

Trabajo Fin de Grado
Grado en Ingeniería de las Tecnologías de
Telecomunicación

Visualización de datos electromagnéticos
geolocalizados usando Matlab

Autor: Silvia Martín Carmona

Tutora: María José Madero Ayora

Dpto. Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021



Trabajo Fin de Grado
Ingeniería de Telecomunicación

Visualización de datos electromagnéticos geolocalizados usando Matlab

Autor:

Silvia Martín Carmona

Tutor:

María José Madero Ayora

Profesora Titular de Universidad

Dpto. de Teoría de la Señal y Comunicaciones

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2021

Trabajo Fin de Grado: Visualización de datos electromagnéticos geolocalizados usando Matlab

Autor: Silvia Martín Carmona

Tutora: María José Madero Ayora

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2021

El Secretario del Tribunal

A mi familia

A mis amigos

A mis maestros

Agradecimientos

En primer lugar, me gustaría mencionar a mi familia, que ha estado junto a mí y me ha apoyado incondicionalmente desde siempre y, concretamente, estos últimos años de grado. Ellos han sido los que me han proporcionado las fuerzas para seguir adelante en los momentos más complicados y los que han compartido mi felicidad en los momentos dichosos. Gracias a su motivación, comprensión y dedicación he podido ir avanzando con paso firme ante cada adversidad y me enorgullece ver cómo al final el esfuerzo tiene su recompensa, porque el conseguir llegar donde he llegado es, también, mi humilde manera de agradecerles toda su preocupación y de demostrarles mi inmenso cariño y gratitud.

Como no podría ser de otra manera, debo hacer un agradecimiento especial a mi segunda familia, mis amigos, no sólo aquellos que ya formaban parte de mi vida, sino también a todas las personas extraordinarias que he ido conociendo a lo largo de estos años de carrera y que, por supuesto, permanecerán conmigo durante mucho tiempo. El haber podido compartir experiencias tanto positivas como negativas durante tanto tiempo es algo que crea un vínculo especial y que une a personas que no se conocían de nada de una manera mágica.

Sin duda, dentro de este grupo quiero destacar al que para mí ha sido el descubrimiento más maravilloso de la carrera, mi otra mitad en esta aventura, y creo que no es difícil saber a quién me estoy refiriendo. Kike, no sabes cuánto he disfrutado de cada momento que hemos compartido, cada charla, cada risa, cada comida previa a una práctica y todos esos nervios previos a los exámenes que solo tú eras capaz de apaciguar. Gracias por haberme aguantado tanto tiempo y por enseñarme a ser mejor persona, a luchar y enfrentarse a la vida porque por muchos obstáculos que nos ponga, somos más que capaces de ganar la batalla. Aquí tendrás una amiga y una gran admiradora siempre que lo necesites.

Por último, quería agradecer a mi tutora, María José, todas esas horas que ha invertido en mí, escuchándome y aconsejándome. Gracias por tu infinita paciencia y por tu comprensión, porque me has apoyado en todo el transcurso del TFG y eso que a veces no te lo he puesto muy fácil. Ha sido un verdadero placer trabajar contigo y aprender tanto de ti, tanto en lo personal como en lo profesional.

Silvia Martín Carmona

Sevilla, 2021

El incremento exponencial, experimentado en los últimos años, en el uso de dispositivos inteligentes tales como *smartphones* u ordenadores ha llevado a la generación masiva de datos, con los que se podría reconstruir, sin ningún problema, todos los movimientos realizados por una persona en su día a día. Y es que cualquier acción que hagamos en la actualidad queda registrada: nuestras compras, transacciones, nuestro recorrido por la ciudad, y un sinfín de acciones más. La existencia de extensas redes de sensores conectados entre sí contribuye también a esta creación masiva de información, que es necesario gestionar y analizar de la manera adecuada.

En este contexto surge lo que hoy se conoce como *Big Data*, grandes y complejos conjuntos de datos, procedentes de fuentes muy diversas, que exceden las capacidades informáticas tradicionales de análisis y procesamiento. Es en este mundo de existencia masiva de información en el que cobra especial importancia una correcta visualización o representación gráfica de los datos. Si el análisis tradicional no es válido, se debe recurrir a estas otras herramientas que permitan observar y estudiar los datos, con el objetivo de extraer el conocimiento o las conclusiones necesarias que deriven en una acción posterior, como puede ser la toma de una decisión.

En el presente trabajo se va a hacer uso de un conjunto de datos, en concreto, de las medidas de nivel de potencia recibida en varias ubicaciones de una ciudad. Haciendo uso de esta información se presentarán diferentes maneras de visualizar los datos y la importancia de cada una de ellas para facilitar el análisis, la comprensión y la extracción de resultados a partir un conjunto de datos que, inicialmente, parecen inconexos.

The exponential increase, lived in the last years, in the use of smart devices, such as smartphones or computers has led to a massive generation of data, with which it is possible to reconstruct, without any problem, all the movements made by a person in his/her day-to-day life. And it is that any action we do today is recorded: our purchases, transactions, our steps across the city, and endless more actions. The existence of extensive sensor networks connected to each other also contributes to this massive creation of information, which must be managed and analyzed in the appropriate way.

In this context, what is known as *Big Data* emerges, large and complex sets of data, from very different sources, which exceed traditional computing capabilities for analysis and processing. It is in this world of massive existence of information where a correct visualization or graphic representation of the data takes on special importance. If the ordinary analysis is not valid or available, it is necessary to appeal to these other tools that allow observing and studying the data, with the aim of extracting the required knowledge or conclusions that result in a subsequent action, such as making a decision.

In the present project, we are going to make use of a data set, specifically, of the measurements of the received power level in several locations in a city. With this information, different ways of visualizing the data and the importance of each of them will be presented in order to make easier the analysis, understanding and results extraction from a set of data that, initially, seem unrelated.

Agradecimientos	ix
Resumen	xi
Abstract	xiii
Índice	xv
Índice de Tablas	xvii
Índice de Ilustraciones	xix
Notación	xxiii
1 Introducción a la visualización de datos	1
1.1. <i>Objetivos</i>	1
1.2. <i>Alcance</i>	1
1.3. <i>Estructura</i>	2
2 El origen: los datos	3
2.1. <i>Tipos de datos</i>	4
2.2. <i>Fuentes de datos</i>	5
3 Visualización de datos	8
3.1. <i>Tipos de representaciones de datos</i>	9
3.1.1. <i>Diagrama de Barras (Bar Chart)</i>	9
3.1.2. <i>Gráfico de Líneas (Line Chart)</i>	11
3.1.3. <i>Diagrama Circular (Pie Chart)</i>	12
3.1.4. <i>Diagrama de Puntos (Dot Plots)</i>	13
3.1.5. <i>Gráfico de Intervalos (Interval Plots)</i>	14
3.1.6. <i>Diagrama de Dispersión (Scatterplot)</i>	15
3.1.7. <i>Gráfico de Burbujas (Bubble Plot)</i>	17
3.1.8. <i>Gráfico de Superficie (Surface Plot)</i>	17
3.1.9. <i>Diagrama de Contorno (Contour Plot)</i>	18
3.1.10. <i>Mapas de Procesos (Process Maps)</i>	19
3.1.11. <i>Mapas Geográficos (GIS)</i>	21
3.1.12. <i>Histogramas: Distribución de Frecuencia</i>	22
3.1.13. <i>Infográficos (Infographics)</i>	23
3.2. <i>Herramientas de Visualización</i>	26
3.2.1. <i>Google Charts</i>	26
3.2.2. <i>Tableau</i>	28
3.2.3. <i>MATLAB</i>	32
3.3. <i>Tendencias en la Visualización de datos: Big Data</i>	34
4. Desarrollo práctico del trabajo	37
4.1. <i>Recolección de datos</i>	37
4.2. <i>Volcado de los datos en una tabla</i>	40
4.2.1. <i>Creación de tablas auxiliares</i>	42
4.3. <i>Representaciones gráficas</i>	43
4.3.1. <i>Gráficas dependientes del canal elegido</i>	43

4.3.2.	Gráficas dependientes de la ubicación elegida	46
4.3.3.	Gráficas de carácter general	¡Error! Marcador no definido.
5.	Interfaz Gráfica	51
5.1.	<i>Manual de Usuario</i>	51
5.2.1.	Menú Principal	52
6.	Conclusiones y líneas futuras	67
	Referencias	68

ÍNDICE DE TABLAS

3.1	Muestra de la frecuencia de cada intervalo en un Histograma	22
3.2	Creación de una tabla con los datos a representar	27
4.1	Equipamiento y características	38
4.2	Canales utilizados para la medida	39

ÍNDICE DE ILUSTRACIONES

2.1	Diseño de la información	3
2.2	Tipos de datos	5
2.3	Variabilidad existente de fuentes de datos	7
3.1	La Visualización como herramienta de comunicación	8
3.2	Ejemplo de Diagrama de Barras, con orientación vertical	10
3.3	Ejemplo de Diagrama de Barras con información agrupada	10
3.4	Ejemplo de Diagrama de Barras Apiladas, con orientación horizontal	11
3.5	Ejemplo de un Gráfico de Líneas	11
3.6	Ejemplo de Diagrama Circular	12
3.7	Comparación de un Diagrama de Barras y un Diagrama Circular	13
3.8	Ejemplo de <i>Pie of Pie</i>	13
3.9	Ejemplo de Diagrama de Puntos	14
3.10	Ejemplo de Gráfico de Intervalos	14
3.11	Ejemplo de Diagrama de serie temporal	15
3.12	Ejemplo de Diagrama de Dispersión	15
3.13	Ejemplo de Diagrama de Dispersión, con relación negativa de las variables	16
3.14	Ejemplo de Diagrama de Dispersión No Lineal	16
3.15	Ejemplo de Gráfico de Burbujas	17
3.16	Ejemplo de Gráfico de Superficie	18
3.17	Diagrama de Superficie de un material	18
3.18	Ejemplo de Diagrama de Contorno	19
3.19	Mapa de Procesos	19
3.20	Ejemplo de Mapa de Procesos	20
3.21	Leyenda utilizada para el Mapa de Procesos	21
3.22	Ejemplo de Mapa Geográfico	22
3.23	Ejemplo de Histograma	23
3.24	Diferencias entre Infografía y Visualización de Datos	24
3.25	Ejemplo de Infográfico	25
3.26	Logotipo de <i>Google Charts</i>	26
3.27	Ejemplo de representación con la herramienta <i>Google Charts</i>	28
3.28	Logotipo de la herramienta <i>Tableau</i>	29
3.29	Interfaz mostrada por la herramienta <i>Tableau</i>	29
3.30	Ejemplo de representación con la herramienta <i>Tableau</i>	31

3.31	Ejemplo de <i>Dashboard</i> creado con <i>Tableau</i>	32
3.32	Logotipo de la herramienta MATLAB	32
3.33	Interfaz mostrada por la herramienta MATLAB	33
3.34	BIG DATA, tendencia actual en el campo de los datos	34
4.1	Logotipo del IEEE DataPort	37
4.2	Diagrama general de un Nodo IoT/WSN alimentado por RFEH	37
4.3	Configuración del <i>Drive Test</i>	39
4.4	Fichero KML con los datos obtenidos	40
4.5	Variable que contiene los datos de entrada	41
4.6	Muestra de la variable que contiene los datos a analizar	41
4.7	Muestra de las variables <i>LatitudFinal</i> y <i>LongitudFinal</i>	42
4.8	Muestra de la variable <i>PotenciaGlobal</i>	42
4.9	Muestra de la variable <i>FrecuenciasGlobal</i>	43
4.10	Representación Geográfica	43
4.11	Selección de un punto para su análisis	44
4.12	Variación de la potencia en función de la latitud y longitud	45
4.13	Representación dependiente de la ubicación escogida	46
4.14	Selección de un punto para su análisis	47
4.15	Gráficas de carácter general	48
4.16	Visualización de potencia recibida en un punto concreto	49
4.17	Variación del promedio de potencia	50
5.1	Elaboración de una GUI	51
5.2	Menú Principal de la GUI	52
5.3	Menú desplegable con opciones de zona a elegir	52
5.4	Mensaje de error por la opción introducida	53
5.5	Interfaz gráfica correspondiente a la Avenida	53
5.6	Menú desplegable para elegir el canal deseado	54
5.7	Selección del canal 5 y del botón “Visualización gráficas dependientes del canal”	54
5.8	Representaciones dependientes del canal elegido	55
5.9	Menú desplegable para elegir ubicación deseada	55
5.10	Selección de la ubicación y del botón “Visualización gráficas dependientes de la ubicación”	56
5.11	Representaciones dependientes de la ubicación elegida	56
5.12	Selección del botón “Visualización Global”	57
5.13	Mensaje de error al no seleccionar una ubicación válida	57
5.14	Representación global de la Avenida bajo estudio	58
5.15	Menú General para la opción “Espacios Abiertos”	59
5.16	Interfaz gráfica “Espacio Abierto número 2”	59

5.17	Proceso de selección del canal 1 para el “Espacio Abierto número 2”	60
5.18	Gráficas del canal 1 para el “Espacio Abierto número 2”	61
5.19	Selección de la ubicación y el botón “Visualización gráficas dependientes de la ubicación”	61
5.20	Gráfica de la ubicación 4 del “Espacio Abierto número 2”	62
5.21	Gráficas de carácter general del “Espacio Abierto número 2”	62
5.22	Menú General para la opción “Calles”	63
5.23	Interfaz gráfica “Conjunto de Calles 1”	63
5.24	Proceso de selección del canal 10 en el “Conjunto de Calles 1”	64
5.25	Gráficas del canal 10 para el “Conjunto de Calles 1”	65
5.26	Selección de la ubicación 65 y el botón “Visualización gráficas dependientes de la ubicación”	65
5.27	Gráfica de la ubicación 65 del “Conjunto de Calles 1”	66
5.28	Gráficas de carácter general del “Conjunto de Calles 1”	66

NOTACIÓN

RAE	Real Academia Española
E-commerce	Comercio electrónico
World Wide Web	Red informática mundial
2D	Representaciones en 2 dimensiones
3D	Representación en tres dimensiones
GIS	Sistema de información geográfica (Geographic Information System)
HTML	Lenguaje de marcas de hipertexto (HyperText Markup Language)
VizQL	Lenguaje de consulta visual (Visual Query Language)
SQL	Lenguaje de consulta estructurado (Structured Query Language)
MATLAB	Laboratorio de matrices (MATrix LABoratory)
Machine Learning	Aprendizaje automático
HW	Hardware
FIR	Respuesta finita al impulso (Finite Impulse Response)
IIR	Respuesta infinita al impulso (Infinite Impulse Response)
5G	Quinta Generación de comunicaciones móviles (Fifth Generation)
LTE	Evolución a Largo Plazo (Long Term Evolution)
WLAN	Red Inalámbrica de Área Local (Wireless Local Area Network)
RF	Espectro de radiofrecuencia (RadioFrequency)
M2M	Máquina a Máquina (Machine to Machine)
KPI	Indicador clave de desempeño (Key Performance Indicator)
SW	Software
BI	Inteligencia empresarial (Business Intelligence)
IoT	Internet de las cosas (Internet Of Things)
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos
Energy Harvesting	Recolección de energía
WSN	Red Inalámbrica de Sensores (Wireless Sensor Network)
WIFI	Fidelidad inalámbrica (Wireless Fidelity)
RFEH	Recolección de energía de RF (Radio Frequency Energy Harvesting)
Drive Test	Prueba de conducción
HSA	Analizador de espectro portátil (Handheld Spectrum Analyzer)
SMA	Conector roscado para cable coaxial (SubMiniature versión A)
GPS	Sistema de posicionamiento Global (Global Positioning System)

RHCP	Polarización circular derecha (Right-Handed Circular Polarization)
KML	Lenguaje de marcado visual (Keyhole Markup Language)
MHz	MegaHercios (MagaHertz)
GUI	Interfaz gráfica de usuario (Graphical User Interface)

1 INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS

El gran mito de nuestro tiempo es que la tecnología es comunicación.

Libby Larsen

En la actualidad, estamos experimentando un cambio de paradigma en nuestro entorno personal. Aumenta cada vez más el uso de dispositivos inteligentes que, gracias a tecnologías como el IoT, están permanentemente conectados, generando y compartiendo gran cantidad de datos. Esta producción masiva de datos, con unas características particulares, se recogen bajo conceptos emergentes tales como el *Big Data*. Las tecnologías de *Big Data* pretenden sacar partido a esos grandes conjuntos de datos, que se presentan de una manera aparentemente inconexa. Estas técnicas tratan de agrupar los datos que muestran informaciones similares con el objetivo de obtener conclusiones sobre diversos campos de nuestra vida diaria.

Aún así, el análisis computacional tradicional no dispone, en la mayoría de las ocasiones, de la capacidad necesaria para almacenar, procesar y tratar este gran volumen de datos. Y, aunque pueda, los resultados que ofrece no son fáciles de comprender o interpretar, pues se presentan en formatos en los que percibir posibles patrones o tendencias en los datos se convierte en una tarea ardua. De ahí que cada vez adquiera mayor importancia la visualización de los datos. Crear gráficos que muestren los datos más relevantes y que permitan, de un simple vistazo, obtener conclusiones o juicios sobre un determinado conjunto de datos, facilita enormemente su comprensión, además de contribuir a la toma de decisiones basada en dichas conclusiones. Por tanto, este concepto de visualización implica dos aspectos: por un lado, la comprensión de los datos y, por otro, el empleo de los resultados como motor de acciones adicionales.

1.1. Objetivos

El objetivo principal de este proyecto es el de mostrar el notable valor que poseen hoy en día las herramientas de visualización como base para el análisis de datos. Para ello se va a emplear una base de datos, obtenida del repositorio *IEEE DataPort*, en la que se recogen, a partir de un conjunto de equipos tales como antenas y analizadores de espectro, datos relativos al nivel de potencia recibido en diferentes puntos geográficos de una ciudad. Esta potencia proviene de la transmisión de otros dispositivos colocados en la zona, como pueden ser radios o equipos inteligentes. El objetivo último del estudio será determinar en qué zonas convendrá el despliegue de técnicas de *Energy Harvesting* que permitan a otros dispositivos alimentarse a partir de la potencia presente en el ambiente.

Con esos datos disponibles, el objetivo del presente Trabajo Fin de Grado será realizar una cantidad determinada de representaciones de manera que sea posible analizar el nivel de potencia recibido en cada punto para estimar la validez, o no, de esa ubicación como sede de los dispositivos a alimentar.

Tanto para el análisis de los datos como para el trazado de las gráficas se va a utilizar como herramienta MATLAB, que posee una gran cantidad de funciones de representación y manejo de datos, además de resultar de fácil acceso y utilización.

1.2. Alcance

Se propone en este proyecto un alcance tal que cubra todos los aspectos relacionados, por una parte, con los datos, y, por otra, con todo lo referente a la visualización efectiva de éstos. Para la primera parte, se incluyen los diferentes tipos de datos que se recogen y las diversas fuentes de las que pueden proceder. Para la segunda, se

recogen las representaciones más comunes que se pueden realizar con un conjunto de datos dado, las principales herramientas que un usuario tiene a su disposición para la creación de gráficos visuales y, por último, se incluye un breve análisis del concepto conocido como *Big Data*. Además, para facilitar el análisis de los datos recogidos, se presenta una interfaz gráfica que permitirá al usuario interactuar y trabajar con los datos de una manera óptima.

1.3. Estructura

La estructura presente en el documento se puede dividir en 6 capítulos diferentes. El primero de ellos, en el que se encuentra el lector, supone una pequeña introducción acerca de los objetivos, alcance y la estructura seguida para el desarrollo completo del proyecto.

El segundo capítulo está dedicado al campo de los datos. En esta parte se exponen con detalle los diferentes tipos de datos que existen, partiendo de dos tipos de datos básicos: cualitativos y cuantitativos. Además, se explica una posible clasificación acerca de los tipos de fuentes de las que pueden proceder dichos datos, con especial detenimiento en las principales diferencias entre unas fuentes y otras.

El tercer capítulo está centrado en la visualización de datos. En este capítulo se presentan las principales ventajas que conlleva la correcta representación de los datos, haciendo hincapié en los diferentes diagramas y gráficos que se pueden crear. Para una mejor comprensión de cada tipo de gráfico, se incluyen aquí numerosos ejemplos, a partir de los cuáles se intentarán sacar las conclusiones oportunas de los datos representados. Por último, en esta parte se incluye una breve explicación del *Big Data*, mostrando sus principales características y ámbitos de aplicación actuales.

Una vez hecha la introducción teórica del proyecto, el cuarto capítulo muestra el desarrollo práctico llevado a cabo. Este desarrollo abarca desde la explicación de la base de datos escogida hasta la presentación de los resultados gráficos obtenidos a partir de dichos datos. Este capítulo tiene, por tanto, 3 partes diferenciadas: el proceso de recolección de datos, el tratamiento de éstos y, por último, el proceso de representación gráfica usando como herramienta MATLAB.

En sintonía con el cuarto capítulo, el quinto es el encargado de introducir al lector en la interfaz gráfica creada. En esta parte, se recoge una guía que le servirá al usuario para la utilización de la GUI, de una manera sencilla e interactiva.

Por último, el sexto capítulo contiene las conclusiones obtenidas a partir del trabajo realizado y una serie de posibles líneas futuras del mismo.

2 EL ORIGEN: LOS DATOS

El producto más valioso que conozco es la información.

Gordon Gekko

Actualmente vivimos en un mundo complejo. Nuestra sociedad debe enfrentarse diariamente a un escenario poco conocido hasta el momento, la sobreabundancia de datos y de información. Y es que vivimos rodeados de datos y en un mundo digitalizado en el que cualquiera de las acciones que llevamos a cabo origina o recopila, aún, más cantidad de información.

Estos datos se van a convertir en la base de cualquier proyecto que se quiera crear o de cualquier actuación que se quiera realizar, de ahí que sean sumamente importantes. Si recurrimos a la RAE, esta entidad define los datos como “Información dispuesta de manera adecuada para su tratamiento por una computadora” [25]. En esta definición se puede encontrar un aspecto que llama notablemente la atención: los datos, como se ha mencionado antes, se pueden considerar como un registro de actividad. Sin embargo, estos datos, aisladamente, no encierran información destacable para la sociedad, ya que, desde el punto de vista humano, para poder comprender dicha información es condición ineludible colocarla en su contexto. De esta manera, solo cuando somos capaces de contextualizar los datos, estructurarlos y procesarlos computacionalmente, los convertimos en información relevante que, visualizada de la manera adecuada en cada caso, puede ser comprendida y asimilada.



Ilustración 2.1. Diseño de la información

El análisis y procesamiento de los datos contribuye, a su vez, al descubrimiento de [1]:

- **Tendencias:** el término tendencia puede definirse en este ámbito como el modelo de comportamiento que siguen los elementos de un determinado entorno durante un período de tiempo concreto.
- **Patrones:** modelo cuyos componentes se van a repetir de una forma previsible.
- **Anomalías:** propiedad que se aplica, desde un punto de vista subjetivo, a aquellos datos que presentan condiciones extrañas o inusuales.
- **Comparaciones:** consiste en, como su nombre indica, confrontar dos o más variables, esto es, analizarlas con detenimiento con el fin de encontrar sus semejanzas o diferencias.
- **Conexiones:** con este concepto se hace referencia a la posibilidad de encontrar relaciones establecidas entre dos o más variables.
- **Geolocalizaciones:** consiste en la localización o ubicación geográfica de la información atendiendo a unos datos de latitud y longitud.
- **Procesos:** conjunto de acciones relacionados entre sí cuya interacción modifica una serie de elementos de entrada para conseguir unos resultados o elementos de salida determinados.

El gran avance que ha experimentado la tecnología en los últimos años ha hecho posible recoger una cantidad masiva de datos procedentes de las diversas actividades que realizamos: *e-commerce*, las webs que visitamos, las compras que hacemos en tiendas físicas, las transacciones de nuestras tarjetas de crédito...Estamos

monitorizados en todo momento.

Toda esta cantidad de datos originados, que son accesibles de manera instantánea, supone también un nuevo reto para las empresas, quienes se ven cada vez más presionadas para mejorar la calidad de sus productos y servicios con el objetivo de seguir siendo un referente en un mercado muy competitivo. A dichas empresas se les abre ahora un nuevo paradigma en el que tienen que ser capaces de analizar una cantidad masiva de datos para encontrar en ellos patrones o elementos que les puedan servir de utilidad a la hora de satisfacer las necesidades de sus clientes. El procesamiento y análisis de estos grandes conjuntos de datos se recogen en el nuevo concepto emergente denominado *Minería de Datos*. Dicha disciplina se encargará de procesar abundantes cantidades de datos usando para ello técnicas y algoritmos que darán como resultado patrones complejos, que anteriormente eran desconocidos.

Una vez hecha esta breve introducción, en las secciones presentes en este capítulo se hará hincapié en los diferentes tipos de datos que existen y también las diversas fuentes de las que pueden proceder.

2.1. Tipos de datos

Dentro de un entorno científico, los datos se pueden definir como el resultado de la generación, recolección y observación de elementos, eventos o procesos con el objetivo de poder realizar un análisis posterior. Desde el punto de vista de dicho análisis, se pueden dividir los datos en dos grandes clases [1] [2]:

- **Datos Cualitativos:** se engloban dentro de este grupo todos aquellos datos que hacen referencia a cualidades, de manera que se incluyen aquí los datos que indican color, sentimientos, experiencias, etc. En definitiva, todas aquellas propiedades que no pueden expresarse de una manera numérica. A menudo, a los datos cualitativos también se les denomina datos categóricos, definidos como aquellos que pueden ser asignados a una categoría. Así, en el sector automovilístico, podemos decir que un vehículo puede ser “nuevo”, “de ocasión” o “de segunda mano”. Todos estos adjetivos se corresponderían con categorías aplicables a cada coche.
- **Datos Cuantitativos:** son aquellos datos que responden a cantidades numéricas. Por ejemplo, la altura de una persona, su peso, el precio de un determinado artículo, las notas de un examen, etc.

Dentro de la clase de datos cualitativos se pueden encontrar dos subclases diferentes:

- **Datos Nominales:** son datos “etiquetados” que pueden dividirse en varios grupos. Por lo general, consiste en crear nombres para la información en función de algunas de sus características. En este caso, los datos nominales no pueden ser cuantificados, es decir, no se pueden medir ni evaluar, sino que se asignan a varios grupos. Estos datos tampoco siguen un orden definido y son típicamente alfabéticos. Por ejemplo, si en una encuesta encontramos una pregunta referente al sexo del participante, las respuestas podrán ser “Hombre”, “Mujer”, etc. Todas las opciones designan a diferentes grupos y no siguen ningún orden.
- **Datos Ordinales:** dentro de este grupo los valores se pueden contar y establecen una jerarquía u orden, en base a una escala, que resulta crítico. Se puede establecer una relación clara entre cada uno de los datos y los números naturales. Un ejemplo claro de este tipo de datos son los resultados de una carrera de atletismo: primero, segundo, tercero, etc.

Dentro de los datos cuantitativos se pueden encontrar también dos subclases:

- **Datos Discretos:** suelen ser números que presentan una cierta característica y es que, entre un valor posible y el siguiente, no hay ninguno intermedio. Por lo general son números enteros que solo se pueden expresar con una cifra. El número de hermanos en una familia o el número de alumnos de una clase se consideran ejemplos pertenecientes a este tipo de datos.
- **Datos Continuos:** al contrario que los elementos discretos, en esta categoría los datos pueden tomar cualquier valor, sin tener por qué ser éste un valor entero. Esto es, los datos pueden tomar cualquier

valor dentro de un rango o escala que puede ser dividida infinitas veces. El tiempo que se tarda en completar una carrera o la temperatura que hace en una determinada habitación podrían incluirse en este grupo.

De manera esquemática, la clasificación realizada de los datos se muestra en la Ilustración 2.2.

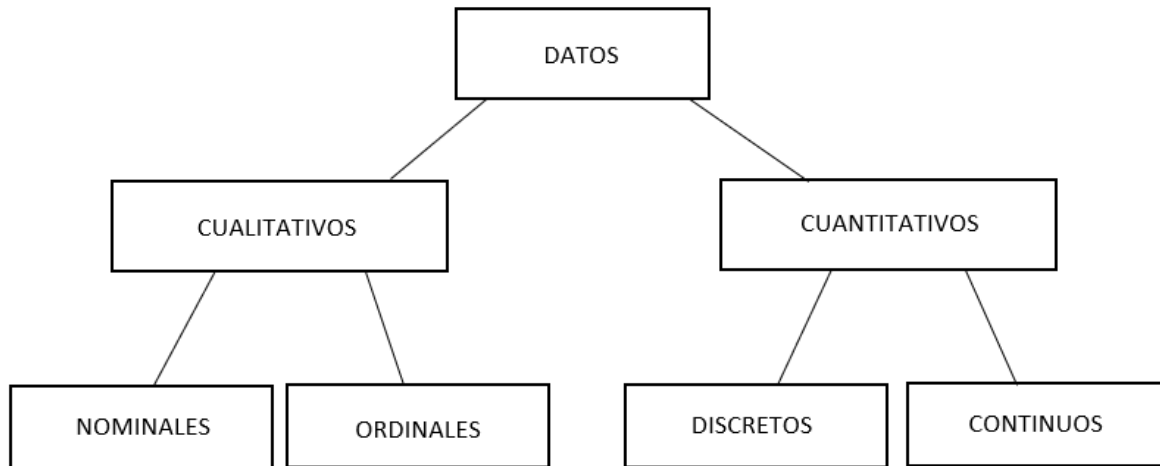


Ilustración 2.2. Tipos de datos

Atendiendo ahora a la configuración presente en los datos, se pueden diferenciar los siguientes grupos de datos [4]:

- **Datos estructurados:** engloba todas aquellas informaciones que forman parte de bases de datos. Suelen consistir en archivos, normalmente de texto, ordenados en filas y columnas. Este tipo de datos son fácilmente procesados por cualquiera de las herramientas de la *Minería de datos*, así como por cualquier ordenador.
- **Datos no estructurados:** estos datos no presentan una estructura interna claramente identificable. Es un conjunto masivo, y sin ninguna organización, de objetos que no aportan valor hasta que son ordenados en base a cierto criterio. Ejemplos de datos no estructurados pueden ser las opiniones recogidas en un determinado sitio web de una empresa, archivos PDF, imágenes descargadas de un repositorio o incluso la información que manejamos diariamente en nuestros *smartphones*: registro de llamadas, correos electrónicos, búsquedas realizadas en un navegador web, etc.
- **Datos semiestructurados:** estos datos son una mezcla de los dos anteriores. No presentan una estructura completamente definida en filas y columnas, pero sí tienen elementos que facilitan la identificación y separación de campos y registros. Estos elementos pueden ser, por ejemplo, comas o puntos. Muestran o describen objetos y sus relaciones en un formato que ha sido aceptado por convención. En este grupo se pueden incluir como ejemplo los formatos de HTML o JSON.

En función del tipo de datos con el que se trabaje, para su análisis se necesitarán unas determinadas herramientas que permitan extraer de los datos las conclusiones necesarias que sirvan como base para, por ejemplo, una posterior toma de decisiones.

2.2. Fuentes de datos

Una fuente de datos se puede definir como todo aquel instrumento y recurso que tiene la capacidad de satisfacer las necesidades de información o conocimiento de cualquier persona. Las fuentes pueden ser, a su vez, de muchos tipos: impresas o electrónicas, como libros, periódicos, revistas, bases de datos, etc.

Una primera clasificación que se puede realizar de las fuentes es atendiendo al origen de la información. Desde

este punto de vista, se pueden distinguir 3 clases de fuentes:

- **Fuentes de información personales:** se clasifican en esta clase todas aquellas fuentes que pueden proceder de una colectividad o de un individuo, es decir, de una o más personas relacionadas que proporcionan información, que puede resultar de interés, originalmente de forma oral. Este tipo de fuentes se caracteriza por su inmediatez y por su actualidad, aunque también se debe destacar su informalidad, puesto que la información que ofrece no aparece estructurada como sí lo haría, por ejemplo, en un documento.
- **Fuentes de información institucionales:** son todas aquellas fuentes que proporcionan información sobre una institución. La información mostrada suele hacer referencia a aspectos propios de la institución tales como su organización, su funcionamiento o los servicios que ofrece. El acceso a la información se puede producir de forma oral o a través de, por ejemplo, directorios, memorias o informes creados por la propia institución. Cabe destacar que el acceso a este tipo de información puede resultar arduo, ya que las instituciones pueden ser de carácter público o de carácter privado y tener sus propias normas a la hora de permitir el acceso a la información.
- **Fuentes de información documentales:** estas fuentes son las que proporcionan información a partir de un documento. Se caracterizan por ser fuentes de consulta rápida e inmediata, que satisfacen necesidades informativas directa o indirectamente. Esto es, los documentos pueden contener los datos requeridos o bien pueden remitir a otras fuentes que contienen la información necesitada. Dentro de este tipo de fuentes se encuentran, a su vez, 3 subgrupos:
 - **Fuentes primarias:** constituidas por publicaciones, documentos o trabajos que realizan, de primera mano, las mismas personas o instituciones que han sido los encargados, también, de recolectar los datos.
 - **Fuentes secundarias:** la información mostrada en las fuentes secundarias procede del tratamiento de los datos proporcionados por una fuente primaria. Este tratamiento puede consistir en el análisis, síntesis o interpretación de los datos, recogiendo el resultado en publicaciones o trabajos que han sido elaborados por terceras personas o entidades, que no han tenido una colaboración directa en la obtención de los datos.
 - **Fuentes terciarias:** se incluyen en este grupo una recopilación de los dos tipos anteriores de fuentes, incluyendo, por ejemplo, bibliografías, artículos o catálogos.

Los datos, además, pueden ser generados en diversos ámbitos o áreas: la industria, la ciencia, la actividad empresarial e, incluso, la sociedad, genera grandes cantidades de información. Una vez que esos datos son recogidos y almacenados, puede surgir la necesidad de querer hacer uso de ellos. En ese caso, será necesario recurrir a una fuente que los proporcione. A grandes rasgos, las fuentes más comunes para acceder a los datos son dos [2]:

- **Repositorios digitales:** también se los conoce con el nombre de biblioteca de datos. Sistemas de información encargados de ordenar y organizar los datos de cara a fines tales como la investigación, el desarrollo, etc., al mismo tiempo que garantizan el fácil acceso a dichos datos. La clasificación más común de los repositorios es en públicos y privados. Los repositorios públicos son aquellos que instituciones públicas ponen a nuestro alcance en la red. Los datos aquí contenidos se denominan datos abiertos (*open data*). Esta práctica surge con el objetivo de que ciertos tipos de datos estén siempre disponibles para toda la sociedad, obviando un determinado número de restricciones tales como derechos de autor o patentes. Son datos que pueden ser utilizados y compartidos por cualquier persona, quien debe, ineludiblemente, respetar siempre la seguridad y privacidad de la información. La definición de datos abiertos entraña, por tanto, una serie de aspectos elementales: la información debe ser de acceso libre y, en el caso, a un costo razonable, y, además, dicha información debe poder ser modificada, reutilizada, distribuida y de fácil integración con otros conjuntos de datos. Los repositorios privados, en cambio, serán aquellos que un organismo o persona posea en propiedad y en los que tiene la libertad de establecer el modo de acceso y compartición de los datos contenidos.

3 VISUALIZACIÓN DE DATOS

Un buen dibujo es siempre mejor que un discurso largo.

Napoleón Bonaparte

Una vez que se dispone de los datos, el paso más importante es el que tiene que ver con la visualización de estos, ya que dicha visualización será la que permita analizar y sacar las conclusiones oportunas en cada caso. De ahí que se acuda constantemente al famoso refrán “una imagen vale más que mil palabras”.

El objetivo principal de esta sección será la de conocer los beneficios de la presentación visual de los datos, así como los diferentes tipos de representaciones que se pueden realizar y algunas de las múltiples herramientas a nuestra disposición.

Se podría definir la visualización de datos como aquella variación de la información, convirtiéndola en imágenes, que ayuda a obtener y sacar un significado concreto. Básicamente consiste en la representación de los datos usando una aproximación visual en lugar de los tradicionales métodos de análisis. Como se ha visto en capítulos anteriores, estos datos pueden ser de diferentes tipos e, incluso, muchas veces pueden ser una mezcla de varios tipos.

Inherente a la definición mostrada existe, además, otra idea importante. Y es que la visualización de datos no sólo se emplea como un método nuevo de análisis, sino que también es utilizada como una herramienta de comunicación. Y, a su vez, este concepto de comunicación visual no es nuevo, la sociedad ha utilizado las ilustraciones para manifestarse desde tiempos remotos: pictogramas en los interiores de las cuevas o los jeroglíficos egipcios, son algunos de los ejemplos que podemos encontrar [6].

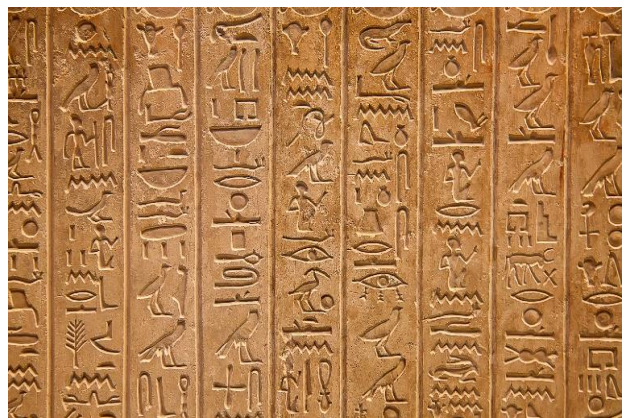


Ilustración 3.1. La Visualización como herramienta de comunicación

Hoy en día, estamos siendo testigos de un aumento significativo del uso de la visualización de datos. Vivimos rodeados de gráficos, mapas, iconos, etc., herramientas que contribuyen de una manera ampliamente beneficiosa de varias formas [1]:

- Evitan que haya un exceso de información, esto es, de la gran cantidad de datos que podemos tener al alcance, estas herramientas visuales nos permiten filtrar aquella información más relevante e importante, apartando, a su vez, aquella que no aporta valor.

