



Universidad de Valladolid

CURSO 2018-2019

Facultad de Filosofía y Letras

Grado en Periodismo

¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados

Alumna: Sara Becerro Alonso

Tutora: Eva Campos Domínguez

Co-tutora: Dafne Calvo Miguel

Universidad de Valladolid

Departamento: Historia Moderna, Contemporánea y de América, Periodismo, Comunicación Audiovisual y Publicidad. Área de Periodismo.

Convocatoria: Ordinaria. 20/06/2019.

Modalidad profesional

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

TÍTULO: ¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados.

AUTORA: Sara Becerro Alonso

TUTORA: Eva Campos Domínguez

CO-TUTORA: Dafne Calvo Miguel

RESUMEN:

Con un sistema democrático instaurado hace 42 años en España, el sistema de representación parlamentaria no representa igualmente a todos los grupos de la sociedad, cuestión que se ha observado en los últimos años con movimientos sociales como el 15M que han dado lugar al surgimiento de nuevos partidos políticos y a cambios en cuanto a lo que la representación dentro del órgano más importante del país, el Congreso de los Diputados, se refiere.

El descontento con la ley electoral vigente, las desigualdades e infrarrepresentaciones o la dificultad de encontrar información básica sobre los mandatarios, es un hecho. Todo ello, junto con el cambio en la comunicación desde la llegada de Internet y la reciente inclusión y auge del “periodismo de datos” en los perfiles de los profesionales y las redacciones, tienen cabida en las páginas que aquí se contemplan.

Por ello, el foco de la investigación es mostrar la representación que tiene la población española en el Congreso de los Diputados durante la XII Legislatura, comenzada en 2016, y analizar si la apertura del ejercicio político a mayor parte de la población se traduce en mayor representatividad. Estas preguntas se responden a través de un reportaje, a través de la aproximación al periodismo de datos, y en formato blog para que permita desarrollar diversas habilidades adquiridas en el Grado. Además, se incluye la figura del programador, muy presente e indispensable en este tipo de periodismo y se experimenta con el lenguaje de programación R para la inclusión de la interactividad en las visualizaciones gráficas que se proporcionan.

PALABRAS CLAVE: Representación, Ley D’Hondt, Población, Congreso, Periodismo de datos

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

TITLE: Who represents us? Journalistic analysis of data on the representation of the spanish population in the Congress.

AUTHOR: Sara Becerro Alonso

TUTOR: Eva Campos Domínguez

CO-TUTOR: Dafne Calvo Miguel

ABSTRACT:

With a democratic system established 42 years ago in Spain, the system of parliamentary representation does not represent all groups of society equally, an issue that has been observed in recent years with social movements such as the 15M that have led to the emergence of new political parties and changes in what representation within the country's most important body, the Congress of Deputies, refers to.

Dissatisfaction with the current electoral law, inequalities and under-representations, or the difficulty of finding basic information about the presidents, is a fact. All of this, together with the change in communication since the arrival of the Internet and the recent inclusion and rise of "data journalism" in the profiles of professionals and newsrooms, have a place in the pages contemplated here.

Therefore, the focus of the investigation is to show the representation of the Spanish population in the Congress of Deputies during the XII Legislature, begun in 2016, and to show whether the opening of the political exercise to a greater part of the population translates into greater representation. It will be reflected in a report, using and testing data journalism, and in blog format to develop various skills acquired in the Degree. It also includes the figure of the programmer, very present and indispensable in this type of journalism and experiments with the programming language R for the inclusion of interactivity in the graphic visualizations that are provided.

KEY WORDS: Representation, Law D'Hondt, Population, Congress, Data journalism

ÍNDICE

1. Introducción.....	6
1.1 Justificación.....	7
1.2 Objetivos.....	13
1.3 Marco teórico. Fundamentos teórico-académicos.....	13
1.3.1 Periodismo de datos.....	15
1.3.2 La ley electoral en España: el sistema D´Hondt.....	20
1.3.3 Representación en la política parlamentaria.....	21
2. Plan de trabajo: metodología.....	27
2.1 Fases de la investigación.....	28
2.1.1 Extracción de datos del Congreso de los Diputados.....	28
2.1.2 Extracción de datos del INE.....	29
2.1.3 Análisis de datos: uso del lenguaje de programación R.....	32
a) Elección del software para análisis y visualización.....	32
b) Pasos en la elaboración de un análisis de datos.....	34
b.1) Lectura de datos.....	35
b.2) Limpieza de datos.....	35
b.3) Transformación (Análisis).....	38
b.4) Comunicación interactiva.....	42
2.1.4 Redacción del contenido web y del presente trabajo académico.....	46
2.1.5 Diseño de la web.....	48
3. Conclusiones.....	49
4. Bibliografía.....	52
5. Anexos.....	55

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

A mis padres, Paloma y Víctor, por el apoyo y la lucha incondicional.

A Dani, por ser mi compañero en este trabajo. Y mi hermano en la vida.

A Rodri, por estar siempre al pie del cañón. Por el cariño y la paciencia.

A mi hermana, María, por enseñarme el significado de constancia y esfuerzo.

A mi abuelo Ani y a mi padrino Maxi. Porque sólo muere lo que se olvida.

A todxs los que han estado apoyándome, cerca y lejos, en este tiempo.

A Dafne y Eva, mis tutoras, porque sin ellas la pasión no hubiese sido la misma y este trabajo no hubiese sido posible.

1. Introducción

Albert Einstein (1879-1955): “Mi ideal político es el democrático. Cada uno debe ser respetado como persona y nadie debe ser divinizado”.

La importancia del sufragio universal dentro del sistema actual ha quedado demostrada a lo largo de los años, ya que “[...] no podemos entender como democracia a un sistema de gobierno que no reconozca el sufragio universal” (Abenza et al., 2017), y que está establecido por la ley, que cumplió 126 años el mismo 26 de junio en el que se realizaron las elecciones de la Legislatura que es objetivo de esta investigación.

“A lo largo de la historia hemos constatado que la democracia es el menos malo de los sistemas de gobierno. El rumbo de la sociedad en las democracias lo señalan las corrientes mayoritarias que se manifiestan y se movilizan en las poblaciones a través de representantes elegidos periódicamente mediante convocatorias electorales” (Pérez, 2014¹).

Pero, el voto de los ciudadanos españoles, no vale lo mismo en todas las circunscripciones y los escaños se distribuyen en función de diversos factores que establece la ley electoral, conocida como Ley D’Hondt. Aunque, la representación va más allá y existen sectores y grupos de la sociedad que no se sienten representados, aspectos que durante los últimos años se han visto en movimientos sociales como el 15 M, mostrando una crisis en la legitimidad política.

“La acción de los medios de comunicación, los liderazgos entre bambalinas y el efecto anestesia de la democracia representativa forman parte de las críticas lanzadas por el 15 M, que a mediados de 2011 ocupó las plazas de España para gritar ‘¡No nos representan!’” (Yanes, 2017²).

Por ello la cuestión es: ¿sí es correcta la representación política que tiene España actualmente. ¿Están infrarrepresentados o sin voz algunos grupos sociales?

¹ Se puede encontrar en:

https://www.infolibre.es/noticias/club_info_libre/librepensadores/2014/12/18/la_democracia_representativa_espana_25565_1043.html. Última consulta: 14-06-2019.

² Se puede encontrar en: <https://politikon.es/2017/12/14/los-conceptos-de-representacion-en-el-15-m/>. Última consulta: 14-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

¿Tienen algo que ver las características de la población con las de los mandatarios?

Y por ello, ¿se debe realizar una apertura y diversificación del ejercicio político a más sectores de la sociedad? ¿Qué futuro pueden esperar los jóvenes, las mujeres, los extranjeros y otros colectivos? ¿Es necesaria, por ejemplo, la alta cualificación académica de los dirigentes para ejercer la función pública de representante político?

Y, dentro del campo de la información y el periodismo, ¿en qué medida es importante la figura del periodista y cómo ha afectado al periodismo? ¿El llamado “periodismo de datos” es el futuro?

Mediante distintas investigaciones y en los resultados obtenidos de ésta misma, se ha intentado dar respuesta a las anteriores preguntas.

Por ello, como cuestión principal, se ha realizado esta investigación en la que a través de los distintos capítulos dentro del marco teórico se ofrecerán los fundamentos teórico-académicos objeto de ella.

La investigación como tal, plasmará en un reportaje en formato multimedia, ordenado en distintos artículos donde se incluye: la extensa recogida, limpieza e interpretación de los datos y el trabajo profesional realizado que consta de distintos elementos, entre ellos mapas interactivos, gráficos e información de alcance ciudadano y totalmente basada en los datos obtenidos a través de varios organismos oficiales.

1.1 Justificación

Durante los últimos años en España, la política, los dirigentes que componen los cargos públicos, sus estudios y otros factores han sido muy discutidos, puestos en duda y nombrados en los medios de comunicación, ya que existe una sensación de “no representación”, entre gran parte de los ciudadanos que conforman el país.

Ese, fue el lema que encabezó uno de los grandes movimientos sociales en España hace ocho años, el 15 M, junto a la reivindicación de una “democracia

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

real ya”: “los ciudadanos querían participar en la vida política y en las decisiones que afectaban a su día a día y a su futuro sin tener que formar parte de ninguna organización política” (Público, 2017³).

La crisis de representación surgida con el 15-M mostró la sensación de una parte de la población de que la política estaba desprotegida frente a un poder económico que no rendía cuentas en las urnas; lo que se traducía como la falta de autonomía del poder político y la incapacidad de dichos políticos de actuar siguiendo las preferencias de los ciudadanos (León, 2018⁴).

Tras las movilizaciones, el surgimiento de nuevos partidos políticos que daban voz a algunos colectivos hasta entonces sin ella, el auge del feminismo en los últimos años, el posicionamiento de los jóvenes, la exclusión del voto de los extranjeros en las elecciones generales o la importancia de la trayectoria académica de los dirigentes, unido al avance de las nuevas tecnologías y el desarrollo de la política mediante las redes sociales, también han afectado al periodismo.

Dentro del marco de la información, han aparecido nuevos métodos de trabajo y especialización, como el Periodismo de Datos que en sus inicios fue denominado periodismo de precisión. Guarda una estrecha relación con el periodismo de investigación y aunque tiene un surgimiento aún reciente, es anterior a Internet y poco a poco resulta más necesaria la especialización y realización de este tipo de trabajos; además de la gran demanda existente en las redacciones de profesionales en este campo.

“Este cambio tecnológico desafía al periodista. Ante tanta información almacenada en bases de datos y con tantas historias que contar, es necesario formar a los periodistas en el tratamiento de esos datos” (López, 2018, pp. 13).

Como afirmó Dader (1994) refiriéndose al periodismo de precisión, es “irregular y disperso” y aún no cuenta con los suficientes respaldos, además de ser “muy

³ Se puede encontrar en: <https://www.publico.es/sociedad/15-m-cambio-vida.html>. Última consulta: 12-06-2019.

⁴ Se puede encontrar en: https://elpais.com/elpais/2018/04/03/opinion/1522763915_939400.html. Última consulta: 12-06-2019.

deficitario desde el punto de vista de la formación metodológica”. Por ello, cada vez más, este tipo de periodismo que muestra historias de interés ciudadano con datos es reclamado por la población y se está haciendo necesario.

“El periodismo de precisión guarda una indudable relación con el de investigación, pero no se identifican entre sí. Ni todo el periodismo de investigación aplica las estrategias metodológicas y conocimientos instrumentales del periodismo de precisión, ni este último se aplica tan sólo en proyectos de investigación periodística. La línea divisoria entre ambos está en el uso sistemático de métodos de investigación social por parte de la nueva especialidad” (Dader y Gómez, 1993, pp. 102).

Con el avance de las nuevas tecnologías y las grandes posibilidades que ofrece Internet, los políticos y los Gobiernos utilizan este “reciente” medio para la interacción con la ciudadanía e incluso para sus campañas electorales: conceptos como el de ciberdemocracia, transparencia o gobierno abierto, cada vez son más enunciados e incluidos en la vida y actividad política de los representantes.

También, se ha asistido en los últimos tiempos a otra especialización: la de los periodistas en Internet; y a un desarrollo de la prensa especializada en este medio, que cada vez rige más nuestra sociedad mediante, por ejemplo, la influencia de las redes sociales. Internet cuenta con cuatro elementos que han cambiado la comunicación: hipertextualidad, multimedialidad, interactividad y frecuencia de actualización.

“Un contexto que propicia la necesidad de redefinir determinados aspectos del ejercicio periodístico ante el nuevo Periodismo digital que transforma la figura tradicional del informador en un periodista multimedia, con rutinas nuevas en diferentes soportes, multitarea y con uso de multilinguaje” (Sánchez; Campos y Berrocal, 2015, pp. 188).

Por ello, el planteamiento de este Trabajo de Fin de Grado también consiste en un proyecto profesional en Internet: un reportaje periodístico de datos plasmado en una web alojada en Internet, en formato blog donde se visualizará todo el contenido de la investigación. La interactividad es uno de los objetivos de este

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

proyecto, además de acompañar el contenido con hipertextos, visualizaciones y aprovechar las distintas opciones y potenciales que Internet proporciona.

“Desde hace un tiempo el periodismo está sufriendo una fuerte reconversión, adaptándose a las nuevas tecnologías de la información y comunicación y volviendo a su más pura esencia, al fin y al cabo, reinventándose.

El periodismo digital lleva poco más de una década en el panorama de los medios de comunicación, pero en escaso tiempo ha conseguido hacer sombra a su mayor competidor, el gigante de la prensa escrita en papel [...]. Ha conseguido consolidarse satisfactoriamente en el panorama comunicativo como un medio más, de referencia y de enorme difusión” (Blanco, 2016⁵).

Tras las pasadas elecciones del año 2016, el 19 de julio de ese mismo año se inicia la XII Legislatura y el Congreso de los Diputados aprueba la investidura de Mariano Rajoy como presidente del Gobierno el 29 de octubre de 2016, aunque no acaba su mandato.

En un clima social de cambios en la población y con una convocatoria de elecciones para el 28 de abril de 2019, tras la primera moción de censura que prospera con su respectivo cambio de presidente y diputados se realizará: un análisis periodístico de datos de los representantes iniciales que formaron parte de la XII Legislatura que comenzó el 19 de julio de 2016 y que terminó el 5 de marzo de 2019, frente a los datos registrados, de la población española.

Desde el inicio de la Legislatura hasta la convocatoria de elecciones en marzo de 2019 ha habido cambios de 43 de los diputados que componen el Congreso, debido a los acontecimientos ya citados, y esta investigación ha contado con diversas controversias en cuanto a los datos, que serán expuestas a lo largo de estas páginas.

