Tufts Data Lab

International Census Tutorial: Downloading, Joining & Mapping International Census Data

Updated 2/5/2018 by Ash Venkat and Carolyn Talmadge for 10.5.1

Table of Contents

Setting Up & Mapping Network Drives – Skip this session if you're in the Data Lab

1. If they are not already mapped, map out M, S, and H drives.

    Click on **File Explorer** → *This PC*. From the menu bar, select *Map Network Drive*.

2. Follow the directions on the [Tufts Data Lab Website for mapping Network Drives.](#)

## International Census Data Sources

International Data

1. The Tufts Data Lab website provides connections to many data sources, including international data sources. Start at [http://gis.tufts.edu.](http://gis.tufts.edu) Click on the **Data Sources** tab → **Online GIS Data Sources**.

2. Under **Geographic Region**, click **International.** Then click on the first link for **Census Dates for Countries and Areas of the World: 1945 to 2014** – this will take you to the US Census Bureau's resources on International census data.

    This page provides general international population and census data for review and download, as well as links to additional resources. Note when this site was last updated.

3. Click on the **Africa** under *Tables of census dates by region*. Here you can see the dates that censuses were conducted in each of the African countries listed. Find South Africa and determine when their most recent census was conducted.

4. Return to the [International GIS Data Sources page](#) on the Tufts GIS Website and select **United Nations Statistics Division, 2010 World Population and Housing Census Programme**. This is another site which provides details on the dates of international censuses.

5. Return to the International GIS Data Sources page on the Tufts GIS Website, and click on the Index of International Statistical Agencies maintained by the US Census Bureau.

6. We are going to look at **South Africa**, so find and click the link for South Africa's data in the list – this will take you to the "Statistics South Africa" page. To analyze census information in GIS, we need to gather both **geographic boundaries** and **tabular data**. We will start with tabular data.

Tabular Census Data

1. Unlike the geographic boundary files, **tabular census data** is available for direct download from the "Statistics South Africa" website.

2. On the Statistics South Africa Agency's homepage, scroll to the bottom of the page. Under **Tools**, click on **SuperWeb2.**

3. This directs you to a login page. Click on **Guest.**

4. On the left, expand *Statistics South Africa*. You will see several **Databases** listed. There is a ton of data within each folder to view and download. Take a few minutes to expand some of the folders and explore the available datasets. You will see that they have several datasets from multiple dates as well. This allows you to do things such as calculate change over time.

5. Expand *Community Profiles* → **Census 2011** →**Household Services**. Double click on **Household Services**. This now brings you to a new search showing all the available datasets related to household Services.

6. Click **once** on *Household Services_wd* . Now you see that there is information collected on everything from *refuse disposal* to the *toilet facilities.* Click **on the folder icon** for *Source of Water.* Notice that there are several very useful datasets within each category!
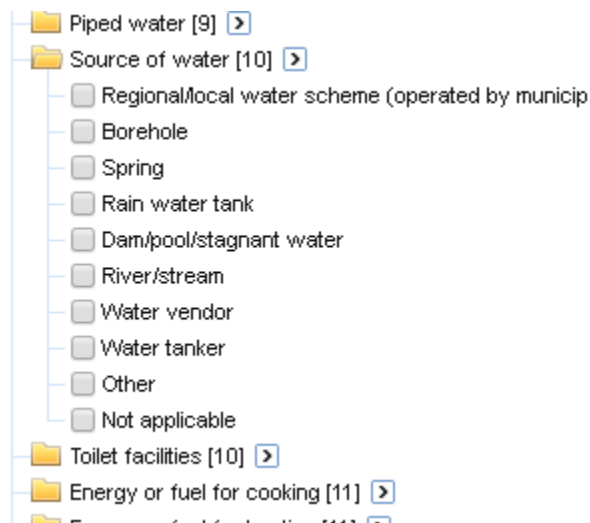

