

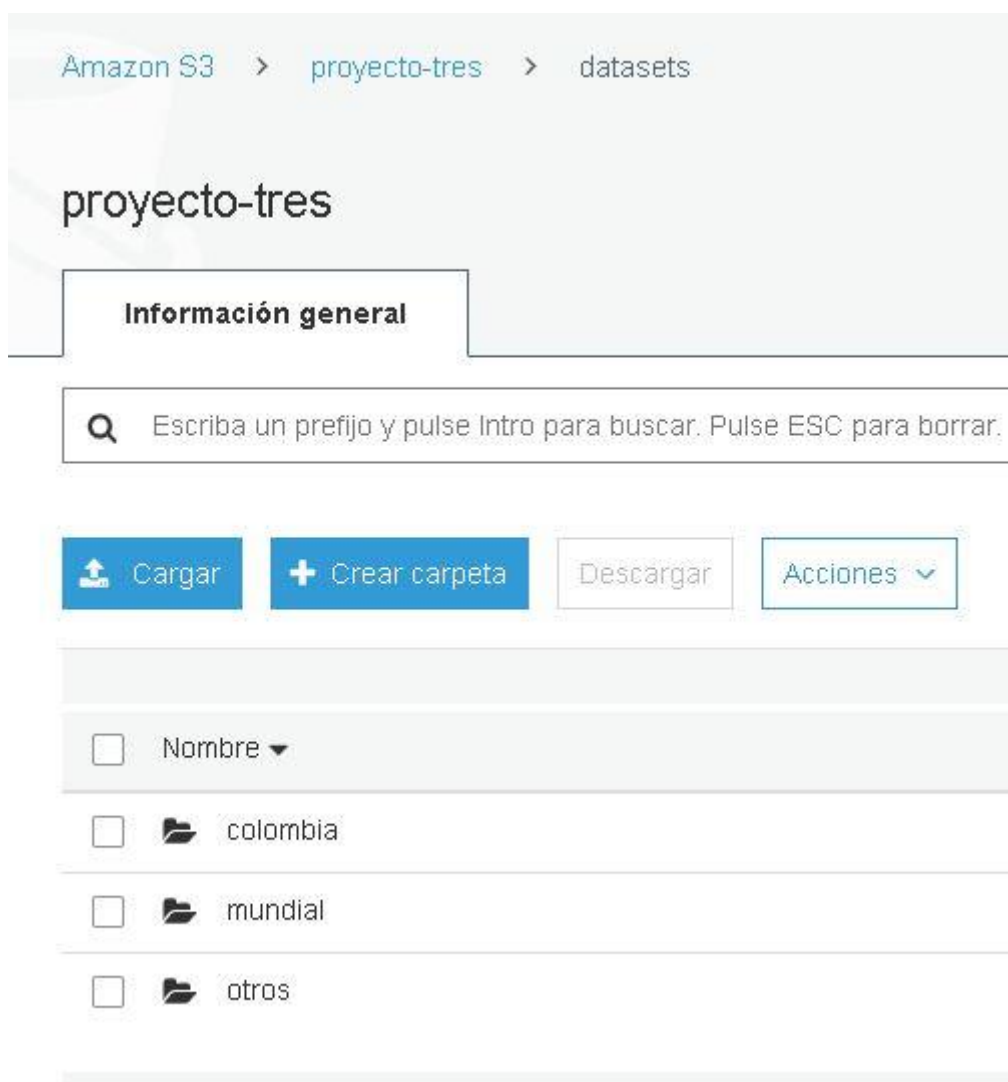
TRABAJO 3 - SPARK sobre COVID-19

1. Fuentes de datos

Se descargaron los datasets que se encontraban en los enlaces provistos. Estos se pueden consultar en el directorio [datasets](#)

2. Ingesta y Almacenamiento de datos

Se creó un bucket en AWS S3 y se subieron los datos de los correspondientes datasets.



