

Introducción al uso de datos abiertos para el análisis espacial con SIG

Objetivo

En los últimos años se ha extendido el uso del manejo, análisis y visualización de grandes set de datos en lo que se dio a llamar Big Data o Ciencia de Datos. En este contexto, el análisis espacial está llamado a ocupar un lugar central. El espacio en primer lugar permite estructurar estos enormes set de datos que en primer lugar parecen no estar estructurados. Pero al mismo tiempo el espacio, a la vez producto y productor de relaciones sociales, es una variable que tiene el potencial de enriquecer cualquier análisis. En particular para el estudio de los procesos urbanos o la evaluación de políticas públicas que tengan lugar en las grandes ciudades. El objetivo general de este taller es introducir a profesionales de las diversas disciplinas (economistas, sociólogos, polítólogos, estadísticos, etc.) en el uso de herramientas de análisis espacial aplicado al procesamiento de datos abiertos. Para ello se realizará un taller práctico en base al aprovechamiento de los datos abiertos de la Ciudad de Buenos Aires para el estudio de la movilidad multimodal (combinando bicicletas públicas y subte) para mejorar la accesibilidad y los tiempos de viaje.

El objetivo del trabajo es poder ubicar estaciones de bicicletas públicas compartidas en las zonas que se encuentran relativamente cerca del subte de modo tal que las personas que viven allí no pueden caminar a una estación porque está muy lejos, pero si podrían tomarse una bicicleta, dejarla en la estación y de allí seguir en subte. Esto implicaría ahorrarse tiempo de viaje.

Esta guía servirá de paso a paso para seguir durante el taller y que quede como un apunte para aquellos que les interese repetir y adaptar lo que hicimos, utilizando estas herramientas en otros contextos.

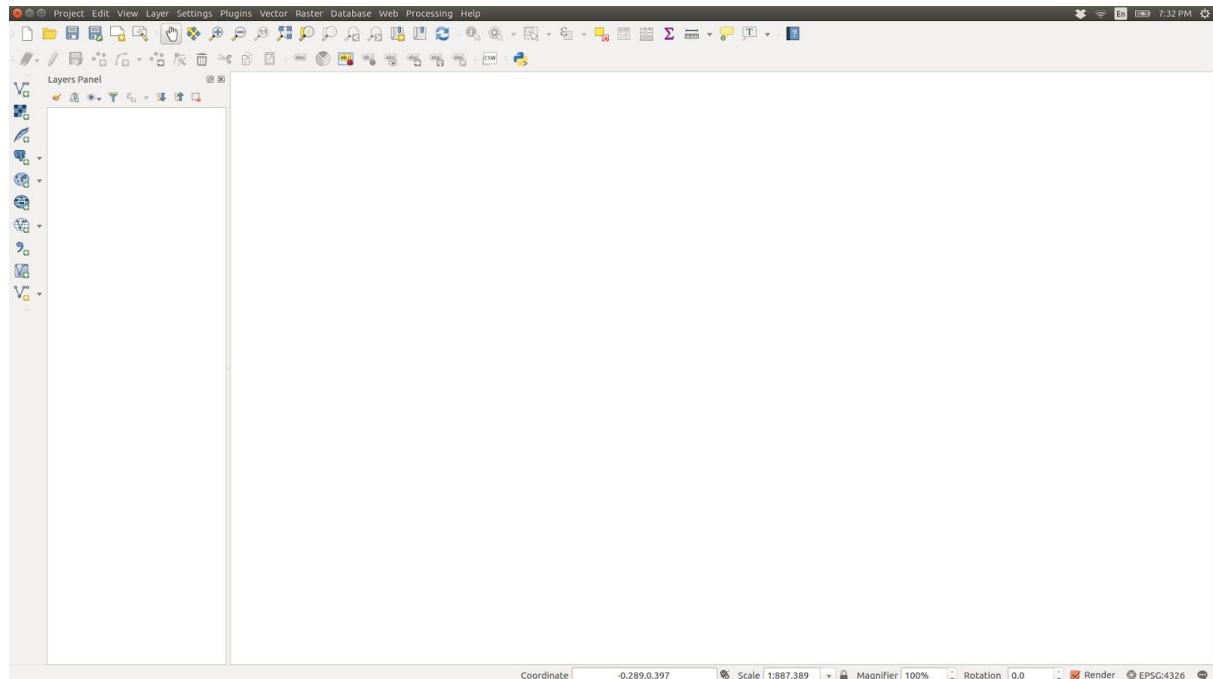
1 - Obtener las herramientas de software



QGIS es la herramienta escogida para llevar adelante el taller. Si bien el objetivo es conocer las técnicas de un modo independiente del software que se utilice e ir más allá de la pedagogía de *point and clic* donde sólo se siguen instrucciones, QGIS cuenta con numerosas ventajas que lo hacen un candidato ideal para los primeros pasos en el análisis espacial. A su vez no contamos con tiempo para realizar el trabajo en simultáneo con ArcGIS. Pero confiamos en que expondremos los contenidos de un modo que se entienda el concepto en sí y pueda ser trasladado a cualquier herramienta.

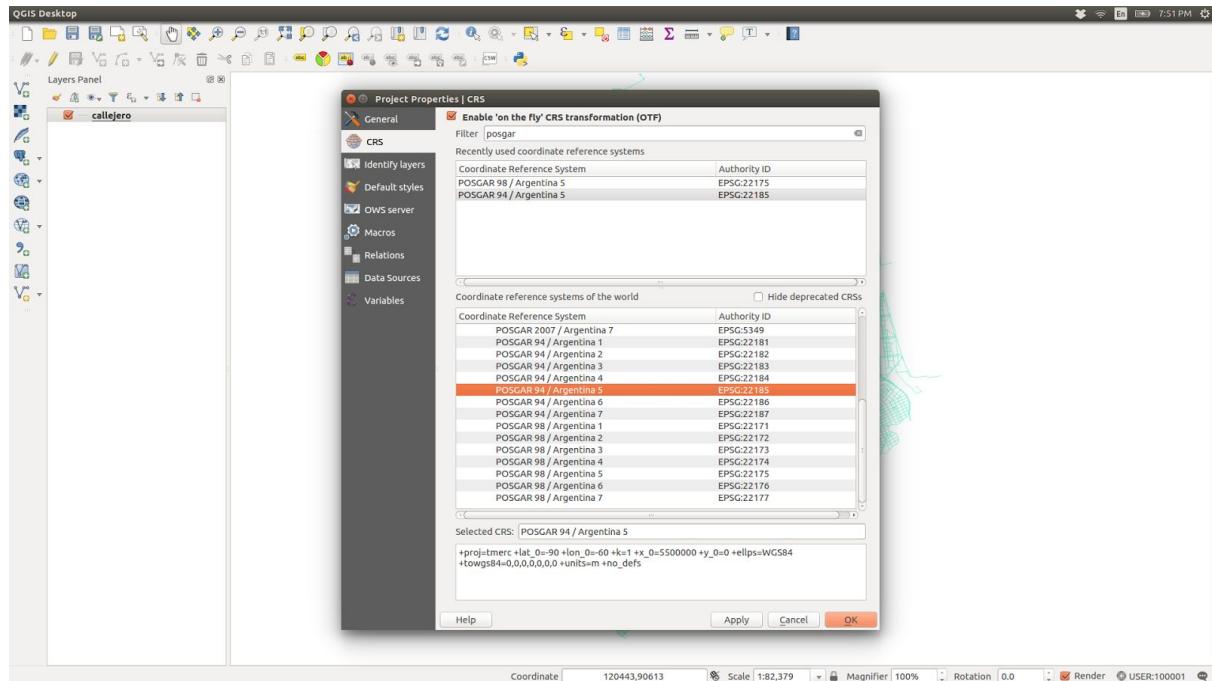
Para descargar el software pueden ir al [sitio de descargas de QGIS](#) y descargar la versión para su sistema operativo. Aquí utilizaremos la versión 2.18, nombre clave Las Palmas

Una vez descargado el software lo instalamos y abrimos. La primera pantalla puede variar de acuerdo a la configuración y el sistema operativo.



Lo que hacemos a continuación es establecer la [proyección de coordenadas](#) para nuestro trabajo. Siempre es el mejor modo de comenzar a trabajar, estableciendo la proyección deseada de modo consciente.

Nosotros utilizaremos el sistema POSGAR 94 / Argentina 5, cuyo código EPSG es 22185. Una de las propiedades más importantes del sistema de proyección, como veremos más adelante, es la unidades de medida. POSGAR utiliza metros.



Para ello hacemos clic en dónde dice EPSG y un número. [EPSG](#) es un estándar ([European Petroleum Survey Group](#)) de la [International Association of Oil and Gas Producers](#). El mismo establece un sistema de referencia de coordenadas (SRC o coordinate reference system - CRS en inglés) local, regional o global que define una proyección utilizada y cómo pasar de una a otra proyección.

2 - Descargar los datos

A continuación nos dirigimos al [portal de datos](#) de la Ciudad de Buenos aires. Esta será nuestra principal fuente de información para este trabajo.

The screenshot shows the official website for Buenos Aires Data. At the top, there's a navigation bar with links for File, Edit, View, History, Bookmarks, Tools, and Help. The address bar shows the URL https://data.buenosaires.gob.ar. The main header features the "Buenos Aires Data" logo and the text "Iniciativa de Datos Públicos y Transparencia de la Ciudad Autónoma de Buenos Aires." Below the header is a search bar with the placeholder "Buscá entre los 190 datasets" and a magnifying glass icon. A yellow button labeled "Ver catálogo completo" is visible. The page is divided into several sections: a grid of icons representing different datasets, a sidebar with a cartoon character, and a video player showing a presentation by Andy Freire.

Una vez allí podemos buscar el dataset de calles. Elegimos descargar el archivo comprimido que contiene los archivos fundamentales de todo sistema SIG para tipos vectoriales (siempre en este taller vamos a trabajar con archivos de tipo vectoriales). El estándar para archivos vectoriales hoy se denomina *shapefile* en base a un modelo de datos desarrollado por la compañía [ESRI](#) creadora del software ArcGIS. Existen alternativas como GeoJSON y otros que no abordaremos en este taller.

Los archivos indispensables de un shapefile son:

- **.shp** - Es el archivo con la geometría en sí misma
- **.prj** - Es el archivo que guarda la información referida al [sistema de coordenadas](#) en formato [WKT](#).
- **.dbf** - Almacena la tabla de datos correspondiente a los atributos (variables) para cada entidad (caso)
- **.shp.xml** - Almacena los [metadatos](#) del shapefile.
- **.shx** - Almacena el [índice espacial](#) de las entidades.

The screenshot shows a web browser window with the URL <https://data.buenosaires.gob.ar/dataset/calles>. The page title is "Recursos y descargas". There are two main download sections:

- Callejero**: CSV file.
- Callejero (RAR)**: ZIP file.

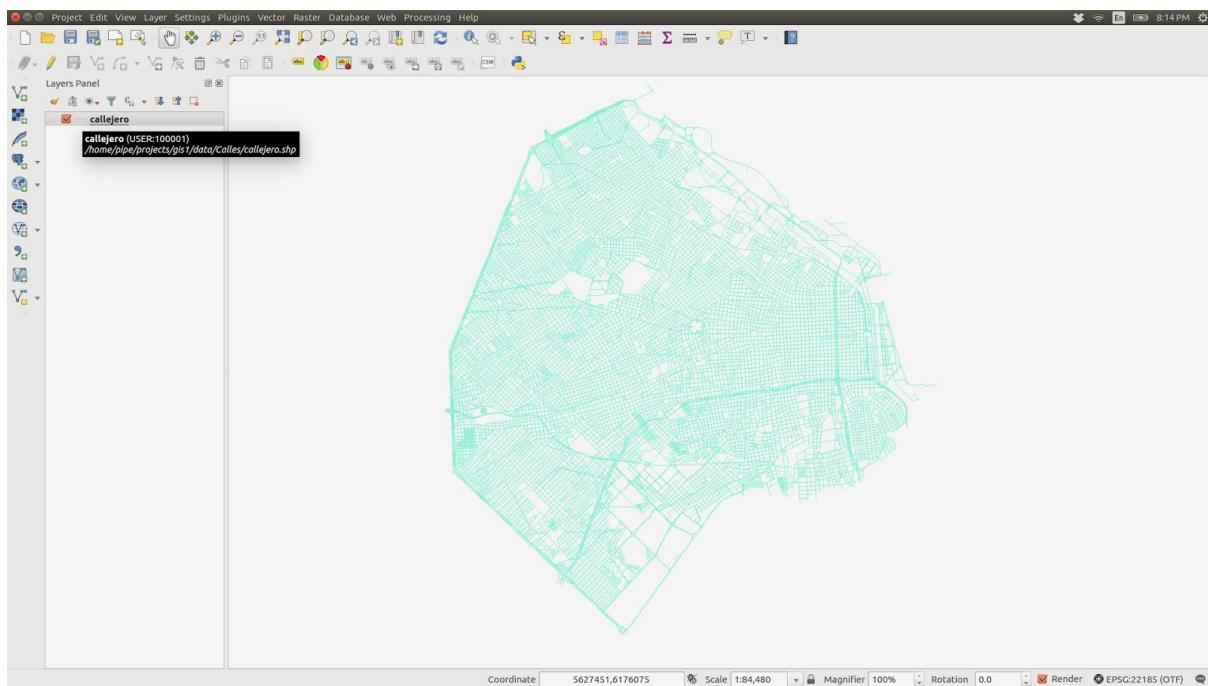
Below the downloads, there is a comment section with 12 comments, social sharing links (D, f, t, G, a), and a login/register form.

Repetimos este proceso con todos los datos de nuestro interés:

- [Calles](#)
- [Estaciones de subte](#)
- [Recorrido de subte](#)
- [Estaciones Ecobici](#)
- [Ciclovías](#)
- [Radios censales](#)

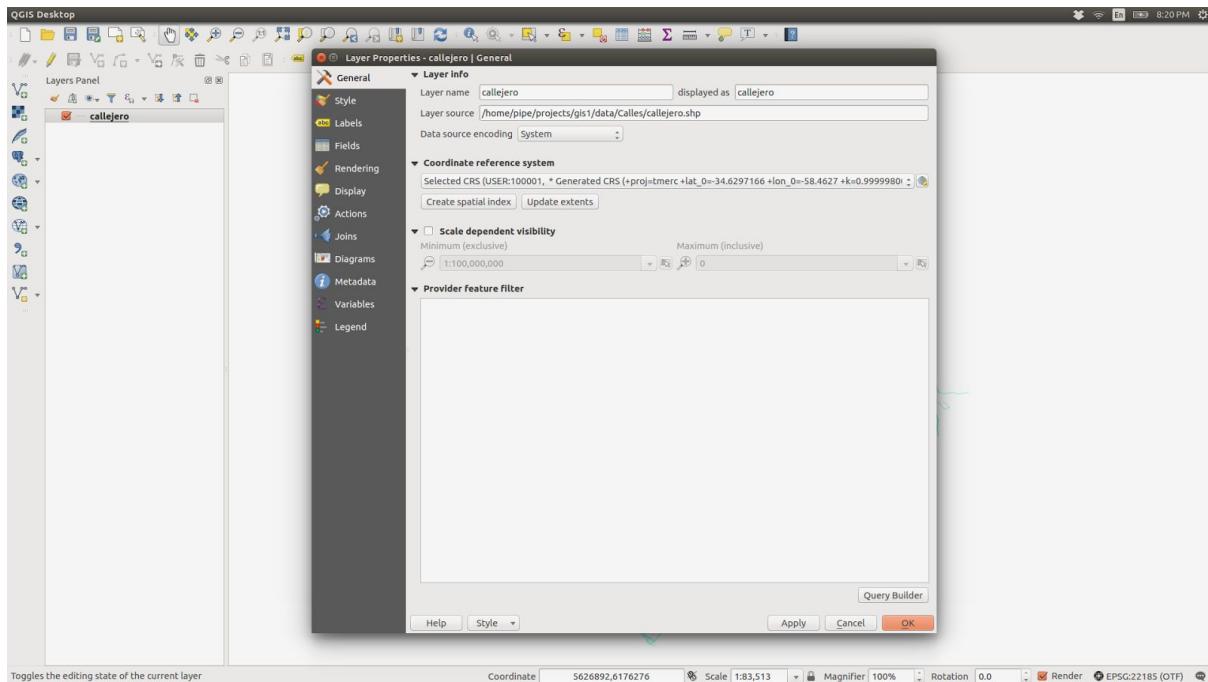
3 - Abrir los datos en QGIS

Comenzaremos con las calles para que nos sirvan de referencia para el resto del proyecto. Una vez descargado el archivo comprimido, lo descomprimimos en el directorio de nuestra elección.



