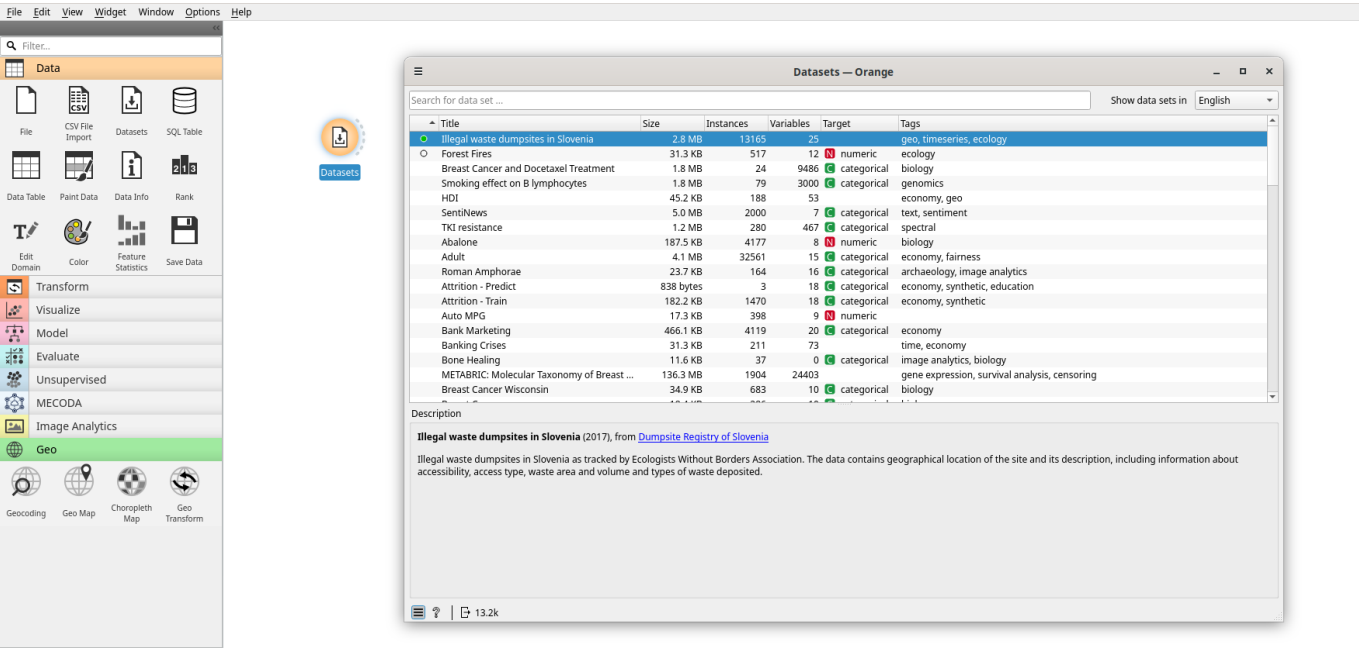


# Ejemplo genérico de funcionamiento de Orange

## (tiempo estimado: 30min)

Usaremos los widgets propios de Orange para familiarizarnos con la aplicación.

Vamos a descargar uno de los datasets existentes en Orange para realizar un análisis sencillo. Para ello, seleccionamos el widget "Datasets" de la sección "Data", de la columna izquierda donde se encuentran todos los widgets disponibles.



Hacemos doble click sobre el widget para que se abra la ventana que nos permitirá seleccionar un dataset de los que tiene cargados Orange para practicar. En este caso elegimos el conjunto de datos "Illegal waste dumpsites in Slovenia", haciendo doble click sobre él para que se descargue a nuestra aplicación. Aparecerá un círculo verde junto al nombre del dataset cuando se haya descargado.

Vamos a explorar los datos que hemos obtenido. Para ello usamos el widget *Data Table*, conectándolo con nuestro widget *Datasets*. Vemos que la conexión entre los dos widgets es una línea continua, que indica que los datos se han transmitido correctamente. Haciendo doble click en *Data Table* vemos los datos que tenemos:

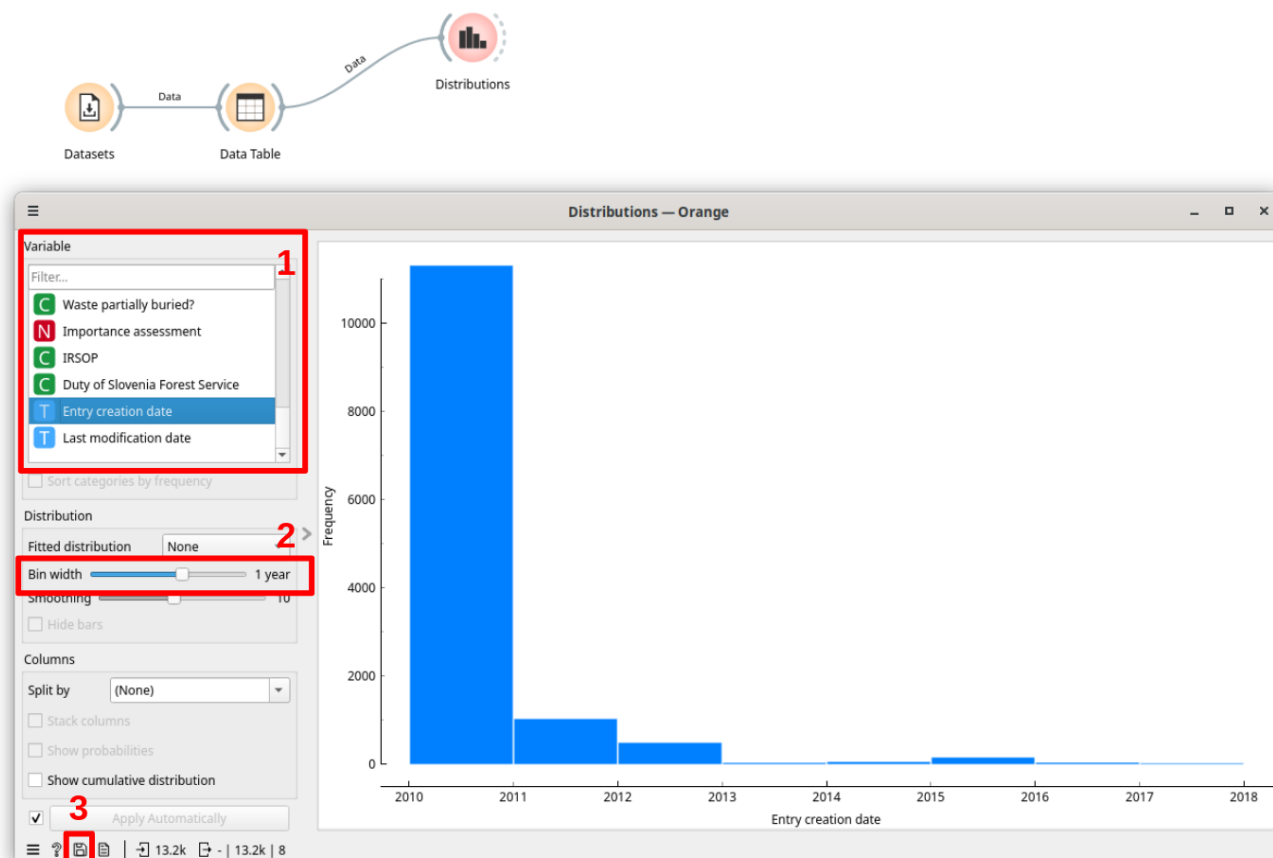


Data Table — Orange							
Info	Geopedia URL	ID	Name	/quantity of hazar	Municipality	d registry propert	
13165 instances 23 features (24.0 % missing data) No target variable. 6 meta attributes (18.7 % missing data)	1 http://www.geopedia.si/?params=T1199_L6357_F4	4 ?		cca. 10 računal...	Slovenska Bist...	618-1004/3	
	2 http://www.geopedia.si/?params=T1199_L6357_F5	5 ?		gradbeni mate...	Hrpelje-Kozina	1415-2175/1	
	3 http://www.geopedia.si/?params=T1199_L6357_F7	7 ?		?	Celje	2171-416/4	
	4 http://www.geopedia.si/?params=T1199_L6357_F9	9 ?		Gume; skrinja,...	Ivančna Gorica	1836-1009	
	5 http://www.geopedia.si/?params=T1199_L6357_F10	10 ?		Odvržene star...	Ajdovščina	1251-824	
	6 http://www.geopedia.si/?params=T1199_L6357_F13	13 ?		?	Koper	747-1835/4	
	7 http://www.geopedia.si/?params=T1199_L6357_F14	14 ?		?	Šentjernej	1710-592	
	8 http://www.geopedia.si/?params=T1199_L6357_F15	15 ?		dostop je z avt...	Koper	637-6005/2	
	9 http://www.geopedia.si/?params=T1199_L6357_F16	16 ?		?	Koper	637-6003	
	10 http://www.geopedia.si/?params=T1199_L6357_F17	17 ?		?	Koper	1105-1201	
	11 http://www.geopedia.si/?params=T1199_L6357_F18	18 ?		?	Koper	1955-740/1	
	12 http://www.geopedia.si/?params=T1199_L6357_F19	19 ?		?	Koper	2532-3168/1	
	13 http://www.geopedia.si/?params=T1199_L6357_F20	20 ?		razpršeni kom...	Koper	637-3882/1	
	14 http://www.geopedia.si/?params=T1199_L6357_F855	855 ?		?	Semič	2176-4170	
	15 http://www.geopedia.si/?params=T1199_L6357_F52	52 ?		?	Slovenska Bist...	2501-1495/2	
	16 http://www.geopedia.si/?params=T1199_L6357_F24	24 ?		?	Dravograd	524-40/1	
	17 http://www.geopedia.si/?params=T1199_L6357_F25	25 ?		?	Dravograd	524-936	
	18 http://www.geopedia.si/?params=T1199_L6357_F26	26 ?		?	Dravograd	1740-1196/64	
	19 http://www.geopedia.si/?params=T1199_L6357_F27	27 ?		sežiganje neva...	Kranj	1994-1039/1	
	20 http://www.geopedia.si/?params=T1199_L6357_F28	28 ?		odlaganje v ka...	Kranj	318-744/1	
	21 http://www.geopedia.si/?params=T1199_L6357_F29	29 ?		odlaganje v ka...	Kranj	318-744/1	
	22 http://www.geopedia.si/?params=T1199_L6357_F31	31 ?		10 letovstari	Kočevje	2702-237	

Vemos que se trata de 13165 líneas (instances). Exploramos las columnas que tenemos para hacernos una idea de los datos. Ya podemos analizarlos.

¿Cuántos incidentes se han registrado cada año?

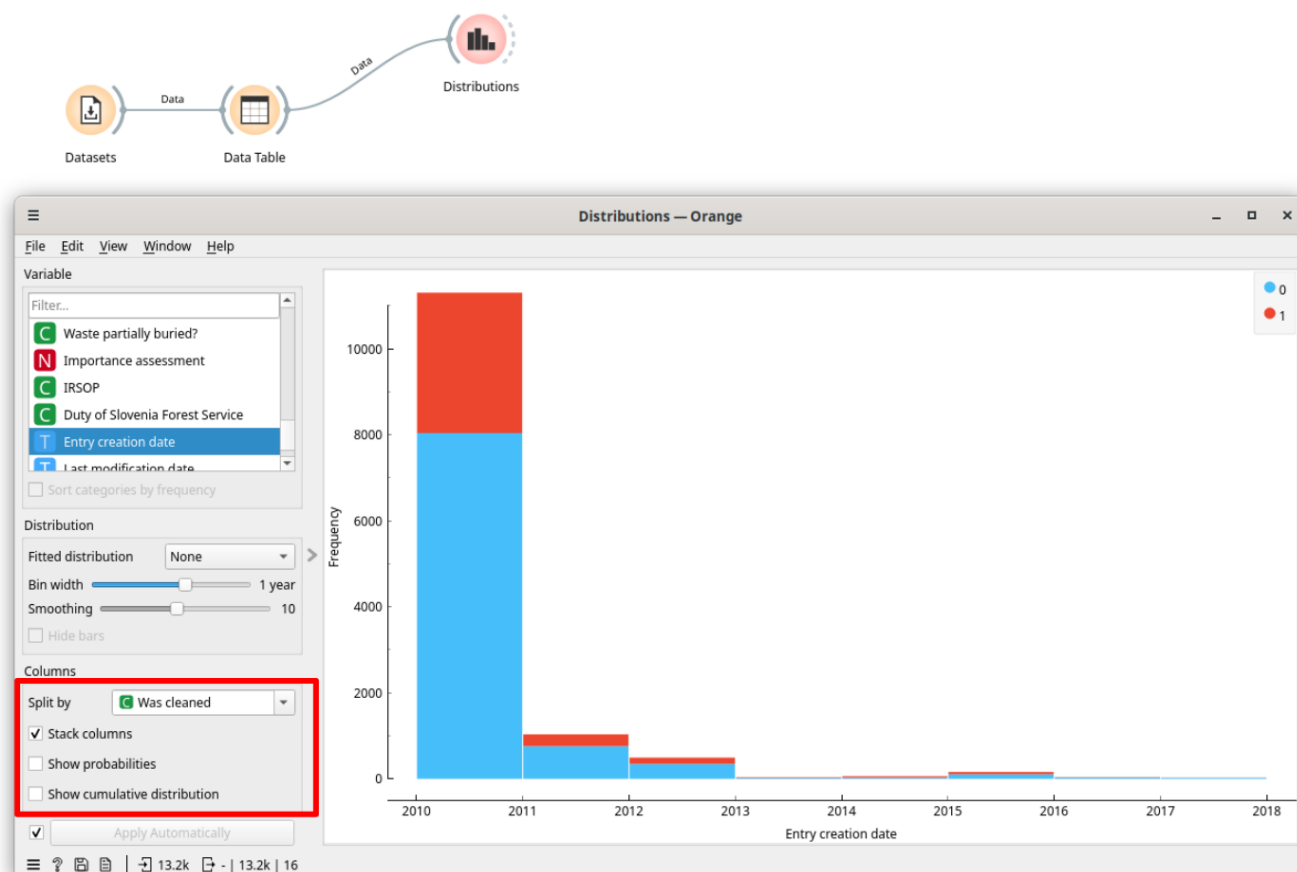
Existe un campo llamado "Entry creation date" que nos da la fecha en que se incorporó el registro. Lo tomaremos como la fecha de la infracción. Queremos ver cuántos registros hay cada año. Conectaremos nuestra tabla con el widget que nos permite visualizar distribuciones, llamado *Distributions* y hacemos doble click sobre él para ver qué opciones nos da:



Lo primero es seleccionar la variable (la columna) que queremos representar. De la lista de nombres de columnas marcada con el número 1, seleccionamos *Entry creation date*. Veremos una representación de los datos en fechas. El campo con el número 2, llamado *Bin width*. Aquí podremos decidir que rango temporal queremos para nuestra agrupación. Elegiremos *1 year*, para sacar la distribución año a año.

Ahora vemos un gráfico de columnas con el número de registros que tenemos cada año, y llama la atención el número de eventos que tenemos de 2010. Pensemos que queremos descargarnos este primer gráfico. Nos vamos a la parte inferior izquierda de la ventana y veremos un pequeño disquete (identificado con el número 3). Pinchando en él nos podremos descargar la imagen del gráfico.

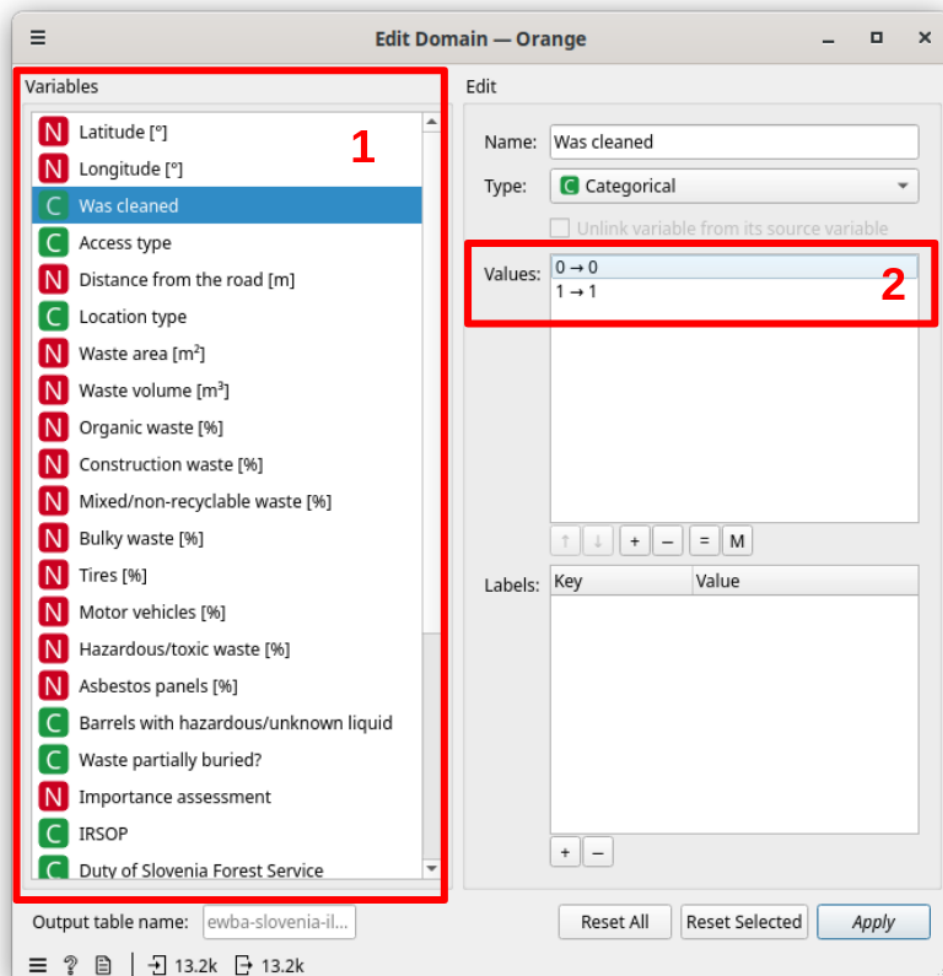
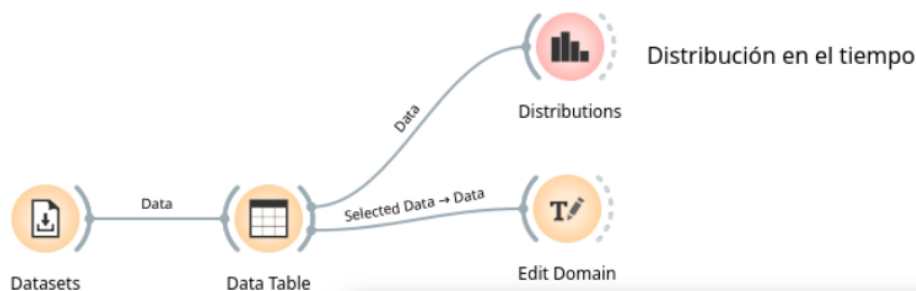
Pero podemos hacer algo más con nuestro gráfico de distribución. Podemos querer dividir los registros de cada columna por otra variable:



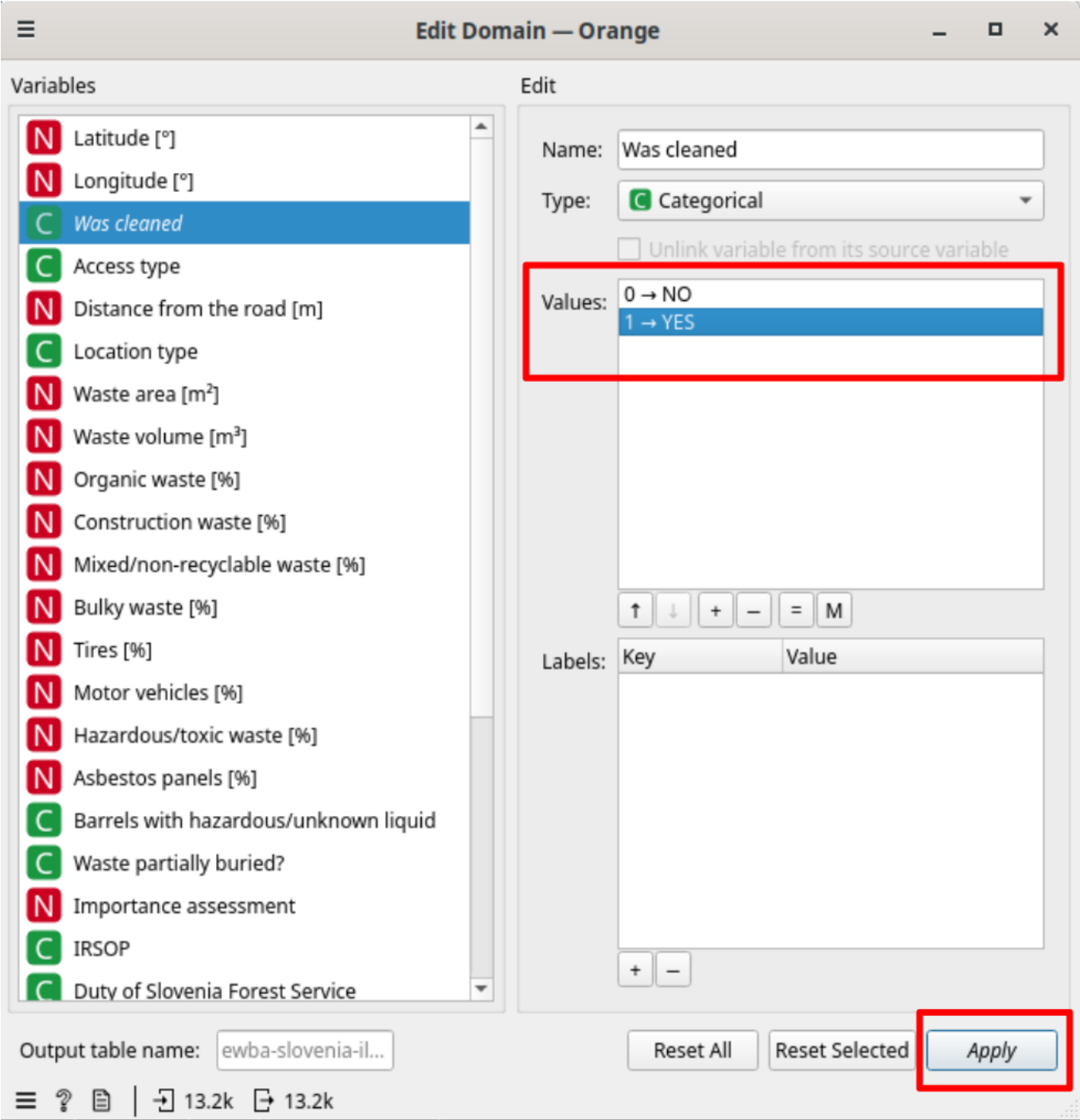
En este caso tenemos una columna llamada *Was cleaned* que nos indica si el vertido ha sido limpiado o no. Vamos a usarlo para ver cuántos de los registros de cada año se han limpiado. Seleccionamos el nombre de la columna en el campo *Split by*. Nos mostrará dos columnas por cada año, uno para los registros que tiene valor 0 en *Was cleaned* y otra con lo que tienen valor 1. De forma general, se considera que 0 equivale a Falso y 1 a Verdadero. Esta vista de dos columnas por año puede ser un poco confusa para analizar. Si preferimos que ambas aparezcan en columnas apiladas, seleccionaremos *Stack columns*, debajo del desplegable donde hemos elegido la columna. Nos quedará un gráfico muy ilustrativo.

## Cambiar valores de una columna

Digamos que el hecho de que la columna *Was cleaned* tenga valores 0 y 1 no es muy claro y queremos que en lugar de eso indique False/True, para que nos ayude a interpretar los resultados. Podemos hacerlo usando el widget *Edit Domain*:



Ahí seleccionamos la columna que queremos tratar, en el apartado marcado con un 1. Vemos los valores que puede tomar la columna *Was cleaned*, que puede ser 0 o 1. Si hacemos doble click sobre 0 podemos darle el valor que queremos que tenga, por ejemplo, "NO" para 0, y "YES" para 1. Después de escribir el valor que queremos que tome, pulsamos Intro para que se haga el cambio. Veremos algo así:

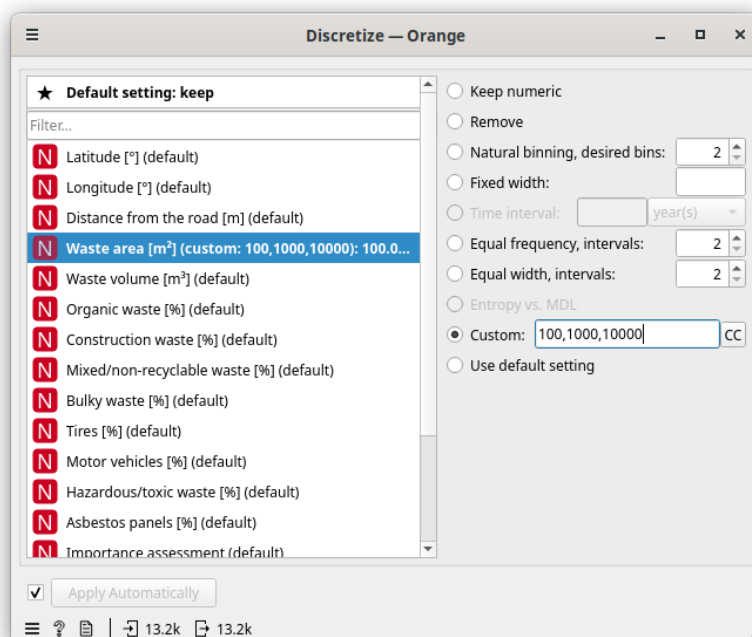


Ya podemos utilizar esta salida para representar correctamente estos datos.

## Crear una columna a partir de otros datos

Volvamos al dataset original. Hay una columna llamada *Waste area (m2)* muy interesante para nuestro análisis, pero nos gustaría obtener a partir de ella los registros que se corresponden a áreas pequeñas, grandes y muy grandes.

Vamos a tratar esa columna numérica. Usamos el widget *Discretize* que se encuentra en la sección *Transform*. Abrimos ese widget, elegimos la columna que queremos convertir y luego la forma de transformación. En este caso elegiremos directamente los rangos que queremos para nuestra columna (opción Custom) y estos serán: 100m2, 1000m2 y 10000m2:



¿Cómo queda nuestra tabla después de hacer esto?

Info

13165 instances  
23 features (24.0 % missing data)  
No target variable.  
6 meta attributes (18.7 % missing data)

Variables

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

Selection

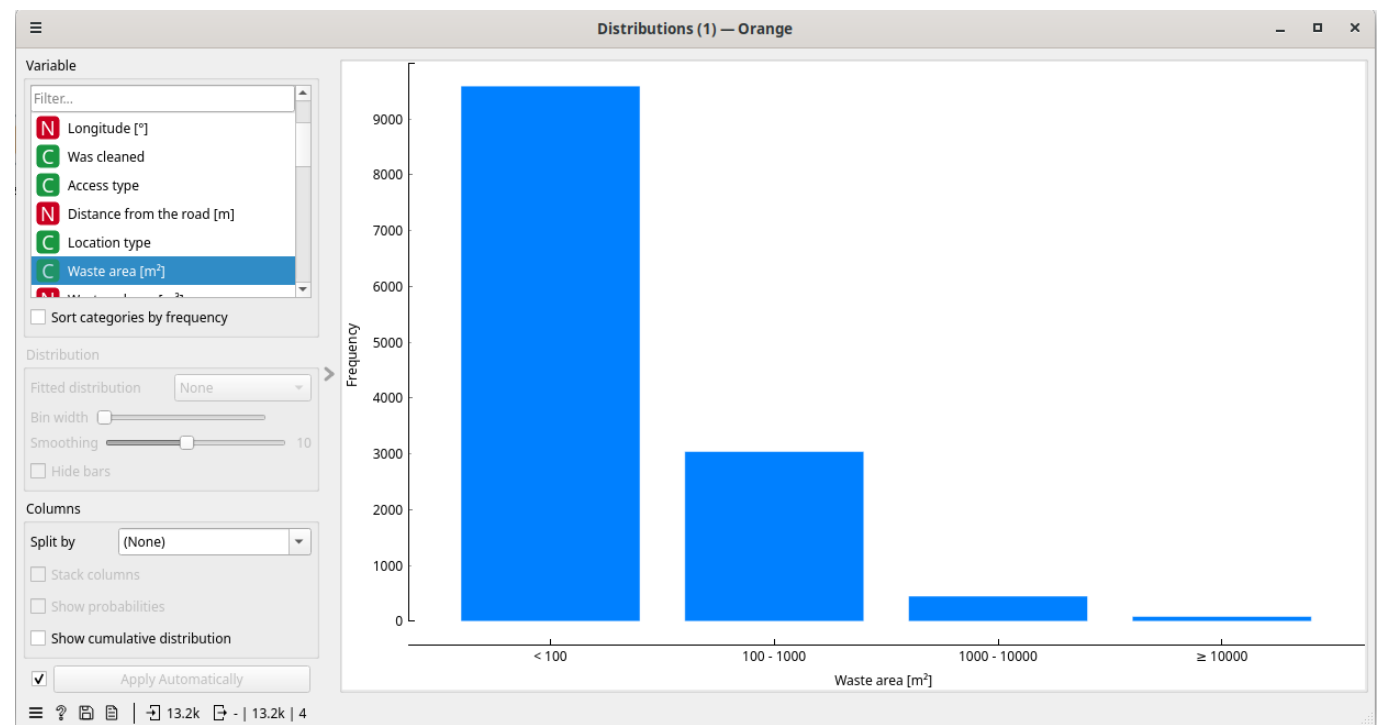
☒ Select full rows

Restore Original Order

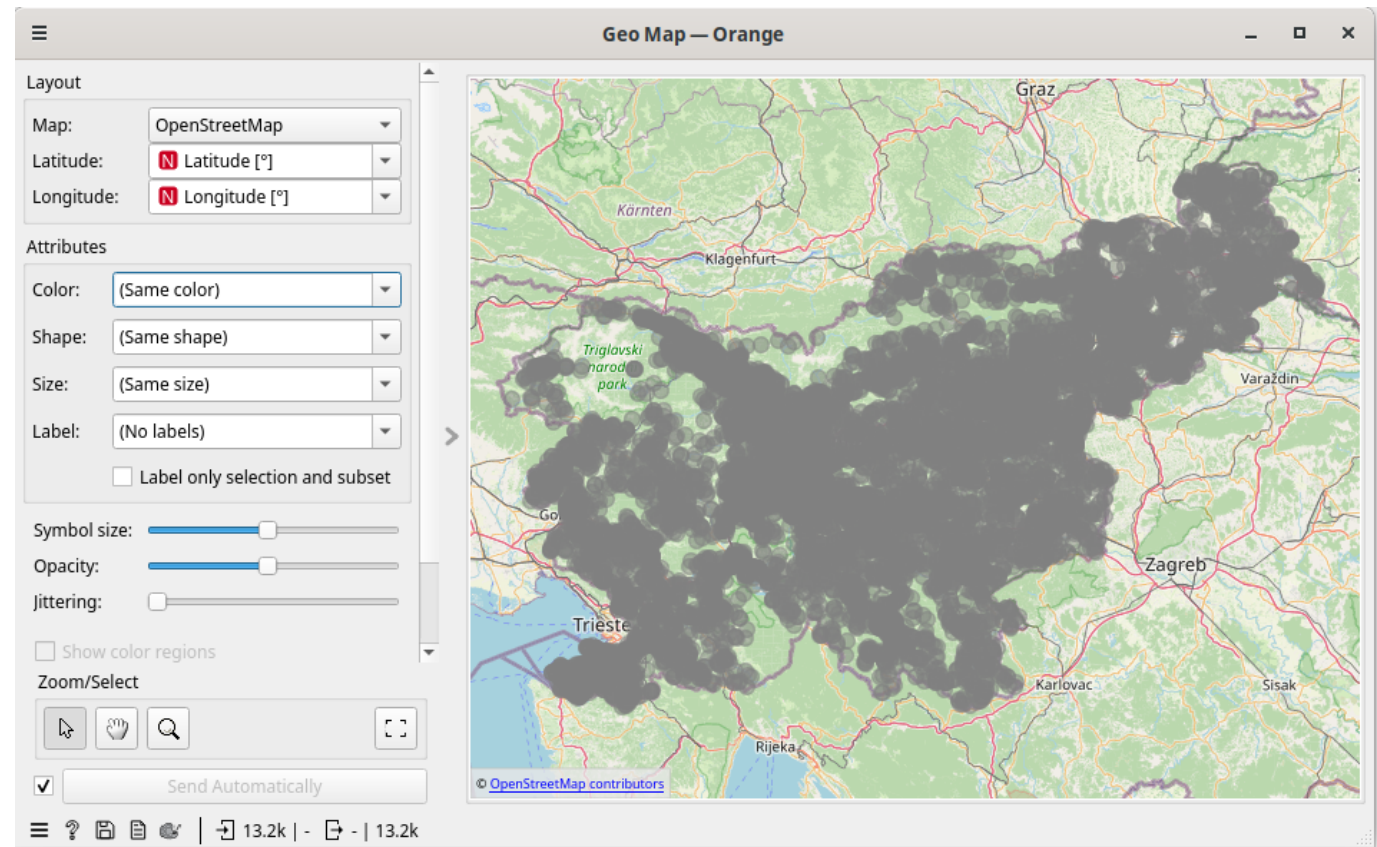
☒ Send Automatically

	Waste area [m²]	Waste volume [m³]	Organic waste [%]	Construction waste [%]	Non-recyclable waste [%]	Bulky waste [%]	Tires [%]	Motor vehicles [%]
1	< 100	10	?	30	10	10	?	?
2	100 - 1000	500	60	40	?	?	?	?
3	< 100	50	?	80	20	?	?	?
4	< 100	10	?	20	?	50	?	?
5	< 100	10	?	?	?	100	?	?
6	100 - 1000	2	?	?	80	20	?	?
7	100 - 1000	1500	50	30	?	20	?	?
8	< 100	2	?	100	?	?	?	?
9	< 100	1	?	50	?	50	?	?
10	< 100	2	?	20	?	80	?	?
11	< 100	2	?	?	100	?	?	?
12	< 100	2	?	?	100	?	?	?
13	< 100	1	?	?	100	?	?	?
14	100 - 1000	250	30	30	20	20	?	?
15	< 100	10	?	?	100	?	?	?
16	100 - 1000	50	?	50	?	50	?	?
17	< 100	25	?	50	50	?	?	?
18	< 100	10	50	50	0	0	?	?
19	100 - 1000	100	?	100	?	?	?	?
20	100 - 1000	25	?	70	10	20	?	?
21	< 100	25	10	50	10	30	?	?
22	< 100	10	?	?	?	100	?	?

Vemos que ahora tenemos una columna en la que aparece el rango del área de basura, según los parámetros que le dimos: menos de 100m2, entre 100 y 1000m2, entre 1000-10000 y más de 10000. Ya podemos ver el número de registros de cada rango. Sólo tenemos que conectar nuestra nueva tabla con el widget *Distribution*, como hemos hecho antes, para ver el número de registros en cada franja.







Modificando los atributos, podremos crear un mapa que comunique información. Podemos elegir el color de los puntos según una columna, como por ejemplo *Was cleaned*, *Waste partially buried?* o cualquiera de tipo categoría. Podemos darle tamaño (size) en función de una columna numérica, como *Waste area [m2]* o *Waste volume [m3]*. Podemos guardar el mapa de la misma forma que guardamos un gráfico, o hacer una selección de una zona para quedarnos con esos registros y poder analizarlos por separado.

El resultado de todo esto es un workflow que podemos guardar, en formato propio de .ows y reproducir en cualquier momento.

