



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

***WorldCupViz*: Aplicación shiny para el análisis y visualización
de datos de competiciones de fútbol**

Autor: Luis Hernández Díez
Tutora: María Teresa González Arteaga
Año: 2023/2024

Resumen

La estadística aplicada al deporte ha experimentado avances significativos que han transformado tanto la comprensión como la gestión de las actividades deportivas a lo largo de las últimas décadas. En el contexto contemporáneo, la aplicación de métodos estadísticos sofisticados ha permitido analizar de manera más profunda y precisa diversos aspectos del rendimiento atlético, la estrategia competitiva y la gestión de equipos.

En el presente documento se presenta un análisis detallado de datos de las recientes copas mundiales de fútbol masculina y femenina celebradas en Qatar y Australia/Nueva Zelanda en los años 2022 y 2023, respectivamente. Este análisis se lleva a cabo mediante una aplicación web desarrollada en shiny, denominada WorldCupViz. Se persigue el objetivo de proporcionar a los usuarios una exploración exhaustiva de ambas competiciones, focalizándose en aspectos tanto individuales como colectivos. También se incluye en este trabajo una clasificación de las selecciones nacionales femeninas según su estilo de juego mediante un análisis clúster.

WorldCupViz emplea visualizaciones interactivas y herramientas de análisis estadístico avanzado para examinar diversos aspectos del rendimiento en el fútbol durante estas competiciones. La aplicación permite a los investigadores, analistas y aficionados del deporte explorar datos complejos de los partidos disputados, tales como estadísticas de jugadores, estrategias de equipos, resultados de encuentros y tendencias emergentes.

Abstract

Statistics applied to sport has undergone significant advances that have transformed both the understanding and management of sporting activities over the last decades. In the contemporary context, the application of sophisticated statistical methods has allowed a deeper and more accurate analysis of various aspects of athletic performance, competitive strategy, and team management.

This document presents a detailed data analysis of the recent men's and women's World Cups held in Qatar and Australia/New Zealand in the years 2022 and 2023, respectively. This analysis is conducted using a web application developed in shiny, called WorldCupViz. The objective is to provide users with a comprehensive exploration of both competitions, focusing on both individual and collective aspects. Also included in this work is a classification of the women's national teams according to their style of play by means of a cluster analysis.

WorldCupViz employs interactive visualizations and advanced statistical analysis tools to examine various aspects of soccer performance during these competitions. The application allows researchers, analysts, and sports fans to explore complex match data such as player statistics, team strategies, match results and emerging trends.

Agradecimientos

Quiero expresar mi profundo agradecimiento a todos los profesores que han sido parte fundamental de mi formación durante estos últimos cuatro años. En particular, agradezco de manera especial a mi tutora, Teresa, no solo por su invaluable apoyo en la elaboración de este trabajo de fin de grado, sino también por permitirme explorar un tema que despierta un gran interés personal.

Mi más sincero agradecimiento va dirigido también a mis amigos, quienes han sido un apoyo incondicional en los momentos más difíciles y han sido fundamentales para llevar adelante este proyecto académico.

Agradecer también a mis compañeros de estudio por su constante ayuda y por convertir estos años en una etapa inolvidable, llena de experiencias compartidas y amistades que valoraré siempre.

Por último, quiero expresar mi profunda gratitud a mi familia por su constante esfuerzo y apoyo a lo largo de estos años. Gracias a ellos, he podido concentrarme plenamente en mi formación académica, sabiendo que siempre contaba con su respaldo y comprensión.

Índice

CAPÍTULO 1	1
INTRODUCCIÓN	1
1.1. OBJETIVOS.....	1
1.2. DISPOSICIÓN DE LA MEMORIA.....	2
CAPÍTULO 2	3
HISTORIA DE LOS MUNDIALES DE FÚTBOL	3
2.1. HISTORIA DEL FÚTBOL Y DE LAS COPAS MUNDIALES MASCULINAS	3
2.1.1. <i>Datos relevantes en la historia de los mundiales masculinos</i>	4
2.2. HISTORIA DEL FÚTBOL FEMENINO Y DE LAS COPAS MUNDIALES FEMENINAS	5
2.2.1 <i>Datos relevantes en la historia de los mundiales femeninos</i>	6
2.3. ESTRUCTURA DE LA COMPETICIÓN	7
CAPÍTULO 3	9
CONJUNTO DE DATOS Y EXPLORACIÓN	9
3.1. ¿QUÉ ES STATS BOMB?.....	9
3.2. FICHEROS DE DATOS.....	10
3.3. EXPLICACIÓN DEL CONJUNTO DE DATOS.....	10
3.4. EXPLORACIÓN DE LOS DATOS.....	13
CAPÍTULO 4	19
CLASIFICACIÓN DE LOS EQUIPOS DE LA COMPETICIÓN FEMENINA SEGÚN SU ESTILO DE JUEGO	19
4.1. CLUSTERING JERÁRQUICO: MARCO TEÓRICO	19
4.2. CLUSTERING JERÁRQUICO: PROCEDIMIENTO.....	20
4.3. APLICACIÓN DEL CLUSTERING JERÁRQUICO AL CONJUNTO DE DATOS DE LA COMPETICIÓN FEMENINA.....	20
CAPÍTULO 5	27
APLICACIÓN SHINY: WORLDCUPVIZ	27
5.1. SHINY Y PAQUETES UTILIZADOS	27
5.1.1 <i>Estructura básica de una aplicación construida con shiny</i>	28
5.1.2 <i>Funcionamiento de una aplicación shiny</i>	29
5.1.3 <i>Otros paquetes de R utilizados</i>	30
5.2. ESTRUCTURA DE LA APLICACIÓN	30
<i>Portada de la aplicación</i>	31
<i>Menú de la competición</i>	32
<i>Página Resultados globales</i>	33
<i>Página Análisis Ofensivo</i>	34
<i>Página Partidos</i>	36
<i>Página Análisis defensivo</i>	39
<i>Página Estadísticas individuales</i>	41
<i>Creación radar y ranking por jugador</i>	43
<i>Página Balance final</i>	45
<i>Página de Ayuda</i>	45
5.3. IMÁGENES Y MENCIONES EXTERNAS.....	46

CAPÍTULO 6	47
CONCLUSIONES	47
6.1. COMPLEJIDAD Y DIFICULTADES DEL TRABAJO	47
6.2. CONOCIMIENTOS APLICADOS	47
6.3. TRABAJO A FUTURO	48
BIBLIOGRAFÍA	49

Índice de figuras

FIGURA 1. TABLA DE RÉCORDS DE SELECCIONES EN LA HISTORIA DE LOS MUNDIALES MASCULINOS DE FÚTBOL.....	4
FIGURA 2. TABLA DE RÉCORDS INDIVIDUALES EN LA HISTORIA DE LOS MUNDIALES MASCULINOS DE FÚTBOL.....	4
FIGURA 3. TABLA DE RÉCORDS DE SELECCIONES EN LA HISTORIA DE LOS MUNDIALES FEMENINOS.	6
FIGURA 4. TABLA DE RÉCORDS INDIVIDUALES EN LA HISTORIA DE LOS MUNDIALES FEMENINOS	6
FIGURA 5. CÓDIGO DE IMPORTACIÓN DE LOS DATOS STATS BOMB.....	10
FIGURA 6. GRÁFICO DE BARRAS CON LAS DIFERENTES ACCIONES MÁS RELEVANTES DE LAS SELECCIONES MASCULINAS Y FEMENINAS	13
FIGURA 7. RANKING DE LOS MEJORES PASADORES DE LA COMPETICIÓN MASCULINA Y FEMENINA	14
FIGURA 8. RANKING DE LAS MEJORES SELECCIONES PASADORAS DE LA COMPETICIÓN MASCULINA Y FEMENINA	14
FIGURA 9. GRÁFICO DE CORRELACIONES ENTRE DIFERENTES VARIABLES SOBRE ACCIONES OFENSIVAS EN AMBAS COMPETICIONES.....	15
FIGURA 10. GRÁFICO DE SECTORES SOBRE ACCIONES DEFENSIVAS DE AMBAS COMPETICIONES ..	16
FIGURA 11. MATRIZ DE CORRELACIÓN DE VARIABLES DE ACCIONES DEFENSIVAS	17
FIGURA 13. GRÁFICO DENDOGRAMA	21
FIGURA 14. DENDOGRAMA CON EL NÚMERO FINAL DE CLÚSTERS ELEGIDOS	22
FIGURA 15. GRÁFICOS DE DISPERSIÓN ENTRE LAS VARIABLES QUE MUESTRAN LA CLASIFICACIÓN EN 3 GRUPOS.....	23
FIGURA 16. GRÁFICO DE 2 DIMENSIONES TRAS APLICAR ACP.....	23
FIGURA 17: TABLA DE VALORES PROPIOS ASOCIADOS A CADA COMPONENTE PRINCIPAL	24
FIGURA 18. DENDOGRAMA TRAS APLICAR ACP.....	25
FIGURA 19. GRÁFICO DE BARRAS AGRUPADOS POR CLÚSTERS.....	26
FIGURA 20. PORTADA DE LA APLICACIÓN WORLDCUPVIZ.....	31
FIGURA 21. MENÚ DE LA COMPETICIÓN	32
FIGURA 22. PAGINA RESULTADOS GLOBALES, PESTAÑA “CUADRO GENERAL”	33
FIGURA 23. PÁGINA RESULTADOS GLOBALES, PESTAÑA “FASES Y RESULTADOS”	34
FIGURA 24. PÁGINA ANÁLISIS OFENSIVO, PESTAÑA “GLOBAL”	34
FIGURA 25. PÁGINA ANÁLISIS OFENSIVO, PESTAÑA “DATOS OFENSIVOS POR SELECCIÓN”	35
FIGURA 26. PÁGINA ANÁLISIS OFENSIVO, PESTAÑA “MAPA OFENSIVO”	36
FIGURA 27. PÁGINA PARTIDOS, PESTAÑA “ESTADÍSTICAS GENERALES”	37
FIGURA 28. PÁGINA PARTIDOS, PESTAÑA “INFLUENCIA DE PASES”	37
FIGURA 29. PÁGINA PARTIDOS, PESTAÑA “MÉTRICAS XG+XGA”	38
FIGURA 30. PÁGINA PARTIDOS, PESTAÑA “MVP DEL PARTIDO”	38

FIGURA 31. PÁGINA ANÁLISIS DEFENSIVO, PESTAÑA “ACCIONES DEFENSIVAS”	39
FIGURA 32. PÁGINA ANÁLISIS DEFENSIVO, PESTAÑA “ESTADÍSTICAS DEFENSIVAS GLOBALES”	40
FIGURA 33. PÁGINA ANÁLISIS DEFENSIVO, PESTAÑA “FALTAS Y TARJETAS”	40
FIGURA 34. PÁGINA ESTADÍSTICAS INDIVIDUALES, PESTAÑA “LOCALIZACIÓN Y PASES”	41
FIGURA 35. PÁGINA ESTADÍSTICAS INDIVIDUALES, PESTAÑA “ACCIONES INDIVIDUALES”	42
FIGURA 36. PÁGINA ESTADÍSTICAS INDIVIDUALES, PESTAÑA “RADAR GLOBAL”	42
FIGURA 37. EJEMPLO DE RADAR DE JUGADOR Y DE PORTERO.....	44
FIGURA 38. PÁGINA BALANCE FINAL.....	45
FIGURA 39. PÁGINA AYUDA	46

Capítulo 1

Introducción

En la última década, el ámbito deportivo ha experimentado una transformación sin precedentes impulsada por los avances tecnológicos. En particular, la aplicación del *Big Data* ha revolucionado la manera en que los equipos y organizaciones deportivas comprenden y mejoran el rendimiento de los atletas. Este fenómeno no solo ha redefinido las estrategias de entrenamiento y tácticas dentro de disciplinas como el fútbol, sino que también ha abierto nuevas oportunidades para la innovación tecnológica en este campo.

Entre los avances destacados, se encuentra el uso extensivo de análisis de datos para evaluar el desempeño individual y colectivo de los atletas, identificar patrones de juego, predecir resultados y optimizar entrenamientos. La incorporación de tecnologías avanzadas, como sensores de movimiento, sistemas de seguimiento GPS y cámaras de alta velocidad, ha proporcionado datos detallados y en tiempo real que son cruciales para los análisis estadísticos en el deporte.

En este contexto, surge *WorldCupViz* como una aplicación diseñada para aprovechar las capacidades del Big Data en el análisis de los eventos deportivos más importantes, como la **Copa Mundial de la FIFA**. Desarrollada como parte de mi trabajo de fin de grado, esta herramienta no solo busca ofrecer percepciones profundas y visualizaciones claras sobre datos complejos, sino también proporcionar a entrenadores, analistas y aficionados una plataforma intuitiva para explorar y entender mejor el rendimiento de jugadores y selecciones.

Este trabajo se centra en exponer el proceso de desarrollo de la aplicación web *WorldCupViz*, desde su concepción inicial hasta su implementación práctica, destacando cómo la integración de tecnologías emergentes puede potenciar tanto el rendimiento deportivo como la experiencia del espectador.

<http://shiny1.eio.uva.es:3838/luis/App/WorldCupViz/>

1.1. Objetivos

El objetivo principal de este trabajo es el desarrollo de la aplicación, *WorldCupViz*, centrada en la visualización de datos de la copa mundial de fútbol masculina y femenina disputadas en Qatar y Australia/Nueva Zelanda en los años 2022 y 2023 respectivamente. Esta herramienta tiene como propósito ofrecer una representación clara y accesible de estadísticas y tendencias clave durante los torneos, facilitando así una comprensión profunda del desempeño de equipos y jugadores a lo largo de los eventos deportivos más importantes a nivel global. Para conseguir este objetivo se ha necesitado alcanzar los siguientes objetivos parciales:

- Comprensión del conjunto de datos para realizar un correcto tratamiento de estos.
- Selección de los diferentes paquetes a utilizar.
- Análisis exploratorio de los conjuntos de datos de las competiciones de fútbol tanto la masculina como la femenina.
- Creación de diferentes tablas y gráficos para una mejor comprensión.

- Realización de clustering jerárquico aplicado al conjunto de datos la copa mundial femenina para visualizar similitudes entre diferentes selecciones.
- Creación y diseño de una interfaz de fácil manejo e intuitiva para que todo tipo de usuarios puedan hacer uso de ella.

1.2. Disposición de la memoria

La memoria está estructurada de la siguiente manera:

- **Capítulo 1: Introducción:** Breve explicación de lo que trata este trabajo, objetivos y disposición de la presente memoria.
- **Capítulo 2: Historia de los mundiales de fútbol:** Se describe la historia de ambas competiciones de fútbol tanto la masculina como la femenina. En este capítulo también se detalla información relevante de cada una de ellas.
- **Capítulo 3: Conjunto de datos y exploración.** En este capítulo se realiza la explicación de los conjuntos de datos analizados, origen, variables utilizadas y pequeña exploración de los conjuntos de datos.
- **Capítulo 4: Clasificación de los equipos de la competición femenina según su estilo de juego:** Se realiza un análisis clúster jerárquico sobre el conjunto de datos de la competición femenina y observar similitudes entre selecciones.
- **Capítulo 5: WorldCupViz Aplicación shiny:** Capítulo en el que se detalla la estructura de la aplicación, su funcionamiento y contenido. En este punto también se comenta los paquetes utilizados y demás información utilizada.
- **Capítulo 6: Conclusiones:** Para finalizar esta memoria se exponen las diferentes dificultades encontradas para realizar este trabajo y posibles implementaciones que pueden realizarse en el futuro.

Capítulo 2

Historia de los mundiales de fútbol

El fútbol es un deporte ampliamente conocido, pero ¿cuántos realmente conocen la historia detallada de los mundiales de fútbol, tanto masculinos como femeninos?

Actualmente, esta competición, conocida como la **Copa Mundial de la FIFA**, se celebra cada cuatro años en distintos países alrededor del mundo. La primera edición masculina tuvo lugar en 1930 en Uruguay, mientras que la primera edición femenina se celebró en Italia en 1970.

Para participar en las fases finales de cada copa mundial, los equipos deben clasificarse mediante diversos métodos. Las formas de clasificación y el número de participantes varían entre las secciones masculina y femenina, reflejando las diferencias y evoluciones históricas de ambas competiciones.

2.1. Historia del fútbol y de las copas mundiales masculinas

El fútbol es el deporte más importante y practicado del mundo, aunque sus orígenes son disputados entre culturas como la china, griega y maya. En la Inglaterra medieval se jugaban juegos similares, pero fue en el siglo XIX cuando se formalizó. En 1848, se estableció el código de Cambridge, y en 1863 en Londres, se oficializó el fútbol moderno. [1]

El fútbol ha evolucionado, con cambios en las reglas y la velocidad del juego. La introducción de tarjetas en 1970 y la regla del fuera de juego en 1952 han sido claves para mejorar la justicia y el equilibrio en el juego. La implementación del video arbitraje (VAR) ha revolucionado la toma de decisiones durante los partidos, mejorando la precisión arbitral.

El crecimiento del fútbol ha sido impresionante, con 270 millones de participantes diarios y más de 4.000 millones de aficionados. La FIFA, fundada en 1904, y la International Football Association Board en 1886, han sido cruciales en su expansión.

La primera copa mundial masculina en 1930 en Uruguay no tuvo proceso de clasificación y fue ganada por el país anfitrión. Tras la Segunda Guerra Mundial, las ediciones de 1942 y 1946 no se disputaron, retomándose en 1950 con el famoso “*Maracanazo*”. La primera copa mundial televisada fue en Suiza en 1954. En 1966 se introdujeron controles de dopaje.

La final del Mundial de 1970 en México, donde Brasil liderado por Pelé fue campeón, es considerada por los expertos de fútbol una de las mejores copas mundiales. En 1974, se incorporó una segunda fase de grupos, eliminada en 1982. La edición de 1994 en EE. UU. introdujo varias novedades, como otorgar 3 puntos por victoria y la inclusión de nombres en las camisetas.

En 2010, España ganó su primer Mundial en Sudáfrica. En 2018, en Rusia, se usó el VAR por primera vez en un Mundial y en la última edición de 2022 Argentina ganó su 3ª copa mundial.

