



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**ESTUDIO Y VISUALIZACIÓN DE DATOS RELATIVOS AL
MOVIMIENTO NATURAL DE LA POBLACIÓN
ESPAÑOLA**

Autor:

Raúl Gutiérrez Martín

Tutores:

Alfonso Gordaliza Ramos

Francisco Rodríguez Redondo

Año: 2022

AGRADECIMIENTOS

Quiero agradecer el apoyo que he recibido por todas las personas cercanas a mí, en especial de mi familia, de Esther y de mis compañeros, ya que sin alguno de ellos no habría podido llegar hasta aquí.

También destacar la ayuda que me han proporcionado mis dos tutores Alfonso Gordaliza y Francisco Rodríguez por el apoyo y la aportación de conocimientos, así como al Instituto Nacional de Estadística por la aportación de datos relevantes.

Índice general

1. Introducción	7
1.1. Objetivo	9
2. Movimiento Natural de la Población	10
2.1. Proceso Estadístico	10
2.2. Nacimientos	11
2.3. Matrimonios	11
2.4. Defunciones	11
2.5. Acceso a los microdatos	12
3. Marco teórico	14
3.1. Tasas y probabilidades	14
3.1.1. Tasa bruta de mortalidad	15
3.2. Estandarizaciones	15
3.2.1. Estandarización directa o población-tipo	15
3.2.2. Estandarización indirecta o mortalidad-tipo	16
3.3. Tabla de mortalidad	17
3.3.1. Funciones de la tabla de mortalidad	17
3.4. Mortalidad infantil	20
3.4.1. Mortalidad infantil y causas de muerte	20
3.4.2. Medidas de mortalidad infantil	21
4. Base de datos	22
4.1. Fuentes de información y obtención de datos	22
4.1.1. Datos cartográficos	22
4.1.2. Datos padrón continuo	23
4.1.3. Datos defunciones	23
4.1.4. Datos sobre natalidad	23
4.1.5. Datos mujeres	24
4.2. Depuración y adecuación de los datos	24
4.2.1. Datos cartográficos	24
4.2.2. Datos padrón continuo	25
4.2.3. Datos defunciones	26
4.2.4. Datos natalidad	27
4.2.5. Datos mujeres	27

5. Creación de la aplicación	28
5.1. Datos geográficos	28
5.1.1. mapa.españa_prov y mapa.españa_com	29
5.2. Unificación de los datos	29
5.2.1. TBM_prov y TBM_com	30
5.2.2. TMI_prov y TMI_com	31
5.2.3. Causas de muerte	32
5.2.4. defunciones_prov y defunciones_com	33
5.2.5. padron_prov y padron_com	33
5.3. Librerías de R	34
5.3.1. 'simple features'	34
5.3.2. 'ggplot2'	35
5.3.3. 'ggiraph'	36
5.3.4. 'gt'	37
6. TAMEApp	39
6.1. Estructura Shiny	40
6.2. Estructura gráfica	42
6.2.1. Mensaje emergente	42
6.2.2. División pantalla	43
6.2.3. Utilización de pestañas	52
7. Conclusiones y trabajos futuros	57
Bibliografía	59

RESUMEN

Nos encontramos actualmente en una sociedad en la que las nuevas tecnologías, la búsqueda de información y la representación gráfica de la misma, están a la orden del día. Es por ello que la creación de aplicaciones que faciliten al usuario la consulta de información gana cada vez más terreno a los clásicos libros de consulta.

Dada la situación actual de pandemia en la que nos encontramos en el momento de la realización de este trabajo, ha cobrado más importancia si cabe, o mayor curiosidad, conocer las cifras de mortalidad que se producen en la sociedad. Motivado por este interés social y en base a la gran importancia que tiene el estudio de la evolución de la mortalidad en base a sus causas y territorios, el propósito de este Trabajo Fin de Grado consiste en desarrollar una aplicación Shiny en R para ofrecer una herramienta intuitiva, dinámica y rápida con la que acceder a los diferentes datos de mortalidad desde 2003 a 2020. Y detallar la evolución de la mortalidad en el territorio español, centrandolo el interés en la escala provincial y por comunidades autónomas por medio de mapas interactivos.

Palabras clave: Mortalidad, población, demografía, R, Shiny, tasas demográficas

ABSTRACT

We are currently in a society in which new technologies, the search for information and its graphical representation, are in everyday life. That is why the creation of applications that make it easier for the user to consult information is gaining more and more space over the classic reference books.

In the current pandemic situation in which we find ourselves at the time of carrying out this work, it has become even more important, or increased curiosity, to know the mortality figures that occur in society. Motivated by this social interest and based on the great importance of studying the evolution of mortality based on its causes and territories, the purpose of this Final Degree Project is to develop a Shiny application in R to offer a tool intuitive, dynamic and fast to have access to the different mortality data from 2003 to 2020. And detail the evolution of mortality in the Spanish territory, focusing on the provincial scale and by autonomous communities through interactive maps.

Keywords: Mortality, population, demography, R, Shiny, demographic rates

Capítulo 1

Introducción

Los registros de mortalidad no son algo nuevo o de corta trayectoria, sino que se remontan muchos años atrás. Se considera a Graunt (1662) uno de los pioneros en estudiar la mortalidad desde un punto de vista estadístico, aunque su estudio fue únicamente descriptivo. El primer acercamiento analítico a los cambios en la mortalidad se lo podemos atribuir a Malthus (1766- 1834), quien puso en relación la población con los recursos necesarios para mantenerla, atribuyendo así el concepto de “subsistencia” como causa de variación de la población.

Muchas han sido las investigaciones y conclusiones extraídas en base al estudio de la mortalidad desde los inicios, pero, ¿cómo nos condiciona la evolución de la mortalidad?

La información estadística obtenida sobre los fallecimientos a nivel nacional, o desagregada a niveles territoriales inferiores, nos permite conocer las causas que provocan dichas defunciones, su evolución temporal, las diferencias territoriales, etc y contribuye de manera decisiva a la toma de decisiones políticas dirigidas a la mejora de los sistemas de salud y de protección social.

Esto queda patente analizando los datos del último siglo, en el que en base a los cambios en la forma de vida, la mejora de los sistemas de salud, el avance en tratamientos... , han permitido que el aumento de la protección social y la esperanza de vida al nacer en nuestro país sea mucho más elevada.

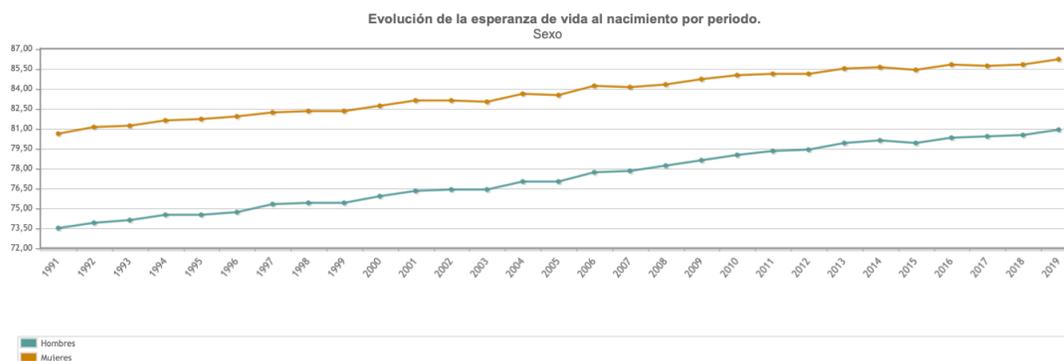


Figura 1.1: Información extraída de la página web del Instituto Nacional de Estadística

Cabe destacar también el periodo de pandemia por el que todavía estamos pasando en el momento de realizar este trabajo, en el que tanto los médicos como los investigadores y analistas de la mortalidad han tenido una gran importancia permitiendo a las administraciones tomar drásticas decisiones para minimizar en lo posible dicha tasa de mortalidad.

Así bien, podemos decir que las estadísticas de mortalidad suponen una fuente de información muy completa e importante sobre la evolución de la salud de la población.

Las reflexiones anteriores motivan la realización de este Trabajo Fin de Grado, en el que se pretende desarrollar una herramienta informática de visualización cartográfica de los indicadores más significativos que se utilizan para describir y analizar la mortalidad como uno de los fenómenos más importantes dentro de lo que se conoce como el movimiento natural de la población.

1.1. Objetivo

Dentro del rango de fenómenos que abarca el denominado movimiento natural de la población, este Trabajo de Fin de Grado se dirige a recopilar y agrupar los datos de las defunciones en España, resumirlos calculando unos estadísticos determinados, y finalmente ofrecer la información a cualquier usuario para conocimiento directo del fenómeno en estudio o bien que pueda ser utilizada para futuras investigaciones o trabajos.

Además, analizando la evolución de las diferentes causas de muerte haremos especial incidencia en el impacto de la pandemia causada por el SARS-COV-2.

Para llevar a cabo todo ello debemos recopilar toda la información necesaria y utilizando el paquete 'Shiny' de 'R', crear una aplicación intuitiva y rápida de utilizar en la cual se reflejen los datos en forma de mapa o tablas estadísticas que resuman la información que queremos mostrar.

Así pues, la componente de visualización cartográfica de la información estadística jugará un peso muy importante en este trabajo, en línea con las tendencias actuales propiciadas por la disponibilidad tecnológica del momento.

Capítulo 2

Movimiento Natural de la Población

El Movimiento Natural de la Población es una estadística realizada por el Instituto Nacional de Estadística (en adelante INE), hace referencia a los nacimientos, defunciones y matrimonios que ocurren en la totalidad del territorio Español. Estos datos se elaboran gracias a la colaboración de las comunidades autónomas y los distintos Registros Civiles, los cuales mensualmente hacen llegar al Instituto Nacional de Estadística sus boletines de parto, defunción y matrimonio.

Para el INE, el Movimiento Natural de la Población supone uno de los trabajos que más tiempo se lleva realizando. El primer volumen que publicó la Junta General de Estadísticas del Reino fue en el año 1863, el cual contenía datos del periodo 1858-1861. Desde esa fecha se han hecho publicaciones ininterrumpidamente a excepción del periodo 1871-1885 sobre los fenómenos demográficos nombrados.

A lo largo de todos estos años lo único que ha cambiado ha sido la fuente de información de la cual provenían los datos, pero los conceptos demográficos siguen siendo prácticamente los mismos.

2.1. Proceso Estadístico

Como se ha comentado al principio, la fuente administrativa que proporciona los datos al Instituto Nacional de Estadística es el Registro Civil.

En el artículo 20 del Reglamento de la Ley del Registro Civil se cita que los encargados comunicarán a los órganos oficiales, sin necesidad de petición especial, los datos exigidos por Ley, Real Decreto o por la Dirección General. De la misma forma remitirán al Instituto Nacional de Estadística, a través de sus Delegaciones y a los servicios de Estadística Municipal, los boletines sobre nacimientos, defunciones, matrimonios y abortos u otros hechos inscribibles.

La recogida de los datos se lleva a cabo rellenando los correspondientes boletines creados por el INE de parto, defunción y matrimonio y remitiéndolos a sus respectivas delegaciones provinciales del INE.

2.2. NACIMIENTOS

Mensualmente, todos los boletines serán corregidos si fuera necesario y grabados para después poder proceder a comunicar las defunciones y los nacimientos tanto al Censo como a cada Ayuntamiento para poder actualizar el Padrón Municipal.

Una vez terminado este proceso se hacen llegar todos los datos a los Servicios Centrales del Instituto Nacional de Estadística donde se estudiarán de una forma más detallada y global.

2.2. Nacimientos

El nacimiento antes del año 1975 estaba considerado según el artículo 30 del Real Decreto de 24 de julio de 1889 como "feto que tuviere figura humana y viviere veinticuatro horas enteramente desprendido del seno materno". A partir de esa fecha, este concepto se modifica y se pasa a identificar el nacimiento con el concepto biológico "nacido en vida", el cual es utilizado en el campo demográfico.

2.3. Matrimonios

La recogida de información de los matrimonios también ha tenido un gran cambio en el año 2005 con la entrada en vigor de la Ley 13/2005 de 1 de julio. Dicha ley permite ahora contraer matrimonio entre personas del mismo o distinto sexo con plenitud e igualdad de derechos y obligaciones. Gracias a esta modificación se analizan las características tanto de matrimonios de distinto sexo como de matrimonios del mismo sexo.

2.4. Defunciones

Debido a la modificación de la ley en 1975 antes nombrada, las defunciones anteriores a esta fecha estaban consideradas como los fallecidos con más de 24 horas de vida. A partir de 1975 este concepto cambia y las defunciones incorporan a los nacidos con vida fallecidos durante las primeras 24 horas. Esta modificación no genera un gran impacto en las cifras brutas pero supone la adopción del criterio de defunción recomendado por los Organismos Internacionales.

Seguramente las defunciones sean el fenómeno demográfico de mayor interés por su importancia en la sociedad. En la recogida de estos datos se tiene muy en cuenta la edad y el año de nacimiento para de esta forma poder calcular la edad en el momento del fallecimiento.

2.5. ACCESO A LOS MICRODATOS

Si se necesitan los datos más detallados y con un mayor número de variables, como es el caso de este trabajo para los datos de mortalidad, se pueden conseguir también gratuitamente haciendo una solicitud y firmando un acuerdo de confidencialidad, eso sí, siempre cumpliendo con la Ley 12/1989 de 9 de mayo de la Función Estadística Pública que obliga al INE a no difundir en ningún caso los datos personales cualquiera que sea su origen (es lo que se conoce como secreto estadístico). Se entiende como datos personales los referentes a personas físicas o jurídicas que permiten la identificación inmediata.

