



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

***Estudio comparativo de resultados de la EBAU:
aplicación shiny VisionEbau para su visualización***

Autor: Iñigo Aranzadi Bouzas
Tutora: María Teresa González Arteaga
2024-2025

Agradecimientos

A mi madre, por su sacrificio para que pueda formarme en lo que más me apasiona, y por animarme siempre a seguir aprendiendo y adquiriendo nuevos conocimientos.

A mi familia y amigos, por su apoyo y ayuda incondicional en todo momento.

A Lara e Irene, por “obligarme” a terminar este proyecto.

A mi tutora, María Teresa González Arteaga, por compartir generosamente sus conocimientos conmigo.

Gracias a todos.

Resumen

El propósito principal de este Trabajo de Fin de Grado es el desarrollo de una aplicación web interactiva denominada *VisionEbau*, cuyo objetivo es facilitar la visualización de los datos relacionados con la Evaluación de Bachillerato para el Acceso a la Universidad (EBAU). Para su implementación, se ha empleado la librería Shiny del lenguaje de programación R, lo que permite una presentación clara, dinámica e intuitiva de la información .

Además, se ha llevado a cabo un análisis estadístico que incluye la aplicación del test chi-cuadrado y un análisis de correspondencias simples, con el fin de comparar los resultados de las calificaciones obtenidas en la EBAU antes y después del período de pandemia.

El documento detalla también las herramientas empleadas a lo largo del proceso, así como las distintas fases que conforman el flujo de trabajo en un proyecto de visualización de datos. Estas etapas abarcan desde la extracción y el procesamiento de los datos, hasta su integración en una interfaz web funcional, incluyendo además la explicación estructurada de los componentes que conforman la aplicación web *VisionEbau*.

Palabras clave: EBAU, Shiny, R, aplicación web, visualización, análisis de correspondencias, test chi-cuadrado.

Abstract

The main objective of this Final Degree Project is the development of an interactive web application called *VisionEbau*, aimed at facilitating the visualization of data related to the Spanish University Entrance Examination (EBAU). The application has been implemented using the Shiny library of the R programming language, allowing for a clear, dynamic, and intuitive presentation of the information.

Additionally, a statistical analysis has been conducted, which includes the application of the chi-square test and a simple correspondence analysis. These analyses are intended to compare EBAU exam scores before and after the COVID-19.

The document also details the tools used throughout the process, as well as the various stages that make up the workflow in a data visualization project. These stages range from data extraction and processing to its integration into a functional web interface, including a structured explanation of the components that constitute the *VisionEbau* web application.

Keywords: EBAU, Shiny, R, web application, data visualization, correspondence analysis, chi-square test.

Índice general

1. Introducción	1
1.1. Objetivos de este Trabajo de Fin de Grado	1
1.2. Estructura de la memoria	2
2. Prueba de Evaluación de Bachillerato para el Acceso a la Universidad	3
2.1. Origen EBAU	3
2.2. Convocatorias	4
2.3. Estructura	5
2.4. Calificación	6
3. Datos	7
3.1. Extracción de los datos	7
3.1.1. Fuentes originales	7
3.1.2. Identificador y descripción de cada dato	8
3.1.3. Recopilación de archivos de datos	10
3.2. Filtrado de los datos	11
3.3. Exploración inicial de los datos	11
3.3.1. Valores ausentes (missing values)	11
3.3.2. Valores atípicos (Outliers)	12
3.4. Transformación de los datos	13
4. Recursos empleados	17
4.1. R	17
4.2. RStudio	18
4.3. Shiny	18
4.4. ggplot2	19
4.5. Plotly	19
5. Estructura y funcionamiento de una aplicación Shiny	21
5.1. Estructura de una aplicación Shiny	21
5.1.1. UI. Inputs y outputs	23
5.1.2. Server. Conexión entre inputs y outputs	24
5.2. Programación reactiva	25
5.3. Diseño y Organización de Aplicaciones Shiny	25
5.4. Paquete <i>Shinydashboard</i>	27
6. <i>VisionEbau</i>: aplicación Shiny de visualización de datos de EBAU	29
6.1. Página de <i>Inicio</i>	29
6.2. Menú de navegación	30
6.3. Página de <i>Resultados por Comunidad</i>	31
6.3.1. Apartado de <i>Aprobados</i>	31
6.3.2. Apartado de <i>Presentados</i>	32
6.3.3. Apartado de <i>Matriculados</i>	32
6.3.4. Apartado de <i>Evolución por Comunidad</i>	33

6.3.5.	Apartado de <i>Mapa</i>	34
6.3.6.	Apartado de <i>Mapa %</i>	34
6.4.	Página de <i>Nota media de Bachillerato</i>	35
6.4.1.	Apartado de <i>Comunidades</i>	35
6.4.2.	Apartado de <i>Tabla</i>	36
6.4.3.	Apartado de <i>Comparación comunidades por sexo</i>	36
6.4.4.	Apartado de <i>Comparación comunidades por centro</i>	37
6.5.	Página de <i>Distribución nota de Bachillerato</i>	37
6.5.1.	Apartado de <i>Comunidades</i>	37
6.5.2.	Apartado de <i>Tabla</i>	38
6.5.3.	Apartado de <i>Evolución</i>	38
6.6.	Páginas de <i>Nota media de Ebau</i> y <i>Nota media de Acceso</i>	39
6.7.	Páginas de <i>Distribución nota de Ebau</i> y <i>Distribución nota de Acceso</i>	39
6.8.	Página de <i>Información</i>	40
7.	Comparación de los resultados de la EBAU	41
7.1.	Descripción y preparación de los datos	41
7.2.	Aplicación Test chi-cuadrado	42
7.2.1.	Metodología	42
7.2.2.	Evaluación de los resultados	42
7.3.	Aplicación de un análisis de correspondencias	44
7.3.1.	Metodología	44
7.3.2.	Generación de la matriz de correspondencias	45
7.3.3.	Visualización de resultados mediante Biplot	45
7.3.4.	Interpretación de los resultados	46
8.	Conclusiones	51
	Anexos	53
	A. Extracción Datos	55
	B. Imágenes de la aplicación <i>VisionEbau</i>	59
	C. Construcción tablas de contingencia	77
	D. Tablas de residuos estandarizados y gráficos de mosaicos	79
	E. Biplots análisis de correspondencias	91
	Bibliografía	98

Índice de figuras

3.1. Diagramas de cajas para la búsqueda de valores atípicos en las distribuciones de calificaciones . . .	12
3.2. Diagramas de cajas para la búsqueda de valores atípicos en notas por sexo y por centro	13
3.3. Diagramas de cajas para la búsqueda de valores atípicos en matriculados, presentados y aprobados	13
3.4. Cálculo conjunto de datos porcentaje aprobados	14
3.5. Cálculo conjunto de datos para la representación del mapa	15
4.1. Interfaz de RStudio	18
5.1. Esquema de una aplicación Shiny: interfaz de usuario (UI) y servidor (Server)	22
5.2. Componentes principales de una aplicación Shiny	22
5.3. Resultado de la ejecución de la aplicación Shiny	22
5.4. Listado principales Inputs	24
5.5. Listado principales Outputs	24
5.6. Relación entre Outputs y funciones de renderización	25
5.7. Ejemplo principales layouts	27
5.8. Estructura aplicación diseñada con Shinydashboard	28
6.1. Página de Inicio	29
6.2. Menú de navegación	31
6.3. Resultados por comunidad. Pestaña de Aprobados	32
6.4. Resultados por comunidad. Pestaña de Matriculados	33
6.5. Resultados por comunidad. Pestaña de Evolución por Comunidad	33
6.6. Resultados por comunidad. Pestaña de Mapa	34
6.7. Nota media de Bachillerato. Pestaña de Comunidades	35
6.8. Nota media de Bachillerato. Pestaña de Tabla	36
6.9. Nota media de Bachillerato. Pestaña de Comparación comunidades por sexo	37
6.10. Distribución nota de Bachillerato. Pestaña de Comunidades	38
6.11. Distribución nota de Bachillerato. Pestaña de Evolución	39
6.12. Página de <i>Información</i>	40
7.1. Tabla de contingencia para Castilla y León	42
7.2. Gráfico de mosaico para la comunidad de Castilla y León	43
7.3. Tabla de coordenadas de las columnas para Castilla y León	45
7.4. Tabla de coordenadas de las filas para Castilla y León	45
7.5. Biplot para la comunidad de Extremadura	46
7.6. Biplot para la comunidad de Andalucía	47
7.7. Biplot para la comunidad de Castilla y León	48
7.8. Biplot para la comunidad de las Islas Canarias	49
A.1. Definición de columnas y valores de las observaciones	55
A.2. Bucle <code>for</code> para generar las observaciones en Python	55
A.3. Observaciones exportadas a un archivo CSV desde Python	56
A.4. Vista inicial de los datos descargados en formato <code>xlsx</code>	56
A.5. Formato final de los datos tras replicar las variables	56
A.6. Código en R para transformar, transponer y unificar los datos	57

A.7. alida final de los datos exportados en CSV	57
B.1. Página de Inicio	59
B.2. Menú de navegación	60
B.3. Página de Resultados por Comunidad. Pestaña de <i>Aprobados</i>	60
B.4. Página de Resultados por Comunidad. Pestaña de <i>Aprobados</i> . Hombres y Mujeres	61
B.5. Página de Resultados por Comunidad. Pestaña de <i>Presentados</i>	61
B.6. Página de Resultados por Comunidad. Pestaña de <i>Presentados</i> . Hombres y Mujeres	62
B.7. Página de Resultados por Comunidad. Pestaña de <i>Matriculados</i> . Hombres y Mujeres	62
B.8. Página de Resultados por Comunidad. Pestaña de <i>Evolución por Comunidad</i> . Hombres y Mujeres	63
B.9. Página de Resultados por Comunidad. Pestaña de <i>Mapa %</i>	63
B.10. Página de Nota media de Bachillerato. Pestaña de <i>Comparación comunidades por sexo</i>	64
B.11. Página de Nota media de Bachillerato. Pestaña de <i>Comparación comunidades por centro</i>	64
B.12. Página de Nota media de Bachillerato. Pestaña de <i>Comparación comunidades por centro</i>	65
B.13. Página de Distribución nota de Bachillerato. Pestaña de <i>Tabla</i>	65
B.14. Página de Nota media de Ebau. Pestaña de <i>Comunidades</i>	66
B.15. Página de Nota media de Ebau. Pestaña de <i>Tabla</i>	66
B.16. Página de Nota media de Ebau. Pestaña de <i>Comparación comunidades por sexo</i>	67
B.17. Página de Nota media de Ebau. Pestaña de <i>Comparación comunidades por sexo</i>	67
B.18. Página de Nota media de Ebau. Pestaña de <i>Comparación comunidades por sexo</i>	68
B.19. Página de Nota media de Ebau. Pestaña de <i>Comparación comunidades por centro</i>	68
B.20. Página de Nota media de Ebau. Pestaña de <i>Comparación comunidades por centro</i>	69
B.21. Página de Nota media de Acceso. Pestaña de <i>Comunidades</i>	69
B.22. Página de Nota media de Acceso. Pestaña de <i>Tabla</i>	70
B.23. Página de Nota media de Acceso. Pestaña de <i>Comparación comunidades por sexo</i>	70
B.24. Página de Nota media de Acceso. Pestaña de <i>Comparación comunidades por sexo</i>	71
B.25. Página de Nota media de Acceso. Pestaña de <i>Comparación comunidades por sexo</i>	71
B.26. Página de Nota media de Acceso. Pestaña de <i>Comparación comunidades por centro</i>	72
B.27. Página de Nota media de Acceso. Pestaña de <i>Comparación comunidades por centro</i>	72
B.28. Página de Distribución nota de Ebau. Pestaña de <i>Comunidades</i>	73
B.29. Página de Distribución nota de Ebau. Pestaña de <i>Tabla</i>	73
B.30. Página de Distribución nota de Ebau. Pestaña de <i>Evolución</i>	74
B.31. Página de Distribución nota de Acceso. Pestaña de <i>Comunidades</i>	74
B.32. Página de Distribución nota de Acceso. Pestaña de <i>Tabla</i>	75
B.33. Página de Distribución nota de Acceso. Pestaña de <i>Evolución</i>	75
D.1. Gráfico de mosaico para la comunidad de Andalucía	83
D.2. Gráfico de mosaico para la comunidad de Aragón	83
D.3. Gráfico de mosaico para la comunidad de Asturias	83
D.4. Gráfico de mosaico para la comunidad de las Islas Baleares	84
D.5. Gráfico de mosaico para la comunidad de las Islas Canarias	84
D.6. Gráfico de mosaico para la comunidad de Cantabria	85
D.7. Gráfico de mosaico para la comunidad de castilla y León	85
D.8. Gráfico de mosaico para la comunidad de Cataluña	86
D.9. Gráfico de mosaico para la comunidad de Extremadura	86
D.10. Gráfico de mosaico para la comunidad de Galicia	86
D.11. Gráfico de mosaico para la comunidad de Madrid	87
D.12. Gráfico de mosaico para la comunidad de Castilla la Mancha	87
D.13. Gráfico de mosaico para la comunidad de la Región de Murcia	87
D.14. Gráfico de mosaico para la comunidad de Navarra	88
D.15. Gráfico de mosaico para la comunidad de la Rioja	88
D.16. Gráfico de mosaico para la comunidad Valenciana	89
D.17. Gráfico de mosaico para la comunidad del Pais Vasco	89
E.1. Biplot para la comunidad de Aragón	91
E.2. Biplot para la comunidad de Asturias	91

E.3. Biplot para la comunidad de las Islas baleares	92
E.4. Biplot para la comunidad de Cantabria	92
E.5. Biplot para la comunidad de Cataluña	92
E.6. Biplot para la comunidad de Extremadura	93
E.7. Biplot para la comunidad de Galicia	93
E.8. Biplot para la comunidad de Madrid	93
E.9. Biplot para la comunidad de Castilla la Mancha	94
E.10. Biplot para la comunidad de Murcia	94
E.11. Biplot para la comunidad de la Rioja	94
E.12. Biplot para la comunidad de Valencia	95
E.13. Biplot para la comunidad del Pais Vasco	95
E.14. Biplot para la comunidad de Navarra	95

Capítulo 1

Introducción

La EBAU es la prueba obligatoria que da acceso a las universidades en España, de ahí su gran interés. Esta prueba de acceso se realiza todos los años en cada una de las comunidades autónomas de nuestro país por todas las personas que quieren acceder a una universidad pública o privada. Los resultados suponen cada año un gran volumen de datos. Estos datos pueden ser utilizados por los profesores, coordinadores educativos y las autoridades educativas para evaluar la dificultad de dicha prueba, preparación de los alumnos e identificación de distintos aspectos a mejorar relativos a los resultados.

Las aplicaciones web han cobrado mucha importancia en los últimos años debido a su fácil accesibilidad, se puede acceder desde cualquier dispositivo con navegador web y una conexión a internet. Esto, además implica que no sea necesario la instalación una aplicación para su ejecución por el usuario final.

Uno de los entornos de programación mas utilizados para el análisis estadístico, visualización de datos y computación numérica es el entorno de programación R. En este entorno podemos encontrar un paquete llamado Shiny el cual permite desarrollar visualizaciones web.

Este TFG presenta *VisionEbau*¹, una aplicación de visualización de datos de la EBAU que permite explorar información detallada sobre los matriculados, presentados y aprobados en cada comunidad autónoma y ofrece datos relativos a las calificaciones de Bachillerato, la EBAU y la nota de acceso a la universidad, proporcionando una visión amplia del rendimiento académico en cada región.

1.1. Objetivos de este Trabajo de Fin de Grado

El propósito principal de este TFG es realizar una visualización para comparar los resultados de la Evaluación de Bachillerato para el Acceso a la Universidad (EBAU) en las diferentes comunidades autónomas durante los años previos y posteriores a la pandemia de COVID-19.

Se planea realizar la preparación de los datos, así como el diseño y desarrollo de una aplicación interactiva basada en Shiny. *VisionEbau* tiene la capacidad de visualizar datos asociados con la EBAU, generación de tablas resumen y gráficos descriptivos, todo ello a través de una interfaz moderna, intuitiva y de fácil uso.

Asimismo, se prevé la aplicación de un test chi-cuadrado y un análisis de correspondencias con el objetivo de comparar las calificaciones en la EBAU para los distintos años.

¹<http://shiny1.eio.uva.es:3838/aranzadi/visionebau/>

1.2. Estructura de la memoria

La presente memoria esta estructurada de la siguiente manera:

- **Capítulo 1. Introducción.** Contextualización del Trabajo de Fin de Grado, objetivos a alcanzar y estructuración de la memoria.
- **Capítulo 2. Prueba de Evaluación de Bachillerato para el Acceso a la Universidad.** Exposición sobre cuestiones generales y contexto de la Evaluación de Bachillerato para el Acceso a la Universidad.
- **Capítulo 3. Datos.** Explicación de los datos utilizados en la aplicación, las fuentes de procedencia de los datos, el análisis para observar la calidad de los datos obtenidos y su posterior tratamiento.
- **Capítulo 4. Recursos empleados.** Explicación de las herramientas utilizadas para el desarrollo de la visualización: RStudio, Shiny..
- **Capítulo 5. Estructura y funcionamiento de una aplicación Shiny.** Se presenta la estructura y programación básica de una aplicación Shiny.
- **Capítulo 6. *VisionEbau*: aplicación Shiny de visualización de datos de EBAU.** Se muestra la programación detallada de la aplicación *VisionEbau* así como sus distintas funcionalidades.
- **Capítulo 7. Comparación de los resultados de la EBAU** Se comparan las calificaciones en la EBAU considerando los intervalos de notas de acceso correspondientes al periodo comprendido entre los años 2018 y 2022.
- **Capítulo 8. Conclusiones.** Se exponen las conclusiones obtenidas del Trabajo de Fin de Grado y se comentan algunas posibles líneas de mejora futuras para la aplicación.

Capítulo 2

Prueba de Evaluación de Bachillerato para el Acceso a la Universidad

La prueba de acceso a la universidad hasta el último Real decreto 534/2024 del 11 de junio recibía distintas denominaciones dependiendo de la comunidad autónoma. En el presente trabajo de fin de grado, nos referimos a dicha prueba con el nombre que se le otorga en la Comunidad de Castilla y León: EBAU (Evaluación del Bachillerato para el Acceso a la Universidad).

El propósito formal de la Evaluación de Bachillerato para el Acceso a la Universidad, EBAU, consiste en evaluar los conocimientos y competencias adquiridos por los estudiantes durante su etapa de educación secundaria referente a los cursos de primero y segundo de bachillerato.

Esta prueba escrita tiene como objetivo determinar la capacidad de los estudiantes para acceder a una educación superior universitaria. Asimismo, la EBAU cumple una función fundamental como criterio de selección por parte de las universidades, asegurando un proceso equitativo para la admisión de estudiantes en sus programas académicos al establecer un estándar común de evaluación para todos los aspirantes.

2.1. Origen EBAU

En esta sección se aborda el comienzo y la evolución de la prueba de acceso a la universidad a lo largo de los años, analizando los cambios y adaptaciones que ha experimentado este proceso en respuesta a las necesidades educativas y sociales de cada época.

Nos remontamos a 1857 con la promulgación de la Ley Moyano, la cual estableció la organización del sistema educativo español. Esta legislación reguló la enseñanza en tres niveles: la enseñanza primaria, la segunda enseñanza o enseñanza media, que incluía institutos y escuelas normales, y la enseñanza superior universitaria, cuya administración estaba a cargo del Estado [1].

En el artículo 26 de la Ley Moyano se establecía el requisito de poseer el título de Bachiller en Artes para poder inscribirse en las facultades. Dicho título se obtenía mediante el Examen de Grado de Bachiller en Artes, el cual abarcaba los contenidos de las materias cursadas en la Segunda Enseñanza.

Con el paso del tiempo, se realizaron múltiples modificaciones en la organización del sistema educativo, tanto a nivel estructural como en cuanto al tipo de pruebas llevadas a cabo en cada curso.

En la década de los setenta del siglo XX, la educación universitaria experimentó un aumento en su popularidad, lo que planteó el desafío de gestionar la selección de los estudiantes que buscaban acceder a la universidad.

Las primeras pruebas de Selectividad en España se celebraron en el año 1975, estas pruebas fueron fruto de la Ley 30/1974, de 24 de julio, sobre pruebas de aptitud para acceso a las Facultades, Escuelas Técnicas Superiores, Colegios Universitarios y Escuelas Universitarias.

Una de las principales novedades de esta prueba consistía en que los alumnos debían aprobar el Curso de Orientación Universitaria (COU) para poder presentarse a ella. Este requisito establecía una etapa previa indispensable en el proceso de acceso a la universidad, garantizando que los estudiantes alcanzaran el nivel académico necesario antes de someterse a la Selectividad.

El objetivo de esta prueba era asegurarse que los alumnos que accedían a la universidad tenían la vocación, conocimientos y preparación para su ingreso en la universidad. Esta prueba trajo consigo nuevas modificaciones:

- Esta prueba aseguraría el anonimato, uniformidad y objetividad en la corrección.
- Los alumnos realizarían la prueba en la universidad a cuyo distrito perteneciese el centro donde habían cursado COU.
- Las materias a evaluar eran las cursadas durante el año de COU.
- Ninguna prueba sería de carácter eliminatorio
- La puntuación total del alumno para acceder a la universidad tendría en cuenta el historial académico del estudiante.

En el año 2006, con la implementación de la Ley Orgánica de Educación (LOE), la prueba adoptó la denominación de Prueba de Acceso a la Universidad (PAU). Posteriormente, a partir de 2010, se produjo un nuevo cambio de nomenclatura para adaptarse a las directrices del Espacio Europeo de Educación Superior, pasando a denominarse Prueba de Acceso a los Estudios de Grado (PAEG) en algunas Comunidades Autónomas. Este cambio no solo introdujo una modificación en el nombre, sino que también introdujo otro cambio bastante significativo, la calificación máxima se amplió de un 10 a un 14.

En el año 2017, con la implementación de la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE), la prueba de acceso a la universidad experimentó un nuevo cambio de denominación. En esta ocasión, pasó de llamarse PAEG a Evaluación para el Acceso a la Universidad (EvAU), o en algunas Comunidades Autónomas, EBAU (Evaluación de Bachillerato para el Acceso a la Universidad).

En 2020, debido a la pandemia de COVID-19, se implementó un modelo excepcional para la Evaluación para el Acceso a la Universidad (EvAU). Este modelo adaptó la prueba a las circunstancias especiales generadas por la crisis sanitaria, permitiendo a los estudiantes presentar exámenes en condiciones modificadas. Hasta el año 2023-2024, se aplicó el modelo excepcional establecido en 2020 en respuesta a la pandemia de COVID-19. Actualmente se encuentra en desarrollo un nuevo modelo publicado en el Real decreto 534/2024.

2.2. Convocatorias

La Evaluación para el Acceso a la Universidad (EBAU) se compone de dos convocatorias anuales. La primera corresponde a la fase ordinaria, que se lleva a cabo en junio, mientras que la segunda corresponde a la fase extraordinaria, que se realiza en julio [2].

Los estudiantes que hayan aprobado todas las asignaturas de bachillerato en el periodo ordinario podrán presentarse a la convocatoria ordinaria. En caso contrario, deberán aprobar las asignaturas pendientes para poder acceder a la fase extraordinaria.

Asimismo, es importante señalar que los estudiantes que hayan aprobado previamente la prueba tendrán la opción de presentarse nuevamente para mejorar su calificación en las convocatorias posteriores, si así lo desean.

Estos estudiantes tendrán la opción de presentarse únicamente a la fase específica, manteniendo como nota

de la calificación general la última obtenida. También podrán optar por presentarse únicamente a la fase general, donde la nueva calificación general solo se calculará si mejora la calificación anterior. Alternativamente, podrán presentarse a ambas fases, considerando como nueva calificación final la ponderación de ambas [3].

2.3. Estructura

La prueba consta de dos fases, la fase general y la fase específica. En ambas fases cada examen tiene una duración máxima de 90 minutos con un descanso entre prueba y prueba de al menos 45 minutos [2] [3].

La fase general es de carácter obligatorio para todos los alumnos matriculados. Esta fase está formada por cuatro ejercicios, tres de ellos se corresponden con tres de las cuatro materias comunes de 2º de Bachillerato. El cuarto ejercicio está asociado a la materia específica elegida por el estudiante para la modalidad cursada en la prueba.

Las asignaturas comunes para todos los alumnos son:

- Lengua castellana y literatura.
- Lengua extranjera.
- Historia de España o Historia de la filosofía.

Además, el alumno puede elegir entre los siguientes idiomas: inglés, francés, alemán, italiano o portugués como prueba de evaluación de la Lengua extranjera.

En relación con la asignación de la materia específica según la modalidad seleccionada, que constituye el cuarto componente de esta fase, puede variar entre las distintas comunidades autónomas, aunque la mayoría coincide en las siguientes asignaturas.

- Matemáticas II para la modalidad de Ciencias.
- Latín es la asignatura correspondiente para la modalidad de bachillerato de Humanidades.
- Matemáticas Aplicadas a las Ciencias Sociales es la prueba a realizar para la modalidad de bachillerato de Ciencias sociales.
- Fundamentos de Arte es la asignatura a evaluar para la modalidad de bachillerato de Artes.

La fase específica es opcional para todos los alumnos. En esta fase el alumno puede examinarse de hasta cuatro exámenes correspondientes con materias de opción del bloque de las asignaturas troncales de segundo curso de Bachillerato elegidas para la prueba. No es necesario que el alumno haya cursado la asignatura para presentarse a su evaluación. Esta fase está formada por las siguientes asignatura:

- Biología.
- Dibujo Técnico II.
- Física.
- Geología.
- Química.
- Economía de la Empresa.
- Geografía.
- Griego II.
- Historia del Arte.
- Historia de la Filosofía.
- Artes Escénicas.
- Cultura Audiovisual II.

Hasta el año previo al COVID cada enunciado de una asignatura estaba compuesta por dos opciones: la opción A y la opción B. El estudiante debía seleccionar una de estas dos opciones y responder únicamente a las preguntas correspondientes a la opción elegida.

Tras el año del COVID, todas las comunidades autónomas realizaron ligeros ajustes en la estructura del examen, reemplazando la opción A y B por bloques de preguntas entre los cuales el alumno debe seleccionar cuáles realizar. Este nuevo modelo covid se ha mantenido desde el curso 2019-2020 hasta el curso 2023-2024.

2.4. Calificación

En esta sección, abordaremos el proceso mediante el cual se determina la calificación final [2].

La evaluación de cada materia se califica en una escala de 0 a 10 puntos, donde 0 representa la calificación mínima y 10 la calificación máxima posible. Para que el estudiante supere la asignatura, deberá alcanzar una puntuación mínima de 5 en dicha prueba.

La calificación máxima final de la prueba de acceso a la universidad es de 14 puntos en lugar de 10. Esto se logra asignando hasta seis puntos por el promedio de calificaciones del bachillerato, hasta cuatro puntos por la puntuación obtenida en la prueba general, y hasta cuatro puntos adicionales por la fase específica.

Por lo tanto, en lo que respecta a la calificación final, es decir, la nota de acceso, se considera que el 60 % corresponde al promedio de calificaciones del bachillerato, y el 40 % restante se basaba en la puntuación obtenida en la prueba de acceso a la universidad.

La fórmula de la nota de acceso quedaría de la siguiente manera:

$$\text{Nota de acceso} = 0.6 \cdot NB + 0.4 \cdot CG + a \cdot M1 + b \cdot M2$$

Considerando la fórmula anterior, donde NB se corresponde con la nota media de los dos cursos de bachillerato, CG es la nota media obtenida en la fase general, $M1$ y $M2$ representan las dos mejores notas obtenidas en la fase específica, y los parámetros a y b indican la ponderación de las materias específicas.

Es relevante señalar que la ponderación de las materias específicas varía en función de la titulación a la que se aspire y de la universidad en que se curse, estando relacionada la materia a ponderar con la modalidad del Grado a la que se desea acceder. Estos valores pueden ser de 0.1 o 0.2.

Capítulo 3

Datos

En este capítulo se introducen los datos fundamentales para la aplicación web desarrollada en este TFG, ya que proporcionan la información necesaria para generar las visualizaciones.

Generalmente, se comienza por un proceso de extracción de datos, que consiste en recopilar y almacenar los datos de sus fuentes originales; un proceso de depuración, donde se aplican una serie de reglas para garantizar la calidad de los datos; y, por último, un proceso de carga, en el cual los datos ya introducidos al formato correcto se transfieren al entorno de explotación para permitir su análisis posterior.

A continuación, procedemos a detallar de manera más exhaustiva cada uno de los pasos mencionados en este apartado.

3.1. Extracción de los datos

El proceso de extracción de datos consiste en recopilar los datos de diversas fuentes para su posterior análisis y utilización. Determinando de dónde se van a extraer los datos, extrayendo los datos desde las fuentes y almacenando los datos en un repositorio adecuado para su análisis posterior.

3.1.1. Fuentes originales

Los datos que utilizaremos están proporcionados por el Ministerio de Ciencia, Innovación y Universidades. Aquí encontramos diversos repositorios, nosotros utilizamos datos que se encuentran en la sección de *Estadística de las pruebas de acceso a la universidad (PAU)*. Destacar que no disponemos de los microdatos ya que ni el INE ni el ministerio los ha proporcionado.

La sección de *Estadística de las pruebas de acceso a la universidad (PAU)* esta formada por un total de 54 conjuntos de datos, divididos en tres secciones: Resultados nacionales, Resultados por comunidad autónoma de realización y Resultados por universidad de realización.

Para este Trabajo de Fin de Grado extraemos los conjuntos de datos que pertenecen a la sección de *Resultados por comunidad autónoma de realización*.

En esta sección se presentan 27 conjuntos de datos. Los que utilizamos en este trabajo son los siguientes:

- Nota de Bachiller por tipo de matrícula, convocatoria y sexo.
- Nota de Bachiller por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP.
- Nota en la fase general de los aptos por tipo de matrícula, convocatoria y sexo.
- Nota en la fase general de los aptos por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP.

- Nota de Acceso al grado de los titulados de Bachiller por tipo de matrícula, convocatoria y sexo.
- Nota de Acceso al grado de los titulados de Bachiller por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP.
- Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo.
- Nota de Bachiller. Distribución por tipo de matrícula, convocatoria y sexo.
- Nota en la fase general de los aptos. Distribución por tipo de matrícula, convocatoria y sexo.
- Nota de Acceso al grado de los titulados de Bachiller. Distribución por tipo de matrícula, convocatoria y sexo.

3.1.2. Identificador y descripción de cada dato

Los conjuntos de datos que utilizaremos, mencionados en el punto anterior, comparten en todos los casos tres variables comunes.

- **Comunidad**, como su propio nombre indica, corresponde al lugar donde se ha realizado la prueba, toma los valores de las 17 CCAA de España y el valor de 'Estado'.
- **Convocatoria**, identificador del tipo de convocatoria, puede tomar los valores 'Total', 'Ordinaria' y 'Extraordinaria'.
- **Periodo**, identificador del año al que hacen referencia los datos. Puede tomar los valores comprendidos entre 2018 y 2022 ambos incluidos.

A continuación, se detalla las variables específicas de cada conjunto de datos para ofrecer una explicación exhaustiva de sus particularidades.

Para el conjunto de datos *Nota de Bachiller por tipo de matrícula, convocatoria y sexo* y además de las variables comunes tenemos adicionalmente estas otras:

- Sexo, corresponde al sexo de las personas consideradas para el cálculo de las medias y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Nota, representa la media de la calificación de la nota de bachillerato, toma un valor numérico entre 5 y 10.
- Matrícula, se corresponde con la fase en la prueba. Toma los valores 'Fase general y específica', 'Solo fase general' o 'Solo fase específica'.

En el conjunto de datos *Nota de Bachiller por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP* se incorporan además de las variables comunes las siguientes variables:

- Centro, hace referencia a la naturaleza del centro donde estudia bachillerato, toma los valores 'Total', 'Publico', 'Privado', 'Concertado', 'No consta'.
- Nota, representa la calificación de la nota media de bachillerato, toma un valor numérico entre 5 y 10.
- Matrícula hace referencia a la fase del examen en la que se inscribe el estudiante. Puede tomar los valores 'Fase general y específica', 'Solo fase general' o 'Solo fase específica'.

El conjunto de datos *Nota en la fase general de los aptos por tipo de matrícula, convocatoria y sexo*, además de las variables comunes, se incluyen las siguientes variables:

- Sexo, corresponde al género de la persona evaluada y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Nota, representa la calificación de la nota media de la prueba EBAU, toma un valor numérico entre 4 y 10.

- Matrícula, se corresponde con la fase en la prueba. Toma los valores 'Fase general y específica' o 'Solo fase general'.

Para los datos correspondientes al conjunto *Nota en la fase general de los aptos por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP*, adicionalmente de las variables comunes, encontramos las siguientes variables:

- Centro, esta variable representa la naturaleza del centro donde estudian bachillerato, toma los valores 'Total', 'Publico', 'Privado', 'Concertado', 'No consta'.
- Nota, representa la calificación de la nota media de la prueba EBAU, toma un valor numérico entre 4 y 10.
- Matrícula, se corresponde con la fase en la prueba. Toma los valores 'Fase general y específica' o 'Solo fase general'.

Para el conjunto de datos *Nota de Acceso al grado de los titulados de Bachiller por tipo de matrícula, convocatoria y sexo* y además de las variables comunes tenemos adicionalmente estas otras:

- Sexo, corresponde al género de la persona evaluada y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Nota, representa la calificación de la nota media de acceso donde estudian bachillerato, toma un valor numérico entre 5 y 10.
- Matrícula hace referencia a la fase del examen en la que se inscribe el estudiante. Puede tomar los valores 'Fase general y específica' o 'Solo en la fase general'.

En el conjunto de datos *Nota de Acceso al grado de los titulados de Bachiller por tipo de matrícula, convocatoria y naturaleza del centro de secundaria/FP* se incorporan además de las variables comunes las siguientes variables:

- Centro, esta variable representa la naturaleza del centro donde estudian bachillerato, toma los valores 'Total', 'Publico', 'Privado', 'Concertado', 'No consta'.
- Nota, representa la calificación de la nota media de acceso, toma un valor numérico entre 5 y 10.
- Matrícula hace referencia a la fase del examen en la que se inscribe el estudiante. Puede tomar los valores 'Fase general y fase específica' o 'Solo en la fase general'.

Para los datos correspondientes al conjunto *Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo*, adicionalmente a las variables comunes, encontramos las siguientes variables:

- Indicador, hace alusión al estado del alumno, y puede tomar los valores: 'Matriculados', 'Presentados' o 'Aprobados'.
- Sexo, corresponde al sexo de la persona participante y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Cifra, indica el numero de alumnos por indicador, sexo, comunidad autonoma, periodo y matricula.
- Acceso, esta variable únicamente adopta el valor 'PAU genérica', ya que los únicos datos de interés corresponden a esta categoría.

En el conjunto de datos *Nota de Bachiller. Distribución por tipo de matrícula, convocatoria y sexo* se incorporan además de las variables comunes las siguientes variables:

- Sexo, corresponde al sexo de la persona evaluada y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'. Rango, cuyos valores pueden ser 'De 5 a 6', 'De 6 a 7', 'De 7 a 8', 'De 8 a 9' y 'De 9 a 10' haciendo referencia al intervalo en el que se encuentra la nota para dicha distribución.
- Distribución, hace referencia a cuantos aprobados caen en cada intervalo.

- Matrícula, se corresponde con la fase en la prueba. Toma los valores 'Fase general y específica', 'Solo fase general' o 'Solo fase específica'.

Para los datos correspondientes al conjunto *Nota en la fase general de los aptos. Distribución por tipo de matrícula, convocatoria y sexo*, adicionalmente a las variables comunes, encontramos las siguientes variables:

- Rango, indica los tramos en los que se agrupan las calificaciones para la distribución, pudiendo tomar los valores 'De 4 a 5', 'De 5 a 6', 'De 6 a 7', 'De 7 a 8', 'De 8 a 9' y 'De 9 a 10'.
- Sexo, corresponde al sexo de la persona evaluada y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Distribución, se refiere al número de aprobados que se ubican dentro de cada intervalo definido de notas.
- Matrícula, se corresponde con la fase en la prueba. Toma los valores 'Fase general y específica' o 'Solo fase general'.