- Facilitan la comprensión. La visualización hace que grandes cantidades de datos complejos puedan ser entendidos y asimilados por la sociedad, pues ésta tiende a interpretar las imágenes de una manera más rápida y eficaz que los textos. Además, el hecho de que, a partir de los mismos datos, se puedan realizar multitud de representaciones facilita el procesamiento y análisis de estos, permitiendo una visión más amplia y profunda de la naturaleza de cualquier problema y el planteamiento de su solución.
- Contribuye a la distribución del contenido. Un contenido que sea fácil de entender, visualmente atractivo y que aporte nuevos conocimientos tiene mayores posibilidades de ser compartido de manera masiva. Este hecho puede ayudar, por ejemplo, a crear un punto de vista común, entre un grupo de personas, sobre una situación o escenario concreto, alineando a los afectados para llevar a cabo las acciones necesarias al respecto.

3.1 Tipos de representaciones de datos

Aunque dispongamos de una cantidad masiva de datos de una manera prácticamente instantánea, la mayoría de dichos datos pueden resultar incoherentes o simplemente inconexos. Por tanto, el primer paso antes de proceder a la representación de la información debe ser entender el problema a solucionar y saber qué queremos transmitir, es decir, cuál debe ser el objetivo último de nuestra representación. Sólo después de haber realizado este paso, contaremos con la capacidad de elegir el diseño más conveniente, y de más fácil entendimiento, para los datos que nos interesen en cada momento.

La forma gráfica, por tanto, vendrá determinada por el tipo de dato con el que se trabaja y, también, por el análisis que hayamos hecho de los mismos. Esta gráfica deberá presentar un equilibrio entre dos propiedades importantes: funcionalidad y estética [3]. La primera de ellas hace referencia a la necesidad de que la representación sea un fiel reflejo de los datos, de manera que permita la extracción de conclusiones efectivas. La componente estética aparece como un valor añadido y es que aquellas representaciones que presenten colores vistosos, espacios coherentes o una estructura simétrica, serán más fáciles de recordar y entender para cualquier persona que lo vea.

De esta manera, cuando sea necesario visualizar algún tipo de información siempre será oportuno recurrir a elementos tales como [2]:

- **Tablas:** una tabla no es más que la presentación, de manera organizada en filas y columnas, de una serie de datos cuantitativos englobados dentro de un conjunto de etiquetas. Es principalmente útil en situaciones en las que se requiere buscar u obtener un valor específico.
- **Gráficos:** un gráfico se va a emplear como método para mostrar las relaciones existentes entre los datos utilizados y otros valores que van a estar codificados como objetos visuales. Algunos de estos objetos visuales serán líneas, puntos, barras... Las representaciones gráficas estarán englobadas dentro de un área, delimitada por lo que se conoce como ejes. Dichos ejes proporcionan las escalas que se emplean en la asignación de valores a los objetos visuales.

Los gráficos constituyen un poderoso medio para concentrar la información y presentarla de una manera que la mayoría de la sociedad considere fácil de entender. Permiten, por tanto, extraer, a partir de una gran cantidad de datos, las características y conclusiones principales de estos. Además, no sólo son una herramienta para presentar los resultados numéricos de un estudio, sino que también proporcionan, por ejemplo, los patrones existentes en los datos, que serán de una importancia crítica para el análisis o la toma de decisiones.

Aunque existen multitud de gráficos diferentes, la mejor opción será, teniendo en cuenta la información que se quiere analizar, elegir siempre el gráfico que, siendo el más sencillo posible, represente mejor los datos a tratar. En las siguientes subsecciones se explicarán algunos de los gráficos más usados y el aspecto visual que presentan.

3.1.1. Diagrama de Barras (*Bar Chart*)

Este diagrama es uno de los más empleados a la hora de mostrar datos categóricos o cualitativos y tiene una

especial importancia en la realización de comparaciones. Se trata de la representación de un conjunto de datos a través de una serie de barras rectangulares, que pueden tener una orientación horizontal o vertical, y una longitud que será proporcional a los valores de los datos mostrados [2].

Un ejemplo del uso de este tipo de gráfico es el que aparece en la Ilustración 3.2, en la que se muestran las ventas mensuales de una compañía.

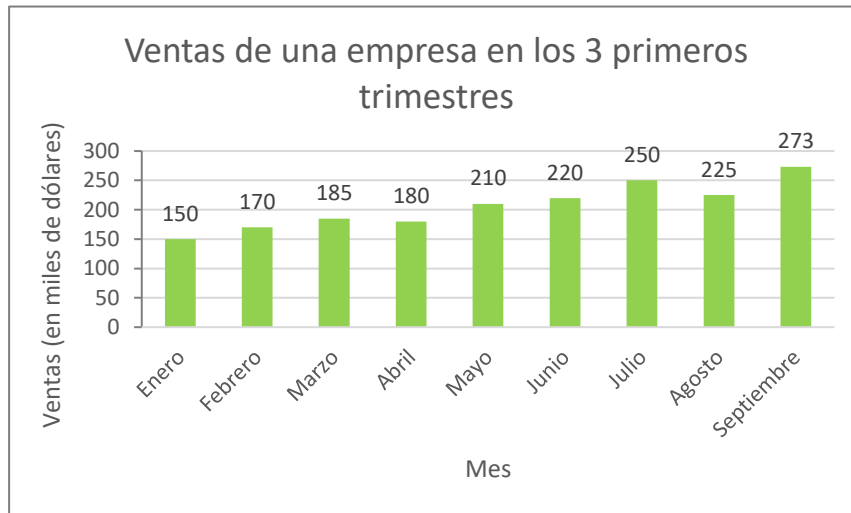


Ilustración 3.2. Ejemplo de Diagrama de Barras, con orientación vertical [2]

En este tipo de representaciones, hay otras consideraciones a tener en cuenta: las barras deben estar espaciadas una distancia tal que sea igual siempre, y el ancho de la barra debe ser idéntico para todas las barras. Lo único que va a ir cambiando es, como se ha comentado antes, la longitud de cada barra, que dependerá de los datos incluidos.

El diagrama de barras, debido a su amplio uso, ha derivado en otras representaciones similares. Así, otro tipo de diagrama es el que muestra la información agrupada (*Cluster Bar Chart*), con el objetivo de comparar varias categorías de datos. Siguiendo con el ejemplo anterior, este nuevo gráfico representaría, por ejemplo, las ventas cuatrimestrales de una compañía en tres zonas diferentes que podrían ser las diferentes sedes en las que opera. Este ejemplo se muestra en la Ilustración 3.3. A partir de esta representación, desde el punto de vista empresarial se puede comprobar, por ejemplo, qué zonas resultan más productivas y cuáles menos, o en qué épocas del año podría interesar lanzar campañas para captar clientes.

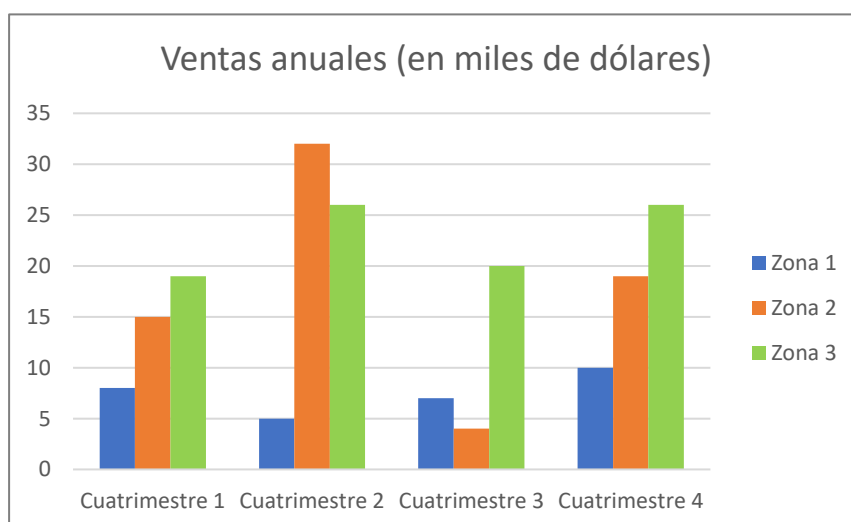


Ilustración 3.3. Ejemplo de Diagrama de Barras con información agrupada [2]

Una última gráfica, derivada de este tipo, es la que se conoce como Diagrama de Barras Apiladas (*Stacked Bar Chart*). En este caso, se muestra cómo una categoría amplia se puede dividir en otras más pequeñas y cuál es la

relación o el peso de cada una de las partes pequeñas dentro de la categoría global [2]. Para elaborar el gráfico de la Ilustración 3.4 se toma como ejemplo las ventas producidas por una determinada tienda física en una semana. Se muestra, por un lado, las edades entre las que se distribuyen los compradores y, por otro lado, su sexo, distinguiendo entre hombres y mujeres. Igual que sucedía en los casos anteriores, es posible extraer conclusiones del gráfico como, por ejemplo, se puede comprobar que la mayoría de los clientes son mujeres de entre 20 y 30 años.

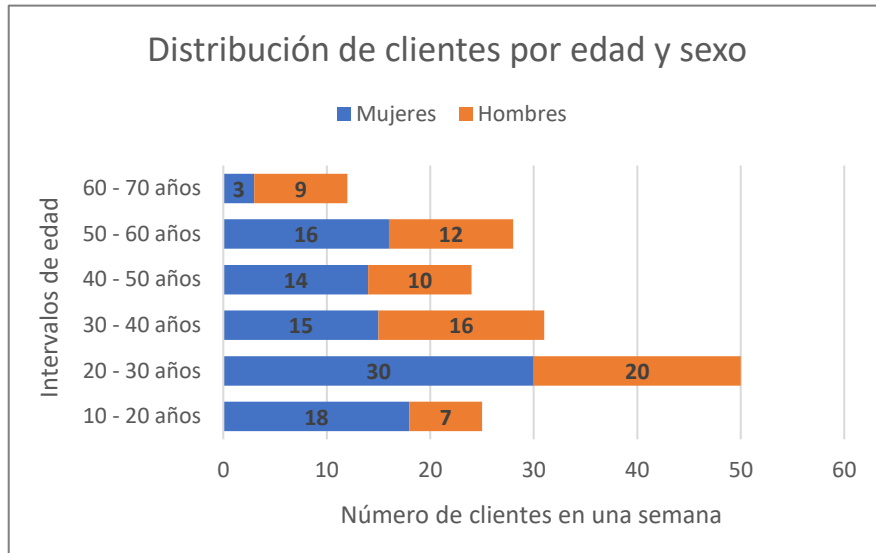


Ilustración 3.4. Ejemplo de Diagrama de Barras Apiladas, con orientación horizontal

3.1.2. Gráfico de Líneas (Line Chart)

En este tipo de representaciones, los valores de cada uno de los datos van a estar contenidos en dos ejes cartesianos perpendiculares entre sí, y van a estar conectados mediante una línea. Los gráficos de líneas son especialmente útiles para la visualización de la tendencia que siguen los datos y son fáciles de entender y, también, de crear [4]. Como ejemplo se muestra la Ilustración 3.5, en la que se observa el mismo ejemplo que para el diagrama de barras, las ventas de una empresa a lo largo de los tres primeros trimestres del año.

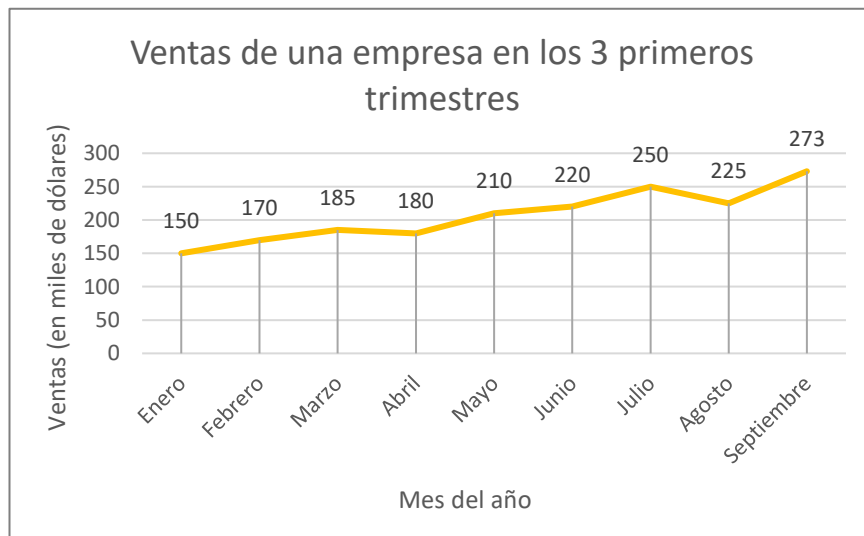


Ilustración 3.5. Ejemplo de un Gráfico de Líneas [2]

Para este ejemplo se utilizan los mismos datos que para la figura mostrada en la Ilustración 3.2. La diferencia fundamental que se consigue empleando un gráfico de líneas es que es posible, además de observar los valores exactos de las ventas producidas, percibir la tendencia que sigue la empresa, que, en este caso, es positiva y

ascendente, evidenciando que las ventas van aumentando paulatinamente con el paso de los meses, aunque existan algunas excepciones.

3.1.3. Diagrama Circular (*Pie Chart*)

La utilización de este tipo de representaciones ha generado un gran debate en los últimos tiempos, al considerarse menos adecuado a la hora de comprender la información comparado, por ejemplo, con los dos gráficos anteriores. El motivo es que algunos expertos en diseño consideran que el ojo humano tiene peores capacidades para estimar áreas a simple vista [4]. Aun así, sigue siendo uno de los diagramas más usados actualmente.

Un gráfico circular no es más que un círculo dividido en partes o secciones. Cada una de las secciones representa un grupo de datos y la anchura de cada sección muestra el valor concerniente de los datos dentro de un conjunto. En otras palabras, este tipo de gráfico ofrece la capacidad de apreciar el peso relativo de cada parte respecto del todo [2]. De esta manera, con un simple vistazo, el que analiza la imagen es capaz de ver cuáles de las secciones tienen un mayor o menor peso en el conjunto global. Para calcular la anchura que debe tener cada sección, se emplea la siguiente fórmula:

$$\text{Amplitud de la sección (}^\circ\text{)} = 360^\circ \times f_i$$

La ecuación consiste en el producto del valor 360° , por ser los grados que hay en un círculo, y el valor f_i , que hace referencia a la frecuencia relativa de cada sección de datos, es decir, el número de veces que aparece ese grupo de datos en el conjunto total. El resultado de la ecuación es el número de grados que debe ocupar esa sección en el círculo global.

Estos gráficos ofrecen enormes posibilidades de representación, pero requieren tener en cuenta dos aspectos importantes [4]:

- El número de secciones a representar no debe ser excesivo, puesto que, si se incluyen demasiadas secciones y algunas de ellas son de tamaño muy reducido, lo único que hará será distraer al observador, dificultándole el análisis de los aspectos más importantes. Por ejemplo, si se quiere analizar el porcentaje de turistas recibidos por cada comunidad autónoma en un año, se podría obtener el gráfico de la Ilustración 3.6.

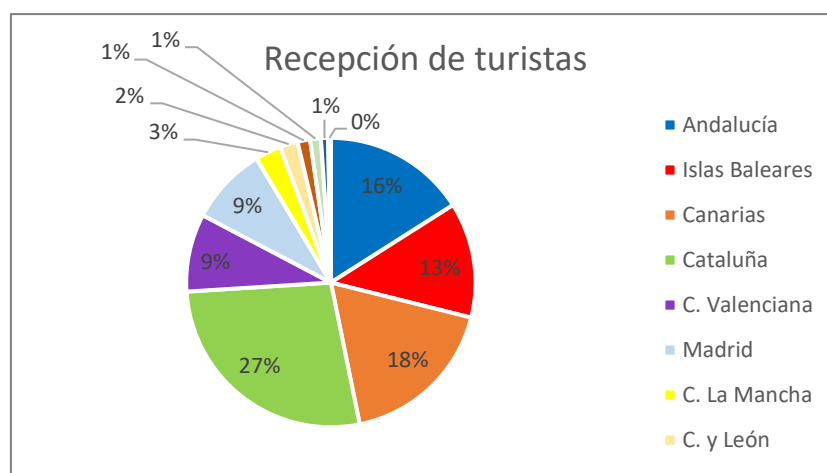


Ilustración 3.6. Ejemplo de Diagrama Circular

En estos casos resulta preferible agrupar todas aquellas secciones con poco peso en una sola, de manera que destaquen más aquellas secciones con un mayor peso en el conjunto.

- Cuando existen valores de amplitud muy similares para varias secciones, puede resultar difícil comparar la contribución de esas secciones individualmente, de ahí que en estos casos resulta más conveniente utilizar, por ejemplo, un diagrama de barras. Este hecho se hace evidente en las figuras de la Ilustración 3.7. Se puede observar cómo, con un diagrama de barras, se puede apreciar mejor la distribución del peso relativo de las diferentes secciones.

Por tanto, a la hora de crear un gráfico que muestre un conjunto de datos, será necesario analizar qué tipo de representación es la más adecuada para esos datos en concreto.

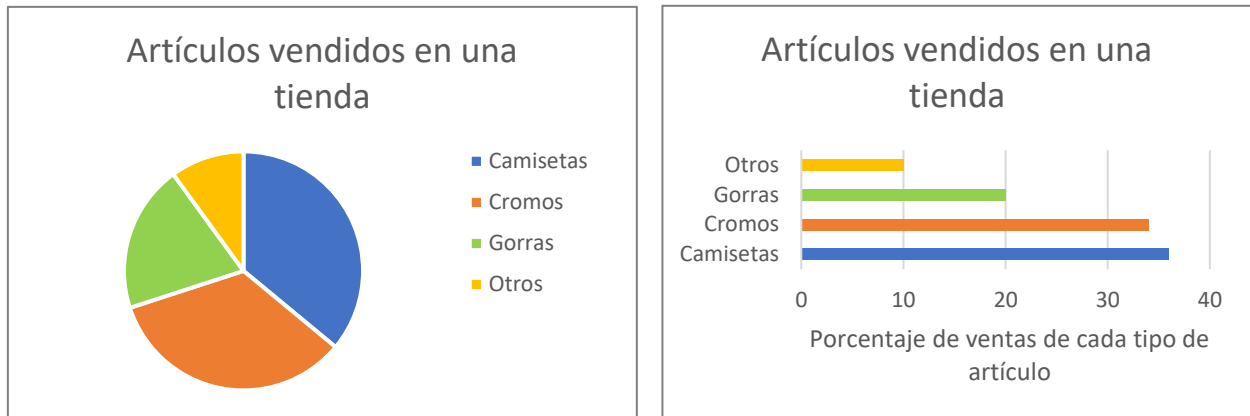


Ilustración 3.7. Comparación de un Diagrama de Barras y un Diagrama Circular

Sin embargo, dos tipos diferentes de gráficas también pueden complementarse. Una variante de este tipo es lo que se conoce como *Pie of Pie*, cuyo propósito es el de mostrar, mediante un diagrama circular, las diferentes componentes presentes dentro de una de las secciones de otro diagrama circular. Por ejemplo, en la Ilustración 3.8 se puede observar el consumo de energía de un país y, en una de las secciones, cómo se distribuyen sus componentes. En esta línea, también se encuentra el *Bar of Pie*, en el que lo único que varía es el gráfico empleado para el análisis de la sección concreta, que, en este caso, es un diagrama de barras [2].

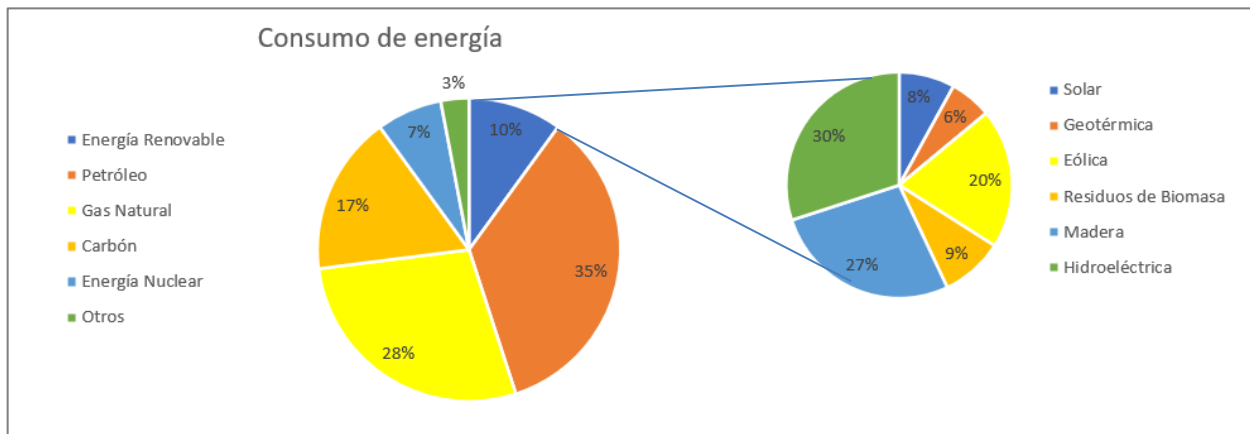


Ilustración 3.8. Ejemplo de *Pie of Pie*

Como se puede observar, con el diagrama circular de la derecha se presentan los diferentes elementos que forman parte de la sección denominada “Energía Renovable”, y los pesos relativos de cada elemento en el conjunto total.

3.1.4. Diagrama de Puntos (*Dot Plots*)

Este tipo de representaciones se suelen utilizar para analizar la forma de distribuirse, o para comparar, dos o más conjuntos de datos. En un diagrama de puntos va a existir un solo eje, el eje horizontal, que representa el rango de valores entre el que están comprendidos los datos. Cada ocurrencia de un valor dentro de ese rango se va a representar como un punto sobre el eje horizontal en su lugar correspondiente. De esta manera, si el valor se repite, los puntos se irán apilando unos encima de otros, mostrando un punto por cada vez que se repita el valor. Es aconsejable reservar representaciones de este tipo para el caso de conjuntos de datos reducidos o relativamente pequeños. A modo de ejemplo se presenta la Ilustración 3.9 [2].

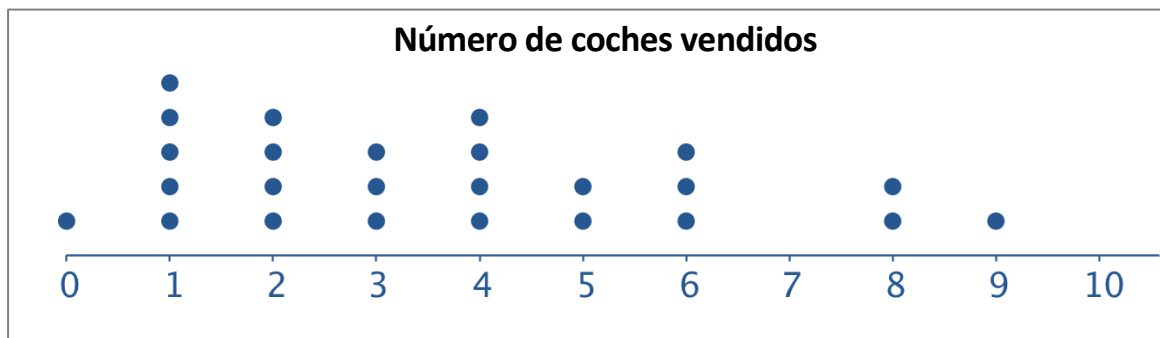


Ilustración 3.9. Ejemplo de Diagrama de Puntos

En la gráfica anterior se muestran el número de coches vendidos por una empresa en un período de 25 días. El eje horizontal representa, por tanto, un rango de valores que se corresponde con el número de ventas que se han producido en los diferentes días. Cada punto en la gráfica representa un día, de manera que se puede leer la información de la siguiente manera: hay 1 día en el que la cantidad de coches vendidos fue 0 (porque se aprecia 1 punto sobre el valor 0), 5 días en los que se vendió 1 vehículo, 4 días en los que vendieron 2, etc. El número total de puntos representados es 25, que se corresponden con los 25 días que se quieren analizar.

3.1.5. Gráfico de Intervalos (*Interval Plots*)

Los gráficos de intervalos se convierten en representaciones de gran interés a la hora de medir tendencias o la variabilidad de los datos. Principalmente, estos diagramas incluyen medias y/o intervalos de confianza de una o más variables, mostrando así los diferentes valores que puede tomar una variable.

Para mostrar esta definición con mayor claridad, se va a recurrir a un ejemplo [2]: una empresa que dispone 5 líneas diferentes de producción de un determinado tipo de bebidas en latas decide hacer un estudio para analizar si la media de líquido presente en cada lata varía de una línea de producción a otra. Para lograr esto, se eligen aleatoriamente 5 latas de cada una de las líneas de producción, y se mide el contenido de las latas. El resultado del análisis se recoge en la Ilustración 3.10.

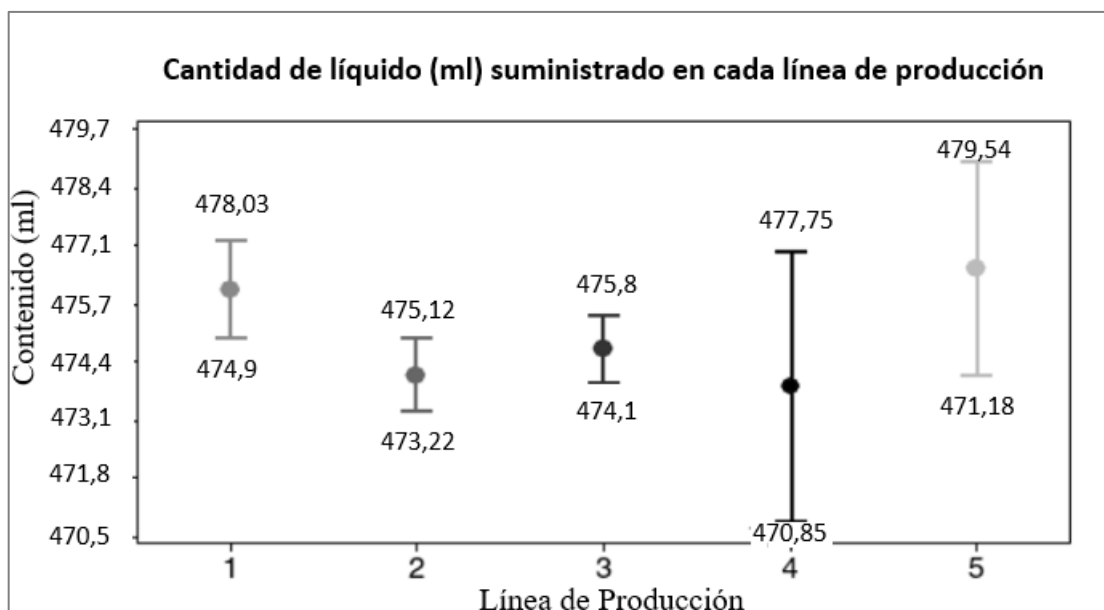


Ilustración 3.10. Ejemplo de Gráfico de Intervalos [2]

Como se puede observar, en el eje X se incluyen cada una de las diferentes líneas de producción. Se puede ver en cada línea, para las 5 latas estudiadas, la menor cantidad de líquido presente y también la cantidad mayor, estimándose con esos valores una media de líquido contenido en las latas para cada línea de producción,

representado por la señal colocada en el punto medio de los intervalos.

Otro tipo de representaciones incluidas en este apartado son las Series Temporales. En estas gráficas lo que se hace es representar los datos sobre el tiempo, variable representada en el eje X. Se presentan las parejas de puntos (X,Y) y se unen mediante líneas rectas con el objetivo de facilitar la detección de patrones o tendencias en los conjuntos de datos. Véase la Ilustración 3.11.

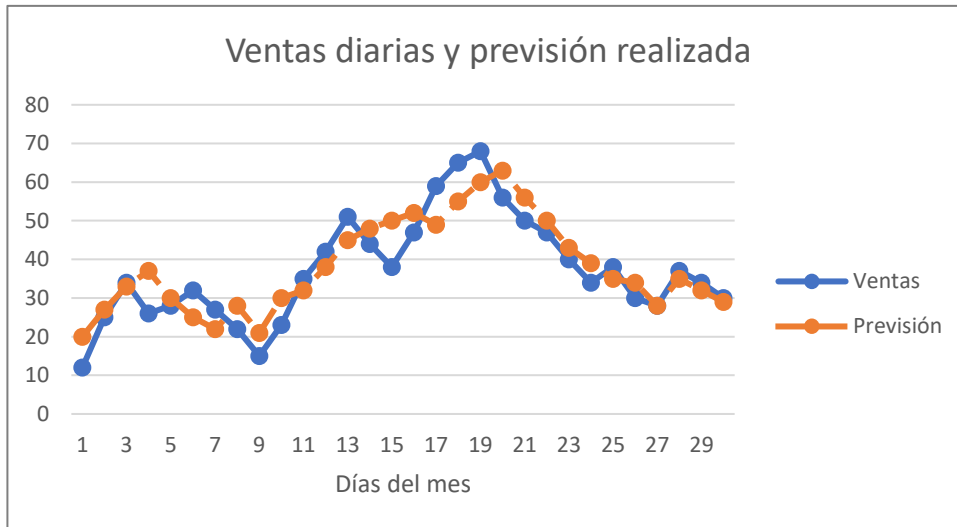


Ilustración 3.11. Ejemplo de Diagrama de Serie Temporal

En esta ilustración se recogen las ventas de una empresa durante un mes, considerado como 30 días, así como la previsión de ventas esperadas en dicho tiempo. Con este tipo de representaciones, la empresa será capaz de evaluar su actividad, así como tomar decisiones relacionadas con, por ejemplo, el incremento o reducción de la producción, el lanzamiento de campañas que fomenten las ventas en otros sectores de la población, etc. Sin una gráfica de este tipo sería imposible detectar la tendencia mostrada, aunque ésta esté contenida en los datos.

3.1.6. Diagrama de Dispersión (*Scatterplot*)

Los diagramas de dispersión resultan una herramienta muy útil a la hora de comparar dos variables. Se trata de una representación en 2D, en la que una variable es trazada a lo largo del eje vertical (variable dependiente), y la otra a lo largo del eje horizontal (variable independiente). De esta manera, cada uno de los puntos representados, con coordenadas (X,Y), muestran visualmente la relación existente entre ambas variables. En la Ilustración 3.12 se recoge un ejemplo de representación [2]. En este caso, se va a estudiar la correspondencia existente entre las ventas de una determinada empresa y la cantidad de inversión realizada en publicidad y marketing.

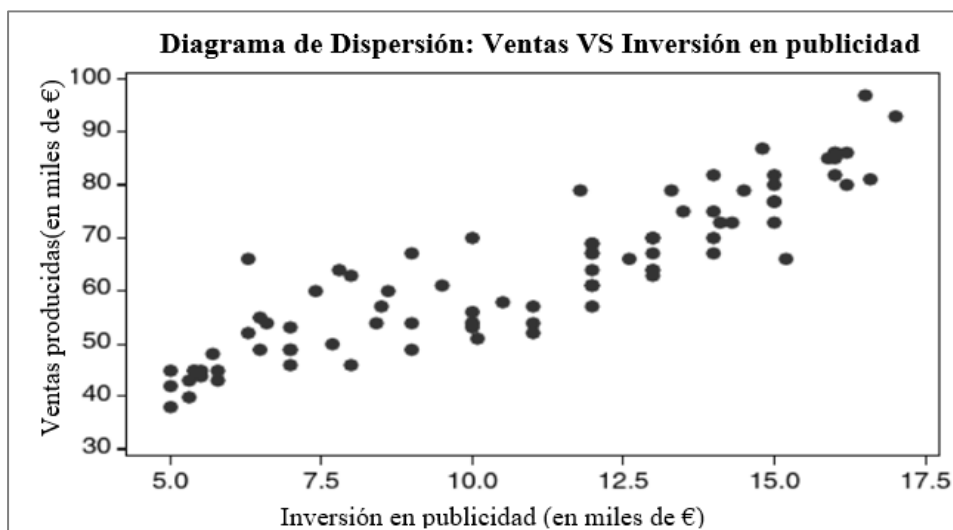


Ilustración 3.12. Ejemplo de Diagrama de Dispersión [2]

Como resultado de la representación gráfica se observa cómo el nivel de ventas aumenta al incrementar la inversión en publicidad. Esto es lo que se conoce como una relación positiva entre las variables, al incrementar una, va a incrementar también la otra.

Esta tendencia no tiene por qué siempre ser así, y como ejemplo se muestra la Ilustración 3.13. En esta gráfica se representa el gasto en calefacción de un hogar, en función de la temperatura exterior.

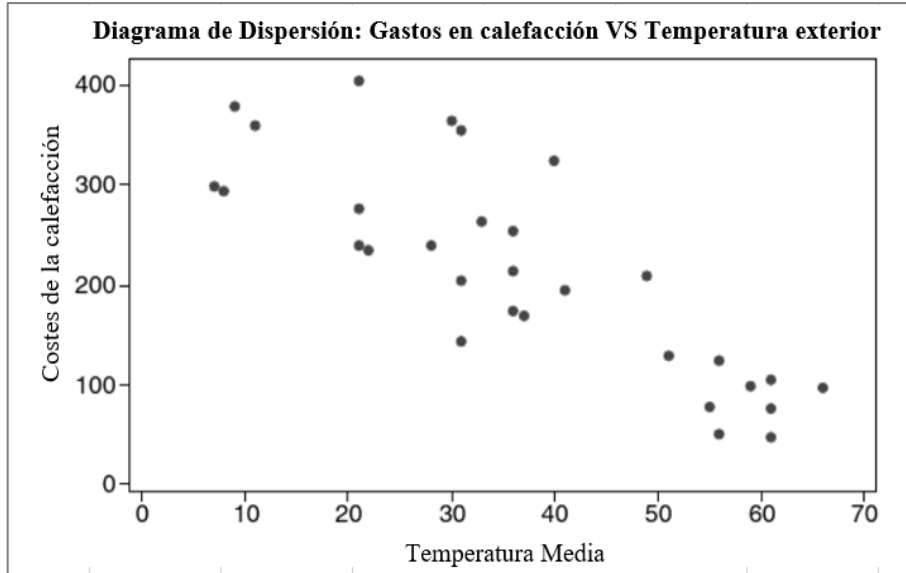


Ilustración 3.13. Ejemplo de Diagrama de Dispersión, con relación negativa de las variables [2]

De manera lógica se ve que, cuanto mayor es la temperatura en el exterior, menor es la necesidad de usar la calefacción y por tanto menor es el coste asociado. Es evidente que, en este caso, la tendencia presenta una pendiente negativa, es el caso de una relación negativa o relación inversa entre dos variables, al aumentar una, la otra va a disminuir. También podría aparecer el caso de que la relación entre las dos variables fuera muy débil o inexistente. En tal caso, se apreciaría un diagrama de dispersión con puntos que no siguen una distribución fija ni ninguna tendencia reconocible.

Los tres casos anteriores muestran la situación en la que la conexión entre las dos variables es lineal, es decir, se puede aproximar por la ecuación de una recta. En la Ilustración 3.14 se muestra una relación no lineal entre las dos variables de estudio. En este caso, se representa el rendimiento o productividad de un determinado proceso químico en función de la temperatura que alcanza.

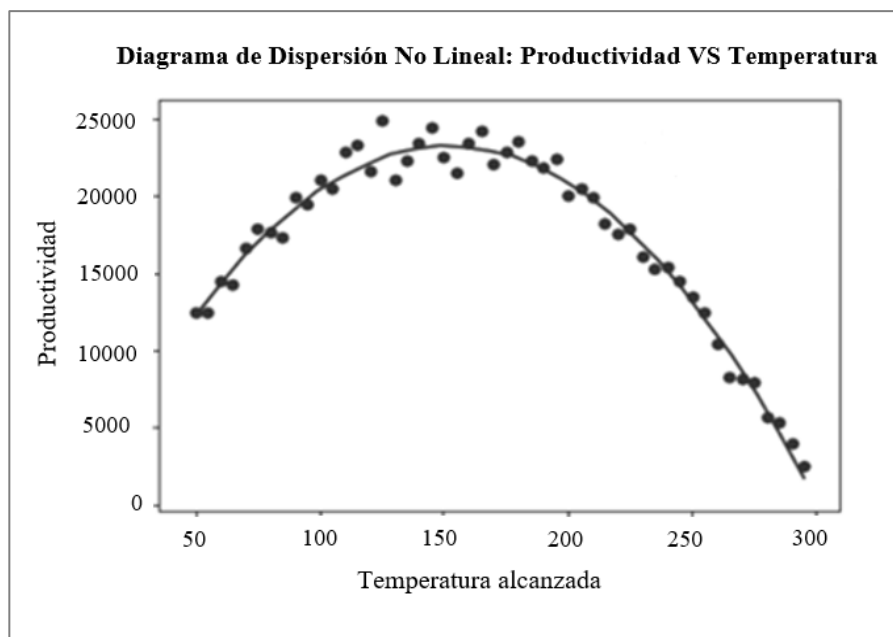


Ilustración 3.14. Ejemplo de Diagrama de Dispersión No Lineal [2]

En este caso se aprecia una relación entre variables que puede ser aproximada mediante una ecuación cuadrática. Este tipo de representaciones se utiliza de manera frecuente para calcular lo que se conoce como coeficiente de correlación [2]. Este coeficiente, cuyo valor oscila entre -1 y 1, mide el nivel de dependencia lineal entre dos variables cuantitativas. Concretamente mide el grado de relación de dos variables cuantitativas y, además, continuas. Teniendo en cuenta las gráficas mostradas anteriormente, si el diagrama de dispersión muestra una relación positiva entre las dos variables, esto es, la gráfica presenta una pendiente positiva, el coeficiente de correlación adquirirá un valor positivo, siendo 1 el valor óptimo. Por el contrario, si la relación mostrada es negativa, el coeficiente de correlación tomará un valor negativo, pero mayor que -1. Si el valor del coeficiente de correlación se aproxima o es cero, este hecho indica una relación débil o incluso inexistente entre las variables bajo estudio.