7. Click the back button once so you can view the *Community Profiles* folder again.

8. In the menu on the left, expand *Community Profiles* → *Census 2011* → *Family.*
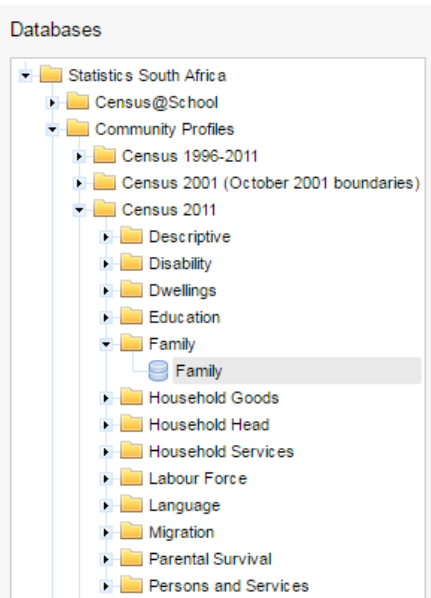
9. Double click on the *Family* database (⊜) symbol to bring up a new page again. Now, we will actually go through the process of building a table for download.

10. We will look at *monthly household incomes* at the <u>*municipality*</u> level. Click on **Family_wd** to expand it so that all the *folders* are visible.

11. Then click on the folder next to ***Geography.*** We need to first tell the system what <u>administrative boundary</u> we care about so we can get the data aggregated to the correct level. After clicking **Geography**, you'll see all the ***Provinces*** are visible. You could continue to expand the ***provinces*** so that the ***Districts*** are visible, and then so and so on to show the ***municipalities*** *and lastly the* ***wards.***
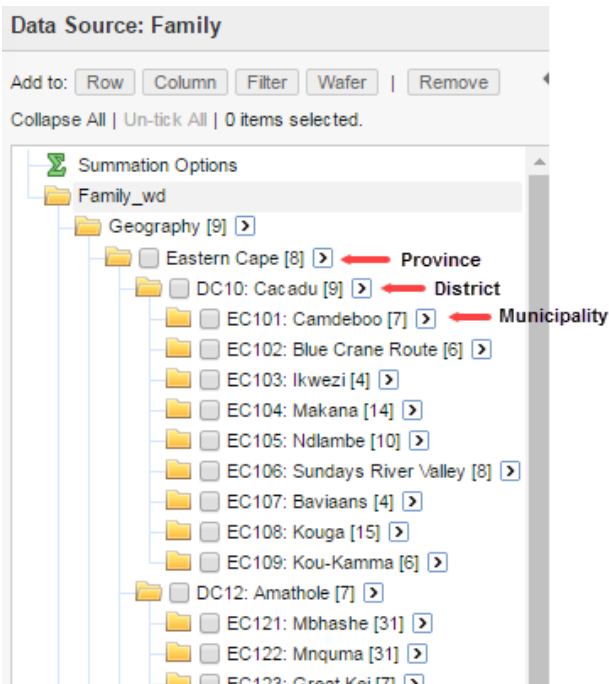
12. In this view, you can interactively select only the municipalities you care about. However, if you want to select **ALL** the municipalities in South Africa this could take quite some time.

13. A quick way to select all the **municipalities** within south Africa is to click on the carrot arrow next to **Geography** and then select **Municipal Boundaries.** This will check all the municipal boundaries without having to expand every single folder down to this level and manually checking them. Phew!
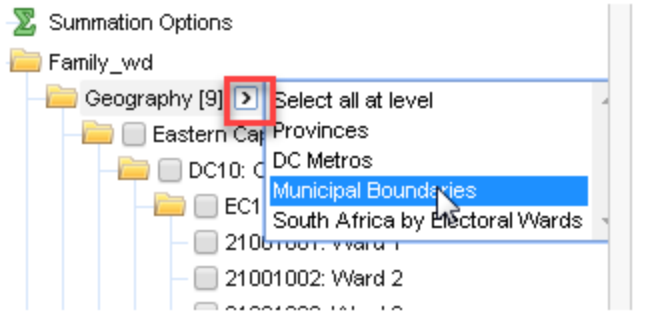
14. Now you will notice that only the **municipal** layers are checked. Much easier! Now, all the data we select will be aggregated to this level. Each row in the table will represent a municipality (just like an attribute table!), so click the **Row** button at the top left to populate the table with your municipalities.