Para los datos correspondientes al conjunto *Nota en la fase general de los aptos. Distribución por tipo de matrícula, convocatoria y sexo*, adicionalmente a las variables comunes, encontramos las siguientes variables:

- Rango, cuyos valores pueden ser 'De 5 a 6', 'De 6 a 7', 'De 7 a 8', 'De 8 a 9' y 'De 9 a 10' haciendo referencia al intervalo en el que se encuentra la nota para dicha distribución.
- Sexo, corresponde al sexo de la persona evaluada y puede asumir los valores 'Ambos sexos', 'Mujeres', 'Hombres'.
- Distribución, hace referencia a cuantos aprobados caen en cada intervalo.
- Matrícula hace referencia a la fase del examen en la que se inscribe el estudiante. Puede tomar los valores 'Fase general y fase específica' o 'Solo en la fase general'.

En los conjuntos de datos correspondientes a las distribuciones, la variable *Comunidad* también puede asumir el valor Total para recoger los datos a nivel nacional.

3.1.3. Recopilación de archivos de datos

La tarea de descarga y recopilación de los archivos de datos se ha llevado a cabo mediante dos métodos. El repositorio ofrece la opción de descargar los datos en formatos CSV y XLS, lo cual facilitaba el proceso inicial de obtención. Sin embargo, la disposición en la que se encontraban los datos descargados no es adecuada para su posterior utilización, lo que hace necesario emplear dos enfoques distintos para su descarga y transformación a un formato correcto.

Para todos los archivos de datos, excepto para el conjunto de datos *Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo* y los conjuntos de datos de las distribuciones, se ha utilizado un script en Python con el fin de organizar los datos en el formato requerido. Para ello, dichos archivos fueron descargados en formato CSV con separación por comas. El programa implementado en Python organiza los datos en una lista y, posteriormente, una función los procesa mediante un bucle, generando de forma iterativa las observaciones necesarias.

Como se puede apreciar en la Figura A.1 del Anexo A, se han definido las columnas con los valores que pueden tomar las observaciones para su posterior utilización en un bucle, el cual se muestra en la Figura A.2 del Anexo A. Este bucle se encarga de generar todas las observaciones necesarias.

Una vez generadas las observaciones, tal como se muestra en la Figura A.3 del Anexo A, los datos se exportan a un archivo CSV para su posterior utilización en el programa R.

Para el conjunto de datos *Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo* se ha desarrollado un programa en R con el objetivo de disponer las observaciones en el formato adecuado. En este caso, el archivo de datos se descarga en formato xlsx. Inicialmente, el archivo presentaba el aspecto que se muestra en la Figura A.4 del Anexo A. Posteriormente, las variables se replicaron

hasta obtener el formato final, como se ilustra en la Figura A.5 del Anexo A.

Una vez que el archivo se encuentra en el formato adecuado, este es importado en el programa R para generar la transpuesta correspondiente a cada año, unificar todos los periodos en un único archivo de datos y asignar los nombres de las variables, tal como se muestra en la Figura A.6 del Anexo A. Una vez generada la salida esperada, como se muestra en la Figura A.7 del Anexo A, los datos se exportan a un archivo en formato *CSV* para su posterior procesamiento.

3.2. Filtrado de los datos

En esta sección previa al análisis exploratorio de los datos, se detalla el proceso de filtrado aplicado a los distintos conjuntos de datos con el objetivo de conservar únicamente la información relevante para este estudio.

Una parte importante del filtrado se realizó antes de la descarga, seleccionando únicamente los datos de interés. Así, aunque la fuente original abarcaba el período comprendido entre 2015 y 2022, se decidió limitar la descarga a los años 2018 a 2022, considerados los más pertinentes para el análisis por ser más recientes y permitir comparar el periodo antes de la pandemia y después de la pandemia.

Posteriormente, se aplicaron filtros adicionales. En primer lugar, se seleccionaron únicamente los registros en los que la variable *Matrícula* toma el valor 'Fase general y fase específica', ya que este permite obtener una visión completa del fenómeno de estudio.

En segundo lugar, en los conjuntos de datos que distinguen por tipo de centro educativo, se eliminaron los registros correspondientes a los valores 'Concertado' y 'No consta', dado que en España no existe Bachillerato en centros de educación secundaria concertada.

Finalmente, se excluyeron los registros con el valor 'Estado' en la variable *Comunidad*, debido a que no se cuenta con una definición clara de a qué se refieren dichos datos. Asimismo, en los conjuntos de datos relacionados con distribuciones, se eliminaron los registros con el valor 'Total', dado que esta categoría solo aparece en esos tres conjuntos, y su exclusión permite mantener la coherencia con el resto de la base de datos.

3.3. Exploración inicial de los datos

Antes de proceder al análisis de los datos, se lleva a cabo una exploración preliminar con el fin de identificar posibles valores ausentes, detectar valores atípicos.

En esta subsección se procederá a realizar un análisis exhaustivo de los conjuntos de datos previamente mencionados.

3.3.1. Valores ausentes (missing values)

El primer paso consiste en identificar la presencia de valores ausentes en cada conjunto de datos y calcular el porcentaje correspondiente. Para ello, se emplea el siguiente script en R, ejecutado individualmente para cada conjunto de datos. La salida obtenida proporciona el porcentaje de valores ausentes presentes en cada uno de ellos.

```
1 #matriculados, presentados, aprobados
2 datos1 <- read.csv("C:/Users/iaranzadibouzas/OneDrive - Deloitte (0365D)/Documents/TFG
   estadistica/datos/Datos finales/matriculados_presentados_aprobados.csv", sep = ",",
   header=TRUE)
3 datos1$Distribucion<- as.numeric(datos1$Cifra)
4 datso1Nan <- round (( colMeans ( is.na( datos1 ) ) ) *100 ,1)
5 print ( sort ( datso1Nan [ datso1Nan >0] , decreasing = T ) )
```

Listing 3.1: Código en R

Hemos identificado la presencia de valores ausentes en los siguientes conjuntos de datos que se muestran en el Cuadro 3.1.

Conjunto datos	Porcentaje
Nota acceso por centro	9.4
Nota acceso distribucion por sexo	1
Nota bachiller distribucion por sexo	0.4
Nota bachillerato por centro	9.4
Nota selectividad por centro	9.4
Nota selectividad distribucion por sexo	1.7

Cuadro 3.1: Tabla porcentaje valores ausentes

Existen múltiples enfoques para el tratamiento de valores ausentes, como la eliminación de la variable afectada o la imputación mediante medidas estadísticas como la media o la moda. No obstante, en este caso, se descarta la eliminación de la variable, dado que contiene información esencial que no puede ser prescindida. De igual manera, se descartan las técnicas de imputación, ya que sustituir los valores ausentes por la media o la moda no resulta adecuado ni representativo para este contexto.

Por consiguiente, dado lo expuesto anteriormente y considerando que el porcentaje de valores ausentes es muy reducido, se procederá a trabajar con los datos en su estado original.

3.3.2. Valores atípicos (Outliers)

Para identificar la presencia de posibles valores atípicos, se ha optado por utilizar diagramas de cajas para cada variable numérica. Este método permite visualizar de forma clara la distribución de los datos y detectar posibles observaciones que se encuentren fuera de los rangos esperados.

A continuación, en los conjuntos de datos relacionados con las distribuciones, se identifican valores atípicos (ver Figura 3.1). Sin embargo, estos se consideran válidos para nuestro estudio, ya que responden a una concentración natural de las calificaciones en los intervalos intermedios (por ejemplo, *De 6 a 7* o *De 7 a 8*) y no en los extremos.

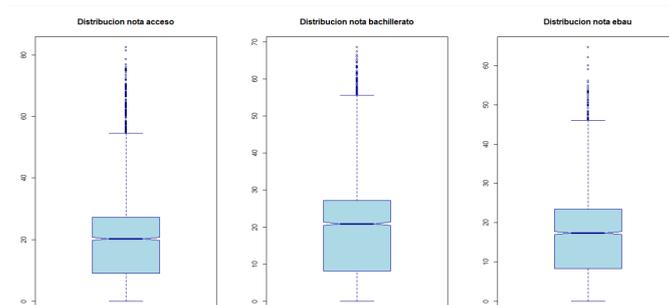


Figura 3.1: Diagramas de cajas para la búsqueda de valores atípicos en las distribuciones de calificaciones

Por otro lado, se presentan los diagramas de cajas correspondientes a los diez conjuntos de datos, centrados en sus respectivas variables numéricas. Se observa que en la Figura 3.2, relativa a los conjuntos de datos de notas por sexo y por centro, no existen valores atípicos destacables.

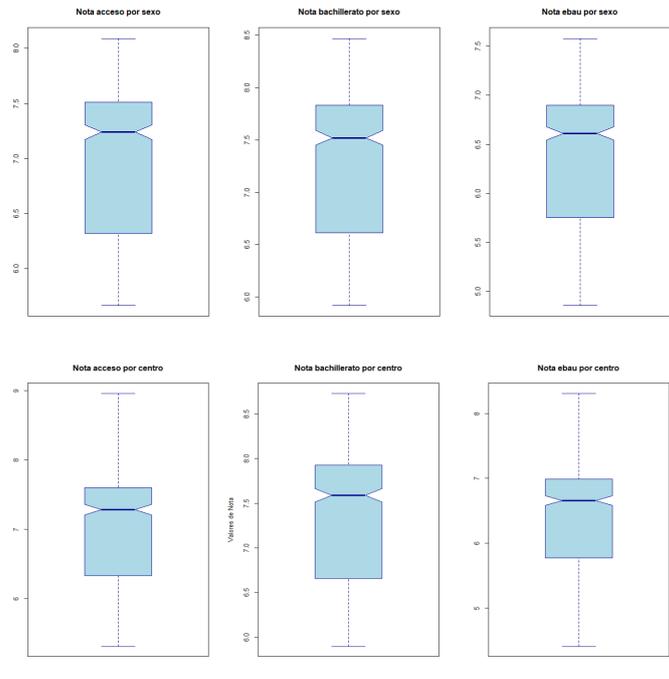


Figura 3.2: Diagramas de cajas para la búsqueda de valores atípicos en notas por sexo y por centro

Finalmente, en el conjunto de datos de matriculados, presentados y aprobados, también se detectan valores atípicos (ver Figura 3.3) . Esta situación se explica por las diferencias en la población de las distintas comunidades autónomas, lo que provoca que la variable *Cifra* adopte valores significativamente mayores en las comunidades más grandes y más bajos en las más pequeñas.

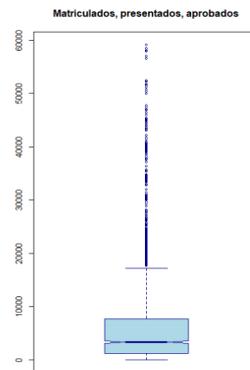


Figura 3.3: Diagramas de cajas para la búsqueda de valores atípicos en matriculados, presentados y aprobados

Dado que estos valores atípicos no corresponden a errores de medición, se ha decidido mantenerlos en los conjuntos de datos empleados para el análisis.

3.4. Transformación de los datos

Como se mencionó en la sección anterior, se ha optado por no realizar transformaciones en los datos respecto a valores ausentes o valores atípicos (outliers). Sin embargo, se han aplicado cuatro transformaciones para la obtención de nuevos conjuntos de datos a partir de los iniciales.

La primera transformación consiste en la reasignación de los nombres de las comunidades autónomas, ajustándolos a un formato más adecuado y consistente para todos los conjuntos de datos.

La segunda transformación consiste en la generación de dos nuevos conjuntos de datos. En el primero de ellos, se calcula el porcentaje de aprobados respecto a los presentados. Este conjunto de datos está compuesto por las columnas *Sexo*, *Comunidad*, *Período* y *Convocatoria*, que fueron previamente descritas. Además, se incluyen dos nuevas variables: *Categoría*, que puede tomar los valores 'Porcentaje_Aprobados' o 'Porcentaje_No_Aprobados', y *Cifra*, que representa el porcentaje de aprobados o no aprobados.

El segundo conjunto de datos se ha generado calculando el porcentaje de presentados respecto a los matriculados. De manera similar al conjunto anterior, contiene las variables *Sexo*, *Comunidad*, *Período* y *Convocatoria*, y además incorpora la variable *Categoría*, que toma los valores 'Porcentaje_Presentados' o 'Porcentaje_No_Presentados', y la variable *Cifra*, que refleja el porcentaje de presentados o no presentados.

En la Figura 3.4 se muestra el cálculo del porcentaje para uno de los conjuntos de datos. A partir del conjunto denominado *Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo*, se han seleccionado las categorías de *presentados* y *matriculados* para cada comunidad autónoma.

```
# Calcular porcentaje de Presentados respecto a Matriculados
calcular_porcentaje_presentados <- reactive({
  # Filtrar los datos de Presentados y Matriculados
  matriculados <- datos_filtrados() %>% filter(Indicador == "Matriculados")
  presentados <- datos_filtrados() %>% filter(Indicador == "Presentados")

  # Unir los dos datos por Comunidad
  df_combinado <- left_join(matriculados, presentados, by = c("Comunidad", "Sexo"), relationship = "many-to-many", suffix = c("_mat", "_pre"))

  # Calcular el porcentaje de presentados respecto a los matriculados
  df_combinado <- df_combinado %>%
    mutate(Porcentaje_Presentados = (Cifra_present / Cifra_mat) * 100) %>%
    select(Comunidad, Sexo, Porcentaje_Presentados, Cifra_present, Cifra_mat)

  # Calcular los porcentajes
  df_combinado <- df_combinado %>%
    mutate(Porcentaje_Presentados = (Cifra_present / Cifra_mat) * 100,
           Porcentaje_No_Presentados = 100 - Porcentaje_Presentados) %>% # Calculamos los no aprobados
    select(Comunidad, Sexo, Porcentaje_Presentados, Porcentaje_No_Presentados) # Conservamos solo columnas necesarias

  # Convertir a formato largo para ggplot
  df_long <- df_combinado %>%
    pivot_longer(cols = c(Porcentaje_Presentados, Porcentaje_No_Presentados),
                 names_to = "Categoría",
                 values_to = "Porcentaje")
})
```

Figura 3.4: Cálculo conjunto de datos porcentaje aprobados

La tercera transformación consiste en la generación de un nuevo conjunto de datos a partir del conjunto original *Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo*, con el objetivo de representar la información en un mapa de las comunidades autónomas de España.

Como se puede observar en la Figura 3.5, en primer lugar, se obtiene el mapa cartográfico base de las comunidades autónomas. A continuación, se construye una tabla de correspondencias que vincula los nombres de las comunidades autónomas presentes en los datos originales con sus respectivos códigos ISO oficiales. Posteriormente, se realiza la combinación de los datos propios con esta tabla de equivalencias, y se integran con el objeto geoespacial del mapa.

Finalmente, estos datos transformados están listos para ser utilizados en la generación de un gráfico tipo mapa que permita visualizar la información de forma geográfica.

```

#datos para generar mapa

# Obtener mapa de comunidades
mapa_ccaa <- esp_get_ccaa()
mapa_ccaa$iso2.ccaa.code

# Tabla con correspondencias entre tus datos y los de mapSpain
equivalencias <- data.frame(
  Comunidad = c("Andalucía", "Aragón", "Asturias", "Islas Baleares", "Islas Canarias", "Cantabria", "Castilla la Mancha", "Castilla y León",
    "Extremadura", "Galicia", "Madrid", "Región de Murcia", "Navarra", "País Vasco", "La Rioja"),

  ccaa.code = c("ES-AN", "ES-AR", "ES-AS", "ES-IB", "ES-CN", "ES-CB", "ES-CM", "ES-CL", "ES-CT", "ES-VC", "ES-EX", "ES-GA", "ES-MD", "ES-M",
    "ES-PV", "ES-RI" )
)

df_mapa <- df %>%
  left_join(equivalencias, by = "Comunidad")

# Unir datos al mapa
df_mapa <- mapa_ccaa %>%
  left_join(df_mapa, by = c("iso2.ccaa.code" = "ccaa.code"))

```

Figura 3.5: Cálculo conjunto de datos para la representación del mapa

Por último, se ha calculado un conjunto de datos específico para la representación en el mapa, incorporando los porcentajes de aprobados respecto a presentados y de presentados respecto a matriculados. Para obtener el conjunto de datos final, se aplicaron los procedimientos descritos previamente para el cálculo de los porcentajes, y posteriormente se generaron los datos incluyendo las comunidades autónomas y sus correspondientes códigos de equivalencia. Este conjunto de datos se utiliza para una visualización clara y comparativa de los resultados en el mapa geográfico.

Capítulo 4

Recursos empleados

En este capítulo se describen brevemente las herramientas utilizadas durante la realización de este Trabajo de Fin de Grado.

4.1. R

R es un entorno de software ampliamente utilizado para la manipulación de datos, cálculos estadísticos y generación de gráficos. Se considera una reimplementación del lenguaje S, desarrollado en AT&T por Rick Becker, John Chambers y Allan Wilks[4]. Entre sus principales características se destacan las siguientes:

- Almacenamiento y gestión eficiente de datos.
- Operadores especializados para el cálculo con estructuras indexadas, como matrices y arreglos (arrays).
- Una amplia, coherente e integrada colección de herramientas para el análisis de datos.
- Capacidades gráficas avanzadas para la visualización de datos, tanto en pantalla como en dispositivos de impresión.
- Un lenguaje de programación bien estructurado, intuitivo y eficiente, que incluye estructuras de control como condicionales, bucles, funciones recursivas y gestión de entradas y salidas.

Además, R constituye una plataforma clave para el desarrollo de nuevos métodos de análisis interactivo de datos. Su evolución constante lo convierte en un software dinámico y en permanente actualización.

Aunque R puede considerarse un lenguaje de programación, los scripts desarrollados en este entorno deben entenderse como herramientas flexibles y adaptables, en lugar de soluciones rígidas y definitivas.

A pesar de que R no se define exclusivamente como un sistema estadístico, es ampliamente utilizado con este propósito. Más que un software estadístico, se concibe como un entorno que integra una amplia variedad de técnicas estadísticas, tanto clásicas como modernas. Algunas de estas herramientas están incluidas en la versión base de R, mientras que muchas otras se encuentran disponibles en forma de bibliotecas o packages, accesibles a través de CRAN (Comprehensive R Archive Network).

A diferencia de otros programas estadísticos como SAS o SPSS, que generan automáticamente salidas detalladas para cada análisis, R sigue un enfoque más modular. Los análisis se desarrollan en una secuencia de pasos, almacenando los resultados intermedios en objetos que pueden ser examinados y procesados según sea necesario. Este enfoque proporciona mayor flexibilidad y control sobre el proceso analítico, permitiendo una exploración de datos más estructurada y personalizada.

En el momento de escribir esta memoria la última versión estable de R es la 4.4.2, denominada "Pile of Leaves", lanzada el 31 de octubre de 2024 pero para este trabajo se ha utilizado la versión 4.4.0

4.2. RStudio

RStudio es el entorno de desarrollo integrado (IDE) gratuito y de código abierto más utilizado para trabajar con R. En términos generales, un IDE es un software que integra diversas herramientas para facilitar la programación y el análisis de datos [5].

Específicamente, un IDE está compuesto por cuatro elementos fundamentales que se muestran en la Figura 4.1 y que pasaremos a describir a continuación [6]:

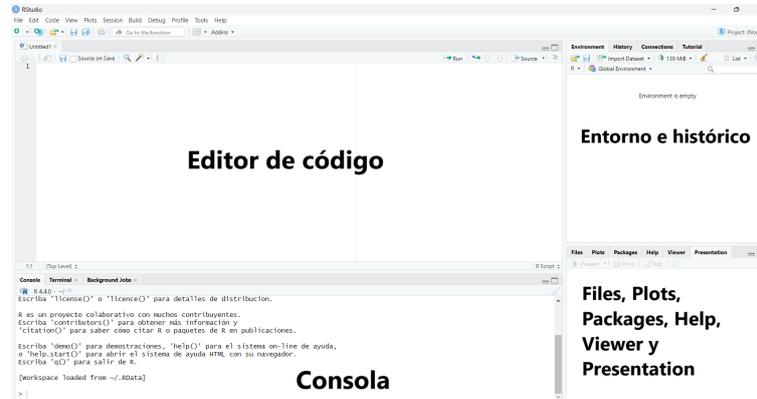


Figura 4.1: Interfaz de RStudio

- **Editor de código:** en este entorno se edita la sintaxis del código antes de su ejecución, permitiendo a los usuarios desarrollar y gestionar sus scripts de manera eficiente ya que permite escribir instrucciones u órdenes de forma organizada, sin aplicar formatos, es decir, permite la creación de archivos en formato de texto plano. Además, facilita la visualización y edición de diversos archivos relacionados con el código.
- **Entorno e histórico:** en la pestaña entorno se presentan los conjuntos de datos y los distintos objetos generados, incluyendo resultados, variables y gráficos, a medida que se ejecutan los análisis. En la pestaña histórico se muestran los comandos que han sido ejecutados en la sesión actual de trabajo.
- **Consola:** representa la versión fundamental del software R, en la cual se procesan y ejecutan las instrucciones introducidas a través del editor de sintaxis. También, muestra los resultados generados tras la ejecución del código de los distintos scripts.
- **Files, Plots, Packages, Help, Viewer y Presentation:** la pestaña *Files* permite visualizar el histórico de archivos trabajados con el programa e indica el directorio de trabajo. La pestaña *Plots* muestra la visualización de los diferentes gráficos generados. La pestaña *textitPackages* permite visualizar los paquetes descargados y almacenados en el disco duro, así como gestionar su instalación y actualización. La pestaña *Help* permite acceder a la página oficial del software que proporciona recursos como manuales, cursos, información general y gestión de paquetes. Incluye un buscador que permite acceder rápidamente a la documentación de los paquetes instalados sin necesidad de conexión a Internet. La ventana *Viewer* muestra los resultados interactivos generados al crear reportes utilizando funcionalidades como RMarkdown. La pestaña *Presentation* se utiliza para visualizar diapositivas en formato HTML generadas mediante el paquete *revealjs*.

4.3. Shiny

Shiny es un framework desarrollado por RStudio para la creación de aplicaciones web interactivas. En la práctica, opera como un paquete adicional de R, similar a dplyr o ggplot2, pero con una interfaz específica para la visualización de resultados permitiendo construir aplicaciones web interactivas a partir de los scripts de R [7].

Asimismo, Shiny permite su integración con lenguajes de desarrollo web como CSS y JavaScript, lo que facilita la mejora tanto estética como funcional de la aplicación. Esta combinación entre paquetes interactivos y una arquitectura orientada a la interactividad proporciona un enfoque eficaz para la creación de visualizaciones con un nivel de dinamismo y versatilidad comparable al de un cuadro de mando en Power BI.

En el siguiente capítulo explicamos en detalle el funcionamiento y estructura de una aplicación Shiny.

4.4. ggplot2

ggplot es la abreviatura de *Grammar of Graphics plot*. Este paquete forma parte del ecosistema de herramientas incluidas en la librería *tidyverse*, que engloba, entre otras, *ggplot2*, *dplyr*, *tidyr*, *readr*, *purrr*, *tibble*, *forcats* y *lubridate* [8].

Su principal objetivo es integrar la creación de gráficos como un componente fundamental en el análisis y modelado de datos, siguiendo el enfoque denominado *Gramática de los Gráficos*, desarrollado por Leland Wilkinson. Este enfoque permite combinar diversos elementos dentro de un gráfico en forma de capas superpuestas, facilitando su vinculación con los datos. Además, posibilita el mapeo de atributos estéticos, como el color, la forma y el tamaño, así como la representación de objetos geométricos, tales como puntos, barras o líneas.

No todas las capas son necesarias para construir una visualización en *ggplot2* pero cada una de ellas contribuye al gráfico final aportando elementos diferenciados. A continuación, se describen las principales capas y su función:

- Data: constituye la primera capa necesaria para generar la visualización. Representa la base fundamental sobre la cual se construye el gráfico y, por tanto, es una parte obligatoria en la estructura de cualquier visualización en este paquete. Esta capa hace referencia al conjunto de datos, que debe estar contenido en un objeto de clase *data.frame* o *tibble*.
- Mapeado de elementos (aesthetics): esta capa define las variables del *data.frame* que se asignarán a los distintos elementos del gráfico, como los ejes (x, y), el color, el tamaño o la forma. Es esencial, ya que establece el vínculo entre los datos y su representación visual. En *ggplot2*, esta capa se denomina *aes*.
- Elementos geométricos (geom): esta capa le indica a R el tipo de visualización que se desea, es decir, la geometría a utilizar para mostrar los datos, como barras, líneas, puntos, histogramas, entre otros. También es una capa obligatoria y constituye el tercer pilar indispensable.
- Componentes gráficas (facet): permite dividir un gráfico en subgráficos (facetas) según una variable. No constituye una capa obligatoria.
- Transformaciones estadísticas (stat): permite agregar información estadística que resume los datos, como una línea de tendencia. No es obligatoria.
- Sistema de coordenadas (coord): define cómo se combinan las variables de los ejes x e y según la estética, determinando el espacio donde se representan los datos. Permite, por ejemplo, fijar límites, intercambiar ejes o igualar sus magnitudes.
- Temas (theme): define la apariencia final del gráfico, incluyendo fondo, ejes, tamaños, grids y posiciones de los nombres. *ggplot2* ofrece temas predeterminados, y es posible crear o incorporar otros mediante paquetes adicionales. No es obligatoria.

4.5. Plotly

El paquete *plotly* ofrece una poderosa herramienta para transformar gráficos estáticos elaborados con *ggplot2* en visualizaciones interactivas, lo que permite mejorar significativamente la exploración y comunicación de los datos [9]. Esta interactividad se traduce en la posibilidad de desplazarse por el gráfico, hacer zoom, mostrar etiquetas emergentes al pasar el cursor sobre los elementos y seleccionar subconjuntos de datos de manera dinámica, entre otras funcionalidades.

Capítulo 5

Estructura y funcionamiento de una aplicación Shiny

En el capítulo anterior se presentó una descripción general del paquete Shiny. En este capítulo, se aborda con mayor detalle la estructura y el funcionamiento de una aplicación desarrollada con Shiny, analizando tanto sus componentes principales como su interacción para lograr un correcto desempeño.

Una aplicación desarrollada con Shiny se estructura fundamentalmente en dos componentes principales: la interfaz de usuario y el motor del servidor. Esta división corresponde al modelo comúnmente conocido en el ámbito del desarrollo web como front-end y back-end [10].

Para desarrollar un aplicación con Shiny primero se descarga e instala el paquete desde el repositorio oficial de R y después se carga el paquete en la sesión actual de R, permitiendo acceder a todas sus funciones.

5.1. Estructura de una aplicación Shiny

Una aplicación Shiny consta de al menos dos archivos [11]:

- **UI:** script para la interfaz usuario, está orientado a la construcción de los elementos gráficos e interactivos que permiten al usuario explorar y manipular los datos. Este componente es responsable de recibir las entradas (*inputs*) proporcionadas por el usuario y mostrar las salidas (*outputs*) generadas por el servidor. En términos generales, constituye el front-end de la aplicación.
- **Server:** es el motor del servidor, responsable de gestionar todos los procesos relacionados con el tratamiento, análisis y preparación de la información. Su función principal es realizar los cálculos necesarios y ejecutar la lógica del programa. Constituye el back-end de la aplicación, obteniendo las entradas del usuario (*inputs*) y enviando los resultados (*outputs*) a la interfaz de usuario para su visualización.

Esta división en dos componentes, *ui* y *server*, permite una organización clara y modular del código, lo que facilita tanto el desarrollo como el mantenimiento de las aplicaciones interactivas. Al separar las responsabilidades de la interfaz de usuario y la lógica del servidor, se mejora la comprensión del flujo de datos y se hace más sencillo realizar modificaciones y actualizaciones sin afectar a otras partes del sistema [12].

En la Figura 5.1 se muestra un esquema de una aplicación Shiny, destacando su división en dos componentes fundamentales: **la interfaz de usuario (UI)**, que define la apariencia y disposición de la aplicación, y el **servidor (Server)**, que se encarga de la lógica y el procesamiento de los datos necesarios para el funcionamiento de la aplicación.

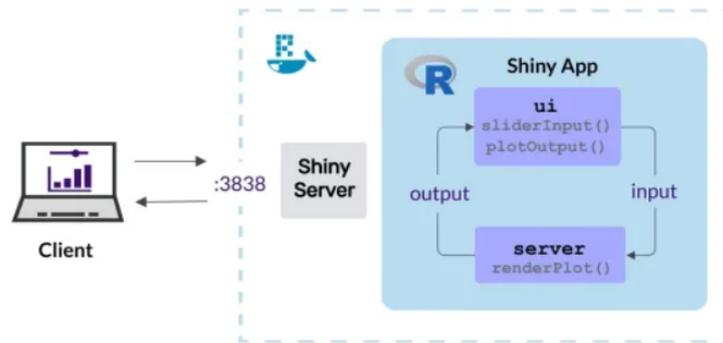


Figura 5.1: Esquema de una aplicación Shiny: interfaz de usuario (UI) y servidor (Server)

En la Figura 5.2 se muestra el código de una aplicación Shiny junto con el resultado de su ejecución. Se puede observar que, en primer lugar, se carga la librería shiny en ambos scripts. Posteriormente, en el componente UI se define la apariencia de la aplicación, mientras que en el componente Server se implementa toda la lógica del programa.

```

library(shiny)

# UI
shinyUI(fluidPage(

  # Título de la aplicación
  titlePanel("Los histogramas"),

  # Columna lateral con una barra deslizable para el número de
  # intervalos
  sidebarLayout(
    sidebarPanel(
      sliderInput("bins",
                 "Número de intervalos:",
                 min = 1,
                 max = 50,
                 value = 30)
    ),

    # Panel principal con el gráfico generado
    mainPanel(
      plotOutput("distPlot")
    )
  )
)

```

(a) Componente Ui

```

library(shiny)

# SERVER
shinyServer(function(input, output) {

  # La expresión que obtiene el histograma es
  # un renderPlot que quiere decir que:
  #
  # 1) Es "reactiva" y debe re-ejecutarse automáticamente
  # cuando el input cambie.
  # 2) El output es un gráfico.

  output$distPlot <- renderPlot({
    x <- faithful[, 2] # Old Faithful Geyser data
    bins <- seq(min(x), max(x), length.out = input$bins + 1)

    # Dibujar el histograma con dicho número de intervalos
    hist(x, breaks = bins, col = 'darkgray', border = 'white',
         main='Histograma')
  })
})

```

(b) Componente Server

Figura 5.2: Componentes principales de una aplicación Shiny

En este ejemplo concreto, el resultado es un histograma interactivo en el que el usuario puede seleccionar el número de intervalos mediante un control deslizante (*slider*), como se puede observar en la Figura 5.3.

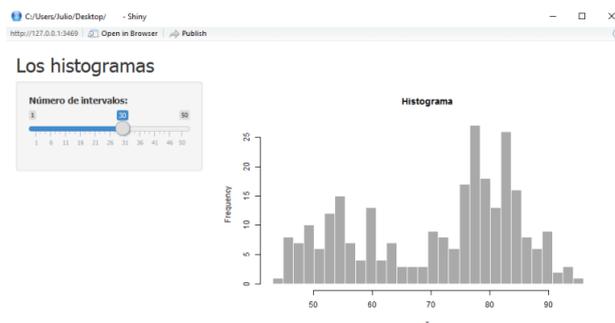


Figura 5.3: Resultado de la ejecución de la aplicación Shiny

La orden utilizada para unir las dos componentes principales de la aplicación es:

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

Sin embargo, cabe destacar que el código completo de la aplicación Shiny puede integrarse dentro de un único script, combinando la definición de la interfaz de usuario (*UI*) y del servidor (*Server*) en un mismo archivo, como se muestra a continuación:

```
1 library(shiny)
2
3 # UI
4 u <- shinyUI(TipoDePagina(
5   # elementos de la interfaz
6 ))
7
8 # SERVER
9 s <- shinyServer(function(input, output) {
10   # logica del servidor
11 })
12
13 # Lanzar la aplicacion
14 shinyApp(ui = u, server = s)
```

Listing 5.1: Código en R

Para este Trabajo de Fin de Grado se ha optado por la opción de integrar tanto la interfaz de usuario (*UI*) como el servidor (*Server*) en un único script. Esta decisión permite simplificar la estructura del código y facilita su mantenimiento y comprensión durante el desarrollo de la aplicación.

5.1.1. UI. Inputs y outputs

Como se ha mencionado anteriormente, las aplicaciones Shiny se conforman por elementos de entrada y elementos de salida. A continuación, se procede a detallar cada uno de ellos [13]:

- **Inputs:** todas las funciones que definen elementos de entrada comparten un primer argumento denominado *inputId*. Este argumento actúa como identificador, permitiendo conectar la interfaz de usuario con la lógica de servidor. Así, si en la interfaz se ha definido un elemento con `inputId = "name"`, el servidor podrá acceder a su valor mediante `input$name`.

El uso de `inputId` está sujeto a dos condiciones fundamentales:

- Debe consistir únicamente en letras, números y guiones bajos, sin admitir espacios, guiones, puntos u otros caracteres especiales. En resumen, debe seguir las reglas con las que se define una variable en R.
- Debe ser único dentro de la aplicación. En caso de duplicidad, el servidor no podrá identificar correctamente a qué control se hace referencia, lo que puede provocar errores en la ejecución.

La mayoría de las funciones de entrada incluyen además un segundo argumento denominado `label`, destinado a proporcionar un rótulo legible por el usuario. Aunque **Shiny** no establece restricciones específicas sobre este texto, resulta recomendable seleccionarlo cuidadosamente para garantizar que la aplicación sea comprensible y fácil de usar.

El tercer argumento habitual es `value`, el cual, siempre que sea posible, permite asignar un valor predefinido al control. Los argumentos adicionales varían según el tipo específico de entrada y las funcionalidades que esta ofrezca.

En la Figura 5.4 se muestran los ejemplos más representativos de inputs

- **Outputs:** actúan como marcadores de posición que posteriormente son completados por la función del servidor. Al igual que los elementos de entrada, cada elemento de salida requiere un identificador único como primer argumento. Por ejemplo, si en la interfaz se define un elemento de salida con ID = "plot", este podrá ser accedido desde la función del servidor mediante `output$plot`.

Cada función de salida en el frontend está asociada a una función de renderizado (*render function*) en el backend. En términos generales, existen tres tipos principales de salidas, que corresponden a los componentes habitualmente presentes en un informe: texto, tablas y gráficos.

El listado de los outputs mas representativos se observa en la Figura 5.5.