3.1.7. Gráfico de Burbujas (*Bubble Plot*)

Hasta ahora, todas las representaciones anteriores se utilizan para comparar o establecer una relación entre dos variables. En este caso, el gráfico de burbujas permitirá establecer conexiones entre tres variables diferentes, una de ellas se recogerá en el eje X, otra en el eje Y, y se hará uso de “burbujas” cuyo tamaño determinará el valor de la tercera variable empleada [2].

A modo de ejemplo se presenta la Ilustración 3.15. En este caso se van a analizar las visitas que recibe, durante 5 días, una determinada página web de comercio electrónico. El número de visitas se recoge en el eje X, mientras que en el eje Y se presenta el número de ventas que registra la página web en esos días. El tamaño de las burbujas determinan el volumen de ingresos, en términos monetarios, que recibe la empresa propietaria de la página web.

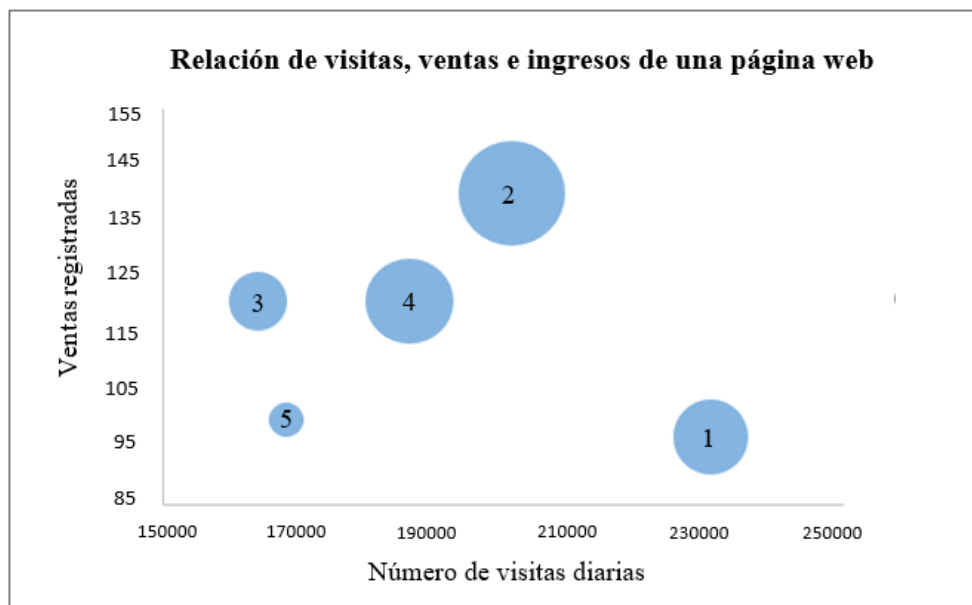


Ilustración 3.15. Ejemplo de Gráfico de Burbujas

Observando la ilustración anterior se pueden obtener algunas conclusiones de interés. Se puede apreciar cómo el día que más visitas recibe la web, burbuja número 1, no se corresponde con el mayor número de ventas ni de ingresos. Igualmente, llama la atención cómo el día que se producen menos ventas, etiquetado como burbuja número 3, no es el día que menos ingresos obtiene la compañía. Este tipo de información obtenida puede resultar de interés a las empresas a la hora de, por ejemplo, comprobar el funcionamiento o rendimiento de determinados servicios web que ofrezcan.

3.1.8. Gráfico de Superficie (*Surface Plot*)

Dentro de todos los tipos de diagramas vistos hasta el momento, el gráfico de superficie es el primero que va a realizar una representación en 3D, y, al igual que el gráfico de burbujas, va a establecer la relación entre tres variables diferentes (X,Y,Z), dos de ellas independientes y una tercera que va a depender de las otras dos.

Para elaborar este tipo de gráficos, se construye una cuadrícula (expresada en inglés por el término *grid*), usando las dos variables independientes, cuyo rango de valores debe coincidir con el de la tercera variable. A continuación, se calcula el valor de la tercera variable en cada punto de la cuadrícula. Este valor será el resultado de un promedio ponderado de todos los valores de datos que están cercanos a ese punto. Se utiliza, así, la interpolación para producir una superficie continua [2].

Un ejemplo de uso de este tipo de representaciones puede ser el análisis de la resistencia a la tensión que experimenta un determinado material en función, también, de la temperatura a la que se le somete. De esta manera se obtiene una figura como la mostrada en la Ilustración 3.16.

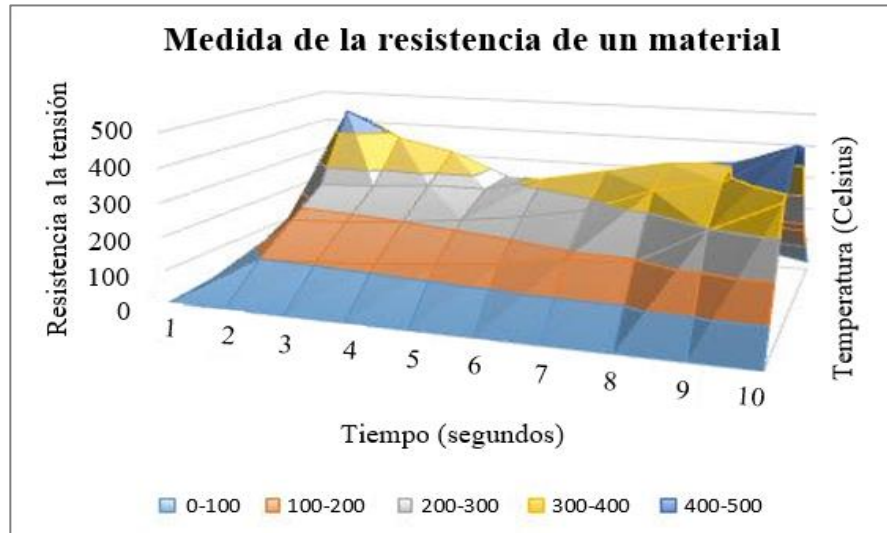


Ilustración 3.16. Ejemplo de Gráfico de Superficie

En base a esta representación, una empresa podrá tener un mayor conocimiento del producto con el que trabaja y le será de ayuda, por ejemplo, en la toma de futuras decisiones sobre lo apropiado (o no) de su uso.

3.1.9. Diagrama de Contorno (*Contour Plot*)

El gráfico de contorno se presenta como alternativa al de superficie para presentar la relación entre tres variables. Consiste en una técnica gráfica para representar una superficie tridimensional en un formato bidimensional. En este tipo de representaciones, dos variables se trazan en los ejes X e Y respectivamente y, con cada valor de la tercera variable, la respuesta Z, se dibujan líneas que conectan las coordenadas (X,Y) donde ese valor de Z ocurre y se repite, generándose regiones sombreadas denominadas contornos [2]. Por ejemplo, si se toma como referencia una compañía que se dedique a la creación de aleaciones, puede resultarle de gran utilidad una gráfica de superficie como la que se muestra en la Ilustración 3.17.

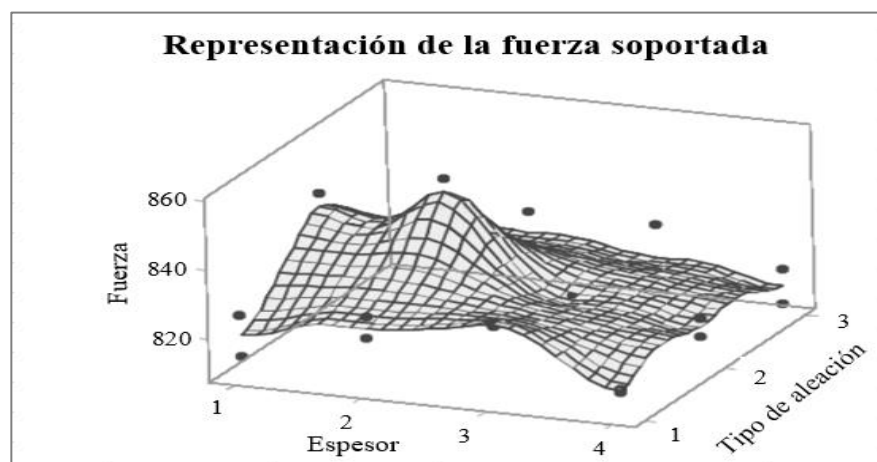


Ilustración 3.17. Diagrama de Superficie de un material [2]

Esta figura muestra, en función del tipo de aleación y de su grosor, la fuerza que va a ser capaz de soportar el material analizado. Esta gráfica resulta, por tanto, muy útil. Sin embargo, puede ocurrir que el encargado de analizar estos datos no disponga a su alcance una herramienta que realice representaciones en 3D. Para estos casos puede tener una gran importancia el diagrama de contorno. Para el mismo ejemplo, la gráfica tendría el aspecto mostrado en la Ilustración 3.18.

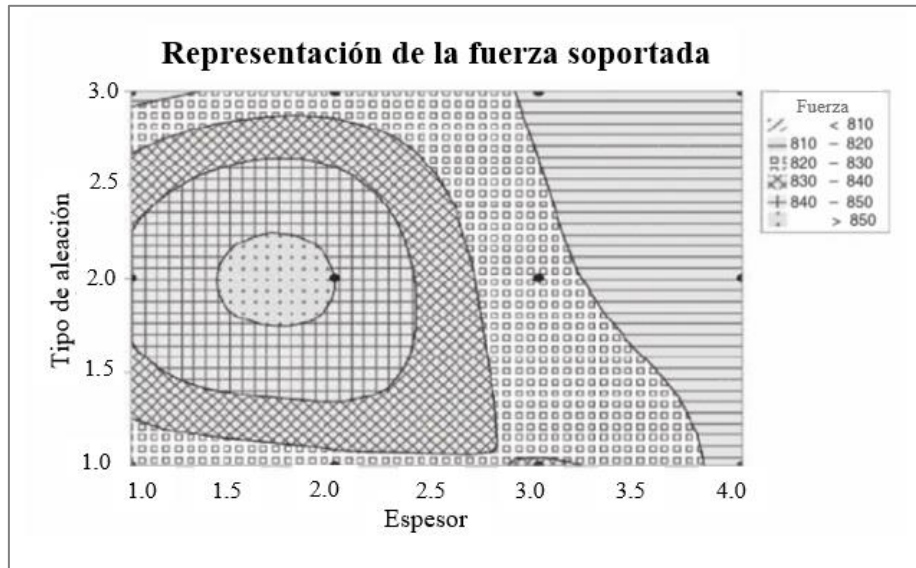


Ilustración 3.18. Ejemplo de Diagrama de Contorno [2]

Al igual que en la Ilustración 3.17, se puede apreciar la relación entre el grosor, el tipo de aleación y la fuerza capaz de soportar la aleación. Este gráfico presenta una manera más sencilla y eficaz de obtener el valor óptimo de la variable de respuesta, en este caso, la fuerza soportada por el material.

3.1.10. Mapas de Procesos (Process Maps)

Un mapa de proceso se define como un diagrama de flujo de un proceso productivo o de un determinado servicio [8]. Es un modelo de operación que muestra los diferentes pasos que suceden en el proceso de transformación de una determinada entrada en una determinada salida, para lo que es necesario disponer de una serie de recursos. Este flujo de actividades se muestra en la Ilustración 3.19.

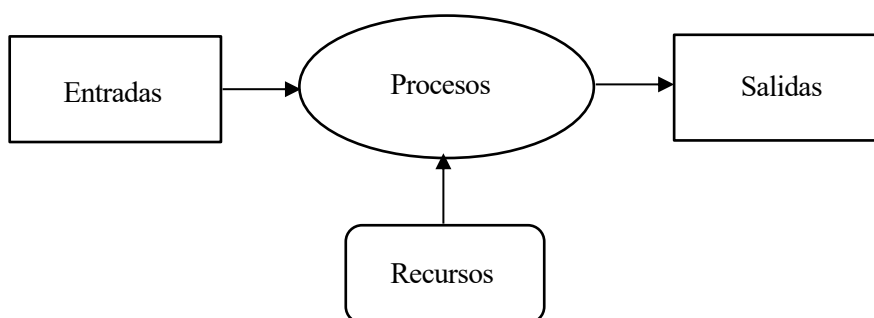


Ilustración 3.19. Mapa de Procesos [8]

Por tanto, en un mapa de procesos siempre estarán presentes una serie de elementos y conceptos [8]:

- **Entradas:** hacen referencia a los objetos, tangibles o intangibles, obtenidos del entorno y que sirven como punto inicial del proceso.
- **Salidas:** son los productos terminados, resultado de la transformación de las entradas.

- **Unidades de flujo:** constituyen los elementos cuyo flujo se analiza. Pueden ser productos, clientes, etc. Por ejemplo, en un servicio de atención al cliente, la unidad de flujo sería el cliente desde que se contacta con él hasta que la comunicación termina.
- **Secuencia de actividades:** se trata del diagrama de flujo del proceso completo. Estas actividades son las encargadas de transformar las entradas en salidas, con ayuda de los recursos.
- **Recursos:** se incluyen dentro de este grupo todos los activos fijos tales como edificios, equipamiento y maquinarias necesarias para el desarrollo del proceso, así como los operarios y trabajadores que llevan a cabo las actividades.
- **Flujo de información:** se incluye en este concepto aquella información necesaria a la hora de tomar decisiones o gestionar una determinada actividad, incluyéndose en un mapa de procesos mediante una línea de puntos.

De esta manera, con los mapas, se hará sencilla la visualización del proceso completo, así como el análisis de posibles riesgos o puntos débiles y la localización de oportunidades de mejora. Un ejemplo de este tipo de gráfico se recoge en la Ilustración 3.20, donde se explica mediante un diagrama de flujo el funcionamiento de un determinado servicio médico.

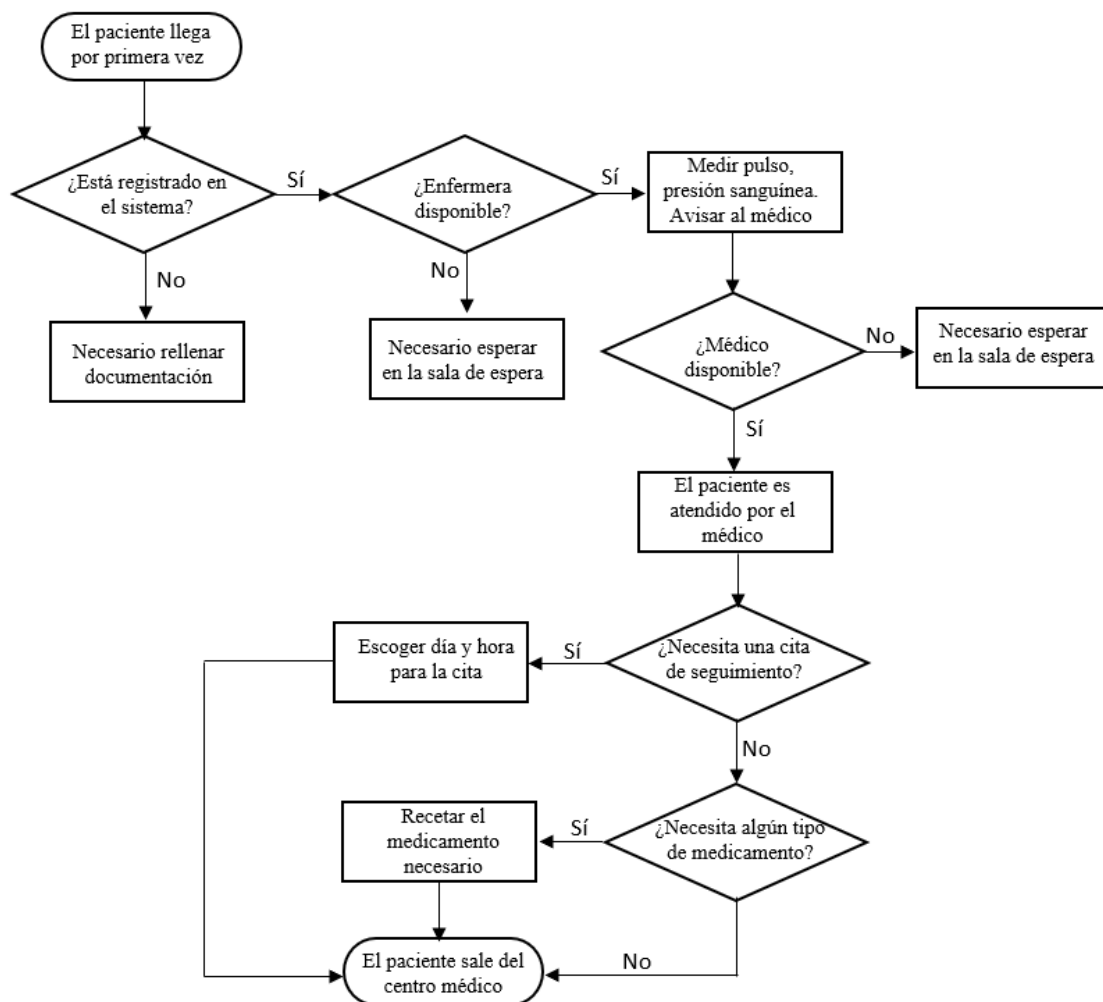


Ilustración 3.20. Ejemplo de Mapa de Procesos [8]

Es habitual en este tipo de representaciones describir un criterio que haga que las diferentes formas geométricas, rectángulos, rombos, etc., tengan un significado concreto. Para el ejemplo mostrado, la leyenda de significados es la que aparece en la Ilustración 3.21.

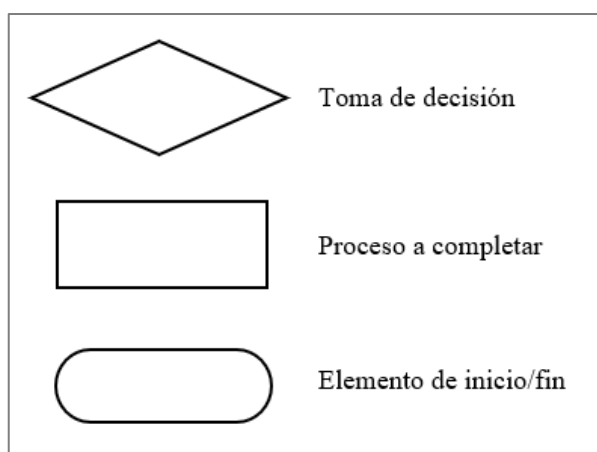


Ilustración 3.21. Leyenda utilizada para el Mapa de Procesos [8]

El objetivo de este tipo de gráficos es, por tanto, el de presentar de una manera sencilla, mediante diagramas de flujo, el funcionamiento de un proceso concreto, incluyendo cada uno de los pasos que se recorren, las decisiones a tomar al respecto y, también, los elementos que marcan cuál es el inicio y el fin del proceso.

3.1.11. Mapas Geográficos (GIS)

El uso de mapas geográficos como base para la visualización de los datos es lo que se conoce como Sistemas de Información Geográfica (GIS). En estos sistemas la clave es enlazar la información, cuantitativa o cualitativa, a marcadores o localizaciones geográficas [3], permitiendo a los usuarios integrar, analizar y representar aquella información que se encuentra asociada a un territorio geográfico concreto. El uso de este tipo de representaciones será muy útil en implementaciones como [10]:

- Predecir futuras áreas en las que va a existir una necesidad y analizar los recursos disponibles para solucionar esa necesidad.
- Comparar situaciones temporales o espaciales distintas de alguna propiedad en concreto, detectando así una posible evolución o cambio.
- Comparar dos o más variables, detectando posibles pautas de comportamiento.
- Detectar patrones en los datos de una manera sencilla.
- Obtener las características y propiedades de un lugar concreto.
- Cálculo de rutas óptimas entre dos puntos, siendo el ejemplo más claro el uso de programas como *Google Maps* en la actualidad.

De esta manera, los ámbitos en los que los mapas geográficos pueden ser empleados son muy amplios, concretamente, en cualquier campo que presente una componente espacial: gestión de recursos y de activos, cartografía, sociología o marketing, por ejemplo. Serán utilizados, a su vez, como herramientas que permitan la gestión y la toma de decisiones.

El GIS va a utilizar un sistema con un funcionamiento similar al de una base de datos para crear los mapas. De esta manera, el formato de los datos geográficos va a parecerse al de una base de datos, pero va a incluir algunas variables tales como una dirección, coordenada GPS, nombre de ciudad, etc., elementos que contribuyen a la visualización de los datos sobre el mapa geográfico. Además, este sistema separa la información en diferentes capas, que son almacenadas de manera independiente, permitiendo trabajar de una manera mucho más sencilla y eficaz.

En este tipo de representaciones cobran especial importancia las escalas de colores y escalas geográficas empleadas, puesto que afectarán al tipo de datos que pueden ser accedidos e incorporados dentro de los mapas. Así, habrá, por ejemplo, variables que solo recojan información a nivel estatal o variables que recojan información a nivel nacional. Esta variabilidad posible en las escalas permite la creación de mapas con diferentes grados de resolución y de detalle. Estos mapas permitirán resaltar la información en aquellos puntos que sean

de especial relevancia. A modo de ejemplo, en la siguiente ilustración se puede observar la mortalidad producida por el cáncer en los hombres en Estados Unidos entre los años 1990 y 1994 [10].

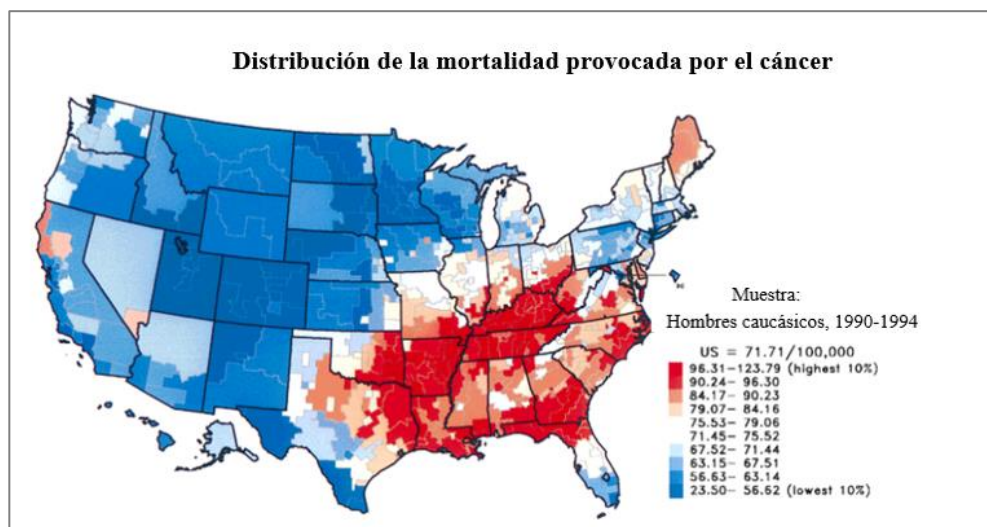


Ilustración 3.22. Ejemplo de Mapa Geográfico [10]

En el mapa observado, las zonas con colores más rojizos se corresponden con las áreas que presentan mayor mortalidad, mientras que, en las zonas en azul, la tasa de mortalidad desciende. A partir de este estudio se podría, por ejemplo, analizar el motivo por el cuál en las zonas en rojo afecta más esta enfermedad, y establecer medidas preventivas que permitan reducir la tasa de mortalidad existente.

3.1.12. Histogramas: Distribución de Frecuencia

Un gráfico de distribución de frecuencia ofrece una representación compacta de los datos, consistente en la organización de estos en grupos o intervalos de clases, normalmente del mismo ancho, para, posteriormente, realizar el cómputo, en los datos, del número de apariciones de cada intervalo, lo que se va a denominar frecuencia de clase. Por ejemplo, si se mide la velocidad con la que pasan por un radar 100 vehículos distintos, se puede obtener la siguiente tabla [2]:

Tabla 3.1. Muestra de la frecuencia de cada intervalo

Intervalo de clase (Km/h)	Frecuencia de clase
45 – 48	1
48 – 51	3
51 – 54	7
54 – 57	15
57 – 60	21
60 – 63	26
63 – 66	14
66 – 69	3
69 – 72	9
72 – 75	1

La lectura de esta tabla es sencilla: de los 100 vehículos que pasan por el radar, 1 pasa con una velocidad de entre 45 y 48 Km/h, 3 pasan entre 48 y 51 km/h y así sucesivamente. Como se puede observar, hay un total de 10 intervalos diferentes, cada uno de ancho 3 kilómetros por hora. Para calcular el valor de este ancho simplemente basta con coger los valores máximo y mínimo de los datos y aplicar la siguiente fórmula:

$$\text{Ancho del intervalo} = \frac{\text{Valor máximo} - \text{Valor mínimo}}{\text{Número de intervalos}}$$

Al aplicar esta expresión el resultado es un ancho de 3 unidades. Cada uno de los intervalos puede describirse, por tanto, de la siguiente manera:

$$45 \leq X < 48$$

En este caso, el valor 45 se conoce como el límite inferior del intervalo y el valor 48 es el límite superior. Como se puede observar en la definición, en este intervalo estarían recogidos los valores 45, 46, 47 pero no el 48, dado que se ha concluido que el ancho es 3, es decir, cada intervalo recoge 3 valores de velocidad, por lo que el límite superior queda excluido.

Un histograma se define como el método gráfico empleado para ilustrar una distribución de frecuencias para variables cuantitativas continuas. En este tipo de representaciones, en el eje X se trazan los diferentes intervalos, y las frecuencias de aparición se muestran en el eje Y. Así, se trata de la representación gráfica de una variable en forma de barras cuyas alturas están condicionadas por la frecuencia de aparición en los datos de cada uno de los intervalos y cuyo ancho es proporcional al rango de valores dentro de cada intervalo [2].

Tomando como partida la tabla anterior, se puede obtener la siguiente figura de la Ilustración 3.23.

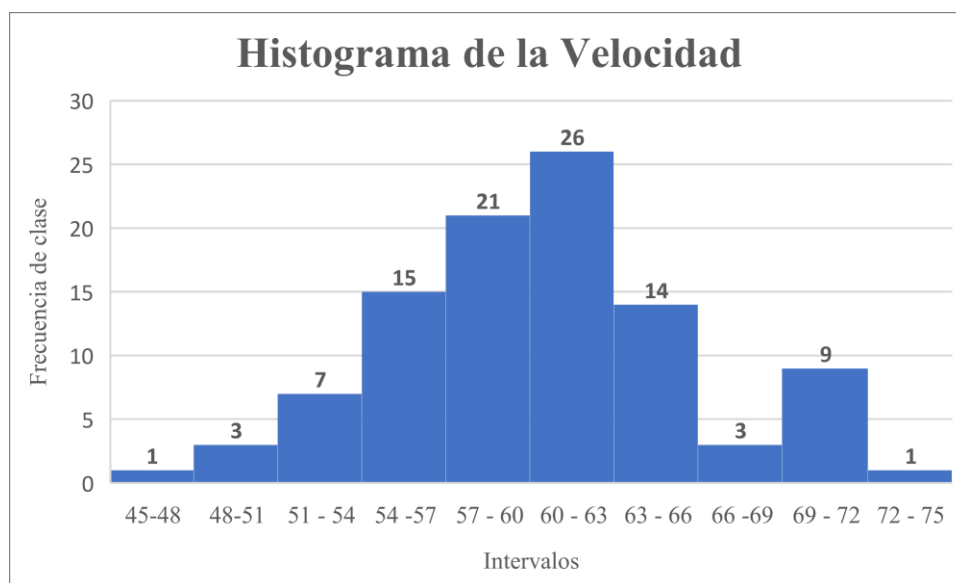


Ilustración 3.23. Ejemplo de Histograma

Mediante esta figura se puede observar el intervalo de velocidades más repetido entre los vehículos. Se podría también, por ejemplo, hacer un estudio que analizara la probabilidad de que un vehículo activara el radar. Bastaría con establecer la velocidad límite del radar y calcular, de cada 100 vehículos, cuántos de ellos estarían incumpliendo la velocidad establecida. Por tanto, los histogramas permiten resumir grandes cantidades de datos y son especialmente útiles en el análisis de probabilidades. Además, es posible apreciar el grado de homogeneidad o de dispersión presente entre todos los valores que componen la muestra, obteniendo la forma de la distribución, ya sea simétrica, desigual, o de cualquier otro tipo.

3.1.13. Infográficos (Infographics)

El término infografía se define como el diseño gráfico en el que se incluyen elementos de visualización de datos, ilustraciones, descripciones y textos en un formato que cuenta una historia completa. De esta manera, un infográfico es un medio de comunicar información gráficamente, más visual que un simple texto, que permite

al usuario entender dicha información de una manera más sencilla, de un solo vistazo.

Históricamente, los conceptos de infografía y visualización de datos se han considerado sinónimos, aunque cabe destacar algunas diferencias entre ellos. Cuando se hace referencia a la visualización de datos, este concepto implica la representación de datos numéricos a través de cuadros o gráficos. Estos gráficos tienen un comportamiento dinámico y permiten la interacción con los datos, además de una observación más exhaustiva de estos. Bajo el término infografía, se recoge la representación visual, de forma estática, de datos abstractos, tanto numéricos como no numéricos, teniendo como objetivo último el entendimiento y asimilación de un mensaje por parte del usuario [4].

En la Ilustración 3.24 se muestran dos figuras diferentes, que pretenden ejemplarizar la diferencia existente entre los conceptos de infografía y visualización de datos.

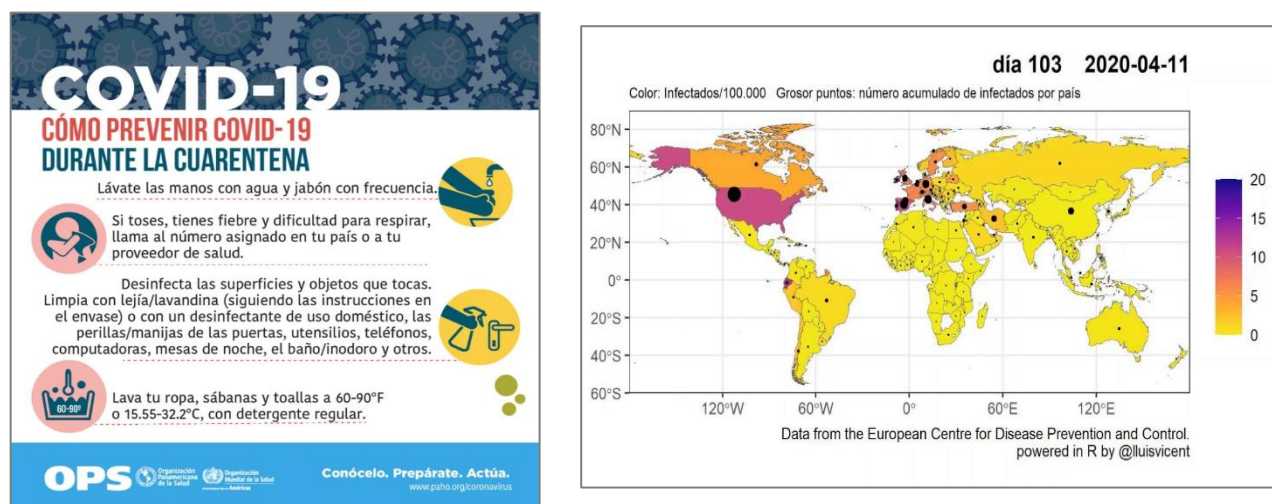


Ilustración 3.24. Diferencias entre Infografía (figura izquierda) y Visualización de datos (figura derecha)

El uso de infográficos se encuentra actualmente ampliamente extendido en diversos campos, entre los que destacan el periodismo, el turismo o las redes sociales. Por tanto, a la hora de crear una infografía, el primer paso va a ser conocer su finalidad: analizar un conjunto de datos, describir un producto, explicar un determinado concepto, etc. A continuación, será necesario recabar la información y filtrarla, eliminando aquellos datos que no sean de interés o no aporten valor al objetivo final.

Como ya se ha mencionado, el infográfico va a presentar los datos con el formato de una historia usando para ello la combinación de múltiples elementos visuales (imágenes, gráficos) y también elementos textuales. Debido a esta analogía, en el infográfico se van a poder distinguir tres secciones diferenciadas [1] [6] :

- **Introducción:** esta sección suele estar formada por un título y un breve párrafo textual, que tienen el objetivo de presentar al espectador el tema de la infografía y convencerlo de la relevancia de los datos mostrados en ella. La información aquí incluida debe ser clara y comprensible, pues será la base de los conceptos y temas tratados en secciones posteriores.
- **Cuerpo:** constituye la parte más importante del gráfico. En esta sección, el infográfico adquiere un significado completo y concreto para el espectador, por lo que debe ser efectivo y singular. Con este fin, será necesario incluir en esta parte información que sea desconocida hasta el momento y que resulte relevante y de novedad para el espectador. Este punto del gráfico es el que se pretende que el espectador recuerde, de ahí que se incluyan aquí la mayoría de las ilustraciones e imágenes, que pretenden impactar visualmente en el lector y captar su atención.
- **Conclusión:** esta sección actúa como cierre del gráfico y permite al diseñador volver a resaltar el mensaje que se quiere hacer llegar al espectador. Se incluye, además, en los casos que sean oportunos, lo que se conoce como “una llamada a la acción”, esto es, se debe incitar al lector a llevar a cabo cualquier actividad que el diseñador considera como necesaria después de conocer la información mostrada en el gráfico. Por ejemplo, se puede empujar al usuario a visitar una cierta página web, comprar un producto, firmar una petición, etc.

Otra sección opcional que se podría incluir en la estructura sería el pie, donde es recomendable incluir las diferentes fuentes de datos de donde se han obtenido la información mostrada y la cita del autor del infográfico. Para mostrar un ejemplo de la estructura de un infográfico, se incluye la Ilustración 3.25.



Ilustración 3.25. Ejemplo de Infográfico

En el infográfico mostrado se pretende alertar del peligro que supone para la fauna marina el vertido de plásticos en el océano. La introducción, observada en la parte superior de la imagen, está constituida por un título que introduce al lector en el tema tratado. A continuación, aparece el cuerpo, con una gran componente visual para la representación de los datos que aporta el diseñador. En la parte inferior del cuerpo aparece, además, la fuente utilizada para obtener dichos datos, sección a la que se ha denominado pie. Por último, se incluye la conclusión, formada por la acción a realizar por parte del usuario que es, en este caso, el intentar reducir la contaminación producida por el vertido de productos plásticos a los ecosistemas marinos.

Una vez vista la estructura, una primera clasificación de las infografías que se puede hacer es atendiendo a su elaboración. Así, una infografía puede ser [1] [4]:

- **Cronológica:** teniendo como base una línea temporal, muestra la evolución de un cierto conjunto de datos en el tiempo.
- **Comparativa:** utilizadas para comparar los valores de dos o más variables.
- **Diagrama de flujo:** muestra las actividades de control de un determinado proceso.
- **Descriptiva:** empleada para definir las características o propiedades de una determinada información.
- **Estadística:** presta especial interés a las métricas, encuestas, estudios de mercado, tendencias o resultados.
- **Geográfica:** relaciona la información con un lugar geográfico en un mapa.
- **Jerárquica:** cataloga la información en un conjunto de clases y subclases, estableciendo un orden en los datos.

Una segunda clasificación que se suele emplear es aquella que atiende a los diferentes objetivos que persigue el infográfico. De esta manera, pueden tener finalidades muy variadas [6]:

- **Periodística:** se aborda la infografía desde un punto de vista narrativo, apoyado mayormente en imágenes en detrimento de los datos numéricos.
- **Didáctica:** la finalidad de este tipo de infografías es la enseñanza, ya sea de las diferentes fases que componen un proceso de producción, los distintos productos que existen o cómo se utilizan, etc.

- **Informativa:** se pueden incluir en esta categoría todos aquellos gráficos que presentan datos e información que resulta ser de relevancia.
- **Empresarial:** presentan una marcada componente publicitaria, incluyendo, por ejemplo, presentación de productos, encuestas o estudios de mercado.
- **Persuasiva:** en estos gráficos, la información y datos representados tienen como objetivo el conducir al lector a una determinada conclusión de manera que le lleve a realizar una determinada acción posterior.

