

Análisis de datos usando gráficos estadísticos desde un enfoque exploratorio

Maileen Julissa Hoyos Castellanos

Trabajo de grado presentado como requisito para optar al título de Licenciada en Matemáticas

Directora

Tulia Esther Rivera Flórez

Magister en Estadística

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemáticas

Licenciatura en matemáticas

Bucaramanga

2024

Dedicatoria

A todos aquellos que sembraron en mí su enseñanza con paciencia y amor.

Por ser los arquitectos de quien soy como persona y como profesional.

Agradecimientos

A mis padres y abuelos, pilar en el que me sostuve desde las primeras letras que tracé en papel, hasta el último examen que selló mi formación. Su apoyo incondicional, sacrificios y fe inquebrantable me han impulsado en cada desafío.

Al amor de mi vida, Miguel, porque has sido refugio en las tormentas y alegría en los días soleados. A tu lado, las horas de estudio se tornaron más ligeras y los teoremas y demostraciones, por muy complejos que fueran, se hicieron más llevaderos. Gracias por sumarte a mis desvelos, por ser mi cómplice en cada risa y cada lágrima. Tu amor y paciencia han sido la fuerza que me ha impulsado a seguir adelante.

A mis profesores, *Tulia y Alexander*, sus palabras de aliento llegaron en los momentos en que más lo necesitaba, siendo como un faro en mis noches más oscuras. A través de su sabiduría y empatía, me enseñaron no solo lecciones académicas, sino también el valor de la persistencia y la pasión. Sin su guía, este sueño quizás hubiera quedado en el camino.

A *Tatiana*, cuya aparición oportuna en mi vida me permitió ser. Por enseñarme el verdadero significado de la amistad, haciendo más llevaderos mis días con su apoyo constante y su generosidad aliviando el peso, dándome el aliento necesario para continuar.

A todos ustedes, mi eterna gratitud. Gracias por creer en mí, incluso cuando yo misma dudaba. Sin ustedes, este logro no tendría significado.

Tabla de Contenido

	Pág.
Introducción	11
1. Descripción del problema.....	13
1.1 Lineamientos orientadores para la formación en Estadística.....	13
1.2 Aprendizaje activo y nuevas tecnologías.....	15
1.3 Limitaciones con el uso de recursos disponibles en la web.....	17
1.4 El papel de los libros de texto.....	19
2. Objetivos	21
2.1 Objetivo General.....	21
2.2 Objetivos Específicos.....	21
3. Justificación.....	22
4. Antecedentes.....	25
5. Marco teórico y conceptual.....	30
5.1 Elementos de diseño gráfico.....	30
5.2 Diseño editorial.....	31
5.3 Elementos básicos.....	31
5.3.1 Diseño interactivo.....	32
5.4 Herramientas para la programación de un libro interactivo.....	33
5.5 Elementos de didáctica de la estadística.....	34
5.5.1 Aprendizaje activo.....	35
5.5.2 Enseñanza basada en proyectos.....	36
5.5.3 Análisis exploratorio de datos.....	36

6. Metodología	37
7. Resultados	38
7.1. Estructura del capítulo	41
7.1.1 Introducción	42
7.1.2 Validando un supuesto distribucional	43
7.1.3. Pruebas gráficas	45
7.1.3.1. Diagrama de caja y bigotes	45
7.1.3.2. Histograma.	47
7.1.3.3. PP y QQ plot.	49
7.1.3.4. Ejercicios propuestos.	51
7.1.3.5. Comunicando resultados.	53
7.2. Pruebas de Usuario	54
7.2.1. Procedimiento de las pruebas.....	54
8. Conclusiones	56
Referencias bibliográficas	59
Apendices	62

Listado de Figuras

<i>Figura 1 Primera sección: Introducción</i>	43
<i>Figura 2 Sección 1: Cómo validar un supuesto distribucional</i>	44
<i>Figura 3 Sección 2, pruebas gráficas. Subsección, diagrama de caja y bigotes. Mapa de ruta y señales de kilometraje</i>	45
<i>Figura 4 Sección 2, pruebas gráficas. Subsección, diagrama de caja y bigotes. Boxplot interactivo población por departamento.</i>	46
<i>Figura 5 Sección 2, pruebas gráficas. Subsección, diagrama de caja y bigotes. Gráfico de caja y bigotes</i>	47
<i>Figura 6 Sección 2, pruebas gráficas. Subsección, Histograma. Algoritmo para la construcción de un histograma</i>	47
<i>Figura 7 Sección 2, pruebas gráficas. Subsección, Histograma. ¿Cómo interpretar un histograma?</i>	48
<i>Figura 8 Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. Comparación de cuartiles teóricos y muestrales</i>	49
<i>Figura 9 Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. QQ plot interactivo</i> ...	50
<i>Figura 10 Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. Comparación curva normal y Dotplot</i>	51
<i>Figura 11 Sección 2, pruebas gráficas. Subsección, Comunicando Resultados. Video tutoriales para realizar Dashboard en Excel</i>	53
<i>Figura 12 Comparación de interfaz gráfica antes y después de pruebas de usuario.</i>	54
<i>Figura 13 Pruebas de usuario, segunda sesión</i>	55

Lista de Tablas

	Pág.
<i>Tabla 1. Etapas que incluyó el diseño y desarrollo del producto final</i>	<i>37</i>

Lista de Apéndices

Apendice a Comentarios/Sugerencias estudiantes 62

Apendice b Código fuente 68

Resumen

Título: Análisis de datos usando gráficos estadísticos desde un enfoque exploratorio *

Autor: Maileen Julissa Hoyos Castellanos**

Palabras Clave: Libro interactivo, enfoque exploratorio, R Studio, Shiny, Estadística.

Descripción:

Esta tesis abarca el diseño y programación de un capítulo de libro en formato digital, orientado a la enseñanza de la estadística con un enfoque especial en el análisis de gráficos estadísticos desde una perspectiva exploratoria. A través de una revisión bibliográfica, se identificó una tendencia en la educación estadística hacia métodos de enseñanza tradicionales y la falta de recursos didácticos enfocados en el autoaprendizaje y la interpretación activa de datos. Ante esta observación, la tesis propone una innovación en la educación estadística mediante el desarrollo de material didáctico interactivo que centra la experiencia de aprendizaje en el estudiante.

El capítulo combina teoría y ejercicios prácticos para ofrecer una experiencia de aprendizaje enriquecedora. La programación en R y la utilización de R Shiny, que es conocida por su capacidad de crear aplicaciones interactivas, permite que el contenido se presente de una manera dinámica y atractiva. Los usuarios tienen la oportunidad de interactuar con gráficas y datos en tiempo real, lo que facilita una comprensión más profunda de los conceptos estadísticos y su aplicación en escenarios del mundo real. Este enfoque práctico y visual, potenciado por las capacidades interactivas de R y R Shiny, hace que el aprendizaje sea más atractivo y significativo contribuyendo a la educación estadística, mostrando cómo estos métodos centrados en el estudiante pueden enriquecer el proceso de aprendizaje y aumentar el interés en el análisis de datos.

* Trabajo de Grado

** Facultad de Ciencias Escuela de Matemáticas. Licenciatura en Matemáticas. Director: MSc. Tulia Esther Rivera Flórez.

Abstract

Title: An Exploratory Approach to Statistical Graph Analysis *

Author(s): Maileen Julissa Hoyos Castellanos **

Key Words: Interactive Book, Exploratory Approach, RStudio, Shiny, Statistics

Description:

This thesis focuses on the design and programming of a digital book chapter dedicated to teaching statistics, with a special emphasis on the statistical graph analysis from an exploratory perspective. A literature review identified a trend in statistical education towards traditional methods and a lack of educational resources focused on self-learning and active data interpretation. Considering this observation, the thesis proposes an innovation in statistical education through the development of interactive educational material that centers the learning experience on the student.

The chapter combines theory and practical exercises to provide an enriching learning experience. Thanks to programming in R and the use of R Shiny, known for its ability to create interactive applications, the content is presented dynamically and attractively. Users can interact with graphs and data in real time, facilitating a deeper understanding of statistical concepts and their application in real-world scenarios. This practical and visual approach, enhanced by the interactive capabilities of R and R Shiny, makes learning more engaging and meaningful, contributing to statistical education by showing how these student-centered methods can enrich the learning process and increase interest in data analysis.

* Bachelor Thesis

**Science Faculty. Mathematics School. Director: MSc. Tulia Esther Rivera Flórez.

Introducción

En la última década, la enseñanza de la Estadística se ha constituido como un tema de gran importancia en el sistema educativo gracias a la demanda de profesionales con formación en esta disciplina. Desafortunadamente su enseñanza sigue estando dominada por un enfoque tradicional, principalmente por la falta de cualificación de los docentes que asumen esta responsabilidad. Este no es el escenario que esperamos observar como ciudadanos de este siglo dado el marco conceptual en Didáctica de la Estadística, la variedad de recursos tecnológicos y de softwares especializados existentes.

Otra de las acciones que se requiere en nuestro país es la revisión de los propósitos de los cursos de Estadística en el nivel superior similar al proceso vivido en la educación básica en la que desde 1998 se cuenta con directrices a nivel nacional como son los Lineamientos y Estándares curriculares que lentamente han logrado posicionar la Estadística y Probabilidad dentro de la matemática escolar y orientar su formación en términos de competencias asociadas a consolidar el pensamiento aleatorio. Contrario a esto, en niveles superiores, cada Institución educativa tiene libertad de hacer un diseño propio y en la mayoría de programas hay un único curso en los planes de estudio que se orienta a revisar elementos básicos en estadística descriptiva e inferencial porque los prerrequisitos con los que llegan los estudiantes a la Universidad en la mayoría de los casos son bajos o ninguno.

En cuanto a recursos didácticos, el libro de texto solía tener un uso preponderante en el aula superado sólo por el tablero. Hoy día, el acceso a internet ha posicionado otras opciones como páginas web, canales de YouTube y applets. Sin embargo, los recursos tecnológicos que nos ha

traído la modernidad nos hacen pensar que es necesario analizar cómo se puede mejorar su uso para que realmente sean una herramienta cognitiva que potencie el aprendizaje del estudiante. Consideramos que es posible hacer trascender el papel del libro de texto en el aula y no limitarlo a ser un simple medio de transmisión de información. Los nuevos formatos pueden hacer posible más formas de interacción, la variedad de recursos que incluyen pueden orientarse a: promover un aprendizaje más autónomo, permitir la exploración y análisis de situaciones reales, incentivar la creatividad, potenciar el pensamiento crítico y dar respuesta a diferentes estilos de aprendizaje y necesidades de los estudiantes.

Este proyecto tuvo como principal motivación contribuir al mejoramiento del proceso de enseñanza-aprendizaje en cursos de formación en Estadística a nivel universitario a través de una mediación que se aparta del enfoque tradicional basado en la ejercitación de fórmulas al resolver misceláneas de ejercicios rutinarios. En particular queremos aportar a la comunidad educativa nacional el libro de texto “Análisis de datos desde un enfoque exploratorio”, en este proyecto se presenta un capítulo de éste cuyo tema central es el análisis de datos a través de gráficos estadísticos, para su diseño se combinaron tres elementos: resolución de problemas que preferiblemente incorporen datos del mundo real, énfasis en la exploración dinámica de datos y la incorporación de herramientas tecnológicas tanto para la edición del material digital como para las actividades de enseñanza propuestas.

Como objetivo secundario se consideró la creación de un recurso educativo de acceso libre, comprensible para todo tipo de público y que fomente el interés por la Estadística en los estudiantes al plasmar el trabajo real de un analista de datos moderno.

1. Descripción del problema

Desde el punto de vista de la enseñanza de la Estadística a nivel de Educación superior existen varios elementos del proceso formativo que se deben analizar en relación con la efectividad que pueden tener este tipo de cursos. A continuación presentamos un análisis para algunos de ellos que tomamos en consideración para plantear una alternativa de solución con nuestra propuesta.

1.1 Lineamientos orientadores para la formación en Estadística

Desde la Educación Matemática se promueve hoy día centrar la formación en el desarrollo de habilidades relacionadas con pensamiento matemático más que pensar en contenidos específicos e incluso en habilidades independientes. Para ello, cada país ha definido una estructura de formación alrededor de esta idea. En general, se establecen unos principios que orientan la formación en matemáticas con ideas curriculares actualizadas que incorporan el estado del arte de la investigación en Educación Matemática, unos procesos generales que enfatizan focos de intervención a nivel cognitivo y unas competencias claves asociadas a ellos que orientan tanto la selección de contenidos como los indicadores para evaluación.

A Nivel de Estados Unidos (EUA), desde 2007 se tienen los estándares National Council Teaching of Mathematics (NCTM) que proponen como procesos de pensamiento estándar en matemáticas : 1) Representación, 2) Razonamiento y prueba, 3) Comunicación, 4) Resolución de problemas, y 5) Conexiones. Este documento es el resultado del primer movimiento gestado en procura de formular estándares para la enseñanza de la matemática motivado en gran medida por el bajo desempeño en pruebas internacionales en un país donde la educación no está regulada por el gobierno central. Por su parte en Colombia, el Ministerio de Educación Nacional (MEN) determinó cinco dimensiones que integran el pensamiento matemático: Pensamiento espacial,

Pensamiento numérico, Pensamiento variacional, Pensamiento aleatorio y sistemas de datos y Pensamiento métrico, esto se hizo bajo la figura de dos directrices nacionales, inicialmente los Lineamientos curriculares promulgados en 1998 y luego en 2006 los Estándares básicos de competencia; estos documentos han logrado posicionarse como el marco legal para estructurar la formación en matemáticas en los niveles básicos, sin perder vigencia frente a la publicación de otro documento de referencia como son los Derechos Básicos de Aprendizaje (DBA) expedido en 2016.

Entre los aportes conseguidos con estas directrices reconocemos que además de reestructurar la formación en matemáticas proyectándola hacia el nuevo milenio, ellos han conseguido promover la enseñanza de áreas antes descuidadas como fue el caso de la geometría y en años recientes la Estadística.

Desafortunadamente para el nivel superior no se tiene una directriz tan específica que oriente la formación en áreas transversales como son el Lenguaje y Matemáticas. En Colombia se podrían citar como únicos elementos relacionados con la formulación de competencias a evaluar la prueba Saber Pro en su sección de razonamiento cuantitativo y de manera específica, el aporte de organizaciones como la Asociación Colombiana de Facultades de Ingeniería (ACOFI) que han implementado el Examen de Ciencias Básicas (EXIM) para estudiantes de Ingeniería de mitad de carrera, este examen abarca cinco dominios en Matemáticas: Variación y cambio, Medición, Convergencia, Estructuras y Aleatorio, sin embargo, la participación en este examen está a discreción de cada universidad así como la elección de la muestra de participantes lo cual afecta la representatividad y generalidad de los resultados.

Teniendo en cuenta lo anterior, en el contexto universitario colombiano las instituciones disfrutaban de libertad para diseñar sus currículos lo que permite incluir una identidad institucional

y mantener cierta flexibilidad; no obstante, la carencia de directrices ha llevado a un cierto descuido en la formación estadística. Muchos programas aún se basan en métodos tradicionales enfocados en el uso de fórmulas y la inclusión de temas obsoletos sin relevancia práctica actual. Como evidencias de esta situación podemos citar el cálculo manual de estadísticas descriptivas para datos agrupados, aproximación de la distribución binomial a través de una distribución normal, excesivo énfasis en la distribución normal y la estandarización de variables ignorando otros modelos probabilísticos, inferencia basada en muestras pequeñas y pruebas paramétricas.

De esta manera, es necesaria la actualización de los cursos de estadística en las universidades colombianas, considerando las necesidades y circunstancias actuales. Se debe incentivar la renovación de estos programas tradicionalmente enmarcados en enfoques anticuados. Hoy en día se cuenta con una amplia gama de tecnologías y software especializado que pueden ser excelentes recursos para la enseñanza. Además, los nuevos desafíos en la formación estadística apuntan hacia la analítica de datos, lo cual implica un mayor énfasis en la programación para analizar grandes bases de datos y la capacidad de integrar fuentes de información no estructuradas.

1.2 Aprendizaje activo y nuevas tecnologías

En la actualidad, se espera que los enfoques educativos promuevan un rol activo del estudiante en su proceso de aprendizaje, esto implica fomentar la participación y el compromiso del alumno en su proceso formativo de manera que pueda desarrollar las competencias y conocimientos esenciales para enfrentar, en primera instancia, los desafíos que le presenta su entorno cercano y a largo plazo los retos a nivel profesional. El promover un enfoque educativo centrado en el estudiante hace que los contenidos sean secundarios en una ruta de aprendizaje, la prioridad entonces es preparar a los jóvenes para interactuar con la realidad de su entorno al tiempo

que se forman para ser ciudadanos activos y críticos en un mundo cada vez más globalizado, complejo y cambiante.

Para complementar el enfoque propuesto en el párrafo anterior, los investigadores y docentes que se centran en el desarrollo de habilidades de pensamiento deben considerar otro elemento importante, de acuerdo con Garfield (2002) y Garfield et al. (2005), el razonamiento estadístico es fundamental para comprender ideas estadísticas y poder interpretar información estadística. Este tipo de razonamiento implica conectar conceptos y combinar ideas sobre datos, lo que incluye la capacidad de entender y explicar procesos estadísticos, así como interpretar los resultados de pruebas de significancia y de intervalos de confianza. Desarrollar habilidades de razonamiento estadístico debe ser uno de los objetivos principales de un curso de estadística.

A nivel nacional Behar y Ojeda (2013), reafirman el anterior planteamiento sugiriendo que el estudiante sea el protagonista del proceso educativo y en el caso específico de resultados de investigación en Enseñanza de la Estadística, una mediación efectiva en este sentido involucra el uso de recursos tecnológicos variados tanto en lo metodológico como en lo evaluativo, enfoque muy distante del tradicional que se ha otorgado a los cursos de Estadística a nivel universitario en nuestro país centrados en la ejercitación de fórmulas y algoritmos.

Particularmente, el uso de la simulación como herramienta didáctica en la clase de Estadística hoy día es imprescindible, a su favor no sólo podemos citar la posibilidad de que el estudiante vivencie un proceso de aprendizaje diferente centrado en la exploración, sino porque además permite mostrar la validez de los resultados teóricos de manera experimental sin requerir de tediosos desarrollos analíticos que demandan formación en temas de probabilidad y matemática, también se debe notar el amplio rango de experimentos aleatorios que permite abordar.

1.3 Limitaciones con el uso de recursos disponibles en la web

Ser ciudadano y profesional de este siglo demanda reconocer el impacto que tiene la tecnología en nuestra vidas. Particularmente en el sector educativo la transformación es necesaria tal como se viene promulgando por las asociaciones académicas, "La tecnología está cambiando la forma en que enseñamos y aprendemos matemáticas" (National Council of Teachers of Mathematics, 2016). No obstante, existe temor por abandonar ciertas condiciones de los entornos tradicionales aun cuando se abre un horizonte de posibilidades, tal como lo plantean Santos y Camacho (2018):

El uso de diversas tecnologías digitales puede ofrecer diversas oportunidades a los estudiantes en términos de analizar las formas de construir y explorar representaciones de los problemas, en los procesos de formulación de conjeturas y relaciones, en la búsqueda de argumentos para sustentar alguna conjetura y en las maneras de discutir y comunicar resultados. (p.24)

En la actualidad, podemos hablar de software especializado para todo campo del conocimiento, para el caso particular de enseñanza de la Estadística, además de software como Fathom, Tinker plots, Estat y CODAP, existen varios recursos en línea como calculadoras, libros de texto, escenarios con simulaciones, cursos interactivos, blogs, tutoriales interactivos, colecciones de bases de datos, recursos para calcular y graficar en línea, juegos, acceso a journals, páginas con ideas para proyectos o retos, etc.; entre estos, los más llamativos pueden ser los denominados applets, que son aplicaciones que de manera interactiva y visual presentan una idea estadística concreta. Hay colecciones de applets de amplia difusión y prestigio como son: Rossman/Chance, Shoder, Illuminations, Stat key, Cause web, Stat lets, Stat Crunch, Seeing Theory, entre otras.

No obstante, la utilidad de los applets como mediadores en el aula admite puntos de discusión. Por ejemplo, en la mayoría de los casos estos recursos son desarrollados de manera independiente, luego se liberan de forma gratuita sin pasar por una revisión de contenido o de diseño, de otro lado, no siempre es claro el enfoque metodológico al que responde y su potencial didáctico queda para ser descubierto por el profesor o usuario de este. Más aún, la tendencia actual, como se ha dicho anteriormente, es incentivar un proceso de formación más autónomo por parte del estudiante, pero ante lo señalado consideramos que se requiere una guía suficiente que permita captar el máximo provecho del recurso.

Otro problema con algunos recursos digitales de estadística es la informalidad con que se presentan conceptos estadísticos, muchos no utilizan la notación y el lenguaje propio de la disciplina, carecen de definiciones claras de los términos y conceptos utilizados. Es común que trabajen con muestras pequeñas de números generados aleatoriamente o colecciones de datos aportadas por los autores, los contextos utilizados no siempre son relevantes o bien documentados, además suelen incluir información excesiva o irrelevante en sus presentaciones lo que puede distraer al usuario. En general, el diseño de estos recursos no siempre facilita la comprensión de los conceptos ni su aplicación a situaciones reales.

Adicionalmente, las colecciones de applets que han sido desarrolladas por educadores estadísticos no siempre guardan un orden lógico en la presentación de conceptos o son suficientes para cubrir los propósitos de un curso introductorio de Estadística. En trabajos como el caso de Statkey, esta colección de applets acompaña al libro de texto “Statistics: unlocking the power of data” con lo cual los aplicativos guardando un orden temático. Finalmente indicar que los recursos mencionados no han sido diseñados para adaptarse y atender factores como inclusión y diversidad, aun cuando atender algunos niveles de discapacidad física en un formato digital podría hacerse de

manera eficiente sin requerir de un gran despliegue tecnológico como es la posibilidad de cambiar el tamaño de la letra o el acceso a la información en formato audiovisual.

1.4 El papel de los libros de texto

A nivel de Educación superior el libro de texto sigue siendo un recurso de apoyo central cuyo protagonismo es sólo superado a veces por el tablero o una presentación digital; es claro que la elección de este recurso influye directamente en las decisiones que toma el profesor en aspectos curriculares como son el enfoque metodológico, el alcance y propósitos del curso, la secuencia de contenidos, los ejemplos o aplicaciones, el software de apoyo y hasta el tipo de evaluación.

Entre los aspectos que limitaban el acceso de los estudiantes a libros de texto impresos o físicos en nuestro medio podemos citar el alto costo y la dificultad para su manipulación en el aula, para las nuevas generaciones éste dejó de ser un recurso motivador y está siendo superado por otros medios de consulta disponibles en la web gratuitos y más versátiles.

Por otra parte, la estadística es un campo disciplinar que toma todo el entramado conceptual y simbólico de la Matemática, pero le adiciona elementos propios. Muchas fuentes bibliográficas no acogen la notación estándar generando confusiones en los estudiantes cuando hacen uso de varias fuentes bibliográficas o recursos en línea. Así, por ejemplo, un mismo concepto pueden ser llamados de manera diferente en diferentes recursos, el estadístico "media" puede ser referenciado como "promedio", "media aritmética", "media muestral", "valor esperado" o "esperanza matemática". Además, las notaciones matemáticas pueden variar entre recursos lo que puede dificultar el seguimiento de las explicaciones y ejercicios. Por citar algún ejemplo, una variable aleatoria puede ser representada por "x" en un recurso y por "X" en otro. Cuando se promueve el estudio independiente, esta falta de estandarización puede generar problemas en los estudiantes

que están tratando de entender los conceptos y aplicaciones de la estadística en diferentes contextos.

Finalmente, los libros digitales adquieren el compromiso de proporcionar un entorno personalizable y adaptable a las necesidades y características de los alumnos (Solís et al, 2022) y de ser accesibles desde cualquier lugar vía internet, esto significa que los estudiantes pueden hacer uso de ellos desde cualquier dispositivo móvil o computadora (National Council of Teachers of Mathematics, 2016) y que una persona con algún tipo de discapacidad debería no sentirse excluido al ser usuario de un libro de texto.

2. Objetivos

2.1 Objetivo General

Diseñar un capítulo de libro de texto en formato digital para la enseñanza de Estadística cuyo tema central es la descripción de datos usando gráficas estadísticas.

2.2 Objetivos Específicos

Incorporar una mediación basada en resolución de problemas del mundo real, exploración dinámica de datos y una interacción basada en recursos tecnológicos variados para la enseñanza de gráficas estadísticas.

Diseñar un capítulo de libro haciendo uso de las herramientas Rmarkdown y Shiny que el entorno de R ofrece para la edición de textos de código abierto que permiten combinar texto, imágenes, comandos en R y aplicaciones.

3. Justificación

La enseñanza de la estadística es esencial en la formación de individuos como ciudadanos de la sociedad del conocimiento, aunado al desarrollo de las competencias propias que aporta la formación en estadística en cuanto a la solución de problemas a partir de datos, hoy día se adiciona el compromiso de formar pensamiento crítico y fundamentar los procesos de toma de decisiones informadas en contraposición de un pensamiento basado en posturas personales carentes de formalidad, por tanto cualquier esfuerzo por mejorar los procesos de enseñanza de esta disciplina es necesario y urgente.

La clase de estadística ha evolucionado positivamente en los últimos años incorporando el uso de software y el trabajo exploratorio a través de simulación, estos elementos han permitido a los estudiantes tener una experiencia más práctica y visual en el aprendizaje de los conceptos estadísticos. El uso de programas especializado como SAS, SPSS, Stata, Minitab, Matlab , R o Python han posibilitado que los estudiantes realicen sus propios análisis de datos y produzcan informes técnicos que ilustren sus hallazgos con gran versatilidad y eficiencia; además, la simulación como herramienta didáctica ha permitido que los estudiantes experimenten con distintas situaciones y vean cómo los resultados cambian al manipular diferentes variables, esto sin duda ha permitido avanzar en la comprensión de los elementos teóricos en estadística y probabilidad, al tiempo que se otorga un sentido práctico a las metodologías requeridas en un análisis de datos las cuales fácilmente se conectan con la solución a problemas del mundo real sin dejar de mencionar la posibilidades para reforzar la visualización y opciones para comunicar ideas.

No obstante, el papel preponderante de la tecnología y recursos web, el diseño y creación de un libro de texto que de soporte a un curso de Estadística es importante por varias razones. En primer lugar, permite a los estudiantes acceder a la información de manera organizada y eficiente,

hemos pasado de tener que llevar un libro físico a todas partes a la facilidad de acceder a un libro digital desde cualquier dispositivo con conexión a internet, poder estudiar en cualquier tiempo y lugar es una ventaja que debe contribuir a mejorar significativamente la capacidad de autoaprendizaje que bien manejada debería reflejarse en un mejor rendimiento académico.

En segundo lugar, un libro de estadística en formato digital puede ser más interactivo y dinámico que un resumen en el tablero o la consulta de un libro impreso debido a la variedad de recursos que puede incluir: videos, animaciones, herramientas para graficar y hacer cálculos online, ejercicios interactivos, prácticas evaluativas, acceso libre a colecciones de datos por mencionar las más conocidas, todos ellos pueden hacer un aporte muy importante en el desarrollo de las habilidades básicas del pensamiento estadístico. Desde la óptica de la enseñanza, Santos y Camacho (2018) mencionan que el uso de tecnologías digitales tiene un efecto en la comunicación entre los estudiantes, el maestro y los materiales didácticos, cambia el rol de los estudiantes brindándoles la oportunidad para que los estudiantes investiguen las propiedades y características de estos objetos, identificando aquellos aspectos que permanecen constantes y generando hipótesis sobre su comportamiento que fácilmente pueden ser validadas con datos.

De otro lado, a nivel actitudinal, Solís et al (2022) comentan que los materiales virtuales suponen una serie de ventajas respecto a los libros de texto en físico (o digital en PDF). Se resaltan aportes como el aumento en la motivación de los estudiantes, más ayudas para la comprensión de conceptos abstractos y favorecer la adquisición de la competencia digital. En general, se asegura un aporte en la componente emotiva ya que posibilita alcanzar logros académicos mejor y más rápido; de nuestra parte destacamos otros aportes positivos como la posibilidad de recibir retroalimentación individualizada, incentivar el pensamiento crítico y habilidades relacionadas con

la capacidad propositiva y argumentativa, así como poner en juego la creatividad para favorecer la visualización y presentación de resultados.

Es claro que el rol del profesor moderno también se está viendo afectado ya que debe pasar de ser un transmisor a ser un auténtico facilitador del talento individual, generando la necesidad de renovar enfoques metodológicos y ámbitos de evaluación y seguimiento. Adicionalmente, la modalidad de estudio empieza a diversificarse y combinarse. La no obligatoriedad de la presencialidad en los procesos formativos hace que cobren valor los recursos para la gestión de la información y de comunicación como son las plataformas, softwares y recursos web pero en todo caso, un libro de texto guía digital será un complemento ideal en un proceso de formación autónomo.

En cuarto lugar, un libro en formato digital es más fácil de actualizar y mejorar, a diferencia de un libro impreso, permitiendo ser actualizado en tiempo real para incluir información más reciente o para corregir errores. Esto asegura que los estudiantes estén recibiendo información precisa y actualizada lo que es esencial para un tema en constante evolución como la estadística y la innovación tecnológica.

También, un libro digital en general puede ser más accesible para estudiantes con discapacidades. Implementar funciones como audio y la opción de aumentar el tamaño de letra para asistir a quienes tienen discapacidades visuales es una acción sencilla, pero fundamental para promover la igualdad. Otras posibilidades incluyen la traducción a diversos idiomas, beneficiando a estudiantes no nativos del idioma original del libro, esto también facilita recoger retroalimentación de entornos educativos tanto nacionales como internacionales.

Con relación a los libros de estadística, Berihuete et al (2017) señalan que muchos libros de texto no incluyen contenido dinámico con el que el estudiante pueda interactuar, a pesar de que

se sabe que el material interactivo hace que los estudiantes profundicen más en los conceptos y se hagan preguntas del tipo "¿qué pasaría si...?", cuando interactúan con un elemento creado con *Shiny*, por ejemplo. Estas preguntas y la respuesta inmediata del material interactivo optimizan las clases prácticas de estadística, centrándose más en el entendimiento de los conceptos y su uso aplicado.

En síntesis, diseñar un libro virtual para enseñar estadística presenta una gran oportunidad para mejorar el aprendizaje de los estudiantes. El uso de tecnologías digitales permite una mayor interacción entre los materiales, los estudiantes y el maestro, genera oportunidades para investigar el comportamiento de los objetos o ideas en estudio, además permite a los estudiantes acceder a la información de manera más agradable y eficiente, hace que el aprendizaje sea más interesante y atractivo, facilita la actualización y mejora del contenido y hace el aprendizaje más accesible para estudiantes con discapacidades. Por lo tanto, diseñar libros de texto sigue siendo una opción valiosa que debe ser considerada en toda comunidad educativa.

4. Antecedentes

La evolución de los libros de texto educativo ha seguido un camino similar al de los libros en general. Sus inicios se ubican con la civilización egipcia, los libros eran manuscritos copiados a mano en los antiguos rollos de papiro o pergamino, y distribuidos en forma limitada, posteriormente el descubrimiento del papel dio más variedad; aparecieron luego las Universidades y con ellas importantes cambios en la edición de los libros principalmente de presentación como fueron la escritura gótica en columnas, organización por capítulos y las ilustraciones. En las últimas décadas, el desarrollo tecnológico permitió pasar del formato físico a la posibilidad de

almacenamiento en otros medios y dispositivos como el CD, y ahora la accesibilidad por diferentes dispositivos electrónicos a través del internet.

En la actualidad, los libros educativos interactivos digitales son una evolución más, se pasó del texto digital que sólo presentaba texto plano a incluir formatos dinámicos con recursos variados; como un importante referente que da cuenta de esta evolución es el Texto “*Teoría y problemas de Probabilidad y Estadística*” de la editorial Mac Graw Hill en su serie Schaum, este solía ser un libro de texto muy utilizado en las clases a nivel universitario en Colombia, su primera edición en inglés circuló en 1975. Explorar el contenido de este texto evidencia el enfoque característico de la época: como prerrequisito advertía la necesidad de tener formación en Cálculo, en cuanto a temáticas cubría sólo la parte de Probabilidad y Estadística, el capítulo de modelos de probabilidad el enfoque era claramente orientado a la ejercitación y cálculos, en lo concerniente a inferencia se entatiza en las distribuciones muestrales y teoría de estimación incluyendo ejercicios numéricos y demostraciones, no se incluía una revisión de estadística descriptiva sino que en la medida que se requerían se iban introduciendo las medidas básicas de resumen, como aporte al lector se publicitaba el ofrecer 760 problemas resueltos y una nutrida sección de ejercicios propuestos.

Posteriormente en 1991, la segunda edición anunciaba más énfasis en la comprensión en los métodos de resolución de problemas prácticos y cubrir aspectos teóricos esenciales de Estadística, se incluyen entonces cinco capítulos para tratar temas de estadística descriptiva, se reduce la profundidad en los temas probabilidad, se incluye las pruebas no paramétricas y un capítulo final sobre series de tiempo.

Para 2009 sale una cuarta edición, se ofrecían 486 problemas resueltos paso a paso, 660 problemas de práctica actualizados con relación a la primera edición, y se introducen salidas de

los softwares Excel, Minitab, SAS, SPSS y STATISTIX con sus respectivos comandos para que sean ejecutados por el lector, se introduce el valor p en lo relacionado a inferencia, agregan un capítulo de Control estadístico de procesos y elimina el de número índices. Finalmente en la sexta edición que salió en 2020 incluye como novedad vídeos en línea con las explicaciones de los problemas resueltos.

Otra referencia bibliográfica muy común en la enseñanza de la estadística a nivel universitario en Colombia es el texto de Mendenhall, Beaver y Beaver cuyo título es “Introducción a la probabilidad y estadística”, hoy día va por la décimo tercera edición; su diseño mantiene la idea de ofrecer un número variado y amplio de aplicaciones que suman 1300 nuevos problemas, estos son basados en datos reales extractados de periódicos o revistas científicas recientes y con diferentes niveles de dificultad; también los autores afirman que se han introducido como cambios significativos en cuanto a la simplificación y claridad del lenguaje aportando un estilo más legible y amigable sin sacrificar la integridad estadística de las ideas, se proponen entrenar el cerebro más que los procedimientos y explicar cómo describir un conjunto de datos reales, interpretar los resultados en términos prácticos, evaluar la validez de los supuestos y qué hacer en esos casos, después del décimo capítulo se puede alterar el orden de revisión de los capítulos.

Desde la edición anterior se introdujo el uso de valor p generados por computadora notando la diferencia entre significancia estadística y práctica, para los profesores se ofrece un sitio web con todos los ejercicios de las ediciones anteriores. Entre lo nuevo se destacan las secciones “*Mi entrenador personal*” que muestra el paso a paso en la solución de un problema al que le siguen una serie de ejercicios repetitivos que se resuelven similarmente, esto pretende emular la preparación que sigue un atleta, algunos ejercicios traen respuesta, esto para los primeros diez capítulos del libro porque más allá las soluciones se basan en el uso de software donde se

proponen incentivar un método al que llaman “*integral*” en el que se da una síntesis de los resultados de un análisis completo que incluye conclusiones y recomendaciones para el experimentador. También incluye el acceso a applets que muestran experimentos, estos permiten interacción con gráficas y tablas así como insertar datos. Al final del capítulo hay series de ejercicios que se resuelven con un mismo applet.

En cuanto al uso de tecnología escogieron MINITAB para las salidas requeridas, pero se aíslan de los comandos que son puestos en una sección especial para favorecer a quienes no les interesa este tema o usan otros programas. Finalmente, como material de apoyo para estudiar se resalta que cada ejercicio tiene un nombre para relacionarlo fácilmente con el contexto, ejercicios con espacios en blanco que pueden ser llenados para recibir retroalimentación inmediata, hay sugerencias prácticas en la sección Mi Consejo que aparece en el margen del texto, otras secciones son las de conceptos claves y fórmulas a manera de repaso al final de un capítulo. Adicional a los applets, en el sitio Companion Website se accede a diapositivas en Power Point para cada capítulo, bases de datos para los ejercicios en diferentes formatos, calculadora graficadora manual y los ejercicios de práctica con corrección, a todo lo anterior se suma una opción para el profesor que le permite el manejo de tareas en línea.

Un ejemplo de libro nacido totalmente digital es “*Unlocking the Power of Data by Lock, Lock, Lock, and Lock*” el cual cuenta con una página web de soporte que contiene la colección de applets y bases de datos que dan soporte a las actividades propuestas en el texto, incluye recursos para todos los temas de un primer curso de Estadística.

En general, los libros de estadística cubren conceptos estadísticos básicos como probabilidad, estadística descriptiva, distribuciones de probabilidad, variables aleatorias e inferencia estadística, incluyen las ideas básicas en regresión lineal y capítulos relacionados que

resaltan la importancia de estos temas para la práctica de la estadística. Sin embargo, también hay algunas diferencias notables; algunos libros se centran más en la aplicación de métodos estadísticos a campos específicos, como la ingeniería o las matemáticas, mientras que otros se centran en la aplicación de métodos estadísticos a la investigación científica general. Algunos libros incluyen capítulos adicionales sobre temas más avanzados como el análisis de datos de supervivencia, diseño experimental y regresión múltiple.

En particular el fondo bibliográfico de la UIS cuenta de una colección suficiente de títulos como textos de Estadística tanto en físico como en versión electrónica, los nombres corresponden a textos clásicos a nivel universitario como son Schaum, Walpole, Navidi, Evans Wonnacott, Freund, Montgomery, Neter, Daniel. Como novedad se puede resaltar que el acceso a plataformas como Connect de Mac Graw Hill, que consta de una plataforma donde el profesor crea el curso para que los estudiantes puedan acceder a usar los recursos tanto en clase como de manera independiente, ofrece recursos para la clase, exámenes, multimedia, reportes de progreso, grabación de clases, herramientas para calificación, desafortunadamente contar con demanda un costo semestral por curso donde se utilice dicho servicios.

Finalmente, elegir el libro de estadísticas correcto dependerá del campo de aplicación, el nivel de detalle y la profundidad de los conceptos que desea aprender. Algunos libros pueden ser más adecuados para principiantes, mientras que otros pueden ser más adecuados para estudiantes o investigadores avanzados. Existe gran cantidad de textos en el mercado que ofrecen un recorrido por los temas básicos, el formato va desde el texto plano con alusiones a tecnología hasta los libros interactivos cargados de actividades mediadas con tecnología como se indicó en los párrafos precedentes al anterior.

5. Marco teórico y conceptual

El tema de Estadística que trata esta propuesta son las gráficas estadísticas, no obstante, acorde a las características del proyecto para su desarrollo se requiere combinar principalmente conocimientos en tres áreas: Didáctica de la Estadística, Herramientas para diseño web disponibles en R y Elementos básicos en escritura y edición de libros. Por ello a continuación presentaremos las ideas básicas que soportarán el desarrollo de esta propuesta.

5.1 Elementos de diseño gráfico

Según Ghinaglia (2009) a la hora de producir material que incluye un componente gráfico se deben tener en cuenta los siguientes elementos:

- **Tipografía:** Se refiere a la selección y uso de fuentes o tipos de letra para transmitir un mensaje. Las fuentes pueden variar en tamaño, estilo, espaciado y color. Elegir la tipografía adecuada es muy importante para la legibilidad del texto y para proyectar el tono y la personalidad deseados.
- **Imagen:** Es un elemento visual utilizado en diseño gráfico para transmitir una idea o mensaje. Las imágenes pueden ser fotografías, ilustraciones, gráficos o cualquier otra forma de arte visual. Las imágenes pueden complementar o resaltar el contenido del diseño y pueden usarse de forma creativa para captar la atención del espectador.
- **Color:** Este es un elemento clave del diseño gráfico y se puede utilizar para transmitir emociones, crear una jerarquía visual, resaltar elementos importantes o simplemente captar la atención del espectador. Elegir los colores correctos y sus combinaciones es fundamental para el éxito de su diseño.

5.2 Diseño editorial

La rama del diseño gráfico enfocada en la presentación y disposición de diversas publicaciones, como libros, revistas, periódicos, catálogos y folletos, se conoce como diseño editorial. Esto incluye tanto el aspecto visual interno como externo del texto, teniendo en cuenta el concepto estético que define cada publicación y las condiciones de impresión y recepción (Rosas, 2012).

Los profesionales del diseño editorial buscan crear una armonía entre el texto, la imagen y la disposición de los elementos, con el objetivo de transmitir el mensaje del contenido de manera estética y atractiva para el público, con un impacto comercial positivo. Es importante tener en cuenta que un buen diseño editorial debe cumplir con ciertos objetivos como:

- Adaptarse al público objetivo
- Establecer una buena comunicación con el usuario
- Facilitar el proceso creativo
- Ser diferente a otras publicaciones del mismo tipo
- Alcanzar al público objetivo deseado.

Como se ve, incorporar un diseño editorial en el diseño de un libro de texto para la enseñanza de la Estadística es esencial, ya que su objetivo es organizar y presentar la información de manera coherente y atractiva, combinando texto e imagen para que de esta manera el usuario pueda interactuar con la información de manera cómoda y amena.

5.3 Elementos básicos

En el contexto de una creación editorial las ideas que debe atender un autor son planteadas por Iñigo y Makhoulf (2014):

- **Párrafo:** es la unidad básica de escritura y consiste en una serie de oraciones interconectadas que se unen para formar una idea o argumento. Los párrafos están separados por espacios y suelen tener una estructura clara y coherente.
- **Legibilidad:** Se refiere a la facilidad con la que el lector lee y comprende el texto. La legibilidad depende de varios factores, como la fuente, el tamaño de la fuente, la alineación y la estructura del texto.
- **Portada:** es la parte exterior de un libro que protege y muestra el contenido. Las portadas pueden incluir títulos, autores, imágenes o ilustraciones, y otros elementos gráficos o informativos que llamen la atención del lector.
- **Maquetación:** se refiere a la presentación visual de las páginas del libro, incluida la selección y organización del texto y los elementos gráficos de cada página. El diseño afecta la legibilidad del texto y puede incluir elementos como márgenes, sangrías, números de página, encabezados y pies de página.
- **Retícula:** es una estructura invisible en un diseño que se utiliza para dividir una página en columnas y filas. Las retículas ayudan a crear una jerarquía visual de los elementos de la página y pueden mejorar la legibilidad y la organización del contenido.

5.3.1 Diseño interactivo

Este tipo de diseño es caracterizado por Aubia quien es citado por Serrano y Pugo (2021) porque éste busca soluciones en las que el lector se relaciona con el contenido de manera natural a través de herramientas que le permiten realizar acciones o navegar entre el contenido.

La interactividad en un libro puede ocurrir de diversas maneras, dependiendo del tipo de libro y del enfoque de la interactividad que se quiera implementar. Algunos ejemplos de cómo se puede lograr la interactividad en un libro son:

- **Hipervínculos:** Los hipervínculos son enlaces en un libro que permiten a los lectores acceder a más información sobre un tema en particular. Estos enlaces pueden dirigir a los lectores a otras partes del libro, sitios web relacionados, videos, imágenes, etc.
- **Animaciones:** algunos libros incluyen animaciones para ilustrar ciertos conceptos a los lectores de una manera más visual y atractiva.
- **Juegos y actividades interactivas:** algunos libros incluyen juegos o actividades interactivas que permiten a los lectores aplicar lo aprendido en el libro de una manera más interesante.

5.4 Herramientas para la programación de un libro interactivo

R Studio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R; este software ofrece una interfaz amigable y fácil de usar para escribir y ejecutar código, así como para visualizar y analizar datos. Lo anterior aunado a su carácter de software libre lo convierte en una herramienta ideal para enseñar estadística. Una de las principales ventajas de R Studio es que permite a los estudiantes ver los resultados de su código de manera inmediata sin necesidad de aprender comandos y sintaxis, esto les ayuda a concentrarse más en los conceptos estadísticos y su respectiva interpretación además que permite adoptar un enfoque experimental en el que se validan conjeturas y opiniones hasta llegar a consolidar unas conclusiones válidas.

Un rasgo característico de R es el trabajo a través de librerías y paquetes disponibles para realizar tareas estadísticas específicas, como análisis de regresión, el análisis de clúster, métodos multivariados, muestreo y el análisis de datos categóricos entre otros. Estas librerías son fáciles de instalar y utilizar, lo que permite a los estudiantes acceder a una variedad de herramientas avanzadas sin necesidad de escribir un código complejo. Así también, cada día hay más recursos para crear gráficos y visualizaciones atractivas en R Studio, lo que ayuda a que los estudiantes

comprendan mejor y con más profundidad los datos permitiendo la comunicación de sus hallazgos de manera efectiva.

Adicionalmente, el entorno de RStudio puede incluir a Rshiny y RMarkdown, las cuales son dos herramientas muy útiles para crear libros virtuales para la enseñar estadística. Ambos son complementos de R de los que se puede resaltar:

- RShiny es una herramienta para crear aplicaciones web interactivas a partir de la potencia computacional de R y la interactividad ganada a través del internet que permite a los profesores crear aplicaciones que posibilitan a los estudiantes explorar y visualizar datos de manera dinámica vinculado recursos de visualización modernos como son los datos geo referenciados, trabajo con datos no estructurados como sería una colección de tweets, creación de dashboards y en general usar simulación computacional. Claramente estos elementos se convierten en un recurso didáctico que debe explotarse al máximo.
- *RMarkdown* es una herramienta para crear documentos dinámicos que tengan código de R, con ella es posible crear diferentes formatos como presentaciones dinámicas, informes y libros digitales que incluyen texto, imágenes, tablas, gráficos y código R que se compila para hacer visible la salida y su modificación.

5.5 Elementos de didáctica de la estadística

El pensamiento estadístico se basa en tres principios esenciales: todo trabajo ocurre en un sistema de procesos interconectados, la variación existe en todos los procesos y la clave del éxito se alcanza comprendiendo y reduciendo la variación del proceso (Carballo y Domínguez, 2010). Por su parte, Estrella (2017) nos dice que el pensamiento estadístico tiene cuatro características fundamentales: reconocer la necesidad de datos, tras numeración, percepción de la variación y razonamiento con incertidumbre.

Por lo anterior es claro que el aprendizaje del pensamiento estadístico no es unidimensional, Behar y Grima (2004) sugieren que entender la estadística va más allá de un enfoque lineal o limitado, destacan la necesidad de adoptar una perspectiva amplia, considerando diferentes aspectos y métodos para capturar la esencia completa de la estadística en el contexto educativo superior. Esto implica un reconocimiento de la estadística no solo como un conjunto de técnicas, sino como una disciplina rica y profunda.

5.5.1 Aprendizaje activo

George Cob (1992), define una nueva línea para la enseñanza de los cursos introductorios de estadística universitarios, proponiendo las metas para la enseñanza de la estadística y recomendando: a) enfatizar en el pensamiento estadístico, b) más datos y conceptos y menos teoría y recetas, y c) fomentar el aprendizaje activo en el sentido de Valverde (2013).

El aprendizaje activo se refiere a una estrategia de instrucción en la que los estudiantes participan de manera directa en su proceso de aprendizaje a través de una serie de actividades y técnicas diseñadas para involucrarlos. La estrategia se enfoca en crear un ambiente de aprendizaje interactivo y participativo donde los estudiantes tengan la oportunidad de explorar, descubrir y construir conocimiento a través de la resolución de problemas y la colaboración con sus compañeros. El objetivo principal del aprendizaje activo es hacer que los estudiantes intervengan de manera autónoma y efectiva, siendo capaces de aplicar lo aprendido en situaciones reales y resolver problemas complejos de manera más efectiva. Algunas estrategias de aprendizaje activo incluyen trabajo en equipo, discusiones grupales, resolución de problemas, uso de tecnología educativa, retroalimentación continua y aplicación práctica de conceptos aprendidos en situaciones de la vida real.

5.5.2 Enseñanza basada en proyectos

Batanero (2004) destaca varias razones por las que la enseñanza basada en proyectos es beneficiosa para los cursos de estadística. En primer lugar, la estadística está estrechamente relacionada con sus aplicaciones, lo que en última instancia se justifica por su papel en la resolución de problemas que van más allá de la propia estadística. El aprendizaje a través de proyectos permite a los estudiantes aplicar los conocimientos estadísticos a situaciones del mundo real, aumentando su pertinencia. Además los estudiantes desarrollan habilidades de pensamiento crítico y resolución de problemas al pensar en temas actuales alrededor de los cuales se plantean los proyectos.

Batanero (2004) también señala que un proyecto de estadística puede aumentar la motivación de los estudiantes porque les permite elegir temas de interés y utilizar datos reales y significativos, aprender sobre datos e ideas del mundo real, que favorecen el uso de conceptos como precisión, variabilidad, confiabilidad, mensurabilidad y sesgo que no estarían presentes en unos "datos inventados por el maestro". En general, un enfoque basado en proyectos puede ser una forma eficaz y motivadora de enseñar estadística a los estudiantes.

5.5.3 Análisis exploratorio de datos

Propuesto por Tukey como una primera etapa que debiera seguir todo análisis estadístico, el propósito inicial es conocer los datos desde múltiples perspectivas apoyándose para ello en herramientas de Estadística Descriptiva, luego vienen procesos como la limpieza de datos, validación de supuestos y formulación de hipótesis. No obstante, de ser estos los resultados esperados para este tipo de análisis, desde el punto de vista didáctico se destaca que adoptar este tipo de análisis en el aula promueve una actitud positiva hacia la investigación científica. Se aclara que, aunque el término se asocia para el trabajo con grandes bases de datos, nosotros en este

proyecto lo adoptamos como un enfoque didáctico para el trabajo con objetos estadísticos como gráficas y tablas.

6. Metodología

Esta propuesta de investigación es de carácter aplicado y se enmarca en la modalidad de producción de material didáctico, para su desarrollo se plantea un trabajo que ha sido organizado a través de cinco fases:

Tabla 1.

Etapas que incluyó el diseño y desarrollo del producto final

Fase	Fechas	Actividad
1	Octubre de 2022 - enero de 2023	Consulta bibliográfica en temas como: Educación estadística, Diseño y edición de textos digitales, programación y práctica con herramientas para edición de textos interactivos en Estadística (R, RMarkdown, RShiny).
2	Febrero de 2023-abril de 2023	Diseño y programación de recursos digitales tomando como base el material de clase utilizado por la directora del Trabajo de grado.
3	Mayo de 2023 – junio de 2023	Prueba piloto con una muestra de estudiantes del curso de Estadística I
4	Julio de 2023 – agosto de 2023	Revisión de estilo y de contenido
5	Agosto – septiembre de 2023	Ajuste y edición del capítulo de libro

7. Resultados

El producto de este proyecto es un capítulo de libro digital que gira en torno al tema de análisis de datos a través de gráficos estadísticos desde un enfoque exploratorio. Este recurso, integra de manera efectiva el aprendizaje teórico y práctico, proporcionando a los usuarios una plataforma dinámica para explorar conceptos estadísticos a través de gráficos y simulaciones programadas en R. El código del programa construido en *Shiny* se ha dispuesto en el Anexo 1 al final de este documento.

En el desarrollo del libro, se recurrió a una serie de herramientas de programación para proporcionar una experiencia de aprendizaje enriquecedora y accesible. Se destacó la integración de librerías especializadas como *Shiny*, que facilita la creación de aplicaciones web interactivas; *ggplot2* y *plotly*, para la elaboración de visualizaciones de datos detalladas y dinámicas; *leaflet*, para incorporar mapas interactivos; y *MathJax*, para la inclusión precisa de ecuaciones matemáticas. Con el objetivo de asegurar una experiencia de usuario óptima, se personalizó el *CSS* de la aplicación, ajustando elementos visuales como la disposición de botones flotantes, y se utilizó *JavaScript* para añadir funcionalidades específicas, tales como la lectura de texto en voz alta y la adaptación del tamaño del texto. Estas medidas reflejan un compromiso con la mejora de la accesibilidad y la interactividad, garantizando que el contenido sea no solo comprensible, sino también atractivo y fácil de manejar para todos los usuarios.

Un ejemplo de una aplicación interactiva incorporada en el libro, es el histograma de la distribución de puntuaciones de coeficiente intelectual (CI). En la interfaz de usuario (UI) se establece la estructura y los elementos interactivos que utilizará el usuario para interactuar con el histograma como se muestra a continuación:

```
titlePanel("Distribución de Puntuaciones de CI"),
```

```
fluidRow(  
  width = 8,  
  plotOutput("interactiveHistogram"),  
  sliderInput("bins", "Número de barras:", min = 1, max = 20, value =  
    15),
```

En donde:

- **titlePanel:** Añade un título a la aplicación Shiny, que en este caso es "Distribución de Puntuaciones de CI"
- **fluidRow:** Organiza los elementos en una fila que se adapta al ancho del navegador. El parámetro `width = 8` indica que el histograma y el control deslizante ocuparán 8/12 del ancho disponible.
- **plotOutput:** Crea un espacio donde se mostrará el histograma interactivo. El identificador "interactiveHistogram" se vincula con el servidor para mostrar el gráfico generado.
- **sliderInput:** Genera un control deslizante que permite a los usuarios ajustar el número de barras en el histograma. Los valores van de 1 a 20, y el valor predeterminado es 15.

Ahora bien, en el servidor, se realizan los cálculos y se genera el histograma basado en los datos y las entradas de los usuarios como se muestra a continuación.

```
output$interactiveHistogram <- renderPlot({  
  ci_scores <- rnorm(1000, mean = 100, sd = 15)  
  ci_ranges <- data.frame(  
    bins = 15,  
    ci_scores = ci_scores,  
    ci_ranges = ci_ranges
```

```

start = c(55, 70, 85, 115, 130),
end = c(70, 85, 115, 130, 145),
label = c("Discapacidad", "Inferior", "Promedio", "Superior", "Excepcional"),
y_position = c(0.015, 0.015, 0.015, 0.015, 0.015),
x_position = c(62.5, 77.5, 100, 122.5, 137.5)
)
ggplot(data = data.frame(ci_scores), aes(x = ci_scores)) +
  geom_histogram(aes(y = ..density..),
    binwidth = 160/input$bins,
    fill = "#9F82F7",
    alpha = 0.5) +
  stat_function(fun = dnorm,
    args = list(mean = mean(ci_scores), sd = sd(ci_scores)),
    color = "red",
    size = 1) +
  geom_vline(data = ci_ranges, aes(xintercept = start), linetype = "dashed",
    color = "gray") +
  geom_vline(data = ci_ranges, aes(xintercept = end), linetype = "dashed",
    color = "gray") +
  geom_text(data = ci_ranges, aes(x = x_position, y = y_position, label = label),
    angle = 90, vjust = 0.5, hjust = 0.5) +
  labs(x = "Puntuación de CI", y = "Densidad") +
  xlim(40, 160)

```

```
})
```

En el cual:

- **output\$interactiveHistogram:** Define un objeto de salida que estará vinculado con plotOutput en la UI.
- **renderPlot:** Contiene el código para generar el gráfico. Este bloque se ejecutará cada vez que el usuario cambie el número de barras con el control deslizante.
- **rnorm:** Genera un conjunto de datos simulados de puntuaciones de CI, asumiendo una distribución normal con una media de 100 y una desviación estándar de 15.
- **ci_ranges:** Define un marco de datos con los rangos de CI y las etiquetas correspondientes, así como las posiciones para las líneas y el texto en el gráfico.
- **ggplot:** Inicializa la construcción del gráfico con ggplot2. Se asigna la columna ci_scores como el eje x.
- **geom_histogram:** Añade barras al gráfico para el histograma, con una transparencia y color específicos, y ajusta el ancho de las barras en función de la entrada del usuario.
- **stat_function:** Añade una curva de densidad normal sobre el histograma.
- **geom_vline y geom_text:** Añaden líneas verticales y etiquetas de texto para los rangos de CI.
- **labs:** Establece las etiquetas de los ejes x e y.
- **xlim:** Define los límites del eje x.

A continuación se presenta de manera detallada la estructura del capítulo del libro

7.1. Estructura del capítulo

Capítulo 1: Análisis de datos a través de gráficos estadísticos

Introducción

1.1. Validando un supuesto distribucional

1.2 Pruebas gráficas

1.2.1 Diagrama de caja y bigotes

1.2.2 Histograma

1.2.3 PP plot y QQ plot

1.3 Comunicando los resultados

1.4 Ejercicios propuestos

A continuación, se hará una descripción de cada sección del capítulo enfatizando las consideraciones y enfoque que se tuvieron en cuenta para el diseño y selección del material allí presentado.

La unidad didáctica ha sido diseñada no solo para proporcionar conocimientos teóricos sino también para asegurar que el lector pueda aplicar estos conceptos en situaciones reales. El enfoque didáctico combina tres elementos: exploración de datos reales, interactividad y la preparación para la comunicación efectiva de resultados. En resumen, se busca preparar al lector para enfrentar desafíos del mundo real en el ámbito del análisis de datos.

7.1.1 Introducción

El objetivo de esta sección es sensibilizar al lector sobre la importancia de seleccionar medidas de tendencia central adecuadas, a partir de un ejemplo práctico basado en datos de población de las ciudades capitales en Colombia, tal como se observa en la figura 1, se ilustra cómo el uso indiscriminado del promedio puede conducir a presentar descripciones erradas del comportamiento de una variable, por lo cual, el analista debe tener criterios técnicos para

incorporar adecuadamente esta medida de resumen; de manera poco usual, se motivará el uso de gráficos estadísticos desde la validación de un supuesto distribucional como puede ser la normalidad. La sección no es rigurosa a nivel teórico y es principalmente textual, el objetivo es preparar al lector para futuros elementos interactivos que se encontrarán en el resto del capítulo y valorar el uso de los gráficos como pruebas estadísticas esenciales en un análisis de datos.

Figura 1

Primera sección: Introducción



El ejemplo escogido ilustra una situación muy frecuente en el mundo real, con lo cual se espera que el tema a motivar resulte relevante, sirve al mismo tiempo como un llamado a que los lectores sean más críticos y metódicos en su aproximación a un análisis de datos. Esta introducción sienta las bases para temas más avanzados que se abordarán en capítulos posteriores, como pruebas estadísticas de normalidad, medidas de tendencia central y técnicas más avanzadas de análisis de datos como la detección de valores extremos o ajuste de modelos.

7.1.2 Validando un supuesto distribucional

En esta sección se pretende guiar al lector en cómo validar supuestos distribucionales en diferentes contextos de análisis de datos. Establece un marco conceptual para entender por qué es

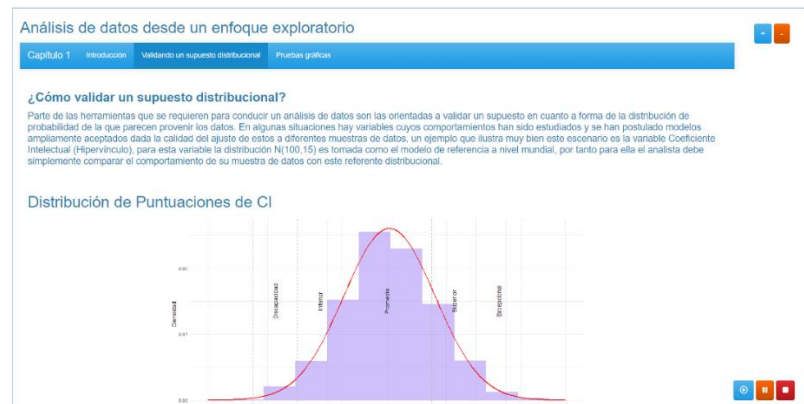
esencial validar la forma de la distribución de una variable, tanto en situaciones donde ya existe un modelo preestablecido como cuando no lo hay.

En cuanto a contenido, se introduce el concepto de supuestos distribucionales, destacando su importancia en la estadística. Además, se brinda una revisión sobre cómo y cuándo aplicar modelos distribucionales preestablecidos, usando datos del Coeficiente Intelectual lo cual se ejemplifica en la figura 2.

También, se aborda la complejidad de trabajar con datos que no tienen un modelo preestablecido, enfatizando la tarea del analista en identificar la distribución que mejor se ajusta a los datos.

Figura 2

Sección 1: Cómo validar un supuesto distribucional



La sección se enriquece con histogramas ajustables, elementos que permiten al lector explorar más profundamente los conceptos y aplicarlos a diferentes conjuntos de datos.

Esta sección es crucial para cualquiera que esté interesado en análisis de datos, ya que establece la base para entender la importancia de validar supuestos distribucionales. Al hacerlo, permite al lector evitar conclusiones erróneas o interpretaciones incorrectas.

Se introduce al lector en el uso de diversas herramientas gráficas para la validación de supuestos, como histogramas, boxplot y gráficos PP y QQ, entre otros. Además, se invita al lector a considerar la inspección de otras opciones gráficas menos populares, pero igualmente útiles.

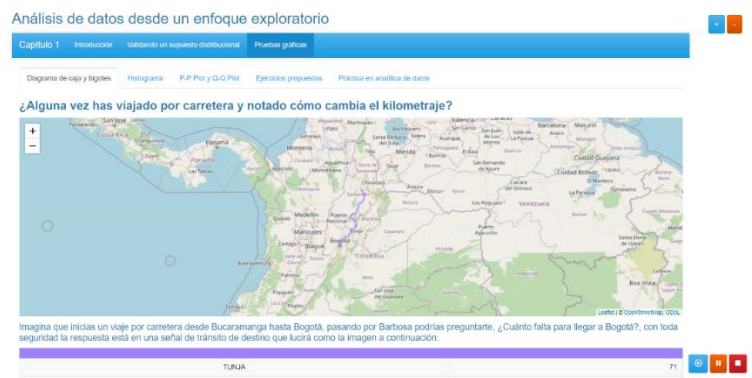
7.1.3. Pruebas gráficas

El objetivo de la Sección 2 es dotar al lector de las herramientas necesarias para visualizar y entender conjuntos de datos estadísticos a través de diferentes gráficos y representaciones. Específicamente, se espera que el lector aprenda a crear y entender diagramas como el histograma, pero se dedicará tiempo para introducir gráficos menos populares como son caja y bigotes, PP y QQ plots, los cuales son introducidos a partir de un gráfico que técnicamente no existe pero se ideó para facilitar el entendimiento del PP y QQ plots, a este nos referiremos como CC plot el cual se construye a partir de los cuartiles empíricos (muestrales) y los teóricos.

7.1.3.1. Diagrama de caja y bigotes. Se inicia utilizando una analogía con un recorrido por una ruta de viaje (ver figura 3). Así, relacionar el uso de los estadísticos de orden para resumir un conjunto de datos con la señalización vial sobre distancia recorrida en kilómetros puede facilitar al lector la interpretación y uso práctico de estos recursos.

Figura 3

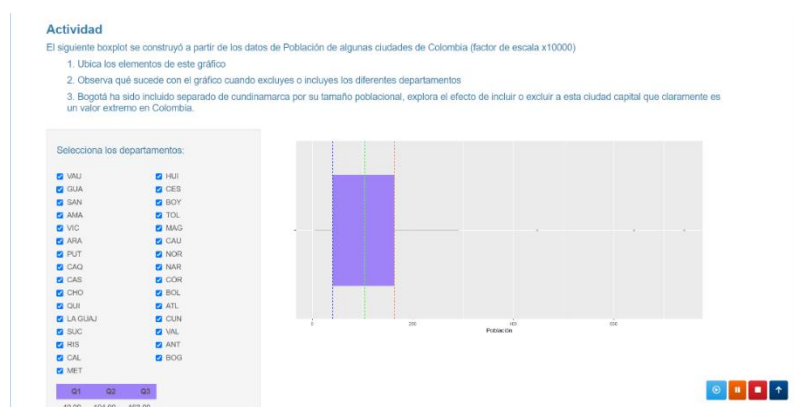
Sección 2, pruebas gráficas. Subsección, diagrama de caja y bigotes. Mapa de ruta y señales de kilometraje



Para el desarrollo del tema, se presenta un escenario interactivo que permite visualizar diferentes configuraciones de un diagrama de caja y bigotes construidos a partir de la variable población de las ciudades capitales colombianas tal como se observa en la figura 4, la aplicación permite hacer modificaciones al conjunto inicial de 32 datos a través de la activación o desactivación de un dato particular.

Figura 4

Sección 2, pruebas gráficas. Subsección, diagrama de caja y bigotes. Boxplot interactivo población por departamento.



El propósito de esta actividad es identificar los elementos claves de este gráfico a través del código de color que relaciona las definiciones y su ubicación en el gráfico. Luego, a través de actividades prácticas, los lectores pueden aplicar conceptos teóricos para calcular cuartiles, mediana y rango intercuartílico utilizando software estadístico como R.

Sección 2, pruebas gráficas. Subsección, Histograma. Algoritmo para la construcción de un histograma



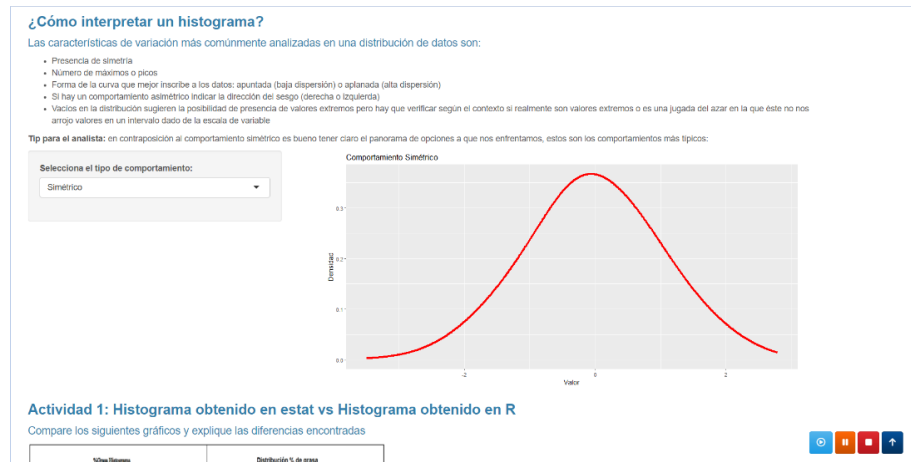
Dado el auge del análisis de datos en diversas áreas, desde la ciencia hasta los negocios, la habilidad de construir e interpretar histogramas es fundamental. Este enfoque es particularmente útil para evitar errores comunes que los analistas principiantes pueden cometer al depender únicamente de medidas resumidas como el promedio.

La sección utiliza un enfoque didáctico y práctico, proporcionando un algoritmo paso a paso para construir histogramas manualmente, así como un tutorial de cómo hacerlo utilizando el lenguaje de programación R. Se basa tanto en teoría estadística como en aplicaciones prácticas para explicar su relevancia y utilidad.

Se incluyen múltiples visualizaciones para ilustrar los conceptos estadísticos claves, como se muestra en la Figura 7, se programó una herramienta que ilustra diferentes gráficas de densidad de una distribución y ayuda a identificar las características de variación más comunes. Cada gráfico viene acompañado de una leyenda clara y una breve descripción para asegurar que sean fáciles de entender.

Figura 7

Sección 2, pruebas gráficas. Subsección, Histograma. ¿Cómo interpretar un histograma?

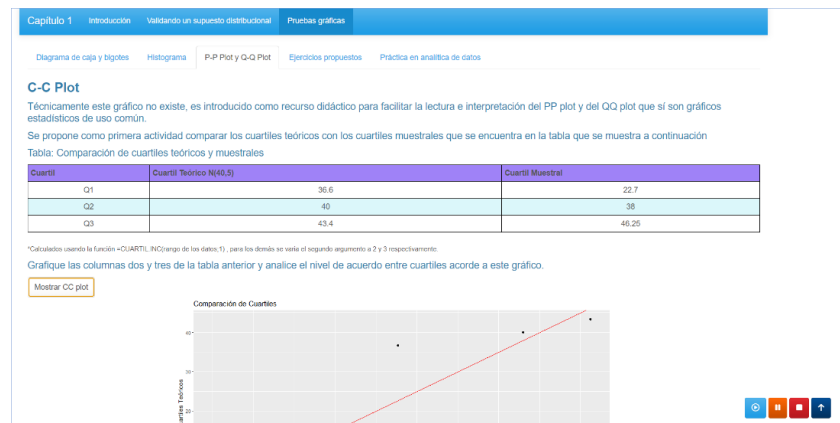


Esta sección sirve como un puente entre las medidas de tendencia central, que serán tratadas en otros capítulos del texto en construcción y, temas más avanzados como pruebas de hipótesis, que serán abordados en secciones posteriores. Al dominar la construcción e interpretación de histogramas, el lector estará mejor preparado para comprender conceptos más complejos.

7.1.3.3. PP y QQ plot. La sección tiene como objetivo principal enseñar diferentes métodos gráficos para evaluar la normalidad de una muestra de datos. Específicamente se introduce una secuencia de tres tipos de gráficos: Gráfico cuantil-cuantil o C-C Plot, aspecto novedoso que ha mostrado ser efectivo para la interpretación de los dos siguientes tipos de gráficos, el percentil-percentil o P-P Plot y el cuantil-cuantil o QQ Plot. Se proporcionan instrucciones detalladas para la creación e interpretación de estos gráficos (ver figura 8), con el fin de determinar si los datos pueden considerarse procedentes de una distribución normal.

Figura 8

Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. Comparación de cuartiles teóricos y muestrales



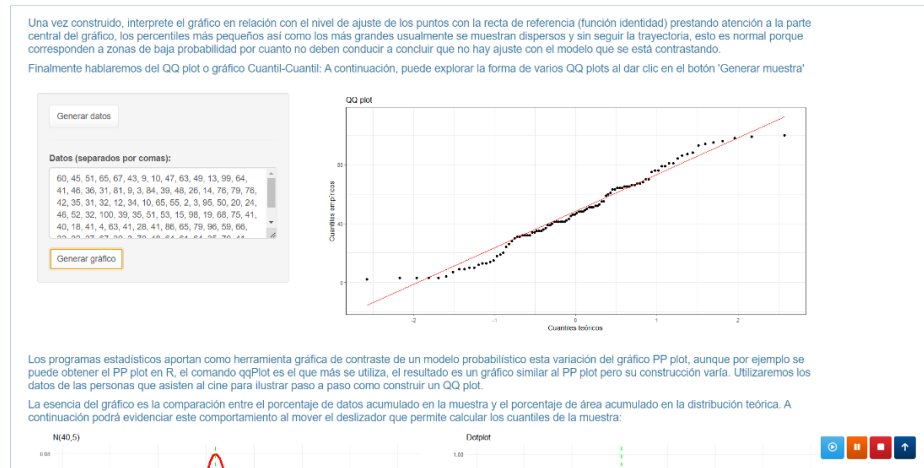
La sección es especialmente relevante para analistas de datos, estadísticos y cualquier persona interesada en verificar los supuestos de normalidad en un conjunto de datos.

Se utiliza una combinación de instrucciones paso a paso, ejemplos y ejercicios prácticos. La sección también incluye fórmulas matemáticas para calcular cuantiles y un código R para generar un QQ Plot. Los ejemplos se derivan de un archivo de Excel específico y datos simulados.

La sección hace uso de varios recursos estadísticos: una tabla que compara cuantiles teóricos y muestrales, un gráfico que muestra la relación entre estos cuantiles, y gráficos P-P y QQ. También hay un componente interactivo que permite a los usuarios generar un gráfico QQ Plot introduciendo su propia muestra de datos, tal como se evidencia en la figura 9.

Figura 9

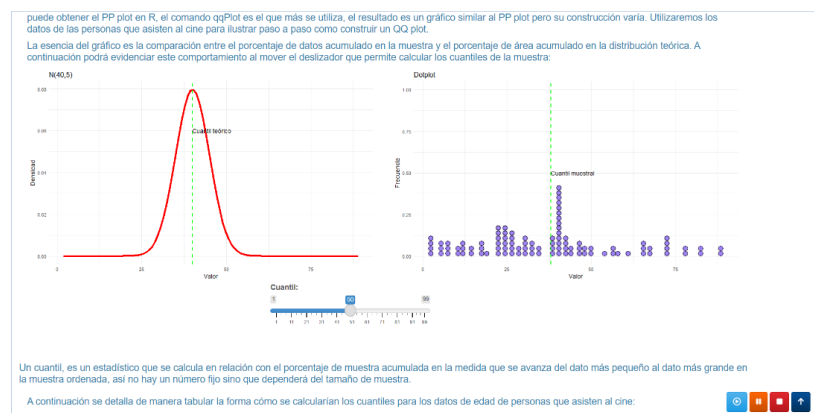
Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. QQ plot interactivo



En la figura 10, se ilustra la esencia de un gráfico QQ plot que es facilitar la comparación del porcentaje de datos acumulados en una muestra contra el porcentaje de área acumulada en una distribución teórica. Esta comparativa visual es importante para observar cómo los datos se ajustan a una distribución esperada y es fundamental para el análisis de la normalidad en un conjunto de datos.

Figura 10

Sección 2, pruebas gráficas. Subsección, P-P Plot y Q-Q Plot. Comparación curva normal y Dotplot



7.1.3.4. Ejercicios propuestos. La sección de "Ejercicios Propuestos" busca profundizar en la comprensión y aplicación práctica de métodos gráficos y análisis estadísticos poniendo

especial énfasis en el uso de diferentes softwares para la visualización y comparación de datos. En esta sección, se invita a los usuarios a confrontar y conciliar histogramas construidos a partir de la misma muestra pero generados con softwares distintos, como "ESTAT" y "R".

Especialmente diseñada para estudiantes, analistas y profesionales del campo estadístico, esta sección pone a prueba la habilidad del lector para interpretar y comparar visualmente las distribuciones. Es esencial en el mundo del análisis de datos poder corroborar y validar los resultados obtenidos independientemente del software utilizado.

Los ejercicios propuestos se centran en la interpretación de histogramas que representan, por ejemplo, la distribución del porcentaje de grasa corporal. Esta métrica, vital en muchos campos como la medicina y el acondicionamiento físico, se analiza en detalle, considerando la relación entre la grasa corporal y otras medidas físicas.

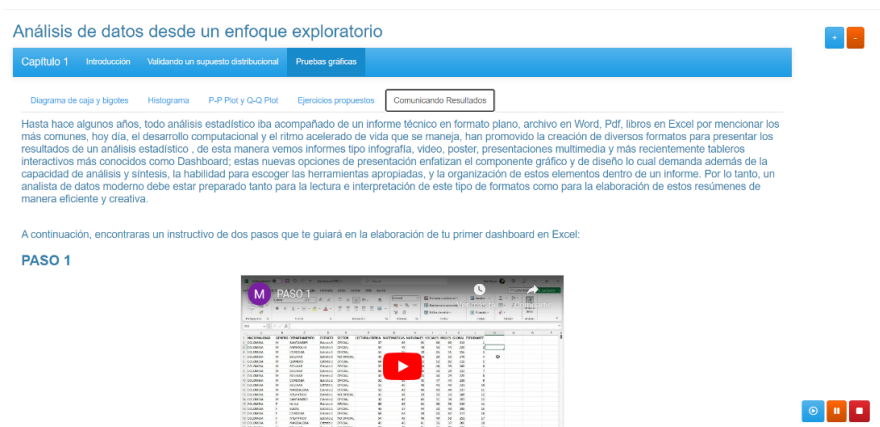
La sección es enriquecida con ejemplos prácticos y situaciones reales, derivados de una muestra de 252 hombres, y desafía al usuario a reflexionar sobre el comportamiento probabilístico de los datos. Acompañados de instrucciones claras, gráficas detalladas y contextos relevantes, los ejercicios buscan reforzar el conocimiento adquirido y preparar al lector para futuros análisis estadísticos.

Adicionalmente, la sección incorpora herramientas interactivas, permitiendo a los usuarios no solo observar sino también interactuar, comparar y aprender de manera más efectiva. El capítulo saca el máximo provecho de su formato interactivo al ofrecer actividades donde los lectores pueden ingresar sus propios datos, calculadoras en línea para realizar cálculos de estadísticas resumen, y un cuadro de entrada que ejecuta código en R. También presenta problemas y preguntas para estimular el pensamiento crítico y la aplicación práctica de los conceptos discutidos.

7.1.3.5. Comunicando resultados. En esta sección de "Comunicando resultados", añadimos un componente adicional: video tutoriales (ver figura 11). Estos videos son esenciales para guiar al usuario en el proceso de creación de dashboards con gráficos dinámicos construidos en Excel, una herramienta ampliamente accesible y utilizada en el mundo profesional. Esta instrucción paso a paso asegura que incluso aquellos con conocimientos básicos de Excel, puedan crear presentaciones impactantes y profesionales que aporten elementos claves para realizar un análisis estadístico eficiente.

Figura 11

Sección 2, pruebas gráficas. Subsección, Comunicando Resultados. Video tutoriales para realizar Dashboard en Excel



Este capítulo supone un conocimiento previo en probabilidad, como es el manejo de la distribución normal y elementos básicos de estadística descriptiva como medidas de tendencia central, de dispersión y de posición.

7.2. Pruebas de Usuario

Las pruebas de usuario que se realizaron para el capítulo del libro, se llevaron a cabo mediante una prueba piloto con estudiantes del curso de Estadística I. Estas pruebas se llevaron a cabo en dos sesiones distintas de dos horas cada una, permitiendo una evaluación de la interacción del usuario con el material.

7.2.1. Procedimiento de las pruebas

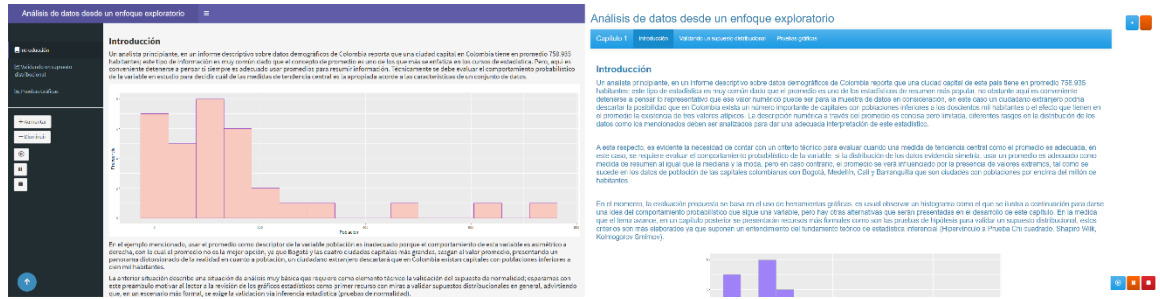
En la primera sesión, se permitió a los estudiantes interactuar con el libro bajo la supervisión y guía de la docente. Se les pidió que se enfocaran en probar diversas funcionalidades incorporadas, como ejercicios interactivos o herramientas gráficas.

Al finalizar la sesión se solicitó a los estudiantes realizar comentarios escritos con respecto a las funcionalidades exploradas. Estos comentarios se recopilaron con el objetivo de tener una retroalimentación valiosa sobre la interfaz de usuario y la utilidad de las herramientas que fueron integradas en el capítulo. Para una revisión detallada de las observaciones proporcionadas por los estudiantes, se remite al apéndice A.

Tras el análisis de los comentarios recibidos por los estudiantes, se realizaron ajustes significativos al capítulo del libro. Entre las mejoras más notables, se refinó la interfaz de usuario, se incluyeron más entornos interactivos y se mejoró el contenido para facilitar un aprendizaje autónomo y efectivo tal como se ilustra en la figura 12.

Figura 12

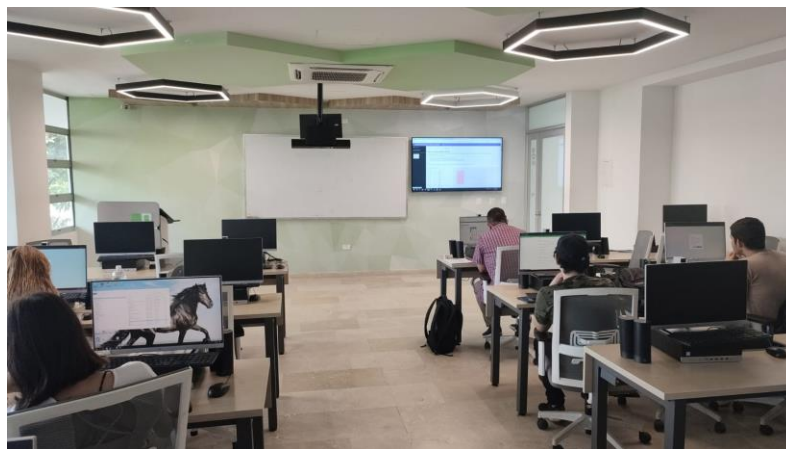
Comparación de interfaz gráfica antes y después de pruebas de usuario.



