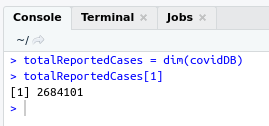
**Libreras y Bases de datos**

**Análisis de bases de datos en R**

*Julian Camilo Cruz Sanchez*

1. **¿Cuántos casos positivos de COVID-19 han sido reportados por el Instituto Nacional de Salud en Colombia?**

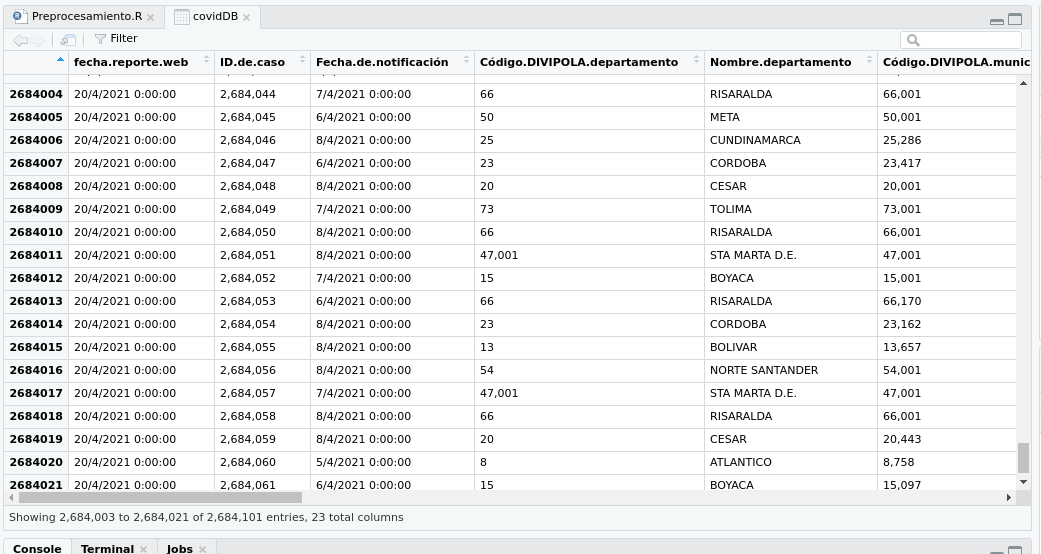
Han sido reportados 2,684,101 casos



1. **¿Qué información se tiene de los casos reportados? Revisar las variables que tiene la base.**

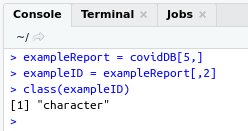
La informacion que tiene sobre cada informe de contagio es la siguiente:

* Fecha de reporte web
* ID del caso
* Fecha de notificacion
* Codigo departamento
* Nombre departamento
* Codigo municipio
* Nombre municipio
* Edad
* Sexo
* Tipo de contagio
* Ubicacion del contagio
* Estado
* Recuperacion
* Otros datos no relevantes

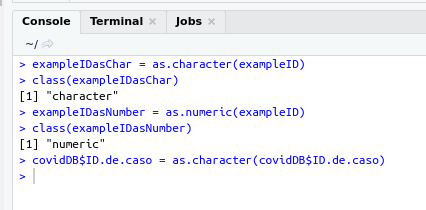


1. **¿Cuál es la clase de la variable “ID de caso”?**

La variable ID de caso es de tipo carácter



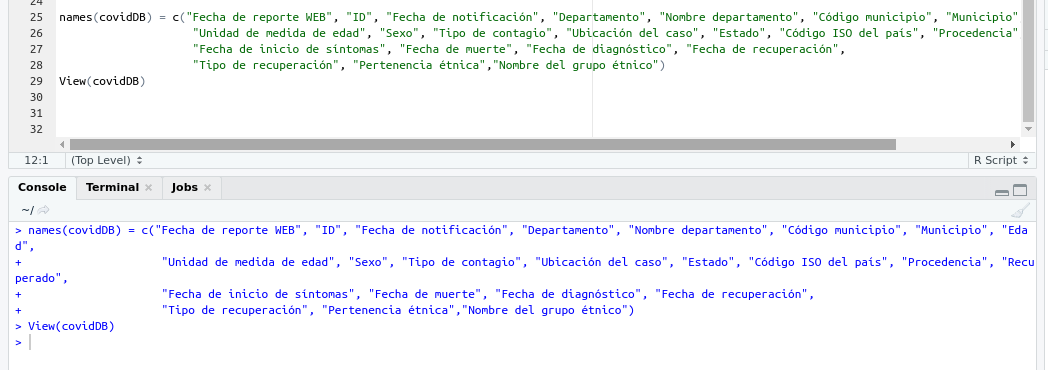
1. **La variable “ID de caso” denota una etiqueta, en tanto, no debe ser numérica. Escriba el código para hacer que ésta variable sea tipo carácter.**



