



# **HERRAMIENTA DE VISUALIZACIÓN DE DATOS PROVENIENTES DEL ANÁLISIS MASIVO DE IMÁGENES MÉDICAS EN LA PLATAFORMA DE QUIBIM PRECISION®**

**Jorge Javier Galarza Zambrano**

**Tutor: Ignacio Bosch Roig**

**Cotutor: Ana Jiménez Pastor**

Trabajo Fin de Grado presentado en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Graduado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Curso 2018-19

Valencia, 6 de septiembre de 2019



*Dedicado a mis padres, a mi hermano, a mi amigo peludo  
y a todos los compañeros y amigos  
que han estado conmigo estos años.*



## **Agradecimientos**

Me gustaría agradecer a Ana por su ayuda y paciencia durante estos meses en el desarrollo de este TFG, y a Ángel y a Nacho por ofrecerme esta oportunidad para poder llevarlo a cabo con QUIBIM. También quisiera agradecer a Edu y a Alejandro del equipo de QUIBIM por su colaboración en este proyecto.

Finalmente, quiero agradecer a mis padres por el apoyo incondicional y la tranquilidad que me han transmitido durante todos estos años.

Muchas gracias a todos.



## Resumen

La plataforma de QUIBIM Precision® incluye una herramienta de visualización de datos o *Data Miner*. Esta herramienta permite al usuario realizar un análisis estadístico de cada uno de los biomarcadores de imagen analizados previamente con el objetivo de, dada una población, extraer conclusiones con valor diagnóstico, pronóstico y terapéutico.

Previamente a la realización de este trabajo, el *Data Miner* incluía representaciones tales como un diagrama de dispersión y un diagrama de cajas, desarrolladas empleando el framework<sup>1</sup> *AngularJS* como directiva para la librería *angular-nvd3.js* de JavaScript, basada en *D3.js*. Esta librería nos permite trabajar con grandes bloques de datos estructurados en formato JSON para la representación de gráficos estadísticos. Por lo tanto, el principal objetivo del proyecto consistiría en el desarrollo de nuevos modos de visualización utilizando la librería que hemos indicado, con la finalidad de ampliar la variedad de visualizaciones enfocadas al entorno médico y, ofrecer una interfaz accesible al usuario que pueda ser utilizado sin necesidad de un manual.

## Resum

La plataforma de QUIBIM Precision® inclou una eina de visualització de dades o *Data Miner*. Aquesta eina permet a l'usuari realitzar una anàlisi estadística de cada un dels biomarcadors d'imatge analitzats prèviament amb l'objectiu de, donada una població, extreure conclusions amb valor diagnòstic, pronòstic i terapèutic.

Prèviament a la realització d'aquest treball, el *Data Miner* incloïa representacions com ara un diagrama de dispersió i un diagrama de caixes, desenvolupades emprant el framework *AngularJS* com a directiva per la llibreria *angular-nvd3.js* de JavaScript, basada en *D3.js*. Aquesta llibreria ens permet treballar amb grans blocs de dades estructurades en format JSON per a la representació de gràfics estadístics. Per tant, el principal objectiu del projecte consistiria en el desenvolupament de noves maneres de visualització utilitzant la llibreria que hem indicat, amb la finalitat d'ampliar la varietat de visualitzacions enfocades a l'entorn mèdic i, oferir una interfície accessible a l'usuari que pugui ser utilitzat sense necessitat d'un manual.

## Abstract

QUIBIM Precision® platform includes a data visualization tool or *Data Miner*. This tool allows the user to perform a statistical analysis of each of the imaging biomarkers previously analyzed in order to, given a population, draw conclusions with diagnostic, prognostic and therapeutic value.

Prior to the implementation of this project, the *Data Miner* included representations such as a scatter plot and a box plot, developed using the framework *AngularJS* as a directive for the *Angular-nvd3.js* JavaScript library, based on *D3.js*. This library allows us to work with large blocks of data structured in JSON format for the representation of statistical graphs. Therefore, the main objective of the project is the development of new display modes using the library that we have indicated, with the goal of extending the variety of visualization tools focused on the medical environment and offer an accessible user interface that can be used without the need of a user manual.

---

<sup>1</sup> Framework: Se trata de un entorno de trabajo que sirve de base para la organización y desarrollo de software.



## Índice

Capítulo 1.	Introducción .....	7
1.1	Visualización de datos.....	7
1.1.1	Herramientas .....	8
1.1.2	Visualización de datos de imágenes médicas .....	8
1.2	QUIBIM Precision® .....	9
1.2.1	Biomarcador de imagen .....	9
1.2.2	Análisis de imagen a través de la plataforma QUIBIM Precision®.....	11
1.3	QUIBIM Precision® <i>Data Miner</i> .....	11
1.3.1	Modalidades de imagen.....	12
1.3.2	Biomarcadores de imagen .....	14
1.3.3	Herramienta de visualización .....	20
Capítulo 2.	Motivación .....	25
2.1	Objetivos .....	25
2.2	Estructura de la memoria.....	26
Capítulo 3.	Metodología .....	27
3.1	Gestión del proyecto.....	27
3.2	Distribución en tareas.....	27
3.2.1	Familiarización con el entorno .....	27
3.2.2	Desarrollo de los modos de visualización .....	27
3.2.3	Testeo de la plataforma .....	27
Capítulo 4.	Contexto del proyecto .....	28
4.1	Conceptos.....	28
4.1.1	DICOM .....	28
4.1.2	DOM .....	28
4.1.3	JSON .....	28
4.1.4	Canvas .....	28
4.1.5	SVG.....	29
4.1.6	GitHub.....	29
4.2	Lenguajes utilizados.....	29
4.2.1	CoffeeScript .....	29
4.2.2	Jade.....	30
4.2.3	Less .....	32
4.3	Librerías de visualización de datos .....	32
4.3.1	D3.js .....	32
4.3.2	Angular-nvD3.js.....	33



Capítulo 5. Desarrollo y resultados.....	34
5.1 Estructura y sintaxis .....	34
5.1.1 prepareData .....	34
5.1.2 buildData .....	35
5.1.3 buildOptions .....	35
5.2 Diagrama de dispersión.....	35
5.2.1 Desarrollo.....	37
5.2.2 Resultado.....	39
5.3 Diagrama de cajas .....	40
5.3.1 Desarrollo .....	40
5.3.2 Resultado.....	41
5.4 Gráfico circular .....	41
5.4.1 Desarrollo .....	42
5.4.2 Resultado.....	43
5.5 Diagrama de barras.....	43
5.5.1 Selección múltiple .....	43
5.5.2 Bidimensional.....	45
5.6 Histograma .....	47
5.6.1 Datos múltiples.....	47
5.6.2 Datos agrupados .....	50
5.7 Bland-Altman .....	52
5.7.1 Desarrollo .....	52
5.7.2 Resultado.....	55
5.8 Kaplan-Meier .....	55
5.8.1 Desarrollo .....	56
5.8.2 Resultado.....	60
5.9 Mapa de calor .....	61
5.9.1 Desarrollo .....	61
5.9.2 Resultado.....	62
5.10 Paleta de colores.....	63
5.11 Captura de la gráfica .....	64
5.11.1 Librerías .....	64
5.11.2 Desarrollo y resultado .....	65
Capítulo 6. Conclusiones .....	66
6.1 Resultado del proyecto.....	66
6.2 Planes futuros .....	66
Bibliografía .....	67



## Índice de tablas

Tabla 1. Resultados de volumetría cerebral con parcelación. [7].....	14
Tabla 2. Resultados de las lesiones de sustancia blanca. [7].....	15
Tabla 3. Resultados de concentración de hierro y grasa. [7].....	15
Tabla 4. Resultados de microarquitectura 3D del hueso. [7] .....	16
Tabla 5. Resultados de perfusión – Modelado farmacocinético (DCE-MRI). [7] .....	17
Tabla 6. Resultados de perfusión – Análisis semicuantitativo (DCE-MRI). [7].....	17
Tabla 7. Resultados de imágenes ponderadas por difusión (DWI). [7].....	18
Tabla 8. Resultados del modelo de IVIM. [7].....	18
Tabla 9. Resultados de relajometría – mapeado por parámetros T1 y T2. [7] .....	19
Tabla 10. Resultados de enfisema pulmonar. [7] .....	19
Tabla 11. Resultados de baja densidad funcional pulmonar. [7].....	20
Tabla 12. Cálculo de la probabilidad de supervivencia.....	57
Tabla 13. Cálculo del intervalo de confianza. ....	58



## Índice de figuras

Figura 1. Interfaz de Tableau. [2].....	8
Figura 2. Flujo de desarrollo de los biomarcadores de imagen. [4] .....	9
Figura 3. Análisis avanzado de cáncer de próstata.....	10
Figura 4. Vista de QUIBIM Precision® <i>Data Miner</i> . .....	11
Figura 5. Vista de la tabla del <i>Data Miner</i> . .....	12
Figura 6. Sección transversal y sagital de una resonancia cerebral. [5] .....	13
Figura 7. Radiografía de tórax. [6].....	13
Figura 8. Menú desplegable de la herramienta de visualización.....	20
Figura 9. Ejemplo del diagrama de dispersión.....	21
Figura 10. Ejemplo del diagrama de cajas. ....	21
Figura 11. Ejemplo de gráfico circular.....	21
Figura 12. Ejemplo de diagrama de barras de selección múltiple.....	22
Figura 13. Ejemplo del diagrama de barras bidimensional. ....	22
Figura 14. Ejemplo del histograma con selección de múltiples datos.....	22
Figura 15. Ejemplo del histograma con selección por datos agrupados.....	23
Figura 16. Ejemplo del método de Bland-Altman. ....	23
Figura 17. Ejemplo del gráfico Kaplan-Meier. ....	24
Figura 18. Ejemplo del diagrama de calor. ....	24
Figura 19. Ejemplo del registro de algunos cambios realizados en el proyecto.....	29
Figura 20. Diferencia entre los operadores en JavaScript y CoffeeScript.....	30
Figura 21. Inspector de página de Google Chrome.....	31
Figura 22. Gráficos reutilizables de <i>Angular-nvD3.js</i> . [8] .....	33
Figura 23. Propiedades de los diagramas disponibles.....	34
Figura 24. Objeto proveniente de la base de datos.....	34
Figura 25. Muestra del bloque de datos preparado. ....	34
Figura 26. Datos listos del diagrama de dispersión.....	35
Figura 27. Opciones del diagrama de dispersión. ....	36
Figura 28. Advertencia de la regresión potencial.....	36
Figura 29. Opciones de la regresión polinomial.....	36
Figura 30. Registro de los datos preparados del diagrama de dispersión.....	37
Figura 31. Datos listos para ser representados del diagrama de dispersión. ....	37
Figura 32. Plantilla del registro para el diagrama de dispersión. ....	37
Figura 33. Datos del diagrama de dispersión y la línea de tendencia (Regresión lineal).....	37
Figura 34. Propiedad con los datos de la regresión logarítmica.....	38
Figura 35. Propiedad con los datos de la regresión potencial. ....	38



Figura 36. Propiedad con los datos de la regresión exponencial.....	38
Figura 37. Propiedad con los datos de la regresión polinomial.....	38
Figura 38. Registro con la propiedad de grupo del diagrama de dispersión. ....	39
Figura 39. Bloques diferenciados por grupos del diagrama de dispersión.....	39
Figura 40. Resultado del diagrama de dispersión.....	39
Figura 41. Opciones del diagrama de cajas.....	40
Figura 42. Muestra de los datos del diagrama de cajas. ....	40
Figura 43. Plantilla de los datos listos para representar el diagrama de cajas.....	40
Figura 44. Resultado del diagrama de cajas. ....	41
Figura 45. Opciones del gráfico circular. ....	41
Figura 46. Rangos predeterminados del gráfico circular.....	42
Figura 47. Propiedad con los valores de los cálculos realizados en el rango.....	42
Figura 48. Resultado del gráfico circular. ....	43
Figura 49. Opciones del diagrama de barras de selección múltiple. ....	43
Figura 50. Objetos de los datos seleccionados en el diagrama de barras de selección múltiple. ....	44
Figura 51. Datos ordenados para el diagrama de barras de selección múltiple.....	44
Figura 52. Plantilla de los datos listos para el diagrama de barras de selección múltiple. ....	44
Figura 53. Resultado del diagrama de barras de selección múltiple. ....	45
Figura 54. Opciones del diagrama de barras bidimensional. ....	45
Figura 55. Muestra de un dato preparado para el diagrama de barras bidimensional. ....	46
Figura 56. Plantilla de los datos listos para representar el diagrama de barras bidimensional....	46
Figura 57. Objeto con los datos de la barra seleccionada. ....	46
Figura 58. Resultado del diagrama de barras bidimensional.....	47
Figura 59. Opciones del histograma con selección de múltiples datos. ....	47
Figura 60. Datos ordenados para el histograma de selección múltiple. ....	48
Figura 61. Muestra con las propiedades del histograma de selección múltiple. ....	48
Figura 62. Plantilla de los datos listos para representar el histograma de selección múltiple. ....	49
Figura 63. Resultado del histograma con selección de múltiples datos. ....	49
Figura 64. Opciones del histograma con selección por datos agrupados. ....	50
Figura 65. Datos ordenados para el histograma por datos agrupados. ....	50
Figura 66. Muestra con las propiedades del histograma por datos agrupados. ....	51
Figura 67. Plantilla de los datos listos para representar el histograma por datos agrupados.....	51
Figura 68. Resultado del histograma por datos agrupados (Agrupado). ....	51
Figura 69. Resultado del histograma por datos agrupados (Apilado). ....	52
Figura 70. Opciones del método de Bland-Altman.....	52
Figura 71. Plantilla de los datos listos para representar el gráfico Bland-Altman. ....	52
Figura 72. Estructura de un gráfico Bland-Altman. [10] .....	53



Figura 73. Elemento <i>Scatter</i> del Bland-Altman. ....	53
Figura 74. Elemento <i>Bias</i> del Bland-Altman. ....	53
Figura 75. Elemento <i>Zero</i> del Bland-Altman. ....	54
Figura 76. Elemento <i>Limit Plus</i> del Bland-Altman. ....	54
Figura 77. Elemento <i>Limit Minus</i> del Bland-Altman. ....	54
Figura 78. Resultado del método de Bland-Altman. ....	55
Figura 79. Opciones del gráfico Kaplan-Meier. ....	55
Figura 80. Muestra de los datos preparados para el análisis de supervivencia (Registro). ....	56
Figura 81. Plantilla de los datos listos para representar el análisis de supervivencia (Registro). ....	56
Figura 82. Elemento <i>Kaplan-Meier</i> del análisis de supervivencia. ....	56
Figura 83. Muestra de los valores representados en el análisis de supervivencia (Registro). ....	57
Figura 84. Elemento <i>+95% CI</i> del análisis de supervivencia. ....	57
Figura 85. Elemento <i>-95% CI</i> del análisis de supervivencia. ....	58
Figura 86. Elemento <i>50%</i> del análisis de supervivencia. ....	59
Figura 87. Muestra de los datos del análisis de supervivencia (Sexo del paciente). ....	59
Figura 88. Plantilla de los datos del análisis de supervivencia (Sexo del paciente). ....	59
Figura 89. Muestra de los valores del análisis de supervivencia (Sexo del paciente). ....	59
Figura 90. Resultado del gráfico Kaplan-Meier (Registro). ....	60
Figura 91. Resultado del gráfico Kaplan-Meier (Sexo del paciente). ....	60
Figura 92. Opciones del diagrama de calor. ....	61
Figura 93. Repositorio <i>heatmap</i> del usuario <i>pieterprovoost</i> . ....	61
Figura 94. Muestra de los datos preparados para el mapa de calor. ....	62
Figura 95. Parámetros del mapa de calor. ....	62
Figura 96. Opciones del mapa de calor. ....	62
Figura 97. Resultado del mapa de calor. ....	62
Figura 98. Opciones de selección de paleta. ....	63
Figura 99. Paleta de colores de la escala <i>ColorBrewer</i> . [12] ....	63
Figura 100. Ejemplo del objeto al seleccionar una paleta. ....	64
Figura 101. Botón de captura de la gráfica. ....	64
Figura 102. Resultado de la captura de la gráfica. ....	65



## Capítulo 1. Introducción

Este proyecto final de grado ha consistido en el desarrollo de una herramienta de visualización de datos para la plataforma QUIBIM Precision® como proyecto para la empresa QUIBIM SL durante el desarrollo de las prácticas.

QUIBIM S.L. se creó en octubre de 2012 como una iniciativa de un grupo de Ingenieros Biomédicos y Médicos con amplia experiencia y carreras científicas reconocidas en el campo de la imagen médica. El nombre **QUIBIM** proviene de **Quantitative Imaging Biomarkers In Medicine** y como empresa innovadora, está especialmente dedicada al procesado de imagen médica y la extracción de biomarcadores de imagen en el flujo de trabajo radiológico. La empresa nació como un ‘spin-off’ del Instituto de Investigación Sanitaria La Fe en Valencia y es una de las empresas emergentes seleccionadas en la 3ª edición del programa Lanzadera, impulsado por Juan Roig. QUIBIM tiene el objetivo de liderar un programa de calidad y rentabilidad radiológica en las vertientes profesional y científica del procesado de la imagen médica y la cuantificación.

La aplicación de técnicas avanzadas de análisis y procesado de imagen permiten la extracción de parámetros y características de las imágenes médicas, conocidos como biomarcadores de imagen. Estos biomarcadores de imagen aportan información adicional a los informes radiológicos proporcionando información objetiva con un elevado valor diagnóstico, pronóstico y terapéutico.

QUIBIM ofrece un servicio integral para la extracción de información cuantitativa de las imágenes radiológicas con el objetivo de mejorar los procesos de diagnóstico de enfermedades con alta incidencia y evaluar adecuadamente los cambios que producen los tratamientos farmacológicos en el organismo. Para poder integrar estos biomarcadores en uso clínico QUIBIM ha desarrollado **QUIBIM Precision®** la primera Plataforma Cloud que permite la gestión automatizada de estudios de imagen médica y el lanzamiento de análisis por parte del usuario. QUIBIM Precision® devuelve para cada análisis realizado un informe estructurado con los resultados más significativos.

Además, como empresa experta en imagen médica, ofrece servicios de consultoría centrados en la **optimización de la adquisición** de las imágenes a aquellos clientes en los que las imágenes médicas no cumplan con los criterios de calidad mínimos requeridos para la extracción de los biomarcadores de imagen, y el asesoramiento sobre las diferentes estrategias de **análisis estadístico** de los datos, con el objetivo de minimizar sesgos y aumentar la significación de los resultados.

### 1.1 Visualización de datos

En los últimos años, debido al crecimiento exponencial de la cantidad de datos que se generan día a día, surgen necesidades, cada vez más importantes, para el tratamiento de estos datos. De entre todas las técnicas que comprenden el mundo de la ciencia de datos, la visualización de datos nos permite, de una manera sencilla y eficaz, conocer cómo son los datos con los que estamos trabajando, paso fundamental para el posterior desarrollo de modelos predictivos.

La visualización de datos se basa, fundamentalmente, en representaciones capaces de comunicar la relación de unos datos. Para una comunicación clara y eficiente, la visualización de datos utiliza gráficos estadísticos, diagramas y otras herramientas de la misma índole. Una visualización efectiva ayuda a los usuarios a analizar y razonar sobre los datos representados, haciendo que datos complejos sean más accesibles, comprensibles y utilizables.

Sin embargo, la visualización de datos es una herramienta que ayuda al análisis, pero no es un sustituto de la habilidad analítica ni a las estadísticas. Disponer de una buena herramienta de visualización de datos no es suficiente, también es necesario disponer de conocimientos de negocio, estadística, teoría del color, composición gráfica e inteligencia emocional [1].

Las visualizaciones de datos generalmente se basan en herramientas con plantillas. Pueden basarse en herramientas online gratuitas o de pago, o incluso en entornos de trabajo de código abierto. Propondremos un ejemplo para cada caso.

Dada la importancia que ha adquirido la visualización de datos, varias son las herramientas que se han desarrollado para ello con el objetivo de facilitar esta labor a los usuarios.

### 1.1.1 Herramientas

#### 1.1.1.1 Tableau

Tableau es un software de visualización de datos ampliamente utilizado por los usuarios debido a la facilidad de uso de todas sus funciones. Existe una versión de pago, pero también se puede utilizar de forma gratuita, con algunas limitaciones. Tableau se especializa en técnicas de visualización para analizar y explorar bases de datos enfocados en inteligencia empresarial.

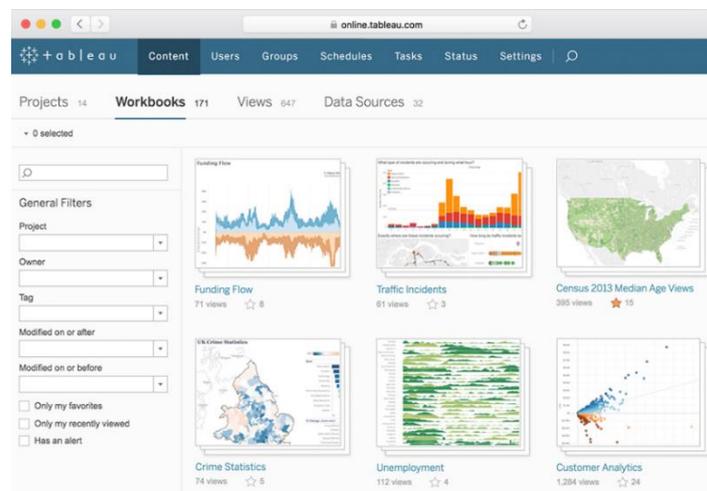


Figura 1. Interfaz de Tableau. [2]

#### 1.1.1.2 Plotly

Plotly ofrece visualizaciones más complejas y sofisticadas que Tableau. Está desarrollado sobre *D3.js* y *stack.gl*. A través de GitHub, ofrece librerías de alto nivel para la representación de gráficos científicos en Python, JavaScript, R, MATLAB, Perl, Julia, Arduino.

### 1.1.2 Visualización de datos de imágenes médicas

La obtención de imágenes médicas, como la resonancia magnética o la tomografía computarizada, se utilizan de forma rutinaria en radiología para la evaluación y el tratamiento de enfermedades.

Los avances técnicos en la adquisición de imágenes han permitido una nueva visualización en los tejidos y partes de los órganos que antes apenas eran accesibles. Todo esto ha permitido el desarrollo de modelos computacionales avanzados para la extracción de información cuantitativa de la imagen (biomarcadores de imagen).

La radiómica es un novedoso campo de estudio médico que tiene como objetivo extraer características cuantitativas de imágenes médicas utilizando algoritmos de caracterización de datos. Estas características, denominadas **características radiómicas**, tienen el potencial de descubrir características de la enfermedad que no se pueden apreciar a simple vista. Estos procedimientos pueden ser útiles en términos de respuesta terapéutica, pronóstico y para diseñar un tratamiento personalizado para patologías concretas. [3]

Por tanto, estas características dan lugar a unos datos que podemos evaluar y utilizar en la visualización de datos de la herramienta en que consiste este proyecto desarrollado para la plataforma QUIBIM Precision®.

## 1.2 QUIBIM Precision®

La plataforma QUIBIM Precision®, desarrollada por QUIBIM, es una aplicación web para la visualización y el análisis no invasivo de imágenes médicas basadas en el estándar de comunicación DICOM<sup>2</sup>.

El uso de la plataforma está enfocado a especialistas de la salud, con el objetivo de medir los cambios producidos por una enfermedad o un tratamiento en tejidos y órganos, ofreciendo información adicional cuantitativa en complemento a la evaluación cualitativa realizada por el especialista.

La aplicación consiste en un software de procesamiento de imagen diseñado para funcionar en un navegador, de manera que, pese a estar enfocado a especialistas, su uso no es limitado para médicos o tecnólogos. Se puede utilizar un equipo común para realizar visualización, procesamiento y análisis de imágenes médicas.

En QUIBIM Precision® están integrados los diferentes módulos de análisis que permiten el análisis cuantitativo de las imágenes médicas en diferentes aplicaciones clínicas, distribuidas en cinco especialidades diferentes: oncología, neuroimagen, abdomen, tórax e imagen musculoesquelética. Además, gracias al protocolo DICOM, la plataforma es interoperable con la mayoría de los sistemas de salud. La plataforma ofrece una amplia gama de funcionalidades. Las características principales del software son:

- Lectura de estudios DICOM.
- Anonimización, almacenamiento y visualización de imágenes DICOM.
- Análisis de imágenes a través de distintos módulos de análisis:
  - Módulo de volumetría cerebral con parcelación.
  - Módulo de concentración de hierro y grasa en el hígado.
  - Módulo de morfometría del hueso trabecular y análisis mecánico.
  - Módulos de perfusión semicuantitativa y farmacocinética.
  - Módulos de difusión.
  - Módulo de postprocesado de relajometría.
  - Módulos de enfisema pulmonar y baja densidad funcional pulmonar.

### 1.2.1 Biomarcador de imagen

Hasta ahora hemos mencionado los biomarcadores de imagen en varias ocasiones, de manera que en este apartado vamos a hacer énfasis en la explicación de este término.

Un biomarcador de imagen se puede definir como una característica extraída de las imágenes médicas radiológicas adquiridas de un paciente mediante técnicas computacionales. Puede medirse de manera objetiva y se comporta como un indicador para un proceso biológico normal, una enfermedad o la respuesta a un tratamiento.

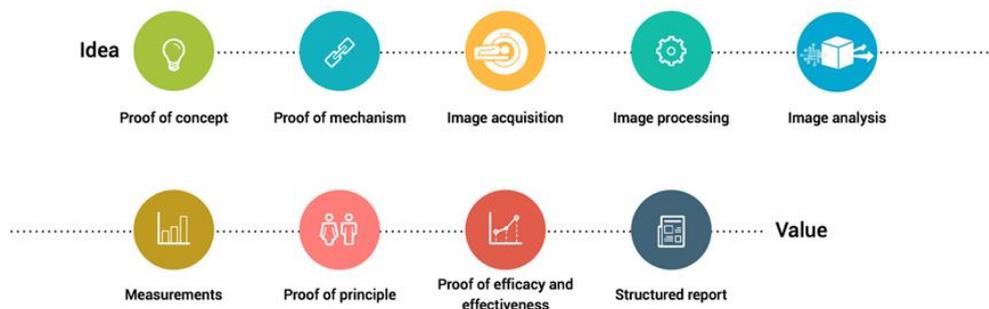


Figura 2. Flujo de desarrollo de los biomarcadores de imagen. [4]

<sup>2</sup> Digital Imaging and Communications in Medicine.

En la Figura 2 podemos observar el flujo óptimo para la cuantificación de los biomarcadores de imagen:

1. **Prueba de concepto.** Detecta una necesidad clínica relevante para la enfermedad que puede estudiarse utilizando las modalidades de imágenes disponibles y las técnicas de procesamiento de imágenes.
2. **Prueba del mecanismo.** Explica la relación entre este cambio y la presencia, grado y respuesta al tratamiento de la enfermedad.
3. **Adquisición de imágenes.** Las imágenes necesarias para extraer el biomarcador deben ser apropiadas en términos de resolución espacial, relación señal a ruido y contraste.
4. **Preprocesamiento de imágenes.** Consiste en mejorar la calidad de la imagen mediante procesos de segmentación, filtrado o interpolación.
5. **Análisis y modelado de la señal.** Ya sea por métodos de análisis estructurales o funcionales, permiten extraer información relevante.
6. **Técnicas de medición.** Mediante imágenes paramétricas obtenidas a partir del modelado estadístico de los parámetros más relevantes.
7. **Prueba de principio.** Se realiza una prueba en una muestra controlada de sujetos para validar el método y evaluar posibles sesgos relacionados con la edad, el sexo o cualquier otra naturaleza.
8. **Pruebas de eficacia y efectividad.** Se realizan en grupos de pacientes más grandes y mejor definidos. Mostrarán la capacidad de un biomarcador para medir el ensayo clínico.

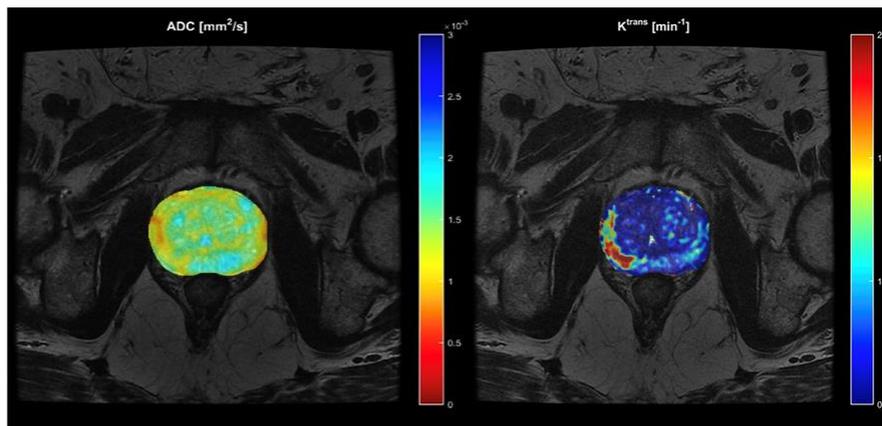


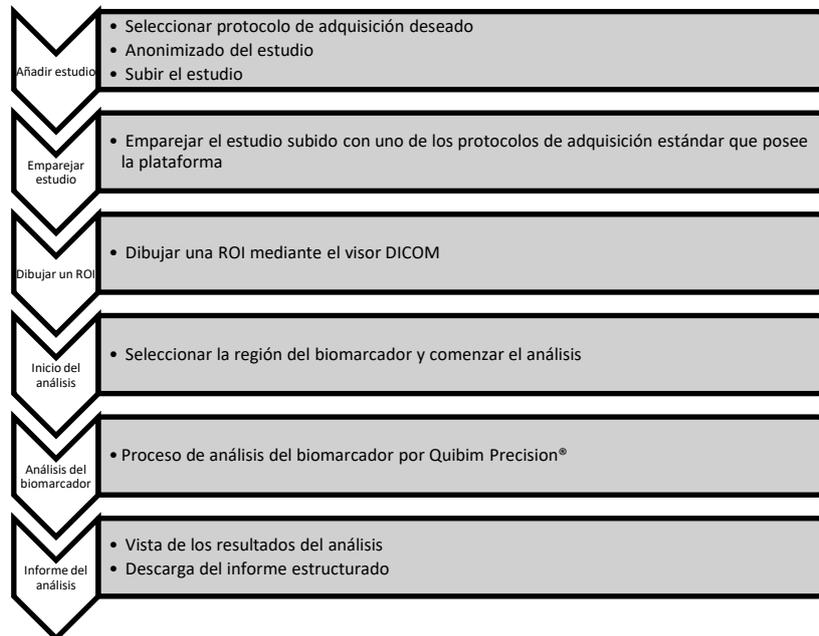
Figura 3. Análisis avanzado de cáncer de próstata.

En la Figura 3 se muestra el análisis cuantitativo de la difusión y la perfusión de la próstata en un paciente con cáncer de próstata. En el análisis de difusión se obtiene un coeficiente aparente de difusión (ADC) más bajo en aquellas zonas del tejido en el que el movimiento de las moléculas de agua se encuentra más restringido. Esta restricción de las moléculas de agua es debido, entre otras causas, a un aumento de la celularidad, característica típica en tejidos tumorales. Por otro lado, en el análisis de perfusión se obtiene un valor de  $K^{trans}$  elevado en aquellas zonas de tejido que presentan una mayor permeabilidad vascular. Este aumento de la permeabilidad es debido al aumento del número de vasos en una determinada región, proceso conocido como angiogénesis o neovascularización, típico de los tejidos tumorales.

Por lo tanto, sin alterar la exploración radiológica y extrayendo la información contenida en las imágenes mediante métodos de procesamiento de imagen, los biomarcadores de imagen ofrecen información adicional objetiva relacionada con el proceso patológico bajo estudio.

### 1.2.2 Análisis de imagen a través de la plataforma QUIBIM Precision®

La plataforma permite subir estudios radiológicos para su posterior análisis. El análisis de las imágenes se puede realizar o bien de manera completamente automática, sin necesidad de ningún tipo de interacción manual, o bien tras la delineación de una región de interés (ROI<sup>3</sup>).



En el diagrama de flujo se muestra la metodología que se ha de seguir para el análisis de las imágenes, siendo el tercer punto el que desaparece cuando se trata de un análisis automático.

### 1.3 QUIBIM Precision® Data Miner

En la Figura 4 podemos observar la interfaz de QUIBIM Precision®. En el lado izquierdo observamos la barra de navegación con las secciones de Estudios (donde se encuentra una lista con todos los estudios que un usuario ha subido a la plataforma), Administración (Solo administradores), Minería de datos (*Data Miner*), Biomarcadores (vista general de los métodos de análisis disponibles) y Cuenta (para inclusión de datos del usuario).

El desarrollo de la herramienta de visualización se ha hecho exclusivamente para el DataMiner, de manera que nos centraremos en esta sección de la plataforma.



Figura 4. Vista de QUIBIM Precision® Data Miner.

<sup>3</sup> Region Of Interest.

La vista del *Data Miner* es una herramienta diseñada para realizar análisis estadísticos sobre problemas clínicos y grupos de interés para la investigación y explotación científica. Esta herramienta trabaja con una base de datos de biomarcadores provenientes del análisis cuantitativo de imágenes médicas.

Explore los resultados de sus análisis

Descargar csv Descargar xls

Seleccione el biomarcador Lesiones de sustancia blanca (CNN) Búsqueda global  Mostrar borrados

Regist...	Nombre de paciente	Sexo del pacie...	Edad del paciente	Fecha del análisis	Modalidad	Etiquetas	ROI	Nombre del estudio	Fec...
			Min <input type="text"/> Max <input type="text"/>	Desde <input type="text"/> Hasta <input type="text"/>					Desde <input type="text"/> Hasta <input type="text"/>
1	LSTPrueba	Femenino	31	18/06/2018	MR			RM111C - RM CEREBRAL CON C...	12
2	TestWMLCarola01	Masculino	36	18/06/2018	MR			CEREBRO/C	25
3	TestWMLCarola01	Masculino	36	18/06/2018	MR			CEREBRO/C	25
4	TestWMLCarola02	Masculino	57	18/06/2018	MR			CEREBRO/C	05
5	TestWMLCarola05	Femenino	73	18/06/2018	MR			CEREBRO	04
6	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
7	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
8	TestWMLCarola05	Femenino	73	18/06/2018	MR			CEREBRO	04
9	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
10	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
11	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
12	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
13	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
14	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
15	TestWMLCarola03	Masculino	63	18/06/2018	MR			CEREBRO	05
16	TestWMLCarola01	Masculino	36	18/06/2018	MR			CEREBRO/C	25

quibim MANUAL TÉRMS PRIVACY

Figura 5. Vista de la tabla del *Data Miner*.

El usuario podrá ver los biomarcadores disponibles en la plataforma. A partir del biomarcador seleccionado, se mostrará una tabla con, aparte de los valores cuantitativos extraídos del análisis de las imágenes médicas, algunos datos de la cabecera DICOM, de los cuales los que a nosotros nos importan son: El identificador del **paciente**, el **sexo**, la **edad**, la **modalidad**, la etiqueta asociada a la región de interés (**ROI**), la **fecha del análisis**, el **nombre del estudio**, la **fecha del estudio** y el **identificador del estudio**.

Por lo tanto, en función del biomarcador que haya sido seleccionado, los datos que se mostrarán en la tabla, y con los que posteriormente trabajaremos en la herramienta de visualización, responderán a dos criterios principales: El protocolo de adquisición y los resultados de los biomarcadores.

### 1.3.1 Modalidades de imagen

Los módulos de análisis de la plataforma QUIBIM Precision®, trabajan con tres modalidades principales: Resonancia Magnética (RM), Tomografía Computarizada (TC) y Rayos X.

El estándar DICOM permite que la plataforma también sea interoperable con otras modalidades, como pueden ser:

- Ultrasonidos (US).
- Tomografía por emisión de positrones (PET<sup>4</sup>)
- PET/CT.
- PET/MR.

Nos vamos a centrar en los tres tipos que están presentes en nuestros biomarcadores.

#### 1.3.1.1 Resonancia Magnética (RM)

La imagen de resonancia magnética es una técnica no invasiva que permite obtener información sobre la estructura y composición del cuerpo mediante el fenómeno de resonancia magnética, que consiste en un campo magnético constante y frecuencias de radio que rebotan en las moléculas de grasa y agua en el cuerpo. Las ondas de radio se transmiten a un receptor en la máquina y se traducen en una imagen del cuerpo.

<sup>4</sup> Positron Emission Tomography.

Esta técnica permite obtener secciones secuenciales de los planos transverso, sagital y coronal de la parte del cuerpo que se haya analizado, así pudiendo crear imágenes tridimensionales a través de la que se puede navegar con el objetivo de encontrar posibles alteraciones en los tejidos, detectar cáncer y otras enfermedades. Generalmente se utiliza para diagnosticar patologías en el cerebro, los senos, el corazón, los vasos sanguíneos y las articulaciones.

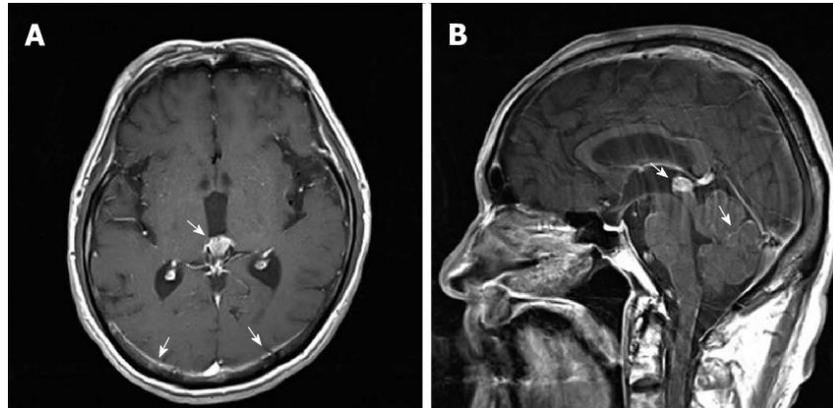


Figura 6. Sección transversal y sagital de una resonancia cerebral. [5]

### 1.3.1.2 Tomografía Computarizada (TC)

La tomografía computarizada tiene el mismo objetivo que la resonancia magnética, es decir, permite obtener secciones del cuerpo para su posterior análisis. La diferencia fundamental que posee respecto al RM es que utiliza radiación ionizante, es decir, realiza las secciones mediante rayos X.

El TC permite obtener una mayor nitidez de la imagen además de tratarse de un equipo más económico, el principal inconveniente es el uso de radiación ionizante. En la mayoría de los casos se recibirá una dosis despreciable, pero esto variará en función del número de radiografías (secciones) que se realicen del cuerpo. Su uso va enfocado generalmente en fracturas de hueso, detección de tumores en el cerebro y el abdomen, monitoreo del cáncer y búsqueda de sangrado interno.

### 1.3.1.3 Rayos X

Comúnmente conocida como radiografía convencional. Consiste en la exposición de un receptor de imagen radiográfica a una fuente de rayos X que, al interponer un objeto entre la fuente de radiación y el receptor, cada región aparecerá de diferentes tonos en una escala de grises en función de la radiación absorbida.



Figura 7. Radiografía de tórax. [6]

En la radiografía aparecerá en tonos más hiperintensos aquellas estructuras que absorben más radiación, como sería el hueso, y en tonos más hipointensos aquellas estructuras que ofrecen menos resistencia al paso de los rayos X y, por tanto, absorben menos radiación, como sería el caso del aire (los pulmones aparecen hipointensos en una radiografía de tórax).

Las radiografías se utilizan generalmente para el análisis de fracturas de hueso, pero también permite visualizar algunos tipos de cáncer. Al igual que la TC, los rayos X exponen el cuerpo a una dosis de radiación ionizante, cuyo impacto variará en función de la parte del cuerpo analizada.

### 1.3.2 Biomarcadores de imagen

La herramienta de visualización de datos o *Data Miner* permite a los usuarios de la plataforma la visualización, de una manera sencilla, de los resultados obtenidos a través de los distintos módulos de análisis. Así, se puede realizar comparaciones de poblaciones, buscar tendencia, etc. empleando los biomarcadores de imagen.

Vamos a clasificar los biomarcadores en función de cada módulo de análisis localizado en las distintas partes del cuerpo que permite analizar la plataforma.

#### 1.3.2.1 Cerebro

##### Volumetría cerebral con parcelación

Este módulo está diseñado para calcular el volumen de los tejidos cerebrales y sus regiones. Permite calcular el volumen de materia gris y blanca, líquido cefalorraquídeo y su distribución en diferentes regiones, además de la fracción del parénquima cerebral por imágenes de resonancia magnética (RM).

La determinación de los volúmenes cerebrales puede ser relevante para la detección temprana de enfermedades neurodegenerativas y desmielinizantes. Por tanto, su aplicación está determinada en la detección de patologías como el **deterioro cognitivo leve**, el **Alzheimer**, la **esclerosis múltiple**, la **demencia**, la **atrofia cerebral** o la **esquizofrenia**. Este módulo también se puede utilizar para el seguimiento de las enfermedades.

Resultado	Unidad de medida
Fracción del parénquima cerebral	Porcentaje [%]
Volumen absoluto y relativo de la materia gris	Valor [mL] y porcentaje de volumen intracraneal (ICV <sup>5</sup> ) [%]
Volumen absoluto y relativo de la materia blanca	Valor [mL] y porcentaje de ICV [%]
Volumen absoluto y relativo del líquido cefalorraquídeo	Valor [mL] y porcentaje de ICV [%]
Volumen absoluto y relativo del cerebro por regiones	Valor [mL] y porcentaje de ICV [%]

Tabla 1. Resultados de volumetría cerebral con parcelación. [7]

##### Lesiones de sustancia blanca

Este módulo permite la segmentación automática presentes en la sustancia blanca cerebral. Tras la segmentación automática, podemos obtener estadísticos relacionados con la carga lesional, como el volumen total de lesión en el cerebro.

<sup>5</sup> Intracranial Volume.

Esto se consigue empleando imágenes de resonancia magnética, en concreto empleando la secuencia FLAIR (del inglés, *Fluid Attenuated Inversion Recovery*) donde estas lesiones aparecen como hiperintensidades en la imagen. Estas lesiones aparecen en pacientes con esclerosis múltiple, infarto cerebral, tumores, lesiones vasculares y también son típicas de la edad.

Resultado	Unidad de medida
Área media de lesiones	Media [mm <sup>2</sup> ]
Máxima entropía	Máximo [Adimensional]
Número de lesiones	Valor [Número]
Promedio distancias	Media [mm]
Promedio entropía	Media [Adimensional]
Volumen absoluto lesional dominante	Porcentaje [%]
Volumen lesional dominante	Valor [mL]
Volumen lesional total	Valor [mL]
Volumen medio de lesiones	Media [mL]

Tabla 2. Resultados de las lesiones de sustancia blanca. [7]

### 1.3.2.2 Abdomen

#### Concentración de hierro y grasa hepático

Este módulo permite la cuantificación simultánea de hierro y grasa que se encuentra en el hígado mediante el modelado matemático de la señal de RM, teniendo en cuenta la distribución heterogénea de los depósitos de hierro y grasa para evitar factores de confusión. Estos depósitos se calculan vóxel a vóxel, es decir, por las unidades cúbicas que componen el objeto tridimensional, de todo el parénquima hepático. La determinación de la concentración de hierro y grasa puede ser útil en el diagnóstico y el seguimiento de las enfermedades difusas hepáticas.

Se conoce como una biopsia virtual en la que podemos conocer la concentración de ambas componentes en cada punto del parénquima hepático. Su aplicación va enfocada a **enfermedades difusas hepáticas**, como la **esteatohepatitis**, y casos de **cirrosis o fibrosis**.

Resultado	Unidad de medida
Densidad protónica de fracción grasa (PDFF <sup>6</sup> )	Porcentaje [%]
Densidad protónica de fracción de agua (PDWF <sup>7</sup> )	Porcentaje [%]
Relajometría R2* de hierro en el agua	Valor [s <sup>-1</sup> ]
Concentración de hierro en el agua	Valor [μmol/g, mg/g]
Relajometría R2* de hierro en la grasa	Valor [s <sup>-1</sup> ]
Concentración de hierro en la grasa	Valor [μmol/g, mg/g]

Tabla 3. Resultados de concentración de hierro y grasa. [7]

<sup>6</sup> Proton Density Fat Fraction.

<sup>7</sup> Proton Density Water Fraction.

Para conocer la distribución de cada biomarcador en el parénquima hepático, las principales estadísticas de cada biomarcador que se almacenan en la base de datos son: media, mediana, desviación típica, percentil 25 y percentil 75.

### 1.3.2.3 Hueso

#### Microarquitectura 3D del hueso

Este módulo permite calcular propiedades morfológicas, de irregularidad y de resistencia mecánica a partir de imágenes de RM o TC del hueso trabecular, mediante el análisis del porcentaje del volumen de hueso (BV/TV<sup>8</sup>), espesor (Tb.Th<sup>9</sup>) y separación (Tb.Sp<sup>10</sup>) trabecular, dimensión fractal (D<sup>2D</sup>/D<sup>3D</sup>) y módulo de elasticidad (E<sub>x,y,z</sub>). Este módulo también se puede aplicar a imágenes 2D de rayos X, mediante el análisis de características trabeculares en un único plano.

Estos parámetros pueden ser útiles para evaluar la calidad ósea y el riesgo de fractura y ayudar en el diagnóstico, el seguimiento y la evaluación de la respuesta al tratamiento en diferentes trastornos musculoesqueléticos. La principal aplicación del módulo es para casos de **osteoporosis**.

Resultado	Unidad de medida
Volumen de hueso a volumen total (BV/TV)	Porcentaje [%]: Mide el volumen total de hueso contenido en la región de interés delimitada (ROI).
Espesor trabecular (Tb.Th)	Valor [µm]: Mide el grosor promedio de las trabéculas contenidas en la ROI.
Separación trabecular (Tb.Sp)	Valor [µm]: Mide el tamaño promedio de los poros.
Número de trabéculas (Tb, N)	Valor [µm <sup>-1</sup> ]: Mide la relación entre (BV/TV) y el grosor trabecular (Tb.Th).
Dimensión fractal (D2D, D3D)	Valor: Mide la complejidad de la estructura trabecular. Cuanto más altos son estos índices, más compleja es la estructura y, a priori, mejor es la calidad y la resistencia del hueso.
Calidad de la estructura trabecular (QTS <sup>11</sup> )	Valor: Parámetro que tiene en cuenta todos los biomarcadores previos para dar un índice de la calidad del hueso

**Tabla 4. Resultados de microarquitectura 3D del hueso. [7]**

### 1.3.2.4 Oncología

Un análisis temprano del entorno microvascular del tumor conduce a una mejor evaluación de la agresividad de la lesión y la evaluación de la respuesta al tratamiento.

#### Perfusión – Modelado farmacocinético (DCE-MRI<sup>12</sup>)

Este módulo está diseñado para calcular los diferentes parámetros relacionados con el entorno vascular del tejido los cuales se basan en la evaluación farmacocinética cuantitativa de imágenes de resonancia magnética mejoradas con contraste dinámico (DCE-MRI).

<sup>8</sup> Bone Volume to Total Volume.

<sup>9</sup> Trabecular Thickness.

<sup>10</sup> Trabecular Separation.

<sup>11</sup> Quality of Trabecular Structure.

<sup>12</sup> Dynamic Contrast - Enhanced Magnetic Resonance Images.

La constante de transferencia ( $K^{\text{trans}}$ ), la tasa de extracción ( $k_{\text{ep}}$ ) y el volumen del espacio extravascular-extracelular ( $v_e$ ) se dan para evaluar los procesos de neovascularización en los tejidos y la angiogénesis tumoral.

Permite aplicar un modelo farmacocinético de 2-compartimentos y 2-entradas a las curvas de captación de contraste en las adquisiciones dinámicas de RM. Las aplicaciones de este módulo van enfocadas a tumores sólidos.

Resultado	Unidad de medida
Coeficiente de permeabilidad vascular ( $K^{\text{trans}}$ )	Valor [ $\text{min}^{-1}$ ]
Coeficiente de extracción - lavado ( $k_{\text{ep}}$ )	Valor [ $\text{min}^{-1}$ ]
Volumen extravascular-extracelular ( $v_e$ )	Porcentaje [%]

**Tabla 5. Resultados de perfusión – Modelado farmacocinético (DCE-MRI). [7]**

Estos biomarcadores se extraen para cada punto del tumor, siendo los estadísticos que se calculan y se almacenan en la base de datos: media, mediana, desviación típica, percentil 25 y percentil 75.

#### Perfusión – Análisis semicuantitativo (DCE-MRI)

Este módulo está diseñado para calcular las características de las curvas de concentración sin ningún modelo subyacente, son cuantitativas o semicuantitativas.

Estos descriptores de curva permiten evaluar las características morfológicas de la curva de contraste para la identificación de patrones relacionados con alguna malignidad. Igual que el anterior, su uso se enfoca en tumores sólidos.