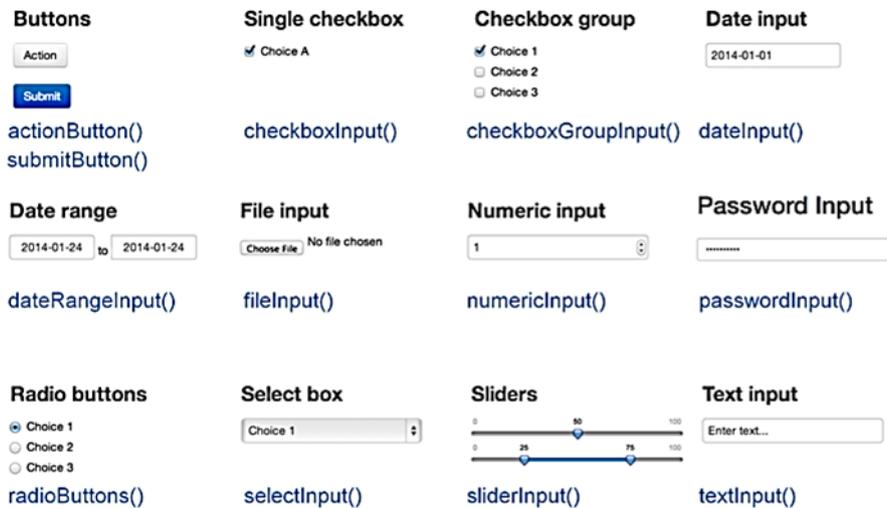


Figura 5.4: Listado principales Inputs

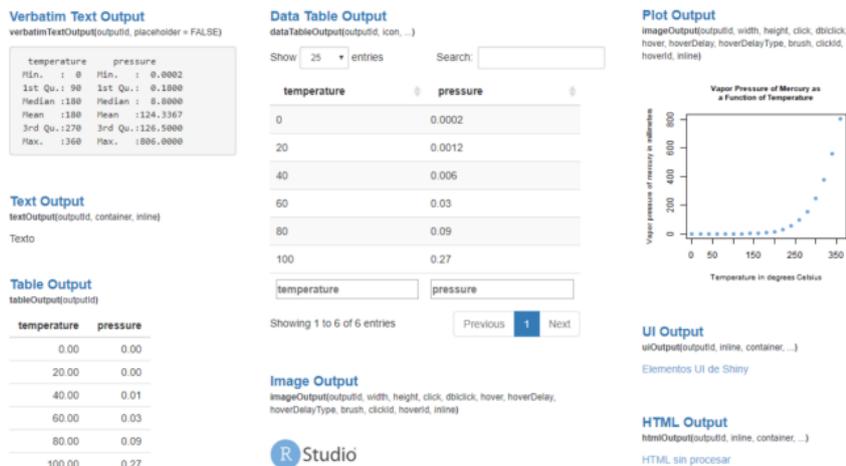


Figura 5.5: Listado principales Outputs

5.1.2. Server. Conexión entre inputs y outputs

Los outputs se especifican en el servidor (server), mientras que en la interfaz de usuario (UI) se configura cómo se visualizan. Para definir los parámetros de salida, es necesario seguirlos siguientes principios [13]:

1. **Definir los outputs:** Se debe definir cada uno de los outputs utilizando la sintaxis `output$nombre_output`. Esto permite especificar qué componentes serán visibles en la interfaz de usuario, como gráficos, textos o tablas.

2. **Conectar los inputs con los outputs:** Es necesario vincular los inputs con los outputs mediante las funciones `render*`, como `renderPlot`, `renderText`, `renderTable`, entre otras. Dentro de estas funciones, los valores de los inputs deben ser referenciados utilizando la sintaxis `input$nombre_del_input`. De este modo, el servidor podrá generar los outputs adecuados en función de los valores actuales de los inputs. Esta relación de dependencia puede observarse en la Figura 5.6.

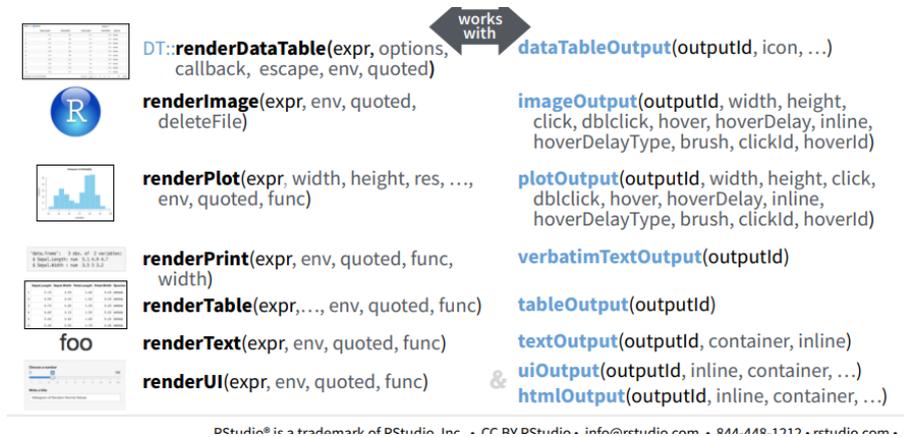


Figura 5.6: Relación entre Outputs y funciones de renderización

3. **Aprovechar la reactividad:** El sistema reactivo de Shiny es crucial para asegurar que los outputs se actualicen de manera automática cada vez que los valores de los inputs cambian. Este mecanismo permite que los componentes de salida se ajusten en tiempo real a las interacciones del usuario, sin necesidad de intervención manual. Para lograrlo, se accede a los valores de los inputs mediante la sintaxis `input$nombre_del_input`.

5.2. Programación reactiva

Shiny utiliza el paradigma de programación reactiva [14], un enfoque que permite que los outputs se actualicen automáticamente cuando los inputs cambian. Esto implica que cualquier componente que dependa de una variable reactiva será recalculado cada vez que dicha variable se modifique. En Shiny, todos los inputs son gestionados de forma reactiva, lo que facilita la creación de aplicaciones dinámicas y responsivas, sin necesidad de intervención manual para actualizar los resultados.

A continuación, se describen los tres tipos principales de objetos involucrados en la programación reactiva de Shiny [15]:

- **Fuentes reactivas:** Se refiere a las entradas proporcionadas por el usuario a través de la interfaz (UI). El servidor Shiny tiene acceso a estas entradas mediante los widgets programados, y cada vez que sus valores cambian, dicha información es transmitida automáticamente al servidor.
- **Conductores reactivos:** Son objetos que existen exclusivamente dentro del servidor Shiny. Aunque no son necesarios en aplicaciones simples, permiten generar objetos internos que pueden ser utilizados en otras operaciones dentro del servidor. Generalmente, estos conductores dependen de las fuentes reactivas.
- **Puntos finales:** Son las salidas generadas por el servidor y enviadas a la interfaz de usuario. Por ejemplo, en una aplicación que produce una gráfica, el punto final sería el gráfico renderizado que se presenta al usuario.

5.3. Diseño y Organización de Aplicaciones Shiny

El diseño de la interfaz de usuario de una aplicación Shiny se basa en diferentes estructuras que permiten organizar los elementos de manera eficiente y adaptable a distintos contextos. Existen principalmente tres tipos de estructuras que pueden emplearse [16]:

- `pageWithSidebar` (obsoleto): Fue una de las primeras formas de estructurar aplicaciones Shiny, dividiendo la página en una barra lateral y un panel principal. Sin embargo, este enfoque ha quedado obsoleto y ya no se recomienda, ya que presenta limitaciones en comparación con las opciones más actuales.
- `fluidPage`: Es la opción más recomendada en la actualidad. Permite construir interfaces dinámicas y responsivas que se adaptan automáticamente al tamaño de la ventana del navegador. Gracias a su flexibilidad, es ideal para aplicaciones que deben ser visualizadas en diferentes dispositivos, como computadoras, tablets o teléfonos móviles.
- `fixedPage`: Similar a `fluidPage`, pero con la diferencia de que establece un ancho fijo para la página, sin importar el tamaño de la ventana del navegador. Esto puede ser útil cuando se desea mantener un formato constante y controlado en el diseño, evitando que los componentes se redimensionen.

Independientemente de la estructura elegida, toda aplicación Shiny sigue un esquema común que se organiza a través de componentes básicos que definen su contenido y distribución [17]:

- El componente `titlePanel` se utiliza para especificar el título principal de la aplicación. Este título generalmente se coloca en la parte superior de la interfaz y tiene como objetivo proporcionar contexto al usuario, permitiéndole identificar rápidamente el propósito de la aplicación. Es un elemento estático que no cambia con las interacciones del usuario, sirviendo como punto de referencia para el usuario.
- El componente `layout` organiza y estructura los elementos en la interfaz de usuario de manera eficiente. A continuación, se describen algunos de los principales layouts disponibles:
 - `sidebarLayout`: Es un contenedor que organiza la disposición de la página en dos áreas principales. La barra lateral (`sidebarPanel`) donde se colocan los controles de entrada que permiten al usuario interactuar y modificar los parámetros de análisis o visualización y el panel principal (`mainPanel`) destinado a mostrar los resultados derivados de las interacciones del usuario, tales como gráficos, tablas, textos interpretativos, mapas, etc
 - `fluidRow`: Es un contenedor flexible que organiza los elementos en una fila horizontal adaptable al ancho de la ventana del navegador. Dentro de cada `fluidRow`, se pueden incluir columnas para dividir el contenido en secciones proporcionadas.
 - `flowLayout`: Organiza varios elementos de forma automática en filas consecutivas, alineándolos horizontalmente. Si el espacio disponible no es suficiente, los elementos se reubican automáticamente en la siguiente línea, adaptándose dinámicamente al tamaño de la pantalla.
 - `splitLayout`: Distribuye varios objetos en columnas de igual ancho dentro de un solo contenedor. Esta función es útil para mostrar elementos alineados uno al lado del otro, como gráficos, tablas o cuadros de texto, garantizando un reparto uniforme del espacio disponible.
 - `verticalLayout`: Organiza los elementos de forma vertical, ubicándolos uno debajo del otro en una secuencia lineal. Este diseño es sencillo y resulta adecuado cuando se desea presentar los componentes de manera ordenada a lo largo de la página, ocupando todo el ancho disponible.

En la Figura 5.7 se muestran los distintos tipos de layouts mencionados anteriormente.

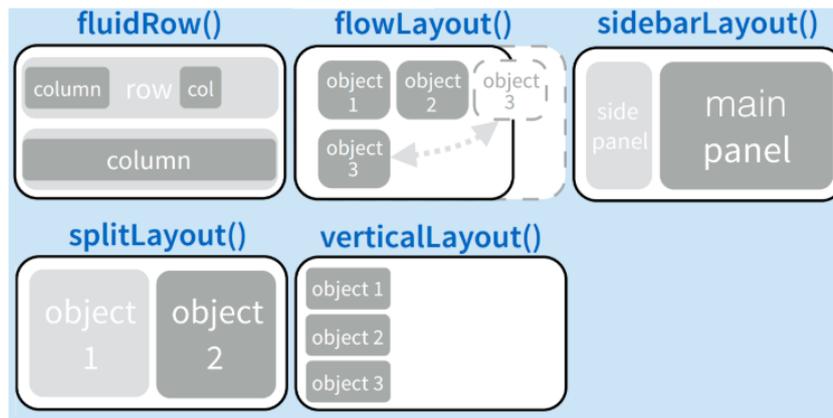


Figura 5.7: Ejemplo principales layouts

La correcta definición y combinación de estas estructuras y componentes permite desarrollar aplicaciones Shiny intuitivas, dinámicas y bien organizadas, facilitando tanto la interacción del usuario como la presentación clara de los resultados.

5.4. Paquete *Shinydashboard*

En esta sección se presenta el paquete *shinydashboard*, una biblioteca complementaria de R diseñada para ampliar las capacidades estándar de Shiny [18]. Este paquete permite desarrollar aplicaciones web interactivas con una interfaz tipo *dashboard* o tablero de control, ofreciendo una disposición visual moderna, estructurada y eficiente. *Shinydashboard* facilita la creación de aplicaciones que integran múltiples secciones, menús de navegación y elementos visuales, permitiendo una organización clara del contenido. Además, destaca por su capacidad de personalización, ya que admite la incorporación de temas y estilos CSS, lo que facilita adaptar la apariencia del tablero a las necesidades y características específicas de cada proyecto.

La estructura principal de una aplicación desarrollada con *shinydashboard* se establece a través de la función `dashboardPage()`, la cual organiza y coordina los componentes fundamentales que conforman la interfaz [19].

```
dashboardPage(
  dashboardHeader(),
  dashboardSidebar(),
  dashboardBody()
)
```

En esta función podemos observar tres componentes fundamentales:

- Barra lateral (*sidebar*): Es el área donde se colocan los menús de navegación y los controles de entrada para interactuar con la aplicación. Generalmente, un *sidebar* se utiliza para una navegación rápida y puede contener elementos como `menuItem`s que funcionan como pestañas (similar a `tabPanel`), así como entradas de Shiny, por ejemplo deslizadores (*sliders*) y campos de texto.
- Encabezado superior (*header*): Es la parte superior de la aplicación que permite mostrar el título principal, iconos, mensajes o notificaciones. El encabezado contribuye a dar identidad a la aplicación y puede incorporar menús desplegables (*dropdowns*). Un ejemplo básico sería:

```
dashboardHeader(title = "My Dashboard")
```

- Cuerpo principal (*body*): Es el área donde se presentan las salidas principales de la aplicación, tales como gráficos, tablas, cuadros de texto, indicadores u otras visualizaciones. Estos elementos se organizan en pestañas, filas o cajas. Aunque el *body* puede contener cualquier contenido estándar de Shiny, lo habitual

en un *dashboard* es utilizar estructuras más organizadas, siendo el bloque básico el *box*, que a su vez puede incluir cualquier tipo de contenido.

En la Figura 5.8 se muestra de forma visual lo explicado anteriormente, permitiendo ver claramente cómo se organiza la estructura de una aplicación con *shinydashboard*.

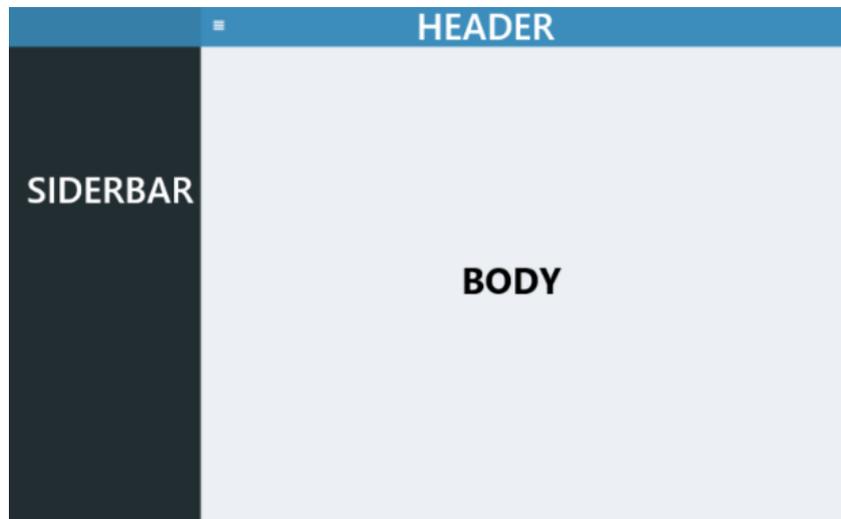


Figura 5.8: Estructura aplicación diseñada con Shinydashboard

Gracias a esta organización modular, *shinydashboard* permite construir aplicaciones con un diseño claro, jerárquico y altamente personalizable, ideal para reportes interactivos, paneles de monitoreo y aplicaciones analíticas.

Capítulo 6

VisionEbau: aplicación Shiny de visualización de datos de EBAU

En este capítulo se presenta VisionEbau, una aplicación desarrollada en Shiny orientada a la visualización interactiva de datos relacionados con la Evaluación del Bachillerato para el Acceso a la Universidad (EBAU).

La aplicación muestra información detallada sobre el número de matriculados, presentados y aprobados, así como sobre la nota media de Bachillerato, la calificación obtenida en la prueba de la EBAU, y la nota de acceso a la universidad. Además, se presenta la distribución por intervalos de las distintas calificaciones. Todos estos datos se encuentran desglosados por comunidad autónoma, sexo, convocatoria, tipo de centro educativo y año.

El acceso a la aplicación está disponible en el siguiente enlace: <http://shiny1.eio.uva.es:3838/aranzadi/visionebau/>

Resaltar que los gráficos mostrados en la aplicación Shiny son de carácter interactivo: al pasar el cursor sobre ellos se visualizan los valores correspondientes, y además permiten acciones como ampliación y alejamiento para una mejor exploración de los datos. En el Anexo C se han incluido capturas complementarias que ilustran la interfaz de la aplicación.

6.1. Página de *Inicio*

En la Figura 6.1 se muestra la página de *Inicio*, donde se presenta una breve descripción de los contenidos de la aplicación.



Figura 6.1: Página de Inicio

6.2. Menú de navegación

En la Figura 6.2 se muestra el menú de navegación de la aplicación, el cual se presenta como un panel lateral situado en la parte izquierda de la pantalla. Este menú es ocultable mediante la selección del icono de las tres líneas horizontales, lo que permite ampliar el área de visualización del panel principal, donde se presentan las tablas y los gráficos. Con el fin de mejorar la experiencia del usuario y facilitar la identificación visual de las diferentes secciones, se han incorporado iconos representativos en cada una de las pestañas.

En el menú de navegación se pueden encontrar nueve pestañas, cada una acompañada de su respectivo icono representativo. A continuación, se describen brevemente los contenidos de cada una de ellas:

- **Inicio:** muestra una breve descripción del contenido y funcionalidad general de la aplicación.
- **Resultados por Comunidad:** presenta información sobre los estudiantes matriculados, presentados y aprobados en la prueba de acceso a la universidad (EBAU) para cada comunidad autónoma. El usuario puede seleccionar una comunidad específica y observar la evolución de estos indicadores a lo largo de los años. Además, es posible comparar los datos entre comunidades en función del año, la convocatoria y el sexo seleccionados. Más información sobre esta sección se puede encontrar en la Sección 6.3.
- **Nota media de Bachillerato:** ofrece información sobre la evolución de la nota media de Bachillerato. El usuario puede seleccionar una comunidad y visualizar su evolución por sexo o por tipo de centro. También se proporciona la opción de comparar varias comunidades simultáneamente. Tablas detalladas acompañan a los gráficos. Para más detalles, véase la Sección 6.4.
- **Distribución de la nota de Bachillerato:** muestra gráficos que representan la distribución de estudiantes por intervalos de nota media de Bachillerato, diferenciados por sexo. También se incluye una tabla con los valores numéricos utilizados en los gráficos. Más información en la Sección 6.5.
- **Nota de EBAU:** presenta la evolución de la calificación obtenida en la prueba de EBAU, permitiendo filtrar por comunidad, sexo y centro. Al igual que en las pestañas anteriores, se incluyen gráficos y tablas que permiten comparar entre comunidades. Para más detalles, consulte la Sección 6.6.
- **Distribución de la nota de EBAU:** esta pestaña muestra la distribución por intervalos de la nota obtenida en la EBAU, diferenciada por sexo. Se ofrecen representaciones gráficas y tablas de apoyo para una mejor interpretación de los datos. Para más información, remítase a la Sección 6.7.
- **Nota de Acceso:** muestra información relativa a la nota final de acceso a la universidad, resultado de la ponderación de la nota de Bachillerato y la prueba de EBAU. Se pueden visualizar gráficos y tablas por comunidad, sexo y tipo de centro. Véase la Sección 6.6 para mayor detalle.
- **Distribución de la nota de Acceso:** presenta la distribución por intervalos de la nota de acceso, diferenciada por sexo. Incluye tanto gráficos como tablas numéricas. Más detalles se ofrecen en la Sección 6.7.
- **Información:** en esta pestaña se muestra contenido relacionado con definiciones de interés, un enlace a la fuente de datos utilizada y una breve descripción de las herramientas empleadas para el desarrollo de la aplicación. Mas detalles en la Sección 6.8.



Figura 6.2: Menú de nevgación

6.3. Página de *Resultados por Comunidad*

Esta página muestra datos sobre los estudiantes presentados, aprobados y matriculados en la EBAU para cada comunidad autónoma entre los años 2018 y 2022. Está organizada en seis secciones principales: *Aprobados*, *Matriculados*, *Presentados*, *Evolución por Comunidad*, *Mapa* y *Mapa %*.

Para su diseño se ha utilizado un entorno *fluidRow* que permite situar los selectores en la parte superior y reservar la zona inferior para las distintas pestañas, implementadas mediante un *tabsetPanel*.

En la parte superior de la página se incluyen tres selectores de entrada única (*selectizeInput()*), comunes a todas las pestañas, que permiten filtrar los datos por año, convocatoria y sexo. Sólo en la pestaña *Evolución por Comunidad* se sustituye el selector de año por un selector de comunidad autónoma, con el fin de comparar directamente la evolución de los indicadores entre distintas regiones.

Todos los gráficos representados en esta página son interactivos, gracias a la integración de *ggplot2* con la librería *plotly*.

6.3.1. Apartado de *Aprobados*

La primera pestaña que se visualiza al acceder a la sección *Resultados por Comunidad* es *Aprobados*, la cual contiene dos representaciones gráficas (Figura 6.3).

En la parte superior se encuentra un diagrama de barras que muestra el número de estudiantes aprobados por comunidad autónoma, según los filtros seleccionados de año, convocatoria y sexo.

En la parte inferior se presenta un gráfico de barras apiladas. En este, se visualiza en color verde el porcentaje de aprobados respecto al total de presentados, y en rojo el porcentaje de suspensos, también con base en el total de presentados. Ambos gráficos se actualizan dinámicamente en función de los valores seleccionados para año, convocatoria y sexo.

Cuando se selecciona la opción 'Hombres y Mujeres' en el filtro de sexo, el histograma superior muestra dos barras por comunidad: una correspondiente a los hombres y otra a las mujeres. En el gráfico inferior, se presentan dos gráficos adyacentes, uno para cada sexo. Esta funcionalidad puede observarse en mayor detalle en la Figura B.4 del Anexo B.

Cabe destacar que, de forma coherente en toda la aplicación, se ha utilizado el color azul para representar a los hombres, el color naranja para las mujeres, y el color amarillo para el total.



Figura 6.3: Resultados por comunidad. Pestaña de Aprobados

6.3.2. Apartado de *Presentados*

En esta pestaña se muestran los datos de los presentados para cada comunidad. Al igual que en la pestaña anterior, se incluyen dos gráficos: el diagrama de barras en la parte superior muestra el número de presentados para cada comunidad según los filtros seleccionados de año, convocatoria y sexo (Figura B.5 del Anexo B).

El gráfico de barras apiladas en la parte inferior representa en verde el porcentaje de presentados respecto al total de matriculados, y en rojo el porcentaje de no presentados respecto a los matriculados.

De manera análoga a la pestaña de Aprobados, si se selecciona la opción de sexo *Hombres y Mujeres*, el gráfico superior se transforma en un gráfico de barras agrupadas, mostrando una barra para cada sexo por comunidad. En la parte inferior se visualizan dos gráficos de barras apiladas, uno para hombres y otro para mujeres. Consultar la Figura B.6 del Anexo B.

6.3.3. Apartado de *Matriculados*

En esta pestaña se representa un único diagrama de barras que muestra el número de matriculados para cada comunidad autónoma en función del año, la convocatoria y el sexo seleccionados. En el eje horizontal (X) se disponen las distintas comunidades autónomas, mientras que en el eje vertical (Y) se representa el número de matriculados. Este comportamiento se ilustra en la Figura 6.4.

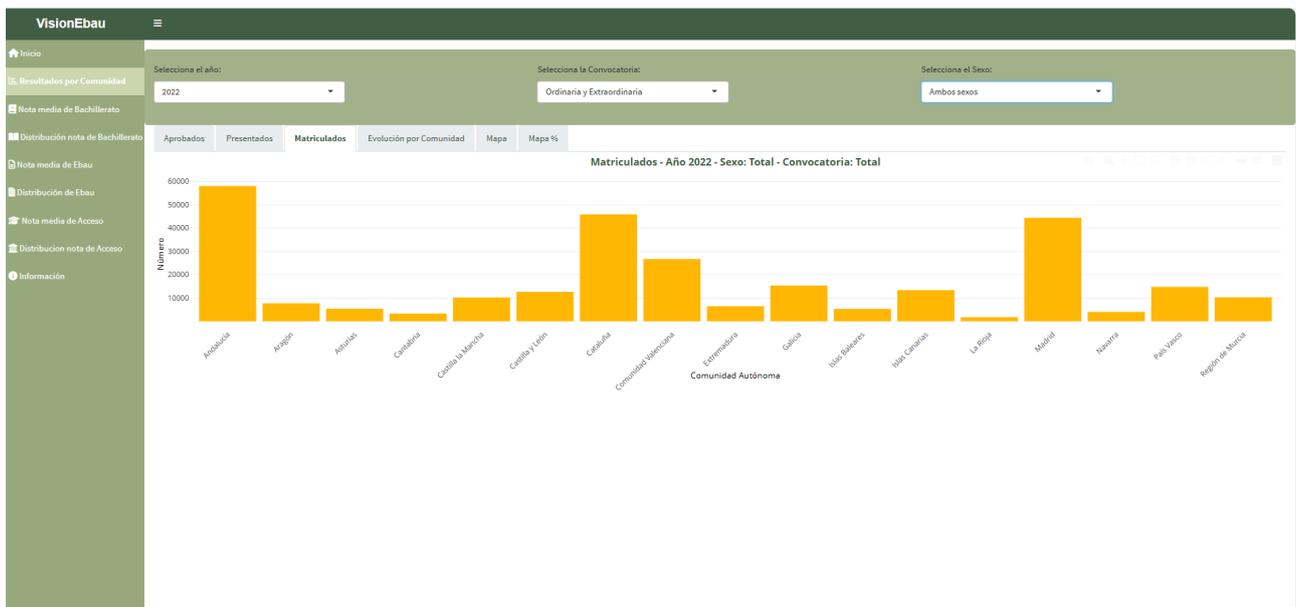


Figura 6.4: Resultados por comunidad. Pestaña de Matriculados

Al igual que en las dos pestañas anteriores, si se selecciona la opción de 'Hombres y Mujeres', se presenta un gráfico de barras agrupadas. Este comportamiento se puede consultar en la Figura B.7 del Anexo B.

6.3.4. Apartado de *Evolución por Comunidad*

En esta pestaña se presenta la evolución a lo largo de los años del número de matriculados, presentados y aprobados para cada comunidad autónoma.

A diferencia de las pestañas anteriores, el selector de años ha sido sustituido por un selector de comunidades autónomas, permitiendo al usuario elegir la comunidad, la convocatoria y el sexo que desea visualizar. Posteriormente, se genera un gráfico de líneas que representa dicha evolución para la comunidad seleccionada. Para ello, se ha utilizado la función `geom_line()` del paquete `ggplot2`, añadiendo también los puntos correspondientes mediante `geom_point()`. En el eje X se representan los años, mientras que en el eje Y se muestra el número de matriculados, presentados o aprobados, según corresponda.

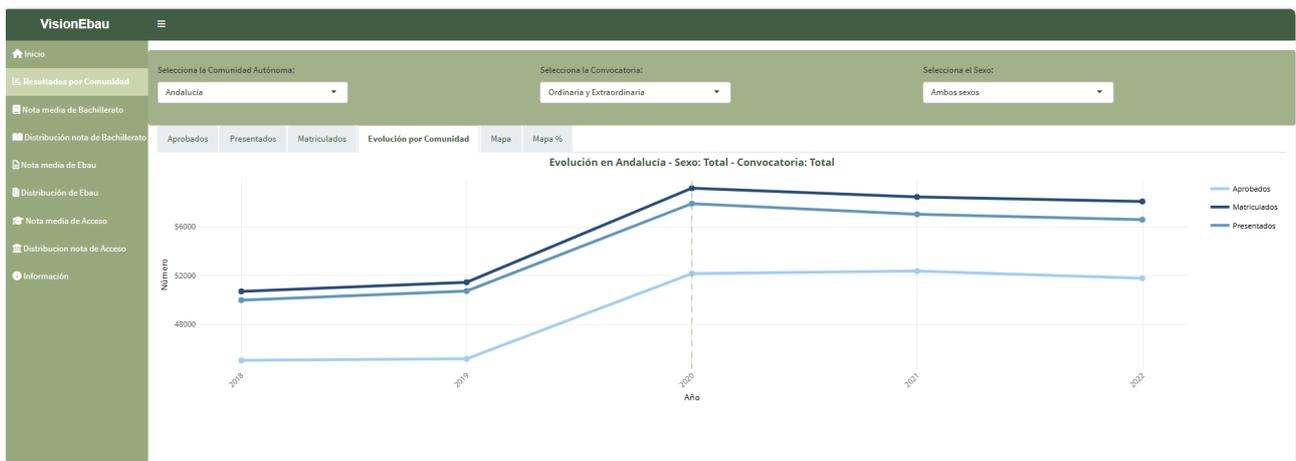


Figura 6.5: Resultados por comunidad. Pestaña de Evolución por Comunidad

También se puede observar en la Figura 6.5 que en el gráfico se ha dibujado una línea vertical discontinua en el año 2020, con el fin de hacerlo más visual y facilitar la diferenciación de los años previos y posteriores al impacto del COVID-19.

Como en los casos anteriores, si se selecciona la opción *Hombres y Mujeres* en el selector de sexo, se generan dos gráficos de líneas con puntos que representan la evolución de los datos para la comunidad autónoma seleccionada: uno correspondiente a la población masculina y otro a la femenina. Esta visualización comparativa permite analizar las posibles diferencias entre ambos sexos a lo largo del tiempo. Para una ilustración detallada de este comportamiento, puede consultarse la Figura B.8 del Anexo B.

6.3.5. Apartado de *Mapa*

En esta pestaña se incluyen nuevamente los selectores de *Año*, *Convocatoria* y *Sexo*, que permiten filtrar la información representada en los gráficos. Como se puede observar en la Figura 6.6, se muestran tres mapas, en los cuales se representa, respectivamente, el número de aprobados, presentados y matriculados por comunidad autónoma.

Estos mapas han sido generados utilizando la librería *mapSpain*, que facilita la representación geográfica de datos estadísticos sobre el territorio español, proporcionando la cartografía de mapas detallados y actualizados de las distintas comunidades autónomas.

Además, los gráficos son interactivos: al situar el cursor sobre una comunidad autónoma, se muestra una etiqueta emergente (*tooltip*) con el nombre de la comunidad y el número correspondiente de aprobados, presentados o matriculados, en función del mapa que se esté visualizando.

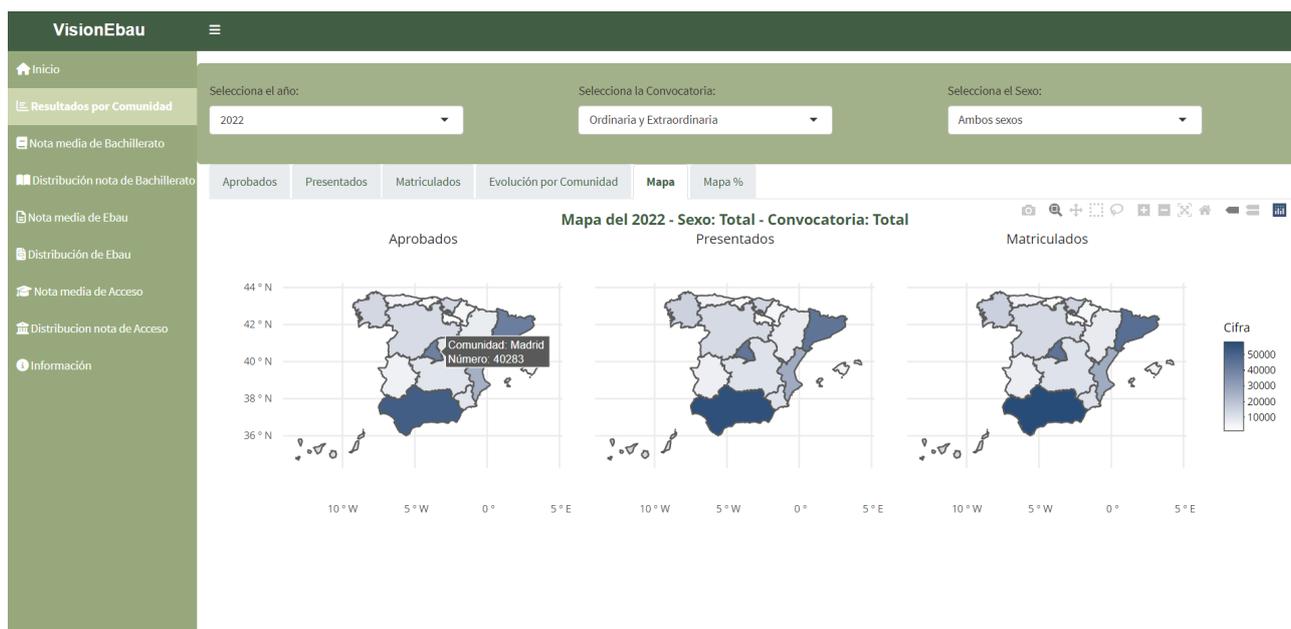


Figura 6.6: Resultados por comunidad. Pestaña de Mapa

6.3.6. Apartado de *Mapa %*

Esta pestaña presenta los mismos selectores que la anterior: *Año*, *Convocatoria* y *Sexo*. En ella se muestran dos mapas de España generados de la misma forma que en la pestaña anterior, utilizando la librería *mapSpain*. La principal diferencia radica en la información representada: el mapa de la izquierda muestra el porcentaje de aprobados sobre presentados, mientras que el mapa de la derecha representa el porcentaje de presentados sobre matriculados.

Ambos gráficos son interactivos, permitiendo al usuario visualizar, al pasar el cursor sobre una comunidad autónoma, su nombre y el valor del porcentaje correspondiente. Además, se ha añadido un pequeño comentario explicativo para facilitar la interpretación de los datos representados. Una visualización de estos mapas puede consultarse en la Figura B.9 del Anexo B.

6.4. Página de *Nota media de Bachillerato*

Esta página ha sido diseñada utilizando la función *sidebarLayout()*, lo cual permite ubicar los selectores en la parte izquierda de la interfaz, mientras que el contenido principal se muestra a la derecha mediante el *mainPanel*.

Se incluyen cuatro pestañas principales: *Comunidades*, *Tabla*, *Comparación comunidades por sexo* y *Comparación comunidades por centro*. Cada una de ellas permite al usuario explorar distintos aspectos relacionados con la nota media de Bachillerato en el período comprendido entre los años 2018 y 2022.

6.4.1. Apartado de *Comunidades*

En esta pestaña se incluyen dos selectores: uno correspondiente a la comunidad autónoma y otro a la convocatoria. Una vez aplicados los filtros, se presentan dos gráficos de líneas y puntos que muestran la evolución de la nota media de Bachillerato para la comunidad seleccionada.

En el gráfico superior se representa la evolución de la nota media de Bachillerato diferenciada por sexo. Se muestran tres líneas: una para el total del alumnado, otra para hombres y otra para mujeres. Se siguen los esquemas de color definidos previamente, donde el color azul representa a los hombres, el color naranja a las mujeres y el color amarillo al total.

En la parte inferior, se muestra un segundo gráfico de líneas y puntos que refleja la evolución de la nota media de Bachillerato, esta vez diferenciada por la naturaleza del centro educativo (público, privado o total).

Al igual que antes ambos gráficos son interactivos, gracias al uso combinado de las librerías *ggplot2* y *plotly*, lo que permite mejorar la experiencia del usuario. Además, también se ha incorporado en ambos gráficos una línea discontinua vertical en el año 2020.

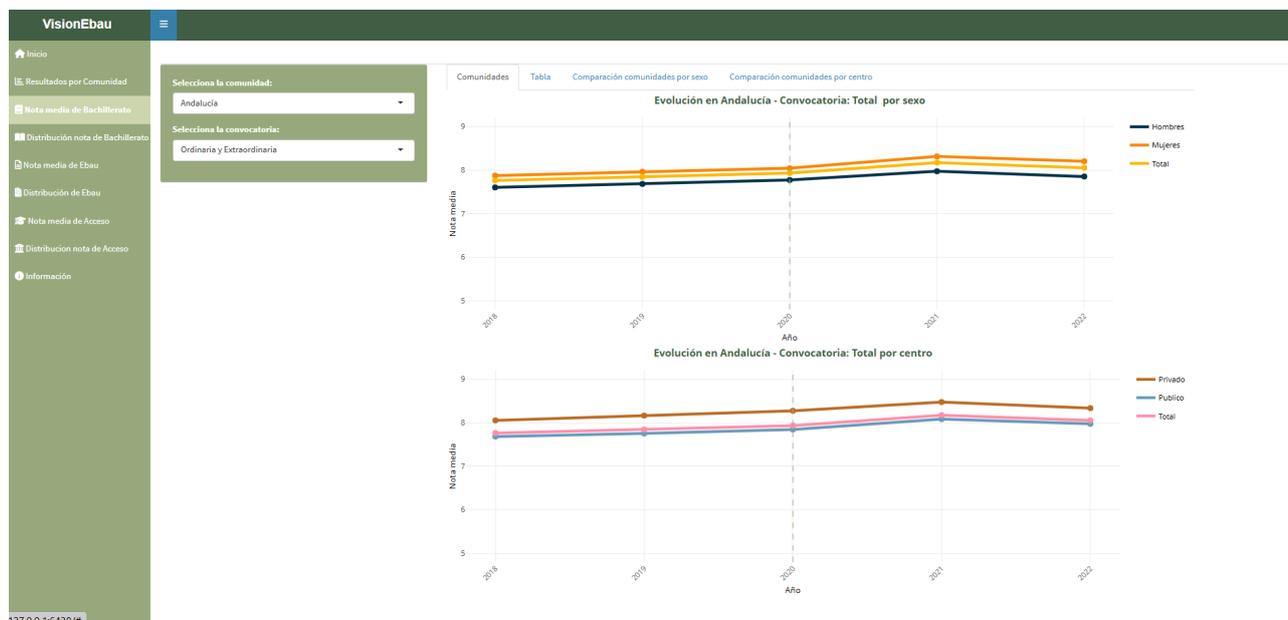


Figura 6.7: Nota media de Bachillerato. Pestaña de Comunidades

6.4.2. Apartado de *Tabla*

En esta pestaña se han incluido dos tablas interactivas empleando la librería *DT*, las cuales muestran información relacionada con la nota media de Bachillerato, complementando los gráficos presentados en la pestaña anterior.

La tabla situada en la parte superior presenta los datos correspondientes a la nota media de Bachillerato desglosados por sexo. Por su parte, la tabla ubicada en la parte inferior muestra la información relativa a la nota media de Bachillerato diferenciada según la naturaleza del centro educativo (público, privado o total).

Ambas tablas ofrecen funcionalidades interactivas como la posibilidad de ordenar las columnas, seleccionar el número de entradas a visualizar por página, cambiar de página y realizar búsquedas personalizadas. Estas funcionalidades pueden observarse en la Figura 6.8.

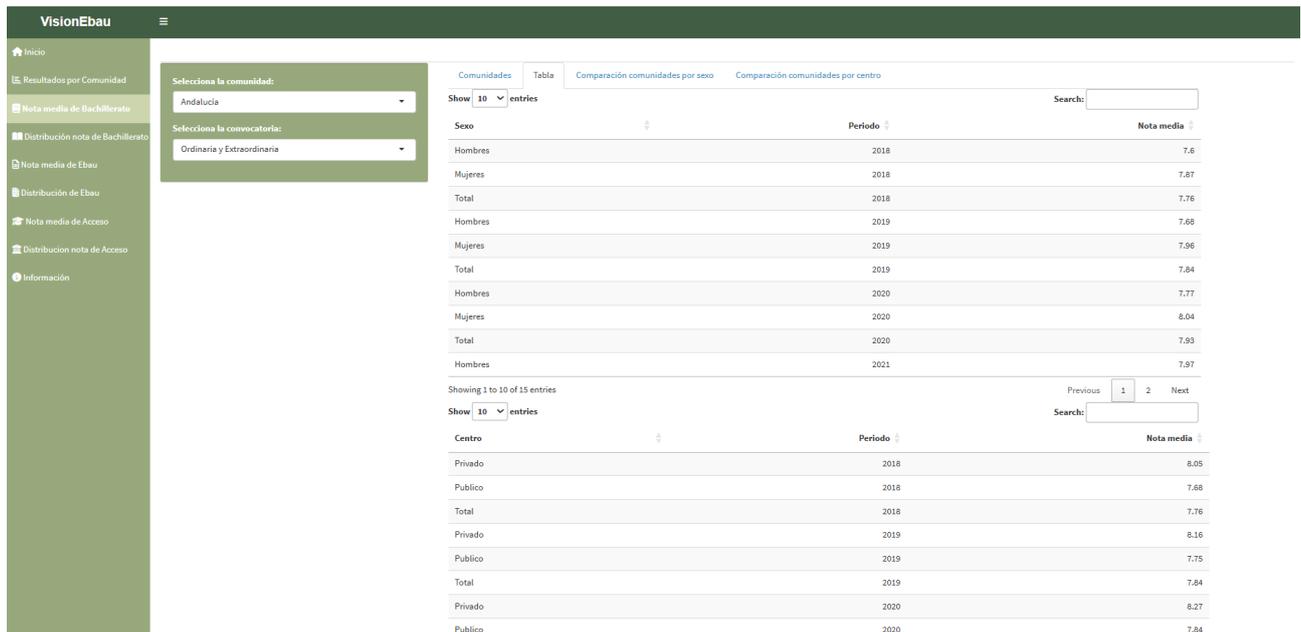


Figura 6.8: Nota media de Bachillerato. Pestaña de Tabla

6.4.3. Apartado de *Comparación comunidades por sexo*

En esta pestaña se incluyen dos selectores: uno destinado a la elección del sexo y otro correspondiente a la convocatoria. Adicionalmente, se incorpora un *checkboxGroupInput* que permite la selección múltiple de comunidades autónomas.

El propósito de esta sección es ofrecer al usuario la posibilidad de comparar la evolución de la nota media de Bachillerato entre varias comunidades, filtrando los datos según el sexo y la convocatoria seleccionados. Para ello, se genera un gráfico de líneas y puntos, donde el eje horizontal (X) representa los años y el eje vertical (Y) muestra el valor de la nota media de Bachillerato.

El gráfico resultante representará tantas líneas como comunidades hayan sido seleccionadas por el usuario, permitiendo así una comparación visual clara entre ellas. En la Figura 6.9 se muestra un ejemplo con dos comunidades seleccionadas, mientras que en la Figura B.10 del Anexo B se presenta un ejemplo con cuatro comunidades.

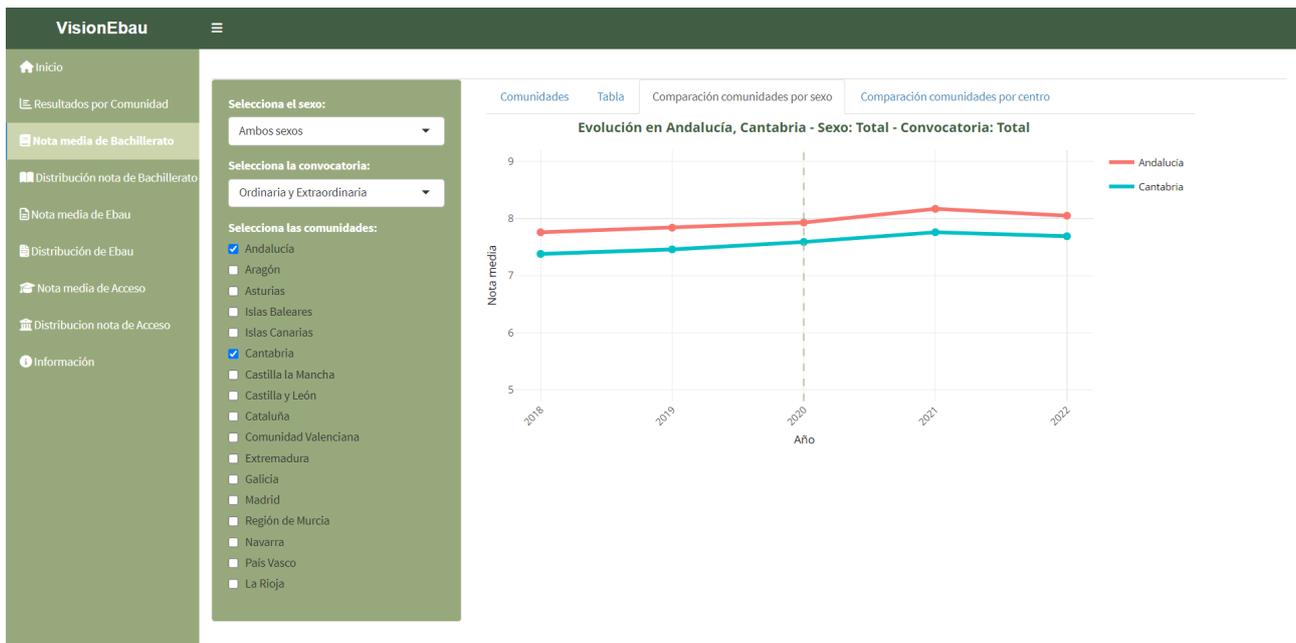


Figura 6.9: Nota media de Bachillerato. Pestaña de Comparación comunidades por sexo

6.4.4. Apartado de *Comparación comunidades por centro*

El objetivo de esta pestaña es el mismo al de la descrita anteriormente, con la diferencia de que, en este caso, la comparación se realiza en función de la naturaleza del centro educativo en lugar del sexo.

En esta sección se presentan dos selectores: uno para seleccionar la naturaleza del centro (público, privado o total) y otro correspondiente a la convocatoria, igual que en la pestaña anterior. Además, se incorpora el mismo `checkboxGroupInput` que permite al usuario seleccionar múltiples comunidades autónomas para su comparación.

El gráfico generado es de líneas y puntos, donde el eje horizontal (X) representa los años y el eje vertical (Y) muestra la nota media de Bachillerato. Este gráfico permite comparar visualmente la evolución de la nota media según el tipo de centro en las distintas comunidades seleccionadas.

En la Figura B.11 se presenta un ejemplo de comparación entre comunidades, y en la Figura B.12 del Anexo B se puede observar un caso con un mayor número de comunidades seleccionadas.

6.5. Página de *Distribución nota de Bachillerato*

Esta página ha sido diseñada de forma similar a la anterior, utilizando un `sidebarLayout()`, lo que permite disponer los controles de usuario en la parte izquierda de la interfaz y el `mainPanel` con los resultados gráficos en la parte derecha.

El objetivo principal de esta sección es representar gráficamente la distribución de los intervalos de notas de Bachillerato. Para ello, se han diseñado distintos gráficos que permiten analizar cómo se distribuyen las calificaciones entre los estudiantes en función de diversos criterios seleccionados por el usuario.

6.5.1. Apartado de *Comunidades*

En esta pestaña se presentan tres selectores: uno para el año, otro para la comunidad autónoma y un tercero para la convocatoria (ver Figura 6.10). A partir de estos filtros, se generan dos gráficos que se muestran en el panel principal.

El gráfico superior se corresponde con un gráfico de barras en el que se representa la distribución de notas de Bachillerato para ambos sexos. En este gráfico, el eje horizontal (X) muestra los intervalos de calificaciones, mientras que el eje vertical (Y) representa la frecuencia de cada intervalo.

El gráfico inferior se trata de un histograma doble (*back-to-back histogram*), diseñado para comparar la distribución de notas entre hombres y mujeres. En este caso:

- El eje vertical (Y) muestra los intervalos de calificaciones.
- El eje horizontal (X) representa la frecuencia dentro de cada grupo.
- Las barras hacia la izquierda reflejan la distribución para el sexo femenino, mientras que las barras hacia la derecha corresponden al sexo masculino.

Este tipo de representación facilita la comparación visual directa entre ambos sexos, permitiendo detectar diferencias en la distribución de calificaciones dentro de cada intervalo.



Figura 6.10: Distribución nota de Bachillerato. Pestaña de Comunidades

6.5.2. Apartado de *Tabla*

En esta pestaña se presenta una tabla complementaria que muestra los datos numéricos correspondientes a los gráficos de la pestaña anterior sobre la distribución de notas de Bachillerato.

Al igual que en el resto de tablas de esta aplicación, esta tabla ha sido implementada con la librería *DT*, lo que permite al usuario reordenar las filas por cualquier columna, modificar el número de entradas visibles y realizar búsquedas.

Esta funcionalidad mejora notablemente la interactividad y usabilidad de la aplicación, permitiendo una exploración detallada de los datos.

Un ejemplo de esta tabla puede observarse en la Figura B.13 del Anexo B.

6.5.3. Apartado de *Evolución*

El objetivo de esta pestaña es representar la evolución temporal de un intervalo específico de calificación para una comunidad autónoma seleccionada (ver Figura 6.11).

Para ello, se incluyen tres selectores que permiten filtrar los datos según la comunidad, la convocatoria y el intervalo de notas.

Una vez seleccionados los filtros, se genera un gráfico de líneas y puntos que muestra la evolución de la frecuencia de dicho intervalo a lo largo de los años. El gráfico incluye tres líneas diferenciadas: una para el total, otra para hombres y otra para mujeres.

En el eje de abscisas (X) se representan los años, mientras que en el eje de ordenadas (Y) se muestra la frecuencia del intervalo seleccionado. Un ejemplo visual de este gráfico puede consultarse en la Figura B.13.

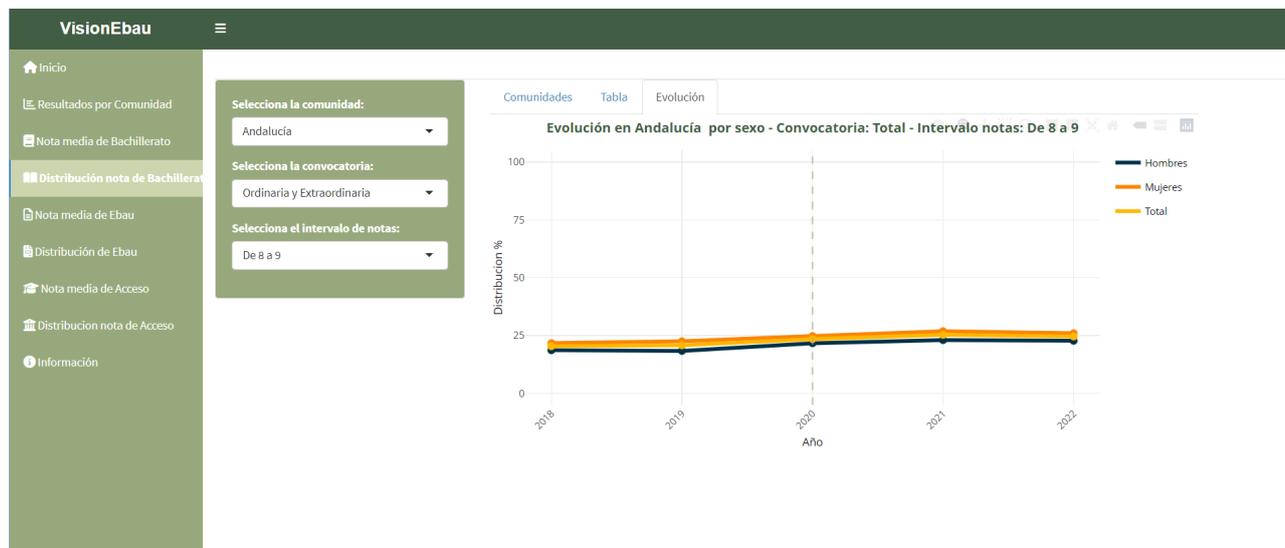


Figura 6.11: Distribución nota de Bachillerato. Pestaña de Evolución

6.6. Páginas de *Nota media de Ebau* y *Nota media de Acceso*

Las páginas correspondientes a la *Nota media de la Ebau* y a la *Nota media de acceso* comparten la misma estructura y disposición gráfica que la página de *Nota media de Bachillerato*, descrita anteriormente.

- En la página de *Nota media de la Ebau*, los datos representados hacen referencia a las calificaciones medias obtenidas en la prueba de evaluación. La disposición y el contenido visual de esta sección pueden consultarse en las Figuras B.14 a B.20 del Anexo B.
- En la página de *Nota media de acceso*, los datos se refieren a las calificaciones medias de acceso a la universidad. Para una referencia visual de esta página, véanse las Figuras B.21 a B.27 del Anexo B.

6.7. Páginas de *Distribución nota de Ebau* y *Distribución nota de Acceso*

Las páginas correspondientes a la *Distribución de nota de la EBAU* y a la *Distribución de nota de acceso* siguen una estructura similar a la de la página de *Distribución de nota de Bachillerato*, tanto en términos de disposición como de gráficos representados.

- En la página de *Distribución de nota de la EBAU*, los datos representados hacen referencia a la distribución de los intervalos de las notas de la prueba. En esta pestaña, además de los intervalos previos, se incluye el intervalo de *4 a 5*, dado que una persona puede suspender la prueba pero, tras aplicar la ponderación, puede obtener una calificación aprobatoria. Para ver los gráficos correspondientes, consulte las Figuras B.28 a B.30 del Anexo B.

- En la página de *Distribución de nota de acceso* se presenta la distribución de los intervalos de la nota de acceso. Para ver los gráficos correspondientes, consulte las Figuras B.31 a B.33 del Anexo B.

6.8. Página de *Información*

En esta página, su finalidad es la de explicación y aclaración de puntos importantes para el usuario. Se pueden encontrar tres apartados:

- **Definiciones de interés:** En este apartado se explican conceptos clave para un mejor entendimiento de los datos.
- **Fuente de datos:** Se indica cuál es la fuente de donde se han extraído los datos y se proporciona un enlace a ella.
- **Herramientas:** En este apartado se explican las herramientas utilizadas para el desarrollo de esta aplicación, y se incorpora un logo de Shiny.

VisionEbau ☰

Inicio

Resultados por Comunidad

Nota media de Bachillerato

Distribución nota de Bachillerato

Nota media de Ebau

Distribución de Ebau

Nota media de Acceso

Distribución nota de Acceso

Información

Definiciones de interés

A continuación, se explican las diferencias entre los tres tipos de calificaciones implicadas en el análisis de esta visualización:

- Nota de Bachillerato:** corresponde a la calificación media obtenida por el estudiante en sus estudios de Bachillerato.
- Nota de la EBAU:** es la puntuación alcanzada en la Evaluación del Bachillerato para el Acceso a la Universidad (EBAU), también conocida como Selectividad.
- Nota de Acceso:** se obtiene aplicando una ponderación del 60% a la nota de Bachillerato y del 40% a la nota de la EBAU. Esta calificación final se utiliza para acceder a los estudios universitarios.

El rango de valores que pueden alcanzar estas tres calificaciones está entre 4 y 10.

Fuente de datos

La fuente de donde han sido extraídos los datos esta proporcionada por el Ministerio de Ciencia, Innovación y Universidades. En esta fuente de datos encontramos diversos repositorios, nosotros utilizamos datos que se encuentran en la sección de Estadística de las pruebas de acceso a la universidad (PAU).
estadisticas.universidades.gob.es

Herramientas

Esta aplicación ha sido creada utilizando el lenguaje R dentro del entorno de desarrollo RStudio. Para su funcionamiento se han empleado diversos paquetes, entre ellos Shiny y Shinydashboard para la interfaz interactiva, MapSpain para la representación geográfica, Dplyr para el manejo y transformación de datos, y Plotly para la visualización dinámica, entre otros complementos adicionales.



Figura 6.12: Página de *Información*

Capítulo 7

Comparación de los resultados de la EBAU

Este capítulo está dedicado al análisis de los intervalos de las notas de acceso correspondientes al período comprendido entre los años 2018 y 2022, diferenciando entre las diversas comunidades autónomas.

7.1. Descripción y preparación de los datos

Los datos utilizados para el presente análisis provienen de dos conjuntos principales mencionados y descritos en el Capítulo 3. No obstante, dado que no se requiere la totalidad de la información contenida en ambos conjuntos, a continuación se detalla únicamente el proceso de filtrado aplicado a dichos datos, con el propósito de seleccionar exclusivamente aquellas variables y registros relevantes para la comparación mediante el test chi-cuadrado y el análisis de correspondencias:

- **Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo:** De este conjunto, se ha seleccionado exclusivamente la información correspondiente al número de aprobados y presentados por comunidad autónoma, considerando los años comprendidos entre 2018 y 2022. Además, se ha optado por trabajar únicamente con los valores agregados bajo las categorías de *convocatoria=total* y *sexo=total* con el objetivo de obtener una visión general y evitar subdivisiones excesivas.
- **Nota de acceso al grado de los titulados de Bachiller. Distribución por tipo de matrícula, convocatoria y sexo:** En este caso, se ha retenido la información correspondiente a cada intervalo de calificación, comunidad autónoma y año, específicamente aquella asociada a la matrícula en la *fase general y específica*, también agregada bajo las categorías de *sexo=total* y *convocatoria=total*.

La integración de ambos conjuntos de datos, tras este proceso de selección y simplificación, permite disponer de una base unificada y general, adecuada para la construcción de la tabla de contingencia necesaria para el análisis de correspondencias. Este enfoque asegura que los datos sean consistentes, comparables y representativos, facilitando así una interpretación precisa y coherente de los resultados obtenidos.

Construcción de las tablas de contingencia

El procedimiento para obtener las tablas de contingencia se describe en el Anexo C. Para cada comunidad se genera una tabla de contingencia completa. En ella se recogen tanto los suspensos (calculados como la diferencia entre los presentados y los aprobados) como la distribución de los aprobados según los intervalos de calificación previamente definidos. Este paso es fundamental, ya que establece las bases necesarias para llevar a cabo el posterior análisis de correspondencias.

A continuación, en la Figura 7.1 se muestra un ejemplo de la tabla de contingencia correspondiente a la comunidad de Castilla y León, que refleja los datos de calificación desglosados por intervalos y años.

```

> tablas_comunidad[['Castilla y León']]
Intervalo 2018 2019 2020 2021 2022
1 <5 830 779 765 389 379
2 De 5 a 6 2060 1707 1616 1033 1119
3 De 6 a 7 3453 3042 3121 2910 2910
4 De 7 a 8 2841 2929 3188 3314 3113
5 De 8 a 9 2160 2337 2847 3134 3064
6 De 9 a 10 1007 1239 1834 2104 1986

```

Figura 7.1: Tabla de contingencia para Castilla y León

7.2. Aplicación Test chi-cuadrado

En esta sección, se usa la prueba chi-cuadrado sobre las tablas de contingencia correspondientes a cada comunidad autónoma definidas en la Sección 7.1 .

7.2.1. Metodología

La prueba chi-cuadrado es un método estadístico no paramétrico que permite analizar la relación entre variables categóricas [20]. En este procedimiento a partir de la clasificación de los datos en categorías se calcula un estadístico basado en la diferencia entre las frecuencias observadas y las frecuencias esperadas bajo la hipótesis nula de homogeneidad entre los grupos.

En este análisis, el objetivo específico de la prueba chi-cuadrado es evaluar la homogeneidad de las distribuciones, es decir, determinar si la distribución de los intervalos de calificaciones se mantiene consistente a lo largo de los distintos cursos académicos. En otras palabras, se busca comprobar si las proporciones de estudiantes en cada categoría de notas son estadísticamente equivalentes entre los diferentes años considerados.

En esta sección se hace especial énfasis en el análisis de los residuos estandarizados [21], ya que estos permiten identificar con mayor precisión qué celdas de la tabla contribuyen significativamente a las discrepancias entre las frecuencias observadas y las esperadas bajo la hipótesis nula. De este modo, los residuos estandarizados resultan particularmente útiles para interpretar el origen de la no homogeneidad detectada entre las variables analizadas.

Los residuos estandarizados se calculan como la diferencia entre la frecuencia observada y la frecuencia esperada, dividida por la raíz cuadrada de la frecuencia esperada [22]:

$$r_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

- n_{ij} : frecuencia observada en la celda (i, j) .
- e_{ij} : frecuencia esperada en cada celda (i, j) de la tabla de contingencia: $e_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n}$

Estos residuos permiten detectar desviaciones significativas de las frecuencias esperadas, indicando celdas donde la diferencia es más notable. Los residuos estandarizados con valores absolutos superiores a 2 son considerados estadísticamente significativos, lo que indica que la diferencia entre la frecuencia observada y la esperada en esa celda no es atribuible al azar.

7.2.2. Evaluación de los resultados

Se ha realizado el test chi-cuadrado y se ha obtenido la tabla de residuos estandarizados para todas las comunidades. Dado que los resultados para todas las comunidades son muy similares, se procede a analizar únicamente los resultados correspondientes a la tabla de contingencia de la Figura 7.1. El resto de los resultados están disponibles en el *Anejo D*.

Para Castilla y León el valor del estadístico chi-cuadrado es 1707.475 y el p-valor correspondiente es muy cercano a 0, por lo tanto las calificaciones no son homogéneas en los distintos cursos.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	9.97	18.19	10.20	-3.86	-11.88	-17.62
2019	8.43	8.52	2.44	0.01	-5.95	-9.55
2020	4.50	0.66	-3.27	-1.51	-0.42	3.07
2021	-11.42	-15.32	-5.43	4.10	8.99	12.91
2022	-11.27	-11.67	-3.71	1.24	9.01	10.73

Cuadro 7.1: Residuos estandarizados para la comunidad de Castilla y León

Para el 2018 se puede observar que los residuos estandarizados son positivos en los rangos de calificaciones más bajas (<5, De 5 a 6, De 6 a 7), lo que indica una sobrerrepresentación de estudiantes en estos intervalos. En cambio, los residuos son negativos en los rangos más altos (De 7 a 8, De 8 a 9, y De 9 a 10), lo que indica una subrepresentación de estudiantes en estas categorías (ver Cuadro 7.1) respecto a lo esperado en caso de homogeneidad de los cursos.

Para el 2019 al igual que en 2018, se observa una sobrerrepresentación en los rangos bajos (<5, De 5 a 6, De 6 a 7), especialmente en el primer rango (<5). En las calificaciones más altas (especialmente en De 9 a 10), hay una subrepresentación notable, indicando que menos estudiantes se agrupan en estas categorías.

En el año 2020, los residuos estandarizados tienden a estar más cerca de cero en varias categorías, lo que sugiere una distribución más equilibrada de las calificaciones. Sin embargo, hay una ligera sobrerrepresentación en los rangos de calificación media (De 5 a 6, De 6 a 7) y una ligera sobrerrepresentación en las categorías más altas (De 9 a 10).

Para el año 2021 se observa una clara subrepresentación en los rangos de calificaciones bajas (<5, De 5 a 6, De 6 a 7), indicando que menos estudiantes se agruparon en estos intervalos. Por otro lado, las calificaciones más altas (De 8 a 9, De 9 a 10) muestran una sobrerrepresentación, lo que sugiere un aumento en la proporción de estudiantes en estos intervalos.

Por último, el año 2022 es Similar a 2021, los residuos estandarizados muestran una subrepresentación en los rangos más bajos (<5, De 5 a 6, De 6 a 7) y una sobrerrepresentación en los rangos de calificaciones altas (De 8 a 9, De 9 a 10).

El gráfico de mosaico de la Figura 7.2 representa la distribución de los intervalos de calificaciones a lo largo de los distintos años académicos en la comunidad de Castilla y León.

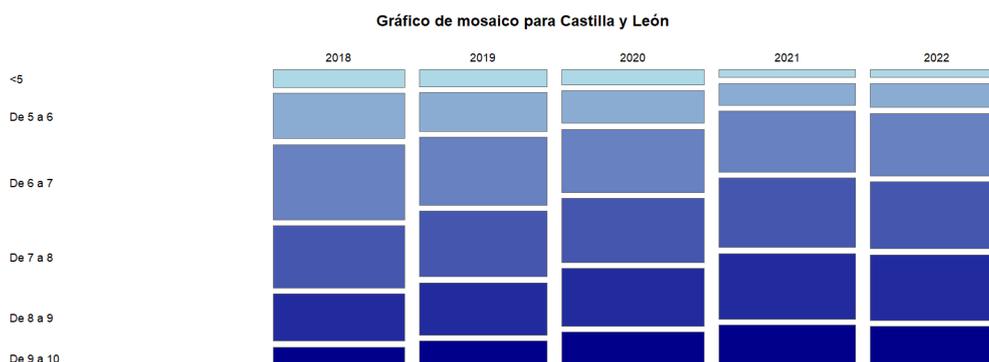


Figura 7.2: Gráfico de mosaico para la comunidad de Castilla y León

Durante los años 2018 y 2019 se observa una mayor concentración de calificaciones en los intervalos bajos y medios, especialmente en los tramos 'De 5 a 6' y 'De 6 a 7', mientras que las calificaciones altas ('De 8 a 9'

y 'De 9 a 10') aparecen con mucha menor frecuencia. En contraste, los años 2021 y 2022 presentan una clara tendencia hacia una mayor proporción de calificaciones altas, junto con una disminución significativa de los intervalos más bajos.

El año 2020 muestra una distribución más equilibrada, sin grandes desviaciones en ninguna dirección.

Estos resultados reflejan una ruptura en la homogeneidad de las calificaciones entre los distintos años académicos, indicando un cambio sustancial en la tendencia de evaluación a lo largo del tiempo.

7.3. Aplicación de un análisis de correspondencias

En esta sección se detalla la utilización de un Análisis de Correspondencias aplicado a los datos de las notas de acceso de los años 2018 a 2022. El objetivo principal es explorar si existe alguna relación o asociación entre las diferentes categorías de las notas de acceso (intervalos: $[< 5]$, $[5-6)$, $[6-7)$, $[7-8)$, $[8-9)$, $[9-10]$) a lo largo de los años estudiados.

7.3.1. Metodología

El análisis de correspondencias es una técnica estadística multivariante que permite examinar gráficamente la relación entre dos variables nominales en un espacio multidimensional [23]. Su principal objetivo es representar, de forma simplificada, las asociaciones existentes entre las categorías de dichas variables, facilitando la interpretación visual de patrones y relaciones.

Este procedimiento se basa en el cálculo de las puntuaciones asociadas a las filas y columnas de la tabla de contingencia, a partir de las cuales se generan gráficos que reflejan la proximidad entre categorías. En estos gráficos, las categorías que presentan patrones similares aparecen ubicadas próximas entre sí, lo que permite identificar afinidades o similitudes tanto dentro de cada variable como entre las dos variables consideradas[24].

A continuación, se procede a describir detalladamente los tres principales componentes del análisis de correspondencias [25] [26] :

- **Tabla de contingencia:** Constituye el punto de partida del análisis. En ella se recogen las frecuencias observadas para cada combinación de categorías de las variables cualitativas analizadas. Esta tabla resume la información conjunta de las dos variables y permite identificar patrones iniciales de asociación.
- **Matriz de correspondencias:** Se obtiene mediante una transformación a partir de la tabla de contingencia, y permite calcular las coordenadas de las filas y columnas en un espacio de dimensiones reducidas. Esta matriz refleja la relación subyacente entre las categorías de ambas variables, facilitando la interpretación de las asociaciones presentes.
- **Biplot o gráfico de correspondencias:** Representa visualmente los resultados del análisis. En este gráfico, cada categoría de las filas y columnas se representa como un punto en el espacio reducido, y las distancias entre puntos indican la intensidad de la asociación entre las categorías. Este recurso visual es fundamental para comprender de manera intuitiva las relaciones entre variables y detectar agrupamientos o patrones relevantes.

Cabe destacar que existen dos tipos principales de análisis de correspondencias, cuya elección depende de la naturaleza y el número de variables consideradas:

- El análisis de correspondencias simple permite analizar la relación entre dos variables cualitativas a partir de una tabla de contingencia, representando gráficamente las asociaciones entre sus categorías [27].
- El análisis de correspondencias múltiple permite analizar simultáneamente más de dos variables cualitativas, utilizando tablas de Burt o matrices disyuntivas completas, mostrando las relaciones globales entre categorías y los patrones conjuntos presentes en los datos [28].

En el presente trabajo se ha optado por utilizar un análisis de correspondencias simple, ya que el objetivo principal es examinar la relación entre dos variables específicas, permitiendo una interpretación clara y directa de los resultados obtenidos.

7.3.2. Generación de la matriz de correspondencias

Una vez que hemos obtenido la tabla de contingencia, el siguiente paso es generar la matriz de correspondencias. Para ello, utilizamos la función `CA` de la librería `FactoMineR` en R, que es específicamente diseñada para realizar un análisis de correspondencias [29].

La función `CA` lleva a cabo el análisis de correspondencias de una tabla de contingencia y calcula las coordenadas de las filas y las columnas en un espacio de menor dimensión. Estas coordenadas son fundamentales para representar visualmente las relaciones entre las categorías de las variables involucradas en el análisis. La salida de la función `CA` proporciona tanto los autovalores (que permiten conocer la importancia de cada dimensión) como las coordenadas de las filas y las columnas, lo que nos ayuda a interpretar las asociaciones entre las variables.

Una vez que tenemos la tabla de contingencia preparada, ejecutamos el siguiente código en R para generar la matriz de correspondencias para Castilla y León:

```
# Cargar la librería FactoMineR
library(FactoMineR)

# Convertir a dataFrame nuestra tabla de una comunidad determinada
tabla_analisis <- as.data.frame(tablas_comunidad[['Castilla y Leon']])

# Generar la matriz de correspondencias con la tabla de contingencia para una comunidad
res.ca <- CA(tabla_analisis[, -1], graph = FALSE) # Excluye la columna 'Intervalo'
           porque es cualitativa

# Visualizar la matriz de correspondencias
res.ca$col$coord # Coordenadas de las columnas
res.ca$row$coord # Coordenadas de las filas
```

La salida que obtenemos se puede observar en la Figura 7.3, donde se presentan las coordenadas de las columnas, y en la Figura 7.4, donde se muestran las coordenadas de las filas. Esta matriz de correspondencias nos permite observar las relaciones entre las categorías de las variables de manera más clara y directa, facilitando el análisis posterior.

```
> res.ca$col$coord # Coordenadas de las columnas
      Dim 1      Dim 2      Dim 3      Dim 4
2018 0.234260355 0.030260015 -0.007254292 -0.002907712
2019 0.125900104 -0.012807660 0.018315854 0.004466172
2020 0.002703938 -0.045768947 -0.011057395 -0.001445446
2021 -0.187796871 0.009576512 0.010492101 -0.005636944
2022 -0.161075894 0.021395747 -0.009396923 0.005896520
```

Figura 7.3: Tabla de coordenadas de las columnas para Castilla y León

```
> res.ca$row$coord # Coordenadas de las filas
      Dim 1      Dim 2      Dim 3      Dim 4
<5      0.31620894 -0.092641790 0.008286554 0.0001442662
De 5 a 6 0.26918167 0.001656209 -0.013731227 0.0022144322
De 6 a 7 0.07424216 0.030046744 -0.003588374 -0.0030997152
De 7 a 8 -0.03254353 0.002082931 0.018154651 -0.0016643461
De 8 a 9 -0.12533870 0.001858770 -0.001594213 0.0074388491
De 9 a 10 -0.24110104 -0.029671716 -0.015287870 -0.0054373020
```

Figura 7.4: Tabla de coordenadas de las filas para Castilla y León

De forma análoga se puede disponer de las tablas de coordenadas correspondientes para cada comunidad autónoma.

7.3.3. Visualización de resultados mediante Biplot

En este apartado procedemos a la representación gráfica de los resultados de la matriz de correspondencias mediante un *biplot*. Este gráfico es una representación gráfica bidimensional que permite visualizar simultánea-

mente las filas y columnas de una tabla de contingencia [30].

En primer lugar, el gráfico se construye generalmente a partir de dos ejes principales, denominados dimensiones, que corresponden a las componentes extraídas del análisis de correspondencias. Cada una de estas dimensiones lleva asociado un porcentaje que indica la proporción de la variabilidad total que explica. La suma de estos porcentajes proporciona el valor total de variabilidad explicada en el plano bidimensional, lo que permite evaluar el grado de fidelidad con el que el *biplot* representa la estructura de los datos originales.

La Dimensión 1 suele recoger la mayor parte de la variabilidad y representa las asociaciones predominantes entre las categorías de las variables, mientras que la Dimensión 2 proporciona información adicional sobre asociaciones secundarias.

En el gráfico, se representan mediante puntos tanto las categorías de las filas (en nuestro caso los intervalos de calificaciones) como las de las columnas (en nuestro caso los años). La distancia entre puntos de distinto tipo (fila y columna) indica el grado de asociación entre dichas categorías: cuanto más cercanos estén, más fuerte es su relación o comportamiento conjunto. Por otro lado, la proximidad entre puntos del mismo tipo sugiere que dichas categorías comparten un perfil similar en relación con las categorías de la variable opuesta.

Otro aspecto clave es la distancia al origen (es decir, al centro del gráfico). Cuanto mayor sea la distancia de un punto al origen, mayor es su contribución a las dimensiones representadas, y por tanto más relevante es su papel en la estructura de las asociaciones observadas. Por el contrario, los puntos próximos al centro tienen un perfil más neutro o menos característico, y no están fuertemente asociados con ninguna categoría específica.

Para la construcción del *biplot*, es necesario ejecutar el siguiente código en R, el cual genera automáticamente el gráfico correspondiente a partir de los resultados del análisis de correspondencias realizado previamente. Dicho gráfico permite visualizar de manera conjunta las coordenadas de las filas (intervalos de calificaciones) y de las columnas (años) en el espacio bidimensional, facilitando así la identificación de las asociaciones y patrones relevantes presentes en los datos.

```
# Generacion del biplot para la comunidad seleccionada
library(FactoMineR)
library(factoextra)

# Grafico biplot
fviz_ca_biplot(res.ca, repel = TRUE)
```



Figura 7.5: Biplot para la comunidad de Extremadura

7.3.4. Interpretación de los resultados

En esta sección se interpretan los resultados del análisis de correspondencias analizando las asociaciones entre años e intervalos de notas en varias comunidades autónomas.

Para todas las comunidades autónomas analizadas, la Dimensión 1 del Análisis de Correspondencias es la que explica la mayor parte de la variabilidad de los datos, con un porcentaje notablemente elevado. Por ejemplo, en el caso de Andalucía esta dimensión representa el 97,5% de la variabilidad, en Castilla y León el 96,6%, y en Canarias el 88,5%. Este patrón se repite de forma consistente en todas las comunidades, lo que permite afirmar que la estructura de los datos es similar en todos los territorios considerados.

A continuación se analizarán con mayor detalle los gráficos correspondientes a Andalucía, Castilla y León y Canarias. El resto de los gráficos biplot puede consultarse en el *Anexo E*, para aquellos lectores interesados en profundizar en los resultados de las demás comunidades.

Análisis de correspondencias para Andalucía

En el análisis de correspondencias realizado para el caso de Andalucía, la Dimensión 1 explica el 97,5% de la variabilidad, lo que indica que prácticamente toda la información relevante se concentra en dicho eje, por otro lado la Dimensión 2, que representa solo el 1,6% de la variabilidad, apenas aporta información adicional en el análisis. En consecuencia, la verticalidad de los puntos en el gráfico tiene poco peso en la interpretación general.

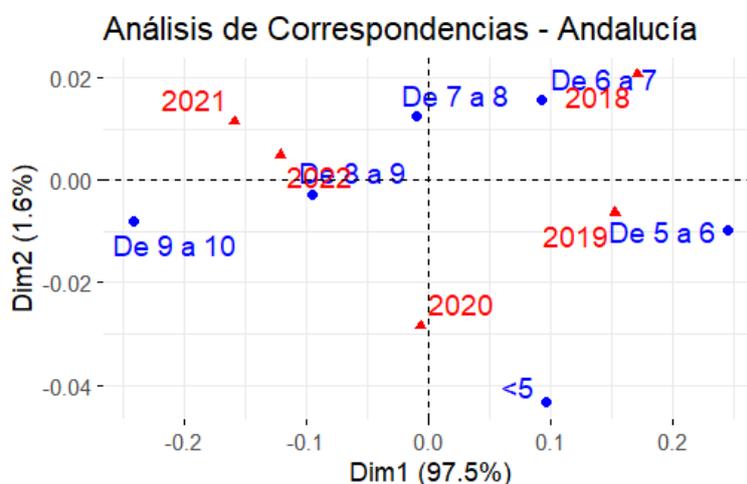


Figura 7.6: Biplot para la comunidad de Andalucía

Inicialmente se puede observar una clara diferencia entre los años antes de la pandemia y después de la pandemia.

En cuanto a la proximidad entre los puntos de la Figura 7.6, si un año está cerca de un intervalo de calificación, esto significa que dicho intervalo fue más frecuente o característico en ese año en Andalucía. Los ejemplos específicos son los siguientes:

- El año 2018 está cerca del intervalo de 6 a 7, lo que sugiere que en 2018 fue común este rango de notas.
- El año 2021 se aproxima al intervalo de 9 a 10, lo que indica una mayor concentración de notas altas en ese año.
- El año 2020 se encuentra cerca del centro del gráfico, sin acercarse demasiado a ningún intervalo específico, lo que podría sugerir una distribución más uniforme o ambigua de las calificaciones.
- El año 2022 está junto al intervalo de 8 a 9, lo que sugiere similitudes una gran cantidad de notas de ese intervalo ese año.
- El año 2019 se asocia principalmente con el intervalo de 5 a 6.

Destacar que al observar el gráfico biplot de la Figura 7.6 , se aprecia una clara asociación entre los años 2021 y 2022 con los intervalos de calificaciones más altos, especialmente aquellos comprendidos entre 9 y 10. Por el contrario, los años 2018 y 2019 se ubican cerca de los intervalos de calificación más bajos, tales como el rango de 5 a 6 o 6 a 7, lo que sugiere una concentración de calificaciones más reducidas en esos años. El año 2020, por su parte, se sitúa en una posición central y alejada de los extremos, lo que puede interpretarse como un comportamiento intermedio.

Análisis de correspondencias para Castilla y León

Al igual que en el caso de Andalucía se puede observar que la mayor variabilidad esta explicada por la dimensión 1 con un 96.6% mientras que la dimensión 2 solo explica un 2.8%.

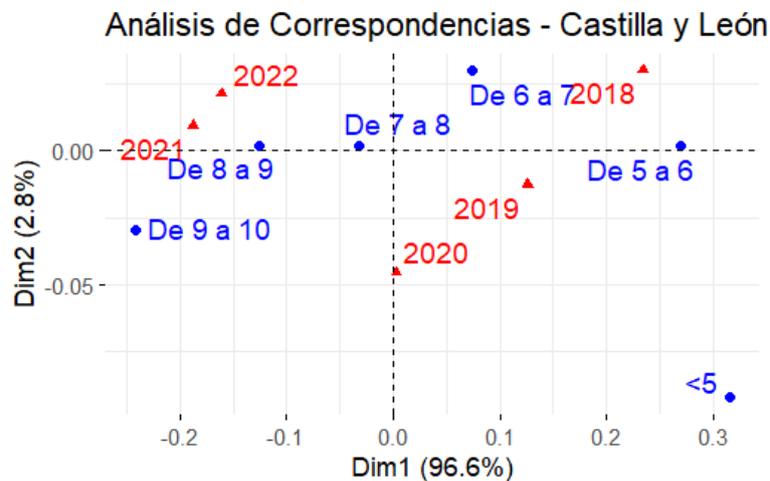


Figura 7.7: Biplot para la comunidad de Castilla y León

Se observa una cierta tendencia que vincula los años 2021 y 2022 con los intervalos de calificaciones más altos, mientras que los años 2018 y 2019 tienden a asociarse con categorías inferiores. Sin embargo, en este gráfico las distancias entre los años y los intervalos de calificación no son lo suficientemente reducidas como para establecer asociaciones categóricas sólidas. Se podría decir que el año 2022 se asocia de manera clara con el intervalo “De 8 a 9” como sucedía con el gráfico de Andalucía (Figura 7.6).

Análisis de correspondencias para las Islas Canarias

Una vez más se puede observar que la Dimensión 1 es la que explica la mayor variabilidad (88.5%) mientras que la Dimensión 2 explica solo un 10% (ver Figura 7.8).

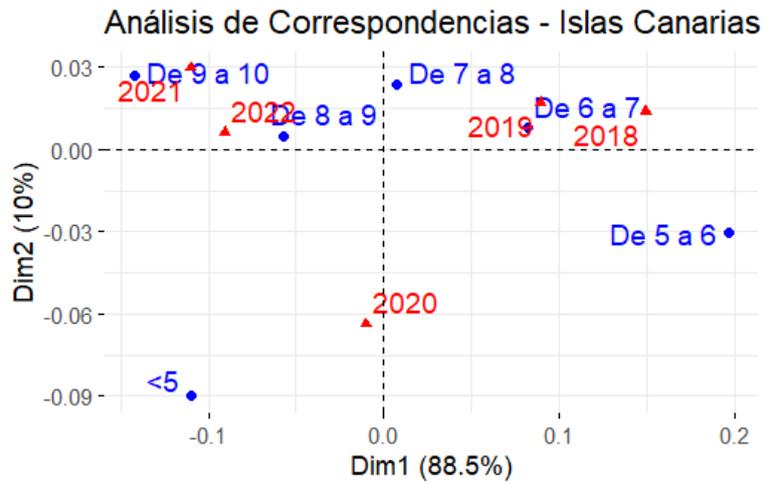


Figura 7.8: Biplot para la comunidad de las Islas Canarias

Los años posteriores al 2020 se asocian con una tendencia hacia calificaciones más altas, mientras que los años previos a 2020 presentan una mayor vinculación con rangos de calificación más bajos.

Como se ha podido observar en los casos anteriores, existe una tendencia común: los años 2021 y 2022 se asocian con los intervalos de notas más altos, mientras que los años 2018 y 2019 tienden a relacionarse con los intervalos de notas más bajos. El año 2020, por su parte, se presenta como un punto neutro o intermedio, probablemente debido a los efectos de la pandemia por COVID-19 y a las modificaciones introducidas en los métodos de evaluación durante ese periodo excepcional.

Capítulo 8

Conclusiones

En este último capítulo se presentan las conclusiones derivadas del desarrollo del TFG, así como las posibles líneas de trabajo que podrían abordarse en el futuro.

Los objetivos establecidos en el apartado 1.1 se han alcanzado satisfactoriamente. Se ha desarrollado una aplicación web interactiva (llamada *VisionEbau*) centrada en el análisis de los datos de la EBAU, dotada de una interfaz intuitiva, accesible, funcional y visualmente atractiva. Esta herramienta integra múltiples gráficos y tablas que permiten al usuario una interpretación rápida, clara y eficaz de la información representada permitiendo analizar y descubrir patrones en los años antes y después de la pandemia.

Asimismo, se ha implementado un análisis de comparación de las calificaciones de la EBAU para los distintos años mediante la aplicación del test chi-cuadrado y un análisis de correspondencias simples, lo que también ha permitido identificar similitudes de comportamiento entre los resultados de la EBAU en los cursos previos a la pandemia y por otro lado los resultados de los cursos posteriores. Llegando a la conclusión que existe una diferencia significativa entre los cursos de antes y después de la pandemia.

Dificultades encontradas

La principal dificultad durante la realización de este proyecto ha sido la extracción y preparación de los datos. En un primer momento, la obtención de los mismos parecía sencilla, dado que la fuente de datos permite la descarga en distintos formatos. Sin embargo, el formato en el que se proporcionaban los archivos no era adecuado para su uso directo en la aplicación, lo que requirió realizar un tratamiento exhaustivo y laborioso de los datos para poder estructurarlos correctamente y hacerlos funcionales dentro del entorno de trabajo.

Adicionalmente, el desarrollo de la aplicación también presentó ciertos desafíos, especialmente en lo relativo al aprendizaje y uso del paquete Shiny. Al no contar con experiencia previa en esta herramienta, fue necesario dedicar un tiempo considerable a su estudio y comprensión para poder implementar de forma efectiva una interfaz dinámica y funcional que respondiera a los objetivos del proyecto.

Trabajo futuro

En cuanto a las posibles líneas futuras de trabajo, se identifican diversas ampliaciones y mejoras que podrían enriquecer notablemente el presente proyecto:

- **Aplicación de un análisis de correspondencias múltiples (ACM):** Este tipo de análisis permitiría estudiar simultáneamente la relación entre más de dos variables categóricas. En el contexto de este proyecto, se podrían considerar variables como el *año*, el *intervalo de nota* y la *comunidad autónoma*, con el objetivo de identificar patrones de comportamiento comunes o diferencias relevantes entre regiones y periodos temporales.
- **Análisis de calificaciones por asignaturas específicas:** Una línea de trabajo interesante consistiría en desglosar los resultados académicos por materia de la EBAU. Esto permitiría analizar el rendimiento

de los estudiantes en áreas concretas del conocimiento y detectar posibles desigualdades o fortalezas por comunidad, sexo o tipo de centro educativo.

- **Actualización automática de los datos:** Con el objetivo de mantener la herramienta útil y vigente en el tiempo, se podría automatizar el proceso de actualización de los datos a medida que se publiquen los resultados de cursos futuros. Esto garantizaría la escalabilidad y sostenibilidad del proyecto a largo plazo.

Anexos

Anexo A

Extracción Datos

```
6
7 # Definir las columnas
8 tipos_convocatoria = ['Fase general y especifica', 'Fase general']
9 tipo centro = ['Total', 'Publico', 'Privado', 'Concertado', 'No consta']
10 periodo_2022 = ['2022']
11 periodo_2021=['2021']
12 periodo_2020=['2020']
13 periodo_2019=['2019']
14 periodo_2018=['2018']
15 comunidades = ['Andalucía', 'Aragón', 'Asturias', 'Balears', 'Canarias', 'Cantabria', 'Castilla la Mancha', 'Castilla y León', 'Cataluña', 'Comunitat Valenciana', 'Estado', 'Extrema
16 tipos_matricula = ['Total', 'Ordinaria', 'Extraordinaria']
17
18 # Las notas proporcionadas para cada combinación
19 notas_acceso_2022 = [
20 [
21 [
22 [7.7,7.47,7.75,7.08,7.64,7.53,7.66,7.39,7.27,7.31,7.15,7.61,7.34,7.46,7.72,7.4,7.62,7.46], # Fase general y especifica , total, convocatoria total, 2022
23 [7.83,7.6,7.89,7.25,7.78,7.67,7.8,7.54,7.33,7.46,7.33,7.8,7.47,7.59,7.84,7.5,7.7,7.65], # Fase general y especifica , total, convocatoria ordinaria, 2022
24 [6.47,6.37,6.39,5.89,6.42,6.23,6.26,6.13,6.05,6.21,6.2,6.39,6.16,6.2,6.53,6.04,6.3,6.02] # Fase general y especifica , total, convocatoria extraordinaria, 2022
25 ],
26 [
27 [7.61,7.35,7.65,7.06,7.57,7.44,7.58,7.34,7.16,7.26,0.0,7.58,7.28,7.29,7.65,7.31,7.55,7.42], # Fase general y especifica , publico, convocatoria total, 2022
28 [7.75,7.51,7.79,7.22,7.7,7.59,7.74,7.49,7.22,7.41,0.0,7.77,7.42,7.44,7.78,7.42,7.66,7.63], # Fase general y especifica , publico, convocatoria ordinaria, 2022
29 [6.34,6.24,6.49,5.93,6.26,6.15,6.25,6.09,6.0,6.17,0.0,6.31,6.1,6.16,6.46,6.08,6.25,6.03]# Fase general y especifica , publico, convocatoria extraordinaria, 2022
30 ],
31 [
32 [7.97,8.07,7.96,0.0,8.09,0.0,7.86,8.05,7.46,7.54,0.0,8.13,7.55,7.68,8.01,0.0,0.0,8.09], # Fase general y especifica , privado, convocatoria total, 2022
33 [8.07,8.12,8.09,0.0,8.25,0.0,7.96,8.22,7.51,7.68,0.0,8.18,7.65,7.77,8.16,0.0,0.0,8.11], # Fase general y especifica , privado, convocatoria ordinaria, 2022
34 [6.71,6.51,6.05,0.0,6.19,0.0,6.3,7.0,6.16,6.29,0.0,6.94,6.43,6.27,6.81,0.0,0.0,7.64] # Fase general y especifica , privado, convocatoria extraordinaria, 2022
35 ],
36 [
37 [0.0,7.67,0.0,0.0,7.91,7.91,7.8,7.61,6.73,7.41,0.0,7.73,0.0,7.41,7.91,7.52,7.69,7.52], # Fase general y especifica , concertado, convocatoria total, 2022
38 [0.0,7.74,0.0,0.0,7.98,7.98,7.9,7.76,6.73,7.52,0.0,7.87,0.0,7.52,8.03,7.59,7.74,7.65], # Fase general y especifica , concertado, convocatoria ordinaria, 2022
39 [0.0,6.84,0.0,0.0,6.67,6.88,6.25,6.12,0.0,6.33,0.0,6.79,0.0,6.15,6.74,5.96,6.38,5.82]# Fase general y especifica , concertado, convocatoria extraordinaria, 2022
40 ],
41 ]
42 ]
43 ]
```

Figura A.1: Definición de columnas y valores de las observaciones

```
8
9 # Generar las combinaciones y las notas correspondientes para 2022
10 for tipo_conv in tipos_convocatoria:
11     for centro in tipo_centro:
12         for periodo in periodo_2022:
13             for comunidad in comunidades:
14                 for tipo_matr in tipos_matricula:
15                     # Obtener las notas correspondientes a cada combinación
16                     nota = notas_acceso_2022[tipos_convocatoria.index(tipo_conv)][tipo_centro.index(centro)][tipos_matricula.index(tipo_matr)][comunidades.index(comunidad)
17                     data_2022.append([tipo_conv, centro, periodo, comunidad, tipo_matr, nota])
```

Figura A.2: Bucle for para generar las observaciones en Python

	Convocatoria	Centro	Periodo	Comunidad Autónoma	Matrícula	Nota
0	Fase general y específica	Total	2022	Andalucía	Total	7.7
1	Fase general y específica	Total	2022	Andalucía	Ordinaria	7.83
2	Fase general y específica	Total	2022	Andalucía	Extraordinaria	6.47
3	Fase general y específica	Total	2022	Aragón	Total	7.47
4	Fase general y específica	Total	2022	Aragón	Ordinaria	7.6
..
535	Fase general	No consta	2018	País Vasco	Ordinaria	
536	Fase general	No consta	2018	País Vasco	Extraordinaria	
537	Fase general	No consta	2018	La Rioja	Total	7.14
538	Fase general	No consta	2018	La Rioja	Ordinaria	7.14
539	Fase general	No consta	2018	La Rioja	Extraordinaria	

Figura A.3: Observaciones exportadas a un archivo CSV desde Python

Estadística de las Pruebas de Acceso a la Universidad (EPAU)										
Resultados por comunidad autónoma de realización										
Matriculados, presentados y aprobados en las PAU por procedimiento de acceso, convocatoria y sexo										
Unidades:										
										Matriculados 2022
Total PAU genérica	Andalucía PAU genérica	Aragón PAU genérica	Asturias (Principado de) PAU genérica	Balears (Illes) PAU genérica	Canarias PAU genérica	Cantabria PAU genérica	Castilla-La Mancha PAU genérica	Castilla y León PAU genérica	Cataluña PAU genérica	Ciudad de Ceuta y Melilla PAU genérica

Notas:

- Las PAU genéricas la realizan: estudiantes de título de bachiller o equivalente, estudiantes con el título superior de formación profesional, de técnico superior de artes plásticas y diseño o de técnico deportivo superior o equivalente, estudiantes de otros centros de enseñanza superior.
- Matriculados: Recuento de las matriculaciones que se realizan en las pruebas de acceso a la universidad
- Presentados: Recuento de los presentados en las pruebas de acceso a la universidad
- Aprobados: Recuento de los aprobados en las pruebas de acceso a la universidad
- (.): Dato no disponible. La Comunidad autónoma de Andalucía no envía información sobre los estudiantes aprobados en el procedimiento de acceso para mayores de 40 años con experiencia laboral o profesional.
- (...): Dato omitido para preservar el secreto estadístico.

Figura A.4: Vista inicial de los datos descargados en formato xls

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Total	Total	Total	Total	Total	Total	Total	Total	Total	Total						
Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos	Ambos sexos						
Matriculados	Matriculados	Matriculados	Matriculados	Matriculados	Matriculados	Matriculados	Matriculados	Matriculados	Matriculados						
2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
Andalucía	Aragón	Asturias	Balears	Canarias	Cantabria	Castilla-La Mancha	Castilla y León	Cataluña	Comunitat Valenciana	Estado	Extremadura	Galicia	Madrid	Murcia	Navarra
PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica	PAU genérica						
58.060	7.719	5.410	5.316	13.372	3.329	10.213	12.635	45.915	26.729	8.916	6.465	15.359	44.430	10.301	4.053

Figura A.5: Formato final de los datos tras replicar las variables

```

#matriculados, presentados y aprobados 2019
data2019 <- read.csv("C:/Users/iaranzadibouzas/OneDrive - Deloitte (0365D)/Documents/TFG estadística/datos/indicador/seleccion2019.csv")
traspuesta2019<-t(data2019)
traspuesta2019<-as.data.frame(traspuesta2019)
traspuesta2019

#matriculados, presentados y aprobados 2018
data2018 <- read.csv("C:/Users/iaranzadibouzas/OneDrive - Deloitte (0365D)/Documents/TFG estadística/datos/indicador/seleccion2018.csv")
traspuesta2018<-t(data2018)
traspuesta2018<-as.data.frame(traspuesta2018)
traspuesta2018
|

# Lista de data frames
dfs <- list(traspuesta2022, traspuesta2021, traspuesta2020, traspuesta2019, traspuesta2018)

# Combinar todos usando rbind()
resultado <- do.call(rbind, dfs)

colnames(resultado) <- c("Matriculada", "Sexo", "Indicador", "Periodo", "Comunidad", "Acceso", "Cifra")
resultado
write.csv(resultado, "C:/Users/iaranzadibouzas/OneDrive - Deloitte (0365D)/Documents/TFG estadística/datos/indicador/matriculados_presc
getwd()

```

Figura A.6: Código en R para transformar, transponer y unificar los datos

	Matriculada	Sexo	Indicador	Periodo	Comunidad	Acceso	Cifra
V1	Total	Ambos	sexos	Matriculados	2022	Andalucía	PAU generica 58.060
V2	Total	Ambos	sexos	Matriculados	2022	Aragón	PAU generica 7.719
V3	Total	Ambos	sexos	Matriculados	2022	Asturias	PAU generica 5.410
V4	Total	Ambos	sexos	Matriculados	2022	Baleares	PAU generica 5.316
V5	Total	Ambos	sexos	Matriculados	2022	Canarias	PAU generica 13.372
V6	Total	Ambos	sexos	Matriculados	2022	Cantabria	PAU generica 3.329
V7	Total	Ambos	sexos	Matriculados	2022	Castilla-La Mancha	PAU generica 10.213
V8	Total	Ambos	sexos	Matriculados	2022	Castilla y León	PAU generica 12.635
V9	Total	Ambos	sexos	Matriculados	2022	Cataluña	PAU generica 45.915
V10	Total	Ambos	sexos	Matriculados	2022	Comunitat Valenciana	PAU generica 26.729
V11	Total	Ambos	sexos	Matriculados	2022	Estado	PAU generica 8.916
V12	Total	Ambos	sexos	Matriculados	2022	Extremadura	PAU generica 6.465
V13	Total	Ambos	sexos	Matriculados	2022	Galicia	PAU generica 15.359
V14	Total	Ambos	sexos	Matriculados	2022	Madrid	PAU generica 44.430
V15	Total	Ambos	sexos	Matriculados	2022	Murcia	PAU generica 10.301
V16	Total	Ambos	sexos	Matriculados	2022	Navarra	PAU generica 4.053
V17	Total	Ambos	sexos	Matriculados	2022	País Vasco	PAU generica 14.776
V18	Total	Ambos	sexos	Matriculados	2022	La Rioja	PAU generica 1.759
V19	Total	Ambos	sexos	Presentados	2022	Andalucía	PAU generica 56.563
V20	Total	Ambos	sexos	Presentados	2022	Aragón	PAU generica 7.663
V21	Total	Ambos	sexos	Presentados	2022	Asturias	PAU generica 5.354
V22	Total	Ambos	sexos	Presentados	2022	Baleares	PAU generica 5.238
V23	Total	Ambos	sexos	Presentados	2022	Canarias	PAU generica 13.019

Figura A.7: alida final de los datos exportados en CSV

Anexo B

Imágenes de la aplicación *VisionEbau*



Figura B.1: Página de Inicio



Figura B.2: Menú de navegación

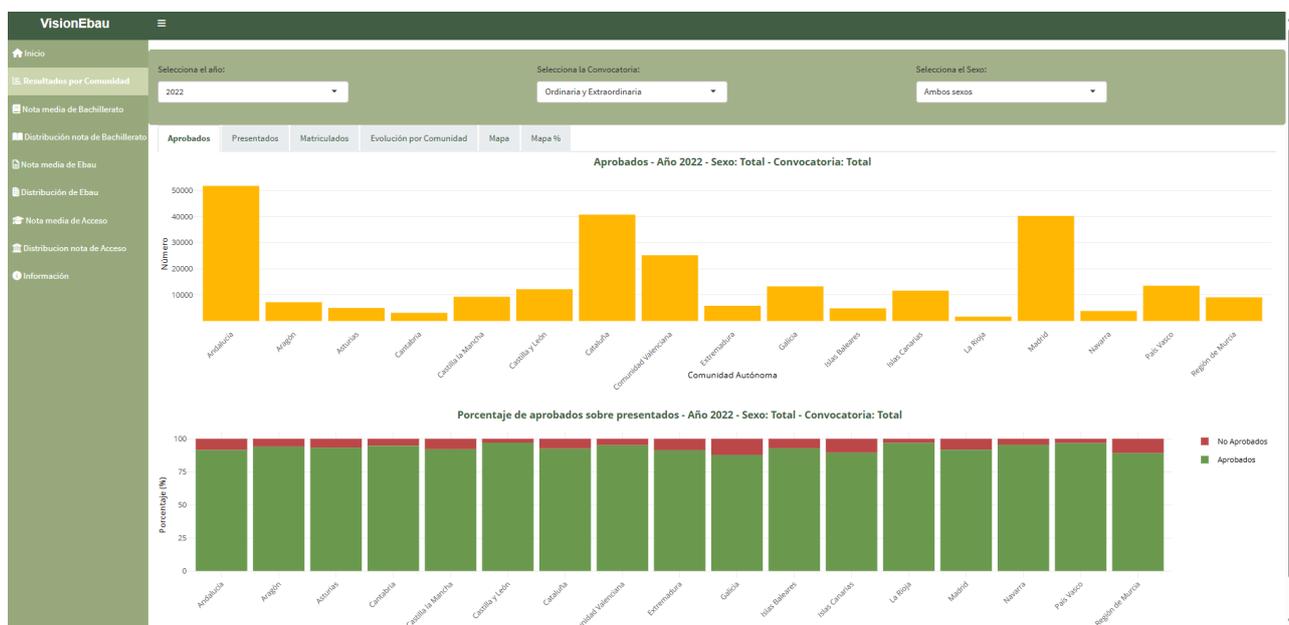


Figura B.3: Página de Resultados por Comunidad. Pestaña de *Aprobados*

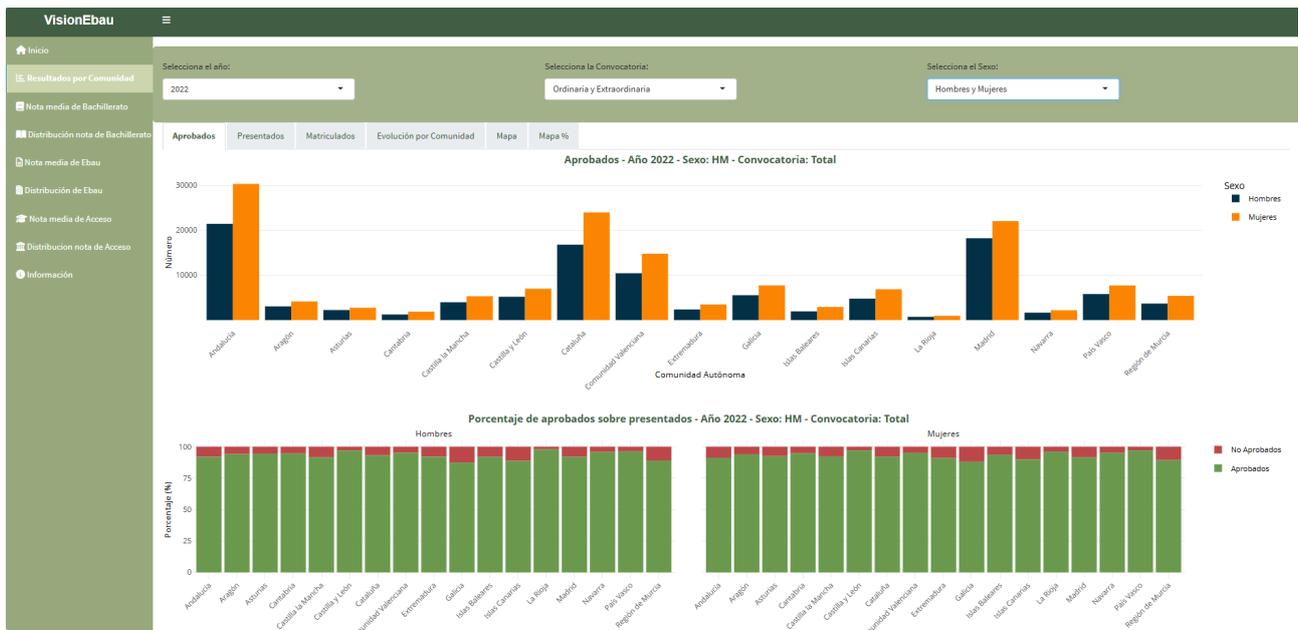


Figura B.4: Página de Resultados por Comunidad. Pestaña de *Aprobados*. Hombres y Mujeres



Figura B.5: Página de Resultados por Comunidad. Pestaña de *Presentados*

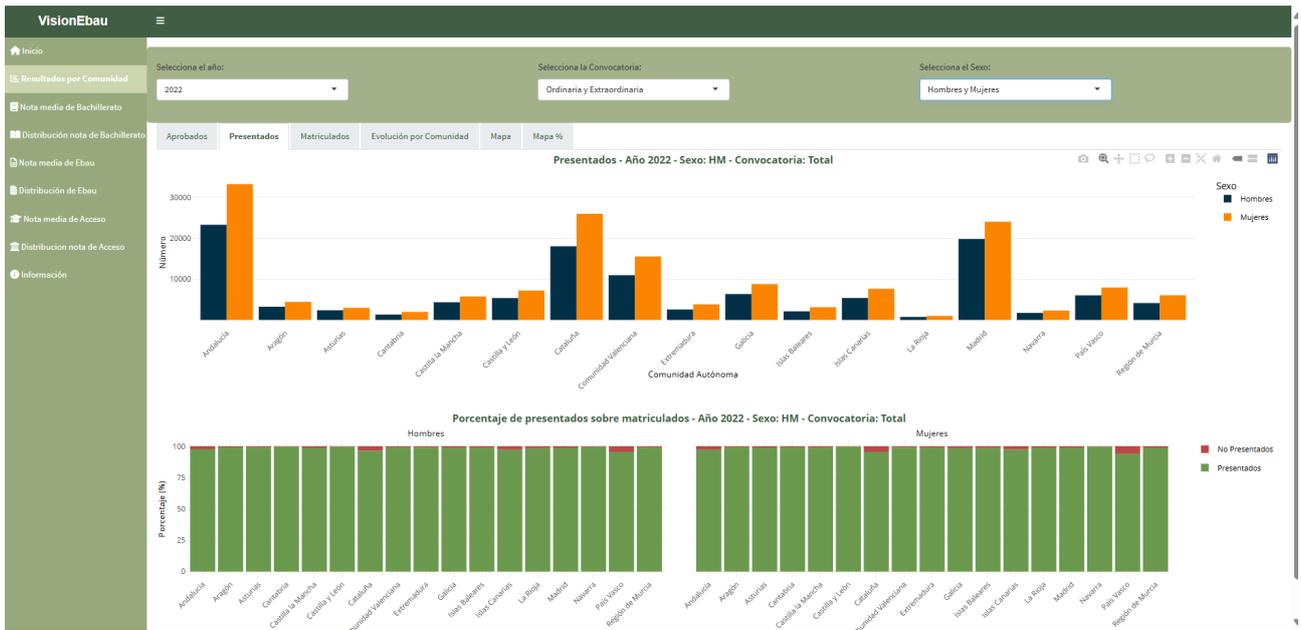


Figura B.6: Página de Resultados por Comunidad. Pestaña de *Presentados*. Hombres y Mujeres

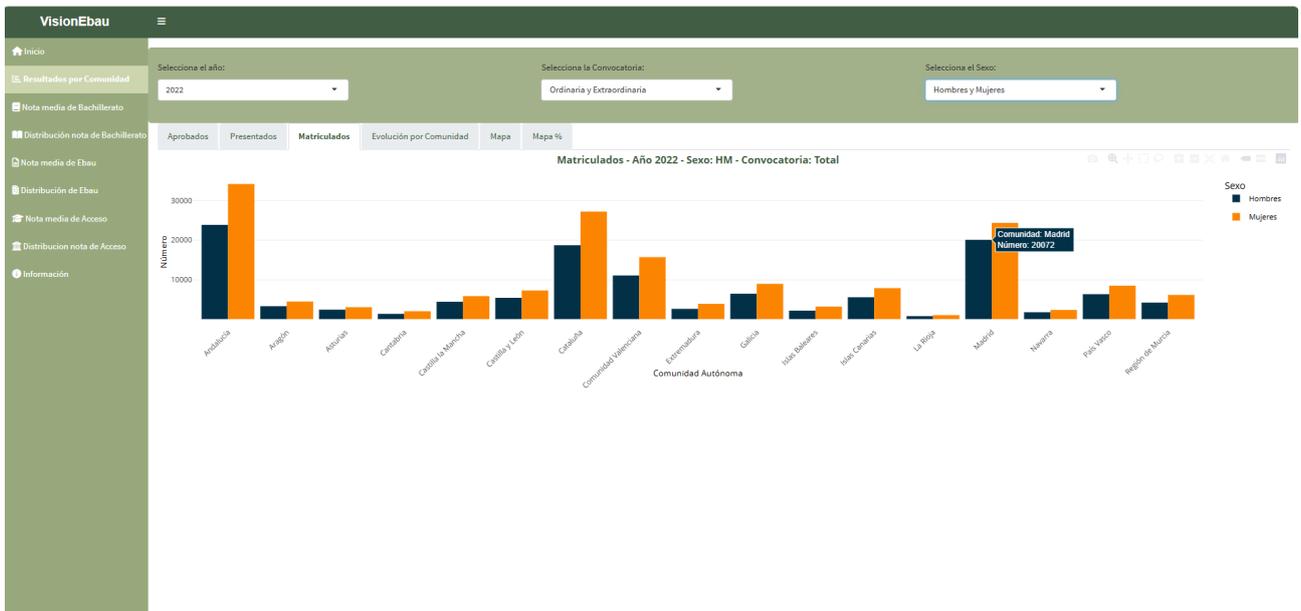


Figura B.7: Página de Resultados por Comunidad. Pestaña de *Matriculados*. Hombres y Mujeres

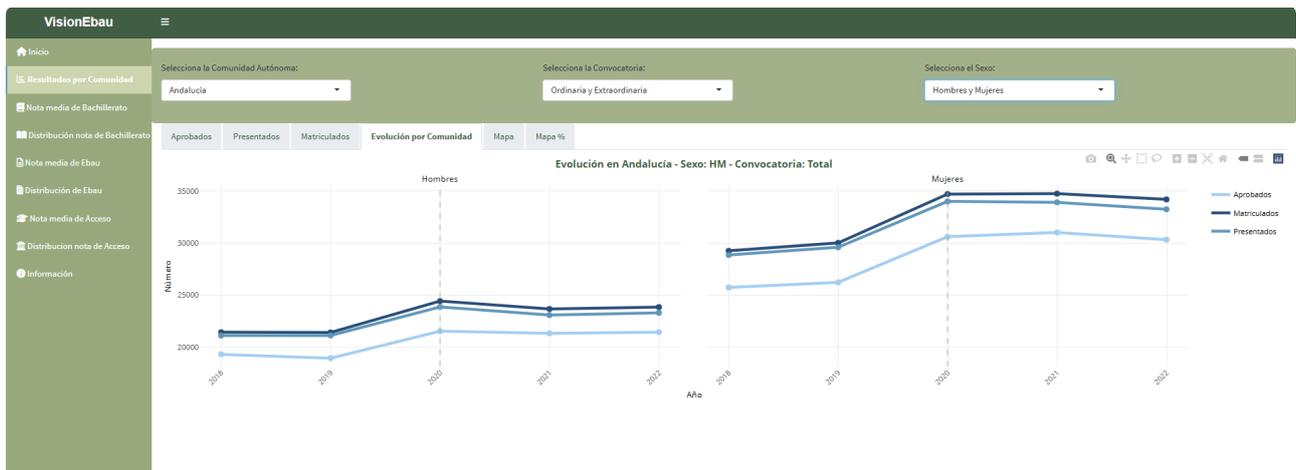


Figura B.8: Página de Resultados por Comunidad. Pestaña de *Evolución por Comunidad*. Hombres y Mujeres

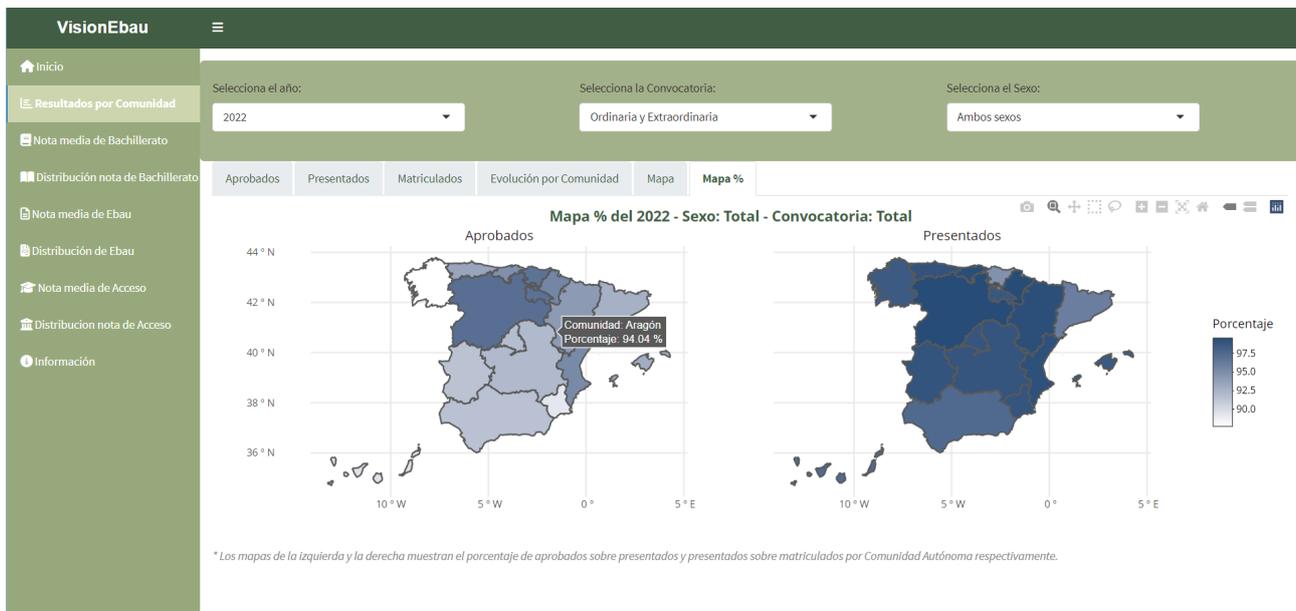


Figura B.9: Página de Resultados por Comunidad. Pestaña de *Mapa %*

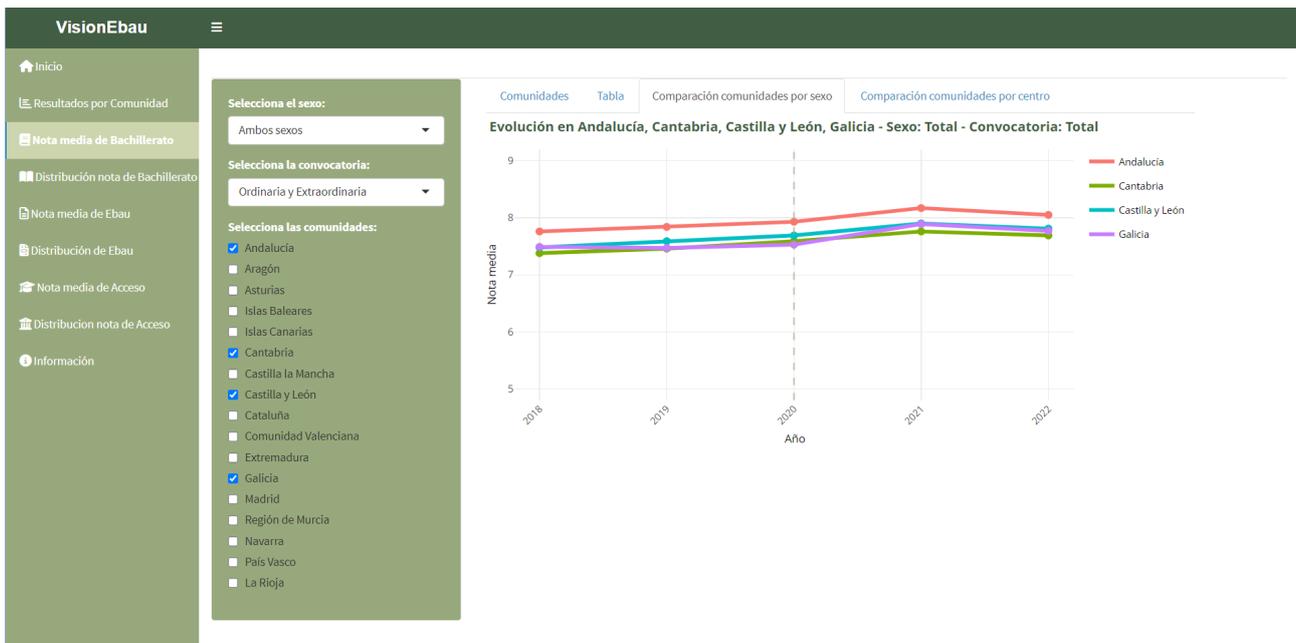


Figura B.10: Página de Nota media de Bachillerato. Pestaña de *Comparación comunidades por sexo*

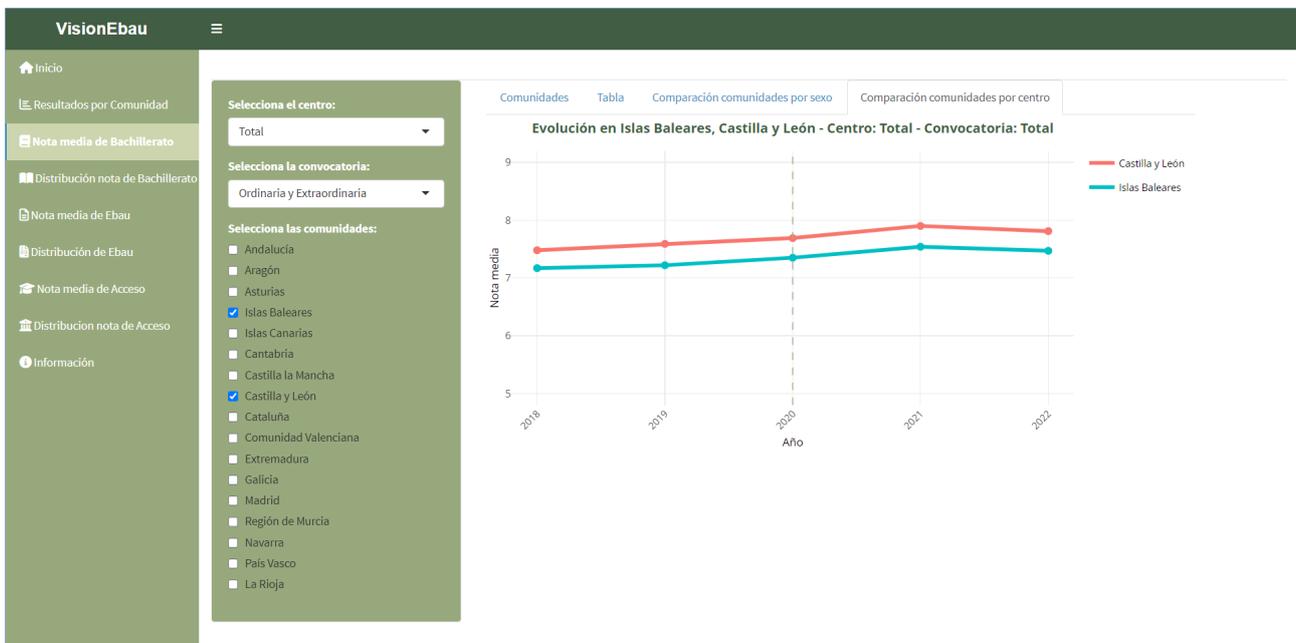


Figura B.11: Página de Nota media de Bachillerato. Pestaña de *Comparación comunidades por centro*



Figura B.12: Página de Nota media de Bachillerato. Pestaña de *Comparación comunidades por centro*

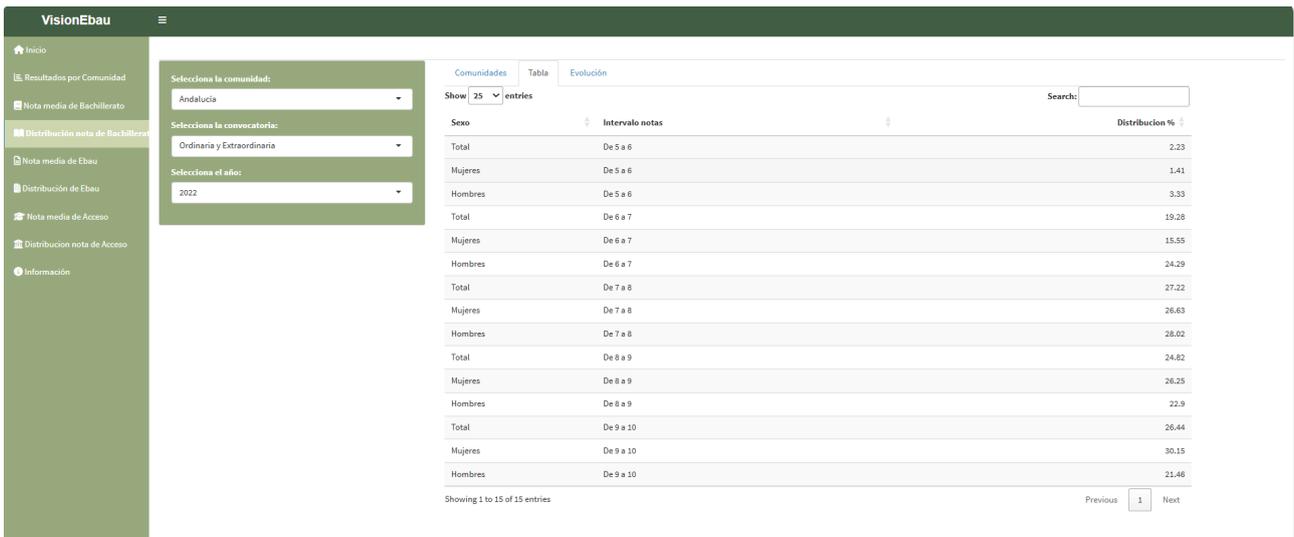


Figura B.13: Página de Distribución nota de Bachillerato. Pestaña de *Tabla*

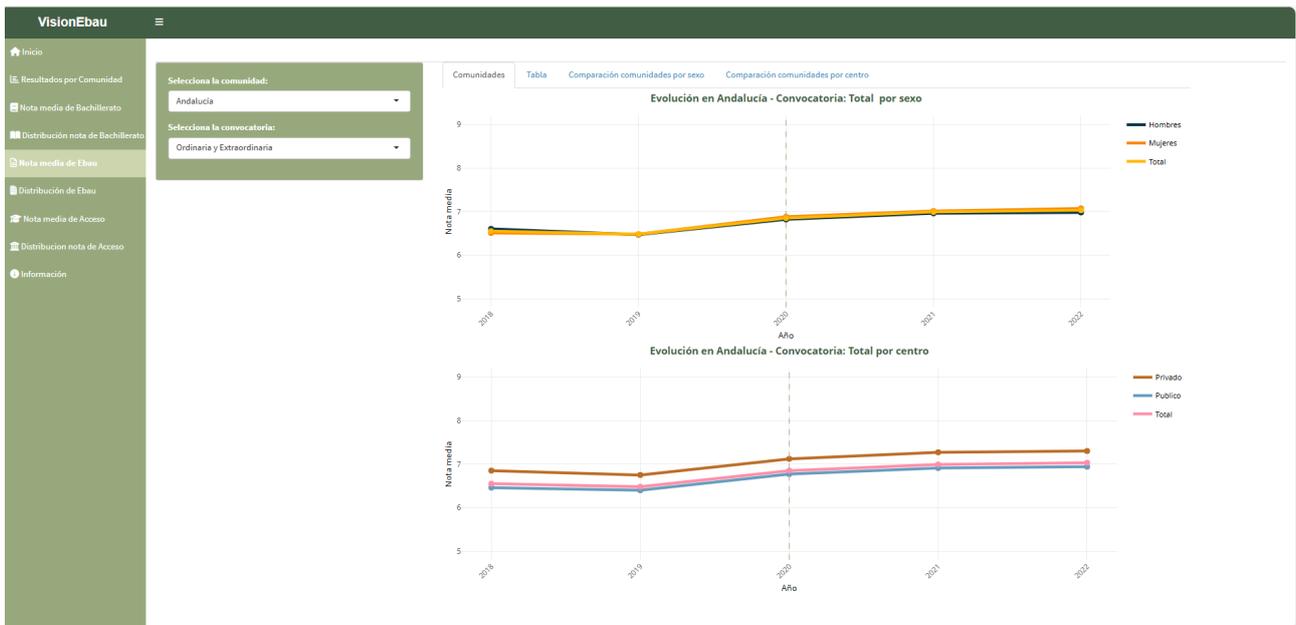


Figura B.14: Página de Nota media de Ebau. Pestaña de *Comunidades*

Comunidades Tabla Comparación comunidades por sexo Comparación comunidades por centro

Show 10 entries Search:

Sexo	Período	Nota media
Hombres	2018	6.6
Mujeres	2018	6.51
Total	2018	6.55
Hombres	2019	6.47
Mujeres	2019	6.48
Total	2019	6.48
Hombres	2020	6.82
Mujeres	2020	6.88
Total	2020	6.85
Hombres	2021	6.96

Showing 1 to 10 of 15 entries Previous 1 2 Next Search:

Centro	Período	Nota media
Privado	2018	6.85
Publico	2018	6.46
Total	2018	6.55
Privado	2019	6.75
Publico	2019	6.4
Total	2019	6.48
Privado	2020	7.12
Publico	2020	6.77

27.0.0.16438/#shiny-tab-paginainicio

Figura B.15: Página de Nota media de Ebau. Pestaña de *Tabla*

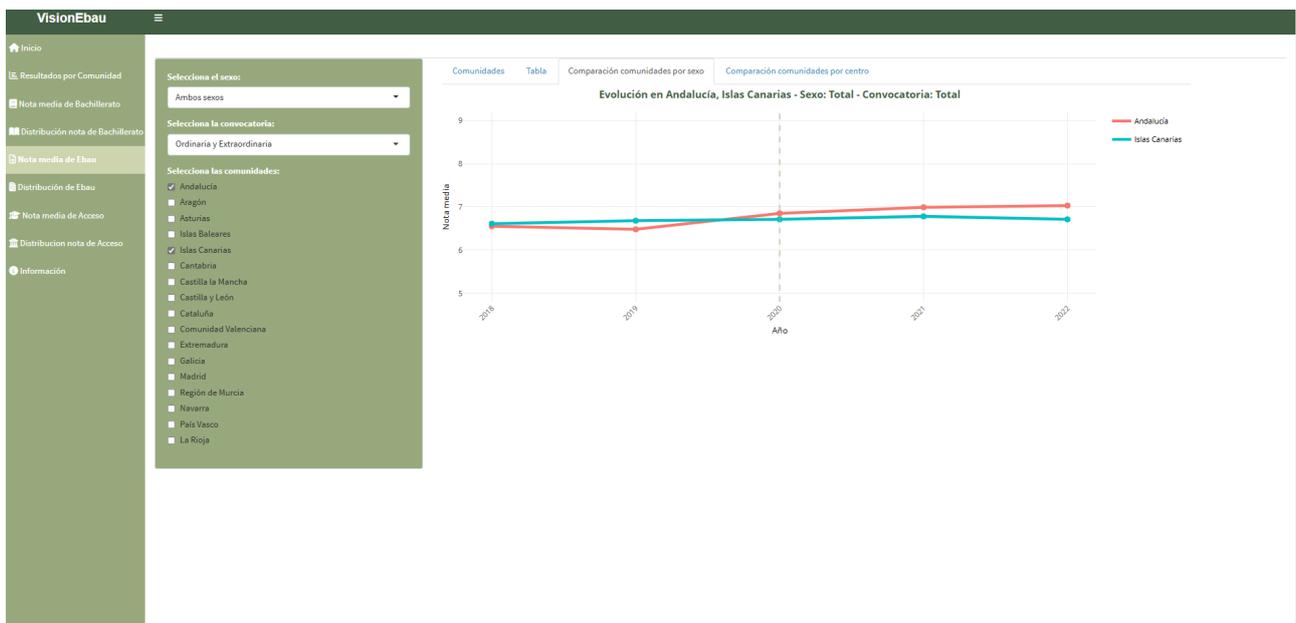


Figura B.16: Página de Nota media de Ebaeu. Pestaña de *Comparación comunidades por sexo*

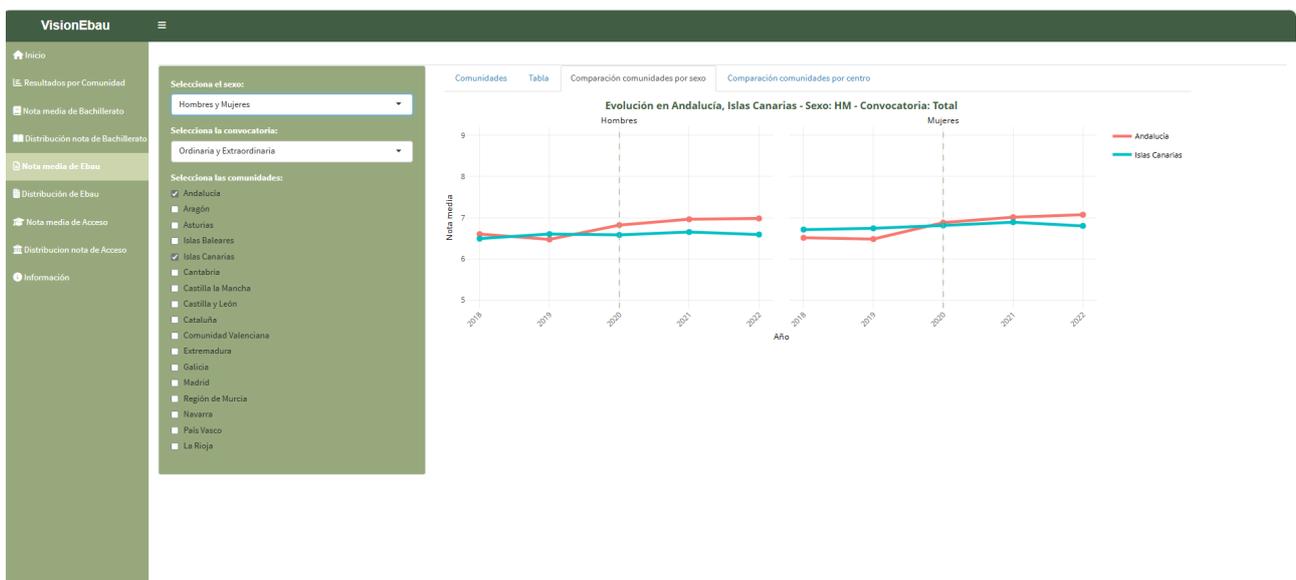


Figura B.17: Página de Nota media de Ebaeu. Pestaña de *Comparación comunidades por sexo*

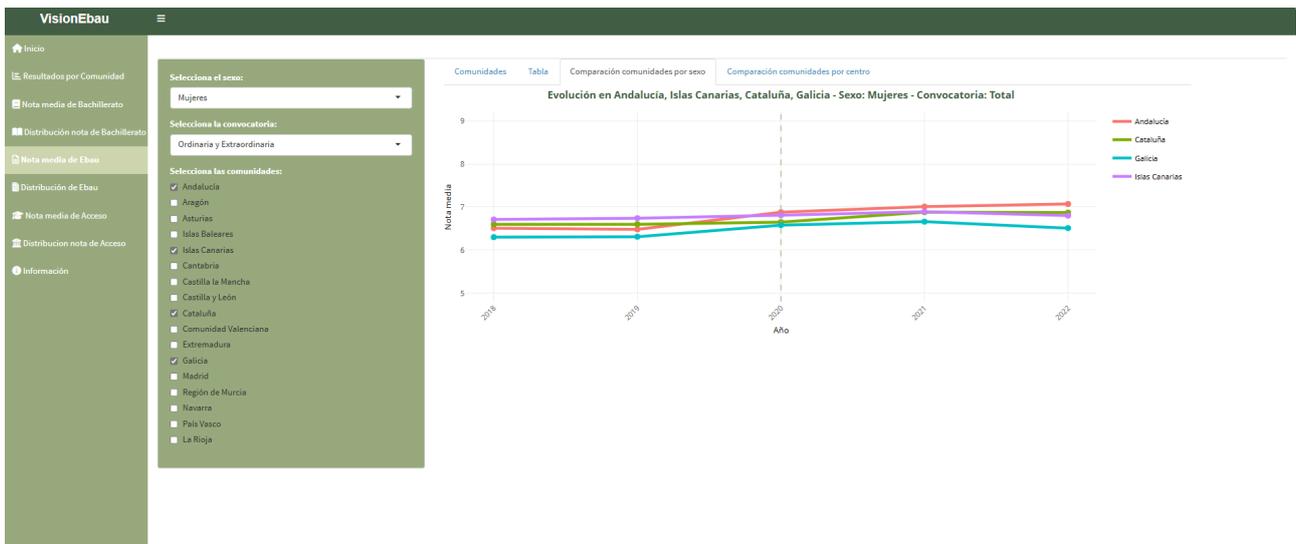


Figura B.18: Página de Nota media de Ebau. Pestaña de *Comparación comunidades por sexo*

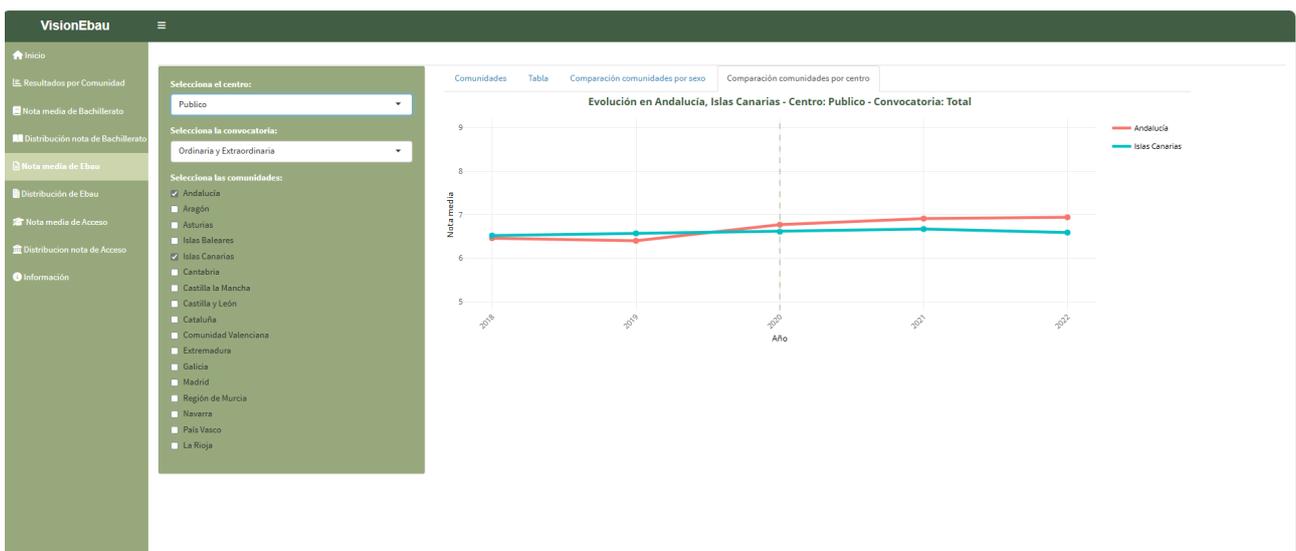


Figura B.19: Página de Nota media de Ebau. Pestaña de *Comparación comunidades por centro*

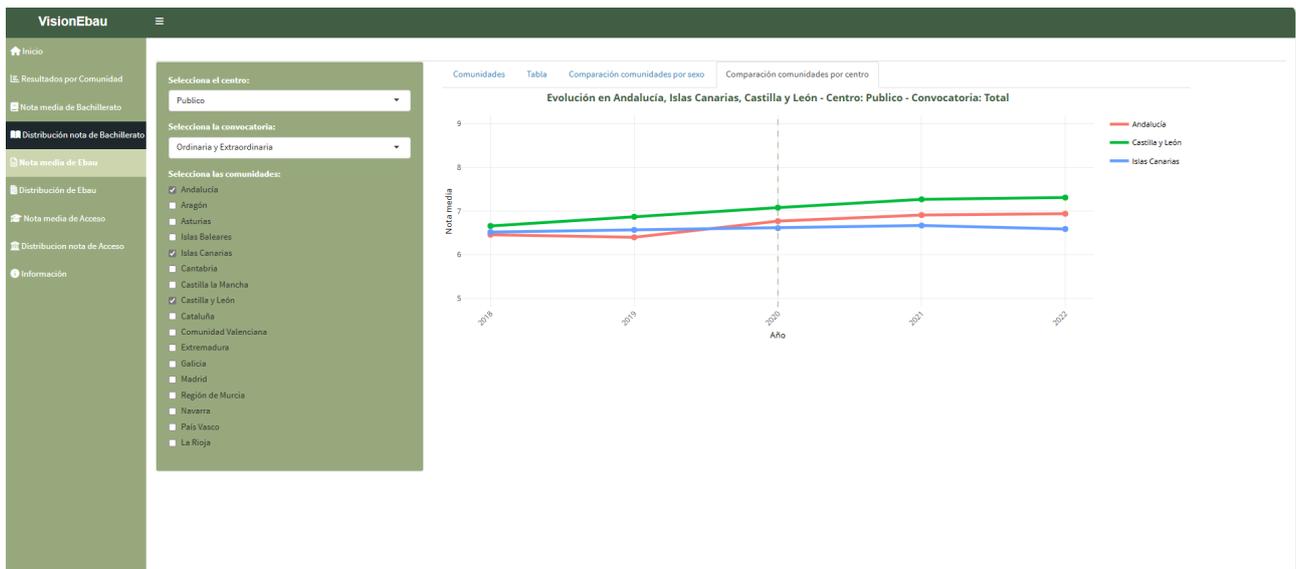


Figura B.20: Página de Nota media de Ebau. Pestaña de *Comparación comunidades por centro*

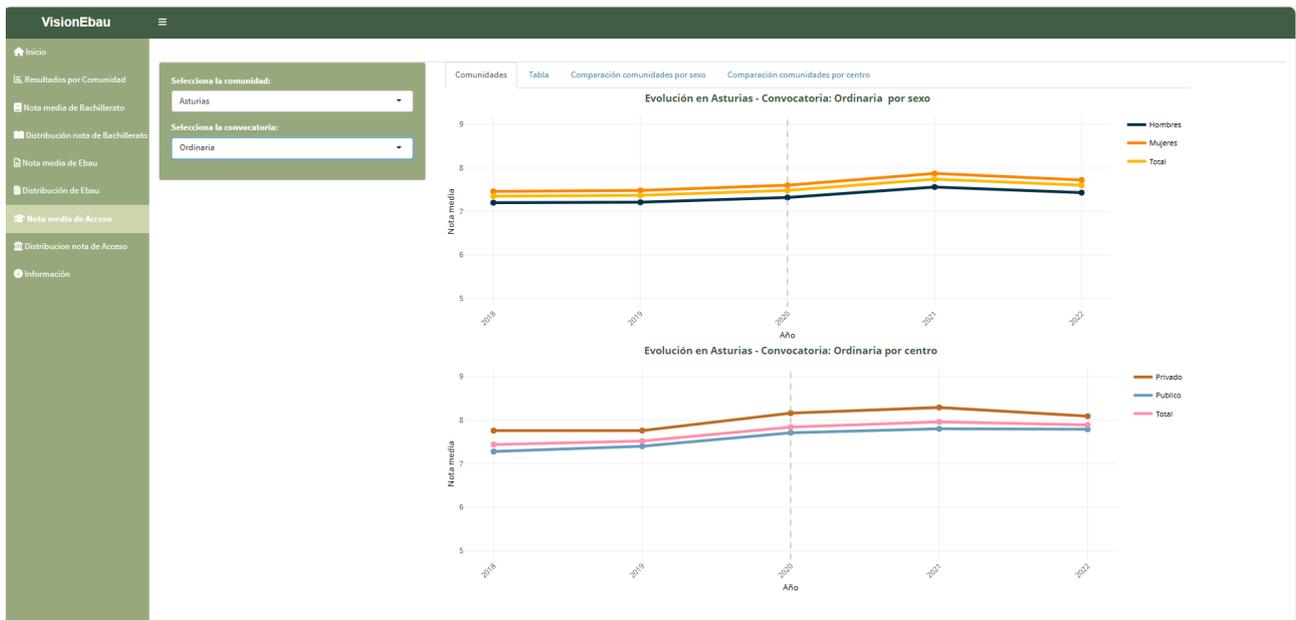


Figura B.21: Página de Nota media de Acceso. Pestaña de *Comunidades*

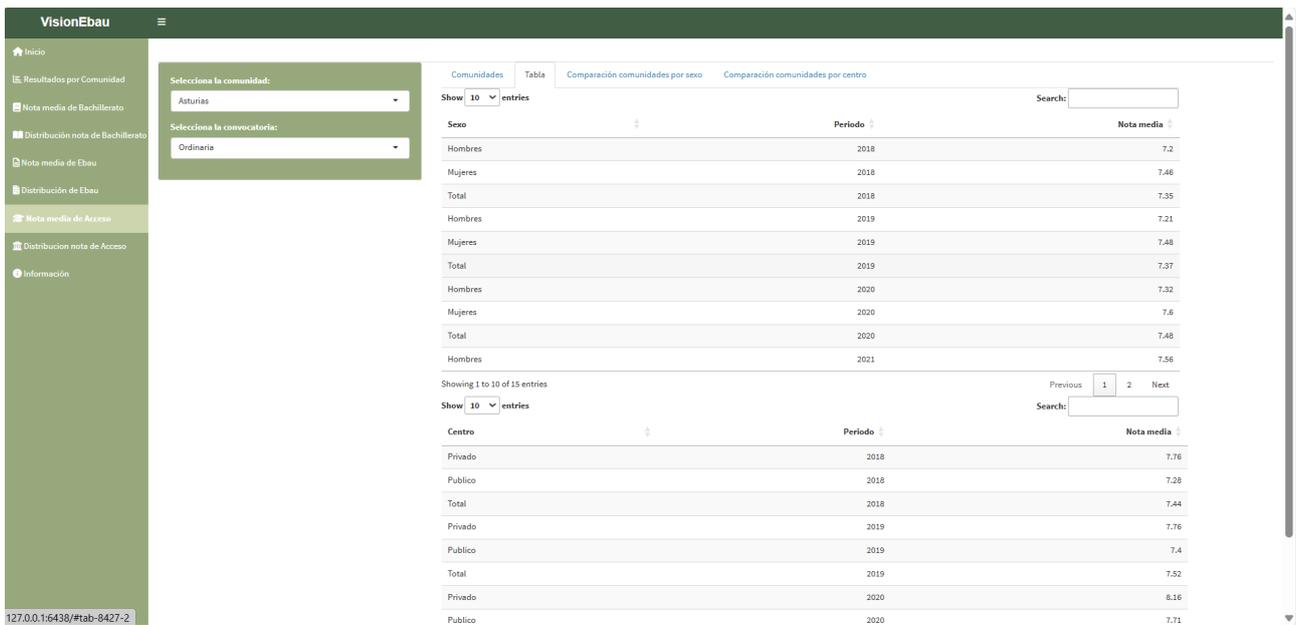


Figura B.22: Página de Nota media de Acceso. Pestaña de *Tabla*

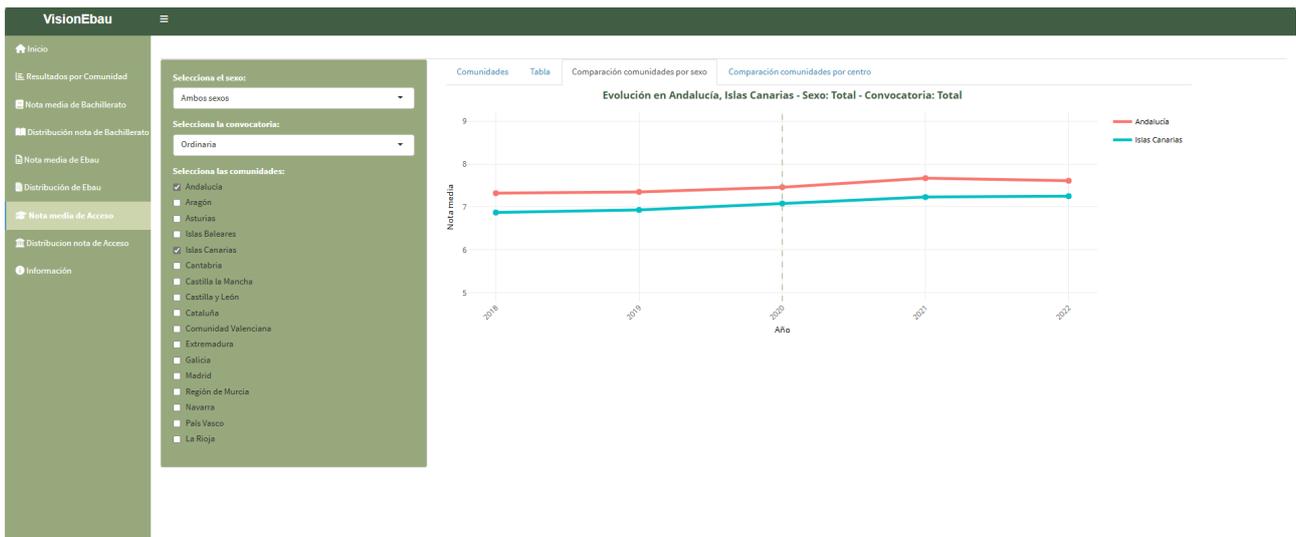


Figura B.23: Página de Nota media de Acceso. Pestaña de *Comparación comunidades por sexo*

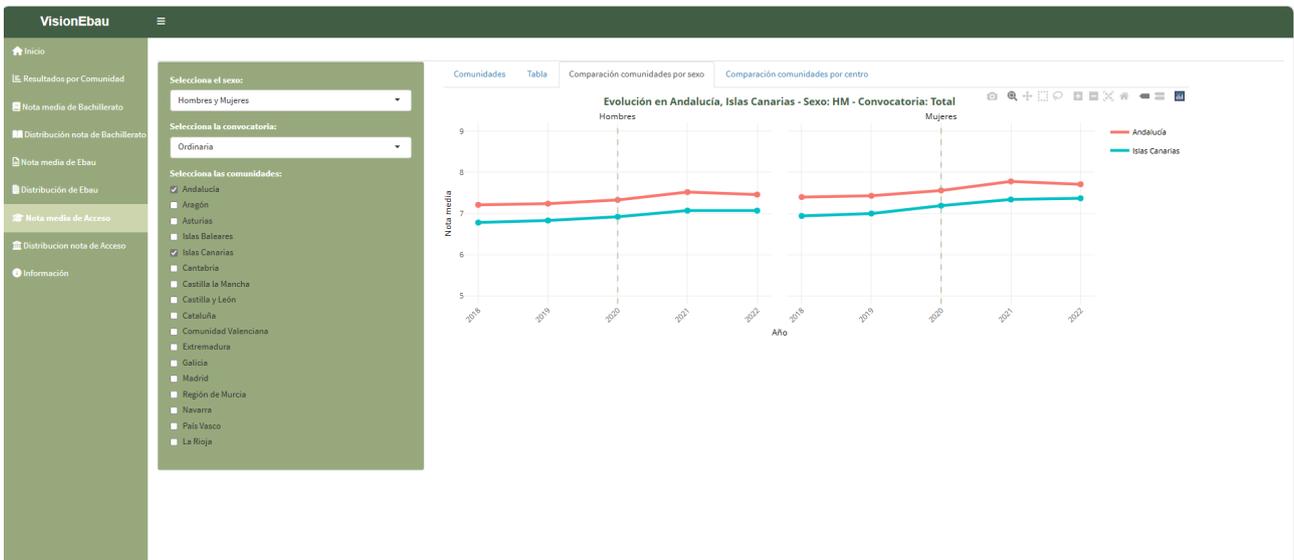


Figura B.24: Página de Nota media de Acceso. Pestaña de *Comparación comunidades por sexo*