1. **¿La base de datos es un tidy data? Explique su respuesta.**

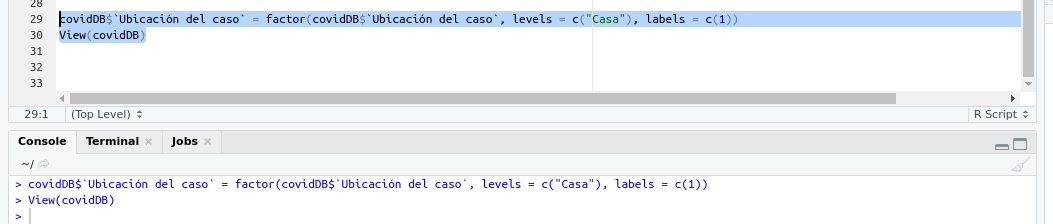
La base de datos obtenida de la pagina de datos abiertos y exportada en el workspace de R es un tidy data, debido a que su organizacion y estructura demustra que se le ha hecho una fase de limpieza de datos. Los tipos de datos para las variables son los correctos y permitirian crear informes y graficos correctamente. Los cambios que se pueden hacer son minimos y mayormente esteticos.

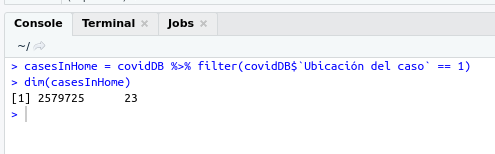
1. **Separe la columna “Fecha de diagnóstico” en tres columnas “Día”, “Mes”, “Año”. Guarde los cambios en la base original**
2. **Unifique la columna “Año”, debe aparecer solo 2020 o 2021, según corresponda. Guarde los cambios en la base original.**
3. **Escriba un código que le permita, cambiar el nombre de las columnas “ID de caso”, “Nombre municipio”, “Nombre departamento” y “Nombre del país” por “ID”, “Municipio”, “Departamento” y “Procedencia”, respectivamente. Guarde los cambios en la base original.**



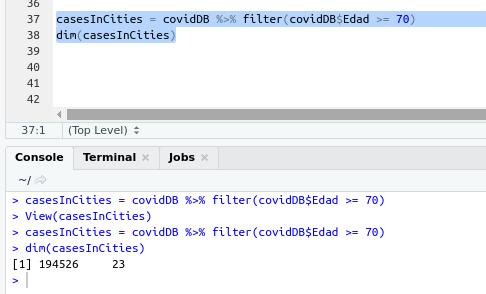
1. **Convierta la variable “Ubicación del caso” a tipo factor. ¿Cuántos casos se han clasificado en casa?**

Se han generado 2,684,101 casos en casa





1. **Haga un filtro que permita saber ¿Cuántas personas se han reportado mayores de 70 años en la ciudad de Bogotá o Medellín? No guarde los cambios. NOTA: la ciudad se ubica en la variable Municipio.**



1. **Haga un filtro que le permita seleccionar las tres personas de mayor edad que han sido reportadas. No guarde los cambios en la base original y describa en dos líneas las características de esas personas.**



1. **Escriba un código que le permita conocer y describir las características del individuo reportado como número 100.**



1. **Escriba un código que le permita ordenar de menor a mayor por edad. No guarde los cambios en la base original.**



1. **Escriba un código que le permita conocer solo el “Sexo” y “Procedencia” de los 10 primeros registros. En la mayoría de los casos ¿cuál fue la procedencia? No guarde los cambios en la base original.**



**Referencias**

Datos abiertos (2021). Casos positivos de COVID-19 en Colombia. Document avaliable in: <https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr>