

Taller de análisis espacial avanzado

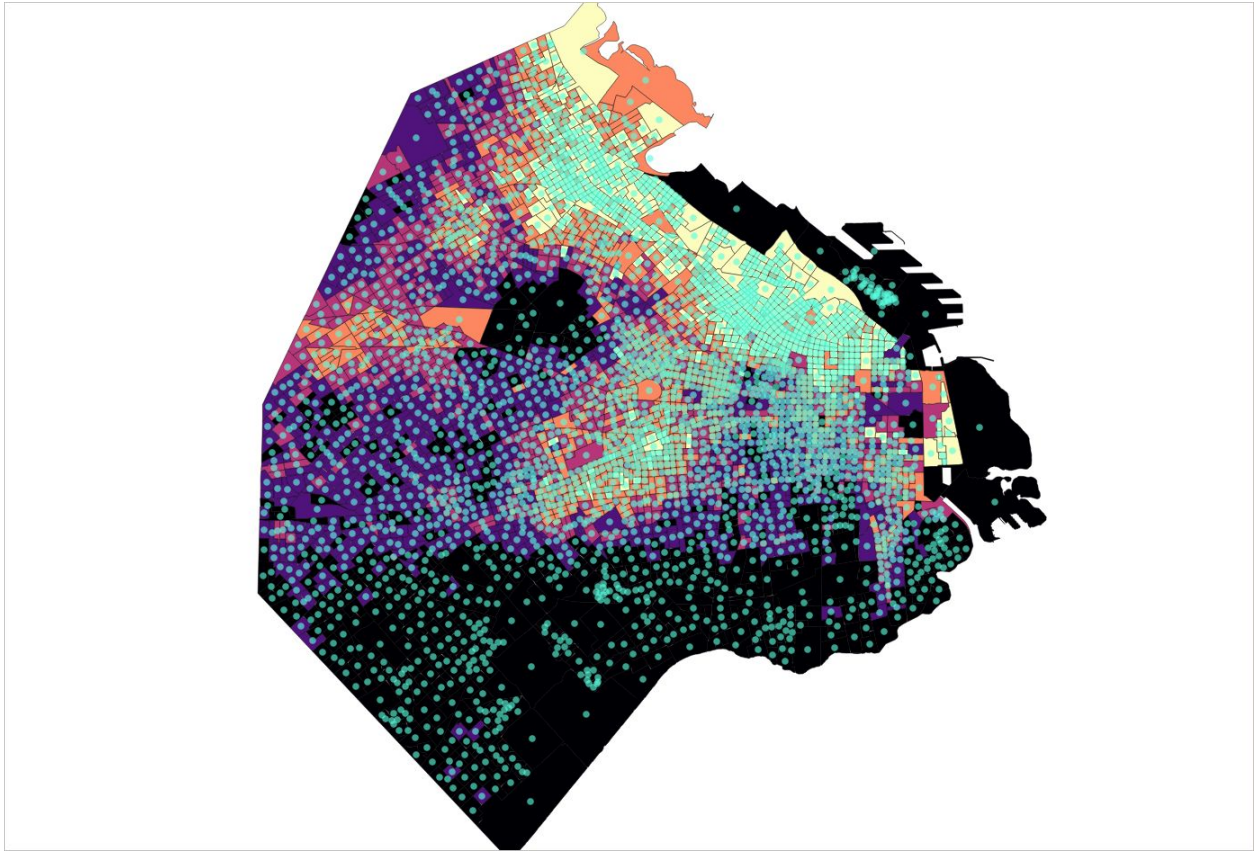
Clusters espaciales y hotspots con SIG - Paso a paso clase 3.
Heatmaps y mallas

Felipe Gonzalez
felipegonz@gmail.com
twitter: @lephcero
github: /alephcero

1 - Heat maps

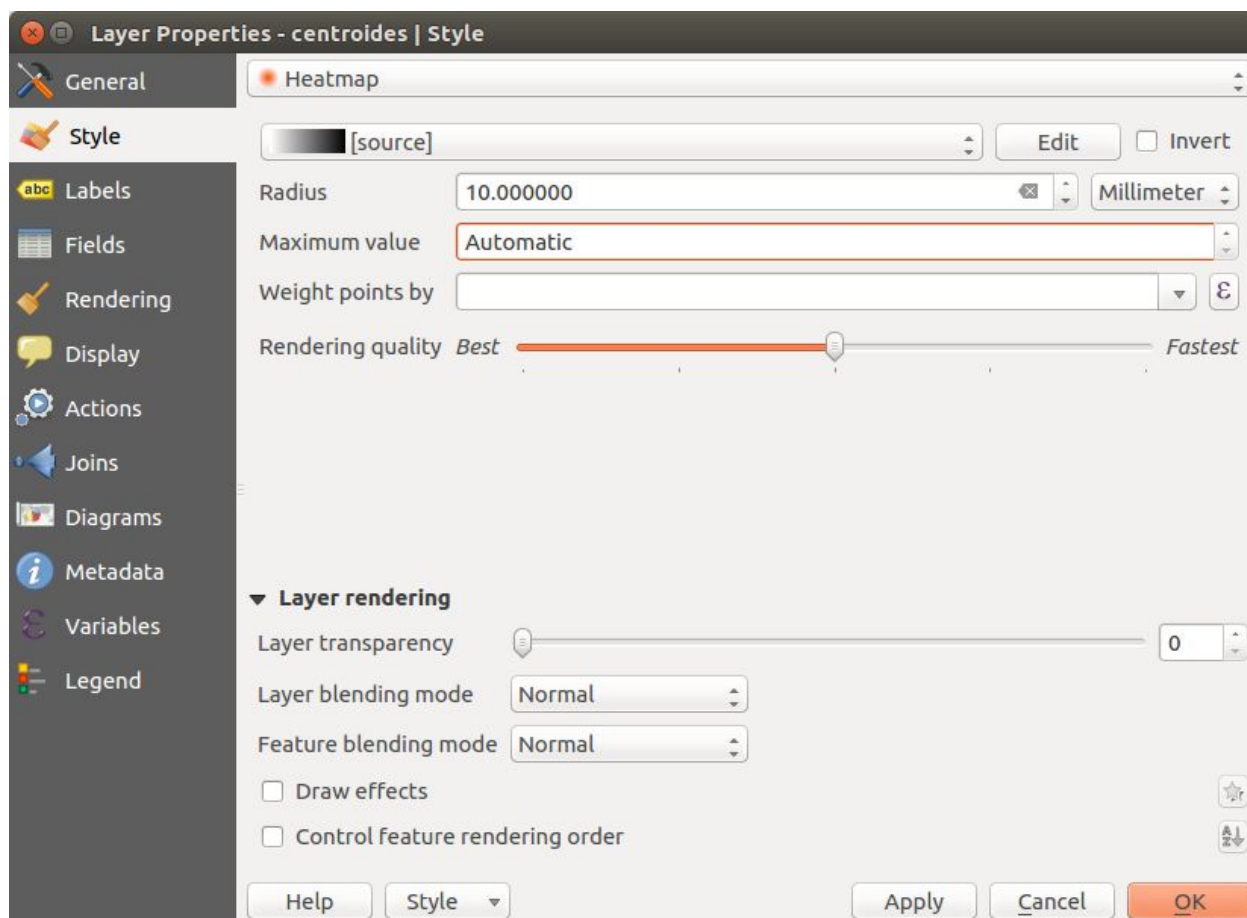
Para poder realizar un heatmap primero debemos modificar nuestras geometrías. Como estamos utilizando radios censales, estamos trabajando con polígonos. Los heatmaps, al estimar densidad de puntos, necesitan puntos para trabajar. Por eso realizaremos la transformación de polígonos a puntos. Para ellos utilizaremos el centroide de cada polígono para representar al radio. Para obtener los centroides elegimos en el menú:

Vector -> Herramientas de geometría -> Polígonos a centroides



El producto de esto es un nuevo archivo vectorial con puntos en lugar de polígonos y los mismos datos en la tabla. Siempre es conveniente guardar la capa en un archivo separado con la misma proyección de coordenadas en la cual decidimos trabajar. Una vez que tenemos los puntos, podemos crear el heatmap.

