

UNIVERSIDAD DE GRANADA

MÁSTER EN ESTADÍSTICA APLICADA

TRABAJO FIN DE MÁSTER

**Estudio de la evolución mundial del Covid-19 mediante
análisis de datos funcionales**

Presentado por:

D./D^a. LUIS MURILLO GONZÁLEZ

Tutor/a:

Ana María Aguilera del Pino

Manuel Escabias Machuca

Curso académico 2020/ 2021

Agradecimientos

Me gustaría agradecer en primera instancia a mi tutores Manuel Escavias Machuca y Ana María Aguilera del Pino, quienes de forma desinteresada y vocacional me han dedicado tiempo fuera de las horas lectivas de este trabajo para poder compatibilizar mi vida profesional junto con la académica y obtener los objetivos que buscaba.

Del mismo modo me gustaría agradecer a la Universidad de Granada por ofrecerme el material académico básico para obtener los conocimientos necesarios y poder así aplicar el proceso de investigación.

Finalmente, me gustaría agradecer a todos mis compañeros de trabajo y a mi familia por su apoyo y preocupación. De manera particular a mi madre, que siempre estuvieron ahí para darme ánimos y energías para continuar.

Gracias a todos.

Índice

1. Abstrac	4
1.1 Introducción	5
1.2 Objetivos	6
2. Datos Funcionales	7
2.1 Definición	7
2.2 Objetivo de los datos funcionales	8
2.3 Marco teórico	9
2.4 Representación en bases de funciones	10
2.4.1 Tipos de bases	12
2.5 Estimación de los coeficientes básicos	14
2.5.1 Aproximación por mínimos cuadrados	14
2.5.2 Interpolación B-splines	16
2.5.3 Proyección Ortogonal	17
2.6 Media Funcional y Varianza	19
3. Análisis Funcional de componentes principales	20
3.1 AFPC	20
3.2 Análisis funcional de componentes principales: curvas de componentes principales	21
3.3 Análisis de componentes principales funcionales: expansión FPC	21
4. Analisis clúster	22
4.1 Analisis clúster multivariante	22
4.1.2 Clasificación cluster	23
4.1.2.1 Metodos Jerárquicos	23
5. Metodología	29
5.1 Data set	29
6. Análisis práctico	35
6.1 Descripción del proceso	35
7. Agrupación clúster	44
7.1 Clustering	45
7.2 Análisis curva media	54
8. Conclusiones	60
9. Referencias	62
10. Anexo	64

1. Abstract

This dissertation analyses different variables caused by SARS-CoV-2 (COVID-19) in 106 countries through functional data analysis (FDA).

The countries are grouped around two reasons, one natural and the other through the cluster methodology. The natural group is each country within its own continent, while the cluster methodology follows the k-means algorithm.

Different analysis methods from classical statistics will be used to analyse mean curves and variability. The FDA process and cluster analysis have been developed within statistical computing environments to simplify and obtain as much information as possible in this study.

Keyword

Functional Data Analysis, Mean Curve, Clustering, K-means, Analysis Component Principal, Scores, Coefficients, basis.

Resumen

Este trabajo aproxima el análisis de datos funcionales (FDA) al estudio de diferentes variables originadas por el SARS-CoV-2 (COVID-19), para un total de 106 países.

Tras una amplia síntesis sobre la base teórica del análisis funcional y la puesta en marcha de nuestro análisis, los países se agrupan en torno a dos criterios, uno natural y otro a través de la metodología clúster. La agrupación natural atiende a cada individuo respecto de su continente, mientras la metodología clúster ha seguido los pasos del algoritmo k-medias.

A partir de aquí, el trabajo utiliza diferentes métodos de análisis importados de la estadística clásica con el fin de ver la evolución y conocer más acerca del comportamiento de cada uno de los diferentes grupos de países. Estos métodos se han basado principalmente en cálculos de las curvas medias y análisis de variabilidad.

Tanto el proceso FDA como el análisis clúster se han desarrollado dentro de los entornos de computación estadística, con el fin de simplificar y obtener el mayor rendimiento posible al estudio.

Palabras Clave

Análisis de datos funcionales, Curva media, Clustering, K-medias, Análisis de Componentes Principales

1.1 Introducción

Nada ha agitado el mundo tras la Segunda Guerra Mundial como lo ha hecho el SARS-CoV-2 (COVID-19). Para poder afrontar este enorme reto, la humanidad se ha servido de todas sus capacidades con el fin de atajar este hecho histórico. Una de estas capacidades ha sido el análisis de datos, el cual ha jugado un papel esencial para contener bajo control la evolución de la pandemia.

Dentro del análisis masivo de datos podemos encontrar numerosas bibliografías las cuales nos ayudan a desarrollar y aplicar distintas metodologías de análisis, siempre con el objetivo de obtener la mayor información posible.

Una de estas metodologías es el Análisis de Datos Funcionales (FDA con sus siglas en inglés) y en ella es posible analizar las observaciones que se caracterizan por poder observarse como curvas naturales.

Si por ejemplo observamos la evolución acumulada del total de casos o muertes desde el inicio de la pandemia hasta nuestros días, podremos definir estos datos discretos como curvas. A partir, se pueden obtener unas determinadas funciones para poder realizar el análisis que deseemos sobre ellas.

Una de las ventajas que ofrece el FDA es que podemos complementarlo con distintos métodos que normalmente se aplican dentro de la metodología clásica. Un ejemplo de esto sería estudiar medidas centrales o de variabilidad sobre dichas funciones.

Otro método clásico que podría complementar al FDA sería el análisis clúster. Esta metodología de análisis nos permite entender cómo se agrupan y comportan unos determinados individuos (en nuestro caso, países). Unos de los campos donde popularmente se utiliza esta técnica es la medicina, concretamente dentro de la epidemiología Salvatore (2015).

Desde los trabajos de Ramsay y Silverman (1997) el Análisis de Datos Funcionales se ha ido convirtiendo en una herramienta cada vez más valorada dentro de la ciencia de datos, la cual ha ido abriendo nuevas oportunidades de estudio e investigación.

Hoy en día, la versatilidad de este tipo de análisis es muy amplia, a lo largo de los capítulos pondremos los conceptos del FDA con algunos ejemplos.

1.2 Objetivos

Uno de los principales objetivos que se han buscado dentro de este trabajo ha sido poner en valor y en práctica, aquellos conocimientos que se han ido adquiriendo a lo largo del curso académico.

A pesar de que el tema central gira en torno al Funcional Data Analysis (FDA) y la asignatura de *Análisis de Datos Funcionales*, la consecución de los objetivos no hubiera sido posible sin cursar otras asignaturas como *Entornos de Computación Estadística* o *Modelos de Respuesta Discreta*.

La propuesta de este trabajo se ha basado principalmente en abordar las principales variables discretas recogidas en el apartado COVID-19 de la base de datos *Our Data in World* desde la perspectiva funcional descrita.

A partir de esto, hemos querido abordar dos objetivos principales los cuales a su vez presentaban una serie de retos. Estos objetivos han sido diseñados en función del marco teórico disponible y de los resultados que se han ido obteniendo a lo largo del trabajo, los cuales definimos de la siguiente manera:

Por un lado, tenemos un primer objetivo de agrupar, documentar y poner en valor parte de la teoría disponible que envuelve los conceptos y objetivos de datos funcionales, la estimación sobre los coeficientes básicos, la realización de un análisis de componentes principales o ACP, la comprensión y diferenciación sobre los distintos tipos de clúster, así como su posterior elección óptima.

Y por otro lado, realizar un profundo estudio funcional junto a una metodología clásica para poder entender cómo se comportan y agrupan los países. Además de esto, el trabajo busca ir más allá del análisis clúster y también busca realizar una comparativa de los resultados a través de la utilización de distintos argumentos para el algoritmo k-medias. Los distintos argumentos utilizados son los siguientes: datos discretos, coeficientes básicos y scores.

Otros de los retos que este trabajo asume es la simplificación y la selección de la información adecuada, trabajar con este tipo de bases de datos requiere de una profunda comprensión sobre el proceso computacional.

La consecución de estos objetivos se ha llevado a cabo en base a un procedimiento científico, comprendiendo cada uno de los pasos que se realiza dentro de la metodología funcional y clúster, tanto dentro del marco teórico como computacional.

2. Datos Funcionales

2.1 Definición

Los primeros autores en definir el concepto de datos funcionales fueron Ramsay y Dalzell (1991) y Ramsay & Siverman (1997). Los datos funcionales se dan cuando los valores de una variable pueden observarse de una manera natural, como una curva o una función suave.

De acuerdo con la definición formal que recogen Ferraty y Vieu (2006) los datos funcionales se pueden describir de la siguiente manera:

Definición Nº.1 *Como una variable aleatoria X que toma valores en un espacio funcional \mathcal{E} (Espacio normado o seminormado completo).*

Definición Nº.2 *Un conjunto de datos funcionales $\{X_1, \dots, X_n\}$ es la observación de n variables funcionales X_1, \dots, X_n idénticamente distribuidas.*

Al comienzo de su historia, los datos funcionales fueron planteados como observaciones independientes en un proceso estocástico en un espacio Hilbert $L^2(T)$. Las variables se identificaban como curvas para un intervalo $T = [0, T]$, las cuales tienen la característica de poder ser suavizadas.

A través de estos condicionantes, el Análisis de Datos Funcionales se ha visto envuelto un desarrollo importante, permitiendo así convertirse en una herramienta metodológica realmente útil en diferentes campos de la investigación, como podría ser dentro de las ciencias de la salud Salvatore (2015), la neurociencia o en la economía, la evolución de los mercados.

Son numerosas las metodologías que se utilizan dentro del FDA, al poderse combinar con análisis clásico se puede exportar de aquí muchas de ellas como por ejemplo los modelos de regresión, análisis multivariantes (Análisis de Componentes Principales), clustering, estadística espacial, o estudio de series temporales, entre otros.

Pero antes de nada, se debe entender que partimos de datos discretos y que debemos realizar una conversión a datos funcionales. Si los datos observados no presentan errores simplemente se realiza una interpolación, pero si en los datos discretos se observan alguna serie de errores de cualquier tipo, entonces se debe aplicar un suavizado.

2.2 Objetivo de los datos funcionales

A pesar de ser una metodología relativamente nueva Ramsay y Silverman (1997), el análisis de datos funcionales se enfrenta a los mismos problemas que cualquier metodología. De forma general, algunos de estos objetivos se recoge en este libro los cuales podemos agrupar de forma resumida en los siguientes puntos:

- Representar los datos de manera que ayuden a obtener más detalles e información.
- Mostrar los datos para resaltar varias características que sean de interés.
- Estudiar fuentes importantes de patrones y variaciones entre los datos.
- Explicar la variación en un resultado o variable dependiente, usando información de entrada o variable independiente.
- Comparar dos o más conjuntos de datos con respecto a ciertos tipos de variación, donde dos conjuntos de datos pueden contener diferentes conjuntos de réplicas de las mismas funciones, o diferentes funciones para un conjunto común de réplicas.

Cada uno de estos objetivos anteriormente mencionados cuenta con un proceso propio. Con tal de afrontar esta problemática, se puede desarrollar una estrategia de tipo exploratorio, confirmatorio o predictivo (este trabajo solo se centra en la parte exploratoria).

La metodología exploratoria, estudia con un carácter general la información que se puede obtener sobre unos datos. Tenemos entonces que esta técnica permite obtener aspectos nuevos e interesantes sobre las observaciones, permitiendo así obtener nuevas posibilidades de estudio.

Por otra parte, la metodología confirmatoria abarca más la inferencia estadística y está sujeta a reglas más estrictas estadísticamente hablando. Este tipo de normas vienen dadas por ejemplo por el tipo de estructura que presentan los datos, el tipo de hipótesis que se pretende hacer y el mismo enunciado.

La principal diferencia entre la metodología exploratoria y la metodología confirmatoria es que esta segunda se apoya en la teoría de la probabilidad.

Por su parte, los estudios predictivos son menos populares y se centran en usar dos datos que están disponibles para afirmar o predecir la información no observada y que se proyectará en un futuro.

2.3 Marco teórico

Además de Ramsay y Silverman, han sido numerosos los autores que han continuado y desarrollando la teoría en la que se envuelve el análisis de datos funcionales. Los autores Jacques y Preda (2014) publican un artículo donde recogen muy bien la síntesis metodológica en la que se basa el análisis de datos funcionales.

Si partimos de una variable funcional aleatoria X , esta es una variable aleatoria con valores en un espacio infinito-dimensional, entonces los datos funcionales que representan un conjunto de observaciones será $\{X_1, \dots, X_n\}$ de X .

El modelo subyacente de X_i 's es generalmente un ejemplo de muestras de variables aleatorias extraídas de la misma distribución que X .

Un modelo aceptado para este tipo de datos, es considerarlos como un proceso estocástico donde $X = \{X_t\}_{t \in \mathcal{T}}$ que toma valores en un espacio de Hilbert L^2 de funciones definidas en algún conjunto \mathcal{T} . Generalmente, \mathcal{T} representa un intervalo de tiempo, de longitudes de onda o cualquier otro subconjunto continuo de \mathbb{R} .

La principal dificultad cuando tratamos datos funcionales consiste en el hecho de que se supone que las observaciones pertenecen a un espacio infinito-dimensional, mientras que en la práctica solo se han muestreado las curvas observadas en un conjunto finito de puntos temporales. De hecho, es habitual que solo tengamos observaciones discretas X_{ij} para cada ruta de muestra $X_i(t)$ en un conjunto finito de nudos $\{t_{ij} : j = 1, \dots, m_i\}$.

Debido a esto, el primer paso de la FDA suele ser la reconstrucción de la forma funcional de los datos a partir de observaciones discretas. La solución más común a este problema es considerar que las trayectorias pertenecen a un espacio de dimensión finita generado por alguna base de funciones.

2.4 Representación en bases de funciones.

Una base de funciones es un conjunto de funciones conocidas ϕ_k que son matemáticamente independientes entre sí y que tienen la ventaja de poder describir de forma particular, a partir de la suma ponderada o combinación lineal de un determinado número de K funciones (Curtis, 2014).

A modo de ejemplo tenemos que, el sistema de funciones de base más familiar es la colección de monomios la cual se utiliza para construir unas series de potencias, como podemos ver a continuación,

$$1, t, t^2, t^3, t^4, \dots$$

Esto también se puede representar como un proceso de las bases en un función x para una expansión lineal, y que podemos observar de la siguiente manera:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t)$$

En términos de K funciones de base conocidas ϕ_k

Donde c indica el vector de longitud K de los coeficientes c_k y ϕ como el vector funcional de cuyos elementos funcionales son las bases ϕ_k . Aquí también podemos expresar la expresión anterior de la siguiente forma:

$$x = c' \phi = \phi' c$$

La expansión de las bases nos permiten representar la dimensión infinita perteneciente a las funciones sujetas a un espacio de vectores finitos como c . La dimensión de dicha expansión la llamamos K .

Se tiene una interpolación real cuando $K = n$, en este sentido, tenemos que elegir los coeficientes c_k para obtener $x(t_j) = y_j$ de cada j .

Entonces tenemos que la manera en que los valores de y_j se suavizan en vez de interpolar, estará determinado por el K número de funciones básicas. Siendo K el mismo parámetro que elegimos de acuerdo a las características con la que contamos los datos.

El objetivo de las bases es que deben recoger las mismas características para aquellas funciones que están estimando. Cuanto menor sea K y mejor se observe las bases de las funciones, más grados de libertad tendremos para probar nuestras hipótesis (en el supuesto de trabajar este área).

De la misma manera, existen bases que están listas para trabajar, como los polinomios, series de Fourier o B-spline. Algunas de estas pueden ser desaconsejadas para algunas aplicaciones, pero no existe aún un marco universal.

La elección de la base es realmente importante y en el caso de que necesitemos estimar su derivada tendríamos que,

$$D\hat{x}(t) = \sum_k^K \hat{c}_k D\phi_k(t) = \hat{c}' D\phi(t).$$

Aunque no es el objetivo de este trabajo, me gustaría decir que para las bases que trabajan a la perfección nos permiten estimar funciones resultantes que pueden dar resultados de derivadas bastante débiles. Es debido a que una representación precisa de las observaciones, puede obligar a \hat{x} a tener oscilaciones pequeñas pero de alta frecuencia y que tienen terribles consecuencias para sus derivadas. El capítulo 21 de Ramsay y Silverman (1997) trata más en profundidad este apartado pero por ahora, discutimos algunas bases populares.

Continuando con nuestra bibliografía, tomamos nota de la síntesis que hace (Curtis & Barcenas Sánchez, 2014) en su publicación *“Estudio de Perfiles de Olas vía Análisis de Datos Funcionales”*.

Para convertir los datos discretos en funciones, tenemos que abordar dos etapas:

- La primera es elegir una base adecuada de acuerdo a la naturaleza de los datos.
- La segunda es concluir los coeficientes que definen a la función como una combinación lineal de esas funciones base.

Destacar en la primera etapa que existen distintos tipos de bases y se consideran los siguientes criterios para su elección: Que la base represente bien a las funciones deseadas y una base de Fourier para datos periódicos o una base B-spline para datos no periódicos.

Que se logre una buena estimación con un número K de funciones base relativamente pequeño y que las derivadas proporcionen información relevante al análisis.

A continuación, podremos obtener más información sobre los principales tipos de bases que existen y con qué características cuentan.

2.4.1 Tipos de bases

Fourier

Una de las funciones más populares y empleadas es la función de Fourier, donde las bases son senos y cosenos de frecuencia creciente:

$$\begin{aligned}\phi_1(t) &= 1, \\ \phi_2(t) &= \text{sen}(\omega t), \\ \phi_3(t) &= \text{cos}(\omega t), \\ \phi_4(t) &= \text{sen}(2\omega t), \\ \phi_5(t) &= \text{cos}(2\omega t), \dots\end{aligned}$$

Siendo $\omega = 2\pi/P$ donde P resulta ser el periodo en el que oscila.

Este tipo de bases es utilizada y definida cuando los datos que tenemos disponibles son periódicos o cercanamente periódicos. Si los valores de t_j son iguales en el espacio T y el periodo es igual a la longitud del intervalo T entonces la base es ortogonal, en el sentido que obtiene el producto de la matriz diagonal $\phi'\phi$ y que puede ser calculada igual para la identidad si dividimos las funciones bases por las constantes adecuadas \sqrt{n} para $j = 0$ y $\sqrt{n/2}$ para todo j .

La transformada rápida de Fourier (FFT en sus siglas en inglés) permite encontrar todos los coeficientes de manera extremadamente eficiente, cuando n es una potencia de 2 y los argumentos están igualmente espaciados y para este caso, podemos encontrar tanto los coeficientes c_k como todos los n valores suaves en $x(t_j)$ en operaciones $O(n \log n)$. Esta es una de las características que ha convertido a la serie de Fourier en la base tradicional de elección para series de tiempo prolongado, pero las técnicas más nuevas, como B-splines y wavelets, pueden igualar e incluso superar esta eficiencia.

Base B-Splines

Esta metodología ha sido puesta en marcha por De Boor (1977) donde tiene en cuenta una serie de bases tipo $\phi_k(t)$ para cumplir con los siguientes puntos:

- Cada una de las $\phi_k(t)$ es una función spline definida por un orden m y una sucesión de nodos τ .
- Cualquiera de las diferentes combinaciones lineales que hagamos con esta serie de funciones base $\phi_k(t)$ será un spline.
- Dentro de cualquiera función spline que haya sido definida por su orden m con nodos τ , puede expresarse como una combinación lineal de este tipo de funciones.

La piedra angular de las funciones B-spline es que son funciones positivas. Esto quiere decir que una B-spline de cierto orden m será positivo en no más de m intervalos vecinos, por lo que será más sencillo calcular los coeficientes.

La base de B-splines se usa principalmente en datos no periódicos cuya función es suave y sus coeficientes se pueden calcular rápidamente con una aproximación muy buena, donde además puede lograrse con un número K de funciones pequeño.

Otras bases

- **Wavelets:** Se utilizan como funciones básicas cuando representamos los datos funcionales. Estas bases son obtenidas mediante dilataciones y traslaciones de una función ϕ .

$$\phi_{kj}(t) = 2^{k/2}\phi(2^k t - k),$$

Siendo j y k número enteros.

Me gustaría destacar que para este tipo de funciones no tiene en cuenta que los datos sean periódicos.

- **Potenciales:** Son bases las cuales están compuestas por monomios y las cuales se utilizan para construir una serie de potencias de las que obtenemos la siguiente manera:

$$1, t, t^2, t^3, \dots, t^k. \quad \text{siendo } k = 1, 2, \dots,$$

2.5 Estimación de los coeficientes básicos.

Una vez seleccionada la base oportuna para calcular nuestro objeto funcional, las observaciones se pueden presentar de dos maneras, con y sin error.

Si tenemos en cuenta que el predictor observado es con error, este tendría la siguiente forma:

$$X_{ik} = X_i(t_{ik}) + \varepsilon_k \quad k = 0, \dots, m_i,$$

Suponemos que los errores tienen media cero, varianza constante, y la covarianza cero para los valores que presenten argumentos distintos.

