Chapter 6

Building Maps with Carto (formerly CartoDB)

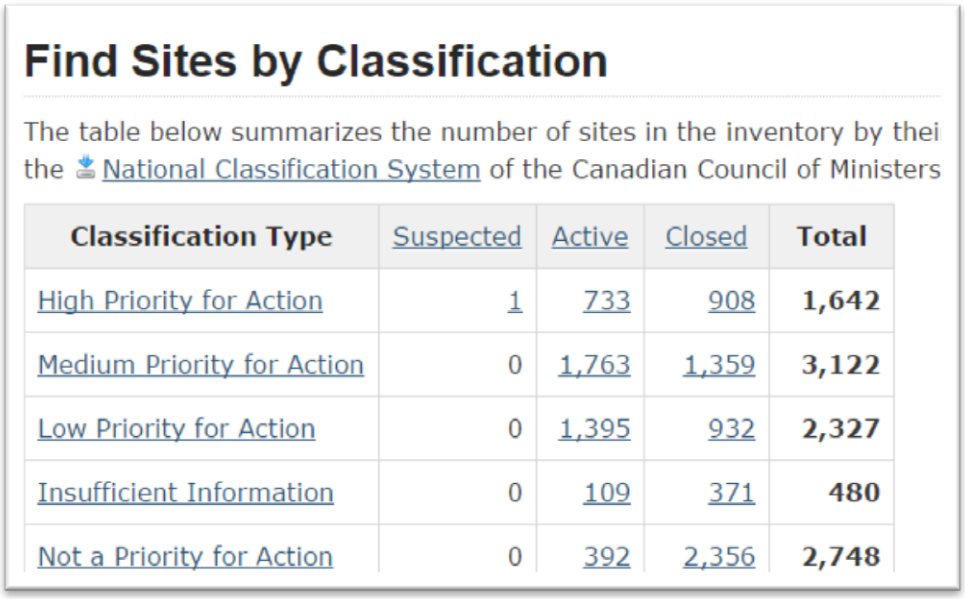
Skills you will learn:

1. Upload and create a simple map based on point features.
2. Create a choropleth map using uploaded tabular data and polygon features.

Summary: In this tutorial, we'll learn how to construct two types of maps; one with points, the other with polygons. You'll need a Carto account, which you can set up by going to Carto.com. The free account should be more than sufficient. For our simple points map, we'll be using a table containing the list of federal contaminated sites in Canada. For the choropleth map, we will use obesity prevalence data form the U.S. Centres for Disease Control and display the prevalence by county.

Making a map from simple points

You can find the contaminated sites data at this link: http://www.tbs-sct.gc.ca/fcsi-rscf/classification-eng.aspx A page like this will open.

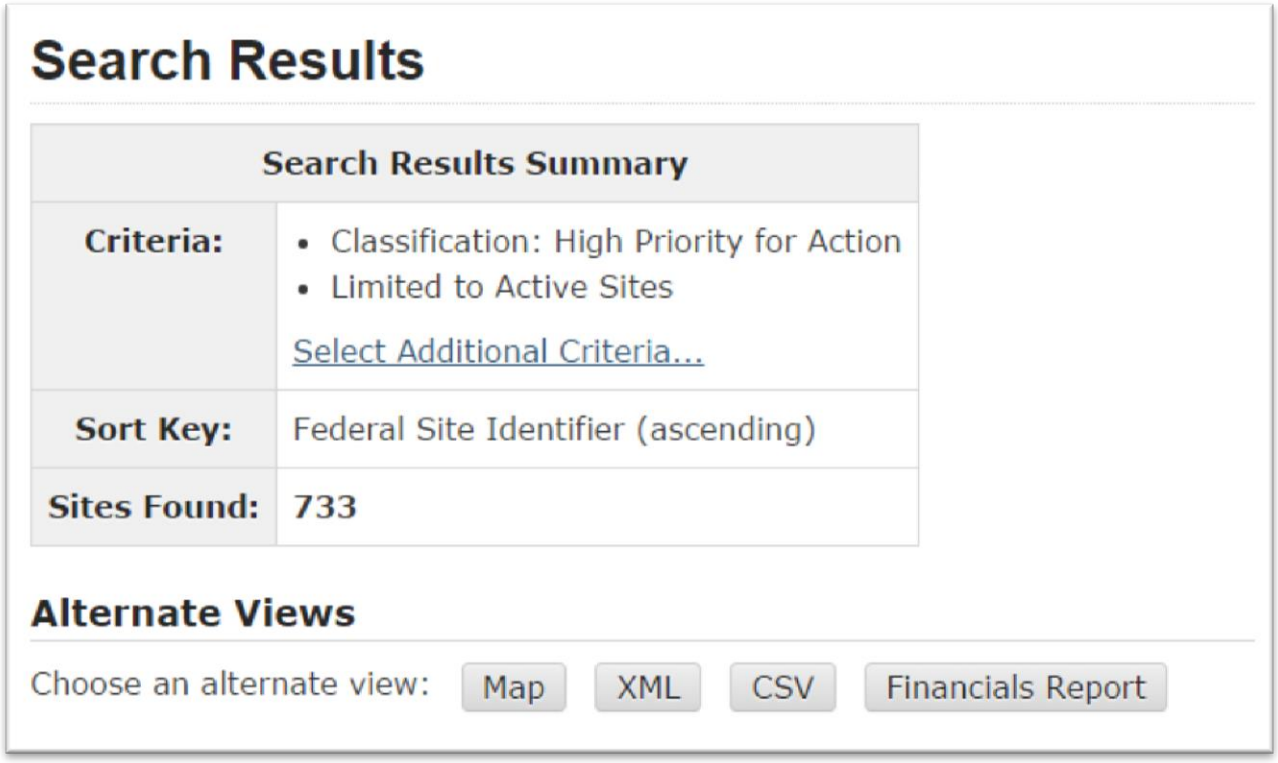


Find Sites by Classification

The table below summarizes the number of sites in the inventory by thei
the National Classification System of the Canadian Council of Ministers

Classification Type	Suspected	Active	Closed	Total
High Priority for Action	1	733	908	**1,642**
Medium Priority for Action	0	1,763	1,359	**3,122**
Low Priority for Action	0	1,395	932	**2,327**
Insufficient Information	0	109	371	**480**
Not a Priority for Action	0	392	2,356	**2,748**

The file we'll download "Active" sites in the first category, "High Priority for Action." Click the number (733 in the illustration above, but subject to change). That will bring up this page.



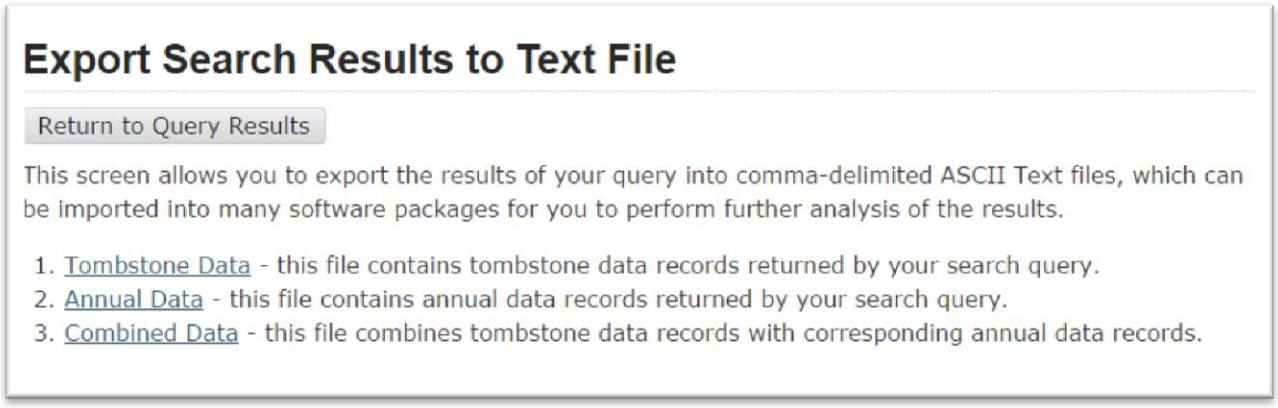
Search Results

Search Results Summary	
Criteria:	• Classification: High Priority for Action • Limited to Active Sites Select Additional Criteria...
Sort Key:	Federal Site Identifier (ascending)
Sites Found:	**733**

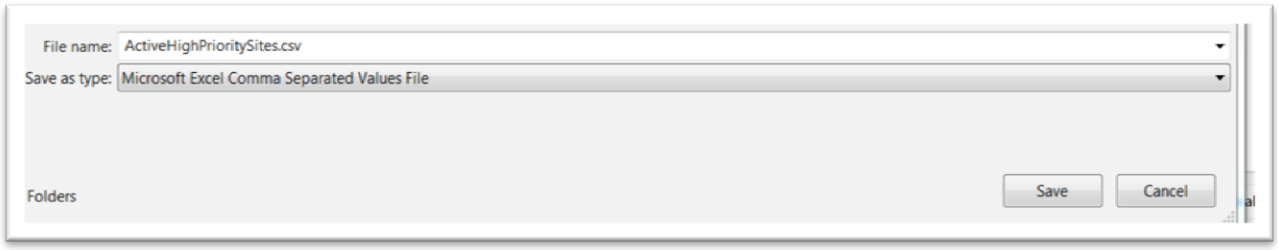
Alternate Views

Choose an alternate view: Map XML CSV Financials Report

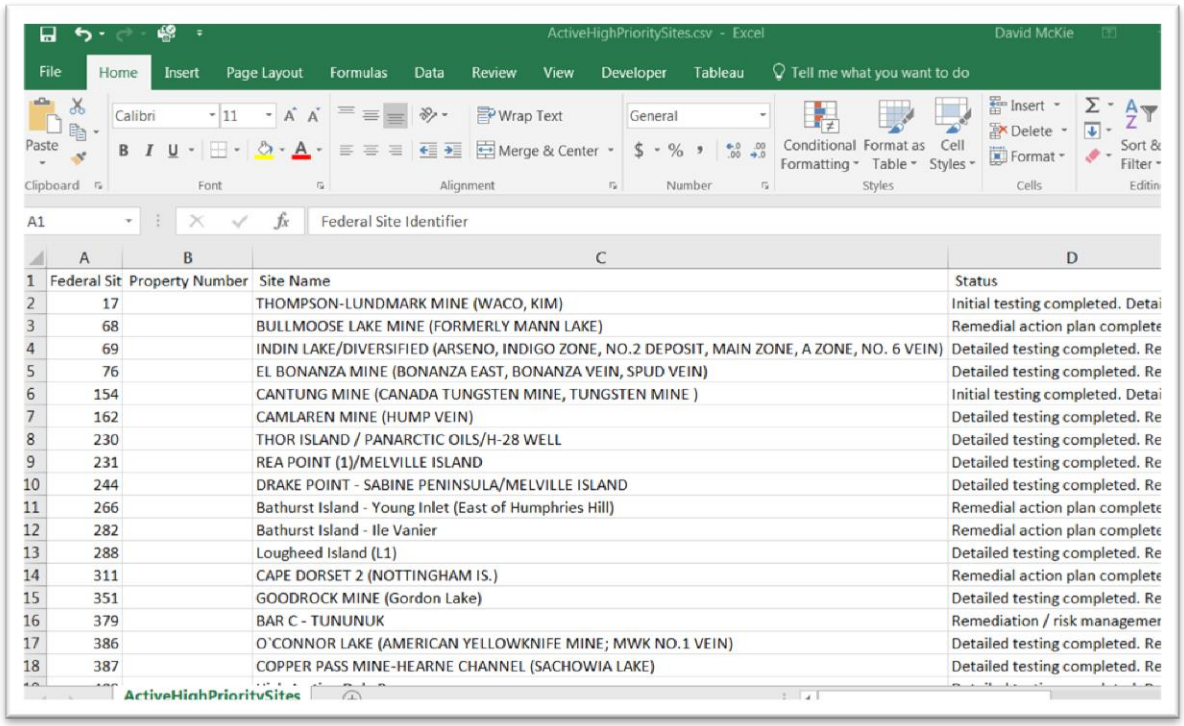
Select CSV.



Select "Tombstone Data and save the file to the folder on your hard drive that you'll be using for this tutorial.

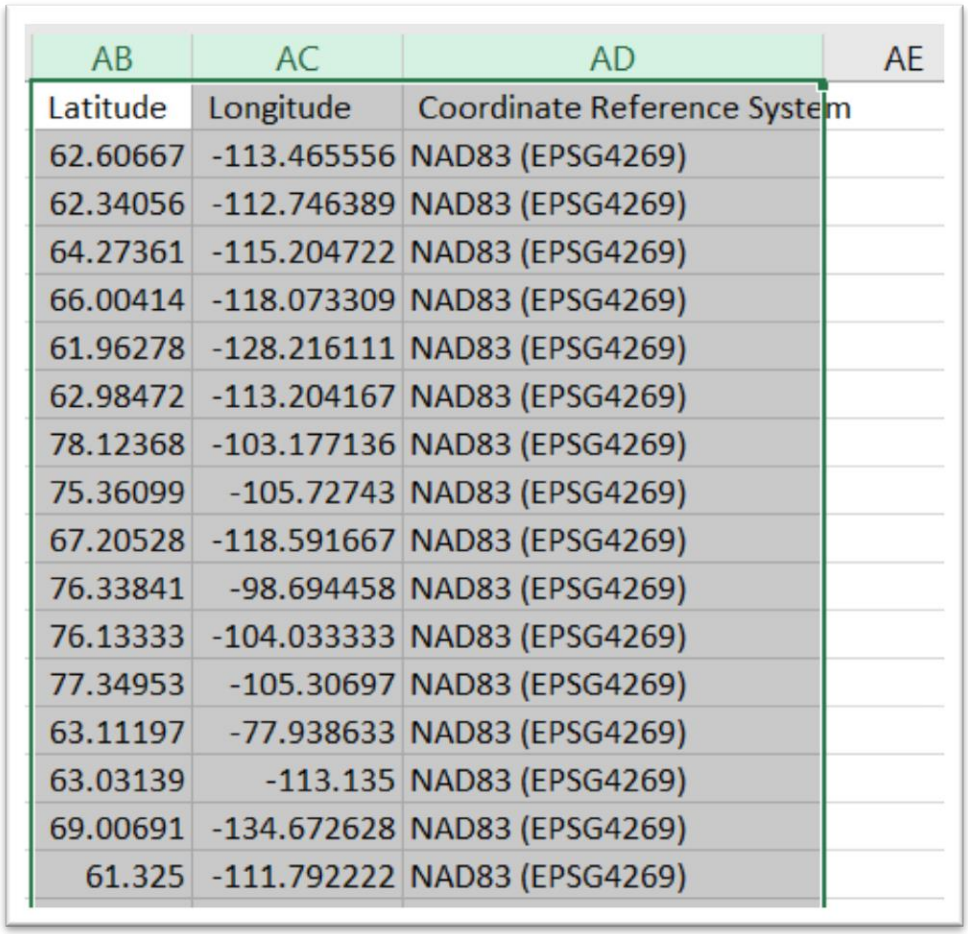


The file, as seen here in Microsoft Excel, contains 30 columns and 733 rows, with each one representing a site.



	A	B	C	D
1	Federal Sit	Property Number	Site Name	Status
2	17		THOMPSON-LUNDMARK MINE (WACO, KIM)	Initial testing completed. Detai
3	68		BULLMOOSE LAKE MINE (FORMERLY MANN LAKE)	Remedial action plan complete
4	69		INDIN LAKE/DIVERSIFIED (ARSENO, INDIGO ZONE, NO.2 DEPOSIT, MAIN ZONE, A ZONE, NO. 6 VEIN)	Detailed testing completed. Re
5	76		EL BONANZA MINE (BONANZA EAST, BONANZA VEIN, SPUD VEIN)	Detailed testing completed. Re
6	154		CANTUNG MINE (CANADA TUNGSTEN MINE, TUNGSTEN MINE)	Initial testing completed. Detai
7	162		CAMLAREN MINE (HUMP VEIN)	Detailed testing completed. Re
8	230		THOR ISLAND / PANARCTIC OILS/H-28 WELL	Detailed testing completed. Re
9	231		REA POINT (1)/MELVILLE ISLAND	Detailed testing completed. Re
10	244		DRAKE POINT - SABINE PENINSULA/MELVILLE ISLAND	Detailed testing completed. Re
11	266		Bathurst Island - Young Inlet (East of Humphries Hill)	Remedial action plan complete
12	282		Bathurst Island - Ile Vanier	Remedial action plan complete
13	288		Lougheed Island (L1)	Detailed testing completed. Re
14	311		CAPE DORSET 2 (NOTTINGHAM IS.)	Remedial action plan complete
15	351		GOODROCK MINE (Gordon Lake)	Detailed testing completed. Re
16	379		BAR C - TUNUNUK	Remediation / risk managemer
17	386		O`CONNOR LAKE (AMERICAN YELLOWKNIFE MINE; MWK NO.1 VEIN)	Detailed testing completed. Re
18	387		COPPER PASS MINE-HEARNE CHANNEL (SACHOWIA LAKE)	Detailed testing completed. Re

Scroll to the far right, and you'll see the longitude and latitude coordinates for the sites.

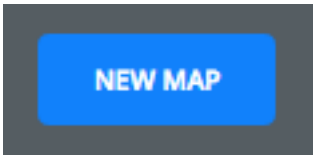


AB	AC	AD	AE
Latitude	Longitude	Coordinate Reference System	
62.60667	-113.465556	NAD83 (EPSG4269)	
62.34056	-112.746389	NAD83 (EPSG4269)	
64.27361	-115.204722	NAD83 (EPSG4269)	
66.00414	-118.073309	NAD83 (EPSG4269)	
61.96278	-128.216111	NAD83 (EPSG4269)	
62.98472	-113.204167	NAD83 (EPSG4269)	
78.12368	-103.177136	NAD83 (EPSG4269)	
75.36099	-105.72743	NAD83 (EPSG4269)	
67.20528	-118.591667	NAD83 (EPSG4269)	
76.33841	-98.694458	NAD83 (EPSG4269)	
76.13333	-104.033333	NAD83 (EPSG4269)	
77.34953	-105.30697	NAD83 (EPSG4269)	
63.11197	-77.938633	NAD83 (EPSG4269)	
63.03139	-113.135	NAD83 (EPSG4269)	
69.00691	-134.672628	NAD83 (EPSG4269)	
61.325	-111.792222	NAD83 (EPSG4269)	

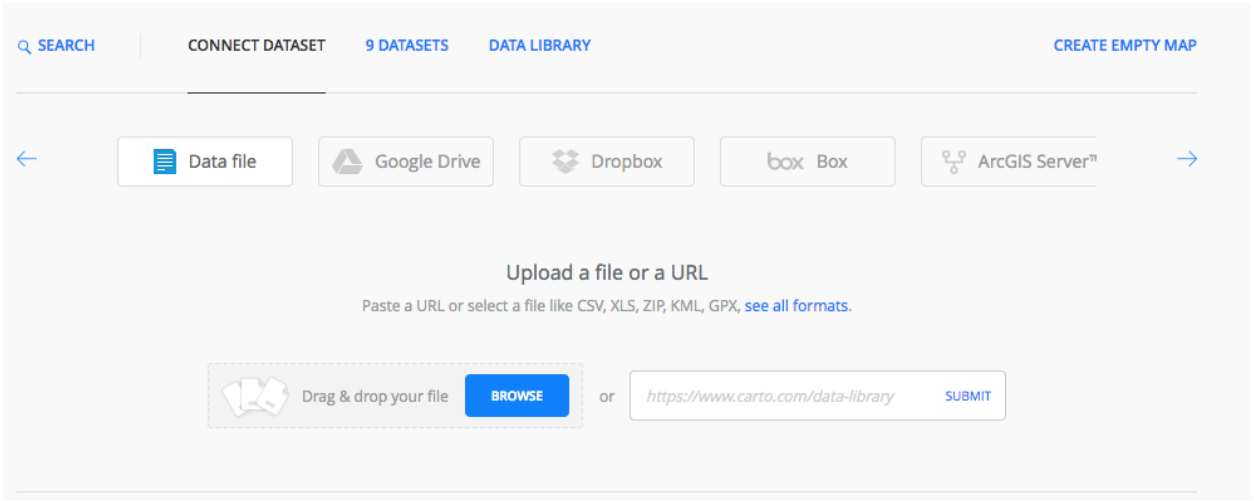
Column AD indicates that the latitudes and longitudes are based on the NAD83 datum. Latitude and longitude coordinates in NAD83 are slightly out of alignment with the same latitude and longitude points as positioned by the WGS84 datum used by Carto and other web mapping services. But for the purposes of this exercise, the offset of up to a metre or so is not meaningful, so we'll go ahead and upload the data. You'll find a discussion of datums and reference systems in Chapter 7 of *The Data Journalist.*

Carto makes it extremely easy to upload data that contains geographic coordinates such as latitudes and longitudes or UTM coordinates.

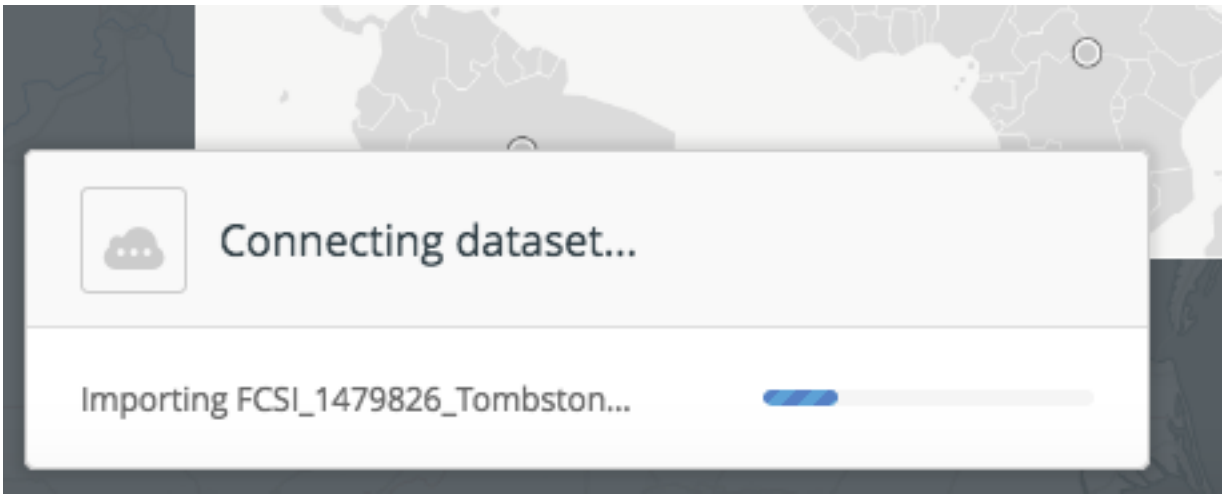
To begin, log into your Carto account, then click on New Map.



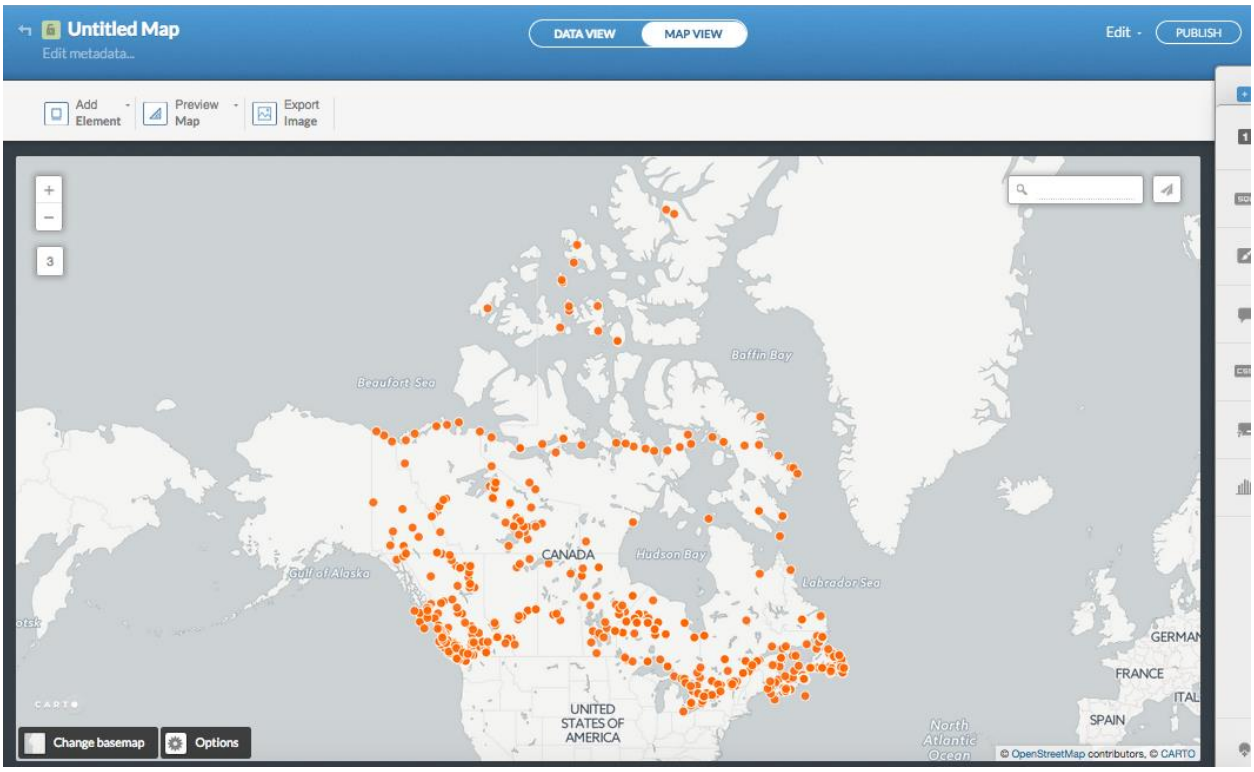
Now, click Connect Dataset and browse to the location of the file on your computer. In this case, we are uploading the CSV file of contaminated sites, but you can also upload KML files, zipped shapefiles, and a number of other formats, as well as connect to files contained in your DropBox or Google Drive. Carto is more flexible in this regard than other cloud mapping services such as Google Fusion Tables and ArcGIS Online.



In our case, we will browse for the contaminated sites data that we downloaded from the Treasury Board site. Once the file is selected, it will appear in the centre of the dialogue window. Click Connect Dataset to complete the process. You will see a progress meter.

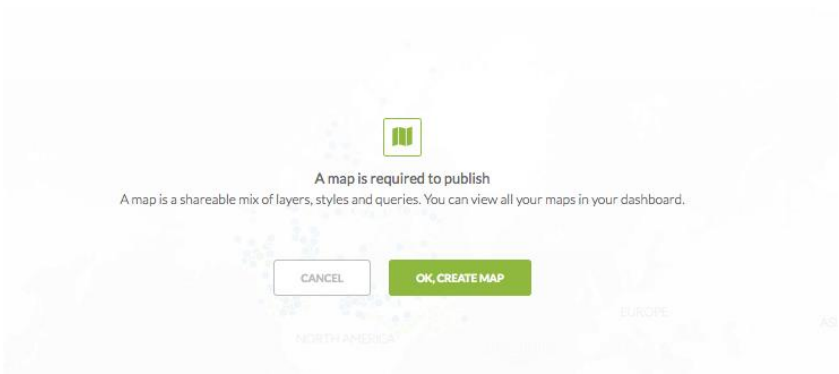


When the upload is completed, and if the latitude and longitude columns contained properly formatted coordinates, a new map will automatically be populated with the points.

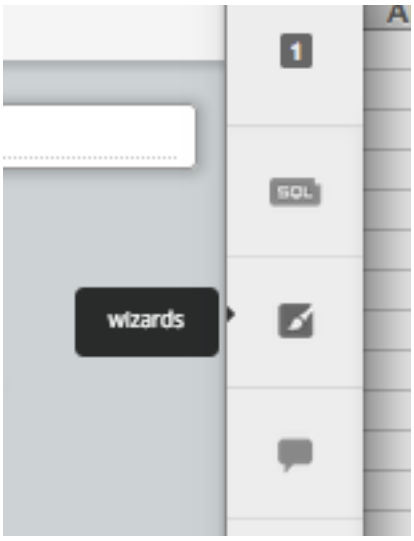


You now have a number of options for customizing your map before you publish it.

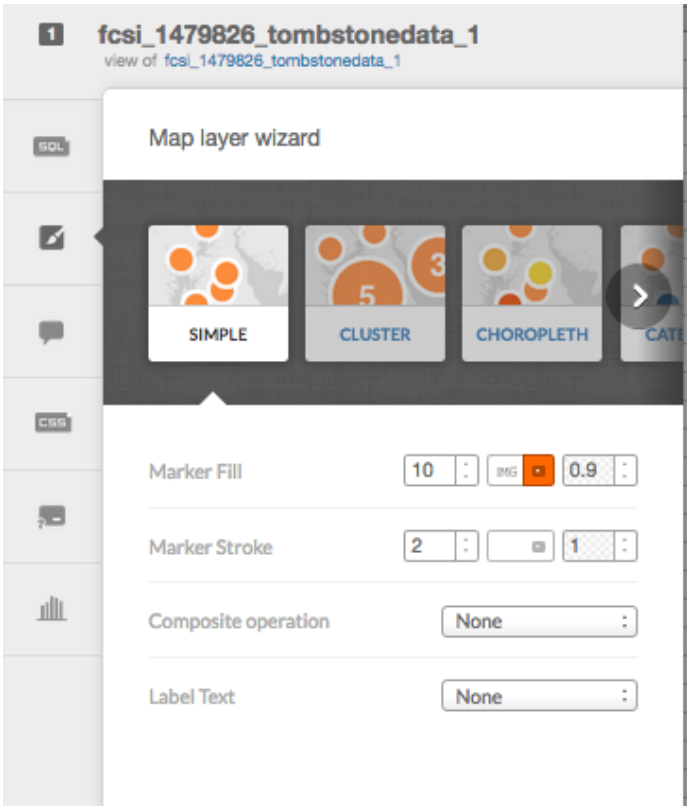
The first step is to click on Visualize. This will create a map.



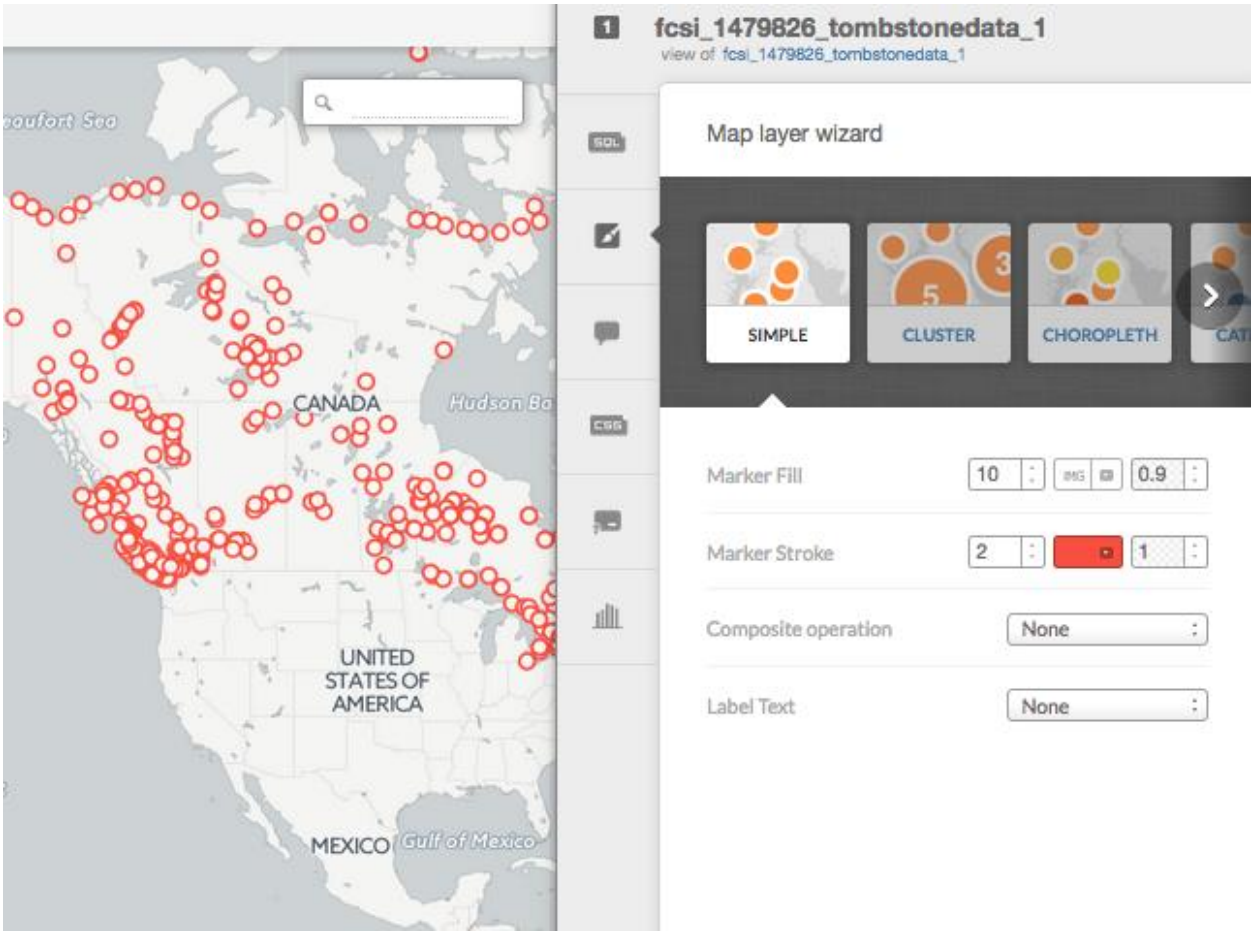
At first, the map will be in what Carto calls “simple” mode, with one dot for each of the points in the uploaded data. To the right of the screen, you will see a sidebar with various icons. These open different tools that you can use to manipulate the look of the map. If you hover over any of the icons, a tooltip appears with a label for that tool. We’ll click on the wizards icon.



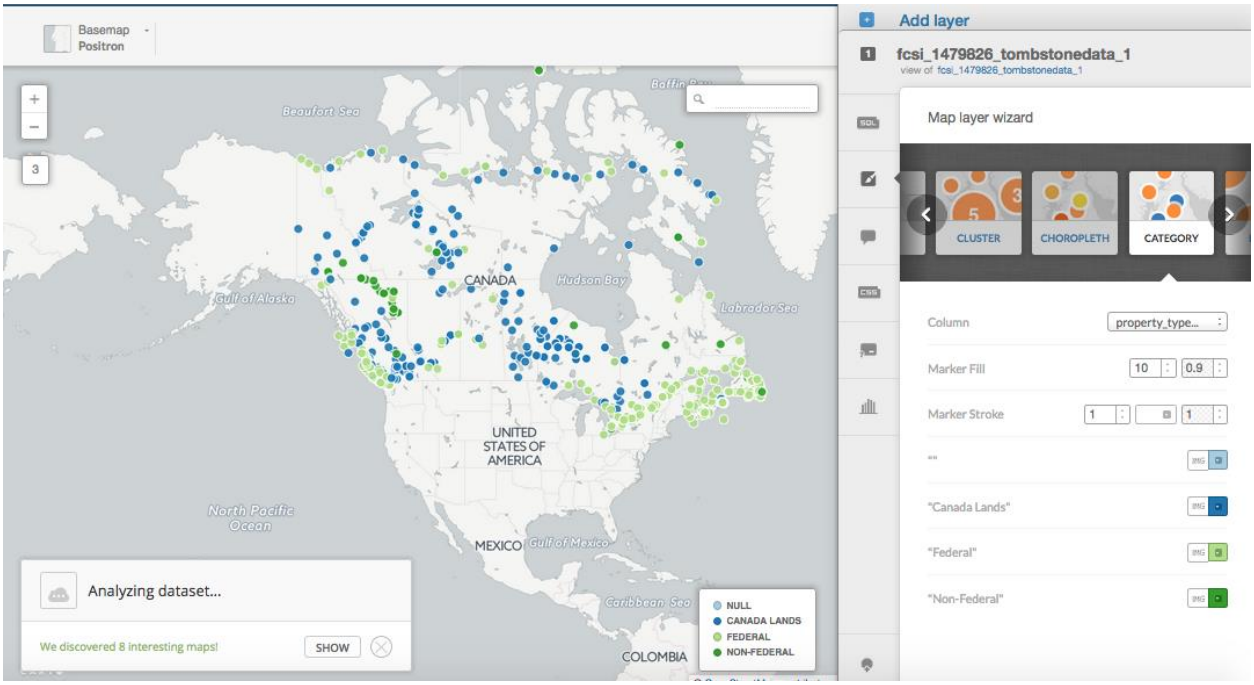
This will open a new dialogue that offers a number of quick ways to style your map.



As you can see, even with the simple dot option, you can alter the styles on the map. Here, we have increased the stroke on the dots, which is an outline around the centre dot. You can set the colour of both the fill for the central dot, and for the stroke, allowing you to create effects such as an outline effect, as seen below.



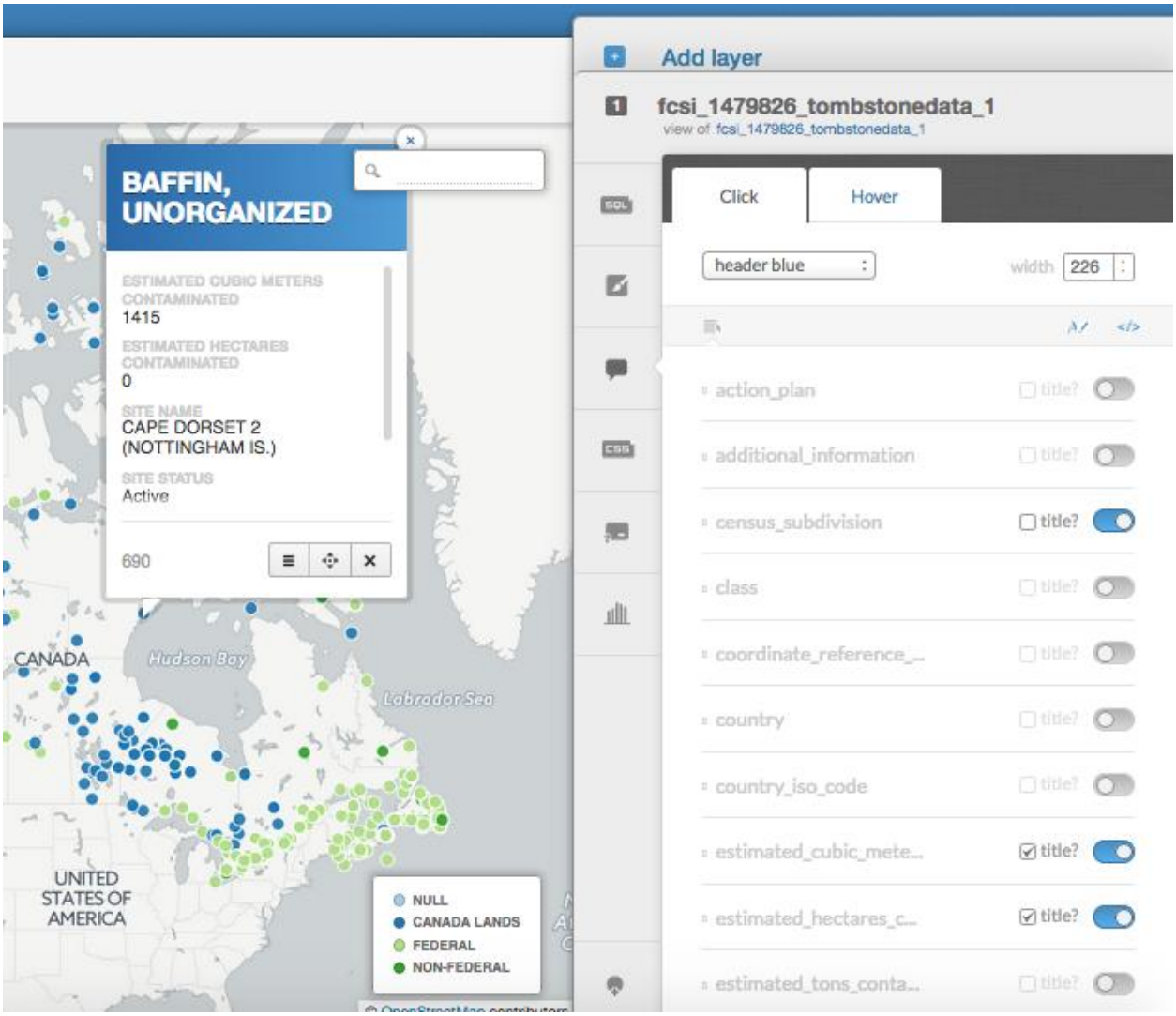
There are a number of other choices available among the wizards. Cluster will merge points close together into clusters numbered by the number of points contained. The closer you zoom in, the more the points will disaggregate. Choropleth allows you to colour the points based on values in a numeric field, while heat map creates a heat map that displays the intensity of number of points. The category option allows you to colour the points based on a categorical field, preferably one with a limited array of values. In the next image, you can see a map with points coloured by the type of land.



You'll notice that Carto automatically created a legend for us. Again, you can set the stroke for each point, as well as the fill colour.

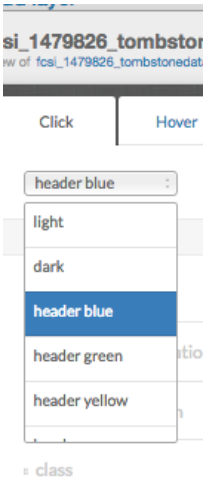
Along with changing the display of features on the map, you can also change the styling of the info window that pops up when you click on a feature. To start, click on the infowindow icon.

In the infowindow wizard dialogue you can set the behaviour for when the user clicks on a feature, or hovers over the feature. Each must be set separately.



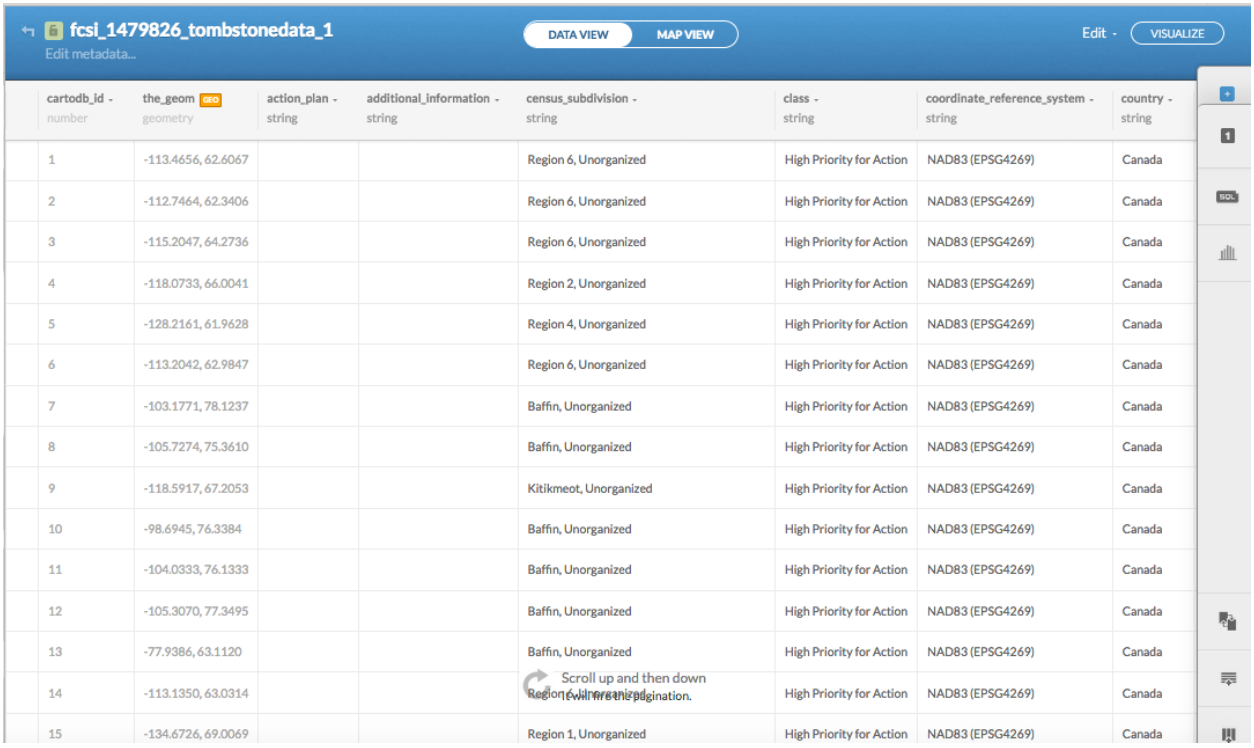
In the above image, a selection of fields has been set to appear in the info window. Clicking the small title? checkbox causes a title to appear with the field's value.

The dropdown list at the top of the dialogue permits you to change the overall style of the info window. The selected choice, header blue, creates the style seen in the image above. You can also adjust the width of the info window, in pixels.



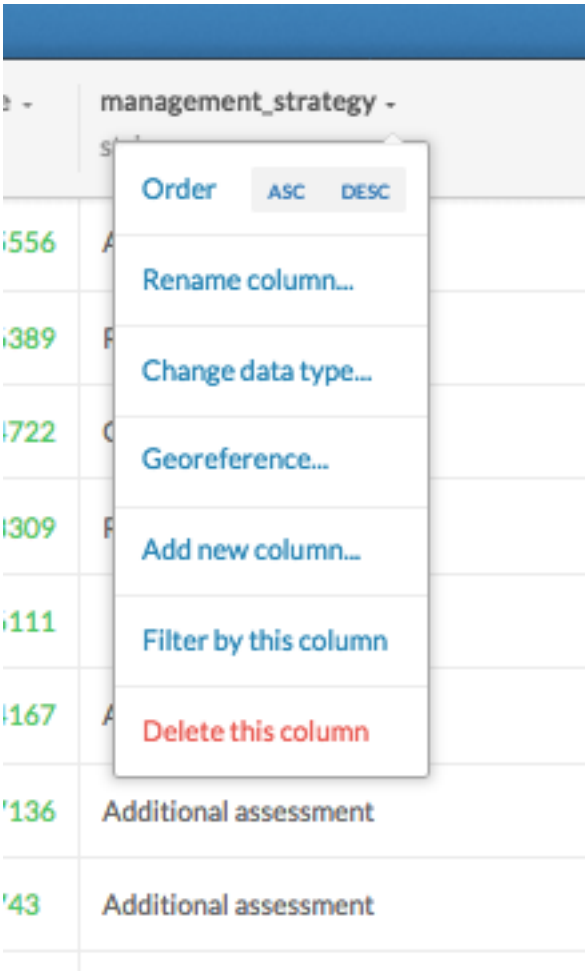
It is also possible to include an image in the header.

Clicking on DATA VIEW will display the data behind the map. Along with the name of each field, you will also see its data type.

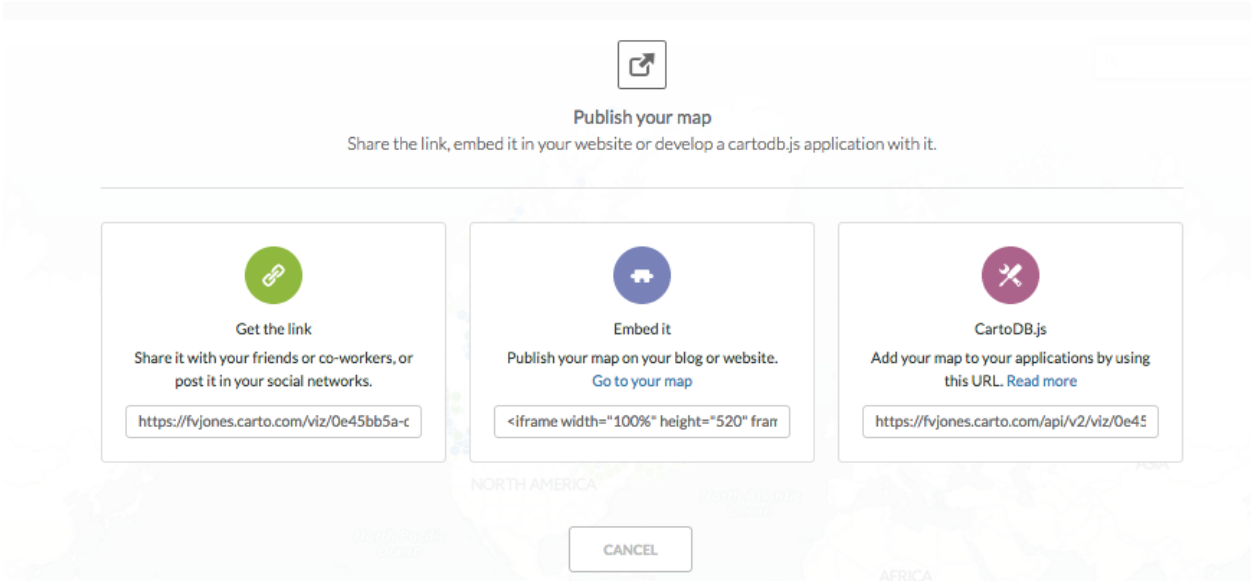


cartodb_id number	the_geom geometry	action_plan string	additional_information string	census_subdivision string	class string	coordinate_reference_system string	country string
1	-113.4656, 62.6067			Region 6, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
2	-112.7464, 62.3406			Region 6, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
3	-115.2047, 64.2736			Region 6, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
4	-118.0733, 66.0041			Region 2, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
5	-128.2161, 61.9628			Region 4, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
6	-113.2042, 62.9847			Region 6, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
7	-103.1771, 78.1237			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
8	-105.7274, 75.3610			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
9	-118.5917, 67.2053			Kitikmeot, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
10	-98.6945, 76.3384			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
11	-104.0333, 76.1333			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
12	-105.3070, 77.3495			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
13	-77.9386, 63.1120			Baffin, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
14	-113.1350, 63.0314			Region 6, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada
15	-134.6726, 69.0069			Region 1, Unorganized	High Priority for Action	NAD83 (EPSG4269)	Canada

Clicking the small dropdown arrow at the top of a field allows you to sort a field in ascending and descending order, and perform other actions such as renaming the column and filtering so only some of the values are visible (similar to the Excel filter).



When you are ready to publish your map, click on PUBLISH.



You'll most likely be interested in the middle option, Embed it, which will provide

you with an embed code that will allow you to embed you map in a story, such as in a WordPress or other CMS post. Here is a typical Carto embed code:

<iframe width="100%" height="520" frameborder="0" src="https://fvjones.carto.com/viz/0e45bb5a-d8d5-11e6-ad5b-0e233c30368f/embed_map" allowfullscreen webkitallowfullscreen mozallowfullscreen oallowfullscreen msallowfullscreen></iframe>

You can manually change the parameters of the iframe here, such a its width and height.

If you have difficulty embedding the iframe in WordPress, you can use the iframe plugin.

Creating a choropleth map by joining two datasets

Making a choropleth map using two datasets is also easy with Carto. You'll need a polygon layer and an associated dataset containing values you wish to display within the polygon boundaries. For our example, we will build a map that shows the age-adjusted percentage prevalence of obesity by U.S. county, as measured by the U.S. Centers for Disease control.

To begin we will need to upload our two datasets. You will use two files. A csv file of obesity numbers and percentages for each U.S. county in 2013, and a map layer of U.S. counties from the U.S. Census Bureau. The two files are available for download from the companion website to *The Data Journalist.*

If you wish to work with the original data, you can download it from the CDC and Census websites, though we have modified each of the files slightly to make the process smoother.

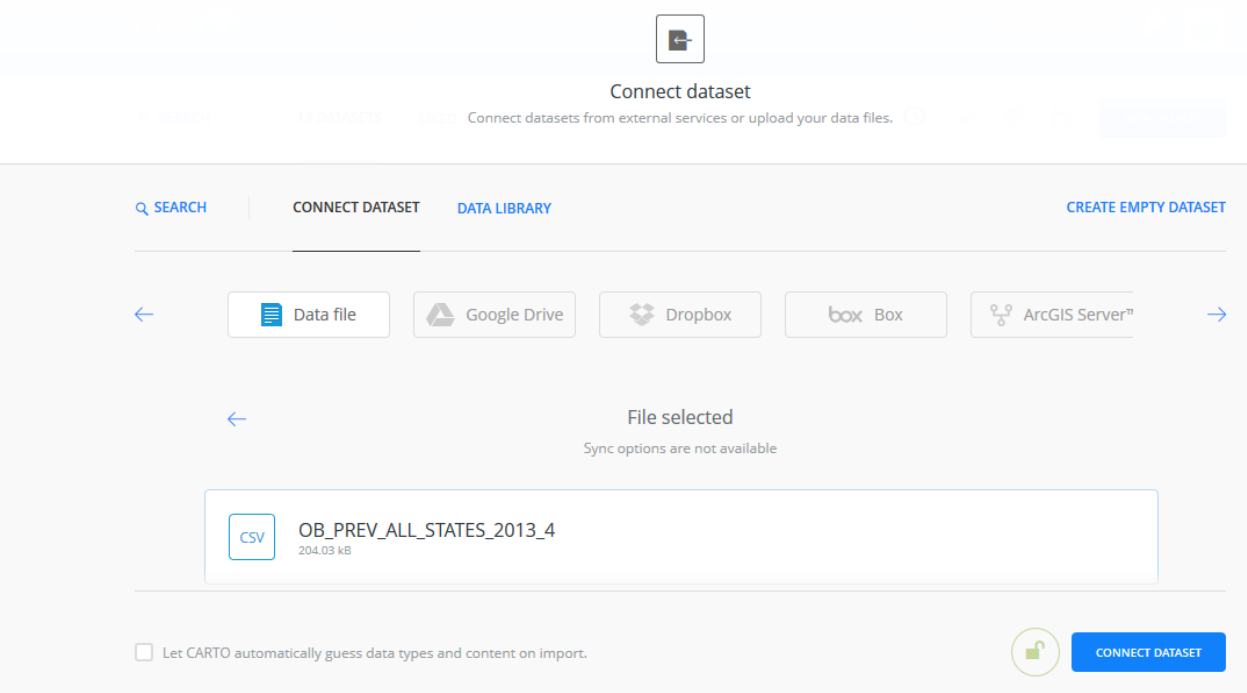
The obesity numbers come from the U.S. Centres for Disease control and the original data can be downloaded from https://www.cdc.gov/diabetes/data/countydata/countydataindicators.html

Download the AllStates.XLSX file under the Obesity Prevalence tab. Our file has been modified to include only 2013 numbers, and to eliminate spaces in field names. We also converted the data to CSV format.

The counties file is available from https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html

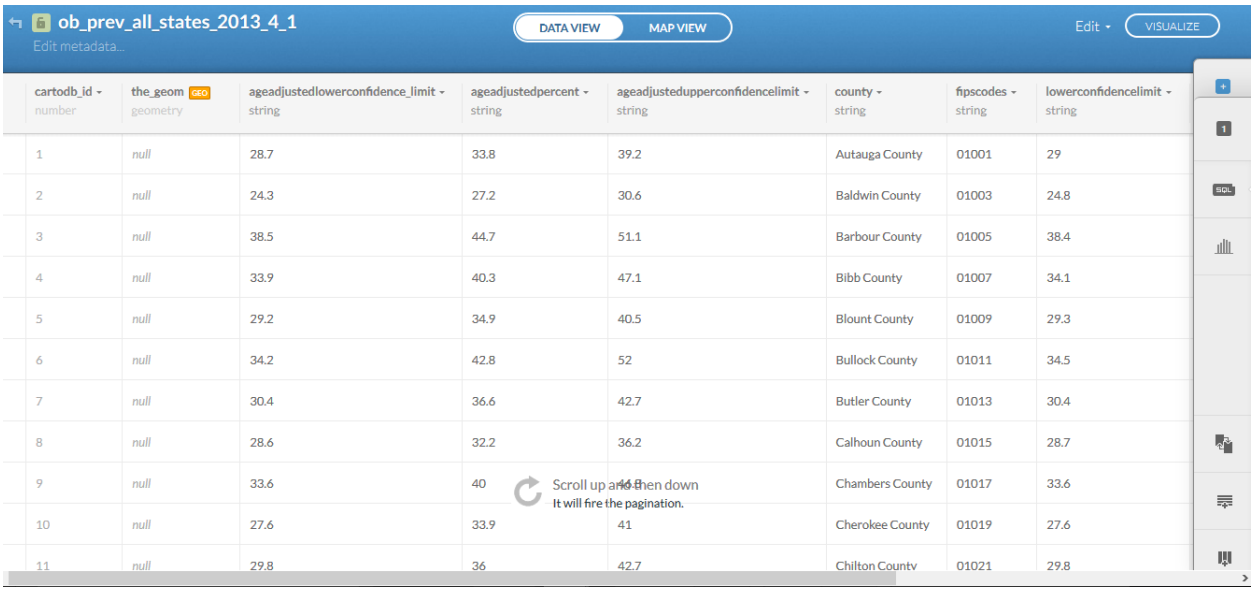
Download the 1:500k scale file. The file available for download from the companion website has been modified to combine the state and county unique identifiers into one field, so as to be able to join easily with the obesity data.

To begin, we will upload our CSV file of obesity prevalence. Click on the New Dataset icon, then the Browse icon. Browse to the file, select it, and choose Open.



You can often let Carto guess the field data types, but in this case we will uncheck the *Let CARTO automatically guess data types and content on import* checkbox because the unique identifiers for states/counties have leading 0s. By unchecking the checkbox, the fields will all be imported as text. This will preserve the leading 0s in the county id. We can modify the data types later so numeric columns have the number data type.

Click Connect Dataset. The data will be uploaded.

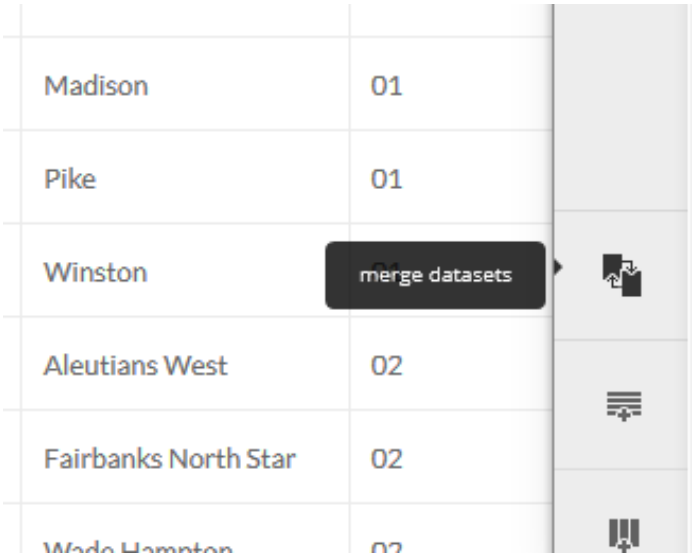


From here, we can modify the columns that are meant to be numbers, such as the number and percentage columns. To do this, click on the small downward pointing arrow at the top of the field you wish to modify, chose Change Data Type from the menu, then Number. When asked to confirm the change, do so. You can make the same change to other fields that need to be numeric.

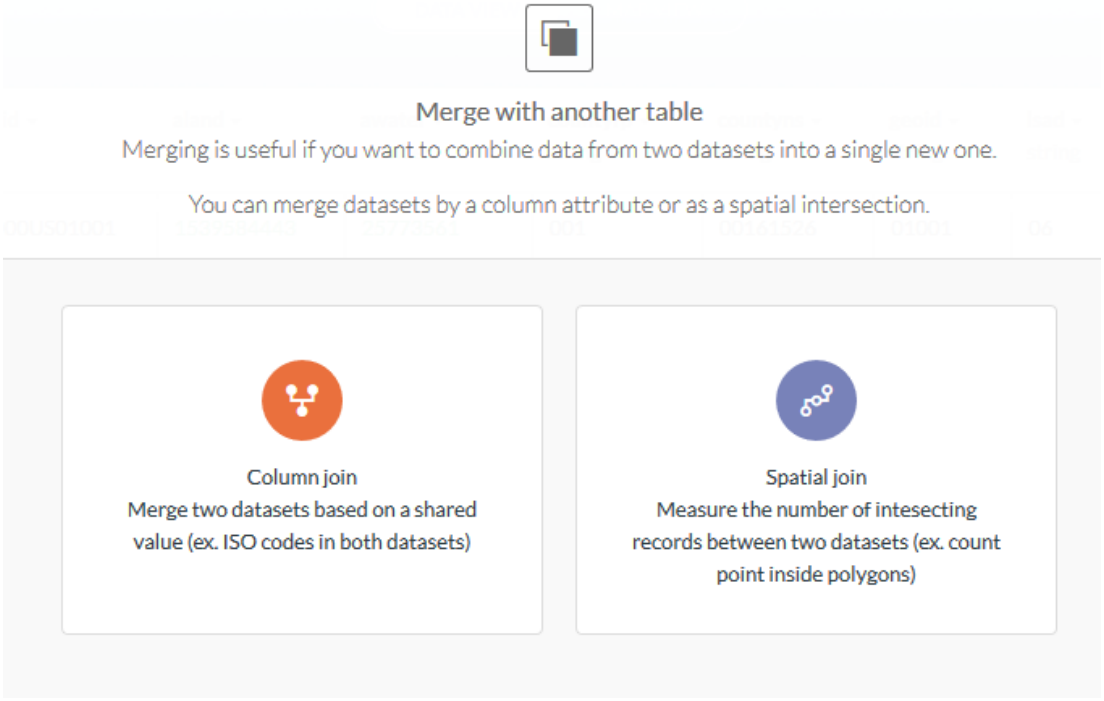
When you are done, click on the small arrow at the top left of the dataset, to go back to the list of datasets.

Now, upload the KML file of U.S. counties, or the zipped shapefile. Either will do.

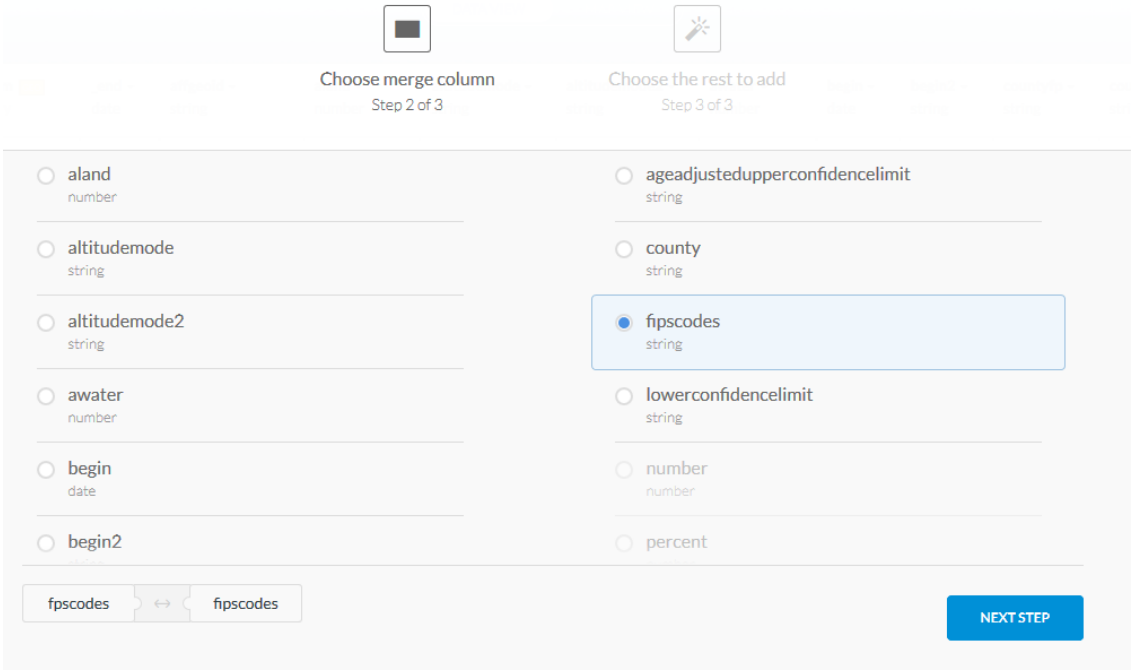
Once you have both datasets uploaded, go to one of the datasets, in Data View, and click on the Merge Datasets icon to the right of the screen.



On the screen that opens, click on Column Join.



This will open a page on which you can choose the second dataset you will join to, and the columns on which the join will be based in the two datasets.

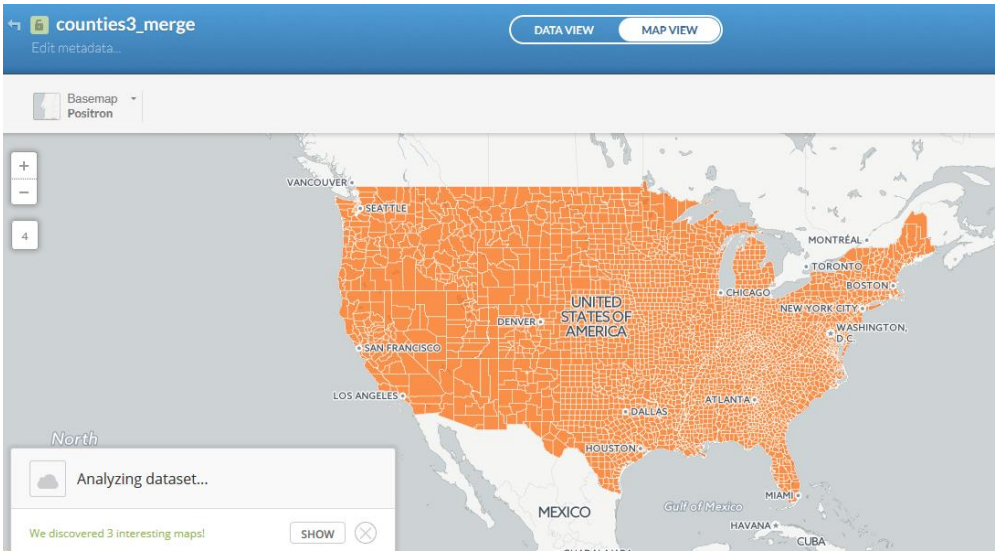


The chosen join will appear in the lower left. When you are satisfied that you have the appropriate fields, that is fields that contain the identical fields in the two datasets on which the join will be based, choose Next Step.

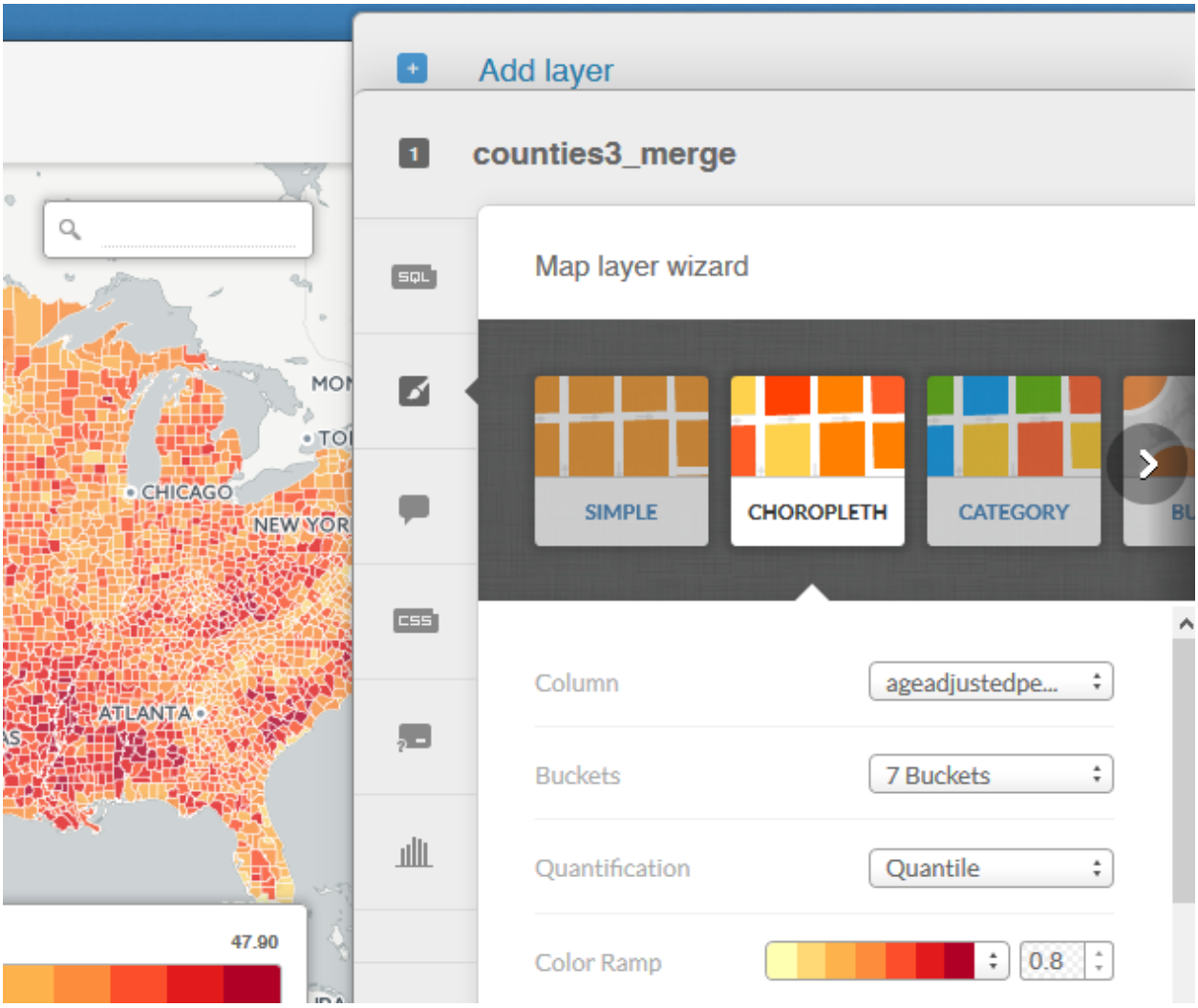
The next page allows you to choose which fields you would like to have in the merged datasets. Select some or all of the fields, though in most cases there is no reason not to select them all.

When you are satisfied, choose Merge Datasets. A new dataset, containing the joined attributes from the two predecessor datasets will be created.

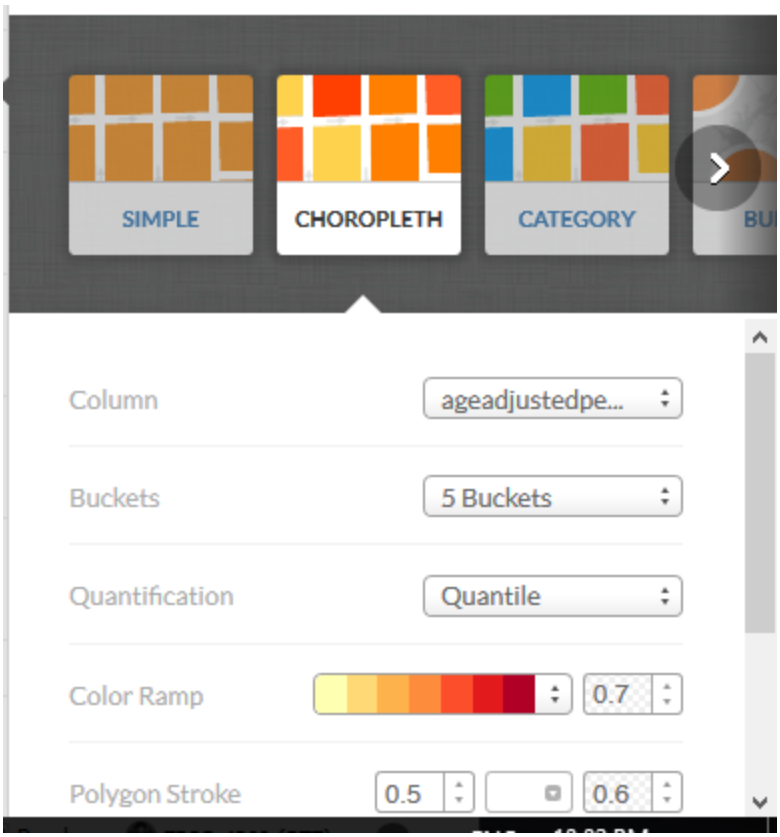
You can now switch to Map View.



On the right hand side of the page, choose the Wizards icon, and choropleth.



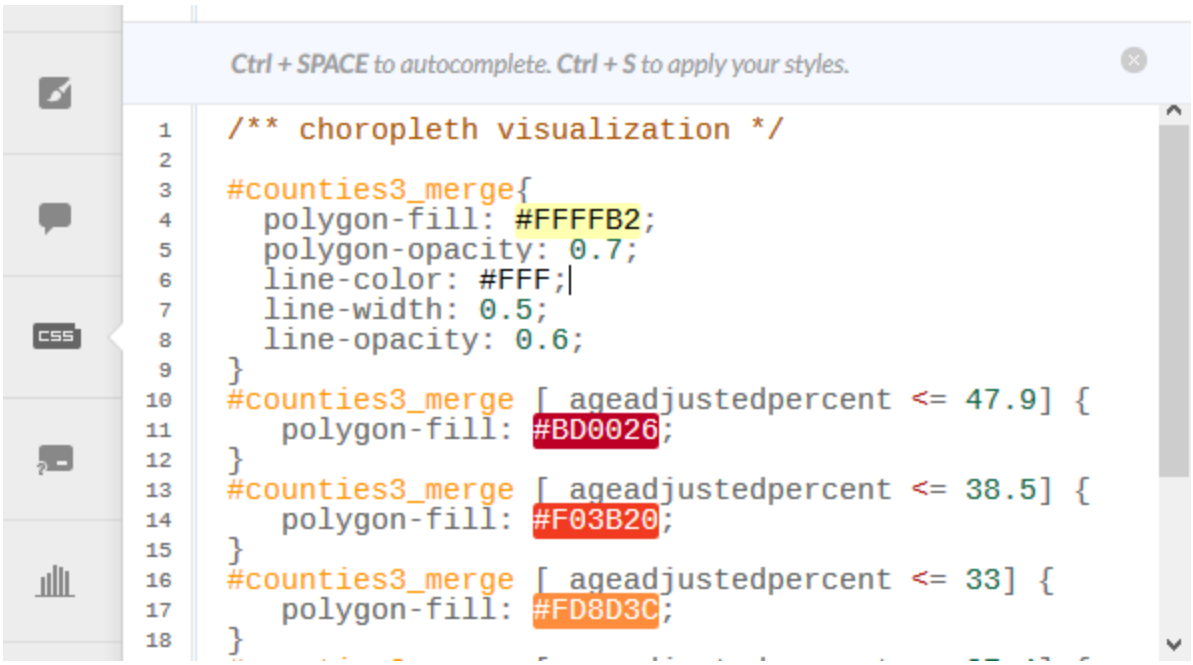
From here, you can choose the options for your map.



In the above image, we have selected the age adjusted obesity prevalence field as the value column. We are choosing this field for two reasons. First, it is not a raw number. As we discussed in Chapter 6, you should never create a choropleth map with raw numbers because different geographies will not be able to be compared on an apples-to-apples basis. The number of obese people will be affected most dramatically by the size of the population itself. For example New York County (Manhattan) will have a great many obese individuals, but the percentage of obese people may be lower than in a county with a small population in a southern, rural area. The other reason we chose this field is that the age-adjusted percentage takes into account differing age profiles in different counties. Many people weigh more as they get older.

We have set the number of buckets (numeric ranges) to five, quantile as the classification method, a light amber to dark red colour ramp, and a white stroke between polygons of .5 pixels in width.

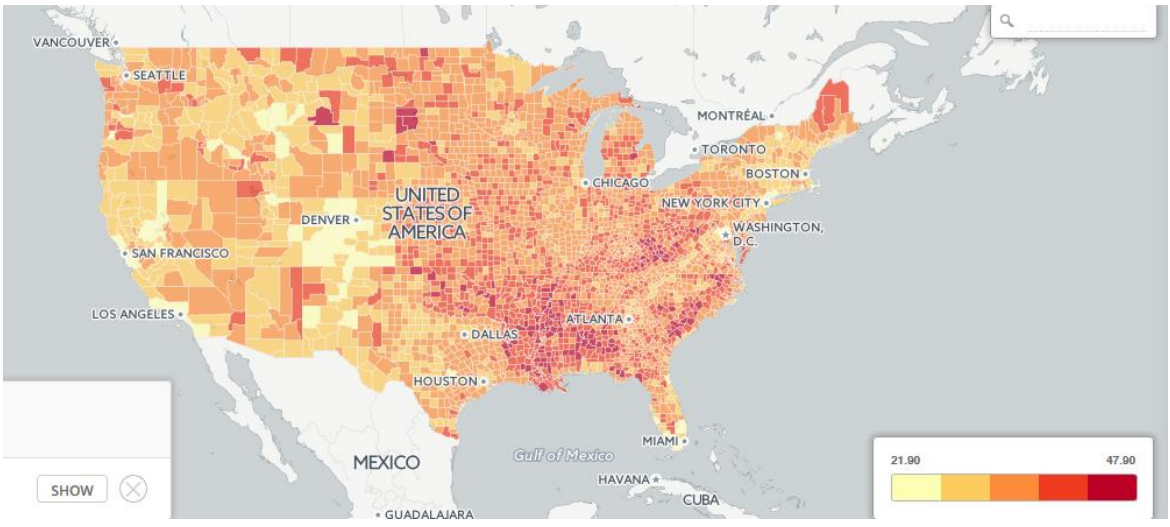
If you want even more fine grained control of the colour and stroke, you can click on the CSS icon and enter values directly.



Chapter 6 of *The Data Journalist* has a discussion on how to choose appropriate colours on choropleth maps. You can also use the colorbrewer tool discussed in Chapter 6 to make more sophisticated colour choices, changing the Hexadecimal colour values in the CSS.

The quantile classification is valid in this case because it neatly divides up the counties into one fifth of the values each. For a longer discussion of classification schemes and colour choices, see Chapter 6 of *The Data Journalist.*

Our finished map shows graphically how obesity varies across the U.S.



When you are done, you can publish your map as discussed earlier in this tutorial.

Going beyond

This tutorial has been a basic introduction to Carto. It can do much more, however, including building JavaScript based interactive applications. We would encourage you to explore the online documentation to learn more about this versatile mapping solution.