

Desarrollo sostenible: Un análisis con ACP Robusto y K-Medias

Lara María Elisa y Porras Hugo

Marzo de 2020



Definición del desarrollo sostenible

- El desarrollo sostenible se define como “la satisfacción de las necesidades de la generación presente sin comprometer la capacidad de la generación futura de satisfacer sus propias necesidades” (Comisión Mundial sobre el Medio Ambiente y Desarrollo, 1987).

Un poco de historia

Comité Económico y Social Europeo, 2018:

- 1922, Río de Janeiro: “Cumbre de la Tierra de Río” - Programa 21.
- 2000, Nueva York: Declaración del Milenio de las Naciones Unidas. Reducción de la pobreza extrema a través del establecimiento de ocho objetivos, conocidos como los “Objetivos del Desarrollo del Milenio”.
- 2012, Río de Janeiro: “Río+20”. Objetivos de Desarrollo del Milenio + nuevas temáticas como el manejo de los recursos naturales, la producción y consumo sostenible, instituciones efectivas entre otros.
- 2015, Asamblea General de las Naciones Unidas: “Agenda 2030 para el desarrollo sostenible”.

¿Qué incluye la agenda actual?

- La Agenda 2030: "un compromiso universal adquirido tanto por países desarrollados, como en desarrollo en el marco de una alianza reforzada, que toma en cuenta los medios de implementación para realizar el cambio y la prevención de desastres por eventos naturales extremos, así como la mitigación y adaptación al cambio climático" (ONU y CEPAL, 2016)
- Cinco elementos: inclusión social, crecimiento económico, protección ambiental, paz y asociatividad.
- Instrumentación a través de 17 objetivos de desarrollo sostenible (ODS), 169 metas y más de 370 indicadores (ONU, 2015).

Tópicos de cada objetivo



Variables adicionales

Cinco variables: Coeficiente de Gini (por ingreso disponible e ingreso de mercado), Tamaño de la población, Fuerza Laboral, Territorio.

- Coeficiente de Gini: Si bien el ODS 10 plantea como variables de medición a la tasa promedio de crecimiento de consumo per cápita, el costo promedio de transacción de las remesas y la asistencia oficial neta para el desarrollo, no existen medidas de disparidad económica, como el coeficiente de Gini - medida de desigualdad más popular y usada a nivel mundial, dada su interpretación intuitiva (Chibba, 2016; Holden, Linnerud, Banister, 2014; Organización de Naciones Unidas, 2010; Allison, 1978).

Variables adicionales

Cinco variables: Coeficiente de Gini (por ingreso disponible e ingreso de mercado), Tamaño de la población, Fuerza Laboral y Territorio.

- Tamaño de la población: Tezanos (2012) sugiere que para establecer una correcta clasificación de los países es necesario ponderar sus diferentes dimensiones, para lo cual se usa como variable el tamaño de la población.
- ODS3 y Crecimiento poblacional: Una reducción en la tasa de mortalidad y fecundidad tendría efecto negativo en el crecimiento de la población. A su vez, una disminución en el crecimiento poblacional afecta de forma positiva al desarrollo sostenible en países en vías de desarrollo (Güney, 2017; Abel, Barakat, Kc, Lutz, 2016).

Variables adicionales

Cinco variables: Coeficiente de Gini (por ingreso disponible e ingreso de mercado), Tamaño de la población, Fuerza Laboral y Territorio.

- Respecto al efecto que tiene la población y el territorio en el desarrollo sostenible, la Organización de Naciones Unidas (2012) establece que no es posible determinar una relación lineal, positiva o negativa entre estos, en tanto esta depende de factores como las características de la población, el uso de la tierra, los patrones de producción y consumo, y la capacidad de resiliencia del ecosistema.
- Sin embargo, una alta concentración de la población puede deteriorar el ecosistema ocupado y saturar los límites de la cadena productiva.

Variables adicionales

Cinco variables: Coeficiente de Gini (por ingreso disponible e ingreso de mercado), Tamaño de la población, Fuerza Laboral y Territorio.

- Koirala y Pradhan (2019) señalan que el desarrollo sostenible puede ser expresado, entre otros, en función del crecimiento de los factores productivos (capital, trabajo y recursos naturales) y la participación de la fuerza laboral.
- Fuerza laboral = Empleados + Desempleados en edad de trabajar. Complementario al ODS 8, Cordera (2017) y Nikolaevna y Vladimirovna (2017) sugieren que para determinar la “brecha laboral” existente se debe considerar el total de la fuerza laboral, pues incluye a la población en edad de trabajar no ocupada disponible y consiste en uno de los factores económicos asociados al desarrollo sostenible.