El trabajo, adscrito a la modalidad profesional, surge con el objetivo de mostrar la representatividad real a partir de datos obtenidos del Instituto Nacional de

⁵ Se puede encontrar en: https://elpais.com/elpais/2016/04/13/opinion/1460540302_620130.html.
Última consulta: 11-05-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Estadística⁶ -para la población-y el Congreso de los Diputados⁷ -para los dirigentes -; mediante la comparación de una serie de factores que son: a) edad, b) sexo, c) estado civil, d) procedencia, e) nivel académico y que serán representados a través de distintos artículos con gráficos, datos e infografías explicativas, que darán paso a los resultados obtenidos. Se podrá ver, quiénes de los políticos se parecen en algo a la población española, si realmente la representan o si el sistema necesita un cambio.

Dentro del blog se tratarán las infrarrepresentaciones de diversos colectivos, como el LGTBI, el papel de las mujeres dentro de la sociedad y la política, el de los jóvenes, la situación de los extranjeros o el reparto de votos según la procedencia, en entradas con datos reales, que parten de una base documentada mediante una bibliografía explícita.

Además, el formato web donde se plasmará el reportaje, me ha permitido desarrollar diversos conocimientos que he desarrollado durante el Grado en Periodismo en la Universidad de Valladolid, entre los que se mezclan: la investigación y documentación a través de diversas fuentes, además de la redacción periodística extensa y la elaboración de distintas visualizaciones sobre los datos:

- Ciberperiodismo: a la hora de realizar la web y de manejar y crear un medio digital para plasmar un reportaje de datos.
- Redacción Periodística: en la realización del contenido de los artículos que van dentro del blog.
- Géneros Periodísticos: para diferenciar los distintos géneros dentro del periodismo y elegir el adecuado para este trabajo.
- Ética y Deontología Periodística: en el tratamiento de ciertos temas como son los extranjeros o las mujeres, y establecer los límites desde la objetividad y la veracidad que aportan los datos.

⁶ Consultado a partir de: <https://www.ine.es/>. Última consulta: 10-06-2019.

⁷ Consultado a partir de: <http://www.congreso.es/portal/page/portal/Congreso/Congreso>. Última consulta: 16-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

- Periodismo de Investigación: en la utilización de herramientas para desarrollar mapas interactivos, infografías explicativas y diversas plataformas para tener datos limpios y claros. También a la hora de realizar una investigación ha sido primordial.

Al adquirir esta modalidad, hemos contado con la colaboración del Grado de Telecomunicaciones de la Uva, a través del asesoramiento del alumno Daniel Álvarez Hernández, a la hora de aportar los conocimientos sobre la parte informática y de programación del trabajo, ya que como bien afirma Crucianelli el periodismo de datos:

“Se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella, se trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico” (2013, pp.106).

Ha aportado los conocimientos técnicos para las tareas de programación, para lo que se ha utilizado lenguaje “R” que es “una serie de paquetes R y un entorno para el cálculo estadístico con grandes volúmenes de datos mediante el uso de la programación de alto rendimiento estadístico” (Wikipedia, 2018⁸).

Ha dado lugar a la recogida y rápida clasificación de datos en tablas y las divisiones de los filtros de los componentes del Congreso de los Diputados, además de para realizar los gráficos interactivos recogidos dentro de la web con los que la población podrá marcar los filtros correspondientes y saber quiénes les representan, que se explicará con más detalle en la parte de la metodología.

En definitiva, se trata de un trabajo de investigación de la representación de la población en el Congreso de los Diputados complementado con el desarrollo de diversos artículos dentro de un reportaje periodístico de datos plasmado en formato blog en la web.

⁸ Se puede encontrar en: https://es.wikipedia.org/wiki/Programaci%C3%B3n_con_datos_masivos_en_R. Última consulta: 18-05-2019.

1.2 Objetivos

El objetivo principal de esta investigación es realizar un análisis periodístico de datos incluyendo la programación, presentado en un formato de web/blog, para mostrar la representación que tiene la población española dentro del Congreso de los Diputados a través de distintas variables.

Dicho objetivo, se completa y complementa con otros secundarios que son los siguientes:

- 1- Conocer quiénes fueron los representantes en el Congreso de los Diputados durante la XII Legislatura y si la población está representada en ese período o mandato.
- 2- Identificar esa representación de la población en torno a las variables de sexo, edad, orientación, estado civil y estudios.
- 3- Comparar el grado de representación, según dichas variables, por cada partido político dentro de la cámara.
- 4- Experimentar con el periodismo de datos y con lenguajes de programación para comprobar su aplicación en el ejercicio periodístico utilizando una web/blog para la visualización de los contenidos y resultados del extenso análisis de datos recogidos.

1.3 Marco teórico. Fundamentos teórico-académicos

El periodismo y el motivo de esta investigación podría resumirse con esta frase de Navarro (2010):

“Tengo un defecto gravísimo para el ejercicio de mi profesión. Creo que el periodismo es un servicio público donde debe primar la verdad y no los intereses de los políticos, de los empresarios, de los banqueros, de los sindicatos o del que me paga” (2010, pp 2).

Dentro de un marco de inestabilidad política dentro del país y de desconcierto de los ciudadanos, con grandes cambios dentro de la vida política, la primera

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

moción de censura en España que sale adelante, y unas elecciones controvertidas, el trabajo que aquí se contempla y el objetivo de esta investigación se basa en distintas premisas.

Además de la obtención e interpretación de los datos obtenidos para la web, el trabajo realizado también ha venido precedido anteriormente por una revisión bibliográfica acerca de otras pautas fundamentales para la investigación periodística.

Por tanto, las motivaciones principales de éste trabajo han sido: el interés por la política y la representación en mi país, España, de los que están al frente de los órganos de decisión; la desmotivación y el descontento, además de la crispación, observados en un porcentaje alto de la población, sobre todo entre los ciudadanos más jóvenes, un tema controvertido y que, como afirma Arias:

“Cada vez es más difícil “politizar” a los jóvenes porque para ellos la política clásica hace tiempo que ha dejado de interesarles. Se balancean entre la indiferencia y el rechazo al sistema” (2014)⁹.

Por otro lado, el interés personal por la modalidad surgida hace relativamente pocos años dentro del periodismo, el llamado periodismo de datos, que está en crecimiento y auge aunque cuenta con pocos conocimientos por parte de los nuevos profesionales que salen cada año de las universidades y es poco explorado por los periodistas que llevan tiempo en el mundo de la información; pero que va tomando fuerza paulatinamente y que puede ser uno de los futuros más beneficiosos que pueda tener la profesión.

Como se recoge en *infoLibre*¹⁰ en 2013, Mar Cabra, una de las pioneras en el periodismo de datos en España, afirmaba que:

“Lo bueno de los datos es que permiten hablar de problemas y de patrones que ocurren de manera sistemática. Ya no se trata de algo que

⁹ Se puede encontrar en:

https://elpais.com/internacional/2014/02/25/actualidad/1393358639_000137.html. Última consulta: 10-05-2019.

¹⁰ Se puede encontrar en:

https://www.infolibre.es/noticias/medios/2013/04/17/datos_periodismo_evidencia_2621_1027.html. Última consulta: 10-05-2019.

te dice una fuente, que te cuenta un conocido o un político, sino que los datos posibilitan hablar de algo que es un hecho y que puedes demostrar.”

La dificultad de encontrar información por parte de los ciudadanos, de aspectos tan básicos como son la edad, el porcentaje de hombres y mujeres en el Congreso de los Diputados, la representación de colectivos como el LGTBI, la provincia por dónde se presenta cada representante, o el grado de estudios con el que cuentan, por la gran cantidad de datos que supone ese análisis, debe ser transformada para el fácil entendimiento de la población.

La sociedad cuenta con que, como afirmó Campos (2011, pp. 62) citando a Font (2001), “siguen predominando las reticencias hacia la apertura de nuevos espacios participativos por parte de quienes ostentan el poder”, pero con los distintos cambios, entre ellos la Ley de Transparencia, como afirma López:

“La apertura de información pública a los ciudadanos desde las diferentes administraciones públicas, genera un nicho de grandes volúmenes de datos. Información que los periodistas, mediante herramientas de limpieza y análisis, pueden utilizar para encontrar noticias exclusivas y de gran repercusión mediática. Esta es, sin dudas, la edad de oro de la información masiva” (2018, pp.19).

1.3.1 Periodismo de datos

El periodismo de datos, es una disciplina relativamente moderna y es definida por López como:

“Periodismo que surge como un nuevo método para obtener informaciones que se encuentran escondidas en grandes bases de datos. Mediante tratamientos informáticos se extraen los datos que luego se muestran en diferentes tipos de noticias, reportajes, visualizaciones o aplicaciones web, entre otros.

Una disciplina que tiene su origen en el Periodismo de precisión y en el CAR (*Computer Assisting Reporting*) o lo que es lo mismo, Periodismo Asistido por Computadora; corrientes innovadoras de trabajar la

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

información que aparecieron gracias al periodismo de investigación” (2016, pp. 8).

Como bien afirma la misma autora y centrándonos en su investigación, mediante este tipo de periodismo los profesionales de la información, “hacen comprensibles” los datos que obtienen en grandes cantidades “mediante artículos, visualizaciones y aplicaciones web” (2016, pp. 6). Esta disciplina tiene su origen en Estados Unidos y se ha ido implantando en los medios de comunicación, también en España, gracias a los avances en las tecnologías informáticas y a Internet, aunque su surgimiento es anterior.

Se pueden cercar sus antecedentes en el periodismo de precisión de Philip Meyer, “un nuevo método periodístico que hará hincapié en la revisión técnica de la metodología científica empleada en cualquier tipo de cuantificación susceptible de trascendencia noticiosa” (Dader y Gómez 1993, pp. 101), y en su primer trabajo realizado en 1959.

Evolucionó, diez años más tarde, “al periodismo asistido por computadora o CAR. Una disciplina periodística que introduce el uso de hojas de cálculo para procesar la información” (López, 2016, pp. 11). En constante evolución, llegó hasta el “Data Journalism” o periodismo de datos. “Paul Bradshaw, profesor de la Universidad de Birmingham y Adrian Holovaty, periodista estadounidense, fueron los grandes impulsores de esta disciplina periodística” (López, 2016, pp. 12).

Como se afirma desde una de las presentaciones a los cursos de verano sobre periodismo de datos, de la Universidad Complutense de Madrid:

“Desde la entrada en vigor en 2014 de la Ley de Transparencia del Gobierno de España y el emergente concepto de los datos masivos (Big Data), muchos periodistas se vienen especializando en el Periodismo de Datos” (López y Vivar, 2015¹¹).

Tras un trabajo de investigación y buceo entre reportajes, artículos y demás piezas periodísticas que dieran impulso a la idea establecida para este trabajo,

¹¹ Se puede encontrar en: <https://www.ucm.es/data/cont/docs/71-2015-04-15-71311.pdf>. Última consulta: 15-04-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

un reportaje¹² realizado por el periódico americano *The Guardian*, uno de los pioneros y con mejores resultados en este tipo de periodismo, fue el detonante.

“Esta disciplina está siendo un referente en grandes medios de comunicación a nivel internacional como *The New York Times* o el británico *The Guardian* que están apostando fuertemente por los datos y las visualizaciones como diferenciador de sus noticias y clave para atraer lectores en un momento difícil para el periodismo” (López y Vivar, 2015¹³).

El reportaje, realizado por Nadja Popovich y Feilding Cage, en noviembre de 2014, muestra los parecidos mediante un mapa y unos filtros interactivos, de la población estadounidense con sus representantes, y fue el modelo a seguir para el caso del Congreso de los Diputados en España. También realizan otro reportaje¹⁴, Nick Evershed, Ri Liu y Anna Livseycon, con características parecidas sobre el el 45º parlamento de Australia.

Figura 1: captura propia de la web. Inicio del reportaje: aparecen los filtros para marcar y la introducción; los autores y la fecha de publicación.



The screenshot shows the top navigation bar of The Guardian website with categories: News, Opinion, Sport, Culture, Lifestyle, and More. Below the navigation bar, the article title is "Are you reflected in the new Congress?" by Nadja Popovich and Feilding Cage. The article is dated "Thu 6 Nov 2014 16:39 GMT" and has 3,566 views. The main text begins with: "Despite the record number of women and the first black senator elected in the south since Reconstruction, the new US Congress is still largely male and largely white. A person's gender, race, sexual orientation, or age doesn't necessarily mean they represent the views of a whole demographic, but lack of diversity could result in certain concerns not being heard - or not heard loudly enough. Click the categories below to find yourself in the new Congress. This graphic was last updated on 7 January 2015". Below the text, there are interactive filters for "If you are" (Female, Male, and Hispanic, Black, Asian, White*, Native American, and LGB, Heterosexual, and under 45, 45-64, over 64) and "and have" (not attended college, some college, associate's degree, bachelor's degree, higher degree).

Fuente: The Guardian.

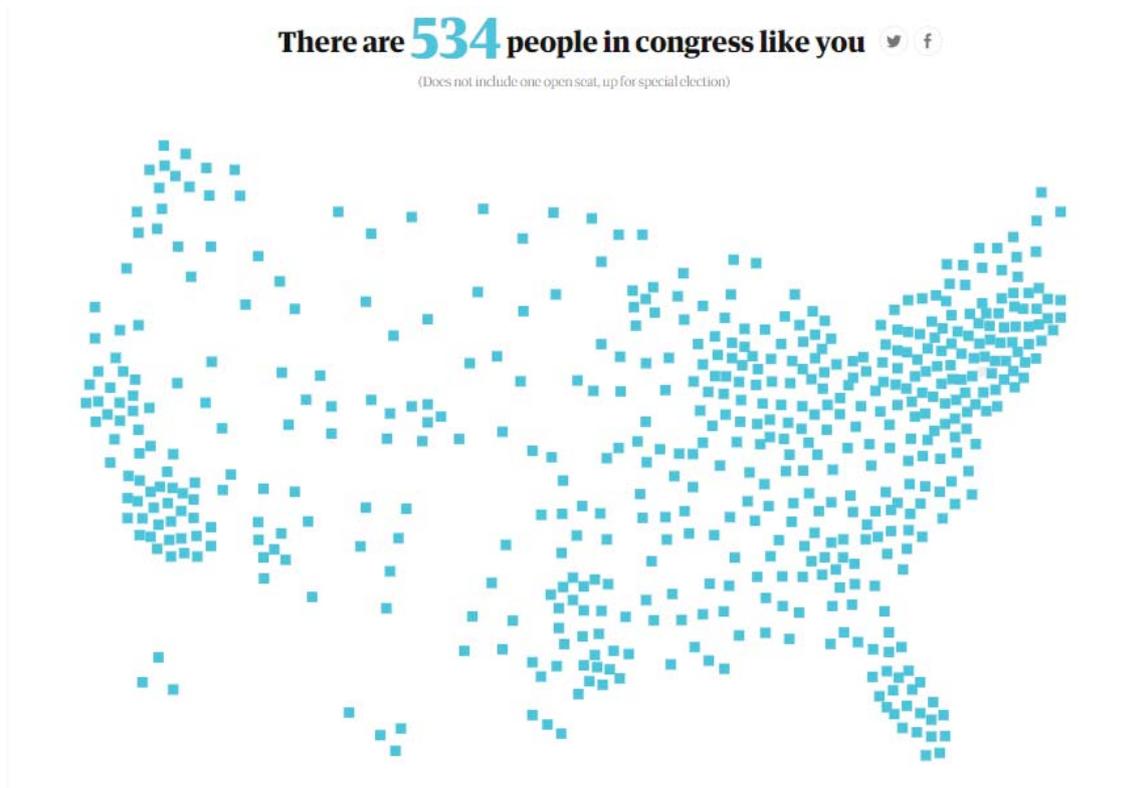
¹² El reportaje se puede visualizar en: <https://www.theguardian.com/us-news/ng-interactive/2014/nov/06/-sp-congress-diversity-women-race-lgbt-are-you-represented>. Última consulta: 10-06-2019.