En la segunda sesión de pruebas, los estudiantes utilizaron una versión actualizada del libro, la cual incluía las mejoras derivadas de la primera sesión, relacionadas a los ejercicios interactivos y contenido en general tal como se observa en la figura 13.

Figura 13

Pruebas de usuario, segunda sesión.



En esta etapa, se presta atención al uso del recurso de manera independiente, es decir, sin la instrucción directa de un profesor. A partir de allí, se continuó trabajando en la mejora de las

instrucciones y la navegación, para que los estudiantes puedan aprovechar al máximo el material en un contexto autodidacta.

8. Conclusiones

Nuestra revisión bibliográfica de libros de texto en estadística básica nos ofreció una perspectiva valiosa sobre la evolución de estos recursos educativos. Detectamos una carencia significativa en el diseño pedagógico de las actividades y materiales complementarios; específicamente, notamos la ausencia de una estructura didáctica intencionada que conduzca al estudiante a través del proceso de interpretación de datos. Estos libros tienden a seguir un enfoque educativo tradicional, centrado en el profesor, en lugar de fomentar el aprendizaje autónomo. Esta observación es especialmente relevante en el contexto de la educación virtual, donde el autoaprendizaje y la autogestión son cruciales para el éxito académico.

La unidad didáctica, que forma parte del libro interactivo desarrollado como producto de esta tesis, al ser pilotada con un grupo de estudiantes, reveló un incremento notorio en la versatilidad y autonomía al seleccionar gráficos estadísticos para apoyar sus presentaciones en el proyecto final de clase, que se centraba en adaptar un modelo de probabilidad a un conjunto de datos. Aunque estos resultados no son lo suficientemente concluyentes como para generalizar, sí ofrecen un indicativo prometedor del valor que el enfoque exploratorio y la incorporación activa de gráficos estadísticos pueden aportar a las etapas preliminares de cualquier análisis de datos.

La combinación de paquetes de R como RShiny, ShinyJS, junto con HTML, representa una innovación en la creación de libros digitales interactivos centrados en el aprendizaje de la estadística. Estas herramientas ofrecen flexibilidad y funcionalidad, permitiendo la creación de un

entorno de aprendizaje interactivo y dinámico. Estos recursos tecnológicos no solo transforman la manera en que los contenidos estadísticos se presentan y exploran, sino que también abren nuevas posibilidades para un aprendizaje más autónomo, crítico y enfocado en problemas del mundo real.

La limitación en la capacidad de programación puede presentar un obstáculo significativo en el desarrollo de proyectos digitales, especialmente cuando se requieren el uso de paquetes y lenguajes especializados como R. En este contexto, la inteligencia artificial emerge como una solución invaluable para superar estas barreras. Aunque una base en programación es ciertamente beneficiosa, la colaboración con sistemas de inteligencia artificial generativa puede agilizar considerablemente el proceso, llenando lagunas de conocimiento y permitiendo una mayor eficiencia en la ejecución del proyecto.

La retroalimentación obtenida tanto en la muestra piloto como en la participación como ponentes del 31° Simposio Internacional de Estadística resalta la efectividad de nuestro enfoque didáctico. El aspecto más destacado fue la integración armónica de la teoría, el contexto y la programación en R, lo cual resultó ser especialmente valioso para dos grupos demográficos clave: estudiantes que están comenzando su formación en estadística y profesionales de otras disciplinas interesados en aplicar resultados estadísticos en sus respectivos campos. Así, las apreciaciones recibidas validan nuestro enfoque y sugieren un alto potencial para su implementación más amplia en la enseñanza de la estadística.

Adaptar actividades de aula a un libro interactivo fue un desafío especialmente en sustituir la guía del profesor en tiempo real, en el capítulo de libro se buscó reemplazar esta guía con

instrucciones escritas y herramientas adecuadas para orientar al lector. Se resalta la necesidad de instrucciones claras, ejemplos y elementos interactivos para facilitar el aprendizaje autónomo. A pesar de la complejidad, es esencial innovar en las técnicas educativas para adaptarse a los formatos modernos y el aprendizaje independiente.

Este proyecto representa un punto de partida para la evolución continua de esta herramienta. La siguiente fase ya se encuentra planificada: se contempla la extensión de la prueba piloto del primer capítulo a otras muestras de estudiantes, esto nos permitirá recolectar más datos y obtener retroalimentación lo que será invaluable para refinar y optimizar esta primera versión del libro interactivo. Adicionalmente, ya se encuentra en fase de diseño el desarrollo del segundo capítulo del libro que se enfocará en las medidas de asociación para variables tanto numéricas como cualitativas.

Esta progresión planificada resalta el carácter dinámico y evolutivo del proyecto, reafirmando nuestro compromiso con la mejora continua y la adaptabilidad a las necesidades cambiantes tanto de los estudiantes como del campo de la estadística en sí. Esperamos que cada nueva versión no solo mejore el contenido existente, sino que también amplíe la gama de temas cubiertos, fortaleciendo así la utilidad y aplicabilidad del libro en diferentes contextos educativos y profesionales.

Referencias bibliográficas

- Batanero, C., & Díaz, C. (2004). El papel de los proyectos en la enseñanza y aprendizaje de la estadística. Aspectos didácticos de las matemáticas, 125-164.
- Behar Gutiérrez, R y Ojeda Ramírez, M. (2013). El proceso de aprendizaje de la estadística.
- Behar Gutiérrez, R. & Grima Cintas, P. (2004). La Estadística en la Educación Superior ¿Formamos Pensamiento Estadístico? Ingeniería y Competitividad, 5(2).
- Berihuete Macías, Á., Ramos García, C. D., & García Ramos, J. A. (2018). Libro dinámico y colaborativo de Estadística descriptiva. Proyectos de Innovación y Mejora Docente 2017/2018. Universidad de Cádiz.
- Burbano-Pantoja, V. M. A., Valdivieso-Miranda, M. A., & Aldana-Bermúdez, E. (2017) Conocimiento base para la enseñanza: un marco aplicable en la didáctica de la probabilidad. Rev.investig. desarro.innov, 7(2), 269-285. Doi: 10.19053/20278306.v7.n2.2017.6070
- Carballo Monsivaisa, C.A. & Domínguez Domínguez, J. (2010). Impulsando el Pensamiento Estadístico con Seis Sigma. En Estadística Aplicada: “Didáctica de la Estadística y Métodos Estadísticos en Problemas Socioeconómicos”
- Cuétara Hernández, Y., Salcedo Estrada, I. M., & Hernández Díaz, M. (2016). La enseñanza de la estadística: antecedentes y actualidad en el contexto internacional y nacional. Atenas, 3(35), 125-140.
- Devore, J. L. (2019). Introducción a la probabilidad y estadística para ingeniería y ciencias. Cengage Learning.
- Estrella, Soledad. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico.

- Garfield, J. et al., 2005. Guidelines for Assessment and Instruction in Statistics Education (GAISE): College report. Alexandria, Virginia: The American Statistical Association.
- Garfield, J., Hogg, B., Schao, C., y Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, 10(2), 1-14.
- Ghinaglia, D. (2009). Taller de diseño editorial: entre corondeles y tipos. *Diseño en Palermo. IV Encuentro Latinoamericano de Diseño*.
- Guerrero Reyes, L. (2016). *El Diseño Editorial. Guía para la realización de libros y revistas* [Trabajo de fin de máster, Universidad Complutense de Madrid]. Repositorio de la Universidad Complutense de Madrid.
- Iñigo Dehud, L. S., & Makhlof Akl, A. (2014). *Diseño Editorial: Manual de conceptos básicos* (1st ed.). Universidad Autónoma del Estado de Morelos.
- Kapr, A. (1985). *101 reglas para el diseño de libros*. Empresa Editoriales de Cultura y Ciencia.
- Los libros electrónicos: la tercera ola de la revolución digital
- Mendenhall, W., Beaver, B. M., Beaver, R. J. (2015). *Introducción a la probabilidad y estadística*. Cengage Learning.
- Ministerio de Educación Nacional, (2006). *Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas: Guía sobre lo que los estudiantes deben saber y saber hacer con lo que aprenden*
- Navidi, W. (2022). *Estadística para ingenieros y científicos*. McGraw-Hill Interamericana.
- Serrano Quezada, K. J., & Pugo Guamán, M. E. (2021). *Diseño de un libro interactivo y material didáctico para generar conciencia en los niños y niñas sobre el cambio climático*.
- Software de Analítica y Soluciones “SAS”, https://www.sas.com/es_co/home.html
- Stat Key <https://www.lock5stat.com/StatKey/index.html>

Statistics: Unloking the power of data <https://www.lock5stat.com/>

The R Project for Statistical Computing, “R,” <https://www.r-project.org/>

Valverde, M. M. M. (2013) Aportes de la investigación en educación estadística a la enseñanza de la estadística en la universidad

Apendices

Apendice a Comentarios/Sugerencias estudiantes

Histograma

Al terminar la actividad, creo que lo único que hace falta es un apartado para realizar las operaciones ya que me dice de donde salen los resultados y el aplicativo crea el histograma automáticamente.

PERO seria de utilidad donde hacer los cálculos rapidamente y ver que el histograma es veridico.

En el paso 3, la casilla de incremento, sin importar el valor que coloque, no genera ningún cambio, esto hace que no quede claro de lo que es el incremento.

P-P Plot y Q-Q Plot

Al principio me dan una $N(50,5)$ y creo que seria de ayuda tener el grafico de la distribución para ir comparando y apoyándose en el gráfico.

Histograma

- Observaciones^o. En el paso 2) Indica la búsqueda de criterios para la muestra de clase en una IA. (chat GPT), Tener en cuenta que a veces falla con lo que dice y en oraciones no arroja ningún resultado cuando se pregunta.
- En el paso 3) No especifica si el incremento puede tomar o no valores decimales (como en el número de clases solo se toma la parte entera, se tiende a hacer la misma analogía), especificar que si puede tomarlos. Tampoco podemos evidenciar a qué hace referencia esta.

P-P Plot y Q-Q plot

- Observaciones^o. En el gráfico: "la esencia del gráfico es la siguiente comparación", el texto debajo de este no es muy claro (de difícil comprensión).
- Supervisar que todos vayan trabajando, algunos por pereza a preguntar, preferir no hacerlo y quedan estancados.

- ① Al inicio se le podría dar la opción al estudiante de tomar una base de datos diferente a la propuesta, para incentivar la investigación y la toma de datos
- ② No establecer un medio de búsqueda porque a veces el programa no lo sabemos utilizar y un niño no lo va a buscar en otro programa porque a son lineales
- ③ Cuando empiece la parte "Construyamos el histograma con R" se debería explorar y experimentar más con comandos para que el estudiante le vea lo bello del histograma

P-P Plot

- ① Se debería dar el rango de datos para encontrar el cuartil, Muestra en la primera parte
- ② La parte del QQ plot no entendí muy bien la gráfica

Ejercicios propuestos

- ① En la actividad 1 en el inciso c la palabra usando está repetida
- ② Sería chévere que se pudiera generar ahí mismo todo (Actividad 1) cómo cuando se propuso generar el histograma
- ③ Cuando se le pregunta el concepto a un estudiante, sería bueno poner un cuadro en donde interprete lo que leyó

Diagrama de Caja y Bigotes - Boxplot

- ① En la parte uno de reconocimiento de la caja y bigotes - Boxplot se le podría preguntar, los cambios entre los datos y la gráfica ^{se} puede ver o con un "¿Qué pasa si? los departamentos de BOG y ATL no están...

• Cuando se ingresa en número 0 en las clases, no entiendo porque salen 16 barras, no deberían estar.

- el programa colapsa al poner un número grande (9999)
- "1. La distribución a contrastar el la normal", no se queda clara la condición.

Histograma

- el texto está bien, pero el tema de graficar, usar r, el uso de los datos tampoco me queda claro
- el color pssia
- la 3 imagen de P-P y q-q no siento que sea la mejor

Observaciones

- No entendi como construir el histograma con R
no dice que digite el código y que cambie las comillas
Sencillas por comillas dobles, pero no me parece
no luego, y doy ejecutar código y nada.
Como explicar o detallar un poco más

Escribir el cuartil en decimales

No entiendo el 18*

La condición
Boxplot

No Ordena en la Consola los datos que ingreso el usuario.

El grafico con Bigotes no cambia a pesar que
Cambian los datos.

No se actualiza el boxplot con bigotes.

- ¿cómo a través del ejemplo el lector puede observar un modelo diferente?
- En los ejemplos que se encuentra texto en inglés, podría traducirse a español.
- En el ejemplo 2 no quedan específicos algunos datos; por ejemplo, qué jugadores tomaron, si fue de alguna competición o el año de recolección de los datos.
- No están claros conceptos como: rango, número de clase e incremento.
- Digitando el segundo comando no hubo necesidad de cambiar las comillas sencillas.
- En el apartado "¿cómo interpretar un histograma" en el último inciso "este" va sin tilde y "arajo" lleva tilde en la última "o".
- No queda clara los momentos en el que se habla de D-Q plot y P-P plot.
- Utilizar el mismo concepto o especificar que son similares el concepto de cuartil y cuantil.
- El orden de las actividades; por ejemplo que la actividad 4 este siguiente a la actividad 1, ya que guardan relación
- Como en la actividad 2 sugerir un contexto que ayude a extraer los datos en la actividad 3
- En general la sección "ejercicios propuestos" presentan ausencia de ortografía en algunas palabras, pero las actividades tienen un buen planteamiento.
- Buen uso de los colores para relacionar los conceptos con su representación gráfica
- En el primer ejemplo no hay valores verticalmente
- En términos personales haciendo los ejercicios de la actividad 4 no logré comprender cómo graficar un boxplot

- Si tuvieramos muchísimas frecuencias acumuladas y se necesita hacer un gráfico es mejor un histograma o una ojiva?
- Que sucede si llega a ocurrir que hay bastantes valores atípicos en un diagrama de caja y bigotes. ¿sucedería?
- Porque en el Boxplot interactivo yo coloco unos datos separados por comas pero si te doy un espacio entre la coma y el siguiente número me aparece una línea roja extendida en todo el gráfico pero si no lo separo (copio así seguidos los datos) me aparece el bigote dentro de la caja sin extenderse por todo el gráfico?

Histogramas:

- En la parte del enlace para descargar el archivo de base de datos si se le da click directamente arroja un archivo donde sale que el navegador no es compatible, hay de dar click derecho y abrirlo en otra pestaña para que si registre. Se debe especificar porque algunos lectores lo pueden tomar como que no funciona la base de datos.
- En la sección de construycamos con R, hay una notación de cambio las comillas sencillas por dobles, en la consola se copia y pega lo dado por el autor, si se cambia por comilla doble genera un error en la consola.

P-P Plot y Q-Q Plot.

- En la primera sección donde registra la consola dice escribe el cuartil a calcular, por más que un ingrese valores, sean pequeños, grandes, con decimales siempre arroja un error. Con lo que a mi respecta no entendí muy bien lo que debía hacer en esa sección.

Ejercicios propuestos.

En esta sección hay ejercicios muy interesantes para mirar y hacer, sería mejor donde ~~hubiera~~ hubiera un simulador donde realizar estos ejercicios ahí mismo, para rectificar que lo que se esté haciendo se vaya comprendiendo.

Diagrama de Caja y Bigotes - Boxplot

En la parte de actividades cuando dice genere se refiere o se entiende a que la persona pueda colocar una base de datos y se generen los resultados, cosa que no pasa ya que ya tiene unas bases de datos predeterminadas.

En conclusión es un libro bastante bueno, tiene muchas partes interactivas las cuales hacen que el aprendizaje se facilite, es un libro muy detallado explica como se debe realizar cada proceso, cada concepto, una persona con conocimientos previos se facilita su entendimiento.

Para el ejemplo #2, en el histograma con la variable edad, creo que sería conveniente agregar los intervalos de cada una de las distribuciones.

Texto #2 = P-P plot y Q-Q plot

No encuentre ninguna observación negativa.

Texto #3 Ejercicios propuestos.

En la actividad #5 sería interesante que dejaran, en apartado, en donde los estudiantes puedan realizar los histogramas de frecuencia y la ojiva.

Texto #4 (Box) Boxplot

En la actividad 7 después de la instrucción número 4, vuelve aparecer una instrucción con el número 3.

El resto de actividades me parecen que están bien, me gustaría que implementaran una opción para calcular los cuartiles y hallar los límites izquierdo y límite derecho y dejar que el estudiante realice la caja y los bigotes para un mejor entendimiento.

Apendice b Código fuente

```
if (!require("shiny")) install.packages("shiny")

if (!require("shinydashboard")) install.packages("shinydashboard")

if (!require("ggplot2")) install.packages("ggplot2")

if (!require("shinyjs")) install.packages("shinyjs")

if (!require("readxl")) install.packages("readxl")

if (!require("plotly")) install.packages("plotly")

if (!require("gridExtra")) install.packages("gridExtra")

if (!require("patchwork")) install.packages("patchwork")

if (!require("shinyAce")) install.packages("shinyAce")

if (!require("shinyjs")) install.packages("shinyjs")

if (!require("googleway")) install.packages("googleway")

if (!require("ggmap")) install.packages("ggmap")

if (!require("httr")) install.packages("httr")

if (!require("leaflet")) install.packages("leaflet")

if (!require("osrm")) install.packages("osrm")

if (!require("DT")) install.packages("DT")

if (!require("shinythemes")) install.packages("shinythemes")

#if (!require("")) install.packages("")

library(shiny)

library(shinydashboard)

library(ggplot2)
```

```
library(shinyjs)
library(readxl)
library(plotly)
library(gridExtra)
library(patchwork)
library(shinyAce)
library(shinyjs)
library(rsconnect)
library(googleway)
library(ggmap)
library(httr)
library(googlePolylines)
library(leaflet)
library(osrm)
library(DT)
library(shinythemes)

#library()

css <- "
#floating-buttons {
  position: fixed;
  bottom: 20px;
  right: 20px;
```

```
z-index: 1000;
}
#back_to_top {
  display: none; /* Ocultar hasta que se haga scroll */
}
"

ui <- fluidPage(
  useShinyjs(), # Inicializa shinyjs
  inlineCSS(css), # Añade el CSS directamente
  theme = shinytheme("cerulean"),
  #Hoja de presentación con imagen de fondo
  div(id = "landing_page",
    style = "background-image: url('portada.png'); height: 100vh; width: 100vw;
background-size: cover; background-position: center; position: relative;",
    div(class = "welcome-page",
      actionButton("start_btn", "Capítulo 1", class = "btn btn-primary btn-lg",
        style = "position: absolute; top: 60%; left: 50%; transform: translate(-50%,
-50%); padding: 10px 20px;")
    )
  ),
```

```
#PARA HACER ECUACIONES

tags$head(

  tags$script(src

=

"https://cdnjs.cloudflare.com/ajax/libs/mathjax/2.7.5/MathJax.js?config=TeX-MML-

AM_CHTML"),
```

```
tags$style(HTML("

.shiny-plot-output {

  max-width: 800px; /* Establece el ancho máximo de los gráficos */

  margin-left: auto;

  margin-right: auto;

}

")),
```

```
#Lector de voz
```

```
tags$script(HTML("

var utterance = new SpeechSynthesisUtterance();

var isPaused = false;

function updateTextToRead() {

  // Encuentra la pestaña activa dentro de la pestaña principal.
```

```
var activeTabContent = document.querySelector('.tab-pane.active .tab-pane.active');  
  
// Verifica si se ha encontrado contenido activo y actualiza el texto de la utterance.  
  
if (activeTabContent) {  
  
    utterance.text = activeTabContent.textContent || activeTabContent.innerText;  
  
} else {  
  
    // Si no hay subpestaña activa, simplemente toma el contenido de la pestaña principal.  
  
    var mainTabContent = document.querySelector('.tab-pane.active');  
  
    utterance.text = mainTabContent ? mainTabContent.textContent ||  
mainTabContent.innerText : "  
  
    }  
  
}  
  
  
window.readText = function() {  
  
    updateTextToRead();  
  
    if (!speechSynthesis.speaking) {  
  
        speechSynthesis.speak(utterance);  
  
    }  
  
};  
  
  
window.pauseResume = function() {  
  
    if (isPaused) {  
  
        speechSynthesis.resume();  
  
        isPaused = false;  
  
    }  
  
};
```

```
    } else {  
        speechSynthesis.pause();  
        isPaused = true;  
    }  
};  
  
window.stopReading = function() {  
    speechSynthesis.cancel();  
    isPaused = false;  
};  
  
// Asegúrate de llamar a updateTextToRead() cuando se cambie de pestaña.  
$('.nav-tabs a').on('shown.bs.tab', function(event) {  
    updateTextToRead();  
});  
)),  
  
#acaba lector  
  
#Tamaño de letra - botones  
  
tags$style(HTML("  
#text-size-controls {  
    position: absolute;  
    top: 30px;
```

```
    right: 30px;
  }
  ")),
  #Tamaño de letra fin
  ),

div(id = "text-size-controls",
  actionButton("increase_text", label = "+", class = "btn btn-primary btn-floating"),
  actionButton("decrease_text", label = "-", class = "btn btn-warning btn-floating")
),

div(style = "margin-right: 120px;",

tags$style("h3 { font-weight: bold; }"),
useShinyjs(),
tags$head(tags$style(".content-wrapper { background-color: #f5f5f5; }")),
tags$style("
.boxplot-box { fill: #FFA07A; }
.boxplot-median { stroke: #DC143C; }
.boxplot-whiskers { stroke: #6495ED; }
.boxplot-outliers { fill: #228B22; }
.boxplot-q1 { color: #FFA07A; }
.boxplot-median-label { color: #DC143C; }
```

```
.boxplot-q3 { color: #6495ED; }  
.boxplot-outlier { color: #228B22; }  
  
"),  
  
# Nueva página  
  
hidden(  
  
  div(id = "new_page",  
  
    h3("Instructivo de Navegación y Uso del Libro Interactivo"),  
  
    tags$ul(  
  
      h4(tags$b("Navegación Básica:")),  
  
      h5("En el contenido principal encontrarás una variedad de secciones o capítulos que  
abarcan diferentes temas. Para desplazarte entre ellos, utiliza los botones o enlaces que se  
encuentran claramente marcados en cada página. Estos enlaces están diseñados para facilitar una  
navegación fluida y coherente a través del material, permitiéndote avanzar, retroceder o saltar a  
secciones específicas según tu interés o necesidad de estudio."),  
  
      h4(tags$b("Funciones Especiales:")),  
  
      tags$ul(  
  
        tags$li(h5("Si necesitas un texto más grande o más pequeño para una lectura  
cómoda, simplemente haz uso de los botones con los signos '+' y '-'. Estos botones te permitirán  
personalizar el tamaño del texto a tu preferencia, garantizando así una experiencia de usuario más  
accesible y agradable. Los puedes encontrar en la parte superior derecha de la pantalla")),
```

tags\$li(h5("Para una experiencia más dinámica y accesible, especialmente si prefieres escuchar en lugar de leer, activa nuestro lector de voz. Con botones para reproducir, pausar y detener, puedes controlar la narración del contenido a tu ritmo, facilitando la comprensión y el aprendizaje en diferentes contextos y situaciones.Los puedes encontrar en la parte inferior derecha de la pantalla "))

),

h4(tags\$b("Consejos para una mejor experiencia:")),

tags\$ul(

tags\$li(h5("Tómate tu Tiempo: Este recurso está diseñado para que lo explores a tu propio ritmo. No hay prisa, así que siéntete libre de detenerte, reflexionar y digerir la información presentada. Esta flexibilidad es clave para una comprensión profunda y duradera de los temas tratados.")),

tags\$li(h5("Interactúa con los Contenidos: Te animamos a participar activamente en tu aprendizaje. Las funciones interactivas como gráficos, ejercicios y cuestionarios están aquí para enriquecer tu experiencia. A través de esta interacción, podrás obtener una comprensión más completa y práctica de los conceptos presentados.")),

tags\$li(h5("Utiliza los Recursos Adicionales: Para ampliar tu comprensión de los temas, te ofrecemos una variedad de enlaces externos y recursos adicionales. Estos materiales complementarios son una excelente manera de profundizar en los temas y explorar diferentes perspectivas y aplicaciones."))

)

```
),  
  h4(tags$b("¡Disfruta explorando y aprendiendo de una manera nueva y  
emocionante!")),  
  actionButton("go_main", "¡Vamos!",class = "btn btn-primary btn-floating")  
),  
# Contenido principal  
hidden(  
  div(id = "main_content",  
  
  titlePanel("Análisis de datos desde un enfoque exploratorio"),  
  
  navbarPage(  
    "Capítulo 1",  
    tabPanel("Introducción",  
      div(id = "introduction-content",  
        h3("Introducción"),  
        h4("Un analista principiante, en un informe descriptivo sobre datos  
demográficos de Colombia reporta que una ciudad capital de este país tiene en promedio 758.935  
habitantes; este tipo de estadística es muy común dado que el promedio es uno de los estadísticos  
de resumen más popular, no obstante aquí es conveniente detenerse a pensar lo representativo que  
ese valor numérico puede ser para la muestra de datos en consideración, en este caso un ciudadano
```

extranjero podría descartar la posibilidad que en Colombia exista un número importante de capitales con poblaciones inferiores a los doscientos mil habitantes o el efecto que tienen en el promedio la existencia de tres valores atípicos. La descripción numérica a través del promedio es concisa pero limitada, diferentes rasgos en la distribución de los datos como los mencionados deben ser analizados para dar una adecuada interpretación de este estadístico. "),

br(),

h4("A este respecto, es evidente la necesidad de contar con un criterio técnico para evaluar cuando una medida de tendencia central como el promedio es adecuada; en este caso, se requiere evaluar el comportamiento probabilístico de la variable, si la distribución de los datos evidencia simetría, usar un promedio es adecuado como medida de resumen al igual que la mediana y la moda, pero en caso contrario, el promedio se verá influenciado por la presencia de valores extremos, tal como se sucede en los datos de población de las capitales colombianas con Bogotá, Medellín, Cali y Barranquilla que son ciudades con poblaciones por encima del millón de habitantes. "),

br(),

h4("En el momento, la evaluación propuesta se basa en el uso de herramientas gráficas, es usual observar un histograma como el que se ilustra a continuación para darse una idea del comportamiento probabilístico que sigue una variable, pero hay otras alternativas que serán presentadas en el desarrollo de este capítulo. En la medida que el tema avance, en un capítulo posterior se presentarán recursos más formales como son las pruebas de hipótesis para validar un supuesto distribucional, estos criterios son más elaborados ya que suponen un entendimiento del fundamento teórico de estadística inferencial ([Hipervínculo a Prueba Chi cuadrado, Shapiro Wilk, Kolmogorov Smirnov](#)). "),

```
plotOutput("histogramPoblacion"),
```

```
h4("Es conveniente concluir diciendo que para el ejemplo en consideración, dada la asimetría a derecha que presentan los datos de población, un análisis descriptivo básico consistirá en reportar la mediana como medida de resumen la cual debe ser acompañada del reporte de una medida de dispersión apropiada como lo es el rango intercuartílico (Hipervínculo a glosario), de manera alternativa si se quiere ofrecer una descripción más detallada, el analista podría segmentar la muestra de ciudades en subgrupos más homogéneos, por ejemplo, se podría considerar tres subconjuntos de ciudades, Capitales grandes (más de un millón de habitantes), Ciudades intermedias y Capitales pequeñas (menos de 100.000 habitantes) y presentar un análisis descriptivo para cada uno de ellos. ")
```

```
)),
```

```
tabPanel("Validando un supuesto distribucional",
```

```
h3("¿Cómo validar un supuesto distribucional?"),
```

```
h4("Parte de las herramientas que se requieren para conducir un análisis de datos son las orientadas a validar un supuesto en cuanto a forma de la distribución de probabilidad de la que parecen provenir los datos. En algunas situaciones hay variables cuyos comportamientos han sido estudiados y se han postulado modelos ampliamente aceptados dada la calidad del ajuste de estos a diferentes muestras de datos, un ejemplo que ilustra muy bien este escenario es la variable Coeficiente Intelectual (Hipervínculo), para esta variable la distribución  $N(100,15)$  es tomada como el modelo de referencia a nivel mundial, por tanto para ella el analista debe simplemente comparar el comportamiento de su muestra de datos con este referente distribucional. "),
```

```
br(),
```

```

titlePanel("Distribución de Puntuaciones de CI"),
fluidRow(
  width = 8,
  plotOutput("interactiveHistogram"),
  sliderInput("bins", "Número de barras:", min = 1, max = 20, value = 15)
,
br(),

```

h4("No obstante, existen otras variables para las que no existe un modelo postulado para describir su comportamiento probabilístico, en estos casos, el trabajo del analista es encontrar cuál de las distribuciones de probabilidad existentes exhibe un mejor ajuste a los datos, claramente esta labor es demandante porque requiere tener un conocimiento del comportamiento de los diferentes modelos a disposición y evaluar la calidad del ajuste para los que se considere sean los más apropiados. Exploremos un escenario real de análisis ayudados por el siguiente applet, el objetivo es que basado en las características del histograma desplegado a derecha confirme si un modelo normal es el adecuado para describir la variable Porcentaje de grasa: "),

```

br(),
sidebarLayout(
  sidebarPanel(
    textAreaInput("dataInput",
      "Ingrese sus datos aquí (separados por comas):",
      value = "25.3, 29.3, 37.7, 32.8, 24.6, 26.5, 21.2, 28.4, 24, 28.7, 37.4,
30.7, 36.7, 28.2, 26.4, 37.1, 31.1, 43.1, 34.1, 26.7, 30.9, 30, 22.1, 24.4, 22.5, 24.9, 18.2, 27.3, 25.9,

```

28.3, 18.7, 22.4, 23.6, 26.8, 17.8, 27.4, 16.8, 26.1, 20.8, 22.7, 20.2, 20.3, 31.9, 22.9, 25.3, 17.3, 41, 25.8, 36.6, 27.8, 39.4, 36.4, 46.8, 40.5, 43, 39.5, 39.4, 24.8, 35, 25.3, 42.5, 27.8, 35.8, 39.2, 38.3, 34.1, 39.9, 32.5, 29.9, 32, 23, 28.5, 26.4, 33.9, 29.4, 29.4, 21, 22.4, 25.3, 20.2, 22.5, 23.4, 29.5, 21.9, 19.2, 28.6, 23.4, 23.2, 18.1, 31.4, 22.9, 25.6"),

```
selectInput("binsInput",
            "Seleccione el número de clases (bins) para el histograma:",
            choices = 1:14,
            selected = 5)
```

```
),
```

```
mainPanel(
  plotOutput("histPlot")
```

```
)
```

```
),
```

```
br(),
```

h4("Dejando los datos tal como fueron cargados en el applet, cambie el número de clases (rectángulos) y visualice el efecto que tiene esta variación en el aspecto del gráfico. Puede ahora cambiar la muestra dada al digitar o copiar otros datos y variar el número de clases según las opciones de la lista para analizar el comportamiento de los histogramas. "),

```
br(),
```

h4("Como conclusión a esta actividad se espera que el lector vaya recordando o familiarizándose con aspectos como el efecto del tamaño de muestra en la imagen de la distribución de datos que proyecta el histograma, muestras pequeñas, $n < 20$ (Hipervínculo), pueden ser problemáticas de analizar gráficamente, también que el número de intervalos o rectángulos que

contiene un histograma es un parámetro del gráfico que puede ser manipulado por el analista y en ese caso será mejor contar con criterios técnicos para decidir su valor óptimo (Hervinculo) "),

br()),

h4("Un ámbito donde se requiere evaluar comportamientos distribucionales con rigor es como paso intermedio de un análisis estadístico inferencial, la validez del uso de ciertas técnicas estadísticas requiere del comportamiento normal de las variables, un primer encuentro con este tipo de análisis se dará al usar las pruebas sobre la media y posteriormente en la elaboración de pronósticos a partir de un modelo lineal. Claramente la exigencia de supuestos distribucionales como la normalidad requiere de contar con tamaños de muestra apropiados, muestras pequeñas revisten dificultad porque recursos como los gráficos pueden no mostrar los rasgos característicos de un modelo de manera clara. Hoy día la disponibilidad de grandes muestras de datos es cada vez más frecuente pero no siempre corresponden a muestras aleatorias, la presencia de datos no estructurados supone un ámbito de análisis que demanda adecuar los análisis que aquí se ilustrarán. "),

h3("1. Elementos teóricos para contrastar un supuesto de Normalidad"),

br()),

h4("Empecemos por revisar este supuesto distribucional particular, si bien la distribución normal junto con el modelo binomial son los dos modelos más estudiados cuando se estudia el apartado de probabilidad, es válido preguntarse, ¿qué tan frecuente observamos variables en nuestro entorno que sigan un comportamiento normal?, A continuación, les ofrecemos algunos ejemplos para que usted mismo vaya formándose una respuesta a ese interrogante, usaremos como recurso para analizar el comportamiento de los datos la visualización de histogramas, en las

secciones siguientes se profundizara en la construcción e interpretación de este tipo de gráfico estadístico. "),

```
br(),
```

```
h3("Ejemplo 1:"),
```

```
h4("Distribución de la Edad de las ganadoras del premio Oscar desde 1970. "),
```

```
br(),
```

```
plotOutput("histPlot2"),
```

```
h4("En primera instancia, puede pensarse en abordar ese interrogante basado en nuestra percepción personal y entonces las respuestas pueden ser variadas y subjetivas, si queremos ser un poco más formales debemos empezar por considerar si los valores obtenidos al medir la variable en una población exhiben dos características en su distribución acorde al modelo normal: ",
```

```
strong("una alta concentración en la región central (Tendencia central) y una concentración de datos cada vez menor pero simétrica conforme nos movemos a izquierda o derecha de ese centro"),
```

```
" tal como se ilustra en esta gráfica: "),
```

```
br(),
```

```
imageOutput("Imagen2"),
```

```
br(),
```

```
h4("Note que para los datos de edades de ganadoras al Oscar la concentración de datos no se da en la región central sino en el extremo izquierdo de la escala, la distribución no exhibe rasgos de simetría, sino que es sesgada a derecha, comportamiento caracterizado por la presencia de valores atípicos a derecha como lo son aquí las edades contenidas en el intervalo entre
```

60 y 80 años. Si superpusiéramos una curva a mano alzada sobre este histograma, veríamos lo que se denomina, cola a derecha. "),

```
br(),
plotOutput("histPlotWithDensity"),
```

```
h3("Ejemplo 2:"),
```

```
br(),
```

```
h4("Esta fuente de datos ",
```

```
tags$a(href = "https://sofifa.com", target = "_blank", "SoFIFA.com"),
```

" nos muestra varias variables de interés para los jugadores del juego FIFA o simples aficionados al futbol en general. Examine el histograma para cada una de ellas y concluya cuál o cuáles evidencian simetría en su distribución, por lo que podrían ser modeladas a través de una normal."),

```
br(),
```

```
br(),
```

h4("Variables de interés para los jugadores del juego FIFA o simples aficionados al futbol"),

```
selectInput("variable",
```

```
"Selecciona una variable:",
```

```
choices = c("EDAD", "CALIFICACION GENERAL",
```

```
"POTENCIAL", "ALTURA",
```

```
"PESO", "VALOR(M)", "SALARIO (K)", "ESTADÍSTICAS
```

```
TOTALES"),
```

```

        selected = "EDAD"
    ),

    # Gráfico del histograma
    plotOutput("histogram"),

    br(),

    h4("Para más ejemplos puede consultar la siguiente referencia:"),

    # Hipervínculo
    tags$a(href = "https://studiousguy.com/real-life-examples-normal-
distribution/", target="_blank", "Otros ejemplos"),

    br(),

    h4("Como primer recurso para validar un supuesto distribucional estamos
considerando la inspección visual de gráficas estadísticas, principalmente se suele utilizar el
Histograma pero existen otras opciones como el gráfico de Caja y bigotes o Boxplot, los gráficos
de probabilidad como el PP plot o QQ plot y, aunque no se desarrollaran en este capítulo, también
puede usarse gráficos como las Ojivas, Gráficos sin tendencia y gráfico de Tallo y hojas, cada uno
de los mencionados recursos tiene un potencial diferente en cuanto a que visibiliza ciertos rasgos
de la distribución de una mejor manera; en la práctica, el analista puede hacer uso de más de un
recurso para tomar una decisión acertada, pero al igual que con todo recurso estadístico es bueno
asumir su uso con sentido crítico en cuanto a reconocer las posibilidades y limitaciones que ofrecen
en torno a un objetivo de análisis particular. ")

    )),

    tabPanel("Pruebas gráficas",

```

```
tabsetPanel(  
  id = "subTabs",  
  tabPanel("Diagrama de caja y bigotes",  
    h3("¿Alguna vez has viajado por carretera y notado cómo cambia el  
kilometraje? "),  
    #Mapa  
  
    leafletOutput("map"),  
  
    #Termina mapa  
  
    h4("Imagina que inicias un viaje por carretera desde Bucaramanga hasta  
Bogotá, pasando por Barbosa podrías preguntarte, ¿Cuánto falta para llegar a Bogotá?, con toda  
seguridad la respuesta está en una señal de tránsito de destino que lucirá como la imagen a  
continuación: "),  
    #imagen  
    DTOutput("myTable"),  
  
    h4("En este caso las señales informan del kilometraje que falta para llegar  
a un destino pero de manera complementaria se sabe la distancia recorrida. Queremos que ahora  
usted relacione el papel informativo que cumplen este tipo de señales en la vía con el propósito  
que cumplen los denominados estadísticos de orden. "),
```

```
br(),  
h3("Medidas de ubicación: recursos estadísticos para describir una  
distribución de datos"),  
h4("Las medidas de posición son valores en la escala de la variable que  
permiten ubicar un valor en relación con el resto de unidades que integran la muestra, los más  
comunes son: "),  
tags$ul(  
  tags$h4(tags$h4(tags$span("Mediana: ", style = "color: #DC143C;")),  
"punto en la escala de la variable que divide el conjunto de datos en dos. Nos permite hablar en  
términos de 50% Notación: Q2" ),  
  div(tags$img(id="ImagenM", src="Mediana.png", class="custom-  
img", width=300, height=130)),  
  tags$h4(tags$h4(tags$span("Cuartiles: ", style = "color: #6495ED;")),  
"puntos en la escala de la variable que divide el conjunto de datos  
en 4. Nos permite hablar en términos del 25%, 50% y 75%. Notación: Q1,Q2,Q3."  
),  
br(),  
h3("Gráfico de caja y bigotes"),
```

h4("Un boxplot es una representación gráfica que resume la distribución de un conjunto de datos a partir de los cuartiles. Es útil para visualizar medidas de tendencia central, dispersión y la presencia de valores atípicos."),

h4("Los elementos clave de un boxplot son:"),

tags\$ul(

tags\$h4(tags\$span("Caja: ", style = "color: #9F82F7;"), "representa el 50% de la muestra, el ancho de la caja corresponde al rango intercuartílico que sirve de medida de dispersión y se calcula haciendo $Q3-Q1$. La caja permite ubicar el subconjunto de datos 'centrales' o más representativos de la muestra."),

tags\$h4(tags\$span("Línea divisoria en la caja: ", style = "color: #DC143C;"), "Está ubicada en la posición de la mediana ($Q2$), divide la parte central de la muestra en dos subconjuntos de datos que incluyen un 25% de la muestra cada uno."),

tags\$h4(tags\$span("Bigotes: ", style = "color: #6495ED;"), "líneas que se extienden desde la caja hasta los valores mínimos y máximos de la muestra que no sobrepasen una distancia de 1.5 veces el rango intercuartílico, por lo anterior habrá un bigote a izquierda y otro a derecha. Su amplitud informa sobre la dispersión en los datos en la parte de la muestra que representan."),

tags\$h4(tags\$span("Valores atípicos: ", style = "color: #228B22;"), "puntos fuera de los bigotes que representan valores atípicos en el conjunto de datos, valores que se ubiquen a más de 3 veces el rango intercuartílico son catalogados de esta forma.")

),

h4("Tip para el analísta principiante:"),

```
h5("No todo gráfico de caja y bigotes exhibirá los elementos anteriores,  
más adelante propondremos un ejercicio para ilustrar este comentario."),
```

```
br(),
```

```
br(),
```

```
h3("Actividad"),
```

```
h4("El siguiente boxplot se construyó a partir de los datos de Población  
de algunas ciudades de Colombia (factor de escala x10000)'),
```

```
tags$ul(tags$h4(tags$span("1. Ubica los elementos de este gráfico")),
```

```
tags$h4(tags$span("2. Observa qué sucede con el gráfico cuando  
excluyes o incluyes los diferentes departamentos ")),
```

```
tags$h4(tags$span(("3. Bogotá ha sido incluido separado de  
cundinamarca por su tamaño poblacional, explora el efecto de incluir o excluir a esta ciudad capital  
que claramente es un valor extremo en Colombia."))),
```

```
)),
```

```
br(),
```

```
sidebarLayout(  
  sidebarPanel(  
    tags$h4("Selecciona los departamentos:"),  
    splitLayout(  
      cellWidths = c("50%", "50%"),  
      checkboxGroupInput(  
        "deptosSeleccionados",
```

```
    ""  
    choices = NULL, # Se actualizará con los nombres de los  
departamentos  
    selected = NULL  
  ),  
  checkboxGroupInput(  
    "deptosSeleccionados2",  
    ""  
    choices = NULL, # Se actualizará con los nombres de los  
departamentos  
    selected = NULL  
  )  
),  
tableOutput("cuartilesTabla") # Agregar la tabla de cuartiles aquí  
)  
mainPanel(  
  plotOutput("boxplotPoblacion"),  
  div(style = "width: 100%; overflow-x: auto;", # Ajustar el tamaño de  
la tabla al ancho del gráfico  
  tableOutput("datosTranspuestos") # Establecer el ancho al 100%  
del contenedor  
)  
)
```

),

h3("Interpretación y usos comunes del Boxplot"),

h4("La lectura de un boxplot se basa en ubicar e interpretar sus elementos: "),

tags\$ul(

tags\$h4("- Visualización de la simetría o asimetría de los datos: si la mediana divide la caja en dos partes iguales y los bigotes tienen una amplitud similar, hay evidencia de simetría, de lo contrario se debe examinar la dirección del sesgo (lado de la caja más ancho o bigote más amplio) para determinar la dirección en que se presenta la asimetría que puede ser a derecha o izquierda"),

tags\$h4("- Evaluación de la dispersión de los datos: a partir de la observación del rango intercuartilico (ancho de caja) y amplitud de bigotes."),

tags\$h4("- Identificación de valores atípicos en el conjunto de datos: presencia de símbolos o caracteres en los extremos del gráfico, cada software tiene una convención diferente para representarlos."),

tags\$h4("Comparación de distribuciones para diferentes grupos de datos: este gráfico permite presentar la comparación de la forma como se distribuyen los datos para una misma variable en dos o más poblaciones o subgrupos formados a partir de una variable

categoría, este análisis resulta fácil pues depende de la ubicación de las cajas y la comparación de los demás elementos que contiene el gráfico"),

```
),
```

```
h3("Tip para el analista principiante:"),
```

```
h4("En el caso de querer representar en un mismo gráfico dos variables se debe tener cuidado por el efecto que puede tener la diferencia de escala de medida en el despliegue del gráfico. Piense, ¿qué pasaría si presentamos la edad y los salarios en un mismo gráfico? "),
```

```
br(),
```

```
h3("Actividades"),
```

```
h3("Actividad 1: Dibujando 'manualmente' un gráfico de caja y bigotes:"),
```

```
h4("Siga las siguientes instrucciones y en una hoja de papel vaya dibujando el gráfico, no olvide definir una escala apropiada. "),
```

```
h3("Instrucciones:"),
```

```
h4("1. Introduzca los datos o genere una muestra aleatoria"),
```

```
textInput("user_nums", "Introduzca los números separados por comas:", value = "", width = "100%"),
```

```
actionButton("generate_nums", "Generar muestra aleatoria"),
```

```
actionButton("insert_data", "Insertar datos"), # Botón para insertar los datos
```

```
h4("2. Calcular cuartiles"),  
h5("Puedes hacerlo manualmente, utilizando excel, u otros softwares  
para calcularlos; sin embargo, debes tener en cuenta que puedes obtener resultados 'diferentes' si  
observamos las cifras decimales que considera cada uno. En esta oportunidad, lo calcularemos  
usando r así:"),
```

```
br(),  
shinyAce::aceEditor("code3",  
value = '# Crea algunos datos  
  
datos <- c("escribe tu muestra ordenada aquí")  
  
# Calcula los cuartiles  
  
cuartiles <- quantile(datos, probs=c(0.25, 0.5, 0.75))  
  
# Imprime los cuartiles  
  
print(cuartiles)',  
height = "100px"),  
actionButton("run_code3", "Ejecutar código"),  
uiOutput("code3_result"),  
uiOutput("quartiles_button"),  
tags$p("En el siguiente ",  
tags$a(href="https://www.calcvio.com/cuartiles", "enlace"),  
", pega tu muestra de datos y observa las diferentes formas de  
calcular los cuartiles. En el caso de R, ¿qué método utiliza?"),
```

```
h4("3. Ver datos divididos en cuartiles"),
uiOutput("quartiles_table"),
h4("4. Calcule el rango intercuartílico (RI)"),
h4("RI = Q3 - Q1 "),
h4("3. Dibuje la caja"),
actionButton("show_boxplot2", "Mostrar Gráfico"),
plotOutput("boxplotp", height = "200px"),
br(),

h4("5. Calcule los límites para los bigotes"),
h4("Límite Izquierdo = Q1 - 1,5*RI "),
h4("Límite Derecho = Q3 + 1,5*RI "),
br(),

h4("6. Dibuje los bigotes"),
actionButton("show_boxplot_whiskers", "Mostrar Gráfico con
Bigotes"), # Botón para mostrar el gráfico
plotOutput("boxplot_whiskers", height = "200px"),

h4("7. Calcule los límites para separar los valores extremos de los
inusuales"),
```

```
h4("Límite Izquierdo =  $Q1 - 3*RI$  "),  
h4("Límite Derecho =  $Q3 + 3*RI$  "),  
actionButton("reset_button", "Reiniciar"),
```

```
br(),
```

```
h3("Actividad 2:"),
```

h4("A continuación aparecerá una variable a analizar y una muestra de datos (para cambiar de variable de clic en 'Mostrar otro problema'), el objetivo de esta actividad es comparar el gráfico de caja y bigotes elaborado manualmente con el gráfico que produce R"),

```
br(),
```

h4("1. Dibuje manualmente el gráfico de caja y bigotes y describa las características que presentan"),

h4("2. Compare el gráfico anterior con el que se despliega a través del comando boxplot en R para ello solo debe dar clic en el botón 'Mostrar boxplot'"),

```
h4("Problema:"),
```

```
verbatimTextOutput("problema"),
```

```
h3("Datos"),
```

```
verbatimTextOutput("datos"),
```

```
actionButton("nuevo_problema", "Mostrar otro problema"),
```

```
actionButton("mostrar_boxplot", "Mostrar Boxplot"),
```

```
    uiOutput("plot_ui"),  
  
    br(),  
  
    h3("Actividad 3: Ubicando estadísticos para medir tendencia central"),  
    h4("3.1. El objetivo de esta actividad es 'adivinar' la ubicación de la  
media y la mediana para una muestra de datos a partir de tres tipos de gráficos, el gráfico de  
probabilidad normal y el histograma serán tratados con detalle en las próximas secciones por  
cuanto si no está familiarizado con ellos, puede omitir su uso."),  
  
    h5("1. Digite el número de datos que desea que tenga la muestra  
aleatoria y a continuación presione el botón 'generar datos' para visualizarlos en el gráfico"),  
  
    h5("2. Seleccione el tipo de gráfico del que desee adivinar la media y la  
mediana"),  
  
    h5("3. Seleccione la casilla 'mostrar deslizadores' y moviendo las líneas  
intente adivinar donde cree que se encuentran la media y la mediana de la distribución" ),  
  
    h5("4. De clic en el botón 'mostrar estadísticos reales', ¿qué tan cerca  
estuvo su conjetura?"),  
  
    h5("5. Realice la misma actividad para muestras diferentes, compare y  
concluya"),  
  
    br(),  
  
    sidebarLayout(  
      sidebarPanel(  
        numericInput("numDataPoints", "Número de datos:", 100),  
        actionButton("generateData", "Generar datos"),
```

```

        selectInput("graphType", "Tipo de gráfico:", c("Dotplot",
"Histograma", "Gráfico de probabilidad normal")),

        checkboxInput("showBoxplot", "Mostrar Boxplot", value = FALSE),
        checkboxInput("showSliders", "Mostrar deslizadores", value =
FALSE),

        conditionalPanel(
            condition = "input.showSliders",
            sliderInput("meanSlider", "Adivina la media:", -3, 3, 0, step = 0.01),
            sliderInput("medianSlider", "Adivina la mediana:", -3, 3, 0, step =
0.01)

        ),

        actionButton("showRealLines", "Mostrar estadísticos reales")

    ),

    mainPanel(
        plotOutput("mainPlot")
    )

),

br(),

h4("3.2 Si graficamos la media de la muestra de datos, ¿ésta siempre va
a estar ubicada dentro de la caja?"),

```

```
h4("Use el siguiente aplicativo para comprobarlo"),
sidebarLayout(
  sidebarPanel(
    textInput("data_input", "Ingresa los datos separados por comas", ""),
    actionButton("random_button", "Generar datos aleatorios"),
    checkboxInput("show_mean", "Mostrar media", FALSE)
  ),
  mainPanel(
    plotOutput("boxplot_output1")
  )
),
```

h3("Actividad 4: ¿Qué características de la distribución de los datos son visibles a través de un gráfico de caja y bigotes?"),

```
h4("Indique cuál de los siguientes gráficos exhibe rasgos de simetría"),
h4(""),
br(),
sidebarLayout(
  sidebarPanel(
    radioButtons("choice", "Elige un gráfico:",
      choices = c("Gráfico 1", "Gráfico 2", "Gráfico 3", "Gráfico
4")),
```

```
        actionButton("submit", "Enviar")
    ),

    mainPanel(
        column(6,
            plotOutput("boxplot1", height = "300px"),
            plotOutput("boxplot3", height = "300px")
        ),
        column(6,
            plotOutput("boxplot2", height = "300px"),
            plotOutput("boxplot4", height = "300px")
        ),
        textOutput("result")
    )
),
br(),
h3("Actividad 5: Exploremos cómo cambian las características de un
grafico de caja y bigotes al manipular la muestra de datos"),

h4("5.1 Haga uso del escenario dado para contestar las siguientes
preguntas, recuerda que puedes editar la muestra de datos:"),
tags$ul(
```

tags\$li("¿Es posible construir un gráfico que no tenga bigote izquierdo ?"),

tags\$li("¿Es posible construir un gráfico de caja y bigotes en el que no se vea la caja ?"),

),

fluidPage(

titlePanel(""),

sidebarLayout(

sidebarPanel(

actionButton("generate_data_colored", "Generar datos"),

tags\$hr(),

textAreaInput("data_input_colored", "Datos (separados por comas):", rows = 5),

actionButton("generate_plot_colored", "Generar gráfico")

),

mainPanel(

plotOutput("boxplot_colored"),

verbatimTextOutput("boxplot_info_colored")

)

)

),

h4("5.2. Observe la muestra de datos, el gráfico de caja y bigotes y el gráfico de puntos para tener más información sobre cómo se distribuyen estos datos"),

```

        sidebarLayout(
          sidebarPanel(
            textAreaInput("user_data", "Ingresa tus datos aquí:",
"0,0,1,1,1,2,3,3,4,9", rows = 10),
            helpText("Nota: Los datos deben estar separados por comas. Por
ejemplo:"),
            verbatimTextOutput("example")
          ),
          mainPanel(
            plotOutput("plot")
          )
        ),
        h4("5.3 ¿Si se desea agregar el 11avo dato de la muestra a la muestra
con que se hizo el anterior gráfico, ¿basta con ubicarlo en la zona de valores extremos de la
derecha?. Recuerda que puedes editar la muestra que aparece en la parte izquierda del applet."),
        br(),
        h4("5.4 ¿Cómo lucirá el gráfico si retiramos del estudio el valor
extremo? "),
        tags$sul(
          tags$li("Se queda igual, sólo se eliminaría el punto a la derecha ")

```

```
tags$li("Cambiará considerablemente, ahora se elimina el sesgo a  
derecha"),
```

```
tags$li("El bigote de la derecha ahora es más corto"),
```

```
tags$li("La distribución sigue siendo asimétrica a derecha pero cambia  
la ubicación de la mediana")
```

```
),
```

```
h3("Actividad 6: Comparando gráficos producidos por diferentes  
softwares"),
```

```
h4("Observe el siguiente boxplot"),
```

```
plotOutput("boxplotejem", height = "300px", width = "300px"),
```

```
br(),
```

```
h4("Los datos utilizados para realizar el grafico fueron 0, 0, 1, 1, 1, 2,  
3, 3, 4, 9, dirijase a excel y realice el diagrama de caja y bigotes con dichos datos"),
```

```
tags$ul(  
  tags$li("Ahora, dirijase al siguiente", tags$a(href =
```

```
"http://www.estat.me/estat/eStat/", "enlace"), ", alli realice el gráfico de caja y bigotes de los datos,  
compare los tres gráficos construidos"),
```

```
tags$li("¿Cómo explicar que, para una misma muestra de datos, los  
gráficos producidos por softwares diferentes pueden no ser iguales?")
```

```
),
```

```
br(),
```

```

    )

),
  tabPanel("Histograma",
    h3("Histograma"),
    h4("En lugar de iniciar por la definición, vamos a revisar uno a uno los
pasos para realizar un histograma, se debe aclarar que este es un algoritmo pero no es el único que
hay para construir este gráfico. Realizaremos el histograma de forma manual y luego
aprenderemos a realizar el código en R para realizar histogramas. "),
    br(),
    h4("En el", tags$a(href =
"https://docs.google.com/spreadsheets/d/1sXxP3DAO8EsqgC_meQg9_3C7evPAv_a2/edit?usp=
sharing&ouid=104136984824170975752&rtpof=true&sd=true", "enlace"), ", descargue la base
de datos EdadesPersonas y úsela para realizar la siguiente actividad:"),

    h3("Algoritmo para construcción de un histograma"),
    strong("PASO 1:"),