Como se puede observar, en la configuración de QGIS que tenemos podemos ver a la izquierda los archivos que tenemos abiertos (solo el de calles por ahora) y en el panel principal la visualización de los mismos. Al posar el mouse, vemos que dice USER:100001 como proyección. Debemos hacer coincidir la proyección del archivo con la proyección general del proyecto.

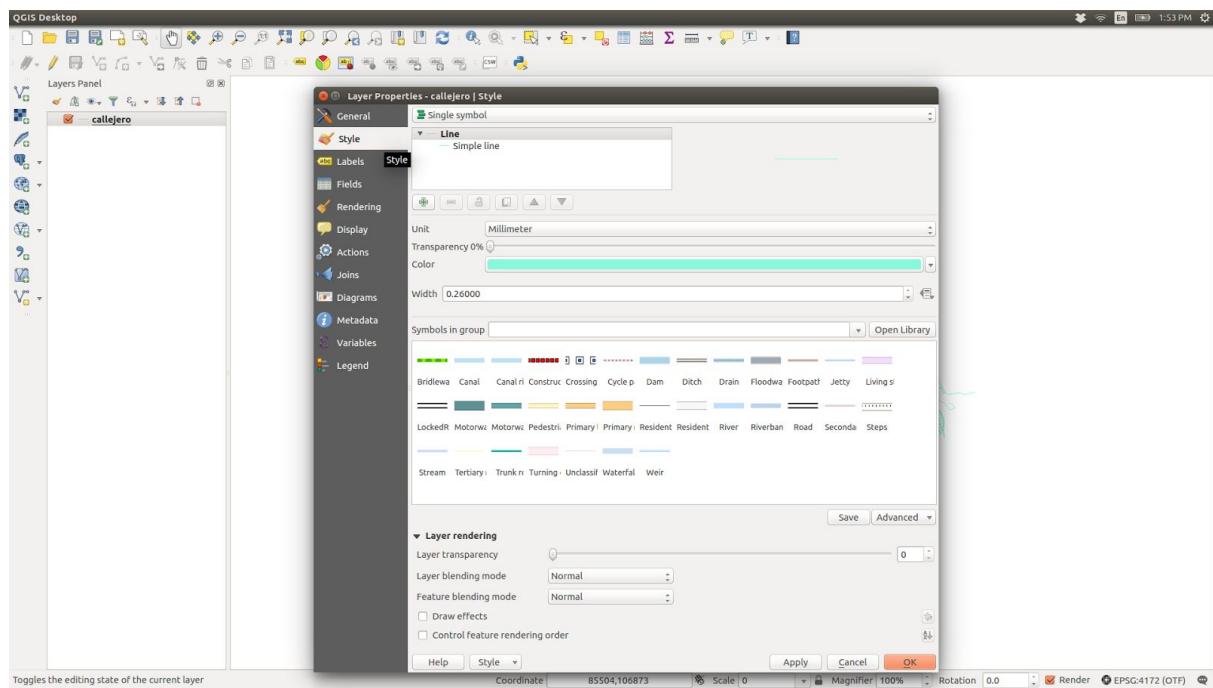
Para eso hacemos clic derecho en el archivo > *Propiedades* y vamos a la pestaña *General*.



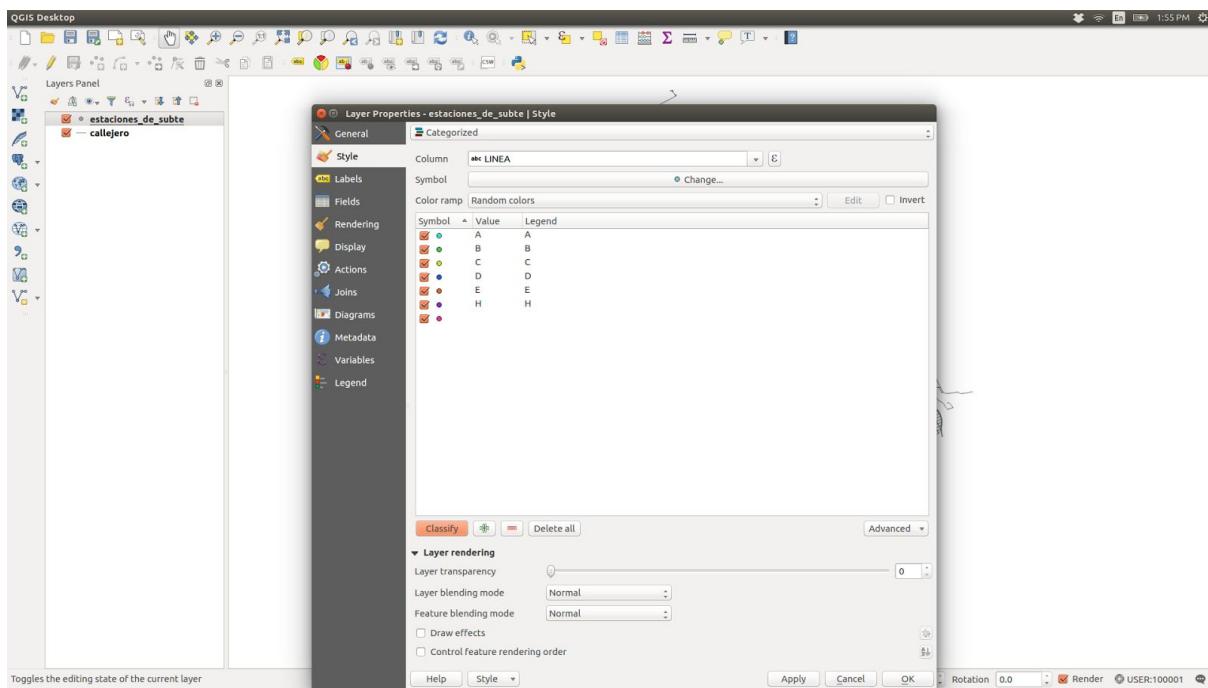
Ahí en el CRS elegimos POSGAR 94 / Argentina 5. Las calles no desaparecieron, si no que se movieron a otro lugar en las coordenadas. Para encontrarlas solo hay que hacer clic derecho en el archivo y elegir *Zoom en la capa*.

5 - Visualización

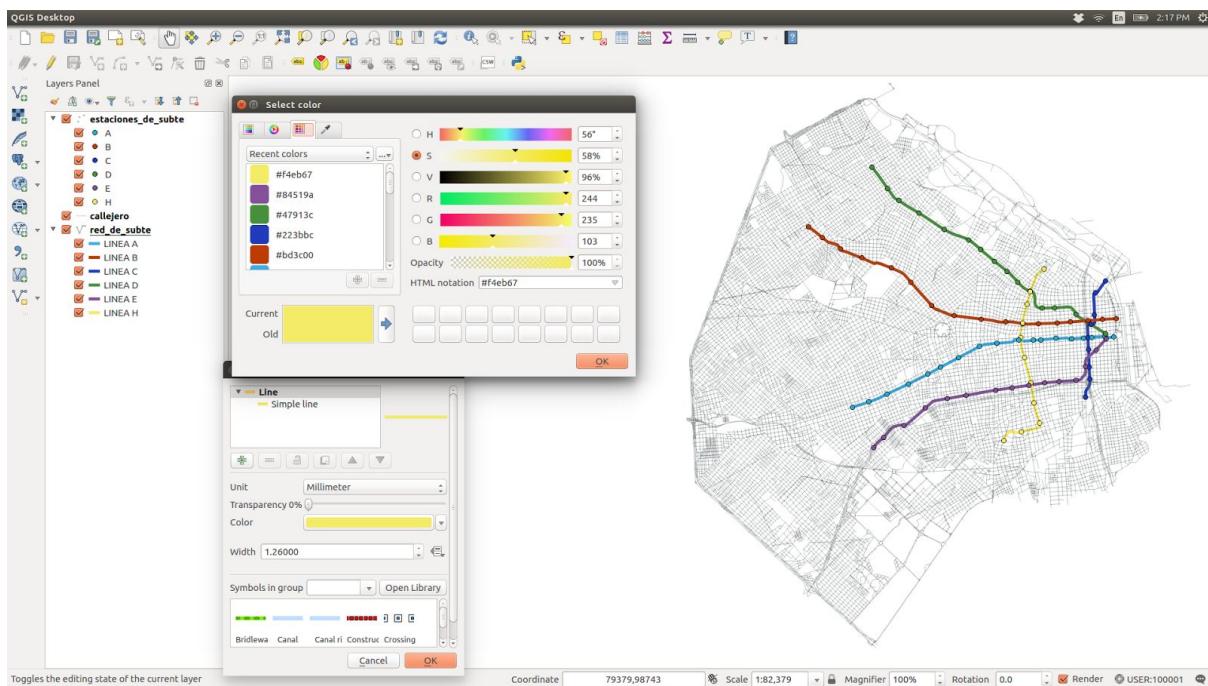
Las propiedades de visualización depende en gran medida del tipo de objeto espacial que queremos visualizar (punto, línea, polígono) y el nivel de medición de las variables o atributos (nominal, ordinal, intervalar). Comenzamos editando las calles, limitandonos a darle un color que nos agrade, digamos gris oscuro accediendo a las Propiedades de la capa.



Continuamos con los puntos. Los cargamos y les damos la misma proyección con la que estamos trabajando. Una vez localizados en el mapa, accedemos a sus *Propiedades*.

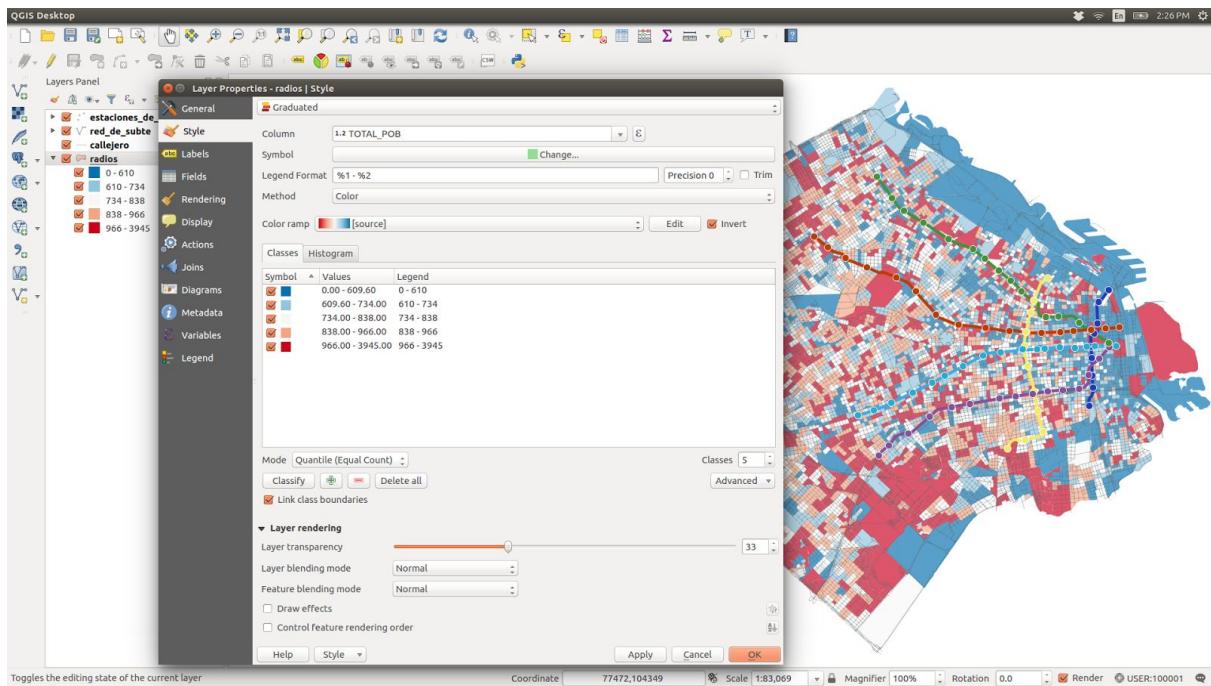


En las herramientas SIG podemos visualizar los objetos en función de sus propiedades. Para los puntos podemos variar su color en función de la línea a la que pertenecen. Dado que es una variable nominal podemos categorizarlos y darles un color en función de dicha categoría. Qué color darles implica toda una estrategia de visualización, no es una decisión neutra. El color construye sentido. En este caso podemos utilizar los reconocidos colores de las líneas de subte, lo que ayuda a nuestro mapa a comunicar mejor. Pero las decisiones para otros temas son más complejas.



Agregamos las líneas. Podemos aumentar el ancho, darles el mismo color con la herramienta de seleccionar los colores dentro de los más recientes. O tener anotados los colores por sus códigos alfanuméricos: #f4eb67.

Finalmente agregamos los radios censales con información y coloreamos en función de alguna variable como puede ser el total poblacional. Elegimos la cantidad de grupos, la paleta de colores (recordando que el color es un portador de sentidos) y *fundamentalmente*, un criterio para colorear.



QGIS ofrece por defecto 5 criterios:

- Intervalo igual: como su nombre indica, este método creará clases que tengan el mismo tamaño. Si nuestros datos van de 0 a 100 y queremos 10 clases, este método crearía una clase de 0-10, 10-20, 20-30 y así sucesivamente, manteniendo cada clase del mismo tamaño de 10 unidades.
- Quantile: este método decidirá las clases en las que dicho número de valores en cada clase sea el mismo. Si hay 100 valores y queremos 4 clases, el método de cuantiles decidirá las clases de modo que cada clase tendrá 25 valores.
- Quiebres naturales de Jenks: este algoritmo trata de encontrar agrupaciones naturales de datos para crear clases. Las clases resultantes serán tales que habrá una varianza máxima entre clases individuales y la menor varianza dentro de cada clase.
- Desviación estándar: este método calculará la media de los datos y creará clases basadas en la desviación estándar de la media.
- Pretty Breaks - Esto se basa en el algoritmo bonito del paquete estadístico R. Es un poco complejo, pero lo 'bonito' en el nombre significa que crea límites de clase que son números redondos.