De la misma manera, si las observaciones se observarían sin algún tipo de error el método utilizado sería alguna interpolación como por ejemplo s-plines cúbicos González (2016), donde tenemos la siguiente forma:

$$X_{ik} = X_i(t_{ik}) \quad k = 0, \dots, m_i,$$

Cada uno de estos métodos nos ayuda a la obtención de la forma funcional de las trayectorias muestrales mediante la aproximación de los coeficientes básicos c_k

$$x(t) = \sum_{k=1}^K c_k \phi_k(t)$$

A partir de las observaciones discretas.

2.5.1 Aproximación por mínimos cuadrados.

La aproximación por mínimos cuadrados tiene por objetivo encontrar una función continua en base al mínimo error cuadrático (Patricia, 2013).

Si partimos de que tenemos observaciones para una curva, tenemos que:

$$y_i = x_{t_i} + \varepsilon$$

Y deseamos aproximar a,

$$x(t) \approx \sum_{k=1}^K c_k \phi_k(t)$$

Entonces sabemos que lo que se busca es minimizar la sumas de estos errores al cuadrado, entonces observamos lo siguiente:

$$SSE = \sum_{i=1}^n (y_i - x(t_i))^2 = \sum_{i=1}^n (y_i - \phi(t_i)c)^2$$

Para la realización del ajuste de curvas a los datos discretos X_{ik} , utilizamos la siguiente expresión con error

$$X_{ik} = X_i(t_{ik}) + \varepsilon_k \quad k = 0, \dots, m_i,$$

Aquí los errores son independientes y distribuidos de manera idéntica. Si lo expresamos en términos de funciones,

$$x_i(t_k) = \sum_{p=1}^P a_{ip} \phi_p(t_k) \quad k = 0, 1, \dots, K \quad i = 1, 2, \dots, N$$

Donde la forma de la matriz nos quedaría como:

$$x_i = \Phi_i a_i$$

Donde $a_i = (a_{i1}, \dots, a_{ip})'$ y $\Phi_i = (\Phi(t_k))_{K \times P}$, mientras que x_i es el vector que muestra los vectores de los datos discretos $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})'$, en el que los coeficientes de a_{ip} se aproxima mediante cuadrados mínimos, cuya matriz quedaría la siguiente:

$$ECM(x_i | a_i) = (x_i - \Phi_i a_i)'(x_i - \Phi_i a_i)$$

Si derivamos respecto de a_i , tenemos que:

$$2\Phi_i' a_i - 2\Phi_i' x_i = 0$$

Para el siguiente paso, tenemos que el estimador de mínimos cuadrados del vector de coeficientes básicos \hat{a}_i vendrá expresado por:

$$\hat{a}_i = (\Phi_i' \Phi_i)^{-1} \Phi_i' x_i$$

Y finalmente se calcula \hat{x}_i con los valores aproximados, donde la curva está definida por,

$$\hat{x}_i = \Phi_i \hat{a}_i = \Phi_i (\Phi_i' \Phi_i)^{-1} \Phi_i' x_i$$

De modo que cada curva se aproxima por,

$$\hat{x}_i(t) = \widehat{a}_i' \Phi(t)$$

2.5.2 Interpolación B-splines

Este tipo de interpolación son las más utilizadas cuando los datos no son cíclicos y cuando los datos han sido observados sin error, observamos la expresión anterior:

$$X_{ik} = X_i(t_{ik}) \quad k = 0, \dots, m_i,$$

Cuando realizamos algún tipo de interpolación, la curva ajustada cruza todos los puntos, mientras que la dimensión de la base es la misma que para el número de observaciones de cada curva más dos. Entonces este tipo de funciones splines cúbica interpola estas trayectorias $X_i(t)$ sobre los nodos de observación. Podemos obtener en términos de la base B-Splines cúbicos $\{B_{-1}(t), B_0(t), B_1(t), B_2(t), B_3(t), \dots, B_k(t), B_{k+1}(t)\}$ la expresión:

$$x_i(t) = \sum_{j=-1}^{k+1} a_{ij} B_j(t),$$

De donde podemos obtener los coeficiente básicos resolviendo la siguiente expresión,

$$x_{ik} = x_i(t_k) = \sum_{j=-1}^{k+1} a_{ij} B_j(t_k), \quad \text{donde } k = 0, 1, \dots, m$$

Para esta ecuación tendremos una serie de condiciones determinadas, entonces la interpolación que hagamos recibirá un nombre diferente dependiendo de dichas características.

- **Spline cúbico natural:** Establece que la derivada segunda del spline en la interpolación sea igual a 0 para los nodos límites del intervalo $[0, T]$,

$$x''_i(t_0) = 0$$

$$x''_i(t_k) = 0$$

Así que tenemos en cuenta las curvas muestrales para nuestro método $x_1(t), x_2(t), \dots, x_n(t)$ observado en los nodos t_0, t_1, \dots, t_m junto con la siguiente expresión anteriormente citada

$$x_{ik} = x_i(t_k) = \sum_{j=-1}^{k+1} a_{ij} B_j(t_k), \quad \text{donde } k = 0, 1, \dots, m$$

Y las condiciones dadas

$$x''_i(t_0) = 0$$

$$x''_i(t_k) = 0$$

Tenemos que $X_I = AB'$ donde A es la matriz de las filas pertenecientes a los coeficientes básicos de cada una de las trayectorias muestrales y que están interpoladas de la siguiente manera $x_i(t), X_I = (0|X|0)$ con $0 = (0, \dots, 0)'$ y $X_{nx(m+1)} = x_i(t_k)$, siendo $i = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, m$ y con B expresada por,

$$B = \begin{bmatrix} B''_{-1}(t_0) & B''_0(t_0) & \dots & B''_{m+1}(t_0) \\ B_{-1}(t_0) & B_0(t_0) & \dots & B_{m+1}(t_0) \\ \vdots & \vdots & \vdots & \vdots \\ B_{-1}(t_m) & B_0(t_m) & \dots & B_{m+1}(t_m) \\ B''_{-1}(t_m) & B''_0(t_m) & \dots & B''_{m+1}(t_m) \end{bmatrix}$$

Donde $B''_i(t)$ es la segunda deriva del i -ésimo B-spline en el tiempo t . Si resolvemos la ecuación anterior, podemos observar la matriz A con el resultado de la aproximación de los coeficientes básicos de todas las trayectorias muestrales, expresadas por:

$$\hat{A} = X(B')^{-1}$$

- **Spline cúbico periódico:** Lo denominaremos de esta manera si las características iniciales son las siguientes:

$$\begin{aligned} x'_i(t_1) &= x'_i(t_k) \\ x''_i(t_1) &= x''_i(t_k) \end{aligned}$$

2.5.3 Proyección Ortogonal

Una de las alternativas para suavizar y estimar los coeficientes básicos es realizar una proyección ortogonal. Esta técnica consiste en acercarnos a cada una de las curvas muestrales a través de su proyección en el subespacio de $L^2(T)$ donde es resuelto por una base ortonormal $\{\phi_1(t), \phi_2(t), \dots, \phi_p(t)\}$.

La proyección de cada curva muestral $x_i(t)$ en un subespacio p -dimensional vendrá dado por la siguiente expresión:

$$P_p(x_i(t)) = \sum_{j=1}^p \left[\int_T x_i(t) \phi_j(t) dt \right] \phi_j(t) \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$$

La dificultad vendrá dada por la estimación de los coeficiente base, que se expresa de esta manera,

$$a_{ij} = \int_T x_i(t) \phi_j(t) dt$$

Desde las observaciones para el tiempo discreto de las trayectorias muestrales. Una manera de resolver esto, sería aplicar métodos de cuadratura compuesta, como el método Romberg.

Si deseamos obtener más información de cómo aplicar este método y ver cómo dichos valores pueden ser aproximados véase, (Aguilera, Ocana, & Valderrama, 1995).

2.6 Media Funcional y Varianza

La gran mayoría de las técnicas empleadas dentro de la estadística clásica puede ser exportadas a otros campos, también dentro del Análisis de Datos Funcionales.

Una de las medidas más utilizadas es la media, la varianza y la covarianza. Utilizamos la media como medida de tendencia central y la varianza y covarianza como medidas de variación individual y conjunta.

Tenemos que la función media es,

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p a_{ij} \phi_j(t) = \sum_{j=1}^p \bar{a}_j \phi_j(t), \bar{a}_j = \frac{1}{n} \sum_{i=1}^n a_{ij}$$

A través de la función media podemos calcular la varianza que viene expresada por la siguiente función,

$$var_x(t) = (N - 1)^{-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2,$$

y la función de desviación estándar es la raíz cuadrada de la varianza función.

La covarianza por su parte la dependencia de registros a través de diferentes valores de argumento, y se calcula para todo s y t .

A partir de un conjunto de curvas $x_1(t), x_2(t), \dots, x_n(t)$ con curva media $\bar{x}(t)$ se define la superficie de covarianza

$$C(s, t) = \frac{1}{n - 1} \sum_{i=1}^n (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$$

La función de correlación asociada es,

$$r(s, t) = \frac{C(s, t)}{\sqrt{C(s, s)C(t, t)}}$$

En definitiva estas son funciones análogas para el análisis de datos multivariante. En el capítulo 2.3 *Summary statistics for functional data del libro de Ramsay & Siverman (1997)* podemos obtener más información sobre este tipo de medidas centrales y de variabilidad.

3. Análisis Funcional de componentes principales

3.1 AFPC

El Análisis Funcional de Componentes Principales (AFPC) es una extensión del Análisis de Componentes Principales (ACP) en donde los componentes principales son vectores Gonzalez (2016). Una de las principales de estas aportaciones al desarrollo de este concepto ha sido (Hotelling, 1936), donde define los componentes principales como aquellas combinaciones lineales de las variables originales para obtener un nuevo conjunto y que lo denominan componentes principales.

Si el ACP clásico busca convertir un grupo de variables en un grupo de menor dimensión llamadas componentes principales. El AFPC tiene también el mismo objetivo de reducir la dimensión de las variables del proceso.

Siendo $x_1(t), \dots, x_n(t)$, un conjunto de curvas con media

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

Y covarianza

$$C(s, t) = \frac{1}{n-1} \sum_{i=1}^n (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$$

El análisis funcional en componentes principales es definido como un vector cuyos elementos son obtenidos de una combinación lineal de curvas.

$$\xi_i = \int_T (x_i(t) - \bar{x}(t))f(t)dt, i = 1, 2, \dots, n$$

Donde maximiza la varianza de ξ_1, \dots, ξ_n y no correlaciona. Al imponer esta condición, los componentes principales funcionales son las soluciones de esta ecuación.

$$\int_T C(s, t)f(s)ds = \lambda f(t)$$

Siendo λ la la varianza de las componentes principales.

Cuando las curvas son expresadas en términos de la base de las funciones tenemos la expresión anterior,

$$x_i(t) = \sum_{j=1}^k a_{ij}\phi_j(t), i = 1, \dots, n$$

La ecuación anterior tiene p soluciones para p valores de λ que verifica $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Cada una de las λ_j está asociada a $f_j(t)$ funciones que definen los componentes principales y conocidos como funciones propias o curvas de componentes principales.

$$\xi_{ij} = \int_T (x_i(t) - \bar{x}(t))f_j(t)dt, \quad j = 1, 2, \dots, p, \quad i = 1, 2, \dots, n$$

Así que tenemos que cada componente funcional principal ξ_j es un vector de dimensión n . En la forma de la matriz, funcional de componentes principales son usualmente considerados en una matriz Γ de dimensión $n \times p$. Además de lo anterior, cada componente principal funcional acumula una proporción de variabilidad dada por,

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j}$$

La variabilidad total que acumula los componentes principales funcionales es igual a la variabilidad de todas las curvas.

3.2 Análisis funcional de componentes principales: curvas de componentes principales

Cuando las curvas se expresan en términos de funciones básicas, tenemos que

$$x_i(t) = \sum_{j=1}^p a_{ij} \phi_j(t), \quad i = 1, \dots, n$$

Las funciones propias está también expresadas en la misma base de función

$$f_j(t) = \sum_{k=1}^p F_{jk} \phi_k(t), \quad j = 1, \dots, n$$

3.3 Análisis de componentes principales funcionales: expansión FPC

Las curvas originales se pueden aproximar mediante el uso de un conjunto reducido de componentes principales funcionales

$$x_i(t) = \sum_{j=1}^{q < p} \xi_{ij} f_j(t), \quad i = 1, \dots, n$$

Para poder expresar las curvas de componentes principales $f_j(t)$ en términos de las funciones básicas, tenemos una aproximación de las curvas originales en términos de funciones base, es decir, conociendo sus coeficientes base tendríamos:

$$x_i(t) = \sum_{j=1}^{q < p} \xi_{ij} \sum_{k=1}^p F_{jk} \phi_k(t) =, \quad i = 1, \dots, n$$

Para más información acerca de estos apartados véase Hotelling (1936).

4. Analisis clúster

La mayoría de los métodos clásicos del análisis multivariante adaptado al análisis funcional pasan por:

- Aplicar métodos clásicos del análisis multivariante a las observaciones discretas de las curvas.
- Aplicar los métodos multivariantes a los coeficientes básicos de la representación básica de las curvas.
- Aplicar métodos multivariantes clásicos a los componentes principales funcionales.

En este trabajo vamos a utilizar estas tres aproximaciones para el análisis clúster, intentando detectar posibles diferencias entre unos y otros.

Antes de hacer esa aplicación, debemos ver en qué consiste el análisis clúster multivariante.

4.1 Analisis clúster multivariante.

El Análisis Cluster lo definimos como un proceso estadístico multivariante que parte un conjunto de observaciones, las cuales contienen una determinada información y que viene dada por las variables donde se intentan colocar de forma homogénea en los grupos que llamamos clusters.

El análisis cluster a diferencia de otros métodos parte de muy poca o ninguna información, por lo que se convierte en un reto seleccionar la estructura real de las categorías para su posterior ordenación.

Las distintas maneras en las que se puede clasificar m valores en k grupos es el número de la segunda especie de Stirling (Abramowitz y Stegun, 1968).

$$S_m^{(k)} = \frac{1}{k!} \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^m$$

Si aún no se conoce el número de grupos las posibilidades será la suma de número de Stirling; así, por ejemplo en el caso de,

$$\sum_{j=1}^m S_m^{(j)}$$

Pero este caso no puede ser funcional ya que nos da como resultado un número desorbitado, por ejemplo si tenemos 30 observaciones se tiene,

$$\sum_{j=1}^{30} S_{30}^{(j)}$$

Por lo que debemos buscar alguna otra alternativa.

4.1.2 Clasificación cluster.

Cualquier análisis cluster parte de un conjunto de datos donde cada matriz recoge las distintas variables y los individuos u observaciones que deseamos estudiar. El objetivo del análisis se traduce en clasificar las observaciones de dicha matriz.

De manera general podemos destacar dos grandes bloques que agrupan los distintos métodos cluster, estos son: Métodos Jerárquicos y Métodos no Jerárquicos.

4.1.2.1 Metodos Jerárquicos

Los llamados Métodos Jerárquicos tienen por objetivo agrupar clusters para formar uno nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o bien se maximice alguna medida de similitud.

Dentro de los Métodos Jerárquicos encontramos principalmente dos grupos, estos son: Métodos Jerárquicos Disociativos y Métodos Jerárquicos Aglomerativos.

Los Métodos Jerárquicos Disociativos nos resultan los más interesantes para el estudio ya que parten de una aglomeración que recoge todas las observaciones, es a partir de este conjunto donde se realizan continuas sucesiones mientras se forman los grupos. Se obtendrá una vez completado de este proceso tantos grupos como observaciones han sido tratadas.

Los Métodos Jerárquicos Aglomerados por su parte presentan el caso opuesto al anterior. Parte del estudio de tantos grupos como observaciones existen. A partir de las observaciones iniciales se van formando grupos hasta que todos los objetos están aglutinados en un mismo conjunto.

Actualmente también es interesante estudiar los Métodos Jerárquicos Aglomerados, ya que este tipo tiene la característica de que si se forman dos clusters en cierto nivel, estos quedarán automáticamente organizados para el resto de los niveles.

4.1.2.1.1 Métodos Jerárquicos Aglomerativos

Dentro de este apartado enumeramos algunas de las maneras y criterios que son útiles a la hora de formar clústeres. Decir que cada una de ellas son diferentes y que se deja de manera objetiva la elección de las mismas; ninguna es más eficiente que el resto.

- **Single Linkage:** Se encarga de tomar la distancia entre dos clústeres obteniendo la distancia entre los individuos más próximos o la similitud entre sus individuos más semejantes.
- **Complete Linkage:** Esta metodología nos anuncia que la similitud entre clústeres viene dada por la distancia más dispar entre sus elementos o lo que es lo mismo su máxima distancia entre sus elementos.
- **Métodos basados en el centroide y de la mediana:** Para esta metodología, la semejanza de los clústers se da en la semejanza de sus centroides. Y aquí podemos apreciar dos tipos, Método del Centroide Ponderado y Método del Centroide No Ponderado.

En el Método del Centroide Ponderado los tamaños de los clusters son considerados a la hora de efectuar cualquier tipo de cálculo, mientras que en el Método del Centroide No Ponderado el tamaño no lo está.

- **Método de Ward:** Es uno de los procedimientos jerárquicos más conocidos, en cada una de las partes se unen dos grupos para la que se observa un incremento menor en el valor total de la suma de los cuadrados de las diferencias, dentro de cada cluster va del individuo al centroide del cluster.

4.1.2.1.2 Métodos jerárquicos Disociativos.

Como se mencionó anteriormente, los métodos disociativos son una metodología opuesta a los métodos aglomerativos. Parte de un grupo el cual cuenta con todos los casos tratados, es a partir de dicho grupo cuando se van formando grupos cada vez menores. Al final de este proceso se obtendrá tantos grupos como casos han sido tratados.

En cuanto a la clasificación de este tipo de métodos (disociativos) decimos que la síntesis de los métodos aglomerativos puede mantenerse en lo que atañe a la forma de calcular la distancia entre los grupos, siendo lógico, al partir de un grupo único que hay que subdividir, y seguirá la estrategia de maximizar las distancias, o minimizar las similitudes, ya que buscamos ahora los individuos menos similares para separarlos del resto del conglomerado.

Esta clase de métodos son esencialmente de dos tipos:

1. **Monotéticos:** Son aquellos que dividen los datos sobre la base de un solo atributo y suelen emplearse cuando los datos son de tipo binario.
2. **Politéticos:** Estos son cuyas divisiones se basan en los valores tomados por todas las variables.

Se debe tener especial atención al hecho de que cuando un clúster determinado debe dejar de dividirse para proceder con la división de otro conglomerado distinto.

Dicha cuestión puede resolverse con la siguiente variante expuesta por MacNaughton-Smith en 1964 y que está concebida para aquellas medidas de asociación que sean positivas. (Williams, Dale, & Macnaughton-Smith, 1964).

Esta metodología está diseñada para aglutinar a los individuos en una agrupación de K clústeres, donde K se determina antes del proceso, convirtiéndose esta particularidad en la principal diferencia respecto de los métodos jerárquicos.

Otra particularidad es que este tipo de metodologías parte de la matriz de datos original, por lo que no es necesario pasos alternativos como ocurre con las matrices de distancias y similitudes.

Los siguientes métodos parten de una agrupación inicial en donde pueden formarse los clústeres correspondientes. En la mayoría de los casos, la metodología para encontrar una partición inicial es parte del algoritmo.

4.1.2.2.1 Métodos de reasignación

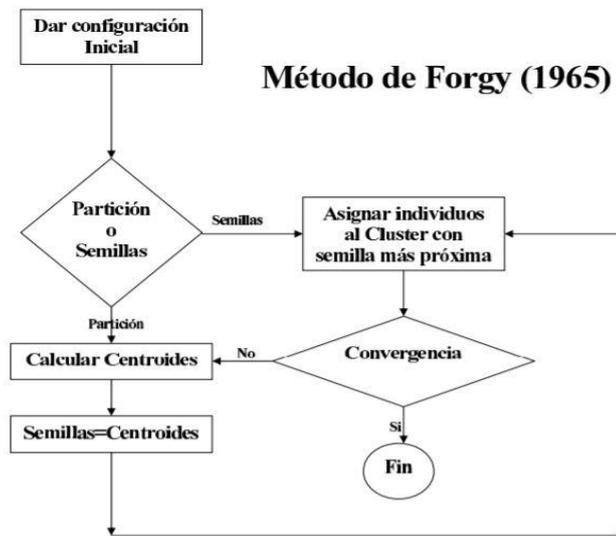
Este método permite recolocar a un individuo dentro de un grupo a otro a través de un determinado proceso de reasignación. Este método finaliza cuando dicha recolocación ha sido optimizada. Particularmente, dentro de este tipo de metodología podemos encontrar: El método de Forgy, El método de K-Medias, Quick-Cluster y el método de nubes dinámicas.

Nosotros principalmente nos centramos en los dos primeros, si deseamos obtener más información véase (Gallardo, 2016).

El método Forgy : Propuesto en 1965, constituye la aproximación más simple al clustering particular. Al igual que el método de McQueen utiliza el concepto de centroide. Forgy planteó un proceso secuenciado en los siguientes pasos (Forgy, 1965).

1. Comenzar con una configuración inicial. Ir al paso segundo si se comienza con un conjunto de puntos semilla. Ir al paso tercero si se comienza con una partición de los casos.
2. Aproximar cada individuo respecto de su semilla más cercana.
3. Obtener el centroide de los clústeres a partir de los puntos semilla.
4. Combinar los dos pasos anteriores hasta que ningún individuo cambie de cluster.

En el proceso de elección de k clústeres se necesita mK operaciones y $m(K - 1)$ comparativas en las distancias. De manera que K en promedio es menor que m y el total de ciclos cuando se consigue la convergencia, no es el esperado, se deja al investigador/a que analice de manera constante el número de clústeres enlazados a los valores de K .



El método de K-Medias: Estudiado por MacQueen (1967), el método de k-medias o k-means relaciona a cada uno de los individuos con el cluster (previamente fijados) que tenga el centroide más próximo. Lo interesante de este método es que el centroide ha sido resuelto desde cada uno de los miembros del cluster a partir de cada asignación en vez del final del ciclo. Este método propone lo siguiente:

1. Los primeros serán clústeres unitarios
2. Al enlazar los $m - K$ individuos al cluster con el centroide más próximo, debemos volver a estimar el centroide después de cada asignación
3. Una vez están todos los individuos, el siguiente paso es volver a seleccionar los centroides de los clusters actuales como puntos de semilla y volver a calcular los valores asignados cada uno, al punto de semilla más cercano.

El procedimiento final es similar al método de Forgy, la única diferencia es que al reasignar se efectúa sin esperar que haya convergencia. Si tenemos los primeros individuos como puntos semilla, será el método más sencillo hasta ahora planteado. El conjunto de operaciones se resumen en cálculo de las distancias, comparaciones de las distancias y el cálculo de los centros.

En la segunda etapa el conjunto de clusters que han sido calculados estarán definidos por la distribución o la manera en que los individuos han sido procesados. En este sentido en 1967, MacQueen concluyó que si los datos son ordenados solo deberían tener un efecto marginal cuando los clusters sean separados.

Desde el último apartado se puede establecer un proceso convergente al que se le conoce como método de k-medias convergente.

1. Partimos con la agrupación de los individuos en clústeres iniciales. Para partir este grupo inicial debe ser realizado usando el primer y segundo paso anteriormente mencionado.
2. Seguir cada uno de los pasos de forma continuada y obtener las distancias para el total de los centros de los clústeres. Si en el cálculo del centroide más cercano no está en torno al cluster padre, reasignar dicho caso al cluster con centroide más próximo y recalcular los centroides de los clusters afectados en el proceso.
3. Continuar con la segunda etapa hasta completarla y se obtenga la convergencia.

4.1.2.2.2 Otros métodos

Métodos de búsqueda de la densidad: Para este tipo de métodos están los que ofrecen una aproximación tipológica y probabilística.

Para el método tipológico, los grupos se forman buscando las áreas en las cuales se da una mayor concentración de individuos. Algún ejemplo de esto puede ser el análisis del modelo de Wishart, el método de Taxamp o el método Fortin.

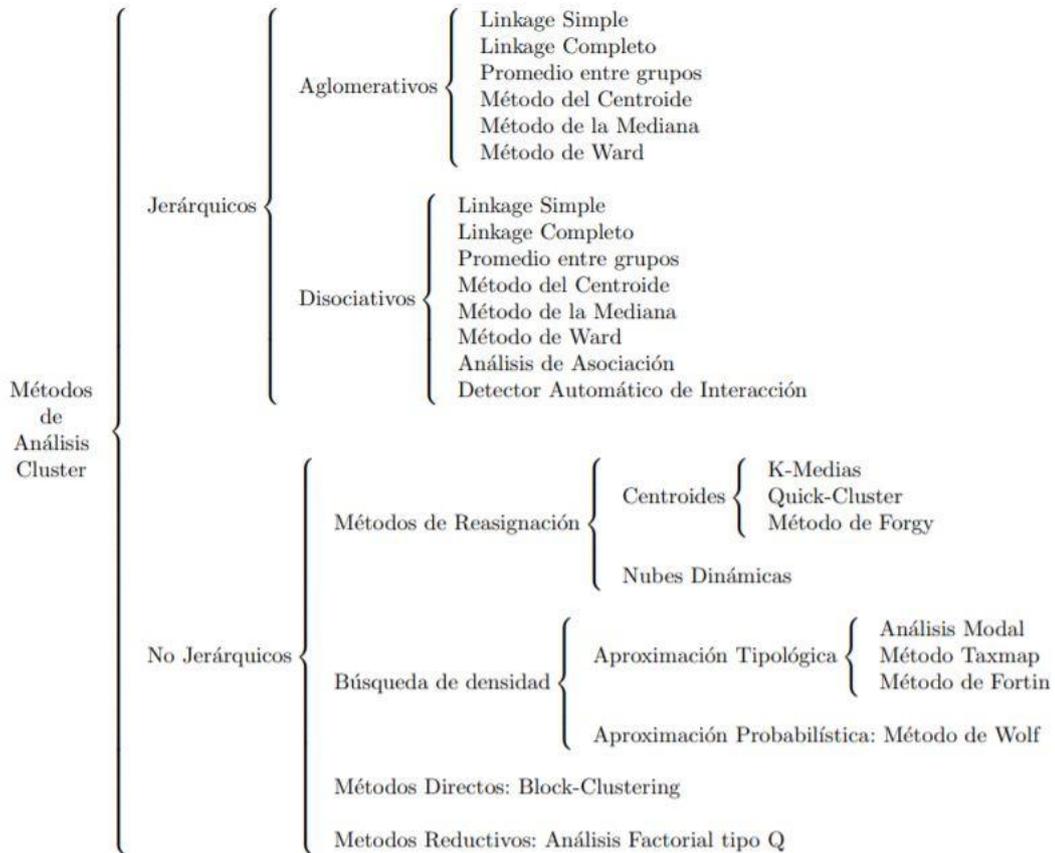
Para el segundo método tenemos que las variables siguen unas reglas de probabilidad en la que los parámetros varían de un clúster a otro. El objetivo es encontrar aquellos individuos que pertenecen a la misma distribución, uno de los métodos más conocidos es el de Wolf.

Métodos directos: Este tipo de metodologías nos ayudan a encasillar al mismo tiempo los individuos y las variables. Dentro de este tipo de metodología podemos destacar el algoritmo Block – Clustering.

El Block-Clustering se apoya en el desarrollo de la minería de datos, el cual le ayuda a establecer agrupamientos al mismo tiempo de las filas y las columnas en una matriz. Este concepto fue planteado por primera vez por (Mirkin, 2012).

Este tipo de métodos recoge una amplia cantidad de librerías sobre las que trabajar en términos de computación. Una de las más populares es la librería Blockcluster de R.

Métodos de reducción de dimensiones: Esta metodología consiste en buscar una serie de factores en el espacio de los individuos, cada factor corresponde a un grupo y se les conoce como análisis factorial tipo Q.



5. Metodología

5.1 Data set

Our Data World y SDG-Tracker colaborando de forma constante con investigadores de la Universidad de Oxford han desarrollado una poderosa base de datos, donde recoge información sobre las condiciones de vida de todo el mundo.

El objetivo de esta herramienta es poder mostrar mediante gráficos, tablas y mapas interactivos la evolución y los cambios que experimentan los países y regiones sobre diferentes disciplinas.

Una de estas disciplinas recoge la evolución del SARS-CoV-2 (COVID-19) de forma acumulada y diaria para cada uno de los países y regiones del mundo.

Se trata de un recurso accesible (código abierto) y útil para el público, ya que unos de los principales fines de esta web es divulgar una información veraz y que represente la evolución de la pandemia dentro del contexto global.

La base de datos recoge al menos 200 regiones entre países y comunidades, donde se monitorizan 55 variables diferentes para cada uno de los países hasta el 1 de Julio del 2021. El manejo de esta base de datos se hace complicado por el gran volumen de información que contiene, es por ello por lo que desechamos aquella información que no nos resultará útil en nuestro análisis.

Sin perder las características de la base datos original, se ha ido seleccionando cuidadosamente las variables y los países que serán objeto de estudio. Estos procesos de selección han sido llevado a cabo a través de la sentencia *subset* en el software Rstudio.

Dentro del conjunto de países y variables que han sido seleccionadas, observamos que algunos carecen de observaciones, por lo que resulta conveniente eliminarlas ya que no podemos obtener información de ellos.

Además de analizar las variables presenten en la base de datos, nos resultaría muy interesante calcular otras a partir de estas, como por ejemplo sería el total de casos por cada 100.000 habitantes o a 14 días con el fin de observar cómo se van comportando las diferentes olas a lo largo del periodo que establezcamos.

Para poder hacernos una mejor idea de aquellos países con los que contamos, en la tabla Nº.1 podemos ver el total de países, estos aparecen agrupados en cada uno de sus respectivos continentes.

Tabla N°.1: Agrupación de los países objeto de estudio respecto de sus continentes.

Asia	Africa	Europe	North America	South America	Oceania
Afghanistan	Cape Verde	Albania	Canada	Argentina	Australia
Bahrain	Gambia	Andorra	Costa Rica	Bolivia	New Zealand
Bangladesh	Ghana	Austria	Dominican Republic	Chile	-
India	Kenya	Belarus	El Salvador	Colombia	-
Indonesia	Malawi	Belgium	Guatemala	Ecuador	-
Iran	Morocco	Bulgaria	Jamaica	Paraguay	-
Iraq	Mozambique	Croatia	Mexico	Peru	-
Israel	Namibia	Cyprus	Panama	Uruguay	-
Japan	Nigeria	Czechia	Trinidad and Tobago	Argentina	-
Jordan	Rwanda	Denmark	United States	Bolivia	-
Kazakhstan	Senegal	Estonia	Canada	Chile	-
Kuwait	South Africa	Finland	Costa Rica	Colombia	-
Malaysia	South Sudan	France	Dominican Republic	Ecuador	-
Maldives	Togo	Germany	El Salvador	-	-
Myanmar	Tunisia	Greece	Guatemala	-	-
Nepal	Uganda	Hungary	Jamaica	-	-
Oman	Zambia	Iceland	Mexico	-	-
Pakistan	Zimbabwe	Ireland	Panama	-	-
Palestine	Cape Verde	Italy	-	-	-
Philippines	Gambia	Latvia	-	-	-
Qatar	Ghana	Liechtenstein	-	-	-
Singapore	Kenya	Lithuania	-	-	-
South Korea	Malawi	Luxembourg	-	-	-
Sri Lanka	Morocco	Malta	-	-	-
Taiwan	Mozambique	Netherlands	-	-	-
Thailand	Namibia	North Macedonia	-	-	-
Turkey	Nigeria	Norway	-	-	-
United Arab Emirates	-	Poland	-	-	-
Vietnam	-	Portugal	-	-	-
-	-	Romania	-	-	-
-	-	Russia	-	-	-
-	-	Serbia	-	-	-
-	-	Slovakia	-	-	-
-	-	Slovenia	-	-	-
-	-	Spain	-	-	-
-	-	Sweden	-	-	-
-	-	Switzerland	-	-	-
-	-	Ukraine	-	-	-
-	-	United Kingdom	-	-	-

Por otra parte, una vez realizado los pasos anteriores nos quedamos con unas series de variables que podemos definir de la siguiente manera:

Total de casos confirmados: Corresponde al número acumulado y diario de personas testadas como positivas en SARS-CoV-2 (COVID-19).

Nuevos casos: Corresponden al total acumulado y diario de nuevas personas testadas como positivas en SARS-CoV-2 (COVID-19).

Total de muertes: Corresponde al número acumulado y diario de muertes asociadas a SARS-CoV-2 (COVID-19).

Nuevas muertes: Corresponde al número diario y acumulado total de nuevas muertes asociadas al SARS-CoV-2 (COVID-19).

Total de casos por millón de habitantes: Corresponde al número acumulado y diario de personas testadas como positivas en SARS-CoV-2 (COVID-19) por cada millón de habitantes.

Nuevos casos por millón de habitantes : Corresponde al total acumulado y diario de nuevas personas testadas como positivas en SARS-CoV-2 (COVID-19) por cada millón de habitantes.

Total de muertes por millón de habitantes: Corresponde al número acumulado y diario de muertes asociadas a SARS-CoV-2 (COVID-19) por millón de habitantes.

Nuevas muertes por millón: Corresponde al número total acumulado y diario de nuevas asociadas a SARS-CoV-2 (COVID-19) por millón de habitantes.

Tasa de reproducción: La tasa de reproducción representa el número promedio diario de nuevas infecciones causadas por un solo individuo infectado. Si la tasa es superior a 1, el virus se transmite de manera comunitaria (de 1 a 1) por lo que este nivel debe estar por debajo de la unidad.

Pacientes en UCIS: Total acumulado y diario de pacientes en Unidad de Cuidados Intensivos (UCI) asociadas a SARS-CoV-2 (COVID-19).

Pacientes en UCIS por cada millón de habitantes: Total acumulado y diario de pacientes en Unidad de Cuidados Intensivos (UCI) asociadas a SARS-CoV-2 (COVID-19) por cada millón de habitantes.

Pacientes hospitalizados: Total de pacientes hospitalizados debido a SARS-CoV-2 (COVID-19).

Pacientes hospitalizados por millón de habitantes: Total de pacientes hospitalizados debido a SARS-CoV-2 (COVID-19) por millón de habitantes.

Admisiones semanales en UCIS: Total acumulado y diario de nuevos pacientes en Unidad de Cuidados Intensivos (UCI) cada semana debido al SARS-CoV-2 (COVID-19).

Admisiones semanales en UCIS por millón de habitantes: Total acumulado y diario de nuevos pacientes en Unidad de Cuidados Intensivos (UCI) cada semana debido al SARS-CoV-2 (COVID-19) por millón de habitantes.

Pacientes semanales hospitalizados: Total acumulado y diario de nuevos pacientes hospitalizados cada semana debido a SARS-CoV-2 (COVID-19).

Pacientes semanales hospitalizados por millón de habitantes: Total de nuevos pacientes hospitalizados cada semana debido a SARS-CoV-2 (COVID-19) por millón de habitantes.

Nuevos test: Total acumulado y diario de nuevos test realizados de forma diaria (incluye cualquier tipo) para detectar SARS-CoV-2 (COVID-19).

Total de test: Total acumulado y diario de test realizados (incluye cualquier tipo) para detectar SARS-CoV-2 (COVID-19).

Total de test por mil habitantes: Total acumulado y diario de test realizados (incluye cualquier tipo) para detectar SARS-CoV-2 (COVID-19) por mil habitantes.

Nuevos test por mil habitantes: Total acumulado y diario acumulado y diario de nuevos test realizados de forma diaria (incluye cualquier tipo) para detectar SARS-CoV-2 (COVID-19) por mil habitantes.

Ratio de test positivos: Total de personas diarias testadas en SARS-CoV-2 (COVID-19) entre el número de pruebas realizadas.

Test por casos : Total acumulado de test diarios realizados en SARS-CoV-2 (COVID-19) por cada positivo.

Total de vacunas: Total acumulado y diario de vacunas disponibles contra la SARS-CoV-2 (COVID-19)

Total de personas vacunadas: Total acumulado y diario de personas que al menos han recibido una dosis contra SARS-CoV-2 (COVID-19).

Total de personas con dosis completa: Total acumulado y diario de personas con la pauta completa contra la SARS-CoV-2 (COVID-19).

Nuevas vacunaciones: Total acumulado y diario de vacunas inyectadas contra la SARS-CoV-2 (COVID-19) por días.

Total de vacunas por mil habitantes: Total acumulado y diario de vacunas disponibles contra la SARS-CoV-2 (COVID-19) por mil habitantes.

Total de personas vacunadas por mil habitantes: Total acumulado y diario de personas que al menos han recibido una dosis contra SARS-CoV-2 (COVID-19) por mil habitantes.

Total de personas con dosis completa por mil habitantes: : Total acumulado y diario de personas con la pauta completa contra la SARS-CoV-2 (COVID-19) por mil habitantes.

Índice de rigurosidad: Indica cómo de responsable se hacen los legisladores a la hora de tomar medidas contra SARS-CoV-2 (COVID-19)

Población: Total de la población residente en el país en cuestión.

Densidad de población: Población en función del territorio.

Edad media: Edad media de los residentes de dicho país.

Mayores de 65 años: Total de la población que cuenta con 65 años o más.

Mayores de 70 años: Total de la población que cuenta con 70 años o más.

Producto Interno Bruto per cápita: Valor promedio de los bienes y servicios que produce cada ciudadano.

Pobreza extrema: Comprende a las personas cuyos hogares tienen ingresos o consumos per cápita inferiores al valor de una cesta de consumo mínima.

Ratio de muerte cardiovascular: Número de muertes asociadas a problemas cardiovasculares entre la población.

Prevalencia a la diabetes: Proporción de personas con diabetes diagnosticada.

Mujeres fumadoras: Total de mujeres fumadoras.

Hombres fumadores: Total de hombres fumadores

Higiene: Indicador de acceso a la higiene.

Camas de hospital por mil habitantes: Total de camas disponibles por cada mil habitantes.

Esperanza de vida: Es la media del número de años que vive una determinada población vive en promedio.

Índice de desarrollo humano: Mide los adelantos medios de un país en tres aspectos básicos del desarrollo humano.

Exceso de mortalidad: Registra las variaciones de muerte respecto del promedio de años anteriores.

Los principales análisis de nuestro trabajo se han focalizado en **el total de casos, total de muertes y ratio de reproducción**, además como anteriormente hemos mencionado se ha calculado otras variantes como el total de casos y muertes por 100.000 habitantes y a 14 días, donde para ello hemos utilizado otras variables disponibles en la base de datos como la población.

La definición resultante sería la siguiente:

Total de casos confirmados por cada 100.000 habitantes: Corresponde al número acumulado y diario de personas testadas como positivas en SARS-CoV-2 (COVID-19) por cada 100.000 habitantes.

Total de las muertes por cada 100.000 habitantes: Corresponde al número acumulado y diario de muertes asociadas a SARS-CoV-2 (COVID-19) por cada 100.000 habitantes.

Total de casos confirmados por cada 100.000 habitantes en 14 días: Corresponde al número acumulado y diario de personas testadas como positivas en SARS-CoV-2 (COVID-19) por cada 100.000 habitantes en 14 días.

Total de las muertes por cada 100.000 habitantes en 14 días: Corresponde al número acumulado y diario de muertes asociadas a SARS-CoV-2 (COVID-19) por cada 100.000 habitantes en 14 días.

Si se desea obtener más información acerca de cómo se ha calculado este proceso o qué variables se han tomado, por favor véase **anexos**.

6. Análisis práctico

6.1 Descripción del proceso

Para poder llevar a cabo este proceso solo es posible si se han realizado los pasos descritos en los capítulos anteriores.

Tenemos en cuenta que partimos de un conjunto de variables, las cuales recogen una lista de 106 países y de la que será objeto de nuestro análisis. La mayoría de estas variables vienen medidas en términos acumulativos y de forma diaria, tan solo el ratio de reproducción viene medido por un cociente y que no se observa de forma acumulada tan solo de forma diaria.

El paquete *fda* de Ramsay, Wickham, Graves, & Hooker (2014) nos ayuda a procesar toda la información obtenida de la base de datos a través de las diferentes sentencias con el fin de obtener las funciones y los resultados deseados.

Una vez está preparado nuestro *data frame*, lo primero que debemos hacer es seleccionar qué tipo y número de bases necesitamos. En nuestro caso particular, hemos seleccionado 15 funciones B-Spline, ya que este tipo puede explicar de manera más eficiente la trayectoria y evolución de nuestras curvas.

La sentencia *Data2fd* nos permite crear un objeto funcional a partir de un conjunto de datos discretos dentro de un espacio determinado y que este trabajo establece en $[0,1]$ con la base 15 de funciones B-Spline como hemos mencionado anteriormente. Este proceso computacional, es repetido para cada una de las variables independientemente del periodo en el que comienzan las observaciones de los países (hace homogéneo el espacio).

Una vez creado el objeto funcional seleccionamos los coeficientes (que nos servirán de argumento para el siguiente paso) el nombre de cada uno de nuestros países, con el fin de identificar a los individuos cuando realicemos el análisis. Fruto de esto, nos queda una matriz 106×15 , cuyas columnas son cada país y cada fila es la base correspondiente.

Una vez obtenidos los pasos anteriores, mediante la sentencia *fd* pasamos a obtener la función de nuestros objetos funcionales.

Resultados de la representación funcional

Ya creadas nuestras funciones, estamos dispuestos a trabajar con ellas y es por esto que representamos las curvas oportunas para cada una de las variables que hemos seleccionado. Poder visualizar las curvas nos permite ver de forma homogénea como se han comportado los países para un intervalo de tiempo determinado.