2.1.1. Datos relevantes en la historia de los mundiales masculinos

A lo largo de sus 22 ediciones, la *Copa Mundial de la FIFA* ha sido escenario de la consecución de récords excepcionales por parte de diversas selecciones nacionales, muchos de los cuales resultan extremadamente difíciles de superar. Entre los más destacados se incluyen: (Véase la Figura 1) [2]

Selección con más títulos	Brasil; 5 títulos
Selección con más finales jugadas	Alemania; 8 finales
Selección con más partidos jugados	Brasil; 114 partidos
Selección con más victorias	Brasil; 75 victorias
Selección con más victorias en un Mundial	Brasil; 7 victorias (2002)
Selección con más goles	Brasil; 237 goles a favor
Selección con más goles en un Mundial	Hungría; 27 goles (1954)
Selección con más Mundiales disputados	Brasil, 22 mundiales

Figura 1. Tabla de récords de selecciones en la historia de los mundiales masculinos de fútbol

Se han mencionado algunas de las estadísticas más destacadas en relación con las selecciones nacionales. A continuación, se presentan datos significativos sobre el desempeño individual de diversos jugadores a lo largo de la historia de las copas del mundo. (Véase la Figura 2)

Jugador con más títulos	Pelé (Brasil); 3.
Jugador con más partidos jugados	Lionel Messi (Argentina); 26
Jugador con más goles convertidos	Miroslav Klose (Alemania); 16
Jugador con más goles convertidos en un partido	Oleg Salenko (Rusia); 5 vs. Camerún en 1994
Jugador con más goles convertidos en un mismo Mundial	Just Fontaine (Francia); 13 (1958)
Jugador con más asistencias	Pelé (Brasil); 10
Jugador con más victorias en los Mundiales	Miroslav Klose (Alemania); 17
Entrenador con más partidos dirigidos	Helmut Schön (Alemania); 25

Figura 2. Tabla de récords individuales en la historia de los mundiales masculinos de fútbol

2.2. Historia del fútbol femenino y de las copas Mundiales femeninas

A lo largo de los años, el fútbol femenino ha ganado cada vez más fuerza, atrayendo a un número creciente de mujeres que practican este deporte tanto de manera *amateur* como profesional. Sin embargo, la historia de las copas del mundo organizadas por la FIFA en esta disciplina tiene un recorrido relativamente corto, a pesar de la existencia de otros torneos de carácter mundial en épocas anteriores. [3]

El origen del fútbol femenino se remonta a finales del siglo XIX, cuando en Inglaterra se fundó el primer equipo femenino de la historia, el "*British Ladies Football Club*", creado por una activista de los derechos de la mujer. Un hito significativo ocurrió en 1920, cuando se estableció un récord de asistencia a un partido entre clubes de fútbol femenino en Inglaterra. No obstante, en 1921, la Federación Inglesa, seguida por otras federaciones, prohibió el fútbol femenino, argumentando que su práctica "no era adecuada para mujeres" y temiendo que afectara la asistencia a los partidos masculinos. Esta prohibición se mantuvo durante 50 años.

En 1970, en Turín, Italia, se fundó la Federación Internacional de Fútbol Femenino, que promovió el fútbol femenino mediante la organización de un Mundial no oficial, no reconocido por la FIFA. En este torneo participaron siete selecciones en donde la selección danesa se alzó con la copa.

En España, la selección femenina de fútbol fue reconocida en 1983, aunque las jugadoras no fueron identificadas como profesionales. Años después, en 1991, se celebró en China el primer mundial de fútbol femenino organizado por la FIFA, en el que Estados Unidos se llevó la copa. A partir de entonces, esta competición se ha celebrado cada cuatro años en diferentes sedes, al igual que el torneo masculino.

En la era moderna, tras la oficialización de la copa del mundo Femenina por parte de la FIFA, Estados Unidos apostó decididamente por el fútbol femenino, brindando a las jugadoras mayores facilidades para profesionalizar su deporte. En 1999, se disputó uno de los mundiales más recordados en este país, con una gran asistencia a los estadios y un récord de 90,000 espectadores en la final, en la que nuevamente se coronó la selección anfitriona.

Progresivamente, instituciones deportivas como la *UEFA* empezaron a oficializar competiciones femeninas, como la "*Champions*" en 2001. En España, clubes nacionales como el FC Barcelona y el Atlético de Madrid comenzaron a invertir en el fútbol femenino, convirtiéndose en algunos de los clubes más importantes en esta disciplina.

Si bien el año 2010 fue crucial para el fútbol masculino español, con la consecución de la copa del mundo, el año 2023 ha sido igual de trascendental para el fútbol femenino. En este año, la selección española femenina ganó su primera copa mundial de fútbol, consolidando a España como uno de los países más influyentes en el fútbol femenino a nivel mundial.

2.2.1 Datos relevantes en la historia de los mundiales femeninos

Tras su oficialización en 1991, se han disputado un total de 9 ediciones. Datos que destacar en cuanto a nivel de selecciones son los siguientes, en donde Estados Unidos acapara gran parte de estos récords: [4] (Véase la Figura 3)

Selección con más títulos	Estados Unidos; 4 títulos
Selección con más finales jugadas	Estados Unidos; 5 finales
Selección con menos goles encajados en un Mundial	Alemania; 0 goles (2007)
Mundial con más goles anotados	2023 Australia/Nueva Zelanda
Selección con más goles encajados	Nigeria; 65 goles encajados
Selección con más goles	Estados Unidos; 142 goles a favor
Selección con más goles en un Mundial	Estados Unidos; 26 goles (2019)
Selección con más goles en un partido	Estados Unidos 13 Goles; (2019)

Figura 3. Tabla de récords de selecciones en la historia de los mundiales femeninos

En cuanto a récords individuales existen datos sorprendentes que diferentes jugadoras han logrado y han conseguido aportar grandes actuaciones a su selección: (Véase la Figura 4)

Jugadora con más partidos jugados	Kristine Lilly (Estados Unidos); 30 partidos
Jugadora con más goles convertidos	Marta Viera Silva (Brasil); 17 goles
Jugadora con más mundiales disputados	Formiga (Brasil); 7 mundiales
Jugadora más joven en debutar	Casey Phair; 16 años y 26 días (Corea del Sur)
Jugadora en ser expulsada más temprano	Alicia Ferguson; 2 minutos (Estados Unidos)
Jugadora más veterana en disputar un partido	Christie Rampone (Estados Unidos); 40 años y 11 días
Entrenadora que más veces ha ganado	Jill Ellis (Estados Unidos); 2

Figura 4. Tabla de récords individuales en la historia de los mundiales femeninos

2.3. Estructura de la competición

En la actualidad, la **Copa Mundial de la FIFA** está compuesta de dos etapas: una fase clasificatoria y una fase final, compartiendo este formato tanto en la competición masculina como en la femenina, y llevándose a cabo cada cuatro años. Para garantizar la legitimidad de la competición y de todos sus partidos, es fundamental que tanto la organización del evento como los encuentros cumplan con las normas y regulaciones establecidas por la **FIFA** y el “**International Football Association Board**” (**IFAB**), organismos responsables de las reglas del juego.

Estructura de la fase clasificatoria en el formato masculino

La FIFA determina la cantidad de plazas de participación asignadas a cada confederación para la copa mundial, con algunas plazas otorgadas automáticamente y otras disputadas en playoffs. El número de equipos participantes ha variado a lo largo del tiempo. Por ejemplo, en 1934 hubo 32 equipos compitiendo por 16 lugares, mientras que en 2022, 211 equipos compitieron por 31 plazas. Para 2026, el torneo se ampliará a 48 equipos, siendo este el primer incremento desde 1998, cuando el número de selecciones aumentó de 24 a 32.

Dependiendo de la zona geográfica, se asigna un número específico de plazas para clasificar a las rondas finales. A continuación, se presentan los cupos correspondientes para el próximo Mundial de Fútbol Masculino, que se celebrará en 2026 en Canadá, Estados Unidos y México.

- **AFC**, *Confederación de Fútbol de Asia*: 8 plazas directas y 1 plaza para la repesca
- **CAF**, *Confederación Africana de Fútbol*: 9 plazas y 1 plaza para la repesca
- **CONCACAF**, *Confederación de Fútbol Asociación de Norte y el Caribe*: 6 plazas (3 de ellas asignadas automáticamente a los anfitriones) y 2 plazas para la repesca
- **CONMEBOL**, *Confederación Sudamericana de Fútbol*: 6 plazas y 1 plaza para la repesca
- **OFC**, *Confederación de Fútbol de Oceanía*: 1 plazas y 1 plaza para la repesca
- **UEFA**, *Unión de Federaciones Europeas de Fútbol*: 16 plazas

Las selecciones que no hayan conseguido clasificar directamente a la fase final y tengan que jugar repesca se verán envueltas en un pequeño torneo de 6 selecciones en el que solamente 2 de ellas conseguirán una plaza para la fase final a disputar en 2026.

Estructura de la fase clasificatoria en el formato femenino

Al igual que en el formato masculino, antes de la fase final de la copa mundial, se organiza una fase clasificatoria en cada continente, una práctica que se ha mantenido desde la primera edición en 1991.

Cada confederación es responsable de organizar su propio proceso de selección de equipos para la copa mundial, siguiendo el número de cupos asignados por la FIFA. Algunos de estos cupos se otorgan automáticamente, mientras que otros se disputan en partidos de repesca entre equipos de diferentes confederaciones. Estos encuentros son cruciales para determinar qué equipos avanzan a la fase final del torneo.

En la pasada edición de 2023, la FIFA designó un número de plazas para cada Confederación:

- **AFC**, *Confederación de Fútbol de Asia*: 6 plazas directas y 2 plazas para la repesca
- **CAF**, *Confederación Africana de Fútbol*: 4 plazas directas y 2 plazas para la repesca
- **CONCACAF**, *Confederación de Fútbol Asociación de Norte y el Caribe*: 4 plazas directas y 2 plazas para la repesca
- **CONMEBOL**, *Confederación Sudamericana de Fútbol*: 3 plazas directas y 2 plazas para la repesca
- **OFC**, *Confederación de Fútbol de Oceanía*: 1 plaza directa y 1 plaza para la repesca
- **UEFA**, *Unión de Federaciones Europeas de Fútbol*: 11 plazas directas y 1 plaza para la repesca

De todos los países que jugaron la repesca, solamente 3 fueron capaces de lograr llegar a la fase final del mundial disputado en Australia/Nueva Zelanda en el año 2023.

Estructura de la Fase Final de la Copa Mundial de la FIFA

La fase final de la copa mundial tanto para el formato masculino como femenino, organizada en la sede elegida por la FIFA, se divide en dos rondas:

Fase de Grupos

Las selecciones clasificadas se distribuyen en grupos de cuatro equipos cada uno, mediante un sorteo en el que cada grupo cuenta con un cabeza de serie designado por la FIFA. En esta fase, no puede haber más de dos selecciones de cada confederación en un mismo grupo, asegurando enfrentamientos entre equipos de diferentes continentes.

Cada selección juega un total de tres partidos. Los dos primeros clasificados de cada grupo avanzan a la fase de eliminación directa. Sin embargo, en el próximo mundial de 2026, los ocho mejores terceros de los doce grupos también avanzarán a la siguiente ronda.

Fase Eliminatoria

En esta fase, solo la mitad de las selecciones avanzan. Se enfrentan el primer clasificado contra el segundo clasificado del grupo correspondiente y viceversa, en rondas que incluyen los octavos de final, cuartos de final, semifinales y la final.

A diferencia de la fase de grupos, donde un empate en el tiempo reglamentario otorga un punto a cada equipo, en la fase eliminatoria, en caso de empate tras el tiempo reglamentario, se jugará una prórroga de 15 minutos por cada lado. Si el empate persiste, el partido se decidirá mediante una tanda de penaltis para determinar el ganador.

Capítulo 3

Conjunto de datos y exploración

En este capítulo se presenta una exposición detallada de los conjuntos de datos utilizados en el presente trabajo. En primer lugar, se introduce la fuente de procedencia de los datos, *StatsBomb*, una reconocida plataforma de análisis y recopilación de datos deportivos. Más tarde se proporciona una descripción exhaustiva del contenido del conjunto de datos, incluyendo las variables que lo componen, lo que permite una comprensión integral de la información disponible para el análisis.

3.1. ¿Qué es StatsBomb?

StatsBomb es una empresa de análisis de datos deportivos, especializada en la recopilación y el análisis avanzado de estadísticas de fútbol. Fundada en 2017, StatsBomb proporciona datos detallados y análisis para clubes de fútbol, analistas, entrenadores, y otros profesionales de la industria del deporte. [5]

Características Principales de StatsBomb

1. **Datos Detallados:** StatsBomb recolecta una gran cantidad de datos en cada partido, incluyendo pases, tiros, regates, intercepciones, y más. Estos datos son mucho más detallados que los que suelen recopilar otras empresas del sector.
2. **Modelos Avanzados:** Utiliza modelos estadísticos avanzados y técnicas de aprendizaje automático para proporcionar insights profundos sobre el rendimiento de jugadores y equipos.
3. **Análisis Táctico:** Ofrece análisis táctico avanzado, ayudando a los equipos a comprender mejor su rendimiento y el de sus oponentes.
4. **Visualizaciones:** Proporciona herramientas de visualización de datos que facilitan la interpretación de los datos complejos, haciendo más accesible la información tanto para analistas profesionales como para aficionados.
5. **Consultoría:** Además de los datos y análisis, StatsBomb también ofrece servicios de consultoría para ayudar a los equipos a mejorar su rendimiento basado en la interpretación de los datos.

StatsBomb ha logrado una reputación significativa en el mundo del fútbol gracias a la profundidad y precisión de sus datos, y es utilizada por numerosos equipos de primer nivel alrededor del mundo para ganar una ventaja competitiva basada en análisis estadísticos avanzados. Cada cierto tiempo, StatsBomb ofrece algunos datos gratis sobre fútbol y ha desarrollado el paquete de R *StatsBombR* para facilitar el uso de estos.

3.2. Ficheros de datos

A continuación, se describe el proceso de recopilación de datos de la copa mundial masculina celebrada en 2022 en Qatar y la copa mundial femenina celebrada en 2023 en Australia/Nueva Zelanda.

Inicialmente, para trabajar con estos datos en R, es esencial tener instalados los siguientes paquetes:

- *tidyverse*: Conjunto de paquetes para manipular los datos de una forma más sencilla. [12]
- *devtools*: Permite descargar paquetes desde Github. [13]
- *ggplot2*: Uno de los paquetes más relevantes en cuanto a la visualización de datos. [14]
- *StatsBombR*: Paquete para analizar los datos de StatsBomb. [15]

Proceso de extracción de los datos

Al ejecutar el siguiente código, (véase Figura 5), se carga en nuestro entorno de R el conjunto de datos correspondiente a la competición masculina.

```

Competicion <- FreeCompetitions() %>%
  filter(competition_id==43 & season_id==106)

Partidos <- FreeMatches(Competicion)

Mundial_2022_360 <- free_allevents(PartidosDF = Partidos, Parallel = T)

Mundial_2022_360 = allclean(StatsBombData)

```

Figura 5. Código de importación de los datos StatsBomb

De forma similar se cargan los datos de la competición femenina modificando la variable *competition_id==72* y *season_id==107*.

3.3. Explicación del conjunto de datos

En este punto, se explica el contenido del conjunto de datos y se identifican las variables más relevantes que se utilizan en la creación de la aplicación.

Ambos conjuntos de datos contienen 187 variables de diversos tipos, no todas serán consideradas relevantes para el análisis que se realiza en este TFG. A continuación, se enumeran y se dividen las variables en tres categorías distintas: variables de información general, variables relativas a acciones ofensivas y variables relativas a acciones defensivas.

Variables de información general

En esta categoría se colocan las variables que aportan información más general como por ejemplo minuto del partido, acción realizada, equipos disputando el partido, localización del jugador en el campo, nombre del jugador...:

- ***period***: Variable de tipo categórica que indica la parte en la que se está realizando la acción. 1 indica la primera parte, 2 la segunda parte, 3 la primera parte de la prórroga, 4 la segunda parte de la prórroga y 5 la tanda de penaltis.
- ***minute***: Variable de tipo numérica que indica el minuto de cada acción.
- ***possession***: Variable categórica que indica el equipo que tiene la posesión en la acción.
- ***type.name***: Variable cualitativa que indica el nombre de cada acción. Algunos de los tipos más relevantes son: Pase, tiro, presión...
- ***team.name***: Variable cualitativa que indica el equipo que mantiene la posesión.
- ***player.name***: Variable cualitativa que indica el nombre del jugador que realiza la acción.
- ***position.name***: Variable cualitativa que indica la posición del jugador dentro del campo.
- ***play_pattern.name***: Variable cualitativa que indica el nombre de donde procede la acción.
- ***OpposingTeam***: Variable que indica el equipo contrario
- ***location.x***: Coordenada x del campo donde se está llevando a cabo la acción.
- ***location.y***: Coordenada y del campo donde se está llevando a cabo la acción.
- ***match_id***: Identificador único que posee cada partido
- ***player.id***: Identificador único que posee cada jugador

Variables relativas a acciones ofensivas

En esta categoría de variables se incluyen aquellas que proporcionan información sobre aspectos ofensivos del juego, y que serán objeto de análisis posteriormente:

- ***pass.length* , *pass.angle* , *pass.end_location***: Variables que indican la dirección del pase y su duración. Estas variables numéricas se utilizan más adelante en la creación de diferentes mapas.
- ***pass.shot_assist* , *pass.goal_assist***: Variable de tipo Verdadero/Falso que nos indican si el pase previo al tiro ha finalizado en tiro y si ha terminado en gol.
- ***pass.type.name***: Variable categórica que indica el tipo de pase. Algún ejemplo es córner, recuperación, falta...
- ***pass.outcome.name***: Variable categórica que indica como ha sido el pase.
- ***dribble.outcome.name***: Variable categórica que indica si el regate realizado se ha completado correctamente o no.