¹³ Se puede encontrar en: <https://www.ucm.es/data/cont/docs/71-2015-04-15-71311.pdf>. Última consulta: 15-04-2019.

¹⁴ Visualizado en: <https://www.theguardian.com/australia-news/datablog/ng-interactive/2016/aug/31/are-you-reflected-in-the-new-parliament-diversity-survey-of-australian-politics>. Última consulta: 10-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Figura 2: mapa donde en función de los filtros que marques anteriormente aparecerán los representantes correspondientes. En la captura, no hay marcado ningún filtro. Además, pasando el cursor por encima de los cuadros aparece información sobre el representante.



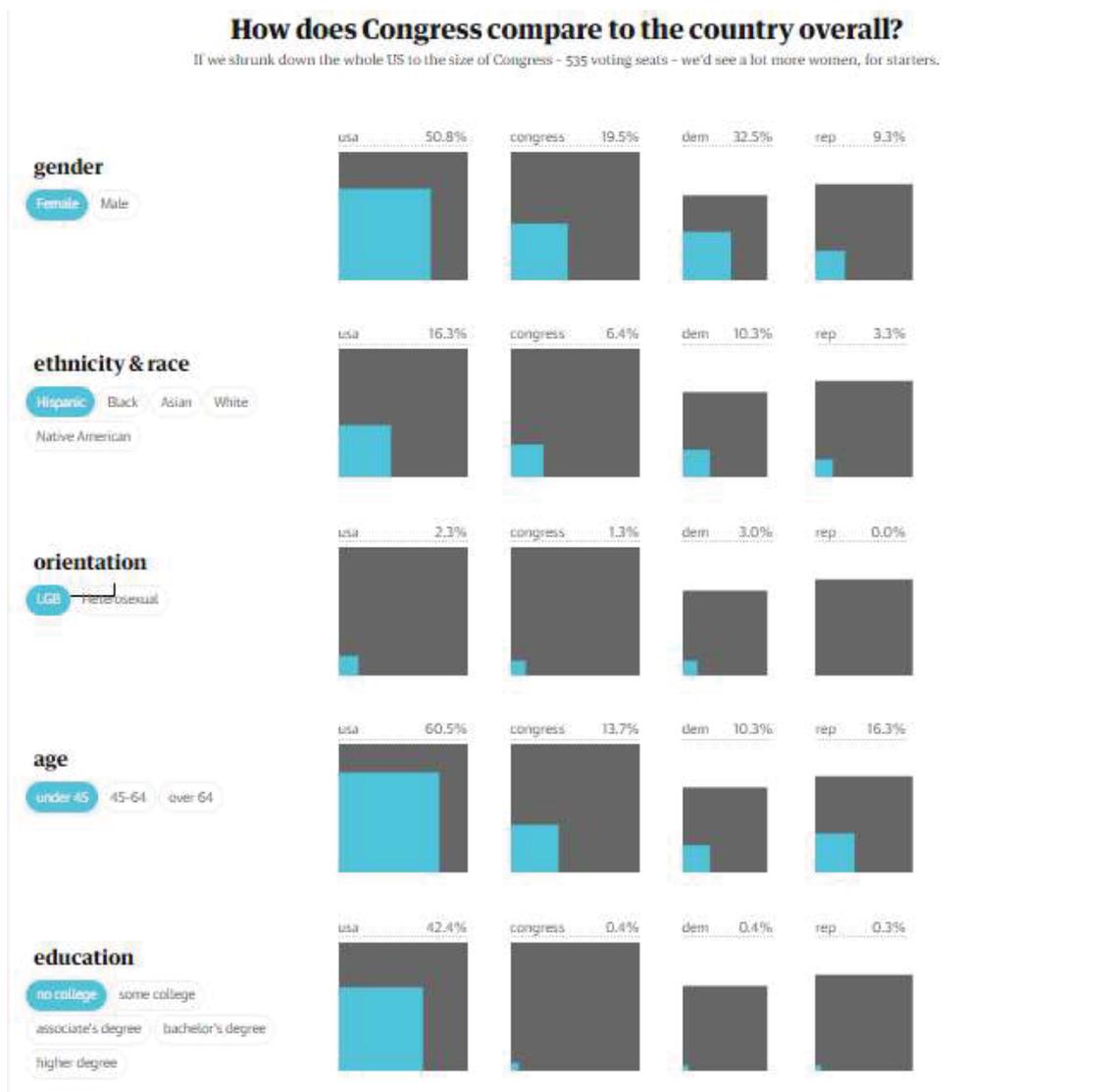
Fuente: The Guardian.

Para ello, utiliza una serie de datos como son el sexo, la edad o los estudios mediante los que, al visitar la web, puedes marcar obteniendo como resultado a los senadores y representantes americanos que tienen exactamente esas mismas características. Aparecen ubicados según el lugar al que representan y además muestran todos los porcentajes y gráficos con la información de cada uno de los filtros desglosado.

Se ha de aclarar, que por algún tipo de problema algunas veces la web del medio *The Guardian* no realiza correctamente la interacción y se queda bloqueado sin cambiar los porcentajes o el resultado total, aunque rara vez da este tipo de problemas, por la frecuencia de consulta ha aparecido en alguna ocasión este error.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Figura 3: porcentajes mediante gráficos de los resultados obtenidos con todo lo anterior. Igualmente, en función de lo que se marque aparecen los resultados correspondientes.



Fuente: The Guardian.

Para todo ello, se necesita una recopilación de grandes cantidades de datos, además de su interpretación por parte del periodista, pero también se requieren conocimientos informáticos, de limpieza de datos y programación para la parte interactiva, de visualización y de compresión de todos esos datos en porcentajes sencillos.

“El periodista generalmente ha destacado por ser una persona individualista en el ejercicio de su profesión; pero, con esta nueva disciplina se crea la necesidad del trabajo en equipo. El profesional de la

comunicación necesita la ayuda de un programador y un estadístico para elaborar bien la información” (López, 2018, pp. 14).

1.3.2 La ley electoral en España: el sistema D’Hondt

Uno de los aspectos que más concierne a la población española y que es el motor de la elección de gobernantes y partidos políticos es la ley electoral con la que cuenta el país, la conocida Ley D’Hondt. Creada por el jurista y profesor belga Victor D’Hondt en 1878, aún en nuestros días España se rige por ella.

“La Ley D'Hont es un sistema electoral de cálculo proporcional que divide el número de votos emitidos para cada partido entre el número de cargos electos con los que cuenta cada circunscripción. En unas elecciones generales, la legislación española descarta las candidaturas que hayan obtenido menos del 3% de las papeletas. El resultado de cada una de las otras habrá de dividirse por el número de escaños en juego. Los representantes se conceden a estos cocientes por orden decreciente” (Velasco, 2017¹⁵).

Atendiendo a esta definición y a los resultados obtenidos en esta investigación, la ley electoral presente en España no es concretamente la más correcta, la que mejor representa, ni tampoco es totalmente justa, como se podrá observar a lo largo de la investigación.

A pesar de ello, algunos autores como Bilbao Arrese afirman que:

“En nuestro país, debido a la presencia de partidos nacionalistas con apoyo electoral, necesitamos un sistema electoral que combine la representatividad y la capacidad de generar gobiernos estables. El sistema de proporcionalidad corregida de D'Hondt, aplicado en circunscripciones de tamaño reducido, es la mejor solución conocida a las citadas necesidades” (1994, pp. 320).

¹⁵ Se puede encontrar en: <https://wsimag.com/es/economia-y-politica/18650-que-es-y-en-que-consiste-la-ley-dhondt>. Última consulta: 16-06-2019.

Entre los estigmas de la población ante esta ley se encuentra en que una parte de la población no se siente representada con este sistema. Contando con los resultados obtenidos, dicha ley beneficiará por tanto a los partidos mayoritarios y que llevan más años en el campo electoral, perjudicando así a las nuevas formaciones.

De la misma manera, favorece a los partidos regionalistas con mucho apoyo en sus circunscripciones y perjudica a partidos de índole nacional pero que son apoyados dispersamente en distintas comunidades autónomas.

“El sistema electoral español ha recibido frecuentes críticas por sus tasas importantes de desproporcionalidad, que en una palabra quiere decir que los votos no tienen el mismo valor, sino que éste depende de una serie de circunstancias, y que el coste en votos de los escaños es asimismo desigual (algunos necesitan más votos que otros). España se caracteriza por ser uno de los países con más alta tasa de desproporcionalidad” (Soriano y Alarcón, 2001, pp. 119).

Por lo tanto, ¿debería cambiar el sistema de voto en nuestro país? ¿Realmente los ciudadanos votan al partido que quieren? O, por el contrario, ¿votan en función de dónde viven y del llamado “voto útil”, al que se apela constantemente en los medios¹⁶ y por parte de los partidos políticos en campaña electoral? En el reportaje, se puede encontrar una entrada que habla extensamente acerca de esta premisa.

1.3.3. Representación en la política parlamentaria

Entre uno de los centros más importantes de la investigación, relacionado con la representación en el Congreso de los Diputados, se encuentra el libro *“El muro invisible. Las dificultades de ser joven en España”* de los autores Luis Abenza, Berta Barbet, Silvia Claveria, Elena Costas, Jorge Galindo, Kiko Llaneras, Octavio Medina, María Ramos y Pablo Simón. Es una de las importantes

¹⁶El voto útil. Artículo de La Marea. <https://www.lamarea.com/2019/05/20/es-legitimo-llamar-al-voto-util/>. Última consulta: 12-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

aportaciones bibliográficas y de resultados en este trabajo, y se define propiamente como:

“Un proyecto común de un grupo de jóvenes académicos y profesionales independientes que se reunieron en 2010 con el fin de promover análisis, debates y políticas basados en el conocimiento de las ciencias sociales”.

Un aspecto importante: la edad. ¿Qué implica ser joven?

Uno de los aspectos analizados dentro de este análisis y que trata con claridad y estudios este libro, ha sido la edad. Los resultados de la investigación han sido los de un Congreso con personas de avanzada edad en la mayoría de los partidos políticos y poca representación de la población joven.

Como ya se afirma en una de las partes del reportaje, la franja de edad más numerosa entre la población española es la de entre 40 y 44 años, según desvelan las cifras de población de 2018 del Instituto Nacional de Estadística (INE), y en el Congreso de los Diputados está representada por 38 diputados, cerca de un 11% de su composición completa. Una infrarrepresentación tratada extensamente dentro de la web.

Como se afirma en el libro objeto de referencia en este epígrafe, “la generación nacida entre los años ochenta y el 2000”, aunque con un país “más próspero, más abierto y más libre [...], se ha dado de bruces con una triple crisis-económica, social e institucional-[...]. Esto ha tenido sus consecuencias en política, como la aparición de nuevos partidos” (Abenza, et al., 2017, pp.13).

Además, de cómo comentan algunos expertos en relación a los jóvenes, como el ya citado anteriormente, Arias, nadie se ha preocupado del interés de los jóvenes en cuanto a los temas políticos y podría haber alguna que otra sorpresa ya que como explica el autor:

“En una gran mayoría, son apolíticos ya que no confían en los partidos. Los consideran anticuados, lo que no significa que aborrezcan la democracia. Mal distinguen ya entre izquierdas y derechas. Son pragmáticos y pospolíticos. No ven excesiva diferencia entre progresistas

y conservadores. Para ellos son todos iguales, o casi. Y, sobre todo, no les tienen miedo” (2014).

El sentimiento surgido en esta generación de “precariedad, la falta de oportunidades, la ausencia de posibilidades para emanciparse y construir un hogar, y las deficiencias del sistema educativo”, unidas a “la ignorancia, la falta de interés (en el mejor de los casos) o la corrupción (en el peor) de nuestros servidores públicos a la hora de buscar soluciones”, puede tener el resultado de “una generación que podría acabar viviendo peor que la de sus padres” (Abenza, et al., 2017, pp. 14).

“Hay un pacto intergeneracional que está roto y que debemos recomponer”. Y como mencionan los anteriores autores y se concluye también en este trabajo “este proceso tiene que empezar desde la política, sin duda, pero muy especialmente desde las políticas (públicas)” (Abenza, et al., 2017, pp. 15).

Con mucha frecuencia dentro de la sociedad, hay un pensamiento mayoritario de que los jóvenes no participan en la vida política ni muestran ningún interés por ella, a la vez que reciben críticas cuando recurren a las movilizaciones o concentraciones callejeras para expresarse en relación a la política; pero si hay algo que ha quedado demostrado es que mayores y jóvenes “votan cada vez a formaciones políticas más diferentes”. Aunque, según las encuestas, son los de menor edad los que menos participan en las jornadas electorales.

Y, ¿qué pasa con las mujeres?

Como expresan los autores de Politikon:

“Es frecuente olvidar que un pilar del Estado de Bienestar de la posguerra era la familia tradicional, algo que se siente aún en nuestros días. [...] con las mujeres dedicadas al trabajo doméstico y al cuidado de los hijos, y los hombres al mercado laboral. [...] la sociedad estaba muy segregada en términos de género” (2017, pp. 130).