Para entender mejor cómo se comportan los países de manera natural, los hemos agrupado respecto de sus propios continentes. De este modo, nos quedan 6 curvas las cuales permiten comprar a los continentes en los mismos instantes de tiempo y ver si el factor geográfico ha jugado un papel importante en el desarrollo de su pandemia, independientemente de cuando haya comenzado.

En los siguientes gráficos podemos ver cómo evoluciona el ritmo del total acumulado de casos medios, total acumulado de muertes medias por cada 100.000 habitantes y cada 100.000 habitantes a 14 días, además del ratio de reproducción.

Antes de representar las variables por cada 100.000 habitantes y 100.000 habitantes a cada 14 días debemos aclarar que estas se han seleccionado debido a que se asemejan lo más posible a la realidad. Otras de las ventajas con las que cuenta es que cada 14 días podemos ver la evolución por ciclos, la cual es muy útil para identificar las diferentes olas.

Cada una de las curvas se representa en un espacio entre $[0,1]$ y todas las variables serán medidas por miles de habitantes, excepto el ratio de reproducción que se trata de un cociente y cuyo valor está comprendido mayormente entre 0 y 3 normalmente.

Para poder simplificar la representación visual del análisis hemos asignado un color para cada continente:

Asia: Color negro

África: Color rojo

Europa: Color verde

América del Norte: Color azul oscuro

América del Sur: Color azul claro

Oceanía: Color rosa

Total acumulado de casos medios por cada 100.000 habitantes

En la gráfica Nº.1 podemos observar la evolución acumulada media del total de casos por cada 100.000 habitantes para cada uno de los continentes.

A pesar de que cada continente tiene unas características diferentes, observamos que se ha producido una evolución muy acelerada a partir del punto 0.2 y que se intensifica en la mitad del intervalo [0,1] en algunos continentes, como es el caso de Europa y Sudamérica.

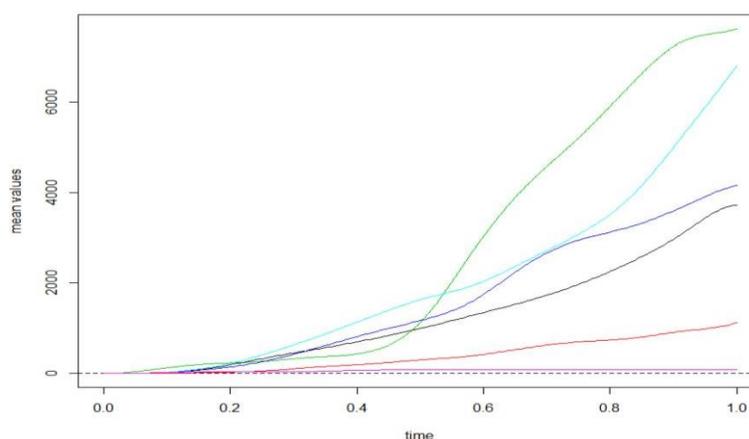
No todas las curvas medias han evolucionado con la misma intensidad, por lo que si dividimos el intervalo [0,1] en tres partes, tenemos lo siguiente:

En una primera parte del espacio [0,0.3] podemos ver como todas las curvas medias evolucionan con la misma suavidad. Los países europeos registran un mayor número de casos, hasta que consiguen controlar la evolución de su pandemia. Al mismo tiempo, otros continentes toman el relevo de esta escalada de casos como por ejemplo América del Norte o América del sur.

Si observamos el intervalo [0.4,0.7] podemos ver como las curvas medias de América, Europa y Asia evolucionan de una manera más rápida que las curvas de África y Oceanía. La evolución de la curva europea a mitad del intervalo [0,1] nos hace tener una idea más clara de cómo de rápido ha evolucionado el ritmo de contagios para las nuevas variantes del virus y que podemos obtener más información en (Hadfield et al., 2018).

Se observa para final del intervalo [0.8,1] una dispersión de las curvas, quizás la más significativa sea la curva europea, la cual sigue acaparando el mayor número de casos. También el comportamiento anticíclico de América del Sur con su vecino América del Norte, mientras una presenta una desaceleración de su curva (América del norte) otra experimenta un fuerte crecimiento.

Gráfica Nº.1: Total acumulado de casos medios por cada 100.000 habitantes.



Visualización: Asia: Negro, África: Rojo, Europa: Verde, Norteamérica: Azul Oscuro, Sudamérica: Azul Claro y Oceanía: Rosa.

Total acumulado de casos medios por cada 100.000 a 14 días

En la gráfica N^o.2 representamos el total de casos acumulados medios por cada 100.000 habitantes a 14 días.

Para poder analizar las diferentes olas, observamos dos partes dentro de nuestra gráfica. En una primera instancia vemos que para el intervalo [0,0.4] todas las curvas medias observadas evolucionan de una manera suave y constante.

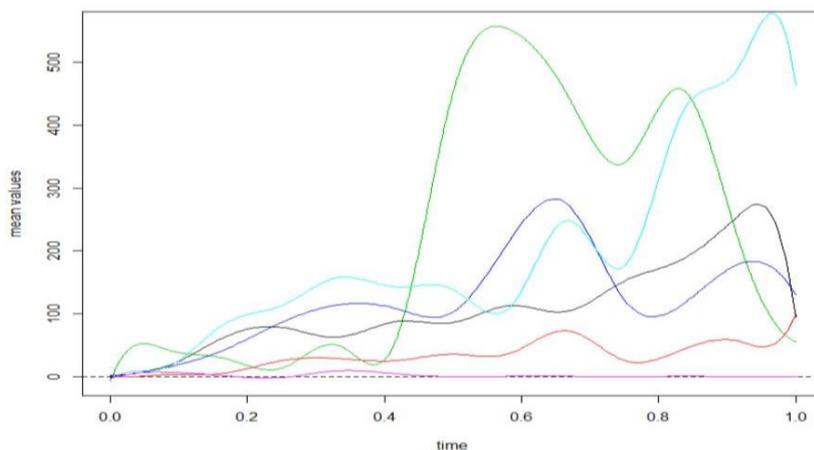
La ola europea experimenta una ola inicial muy fuerte y que es controlada rápidamente, sin embargo continentes como América y Asia presentan olas iniciales inferiores, pero que acaban alcanzando niveles superiores. El continente Africano a comparación del resto experimenta una ola inicial poco acelerada y además constante a lo largo de dicho periodo.

Por otra parte, en la segunda mitad del intervalo [0.5,1] la evolución de las curvas medias se vuelve más dispares entre ellas. Hasta este momento, cada continente ha registrado entre tres y cuatros olas a lo que la evolución del ritmo de contagios se refiere.

La curva europea experimenta la evolución más singular, que pasa de tener prácticamente la situación bajo control a experimentar un fuerte crecimiento. Con retardo y menor intensidad las curvas medias de América y África experimentan el mismo pico en el punto 0,6. Asia y Oceanía son los únicos que presentan unas curvas medias de menor intensidad y con ello un menor número de olas dentro del intervalo.

Centrándonos en la parte final, observamos como la mayoría de los países comienzan a revertir las tendencias de sus curvas medias, sin embargo África sigue experimentando una evolución fuerte hasta el final del intervalo, donde no da signos de un descenso en el total de casos.

Gráfica N^o.2: Total acumulado de casos por cada 100.00 habitantes a 14 días.



Visualización: Asia: Negro, África: Rojo, Europa: Verde, Norteamérica: Azul Oscuro, Sudamérica: Azul Claro y Oceanía: Rosa.

Total acumulado de muertes medias por cada 100.000 habitantes

En la gráfica N.º.3 se puede observar la evolución del total de muertes por cada 100.000 habitantes.

De manera general podemos decir que las curvas medias del Norte América y Europa se posicionan de forma permanente sobre el resto de las curvas. Ambos continentes presentan una evolución similar, mientras que el resto de continentes presenta una evolución más lineal y suave a lo largo de todo el periodo [0,1].

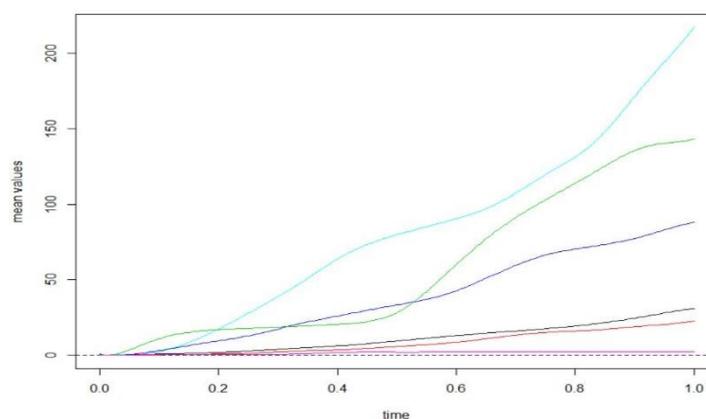
Como en la variable anteriormente comentada, también observamos dos partes, una está comprendida en el intervalo [0,0.5] y la otra para el periodo [0.6,1].

En esta primera parte del espacio, destacamos el comportamiento anticíclico que se observa en la curva sudamericana respecto de la europea, manteniéndose la curva media de Norteamérica entre ambas. Destacar que la curva media europea ha sido la que mayor crecimiento ha experimentado para este periodo, seguido de Norteamérica y Sudamérica. En el lado opuesto a ello podemos observar a Oceanía, Asia y África.

Para el segundo tramo de la gráfica, observamos como la evolución diferente para cada una de las curvas. Las curvas medias que registran más muertes por cada 100.000 habitantes son la curva media europea y sudamericana, mientras que las de menor registro presentan son las curvas perteneciente a África, Asia y Oceanía.

Otra de las cosas que hemos observado en este tramo es que las curvas siguen una trayectoria cíclica, a pesar de que no todas presentan la misma intensidad en su evolución (véase la curva perteneciente Sudamérica y Europa para más información).

Gráfica N.º.3: Total acumulado de muertes por cada 100.00 habitantes.



Visualización: Asia: Negro, África: Rojo, Europa: Verde, Norteamérica: Azul Oscuro, Sudamérica: Azul Claro y Oceanía: Rosa.

Total de muertes por cada 100.000 habitantes a 14 días

En el Gráfico N°.4 podemos ver el total acumulado de muertes por cada 100.000 habitantes en 14 días.

En términos generales podemos observar como la mayoría de las curvas evolucionan de manera diferente para el espacio $[0,1]$. De una manera hay curvas medias con unas olas muy volátiles y de otra hay curvas medias que presentan unas olas mucho más suaves.

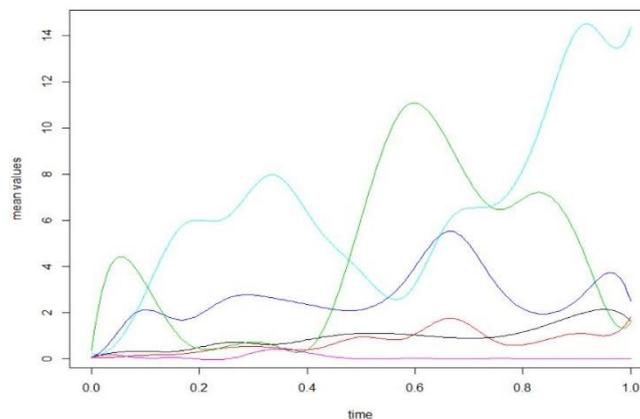
Las olas que observadas en las curvas medias de Europa y Sudamérica son con carácter general anticíclicas, mientras que si atendemos al resto de continentes (África, Asia y Oceanía) como un grupo de curvas independientes, observamos que experimentan una evolución cíclica (véase la primera mitad del gráfico) y anticíclica (véase la segunda mitad del gráfico) en el espacio.

En la primera mitad de gráfico vemos como la curva europea evoluciona de manera pronunciada, al mismo tiempo, pero algo retardada lo hace el continente americano (ambos hemisferios). A diferencia de la europea, las curvas americanas no consiguen controlar la crisis sanitaria, en términos comparados. El resto de continentes por su parte sí lo hace, estos presentan unas evoluciones más suaves y menos volátiles.

Para la segunda mitad del espacio, la evolución de las curvas medias se dispersan unas respecto de otras al igual que para el total de casos. La evolución de las olas son más acusadas (en cuanto a incrementos y disminuciones se refiere) para menores periodos de tiempo. Esto se puede observar muy bien en la curva europea o sudamericana.

Finalmente, destacar que para la mayoría de las curvas se observa una ola en el mismo periodo, véase el punto 0,7.

Gráfica N°.4: Total acumulado de muertes por cada 100.00 habitantes a 14 días



Visualización: Asia: Negro, África: Rojo, Europa: Verde, Norteamérica: Azul Oscuro, Sudamérica: Azul Claro y Oceanía: Rosa.

Tasa de Reproducción

En la gráfica Nº.5 podemos ver representadas las curvas medias del ratio de reproducción para cada continente.

Para todo el espacio, la evolución de esta variable se ha movido principalmente entre 0 y 2 en cada uno de los continentes. Si observamos de manera general el gráfico, podemos ver como todas las curvas medias del ratio de reproducción han ido evolucionando de manera cíclica.

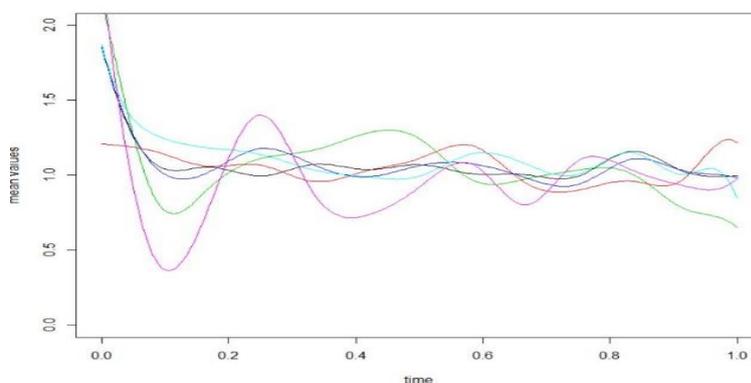
Para esta variable observamos como todas las curvas tienden a situarse por debajo de la unidad. Sin embargo, sí que podemos destacar que los incrementos que se observan en cada una de las curvas corresponden con los diferentes incrementos observados en el total acumulado de casos y muertes.

De manera general, observamos que en la primera mitad del gráfico las curvas parten de un valor superior a 2 y que este tienda a estar por debajo de la unidad, pero no todas las curvas tienden a ello de la misma manera. Si observamos a Oceanía, el ratio de reproducción generalmente inferior que para el resto de curvas, mientras que en América o Europa el valor observado está generalmente muy por encima de la unidad.

Estas fluctuaciones se irán diluyendo a medida que nos acercamos al límite del intervalo [0,1]. Para ello véase el tramo final donde todas las curvas se sitúan por debajo y en torno a 1, exceptuando África que repunta en ratio de contagios (relacione este dato con la última ola para el total de muertes por 100.000 habitantes a 14 días).

La tendencia esperada de todas estas curvas es que a medida que avancen los ritmos de vacunación estos valores tiendan a situarse por debajo de 1 y que la dimensión de las olas sean cada vez menores.

Gráfica Nº.5: Ratio de reproducción



Visualización: Asia: Negro, África: Rojo, Europa: Verde, Norteamérica: Azul Oscuro, Sudamérica: Azul Claro y Oceanía: Rosa.

La relación entre los tres tipos de variables, es clara. Si observamos los incrementos iniciales en el ratio de reproducción, podemos interpretar la evolución del total de contagios y este a su vez el número de muertes.

En términos generales para el total de las curvas, en cada una de las variables estudiadas se puede decir que parten o experimentan un crecimiento acelerado normalmente al inicio del intervalo $[0,1]$. Las curvas que mayor crecimiento presentan son las europeas, americanas y asiáticas frente a las africanas y las oceánicas las cuales parten de unos inicios mucho más suaves.

Seguidamente, en la parte intermedia del intervalo las curvas experimentan evoluciones y relaciones diferentes, dependiendo del tipo de variable que analizamos. Salvaguardado el ratio de reproducción, el resto de variables experimentan fuertes crecimientos y fluctuaciones de sus curvas.

En el tramo final, ocurre normalmente lo contrario de la primera mitad del intervalo $[0,1]$. La mayoría de curvas tienden a disminuir o suavizar la aceleración de sus curvas. Sin embargo, las curvas africanas y oceánicas presentan casi la misma tendencia desde inicio a fin, a excepción de algunos periodos concretos.

Otra de las visiones generales que nos queda sobre las curvas medias, es que podemos apreciar lo siguiente:

Mientras que Europa en breves pero intensos periodos de tiempo ha sido capaz de controlar dicha evolución, en el resto de continentes no les ha resultado tan fácil (véase sudamérica en el gráfico N^o.1). Otros continentes partiendo también de una situación menos favorable que Europa o Sudamérica, acaban experimentando una curva menor debido a los criterios adoptados para el control de la pandemia (hablamos de Asia).

Lo mismo ocurre si comparamos la evolución de los casos y del total de muertes. Podemos observar que para algunas curvas, las pequeñas variaciones en el total de casos produce grandes variaciones en el total de muertes. Esto se observa claramente en las curvas medias de América, en la medida que aumentan el total de casos también lo hace en mayor proporción el total de muertes, en el caso opuesto tendríamos el continente africano.

De manera individual podemos ver que Europa ha sido uno de los continentes más golpeados por la pandemia. Los niveles de contagios han sido altos para todo el periodo, así como un alto registro de las tasas del total de muertes. También ha sido una de las primeras en invertir sus curvas medias al final del periodo $[0,1]$, posiblemente debido a los avances en los planes de vacunación.

Por otra parte, las curvas que representan al continente asiático (punto inicial de la pandemia) han evolucionado en términos generales de forma más suave que las europeas y americanas pero algo más volátiles que las africanas y oceánicas. En mi opinión, sus respectivas curvas asiáticas se ven más semejantes a las curvas de Europa y Sudamérica que a las de África u Oceanía.

Las curvas medias del continente africano experimentan una evolución más suave y menos volátil que el resto de continentes. Sin embargo, atendiendo a los pequeños ciclos que experimentan en cada una de las curvas podemos observar que tiene ciertas similitudes con las curvas medias de Sudamérica.

Tanto el norte como el sur del continente americano experimentan una evolución similar, a pesar de esto, el sur se ha visto en promedio más afectado que el norte. Para el inicio de la pandemia ha experimentado un comportamiento anticíclico respecto de otros continentes como Europa, sin embargo estos ciclos han tendido a ser homogéneos a medida que ha ido avanzando en el espacio.

Finalmente, un caso particular ha sido Oceanía donde sus curvas han sido mucho más suaves que para cualquier continente. El número de olas tanto para el total de casos como para el total de muertes han sido menores. Uno de los continentes con lo que refleja similitudes es con el continente africano, donde para la mayoría de las olas en las respectivas curvas presentan posiciones similares.

7. Agrupación clúster

Una vez expuesto el marco teórico, pasamos a la obtención de nuestros clústeres a partir de la información que disponemos y para llevar a cabo este cálculo nos basamos en la metodología k-medias la cual pertenece al conjunto no jerárquico.

El objetivo de este análisis clúster es poder hacer una comparativa a través de los distintos argumentos discretos, coeficientes básicos y scores y ver cómo se agrupan los 106 países. Además, queremos ver si existe, y qué tipo de diferencia presentan las agrupaciones cuando aplicamos diferentes argumentos al algoritmo k-medias.

Como tenemos que seleccionar un determinado número de clúster para poder trabajar con el método K-medias, buscamos una manera que optimice este número de clústeres con el fin de no incurrir en un error de exceso y defecto.

La librería más eficiente y disponible para trabajar en R es, *NbClust* y esta nos indica cual es el número óptimo de clústeres para nuestro conjunto de datos. Esta sentencia es compatible con K-medias por lo que no existe ningún inconveniente a la hora de realizar nuestro análisis.

Debemos aclarar, que para aquellos casos en los que ha resultado el número óptimo de clústeres ha sido *dos*, hemos decidido añadir un clúster con el fin de obtener más información sin incurrir demasiado en un análisis ineficiente.

Una vez obtenido los grupos para cada una de las variables (total de casos acumulados por cada 100.000 habitantes, total de muertes acumuladas por cada 100.000 habitantes y ratio de reproducción) se han representado sus curvas medias con el fin de obtener una mejor visualización, manejo y análisis. Este proceso se ha repetido con los tres argumentos anteriormente señalados.

Al igual que para el apartado anterior las curvas están definidas por un espacio entre [0,1] en donde cualquier instante puntual en el que se mida una determinada curva se podrá hacer lo mismo para el resto.

7.1 Clustering

Total de casos acumulados por cada 100.000 habitantes.

Los resultados obtenidos de aplicar los distintos argumentos al algoritmo k-medias, han sido los mismos en cada uno de ellos. La única excepción es Portugal que pasa del grupo Nº.1 al Nº.2 para clúster obtenido mediante los coeficientes básicos. En total, observamos que el tipo discreto y scores cuenta un primer grupo de catorce países, un segundo de cuarenta y un tercero con cincuenta y dos. Para los coeficientes básicos, el primer grupo cuenta con trece países el grupo dos con cuarenta y uno (se integra Portugal en este grupo), mientras que el tercero mantiene los valores (cincuenta y dos países).

