

Exercise 8b – Generate summary statistics from raster data

Learning objectives: The purpose of this exercise is to outline a method of generating summary statistics (min, max and mean) from a raster dataset by aggregating pixel values to a defined area (typically administrative or census units). This procedure will introduce the Zonal Statistics geoprocessing tool.

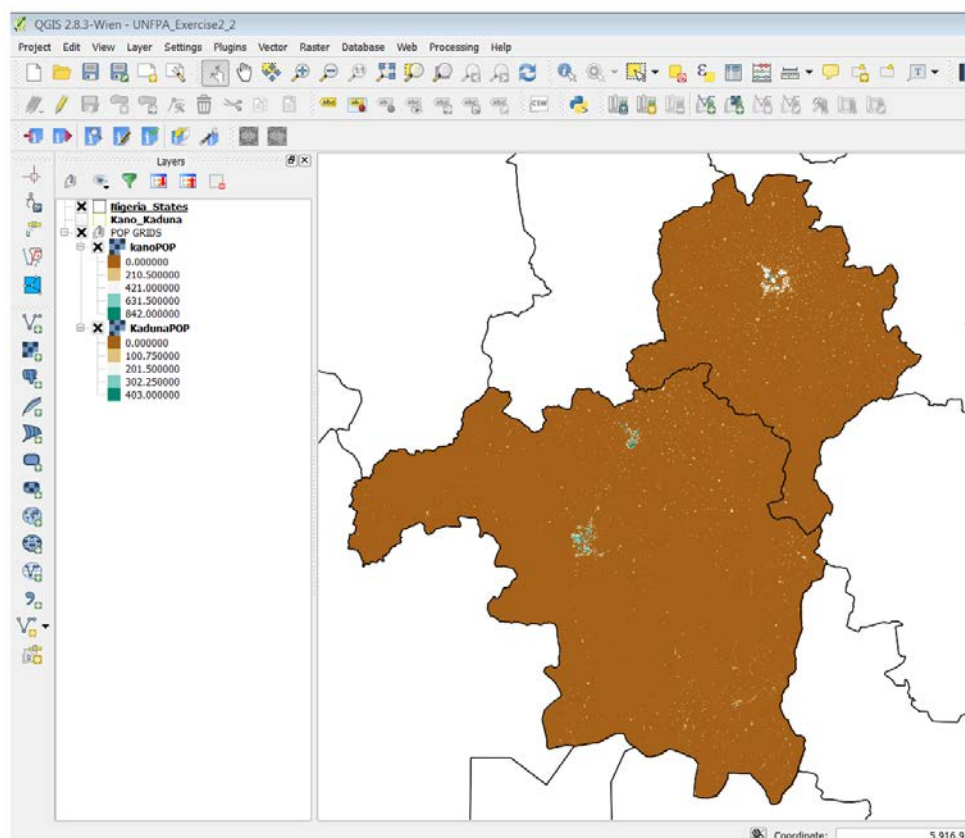
We will use population count grid data for two provinces in Nigeria. These datasets have been generated using the methods under discussion in this workshop. Initially we will prepare the data by combining raster datasets of two provinces into one, using the Merge geoprocessing tool. After generating zonal statistics we will produce a thematic population map.

Skills you will learn

- Merge two datasets together
- Zonal statistics
- Vector styling
- Buffer Tool

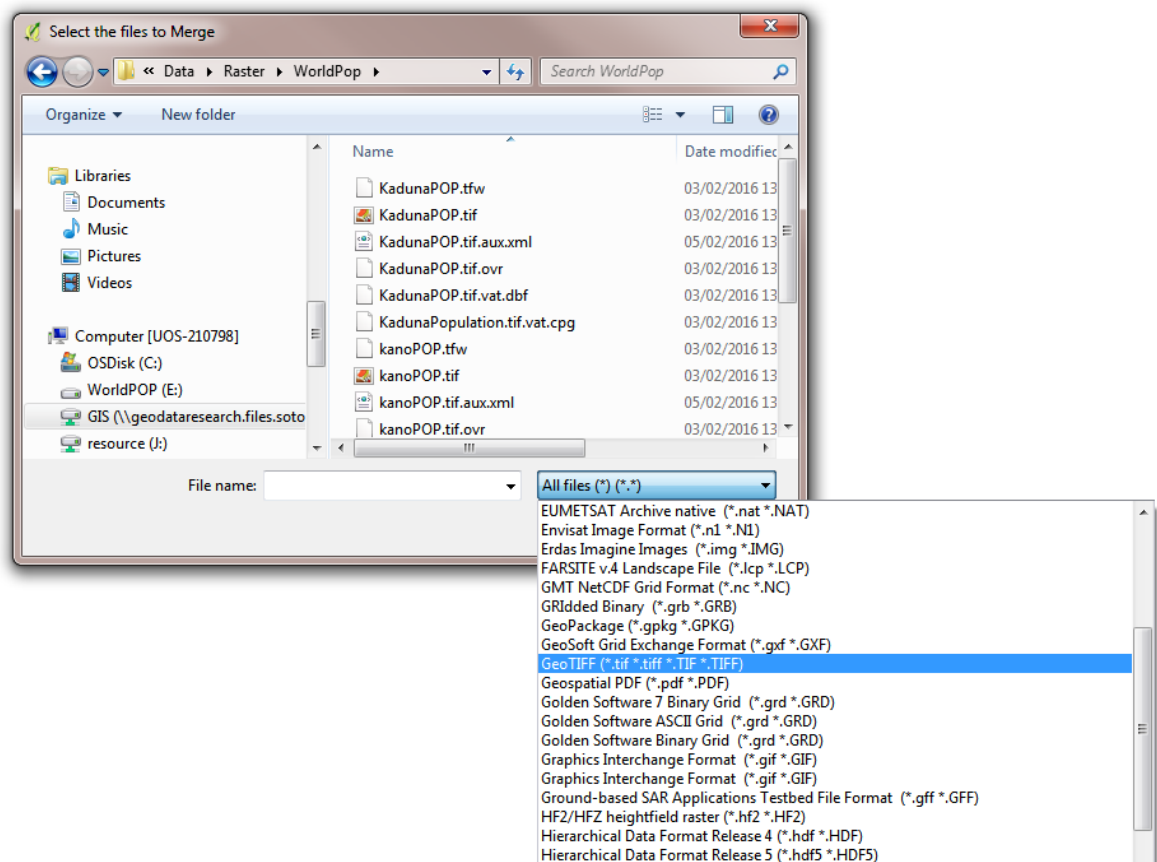
Part 1 – Merge (mosaic) raster datasets.

1. Open an existing QGIS project file from the following location:
 - C:\Intro_Quantum_GIS\Exercises\Exercise_8b.qgs

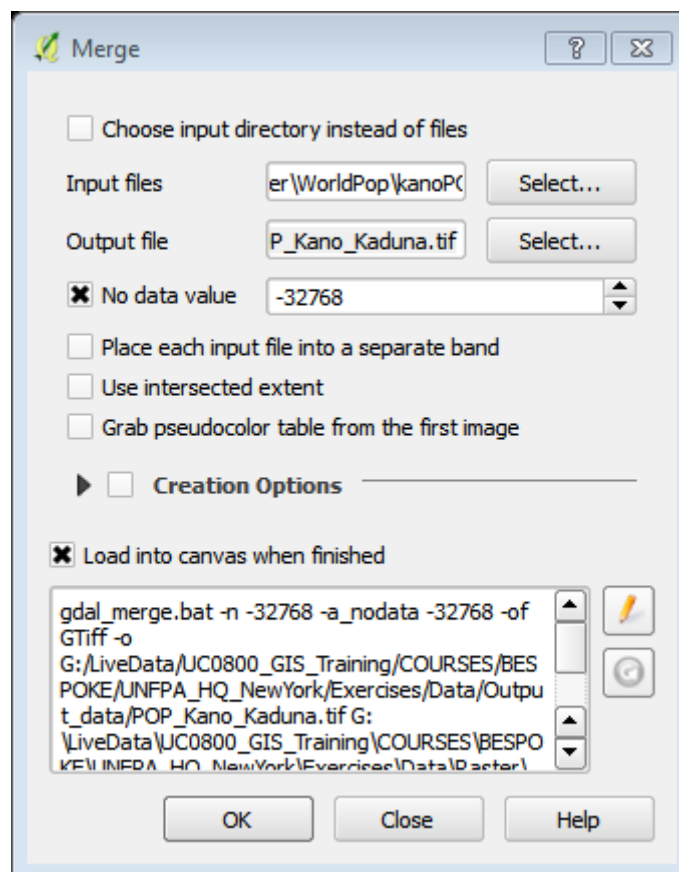


The raster datasets within the POP GRIDS folder represent two neighbouring provinces in Nigeria. The maps represent people per pixel in these two areas and consequently there are two separate scales of values (0 to 403 and 0 to 842), as displayed in the legend. Converting the two areas into a single raster would establish a single standard scale. This would clearly provide a better representation of population values across the combined region. In addition, it is an essential step in generating zonal statistics for the whole area.

2. Merging (mosaic) multiple datasets into a single dataset
 - From the **Raster** drop-down menu, select **Miscellaneous > Merge**
 - In the dialog box that appears, click **Select** button, to the right of **Input files**
 - Browse the following location:
C:\Intro_Quantum_GIS\Exercises\Data\Raster\WorldPop

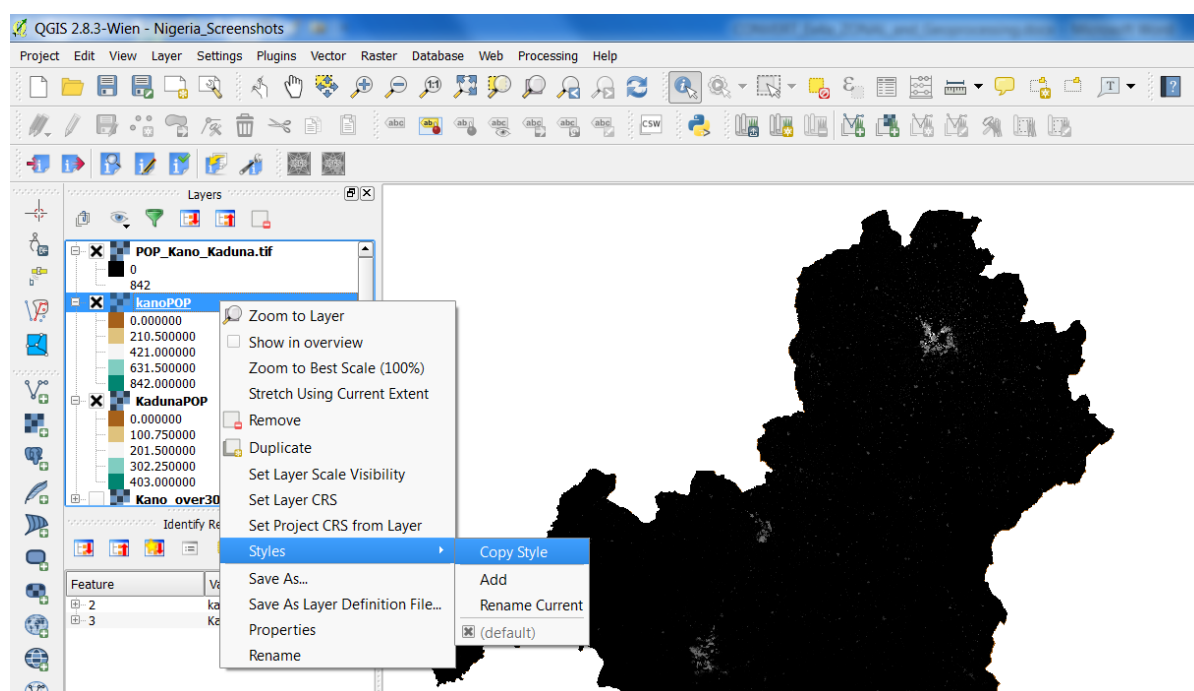


- Click the **All files** drop-down menu and select **GeoTIFF**
- Hold down **Ctrl** on your keyboard and select **KadunaPOP.tif** and **KanoPOP.tif**
- Click **Select** button, to the right of **Output file** and save the output as follows:
C:\Intro_Quantum_GIS\Exercises\Data\Output_data\POP_Kano_Kaduna.tif
- Click to enter an **X** in the **No data value** option, and type -32768 (this is the designated value assigned to no data pixels in the input rasters)
- Your dialog box should look as follows:



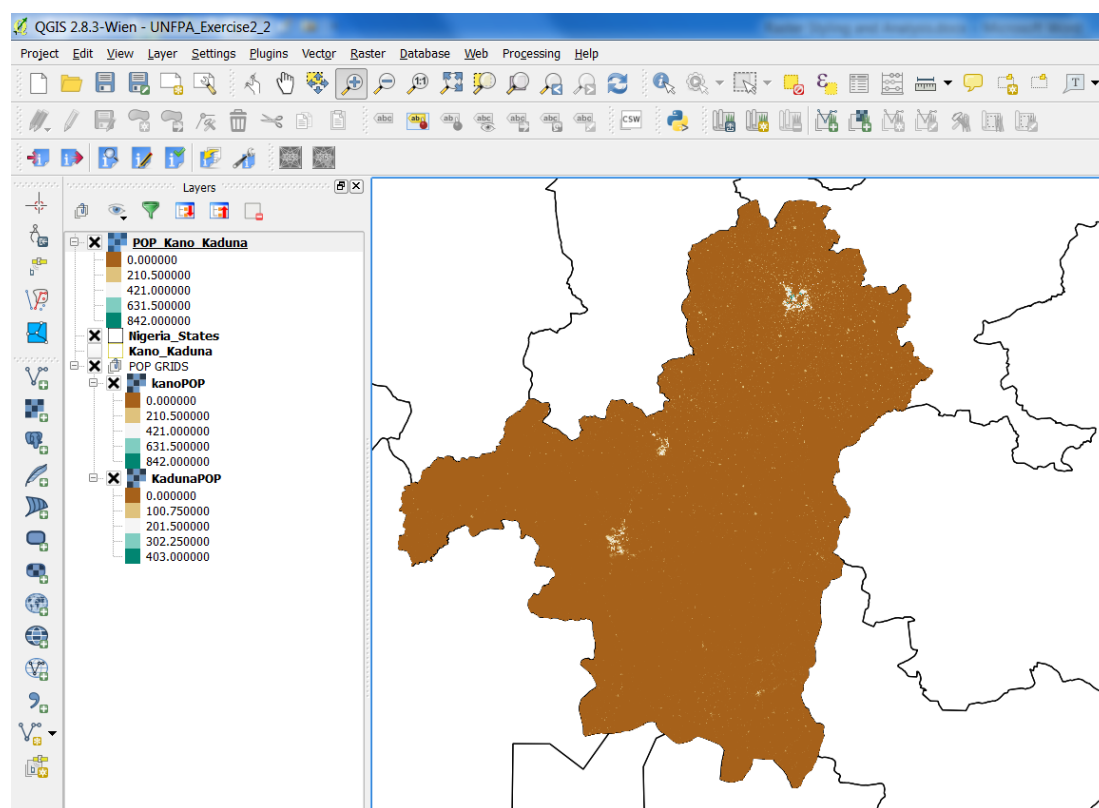
- Click **OK**

As you will see the merged dataset is presented in a black to white 'stretched' style. Follow the procedure illustrated in the screenshot to copy the style from KanoPOP to the new dataset:



- Right-click KanoPOP > Styles > Copy Style

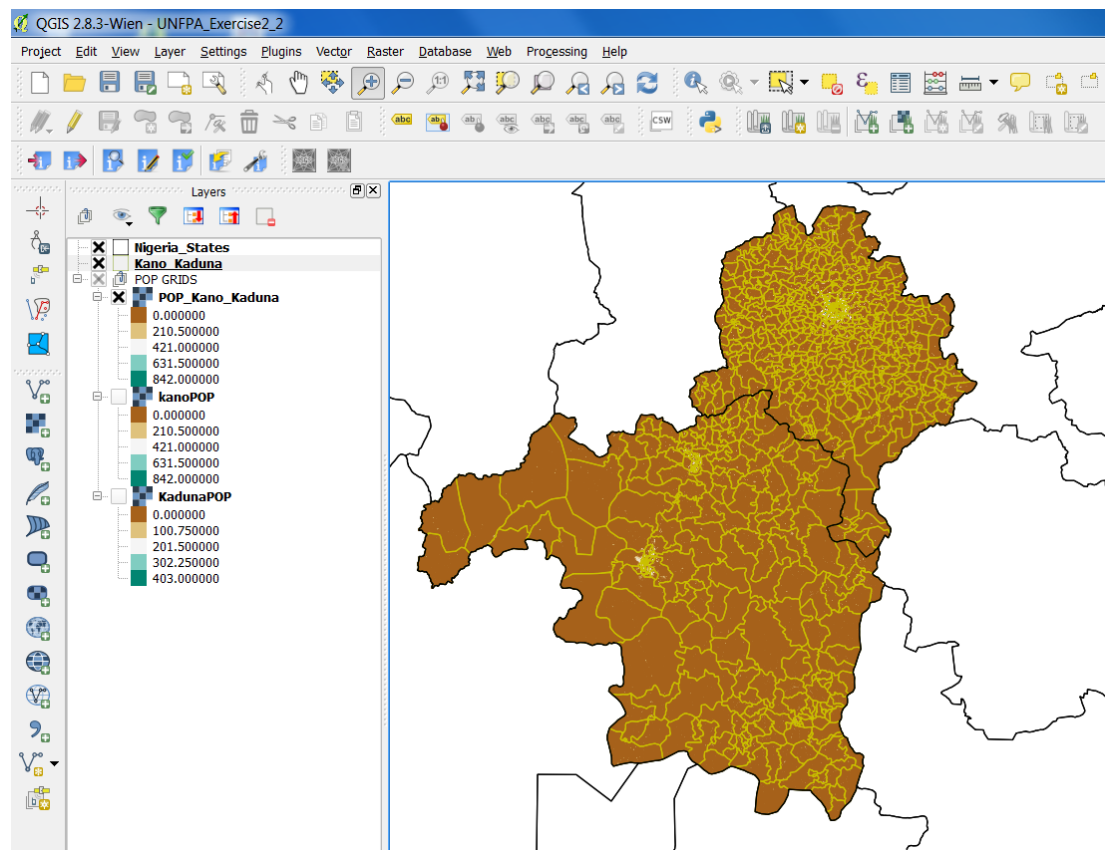
- Right-click POP_Kano_Kaduna > Styles > Paste Style



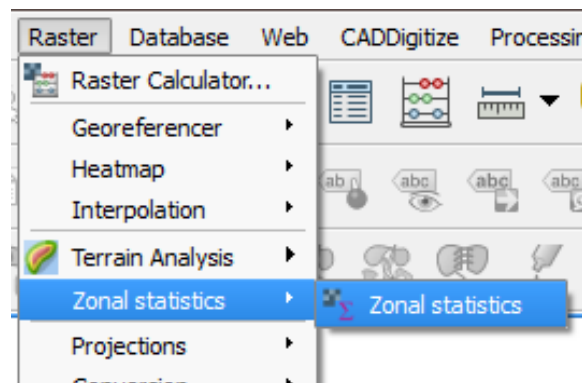
The raster dataset now presents a more accurate distribution of population counts across these two states using a single consistent value range (0 to 842 persons); the raster can now be fed into the zonal statistics tool.

Part 2 – Analyse raster population data using Zonal Statistics tool

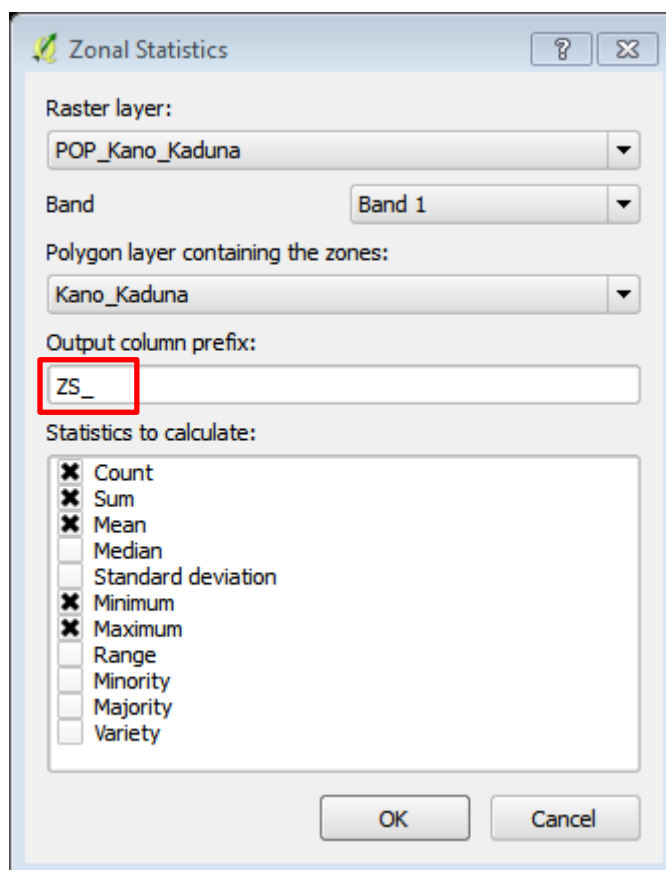
1. We will use the **Zonal statistics** tool to calculate mean, min and max pixel values per **Ward** administrative level. The tool will calculate statistics using the raster data and return the results into the attribute table of the ward boundary data
 - First of all, drag layer **POP_Kano_kaduna** into the group folder **POP GRIDS**
 - Then make the following dataset visible: **Kano_Kaduna**



- Go to drop-down menu **Raster > Zonal Statistics > Zonal Statistics**



- Complete the dialogue box as below
- Any text entered into the **Output column prefix** box will prefix the new fields which will be generated in the attribute table of the Kano_Kaduna Ward level boundary data. The suggested prefix **ZS_** will help to differentiate these calculated fields from others in the table



Zonal Statistics

Raster layer: POP_Kano_Kaduna

Band: Band 1

Polygon layer containing the zones: Kano_Kaduna

Output column prefix: ZS_

Statistics to calculate:

- ☒ Count
- ☒ Sum
- ☒ Mean
- ☐ Median
- ☐ Standard deviation
- ☒ Minimum
- ☒ Maximum
- ☐ Range
- ☐ Minority
- ☐ Majority
- ☐ Variety

OK Cancel

- We will accept the default statistics but feel free to select additional stats, if you wish
- Click **OK** to run the tool
- Open up the attribute table for **Kano_Kaduna** and clarify the calculations that have been performed:
- **Name:** unique ID for each grid square
- **ZS_count:** count of the number of pixels from the population data that fall within each ward
- **ZS_sum:** sum total of the population values for all pixels within each ward
- **ZS_mean:** mean of the population values for all pixels within each ward
- **ZS_min & ZS_max:** min and max individual pixel value within each ward

Kano_Kaduna :: Features total: 736, filtered: 736, selected: 0

	HAPE_Leng	SHAPE_Area	ZS_count	ZS_sum	ZS_mean	ZS_min	ZS_max
1			0.372364510304...	0	0	0	0
2	0096995436...	0.006288655305...	9048	12858	1.4210875331565	0	102
3	1300385994...	0.012101818660...	17427	10683	0.613014288173...	0	109
4	0683247551...	0.008322307569...	11989	7558	0.630411210276...	0	113

Show All Features

Part 3 – Thematic symbology

1. Symbolise the data to show the sum (or total) population for each ward
 - Right-click **Kano_Kaduna** and select **Properties**
 - Select **Style** tab
 - Complete the configuration as shown below; alternatively decide for yourself regarding colour ramp, number of classes and mode of value distribution

Layer Properties - Kano_Kaduna | Style

General | **Style** | Labels | Fields | Rendering | Display | Actions | Joins | Diagrams | Metadata | Variables | Legend

Style

Column: **1.2 ZS_sum**

Symbol: Change...

Legend Format: %1 - %2 Precision 0 Trim

Method: Color

Color ramp: **YlOrRd** Edit Invert

Classes Histogram

Symbol	Values	Legend
0.00 - 14705.00	0 - 14705	
14705.00 - 24130.00	14705 - 24130	
24130.00 - 34911.00	24130 - 34911	
34911.00 - 50044.00	34911 - 50044	
50044.00 - 70569.00	50044 - 70569	
70569.00 - 99577.00	70569 - 99577	
99577.00 - 146973.00	99577 - 146973	
146973.00 - 196797.00	146973 - 196797	
196797.00 - 293805.00	196797 - 293805	
293805.00 - 532049.00	293805 - 532049	

Mode: **Natural Breaks (Jenks)** Classes: **10**

Classify + Delete all Advanced

☒ Link class boundaries

Layer rendering

Layer transparency: 0

Layer blending mode: Normal

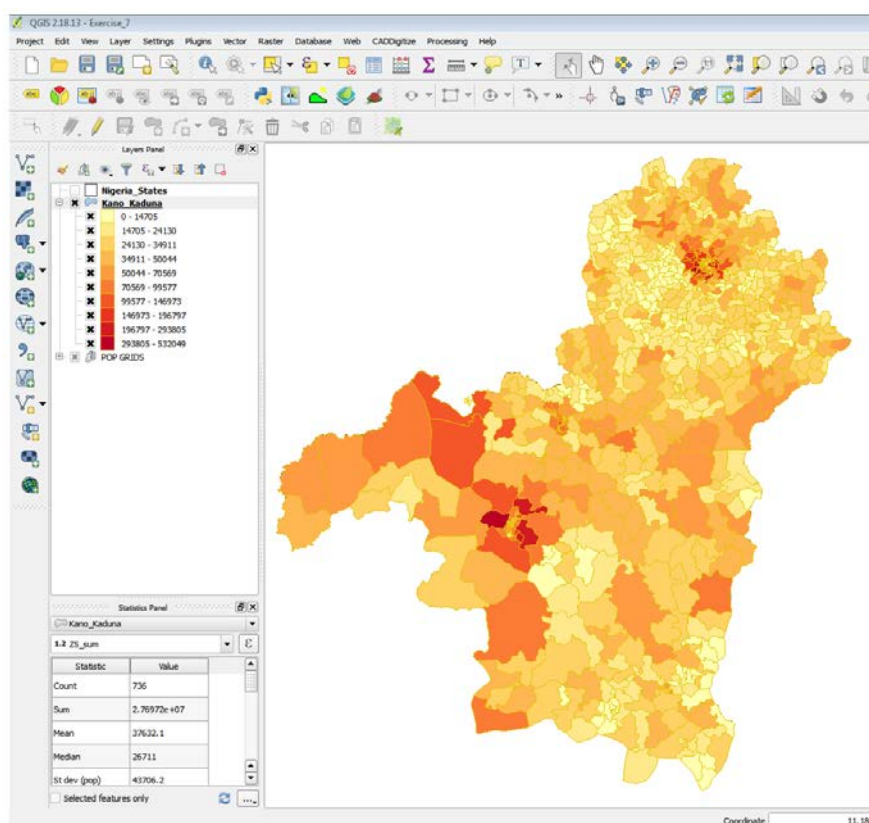
Feature blending mode: Normal

☐ Draw effects

☐ Control feature rendering order

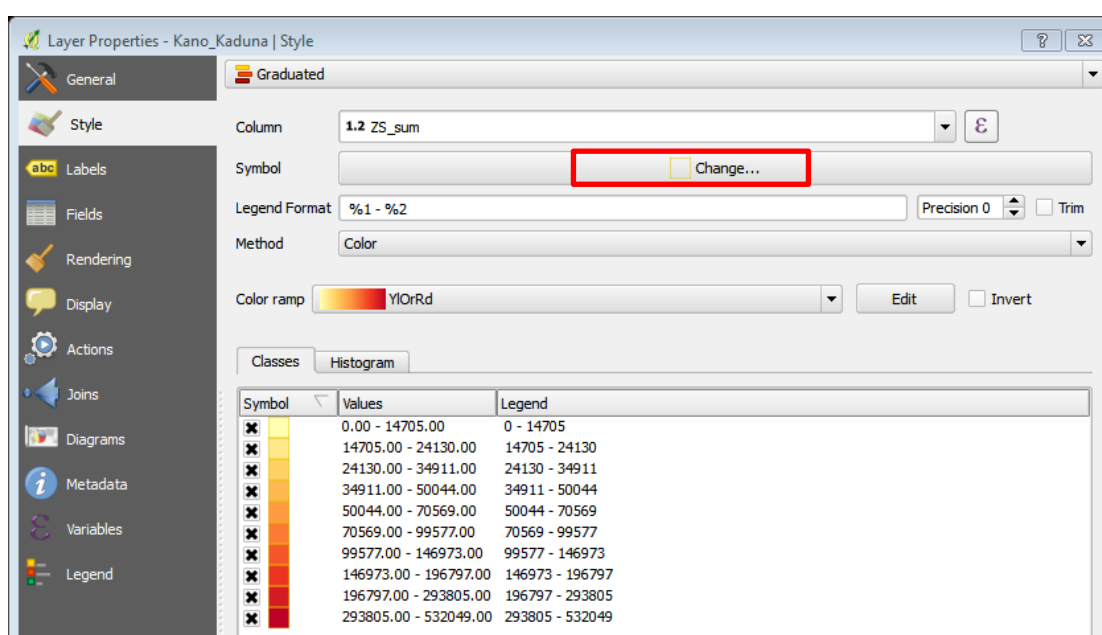
Style OK Cancel Apply Help

- The suggested settings should generate a map like this:

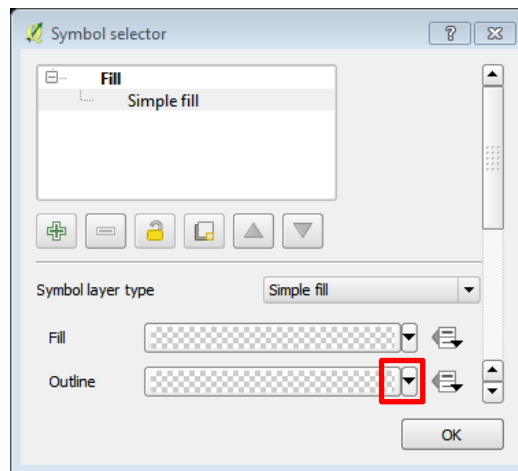


With thematic maps such as this, displaying polygons of relatively detailed spatial resolution, it is often useful to remove polygon borders to better illustrate the distribution, so we'll return to the **Style** tab.

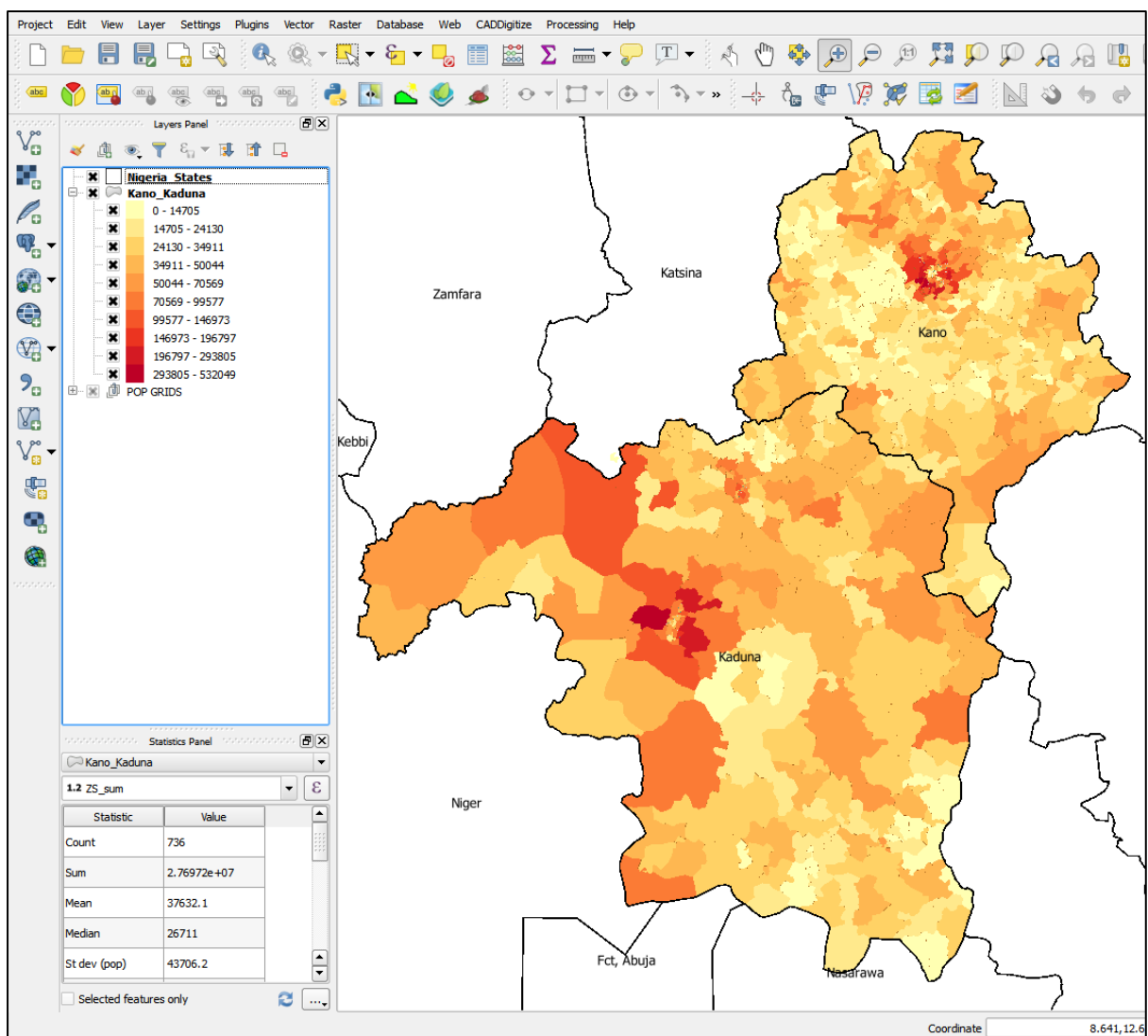
- Right-click **Kaduna_Kaduna** and select **Properties**
- Select **Style** tab
- Click the **Change** button, to the right of **Symbol**



- In the dialog box that appears, click **Simple fill**, to bring up **Fill** and **Outline** options
- Click the down arrow for the **Outline** drop-down menu and select **Transparent Border**
- Click **OK**



- This produces a much cleaner image, as below

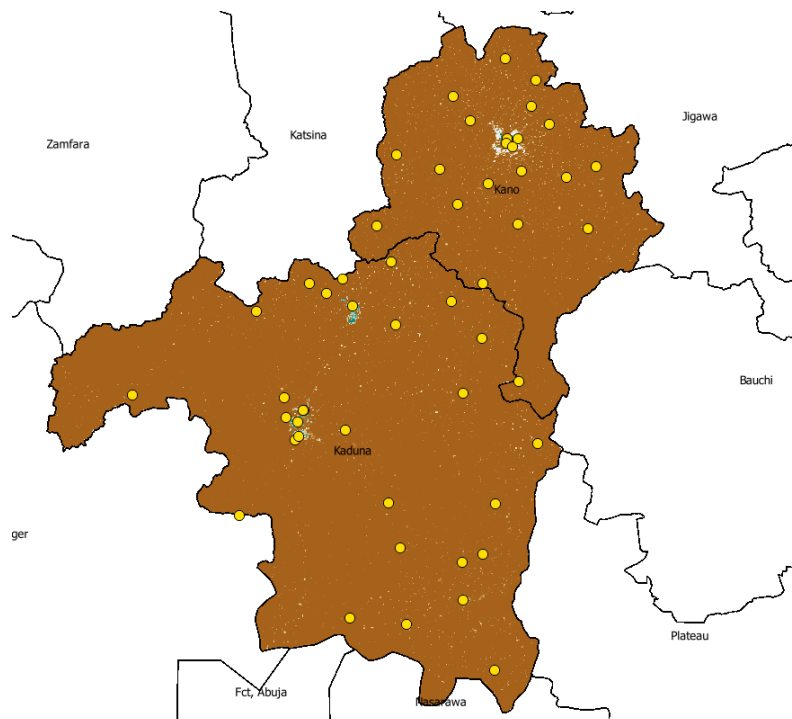


Part 4 - Taking it a step further...

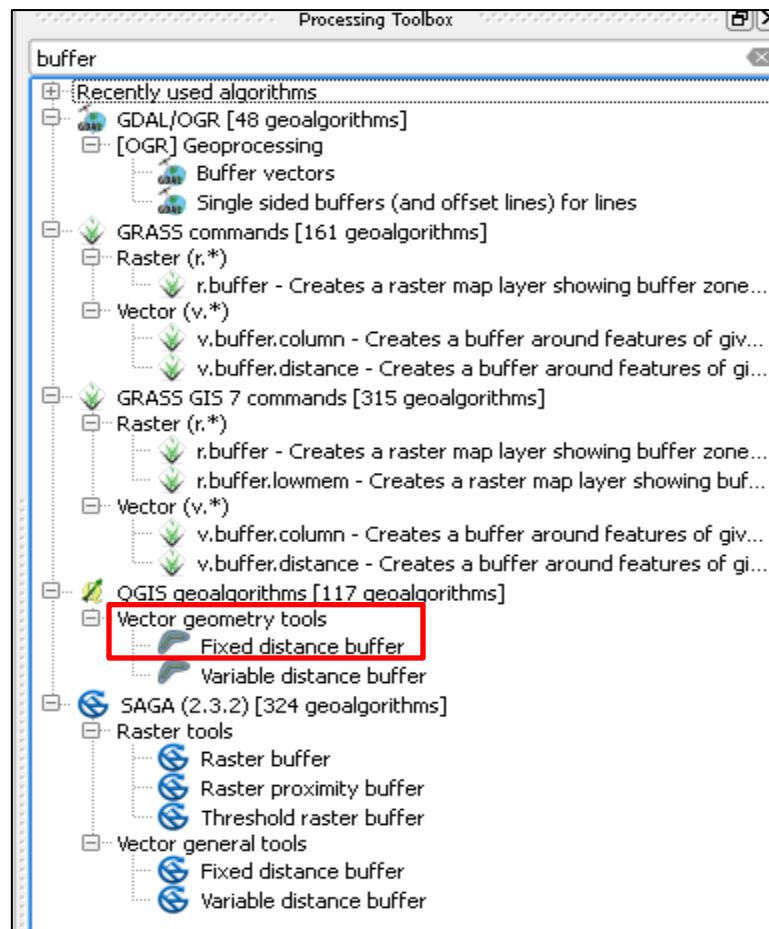
Zonal Statistics is a very useful tool but often you will want to define your own boundaries. We will now bring in some vector point data for hospital locations in Kano and Kaduna states. We will generate a buffer around each one and then apply zonal statistics to help us estimate catchment populations for each buffer.

1. Add vector point data

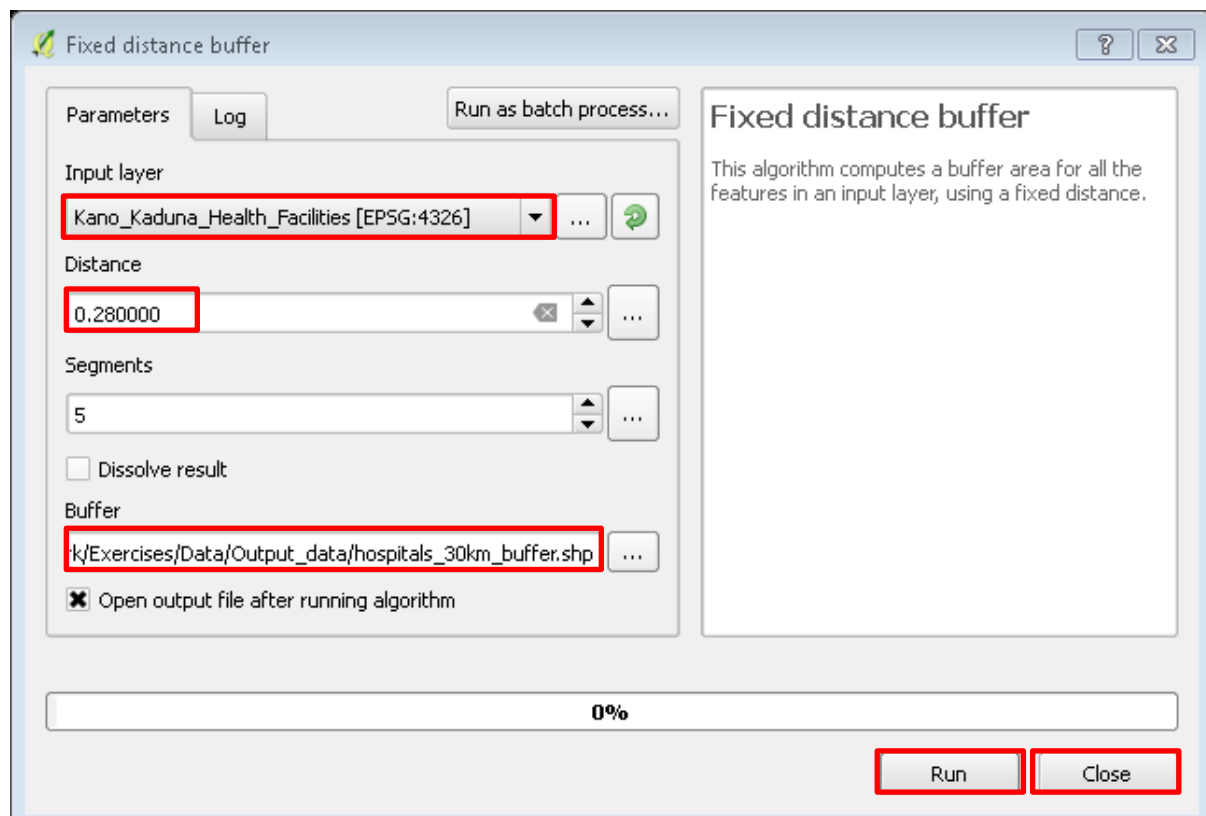
- Go to **Layer** drop-down menu > **Add Layer** > **Add Vector Layer**
- Browse to the following location and select these data:
`C:\Intro_Quantum_GIS\Exercises\Data\Vector\Nigeria\Kano_Kaduna_Health_Facilities.shp`



- The points represent main hospitals in the two states
- ### 2. Buffer the hospitals to approximately 30km. (Estimates for health facility catchment vary but 30km is a good measure for assuming a combination of walking and some form of motorised transport to the health facility.)
- Open up the Processing toolbox if it is not already present: Click on the **Processing** drop-down menu > **Toolbox**
 - Within the 'Search...' window of the toolbox, type in *Buffer*
 - From the filtered results, double-click **Fixed Distance Buffer**



- In the tool window that appears, complete the parameters as follows:



- Note that *0.28* is a measurement in angles (degrees of Latitude/Longitude) and equates to approximately 30km
- Select an output location for your buffer shapefile, as suggested
- Click **Run**
- Click **Close**
- Now inspect the hospital buffers which will have appeared in the layer list (note that it's display name will be simply *Buffer*)

Challenge

This polygon dataset is now the equivalent of the *Kano_Kaduna* admin areas layer which you used in the zonal statistics section earlier (part 2). Using what you have learnt, refer back to the procedure in Part 2 and generate zonal statistics for each hospital catchment.

- According to the population data, which hospital has the largest catchment population, and which has the smallest?