conjunto de países al considerar los nueve indicadores de manera simultánea.

Los resultados indican (ver Cuadro 3) que el *clúster* formado únicamente por China, Georgia y Rusia tiene el índice más alto de los seis grupos, el cual es igual a 10. El resultado es congruente ya que estos tres países ocupan las primeras posiciones en varios de los indicadores como en la penetración de telefonía móvil o en la accesibilidad de banda ancha fija y móvil.

En el caso del *clúster* que conforman Bulgaria, Malasia, Serbia este grupo tiene un índice igual a 5.3. Estas naciones destacan, en especial Bulgaria, en los indicadores del porcentaje de población con cobertura 4G y ancho de banda internacional de internet por usuario de internet.

El grupo formado por Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México y Turquía tiene un índice de 2.3 y se clasifica en la tercera posición de los seis *clústers*. Por su parte, Sudáfrica aparece solitaria en el *clúster* 4, ya que alcanza un índice de variables estandarizadas igual a 0.4. En tanto que el quinto *clúster* se encuentra integrado por Armenia, Indonesia y Tailandia, y su índice de variables estandarizadas es igual a -1.7. Por último, el grupo de Botswana y Namibia quedó posicionado en el último puesto con un índice de -7.5.

⁹ La estandarización consiste en una transformación lineal de las nueve variables con objeto de que tengan un promedio de cero y una desviación estándar igual a uno. Este proceso es necesario para poder sumar variables expresadas en unidades diferentes.

Aunado a lo anterior, el análisis se enriquece si se toma en consideración un indicador del nivel de bienestar social de la población. Específicamente se emplea en este reporte el Índice de Desarrollo Humano del Programa de Naciones Unidas para el Desarrollo¹⁰ (IDH, por sus siglas), el cual mide el nivel de desarrollo de cada país atendiendo a variables como la esperanza de vida, la educación o el ingreso per cápita.

Para cada agrupamiento se obtiene el promedio del IDH de los países que integran a cada clúster (ver Cuadro 3). El primer clúster conformado por China, Georgia y Rusia tiene un IDH igual a 0.799. Por su parte, el segundo clúster que conforman Bulgaria, Malasia y Serbia tiene IDH promedio igual a 0.810. El tercer clúster incluye a Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México y Turquía, y su IDH promedio es igual 0.767. Por su parte, el cuarto clúster que conforman Sudáfrica, cuyo IDH es 0.709.

El quinto clúster con Armenia, Indonesia y Tailandia tiene IDH promedio de 0.757. Finalmente, Botswana y Namibia conforman el sexto clúster con un IDH promedio de 0.690.

En suma, se observa que los países que en general presentan los mejores indicadores de avance en las telecomunicaciones también presentan un mayor desarrollo humano.

Cuadro 3. Países según agrupamiento

Agrupamiento	Países	Índice de variables estandarizadas	IDH
1	China, Georgia, Rusia	7.5	0.799
2	Bulgaria, Malasia, Serbia	5.3	0.810
3	Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México, Turquía	2.3	0.767
4	Sudáfrica	0.4	0.709
5	Armenia, Indonesia, Tailandia	-1.7	0.757
6	Botswana, Namibia	-7.5	0.690

¹⁰ <http://hdr.undp.org/en/content/2019-human-development-index-ranking>

Conclusiones

El presente Reporte permite obtener una perspectiva del desarrollo relativo de México respecto a otros países con economías en desarrollo pertenecientes a diferentes continentes, y con un sector de las telecomunicaciones que igualmente se encuentra en desarrollo. Lo anterior, mediante un análisis de nueve variables las cuales incluyen la penetración de servicios o de tecnologías del sector de las telecomunicaciones, así como la accesibilidad a la banda ancha fija y móvil, utilizando dos técnicas de agrupamiento de *clústers*: el método de K-medias y el método Ward. Asimismo, se opta por seleccionar seis *clústers* o agrupamientos.

El reporte permite establecer que México mantiene un nivel de desarrollo de las telecomunicaciones al nivel de naciones como Brasil, Colombia o Turquía.

La creciente importancia de las telecomunicaciones en las actividades económicas y en el nivel de vida de las personas hace imperativo cerrar la brecha en el equipamiento de los servicios de telecomunicaciones en México con respecto a otros países o regiones.

Con base en los resultados obtenidos en el presente estudio se recomienda continuar el diseño de políticas públicas y regulación para impulsar una mayor penetración y accesibilidad de los servicios.

ANEXO METODOLÓGICO

Aplicación de técnicas de agrupamiento a variables de telecomunicaciones de países con economías emergentes

El objetivo del presente trabajo es identificar las similitudes o disimilitudes que guardan 20 países con economías emergentes, incluyendo a México, con respecto a la penetración de los diferentes servicios de telecomunicaciones. Para la consecución de este objetivo se aplicaron dos técnicas de agrupamiento. La primera es el método Ward, el cual usa una técnica jerárquica que se representa mediante un dendograma¹¹, el segundo se refiere a una técnica de partición conocida como K-medias. Ambas técnicas se engloban en lo que se conoce como análisis de *clúster*, y consisten en integrar casos en grupos homogéneos de acuerdo con sus similitudes.

Para aplicar las técnicas de *clúster* se estandarizan los datos, por lo que para cada variable se calcula la media y la desviación estándar de cada serie de datos; posteriormente, se resta a cada observación la media y se divide entre la desviación estándar de esa variable¹². Una vez hecho lo anterior, el primer paso consiste en calcular la *Suma de Cuadrados al interior de los Grupos* (Sum of

¹¹ Dendograma. Gráfico que muestra en qué orden se han unido los clústers y cuál es el grado de proximidad que tienen los mismos. Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to Data Mining*. Addison-Wesley, 2006.

¹² $(x_i - \bar{x}) / \text{Std. Dev}(x)$.

Squares Within Groups o *SSW* por sus siglas en inglés) para los grupos. Posteriormente se grafica la *SSW* con respecto al número de grupos para minimizar la *SSW*. A este tipo de gráfica se le conoce como *Screeplot* (Gráfica A1), y permite determinar, de una forma visual, el número de *clústers* óptimo a utilizar. La regla consiste en observar en qué parte se rompe la estructura de la curva para volverse cada vez más plana. Una vez que se tiene el número de *clústers* se procede al agrupamiento de los países.

La primera técnica de agrupamiento que se utiliza es la jerárquica mediante el método Ward¹³, la cual arroja un dendograma. Posteriormente, se emplea la técnica de partición mediante la utilización del algoritmo de K-Medias, la cual arroja un agrupamiento.

Metodología y Análisis de clústers.

En esta sección se presenta el análisis de clúster que es utilizado en este reporte analítico, iniciando con un gráfico *Screeplot*¹⁴ (Gráfica A1) el cual es de utilidad para determinar el número de grupos a utilizar. La forma de interpretar la gráfica *Screeplot*¹⁵ consiste en observar en qué parte se rompe la curva formándose una especie de “codo”, lo cual es indicativo de que la calidad del modelo no se incrementa conforme van aumentando la cantidad de clústers. Es así que la gráfica, así como la experiencia en estos estudios sugieren trabajar con seis clústers.

Una vez determinado el número de clústers con el cual se va a trabajar se inicia con la aplicación de un método jerárquico conocido como el método Ward¹⁶. Este método construye una estructura conocida como dendograma, en la cual muestra en qué orden se han unido los clústers y cuál es su grado de proximidad. En este sentido se obtienen dos dendogramas para cada clúster en virtud de que utilizamos dos tipos de distancias en el algoritmo para darle robustez a los resultados: la Euclidiana y la Manhattan. Por una parte, la distancia Euclidiana se refiere a que la distancia más corta que une a dos puntos es una línea recta; en tanto que la distancia Manhattan consiste en que la distancia más corta viene dada por la suma de los dos catetos de un triángulo rectángulo¹⁷.

Al utilizar el método Ward con la distancia Euclidiana y seis clústers (Gráfica A2) se observa que México forma un agrupamiento junto Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia y Turquía.

Para dar robustez al análisis técnico, se sustituye la técnica jerárquica por la técnica de partición mediante el algoritmo de K-Medias repitiendo la estimación para seis clústers (Gráfica A4). La idea básica detrás del algoritmo de K-medias consiste en definir los agrupamientos de tal manera que la variación intra-grupo sea minimizada. El algoritmo de K-medias se puede resumir de la siguiente forma:

i. Se especifica el número de *clústers* (*k*) que serán utilizados.

¹³ Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

¹⁴ *Screeplot*. Diagrama en el cual se grafican los eigenvalores de los factores o de los componentes principales. Se utiliza para determinar el número de factores o componentes principales a mantener. Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to Data Mining*. Addison-Wesley, 2006.

¹⁵ <https://campus.datacamp.com/courses/unsupervised-learning-in-r/unsupervised-learning-in-r?ex=9>

¹⁶ Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

¹⁷ <https://xlinux.nist.gov/dads/HTML/manhattanDistance.html>

ii. A partir de los datos se seleccionan de forma aleatoria k puntos como los centros iniciales (centroide¹⁸) o medias.

iii. Cada observación es asignada al centroide más cercano, basado en la distancia Euclidiana entre el punto y el centroide.

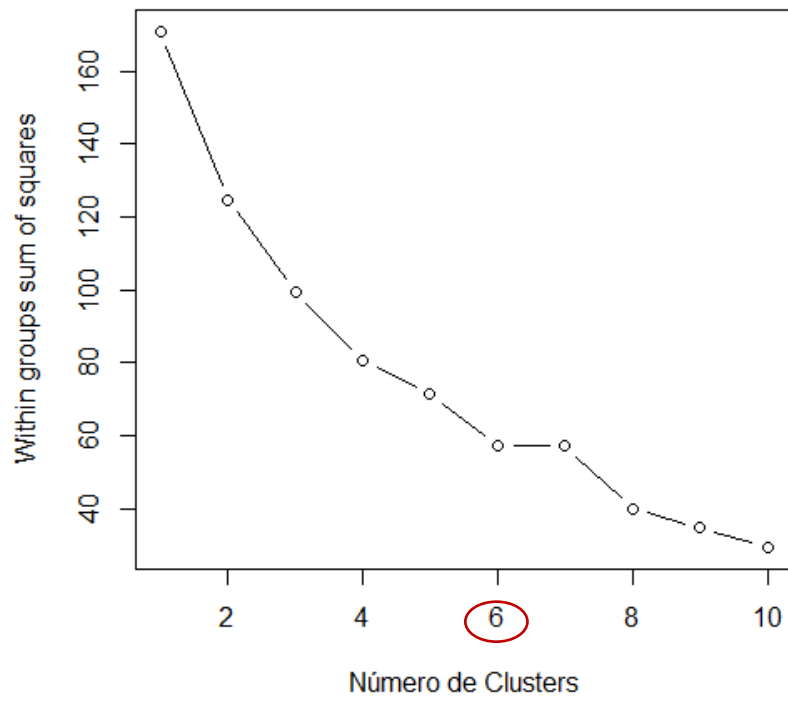
iv. Para cada uno de los k *clústers* se actualiza el centroide del *clúster* por medio del cálculo del nuevo promedio para todos los puntos en el *clúster*. El centroide de un clúster K es un vector de longitud p que contiene las medias de todas las variables para las observaciones del K -ésimo clúster, p se refiere al número de variables.

v. Se minimiza de forma iterativa el total de la suma de cuadrados. Esto es, se iteran los pasos iii y iv hasta que el *clúster* deje de cambiar o se alcance el número máximo de iteraciones.

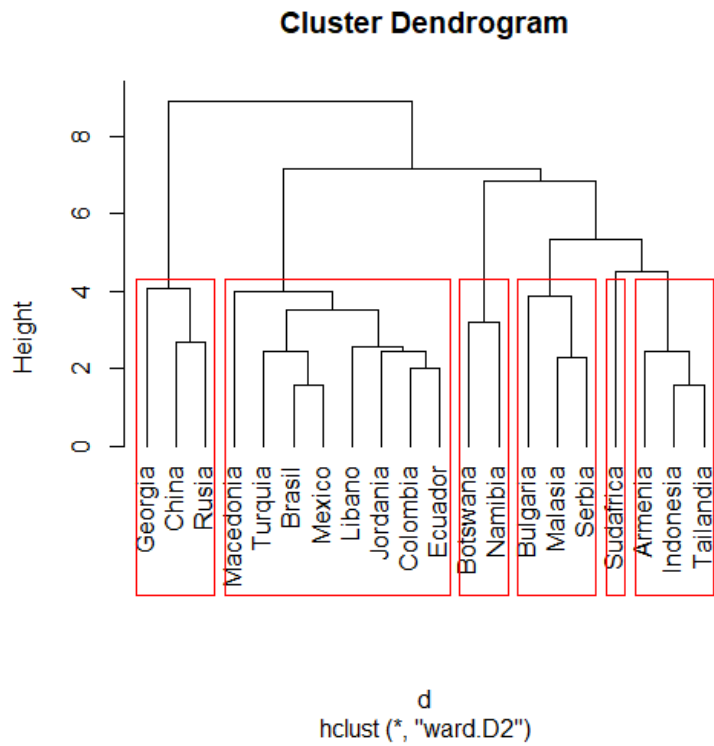
Aplicando este método para definir seis *clústers* (Gráfica A4) se observa que el México se agrupa junto a Brasil, Colombia, Ecuador, Jordania, Líbano y Macedonia. Los resultados se resumen en el Cuadro A1. El ejercicio estadístico completo considera que seis es el número óptimo de *clústers* a considerar.

Gráfica A1. Screeplot

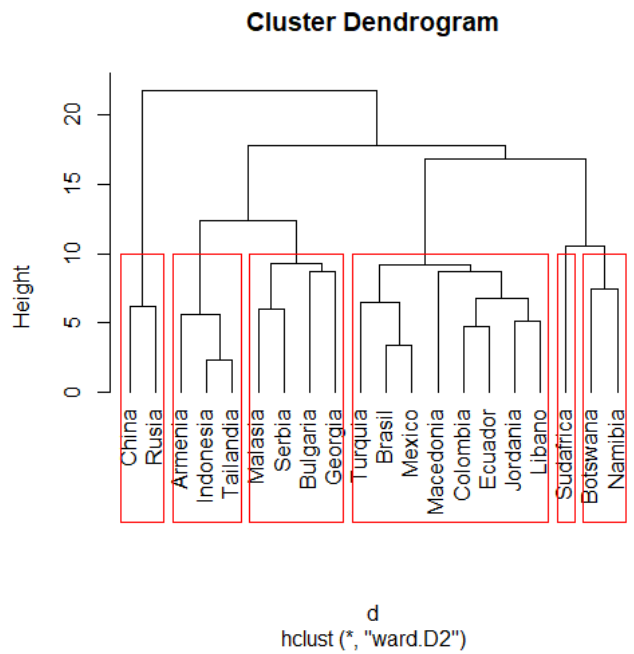
¹⁸ El centroide de un clúster se define como el punto equidistante de los objetos pertenecientes a dicho clúster. Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to Data Mining*. Addison-Wesley, 2006.



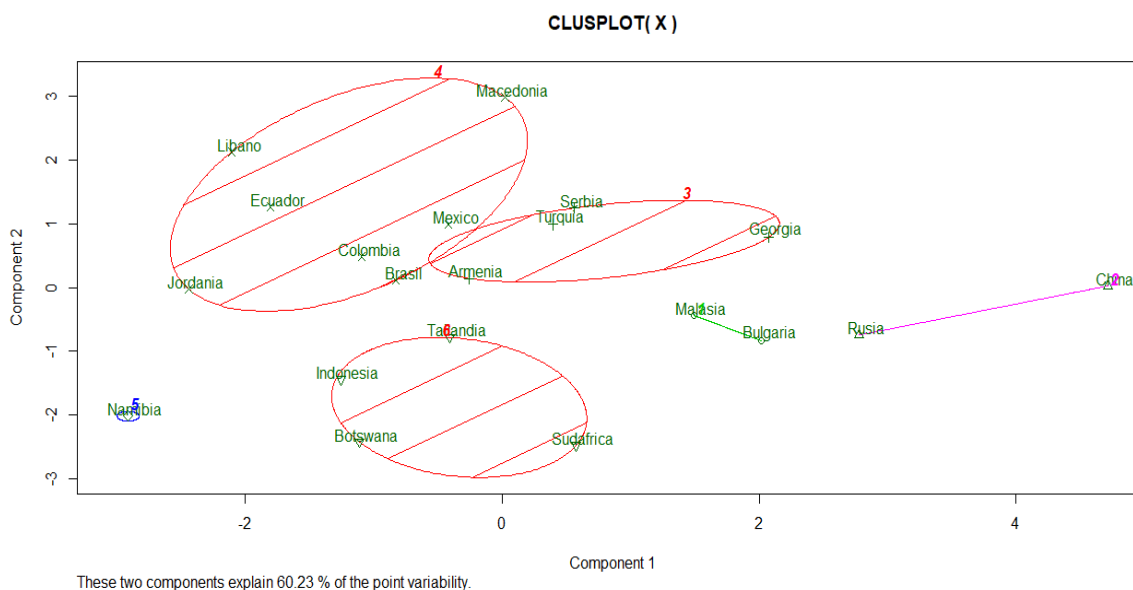
Gráfica A2. Método Ward. Dendrograma-Distancia Euclidiana con seis Clústers



Gráfica A3. Método Ward. Dendrograma-Distancia Manhattan con seis Clústers



Gráfica A4. Método K-Medias con seis Clústers



Cuadro A1. Resultados para seis agrupamientos según metodología de *clústers*

Método Ward		K-medias
Distancia Euclidiana	Distancia Manhattan	
China, Georgia, Rusia	China, Rusia	China, Rusia
Bulgaria, Malasia, Serbia	Bulgaria, Malasia, Serbia	Bulgaria, Malasia
Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México, Turquía	Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México, Turquía.	Brasil, Colombia, Ecuador, Jordania, Líbano, Macedonia, México
Sudáfrica	Sudáfrica	Armenia, Georgia, Serbia, Turquía
Armenia, Indonesia, Tailandia	Armenia, Indonesia, Tailandia	Botswana, Indonesia, Tailandia, Sudáfrica
Botswana, Namibia	Botswana, Namibia	Namibia