

## Seminario sobre GeoKettle

### Actividad 8

#### Enunciado del problema.

Se dispone de **varios archivos csv con información de la red social twitter**. Estos archivos contienen un conjunto de tweets georeferenciados. Se desea volcar esa información en una base de datos PostGIS para su posterior carga y visualización en QGIS.

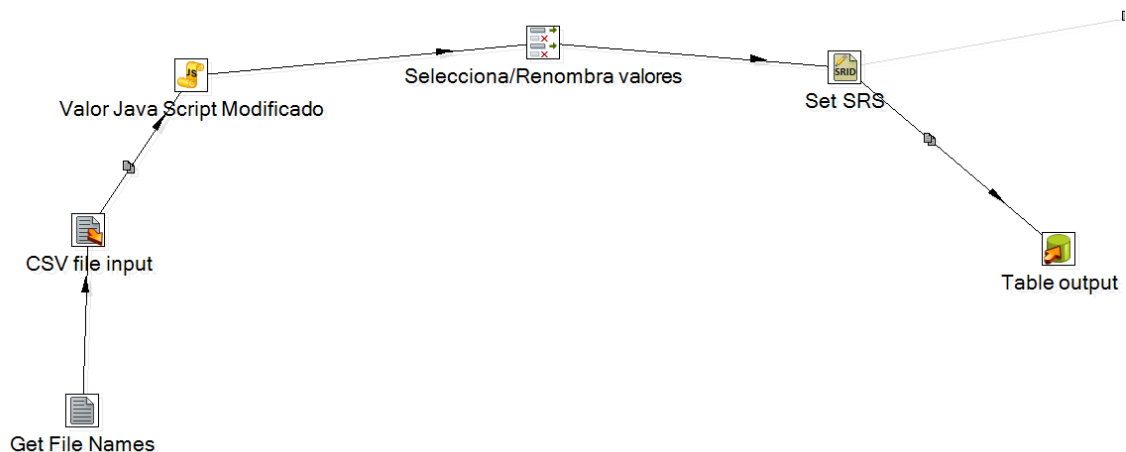
Se desea **fundir en un solo archivo** todo el contenido eliminando la primera línea que contiene los nombres de los campos (columnas) en todos los archivos.

#### Solución inmediata:

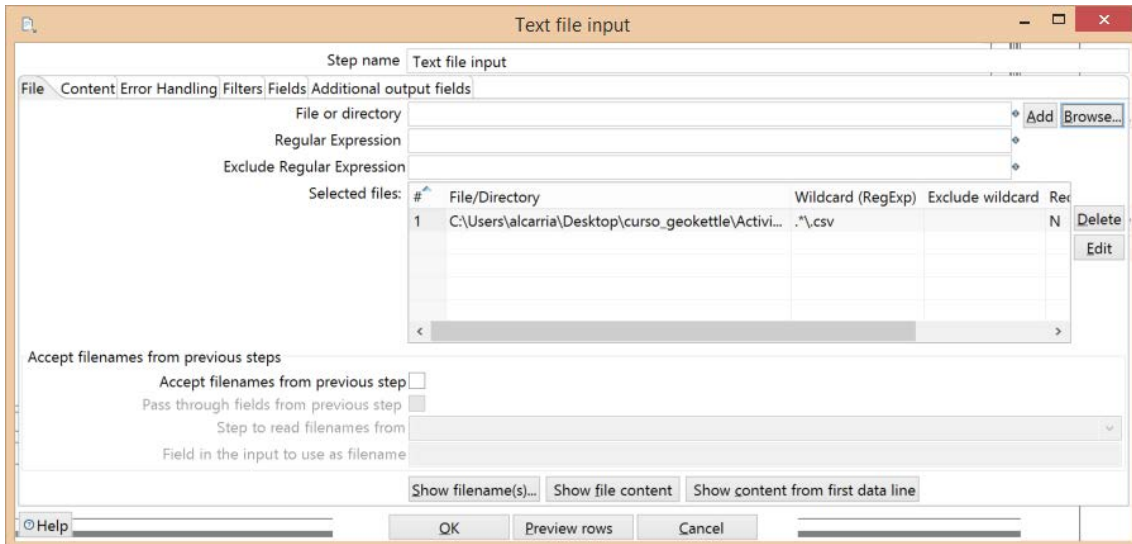
Utilizar un editor de texto y quitar de todos los archivos excepto del primero la primera fila. Copiar el contenido de los archivos en uno sólo y posteriormente abrir con QGIS y utilizar una extensión para publicar en PostGIS.

#### Solución basada en GeoKettle:

Crear una transformación con los siguientes pasos:

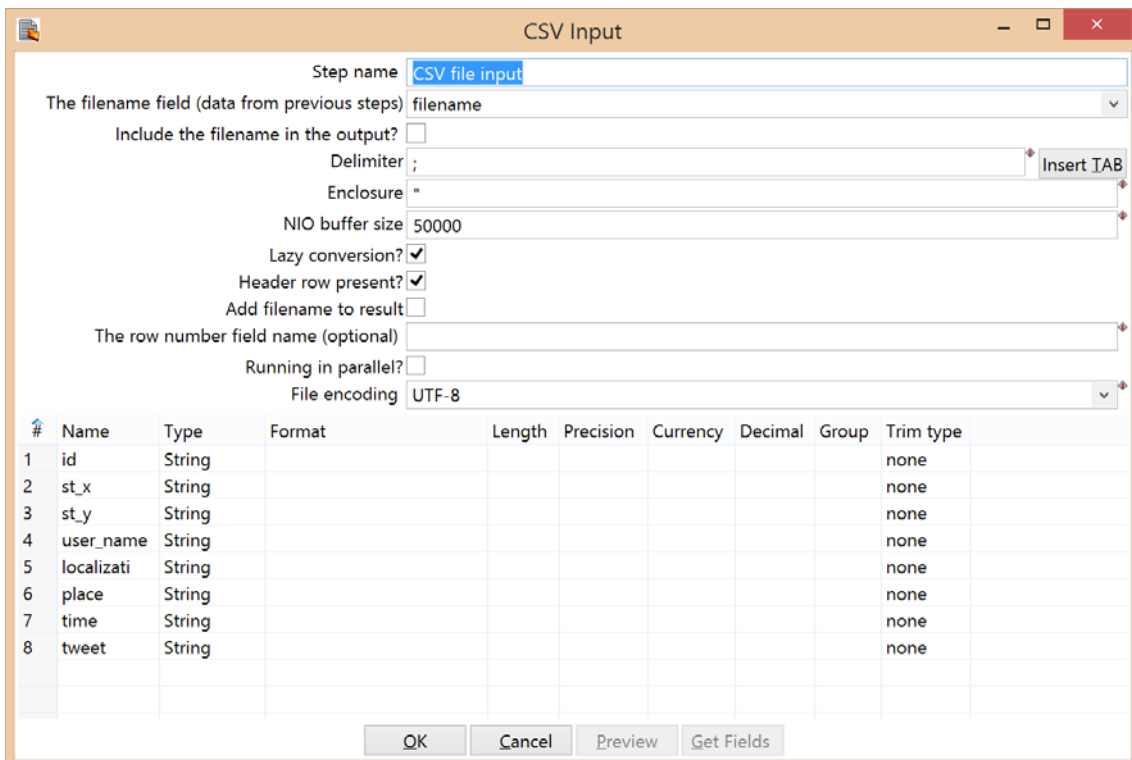


Un paso de entrada Fichero de Texto para recuperar el nombre de todos los archivos CSV. Conectamos este paso con un paso de CSV File Input.



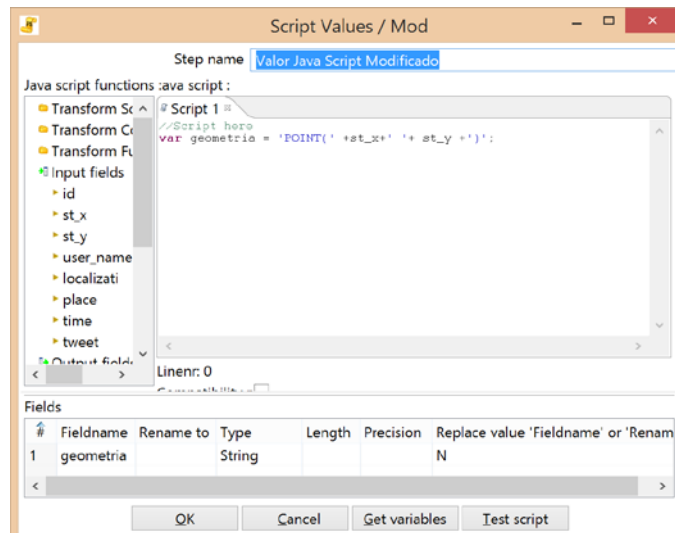
En el paso CSV File Input incluimos como nombre del fichero 'filename'. Elegimos como delimitador el ; y como cierre de cadenas de texto el ", file encoding UTF-8

No podemos recuperar los valores de esta forma, así que tenemos que incluirlo a mano.