3. Procesamiento: Análisis exploratorio de datos con pyspark

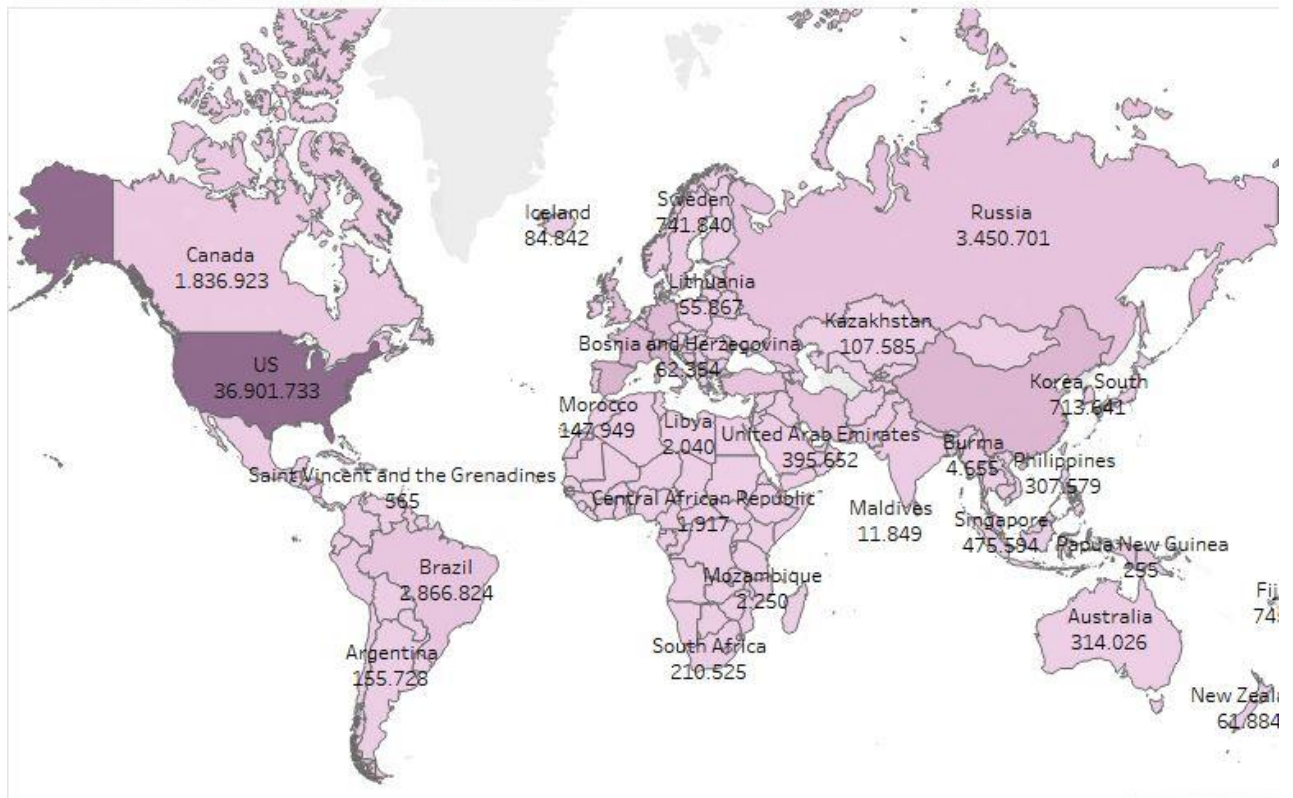
El análisis descriptivo exploratorio se encuentra en el notebook, donde se realizaron consultas para conocer los datos y compararlos en los diferentes escenarios.

Se pueden ver las consultas en el archivo [notebook](#)