La agrupación de los países en torno a cada uno de sus respectivos grupos se ha realizado de una manera muy peculiar.

A grandes rasgos podemos describir que para el primer clúster se agrupan países de un tamaño reducido y con alto nivel de desarrollo como es el caso de Bélgica, Países Bajos, Luxemburgo o Andorra a excepción de EE.UU que tiene un tamaño mayor y aparece en el grupo.

En un segundo clúster aparecen la mayoría de los países europeos y algunos países de diversos continentes (exceptuando África) pero de similar nivel de desarrollo económico. Es en este grupo donde aparecen España, Italia, Francia, Reino Unido, Emiratos Árabes o Argentina a modo de ejemplo.

En el tercer grupo agrupan a todos los países del continente Africano, además de países de centro América (México o Guatemala), Asia (Japón o Corea del Sur) y Europa (Finlandia o Noruega).

Unas de las singularidades que aparecen para cada uno de los diferentes argumentos es que EE.UU siempre se agrupa en torno países con una dimensión muy reducida, incluso siendo estos mayoritariamente europeos. Por otro lado, también nos llama la atención como la mayoría de los países del norte de Europa se concentran con la mayoría de los países del continente africano y Asia occidental.

Finalmente, destacar que España se enmarca en el mismo clúster que el resto de las economías europeas y de su entorno, excepto Portugal la cual aparece situada en el primer clúster junto a varios microestados. Véase la tabla Nº.2 y Nº.3 para más información.

Tabla Nº.2: Clúster Discreto y Clúster Scores para el total de casos acumulados por cada 100.000 habitantes

Grupo Nº.1			
Belgium	Czechia	Israel	Lithuania
Netherlands	Panama	Portugal	Serbia
Sweden	United States	Andorra	Bahrain
Luxembourg	Slovenia	-	-
Grupo Nº.2			
Argentina	Austria	Belarus	Bulgaria
Cape Verde	Chile	Colombia	Costa Rica
Croatia	Cyprus	Denmark	Estonia
France	Germany	Hungary	Ireland
Italy	Jordan	Kuwait	Latvia
Liechtenstein	Maldives	Malta	North Macedonia
Oman	Palestine	Paraguay	Peru
Poland	Qatar	Romania	Slovakia
Spain	Switzerland	Turkey	Ukraine
United Arab Emirates	United Kingdom	Uruguay	Albania
Grupo Nº.3			
Afghanistan	Australia	Bangladesh	Bolivia
Canada	Dominican Republic	Ecuador	El Salvador
Finland	Gambia	Ghana	Greece
Guatemala	Iceland	India	Indonesia
Iran	Iraq	Jamaica	Japan
Kazakhstan	Kenya	Malawi	Malaysia
Mexico	Morocco	Mozambique	Myanmar
Namibia	Nepal	New Zealand	Nigeria
Norway	Pakistan	Philippines	Russia
Rwanda	Senegal	Singapore	South Africa
South Korea	South Sudan	Sri Lanka	Taiwan
Thailand	Togo	Trinidad and Tobago	Tunisia
Uganda	Vietnam	Zambia	Zimbabwe

Tabla Nº.3: Clúster Coeficientes básicos para el total de casos acumulados por cada 100.000 habitantes.

Grupo Nº.1			
Andorra	Bahrain	Belgium	Czechia
Lithuania	Luxembourg	Netherlands	Panama
Slovenia	Sweden	United States	Israel
Serbia	-	-	-
Grupo Nº.2			
Albania	Argentina	Austria	Ukraine
Belarus	Bulgaria	Cape Verde	Uruguay
Chile	Colombia	Costa Rica	Switzerland
Croatia	Cyprus	Denmark	United Arab Emirates
Estonia	France	Germany	Slovakia
Hungary	Ireland	Italy	Turkey
Jordan	Kuwait	Latvia	United Kingdom
Liechtenstein	Maldives	Malta	Spain
North Macedonia	Oman	Palestine	Romania
Paraguay	Peru	Poland	Qatar
Portugal	-	-	-
Grupo Nº.3			
Afghanistan	Australia	Bangladesh	Togo
Bolivia	Canada	Dominican Republic	Uganda
Ecuador	El Salvador	Finland	Zimbabwe
Gambia	Ghana	Greece	Trinidad and Tobago
Guatemala	Iceland	India	Vietnam
Indonesia	Iran	Iraq	Tunisia
Jamaica	Japan	Kazakhstan	Zambia
Kenya	Malawi	Malaysia	South Sudan
Mexico	Morocco	Mozambique	Thailand
Myanmar	Namibia	Nepal	South Korea
New Zealand	Nigeria	Norway	Taiwan
Pakistan	Philippines	Russia	South Africa
Rwanda	Senegal	Singapore	Sri Lanka

Total de muertes acumuladas por cada 100.000 habitantes.

Los resultados de aplicar los diferentes argumentos al algoritmo k-medias han sido los mismos para los tres tipos. En total, observamos cuatro clústeres en los cuales el primer grupo cuenta con un país, el segundo con veinte y cinco, el tercero con veinte y seis y el cuarto con cincuenta y cuatro.

Lo más curioso de este análisis es que para nuestro primer clúster sólo se observa un país, Perú. Esto se puede observar en la tabla N.º.4, donde el algoritmo lo agrupa de forma individual para los tres tipos de argumento.

El segundo grupo aglutina a los países que mayor niveles de muertes presentan, estos son principalmente europeos como Reino Unido, España o Italia pero también aparecen del continente americano, como por ejemplo EE.UU, México o Argentina. Si seguimos observando sus características de manera más individual, podemos ver que por tamaño destacamos a Andorra.

Dentro del tercer clúster, la gran parte de los países son procedentes del este y centro de Europa como Alemania, Luxemburgo o Ucrania; sin embargo también aparecen países de centro América como Tunisia, Uruguay y Ecuador. Destacar que dentro de este grupo, la mayoría de estos países aparece junto a un país vecino, es decir, que para el caso de Austria observamos Alemania y para Israel encontramos a Palestina.

Para el grupo número N.º.4 al igual que para el total de casos, encontramos a todos los países del continente africano y junto a ellos también se observan una amalgama de países procedentes de diferentes continentes, entre los que tenemos que destacar a Finlandia y Noruega como las economías más importantes del grupo.

De forma resumida, decimos que no aparecen variaciones entre los diferentes tipos de argumentos utilizados. Para el primer grupo aparece un valor aislado y que resultaría interesante indagar en (Hadfield et al., 2018) por si alguna variante haya podido determinar esta característica.

Finalmente, destacar que observamos a países como Argentina donde para el total acumulado de casos por 100.000 habitantes se rodea de países con niveles más bajos de casos acumulados, mientras que en esta variable se rodea de los países con más muertes acumuladas por 100.000 habitantes del mundo. Esto quiere decir que ha debido haber algún factor que ha determinado un número mayor de muertes de lo que se esperaba.

Tabla Nº.4: Clúster Discreto, Scores y Coeficientes básicos para el total de muertes acumuladas por cada 100.000 habitantes.

Grupo Nº.1				
Peru	-	-	-	-
Grupo Nº.2				
Argentina	Belgium	Bulgaria	Chile	Colombia
Croatia	Czechia	France	Hungary	Italy
Liechtenstein	Lithuania	Mexico	North Macedonia	Panama
Poland	Portugal	Romania	Slovakia	Slovenia
Spain	Sweden	United Kingdom	United States	Andorra
Grupo Nº.3				
Albania	Austria	Bolivia	Canada	Costa Rica
Estonia	Germany	Greece	Iran	Ireland
Jordan	Latvia	Luxembourg	Malta	Netherlands
Paraguay	Russia	Serbia	South Africa	Switzerland
Ukraine	Uruguay	Ecuador	Israel	Palestine
Tunisia	-	-	-	-
Grupo Nº.4				
Afghanistan	Australia	Bahrain	Bangladesh	Singapore
Belarus	Cape Verde	Cyprus	Denmark	Taiwan
Dominican Republic	El Salvador	Finland	Gambia	Turkey
Ghana	Guatemala	Iceland	India	Zambia
Indonesia	Iraq	Jamaica	Japan	South Korea
Kazakhstan	Kenya	Kuwait	Malawi	Thailand
Malaysia	Maldives	Morocco	Mozambique	Uganda
Myanmar	Namibia	Nepal	New Zealand	Zimbabwe
Nigeria	Norway	Oman	Pakistan	Trinidad and Tobago
Philippines	Qatar	Rwanda	Senegal	Vietnam
United Arab Emirates	Togo	South Sudan	Sri Lanka	-

Ratio de reproducción

Tras observar los diferentes resultados en cada uno de los argumentos descritos podemos decir que todos los países se agrupan en torno a tres clústeres, a pesar de que cada argumento agrupa a los países de manera diferente.

De manera breve podemos apreciar que para el argumento discreto tenemos un primer grupo con treinta y un países, un segundo con treinta y dos países y un tercero con cuarenta y tres países. Mientras que para los scores tenemos tres grupos un primero con treinta, un segundo treinta y dos, y un tercer con cuarenta y cuatro países respectivamente. Finalmente, para los coeficientes básicos tenemos otros tres clústeres, un primero con treinta y cuatro, un segundo con treinta y tres y un tercero con treinta y nueve países.

Para el tipo discreto observamos como en el primer clúster aglutina diferentes países procedentes de distintos continentes, exceptuando África. Se trata de países y economías desarrolladas o en vías de desarrollo, que cuentan con una población relativamente baja respecto de otros países. Un segundo grupo, recoge las principales economías europeas como Reino Unido, Italia o España, mientras que el tercero recoge los países africanos y algunos asiáticos como Japón o Singapur. En este último ya no aparecen como solían hacerlo las economías del norte de Europa, ya que ahora se sitúan en torno al primer y segundo clúster.

Por otra parte, los grupos de países observados para los scores son prácticamente los mismos, la única diferencia presentada es Tailandia que se va al primer grupo y Chile al tercero grupo, si realizamos la comparación con el clúster tipo discreto.

Del mismo modo, para los coeficientes básicos tenemos que el total de grupos han sido tres, pero siendo la agrupación de los países observados por cada grupo muy diferentes entre sí.

Además, para los coeficientes básicos podemos decir que los países se agrupan de una manera más heterogénea para cada uno de tres grupos respecto de los otros argumentos. Tan solo destacar que es en el tercer clúster donde se agrupan las principales economías del mundo como puede ser el G5 europeo, EE.UU. o Rusia.

Tabla Nº.5: Clúster Discreto para el ratio de reproducción

Grupo Nº.1			
Albania	Andorra	Bulgaria	South Sudan
Croatia	Cyprus	Denmark	Trinidad and Tobago
Estonia	Finland	Greece	Vietnam
Hungary	Iceland	Jamaica	Taiwan
Jordan	Latvia	Liechtenstein	Tunisia
Lithuania	Luxembourg	Malta	Togo
Myanmar	New Zealand	North Macedonia	Uruguay
Palestine	Slovakia	Slovenia	-
Grupo Nº.2			
Afghanistan	Argentina	Bahrain	Singapore
Bolivia	Cape Verde	Colombia	Uganda
Costa Rica	Dominican Republic	El Salvador	Zimbabwe
Gambia	Ghana	Guatemala	South Africa
India	Indonesia	Iraq	United Arab Emirates
Japan	Kazakhstan	Kenya	Senegal
Kuwait	Malawi	Maldives	Sri Lanka
Mexico	Morocco	Mozambique	Zambia
Namibia	Nepal	Nigeria	Peru
Oman	Pakistan	Panama	Rwanda
Paraguay	Philippines	Qatar	-
Grupo Nº.3			
Australia	Austria	Bangladesh	Belarus
Canada	Chile	Czechia	Ecuador
Germany	Iran	Ireland	Israel
Malaysia	Netherlands	Norway	Poland
Romania	Russia	Serbia	South Korea
Sweden	Switzerland	Thailand	Turkey
United Kingdom	United States	Spain	Italy
France	Belgium	Ukraine	Portugal

Tabla N.º.6: Clúster Scores para el ratio de reproducción

Grupo N.º.1			
Albania	Andorra	Bulgaria	Croatia
Cyprus	Denmark	Estonia	Finland
Greece	Hungary	Iceland	Jamaica
Jordan	Latvia	Liechtenstein	Lithuania
Luxembourg	Malta	Myanmar	New Zealand
North Macedonia	Palestine	Slovakia	Slovenia
South Sudan	Taiwan	Togo	Trinidad and Tobago
Tunisia	Uruguay	Vietnam	Thailand
Grupo N.º.2			
Afghanistan	Argentina	Bahrain	Bolivia
Cape Verde	Colombia	Costa Rica	Dominican Republic
El Salvador	Gambia	Ghana	Guatemala
India	Indonesia	Iraq	Japan
Kazakhstan	Kenya	Kuwait	Malawi
Maldives	Mexico	Morocco	Mozambique
Namibia	Nepal	Nigeria	Oman
Pakistan	Panama	Paraguay	Peru
Philippines	Qatar	Rwanda	Senegal
Singapore	South Africa	Sri Lanka	Uganda
United Arab Emirates	Zambia	Zimbabwe	Chile
Grupo N.º.3			
Belarus	Belgium	Canada	Czechia
Ecuador	France	Germany	Iran
Israel	Italy	Malaysia	Netherlands
Poland	Portugal	Romania	Russia
South Korea	Spain	Sweden	Switzerland
Turkey	Ukraine	United Kingdom	Australia
Bangladesh	United States	Serbia	Norway
Austria	Ireland	-	-

Tabla N.º.7: Cluster Coeficientes básicos para el ratio de reproducción

Grupo N.º.1			
Albania	Andorra	Bulgaria	Togo
Costa Rica	Cyprus	Denmark	Vietnam
Estonia	Finland	Greece	Taiwan
Hungary	Iceland	Iraq	Tunisia
Jamaica	Jordan	Latvia	Sri Lanka
Liechtenstein	Lithuania	Maldives	Trinidad and Tobago
Malta	Myanmar	New Zealand	South Sudan
North Macedonia	Palestine	Paraguay	Slovenia
Slovakia	-	-	-
Grupo N.º.2			
Afghanistan	Argentina	Bahrain	Singapore
Bolivia	Cape Verde	Colombia	United Arab Emirates
Dominican Republic	El Salvador	Gambia	Zimbabwe
Ghana	Guatemala	Indonesia	South Africa
Japan	Kazakhstan	Kenya	Uruguay
Kuwait	Malawi	Mexico	Uganda
Mozambique	Namibia	Nepal	Zambia
Nigeria	Oman	Peru	Senegal
Qatar	Rwanda	-	-
Grupo N.º.3			
Australia	Austria	Bangladesh	Belarus
Belgium	Canada	Chile	Croatia
Czechia	Ecuador	France	Germany
India	Iran	Ireland	Israel
Italy	Luxembourg	Malaysia	Morocco
Netherlands	Norway	Pakistan	Panama
Philippines	Poland	Portugal	Romania
Russia	Serbia	South Korea	Spain
Sweden	Switzerland	Thailand	Turkey
Ukraine	United Kingdom	United States	-

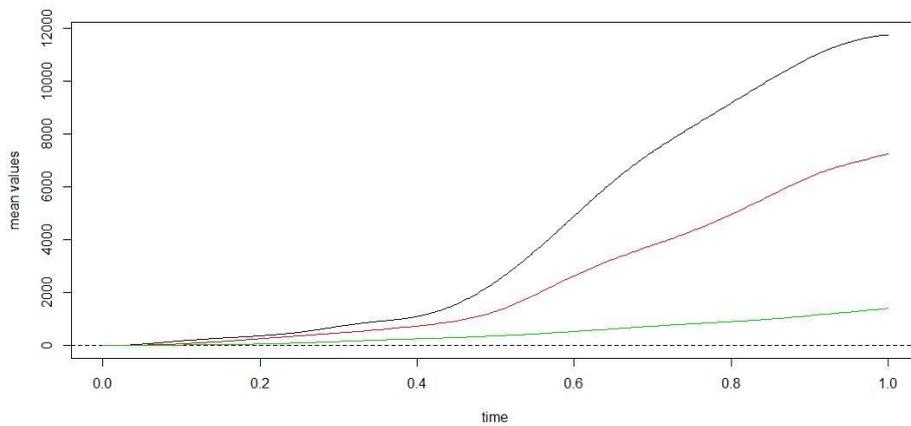
7.2 Análisis curva media

En este apartado calculamos las curvas medias y su posterior representación para cada uno de distintos grupos calculados por los distintos argumentos computados en el algoritmo k-medias en cada una de las variables ya señaladas.

Total de casos acumulados por 100.000 habitantes.

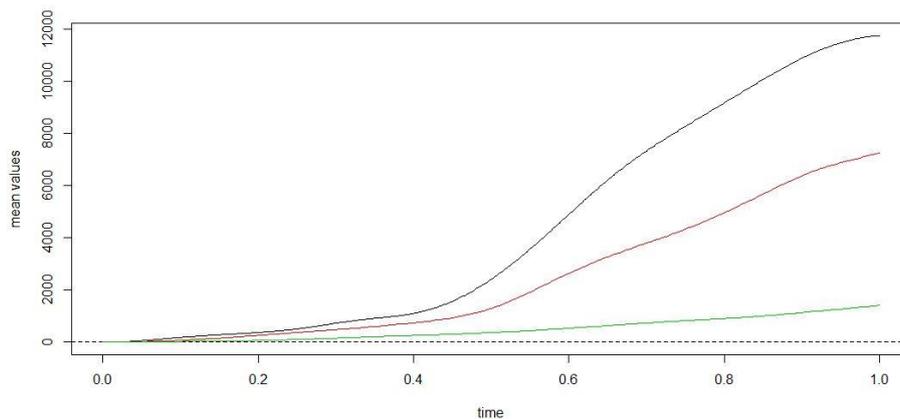
Una vez hemos obtenido los distintos grupos para el total de casos acumulados por 100.000 habitantes y atendemos al gráfico N.º.6 y N.º.7 podemos observar tres curvas donde cada una corresponde a un grupo diferente.

Gráfica N.º.6: Curvas medias acumuladas del total de casos por cada 100.000 habitantes por Continentes para análisis discreto y scores.



Visualización: Grupo N.º.1: Negro, Grupo N.º.1: Rojo, Grupo N.º.1: Verde

Gráfica N.º.7: Curvas medias acumuladas del total de casos por cada 100.000 habitantes por continentes para análisis por coeficientes básicos



Visualización: Grupo N.º.1: Negro, Grupo N.º.1: Rojo, Grupo N.º.1: Verde

La primera característica que nos llama la atención es que si observamos ambos gráficos, podemos decir que estamos ante curvas similares. Esto es algo que se espera, ya que observamos prácticamente los mismos grupos de países, tan solo Portugal varía de un grupo a otro y dicho movimiento no hace que las trayectorias de las curvas varíen.

Si nos centramos en cualquiera de los dos gráficos, tenemos que en el grupo N^o.1 se observa la mayor curva media. Intuimos que esta evolución viene dada por uno de sus países, EE.UU. El país Norte Americano ha sido uno de los más golpeados por la pandemia debido a escasas políticas para contener la expansión del virus.

Además de esto, el primer grupo presenta la curva con mayor pendiente respecto de los otros grupos, observe cómo se produce un cambio de la pendiente en el punto 0.5, gráfico N^o.6. Del mismo modo, el grupo N^o.2 formado por la mayoría de las economías europeas experimenta una evolución similar al grupo N^o.1, pero en menor promedio. Cabría reseñar que dicho grupo al estar formado por la mayoría de las economías europeas podemos ver similitudes con las curvas medias del continente europeo analizadas en el capítulo anterior.

Sin embargo, la evolución que se observa en el tercer grupo de países es mucho más suave. Aquí está representado principalmente a los países del continente africano y oeste asiático, además de los países del norte de Europa. Si comparamos este grupo con el primero, podemos interpretar cómo algunos de estos países han pasado de “puntillas” sobre los efectos de la pandemia.

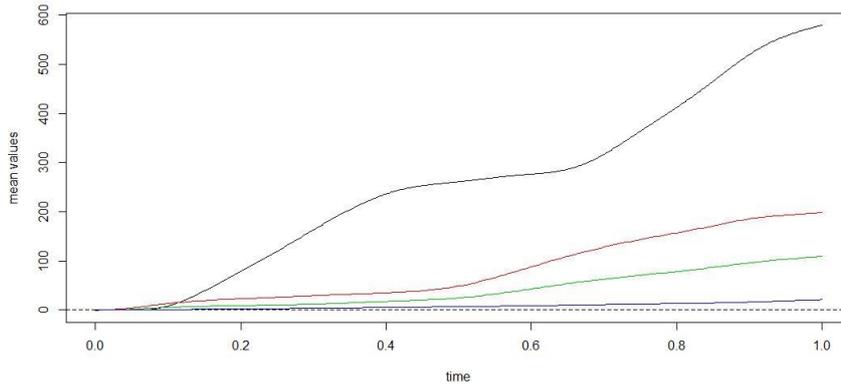
Si hay algo que tienen en común estos tres grupos es a partir del instante 0.5, donde sus curvas medias presentan una aceleración mucho mayor que en la primera mitad del espacio. A pesar de la magnitud de crecimiento, dentro de la segunda mitad se puede observar ritmos de crecimiento diferentes pero con las mismas olas para los mismos instantes de tiempo, permitiendo así observar una evolución cíclica.

