



Universitat Oberta
de Catalunya

uoc.edu



Universitat Oberta
de Catalunya

Universitat Oberta de Catalunya (UOC)
Máster en Ciencia de Datos

TRABAJO FIN DE MÁSTER

Calidad Asistencial Sanitaria Europea

Alicia Perdices Guerra

Máster en Ciencia de Datos

Área Medicina (TFM-Med), Salud Pública y Medical Devices.

Nombre Consultor/a Susana Pérez Álvarez

Nombre Profesor/a responsable de la asignatura: Àngels Rius Gavidia.

Fecha de entrega PEC3 (entrega parcial): **13 de abril 2021**



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-

SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Calidad Asistencial Sanitaria Europea</i>
Nombre del autor:	<i>Alicia Perdices Guerra</i>
Nombre del consultor/a:	<i>Susana Pérez Álvarez</i>
Nombre del PRA:	<i>Àngels Rius Gavidia</i>
Fecha de entrega (mm/aaaa):	04/2021
Titulación:	<i>Máster Universitario en Ciencia de Datos</i>
Área del Trabajo Final:	<i>Área Medicina (TFM-Med)</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Calidad asistencial europea, Estadística de Gasto, Recursos y Actividades, Predicción y patrones por países.</i>

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El Trabajo fin de máster tiene como objetivo determinar la situación asistencial sanitaria europea a través del **análisis estadístico de los gastos** generados por diferentes tipos de hospitales y países sobre los diversos cuidados necesarios por enfermedades, **de los recursos** utilizados como personal médico, enfermería e investigadores, así como **de las actividades y/o procedimientos** relacionados con altas hospitalarias en diferentes ratios y tasa de ocupación de camas, tipos de operaciones quirúrgicas y tratamientos, tecnología médica, servicios preventivos, y cuidados domiciliarios.

Para ello se ha utilizado los datos disponibles en el repositorio de **Eurostat**. Las tablas contendrían información sobre el cuidado sanitario relacionado con el **gasto, recursos y actividades hospitalarias**.

Después de realizar **tareas de limpieza y combinación de tablas**, se aplican métodos **estadísticos descriptivos** que reflejan la realidad de la calidad asistencial en los distintos países europeos. A continuación, se realizan predicciones atendiendo a distintos factores y se obtienen patrones entre los distintos países estudiados. Tras la investigación se persigue obtener la información necesaria para mejorar la calidad asistencial en hospitales gestionando el gasto y los recursos adecuadamente, optimizando los presupuestos para disminuir las estancias hospitalarias con una tasa de ocupación equilibrada, mejorar los servicios preventivos y cuidados domiciliarios, reduciendo de este modo el número de muertes y mejorando la calidad de vida de los pacientes.

Abstract (in English, 250 words or less):

The objective of the present study that I carry out is to determine the European healthcare status through the statistical analysis of the **expenses** applied by different types of hospitals and countries to the various necessary care for diseases, of the **resources** used as medical personnel, nurses and researchers , as well as **activities** related to procedures such as hospital discharges in different ratios and bed occupancy rates, types of surgeries and treatments, medical technology, preventive services, causes of death and home care.

For this, data has been made available in the Eurostat repository. The tables would contain information on Health care related to hospital expenses, resources and activities.

After cleaning and combining the tables, descriptive statistical methods are applied that reflect the reality of healthcare quality in different European countries. Then, predictions are made based on different factors and patterns are obtained between the different studied countries.

After the investigation, the aim is to obtain the necessary information to improve the quality of care in hospitals, managing spending and resources appropriately, optimizing budgets to reduce hospital stays with a balanced occupancy rate, improve preventive services and home care, thus reducing the number of deaths and improving the quality of patient's life.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	1
1.3 Enfoque y método seguido	2
1.4 Planificación del Trabajo	3
1.5 Breve resumen de productos obtenidos	6
1.6 Breve descripción de los otros capítulos de la memoria	6
2. Estado del Arte.....	7
3. Diseño e implementación del trabajo.....	13
3.1 Especificaciones de Diseño.....	13
3.2 Proceso de obtención y Limpieza de Datos (ETL, Extract, Transform, Load).....	15
3.3 Análisis Exploratorio de datos (EDA, Exploratory Data Analysis).....	24
3.4 Modelización: Modelos de Regresión.....	40
3.5 Modelización: Contraste de muestras.....	44
3.6 Modelización Avanzada de los datos.....	48
4. Conclusiones.....	54
5. Glosario	58
6. Bibliografía	59
6. Anexos	61

Lista de figuras y tablas

Figuras

- [Figura 1.](#) Casafranca, Johana. **La agilidad es ahora.** Recuperado de: <https://agileisnow.com/2021/02/21/por-que-usar-el-tablero-kanban/>
- [Figura 2.](#) Tic. Portal. **Scrum: una manera de ejecutar los principios de la metodología Agile.** Recuperado de: <https://www.ticportal.es/glosario-tic/scrum-implementacion-proyectos>
- [Figura 3.](#) Diagrama de Gantt.
- [Figura 4.](#) Estadísticos de Salud por Regiones Europeas Eurostat Statistics Explained. **Health statistics at regional level.2021.** Recuperado de: https://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level#Health_status_and_health_care.
- [Figura 5.](#) Cinco formas de hacer que la Sanidad sea Efectiva, Accesible y Resistente. **EU_HEALTH. State of Health in the EU 2017. 5 ways to make healthcare effective, accessible and resilient.** 2017. Recuperado de: https://twitter.com/eu_health/status/933684222614753280
- [Figura 6.](#) Estructura de los datos.
- [Figura 7.](#) Diagrama de flujo de la Extracción, Transformación y Carga de Datos (ETL).
- [Figura 8.](#) Análisis Exploratorio de Datos.
- [Figura 9.](#) Regresión Logística.
- [Figura 10.](#) Contraste de Muestras.
- [Figura 11.](#) Modelización Avanzada de los Datos.
- [Figura 12.](#) Diagramas de Dispersión para comprobar dependencia lineal
- [Figura 13.](#) Curvas Q-Q y relación entre variables. (Diálisis, Trasplante)
- [Figura 14.](#) Curvas Q-Q y relación entre variables. (Mortalidad Tratable, Prevenible)

Tablas

- [Tabla 1.](#) Evolución del Gasto Sanitario, Personal Sanitario y Mortalidad.
- [Tabla 2.](#) Evolución de las Actividades Preventivas, Diálisis y Trasplante en el Tiempo.
- [Tabla 3.](#) Gasto Sanitario, Personal Sanitario y Mortalidad por Países.
- [Tabla 4.](#) Actividades Preventivas, Diálisis y Trasplante por Países
- [Tabla 5](#) Tabla resumen de la Evolución en el tiempo de la Mortalidad y la Atención primaria.
- [Tabla 6](#) Tabla resumen de Países con la mayor media en Recursos en un periodo de tiempo.
- [Tabla 7](#) Curvas de Densidad y Q-Q

1. Introducción

1.1 Contexto y justificación del Trabajo

Lo que se pretende estudiar con este trabajo es la situación real **de la asistencia sanitaria en Europa y cómo se responde a las necesidades cotidianas en los distintos países**. Es de vital importancia poder darse cuenta si los pacientes son atendidos correctamente, si se están utilizando los recursos necesarios, o si se dispone de ellos, y si su gestión se está realizando correctamente.

En la situación actual que estamos viviendo con la pandemia de la COVID-19, se está poniendo de manifiesto las carencias de los sistemas sanitarios mundiales, evidenciando en muchos casos, no solo falta de recursos, sino una gestión deficiente de éstos.

Por este motivo, es primordial tener un plan de actuación en relación con las distintas situaciones sanitarias que pueden vivir los distintos países, saber qué recursos son los más solicitados, y cómo gestionarlos para evitar un colapso hospitalario y que todos los pacientes, puedan ser tratados en tiempo y forma.

En lo que respecta a la **motivación** para elegir esta línea de investigación, tengo que decir, que mi experiencia vivida durante muchos años en hospitales mientras me curaba de una leucemia y recibía diálisis debido a que la quimioterapia dañó mis riñones, hace que sienta cierta implicación y necesidad de compartir lo aprendido y seguir aprendiendo para poder conseguir que todos los pacientes no solo sean atendidos, sino que se haga a tiempo y en las mejores condiciones posibles. Que todo paciente sienta que está en las mejores manos, que se sienta seguro y confiado, y que se haga por él todo lo posible. Esto solo lo podremos conseguir si se tienen recursos y se gestionan bien.

1.2 Objetivos del Trabajo

- El objetivo principal es **descubrir la situación real de los hospitales** y en general del **sistema sanitario**, en relación con los **gastos presupuestados, recursos disponibles, actividades y estrategias preventivas** y hacer una **comparativa por países**, poniendo el foco **en aquellos con más recursos**, pudiendo encontrar las claves que llevarán a modificar los aspectos necesarios al resto de países. Además de establecer **predicciones** sobre el futuro de esta calidad asistencial que marcarán las pautas a seguir.

- Analizar **qué países han aplicado medidas de prevención sanitarias** a su población para la detección de enfermedades como el cáncer de mama y cérvix uterino, por ejemplo.
- Obtener **patrones por países** que ayuden a explicar el **éxito o fracaso conseguido**, es decir, que expliquen **la mortalidad evitable (tratable y/o prevenible) en función de los recursos y estrategias de prevención** para modificar la gestión realizada en aquellos con peores resultados.
- Realizar **predicciones** sobre la evolución de ciertos índices (mortalidad y tratamientos renales) para poder rectificar en tiempo y forma todos los aspectos negativos que encuentre.
- Realizar **varias visualizaciones** sobre todo lo investigado para facilitar la comprensión del estudio.

1.3 Enfoque y método seguido

El método de trabajo está basado en **metodologías ágiles** que son un conjunto de prácticas innovadoras para el desarrollo de proyectos que prioriza la **agilidad y la flexibilidad** en la evolución de éstos. El tipo de metodología ágil seguido es **un tablero Kanban**.

Un tablero Kanban, es una herramienta visual que ofrece una visión general del estado actual del trabajo, y así, poder optimizar y mejorar todas las tareas planificadas para lograr los objetivos buscados.

En este tipo de tableros, se asignan **tareas o elementos colocados en columnas**. Estas columnas son una secuencia de los pasos a seguir de principio a fin en el proyecto o trabajo. Los elementos se pueden organizar por colores para facilitar conseguir los objetivos marcados.

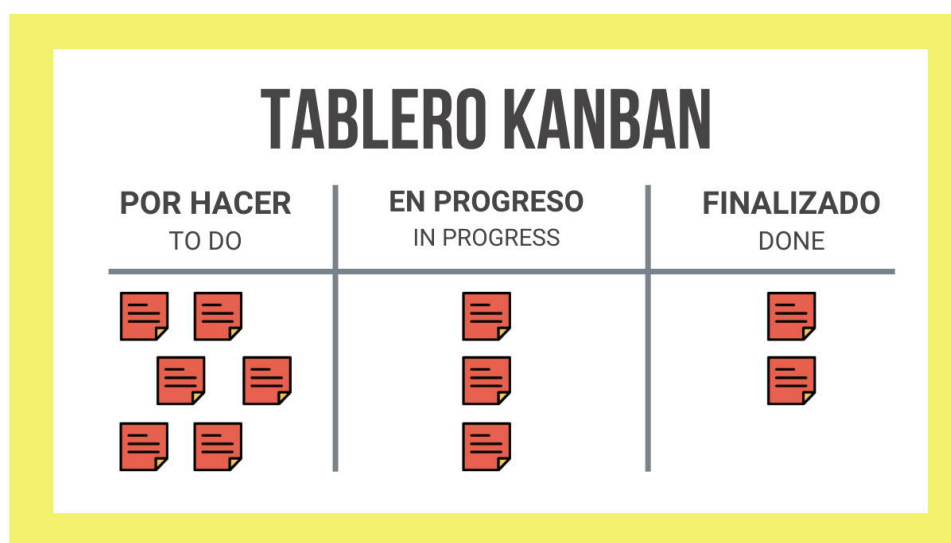


Fig. 1 Ejemplo de Tablero Kanban.

Existen otros tipos de metodologías ágiles como **SCRUM**. Ésta, se caracteriza por planificar tareas de un proyecto complejo en distintas fases con el fin de obtener resultados relativamente pronto ofreciendo un trabajo de calidad. En SCRUM, los proyectos se ejecutan en ciclos temporales cortos y de duración fija en los que cada iteración tiene que proporcionar un resultado completo para poder ser entregado. Ésta es la razón por la que no he seguido esta metodología, ya que el proyecto desarrollado en el TFM no lo permite de ese modo.

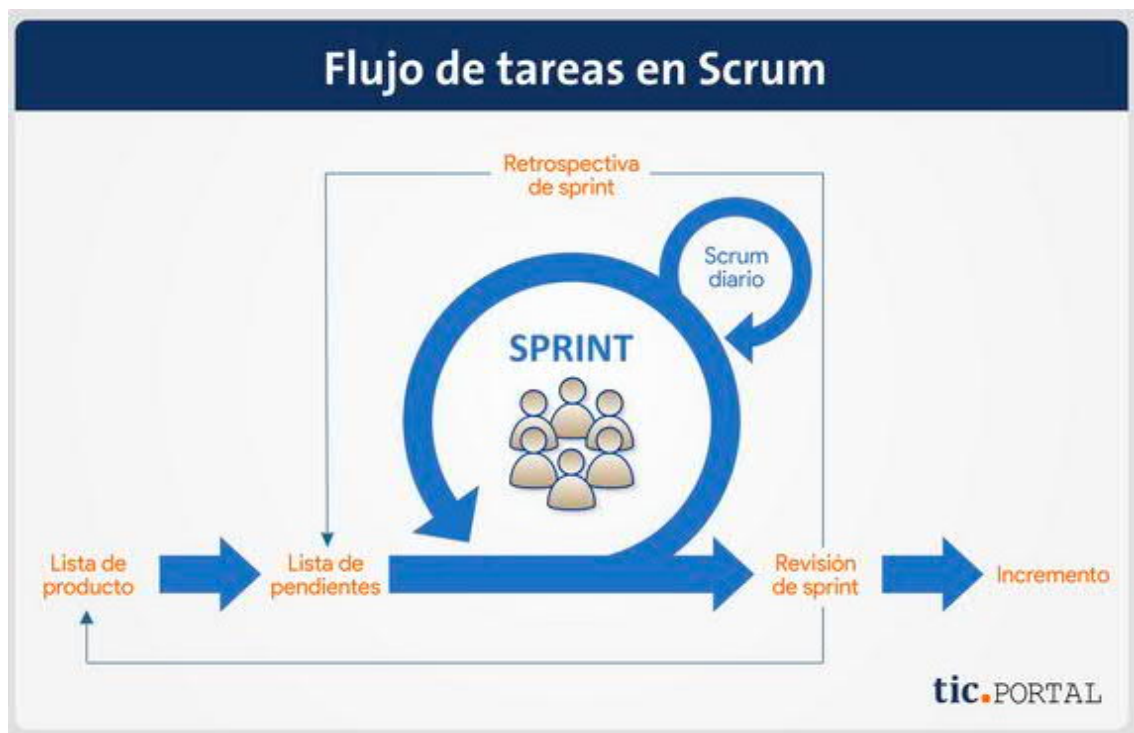


Fig.2 Flujo de Tareas Scrum

1.4 Planificación del Trabajo

La planificación total abarcaría 20 semanas dividida en varios bloques claves:

- **Definición y Planificación del trabajo Final de Máster** en la que se concretan el contexto, justificación y motivación personal. Por otro lado, se marcan los objetivos, el método y la planificación del trabajo en una línea temporal. Se ha empleado 2 semanas.
- **Estado del Arte.** En el que se plasma la investigación llevada a cabo en el pasado intentando resolver la problemática del trabajo que se está realizando. Se ha empleado 3 semanas
- **Diseño e implementación del trabajo.** En el que se diseña, desarrolla e implementa el producto. Esta fase del proyecto está dividida en varias tareas:

- **Obtención y limpieza de datos (ETL, Extract Transform and Load).** Que abarcan 2 semanas cada una, aunque se han simultaneado en los últimos días dedicados a estas tareas. Se estudian valores extremos, inconsistencias y valoración de posibles imputaciones.
- **Análisis de los Datos:**
 - **Análisis Descriptivo y Exploratorio de datos (EDA, Exploratory Data Analysis).** Es una de las fases a la que más tiempo se ha dedicado, 3 semanas.
 - Una vez terminado el análisis, se pasó a las siguientes fases de creación de **Modelos de Regresión** y **patrones** mediante **Machine Learning**.
- Por último, se sacaron las **conclusiones** de la investigación y el análisis
- **Redacción y presentación de la Memoria.** Que se completaron en 4 semanas y que incluye la explicación de la necesidad del proyecto, los objetivos, el estado del arte o análisis de mercado, el diseño e implementación del proyecto y sus conclusiones.

La división de tareas quedaría plasmada en el siguiente diagrama temporal:

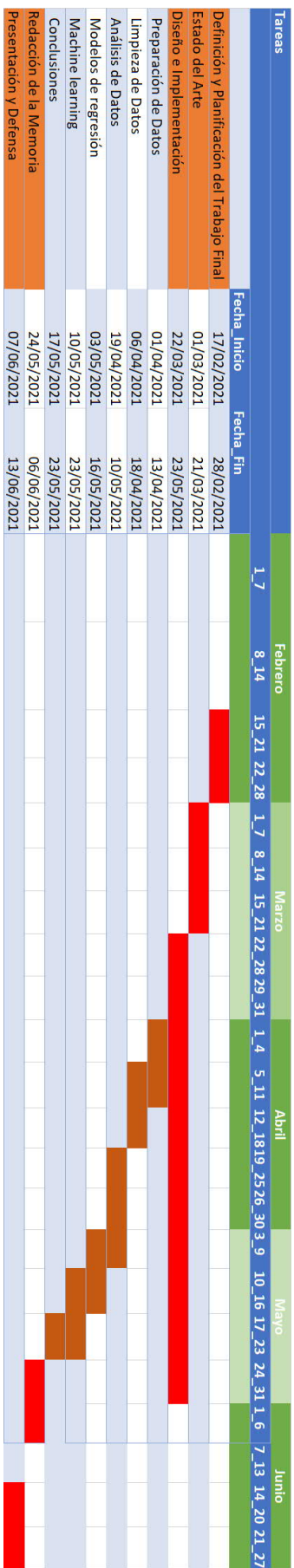


Fig. 3 Diagrama de Gantt

1.5 Breve resumen de productos obtenidos

Los productos generados en el trabajo fin de máster incluyen:

- Archivos RMD en código R en relación con:
 - La limpieza de datos.
 - Análisis exploratorio de los datos (EDA).
 - Modelización. Regresión.
 - Modelización. Contraste de Muestras.
 - Archivos PDF y HTML con el código en R.
- Notebook en código Python en relación con:
 - La modelización avanzada. Clasificación no supervisada.
 - Archivo HTML con el código en Python.
- Conjunto de datos originales extraídos de la base de datos Eurostat en formato “csv”.
<https://ec.europa.eu/eurostat/web/health/data/database>
- Archivos “csv” generados en la extracción, transformación y carga de datos (ETL).
- Memoria. Documento PDF que contiene el desarrollo completo del TFM.
- Visualizaciones en Tableau Public en relación con el análisis de datos.
<https://public.tableau.com/app/profile/alicia3628/viz/CalidadAsistencialSanitariaEuropea/GastoSanitario>

1.6 Breve descripción de los otros capítulos de la memoria

A continuación, se detalla los contenidos de los capítulos de la memoria:

1. **Introducción:** Se describen el contexto y justificación del Trabajo relacionado con la calidad asistencial sanitaria europea. Los objetivos, la motivación y la planificación del trabajo.
2. **Estado del arte.** En este apartado se detallan los trabajos o artículos que han estudiado todos los aspectos de la calidad asistencial sanitaria con anterioridad.
3. **Diseño e implementación del trabajo.** Se desarrollan las etapas detalladas en la planificación, el proceso de extracción, transformación y carga de datos (ETL), Análisis exploratorio de los datos (EDA) y modelización en relación con modelos de regresión, contraste de muestras y modelización avanzada de clasificación no supervisada.
4. **Conclusiones.** Se resumen los resultados obtenidos contrastándolos con los objetivos fijados en el inicio.

2. Estado del Arte

La calidad asistencial sanitaria ha sido objeto de estudio en múltiples ocasiones, realizado desde ópticas distintas.

Entre los estudios a los que se hace referencia encontramos el artículo **La calidad Asistencial: concepto y medida** de María del Mar Villegas e Isabel María Rosa Díaz [\[1\]](#).

Se ha intentado definir su concepto con claridad y sobre todo descubrir cuál es la medida óptima que acredite si la calidad es buena o mala, pero **nunca existe un único criterio** que nos ayude a medir la calidad de la atención al paciente.

En general, la calidad asistencial responde a la interacción de un conjunto de dimensiones que son:

- ✓ **Científico-Técnicas** en relación con el correcto diagnóstico, tratamiento y destreza médicas.
- ✓ **Interpersonales.** Relacionadas con:
 - La confianza y comunicación entre médico y paciente.
 - Empatía, teniendo especial cuidado en el tacto a la hora de tratar con el paciente.
 - Trato también a familiares.
- ✓ **Ambientales**, entorno, relacionadas con el confort, servicios prestados, intimidad en los ingresos...
- ✓ **Sociales.** Poniendo el foco en toda la infraestructura necesaria para que familias socialmente desfavorecidas reciban los tratamientos que necesitan con las relaciones interpersonales óptimas.
- ✓ **Económicas.**

La dimensión científico-técnica es algo que se da por hecho, ser atendido y recibir el tratamiento adecuado para la resolución del problema médico.

Sin embargo, en los últimos años, **las expectativas y las necesidades de los pacientes están jugando un papel muy importante** a la hora de valorar la calidad de la asistencia sanitaria.

El punto de vista del paciente está siendo considerado como un factor esencial hasta tal punto que está produciendo cambios en los sistemas sanitarios regionales o nacionales. Por ejemplo, el plan de Calidad de la Conserjería de salud de la junta de Andalucía explica que *“El ciudadano es el centro del sistema sanitario público, por lo tanto, la satisfacción de sus necesidades, demandas y expectativas se convierten en objetivos fundamentales en toda la política sanitaria andaluza”*.

Aunque hay que destacar, que **todavía las opiniones de los pacientes no provocan cambios realmente importantes** en los aspectos organizativos de nuestro sistema de salud.

Del mismo modo, en la investigación realizada por Miguel López-Coronado, Isabel de la Torre, Jesús Herreros y Javier Cabo se evidencia una mejora de la calidad asistencial mediante la **telemedicina y la teleasistencia** debido en muchos casos a la falta no solo de personal especializado sino además de su disponibilidad [2].

Por otro lado, en el artículo “La aplicación del modelo europeo de gestión de calidad Total al Sector Sanitario” [3], se explica cómo la **EFQM (The European Foundation for Quality Management)** ha elaborado un modelo que es entendido como un conjunto de procesos de mejora en los que todo el personal sanitario está involucrado con el objetivo de obtener la calidad asistencial deseada. Actualmente existen dos formas distintas de enfocarlo:

- ✓ **Monitorización** de dimensiones como la mortalidad, cirugía inapropiada, tratamientos inadecuados etc....
- ✓ Y lo que se llama **Individualización**, que pone el foco en las necesidades de los pacientes y/o profesionales. Aunque sigue siendo **más frecuente el enfoque en la calidad científico-técnico**, las expectativas de los pacientes y profesionales se abren camino rápida y eficazmente.

Es decir, para la gestión de la calidad total de la EFQM se ha elaborado un modelo que se basa en la siguiente premisa: *“la satisfacción del cliente, de los empleados y el impacto positivo en la sociedad se consiguen mediante el liderazgo en la política y estrategia, gestión de personal, el uso eficiente de los recursos y una adecuada definición de los procesos”*. Esto es, constituye una herramienta que posibilita a cualquier organización ya sea sanitaria o no, orientar su gestión hacia la **satisfacción del cliente**.

Y para finalizar con una aplicación específica, el estudio sobre la **evaluación y seguimiento de la calidad asistencial, ofrecida a los intoxicados en los Servicios de Urgencias** [4], revelaron que, con la disponibilidad de un protocolo específico de tratamiento de las intoxicaciones, de material adecuado para llevar a cabo estos procesos como sondas especiales, stock de antídotos suficientes y realización de análisis específicos, se obtuvieron resultados satisfactorios. Por lo que concluimos que la **existencia de protocolos y recursos, redundan en una mejora de la calidad científico- técnicas** y en este estudio en concreto en la **calidad de satisfacción del paciente medida por el número de reclamaciones** que se hicieron que no superó la cifra general de las reclamaciones del Servicio de Urgencias.

En los diferentes artículos mencionados hasta ahora, se comprueba que la calidad asistencial sanitaria no se puede medir o valorar únicamente por un único criterio, aunque si existe un foco común en los procesos de mejora que se llevan a cabo, la **satisfacción del paciente**.

Hasta ahora se ha analizado qué variables se tienen en cuenta para valorar la calidad asistencial sanitaria ofrecida a los pacientes y cómo la satisfacción de éstos es el foco de los procesos de mejora que se llevan a cabo.

Para medir la satisfacción de los pacientes se tienen en cuenta 5 puntos básicos, recogidos en el artículo **Health Statistics at Regional Level** [5]:

- Estado físico.
- Estado Emocional.
- Situación Social.
- Síntomas
- Medidas biométricas.

El estudio arroja los siguientes datos:

- Casi el 70% de la población adulta europea percibe su salud como buena o muy buena, encontrando diferencias por edades y por regiones (zonas rurales o urbanas), siendo la población adulta más joven que vive en las ciudades la que se considera en mejor estado de salud.
- La población de las zonas rurales no accede de forma satisfactoria a los servicios sanitarios bien porque son caros, por la imposibilidad de desplazarse a los hospitales por largas distancias o por falta de medios de transporte, o debido a las largas lista de espera para la realización de ciertas pruebas médicas que disuaden a los pacientes de realizárselas.
- Como media, en 2017 había 1 médico por cada 266 habitantes.

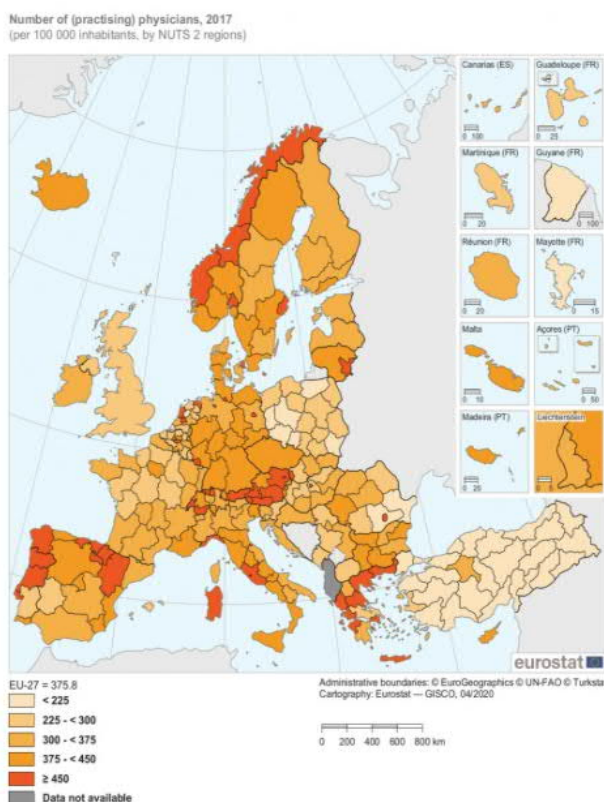
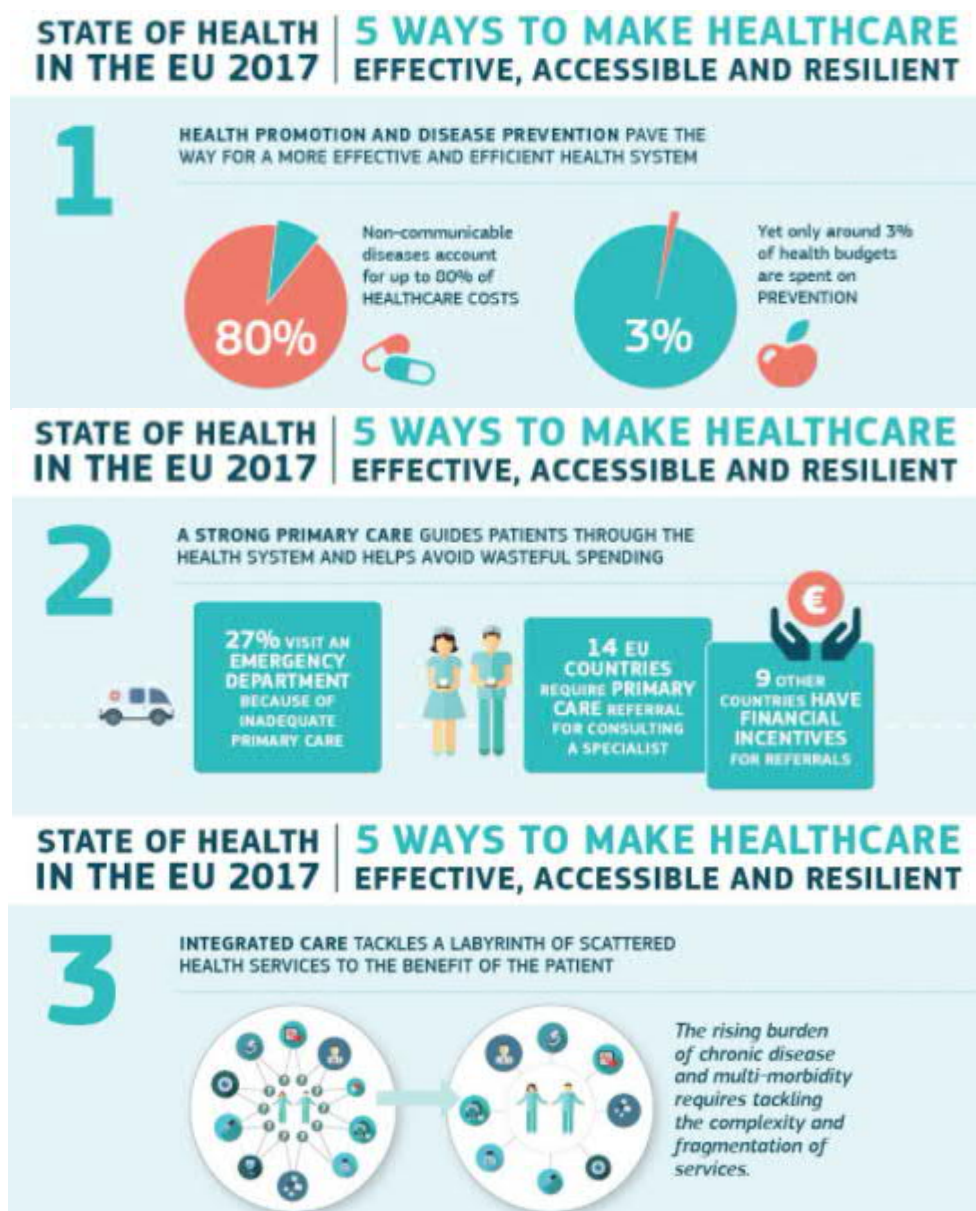


Fig. 4. Estadísticos de Salud por Regiones Europeas

De entre los procesos de mejora, cabe mencionar las recomendaciones de la **Comisión Europea para la Salud y Seguridad Alimentaria (SANTE, EU Commission's DG Health & Food Safety)** que encontramos resumidas en los siguientes gráficos [\[6\]](#):

- Promover la Prevención.
- Fortalecer la Asistencia Médica primaria en los Centros de Salud.
- Fragmentación de los Servicios Sanitarios en beneficio del paciente.
- Previsión y Planificación proactiva.
- Base de datos centralizada de los pacientes.



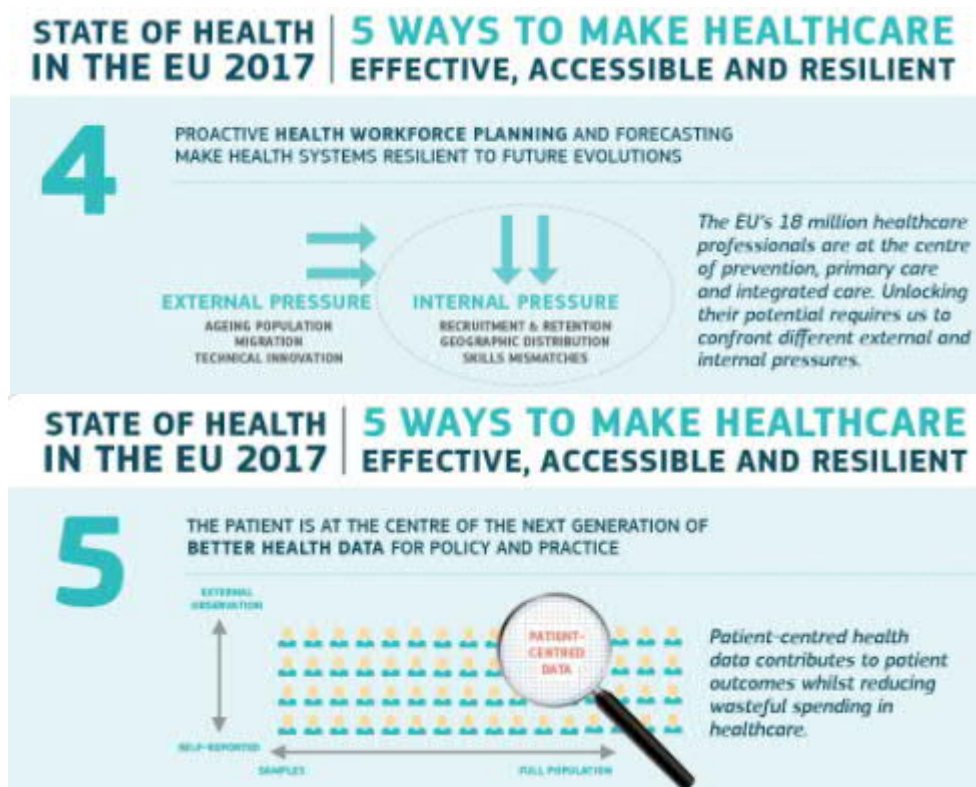


Fig. 5 Cinco formas de hacer que la Sanidad sea Efectiva, Accesible y Resistente.

Y para finalizar, a modo de resumen, se describen las instituciones que integran los Sistemas Sanitarios de varios Países Europeos y se comparan con el establecido en USA. **Howard SJ y Davison P** lo resumen en el artículo: **Comparing Health protection systems across six European nations and the USA** [7].

- **Finlandia.** El Sistema de Salud se integra en el Instituto Nacional de Salud y Bienestar (**THL**) financiado en parte por empresas privadas y farmacéuticas, con un modelo que integra la investigación y la práctica.
- **Francia.** Como consecuencia de las reformas en el sistema de protección de la Salud, se establece una nueva organización llamada **l'institute** que engloba diferentes instituciones. Además, existen 26 Agencias Regionales de Salud.
- **Alemania.** El instituto **Robert Koch (RKI)** como responsable del control y la prevención de enfermedades, la **Oficina Federal de Estadísticas (SB)** con funciones epidemiológicas y de vigilancia, el **instituto Paul Ehrlich (PEI)** con responsabilidad en los programas de vacunación integran el sistema de protección para la salud alemán.
- **Hungría.** La responsabilidad nacional de protección de la salud recae en el director médico, que supervisa ocho institutos nacionales de salud y salud pública dentro de una organización conocida como **ÁNTSZ**. Cabe destacar que en Hungría no hay un programa obligatorio de vacunación.
- **Países Bajos.** Las funciones para la protección de la salud a nivel regional son realizadas por los **servicios municipales**. Sin embargo, a

nivel nacional, es el **Centro para el Control de Enfermedades Infecciosas (Cib)** el encargado de estas funciones.

- **USA.** La protección de la salud se proporciona a nivel de nacional en los Centros para la Prevención y el Control de Enfermedades (CDC) que se organizan en Oficinas (Ej. Oficina de Enfermedades Infecciosas) y subdividen en Centros (Ej. Centro nacional de Enfermedades infecciosas Emergentes y Zoonóticas).

3. Diseño e implementación del trabajo.

En el apartado de Diseño e Implementación, se indican las especificaciones de Diseño en relación con la estructura de los Datos, su proceso de Extracción, Transformación y Carga (ETL), Análisis Exploratorio (EDA) y Modelización.

3.1 Especificaciones de Diseño.

La implementación del trabajo cumple las siguientes especificaciones de **diseño**:

Se deberán cumplir los objetivos marcados, es decir principalmente **descubrir la situación real del sistema sanitario** en relación con los **gastos presupuestados, recursos disponibles y actividades** y hacer una **comparativa por países**.

Para ello, se obtiene la información necesaria para mejorar la calidad asistencial en hospitales, gestionando los recursos adecuadamente y de este modo mejorar la calidad de vida de los pacientes.

De entre los programas y lenguajes utilizados tanto en el proceso de extracción, transformación y carga de datos (ETL), así como el Análisis Exploratorio de los Datos (EDA) y la aplicación de Modelos, destacamos:

- El entorno de desarrollo **RStudio** en lenguaje de programación **R**.
- El programa **Anaconda** en lenguaje de programación **Python**.
- Por otro lado, se ha utilizado el sistema de control de versiones de código **GitHub** para almacenar, los datos, y los ficheros con el código en R y Python tanto del proceso ETL, EDA y Modelización.
- La plataforma **Tableau Public** para la realización de visualizaciones.

La estructura de los datos disponibles es la siguiente:

ESTRUCTURA DE LOS DATOS		
Nombre del Bloque	Sub-Bloque y Descripción de Archivos	
1. Bloque Gasto Sanitario (4 archivos)	1. Financiación del Gasto Sanitario. 2. Gasto Sanitario por Función (cuidados curativos, rehabilitación...). 3. Gasto Sanitario por proveedor (destinatario del gasto sanitario). 4. Ingresos en el Sector Sanitario por Países y Organizaciones.	
2. Bloque Recursos Sanitarios •(Staff: 5 archivos,	Staff (Personal Hospitalario)	Facilities (Recursos Auxiliares, que facilitan las actividades

•Facilities: 4 archivos)		hospitalarias)
	<ol style="list-style-type: none"> 1. Médicos por especialidad 2. Médicos por Países. 3. Personal de Enfermería y Cuidados. 4. Personal Sanitario Hospitalario. 5. Personal Sanitario Hospitalario exceptuando Enfermería. 	<ol style="list-style-type: none"> 1. N.º de camas por propiedad. 2. Recursos Técnicos Hospitalarios. 3. Tecnología Médica. 4. Tipos de Camas Hospitalarias.
3. Bloque de Actividades (4 archivos)	<ol style="list-style-type: none"> 1. Altas, Media de Estancia Hospitalaria. 2. Ocupación de camas. 3. Pacientes Trasplantados y en Diálisis. 4. Ratio de Altas en el Hospital de Día por Diagnóstico. 	
4. Bloque de Prevención (3 archivos)	<ol style="list-style-type: none"> 1. Detección de Cáncer de Mama y Cérvix Uterino. 2. Detección de Enfermedades Cardiovasculares y Diabetes. 3. Vacunación a la población mayor de 65 años. 	
5. Bloque de Causas de Mortalidad (2 archivos)	<ol style="list-style-type: none"> 1. Mortalidad Tratable y Prevenible 2. Muertes por Enfermedades Infecciosas. 	
6. Cuidados Domiciliarios (1 archivo)	<ol style="list-style-type: none"> 1. Cuidados Domiciliarios. 	
7. Estado de Salud (1 archivo)	<ol style="list-style-type: none"> 1. Estado de Salud. 	

Fig. 6 Estructura de los Datos

3.2 Proceso de obtención y Limpieza de Datos (ETL, Extract, Transform, Load).

A continuación, se muestra el diagrama de flujo del proceso ETL:

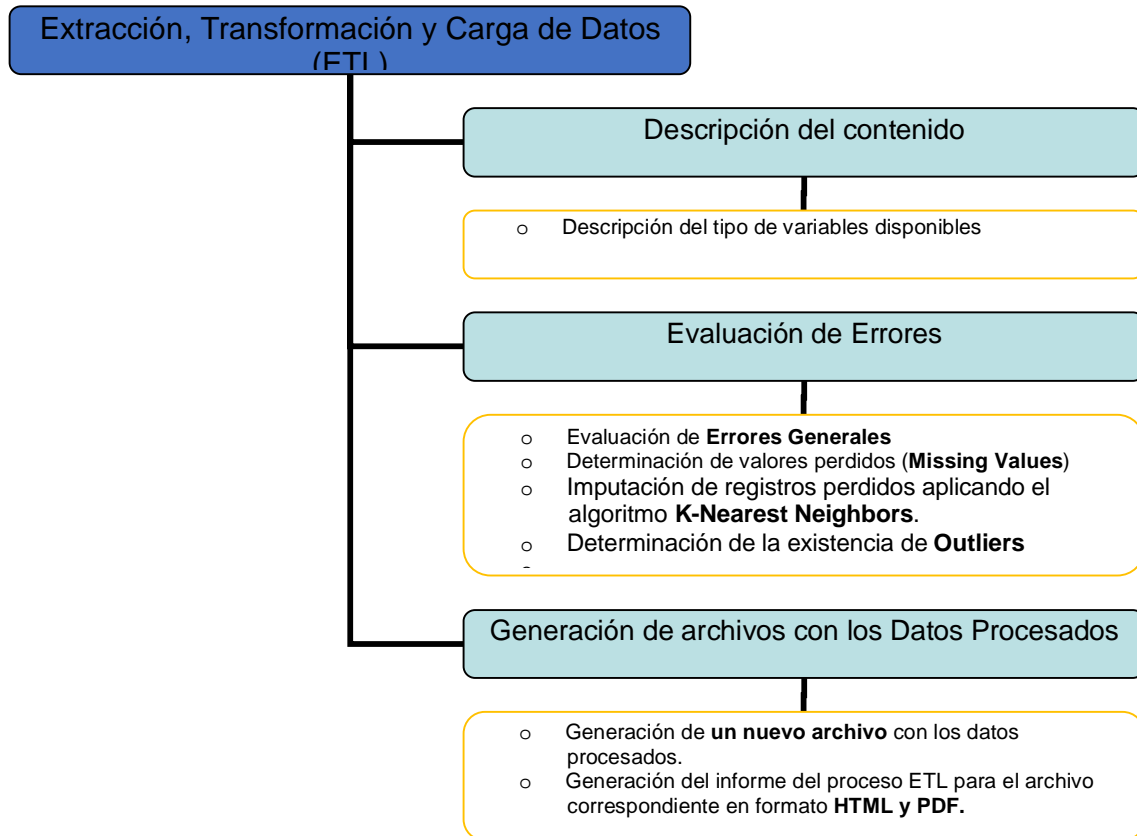


Fig. 7 Diagrama de flujo de la Extracción, Transformación y Carga de Datos (ETL)

Por otro lado, se describe cada bloque de datos en relación con las fases del proceso ETL.

- Los bloques de datos son:
 - Bloque de **Gasto Sanitario**.
 - Bloque de **Recursos Sanitarios**.
 - Personal del Hospital (Staff)**.
 - Recursos Auxiliares (Facilities)**.
 - Bloque de **Actividades**.
 - Bloque de **Prevención**.
 - Bloque de **Causas de Muerte**.
 - Bloque de **Cuidados Domiciliarios**.
 - Bloque de **Estado de Salud**.
- Los variables genéricas a todos los archivos de todos los bloques están descritas a continuación:
 - TIME**: variable cuantitativa. Indica el año en el que se ha realizado la medida, en este caso el valor de la variable "Value". Se carga bien como número entero.

2. **GEO:** variable cualitativa. Indica el país o región en el que se ha realizado la medida. Se carga bien como factor.
 3. **UNIT:** variable cualitativa. Indica la medida de la variable valor. Se carga bien como factor.
 4. **Value:** Variable cuantitativa. Indica el número del recurso en cuestión (Ya sea gasto sanitario, número de recursos tecnológicos etc.). Es una variable numérica, pero se carga como factor. Se hace transformación.
 5. **Fal.and.footnotes:** Notas sobre etiquetas. Eliminamos esta columna.
- Se ha utilizado el algoritmo K-Nearest Neighbors para la imputación de Missing Values, por la gran versatilidad que presenta para estimar todo tipo de patrones de valores perdidos ya sean continuos o discretos. Funciona del siguiente modo:
Dado un conjunto de datos incompleto, se seleccionan K casos más cercanos del conjunto de datos conocido que estén más cercanos al valor desconocido de tal forma que minimicemos una métrica de distancia.

En cada bloque se describen las variables propias de cada archivo a considerar. Encontramos entre ellas variables que hacen referencia a códigos internacionales sobre clasificaciones de los planes de financiación sanitaria. Se puede consultar el siguiente enlace para más información: European Environment Agency. **Vocabulary: Classification of health care financing schemes- SHA 2011.**[\[8\]](#)

1. BLOQUE GASTO SANITARIO

1) Financiación del Gasto Sanitario.

I. Descripción del contenido.

1. **ICHA11_HF:** variable cualitativa. Indica el organismo de financiación, ya sea gubernamental o por seguros privados etc.

II. Evaluación de Errores

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 513 valores perdidos de 2000 disponibles.
3. Se comprueba que existen **outliers** en los valores de la variable **"Value"**.

III. Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Financiacion_clean.csv**

2) Gasto Sanitario en relación con la Función

I. Descripción del contenido:

1. **ICHA11_HC:** variable cualitativa. Indica cómo se aplica el gasto sanitario por función.

II. Evaluación de errores

1. Evaluación de **Errores Generales**.

2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 595 valores perdidos de 2000 disponibles.
 3. Se comprueba que existen **outliers** en los valores de la variable **“Value”**.
 - III. **Generación de archivos con la información corregida.**
 1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Funcion_clean.csv**.
- 3) **Gasto Sanitario en relación con la Entidad sobre la que se aplica el gasto sanitario.**
- I. **Descripción del contenido:**
 1. **ICHA11_HP**: variable cualitativa. Indica la entidad sobre la que se aplica el gasto sanitario.
 - II. **Evaluación de Errores.**
 1. Evaluación de **Errores Generales**.
 2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 694 valores perdidos de 2000.
 3. Se comprueba que existen **outliers** en los valores de la variable **“Value”**.
 - III. **Generación de archivos con la información corregida.**
 1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Proveedor_clean.csv**.
- 4) **Ingresos Sanitarios por países.**
- I. **Descripción del contenido:**
 1. **ICHA11_FS**: variable cualitativa. Indica los ingresos por países.
 - II. **Evaluación de Errores.**
 1. Evaluación de **Errores Generales**
 2. **Determinación de Datos faltantes o Missing Values**.
 - Se observan 62 valores perdidos de 220.
 3. Se comprueba que existen **outliers** en los valores de la variable **“Value”**.
 - III. **Generación de archivos con la información corregida.**
 1. Se genera el archivo resultado del proceso de los datos: **IngresosSanitarios_Financiacion_clean.csv**.

2. BLOQUE RECURSOS SANITARIOS

STAFF

1) **Médicos por Edad y Países.**

- I. **Descripción del contenido**
 1. **AGE**: Variable cualitativa. Indica el rango de edad del número de médicos disponibles por países. Variable factor.
 2. **SEX**: Variable cualitativa. Solo un valor “Total”. Variable total.
- II. **Evaluación de Errores.**

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 419 valores perdidos de 1900.
3. Se comprueba, que la variable “**Value**” tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **Medicos_Por_Paises_clean.csv**.

2) Médicos por Especialidad.

I. Descripción del contenido

1. **MED_SPEC**: Variable cualitativa. Indica la especialidad médica. Variable factor.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Observamos 390 valores perdidos de 1800.
3. Se comprueba, que la variable “**Value**” tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero limpio: **Medicos_x_especialidad_clean.csv**.

3) Personal Sanitario Hospitalario.

I. Descripción del contenido

1. **ISCO08**: Variable cualitativa. Indica el tipo de personal al que se refiere la variable “Value”, esto es, enfermeros/as, personal de cuidados, matronas, Médicos, Matronas, o Personal hospitalario.
2. **WSTATUS**: Variable cualitativa. Indica el estatus laboral del personal sanitario hospitalario.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 883 valores perdidos de 1800.
3. Se comprueba, que la variable “**Value**” tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **PersonalSanitarioHospitalario_clean.csv**.

4) Personal de Enfermería y Cuidados.

I. Descripción del contenido

1. **WSTATUS**: Variable cualitativa. Indica el estatus laboral del personal de Enfermería y Cuidados.
2. **ISCO08**: Variable cualitativa. Indica el tipo de personal al que se refiere la variable “Value”, esto es, enfermeros/as, personal de cuidados, matronas...

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 9347 valores perdidos de 17100.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos:

Personal_Enfermeria_Cuidados_clean.csv.

5) Personal Sanitario Exceptuando Enfermería.

I. Descripción del contenido

1. **WSTATUS**: Variable cualitativa. Indica el estatus laboral del personal sanitario (Exceptuando Enfermería).
2. **ISCO08**: Variable cualitativa. Indica el tipo de personal al que se refiere la variable **“Value”**, esto es, Médicos, Dentistas, Fisioterapeutas o Farmacéuticos.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 847 valores perdidos de 1600.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos:
PersonalSanitario_No_enfermeria_clean.csv.

FACILITIES

1) Número/Ratio de Camas por Propiedad y Países.

I. Descripción del contenido

1. **OWNER**: Variable cualitativa. Indica la propiedad de las camas hospitalarias.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 786 valores perdidos de 2790.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos:
Camas_Propiedad_clean.csv.

2) Recursos Técnicos Hospitalarios.

I. Descripción del contenido

1. **FACILITY**: Variable cualitativa. Indica el tipo de recurso hospitalario.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.

- Se observan 1030 valores perdidos de 1860.
 - 3. Se comprueba, que la variable “**Value**” tiene valores extremos.
 - III. Generación de archivos con la información corregida.**
 - 1. Se genera el archivo resultado del proceso de los datos: **RecursosTecnicos_hospitalarios_clean.csv.**
- 3) Tecnología Médica.**
- I. Descripción del contenido**
 - 1. **FACILITY:** Variable cualitativa. Indica el tipo de Tecnología Médica.
 - 2. **ICHA_HP:** Variable cualitativa. Indica dónde se usa el recurso tecnológico: Hospitales, Ambulatorios, etc....
 - II. Evaluación de Errores.**
 - 1. Evaluación de **Errores Generales.**
 - 2. Determinación de **Datos faltantes o Missing Values.**
 - Se observan 7005 valores perdidos de 15750.
 - 3. Se comprueba, que la variable “**Value**” tiene valores extremos.
 - III. Generación de archivos con la información corregida.**
 - 1. Se genera el fichero con los datos procesados: **TecnologiaMedica_clean.csv.**
- 4) Tipos de camas hospitalarias.**
- I. Descripción del contenido**
 - 1. **FACILITY:** Variable cualitativa. Indica el tipo de cama hospitalaria.
 - II. Evaluación de Errores.**
 - 1. Evaluación de **Errores Generales.**
 - 2. Determinación de **Datos faltantes o Missing Values.**
 - Se observan 1604 valores perdidos de 5850.
 - 3. Se comprueba, que la variable “**Value**” tiene valores extremos.
 - III. Generación de archivos con la información corregida.**
 - 1. Se genera el fichero con los datos procesados: **TiposCamasHospitalarias_clean.csv.**

3. ACTIVIDADES

1) Altas, Estancias y Camas Hospitalarias en Pacientes Ingresados y por Servicios Curativos.

- I. Descripción del contenido**
 - 1. **OWNER:** Variable cualitativa. Indica la propiedad de las camas hospitalarias.
- II. Evaluación de Errores.**
 - 1. Evaluación de **Errores Generales.**
 - 2. Determinación de **Datos faltantes o Missing Values.**
 - Se observan 786 valores perdidos de 2790.
 - 3. Se comprueba, que la variable “**Value**” tiene valores extremos.
- III. Generación de archivos con la información corregida.**
 - 1. Se genera el fichero con los datos procesados: **Camas_Propiedad_clean.csv.**

2) Recursos Técnicos Hospitalarios.

I. Descripción del contenido

1. **FACILITY:** Variable cualitativa. Indica el tipo de recurso hospitalario.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 1030 valores perdidos de 1860.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
RecursosTecnicos_hospitalarios_clean.csv.

3) Tecnología Médica.

I. Descripción del contenido

1. **FACILITY:** Variable cualitativa. Indica el tipo de Tecnología Médica.
2. **ICHA_HP:** Variable cualitativa. Indica dónde se usa el recurso tecnológico: Hospitales, Ambulatorios, etc....

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 7005 valores perdidos de 15750.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
TecnologiaMedica_clean.csv.

4) Tipos de camas hospitalarias.

I. Descripción del contenido

1. **FACILITY:** Variable cualitativa. Indica el tipo de cama hospitalaria.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 1604 valores perdidos de 5850.
3. Se comprueba, que la variable **“Value”** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
TiposCamasHospitalarias_clean.csv.

4. PREVENCIÓN

1) Detección del Cáncer de Mama y Cérvix Uterino.

I. Descripción del contenido

1. **SOURCE:** Variable cualitativa. Indica la fuente del estudio.

2. **ICD10:** Indica el tipo de cáncer foco de estudio: De mama o Cérvix Uterino.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 915 valores perdidos de 1440.
3. Se comprueba, que la variable “**Value**” tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **Deteccion_Cancer_Mama_Cervix_clean.csv**.

2) Detección de Enfermedades Cardiovasculares y Diabetes.

I. Descripción del contenido

1. **HLTHCARE:** Variable cualitativa. Indica el tipo de medida.
2. **ISCED11:** Variable cualitativa. Estándar en estadísticas de educación en el que se hacen las mediciones.
3. **DURATION:** Variable cualitativa. Indica el tiempo durante el cual se hace la medición.
4. **SEX:** Variable cualitativa. En cómputo total. No hay distinción entre sexos.
5. **AGE:** Variable cualitativa. En cómputo total. No hay distinción entre edades.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - No se observan valores perdidos.
3. Se comprueba, que la variable “**Value**” no tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **Deteccion_enfCardio_Diabetes_clean.csv**.

3) Vacunación a la población mayor de 65 años.

I. Descripción del contenido

1. Variables comunes descritas al inicio.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 119 valores perdidos de 380.
3. Se comprueba, que la variable “**Value**” no tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **Vacunacion_+65_clean.csv**.

5. CAUSAS DE MUERTE

1) Mortalidad Tratable y Prevenible.

I. Descripción del contenido

1. **MORTALIT**: Variable cualitativa. Indica el tipo de mortalidad: Tratable o Prevenible.
2. **ICD10**: Variable Cualitativa. En la clasificación de enfermedades, se indica el total de ellas.
3. **SEX**: Variable cualitativa. Indica el sexo de la población estudiada.

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 54 valores perdidos de 4536.
3. Se comprueba, que la variable **"Value"** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **Mortalidad_Tratable_Prevenible_clean.csv**.

2) Mortalidad por Enfermedades infecciosas.

I. Descripción del contenido

1. **ICD10**: Variable Cualitativa. Indica la clasificación de enfermedades infecciosas que se han estudiado en la población.
2. **SEX**: Variable cualitativa. Indica el sexo de la población estudiada. Total (No hay diferencia entre sexos).

II. Evaluación de Errores.

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 156 valores perdidos de 1332.
3. Se comprueba, que la variable **"Value"** tiene valores extremos.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **Deteccion_enfCardio_Diabetes_clean.csv**.

6. BLOQUE CUIDADOS DOMICILIARIOS

1) Cuidados Domiciliarios.

I. Descripción del contenido.

1. **ISCED11**: variable cualitativa. Indica estándar en estadísticas de educación en el que se hacen las mediciones.
2. **SEX**: Variable cualitativa. Indica el sexo de la población estudiada.
3. **AGE**. Variable cualitativa. En cómputo total. No hay distinción entre edades.

II. Evaluación de Errores

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - o Se observan 12 valores perdidos de 396.
3. Se comprueba que existen outliers en los valores de la variable **"Value"**.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
CuidadosDomiciliarios_clean.csv.

7. BLOQUE ESTADO DE SALUD

1) Estado de salud por Sexo y por Países.

I. Descripción del contenido.

1. **INDIC_HE**: variable cualitativa. Explica el valor de la variable “Value”, esto es, años de vida sana en valores absolutos desde el nacimiento.
2. **SEX**: Variable cualitativa. Indica el sexo de la población estudiada.

II. Evaluación de Errores

1. Evaluación de **Errores Generales**.
2. Determinación de **Datos faltantes o Missing Values**.
 - Se observan 18 valores perdidos de 990.
3. Se comprueba que existen outliers en los valores de la variable “Value”.

III. Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
EstadoDeSalud_Sexo_clean.csv

3.3 Análisis Exploratorio de datos (EDA, Exploratory Data Analysis).

Estudio de los distintos aspectos de la calidad sanitaria por países.

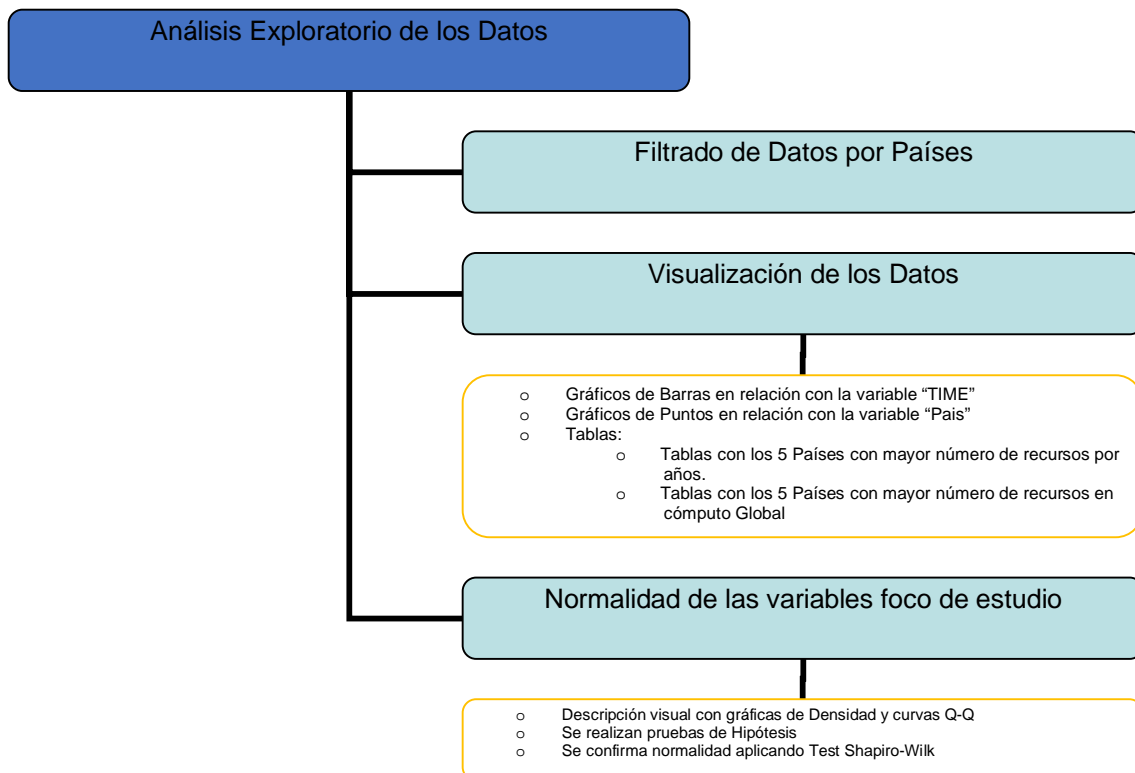


Fig. 8 Análisis Exploratorio de Datos

1) Filtrado de la información.

A continuación, se explica la motivación de la elección y filtrado de la información para el estudio.

A. Filtrado de la información, **seleccionando los países foco de estudio** (Se eliminan los campos que hacen referencia a los países de la Unión Europea de forma global en un período de tiempo determinado, ya que **el objetivo es obtener información de forma individualizada para poder establecer comparaciones entre países y descubrir patrones de similitud entre ellos.**)

- i. European Union – 27 countries (from 2020)
- ii. European Union – 28 countries (2013-2020)
- iii. European Union – 27 countries (2007-2013)
- iv. European Union – 15 countries (1995-2004)
- v. Euro area – 19 countries (from 2015)
- vi. Euro area – 18 countries (2014)
- vii. Euro area – 12 countries (2001-2006)

B. Filtrado particular de cada archivo en relación con la información foco de estudio:

- **Gasto Sanitario:** Se selecciona la información relativa al gasto destinado a **Hospitales, Hospitales Psiquiátricos y Hospitales Especiales**, además del gasto por función destinado a **Cuidados Sanitarios y Rehabilitación** en términos generales y para **pacientes ingresados**, ya que interesa comparar el gasto destinado a asistencia hospitalaria y ambulatoria porque la calidad asistencial debe poder medirse en situaciones en las que un paciente requiere hospitalización y en las que no.
- **Financiación:** Se filtra la información de una forma general (**Todos los sistemas de financiación**), porque interesa tener una visión global de la evolución de la financiación en el tiempo.
- **Recursos Sanitarios:**
 - Si se atiende a los recursos auxiliares o **Facilities**, se obtiene información sobre **número de quirófanos, y salas de cirugía y cuidados llamados “de día”** que no requieren de hospitalización para de nuevo establecer diferencias entre los recursos destinados a asistencia hospitalaria y ambulatoria.
 - Con respecto al personal sanitario o **Staff Sanitario** y/o hospitalario se obtiene información sobre el número de **médicos de atención primaria, especialistas y pediatras** englobando las necesidades básicas que cualquier paciente pueda tener, además del **número de médicos por edades** para poder plasmar la experiencia, y/o necesidad de renovación de plantillas. Por otro lado, se especifica el **número de enfermeras, matronas y auxiliares**, personal esencial en los cuidados y tratamientos a enfermos y del número de especialistas en otras áreas, como

farmacéuticos, dentistas o fisioterapeutas, servicios que tienen un rol importante en el estado de salud de cada individuo.

- **Prevención:**

Para el análisis se ha seleccionado información relativa a:

- ✓ La **Detección del Cáncer de Mama** ya que es uno de los procesos oncológicos con mayor incidencia y mortalidad (véase referencia en el enlace del National Cancer Institute [\[9\]](#)).
- ✓ La **Detección de Cáncer de Cérvix Uterino**, ya que es el segundo cáncer más frecuente después del de mama (véase referencia en el enlace de la Gaceta Médica [\[10\]](#)).
- ✓ La **Vacunación a la población mayor de 65 años** con el fin de proteger a los colectivos más vulnerables de complicaciones en procesos gripales y/o neumonías.

Las pruebas preventivas que llevan a cabo los distintos países juegan un papel importante en las curvas de supervivencia de estas enfermedades por lo que son relevantes si se pretende establecer dependencias con el índice de mortalidad existente en ellos.

- **Causas de Mortalidad:**

Para el análisis de la mortalidad en los países foco de estudio, se ha seleccionado información sobre la **mortalidad tratable, prevenible y por enfermedades infecciosas**.

El índice de la mortalidad evitable agrupada en tratable y prevenible refleja la capacidad del sistema Sanitario de cada país para poder dar una respuesta eficaz en tiempo y forma a las necesidades de los pacientes.

Por otro lado, se ha examinado el perfil de la mortalidad por enfermedades infecciosas, para poder establecer comparaciones entre las capacidades de los países para enfrentarse a infecciones víricas y así poder plasmar su situación antes de la pandemia de COVID-19. Si un país tiene un número alto de fallecimientos por enfermedades infecciosas, podría evidenciar carencias en la estrategia de vacunación de neumococo y gripe, falta de recursos para llevar a cabo actividades de prevención necesarias o simplemente falta del personal para poder llevarlas a cabo.

Todas estas variables podrían explicar el éxito o fracaso a la hora de enfrentarse a una situación límite como la pandemia contra la que aún estamos luchando.

- **Estado de Salud.**

El estado de salud de los ciudadanos de un país, es decir, la edad máxima en la que un ciudadano tiene una salud plena o lo suficientemente buena, refleja no solo la calidad de vida en una región en concreto, ya por factores ambientales (ausencia de Centrales Eléctricas o Fábricas que aumenten la contaminación) o alimenticios, sino por una buena gestión sanitaria focalizada sobre todo en prevención, y recursos.

Se ha establecido diferencias entre países por sexo.

- **Cuidados Domiciliarios.**

Aunque para este servicio sanitario solo se tiene información del año 2014, se ha querido establecer diferencias entre países sobre el porcentaje de población por sexos que ha requerido de una **cobertura sanitaria distinta, atendiendo a necesidades especiales.**

De entre todos los procesos que hacen que la salud de ciudadano pueda verse comprometida, se pone especial atención con este estudio en aquellos que no precisan de hospitalización, pero sí asistencia domiciliaria, debido por un lado por las características del tratamiento como pueda ser por ejemplo diálisis peritoneal o por una determinada situación social o sanitaria desfavorecida. Es importante tenerlo en cuenta para poder valorar la calidad asistencial proporcionada.

- **Actividades Sanitarias.**

Se ha seleccionado para el estudio información relativa al ratio de pacientes en **diálisis y trasplantados.**

Generalmente los tratamientos de diálisis se alargan mucho en el tiempo (en la mayoría de los casos varios años) debido no solo a las características propias de los pacientes que los reciben, que son en su mayoría personas mayores con problemas de salud asociados a la edad, sino también por el tiempo de espera para recibir el trasplante de un órgano. Esto hace que este tipo de tratamientos ponga a prueba la solidez del sistema sanitario en cuanto a recursos disponibles, como dializadores y máquinas de diálisis y personal sanitario. Todos estos factores, hacen que sea interesante observar la evolución en el tiempo de este tipo de pacientes para poder hacer una previsión de los recursos necesarios y así, poder proporcionar un tratamiento adecuado.

2) Visualización y Distribución de los datos.

En relación con la visualización de la información se ha graficado la variable “value” del siguiente modo:

Value: Variable que en cada archivo nos indica el valor del aspecto considerado a investigar.

1. Evolución en el tiempo mediante GRÁFICAS DE BARRAS.

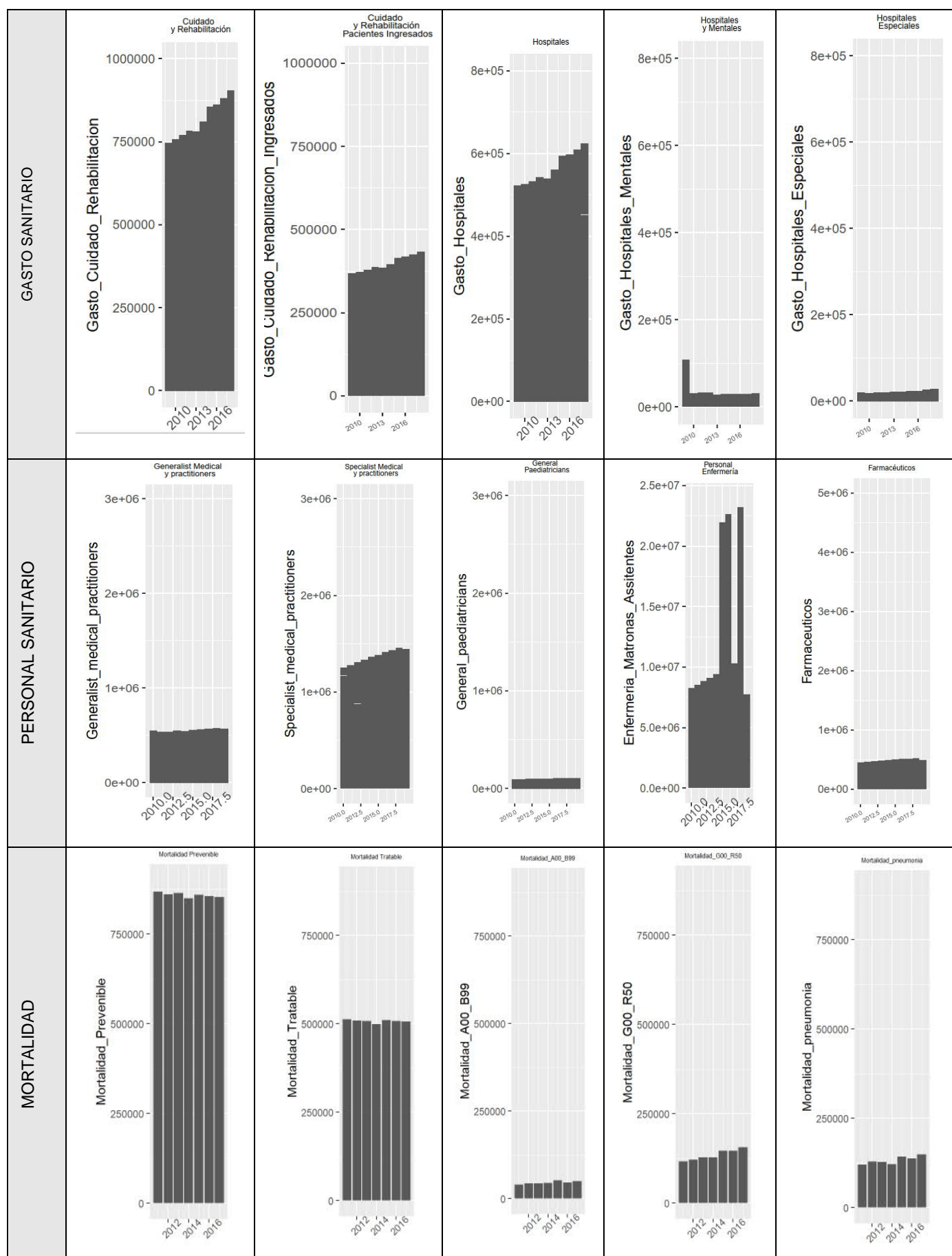


Tabla 1. Evolución del Gasto Sanitario, Personal Sanitario y Mortalidad en el Tiempo

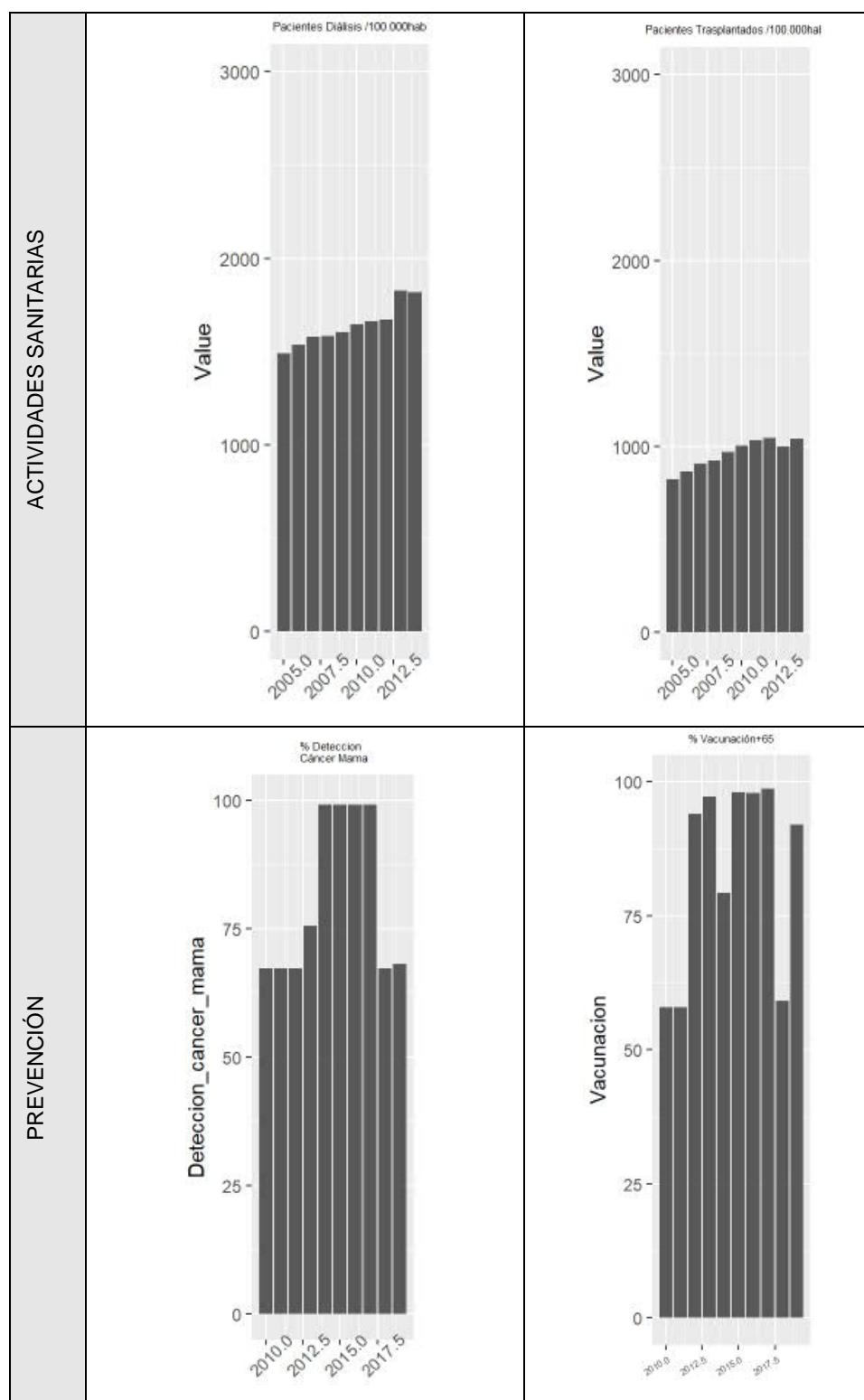


Tabla 2. Evolución de las Actividades Preventivas, Diálisis y Trasplante en el Tiempo

2. Distribución de los datos por Países mediante GRÁFICAS DE PUNTOS.

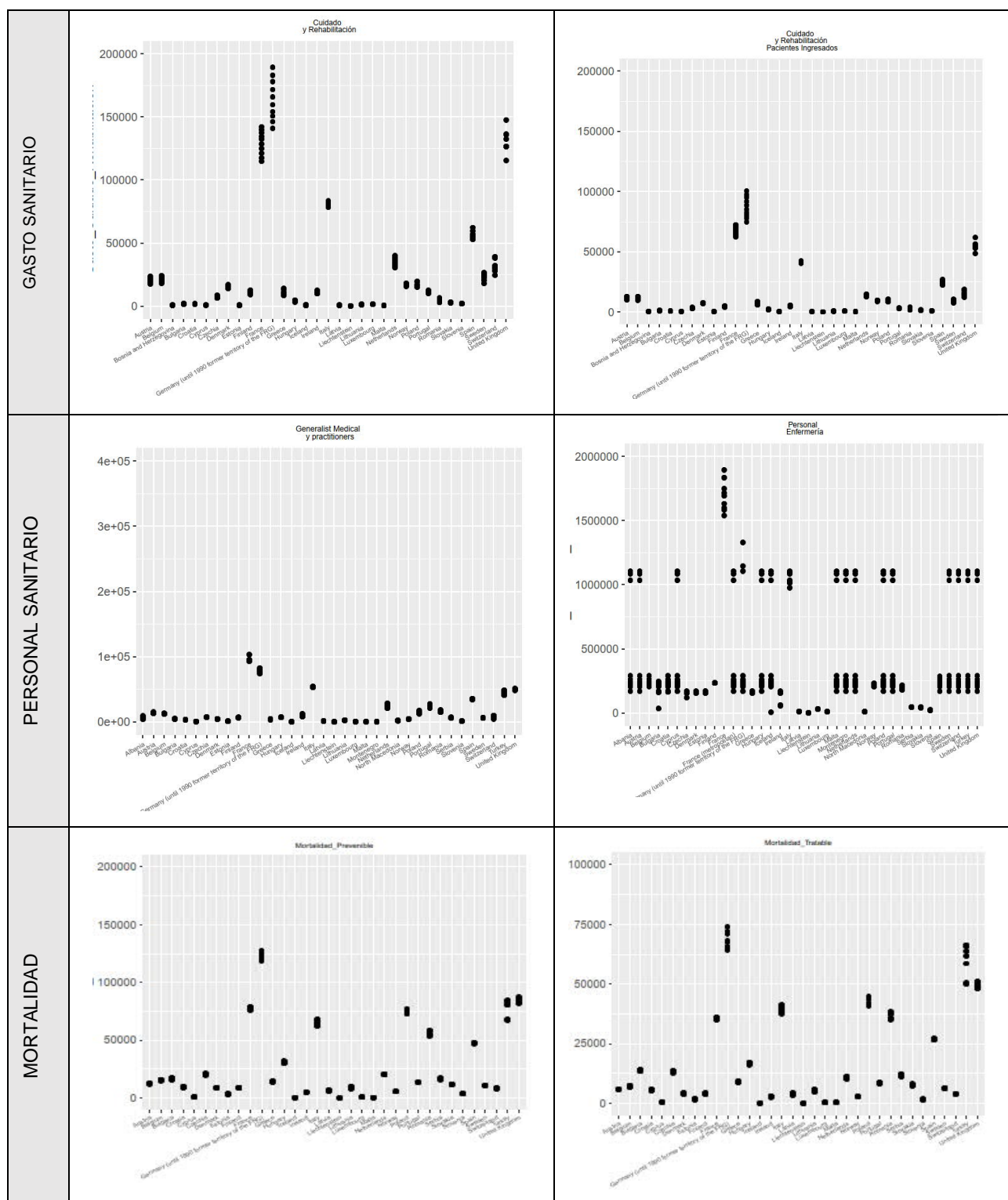


Tabla 3. Gasto Sanitario, Personal Sanitario y Mortalidad por Países

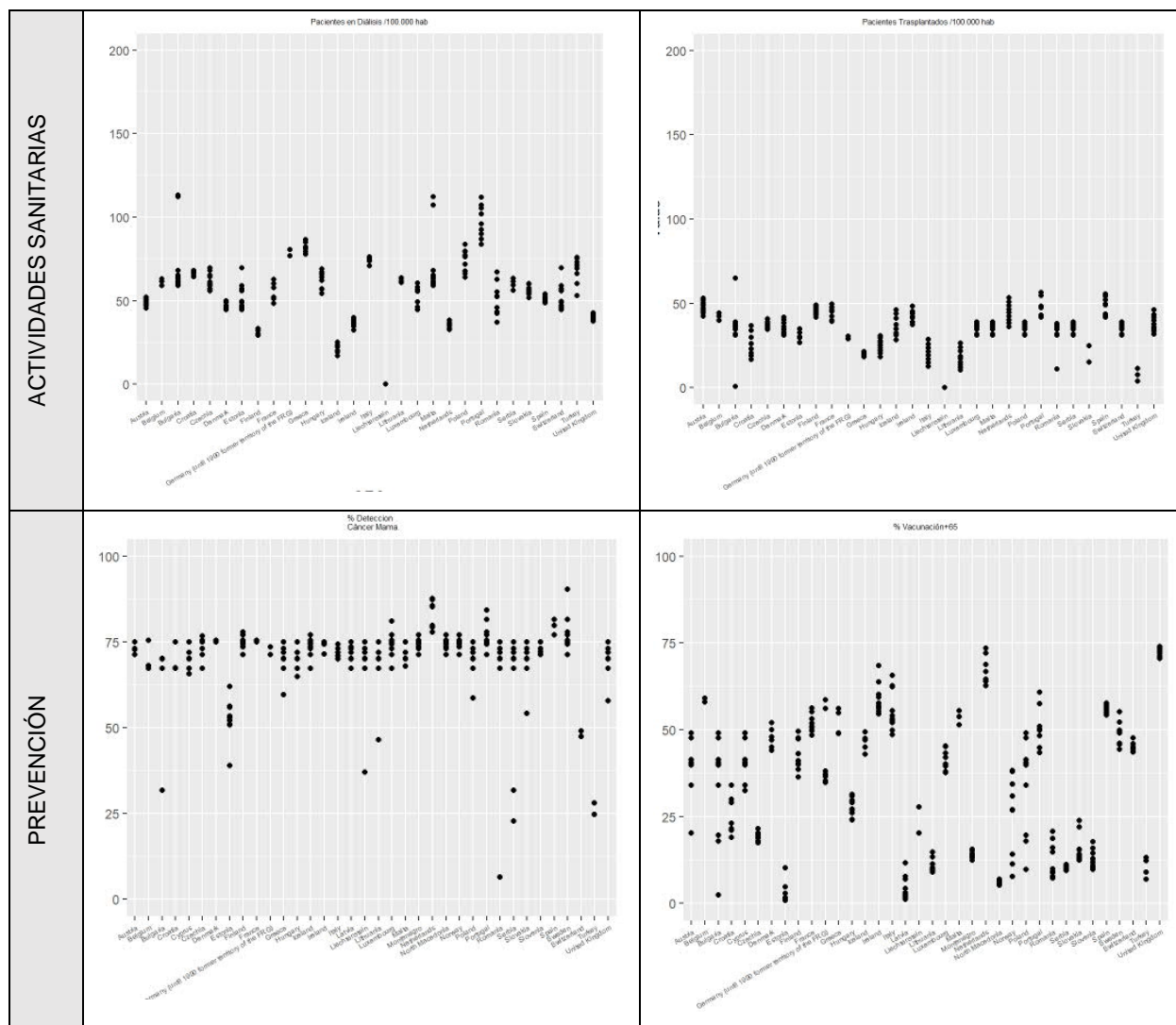


Tabla 4. Actividades Preventivas, Diálisis y Trasplante por Países

3. Obtención de tablas con los 5 países con mayor recuento en el aspecto considerado en cuestión, en cada año de forma individual (en 2010, 2011...)

Esta tabla recoge a modo de resumen la información relativa a la mortalidad tratable, prevenible y a los médicos de atención primaria.