- ***shot.statsbomb_xg***: Variable numérica que comprende de 0 a 1 que indica el valor esperado de gol por cada tiro realizado.
- ***shot.end_location***: Variable que indica las coordenadas x e y para señalar dónde finaliza cada disparo realizado.
- ***shot.type.name***: Variable categórica que indica el resultado del tiro.
- ***shot.outcome.name***: Variable categórica que indica como ha finalizado el tiro.
- ***shot.key_pass_id***: Identificador único para cada pase previo a un tiro
- ***pass.goal_assist***: Variable de tipo Verdadero/Falso que indica si el último pase antes del tiro ha terminado en gol o no.

Variables relativas a acciones defensivas

En esta categoría de variables se encuentran aquellas que ofrecen detalles sobre los aspectos defensivos del juego y que serán examinadas más adelante:

- ***duel.outcome.name***: Variable categórica que indica el resultado del duelo entre dos jugadores.
- ***foul_committed.card.name***: Variable categórica que indica el tipo de tarjeta señalada a un jugador.
- ***bad_behaviour.card.name***: Variable categórica que indica el tipo de tarjeta señalada por un mal comportamiento dentro del campo
- ***foul_committed.advantage***: Variable de tipo Verdadero/Falso que indica si se ha obtenido ventaja de la falta cometida.
- ***goalkeeper.type.name***: Variable categórica que indica la acción realizada por el portero
- ***ball_recovery.recovery_failure***: Variable de tipo Verdadero/Falso que indica si la recuperación por parte del jugador es efectiva o no.
- ***interception.outcome.name***: Variable categórica que indica el tipo de intercepción realizada por el jugador

Todas estas variables del conjunto de datos nombradas anteriormente se usan en el desarrollo de la aplicación web WorldCupViz.

3.4. Exploración de los datos

Una vez presentadas todas las variables con las que se trabajan, se proporciona información adicional sobre algunas de ellas y realizar un análisis exploratorio. Este proceso nos permite identificar posibles errores, descubrir patrones en las variables y establecer relaciones entre ellas.

Al extraer los conjuntos de datos tanto de la competición masculina como la femenina de la misma fuente, es posible realizar un análisis sobre cada uno de ellos y ver similitudes y diferencias entre ellos.

Análisis exploratorio de ambos conjuntos de datos

Existen diversos tipos de acciones dentro de un partido de fútbol, tales como tiros, pases, recuperaciones, y paradas. En el apartado anterior, hemos clasificado todas estas variables en tres grupos distintos. A continuación, en la Figura 6, se muestran algunas de las variables más relevantes del mundial masculino y femenino fútbol .

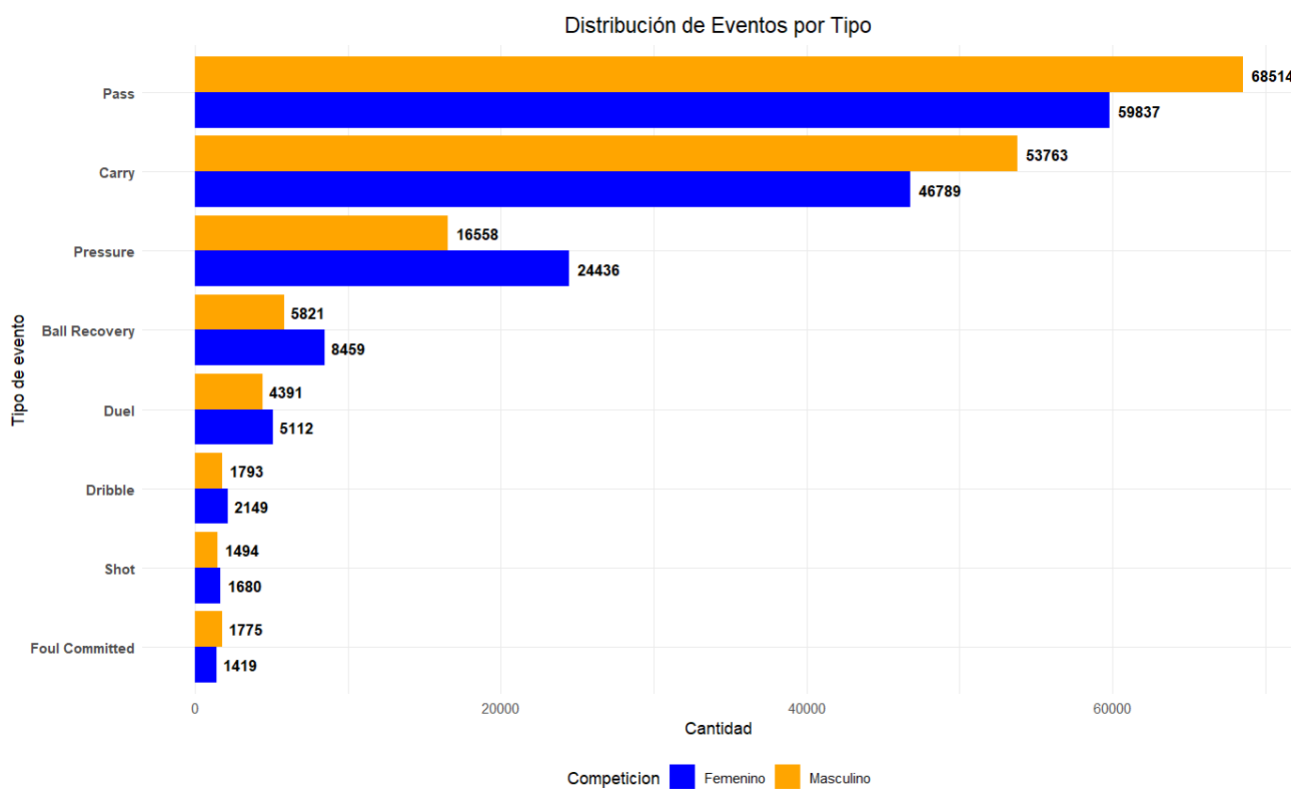


Figura 6. Gráfico de barras con las diferentes acciones más relevantes de las selecciones masculinas y femeninas

En la Figura 6, se visualizan las acciones más relevantes de todas las selecciones en la competición. Analizando la diferencia entre ambas competiciones, las cuales tienen el mismo número de partidos, se puede sacar en claro que en la competición masculina se han realizado un mayor número de pases, conducciones con el balón y faltas realizadas que la competición femenina. En cambio las mujeres destacan por realizar mayores acciones de presión, recuperaciones de balón, duelos, regates y tiros.

A continuación, en la Figura 7, se analiza la relación de algunas de estas acciones de los equipos y jugadores de la competición.

Ranking masculino mejores pasadores			Ranking femenino mejores pasadoras				
Jugador	% Pases acertados	% Pases fallados	Jugador	% Pases acertados	% Pases fallados		
1	António João Pereira	98.30	1.70	1	Rocío Gálvez Luna	94.40	5.60
2	Leander Dendoncker	98.30	1.70	2	Luana Bühler	94.00	6.00
3	Marcos Aoás Corrêa	96.10	3.90	3	María Isabel Rodríguez Rivero	92.90	7.10
4	Axel Witsel	96.00	4.00	4	Elena Linari	91.80	8.20
5	John Stones	95.60	4.40	5	Sara Doorsoun-Khajeh	91.70	8.30
6	Thiago Emiliano da Silva	95.50	4.50	6	Francesca Durante	91.20	8.80
7	Rodrigo Hernández Cascante	94.80	5.20	7	Laia Codina Panedas	90.80	9.20
8	Leandro Daniel Paredes	94.60	5.40	8	Emily Ann Sonnett	90.10	9.90
9	Andreas Christensen	94.40	5.60	9	Ella Toone	89.60	10.40
10	César Azpilicueta Tanco	94.40	5.60	10	Maren Nævdal Mjelde	89.50	10.50
11	Victor Nelsson	94.40	5.60	11	Kathrin Julia Hendrich	89.20	10.80
12	Aymeric Laporte	94.20	5.80	12	Keira Walsh	88.70	11.30
13	Ismael Koné	94.10	5.90	13	Saki Kumagai	88.50	11.50
14	Josip Šutalo	93.90	6.10	14	Millie Bright	88.40	11.60
15	Kamil Glik	93.80	6.20	15	Hana Takahashi	88.40	11.60

Figura 7. Ranking de los mejores pasadores de la competición masculina y femenina

En la Figura 8 se presenta un ranking que muestra el porcentaje de pases acertados y fallados realizados por diversos jugadores y jugadoras durante las pasadas copas mundiales de fútbol. Para la elaboración de esta tabla, se ha establecido el requisito de haber realizado un mínimo de 50 pases, con el fin de asegurar la objetividad y representatividad de los datos.

Se puede observar que los jugadores masculinos poseen una mayor efectividad de pases que las jugadoras.

Ranking masculino mejores selecciones pasadoras			Ranking femenino mejores selecciones pasadoras				
Selección	% Pases acertados	% Pases fallados	Selección	% Pases acertados	% Pases fallados		
1	Spain	90.40	9.60	1	Spain	83.20	16.80
2	Brazil	86.90	13.10	2	England	82.40	17.60
3	England	86.50	13.50	3	Germany	81.90	18.10
4	Germany	85.70	14.30	4	Switzerland	79.70	20.30
5	Belgium	85.60	14.40	5	Japan	79.10	20.90
6	Argentina	85.40	14.60	6	Netherlands	78.80	21.20
7	Portugal	84.90	15.10	7	Brazil	77.50	22.50
8	Croatia	84.10	15.90	8	Norway	76.90	23.10
9	Canada	83.90	16.10	9	Canada	76.80	23.20
10	France	83.80	16.20	10	Italy	76.70	23.30
11	Denmark	83.70	16.30	11	France	76.20	23.80
12	Netherlands	83.20	16.80	12	Sweden	74.70	25.30
13	United States	82.80	17.20	13	Portugal	74.60	25.40
14	Switzerland	82.20	17.80	14	Argentina	72.90	27.10
15	South Korea	81.30	18.70	15	Denmark	72.50	27.50

Figura 8. Ranking de las mejores selecciones pasadoras de la competición masculina y femenina

De la competición masculina se observa que siete de las ocho selecciones que alcanzaron, como mínimo, los cuartos de final de la competición se encuentran dentro del top 10 de este ranking. Este dato sugiere que un alto porcentaje de pases acertados contribuye significativamente al rendimiento exitoso de una selección en la competición. En la competición femenina este dato es menor, 5 de las 8 selecciones clasificadas para los cuartos de final se encuentran en el top 10 del ranking.

En la Figura 9, se analiza la relación entre la efectividad de los pases, el valor de goles esperados y otras variables relevantes.

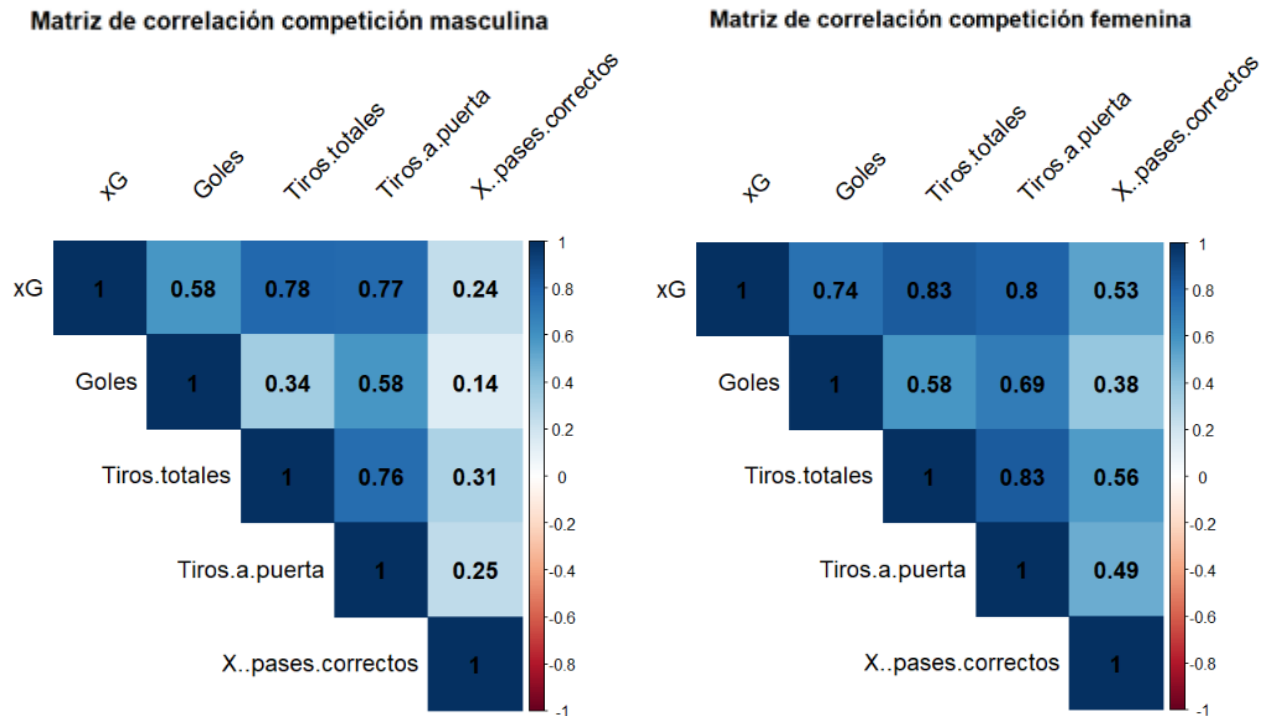


Figura 9. Gráfico de correlaciones entre diferentes variables sobre acciones ofensivas en ambas competiciones

Para la construcción del gráfico de correlaciones anterior, se analizan cinco variables relacionadas con acciones ofensivas: *xG*, *Goles*, *Tiros totales*, *Tiros a puerta* y *Porcentaje de pases correctos* por selección y partido.

Se observa que la mayoría de las variables presentan una relación lineal positiva, ya que el **coeficiente de correlación de Pearson** es superior a 0. Sin embargo, algunas variables están más fuertemente relacionadas entre sí. Un ejemplo notable es la relación entre los *Tiros* y los *Goles esperados* (*xG*), como era de prever. Una mayor cantidad de tiros, tanto a puerta como en total, incrementa el valor de *xG*.

Al fijarse en ambos gráficos, se visualiza que las 5 variables sobre acciones ofensivas, conjuntamente, están más fuertemente correlacionadas entre sí en la competición femenina que en la masculina. Por ejemplo, la correlación entre *Pases correctos* y *xG* es de 0.53 en la competición femenina, frente a 0.28 en la competición masculina, más del doble.

Respecto a variables que explican acciones relativas a acciones defensivas:

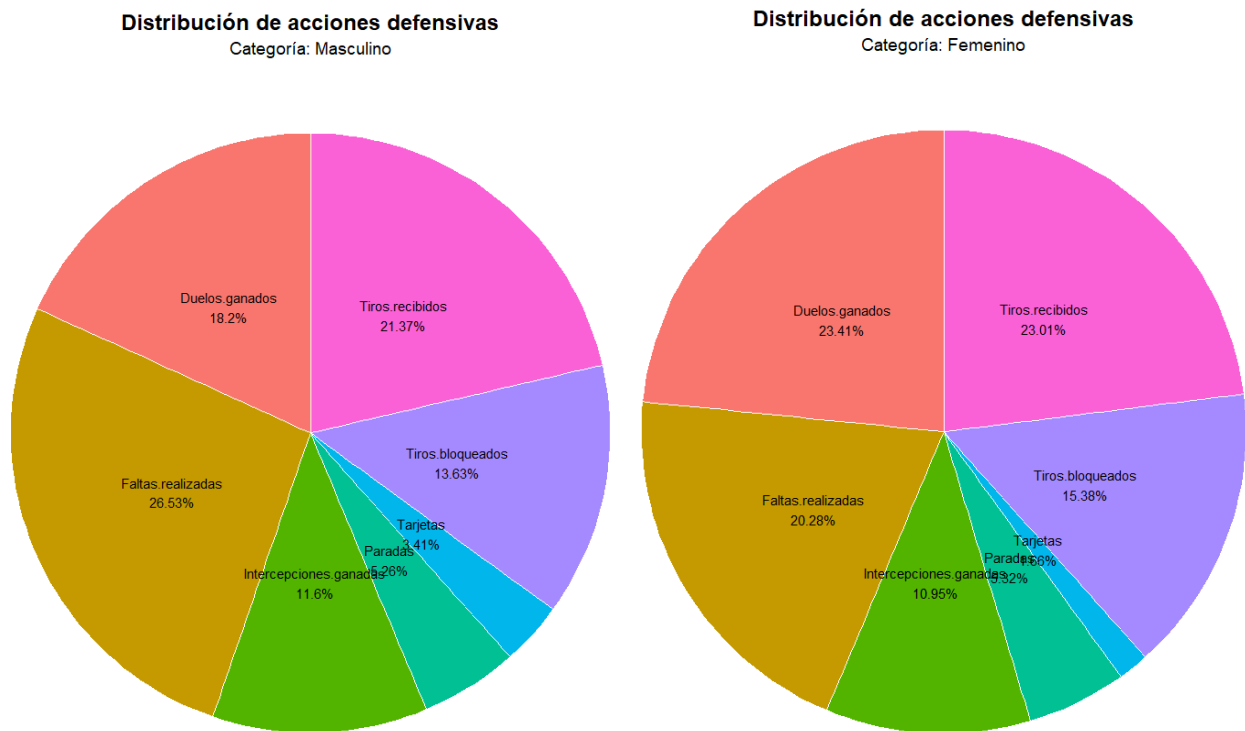


Figura 10. Gráfico de sectores sobre acciones defensivas de ambas competencias

En la Figura 10, se observa la proporción de acciones defensivas de ambas competencias. Existen algunas objeciones que vale la pena destacar:

- Por norma general, la mayor acción defensiva realizada por las selecciones masculinas son las **Faltas realizadas**.
- La mayor acción defensiva realizada por selecciones femeninas son los **Duelos ganados**.
- El porcentaje de **Duelos ganados** por parte de las selecciones femeninas es de 5 puntos superior al de las selecciones masculinas.
- Las selecciones masculinas realizan un mayor número de **Faltas** que las mujeres, 6 puntos superior.
- El resto de las variables defensivas observadas en la Figura 10 tienen unos valores muy parecidos entre ambas competencias, variando 2 puntos arriba o abajo.