Capítulo 3

Marco teórico

En este estudio, como hemos comentado anteriormente, se trabajará sobre la mortalidad en la geografía española. En el sentido más general, según la Real Academia Española, la mortalidad es la cualidad de mortal que poseen todos los seres vivos, la cual es inevitable, no repetible e irreversible. Todas estas características la hacen que sea un evento muy interesante para su estudio demográfico. En un sentido más específico, la mortalidad es un indicador estadístico de defunciones en un tiempo y espacio determinados.

El número bruto de las defunciones en una determinada población nos aporta poca información. Es por ello, que en este trabajo haremos uso de los principales indicadores los cuales serán presentados a continuación y nos serán útiles para poder comparar poblaciones entre sí de una forma más correcta.

3.1. Tasas y probabilidades

Una tasa es el cociente que resulta de dividir un número de acontecimientos sucedidos durante un periodo de tiempo por la población media existente durante ese periodo. Puede interpretarse como la frecuencia relativa con la que se produce un cierto acontecimiento en relación a la población media existente durante el tiempo que se han registrado los acontecimientos.

Una probabilidad es el cociente entre los acontecimientos experimentados por una población durante un periodo de tiempo y la población inicial de dicho periodo susceptible de experimentar tales acontecimientos. Puede interpretarse como la relación numérica entre los sujetos susceptibles de experimentar un determinado fenómeno y los fenómenos acontecidos después de un cierto tiempo.

3.1.1. Tasa bruta de mortalidad

El nivel de mortalidad depende en gran medida del tamaño de la población que se esté estudiando, por eso, el primer indicador y más básico que utilizaremos será la tasa bruta de mortalidad, la cual es la relación entre el número de defunciones en un año concreto y la población media de dicho año. La media de la población será la semisuma de las poblaciones en el comienzo y final del periodo elegido.

$$TBM = \frac{D^{t,t+1}}{0,5 * (P^t + P^{t+1})} * 1000$$

En realidad, lo que mide la tasa bruta de mortalidad es la frecuencia relativa de las defunciones en una determinada población, de tal modo que esta no es un buen indicador para llevar a cabo comparaciones de mortalidad entre dos poblaciones determinadas, ya que depende de la estructura de la población sobre la que se mide. Por ejemplo, una población podrá tener una menor tasa de mortalidad únicamente porque es una población más joven. Esta será la razón por la cual tendremos que dividir la población en diferentes grupos de edad para su estudio. El procedimiento será el mismo solo que teniendo en cuenta la población y las defunciones de esos grupos de edad, obteniendo de esta manera las denominadas tasas específicas de mortalidad.

3.2. Estandarizaciones

Como ya se ha comentado, no es del todo correcto comparar dos poblaciones directamente por la tasa bruta de mortalidad. La diferenciación en grupos de edad podría solucionar en cierta parte este problema, pero esta diferenciación también tiene sus inconvenientes, y es que se pierde la capacidad de comparar el conjunto de la población, y resultaría mucho más complicado comparar individualmente cada grupo de edad, además de que podríamos llegar a conclusiones muy diferentes. Es por esto por lo que se busca un método para poder comparar una población con otra directamente.

3.2.1. Estandarización directa o población-tipo

Este método consiste en seleccionar una población tipo que tomaremos como referencia, a la que se le aplicarán las estructuras de las poblaciones las cuales queremos comparar. El resultado serán unos valores brutos los cuales son estimaciones que individualmente no nos aportan ningún tipo de información, pero sí que los podemos utilizar para comparar estos valores con los de las diferentes poblaciones que queremos estudiar.

3.2. ESTANDARIZACIONES

Edades	Población Española	Tasa Mortalidad Andalucía	Defunciones Estimadas	Tasa Mortalidad Aragón	Defunciones Estimadas
0	426992	0.01022	4362	0.00838	3578
1-4	1898492	0.00054	1028	0.00044	843
5-9	3091517	0.00029	910	0.00025	765
10-14	3286886	0.00030	997	0.00030	1000
...

Tabla 3.1: Información extraída de: Vinuesa, J.(Ed.).(1994). Demografía, análisis y proyecciones. Editorial Síntesis

A modo de ejemplo ilustrativo de la utilidad de la estandarización, en la tabla anterior podemos observar cómo se están comparando las comunidades autónomas de Aragón y Andalucía teniendo como población-tipo la Española. Así podemos comparar en términos absolutos ambas comunidades, pero como hemos comentado anteriormente el resultado será únicamente comparativo.

El mayor inconveniente de la estandarización directa es la gran cantidad de información que necesitamos conocer y que no siempre se encuentran disponibles.

3.2.2. Estandarización indirecta o mortalidad-tipo

Como referencia se seleccionará una población de la cual vamos a obtener sus tasas de mortalidad. Estas tasas se aplicarán de igual manera a las poblaciones que queramos comparar. Como resultado tendremos unos valores de mortalidad brutos esperados. Estos valores los tendremos que comparar con los reales, para de esta forma tener el Índice de Mortalidad Estándar que será el cociente entre las defunciones observadas y esperadas. Este índice tendrá carácter únicamente comparativo y será especialmente útil para situar la mortalidad de poblaciones pequeñas como municipios.

También podemos utilizar la estandarización indirecta para comparar dos poblaciones sin la necesidad de conocer los datos de mortalidad de estas, solamente necesitaríamos conocer las defunciones de la población tipo de la cual obtenemos las tasas.

Edades	Tasa Mortalidad España	Población Andalucía	Defunciones Estimadas	Población Aragón	Defunciones Estimadas
0	0.00922	96314	888	11575	107
1-4	0.00059	424560	251	51791	31
5-9	0.00028	597748	170	80779	23
10-14	0.00027	623388	170	88763	24
...

Tabla 3.2: Información extraída de: Vinuesa, J.(Ed.).(1994). Demografía, análisis y proyecciones. Editorial Síntesis

En este caso la información que necesitamos es la Tasa Bruta de Mortalidad de España y la población de las comunidades autónomas de Andalucía y Aragón.

3.3. Tabla de mortalidad

La tabla de mortalidad o también denominada tabla de vida, es el mejor instrumento para conocer las condiciones de mortalidad de una población. En este trabajo utilizaremos las denominadas tablas de momento, que como su propio nombre indica, resumirán las condiciones de mortalidad de una población en un momento determinado.

Resultarán de aplicar el patrón de mortalidad de una determinada población sobre una cohorte o generación ficticia de individuos, de esta forma será mucho más fácil la posterior comparación con otras poblaciones.

3.3.1. Funciones de la tabla de mortalidad

La elaboración de una tabla de mortalidad puede llegar a ser muy complicada. Pero existen unos planteamientos básicos que permiten alcanzar una primera aproximación de forma sencilla y al mismo tiempo conceptualmente rigurosa. A continuación presentaremos los elementos que se incluyen en la tabla.

Tasas Específicas por edad

La tasa específica por edad o la tasa bruta de mortalidad será el primer elemento que introduciremos en la tabla de mortalidad y que nos servirá de base para calcular las futuras series.

$${}_n m_x = \frac{{}_n D_x}{{}_n P_x}$$

Será el cociente entre el número bruto de defunciones y la población desde una edad x hasta una edad $x + n$.

Probabilidades de muerte o cocientes de mortalidad por edad

El primer elemento de una tabla de mortalidad está compuesto por las probabilidades de muerte entre edades exactas o cocientes de mortalidad por edad.

3.3. TABLA DE MORTALIDAD

$${}_nq_x = \frac{2n \cdot m_x}{2 + (n \cdot m_x)}$$

El denominador es el efectivo de la población sometido a riesgo al principio del intervalo el cual estamos considerando. La forma más conocida de su cálculo asume un reparto lineal de la mortalidad a lo largo del intervalo.

Probabilidad de supervivencia

$${}_np_x = 1 - {}_nq_x$$

Es el número complementario de la probabilidad de muerte presentada anteriormente. Mide el riesgo de que una persona de una determinada edad continúe con vida después de un número de años concretos.

Supervivencia a edades exactas

$$l_{x+n} = l_x - (l_x \cdot q_x)$$

Las probabilidades de muerte o cocientes de mortalidad, nos hacen posible obtener una serie de supervivientes a edades exactas sucesivas. Se suele utilizar como primera cantidad inicial de la generación l_0 una potencia de base 10, que se denomina raíz de la tabla.

Defunciones entre edades exactas

$${}_nd_x = l_x - l_{x+n}$$

La diferencia entre los supervivientes entre dos edades determinadas, es el número esperado de defunciones que han ocurrido en el intervalo de esas dos edades. En el último grupo de edad, al ser un intervalo abierto, todos los individuos deben morir en él.

Población estacionaria de la tabla

La población estacionaria asociada a la tabla de mortalidad representa el tiempo que vive toda la generación entre las edades x y $x + n$.

$${}_nL_x = n * ((l_n * {}_n a_x) + l_{x+n} * (1 - {}_n a_x))$$

Este valor se calcula como el tamaño del intervalo n por el coeficiente de reparto de defunciones y multiplicado por la suma de los supervivientes a edades exactas x y $x + n$.

Coeficiente de reparto de las defunciones

Indica o estima la fracción de intervalo de tiempo vivida por aquellos que fallecen durante el periodo de tiempo al que hace referencia. Se podrán conocer los valores gracias a los supervivientes a edades exactas y la población estacionaria de la tabla.

$${}_n a_x = \frac{{}_nL_x - (n * l_{n+x})}{l_x - l_{x+n}}$$

En tablas de mortalidad muy elaboradas suelen aplicarse valores de ${}_n a_x$ variables en función de la edad, ya sean estos coeficientes extraídos de los datos o estimados mediante complejos métodos matemáticos. Pero se suele utilizar una simplificación aceptable en términos generales la cual da un valor de ${}_n a_x = 2,5$ si se trabaja con grupos quinquenales de edad, lo que presupone un reparto lineal de las defunciones dentro del intervalo.

Tiempo vivido

$$T_x = \sum_x^{\infty} {}_nL_x$$

Será la suma de la serie desde una edad determinada hasta el final de la tabla. Expresa el tiempo vivido por la población de la edad determinada. El último valor de la tabla será el mismo para T_x que para ${}_nL_x$.

Esperanza de vida a edades exactas o expansión de vida

La esperanza de vida a una determinada edad es uno de los indicadores que más se utiliza cuando hablamos de la mortalidad. Se define como la cantidad media de años que una persona puede esperar vivir dentro de la generación y población a la que pertenece. Aunque el indicador es al fin y al cabo ficticio, es el indicador que mejor imagen ofrece de las condiciones de mortalidad de una población.

$$e_x = \frac{T_x}{l_x}$$

La expresión de la fórmula será la división entre la suma del tiempo vivido en un intervalo y la supervivencia a edades exactas.

Probabilidad perspectiva de paso

Principalmente su utilidad reside en las proyecciones demográficas. El concepto es equiparable con la probabilidad de supervivencia, pero ${}_nZ_x$ se estimará entre intervalos de edad tomando la población estacionaria de la tabla calculada previamente.

$${}_nZ_x = \frac{{}_nL_{x+n}}{{}_nL_x}; Z_N = \frac{L_0 + {}_4L_1}{5 * l_0}; Z_0 = \frac{{}_5L_5}{L_0 + {}_4L_1}; Z_{95} = \frac{L_{100+}}{{}_5L_{95} + L_{100+}}$$

3.4. Mortalidad infantil

El concepto demográfico de mortalidad infantil se refiere estrictamente a las defunciones que ocurren justo desde el nacimiento hasta el primer año de vida.

La mortalidad infantil se utiliza con frecuencia para comprender el nivel de desarrollo y la salud general de una población, ya que muchos factores que contribuyen a las defunciones de los bebés también están influenciados por la salud de la población general, como la salud de los padres, las condiciones higiénicas domésticas, la calidad de la sanidad, etc.

3.4.1. Mortalidad infantil y causas de muerte

Históricamente las causas más comunes de la mortalidad infantil han sido las relacionadas con los agentes externos, factores genéticos y los riesgos correspondientes en el momento del

parto. Pero la participación de estos grupos en el último siglo ha cambiado enormemente en los países más desarrollados.

La disminución de las defunciones debidas a los agentes externos se ha reducido enormemente, siendo ahora las otras causas nombradas las más comunes. La consecuencia de esto es que las defunciones se concentran en los primeros días y semanas de vida. Con carácter general, el descenso de la mortalidad ocurrido en la mayoría de los países en las últimas décadas se ha visto favorecido por la práctica desaparición de la mortalidad infantil.

3.4.2. Medidas de mortalidad infantil

La medida más correcta, como en cualquier grupo de edad, es la que hemos visto previamente, es decir, el cociente entre las defunciones de una determinada población, dividida en este caso por el total de los nacidos en esa misma población.

$$TMI_g = \frac{{}_g d_0}{\text{Nacidos vivos}_g}$$

Una pequeña aclaración en este apartado es que para un año concreto contamos con dos generaciones diferentes, es decir, los nacidos en el año anterior que aún no han cumplido un año de vida en el actual y los nacidos en el propio año de estudio. Esto es difícil de solucionar ya que normalmente no disponemos de los datos por generaciones. En el caso de que esto ocurra no será tan preciso pero sí que se tendrá una buena estimación con los datos en años naturales.

$$TMNeonatal_g = \frac{{}_g d_{0-27}}{\text{Nacidos vivos}_g}$$

$$TMPostneonatal_g = \frac{{}_g d_{28-365}}{\text{Nacidos vivos}_g}$$

La importancia de la separación de la mortalidad infantil en sus componentes neonatal, menos de 27 días de vida y postneonatal, entre 28 días y un año de vida, radica en la diferente naturaleza de las causas que determinan la defunción del niño en cada uno de estos tramos de edades. Por lo tanto, si es posible tener acceso a los datos en los que conocemos los días o meses de vida, es esencial dividir estos datos.

Capítulo 4

Base de datos

La base de datos es el pilar fundamental de cualquier estudio estadístico, y así lo será para este. El trabajo de la creación de la base de datos no consiste únicamente en recopilar toda la información y almacenarla. Es necesario un preprocesado inicial y su correcta organización, no solo para poder trabajar más cómodamente, sino para facilitar el trabajo a los posteriores investigadores que quieran hacer uso de esos datos.