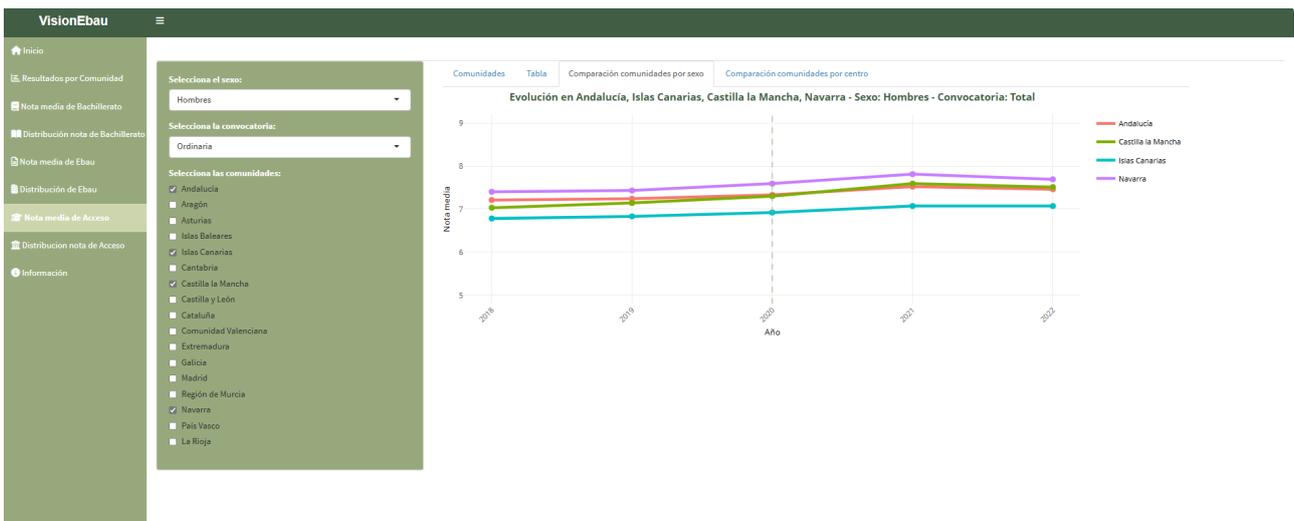


Figura B.25: Página de Nota media de Acceso. Pestaña de *Comparación comunidades por sexo*

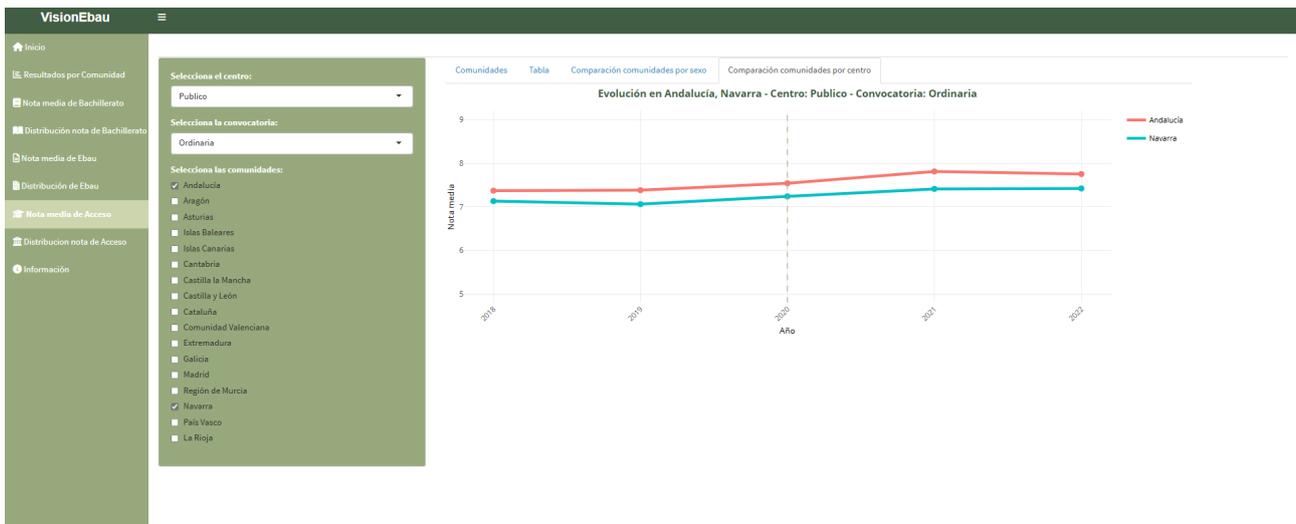


Figura B.26: Página de Nota media de Acceso. Pestaña de *Comparación comunidades por centro*

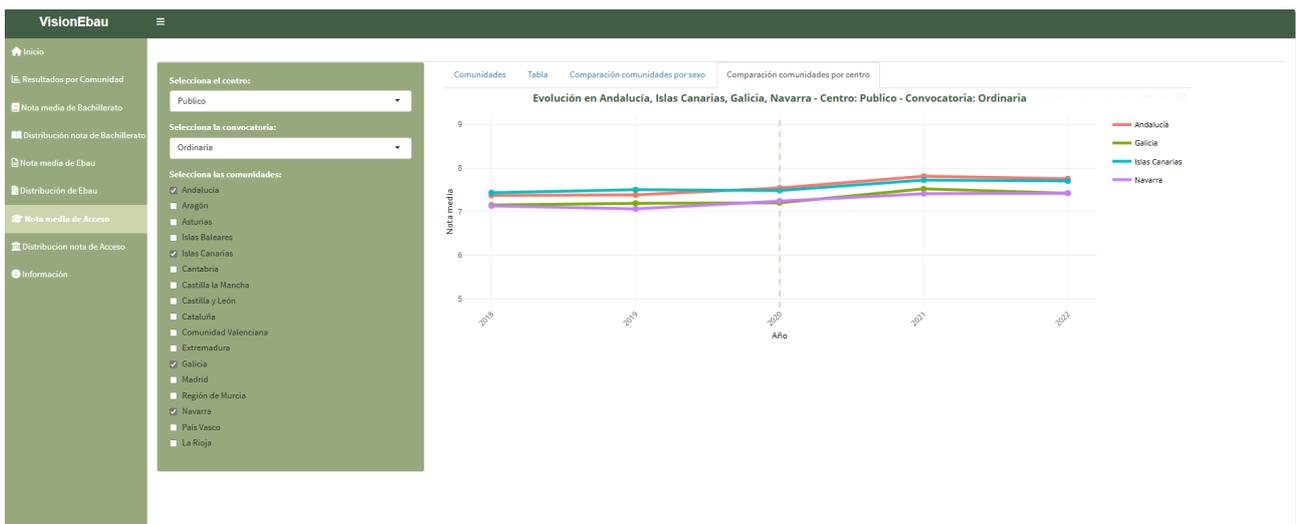


Figura B.27: Página de Nota media de Acceso. Pestaña de *Comparación comunidades por centro*

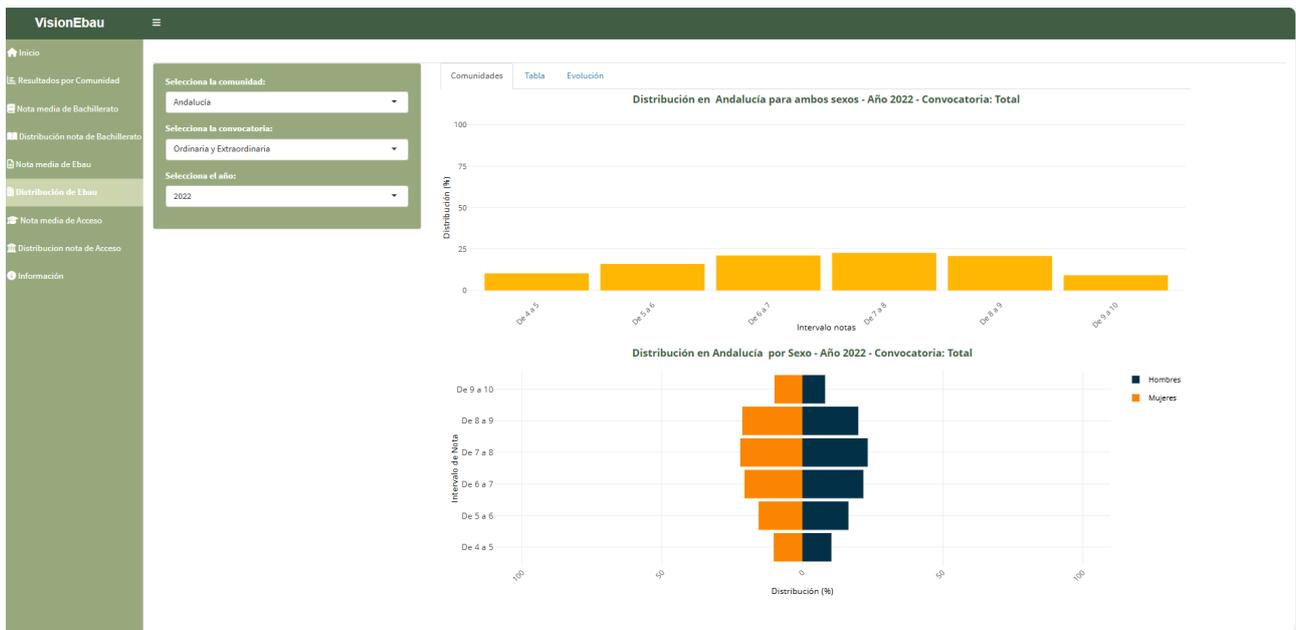


Figura B.28: Página de Distribución nota de Ebau. Pestaña de *Comunidades*

Sexo	Intervalo notas	Distribución %
Total	De 4 a 5	10,28
Mujeres	De 4 a 5	10,21
Hombres	De 4 a 5	10,37
Total	De 5 a 6	15,97
Mujeres	De 5 a 6	15,62
Hombres	De 5 a 6	16,45
Total	De 6 a 7	21,1
Mujeres	De 6 a 7	20,61
Hombres	De 6 a 7	21,77
Total	De 7 a 8	22,67
Mujeres	De 7 a 8	22,17
Hombres	De 7 a 8	23,33
Total	De 8 a 9	20,8
Mujeres	De 8 a 9	21,44
Hombres	De 8 a 9	19,96
Total	De 9 a 10	9,17
Mujeres	De 9 a 10	9,95
Hombres	De 9 a 10	8,12

Figura B.29: Página de Distribución nota de Ebau. Pestaña de *Tabla*

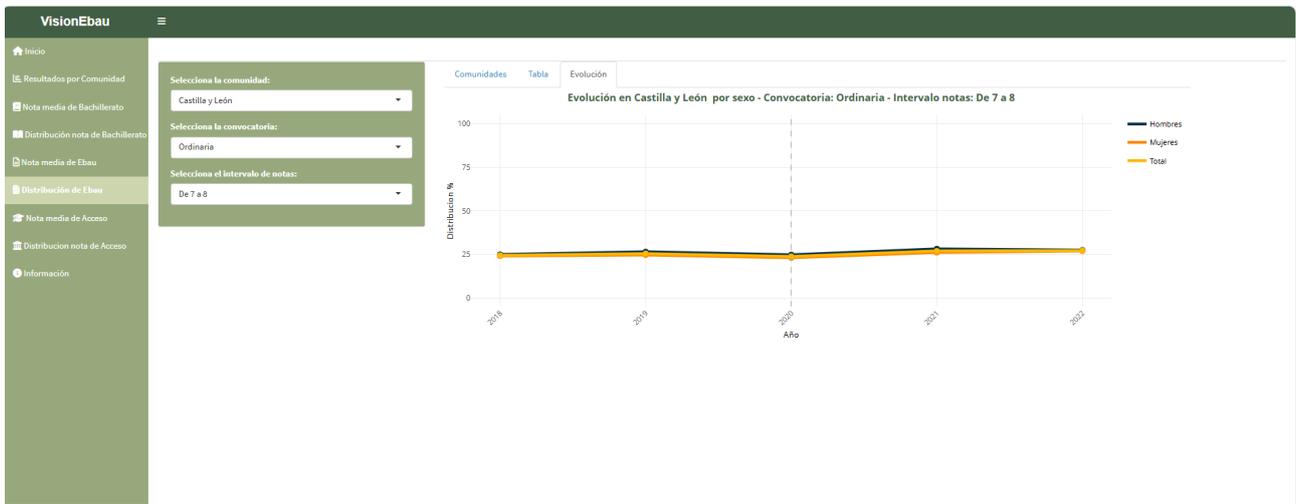


Figura B.30: Página de Distribución nota de Ebau. Pestaña de *Evolución*

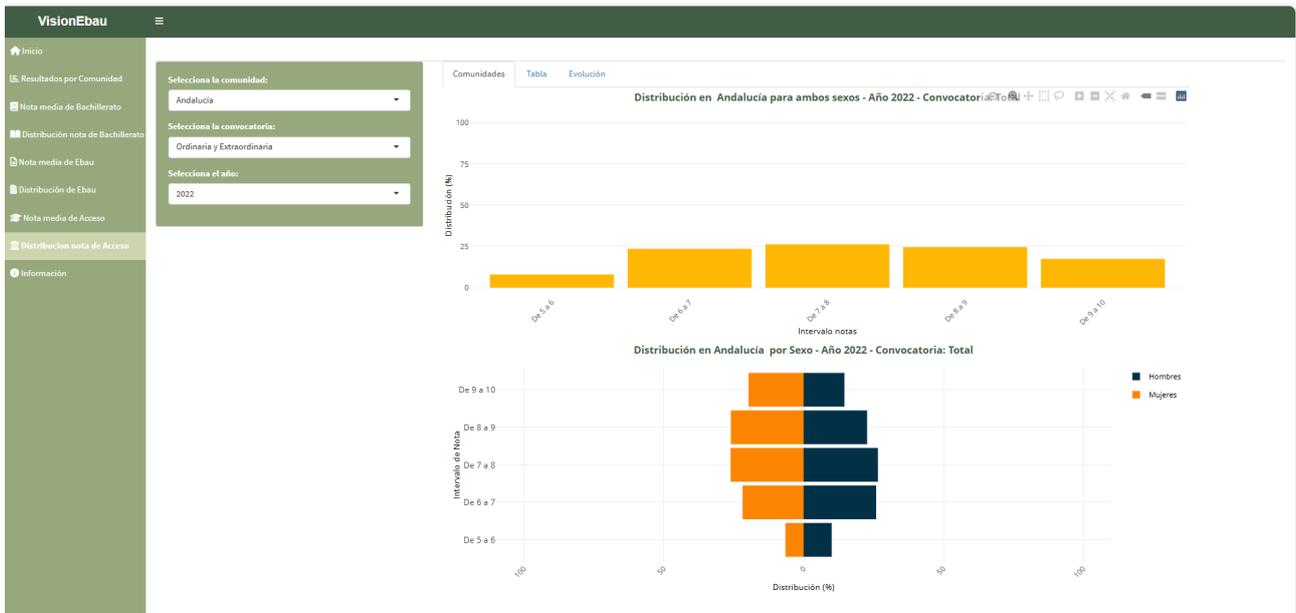


Figura B.31: Página de Distribución nota de Acceso. Pestaña de *Comunidades*

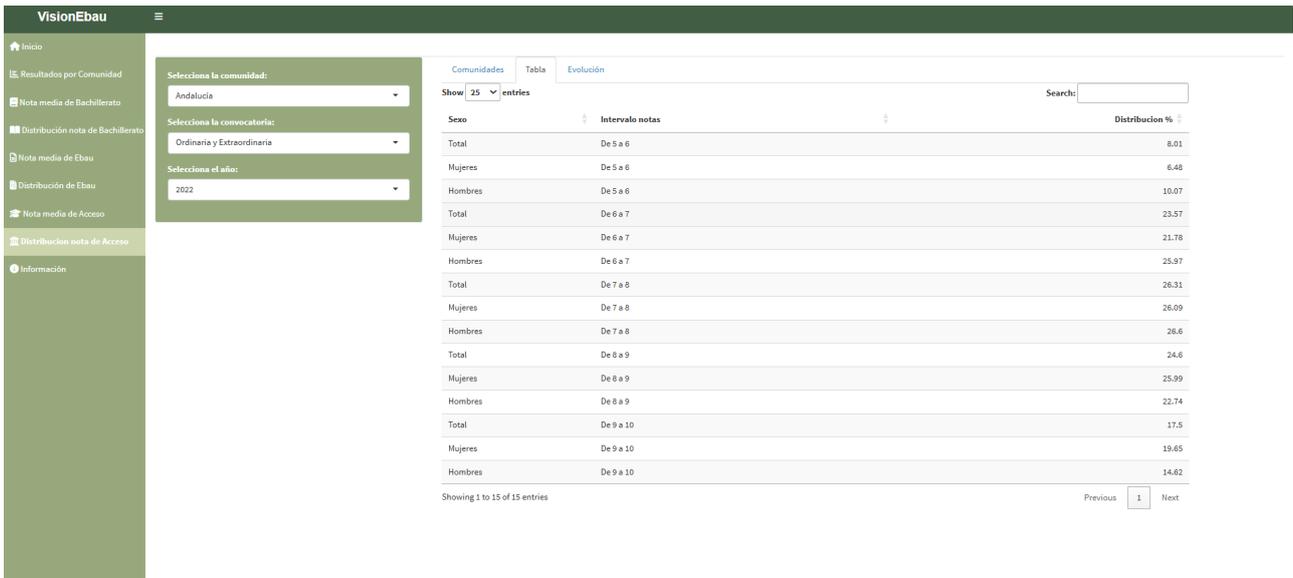


Figura B.32: Página de Distribución nota de Acceso. Pestaña de *Tabla*

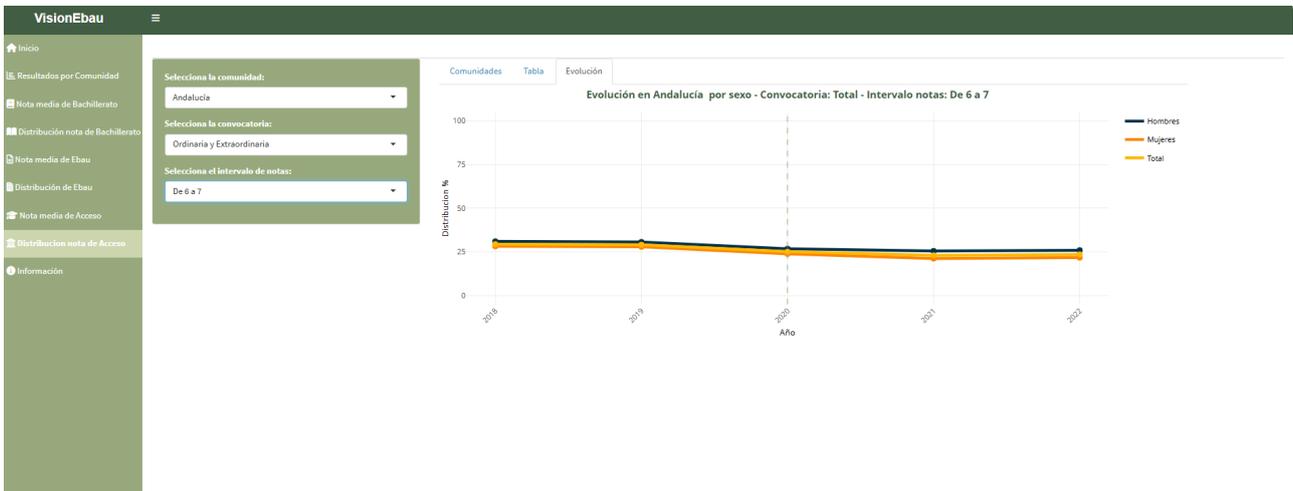


Figura B.33: Página de Distribución nota de Acceso. Pestaña de *Evolución*

Anexo C

Construcción tablas de contingencia

A continuación, se describen de manera detallada los pasos para la construcción de las tablas de contingencia.

Primero, se procede a definir los intervalos de calificaciones y los años de estudio, así como a inicializar una lista para guardar las tablas generadas. El objetivo es construir una tabla de contingencia para cada comunidad autónoma. El siguiente código en R muestra cómo se realiza esta etapa inicial:

```
1 # Definición de intervalos y periodos
2 intervalos <- c('<5', 'De 5 a 6', 'De 6 a 7', 'De 7 a 8', 'De 8 a 9', 'De 9 a 10')
3 periodos <- 2018:2022
4 # Lista para guardar tablas por comunidad
5 tablas_comunidad <- list()
```

Listing C.1: Código en R para la definición de intervalos y períodos

Una vez establecida esta estructura, se procede a rellenar las tablas. Para el intervalo <5 en cada año, la observación que se registrará será el número de presentados menos el número de aprobados, ya que esto corresponde al número de suspensos. Es importante señalar que para este intervalo no se dispone de la distribución de la nota de acceso, dado que el conjunto de datos solo incluye información de los aprobados; por ello, es necesario calcularlo indirectamente.

Para el resto de los intervalos, la observación se calcula utilizando la distribución porcentual disponible para ese año e intervalo, multiplicándola por el número de aprobados correspondiente y dividiéndola entre 100. De esta manera, se obtiene el número estimado de personas que han obtenido una nota comprendida en cada intervalo para cada año.

Estos cálculos se realizan mediante un bucle `for` que recorre cada comunidad autónoma, como se muestra en el siguiente código:

```
1 comunidades <- unique(df$Comunidad)
2 for (com in comunidades) {
3   # Crear un dataframe vacío con los intervalos como filas y los periodos como columnas
4   tabla <- data.frame(Intervalo = intervalos)
5
6   # Crear las columnas para los periodos 2018-2022
7   for (per in periodos) {
8     tabla[[as.character(per)]] <- NA # Inicializamos las columnas con NA
9   }
10
11  # Rellenar la tabla con los valores correspondientes
12  for (per in periodos) {
13    # Presentados y aprobados para <5
14    presentados <- df %>%
```

```

15     filter(Comunidad == com, Periodo == per, Indicador == 'Presentados') %>%
16     summarise(total = sum(Cifra, na.rm = TRUE)) %>%
17     pull(total)
18
19     aprobados <- df %>%
20     filter(Comunidad == com, Periodo == per, Indicador == 'Aprobados') %>%
21     summarise(total = sum(Cifra, na.rm = TRUE)) %>%
22     pull(total)
23
24     # Rellenar el valor de la columna '<5'
25     tabla[tabla$Intervalo == '<5', as.character(per)] <- presentados - aprobados
26
27     # Rellenar los demas intervalos
28     for (int in intervalos[-1]) {
29         dist <- dfacesodis %>%
30         filter(Comunidad == com, Periodo == per, Rango == int) %>%
31         summarise(total = sum(Distribucion, na.rm = TRUE)) %>%
32         pull(total)
33
34         # Calcular la distribucion
35         tabla[tabla$Intervalo == int, as.character(per)] <- round((dist * aprobados) / 100,
36         0)
37     }
38
39     # Asignar la tabla creada a la lista de tablas por comunidad
40     tablas_comunidad[[com]] <- tabla
41 }

```

Listing C.2: Código en R para calcular las tablas de contingencia

El resultado obtenido es una tabla de contingencia para cada comunidad autónoma.

Anexo D

Tablas de residuos estandarizados y gráficos de mosaicos

Se muestran las tablas de residuos estandarizados tras realizar el test chi-cuadrado y los gráficos de mosaicos para cada comunidad autónoma.

hazme una tabla como esta

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	3.73	26.67	18.14	-1.29	-15.10	-29.24
2019	12.98	18.62	14.64	-0.02	-14.57	-28.14
2020	4.35	2.64	-5.11	-4.40	1.91	3.61
2021	-11.60	-27.04	-14.42	3.59	15.56	28.23
2022	-8.77	-18.88	-11.75	2.08	10.86	22.96

Cuadro D.1: Residuos estandarizados para la comunidad de Andalucía

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	-1.0970901	11.8490980	4.2279966	-2.1918981	-5.9261273	-7.0079526
2019	-1.2451928	7.7734802	6.5434538	-3.1010717	-3.6304790	-7.5564453
2020	7.0375529	-0.5719268	-4.4407627	1.2316401	-2.1324739	2.7159729
2021	-6.2235295	-13.7460198	-4.5171752	2.9282142	9.4600242	9.2584417
2022	1.3766267	-4.6073950	-1.4163470	0.9451100	1.9129171	2.0723634

Cuadro D.2: Residuos estandarizados para la comunidad de Aragón.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	12.0590057	11.7774796	5.0846530	-2.7068076	-6.1824676	-14.8127382
2019	10.1531750	6.9019819	3.6116279	-1.8475055	-3.9602028	-11.1414742
2020	-7.7518292	-2.9824736	-5.0627037	0.8728169	4.3671110	8.2804676
2021	-6.6905272	-10.1647037	-2.9389947	1.3073276	3.1172730	11.9229379
2022	-7.0762147	-4.8851466	-0.4081925	2.2346901	2.3340097	4.8765490

Cuadro D.3: Residuos estandarizados para la comunidad de Asturias.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	7.9129403	8.4160035	-1.4498752	-2.9090256	-6.3632382	-8.8888249
2019	6.8888947	4.1028229	2.3373187	-1.7268797	-7.7330658	-7.0527024
2020	-1.1844620	-0.8220644	-1.6684513	-1.4416365	3.1545253	4.4135629
2021	-2.9155853	-6.8649573	0.0590717	1.1937376	4.4176673	7.8739879
2022	-10.2943110	-4.3228216	0.7744103	4.8096247	6.0186621	2.9313307

Cuadro D.4: Residuos estandarizados para la comunidad de las Islas Baleares.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	-7.6672263	11.2031252	9.0042776	0.7279933	-6.1752823	-8.1622847
2019	-5.8037263	6.3810575	4.1387129	1.9158683	-1.3708046	-6.5988021
2020	7.6028447	1.8299528	-1.9597673	-3.0429546	-0.3903569	-1.5670686
2021	1.7385221	-9.7142916	-6.6290823	1.2764359	3.4895559	9.7521073
2022	3.5599403	-8.8918607	-3.9664744	-0.7758633	4.0947812	5.9064431

Cuadro D.5: Residuos estandarizados para la comunidad de las Islas Canarias.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	6.2316943	5.8989841	5.0010703	-2.2104988	-5.0564495	-8.8112729
2019	4.7846599	8.5793516	2.3263562	-1.4694631	-4.1864948	-9.0398167
2020	-1.0902690	-2.3612560	-0.8734019	-2.5716955	2.2731003	5.1065432
2021	-4.9367239	-6.1876721	-4.0158636	1.4760589	5.3342858	7.4074466
2022	-4.5861074	-5.3869523	-2.1693556	4.6565594	1.2844456	4.6648037

Cuadro D.6: Residuos estandarizados para la comunidad de Cantabria.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	-2.79730778	9.28600460	3.70344889	-0.03639441	-4.00789597	-6.25379510
2019	2.46953203	7.07542507	3.84480301	-2.36615645	-2.81777749	-7.32690491
2020	7.65554107	1.30960542	-0.47728485	-2.52762994	-3.68098654	0.21861366
2021	-5.66405974	-10.06673255	-5.10301123	2.09823967	6.23643097	10.74482939
2022	-1.86529665	-7.32433792	-1.80757780	2.86968721	4.22365842	2.33509228

Cuadro D.7: Residuos estandarizados para la comunidad de Extremadura.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	4.9473282	13.9690226	4.9082820	-2.9367948	-7.5453262	-13.0261429
2019	3.0714777	4.5148964	2.8539250	-0.3275780	-3.0773528	-7.1883139
2020	0.8355280	0.3013023	-2.9264425	-2.1685133	1.9807441	3.3137244
2021	-9.1992229	-15.2618871	-5.7437126	2.8994880	9.0427274	17.3715390
2022	0.3820268	-3.3787814	1.0071771	2.5277442	-0.4968951	-0.6691821

Cuadro D.8: Residuos estandarizados para la comunidad de Castilla la Mancha.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	9.972478833	18.193634210	10.200595761	-3.861037538	-11.883678145	-17.623598260
2019	8.431413819	8.523198506	2.441242549	0.005372818	-5.949267251	-9.551765479
2020	4.497411211	0.664220224	-3.269519284	-1.507761471	-0.417295280	3.070767738
2021	-11.422156612	-15.322228747	-5.430650686	4.097707025	8.994222506	12.912055689
2022	-11.272842893	-11.671667026	-3.707937139	1.237222260	9.005834910	10.728052791

Cuadro D.9: Residuos estandarizados para la comunidad de Castilla y León.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	7.4935487	10.6529418	10.8486090	-4.7523038	-11.9251021	-17.6895176
2019	2.4290684	12.4413423	7.1306278	-2.2266306	-11.9659300	-12.1878504
2020	4.9404741	10.7456820	-6.1444831	-3.4225774	-3.1943083	0.6559386
2021	-7.8907177	-19.2883113	-5.5754841	5.2049588	15.9869409	14.2498049
2022	-6.4722305	-13.4226674	-5.4160360	4.8448635	9.9590101	13.5345018

Cuadro D.10: Residuos estandarizados para la comunidad de Cataluña.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	6.2045978	15.6438844	3.4629334	-3.4073858	-11.1702797	-10.4223489
2019	14.6832428	18.3822541	9.3315858	-9.0337051	-15.2289760	-16.2706811
2020	-9.9614677	-2.5752089	-0.1199769	1.2291955	1.6245828	7.8459746
2021	-4.1882483	-18.9621349	-10.9383875	4.7299202	15.5241166	17.2046727
2022	-6.2867943	-11.2446372	-1.1910846	6.0654414	8.2639309	0.7740333

Cuadro D.11: Residuos estandarizados para la comunidad Valenciana.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	4.5323	11.2439	1.6188	-3.9511	-5.7902	-8.6678
2019	6.3066	9.4580	4.5341	-3.3181	-8.9215	-9.7282
2020	1.8415	5.2645	-3.0928	-3.4549	0.6489	0.1742
2021	-11.9192	-14.9817	-2.0989	5.4432	10.8018	13.9754
2022	-0.3108	-10.1763	-0.7040	5.0224	2.6590	3.4682

Cuadro D.12: Residuos estandarizados para la comunidad de Galicia.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	16.7564	21.4578	4.8899	-6.1410	-14.9541	-19.2790
2019	5.4549	13.7623	6.7995	-4.9655	-10.0415	-10.4067
2020	2.2143	-3.2897	-7.3339	0.3684	3.5421	6.9124
2021	-12.8805	-18.0080	-1.1274	6.6569	10.1174	11.8122
2022	-10.6718	-12.5121	-2.7132	3.6483	10.3198	9.7242

Cuadro D.13: Residuos estandarizados para la comunidad de Madrid.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	2.6075	12.4913	8.7056	-0.4316	-8.1178	-12.7641
2019	1.9686	5.1547	5.0264	2.5658	-2.7517	-11.5215
2020	-0.6214	-2.7323	-0.0153	-0.5806	0.4532	2.9399
2021	-8.9798	-7.8397	-7.0934	-0.2932	4.5011	17.1969
2022	5.2065	-6.1638	-5.9439	-1.1787	5.3484	3.0137

Cuadro D.14: Residuos estandarizados para la Región de Murcia.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	4.4980	7.1530	2.0182	-2.3090	-5.2423	-5.0423
2019	0.1299	7.9372	5.7519	-3.1710	-5.9004	-6.3012
2020	-1.4911	-0.8983	-1.0206	0.7351	0.7390	1.8777
2021	-1.0016	-7.1439	-3.5476	2.3787	4.8799	5.1015
2022	-1.8548	-6.1976	-2.7985	2.0673	4.9052	3.7363

Cuadro D.15: Residuos estandarizados para la cominidad de Navarra.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	0.1196	17.0366	11.1270	-1.3653	-10.6571	-15.0683
2019	5.1865	12.3640	9.2974	-2.5694	-7.5035	-13.5783
2020	2.2848	-5.9331	-3.9575	0.9776	2.2811	5.2690
2021	-7.2596	-14.1131	-10.5098	0.9116	10.4003	16.4860
2022	-0.1114	-8.4622	-5.3299	1.9492	4.8993	5.9850

Cuadro D.16: Residuos estandarizados para la cominidad del País Vasco.

Año	<5	De 5 a 6	De 6 a 7	De 7 a 8	De 8 a 9	De 9 a 10
2018	-0.7230	6.4432	3.0069	-3.3042	-2.5463	-4.0228
2019	1.8578	2.8249	2.9978	0.2399	-3.6603	-5.2862
2020	7.2966	3.3537	-2.5158	-1.7611	-3.4866	1.2941
2021	-3.6572	-6.9405	-0.2999	2.0102	5.0825	2.3439
2022	-4.7164	-5.3165	-2.9165	2.6868	4.3592	5.2577

Cuadro D.17: Residuos estandarizados para la cominidad la Rioja.

