

Paso a paso clase 1

1. Bajar datos de indec (lista de área, sexo con filtro universo jefe universitario completo)

Para descargar los datos Censales del año 2010 vamos a descargar la lista por área según sexo filtrando a los jefes de hogar con universitario completo. La consulta en Redatam se realiza desde la siguiente pagina y el filtro a aplicar es el siguiente:

<http://200.51.91.245/argbin/RpWebEngine.exe/PortalAction?&MODE=MAIN&BASE=CPV2010B&MAIN=WebServerMain.inl>

VARIABLES A PROCESAR

Definir Parámetros

Título de la Tabla

Variable(s) a desplegar
Al menos un indicador NBI
Relación o parentesco con el jefe(a) del hogar
Sexo
Edades quinquenales
Edad en grandes grupos
En que país nació
Sabe leer y escribir
Condición de asistencia escolar
Nivel educativo que cursa o cursó

Corte de Área

Incluir Totales de ☒ Filas ☐ Columnas

Área Geográfica

Definición del Universo

PERSONA.P01 = 1 AND ((PERSONA.P10 = 1 AND PERSONA.P09 = 7) OR PERSONA.P09 = 8)

Tipo de salida

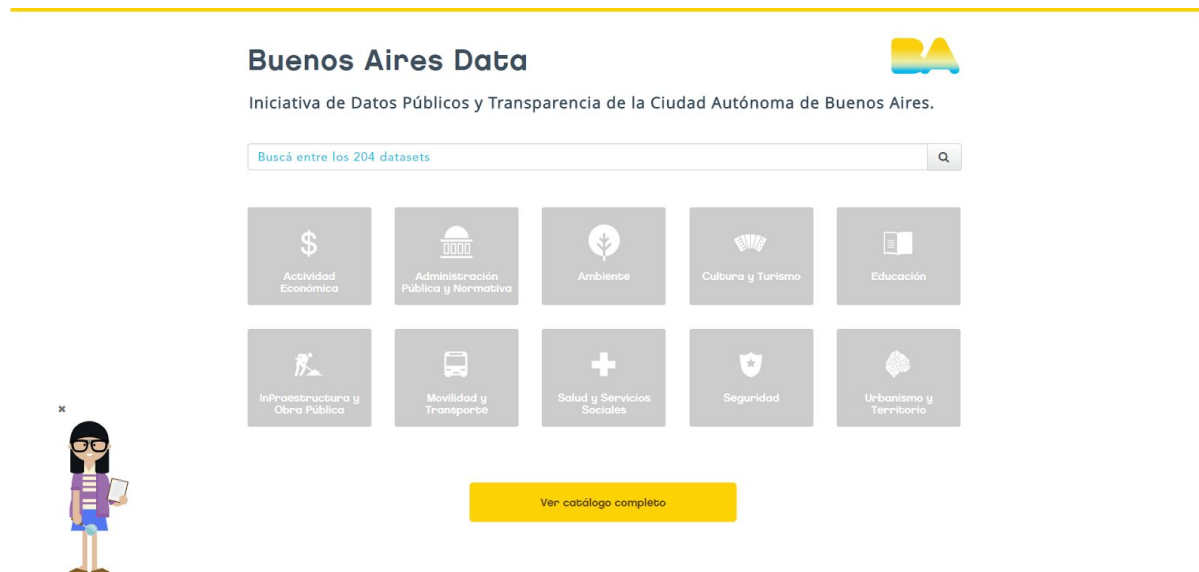
Filtro:

`PERSONA.P01 = 1 AND (PERSONA.P10 = 1 AND PERSONA.P09 = 7) OR PERSONA.P09 = 8`

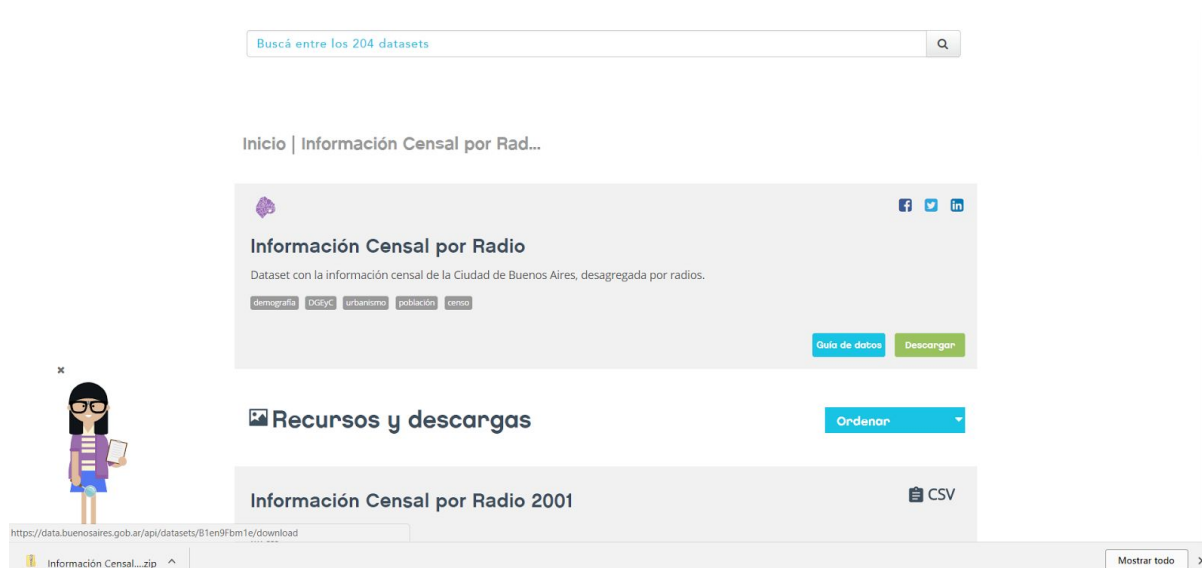
Descargar la tabla que resulta de nuestra consulta a la base de datos REDATAM.

2. Bajar cartografía radios censal CABA

Para descargar la cartografía vamos a utilizar como fuente el [portal de datos](#) de la Ciudad de Buenos Aires.



Una vez allí podemos buscar el dataset de [Información Censal por RADIO](#). 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 desarrollados 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)
- .xml - Almacena los metadatos del shapefile.
- .shx - Almacena el índice espacial de las entidades.

3. Hacer un join (hay que crear la variable que haga el join con radio fracción y comuna)

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 nos 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.

Elementos a estar atento antes de unir:

- Tener en claro el set de datos objetivo o destino, es decir, en que set de datos finalmente queremos que queden almacenadas las nuevas variables.
- Tener en claro que 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.

Primero, para realizar este procedimiento, es necesario trabajar en el archivo Excel que descargamos de redatam. Debido a que el código para cada radio difiere de los datos descargados de INDEC de la de CABA, debemos crear un campo que sea idéntico para luego poder unirlos en el QGIS.

Para esto, abrimos el archivo y por la forma en que se descarga el archivo lo mas recomendable es guardarlo como archivo .xlsx, luego en las opciones de Excel se debe modificar en la pestaña “formulas” el estilo de referencia que se trabaja, ya que al descargarlo las columnas y filas nos quedan ambos numéricos, sin la forma tradicional de letras para columnas y números para filas.

El campo código es el que nos va a permitir construir la variable CO_FRAC_RA para realizar la unión con el shape que tenemos de los radios, para esto debemos utilizar la formula de entero, izquierda y derecha, para poder sacar del código que recibimos el ID de cada una de los radios, las fracciones y las comunas.

A modo de ejemplo, la función que vamos a utilizar es:

- Radio: =ENTERO(DERECHA(A2;2))
- Fracción: =ENTERO(IZQUIERDA(DERECHA(A2;4);2))
- Comuna: =ENTERO(DERECHA(IZQUIERDA(A2;4);2))

Con estas funciones le pedimos que nos vaya desagregando el código y rearmar la variable que necesitamos. Ejemplo:

Código	Varón	Mujer	Total	RADIO	FRACCION	COMUNA
20010101	30	11	41	1	1	1

20010201	8	10	18	1	2	1
20010202	3	5	8	2	2	1
20010203	11	6	17	3	2	1

Aquí debemos tener en cuenta que esta es una de las formas de obtener el código de cada uno de estos, se pueden utilizar otras funciones o herramientas de Excel.

El siguiente paso es crear la variable CO_FRAC_RA para que coincida con el shape que tenemos.

La función que vamos a utilizar es la de CONCATENAR, que nos permite unir texto de diferentes celdas y agregarle el separador que deseamos, en este caso el guion bajo.

Seria la siguiente función: =CONCATENAR(G2;"_";F2;"_";E2)

Dando como resultado:

Código	Varon Univ Com	Mujer Univ Com	Total Univ Com	RADIO	FRACCION	COMUNA	CO_FRAC_RA
20010101	30	11	41	1	1	1	1_1_1
20010201	8	10	18	1	2	1	1_2_1
20010202	3	5	8	2	2	1	1_2_2
20010203	11	6	17	3	2	1	1_2_3

Variante en QGIS

Podemos utilizar la funcion substr(), que nos devuelve un substring o porción de un texto, si le damos el caracter de inicio y del fin.

substr(*a*,desde,cantidad de caracteres)

COMUNA = substr("codigo",2,3)

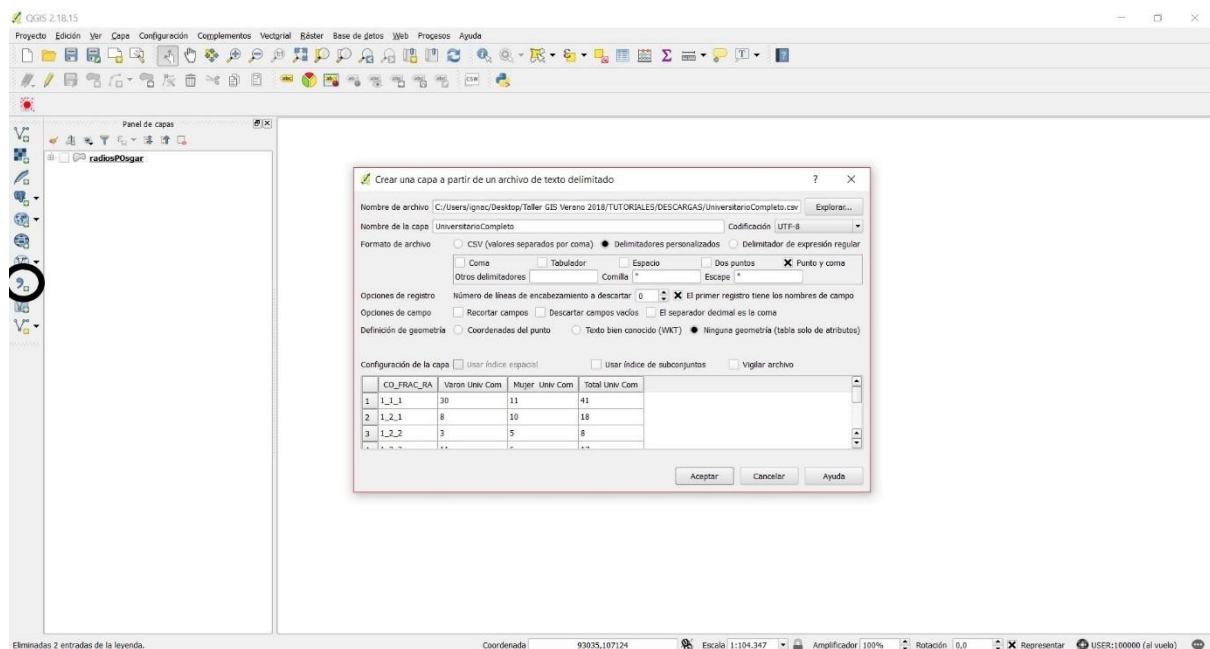
FRACCION = substr("codigo",5,2)

RADIO = substr("codigo",7,2)

CO_FRAC_RA = concat("comuna" ,'_',"fraccion",'_',"radio")

Una vez finalizado este cambio y creación del nuevo campo, podemos directamente quedarnos solo con las columnas que necesitamos, es decir, "Varon Univ Com", "Mujer Univ com". "Total Univ Com" y "CO_FRAC_RA". Las primeras 3 corresponden a los jefes de hogar según sexo con universitario completo, el total de ellos y luego el campo necesario para poder realizar la unión. Este archivo debemos guardarlo en .csv, es decir en un archivo que se encuentre separado por comas que luego podemos trabajarlo desde el QGIS.

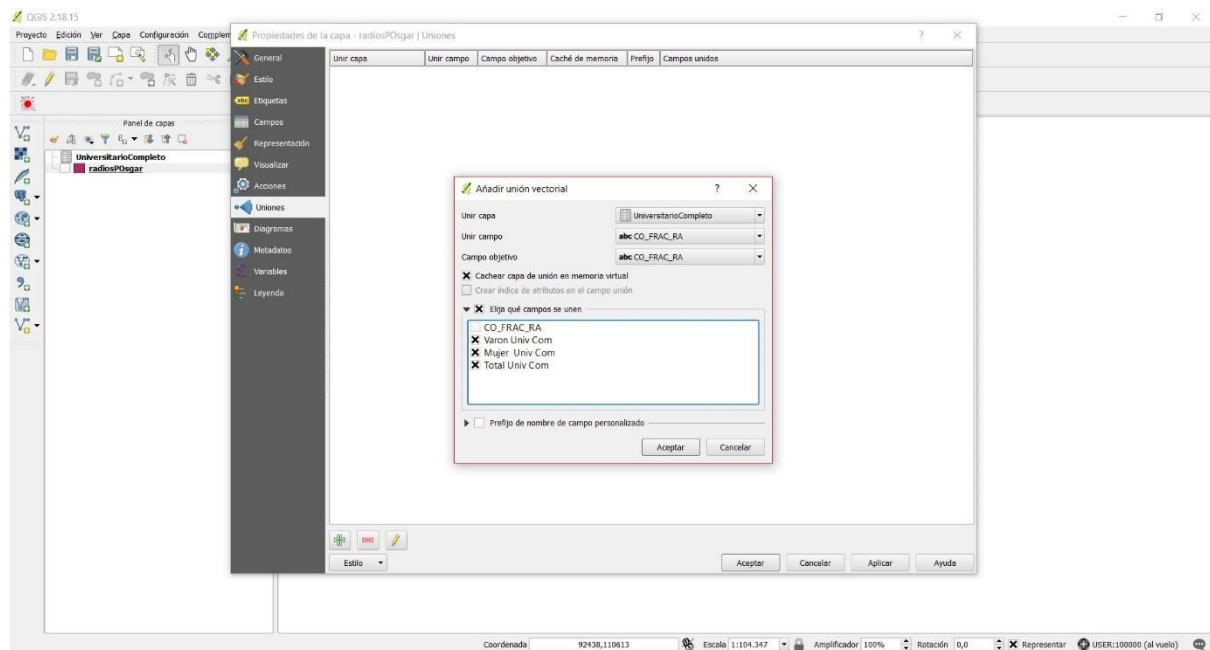
Para abrir en QGIS el csv, debemos desde la barra de herramientas de la izquierda clicar el icono con una coma, en la pantalla que se nos abre debemos seleccionar el archivo que desamos abrir y seleccionar las opciones que correspondan al archivo que deseamos abrir. En este caso, debemos usar delimitadores personalizados y elegir la opción de punto y coma, dejar tildada la opción de que el primer campo es el título de cada columna y en la definición de geometría en este caso debemos poner la opción de ninguna geometría ya que no se encuentra georreferenciado el dato que estamos importando a QGIS y deseamos unir al que si tenemos con propiedades geográficas que descargamos de la pagina de CABA.