Uno puede también elegir manualmente los cortes. Cada estrategia tiene ventajas y desventajas y el talento está en saber utilizar cada una en el contexto que se necesite

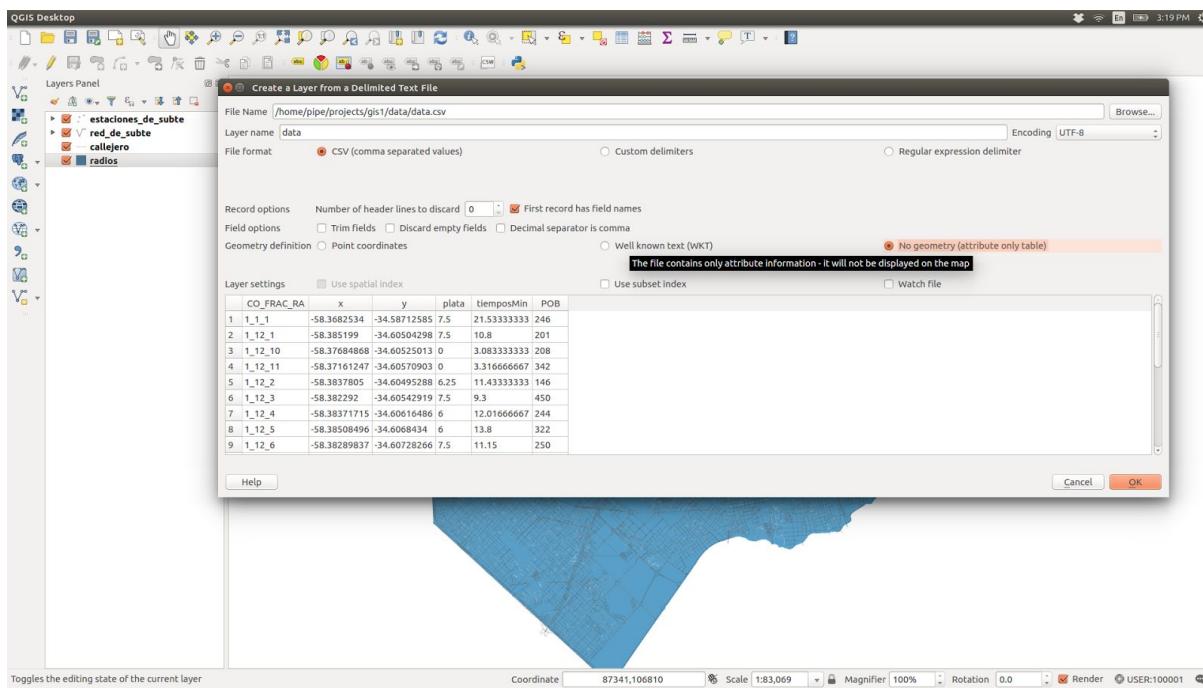
Aquí podemos ver un mismo set de datos con diferentes estrategias de visualización

6 - Uniones o Joins

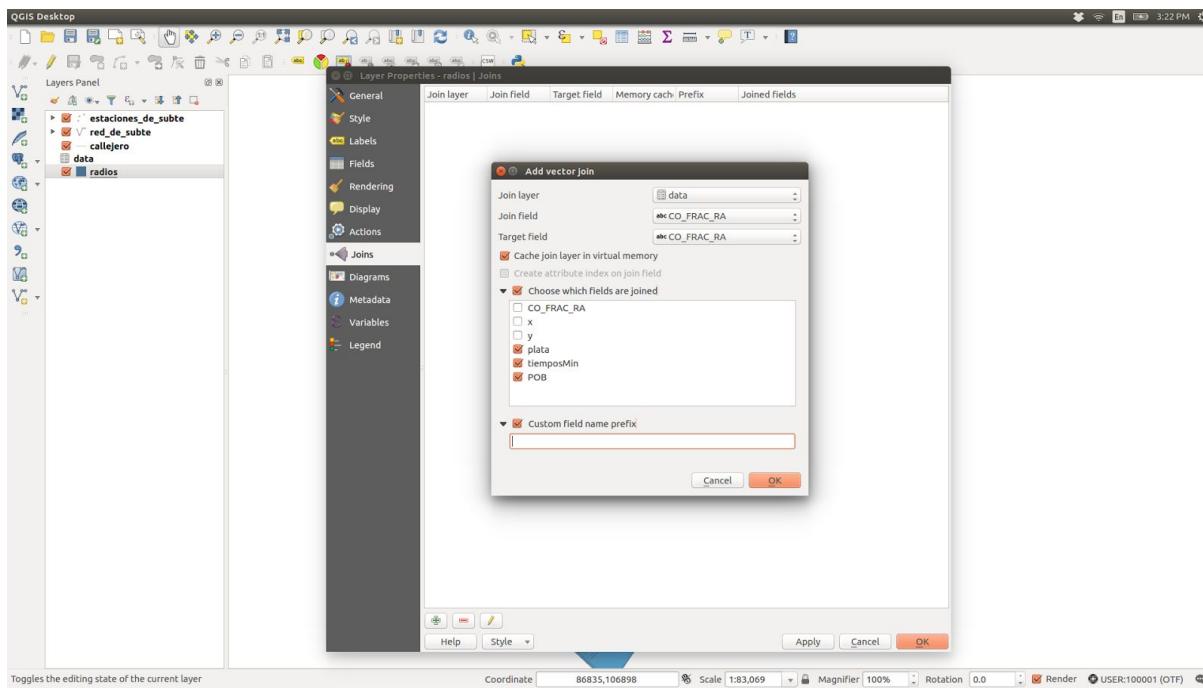
La otra fuente de datos que se puede utilizar en un sistema SIG es una tabla de datos tradicional. La misma puede estar en formato csv o Excel. En este caso no tendrán un índice espacial que las ubique en un espacio determinado, pero cada caso o registro deberá tener una variable identificadora única que coincida con la que tenemos en nuestra geometría. Esto no permitirá hacer una [unión o join](#) en base a ese atributo común. Para los radios censales este campo CO_FRAC_RA, que une los códigos de Comuna, Fracción y Radio. Usaremos un set de datos ya existente que contiene el costo de transportarse en pesos desde cada radio al centro, el tiempo en minutos que lleva y la población entre 18 y 55 años (*plata,tiemposMin,POB*)

Elementos a estar atento antes de unir:

- Tener en claro el set de datos objetivo o destino, es decir, en qué set de datos finalmente queremos que queden almacenadas las nuevas variables
- Tener en claro qué variable sirve como clave única de identificación de los casos
- El tipo de cada variable (numérica, texto, decimales, etc) para la clave única y para las otras variables a unir.
- Cómo lidiar con los casos que están en una tabla y no en la otra.
- Una unión NO modifica el set de datos original necesariamente. Se debe volver a guardar como un nuevo dataset



Para hacer la unión en QGIS se accede a las Propiedades de la capa a la que queremos unirle nuevos atributos, es decir nuestra capa destino u objetivo (los radios censales descargados). Seleccionamos la operación de Unión.



Seleccionamos la tabla de datos que contiene las variables que queremos agregar (*data.csv* en nuestro ejemplo)

Seleccionamos la variable que contiene la clave única de identificación de casos común a ambos set de datos (CO_FRAC_RA en nuestro ejemplo). Elegimos dicha variable en ambos set de datos ya que puede aparecer con diferentes nombres en el shapefile que en el csv.

Luego elegimos qué variables son las que deseamos unir (*plata,tiemposMin,POB*) y si queremos identificar a las mismas con algún prefijo (lo dejamos vacío).

Es importante chequear luego de la unión que hayan sido almacenadas con el mismo tipo de datos, es decir que por más que hayan sido originalmente números enteros no hayan sido copiadas como texto. Para esto podemos ver la *Propiedades de la capa* (en *Campos*)

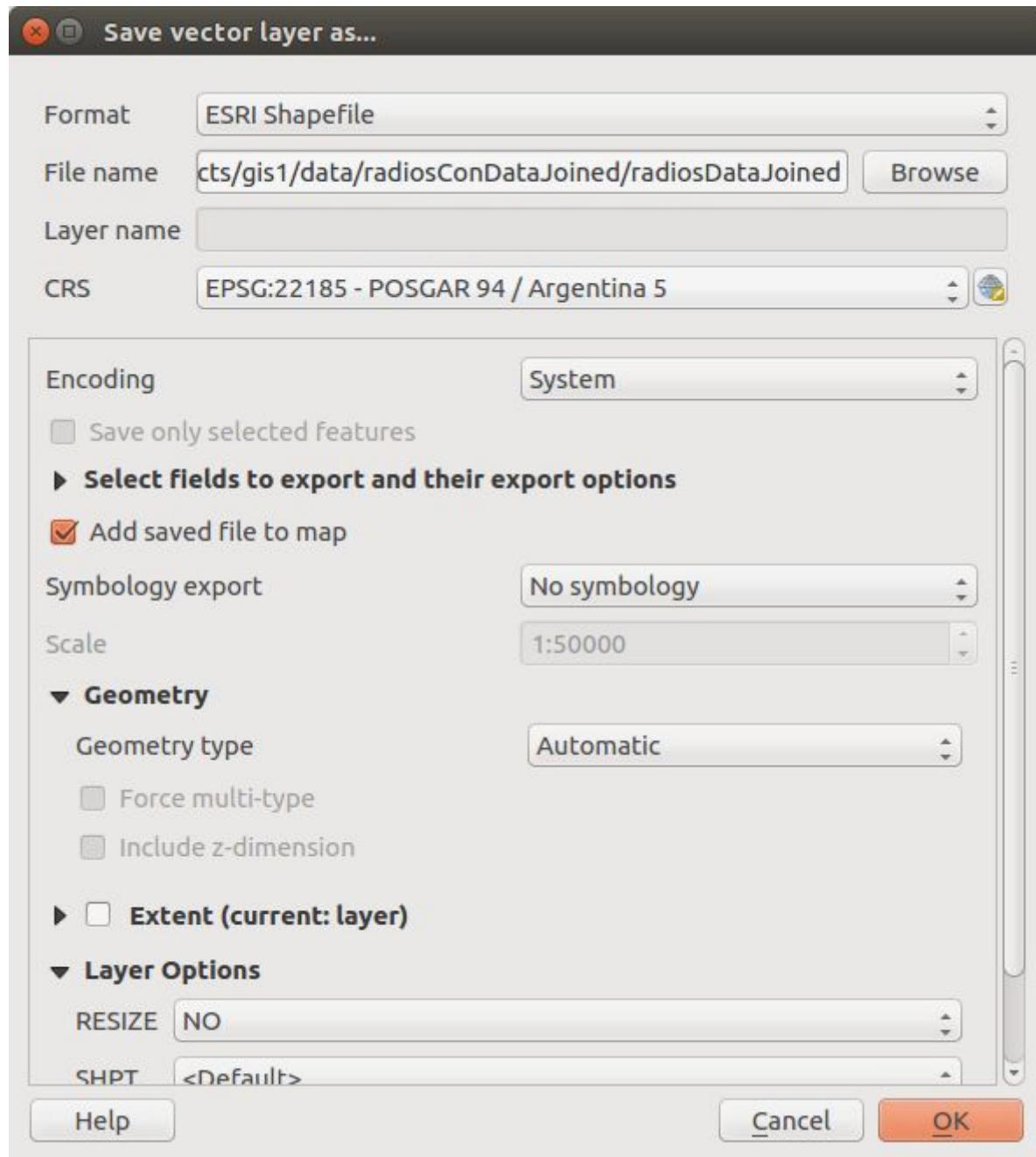
The screenshot shows the 'Layer Properties - radios | Fields' dialog in QGIS. The left sidebar contains icons for General, Style, Labels, Fields, Rendering, Display, Actions, Joins, Diagrams, Metadata, Variables, and Legend. The main area has tabs for Attribute editor layout (Autogenerate) and Python Init function. The 'Fields' tab is selected, displaying a table of attributes:

ID	Name	Edit widget	Alias	Type	Type name	Length	Precision	Comm
123 0	ID	Text Edit		int	Integer	10	0	
abc 1	CO_FRAC_RA	Text Edit		QString	String	40	0	
123 2	COMUNA	Text Edit		int	Integer	2	0	
123 3	FRACCION	Text Edit		int	Integer	10	0	
123 4	RADIO	Text Edit		int	Integer	2	0	
1.2 5	TOTAL_POB	Text Edit		double	Real	11	0	
1.2 6	T_VARON	Text Edit		double	Real	11	0	
1.2 7	T_MUJER	Text Edit		double	Real	11	0	
1.2 8	T_VIVIENDA	Text Edit		double	Real	19	0	
1.2 9	V PARTICULAR	Text Edit		double	Real	15	0	
1.2 10	V COLECTIV	Text Edit		double	Real	19	0	
1.2 11	T_HOGAR	Text Edit		double	Real	19	0	
1.2 12	H_CON_NBI	Text Edit		double	Real	19	0	
1.2 13	H_SIN_NBI	Text Edit		double	Real	19	0	
1.2 14	plata	Text Edit		double	double	0	0	
1.2 15	tiemposMin	Text Edit		double	double	0	0	
1.2 16	POB	Text Edit		int	integer	0	0	

Below the table, the 'Relations' tab is shown. At the bottom, there are buttons for Help, Style, Apply, Cancel, and OK. A note says 'Suppress attribute form pop-up after feature creation Default'. The 'OK' button is highlighted in orange.

Podemos ver que las alacena como números con decimales (double) o redondos (integer). Al mismo tiempo podemos ver que están indicadas con un ícono especial. Esto quiere decir que los datos no están almacenados en el set de datos, si no que hay un mero vínculo o unión entre dos set de datos diferentes. Si borramos data.csv o lo cambiamos de lugar en nuestros directorios, estos datos se van a perder. Para que esto no suceda, guardamos el shapefile como un archivo aparte. Esto va a resguardar los datos de manera segura.

Con clic derecho en la capa *Guardar como*. Guardamos en otro archivo como ESRI Shapefile, indicamos la ruta y **ELEGIMOS UNA PROYECCIÓN**.



7 - Buffers o áreas de influencia

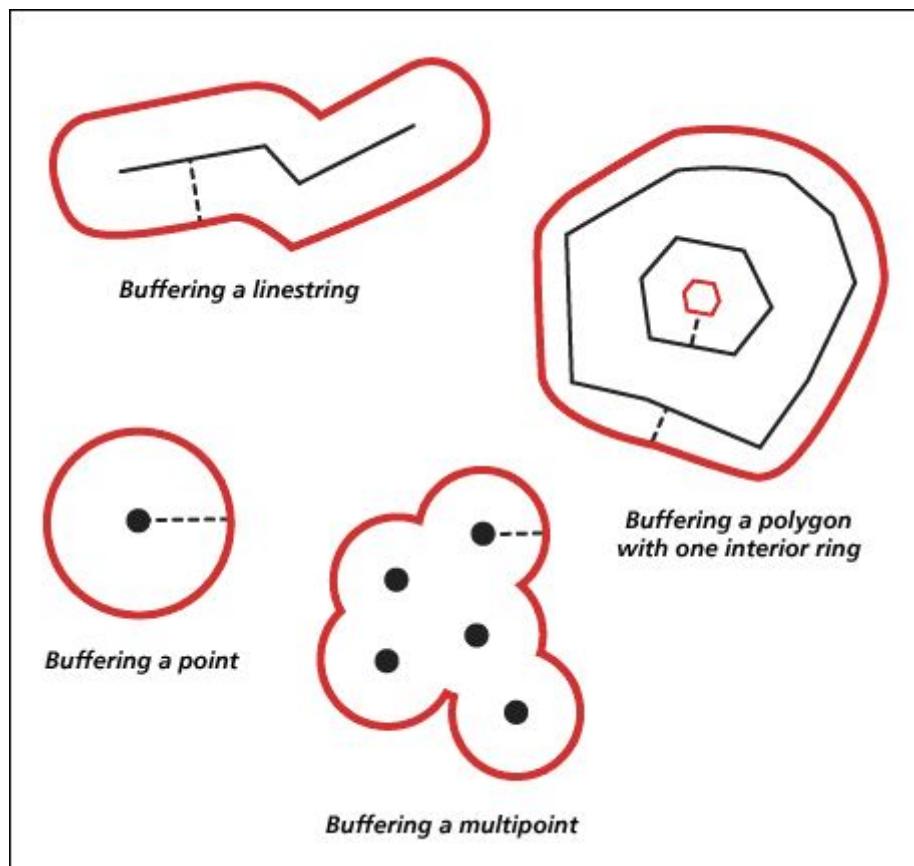
Ya tenemos nuestros datos fundamentales, la materia prima para nuestro análisis. Ahora vamos a comenzar con el mismo. Como es un taller introductorio vamos a comenzar por el análisis espacial más fundamental en torno a una de las principales propiedades del espacio: las distancias. Vamos a pensar este problema en función de elementos que están cerca y lejos.