Aunque actualmente España encabeza la lista de países con más mujeres dentro de sus posiciones de poder, concretamente el quinto puesto a nivel mundial, y

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

tras los últimos comicios de 2019 se ha alcanzado la composición más igualitaria de la historia, la desigualdad es un hecho en muchos aspectos sociales.

“Las mujeres hasta hace muy poco han formado parte, mayoritariamente, de la parte invisible de la historia y la cultura” (Dader, 2000¹⁷).

En los últimos años hemos asistido al auge de las posiciones feministas y la lucha por la igualdad real entre sexos, lo que se ha convertido en un tema de rigurosa actualidad. Partidos políticos y medios de comunicación se hacen eco de todo lo relacionado con el sexo femenino y lo toman como prioridad. Politikon también hace un hueco entre sus páginas para tratar el tema:

“La conquista del espacio público por parte de las mujeres corresponde con una visión feminista y progresista que apuesta por un mundo con roles de género simétricos” (2017, pp. 137).

Explicado dentro de la web y en consenso con los resultados obtenidos y las distintas investigaciones realizadas por expertos, podemos llegar a la conclusión de que los partidos de izquierdas en España mantienen una fuerte apuesta por romper los techos de cristal y por la defensa de una sociedad feminista que, por el contrario, los partidos de derechas apuestan por políticas natalistas, el modelo de familia tradicional y un atraso en las libertades.

Como relatan Uriarte y Ruíz en su investigación:

“Los partidos de izquierda han asumido de forma mucho más importante el discurso feminista que los partidos conservadores, básicamente porque el discurso de la igualdad del feminismo encaja en las ideologías de izquierdas, pero de forma mucho más complicada en las ideologías de derechas.

El caso español es un ejemplo de la importancia de la existencia de partidos de izquierda fuertes para el aumento de la presencia de mujeres

¹⁷ Se puede encontrar en: <https://docplayer.es/58008031-La-mujer-como-categoria-emergente-en-la-produccion-de-noticias-pros-y-contras-de-una-nueva-visibilidad-periodistica.html>. Última consulta: 5-5-2019.

en las élites” (1999, pp. 211); conclusión a la que se ha llegado también con la presente investigación.

Como las mismas autoras afirman, España se ha situado entre los primeros países a la hora de equiparar a las mujeres dentro de los poderes políticos, pero, aun así, las diferencias entre ambos sexos siguen siendo de gran índole dentro de las elites políticas y esta desigualdad está todavía presente.

En los últimos tiempos se han venido produciendo cambios en la sociedad española, como ya se ha reflejado anteriormente y en mayor medida vienen dados por la política y la representación en el caso de España, en el Congreso de los Diputados.

Por lo tanto, el reportaje en web pretende interactuar e informar mediante datos sobre las distintas variables analizadas, a la población, ya que según afirmaciones recogidas por Campos citando a Held:

“Si las personas son sistemáticamente marginadas y/o pobremente representadas es probable que crean que rara vez se tomarán en serio sus opiniones y preferencias” (2001, pp.302).

Uno de esos casos es el del colectivo LGTBI, que ha cogido numerosa fuerza en la lucha por sus derechos y por la no discriminación en diversos aspectos dentro de la sociedad durante los últimos años; aunque por todos es sabido, que se ha discriminado a este colectivo durante muchos años y que ha sido difícil alcanzar el respeto y la normalidad en cuanto al tema, en unos países más que en otros.

Aun así, se puede afirmar que la situación de discriminación no ha desaparecido y actualmente en España, no se reconoce esta variable ni siquiera dentro de las propias categorías de clasificación del INE en cuanto a la población y hay una existencia mínima de diputados cuya condición sexual de LGTBI haya sido mostrada públicamente.

Citando a el Alto Comisionado de las Naciones Unidas para los Refugiados:

“La denominación de una persona como lesbiana, gay, bisexual, transgénero o intersex (LGBTI) asegura el reconocimiento legal de la orientación sexual o de la identidad de género como condición esencial

de la persona a ser protegida. La noción de LGBTI como un sector específico permite visibilizar y reconocer la discriminación histórica a la que han estado sometidas las personas que se encuentran en este supuesto, contribuyendo a ofrecer protección acorde a sus necesidades particulares” (ACNUR, 2014¹⁸).

Otro de los temas o infrarrepresentaciones que se tratan, es el de las personas de otras nacionalidades que viven en España, los cuáles solo tienen derecho a voto en los municipios donde estén empadronados. Existen acuerdos de reciprocidad con 12 países, pero solo afectan a las elecciones municipales por lo que, existen colectivos como los extranjeros residentes en el país que no tienen ninguna representación ni pueden votar para elegir a los representantes del Congreso de los Diputados.

Es decir, si en España se contabilizaron en 2018 más de medio millón de personas marroquíes, por ejemplo, son las mismas personas que no tienen ninguna representación a nivel nacional y, que sólo pueden ejercer el derecho a voto en las elecciones municipales.

Y, por último, como se plantea en la propia introducción, con la variable de la formación académica, hasta dónde llega la necesidad de una alta cualificación en los diputados/as. ¿Es necesario para el ejercicio de la política una cualificación específica?

A ello responden Coller, Jaime y Santana clara y concisamente en un artículo para *El País*:

“No es necesario puesto que es de las pocas profesiones para las que no se requiere una capacitación especial y sobre las que no hay controles formales de entrada” (2012¹⁹).

¹⁸ Se puede encontrar en: <https://www.acnur.org/fileadmin/Documentos/Publicaciones/2014/9872.pdf>.
Última consulta: 3-4-2019.

¹⁹ Se puede encontrar en: https://elpais.com/elpais/2012/07/25/opinion/1343235375_328774.html.
Última consulta: 8-4-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Pero, aun así, en nuestra política cada vez hay más porcentaje de licenciados, doctores o representantes frente a personas con menos títulos educativos o con otras experiencias en el mundo laboral y social.

Por ello, se ha planteado esta cuestión que se trata más extensamente en el reportaje, a partir de varios autores que han hablado sobre el tema, además de la comparativa con los estudios con los que cuenta la población. Con ello, se intentará dar respuesta una vez más, a que si la apertura del ejercicio político a personas con otras características, en este caso con otro tipo de estudios que no sean de nivel alto, puede suponer una mayor igualdad y representatividad de la población.

2. Plan de trabajo: metodología

Para llevar a cabo esta investigación, además de la consulta bibliográfica y la amplia recopilación de información, se ha realizado una web en formato de blog que consiste en distintas entradas que comparan los datos, según unas características básicas: a) edad, b) sexo, c) estado civil, d) procedencia y e) nivel académico), de la población en 2018 obtenidos del Instituto Nacional de Estadística²⁰, y los extraídos en la web del Congreso de los Diputados²¹, que han aportado las referencias sobre los 350 representantes, dentro de un reportaje web.

Un tema que apenas se trata desde el punto de vista académico y que los medios de comunicación muchas veces tratan desde perspectivas poco objetivas pero que, al contrario, está entre las principales preocupaciones de la población.

El trabajo, comenzado a finales del año 2018 (noviembre-diciembre) ha contado con dificultades en cuanto a los datos: 43 de los diputados y también las posiciones en el Congreso han ido cambiando con el desarrollo de la investigación y con ello han diferido porcentajes y datos o la moción de censura

²⁰ Consultado en: <https://www.ine.es/>. Última consulta: 18-06-2019.

²¹ Consultado en: www.congreso.es. Última consulta: 18-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

al gobierno del Partido Popular que presidía Mariano Rajoy, que propició un cambio de gobierno y por ello, de representantes y datos.

Por tanto, el análisis se realizó de los diputados que comenzaron la legislatura y las características que tienen en función a esas mismas variables respecto de la población española.

Tras la decisión del tema y de enmarcar, junto a las dos tutoras de esta investigación, el Trabajo de Fin de Grado en la modalidad profesional con la idea de desarrollar una web e incluir la ayuda de un programador para el análisis y limpieza de datos se pasó a la lectura de bibliografía explícita, para adquirir conocimientos técnicos y teóricos sobre el periodismo de datos.

Después, tras la información obtenida se decidieron las variables para analizar dentro de la investigación que se fijaron en las siguientes: a) edad, b) sexo, c) estado civil, d) procedencia y e) nivel académico. Todo el desarrollo de la investigación en los siguientes meses hasta junio de 2019 y los pasos que se han seguido se desarrollan a continuación.

2.1 Fases de la investigación.

2.1.1 Extracción de datos del Congreso de los Diputados.

La primera parte, tras aclarar el tema y con la idea en mente, fue la extracción de los datos del portal web del Congreso de los Diputados.

Uno a uno, se fue obteniendo la información de los 350 representantes y manualmente, se fueron recogiendo los datos en un libro de Excel²² separado por las distintas variables por cada uno de los diputados como se ve en la figura siguiente.

Figura 1: captura propia de una parte del libro de Excel donde se encuentra toda la información recogida por variables de los diputados/as del Congreso.

²² Se encuentra alojado completo en la web para descargar: <https://sarabecerro.wixsite.com/quien-nos-represent>.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	NOMBRE	SEXO	AÑO DE NAC	GRUPO PARLAMENTARIO	CIVIL	NIVEL ACAD	ORIGEN	NIVEL ACAD	CARGO							
2	Pedro, Aceo	Hombre	1935	Grupo Popul Separado		Técnico en P. Badajoz		Educación Su	Diputado. Alcalde de Mérida durante 16 años. Diputado autonómico de Extremadura durante ocho años.							
3	Joselita Ando	Hombre	1966	Grupo Vasco-Casado		Licenciado e- Gipuzkoa		Educación Su	Diputado. Concejal en Ondarra (1987-1990). Miembro de la Asamblea Nacional del PNV.							
4	Ernesto, Agu	Hombre	1961	Grupo Popul Casado		Técnico espe Santa Cruz		Educación Su	Diputado. Concejal del Ayuntamiento de Breña Baja (1996-1999). Consejero del Cabildo Insular de La Pa							
5	Famón, Agui	Hombre	1933	Grupo Popul Casado		Directivo Sar Guadaluja		No consta	Diputado. Presidente de la Comisión del Pacto de Toledo. Presidente del Instituto de Cris							
6	Naywa Miran	Mujer	1990	Grupo Confe Soltera		Licenciada e- Gipuzkoa		Educación Su	Diputada. Secretaria General de Podemos Euzkadi							
7	Maria Dolore	Mujer	1964	Grupo Popul Casado		Licenciada e- Alicante		Educación Su	Diputada. Concejal del Ayuntamiento de Alcoy hasta 2007, en la oposición y en el Gobierno							
8	Joaquín, Aba	Hombre	1973	Grupo Popul Casado		Licenciado e- Alicante		Educación Su	Diputado. Concejal del Ayuntamiento de Torreveja							
9	Miriam, Alco	Mujer	1978	Grupo Social Casado		Abogada y A. Cádiz		Educación Su	Diputada. Delegada Territorial de Igualdad, Salud y Políticas Sociales en Cádiz de la Junta de Andalucía (2							
10	Rigo Jesús,	Hombre	1973	Grupo Mixto Divorciado		Licenciado e- Navarra		Educación Su	Diputado. Fundador de la Asociación Síndrom UPI por la Discapacidad. Ex-Director de Capital Riesgo. Ex-							
11	Felix, Alorc	Hombre	1939	Grupo Confe Divorciado		Licenciado e- Tarragona		Educación Su	Diputado. Alcalde de Altafulla. Responsable de actividad sindical del Sindicato de Periodistes de Catalu							
12	Rosa Ana, Al	Mujer	1970	Grupo Confe Divorciado		Maestra de E Cantabria		Educación Su	Diputada. Secretaria de Relación con la Sociedad Civil y los Movimientos Sociales del Consejo Ciudadan							
13	José Jaime,	Hombre	1974	Grupo Popul Casado		Diplomado e Toledo		Educación Su	Diputado. Teniente Alcalde del Ayuntamiento de Fuemalvid (Toledo)							
14	Ana Isabel,	Mujer	1969	Grupo Popul Casado		Licenciada e- Huesca		Educación Su	Diputada. Concejala del Ayuntamiento de Huesca (2003-2016). Alcaldesa del Ayuntamiento de Huesca (2							
15	Ángelies, Ál	Mujer	1961	Grupo Social Casado		B.I.P. Impar Madrid		2ª Etapa de E	Diputada. Concejala electa del Ayuntamiento de Madrid. Ha trabajado en Fundación Mujeres como resp							
16	Felix, Álava	Hombre	1990	Grupo Ciudad Casado		Técnico Espe Cantabria		2ª Etapa de E	Diputado. Portavoz de la Comisión de Cultura y Deporte desde el 21/06/2018 al 03/03/2019. Secretario Si							
17	Maria Teresa	Mujer	1993	Grupo Popul Casado		Licenciada e- Badajoz		Educación Su	Diputada. Portavoz de Sanidad en la II Legislatura. Diputada autonómica en la Asamblea de Extremadura							
18	Javier, Antó	Hombre	1972	Grupo Social Casado		Diplomado e Soria		Educación Su	Diputado. Concejal en el Ayuntamiento de Soria. Secretario de Organización del PSOE Soria.							
19	Maria Teresa	Mujer	2000	Grupo Confe Soltera		No consta	Albacete	No consta	Diputada. Asistente en el Grupo Parlamentario Podemos-En Comú Podem-En Marea en la Legislatura XI.							
20	Pedro, Arroy	Hombre	1951	Grupo Confe Casado		Doctor en CI- Zaragoza		Educación Su	Diputado. Vocal de la Comisión de Agricultura, Pesca y Alimentación desde el 11/07/2018 al 05/03/2019							
21	José Luis, Ay	Hombre	1970	Grupo Popul Divorciado		Licenciado e- Madrid		Educación Su	Diputado. Director de Gabinete de la Presidencia del Gobierno (2011). Secretario de Estado de Relación							
22	Maria Luz, Be	Mujer	1964	Grupo Popul Soltera		Licenciada e- Madrid		Educación Su	Diputada. Gabinete de prensa del Partido Popular (1996-1996). Directora de Comunicación en los siguien							
23	Juan, Baldo	Hombre	1958	Grupo Mixto Casado		Maestro	Valencia	Educación Su	Diputado. Alcalde de Sueca (2007-2011). Diputado de Compromís-Q (2001-2015).							

Fuente: captura propia.

Se establecieron las categorías de: A) Nombre; B) Sexo; C) Año de nacimiento (que se corresponde con la variable edad dentro de la investigación); D) Grupo Parlamentario (variable que se utiliza para comparar con el resto, en función del partido político se representa el sexo, la edad, etc); E) Estado civil; F) Nivel académico; G) Origen; H) Nivel académico según los parámetros del INE; e I) Cargo que ocupa cada dirigente.

En cuanto a las dificultades, en esta parte se tuvo problemas con los datos de los diputados y por ello, en todas las variables contamos con la opción de “No consta” para toda aquella información sobre algunos representantes que no se encuentra, o no está muy clara dentro de la propia web.

Entre otros problemas, durante la XII Legislatura, objetivo de esta investigación, 43 de los diputados/as de diferentes formaciones políticas han sido sustituidos. Dentro del libro de Excel, en una hoja aparte, se encuentran también los 43 nuevos diputados que pasaron a formar parte de la cámara con las mismas categorías, y añadiendo el diputado/a al que sustituye. Dato a dato se han ido recopilando toda la información para cada clasificación sobre los diputados.

La investigación, ha contado con un extenso trabajo de campo y distintos documentos realizados propiamente con los resultados de cada variable adaptados a la forma en que se ha realizado el trabajo, además de la comparativa entre población y representantes o sacar los distintos porcentajes, que luego fueron comparados con los resultados en las tablas que proporcionó

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

el lenguaje; siendo esta, una manera de asegurar que los resultados obtenidos con R eran los mismos que obtenidos manualmente y aportar fiabilidad y rigor a la investigación, sin dejar todo en manos de la informática.

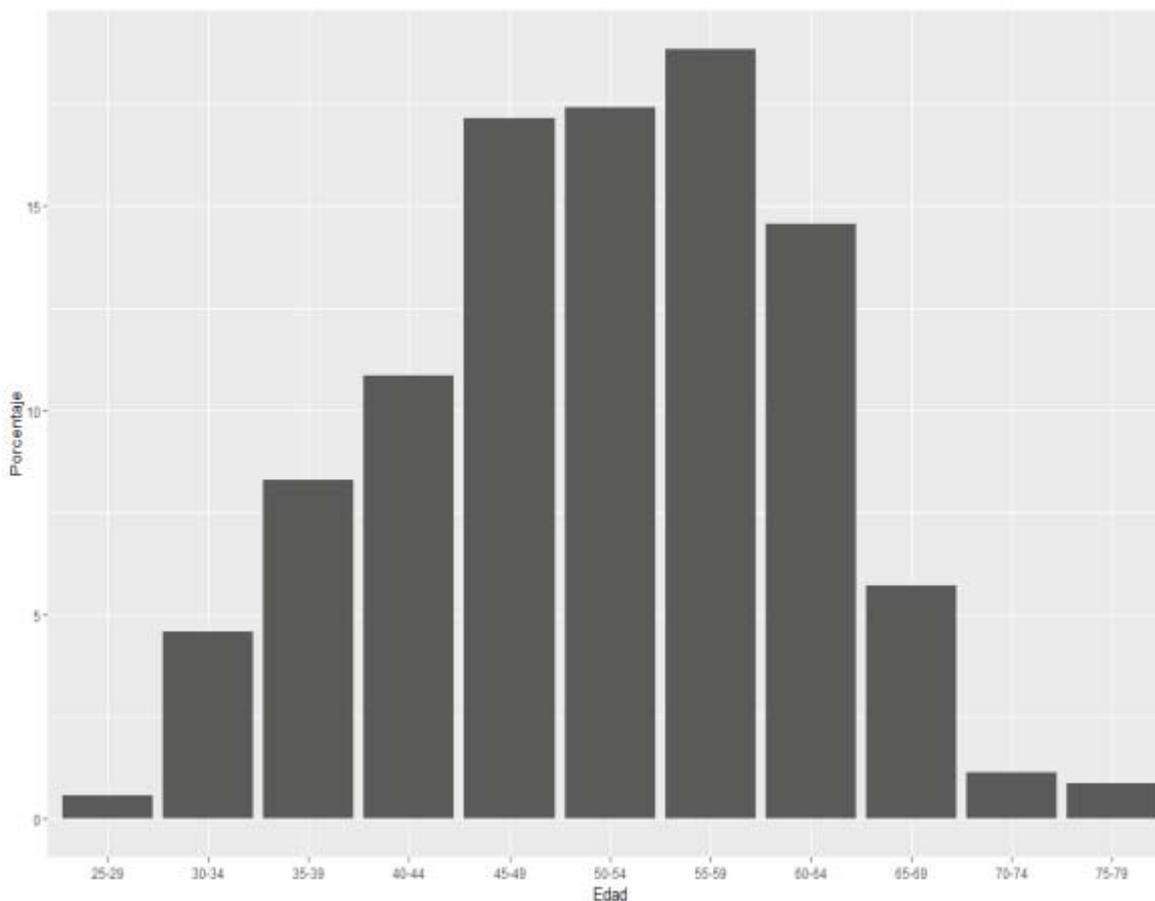
2.1.2 Extracción de datos del INE

El siguiente paso era realizar lo mismo, pero con los datos de la población. Para ello se recurrió al Instituto Nacional de Estadística (INE) y a las bases de datos con las que este organismo cuenta acerca de la población. Variable por variable, se fueron obteniendo las diferentes tablas que daban la información y los datos sobre los registros existentes de la población en el período de 2018.

Cada variable (sexo, edades) viene dada de una manera distinta: en miles de personas, en tasas o en porcentajes, entre otras. Además, la variable por ejemplo de la edad, se había cercado entre los 20 y los 84 años ya que esas eran las edades entre las que se comprenden los diputados que forman el Congreso y para que la investigación sea real hay que cercar también a la población entre esas edades.

El INE, dependiendo de la variable, ofrece la información desglosada de distintas formas, pero en cuanto a la edad de la población agrupa, desde los 0 años, las edades de 4 en 4 (0-4; 5-9; 20-24; 30-34; 35-39) y en otros casos va desarrollando por años de uno en uno, por lo que en esta investigación se ha utilizado la agrupación de cuatro en cuatro años.

Figura 2: Ejemplo de gráfico elaborado con lenguaje R en el que se representan los grupos de edad de los diputados y diputadas, agrupados, y el porcentaje que representan dentro del Congreso de los Diputados.



Fuente: captura de pantalla propia.

El análisis sobre la población ha llevado un trabajo costoso de recopilación de datos, sumas de cantidades, porcentajes y tasas, además de la adaptación en función a los resultados del Congreso de los Diputados, que se ha realizado manualmente primero y con cruce de datos con R, después.

En los anexos de esta investigación se adjuntas las tablas recopiladas del INE de dónde se han sacado los datos y el Excel completo de los Diputados.

Tras tener los datos desglosados, había que combinarlos y sacar información de ellos, realizando una comparativa entre todos ellos que se encuentra alojada en el blog en distintos artículos según la variable. La compilación, limpieza y cruce de datos no hubiese sido posible sin la inclusión del programador y el lenguaje R, que se desarrolla a continuación.

2.1.3 Análisis de datos: uso del lenguaje de programación R

El análisis de datos es la disciplina encargada de limpiar, ordenar, y transformar un conjunto de información ya recolectada para poder obtener una serie de conclusiones o tomar unas decisiones óptimas en campos como los negocios, la ingeniería o el caso que nos ataña: la sociedad.

En definitiva, no deja de ser una herramienta; una métrica intermedia para poder apoyar estadísticamente las preguntas suscitadas por nuestro foco de estudio: ¿Cómo está representada la sociedad española en el congreso de los diputados? Es aquí donde radica el gran poder y responsabilidad del periodismo de datos: democratizar los resultados estadísticos, hacerlos comprensibles al gran público a través de ensayos, infografías, graficación de datos o aplicaciones interactivas.

Los datos son el punto de partida. Como ya se ha comentado, el análisis de datos no es el objetivo mismo de este proyecto, más bien se trata de una herramienta. Por ello se realiza una explicación de los pasos a seguir hasta llegar al objetivo de crear un canal interactivo entre la información y el lector.

Antes de exponer los pasos a seguir propios de un análisis de datos hay que decidir con qué herramientas informáticas se realizará dicho análisis.

a) Elección del software para el análisis y visualización

Existen una gran cantidad de soluciones online para la graficación de tablas de datos y la obtención de estadísticas acerca de los mismos. Sin embargo, estas soluciones suelen estar muy limitadas y no ofrecen resultados que se ajusten a nuestras necesidades además de no aportar un enfoque profesional a la realidad que exige un verdadero análisis de datos.

Por ello la mejor opción es utilizar un lenguaje de programación creado única y exclusivamente para el análisis estadístico y la graficación. El lenguaje utilizado para el tratamiento de la información es el lenguaje R.

¿Por qué se ha elegido el lenguaje R?

Como ya se ha comentado, una de las principales razones de utilizar este lenguaje es que ha sido creado exclusivamente para el análisis de datos.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

R es de los lenguajes más empleados en bioestadística, matemática financiera o en investigaciones sociológicas, entre otros campos. Otra razón principal es su sencillez puesto que su sintaxis ha sido creada para utilizarse sin necesidad de poseer profundos conocimientos de programación.

Frente a otras opciones que puede ofrecer por ejemplo Excel, en R la organización de proyectos, la corrección de errores o la graficación de la información es mucho más sencilla y eficiente.

Un aspecto muy importante a la hora de elegir un lenguaje de programación es la comunidad de usuarios que este tenga detrás. Al tratarse de un lenguaje de código abierto la comunidad de R es muy extensa y ofrece una gran variedad de respuestas y funciones ya existentes para poder llegar al éxito de una forma más temprana y sustentada en sólidas soluciones ya probadas libres de errores.

Otra faceta que definitivamente a desequilibrado la balanza a favor de R son sus potentes herramientas para crear una aplicación web donde el usuario pueda interactuar con los gráficos generados, es decir, un canal donde el público pueda interactuar libremente con la información ya procesada. Más adelante se hará más énfasis en este tema.

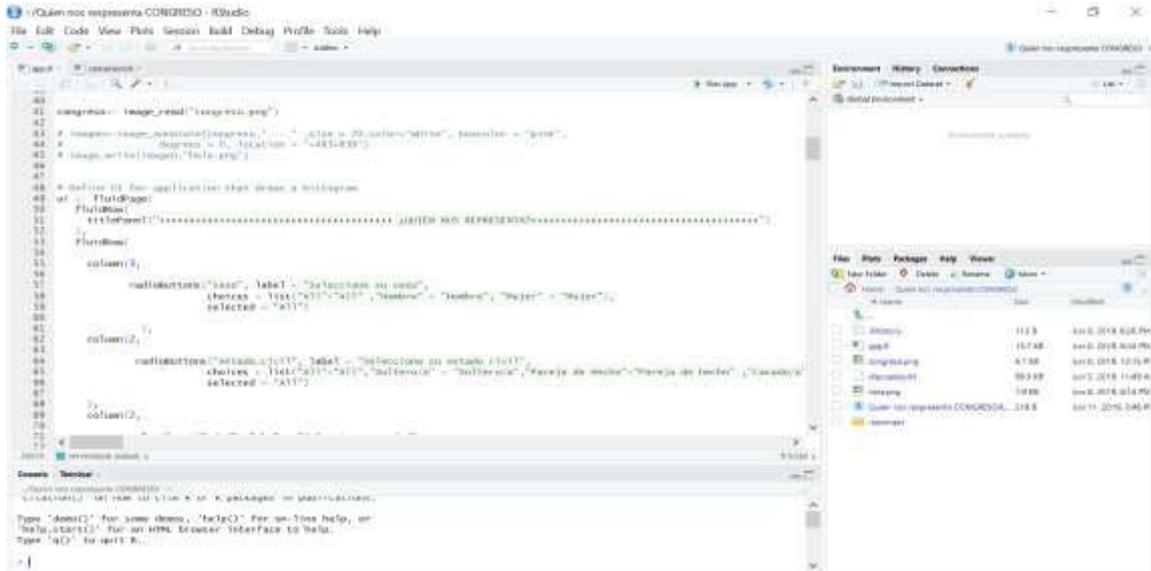
Todo lenguaje importante de programación debe de tener un potente entorno de desarrollo integrado o IDE (en inglés *Integrated Development Environment*). Cualquier código de un lenguaje de programación no deja de ser un texto plano que puede ser escrito en cualquier aplicación de procesador de texto como puede ser el bloc de notas.

Sin embargo, un IDE es una aplicación software donde es mucho más intuitivo escribir el código puesto que los errores son depurados, se incluyen tabulaciones en el texto automáticamente (lo que hace más legible su contenido), ciertas palabras especiales como las funciones aparecen escritas en colores diferentes, etc.

Para programar se ha utilizado el entorno *RStudio* que da un soporte oficial al lenguaje R. En la siguiente figura se puede apreciar el aspecto de dicho IDE.

Figura 3: Entorno de desarrollo integrado *RStudio* para el lenguaje R. Captura propia mediante el proceso dentro del programa.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.



Fuente: captura propia en *Rstudio*.

Existen otros entornos integrados de desarrollo para programar con R, sin embargo, se ha decidido utilizar *RStudio* porque aparte de ser el IDE oficial no requiere el pago de licencias para su utilización ya que es de código abierto.

b) Pasos en la elaboración de un análisis de datos

Figura 4: Etapas en el análisis de datos.



Fuente: elaboración propia, realizado con *Paint*.

En la figura 4 aparecen esquematizadas las distintas etapas que tienen lugar en nuestro análisis de datos. A continuación, se exponen con más detalle cómo han

sido implementadas con el lenguaje R hasta llegar al objetivo final de comunicar los resultados mediante la graficación interactiva de los mismos.

Sin embargo, puesto que esto no es un manual de programación las explicaciones no se centrarán en cada detalle de la sintaxis de R o aspectos básicos de programación: se centrarán en aspectos generales de cómo se leen las tablas, qué limpieza de datos se ha efectuado y qué librerías ya existentes se han utilizado para ello, así como qué librerías se han empleado para la visualización gráfica de los resultados.

b.1) Lectura de datos

Los datos ya se han recolectado en una tabla Excel. Aunque es posible leer desde R el archivo Excel se ha optado por guardar la base de datos como un archivo “.txt” para hacer el proceso más estándar.

Para la lectura se ha utilizado una de las funciones que ofrece R para tal efecto:

```
diputados<-read.csv2("diputados.txt", encoding = "latin1")
```

Lo que quiere decir esta sintaxis es que en el objeto “diputados” se ha guardado el marco de datos con la tabla de todos los diputados del congreso.

El argumento “encoding=latin1” hace referencia al formato de los caracteres guardados en el archivo “diputados.txt” ya que estos son caracteres con tildes o caracteres conformados por la letra “ñ” los cuales son propios del idioma español. Si no se realiza esta indicación pueden ocurrir problemas a la hora de visualizar los datos en una aplicación web o en las gráficas estáticas generadas.

b.2) Limpieza de datos

Una vez se obtuvo la información de la base de datos se procede a su limpieza. ¿Qué se quiere decir realmente con “limpieza de datos”?

Es muy probable que a la hora de recolectar y efectuar la base de datos se produjeran errores humanos en la redacción de algunas variables como el sexo o el estado civil de un diputado. Además, es posible que en la tabla recolectada aparezcan observaciones que no son de nuestro interés.

También a la hora de volcar la información en R se han producido errores de formato como: duplicaciones, malas interpretaciones por parte del programa de espacios en blanco, etc.

En definitiva, limpiar los datos ha consistido en homogeneizarlos en formato, eliminar duplicados o errores que introduzcan más errores en etapas futuras, así como en eliminar variables que no aportan información útil y crear aquellas otras nuevas variables que sí aporten utilidad para hacer más fácil y legible su procesamiento en la siguiente etapa de transformación.

Para realizar todas estas tareas se pueden emplear técnicas típicas de cualquier lenguaje de programación de propósito general, como bucles o sentencias condicionales que requieren conocimientos más profundos sobre dicha disciplina. Sin embargo, como se ha remarcado anteriormente, uno de los motivos por los que se ha empleado R es su facilidad a la hora de realizar estas tareas sin necesidad de tener amplios conocimientos sobre como programar.

Por tanto, para esta finalidad se emplearon librerías que contienen funciones que realizan automáticamente ciertas tareas de limpieza de datos. De esta forma el analista únicamente debe de indicarle a las funciones las variables con las que quiere trabajar sin necesidad de conocer el funcionamiento interno de tales funciones.

La librería en concreto que se ha utilizado es “Tydiverse” (Se puede traducir al español como “Universo Ordenado”). La base de datos inicial sin limpiar consta de 8 variables: NOMBRE; SEXO; AÑO.DE.NACIMIENTO; G.PARLAMENTARIO; ECIVIL; NIVEL.ACADEMICO; ORIGEN; CARGO.

En la variable “SEXO” debían aparecer las dos categorías “Hombre” y “Mujer”. La variable sin limpiar contiene más de dos categorías por errores de transcripción, como por ejemplo “Hombr” o “Mujr”. Si no se corrige este aspecto, a la hora de contabilizar cuantos hombres y cuantas mujeres hay en el congreso obtendremos resultados falseados, puesto que, si por ejemplo existen 100 hombres, se contarán 90 a partir de la categoría correcta “Hombre” y los otros 10, se asignarán a la categoría errónea “Hombr”. También tiene lugar un error

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

de formato puesto que R detecta espacios en blanco. Si la categoría esta seguida de un espacio en blanco también se contabilizará como una categoría diferente.

En la figura 3 aparece un ejemplo del código para la limpieza de “SEXO” utilizando la función “mutate()” de la librería “tidyverse” antes mencionada.

Figura 5: Ejemplo para la limpieza de “SEXO” durante el proceso, en el propio programa.

```
#####  
#####LIMPIEZA DE SEXO#####  
#####  
diputados<-mutate(diputados,SEXO =ifelse(SEXO=="Huer","Mujer",ifelse(SEXO=="Mujer","Mujer",ifelse(SEXO=="Mujer ","Mujer",  
ifelse(SEXO==" Hombre","Hombre",ifelse(SEXO=="Hombre","Hombre",ifelse(SEXO=="Hombre ","Hombre","No consta"))))))  
diputados$SEXO<-factor(diputados$SEXO)
```

Fuente: captura propia en *Rstudio*.

Para el resto de las variables se procede de una forma análoga. Mención especial para el caso de la variable “G.PARLAMENTARIO”. Además de corregir los errores expuestos también se decidió sustituir el nombre completo de las formaciones políticas por sus siglas o nombres más representativos para una legibilidad mayor en la siguiente etapa de análisis.

Por ejemplo, el nombre "Grupo Confederal de Unidos Podemos-En Comú Podem-En Marea (GCUP-EC-EM)" es demasiado extenso para trabajar de una forma más eficiente e intuitiva a la hora de calcular tablas porcentuales o representaciones gráficas. Por ello se sustituyó por el nombre “Podemos”.

Otro ejemplo, pero empleando siglas, es el caso de “Grupo Popular”. Aunque el nombre no es especialmente extenso es más fácil e intuitivo trabajar con el nombre “PP”.

Para finalizar la limpieza se creó una nueva variable llamada “RANGO.EDAD” puesto que en etapas siguientes es necesario evaluar la pirámide poblacional del congreso según ciertos rangos de edad coincidentes con los rangos de edad que utiliza el INE para la población total española.

Esta variable se creó a partir de la ya existente “AÑO.DE.NACIMIENTO”. Según el año de nacimiento a cada diputado se le asigna un rango de edad. Para realizar esta tarea además de utilizar las funciones de la librería “tidyverse” ha

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

hecho falta emplear otros métodos de programación más avanzados como son los bucles y las sentencias condicionales.

En la figura 4 aparece una porción del código utilizado para la confección de esta nueva variable.

Figura 6: Creación de la variable “RANGO.EDAD” durante el proceso, en el propio programa.

```
##### Limpieza de Edad #####
#####

for(i in 1:350) {
  if(diputados$AÑO.DE.NACIMIENTO[i]<=1999&&diputados$AÑO.DE.NACIMIENTO[i]>=1995){diputados$RANGO.EDAD[i]="20-24"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1994&&diputados$AÑO.DE.NACIMIENTO[i]>=1990){diputados$RANGO.EDAD[i]="25-29"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1989&&diputados$AÑO.DE.NACIMIENTO[i]>=1985){diputados$RANGO.EDAD[i]="30-34"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1984&&diputados$AÑO.DE.NACIMIENTO[i]>=1980){diputados$RANGO.EDAD[i]="35-39"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1979&&diputados$AÑO.DE.NACIMIENTO[i]>=1975){diputados$RANGO.EDAD[i]="40-44"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1974&&diputados$AÑO.DE.NACIMIENTO[i]>=1970){diputados$RANGO.EDAD[i]="45-49"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1969&&diputados$AÑO.DE.NACIMIENTO[i]>=1965){diputados$RANGO.EDAD[i]="50-54"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1964&&diputados$AÑO.DE.NACIMIENTO[i]>=1960){diputados$RANGO.EDAD[i]="55-59"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1959&&diputados$AÑO.DE.NACIMIENTO[i]>=1955){diputados$RANGO.EDAD[i]="60-64"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1954&&diputados$AÑO.DE.NACIMIENTO[i]>=1950){diputados$RANGO.EDAD[i]="65-69"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1949&&diputados$AÑO.DE.NACIMIENTO[i]>=1945){diputados$RANGO.EDAD[i]="70-74"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1944&&diputados$AÑO.DE.NACIMIENTO[i]>=1940){diputados$RANGO.EDAD[i]="75-79"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1939&&diputados$AÑO.DE.NACIMIENTO[i]>=1935){diputados$RANGO.EDAD[i]="80-84"}
  else{if(diputados$AÑO.DE.NACIMIENTO[i]<=1934&&diputados$AÑO.DE.NACIMIENTO[i]>=1930){diputados$RANGO.EDAD[i]="85-89"}
  else{diputados$RANGO.EDAD[i]="No"}
  }
  }
}
```

Fuente: captura propia en *Rstudio*.

Una vez se ha limpiado toda la tabla se puede volver a guardar como un archivo Excel o en este caso como un archivo “.txt”. Para ello se ha utilizado la función inversa a la de lectura:

```
write.csv2(diputados,file="diputados_final.txt",row.names = FALSE)
```

b.3) Transformación (Análisis)

En esta etapa, los datos fueron sometidos a métodos y cálculos para obtener estructuras de información más organizada que aporten respuestas a las cuestiones planteadas por la investigación que concierne.

Se hará llegar toda esta información útil al usuario final a través de gráficas y tablas probabilísticas. En la anterior “Figura 4” las etapas de transformación y visualización se han representado con una retroalimentación porque a medida que se representan resultados estadísticos surgen nuevas ideas acerca de qué tipos de cálculos se pueden efectuar sobre los datos y viceversa.

A partir de la información obtenida de esta etapa es cuando el investigador puede exponer unas conclusiones en una discusión, artículo, infografía, etc. acerca de lo que significan tales resultados.

Para la elaboración de las tablas probabilísticas utilizaremos funciones de la librería “gmodels”. De esta forma, la función “table()” remitirá automáticamente la contabilidad de las categorías de cada variable, sin necesidad de que el analista realice a mano uno a uno los cálculos.

Por ejemplo, para conocer el número de hombres y mujeres totales del congreso basta con ejecutar la siguiente línea de código:

```
table(diputados$SEXO)
```

Mediante dicha sentencia se puede conocer de forma aislada el número de personas divorciadas, el número de personas procedentes de la provincia de Asturias, etc. Sin embargo, aunque estos datos son útiles realmente de por si no nos aportan demasiada información.

La mayor distinción del conjunto de diputados que conforman el congreso es su división por partidos políticos. Por tanto, una información que es verdaderamente útil será aquella tabla en la que aparezcan cruzadas las distintas variables categóricas (“SEXO”, “ECIVIL”, “ORIGEN”, etc.) con la variable “G.PARLAMENTARIO”:

```
table(diputados$SEXO, diputados$G.PARLAMENTARIO)
```

```
table(diputados$ECIVIL, diputados$G.PARLAMENTARIO)
```

```
table(diputados$ORIGEN, diputados$G.PARLAMENTARIO)
```

De esta forma se puede conocer cuántas personas del total del congreso son mujeres, cuantas personas de, por ejemplo, Ciudadanos son mujeres o cuantas mujeres del total del congreso son de Ciudadanos. Este ejemplo es aplicable al resto de las variables que hemos cruzado con el grupo parlamentario.

Sin embargo, los valores absolutos de números concretos de diputados pueden quedarse a medias a la hora de valorar su repercusión respecto al número total de diputados existentes. Por ello, se confeccionó la tabla definitiva que aporta

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

una información útil más completa donde se incluyen valores absolutos y porcentuales de cada variable respecto a cada grupo parlamentario.

En la figura 7 se aprecia una parte de la tabla definitiva para el caso de “SEXO” por cada partido político. Para generarla se ejecutó el siguiente código con la función “CrossTable()”:

```
CrossTable(diputados$SEXO,diputados$G.PARLAMENTARIO)
```

Figura 7: muestra de la tabla de cada grupo parlamentario según el sexo.

diputados\$SEXO	diputados\$G.PARLAMENTARIO		
	Bildu	C-P-EUPV	Ciudadanos
Hombre	1	3	25
	0.005	0.014	0.118
	0.500	0.750	0.781
	0.003	0.009	0.071
Mujer	1	1	7
	0.007	0.007	0.051
	0.500	0.250	0.219
	0.003	0.003	0.020
Column Total	2	4	32
	0.006	0.011	0.091

Fuente: Captura propia en el txt completo.

Los resultados porcentuales de la tabla están expresados en tanto por 1 en lugar de en tanto por 100. Los valores enteros de la tabla hacen referencia al número de diputados de cada formación política. Cada fila de la tabla para “Hombre” y para “Mujer” se lee de la siguiente manera:

- Fila 1: Número de diputados del partido político. Ejemplo: en la sección Hombres, existe 1 diputado en el partido Bildu.
- Fila 2: Porcentaje de Hombres/Mujeres del total del congreso que pertenecen al partido político. Ejemplo: el 0.7 % (0.007) de mujeres del congreso están en Bildu.
- Fila 3: Porcentaje del partido político que son Hombres/Mujeres. Ejemplo: el 50% (0.500) de Bildu son Hombres.
- Fila 4: Igual que en la fila 1, número de diputados expresados porcentualmente respecto del total de diputados en el congreso. Ejemplo: en Bildu hay 1 Mujer

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

(Fila 1). Esa única mujer de Bildu supone el 3% (0.003) del total de diputados existentes en el congreso (350 diputados).

De cara al público general, es mucho más intuitivo presentar estos resultados gráficamente, por ejemplo, mediante gráficas de barras.

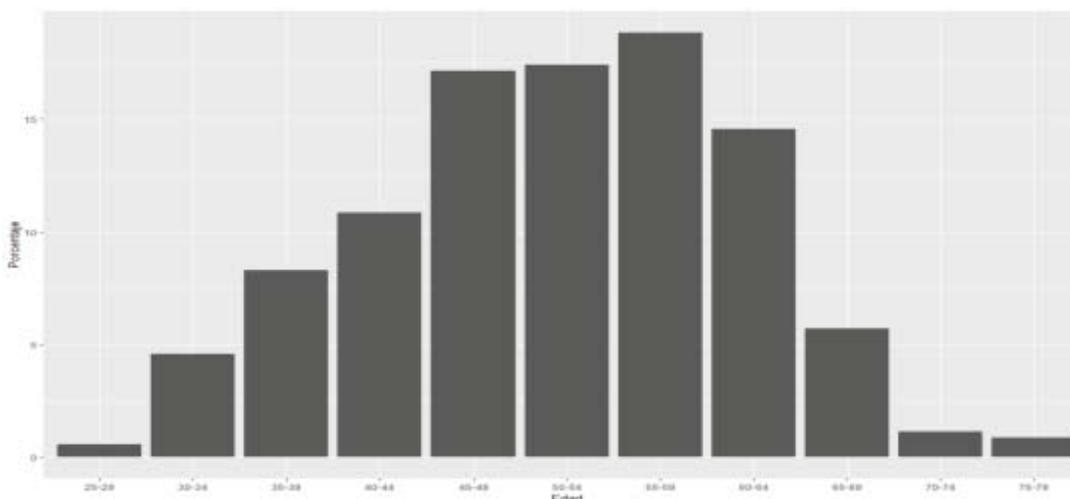
El lenguaje R es de los más utilizados para estas representaciones por sus potentes librerías gráficas y sobre todo su facilidad de uso intuitivo. La librería utilizada en cuestión se llama “ggplot2”. De forma análoga al caso de las tablas se pueden representar las variables de forma individual.

Para el ejemplo de la variable “RANGO.EDAD” se puede ejecutar el siguiente código:

```
ggplot(data=diputados, aes(x=RANGO.EDAD,y=100*..count../350 )) +  
  geom_bar(stat="count") +ylab("Porcentaje")+xlab("Edad")
```

El resultado puede apreciarse en la figura 8.

Figura 8: Gráfica de barras representando el porcentaje de cada rango de edad.

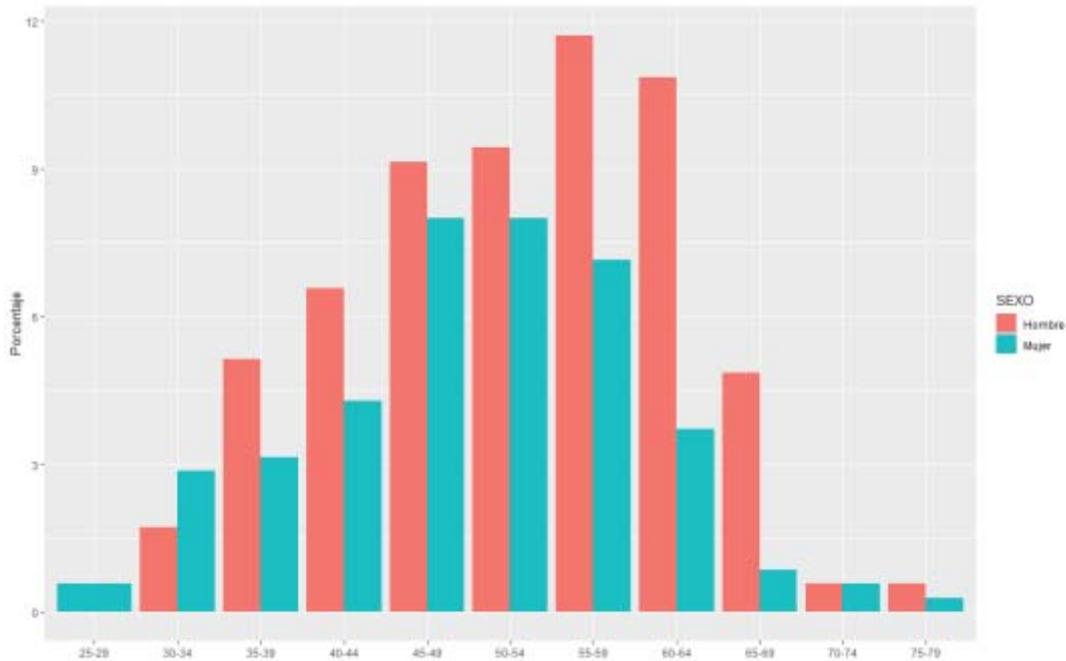


Fuente: captura propia en parte interactiva.

También se puede realizar el cruce de variables de distintas formas. En la figura 9 aparece el mismo gráfico para la edad, pero incluyendo la distinción del sexo para cada rango.

Figura 9: Porcentaje de hombres y mujeres para cada rango de edad.

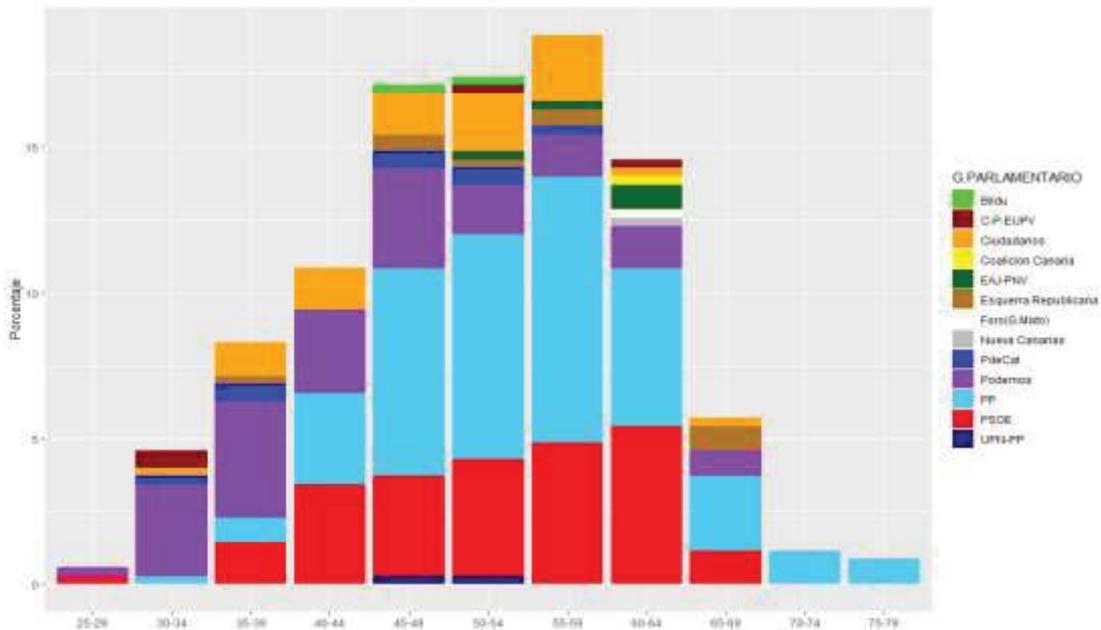
Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.



Fuente: captura propia en parte interactiva.

Asimismo, se pudo realizar el cruce de las variables con “G.PARLAMENTARIO”. Para el ejemplo del rango de edad el resultado puede observarse en la figura 10.

Figura 10: Porcentaje de diputados de cada partido político por cada rango de edad. Captura propia.



Fuente: captura propia en parte interactiva.

b.4) Comunicación interactiva

Una buena forma de conseguir una mayor inmersión entre el público y la información generada tras el análisis, es la creación de un canal donde el lector

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

El usuario puede seleccionar diferentes casillas de modo que la gráfica represente la información que se ha elegido. Además, acercando el cursor a las diferentes barras se muestra información más detallada como el porcentaje de diputados. También se puede efectuar un zoom o pinchar sobre las diferentes categorías de las variables para que en la gráfica solo aparezcan las barras o colores que más nos interesen.

El código elaborado en el IDE *RStudio* para construir la aplicación consta de dos secciones:

- Una primera sección que hace referencia a la interfaz de usuario. En esta sección aparece el código referido a todo aquello que se verá en el navegador web a modo de fachada; como los *widgets* que recogen las diferentes opciones que debe escoger el usuario. El código necesario para generar el *widget* ya lo proporciona *Shiny* en una galería donde aparecen diferentes ejemplos. El analista simplemente debe de adecuarlo a las opciones que desea mostrar. En la figura 12 se muestra una parte de código referente a esta sección.

Figura 12: Parte del código de la sección Interfaz de Usuario en el programa durante la elaboración.

```
#####  
# Define la interfaz de usuario para gestionar los datos de entrada y los datos de salida #  
#####  
ui <- fluidPage(  
  fluidRow(  
    column(10,  
      titlePanel("1. REPRESENTACIÓN POR SEXO")  
    )  
  ),  
  fluidRow(  
    column(9,  
      wellPanel(  
        checkboxGroupInput("partido", "Seleccione uno o varios Partidos Políticos",  
          c("Partido Popular"="PP", "Partido Socialista"="PSOE", "Unidos Podemos"="Podemos", "Ciudadanos"="Ciudadanos", "E.Republ  
        )  
      ),  
      fluidRow(  
        textOutput("etiqueta_1")  
        #verbatimTextOutput("etiqueta_1")  
      ),  
      fluidRow(  
        column(12,  
          plotlyOutput("sexo.congreso")  
        )  
      )  
    )  
  )  
)
```

Fuente: captura propia en *Rstudio*.

- Una segunda sección donde se procesa todo el código que da lugar a las gráficas interactivas. A esta sección se le denomina “*server*” puesto que su código es el

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

que continuamente es procesado en el servidor web a medida que el usuario cambia las opciones de los widgets.

En esta sección básicamente se incluye el código ya empleado en las etapas anteriores del análisis donde se generaron las tablas probabilísticas y las gráficas de las variables y sus categorías. Simplemente se incluyen unas pequeñas modificaciones en el código para convertir dichas gráficas en interactivas. Si la primera sección actúa como una fachada, esta segunda sección actúa como el interior de la casa donde realmente se confeccionan y procesan los datos.

Es en esta sección donde entran en juego las distintas funciones ya mencionadas de la librería “ggplot2”. En la figura 13 aparece una parte del código de esta sección.

Figura 13: parte del código perteneciente a la sección “serve” en el programa durante la elaboración.

```
#####  
####Gestion de Sexo#####  
#####  
  
#output$sexo.etiqueta<-renderText({  
# c(paste(paste(round(100*test$prop.co),1,ifelse(input$partido[length(input$partido)]=="PP",1,ifelse(input$partido[length(input$partido)]=="PSOE",  
# ))  
  
output$sexo.congreso <- renderPlotly({  
  ggplot(data=diputados, aes(x=G.PARLAMENTARIO,y=..count..,fill=SEXO,label=paste(100*..count../350,"%")))+  
  geom_bar(position="dodge")>ylab("Nº Diputados")>xlab("Partido Político")>scale_x_discrete(limit = input$partido)  
})  
output$event <- renderPrint({  
  d <- event_data("plotly_hover")  
  if (is.null(d)) "Hover on a point!" else d  
})  
  
output$sexo.general<-renderPlotly({  
  ggplot(data=df, aes(x=Lugar, y=Porcentaje, fill=Sexo)) +  
  geom_bar(stat="identity", position="dodge")>scale_x_discrete(limit = input$lugar)  
})  
  
#####
```

Fuente: captura propia en *Rstudio*.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Aplicación web con mapa de diputados coincidentes interactiva:

En la figura 14 se aprecia el aspecto de esta aplicación en un navegador web.

Figura 14: Aplicación web con mapa de diputados coincidentes interactiva. ²⁴



Fuente: Captura propia en parte interactiva.

En esta aplicación el usuario selecciona las opciones que representen sus preferencias de forma que los diputados que coincidan con tales valores aparecen en una tabla interactiva que muestra la información relativa a cada uno de ellos.

El código necesario para construir esta aplicación sigue las mismas premisas que la otra aplicación interactiva salvando las diferencias de que en lugar de gráficas aparece una tabla.

Para poder utilizar los servidores gratuitos de *RStudio* donde se ejecutará el código previamente se debe de crear una cuenta en el sitio shinyapps.io

Desde el entorno de desarrollo integrado *RStudio* utilizado durante todo el proceso se brinda la opción mediante un simple botón de cargar el código automáticamente en el servidor de modo que todo el proceso de subida se simplifica considerablemente.

²⁴ Se puede encontrar en: <https://quienosrepresenta.shinyapps.io/congreso/>

2.1.4 Redacción del contenido web y del presente trabajo académico

Tras las distintas etapas que conciernen al análisis de datos con R, se sincronizó la realización de la parte anteriormente explicada con la redacción de cada artículo para la web, además de la incansable búsqueda de bibliografía explícita y referencias de autores para la también simultánea realización de la memoria que aquí se contempla.

Las últimas partes realizadas han sido las conclusiones, metodología, bibliografía y anexos de la memoria ya que como se indica anteriormente el desarrollo del marco teórico y las distintas partes de la memoria académica, se han ido desarrollando alternando con el resto del trabajo que se ha realizado.

Tras analizar las distintas variables se decidió realizar una web con una plantilla de blog para lo que se optó por la opción de Wix²⁵ ya que era una plataforma con la que se había trabajado anteriormente en el grado y da múltiples opciones en su versión gratuita. La decisión fue dividir dicho blog en distintos apartados en el siguiente orden:

Introducción: dentro de esta parte se contempla una pequeña introducción al tema que se va a tratar en la investigación, con algunas curiosidades e informaciones en relación con el tema. La situación política en la que se enmarca la investigación, entre otras cosas.

Reportaje: es la parte más importante ya que contiene los distintos artículos que se han ido elaborando a través de las variables. Al inicio, contemplamos dos enlaces externos, donde se alejan dos partes interactivas para la participación y obtención de información sobre los resultados obtenidos en la investigación: gráficos interactivos por cada variable, comparados los resultados del Congreso con los de la población y con una relación entre las distintas variables y de los que se pueden obtener resultados de todos los partidos políticos o que solamente se muestren los que a la persona le interese conocer. También incluye un mapa interactivo con los escaños del Congreso, por colores según el partido

²⁵ <https://es.wix.com/>. Última consulta: 16-06-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

político, y dónde se encuentra toda la información de cada diputado que se recogió anteriormente en el libro de Excel.

Tras ello, se encuentran los distintos artículos periodísticos en los que se habla de los resultados y comparativas obtenidas de la investigación con gráficos explicativos y referencias de otras investigaciones.

En esta parte, también se cuenta con las declaraciones de distintos ciudadanos, que dan su visión para los distintos aspectos que se tratan en los distintos artículos: la edad, la nacionalidad o la orientación sexual, por ejemplo.

Objetivos: se muestran los objetivos planteados al inicio de la investigación.

Metodología: se incluye una explicación de todas las partes del proceso llevado a cabo.

Autoría del proyecto: se puede observar una pequeña descripción tanto de la alumna que realiza el trabajo, como del estudiante que ha prestado sus conocimientos con el lenguaje de programación, un elemento imprescindible sin el que no hubiese sido posible realizar esta investigación.

Contacto: se puede encontrar las referencias para contactar con la autora.

Tras distintas correcciones por parte de las tutoras de este TFG, se fue completando el texto cuyo contenido va alojado en el blog, con las correspondientes revisiones de literatura, estructura del reportaje o referencias, además del contenido académico que aquí se incluye y poco a poco se fue desarrollando y perfeccionando gracias a sus aportaciones.

2.1.5 Diseño de la web

Tras tener toda la parte académica, la parte de programación con R y el contenido de la web se procedió al diseño de la web en formato blog.

Tras haber elegido Wix, se elige una plantilla sencilla, con funciones periodísticas y que dé cabida a distintos apartados para alojar las diferentes partes. Las opciones ya mencionadas de interactividad o hipertextualidad que nos ofrece el ciberespacio son uno de los puntos fuertes.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

La elección de los colores blancos y negros, alternando con algún otro, aunque en pocas ocasiones, se hace para aportar fluidez y claridad en la lectura además de seriedad y limpieza en la organización del trabajo.

Dentro del texto se han incluido distintos y numerosos hipervínculos que llevan a noticias, entrevistas o reportajes de autores citados; a perfiles en redes sociales o a estudios académicos, entre otras cosas.

También se cuenta con la parte interactiva en la que cualquier persona que visite la página puede acceder y observar los resultados del tema, aspecto, partido político, o variable que le interese en función al Congreso, o a la población. E incluso la comparación entre ambas, por variables.

La multitud de opciones que ofrece Internet y el poder alojar un reportaje extenso de datos fueron uno de los motivos de, al ver la cabida que podía tener la investigación, elegir realizar esta web en formato blog donde contemplar dos perfiles que van de la mano en el futuro del periodismo: el del periodista de datos en Internet.

El diseño estético de la web ha venido dado sobre la marcha en la realización, con criterios propios de además de autora, también lectora y visitante de webs. Se ha realizado bajo gustos propios y pensando siempre en proporcionar fluidez y facilidad de lectura, además de claridad tanto en los datos, como en el resto del contenido que se aloja.