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 (CO_FRAC_RA 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 sets 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 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).

Podemos ver que las almacena 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, sino que hay un mero vínculo o unión entre dos sets 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.

4. Hacer una coropleta con 5 niveles y los diferentes criterios (quintiles, quiebres jenks, desvío estándar, intervalos iguales de 0 20, 20 40 y así)

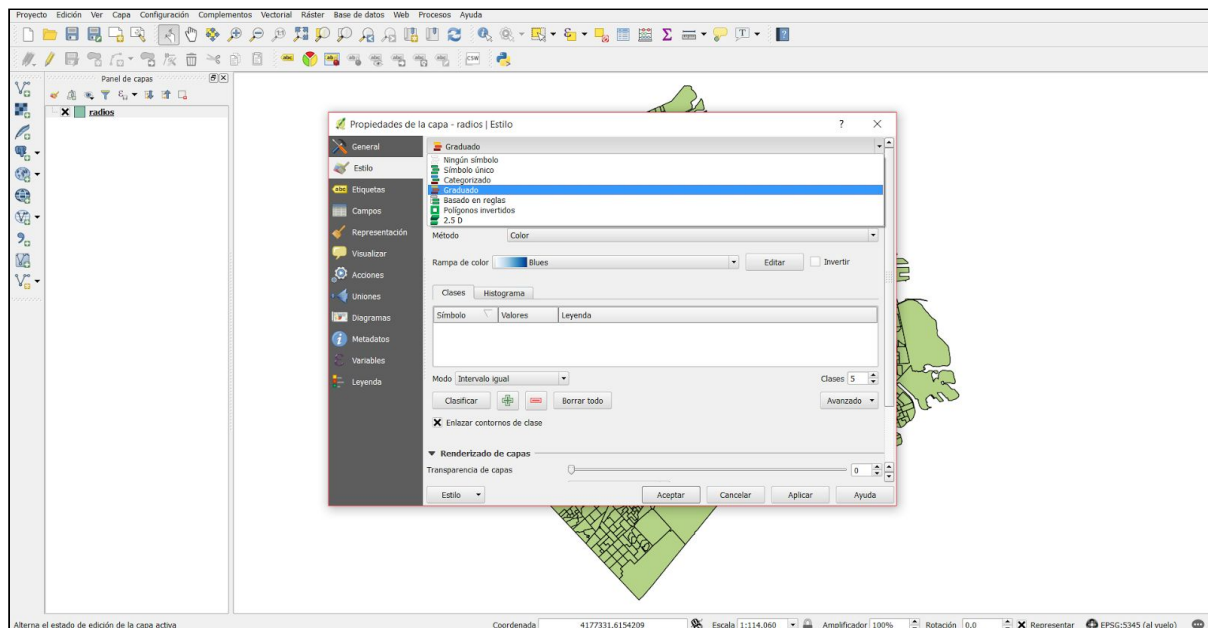
Una vez finalizada la unión, la capa nos va a quedar lista para poder realizar según diversos criterios una cloropleta.

Para esto debemos ir a las propiedades de la capa a la pestaña “Estilo” donde seleccionaremos la opción de “Graduado” ahí se nos abrirá las opciones para elegir los criterios para tener en cuenta para realizar la graduación de colores según el indicador que deseemos usar.

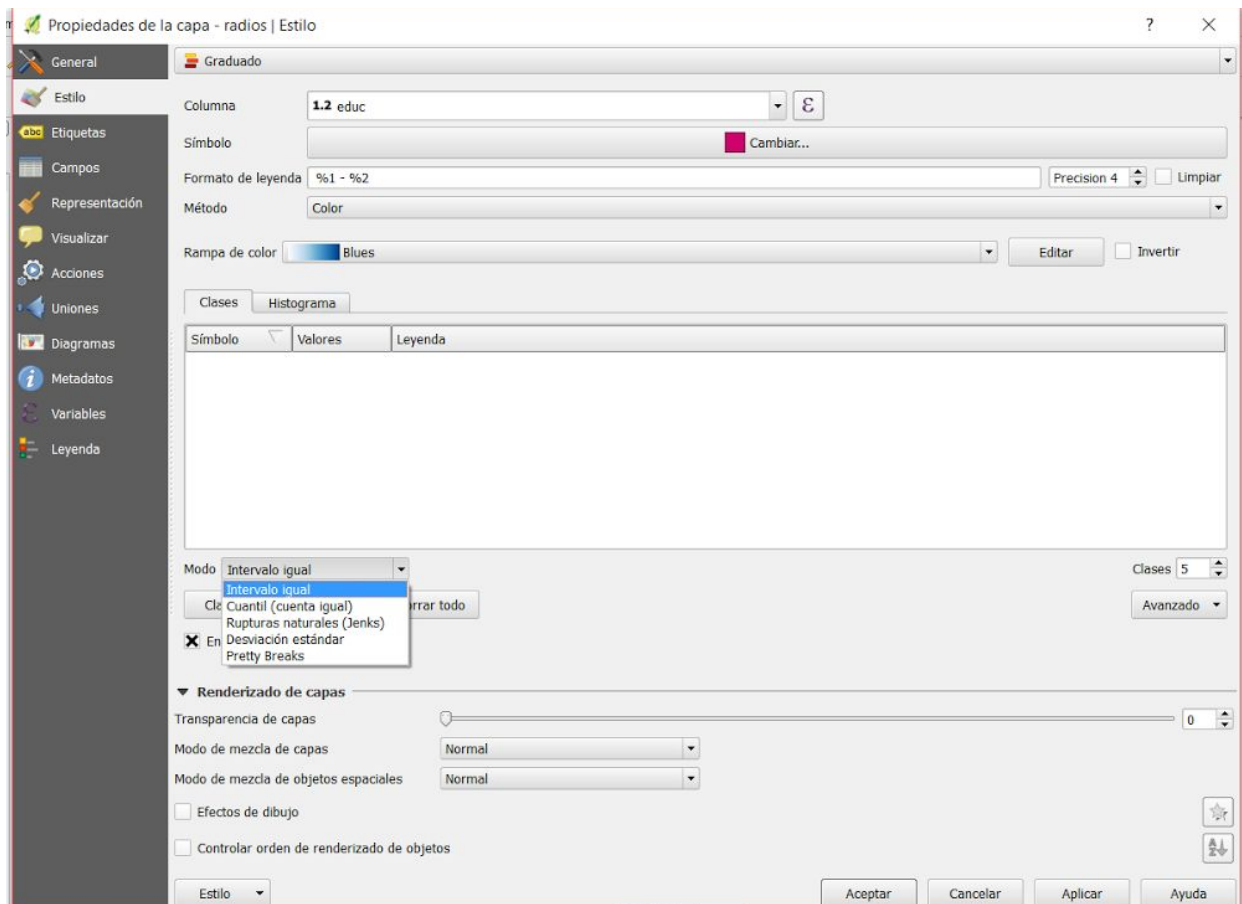
QGIS ofrece por defecto 5 criterios:

- 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.

- **Quintiles:** 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.



En este caso vamos a elegir la columna “educ” para realizar la cloropleta, seleccionamos la columna, luego la rampa de colores que realizara el graduado según los valores de dicha columna y en la opción de “MODO” elegimos el criterio por el cual realizar esta acción. Clickeamos clasificar y aplicamos los cambios realizados.



Repetimos esta acción con los siguientes modos: Quintiles, Rupturas naturales (Jenks), Desviación Estándar y por intervalos iguales (en este último somos nosotros los que decidimos de cuanto será el intervalo de cada categoría).

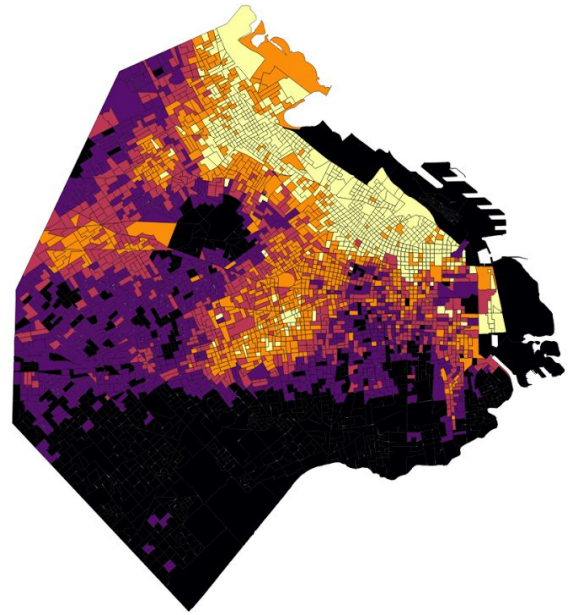
Elegimos la cantidad de grupos, la paleta de colores (recordando que el color es un portador de sentidos) y fundamentalmente, un criterio para colorear.

Qué color darles implica toda una estrategia de visualización, no es una decisión neutra.

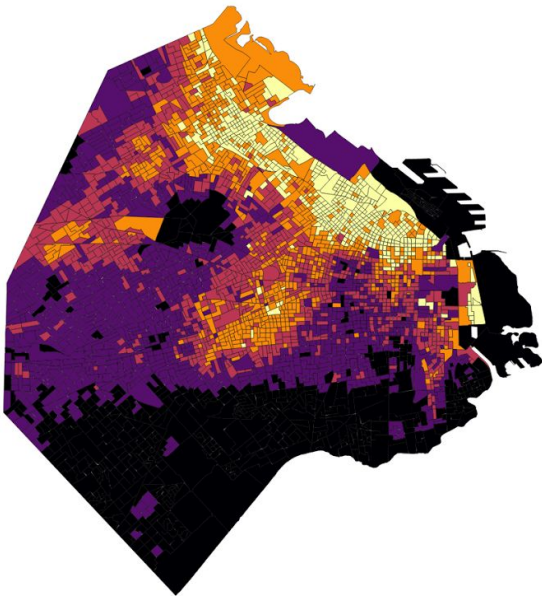
El color construye sentido.



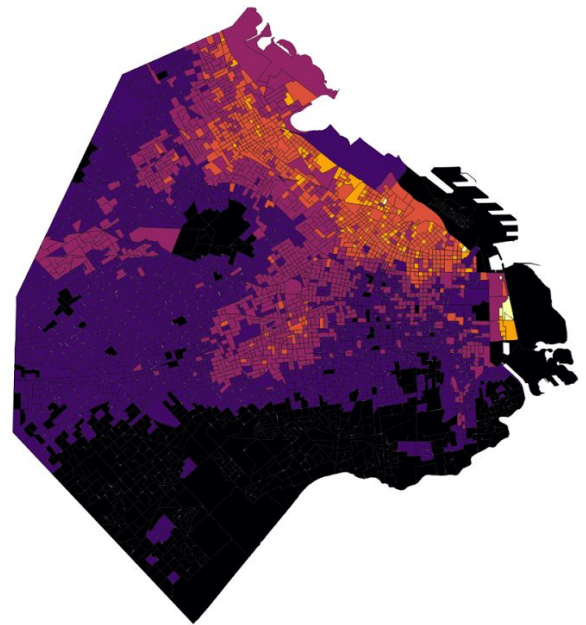
Intervalo igual



Quintiles



Quiebres Naturales Jenks



Desvíos estándar