Finalmente, destacar cómo la evolución de cada una de las curvas en esta variable guarda relación con cada uno de los grupos que se han representado en el capítulo anterior, si se observa de manera abierta los gráficos podemos ver como las curvas trazan el mismo dibujo en cada uno de ellos.

Total de muertes acumuladas por cada 100.000 habitantes.

De la misma manera que en el apartado anterior, hemos calculado la curva la curva media para el total de muertes acumuladas por 100.000 habitantes en cada uno de los distintos argumentos. Observamos que la agrupación de los países para cada uno de los distintos argumentos ha sido la misma, entonces se espera que la representación gráfica sea idéntica.

Gráfica N.º.8: Total acumulado de muertes medias por cada 100.000 habitantes para el análisis discreto, scores y por coeficientes básicos.



Visualización: Grupo N.º.1: Negro, Grupo N.º.2: Rojo, Grupo N.º.3: Verde

Cada una de las curvas experimenta una evolución diferente, pero prestamos especial atención al grupo N.º.1 (Perú), donde se observa una curva media muy superior respecto del resto de grupo en todo el periodo. En su curva se ve un mínimo de cuatro ciclos, los cuales presentan una pendiente mayor en el intervalo [0.1,0.4] y [0.7,0.9].

Como bien decimos anteriormente, la evolución del total de muertes es similar al total de casos ya que una es causa de la otra. Dentro de las curvas del grupo N.º.2 se observa a la mayoría de los países europeos donde el número de muertes ha sido mayor que en los grupos N.º.3 y N.º.4. En este N.º.2 observamos a países como Reino Unido, Italia y España entre otros.

Por otra parte, destacar cómo el clúster N.º.4 sigue una evolución en su curva media similar a la observada en el apartado anterior (suave y constante). Dentro de esta agrupación podemos observar cómo hay países del entorno europeo (Noruega y Finlandia) que han registrado una menor incidencia respecto de sus socios comunitarios.

Finalmente, destacar que si comparamos las curvas observadas en este apartado con las del punto anterior, podemos ver cómo las trayectorias de las curvas son diferentes. Esta discrepancia puede deberse a que existen países que han presentado unos niveles altos de casos acumulados y un menor número de muertes acumuladas, mientras que otros han presentado pocos casos acumulados y un número mayor de muertes acumuladas.

Ratio de reproducción

Una vez realizado el mismo proceso que en los pasos anteriores, los grupos obtenidos son muy similares a los calculados en la variable total de casos acumulados por 100.000 habitantes a excepción de dos países, los cuales son Filipinas y Perú.

Si representamos las curvas medias de los grupos obtenidos por el argumento discreto y scores respecto del coeficientes básicos podemos ver que en esta última es donde se observa de manera más definida la evolución de los ciclos en las curvas.

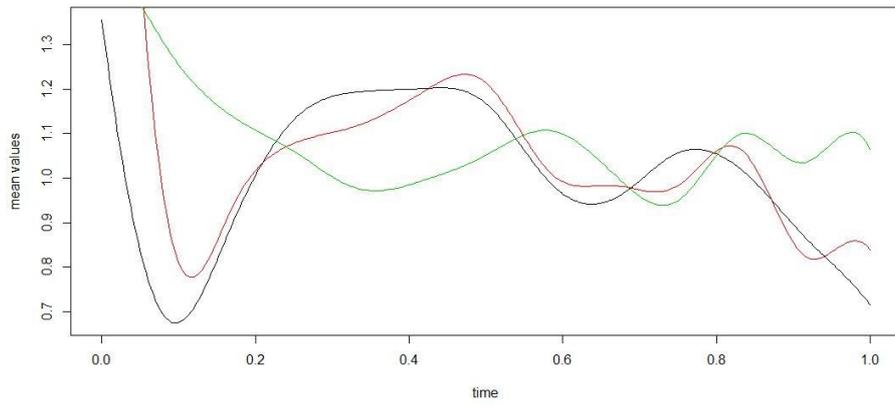
De manera más concreta, las curvas grupo N^o.1 y N^o.2 ha presentado generalmente un ratio de reproducción menor que la curva del grupo N^o.3, a excepción del intervalo [0.3,0.5]. Los países que podemos encontrar dentro del grupo N^o.3 son también los que mayor total de casos y de muertes han reportado en el mundo, como es el caso de EE.UU y Reino Unido.

Las curvas medias del grupo N^o.1 y N^o.2 presentan una evolución más cíclica que el grupo N^o.3, que sigue una evolución independiente respecto de las otras dos para el intervalo [0,0.7].

Por otra parte, si seguimos observando la evolución de estas curvas podemos ver como para el espacio [0,0.3], las curvas del N^o.1 y N^o.2 presentan una volatilidad en el ratio de reproducción más suave respecto del grupo N^o.3. Sin embargo, es a partir de aquí cuando la ciclicidad de las curvas para ambas metodologías se suavizan hasta registrar una pendiente negativa para el último tramo del periodo [0,1].

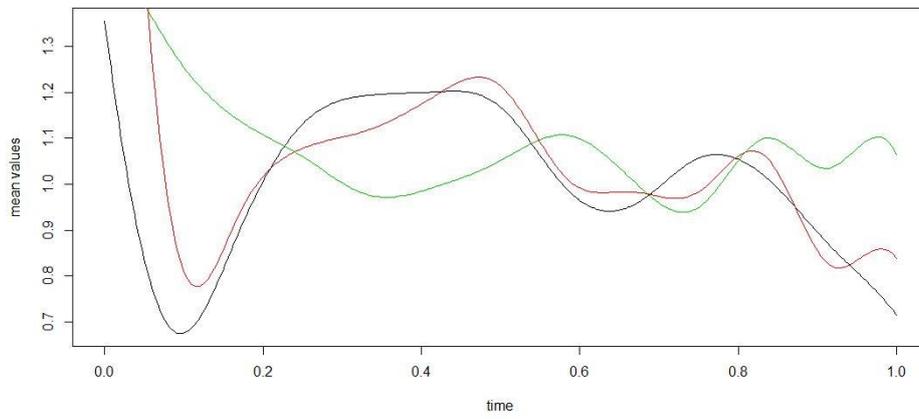
No ocurre lo mismo para el grupo N^o.3 donde la evolución no ha sido tan volátil, pero si algo más constante y para el último tramo del periodo [0,1] donde se observa que la curva evoluciona en picos de sierra.

Gráfica Nº.9: Ratio medio de reproducción para el análisis discreto.



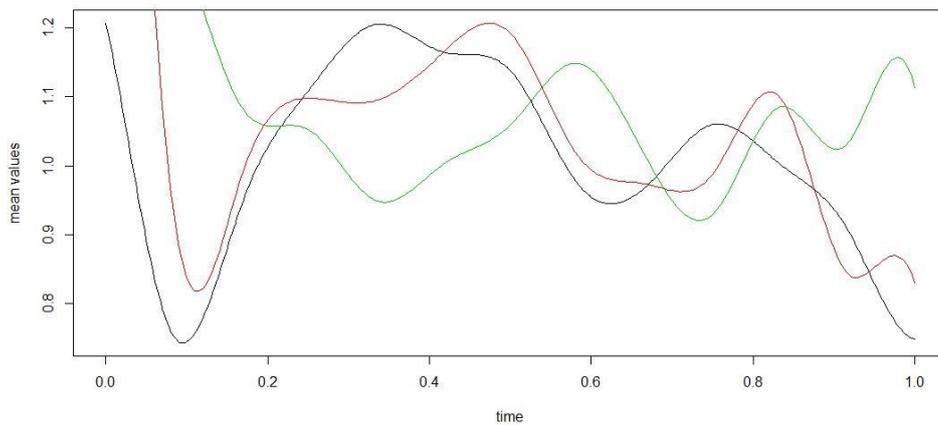
Visualización: Grupo Nº.1: Negro, Grupo Nº.1: Rojo, Grupo Nº.1: Verde

Gráfica Nº.10: Ratio de reproducción para el análisis scores



Visualización: Grupo Nº.1: Negro, Grupo Nº.1: Rojo, Grupo Nº.1: Verde

Gráfica Nº.11: Ratio de reproducción para el análisis de coeficientes



Visualización: Grupo Nº.1: Negro, Grupo Nº.1: Rojo, Grupo Nº.1: Verde

De manera general concluimos que en este capítulo hemos observado como el ratio de contagios determina el total de casos y este a su vez el total de muertes, y que detallamos en las siguientes anotaciones.

Se parte de un ratio de reproducción elevado (para cada argumento) dentro del periodo $[0.1, 0.3]$ y es a partir de este, donde vemos un aumento del total casos medios acumulados y a su vez un impacto en el total muertes acumuladas en el punto $[0.5]$ y que no logra estabilizarse la evolución de dicha curva hasta finales del periodo.

Dicho esto, de forma global vemos que para todo el intervalo $[0, 1]$ tenemos que la mayoría de las curvas medias (en cualquier grupo) suelen finalizar con una pendiente negativa (véase ratio de reproducción) o una desaceleración en sus curvas (véase total acumulado de casos y muertes). Esto puede ser un claro indicativo de que la pandemia ha perdido ritmo de crecimiento y que se está produciendo un punto de inflexión al final de periodo.

De manera más particular podemos observar que la evolución del grupo principalmente formado por países de Europa y Sudamérica registra los mayores niveles de contagios, de muertes y de ratios de reproducción.

Al grupo al que pertenecen el conjunto de microestados junto con EE.UU, también se registra una de las peores partes de la pandemia. Si a este grupo le añadimos los valores del grupo anterior, tenemos que estos países acumularían la mayor parte de la incidencia mundial.

Finalmente, destacar que el grupo principalmente formado por África y países nórdicos presentan unas curvas mucho menores y más suaves en cada una de las variables, por lo que podemos concluir que la presión que ha ejercido la pandemia sobre estos territorios ha sido mucho menor.

8. Conclusiones

Tras leer y analizar los resultados, en este capítulo resumimos algunas de las conclusiones más significativas que hemos obtenido.

Por un lado, dentro del análisis FDA hemos podido ver como las curvas medias del total acumulado de casos ha evolucionado de una manera diferente para cada uno de los continentes. Por cada 100.000 habitantes, Europa ha sido el continente con un mayor número de casos confirmados por Covid-19, mientras que África y Oceanía han sido los que menos.

El impacto de esta evolución sobre cada uno de los continentes se puede observar de manera más clara cada 14 días, donde podemos ver como las olas más grandes corresponden con aquellos continentes que mayor número de casos acumulado notifica. Destacamos la evolución de las olas en Europa donde se llegan alcanzar valores tan altos como África, Asia y Oceanía juntos.

Sin embargo, al observar que Europa registra el mayor número de contagios, esperamos que también registre el mayor número de muertes pero esto no es así. Sudamérica se convierte en el país con más muertes acumuladas por 100.000 habitantes y que podemos corroborar siguiendo la evolución de sus olas, las cuales llegan a ser mucho más agresivas que las europeas. En el lugar opuesto tenemos a África y Oceanía donde se registra una evolución mucho más suave de sus curvas medias en comparación con el resto de continentes.

Algo similar ocurre en el ratio de reproducción, las curvas que representan mayores niveles de ratio presentan corresponden a Europa y Sudamérica. A pesar de ello, otros continentes como África u Oceanía presentan valores superiores a 1 de forma constante pero no se ve reflejado en la evolución de casos acumulados por cada 100.000 habitantes (una de las razones que explicaría esto sería las distintas densidades de población así como las conexiones interterritoriales con las que cuenta cada uno de los continentes).

Por otro lado, dentro del análisis clúster hemos podido observar cómo cada uno de los países se han ido posicionando en torno a sus respectivos grupos. A grandes rasgos para cada una de las variables tenemos entre 3 y 4 grupos (total acumulado de casos por 100.000 habitantes y ratio de reproducción y total acumulado de muertes por 100.000 habitantes respectivamente).

Una de las singularidades que hemos observado en el total de casos acumulados por 100.000 habitantes es que para cada uno de los distintos tipos de análisis tenemos lo siguiente:

Un primer grupo agrupa a un determinado conjunto de microestados o países con una dimensión reducida donde junto a ellos aparece EE.UU.

Un segundo aglutina la mayoría de los países del continente europeo, así como algunas economías de categoría similar de otros continentes como es Chile.

Y un tercer grupo donde aparecen la mayoría de los países del continente Africano y asiático, aunque es interesante poder destacar que junto a ellos aparecen países del norte de Europa como puede ser Finlandia.

Del mismo modo, algo similar hemos podido observar para el total de muertes acumuladas por cada 100.000 habitantes, los países se agrupan en torno a 4 clústeres diferentes. Uno de estos grupos cuenta con un país observado, Perú. Para el resto de grupos, reúne en términos generales las mismas características (en general también los mismos países) que el total acumulado de casos por 100.000 habitantes.

De manera diferente se han agrupado los países en el ratio de reproducción, en cada uno de los distintos argumentos aplicados. La agrupación ha sido mucho más heterogénea en cuanto a que los países que representan cada grupo no reúne características socio-económicas generales. Sin embargo, esta vez aparece la mayoría de los países del norte de Europa en torno a otros de su misma condición política y geográfica (Comunidad Europea).

Por otra parte si comparamos las trayectorias de las curvas de la agrupación natural frente a las agrupadas mediante la metodología clúster podemos observar cómo para el total de casos acumulados por 100.000 habitantes se observan las mismas trayectorias, sin embargo no ocurre lo mismo para el total de muertes acumuladas por 100.000 habitantes ya que algunos países se fugan de un grupo a otro al no afrontar el tratamiento de los casos de la misma manera.

Finalmente, la conclusión que obtenemos a partir de utilizar diferentes argumentos para el análisis clúster es que los resultados varían entre unos y otros, a pesar de que los grupos cuentan casi con los mismos países, estos no son del todo iguales y por tanto nuestras interpretaciones pueden variar.

A pesar de todo, cualquiera de los métodos utilizados en el análisis es válido. Cada uno utiliza una técnica y una base teórica diferente, sin dejar atrás la funcionalidad a la hora de obtener los resultados.

9. Referencias

- Abraham, C., Cornillon, P.A., Matzner-Løber, E.R.I.C. and Molinari, N., 2003. Unsupervised curve clustering using B-splines. *Scandinavian journal of statistics*, 30(3), pp.581-595.
- Abramowitz, M. and Stegun, I. (1968). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, pp. 882-884
- Aguilera, A.M., Gutiérrez, R., Ocaña, F.A. and Valderrama, M.J., 1995. Computational approaches to estimation in the principal component analysis of a stochastic process. *Applied Stochastic Models and Data Analysis*, 11(4), pp.279-299.
- Aguilera Morillo, M., 2009. Estimación Penalizada con Datos Funcionales. Master. Universidad de Granada.
- Bárceñas Curtis, R., 2014. Estudio de Perfiles de Olas vía Análisis de Datos Funcionales. Tesis. Centro de Investigación en Matemáticas, A.C.
- Berrendero, J.R., Justel, A. and Svarc, M., 2011. Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9), pp.2619-2634.
- De Boor, C., 1977. Package for calculating with B-splines. *SIAM Journal on Numerical Analysis*, 14(3), pp.441-472.
- Ferraty, F. and Vieu, P., 2006. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.
- Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21, pp.768-769
- González, O., 2016. *Modelización de curvas de rentabilidad del IBEX 35 basada en Análisis en componentes principales funcional*. Master. Universidad de Granada.
- Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T. and Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23), pp.4121-4123.
- Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), pp.100-108.
- Jacques, J. and Preda, C., 2014. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3), pp.231-255.
- Lloyd, S., 1982. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), pp.129-137.
- MacQueen, J., 1967, June. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).
- Mirkin, B. (2012). *Clustering: A data Recovery Approach*. Londres: Department of Computer Science and Information Systems Birkbeck University of London, UK.
- Ramsay, J.O. and Silverman, B.W., 2002. *Applied functional data analysis: methods and case studies* (Vol. 77, pp. 115-130). New York: Springer.
- Ramsay, J.O. and Silverman, B.W., 2005. Principal components analysis for functional data. *Functional data analysis*, pp.147-172.

Ramsay, J.O., Wickham, H., Graves, S. and Hooker, G., 2014. fda: Functional data analysis. *R package version*, 2(4), p.142 <http://CRAN.R-project.org/package=fda>

Rdocumentation.org. 2021. eval.fd function - RDocumentation. [online] Available at: <https://www.rdocumentation.org/packages/fda/versions/5.1.9/topics/eval.fd> [Accessed 30 August 2021].

Rdocumentation.org. 2021. pca.fd function - RDocumentation. [online] Available at: <https://www.rdocumentation.org/packages/fda/versions/5.1.9/topics/pca.fd> [Accessed 30 August 2021].

Role, F. and Nadif, M., 2011, October. Handling the impact of low frequency events on co-occurrence based measures of word similarity. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR-2011)*. Scitepress (pp. 218-223).

Salvatore S, Bramness JG, Reid MJ, Thomas KV, Harman C, Røislien J (2015) Wastewater-Based Epidemiology of Stimulant Drugs: Functional Data Analysis Compared to Traditional Statistical Methods. *PLoS ONE* 10(9): e0138669. <https://doi.org/10.1371/journal.pone.0138669>

Williams, W.T., Dale, M.B. and Macnaughton-Smith, P., 1964. An objective method of weighting in similarity analysis. *Nature*, 201(4917), pp.426-426

10. Anexo

```
WorldCovid <- read.csv("owid-covid-data.csv")

#ANÁLISIS FDA.

WorldStatistic<-
subset(WorldCovid,location==names(table(WorldCovid$location))[1])[1,c("iso_code","contine
nt","location","population", "population_density", "median_age", "aged_65_older",
"aged_70_older", "gdp_per_capita", "extreme_poverty", "cardiovasc_death_rate",
"diabetes_prevalence", "female_smokers", "male_smokers", "handwashing_facilities",
"hospital_beds_per_thousand", "life_expectancy", "human_development_index")]

for (i in 2:length(table(WorldCovid$location))){

  WorldStatistic<-
  rbind(WorldStatistic,subset(WorldCovid,location==names(table(WorldCovid$location))[i])[1,c("
iso_code","continent","location","population", "population_density", "median_age",
"aged_65_older", "aged_70_older", "gdp_per_capita", "extreme_poverty",
"cardiovasc_death_rate", "diabetes_prevalence", "female_smokers", "male_smokers",
"handwashing_facilities", "hospital_beds_per_thousand", "life_expectancy",
"human_development_index"])]

}

CountriesCovid<-
subset(WorldCovid,iso_code!="OWID_AFR"&iso_code!="OWID_ASI"&iso_code!="OWID_CYN"
&iso_code!="OWID_EUR"&iso_code!="OWID_EUN"&iso_code!="
"OWID_INT"&iso_code!="OWID_NAM"&iso_code!="OWID_OCE"&iso_code!="OWID_SAM"&is
o_code!="OWID_WRL")

CountriesCovid$iso_code<-as.factor(as.character(CountriesCovid$iso_code))

CountriesCovid$continent<-as.factor(as.character(CountriesCovid$continent))

CountriesCovid$location<-as.factor(as.character(CountriesCovid$location))

CountriesStatistic<-
subset(WorldStatistic,iso_code!="OWID_AFR"&iso_code!="OWID_ASI"&iso_code!="OWID_CY
N"&iso_code!="OWID_EUR"&iso_code!="OWID_EUN"&iso_code!="
"OWID_INT"&iso_code!="OWID_NAM"&iso_code!="OWID_OCE"&iso_code!="OWID_SAM"&is
o_code!="OWID_WRL")
```

```

CountriesStatistic$iso_code<-as.factor(as.character(CountriesStatistic$iso_code))

CountriesStatistic$continent<-as.factor(as.character(CountriesStatistic$continent))

CountriesStatistic$location<-as.factor(as.character(CountriesStatistic$location))

#Variables Funciónales a analizar

#"total_cases"

#"total_deaths"

#"reproduction_rate"

#"positive_rate"

#"total_vaccinations"

#"people_vaccinated"

#"people_fully_vaccinated"

#Nos quedamos con las variables que podemos analizar.

CountriesCovid<-
CountriesCovid[,c("iso_code","continent","location","date","total_cases","total_deaths","reproduction_rate","icu_patients","hosp_patients","total_tests","positive_rate","total_vaccinations","people_vaccinated","people_fully_vaccinated","population")]

sort(table(CountriesCovid$location))

#Eliminamos lo países que no nos interesan

CountriesCovid<-
subset(CountriesCovid,location!="Greenland"&location!="Montserrat"&location!="Saint
Helena"&location!="Anguilla"&location!="Falkland
Islands"&location!="Macao"&location!="Faeroe
Islands"&location!="Guernsey"&location!="Jersey"&location!="Turks and Caicos
Islands"&location!="Isle of Man"&location!="Micronesia
(country)"&location!="Gibraltar"&location!="Bermuda"&location!="Cayman
Islands"&location!="Hong
Kong"&location!="Samoa"&location!="Vanuatu"&location!="Marshall
Islands"&location!="Solomon
Islands"&location!="Vatican"&location!="Tonga"&location!="Saint Kitts and