    If you have made any errors, you can use the **Clear Table** button above the table you are creating.

15. Your geography selection should now be un-checked. If it is not, click *Un-tick All*. Next, minimize **Geography** by clicking on the folder*.*

16. Expand **Individual monthly income**. This shows all the different categories of income collected by the census. We want to add all of them to the table as columns. Click the arrow next to *Individual Monthly Income* and click on **P16_INCOME** so that all subcategories are selected.

17. Now, click the **Column** button in the upper left to add it to the table. Your table should now have geography data and income data and look very similarly to an Attribute Table. If you only cared about one income bracket, you could manually check just one and add it as a singular column.

18. The table is currently empty. In order to populate the table with data, you need to click the **Retrieve Data** button.

19. To save your table as an Excel file, ensure that it says "Excel 2007(.xlsx)" in the upper right hand corner. Then press **Go.**

20. Open the newly downloaded table in excel and examine it. What needs to happen for it to be joined in ArcMap? A lot of data cleaning!

Working with 2011 Tabular Census Data

1. For the remainder of this exercise we are going to use 2011 census data on household income, which has been previously downloaded from this website and saved in *S:\Classes\DHP_P207\Census Exercise Materials.*

2. Copy the **Census Exercise Materials** folder and paste it to your **H: drive**. In your **H drive**, right click on the folder and open the properties. Uncheck "read only". Click "OK' on the dialog box that appears as a result.

3. Open the file ***SAfrica_HH_income_2011.xlsx*** and familiarize yourself with the data. This is the data we just downloaded from the website, but already cleaned up significantly. What data is here? Is there metadata?

Preparing the Geographic Data

1. The **geographic census boundaries of South Africa** (the spatial data) necessary for this join are not posted on the Statistics South Africa website. This data was previously acquired by Tufts and saved on the ***M:*** drive. It took about three months to receive this data.

2. **Open a blank ArcMap.** In the **Catalog,** navigate to ***M:\Country\SouthAfrica\Census_AdminBoundaries_2011.mdb.*** Here you will find several data layers for South Africa, which were requested and received from "Statistics South Africa".

   Note: This silver cylinder is called a ***Geodatabase***. It is simply a *folder* for holding GIS data. The files inside the geodatabase are called **Feature Classes** and are similar to a **shapefile** with more advanced properties. **Feature Classes inside a Geodatabase** are newer and increasingly popular file types for spatial data. Geodatabases are helpful for keeping similar data in one location. The term shapefile and feature class are sometimes used interchangeably.
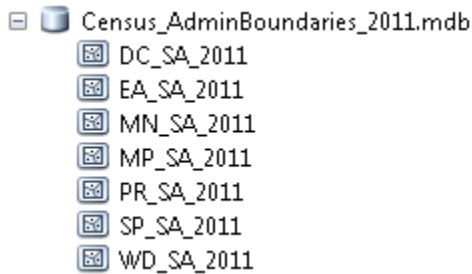
3. One by one, drag in all the files inside this **Census_AdminBoundaries_2011 Geodatabase** to add them to your map document.

4. In <u>Windows folder (This PC)</u>, navigate to ***M:\Country\SouthAfrica\Census2011*** and open the document called ***Census 2011 Spatial Metadata.doc***. This file contains information about the contents of the feature classes/shapefiles in this folder/geodatabase. Which **feature class/shapefile** will we be using for this join?

5. **Rename** each of the layers in the *table of contents* to match the full name of that administrative boundary that is provided in the metadata file. For example:

   DC_SA_2011 → District Councils 2011
   EA_SA_2011 → Enumeration Areas 2011
   MN_SA_2011 → Municipalities 2011
   etc…

Note: A quick way to rename layers and headings is to click the text you want to edit and press the **F2** button.

6. Once you've renamed all the layers, drag in the **cities_and_towns.shp** and the **main_rivers.shp** from *M:\Country\SouthAfrica.*

7. Take a few minutes to explore the data layers by turning on the map layers and viewing the *attribute tables*.

8. Use the attribute tables and data layers to determine the relative scales of the polygon layers. Which are the biggest administrative boundaries and which are the smallest? **Re-order** them in the *Table of Contents* so that they are arranged from the smallest to the largest geographic unit. Make sure your points and lines are at the TOP of your table of contents!

Joining Census Data to Geographic Boundaries

1. In ArcMap, turn on the ***Municipalities 2011*** layer in the *Table of Contents*. Open the Attribute table to view all the data contained in this layer.

   This layer contains codes, names, and identifiers for geographic municipal administrative boundary data. To attach the census data downloaded from the website, we will **join** the income data file TO this layer of municipality boundaries.

2. In order to **join** data in ArcMap, it must be *clean* and in a *specific format*. These rules hold for all joins in ArcMap!
   - All **titles**, **table footnotes** and **extraneous info** need to be deleted from the excel table. The bottom row needs to be deleted if it is a "total" row.
   - **Row 1** must be the column field names for the data you are importing.
   - Column field names must be **unique**.
   - Column field names cannot be **empty** and **cannot contain spaces**!
   - Column field names cannot contain strange characters such as **-,*, $, etc**. The ONLY character that is okay are **underscores** (_).
   - Column field names can be no longer than **11 characters**.
   - Make sure there are no **leading** or **trailing** spaces in your field names.
   - The first **8 rows** of data values will determine field type when brought into ArcGIS. If there is a mix of numbers and text in the data, the field type in ArcMap will be set as **TEXT**.

   Note: If you do not follow all these rules exactly, the join will fail!

3. Edit the open Excel file, *SAfrica_HH_Income_2011.xlsx,* per the rules above. A suggestion for the income table is to re-label the fields as follows:
   - Name the first column *muni_code*
   - *No income → no_inc*
   - *R1 – R400 → R400*
   - *R401 – R800 → R800*
   - *…*
   - *R204 801 or more → more*

Double check that ALL the above rules are followed, there are a few curveballs hidden!

4. Save your new excel file in your *H drive/Census Exercise Materials* folder as **SA_HH_Income2011_clean.xlsx**. Make sure it is an .xlsx. Change the file type.

5. In order to join the **excel data** to an **administrative boundary layer**, the files need to share a common unique identifier column/field. In looking at the **municipality layer** and the **excel tabular file**, what field matches exactly in both?

    Note: The column heading names can be different, but the contents need to be IDENTICAL. Write down these two field names!

6. We find that the **MN_MDB_C** field heading in the *Municipalities 2011* attribute table and the **muni_code** field heading in the *SA_HH_Income2011_clean.xls* match exactly.

7. Close the **excel file.** Excel cannot be open when you are trying to join it to ArcMap!!

8. Turn off all layers in ArcMap except for **Municipalities 2011** to speed up processing. Having a lot of layers always on can really slow things down!

9. In **Catalog**, navigate to the *Census Exercise Materials* folder in your **H drive**. Find your new **SA_HH_Income2011_clean.xlsx** and double click on it to show the **sheet1$** within the excel file. (If it doesn't show up right-click on **Census Exercise Materials** and click **Refresh**.)
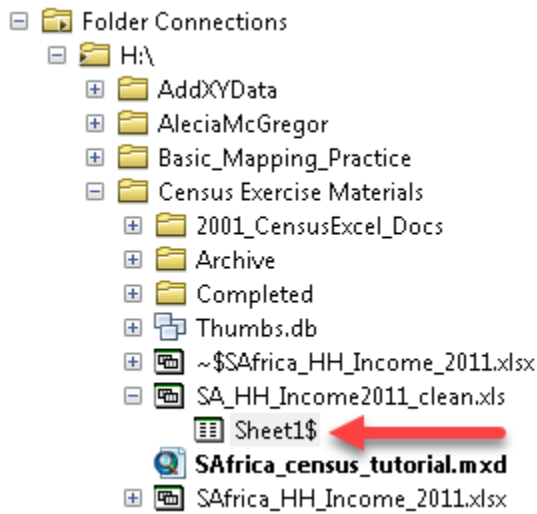
10. Pull the **Sheet1$** into ArcMap like you would a shapefile/feature class.