Recordemos nuestro objetivo. Ubicar estaciones de bicicletas públicas y compartidas en lugares que estén dentro de una distancia *bicicleable* del subte, pero no tan cerca que sea una distancia *caminable*.

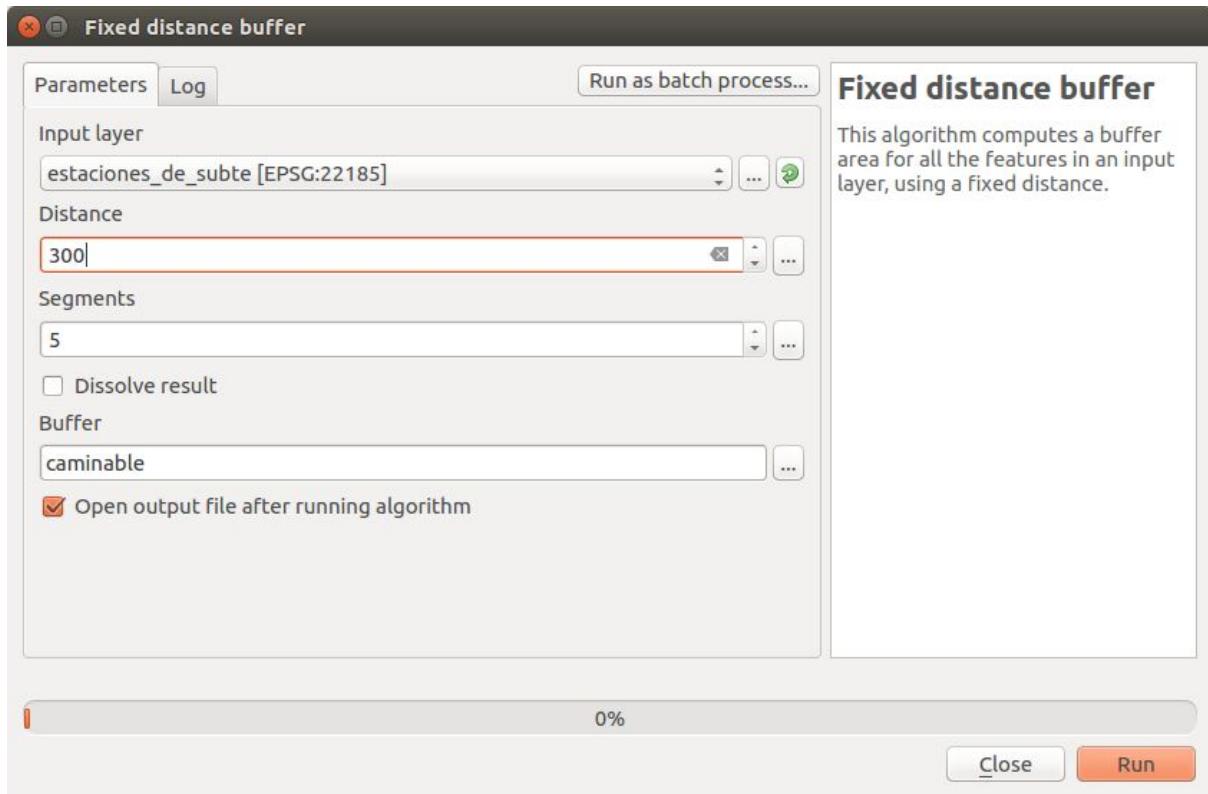
Ya sabemos que la gente que vive en Villa Soldati o Villa Pueyrredón no va a ir en bicicleta hasta el subte. Pero quizás los de Paternal si. Entonces comenzemos por eliminar del análisis todos esos radios que están **FUERA** de una distancia *bicicleable* del subte y a aquellos que están **DENTRO** de una distancia *caminable* del subte (esos ya sabemos que directamente caminarán al subte). Nos quedan algunas preguntas antes: qué significa *caminable*? qué significa *bicicleable*?

Vamos a explicitar algunas de nuestras suposiciones. Primero suponemos que todo el mundo camina hasta 3 cuadras máximo para ir al subte. Si vive a más de 5 cuadras del subte, asumimos que no va a caminar al subte y va buscar un colectivo que pase más cerca. También vamos a suponer que todo el mundo está dispuesto cada mañana a andar en bicicleta hasta 20 cuadras. Entonces *caminable* es 5 cuadras y *bicicleable* es 20 cuadras.

La principal herramienta de análisis de los sistemas SIG para esto es la zona de influencia o *buffer*. Básicamente lo que hace es trazar una distancia constante a partir de un objeto espacial. Se puede hacer un buffer sobre un punto, una línea o un polígono.



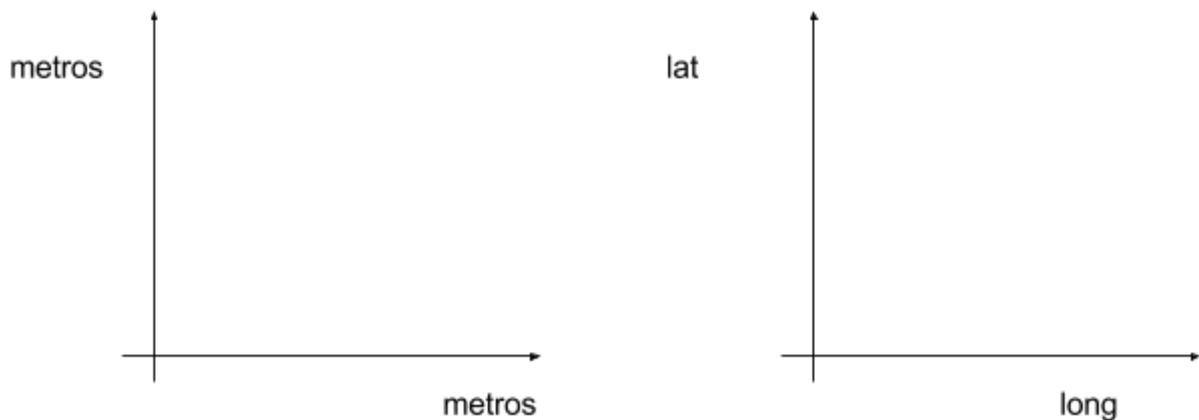
Para esto en QGIS podemos escoger Buffer en las Herramientas de Geoprocreso. Esta herramienta va a extender un área alrededor de un objeto espacial en base a una distancia. Esto nos va a permitir saber las áreas *caminables* y *bicicletables*.



Comencemos con el área *caminable*. Debemos escoger cual es el objeto espacial en torno al cual queremos extender un área de influencia. Sabemos que caminable es 3 cuadras del subte. Pero, qué significa *del subte*? Uno puede vivir cerca de donde *pasa* el subte, pero no puede saltar al medio del túnel. Entonces, más allá de que técnicamente podamos aplicar un buffer por el recorrido del subte, por una línea, no significa que debamos.

En este punto es donde debemos poner las herramientas SIG en el contexto de nuestro análisis. Entender qué significa un punto o una línea en nuestra pregunta. Separar la sintaxis del SIG (puntos, líneas, polígonos) de la semántica (estaciones, recorridos, etc).

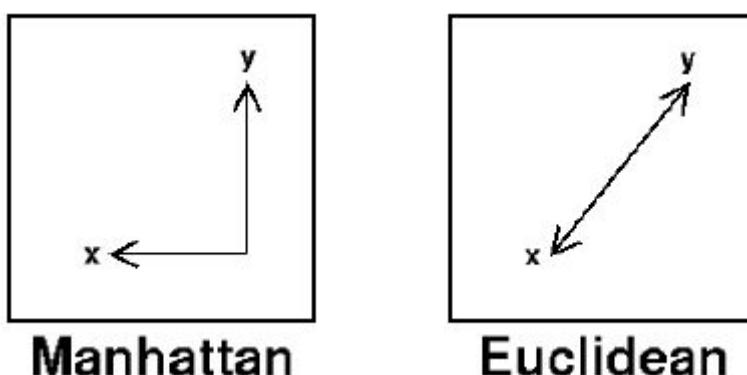
Entonces escogemos las estaciones de subte, los puntos. También necesitamos calcular una distancia. Ahora es dónde todo el tema proyecciones cobra sentido. Porque las distancias debemos darlas en *unidades de medida*. Dijimos caminable es 3 cuadras. Pero en qué unidades de medida está nuestra proyección, qué hay en las coordenadas de nuestros ejes x y y?



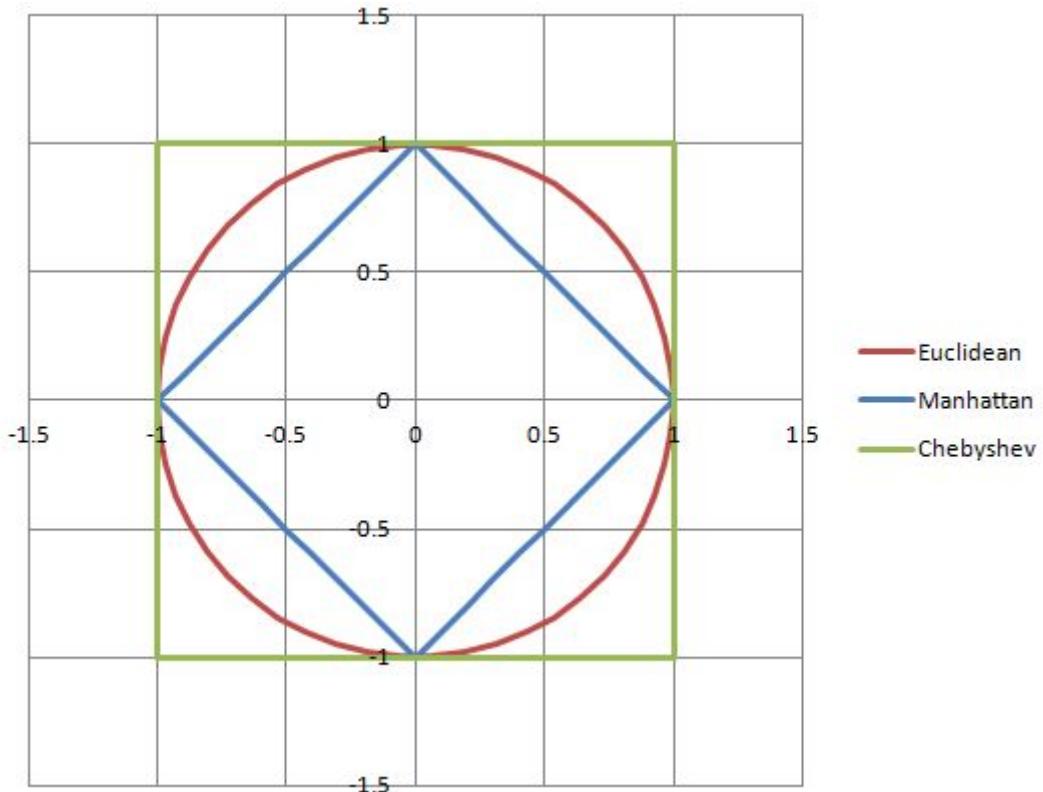
Esto también nos exige comprender nuestro problema, nuestra proyección y nuestras unidades de medida. Porque sin eso, lo que pongamos en un recuadro de texto de un programa de computadora no tendrá sentido. Sabemos que 3 cuadras son 300 metros o 0.3 kilómetros. POSGAR no está en cuadras claramente, con lo cual no podemos poner 3. Si estuviese en latitud y longitud habría que poner una distancia en grados minutos y segundos que equivalga a 300 metros. Las proyecciones para Estados Unidos utilizan el sistema imperial y están en pies y millas. Afortunadamente POSGAR está en metros, con lo cual podemos ingresar 300 directamente.

Qué tipo de distancia?

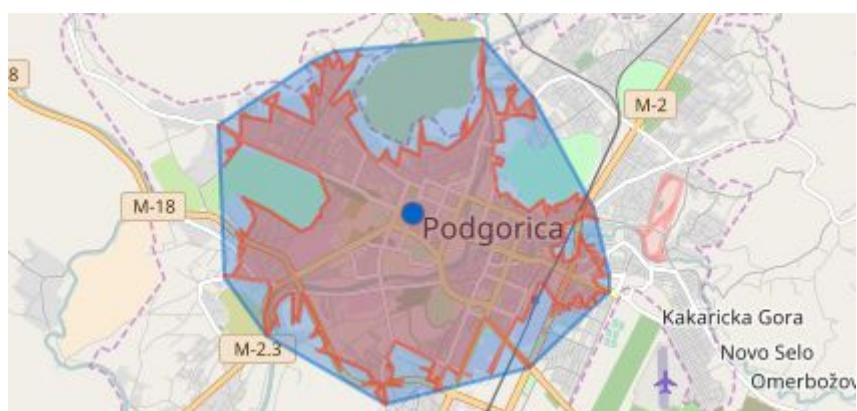
Hilar fino sobre el tipo de distancia excede los límites de este curso. Pero es importante que sepamos que existen varios tipos de distancia y ser críticos con las herramientas que utilizamos, para conocer sus limitaciones. Normalmente utilizamos la distancia euclíadiana y lineal. Esta es la que comúnmente se llama distancia a la redonda, y el área que aroja un buffer de distancia lineal es tiene forma redonda. Pero en contextos urbanos, uno no puede recorrer una distancia lineal, no puede saltar por los edificios. Normalmente uno camina por las calles y recorre distancias mayores a la lineal, directa o euclidiana. Existe otra distancia que se llama Manhattan en base a las calles en grillas de la ciudad de Estados Unidos.



