Forma

Descripción generada automáticamente con confianza media

**Análisis multivariado del Índice de Prosperidad Legatum: Un enfoque basado en componentes principales y agrupamiento no jerárquico**

**PRESENTADO POR:  
Andrés Arturo Pérez Moreno y Laura Vanessa Tejada Trujillo**

**PRESENTADO A:   
PhD. Mario Pacheco**

**ESTADÍSTICA MULTIVARIADA  
FACULTAD DE ESTADISTICA  
BOGOTÁ D.C  
2023**

**Tabla de Contenido**

[Introducción 3](#_Toc151328161)

[Objetivos 4](#_Toc151328162)

[Objetivo General 4](#_Toc151328163)

[Objetivos Específicos 4](#_Toc151328164)

[Justificación 4](#_Toc151328165)

[Marco Teórico 5](#_Toc151328166)

[Metodología 5](#_Toc151328167)

[Descripción de los datos 6](#_Toc151328168)

[Resultados 7](#_Toc151328169)

[Análisis de Componentes Principales para el Índice de Prosperidad 7](#_Toc151328170)

[Indicador de Prosperidad 17](#_Toc151328171)

[Agrupamiento no Jerárquico del Índice de Prosperidad 17](#_Toc151328172)

[Conclusiones 17](#_Toc151328173)

[Referencias 18](#_Toc151328174)

[Anexo 19](#_Toc151328175)

# Introducción

El Índice de Prosperidad Legatum es una herramienta integral y multidimensional que evalúa el bienestar y el progreso de las naciones a través de una variedad de indicadores clave. Desarrollado por el Instituto Legatum, este índice ofrece una perspectiva amplia que va más allá de las métricas económicas tradicionales, abarcando aspectos fundamentales como la salud, la educación, la gobernanza, la economía y la calidad de vida. Su enfoque holístico permite comprender la prosperidad en su contexto más completo, proporcionando una visión detallada de los factores que contribuyen al éxito y la estabilidad de una sociedad.

Conforme Legaum Institute (2023) se compone de 12 pilares de prosperidad divididos en 67 elementos enfocados en políticas, agrupados en tres dominios esenciales para la prosperidad: Sociedades Inclusivas, Economías Abiertas y Personas Empoderadas. Utiliza más de 300 indicadores diferentes de más de 70 fuentes de datos distintas para construir el índice. El proceso involucra la selección de indicadores relevantes para cada elemento de prosperidad con la ayuda de expertos, seguido de la normalización de los indicadores para su comparación entre países. Se asignan pesos a cada indicador y elemento según su importancia relativa para la prosperidad, determinados por la literatura académica y la opinión de expertos, así como su significado estadístico para el bienestar económico y social. Los puntajes se calculan a nivel de elemento y pilar utilizando una suma ponderada de los indicadores, y los países se clasifican según sus puntajes en cada pilar. Los puntajes de dominio y el índice general de prosperidad se determinan asignando igual peso a cada dominio y promediando los puntajes.

Para el desarrollo de este trabajo, se emplean dos métodos estadísticos para desentrañar patrones y tendencias significativas, el Análisis de Componentes Principales (ACP) y el Análisis de Agrupamiento No Jerarquizado. El primero, el ACP, permite identificar las variables más relevantes que explican la variabilidad entre los países evaluados, reduciendo la complejidad de los datos y destacando las dimensiones clave de la prosperidad. El segundo, también conocido como análisis de clustering, agrupa países con características similares, revelando similitudes y diferencias en la prosperidad entre distintas regiones o grupos de naciones. Estos métodos estadísticos proporcionan una comprensión profunda y estructurada de las dinámicas de prosperidad a nivel global, se abordarán de forma más detallada más adelante.

La aplicación de estos métodos estadísticos al Índice de Prosperidad Legatum no solo permite una comprensión más profunda de los factores que impulsan el progreso de una nación, sino que también facilita la identificación de áreas de mejora y desarrollo. Al desglosar las complejas interrelaciones entre los indicadores de prosperidad, el análisis de componentes principales y el análisis de agrupamiento no jerarquizado ofrecen una visión estratégica para los responsables de la toma de decisiones a nivel gubernamental, institucional y empresarial. Esta investigación estadística no solo revela la posición relativa de los países en términos de prosperidad, sino que también brinda una hoja de ruta para el diseño de políticas públicas más efectivas y la implementación de estrategias orientadas a fortalecer áreas específicas en busca de un progreso sostenible y equitativo.

En ese sentido, la estructura de este trabajo parte del objetivo y la justificación de la investigación, seguido por el desarrollo de un marco teórico. Posteriormente, se detallan las metodologías empleadas, incluyendo la descripción de los datos utilizados. Por último, se presentan los resultados obtenidos y se derivan las conclusiones pertinentes.

# Objetivos

## Objetivo General

Llevar a cabo un análisis multivariado al índice de prosperidad Legatum para el año 2019 a partir de las metodologías de componentes principales y el agrupamiento no jerárquico.

## Objetivos Específicos

1. Realizar un análisis de componentes principales con el fin de entender cómo se relacionan las variables y cómo influyen en el comportamiento de los individuos.
2. Construir un indicador Prosperidad
3. Realizar un agrupamiento no jerárquico de los individuos explicando las características relevantes de los grupos resultantes.

# Justificación

El Índice de Prosperidad Legatum se destaca por su amplia mirada hacia el bienestar integral de las naciones. Al abarcar aspectos económicos, sociales, gubernamentales y de calidad de vida, ofrece una perspectiva completa sobre el progreso de una sociedad más allá de indicadores financieros. Este enfoque holístico es crucial en un mundo donde la riqueza material no es el único indicador de éxito. Al considerar factores como salud, educación, gobernanza y seguridad, este índice permite comprender la calidad de vida y el estado de desarrollo de un país, brindando una base sólida para la toma de decisiones políticas, el diseño de políticas públicas y la identificación de áreas de mejora.

La importancia del Índice de Prosperidad Legatum radica en su capacidad para ofrecer una visión más completa y equilibrada del estado de una nación. Al trascender la mera medición del crecimiento económico, permite evaluar el bienestar de las personas, identificar áreas de fortaleza y debilidad, y guiar el camino hacia un desarrollo más inclusivo y sostenible, considerando el impacto de las políticas en la vida cotidiana de los ciudadanos.

# Marco Teórico

De acuerdo con SELA (2020) el Índice de Prosperidad Legatum ofrece una perspectiva única sobre cómo evoluciona la prosperidad a nivel global. Este índice anual evalúa 149 países mediante 104 variables, distribuidas en 9 dimensiones clave. Define la prosperidad como una combinación de riqueza y bienestar, siendo una herramienta integral para comprender el progreso mundial y cómo este se forma, cambia y puede ser fortalecido a lo largo del tiempo. El indicador utiliza un método sofisticado que va desde la selección y ponderación de variables hasta técnicas estadísticas avanzadas, como la regresión, para proporcionar resultados más precisos y ajustados a la realidad. Este índice se convierte así en una herramienta poderosa para líderes y tomadores de decisiones en la formulación e implementación de políticas a nivel mundial.

Phélan (2018), señaló que el índice adopta una perspectiva integral, considerando dimensiones como crecimiento económico, familias sólidas, libertades políticas y religiosas, acceso a la educación, oportunidades de vida y un entorno saludable, tanto a nivel individual como colectivo. Phélan evaluó el indicador en 2017 destacando a Uruguay, Costa Rica y Chile como los mejor clasificados en 2017, mientras que Guatemala, El Salvador y Venezuela obtuvieron puntuaciones más bajas, este enfoque multidimensional y la combinación de indicadores objetivos y subjetivos brindan una visión integral del estado de prosperidad de un país, siendo una herramienta valiosa para comprender el bienestar de sus ciudadanos y orientar políticas de desarrollo.

# Metodología

El desarrollo de este trabajo se realizó en el software RStudio y librerias “FactoMineR2”, “factoextra” y “fpc” que comprenden funciones de metodologías del análisis multivariado de datos. Enseguida se describe cada una de las metodologías utilizadas.

Un primer método, es el Análisis de Componentes Principales (PCA por sus siglas en ingles Principal Component Analysis), conforme Beattie y Esmonde-White (2021) esta metodología es útil para reducir la dimensionalidad de conjuntos de datos lo que permite simplificar s visualización y análisis cuantitativo, el PCA permite filtrar múltiples variables en un número más manejable mientras retiene la mayor parte de la información original para explorar posibles asociaciones entre ellas.

Por su parte, Fernández et al. (2022) resalta que el PCA es más efectivo cuando hay una estructura de correlación marcada entre las variables originales. Si no existe asociación entre las variables, el PCA no resulta útil. Los autores mencionan los diferentes campos, como en estudios de fermentación de anticuerpos monoclonales, procesos de purificación biofarmacéutica, análisis de causas en ciencias biomédicas y estudios de sistemas mecánicos dinámicos.

Una componente principal se construye a partir de la combinación lineal de las variables originales de la base, en ese sentido la j-ésima componente principal se calcula a partir de la siguiente expresión:

Donde:

Corresponden a los vectores propios con valores propios asociados ….

El segundo método corresponde al algoritmo de agrupamiento no jerárquico, de acuerdo con Bettolli, Rivera y Penalba (2019) esta metodología permite reasignar los miembros de los grupos en cada etapa del proceso, un ejemplo es el método de k-medias, un enfoque de agrupamiento no-jerárquico que busca la mejor partición para dividir objetos en k grupos. Su objetivo es minimizar la varianza dentro de cada grupo y maximizar la varianza entre grupos. Se inicia determinando el valor de k, ya sea por técnicas específicas o decisiones subjetivas. Las observaciones iniciales se eligen para maximizar las distancias entre grupos.

De otra parte, Ayala y Parra (2023) el agrupamiento no jerárquico crea grupos nuevos sin seguir una estructura de árbol, a diferencia del agrupamiento jerárquico. En este enfoque, se forman grupos para maximizar o minimizar ciertos criterios de evaluación, sin fusionar ni dividir grupos existentes.

En resumen, aunque los métodos no jerárquicos son ampliamente utilizados debido a su eficiencia y simplicidad, pueden presentar desafíos al tratar con ciertos tipos de datos y formas de agrupación. La selección de la técnica adecuada depende en gran medida de la naturaleza y la distribución de los datos que se están analizando.

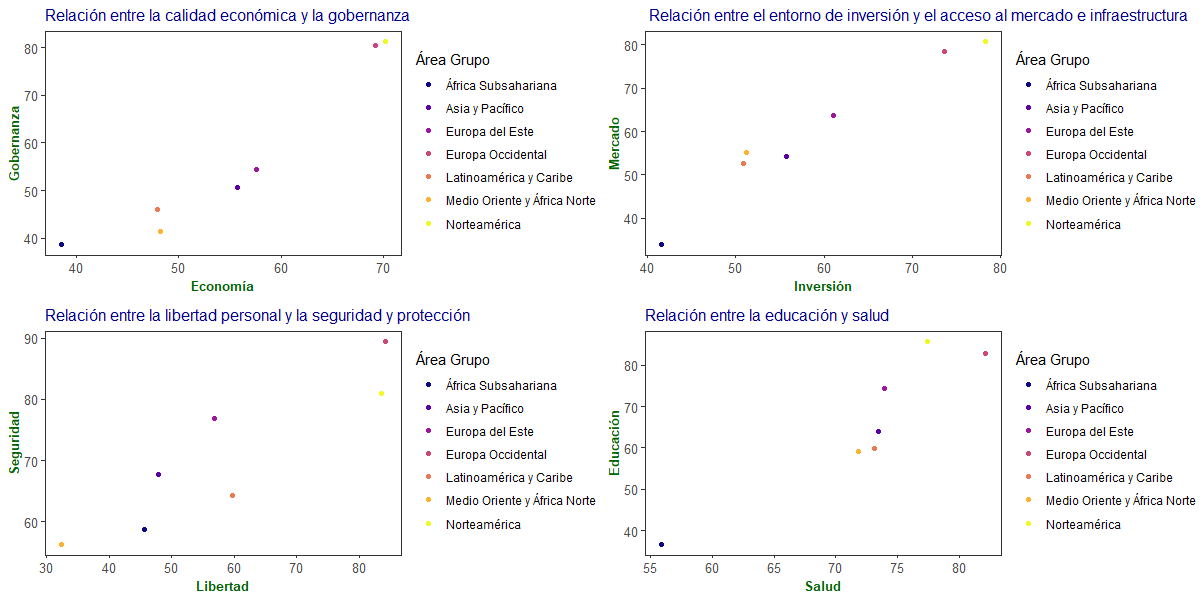
## Descripción de los datos

La base “6. Legatum Prosperity Index 2019” contiene 168 observaciones y 15 variables, donde 3 son cadenas de texto y el restante corresponde a numéricas. El conjunto de datos contiene información de 167 países, agrupados por su ubicación geográfica (7 categorías) acerca de las condiciones socioeconómicas y políticas como educación, salud, seguridad, infraestructura e información, gobernanza, inversión, calidad económica, etc.

Para realizar un análisis exploratorio se agruparon los datos conforme la variable “area\_group” y se promedió cada de las variables numericas para poder mostrar los gráficos de dispersión (Ilustración 1), así mismo para mostrar cada gráfico se unieron variables con similitudes factores económicos y de gobernanza, de inversión y mercado, de libertad y seguridad y de derechos universales como educación y salud.

Conforme la Ilustración 1, se obtuvo que las áreas de Norteamérica y Europa Occidental muestran los mejores desempeños en las variables, lo que sugiere que estos países de este territorio tiene mejores de condiciones de derechos, libertad, gobernanzas y factores económicos a diferencia de regiones como África Subsahariana. Se observa también que la unión de las variables asociadas a la seguridad – protección y libertan tiene una dispersión importante entre los diferentes territorios, siendo Medio Oriente y África Norte el área con el peor desempeño promedio.

Ilustración 1. Legatum Prosperity Index 2019



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

# Resultados

Dado que el objetivo de este estudio esta centrado en tres temáticas, se mostrarán los hallazgos de forma independiente.

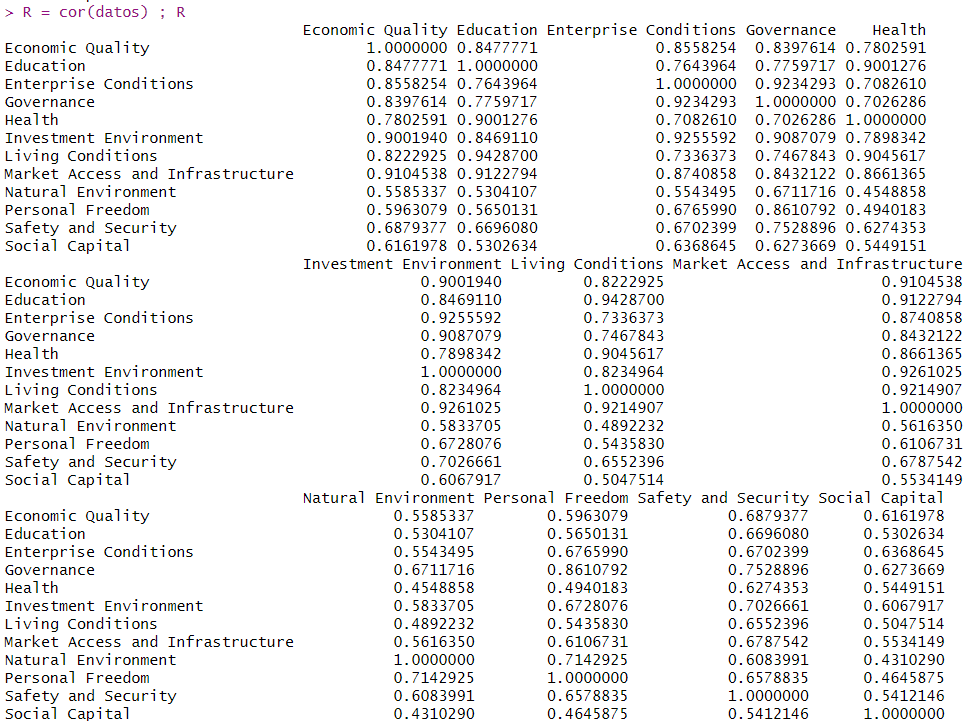
## Análisis de Componentes Principales para el Índice de Prosperidad

Dado que la metodología de componentes principales establece que deben utilizarse los datos estandarizados, se presenta la correlación entre las variables a partir de la Ilustración 2, donde se observa la relación entre 'Education' y 'Health', la correlación es de 0.9001276, lo que sugiere una correlación positiva bastante fuerte entre una educación de calidad y la salud.

También se puede notar que 'Investment Environment' tiene correlación positiva con varias variables, como 'Economic Quality', 'Enterprise Conditions', 'Governance', entre otros, lo que implica que un buen entorno de inversión está relacionado con un mejor desempeño en esos aspectos.

Por otro lado, 'Natural Environment' parece tener correlaciones más bajas en general con otras áreas, lo que podría indicar una relación más débil o menos directa con los demás aspectos evaluados.

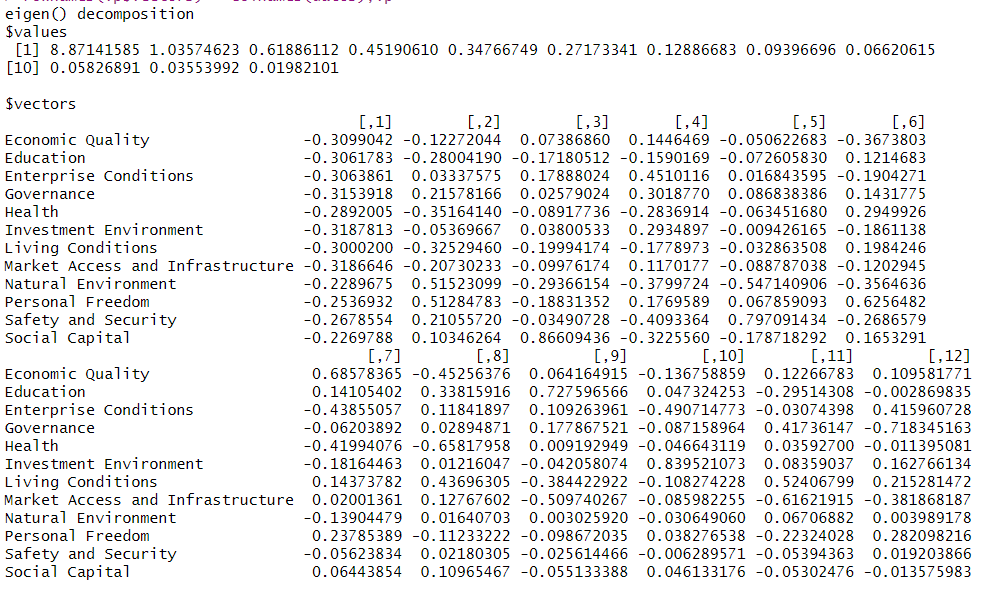
Ilustración 2. Correlación de las variables

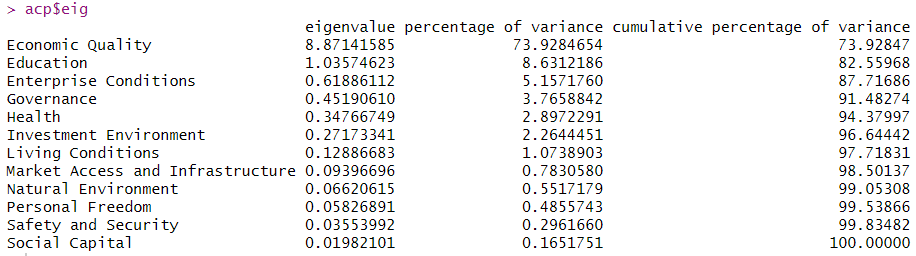


*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El siguiente paso comprende la generación de los valores y vectores propios, dado que se tienen doce variables, se espera que los valores propios tengan la misma cantidad. A partir de la Ilustración 3, se construirán a continuación las primeras tres componentes:

Ilustración 3. Valores y vectores propios para las variables





*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Se observa en la salida de la inercia la primera componente explica el mayor porcentaje de varianza en un 74% seguida del segunda por 8.7% ambas variables acumulan el 82.5% (inercia acumulada). Después de generar los componentes principales para la variable se hará para los individuos. (Ilustración 4)

Ilustración 4. Valores y vectores propios para los países

Imagen que contiene Tabla

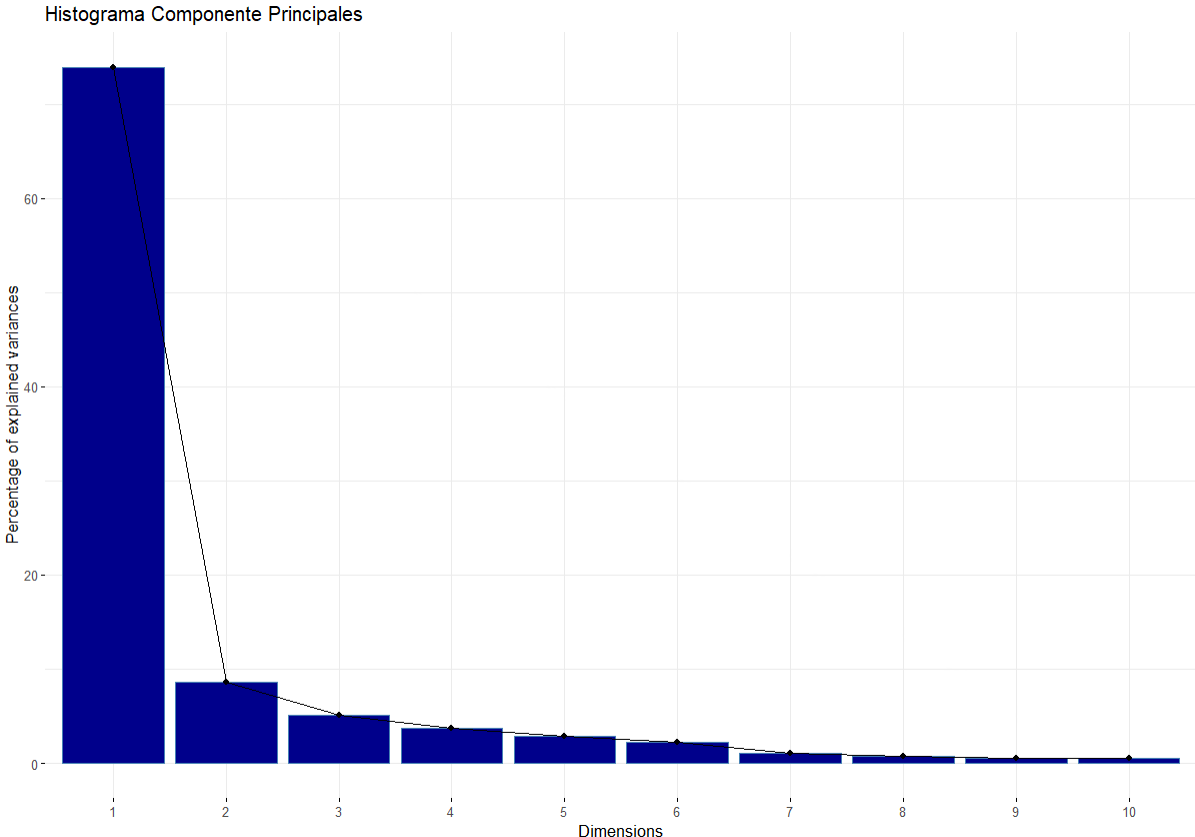
Descripción generada automáticamente  
Imagen que contiene Texto

Descripción generada automáticamente

*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

De forma gráfica se evidencia la fuerte caída que se genera luego del primer componente, a partir del segundo la caída es progresiva.

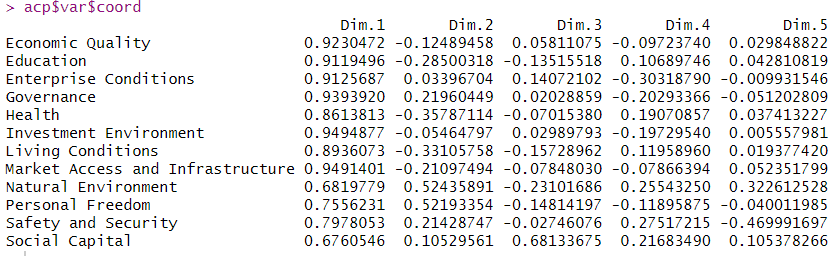
Ilustración 5. Histograma de los componentes principales



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Conforme la Ilustración 6, se observa que la componente explica todas las variables con encima del 67%, donde “Economic Quality” tiene una fuerte influencia positiva en la Dimensión 1. Indica que un aumento en la calidad económica está altamente asociado con valores altos en esta dimensión, caso similar ocurre en variables como “Education”, “Enterprise Conditions” y “Governance”. En cuanto a la componente 2, “Economic Quality” tiene una influencia negativa, lo que sugiere una relación inversa con la Dimensión 1.

Ilustración 6. Aporte de las componentes principales



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Posteriormente, se calculan las medidas de calidad. Por un lado, los cosenos cuadrados, que hace referencia a la correlación de las variables elevada al cuadrado. Donde “Economic Quality” está muy bien representada en la Dimensión 1, lo que significa que explica la mayor parte de su variabilidad en esta dimensión situación que se repite en “Education”, “Enterprise Conditiones”, y “Gobernance”. En caso de la componente 2, la variable “Personal Freedom y “Natural Enviroment” son las de mayor variabilidad en esta dimensión.

Ilustración 7. Cosenos al cuadrado

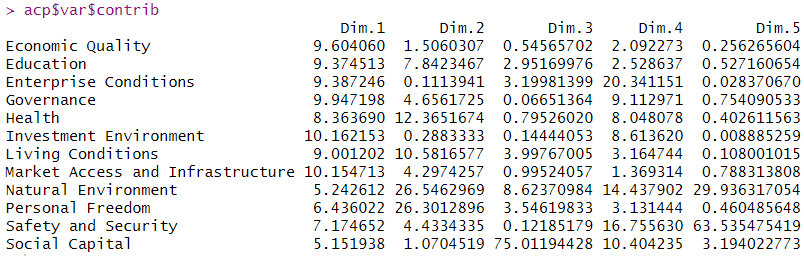
Texto

Descripción generada automáticamente

*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Por otra parte, la medida de calidad asociada a las contribuciones en la Ilustración 8, se tiene que las variables “Investment Environment”, “Market Access and Infrastructure” y “Governance” contribuyen 10.16, 10,15 y 9.94% a la componente 1 mientras que variables como “Natural Environment” y “Social Capital” solo explican el 5.24 y 5.15% respectivamente. En la componente 2 las variables “Natural Enviroment” y “Personal Freedom” contribuye más el 26% y la componente 3 es explicada en un 75% por la variable “Social Capital”

Ilustración 8. Contribución de las variables en las componentes



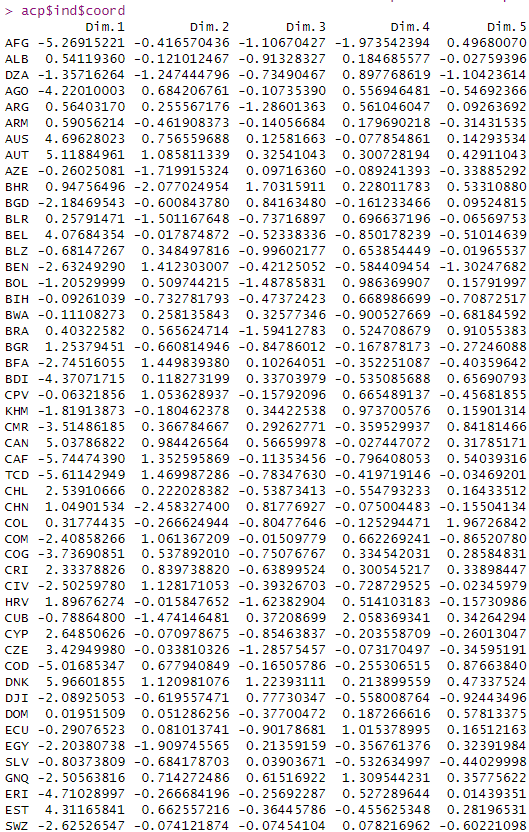
Gráfico, Gráfico de barras

Descripción generada automáticamente

*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

En cuanto a los individuos, en la Ilustración 9 se observa que países como NOR (Noruega), DNK (Dinamarca) y CHE (Suiza) contribuyen más a las componente 1 a diferente de países como AFG (Afganistán), COD (República Democrática del Congo) y SSD (Sudán del Sur)

Ilustración 9. Contribución de los individuos en las componentes

 Tabla

Descripción generada automáticamente  
Tabla

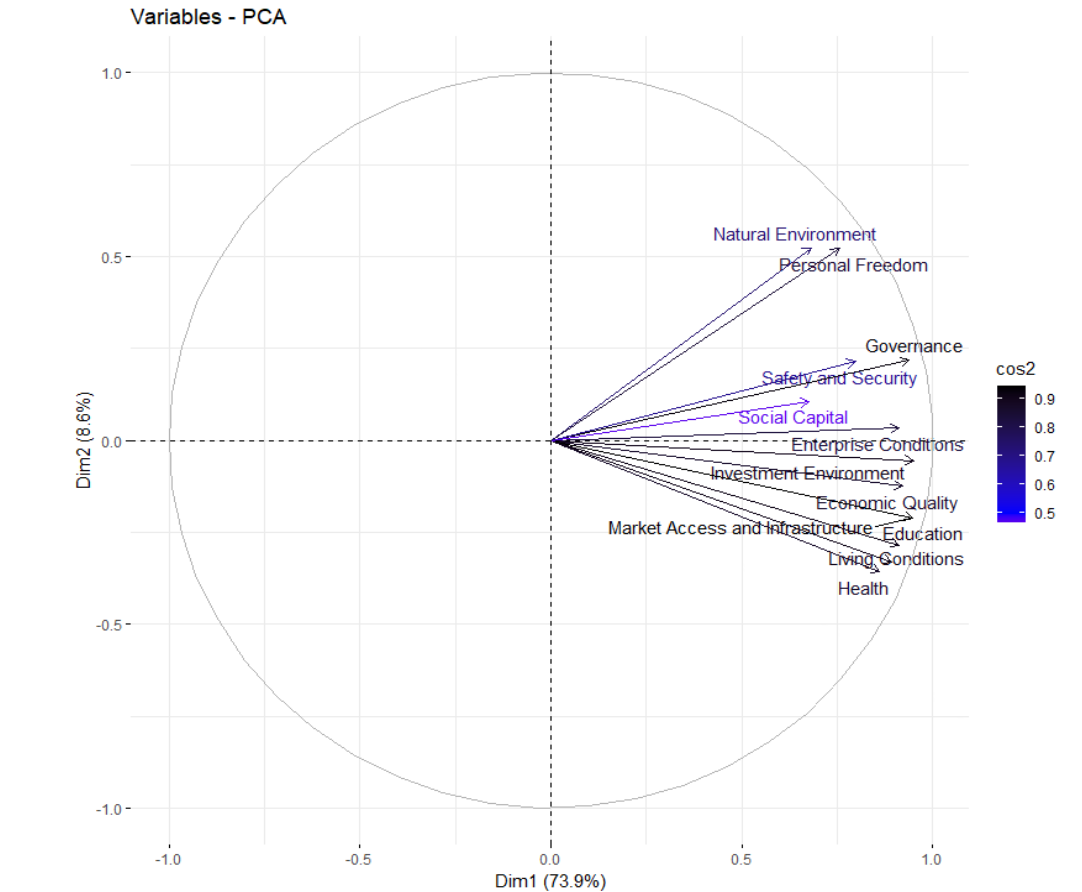
Descripción generada automáticamente Tabla

Descripción generada automáticamente

*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Finalmente, para tomar la decisión se realizó un plano principal con la primera y segunda componente para representar de manera bidimensional los individuos y las variables del estudio (Ilustración 10)

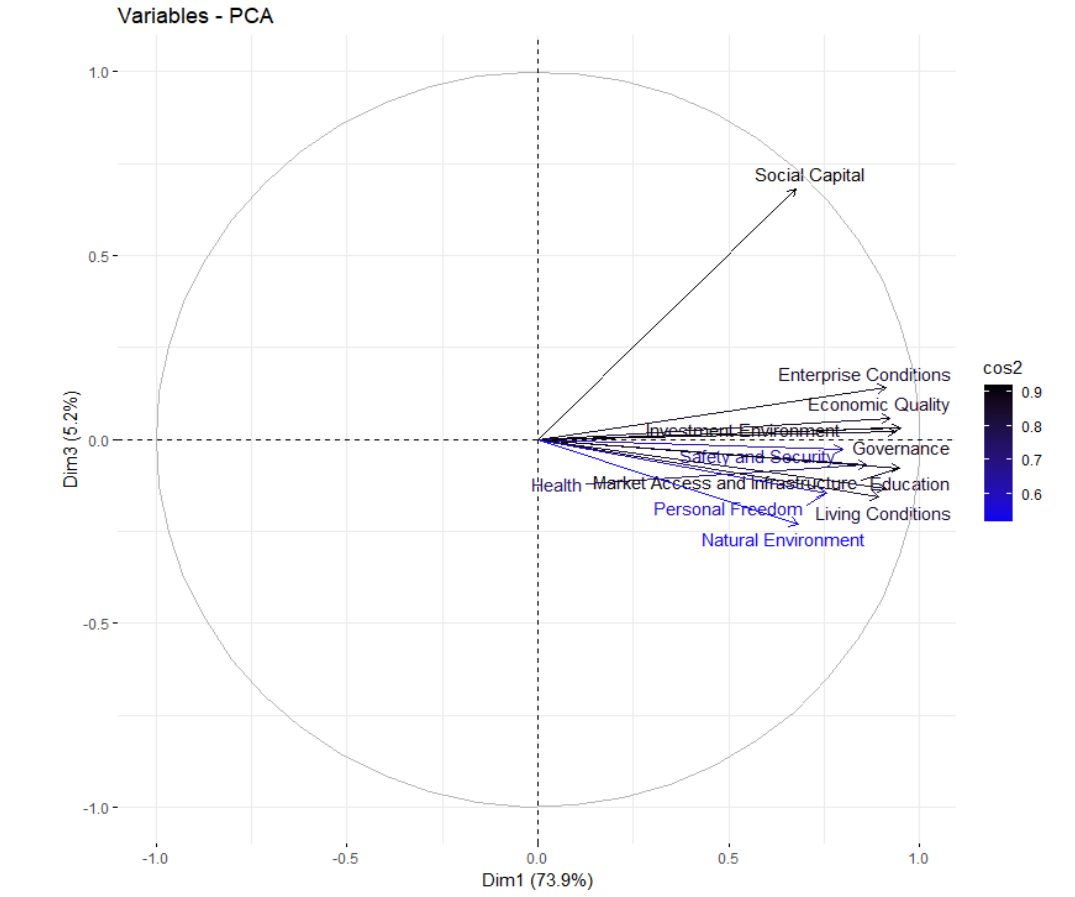
Ilustración 10. Gráfico de Primera y Segunda Componente Principal y Variables.



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

En la primera componente principal, se pueden ver casi todas las componentes de forma robusta, menos capital social y seguridad. La segunde componente principal solo agrupa un 8% de la varianza de las variables.

Ilustración 11. Gráfico de Primera y Tercera Componente Principal y Variables.

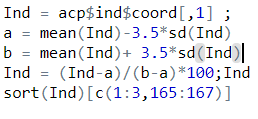


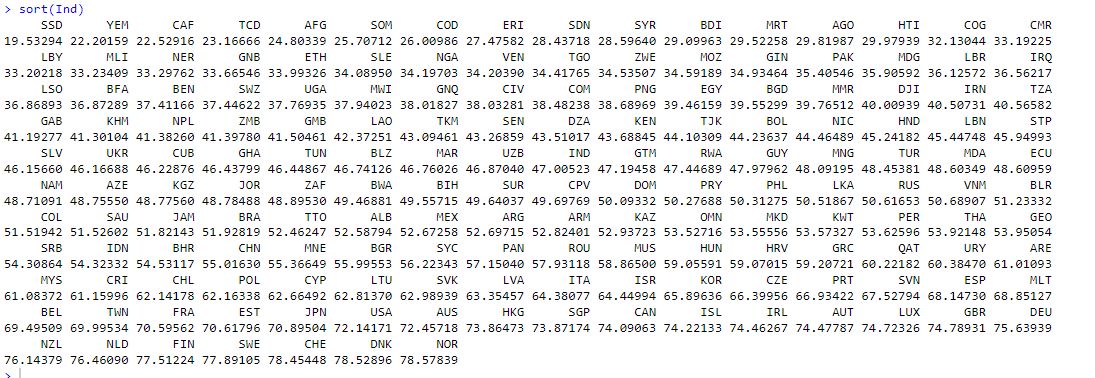
*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

En la Tercera Componente Principal, representa la explicación que parece hacer falta para la variable Capital Social, no se presentara el grafico de la componente 2 y 3 ya que no agrega ninguna información adicional relevante.

## Indicador de Prosperidad

*Ilustración 12. Estandarización del indicador para evitar su incorrecta lectura.*

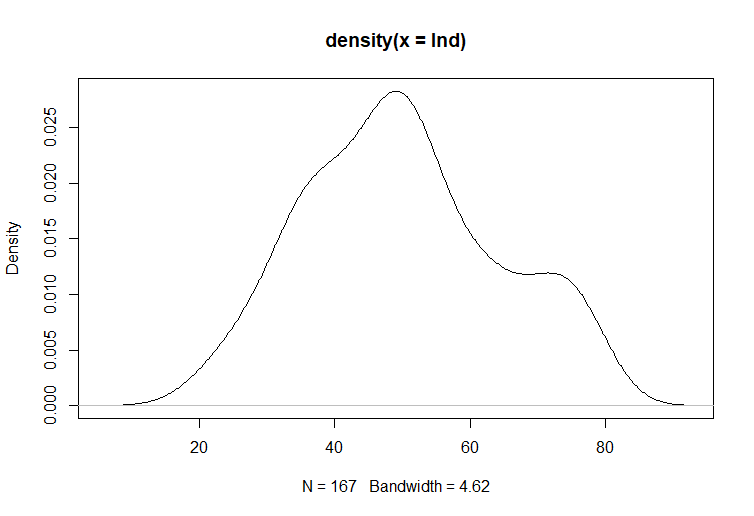
**



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El indicador pasa de estar de 0 a 100 a estar entre 19.53294 y 78.57839 esto para evitar una incorrecta lectura de que podría ser sudan del sur está en 0 algo terrible y noruega está en un 100 un indicador perfecto.

*Ilustración 13. Gráfico de densidad del indicador.*



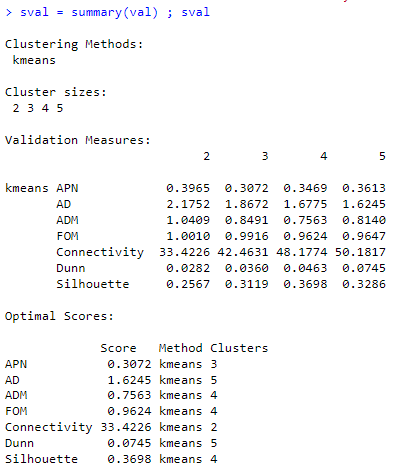
*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El indicador aparenta sesgo y kurtosis, parece tener una cola más larga hacia la derecha y una cola más corta y más alta hacia la izquierda, esto mostrando que hay posiblemente más países, con un indicador bajo que con un indicador alto.

## Agrupamiento no Jerárquico del Índice de Prosperidad

## Método k medias

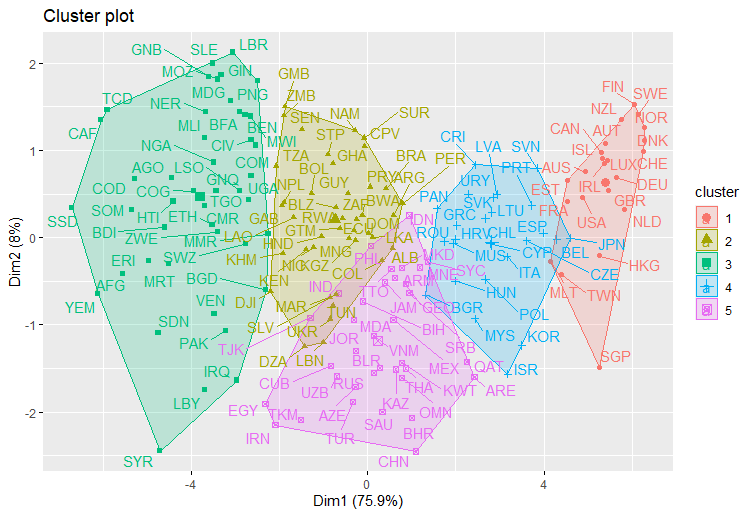
*Ilustración 14. Evacuación de numero de clústeres para K medias.*



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El número de clústeres óptimos por la prueba aparente es de 5 clústeres, en parte por los resultados en parte para poder ser comparado más directamente con el segundo método de k-mediodes.

*Ilustración 15. Grafico agrupaciones por K-Medias*

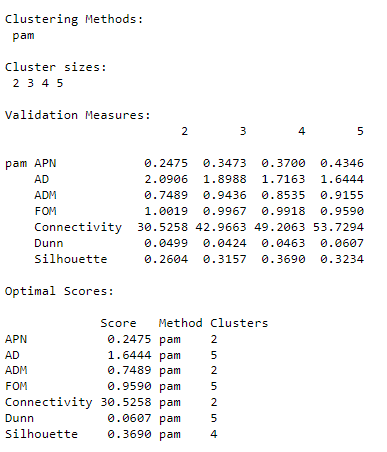


*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El método seleccionado es de 5 clústeres que en su mayoría discriminan por el Componente Principal 1, para grupos como el 5,2 y 4 parece ser importante la dimensión 2 también, para su discriminación entre sí.

## Método k mediodes

*Ilustración 16. Selección de numero de clústeres por K-mediodes*

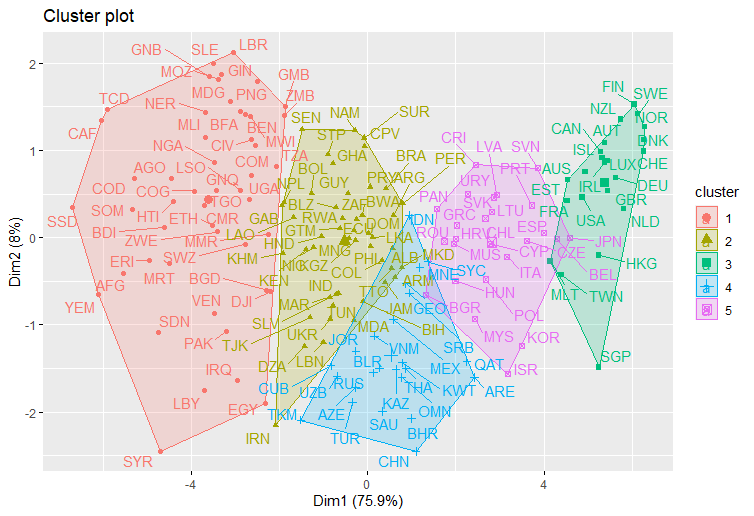


*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El número de clústeres óptimos por la prueba aparente es de 5 clústeres, en parte por los resultados en parte para poder ser comparado más directamente con el primer método k-mediodes.

Adentrar un poco en el metodo de pam

*Ilustración 17. Grafico agrupaciones por K-Medianas*

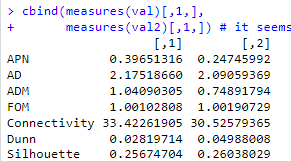


*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

El método agrupa de forma parecida a k- Medias, teniendo una sobrexposición aparente entre clústeres, el grupo 4 y el grupo 5 parece tener las diferencias más aparentes en tamaño frente a los otros grupos generados en K-Medianas.

*Ilustración 18. Grafico agrupaciones por K-Medianas*

## Selección del mejor modelo

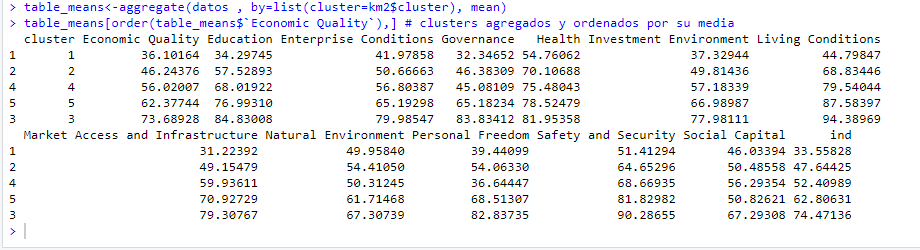


*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Por las métricas evaluadas podemos decidir usar el método 2 el método por K-Mediodes ya que en la mayoría de métricas es superior.

## Caracterización de los clústeres

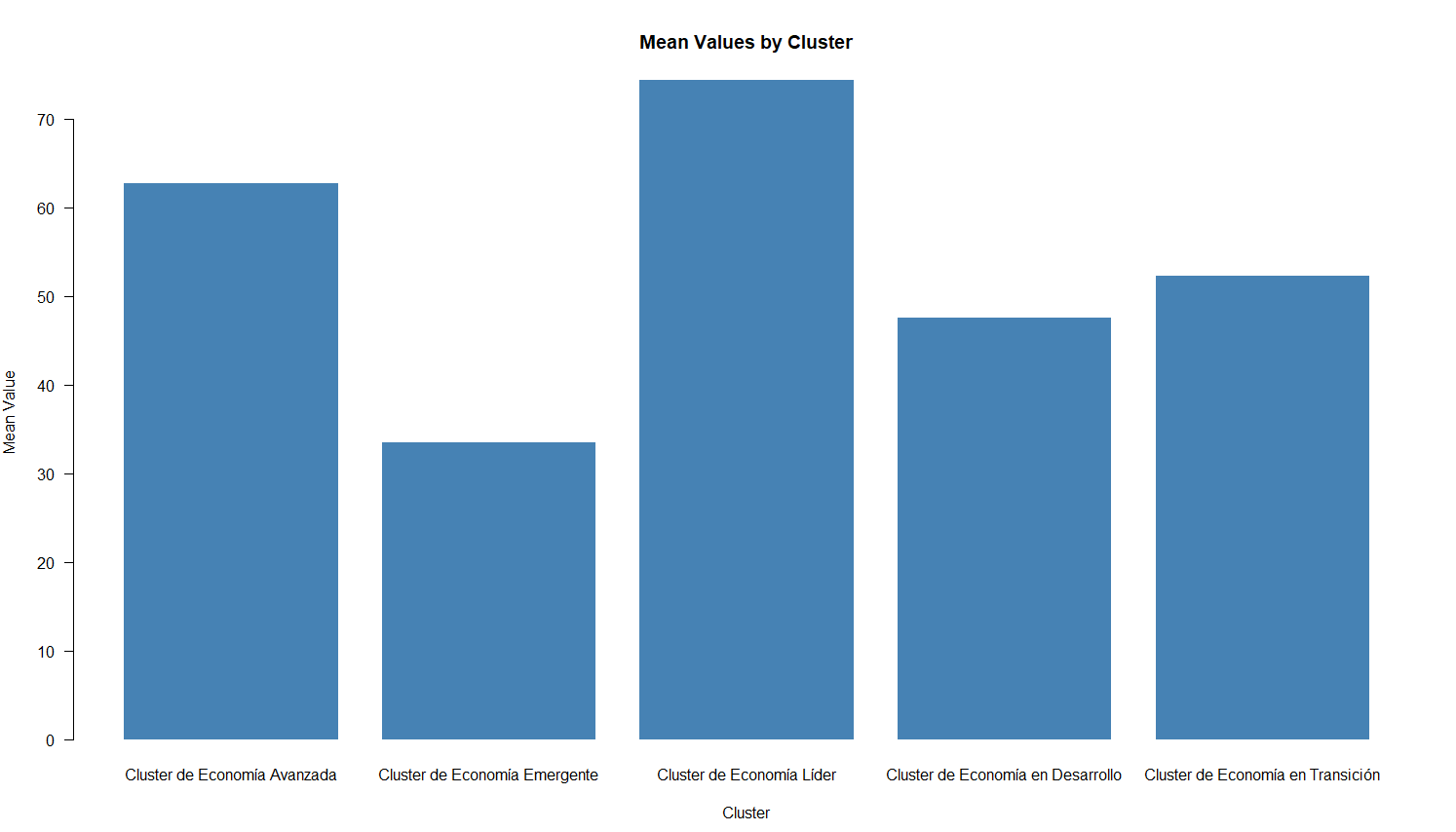
*Ilustración 19. Media Por Clúster Por Media Ordenadas, De menor a Mayor.*



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

Hay un orden jerárquico de los clústeres para cada una de las variables de deseamos presentar en el indicador por los clústeres, por esto podemos darle nombres, para otorgarles un mayor sentido.

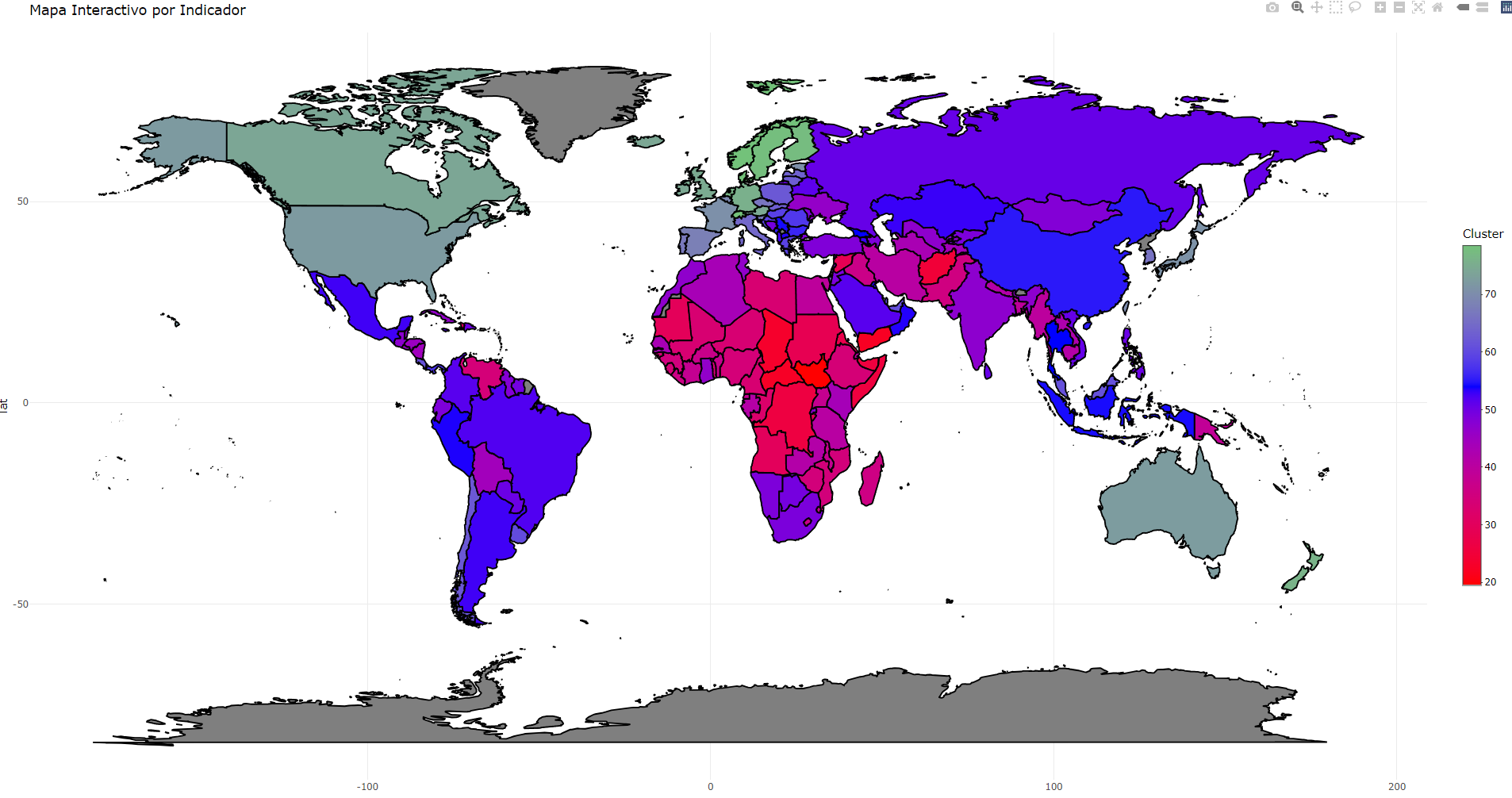
*Ilustración 19. Gráfico de barras media de indicador por clústeres.*



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

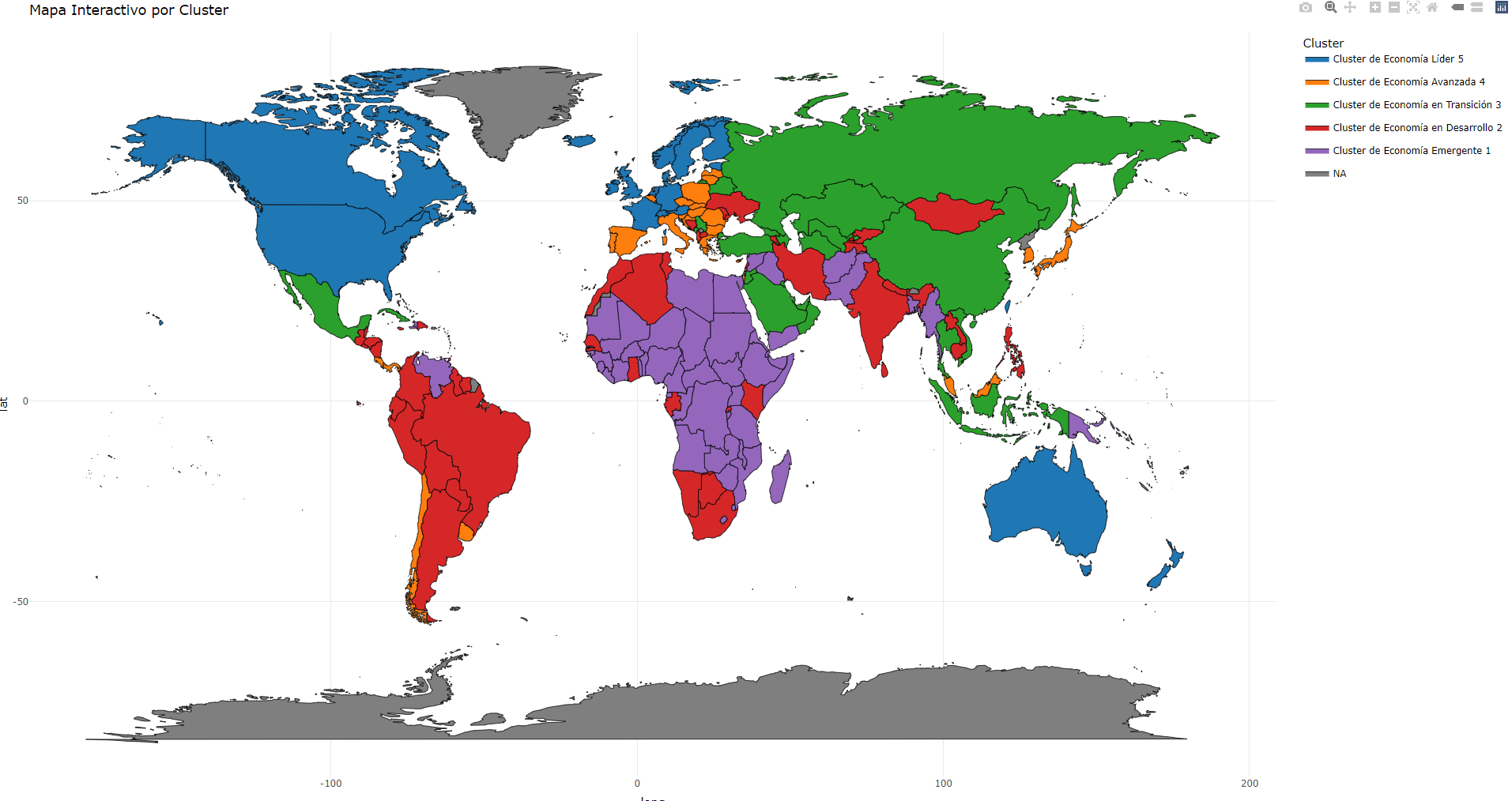
## Índice de Prosperidad en Mapas

*Ilustración 20. Mapa mundial Países por Indicador*



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

*Ilustración 21. Mapa mundial Países por Indicador*



*Fuente: Base de datos* *Legatum Prosperity Index 2019. Diseño proprio*

En los mapas es evidente que los niveles económicos están presentes en diferentes puntos geográficos

# Conclusiones

Conforme los resultados

# Referencias

Ayala, N., & Parra, C. (2023). Algoritmos de agrupamiento jerárquico para el control de la susceptibilidad antibiótica. *Revista Chilena de Enfermedades Respiratorias*, 120-121.

Beattie, J., & Esmonde-White, F. (2021). Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. *Applied Spectroscopy. 2021;75(4)*, 361-375.

Bettolli, M., Rivera, J., & Penalba, O. (2019). REGIONALIZACIÓN DE LOS DÍAS SECOS EN ARGENTINA. UN ENFOQUE METODOLÓGICO. *Metereologica Vol 35 (2)*, 67-80.

Fernández-Chuaire, L., Rangel, L., Varela, M., Pino, J., Del Pozo, J., & Lim, N. (2022). Analysis of Main Components, an Effective Tool in Agricultural Technical Sciences. *Revista Ciencias Técnicas Agropecuarias, 31(1)*.

Phélan, M. (2018). Revisión y comparación metodológica de cinco Índices Alternativos de Desarrollo. Resultados para países de América Latina. *Espacio Abierto*, 21-45.

(2020). *SELA.* Caracas: Sistema Económico Latinoamericado y del Caribe.

# Anexo