Fuente de los datos

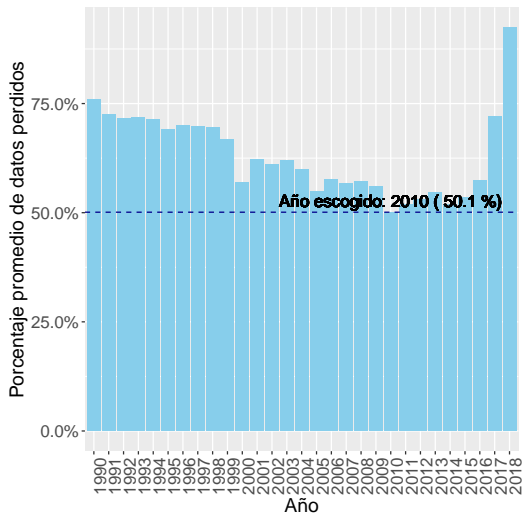
Datos de ODS Agenda 2030:

- Banco Mundial: objetivos de desarrollo sustentable.
- Existen otros caminos: API del Banco Mundial (librería **WDI**, creada por Vincent Arel-Bundock de la Universidad de Montreal).
- 217 economías, año 2010.
Economías del mundo por nivel de ingreso

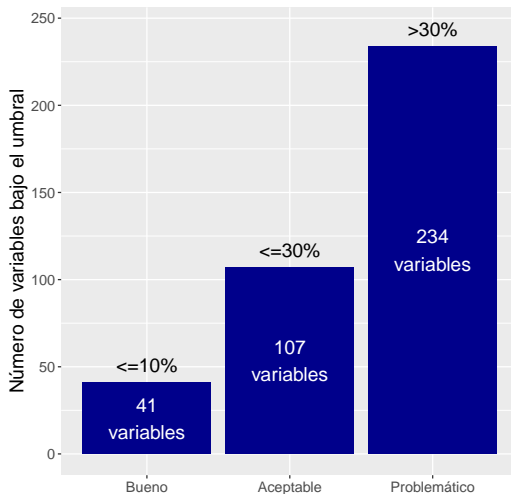
Datos de la SWIID:

- Frederick Solt.

Selección del año de estudio



Selección del umbral de datos perdidos



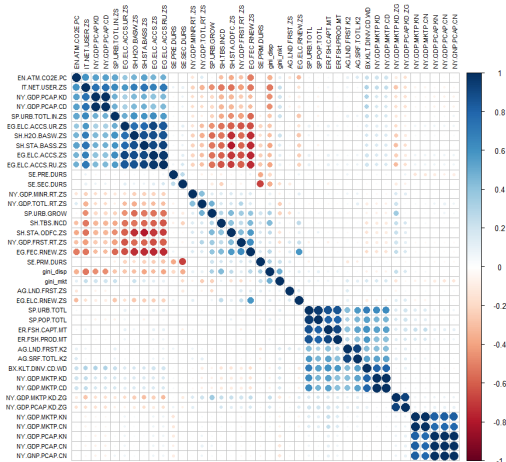
Variables seleccionadas

	MisPor	Variable	GruMis	Porcentaje de datos perdidos
1	0.267	GINI (ingreso disponible)	Aceptable	26.7 %
2	0.267	GINI (ingreso de mercado)	26.7 %	
3	0.097	Ingreso bruto per cápita (moneda nacional corriente)	Bueno	9.7 %
4	0.074	Acceso a electricidad, rural (% de la población rural)	Bueno	7.4 %
5	0.074	Inversión extranjera directa, ingresos netos (dólares corrientes, balanza de pago)	Bueno	7.4 %
6	0.069	PIB (moneda nacional constante)	Bueno	6.9 %
7	0.069	Tasa de crecimiento del PIB (% anual)	Bueno	6.9 %
8	0.069	PIB per cápita (moneda nacional constante)	Bueno	6.9 %
9	0.069	Tasa de crecimiento del PIB per cápita (% anual)	Bueno	6.9 %
10	0.065	Emisiones de CO2 (toneladas métricas per cápita)	Bueno	6.5 %
11	0.065	PIB (dólares constantes del 2010)	Bueno	6.5 %
12	0.065	PIB (dólares corrientes)	Bueno	6.5 %
13	0.065	PIB per cápita (dólares constantes del 2010)	Bueno	6.5 %
14	0.065	PIB per cápita (dólares corrientes)	Bueno	6.5 %
15	0.065	Acceso a internet (% de la población)	Bueno	6.5 %
16	0.065	Rentas obtenidas de la explotación de recursos naturales (% del PIB)	Bueno	6.5 %
17	0.060	Rentas forestales (% del PIB)	Bueno	6 %
18	0.060	Rentas minerales (% del PIB)	Bueno	6 %
19	0.060	Duración de la educación pre primaria (años)	Bueno	6 %
20	0.060	Duración de la educación secundaria (años)	Bueno	6 %