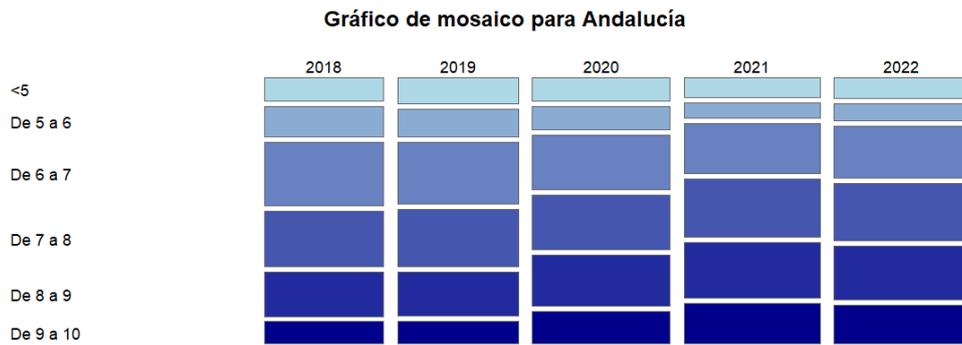


Figura D.1: Gráfico de mosaico para la comunidad de Andalucía

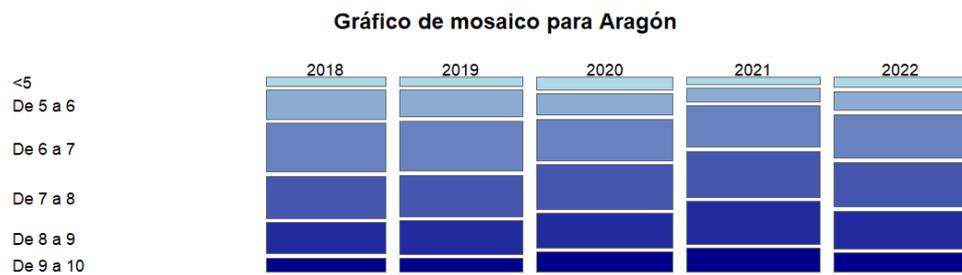


Figura D.2: Gráfico de mosaico para la comunidad de Aragón

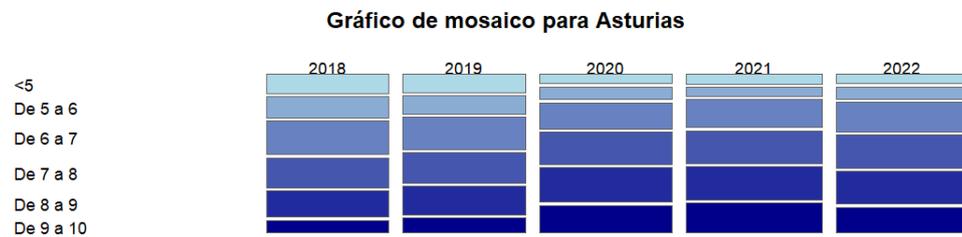


Figura D.3: Gráfico de mosaico para la comunidad de Asturias

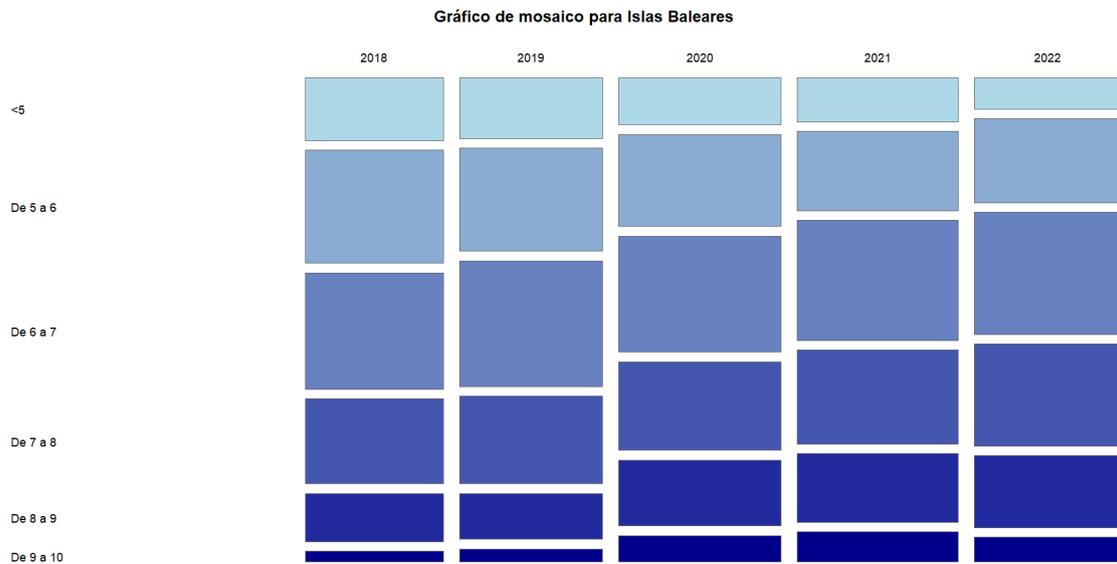


Figura D.4: Gráfico de mosaico para la comunidad de las Islas Baleares



Figura D.5: Gráfico de mosaico para la comunidad de las Islas Canarias

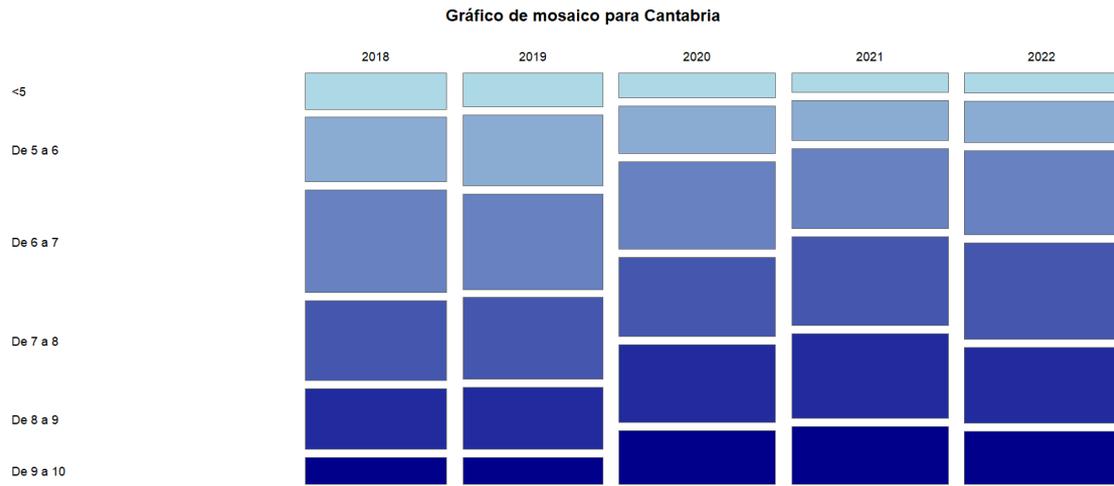


Figura D.6: Gráfico de mosaico para la comunidad de Cantabria



Figura D.7: Gráfico de mosaico para la comunidad de castilla y León

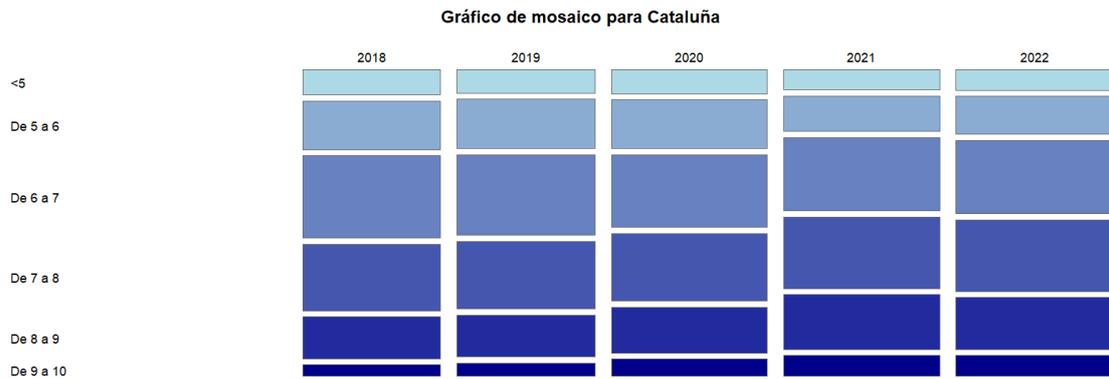


Figura D.8: Gráfico de mosaico para la comunidad de Cataluña

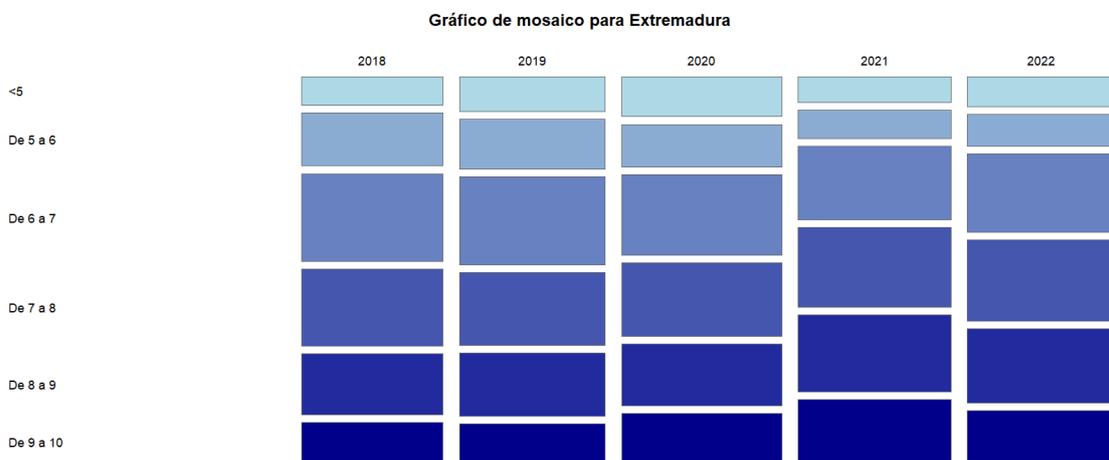


Figura D.9: Gráfico de mosaico para la comunidad de Extremadura

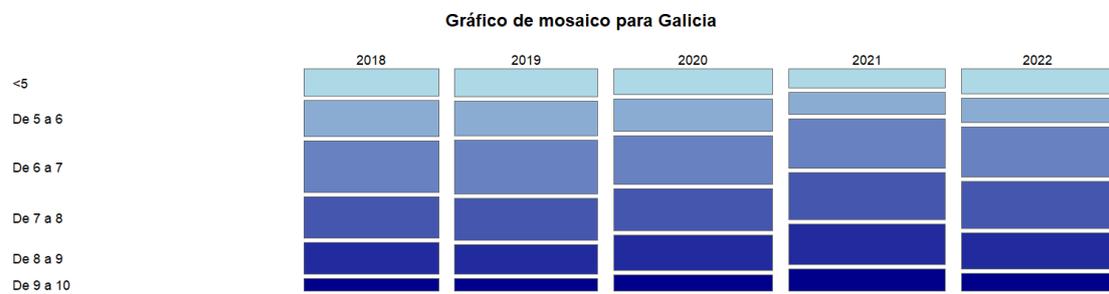


Figura D.10: Gráfico de mosaico para la comunidad de Galicia

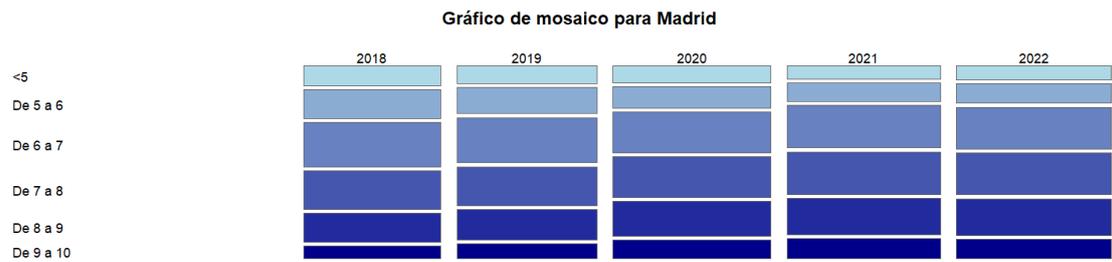


Figura D.11: Gráfico de mosaico para la comunidad de Madrid



Figura D.12: Gráfico de mosaico para la comunidad de Castilla la Mancha

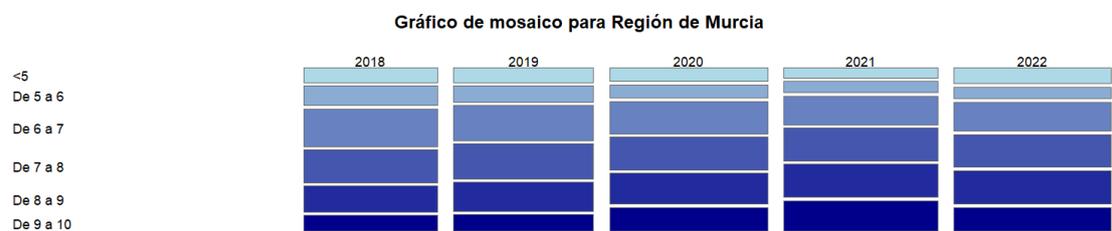


Figura D.13: Gráfico de mosaico para la comunidad de la Región de Murcia

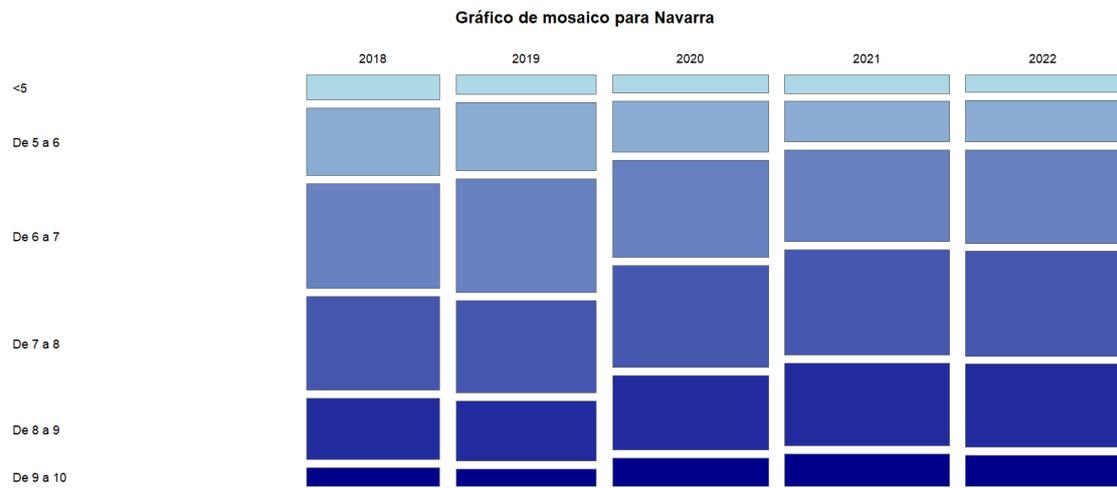


Figura D.14: Gráfico de mosaico para la comunidad de Navarra

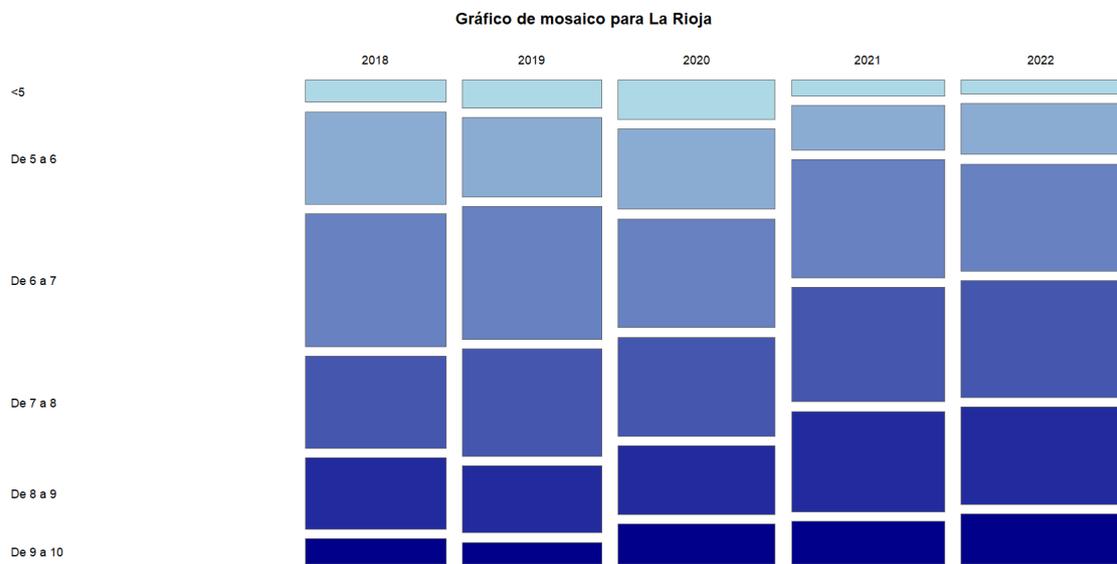


Figura D.15: Gráfico de mosaico para la comunidad de la Rioja

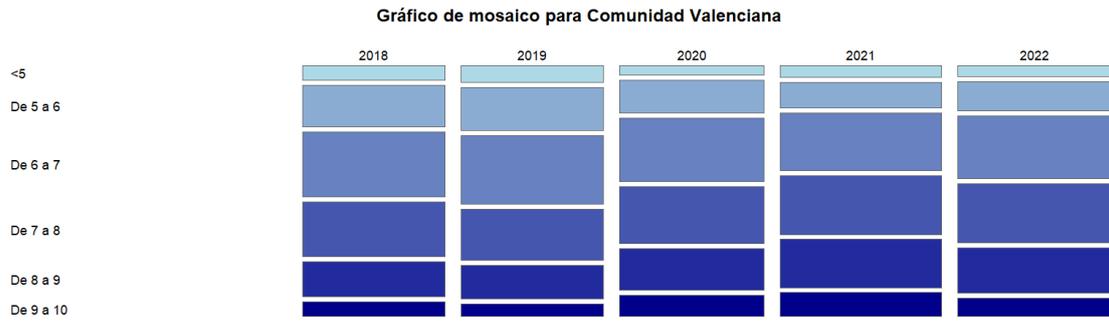


Figura D.16: Gráfico de mosaico para la comunidad Valenciana



Figura D.17: Gráfico de mosaico para la comunidad del País Vasco

Anexo E

Biplots análisis de correspondencias

A continuación se muestran los biplots para las distintas comunidades autónomas tras realizar el análisis de correspondencias.

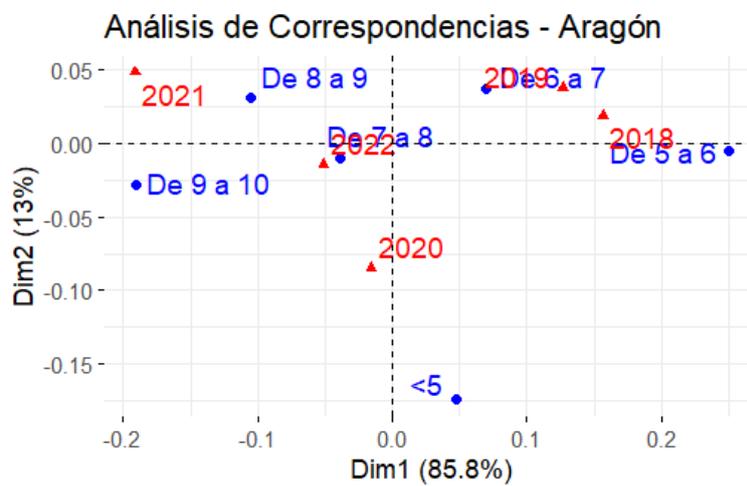


Figura E.1: Biplot para la comunidad de Aragón

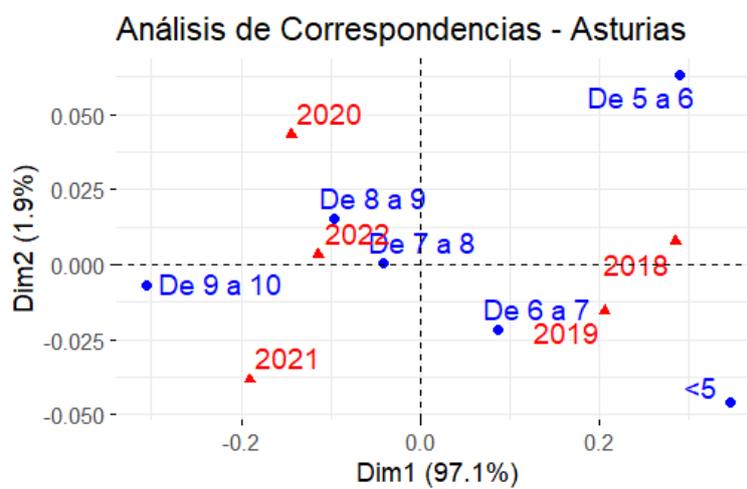


Figura E.2: Biplot para la comunidad de Asturias

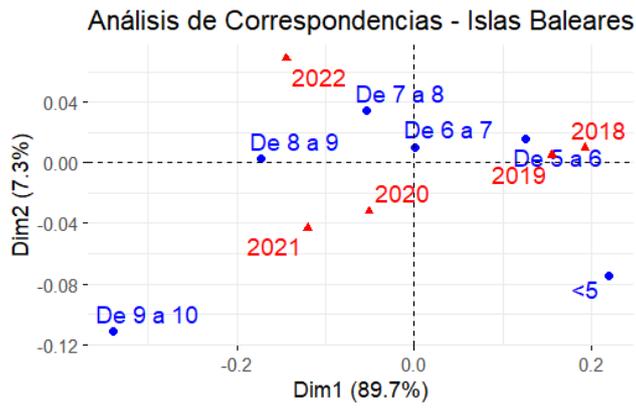


Figura E.3: Biplot para la comunidad de las Islas baleares

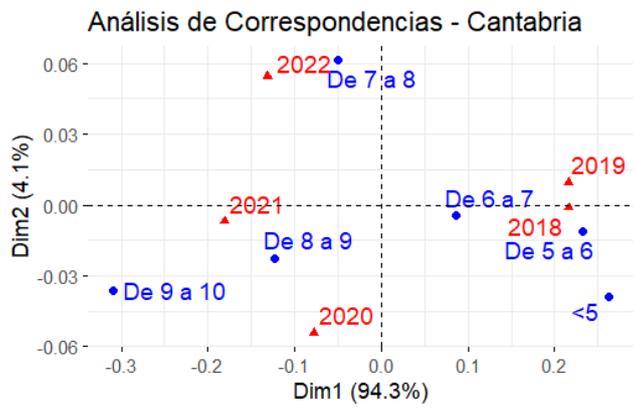


Figura E.4: Biplot para la comunidad de Cantabria

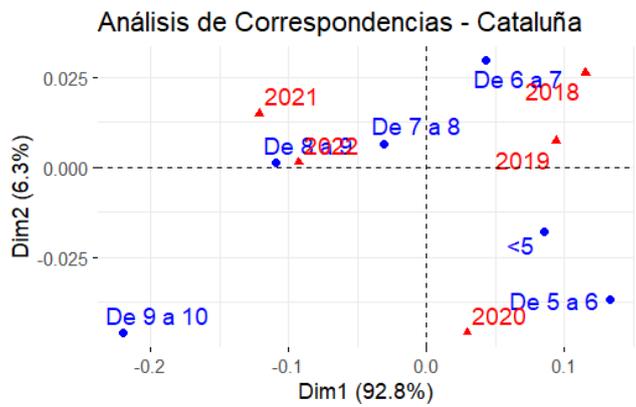


Figura E.5: Biplot para la comunidad de Cataluña

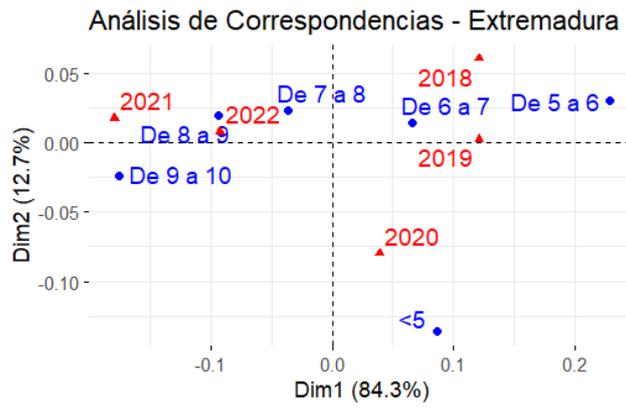


Figura E.6: Biplot para la comunidad de Extremadura

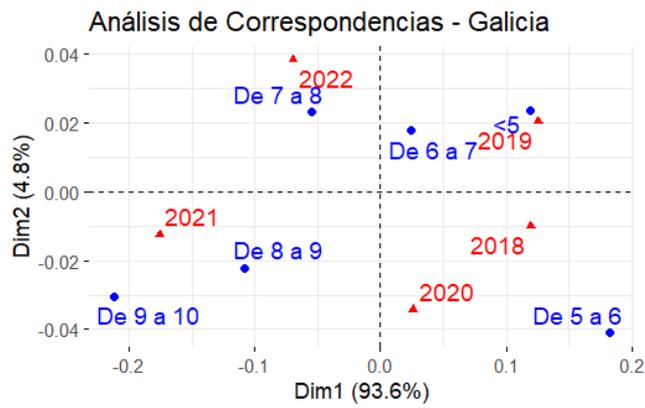


Figura E.7: Biplot para la comunidad de Galicia

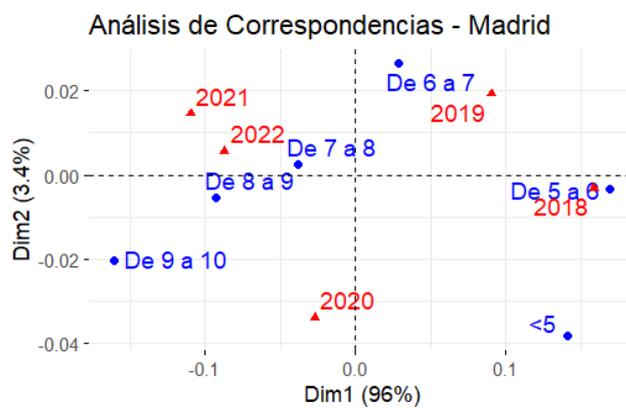


Figura E.8: Biplot para la comunidad de Madrid

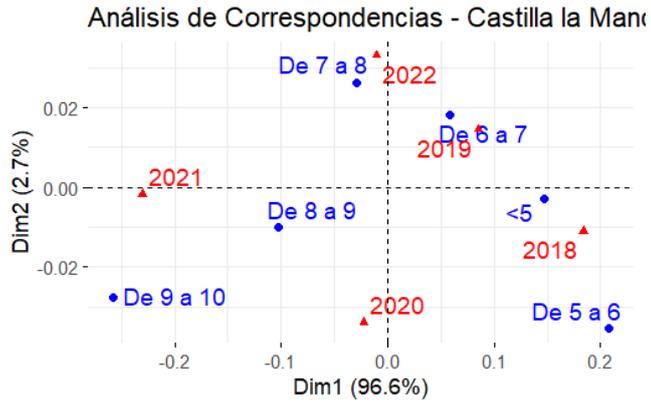


Figura E.9: Biplot para la comunidad de Castilla la Mancha

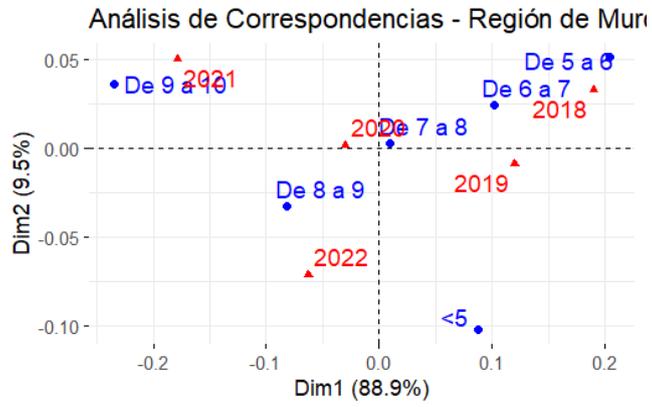


Figura E.10: Biplot para la comunidad de Murcia

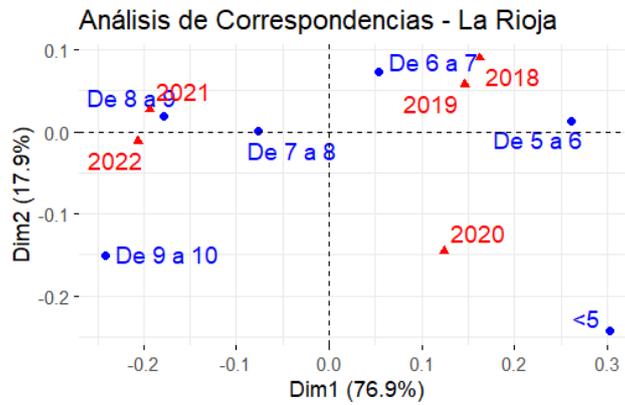


Figura E.11: Biplot para la comunidad de la Rioja

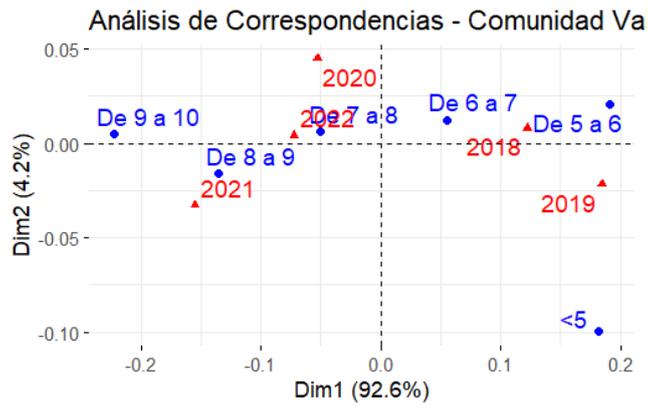


Figura E.12: Biplot para la comunidad de Valencia

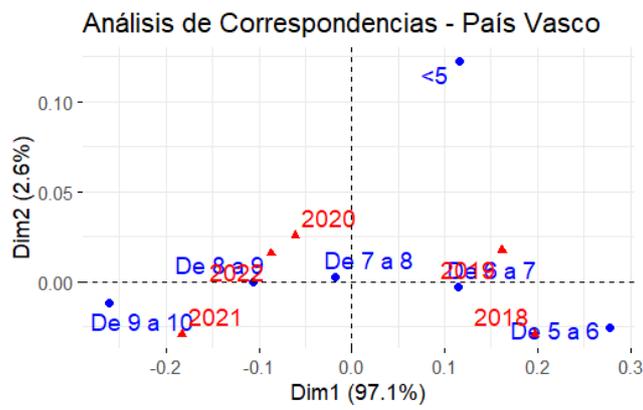


Figura E.13: Biplot para la comunidad del País Vasco

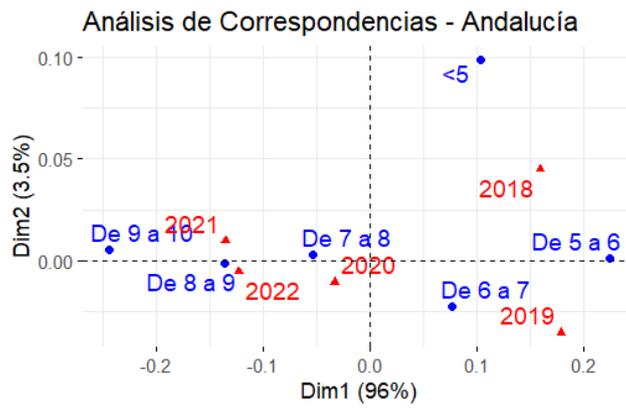


Figura E.14: Biplot para la comunidad de Navarra

Bibliografía

- [1] *¿Qué es la selectividad o EBAU? Así ha cambiado en la última década.* Onda Cero (Último acceso: 28 de mayo de 2025). 2024. URL: https://www.ondacero.es/noticias/sociedad/que-selectividad-llama-ebau-asi-cambiado-prueba-ultima-decada_20240603665d9c526a0d130001e2e085.html.
- [2] *Pruebas de Acceso Universidad de Valladolid.* Universidad de Valladolid (Último acceso: 28 de mayo de 2025). 2024. URL: <https://pruebasdeacceso.uva.es/1.ebau/>.
- [3] *Orden PJC/39/2024, de 24 de enero, por la que se determinan las características, el diseño y el contenido de la evaluación de Bachillerato para el acceso a la universidad, y las fechas máximas de realización y de resolución de los procedimientos de revisión de las calificaciones obtenidas, en el curso 2023-2024.* Boletín Oficial del Estado (Último acceso: 28 de mayo de 2025). 2024. URL: <https://www.boe.es/boe/dias/2024/01/26/pdfs/BOE-A-2024-1471.pdf>.
- [4] *Introducción al R. R Project, traducción al español* (Último acceso: 28 de mayo de 2025). 2004. URL: <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>.
- [5] *Manual de RStudio.* Fedepalma (Último acceso: 28 de mayo de 2025). 2023. URL: <https://repositorio.fedepalma.org/bitstream/handle/123456789/141281/Manual%20REstudio.pdf>.
- [6] *Manual RStudio - Universidad Distrital.* Universidad Distrital (Último acceso: 28 de mayo de 2025). 2023. URL: <https://ftecnologica.udistrital.edu.co/laboratorios/ciencias-basicas/sites/lab-ciencias-basicas/files/practica/2023-03/Manual%20Rstudio.pdf>.
- [7] *Creación de una aplicación Shiny.* The Programming Historian (Último acceso: 28 de mayo de 2025). 2023. URL: <https://programminghistorian.org/es/lecciones/creacion-de-aplicacion-shiny>.
- [8] *An Introduction to ggplot2.* ggplot2.tidyverse.org (Último acceso: 30 de mayo de 2025). 2023. URL: <https://ggplot2.tidyverse.org/articles/ggplot2.html>.
- [9] *Getting Started with Plotly in ggplot2.* Plotly (Último acceso: 30 de mayo de 2025). 2023. URL: <https://plotly.com/ggplot2/getting-started/>.
- [10] *Shiny: the way to create an interactive dashboard in R.* Medium (Último acceso: 28 de mayo de 2025). 2022. URL: <https://medium.com/@dblancbellido/shiny-the-way-to-create-an-interactive-dashboard-in-r-4cb7d5d17178>.
- [11] *Tutorial Shiny.* Universidad de Alicante (Último acceso: 28 de mayo de 2025). 2016. URL: <https://rua.ua.es/dspace/bitstream/10045/54325/1/shiny.pdf>.
- [12] *Shiny Basics - Posit.* Posit (Último acceso: 28 de mayo de 2025). 2024. URL: <https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/>.
- [13] *Desarrollo de aplicaciones con Shiny en R.* Google Books (Último acceso: 28 de mayo de 2025). 2021. URL: <https://books.google.es/books?id=CfErEAAAQBAJ>.
- [14] *Shiny Tutorial - STAT 545.* STAT 545 (Último acceso: 28 de mayo de 2025). 2024. URL: <https://stat545.com/shiny-tutorial.html>.
- [15] *Introducción a Shiny - Epidemiologist R Handbook.* Epi R Handbook (Último acceso: 28 de mayo de 2025). 2024. URL: https://epirhandbook.com/es/new_pages/shiny_basics.es.html.
- [16] *Shiny Cheat Sheet.* Posit (Último acceso: 30 de mayo de 2025). 2022. URL: <https://posit.co/wp-content/uploads/2022/10/shiny-1.pdf>.

- [17] *Clase 3 Shiny - Marcos Armas*. RPubs (Último acceso: 28 de mayo de 2025). 2023. URL: <https://rpubs.com/marcosarmas/shiny-clase-3>.
- [18] *Estructura de shinydashboard*. RStudio (Último acceso: 28 de mayo de 2025). 2024. URL: <https://rstudio.github.io/shinydashboard/structure.html>.
- [19] *shinydashboard*. RStudio (Último acceso: 28 de mayo de 2025). 2024. URL: <http://rstudio.github.io/shinydashboard/>.
- [20] *Chi-square test - IBM SPSS*. IBM SPSS (Último acceso: 28 de mayo de 2025). 2024. URL: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-chi-square-test>.
- [21] *Interpretación de chi-cuadrado - Minitab*. Minitab (Último acceso: 28 de mayo de 2025). 2024. URL: <https://support.minitab.com/es-mx/minitab/help-and-how-to/statistics/tables/how-to/chi-square-test-for-association/interpret-the-results/all-statistics/#standardized-residuals>.
- [22] *Chi-cuadrado - Notas metodológicas*. Scribd (Último acceso: 28 de mayo de 2025). 2018. URL: <https://www.scribd.com/document/388909289/Chi-cuadrado-Notas-metodologicas>.
- [23] *Correspondence Analysis - IBM SPSS*. IBM SPSS (Último acceso: 28 de mayo de 2025). 2024. URL: <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=categories-correspondence-analysis>.
- [24] *Análisis de correspondencias*. Fuenterebollo (Último acceso: 28 de mayo de 2025). 2020. URL: <https://www.fuenterebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>.
- [25] *Análisis de correspondencias*. CDR Book (Último acceso: 28 de mayo de 2025). 2023. URL: <https://cdr-book.github.io/correspondencias.html>.
- [26] *Análisis factorial de correspondencias*. Universidad de Sevilla (Último acceso: 28 de mayo de 2025). 2019. URL: <https://personal.us.es/analopez/afc.pdf>.
- [27] *Análisis de correspondencias en IBM SPSS Statistics 29*. IBM Documentation (Último acceso: 30 de mayo de 2025). 2023. URL: <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=categories-correspondence-analysis>.
- [28] *Multiple Correspondence Analysis - IBM SPSS*. IBM SPSS (Último acceso: 28 de mayo de 2025). 2024. URL: <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=categories-multiple-correspondence-analysis>.
- [29] *Análisis de correspondencias con R*. RPubs (Último acceso: 28 de mayo de 2025). 2017. URL: <https://rpubs.com/kfhidalgo/300863>.
- [30] *Capítulo 13: Análisis de correspondencias*. Fundación BBVA (Último acceso: 28 de mayo de 2025). 2017. URL: https://www.biophilia-fbbva.es/wp-content/uploads/2017/05/dat/greenacre_cap13.pdf.