```

```

h4("Determinar el rango de los datos: RANGO es igual al dato mayor
menos el dato menor; \\(Rango = Máx - Mín\\)"),

```

```

br(),

```

```

textInput("maxInput", "Ingresa el dato mayor aquí:"),

```

```

textInput("minInput", "Ingresa el dato menor aquí:"),

```

```

actionButton("calcRange", "Calcular RANGO"),

```

```

br(),

```

```

verbatimTextOutput("rangeOutput"),

```

```

strong("PASO 2:"),

```

h4("Para tener un criterio objetivo sobre el número de clases o barras que tendrá el histograma existen varios criterios, sin embargo no todos conducen a un mismo valor. Algunos autores recomiendan de cinco a quince clases dependiendo de cómo estén distribuidos los datos y cuántos sean. Un criterio usado frecuentemente es que el número de clases debe ser aproximadamente la raíz cuadrada del número de datos, por ejemplo, la raíz cuadrada de 66 (la cantidad de edades en la muestra) es 8.12 por lo que se deberían seleccionar ocho clases. "),

```

tags$h4("Consulta qué otros criterios hay para obtener el número de
barras del histograma y qué criterios se pueden utilizar para decidir cuál es el más apropiado",
style = "font-size: 18px;"),

```

```

br(),

```

```

strong("PASO 3:"),

```

```

p("Establecer la longitud de clase o Incremento: es igual al rango entre el
número de clases."),

```

```

    br(),
    textInput("rangeInput", "Ingresa el rango aquí:"),
    textInput("classInput", "Ingresa el número de clases aquí:"),
    actionButton("calcClassLength", "Calcular Longitud de Clase"),
    br(),
    verbatimTextOutput("classLengthOutput"),

    strong("PASO 4:"),
    h4("Construir los intervalos de clases: Los intervalos resultan de dividir
el rango de los datos en relación al resultado del PASO 2 en intervalos iguales."),
    h4("Busca el dato menor de la muestra y a ese número le vas a sumar la
longitud de clase, este sería el límite izquierdo del primer intervalo o clase. Para el ejemplo, el
número más pequeño es 10 y la amplitud de clase es 5.5, con lo anterior, el primer intervalo será:
\\([10 – 15.5]\\), el segundo \\((15.5, 21]\\), y así sucesivamente."),
    tags$table(
      tags$thead(
        tags$str(
          tags$th("Intervalo de clase"),
        )
      ),
      tags$tbody(
        tags$str(
          tags$td("10 - 15,5"),

```

```
),  
tags$tr(  
  tags$td("15,5 - 21"),  
),  
tags$tr(  
  tags$td("21 - 26,5"),  
),  
tags$tr(  
  tags$td("26,5 - 32"),  
),  
tags$tr(  
  tags$td("32 - 37,5"),  
),  
tags$tr(  
  tags$td("37,5 - 43"),  
),  
tags$tr(  
  tags$td("48,5 - 54"),  
)  
)  
),  
br(),  
strong("PASO 5:"),
```

h4("Graficar el histograma: Sobre un eje horizontal se dibujará una barra por cada uno de los intervalos anteriormente hallados, la altura de los rectángulos será determinada por la frecuencia que se halla al contar el número de datos de la muestra que pertenece al intervalo dado. Para el ejemplo hay 10 datos que pertenecen al intervalo $[10 - 15.5]$ "),

br(),

actionButton("hist_button", "Mostrar histograma"),

plotOutput("histPlotu"),

h3("Construyamos el histograma con R"),

br(),

h4("R utiliza la función $hist()$ para crear un histograma a partir de una muestra de datos. A continuación, se explica cómo funciona este comando: "),

tags\$ul(

tags\$li("La función $hist()$ toma como argumento principal una muestra de datos que se utilizará para generar el histograma. A este objeto se le llamará muestra y es el principal argumento de este comando. "),

tags\$li("R organiza los datos en intervalos o bins para crear las barras del histograma. Por defecto, R utiliza el método de Sturges para determinar el número de bins. "),

tags\$li("R calcula la frecuencia de los datos en cada bin, es decir, cuenta cuántos datos se ubican dentro de los límites de cada intervalo. "),

tags\$li("Luego, R dibuja las barras del histograma, recuerde que la altura de cada barra representa la frecuencia de los datos en el respectivo intervalo."),

tags\$li("La función $hist()$ ofrece varios argumentos adicionales para personalizar el histograma, como el título ($main$), etiquetas de los ejes ($xlab$) y

\\(ylab\\)), colores, leyendas, entre otros. Puedes consultar la documentación de la función hist() para obtener más información sobre estas opciones. ")

```
),  
br(),  
h4("Realicemos ahora la construcción del histograma con R"),  
br(),  
h5("1. Digite el siguiente código en la consola para generar el  
histograma"),  
  
h5("datos <- read_excel('EdadesPersonas.xlsx', sheet = 1)"),  
h5("hist(datos$Edades, col = 'blue', main = 'Histograma de Edades', xlab  
= 'Edades')"),  
  
h4("*Tip para el programador principiante: Cambie las comillas sencillas  
( ) por dobles"),  
  
br(),  
fluidRow(  
  column(width = 6,  
    h3("Generar histograma"),  
  
    shinyAce::aceEditor("code", value = "", height = "300px"),  
  
    br(),  
  
    actionButton("run_code", "Ejecutar código")  
  ),  
)
```

```

column(width = 6,
      plotOutput("histogram3")
    )
),
br(),
h4("¿Qué diferencia hay entre el histograma realizado manualmente y el
realizado por R?¿A qué se debe esta diferencia?"),

```

```

br(),
h4("Un histograma es un resumen gráfico que describe como se presenta
la variación en un conjunto de datos. Cada gráfico permite visualizar características particulares
de la distribución de los datos, en el caso del histograma éste nos permite evidenciar de manera
directa si hay simetría, en caso de que ese no sea el caso hablaremos de una distribución asimétrica
o sesgada y también nos indicara en que dirección se da esta condición (derecha o izquierda) . "),

```

```

h3("¿Cómo interpretar un histograma?"),
h4("Las características de variación más comúnmente analizadas en una
distribución de datos son: "),

```

```

tags$ul(
  tags$li("Presencia de simetría"),
  tags$li("Número de máximos o picos"),

```

```
tags$li("Forma de la curva que mejor inscribe a los datos: apuntada
(baja dispersión) o aplanada (alta dispersión)'),
```

```
tags$li("Si hay un comportamiento asimétrico indicar la dirección del
sesgo (derecha o izquierda)'),
```

```
tags$li("Vacíos en la distribución sugieren la posibilidad de presencia
de valores extremos pero hay que verificar según el contexto si realmente son valores extremos o
es una jugada del azar en la que éste no nos arroja valores en un intervalo dado de la escala de
variable")
```

```
),
```

```
tags$p(
```

```
tags$b("Tip para el analista:"),
```

```
" en contraposición al comportamiento simétrico es bueno tener claro el
panorama de opciones a que nos enfrentamos, estos son los comportamientos más típicos:"
```

```
),
```

```
sidebarLayout(
```

```
sidebarPanel(
```

```
selectInput("tipo", "Selecciona el tipo de comportamiento:",
```

```
choices = c("Simétrico", "Sesgado a la derecha",
```

```
"Sesgado a la izquierda", "Forma de J",
```

```
"J invertida", "Forma de U", "Bimodal",
```

```
"Multimodal"))
```

```
),
```

```
mainPanel(  
  plotOutput("distPlot")  
)  
,  
h3("Actividad 1: Histograma obtenido en estat vs Histograma obtenido  
en R"),  
h4("Compare los siguientes gráficos y explique las diferencias  
encontradas"),  
imageOutput("Imagen8"),  
h4("Nota para el analísta principiante: a pesar que softwares diferentes  
producen gráficos con diferentes características, note que los rasgos importantes de la distribución  
de los datos son visibles y esta es la tranquilidad que necesita un analista. En el ejemplo vemos  
que la asimetría (cola en la distribución a derecha) es evidente en cualquiera de los dos gráficos."),  
br(),  
br(),  
br(),  
h3("Actividad 2"),  
h4("Usando los datos de la siguiente tabla, recomendamos dibujar  
manualmente los gráficos tal como se indican a continuación, hacerlo de esta manera le brinda  
más conocimiento de los elementos y detalles que contiene el gráfico. "),
```

h4("Tabla 1: Distribución de Frecuencias- Edad de personas que asisten
a un cine "),

br(),

#TABLA

tags\$head(

tags\$style(HTML("

table {

width: 100%;

text-align: center;

}

table thead tr {

background-color: #9F82F7;

}

table tbody tr:nth-child(even) {

background-color: #DBF7FA;

}

"))

),

tags\$table(

tags\$thead(

tags\$str(

```
tags$th("Intervalo de clase"),
tags$th("Marca de clase*"),
tags$th("Frecuencia"),
tags$th("Frecuencia Relativa**"),
tags$th("Frecuencia Relativa Acumulada")

)

),
tags$body(
tags$tr(
tags$td("10 - 15,5"),
tags$td("12,75"),
tags$td("10"),
tags$td("0,15"),
tags$td("0,15")
),
tags$tr(
tags$td("15,5 - 21"),
tags$td("18,25"),
tags$td("18"),
tags$td("0,27"),
tags$td("0,42")
),
```

```
tags$tr(  
  tags$td("21 - 26,5"),  
  tags$td("23,75"),  
  tags$td("9"),  
  tags$td("0,14"),  
  tags$td("0,56")  
)
```

```
tags$tr(  
  tags$td("26,5 - 32"),  
  tags$td("29,25"),  
  tags$td("7"),  
  tags$td("0,11"),  
  tags$td("0,67")  
)
```

```
tags$tr(  
  tags$td("32 - 37,5"),  
  tags$td("34,75"),  
  tags$td("10"),  
  tags$td("0,15"),  
  tags$td("0,82")  
)
```

```
tags$tr(  
  tags$td("37,5 - 43"),
```

```
tags$std("40,25"),
tags$std("7"),
tags$std("0,11"),
tags$std("0,92")

),

tags$str(
tags$std("48,5 - 54"),
tags$std("51,25"),
tags$std("2"),
tags$std("0,03"),
tags$std("1")

)

)

),

br(),

tags$table(
tags$thead(
tags$str(
tags$th("Histograma de frecuencias"),
```

```
        tags$th("Histograma de Frecuencia Relativas")
    )
),
tags$body(
    tags$tr(
        tags$td("Marca de clase vs Frecuencia"),
        tags$td("Marca de clase vs Frecuencia relativa")
    )
)
),
br(),
tags$table(
    tags$thead(
        tags$tr(
            tags$th("Histograma de densidad"),
            tags$th("Histograma de Frecuencia Relativas Acumuladas")
        )
    ),
    tags$body(
        tags$tr(
            tags$td("Marca de clase vs Densidad"),
            tags$td("Marca de clase vs Frecuencia relativa acumulada")
        )
    )
)
```

```

    )
  ),
  h4("Ahora, compare sus gráficos con los obtenidos al dar clic en el botón
'Mostrar Histogramas'"),

```

```

#Botón mostrar histogramas ----
strong("Histogramas:"),
br(),
actionButton("showHistograms", "Mostrar Histogramas"),
br(),
conditionalPanel(
  condition = "input.showHistograms > 0",
  plotOutput("histograma1"),
  plotOutput("histograma2"),
  plotOutput("histograma3"),
  plotOutput("histograma4")
),

#Aqui acaba mostrar histogramas
h3("Actividad 3:"),
h4("Averigüe cómo cambiar la escala del eje vertical de un histograma
en R"),

```

h3("Variaciones"),

h4("En el eje y pueden ir las frecuencias absolutas o las relativas o también se puede utilizar una escala de densidad que se consigue haciendo:"),

h4("Considere la fórmula para hallar el área de un rectángulo, pero con los siguientes remplazos:"),

HTML("\\(Área = Base * Altura\\)"),

br(),

br(),

h4("Reemplace el Área por la 'Frecuencia relativa', la Base por el 'incremento' y de esta ecuación despeje la altura"),

br(),

h5("\\(Frecuencia Relativa = Incremento * Altura\\)"),

br(),

h4(HTML("\\(\\frac{Frecuencia Relativa}{Incremento} = Altura\\)")),

br(),

h4("De esta manera, la altura queda expresada en una escala de Densidad, note que la ganancia al hacer esta transformación es que como la suma de las frecuencias relativas es 1, entonces ahora se tiene que la suma de las áreas de los rectángulos equivale 1, de esta manera los histogramas que se construyen usando en el eje y una escala de densidad siempre representan un área de 1 y al superponer una curva que lo inscriba tendremos una imagen muy cercana a lo que es una distribución de probabilidad real (área bajo la curva es 1); una ventaja adicional es que

de tener varios histogramas contruidos de esta forma, estos son comparables de manera directa ya que se está considerando diferencias en patrones de distribución de áreas iguales."),

h4("Los cálculos requeridos para hacer un histograma se presentan en una Tabla de frecuencias, para los datos del ejemplo que nos ocupa, edad de personas que asisten al cine, tendríamos lo siguiente:"),

br(),

h4("Tabla 1: Distribución de Frecuencias- Edad de personas que asisten a un cine "),

br(),

#TABLA

tags\$head(

tags\$style(HTML("

table {

width: 100%;

text-align: center;

}

table thead tr {

background-color: #9F82F7;

}

table tbody tr:nth-child(even) {

background-color: #DBF7FA;

```
}  
"))  
  
),  
tags$table(  
  tags$thead(  
    tags$str(  
      tags$th("Intervalo de clase"),  
      tags$th("Marca de clase*"),  
      tags$th("Frecuencia"),  
      tags$th("Frecuencia Relativa**"),  
      tags$th("Frecuencia Relativa Acumulada"),  
      tags$th("Densidad")  
    )  
  ),  
  tags$tbody(  
    tags$str(  
      tags$td("10 - 15,5"),  
      tags$td("12,75"),  
      tags$td("10"),  
      tags$td("0,15"),  
      tags$td("0,15"),  
      tags$td("0,03")  
    )  
  ),  
)
```

```
tags$tr(  
  tags$td("15,5 - 21"),  
  tags$td("18,25"),  
  tags$td("18"),  
  tags$td("0,27"),  
  tags$td("0,42"),  
  tags$td("0,05")  
)  
tags$tr(  
  tags$td("21 - 26,5"),  
  tags$td("23,75"),  
  tags$td("9"),  
  tags$td("0,14"),  
  tags$td("0,56"),  
  tags$td("0,02")  
)  
tags$tr(  
  tags$td("26,5 - 32"),  
  tags$td("29,25"),  
  tags$td("7"),  
  tags$td("0,11"),  
  tags$td("0,67"),  
  tags$td("0,02")
```

```
),  
tags$tr(  
  tags$td("32 - 37,5"),  
  tags$td("34,75"),  
  tags$td("10"),  
  tags$td("0,15"),  
  tags$td("0,82"),  
  tags$td("0,03")  
),  
tags$tr(  
  tags$td("37,5 - 43"),  
  tags$td("40,25"),  
  tags$td("7"),  
  tags$td("0,11"),  
  tags$td("0,92"),  
  tags$td("0,02")  
),  
tags$tr(  
  tags$td("48,5 - 54"),  
  tags$td("51,25"),  
  tags$td("2"),  
  tags$td("0,03"),  
  tags$td("1"),
```

```

tags$td("0,01")
)
)
),
br(),
tags$small("*Valor que representar al intervalo de clase, usualmente se
escoge el valor central, pero podría ser cualquiera incluso uno de sus límites."),
tags$br(),
tags$small("**Valor obtenido al dividir la Frecuencia entre el tamaño de
muestra, en este caso n=66"),
br(),
),
tabPanel("P-P Plot y Q-Q Plot",
h3("C-C Plot"),
h4("Técnicamente este gráfico no existe, es introducido como recurso
didáctico para facilitar la lectura e interpretación del PP plot y del QQ plot que sí son gráficos
estadísticos de uso común."),
h4("Se propone como primera actividad comparar los cuartiles teóricos
con los cuartiles muestrales que se encuentra en la tabla que se muestra a continuación"),
h4("Tabla: Comparación de cuartiles teóricos y muestrales"),
#Tabla

```

```
tags$head(  
  tags$style(  
    HTML("  
#cuartilesTable thead th {  
  background-color: #9F82F7;  
}  
#cuartilesTable td, #cuartilesTable th {  
  border: 1px solid black;  
  padding: 5px;  
}  
#cuartilesTable td:not(:last-child),  
#cuartilesTable th:not(:last-child) {  
  border-right: none;  
}  
#cuartilesTable input[type='text'] {  
  width: 60px;  
  height: 20px;  
}  
")  
  )  
),
```

```
uiOutput("cuartilesTable"),  
br(),  
tags$small("*Calculados usando la función =CUARTIL.INC(rango de  
los datos;1) , para los demás se varia el segundo argumento a 2 y 3 respectivamente."),  
br(),
```

```
h4("Grafique las columnas dos y tres de la tabla anterior y analice el nivel  
de acuerdo entre cuartiles acorde a este gráfico."),
```

```
actionButton("show_quartiles", "Mostrar CC plot"),
```

```
br(),
```

```
plotOutput("cuartilesPlot1"),
```

```
h4("Los gráficos de probabilidad constituyen otra importante  
herramienta gráfica para comprobar si un conjunto de datos puede considerarse o no procedente  
de una distribución de probabilidad dada. La idea básica consiste en enfrentar, en un mismo  
gráfico, estadísticos obtenidos a partir de los datos observados frente a estos mismos estadísticos  
pero calculados sobre la distribución que se está validando (Distribución teórica), en este apartado  
consideramos la distribución normal. "),
```

```
h4("En el ejercicio se espera que los cuartiles muestrales sean cercanos  
a los cuartiles de la distribución  $N(40,5)$ , en ese caso los puntos se deben ubicar en torno a la  
línea recta roja que se adiciona al gráfico y corresponde a la función identidad, de lo contrario, la  
evidencia gráfica es que no hay acuerdo y por tanto tendríamos problemas para asumir el supuesto  
de normalidad para esa muestra de datos. "),
```

```
h3("P-P Plot"),
```

```
h4("De igual manera al CC plot, podemos pensar en el PP plot, es decir,  
el gráfico Percentil-Percentil. Usando los datos de la muestra simulada a continuación se muestra  
los insumos para poder construir de manera manual (o ayudado por Excel) el gráfico que ahora en  
lugar de 3 puntos debe ubicar 99 puntos:"),
```

```
plotOutput("percentilesPlot1"),
```

```
br(),
```

```
h4("Una vez construido, interprete el gráfico en relación con el nivel de  
ajuste de los puntos con la recta de referencia (función identidad) prestando atención a la parte  
central del gráfico, los percentiles más pequeños así como los más grandes usualmente se muestran  
dispersos y sin seguir la trayectoria, esto es normal porque corresponden a zonas de baja  
probabilidad por cuanto no deben conducir a concluir que no hay ajuste con el modelo que se está  
contrastando."),
```

```
h4("Finalmente hablaremos del QQ plot o gráfico Cuantil-Cuantil: A  
continuación, puede explorar la forma de varios QQ plots al dar clic en el botón 'Generar muestra'  
"),
```

```
fluidPage(  
  titlePanel(""),
```

```
  sidebarLayout(  
    sidebarPanel(  
      actionButton("generate_data_qq", "Generar datos"),
```

```

tags$hr(),
textInput("data_input_qq", "Datos (separados por comas):",
rows = 5),

actionButton("generate_plot_qq", "Generar gráfico")
),
mainPanel(
plotOutput("qqplot"),
verbatimTextOutput("qqplot_info")
)
),
br(),
h4("Los programas estadísticos aportan como herramienta gráfica de
contraste de un modelo probabilístico esta variación del gráfico PP plot, aunque por ejemplo se
puede obtener el PP plot en R, el comando qqPlot es el que más se utiliza, el resultado es un gráfico
similar al PP plot pero su construcción varía. Utilizaremos los datos de las personas que asisten al
cine para ilustrar paso a paso como construir un QQ plot."),
h4("La esencia del gráfico es la comparación entre el porcentaje de datos
acumulado en la muestra y el porcentaje de área acumulado en la distribución teórica. A
continuación podrá evidenciar este comportamiento al mover el deslizador que permite calcular
los cuantiles de la muestra:"),

fluidRow(

```

```
column(6, plotOutput("densityP")),  
column(6, plotOutput("dotPlot2"))  
)  
  
fluidRow(  
  column(4, offset = 4,  
    sliderInput("percentile", "Cuantil:", min = 1, max = 99, value = 50)  
  )  
)  
  
fluidRow(  
  tags$div(style = "height: 40px;"),  
)  
  
fluidRow(  
  h4("Un cuantil, es un estadístico que se calcula en relación con el  
porcentaje de muestra acumulada en la medida que se avanza del dato más pequeño al dato más  
grande en la muestra ordenada, así no hay un número fijo sino que dependerá del tamaño de  
muestra.")  
)
```

h4("A continuación se detalla de manera tabular la forma cómo se calcularían los cuantiles para los datos de edad de personas que asisten al cine:"),

br(),

h4("Para hallar el nivel de probabilidad:"),

tags\$sp("\\(N.Prob=\\frac{j-\\frac{1}{2}}{n}\\)\"),

h4("Así, en el caso de $(j=1)$ "),

tags\$sp("\\(0.0075=\\frac{1-\\frac{1}{2}}{66}\\)\"),

h4("Para hallar el cuantiles teóricos utilizamos la distribución normal invertida calculada a partir de la $(N(\\overline{X},s))$, es decir aproximamos la media y la desviación estándar usando los valores muestrales."),

h4("Así aquí resolvemos $(P(X<?)=0.0075)$ haciendo en Excel $(=INV.NORM(0,0075;26;11))$ "),

tags\$head(

tags\$style(HTML("

table {

width: 100%;

text-align: center;

}

```
table thead tr {  
    background-color: #9F82F7;  
}  
  
table tbody tr:nth-child(even) {  
    background-color: #DBF7FA;  
}  
"))  
  
    ),  
tags$table(  
    tags$thead(  
        tags$str(  
            tags$th("j"),  
            tags$th("Dato observado"),  
            tags$th("Nivel de probabilidad"),  
            tags$th("Cuantil en N(26,11)")  
        )  
    ),  
tags$body(  
        tags$str(  
            tags$td("1"),  
            tags$td("10"),  
            tags$td("0.0075"),
```

```
tags$std("-0.756")
),
tags$str(
tags$std("2"),
tags$std("11"),
tags$std("0.0227"),
tags$std("3.845")
),
tags$str(
tags$std("3"),
tags$std("11"),
tags$std("0.0378"),
tags$std(" ")
),
tags$str(
tags$std("..."),
tags$std(""),
tags$std(""),
tags$std(""),
tags$std(""),
tags$std("")
),
tags$str(
```

```

tags$std("65"),
tags$std("51"),
tags$std("0.9772"),
tags$std("")
),
tags$str(
tags$std("66"),
tags$std("54"),
tags$std("0.9772"),
tags$std(" ")
)
)
),

```

h4("En la escala de una normal con media $\mu(26.84)$ y desviación estándar $\sigma(11)$, el nivel de probabilidad $\alpha(0.0075 (0.75\%))$ es acumulado a izquierda de $z(-0.756)$ mientras que ese mismo nivel de probabilidad en la muestra de datos es acumulado a la izquierda de $z(10)$, claramente en cuanto al Cuantil 1 no hay bastante diferencia, el cuantil muestral es mayor que el teórico."),

h4("Para el segundo cuantil tenemos que el nivel $\alpha(0.0227 (2.27\%))$ es acumulado en la $N(26,11)$ por el valor $z(3.84)$, mientras que en la muestra es acumulado a la izquierda de 11, otra vez no hay proximidad entre estos dos valores, se sugiere verificar la ubicación de estos dos primeros puntos en los gráficos que se presentan a continuación."),

```
br(),  
("A continuación se detalla la forma cómo se calcularían los cuantiles  
para los datos de edad de personas que asisten al cine usando R:"),
```

```
br(),  
br(),  
sidebarPanel(  
  h3("Código R para QQ plot"),  
  tags$ol(  
    tags$li("Instalar y cargar los paquetes necesarios"),  
    tags$li("Leer el archivo de Excel"),  
    tags$li("Crear el QQ plot")  
  ),  
  shinyAce::aceEditor("code2",  
    value = "  
  
# Cargar paquetes necesarios  
  
library(ggplot2)  
  
library(readxl)  
  
# Leer archivo de Excel  
  
datos <- read_excel('EdadesPersonas.xlsx', col_types = c('numeric'))  
  
# Crear QQ plot  
  
ggplot(data.frame(y = datos$Edades), aes(sample = y)) +
```


br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

br(),

h3("Actividad:"),

h4("1. ¿Qué pasaría en el gráfico si el dato más grande fuese 94 en lugar de 54? "),

h4("2. ¿Qué pasaría en el gráfico si le suma diez a los veinte últimos datos?"),

```
h4("3. ¿Qué pasaría en el gráfico si hay diez datos iguales?"),
```

```
h3("Tips para la interpretación:"),
```

```
h4("Los gráficos Q-Q se obtienen representando los cuantiles respecto a los cuantiles de la distribución normal. Además de permitir valorar la desviación de la normalidad, los gráficos de probabilidad permiten conocer la causa de esa desviación, así Una curva en forma de 'u' o con alguna curvatura, significa que la distribución es asimétrica con respecto a la normal, mientras que un gráfico en forma de 'S' significará que la distribución tiene colas mayores o menores que la normal, esto es, que existen pocas o demasiadas observaciones en las colas de la distribución.")
```

```
),
```

```
tabPanel("Ejercicios propuestos",
```

```
h3("Ejercicios propuestos"),
```

```
h5(HTML("<strong>Actividad 1:</strong> Genere una muestra de datos de una distribución N(50,5) usando el <a href='https://www.socscistatistics.com/utilities/normaldistribution/default.aspx'>simulador</a>.
```

```
El objetivo es usar estos datos para validar el supuesto de normalidad mediante herramientas gráficas, como el histograma y el PP plot.")),
```

```
br(),
```

```
sidebarLayout(
```

```
sidebarPanel(
```

```
textAreaInput("data_input2",
```