```

Nevis"&location!="Nauru"&location!="Laos"&location!="Dominica"&location!="Curacao"&location!="Aruba"&location!="Fiji"&location!="Benin"&location!="Burkina Faso"&location!="Burundi"&location!="Central African Republic"&location!="Chad"&location!="Comoros"&location!="Cuba"&location!="Democratic Republic of Congo"&location!="Djibouti"&location!="Eritrea"&location!="Guinea-Bissau"&location!="Haiti"&location!="Liberia"&location!="Libya"&location!="Madagascar"&location!="Somalia"&location!="Tajikistan"&location!="Tajikistan"&location!="Tanzania"&location!="Yemen"&location!="Saint Vincent and the Grenadines"&location!="Algeria"&location!="Bosnia and Herzegovina"&location!="Ethiopia"&location!="Mongolia"&location!="China"&location!="Saudi Arabia"&location!="Angola"&location!="Antigua and Barbuda"&location!="Azerbaijan"&location!="Bahamas"&location!="Barbados"&location!="Belize"&location!="Botswana"&location!="Brazil"&location!="Brunei"&location!="Cambodia"&location!="Cameroon"&location!="Congo"&location!="Egypt"&location!="Equatorial Guinea"&location!="Eswatini"&location!="Gabon"&location!="Georgia"&location!="Grenada"&location!="Guinea"&location!="Guyana"&location!="Honduras"&location!="Kosovo"&location!="Kyrgyzstan"&location!="Lebanon"&location!="Lesotho"&location!="Mali"&location!="Mauritania"&location!="Mauritius"&location!="Moldova"&location!="Monaco"&location!="Montenegro"&location!="Nicaragua"&location!="Niger"&location!="Papua New Guinea"&location!="Saint Lucia"&location!="San Marino"&location!="Sao Tome and Principe"&location!="Seychelles"&location!="Sierra Leone"&location!="Sudan"&location!="Suriname"&location!="Syria"&location!="Timor"&location!="Uzbekistan"&location!="Venezuela"&location!="Bonaire Sint Eustatius and Saba"&location!="British Virgin Islands"&location!="Cook Islands"&location!="French Polynesia"&location!="New Caledonia"&location!="Pitcairn"&location!="Sint Maarten (Dutch part)"&location!="Turkmenistan"&location!="Tuvalu"&location!="Wallis and Futuna"&location!="Kiribati"&location!="Armenia"&location!="Cote d'Ivoire"&location!="Bhutan")

```
CountriesCovid$iso_code<-as.factor(as.character(CountriesCovid$iso_code))
```

```
CountriesCovid$continent<-as.factor(as.character(CountriesCovid$continent))
```

```
CountriesCovid$location<-as.factor(as.character(CountriesCovid$location))
```

```
CountriesCovid$date<-as.Date(CountriesCovid$date)
```

```
#Si quisiera crear un vector con todas las fechas entre la primera y la última:
```

```

AllDays<-
seq(as.Date(names(table(CountriesCovid$date))[1]),as.Date(names(table(CountriesCovid$date
))[length(table(CountriesCovid$date))]),by="1 day")

ParaMerge<-
data.frame(rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),AllDays,rep
p(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),
rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays
)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)))

names(ParaMerge)<-names(CountriesCovid)

#creamos una lista

CountriesList<-list()

for (i in 1:length(table(CountriesCovid$location))){CountriesList[[i]]<-
subset(CountriesCovid,location==names(table(CountriesCovid$location))[i])}

names(CountriesList)<-names(table(CountriesCovid$location))

for (i in 1:length(CountriesList)){

CountriesList[[i]]<-merge(CountriesList[[i]],ParaMerge,by="date",all=T)

CountriesList[[i]]<-CountriesList[[i]][,1:16]

CountriesList[[i]][,16]<-c(1:length(AllDays))

names(CountriesList[[i]])<-
c("date","iso_code","continent","location","total_cases","total_deaths","reproduction_rate","i
cu_patients","hosp_patients","total_tests","positive_rate","total_vaccinations","people_vacci
nated","people_fully_vaccinated","population","times")

}

dim(CountriesList[[1]])

#Cálculo de las variables Funcionales.

#Total de casos

CountriesCases<-list()

CountriesCases<-list()

```

```

for (i in 1:length(table(CountriesCovid$location))){CountriesCases[[i]]<-
subset(CountriesCovid,location==names(table(CountriesCovid$location))[i])[,c("date","total_ca
ses")]}}

names(CountriesCases)<-names(table(CountriesCovid$location))

#Total de muertes

total_deaths<-list()

for (i in 1:length(table(CountriesCovid$location))){total_deaths[[i]]<-
subset(CountriesCovid,location==names(table(CountriesCovid$location))[i])[,c("date","total_d
eaths")]}}

names(total_deaths)<-names(table(CountriesCovid$location))

#Tasa de reproducción

reproduction_rate<-list()

for (i in 1:length(table(CountriesCovid$location))){reproduction_rate[[i]]<-
subset(CountriesCovid,location==names(table(CountriesCovid$location))[i])[,c("date","reprodu
ction_rate")]}}

names(reproduction_rate)<-names(table(CountriesCovid$location))

#Comenzamos a crear nuestros datos Funcionales e instalamos la librería FDA

library(fda)

#Hemos acordado una base de 15

Base15<-create.bspline.basis(rangeval=c(0,1),nbasis = 15)

#Continentes

#Asia

Asia<-sort(c("Thailand","Taiwan","South
Korea","Japan","Singapore","Vietnam","Malaysia","Nepal","Sri Lanka","United Arab
Emirates","India","Philippines","Iran","Israel","Afghanistan","Bahrain","Iraq","Kuwait","Oman"
,"Pakistan","Qatar","Indonesia","Bangladesh","Jordan","Palestine","Maldives","Turkey","Kazak
hstan","Myanmar"))

#FDA de ratio de reproducción

```

```

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","reproduction_rate")]),c("times
","reproduction_rate")]

TemporalFrame

for (i in 2:length(CountriesList)){print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]][,c("tim
es","reproduction_rate")]),c("times","reproduction_rate"))))}}

#reproduction_rate.Coefs<-Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs

reproduction_rate.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs

for (i in 2 :length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","reproduction_rate")]),c("times",
"reproduction_rate")]

  #reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs)

  reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs)

}

reproduction_rate.Coefs

colnames(reproduction_rate.Coefs)<-names(reproduction_rate)

Asiareproduction_rate.fd<-fd(reproduction_rate.Coefs[,Asia],basisobj = Base15)

plot(mean(Asiareproduction_rate.fd))

Asiareproduction_rate.fd<-mean(Asiareproduction_rate.fd)

#FDA de Casos por cada 100.000 habitantes

```

```

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population")]),c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  CasesPor100.Coefs<-cbind(CasesPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.Coefs)<-names(CountriesCases)

AsiaCasesPor100.fd<-fd(CasesPor100.Coefs[,Asia],basisobj = Base15)

plot(AsiaCasesPor100.fd)

#Total Muertes por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population")]),c(
("times","total_deaths","population"))]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

```

```

total_deathsPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

Asiatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,Asia],basisobj = Base15)

plot(Asiatotal_deathsPor100.fd)

#FDA de Casos por por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times", "total_cases", "population")]),c("t
imes", "total_cases", "population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)
}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,Asia],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

asiaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#Total de muertes por cada por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times", "total_deaths", "population")]),c
("times", "total_deaths", "population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","total_deaths","population"))],c(
"times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)
}

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,Asia],basisobj = Base15)

plot(total_deathsPor100.14Dias.fd)

asiatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

#Africa

Africa<-sort(c("Morocco","South
Africa","Nigeria","Senegal","Togo","Tunisia","Kenya","Ghana","Namibia","Rwanda","Gambia",
"Zambia","Cape Verde","Zimbabwe","Uganda","Mozambique","Malawi","South Sudan"))

#FDA de ratio de reproducción

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","reproduction_rate"))],c("times
","reproduction_rate")]

TemporalFrame

for (i in 2:length(CountriesList)){print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]],c("tim
es","reproduction_rate"))],c("times","reproduction_rate"))))}}

#reproduction_rate.Coefs<-Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs

```

```

reproduction_rate.Coeffs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coeffs

for (i in 2 :length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","reproduction_rate"))],c("times",
"reproduction_rate")]

  #reproduction_rate.Coeffs<-cbind(reproduction_rate.Coeffs,Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coeffs)

  reproduction_rate.Coeffs<-cbind(reproduction_rate.Coeffs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coeffs)

}

colnames(reproduction_rate.Coeffs)<-names(reproduction_rate)

Africareproduction_rate.fd<-fd(reproduction_rate.Coeffs[,Africa],basisobj = Base15)

plot(Africareproduction_rate.fd)

Africareproduction_rate.fd<-mean(Africareproduction_rate.fd)

#Total de casos por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","total_cases","population"))],c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coeffs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coeffs

for (i in 2:length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times", "total_cases", "population")]),c("t
imes", "total_cases", "population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-cbind(CasesPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)
}

colnames(CasesPor100.Coefs)<-names(CountriesCases)

AfricaCasesPor100.fd<-fd(CasesPor100.Coefs[,Africa],basisobj = Base15)

plot(AfricaCasesPor100.fd)

#Total de muertes por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times", "total_deaths", "population")]),c(
("times", "total_deaths", "population"))]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

total_deathsPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times", "total_deaths", "population")]),c(
"times", "total_deaths", "population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

```

```

total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

Africatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,Africa],basisobj = Base15)

plot(Africatotal_deathsPor100.fd)

#Total de casos por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","total_cases","population"))],c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","total_cases","population"))],c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

```

```

}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,Africa],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

africaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#Total de muertes por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

```

```

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,Africa],basisobj = Base15)

plot(total_deathsPor100.14Dias.fd)

africatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

#Europa

Europe<-sort(c("France","Germany","Finland","Italy","Russia","United
Kingdom","Spain","Sweden","Denmark","Slovenia","Belgium","Malta","Greece","Norway","Lu
xembourg","Netherlands","Albania","Austria","Croatia","Estonia","Ireland","Latvia","Switzerla
nd","Czechia","Iceland","North
Macedonia","Portugal","Romania","Serbia","Belarus","Lithuania","Andorra","Ukraine","Hunga
ry","Liechtenstein","Poland","Slovakia","Bulgaria","Cyprus"))

#FDA de ratio de reproducción.

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","reproduction_rate"))],c("times
","reproduction_rate")]

TemporalFrame
for (i in 2:length(CountriesList))
2:length(CountriesList){print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]][c("tim
es","reproduction_rate"))]),c("times","reproduction_rate"))))}

reproduction_rate.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs

for (i in 2:length(CountriesList)){

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","reproduction_rate"))],c("times",
"reproduction_rate")]

reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs)

```

```

}

colnames(reproduction_rate.Coeffs)<-names(reproduction_rate)

Europareproduction_rate.fd<-fd(reproduction_rate.Coeffs[,Europe],basisobj = Base15)

plot(Europareproduction_rate.fd)

Europareproduction_rate.fd<-mean(Europareproduction_rate.fd)

#FDA de casos por cada 100000 hab.

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coeffs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coeffs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population")]),c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  CasesPor100.Coeffs<-cbind(CasesPor100.Coeffs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coeffs)

}

colnames(CasesPor100.Coeffs)<-names(CountriesCases)

EuropaCasesPor100.fd<-fd(CasesPor100.Coeffs[,Europe],basisobj = Base15)

plot(EuropaCasesPor100.fd)

#Total de muertes por cada 100.000 habitantes

```

```

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

total_deathsPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

Europatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,Europe],basisobj = Base15)

plot(Europatotal_deathsPor100.fd)

#Total de casos por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

```

```

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population"))],c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,Europe],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

europaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#Total de muertes por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population"))],c(
("times","total_deaths","population"))

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

```

```

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","total_deaths","population"))],c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)
}

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,Europe],basisobj =
Base15)

plot(total_deathsPor100.14Dias.fd)

europatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

#Norteamérica

NorthAmerica<-sort(c("Mexico","United States","Canada","Guatemala","Dominican
Republic","Costa Rica","Panama","Jamaica","Trinidad and Tobago","El Salvador"))

#FDA de ratio de reproducción

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","reproduction_rate"))],c("times
","reproduction_rate")]

TemporalFrame

```

```

for (i in 2:length(CountriesList)){
  print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]]),c("times",
"reproduction_rate"))),c("times","reproduction_rate"))))}

#reproduction_rate.Coefs<-Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs

reproduction_rate.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs

for (i in 2 :length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","reproduction_rate")),c("times",
"reproduction_rate")]

  #reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs)

  reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs)

}

reproduction_rate.Coefs

colnames(reproduction_rate.Coefs)<-names(reproduction_rate)

Norteamericareproduction_rate.fd<-fd(reproduction_rate.Coefs[,NorthAmerica],basisobj =
Base15)

plot(Norteamericareproduction_rate.fd)

Norteamericareproduction_rate.fd<-mean(Norteamericareproduction_rate.fd)

#Total de casos por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","total_cases","population")),c(
"times","total_cases","population")]

```

```

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-Data2fd(argvals                                     =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population")]),c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  CasesPor100.Coefs<-cbind(CasesPor100.Coefs,Data2fd(argvals                                     =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.Coefs)<-names(CountriesCases)

NorteamericaCasesPor100.fd<-fd(CasesPor100.Coefs[,NorthAmerica],basisobj = Base15)

plot(NorteamericaCasesPor100.fd)

#Total de muertes por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

total_deathsPor100.Coefs<-Data2fd(argvals                                     =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)
}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

Norteamericatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,NorthAmerica],basisobj =
Base15)

plot(Norteamericatotal_deathsPor100.fd)

#Total de casos por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population")]),c("t
imes","total_cases","population")]

```

```

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,NorthAmerica],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

namericaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#total de muertes por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

```

```

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)
}

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,NorthAmerica],basisobj =
Base15)

plot(total_deathsPor100.14Dias.fd)

namericatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

#Sudamerica

SouthAmerica<-
sort(c("Argentina","Peru","Chile","Ecuador","Colombia","Paraguay","Bolivia","Uruguay"))

#FDA de ratio de reproducción

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","reproduction_rate"))],c("times
","reproduction_rate")]

TemporalFrame

for (i in 2:length(CountriesList))
  {print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]][c("times",
"reproduction_rate"))]),c("times","reproduction_rate"))))}

#reproduction_rate.Coefs<-Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs

reproduction_rate.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs

for (i in 2 :length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","reproduction_rate")),c("times",
"reproduction_rate")]

#reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals           =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs)

reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals           =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs)

}

reproduction_rate.Coefs

colnames(reproduction_rate.Coefs)<-names(reproduction_rate)

Sudamericareproduction_rate.fd<-fd(reproduction_rate.Coefs[,SouthAmerica],basisobj   =
Base15)

plot(Sudamericareproduction_rate.fd)

Sudamericareproduction_rate.fd<-mean(Sudamericareproduction_rate.fd)

#Total de casos por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","total_cases","population")),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-Data2fd(argvals           =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj   =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
  CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","total_cases","population")),c("t
imes","total_cases","population")]

```

```

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-cbind(CasesPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.Coefs)<-names(CountriesCases)

SudamericaCasesPor100.fd<-fd(CasesPor100.Coefs[,SouthAmerica],basisobj = Base15)

plot(SudamericaCasesPor100.fd)

#Total de muertes por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","total_deaths","population"))],c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

total_deathsPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","total_deaths","population"))],c
("times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

```

```

Sudamericatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,SouthAmerica],basisobj =
Base15)

plot(Sudamericatotal_deathsPor100.fd)

#Total de casos por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_cases","population")]),c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,SouthAmerica],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

```

```

samericaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#Total de muertes por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_deaths","population")]),c
("times","total_deaths","population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

  total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,SouthAmerica],basisobj =
Base15)

plot(total_deathsPor100.14Dias.fd)

samericatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

```

```

#Oceanía

Oceania<-sort(c("Australia","New Zealand"))

#FDA de ratio de reproducción.

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","reproduction_rate"))],c("times
","reproduction_rate")]

TemporalFrame

for (i in 2:length(CountriesList)){
  print(c(i,dim(CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","reproduction_rate"))]),c("times","reproduction_rate"))))}

#reproduction_rate.Coefs<-Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs

reproduction_rate.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs

for (i in 2 :length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","reproduction_rate"))],c("times",
"reproduction_rate")]

  #reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
TemporalFrame$times,y=TemporalFrame$reproduction_rate,basisobj = Base15)$coefs)

  reproduction_rate.Coefs<-cbind(reproduction_rate.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15)$coefs)

}

colnames(reproduction_rate.Coefs)<-names(reproduction_rate)

Oceaniareproduction_rate.fd<-fd(reproduction_rate.Coefs[,Oceania],basisobj = Base15)

plot(Oceaniareproduction_rate.fd)

```

```

Oceaniareproduction_rate.fd<-mean(Oceaniareproduction_rate.fd)

#Total de casos por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","total_cases","population"))],c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

CasesPor100.Coefs<-Data2fd(argvals                                     =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][c("times","total_cases","population"))],c("t
imes","total_cases","population")]

  TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

  CasesPor100.Coefs<-cbind(CasesPor100.Coefs,Data2fd(argvals                                     =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs)

}

colnames(CasesPor100.Coefs)<-names(CountriesCases)

OceaniaCasesPor100.fd<-fd(CasesPor100.Coefs[,Oceania],basisobj = Base15)

plot(OceaniaCasesPor100.fd)

#Total de muertes por cada 100.000 habitantes

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times","total_deaths","population"))],c(
("times","total_deaths","population"))

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

```

```

total_deathsPor100.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times","total_deaths","population")]),c(
"times","total_deaths","population")]

  TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

  total_deathsPor100.Coefs<-cbind(total_deathsPor100.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)

}

colnames(total_deathsPor100.Coefs)<-names(total_deaths)

Oceaniatotal_deathsPor100.fd<-fd(total_deathsPor100.Coefs[,Oceania],basisobj = Base15)

plot(Oceaniatotal_deathsPor100.fd)

#Total de casos por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times","total_cases","population")]),c(
"times","total_cases","population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj =
Base15)$coefs

for (i in 2:length(CountriesList)){

```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times", "total_cases", "population")]),c("t
imes", "total_cases", "population")]

TemporalFrame$total_cases<-
100000*TemporalFrame$total_cases/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

CasesPor100.14Dias.Coefs<-cbind(CasesPor100.14Dias.Coefs,Data2fd(argvals           =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_cases,basisobj   =
Base15)$coefs)}

colnames(CasesPor100.14Dias.Coefs)<-names(CountriesCases)

CasesPor100.14Dias.fd<-fd(CasesPor100.14Dias.Coefs[,Oceania],basisobj = Base15)

plot(CasesPor100.14Dias.fd)

oceaniaCasesPor100.14Dias.fd<-mean(CasesPor100.14Dias.fd)

#Total de muertes por cada 100.000 habitantes a 14 días

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]][,c("times", "total_deaths", "population")]),c
("times", "total_deaths", "population")]

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-Data2fd(argvals           =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj   =
Base15)$coefs

for (i in 2:length(CountriesList)){

  TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]][,c("times", "total_deaths", "population")]),c(
"times", "total_deaths", "population")]

```

```

TemporalFrame$total_deaths<-
100000*TemporalFrame$total_deaths/TemporalFrame$population

TemporalFrame<-TemporalFrame[15:nrow(TemporalFrame),]-
TemporalFrame[1:(nrow(TemporalFrame)-14),]

total_deathsPor100.14Dias.Coefs<-cbind(total_deathsPor100.14Dias.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$total_deaths,basisobj =
Base15)$coefs)}

colnames(total_deathsPor100.14Dias.Coefs)<-names(total_deaths)

total_deathsPor100.14Dias.fd<-fd(total_deathsPor100.14Dias.Coefs[,Oceania],basisobj =
Base15)

plot(total_deathsPor100.14Dias.fd)

oceaniatotal_deathsPor100.14Dias.fd<-mean(total_deathsPor100.14Dias.fd)

#Representación de las curvas medias

#Casos acumulados por 100.000 habitantes

plot(mean(EuropaCasesPor100.fd), col=3)

lines(mean(AfricaCasesPor100.fd), col=2)

lines(mean(AsiaCasesPor100.fd), col=1)

lines(mean(NorteamericaCasesPor100.fd), col=4)

lines(mean(SudamericaCasesPor100.fd), col=5)

lines(mean(OceaniaCasesPor100.fd), col=6)

#Casos acumulados por 100.000 habitantes a 14 días.

plot(europaCasesPor100.14Dias.fd, col=3)

lines(africaCasesPor100.14Dias.fd, col=2)

lines(asiaCasesPor100.14Dias.fd, col=1)

lines(namericaCasesPor100.14Dias.fd, col=4)

lines(samericaCasesPor100.14Dias.fd, col=5)

```

```

lines(oceaniaCasesPor100.14Dias.fd, col=6)