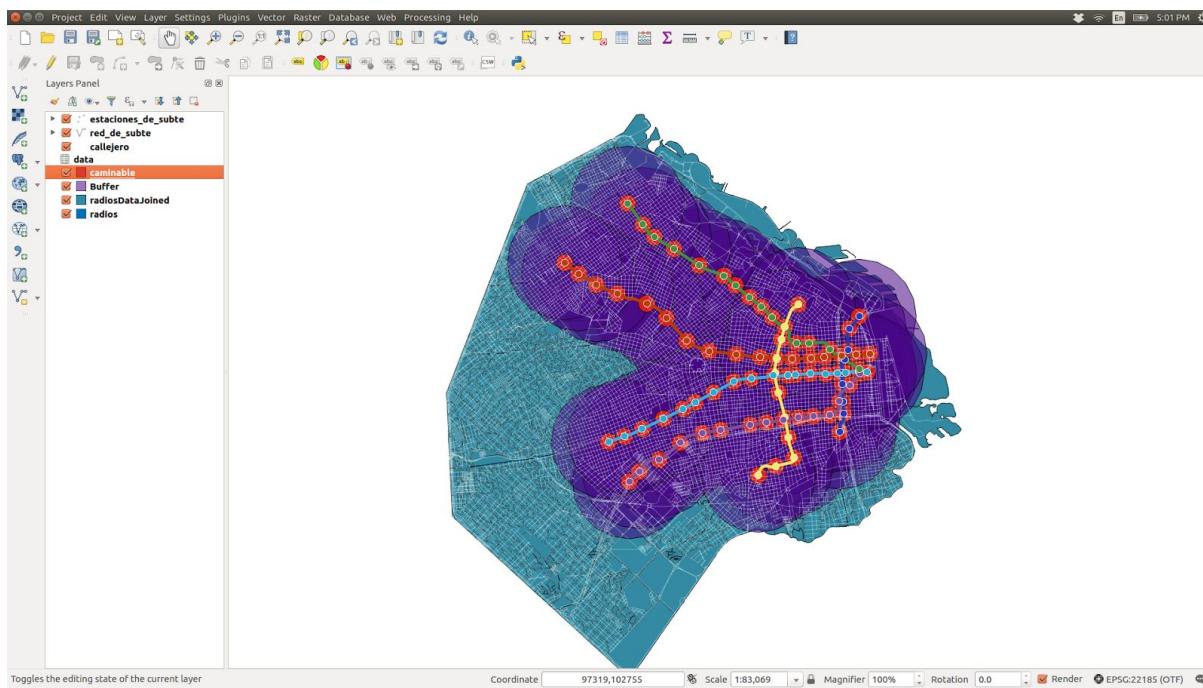
Esas distancias son más apropiadas para contextos urbanos. En especial ante la presencia de barreras urbanas como vías de tren, donde uno tiene que caminar más para recorrer la misma distancia lineal.



Normalmente utilizar distancias de Manhattan arroja otro tipo de polígonos como resultado del buffer. Más alineado con los ejes. Si uno utiliza la red vial en sí misma, esos buffers seguirán a lo largo de esas vías.



Luego de este desvío, obtuvimos los buffer para las distancias caminables y repetimos lo mismo para las distancias bicicleteables.



8 - Selección

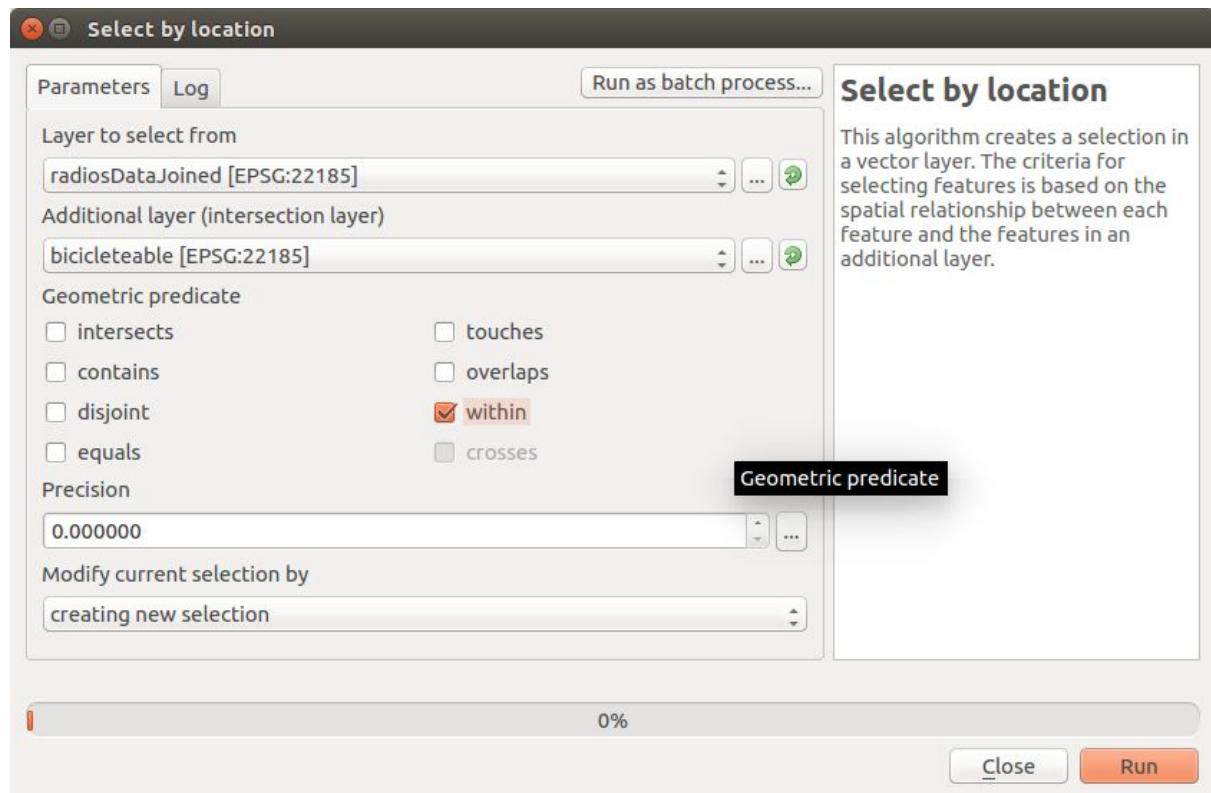
Habiendo determinado las áreas de influencia (buffers) del subte (diferente influencia para caminar que para andar en bici), ahora podemos seleccionar los radios censales que nos interesan: los que están **DENTRO** del área *bicicleable* y los que están **FUERA** del área caminable.

Para eso recurrimos a una selección por ubicación. En las herramientas SIG uno puede seleccionar elementos a partir del valor que toman en una variable (por ejemplo seleccionar todos los puntos que pertenezcan a la línea H) como se hace tradicionalmente en cualquier herramienta de administración de datos. Pero en una herramienta de datos de análisis espacial, queremos seleccionar elementos en base a su ubicación o en base su relación espacial con otros elementos.

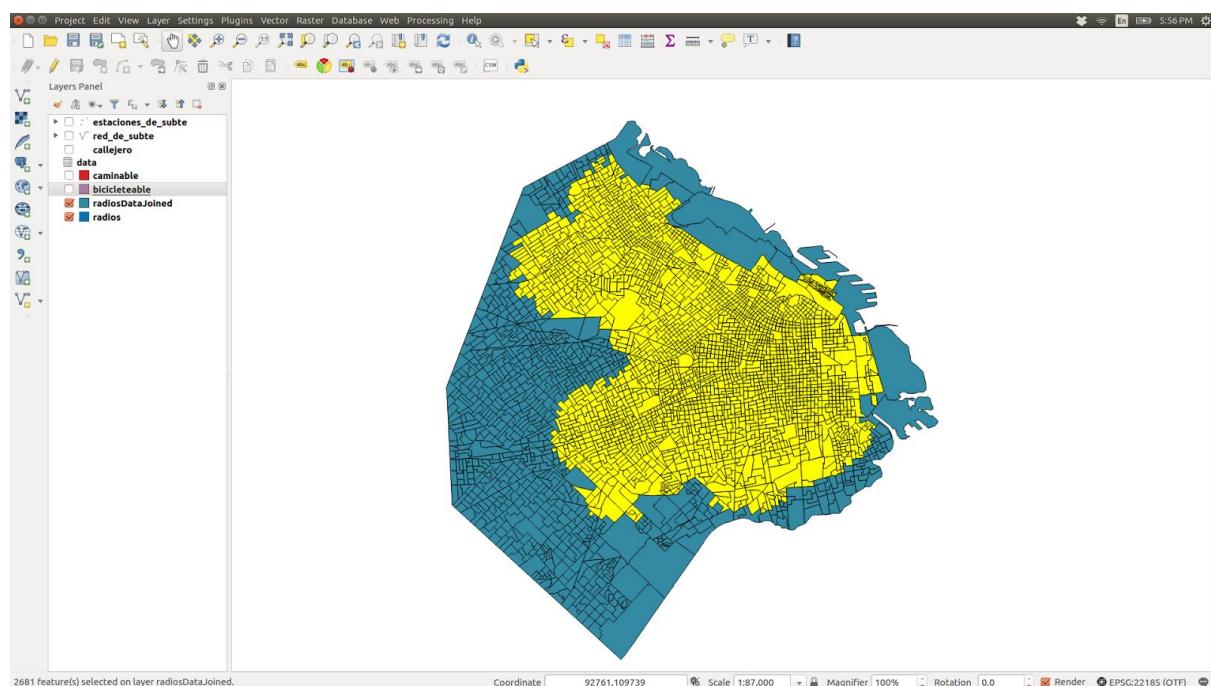
En este caso ya establecimos el criterio. Queremos los radios censales DENTRO del área *bicicleable* y FUERA del área *caminable*. Comencemos con los primeros. Vamos a seleccionar esos elementos desde *Vector > Herramientas de investigación > Seleccionar por ubicación*. Esa ubicación, o relación con otro elemento espacial puede ser diversa y varía de acuerdo a qué queremos relacionar con qué. Toca, contiene, intersecta, son todas opciones diferentes con diferentes sentidos de acuerdo a si queremos relacionar puntos con polígonos, puntos con líneas, polígonos con polígonos, etc.

Elegimos primero los elementos de qué capa queremos seleccionar y luego la capa que va a realizar la intersección. Nosotros queremos seleccionar radios censales, y queremos seleccionar los que están dentro del área bicicleable. Finalmente el criterio con el que se

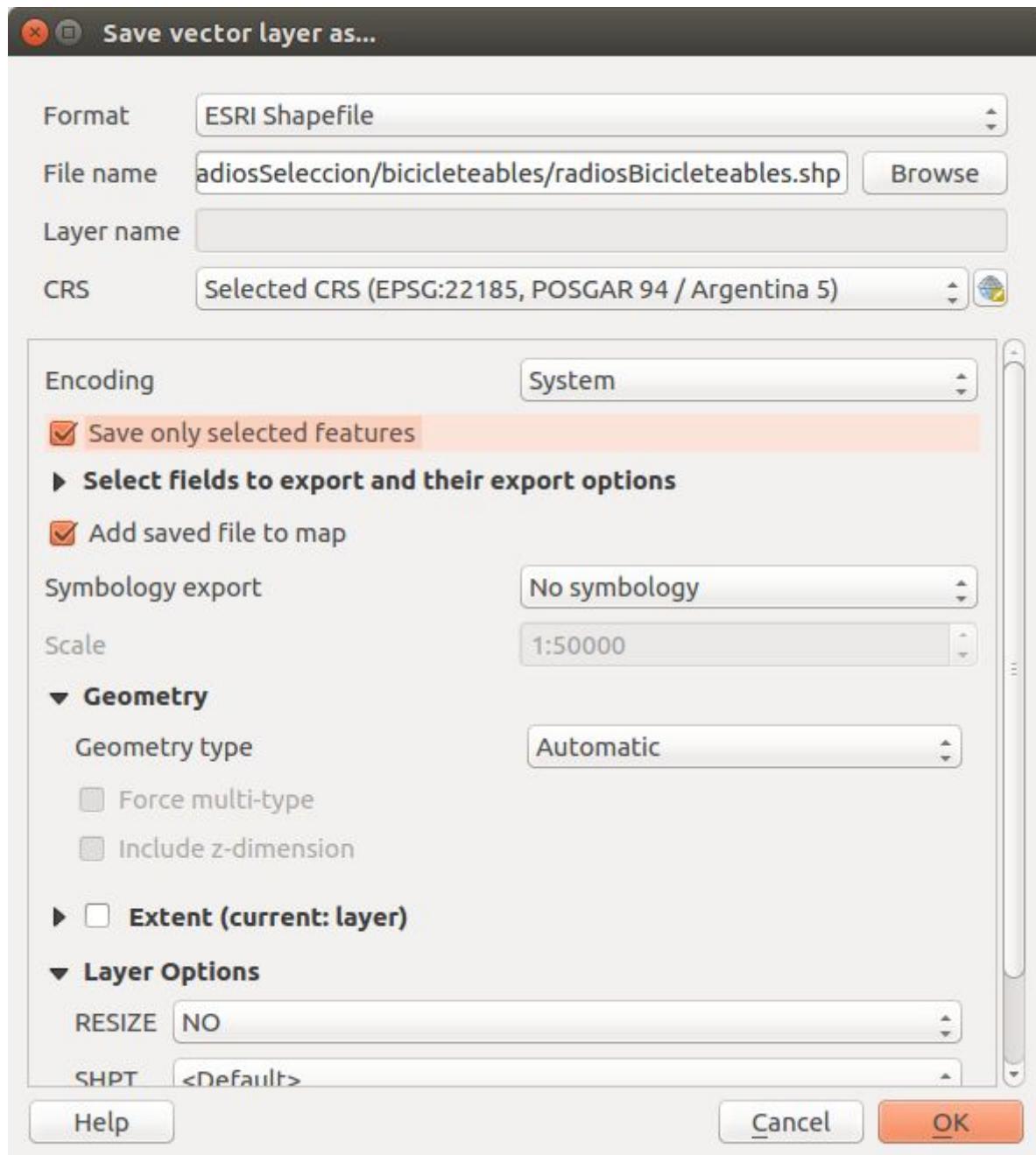
intersecta y el margen de error. Como inicialmente pensamos DENTRO DE, elegimos ese criterio de intersección.



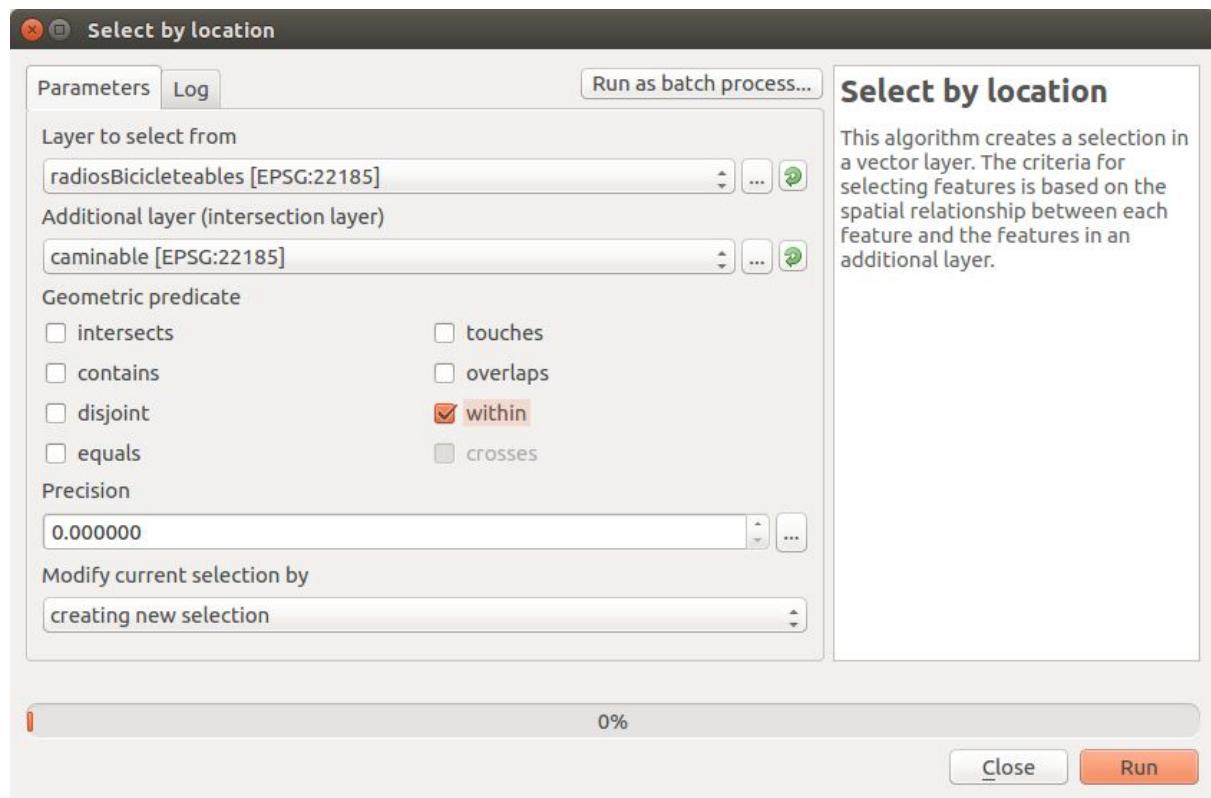
En QGIS los elementos seleccionados quedan representados en amarillo



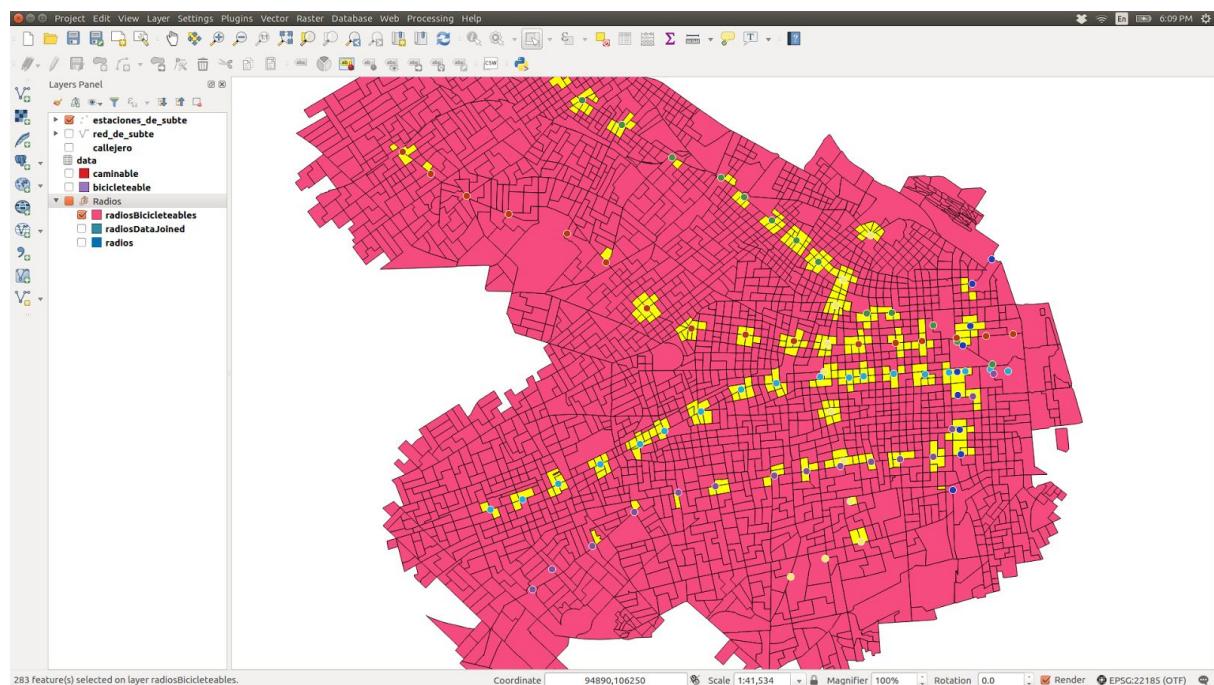
El paso siguiente sería borrar los otros radios. Hay diversas maneras de hacer esto, pero quizás la manera más prolífica y que a su vez permite ir dejando archivos de trabajo intermedios, es guardar esta selección en un shapefile separado. Para eso la guardan se selecciona la opción de guardar solo los elementos seleccionados. Una vez que hicimos esto procuramos eliminar toda selección. Ya que las operaciones en SIG son sensibles a las selecciones que tenemos hechas. Esto se puede hacer con el ícono de la barra de tareas o seleccionando un elemento vacío fuera del mapa.



Luego de guardar esta selección como radiosBicicleables.shp continuamos. Dentro de estos radios, queremos solo los que están FUERA DE la distancia caminable. Esto se puede hacer de diferentes maneras. La que elegimos es repetir el mismo proceso, seleccionando los que están DENTRO de dicha distancia, e invertir esa selección.

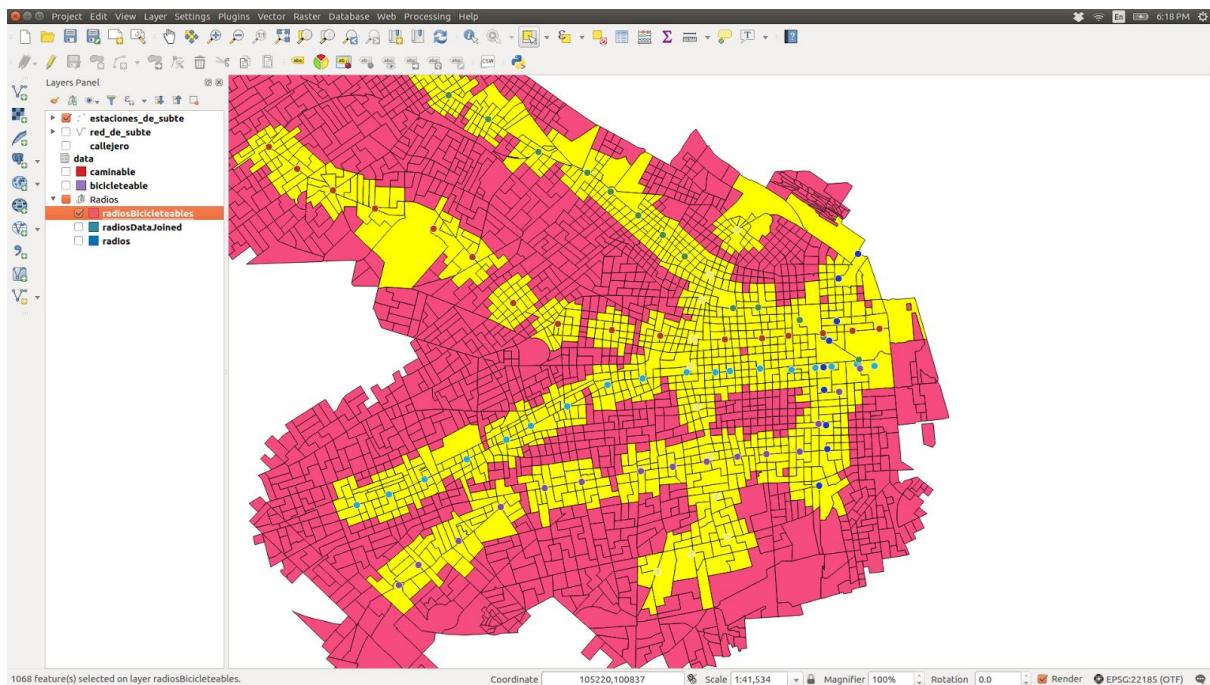


Seleccionamos desde los radios que guardamos previamente, ya que queremos hacer una selección dentro de la selección anterior. Escogemos ahora como capa para interceptar al área caminable.



Primero vemos que no estamos contemplando los radios fuera del área *bicicleable*. Fuera quedaron los radios de Villa Pueyrredon. Segundo, vemos un problema de utilizar el criterio DENTRO. Algunas estaciones, en especial las más cercanas a las terminales, alejadas del centro, con baja densidad poblacional, y por ende con radios muy extensos en área, no logran captar dentro de 300 metros al radio en su totalidad. Dicho de otra manera, como son radios grandes, ninguno entra EN SU TOTALIDAD, en un radio de 300 metros. **Esto tiene que ver con el Problema de la Unidad Espacial Modifiable, del cual habrá un taller el lunes 27/11.**

Repetimos el proceso y utilizamos otro criterio más contemplativo.



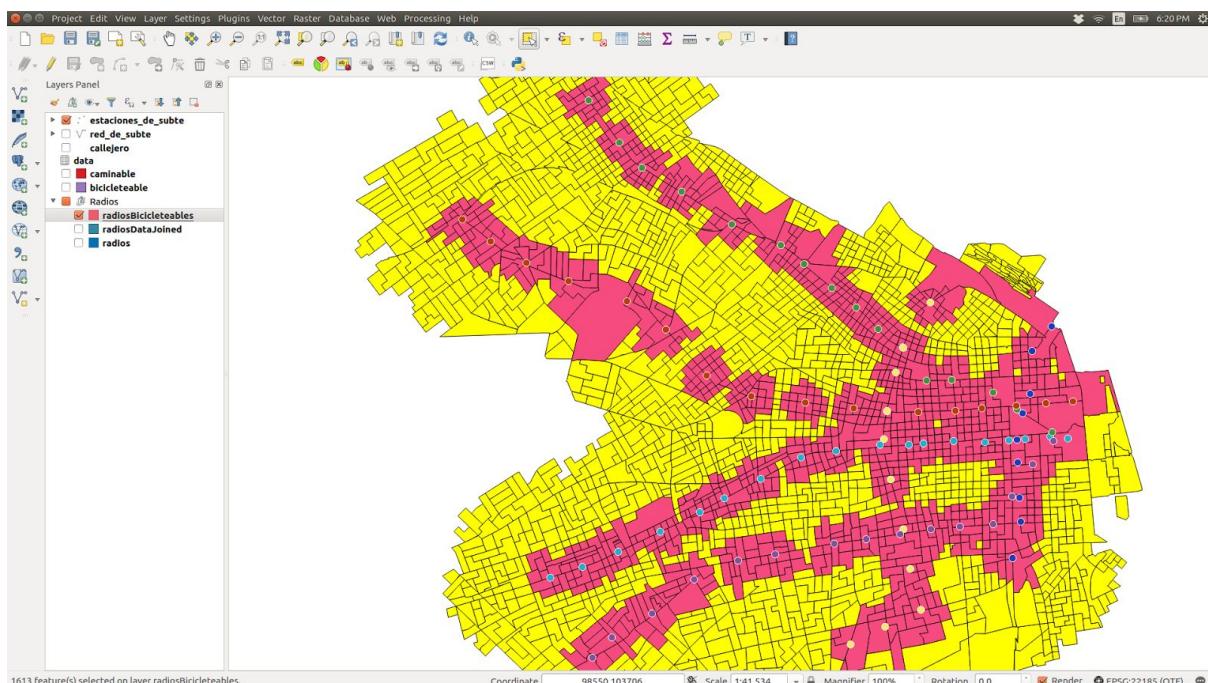
A continuación abrimos la tabla de atributos haciendo clic derecho sobre la capa de *radiosBicicleables* y elegimos *Invertir selección*.

radiosBicicletables :: Features total: 2681, filtered: 2681, selected: 1068

	ID	CO_FRAC_RA	Invert selection (Ctrl+R)	RADIO	TOTAL_POB	T_VARON	T_MUJER	T_VIVIENDA	V PARTICULAR	V COLECTIV	T_HO
1	2_1_12_1			1	12	1	341	184	157	365	361
2	3_1_12_10			1	12	10	296	162	134	629	627
3	4_1_12_11			1	12	11	528	294	234	375	370
4	5_1_12_2			1	12	2	229	101	128	445	445
5	6_1_12_3			1	12	3	723	374	349	744	739
6	7_1_12_4			1	12	4	393	168	225	341	339
7	8_1_12_5			1	12	5	600	264	336	505	504
8	9_1_12_6			1	12	6	472	220	252	504	502
9	10_1_12_7			1	12	7	786	381	405	546	542
10	11_1_12_8			1	12	8	329	162	167	275	269
11	12_1_12_9			1	12	9	1356	650	706	895	885
12	13_1_13_1			1	13	1	586	364	222	404	402
13	14_1_13_4			1	13	4	725	449	276	641	640
14	15_1_13_13			1	13	13	358	226	132	266	262
15	16_1_13_9			1	13	9	426	211	215	252	251
16	17_1_13_10			1	13	10	614	312	302	641	641
17	18_1_13_11			1	13	11	292	146	146	382	382
18	19_1_14_1			1	14	1	622	325	297	538	530
19	20_1_14_10			1	14	10	431	207	224	341	341
20	21_1_14_11			1	14	11	522	254	268	412	412
21	22_1_14_12			1	14	12	511	241	270	309	307

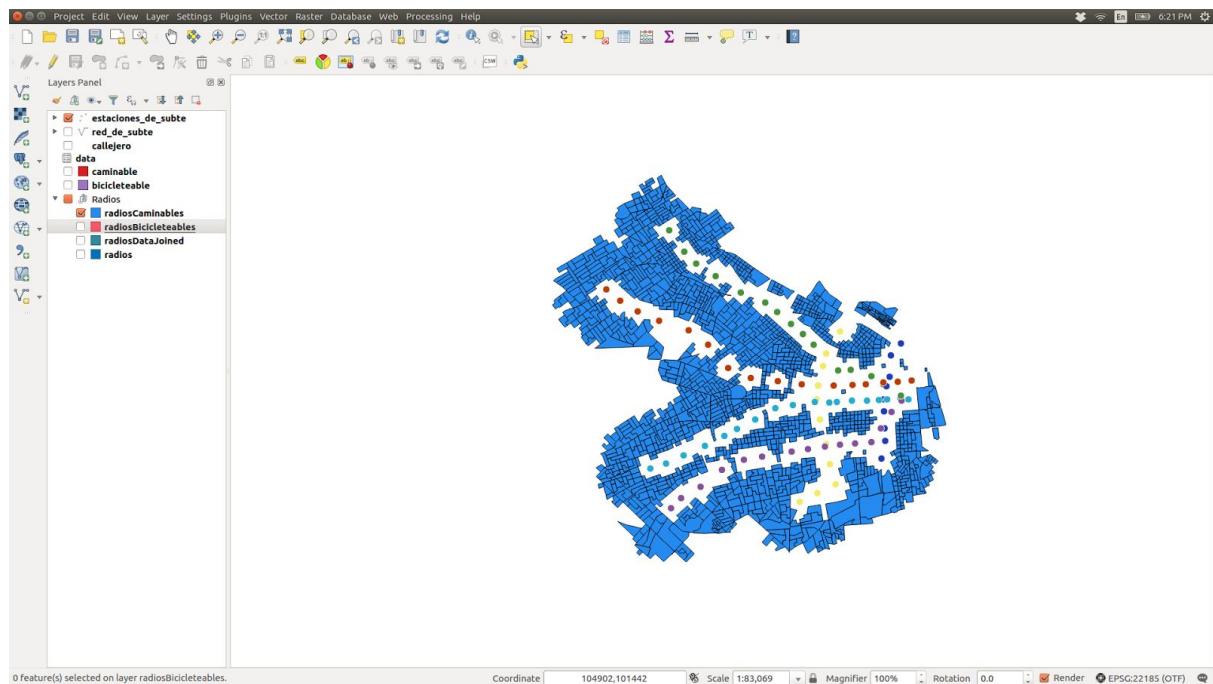
Show All Features

De esta manera nos quedamos con los que están FUERA del área caminable. Repetimos el proceso de guardar esta selección como un shapefile independiente, con la proyección correspondiente y un nombre legible.



Observemos que estamos borrando el polígono del cementerio de Chacarita, que es demasiado grande. La gente que vive del otro lado del cementerio podría hacer uso de la bici para ir al subte en estación Lacroze. Los radios censales que contienen parques,

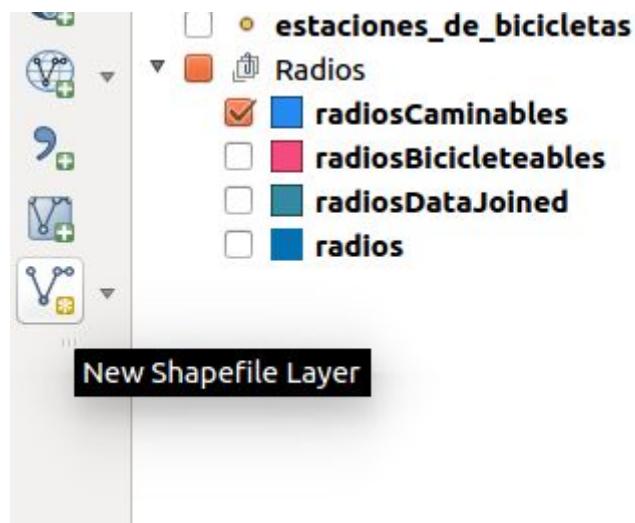
cementerios, prisiones, son problemáticos y siempre merecen ser tratados de un modo especial. Esta es otra expresión del **Problema de la Unidad Espacial Modificable**.

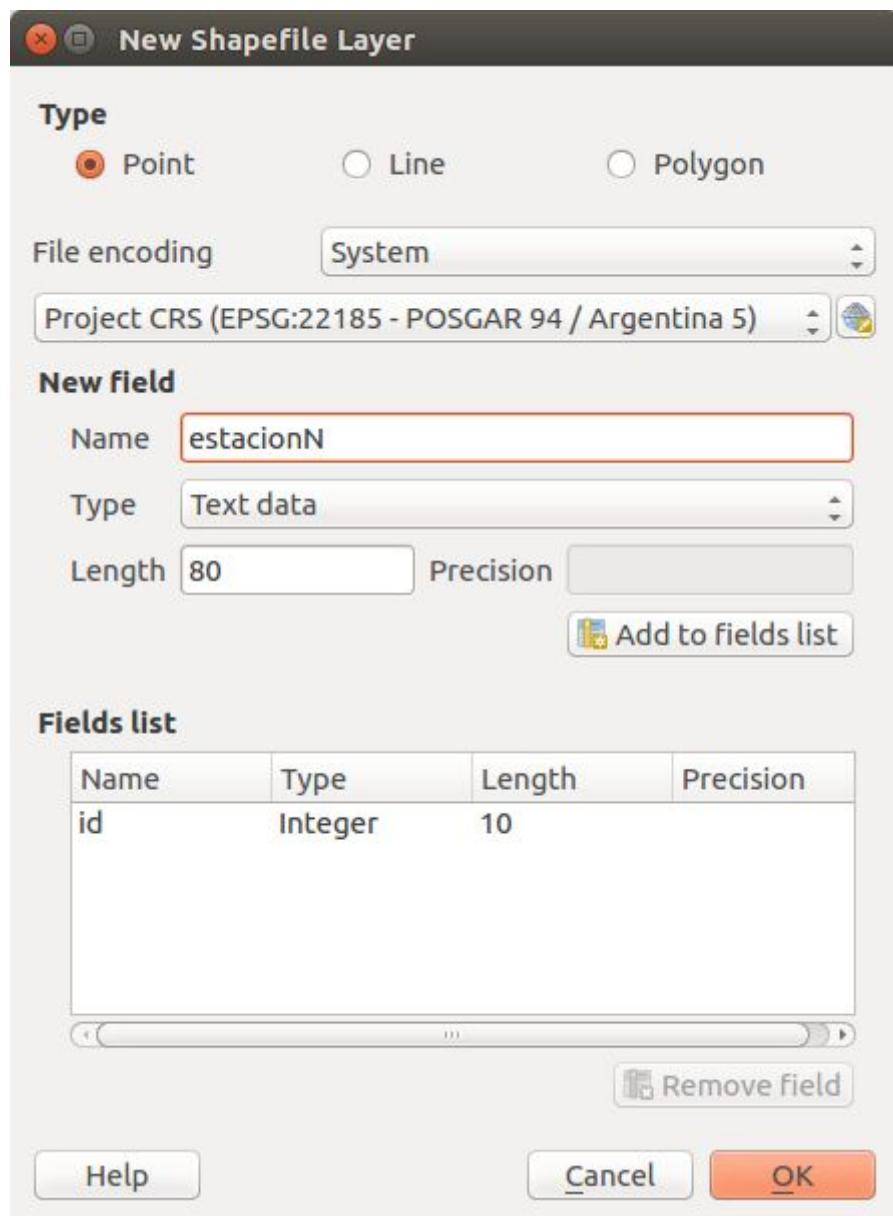


9 - Asignar estaciones

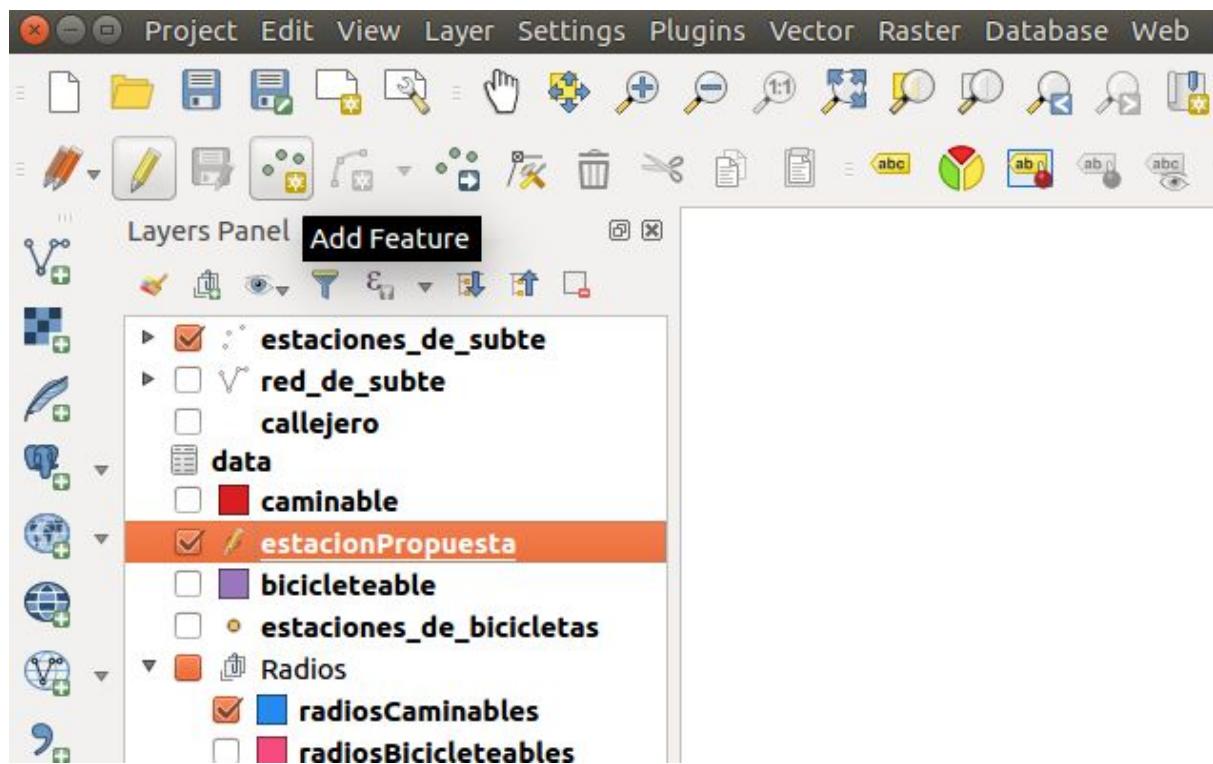
Ya tenemos el conjunto de radios posibles donde podría haber bicicletas públicas compartidas. El paso siguiente es comenzar a asignar posibles lugares para luego, en el último paso del taller, medir el desempeño de cada una.

Para hacer esto vamos a crear nuestra propia capa espacial desde cero. Elegimos como tipo un punto y la proyección correspondiente.

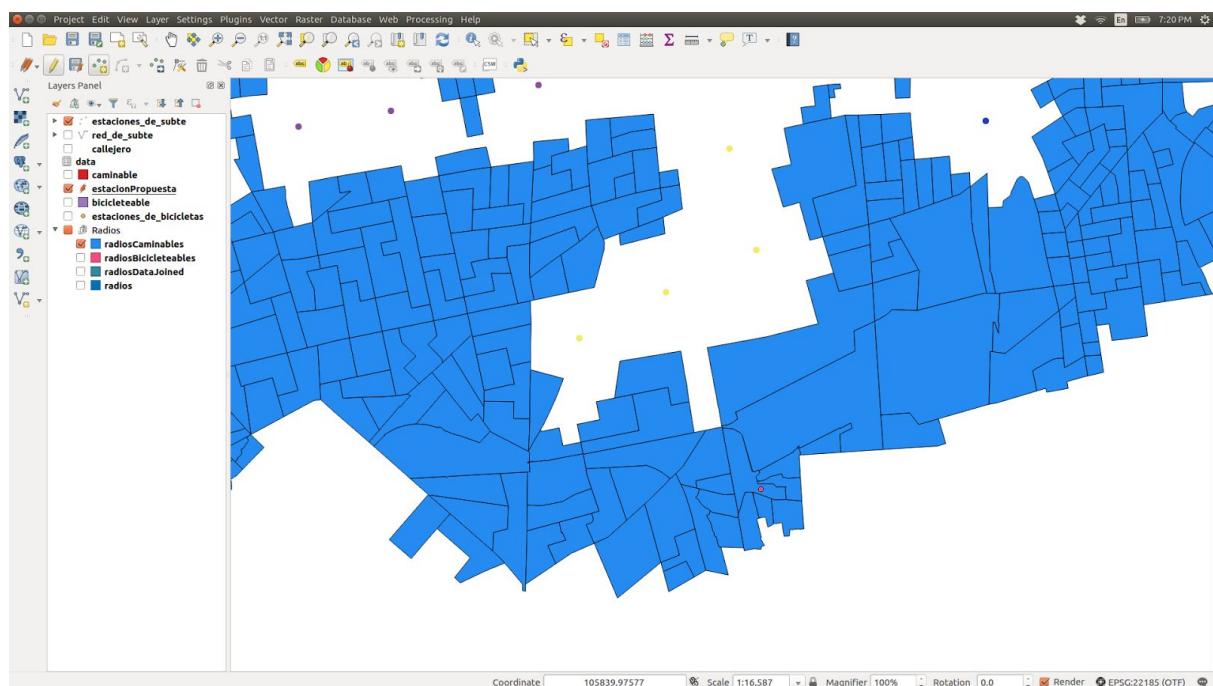




A continuación QGIS abrirá una ventana nueva para pedirnos dónde guardar dicho shapefile. Hasta que no indiquemos eso, no nos podrá dejará continuar. Luego seleccionamos la capa y hacemos clic en el ícono del lápiz que permitirá editar dicho shapefile. Tenemos un shapefile vacío y queremos poblarlo con estaciones nuevas. La edición del shapefile permite crear objetos espaciales y darle valor a sus atributos. Es importante recordar que una vez realizado un cambio no puede volverse atrás. No existe la operación *Deshacer*, con lo cual una vez que se comienza a editar un shapefile hay que ser muy cuidadoso e intentar tener una copia de seguridad por si algo sale mal. Por lo pronto nuestra estación propuesta no tendrá ningún atributo, pero podría agregarse por ejemplo la capacidad.



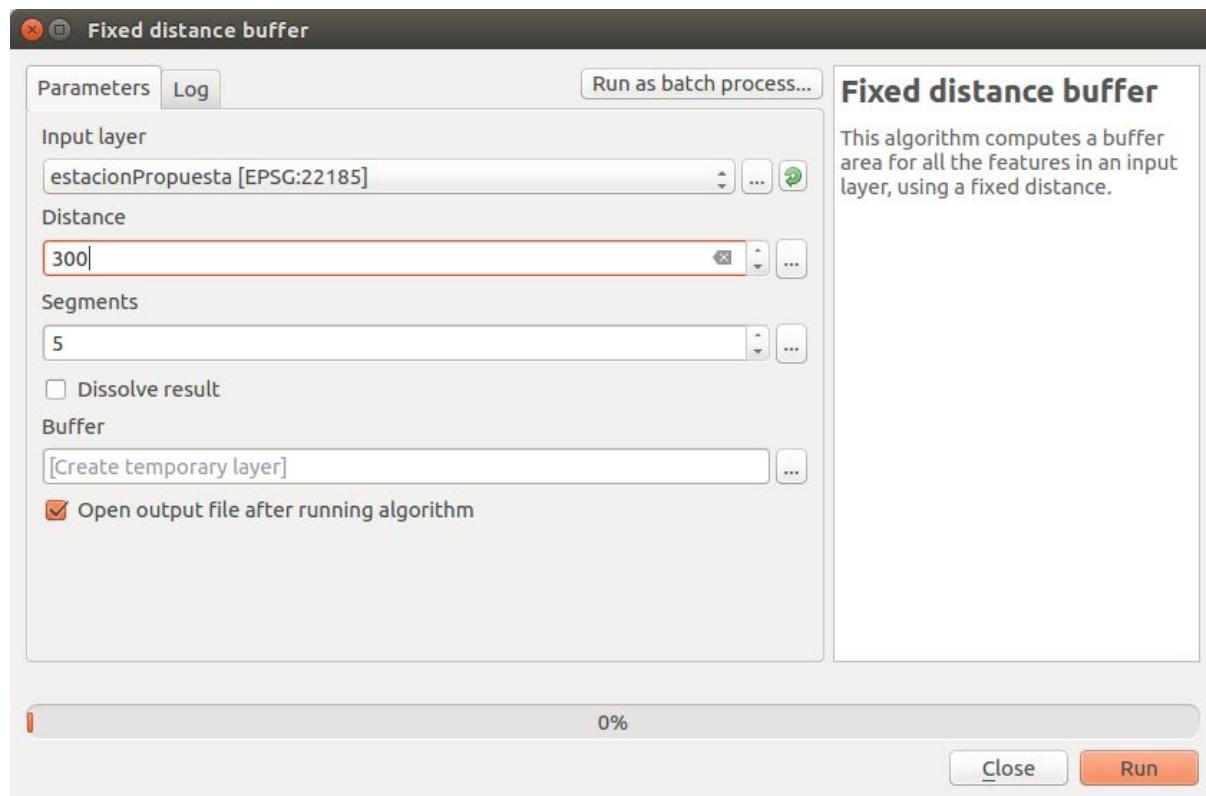
Agregamos la estación donde creamos más conveniente y luego guardamos los cambios y finalizamos la edición. Ya hemos modificado nuestro shapefile.



10 - Uniones espaciales o Spatial Joins

Una vez colocada la estación, el objetivo ahora es ver que tipo de efectividad tendrá nuestra estación. Para ello, podemos asumir un área de influencia o de impacto de la misma. Luego podemos ver dentro de ese área indicadores como cuánta gente vive allí, cuánto demora esa gente en llegar al centro todos los días o cuánto gasta. De este modo podemos aplicar diferentes criterios para asignar estaciones. Podemos directamente decir que queremos influir a la mayor cantidad de gente posible. O podemos querer impactar a aquellos que tienen que gastar más para ir al centro porque tienen que hacer trasbordos.

Para ello, el primer paso es crear un área de influencia basada en lo que creemos que la gente estará dispuesta a caminar desde su casa para acceder a una estación de bici. Supongamos 300 metros.



Creamos el buffer y lo guardamos como shapefile con la proyección correcta.

El proceso que sigue a continuación es otra de las herramientas más comunes del análisis espacial, la **unión o join espacial**. Es decir, tenemos un polígono (nuestra área de influencia) al cual le queremos unir datos de otros polígonos de acuerdo a qué relación en el espacio tienen: se superponen o no y cómo se superponen.

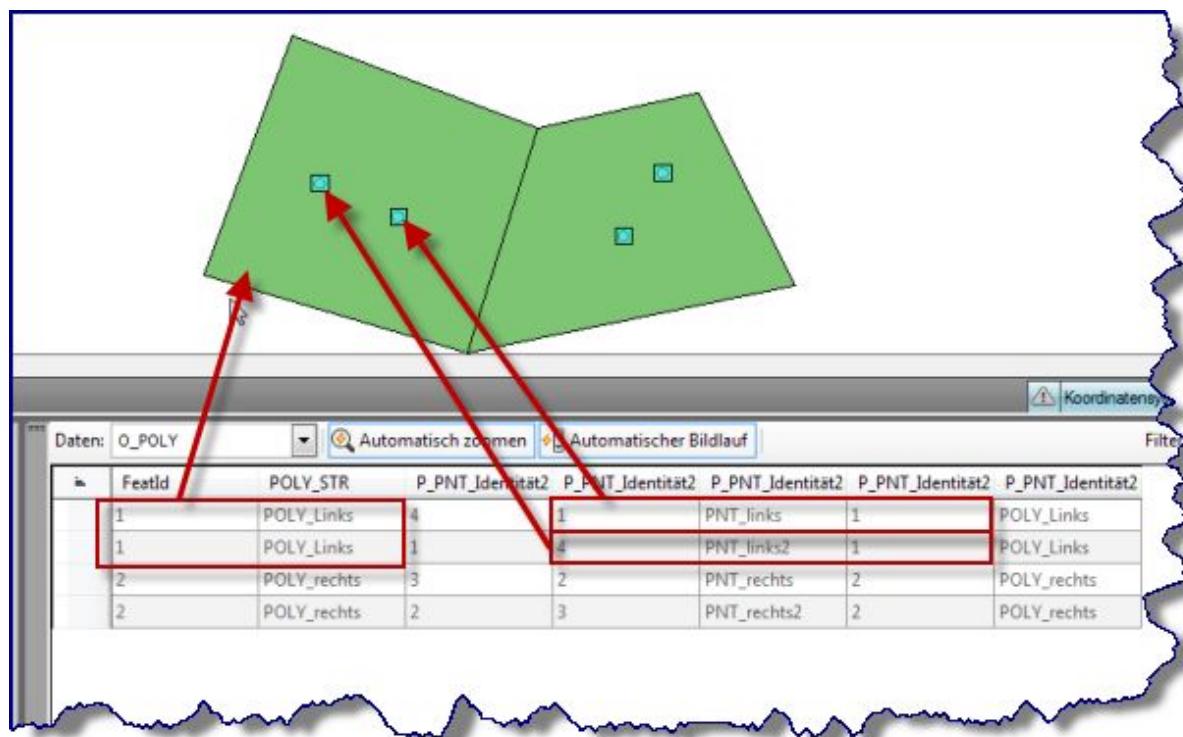


Table 1: Reported Incidents

Date Reported	Blockage 	Flooding 	Odour 
01/02/03	3		4
02/02/03		2	
03/02/03	2	2	1

Geoplan

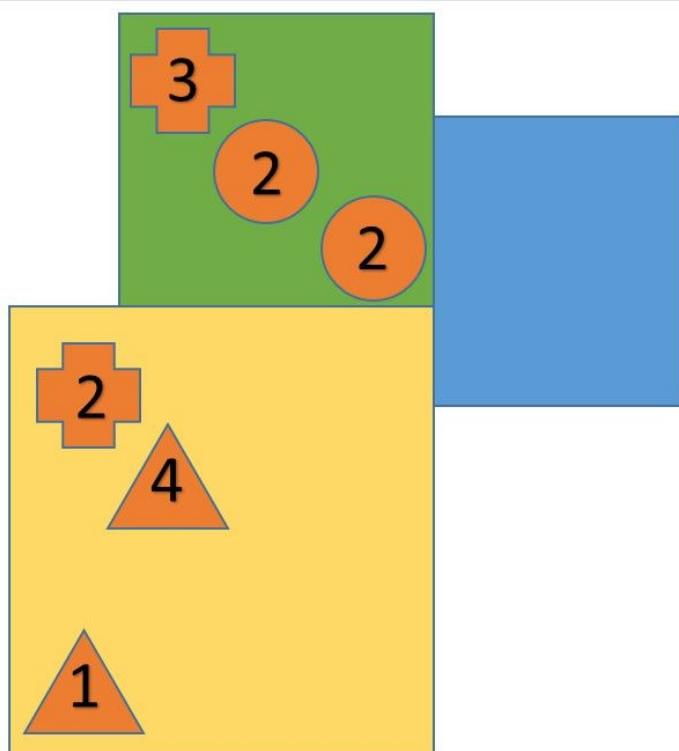


Table 2: No. of Incidents per City

City Names	Blockage Total	Flooding Total	Odour Total
City 1	3	4	
City 2	2		5
City 3			

Estamos unimos datos no por el valor de una variable, sino por la ubicación con respecto a otro objeto. de un modo similar a lo que hicimos con la selección. De hecho es un proceso

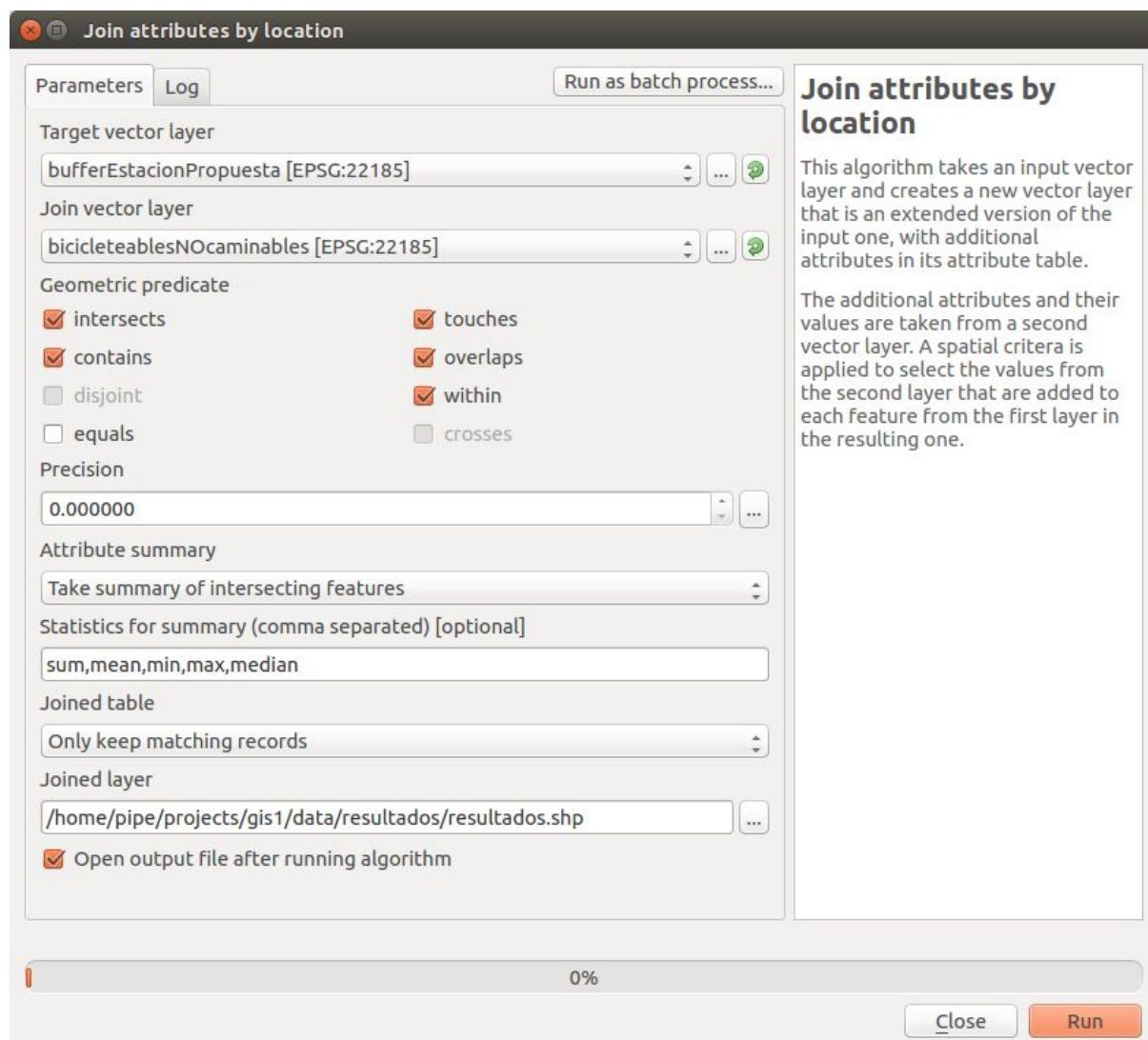
similar a seleccionar los radios que, por ejemplo tocan el área de influencia, y sobre eso realizar algún tipo de estadístico:

- Contar cuántos radios tocan dicha área
- Sumar la población que vive dentro de esos radios
- Obtener una mediana de los tiempos de viaje al centro
- Obtener un promedio de los costos de viaje

Una vez que sepamos esto para cualquier estación podemos asignar diversas estaciones, obtener el indicador que creamos mejor y asignar las estaciones allí donde este indicador tenga sus valores máximos.

Para realizar una unión espacial en QGIS hay que hacerlo desde

Vectorial > Herramienta de gestión de datos > Unir atributos por localización



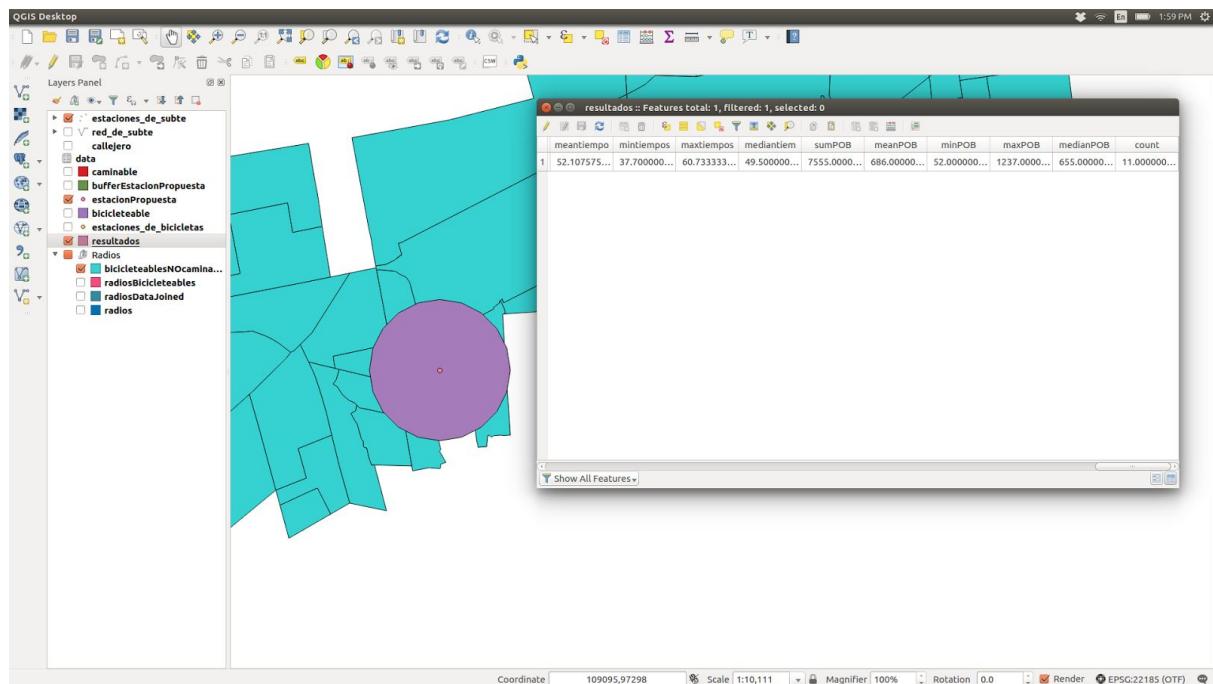
Nuestra capa objetivo es el área de influencia de la nueva estación. Es allí **hacia donde** queremos transferir datos. La capa **desde donde** queremos transferir son los radios

censales. Luego debemos explicitar qué tipo de vínculo espacial queremos. Que la capa objetivo intersecte, contenga, toque, se superponga, etc. Elegimos el criterio más amplio.

Luego escogemos los estadísticos que pueden ser sumatoria, media, mediana, mínimo, máximo, etc.

Por último elegimos que tipo de join o intersección hacer: si queremos mantener todos los datos de la capa objetivo (si tuviésemos un área de influencia que no se superpone con los quedaría en el resultado final pero sin datos), si queremos mantener solo los que se intercepten (si tuviésemos un área de influencia que no se superpone con ningún radio se eliminaría) y otras opciones. Escogemos mantener solo los que se intersectan ya que ese es nuestro interés.

El resultado final es un polígono con la misma geometría, pero con datos diferentes en su tabla de datos. Si accedemos a la tabla de datos de esa geometría podremos ver los estadísticos de esa estación.



11 - Posibles pasos a seguir

Este mismo proceso puede realizarse en variantes mucho más complejas, que superen las limitaciones enumeradas, pero requiere conocimientos más avanzados de técnicas de análisis espacial que procuraremos ir introduciendo en talleres subsiguientes.

- Eliminar los radios que ya están dentro del área de influencia de una estación de bicicletas
- Sumar estaciones de tren, premetro y metrobus.

- Utilizar distancias no lineales que sigan las redes viales por las cuales realmente se transita
- Estimar nuevos patrones de viaje y distancias recorridas para volver a estimar tiempos de viaje y utilizar esto como indicador de performance
- Automatizar el proceso para que realice esta búsqueda de forma automática utilizando herramientas de análisis espacial y SIG mediante librerías de un lenguaje de programación como Geopandas y Python
- Resolver el problema de topología complejo donde al colocar una nueva estación, el universo posible de localizaciones se cambia por la presencia de esta estación nueva