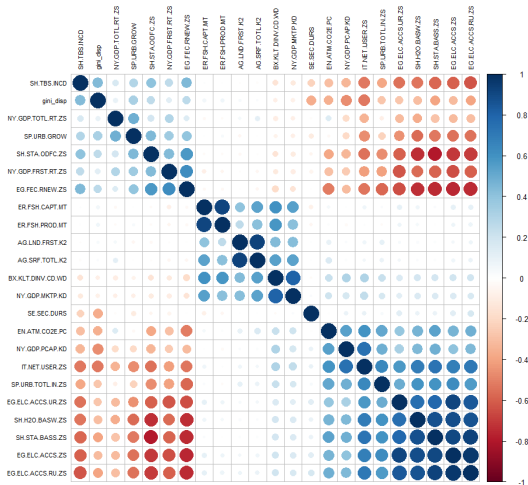
Variables seleccionadas (continuación)

	MisPor	Variable	GruMis	Porcentaje de datos perdidos
21	0.051	PIB (moneda nacional corriente)	Bueno	5.1 %
22	0.051	PIB per cápita(moneda nacional corriente)	Bueno	5.1 %
23	0.046	Duración de la educación primaria (años)	Bueno	4.6 %
24	0.041	Incidencia de tuberculosis (por cada 100,000 personas)	Bueno	4.1 %
25	0.037	Área forestal (% de territorio)	Bueno	3.7 %
26	0.037	Área forestal (kilómetros cuadrados)	Bueno	3.7 %
27	0.028	Personas defecando al aire libre (% de la población)	Bueno	2.8 %
28	0.028	Acceso a servicios sanitarios básicos (% de la población)	Bueno	2.8 %
29	0.028	Consumo de energía renovable (% del consumo final de energía)	Bueno	2.8 %
30	0.018	Producción pesquera, captura (toneladas cuadradas)	Bueno	1.8 %
31	0.018	Acceso a agua potable (% de la población)	Bueno	1.8 %
32	0.018	Producción pesquera total (toneladas métricas)	Bueno	1.8 %
33	0.009	Acceso a electricidad (% de la población)	Bueno	0.9 %
34	0.009	Acceso a electricidad (% de la población urbana)	Bueno	0.9 %
35	0.009	Población urbana	Bueno	0.9 %
36	0.009	Población urbana (% del total)	Bueno	0.9 %
37	0.009	Crecimiento de la población urbana (% anual)	Bueno	0.9 %
38	0.009	Territorio (kilómetros cuadrados)	Bueno	0.9 %
39	0	Código ISO	Bueno	0 %
40	0	Año	Bueno	0 %
41	0	País	Bueno	0 %
42	0	Producción de energías renovables (% de la producción de electricidad total)	Bueno	0 %
43	0	Población total	Bueno	0 %

Análisis de correlaciones



Análisis de correlaciones (continuación)



Variables utilizadas

- Acceso a la electricidad
- Producción pesquera
- Emisiones de CO2 per cápita
- Inversión extranjera directa
- Área total de bosques
- Rentas obtenidas de la explotación de bosques
- PIB en dólares constantes de 2010
- PIB per cápita en dólares constantes de 2010
- Incidencia de la tuberculosis por cada 100.000 personas

Variables utilizadas (continuación)

- Uso de internet
- Porcentaje de la población que defeca al aire libre
- Acceso al agua
- Acceso a servicios sanitarios
- Uso de energía renovable
- Duración de la educación secundaria
- Población urbana
- Territorio
- GINI calculado según ingreso disponible

Análisis de componentes principales robusto (RPCA)

- Desarrollado por Candès et al. (2009) con la finalidad de crear una metodología de componentes principales robusta a datos atípicos o aberrantes.
- Se busca descomponer una matriz de datos con ruido M (i.e. con datos atípicos) en 2 matrices más definidas: una componente dispersa S y una componente de bajo rango L .