Para determinar qué acciones defensivas influyen significativamente en el resultado de un partido y, por ende, en la victoria o derrota de este, se procede a examinar la matriz de correlación, ver Figura 11.

Al evaluar la matriz de correlación, podremos destacar las acciones defensivas que tienen una influencia más significativa en el resultado del partido, lo que proporciona información valiosa para comprender cómo los aspectos defensivos afectan al rendimiento y al resultado del partido.

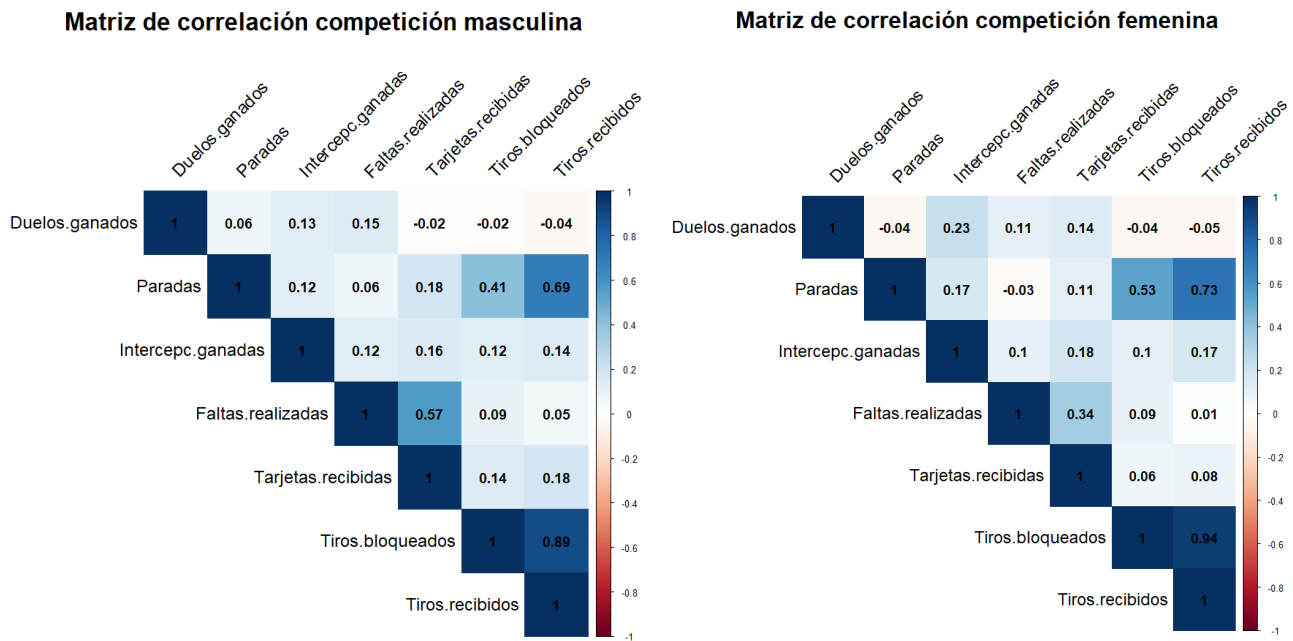


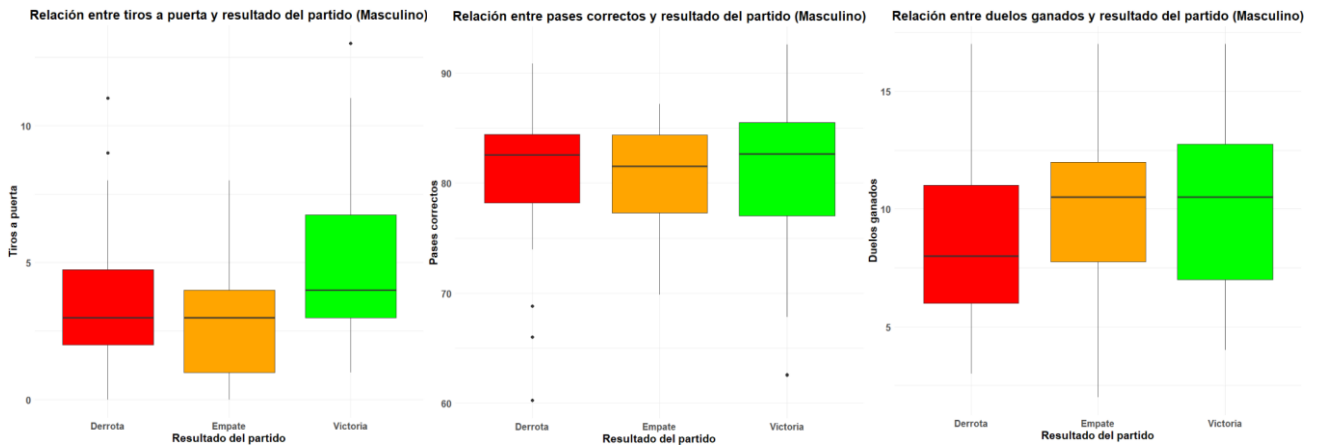
Figura 11. Matriz de correlación de variables de acciones defensivas

En la Figura 11 se puede deducir las siguientes observaciones:

- Las variables **Tiros bloqueados** y **Tiros recibidos** poseen una fuerte relación lineal positiva, con un coeficiente de correlación cercano a 0.90 en ambas competiciones
- La variable **Paradas** también refleja una alta relación lineal positiva junto a la variable **Tiros recibidos** con un coeficiente de correlación cercano a 0.70
- La correlación entre las variables **Faltas realizadas** y **Tarjetas recibidas** difiere dependiendo de la competición. En la competición masculina estas dos variables están más fuertemente correlacionadas entre sí con un coeficiente de correlación cercano al 0.60, mientras que en la competición femenina se acerca al 0.35, por lo que en la competición femenina realizar faltas estaba en una menor medida penalizado con una tarjeta recibida.

A continuación se observa el impacto que tienen algunas de estas variables que explican acciones ofensivas y defensivas sobre el resultado del partido:

Competición masculina:



Competición femenina:

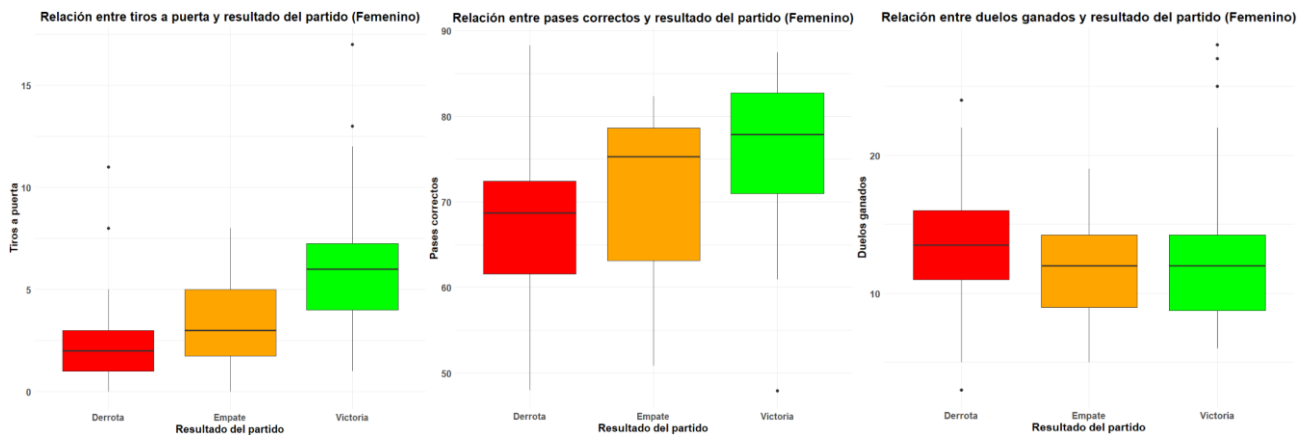


Figura 12. Diagrama de cajas de diferentes variables según el resultado del partido

En la Figura 12 se muestran diferentes diagramas de cajas sobre el resultado del partido tanto para la competición masculina de fútbol como para la femenina.

Destaca lo siguiente:

- En ambas competiciones, se observa en que a medida que aumenta el número de **Tiros a puerta**, el resultado del partido tiende a convertirse en una victoria. Curiosamente, en la competición masculina, cuando el resultado del partido termina en empate tiende a relacionarse con un menor número de **Tiros a puerta** por parte de las selecciones.
- Un mayor porcentaje de **Pases correctos** se relaciona con un mejor resultado final por parte de selecciones de ambas competiciones.
- La relación entre **Duelos ganados** y el resultado del partido difiere entre ambas competiciones. En la competición masculina, un mayor número de duelos ganados se desencadena en una victoria, en cambio en la competición femenina, un mayor número de duelos ganados se relaciona con una derrota por parte de la selección.

Capítulo 4

Clasificación de los equipos de la competición femenina según su estilo de juego

En este capítulo se analiza el conjunto de datos de la competición femenina aplicando la metodología del clustering jerárquico con el objetivo de agrupar selecciones con similitudes entre sí.

4.1. Clustering jerárquico: Marco teórico

El **Clustering Jerárquico**, es un método de minería de datos utilizado para agrupar datos en clústers. Este algoritmo se basa en la distancia entre las observaciones, buscando que los elementos dentro de un mismo clúster sean los más similares entre sí. [26]

En la representación gráfica del clustering jerárquico, los datos se organizan en una estructura jerárquica con forma de árbol, conocida como **dendrograma**. En otras palabras, los elementos se anidan en niveles jerárquicos, mostrando cómo se agrupan de forma progresiva desde clústeres más pequeños hasta formar un único clúster que contiene todos los datos. El proceso de construcción puede ser de dos tipos:

- **Agglomerativo (Bottom-Up)**: Comienza con cada observación en su propio clúster y, en cada paso, combina los dos clústers más cercanos hasta que todas las observaciones están en un solo clúster.
- **Divisivo (Top-Down)**: Comienza con todas las observaciones en un solo clúster y, en cada paso, divide el clúster “más apropiado” hasta que cada observación esté en su propio clúster.

La elección de la métrica de distancia y el método para combinar clústers (como el enlace simple, enlace completo, o promedio) puede afectar significativamente el resultado del agrupamiento.

La técnica de división de clústers utilizada en este trabajo es el dendrograma, un método visual simple en el que se busca el punto en el gráfico en donde hay un gran salto en la distancia de fusión. [7]

Existen otras técnicas para determinar el número óptimo de clústers como por ejemplo:

- **Método del codo**: El punto donde la disminución de la suma de las distancias se vuelve menos pronunciada es el "codo" y puede sugerir el número óptimo de clústeres.
- **Coefficiente de Silueta**: Este coeficiente mide cuán similares son los objetos dentro de un clúster en comparación con los objetos de otros clústeres. Un valor alto de silueta indica una buena separación entre los clústeres.
- **Criterio de Gap**: Este criterio compara la dispersión dentro de los clústeres con una referencia aleatoria, ayuda a determinar el número de clústeres observando la diferencia entre el logaritmo de la dispersión intra-clústeres y la dispersión esperada bajo un modelo aleatorio

4.2. Clustering jerárquico: Procedimiento

Una vez explicado en el punto 4.1 el marco teórico del clustering jerárquico, en esta sección se lleva a cabo la explicación de las diferentes fases que se tienen que realizar con el conjunto de datos para llegar al objetivo de la división de las selecciones en diferentes grupos que posean similitudes entre sí. [6]

1. Una vez cargados los datos y seleccionados las variables necesarias, se estandarizarán estas para garantizar una comparación equitativa entre ellas y para facilitar la interpretación de los resultados.
2. Calcular la distancia Euclidiana entre las diferentes variables ya que proporciona una medida intuitiva y eficiente de la similitud entre observaciones, lo que permite agrupar datos de manera jerárquica y visualizar posteriormente las relaciones entre ellos mediante un dendrograma.
3. En este paso se aplica el clustering jerárquico utilizando el método Ward para minimizar la varianza dentro de cada clúster/grupo.
4. Visualizar el dendrograma tras la aplicación del clustering jerárquico y decidir el número de clústers más apropiado.
5. Aplicar el análisis de componentes principales al conjunto de datos iniciales y decidir si su aplicación mejora el modelo.
6. Revisar los valores propios asociados a cada componente principal ya que nos ayuda en la decisión de la elección del número de clústers.
7. Decidir si la aplicación del análisis de componentes principales distribuye las observaciones de una manera más homogénea y permite una mejor identificación en la decisión de la división de clústers.
8. Exposición de los diferentes clústers elegidos y características de cada uno.

4.3. Aplicación del clustering jerárquico al conjunto de datos de la competición femenina

En este análisis clustering, no se utilizan todas las variables disponibles, sino únicamente las 10 variables siguientes que describen el estilo de juego de las diferentes selecciones:

- *Selección*
- *Duelos ganados*
- *Pases correctos*
- *Tiros a puerta*
- *Goles*
- *xG (Goles esperados)*

- *Intercepciones ganadas*
- *Faltas realizadas*
- *Faltas recibidas*
- *Tiros recibidos*

Se aplica el procedimiento descrito en la sección 4.2 hasta la fase 4 del mismo. El dendograma obtenido es el siguiente:

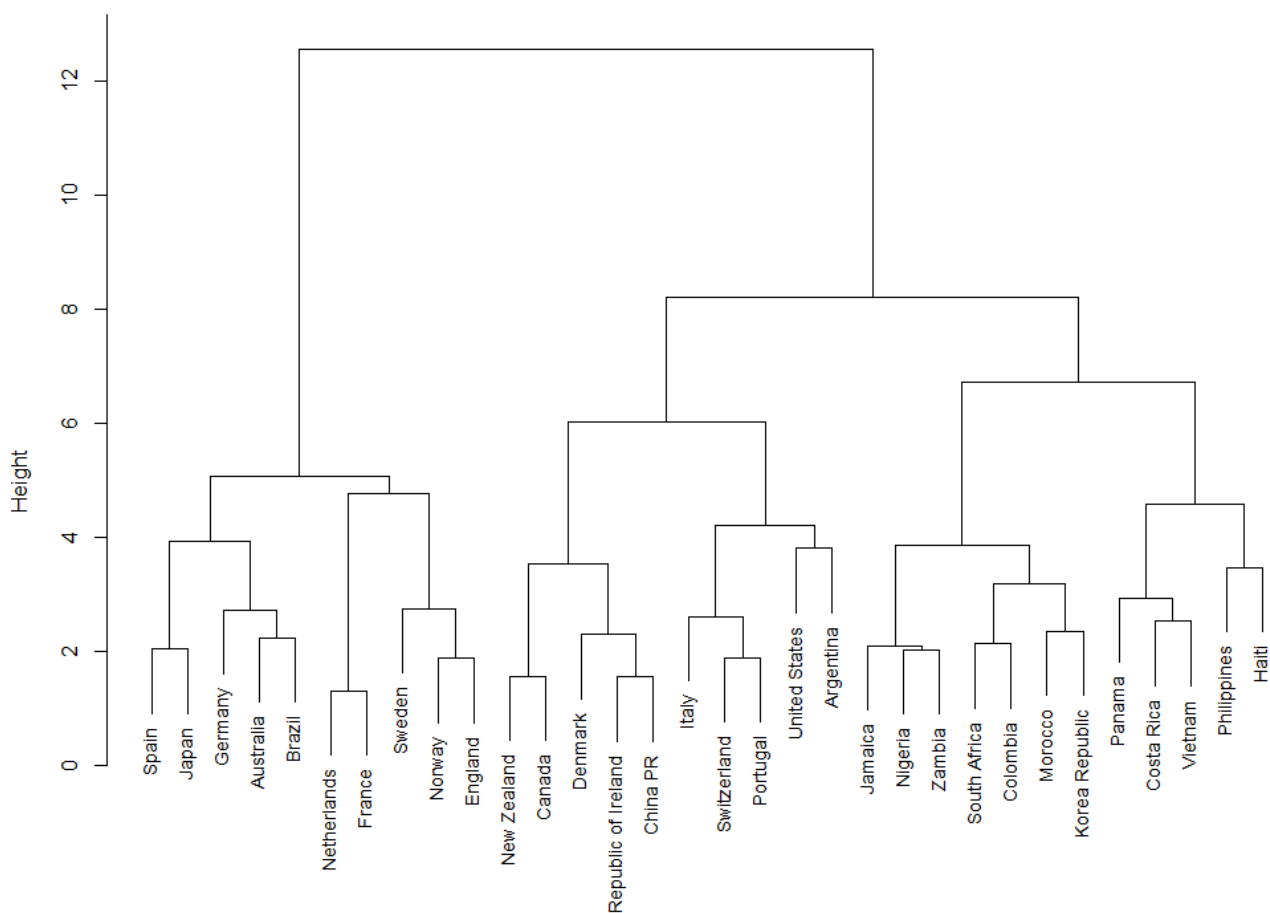


Figura 13. Gráfico Dendograma

Tras observar el dendrograma de la Figura 13, surge la cuestión de elegir entre 3 o 4 clústers, estableciendo el corte a una distancia superior o inferior. Existen diferentes criterios para la elección del número de clúster como se ha explicado en el punto 4.1, pero se decide optar por la selección de 3 clústeres, aplicando la técnica del dendrograma tras observar un nivel en donde la distancia entre clústers aumenta significativamente, formando así “ramas” que están claramente separadas. El corte se establece a una distancia inferior a 8.

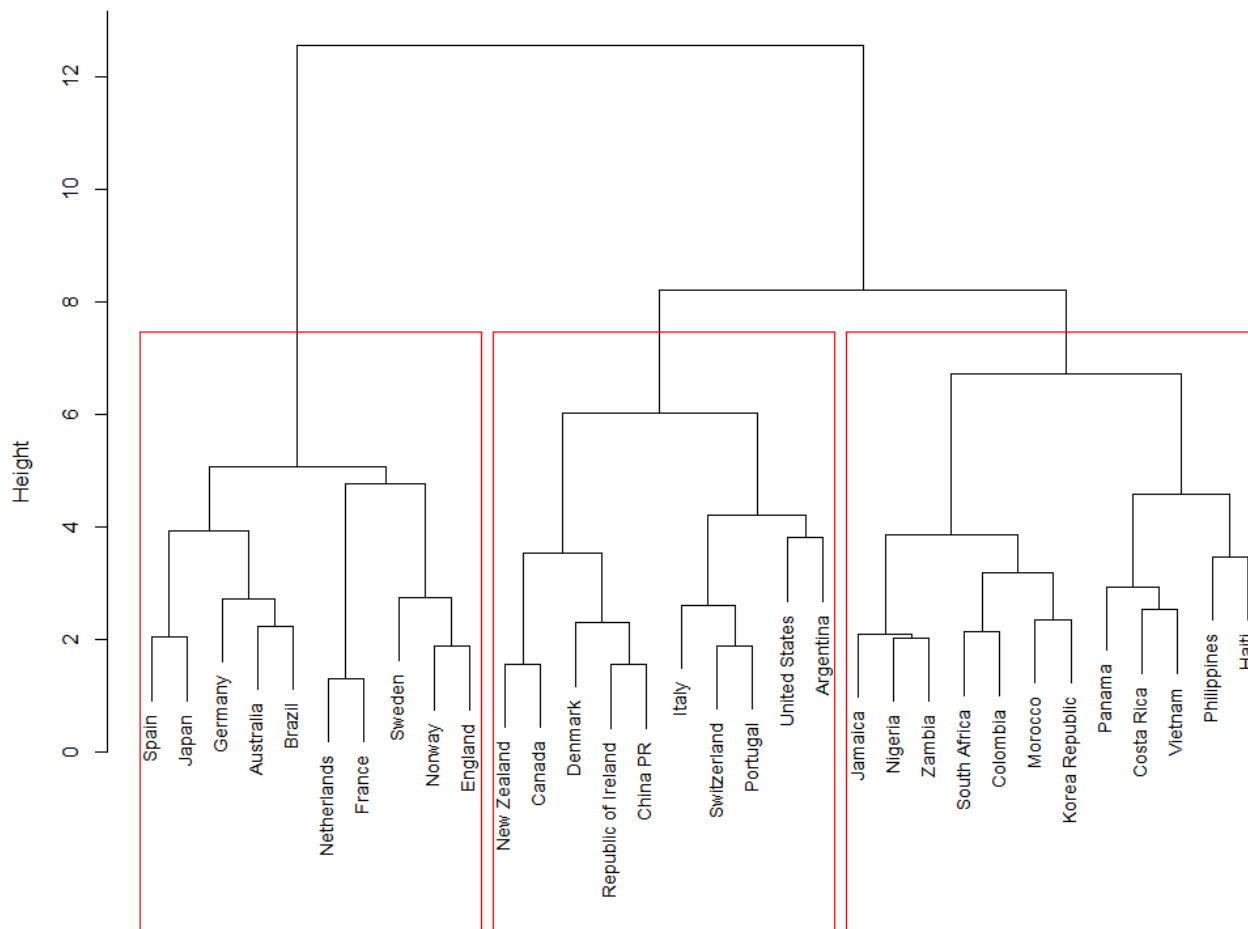


Figura 14. Dendrograma con el número final de clústers elegidos

En la Figura 14 se visualiza el número de clústers en los que se ha dividido el conjunto de datos. 3 clusters compuestos de 10, 10 y 12 selecciones cada uno.

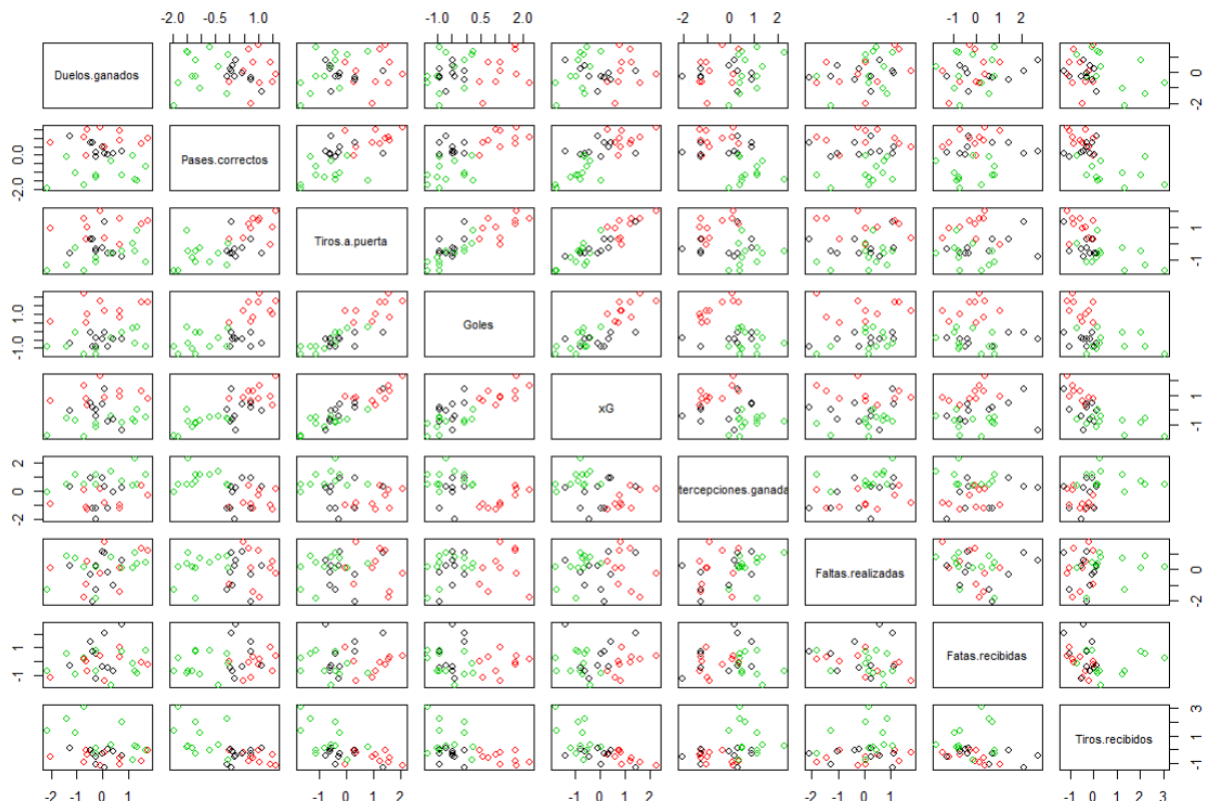


Figura 15. Gráficos de dispersión entre las variables que muestran la clasificación en 3 grupos

En la Figura 15, se observa la agrupación de las selecciones en los 3 grupos establecidos. Algunas variables, como **Goles** y **Pases Correctos**, muestran claramente esta división ya que las selecciones pertenecientes a cada clúster están claramente agrupadas de forma diferenciada.

A continuación, se aplica el análisis de componentes principales (ACP) a este conjunto de datos para evaluar si mejora el agrupamiento. Este ACP tiene como objetivo reducir el número de dimensiones mientras se recoge gran parte de la varianza de los datos.

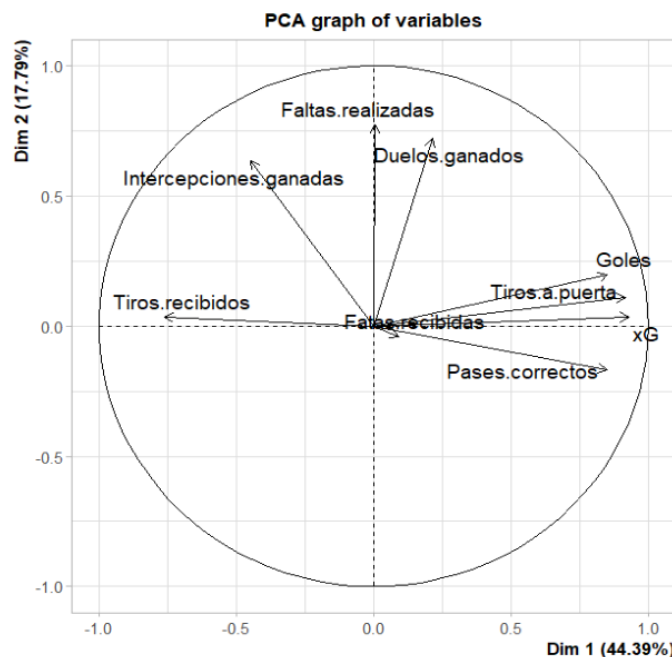


Figura 16. Gráfico de 2 dimensiones tras aplicar ACP

La Figura 16 representa las variables originales en el espacio de las componentes principales. Las variables que están más alejadas del origen y cercanas al círculo unitario están mejor representadas en el plano de los componentes principales seleccionadas. Este gráfico muestra cómo las variables están correlacionadas con los componentes principales.

De este gráfico se deduce que las variables *Goles*, *Tiros a puerta* y *xG* están positivamente asociadas entre sí, mientras que las variables *Tiros recibidos* y *Pases correctos* están negativamente asociadas entre sí.

Se puede observar que utilizando solamente 2 dimensiones **explicamos más del 60% de la variabilidad** del conjunto de datos.

La dimensión 1 explica el 44% de la variabilidad de los datos y agrupa las variables:

- *Goles*
- *Tiros a puerta*
- *xG*
- *Duelos ganados*
- *Faltas realizadas*

La dimensión 2 explica el 17% de la variabilidad de los datos y agrupa las variables:

- *Intercepciones ganadas*
- *Tiros recibidos*
- *Faltas recibidas*
- *Pases correctos*

Como se ha mencionado en el apartado 4.1 de esta memoria, la visualización de los valores propios obtenidos al aplicar el análisis de componentes principales es un método adecuado para determinar el número correcto de grupos en los que dividir el conjunto de datos.

La regla de *Kaiser* explica que hay que retener todas las componentes principales con valores propios mayores que 1. Esto se basa en la idea de que un componente con un valor propio mayor que 1 explica más varianza que una variable original estandarizada.

	Valor propio	Porcentaje de varianza	Porcentaje de varianza acumulada
comp 1	4.00	44.39	44.39
comp 2	1.60	17.79	62.19
comp 3	1.22	13.51	75.70
comp 4	0.69	7.67	83.36
comp 5	0.61	6.80	90.17
comp 6	0.44	4.94	95.11
comp 7	0.22	2.42	97.52
comp 8	0.14	1.57	99.09
comp 9	0.08	0.91	100.00

Figura 17: Tabla de valores propios asociados a cada componente principal

Puesto que la última componente principal con un valor propio mayor a 1 es la 3ª componente, véase Figura 17, elegimos este número de componentes las cuales explican más del 75% de la variabilidad de los datos.

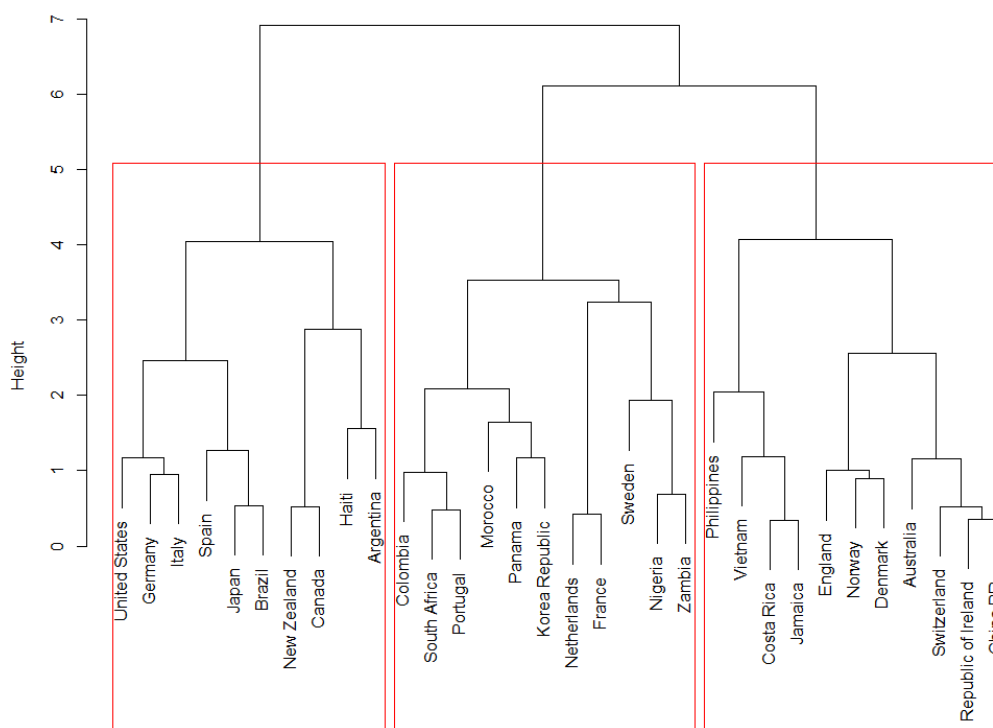


Figura 18. Dendrograma tras aplicar ACP

En la Figura 18 se observa de una manera más clara y concisa que son 3 los grupos en los que se tiene que dividir el conjunto de datos para así agrupar a las selecciones de una mejor forma de acuerdo con las variables analizadas.

Las diferentes selecciones encuadradas en los 3 nuevos grupos creados de acuerdo con las variables analizadas es el siguiente:

- **Grupo 1:** Argentina, Brasil, Canadá, China, Alemania, Haití, Italia, Japón Nueva Zelanda, España y Estados Unidos.
- **Grupo 2:** Australia, China, Costa Rica, Dinamarca, Inglaterra, Jamaica, Noruega, Filipinas, Irlanda, Suiza, Vietnam.
- **Grupo 3:** Colombia, Francia, Corea, Marruecos, Países Bajos, Nigeria, Panamá, Portugal, Sudáfrica, Suecia y Zambia.

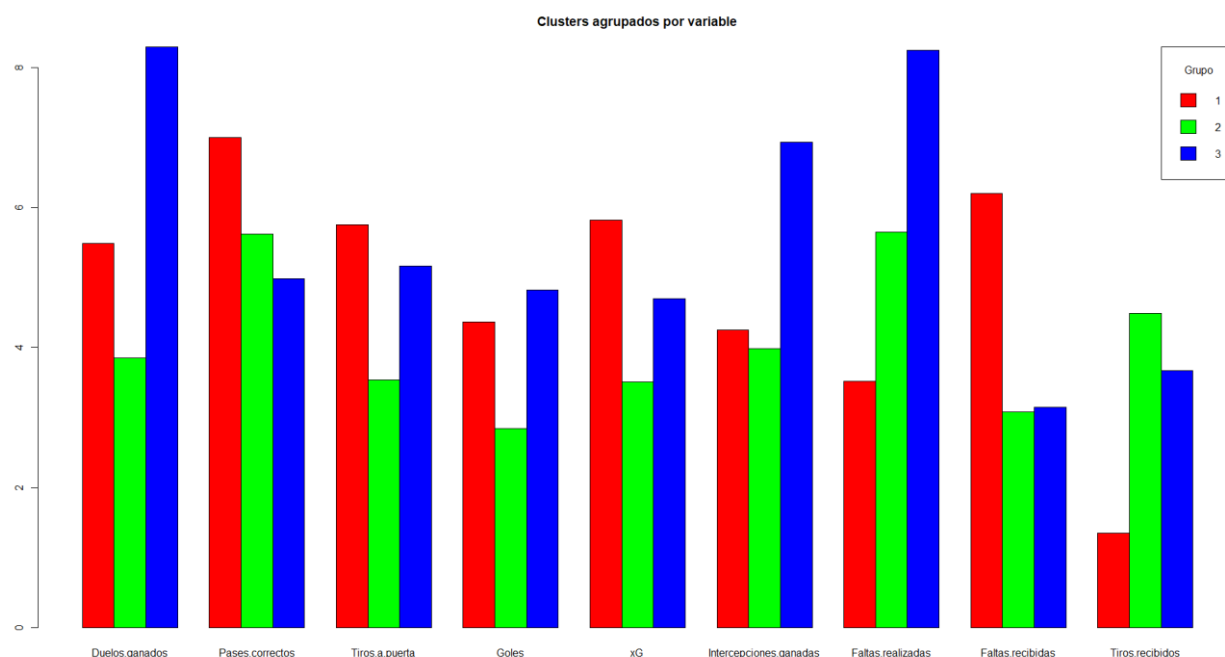


Figura 19. Gráfico de barras agrupados por clústers

Tras haber aplicado el clustering, se ha obtenido el gráfico de la Figura 19. Este muestra qué variables predominan en cada grupo de selecciones.

Se observa que en el grupo 1 de selecciones, donde se encuentra la selección española, campeona del mundial, predominan las acciones ofensivas. Este grupo se caracteriza por un índice alto de **Pases correctos** y **Tiros a puerta**, también se observa un alto número de **Faltas cometidas** y un bajo número de **Tiros recibidos**. Estos indicadores sugieren que las selecciones de este grupo pasaban mucho tiempo en el campo rival buscando el gol.

Por el contrario, las selecciones que pertenecen al grupo 3 adoptan un enfoque más defensivo y duro en su estilo de juego. Se destacan por realizar un alto número de **Faltas** y por liderar en **Intercepciones** y **Duelos ganados**. Sin embargo, este enfoque no implica que sea el incorrecto, ya que estas selecciones muestran un índice de goles significativos, aunque ligeramente menor en comparación con el grupo 1.

Respecto a las selecciones pertenecientes al grupo 2, se caracterizan por ocupar las últimas posiciones en prácticamente todas las variables analizadas. Son las selecciones con el menor número de **Goles**, **Tiros a puerta**, **Duelos ganados** e **Intercepciones**. Además, se destacan por ser las que reciben más tiros por parte de sus oponentes.

Tras analizar esto, podría parecer que pertenecer al grupo 2 implica tener un rendimiento inferior en la competición, dado que estadísticamente las selecciones de este grupo muestran números más bajos. Sin embargo, al dividir el conjunto de datos en solo 3 grupos/clústers y considerar un número limitado de observaciones (32), existen selecciones que destacan dentro de cada grupo, esas son:

- Grupo 1: España (Selección campeona de la competición)
- Grupo 2: Inglaterra (Selección subcampeona de la competición)
- Grupo 3: Suecia (3ª clasificada)

Capítulo 5

Aplicación shiny: WorldCupViz

WorldCupViz es la aplicación shiny desarrollada durante este proyecto académico. Diseñada para analizar y visualizar conjuntos de datos de las recientes *Copas Mundiales de la FIFA* masculinas y femeninas de los años 2022 y 2023, WorldCupViz ofrece una plataforma interactiva para explorar estadísticas detalladas y tendencias clave de los torneos.

La aplicación está estructurada en diversas secciones que permiten a los usuarios investigar el rendimiento de equipos y jugadores a lo largo de las competiciones. Incluye análisis ofensivos y defensivos, estadísticas individuales, así como métricas avanzadas como **xG** (goles esperados) y **xGA** (asistencias esperadas). Cada sección está diseñada para proporcionar una experiencia de usuario intuitiva y enriquecedora.

Además de las visualizaciones estándar como tablas y gráficos, WorldCupViz también incorpora elementos interactivos como mapas de acciones, radar de desempeño individual y comparaciones estadísticas entre jugadores. Estas herramientas permiten a los usuarios profundizar en los detalles tácticos y estratégicos que influyen en los resultados de los partidos.

La aplicación no solo facilita el análisis amplio de datos, sino que también mejora la comprensión general de los eventos de la copa mundial, destacando patrones y actuaciones a lo largo de los torneos.

WorldCupViz representa un paso significativo hacia la visualización avanzada de datos deportivos en el contexto de eventos globales de alta relevancia, proporcionando a los investigadores, analistas y entusiastas del fútbol una herramienta valiosa para explorar y entender mejor los aspectos clave de ambos mundiales, y como se comenta más adelante, es útil para analizar futuras competiciones de fútbol.

A través de este enlace se puede acceder a la aplicación web WorldCupViz:

<http://shiny1.eio.uva.es:3838/luis/App/WorldCupViz/>

5.1. Shiny y paquetes utilizados

La aplicación se implementa utilizando el paquete shiny. Este paquete simplifica considerablemente la creación de aplicaciones web interactivas directamente desde R. Esta elección se fundamenta en la capacidad de shiny para gestionar complejos conjuntos de datos y presentarlos de manera accesible, lo cual es crucial para ofrecer una experiencia fluida y enriquecedora a los usuarios de la aplicación.

Shiny proporciona una estructura robusta y flexible que permite integrar visualizaciones dinámicas, tablas interactivas y herramientas de análisis de datos de manera efectiva. Además, ofrece una experiencia de usuario intuitiva al permitir la actualización en tiempo real de gráficos y resultados según las selecciones y filtros realizados por el usuario.

5.1.1 Estructura básica de una aplicación construida con shiny

Las aplicaciones shiny se encuentran encapsuladas en un único script denominado `app.R`, el cual consta de dos componentes principales y una llamada para ejecutar la aplicación.

Este enfoque permite estructurar de manera organizada tanto la interfaz de usuario interactiva como la lógica de procesamiento de datos detrás de la aplicación. La estructura clara y concisa del script `app.R` facilita la incorporación de elementos visuales, como gráficos y tablas dinámicas, así como la integración de funcionalidades avanzadas de análisis de datos, todo dentro del entorno familiar de R. Esta metodología asegura que el desarrollo y mantenimiento de la aplicación shiny sean eficientes y efectivos, optimizando así la experiencia del usuario final. [8] [25]

1. Archivo `ui.R`

El archivo `ui.R` define la interfaz de usuario (UI) de la aplicación, incorporando elementos como entradas de usuario (botones, selectores, etc.), salidas visuales (gráficos, tablas, texto) y diseños estructurales (páginas, paneles). En esta sección de la aplicación, algunas de las funciones más utilizadas incluyen:

- ***fluidPage()***: Esta función es fundamental para crear una página web adaptable y fluida que se ajuste automáticamente al tamaño del dispositivo del usuario.
- ***selectInput()***: Utilizada para generar un menú desplegable que permite al usuario seleccionar una opción de una lista predefinida, facilitando así la interacción y selección de parámetros.
- ***plotOutput()***: Función esencial para establecer un contenedor en la interfaz de usuario donde se visualizarán gráficos dinámicos generados por el servidor shiny, proporcionando una representación visual de los datos procesados.

Estas funciones no solo estructuran la apariencia visual de la aplicación, sino que también aseguran una experiencia de usuario intuitiva y funcional, optimizando la interacción con los datos y resultados presentados

2. Archivo `server.R`

En el contexto del desarrollo de la aplicación, el archivo encargado de definir la lógica y procesar los datos introducidos por el usuario se conoce como `server.R`. Este archivo despliega funciones clave que manejan las entradas de la interfaz de usuario, procesan estas entradas y generan las salidas correspondientes. A continuación, se especifican algunas de las funciones fundamentales utilizadas en esta sección de la aplicación:

- ***renderTable()***: Esencial para renderizar y mostrar tablas dinámicas en la interfaz de usuario, permitiendo la visualización clara y estructurada de datos procesados.

- *reactive()*: Una función fundamental que facilita la creación de expresiones reactivas. Estas expresiones se diseñan para ejecutarse automáticamente cuando alguna de sus dependencias cambia, asegurando así que los datos y resultados presentados se actualicen dinámicamente en respuesta a las acciones del usuario.
- *observe()*: Utilizada para monitorear y detectar cambios en las entradas de usuario, así como en las expresiones reactivas. Esto permite ejecutar código específico en respuesta a estos cambios, optimizando la interactividad y la capacidad de respuesta de la aplicación.

5.1.2 Funcionamiento de una aplicación shiny

1. Interactividad:

El usuario interactúa con los elementos de la interfaz de usuario, los cuales generan eventos que son capturados por el servidor shiny para ejecutar acciones específicas. Estos eventos pueden incluir desde la selección de opciones en menús desplegados hasta el clic en botones o la introducción de datos en campos de entrada. Cada vez que ocurre un evento, el servidor shiny reacciona procesando los datos introducidos, calculando resultados o actualizando la visualización en la interfaz de usuario.

Los eventos capturados permiten no solo actualizar la interfaz de usuario en tiempo real, sino también mantener la coherencia entre las acciones del usuario y las operaciones realizadas en el servidor para procesar y presentar los datos de manera adecuada. [9]

2. Reactividad:

Los cambios en las entradas de la interfaz de usuario automáticamente generan la reevaluación de las funciones reactivas en el servidor. Este proceso asegura que cualquier modificación en los datos de entrada, como la selección de opciones en menús desplegados o la introducción de nuevos valores, provoque una actualización inmediata de las salidas correspondientes en la interfaz gráfica.

3. Actualización Dinámica:

Cuando ocurren cambios en el servidor, las salidas correspondientes en la interfaz de usuario se actualizan dinámicamente. Esta funcionalidad es muy importante para mantener sincronizados los datos presentados al usuario con las acciones realizadas o los datos introducidos en la aplicación.

5.1.3 Otros paquetes de R utilizados

Para llevar a cabo el desarrollo de la aplicación WorldCupViz, se utiliza una amplia variedad de paquetes para la elaboración tanto de gráficos como de tablas. Estos paquetes son fundamentales para crear visualizaciones interactivas y tablas dinámicas para la presentación y el análisis de los datos de las *Copas Mundiales de la FIFA*.

- **plotly**: Paquete que permite crear gráficos interactivos de alta calidad, desde simples diagramas de dispersión hasta mapas interactivos. [16]
- **gganimate**: Paquete que extiende las capacidades de ggplot2, permitiendo la creación de animaciones a partir de gráficos. [17]
- **DT**: Interfaz para tablas de datos que utiliza una biblioteca de JavaScript ampliamente reconocida, agregando interactividad a las tablas HTML. [18]
- **htmltools**: Herramienta para trabajar con contenido HTML, facilitando la integración y manipulación de elementos HTML en la aplicación. [19]
- **jsonlite**: Herramienta para trabajar con datos en formato JSON, un formato comúnmente utilizado para intercambiar datos en aplicaciones web debido a su simplicidad y capacidad de ser fácilmente leído y escrito tanto por humanos como por máquinas. [20]
- **kableExtra**: Extensión del paquete "knitr" que permite la creación de tablas HTML y LaTeX altamente personalizadas y mejoradas. [21]
- **soccermatics**: Paquete que proporciona herramientas para visualizar el seguimiento espacial y los datos de eventos de partidos de fútbol, utilizado principalmente para generar mapas de calor y visualizaciones de pases. [22]
- **fmsb**: Herramienta para visualizar datos multivariados mediante gráficos de radar. [23]
- **SBpitch**: Paquete que permite crear gráficos en forma de campo de fútbol personalizables utilizando ggplot2. [24]

5.2. Estructura de la aplicación:

La aplicación WorldCupViz está compuesta de diversas páginas dentro de la interfaz web interactiva shiny, donde el usuario puede tomar el control completo para analizar las diferentes estadísticas de la competición. Como se ha mencionado anteriormente, ambos conjuntos de datos, tanto los relativos al campeonato masculino de las pasadas copas mundiales de fútbol como al femenino, están integrados en esta aplicación.

A continuación, se detalla el funcionamiento, la estructura y el resultado esperado de cada página, además de comentar la construcción de diversos gráficos innovadores y de elaboración propia.

Al cargar la aplicación, se visualiza la portada principal. En esta sección se presentan imágenes de los campeones de ambas competiciones de fútbol: la masculina, disputada en 2022 en Qatar, y la femenina, disputada en 2023 en Australia/Nueva Zelanda.

En la portada, (véase la Figura 20), se invita al usuario a seleccionar el contenido que desea visualizar, ya sea el de la competición masculina o femenina. Una vez realizada esta elección, se despliega el menú correspondiente con las diferentes páginas disponibles para que el usuario pueda explorar y analizar el contenido detallado de cada sección.

Portada de la aplicación



Figura 20. Portada de la aplicación WorldCupViz

También aparecen dos botones en la parte superior derecha con la misma funcionalidad que los botones previamente mencionados. La particularidad de estos botones radica en que, sin importar en qué página se encuentre el usuario, permiten cambiar de competición con solo pulsar uno u otro botón. Esto facilita al usuario agilizar el proceso de cambio de competición y la navegación dentro de la aplicación.

Una vez seleccionada la competición deseada, se despliegan las diferentes secciones de la aplicación en el menú situado a la izquierda de la pantalla.

Menú de la competición

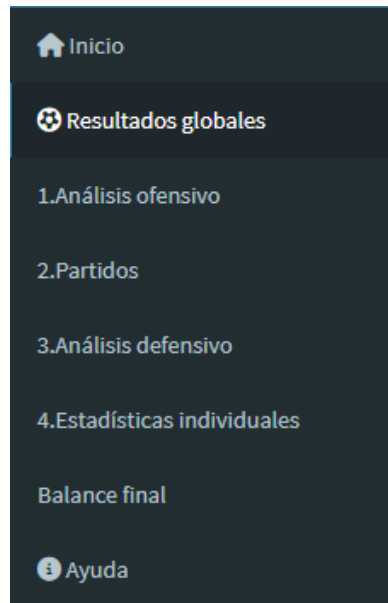


Figura 21. Menú de la competición

Excluyendo la portada, son 7 las secciones que se ofrecen dentro de cada competición seleccionada. (Véase la Figura 21) A grandes rasgos, en cada una de ellas se incluye lo siguiente:

- **Página Resultados globales:** Se incluye información acerca de los resultados de toda la competición tanto de fase de grupos como de la fase final, con diferentes imágenes y desplegables que ayudan al usuario a tener una mejor experiencia dentro de la página
- **Página Análisis Ofensivo:** Proporciona una visión general de las estadísticas ofensivas de todas las selecciones, permitiendo al usuario seleccionar equipos específicos para un análisis más detallado. Los gráficos y tablas presentados muestran las relaciones entre goles, tiros y asistencias, destacando los jugadores más relevantes en estos aspectos.
- **Página Análisis Defensivo:** Permite al usuario explorar las estadísticas acciones defensivas de cada selección, visualizando mapas de acciones defensivas y estadísticas globales de tiros recibidos. La estructura incluye gráficos que representan la cantidad y ubicación de las acciones defensivas y los porcentajes de los diferentes tiros.
- **Página Partidos:** Ofrece una visión detallada de cada partido, presentando resultados, estadísticas generales de cada equipo, mapas de influencia de pases y métricas de xG y xGA. El usuario puede seleccionar partidos específicos y ver gráficos que muestran el rendimiento ofensivo y el jugador más valioso del partido.
- **Página Estadísticas Individuales:** Permite al usuario seleccionar un equipo y un jugador para ver datos básicos y detallados sobre su rendimiento. Incluye gráficos que comparan acciones individuales, localización y precisión de pases, y un radar global que muestra el desempeño del jugador en varias variables.
- **Página Balance final:** Muestra al usuario los logros más relevantes de la competición mediante imágenes.

- **Página Ayuda:** Proporciona al usuario explicaciones acerca de diferentes gráficos, tablas y mapas de la aplicación. Página de gran utilidad para el usuario que no es gran conocedor del mundo futbolístico.

Una vez introducido el contenido de cada página, el siguiente paso es profundizar en cada página comentando sus diferentes gráficos, tablas, información...

Página Resultados globales

Esta página, compuesta por 2 pestañas, muestra los resultados y clasificación de cada selección en la competición de la Copa Mundial de la FIFA



Figura 22. Pagina Resultados globales, pestaña “Cuadro general”

- **Cuadro general:** La primera pestaña de esta página, (véase la Figura 22), muestra una imagen con los resultados de los diferentes partidos disputados en la fase final de la competición, desde octavos de final hasta la gran final.



Figura 23. Página Resultados globales, pestaña “Fases y resultados”

- **Fases y resultados:** La segunda pestaña de esta página, como se puede observar en la Figura 23, presenta todos los resultados de la competición, tanto de la fase final como de la fase clasificatoria. En esta sección, se solicita al usuario que seleccione la fase de grupos o la fase final que desea visualizar.

Página Análisis Ofensivo

Esta página, compuesta por tres pestañas, ofrece información sobre acciones ofensivas tanto de manera general como específica.

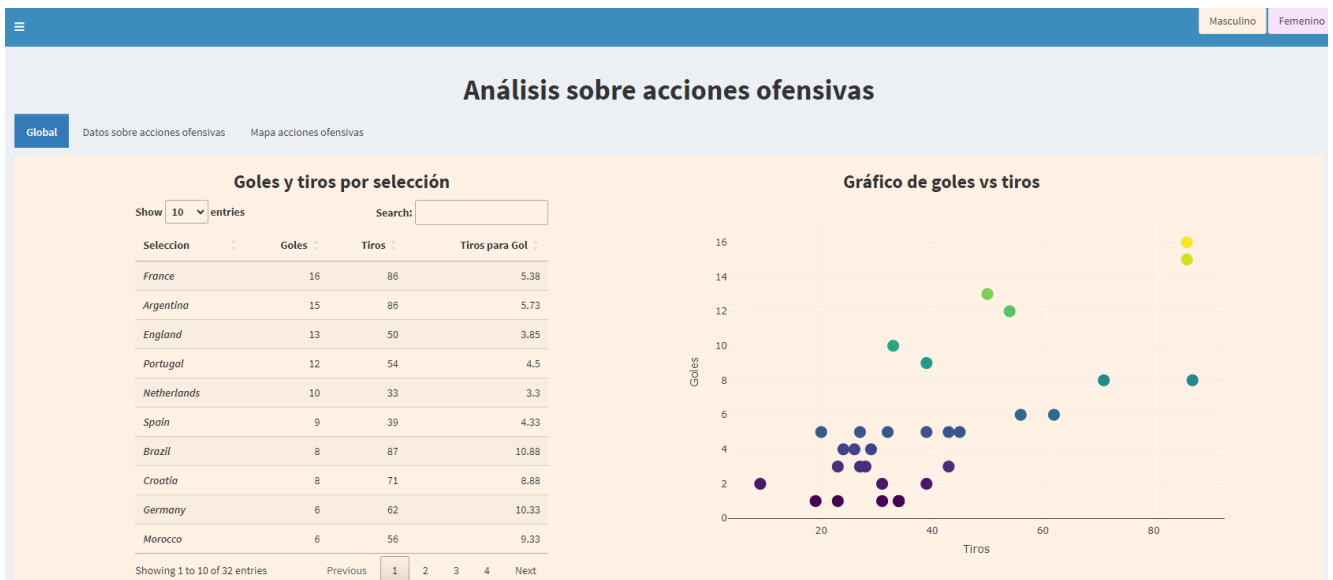


Figura 24. Página Análisis ofensivo, pestaña “Global”

- Global:** En esta primera pestaña, como se ilustra en la Figura 24, se visualiza una tabla y un gráfico relacionados con acciones ofensivas de manera general. Mediante una tabla interactiva y un gráfico de dispersión también interactivo, se presentan acciones como los goles, tiros y tiros esperados para marcar gol. Estos gráficos permiten identificar qué selecciones son las más goleadoras de la competición y cuáles requieren un mayor o menor número de tiros para convertirlos en goles.



Figura 25. Página Análisis Ofensivo, pestaña “Datos ofensivos por selección”

- Datos sobre acciones ofensivas:** La siguiente pestaña de esta página, relativa a las acciones ofensivas, como se ilustra en la Figura 25 se proporciona información más específica sobre las acciones ofensivas de cada selección. Se despliegan dos imágenes correspondientes a los jugadores con el mayor número de goles y asistencias de la selección en la competición elegida por el usuario. Adicionalmente, mediante un gráfico de barras y una tabla interactiva, se visualizan los datos de goles, asistencias y minutos totales disputados por cada jugador a lo largo del torneo.



Figura 26. Página Análisis Ofensivo, pestaña “Mapa ofensivo”

- **Mapa acciones ofensivas:** Por último, en esta pestaña, como se muestra en la Figura 26 se visualiza un mapa de calor y un gráfico. En el mapa de calor se presenta la localización de los goles y disparos del equipo seleccionado. El gráfico de su derecha muestra los tiros esperados necesarios para que la selección convierta una jugada en gol, en comparación con otras selecciones que han requerido un mayor o menor número de disparos.

Página Partidos

En esta página se muestra una ventana dividida en 2 secciones, en la parte superior de ella se informa al usuario tanto del resultado del partido, como del tipo de partido, y en la parte inferior de la página se muestran 4 pestañas en donde se analiza cada uno de los 64 partidos que se han disputado en las pasadas copas mundiales de fútbol.

En ella el usuario elige un partido del que se analizará en profundidad mediante todo tipo de estadísticas, gráficos y tablas.

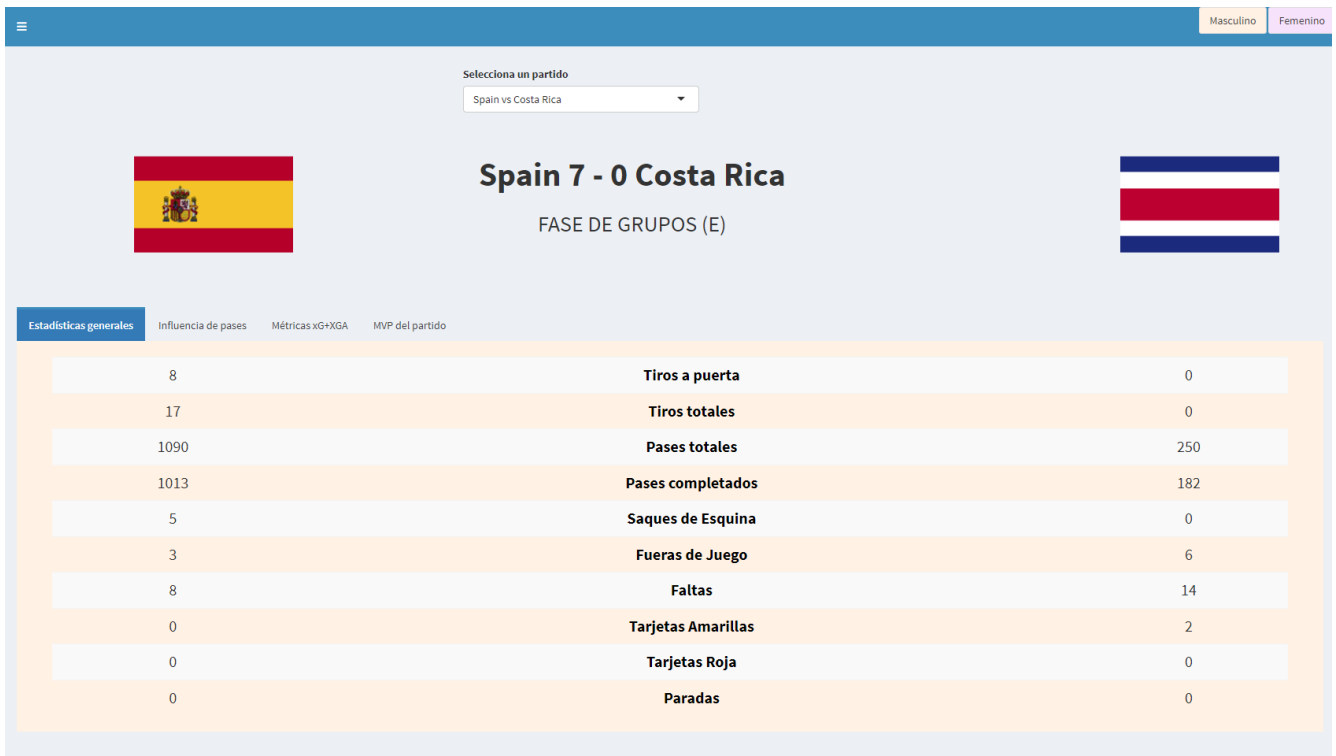


Figura 27. Página Partidos, pestaña “Estadísticas generales”

- **Estadísticas generales:** En esta primera pestaña, véase Figura 27, se muestran 10 de las estadísticas más relevantes del partido.

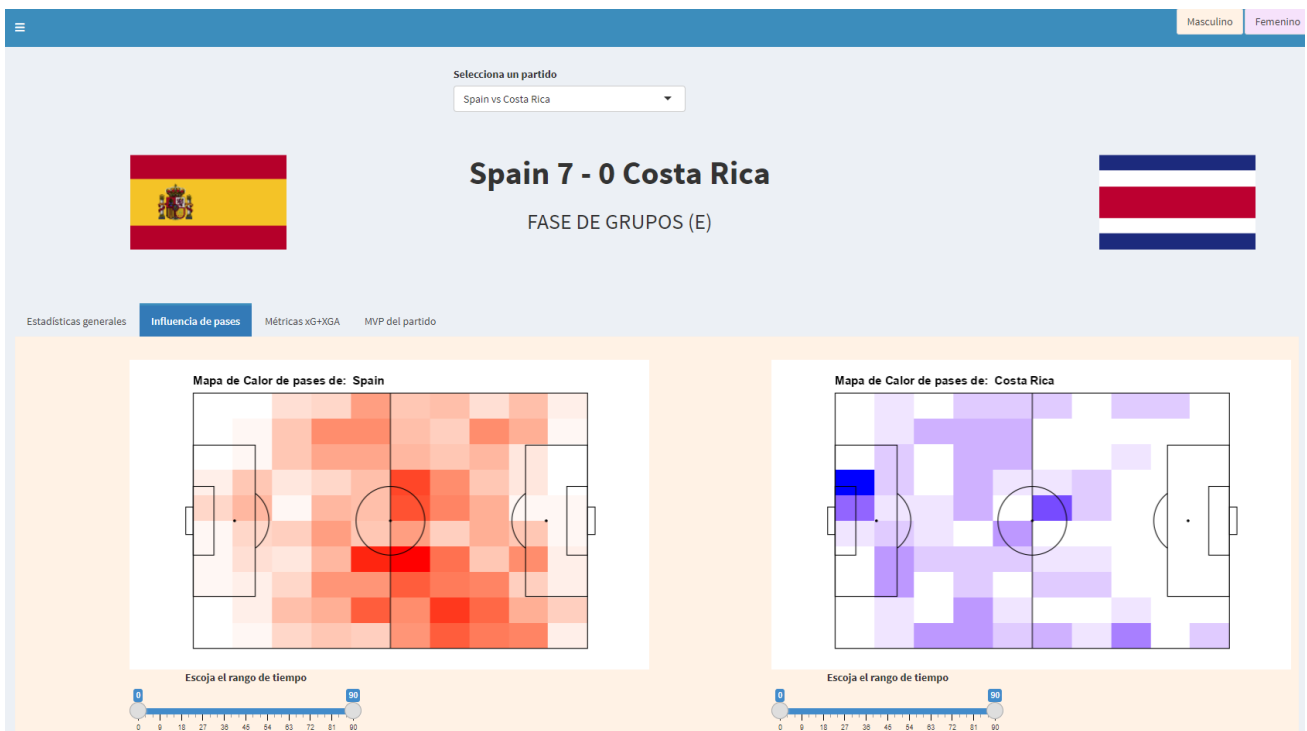


Figura 28. Página Partidos, pestaña “Influencia de pases”

- Influencia de pases:** La segunda pestaña de esta página, como se muestra en la Figura 28, presenta dos mapas de calor que destacan las zonas del campo de juego más frecuentadas por los jugadores de cada equipo desde donde han realizado los pases. Mediante una barra de selección, el usuario puede elegir el rango de tiempo, desde el minuto 0 hasta el minuto 90 de cada partido, y observar las áreas del campo que han predominado más para cada selección.

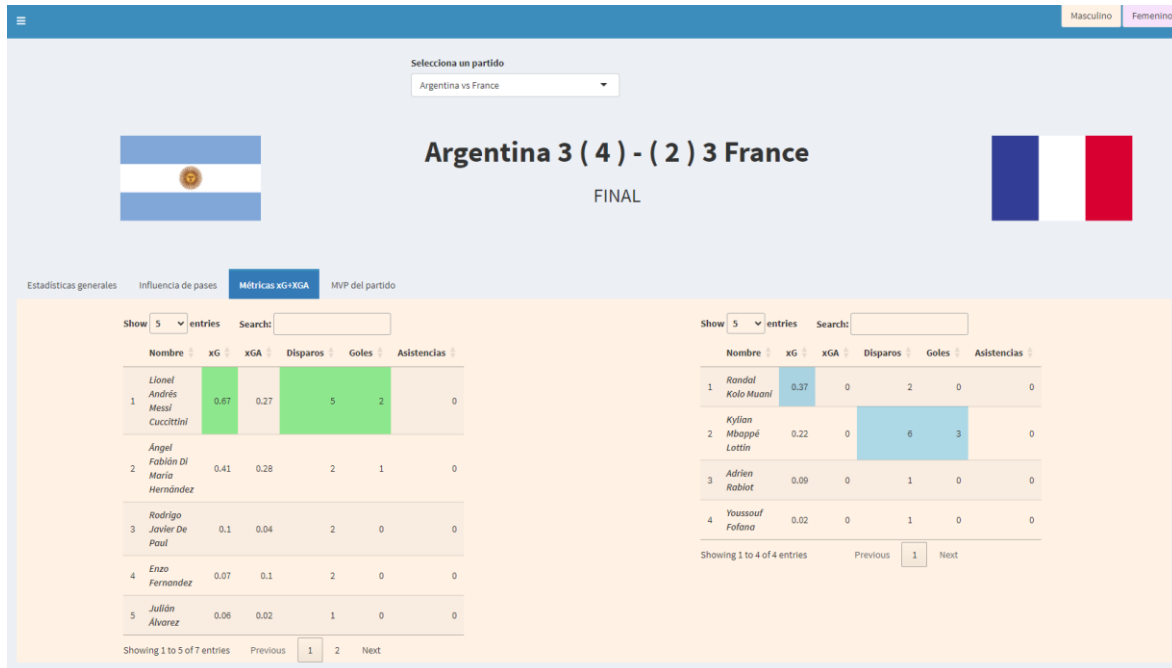


Figura 29. Página Partidos, pestaña “Métricas xG+xGA”

- Métricas xG+xGA:** En esta tercera pestaña, como se muestra en la Figura 29, se presentan métricas xG (Goles esperados) y xGA (Asistencias esperadas), además de otras estadísticas mediante dos tablas interactivas que incluyen información sobre los jugadores más relevantes ofensivamente de cada selección en el partido. En ambas tablas los valores máximos de cada columna tienen un fondo de color para distinguir los jugadores más relevantes en cada acción.

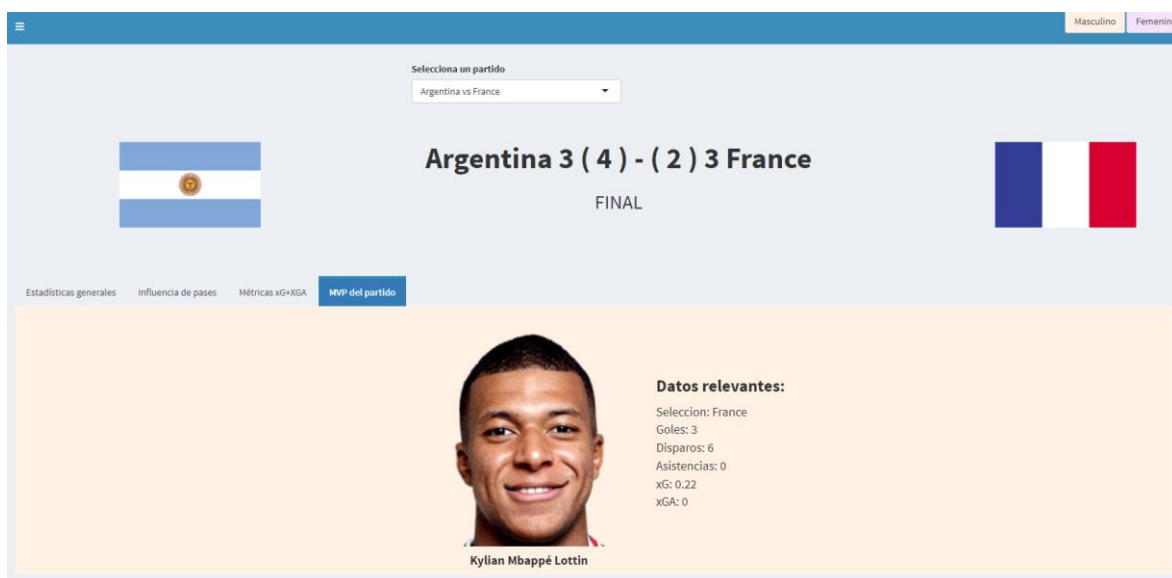


Figura 30. Página Partidos, pestaña “MVP del partido”

- **MVP del partido:** En esta última pestaña, véase Figura 30, se destaca al jugador que ha sobresalido en el partido y ha contribuido significativamente al éxito de su selección. Junto a su imagen, se presentan también sus estadísticas en relación con las acciones ofensivas más relevantes.

Página Análisis defensivo

Esta página, compuesta de tres pestañas, presenta estadísticas detalladas sobre las diversas acciones defensivas realizadas por el equipo a lo largo del torneo. Se solicita al usuario que seleccione un equipo específico para analizar sus acciones defensivas en las diferentes pestañas.



Figura 31. Página Análisis defensivo, pestaña “Acciones defensivas”

- **Acciones defensivas:** En esta primera pestaña, véase Figura 31, se le pide al usuario elegir un partido de entre todos los que ha disputado la selección escogida en el paso anterior. Una vez hecho esto, se exhibe:
 - Mapa con la localización de cada acción defensiva realizada por la selección elegida. Utilizando diferentes colores y formas, se distingue claramente cada tipo de acción.
 - Gráfico de barras donde se visualiza el número total de las acciones defensivas mencionadas anteriormente.



Figura 32. Página análisis defensivo, pestaña “Estadísticas defensivas globales”

Esta pestaña, véase Figura 32, se centra en mostrar al usuario diferentes tipos de tiros y goles recibidos de la selección elegida. Concretamente:

- Mapa de localización de tiros a puerta, tiros fuera y goles recibidos.
- Gráfico de sectores que indica los porcentajes de cada tipo de tiro recibido.
- Información numérica de disparos y goles recibidos por partido.



Figura 33. Página Análisis defensivo, pestaña “Faltas y tarjetas”

- **Faltas y tarjetas:** La última pestaña de esta página, véase Figura 33, muestra datos relativos a faltas y tarjetas de la selección elegida. Concretamente:
 - Tabla interactiva de faltas, tarjetas amarillas y rojas de cada jugador.
 - Información numérica de faltas y tarjetas recibidas por partido.
 - Imágenes de los jugadores que realizan más faltas y han recibido más tarjetas en la competición.

Página Estadísticas individuales

Compuesta de 3 pestañas en esta página se realiza un análisis más exhaustivo de cada jugador del torneo, examinando aspectos clave como la efectividad de los pases, las faltas, las asistencias y las recuperaciones, entre otros.

Al igual que en las páginas anteriores, el usuario selecciona el equipo y el jugador que desea analizar.

Una vez seleccionado el jugador, se presenta primero la imagen de este. A la derecha de esta, se indica el pie dominante con el cual el jugador se siente más cómodo, así como la posición en el campo que el jugador frecuenta con mayor asiduidad.



Figura 34. Página Estadísticas individuales, pestaña “Localización y pases”

- **Localización y pases:** En esta primera pestaña, véase la Figura 34, se presentan dos mapas del campo de fútbol. El primero muestra un mapa de calor que indica con una tonalidad de colores las posiciones más frecuentadas por el jugador, mientras que el mapa situado a la derecha refleja la calidad de cada pase realizado por el jugador, su localización y dirección.

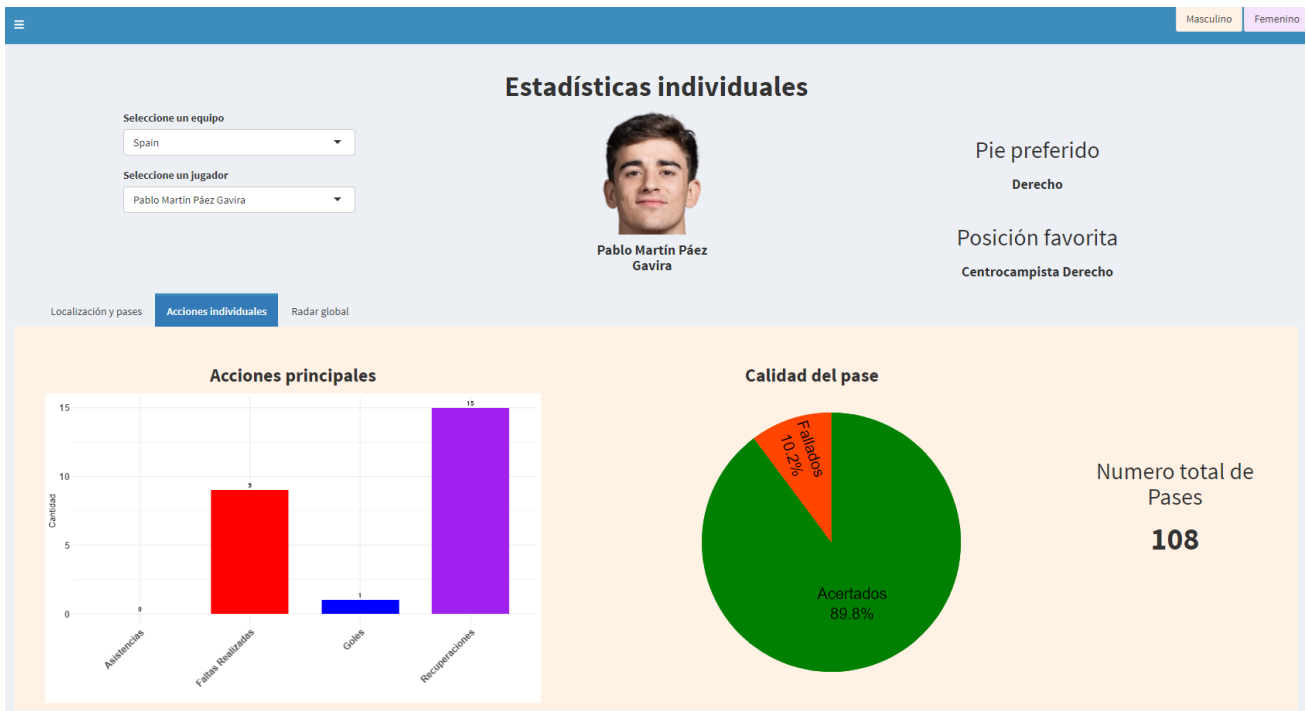


Figura 35. Página Estadísticas individuales, pestaña “Acciones individuales”

- **Acciones individuales:** En la segunda pestaña, véase la Figura 35, se muestran estadísticas relativas a diferentes acciones realizadas por el jugador elegido, concretamente:
 - Gráfico de barras indicando las veces que ha realizado asistencias, faltas, goles y recuperaciones en el torneo.
 - Gráfico de sectores que indica la calidad de los pases realizados por el jugador elegido.
 - Información numérica acerca del total de pases realizados por el jugador elegido

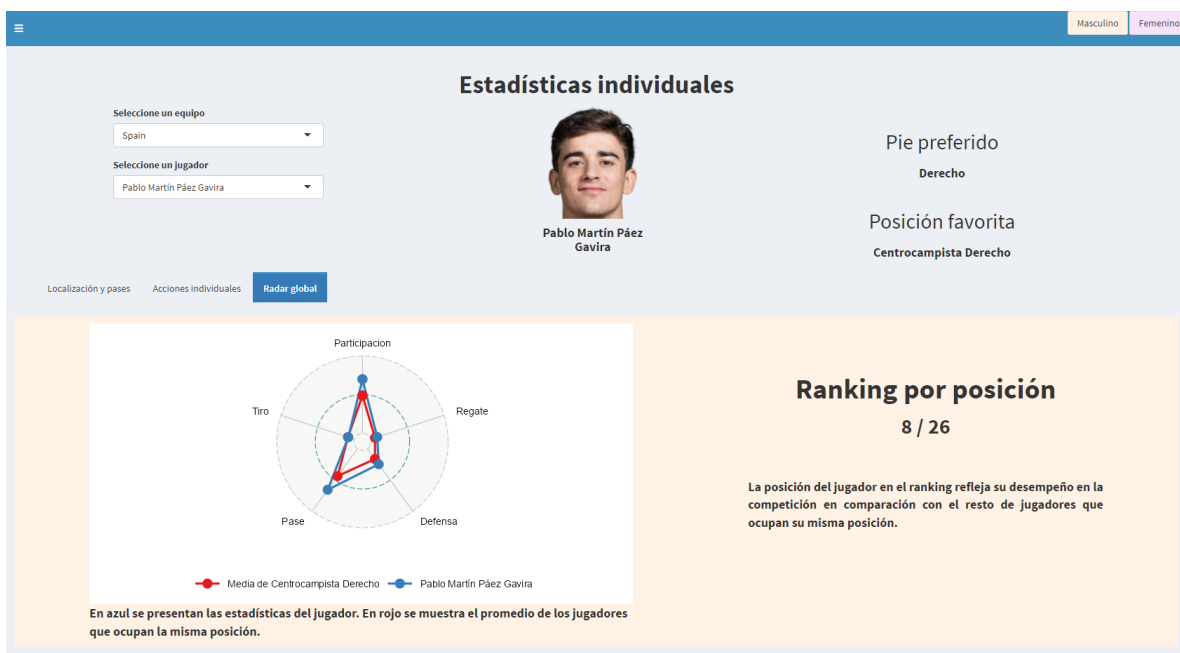


Figura 36. Página Estadísticas individuales, pestaña “Radar Global”

- **Radar global:** En la última pestaña de esta página, véase la Figura 36, se muestra un radar que analiza el rendimiento del jugador en comparación con el resto de los jugadores del torneo.

A la derecha de este radar, se presenta un ranking que mide, en base a las variables analizadas en el radar, el desempeño y clasificación del jugador respecto a los demás jugadores que ocupan su misma posición.

Creación radar y ranking por jugador

En este punto se explica la creación del radar y ranking anteriormente mencionados. [11]

Radar global

Para llevar a cabo la creación de este gráfico, se distingue si el jugador seleccionado es un jugador de campo o un portero, ya que dependiendo de su posición se despliegan diferentes estadísticas. Para los jugadores de campo, se consideran las siguientes variables:

- Participación
- Regate
- Defensa
- Pase
- Tiro

Si el jugador elegido es portero se consideran las siguientes variables:

- Participación
- Pase
- Paradas Exitosas
- Goles Concedidos
- Defensa

Para la elaboración de todas estas variables, se tiene en cuenta los minutos jugados por cada jugador a lo largo de la competición. Esto asegura que la evaluación sea proporcional y justa para todos los jugadores.

A continuación, se detalla la forma de construir la puntuación de las diferentes acciones de cada jugador:

- Variable **Participación:** Se calcula el porcentaje de los minutos totales jugados por el jugador respecto al total de los minutos disputados por su selección.
- Variable **Pase:** Se estima el porcentaje de pases correctos frente al total de pases realizados.

- Variable **Tiro**: Se suman los tiros totales y tiros a puerta. Posteriormente se asigna un valor de 0.5 a los tiros que van fuera y un valor de 1 a los que van a puerta, para que así los tiros a puerta, con una mayor posibilidad de marcar gol, tengan más valor.
- Variable **Regate**: Se suman los regates totales y regates acertados. Los regates realizados de forma errónea reciben un valor de 0.3, mientras que los regates acertados reciben un valor de 1.
- Variable **Defensa**: Para calcular este valor, se suman las acciones defensivas y luego se resta 0.4 por cada tarjeta amarilla y 0.8 por cada tarjeta roja.
- Variable **Paradas Exitosas**: Suma de todas las paradas realizadas por el portero
- Variable **Goles concedidos**: Número total de goles encajados por el portero en la competición.

En las variables Tiro, Regate y Defensa se asigna un valor al resultado de cada acción. Cabe destacar que es una elección arbitraria todos los valores elegido para así poder generar el mejor radar posible de acuerdo con el rendimiento de cada jugador



Figura 37. Ejemplo de radar de jugador y de portero

Una vez se disponen todos estos datos en un archivo Excel, se procede a normalizar los datos por columnas para facilitar un análisis más preciso.

Posteriormente, en el radar, véase la Figura 37, además de mostrar el valor del jugador seleccionado en relación con el total de jugadores, se presenta también la media de todas las estadísticas de los jugadores que ocupan la misma posición que el jugador elegido, lo que permite una comparación más clara y efectiva.

Ranking por posición

Este ranking está estrechamente vinculado con el proceso previo de creación de radares. En este punto, se establece un ranking diferente para todas las posiciones de campo existentes. Es importante recordar que un jugador puede haber jugado en diferentes posiciones, pero se le asigna la posición en la que ha jugado con mayor frecuencia.

A partir del paso anterior, tras la normalización de todas las variables mencionadas, para la elaboración del ranking de cualquier tipo de posición, excepto la de portero, se suman todas las variables, dado que cuanto mayor es el valor de la variable, mejor ha sido el desempeño del jugador. Posteriormente, se

ordenan los nuevos valores y se obtiene el ranking de todos los jugadores en función de su posición más frecuente.

El ranking de porteros presenta una ligera variación en comparación con el de los jugadores de campo, ya que se suman todas las variables que aparecen en su radar, excepto la variable "Goles concedidos". Esta variable, al reflejar un desempeño negativo del portero, se resta.

Página Balance final

En esta página, véase Figura 38 se pueden observar diferentes imágenes sobre los jugadores/selecciones más relevantes del torneo que han logrado cualquier tipo de récord.



Figura 38. Página Balance final

Página de Ayuda

Esta sección final proporciona al usuario un recurso valioso para consultar vocabulario técnico o recibir asistencia relacionada con las páginas analizadas anteriormente. Asimismo, se informa al usuario sobre el autor de la aplicación y se expresan agradecimientos a las fuentes que han contribuido al desarrollo de esta herramienta.



Figura 39. Página Ayuda

En esta página, véase Figura 39, se proporciona asistencia al usuario sobre los gráficos, tablas y diversa información presente en la aplicación. Además, se incluyen imágenes de la fuente que proporcionó las bases de datos gratuitas utilizadas en la elaboración de la aplicación. Se menciona también el paquete shiny y una fuente externa que facilita el despliegue de diversas imágenes en la aplicación.

5.3. Imágenes y menciones externas

Las bases de datos de la aplicación WorldCupViz provienen de la librería StatsBombR, que periódicamente publica datos sobre competiciones de fútbol. En esta aplicación, se muestran numerosas imágenes, las cuales se obtienen de diversas direcciones web, entre ellas:

- Imágenes de resultados globales: Diario Sport <https://www.sport.es/>
- Imágenes de jugadores: SoccerWiki <https://es.soccerwiki.org/>
- Imágenes de jugadoras: Fbref <https://fbref.com>
- Imágenes de banderas de las diferentes selecciones: Diario Marca <https://www.marca.com/>

Capítulo 6

Conclusiones

Para finalizar este trabajo, se proporciona un resumen de los objetivos planteados al inicio y cómo se han alcanzado tras la finalización de este proyecto.

Desde la carga de ambos conjuntos de datos mediante la librería **StatsBombR**, he logrado cumplir con los objetivos establecidos en el punto 1.1 de la memoria. Estos incluyen el análisis exploratorio de los datos, la agrupación de selecciones en diferentes grupos utilizando análisis clúster, y, fundamentalmente, la creación y diseño de la aplicación shiny **WorldCupViz**.

6.1. Complejidad y dificultades del trabajo

El desarrollo del trabajo no ha sido fácil de llevar a cabo, pero ha resultado ser una experiencia sumamente gratificante desde el punto de vista del aprendizaje. He encontrado diversas dificultades a lo largo del proceso, entre ellas:

- **Comprensión de los conjuntos de datos:** Al trabajar por primera vez con datos de la librería **StatsBombR**, he invertido tiempo considerable en analizar y comprender estos conjuntos. Este proceso ha sido crucial para identificar las variables más relevantes y optimizar mi trabajo.
- **Revisión y lectura de diferente documentación:** La búsqueda de paquetes adecuados para la creación de gráficos ha sido un proceso extenso. No todos los paquetes disponibles en R son directamente compatibles con los datos de **StatsBombR**, lo que ha implicado en ocasiones la necesidad de redefinir funciones para asegurar la correcta visualización de los gráficos.
- **Familiarización de la herramienta shiny y creación de la aplicación:** La creación de la aplicación shiny ha representado el mayor desafío de este trabajo. He tenido que dedicar tiempo significativo a familiarizarme con esta herramienta, aprovechando recursos como la bibliografía y videos documentales para avanzar de manera efectiva en su desarrollo.

6.2. Conocimientos aplicados

Durante mi trayectoria en el Grado en Estadística, he aplicado muchos de los contenidos aprendidos en este trabajo, destacando principalmente:

- **Estadística Descriptiva:** He utilizado mi conocimiento para interpretar diversos estadísticos y gráficos, fundamentales para comprender y presentar los datos de manera efectiva.
- **Fundamentos de programación:** A través de esta asignatura, he adquirido habilidades en la creación de bucles básicos y otras estructuras fundamentales en programación, lo cual ha sido útil para manipular datos y automatizar procesos en este proyecto.
- **Computación Estadística:** Mi formación en el lenguaje de programación R ha sido fundamental. He aplicado este conocimiento extensivamente para analizar conjuntos de datos y desarrollar funcionalidades específicas dentro de la aplicación shiny.

- **Análisis multivariante y análisis de datos:** Estas asignaturas han sido cruciales para realizar un análisis exploratorio profundo de los datos, aplicar técnicas como el análisis clúster y los componentes principales, lo que ha enriquecido el análisis y la presentación de resultados en mi trabajo.

Estas asignaturas, junto con otras de Informática y Estadística, han sido pilares fundamentales en mi formación, proporcionándome el conocimiento necesario para desarrollar con éxito este proyecto y la aplicación shiny asociada.

6.3. Trabajo a futuro

Este trabajo presenta numerosas ampliaciones y actualizaciones potenciales que pueden desarrollarse en el futuro:

La aplicación recién creada, **WorldCupViz**, como se mencionó en el capítulo 5 de esta memoria, se enfoca en el análisis de conjuntos de datos de fútbol de las pasadas *Copas Mundiales de la FIFA*, tanto masculinas como femeninas. Los archivos R desarrollados están configurados para analizar cualquier base de datos de la librería StatsBombR y desplegarla en la aplicación.

En adelante, las próximas bases de datos publicadas por StatsBomb, como por ejemplo la Eurocopa 2024 disputada por diversas selecciones europeas en verano, podrán ser analizadas simplemente cargando estos datos en RStudio y realizando ajustes menores en el código, principalmente relacionados con las imágenes de los jugadores.

Con diferentes bases de datos disponibles, se pueden llevar a cabo análisis y actualizaciones adicionales en la aplicación WorldCupViz, tales como:

- **Creación de nuevos radares:** Incluir radares adicionales para analizar el rendimiento de las selecciones en la competición utilizando diversas variables.
- **Desarrollo de nuevas páginas:** Implementar nuevas páginas que analicen cómo la disposición de los jugadores en el campo afecta el resultado del partido.
- **Creación de un paquete R:** Utilizar el código desarrollado en este trabajo para documentar, organizar y crear un paquete de R que contenga funciones nuevas y mejoradas.

Además, tras el análisis clúster realizado para identificar similitudes entre diferentes selecciones femeninas en la copa mundial 2023, se pueden realizar análisis clúster más detallados, no solo de tipo jerárquico, para identificar similitudes y diferencias entre jugadores que ocupan posiciones similares.

Estas ampliaciones y actualizaciones representan oportunidades significativas para expandir y mejorar la aplicación **WorldCupViz**, proporcionando análisis más profundos y completos del rendimiento en competiciones futuras.

Bibliografía

- [1] Wikipedia, "Historia de la Copa Mundial de Fútbol",
https://es.wikipedia.org/wiki/Historia_de_la_Copa_Mundial_de_F%C3%BAbol.
- [2] Goal, "Datos relevantes mundiales masculinos",
<https://www.goal.com/es/noticias/todos-los-mundiales-de-la-historia-de-1930-a-2026/18jq0g6f7vdp61a1vya8e7uivc>.
- [3] Wikipedia, "Historia de las copas mundiales femeninas",
https://es.wikipedia.org/wiki/Copa_Mundial_Femenina_de_F%C3%BAbol.
- [4] N. Geographic, "Curiosidades copas del mundo femeninas",
<https://www.ngenespanol.com/el-mundo/historia-de-la-copa-mundial-femenina-de-la-fifa/>.
- [5] StatsBomb, "Información paquete StatsBombR",
<https://statsbomb.com/es/articulos/futbol/introduccion-al-uso-de-statsbomb-data-con-r/>.
- [6] M. A. J. Cuadrillero, "Clustering Jerárquico en R",
<https://rpubs.com/mjimcua/clustering-jerarquico-en-r>.
- [7] RCharts, "Dendogramas en R",
<https://r-charts.com/es/parte-todo/hclust/>.
- [8] Shiny, "Guía introductoria al paquete Shiny R",
<https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/index.html>.
- [9] E. S. M. Martín, "Interactividad Shiny",
<https://www.r-users.gal/sites/default/files/shiny.pdf>.
- [10] H. Zhu, "Paquete kableExtra",
<https://cran.r-project.org/web/packages/kableExtra/kableExtra.pdf>.
- [11] RCharts, "Creación de radares",
<https://r-charts.com/es/ranking/radar-chart/>.
- [12] Wickham H, Vaughan D, Girlich M (2024). `_tidyr: Tidy Messy Data_`. R package version 1.3.1.,
<<https://CRAN.R-project.org/package=tidyr>>.
- [13] Wickham H, Hester J, Chang W, Bryan J (2022). `_devtools: Tools to Make Developing R Packages Easier_`. R,
<<https://CRAN.R-project.org/package=devtools>>.
- [14] H. Wickham. `ggplot2: Elegant Graphics for Data Analysis`. Springer-Verlag New York, 2016,
<https://ggplot2.tidyverse.org>.
- [15] Yam D (2024). `_StatsBombR: Cleans and pulls StatsBomb data from the API_`,
<https://github.com/statsbomb/StatsBombR>.
- [16] C. Sievert. `Interactive Web-Based Data Visualization with R, plotly, and shiny`. Chapman and Hall/CRC,
<https://plotly-r.com>.

-
- [17] Pedersen T, Robinson D (2024). *_gganimate: A Grammar of Animated Graphics_*,
<https://CRAN.R-project.org/package=gganimate>.
- [18] Xie Y, Cheng J, Tan X (2024). *_DT: A Wrapper of the JavaScript Library 'DataTables'_*,
<https://CRAN.R-project.org/package=DT>.
- [19] Cheng J, Sievert C, Schloerke B, Chang W, Xie Y, Allen J (2024). *_htmltools: Tools for HTML_*,
<https://CRAN.R-project.org/package=htmltools>.
- [20] Ooms J (2014). “The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R,”
<https://arxiv.org/abs/1403.2805>.
- [21] Zhu H (2024). *_kableExtra: Construct Complex Table with 'kable' and Pipe Syntax_*,
<https://CRAN.R-project.org/package=kableExtra>.
- [22] Gallagher J (2024). *_soccermatics: Visualise football (soccer) tracking and event data_*,
<https://github.com/jogall/soccermatics>.
- [23] Nakazawa M (2024). *_fmsb: Functions for Medical Statistics Book with some Demographic Data_*,
<https://CRAN.R-project.org/package=fmsb>.
- [24] rStats F (2024). *_SBpitch: Create a pitch plot ready for Statsbomb data_*,
<https://github.com/FCrSTATS/SBpitch>.
- [25] Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B,
<https://CRAN.R-project.org/package=shiny>.
- [26] C. M. a. J. W. (2007) Guojun Gan, *Data Clustering: Theory, Algorithms, and Applications*, SIAM.