"Ingrese su muestra de datos (separados por comas o saltos de línea):",

```

    height = '200px'),
  selectInput("plot_type2",
    "Seleccione el tipo de gráfico:",
    choices = c("QQ Plot", "Histograma", "Boxplot"))
),

```

```

  mainPanel(
    plotOutput("data_plot2")
  )
),

```

h4("a) Elabore e interprete el histograma para estos datos, ¿sugiere éste que los datos provengan de una distribución normal?, en particular la muestra se tomó de un modelo $N(50,5)$, se evidencia este comportamiento particular claramente?"),

```
br(),
```

h4(HTML("Nota: Aquí puede hacer uso de los estadísticos descriptivos muestrales media y desviación estándar, nuestro recorrido previo nos mostró como calcularlos para la distribución de una variable aleatoria, aquí se requiere el computó de esas dos mismas medidas de resumen, pero a partir de una muestra de datos para hacer el contraste con 50 y 5 respectivamente.")),

```
br(),
```

h4("b) Para la muestra de datos, calcule los cuartiles muestrales usando la función =CUARTIL.EXC de Excel "),

h4("c) Para la muestra de datos, calcule los cuartiles muestrales usando la función quantil de R a partir de cualquiera de estos tres comandos:"),

```
tags$ul(
  tags$li("quantil (nombre de la variable) #sin argumentos devuelve los
cuartiles"),
  tags$li("quantile(nombre de la variable, prob=c(0,0.25,0.5,0.75,1))"),
  tags$li("quantile(nombre de la variable, prob=seq(0, 1, 1/4))"),
),
sidebarLayout(
  sidebarPanel(
    aceEditor("code4",
      value = '# Crea algunos datos

datosCuartil <- c(1,2,3,4,5)

# Calcula los cuartiles

cuartiles <- quantile(datosCuartil, probs=c(0.25, 0.5, 0.75))

# Imprime los cuartiles

print(cuartiles)',
      height = "200px"),
    actionButton("execute", "Ejecutar Código")
```

```

),
mainPanel(
  verbatimTextOutput("results")
)
),

```

h4("d) Compare los valores obtenidos en b y c, concilie los resultados, piense en qué fuentes de diferencia puede haber en el cálculo de estos estadísticos de un software a otro?"),

h4("e) Comparar los percentiles teóricos, es decir los de una distribución N (50,5), con los percentiles muestrales obtenidos en R"),

h4("f) Grafique los percentiles del punto anterior en un PP plot (Percentil-Percentil), interpretar el nivel de ajuste de los puntos en relación con la recta de referencia (Función Identidad) y si hay desviación por qué es que esta se produce. "),

```
br(),
```

h5(HTML("Actividad 2: Analice el comportamiento del PP plot para una muestra de datos sesgada a derecha. Te sugerimos que uses los datos de población de las ciudades capitales de Colombia o las estaturas de una muestra de jugadores de futbol que podría ser una selección o equipo profesional particular (incluir una muestra de al menos 30 datos)")),

```
br(),
```

h5(HTML("Actividad 3: Invente una muestra de datos que claramente sea sesgada a izquierda (grafique el histograma para asegurarse de esto). Contruya el PP plot, ¿cómo luce el gráfico para este tipo de distribución?")),

br(),

h5(HTML("Actividad 4: Para la muestra de datos de la Actividad 1, adicione dos valores muy extremos a derecha, grafique el PP plot, observe cómo afectan estos dos datos al gráfico. ")),

br(),

h5(HTML("Actividad 5: ")),

br(),

h4("A menudo veremos otras formas de presentar un histograma como el que aparece a la izquierda de la siguiente imagen:"),

imageOutput("Imagen7"),

h4("De manera informal podríamos decir que al unir los centros de las bases superiores de los rectángulos de la distribución del histograma de frecuencias acumuladas, da lugar a un polígono de frecuencias acumuladas que al suavizarse genera el gráfico llamado Ojiva. "),

h4("¿Será que la ojiva sirve como gráfico de contraste para el supuesto de normalidad?"),

h4("¿Cómo luce una ojiva para una distribución simétrica?, ¿y para una distribución asimétrica a derecha?"),

br(),

h4(HTML("Actividad 6: Análisis de datos La estimación del porcentaje de grasa corporal hoy día es muy importante en términos de procedimientos de acondicionamiento físico riguroso, no obstante de poder medir el peso o indicadores como el índice de masa corporal, su medición exacta requiere de procedimientos más sofisticados para aislar la medición que corresponde al peso de la masa muscular; por lo anterior, hallar una relación del % de grasa con otras medidas corporales más fáciles de obtener es útil y esto se puede conseguir usando un métodos estadísticos denominado Regresión Lineal múltiple el cual proporciona una “ecuación” para inferir el valor del % de grasa de una manera económica y sin generar incomodidad al paciente. ")),

h4("Este ", tags\$(href = "https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset?resource=download", "enlace"), " aporta una muestra de datos (n=252 hombres) recolectada con el mencionado propósito. Por ahora, nos ubicaremos en el primer paso de todo análisis estadístico, el análisis exploratorio de los datos, allí uno de los interrogantes que se debe responder es: ¿cuáles de las variables disponibles en la base de datos parecen seguir un comportamiento probabilístico Normal? Use las herramientas gráficas vistas para responder."),

),

tabPanel("Práctica en analítica de datos",

h4("Hasta hace algunos años, todo análisis estadístico iba acompañado de un informe técnico en formato plano, archivo en Word, Pdf, libros en Excel por mencionar los

más comunes, hoy día, el desarrollo computacional y el ritmo acelerado de vida que se maneja, han promovido la creación de diversos formatos para presentar los resultados de un análisis estadístico , de esta manera vemos informes tipo infografía, video, poster, presentaciones multimedia y más recientemente tableros interactivos más conocidos como Dashboard; estas nuevas opciones de presentación enfatizan el componente gráfico y de diseño lo cual demanda además de la capacidad de análisis y síntesis, la habilidad para escoger las herramientas apropiadas, y la organización de estos elementos dentro de un informe. Por lo tanto, un analista de datos moderno debe estar preparado tanto para la lectura e interpretación de este tipo de formatos como para la elaboración de estos resúmenes de manera eficiente y creativa. "),

br(),

h4("A continuación, encontraras un instructivo de dos pasos que te guiará en la elaboración de tu primer dashboard en Excel: "),

h3("PASO 1"),

tags\$div(style = "display: flex; justify-content: center;",

tags\$iframe(style="border: 1px solid rgba(0, 0, 0, 0.1);",

width = "560", height = "315",

src = "https://www.youtube.com/embed/vPnI3gtC4m0",

frameborder = "0", allowfullscreen = TRUE)

),

tags\$div(style = "height: 20px;"), #Salto de linea grande

h3("PASO 2"),

tags\$div(style = "display: flex; justify-content: center;",

```
tags$Iframe(style="border: 1px solid rgba(0, 0, 0, 0.1);",
            width = "560", height = "315",
            src = "https://www.youtube.com/embed/zruboUFYkDA",
            frameborder = "0", allowfullscreen = TRUE)
),
br(),
h3("Actividad práctica:"),
h4("Elabore un dashboard que permita analizar el comportamiento
distribucional de algunas variables para los datos de la base de datos que encuentra en el siguiente
enlace.")
)
# Agrega más tabPanel si son necesarios
))
# Agrega más tabPanel según sea necesario
),

# Botones flotantes
div(id = "floating-buttons",
    actionButton("read", label = "", icon = icon("play-circle"), class = "btn btn-primary
btn-floating"),
    actionButton("pause_resume", label = "", icon = icon("pause"), class = "btn btn-
warning btn-floating"),
```

```
        actionButton("stop", label = "", icon = icon("stop"), class = "btn btn-danger btn-  
floating"),  
        actionButton("back_to_top", label = "", icon = icon("arrow-up"), class = "btn btn-  
info btn-floating")  
    )  
  
))))
```

Definir la lógica del servidor

```
server <- function(input, output, session) {
```

```
  observeEvent(input$start_btn, {
```

```
    hide("landing_page")
```

```
    show("new_page")
```

```
  })
```

```
  observeEvent(input$go_main, {
```

```
    hide("new_page")
```

```
    show("main_content")
```

```
  })
```

```
#LECTOR DE VOZ
```

```
observeEvent(input$read, {
```

```
  runjs("readText()")
```

```
})
```

```
observeEvent(input$pause_resume, {  
  runjs("pauseResume()")  
})
```

```
})
```

```
observeEvent(input$stop, {  
  runjs("stopReading()")  
})
```

```
})
```

```
#LECTOR DE VOZ FIN
```

```
#Tamaño letra
```

```
observeEvent(input$increase_text, {  
  runjs(  
    $(".tab-pane").css("font-size", function(i, value) {  
      return parseFloat(value) * 1.1 + "px";  
    });  
  )  
})
```

```
})
```

```
})
```

```
observeEvent(input$decrease_text, {
```

```
runjs(  
  $(".tab-pane").css("font-size", function(i, value) {  
    return parseFloat(value) * 0.9 + "px";  
  });  
)  
})  
  
#Tamaño letra fin  
  
#CONTENIDO  
  
#introduccion e histogramas----  
  
# Generar el histograma poblacion de introduccion  
output$histogramPoblacion <- renderPlot({  
  ggplot(datos_filtrados(), aes(x = Poblacion)) +  
    geom_histogram(fill = "#9F82F7", color = "gray", bins = 15) +  
    labs(title = "",  
          x = "Población",  
          y = "Frecuencia")  
})  
  
observeEvent(input$back_to_top, {
```

```
shinyjs::runjs('window.scrollTo(0, 0);')
})
```

```
text_size <- reactiveValues(size = 14)
```

```
observeEvent(input$increase_text, {
  text_size$size <- text_size$size + 1
  runjs(sprintf("
    $('body').css('font-size', '%dpx');
    $('h1').css('font-size', '%fpx');
    $('h2').css('font-size', '%fpx');
    $('h3').css('font-size', '%fpx');
    $('h4').css('font-size', '%fpx');
    $('h5').css('font-size', '%fpx');
    $('h6').css('font-size', '%fpx');
    ", text_size$size, text_size$size * 2, text_size$size * 1.5, text_size$size * 1.17,
    text_size$size * 1, text_size$size * 0.83, text_size$size * 0.67))
})
```

```
observeEvent(input$decrease_text, {
  text_size$size <- text_size$size - 1
  runjs(sprintf("
    $('body').css('font-size', '%dpx');
    $('h1').css('font-size', '%fpx');
    $('h2').css('font-size', '%fpx');
    $('h3').css('font-size', '%fpx');
    $('h4').css('font-size', '%fpx');
    $('h5').css('font-size', '%fpx');
    $('h6').css('font-size', '%fpx');
    ", text_size$size, text_size$size * 2, text_size$size * 1.5, text_size$size * 1.17,
    text_size$size * 1, text_size$size * 0.83, text_size$size * 0.67))
})
```

```
$(body').css('font-size', '%dpx');  
$(h1').css('font-size', '%fpx');  
$(h2').css('font-size', '%fpx');  
$(h3').css('font-size', '%fpx');  
$(h4').css('font-size', '%fpx');  
$(h5').css('font-size', '%fpx');  
$(h6').css('font-size', '%fpx');  
", text_size$size, text_size$size * 2, text_size$size * 1.5, text_size$size * 1.17,  
text_size$size * 1, text_size$size * 0.83, text_size$size * 0.67))  
})
```

#HISTOGRAMA IQ SCORE

```
output$interactiveHistogram <- renderPlot({  
  # Generar datos aleatorios para simular puntuaciones de CI  
  ci_scores <- rnorm(1000, mean = 100, sd = 15)  
  
  # Definir los rangos de CI  
  ci_ranges <- data.frame(  
    start = c(55, 70, 85, 115, 130),  
    end = c(70, 85, 115, 130, 145),  
    label = c("Discapacidad", "Inferior", "Promedio", "Superior", "Excepcional"),  
    y_position = c(0.015, 0.015, 0.015, 0.015, 0.015),
```

```
x_position = c(62.5, 77.5, 100, 122.5, 137.5)
)

# Crear el histograma con ggplot
ggplot(data = data.frame(ci_scores), aes(x = ci_scores)) +
  geom_histogram(aes(y = ..density..),
    binwidth = 160/input$bins,
    fill = "#9F82F7",
    alpha = 0.5) +
  stat_function(fun = dnorm,
    args = list(mean = mean(ci_scores), sd = sd(ci_scores)),
    color = "red",
    size = 1) +
  geom_vline(data = ci_ranges, aes(xintercept = start), linetype = "dashed", color =
"gray") +
  geom_vline(data = ci_ranges, aes(xintercept = end), linetype = "dashed", color =
"gray") +
  geom_text(data = ci_ranges, aes(x = x_position, y = y_position, label = label),
    angle = 90, vjust = 0.5, hjust = 0.5) +
  labs(x = "Puntuación de CI", y = "Densidad") +
  xlim(40, 160) +
  theme_minimal()
})
```

```
#Termina HISTOGRAMA IQ SCORE
```

```
bins <- reactive({  
  as.integer(input$binsInput)  
})
```

```
#Calculo de histograma manual
```

```
#Rango
```

```
observeEvent(input$calcRange, {  
  maxVal <- as.numeric(input$maxInput)  
  minVal <- as.numeric(input$minInput)  
  rangeVal <- maxVal - minVal  
  output$rangeOutput <- renderText({  
    paste("El RANGO es: ", rangeVal)  
  })  
})
```

```
#Incremento
```

```
observeEvent(input$calcClassLength, {  
  rangeVal <- as.numeric(input$rangeInput)
```

```
classNum <- as.numeric(input$classInput)

classLength <- rangeVal / classNum

output$classLengthOutput <- renderText({
  paste("La Longitud de Clase es: ", classLength)
})
})

#Histograma con calculo manual

observeEvent(input$hist_button, {
  output$histPlotu <- renderPlot({
    datosu <- c(10,11,11,13,14,14,15,15,15,15,16,16,16,17,17,17,18,18,18,19,19,19,
      20,20,20,21,21,21,22,22,23,23,23,24,26,26,26,27,29,30,30,31,32,32,33,
      33,34,34,35,35,36,37,37,37,39,40,40,41,42,42,43,45,46,46,51,54)

    # Intervalos de las barras

    breaks <- c(+10, 15.5, 21, 26.5, 32, 37.5, 43, 48.5, 54)

    h <- hist(datosu, breaks = breaks, main = "Histograma de datos", xlab = "Datos", ylab
= "Frecuencia", plot = FALSE)

    plot(h, col = "#9F82F7", border = "gray", main = "Histograma de edades", xlab =
"Datos", ylab = "Frecuencia")
```

```
# Agregar texto

text(x = h$mids, y = h$counts, labels = paste("[", h$breaks[-length(h$breaks)], ", ",
h$breaks[-1], ")", sep = ""), pos = 3, cex = 0.7)

})

})

output$Imagen2 <- renderImage({

  list(

    src = "histogramanormalysesgado.jpg",

    contentType = "image/jpg",

    width = 600,

    height = 300

  )

})

#Tipos de comportamiento

output$distPlot <- renderPlot({

  if(input$tipo == "Simétrico") {

    datos <- rnorm(1000)

  }

  else if(input$tipo == "Sesgado a la derecha") {

    datos <- rexp(1000, 0.2)
```

```
}  
  
else if(input$tipo == "Sesgado a la izquierda") {  
  datos <- -rexp(1000, 0.2)  
}  
  
else if(input$tipo == "J invertida") {  
  datos <- rweibull(1000, shape = 0.5, scale = 1)  
}  
  
else if(input$tipo == "Forma de J") {  
  datos <- max(rweibull(1000, shape = 0.5, scale = 1)) - rweibull(1000, shape = 0.5, scale  
= 1)  
}  
  
else if(input$tipo == "Forma de U") {  
  datos <- c(rnorm(500, mean = -2), rnorm(500, mean = 2))  
  datos <- datos[datos >= -2 & datos <= 2]  
}  
  
else if(input$tipo == "Bimodal") {  
  datos <- c(rnorm(500, mean = -2), rnorm(500, mean = 2))  
}  
  
else if(input$tipo == "Multimodal") {  
  datos <- c(rnorm(300, mean = -3, sd = 0.5),  
            rnorm(300, mean = 0, sd = 0.5),  
            rnorm(400, mean = 3, sd = 0.5))  
}
```

```
p <- ggplot(data.frame(x=datos), aes(x=x)) +  
  geom_density(color="red", size=1.5, adjust=2) +  
  labs(title=paste("Comportamiento", input$tipo),  
        x="Valor",  
        y="Densidad")
```

```
if(input$tipo == "Forma de U") {  
  p <- p + coord_cartesian(xlim = c(-1, 1))  
}
```

```
print(p)  
})
```

```
output$histPlot2 <- renderPlot({  
  data_numeric2 <- c(34, 34, 26, 37, 42, 41, 35, 31, 41, 33, 30, 74, 33, 49, 38, 61, 21, 41,  
26, 80, 43, 29, 33, 35, 45, 49, 39, 34, 26, 25, 33, 35)
```

```
# Regla de Sturges
```

```
bins <- ceiling(log2(length(data_numeric2)) + 1)
```

```
hist(data_numeric2, breaks = bins, main = "Histograma", xlab = "Edad", ylim = c(0, 40),
col = "#9F82F7", border = "grey")

})
```

```
output$histPlotWithDensity <- renderPlot({

  data_numeric2 <- c(34, 34, 26, 37, 42, 41, 35, 31, 41, 33, 30, 74, 33, 49, 38, 61, 21, 41,
26, 80, 43, 29, 33, 35, 45, 49, 39, 34, 26, 25, 33, 35)
```

```
  # Calcula la media y la desviación estándar de tus datos

  media <- mean(data_numeric2)

  desv_estandar <- sd(data_numeric2)

  # Usa ggplot para crear el histograma y superponer la curva de densidad normal

  ggplot(data = data.frame(Edad = data_numeric2), aes(x = Edad)) +

  geom_histogram(aes(y = ..density..), fill = "blue", alpha = 0.5, bins = 10) +

  stat_function(fun = dnorm, args = list(mean = media, sd = desv_estandar), color =

"red", size = 1) +

  labs(title = "Histograma de Edades con Curva de Distribución Normal",

  x = "Edad",

  y = "Densidad") +

  theme_minimal()

})
```

```
# Crear el histograma basado en la variable seleccionada futbol

output$histogram <- renderPlot({

  # Leer los datos del archivo Excel

  data <- read_excel("FUTBOL.xlsx")

  # Graficar el histograma para la variable seleccionada

  hist(data[[input$variable]], main = input$variable, xlab = input$variable, col =
"#9F82F7", border = "grey")

})
```

```
#histograma con r

output$histogram3 <- renderPlot({

  code <- input$code

  if (!is.null(code) && code != "muestra <- c(10, 12, 15, 13, 20, 18, 11, 14, 17, 16, 19, 12,
14, 15, 13)\nhist(muestra, main = \"Histograma\", xlab = \"Valores\", ylab = \"Frecuencia\")" ) {

    eval(parse(text = code))

  }

})
```

#HISTOGRAMAS CON BOTON

```
datahist <- reactive({

  data.frame(
```

```
Intervalo_de_clase = c("10 - 15,5", "15,5 - 21", "21 - 26,5", "26,5 - 32", "32 - 37,5",  
"37,5 - 43", "48,5 - 54"),
```

```
Marca_de_clase = c(12.75, 18.25, 23.75, 29.25, 34.75, 40.25, 51.25),
```

```
Frecuencia = c(10, 18, 9, 7, 10, 7, 2),
```

```
Frecuencia_Relativa = c(0.15, 0.27, 0.14, 0.11, 0.15, 0.11, 0.03),
```

```
Frecuencia_Relativa_Acumulada = c(0.15, 0.42, 0.56, 0.67, 0.82, 0.92, 1),
```

```
Densidad = c(0.03, 0.05, 0.02, 0.02, 0.03, 0.02, 0.01)
```

```
)
```

```
})
```

```
observeEvent(input$boton_rango, {
```

```
  updateNumericInput(session, "rango", value = input$maximo - input$minimo)
```

```
})
```

```
observeEvent(input$boton_longitud, {
```

```
  updateNumericInput(session, "longitud", value = input$rango / input$clases)
```

```
})
```

```
output$histograma1 <- renderPlot({
```

```
  req(input$showHistograms)
```

```
  ggplot(datahist(), aes(x=Marca_de_clase, y=Frecuencia)) +
```

```
  geom_bar(stat="identity", fill="#9F82F7", color="gray", width=5.5) +
```

```
  ggtitle("Histograma de Frecuencias") +
```

```
  xlab("Marca de clase") +  
  ylab("Frecuencia")  
})
```

```
output$histograma2 <- renderPlot({  
  req(input$showHistograms)  
  ggplot(datahist(), aes(x=Marca_de_clase, y=Frecuencia_Relativa)) +  
  geom_bar(stat="identity", fill="#9F82F7", color="gray", width=5.5) +  
  ggtitle("Histograma de Frecuencia Relativas") +  
  xlab("Marca de clase") +  
  ylab("Frecuencia relativa")  
})
```

```
output$histograma3 <- renderPlot({  
  req(input$showHistograms)  
  ggplot(datahist(), aes(x=Marca_de_clase, y=Densidad)) +  
  geom_bar(stat="identity", fill="#9F82F7", color="gray", width=5.5) +  
  ggtitle("Histograma de Densidad") +  
  xlab("Marca de clase") +  
  ylab("Densidad")  
})
```

```
output$histograma4 <- renderPlot({
```

```
req(input$showHistograms)

ggplot(datahist(), aes(x=Marca_de_clase, y=Frecuencia_Relativa_Acumulada)) +
  geom_bar(stat="identity", fill="#9F82F7", color="gray", width=5.5) +
  ggtitle("Histograma de Frecuencia Relativas Acumuladas") +
  xlab("Marca de clase") +
  ylab("Frecuencia relativa acumulada")
})
```

```
#AQUI INICIA EL SERVER DE PP PLOY Y QQ PLOT ----
```

```
#TABLA
```

```
output$cuartilesTable <- renderUI({
```

```
  # Calcular los cuartiles teóricos y redondear a un decimal
```

```
  cuartil1 <- if(input$show_quartiles > 0) round(qnorm(0.25, mean=40, sd=5), 1) else ""
```

```
  cuartil2 <- if(input$show_quartiles > 0) round(qnorm(0.5, mean=40, sd=5), 1) else ""
```

```
  cuartil3 <- if(input$show_quartiles > 0) round(qnorm(0.75, mean=40, sd=5), 1) else ""
```

```
  table <- tags$table(
```

```
    tags$thead(
```

```
      tags$tr(
```

```
        tags$th("Cuartil"),
```

```
        tags$th("Cuartil Teórico N(40,5)"),
```

```
        tags$th("Cuartil Muestral")
```

```
)  
,  
tags$body(  
  tags$str(  
    tags$td("Q1"),  
    tags$td(cuartil1),  
    tags$td("22.7")  
  ),  
  tags$str(  
    tags$td("Q2"),  
    tags$td(cuartil2),  
    tags$td("38")  
  ),  
  tags$str(  
    tags$td("Q3"),  
    tags$td(cuartil3),  
    tags$td("46.25")  
  )  
)  
)  
)  
  
return(table)  
)
```

```
#Grafico cuartil cuartil

output$cuartilesPlot1 <- renderPlot({

  if(input$show_quartiles == 0) return(NULL) # No mostrar nada antes de hacer clic en
el botón
```

```
# Calcular los cuartiles teóricos

cuartil1 <- qnorm(0.25, mean=40, sd=5)
cuartil2 <- qnorm(0.5, mean=40, sd=5)
cuartil3 <- qnorm(0.75, mean=40, sd=5)

# Crear un dataframe con los cuartiles teóricos y muestrales

data3 <- data.frame(

  Cuartil = c("Q1", "Q2", "Q3"),
  Teorico = c(cuartil1, cuartil2, cuartil3),
  Muestral = c(22.7, 38, 46.25)

)

# Generar el gráfico

ggplot(data3, aes(x = Muestral, y = Teorico)) + # Intercambiar los ejes

  geom_point() +

  geom_abline(slope = 1, intercept = 0, color = "red") + # Línea y = x
```

```
    expand_limits(x = 0, y = 0) + # Hacer que los ejes se crucen en (0,0)
    xlab("Cuartiles Muestrales") + # Actualizar etiquetas de los ejes
    ylab("Cuartiles Teóricos") +
    ggtitle("Comparación de Cuartiles")
  })

#grafico pp plot

muestraplot <- c(40, 38, 22, 31, 26, 23, 42, 56, 12, 12, 25, 42, 46, 40, 22, 38, 6, 57, 50,
40, 15, 48, 68, 73, 54, 24, 5, 40, 26, 44, 10, 42, 27, 89, 24, 23, 30, 50, 33, 65, 83, 47, 46, 30, 2, 66,
18, 41, 34, 24, 26, 39, 43, 2, 11, 30, 32, 28, 40, 40, 29, 27, 12, 7, 7, 25, 78, 40, 61, 73, 38, 14, 5,
40, 48, 73, 65, 33, 40, 72, 40, 88, 44, 78, 22, 25, 40, 67, 22, 40, 2, 40, 35, 58, 17, 8, 18, 82, 19, 3)

output$percentilesPlot1 <- renderPlot({
  if(input$show_quartiles == 0) return(NULL)

  # Calcular los percentiles teóricos y muestrales para varios puntos
  puntos <- seq(0.01, 0.99, by = 0.01)
  teorico <- qnorm(puntos, mean = mean(muestraplot), sd = sd(muestraplot))
  muestral <- quantile(muestraplot, puntos)
```

```
dataplot2 <- data.frame(  
  Percentil = puntos,  
  Teorico = teorico,  
  Muestral = muestral  
)  
  
ggplot(dataplot2, aes(x = Muestral, y = Teorico)) +  
  geom_point() +  
  geom_abline(slope = 1, intercept = 0, color = "red") + # Línea y = x  
  expand_limits(x = 0, y = 0) + # Hacer que los ejes se crucen en (0,0)  
  xlab("Percentiles Muestrales") +  
  ylab("Percentiles Teóricos") +  
  ggtitle("Comparación de Percentiles")  
})
```

#QQ PLOT INTERACTIVO

```
observeEvent(input$generate_data_qq, {
```

```
random_numbers <- sample(1:100, size = 100, replace = TRUE)

updateTextAreaInput(session, "data_input_qq", value = paste(random_numbers,
collapse = ", "))

})
```

```
data_qq <- eventReactive(input$generate_plot_qq, {
  as.numeric(unlist(strsplit(input$data_input_qq, split = ",\\s+")))
})
```

```
output$qqplot <- renderPlot({
  req(data_qq())
  ggplot() +
    geom_qq(aes(sample = data_qq())) +
    geom_qq_line(aes(sample = data_qq()), color = "red") +
    labs(x = "Cuantiles teóricos", y = "Cuantiles empíricos", title = "QQ plot") +
    theme_bw()
})
```

```
output$Imagen5 <- renderImage({
  list(
    src = "Cuantil.png",
    contentType = "image/png",
    width = 600,
```

```
    height = 400
  )
})

observeEvent(input$run_code2, {
  tryCatch(
    {
      eval(parse(text = input$code2))

      output$qqPlot2 <- renderPlot({
        # Cargar paquetes necesarios

        library(ggplot2)

        library(readxl)

        # Leer archivo de Excel

        datos <- read_excel('EdadesPersonas.xlsx', col_types = c('numeric'))

        # Crear QQ plot

        ggplot(data.frame(y = datos$Edades), aes(sample = y)) +
          stat_qq() +
          stat_qq_line() +
          ggtitle('QQ plot de Edades')
      })
    },
```

```
error = function(e) {  
  showNotification(paste("Error:", e$message), type = "error")  
}  
)  
})
```

```
output$Imagen7 <- renderImage({  
  list(  
    src = "ojiva.png",  
    contentType = "image/png",  
    width = 600,  
    height = 400  
  )  
})
```

```
output$Imagen8 <- renderImage({  
  list(  
    src = "hist.png",  
    contentType = "image/png",  
    width = 600,  
    height = 400  
  )  
})
```

```
#Comparacion distrib vs dotplot
```

```
muestraplot2 <- c(40, 38, 22, 31, 26, 23, 42, 56, 12, 12, 25, 42, 46, 40, 22, 38, 6, 57, 50,  
40, 15, 48, 68, 73, 54, 24, 5, 40, 26, 44, 10, 42, 27, 89, 24, 23, 30, 50, 33, 65, 83, 47, 46, 30, 2, 66,  
18, 41, 34, 24, 26, 39, 43, 2, 11, 30, 32, 28, 40, 40, 29, 27, 12, 7, 7, 25, 78, 40, 61, 73, 38, 14, 5,  
40, 48, 73, 65, 33, 40, 72, 40, 88, 44, 78, 22, 25, 40, 67, 22, 40, 2, 40, 35, 58, 17, 8, 18, 82, 19, 3)
```

```
output$densityP <- renderPlot({
```

```
  q_theoretical <- qnorm(input$percentile / 100, mean = 40, sd = 5)
```

```
  ggplot(data = data.frame(x = c(min(muestraplot2), max(muestraplot2))), aes(x)) +
```

```
    stat_function(fun = dnorm, args = list(mean = 40, sd = 5),
```

```
      colour = "red", size = 1.5) +
```

```
    geom_vline(aes(xintercept = q_theoretical), colour = "green", linetype = "dashed", size  
= 1) +
```

```
    annotate("text", x = q_theoretical, y = 0.06, label = "Cuantil teórico", hjust = 0) +
```

```
    theme_minimal() +
```

```
    labs(x = "Valor", y = "Densidad") +
```

```
    ggtitle("N(40,5)")
```

```
  })
```

```
output$dotPlot2 <- renderPlot({
```

```
df <- data.frame(value = muestraplot2)

q_sample <- quantile(muestraplot2, probs = input$percentile / 100)

ggplot(df, aes(x = value)) +

  geom_dotplot(binwidth = 1.5, fill = "#9F82F7", dotsize = 1) +

  geom_vline(aes(xintercept = q_sample), colour = "green", linetype = "dashed", size =

1)+

  annotate("text", x = q_sample, y = 0.5, label = "Cuantil muestral", hjust = 0) +

  theme_minimal() +

  labs(x = "Valor", y = "Frecuencia") +

  ggtitle("Dotplot")

})
```

```
#BOXPLOT ----
```

```
observeEvent(input$generate_data_colored, {

  random_numbers <- sample(1:100, size = 100, replace = TRUE)

  updateTextAreaInput(session, "data_input_colored", value = paste(random_numbers,

collapse = ", "))

})
```

```
data_colored <- eventReactive(input$generate_plot_colored, {
  as.numeric(unlist(strsplit(input$data_input_colored, split = "\\s+")))
})

output$boxplot_colored <- renderPlot({
  req(data_colored())
  ggplot() +
    geom_boxplot(aes(x = 1, y = data_colored()), coef = 1.5,
      fill = "#9F82F7", color = "#6495ED", outlier.color = "#228B22", outlier.size
= 2) +
    geom_hline(aes(yintercept = median(data_colored())), color = "#DC143C", size = 1) +
    labs(x = "", y = "Valores", title = "Boxplot") +
    theme_bw()
})

# Lectura de la base de datos
datos2 <- reactive({
  read_xlsx("Poblacion.por.departamento.xlsx")
})

# Dividir los departamentos en dos listas
lista_deptos <- reactive({
  deptos <- unique(datos2())$Departamento)
```

```
mitad <- ceiling(length(deptos) / 2)

list(deptos1 = deptos[1:mitad], deptos2 = deptos[(mitad + 1):length(deptos)])

})

# Actualizar las opciones del checkboxGroupInput

observe({

  updateCheckboxGroupInput(

    session,

    "deptosSeleccionados",

    choices = lista_deptos()$deptos1,

    selected = lista_deptos()$deptos1

  )

  updateCheckboxGroupInput(

    session,

    "deptosSeleccionados2",

    choices = lista_deptos()$deptos2,

    selected = lista_deptos()$deptos2

  )

})

# Filtrar los datos según los departamentos seleccionados

datos_filtrados <- reactive({

  deptos_sel <- c(input$deptosSeleccionados, input$deptosSeleccionados2)
```

```
datos2() %>%  
  filter(Departamento %in% deptos_sel)  
})  
  
cuartiles <- reactive({  
  quantile(datos_filtrados())$Poblacion, probs = c(0.25, 0.5, 0.75))  
})  
  
# Generar el boxplot horizontal  
output$boxplotPoblacion <- renderPlot({  
  ggplot(datos_filtrados(), aes(x = Poblacion, y = "")) +  
    geom_boxplot(fill = "#9F82F7", color = "gray") + # Cambiar colores aquí  
    geom_vline(aes(xintercept = cuartiles()[1]), color = "blue", linetype = "dashed") +  
    geom_vline(aes(xintercept = cuartiles()[2]), color = "green", linetype = "dashed") +  
    geom_vline(aes(xintercept = cuartiles()[3]), color = "red", linetype = "dashed") +  
    labs(title = "",  
         x = "Población",  
         y = NULL)  
})  
  
# Mostrar la tabla transpuesta  
output$datos2Transpuestos <- renderTable({  
  t(datos_filtrados())  
}, rownames = TRUE)
```

```
# Generar la tabla de cuartiles

output$cuartilesTabla <- renderTable({

  if (nrow(datos_filtrados()) > 0) {

    cuartiles_df <- data.frame(Q1 = cuartiles()[1],

                              Q2 = cuartiles()[2],

                              Q3 = cuartiles()[3])

    colnames(cuartiles_df) <- c("Q1", "Q2", "Q3")

    cuartiles_df

  } else {

    data.frame()

  }

}, rownames = FALSE, colnames = TRUE)

# Carga el archivo de Excel

datos3 <- read_excel("Datosboxplot.xlsx")

# Inicializa un objeto reactiveValues para almacenar el conjunto de datos actual

rv <- reactiveValues(datos_actuales = NULL, boxplot_visible = FALSE)

# Selecciona una fila aleatoria en función del botón

fila_aleatoria <- eventReactive(input$nuevo_problema, {
```

```
rv$boxplot_visible <- FALSE

sample(nrow(datos3), 1)

}, ignoreNULL = FALSE)

# Muestra el problema correspondiente y actualiza el conjunto de datos en rv

output$problema <- renderText({

  fila <- fila_aleatoria()

  rv$datos_actuales <- datos3[fila, -1]

  as.character(datos3[fila, 1])

})

# Muestra los datos correspondientes como texto

output$datos <- renderText({

  paste(rv$datos_actuales, collapse = " ")

})

observeEvent(input$mostrar_boxplot, {

  rv$boxplot_visible <- TRUE

})

# Usa renderUI para generar un panel condicional que muestra el boxplot solo cuando
rv$boxplot_visible es TRUE

output$plot_ui <- renderUI({
```

```
    if (rv$boxplot_visible) {  
      plotOutput("boxplot")  
    }  
  })  
  
# Muestra el boxplot horizontal con colores personalizados  
output$boxplot <- renderPlot({  
  ggplot(data.frame(Datos = unlist(rv$datos_actuales)), aes(x = Datos, y = "")) +  
    geom_boxplot(fill = "#9F82F7", color = "gray7") +  
    theme(axis.title.y = element_blank())  
})  
  
# Generar datos aleatorios  
randomData <- eventReactive(input$generateData, {  
  rnorm(input$numDataPoints)  
})  
  
real_lines_shown <- reactiveVal(FALSE)  
  
observeEvent(input$generateData, {  
  real_lines_shown(FALSE)  
})
```

```
observeEvent(input$graphType, {  
  real_lines_shown(FALSE)  
})
```

```
observeEvent(input$showRealLines, {  
  real_lines_shown(!real_lines_shown())  
})
```

```
# Crear gráfico en función del tipo seleccionado
```

```
output$mainPlot <- renderPlot({  
  data <- randomData()  
  plotType <- input$graphType  
  boxplot_color <- "#9F82F7"  
  show_boxplot <- input$showBoxplot  
  user_mean <- input$meanSlider  
  user_median <- input$medianSlider  
  show_real_lines <- real_lines_shown()  
  show_sliders <- input$showSliders  
  
  base_plot <- ggplot(data = data.frame(data), aes(x = data))  
  
  if (plotType == "Dotplot") {  
    base_plot <- base_plot +
```

```
geom_dotplot(binaxis = "x", stackdir = "center", fill = "#F35FBD") +
xlab("Valores") +
ylab("Frecuencia") +
ggtitle("Dotplot")

if (show_boxplot) {
  base_plot <- base_plot +
  geom_boxplot(data = data.frame(data, y = 0),
              aes(y = y, x = data),
              width = (max(data) - min(data)) / 4, alpha = 0.5, fill = boxplot_color,
orientation = "y")
}
} else if (plotType == "Histograma") {
  base_plot <- base_plot +
  geom_histogram(binwidth = 1, fill = "#9F82F7", color = "black") +
  xlab("Valores") +
  ylab("Frecuencia") +
  ggtitle("Histograma")

if (show_boxplot) {
  base_plot <- base_plot +
  geom_boxplot(data = data.frame(data, y = 5),
              aes(y = y, x = data),
```

```
width = (max(data) - min(data)) / 4, alpha = 0.5, fill = boxplot_color,
orientation = "y")
}
} else if (plotType == "Gráfico de probabilidad normal") {
base_plot <- ggplot(data = data.frame(data), aes(sample = data)) +
  stat_qq() +
  stat_qq_line() +
  xlab("Valores teóricos") +
  ylab("Valores de la muestra") +
  ggtitle("Gráfico de probabilidad normal")
}

if (show_sliders) {
base_plot <- base_plot +
  geom_vline(aes(xintercept = user_mean), color = "blue", linetype = "dashed") +
  geom_vline(aes(xintercept = user_median), color = "red", linetype = "dashed")
}

if (show_real_lines) {
base_plot <- base_plot +
  geom_vline(aes(xintercept = mean(data)), color = "blue") +
  geom_vline(aes(xintercept = median(data)), color = "red")
}
```

```
base_plot
```

```
)
```

```
data <- reactive({
```

```
  strsplit(input$dataInput, ",")[[1]]
```

```
)
```

```
bins <- reactive({
```

```
  as.integer(input$binsInput)
```

```
)
```

```
output$histPlot <- renderPlot({
```

```
  data_numeric <- as.numeric(data())
```

```
  breaks <- seq(min(data_numeric), max(data_numeric), length.out = bins() + 1)
```

```
  data_cut <- cut(data_numeric, breaks = breaks, include.lowest = TRUE)
```

```
  hist(data_numeric, breaks = breaks, main = "% Grasa Histograma", xlab = "%Grasa",
```

```
  ylim = c(0, 40), labels = levels(data_cut), col = "#9F82F7", border = "grey")
```

```
)
```

```
data <- reactive({
```

```
req(input$dataInput)
  strsplit(input$dataInput, ",")[[1]]
})
```

```
output$Imagenp<-renderImage({
  list(
    src="Datosordenados.png",
    contentType="image/png"
  )
},deleteFile=FALSE)
```

```
output$Imagenh <- renderImage({
  list(
    src = "Mediana.png",
    contentType = "image/png",
    width = 300,
    height = 150
  )
}, deleteFile = FALSE)
```

```
#Boxplot ----
```

```
#Mapa
```

```
#AIzaSyA5sU1IGLiLAj1hUR3puvM3GrZLRSdmXI8
```

```
output$map <- renderLeaflet({
```

```
  # Define los puntos de inicio, intermedio y fin
```

```
  start <- c(-73.1198, 7.11392) # Bucaramanga
```

```
  via <- c(-73.361389, 5.540278) # Tunja
```

```
  end <- c(-74.0721, 4.7110) # Bogotá
```

```
  # Calcula la primera ruta de Bucaramanga a Tunja
```

```
  route1 <- osrm::osrmRoute(src = start, dst = via, overview = "full")
```

```
  # Calcula la segunda ruta de Tunja a Bogotá
```

```
  route2 <- osrm::osrmRoute(src = via, dst = end, overview = "full")
```

```
  # Inicializa el mapa
```

```
  map <- leaflet() %>% addTiles()
```

```
  # Verifica si la primera ruta se calculó correctamente y añádela al mapa
```

```
if (!inherits(route1, "osrm-error")) {  
  map <- map %>% addPolylines(data = route1$geometry, color = "#9F82F7")  
}  
  
# Verifica si la segunda ruta se calculó correctamente y añádela al mapa  
if (!inherits(route2, "osrm-error")) {  
  map <- map %>% addPolylines(data = route2$geometry, color = "#9F82F7")  
}  
  
# Establece la vista del mapa para incluir todas las rutas  
map <- map %>% setView(lng = mean(c(start[1], end[1])), lat = mean(c(start[2],  
end[2])), zoom = 6)  
  
# Devuelve el mapa con las rutas añadidas  
map  
})  
  
#Termina mapa  
  
#tABLA RUTA  
output$myTable <- renderDT({  
  # Crear un data.frame con los datos para la tabla  
  data <- data.frame(  

```

```

    Destination = c("TUNJA", "BOGOTÁ"),
    Distance = c(71, 157)
)

# Renderizar la tabla sin encabezados de columna y con estilo personalizado
datatable(data,

    rownames = FALSE, # Ocultar los nombres de las filas
    colnames = c("", ""), # Eliminar los nombres de las columnas
    class = 'cell-border stripe', # Clases de estilo de DataTables
    options = list(
        dom = 't', # Solo mostrar la tabla sin otros elementos como búsqueda o
paginación
        ordering = FALSE, # Deshabilitar la ordenación de columnas
        info = FALSE, # Ocultar información del recuento de filas
        paging = FALSE, # Deshabilitar la paginación
        searching = FALSE, # Deshabilitar la búsqueda
        lengthChange = FALSE, # Deshabilitar la capacidad de cambiar el número de
filas mostradas
        preDrawCallback = JS('function() { Shiny.unbindAll(this.api().table().node());
}'),
        drawCallback = JS('function() { Shiny.bindAll(this.api().table().node()); } ')
    ))
})

```

```
#FIN TABLA RUTA
```

```
nums <- reactiveVal(c())
```

```
observeEvent(input$generate_nums, {
```

```
  random_nums <- sample(1:100, 30)
```

```
  nums(random_nums)
```

```
  updateTextAreaInput(session, "user_nums", value = paste(random_nums, collapse = ",
```

```
"))
```

```
})
```

```
# Insertar los datos en el editor
```

```
observeEvent(input$insert_data, {
```

```
  if (length(nums()) > 0) {
```

```
    user_data <- paste(sort(nums(), decreasing = FALSE), collapse = ", ") # Ordena los
```

```
datos
```

```
    updateAceEditor(session, "code3",
```

```
      value = paste0('# Crea algunos datos
```

```
datos <- c(', user_data, ')
```

```
# Calcula los cuartiles
```

```
cuartiles <- quantile(datos, probs=c(0.25, 0.5, 0.75))

# Imprime los cuartiles
print(cuartiles))
}
})

# Calculate quartiles
cuartiles <- reactive({
  if (length(nums()) > 0) {
    num_data <- sort(as.numeric(unlist(strsplit(input$user_nums, ","))), decreasing =
FALSE)
    return(quantile(num_data, probs=c(0.25, 0.5, 0.75)))
  } else {
    return(NULL)
  }
})

observeEvent(input$run_code3, {
  result <- cuartiles()
  output$code3_result <- renderUI({
    if (!is.null(result)) {
      HTML(paste0(
```

```

"Los cuartiles de la muestra son:<br>",
"<b>Q1</b> = ", result["25%"], "<br>",
"<b>Q2</b> = ", result["50%"], "<br>",
"<b>Q3</b> = ", result["75%"]
))
}
})
# Show "Visualizar cuartiles" button
output$quartiles_button <- renderUI({
  actionButton("show_quartiles", "Visualizar cuartiles")
})
})

observeEvent(input$show_quartiles, {
  output$quartiles_visualization <- renderUI({
    if (!is.null(quartiles())) {
      num_list <- paste(sort(nums(), decreasing = FALSE), collapse = ", ")
      for (q in c(quartiles()["25%"], quartiles()["50%"], quartiles()["75%"])) {
        num_list <- gsub(paste0("\\b", q, "\\b"), paste0("<b>", q, "</b>"), num_list)
      }
      HTML(num_list)
    }
  })
})

```

```
output$quartiles_table <- renderUI({
  if (length(nums()) > 0) {
    num_data <- sort(as.numeric(unlist(strsplit(input$user_nums, ","))), decreasing =
FALSE)

    quartile_1 <- floor(length(num_data)/4)
    quartile_2 <- 2*quartile_1
    quartile_3 <- 3*quartile_1

    quartile_groups <- rep(1:4, times = c(quartile_1, quartile_1, quartile_1,
length(num_data) - 3*quartile_1))

    df <- data.frame("Dato" = num_data, "Cuartil" = quartile_groups)

    # Encuentra la última fila con "Cuartil" = 1
    last_row <- max(which(df$Cuartil == 1))+1

    # Divide el dataframe en dos partes
    df1 <- df[1:last_row, ]
    df2 <- df[(last_row+1):nrow(df), ]

    # Crea una nueva fila
    new_row <- data.frame("Dato" = "<b>Q1</b>", "Cuartil" = "")

    # Une las partes y la nueva fila
```

```
df <- rbind(df1, new_row, df2)
```

```
# Encuentra la última fila con "Cuartil" = 2
```

```
last_row <- max(which(df$Cuartil == 2))+1
```

```
# Divide el dataframe en dos partes
```

```
df1 <- df[1:last_row, ]
```

```
df2 <- df[(last_row+1):nrow(df), ]
```

```
# Crea una nueva fila
```

```
new_row <- data.frame("Dato" = "<b>Q2</b>", "Cuartil" = "")
```

```
# Une las partes y la nueva fila
```

```
df <- rbind(df1, new_row, df2)
```

```
# Encuentra la última fila con "Cuartil" = 3
last_row <- max(which(df$Cuartil == 3))+1

# Divide el dataframe en dos partes
df1 <- df[1:last_row, ]
df2 <- df[(last_row+1):nrow(df), ]

# Crea una nueva fila
new_row <- data.frame("Dato" = "<b>Q3</b>", "Cuartil" = "")

# Une las partes y la nueva fila
df <- rbind(df1, new_row, df2)

# Elimina la columna "Cuartil"
df$Cuartil <- NULL

# Renderiza la tabla como HTML
HTML(knitr::kable(df, "html", escape = FALSE, row.names = FALSE))
```

```
}  
})  
})
```

```
observeEvent(input$show_boxplot2, {  
  output$boxplotp <- renderPlot({  
    # Asegúrate de que nums() contenga los números que quieres graficar  
    if (length(nums()) > 0) {  
      df <- data.frame(nums = as.numeric(unlist(strsplit(input$user_nums, ","))))  
  
      quartiles <- quantile(df$nums, probs=c(0.25, 0.5, 0.75))  
  
      ggplot(df) +  
        # Caja  
        geom_rect(aes(ymin = 0.85, ymax = 1.15, xmin = quartiles["25%"], xmax =  
quartiles["75%"]), fill = "#9F82F7", colour = "gray") +  
        # Línea mediana
```

```
geom_segment(aes(y = 0.85, yend = 1.15, x = quartiles["50%"], xend =
quartiles["50%"])) +
  theme_minimal() +
  theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank()) +
  labs(x = NULL, y = NULL) # Remueve las etiquetas de los ejes
}
})
})

observeEvent(input$show_boxplot_whiskers, {
  output$boxplot_whiskers <- renderPlot({
    if (length(nums()) > 0) {
      # Para rellenar las cajas con color usamos `fill`, y `border` para los bordes
      boxplot(nums(), horizontal = TRUE, outline = FALSE, xlab = "", ylab = "",
              col = "#9F82F7", border = "gray7")
    }
  }, height = 200, width = 500)
})
```

```
#boton reinicio

observeEvent(input$reset_button, {

  # Restablecer las entradas

  updateTextAreaInput(session, "user_nums", value = "")

  # Restablecer las salidas

  output$boxplotp <- renderPlot({NULL})

  output$boxplot_whiskers <- renderPlot({NULL})

  # Restablecer cualquier otro input/output según sea necesario

})

#Actividad selección múltiple

output$boxplot1 <- renderPlot({

  boxplot(rnorm(100, mean = 0, sd = 1), main = "Gráfico 1",col="#9F82F7",

border="gray7")

})

output$boxplot2 <- renderPlot({

  boxplot(rexp(100, rate = 1), main = "Gráfico 2",col="#9F82F7", border="gray7")
```

```
    })

    output$boxplot3 <- renderPlot({
      boxplot(rnorm(100, mean = 0, sd = 2), main = "Gráfico 3",col="#9F82F7",
border="gray7")
    })

    output$boxplot4 <- renderPlot({
      boxplot(rexp(100, rate = 2), main = "Gráfico 4",col="#9F82F7", border="gray7")
    })

    observeEvent(input$submit, {
      if (input$choice == "Gráfico 1") {
        output$result <- renderText("¡Correcto!")
      } else {
        output$result <- renderText("Incorrecto, intenta de nuevo.")
      }
    })

#BOXPLOT CON DOTPLOT

    output$example <- renderText({
```

```

"1,2,2,3,3,3,4,4,4,4,5,5,5,5,5"

})

user_data <- reactive({

  if(input$user_data == "") return(NULL)

  strsplit(input$user_data, ",")[[1]] %>% as.numeric() %>% data.frame(Value = .)

})

output$plot <- renderPlot({

  if(is.null(user_data())) return(NULL)

  ggplot(user_data(), aes(x = factor(1), y = Value)) +

    geom_dotplot(binaxis = "y", stackdir = "up", dotsize = 0.5, aes(x =

factor(0.9)),fill="#9F82F7", color= "#9F82F7") +

    geom_boxplot(alpha = 0.4, fill = "#9F82F7", color = "gray7") +

    coord_flip() +

    theme(axis.title.x = element_blank())

})

```

#Boxplot comparacion excel y otro medio

```

output$boxplotejem <- renderPlot({

  ggplot(data.frame(x = c(0,0,1,1,1,2,3,3,4,9)), aes(x = x)) +

    geom_boxplot(fill = "#9F82F7", color = "gray7") +

```

```
coord_flip() +  
theme_minimal() +  
labs(title = "Gráfico de caja y bigotes", x = "", y = "")  
})
```

```
#Boxplot y linea de media
```

```
# Observamos el botón para generar datos aleatorios
```

```
observeEvent(input$random_button, {
```

```
  # Generamos una muestra aleatoria de 100 números enteros entre 1 y 100
```

```
  random_sample <- sample(1:100, 100, replace = TRUE)
```

```
  # Actualizamos el campo de texto con los nuevos datos
```

```
  updateTextInput(session, "data_input", value = paste(random_sample, collapse = ","))
```

```
}, ignoreNULL = FALSE)
```

```
# Observamos el campo de texto y la checkbox para mostrar la línea de media
```

```
output$boxplot_output1 <- renderPlot({
```

```
  # Si el campo de texto no está vacío, graficamos el boxplot
```

```
  if(input$data_input != "") {
```

```
    # Convertimos la cadena de texto a números
```

```
    data_vector <- as.integer(unlist(strsplit(input$data_input, ",")))
```

```
    boxplot(data_vector, main = "Boxplot", ylab = "Valores", col="#9F82F7",  
border="gray7")
```

```
# Si la checkbox está seleccionada, añadimos la línea de la media
if(input$show_mean) {
  abline(h = mean(data_vector), col = "red", lwd = 2)
}
}
})

#Ejercicios

output$data_plot2 <- renderPlot({
  data_input2 <- input$data_input2

  # Verificar si los datos están separados por comas
  if (grepl(",", data_input2)) {
    data_vector <- as.numeric(unlist(strsplit(data_input2, ",")))
  } else {
    # Si los datos están separados por saltos de línea, se reemplazan por comas
    data_input2 <- gsub("\\s+", ",", data_input2)
    data_vector <- as.numeric(unlist(strsplit(data_input2, ",")))
  }

  # Verificar si los datos son válidos antes de intentar generar un gráfico
```

```
if (all(is.na(data_vector)) || length(data_vector) == 0) {  
  return()  
}  
  
if (input$plot_type2 == "QQ Plot") {  
  qqnorm(data_vector, main="QQ Plot", xlab="Teóricos", ylab="Muestra", pch=19,  
col="deepskyblue4")  
  qqline(data_vector, col="deepskyblue3", lwd=2)  
} else if (input$plot_type2 == "Histograma") {  
  ggplot(data.frame(data_vector=data_vector), aes(x=data_vector)) +  
    geom_histogram(color="deepskyblue4", fill="deepskyblue1") +  
    theme_minimal() +  
    labs(title="Histograma", x="Datos", y="Frecuencia")  
} else if (input$plot_type2 == "Boxplot") {  
  boxplot(data_vector, horizontal = TRUE, col="deepskyblue1", main="Boxplot")  
}  
})  
  
#Consola r para calcular cuartiles  
observeEvent(input$execute, {  
  code5 <- input$code4  
  
  tryCatch({
```

```
    output$results <- renderPrint({
      eval(parse(text = code5))
    })
  }, error = function(e) {
    output$results <- renderPrint({
      print(paste("Error en el código: ", e$message))
    })
  })
})

# Observar eventos de scroll para mostrar/ocultar el botón de volver arriba
runjs("
$(window).scroll(function() {
  if ($(this).scrollTop() > 50) {
    $('#back_to_top').fadeIn();
  } else {
    $('#back_to_top').fadeOut();
  }
});

$('#back_to_top').click(function() {
  $('body,html').animate({ scrollTop : 0}, 500);
  return false;
});
```

```
"")
```

```
}
```

```
# Ejecutar la aplicación
```

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