Análisis de componentes principales robusto (RPCA)

- Formalmente, se busca descomponer o representar la matriz de datos **normalizada** M de n sujetos y k variables en dos matrices. Una matriz L_0 llamada matriz de rango bajo, y otra llamada S_0 o matriz dispersa.
- La matriz L_0 contiene los datos “**comunes**” de los puntos de la matriz M , mientras que S_0 representa las características “**únicas**” de la matriz M .

Análisis de componentes principales robusto (RPCA)

- Este problema es resuelto a través del algoritmo Principal Component Pursuit (PCP), que busca:

$$\text{Minimizar } \|L\|_* - \lambda \|S\|_1 \text{ sujeto a } L + S = M$$

donde $\|*\|_*$ y $\|*\|_1$ son la norma del núcleo y la norma l_1 de una matriz, respectivamente.

Un vistazo a la matriz L

	V1	V2	V3	V4	V5
Brazil	0.644	0.647	0.497	0.277	-0.035
China	0.673	0.667	0.528	0.526	0.037
Ecuador	0.498	0.431	0.421	-0.090	-0.300
India	-0.094	-0.203	-0.066	0.536	-0.562
United States	0.685	0.801	0.501	0.612	0.960

Un vistazo a la matriz S

	ER.FSH.CAPT.MT	EN.ATM.CO2E.PC	NY.GDP.PCAP.KD	SP.POP.TOTL
BRA	0	-0.426	-0.208	1.036
CHN	10.828	0.229	-0.383	9.795
ECU	0.080	-0.111	-0.047	-0.015
IND	2.660	-0.020	-0.037	8.945
USA	2.347	1.095	0.457	1.736

¿Diferencia contra el ACP?

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Brasil	2.289	5.953	1.913	0.096	-5.662
China	4.882	24.231	-4.495	-0.130	5.610
Ecuador	0.583	-0.432	-0.884	1.064	-0.267
India	-0.909	9.728	-3.703	-0.629	2.156
Estados Unidos	6.441	12.493	4.722	-1.310	-2.516

Agrupamiento

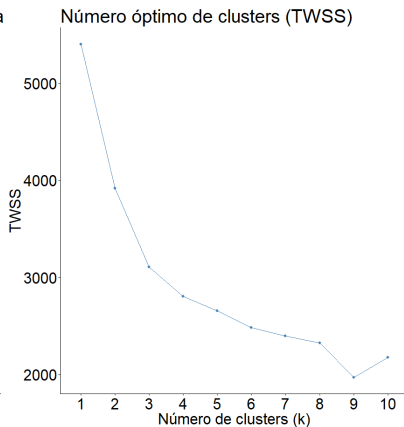
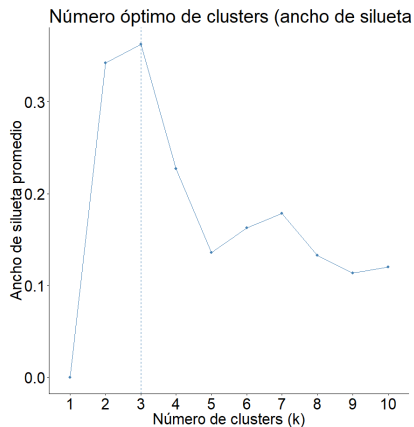
- Por fines didácticos realizaremos dos ejercicios de agrupamiento.
- El primero se realizará sobre la matriz original de datos
- El segundo sobre la matriz de bajo rango L , obtenida del ACP robusto.
- Debido a que todas las variables son continuas, usaremos la técnica k-medias.