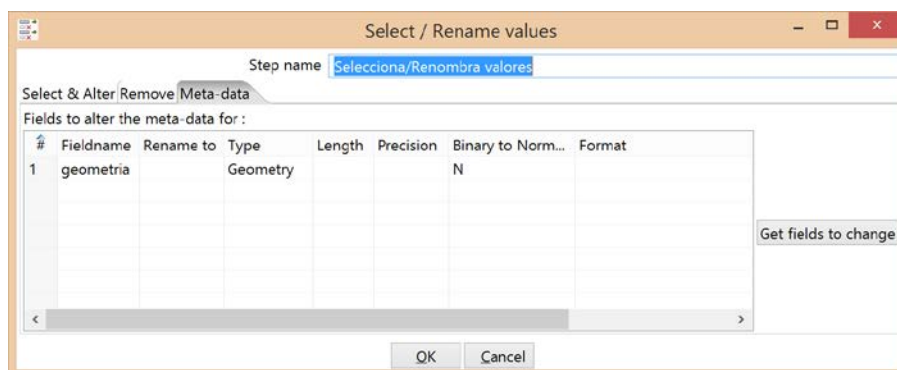


Si solo tenemos un fichero podemos ahorrarnos el paso de Text file input y al cargar directamente el fichero csv a través del caso CSV Input, nos permite traernos los campos.

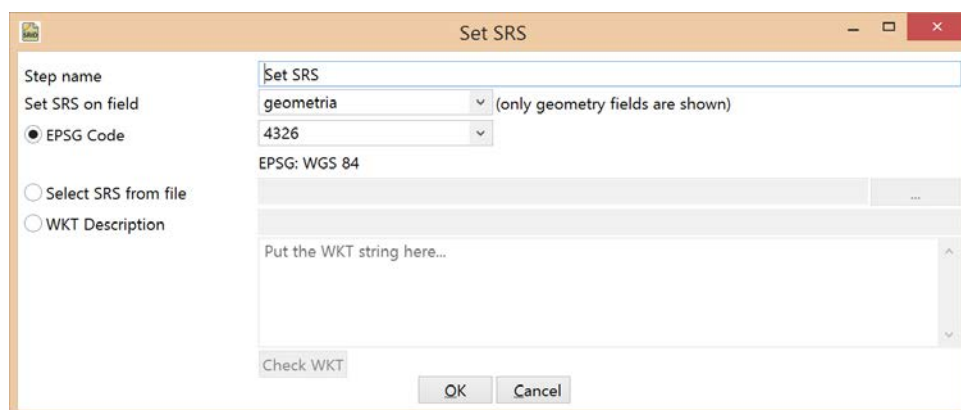
El siguiente paso es un "Valor Javascript modificado" para convertir la geometría a tipo Geometry.



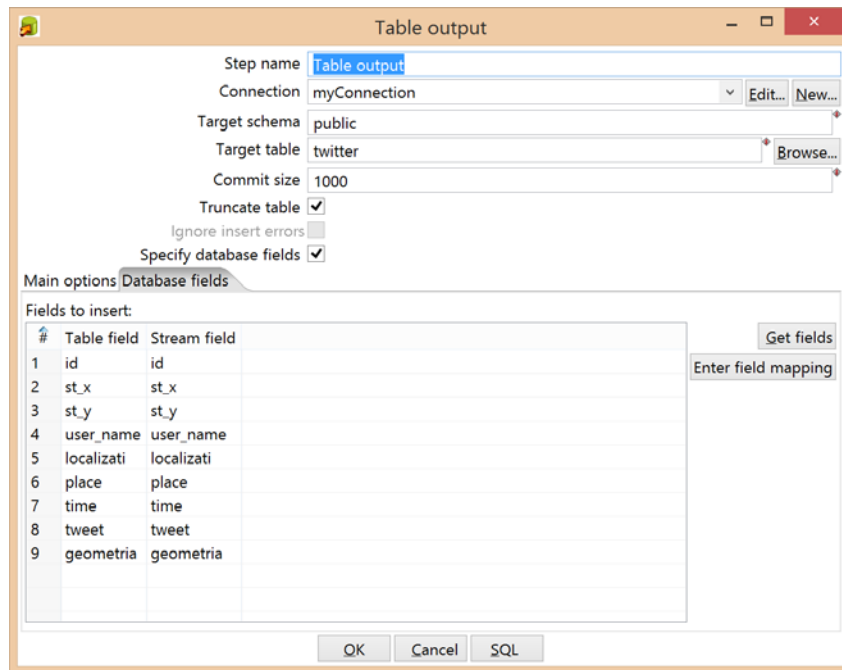
Como hemos dicho, renombramos el valor de la geometría.



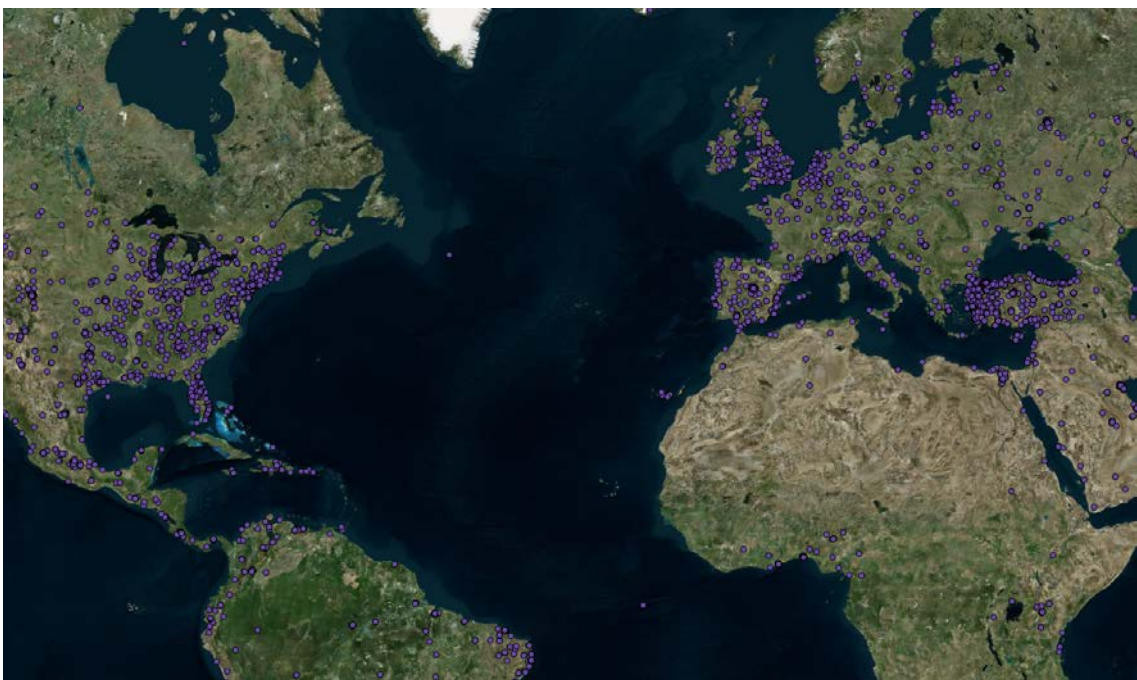
Ofrecemos un sistema de coordenadas. Los datos están en WGS 84, por tanto EPSG:4326



Por último un paso de escritura en tabla:



Por último si cargamos la tabla en el QGIS a través de la opción add “PostGIS layers” podemos ver la localización de los tweets:



Como extensión, podemos poner en paralelo un step para convertir en shapefile, pero si lo ponemos directamente nos va a dar error. Lo que tenemos que hacer es desactivar la opción de “lazy conversions” o conversiones perezosas. Esta opción permitía GeoKettle no tener que convertir los datos a String, Date, Geometry, etc, cuando no es necesario. Haciendo que la ejecución del flujo fuera más rápido, pero a veces se equivoca con algunos módulos, como el de shp, por lo que es necesario desactivarlo.