En este capítulo se detallará todo el trabajo que se ha llevado a cabo para poder tener una base de datos con mucha información y bien estructurada.

4.1. Fuentes de información y obtención de datos

Para este trabajo la única fuente de información ha sido el Instituto Nacional de Estadística (INE) quienes nos han facilitado todos los datos necesarios, vía los microdatos que tiene disponible en su página web.

Para poder representar la información en un mapa, lo primero que necesitamos son los datos cartográficos del territorio del cual queremos hacer el estudio. Necesitaremos también los datos protagonistas para este estudio que serán las defunciones en un año natural, los datos del padrón para poder calcular ciertos indicadores interesantes, así como los datos de nacimientos en un año natural y el número de mujeres por provincia.

A continuación se detallarán los conjuntos de datos mencionados y se explicará la metodología para su tratamiento y posterior utilización.

4.1.1. Datos cartográficos

Para poder representar la información en un mapa lo primero que necesitamos son los datos cartográficos del territorio que queremos representar.

4.1. FUENTES DE INFORMACIÓN Y OBTENCIÓN DE DATOS

Los datos cartográficos para este trabajo serán los correspondientes a todo el territorio español, es decir, la Península, Islas Canarias e Islas Baleares, Ceuta y Melilla.

Nos podemos descargar esta información que estará estructurada en siete ficheros que serán necesarios para tener la información geográfica de cada municipio y nos servirán para poder hacer una representación gráfica, además de otras variables de utilidad.

4.1.2. Datos padrón continuo

El Padrón Municipal es un registro administrativo donde constan las personas empadronadas en cada municipio y una serie de datos sociodemográficos asociados a ellas. Las cifras poblacionales, provenientes de la Revisión del Padrón a 1 de enero de cada año, contiene la información básica de la población clasificada según unas determinadas variables.

Para poder calcular muchas de las tasas necesitamos dichos datos del padrón. Estos datos recogen la población que hay en España en un año determinado por cada municipio y grupo quinquenal de edad.

4.1.3. Datos defunciones

Las defunciones ocurridas a lo largo de todo el año están recogidas por el INE, y es posible acceder gratuitamente a la versión reducida de estos datos. Debido al protagonismo que tienen en el trabajo, se solicitó a esta entidad poder tener la versión ampliada de los mismos, la cual contiene a mayores las causas de muerte.

Los datos están recogidos en un fichero de texto en el cual no están clasificados por variables. Se dispone de datos de las defunciones desde el año 1975 hasta el año 2020, este último año coincidiendo con el comienzo de la pandemia mundial ocasionada por el COVID-19.

4.1.4. Datos sobre natalidad

El número de nacimientos que se producen en España también es una información que el INE proporciona de forma pública y de la cual hemos hecho uso. Estos datos contienen información detallada sobre cada uno de los nacimientos.

El fichero de datos obtenido tiene la misma estructura que los datos de mortalidad, y se hará uso de estos datos para poder calcular todos los estadísticos y tasas relacionados con las personas menores de un año de edad.

4.1.5. Datos mujeres

Para poder calcular las tasas de mortalidad de la causa de muerte por 'Embarazo, parto y puerperio' será necesario conocer la población que está a riesgo, en este caso serán las mujeres de edades comprendidas entre los 15 y 49 años. Estos datos los necesitaremos agrupados por provincias y comunidades.

El fichero de datos obtenido nos proporciona estas cifras para cada provincia divididas en grupos quinquenales de edad.

4.2. Depuración y adecuación de los datos

Esta parte es una de las más laboriosas e importantes en el desarrollo de este trabajo ya que son conjuntos de datos con mucha información la cual hay que tratar y clasificar correctamente. Además se ha tenido que trabajar con ficheros en diferentes formatos y con diferente estructura lo cual dificulta la generalización de procesos.

A continuación pasaremos a explicar detalladamente el trabajo que se ha realizado en cada uno de los conjuntos de datos.

4.2.1. Datos cartográficos

Nuestro objetivo con este conjunto de datos será el de obtener finalmente dos mapas de España, uno a nivel provincial y otro a nivel de comunidades autónomas, ambos válidos para poder representar cualquier información en ellos. Para lograrlo, primero tendremos que realizar un estudio de las variables del fichero para su comprensión. Las variables más importantes que utilizaremos serán:

- **CMUN:** Código del municipio dentro de la provincia. Variable numérica de 3 dígitos.
- **CPRO:** Código de provincia. Es una variable numérica que toma los valores desde 1 a 52, con el que se identifican las provincias de España.
- **CCA:** Código de Comunidad Autónoma. Variable numérica que toma los valores desde 1 al 19 con la que se identifican las comunidades de España.
- **GEOMETRY:** Contiene la información geográfica de cada municipio del país.

Una vez que sabemos las variables que nos interesan nos enfrentamos a un documento en el cual cada línea tiene la información referente a un único municipio.

4.2. DEPURACIÓN Y ADECUACIÓN DE LOS DATOS

Por lo tanto, para crear el mapa a nivel provincial, tendremos que agrupar todos los municipios según el código de la provincia 'CPRO', teniendo en cuenta que se tendrá que reestructurar toda la información de la variable 'GEOMETRY', para obtener un único valor representativo de toda la provincia. El resultado de este procedimiento será un nuevo fichero en el que eliminamos la variable 'CMUN' y el cual tendrá una dimensión de 52.

Creado el anterior fichero, tenemos que realizar un procedimiento muy similar para conseguir el mapa a nivel de comunidades autónomas. En este caso tendremos que agrupar todas las provincias según el código de la comunidad 'CCA', de nuevo teniendo en cuenta que habrá que reestructurar las variables 'GEOMETRY' según la comunidad. Obtendremos entonces un fichero que tendrá 19 líneas.

Para una mejor visualización, los ficheros resultantes se han modificado para poder representar las Islas Canarias más cerca de la península, esto ha sido posible modificando la variable 'GEOMETRY' tanto en el mapa de provincias como en el de comunidades autónomas.

Además se ha utilizado una herramienta online (mapshaper) para simplificar las geometrías de los mapas. Gracias a esta simplificación obtenemos un procesamiento mucho más rápido a la hora de representar los datos, en contrapartida, será menos preciso pero es algo que sacrificaremos para una mejor experiencia del usuario final.

4.2.2. Datos padrón continuo

Tendremos ficheros del padrón municipal por cada año, en los que la población ya está agrupada por grupos quinquenales. Las variables que utilizaremos serán:

- **CUMUN:** Variable numérica de cinco dígitos en la que las dos primeras cifras corresponden con el código de la provincia y las tres últimas corresponden con el código del municipio.
- **Grupos Quinquenales:** Tendremos 21 variables que serán la división de cero a cien años cada 5 años, por ejemplo de 0 a 4 años tendrá la cantidad de individuos de esa edad de cada municipio.
- **Total:** Será la población total de cada municipio del país.

Al tener la información municipal, al igual que en el caso de los datos cartográficos, lo primero que tendremos que hacer será agrupar todos los municipios por provincias, para esto nos ayudaremos de la variable 'CUMUN' y los dos primeros dígitos, que hacen referencia al código de la provincia. De esta forma tendremos un archivo de datos que tiene 52 filas correspondientes a las 50 provincias Españolas y las dos ciudades autónomas, Ceuta y Melilla.

4.2. DEPURACIÓN Y ADECUACIÓN DE LOS DATOS

Para poder hacer los cálculos correspondientes con las comunidades autónomas tendremos que crear otro fichero formado por 19 filas, en las que cada una de ellas tendrá la información de la población de las 17 comunidades autónomas y las ciudades autónomas de Ceuta y Melilla.

Este proceso se tendrá que repetir para todos y cada uno de los años, en nuestro caso será desde el año 2003 hasta el año 2020.

4.2.3. Datos defunciones

Los ficheros de las defunciones son anuales y tienen estructuras diferentes a lo largo de los años ya que se va actualizando la forma de recoger la información y actualmente se guardan muchos más datos que los primeros años, aunque las variables en las que tenemos interés son las mismas. Para cada año disponemos de un documento explicativo de todas las variables proporcionado por el INE, y serán:

- **CPRORE:** Código de provincia de residencia del fallecido.
- **ANOSC:** Número de años que ha cumplido la persona fallecida. Lo utilizaremos para poder agrupar a los fallecidos por grupos quinquenales.
- **MESESC:** Número de meses cumplidos, lo utilizaremos para poder calcular la tasa de mortalidad infantil postneonatal.
- **DIASC:** Número de días cumplidos, lo utilizaremos para calcular la tasa de mortalidad infantil neonatal.
- **CBAS1 o CODCAU1:** Contiene el primer valor de la causa básica de muerte que es una letra mayúscula de la A a la Z.
- **CBAS23 o CODCAU23:** Contiene los dos primeros dígitos de la causa básica de muerte. Se utilizan para hacer divisiones más específicas de estas causas y tiene valores desde 00 a 99.
- **CBAS4 o CODCAU4:** Contiene un último valor de la causa de muerte para proporcionar una mayor clasificación y tiene valores desde 0 a 9.

Al abrir un archivo de microdatos de defunciones nos encontramos que cada entrada corresponde a una defunción y de media en España ocurren unas 400000 defunciones anualmente, esto ya nos indica que disponemos de mucha información, además cada entrada puede llegar a tener hasta 231 caracteres los cuales conforman hasta 157 variables.

Lo primero que tendremos que realizar será dividir cada cadena de caracteres según la longitud de cada variable para tener un documento con todas las variables bien identificadas.

4.2. DEPURACIÓN Y ADECUACIÓN DE LOS DATOS

Una vez disponemos de este documento en el que cada fila será la correspondiente a un único individuo ya podemos filtrar a estos según la variable de interés o los requerimientos en cada caso.

Este proceso se tendrá que realizar también para cada uno de los años que tenemos información.

4.2.4. Datos natalidad

En este caso también disponemos de un archivo de microdatos que tendremos que clasificar. En este caso solo necesitaremos una variable:

- **CPRO:** Será el código de la provincia de nacimiento.

Empezaremos identificando la variable 'CPRO' dentro de la cadena de caracteres, y la tendremos que aislar eliminando todo lo demás. Una vez realizado este proceso tendremos un archivo de datos que contiene únicamente la variable nombrada.

A continuación, si contamos el número de entradas que tiene cada código de provincia, obtendremos el número de nacimientos ocurridos en un año natural en cada provincia. Una vez que tengamos esta información, tendremos que sumar los datos obtenidos según la comunidad a la que pertenezcan para crear así un nuevo fichero con los nacimientos por comunidades autónomas en un año natural.

Al igual que en el caso anterior, este procedimiento se tendrá que repetir con los conjuntos de datos de todos los años del estudio.

4.2.5. Datos mujeres

Gracias a la aplicación de descarga de datos del Instituto Nacional de Estadística nos podremos descargar un documento sobre la población femenina eligiendo las opciones de interés, que en nuestro caso serán las mujeres por provincias con edades comprendidas entre 15 y 49 años por grupos quinquenales.

Una vez tengamos este documento tendremos que transformarlo a un formato adecuado para poder tratar la información y posteriormente crear otro fichero de datos a partir de este, en el que sumemos el número de mujeres según la comunidad autónoma de residencia.

Capítulo 5

Creación de la aplicación

La información recopilada, depurada y almacenada en la base de datos nos va a permitir generar las representaciones que plasmaremos en la aplicación final.

Desde un principio el objetivo ha sido crear una aplicación que nos permita interactuar con ella para recibir un feedback y que la experiencia sea agradable. En el programa estadístico utilizado 'R', hay múltiples opciones para crear gráficos interactivos, pero no todas se adaptan a nuestras necesidades.

A continuación se explicará detalladamente los procedimientos y los paquetes necesarios para crear la aplicación.

5.1. Datos geográficos

El primer problema a solucionar ha sido el representar los datos geográficos. El archivo de datos original contenía mucha precisión geográfica, es decir, el mapa estaba compuesto por unidades muy pequeñas lo que requería un tiempo muy alto de cálculo para generar la representación.

Esto se solucionó gracias a la herramienta online "mapshaper". Lo que conseguimos con esto es simplificar las geometrías representadas, el archivo principal de datos pasa de 97 Mb a 74 Kb.



Figura 5.1: Mapa de España provincial sin reducir



Figura 5.2: Mapa de España provincial tratado con 'mapshaper'

No obstante, la simplificación anterior tiene un aspecto negativo y es la clara pérdida de información, pero como la idea de la aplicación es que sea rápida, este factor se puede sacrificar, y se puede observar cómo la diferencia causada por la pérdida de información es muy pequeña.

5.1.1. mapa_españa_prov y mapa_españa_com

Tendremos dos conjuntos de datos con formato 'shapefile', uno para el mapa de España a nivel provincial y otro para el mapa de España a nivel de comunidades autónomas. Cada uno de ellos formado por tres archivos, con las extensiones '.dbf', '.shp' y '.shx'. Los tres serán necesarios para poder representar correctamente ambos mapas.



Figura 5.3: Datos geográficos cargados en la aplicación

5.2. Unificación de los datos

En la aplicación desarrollada para este trabajo se pueden consultar datos desde el año 2003 hasta el año 2020. Si tuviésemos que cargar cada conjunto de datos dieciocho veces, una por año, la base de datos estaría muy sobrecargada y la gestión de esta sería muy complicada.

5.2. UNIFICACIÓN DE LOS DATOS

Por esta razón, se ha optado por concatenar los datos de los diferentes años, es decir, tener un único archivo que contenga la información de todos los años del estudio.

A continuación se explicará qué archivos existen en el servidor de la aplicación y cuál es su estructura.