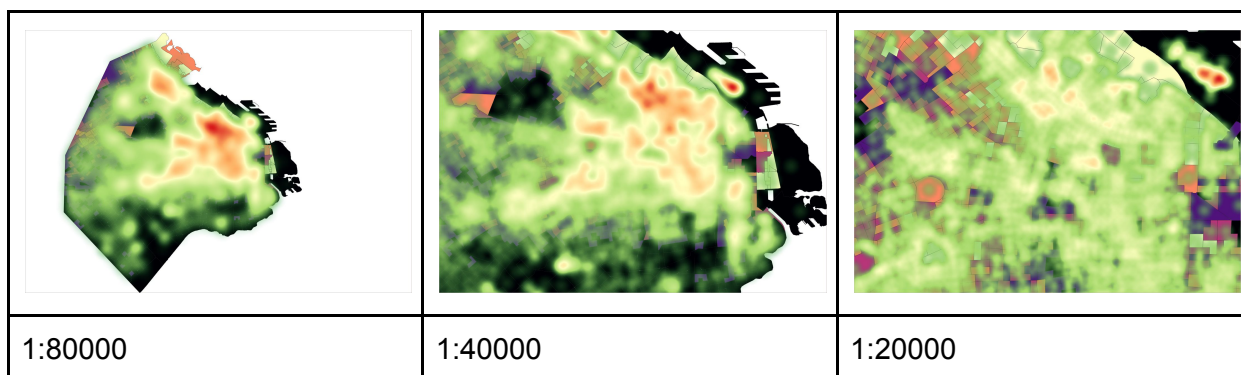
Para ello vamos a las *Propiedades de la capa* y elegimos Heatmap.



Este primer menú ofrece los elementos rudimentarios para construir un heatmap: solo dos de los 3 parámetros indispensables:

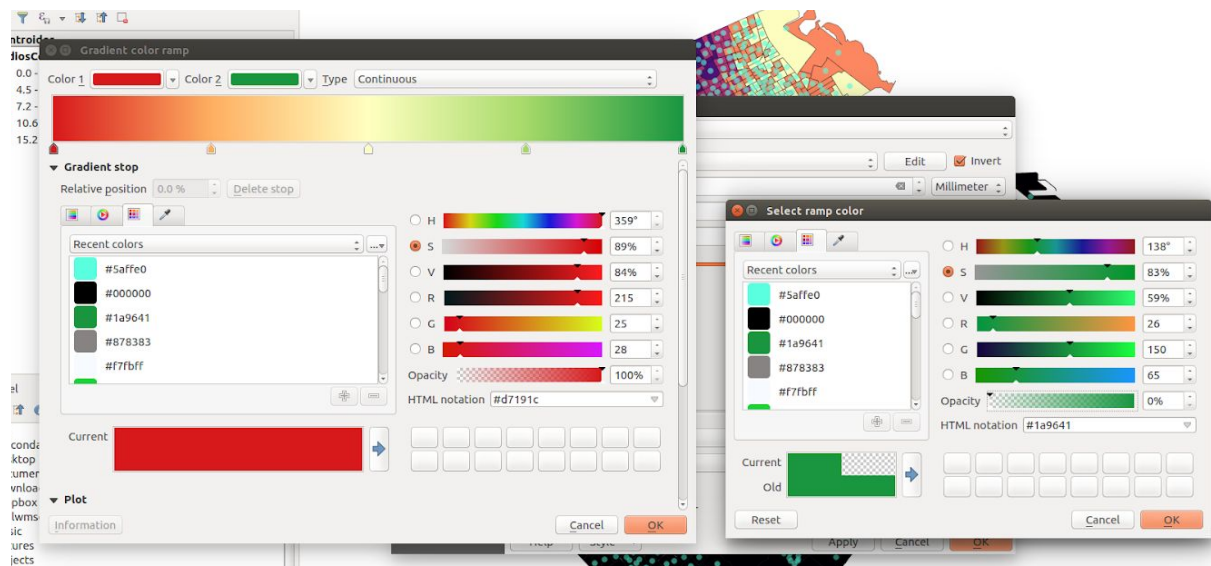
- Radios de búsqueda
- Tamaño de celda

Radios más grandes arrojarán superficies más suaves y menos precisas o granulares y radios más chicos viceversa. Estos radios pueden establecerse en unidades del mapa, milímetros o pixeles. Esto es fundamental dado que cualquier otra que no sea unidades del mapa hará al heatmap sensible no solo al radio de búsqueda sino también a la escala de visualización.

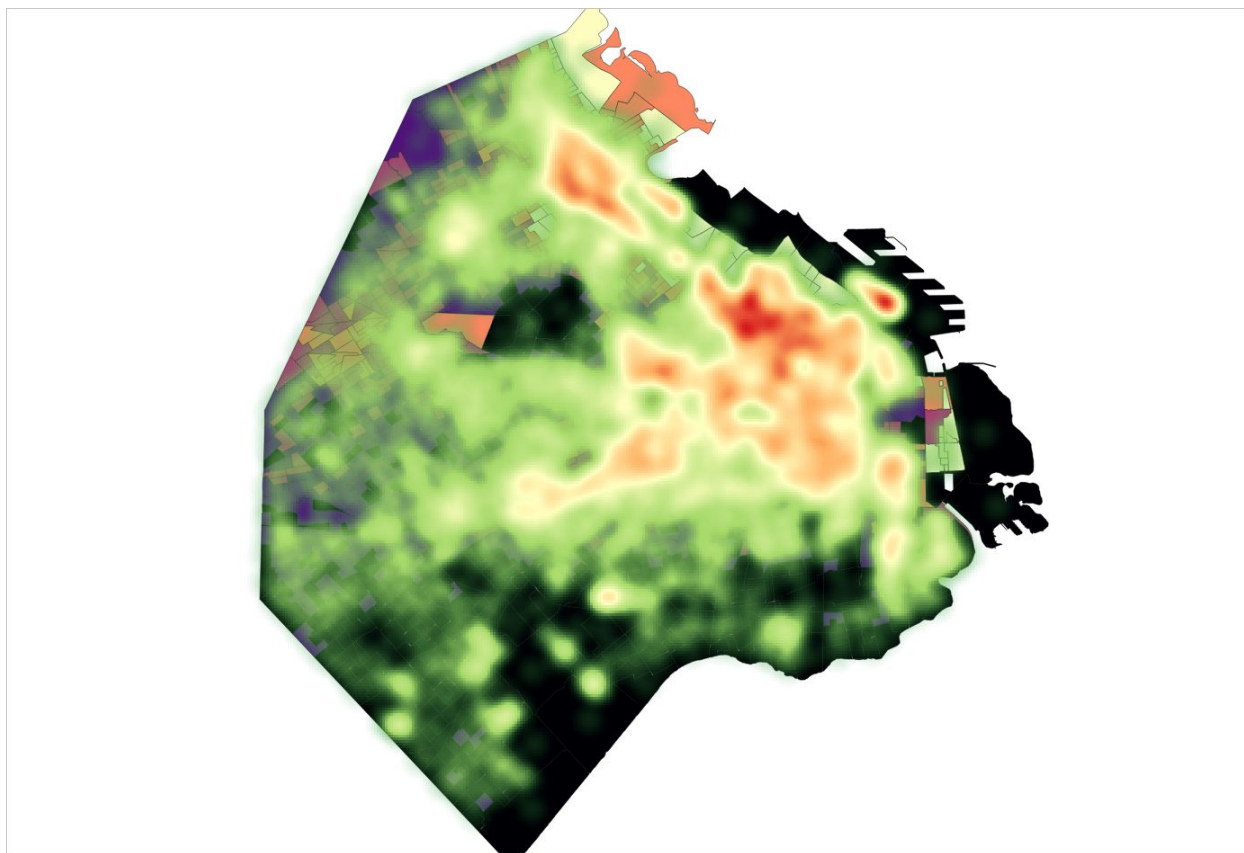


A su vez permite establecer el valor máximo de puntos dentro de un radio de búsqueda o determinar automáticamente este valor.