Resultado	Unidad de medida
Área inicial bajo la curva de contraste a 60 segundos ( $\text{IAUC}^{1360}$ )	Valor [mM]
Pendiente inicial ( $\text{IS}^{14}$ )	Valor [mM/s]
Pico de contraste (Peak)	Valor [mM]
Tiempo hasta el pico ( $\text{TTP}^{15}$ )	Valor [s]

**Tabla 6. Resultados de perfusión – Análisis semicuantitativo (DCE-MRI). [7]**

Estos biomarcadores se extraen para cada punto del tumor, siendo los estadísticos que se calculan y se almacenan en la base de datos: media, mediana, desviación típica, percentil 25 y percentil 75.

#### Difusión – Análisis del coeficiente aparente de difusión ( $\text{ADC}^{16}$ )

Este módulo está diseñado para calcular la difusividad de las moléculas de agua en tejidos, lo que permite detectar difusividad rápida de moléculas de agua en moléculas a partir de moléculas de agua de difusión lenta. El objetivo de este método es proporcionar información relevante sobre la celularidad del tejido y la integridad de las membranas celulares.

<sup>13</sup> Initial Area Under Curve.

<sup>14</sup> Initial Slope.

<sup>15</sup> Time-To-Peak.

<sup>16</sup> Apparent Diffusion Coefficient.

El módulo ADC de difusión calcula el coeficiente de difusión aparente (ADC) a partir de imágenes de dos factores de difusión valor b (b-value) de RM aplicando el modelo mono-exponencial de la señal de difusión. El ADC representa el grado de difusión del agua en el tejido que está inversamente correlacionado con la celularidad del tejido y la integridad de las membranas celulares.

De la misma manera que los módulos previos, su uso se enfoca en tumores sólidos.

Resultado	Unidad de medida
Coefficiente de difusión aparente (ADC)	Valor [ $\text{mm}^2/\text{s}$ ]

Tabla 7. Resultados de imágenes ponderadas por difusión (DWI). [7]

Las principales estadísticas del biomarcador se almacenan como variables en la base de datos: media, mediana, desviación típica, percentil 25 y percentil 75.

#### Difusión – Modelo de movimiento incoherente en el interior de un vóxel (IVIM<sup>17</sup>)

En el modelo de movimiento incoherente en el interior de un vóxel se proporcionan los valores del coeficiente de difusión (D) y pseudodifusión (D\*) y la fracción vascular (f).

Al igual que el método de análisis anterior, mide el movimiento de las moléculas de agua en un tejido. Sin embargo, este modelo se ha demostrado ser más exacto en su aplicación en tejidos más vascularizados, como los riñones, el hígado, el páncreas o la próstata, ya que asume que en un vóxel de la imagen, no sólo tenemos células sino también otros componentes como los vasos.

Por lo tanto, se trata de un modelo bicompartimental en el que se tiene en cuenta la caída de la señal de RM no sólo por la difusión molecular pura, sino también por la microcirculación/perfusión.

Resultado	Unidad de medida
Coefficiente de difusión vóxel a vóxel (D)	Valor [ $\text{mm}^2/\text{s}$ ]
Coefficiente de pseudoperfusión (D*)	Valor [ $\text{mm}^2/\text{s}$ ]
Fracción vascular (f)	Porcentaje [%]

Tabla 8. Resultados del modelo de IVIM. [7]

#### Relajometría – mapeado por parámetros $T_1$ y $T_2$

Las mediciones de relajometría para el cálculo de los parámetros  $T_1$  y  $T_2$  y la generación de mapas paramétricos se integran en módulos para calcular los parámetros relacionados con el tiempo de relajación longitudinal y transversal de RM. Los tiempos de relajación  $T_1$  y  $T_2$  son parámetros físicos determinados por las propiedades biofísicas intrínsecas del tejido.

El **tiempo de relajación longitudinal  $T_1$** , es una constante de tiempo que describe la recuperación de la magnetización de un estado perturbado a su estado de equilibrio. El **tiempo de relajación transversal  $T_2$** , es una constante de tiempo que describe el decaimiento de la magnetización que ha sido excitada por un pulso de radiofrecuencia (RF).

<sup>17</sup> Intra - Voxel Incoherent Motions.

La relajometría cuantitativa aísla las contribuciones de los mecanismos de contraste de RM individuales ( $T_1$ ,  $T_2$ ,  $T_2^*$ ) y proporciona mapas, que son independientes del protocolo de RM y tienen una interpretación física, a menudo expresada en unidades absolutas. El mapeado de  $T_1$  y  $T_2$  proporciona información cuantitativa para la caracterización de macromoléculas y cambios fibróticos en los tejidos.

Las aplicaciones del módulo de relajometría se enfocan en tumores sólidos.

Resultado	Unidad de medida
Tiempo de relajación longitudinal ( $T_1$ )	Valor [ms]
Tiempo de relajación transversal ( $T_2$ )	Valor [ms]

Tabla 9. Resultados de relajometría – mapeado por parámetros  $T_1$  y  $T_2$ . [7]

Las principales estadísticas del biomarcador se almacenan como variables en la base de datos: media, mediana, desviación típica, percentil 25 y percentil 75.

### 1.3.2.5 Pulmón

#### Enfisema pulmonar

El enfisema pulmonar se asocia con la enfermedad pulmonar obstructiva crónica (EPOC), una patología subdiagnosticada que se manifiesta como un agrandamiento de los alvéolos pulmonares que causa una disminución de la función pulmonar. Estos espacios llenos de aire pueden detectarse mediante imágenes de TC y su volumen puede cuantificarse con precisión utilizando técnicas de segmentación de umbral.

Este módulo permite calcular las características morfológicas del pulmón, incluidos los volúmenes anatómicos y patológicos como son los volúmenes de parénquima pulmonar, de enfisema, de vasos sanguíneos a partir de imágenes de TC y baja densidad funcional pulmonar. Toda la información cuantitativa es proporcionada en valores absolutos y relativos con respecto a los volúmenes pulmonares. Se diferencia pulmón derecho e izquierdo y también se tiene en cuenta la tercera diferenciación de las mediciones en el tercio superior, medio e inferior en el pulmón.

Esta herramienta posee un algoritmo que analiza toda la imagen para proporcionar una cuantificación de diferentes biomarcadores de imagen: volumen pulmonar, volumen de vasos sanguíneos, volumen de enfisema y relación a través de su distribución regional. Su aplicación va destinada a casos de enfermedad pulmonar obstructiva crónica (EPOC) y fibrosis quística.

Resultado	Unidad de medida
Volumen pulmonar	Valor [mL]
Volumen de vasos sanguíneos	Valor [mL]
Volumen y porcentaje del enfisema	Valor [mL] y porcentaje [%]

Tabla 10. Resultados de enfisema pulmonar. [7]

Los resultados obtenidos se clasifican en dos grupos: volúmenes absolutos [mL] y volúmenes relativos [%] con respecto al volumen pulmonar total. En ambos casos, los datos se extraen en: totales, por tercio y por pulmón izquierdo y derecho.

### Baja densidad funcional pulmonar

El enfisema y la detección de baja densidad funcional se realizan mediante el uso de umbrales fijos o adaptativos.

Al igual que en el módulo anterior, su aplicación va destinada a casos de enfermedad pulmonar obstructiva crónica (EPOC) y fibrosis quística.

Resultado	Unidad de medida
Volumen pulmonar	Valor [mL]
Volumen de vasos sanguíneos	Valor [mL]
Volumen y porcentaje del enfisema	Valor [mL] y porcentaje [%]
Volumen y porcentaje de baja densidad funcional	Valor [mL] y porcentaje [%]

**Tabla 11. Resultados de baja densidad funcional pulmonar. [7]**

Los resultados obtenidos se clasifican en dos grupos: volúmenes absolutos [mL] y volúmenes relativos [%] con respecto al volumen pulmonar total. En ambos casos, los datos se extraen en: totales, por tercio y por pulmón izquierdo y derecho.

### 1.3.3 Herramienta de visualización

Ya hemos visto qué biomarcadores de imagen y de qué tipo extraen cada uno de los módulos de imagen. La explicación de estos datos servirá como contexto para entender en que se basarán los resultados de las gráficas que se mostrarán en la herramienta de visualización de QUIBIM Precision® *Data Miner*.

#### Visualización de los datos

Seleccione el tipo de gráfico



**Figura 8. Menú desplegable de la herramienta de visualización.**

En la Figura 8 se muestra el menú desplegable con algunos de los modos de visualización. A continuación, vamos a hacer una breve descripción de las gráficas que hemos desarrollado para la herramienta de visualización en este proyecto.

#### 1.3.3.1 Diagrama de dispersión

El diagrama de dispersión es un tipo de gráfica que utiliza las coordenadas cartesianas (Ejes X e Y) para mostrar el tipo de relación (Si existe) entre dos variables cuantitativas seleccionadas de un conjunto de datos.

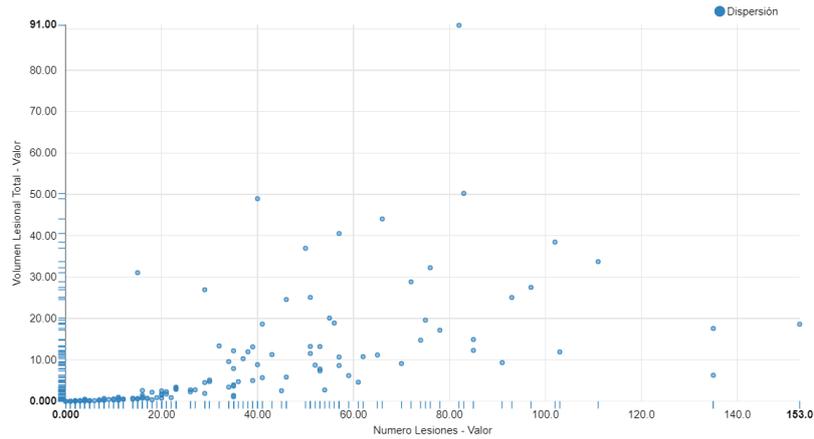


Figura 9. Ejemplo del diagrama de dispersión.

### 1.3.3.2 Diagrama de cajas

Un diagrama de cajas o diagrama de caja y bigote es un tipo de visualización para representar gráficamente los datos numéricos seleccionados a través de sus cuartiles. Este tipo de diagrama muestra a simple vista la mediana y los cuartiles de los datos.

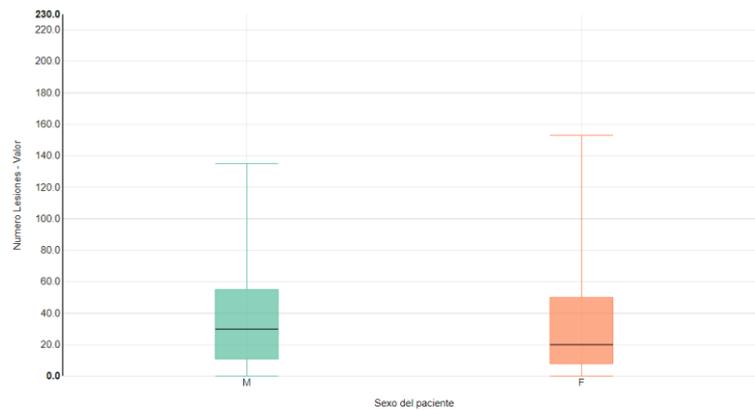


Figura 10. Ejemplo del diagrama de cajas.

### 1.3.3.3 Gráfico circular

Un gráfico circular es un recurso estadístico que generalmente se utiliza para representar porcentajes y proporciones de un bloque de datos.

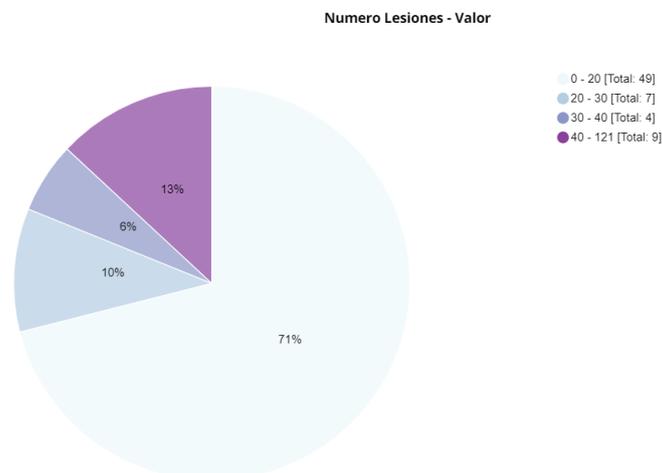


Figura 11. Ejemplo de gráfico circular.

### 1.3.3.4 Diagrama de barras

Un diagrama de barras es una representación gráfica de un conjunto de datos que se representarán como barras rectangulares de longitudes proporcionales al valor de los datos seleccionados. En nuestro caso, presentamos dos tipos distintos de diagrama de barras.

#### Selección múltiple

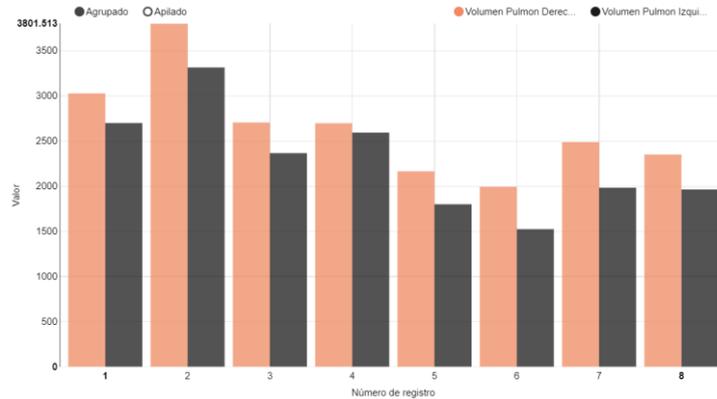


Figura 12. Ejemplo de diagrama de barras de selección múltiple.

#### Bidimensional

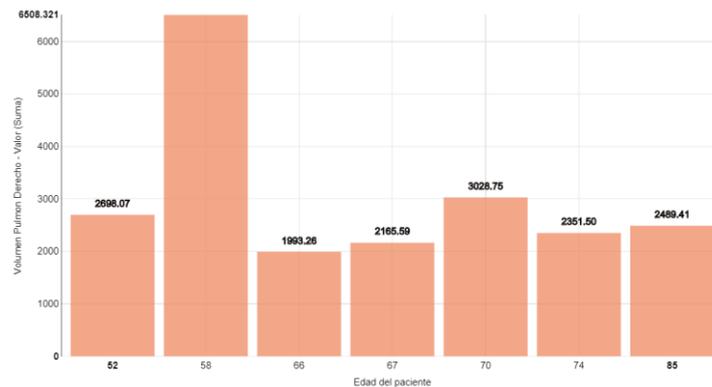


Figura 13. Ejemplo del diagrama de barras bidimensional.

### 1.3.3.5 Histograma

Un histograma se muestra como un diagrama de barras, donde la longitud de la barra representará la frecuencia de los datos seleccionados. Sirven para obtener una vista general de la distribución de la población respecto al dato seleccionado. La herramienta posee dos tipos de histogramas.

#### Datos múltiples

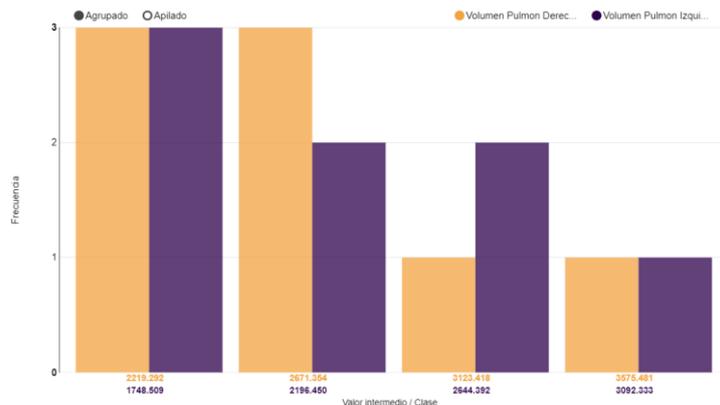


Figura 14. Ejemplo del histograma con selección de múltiples datos.

### Datos agrupados

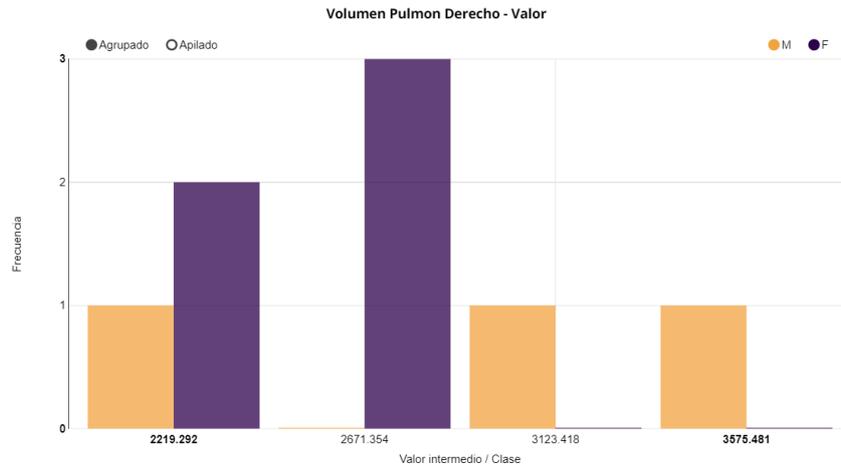


Figura 15. Ejemplo del histograma con selección por datos agrupados.

### 1.3.3.6 Gráfico Bland-Altman

Un gráfico Bland-Altman es un método gráfico que permite comparar dos técnicas de medición sobre variables cuantitativas. En el entorno sanitario es un tipo de visualización muy útil para casos en los que se quiere comparar la diferencia entre un nuevo método con respecto a uno ya establecido, y así poder validar o no el nuevo en función de si resulta fiable y reproducible para el uso previsto.

Para ello, el método de Bland-Altman cuantifica la diferencia media entre ambos métodos (bias o sesgo) y un rango de confianza, entre los cuales se espera que se incluyan el 95% ( $\pm 1,96 \cdot$  Desviación típica ( $s$ )) de las diferencias entre una técnica de medida y la otra.

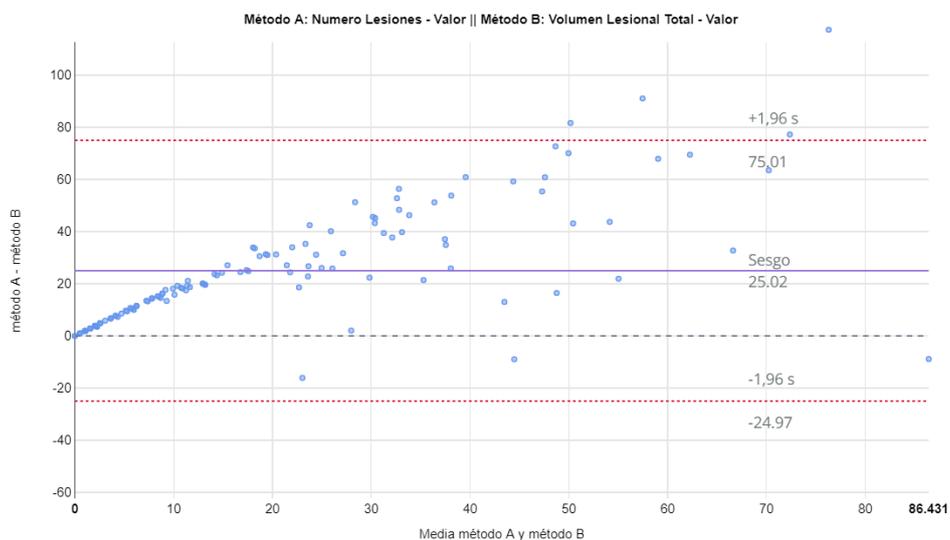


Figura 16. Ejemplo del método de Bland-Altman.

### 1.3.3.7 Gráfico Kaplan-Meier

Mayormente conocido como análisis de supervivencia, esta visualización es un estimador de la función de supervivencia mediante una representación gráfica de una función escalonada. Este tipo de gráfico muestra la probabilidad de supervivencia de los pacientes de un estudio en función del tiempo.

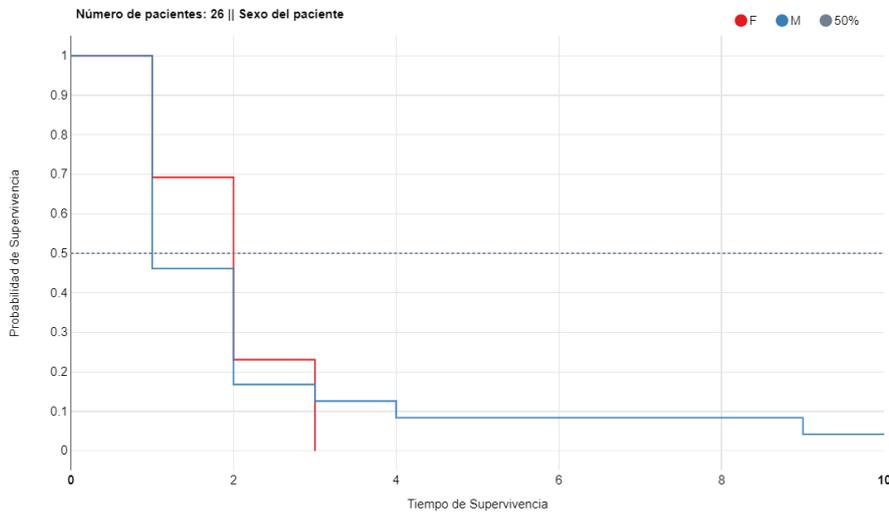


Figura 17. Ejemplo del gráfico Kaplan-Meier.

### 1.3.3.8 Diagrama de calor

El diagrama o mapa de calor es un tipo de representación gráfica donde, a partir la selección de dos datos categóricos distintos, mostrará los valores individuales contenidos de un dato cuantitativo seleccionado en forma de matriz.

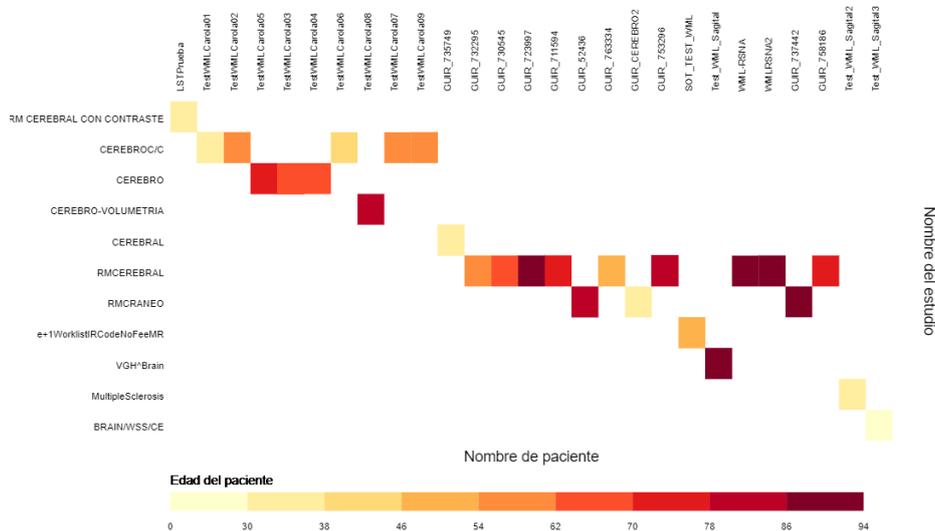


Figura 18. Ejemplo del diagrama de calor.

El dato cuantitativo seleccionado se dividirá en rangos equivalentes de distintos colores. El objetivo de este tipo de visualización es observar la distribución de los valores en el mapa de calor.

## Capítulo 2. Motivación

El desarrollo de este trabajo de fin de grado (TFG) se ha realizado en un entorno de prácticas en la empresa QUIBIM SL. La “*Herramienta de visualización de datos provenientes del análisis masivo de imágenes médicas en la plataforma de QUIBIM Precision®*” aborda la problemática del desarrollo de una herramienta necesaria para análisis estadísticos de datos mediante la visualización de gráficos que estén enfocados en mayor medida al entorno sanitario.

Lo que se plantea en este proyecto es el desarrollo de una aplicación que ofrezca un entorno intuitivo que permita al usuario realizar de manera sencilla una serie de análisis de índole estadístico que trabaje con una base de datos de biomarcadores provenientes de imágenes médicas. Los resultados de dichos análisis serán representados mediante gráficos estadísticos interactivos a elección del usuario, para permitir su interpretación y que ejerzan una influencia en la futura toma de decisiones, como por ejemplo podría ser la elección de un tratamiento frente a otro.

La plataforma QUIBIM Precision® es un software web, de manera que el desarrollo y la integración de la herramienta se ha implementado en CoffeeScript, siendo este un lenguaje de programación que se compila a JavaScript. Como directiva, se ha utilizado el entorno de trabajo de *AngularJS* para la librería *Angular-nvD3.js*, a su vez basada en la librería *D3.js*. Todos los modos de visualización se han desarrollado a través de *Angular-nvD3.js*, a excepción del mapa de calor, que se ha diseñado utilizando enteramente *D3.js*.

### 2.1 Objetivos

El objetivo general de este proyecto ha sido el desarrollo de una herramienta web como complemento de la plataforma QUIBIM Precision® que ofrezca a un especialista (el usuario) un aporte que le permita llegar a conclusiones a partir de los resultados de los análisis.

Se espera que la herramienta cumpla una serie de objetivos específicos:

- Ofrecer una interfaz accesible que resulte agradable y permita un uso intuitivo.
- Integrar distintos modos de visualización para su elección acorde al estudio que el usuario desee realizar.
- Modos de visualización interactivos, de manera que el gráfico resultante no se limite a una imagen que muestre solamente los datos seleccionados.
- Entorno a prueba fallos, con advertencias y casos de error que muestren al usuario si se ha hecho un uso inadecuado de la herramienta.
- Personalización de las gráficas mostradas, además de la opción de descargar las gráficas por el usuario en su equipo.



## 2.2 Estructura de la memoria

El documento de la memoria del trabajo de fin de grado se divide en seis capítulos que recogen toda la información relacionada con el proyecto. Haré un breve resumen de cada uno de los capítulos.

### **Capítulo 1. Introducción**

El primer capítulo sitúa en contexto al lector para que conozca el entorno en el que se ha desarrollado el proyecto y hace énfasis en la plataforma QUIBIM Precision® y las partes de esta que interesa conocer.

### **Capítulo 2. Motivación**

En este capítulo se describe la necesidad de la herramienta desarrollada en el proyecto y expone el porqué de los objetivos que se pretenden alcanzar.

### **Capítulo 3. Metodología**

En complemento al hecho de que dicho proyecto se ha realizado en un entorno de prácticas, en este capítulo se explica cómo se ha gestionado el proyecto y cuáles han sido las tareas establecidas durante el desarrollo de este.

### **Capítulo 4. Contexto del proyecto**

En este capítulo se describen los requisitos que han sido necesarios para el desarrollo de la herramienta y aquellos términos que sean fundamentales para entender algunos elementos de la plataforma.

### **Capítulo 5. Desarrollo y resultados**

En este capítulo se realizará una descripción más técnica de todos los elementos que se han desarrollado para la herramienta de visualización y si el resultado ha sido el deseado.

### **Capítulo 6. Conclusiones**

Finalmente, en el último capítulo se recogen las conclusiones a las que se ha llegado durante el proceso, basándose en los objetivos establecidos. También se planteará el futuro de la herramienta desarrollada.

## Capítulo 3. Metodología

### 3.1 Gestión del proyecto

Durante el desarrollo del trabajo de fin de grado se estableció un seguimiento semanal durante el cual el alumno tuvo reuniones semanales con el tutor de prácticas en QUIBIM, en este caso Ana Jiménez Pastor, para resolver todas las dudas que tenga y establecer los objetivos para la semana siguiente.

También se estableció una sesión mensual en la que el alumno tuvo que mostrar los avances realizados en el desarrollo de la herramienta. En este caso, se hizo una reunión con Ángel Alberich Bayarri, CEO (Chief Executive Officer) de QUIBIM, y una segunda reunión para el departamento de Desarrollo con Alejandro Mañas García, CTO (Chief Technology Officer) de QUIBIM.

En conjunto a las reuniones semanales, también se realizó un memorando semanal para mostrar los avances en el desarrollo del software de manera más visual y descriptiva.

### 3.2 Distribución en tareas

#### 3.2.1 Familiarización con el entorno

Esta primera tarea comenzó de manera previa al inicio de las prácticas, en la que el alumno, con la ayuda del libro *Interactive Data Visualization for the Web* de Scott Murray, desarrolló los conocimientos necesarios para el uso de la librería *D3.js* de JavaScript.

Iniciado el periodo de prácticas, se comenzó una fase de adaptación a la plataforma de Quibim Precision®. Se suministraron todos los ficheros necesarios de la plataforma al alumno, donde destaca el controlador del *Data Miner*. A partir de estos ficheros, el alumno tuvo que aprender las tecnologías utilizadas en la plataforma, siendo los principales:

- CoffeeScript: Lenguaje de programación que se compila a JavaScript.
- Jade: Transpilación de HTML.
- Less: Lenguaje de hojas de estilo que se compila en CSS.

Las tres tienen en común el hecho de ser una versión minificada de otro lenguaje, cuyas características se explicarán más adelante.

#### 3.2.2 Desarrollo de los modos de visualización

Durante la segunda tarea, una vez adquiridos unos conocimientos firmes de las tecnologías, se comenzó con el desarrollo de la herramienta de visualización.

La tutora de prácticas sugirió unas metas al alumno durante las reuniones, y se realizaron los memorandos semanales para mostrar los progresos. Los modos de visualización se desarrollaron mediante la librería *Angular-nvD3.js*, donde cada modo posee un modo peculiar de adquirir los datos de entrada. La mayor parte del trabajo durante este periodo consistió en adaptar los datos provenientes del biomarcador al tipo de visualización que se esperaba realizar.

Este periodo no excluyó la adquisición de nuevos conocimientos, es decir, requirió un aprendizaje constante para lograr llegar a la mejor solución posible de la manera más eficiente, llevando a veces a reformular muchas partes del código para llegar a un mismo objetivo.

#### 3.2.3 Testeo de la plataforma

Esta parte del desarrollo se combinó con la tarea anterior, ya que durante el testeo de la herramienta se encontraron problemas que requerían volver atrás.

Se puso a prueba la herramienta para poder solventar todos los problemas que pudieran ser ocasionados, tanto por los datos provenientes de los biomarcadores, como por el uso forzado del usuario.

## Capítulo 4. Contexto del proyecto

En los capítulos preliminar se han introducido algunos conceptos y se ha hecho mención de las tecnologías que se han utilizado para el desarrollo de este proyecto. En este capítulo se hace una descripción de estos conceptos y se explicarán las características de las tecnologías utilizadas.

### 4.1 Conceptos

#### 4.1.1 DICOM

El estándar DICOM (Digital Imaging and Communication On Medicine) de comunicación es reconocido el manejo de información de imágenes médicas y datos relacionados. El estándar se utiliza para el almacenamiento y el intercambio de imágenes médicas, lo que permite la integración de dispositivos de imágenes médicas como escáneres, servidores, impresoras, hardware de red y sistemas de archivo y comunicación de imágenes (PACS<sup>18</sup>).

Este estándar ha sido ampliamente adoptado por hospitales y está haciendo incursión en ámbitos más pequeños, como pueden ser consultas de dentistas.

Los ficheros DICOM pueden intercambiarse entre dos entidades que sean capaces de enviar y recibir imágenes y datos del paciente en este formato. Los dispositivos adaptados tienen una declaración de conformidad DICOM (Conformance Statements) para establecer que clases DICOM admiten, y el estándar incluye una definición de formato de archivo y un protocolo de comunicaciones de red que utiliza TCP/IP para comunicarse entre sistemas.

#### 4.1.2 DOM

Las siglas de Document Object Model (Modelo de Objetos del Documento). Es esencialmente una interfaz de plataforma que proporciona un conjunto estándar de objetos para representar documentos HTML, XHTML y XML. En definitiva, es la ventana o interfaz que se muestra en una aplicación web, donde se incluyen todos los objetos y elementos visibles y no visibles de una web.

#### 4.1.3 JSON

Acronímico de JavaScript Object Notation (Notación de Objeto de JavaScript). Es un formato de texto para el intercambio de datos. Se trata de un subconjunto de la declaración (Notación literal) de objetos de JavaScript donde se describen los datos con una sintaxis dedicada a gestionar e identificar los datos.

#### 4.1.4 Canvas

Se traduciría como lienzo. Un Canvas es un elemento HTML que permite la generación de gráficos dinámicamente por medio de archivos de procesamiento por lotes, que vendría a ser un archivo de texto plano que se interpreta como un guion.

El Canvas consiste en una región dibujable definida en HTML con atributos de altura y anchura. A través del código JavaScript se puede acceder a la zona mediante un conjunto completo de funciones similares a las de otras APIs<sup>19</sup> comunes de dibujo 2D, permitiendo así que los gráficos sean generados dinámicamente. Algunos de los usos previstos incluyen construcción de gráficos, animaciones, juegos, o la composición de imágenes.

---

<sup>18</sup> Picture Archiving and Communication System.

<sup>19</sup> Application Programming Interface.

#### 4.1.5 SVG

Scalable Vector Graphics (Gáficos Escalables en Formato Vectorial) es un formato que utiliza texto para definir gráficos vectoriales bidimensionales, estáticos o animados. Cada imagen SVG se define con código similar al HTML, por lo que se puede incluir dentro de cualquier documento HTML. Todos los navegadores actuales leen SVG, que al estar basado en XML, tendrá elementos sin etiquetas de cierre.

#### 4.1.6 GitHub

GitHub es una plataforma de desarrollo colaborativo para alojar proyectos utilizando el sistema de control de versiones Git. El propósito de este sistema es llevar un registro de los cambios en los archivos que componen el proyecto, y coordinar el trabajo que realicen varias personas sobre estos archivos compartidos.

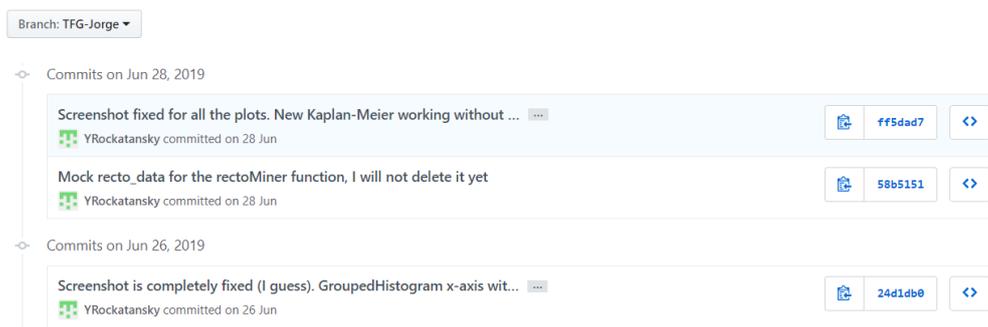


Figura 19. Ejemplo del registro de algunos cambios realizados en el proyecto.

GitHub está pensado para ser utilizado por desarrolladores, hospedando los repositorios de forma pública o privada. Los desarrolladores pueden observar cómo evoluciona el proyecto, los cambios en las versiones y volver a versiones anteriores si es necesario.

## 4.2 Lenguajes utilizados

### 4.2.1 CoffeeScript

CoffeeScript es un lenguaje de programación que se compila a JavaScript, por tanto, primero se hará una descripción de JavaScript.

JavaScript es un lenguaje de programación interpretado orientado a desarrollo web. Se define como orientado a objetos, basados en instancias (Clonación de otros objetos), imperativo, débilmente tipado y dinámico. Se utiliza principalmente en el lado del cliente, implementado como parte de un navegador web, permitiendo mejoras en la interfaz del usuario y en páginas web dinámicas. También posee un uso del lado del servidor, además de uso en aplicaciones externas a la web, como puede ser en documentos PDF y aplicaciones de escritorio.

JavaScript posee una sintaxis similar a C y también adopta nombres y convenciones del lenguaje de programación Java, pero eso no significa que Java y JavaScript estén relacionados, ambos poseen semánticas y propósitos diferentes.

Actualmente todos los navegadores modernos interpretan el lenguaje JavaScript integrado en las páginas web, donde se provee una implementación del DOM al lenguaje JavaScript para la interacción con la página web. Generalmente se venía utilizando en páginas web HTML para realizar operaciones y únicamente en su uso en el marco de la aplicación cliente, sin acceso a funciones de servidor. Este lenguaje se interpreta por el agente del usuario, al mismo tiempo que las sentencias van descargándose con el código HTML.

Definido JavaScript, pasamos a la descripción de CoffeeScript. Este lenguaje añade lo que en programación se llama **azúcar sintáctico**. Este término se refiere a los añadidos en la sintaxis del lenguaje de programación que lo vuelven más fácil de leer y expresar. Esto permite que el lenguaje tenga un uso más agradable para los programadores, ofreciendo un código que se pueda leer de manera más clara y concisa sin afectar a la funcionalidad del programa.

En nuestro caso, CoffeeScript añade estas características más sofisticadas para mejorar la brevedad y legibilidad de JavaScript. Además, la peculiaridad de este lenguaje permite escribir un código reducido en 1/3 menos de líneas de lo que sería en JavaScript, sin afectar al rendimiento ni al tiempo de ejecución.

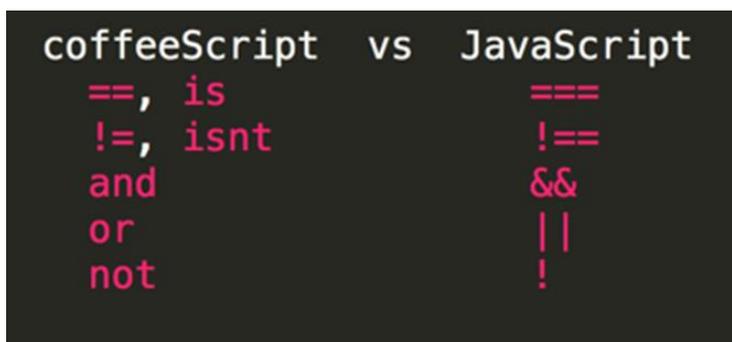


Figura 20. Diferencia entre los operadores en JavaScript y CoffeeScript.

En la Figura 20 se muestra un ejemplo de cómo la sintaxis en las declaraciones de control (`if`, `switch` y `for`) se sustituye por expresiones cuya función es deducible, además de que se puede omitir el uso de paréntesis ya que las llamadas a funciones son implícitas. En su lugar, se utiliza la sangría (Tabulaciones) para denotar bloques de código, esto evita el desarrollo de código desorganizado.

#### 4.2.2 Jade

Jade es un lenguaje de plantillas para HTML, por lo que se hará un inciso para explicar previamente que es HTML.

HTML viene de e HyperText Markup Language (Lenguaje de marcas de hipertexto) y consiste en un lenguaje de marcas para la elaboración de páginas web. No se considera un lenguaje de programación, ya que el lenguaje de marcas no posee funciones aritméticas ni declaración de variables, en lugar de eso, consiste en un documento que incorpora marcas que contienen la información acerca de la estructura de su texto o presentación.

El lenguaje HTML es un estándar para la elaboración de páginas web, que define una estructura básica y un código para la definición de los contenidos de la página, sea texto, imágenes, vídeos o juegos, entre otros. Se considera el lenguaje web más importante, puesto que es un estándar establecido por la World Wide Web (WWW) y todos los navegadores actuales lo han adoptado.

Este lenguaje se basa en las etiquetas. Para añadir un elemento a la página se hace una referencia mediante texto, de este modo, el código contiene solo texto mientras recae en el navegador web, que será el encargado de interpretar el código y unir todos los elementos para visualizar la página final.

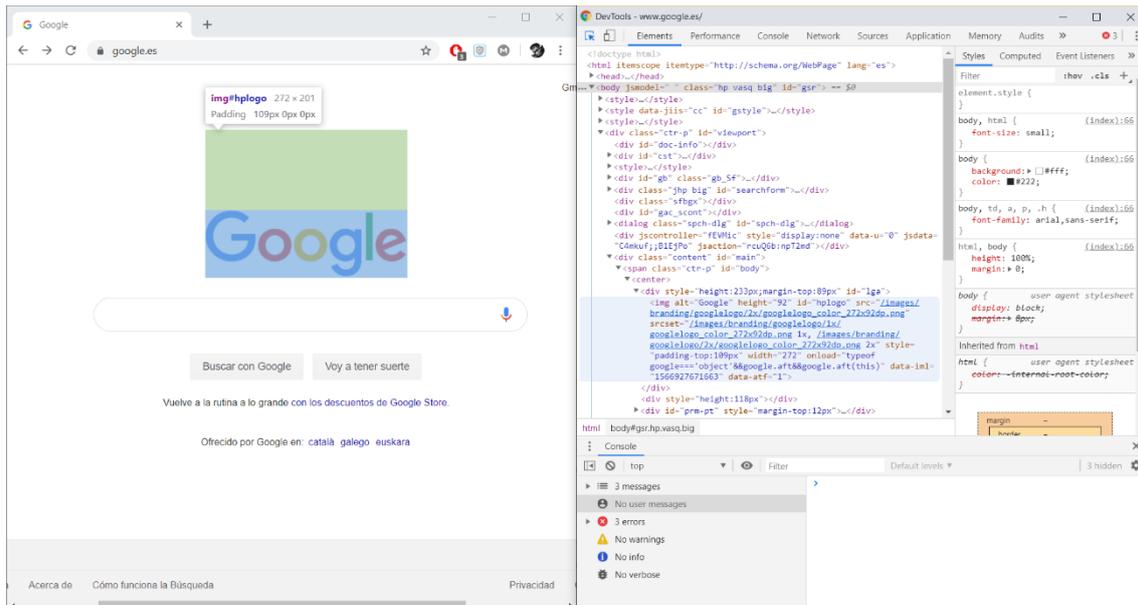


Figura 21. Inspector de página de Google Chrome.

En la Figura 21 se muestra el código fuente de la página web, donde podemos ver cómo están referenciados los elementos que la componen. Ahora que conocemos HTML, pasamos a la definición de Jade.

Jade, que pasó a llamarse Pug debido a un tema legal, pero sigue conociéndose como Jade por los usuarios, es un motor de plantillas de alto rendimiento implementado con JavaScript para Node.js y para navegadores. Jade es una sintaxis limpia y sensible al espacio en blanco para escribir HTML.

Este lenguaje consiste en una forma rápida de generar código, básicamente define una estructura semántica, ordenada y jerárquica utilizando solamente el nombre de las etiquetas HTML. Dicho de otra manera, es una abstracción de HTML plasmada en un formato más sencillo.

```
doctype html
html(lang="en")
  head
    title Jade
  body
    h1 Jade - node template engine
```

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title>Jade</title>
  </head>
  <body>
    <h1>Pug - node template engine</h1>
  </body>
</html>
```

En este ejemplo observamos cómo el código escrito en Jade (Superior) se compila en HTML (Inferior) respetando las etiquetas y las tabulaciones. Posee la misma utilidad que vimos con CoffeeScript, puesto que permite una reducción significativa del código.

### 4.2.3 Less

El último, pero no menos importante, se trata del lenguaje de hojas de estilo Less, que se compila en hojas de estilo en cascada (CSS).

CSS (Cascading Style Sheets) es un lenguaje de diseño gráfico para definir y crear la presentación de un documento estructurado escrito en HTML o XHTML. Junto con HTML y JavaScript, CSS forma el conjunto básico utilizado para el desarrollo de páginas web. La importancia de CSS está en definir estilos para crear páginas visualmente atractivas e interfaces de usuario (tanto para PC, como para móviles) que resulten agradables al usuario.

CSS está diseñado principalmente para separar el contenido del documento y la forma de presentación de este mediante capas, colores y fuentes. Su finalidad es mejorar la accesibilidad del documento, para darle una mayor flexibilidad y control en la especificación de características. De esta forma, varios documentos HTML pueden compartir un mismo estilo utilizando una sola hoja de estilos separada en un archivo `.css`, para así reducir la complejidad y evitar la repetición de código en la estructura del HTML.

Less es una adaptación de sintaxis de formato para CSS. De la misma forma que ocurre con CoffeeScript y con Jade, se caracteriza por poseer una sintaxis indexada que proporciona mecanismos que CSS no ofrece, como creación de variables, anidamiento, operadores y funciones.

## 4.3 Librerías de visualización de datos

En JavaScript, una librería es un archivo con instrucciones que permite agregar diversas funcionalidades. JavaScript interpreta funciones básicas, lo que hacen estas librerías es agregar otros recursos para poder manipular la estructura de las páginas (DOM) de forma dinámica y el estilo visual con CSS. Estos archivos se encuentran en diferentes servicios en la red, como puede ser en GitHub, donde los podemos descargar libremente.

En este apartado nos vamos a centrar en las dos librerías principales que hemos utilizado para el desarrollo de la herramienta de visualización.

### 4.3.1 D3.js

*D3.js* es una librería de JavaScript para visualizar, a partir de datos, diagramas dinámicos e interactivos en navegadores web. *D3.js* permite la representación de los datos mediante SVG, Canvas y HTML. Esta librería combina técnicas de visualización e interacción con un enfoque basado en datos para la manipulación del DOM, ofreciendo todas las capacidades que poseen los navegadores modernos, además de total libertad para diseñar la interfaz visual que sea más adecuada para los datos.

Por tanto, la peculiaridad que tiene *D3.js* frente a otras librerías, es que permite tener el control total sobre el resultado visual final. Por ejemplo, se puede utilizar *D3.js* para generar una tabla HTML a partir de una matriz de números, o utilizar los mismos datos para crear un diagrama de barras SVG interactivo con transiciones e interacción fluida.

El estilo funcional de *D3.js* permite la reutilización de código a través de una colección diversa de módulos oficiales y desarrollados por la comunidad. Entre estos, destacará *Angular-nvD3.js*, que explicaremos en el siguiente apartado.

### 4.3.2 Angular-nvD3.js

*Angular-nvD3.js* es una librería diseñada para facilitar el trabajo con la librería de gráficos reutilizables NVD3. Esta directiva permite al programador personalizar con facilidad los gráficos a través de una JSON API.

NVD3 es una librería de JavaScript basada en *D3.js* con la función de diseñar gráficos reutilizables y componentes de gráficos para *D3.js*, sin perder el potencial que posee *D3.js*. NVD3 es una librería gráfica muy flexible que permite añadir datos dinámicamente.



Figura 22. Gráficos reutilizables de *Angular-nvD3.js*. [8]

La característica clave que tiene *Angular-nvD3.js* frente a NVD3 es que la estructura jerárquica original de los modelos NVD3 se conserva por completo en la estructura directiva JSON. Esto significa que, mientras se crea un gráfico complejo que contenga múltiples modelos de gráficos elementales (Diagrama de barras, de línea...), se pueden personalizar las propiedades de cada modelo independiente contenido en el gráfico complejo, además de permitir modificar las propiedades globales del gráfico complejo. Esto se podría realizar en NVD3, pero se vuelve mucho más fácil con el enfoque JSON.

La utilidad que tiene implementarlo mediante una JSON API es que permite introducir las características de una manera eficiente.

```
"chart": {
  "type": "pieChart",
  "height": 500,
  "showLabels": true,
  "duration": 500,
  "legend": {
    "margin": {
      "top": 5,
      "right": 35,
      "bottom": 5,
      "left": 0
    }
  }
}
```

En este ejemplo de código se muestra cómo se introducen las propiedades de un gráfico circular como si se tratase de un JSON. Es fácil darse cuenta de que esta estructura resulta bastante ordenada, lo que facilita bastante el trabajo al desarrollador.

## Capítulo 5. Desarrollo y resultados

En este capítulo se realizará una explicación técnica de los elementos que se han desarrollado para la herramienta de visualización del *Data Miner*. Veremos los distintos tipos de diagramas que se han desarrollado y, para finalizar el capítulo, algunas funcionalidades que varios tipos de visualización poseen en común.

### 5.1 Estructura y sintaxis

El *Data Miner* se compone de múltiples ficheros que, en su conjunto, dan como resultado esta herramienta web. La mayor parte del desarrollo se ha realizado en el controlador, puesto que es donde está presente el código que realiza las funciones de la herramienta de visualización.

En el apartado de *Angular-nvD3.js* ya mostramos como se introducen las propiedades de los modos de visualización en forma de JSON.

```
▶ bidimensionalBarChart: {id: "bidimensionalBarChart", name: "Barras", img: "multiBarChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ blandAltmanChart: {id: "blandAltmanChart", name: "Bland-Altman", img: "blandAltmanChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ groupedHistogramChart: {id: "groupedHistogramChart", name: "Histograma", img: "histogramChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ heatmapChart: {id: "heatmapChart", name: "Mapa de calor", img: "heatmapChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ histogramChart: {id: "histogramChart", name: "Histograma", img: "histogramChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ kaplanMeierChart: {id: "kaplanMeierChart", name: "Kaplan-Meier", img: "kaplanMeierChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ multiBarChart: {id: "multiBarChart", name: "Barras", img: "multiBarChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ pieChart: {id: "pieChart", name: "Circular", img: "pieChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ plotChart: {id: "plotChart", name: "Cajas", img: "boxPlotChart.png", column_exclusion: {...}, prepareData: f, ...}
▶ scatterChart: {id: "scatterChart", name: "Dispersión", img: "scatterChart.png", column_exclusion: {...}, prepareData: f, ...}
```

Figura 23. Propiedades de los diagramas disponibles.

En la Figura 23 se muestra lo que contiene el JSON: El identificador, el nombre, la miniatura, las columnas excluidas, y lo más importante, las funciones. El controlador posee muchas otras funciones para trabajar con los datos de las gráficas, sin embargo, estas son las tres funciones fundamentales de la estructura de la herramienta.

#### 5.1.1 *prepareData*

El propio nombre indica que esta es la función en la que se preparan los datos. Obtiene un objeto proveniente de la base de datos en función del biomarcador que hayamos seleccionado, y le da legibilidad para que podamos trabajar con él. Según el tipo de gráfica seleccionada, tendrá algunas variaciones en la manera de la que representa los datos preparados.

```
▶ 0: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-000V", visible: true, $$height: 30, ...}
▶ 1: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-000X", visible: true, $$height: 30, ...}
▶ 2: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-000Z", visible: true, $$height: 30, ...}
▶ 3: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-0011", visible: true, $$height: 30, ...}
▶ 4: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-0013", visible: true, $$height: 30, ...}
▶ 5: GridRow {grid: Grid, entity: {...}, uid: "uiGrid-0015", visible: true, $$height: 30, ...}
```

Figura 24. Objeto proveniente de la base de datos.

```
▶ 0: {x: 0.026807, y: 27.60181, patient: "LSTPrueba"}
▶ 1: {x: 18.446, y: 46.799442, patient: "TestWMLCarola01"}
▶ 2: {x: 18.484, y: 51.520473, patient: "TestWMLCarola02"}
▶ 3: {x: 0.099, y: 49.009901, patient: "TestWMLCarola02"}
▶ 4: {x: 2.633, y: 44.015379, patient: "TestWMLCarola05"}
▶ 5: {x: 0.081, y: 52.258065, patient: "TestWMLCarola03"}
```

Figura 25. Muestra del bloque de datos preparado.

En la Figura 24 se observa el aspecto que tienen los datos provenientes del biomarcador que se haya seleccionado, mientras que en la Figura 25 vemos el aspecto que tienen después de haber pasado por la función de *prepareData*. En este caso se trata del diagrama de dispersión y lo que vemos son las coordenadas de los puntos.

### 5.1.2 *buildData*

En esta función trabajaremos el bloque de datos resultante de la función anterior. La complejidad con la que se presente variará según el tipo de visualización que haya sido seleccionada.

```
▼ 0:
  color: "#3182bd"
  ▶ info: {coeffA: "43.486", coeffB: "-0.031774", r2: "0.0074280"}
  key: "Dispersión"
  type: "scatter"
  values: (62)
  yAxis: 1
  ▶ __proto__: Object
▼ 1:
  color: "#35424A"
  key: "Línea"
  type: "line"
  values: (59)
  yAxis: 1
```

Figura 26. Datos listos del diagrama de dispersión.

En la Figura 26 se ha trabajado con los datos que preparamos en la función anterior, podemos ver que contiene los puntos del diagrama de dispersión y los puntos de la línea de regresión.

### 5.1.3 *buildOptions*

Finalmente, en esta función tomaremos los datos que ha trabajado la función anterior e indicaremos las opciones o propiedades de la gráfica que haya sido seleccionada.

```
chart:
  type: 'scatterChart'
  ...
  showDistX: true
  showDistY: true
  ...
  zoom:
    enabled: true
    scale: 1
```

En este pequeño extracto del código mostramos algunas de las opciones que, como ya vimos en el apartado de *Angular-nvd3.js*, indican las propiedades que tendrá la gráfica que se va a representar en un formato JSON.

## 5.2 Diagrama de dispersión

El diagrama de dispersión permite seleccionar dos datos para los ejes X e Y para así mostrar el tipo de relación de los datos seleccionados.

En la Figura 27 podemos observar que las opciones de la gráfica permiten representar una línea de tendencia. Según el tipo de línea seleccionada (Lineal, logarítmica, potencial, exponencial y polinomial) se calculará la ecuación de regresión basándose en los dos bloques de datos seleccionados para cada eje y a continuación dibujará la línea basándose en la ecuación calculada.

En este apartado hemos seleccionado el biomarcador de *Lesiones de sustancia blanca*, que posee 140 registros.

Seleccione los datos del eje X  
Numero Lesiones - Valor

Seleccione los datos del eje Y  
Volumen Lesional Dominante ...

Seleccione la línea de tendencia ✕  
Lineal

Seleccione grupo ⓘ ✕  
Sexo del paciente

Número de puntos de la línea  
Puntos

Regresión lineal  
 $y = 0.14109x - 0.39885$   
 $R^2 = 0.17231$

Ignorar valores menores que 0.

Paleta predeterminada

Restaurar zoom

**Figura 27. Opciones del diagrama de dispersión.**

Las opciones más importantes que posee este modo de visualización son las que se muestran en la Figura 27:

- Selección de datos: donde se selecciona, por un lado, la variable que queremos representar en el eje X y, por otro, la variable que queremos representar en el eje Y.
- Selección de línea de tendencia: donde seleccionaremos el tipo de función al que queremos ajustar los datos.
- Selección de grupos: permite agrupar los datos mostrados en función de una tercera variable.

También existen otras opciones de personalización, como indicar el número de puntos que se desea para la línea de tendencia.

Regresión potencial

Los valores de X e Y han de ser mayores de 0

Ignorar valores menores que 0.

**Figura 28. Advertencia de la regresión potencial.**

La opción de ignorar los valores inferiores a cero puede ser útil para representar las líneas de tendencia que no se pueden representar cuando existen valores negativos o nulos en los bloques de datos seleccionados.

Regresión polinomial

Orden Decimales ⓘ

**Figura 29. Opciones de la regresión polinomial.**

En caso de seleccionar la línea polinomial, también ofrece la opción de indicar el orden y el número de decimales de la ecuación. Ambas opciones son bastante útiles, ya que habrá ocasiones en que la forma predeterminada no mostrará una línea adecuada.

Por último, está la opción de restaurar zoom. El diagrama permite hacer zoom y desplazarse por él, esta opción permite volver a la vista predeterminada.

## 5.2.1 Desarrollo

### 5.2.1.1 Selección de los datos

Recordemos de la Figura 24 la forma que tiene el objeto proveniente de la base de datos. En ese apartado, hemos seleccionado un biomarcador con 140 registros. Hemos seleccionado *Numero Lesiones – Valor* para el eje X y *Volumen Lesional Dominante – Valor* para el eje Y.

```
▶ 82: {x: 40, y: 47.059645, patient: "/w1bb1DWILGuB92V1hwY"}
```

Figura 30. Registro de los datos preparados del diagrama de dispersión.

En la Figura 30 apuntamos al registro 82, y vemos una muestra de los datos preparados. Para llegar a esto, lo que hemos hecho en la selección de datos ha sido inspeccionar los 140 registros del biomarcador, y asignar a cada punto los dos datos correspondientes a X e Y. Además, vemos que también muestra el nombre del paciente anonimizado al que corresponde ese registro.

```
color: "#3182bd"  
key: "Dispersión"  
type: "scatter"  
▶ values: (140) [{...}, {...}, {...}, {...}, {...}, ...  
yAxis: 1
```

Figura 31. Datos listos para ser representados del diagrama de dispersión.

Finalmente, en la construcción de los datos asignamos los 140 registros de las coordenadas de los puntos al diagrama de dispersión.

```
▼ 82:  
patient: "/w1bb1DWILGuB92V1hwY"  
series: 0  
shape: "circle"  
size: 1000  
x: 40  
y: 47.059645
```

Figura 32. Plantilla del registro para el diagrama de dispersión.

Si esta vez inspeccionamos un punto, vemos que contiene los datos fundamentales que nos pide la plantilla del diagrama de dispersión en *Angular-nvd3.js*. En la Figura 32 vemos que contiene la forma, el tamaño y las coordenadas del punto.

### 5.2.1.2 Selección de línea de tendencia

En cuanto a complejidad, el cálculo de la línea de tendencia es la parte del diagrama de dispersión que posee el mayor coste. Afortunadamente en GitHub encontramos la librería *regression-js*, que hemos adaptado a nuestras necesidades para poder combinarlo con la estructura que presenta la herramienta de visualización.

```
▼ 0:  
color: "#3182bd"  
▶ info: {slope: "0.14109", intercept: "-0.39885", r2: "0.17231"}  
key: "Dispersión"  
type: "scatter"  
▶ values: (140) [{...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, ...  
yAxis: 1  
▶ __proto__: Object  
▼ 1:  
color: "#35424A"  
key: "Línea"  
type: "line"  
▶ values: (58) [{...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}, ...  
yAxis: 1  
▶ __proto__: Object
```

Figura 33. Datos del diagrama de dispersión y la línea de tendencia (Regresión lineal).

En la Figura 33 observamos que al seleccionar una línea de tendencia (En este caso la regresión lineal) los datos poseen una mayor complejidad. Ahora se ha añadido una nueva propiedad con los datos de la ecuación para la línea de tendencia, y se añadirá un nuevo bloque que incluye los puntos para la representación de la línea.

Para calcular los puntos de la línea hemos utilizado las funciones adaptadas de la librería *regression-js* [9]. Estas funciones nos proporcionan los coeficientes de la ecuación de la línea de tendencia seleccionada, y nuestro software calculará, a partir de los coeficientes, los puntos de la línea en el área delimitada por el diagrama de dispersión. También podemos ver que para todas las líneas de tendencia se calculará el coeficiente de determinación ( $R^2$ ) correspondiente.

A continuación, mostramos las ecuaciones características de las líneas de tendencia.

#### Regresión lineal

$$y = mx + c \quad (1)$$

Donde el coeficiente  $m$  es la pendiente (*slope*) y el coeficiente  $c$  la ordenada en el origen (*intercept*).

#### Regresión logarítmica

```
info: {coeffA: "3.1684", coeffB: "-5.0008", r2: "0.13425"}
```

Figura 34. Propiedad con los datos de la regresión logarítmica.

$$y = a + b \cdot \ln(x) \quad (2)$$

Donde  $a$  (*coeffA*) y  $b$  (*coeffB*) son los coeficientes característicos de la ecuación logarítmica.

#### Regresión potencial

```
info: {coeffA: "0.0095056", coeffB: "1.4225", r2: "0.056921"}
```

Figura 35. Propiedad con los datos de la regresión potencial.

$$y = ax^b \quad (3)$$

Donde  $a$  (*coeffA*) y  $b$  (*coeffB*) son los coeficientes característicos de la ecuación potencial.

#### Regresión exponencial

```
info: {coeffA: "7.0825", coeffB: "0.013940", r2: "-0.66959"}
```

Figura 36. Propiedad con los datos de la regresión exponencial.

$$y = ae^{bx} \quad (4)$$

Donde  $a$  (*coeffA*) y  $b$  (*coeffB*) son los coeficientes característicos de la ecuación exponencial.

#### Regresión polinomial

```
▼ info:  
  ► equation: (3) [-0.001, 0.271, -2.265]  
  r2: "0.188"  
  string: "y = -0.001x<sup>2</sup> + 0.271x + -2.265"
```

Figura 37. Propiedad con los datos de la regresión polinomial.

$$y = a_n x^n + \dots + a_1 x + a_0 \quad (5)$$

Donde los coeficientes  $[a_n, \dots, a_1, a_0]$  variarán en función del orden de la ecuación que se haya configurado. La ecuación polinomial predeterminada es de orden 2.

### 5.2.1.3 Selección de grupos

La selección de grupos permite diferenciar los puntos del diagrama en la opción seleccionada. Esto resulta bastante útil para observar la dispersión de varias poblaciones distintas que se encuentren en un mismo bloque de datos.

```
82: {x: 40, y: 47.059645, patient: "/w1bb1DWILGuB92V1hwY", group: "M"}
```

Figura 38. Registro con la propiedad de grupo del diagrama de dispersión.

Al seleccionar un grupo la función de `prepareData` añadirá la nueva propiedad, en este caso hemos seleccionado el sexo del paciente. De esta manera, cuando trabaje con los datos para su posterior representación, la herramienta separará los bloques en los grupos.

```
▶ 0: {key: "M", type: "scatter", values: Array(45), yAxis: 1, info: {...}, ...}
▶ 1: {key: "F", type: "scatter", values: Array(93), yAxis: 1, color: "#6baed6"}
▶ 2: {key: "Línea", type: "line", color: "#35424A", values: Array(58), yAxis: 1}
```

Figura 39. Bloques diferenciados por grupos del diagrama de dispersión.

En el ejemplo de la Figura 39 vemos que los 140 registros se han separado en los dos grupos de sexo del paciente.

## 5.2.2 Resultado

### Visualización de los datos

Seleccione el tipo de gráfico

Dispersión

Seleccione los datos del eje X

Numero Lesiones - Valor

Seleccione los datos del eje Y

Volumen Lesional Dominante ...

Seleccione la línea de tendencia

Lineal

Seleccione grupo

Sexo del paciente

Número de puntos de la línea

Puntos

Regresión lineal

$y = 0.14109x - 0.39885$   
 $R^2 = 0.17231$

Ignorar valores menores que 0.

Paleta predeterminada

Restaurar zoom

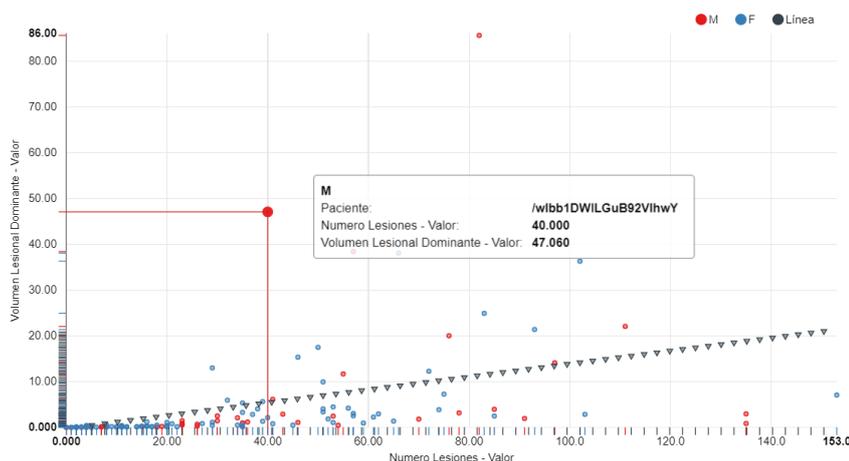


Figura 40. Resultado del diagrama de dispersión.

En la Figura 40 podemos observar el resultado del diagrama de dispersión con el que hemos trabajado en este apartado.

Hemos seleccionado *Numero Lesiones – Valor* para el eje X y *Volumen Lesional Dominante – Valor* para el eje Y. La línea de tendencia seleccionada es la regresión lineal, cuya ecuación podemos observar al lado izquierdo. Se han separado los puntos en función del sexo del paciente, que se puede diferenciar claramente por colores.

Finalmente, en el *tooltip*<sup>20</sup> se muestran los datos del registro que hemos estado utilizando como ejemplo en este apartado.

<sup>20</sup> *Tooltip*: Es una herramienta de ayuda visual, que funciona al situar el cursor sobre algún elemento gráfico.

### 5.3 Diagrama de cajas

El diagrama de cajas permite seleccionar dos datos para los ejes X e Y. Es recomendable seleccionar un dato categórico para el eje X (e.g. sexo del paciente, modalidad de imagen...), y para el eje Y siempre ha de seleccionar un dato cuantitativo, de esta manera podrá hacer el cálculo de las agrupaciones.

Seleccione los datos del eje X  
Sexo del paciente

Seleccione los datos del eje Y  
Numero Lesiones - Valor

Guardar gráfico

Seleccione paleta

Figura 41. Opciones del diagrama de cajas.

En la Figura 41 observamos que esta visualización es bastante más limitada en opciones en comparación a la anterior, pero esto es debido a que este diagrama no permite excesiva personalización.

#### 5.3.1 Desarrollo

##### 5.3.1.1 Selección de datos

El biomarcador seleccionado es el de *Lesiones de sustancia blanca*, que posee 140 registros. Para los ejes X e Y hemos seleccionado los datos del *Sexo del paciente* y *Número Lesiones – Valor* respectivamente.

```
▶ 0: {x: "M", y: 0}  
▶ 1: {x: "M", y: 7}  
▶ 2: {x: "F", y: 2}
```

Figura 42. Muestra de los datos del diagrama de cajas.

En este diagrama, la función de `prepareData` inspeccionará los 140 registros, y los separará en los grupos que hayamos seleccionado en el eje X. Esta es la razón por la que es recomendable utilizar un dato categórico, de lo contrario, obtendríamos una visualización errática. En la Figura 42 podemos ver una muestra de cómo se asigna el grupo al eje X y el dato correspondiente al eje Y.

```
▼ 0:  
  label: "M"  
  ▶ values: {outliers: Array(45), Q1: 11, Q2: 30, Q3: 55, whisker_low: 0, ...}  
  ▶ __proto__: Object  
▼ 1:  
  label: "F"  
  ▶ values: {outliers: Array(93), Q1: 8, Q2: 20, Q3: 50, whisker_low: 0, ...}  
  ▶ __proto__: Object
```

Figura 43. Plantilla de los datos listos para representar el diagrama de cajas.

A continuación, trabajará con los datos seleccionados. Como se puede observar en la Figura 43, el programa tomará los datos correspondientes a cada grupo y los transformará en una matriz numérica (*outliers*) que vamos a utilizar para calcular los percentiles. De esta matriz también tomará el valor mínimo y máximo, y los almacenará en las propiedades correspondientes (*whisker\_low* y *whisker\_high*).

Para el cálculo de los percentiles vamos a utilizar la función `get_percentile`, propia de la herramienta. A esta función se le envía una matriz y se le indica el percentil (25, 50 y 75) que aplica el método de interpolación lineal entre los rangos más cercanos, y nos devuelve el valor del coeficiente.

### 5.3.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Cajas

Seleccione los datos del eje X

Sexo del paciente

Seleccione los datos del eje Y

Numero Lesiones - Valor

Guardar gráfico

Seleccione paleta

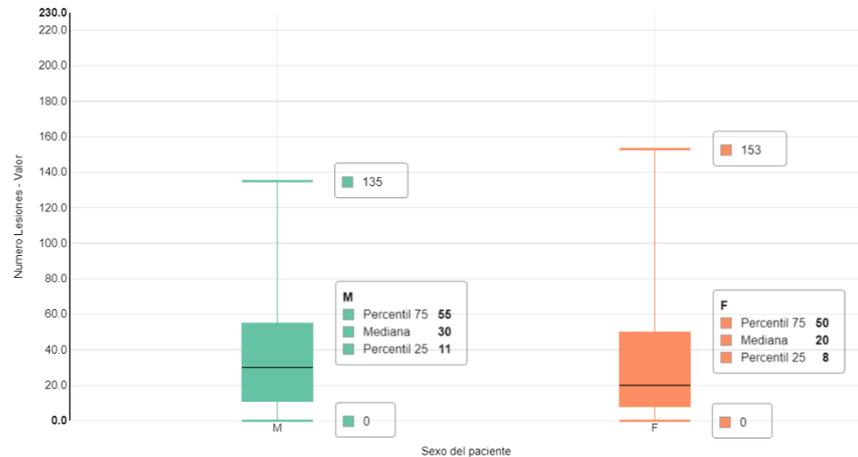


Figura 44. Resultado del diagrama de cajas.

En la Figura 44 hemos combinado el *tooltip* de la caja y de los bigotes para poder visualizarlo en una única captura. Como podemos ver, la caja muestra los percentiles 75, 50 (Mediana) y 25 de la matriz de ese grupo, y en los bigotes se muestran los valores máximo y mínimo que se encuentran dentro de cada matriz.

### 5.4 Gráfico circular

En la plataforma, el gráfico circular está adaptado de tal manera que permite seleccionar un único dato, donde el usuario establecerá unos rangos personalizados para poder visualizar las proporciones de los datos que se encuentran entre los rangos establecidos.

Seleccione los datos

Numero Lesiones - Valor

Define el rango

0	
20	x
40	x
60	x
121	

+    Reiniciar gráfica

Guardar gráfico

Seleccione paleta

Predeterminado

Figura 45. Opciones del gráfico circular.

En la Figura 45 podemos observar las principales opciones de personalización que posee el gráfico: Selección de datos y selección de rangos.

También está la opción de reiniciar gráfica. Es una opción sencilla pero útil, que permite resetear los rangos para no tener que hacerlo uno a uno.

#### 5.4.1 Desarrollo

##### 5.4.1.1 Selección de datos

En este apartado seleccionamos el biomarcador *Lesiones de sustancia blanca (CNN)*, que posee 65 registros. El dato que vamos a utilizar de ejemplo es el de *Numero Lesiones – Valor*.

En el gráfico circular la función de selección de datos se limitará a tomar los valores del dato seleccionado, ya que todo lo demás lo hará en la siguiente función.

##### 5.4.1.2 Selección de rangos

De manera predeterminada, la gráfica que se va a representar tendrá dos rangos, que van del mínimo al valor intermedio y de este al valor máximo. Tanto el valor mínimo como el máximo serán permanentes, por lo que no se podrán modificar por el usuario.

```
▼ 0:
  key: "0 - 60.5"
  ▶ values: {value: Array(62), Q1: 4, Q2: 5, Q3: 18, Q4: 13.370967741935484, ...}
  ▶ __proto__: Object
▼ 1:
  key: "60.5 - 121"
  ▶ values: {value: Array(3), Q1: 83.5, Q2: 96, Q3: 108.5, Q4: 96, ...}
  ▶ __proto__: Object
length: 2
```

Figura 46. Rangos predeterminados del gráfico circular.

De forma similar a la que lo hacía con los grupos, esta visualización inspeccionará el bloque de datos de 65 registros y agrupará los números de manera ordenada en los rangos correspondientes. Además, está pensado para actualizarse de forma dinámica tal que, cada vez que el usuario realice un cambio en los rangos, la gráfica se actualizará automáticamente. Para esto hace uso de la función `update_pie`, propia de este modo de visualización.

La función `update_pie` va ligada a los cuadros de entrada de texto para introducir los rangos, de manera que está siempre a la escucha. El trabajo de dicha función es recalcular los datos para ajustarlos a los rangos que se hayan establecido y adaptarlo a la plantilla de datos de entrada del gráfico circular de *Angular-nvd3.js*. A partir del tamaño que tenga la matriz numérica con los valores, indicaremos el total de registros que se encuentran en este rango y calculamos el porcentaje respecto al número total de registros del biomarcador.