	Mortalidad Tratable	Mortalidad Prevenible	Atención Primaria
2010			Pais Medicos G
			France 103262
			Germany (until 1990 former territory of the FRG) 74113
			Italy 52944
			United Kingdom 48641
2011			Turkey 41077
	Pais Mortalidad_Tratable	Pais Mortalidad_Prevenible	Pais Medicos G
	Germany (until 1990 former territory of the FRG) 74029.0	Germany (until 1990 former territory of the FRG) 127461.0	France 93140
	Turkey 50411.0	United Kingdom 82226.0	Germany (until 1990 former territory of the FRG) 75252
	United Kingdom 48861.0	France 78950.5	Italy 53856
	Poland 44643.5	Poland 77175.5	United Kingdom 49838
	Italy 41124.5	Turkey 68225.0	Turkey 41840

2012	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Germany (until 1990 former territory of the FRG)	72005.5	Germany (until 1990 former territory of the FRG)	125457.5	France	93283
	Turkey	49912.5	United Kingdom	81962.5	Germany (until 1990 former territory of the FRG)	75390
	United Kingdom	48594.5	France	77596.5	Italy	53994
	Poland	44342.0	Poland	76992.0	United Kingdom	50008
	Italy	41222.0	Turkey	67271.5	Turkey	41401
2013	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Germany (until 1990 former territory of the FRG)	71225.0	Germany (until 1990 former territory of the FRG)	126132.0	France	93944
	Turkey	58609.0	United Kingdom	82478.5	Germany (until 1990 former territory of the FRG)	78608
	United Kingdom	48137.5	Turkey	80578.0	Italy	53525
	Poland	43734.0	France	77270.0	United Kingdom	49875
	Italy	39717.0	Poland	75400.0	Turkey	41040
2014	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Germany (until 1990 former territory of the FRG)	67489.5	Germany (until 1990 former territory of the FRG)	120737.5	France	94322
	Turkey	61678.0	United Kingdom	82551.0	Germany (until 1990 former territory of the FRG)	80226
	United Kingdom	48027.0	Turkey	80566.0	Italy	53463
	Poland	40861.5	France	76003.5	United Kingdom	50595
	Italy	38732.0	Poland	72755.5	Turkey	41542
2015	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Germany (until 1990 former territory of the FRG)	68212.0	Germany (until 1990 former territory of the FRG)	122759.0	France	94262
	Turkey	63826.0	United Kingdom	86777.5	Germany (until 1990 former territory of the FRG)	78601
	United Kingdom	49579.5	Turkey	80323.0	Italy	53610
	Poland	41704.5	France	77864.0	United Kingdom	50169
	Italy	40000.0	Poland	72929.5	Turkey	44646
2016	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Turkey	66395.5	Germany (until 1990 former territory of the FRG)	120129.5	France	94328
	Germany (until 1990 former territory of the FRG)	65536.5	United Kingdom	86777.5	Germany (until 1990 former territory of the FRG)	79847
	United Kingdom	50951.5	Turkey	84518.5	Italy	54063
	Poland	42060.5	France	77585.5	United Kingdom	50080
	Italy	37638.0	Poland	73164.5	Turkey	46107
2017	Pais	Mortalidad_Tratable	Pais	Mortalidad_Prevenible	Pais	Medicos G
	Turkey	65810.0	Germany (until 1990 former territory of the FRG)	118506.5	France	94889
	Germany (until 1990 former territory of the FRG)	64417.5	United Kingdom	87159.0	Germany (until 1990 former territory of the FRG)	81411
	United Kingdom	50880.0	Turkey	83229.0	Italy	53691
	Poland	43948.0	France	77585.5	United Kingdom	49824
	Italy	38040.5	Poland	74998.0	Turkey	47999
2018					Pais	Medicos G
					France	94923
					Germany (until 1990 former territory of the FRG)	82757
					Italy	52998
					United Kingdom	49569
					Turkey	48688
2019					Pais	Medicos G
					France	94889
					Germany (until 1990 former territory of the FRG)	81411
					Italy	53114
					United Kingdom	51096
					Turkey	47999

Tabla 5. Tabla resumen de la Evolución en el tiempo de la Mortalidad y la Atención primaria.

4. Obtención de tablas con los 5 países con la mayor media en el aspecto considerado en cuestión, en el rango de años en cómputo global.

	CARACTERÍSTICA	AÑOS	TABLA RESUMEN												
FINANCIACION	Financiación	2009-2018	<table><tr><th>Pais</th><th>Financiacion_Millones de Euros</th></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>3221999.0</td></tr><tr><td>France</td><td>2437454.9</td></tr><tr><td>United Kingdom</td><td>2335692.0</td></tr><tr><td>Italy</td><td>1454905.0</td></tr><tr><td>Spain</td><td>983567.6</td></tr></table>	Pais	Financiacion_Millones de Euros	Germany (until 1990 former territory of the FRG)	3221999.0	France	2437454.9	United Kingdom	2335692.0	Italy	1454905.0	Spain	983567.6
			Pais	Financiacion_Millones de Euros											
			Germany (until 1990 former territory of the FRG)	3221999.0											
			France	2437454.9											
			United Kingdom	2335692.0											
			Italy	1454905.0											
Spain	983567.6														
MORTALIDAD	Mortalidad Tratable	2011-2017	<table><tr><th>Pais</th><th>Mortalidad_Tratable</th></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>48291.5</td></tr><tr><td>Turkey</td><td>41664.2</td></tr><tr><td>United Kingdom</td><td>34503.1</td></tr><tr><td>Poland</td><td>30129.4</td></tr><tr><td>Italy</td><td>27647.4</td></tr></table>	Pais	Mortalidad_Tratable	Germany (until 1990 former territory of the FRG)	48291.5	Turkey	41664.2	United Kingdom	34503.1	Poland	30129.4	Italy	27647.4
			Pais	Mortalidad_Tratable											
			Germany (until 1990 former territory of the FRG)	48291.5											
			Turkey	41664.2											
			United Kingdom	34503.1											
			Poland	30129.4											
Italy	27647.4														

	Mortalidad Prevenible	2011-2017	<table><tr><td>Pais</td><td>Mortalidad_Prevenible</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>86118.30</td></tr><tr><td>United Kingdom</td><td>58783.20</td></tr><tr><td>Turkey</td><td>54471.30</td></tr><tr><td>France</td><td>54285.55</td></tr><tr><td>Poland</td><td>52341.50</td></tr></table>	Pais	Mortalidad_Prevenible	Germany (until 1990 former territory of the FRG)	86118.30	United Kingdom	58783.20	Turkey	54471.30	France	54285.55	Poland	52341.50
Pais	Mortalidad_Prevenible														
Germany (until 1990 former territory of the FRG)	86118.30														
United Kingdom	58783.20														
Turkey	54471.30														
France	54285.55														
Poland	52341.50														
PERSONAL SANITARIO	Médicos de Atención primaria	2010-2019	<table><tr><td>Pais</td><td>Generalist_medical_practitioners</td></tr><tr><td>France</td><td>95124.2</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>78761.6</td></tr><tr><td>Italy</td><td>53525.8</td></tr><tr><td>United Kingdom</td><td>49969.5</td></tr><tr><td>Turkey</td><td>44233.9</td></tr></table>	Pais	Generalist_medical_practitioners	France	95124.2	Germany (until 1990 former territory of the FRG)	78761.6	Italy	53525.8	United Kingdom	49969.5	Turkey	44233.9
	Pais	Generalist_medical_practitioners													
	France	95124.2													
	Germany (until 1990 former territory of the FRG)	78761.6													
	Italy	53525.8													
	United Kingdom	49969.5													
Turkey	44233.9														
Médicos especialistas	2010-2019	<table><tr><td>Pais</td><td>Specialist_medical_practitioners</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>254541.9</td></tr><tr><td>Italy</td><td>182118.7</td></tr><tr><td>United Kingdom</td><td>129527.4</td></tr><tr><td>Spain</td><td>114547.6</td></tr><tr><td>France</td><td>112469.9</td></tr></table>	Pais	Specialist_medical_practitioners	Germany (until 1990 former territory of the FRG)	254541.9	Italy	182118.7	United Kingdom	129527.4	Spain	114547.6	France	112469.9	
Pais	Specialist_medical_practitioners														
Germany (until 1990 former territory of the FRG)	254541.9														
Italy	182118.7														
United Kingdom	129527.4														
Spain	114547.6														
France	112469.9														
Pediatras	2010-2019	<table><tr><td>Pais</td><td>General_paediatricians</td></tr><tr><td>Italy</td><td>17320.7</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>13454.4</td></tr><tr><td>Spain</td><td>12140.4</td></tr><tr><td>United Kingdom</td><td>9934.2</td></tr><tr><td>France</td><td>7741.6</td></tr></table>	Pais	General_paediatricians	Italy	17320.7	Germany (until 1990 former territory of the FRG)	13454.4	Spain	12140.4	United Kingdom	9934.2	France	7741.6	
Pais	General_paediatricians														
Italy	17320.7														
Germany (until 1990 former territory of the FRG)	13454.4														
Spain	12140.4														
United Kingdom	9934.2														
France	7741.6														
Enfermería	2010-2019	<table><tr><td>Pais</td><td>Enfermeria_Matronas_Assitentes</td></tr><tr><td>France</td><td>1706629.6</td></tr><tr><td>Italy</td><td>1045723.4</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>520139.9</td></tr><tr><td>Albania</td><td>484388.1</td></tr><tr><td>Austria</td><td>484388.1</td></tr></table>	Pais	Enfermeria_Matronas_Assitentes	France	1706629.6	Italy	1045723.4	Germany (until 1990 former territory of the FRG)	520139.9	Albania	484388.1	Austria	484388.1	
Pais	Enfermeria_Matronas_Assitentes														
France	1706629.6														
Italy	1045723.4														
Germany (until 1990 former territory of the FRG)	520139.9														
Albania	484388.1														
Austria	484388.1														
Farmacéuticos	2010-2019	<table><tr><td>Pais</td><td>Farmaceuticos</td></tr><tr><td>Italy</td><td>70073.7</td></tr><tr><td>France</td><td>69569.6</td></tr><tr><td>United Kingdom</td><td>53058.5</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>52160.2</td></tr><tr><td>Spain</td><td>51933.4</td></tr></table>	Pais	Farmaceuticos	Italy	70073.7	France	69569.6	United Kingdom	53058.5	Germany (until 1990 former territory of the FRG)	52160.2	Spain	51933.4	
Pais	Farmaceuticos														
Italy	70073.7														
France	69569.6														
United Kingdom	53058.5														
Germany (until 1990 former territory of the FRG)	52160.2														
Spain	51933.4														
Fisioterapeutas	2010-2019	<table><tr><td>Pais</td><td>Fisioterapeutas</td></tr><tr><td>Spain</td><td>178042.2</td></tr><tr><td>Italy</td><td>118364.3</td></tr><tr><td>France</td><td>82274.2</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>69220.1</td></tr><tr><td>United Kingdom</td><td>53058.5</td></tr></table>	Pais	Fisioterapeutas	Spain	178042.2	Italy	118364.3	France	82274.2	Germany (until 1990 former territory of the FRG)	69220.1	United Kingdom	53058.5	
Pais	Fisioterapeutas														
Spain	178042.2														
Italy	118364.3														
France	82274.2														
Germany (until 1990 former territory of the FRG)	69220.1														
United Kingdom	53058.5														
ACTIVIDADES SANITARIAS	Pacientes en Diálisis	2005-2014	<table><tr><td>Pais</td><td>P.Diálisis/100000 habitantes</td></tr><tr><td>Portugal</td><td>98.284</td></tr><tr><td>Greece</td><td>83.198</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>78.059</td></tr><tr><td>Italy</td><td>74.619</td></tr><tr><td>Poland</td><td>74.506</td></tr></table>	Pais	P.Diálisis/100000 habitantes	Portugal	98.284	Greece	83.198	Germany (until 1990 former territory of the FRG)	78.059	Italy	74.619	Poland	74.506
Pais	P.Diálisis/100000 habitantes														
Portugal	98.284														
Greece	83.198														
Germany (until 1990 former territory of the FRG)	78.059														
Italy	74.619														
Poland	74.506														

RECURSOS AUXILIARES	Pacientes Trasplantados	2005-2014	<table><tr><td>País</td><td>P.Trasplantados/100000 habitantes</td></tr><tr><td>Portugal</td><td>50.915</td></tr><tr><td>Spain</td><td>49.227</td></tr><tr><td>Austria</td><td>48.229</td></tr><tr><td>Finland</td><td>46.087</td></tr><tr><td>Netherlands</td><td>45.197</td></tr></table>	País	P.Trasplantados/100000 habitantes	Portugal	50.915	Spain	49.227	Austria	48.229	Finland	46.087	Netherlands	45.197
	País	P.Trasplantados/100000 habitantes													
	Portugal	50.915													
	Spain	49.227													
	Austria	48.229													
Finland	46.087														
Netherlands	45.197														
Quirófanos	2010-2019	<table><tr><td>País</td><td>Quirofanos</td></tr><tr><td>France</td><td>11096.0</td></tr><tr><td>Italy</td><td>6130.2</td></tr><tr><td>Turkey</td><td>5871.6</td></tr><tr><td>Spain</td><td>4375.6</td></tr><tr><td>United Kingdom</td><td>3775.7</td></tr></table>	País	Quirofanos	France	11096.0	Italy	6130.2	Turkey	5871.6	Spain	4375.6	United Kingdom	3775.7	
		País	Quirofanos												
		France	11096.0												
		Italy	6130.2												
Turkey	5871.6														
Spain	4375.6														
United Kingdom	3775.7														
Consultas Psiquiatría	2010-2019	<table><tr><td>País</td><td>S Psiquiatría</td></tr><tr><td>France</td><td>29163.6</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>18440.0</td></tr><tr><td>Spain</td><td>4013.3</td></tr><tr><td>Austria</td><td>2534.3</td></tr><tr><td>Denmark</td><td>2534.3</td></tr></table>	País	S Psiquiatría	France	29163.6	Germany (until 1990 former territory of the FRG)	18440.0	Spain	4013.3	Austria	2534.3	Denmark	2534.3	
		País	S Psiquiatría												
		France	29163.6												
		Germany (until 1990 former territory of the FRG)	18440.0												
		Spain	4013.3												
Austria	2534.3														
Denmark	2534.3														
Consultas Oncológicas	2010-2019	<table><tr><td>Pais</td><td>S Oncologia</td></tr><tr><td>Spain</td><td>7981.6</td></tr><tr><td>France</td><td>6307.2</td></tr><tr><td>Italy</td><td>2645.5</td></tr><tr><td>Belgium</td><td>1474.9</td></tr><tr><td>Bulgaria</td><td>1376.1</td></tr></table>	Pais	S Oncologia	Spain	7981.6	France	6307.2	Italy	2645.5	Belgium	1474.9	Bulgaria	1376.1	
		Pais	S Oncologia												
		Spain	7981.6												
		France	6307.2												
		Italy	2645.5												
Belgium	1474.9														
Bulgaria	1376.1														
Consultas Geriátricas	2010-2019	<table><tr><td>Pais</td><td>S Geriatria</td></tr><tr><td>Spain</td><td>2326.2</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>2214.7</td></tr><tr><td>Austria</td><td>1832.5</td></tr><tr><td>Denmark</td><td>1832.5</td></tr><tr><td>Hungary</td><td>1832.5</td></tr></table>	Pais	S Geriatria	Spain	2326.2	Germany (until 1990 former territory of the FRG)	2214.7	Austria	1832.5	Denmark	1832.5	Hungary	1832.5	
		Pais	S Geriatria												
		Spain	2326.2												
		Germany (until 1990 former territory of the FRG)	2214.7												
		Austria	1832.5												
Denmark	1832.5														
Hungary	1832.5														
PREVENCIÓN	Detección Cáncer de Mama	2010-2019	<table><tr><td>Pais</td><td>% C.Mama</td></tr><tr><td>Netherlands</td><td>82.150</td></tr><tr><td>Spain</td><td>79.698</td></tr><tr><td>Sweden</td><td>77.258</td></tr><tr><td>Portugal</td><td>76.638</td></tr><tr><td>France</td><td>75.120</td></tr></table>	Pais	% C.Mama	Netherlands	82.150	Spain	79.698	Sweden	77.258	Portugal	76.638	France	75.120
	Pais	% C.Mama													
Netherlands	82.150														
Spain	79.698														
Sweden	77.258														
Portugal	76.638														
France	75.120														
Detección Cáncer Cérvix Uterino	2010-2019	<table><tr><td>Pais</td><td>% C Cérvix</td></tr><tr><td>Luxembourg</td><td>78.790</td></tr><tr><td>Italy</td><td>78.330</td></tr><tr><td>Germany (until 1990 former territory of the FRG)</td><td>78.129</td></tr><tr><td>Czechia</td><td>77.401</td></tr><tr><td>Croatia</td><td>77.042</td></tr></table>	Pais	% C Cérvix	Luxembourg	78.790	Italy	78.330	Germany (until 1990 former territory of the FRG)	78.129	Czechia	77.401	Croatia	77.042	
		Pais	% C Cérvix												
		Luxembourg	78.790												
		Italy	78.330												
		Germany (until 1990 former territory of the FRG)	78.129												
Czechia	77.401														
Croatia	77.042														

Tabla 6. Tabla resumen de Países con la mayor media en Recursos en un periodo de tiempo.

3) Estudio de la Normalidad de cada variable foco de estudio (“value”)

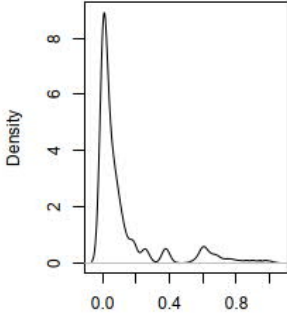
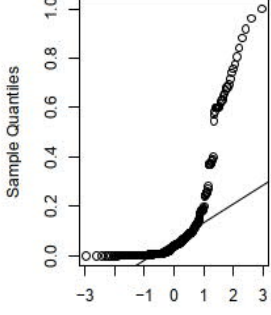
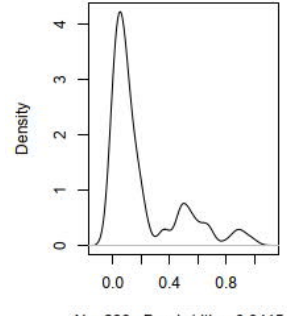
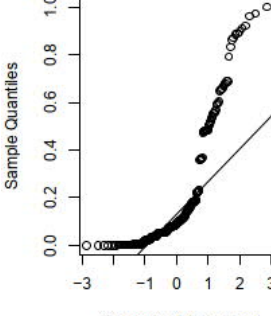
Se ha llevado a cabo el estudio de la normalidad de las variables a estudiar, ya que es importante obtenerla para aplicar los modelos de regresión y contraste de muestras que se van a realizar en el apartado de modelización.

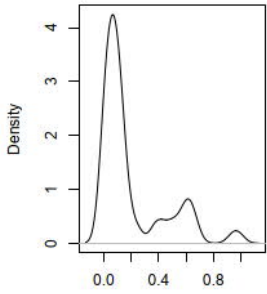
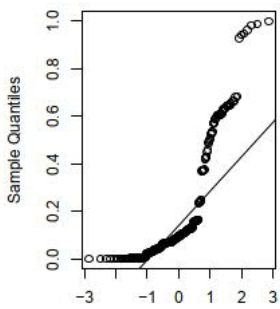
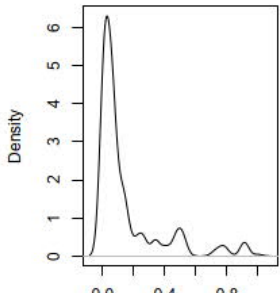
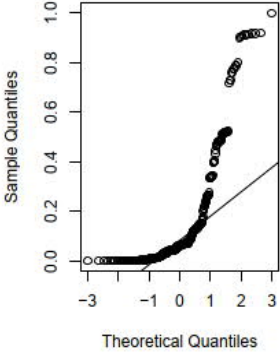
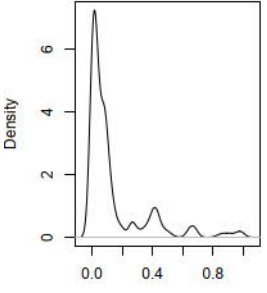
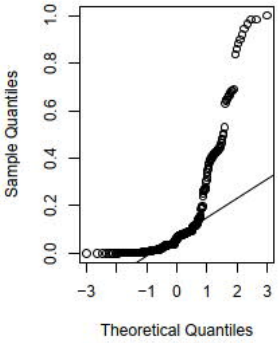
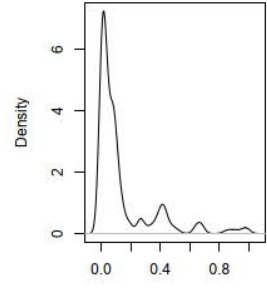
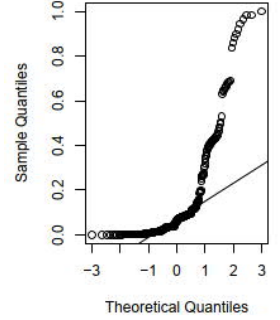
1. Descripción de forma visual de la distribución de característica foco de estudio utilizando gráficas de Densidad y Curvas Q-Q.

