

Mapping with CartoDB (Tanzania Edition)

In this exercise we are going to use election data to build informative and interactive maps of Tanzania.

In order to complete this exercise you need to have a CartoDB account. This can be created at <http://cartodb.com>

This exercise requires a number of key datasets that are available to download from the course website.

1. The regional boundaries of Tanzania. (Source: Tanzania National Bureau of Statistics)
2. Population and region area data. (Source: Tanzania National Bureau of Statistics)
3. Education pass rate and school rank (Source: National Examinations Council of Tanzania)

Step 1 - Preparing the data (already done)

In order to prepare each dataset, several translations have been carried out to extract it from the source and translate each dataset into formats appropriate for use with CartoDB. These steps are briefly described below.

Regional boundary files

Source: Tanzania National Bureau of Statistics (NBS)

Source URL:

<http://www.nbs.go.tz/nbstz/index.php/english/statistics-by-subject/population-and-housing-census/258-2012-phc-shapefiles-level-one-and-two>

These files come as a rather large zip file containing a lot of shape files. Shape files (SHP) are a zip file format developed by ESRI to a “mostly” open specification. The terms of the specification are vague, particularly around usage rights, many online tools choose not to support the format for legal reasons. This means that in order to make the format usable, we need to translate it into something compatible with more tools.

Unwanted data was first removed from the source NBS zip file, leaving only the region data inside the zip file. This zip file was then uploaded to the SHP to KML translator available at <http://www.mapsdata.co.uk/online-file-converter/>.

The resultant KML has been made available on the course website and is the one required for the exercise.



Population and region area data

Source: Tanzania National Bureau of Statistics (NBS)

Source URL: http://www.nbs.go.tz/nbs/takwimu/references/Tanzania_in_Figures_2015.pdf, also available via Wikipedia (https://en.wikipedia.org/wiki/Regions_of_Tanzania)

For the purposes of this exercise the data has been extracted from Wikipedia, however you could use a tool like pdftables.com to extract the data easily from the PDF.

The extracted data is available from the course website in CSV format.

Education pass rate and school rank data

Source: National Examinations Council of Tanzania

Source URL: <http://www.necta.go.tz/brn2015/>

In order for this data to be usable, a pivot table has been created from the data to list each region only once. All values have been summed or averaged in order to work out the overall regional statistics. A pass rate percentage value has also been added by divided the number of candidates that passed by the total number of candidates.

Step 2 - Create a regions map of Tanzania

Once you are logged into CartoDB, in the top right hand corner of the screen you should be able to select **maps** or **datasets**. Click **maps** and then click the **new map** button.

NEW MAP

On the next screen ensure you select **connect dataset** from the available options and then upload a **data file**. Here you can drag and drop or browse to upload the Tanzania_regions.kml file. Click **connect dataset** to upload this file. It may take a while to upload and process.

When complete you should get a map of Tanzania with all the regions outlined. **At the top right, give it a name.**

Step 2 - Add the region names

Unfortunately the source SHP and KML files do not contain the names of the regions, just their boundaries.

In order to combine this dataset with the population and other datasets we will need to add the region names manually.

DATA VIEWMAP VIEW

To do this click the **data view** tab at the top of the screen. On this screen you will be presented with a tabular view of the data. Note that all the long and detailed geometry data is hidden from view in the **the_geom** column and CartoDB simply lists the data as a **Polygon**.

Also in the **data view** you should be able to find the **cartodb_id** and **name** columns. To make the data usable we need to add the region names in the **name** column. A mapping between **cartodb_id** and **region name** is listed below.

Region	Carto_ID	Region	Carto_ID	Region	Carto_ID
Pwani	1	Dodoma	11	Mbeya	21
Simiyu	2	Tanga	12	Songwe	21
Geita	3	Mtwara	13	Arusha	22
Singida	4	Njombe	14	Morogoro	23
Iringa	5	Tabora	15	Katavi	24
Rukwa	6	Kilimanjaro	16	Manyara	25
Kagera	7	Shinyanga	17	Pemba South	26
Kigoma	8	Mara	18	Zanzibar North	27
Lindi	9	Ruvuma	19	Zanzibar South	28
Dar es Salaam	10	Mwanza	20	Zanzibar Urban West	29
				Pemba North	30

To edit the **name** column, simply double click it.

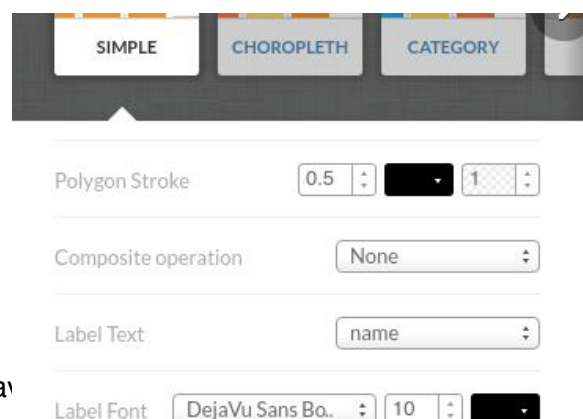
Note: This mapping is also available in the **population dataset**.