    Note: If you try to pull in the whole workbook, it will give you an error that says "Could not add the specified data object to the map." Whenever you see a + sign, it means you must go a layer deeper!

11. Now, you should see your excel table listed in the *Table of Contents.* You are viewing this in the *table contents* because the view changed to **list by source** when you added the excel table. Note that you can no longer move your layers around, but you see the path for each folder and where each layer is saved. This is helpful if you ever forget where you saved a shapefile!
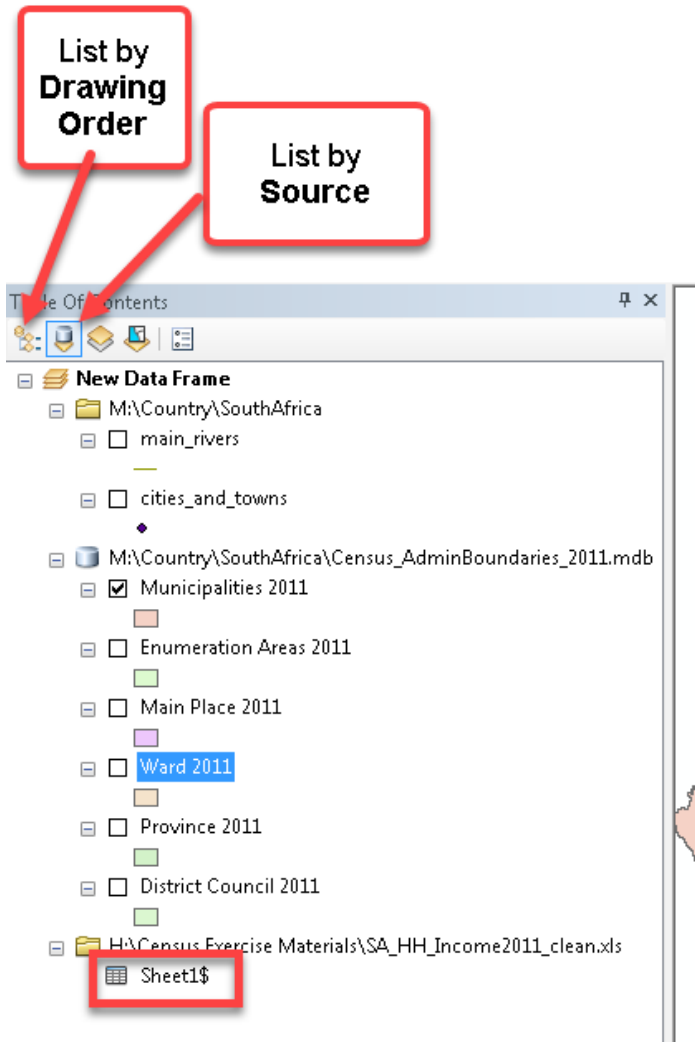
12. Right click on your **Sheet1$** and press **Open**. Make sure everything imported correctly. Note that it now looks very similar to an attribute table!

13. Switch your *table of contents* view back to **list by drawing order** as shown in the above graphic. Your excel table no longer shows up because it is not "drawn" in ArcMap, but it is there! Now you can move your layers around again by dragging and dropping.

14. Right-click on *Municipalities 2011*. Select **Joins and Relates → Join...**
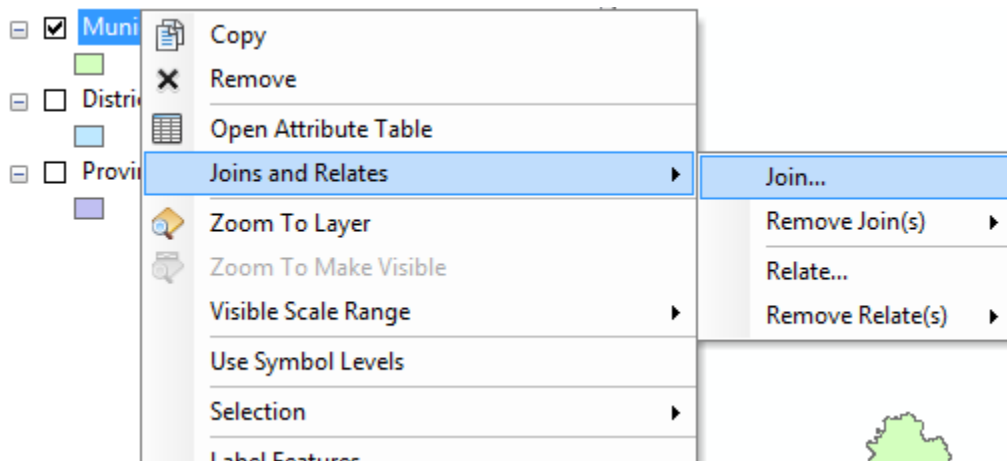
15. Fill out the **Join Data** window as seen below. Make sure that you have "*join attributes from a table*" selected. Note you can also join layers based on a spatial location – that is not what we want to do now.
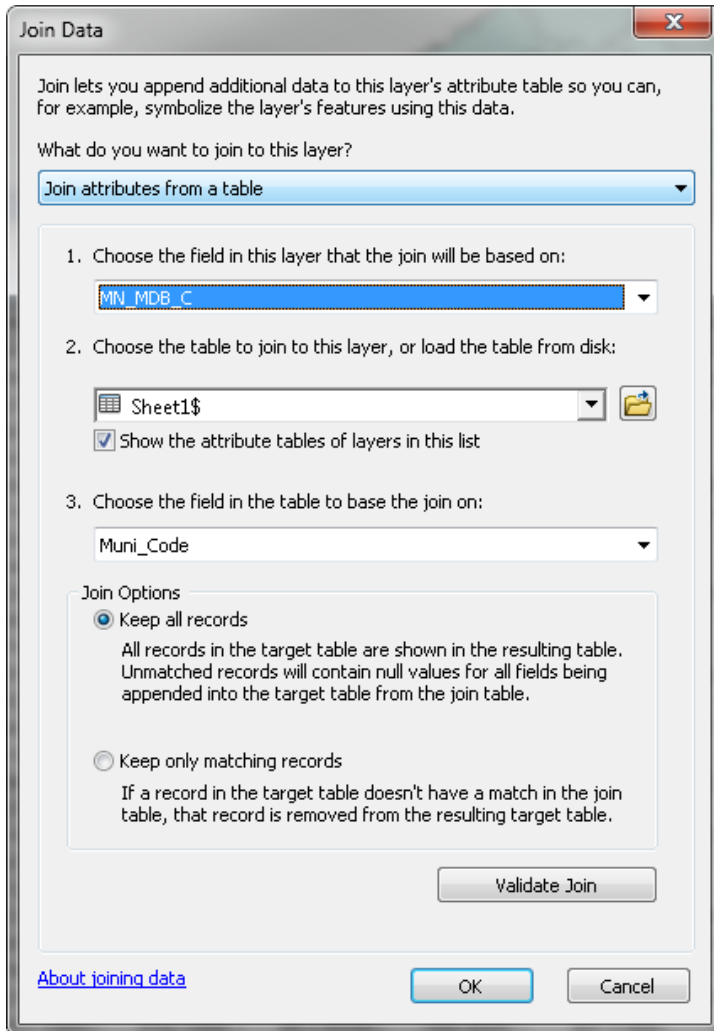
(1) This field is the unique ID from the **feature class/shapefile attribute table**!

(2) The excel file Sheet1$ is already selected because it is the only excel table in our ArcMap session. If we had many excel sheets, we would need to select the correct one.

(3) This is the unique ID in the **excel file** that matches the shapefile!

(4) Select "*Keep All Records*" and click *OK*.

16. It doesn't seem like anything has happened, but open the *attribute table* for **Municipalities 2011** and scroll to the right**.** Now, all of our tabular data has been appended to the end of the municipalities attribute table.