5.2.1. TBM_prov y TBM_com

CPROV	De.0.a.4.años	De.5.a.9.años	De.10.a.14.años	De.15.a.19.años	De.20.a.24.años	De.30.a.34.años	De.35.a.39.años	De.40.a.44.años	De.45.a.49.años	De.50.a.54.años	De.55.a.59.años	De.60.a.64.años	De.65.a.69.años	De.70.a.74.años	De.75.a.79.años	De.80.a.84.años	De.85.a.89.años	De.90.a.94.años	De.95.a.99.años	X100.y.más.años	Total	
1	0.53864799	0.11469205	0.05989817	0.06451197	0.34544701	185.6756	287.5587	551.18110	9.783195													
2	0.56857666	0.16120365	0.00000000	0.14842668	0.23573786	244.1397	369.4196	368.79433	12.537667													
3	0.49157213	0.05460840	0.06832535	0.19751136	0.24730917	198.6317	313.4426	280.37383	9.338322													
4	1.06951872	0.14530308	0.09339248	0.15444015	0.34498053	204.5860	305.1643	233.16062	7.827514													
5	0.38752180	0.00000000	0.13821700	0.43446778	0.27968116	191.1765	332.5387	405.66038	16.091181													
6	0.28787334	0.03121878	0.11370097	0.20256974	0.27947794	211.2584	325.4613	453.41615	11.113806													
7	0.49562515	0.08490983	0.01612513	0.06803184	0.19278037	184.5464	287.2454	298.89299	7.305750													
8	0.43462688	0.05536045	0.05129093	0.17292649	0.22179380	207.4207	344.6090	465.94005	10.386179													
9	0.38461538	0.00000000	0.23750148	0.06275100	0.19059720	185.5391	323.7278	497.51244	13.356634													
10	0.45516614	0.06265664	0.11007760	0.10851283	0.21351553	206.6829	325.0189	322.17573	14.365191													
11	0.47963400	0.03091142	0.07982970	0.14407769	0.25051669	215.7722	321.2544	228.57143	9.022153													
12	0.68495910	0.03404835	0.12385435	0.16798817	0.31413613	205.4688	319.5266	333.33333	9.708158													
13	0.40241449	0.00000000	0.11567380	0.20113440	0.37558685	231.3733	362.2047	427.74566	13.992667													
14	0.67827964	0.07990624	0.09410658	0.24010180	0.14414414	200.3698	320.2805	322.46377	10.619988													
15	0.45835349	0.02178412	0.10814437	0.13114754	0.40389535	169.4013	272.5981	350.55351	11.475154													
16	1.61859918	0.12230920	0.00000000	0.10460251	0.19575218	241.4716	320.2797	424.24242	16.197697													
17	0.74143234	0.14154282	0.06649084	0.19044445	0.26894866	197.8412	348.7673	366.79537	9.222449													
930	1.2222115	0.14255032	0.09537683	0.4635961	0.5328039	234.6887	321.6958	148.40989	9.665142													
931	1.0030090	0.05136107	0.18269011	0.3670129	0.3665401	198.3221	298.7722	220.77922	9.207943													
932	1.2486378	0.07349699	0.06799946	0.3261933	0.4272508	219.0744	307.8125	466.66667	9.387451													
933	1.2300123	0.14865468	0.35005834	0.5884661	0.5569701	158.4071	298.1744	234.56790	12.003726													
934	1.4519752	0.11022928	0.23304591	0.2932551	0.8484445	229.1142	355.6522	337.34940	10.858771													
935	1.2486993	0.20096463	0.00000000	0.00000000	0.4559271	215.1899	342.8571	1000.00000	6.272437													
936	1.8533773	0.19485581	0.61124694	0.5891595	0.1668057	264.1509	189.1892	272.72727	6.675138													

Figura 5.4: Conjunto de datos unificados de la Tasa Bruta de Mortalidad por provincias

El contenido que tenemos en estos dos conjuntos de datos, está formado por las Tasas Brutas de Mortalidad calculadas por provincias 'TBM_prov' y por comunidades 'TBM_com'. Todas las tasas de mortalidad han podido ser calculadas ya que, como hemos comentado antes, teníamos la información de las defunciones y del padrón con la misma estructura.

En cada fila se representa una provincia diferente identificada por la variable 'CPROV', la cual contiene el propio código de la provincia. Las columnas dividen a la población de cada provincia en grupos quinquenales de edad. Por ejemplo, la celda de la fila 1 y columna 'De 0 a 4 años', tendrá la Tasa Bruta de Mortalidad de la provincia de Álava en el rango de edad de 0 a 4 años.

La razón por la que ya están calculadas, es por dotar a la aplicación de una mayor ligereza al no tener que calcular en cada ejecución todas las Tasas de Mortalidad necesarias para ser representadas.

Estos datos se utilizarán para poder representar en el mapa de España, tanto provincial como de comunidades autónomas, las Tasas Brutas de Mortalidad en el intervalo de edad y año elegido.

5.2.2. TMI_prov y TMI_com

	CPROV	Nacimientos	Defunciones.1.año	Defunciones.menos.27.dias	TMI	TMPostneonatal	TMNeonatal
1	1	2125	6	3	2.8235294	1.4117647	1.4117647
2	2	3160	6	5	1.8987342	0.3164557	1.5822785
3	3	13338	27	17	2.0242915	0.7497376	1.2745539
4	4	6860	36	27	5.2478134	1.3119534	3.9358601
5	5	714	1	1	1.4005602	0.0000000	1.4005602
6	6	5004	8	4	1.5987210	0.7993605	0.7993605
7	7	9352	21	15	2.2455090	0.6415740	1.6039350
8	8	43923	90	61	2.0490404	0.6602463	1.3887940
9	9	2145	4	3	1.8648019	0.4662005	1.3986014
10	10	2308	3	2	1.2998267	0.4332756	0.8665511
11	11	8925	20	13	2.2408964	0.7843137	1.4565826
12	12	3967	16	11	4.0332745	1.2603983	2.7728762
13	13	3642	6	3	1.6474465	0.8237232	0.8237232
14	14	5840	18	17	3.0821918	0.1712329	2.9109589
15	15	6958	16	10	2.2995114	0.8623168	1.4371946
16	16	670	9	7	13.4328358	2.9850746	10.4477612
17	17	6002	23	16	3.8320560	1.1662779	2.6657781
930	46	24742	83	51	3.3546197	1.2933473	2.0612723
931	47	4402	14	9	3.1803726	1.1358473	2.0445252
932	48	9982	36	21	3.6064917	1.5027049	2.1037868
933	49	1190	8	5	6.7226891	2.5210084	4.2016807
934	50	8460	39	27	4.6099291	1.4184397	3.1914894
935	51	1156	6	4	5.1903114	1.7301038	3.4602076
936	52	1315	12	7	9.1254753	3.8022814	5.3231939

Figura 5.5: Conjunto de datos unificados de la Tasa de Mortalidad Infantil por provincias

Este conjunto de datos recoge la información necesaria para representar la Tasa de Mortalidad Infantil y sus dos posibles divisiones, la Tasa de Mortalidad Neonatal y Postneonatal. Estos cálculos se han realizado por provincias en 'TMI_prov' y por comunidades 'TMI_com'.

Como se puede observar, las tasas han podido ser calculadas ya que disponemos de los datos de nacimientos y el número de defunciones en un año natural. En la columna 'Nacimientos' se recoge el número de nacimientos que han ocurrido en cada respectiva provincia o comunidad. Asimismo, en las columnas 'Defunciones 1 año' y 'Defunciones menos 27 días' tenemos el número de defunciones en el primer año de vida y en los primeros 27 días respectivamente.

Para poder calcular la Tasas de Mortalidad infantil, columna 'TMI', utilizaremos el número de defunciones en el primer año de vida. En el caso de la Tasa de Mortalidad Neonatal, se hará uso de las defunciones de menos de 27 días de vida para poder calcularla. Y para la Tasa de Mortalidad Postneonatal, se tendrá que calcular la diferencia entre las defunciones en el primer año de vida y los primeros 27 días.

Los cálculos también han sido realizados previamente para que la aplicación no tenga que realizarlos a la hora de la representación.

5.2. UNIFICACIÓN DE LOS DATOS

CPROV	De.0.a.4.años	De.5.a.9.años	De.10.a.14.años	De.15.a.19.años	De.20.a.24.años	De.25.a.29.años	De.90.a.94.años	De.95.a.99.años	X100y.más.años	
1	1	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	1.5046645	3.5211268	7.8740157
2	2	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	1.7456359	4.4642857	0.0000000
3	3	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.010140651	3.3080220	4.5901639	1.8691589
4	4	0.000000000	0.000000000	0.023348120	0.000000000	0.024641467	0.000000000	3.0573248	3.5211268	0.0000000
5	5	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.138102472	1.7507003	4.7675805	0.0000000
6	6	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	1.3091147	3.6900369	0.0000000
7	7	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.026421824	2.9875068	3.2154341	3.6900369
8	8	0.004024323	0.003460028	0.000000000	0.010172147	0.000000000	0.009010822	1.8686548	3.0215933	1.6348774
9	9	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	1.6227912	3.6787247	9.9502488
10	10	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	2.0649521	3.0234316	0.0000000
11	11	0.018447462	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	3.1294007	4.1811847	0.0000000
12	12	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	3.5156250	2.5359256	0.0000000
13	13	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	1.8261505	0.7874016	0.0000000
14	14	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	2.2454101	1.7533606	7.2463768
15	15	0.000000000	0.000000000	0.000000000	0.000000000	0.022438630	0.000000000	2.0149683	2.2740193	7.3800738
16	16	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	2.0066890	2.7972028	0.0000000
17	17	0.027460457	0.000000000	0.000000000	0.023805556	0.000000000	0.000000000	1.7743605	1.8039687	3.8610039
930	46	86.95652	0.009623713	0.009503355	0.026011861	0.000000000	0.056084620	1.2727200	0.4035920	0.0000000
931	47	0.00000	0.000000000	0.000000000	0.000000000	0.000000000	0.078544312	1.4611338	0.6711409	0.0000000
932	48	41.66667	0.000000000	0.000000000	0.022666485	0.018121851	0.051787980	1.4338973	0.3519268	0.7812500
933	49	0.00000	0.000000000	0.000000000	0.000000000	0.000000000	0.079567155	1.8828452	0.4424779	0.0000000
934	50	40.00000	0.000000000	0.000000000	0.025893990	0.022558087	0.016636167	1.0947764	1.1630161	0.0000000
935	51	19.60784	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.0000000	0.0000000	0.0000000
936	52	19.23077	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	2.4213075	0.0000000	0.0000000

Figura 5.6: Conjunto de datos unificados de la Tasa Bruta de Mortalidad por causa de muerte y provincia

5.2.3. Causas de muerte

Según la Clasificación Internacional de Enfermedades, todas las causas de muerte se dividen en 17 grandes grupos que más adelante se nombrarán. Esto quiere decir, que vamos a tener que crear este número de documentos tanto para las provincias 'TM_numerocausa_prov' como para las comunidades 'TM_numerocausa_com'.

Cada uno de estos documentos tendrá las Tasas de Mortalidad de una causa general y de todos los años del estudio. Al igual que en documentos anteriores, tendremos una variable 'CPROV' o 'CCO' con los códigos para identificar las provincias o comunidades respectivamente.

En cada fila se tendrá la información de una provincia o una comunidad según el archivo en el que nos encontramos, y contará con 21 columnas correspondientes a las divisiones de la edad en grupos quinquenales.

Aunque tengamos una gran cantidad de archivos, la aplicación seguirá siendo rápida porque solo tiene que acceder al documento y seleccionar los datos que se quieran representar en el mapa. Será mucho más rápido que tener que hacer los cálculos.

5.2. UNIFICACIÓN DE LOS DATOS

5.2.4. defunciones_prov y defunciones_com

CCO	De.0.a.4.años	De.5.a.9.años	De.10.a.14.años	De.15.a.19.años	De.20.a.24.años	De.90.a.94.años	De.95.a.99.años	X100.y.más.años	Total	
1	1	235	30	36	86	119	11402	3872	572	78461
2	2	24	2	3	10	22	3451	1601	279	16711
3	3	17	3	1	9	12	2798	1225	197	14550
4	4	26	5	1	4	12	1359	536	81	8559
5	5	40	5	9	17	38	2015	759	132	16486
6	6	11	3	3	2	6	1262	516	102	6467
7	7	50	5	12	12	16	7891	3846	683	36197
8	8	48	8	8	12	30	5215	1976	340	25835
9	9	173	24	21	70	90	14546	6263	1137	79784
10	10	112	9	21	44	72	7704	2858	434	48549
11	11	14	2	6	9	14	2392	871	150	13099
12	12	46	3	11	12	34	6272	2722	573	32845
13	13	144	12	34	49	50	12041	5287	997	66648
14	14	51	6	6	8	23	2060	714	102	12392
15	15	14	6	2	10	2	1288	639	111	6662
16	16	45	7	11	9	32	4620	1876	332	24252
17	17	7	2	0	6	4	686	334	71	3700
336	13	392	44	53	140	188	5041	1771	341	42201
337	14	108	6	17	34	72	978	318	50	9811
338	15	18	3	4	11	32	649	239	36	5319
339	16	82	4	7	33	62	2334	750	141	19249
340	17	9	2	3	4	11	319	122	17	2686
341	18	6	1	0	0	3	34	12	3	470
342	19	9	1	3	3	1	42	7	3	457

Figura 5.7: Conjunto de datos unificados de las defunciones por comunidades autónomas

Para poder representar las Tablas de Mortalidad necesitamos un desglose del número de defunciones por grupos quinquenales de edad por provincias 'defunciones_prov' y por comunidades 'defunciones_com'.

Al igual que en los documentos anteriores se identificarán con la variable 'CPROV' o 'CCO' las provincias o comunidades.

5.2.5. padron_prov y padron_com

Las tablas de mortalidad también necesitan la información del número de habitantes que hay en cada provincia 'padron_prov' y en cada comunidad 'padron_com'.

Tiene la misma estructura que los documentos de defunciones anteriores para que sea más sencillo crear todas las tablas de mortalidad.

5.3. LIBRERÍAS DE R

	CCO	De.0.a.4.años	De.5.a.9.años	De.10.a.14.años	De.15.a.19.años	De.20.a.24.años	De.25.a.29.años	De.30.a.34.años	De.35.a.39.años	De.40.a.44.años	De.45.a.49.años	De.50.a.54.años	De.55.a.59.años	De.60.a.64.años	De.65.a.69.años	De.70.a.74.años	De.75.a.79.años	De.80.a.84.años	De.85.a.89.años	De.90.a.94.años	De.95.a.99.años	X100.y.más.años	Total
1	1	373993	440262	493156	457932	444622	484050	54558	12289	2597	846411												
2	2	54113	63010	66512	63731	63547	66854	16726	4648	611	1329391												
3	3	30242	37588	41713	39314	38460	40651	14501	3814	562	1018784												
4	4	52459	58886	62015	58796	62247	75695	7364	1866	271	1171543												
5	5	78268	93995	109994	112718	118195	133553	11337	2705	721	2175952												
6	6	20726	26187	28990	26838	24506	25999	6821	1819	268	582905												
7	7	79684	96609	104975	102739	103144	107862	39822	11604	1631	2394918												
8	8	86469	104037	112955	106948	106730	111359	22602	5582	759	2045221												
9	9	340669	397213	427983	402420	396716	442219	70988	18366	2511	7780479												
10	10	207691	249206	277629	258899	250506	265620	38943	8991	1373	5057353												
11	11	40972	47992	53349	52987	54515	58805	11438	2678	400	1063987												
12	12	91429	108011	116946	110943	108606	122375	36811	9792	1832	2701819												
13	13	306807	346708	360542	341436	344349	398126	58306	15762	2373	6779888												
14	14	76670	86604	92803	86931	84143	88105	9620	2037	254	1511251												
15	15	29740	35207	36247	35037	34442	33685	6783	1911	246	661197												
16	16	87233	105166	108544	104224	99361	102173	24960	6221	731	2220504												
17	17	13303	15836	16657	15680	15106	15352	3707	1084	133	319914												
336	13	248179	262193	275280	314848	441355	556841	25420	6118	1029	5718942												
337	14	72545	69637	75544	80611	109020	120511	4124	811	149	1269230												
338	15	27160	25947	26259	28972	40547	49975	3481	816	109	578210												
339	16	87184	79705	83641	102542	143529	179704	10649	2333	292	2112204												
340	17	11859	11774	13040	14847	20686	24470	1792	384	51	287390												
341	18	4805	4976	5125	5258	6580	6950	158	35	3	74931												
342	19	4856	5132	4908	5092	5995	5904	159	37	11	68463												

Figura 5.8: Conjunto de datos unificados del padrón por comunidades autónomas

5.3. Librerías de R

'R' es un entorno de software libre para computación estadística y gráficos. Pero la enorme potencia de este programa deriva en su capacidad de incorporar en cualquier momento nuevas funciones ya creadas capaces de realizar nuevas tareas.

A continuación se presentarán los 'paquetes' o 'librerías' más importantes que han sido utilizados para poder realizar la aplicación, así como las funciones aplicadas.

5.3.1. 'simple features'

Tendremos que instalar y cargar la librería en nuestra sesión ejecutando el código a continuación.

```
install.packages("sf")  
library(sf)
```

Figura 5.9: Código 'R' de instalación 'sf'

La librería 'simple features' o 'sf', es esencial para poder tratar los datos espaciales. En nuestro caso, va a ser utilizada para leer, modificar y guardar los diferentes datos cartográficos de los que disponemos.

```
mapa_esp_prov <- read_sf("data/mapa_esp_prov.shp")
```

Figura 5.10: Función 'read_sf'

Gracias a la función `read_sf()`, podemos leer el documento que contiene la información geográfica para representar y tratar los mapas.

5.3.2. 'ggplot2'

Utilizando el código que se presenta a continuación, instalamos y cargamos esta librería.

```
install.packages("ggplot2")  
library(ggplot2)
```

Figura 5.11: Código 'R' de instalación 'ggplot2'

Esta es quizás la más conocida del entorno de 'R' para realizar representaciones gráficas. En nuestro caso la utilizaremos por su capacidad para personalizar los gráficos y las facilidades que nos proporciona para poder trabajar con el formato de datos cartográficos. Los métodos que utilizaremos para construir el mapa serán:

- **ggplot():** Función principal con la que inicializamos una nueva representación, y a la que habrá que pasar como argumento el conjunto de datos.
- **geom_sf_interactive():** Pasamos como argumento el archivo de datos que contiene la información geográfica e indicamos cuales son las secciones que se tienen que rellenar y con qué información.
- **theme():** Este método es únicamente personalización y modificación de los elementos visuales del gráfico.
- **scale_fill_carto_c():** Indicamos la paleta de colores que nosotros queramos que previamente ha sido definida.
- **labs():** De esta forma añadimos a nuestro gráfico un título y un subtítulo.

En la siguiente imagen tenemos un extracto del código que se ha utilizado en la aplicación para generar un gráfico con 'ggplot()'. Se observan las diferentes funciones explicadas anteriormente y los argumentos utilizados. En la aplicación este código no se utilizará para

representar el gráfico sino únicamente para crear un objeto que se pase como argumento de la función que explicaremos a continuación.

```
grafico <- ggplot(data = datos) +
  geom_sf_interactive(aes(fill = TBM,
    tooltip = paste(nombres, "\nTBM = ", round(TBM, 3)),
    data_id = TBM,
    color = "black") +
  theme(panel.background = element_rect(fill = "white"),
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    plot.title = element_text(hjust = 0),
    plot.subtitle = element_text(hjust = 0)) +
  scale_fill_carto_c(palette = "SunsetDark") +
  labs(title = paste("Tasa Bruta de Mortalidad", input$anosInput),
    subtitle = paste(" Rango Edad: ", min(input$ageInput), "-", (max(input$ageInput)-1)))
```

Figura 5.12: Código creación de objeto 'ggplot()'

5.3.3. 'ggiraph'

```
install.packages("ggiraph")
library(ggiraph)
```

Figura 5.13: Código 'R' de instalación 'ggiraph'

Gracias a esta librería podremos dotar de interactividad a nuestro gráfico creado con 'ggplot'. La función más característica que se utilizará en este trabajo, será la posibilidad de representar información extra al gráfico cuando se navega con el ratón por encima de las diferentes representaciones.

Con el método 'girafe()', al que pasaremos como primer argumento el objeto creado con 'ggplot()', tendremos diferentes posibilidades de personalización y de ajuste de diferentes parámetros para conseguir nuestra representación deseada.

```
tooltip_css <- "background-color:gray;color:white;font-style:italic;
  font-size:20px;padding:10px;border-radius:5px;"

girafe(ggobj = grafico,
  width_svg = 6,
  height_svg = 3,
  options = list(opts_tooltip(opacity = 0.95,
    use_fill = FALSE,
    use_stroke = FALSE,
    css = tooltip_css),
    opts_hover_inv(css = "opacity:0.5"),
    opts_hover(css = "padding:5pt")))
```

Figura 5.14: Creación del objeto girafe()

5.3. LIBRERÍAS DE R

El resultado de la ejecución de este código nos devolverá, ahora sí, una representación gráfica de los datos solicitados, con todas las características que se le han pasado como argumentos a las funciones. Cabe destacar la variable definida como 'tooltip_css', la cual se utiliza para modificar las opciones gráficas del gráfico en cuestión.

Obtendríamos la siguiente representación:

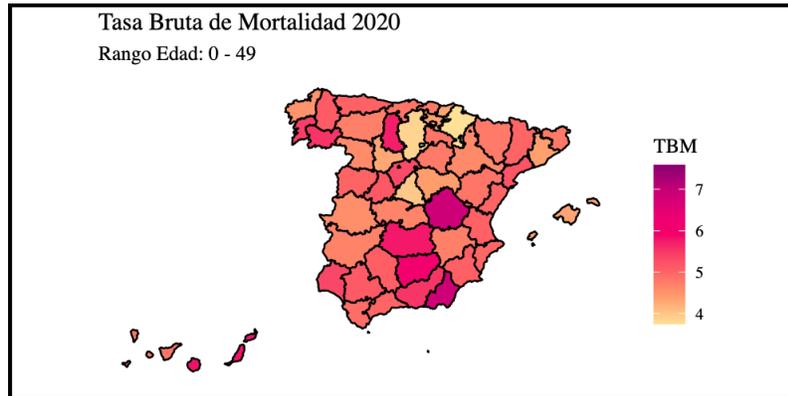


Figura 5.15: Resultado de la ejecución del código de la Figura 5.14

5.3.4. 'gt'

```
install.packages("gt")  
library(gt)
```

Figura 5.16: Código 'R' de instalación 'gt'

Necesitaremos de esta librería para poder representar de una forma mucho más estética y correcta las Tablas de Mortalidad. El funcionamiento de esta es muy sencilla y solo haremos uso del método 'gt()', al que pasaremos como argumento un data frame, y en el cual modificaremos diferentes aspectos visuales.

```
gt(data = df) %>%  
  tab_header(  
    title = md(paste("Tabla de muerte", nombre, input$anosInputtablas)))
```

Figura 5.17: Creación del objeto 'gt()'

5.3. LIBRERÍAS DE R

El resultado de ejecutar este código nos devolverá una tabla de Mortalidad de las características señaladas en la aplicación, que tendrá un aspecto como se muestra en la imagen siguiente.

Tabla de muerte Valladolid 2020								
x	n	Poblacion	Defunciones	Tasas Especificas	Probabilidad Muerte	Probabilidad Supervivencia	Supervivientes	Defunciones Tabla
0	5	18875	17	0.00090066225	0.0044931941	0.99550681	1000	6
5	5	23670	2	0.00008449514	0.0004223865	0.99957761	996	3
10	5	25131	0	0.00000000000	0.0000000000	1.00000000	995	3
15	5	23512	2	0.00008506295	0.0004252243	0.99957478	995	5
20	5	22138	5	0.00022585599	0.0011286427	0.99887136	995	10

Figura 5.18: Creación del objeto 'gt()'

Capítulo 6

TAMEApp

El objetivo de representación gráfica del trabajo es poder visualizar cómo evoluciona la mortalidad en España año por año en las diferentes comunidades y provincias. Para ello se utiliza la Tasa Bruta de Mortalidad y de este objetivo proviene el nombre de la aplicación 'TAMEApp'. Las primeras dos letras 'TA' hacen referencia a la palabra 'Tasa', la siguiente letra, la 'M' de mortalidad, y a continuación la 'E' de España. Además se incluye el acrónimo App para que una primera lectura nos deje claro que es una aplicación.

El proceso de la creación se ha llevado a cabo a través del paquete 'SHINY' del software estadístico 'R', que ofrece un marco para crear aplicaciones web interactivas.

Este paquete es una herramienta orientada a científicos de datos que permite desarrollar una interfaz de usuario utilizando los lenguajes de programación, 'R', 'HTML', 'CSS' o 'JavaScript'. Además nos ofrece una gran variedad de temas y herramientas de personalización de la aplicación.

La aplicación estará en todo momento ejecutada y gestionada a través de 'Shiny' implementado dentro de un servidor de aplicaciones de la Universidad de Valladolid y en el que se trabajará con la versión de 'R 3.6.3'.

Para tener acceso a la aplicación será necesario contar con un enlace directo a la misma, el cual se refleja a continuación. Desde dicho enlace se podrán realizar en cualquier momento las consultas pertinentes.

<http://shiny1.eio.uva.es:3838/users/raul/TAMEApp/>

6.1. Estructura Shiny

En un programa básico de Shiny siempre tendremos que tener dos ficheros obligatoriamente para el correcto funcionamiento de la aplicación.

El primero será el fichero 'ui', el cual se encarga principalmente de la visualización de la propia aplicación y de la creación de la estructura y de los diferentes elementos que están dentro de esta, botones, selecciones, recuadros... De la misma forma, también es la encargada de mostrar el resultado final.

El segundo fichero esencial es 'server', el cual se encarga de tratar todos los datos y los 'Inputs' provenientes de la función 'ui' y así ofrecer unos resultados en forma de 'Outputs' que se puedan mostrar en la aplicación.

```
library(shiny)

ui <- fluidPage(
  xxxxInput(),
  xxxxOutput()
)

server <- function(input, output)
{
 Codigo
}

shinyApp(ui = ui, server = server)
```

Figura 6.1: Estructura básica de un programa 'Shiny'

En el caso de nuestra aplicación se han utilizado dos ficheros más que, aunque no son primordiales, nos facilitan y mejoran la experiencia de la aplicación.

El primero de ellos es un fichero con extensión '.css', indicativo de que es una 'hoja de estilos en cascada', el cual se ocupará de la mayoría de las opciones estéticas tanto de la página en sí como de los elementos que la conforman. Un ejemplo de ello se muestra en la siguiente imagen en la que se dan unos atributos determinados para la cabecera de la página, en concreto el título de la aplicación, y para la visualización de un botón de la aplicación, cuyo identificador del estilo en el código será 'success'.

```

.main-header .logo {
  font-weight: bold;
  font-size: 40px;
}
.btn-success {
  font-size: 18px;
  font-family: inherit;
  padding: 5px 12px;
  height: 50px;
  min-width: 22%;
  margin: 10px 20px 0px 0px;
  text-align: middle;
  color: white;
  background: #57a3eb;
  text-indent: 4px;
  vertical-align: middle;
  border-radius: 4px;
  border-color: black;
}

```

Figura 6.2: Código del archivo con formato '.css'

Este archivo de datos y código se implementará en el fichero 'ui' anteriormente explicado. Primero tendremos que indicar el nombre del archivo de datos al que queremos que se haga referencia, en nuestro caso 'estilo_pagina.css', y una vez realizado este proceso ya podemos indicar los estilos que se aplicarán a los objetos.

```

dashboardBody(
  tags$head(
    tags$link(
      rel = "stylesheet",
      type = "text/css",
      href = "estilo_pagina.css"
    )
  )
)

```

```

bsButton(inputId = "tbm",
  type = "action",
  label = "Tasa Bruta Mortalidad",
  icon = icon("user"),
  style = "success"
)

```

Figura 6.3: Código introducción fichero y personalización

En la imagen de la izquierda tenemos el fragmento de código en el cual se introduce el fichero para su localización, y en la imagen de la derecha en la última línea '*style = "success"*', indicamos que el botón 'tbm' tendrá el estilo indicado anteriormente.

El segundo documento que tenemos que incluir tiene extensión '.html', el cual será necesario para la visualización de una ventana emergente de la aplicación. Para la correcta visualización tendremos que indicar la estructura en lenguaje HTML como se presenta a continuación.



Figura 6.6: Mensaje emergente de la aplicación 'TAMEApp'

En este primer mensaje nos encontramos con una bienvenida y con una pequeña descripción de las utilidades y funciones que se encuentran en la misma.

Una vez leída toda la información, tal y como se señala en la última sección del cuadro, tendremos que hacer click con el ratón fuera de esta ventana para poder ocultar el mensaje y poder comenzar a consultar la información.

6.2.2. División pantalla

Una vez que hemos cerrado la ventana emergente nos encontramos la siguiente página:

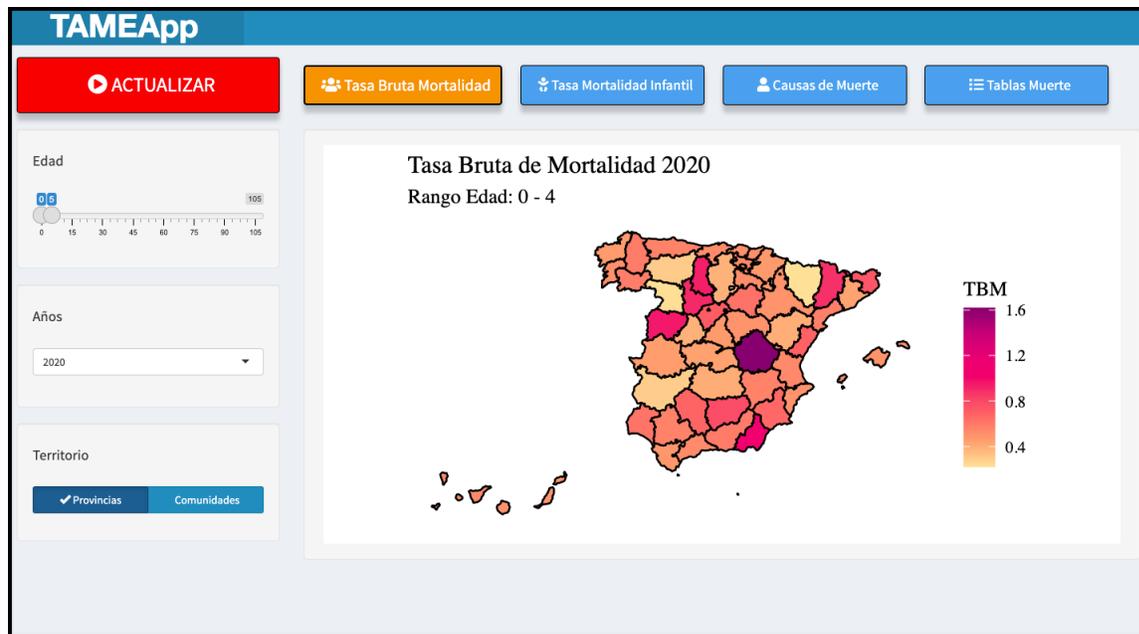


Figura 6.7: Pantalla principal de la aplicación

Se pueden observar en esta primera pantalla cuatro partes principales de la aplicación que serán: la cabecera, las pestañas, los seleccionadores de opciones y, por último, el gráfico principal. A continuación pasaremos a explicar detalladamente el funcionamiento de ellas.

Cabecera de la aplicación



Figura 6.8: Imagen de la cabecera de la aplicación

Este elemento únicamente se utiliza en nuestra aplicación para indicar el nombre de la misma, y se le dan diferentes atributos para una mejor visualización en la página.

```
ui <- dashboardPage(  
  dashboardHeader(  
    title = "TAMEApp",  
    titleWidth = 300  
  ),  
  dashboardSidebar(  
    disable = TRUE  
  ),  
)
```

Figura 6.9: Código con el que se declara la cabecera

6.2. ESTRUCTURA GRÁFICA

Este será el primer elemento que declaremos en el código del fichero 'ui'. La función 'dashboardHeader()' nos facilita tener este formato de cabecera, y la función 'dashboardSidebar()' está implícita siendo utilizada para desplegar una barra que hace las funciones de submenú. En nuestro caso al no ser necesario tenemos que indicar que no queremos que se muestre en pantalla con el atributo 'disable = TRUE'.

Pestañas

En este bloque se pueden observar cuatro botones, 'Tasa Bruta de Mortalidad', 'Tasa Mortalidad Infantil', 'Causas de Muerte' y 'Tablas Mortalidad'.



Figura 6.10: Pestañas de la aplicación con un botón seleccionado

Estos cuatro botones serán los que nos ayuden a navegar por las diferentes secciones de la aplicación en las cuales nos centraremos detalladamente más adelante.

A la hora de interactuar con ellos disponen de dos dinámicas. Si únicamente pasamos el ratón por encima de los botones, estos se iluminarán en verde para tener una sensación de reactividad, pero no será hasta el momento en el que se realice un click cuando se seleccionará la opción.

En el momento en el que hacemos click en una de las cuatro opciones, la seleccionada se iluminará ahora de color naranja para indicarnos en la pestaña en la que nos encontramos. En el ejemplo de la Figura 6.10, la pestaña que esta seleccionada será la 'Tasa Bruta Mortalidad' y tendríamos el ratón encima del botón 'Causas de Muerte'.

Lo que conseguimos al presionar uno de los cuatro botones es cambiar la distribución de los seleccionadores de opciones que se muestran y del gráfico principal. Esto se consigue con las siguientes líneas de código.

```

observeEvent(input$tbm, {
  actualizar_botones("Tbm")
  showElement("panel1")
  hide("panel2")
  hide("panel3")
  hide("panel4")
})

observeEvent(input$infantil, {
  actualizar_botones("Tmi")
  showElement("panel2")
  hide("panel1")
  hide("panel3")
  hide("panel4")
})

```

Figura 6.11: Código para intercambiar entre ventanas

En el código anterior observamos cómo existe un método 'observeEvent()', el cual se ejecuta en el momento en el que se activa 'input\$tbm', o alguno de los cuatro botones.

Una vez se ha presionado el botón, con la función 'actualizar_botones()' modificamos la estética del propio botón que anteriormente hemos mencionado, y se muestra en pantalla el panel correspondiente al botón seleccionado, a la vez que se ocultan los otros tres paneles.

Selección de opciones

En la parte izquierda de la pantalla aparecerán todas los elementos que en cada pestaña nos sirvan para elegir las opciones que vamos a querer representar. Estas irán cambiando según en la pestaña en la que nos encontremos, más adelante nos centraremos específicamente en cada pestaña, pero en este aparatado se desarrollarán todas las posibles opciones para seleccionar los datos que tenemos.

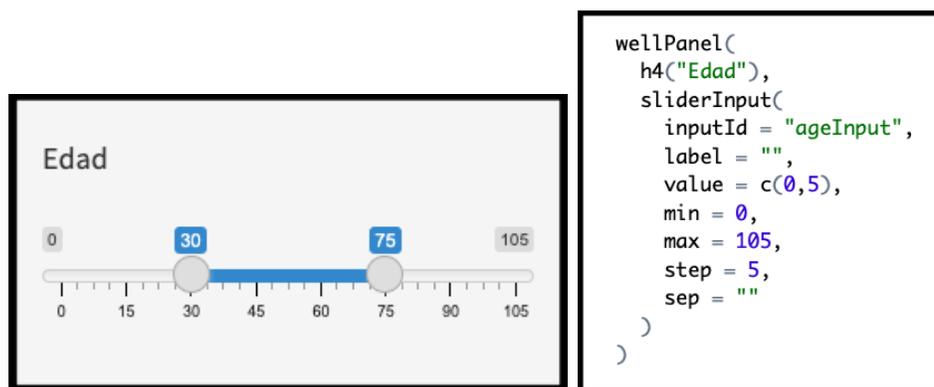


Figura 6.12: Selección del rango de edad

6.2. ESTRUCTURA GRÁFICA

Este primer bloque de selección nos permite escoger el rango de edad que queremos representar en nuestro mapa. Tendremos que arrastrar los dos círculos grises, que tienen entre ellos el intervalo azul, y colocarlos en las edades escogidas.

Para crear este bloque utilizaremos de aquí en adelante la función `'wellPanel()'`, que nos creará una caja específica para representar en este caso el `'sliderInput()'`. A este elemento le pondremos un identificador para poder recoger las selecciones `'inputId'`. Los valores iniciales con los que se representará será 0 y 5 (en la imagen ya están modificados). Cabe destacar que en el caso de nuestra aplicación el máximo es 105 pero esto engloba a todas las personas mayores de 100 años, es decir, si seleccionamos el intervalo 100-105 estamos seleccionando las personas de más de 100 años. Además solo podremos seleccionar los años múltiplos de 5 para crear los grupos quinquenales.

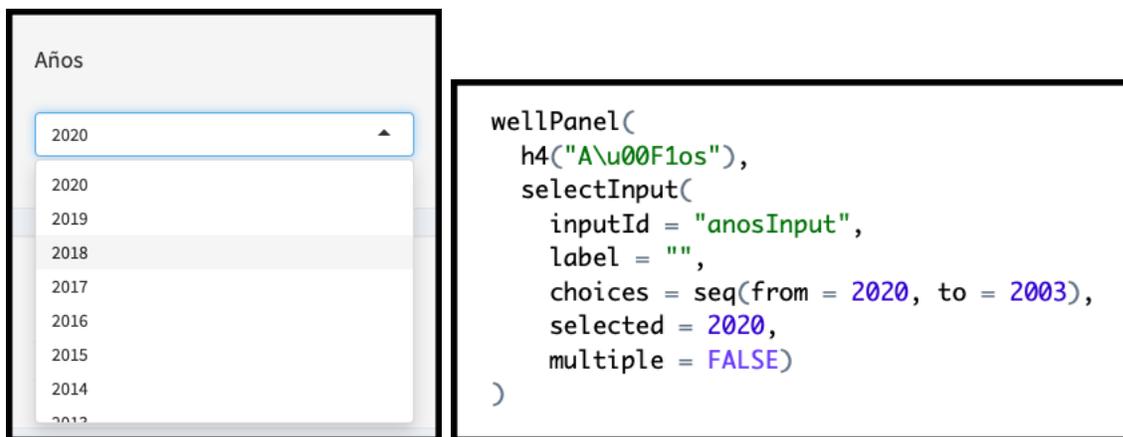


Figura 6.13: Selección del año de los datos

Gracias a este nuevo bloque podremos seleccionar el año del cual queremos que se representen los datos. En este trabajo se disponen datos desde el año 2003 hasta el 2020.

Para poder seleccionar el año debemos hacer click en la flecha negra para poder desplegar un menú de opciones el cual nos deje seleccionar el año deseado haciendo un simple click.

Si nos fijamos en el código, identificaremos este bloque como `'anosInput'`, e indicaremos con la opción `'choices'` el número de años que pueden ser elegidos, en nuestro caso como se ha comentado en repetidas ocasiones será desde el año 2003 hasta el 2020. Como primer año seleccionado siempre apareciera el 2020 y no se podrán seleccionar varios años a la vez.

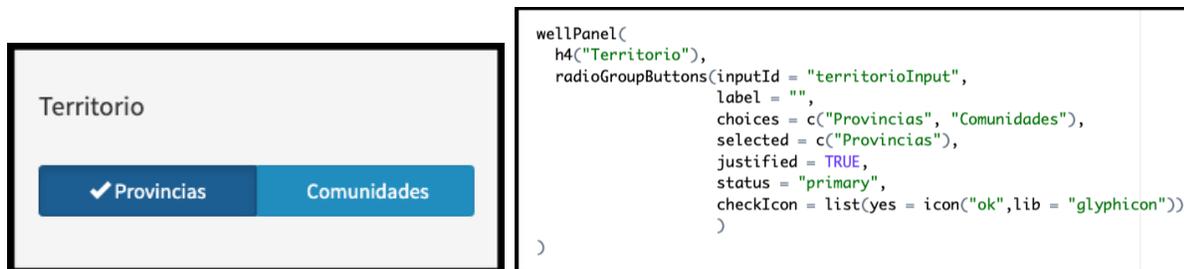


Figura 6.14: Selección de mapa provincial o regional

La siguiente opción nos dará la posibilidad de representar el mapa a nivel de provincias o de comunidades. Si hacemos click en una de las opciones, esta se iluminará en un color más oscuro y aparecerá un símbolo indicando cual está seleccionada.

El código es muy sencillo, y se utiliza el método 'radioGroupButtons()' en el cual indicamos las posibles opciones y la opción que aparecerá seleccionada por defecto. Además tendremos que indicar el estilo del botón, así como el símbolo que se representará en la opción elegida.

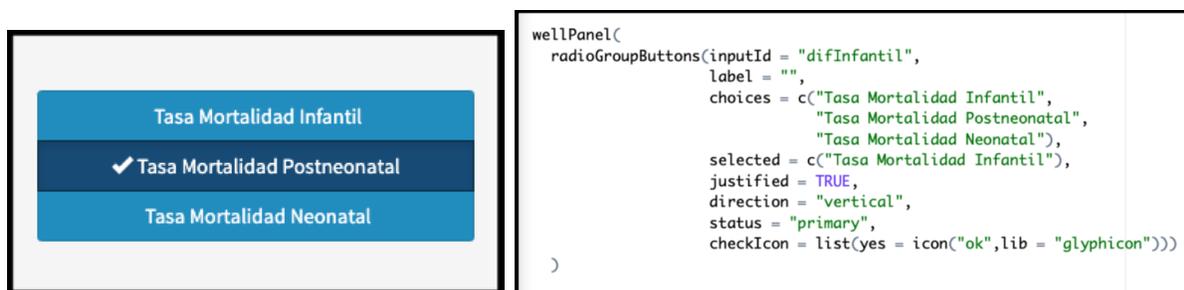


Figura 6.15: Selección Tasas Mortalidad Infantil

Este bloque es muy similar al anterior, y nos ayudará a la hora de indicar qué Tasa de Mortalidad Infantil queremos representar, entre la 'Tasa de Mortalidad Neonatal', 'Tasa Mortalidad Postneonatal' o 'Tasa Mortalidad Infantil'. Esta última será la suma de las dos primeras.

6.2. ESTRUCTURA GRÁFICA

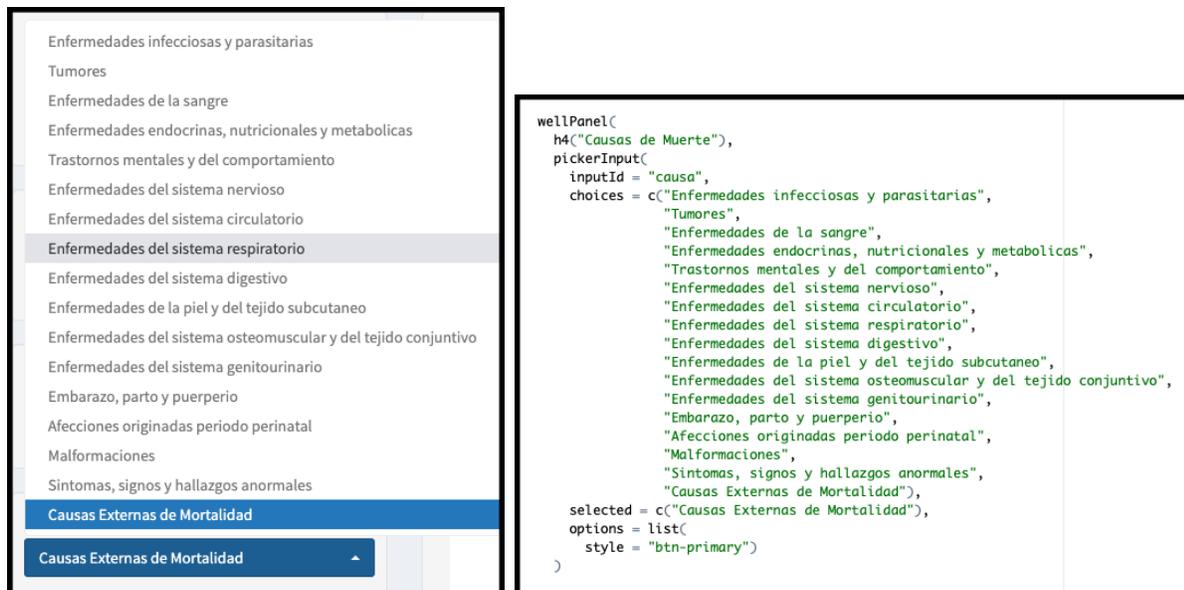


Figura 6.16: Selección de la causa de defunción

Según la Clasificación Internacional de Enfermedades se establecen unos grupos principales de las causas de muerte que aparecen en el CIE-10. En total estas causas principales son 17 y están nombradas para poder representar las tasas de mortalidad según estas causas.

Para el año 2020, en el cual comenzó la pandemia ocasionada por el SARS-COV-2, las defunciones a causa de este virus están introducidas en la causa de muerte 'Enfermedades infecciosas y parasitarias'.

Para crear el código ha sido necesario escribir todas las causas, para que se pudieran representar en la selección de opciones. Se utiliza el método 'pickerInput()' y todas las opciones han sido previamente explicadas.



Figura 6.17: Selección de provincia o comunidad autónoma

6.2. ESTRUCTURA GRÁFICA

A la hora de escoger la tabla de mortalidad, necesitaremos elegir la provincia o comunidad autónoma de la cual queremos la información. Para lograrlo, dispondremos de un desplegable en el cual haciendo un click en la flecha se nos mostrarán las comunidades y seguidamente las provincias, es ahí donde podremos seleccionar la opción elegida con un click.

El código para este bloque es el mismo que para el anterior a diferencia de que en este caso tenemos una subdivisión en el desplegable, en el que se dividen las provincias y las comunidades.



Figura 6.18: Botón para poder descargar la Tabla de muerte

Como último bloque de selección, incluimos un botón para poder descargar directamente la tabla de mortalidad que tengamos en pantalla, simplemente haciendo un click sobre el botón 'Download'.

El método que utilizamos para crear este último bloque es 'downloadBttm()'. Gracias a este podremos descargarnos un documento en formato de imagen '.png'.

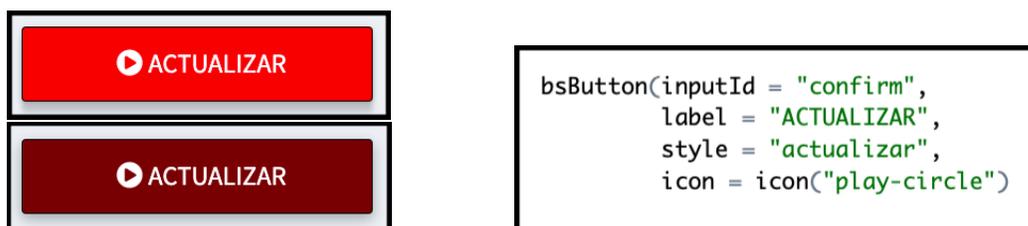


Figura 6.19: Botón para actualizar la página y respectivo código

Para finalizar disponemos de un botón en esta sección 'ACTUALIZAR', el cual tendremos que presionar cada vez que se quiera actualizar la información que queremos representar en la pantalla.

La primera imagen hace referencia al botón cuando no tiene ningún tipo de reacción. En el momento en el que pasamos el ratón por encima, el aspecto pasa a ser como en la segunda imagen. Para poder actualizar la información tendríamos que hacer un click en el mismo.

Gráficos principales

En esta última división de la pantalla se representará el resultado de todo lo comentado anteriormente.

Tendremos dos posibles salidas generales:

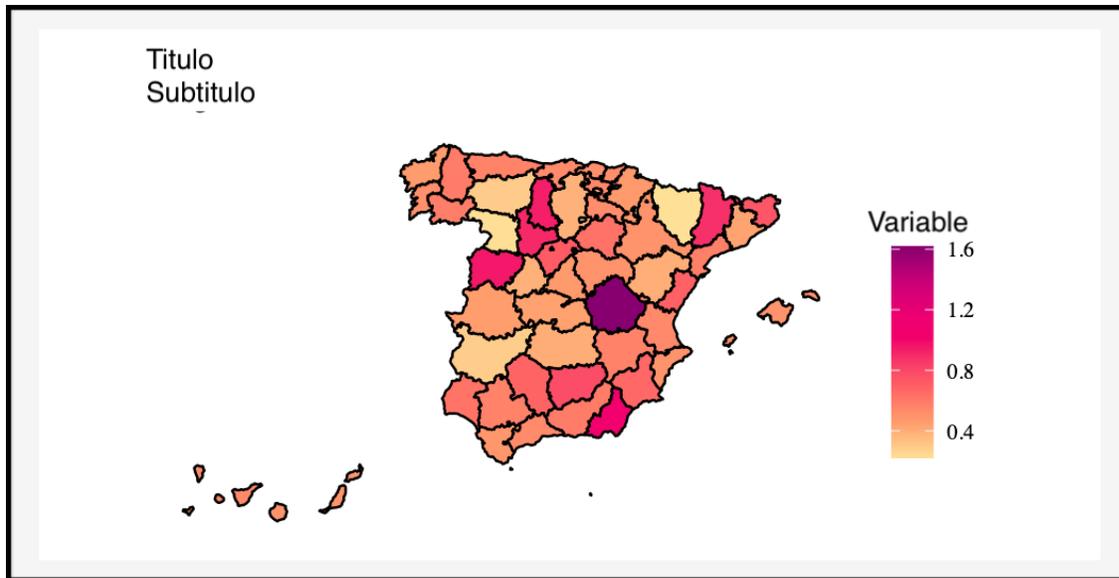


Figura 6.20: Estructura del gráfico representado por la aplicación

Esta primera imagen se corresponde con la salida más importante de este trabajo. Se trata de la representación gráfica en forma de mapa de España.

Estará compuesto por un título, un subtítulo y la parte principal que será el mapa. Además a la derecha del gráfico aparecerá una leyenda con el nombre de la variable que se esté representando.

El segundo tipo de salida que nos proporciona la aplicación serán las tablas de muerte. En el título estará el nombre de la provincia o la comunidad autónoma junto con el año que se seleccione, y a continuación se obtiene la tabla con todos los valores.

6.2. ESTRUCTURA GRÁFICA

Tabla de muerte Álava 2020												
x	n	Población	Defunciones	Tasas Específicas	Probabilidad Muerte	Supervivientes	Defunciones Tabla	Esperanza de vida intervalo	Población estacionaria	Tiempo vivido	Esperanza vida edades exactas	Probabilidad perspectiva de paso
0	1	2125	6	0.002824	0.014100	100000	1410	0.15	98802	8249886	82.4989	3.9895
1	4	12727	2	0.000157	0.000785	98590	77	1.60	394175	8151084	82.6766	1.2492
5	5	17438	2	0.000115	0.000575	98513	57	2.50	492422	7756909	78.7400	0.9996
10	5	16695	1	0.000060	0.000300	98456	30	2.50	492205	7264487	73.7841	0.9997
15	5	15501	1	0.000065	0.000325	98426	32	2.50	492050	6772282	68.8058	0.9990
20	5	14474	5	0.000345	0.001724	98394	169	2.50	491548	6280232	63.8274	0.9990
25	5	15218	1	0.000066	0.000330	98225	33	2.50	491042	5788684	58.9329	0.9992
30	5	18224	5	0.000274	0.001369	98192	134	2.50	490625	5297642	53.9519	0.9980
35	5	23270	12	0.000516	0.002577	98058	253	2.50	489658	4807017	49.0222	0.9967
40	5	28479	23	0.000808	0.004032	97805	394	2.50	488040	4317359	44.1425	0.9941
45	5	27483	43	0.001565	0.007795	97411	759	2.50	485158	3829319	39.3110	0.9898
50	5	25986	66	0.002540	0.012620	96652	1220	2.50	480210	3344161	34.6000	0.9840
55	5	24267	95	0.003915	0.019385	95432	1850	2.50	472535	2863951	30.0104	0.9746
60	5	21577	138	0.006396	0.031477	93582	2946	2.50	460545	2391416	25.5542	0.9598
65	5	18601	188	0.010107	0.049290	90636	4467	2.50	442012	1930871	21.3036	0.9405
70	5	17381	253	0.014556	0.070225	86169	6051	2.50	415718	1488859	17.2784	0.9073
75	5	13293	330	0.024825	0.116872	80118	9364	2.50	377180	1073141	13.3945	0.8401
80	5	9444	440	0.046590	0.208648	70754	14763	2.50	316862	695961	9.8363	0.7109
85	5	7455	724	0.097116	0.390718	55991	21876	2.50	225265	379099	6.7707	0.5172
90	5	3323	617	0.185676	0.634057	34115	21631	2.50	116498	153834	4.5093	0.3117
95	5	852	245	0.287559	0.836463	12484	10442	2.50	36315	37336	2.9907	0.0281
100	5	127	70	0.551181	1.000000	2042	2042	0.00	1021	1021	0.5000	0.0000

Figura 6.21: Ejemplo de Tabla de muerte mostrada en la aplicación

6.2.3. Utilización de pestañas

En esta sección se explicará cómo utilizar cada una de las pestañas y qué información se representa en cada una de ellas.

Tasa Bruta Mortalidad

Como su propio nombre indica en esta sección se van a representar las tasas brutas de mortalidad en España.

En esta pestaña podremos elegir el rango de edad para el cual queremos la información, el año del que queremos estos datos y por último, si queremos que los datos se agrupen por provincias o por comunidades.

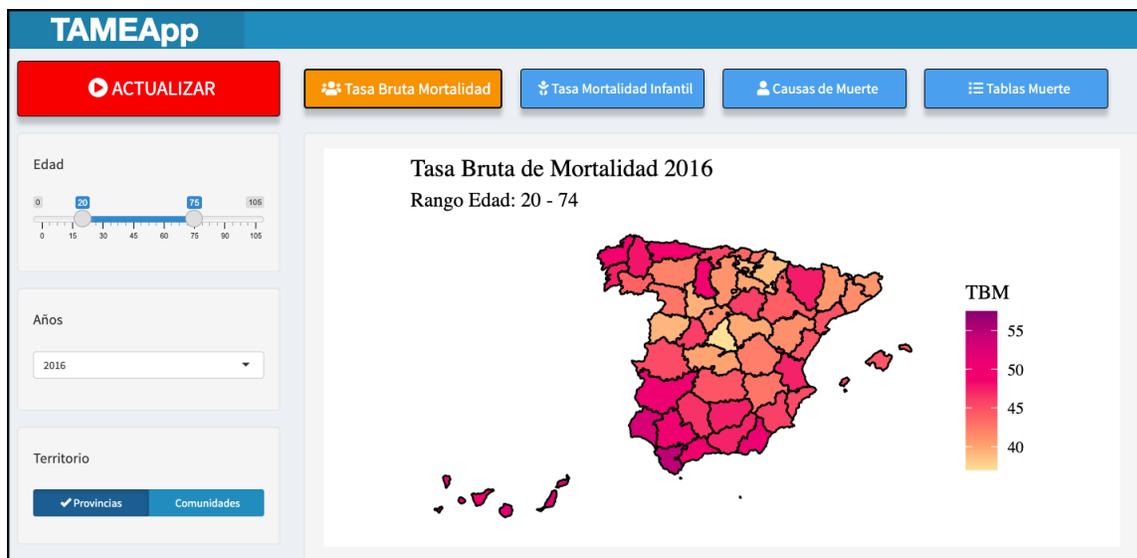


Figura 6.22: Resultado de la pestaña 'Tasa Bruta Mortalidad'

Podemos ver un ejemplo del resultado que obtendríamos en esta pestaña al seleccionar las opciones que aparecen en pantalla.

Se puede observar cómo en el título nos indica el contenido que se está representado junto con el año de los datos. En el subtítulo queda retratado el rango de edad para el cual se están calculando las Tasas Brutas de Mortalidad. Por último, en el margen derecho tenemos la leyenda con la cual nos podemos hacer una idea con un simple vistazo de los valores que tienen en este caso las diferentes provincias.

Pero aquí entra el valor específico de esta aplicación, si pasamos el ratón por encima de las diferentes comunidades o provincias (como se muestra en el ejemplo siguiente) nos mostrará un cuadro de información con el nombre de la comunidad/provincia y el valor de la Tasa Bruta de Mortalidad. Cuando realizamos esta acción el resto de las comunidades/provincias se sombrea para que se pueda apreciar mejor la zona seleccionada.

En este ejemplo en particular dejamos el ratón en la provincia de Soria y observamos que la Tasa Bruta de Mortalidad en el año 2016 es de 46,008.

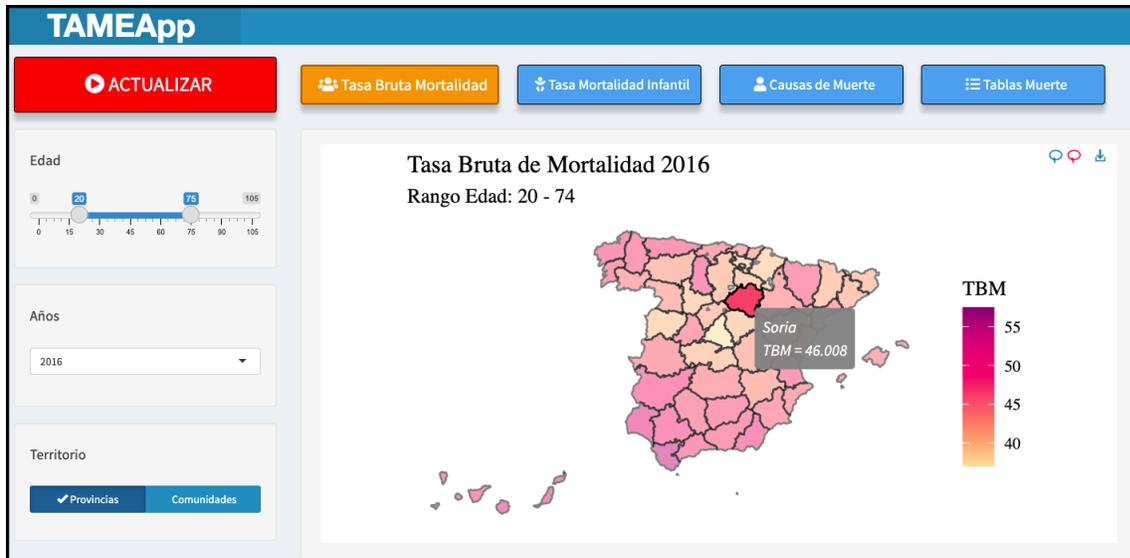


Figura 6.23: Resultado de la pestaña 'Tasa Bruta Mortalidad' con interacción

Tasa Mortalidad Infantil

En esta nueva pestaña no podremos agrupar los datos por rango de edad ya que la mortalidad infantil la dividimos en dos grupos, 'Neonatal' y 'Postneonatal'. Serán estas dos opciones, incluyendo las dos unificadas, 'Tasa Mortalidad Infantil', las que podremos seleccionar junto con el año de los datos y el territorio.



Figura 6.24: Resultado de la pestaña 'Tasa Mortalidad Infantil'

En este ejemplo como se puede comprobar en el título, se está representando la Tasa de Mortalidad Postneonatal del año 2006 por comunidades autónomas.

Causas de Muerte

En esta pestaña tenemos la posibilidad de representar las Tasas Brutas de Mortalidad de España según el rango de edad, el año y la causa de muerte que se desee, además esta información se podrá representar tanto por provincias como por comunidades.



Figura 6.25: Resultado de la pestaña 'Causas de Muerte con interacción'

En el ejemplo representado tenemos la Tasa de Mortalidad de 2020 sobre la causa principal de muerte 'Tumores' y en el rango de edad 20-64 años.

Dejando el cursor sobre la provincia de Burgos, comprobamos que la Tasa de Mortalidad en esta provincia es de 4,456.

Un aspecto común en todos los mapas, es que existe la opción de poder descargar la representación en formato '.png', de esta forma se facilita el poder exportar esta información y poder utilizarla en cualquier otro documento. Dicha opción queda reflejada en la esquina superior derecha del mapa con el icono de una flecha hacia abajo.

Tablas Muerte

Por último, tenemos una pestaña muy diferente a la temática que veníamos viendo. Aquí no tendremos una representación gráfica, sino que tendremos la posibilidad de obtener la tabla de muerte de cualquier provincia o comunidad autónoma y para cualquier año entre el 2003 y el 2020.

6.2. ESTRUCTURA GRÁFICA

Además, se podrá descargar en cualquier momento para, igual que en el caso anterior, incluirla o hacer uso de ella para algún trabajo.

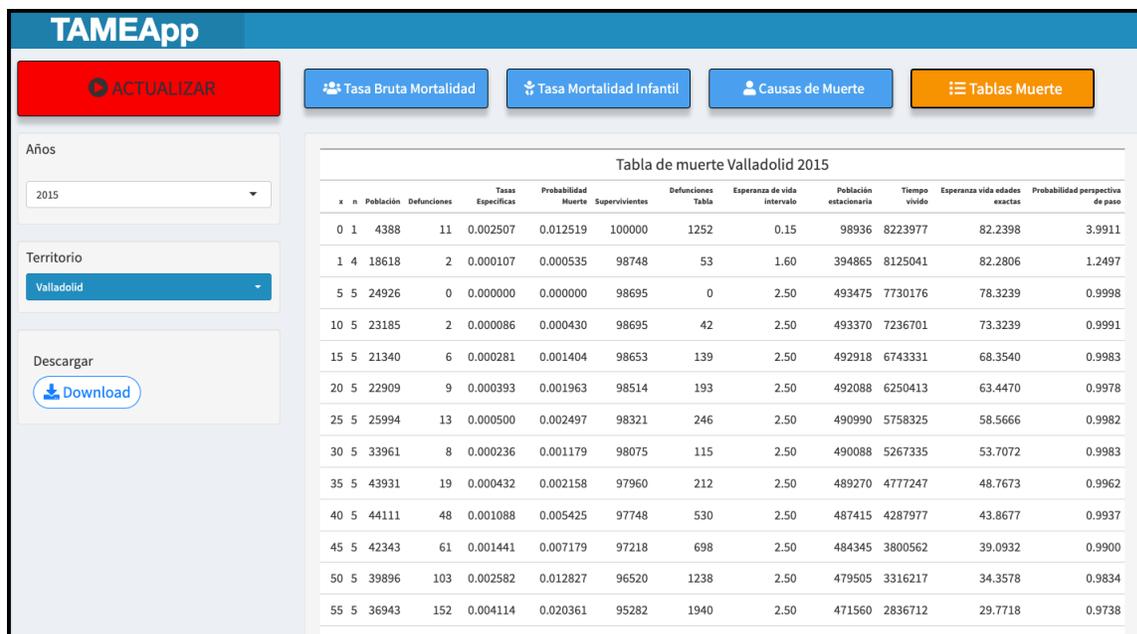


Figura 6.26: Resultado de la pestaña 'Tablas Muerte' para la provincia de Valladolid en 2015

Capítulo 7

Conclusiones y trabajos futuros

La creación de la aplicación 'TAMEApp' ha requerido aplicar conocimientos aprendidos a lo largo de los cursos, en especial el trabajo realizado con el programa estadístico 'R'. No obstante se ha tenido que aprender, estudiar e investigar sobre nuevos conceptos relacionados con los movimientos naturales de la población y la demografía, y más específicamente sobre la mortalidad. De esta manera se complementa la formación de graduado en Estadística con la adquisición en competencias en Estadística Pública y Demografía, parcelas muy importantes en la formación de un estadístico y de gran proyección dada la gran demanda actual de estadísticos por parte de las administraciones públicas.

Todo el trabajo que se ha realizado en estos meses ha dado sus frutos y se ha conseguido el objetivo marcado, consiguiendo así un producto final del cual puede hacer uso cualquier usuario, ya sea de consulta o de investigación.

Este trabajo se ha desarrollado con una estructura que facilita que pueda ser ampliada de manera sencilla en posibles futuros trabajos. Esta posibilidad de ampliación puede crecer en todas las direcciones.

A medida que el Instituto Nacional de Estadística publique microdatos o información actualizada, esta se puede introducir en la aplicación de una forma fácil y sencilla. Del mismo modo, se puede crear una base de datos histórica con información de años anteriores a los utilizados en este trabajo, que permitiría estudiar la evolución de los indicadores demográficos con mayor perspectiva.

Como líneas futuras de trabajo, también cabe plantearse representar otros aspectos sobre la mortalidad para dotar de mayor riqueza a la aplicación o añadir funcionalidades a la misma para que sea más atractiva y completa, incorporando aspectos como la visualización de la evolución dinámica de los indicadores de mortalidad u otros indicadores demográficos.

Como última ampliación también existe la posibilidad de que esta aplicación se utilice como plantilla o ayuda para crear una nueva o bien se implemente en la misma otras facetas del movimiento natural de la población como los nacimientos y los matrimonios.



Bibliografía

- [1] VINUESA, J.(ED). (1994) *Demografía, análisis y proyecciones*. Editorial Síntesis.
- [2] VINUESA, J Y PUGA, D. (2007) *Técnicas y ejercicios de demografía*. Editorial Síntesis. Instituto Nacional de Estadística.
- [3] INSTITUTO GALLEGO DE ESTADÍSTICA. (S.F.). *Indicadores Demográficos Básicos. Metodología*. IGE, p.13-16, de https://www.ige.eu/estatico/pdfs/s3/metodoloxias/met_indicadores_demograficos_es.pdf
- [4] INSTITUTO NACIONAL DE ESTADÍSTICA. (2020, DICIEMBRE) *Indicadores Demográficos Básicos. Metodología*. INEbase. p.11-18, de https://www.ine.es/metodologia/t20/metodologia_idb.pdf
- [5] INSTITUTO NACIONAL DE ESTADÍSTICA. (2016, MAYO) *Tablas de mortalidad. Metodología*. INEbase. <https://www.ine.es/metodologia/t20/t2020319a.pdf>
- [6] INSTITUTO NACIONAL DE ESTADÍSTICA. (ACTUALIZADO 15/11/2021) *Temas de la oposición al Cuerpo de Diplomados en Estadística del Estado. Tema 5*. <https://www.ine.es/index.htm>
- [7] INSTITUTO NACIONAL DE ESTADÍSTICA. (2013 JUNIO) *Estadísticas del Movimiento Natural de la Población. Metodología*. INEbase. <https://www.ine.es/metodologia/t20/t2030301.pdf>
- [8] DECRETO DE 14 DE NOVIEMBRE DE 1958 POR EL QUE SE APRUEBA EL REGLAMENTO DE LA LEY DEL REGISTRO CIVIL. *Boletín Oficial del Estado*, 296, de 11 de diciembre de 1958. Artículo 20. [https://www.boe.es/eli/es/d/1958/11/14/\(1\)/con](https://www.boe.es/eli/es/d/1958/11/14/(1)/con)
- [9] LEY 6/2021, DE 28 DE ABRIL, POR LA QUE SE MODIFICA LA LEY 20/2011, DE 21 DE JULIO, DEL REGISTRO CIVIL. *Boletín Oficial del Estado*, 102, de 29 de abril de 2021. Artículo 20. <https://www.boe.es/eli/es/l/2021/04/28/6/con>
- [10] INSTITUTO NACIONAL DE ESTADÍSTICA. *Cartografía digitalizada de secciones censales*[Digitalización de los contornos georreferenciados de todas las secciones censales, según coordenadas UTM, huso 28, 29, 30 y 31]. INEbase. <https://www.ine.es/uc/TGWSB6oe>
- [11] R CORE TEAM (2019). *R: A language and environment for statistical computing*, R Foundation for Stastitical Computing, Viena, Austria <http://www.R-project.org/>

- [12] RSTUDIO. *R Interface to Leaflet Maps*. Leaflet for R de <https://rstudio.github.io/leaflet/>
- [13] GGIRAPH PACKAGE *makes 'ggplot' graphics interactive*. <https://davidgohel.github.io/ggiraph/index.html>
- [14] KIPP, A. (2017, 30 MAYO). *Getting started with shinyapps.io*. Shiny. <https://shiny.rstudio.com/articles/shinyapps.html>
- [15] ECIE10ES. Edición electrónica de la CIE-10-ES Diagnósticos *Clasificación Internacional de Enfermedades 10.^a Edición Española*. https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_10_mc.html
- [16] MARTÍN, T (2021). *Diseño y utilización de una herramienta de visualización de datos a escala municipal*. Trabajo de Fin de Grado en la Universidad de Valladolid.