También permite elegir una paleta de colores. Elegimos la paleta divergente que va de los rojos a los verdes. Dado que los verdes serán los valores más altos, queremos invertirla haciendo click en el *checkbox* que lo habilita. El problema de dicha paleta es el nivel de transparencia del valor más bajo. Esto hará que el valor más bajo de densidades (es decir, allí donde no haya puntos) estará ocupado plenamente por el color verde, ocultando la capa de fondo. Para ello elegimos editar la paleta y cambiamos el nivel de transparencia del verde.

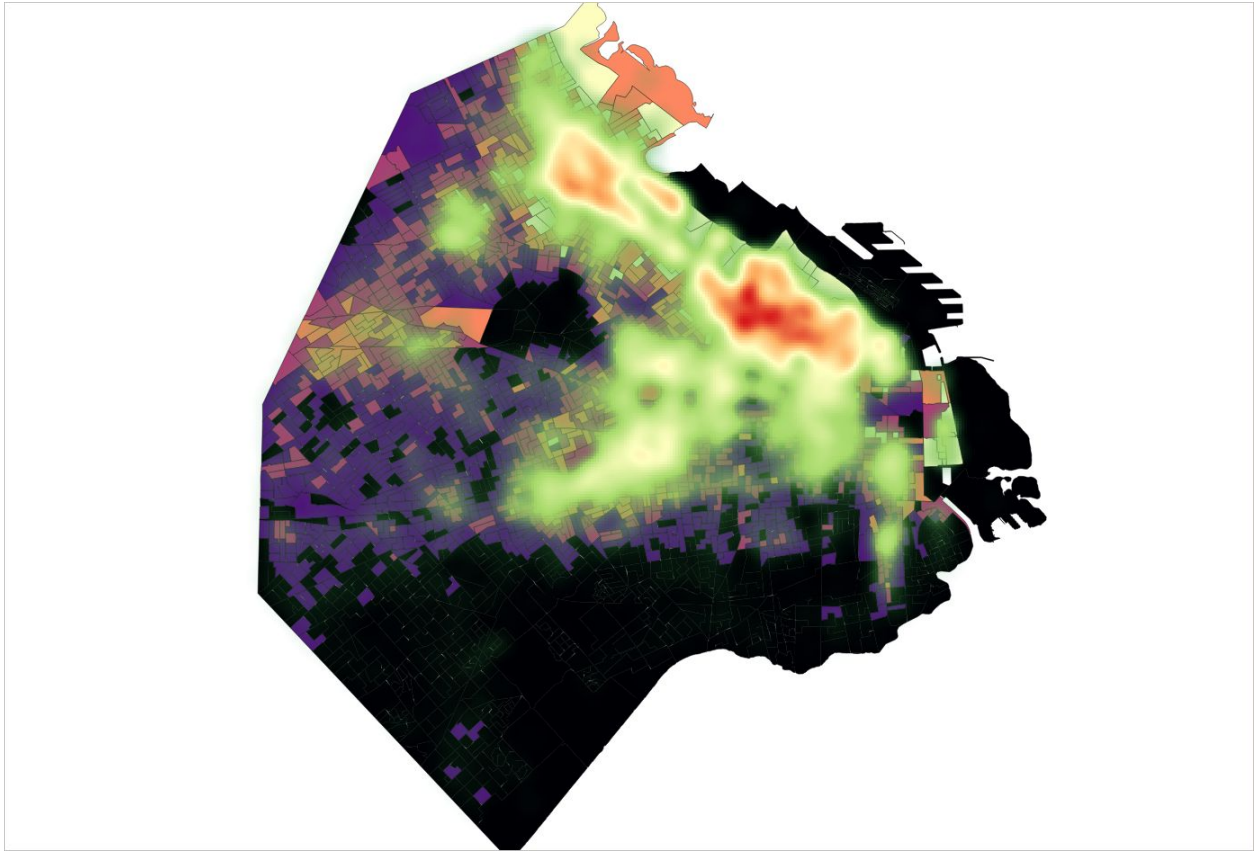


Elegimos 500 metros (en unidades de la proyección elegida) y este es el producto final.

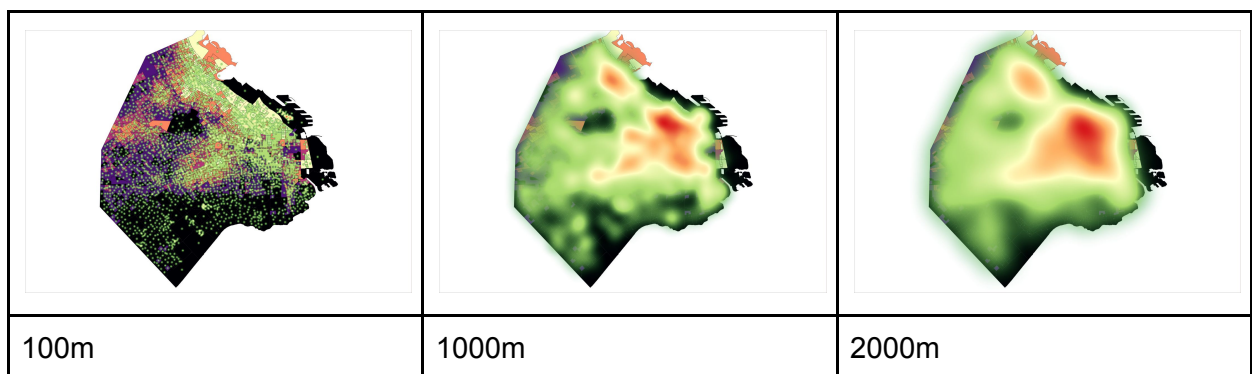


Como fue discutido, el mapa de calor muestra densidades de puntos, no valores de la variable. Como la densidad poblacional se comporta de manera similar a la distribución en el espacio del nivel educativo del/de la jefe/a de hogar, parecería que el heatmap mapea lo mismo, pero no es así. Esto se ve claramente en el caso de la villa 31 que aparece como un punto de calor cuando vimos en la coropleta que su valor se corresponde con el primer quintil de la distribución de nivel educativo del/de la jefe/a de hogar (el fondo es de color negro).

La opción por defecto de QGIS permite ponderar los puntos por un valor de la variable. Si escogemos la variable de % de jefe/a de hogar con universitario completo o más (*unicComp* en el shapefile del repositorio) se puede observar que el mapa cambia, y la villa 31 desaparece. Si bien esto parecería indicar una aproximación al objetivo que buscamos (detectar clusters de valores altos y bajos de una variable) no es así. Por un lado no permite detectar los valores bajos. Por otro, recordemos que simplemente es una interpolación de una superficie en base a densidades, sin ofrecer información sobre la significancia estadística de los resultados.



A su vez, es muy sensible a los parámetros escogidos de radio de búsqueda y tamaño de celda (y al tipo de función de kernel utilizada).



Por último, el resultado es un objeto de tipo Raster. Si bien se puede guardar el resultado en una imagen e incluso preservar el significado de los colores en valores de la densidad de la variable, utilizando el Plug in de QGIS Heatmap, esto ya escapa los objetivos de este curso. Lo que queríamos destacar es que el *heatmap* no realiza un análisis de puntos cálidos y fríos (hot spot - cold spot) ya que se basa en densidades de puntos y no valores de la variable. Existen métodos de interpolación (kriging, etc) para realizar análisis similares, del mismo modo que el PlugIn heatmap de QGIS permite un manejo mucho más preciso y elaborado de la técnica.

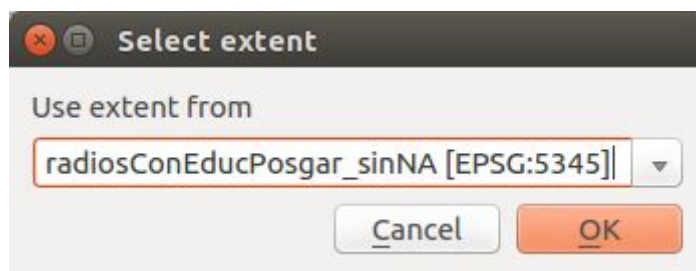
2 - Grillas

Para simular un mapa de calor manteniendo cierto control sobre algunas variables (y para comprender mejor el funcionamiento del mismo) utilizaremos grillas. Las grillas es una forma de subdividir el espacio de análisis en nuevas geometrías de tamaño constante. El tamaño de las mismas debería estar dictado por la teoría. Si por ejemplo deseamos contar cantidad de accidentes de tránsito, y sabemos por la teoría que los mismos se concentran en las esquinas y extienden su radio de influencia 50 metros hacia la mitad de la cuadra, la configuración ideal de grilla sería aquella que tiene el centro en las esquinas y miden 100 metros por 100 metros. Si bien no siempre se podrá hacer coincidir el centro de la celda con las esquinas (existen modos de subdividir áreas irregularmente a partir de teselación para lograr esto pero perdiendo el rango de influencia constante de 50m), se consideraría importante mantener celdas de 100m x 100m. También es posible que se parta de la noción de que se quiere subdividir el espacio en N celdas y ajustar el tamaño de las celdas acorde a eos.

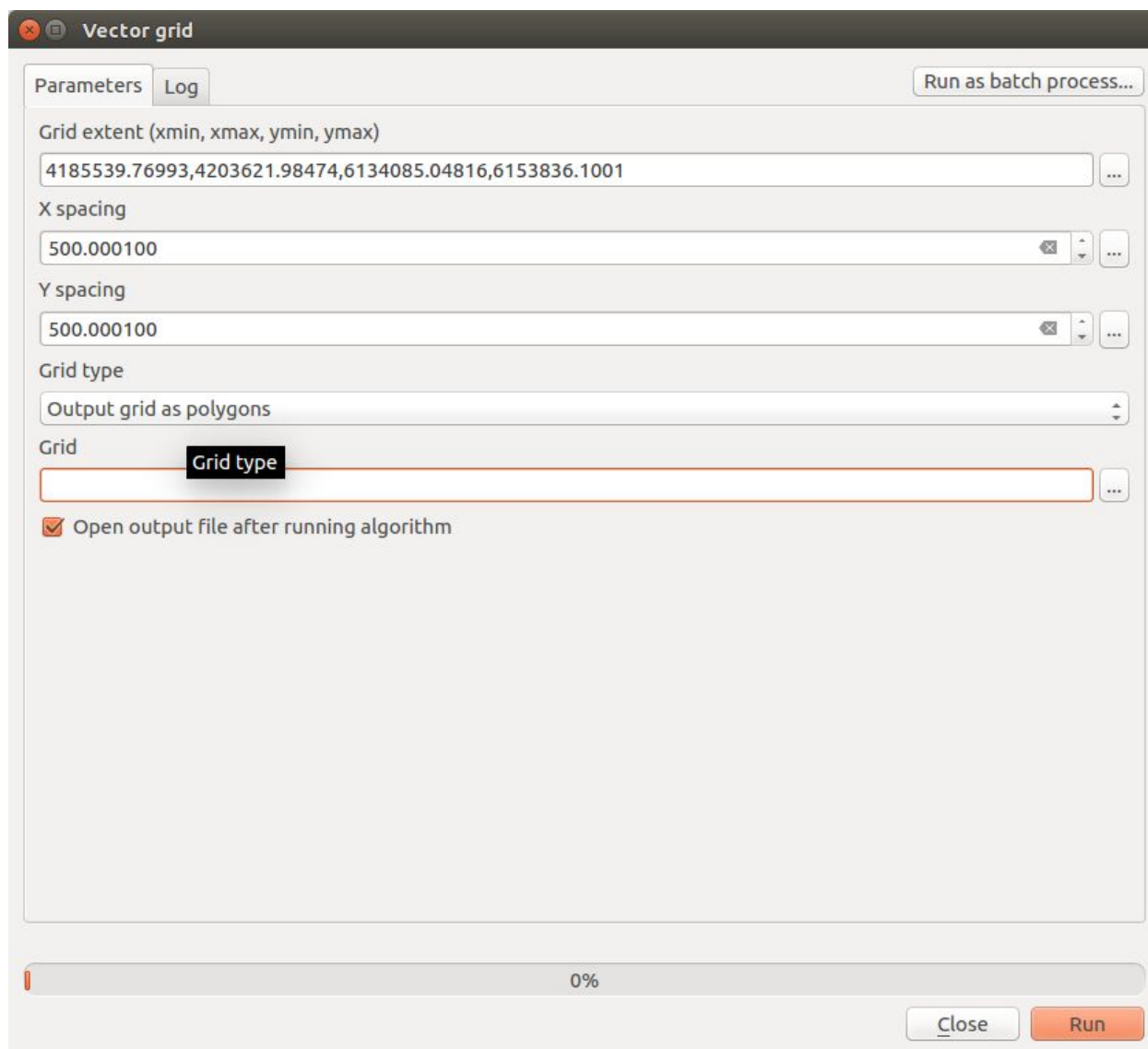
Realizaremos un grillado para la Ciudad de Buenos Aires utilizando diferentes tamaños de celda. Para ello iremos a Vector -> Herramientas de investigación -> Vector Grid

Allí debemos elegir la extensión de la grilla, es decir que tamaño de la grilla en sí. La idea es que cubra el total de la Ciudad de Buenos Aires, por lo cual tomaremos esa extensión desde la capa que tenemos con los radios (radiosconEducPosgar_sinNA en el repositorio).

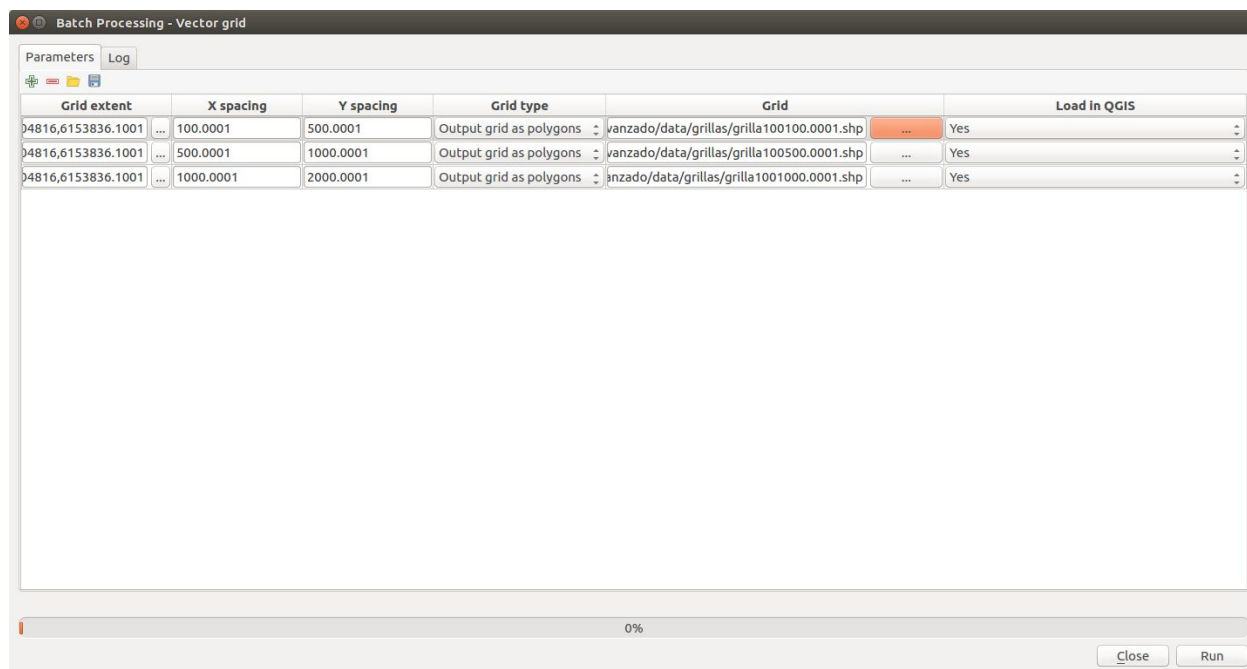




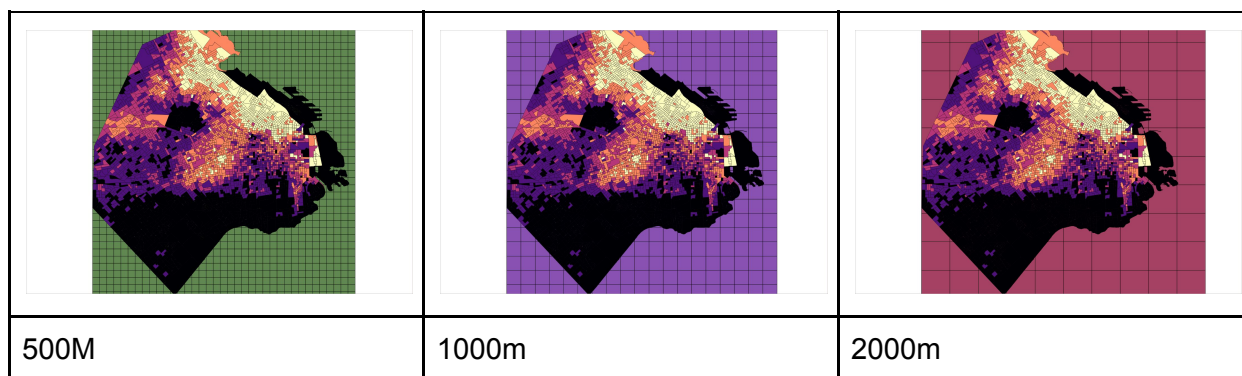
Luego elegimos el alto y ancho de las grillas y si queremos obtener un polígono donde cada celda es una entidad o en cambio queremos líneas. Los diferentes tipos de objetos espaciales permiten diferentes tipos de operaciones. Nosotros elegimos que el output sean polígonos.



También se puede correr este proceso en tandas o batch generando automáticamente diferentes grillas con diferentes tamaños de celdas. **Al momento de crearla y guardarla es vital chequear el sistema de proyección de coordenadas de la grilla.**



El resultado es un grilla sobre la Ciudad don diferentes tamaños de celda



Muchas de estas celdas no tiene relación espacial alguna con la Ciudad. Por lo tanto es mejor eliminarlas del análisis. A su vez hace todo el proceso más performante en términos de eficiencia computacional. Si bien para este ejemplo, no existen muchos datos, para algunas aplicaciones con grandes datos estos procesos pueden ser muy costosos.

Para eliminar las celdas se puede optar por clippear o seleccionar por ubicación. Ambos eliminan las celdas que no se superponen con la ciudad, pero difieren en el tratamiento de aquellas celdas que tienen una porción dentro y otra fuera de la Ciudad. Mientras que clip recorta la porción sobrante, la selección por ubicación la mantiene.

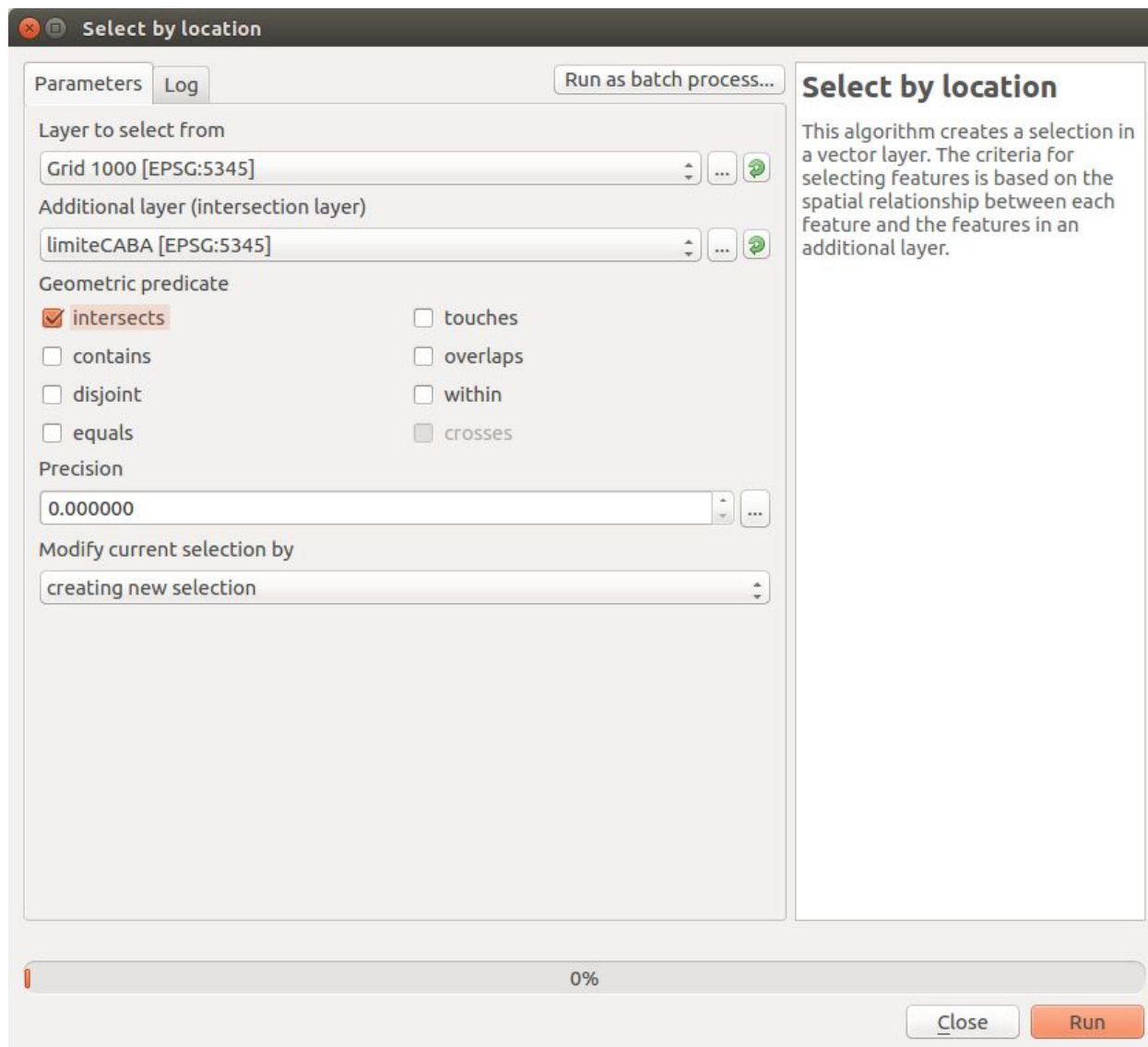
Para recortar se puede utilizar el resultado de disolver los polígonos de comunas que se realizó en la Clase 2. En el repositorio este se encuentra en *limiteCABA*.

Selección por ubicación

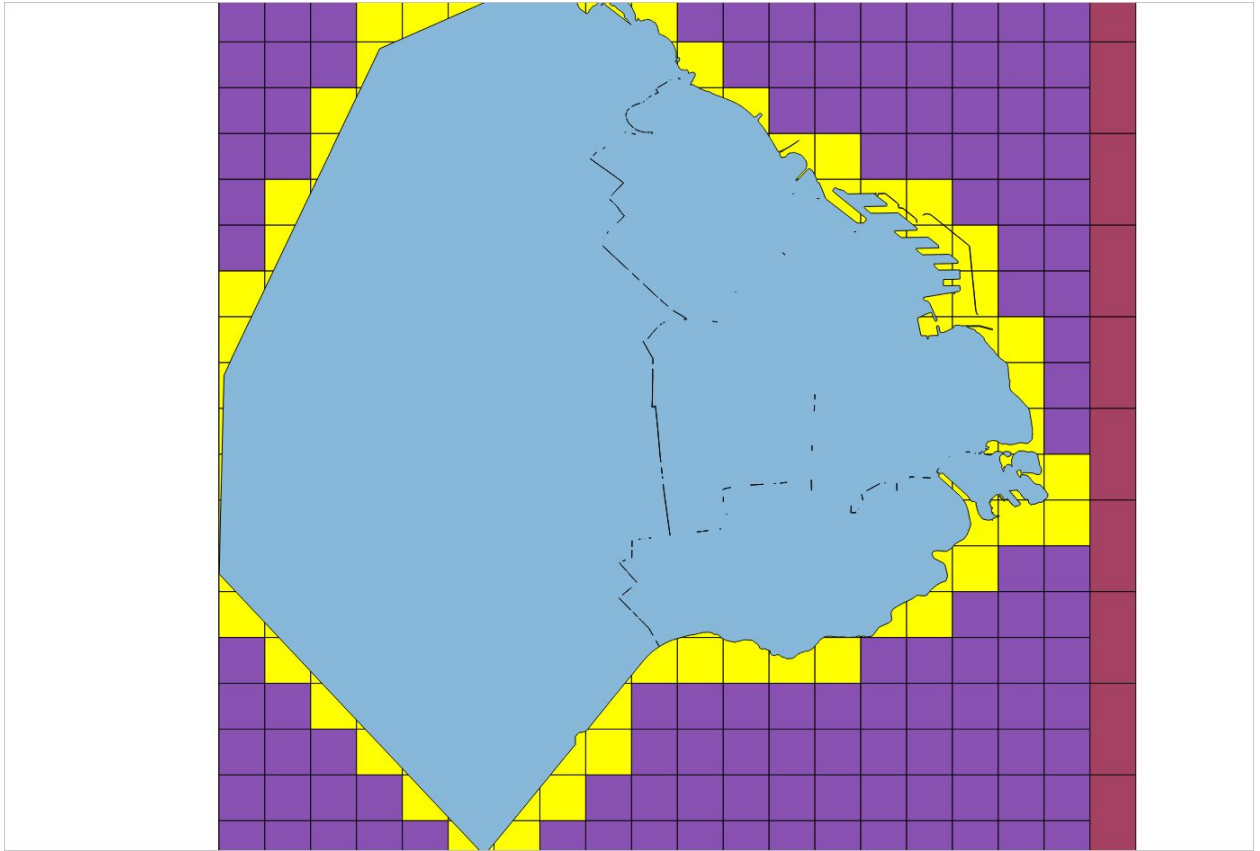
Elegimos Vector -> Herramientas de investigación -> Selección por ubicación.

Seleccionamos la capa de la cual queremos elegir y la capa que hace de intersección.

Seleccionamos “intersección” como operador espacial y aceptamos.



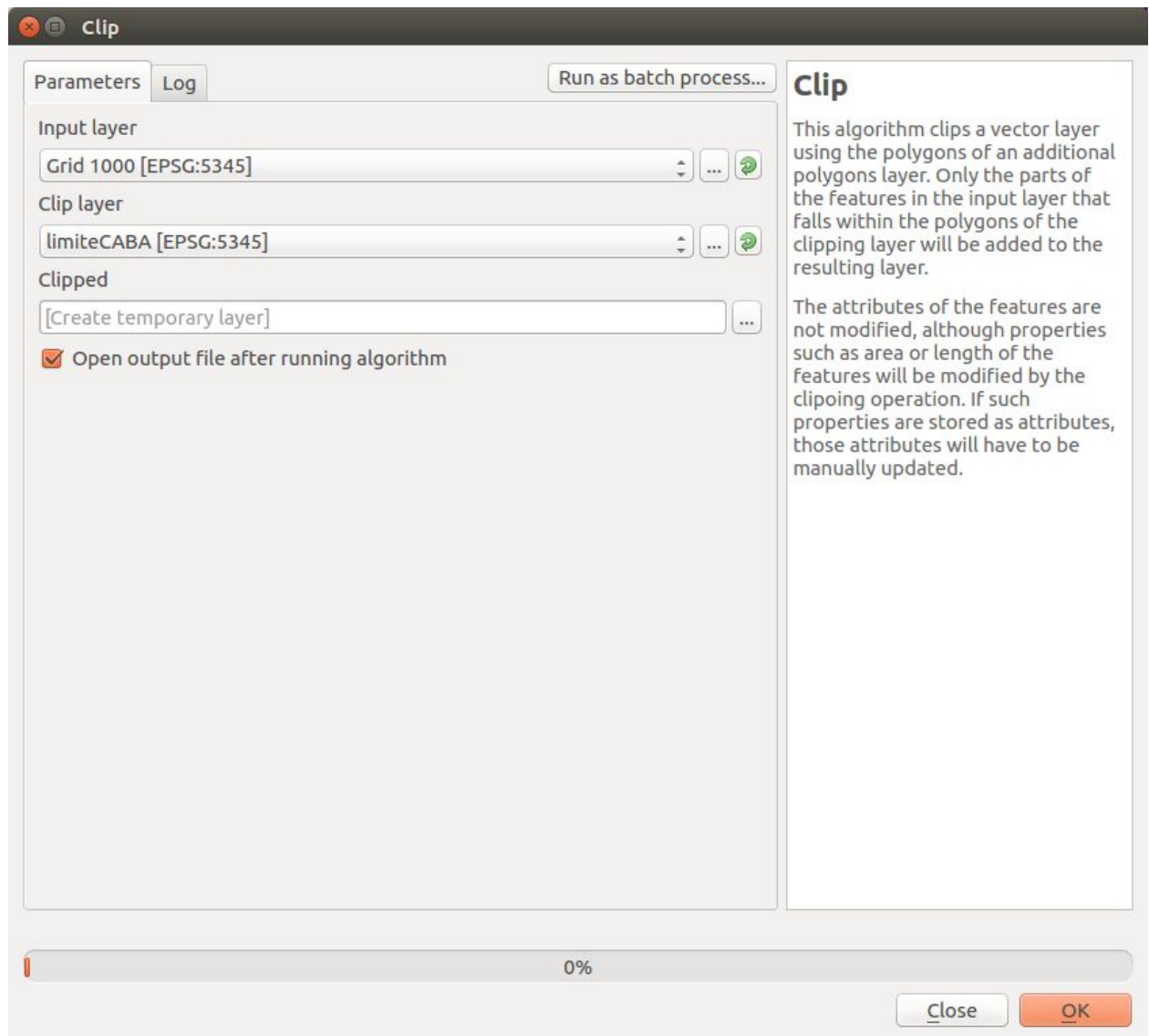
El resultado es una selección de los polígonos grilla que tocan la Ciudad. Algunos de ellos, solo por apenas una pequeña porción, por ejemplo los ubicados en el Sur-este en La Boca o la reserva.



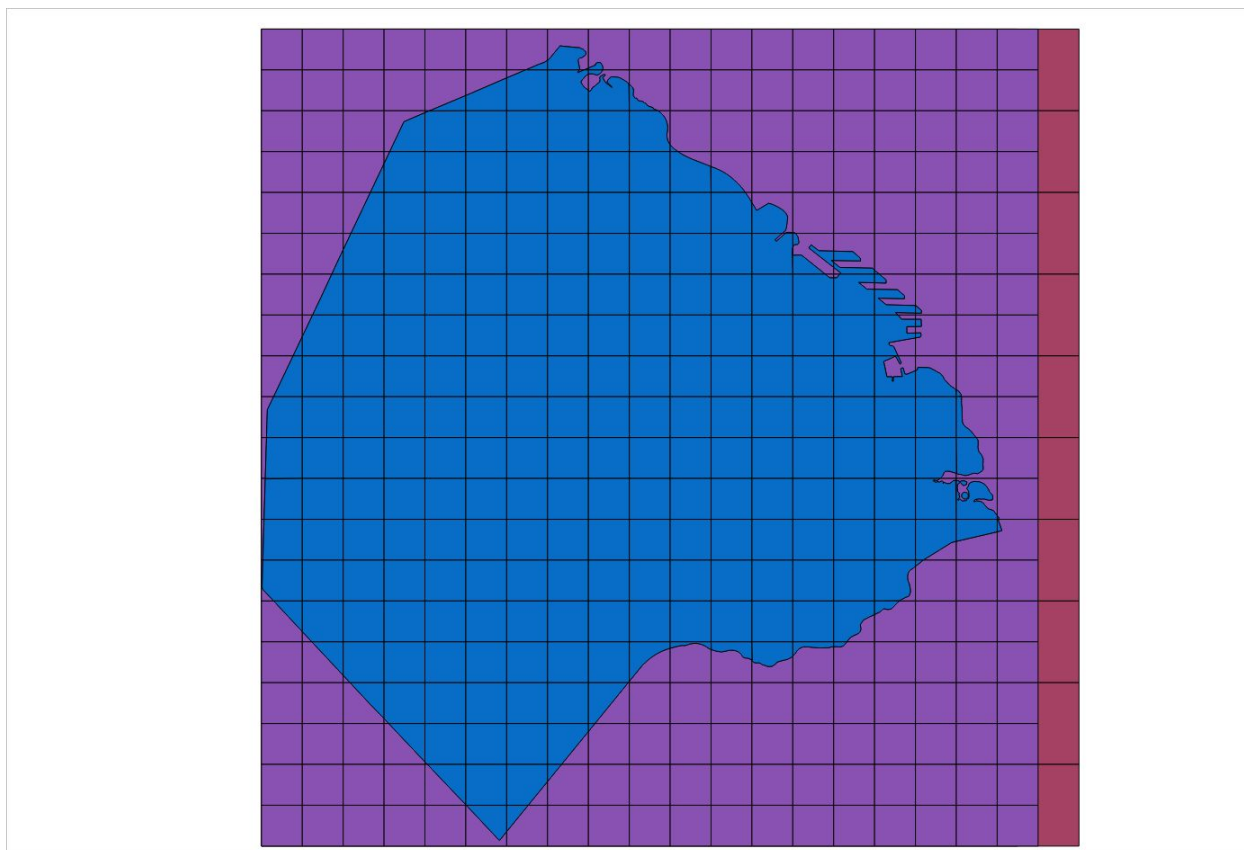
Solo queda guardar la selección en un shapefile separado, siempre teniendo cuidado de la proyección.

Clip

En Vector -> Herramientas de Geoproceso



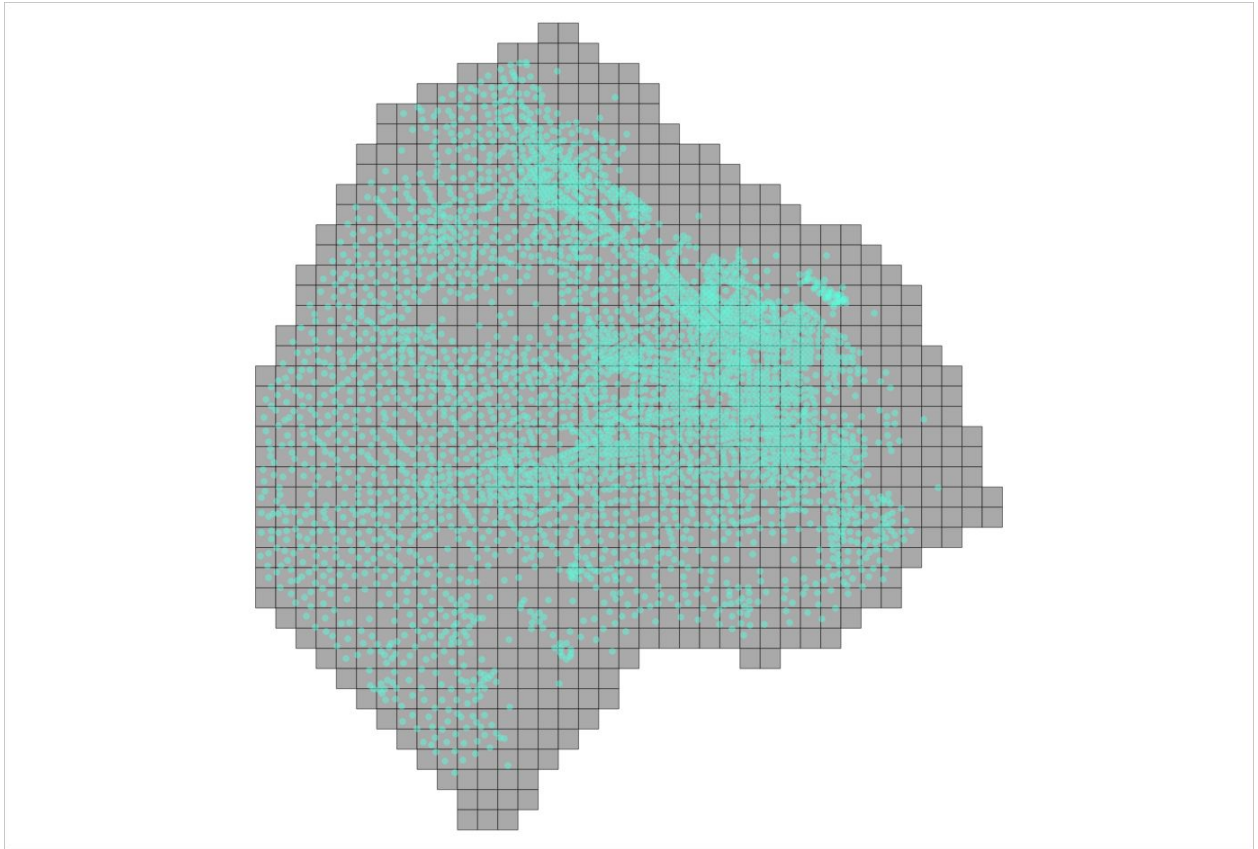
El resultado es las mismas grillas solo que recortadas allí donde una porción de la misma no coincidía con la Ciudad.



Tanto Selección como Clip pueden correrse en tandas.

3 - Uniones espaciales a grillas

A continuación vamos a poblar esas grillas con estadísticos basados en nuestros puntos. Recordemos que esto es una conversión de polígonos (radios) a puntos, con lo cual se encuentra especialmente afectado al Problema de la Unidad Espacial Modificable. Para mover información de una capa a la otra, nos basaremos en el espacio y utilizaremos Uniones Espaciales.



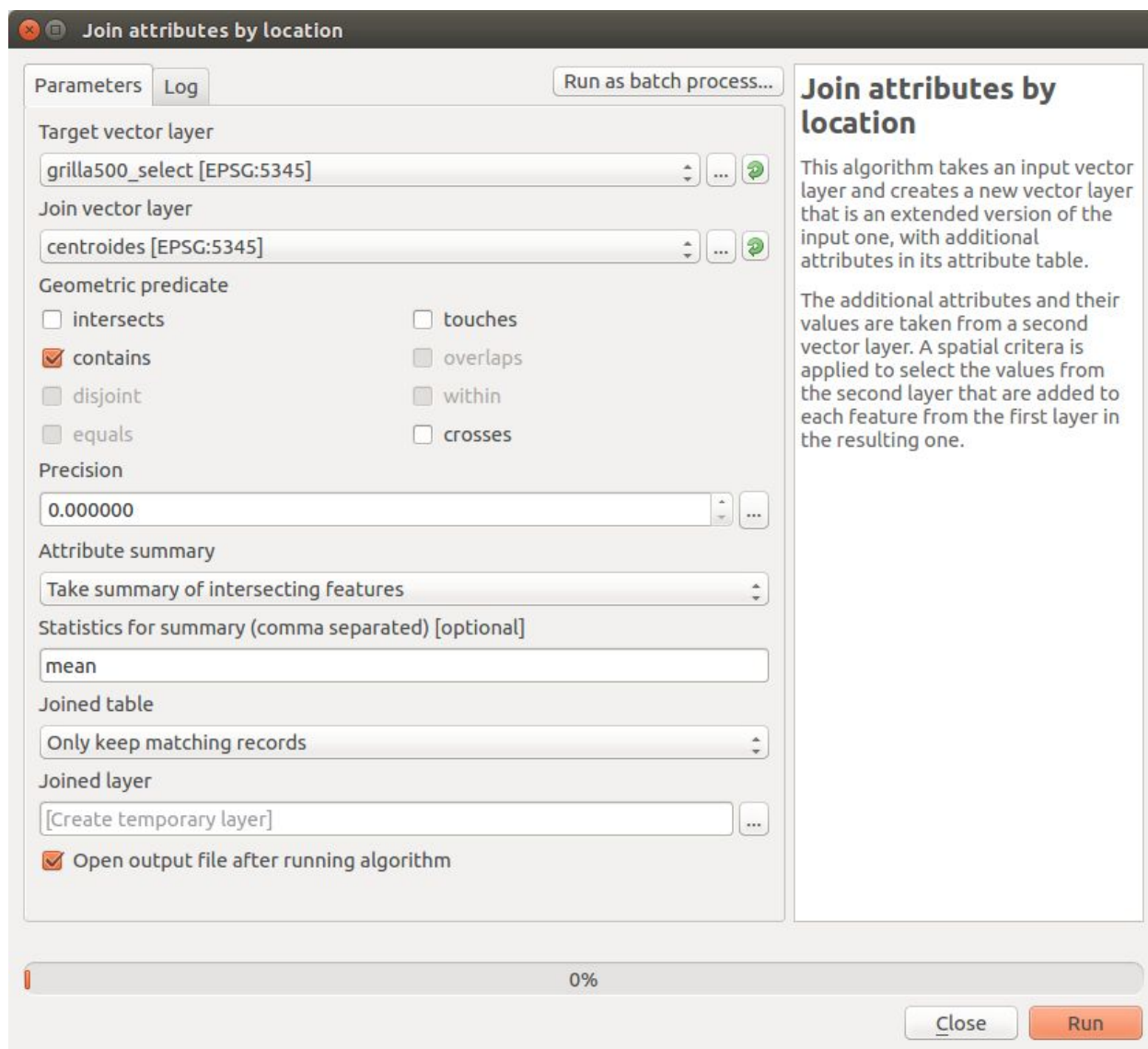
Para realizar una *Unión espacial*, nos dirigimos a Vector -> Herramientas de manejo de datos -> Unir atributos por ubicación. Allí elegimos como capa destino o *target* a nuestra grilla (es la capa a la que queremos transferirle la información). La fuente de información es la capa con los centroides. Recordemos que solo cambió el tipo de objeto geográfico (que polígono a punto) pero la tabla de datos permanece igual. Elegimos como tipo de operador geográfico “*contiene*” ya que un punto puede pertenecer a un y sólo un polígono.

Esta es una manera efectiva aunque poco realista de reconfigurar los datos de un polígono (radios) a otro (celda de grilla) y se encuentra especialmente afectado por el Problema de la Unidad Espacial Modificable, en su aspecto de zonificación. Técnicas más avanzadas permitirían aporcionar el polígono a partir de información contextual como puede ser datos catastrales sobre el reparto de viviendas al interior de un radio censal. Pero dado que no es el objetivo principal del taller, elegimos seguir adelante con esta técnica menos sofisticada.

Deseamos tomar un resumen estadístico de todos los puntos que caen dentro de la celda de la grilla, no solamente el primer caso. Especificamos además que solo queremos el promedio (el conteo de casos lo hace por defecto).

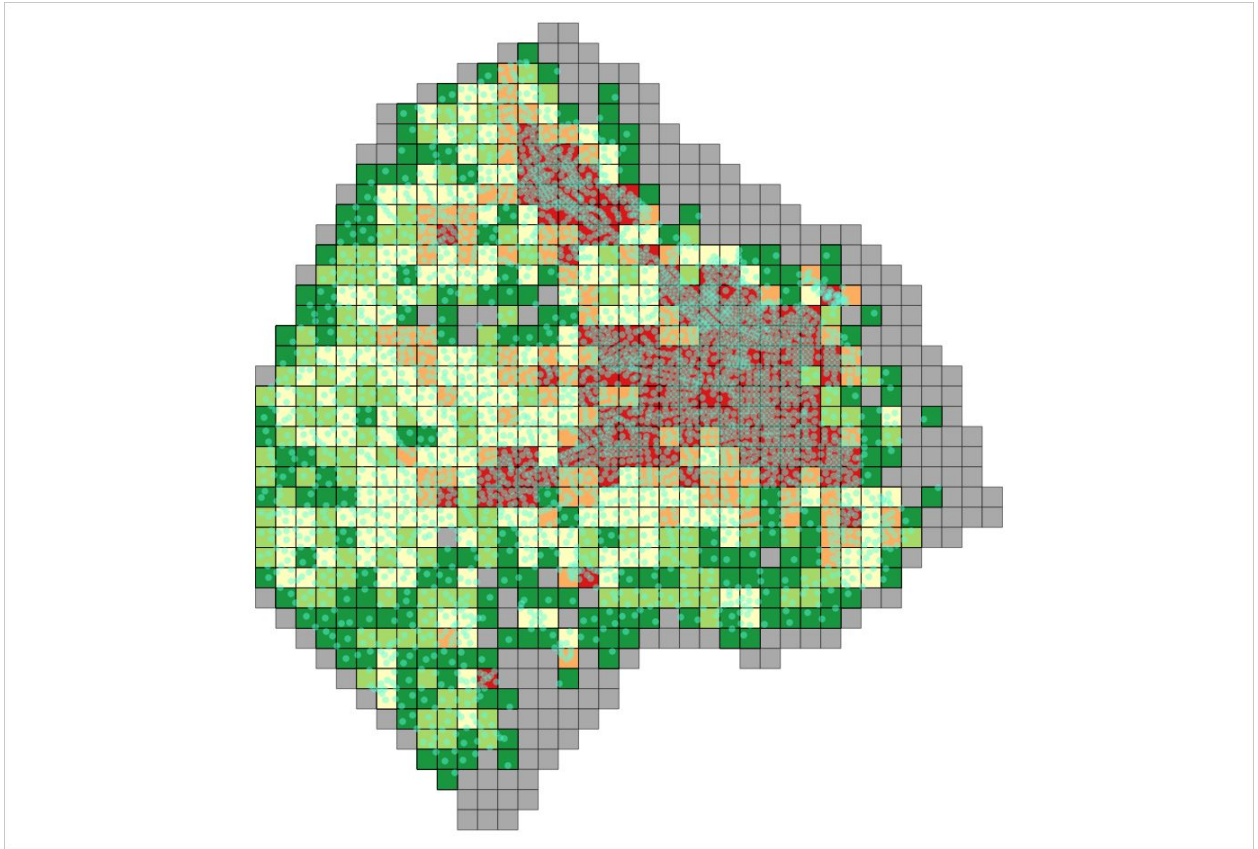
Tenemos la opción de mantener registros que coincidan o todos. Mantener todos los registros mantendrá las celdas donde no recae ningún punto centroeide, mientras que mantener solo los registros que coinciden eliminará estas celdas de la grilla. El efecto de esto es al calcular estadísticos sobre un dominio donde existen celdas con información faltante o que puede tener 0 como dato (por ejemplo, las grillas que no contienen puntos pueden tener 0 como conteo y

faltante como promedio). Como observamos en el caso de los radios censales en la Clase 1, es importante reflexionar sobre la naturaleza de la información faltante y su impacto en el análisis.



3a - Con conteos

Para terminar podemos recrear un mapa de calor visualizando la cantidad de puntos que hay en cada celda al realizar una coropleta con el campo “count” en la capa resultante de la Unión Espacial. Elegimos la misma paleta de colores que el *heatmap* para mostrar la similitud. Esta es una visualización de datos discreta, no continua, o también se podría decir con “tamaños de pixeles grandes” (eventualmente el heatmap discretiza sus valores al darle un valor único a cada pixel del raster).



3b - Con promedios

Sin embargo, es evidente que esta coropleta nada dice de los niveles educativo del jefe de hogar promedio del radio. Pero dado que obtuvimos el promedio de dicha variable para los radios que caen dentro de cada celda en la Unió Espacial, podemos realizar una coropleta que se acerque a ese valor. Recordemos que aquí también juega un papel el PUEM en su faceta de escala, dado que el promedio de los radios dentro de la celda, puede ocultar la heterogeneidad subyacente como se observó en la Clase 2.

Mapeando con la variable del promedio del nivel educativo se obtiene un mapa diferente:

