Taller de análisis espacial avanzado

Clusters espaciales y hotspots con SIG - Clase 2. El problema de la unidad espacial modificable

Del Rigor en la Ciencia
Jorge Luis Borges

En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del Imperio, toda una Provincia. Con el tiempo, estos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el tamaño del Imperio y coincidía puntualmente con él.
Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas.
Suárez Miranda, Viajes de Varones Prudentes, Libro Cuarto, Cap. XLV, Lérida, 1658.

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

Paso a paso clase 2 - El problema de la unidad espacial modificable (PAUM)

Todo análisis estadístico que implica en algún momento utilizar, precisamente, algún estadístico. El mismo se computará a lo largo de un índice i que representa nuestras unidades de análisis y/u observación. Por ejemplo la sumatoria o media aritmética:

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + a_{m+2} + \cdots + a_n$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

El análisis espacial tiene una particularidad. Las unidades de análisis u observación son recortes arbitrarios del espacio. Todo espacio puede ser recortado de infinitos modos y el recorte arbitrario escogido es una opción entre infinitas posibilidades. Estas decisiones arbitrarias son ineludibles en todo proceso científico, pero se debe ser consciente de sus implicaciones en la construcción del conocimiento. En el análisis estadístico, los indicadores utilizados pueden ser muy sensibles a estas arbitrariedades, arrojando un resultado u otro ya se utilicemos un tipo de recorte u otro. De este modo, pueden introducirse sesgos en el análisis. Esto es especialmente cierto para determinados estadísticos utilizados en el análisis espacial. Este problema ha sido denominado el Problema de la Unidad Espacial Modificable o MAUP por sus siglas en inglés (Modifiable Areal Unit Problem) (Openshaw & Taylor, 1979).

Los datos

En este taller vamos a aproximarnos a este problema mediante el análisis de los hogares de la Ciudad de Buenos Aires, deteniendonos en nuestra variable de interés: el nivel educativo de los jefes de hogar. La fuente de esta información es el Censo Nacional de Población, Hogares y Viviendas de la Argentina del año 2010. El mismo no nos permite obtener información de los hogares tomados individualmente en virtud de la Ley 17622 de Secreto Estadístico. Es por eso que el INDEC ofrece información para diferentes agregados: radios, fracciones, departamentos, provincias.

De este modo, sólo podemos saber las características de los hogares que nos interesa analizar a partir de estadísticos computados para esos agregados. Observamos a los hogares a través de esos agregados. Lo que pudimos obtener del censo es el total de hogares de cada radio, y el total de hogares de cada radio con jefe de hogar con universitario completo o más. Con esto, computamos el porcentaje de hogares con jefe con universitario completo o más.

Esto da pie a la pregunta: todos los hogares de ese agregado se comportan del mismo modo en lo que respecta a la variable de interés? Al trabajar con datos agregados que no pueden ser desagregados con mayor granularidad, asumimos que ese dato es representativo de los hogares de dicho radio e incurrimos en lo que se denomina Falacia Ecológica. Esta falacia consiste en hacer inferencia sobre casos individuales (hogares) a partir de datos agregados (radios). *Es especialmente problemático cuando dicho agregado no es homogéneo a su interior para la variable de interés*. Idealmente desearíamos que existiera un recorte del espacio tal que minimice esa heterogeneidad interna del radio y a la vez proteja la privacidad de los datos de los hogares individuales. Sin embargo, dado que esa heterogeneidad u homogeneidad es relativa a la variable de interés (nivel educativo del jefe de hogar) debería haber tantas unidades espaciales como variables de interés.

El INDEC establece la creación de unidades administrativas para la recolección de información con un criterio administrativo basado en una lógica de los recursos humanos utilizados en la recolección de datos para el Censo. De este modo, radios, fracciones, etc no constituyen verdaderas geografías con un significado intrínseco sobre las poblaciones subyacentes y que reflejen el proceso social subyacente que se desea analizar. Debemos utilizar este recorte del espacio y no otro, y aceptar el problema que esto conlleva: el problema de la unidad espacial modificable. Pero debemos ser conscientes de sus limitaciones e implicancias en nuestros análisis.

El problema de la unidad espacial modificable

Este problema podemos dividirlo en sus dos componentes principales:

El efecto de escala: El análisis estadístico basado en datos agregados en áreas de diferentes tamaños producirá resultados diferentes.

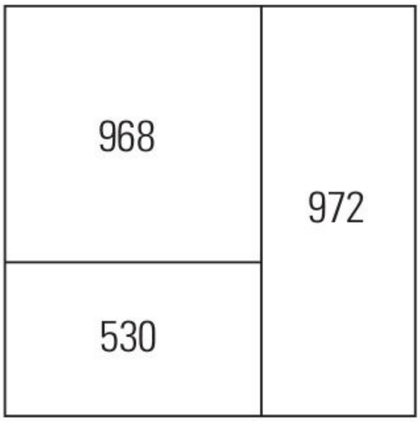
El **efecto de zonificación**: Dos zonas de igual área y forma, pueden producir diferentes resultados de acuerdo a su ubicación.

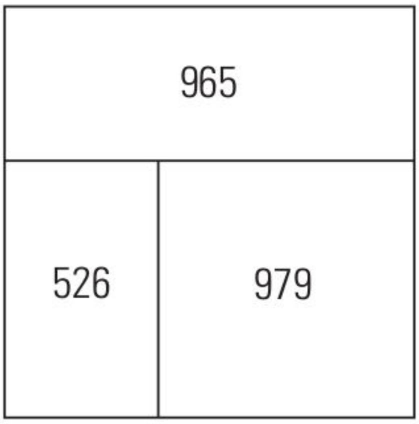
Efecto escala

630	651
548	641

174	162	169	161
150	144	161	160
131	127	162	163
142	148	158	158

Efecto zonificación



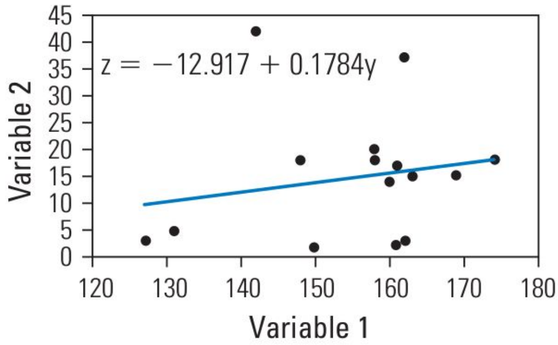


Variable 1

174	162	169	161
150	144	161	160
131	127	162	163
142	148	158	158

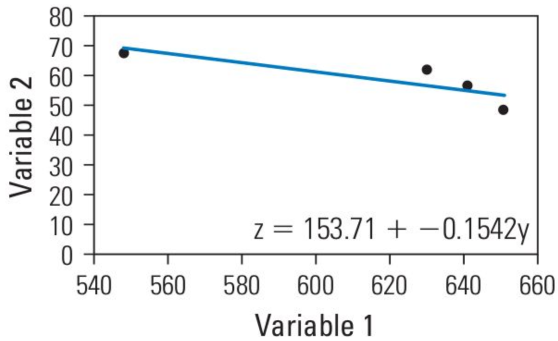
Variable 2

18	37	15	17
2	5	2	14
5	3	3	15
42	18	18	20



630	651
548	641

62	48
68	56



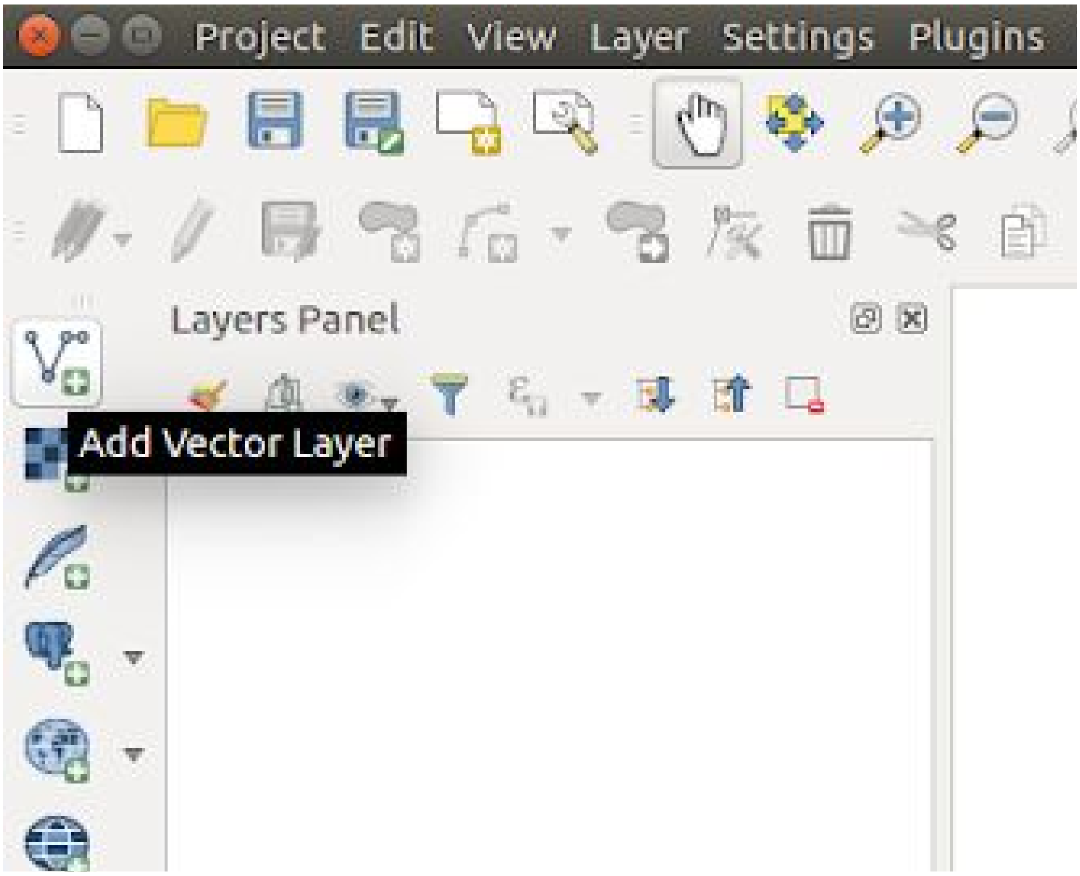
En la primer parte del taller observaremos las implicaciones del efecto escala, haciendo uso de la ventaja que implica que las unidades espaciales censales estén perfectamente anidadas, es decir que se contienen mutuamente. De este modo, podremos observar el estadístico de interés agregado para diferentes unidades y cómo se comporta en términos de heterogeneidad

En la segunda parte analizaremos en efecto zona al utilizar una zonificación no perfectamente anidada con las censales como lo son los circuitos electorales. Esto nos presentará con el problema de partir de unidades espaciales agregadas como radios como fuente de datos en lugar de hogares individuales que nos permitiría distribuirlos en las nuevas zonas (circuitos). Sin embargo, dado que partimos de unidades espaciales agregadas (radios) expresadas en GIS como polígonos, debemos aporcionar dicho polígono. Este aporcionamiento se puede realizar de diversos modos según el estadístico que estemos utilizando, sea una proporción o un total.

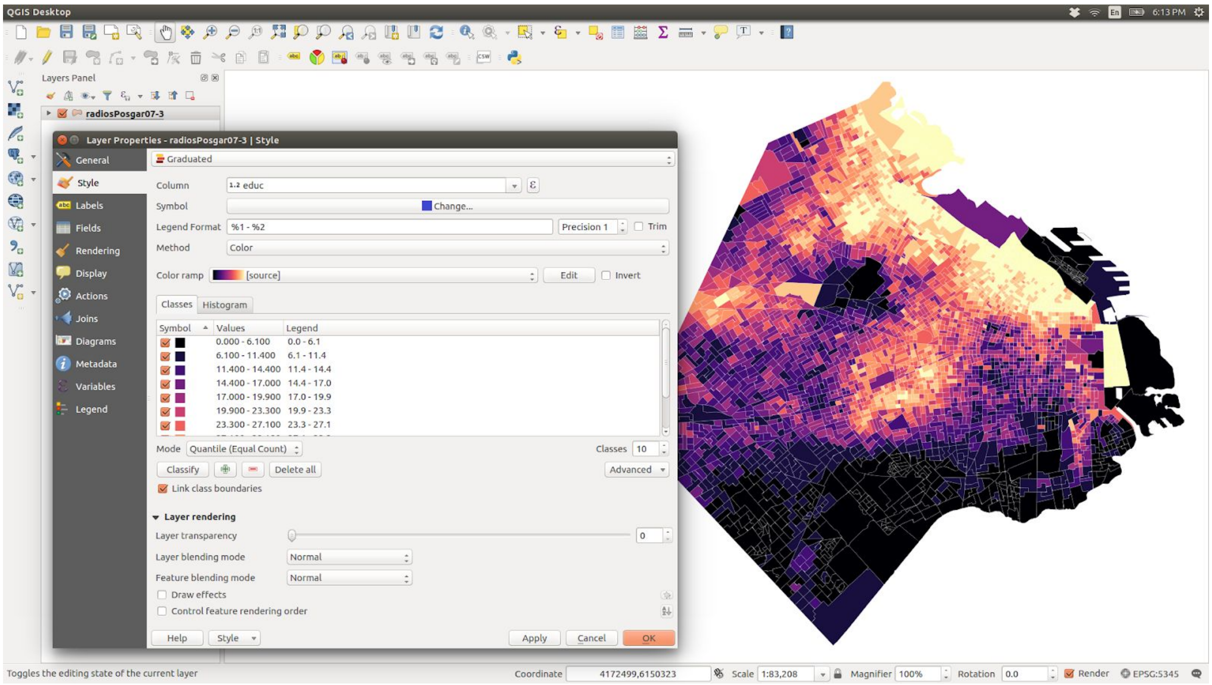
PRIMERA PARTE - EFECTO ESCALA

En esta parte vamos a mapear la variable Porcentaje de Jefe de hogar con Universitario Completo o Más. Luego procederemos a ir agregando esos datos para las unidades espaciales agregadas (fracciones, departamentos). En la CABA, los departamentos coinciden con las comunas.

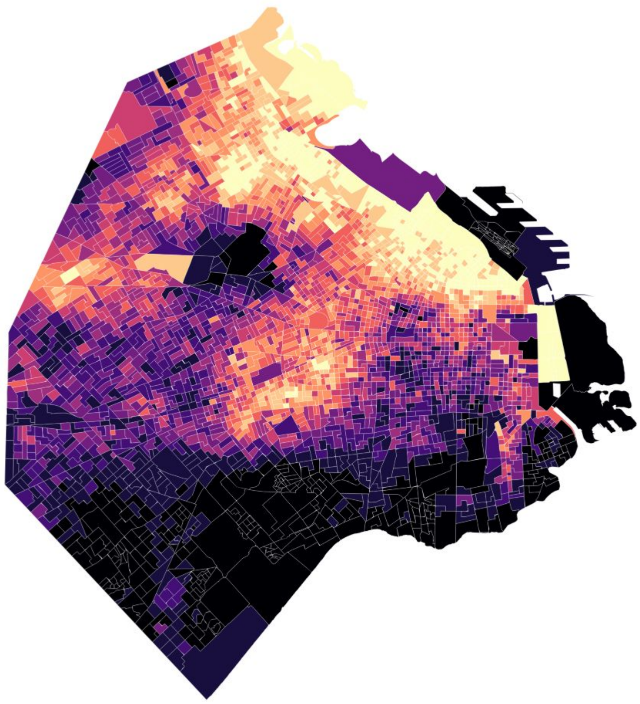
Abrimos el set de datos con los radios censales (unidad espacial más pequeña del censo) desde *Añadir capa vectorial*



Buscamos el shapefile en data/radios/radiosPosgar07-3.shp. El nombre indica la proyección en la que vamos a trabajar. Luego visualizamos los datos a partir de la variable *educ* que contiene el % de jefes de hogar con universitario completo o más por radio censal. Ese porcentaje fue calculado a partir de dividir la cantidad de jefes de hogar con universitario completo o más (*univCompMa*) por el total de jefes de hogar por cada radio (*Total*).



Esta es la distribución espacial de nuestra variable bajo análisis al nivel de radio censal.

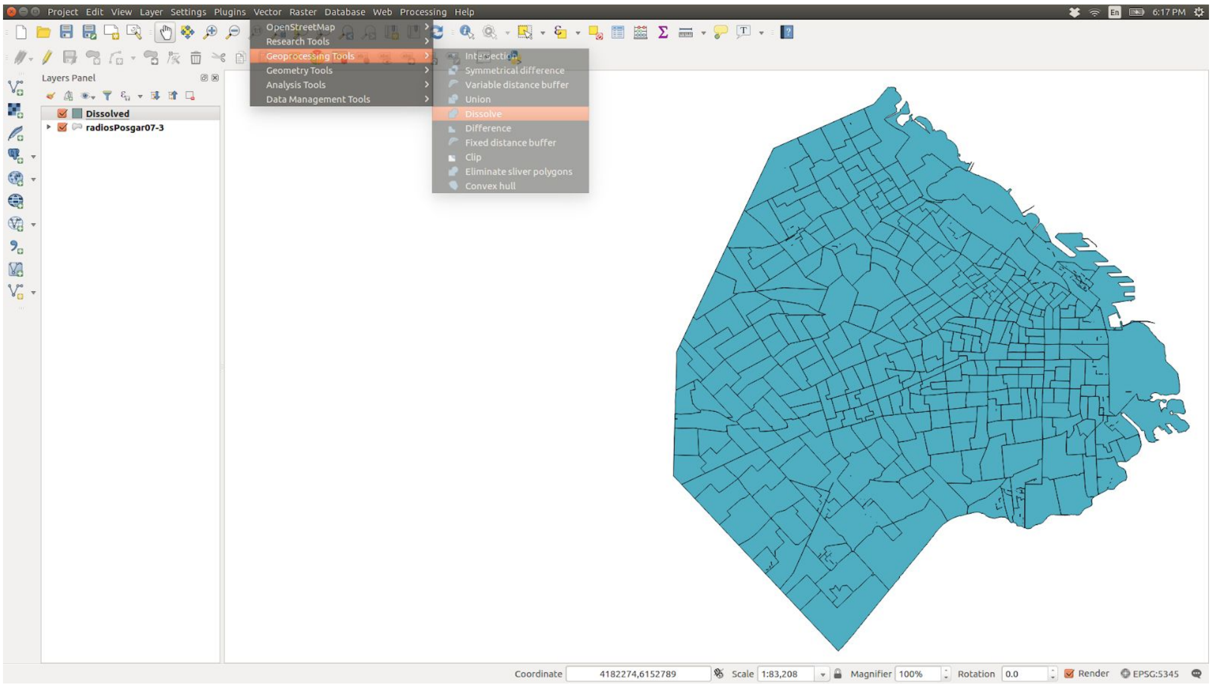


A continuación la idea es observar el PUEM en el aspecto de la escala al computar promedios de esta variable educativa para unidades espaciales más agregadas. De este modo, podremos ver cómo agregar datos oculta la heterogeneidad interna de los datos no agregados.

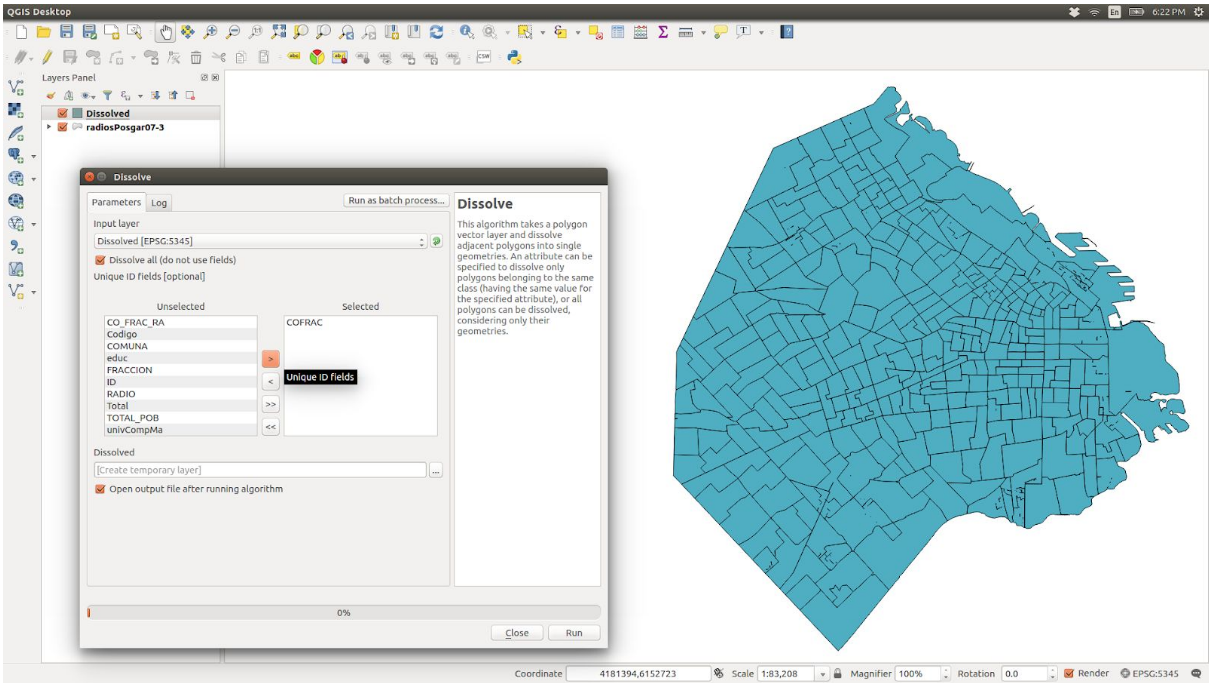
En primer lugar tenemos que comenzar por obtener las geometrías de esas unidades espaciales agregadas (fracciones o departamentos). Si bien es posible descargar shapefiles de las mismas, utilizaremos una herramienta de los software GIS llamada *Disolver*. La misma disuelve en una geometría única las geometrías que en su tabla de datos tengan atributos en común. Dado que en nuestra tabla de datos tenemos dos columnas que identifican a una misma fracción: COFRAC y COMUNA.

Para disolver los radios en fracciones realizamos el siguiente proceso:

Vector -> Herramientas de geoproceso -> Disolver

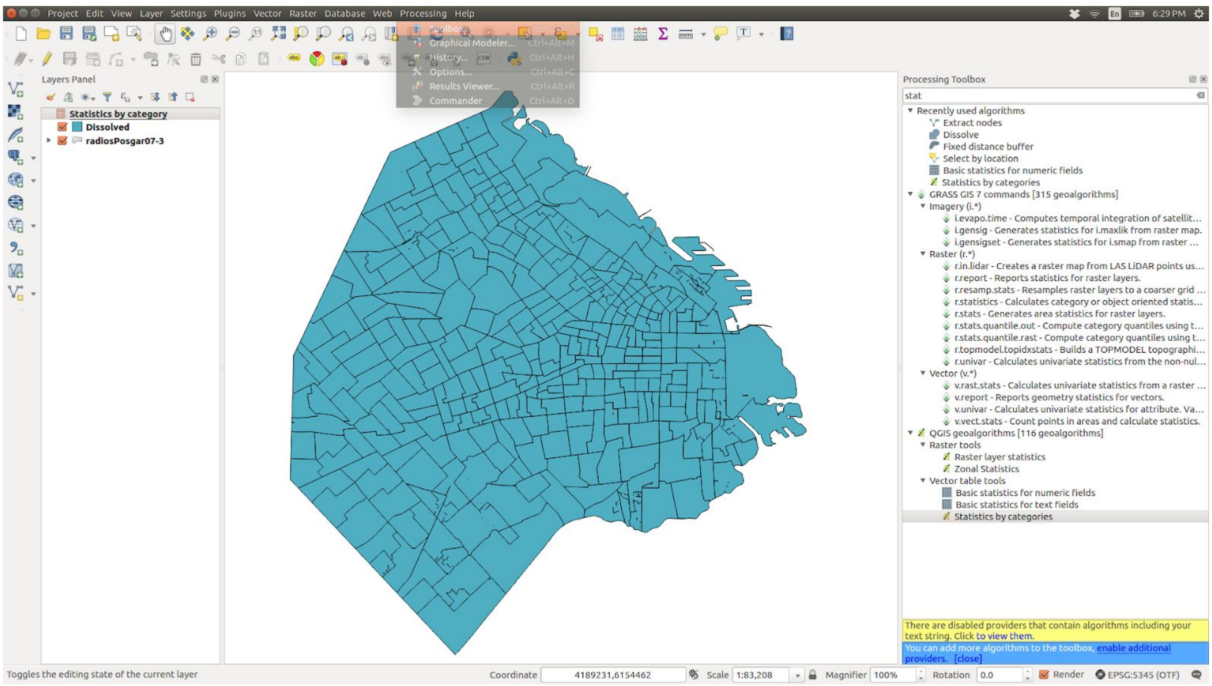


Elegimos el shapefile sobre el que vamos a trabajar, y la variable que es el campo común a los radios que queremos disolver. En este caso es COFRAC. Luego guardamos el archivo resultante. Una versión del mismo se encuentra en *data/fracciones*



Ya tenemos las geometrías para las unidades espaciales agregadas. Lo que necesitamos ahora es obtener el estadístico para las mismas. En este punto vamos a utilizar una media aritmética simple o promedio. Este proceso se puede realizar en cualquier herramienta de manipulación estadística ya que se realiza en una tabla de datos. QGIS cuenta con herramientas para hacerlo, pero esto queda librado a decisión de cada uno y a la herramienta con que trabajen con más comodidad.

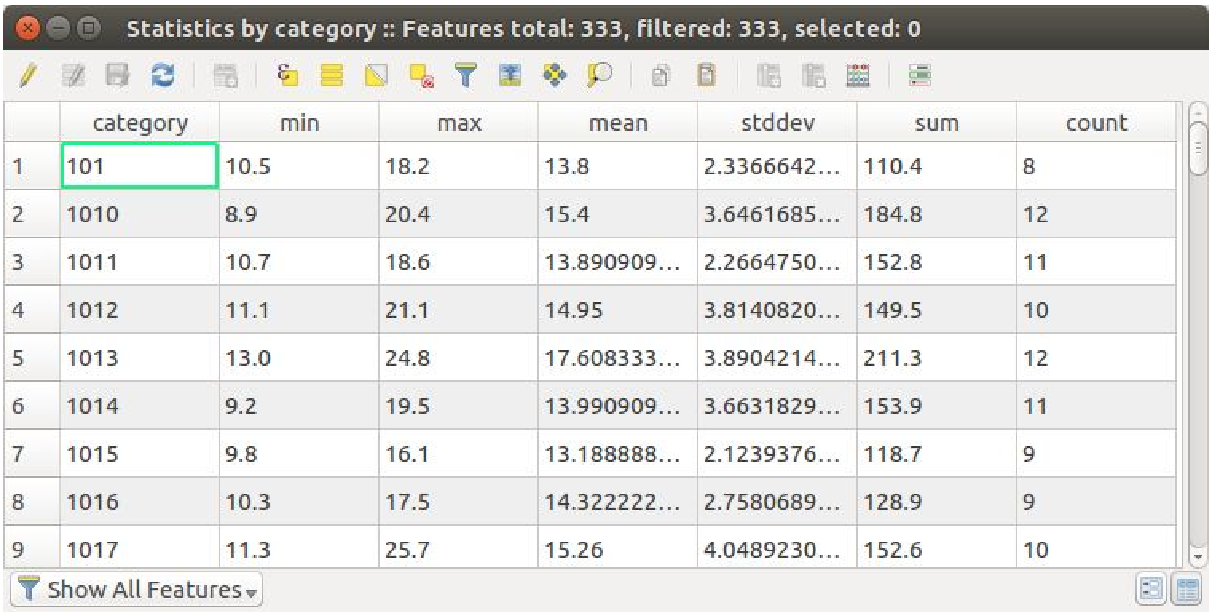
Vamos a Proceso -> Caja de herramientas y buscamos *Estadística por categorías*



Repetimos el proceso anterior, elegimos:

- el shapefile con el que vamos a trabajar
- el campo que identifica el atributo en común (en nuestro caso la fracción, *COFRAC*)
- el campo que contiene la variable de la cual queremos extraer las estadisticos (*educ*)

El resultado es una tabla de datos, donde para cada fracción (identificado con el campo *category*) tenemos una serie de estadísticos (mínimo, máximo, promedio, desvío, suma y conteo)


Statistics by category :: Features total: 333, filtered: 333, selected: 0

	category	min	max	mean	stddev	sum	count
1	101	10.5	18.2	13.8	2.3366642...	110.4	8
2	1010	8.9	20.4	15.4	3.6461685...	184.8	12
3	1011	10.7	18.6	13.890909...	2.2664750...	152.8	11
4	1012	11.1	21.1	14.95	3.8140820...	149.5	10
5	1013	13.0	24.8	17.608333...	3.8904214...	211.3	12
6	1014	9.2	19.5	13.990909...	3.6631829...	153.9	11
7	1015	9.8	16.1	13.188888...	2.1239376...	118.7	9
8	1016	10.3	17.5	14.322222...	2.7580689...	128.9	9
9	1017	11.3	25.7	15.26	4.0489230...	152.6	10

Show All Features

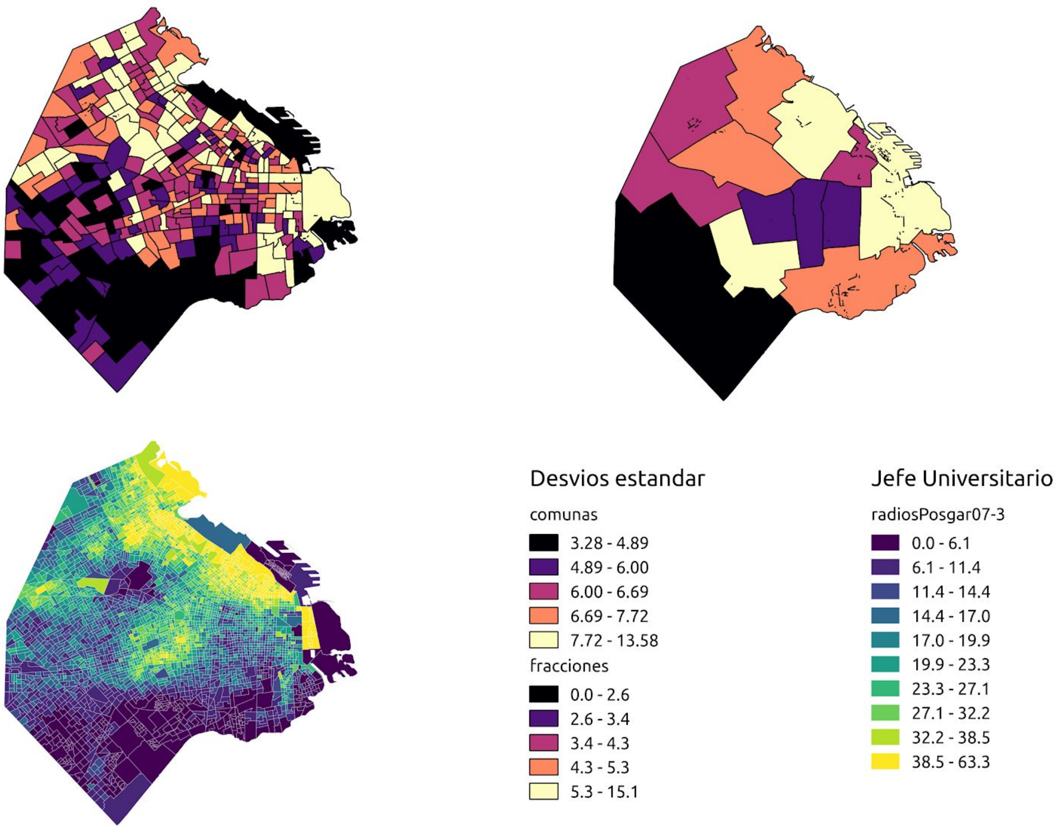
De este modo tenemos un resumen estadístico por categoría (en nuestro caso la categoría es la fracción). De este modo obtenemos un promedio de educación para cada fracción pero al mismo tiempo un desvío estándar. Es decir, cuán heterogéneos son los radios dentro de cada fracción.

A partir de un campo común en la tabla de datos que identifica a los radios de una misma fracción, obtuvimos:

1. Las geometrías de las fracciones (con *Disolver*)
2. Los estadísticos para esas fracciones (con *Estadísticos por categorías*)

Finalmente, hacemos un *join* o unión con las geometrías correspondientes. El campo que une las fracciones con sus datos son *CO_FRAC* y *category*. Los datos guardados en *data/fracciones* ya tienen estos datos almacenados. A su vez en *data/datosCensales* pueden encontrar los archivos csv con la tabla de datos calculados.

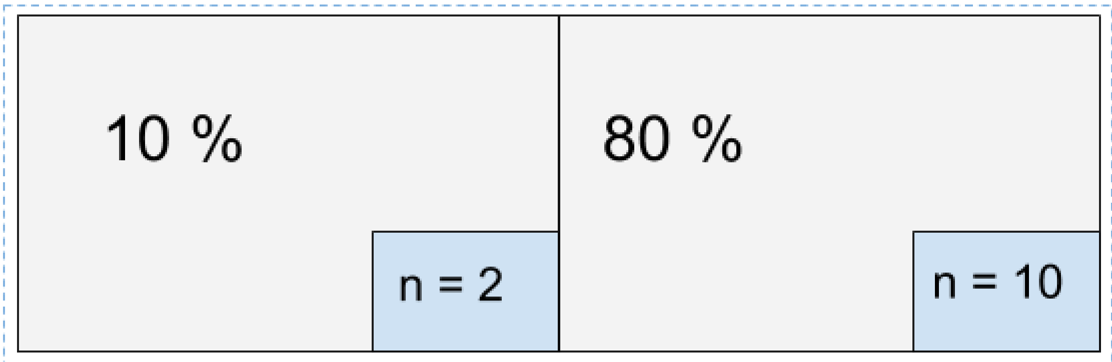
Los resultados son:



Para ver la heterogeneidad de los datos agregados, obtuvimos los totales para cada radio (el total de jefes de hogar y el total de jefes de hogar con universitario completo) a partir de esto, para agregar este dato a nivel de fracción o comuna, obtuvimos un promedio simple de esa proporción. Se podría haber calculado la sumatoria de ambos totales para cada nueva unidad espacial y volver a calcular. Hubiese dado el mismo resultado? Y si hiciésemos un promedio ponderado? Ponderado en base a qué valores?

Vamos a suponer que tenemos 2 microzonas (borde negro), que pueden agregarse en una única macrozona (borde celeste punteado) y están perfectamente anidadas. Una zona tiene una tasa de 10% (supongamos que seguimos trabajando con jefes de hogar con universitario completo o más) y otra tiene el 80%.

Micro y macro zonas : N = 12



El promedio simple daría:

Ps = (10% + 80%) / 2 = 10% * ½ + 80% * ½ = 45 %

El valor de cada una se multiplica por un peso constante: ½.

Ahora queremos usar información contextual que tenemos, como puede ser el total de jefes de hogar en cada microzona (N = 12), lo que nos permite hacer un promedio ponderado en base a qué proporción de los jefes de hogar de la macrozona se encuentra en cada microzona.

Promedio ponderado:

Pp = 10% * 2/12 + 80% * 10/12 = 68.3 %

Como vemos los resultados son diferentes. Como en una microzona tenemos 10 de los 12 jefes, tiene sentido que tenga más peso en el cómputo final. Qué pesos elegir y de acuerdo a si queremos computar un total o una proporción, esa es la cuestión.

Ahora, otro aspecto del PUEM es que se modifican en el tiempo. Por ejemplo, las comunas no existían en el censo 2001, sino que se usaba los distritos escolares. Por lo tanto, las unidades espaciales del censo 2010 se anidan perfectamente en sí mismas, las del censo 2001 también, pero no entre sí. No son compatibles 2001 con 2010. Qué sucede cuando tenemos que trabajar con unidades no anidadas?

SEGUNDA PARTE - EFECTO ZONIFICACIÓN

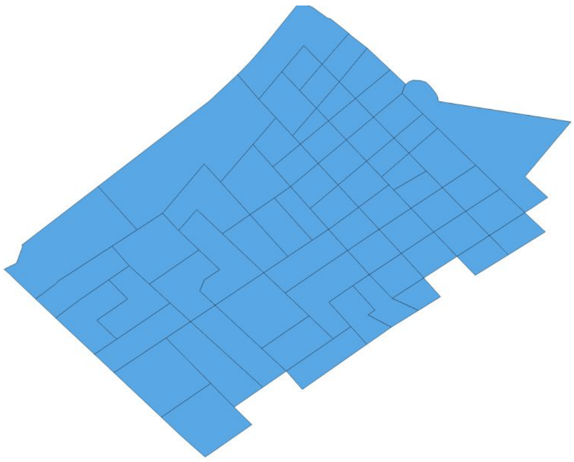
El segundo efecto del Problema de la Unidad Espacial Modificable es el efecto zonificación: Dos zonas de igual área y forma, pueden producir diferentes resultados de acuerdo a su ubicación.

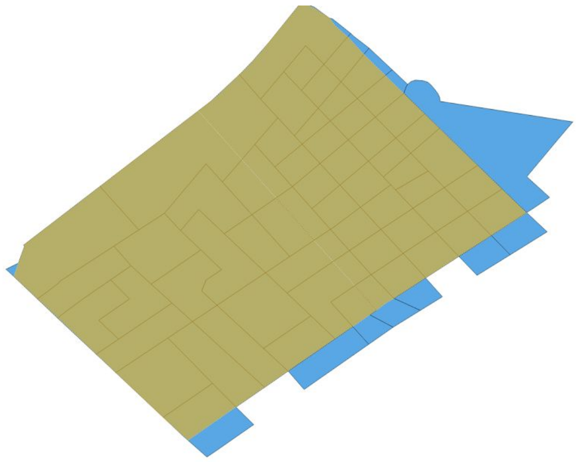
Hasta ahora veníamos trabajando con la proporción de jefes de hogares con universitario completo para radios censales. Los datos eran fácilmente agregables ya que eran unidades anidadas. Que pasa si la nueva zonificación no goza de esa propiedad de “anidamiento” con los radios? En este apartado veremos cómo *aporcionar* esos radios para que coincidan con la nueva geografía y redistribuir los datos del mejor modo, procurando que el efecto zonificación tenga el menor impacto posible.

Llamaremos *aporcionamiento* de polígono al término específico para reagrupar atributos de un conjunto de unidades de área (polígonos) a otro conjunto de unidades de área. Por lo general, usamos la distribución de datos a menor escala para datos de peso desde el polígono de origen (*source*) hasta el polígono de destino (*target*).

Supongamos que trabajamos con la hipótesis de que el nivel educativo promedio del distrito electoral impacta en la proporción de votos al partido oficialista del distrito. Como no tenemos encuestas con una muestra representativa donde para cada votante registre nivel educativo y a quién votó en la última elección, no nos queda otra solución que recurrir a la falacia ecológica y utilizar datos agregados, tenemos que usar los datos censales y el del resultado electoral. Si tuviésemos datos de cada escuela por ejemplo, podríamos hacer una teselación de la CABA con las escuelas como puntos. Esto generaría áreas totalmente irregulares que no coincidiría con ningún radio. La ventaja es que trataríamos con datos individuales de cada escuela y no con datos agregados por circuito electoral. Pero teselación escapa a los alcances de este taller. Trabajaremos con circuitos electorales.

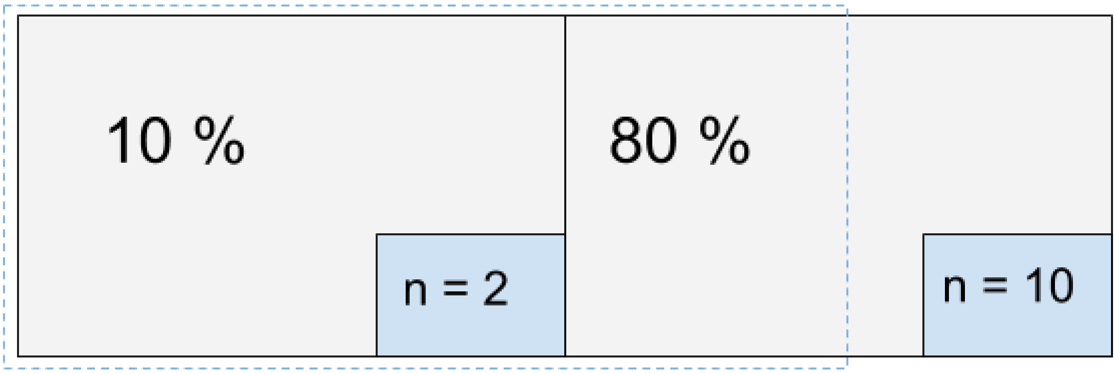
Entonces para esta misma unidad (circuitos electorales) necesitamos caracterizarlos en términos educativos de la zona. No coinciden como el caso anterior de fracciones y comunas, entonces debemos *aporcionar* polígonos. Nos concentraremos en una selección de dos circuitos electorales y los radios con los que se superponen. La misma se encuentra en *data/seleccion*. Allí pueden encontrar el archivo de proyecto de QGIS como los shapes con los datos.





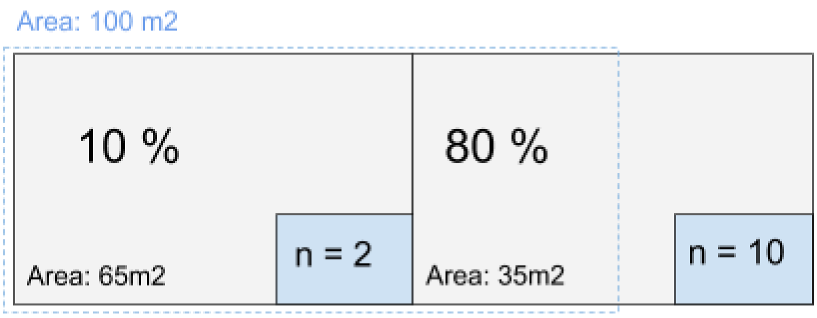
Se pueden observar los circuitos electorales Nº 154 y 156 correspondientes al barrio de Palermo, al Sur-Oeste de Plaza Italia. El radio en el extremo derecho es fácilmente reconocible como la Plaza y las avenidas Santa Fe y Las Heras. En primer lugar se puede ver que algunos radios al Sur-Este se encuentran cortados al medio por los circuitos electorales. Por otro lado, algunos radios (14_13_4, 14_14_7, 14_21_6) pertenecen a ambos circuitos: una porción del radio pertenece al 154 y otra al 156.

Al usar circuitos, deberíamos poder repartir los hogares que conforman un radio de otro modo, de acuerdo a ese circuito. Pero no tenemos los datos para cada hogar. Tenemos un radio, un polígono agregados. lo que debemos hacer es cortar ese polígono en función de la nueva unidad espacial, y redistribuir los datos. Este es un problema que surge todo el tiempo en análisis espacial. Partir una unidad espacial de acuerdo a otras unidades espaciales y transferir sus datos. Pero como estan agregadas, esta no es una tarea trivial. Como se puede hacer? Volvamos al ejemplo anterior, solo que ahora las microzonas y macrozona no están perfectamente anidadas, si no que la microzona de la derecha queda cortada por la macro zona.



La solución es realizar un promedio ponderado, como se hizo en la primera parte, pero el desafío está en qué pesos utilizar.

Pesos basados en areas



Pp = 10% * 65/100 + 80% * 35/100 = 34.5 %

Con pesos basados en áreas, cada radio contribuirá a la proporción del circuito en función del área que representa para el total del circuito. Llamaremos área objetivo o target (*t*) al área **para la cual** vamos a calcular el nuevo valor (en nuestro caso el circuito electoral). Llamaremos área fuente o source (*s*) al área **a partir de la cual** calcularemos el valor del área objetivo o target (en nuestro caso el radio censal).

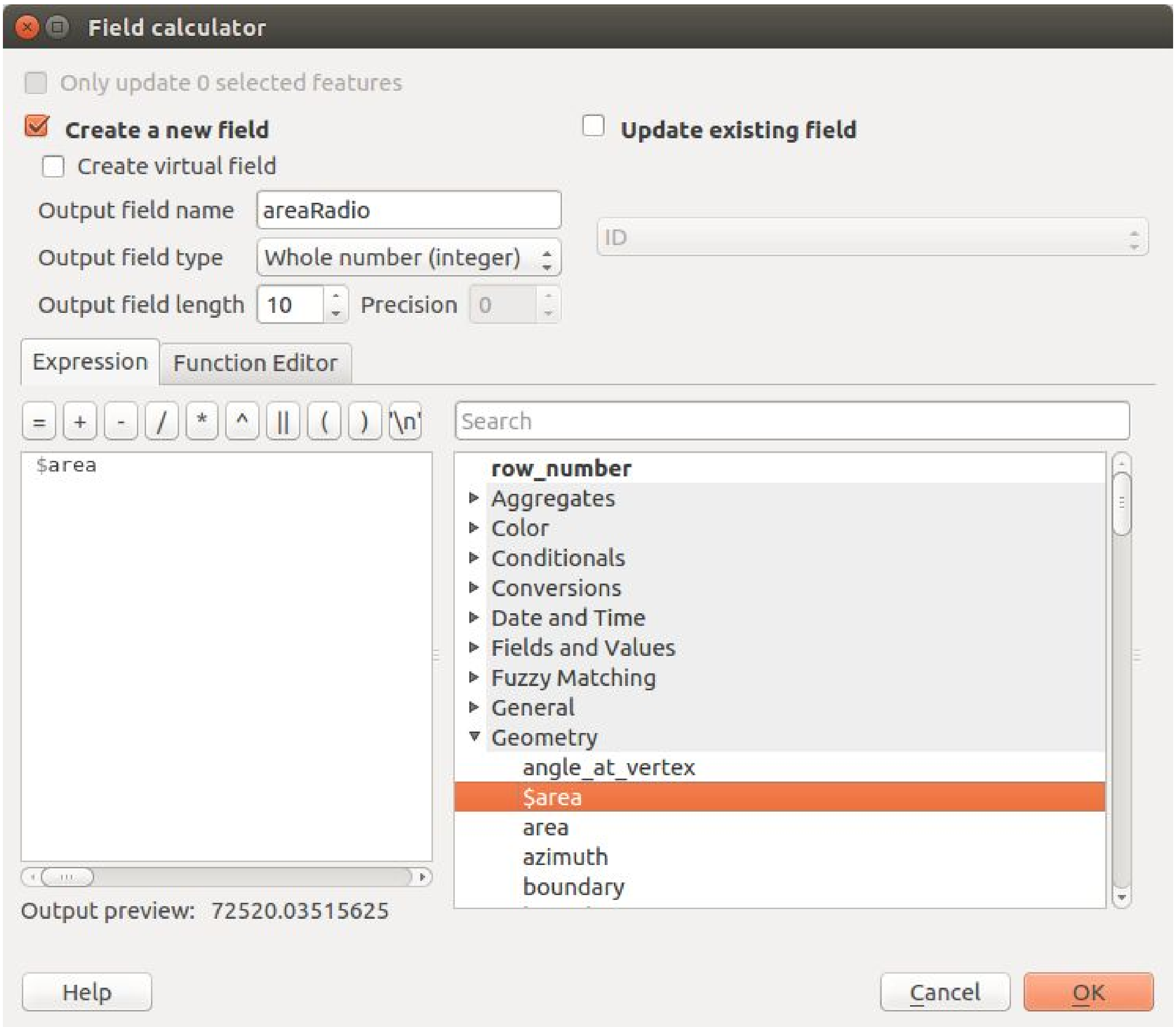
$$V_t = \sum_{i \in t} V_s \frac{a_i}{a_t}$$

V_t = Valor del área objetivo o *target*
V_s = Valor del área fuente o *source*
a_i = área de la porción del área fuente dentro del área objetivo
a_t = área de la porción del área objetivo

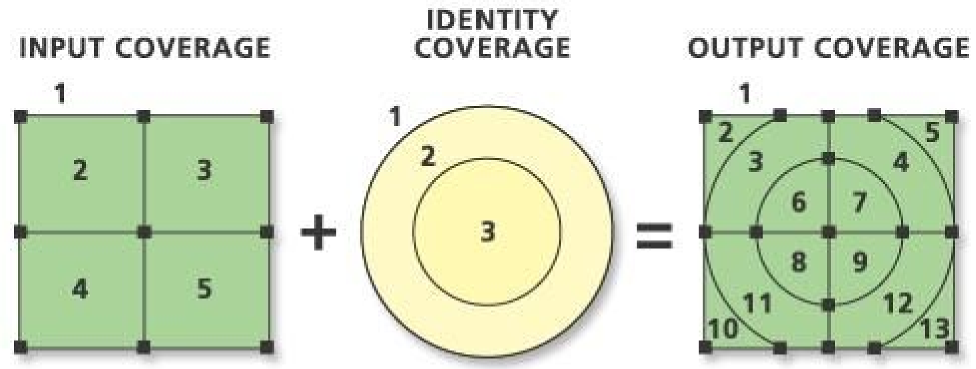
En primer lugar, para poder obtener este promedio ponderado primero tenemos que obtener

- El área de las zonas objetivo (los circuitos electorales)
- El área de las zonas fuente (los radios censales)

Para ello hacemos clic derecho en la capa de los radios censales y abrimos la tabla de atributos. Habilitamos la edición de los datos con el ícono del lápiz arriba a la izquierda de la tabla (SIEMPRE TRABAJAR CON CUIDADO YA QUE CADA CAMBIO QUEDA GRABADO Y NO PUEDE DESHACERSE). Luego abrimos la calculadora de campos (el ícono como un ábaco) creamos un nuevo atributo (o variable) que sea numérico y dentro de la sección geometría, elegimos *$area*. Esto calculará el área para todos los elementos de la geometría, **en las unidades de medida de la proyección en la que estamos trabajando**. Como trabajamos en POSGAR 07 banda 3, será en metros. Repetimos este proceso para los circuitos.



En segundo lugar, tenemos que obtener una geometría donde cada radio quede dividido al superponer los circuitos y que la tabla de datos de la geometría indique a que circuito pertenece esa subárea del radio (a_i). Para ello utilizamos la herramienta Identidad



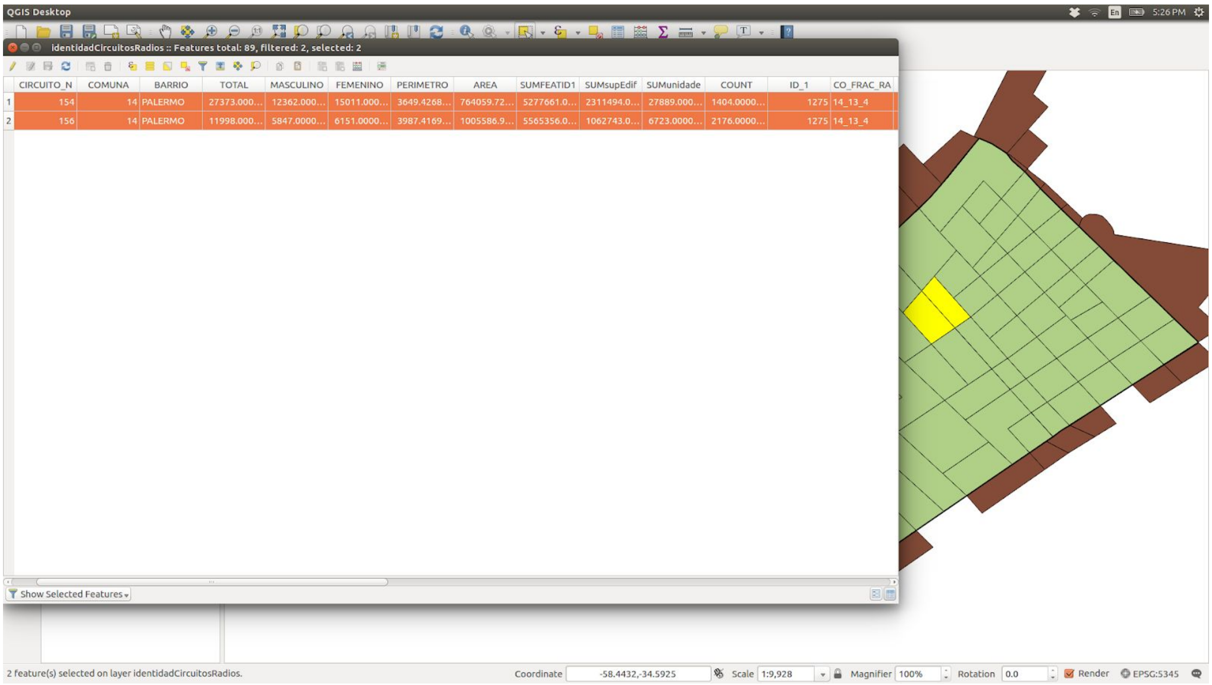
INPUT COVERAGE #	ATTRIBUTE	IDENTITY COVERAGE #	ATTRIBUTE
1		1	
2	A	2	102
3	B	3	103
4	C		
5	D		

OUTPUT COVERAGE #	INPUT COVERAGE #	ATTRIBUTE	IDENTITY COVERAGE #	ATTRIBUTE
1	1		1	
2	2	A	1	
3	2	A	2	102
4	3	B	2	102
5	3	B	1	
6	2	A	3	103
7	3	B	3	103
8	4	C	3	103
9	5	D	3	103
10	4	C	1	
11	4	C	2	102
12	5	D	2	102
13	5	D	1	

Identidad es un algoritmo incluido en SAGA GIS, un programa similar a QGIS. Desde QGIS podemos utilizar sus herramientas (al igual que con GRASS, Python, R, etc). Identidad permite combinar dos geometrías obteniendo como resultado todas las intersecciones posibles, sus geometrías pero también su tabla de datos. Esto es importante, ya que podremos tener para cada registro de la tabla, lo que antes denominamos como a_i, es decir el recorte de cada radio a partir de los circuitos. Para cada uno tendremos a su vez, en la misma fila, todos los datos del circuito al que corresponda.

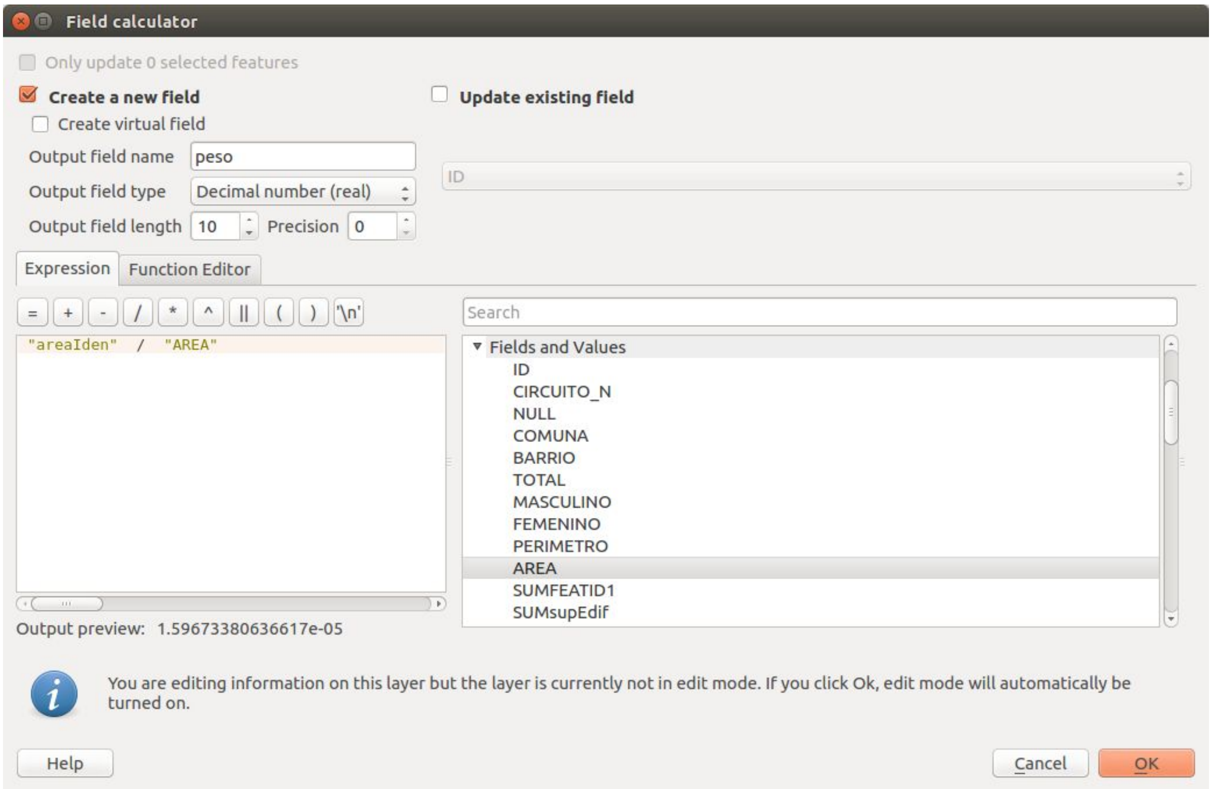
Dejamos en *data/seleccion/identidad* este shapefile con los datos de radios y circuitos.

El orden en el que realizamos estas operaciones es importante. Primero debemos calcular el área de cada radio y circuito, para que estas sean atributos de la tabla que ingresará al proceso de Identidad.



Como se puede observar en la imagen, tenemos el radio 14_13_4, duplicado en la tabla de datos, como dos geometrías diferentes, con dos registros diferentes en la tabla, cada uno asignado a un circuito. En la misma tabla tenemos el área del circuito correspondiente (*AREA*) y el area del radio correspondiente (*areaRadio*).

Creamos nuevamente un nuevo atributo desde la calculadora de atributos con $area para los radios. Llamamos a ese campo *areaIden.* A partir de ahora ya podemos calcular cuánto contribuye la porción del radio 14_13_4 dentro del circuito 154, qué *peso* tiene. Ese peso se obtiene de dividir *areaIden/AREA.* Esto puede hacerse para todos los registros desde la calculadora de atributos y guardarlo en un nuevo atributo llamado *peso*



Luego no tenemos más que multiplicar ese peso por el valor de cada registro (*educ)* en una nueva variable *educPond*.

De este modo obtenemos el valor de cada registro ponderado:

$$v_s \frac{a_i}{a_t}$$

Finalmente no nos queda más que sumar para todo el circuito, utilizando el mismo procedimiento que utilizamos en la primera parte (Proceso -> Caja de herramientas -> *Estadística por categorías*) solo que esta vez nos quedamos con el estadístico SUMA, del campo *educPond* para las categorías en CIRCUITO_N. Esta sumatoria nos da el estadístico buscado:

$$v_t = \sum_{i \in t} v_s \frac{a_i}{a_t}$$

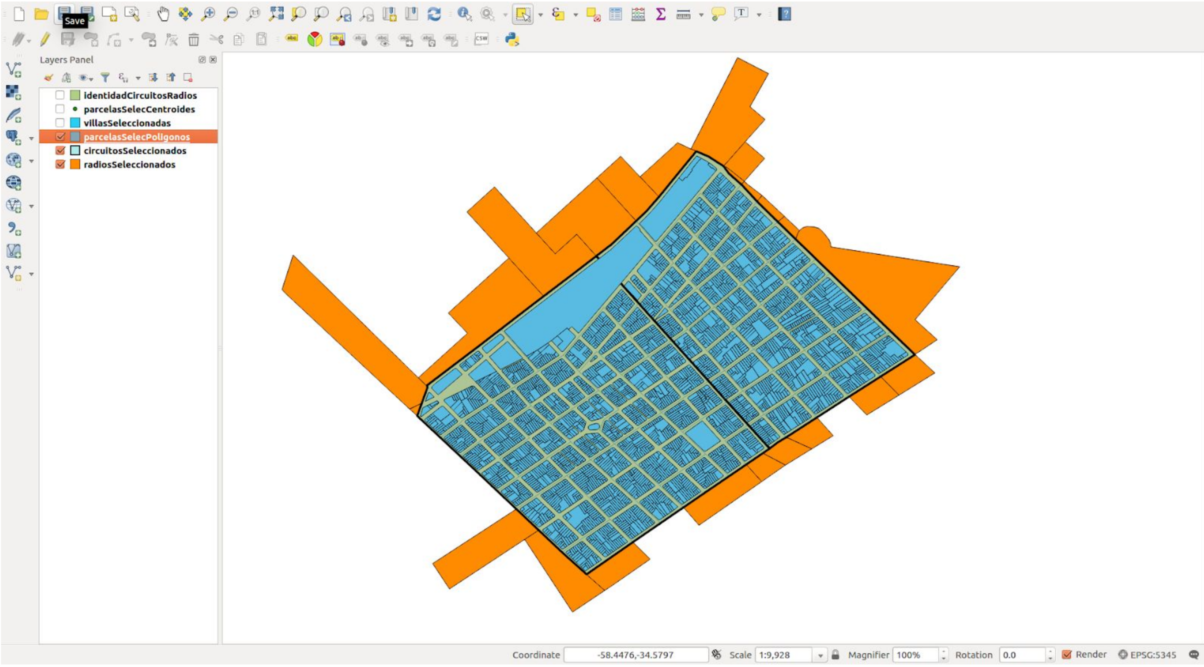
En este punto es importante detenerse a pensar. Utilizamos este recurso como única salida posible frente a la ausencia de datos desagregados. Como todo proceso científico, procedemos a ciegas a partir de las mejores presunciones que podemos conjurar. Para ello hacemos supuestos sobre la realidad subyacente. No es incorrecto, solo que tenemos que ser conscientes de ellos y ser críticos. Qué supuestos asumimos en esta situación? Viendo lo que vimos sobre el problema de la escala y como la agregación de datos en unidades espaciales macro esconde heterogeneidades subyacentes en las unidades espaciales micro, qué podemos decir sobre esta división de un radio en unidades más pequeñas sobre las que no tenemos datos?

Pesos basados en conteos de unidades de vivienda

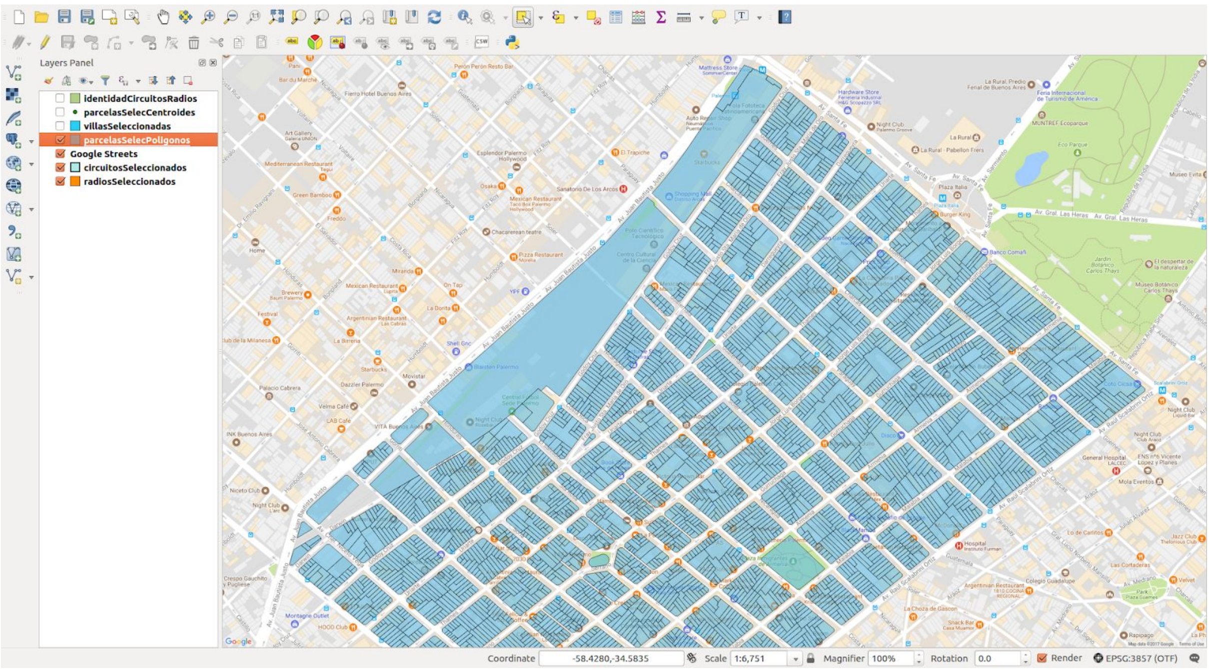
Existe otra forma de procurar pesos. Los radios censales procuran aglomerar un conjunto de viviendas (particulares, colectivas), dentro de las cuales se conforman diferentes hogares. Nosotros observamos al jefe de hogar y su nivel educativo. Si bien no existe relación 1 a 1 entre vivienda, hogar, jefe, podemos presumir que la distribución de unidades de construidas en las parcelas de una manzana guardar cierta relación con la cantidad viviendas dentro de la misma y por ende de los radios.

La ciudad de buenos aires publica un dataset de las parcelas de la ciudad. Si bien no es tal cual existía en 2010, podemos utilizarlo como proxy.

De este modo podemos realizar un peso o ponderador que contenga más información que el área, porque esa misma área bien podría no estar habitada.

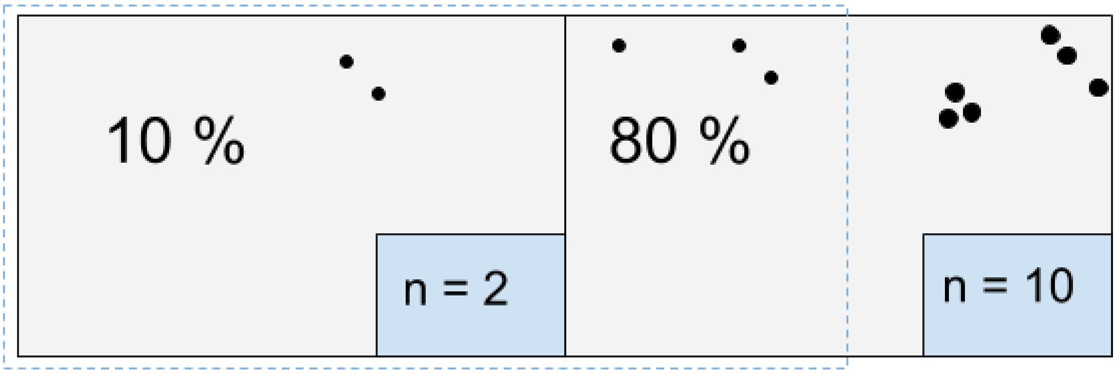


Ese ejemplo puede ser la parcela del norte del área seleccionada que se corresponde con las ex bodegas Ghioldi y terrenos ferroviarios.



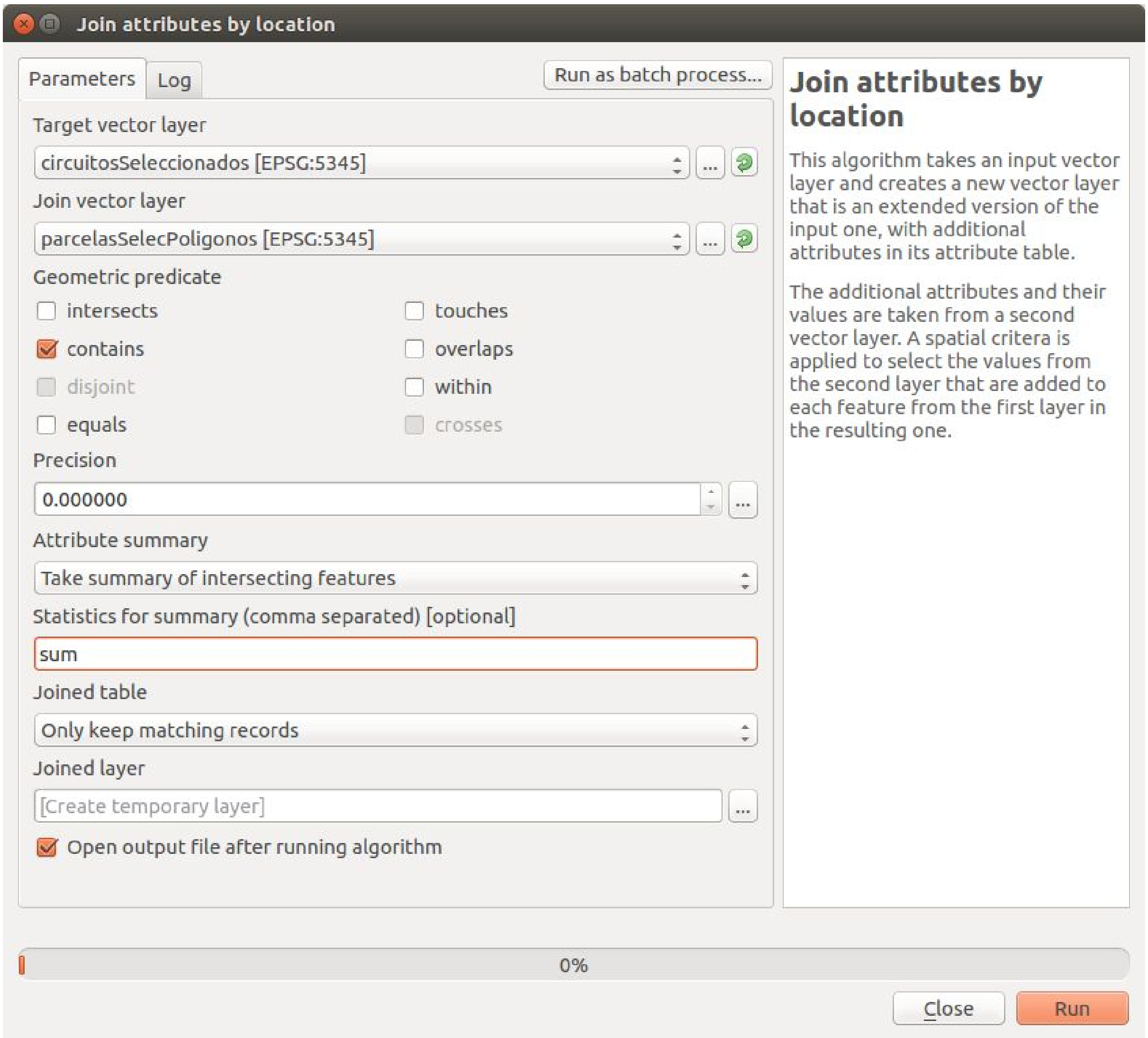
Con estos polígonos de las parcelas de la ciudad podemos realizar pesos basados en unidades de hogares por cada radio y porción de radio. El procedimiento es muy similar solo que en lugar de crear el peso a partir de la porción de area del radio dentro del area del circuito, se consturye a partir de la proporción de hogares del circuito que se haya en la parte del radio que cae dentro del circuito. La formula es muy similar

$$v_t = \sum_{i \in t} v_s \frac{h_i}{h_t}$$



Pp = 10% * $\frac{2}{5}$ + 80 % * $\frac{3}{5}$

Para computar este estadístico tenemos que realizar el mismo proceso. Primero obtenemos el total de parcelas que hay dentro de cada radio y circuito. Para ello obtenemos los centroides de los radios y luego una unión espacial, procurando obtener el conteo.



La capa target u objetivo es la que contiene a los radios (luego hay que hacer lo mismo para los circuitos). La capa fuente o source es la de los centroides de las parcelas. Nos interesa únicamente el conteo de unidades dentro de cada unidad objetivo. Computamos la suma ya que algún estadístico debemos darle a QGIS y el conteo será guardado en la variable de nombre COUNT por defecto. Luego realizamos la identidad (nuevamente el orden es importante, primero se calcula el conteo para radio y circuito y luego se hace Identidad) entre radios y circuitos.

- Crear centroides de parcelas
- Calcular Count de centroides en circuitos (COUNT)
- Calcular Count de centroides en radios (COUNT_1)
- Hacer Identidad con SAGA desde la caja de herramientas con cricuitos y radios
- Calcular Count de centroides en identidades (identidadConParcelas) (COUNT_2)

Al igual que con el área, obtenemos el peso dividiendo la cnatidad de uniades de vivienda dentro de la porción del radio que cae dentro del circuito y la cantidad de viviendas que caen dentro del circuito:

COUNT_2 / COUNT

Qué supuestos utilizamos?

Si podemos observar unidades de vivienda utilizando data de catastro pueden servir de proxy para la población, suponiendo que el tamaño del hogar y las viviendas vacías se distribuyan consistentemente. Heroicamente, ignoraremos la falacia ecológica y supondremos que todos los hogares del radio comparten la misma la tasa de la variable bajo análisis.

Bibliografía

Buzai, Gustavo (2013). Sistemas de Información Geográfica: Teoría y aplicación. Luján: Universidad Nacional de Luján. Capítulo 2. Sistemas de Información Geográfica (SIG). Recuperado de http://www.gesig-proeg.com.ar/documentos/libros/SIG-TEORIA%20Y%20APLICACION%202013.pdf

INDEC (2010). Censo Nacional de Población, Hogares y Viviendas 2010. Cartografía. Recuperado de http://www.indec.gov.ar/codgeo.asp

INDEC (2012). Censo nacional de población, hogares y viviendas 2010. Censo del Bicentenario. Resultados definitivos, Serie B nº 2. Buenos Aires: INDEC.

INDEC (2013a). Definiciones de la base de datos. Censo Nacional de Población, Hogares y Viviendas 2010, Base de datos REDATAM. Buenos Aires: INDEC.

INDEC (2013b). Definiciones de los Indicadores. Censo Nacional de Población, Hogares y Viviendas 2010, Base de datos REDATAM. Buenos Aires: INDEC.

INDEC. Censo Nacional de Población, Hogares y Viviendas 2010. Base de datos REDATAM, Cuestionario básico. CD-ROM.

INDEC (s/f). Unidades Geoestadísticas. Cartografía y códigos geográficos del Sistema Estadístico Nacional. Definiciones. Recuperados de http://geoservicios.indec.gov.ar/codgeo/index.php?pagina=definiciones

Lloyd, C. D. (2010) Spatial Data Analysis - An introduction for GIS users, Oxford University Press, Oxford, Reino Unido

Mera, Gabriela y Marcos, Mariana (2012). Los censos de población como fuente de datos para trabajar a nivel microespacial (1980-2010). Pampa: Revista Interuniversitaria de Estudios Territoriales, (8), 137-162. Recuperado de http://www.dialnet.unirioja.es/descarga/articulo/4695139.pdf

Monmonier, M. (1996) How to lie with maps, The University of Chicago Press, Chicago, Estados Unidos

Openshaw, S. and P. J. Taylor (1979). A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In N. Wrigley, ed. Statistical Applications in the Spatial Sciences, 127–144. London: Pion.