```
key: "0 - 20"
▼ values:
  Q1: 3
  Q2: 5
  Q3: 7
  Q4: 6.653061224489796
  Q5: 5.6696507483703495
  ▶ value: (49)
```

Figura 47. Propiedad con los valores de los cálculos realizados en el rango.

Junto a lo mencionado, como se puede observar en la Figura 47, también hace un cálculo a partir de la matriz de datos que se encuentra en cada rango para mostrar posteriormente en el *tooltip* los percentiles 25 ( $Q1$ ), 50 ( $Q2$ ) y 75 ( $Q3$ ) junto a la media ( $Q4$ ) y a la desviación típica ( $Q5$ ).

Para el cálculo de los percentiles utiliza la función `get_percentile`, que ya comentamos en la visualización anterior. Para calcular la media y la desviación típica introducimos dos nuevas funciones a la herramienta, `get_mean` y `get_standardDeviation`. Ambas reciben la matriz numérica del rango actual y hacen el cálculo correspondiente.

### 5.4.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Circular

Seleccione los datos

Numero Lesiones - Valor

Define el rango

0  
20 ✕  
40 ✕  
60 ✕  
121

+ Reiniciar gráfica

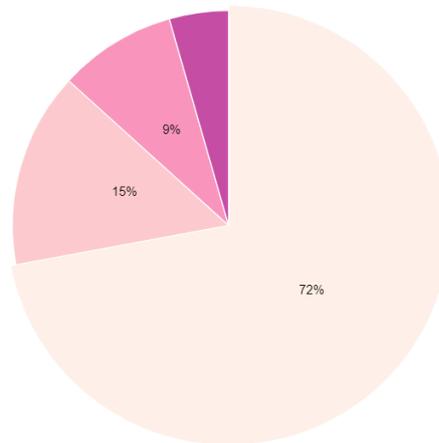
Guardar gráfico

Seleccione paleta

Paleta de colores

Numero Lesiones - Valor

0 - 20 [Total: 49]  
20 - 40 [Total: 10]  
40 - 60 [Total: 6]  
60 - 121 [Total: 3]



Numero Lesiones - Valor:	0 - 20
Total:	49
Porcentaje:	72%
Percentil 75:	3
Mediana:	5
Percentil 25:	7
Media:	6.653
Desviación típica:	5.67

Figura 48. Resultado del gráfico circular.

En la Figura 48 observamos el gráfico que se ha representado a partir de los rangos indicados en las opciones a la izquierda. En la parte superior de la derecha vemos una leyenda que indica los rangos con el total de datos contenidos en cada uno, y en el *tooltip* podemos observar los resultados de los cálculos que se han realizado para el rango que hemos inspeccionado.

## 5.5 Diagrama de barras

Para este diagrama vamos a utilizar el biomarcador de *Air trapping pulmonar*, que al poseer tan solo 8 registros nos facilitará su visualización en las figuras. El diagrama de barras dispone de dos tipos distintos de visualización: selección múltiple y bidimensional.

### 5.5.1 Selección múltiple

Permite la selección de múltiples datos que se ordenarán según el número de registros que existan en el biomarcador seleccionado. El objetivo es seleccionar múltiples datos del mismo registro que estén relacionados para que el usuario puede hacer una comparación.



Figura 49. Opciones del diagrama de barras de selección múltiple.

### 5.5.1.1 Desarrollo

El diagrama de barras de selección múltiple posee algunas diferencias respecto a la manera de la que hemos trabajado con los datos hasta ahora. Para este ejemplo vamos a seleccionar los datos *Volumen Pulmón Derecho – Valor* y *Volumen Pulmón Izquierdo – Valor*.

- ▶ 0: (8) [GridRow, GridRow, GridRow, GridRow, GridRow, GridRow, GridRow, GridRow]
- ▶ 1: (8) [GridRow, GridRow, GridRow, GridRow, GridRow, GridRow, GridRow, GridRow]

Figura 50. Objetos de los datos seleccionados en el diagrama de barras de selección múltiple.

En la Figura 50 apreciamos que esta vez en la función `prepareData` se replicará el objeto proveniente de la base de datos para cada dato que hayamos seleccionado. En este caso hemos seleccionado dos datos, por tanto, tendremos dos objetos con los 8 registros del biomarcador.

```

▼ 0:
  ▼ values: Array(8)
    ▶ 0: {x: 1, y: 3028.747424, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 1: {x: 2, y: 3801.512578, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 2: {x: 3, y: 2706.808169, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 3: {x: 4, y: 2698.065806, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 4: {x: 5, y: 2165.592625, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 5: {x: 6, y: 1993.260219, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 6: {x: 7, y: 2489.414565, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    ▶ 7: {x: 8, y: 2351.501041, series: 0, key: "Volumen Pulmon Derecho - Valor"}
    length: 8
    ▶ __proto__: Array(0)
  ▶ __proto__: Object
▼ 1:
  ▼ values: Array(8)
    ▶ 0: {x: 1, y: 2700.540344, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 1: {x: 2, y: 3316.303932, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 2: {x: 3, y: 2366.797835, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 3: {x: 4, y: 2594.113656, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 4: {x: 5, y: 1800.193382, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 5: {x: 6, y: 1524.538806, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 6: {x: 7, y: 1983.703829, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    ▶ 7: {x: 8, y: 1963.633569, series: 1, key: "Volumen Pulmon Izquierdo - Valor"}
    length: 8
    ▶ __proto__: Array(0)

```

Figura 51. Datos ordenados para el diagrama de barras de selección múltiple.

A continuación, en la misma función, indicaremos en el eje Y el valor del dato correspondiente al registro, y el propio registro se asignará al eje X. También indicaremos a que dato seleccionado pertenece cada registro.

```

▼ 0:
  key: "Volumen Pulmon Derecho - Valor"
  ▼ values: (8) [{...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}]
  ▶ __proto__: Object
▼ 1:
  key: "Volumen Pulmon Izquierdo - Valor"
  ▼ values: (8) [{...}, {...}, {...}, {...}, {...}, {...}, {...}, {...}]
  ▶ __proto__: Object

```

Figura 52. Plantilla de los datos listos para el diagrama de barras de selección múltiple.

Finalmente, en la función `buildData` preparamos los datos para la plantilla del diagrama de barras de *Angular-nvD3.js*.

### 5.5.1.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Barras

Seleccione tipo

Selección múltiple

Seleccione los datos

Volumen Pulmon Derecho - Valor

Volumen Pulmon Izquierdo - Valor

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Paleta de colores

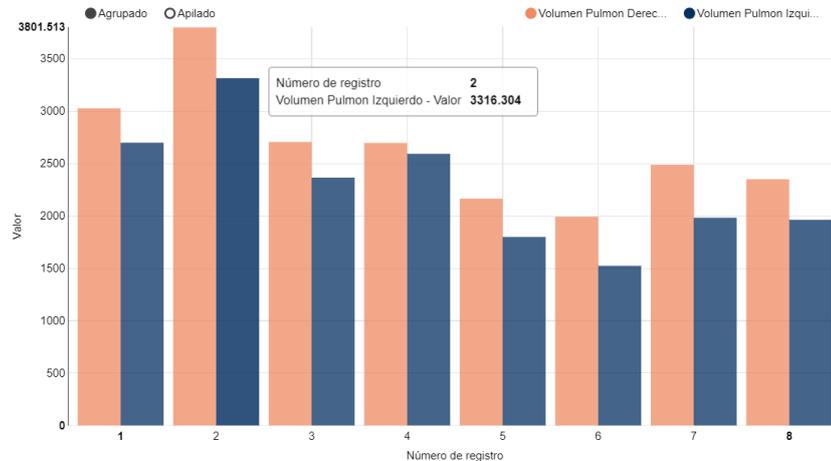


Figura 53. Resultado del diagrama de barras de selección múltiple.

En la Figura 53 apreciamos el resultado de los datos que hemos seleccionado. Si nos fijamos en el *tooltip*, muestra los datos de la barra sobre la que hemos posado el ratón, que se diferencia de las demás al estar un poco más oscurecida.

### 5.5.2 Bidimensional

Esta opción permite seleccionar dos datos, de los cuales es recomendable que el primer dato sea cuantitativo, mientras que el segundo, correspondiente al grupo, es recomendable que sea una variable categórica, de manera que se muestre una gráfica con los datos agrupados.

Seleccione tipo

Bidimensional

Seleccione dato

Volumen Pulmon Derecho - Va...

Seleccione grupo

Nombre de paciente

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Paleta de colores

Figura 54. Opciones del diagrama de barras bidimensional.

#### 5.5.2.1 Desarrollo

El dato que hemos seleccionado es el de *Volumen Pulmón Derecho – Valor* y los hemos agrupado en función del *Sexo del paciente*. De manera similar al diagrama de barras de selección múltiple, se va a replicar el objeto proveniente de la base de datos, solo que esta vez lo hará según el número de grupos. En este caso, lo replicará una vez para cada sexo.

```

▼ 1:
  key: "Sexo del paciente"
  mean: 2482.2764412
  median: 2489.414565
  series: 0
  x: "F"
  y: 12411.382206

```

**Figura 55. Muestra de un dato preparado para el diagrama de barras bidimensional.**

A continuación, aún dentro de la función `prepareData`, se realiza la suma de los datos correspondientes a cada grupo, la cual se asigna al eje Y. Al eje X se le asignan los grupos, que será el número de barras representadas. Además de eso, vamos a utilizar las funciones `get_mean` para obtener la media, y `get_percentile` para obtener la mediana (Percentil 50).

```

▼ [{...}] ⓘ
  ▼ 0:
    key: "Sexo del paciente"
    ▼ values: Array(2)
      ► 0: {x: "M", y: 8823.520220999999, mean: 2941.173407, median: 3028.747424, series: 0, ...}
      ► 1: {x: "F", y: 12411.382206, mean: 2482.2764412, median: 2489.414565, series: 0, ...}
        length: 2
      ► __proto__: Array(0)
    ► __proto__: Object
    length: 1

```

**Figura 56. Plantilla de los datos listos para representar el diagrama de barras bidimensional.**

En la función `buildData` solamente preparamos los datos para la plantilla de datos de entrada del diagrama de barras de *Angular-nvD3.js*.

Finalmente, como peculiaridad de este modo de visualización, en la función de `buildOptions` vamos a hacer una llamada a la función `setOnTopLabel`, propia solamente de este modo de visualización.

```

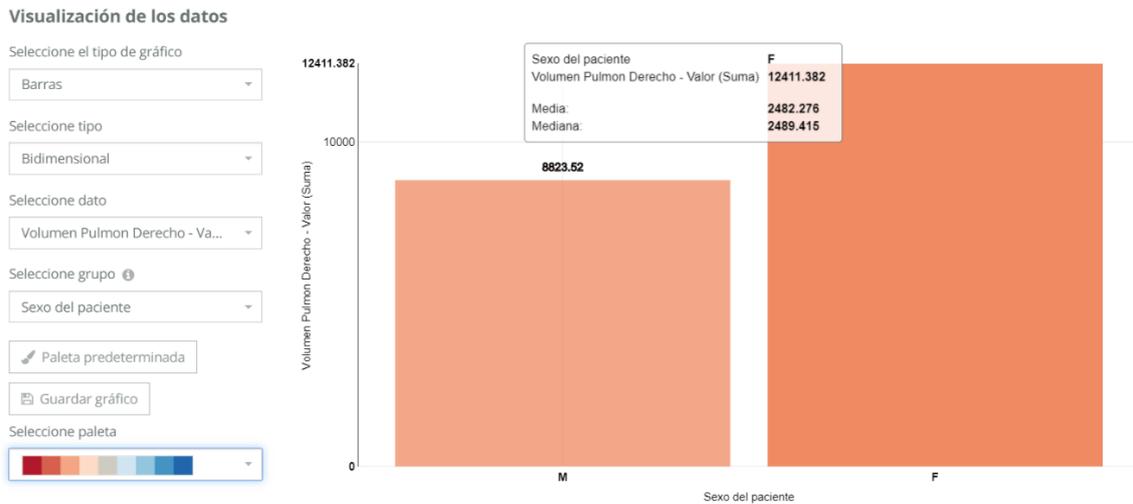
Object
  key: "Sexo del paciente"
  mean: 2482.2764412
  median: 2489.414565
  series: 0
  x: "F"
  y: 12411.382206
  ► __proto__: Object

```

**Figura 57. Objeto con los datos de la barra seleccionada.**

La función `setOnTopLabel` está desarrollada puramente con *D3.js*. Esta función, inspecciona las barras del diagrama ya representado y toma los datos, de los que coge la altura (La propiedad `y`), que representa el valor de la suma de los datos en ese grupo. A continuación, convierte el dato en un elemento visual de texto, espera a que terminen todas las animaciones del diagrama (Para evitar que se solapen animaciones o existan variables con el mismo nombre) y lo adjunta a su barra correspondiente.

### 5.5.2.2 Resultado



**Figura 58. Resultado del diagrama de barras bidimensional.**

En la Figura 58 podemos apreciar el resultado del diagrama de barras. Vemos como en la parte superior de cada barra se muestra la suma del dato dentro de cada grupo, y como la barra más grande hace de límite del máximo para el eje Y del diagrama.

En el *tooltip* se muestran los datos calculados de la media y la mediana de la barra sobre la que posamos el ratón, que se diferencia a la otra al estar un poco más oscurecida.

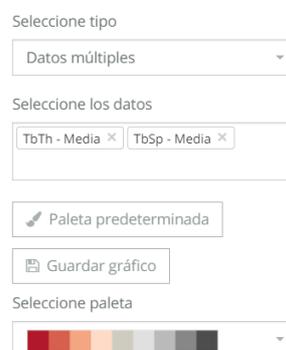
## 5.6 Histograma

El histograma representa la frecuencia absoluta y relativa (siendo la relativa la suma de todas las superficies 1, o sea el 100%) en función de las clases, siendo estas unos rangos del dato seleccionado, cuyo criterio se ha basado en la regla de Sturges<sup>21</sup> para determinar el número de clases (o barras) que hay, y a su vez determinando el rango.

La herramienta posee dos tipos de histogramas: datos múltiples y agrupados. En ambos se representa la frecuencia en función de las clases, que mostrarán el valor intermedio de cada rango. Para este apartado hemos seleccionado el biomarcador *Microarquitectura ósea 2D – QTS score v2*, que posee 60 registros.

### 5.6.1 Datos múltiples

La selección de múltiples datos permite seleccionar todos los datos que puedan estar relacionados para que el usuario pueda realizar una comparación.



**Figura 59. Opciones del histograma con selección de múltiples datos.**

<sup>21</sup> Es una regla práctica acerca del número de clases que deben considerar al elaborarse un histograma.

### 5.6.1.1 Desarrollo

Hemos seleccionado los datos *TbTh – Media* y *TbSp – Media*. De la misma manera que en el diagrama de barras de selección múltiple, el objeto proveniente de la base de datos se replicará por cada dato que seleccionemos, en este caso serán 2.

```
▼ 0:
  ▼ values: Array(7)
    ▶ 0: {LowLim: 178.076538, UppLim: 601.514, relative: 0.483, interval: "178.077 - 601.514", interval_mid: 389.795269, ...}
    ▶ 1: {LowLim: 601.514, UppLim: 1024.952, relative: 0.067, interval: "601.514 - 1024.952", interval_mid: 813.233, ...}
    ▶ 2: {LowLim: 1024.952, UppLim: 1448.39, relative: 0.3, interval: "1024.952 - 1448.390", interval_mid: 1236.671, ...}
    ▶ 3: {LowLim: 1448.39, UppLim: 1871.828, relative: 0.1, interval: "1448.390 - 1871.828", interval_mid: 1660.109, ...}
    ▶ 4: {LowLim: 1871.828, UppLim: 2295.266, relative: 0, interval: "1871.828 - 2295.266", interval_mid: 2083.547, ...}
    ▶ 5: {LowLim: 2295.266, UppLim: 2718.703, relative: 0.017, interval: "2295.266 - 2718.703", interval_mid: 2506.9845, ...}
    ▶ 6: {LowLim: 2718.703, UppLim: 3142.141113, relative: 0.033, interval: "2718.703 - 3142.141", interval_mid: 2930.4220565, ...}
    length: 7
  ▶ __proto__: Array(0)
  ▶ __proto__: Object
▼ 1:
  ▼ values: Array(7)
    ▶ 0: {LowLim: 265.195984, UppLim: 1132.815, relative: 0.483, interval: "265.196 - 1132.815", interval_mid: 699.005492, ...}
    ▶ 1: {LowLim: 1132.815, UppLim: 2000.435, relative: 0.4, interval: "1132.815 - 2000.435", interval_mid: 1566.625, ...}
    ▶ 2: {LowLim: 2000.435, UppLim: 2868.054, relative: 0, interval: "2000.435 - 2868.054", interval_mid: 2434.2445, ...}
    ▶ 3: {LowLim: 2868.054, UppLim: 3735.674, relative: 0.083, interval: "2868.054 - 3735.674", interval_mid: 3301.864, ...}
    ▶ 4: {LowLim: 3735.674, UppLim: 4603.293, relative: 0, interval: "3735.674 - 4603.293", interval_mid: 4169.4835, ...}
    ▶ 5: {LowLim: 4603.293, UppLim: 5470.913, relative: 0, interval: "4603.293 - 5470.913", interval_mid: 5037.102999999999, ...}
    ▶ 6: {LowLim: 5470.913, UppLim: 6338.532227, relative: 0.033, interval: "5470.913 - 6338.532", interval_mid: 5904.7226135, ...}
    length: 7
  ▶ __proto__: Array(0)
  ▶ __proto__: Object
length: 2
```

Figura 60. Datos ordenados para el histograma de selección múltiple.

En la Figura 60 observamos que preparamos los datos con muchas más propiedades de las que tenía el diagrama de barras, y es que todas estas serán necesarias para poder interpretarlo como un histograma.

```
LowLim: 1024.952
UppLim: 1448.39
interval: "1024.952 - 1448.390"
interval_mid: 1236.671
key: "TbTh - Media"
relative: 0.3
series: 0
x: 3
y: 18
```

Figura 61. Muestra con las propiedades del histograma de selección múltiple.

Como se puede observar en la Figura 61, si tomamos una muestra, podemos inspeccionar todas las propiedades que posee cada barra:

- *LowLim*: Límite inferior del rango.
- *UppLim*: Límite superior del rango.
- *interval*: Cadena con el texto del rango.
- *interval\_mid*: Valor intermedio del rango.
- *key*: Bloque de datos al que pertenece.
- *relative*: Frecuencia relativa.
- *series*: Serie a la que pertenece de los bloques de datos.
- *x*: Clase (Eje X).
- *y*: Frecuencia absoluta (Eje Y).

Para obtener todas las propiedades se recurre a la regla de Sturges, que viene dada por la siguiente expresión:

$$c = 1 + \log_2(M) \quad (6)$$

Donde  $M$  es el tamaño de la muestra, en nuestro caso, el número de registros y  $c$  el número de clases, que para nosotros sería el número de barras.

En este apartado nuestro biomarcador posee 60 registros, aplicando Sturges y redondeando el resultado, obtenemos que el número de clases (o barras) será 7.

Basándose en el número de clases, dividiremos el dato seleccionado en intervalos proporcionales. Con la ayuda de un contador que inspecciona los 60 registros del dato, indicamos el número de valores que se encuentran dentro de cada intervalo, siendo esto la frecuencia absoluta. A partir de la frecuencia absoluta, calculamos la frecuencia relativa respecto al número total de registros.

```

▼ 0:
  key: "TbTh - Media"
  ▶ values: (7) [{"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}]
  ▶ __proto__: Object
▼ 1:
  key: "TbSp - Media"
  ▶ values: (7) [{"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}]
  ▶ __proto__: Object
length: 2
  
```

Figura 62. Plantilla de los datos listos para representar el histograma de selección múltiple.

A efectos prácticos, el histograma no deja de ser un diagrama de barras, por lo que en la función de `buildData` preparamos los datos para la plantilla del diagrama de barras de *Angular-nvd3.js*.

### 5.6.1.2 Resultado

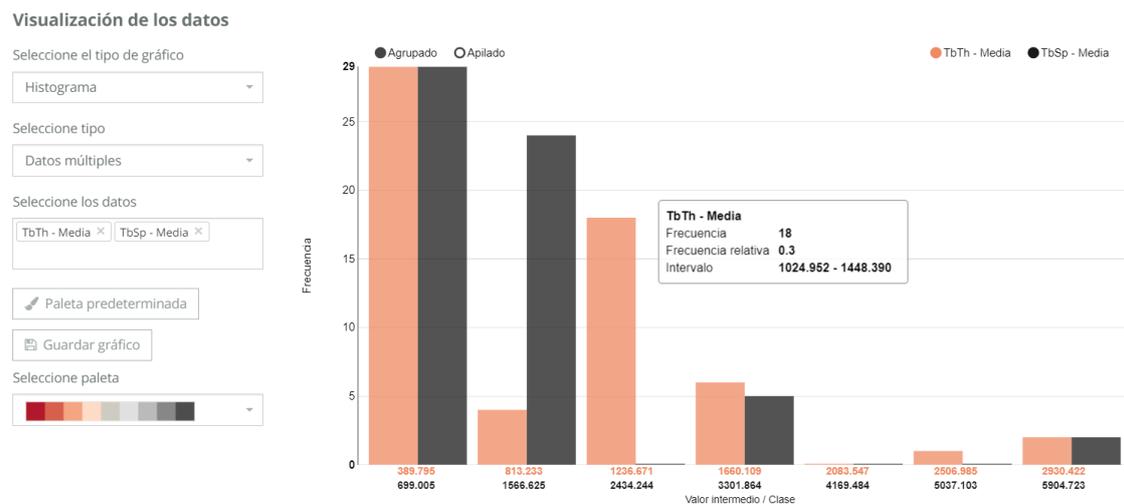


Figura 63. Resultado del histograma con selección de múltiples datos.

En la Figura 63 observamos que los datos se han dividido en 7 barras. En el *tooltip* se pueden ver los datos de la muestra que hemos inspeccionado antes, donde se muestra el intervalo y las frecuencias absoluta y relativa, siendo 18 el 30% de 60. En la gráfica se puede apreciar a simple vista la distribución de los valores.

Si nos fijamos en el eje X, veremos que se muestran los valores intermedios de cada rango de los dos datos seleccionados con el mismo color que las barras. Según el número de datos seleccionados, el eje X se irá adaptando.

Para hacer esta parte, se desarrolló en Jade un cuadro de texto dinámico, ya que *Angular-nvd3.js* no permitía hacer nada parecido, y es un elemento bastante útil para esta visualización.

### 5.6.2 Datos agrupados

La selección de datos agrupados permite al usuario seleccionar dos datos, siendo el primero un dato cuantitativo y el segundo un dato categórico, para así poder clasificar y comparar su distribución en la gráfica.

Figura 64. Opciones del histograma con selección por datos agrupados.

#### 5.6.2.1 Desarrollo

Hemos seleccionado el dato *TbTh – Media* y lo hemos agrupado por el valor categórico de *Modalidad*. Igual que vimos en el diagrama de barras bidimensional, el objeto proveniente de la base de datos se replicará según el número de grupos, que para este biomarcador serán las dos modalidades.

```

▼ 0:
  ▼ values: Array(7)
    ▶ 0: {LowLim: 178.076538, Upplim: 601.514, relative: 0.317, interval: "178.077 - 601.514", x: 1, ...}
    ▶ 1: {LowLim: 601.514, Upplim: 1024.952, relative: 0.033, interval: "601.514 - 1024.952", x: 2, ...}
    ▶ 2: {LowLim: 1024.952, Upplim: 1448.39, relative: 0.05, interval: "1024.952 - 1448.390", x: 3, ...}
    ▶ 3: {LowLim: 1448.39, Upplim: 1871.828, relative: 0, interval: "1448.390 - 1871.828", x: 4, ...}
    ▶ 4: {LowLim: 1871.828, Upplim: 2295.266, relative: 0, interval: "1871.828 - 2295.266", x: 5, ...}
    ▶ 5: {LowLim: 2295.266, Upplim: 2718.703, relative: 0.017, interval: "2295.266 - 2718.703", x: 6, ...}
    ▶ 6: {LowLim: 2718.703, Upplim: 3142.141113, relative: 0.033, interval: "2718.703 - 3142.141", x: 7, ...}
    length: 7
    ▶ __proto__: Array(0)
  ▶ __proto__: Object
▼ 1:
  ▼ values: Array(7)
    ▶ 0: {LowLim: 178.076538, Upplim: 601.514, relative: 0.167, interval: "178.077 - 601.514", x: 1, ...}
    ▶ 1: {LowLim: 601.514, Upplim: 1024.952, relative: 0.033, interval: "601.514 - 1024.952", x: 2, ...}
    ▶ 2: {LowLim: 1024.952, Upplim: 1448.39, relative: 0.25, interval: "1024.952 - 1448.390", x: 3, ...}
    ▶ 3: {LowLim: 1448.39, Upplim: 1871.828, relative: 0.1, interval: "1448.390 - 1871.828", x: 4, ...}
    ▶ 4: {LowLim: 1871.828, Upplim: 2295.266, relative: 0, interval: "1871.828 - 2295.266", x: 5, ...}
    ▶ 5: {LowLim: 2295.266, Upplim: 2718.703, relative: 0, interval: "2295.266 - 2718.703", x: 6, ...}
    ▶ 6: {LowLim: 2718.703, Upplim: 3142.141113, relative: 0, interval: "2718.703 - 3142.141", x: 7, ...}
    length: 7
    ▶ __proto__: Array(0)
  ▶ __proto__: Object
length: 2
  
```

Figura 65. Datos ordenados para el histograma por datos agrupados.

En la Figura 65 observamos que las propiedades de los datos preparados son las mismas que en el histograma de selección múltiple. Los intervalos son los mismos, ya que el dato es el mismo para los dos grupos, pero si nos fijamos en la propiedad *relative*, la frecuencia relativa es distinta, ya que esta vez la hemos calculado en función del grupo al que pertenece cada bloque. El primer bloque pertenece a la modalidad CT (TC) y el segundo bloque a la modalidad DX (Rayos X).