K-Medias

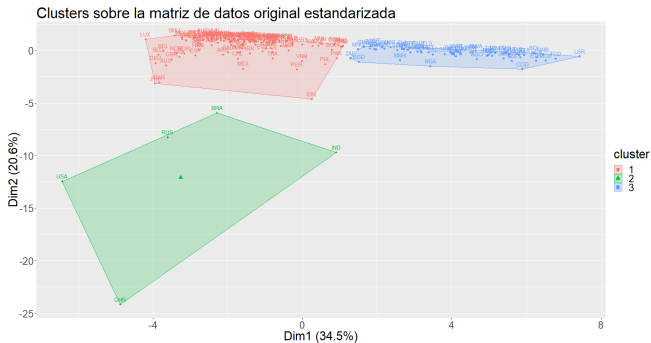
Dado un conjunto de datos X con n observaciones, se busca construir $k \leq n$ grupos mutuamente excluyentes, a fin de minimizar la suma de cuadrados de las distancias S de cada grupo (i.e. la distancia de cada observación en el grupo a la media μ_i de éste), es decir, que las observaciones dentro un grupo sean lo más parecidas posible.

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \mu_i\|^2$$

Agrupamiento sobre los datos originales - Grupos

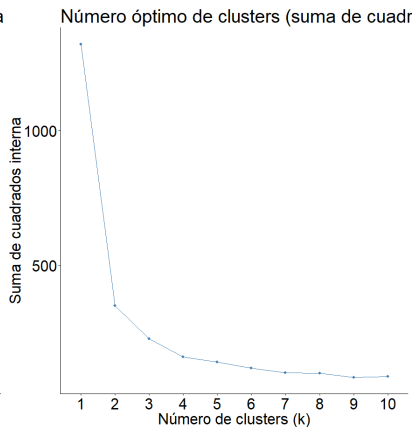
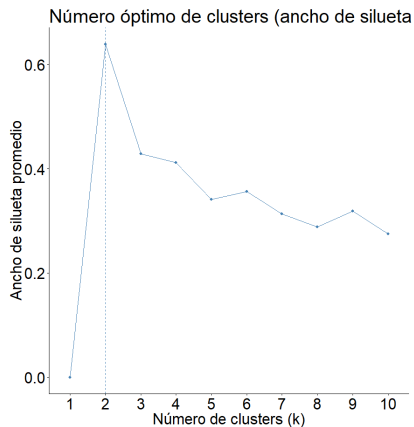


Agrupamiento sobre los datos originales - Resultado

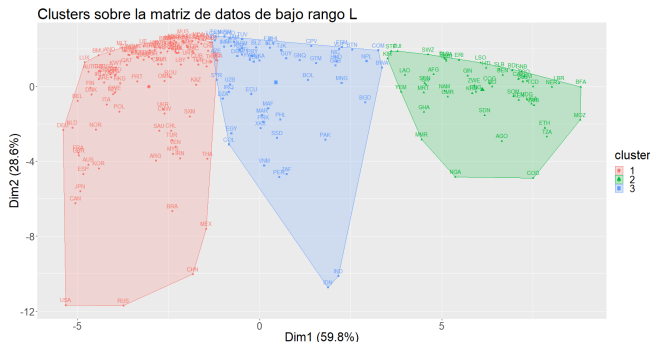


$$\frac{BetweenSS}{TotalSS} = 42.46\%$$

Agrupamiento sobre la matriz L - Grupos



Agrupamiento sobre los datos originales - Resultado



$$\frac{BetweenSS}{TotalSS} = 82.65\%$$

Cluster 1

Cluster de desarrollo alto

Statistic	N	Mean	St. Dev.	Min	Max
PIB per cápita	101	27,750.35	26,937.04	1,631.54	144,569.20
Emisiones de CO2 per cápita	100	8.75	7.05	0.53	40.74
Índice GINI	79	34.56	6.70	24.50	51.50
Acceso a servicios sanitarios básicos	107	94.97	6.59	65.62	100.00

Cluster 2

Cluster de desarrollo bajo

Statistic	N	Mean	St. Dev.	Min	Max
PIB per cápita	49	1,172.96	1,019.36	231.80	5,191.58
Emisiones de CO2 per cápita	50	0.36	0.33	0.02	1.43
Índice GINI	41	41.13	6.44	30.10	60.10
Acceso a servicios sanitarios básicos	50	29.21	14.97	5.69	66.26

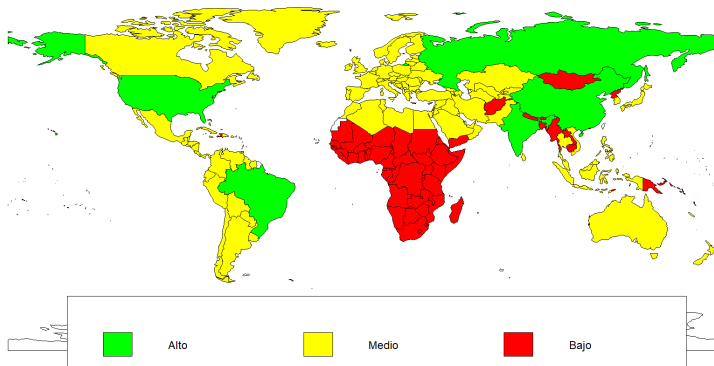
Cluster 3

Cluster de desarrollo medio

Statistic	N	Mean	St. Dev.	Min	Max
PIB per cápita	53	5,388.00	10,038.06	592.18	73,935.68
Emisiones de CO2 per cápita	53	2.08	1.65	0.19	9.19
Índice GINI	39	42.40	6.91	30.20	59.50
Acceso a servicios sanitarios básicos	54	74.28	17.69	32.02	100.00

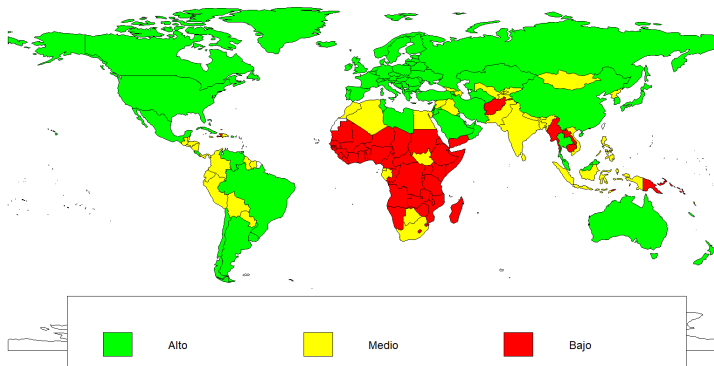
Comparación de resultados (datos originales)

Cluster K-Medias (sobre la matriz de datos original)



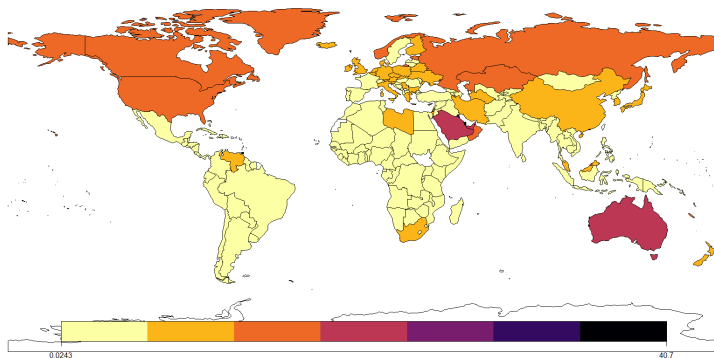
Comparación de resultados (Matriz L)

Cluster K-Medias (sobre la matriz L)