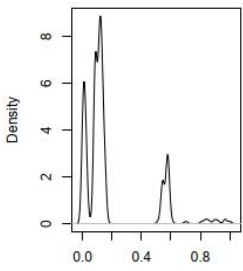
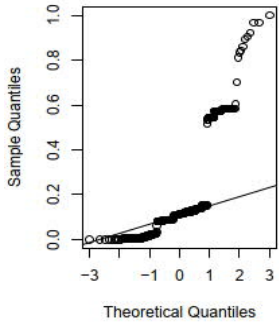
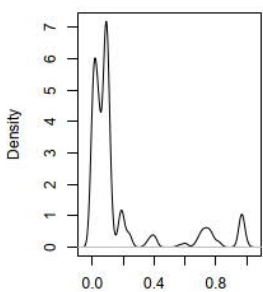
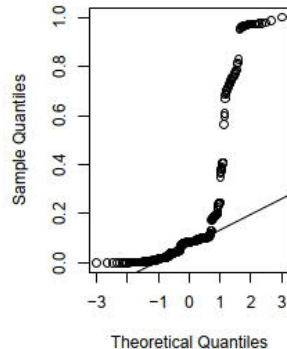
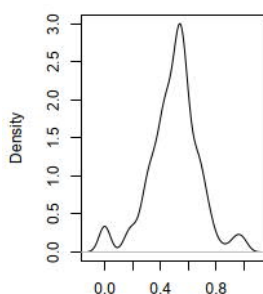
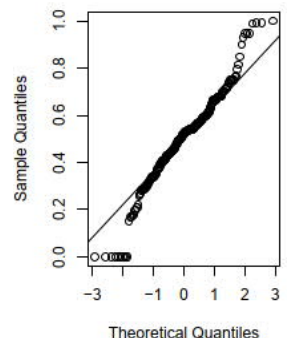
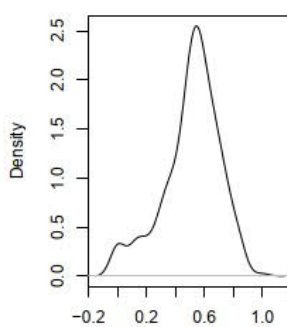
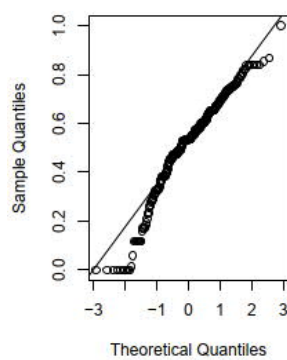
Se han seleccionado para la memoria, ciertas variables a modo de resumen relativas a la: **Financiación, Mortalidad Tratable, Mortalidad Prevenible, Médicos de Atención Primaria, Médicos Especialistas, Pediatras, Personal de Enfermería, Farmacéuticos, Pacientes de Diálisis, Trasplantados, número de Quirófanos, consultas de Psiquiatría y para finalizar pruebas de Detección de Cáncer de Mama y de Cérvix Uterino.** (El estudio completo se encuentra en los anexos)

En primer lugar, definimos que son las gráficas de dispersión y las curvas Q-Q:

- **Gráfica de Dispersión:** diagrama que utiliza las coordenadas cartesianas para mostrar los valores de dos variables.
- **Curvas cuantil-cuantil o Q-Q:** gráfica usada para el diagnóstico entre la distribución de la probabilidad de una población y una distribución usada para la comparación

	Característica	Gráfica Densidad	Curva Q_Q
FINANCIACION	Financiación	<p>Density</p>  <p>N = 330 Bandwidth = 0.02157</p>	<p>Normal Q-Q Plot</p> 
MORTALIDAD	Mortalidad Tratable	<p>Density</p>  <p>N = 238 Bandwidth = 0.04154</p>	<p>Normal Q-Q Plot</p> 

	Mortalidad Prevenible	<p>Density</p>  <p>N = 238 Bandwidth = 0.04388</p>	<p>Normal Q-Q Plot</p> 
PERSONAL SANITARIO	Médicos de Atención primaria	<p>Density</p>  <p>N = 360 Bandwidth = 0.0274</p>	<p>Normal Q-Q Plot</p> 
	Médicos especialistas	<p>Density</p>  <p>N = 360 Bandwidth = 0.0226</p>	<p>Normal Q-Q Plot</p> 
	Pediatras	<p>Density</p>  <p>N = 360 Bandwidth = 0.0226</p>	<p>Normal Q-Q Plot</p> 

ACTIVIDADES SANITARIAS	Enfermería	<p>Density</p>  <p>N = 380 Bandwidth = 0.01129</p>	<p>Normal Q-Q Plot</p> 
	Farmacéuticos	<p>Density</p>  <p>N = 380 Bandwidth = 0.01759</p>	<p>Normal Q-Q Plot</p> 
	Pacientes en Diálisis	<p>Density</p>  <p>N = 290 Bandwidth = 0.04055</p>	<p>Normal Q-Q Plot</p> 
	Pacientes Trasplantados	<p>Density</p>  <p>N = 290 Bandwidth = 0.0509</p>	<p>Normal Q-Q Plot</p> 

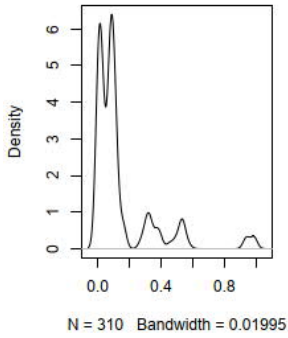
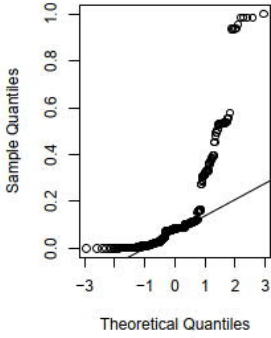
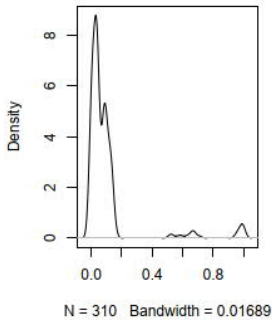
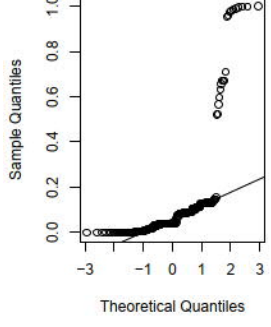
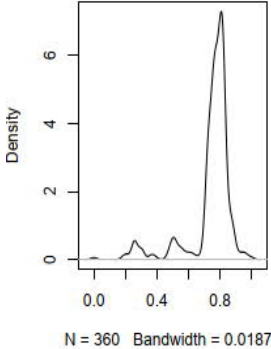
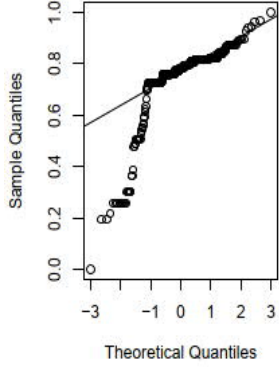
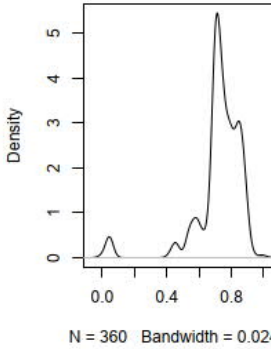
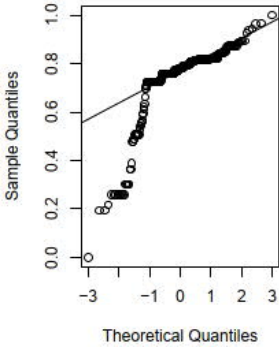
RECURSOS AUXILIARES	Quirófanos	<p>Density</p>  <p>N = 310 Bandwidth = 0.01995</p>	<p>Normal Q-Q Plot</p> 
	Consultas Psiquiatría	<p>Density</p>  <p>N = 310 Bandwidth = 0.01689</p>	<p>Normal Q-Q Plot</p> 
PREVENCIÓN	Detección Cáncer de Mama	<p>Density</p>  <p>N = 360 Bandwidth = 0.01872</p>	<p>Normal Q-Q Plot</p> 
	Detección Cáncer Cérvix Uterino	<p>Density</p>  <p>N = 360 Bandwidth = 0.02407</p>	<p>Normal Q-Q Plot</p> 

Tabla 7. Curvas de Densidad y Q-Q

2. Se realizan pruebas de Hipótesis

Una **prueba de Hipótesis** es una regla que especifica si se puede aceptar o rechazar una afirmación dependiendo de la evidencia proporcionada por una muestra de datos.

Esta regla, estudia dos hipótesis opuestas: **una nula y otra alternativa**.

La hipótesis nula es la afirmación “sin efecto”, lo que queremos desacreditar.

Sin embargo, la hipótesis alternativa es la que responde a nuestra pregunta, la que se establece en base a la evidencia que tenemos. En este caso:

- **HO (Hipótesis Nula):** La muestra proviene de población normal
- **H1 (Hipótesis Alternativa):** la muestra no proviene de una población normal.
- Se confirma normalidad aplicando el test **Shapiro-Wilk**: test que se utiliza para contrastar normalidad de un conjunto de datos, cuya hipótesis nula plantea que la muestra proviene de una población normalmente distribuida.

El test da como resultado un valor para **p-value**, que es la probabilidad de que un valor estadístico calculado sea posible dada una hipótesis nula cierta. Si este valor es menor que el **nivel de significación**, que es la probabilidad de tomar la decisión de rechazar la hipótesis nula cuando ésta es verdadera (Error tipo I, o falso positivo)

Tras analizar las gráficas de Dispersión y Curvas Q-Q (Recordemos que, **Gráfica de Dispersión**: diagrama que utiliza las coordenadas cartesianas para mostrar los valores de dos variables. **Curvas cuantil-cuantil** o **Q-Q**: gráfica usada para el diagnóstico entre la distribución de la probabilidad de una población y una distribución usada para la comparación) observamos que existe un sesgo asociado en la gráfica de densidad sin simetría y, por otro lado, la línea de puntos en las curvas Q-Q no queda perfectamente alineada con los puntos del gráfico por lo que la muestra **no estaría distribuida normalmente**.

Independientemente de la interpretación de las curvas, al aplicar el test Shapiro-Wilk, en todos los casos observamos un **p-valor** inferior al nivel de significación (cualquiera, ya que encontramos valores muy pequeños de p-value) por lo que rechazamos la hipótesis nula, y podemos asumir **NO Normalidad** en todas las variables estudiadas.

3.4 Modelización: Modelos de Regresión.

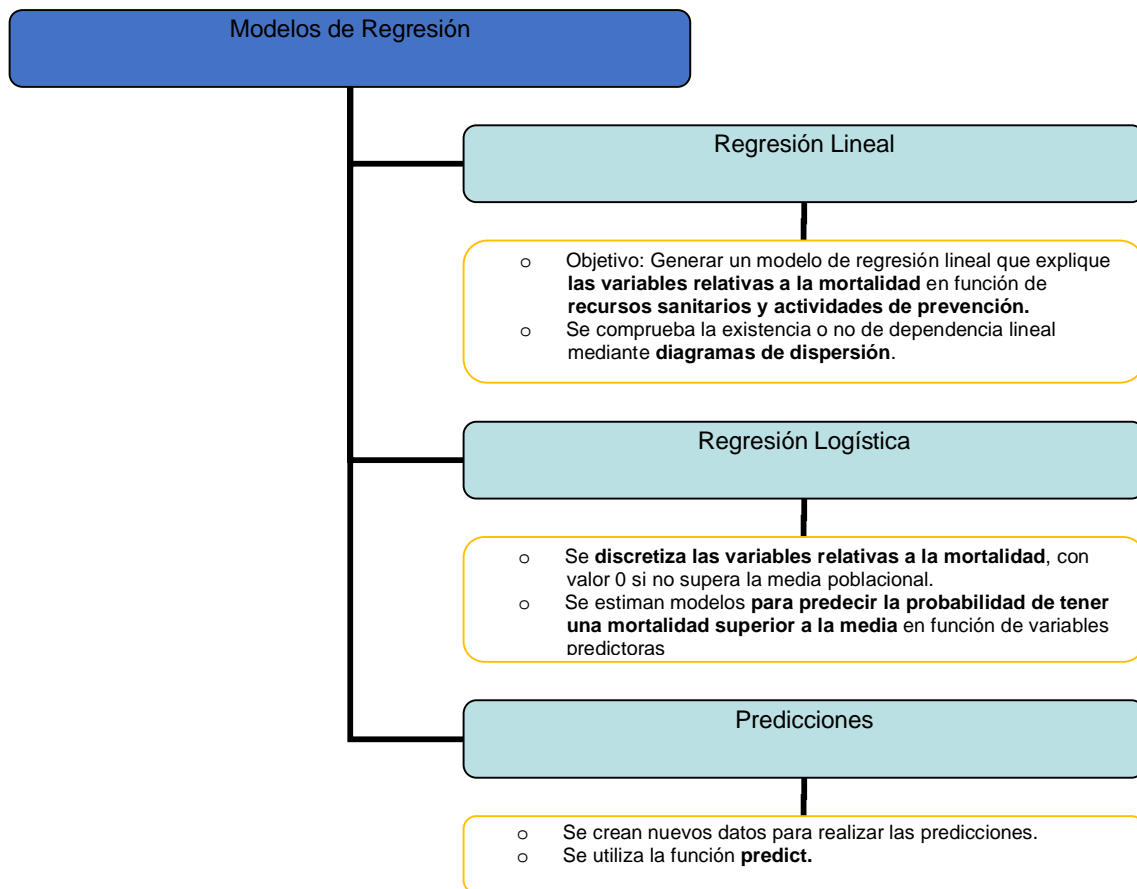


Fig. 9 Modelos de Regresión

1) Análisis Descriptivo. Filtrado de la información.

Una vez analizado la información sobre las causas de Mortalidad, Recursos, Prevención, Actividades, Estado de Salud, Cuidados Domiciliarios y Gastos Sanitarios de forma individual por años y países, se procede a establecer relaciones entre las distintas variables (se ha escogido los 5 países con mayor gasto y recursos sanitarios para hacer el estudio de forma conjunta).

Se procede a estimar un modelo de regresión que explique unas variables en función de otras para después poder hacer predicciones sobre hipótesis de datos nuevos, por ejemplo, si existe relación entre la mortalidad prevenible de un país y el número de médicos disponibles y qué mortalidad cabría esperar si aumentamos o disminuimos estos recursos.

El filtrado de la información se ha realizado seleccionando los países foco de estudio (**Reino Unido, Alemania, Italia, Francia y España**)

1. Regresión Lineal

Se analiza en primer lugar si tiene sentido estimar un modelo de regresión lineal simple entre variables y qué variables serían las adecuadas.

- La relación entre el predictor y la variable respuesta ha de ser lineal. Se puede comprobar con **diagrama de dispersión y graficando los residuos**.
- Los residuos deben distribuirse normalmente en torno a 0. Se puede comprobar con la distribución de cuantiles (normal Q-Q plot) o test de hipótesis de normalidad.
- La variabilidad de los residuos debe ser constante (homocedasticidad): la variabilidad de los datos entorno a la recta de regresión ha de ser aproximadamente constante. Se comprueba graficando los residuos.

Se ha utilizado los **diagramas de dispersión** para comprobar la relación entre variables.

El objetivo es la generación de un modelo de regresión lineal que explique la variable “Mortalidad” en función de:

- Médicos de Atención Primaria.
- Unidades de Resonancia Magnética
- Escáneres Tomográficos,
- Actividades de Diagnóstico de Cáncer de Mama
- Diagnóstico de Cáncer de Cérvix Uterino
- Vacunación

Es decir, si la mortalidad evitable (tratable o prevenible) puede explicarse en función de variables relacionadas con la prevención, por un lado disposición de médicos de atención primaria al ser el primer contacto del paciente ante cualquier dolencia, con variables relacionadas con la disposición de recursos tecnológicos de diagnóstico como puedan ser unidades de resonancia magnética y escáneres tomográficos y sobre todo con actividades de prevención relacionadas con el diagnóstico de cáncer de mama , cervix uterino y campañas preventivas de vacunación a la población mayor de 65 años.

- En primer lugar, se comprueba si la generación del modelo lineal tiene sentido, es decir, si existe una relación lineal entre las variables a estudiar.
 - La comprobación se realiza mediante **diagramas de dispersión** (tipo de diagrama que utiliza las coordenadas cartesianas para mostrar los valores de dos variables de un conjunto de datos).

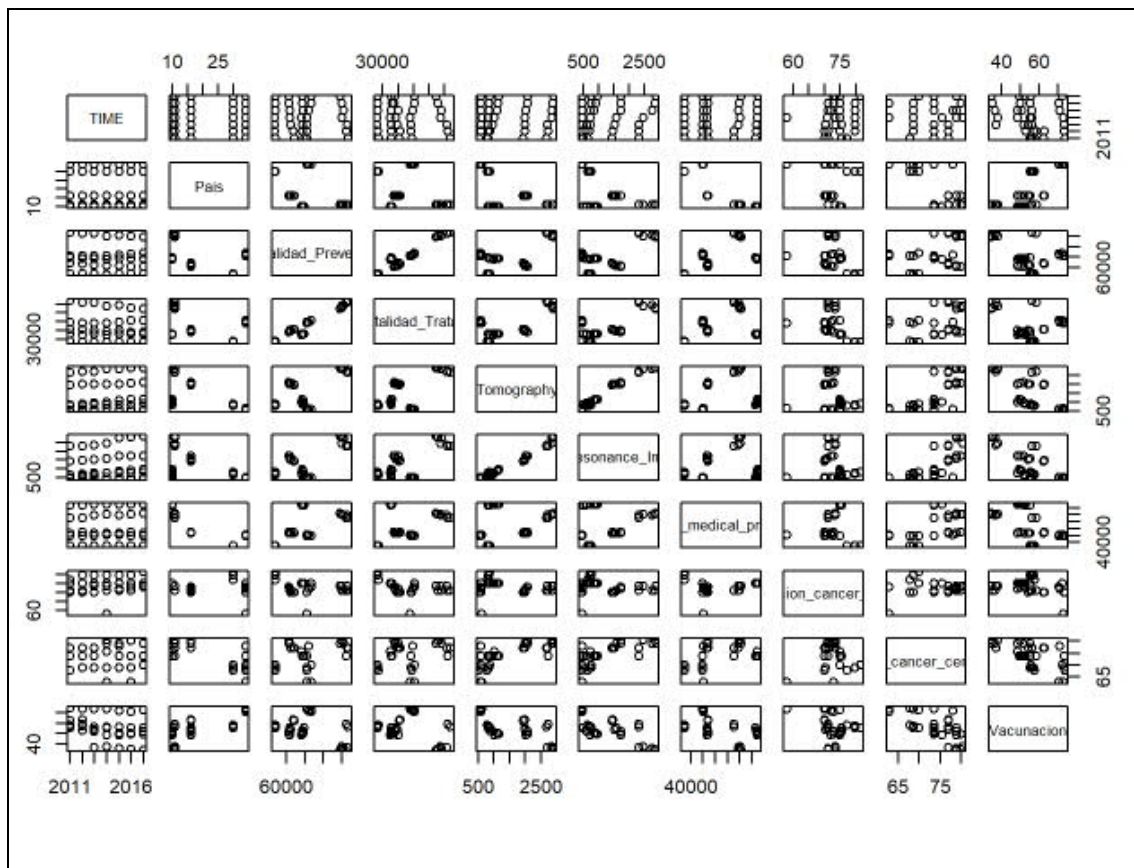


Fig. 12. Diagramas de Dispersión para comprobar dependencia lineal

- Se comprueba que no existe una relación lineal entre variables por lo que se concluye que **la estimación de modelos de regresión lineales no tiene sentido.**
- De haberse podido realizar la estimación, se hubiera generado los modelos para después:
 - Comprobar si el valor de **R-squared** (el coeficiente de determinación ajustado) indica si existe una fuerte dependencia lineal entre las variables estudiadas. Cuanto mayor sea el coeficiente de determinación, mayor dependencia entre variables.
 - Realizar predicciones de los modelos generados sobre nuevos datos para obtener información valiosa, como por ejemplo la variabilidad de la mortalidad en función de incrementos de gastos o personal hospitalario. Posteriormente se hubieran calculado intervalos de confianza sobre las predicciones realizadas.

2. Regresión Logística.

Se procede a estimar un modelo de regresión logística que estime la probabilidad de una variable cualitativa (se transforma las variables **Mortalidad_Tratable** y **Mortalidad_Prevenible** en variable cualitativa, con valores de 0 si el país no supera la media muestral para esa variable y 1 en caso contrario), en función de variables cuantitativas como son Médicos de

Atención Primaria, y de Recursos como disposición de Unidades de Resonancia Magnética y escáneres para diagnóstico, además de las actividades de prevención relacionadas con la detección del cáncer de mama, de cérvix uterino y campañas de vacunación.

- Se discretiza las variables **Mortalidad Prevenible y Mortalidad Tratable** a 0 si no supera la media poblacional de la mortalidad y 1 en caso contrario.
- Se **estiman modelos de regresión logística** para predecir la **probabilidad** de tener una Mortalidad Tratable o Prevenible que supere la media en función de variables predictoras como:
 - Médicos de Atención Primaria.
 - Unidades de Resonancia Magnética
 - Escáneres Tomográficos
 - Detección Cáncer de Mama
 - Detección Cáncer Cérvix Uterino.
 - Vacunación a mayores de 65 años.

Se crean 7 modelos, variando la combinación de variables:

- Mortalidad tratable en función de los médicos de atención primaria.
- Mortalidad prevenible en función de los médicos de atención primaria.
- Mortalidad tratable en función de medios técnicos como escáneres tomográficos.
- Mortalidad tratable en función de medios técnicos como unidades de resonancia magnética.
- Mortalidad tratable en función de unidades de resonancia magnética, escáneres tomográficos, y médicos de atención primaria.
- Mortalidad prevenible en función de unidades de resonancia magnética, escáneres tomográficos, pruebas de detección de cáncer de mama y cérvix uterino
- Mortalidad prevenible en función de escáneres, pruebas de detección de cáncer de mama, cérvix uterino, y vacunación.

Se observa que tanto el protocolo de Detección de Cáncer de Mama, Cérvix Uterino, Vacunación y la adquisición de Escáneres tomográficos disminuyen la probabilidad del aumento de la mortalidad Tratable y Prevenible.

Sin Embargo, las Unidades de Resonancia Magnética y el refuerzo de Medicina de atención primaria, incrementan la probabilidad de aumento de mortalidad. Pero el incremento es tan insignificante (coeficientes muy próximos a cero) que la interpretación más correcta sería indicar que no interfieren en el aumento de la probabilidad de aumentar la mortalidad.

3. Predicciones.

Se propone contestar a las siguientes preguntas:

- ¿Se superaría la Mortalidad Prevenible, si se aumentan las pruebas diagnósticas de detección de Cáncer de Mama y Cérvix, el porcentaje de población vacunada y el número de escáneres?

- ¿Se superaría la Mortalidad Prevenible, si se disminuyen las pruebas diagnósticas de detección de Cáncer de Mama y Cérvix, el porcentaje de población vacunada y el número de escáneres?

Para ello:

- Se crean datos nuevos sobre los que realizar predicciones utilizando los modelos generados.
 - Disminuyendo Recursos
 - Aumentando Recursos
- Se utiliza la función **predict** y se calculan probabilidades.

Se observa:

- Si se aumentan los recursos, la probabilidad de que aumente la mortalidad prevenible es muy baja.
- Por el contrario, si se disminuyen, la probabilidad es máxima.

3.5 Modelización: Contraste de muestras.

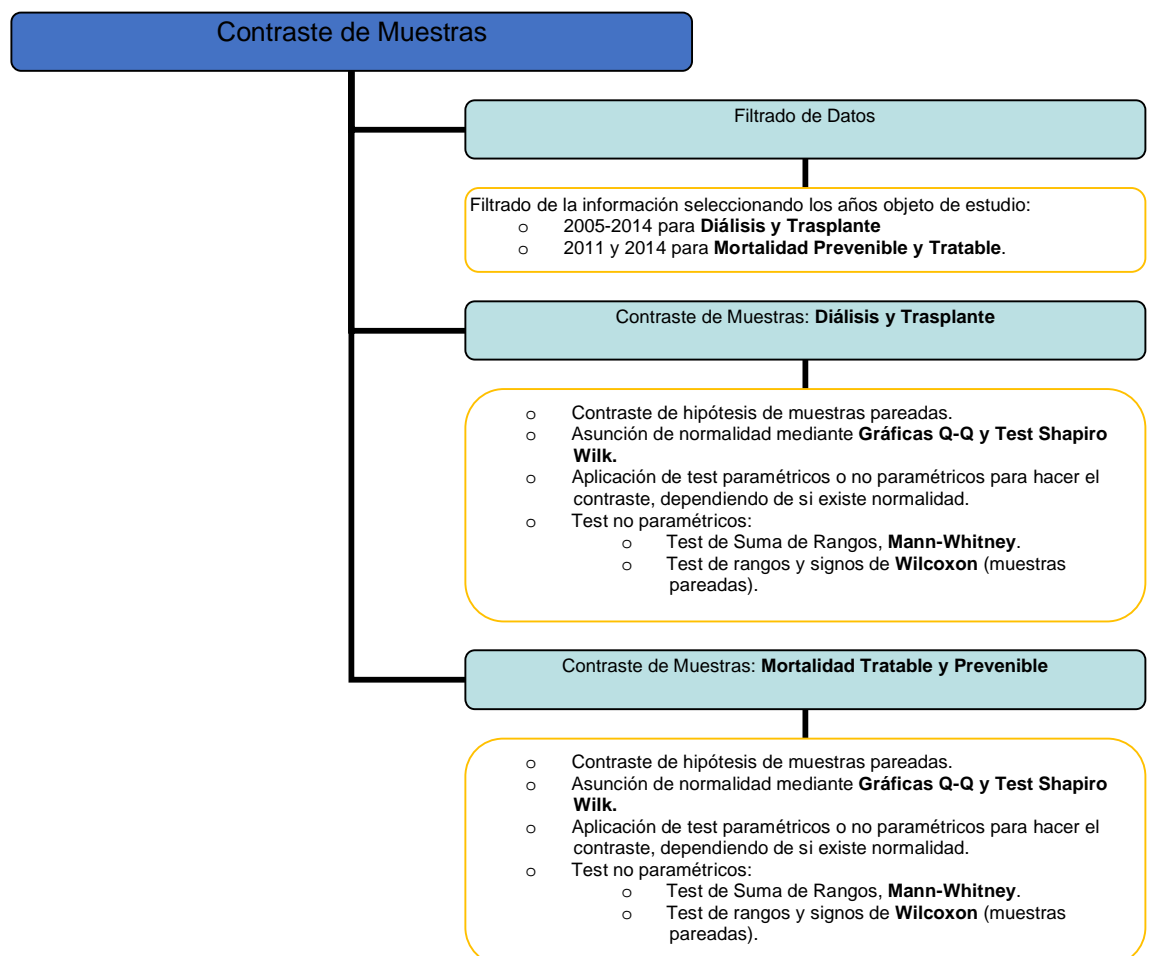


Fig. 10. Contraste de Muestras

1. Análisis Descriptivo. Filtrado de la información.

Se crean **muestras pareadas (mismos países, en distintos años)** con la información de la ratio de enfermos en diálisis y trasplantados en 2005, primer año de información disponible y en 2014, último año de información disponible, **con el fin de poder afirmar o no, si la ratio de pacientes en diálisis o trasplantados aumentarán con los años.**

2. Contraste de Muestras. Diálisis y Trasplante.

Como es una muestra pareada, se calculará la diferencia de muestras y por tanto las hipótesis son:

- $H_0: \mu_{dif}=0$
- $H_1: \mu_{dif}>0$
- Asunción de Normalidad:
 - No se puede aplicar el Teorema del límite Central ya que la muestra no supera los 30 registros.
 - Se interpretan las gráficas Q_Q y se aplica el test de Shapiro-Wilk.
 - Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente.

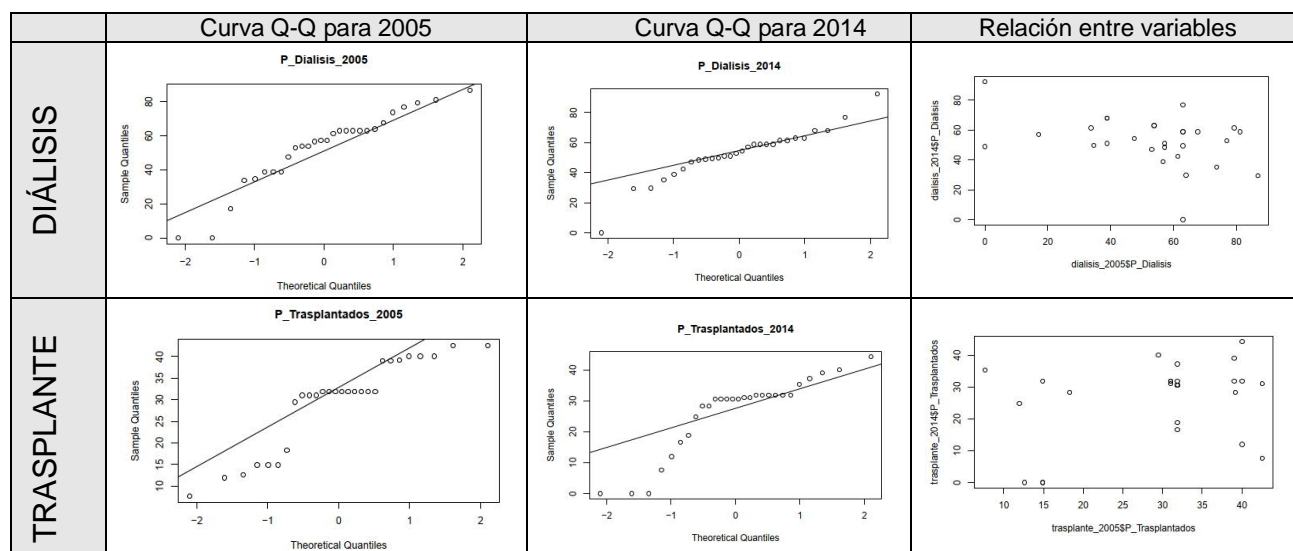


Fig. 13. Curvas Q-Q y relación entre variables.

- **Se utilizan test no paramétricos** para realizar el contraste ya que no puede asumirse normalidad en ninguna muestra, atendiendo al ajuste de los puntos de la muestra a la línea en las curvas Q-Q y al resultado de los test Shapiro-Wilk con p-value muy pequeños.
 - De entre los test no paramétricos tenemos:
 - Test de suma de rangos **Mann-Whitney**.
 - Test de rangos y signos de **Wilcoxon** (muestras pareadas).
- Por lo que usaremos el **de rangos y signos de Wilcoxon** (prueba no paramétrica para comprobar el rango medio de dos muestras relacionadas y determinar si existen diferencias entre ellas.)

Resultados:

- **Diálisis.**

Los gráficos Q-Q muestran una desviación respecto de la normal pronunciada en los extremos sobre todo para la muestra de 2014 y en general para la de 2005.

Por otro lado, de los resultados del test Shapiro-Wilk, con un valor de confianza del 95% podemos rechazar la hipótesis nula de normalidad en la muestra de Diálisis del año 2005 ($p\text{-value} = 0.0251$).

A continuación, y tras realizar el test Wilcoxon, no podemos rechazar la hipótesis nula ($p\text{-value}=0.4233>0.05$), por lo tanto, no podemos afirmar que la ratio de pacientes en diálisis aumente con los años.

- **Trasplante.**

Los gráficos Q-Q muestran una desviación respecto de la normal pronunciada en los extremos tanto para 2005 como para 2014.

Por otro lado, de los resultados del test Shapiro-Wilk, con un valor de confianza del 95% podemos rechazar la hipótesis nula de normalidad en las dos muestras de trasplante ($p\text{-value}<0.05$).

A continuación, y tras realizar el test Wilcoxon, no podemos rechazar la hipótesis nula ($p\text{-value}=0.069>0.05$), por lo tanto, no podemos afirmar que la ratio de pacientes trasplantados aumente con los años.

3. Contraste de Muestras. Mortalidad Tratable y Prevenible.

Como es una muestra pareada, se calculará la diferencia de muestras y por tanto las hipótesis son:

- $H_0: \mu_{\text{dif}}=0$
- $H_1: \mu_{\text{dif}}>0$
- Asunción de Normalidad:
 - No se puede aplicar el Teorema del límite Central ya que la muestra no supera los 30 registros.

- Se interpretan las gráficas Q_Q y se aplica el test de Shapiro-Wilk.
- Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente.

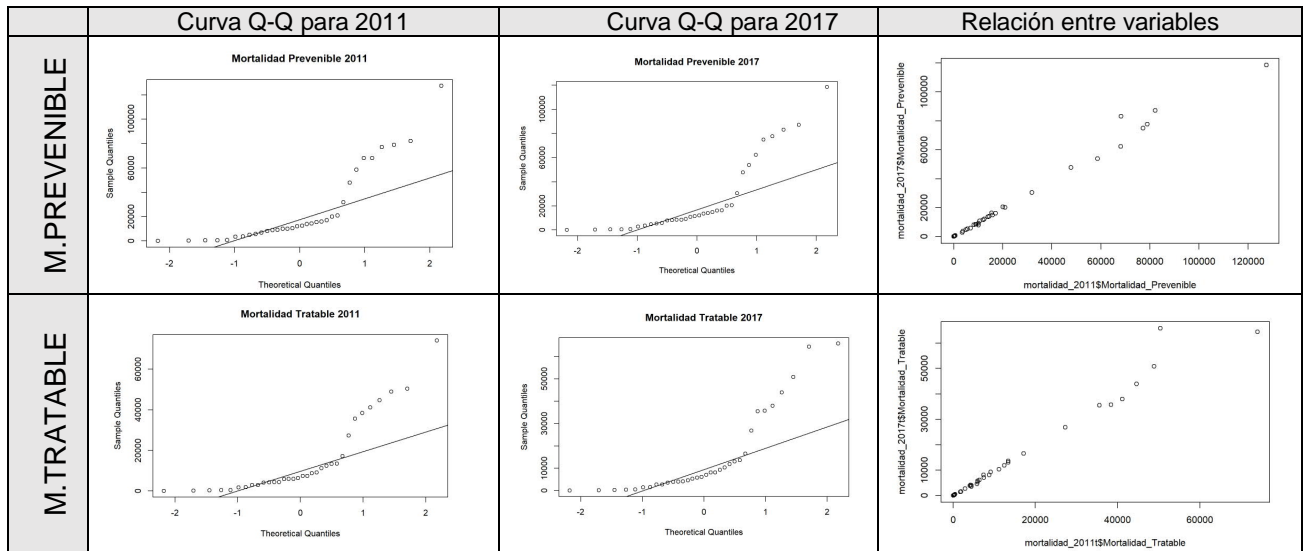


Fig. 14. Curvas Q-Q y relación entre variables.

- **Se utilizan test no paramétricos** para realizar el contraste ya que no puede asumirse normalidad en ninguna muestra, atendiendo al ajuste de los puntos de la muestra a la línea en las curvas Q-Q y al resultado de los test Shapiro-Wilk con p-value muy pequeños por lo que usaremos el test **de rangos y signos de Wilcoxon**.

Resultados:

- **Mortalidad Prevenible.**

Los gráficos Q-Q muestran una desviación respecto de la normal pronunciada en los extremos para ambos años 2011 y 2017.

Por otro lado, de los resultados del test Shapiro-Wilk, con un valor de confianza del 95% podemos rechazar la hipótesis nula de normalidad en ambas muestras (p-value < 0.05).

A continuación, y tras realizar el test Wilcoxon, podemos rechazar la hipótesis nula (p-value < 0.05), por lo tanto, podemos afirmar que la mortalidad prevenible aumenta con los años.

- **Mortalidad Tratable.**

Al igual que en el caso anterior, se rechaza la hipótesis de normalidad en ambas muestras con un valor de confianza del 95%, y tras obtener un p-value en el test de Wilcoxon inferior a 0.05, podemos afirmar que la mortalidad tratable también aumenta con los años.

3.6 Modelización Avanzada de los datos

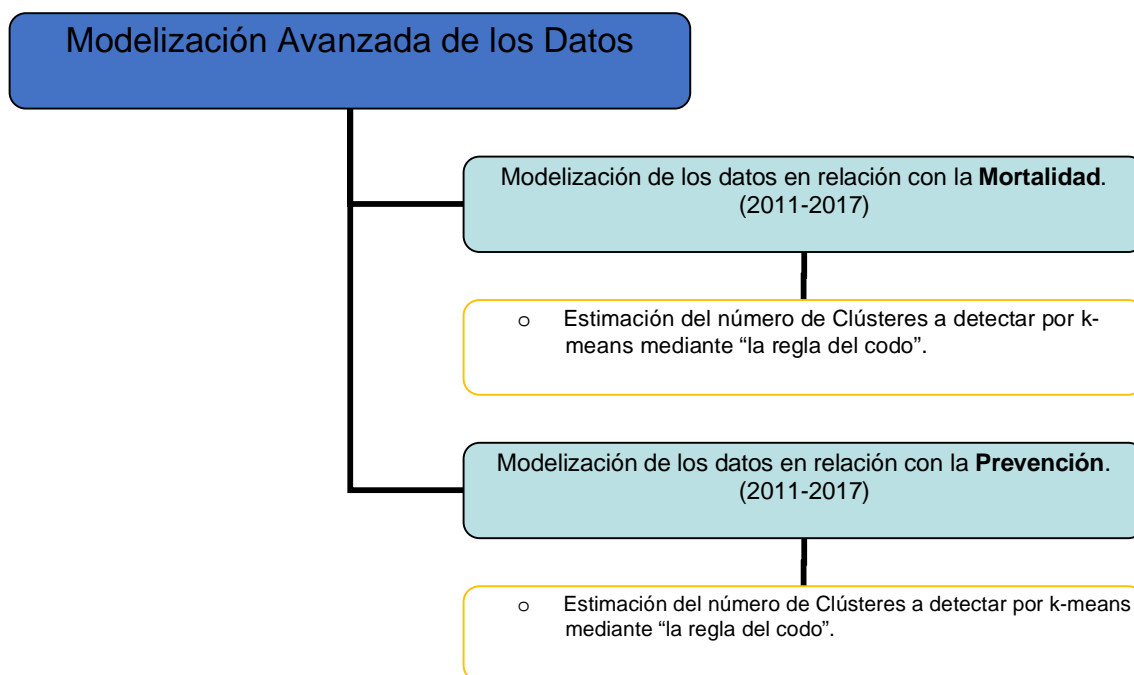


Fig. 11 Modelización Avanzada de los Datos

Se establecen patrones entre los países europeos teniendo en cuenta el índice de mortalidad (tratable, prevenible y por neumonía) y la prevención (Detección de cáncer de mama, de cervix uterino y vacunación) llevada a cabo en los años 2011-2017.

Para ello utilizamos el algoritmo **K-means**:

K-means es un algoritmo de clasificación no supervisada que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o clúster.

El algoritmo consta de tres pasos:

- **Inicialización:** una vez escogido el número de grupos, k, se establecen k centroides en el espacio de los datos de forma aleatoria.
- **Asignación objetos a los centroides:** cada objeto de los datos es asignado a su centroide más cercano.
- **Actualización centroides:** se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso.

Para más información, véase información sobre *El algoritmo k-Means aplicado a clasificación y procesamiento de imágenes* [\[11\]](#)

1. Análisis de la información.

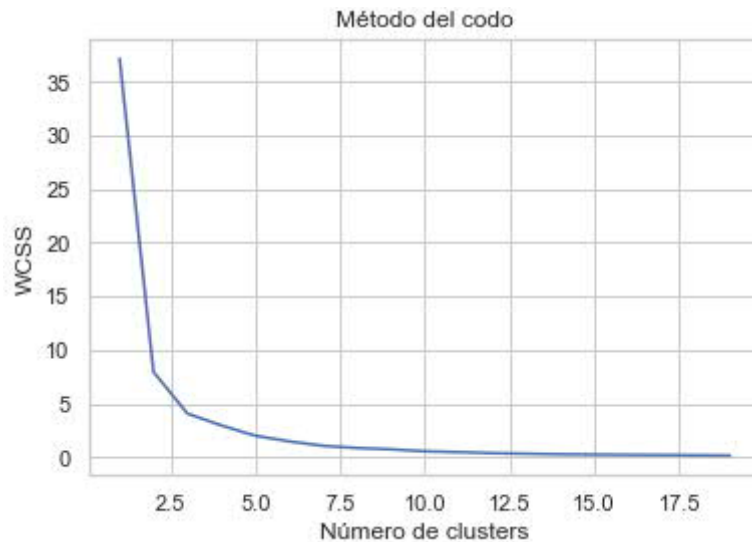
- **Normalización** de las variables para que tengan el **mismo peso** (aplicando la función `MinMaxScaler()` de la librería `sklearn` de Python).
- Para estimar el número de Clústeres a detectar por el algoritmo K-means, se utiliza “**la regla del codo**”:
 - Primero es necesario calcular la suma de los errores cuadráticos (**SSE, Sum of squared of errors**) que consiste en la suma de todos los errores (distancia de cada punto a su centroide asignado) al cuadrado.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} euclidean(x, c_i)^2$$

- Donde K es el número de clústeres a buscar por *k-means*, $x \in C_i$ son los puntos que pertenecen a i-ésimo clúster, c_i es el centroide del clúster C_i (al que pertenece el punto x), y *euclidean* es la distancia euclídea.
- Este procedimiento realizado para cada posible valor k, resulta en una función monótona decreciente, donde el eje X representa los distintos valores de k y el eje Y el SSE. Intuitivamente se observa un significativo descenso del error, que indicará el valor idóneo de k.

2. Modelización. Mortalidad.

- Modelización de datos para agrupar los Países por similitud en relación con la Mortalidad (2011-2017), empleando modelos no supervisados como K-means.
 - Agrupamos la información en relación con la mortalidad prevenible, tratable y por neumonía para establecer patrones con respecto a estas tres variables. (qué patrones siguen los países en relación con la mortalidad evitable y una enfermedad infecciosa común como puede ser la neumonía)
 - Al aplicar la regla del codo observamos que el número de clústeres idóneo es 3.



- Se establecen patrones de similitud por años:

Para 2011-2015:

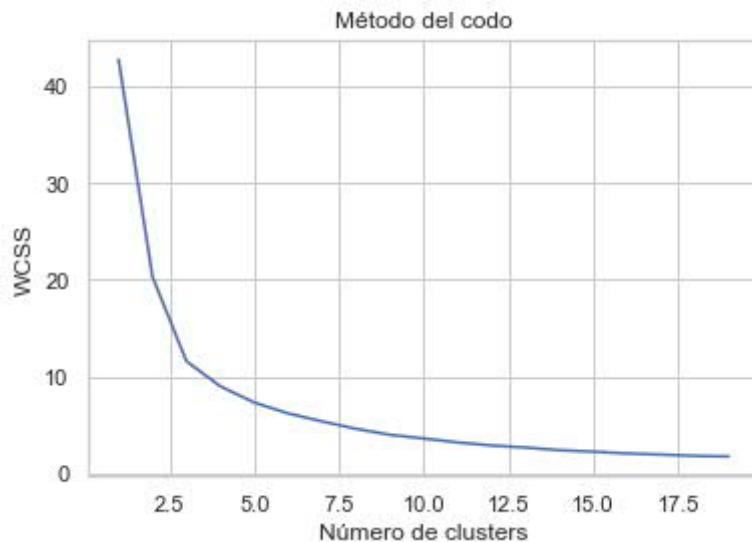
- **Países agrupados en 0:**
['Belgium', 'Bulgaria', 'Czechia', 'Denmark', 'Estonia', 'Ireland', 'Greece', 'Croatia', 'Cyprus', 'Latvia', 'Lithuania', 'Luxembourg', 'Hungary', 'Malta', 'Netherlands', 'Austria', 'Portugal', 'Slovenia', 'Slovakia', 'Finland', 'Sweden', 'Iceland', 'Liechtenstein', 'Norway', 'Switzerland', 'Serbia']
- **Países agrupados en 1:**
['Spain', 'France', 'Italy', 'Poland', 'Romania', 'Turkey']
- **Países agrupados en 2:**
['Germany (until 1990 former territory of the FRG)', 'United Kingdom']

Para 2016-2017:

- **Países agrupados en 0:**
['Belgium', 'Bulgaria', 'Czechia', 'Denmark', 'Estonia', 'Ireland', 'Greece', 'Croatia', 'Cyprus', 'Latvia', 'Lithuania', 'Luxembourg', 'Hungary', 'Malta', 'Netherlands', 'Austria', 'Portugal', 'Slovenia', 'Slovakia', 'Finland', 'Sweden', 'Iceland', 'Liechtenstein', 'Norway', 'Switzerland', 'Serbia']
- **Países agrupados en 1:**
['Spain', 'France', 'Italy', 'Poland', 'Romania']
- **Países agrupados en 2:**
['Germany (until 1990 former territory of the FRG)', 'United Kingdom', 'Turkey']

3. Modelización. Prevención.

- Modelización de datos para agrupar los Países por similitud en relación con la **Prevención** (2011-2017), empleando modelos no supervisados como K-means.
 - Agrupamos la información en relación con la detección del cáncer de mama, cervix uterino y vacunación para establecer patrones con respecto a estas tres variables.
 - Al aplicar la regla del codo observamos que el número de clústeres idóneo es 3.



- Se establecen patrones de similitud por años:

Para 2011:

- **Países agrupados en 0:**

['Czechia', 'Estonia', 'Croatia', 'Latvia', 'Lithuania', 'Hungary', 'Romania', 'Slovenia', 'Slovakia', 'Liechtenstein', 'Norway', 'Montenegro', 'North Macedonia', 'Serbia']

- **Países agrupados en 1:**

['Belgium', 'Bulgaria', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Cyprus', 'Luxembourg', 'Malta', 'Netherlands', 'Austria', 'Poland', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Switzerland', 'United Kingdom']

- **Países agrupados en 2:**

['Turkey']

Para 2012:

- **Países agrupados en 0:**

['Czechia', 'Estonia', 'Croatia', 'Latvia', 'Lithuania', 'Hungary', 'Romania', 'Slovenia', 'Slovakia', 'Norway', 'Montenegro', 'North Macedonia', 'Serbia']

- **Países agrupados en 1:**

['Belgium', 'Bulgaria', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Cyprus', 'Luxembourg', 'Malta', 'Netherlands', 'Austria', 'Poland', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Switzerland', 'United Kingdom']

- **Países agrupados en 2:**

['Turkey', 'Liechtenstein', '']

Para 2013:

- **Países agrupados en 0:**

['Bulgaria', 'Czechia', 'Estonia', 'Croatia', 'Cyprus', 'Latvia', 'Lithuania', 'Hungary', 'Austria', 'Poland', 'Romania', 'Slovenia', 'Slovakia', 'Liechtenstein', 'Norway', 'Montenegro', 'North Macedonia']

- **Países agrupados en 1:**

['Belgium', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Luxembourg', 'Malta', 'Netherlands', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Switzerland', 'United Kingdom']

○ **Países agrupados en 2:**

['Serbia', 'Turkey']

Para 2014:

○ **Países agrupados en 0:**

['Czechia', 'Croatia', 'Cyprus', 'Latvia', 'Hungary', 'Austria', 'Poland', 'Slovenia', 'Slovakia', 'Norway', 'Montenegro', 'North Macedonia']

○ **Países agrupados en 1:**

['Belgium', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Luxembourg', 'Malta', 'Netherlands', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Switzerland', 'United Kingdom']

○ **Países agrupados en 2:**

['Bulgaria', 'Estonia', 'Lithuania', 'Romania', 'Liechtenstein', 'Serbia', 'Turkey']

Para 2015:

○ **Países agrupados en 0:**

['Czechia', 'Estonia', 'Croatia', 'Latvia', 'Lithuania', 'Hungary', 'Romania', 'Slovenia', 'Slovakia', 'Liechtenstein', 'Norway', 'Montenegro', 'North Macedonia', 'Serbia']

○ **Países agrupados en 1:**

['Belgium', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Cyprus', 'Luxembourg', 'Malta', 'Netherlands', 'Austria', 'Poland', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Switzerland', 'United Kingdom']

○ **Países agrupados en 2:**

['Bulgaria', 'Turkey']

Para 2016:

○ **Países agrupados en 0:**

['Czechia', 'Estonia', 'Croatia', 'Latvia', 'Hungary', 'Romania', 'Slovenia', 'Slovakia', 'Liechtenstein', 'Montenegro', 'North Macedonia', 'Serbia']

○ **Países agrupados en 1:**

['Belgium', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Cyprus', 'Luxembourg', 'Malta', 'Netherlands', 'Austria', 'Poland', 'Portugal', 'Finland', 'Sweden', 'Iceland', 'Norway', 'Switzerland', 'United Kingdom']

○ **Países agrupados en 2:**

['Bulgaria', 'Lithuania', 'Turkey']

Para 2017:

○ **Países agrupados en 0:**

['Czechia', 'Estonia', 'Croatia', 'Latvia', 'Lithuania', 'Hungary', 'Romania', 'Slovenia', 'Slovakia', 'Liechtenstein', 'Montenegro', 'North Macedonia', 'Serbia']

○ **Países agrupados en 1:**

['Belgium', 'Denmark', 'Germany (until 1990 former territory of the FRG)', 'Ireland', 'Greece', 'Spain', 'France', 'Italy', 'Cyprus', 'Luxembourg', 'Malta',

'Netherlands', 'Austria', 'Poland', 'Portugal', 'Finland', 'Sweden', 'Iceland',
'Norway', 'Switzerland', 'United Kingdom']
○ **Países agrupados en 2:**
['Bulgaria', 'Turkey']

4. Conclusiones

Fue importante conocer los estudios previos sobre la calidad asistencial sanitaria europea, sobre todo para identificar la medida óptima que acreditara si la calidad asistencial es buena o mala. Se ha concluido que **no existe un único criterio** y que la calidad asistencial sanitaria responde a la interacción de un conjunto de dimensiones como son la **científico-técnica** relacionada con el diagnóstico, tratamiento y destrezas médicas, **interpersonales** relativas a la confianza y empatía, y por otro lado **ambientales, sociales y económicas** relativas esta última al número de recursos (personal, auxiliares y tecnológicos).

El objetivo principal de este trabajo ha sido descubrir la situación real del sistema sanitario en relación con:

- La evolución de la **financiación** en el tiempo por países.
- **Recursos disponibles** (referente tanto al personal como a recursos auxiliares).
- **Actividades sanitarias** como pueden ser tratamiento de diálisis y trasplantes.
- **Estrategias de prevención** como pueden ser realización de pruebas de detección del cáncer de mama , de cérvix uterino y vacunación a la población mayor de 65 años.
- **Mortalidad Evitable**, agrupada en **prevenible o tratable**, o por **infecciones por neumonía**.

Además, se ha estudiado más variables como **Cuidados Domiciliarios, Estado de Salud**, que en las conclusiones no detallaremos ya que no han sido utilizadas en los modelos de regresión ni de clasificación para descubrir dependencias o clasificar países por similitud.

La información ha sido analizada con la siguiente motivación:

- Comprobar la **evolución de la financiación** en el tiempo.
- Plasmar la cantidad de **médicos de atención primaria** como punto de referencia en cualquier estrategia de prevención, y de **unidades de resonancia magnética y escáneres** para poder llevarla a cabo.
- Analizar la **estrategia de prevención** que tienen un efecto directo en las curvas de supervivencia de enfermedades como el cáncer de Mama , y cérvix uterino, además de describir la logística en relación con la vacunación a la población vulnerable mayor de 65 años.
- Describir información sobre la **mortalidad evitable** de los países, agrupada en prevenible y tratable y su evolución en el tiempo.

Del análisis podemos concluir:

1. Los países con una **financiación** mayor (financiación media en los años 2009-2018) en el área sanitaria son por orden decreciente **Alemania, Francia, Reino Unido, Italia y España**:
 - **Alemania** con 3.221.999 millones de euros y una población de 83.157.000 habitantes, tendría una financiación de **3875 millones de euros /100000 hab.**

- **Francia** con 2.437.454,9 millones de euros y una población de 67.320.216 habitantes, financiaría el gasto sanitario con **3621 millones de euros/100000 hab.**
- **Reino Unido** con 2.335.692 millones de euros y una población de 66,65 millones de habitantes, tendría destinado al gasto sanitario **3504 millones de euros/100000 hab.**
- **Italia** con 1.454.905 millones de euros y una población de 59.641.488 habitantes, destinaría al gasto sanitario **2439 millones de euros/100000 hab.**
- **España** con 983.567,6 millones de euros y una población de 47.332.614 habitantes, financiaría el gasto sanitario con **2077 millones de euros/100000 hab.**

Se observa que la financiación aumenta en el tiempo siendo para el caso de:

- **Alemania** un aumento del **28,35%** desde 2009 a 2018.
- **Francia** un aumento del **18%**.
- **Reino Unido** un aumento del **5%**.
- **Italia** un aumento del **6.79%**.
- **España** un aumento del **9.92%**

2. En relación con los **recursos disponibles**:

- **Personal Sanitario (2009-2018):**
 - Países con más **médicos de atención primaria** en orden descendente: **Francia** (141/100000 habitantes), **Alemania** (94,71 /100000 hab), **Italia** (89,74/100000 hab), **Reino Unido** (74,97/100000 hab) y **Turquía** (53,94/100000 hab).
La evolución en el tiempo de los médicos en atención primaria en Europa experimenta poco cambio en los primeros años del rango estudiado(2009-2018), aunque experimenta una leve subida a partir de 2015.
- **Tecnología Médica (2009-2018):**
 - Países con más **Unidades de Resonancia Magnética** son: **Alemania** (6,2/100000 hab), **Italia** (5,36/1000000 hab), **España** (3.01/100000 hab) , **Francia** (2,24/100000 hab), y **Turquía** (1,07/100000 hab). Se observa un aumento progresivo en el tiempo.
 - Países con más **Escáneres** son: **Alemania** (6,75/100000 hab), **Italia** (6,7/1000000 hab), **España** (3.49/100000 hab) , **Francia** (3,04/100000 hab), y **Turquía** (1,56/100000 hab). Al igual que las unidades de resonancia magnética se observa un aumento progresivo en la adquisición de esta tecnología.
- **Estrategia de prevención (2010-2019)**
 - Países con un porcentaje más alto en pruebas de **detección del Cáncer de mama** son por orden descendente: **Holanda**(82,150%), **España** (79,69%), **Suecia** (77,25%), **Portugal** (76,63%) y **Francia** (75,12%).
 - Países con un porcentaje más alto en pruebas de **detección del Cáncer de Cérvix Uterino** son por orden descendente: **Luxemburgo**(78,7%), **Italia** (78.33%), **Alemania** (78,12%), **República Checa** (77,40%) y **Croacia** (77,04%).

- Países con un mayor porcentaje en población mayor de 65 años **vacunada** son por orden descendente: **Reino Unido** (72,45%), **Holanda** (68,04%), **Irlanda** (59,02%), **Bélgica** (58,11%) y **España** (56,05%)
 - **Mortalidad (2011-2017):**
 - Países con mayor número de **fallecimientos tratables**: **Polonia** (79,37/100000 hab), **Alemania** (58,07/100000 hab), **Reino Unido** (51,76/100000 hab), **Turquía** (50,80/100000 hab), **Italia** (46,35/100000 hab).
 - Países con mayor número de **fallecimientos prevenibles**: **Polonia** (137,89/100000 hab), **Alemania** (103,56/100000 hab), **Reino Unido** (88,19/100000 hab), **Francia** (80,63/100000 hab), **Turquía** (66,42/100000 hab).
3. Estudio de los **modelos de regresión, contraste de muestras y clasificación de países** por similitud.
- **Modelos de regresión.**
 - No se observa dependencia lineal entre variables por lo que la estimación de un modelo de regresión lineal no está justificada.
 - En relación con los modelos de regresión logística, se han estimado 6 modelos tratando de calcular las probabilidades de que la mortalidad tratable o prevenible aumente , si aumentamos o disminuimos recursos o actividades como médicos de atención primaria, unidades de resonancia magnética, escáneres, y pruebas de detección de cáncer de mama , cervix uterino y vacunación a población vulnerable.
Se observa que tanto el protocolo de Detección de Cáncer de Mama, Cérvix Uterino, Vacunación y la adquisición de Escáneres tomográficos disminuyen la probabilidad del aumento de la mortalidad Tratable y Prevenible. Sin Embargo, las Unidades de Resonancia Magnética y el refuerzo de Medicina de atención primaria, incrementan la probabilidad de aumento de mortalidad. Pero el incremento es tan insignificante (coeficientes muy próximos a cero) que el aumento en la probabilidad es muy pequeño.
 - **Contraste de muestras.**

Tras analizar las muestras de la mortalidad tratable y prevenible (2011y 2017) y de los pacientes dializados y trasplantados (2005 y 2014) se concluye que:

 - No podemos afirmar que la ratio de pacientes en diálisis y trasplantados aumente con los años. Este dato facilita la gestión del tratamiento de este tipo de pacientes ya que la previsión de necesidades no cambia con los años.
 - Podemos afirmar que la mortalidad tratable y prevenible aumenta con los años. Un mal dato, ya que evidencia una mala estrategia preventiva que no se corrige con los años.
 - **Clasificación de países por similitud.**
 - En relación con la **mortalidad**, hay pocos cambios a lo largo de los años en relación con los países que conforman grupos. Alemania se distancia del resto de países, al igual que España, Francia, Italia,

Polonia y Turquía. El resto de países se clasificaría en un grupo distinto.

- En relación con **actividades preventivas**, los grupos de similitud van cambiando a lo largo de los años, dato que indica al menos cambios en la estrategia de prevención con el fin de mejorarla. Turquía estaría en un grupo distinto a lo largo de los años de estudio (2011-2017), el grueso de los países del Este de Europa en otro grupo (como Croacia, Polonia, Rumanía, Bulgaria etc.), y Europa Central, Norte y Sur Oeste en otro grupo (Alemania, Portugal, España, Italia etc.)

4. Estrategia de mejora.

Debido a que la mortalidad evitable va aumentando con los años, y que el aumento de recursos y pruebas de detección de enfermedades y una buena estrategia preventiva en relación con la vacunación por ejemplo haría que disminuyera, todos los países tendrían que poner foco en este aspecto para cumplir por un lado con la dimensión científico-técnica de la calidad asistencial sanitaria.

5. Glosario

EFQM: The European Foundation for Quality Management.

k-NN: k-Nearest Neighbors algorithm

Python: Lenguaje de programación interpretado, dinámico y multiplataforma.

R: Lenguaje de programación con un enfoque al análisis estadístico.

Sklearn: librería de Python

SSE: Sum of squared of errors.

6. Bibliografía

- [1] Villegas Perriñán María del Mar, Rosa Díaz Isabel María. **La calidad Asistencial: concepto y medida**. 2003. Recuperado de:
[La calidad asistencial: concepto y medida | Villegas Perriñán | Dirección y Organización \(revistadyo.es\)](http://www.calidadasistencial.es/images/gestion/biblioteca/39.pdf)
- [2] López-Coronado Miguel, De la Torre Isabel, Herreros Gonzales Jesús, Cabo Salvador Javier. **Mejora de la calidad asistencial mediante la telemedicina y la teleasistencia**. Ediciones Díaz de Santos.2014. Capítulo 24.
- [3] Mira José Joaquín, Lorenzo Susana, Rodríguez-Marín Jesús, Aranz Jesús, Sitges Esher. **La aplicación del modelo europeo de gestión de la Calidad Total al Sector Sanitario: Ventajas y Limitaciones**.1998. Recuperado de:
<http://www.calidadasistencial.es/images/gestion/biblioteca/39.pdf>
- [4] Nogué-Xarau S, Amigó-Tadín M, Sánchez-Sánchez M y Salmerón Bargo JM. **Evaluación y seguimiento de la calidad asistencial ofrecida a los intoxicados en un Servicio de Urgencias**.2007. Recuperado de:
<https://www.redalyc.org/pdf/919/91924105.pdf>
- [5] Eurostat Statistics Explained. **Health statistics at regional level.2021**. Recuperado de:
https://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level#Health_status_and_health_care.
- [6] EU HEALTH. **State of Health in the EU 2017. 5 ways to make healthcare effective, accessible and resilient**. 2017. Recuperado de:
https://twitter.com/eu_health/status/933684222614753280
- [7] Howard, Simon John. **Comparing health protection systems across six European nations and the USA**. Recuperado de:
https://www.researchgate.net/publication/286611133_Comparing_health_protection_systems_across_six_European_nations_and_the_USA
- [8] European Environment Agency. **Vocabulary: Classification of health care financing schemes- SHA 2011**. Recuperado de:
https://dd.eionet.europa.eu/vocabulary/eurostat/icha11_hf/view
- [9] National Cancer Institute. **Common Cancer Types**. Recuperado de:

<https://www.cancer.gov/types/common-cancers>

- [10] Gaceta Médica. Más de 2000 mujeres son diagnosticadas de cáncer de cérvix cada año en España. 2021. Recuperado de: <https://gacetamedica.com/investigacion/mas-de-2-000-mujeres-son-diagnosticadas-de-cancer-de-cervix-cada-ano-en-espana/>
- [11] Algoritmo K-means aplicado a la clasificación y procesamiento de imágenes. Recuperado de: https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.htm
- Expansión. Datosmacro. UE - Unión Europea. Economía y datos de los países. Recuperado de: [UE - Unión Europea 2021 | datosmacro.com \(expansion.com\)](https://datosmacro.com/expansion.com)

6. Anexos

Los datos utilizados para el análisis se encuentran disponibles en el siguiente enlace:

<https://ec.europa.eu/eurostat/web/health/data/database>

En el propio repositorio de GitHub, se ha creado una carpeta con los datos procesados para el análisis.

Toda la documentación generada durante la fase de diseño e implementación: ficheros pdf y Rmd generados en el estudio estadístico con RStudio en lenguaje R y con Anaconda en lenguaje Python se encuentra alojada en el repositorio GitHub:

[aperdices-g/TFM---Calidad-Asistencial-Sanitaria-Europea \(github.com\)](#)

Esquema del repositorio: **TFM Calidad Asistencial Sanitaria Europea.**

1. **Análisis Exploratorio de Datos (EDA)**
 - a. Análisis Exploratorio por Países.
2. **Datos_Disponibles.** Conjunto de datos utilizados para el análisis.
3. **Limpieza de Datos (ETL)**
 - a. En relación con Actividades
 - b. Causas de Mortalidad
 - c. Cuidados Domiciliarios
 - d. Estados de Salud por Sexos
 - e. Gasto Sanitario
 - f. Prevención
 - g. Recursos Sanitarios
4. **Memoria**
5. **Modelización de datos**
 - a. Modelización Avanzada de Datos
 - b. Modelos de Regresión.
 - c. Contraste de Muestras.