```

LowLim: 1024.952
Upplim: 1448.39
interval: "1024.952 - 1448.390"
key: "DX"
relative: 0.25
series: 1
x: 3
y: 15
  
```

Figura 66. Muestra con las propiedades del histograma por datos agrupados.

Si nos fijamos en la Figura 66, vemos que la única propiedad nueva que poseen ahora los datos es la propiedad *key*. La propiedad *key* indica al grupo que pertenece cada conjunto de datos. Para agruparlos se ha utilizado una metodología similar a la que hemos visto en los demás diagramas que permiten separar los datos en grupos. La única diferencia fundamental ha sido el uso de un filtrado para descartar los valores pertenecientes al otro grupo, para obtener la frecuencia absoluta y relativa a partir del tamaño de la matriz numérica en que se almacenan los valores de su correspondiente grupo.

```

▼ 0:
  key: "CT"
  ▶ values: (7) [{"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}]
  ▶ __proto__: Object
▼ 1:
  key: "DX"
  ▶ values: (7) [{"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}, {"..."}]
  ▶ __proto__: Object
  
```

Figura 67. Plantilla de los datos listos para representar el histograma por datos agrupados.

El cálculo de las demás propiedades ha seguido los mismos criterios que el histograma con selección de múltiples datos. La plantilla generada en la función de `buildData` también es la misma, solamente integra la nueva propiedad *key*.

### 5.6.2.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Histograma

Seleccione tipo

Datos agrupados

Seleccione dato

TbTh - Media

Seleccione grupo

Modalidad

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Paleta de colores

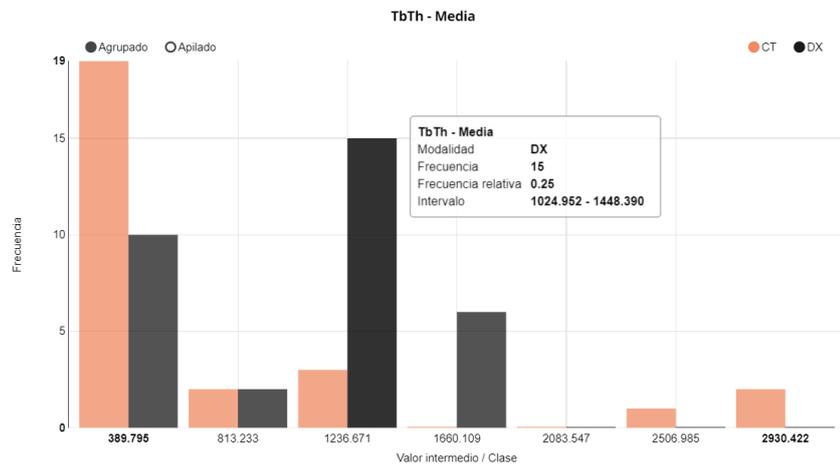


Figura 68. Resultado del histograma por datos agrupados (Agrupado).

En la Figura 68 se muestra la distribución del dato seleccionado en función de los grupos. Observamos que en cuanto a interacción no posee novedades respecto al otro histograma, la única diferencia es que se han hecho dos cálculos para una misma población.

### Visualización de los datos

Seleccione el tipo de gráfico

Histograma

Seleccione tipo

Datos agrupados

Seleccione dato

TbTh - Media

Seleccione grupo

Modalidad

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Paleta de colores

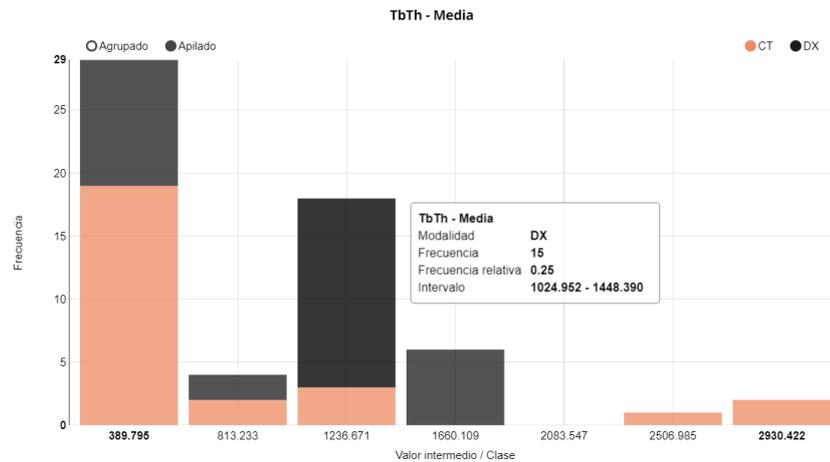


Figura 69. Resultado del histograma por datos agrupados (Apilado).

También está opción de mostrar los datos apilados para mostrar la distribución del dato seleccionado, con la opción de diferenciar los grupos dentro de una misma barra.

## 5.7 Bland-Altman

A diferencia de lo que hemos visto hasta ahora, esta visualización posee unas opciones mucho más limitadas, ya que este tipo de gráfica se limita a mostrar los datos calculados. Para trabajar con esta visualización, hemos seleccionado el biomarcador de *Lesiones de sustancia blanca*, de 140 registros.

Seleccione los datos del método A

Promedio Distancias - Media

Seleccione los datos del método B

Promedio Entropía - Media

Guardar gráfico

Figura 70. Opciones del método de Bland-Altman.

### 5.7.1 Desarrollo

#### 5.7.1.1 Selección de datos

Seleccionamos el dato *Promedia Distancias – Media* para el método A, y el dato *Promedio Entropía – Media* para el método B. Para este modo de visualización hemos utilizado la plantilla de diagramas múltiples de *Angular-nvD3.js*, ya que combina diagramas de dispersión y de líneas en la misma gráfica.

En este modo de visualización la función de `prepareData` será similar a la que vimos en el diagrama de dispersión. Únicamente inspeccionará los 140 registros del objeto proveniente de la base de datos y asignará los valores del dato seleccionado a las propiedades  $x$  (Para el método A) e  $y$  (Para el método B).

- ▶ 0: {key: "Scatter", color: "#6495ED", type: "scatter", values: Array(140), yAxis: 1}
- ▶ 1: {key: "Zero", color: "#708090", classed: "dashed", type: "line", values: Array(2), ...}
- ▶ 2: {key: "Bias", color: "#9370DB", type: "line", values: Array(2), yAxis: 1}
- ▶ 3: {key: "Limit Plus", color: "#DC143C", classed: "dashed-small", type: "line", values: Array(2), ...}
- ▶ 4: {key: "Limit Minus", color: "#DC143C", classed: "dashed-small", type: "line", values: Array(2), ...}

Figura 71. Plantilla de los datos listos para representar el gráfico Bland-Altman.

En la Figura 71 se muestran todos los elementos del diagrama que hemos dejado listos en la función `buildData: Scatter, Zero, Bias, Limit Plus` y `Limit Minus`.

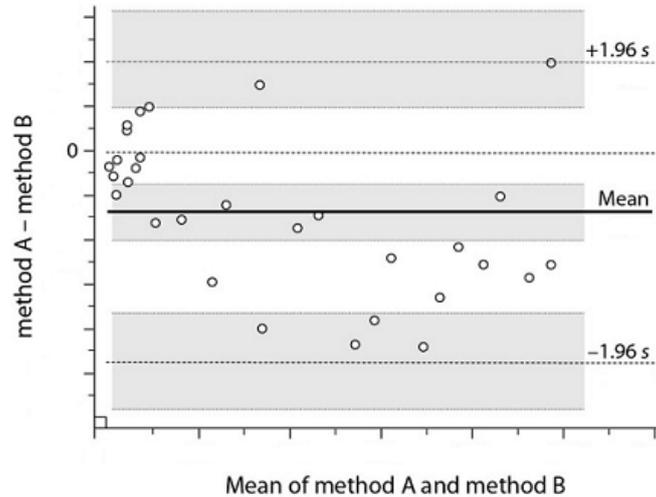


Figura 72. Estructura de un gráfico Bland-Altman. [10]

### Scatter

```
color: "#6495ED"
key: "Scatter"
type: "scatter"
▶ values: (140) [{...
```

Figura 73. Elemento *Scatter* del Bland-Altman.

Contiene los datos del diagrama de dispersión. En esta representación, el eje Y corresponde a las diferencias entre los valores de los métodos A y B ( $A - B$ ), mientras que el eje X representa el respectivo valor de la media de ambos ( $(A + B) / 2$ ).

En la Figura 73 vemos que la propiedad `type` indica a la plantilla de *Angular-nvD3.js* que se trata de un diagrama de dispersión, y vemos que contiene los valores, que vendrían a ser las coordenadas en el diagrama, de los 140 puntos. Para calcular los valores, se ha calculado la diferencia para cada registro entre ambas variables (Asignada al eje Y) y la media (Asignada al eje X) conforme recorremos los datos que hemos preparado en la función `prepareData`.

### Bias

```
color: "#9370DB"
key: "Bias"
type: "line"
▶ values: (2) [{...}, {...}]
```

Figura 74. Elemento *Bias* del Bland-Altman.

El elemento *Bias* contiene la línea donde se encuentra la media en el eje Y. La diferencia media (Sesgo o bias) en el diagrama es el valor medio determinado por la diferencia de los métodos A y B que refleja el error sistemático. Se ha hecho uso de la función `get_mean` con el resultado de la matriz numérica resultante de la resta de los métodos A y B.

En la plantilla se especifica que se trata de una línea. Veremos que en esta y en todas las líneas que vienen a continuación tan solo existen dos valores, ambos valores son las coordenadas de los dos puntos de la línea. En la propiedad `x` le indicamos los límites en el eje X del diagrama (Máximo y mínimo) y en la propiedad `y` el valor del dato que está representando, en este caso la diferencia media.

### Zero

```
classed: "dashed"  
color: "#708090"  
key: "Zero"  
type: "line"  
▶ values: (2) [{...}, {...}]
```

Figura 75. Elemento *Zero* del Bland-Altman.

Contiene los datos de la línea donde se encuentra el cero en el eje Y. Es importante remarcar que, si los métodos A y B obtiene de media valores similares, la diferencia media se situará en cero o próxima a cero, pero si se encuentran lejos de cero, significa que los dos métodos tienen valores muy diferentes.

Igual que con el anterior, en la propiedad *x* le indicamos los límites en el eje X del diagrama y en la propiedad *y* el valor del dato que está representando, en este caso el cero.

### Limit Plus

```
classed: "dashed-small"  
color: "#DC143C"  
key: "Limit Plus"  
type: "line"  
▶ values: (2) [{...}, {...}]
```

Figura 76. Elemento *Limit Plus* del Bland-Altman.

Contiene los datos de la línea del denominado límite de concordancia superior. Los límites de concordancia representan los límites o el intervalo de confianza del 95% para la diferencia de los métodos A y B.

En una representación gráfica, el intervalo de confianza es una técnica de estimación que permite acotar los valores de una población con una probabilidad determinada, en este caso del 95%.

$$CI_+ = +1,96 \cdot s \quad (7)$$

Donde  $CI_+$  es intervalo de confianza superior, aquí determinado límite de concordancia superior, y  $s$  es la desviación típica. Para calcular la desviación típica, utilizamos la función `get_standardDeviation` con la matriz numérica resultante de la resta de los métodos A y B. Finalmente, a partir de este valor se calcula el límite de concordancia superior.

### Limit Minus

```
classed: "dashed-small"  
color: "#DC143C"  
key: "Limit Minus"  
type: "line"  
▶ values: (2) [{...}, {...}]
```

Figura 77. Elemento *Limit Minus* del Bland-Altman.

Contiene los datos de la línea del denominado límite de concordancia inferior.

$$CI_- = -1,96 \cdot s \quad (8)$$

De la misma manera que con el elemento anterior, obtenemos la desviación típica ( $s$ ) con la función `standardDeviation` y calculamos el intervalo de confianza inferior ( $CI_-$ ).

### 5.7.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Bland-Altman

Seleccione los datos del método A

Promedio Distancias - Media

Seleccione los datos del método B

Promedio Entropía - Media

Guardar gráfico

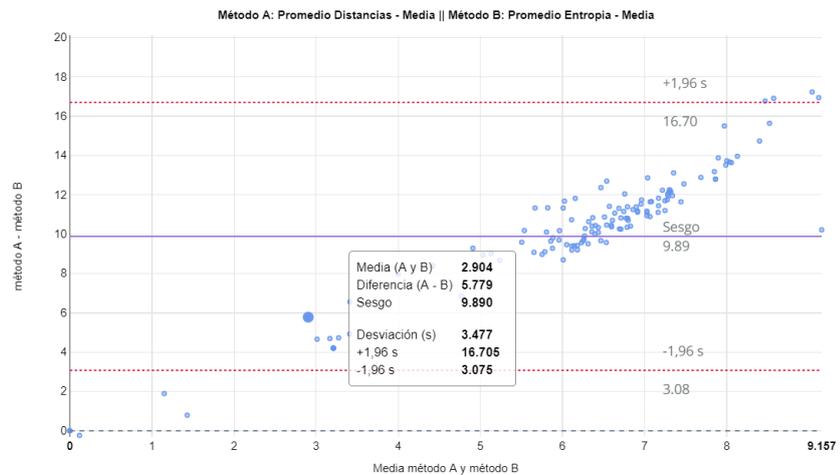


Figura 78. Resultado del método de Bland-Altman.

En la Figura 78 se muestra la distribución de los dos datos seleccionados, donde se puede apreciar que los resultados (valores) de los datos seleccionados son bastante diferentes.

El *tooltip* muestra los datos de todos los elementos que hemos mencionado en el apartado anterior. Además, si nos fijamos en las líneas, podemos ver que también muestra con texto el valor de los límites de concordancia y del sesgo. Ya que la plantilla de *Angular-nvD3.js* no nos permitía hacer nada similar, lo hemos añadido como un texto dinámico mediante Jade, de la misma forma que vimos en el histograma de selección múltiple.

### 5.8 Kaplan-Meier

Los datos de supervivencia que se mostrarán se basarán en el número de registros que tenga cada paciente. Para este ejemplo hemos utilizado el biomarcador de *Lesiones de sustancia blanca (CNN)* que posee 65 registros.

Comparar por ⓘ

Registro

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Predeterminado

Figura 79. Opciones del gráfico Kaplan-Meier.

En la Figura 79 podemos observar que permite la selección de un dato para comparar.

### 5.8.1 Desarrollo

Para representar este diagrama, utilizamos la plantilla de diagramas múltiples. Existen casos en los que se pueden encontrar registros con datos nulos o indefinidos, para un análisis de supervivencia se conocen como censura, si bien cuando se encuentre un dato de censura, lo representaremos con una cruz en el diagrama, por ello necesitamos combinar el diagrama de líneas con un diagrama de dispersión.

Desafortunadamente, los registros de los biomarcadores aún no poseen instantes temporales. Un instante temporal se trata de un dato que nos indica el orden en el que se han ido tomando los registros. Para lidiar con este problema, se ha desarrollado la función `get_Time`, propia de este modo de visualización.

La función `get_Time` requiere dos matrices de datos, los registros de todos los pacientes que hay en el biomarcador, y los nombres únicos de los pacientes que se encuentran dentro del mismo, que los obtiene con un filtrado. A partir de estas dos matrices de datos, la función comenzará un contador que asignará el instante temporal a los pacientes en el orden en que vayan llegando los registros.

#### 5.8.1.1 Selección del registro

En caso de seleccionar el registro, se muestra la función escalonada de los datos de todos los pacientes además del intervalo de confianza del  $\pm 95\%$ .

```
▶ 0: {Group: undefined, Patient: "LSTPrueba", Time: 1}
▶ 1: {Group: undefined, Patient: "TestWMLCarola01", Time: 1}
▶ 2: {Group: undefined, Patient: "TestWMLCarola01", Time: 2}
```

Figura 80. Muestra de los datos preparados para el análisis de supervivencia (Registro).

Si inspeccionamos los datos preparados, como se puede ver en la Figura 80, veremos que mostrará el grupo como indefinido, ya que no se ha seleccionado ningún dato para comparar.

```
▶ 0: {key: "Kaplan-Meier", values: Array(21), type: "line", yAxis: 1, originalKey: "Kaplan-Meier"}
▶ 1: {key: "+95% CI", classed: "dashed", values: Array(21), type: "line", yAxis: 1, ...}
▶ 2: {key: "-95% CI", classed: "dashed", values: Array(21), type: "line", yAxis: 1, ...}
▶ 3: {key: "50%", hiddenLegend: true, values: Array(2), color: "#708090", classed: "dashed-small", ...}
```

Figura 81. Plantilla de los datos listos para representar el análisis de supervivencia (Registro).

En la Figura 81 tenemos los datos listos para su representación, donde encontramos los bloques de datos: *Kaplan-Meier*, *+95% CI*, *-95% CI* y *50%*.

#### Kaplan-Meier

```
key: "Kaplan-Meier"
originalKey: "Kaplan-Meier"
type: "line"
▶ values: (21) [{...}, {...}, {...},
yAxis: 1
```

Figura 82. Elemento *Kaplan-Meier* del análisis de supervivencia.

Para calcular los valores del *Kaplan-Meier*, que se trata del análisis de supervivencia en sí, trabajamos con los instantes temporales. El paciente con el mayor número de instantes temporales marcará el límite del análisis, en este caso será 10.

Para el cálculo de la probabilidad de supervivencia (*St*) trabajamos con tres datos:

- *t*: Instante temporal.
- *Nt*: Número de pacientes en riesgo.
- *Dt*: Número de fallecimientos.

El número de pacientes en riesgo parte del número total de pacientes del estudio, en nuestro caso hay 27 pacientes.

En cada instante temporal que un número de pacientes muere, se resta dicho valor al número de pacientes en riesgo del instante temporal anterior. Para calcular la probabilidad de supervivencia, utilizamos la ecuación:

$$S_{t+1} = S_t \cdot \left( \frac{N_{t+1} - D_{t+1}}{N_{t+1}} \right) \quad (9)$$

Donde t+1 implica el dato que viene a continuación. De esta forma obtenemos la siguiente tabla:

t	Nt	Dt	St
0	27	0	1
1	27	9	0,555
2	18	12	0,185
3	6	3	0,093
4	3	1	0,062
5	2	0	0,062
6	2	0	0,062
7	2	0	0,062
8	2	0	0,062
9	2	1	0,031
10	1	0	0,031

Tabla 12. Cálculo de la probabilidad de supervivencia.

En la tabla se muestra el avance temporal en el análisis de supervivencia. En nuestra visualización vamos a tomar estos datos y adaptarlos a coordenadas, de manera que se pueda representar como una línea escalonada.

```
▶ 2: {x: 1, y: 0.5555555555555556, series: 0}
▶ 3: {x: 2, y: 0.5555555555555556, series: 0}
```

Figura 83. Muestra de los valores representados en el análisis de supervivencia (Registro).

En la Figura 83 vemos una muestra de cómo hemos adaptado el “escalón” del segundo instante temporal.

+95% CI

```
classed: "dashed"
key: "+95% CI"
originalKey: "+95% CI"
type: "line"
▶ values: (21) [...], {...},
yAxis: 1
```

Figura 84. Elemento +95% CI del análisis de supervivencia.

Para calcular el intervalo de confianza, tenemos que averiguar la estimación de supervivencia. Para ello, utilizamos la denominada fórmula de Greenwood:

$$SE(S_t) = S_t \cdot \sqrt{\sum \frac{D_t}{N_t(N_t - D_t)}} \quad (10)$$

Donde la cifra que se encuentran dentro del sumatorio está compuesta de valores que ya conocemos.

A partir de la fórmula obtenemos la siguiente tabla:

$\frac{D_t}{N_t(N_t - D_t)}$	$\sum \frac{D_t}{N_t(N_t - D_t)}$	$S_t \cdot \sqrt{\sum \frac{D_t}{N_t(N_t - D_t)}}$	$1,96 \cdot SE(S_t)$
0	0	0	0
0,018	0,018	0,075	0,147
0,111	0,129	0,066	0,129
0,166	0,296	0,050	0,098
0,166	0,462	0,042	0,082
0	0,462	0,042	0,082
0	0,462	0,042	0,082
0	0,462	0,042	0,082
0	0,462	0,042	0,082
0,5	0,962	0,030	0,059
0	0,962	0,030	0,059

Tabla 13. Cálculo del intervalo de confianza.

La columna final que se muestra en la tabla como  $1,96 \cdot SE(S_t)$  indica el margen de error del intervalo de confianza del 95%, a partir de este valor estimamos los límites del intervalo de confianza, en este caso para el superior:

$$CI_+(95\%) = S_t + 1,96 \cdot SE(S_t) \quad (11)$$

De la misma forma que en el bloque anterior, adaptamos los valores a coordenadas que se puedan representar como una línea escalonada.

Además, si prestamos atención a la Figura 84, veremos que existe una propiedad llamada *classed* en la que se indica la clase *dashed*. Este estilo lo hemos creado con la ayuda de Less, para indicar que esta línea se va a mostrar como discontinua.

-95% CI

```

classed: "dashed"
key: "-95% CI"
originalKey: "-95% CI"
type: "line"
▶ values: (21) [{...}, {...},
yAxis: 1

```

Figura 85. Elemento -95% CI del análisis de supervivencia.

Se trata del límite inferior del intervalo de confianza. Seguimos exactamente los mismos pasos que con el bloque anterior, la única diferencia es que la ecuación que empleamos será la siguiente:

$$CI_-(95\%) = S_t - 1,96 \cdot SE(S_t) \quad (12)$$

50%

```
    classed: "dashed-small"  
    color: "#708090"  
    key: "50%"  
    originalKey: "50%"  
    type: "line"  
  ▶ values: (2) [{...}, {...}]  
  yAxis: 1
```

**Figura 86. Elemento 50% del análisis de supervivencia.**

Finalmente, este elemento consiste en una línea discontinua a la que hemos aplicado el estilo *dashed-small* que posee las coordenadas de los límites del diagrama para representar una línea de donde se encuentra el 50% de la probabilidad de supervivencia.

### 5.8.1.2 Selección de un dato

En caso de seleccionar un dato para comparar, mostrará las líneas escalonadas de los grupos presentes en el dato que se ha seleccionado para comparar.

```
▶ 0: {Group: "F", Patient: "LSTPrueba", Time: 1}  
▶ 1: {Group: "M", Patient: "TestWMLCarola01", Time: 1}  
▶ 2: {Group: "M", Patient: "TestWMLCarola01", Time: 2}
```

**Figura 87. Muestra de los datos del análisis de supervivencia (Sexo del paciente).**

En la muestra de los datos preparados, veremos que ha clasificado los datos según el grupo. En este caso hemos seleccionado el *Sexo del paciente*.

```
▶ 0: {key: "F", values: Array(7), type: "line", yAxis: 1, originalKey: "F"}  
▶ 1: {key: "M", values: Array(21), type: "line", yAxis: 1, originalKey: "M"}  
▶ 2: {key: "50%", values: Array(2), color: "#708090", classed: "dashed-small", type: "line", ...}
```

**Figura 88. Plantilla de los datos del análisis de supervivencia (Sexo del paciente).**

En la Figura 88 tenemos los datos listos para su representación, donde encontramos los bloques de los grupos del dato seleccionado, *F* (Female) y *M* (Male).

El cálculo de los valores para su representación será exactamente el mismo método que vimos en el apartado de selección del registro. La única diferencia que tendrá será que separará los datos por grupos y calculará una función de supervivencia distinta para cada grupo.

Además, en este caso no se calculará ni se representará el intervalo de confianza del 95%, ya que saturaría la gráfica de líneas, y complicaría el uso de este modo de visualización al usuario.