17. As it is right now, ArcMap is connecting the income data from the excel file, but it is **not permanent**! To **save** the join and make the excel data remain permanently in the *Attribute Table*, right-click on the **Municipalities 2011** layer and click **Data → Export Data**. This saves the feature class as a NEW feature class/shapefile with all the joined data as the attribute table.

18. In the window that pops up, select *this layer's source data* to save the coordinate system information. Click on the folder icon and navigate to your **H:** drive and then the Census Exercise Materials.

19. We are going to save this layer with the join as its own feature class. However, to do so, we need to first create a **Geodatabase** to hold the feature class**.** To do so, **click on the geodatabase icon** in the top right.
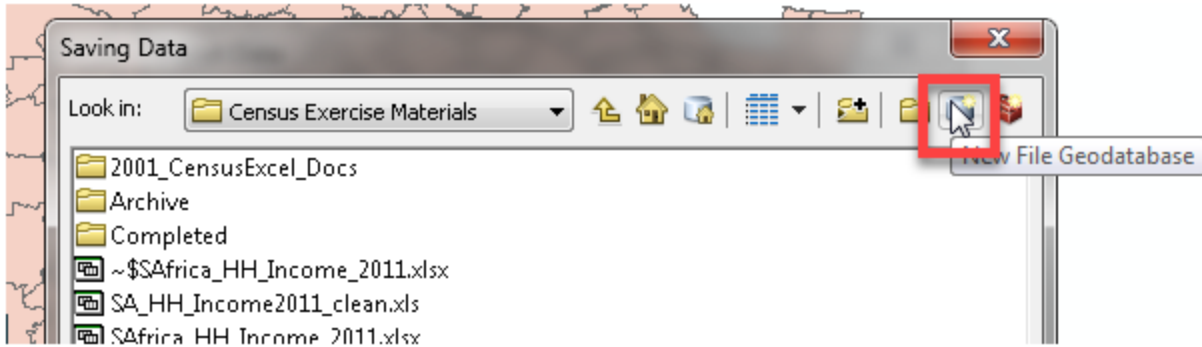
20. Name the Geodatabase **Census_2011.gdb.**
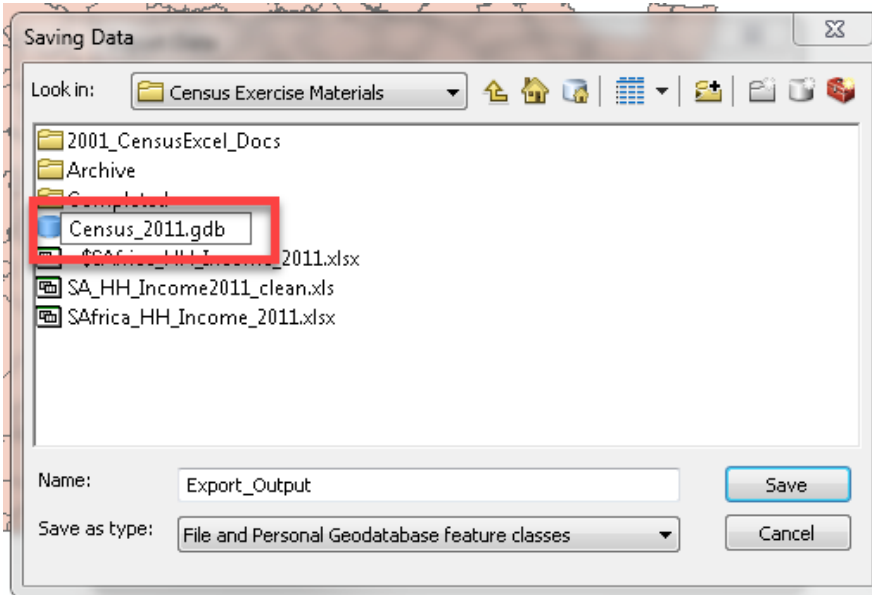
21. Double click on the new geodatabase to open it, just like a folder. Now, save the feature class as ***Municipalities_2011_IncomeJoin.*** Make sure the Save as type is **File and Personal Geodatabase feature class**. Click *Save* and then *OK.* It is important to include both the boundaries, the year and the data joined in the name of the feature class!
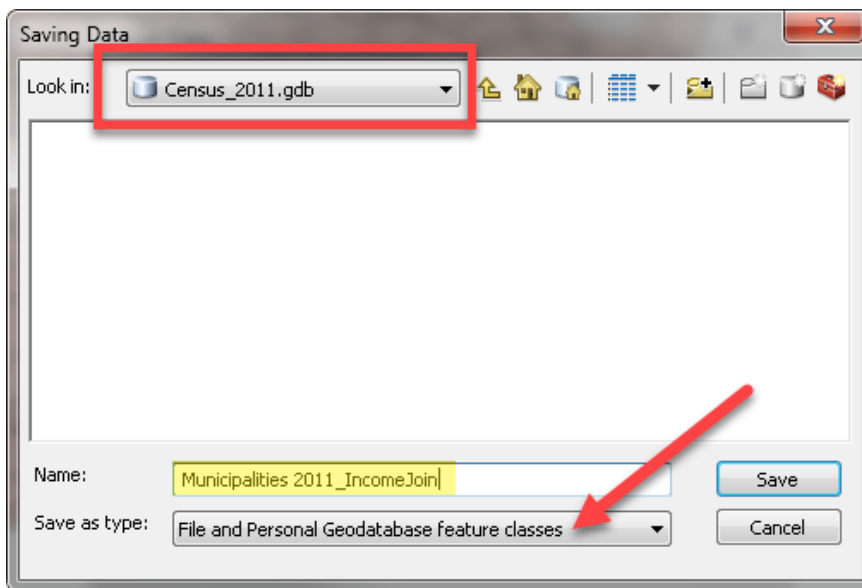
22. When asked if you want to add the new layer to the map, select *Yes* and ***Municipalities 2011_IncomeJoin*** will appear in your *Table of Contents* AND in *Catalog.*