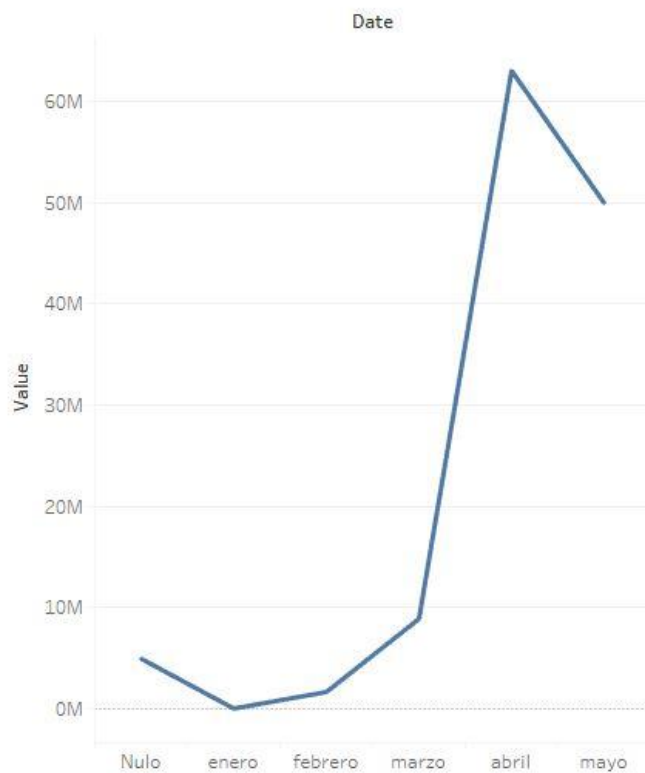
4. Visualización básica de datos

4.1 situación a nivel mundial

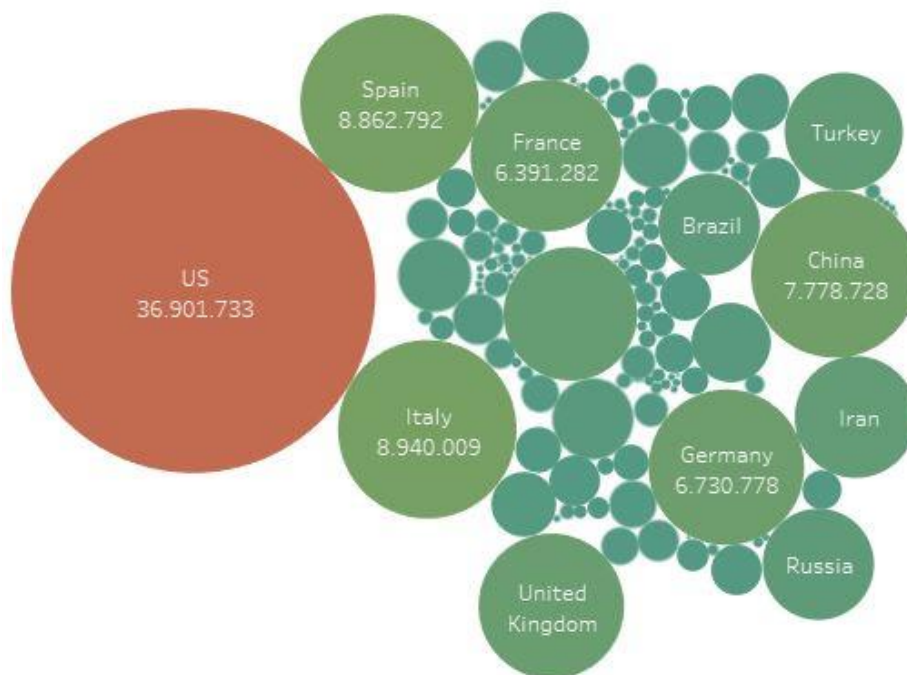
Casos confirmados a nivel mundial