```
▶ 2: {x: 1, y: 0.6923076923076923, series: 0, color: "#3182bd"}  
▶ 3: {x: 2, y: 0.6923076923076923, series: 0}
```

**Figura 89. Muestra de los valores del análisis de supervivencia (Sexo del paciente).**

## 5.8.2 Resultado

### 5.8.2.1 Selección del registro

#### Visualización de los datos

Seleccione el tipo de gráfico

Kaplan-Meier

Comparar por

Registro

Paleta predeterminada

Guardar gráfico

Seleccione paleta

Predeterminado

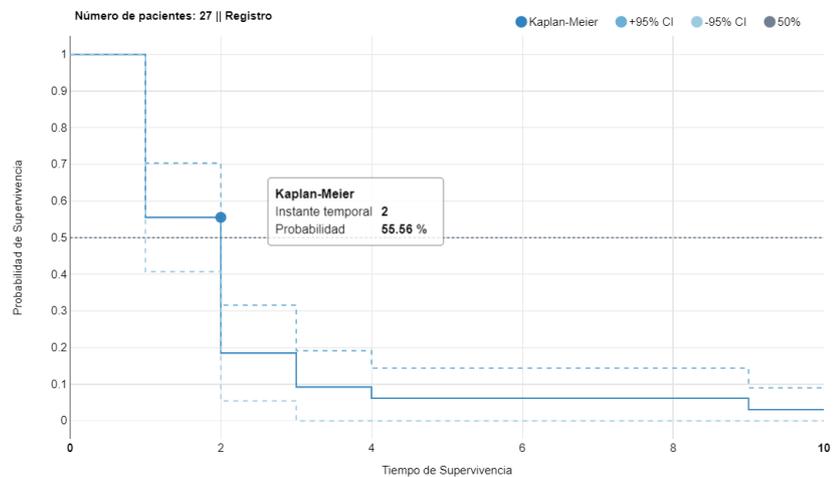


Figura 90. Resultado del gráfico Kaplan-Meier (Registro).

En la Figura 90 se muestra el aspecto de la línea escalonada que representa el diagrama de supervivencia. Podemos ver que los límites del intervalo de confianza se muestran con líneas discontinuas, y en el *tooltip* podemos ver la probabilidad de supervivencia en una escala del 100% en ese instante temporal.

### 5.8.2.2 Selección de un dato

#### Visualización de los datos

Seleccione el tipo de gráfico

Kaplan-Meier

Comparar por

Sexo del paciente

Paleta predeterminada

Guardar gráfico

Seleccione paleta

[Color palette]

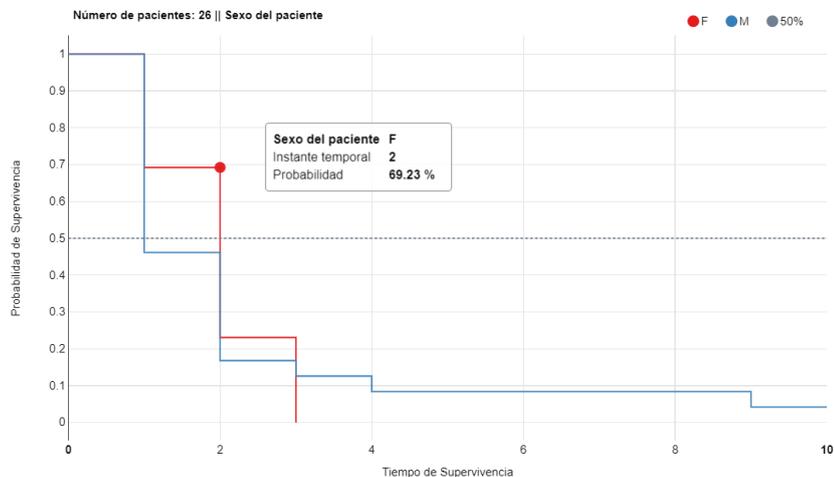


Figura 91. Resultado del gráfico Kaplan-Meier (Sexo del paciente).

Para el caso de la comparación por grupos vemos lo que ya hemos comentado. Se ha evitado la representación del intervalo de confianza para no saturar el diagrama y se muestran las líneas de los grupos, en este caso el sexo del paciente.

El *tooltip* en este caso añade un nuevo dato, en el que indica cuál de los dos grupos se está inspeccionando con el ratón.

## 5.9 Mapa de calor

El objetivo de este tipo de visualización es observar la distribución de los valores en el mapa de calor, que se basará en una paleta de colores gradientes de rangos equivalentes. Para el ejemplo que vamos a mostrar, hemos utilizado el biomarcador de *Estimador de Anormalidad en Radiografías de Tórax* que posee 81 registros.

Seleccione los datos del eje X  
Nombre del estudio

Seleccione los datos del eje Y  
Nombre de paciente

Seleccione valor  
Probabilidad de anormalidad -...

Guardar gráfico

Figura 92. Opciones del diagrama de calor.

### 5.9.1 Desarrollo

En las opciones, este diagrama solamente permite seleccionar valores categóricos para los ejes. Hemos seleccionado *Nombre del paciente* para el eje X y *Nombre del estudio* para el eje Y. El valor que hemos seleccionado para observar la distribución ha sido la *Probabilidad de anormalidad – Valor*.

La peculiaridad de este modo de visualización es que se ha desarrollado utilizando enteramente *D3.js*, es decir, a diferencia de los demás diagramas, no se ha utilizado ninguna plantilla de *Angular-nvD3.js*. Comparte estructura con los demás diagramas, por tanto, también utiliza las tres principales funciones de *prepareData*, *buildData* y *buildOptions*.

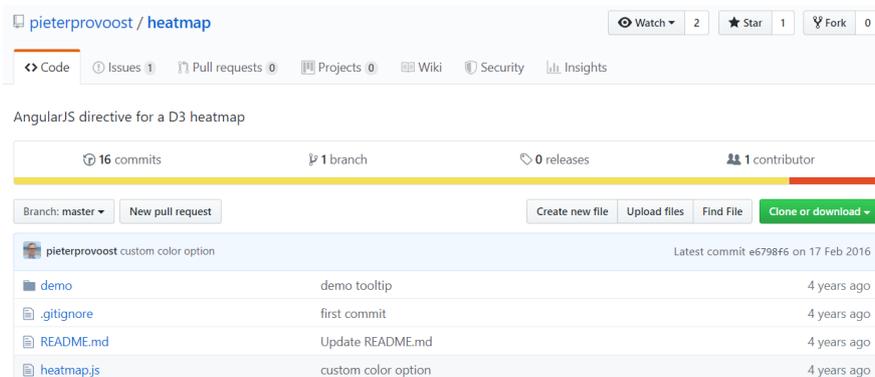


Figura 93. Repositorio *heatmap* del usuario *pieterprovoost*.

En GitHub encontramos el repositorio *heatmap.js* [11]. Hemos tomado como referencia su código para desarrollar nuestro mapa de calor, ya que este repositorio se ha diseñado como directiva de *AngularJS* para *D3.js*, en otras palabras, nos sirve como plantilla de manera similar a las que hemos utilizado de *Angular-nvD3.js*.

El fichero *heatmap.js* mostraba bastantes limitaciones, pero lo hemos modificado de tal manera que pueda recibir los datos de entrada en una estructura JSON por coordenadas, de manera muy similar a las plantillas de *Angular-nvD3.js*. Además, mediante *D3.js* también hemos conseguido que muestre los títulos de los datos seleccionados en el diagrama, además de un *tooltip*, que siempre resulta bastante útil para el usuario.

- ▶ 12: {x: "PruebaToxax3", y: "chest", value: "0.36232126", xIndex: 0, yIndex: 0}
- ▶ 13: {x: "Torax6", y: "RXTORAXPAYLAT", value: "0.5377662", xIndex: 1, yIndex: 1}

Figura 94. Muestra de los datos preparados para el mapa de calor.

Como podemos ver en la Figura 94, la estructura de los datos nos es un muy familiar, nos indica los datos seleccionados para los ejes y el valor que tienen en común. El índice nos dice las coordenadas de su posición, ya que un mapa de calor realmente se trata de una matriz. En este ejemplo existen 8 nombres de paciente y 4 nombres de estudio, por tanto, se consideraría una matriz de 8x4.

```
value: "Probabilidad de anormalidad - Valor"
x_axis: "Nombre de paciente"
y_axis: "Nombre del estudio"
```

Figura 95. Parámetros del mapa de calor.

También es necesario indicarle los nombres de los datos (Figura 95), para poder adjuntarlos como elementos visuales de texto mediante *D3.js* al mapa de calor.

```
buckets: 9
▶ colors: (9) ["#ffffcc", "#ffeda0", "#fed976", "#feb24c", "#fd8d3c", "#fc4e2a", "#e31a1c", "#bd0026", "#800026"]
▼ margin:
  bottom: 100
  left: 150
  right: 40
  top: 150
```

Figura 96. Opciones del mapa de calor.

Finalmente, en la Figura 96 observamos que las opciones para el mapa de calor poseen una estructura muy similar a la que vimos en *Angular-nvD3.js*. La paleta de colores gradientes viene predeterminada a un tamaño de 9 colores. La longitud de la paleta indica en cuantos rangos equivalentes se dividirá el valor.

### 5.9.2 Resultado

#### Visualización de los datos

Seleccione el tipo de gráfico

Mapa de calor

Seleccione los datos del eje X

Nombre de paciente

Seleccione los datos del eje Y

Nombre del estudio

Seleccione valor

Probabilidad de anormalidad - ...

Guardar gráfico

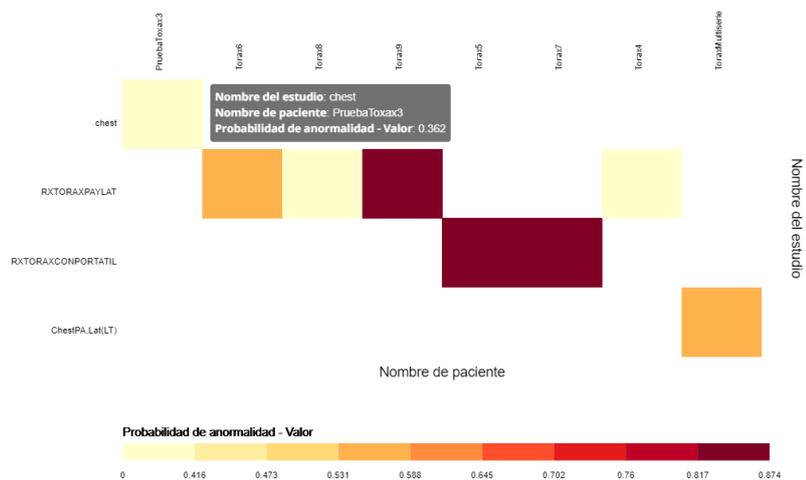


Figura 97. Resultado del mapa de calor.

Finalmente, en la Figura 97 se muestra el ejemplo del diagrama de calor. Podemos diferenciar aproximadamente la distribución del valor seleccionado. En la parte inferior se indican los rangos del gradiente, sin embargo, vemos que existen sectores en blanco. Esto es debido a que no todos los datos que se han seleccionado para los ejes X e Y poseen valores en común.

## 5.10 Paleta de colores

Si hemos estado atentos a las opciones de los modos de visualización, habremos visto que la mayoría de los diagramas poseen una opción para seleccionar una paleta de colores. Las únicas excepciones son el gráfico Bland-Altman y el mapa de calor, ya que ambos poseen una paleta de colores preestablecida para garantizar su funcionamiento óptimo.

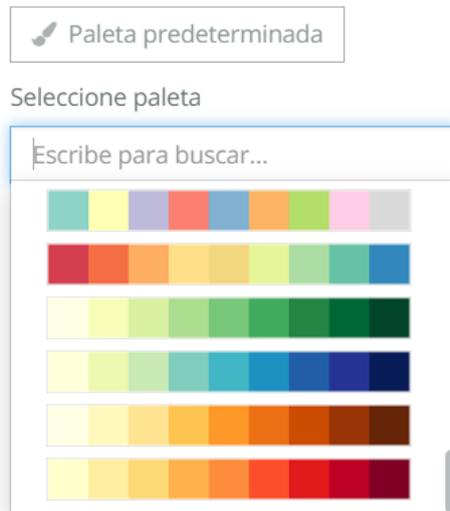


Figura 98. Opciones de selección de paleta.

En la Figura 98 se puede apreciar un menú desplegable con algunas de las opciones que muestra nuestra paleta de colores. Damos esta opción al usuario para permitir una mayor personalización de las gráficas, ya que permite llegar a resultados más agradables.

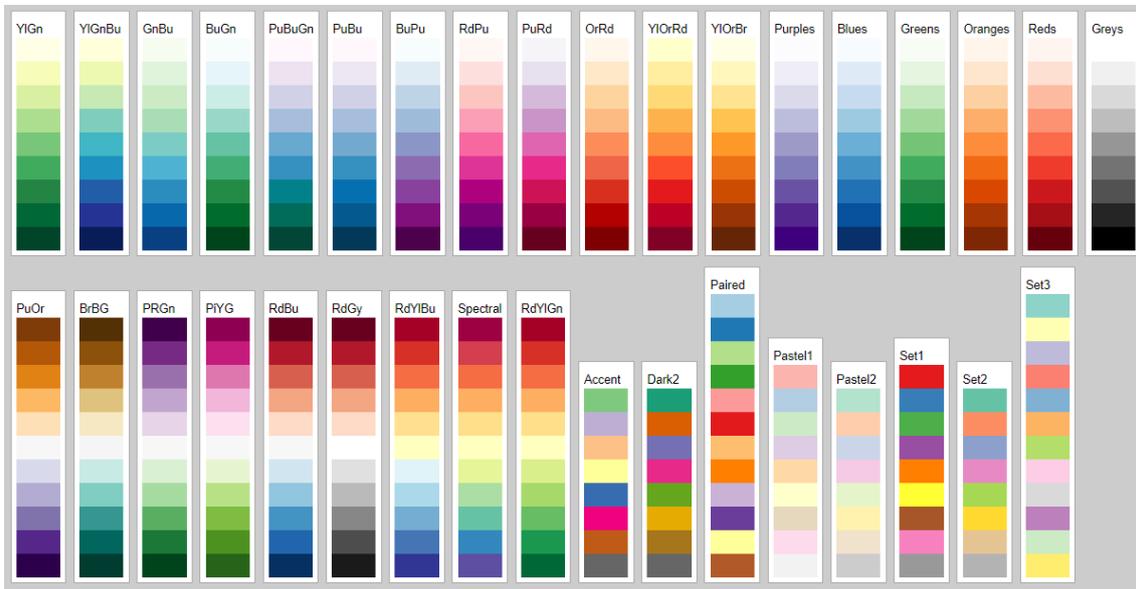


Figura 99. Paleta de colores de la escala *ColorBrewer*. [12]

Para el desarrollo del selector de paletas de colores, hemos utilizado la escala *ColorBrewer*, basada en el estudio de la cartógrafa Cynthia A. Brewer.

Las paletas de la escala *ColorBrewer* están pensadas en lograr un diseño de mapas que resulte agradable, sin embargo, estas paletas gozan de gran popularidad en el diseño de gráficas estadísticas, debido a su riqueza visual.

En la Figura 99 se muestran todas las paletas de la escala *ColorBrewer*. Nosotros hemos tomado estas paletas y las hemos predefinido como un fichero de JavaScript con los nombres que se muestran. El fichero *ColorBrewer.js* dispone de una selección de colores definidos en hexadecimal, que permiten un tamaño de 1 a 9 colores por cada paleta.

En caso de no querer utilizar ninguna de las paletas definidas, existe una paleta predeterminada que admite hasta 20 colores diferentes.

```
key: "Accent"  
▼ value:  
  ▶ 1: ["#7fc97f"]  
  ▶ 2: (2) ["#7fc97f", "#beaed4"]  
  ▶ 3: (3) ["#7fc97f", "#beaed4", "#fdc086"]  
  ▶ 4: (4) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99"]  
  ▶ 5: (5) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0"]  
  ▶ 6: (6) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0", "#f0027f"]  
  ▶ 7: (7) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0", "#f0027f", "#bf5b17"]  
  ▶ 8: (8) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0", "#f0027f", "#bf5b17", "#666666"]  
  ▶ 9: (9) ["#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0", "#f0027f", "#bf5b17", "#666666", "#A3A3A3"]
```

Figura 100. Ejemplo del objeto al seleccionar una paleta.

El menú desplegable de selección de paleta va ligado a la función `logSelection`, propia de la herramienta de visualización. Esta función se encarga de seleccionar el tamaño adecuado de la paleta en función de la cantidad de datos que se requieran representar en la gráfica.

A su vez, la función `logSelection`, dependerá de la función global `set_chart_id`. Esta función se encargará actualizar la gráfica seleccionada en ese momento, ya que no todas las gráficas se actualizan de la misma forma. En caso de encontrarse que la paleta seleccionada no tenga suficientes colores para el número de datos que se quieran representar, saltará a la función `default_palette`.

La función `default_palette` no tiene otra utilidad salvo que regresar a la paleta predeterminada. El botón de paleta predeterminada activará esta misma función.

## 5.11 Captura de la gráfica



Figura 101. Botón de captura de la gráfica.

La única opción común de todas las gráficas se trata del botón de guardar gráfico. Como el nombre indica, la utilidad de este botón es guardar en el equipo una captura de la gráfica que se está mostrando en ese momento en formato PNG. Para el desarrollo de esta opción ha sido indispensable el uso de dos librerías de GitHub: *saveSvgAsPng* [13] y *dom-to-image* [14].

### 5.11.1 Librerías

#### *saveSvgAsPng*

Como el propio nombre indica, permite guardar un elemento SVG del DOM como un PNG. Esta librería resulta enormemente útil, ya que todas las visualizaciones de la herramienta son elementos SVG. Además, esta librería posee otras utilidades, como modificar la proporción, cambiar la calidad de la imagen o cambiar la escala a la que se guarda el PNG.

#### *dom-to-image*

Si recordamos, en algunos diagramas añadíamos elementos dinámicos a la gráfica a través Jade, sin embargo, estos elementos no van enlazados al SVG que muestra la gráfica. Esto resultaba un problema, ya que la librería *saveSvgAsPng* solamente permite guardar elementos SVG.

Para arreglar este problema, hemos utilizado la librería *dom-to-image*. Esta librería posee una utilidad que permite anexas distintos elementos del DOM y transformarlos en un único SVG.

### 5.11.2 Desarrollo y resultado

Para capturar la gráfica, el botón de la Figura 101 va ligado a la función `save`, característica de la herramienta de visualización. Esta función utiliza las librerías que hemos mencionado en los apartados anteriores para hacer una captura de los elementos del DOM.

La única visualización que sólo hace uso de la librería `saveSvgAsPng` es el mapa de calor, ya que todos los elementos que se muestran en esta visualización se han insertado en un único SVG mediante `D3.js`.

El último paso de la función `save` consiste en adjuntar el logo de QUIBIM, mediante las herramientas de `D3.js`, a la gráfica que se haya guardado.

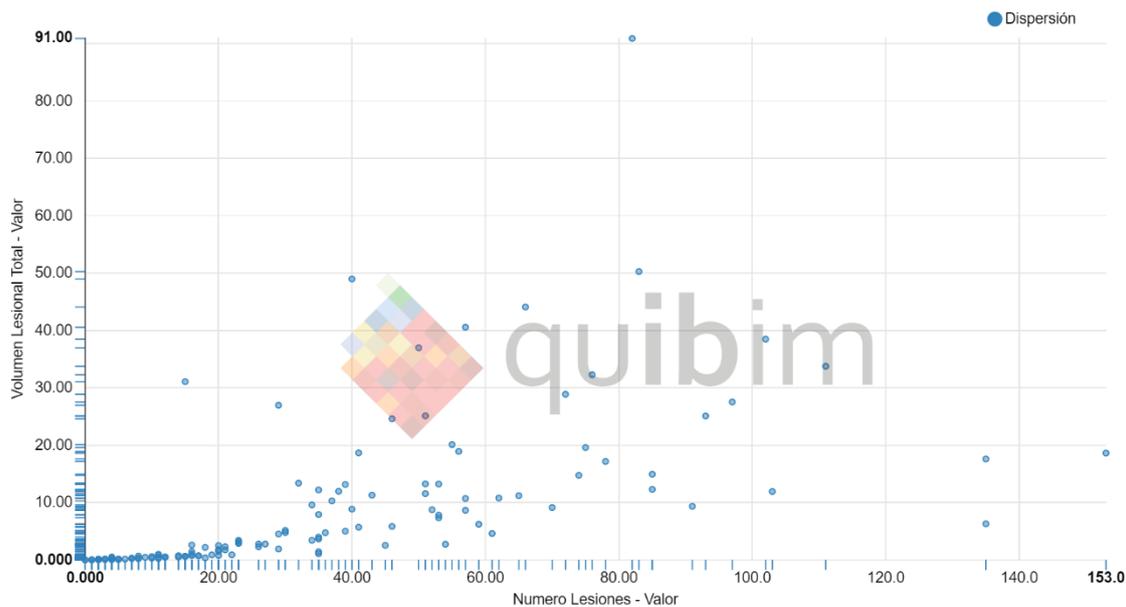


Figura 102. Resultado de la captura de la gráfica.

En la Figura 102 se muestra el resultado de la imagen guardada. Se muestra el logo, y cabe comentar que la resolución de la gráfica posee un tamaño 1,5 veces mayor al que se muestra en la web, para garantizar una resolución óptima al usuario.

## Capítulo 6. Conclusiones

### 6.1 Resultado del proyecto

El desarrollo de la “*Herramienta de visualización de datos provenientes del análisis masivo de imágenes médicas en la plataforma de QUIBIM Precision®*” se llevó a cabo en dos ciclos de prácticas, que fueron del 18 de marzo al 22 de abril, y del 28 de abril al 5 de julio, con una dedicación aproximada de 25 horas semanales durante estos ciclos.

La herramienta de visualización ha conseguido satisfacer todos los objetivos establecidos. Presenta un entorno para un uso profesional que no requiere indicaciones previas. En definitiva, la parte más importante de este proyecto es ofrecer una interfaz versátil y a prueba de fallos, ya que el usuario en realidad desconoce todo lo que hay detrás para garantizar su funcionamiento.

La librería *Angular-nvD3.js* está pensada para trabajar con una base de plantillas bastante específica y limitada, pero con este proyecto se ha podido demostrar que el límite está en el ingenio del desarrollador para “forzar” a la librería a mostrar lo que el desarrollador se ha propuesto.

El desarrollo de este proyecto, a nivel personal, ha supuesto un incremento de conocimientos en el entorno de desarrollo web bastante positivo. Al enfrentarse a un terreno desconocido, ha sido necesaria una amplia búsqueda bibliográfica que ha ocupado una parte importante del tiempo dedicado al proyecto.

Personalmente, como el autor de este proyecto, considero que en el tiempo total dedicado al proyecto podía haber conseguido llegar a presentar un trabajo más atractivo, sin embargo, eso no significa que no esté contento con el resultado.

### 6.2 Planes futuros

Al tratarse de una herramienta pensada para mostrar diferentes tipos de visualización, no se había propuesto un límite en los modos de visualización. En total se ha conseguido desarrollar 8 modos nuevos, y modificar los 2 ya existentes. Esto implica el límite de los modos de visualización que pueda contener la herramienta está en la utilidad que puedan tener para los datos con los que trabaja la plataforma.

Cabe decir, que la herramienta en sí no es exigente y siempre se ha buscado la máxima optimización posible en la implementación del código, sin embargo, nunca se ha testado en un equipo de bajas prestaciones, con lo que se desconocen los requisitos mínimos para ejecutar la herramienta.

La herramienta quedará a disposición de QUIBIM SL para su testeo y posterior integración en la versión actual de la plataforma QUIBIM Precision®.



## Bibliografía

- [1] Gemma Vialcanet. *Visualización de Datos: ¿Qué dicen los Expertos?:* <https://dbi.io/es/blog/visualizacion-de-datos-que-dicen-los-expertos/>
- [2] Imagen tomada de: <https://www.tableau.com/es-es/products/what-is-tableau>
- [3] Stephen SF Yip, Hugo J.W.L. Aerts. *Applications and limitations of radiomics:* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4927328/>
- [4] Alberich-Bayarri Á., Hernández-Navarro R., Ruiz-Martínez E., García-Castro F., García-Juan D., Martí-Bonmatí L. *Development of imaging biomarkers and generation of big data:* <https://www.ncbi.nlm.nih.gov/pubmed/28224398>
- [5] Jong Jin Hyun. *Rare case of pancreatic cancer with leptomeningeal carcinomatosis:* [https://www.researchgate.net/publication/271594471\\_Rare\\_case\\_of\\_pancreatic\\_cancer\\_with\\_leptomeningeal\\_carcinomatosis](https://www.researchgate.net/publication/271594471_Rare_case_of_pancreatic_cancer_with_leptomeningeal_carcinomatosis)
- [6] James Wiggins. *De-identify medical images with the help of Amazon Comprehend Medical and Amazon Rekognition:* <https://aws.amazon.com/es/blogs/machine-learning/de-identify-medical-images-with-the-help-of-amazon-comprehend-medical-and-amazon-rekognition/>
- [7] *QUIBIM Precision® 2.3 User manual:* [https://quibim.com/wp-content/uploads/2018/02/QUIBIM-Precision-2.3-User-Manual\\_20180130.pdf](https://quibim.com/wp-content/uploads/2018/02/QUIBIM-Precision-2.3-User-Manual_20180130.pdf)
- [8] Imagen tomada de: <https://www.thetechieshouse.com/10-angular-4-charts-and-graphs/>
- [9] Repositorio: <https://github.com/Tom-Alexander/regression-js>
- [10] Davide Giavarina. *Understanding Bland Altman analysis:* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4470095/>
- [11] Repositorio: <https://github.com/pieterprovoost/heatmap>
- [12] Imagen tomada de: <http://bl.ocks.org/mhkeller/10504471>
- [13] Repositorio: <https://github.com/exupero/saveSvgAsPng>
- [14] Repositorio: <https://github.com/tsayen/dom-to-image>