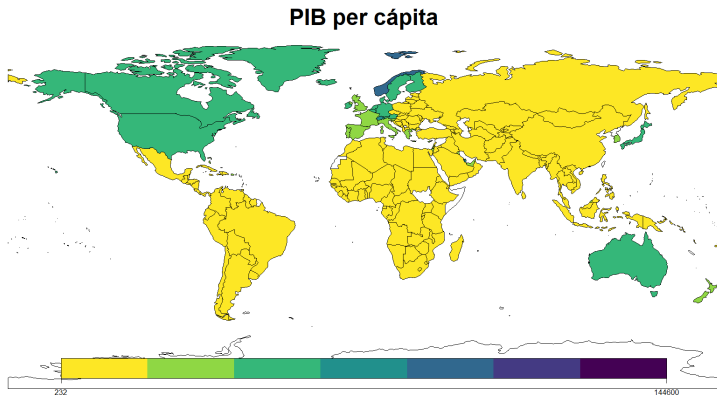


Emisión de CO2 per cápita

Emisiones de CO2 per cápita

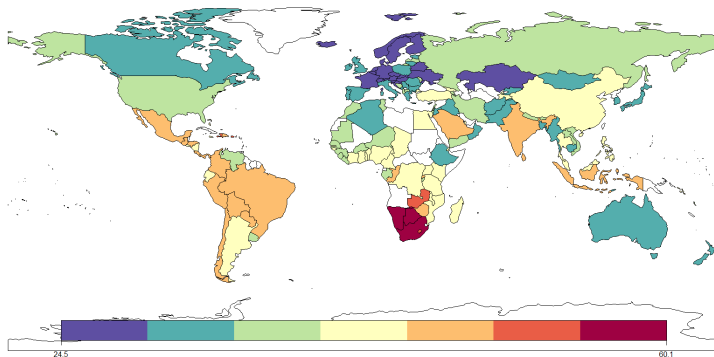


PIB per cápita



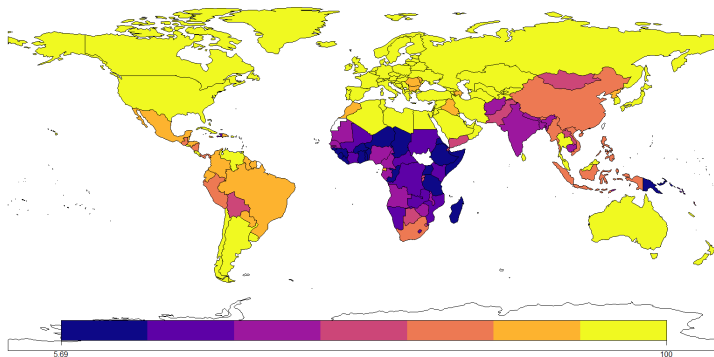
Desigualdad por ingreso

GINI por ingreso disponible



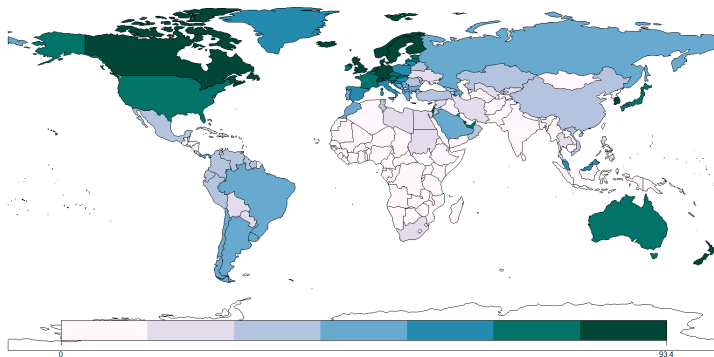
Acceso a servicios sanitarios

Acceso a servicios sanitarios



Uso del internet

Uso de internet



Conclusiones

- 1 Países africanos: variables de acceso a servicios, desigualdad y acceso a tecnología tienen valores bajos en comparación al resto del mundo.
- 2 En tal continente, países como Sudáfrica y Nigeria son puntos que resaltan ya que están emergiendo, alcanzando cada vez mejores valores en varias dimensiones, lo que ha causado que hayan sido clasificados dentro del grupo de desarrollo medio.
- 3 Países europeos y norteamericanos, junto a algunos latinoamericanos como Brasil, han quedado dentro del grupo de desarrollo alto.

Principal deseo

Esperamos inspirar al lector a explorar técnicas nuevas que conlleven a un análisis más preciso y minucioso de temáticas de interés tanto para economistas, decisores y hacedores de política económica, así como para el público en general, donde el conocimiento sea un puente hacia el bienestar personal y social.

“El desarrollo puede concebirse (...) como un proceso de expansión de las libertades reales de que disfrutaban los individuos. El desarrollo exige la eliminación de las principales fuentes de privación de la libertad: la pobreza y la tiranía, la escasez de oportunidades económicas y las privaciones sociales sistemáticas, el abandono en que puedan encontrarse los servicios públicos y la intolerancia o el exceso de intervención de los estados represivos” -Amartya Sen (2000)

ODS



Bibliografía

- 1 Comité económico (2018). SDO work programme/roadmap 2018 – 2020.
- 2 Organización de las Naciones Unidas (2015). Transforming our world: the 2030 Agenda for Sustainable Development.
- 3 Organización de las Naciones Unidas (2016). Summary for Policymakers.
- 4 Solt Frederick (2019). Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database.

Bibliografía (continuación)

- 5 Wickham Hadley (2017). Easily Install and Load the 'Tidyverse': version 1.2.1.
- 6 Hlavac Marek (2018). Stargazer tutorial.
- 7 Candes Emmanuel J., Li Xiaodong, Ma Yi et al. (2009). Robust Principal Componente Analysis?
- 8 Kassambara (2017). Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning.
- 9 Kassambara (2017). Practical Guide to Principal Component Methods in R.