Todos los tipos de gráficos que se han visto en este capítulo suponen una herramienta crucial a la hora de analizar los datos, comprenderlos y obtener las conclusiones buscadas. Cuantas más representaciones se hagan de los datos, más puntos de vistas se podrán obtener, facilitando así la labor de comprensión de estos. Se ha podido observar también a lo largo del capítulo la gran cantidad de campos en los que se usan los datos, aunque con el avance y aparición de nuevas tecnologías, la generación de datos no ha hecho nada más que comenzar.

3.2. Herramientas de Visualización

Tras haber explicado en secciones anteriores la importancia de la visualización y los diferentes tipos de representaciones que se pueden hacer de un conjunto de datos dados, en esta sección el objetivo es presentar algunas de las herramientas más utilizadas para la creación de los gráficos visuales que permitan una mejor comprensión de los datos.

Al explicar cada una de las herramientas se pretende ofrecer al lector una pequeña guía de uso de estas, mostrando cuál será la interfaz que se le expondrá y los diferentes elementos que la componen.

3.2.1. Google Charts

La primera herramienta para analizar es *Google Charts* [24]. Esta aplicación software, de uso libre, se engloba dentro del paquete ofimático que *Google* ofrece a los usuarios. Permite la creación instantánea de una serie de gráficos dinámicos, entre los que se pueden incluir los vistos en apartados anteriores: diagrama de barras, gráficos circulares, diagramas de dispersión, etc.



Ilustración 3.26. Logotipo de *Google Charts*

La forma más usual de utilizar *Google Charts* es con el lenguaje Javascript, cuyo código se va a incrustar en la página web. El proceso de creación del gráfico es sencillo: se cargan las librerías necesarias en *Google Charts*, se listan los datos que se quieren representar y finalmente se crea un objeto gráfico al que se le asigna un identificador. Dicho identificador se utiliza posteriormente en la página web para crear el gráfico que muestra los datos asociados. De esta manera, los gráficos se presentan como clases del lenguaje Javascript.

Para cargar las librerías necesarias se usa un trozo de código Javascript, incluido dentro de la cabecera `<head>` de HTML5, que presenta el siguiente formato [24]:

```
<html>
  <head>
    <script type = "text/javascript"
      src="https://www.gstatic.com/charts/loader.js">
    </script>
    <script type="text/javascript">

    google.charts.load('current', {'packages':['corechart']});
    google.charts.setOnLoadCallback(drawChart);
```

Las dos primeras líneas del código, bajo etiqueta `<script>` son las que determinan que el lenguaje utilizado será Javascript y se encargan de cargar los elementos necesarios para la posterior ejecución correcta de la web.

A continuación, se llama a la función `Google.charts.load` para cargar los paquetes necesarios para un determinado tipo de gráficos. Esta llamada se repetirá tantas veces como tipos de gráficos diferentes haya. El primer parámetro que se le pasa a la función es el nombre o número de la versión que se quiere emplear, de manera que, el parámetro “current” indica que se debe cargar en el sistema la última versión de *Google Charts* que exista. El segundo parámetro, indica que los gráficos que se quieren utilizar pertenecen al paquete “corechart”, que incluye, por ejemplo, los diagramas de barras, diagramas de líneas, histogramas, etc.

Por último, se hace la llamada a la función `Google.charts.setOnLoadCallback`, que hace otra llamada a la función `drawChart`. Esta función debe estar definida de manera global en algún sitio de la página web y va a describir los datos y las representaciones que se van a hacer de ellos. Por ejemplo, para crear un diagrama circular que indique la composición del aire en nuestro entorno, se crea la siguiente función `drawChart` [24].

```
function drawChart() {
  var data = new Google.visualization.DataTable();
  data.addColumn('string', 'Element');
  data.addColumn('number', 'Percentage');
  data.addRows([
    ['Nitrogen', 0.78],
    ['Oxygen', 0.21],
    ['Other', 0.01]
  ]);
  var chart = new
Google.visualization.PieChart(document.getElementById('myPieChart'));
  chart.draw(data, null);
}
</script>
</head>

<body>
  <div id="myPieChart">
</body>
```

Una clase muy importante dentro de esta herramienta es la denominada *DataTable*, que proporciona métodos para ordenar, modificar o filtrar los datos. Se trata de una tabla bidimensional, formada por filas y columnas, donde cada columna recoge un tipo de datos concreto y, opcionalmente, un identificador para esos datos y una etiqueta. Esta clase facilitará el cambio de unos gráficos a otros, permitiendo al usuario obtener la representación que mejor se ajuste a los datos proporcionados. En este caso, se crea una tabla que va a tener 2 columnas: la primera de ellas va a ser de tipo *string* y está etiquetada como “Element”, mientras que la segunda va a ser de tipo numérico y su etiqueta va a ser “Percentage”. A su vez, esta tabla va a tener 3 filas, conteniendo el peso de cada elemento presente en el ambiente. Los datos dentro de esta tabla deben organizarse atendiendo al gráfico o representación que se quiera hacer de ellos. Así, para un diagrama de sectores, se requiere de dos columnas y tantas filas como sectores se pretende que haya. Otro tipo de representaciones pueden requerir tablas más complejas, con más columnas o filas. La que se ha creado presenta el siguiente aspecto:

Tabla 3.2. Creación de una tabla con los datos a representar

Element	Percentage
Nitrogen	0.78
Oxygen	0.21
Other	0.01

A continuación de haber introducido los datos, el siguiente paso es realizar su representación. Para ello, se debe, primero, crear una instancia de la clase de gráfico que se desea utilizar y, a continuación, llamar a la función *draw()*, a la que se le pasa como parámetro los datos introducidos. Tal y como se había comentado anteriormente, cada uno de los gráficos va a constituir una clase del lenguaje Javascript.

En el segmento de código anterior aparece otro aspecto que ya ha sido comentado. Y es que al gráfico se le asigna un identificador, en este caso, “*myPieChart*” al que se llama posteriormente en el cuerpo de la página web a través de la etiqueta `<div>`. Este elemento contiene el gráfico y lo representa en el momento indicado.

De esta manera, *Google Charts* ofrece una gran variedad de tipos de gráficos a emplear, con la ventaja de que permite al usuario personalizarlos de la manera que considere oportuna, ya sea añadiendo un título, etiquetando los ejes, incluyendo una leyenda o combinando a su gusto los colores mostrados. Además, los elementos gráficos son sumamente interactivos y muestran eventos que permiten conectar unos con otros con el objetivo de crear experiencias diferentes para el usuario, integrado todo en una página web. Por ejemplo, algunos de estos eventos se pueden producir cuando el usuario selecciona un campo determinado de un gráfico o cuando pasa el ratón sobre un elemento específico. Para la representación gráfica se utiliza el lenguaje HTML5, muy utilizado en el diseño web, ya que permite la compatibilidad con distintos navegadores [24].

Como ejemplo final, en la Ilustración 3.27. se recoge una representación sencilla que se puede llegar a lograr usando esta herramienta. En concreto, se muestran las ganancias de una empresa en un año, dividido en cuatro trimestres. Se muestran los mismos datos con dos gráficos diferentes, un diagrama de barras y un diagrama circular. Además, se observa el componente interactivo mencionado con anterioridad, al colocar el ratón encima de uno de los sectores del gráfico, se abre una pestaña que indica el trimestre que representa dicho sector, las ganancias de la empresa en ese trimestre y el peso de ese trimestre, sobre el porcentaje total.

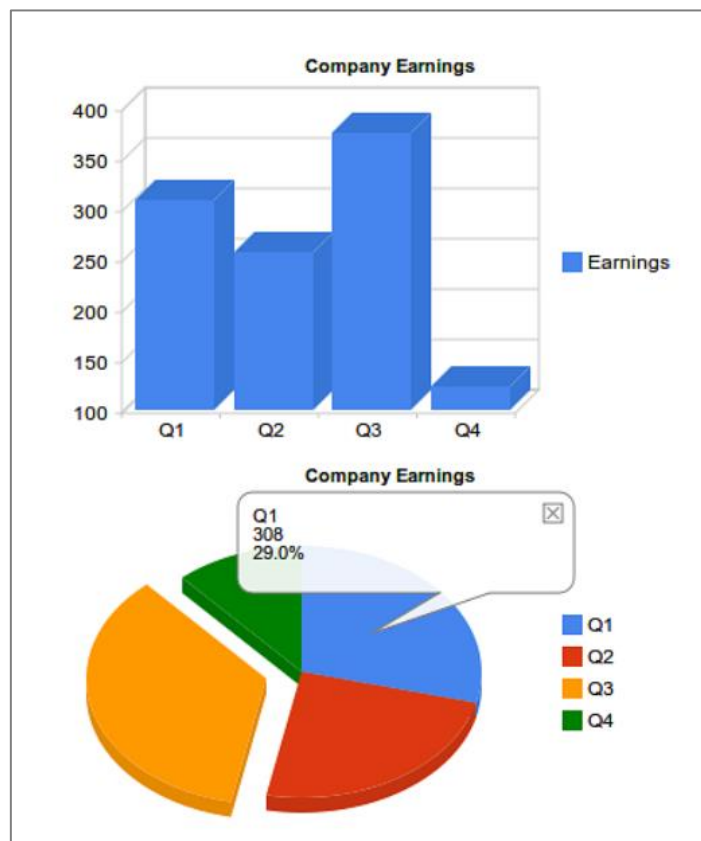


Ilustración 3.27. Ejemplo de representación con la herramienta *Google Charts* [24]

3.2.2. Tableau

Tableau se ha convertido en una de las plataformas de visualización de datos más utilizadas en la actualidad, al permitir, de manera sencilla, el análisis, descubrimiento y presentación de una gran variedad de datos. De esta manera, *Tableau* fue creado para impulsar al usuario a analizar sus datos independientemente de su nivel de conocimiento técnico [22], haciendo asequible el acceso a los datos para cualquier usuario.



Ilustración 3.28. Logotipo de la herramienta *Tableau*

Una de las características que diferencian a esta herramienta es el uso de VizQL, un lenguaje específico para consulta, análisis y visualización de datos. Dicho lenguaje se va a encargar de traducir los actos que el usuario lleva a cabo con el ratón a consultas que se van a hacer a una base de datos, dando como resultado otra acción determinada. De esta manera, el usuario interactuará con los datos dentro de un entorno visual y esta herramienta transformará los movimientos del usuario en las acciones posteriores necesarias. Así, se podrá trabajar, de una manera sencilla, con los datos, obteniendo patrones, relaciones, representando dichos datos en diferentes gráficos que pueden cambiar dinámicamente o compartir los resultados obtenidos.

Esta herramienta permite el uso de datos almacenados en una gran variedad de archivos o bases de datos, de manera que puede trabajar sin complicaciones con, por ejemplo, ficheros Excel o ficheros de texto, o bien datos contenidos en bases de datos tales como SQL u Oracle. El proceso de representación será, en la mayoría de los casos, el mismo, independientemente del tipo de fuente de datos que se emplee. Este proceso de representación comienza, por tanto, enlazando la herramienta *Tableau* con la fuente de datos correspondiente.

Una vez realizado este paso, se abre el directorio de trabajo de la herramienta, cuyo aspecto se presenta en la Ilustración 3.29 [22].

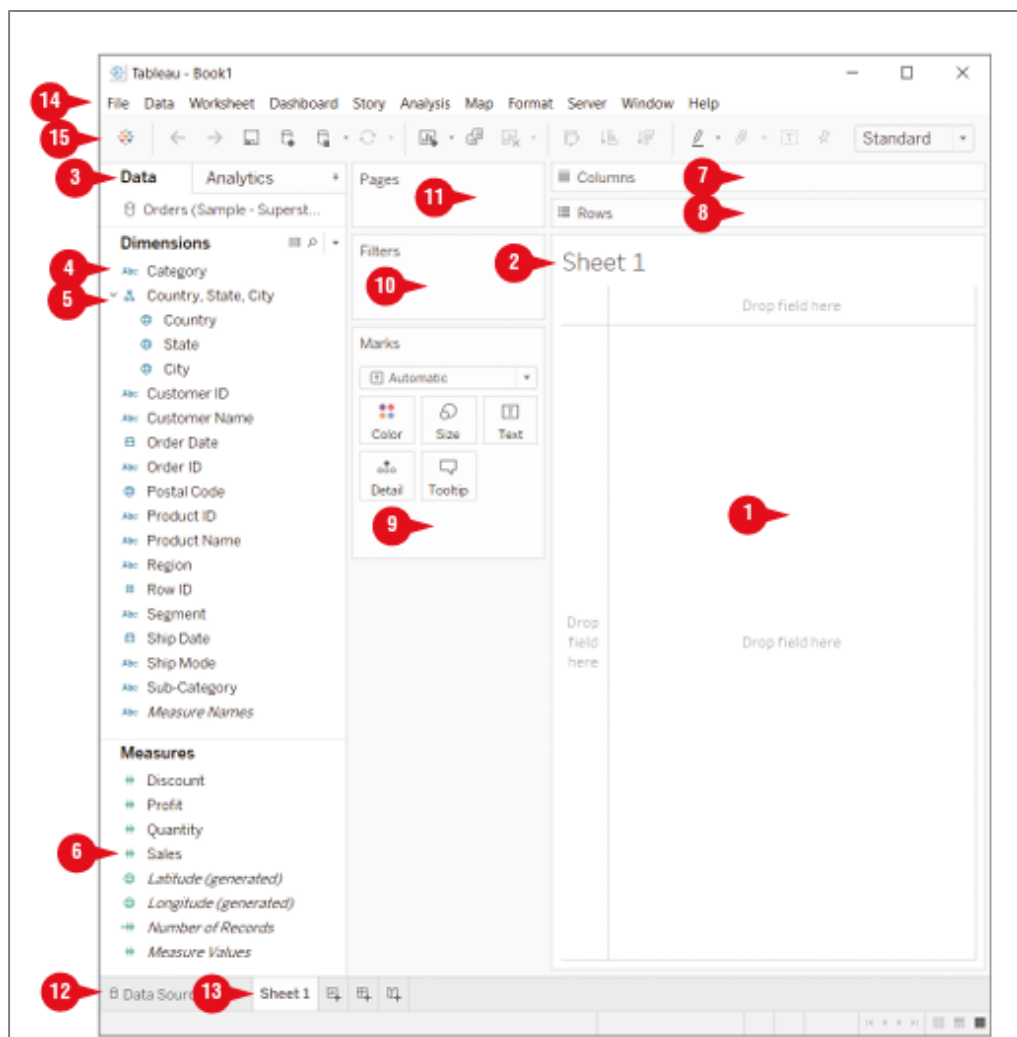


Ilustración 3.29. Interfaz mostrada por la herramienta *Tableau* [22]

En esta interfaz se han marcado 15 puntos diferentes que pueden resultar de interés [22]. El elemento (1) representa la plantilla en blanco, utilizada posteriormente para incluir las representaciones, e incluye el título (2) que va a llevar el gráfico. El elemento etiquetado como (3) hace referencia al panel de datos, que está dividido en dos secciones diferentes:

- Medidas (Measures): se corresponden con las variables numéricas. Añadiendo campos a esta sección, lo que se hace es indicar qué datos pertenecientes al conjunto de datos se quieren visualizar. Por defecto, *Tableau* ejecuta automáticamente funciones tales como la suma o la media aritmética de las diferentes medidas. De esta manera, se podría obtener, utilizando una de las variables mostradas, el descuento (*Discount*) medio obtenido a través de diferentes ventas de productos.
- Dimensiones (Dimensions): en esta sección se engloban las variables descriptivas. El objetivo es permitir la agrupación de los valores de una determinada medida. Por ejemplo, si se calcula el total de ingresos debidos a las ventas, lo cual es una medida, este valor se puede dividir en los ingresos producidos por país, estado o ciudad, que son las diferentes dimensiones.

En esta herramienta, las interacciones consisten arrastrar y soltar (*drag and drop*) los elementos necesarios en la plantilla en blanco. Estos elementos serán tanto las dimensiones (4), independientemente de que sean de tipo jerárquico (5), y las medidas (6). Alternativamente, estos elementos también se pueden añadir en los campos denominados *Columns* (7) y *Rows* (8) de manera que aparezcan en las diferentes visualizaciones.

Los campos del panel de datos también se pueden colocar sobre los espacios denominados *Marks* (9), *Filters* (10) y *Pages* (11). Estas etiquetas serán útiles si, por ejemplo, se quiere cambiar el color o el tamaño de los marcadores o únicamente representar marcas para un subconjunto de los datos, que ha sido filtrado previamente.

En la zona inferior izquierda aparecen los botones que permiten volver al editor de la fuente de datos seleccionada (12) y el que permite moverse a través de las diferentes hojas (13), en cada una de las cuales se representa una visualización.

El elemento etiquetado como (14), situado en la parte superior de la interfaz, presenta la barra de herramientas. Aunque en *Tableau* la mayoría de las acciones se producen interaccionando directamente con el ratón, esta barra de herramientas permitirá llevar a cabo un conjunto de acciones y configuraciones adicionales. Las principales funciones que permite se agrupan en diferentes campos [22]:

- Menú File: en este desplegable se engloban todas aquellas funciones que se pueden realizar sobre un fichero: abrirlo, guardar una versión del archivo en ejecución, cambiar la localización de destino de un determinado archivo, exportar un fichero a una versión determinada, etc.
- Menú Data: dentro de este desplegable destaca la opción *Insert*, que permite añadir, de una manera rápida y eficaz, datos procedentes, por ejemplo, de una tabla englobada dentro de una página web. Este paso convertirá dicha tabla en una nueva fuente de datos para nuestra hoja de trabajo.
- Menú Worksheet: En este campo destacan la opción *Export*, que permite volcar los datos tratados en *Tableau* en una imagen o una base de datos.
- Menú Dashboard: un *dashboard* se define como el documento u hoja que recoge, mediante representación gráfica, las principales métricas o KPIs que aparecen en un determinado estudio. Al seleccionar la opción *Actions* de este menú, se presenta la posibilidad de añadir interactividad a los *dashboards* creados.
- Menú Story: este menú permite crear una historia a partir de las hojas de trabajo y *dashboards* que se hayan creado anteriormente. Básicamente, el contenido se va a organizar de una manera secuencial para su posterior representación y se va a enriquecer con comentarios adicionales que contribuyan a la comprensión de la historia.

Estos menús son los más utilizados y tienen una mayor importancia. Sin embargo, también se presentan otros menús complementarios:

- Menú Map: este menú se utiliza para seleccionar entre diferentes mapas a colocar en el fondo de la plantilla.
- Menú Format: para establecer la alineación, tamaño, fuente y otras opciones relativas al formato de presentación.

- **Menú *Server*:** empleado para compartir los *dashboards* creados a través de la herramienta *Tableau*.
- **Menú *Window*:** incluye las diferentes opciones de presentación de la pantalla, por ejemplo, el modo pantalla completa.
- **Menú *Help*:** mediante este menú el usuario puede acceder a la página de ayuda de *Tableau*, a vídeos explicativos o ejemplos de presentaciones ya realizadas.

Por último, debajo de la barra de herramientas, se presentan una serie de botones (15) que permiten ejecutar directamente ciertas acciones como, por ejemplo, volver a la pantalla de inicio o deshacer el último movimiento realizado, útil en el caso de que se haya cometido un error o se quiera rehacer de otra manera un movimiento.

Así, un ejemplo sencillo del uso de *Tableau* se muestra en la Ilustración 3.30. En esta imagen se presenta el volumen de ventas, de diferentes productos, de una determinada empresa y la clasificación de los diferentes segmentos de clientes que han realizado estas compras. De esta manera, se crea un diagrama de barras en el que cada columna representa un producto diferente que se vende y, dentro de cada barra, el peso de cada segmento en la cantidad total de ventas. *Tableau* cuenta también con elementos interactivos, de manera que, al colocar el ratón sobre el gráfico, se abre una ventana con la información relativa a ese elemento, lo cual puede ser de gran utilidad para el usuario.

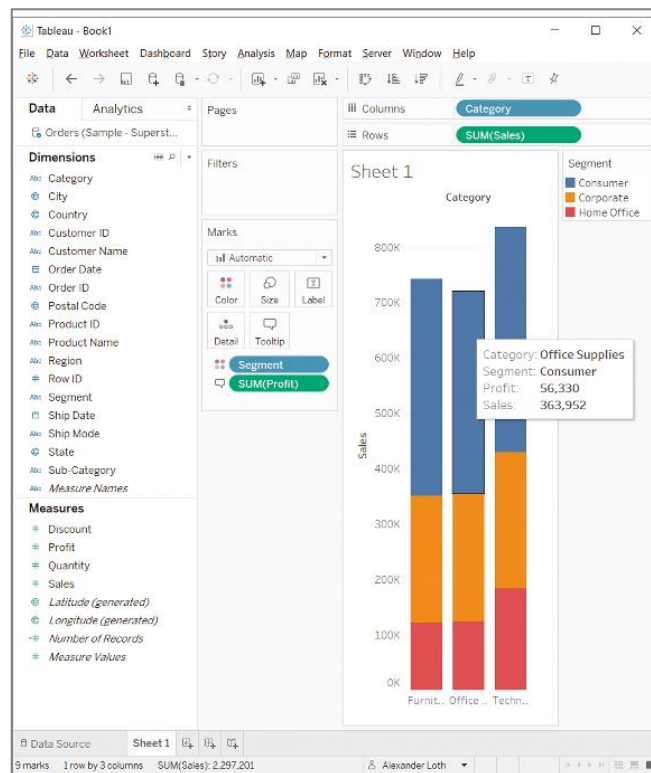


Ilustración 3.30. Ejemplo de representación con la herramienta *Tableau* [22]

Lo mostrado en la Ilustración 3.30 supone un ejemplo sencillo de representación. Sin embargo, también se pueden construir *Dashboards* de gran complejidad y que incluyan más de un tipo de representación. Este hecho se muestra en la Ilustración 3.31 [23]. En este caso, se trata de una empresa que analiza los tipos de envíos de paquetes que realiza, midiendo la demora que tienen, o el número de pedidos que entregan en un período de tiempo determinado.

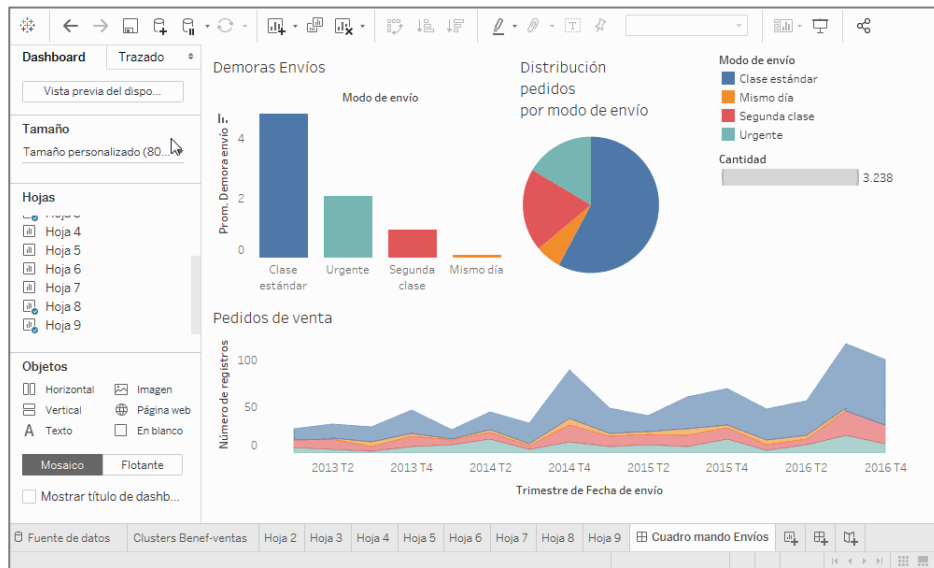


Ilustración 3.31. Ejemplo de *Dashboard* creado con *Tableau*

3.2.3. MATLAB

Aunque MATLAB no está catalogada como una herramienta de visualización pura, se incluye en este apartado por la variedad de representaciones que ofrece y la gran cantidad de tipos de datos que puede manejar, además de por ser la herramienta utilizada en el posterior desarrollo práctico del proyecto.



Ilustración 3.32. Logotipo de la herramienta MATLAB [11]

MATLAB, abreviatura de MATrix LABoratory, se define como una plataforma de programación y cálculo numérico que ofrece un entorno de desarrollo integrado con un lenguaje propio. Es ampliamente utilizada en el análisis de datos, el desarrollo de algoritmos, la creación de aplicaciones, la realización de cálculos a gran escala y la creación de modelos de trabajo. Así, los campos en los que se puede emplear esta herramienta son muy diversos [11]:

- **Sistemas de control:** MATLAB es usada como herramienta en todas y cada una de las diferentes etapas de desarrollo, que se extienden desde la creación del modelo de la planta hasta su posterior diseño, incluyendo los ajustes necesarios de los algoritmos de control y supervisión. Además, contribuye también a la implementación física de estos sistemas gracias a la generación automática de código y a la validación y verificación que hace constantemente del sistema.
- **Machine Learning:** esta herramienta simplifica notablemente el *Machine Learning* gracias a la existencia de sencillas aplicaciones que permiten entrenar y comparar modelos, técnicas avanzadas usadas para el procesamiento de señales, algoritmos de clasificación tanto para el aprendizaje supervisado como el no supervisado, y una ejecución mucho más rápida en la mayoría de los cálculos necesarios en *Machine Learning*.
- **Procesamiento de señales:** MATLAB cuenta con una serie de funciones integradas para el estudio y el preprocesamiento de datos de series temporales, el análisis en el dominio de la frecuencia y la obtención de medidas de una determinada señal. Además, presenta algoritmos que permiten diseñar e implementar filtros digitales, FIR e IIR. Así, proporciona los métodos necesarios para visualizar señales en los dominios del tiempo y de la frecuencia, detectando, de una manera sencilla, posibles patrones y tendencias en las señales.

- **Pruebas y mediciones:** en el entorno de desarrollo que integra cuenta con la posibilidad de poder conectarse y comunicarse con una gran variedad de dispositivos HW, como pueden ser placas de circuitos, sensores, instrumentos de prueba, cámaras web, etc. Además, ofrece la opción de, sin salir del mismo entorno de trabajo, visualizar y analizar, en tiempo real, los datos obtenidos.
- **Procesamiento de imágenes:** actualmente se generan instantáneamente grandes cantidades de datos de imagen y vídeo, lo que hace especialmente necesario el disponer de una herramienta que proporcione un procesamiento y una comprensión de los datos captados. MATLAB dispone de aplicaciones que permiten comunicarse con los datos de una manera interactiva, generando automáticamente nuevo código procedente de esa interacción. Estas aplicaciones proporcionan también otras funciones como la calibración de las cámaras, el etiquetado de imágenes y vídeos y la segmentación de imágenes.
- **Comunicaciones inalámbricas:** ofrece la posibilidad de simular sistemas que transmiten las señales a través del aire, generando formas de onda personalizadas con el objetivo de verificar el cumplimiento de los estándares 5G, LTE, WLAN, entre otros. Además, permite comprobar el comportamiento de un sistema que use tanto elementos digitales, como de RF o antenas.

Cuando el usuario inicia por primera vez la herramienta, le aparece una interfaz como la mostrada en la Ilustración 3.33 [11].

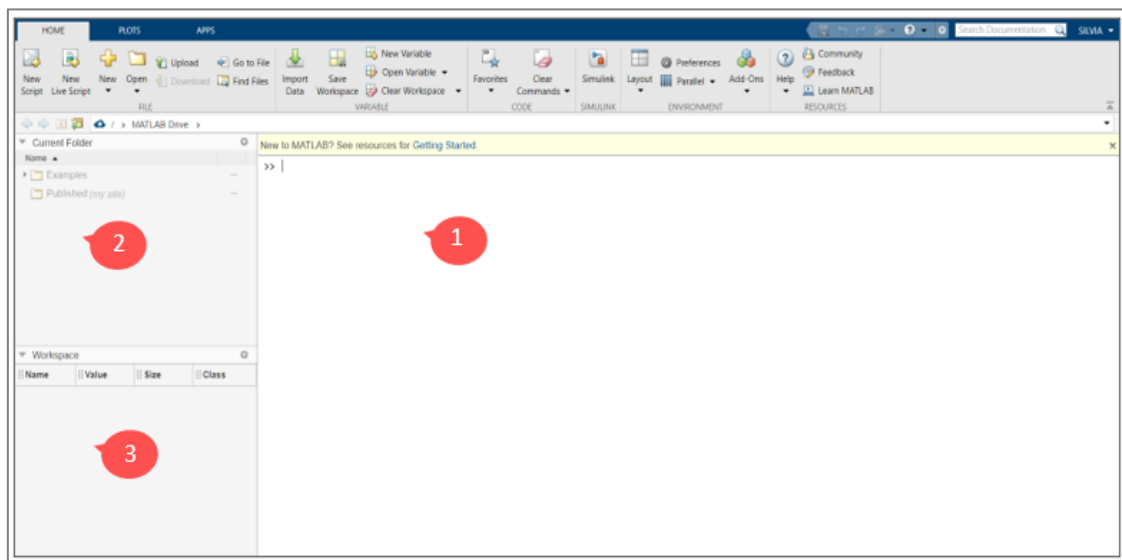


Ilustración 3.33. Interfaz mostrada por la herramienta MATLAB [11]

Como se puede observar, presenta 3 partes diferenciadas:

- **Ventana de comandos (Command Window):** en esta parte, etiquetada como (1), se van a escribir, a continuación del símbolo ‘>>’ todas aquellas instrucciones que se quieran ejecutar. Entre estas instrucciones se pueden incluir cálculos numéricos con matrices, creación de variables o llamadas a funciones.
- **Current Folder:** en esta parte de la interfaz, (2), se muestra la estructura de directorios y ficheros que están por debajo del directorio en el que el usuario se encuentra en ese momento, denominada la carpeta de trabajo. Si se quiere ejecutar un determinado fichero en MATLAB, es necesario situarse en el directorio que contiene dicho fichero, de lo contrario, para poder ejecutarlo, se deberá especificar la ruta de éste.
- **Workspace:** en esta ventana se muestra información de las variables definidas en el espacio de trabajo, especificando su nombre, valor, tamaño y clase.

En la parte superior del entorno de ventanas se muestra la barra de herramientas que permitirá, entre otras acciones, abrir un archivo existente, crear un archivo nuevo, importar datos de un determinado fichero al *Workspace* o configurar el aspecto o las preferencias del entorno de trabajo.

La forma más habitual de utilizar esta herramienta será crear un *script*, esto es, un fichero en el que se van a incluir de manera secuencial todas las instrucciones que se quieren ejecutar al mismo tiempo, en vez de

escribirlas una a una en la ventana de comandos, y posteriormente, ejecutar dicho *script*, mostrándose los resultados en la ventana de trabajo o *Workspace*.

Además de todas las funciones de cálculo numérico, otra propiedad importante de MATLAB es la gran cantidad de métodos que existen para visualizar un conjunto de datos dados. Algunos de los métodos más utilizados son [12]:

- *Plot(X,Y)*: crea un gráfico en 2D de los datos contenidos en la variable Y frente a los valores correspondientes de la variable X.
- *Stem(X,Y)*: representa la secuencia de datos Y como líneas verticales que se extienden sobre los valores especificados por la variable X. El valor de los datos se indica con pequeños círculos que se colocan al final de cada línea vertical.
- *Bar(X,Y)* [17]: este método crea un diagrama de barras, con una barra para cada elemento contenido en Y, colocada sobre cada valor correspondiente en el eje X.
- *Semilogy(X,Y)* [19]: este método utiliza una escala lineal para representar los valores en el eje X, mientras que utiliza una escala logarítmica en base 10 para los valores del eje Y.
- *Scatter(X,Y)*: crea un diagrama de dispersión con círculos que se van a ubicar en los valores correspondientes especificados por X e Y.
- *Mesh(X,Y,Z)*: primero de los métodos vistos que realiza una representación en 3D. Esta función crea una malla con los valores dados por las variables X e Y. El color de la malla en cada punto vendrá dado por los valores contenidos en Z, es decir, se creará una escala de color donde cada color especificará una altura concreta de la superficie.
- *Surf(X,Y,Z)* [13]: este método crea un gráfico de superficie tridimensional, utilizando los valores de Z como las correspondientes alturas situadas en una cuadrícula creada por los valores de las variables X e Y.
- *Contour(X,Y,Z)* [14]: esta gráfica de contorno crea una cuadrícula con los valores de las variables X e Y. A continuación, crea líneas que conectan los puntos de Z que tienen el mismo valor. El resultado es una gráfica en 2D que representa 3 variables diferentes.

Algunos de estos métodos serán los empleados en posteriores capítulos para obtener las representaciones visuales necesarias de los datos que están bajo estudio.

3.3. Tendencias en la Visualización de datos: *Big Data*

Una de las tendencias actuales en el campo de los datos es lo que se conoce como *Big Data*. Este término hace referencia a la existencia de grandes volúmenes de datos procedentes de las más diversas fuentes: sensores utilizados para recopilar algún tipo específico de información, imágenes y vídeos, transacciones realizadas por la sociedad, señales procedentes de teléfonos móviles, datos obtenidos a partir de las interacciones máquina a máquina (M2M), etc. Es tal el volumen de datos que hace que, al referirse al *Big Data*, se esté hablando de Petabytes de información, siendo un Petabyte equivalente a 1024 Terabytes [2].



Ilustración 3.34. *Big Data*, tendencia actual en el campo de los datos

Esta cantidad masiva de información que se obtiene de forma casi instantánea hace que sea muy difícil procesar, manipular o manejar todos estos datos usando los métodos tradicionales de análisis. Además, la mayoría de los datos que se recogen tienen una componente dinámica que hace que cambien constantemente. Estos dos hechos muestran la relevancia que ha adquirido hoy en día la visualización de los datos. Las representaciones visuales, algunas de las cuales han sido explicadas en apartados anteriores, ayudan a describir y comprender de una manera más sencilla y eficaz los datos, permitiendo obtener significativas conclusiones que podrán contribuir a una posterior toma efectiva de decisiones. Los datos, en su forma más pura, no son excesivamente útiles, a menos que sean tratados y analizados, por lo que es, al presentarlos visualmente, cuando existe la posibilidad de detectar patrones o encontrar relaciones entre varias variables simultáneamente.

Como respuesta al problema planteado sobre la dificultad de analizar este gran volumen de datos, que excede las capacidades del SW tradicional, se han desarrollado aplicaciones computacionales más avanzadas con el objetivo de almacenar, analizar, visualizar y compartir grandes cantidades de datos. Sin embargo, *Big Data* difiere, en parte, de lo que se consideran datos convencionales, siendo caracterizado comúnmente por lo que se conoce como “el Modelo de las 5V’s” [2] [4]:

- **Volumen:** hace referencia a la cantidad de datos originados y almacenados alrededor del mundo, cuya magnitud continuará creciendo de manera exponencial en los próximos años. No todos los conjuntos de datos se van a considerar *Big Data*. Su designación o no bajo este término dependerá del valor y el potencial conocimiento que se pueda adquirir de esos conjuntos de datos.
- **Variiedad:** los datos pueden proceder de fuentes heterogéneas: textos, audios, vídeos, imágenes, etc., con lo que aparecen nuevos tipos de datos estructurados y no estructurados. Con este término se hace referencia, por tanto, a la diversidad de tipos y de naturaleza que pueden presentar los datos.
- **Velocidad:** este concepto hace referencia a la rapidez con la que se generan y se procesan los datos, la cual está acelerando, haciendo indispensable la habilidad de analizar datos en tiempo real con el objetivo de cumplir con las diferentes demandas y los retos que se plantean en la sociedad.
- **Veracidad:** los datos obtenidos deben ser de calidad, es decir, deben ajustarse a la realidad, ser fiables. Este hecho es de vital importancia puesto que la falta de veracidad en los datos puede afectar significativamente a los resultados del análisis de estos.
- **Valor:** los datos obtenidos deben ser útiles y aportar nuevos conocimientos.

El conocimiento derivado del estudio del *Big Data* utilizado para la toma inteligente de decisiones, basándose en los datos, es lo que se conoce como *Business Intelligence* (BI) [8]. Esta tendencia ha ayudado a detectar aquellas áreas en las que las empresas deben volcar todos sus esfuerzos si quieren tener éxito. Las compañías de hoy en día tienen la posibilidad de conocer de una manera más exacta los hábitos de cada consumidor individualmente, pudiendo elaborar así una oferta mucho más personalizada. Pueden, además, conocer sus necesidades y crear productos orientados a mejorar la satisfacción del cliente de manera que este pueda, en un futuro, recomendar ese producto a otros. Sin embargo, el ámbito empresarial no es el único en el que el uso del *Big Data* ha tenido una importancia crítica. Actualmente, otras áreas en las que se usa esta tecnología de manera exitosa son [2] [18]:

- **Fabricación e Industria:** en ciertos entornos de producción, como pueden ser las farmacéuticas o las plantas químicas, donde la complejidad de procesado es muy alta e incluye una gran cantidad de factores, la recogida masiva de datos resulta indispensable para valorar y corregir posibles imperfecciones o problemas, incluso en tiempo real. El uso de estas herramientas contribuye a la mejora de los procesos y a la detección de patrones, a partir de datos históricos, que permitan predecir las relaciones entre las entradas y las posteriores salidas de un proceso. Además, es posible detectar los factores con mayor influencia en un proceso, optimizando aquellos que resulten beneficiosos y reduciendo el impacto de los que resulten perjudiciales.
- **Salud:** el uso del *Big Data* en el entorno sanitario está mejorando notablemente el sistema de salud público. Desde la creación automatizada de informes, que muestran el estado de un paciente, hasta los registros de los pacientes en una determinada base de datos, esta herramienta está permitiendo una mejora notable de los beneficios junto con una significativa reducción de los costes. Además, se hace uso de los datos para la predicción de epidemias, la cura o detección temprana de ciertas enfermedades, la mejora de la calidad de vida de un paciente, etc.

- Medios de comunicación: gracias a las grandes cantidades de datos, los profesionales en medios y publicidad pueden elegir como blanco a un segmento específico de clientes, a los que transmitir un determinado mensaje o contenido que esté alineado con sus intereses.
- Deportes: en este campo resulta de gran utilidad el *Big Data* para, por ejemplo, la monitorización, en todo momento, del estado físico de un deportista, pudiendo detectar así anomalías o posibles lesiones en tiempo real.
- Internet de las cosas (IoT): este término hace referencia a una creciente red de objetos conectados, que van a compartir información en tiempo real. Estos objetos van desde máquinas industriales hasta los *smartphones* que usamos diariamente, de manera que, a corto plazo, la ciudad en la que vivimos estará completamente conectada a Internet, lo que mejorará indudablemente la calidad de vida de sus habitantes en muchos sentidos.

Dicho esto, como es evidente, el uso masivo de datos ha mejorado la toma de decisiones en muchos ámbitos. Sin embargo, todas estas mejoras vienen acompañadas de una serie de retos y desafíos a los que todavía hay que enfrentarse, como la existencia de una infraestructura tecnológica inadecuada o la necesidad de garantía de privacidad y seguridad al usar una gran cantidad de datos en muchas ocasiones personales. Esto deja patente que hay mucho camino todavía por recorrer.

4. DESARROLLO PRÁCTICO DEL TRABAJO

La información es el aceite del Siglo XXI, y la analítica es el motor de combustión.

Peter Sondergaard

El objetivo principal de este capítulo es proporcionar una guía sobre el procedimiento que se ha llevado a cabo para trabajar con la información de cara a elaborar, como fin último, una interfaz gráfica que permita al usuario una fácil y rápida visualización de los datos bajo estudio.

El capítulo, como ya se ha mencionado en la estructura, estará dividido en tres partes: la recolección de los datos, la transformación de éstos para que puedan ser manejados por MATLAB y la creación de las diferentes gráficas para visualizarlos. Las pautas dadas en este capítulo sustentarán la creación de la Interfaz Gráfica de Usuario, cuyo funcionamiento se expondrá en un capítulo posterior.

4.1. Recolección de datos

El primer paso en este proyecto consistía en adquirir una base de datos que proporcionara la información a tratar, analizar y visualizar. Para tal motivo, se decidió utilizar un estudio disponible en el repositorio *IEEE DataPort*.



Ilustración 4.1. Logotipo del *IEEE DataPort*

Dicho estudio, titulado “*Energy harvesting: Electromagnetic Site Survey*” [15] fue llevado a cabo con el objetivo de examinar la cantidad de potencia disponible en lugares donde típicamente se podría colocar un dispositivo electrónico usado para aplicaciones como el IoT o WSN. Para alimentar este dispositivo se puede utilizar lo que se conoce como *Energy Harvesting*, o uso de la energía disponible en las ondas electromagnéticas ambientales, que son producidas por elementos tales como torres de telefonía, WIFI, equipamiento inalámbrico, etc. Concretamente, de entre todas estas fuentes, la categoría en la que se realiza el barrido de ondas electromagnéticas se denomina *RF Energy Harvesting (RFEH)* [15].

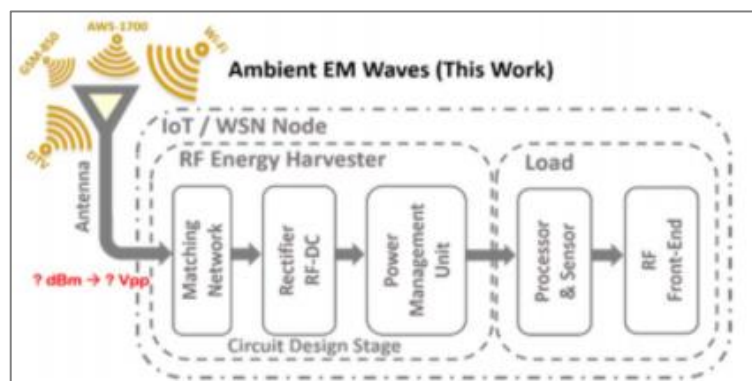


Ilustración 4.2. Diagrama general de un Nodo IoT/WSN alimentado por RFEH [15]

Para realizar este análisis se eligió el municipio de Mayagüez, Puerto Rico, del que se cubrió un área total de 18 km^2 . Dentro de esta área, se clasificaron las diferentes zonas en 3 categorías [15]:

- **Avenidas principales** (*Main Avenues*): grandes espacios públicos en los que se hallan hospitales, vegetación, edificios, etc. En este caso, el estudio incluye una única avenida.
- **Espacios abiertos** (*Open Places*): fundamentalmente se engloban dentro de esta categoría los aparcamientos de diversos centros comerciales de la ciudad y plazas públicas. Así, se cuenta con hasta 9 lugares objeto de estudio.
- **Calles y carreteras convencionales** (*Streets & Roads*): lugares ubicados en densas zonas residenciales y con zonas comerciales a ambos lados. Para el caso de las calles, el estudio cubre un total de 3 áreas distintas.

Una vez clasificadas las diferentes zonas, se instaló el equipamiento en un vehículo con el objetivo de cubrir el área descrita. En la Ilustración 4.3 y en la Tabla 4.1 [15] se puede observar el equipamiento utilizado.



Ilustración 4.3. Configuración del *Drive Test* [15]

Tabla 4.1. Equipamiento y características

Vendedor	Equipamiento	Descripción
KEYSIGHT®	N9343C	Analizador de espectro portátil (HSA) que soporta un amplio rango de frecuencias. Posee una precisión para la amplitud de ± 1.3 dB y un ancho de banda de 3 MHz.
AARONIA®	OmniLOG30800	Antena omnidireccional con polarización lineal. Soporta un rango de frecuencias que va desde los 300 MHz hasta los 8 GHz y presenta una impedancia de 50 Ohmios.
Mini-Circuits®	CBL-6FT-SMSM+	Cable de alta precisión empleado fundamentalmente para la realización de pruebas. Presenta un conector de tipo SMA y garantiza unas pérdidas de inserción, para el rango de frecuencias usado, no superior a 4 dB.
ChangHong®	GPS Antenna	Antena con impedancia de salida de 50 Ohmios, ancho de banda de 10 MHz y ganancia de 28 dB. Polarización RHCP.

Para realizar las medidas de potencia el vehículo se mantiene, en todo momento, con una velocidad entre 25 y 40 km/h con el fin de realizar un barrido en 20 frecuencias distintas de manera precisa. Para asegurar, además, la mayor exactitud factible en los valores obtenidos, todas las fuentes de potencia presentes en el vehículo, tales como radios, teléfonos móviles o módulos Bluetooth, se apagan para evitar posibles interferencias.

Así, se utilizan para medir la potencia recibida un total de 20 canales, cada uno de ellos centrado en una frecuencia diferente y con un ancho de banda igual al 4% de su frecuencia central. De esta manera, en la Tabla 4.2 se muestran estos valores [15].

Tabla 4.2. Canales utilizados para la medida

Canal	Frecuencia Central (MHz)	Ancho de banda (MHz)
1	300	12
2	325	13
3	333	14
4	375	15
5	525	21
6	601	25
7	624	36
8	708	29
9	735	30
10	835	34
11	882	36
12	1733	70
13	1880	76
14	1960	79
15	2133	86
16	2155	87
17	2354	95
18	2437	98
19	2559	103
20	5800	232

El proceso de recolección de datos es, por tanto, muy sencillo. Se elige, en primer lugar, la zona que se va a analizar: Avenida, Espacio Abierto o Calle. A continuación, se recorre dicha zona con el vehículo mostrado, anotando en cada punto recorrido parámetros como: el instante de tiempo en el se toma la medida, las coordenadas geográficas (latitud y longitud) de dicho punto y, por último, la potencia recibida por la antena para cada uno de los 20 canales mostrados en la tabla anterior. De esta manera, y con ayuda de un ordenador portátil, se almacenan los datos y se genera un fichero con extensión KML que, leído usando un editor de texto plano, presenta el aspecto mostrado por la Ilustración 4.4. En este caso se trata de la Avenida bajo estudio. Se presentan las medidas de los 4 primeros canales y de los 5 últimos para una ubicación concreta, dada por los valores de latitud y longitud, dentro de dicha avenida.

```

AV_1_1: Bloc de notas
Archivo Edición Formato Ver Ayuda
<Folder>
  <name>Result Level</name>
  <Placemark>
    <visibility>1</visibility>
    <styleUrl>#34</styleUrl>
    <name></name>
    <description>
      System Time: 09:40:41
      UTC Time: 19:57:59
      Longitude(degree): -67.151245
      Latitude(degree): 18.172382
      Altitude(m): 5.800
      GPS Status: 1
      1-Channel NO: 1
      1-FREQ(HZ): 300000000.000
      1-BAND(HZ): 12000000.000
      1-Channel Power(dBm): -89.285
      2-Channel NO: 2
      2-FREQ(HZ): 325000000.000
      2-BAND(HZ): 13000000.000
      2-Channel Power(dBm): -86.536
      3-Channel NO: 3
      3-FREQ(HZ): 333000000.000
      3-BAND(HZ): 14000000.000
      3-Channel Power(dBm): -86.279
      4-Channel NO: 4
      4-FREQ(HZ): 375000000.000
      4-BAND(HZ): 15000000.000
      4-Channel Power(dBm): -80.793
    </description>
  </Placemark>
</Folder>

```

```

AV_1_1: Bloc de notas
Archivo Edición Formato Ver Ayuda
15-Channel NO: 15
15-FREQ(HZ): 2133000000.000
15-BAND(HZ): 86000000.000
15-Channel Power(dBm): -23.338
16-Channel NO: 16
16-FREQ(HZ): 2155000000.000
16-BAND(HZ): 87000000.000
16-Channel Power(dBm): -26.032
17-Channel NO: 17
17-FREQ(HZ): 2354000000.000
17-BAND(HZ): 95000000.000
17-Channel Power(dBm): -48.149
18-Channel NO: 18
18-FREQ(HZ): 2437000000.000
18-BAND(HZ): 98000000.000
18-Channel Power(dBm): -78.016
19-Channel NO: 19
19-FREQ(HZ): 2559000000.000
19-BAND(HZ): 103000000.000
19-Channel Power(dBm): -68.656
20-Channel NO: 20
20-FREQ(HZ): 5800000000.000
20-BAND(HZ): 232000000.000
20-Channel Power(dBm): -70.605
</description>
<Point><coordinates>-67.151245,18.172382,5.800</coordinates></Point>
</Placemark>

```

Ilustración 4.4. Fichero KML con los datos obtenidos

4.2. Volcado de los datos en una tabla

Como resultado de la Sección 4.1. tenemos un fichero con extensión KML que es fácil de usar y presentar empleando software de programas tales como *Google Earth*. En este proyecto, al manejar MATLAB, se hace evidente, por tanto, la necesidad de generar un código que permita volcar los datos del fichero original en una tabla con la que poder trabajar usando esta herramienta.

Como se puede apreciar en la ilustración anterior, un fichero KML contiene una estructura basada en “etiquetas”, dentro de las cuáles destacan:

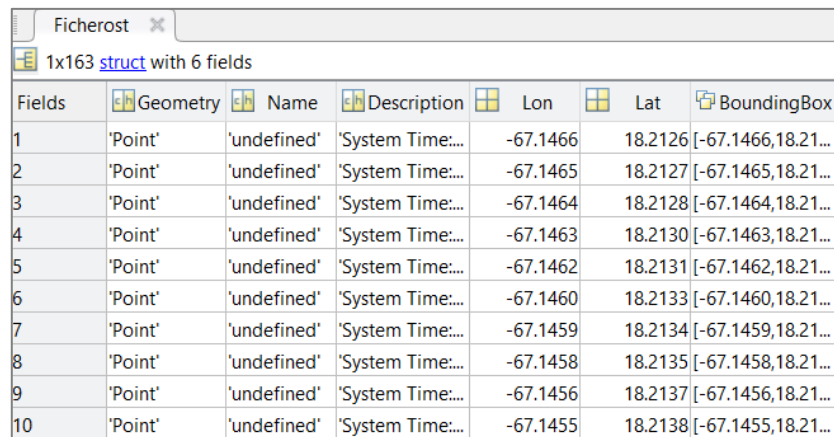
- **<name></name>**: se utiliza para proporcionar un nombre a la variable estudiada.
- **<description></description>**: se emplea para introducir una descripción básica del lugar. En este caso, en este campo es en el que se van a incluir todas las medidas realizadas en cada punto del recorrido.
- **<coordinates></coordinates>**: este campo indica la latitud y longitud de cada punto que se analiza.
- **<Geometry></Geometry>**: indica el tipo de lugar sobre el que se mide, si es un punto, un área concreta, etc.

Aprovechando esta estructura que poseen los archivos KML, se vuelcan los datos en una variable de tipo *struct* que va a contener tantos campos como “etiquetas” diferentes tenga el fichero de datos. Para crear esta variable,

se llama a la función *kml2struct* [27] a través de la siguiente sentencia, en la que se le pasa como parámetro la ubicación del fichero de datos. En este caso, el fichero recoge los datos de una de las Calles, la número 3, y será el que sirva de ejemplo, también, en los apartados posteriores.

```
Ficherost = kml2struct('Streets_Roads\STREETS_3\ST_3_1.kml');
```

Como resultado de esta sentencia se genera la variable *Ficherost*, cuyo aspecto se puede observar en la Ilustración 4.5. Se muestran concretamente las 10 primeras filas, que corresponden a 10 puntos diferentes sobre los que se mide la potencia recibida.



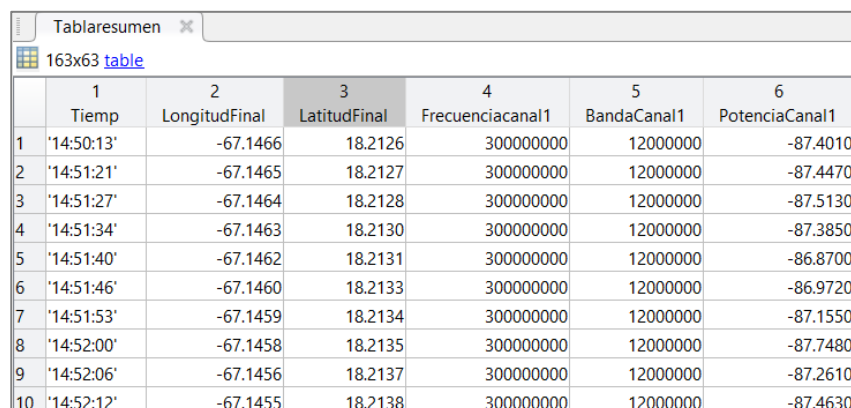
Fields	Geometry	Name	Description	Lon	Lat	BoundingBox
1	'Point'	'undefined'	'System Time:...	-67.1466	18.2126	[-67.1466,18.21...
2	'Point'	'undefined'	'System Time:...	-67.1465	18.2127	[-67.1465,18.21...
3	'Point'	'undefined'	'System Time:...	-67.1464	18.2128	[-67.1464,18.21...
4	'Point'	'undefined'	'System Time:...	-67.1463	18.2130	[-67.1463,18.21...
5	'Point'	'undefined'	'System Time:...	-67.1462	18.2131	[-67.1462,18.21...
6	'Point'	'undefined'	'System Time:...	-67.1460	18.2133	[-67.1460,18.21...
7	'Point'	'undefined'	'System Time:...	-67.1459	18.2134	[-67.1459,18.21...
8	'Point'	'undefined'	'System Time:...	-67.1458	18.2135	[-67.1458,18.21...
9	'Point'	'undefined'	'System Time:...	-67.1456	18.2137	[-67.1456,18.21...
10	'Point'	'undefined'	'System Time:...	-67.1455	18.2138	[-67.1455,18.21...

Ilustración 4.5. Variable que contiene los datos de entrada

Dentro de esta variable, interesa quedarse con el campo *Description* que, como se ha indicado antes, es el que contiene los datos de las medidas del estudio.

Utilizando sobre este campo funciones de manipulación de cadenas en MATLAB [26], tales como la función *split*, para separar cadenas a partir de un cierto delimitador, o la función *str2num*, para convertir las cadenas en variables numéricas de tipo *double*, creamos una variable de tipo *table* que, ahora sí, se puede usar y es perfectamente manejable por MATLAB.

Esta nueva tabla tendrá, por una parte, un número de filas equivalente al número de puntos, en la zona recorrida, sobre los que se mide la potencia recibida. Por otra parte, tendrá un número de columnas tal que se recoja: el instante temporal en el que se realiza la medida, la latitud y longitud del punto que se recorre en cada instante y los valores de frecuencia, ancho de banda y potencia recibida de cada canal empleado. Un ejemplo de variable de este tipo se muestra en la Ilustración 4.6.



	1	2	3	4	5	6
	Tiemp	LongitudFinal	LatitudFinal	Frecuenciacanal1	BandaCanal1	PotenciaCanal1
1	'14:50:13'	-67.1466	18.2126	300000000	12000000	-87.4010
2	'14:51:21'	-67.1465	18.2127	300000000	12000000	-87.4470
3	'14:51:27'	-67.1464	18.2128	300000000	12000000	-87.5130
4	'14:51:34'	-67.1463	18.2130	300000000	12000000	-87.3850
5	'14:51:40'	-67.1462	18.2131	300000000	12000000	-86.8700
6	'14:51:46'	-67.1460	18.2133	300000000	12000000	-86.9720
7	'14:51:53'	-67.1459	18.2134	300000000	12000000	-87.1550
8	'14:52:00'	-67.1458	18.2135	300000000	12000000	-87.7480
9	'14:52:06'	-67.1456	18.2137	300000000	12000000	-87.2610
10	'14:52:12'	-67.1455	18.2138	300000000	12000000	-87.4630

Ilustración 4.6. Muestra de la variable que contiene los datos a analizar

En la imagen anterior se muestran los 10 primeros puntos recorridos y cómo queda la tabla para el canal 1, aunque su formato es extensible para el resto de los canales. Como se puede observar, de un punto a otro, la frecuencia y el ancho de banda de cada canal no van a cambiar, lo que sí se verá alterado será el momento temporal de la medida, las coordenadas geográficas del punto y, también, el valor de potencia medido.

Esta tabla se podrá generar para cada zona deseada y será el punto de partida a la hora de elaborar las gráficas y

representaciones que permitan entender y calibrar el funcionamiento del sistema.

4.2.1. Creación de tablas auxiliares

Con el objetivo de facilitar la comprensión de este capítulo, en esta subsección se va a incluir una breve explicación de las matrices auxiliares que se emplearán en futuras representaciones. Como es de esperar, dichas variables auxiliares derivarán de la tabla obtenida anteriormente en esta sección y lo que harán será englobar todos aquellos datos que sean del mismo tipo.

De esta manera existirá, por una parte, una variable (denominada *LatitudFinal*) que recoja los valores de la latitud para cada ubicación del espacio elegido y otra variable (denominada *LongitudFinal*) para los valores de la longitud. Así, se crean dos tablas auxiliares que tendrán tantas filas como puntos en los que se realice la medida y una sola columna con los respectivos valores. En la Ilustración 4.7 se muestra un ejemplo de ambas variables, para los 5 primeros puntos de la Calle bajo estudio número 3.

	1	2
1	18.2126	
2	18.2127	
3	18.2128	
4	18.2130	
5	18.2131	

	1	2
1	-67.1466	
2	-67.1465	
3	-67.1464	
4	-67.1463	
5	-67.1462	

Ilustración 4.7. Muestra de variables *LatitudFinal* y *LongitudFinal*

Asimismo, se construye otra nueva variable de tipo tabla (denominada *PotenciaGlobal*) que contiene los valores de la potencia recibida en cada punto y para cada canal. Esta tabla tendrá tantas filas como puntos en los que se realice la medida y un total de 20 columnas, cada una de las cuáles recoge la potencia recibida por su respectivo canal. En la Ilustración 4.8 se recoge una muestra de dicha variable. Igual que antes se recogen los 5 primeros puntos estudiados y los valores de potencia en dichos puntos para los canales del 1 al 10.

	1	2	3	4	5	6	7	8	9	10
1	-87.4010	-84.7350	-84.1380	-85.2030	-74.8560	-87.3450	-37.5630	-86.4380	-43.8360	-85.2220
2	-87.4470	-86.1640	-85.0760	-85.5600	-77.5070	-87.5480	-48.6770	-86.6530	-46.8130	-85.1510
3	-87.5130	-85.7050	-84.5160	-86.4920	-79.3510	-87.5100	-48.2960	-86.6600	-53.6950	-84.9670
4	-87.3850	-85.3720	-84.7840	-86.8070	-80.4310	-87.4760	-50.3160	-86.5080	-48.9430	-84.9510
5	-86.8700	-85.5940	-84.5920	-86.6040	-75.8970	-87.5030	-52.5810	-86.4150	-50.7670	-84.8960

Ilustración 4.8. Muestra de la variable *PotenciaGlobal*

Por último, se crea una variable, también de tipo tabla, denominada *FrecuenciasGlobal*, que recoge los valores de las frecuencias centrales de cada canal. Igualmente, el número de filas se corresponderá con el número de puntos a analizar y tendrá 20 columnas, cada una de ellas recogiendo el valor de frecuencia del canal correspondiente. A modo de ejemplo, la Ilustración 4.9 muestra las frecuencias de los 10 primeros canales.

FrecuenciasGlobal										
163x20 cell										
	1	2	3	4	5	6	7	8	9	10
1	300000000	325000000	333000000	375000000	525000000	601000000	634000000	708000000	735000000	835000000
2	300000000	325000000	333000000	375000000	525000000	601000000	634000000	708000000	735000000	835000000
3	300000000	325000000	333000000	375000000	525000000	601000000	634000000	708000000	735000000	835000000
4	300000000	325000000	333000000	375000000	525000000	601000000	634000000	708000000	735000000	835000000
5	300000000	325000000	333000000	375000000	525000000	601000000	634000000	708000000	735000000	835000000

Ilustración 4.9. Muestra de la variable FrecuenciasGlobal

Todas estas nuevas variables recogerán los valores correspondientes a la zona que se esté analizando en cada momento: Avenida, Espacio Abierto o Calle, y servirán de base para realizar los distintos tipos de representaciones que se verán en la siguiente sección. Serán mencionadas en diversos lugares dentro de este documento, por lo que convendría volver a revisar este apartado siempre que fuera necesario.

4.3. Representaciones gráficas

En esta sección se van a estudiar las distintas representaciones que se pueden hacer de los datos recogidos y la información que pueden proporcionar.

En este sentido, para mostrar con mayor claridad los resultados, nos vamos a servir de los datos de una de las zonas catalogadas como Espacios Abiertos, en este caso, el primero de ellos. Así, con el proceso descrito en la Sección 4.2. se obtiene una Tabla resumen con estos nuevos datos y se crean las variables LatitudFinal, LongitudFinal, PotenciaGlobal y FrecuenciasGlobal asociadas a los nuevos datos. De esta manera se tienen todos los datos ordenados convenientemente.

4.3.1. Gráficas dependientes del canal elegido

Como se ha podido intuir a lo largo del desarrollo de esta sección, el canal en el que se realiza la medida de potencia, cuyo valor está comprendido entre 1 y 20, es una variable que se puede elegir según la necesidad o el interés del usuario en cada momento.

En este apartado se va a explicar detenidamente qué tipo de representaciones se pueden realizar en función del canal específico que se quiera estudiar. Aleatoriamente, se va a seleccionar el canal 14 para realizar el análisis. De la misma manera, se seguirá con la observación del comportamiento del Espacio Abierto número 1. Para comenzar con las representaciones, en primer lugar, se encuentra la gráfica mostrada en la Ilustración 4.10.

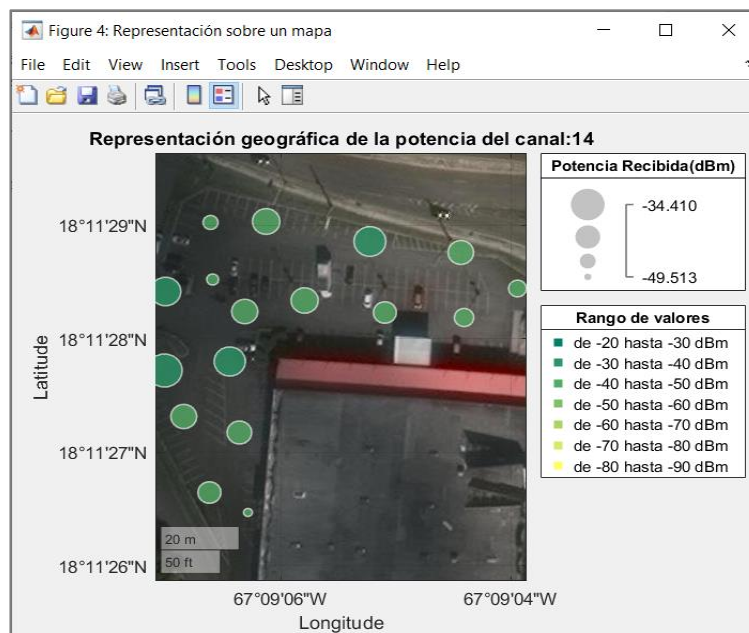


Ilustración 4.10. Representación geográfica

En esta figura se muestra la representación, sobre un mapa geográfico, del denominado Espacio Abierto número 1 que, en este caso, corresponde a un aparcamiento de un centro comercial. Las “burbujas” que se observan indican cada uno de los puntos de esta zona sobre los que se ha medido el nivel de potencia recibida y el tamaño de dichas “burbujas” será menor o mayor en función de si es un lugar donde se reciba más o menos potencia. De esta manera, los puntos que presenten una “burbuja” mayor serán aquellos en los que, en esta zona concreta bajo estudio y para el canal seleccionado, llegue una potencia mayor.

Otra manera similar de interpretar la información de la potencia recibida es haciendo uso de una escala de colores, tal y como muestra la leyenda situada en la parte derecha inferior de la imagen. Siguiendo esta leyenda, se puede ver cómo el color presente en cada “burbuja” indica el rango de valores entre los que se encuentra el nivel de potencia recibida en ese punto. Los colores más oscuros se corresponden con los valores más altos mientras que los colores más claros designan puntos en los que el nivel de potencia es sustancialmente menor.

Ahora, colocando el cursor sobre alguno de los puntos de la imagen, se pueden obtener los datos correspondientes a las coordenadas geográficas de esa ubicación, el nivel exacto de potencia recibida en ese punto y, como información adicional, el rango de valores entre los que se encuentra el nivel de potencia. Esto se puede comprobar visualizando la Ilustración 4.11.

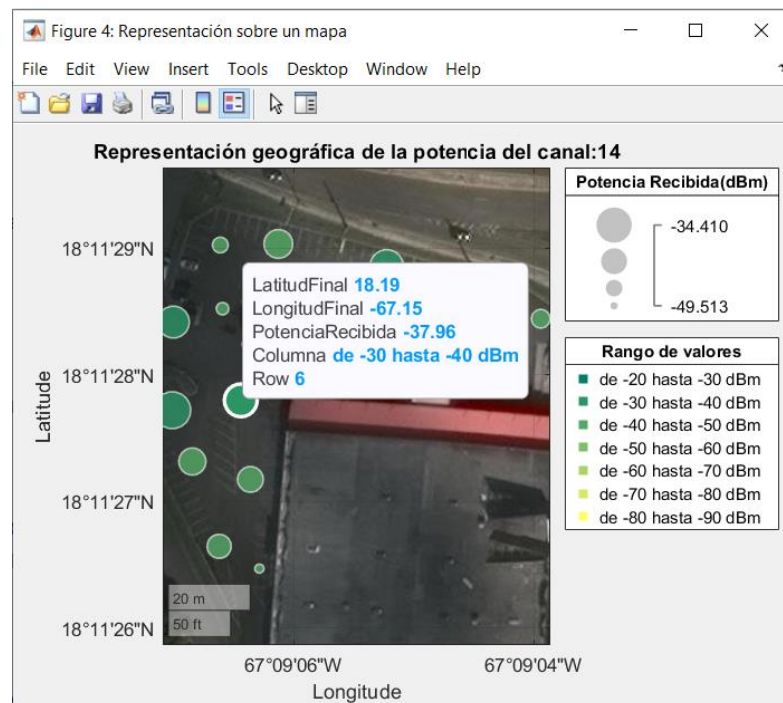


Ilustración 4.11. Selección de un punto para su análisis

Aparece, además de lo mencionado, el campo Row, que indica el número o la posición en la tabla de datos que le corresponde a ese punto en concreto dentro de la zona a analizar. Esto es, como el valor del campo Row del punto seleccionado es 6, esto indica que los datos de latitud, longitud y potencias recibidas en ese punto se corresponderán con la sexta fila de las matrices *LatitudFinal*, *LongitudFinal* y *PotenciaGlobal* respectivamente, como se veía en la Sección 4.2.1.

Como se verá más adelante, ese dato nos será de utilidad para otras posibles representaciones.

Para llevar a cabo la representación anterior se hace uso de la función *geobubble* de MATLAB [21], a la que se le pasan los valores de *LatitudFinal*, *LongitudFinal* y la columna 14 de la variable *PotenciaGlobal*, que es la que corresponde con los valores de potencia obtenidos para el canal 14. Se llama a la función mediante la siguiente sentencia [21]:

```
geobubble(Tabla, 'LatitudFinal', 'LongitudFinal',
          'SizeVariable', 'PotenciaRecibida');
```

Una vez detallada esta primera figura, hay una segunda representación que va a depender del canal elegido y es la mostrada en la Ilustración 4.12. Para su creación, se ha utilizado la función *contourf* de MATLAB [14],

pasándole como argumentos los valores de `LatitudFinal`, `LongitudFinal` y la columna 14 de la variable `PotenciaGlobal`.

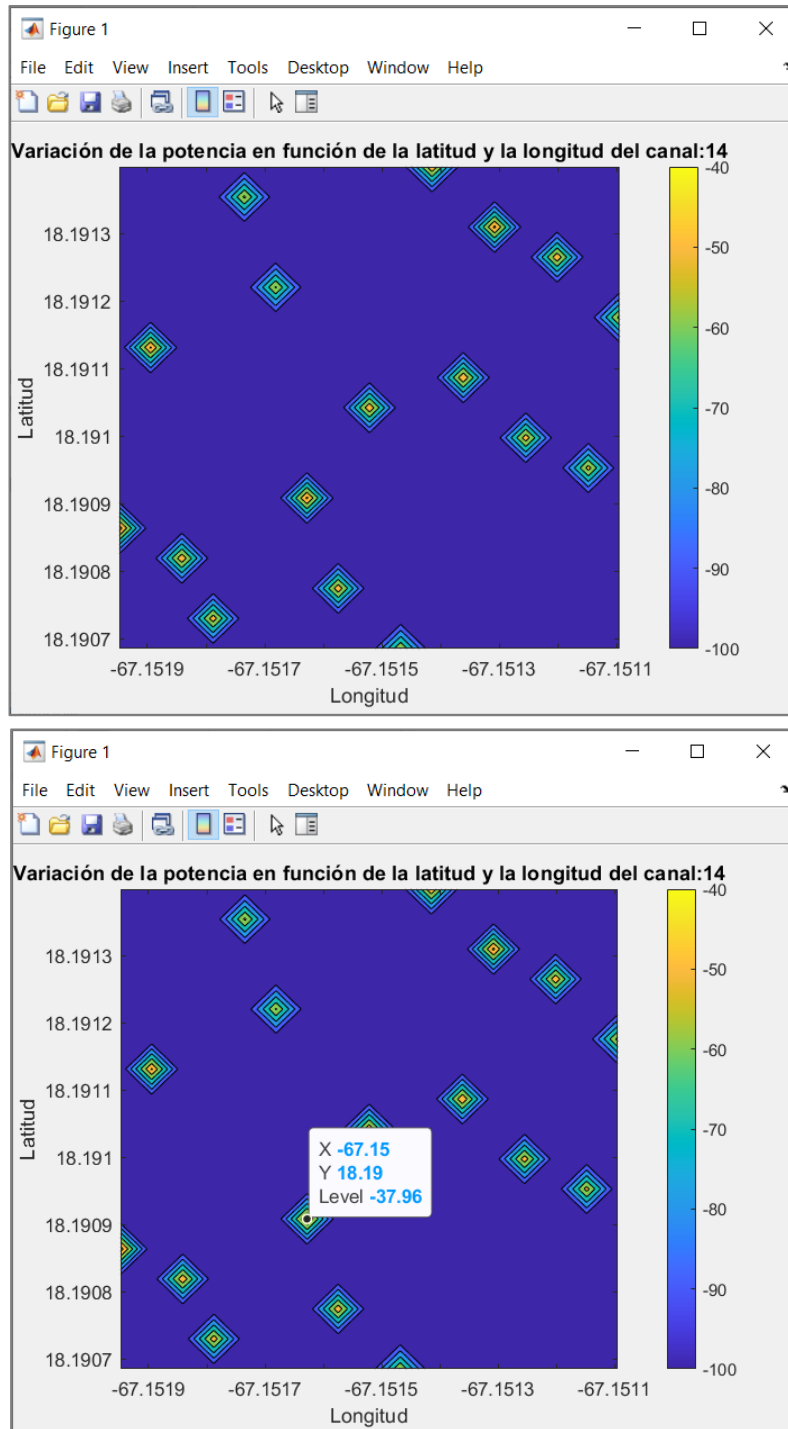


Ilustración 4.12. Variación de la potencia en función de la latitud y longitud

La imagen muestra, en 2D, cómo varía la potencia recibida, haciendo uso del canal 14, en cada una de las ubicaciones del aparcamiento. Cabe destacar cómo la distribución de los puntos en esta segunda gráfica es idéntica a la mostrada en las Ilustraciones 4.10 y 4.11. Este tipo de representaciones serán por tanto útiles o un buen recurso en el caso, por ejemplo, de que no se dispusiera, en el momento del análisis, de una herramienta que permitiera la representación sobre un mapa geográfico.

4.3.2. Gráficas dependientes de la ubicación elegida

Cada zona en la que se realiza el estudio (Avenida, Espacio Abierto o Calle) dispone de un número determinado de puntos en los que se va a medir la potencia recibida, y ese número no tendrá por qué ser igual en todas las zonas. Así, por ejemplo, en el estudio de una Calle el número de puntos en los que se mide es del orden de 100 mientras que existirán Espacios Abiertos en los que la cifra será de unas pocas decenas. Por tanto, al igual que el canal, la ubicación exacta, dentro de cada zona, es otra variable cuyo valor se puede elegir según la conveniencia del usuario en cada momento.

En este segundo apartado se expondrán las diferentes representaciones que se pueden realizar atendiendo a la ubicación exacta que se elija dentro de la zona estudiada. Por continuar de una manera coherente con la explicación, para las figuras de este apartado se seguirá utilizando como zona analizada el denominado Espacio Abierto número 1. En concreto, se va a observar el comportamiento del sistema para la ubicación número 6 dentro de esta zona, que se corresponde con la marcada en la Ilustración 4.11.

Teniendo en cuenta este criterio, para la ubicación 6 se puede obtener la información mostrada en la Ilustración 4.13.

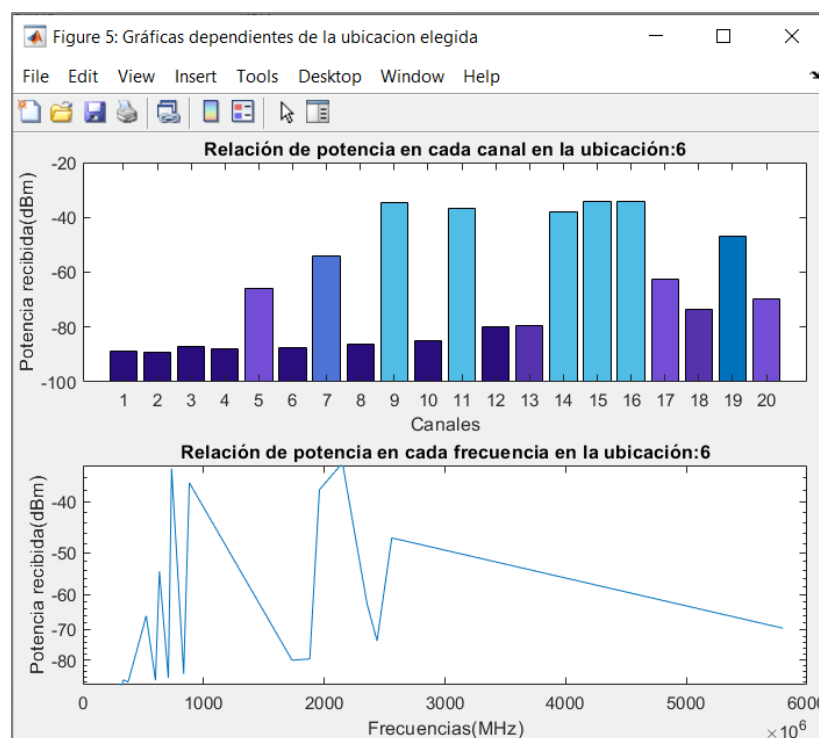


Ilustración 4.13. Representaciones dependientes de la ubicación escogida

En la imagen se pueden distinguir dos tipos de representaciones diferentes. La primera de ellas, situada en la parte superior de la figura, muestra, para la ubicación número 6, la potencia recibida en cada canal. Para ayudar a la lectura de los datos, se emplea un código de colores, de manera que se usan los colores más oscuros para señalar los canales a los que llega menos potencia, y los colores más claros para indicar los canales en los que la potencia recibida es superior. Así, se hace notar que, para ese punto, los canales idóneos para trabajar serán el canal 9, 11 y también los canales del 14 al 16, mientras que, por ejemplo, para los canales del 1 al 4 la potencia recibida es casi despreciable, del orden de -90 dBm.

Para realizar esta representación se emplea la función *bar* de MATLAB [17], pasándole como argumento los canales (enumerados del 1 al 20) y la fila de la matriz *PotenciaGlobal* que corresponde con la ubicación que queremos analizar, en este caso la número 6. La llamada a la función sigue el formato de la siguiente sentencia:

```
bar(Canales, cell2mat(PotenciaGlobal(UbicacionElegida,:)))
```

Por otro lado, en la parte inferior de la imagen, se puede observar una gráfica parecida a la anterior, pero creada con la función *semilogy* [19]. En este caso, lo que se representa es la cantidad de potencia recibida en función de la frecuencia que se utilice para medir. Los picos que se muestran en la gráfica corresponden con las mejores frecuencias de funcionamiento, donde el nivel de potencia recibida es mayor. De esta manera se puede concluir que las frecuencias óptimas son aquellas que tienen un valor alrededor de 1000 y 2200 MHz.

Si analizamos un punto concreto de la gráfica, se observa lo mostrado en la Ilustración 4.14.

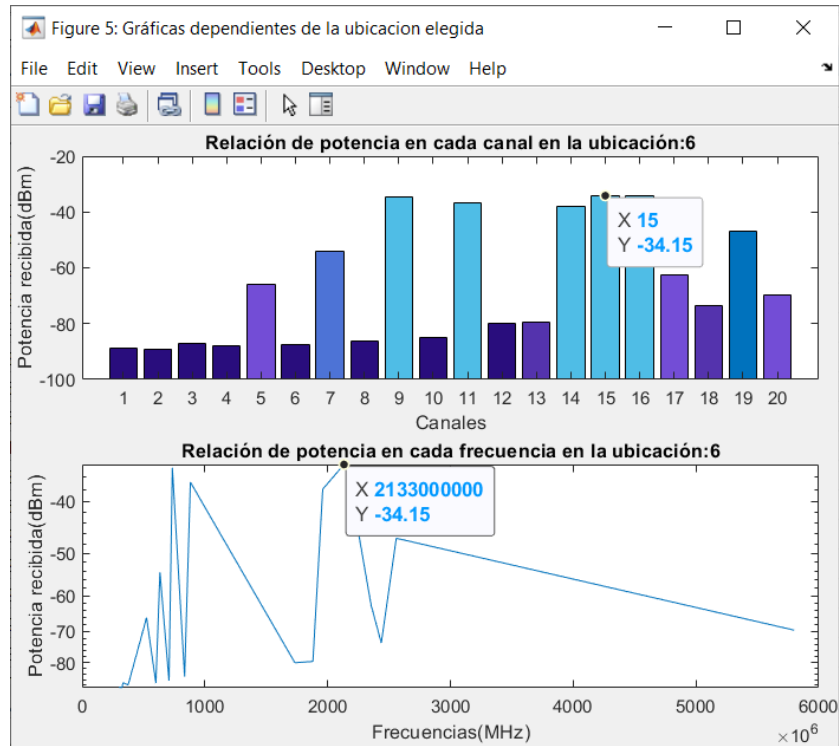


Ilustración 4.14. Selección de un punto para su análisis

En este caso se ha seleccionado, en la ubicación 6, el canal al que le llega mayor potencia, el canal 15 que se corresponde con una frecuencia de 2133 MHz.

4.3.3. Gráficas de carácter general

Se van a incluir dentro de esta subsección aquellas gráficas que recogen información global sobre cómo varía la potencia en cada uno de los canales y en cada una de las ubicaciones o puntos de la zona recorrida. Para continuar con la explicación se seguirá empleando como zona la denominada Espacio Abierto número 1.

En la Ilustración 4.15 se puede observar esa foto global.

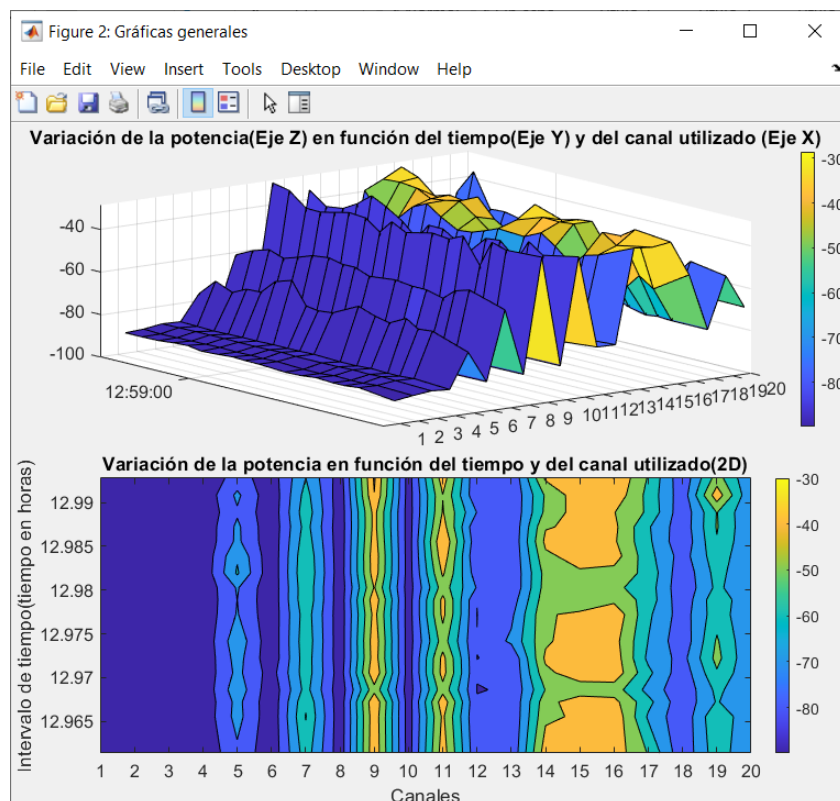


Ilustración 4.15. Gráficas de carácter general

Sobre la misma figura se incluyen dos representaciones diferentes, por lo que es necesario especificar cada una de ellas.

En la gráfica que se encuentra en la parte superior de la figura se observa que, en el eje X, se presentan cada uno de los canales empleados y, en el eje Y, los instantes de tiempo en los que se realiza la medida. Cada instante de tiempo representa un punto sobre el que se mide, punto que lleva asociadas una latitud y una longitud que se pueden conocer fácilmente consultando en la tabla de datos obtenida a partir del fichero original como se indica en el Apartado 4.2. El eje Z recoge el valor de potencia recibida en cada punto y para cada canal.

Como se puede observar, según el código de colores empleado, las zonas azules son las que presentan un nivel de potencia más bajo, mientras que, en las zonas de colores claros, la potencia recibida se hace notablemente mayor.

Para conseguir este resultado gráfico se hace uso de la función *surf* [13] de MATLAB mediante la siguiente sentencia:

```
surf(Canales, TiempoFinal, cell2mat(PotenciaGlobal));
```

Simplemente se le pasa como parámetros los canales y los instantes de tiempo y la matriz *PotenciaGlobal* que recoge las potencias recibidas para todos los canales en cada uno de los puntos.

La representación de la parte inferior es una adaptación en 2D de la gráfica situada en la parte superior. Igualmente, los ejes X e Y presentan los distintos canales e instantes de tiempo, respectivamente, y los valores de las alturas están condicionados por el nivel de potencia recibida en cada punto y para cada canal. Para la representación en 2D se puede hacer uso de la siguiente sentencia en MATLAB [14]:

```
contourf(Canales, Tiempoenhoras, cell2mat(PotenciaGlobal));
```

En este caso, los parámetros que le pasamos a la función son los mismos que para la representación anterior.

Colocando el cursor sobre algún punto de las gráficas, se pueden obtener el valor de potencia de ese punto para un canal específico. En la Ilustración 4.16 se muestra un ejemplo de este hecho.

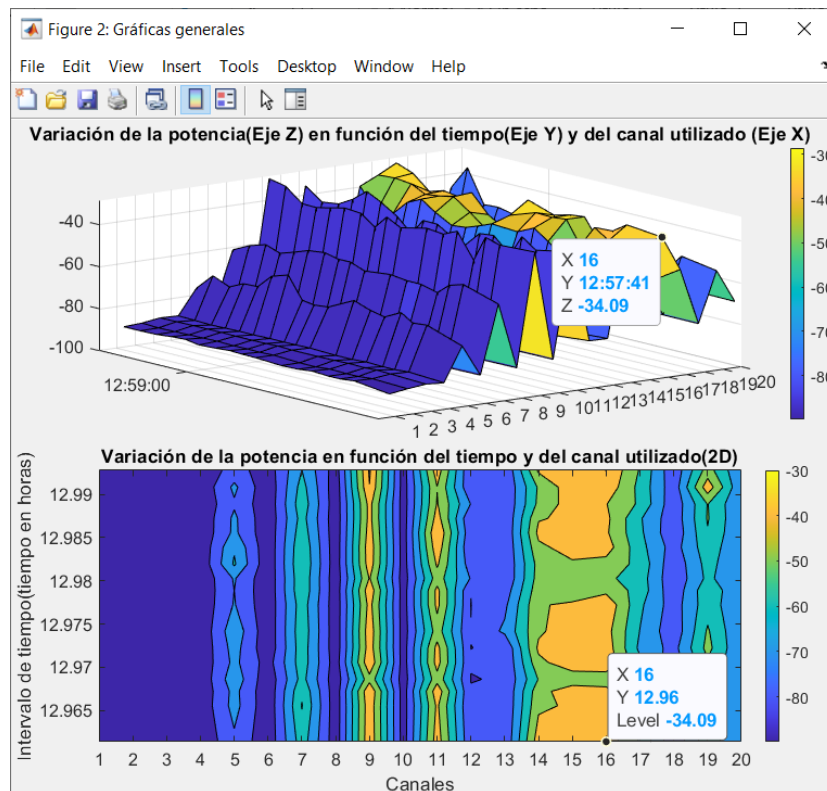


Ilustración 4.16. Visualización de potencia recibida en un punto concreto

Una posible lectura de la información es que, para el punto seleccionado, aquel en el que se tomaron las medidas a las 12:57:41, en el canal 16 se registra un valor de potencia de -34.09 dBm.

Otras conclusiones que se pueden sacar fácilmente es que, para la zona seleccionada, en este caso, el que hemos denominado Espacio Abierto número 1, los canales a los que les llega mayor potencia de señal son el 9, 11, 14, 15 y 16, que son los que, en la imagen inferior, presentan unos colores más claros, mientras que los canales del 1 al 4, el canal 6 y el 8 el nivel de señal es prácticamente despreciable, por lo que se podría considerar que sólo llega ruido.

Una segunda ilustración que podemos incluir en esta parte es aquella que representa, en cada punto, el promedio de la potencia que llega, esto es, en cada punto se muestra la media aritmética de todas las potencias recibidas por los 20 canales empleados. Para calcular el promedio se puede ejecutar la siguiente sentencia en MATLAB:

```
Promedio = mean(cell2mat(PotenciaGlobal),2);
```

La función *mean* así formulada va a devolver un vector columna que contiene la media de los elementos de cada fila de la matriz *PotenciaGlobal*. Cabe recordar en este punto que en dicha matriz cada fila representa una ubicación dentro de la zona que se recorre y cada columna, en total 20, contiene el valor de potencia recibido en ese punto para el canal correspondiente. Con la variable *Promedio* creada, los valores de *LatitudFinal* y *LongitudFinal* y haciendo uso de la función *contourf*, llegamos a la representación mostrada en la Ilustración 4.17.

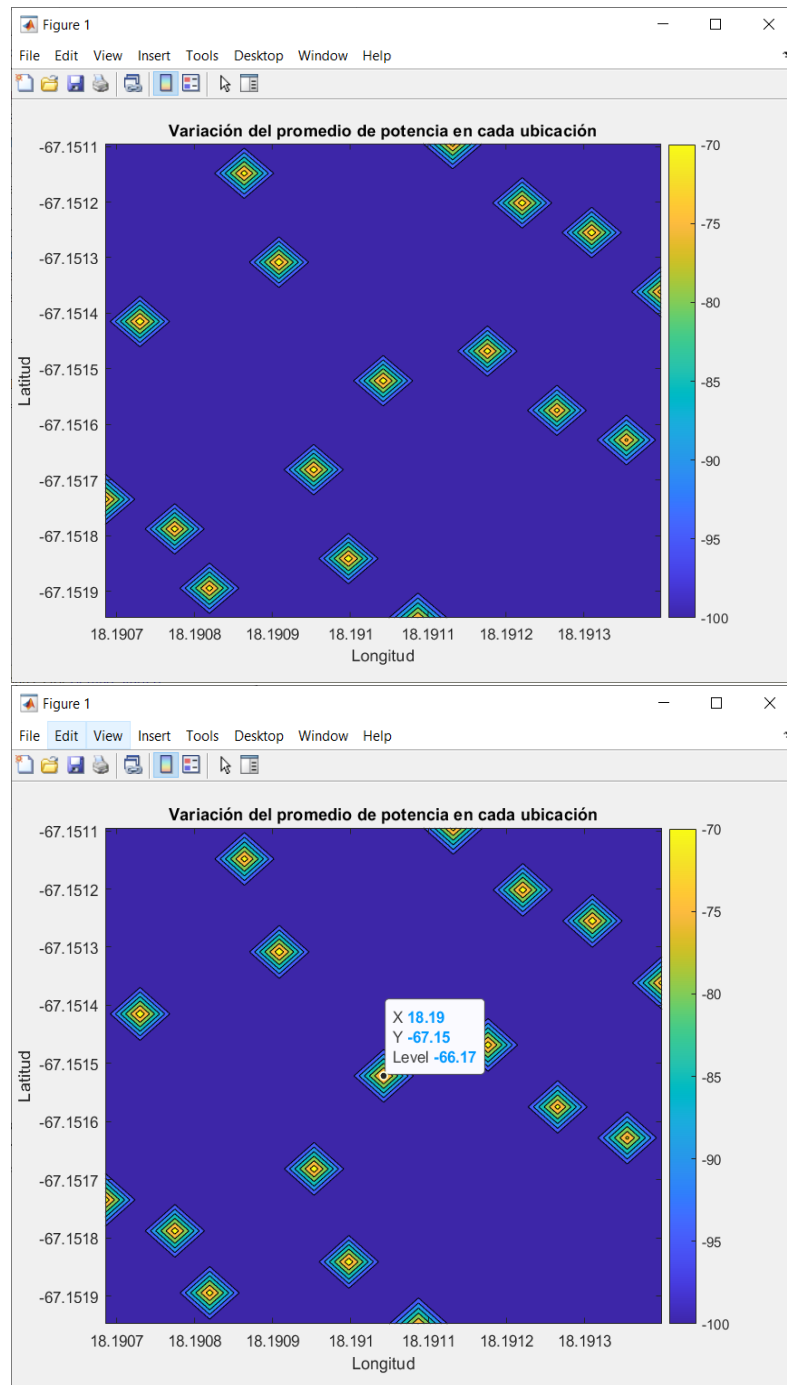


Ilustración 4.17. Variación del promedio de potencia

Como se puede observar, la disposición de los puntos de esta gráfica vuelve a coincidir con la de la gráfica que representa el Espacio Abierto sobre el mapa geográfico, vista por la Ilustración 4.10. Al colocar el cursor sobre cada punto se puede comprobar la potencia media que llega en él (etiqueta Level), así como sus coordenadas geográficas, dadas por los valores, en grados, de la Longitud (Eje X) y Latitud (Eje Y).

5. INTERFAZ GRÁFICA

En el capítulo anterior ya se ha visto cómo manejar los datos que se extraen al hacer el estudio y las diferentes representaciones que podemos hacer de ellos, junto con la información que éstas aportan.

En este nuevo capítulo se va a presentar la Interfaz Gráfica de Usuario (GUI), elaborada con el fin de facilitar al usuario la interacción con los datos y el análisis de estos. Las GUI permiten un control muy sencillo, prácticamente con la única necesidad de usar un ratón, de las aplicaciones software, ahorrándole al usuario el deber aprender un nuevo lenguaje de programación, que en ocasiones puede resultar tremendamente costoso.

MATLAB permite la creación de GUIs mediante el uso del comando *guide* [16], una herramienta cómoda de utilizar, muy intuitiva y que permite al desarrollador incluir en la interfaz gran cantidad de elementos interactivos tales como menús, barras de herramientas, controles deslizantes, botones, etc.

Para la generación de la Intrefaz Gráfica de este capítulo se han utilizado, de manera general, un total de 9 elementos, entre los que se encuentran botones, textos estáticos y ejes para insertar figuras. Insertar cada uno de estos elementos es tan sencillo como arrastrarlos desde un menu de selección hacia una página en blanco y depositarlos en el lugar deseado. En la Ilustración 5.1 se muestra este hecho [20].

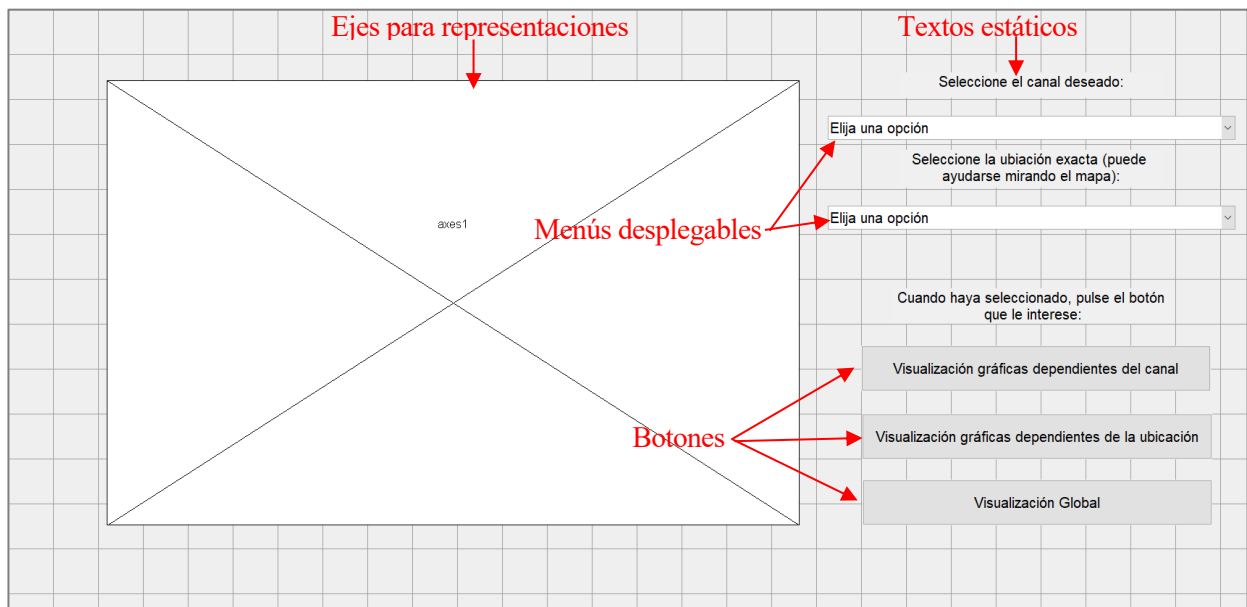


Ilustración 5.1. Elaboración de una GUI

5.1. Manual de Usuario

Una vez introducido el concepto de GUI, en esta sección se pretende mostrar, paso a paso, el recorrido que va a realizar el usuario desde que inicia la aplicación hasta que obtiene la información que necesita, procedente, entre otras fuentes, de las diferentes representaciones que se comentaron en la Sección 4.3 de este documento. Se intentará exponer de la manera más clara y sencilla posible, aportando gran cantidad de ejemplos ilustrativos.

Convendría volver a consultar la sección mencionada en el caso de no recordar la información que proporcionaba cada una de las representaciones allí explicadas y que volverán a aparecer a lo largo de este capítulo.

5.2.1. Menú Principal

Al arrancar por primera vez la aplicación, el usuario se va a encontrar con una interfaz denominada “MenúPrincipal”, mostrada en la Ilustración 5.2.

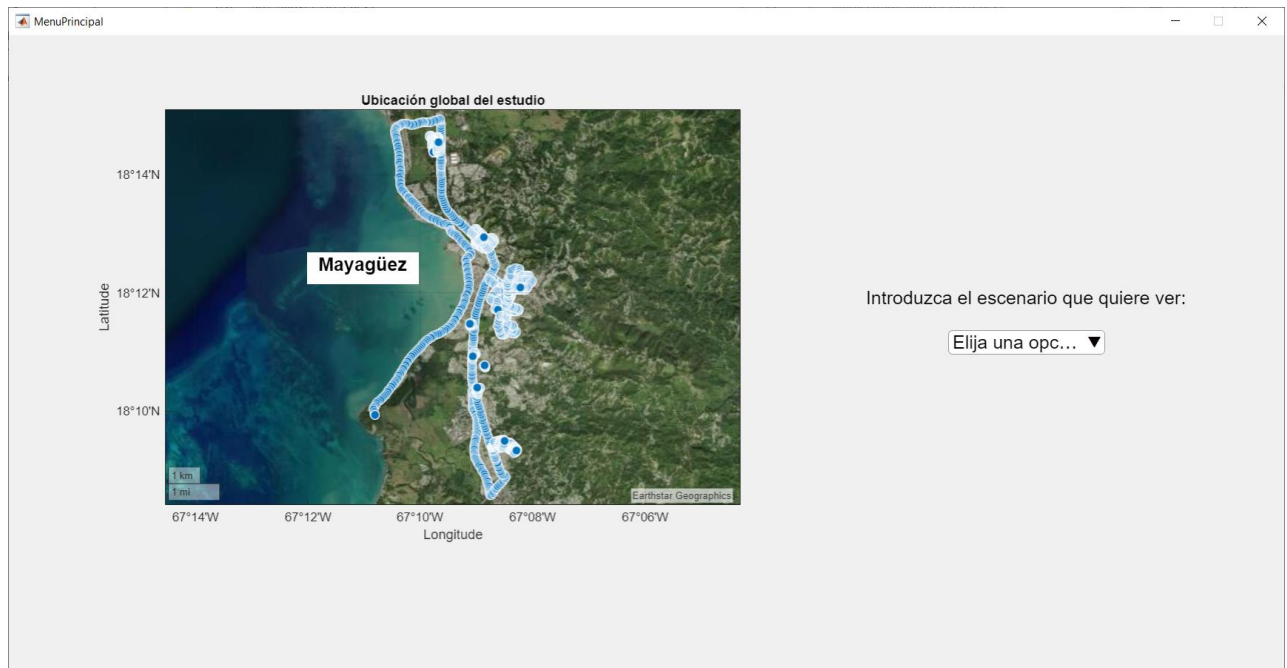


Ilustración 5.2. Menú Principal de la GUI

En la parte izquierda de la imagen se puede observar, sobre un mapa geográfico, todas las zonas bajo estudio. Cada una de las “burbujas” suponen puntos sobre los que se realizan medidas de potencia en los respectivos canales explicados con anterioridad. La figura muestra, en definitiva, la globalidad de zonas que abarca el estudio realizado.

En la parte derecha de la imagen se presenta un menú desplegable en el que el usuario debe elegir qué zona concreta quiere analizar: la Avenida, los Espacios Abiertos o las Calles. Se puede ver el desplegable en la Ilustración 5.3.

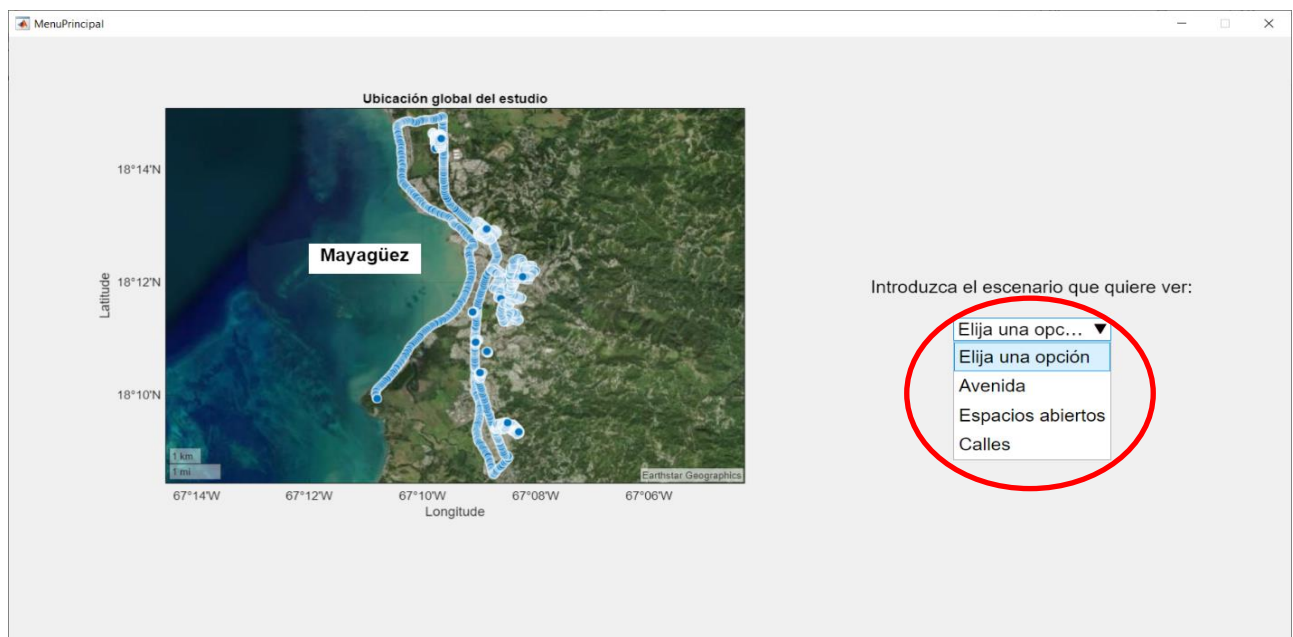


Ilustración 5.3. Menú desplegable con opciones de zona a elegir

Para poder analizar cada una de las zonas, el usuario debe seleccionarla en el menú desplegable. Si por error eligiera la opción etiquetada como “Elija una opción”, el sistema automáticamente le presentaría un mensaje de error como el que se muestra en la Ilustración 5.4.

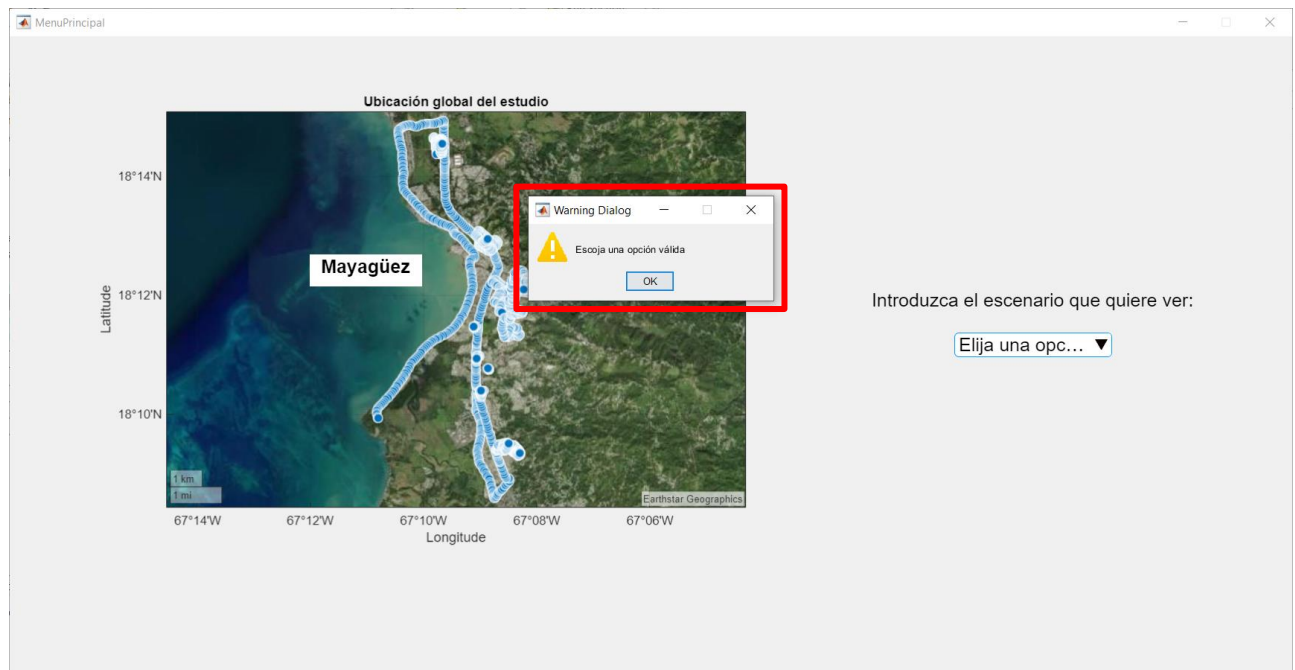


Ilustración 5.4. Mensaje de error en la opción introducida

En los siguientes apartados se verán las diferentes interfaces que se le presentan al usuario en función de la opción que elija del menú desplegable anterior.

5.2.1.1. Opción “Avenida”

En el caso de que el usuario seleccione la opción etiquetada como “Avenida” en el Menú Principal, se le redirigirá una nueva interfaz cuyo aspecto se presenta en la Ilustración 5.5.

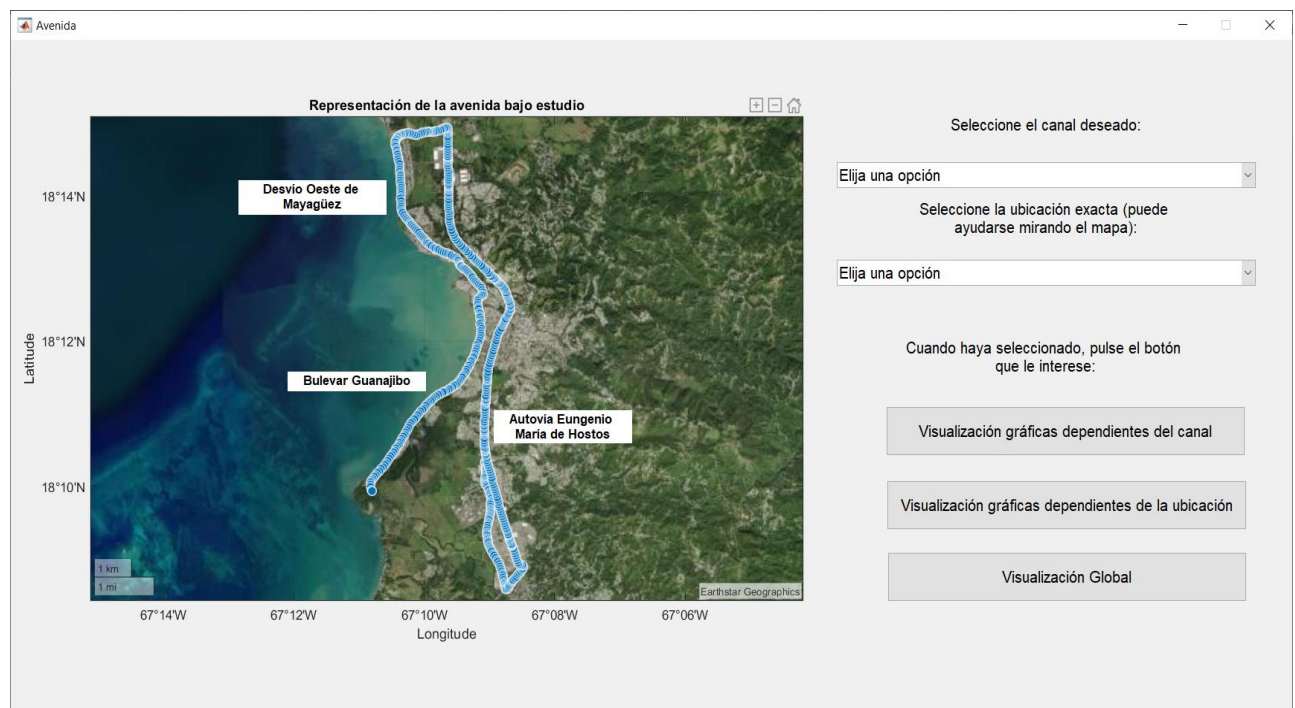


Ilustración 5.5. Interfaz gráfica correspondiente a la Avenida

Siguiendo el mismo formato que para el Menú Principal, en esta nueva interfaz se presenta, en lado izquierdo de la imagen, la representación geográfica de los diferentes puntos de la Avenida en los que se ha medido el nivel de potencia recibido.

En este caso, en la parte derecha, se dispone de 2 menús desplegables diferentes. El primero de ellos le servirá al usuario para seleccionar el canal que quiere analizar. Este desplegable se muestra en la Ilustración 5.6.

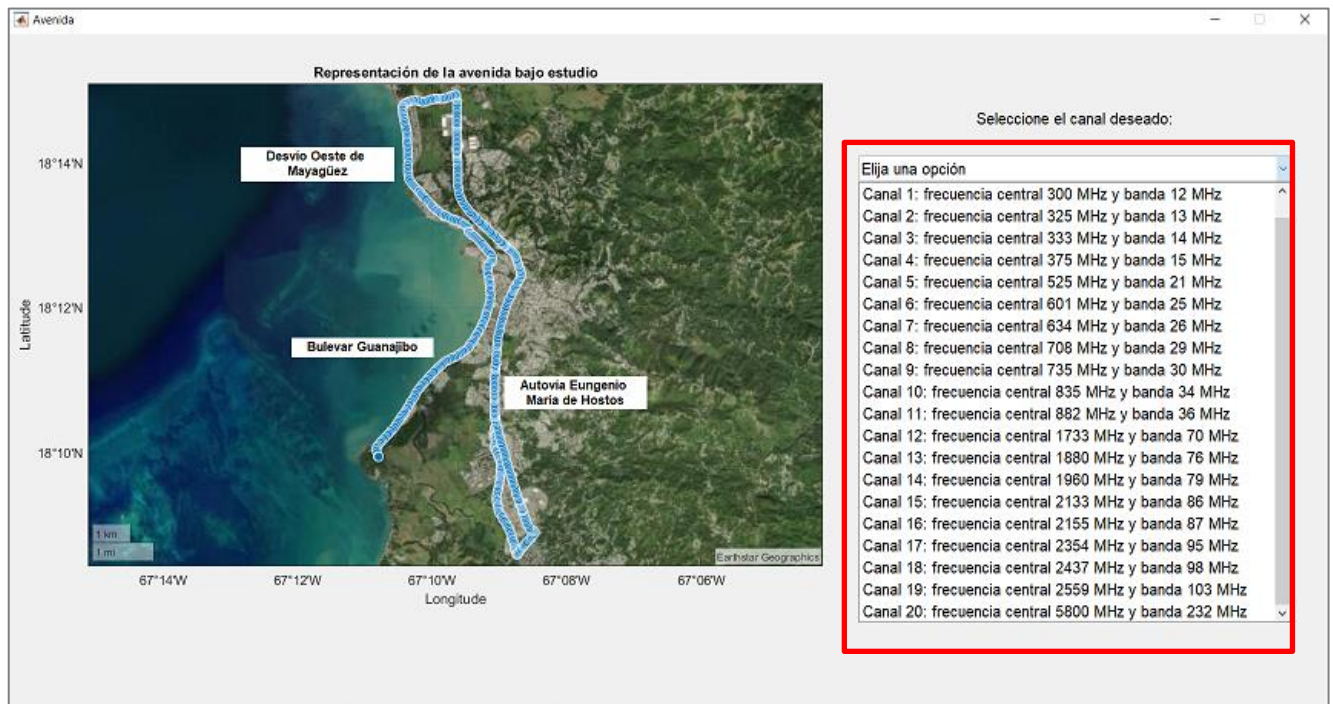


Ilustración 5.6. Menú desplegable para elegir el canal deseado

Una vez que el usuario elige el canal que quiere visualizar, tiene la opción de apretar el botón denominado “Visualización gráficas dependientes del canal”, mostrado en la parte inferior de la Ilustración 5.7. Se ha elegido aleatoriamente el canal 5 para esta imagen.

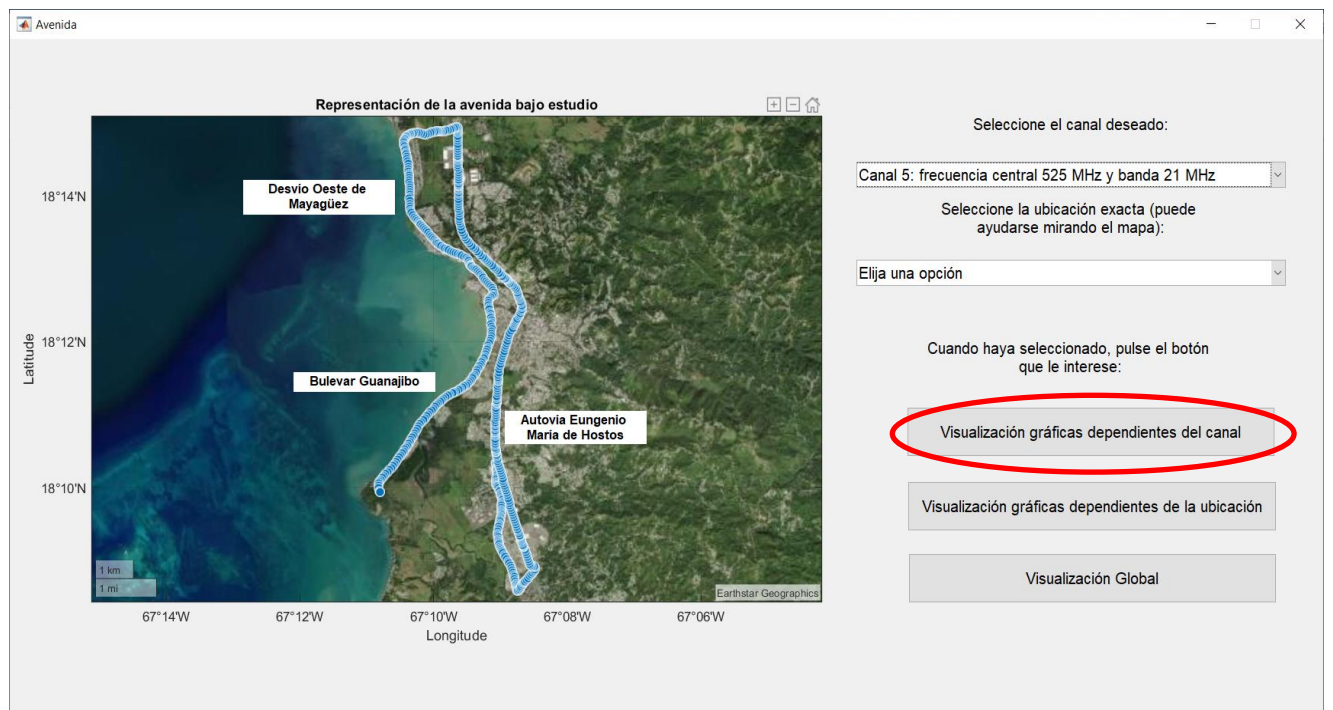


Ilustración 5.7. Selección del canal 5 y del botón “Visualización gráficas dependientes del canal”

A continuación de pulsar el botón mencionado, al usuario se le presentarán en pantalla dos figuras independientes, correspondientes a las representaciones explicadas en el Apartado 4.3.1. y que se muestran en la Ilustración 5.8.

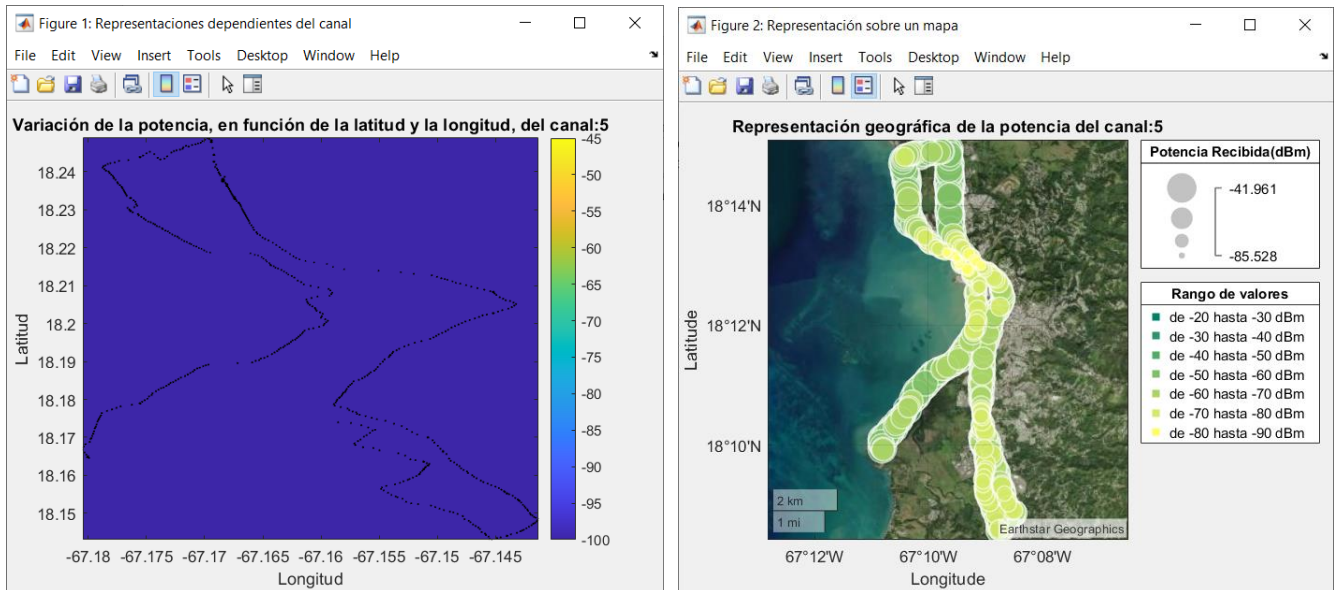


Ilustración 5.8. Representaciones dependientes del canal elegido

Con ayuda de estas gráficas se puede ver en qué puntos de la Avenida llega mayor nivel de potencia empleando el canal 5. Esos puntos son los que presentan, por ejemplo, un tamaño de “burbuja” mayor o un color más oscuro en la gráfica de la derecha.

El segundo menú desplegable, mostrado también en la Ilustración 5.5, es el que permite al usuario seleccionar la ubicación exacta que quiere consultar. La longitud de este menú dependerá del número de puntos sobre los que se haya medido en cada zona. Para el caso de la Avenida, hay un total de 531 ubicaciones posibles.

Para facilitar la elección de un punto exacto, puede ser de ayuda navegar sobre el mapa geográfico de la Avenida. Posando el cursor sobre cada uno de los puntos se puede observar cuál es su latitud y su longitud y también el número de ubicación que lleva asociado. A modo de ejemplo se presenta la Ilustración 5.9.

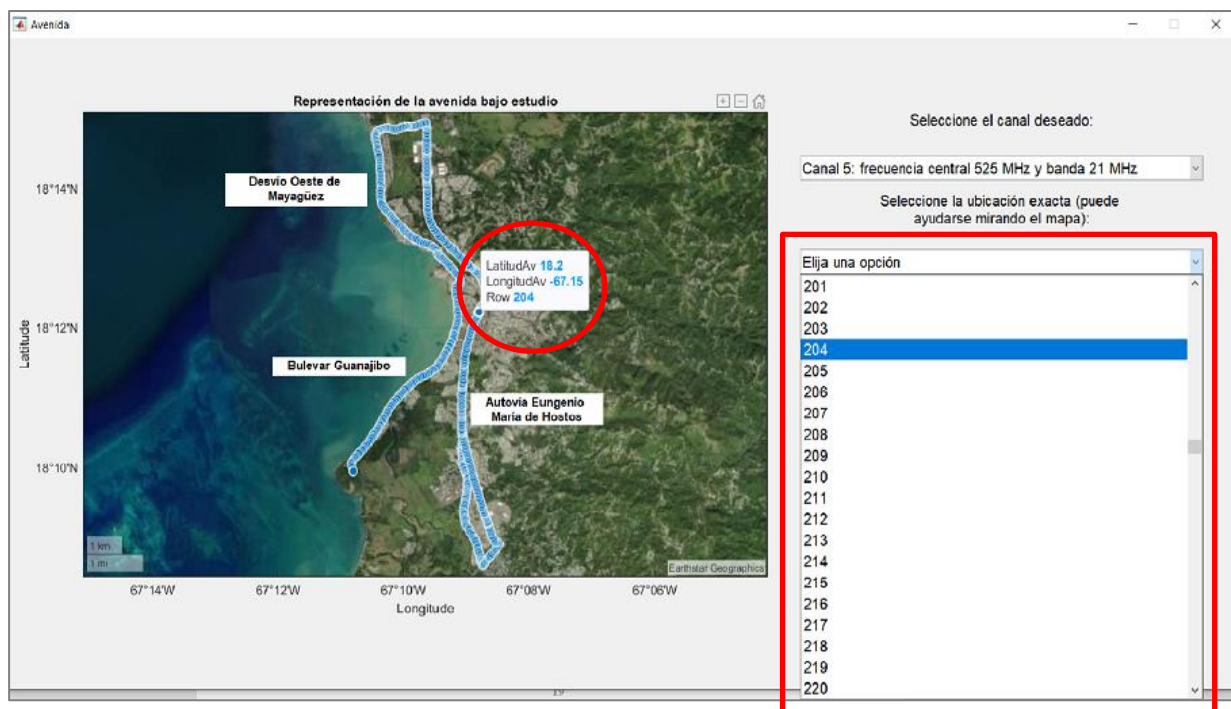


Ilustración 5.9. Menú desplegable para elegir ubicación deseada

Una vez que el usuario elige la ubicación que quiere analizar, tiene la opción de apretar el botón denominado “Visualización gráficas dependientes de la ubicación”, mostrado en la Ilustración 5.10. Se ha elegido aleatoriamente la ubicación 204 para esta imagen.

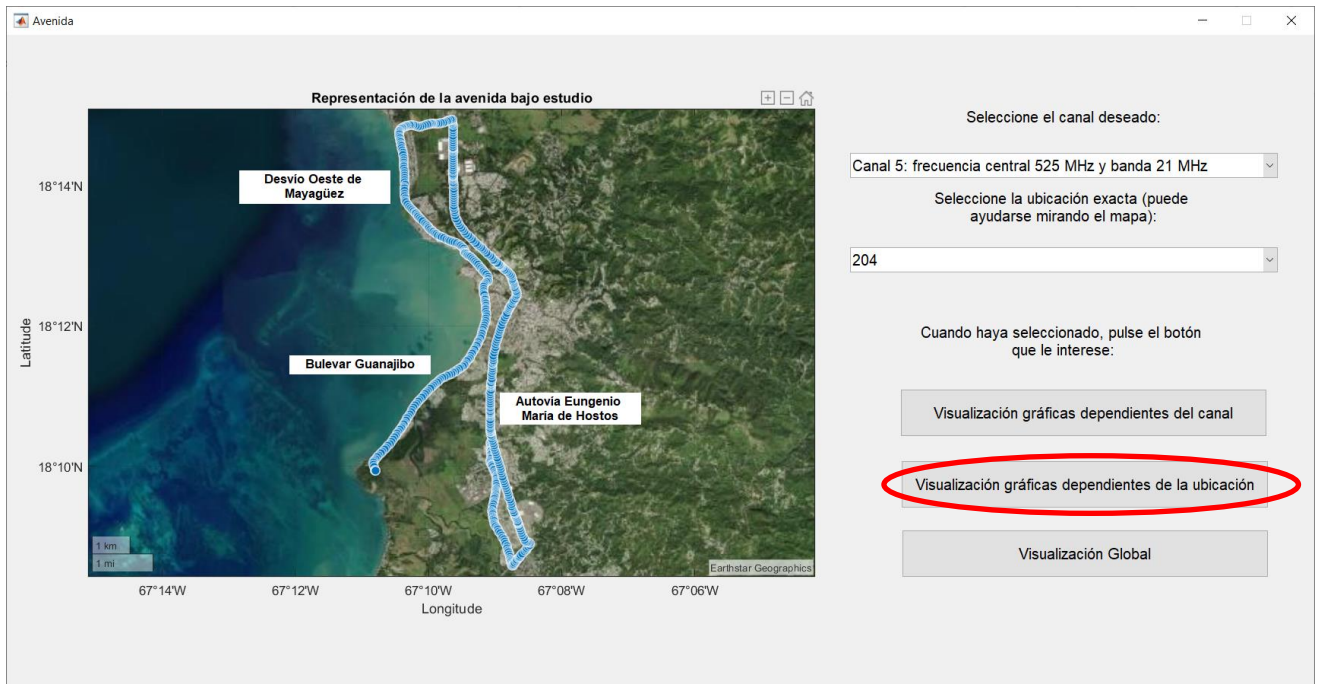


Ilustración 5.10. Selección de la ubicación 204 y del botón “Visualización gráficas dependientes de la ubicación”

Tras pulsar el botón mencionado, se presenta en pantalla la figura explicada en el Apartado 4.3.2 y que se muestra en la Ilustración 5.11.

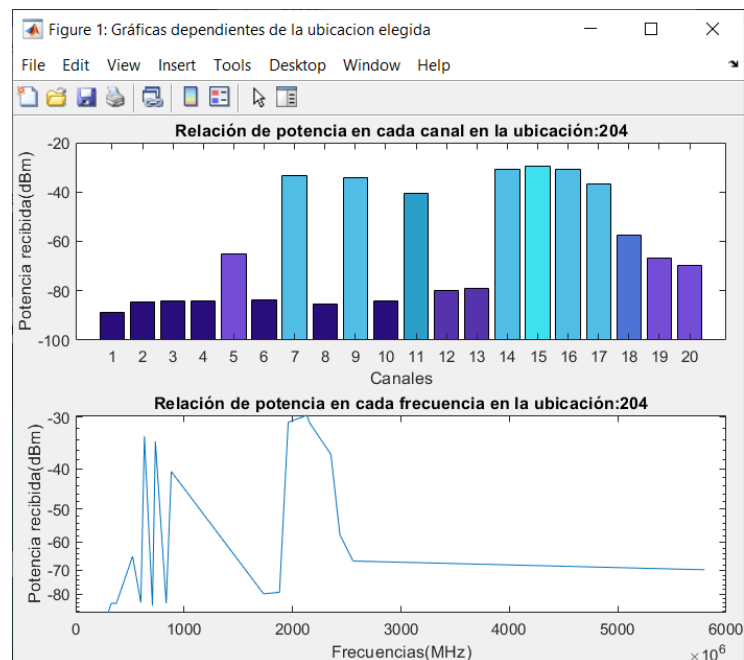


Ilustración 5.11. Representaciones dependientes de la ubicación elegida

Haciendo una lectura similar a la del Apartado 4.3.2 vemos que, para esta ubicación concreta, los canales idóneos para el funcionamiento del sistema son los canales 14, 15 y 16, que son los que reciben mayor potencia, y que corresponden con las frecuencias 1960, 2133 y 2155 MHz respectivamente.

Se ha analizado en estos dos casos anteriores la posibilidad de que el usuario elija un canal o una ubicación

concreta para ver. Este hecho facilita en gran medida el poder comparar el nivel de potencia recibido entre dos canales o dos ubicaciones diferentes, puesto que el número de gráficas a comparar es reducido.

Por último, solo faltaría por explicar el tercer botón existente, denominado “Visualización Global”, que se muestra en la Ilustración 5.12. Este botón permite, una vez escogidos un canal y una ubicación concretos, representar todas las gráficas juntas: las dependientes del canal, las dependientes de la ubicación y también se incluyen aquellas que se han explicado en el Apartado 4.3.3, las denominadas “Gráficas de carácter general”. Por tanto, para poder pulsar este botón el usuario debe haber escogido, previamente, valores tanto para el canal como para la ubicación, de lo contrario se mostrará un mensaje de error como el de la Ilustración 5.13.

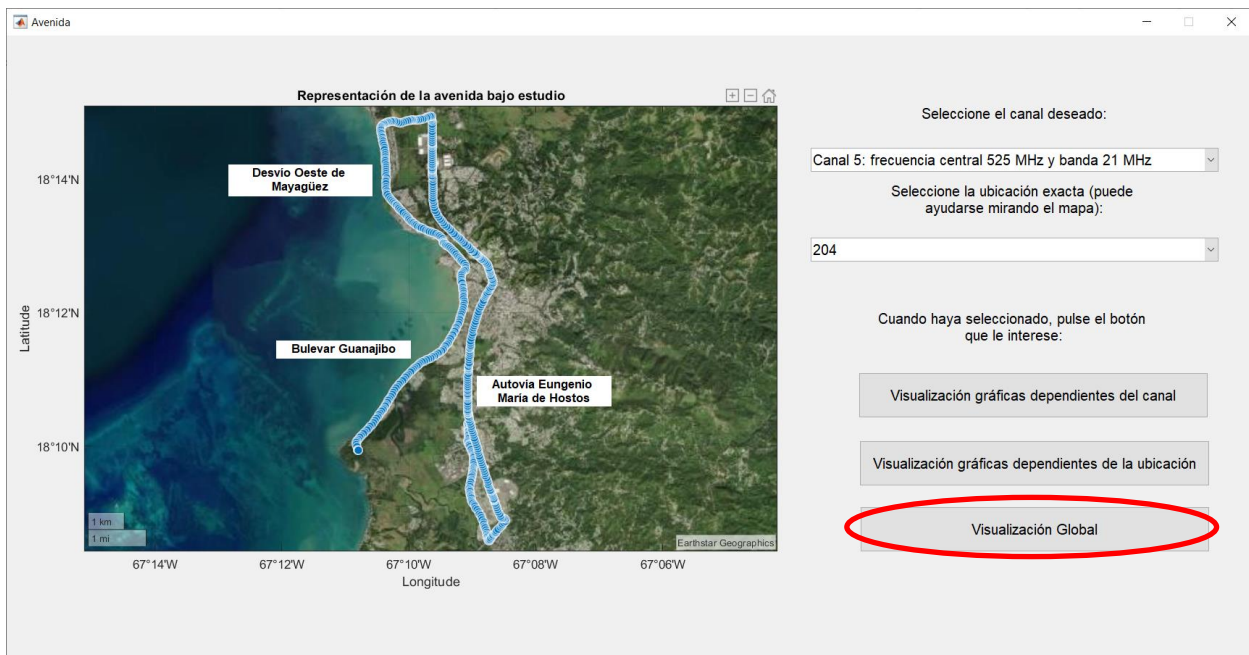


Ilustración 5.12. Selección del botón “Visualización Global”

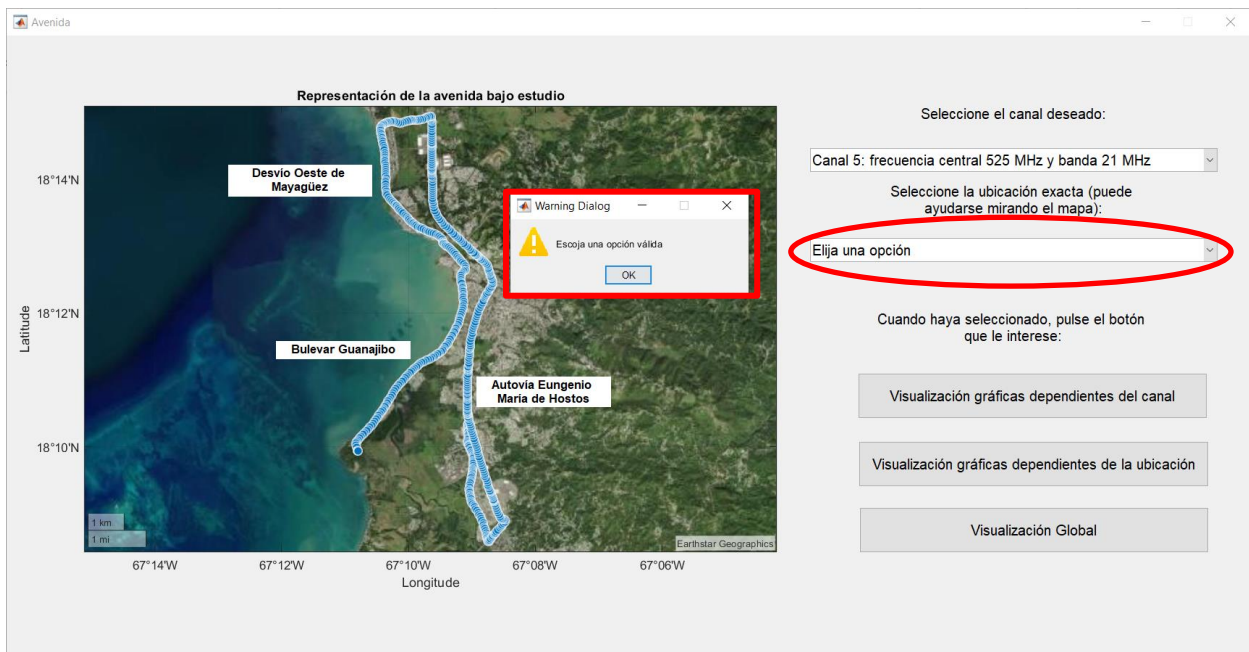


Ilustración 5.13. Mensaje de error al no seleccionar una ubicación válida

Si escoge valores válidos para el canal y la ubicación y pulsa el botón, se le presentarán por pantalla, además de las gráficas dependientes del canal y de la ubicación, las representaciones definidas como de carácter general en el Apartado 4.3.3. De esta manera, al pulsar el botón, se muestran las gráficas de la Ilustración 5.14.

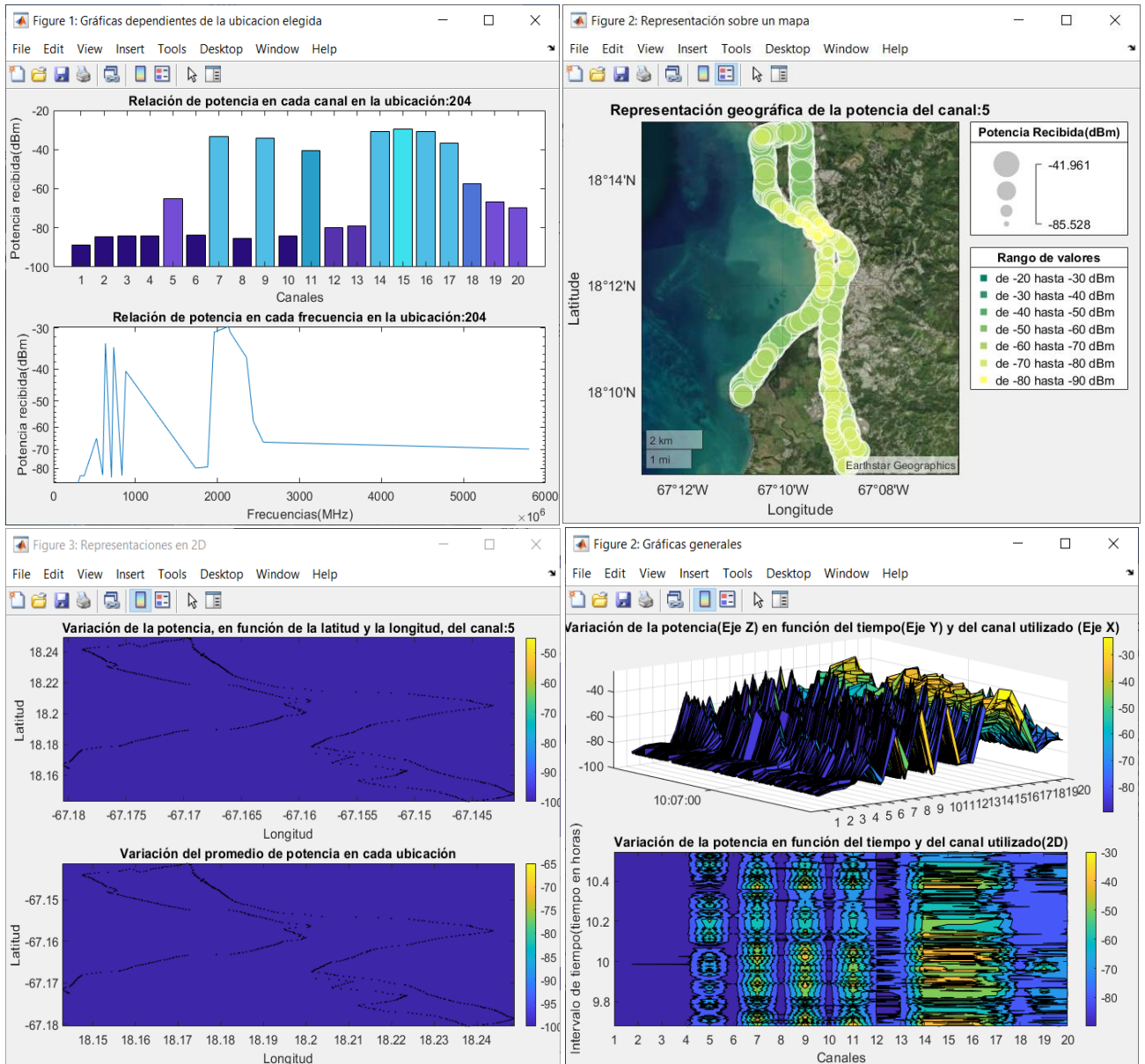


Ilustración 5.14. Representación global de la Avenida bajo estudio

Estas gráficas nos indican que, si por ejemplo queremos instalar en la Avenida un dispositivo que funcione para IoT, será necesario que opere en las frecuencias de los canales 7, 9, 11, 14 o 15, puesto que son a los que va a llegar un nivel de potencia mayor facilitando, por tanto, la alimentación de este dispositivo.

En definitiva, todas estas gráficas son las que un usuario puede obtener de una zona bajo estudio, en este caso, de la Avenida. En los próximos apartados se verá el comportamiento para la elección de los Espacios Abiertos o de las Calles.

5.2.1.2. Opción “Espacios Abiertos”

En el caso de que el usuario seleccione la opción “Espacios Abiertos” en el menú desplegable, se abrirá una nueva interfaz, que se muestra en la Ilustración 5.15.

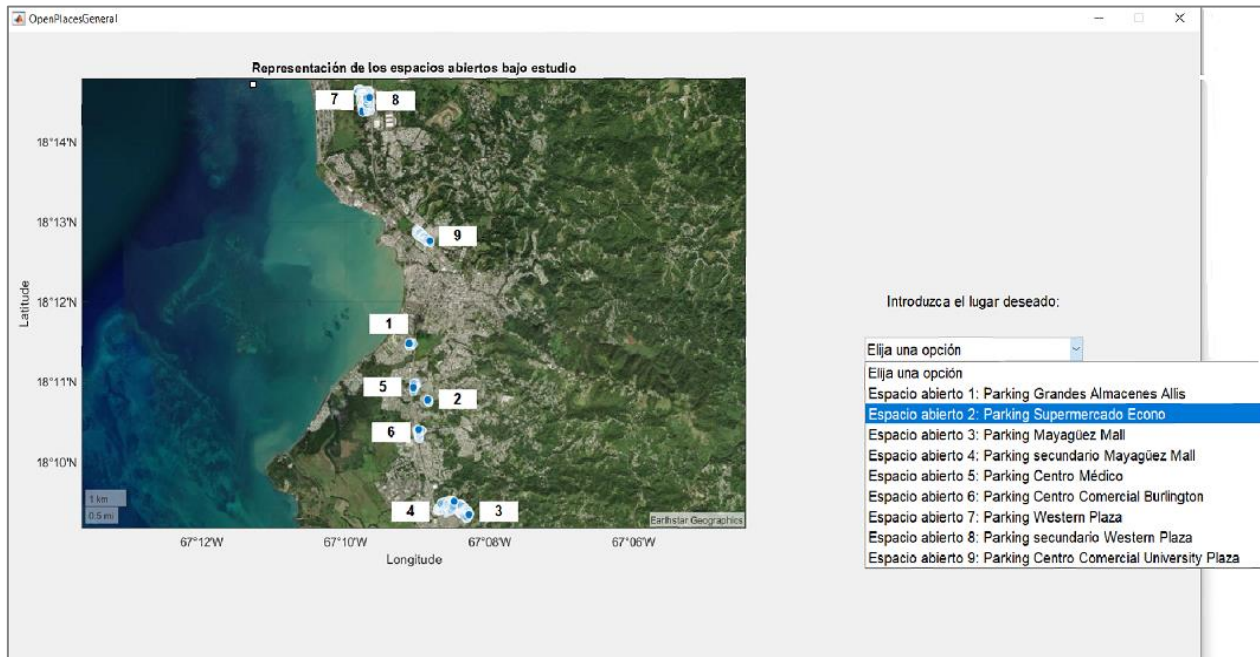


Ilustración 5.15. Menú General para la opción “Espacios Abiertos”

En la imagen anterior se muestra, en la parte izquierda, los diferentes lugares catalogados como Espacios Abiertos, enumerados del 1 al 9 para su fácil localización. Además, aparece en la parte derecha un nuevo menú desplegable en el que el usuario puede seleccionar qué Espacio Abierto quiere analizar concretamente. A modo de ejemplo se va a seleccionar el Espacio Abierto número 2, correspondiente a un aparcamiento de un supermercado. Al seleccionar la opción correspondiente se abre una nueva interfaz, mostrada en la Ilustración 5.16.

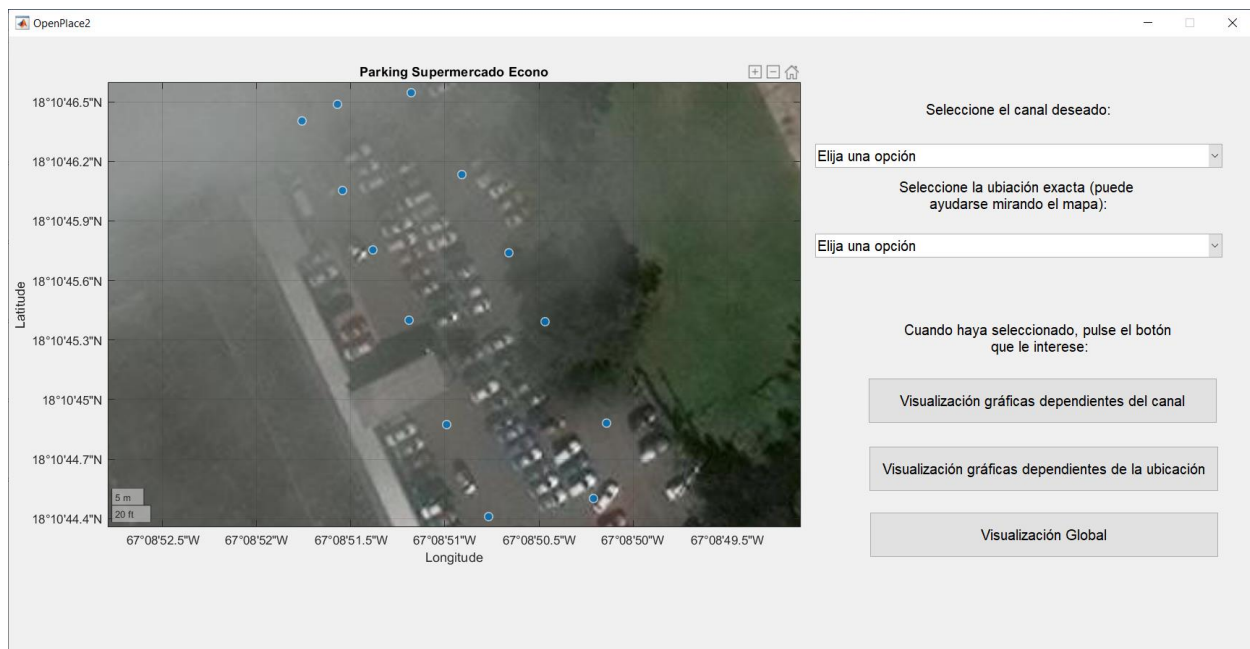


Ilustración 5.16. Interfaz gráfica “Espacio Abierto número 2”

En esta nueva interfaz lo que se muestra es una imagen satélite del Espacio Abierto seleccionado. Las “burbujas” que se observan en la imagen se corresponden con los diferentes puntos en los que se ha medido la potencia recibida. Como se puede observar también, posee los mismos menús desplegables y botones que la interfaz correspondiente a la Avenida. Como ya se ha explicado detenidamente el funcionamiento de cada botón y menú desplegable, no se va a repetir en este apartado. Simplemente se indicarán los pasos para llegar a las representaciones finales.

De esta manera, se selecciona en este caso, el canal 1 (Ilustración 5.17) y se pulsa sobre el botón “Visualización dependientes del canal”.

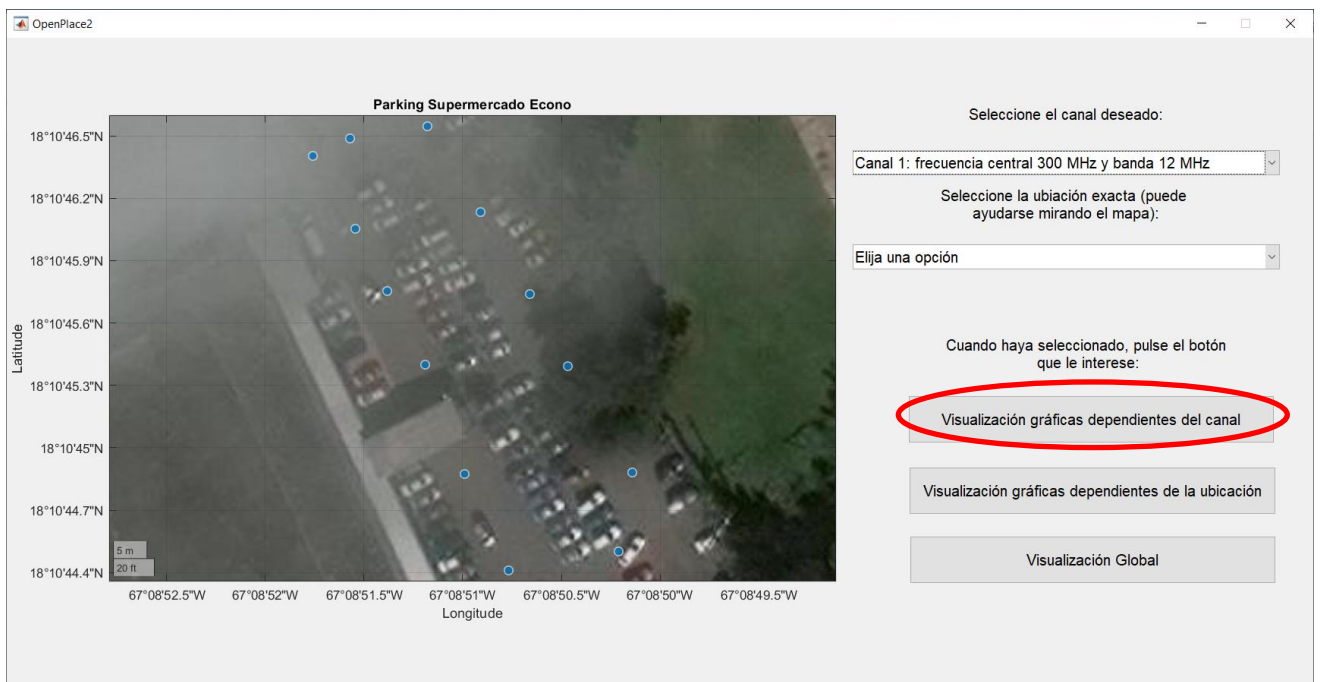
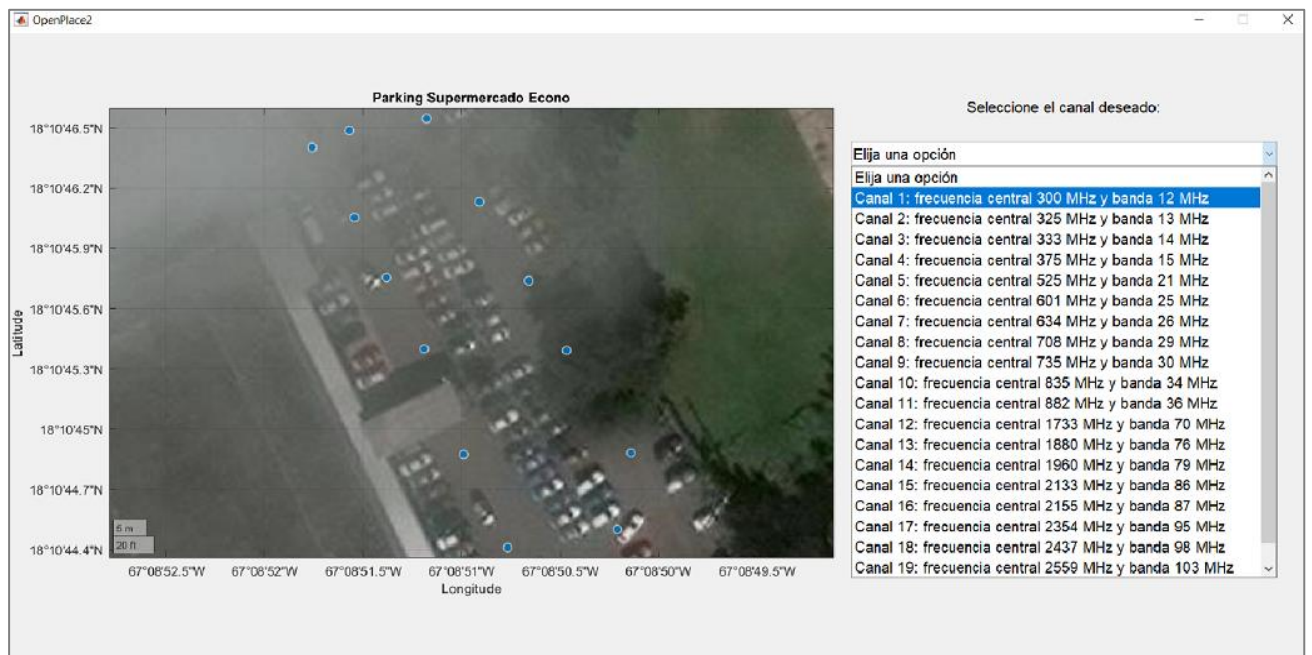


Ilustración 5.17. Proceso de selección del canal 1 para el “Espacio Abierto número 2”

Tras realizar estas acciones, el usuario observará las gráficas correspondientes a la Ilustración 5.18.

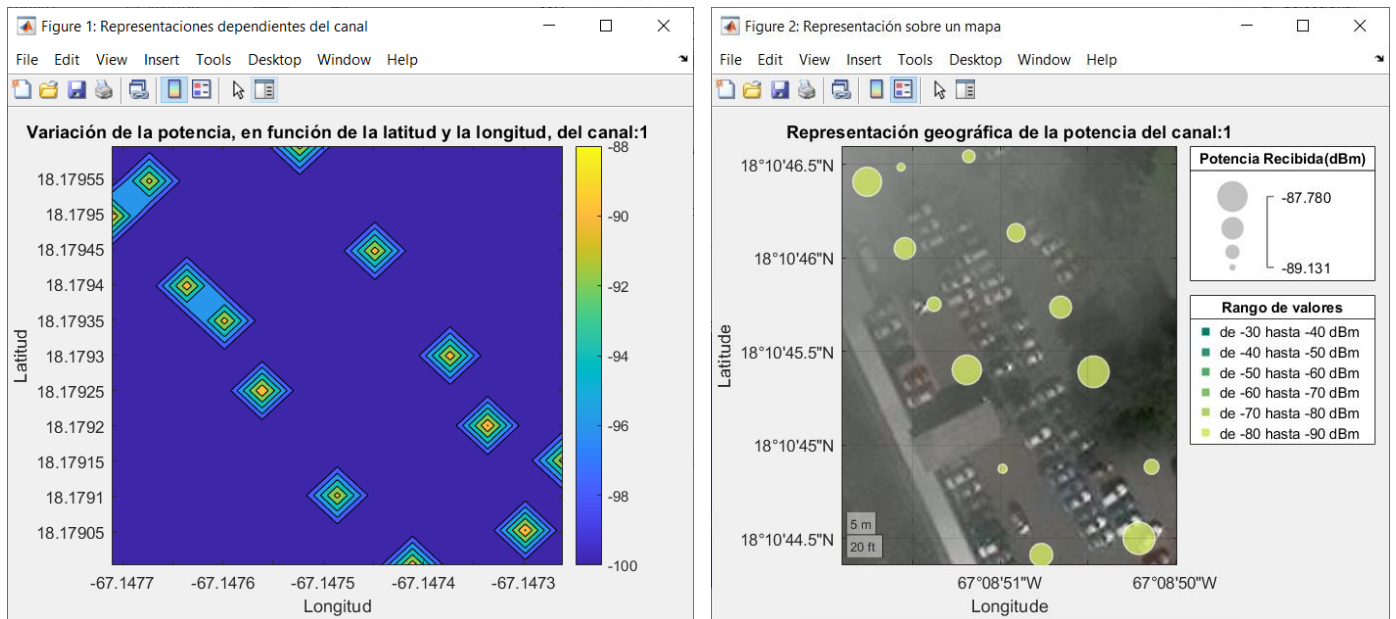


Ilustración 5.18. Gráficas del canal 1 para el “Espacio Abierto número 2”

En este caso, se puede observar que, en todos los puntos, la potencia recibida va a estar comprendida entre -87 y -90 dBm, valores relativamente bajos y que corresponden prácticamente a ruido. Aún así, habrá ubicaciones en las que llegue más potencia que otras. Estas serán aquellas que presenten un tamaño de “burbuja” mayor, como ya se ha explicado.

Repitiendo el mismo proceso para el menú desplegable de la ubicación, elegimos aleatoriamente la ubicación 4, de las 13 posibles, y se pulsa el botón “Visualización dependiente de la ubicación” (Ilustración 5.19).

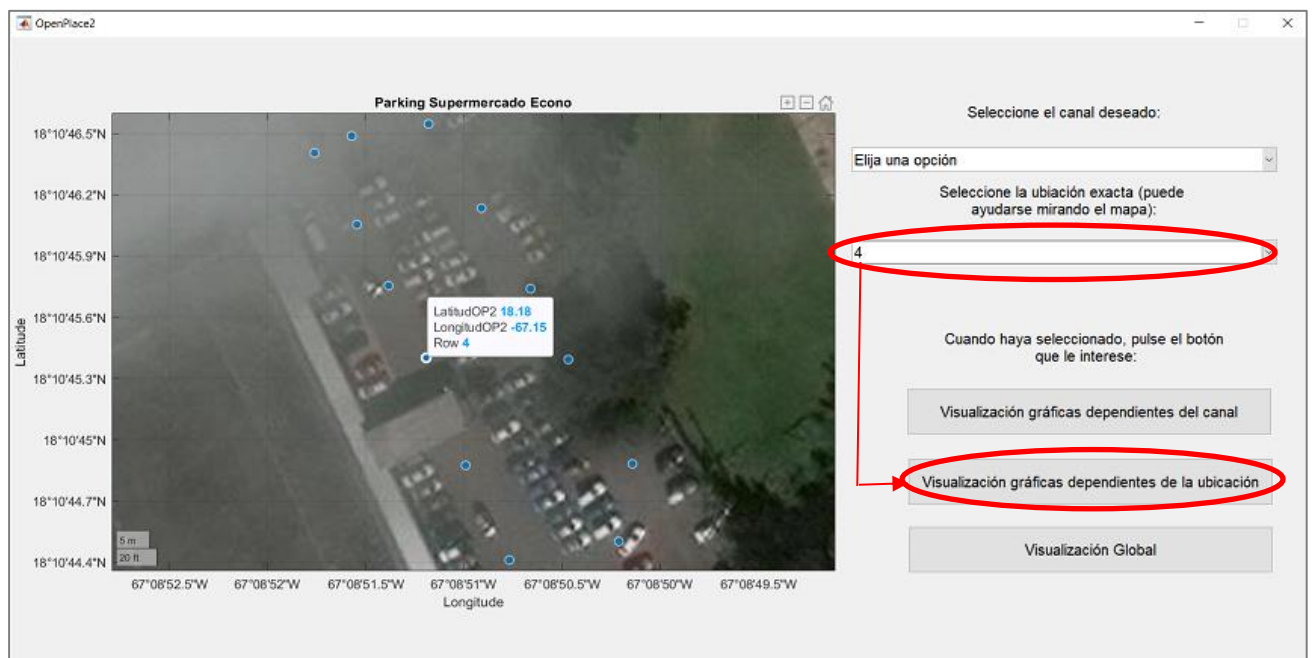


Ilustración 5.19. Selección de la ubicación 4 y el botón “Visualización gráficas dependientes de la ubicación”

Como resultado de este proceso, se presenta en pantalla la Ilustración 5.20.

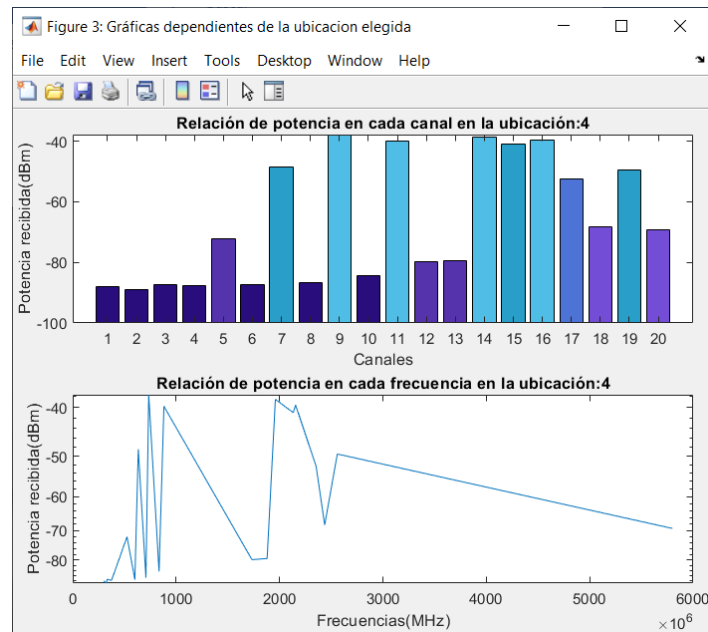


Ilustración 5.20. Gráfica de la ubicación 4 del “Espacio Abierto número 2”

Como se puede comprobar en la imagen, en esta ubicación concreta, los canales a los que les llega una potencia superior son el canal 9, 11, 14 y 16.

Con los valores elegidos para el canal y la ubicación, pulsando el botón “Visualización Global” se obtienen la Ilustración 5.18 y la Ilustración 5.20, y, además, las dos gráficas que se muestran en la Ilustración 5.21.

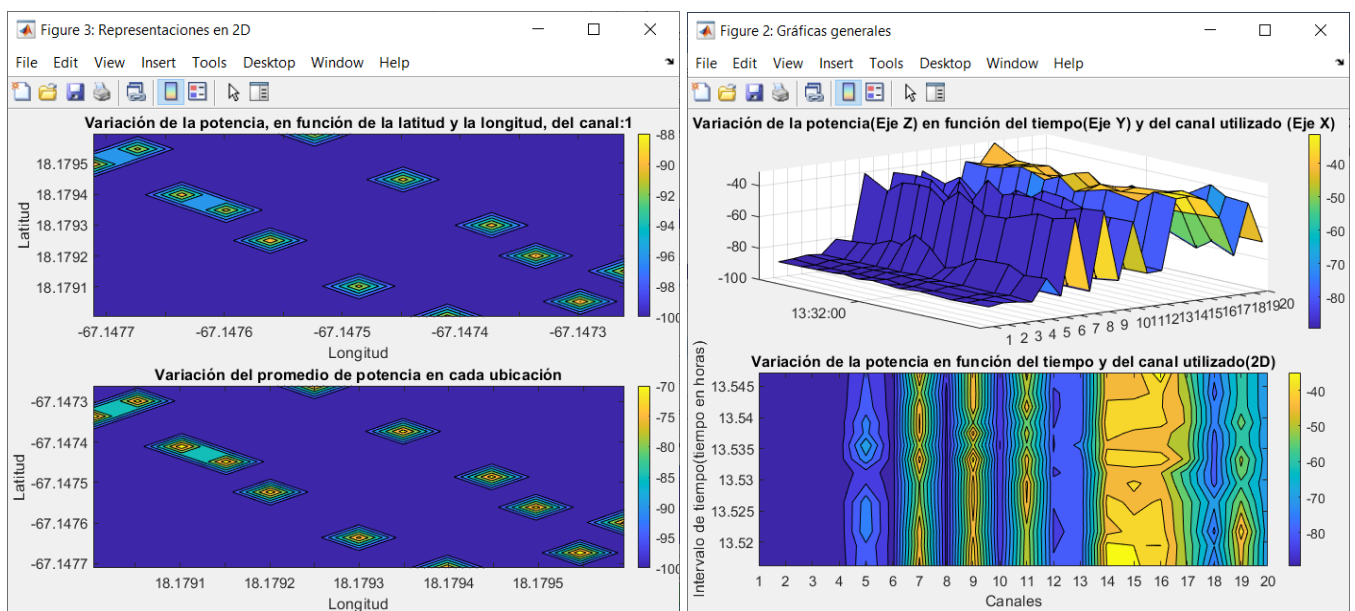


Ilustración 5.21. Gráficas de carácter general del “Espacio Abierto número 2”

De manera global, se puede observar que, si queremos colocar en el Espacio Abierto número 2 un dispositivo que se alimente usando la técnica conocida como *Energy Harvesting*, debería funcionar en las frecuencias correspondientes a los canales 7, 9, 11, 14, 15 o 16, ya que con estos canales conseguiríamos un mejor funcionamiento del sistema.

5.2.1.3. Opción “Calles”

En el caso de que el usuario seleccione la opción “Calles” dentro del Menú Principal, se le abrirá una nueva interfaz, como la mostrada en la Ilustración 5.22.

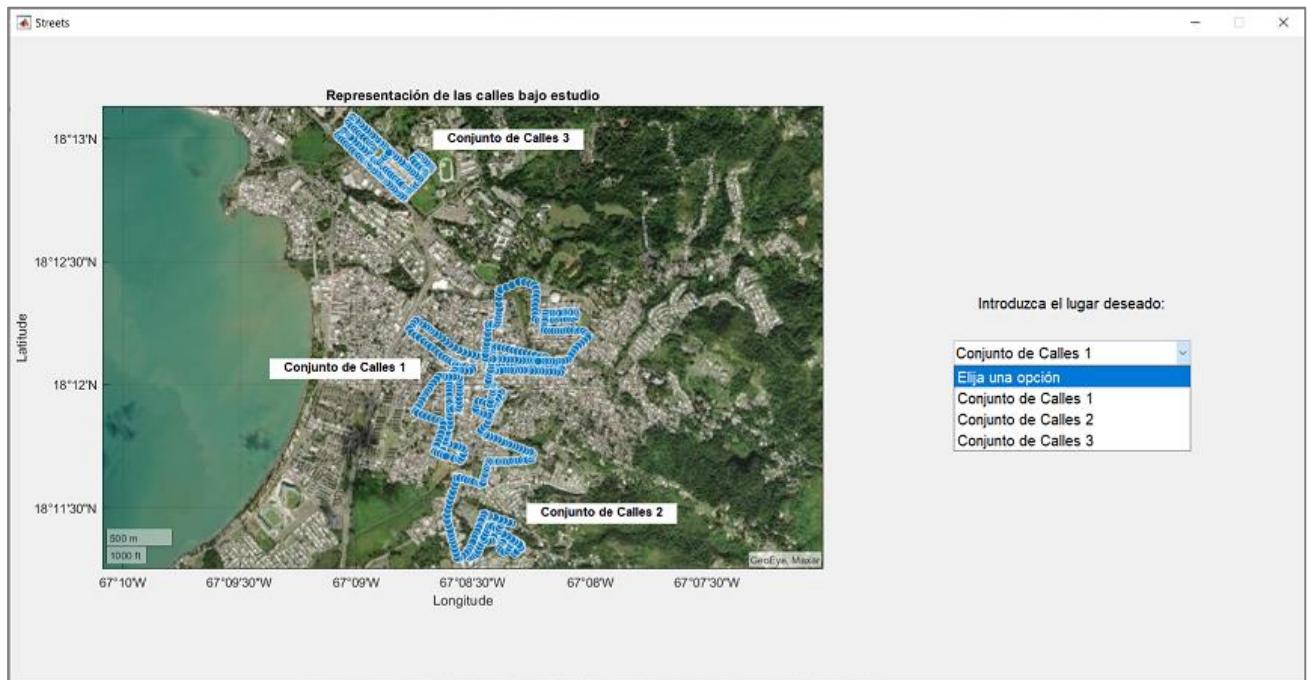


Ilustración 5.22. Menú General para la opción “Calles”

En este caso, el estudio recoge información de tres zonas diferentes, las mostradas sobre el mapa geográfico de la parte izquierda de la imagen. Cada zona va a estar formada por un conjunto de calles y será accesible empleando para ello el menú desplegable de la parte derecha. En este caso, se va a elegir la opción denominada “Conjunto de Calles 1”. Tras seleccionar dicha opción, se le presenta al usuario una nueva interfaz, la mostrada en la Ilustración 5.23.



Ilustración 5.23. Interfaz gráfica “Conjunto de Calles 1”

En esta nueva interfaz lo que se muestra es una imagen satélite de la opción seleccionada. Las “burbujas” que se observan en la imagen se corresponden, nuevamente, con los diferentes puntos en los que se ha medido la potencia recibida. Además, posee los mismos menús desplegables y botones que las interfaces correspondientes a la Avenida y al Espacio Abierto anterior.

Como ya se ha visto, lo primero que debe seleccionar el usuario es el canal. Este proceso se muestra en la Ilustración 5.24, empleando como ejemplo el canal 10. A continuación, se pulsa el botón denominado “Visualización gráficas dependientes del canal”.

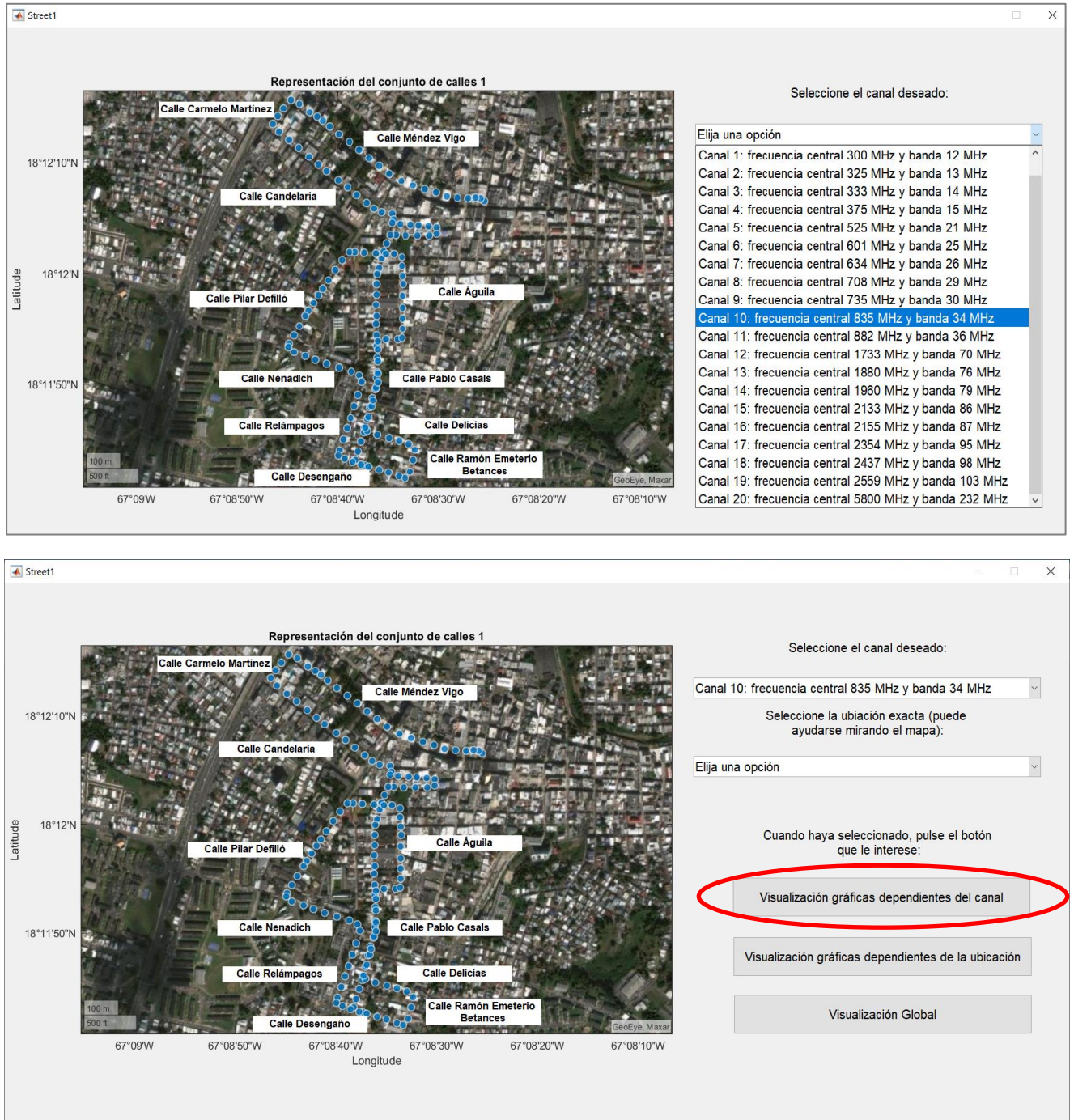


Ilustración 5.24. Proceso de selección del canal 10 en el “Conjunto de Calles 1”

Tras finalizar el proceso anterior, el usuario observa las gráficas correspondientes a la Ilustración 5.25.

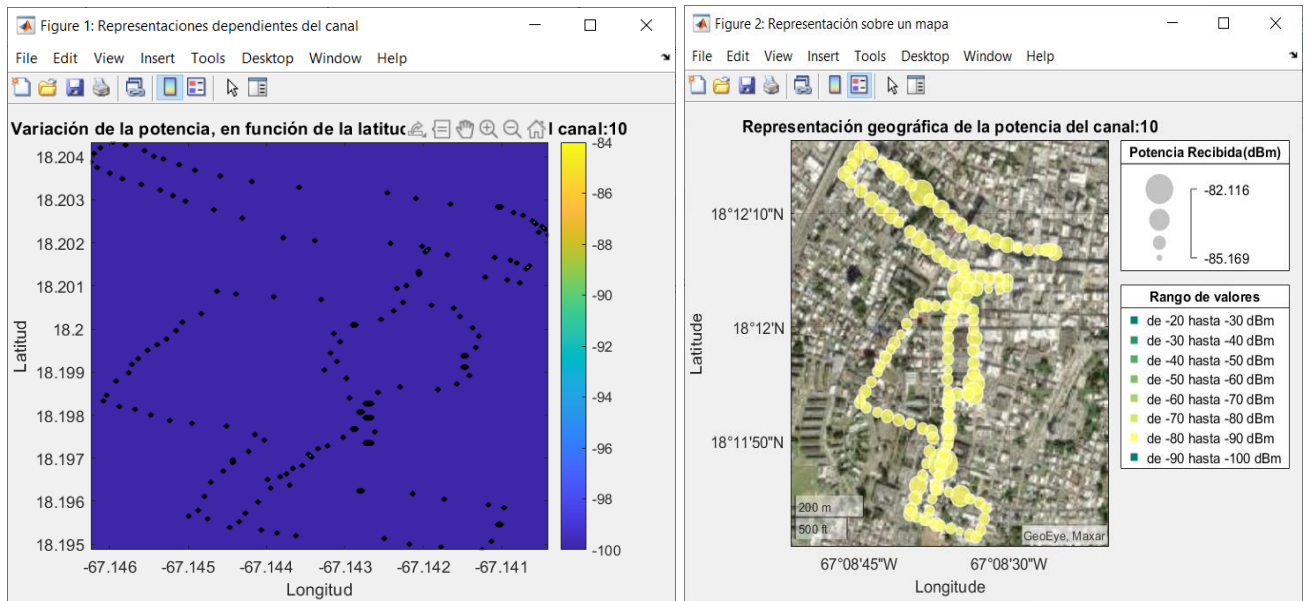


Ilustración 5.25. Gráficas del canal 10 para el “Conjunto de Calles 1”

En este caso, los valores de potencia en esta zona variarán en un rango que va desde los -82 hasta los -86 dBm, de ahí que la gráfica de la derecha presente en todas las “burbujas” un color amarillo. Como se puede observar, los puntos con mayor tamaño de “burbuja” serán en los que se reciba mayor potencia empleando el canal 10.

Repitiendo los mismos pasos para el menú desplegable de la ubicación, elegimos aleatoriamente la ubicación 65, de las 146 posibles, y se pulsa el botón “Visualización dependiente de la ubicación” (Ilustración 5.26).

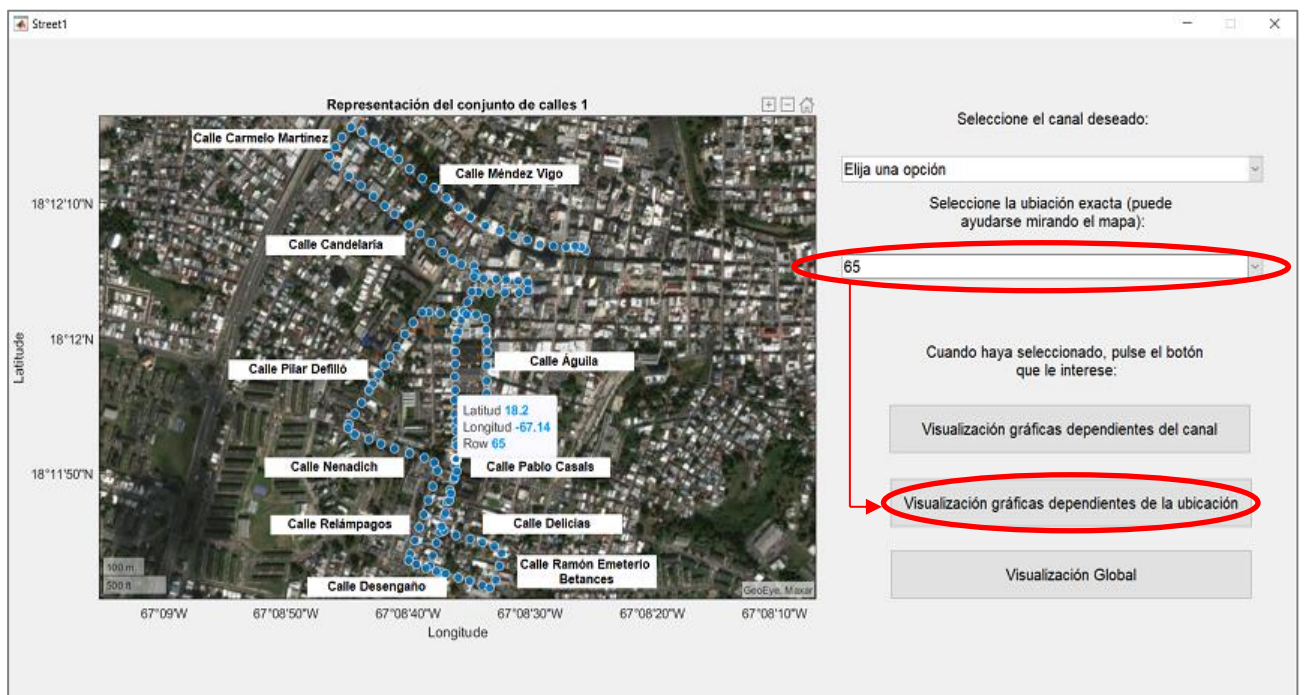


Ilustración 5.26. Selección de la ubicación 65 y el botón “Visualización gráficas dependientes de la ubicación”

Como respuesta, el usuario observa en pantalla la gráfica mostrada en la Ilustración 5.27.

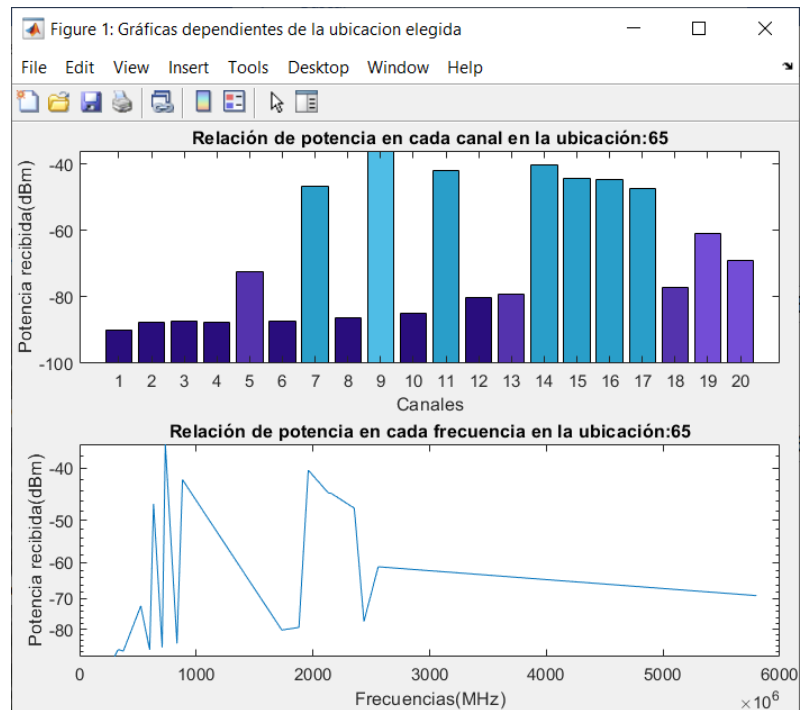


Ilustración 5.27. Gráfica de la ubicación 65 del “Conjunto de Calles 1”

En esta ubicación concreta, se puede observar cómo es el canal 9 el que recoge un nivel de potencia mayor. Este canal corresponde con una frecuencia de 735 MHz.

Por último, con los valores elegidos para el canal y la ubicación, pulsando el botón “Visualización Global” se obtienen la Ilustración 5.28.

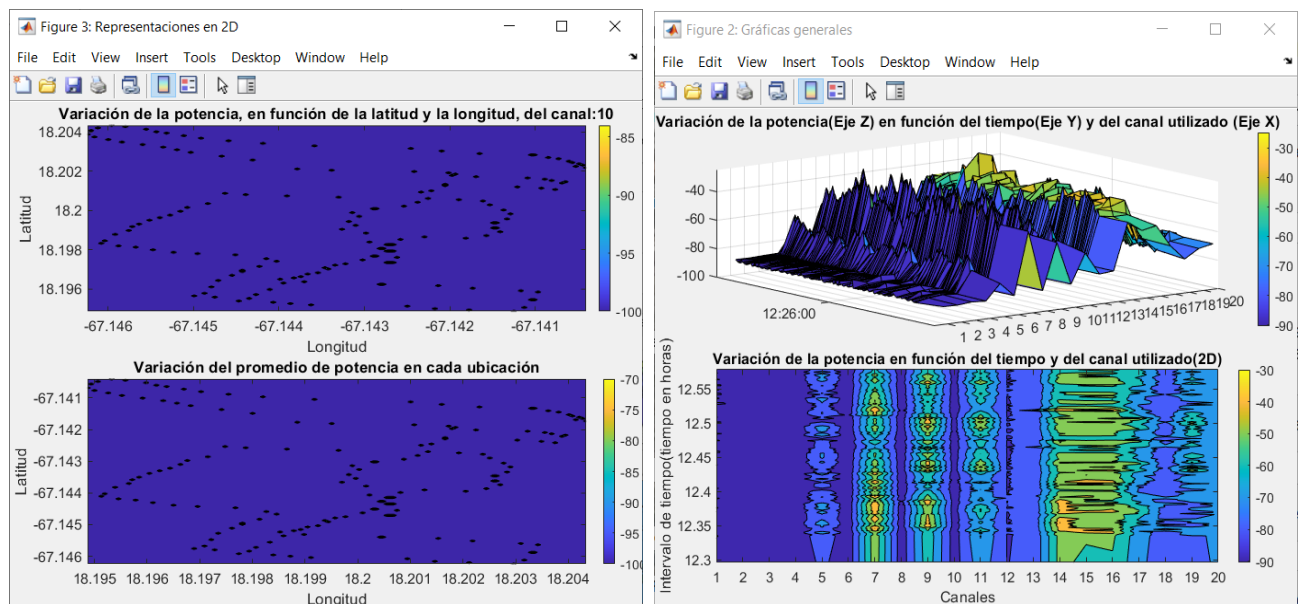


Ilustración 5.28. Gráficas de carácter general del “Conjunto de Calles 1”

De manera global, se puede observar que, si queremos colocar en esta zona un dispositivo, debería funcionar en las frecuencias correspondientes a los canales 7, 9, 11, 14 o 15, ya que con estos canales conseguiríamos un mejor funcionamiento del sistema.

6. CONCLUSIONES Y LÍNEAS FUTURAS

Con lo expuesto en el marco teórico, se hace evidente que vivimos en un mundo en el que existe una sobreabundancia de datos que registran cada movimiento que llevamos a cabo. Y, además, se muestra cómo esta generación masiva de datos va a seguir incrementándose debido a los despliegues de nuevas tecnologías que harán que todo esté conectado, intercambiando información constantemente. Este nuevo paradigma con el que debemos familiarizarnos hace imprescindible la necesidad de desarrollar aplicaciones o herramientas que permitan tanto almacenar como procesar grandes conjuntos de datos de una manera rápida y efectiva. Estas herramientas deben basarse en las representaciones visuales de los datos, pues permitirán a los usuarios una mejor comprensión y análisis de los datos mostrados.

En el presente trabajo se ha llevado a cabo una primera aproximación al tratamiento y representación gráfica de un conjunto de datos recogidos a partir de un cierto estudio. Se ha puesto de manifiesto la importancia de un buen primer análisis de los datos, prestando especial interés a sus diferentes tipos y las fuentes de las que proceden. La naturaleza de los datos será la que marque el modelo de representación a utilizar pues, como se ha visto, para cada tipo de datos existen ciertos gráficos que resultan más apropiados.

Para el desarrollo práctico, es decir, el procesado y representación de los datos se ha empleado la herramienta MATLAB, que permite realizar una gran cantidad de funciones, entre las que se incluye, también, la creación de una interfaz gráfica de usuario, con la que se pretende facilitar la interacción con los datos. Sin embargo, la utilización de una base de datos con un formato de fichero concreto ha resultado ser una tarea ardua, pues MATLAB no es capaz de manejar cualquier tipo de archivo, lo que ha propiciado un gran trabajo previo de adaptación de los datos a un modelo “entendible” por el lenguaje de programación de esta herramienta.

De esta manera, otras posibles líneas futuras se presentan de una manera clara. A partir de este trabajo se podría perseguir el analizar otras bases de datos diferentes, que no estuvieran relacionadas con la obtención de datos electromagnéticos geolocalizados, o la utilización de otro tipo de herramientas gráficas, algunas de las cuales se han presentado en los apartados teóricos. Además, resultaría interesante la creación de, por ejemplo, una aplicación que fuera capaz de analizar grandes volúmenes de datos con el objetivo de detectar patrones, tendencias o anomalías de una manera sencilla.

REFERENCIAS

- [1] ALCALDE, I., 2015. *Visualización de la información: de los datos al conocimiento*. Barcelona: Editorial UOC. ISBN 84-9064-752-6.
- [2] SAHAY, A., 2017. *Data visualization. Volume I, Recent trends and applications using conventional and big data*. First edition. New York, New York (222 East 46th Street, New York, NY 10017): Business Expert Press. ISBN 1-63157-336-5.
- [3] ASSAM, T. y EVERGREEN, S., 2013. *Data visualization. Part 2*. Hoboken, New Jersey: [Jossey-Bass]. ISBN 1-118-83357-0.
- [4] YUK, M. y DIAMOND, S., 2014. *Data visualization for dummies*. Hoboken, New Jersey: John Wiley & Sons, Incorporation. ISBN 1-118-50293-0
- [5] EBRETSEN, M. y KENNEDY, H., 2020. *Data Visualization in Society*. Amsterdam: Amsterdam University Press. ISBN 94-6372-290-4.
- [6] KRUM, R., 2014. *Cool infographic: effective communication with data visualization and design*. 1st edition. Indianapolis, Indiana: Wiley. ISBN 1-118-83715-0.
- [7] SIMON, P., 2014. *The visual organization: data visualization, big data, and the quest for better decisions*. 1st edition. Hoboken, New Jersey: Wiley. ISBN 1-118-85841-7.
- [8] SAHAY, A., 2017. *Data visualization. Volume II, Uncovering the hidden pattern in data using basic and new quality tools*. First edition. New York: Business Expert Press. ISBN 1-63157-731-X.
- [9] CHEN, M., HAUSER, H., RHEINGANS, P. y SCHEUERMANN, G., 2020. *Foundations of Data Visualization*. 1st ed. 2020. Cham: Springer International Publishing. ISBN 3-030-34444-4.
- [10] CHEN, C., HÄRDLE, W.K. y UNWIN, A., 2008. *Handbook of Data Visualization*. 1st ed. 2008. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 1-281-18034-3.
- [11] <https://www.mathworks.com/products/matlab.html>
- [12] MAJUMDAR, N. y BANERJEE, S., 2012. *MATLAB graphics and data visualization cookbook tell data stories with compelling graphics using this collection of data visualization recipes*. Birmingham, UK: Packt Pub. ISBN 1-283-96110-5.
- [13] <https://es.mathworks.com/help/matlab/ref/surf.html>
- [14] <https://es.mathworks.com/help/matlab/ref/contourf.html>
- [15] COBO, N., 2019. *Energy Harvesting: Electromagnetic Site Survey*. IEEE Dataport, doi: <https://dx.doi.org/10.21227/0dfq-ag14>.
- [16] <https://es.mathworks.com/discovery/matlab-gui.html>

- [17] <https://es.mathworks.com/help/matlab/ref/bar.html>
- [18] HURWITZ, J., 2013. *Big data for dummies*. 1st ed. Indianapolis, Ind: John Wiley & Sons. ISBN 1-118-64417-4.
- [19] <https://es.mathworks.com/help/rf/ref/semilogy.htm>
- [20] FUENTES CAMAS, P.L. y NIETO GARCÍA, E.J., 2012. *Programación de una interfaz gráfica de MATLAB para la aplicación del método de los elementos finitos*. S.l: s.n.].
- [21] <https://es.mathworks.com/help/matlab/ref/geobubble.html>
- [22] LOTH, A., 2019. *Visual analytics with Tableau*. 1st edition. Indianapolis, Indiana: Wiley. ISBN 1-5231-2859-3.
- [23] MILLIGAN, J.N., 2015. *Learning Tableau: leverage the power of Tableau 9.0 to design rich data visualizations and build fully interactive dashboards*. 1st edition. Birmingham, England: Packt Publishing. ISBN 1-78439-116-6.
- [24] <https://developers.google.com/chart/>
- [25] <https://dle.rae.es/dato>
- [26] <https://es.mathworks.com/help/matlab/characters-and-strings>
- [27] <https://es.mathworks.com/matlabcentral/fileexchange/35642-kml2struct>

