Centro Universitario de Estudios Medioambientales
Seminarios de la reunión semanal del CUEM
Seminario: 2026-04-27
Expositor: Alfredo Rigalli
Tema: Conociendo el tamaño de los objetos de R

No se existe investigación si no existen datos. Qué son los datos? Que es un dato? Un dato es un valor numérico, cualidad, resultado de análisis y prácticamente cualquier cosa surgida de la aplicación del proceso de medición y observación de un sistema. Un dato se fundamento en 3 pilares. Estos son la objetividad, el contexto y el registro. La objetividad hace referencia a que un dato debe poder ser observado o medido por otro investigador en las mismas condiciones. Por ejemplo, si medimos en el CUEM la concentración de arsénico de una muestra de agua y esa muestra es analizada por otro laboratorio con la misma técnica de medición, debería dar un valor de concentración de arsénico que no discrepe del valor medido en el CUEM. El contexto implica que un dato sin contexto no es un dato. 38ºC no dice nada por sí solo. El contexto también son datos y se suele hacer referencia a ellos como metadato. Los metadatos son tan importantes como los datos. Por ejemplo, medimos la concentración de arsénico de una muestra de agua y obtuvimos el valor de 20 ppm, siendo la muestra de Rosario, del día 2/4/26, de una canilla de la red de distribución, se midió por la técnica de hidrazina, etc. La concentración de 20 ppm es el dato, los demás son los metadatos. El registro significa que un dato tiene que tener un método y un soporte donde el dato persista. Por ejemplo, cada vez que medimos una concentración de cualquier componente, esa medición queda registrada en el cuaderno del CUEM y en la base de datos de Atlantis.
Actualmente se clasifican a los datos en small data (bases de datos pequeñas), medium data (bases de datos de tamaño apreciable) y big data (enormes bases de datos). Llamamos small data a aquellos datos que entran en una planilla de cálculo o papel. se manejan con cualquier programa de computación, incluso a mano. Medium data, son aquellos datos y metadatos que entran en una memoria ram o expansión y requieren programas específicos. no alcanza una planilla de cálculo y es imposible
manejo a mano. Por último big data: no entran en una ram y necesitan computadoras potentes o varias organizadas en clusters, para su procesamiento. Generalizando podemos decir que small data son bases del orden de los kBytes, medium data del orden de los MBytes y bigdata del orden de TeraBytes.
Que es un byte? Un byte es la unidad de información que maneja una computadora y está compuesta por 8 bits. Qué es un bit: Es la mínima unidad de información y puede tomar valor 0 o 1. Lo humanos usamos un lenguaje humano para comunicarnos con la computadora. Existen muchos lenguajes que llamamos lenguajes de programación. R tiene su lenguaje, Arduino tiene el suyo y así sucesivamente. El compilador es un programa o interprete que transforma nuestro "programa" en lenguaje de máquina y se lo pasa al microprocesador de la computadora para que ella lo ejecute. Si el microprocesador hace algo, por ejemplo suma 2 + 2, el interprete transformará el resultado y nos mostrará el resultado en lenguaje humano a través de la pantalla, por ejemplo. La información que se maneja en bytes por las computadoras podemos verla tal cual utilizando lenguajes "no humanos" como el lenguaje binario o el hexadecimal. Solo a modo de ejemplo la letra E, en binario es 01000101 y en hexadecimal: 0x45
Los datos se deben almacenar de una manera adecuada para su análisis y envío. En R existe una infinidad de tipos de datos, para mencionar los de uso cotidiano tenemos los datos: character, integer, numerico, factor y raw. Los character o caracter son letras o cadenas de letras (llamados strings). Los strings pueden ser letras, palabras, oraciones, ejemplo de este tipo de datos es la columna tipoagua de atlantis donde puede existir por ejemplo: "pozo". Los datos integer (entero) son número enteros, en Atlantis por ejemplo cuando preguntamos cuantas personas la consumen, el registro queda con un número entero. Si la persona dice que son 5 personas en la casa, en el registro quedará: 4. El formato numerico (numeric) se utiliza para número con decimales, en general. La concentración de sodio puede ser 120,3 ppm, es un dato de tipo numeric. Los datos de tipo factor, son datos que toman valores diferentes, pero solo algunos dentro de una categoría. El dato de Atlantis que llamamos tipoagua puede ser "pozo", "red", "ósmosis", "envasada" y "otro". Los datos de tipo factor, son los que tienen categorias. Los datos raw, son datos almacenados directamente en formato binario o hexadecimal. Un dato puede guardarse prácticamente en cualquier formato y lo que difiere es el espacio de memoria que utiliza.