Numero de casos confirmados a lo largo del tiempo



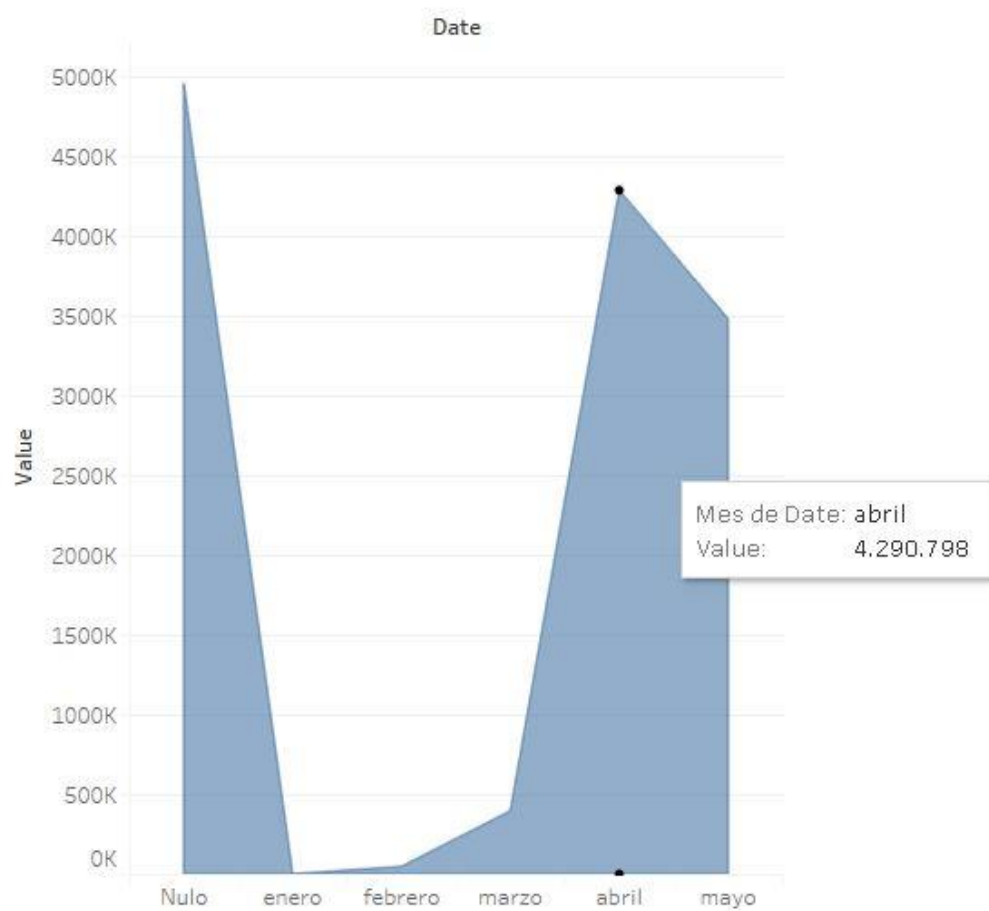
Número de casos confirmados por país



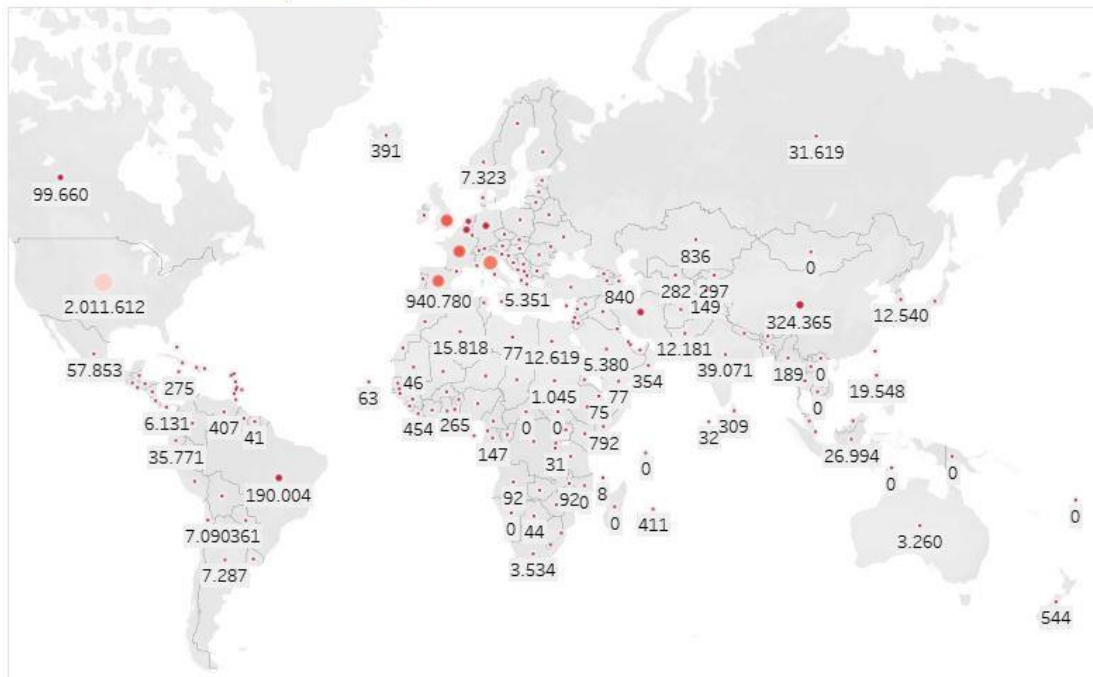
Número de muertos por país



Número de muertos por meses

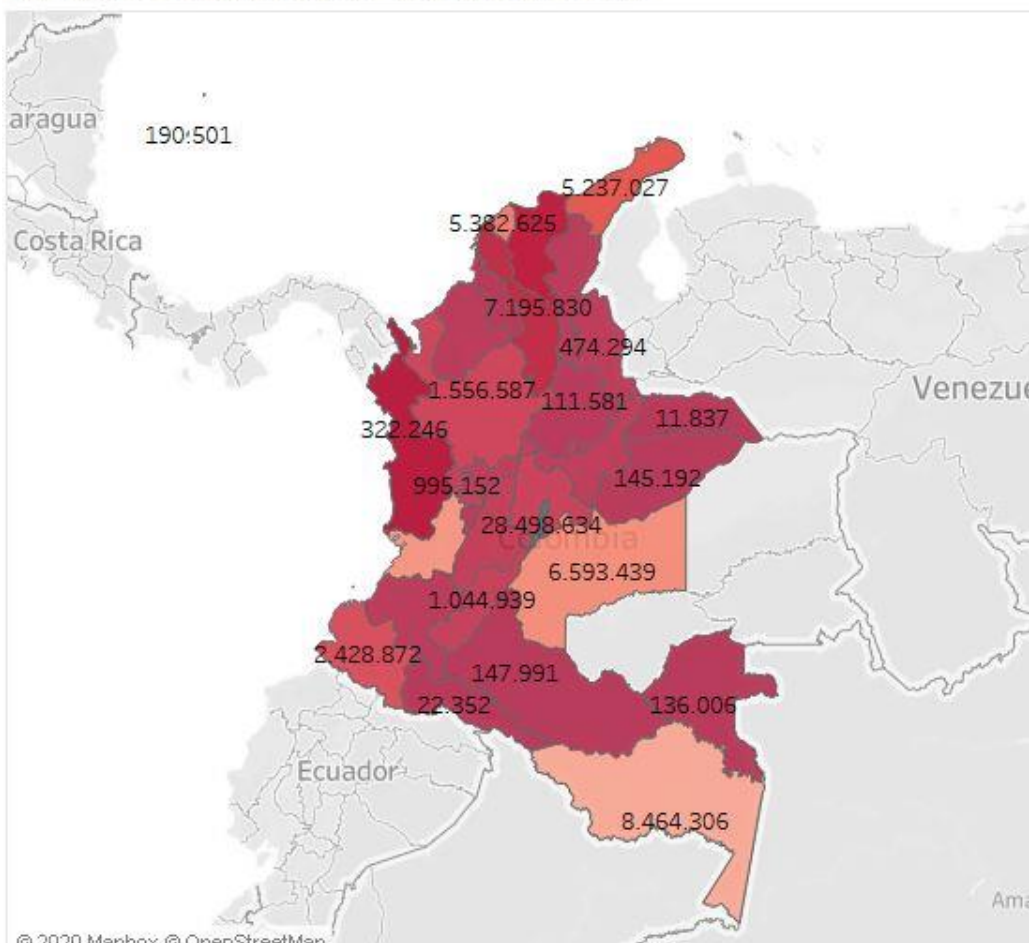


Número de muertos por meses

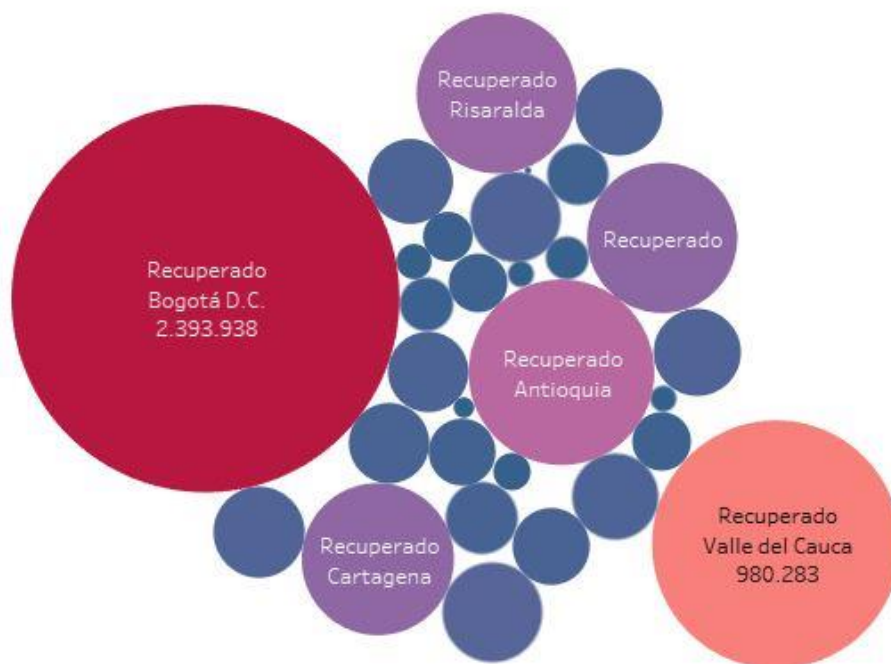


4.2 Situación en Colombia

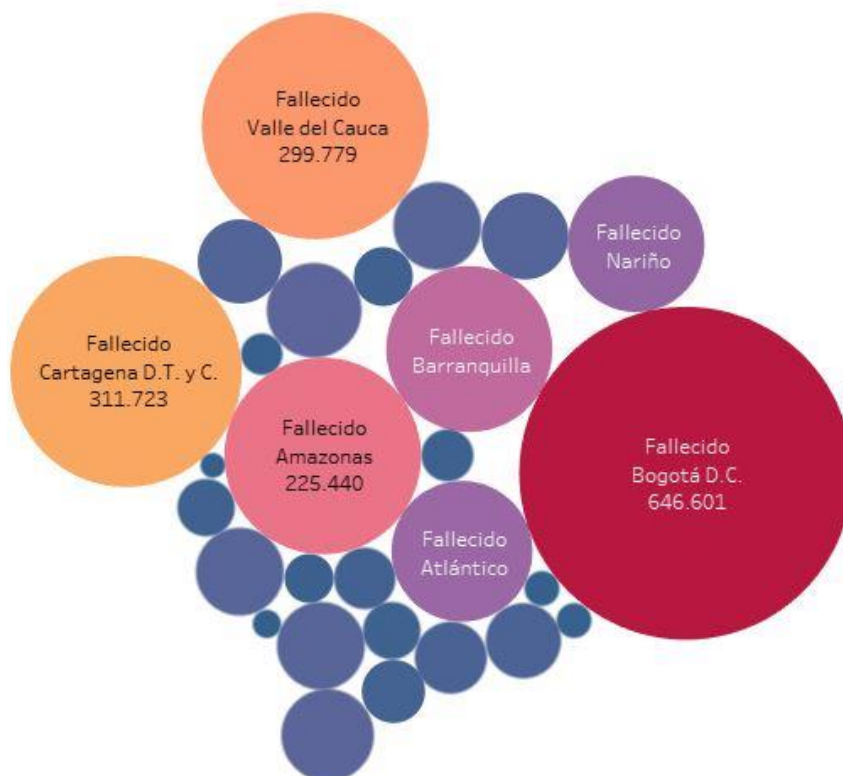
Número de casos por departamento



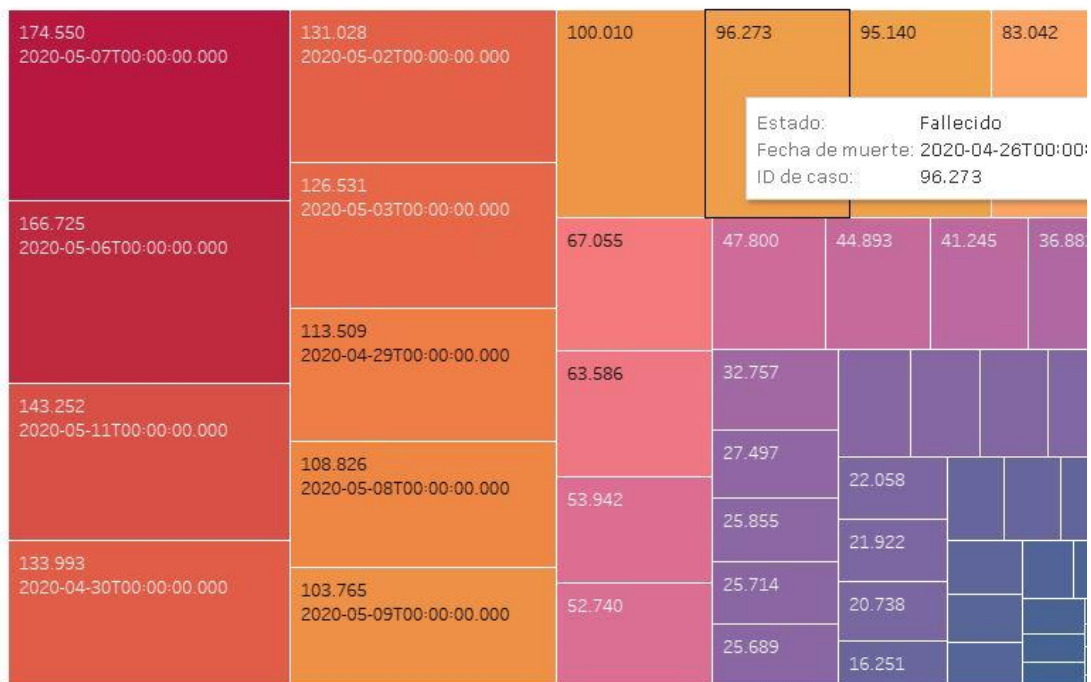
Número de recuperados por departamento



Número de fallecidos por departamento

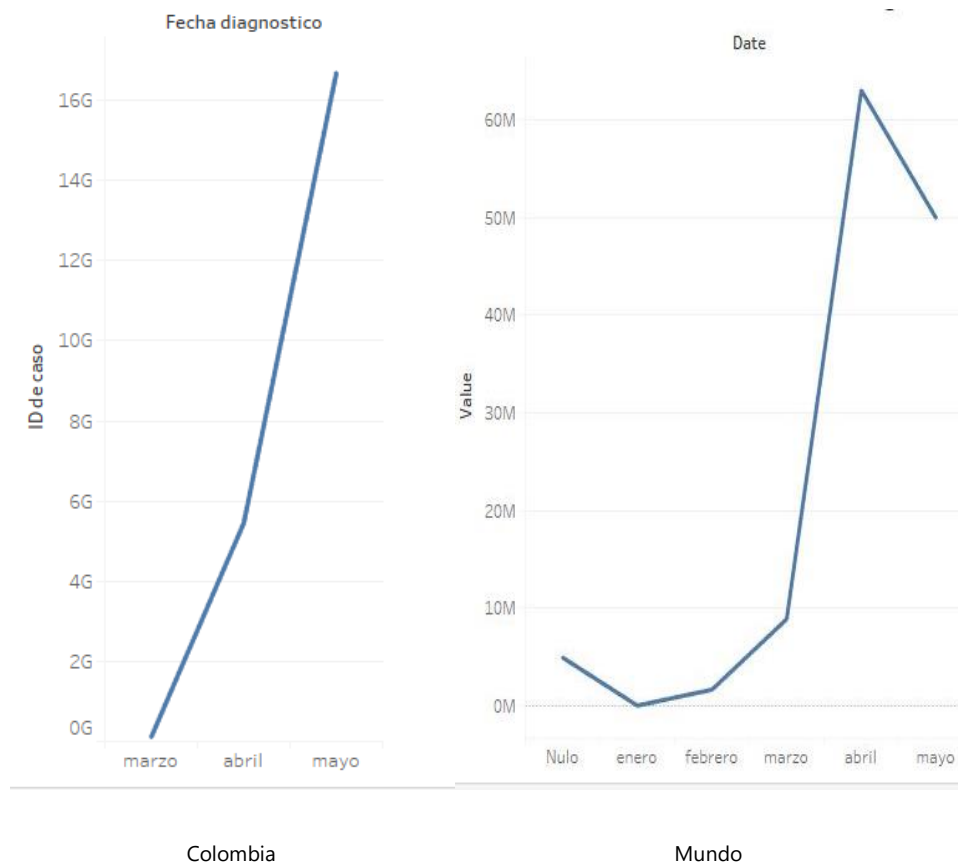


Número de fallecidos por días

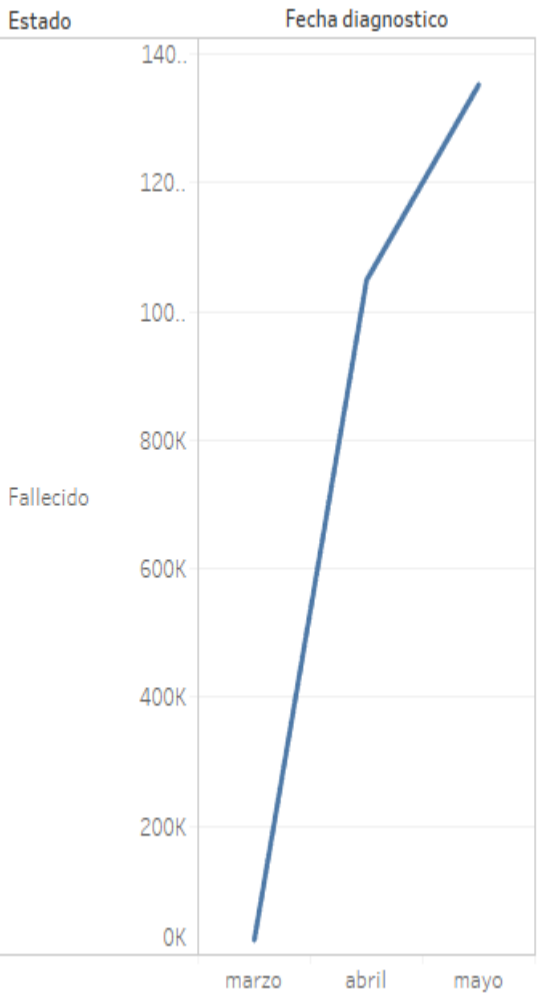


4.3 Colombia vs Mundo

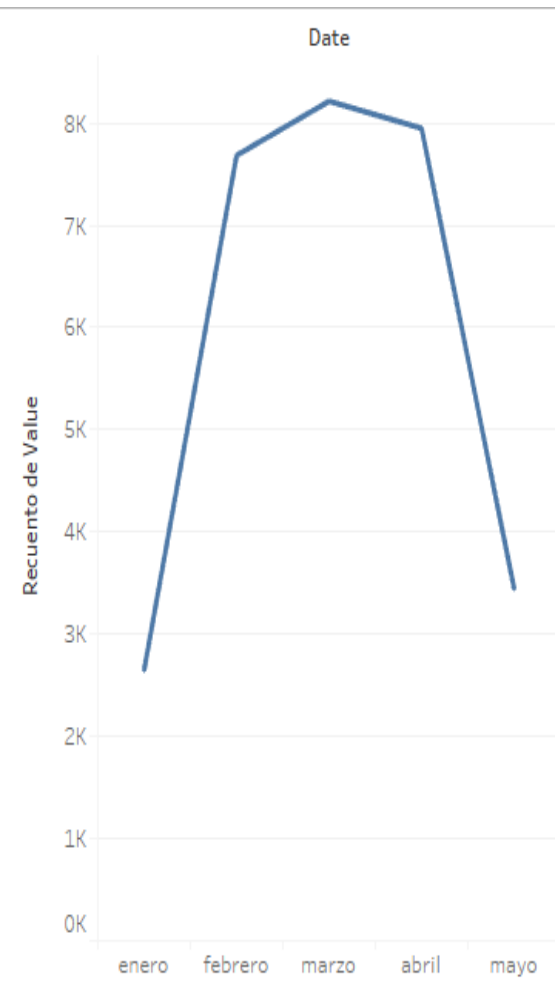
Se puede ver como en el colombia el numero de casos confirmados sigue creciendo mientras en el mundo hubo una reduccion en el mes de mayo.



Podemos ver como los fallecidos en Colombia siguen aumentando mientras que, en el mundo, en el mes de mayo hubo una reducción significativa



Colombia



Mundo