#Muertes acumuladas por 100.000 habitantes

plot(mean(Sudamericatotal_deathsPor100.fd), col=5)

lines(mean(Africatotal_deathsPor100.fd), col=2)

lines(mean(Asiatotal_deathsPor100.fd))

lines(mean(Norteamericatotal_deathsPor100.fd), col=4)

lines(mean(Europatotal_deathsPor100.fd), col=3)

lines(mean(Oceaniatotal_deathsPor100.fd), col=6)

#Muertes acumuladas por 100.000 habitantes a 14 días.

plot(samericatotal_deathsPor100.14Dias.fd, col=5)

lines(africatotal_deathsPor100.14Dias.fd, col=2)

lines(asiatotal_deathsPor100.14Dias.fd)

lines(namericatotal_deathsPor100.14Dias.fd, col=4)

lines(europatotal_deathsPor100.14Dias.fd, col=3)

lines(oceaniatotal_deathsPor100.14Dias.fd, col=6)

#Ratio de reproducción

plot(Oceaniareproduction_rate.fd)

lines(Asiareproduction_rate.fd, col=4)

lines(Europareproduction_rate.fd, col=6)

lines(Norteamericareproduction_rate.fd, col=3)

lines(Sudamericareproduction_rate.fd, col=5)

lines(Africareproduction_rate.fd, col=2)

#ANÁLISIS CLÚSTER

WorldCovid <- read.csv("owid-covid-data.csv")

WorldCovid$iso_code<-as.character(WorldCovid$iso_code)

```

```

WorldCovid$continent<-as.character(WorldCovid$continent)

WorldCovid$location<-as.character(WorldCovid$location)

CountriesCovid<-
subset(WorldCovid,iso_code!="OWID_AFR"&iso_code!="OWID_ASI"&iso_code!="OWID_CYN"
&iso_code!="OWID_EUR"&iso_code!="OWID_EUN"&iso_code!="
"OWID_INT"&iso_code!="OWID_NAM"&iso_code!="OWID_OCE"&iso_code!="OWID_SAM"&is
o_code!="OWID_WRL")

#Variables funcionales a analizar

#"total_cases"

#"total_deaths"

#"reproduction_rate"

CountriesCovid<-
CountriesCovid[,c("iso_code","continent","location","date","total_cases","total_deaths","repr
oduction_rate","population")]

#Y eliminamos paises que no son objeto de estudio.

sort(table(CountriesCovid$location))

CountriesCovid<-
subset(CountriesCovid,location!="Greenland"&location!="Montserrat"&location!="Saint
Helena"&location!="Anguilla"&location!="Falkland
Islands"&location!="Macao"&location!="Faeroe
Islands"&location!="Guernsey"&location!="Jersey"&location!="Turks and Caicos
Islands"&location!="Isle of Man"&location!="Micronesia
(country)"&location!="Gibraltar"&location!="Bermuda"&location!="Cayman
Islands"&location!="Hong
Kong"&location!="Samoa"&location!="Vanuatu"&location!="Marshall
Islands"&location!="Solomon
Islands"&location!="Vatican"&location!="Tonga"&location!="Saint Kitts and
Nevis"&location!="Nauru"&location!="Laos"&location!="Dominica"&location!="Curacao"&loc
ation!="Aruba"&location!="Fiji"&location!="Benin"&location!="Burkina
Faso"&location!="Burundi"&location!="Central African
Republic"&location!="Chad"&location!="Comoros"&location!="Cuba"&location!="Democratic

```

Republic of Congo"&location!="Djibouti"&location!="Eritrea"&location!="Guinea-Bissau"&location!="Haiti"&location!="Liberia"&location!="Libya"&location!="Madagascar"&location!="Somalia"&location!="Tajikistan"&location!="Tajikistan"&location!="Tanzania"&location!="Yemen"&location!="Saint Vincent and the Grenadines"&location!="Algeria"&location!="Bosnia and Herzegovina"&location!="Ethiopia"&location!="Mongolia"&location!="China"&location!="Saudi Arabia"&location!="Angola"&location!="Antigua and Barbuda"&location!="Azerbaijan"&location!="Bahamas"&location!="Barbados"&location!="Belize"&location!="Botswana"&location!="Brazil"&location!="Brunei"&location!="Cambodia"&location!="Cameroon"&location!="Congo"&location!="Egypt"&location!="Equatorial Guinea"&location!="Eswatini"&location!="Gabon"&location!="Georgia"&location!="Grenada"&location!="Guinea"&location!="Guyana"&location!="Honduras"&location!="Kosovo"&location!="Kyrgyzstan"&location!="Lebanon"&location!="Lesotho"&location!="Mali"&location!="Mauritania"&location!="Mauritius"&location!="Moldova"&location!="Monaco"&location!="Montenegro"&location!="Nicaragua"&location!="Niger"&location!="Papua New Guinea"&location!="Saint Lucia"&location!="San Marino"&location!="Sao Tome and Principe"&location!="Seychelles"&location!="Sierra Leone"&location!="Sudan"&location!="Suriname"&location!="Syria"&location!="Timor"&location!="Uzbekistan"&location!="Venezuela"&location!="Bonaire Sint Eustatius and Saba"&location!="British Virgin Islands"&location!="Cook Islands"&location!="French Polynesia"&location!="New Caledonia"&location!="Pitcairn"&location!="Sint Maarten (Dutch part)"&location!="Turkmenistan"&location!="Tuvalu"&location!="Wallis and Futuna"&location!="Kiribati"&location!="Armenia"&location!="Cote d'Ivoire"&location!="Bhutan")

CountriesCovid\$cases_100<-100000*CountriesCovid\$total_cases/CountriesCovid\$population

CountriesCovid\$deaths_100<-

100000*CountriesCovid\$total_deaths/CountriesCovid\$population

Africa<-names(table(subset(CountriesCovid,continent=="Africa")\$location))

Asia<-names(table(subset(CountriesCovid,continent=="Asia")\$location))

Europa<-names(table(subset(CountriesCovid,continent=="Europe")\$location))

AmericaNorte<-names(table(subset(CountriesCovid,continent=="North America")\$location))

Oceania<-names(table(subset(CountriesCovid,continent=="Oceania")\$location))

```

AmericaSur<-names(table(subset(CountriesCovid,continent=="South America")$location))

CountriesCovid$date<-as.Date(CountriesCovid$date)

#Si quisiera crear un vector con todas las fechas entre la primera y la última:

AllDays<-
seq(as.Date(names(table(CountriesCovid$date))[1]),as.Date(names(table(CountriesCovid$date
))[length(table(CountriesCovid$date))]),by="1 day")

ParaMerge<-
data.frame(rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),AllDays,rep
p(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),rep(NA,length(AllDays)),
rep(NA,length(AllDays)),rep(NA,length(AllDays)))

names(ParaMerge)<-names(CountriesCovid)

CountriesList<-list()

for (i in 1:length(table(CountriesCovid$location))){CountriesList[[i]]<-
subset(CountriesCovid,location==names(table(CountriesCovid$location))[i])}

names(CountriesList)<-names(table(CountriesCovid$location))

for (i in 1:length(CountriesList)){

  CountriesList[[i]]<-merge(CountriesList[[i]],ParaMerge,by="date",all=T)

  CountriesList[[i]]<-CountriesList[[i]][,1:11]

  CountriesList[[i]][,11]<-c(1:length(AllDays))

  names(CountriesList[[i]])<-
c("date","iso_code","continent","location","total_cases","total_deaths","reproduction_rate","
population","cases_100","deaths_100","times")

}

dim(CountriesList[[1]])

library(fda)

#Base15<-create.bspline.basis(rangeval = c(1,length(AllDays)),nbasis = 15)

Base15bs<-create.bspline.basis(rangeval=c(0,1),nbasis = 15)

```

```
#FDA de Casos por 100000 habitantes.
```

```
TemporalFrame<-
```

```
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times", "cases_100"))],c("times", "cases_100")]
```

```
Cases.Coefs<-Data2fd(argvals =  
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$cases_100,basisobj =  
Base15bs)$coefs
```

```
Cases.Coefs
```

```
for (i in 2:length(CountriesList)){
```

```
  TemporalFrame<-
```

```
  CountriesList[[i]][complete.cases(CountriesList[[i]][c("times", "cases_100"))],c("times", "cases_100")]
```

```
  Cases.Coefs<-cbind(Cases.Coefs,Data2fd(argvals =  
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$cases_100,basisobj =  
Base15bs)$coefs)
```

```
}
```

```
colnames(Cases.Coefs)<-names(CountriesList)
```

```
Cases.fd<-fd(Cases.Coefs,basisobj = Base15bs)
```

```
Cases.fd
```

```
#FDA de Muertes por 100000 habitantes.
```

```
TemporalFrame<-
```

```
CountriesList[[1]][complete.cases(CountriesList[[1]][c("times", "deaths_100"))],c("times", "deaths_100")]
```

```
deaths.Coefs<-Data2fd(argvals =  
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$deaths_100,basisobj =  
Base15bs)$coefs
```

```
for (i in 2:length(CountriesList)){
```

```

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","deaths_100"))],c("times","death
s_100")]

deaths.Coefs<-cbind(deaths.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$deaths_100,basisobj =
Base15bs)$coefs)
}

colnames(deaths.Coefs)<-names(CountriesList)

deaths.fd<-fd(deaths.Coefs,basisobj = Base15bs)

deaths.fd

#FDA de Tasa de reproduccion.

TemporalFrame<-
CountriesList[[1]][complete.cases(CountriesList[[1]],c("times","reproduction_rate"))],c("times
","reproduction_rate")]

reproduction.Coefs<-Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15bs)$coefs

for (i in 2:length(CountriesList)){

TemporalFrame<-
CountriesList[[i]][complete.cases(CountriesList[[i]],c("times","reproduction_rate"))],c("times",
"reproduction_rate")]

reproduction.Coefs<-cbind(reproduction.Coefs,Data2fd(argvals =
seq(0,1,length=length(TemporalFrame$times)),y=TemporalFrame$reproduction_rate,basisobj
= Base15bs)$coefs)
}

colnames(reproduction.Coefs)<-names(CountriesList)

reproduction.fd<-fd(reproduction.Coefs,basisobj = Base15bs)

reproduction.fd

```

```

# ACP funcional:

Cases.fpca<-pca.fd(Cases.fd,nharm = 15)

deaths.fpca<-pca.fd(deaths.fd,nharm = 15)

reproduction.fpca<-pca.fd(reproduction.fd,nharm = 15)

Cases.fpca

Cases.Discreta<-t(eval.fd(seq(0,1,length=100),Cases.fd))

Cases.CoefBasicos<-t(Cases.fd$coefs)

Cases.Scores<-Cases.fpca$scores

row.names(Cases.Scores)<-names(CountriesList)

names(Cases.Scores)<-paste0("PC",1:15)

Cases.fd$coefs

deaths.Discreta<-t(eval.fd(seq(0,1,length=100),deaths.fd))

deaths.CoefBasicos<-t(deaths.fd$coefs)

deaths.Scores<-deaths.fpca$scores

row.names(deaths.Scores)<-names(CountriesList)

names(deaths.Scores)<-paste0("PC",1:15)

reproduction.Discreta<-t(eval.fd(seq(0,1,length=100),reproduction.fd))

reproduction.CoefBasicos<-t(reproduction.fd$coefs)

reproduction.Scores<-reproduction.fpca$scores

row.names(reproduction.Scores)<-names(CountriesList)

names(reproduction.Scores)<-paste0("PC",1:15)

library(NbClust)

Nb<-NbClust(data=Cases.Discreta, distance = "euclidean", min.nc = 2, max.nc = 10, method =
"kmeans", index = "all")

Cases.Discreta.Clus<-kmeans(Cases.Discreta,centers=3)

```

```

Cases.Discreta.Clus
table(Cases.Discreta.Clus$cluster)

Cluster1.Cases.Discreta<-names(sort(Cases.Discreta.Clus$cluster))[1:14]
Cluster2.Cases.Discreta<-names(sort(Cases.Discreta.Clus$cluster))[15:67]
Cluster3.Cases.Discreta<-names(sort(Cases.Discreta.Clus$cluster))[68:106]

Cluster1.Cases.Discreta
Cluster2.Cases.Discreta
Cluster3.Cases.Discreta

plot(mean(Cases.fd[Cluster1.Cases.Discreta]))
lines(mean(Cases.fd[Cluster2.Cases.Discreta]),col=2)
lines(mean(Cases.fd[Cluster3.Cases.Discreta]),col=3)

Deaths.Discreta.Clust<-kmeans(deaths.Discreta,centers=4)

Deaths.Discreta.Clust
table(Deaths.Discreta.Clust$cluster)

Cluster1.Deaths.Discreta<-names(sort(Deaths.Discreta.Clust$cluster))[1:1]
Cluster2.Deaths.Discreta<-names(sort(Deaths.Discreta.Clust$cluster))[2:26]
Cluster3.Deaths.Discreta<-names(sort(Deaths.Discreta.Clust$cluster))[27:53]
Cluster4.Deaths.Discreta<-names(sort(Deaths.Discreta.Clust$cluster))[54:106]

Cluster1.Deaths.Discreta
Cluster2.Deaths.Discreta
Cluster3.Deaths.Discreta
Cluster4.Deaths.Discreta

plot(mean(deaths.fd[Cluster1.Deaths.Discreta]))
lines(mean(deaths.fd[Cluster2.Deaths.Discreta]),col=2)
lines(mean(deaths.fd[Cluster3.Deaths.Discreta]),col=3)

```

```

lines(mean(deaths.fd[Cluster4.Deaths.Discreta]),col=4)

Nb<-NbClust(data= reproduction.Discreta, distance = "euclidean", min.nc = 2, max.nc = 10,
method = "kmeans", index = "all")

Reproduction.Discreta.Clust<-kmeans(reproduction.Discreta,centers=3)

table(Reproduction.Discreta.Clust$cluster)

Cluster1.Reproduction.Discreta<-names(sort(Reproduction.Discreta.Clust$cluster))[1:31]
Cluster2.Reproduction.Discreta<-names(sort(Reproduction.Discreta.Clust$cluster))[32:77]
Cluster3.Reproduction.Discreta<-names(sort(Reproduction.Discreta.Clust$cluster))[78:106]

Cluster1.Reproduction.Discreta
Cluster2.Reproduction.Discreta
Cluster3.Reproduction.Discreta

plot(Reproduction.Discreta.Clust$cluster)

Nb<-NbClust(data= Cases.Scores, distance = "euclidean", min.nc = 2, max.nc = 10, method =
"kmeans", index = "all")

Cases.Scores.Clus<-kmeans(Cases.Scores,centers=3)

Cases.Scores.Clus
table(Cases.Scores.Clus$cluster)

Cluster1.Cases.Scores<-names(sort(Cases.Scores.Clus$cluster))[1:41]
Cluster2.Cases.Scores<-names(sort(Cases.Scores.Clus$cluster))[42:56]
Cluster3.Cases.Scores<-names(sort(Cases.Scores.Clus$cluster))[56:106]

Cluster1.Cases.Scores
Cluster2.Cases.Scores
Cluster3.Cases.Scores

plot(Bases.Scores.Clus$cluster)

plot(mean(Cases.fd[Cluster1.Cases.Scores]))

```

```

lines(mean(Cases.fd[Cluster2.Cases.Scores]),col=2)

lines(mean(Cases.fd[Cluster3.Cases.Scores]),col=3)

Nb<-NbClust(data=deaths.Scores, distance = "euclidean", min.nc = 2, max.nc = 10, method =
"kmeans", index = "all")

Deaths.Scores.Clus<-kmeans(deaths.Scores,centers=4)

table(Deaths.Scores.Clus$cluster)

Cluster1.Deaths.Scores<-names(sort(Deaths.Scores.Clus$cluster))[1:1]

Cluster2.Deaths.Scores<-names(sort(Deaths.Scores.Clus$cluster))[2:55]

Cluster3.Deaths.Scores<-names(sort(Deaths.Scores.Clus$cluster))[56:81]

Cluster4.Deaths.Scores<-names(sort(Deaths.Scores.Clus$cluster))[82:106]

Cluster1.Deaths.Scores

Cluster2.Deaths.Scores

Cluster3.Deaths.Scores

Cluster4.Deaths.Scores

plot(mean(deaths.fd[Cluster1.Deaths.Scores]))

lines(mean(deaths.fd[Cluster2.Deaths.Scores]),col=2)

lines(mean(deaths.fd[Cluster3.Deaths.Scores]),col=3)

lines(mean(deaths.fd[Cluster4.Deaths.Scores]),col=4)

Nb<-NbClust(data= reproduction.Scores, distance = "euclidean", min.nc = 2, max.nc = 10,
method = "kmeans", index = "all")

Reproduction.Scores.Clus<-kmeans(reproduction.Scores,centers=3)

table(Reproduction.Scores.Clus$cluster)

Cluster1.Reproduction.Scores<-names(sort(Reproduction.Scores.Clus$cluster))[1:30]

Cluster2.Reproduction.Scores<-names(sort(Reproduction.Scores.Clus$cluster))[31:62]

Cluster3.Reproduction.Scores<-names(sort(Reproduction.Scores.Clus$cluster))[63:106]

```

```

Cluster1.Reproduction.Scores

Cluster2.Reproduction.Scores

Cluster3.Reproduction.Scores

plot(Reproduction.Scores.Clus$cluster)

Nb<-NbClust(data=Cases.CoeffBasicos, distance = "euclidean", min.nc = 2, max.nc = 10, method
= "kmeans", index = "all")

Cases.CoeffBasicos.Clust<-kmeans(Cases.CoeffBasicos,centers=3)

table(Cases.CoeffBasicos.Clust$cluster)

Cluster1.Cases.CoeffBasicos<-names(sort(Cases.CoeffBasicos.Clust$cluster))[1:15]

Cluster2.Cases.CoeffBasicos<-names(sort(Cases.CoeffBasicos.Clust$cluster))[16:68]

Cluster3.Cases.CoeffBasicos<-names(sort(Cases.CoeffBasicos.Clust$cluster))[69:106]

Cluster1.Cases.CoeffBasicos

Cluster2.Cases.CoeffBasicos

Cluster3.Cases.CoeffBasicos

plot(Cases.CoeffBasicos.Clust$cluster)

plot(mean(Cases.fd[Cluster1.Cases.CoeffBasicos]))

lines(mean(Cases.fd[Cluster2.Cases.CoeffBasicos]),col=2)

lines(mean(Cases.fd[Cluster3.Cases.CoeffBasicos]),col=3)

Nb<-NbClust(data= deaths.CoeffBasicos, distance = "euclidean", min.nc = 2, max.nc = 10, method
= "kmeans", index = "all")

Deaths.CoeffBasicos.Clust<-kmeans(deaths.CoeffBasicos,centers=4)

table(Deaths.CoeffBasicos.Clust$cluster)

Cluster1.Deaths.CoeffBasicos<-names(sort(Deaths.CoeffBasicos.Clust$cluster))[1:54]

Cluster2.Deaths.CoeffBasicos<-names(sort(Deaths.CoeffBasicos.Clust$cluster))[55:55]

Cluster3.Deaths.CoeffBasicos<-names(sort(Deaths.CoeffBasicos.Clust$cluster))[56:82]

```

```

Cluster4.Deaths.CoeffBasicos<-names(sort(Deaths.CoeffBasicos.Clust$cluster))[83:106]

Cluster1.Deaths.CoeffBasicos

Cluster2.Deaths.CoeffBasicos

Cluster3.Deaths.CoeffBasicos

Cluster4.Deaths.CoeffBasicos

plot(mean(Deaths.fd[Cluster1.Deaths.CoeffBasicos]))

lines(mean(Deaths.fd[Cluster2.Deaths.CoeffBasicos]),col=2)

lines(mean(Deaths.fd[Cluster3.Deaths.CoeffBasicos]),col=3)

lines(mean(Deaths.fd[Cluster4.Deaths.CoeffBasicos]),col=4)

Nb<-NbClust(data=reproduction.CoeffBasicos, distance = "euclidean", min.nc = 2, max.nc = 10,
method = "kmeans", index = "all")

reproduction.CoeffBasicos.Clust<-kmeans(reproduction.CoeffBasicos,centers=3)

table(reproduction.CoeffBasicos.Clust$cluster)

Cluster1.reproduction.CoeffBasicos<-names(sort(reproduction.CoeffBasicos.Clust$cluster))[1:9]

Cluster2.reproduction.CoeffBasicos<-names(sort(reproduction.CoeffBasicos.Clust$cluster))[10:50]

Cluster3.reproduction.CoeffBasicos<-
names(sort(reproduction.CoeffBasicos.Clust$cluster))[51:106]

Cluster1.reproduction.CoeffBasicos

Cluster2.reproduction.CoeffBasicos

Cluster3.reproduction.CoeffBasicos

plot(mean(reproduction.fd[Cluster1.reproduction.CoeffBasicos]))

lines(mean(reproduction.fd[Cluster2.reproduction.CoeffBasicos]),col=2)

lines(mean(reproduction.fd[Cluster3.reproduction.CoeffBasicos]),col=3)

```