Step 3 - Display the region names on the map

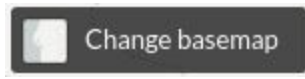
Once you have added the region names, return to the **map view** and click the wizards button from the menu on the right.

Using the wizard you can quickly customise your map, change the region fill color and outline color and label each region with it's name.

In the example on the right we have changed the outline color to black and changed the label



to show the region name. You can also change the basemap if it is too detailed.



Step 4 - Combine with population data

In order to combine the region data with the population data, we first need to upload the population datasets button. To do this, click the **back button** to get back to your dashboard.



From the drop down in the top left where it says **Maps**, click this and select **Your datasets**.

Click **New Dataset** and upload the population data.

NEW DATASET

Once done, click the **back button** again and select your region map from the **maps dashboard** to return to your map.

Ensure you are in **data view**. From the menu on the right, click the **merge datasets** button.



On the next screen select **column join**. In this case the population dataset contains region names, so this is the column we want to join the dataset based upon. Note that if you had a list of geographic locations (e.g lat-long) then you can do a **spatial join** and CartoDB can work out which places are in each region by location boundary.

Select **column join** and on the next screen select your population dataset on the right hand side.

From the **left hand** column select the **name** value and from the **right column** select **region**. These two columns contain the same name and are where we want to merge the datasets.

The screenshot shows two side-by-side panels for selecting columns. The left panel has three unselected options: 'draworder' (number), 'extrude' (number), and 'icon' (string). The 'name' (number) option is selected and highlighted with a blue border. The right panel has four unselected options: 'carto_id' (number), 'population' (number), 'postcode' (string), and 'zone' (string). The 'region' (string) option is selected and highlighted with a blue border. At the bottom left, there is a Creative Commons license icon (CC BY SA) and the text 'The Open Data Institute (UK Node)'.

On the next screen, from the left, **only** select **the_geom** and **name** columns . From the population dataset (right hand side) select **everything** except **the_geom** and **carto_id**.

When done, click **merge datasets**.

This will create an entirely new dataset. It will forget all the customisations we did to the map as we now have a new map as well. The old one is still accessible from your dashboard.

the_geom geometry	<input checked="" type="checkbox"/>	the_geom geometry	<input type="checkbox"/>
_end date	<input type="checkbox"/>	area_km2 number	<input checked="" type="checkbox"/>
altitudemode string	<input type="checkbox"/>	capital string	<input checked="" type="checkbox"/>
begin date	<input type="checkbox"/>	carto_id number	<input type="checkbox"/>
description string	<input type="checkbox"/>	population number	<input checked="" type="checkbox"/>
draworder number	<input type="checkbox"/>	postcode string	<input checked="" type="checkbox"/>
extrude number	<input type="checkbox"/>	region string	<input checked="" type="checkbox"/>
icon string	<input type="checkbox"/>	zone string	<input checked="" type="checkbox"/>
name number	<input checked="" type="checkbox"/>		

Step 5 - A population map

From **map view** select the **wizard** from the side menu.



This time we are going to create a **Choropleth** map using the population data to calculate the color of each region on a continuous scale.

Once uploaded you will be presented with a tabular view of the data in this KML file. Looking through the table you should be able to find a column that has the **geo** icon displayed next to the column header.


Select **population** as the column and then you can adjust the number of buckets, interval type and

also colours.

Step 6 - Adding labels



Click on any region on the map and you will note that the label text is not very helpful.

Before we add labels, note that in this view you can also edit the geometry of the region by clicking the **edit geometry** button . This is not advised on detailed geometries - like this one - as you will likely not have a computer powerful enough to manage all the data in your web browser.

Click **select field** from the popup and choose those you wish to display from the wizard on the right.

Even though this is easy, it doesn't produce very clear labels. Luckily everything in CartoDB can be customised in code.

To customise how the labels look, click the change html  icon.

If you know HTML code, feel free to play around with this. If not, try the following:

```
<div class="cartodb-popup v2">
  <a href="#close" class="cartodb-popup-close-button close">x</a>
  <div class="cartodb-popup-content-wrapper">
    <div class="cartodb-popup-content">
      <h2 style="color: black;">{{name}}</h2>
      <hr style="margin-bottom: 6px;">
      <p><b>Capital</b>: {{capital}}</p>
      <p><b>Area</b>: {{area_km2}}km<sup>2</sup></p>
      <p><b>Population</b>: {{population}}</p>
      <p><b>Postcode</b>: {{postcode}}</p>
    </div>
  </div>
  <div class="cartodb-popup-tip-container"></div>
</div>
```

When done click apply to update the label style.

Extension exercise - Education data

Why not try repeating steps 4 onwards with the education data and create an even bigger dataset. When creating a map, this time you can use a Choropleth map with the **passed percent** column or a category map with the **category** column. Both of these columns were added when the pivot table was created.