Dicha web se puede encontrar y visitar en:
<https://sarabecerro.wixsite.com/quien-nos-represent>

3. Conclusiones

Dentro de las conclusiones obtenidas con este trabajo podemos contar con una primordial: la población española cuenta con muchas infrarrepresentaciones y fallos dentro de su sistema de representatividad parlamentaria en el Congreso de los Diputados y por ello, en base a este estudio, se puede afirmar que la apertura del ejercicio político a mayor parte de la población podría mejorar la representatividad actual.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Existen muchos grupos como es el caso de los extranjeros residentes, del colectivo LGTBI y de algunos grupos de edad que no están, en referencia a los datos y números obtenidos, representados en el mayor órgano de representatividad en España, el Congreso de los Diputados.

¿No sería más justo un sistema con una representación por parte de personas con las que la población se identifique en función a distintas características?

Dichas premisas se encuentran expuestas a lo largo del apartado “Reportaje” alojado en el blog con las conclusiones correspondientes a cada variable de representación.

La figura del periodista, en este caso el periodista de datos, es realmente primordial. Sin la existencia de analistas, que interpreten datos para proporcionar informaciones objetivas, veraces y claras a la población, no existe realmente una democracia realmente representativa de su población.

Por la experiencia general obtenida en este trabajo en su realización, se puede concluir que el desarrollo del periodismo de datos sería mucho más sencillo para los profesionales de la información si no se hubiera extraído manualmente las características de cada diputado.

En un ejercicio de transparencia por parte del gobierno, podría existir un documento que agrupara todas y cada una de las variables y que fuese descargable y accesible públicamente. Y que, además, esos documentos que faciliten puedan cruzarse fácilmente con otros, para que la dificultad a la hora de comparar distintos parámetros sea mucho menor.

Es decir, como conclusión, trabajar con R o con lenguajes de programación es realmente necesario para el periodismo y necesita de una formación especializada en informática y estadística, entre otras materias. Formación, con la que actualmente no se cuenta en los centros educativos y que es necesaria, ya que al realizar este tipo de trabajos hay que contar con la figura de un programador. El periodismo de datos es, probablemente, uno de los futuros más viables para los periodistas.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

En cuanto a Internet, se puede concluir que la interacción y la multimedialidad con la que cuenta, aporta una nueva dimensión narrativa a los relatos periodísticos. Además, ofrece numerosas opciones para la transparencia por parte de los mandatarios y del aporte de información, que, con la realización de esta investigación, se puede decir que tanto el INE como el Congreso de los Diputados deberían ofrecer esa información más claramente, ordenada y desglosada para su fácil accesibilidad y descarga.

La investigación realizada para el Trabajo de Fin de Grado también, ha aportado otra visión hacia la política y los problemas sociales, además de obtener distintas visiones de diversos autores con bases académicas y de investigación respecto a los temas que se tratan. Las infrarrepresentaciones de la sociedad española dentro del Congreso de los Diputados y los cambios sociales y políticos, son un hecho.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

4. Bibliografía

- Arias.J (2014). ¿Por qué la política está perdiendo a los jóvenes? *El País*. Recuperado de: https://elpais.com/internacional/2014/02/25/actualidad/1393358639_000137.html. Última consulta: 15-06-2019.
- Abenza.L, Barbet.B, Clavería.S, Costas.E, Galindo.J, Llaneras.K, Medina.O, Ramos.M, & Simón.P, (2017): *“El muro invisible. Las dificultades de ser joven en España”*. Barcelona: Debate.
- Alto Comisionado de las Naciones Unidas para los Refugiados. ACNUR. (2014). *La protección internacional*. Directrices sobre protección internacional. (9). Recuperado de: <https://www.acnur.org/fileadmin/Documentos/Publicaciones/2014/9872.pdf>. Última consulta: 12-06-2019.
- Bilbao-Arrese, J.M., (1994). Ley electoral y sistema de partidos en España. *Revista de Estudios Políticos (Nueva Época)*,(85), pp 313-321. Recuperado de: <https://recyt.fecyt.es/index.php/RevEsPol/article/view/46886/28367>. Última consulta: 10-06-2019.
- Campos-Domínguez.E. (2004-2008). *La ciberdemocracia en el Congreso de los Diputados*. (Tesis de doctorado). Congreso de los Diputados, Madrid.
- Clemente-Yanes.E. (2017, 14 de diciembre). Los conceptos de representación en el 15-M. *Politikon*. Recuperado de: <https://politikon.es/2017/12/14/los-conceptos-de-representacion-en-el-15-m/>. Última consulta: 9-05-2019.
- Coller,X, Jaime, A.M & Santana, A. (2012, 28 de julio). La educación de nuestros políticos. *El País*. Recuperado de: https://elpais.com/elpais/2012/07/25/opinion/1343235375_328774.html. Última consulta: 7-04-2019.
- Cómo el 15-M te cambió la vida (aunque no te lo parezca). (2017). *Público*. Recuperado de: <https://www.publico.es/sociedad/15-m-cambio-vida.html>. Última consulta: 26-03-2019.
- Crucianelli, S. (2013). ¿Qué es el periodismo de datos? Cuaderno de periodistas, (26), pp.106-124. Recuperado de: <http://www.cuadernosdeperiodistas.com/media/2013/12/106-1241.pdf>. Última consulta: 16-06-2019.
- Dader. J.L (1994). Periodismo de precisión en España. Una panorámica de casos prácticos. Recuperado de: http://www.quadernsdigitals.net/datos/hemeroteca/r_32/nr_452/a_6195/6195.pdf. Última consulta: 15-05-2019.
- Dader. J.L., & Gómez-Fernández.P.(1993). *Periodismo de precisión: una nueva metodología para transformar el periodismo*. Anhlisi (15), pp. 99-116.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Dader, J.L. (2000). La “mujer” como categoría emergente en la producción de noticias: pros y contras de una nueva visibilidad periodística. *Texto presentado como Ponencia en la Jornada sobre “Mujeres, Hombres y Medios de Comunicación” de la Dir. Gral. De la Mujer en Valladolid*. Junta Castilla y León. Recuperado de: <https://docplayer.es/58008031-La-mujer-como-categoria-emergente-en-la-produccion-de-noticias-pros-y-contras-de-una-nueva-visibilidad-periodistica.html>. Última consulta: 19-05-2019.

Galería de Shiny: <https://shiny.rstudio.com/gallery/>.

Las Provincias, (2019). ¿Pueden votar los extranjeros en estas elecciones 2019? Recuperado de: <https://www.lasprovincias.es/elecciones/pueden-votar-extranjeros-elecciones-2019-20190401173056-nt.html>. Última consulta: 30-05-2019.

León.S. (2018, 4 de abril). Políticos impotentes. *El País*. Recuperado de: https://elpais.com/elpais/2018/04/03/opinion/1522763915_939400.html. Última consulta: 12-05-2019.

López-López, P.J. (2016). Antecedentes y fundamentos del Periodismo de datos. *Periodismo de datos. Nuevas narrativas para el Periodismo Especializado*. Recuperado de: <https://riuma.uma.es/xmlui/bitstream/handle/10630/12465/Libro%20de%20actas%20RIUMA%20Cap1.pdf?sequence=3&isAllowed=y>. Última consulta: 10-06-2019.

López-López. P.J (2018). *La enseñanza-aprendizaje del periodismo de datos en las facultades españolas. Estudios y propuestas de viabilidad*. (Tesis doctoral). Universidad Complutense de Madrid. Recuperado de: <https://eprints.ucm.es/50688/1/T40749.pdf>. Última consulta: 10-06-2019.

López-López.P.J., & Flores Vivar, J. M. (2015). *Big Data, Transparencia y Periodismo de Datos*. Cursos de verano Universidad Complutense de Madrid. Recuperado de: <https://www.ucm.es/data/cont/docs/71-2015-04-15-71311.pdf>. Última consulta: 3-04-2019.

Navarro, J. (2010). *Dime quien soy*. Barcelona: Plaza & Janes.

Ortiz, R. (2015). R Pubs - GGPlot2. Gráfico de barras IV - Reordenar barras. Recuperado de: <https://rpubs.com/Rortizdu/140196>. Última consulta: 13-06-2019.

Paradis, E. (2003). R para principiantes. Institut des Sciences de l'Evolution. Universidad de Montpellier II. Recuperado de: https://cran.r-project.org/doc/contrib/rdebuts_es.pdf Última consulta: 13-06-2019.

Pérez-Blanco, P.M (2016, 20 de abril). El periodismo en la era digital. *El País*. Recuperado de: https://elpais.com/elpais/2016/04/13/opinion/1460540302_620130.html. Última consulta: 12-04-2019.

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

Pérez-Martínez.F. (2014, 19 de diciembre). La democracia representativa en España. *Infolibre*. Recuperado de: https://www.infolibre.es/noticias/club_info_libre/librepensadores/2014/12/18/la_democracia_representativa_espana_25565_1043.html. Última consulta: 12-05-2019.

Ruiz-Alaejos.V (2013). Datos, el periodismo de la evidencia. *InfoLibre*. Recuperado de: https://www.infolibre.es/noticias/medios/2013/04/17/datos_periodismo_e_videncia_2621_1027.html. Última consulta: 9-05-2019.

Ruiz, E. (2018). Usando R para la Ciencia de Datos. Recuperado de: <https://resources.rstudio.com/espanol/2018-05-23-13-01-usando-r-para-la-ciencia-de-datos-edgar-ruiz-edited>. Última consulta: 9-06-2019.

Salazar, C. (2018). RPubs - Análisis exploratorio. Recuperado de: https://rpubs.com/camilamila/analisis_exp. Última consulta: 9-06-2019.

Sánchez-García, P., Campos-Domínguez, E., & Berrocal Gonzalo, S. (2015). Las funciones inalterables del periodista ante los perfiles multimedia emergentes. Recuperado de: <https://doi.org/10.4185/RLCS-2015-1042>. Última consulta: 10-04-2019.

Sánchez-Hidalgo, E. (2018). ¿Casado es tan joven? La edad de los políticos españoles al liderar sus partidos. *El País*. Recuperado de: https://verne.elpais.com/verne/2018/07/21/articulo/1532175653_351295.html. Última consulta: 9-05-2019.

Sérvulo-González, J., (2019). ¿Cuántos votos le ha costado un escaño a cada partido? | *El País*. Recuperado de: https://elpais.com/politica/2019/04/29/actualidad/1556505292_395748.html. Última consulta: 7-06-2019.

Soriano, R. & Alarcón, C., (2001). Las elecciones en España: ¿votos iguales y libres? *Revista de Estudios Políticos (Nueva Época)*, (114), pp 115-130. Recuperado de: <https://docplayer.es/52020904-Las-elecciones-en-espana-votos-iguales-y-libres.html>. Última consulta: 27-05-2019.

Trejo, M. (2019). Rescatando la política con el poder transformador de los jóvenes – ACOP. Recuperado de: <https://compolitica.com/rescatando-la-politica-con-el-poder-transformador-de-los-jovenes/>. Última consulta: 11-06-2019

Uriarte.E, & Ruiz.C, (1999). “Mujeres y hombres en las élites políticas españolas: ¿diferencias o similitudes?” *Reis: Revista Española de Investigaciones Sociológicas*, (88), pp.207-232. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=759376>. Última consulta: 20-05-2019.

Villar-Velasco, A., (2017). ¿Qué es y en que consiste la Ley D'Hondt? | *Wall Street International Magazine*. Recuperado de

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

<https://wsimag.com/es/economia-y-politica/18650-que-es-y-en-que-consiste-la-ley-dhondt>. Última consulta: 30-04-2019.

Wikipedia. (2018). *Programación con datos masivos en R*. Recuperado de: https://es.wikipedia.org/wiki/Programaci%C3%B3n_con_datos_masivos_en_R. Última consulta: 15-04-2019.

5. Anexos

Anexo 1: Información recopilada del INE para el análisis de las cifras de población. Se descargaron a través de la web del Instituto, los libros de Excel con la información correspondiente a las distintas variables elegidas. Se pueden encontrar alojadas en la nube, junto con el Excel con toda la información recopilada sobre los diputados y diputadas que forman el Congreso durante la XII Legislatura (que también puede encontrarse en la web para descargar) y se pueden consultar visitando el siguiente enlace:

https://drive.google.com/drive/folders/1hbLzkO4EuXrX_XECgVXQdojOfQsXtUdQ?usp=sharing

El enlace a la web propiamente dicha es: <https://sarabecerro.wixsite.com/quien-nos-represent>

Anexo 2: Código R al completo que se ha utilizado para la creación tanto de los gráficos, como de las dos partes interactivas: el mapa y la página con filtros donde ver la representación según las propias características de la persona. Se adjuntan en la nube por su gran extensión, separados en distintas carpetas donde se puede ver cuál es el código para acción realizada. En el CD que complementa este trabajo se incluye una segunda versión más extensa, desarrollada y con los pequeños detalles en una carpeta llamada “Código opción 2”. Para poder visualizar el código hay que descargarse los archivos que se encuentran en cada carpeta y al abrirlos, utilizar el bloc de notas del ordenador.

<https://drive.google.com/drive/folders/1oZx-BNpCbJr3zfTmo8OiKqPrI3-P4KhF?usp=sharing>

Anexo 3: Tablas obtenidas con R: se ha creado una carpeta con dichas tablas dentro de la nube para que se puedan observar correctamente y ampliar, ya que

Becerro Alonso, Sara (2019). *¿Quién nos representa? Análisis periodístico de datos sobre la representación de la población española en el Congreso de los Diputados*. Universidad de Valladolid. Curso 2018-2019.

la información viene dada en tablas que se han realizado con el propio lenguaje, al tener tantos datos aparecen en pequeño y necesitan ampliarse. Se encuentran en: https://drive.google.com/drive/folders/1yIQ48Q-r6ILbpu7bwMp82_3bF-mj82-c?usp=sharing

Anexo 4: Gráficos, mapas y demás contenido comparativo que se ha realizado a partir del lenguaje de programación R con la información de diputados y población, que se puede encontrar en las dos partes interactivas del reportaje, pinchando en los siguientes enlaces:

https://quiennosrepresenta.shinyapps.io/quien_nos_representa/

<https://quiennosrepresenta.shinyapps.io/congreso/>

Además, se han utilizado también como imágenes estáticas mediante capturas de pantalla. Se encuentran alojadas en el siguiente enlace de la nube:

<https://drive.google.com/drive/folders/1EixEu0S14DvT1zXTSIUa2RUfF-xvVXoW?usp=sharing>