Adding a Field and using Field Calculator

The income data that we have only provides the **number of households** in *each income category*. To make the mapping analysis more useful, we need to know the **total number of households in each municipality** so that we can *normalize* the data. Normalizing the data turns it into a percentage, so we can compare the municipalities.

1. In order to determine the total number of households in each municipality, we need to add all the households from each income category together! However, before we can do that, we ne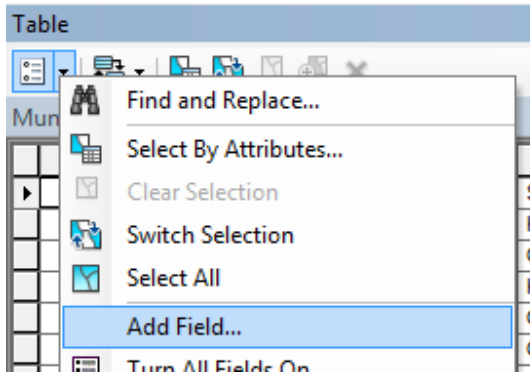ed to create a field in the *attribute table* for this calculation! Open the attribute table of your new shapefile, **Municipalities 2011_IncomeJoin** and click the *Table Options Dropdown Menu* in the upper left of the table. Then click **Add Field…**

2. In the **Add Field** Window, name the new field **Tot_HH** (a.k.a Total Households) and select **Long Integer.** **Note: Long Integer** is a *numeric* field with *no decimal places*. **Double** is a *numerical field* as well but allows for multiple decimal places! Click *OK.*

3. Scroll all the way to the right of your *attribute table* and you will see the new field with the heading **Tot_HH**. All the numbers are currently 0. We need to create an equation to add all the fields together.

4. Right-click on this new *field* and select **Field Calculator**. Click yes if a warning appears. This window allows us to *calculate* values based on the other fields already populated in the *attribute table*.

5. Double-click on the attribute table **fields** in the top left box to add them to the **calculation query box** that says **Tot_HH =**. Press the *plus sign icon* between clicking each field to sum them together. Make sure to include the fields for *no_inc, more, unspecified, and NA* .

   Note: It is important that you use the *buttons* rather than typing in the equation because the query box is VERY sensitive to spaces and formatting. Clicking the buttons will sure everything is entered in properly.
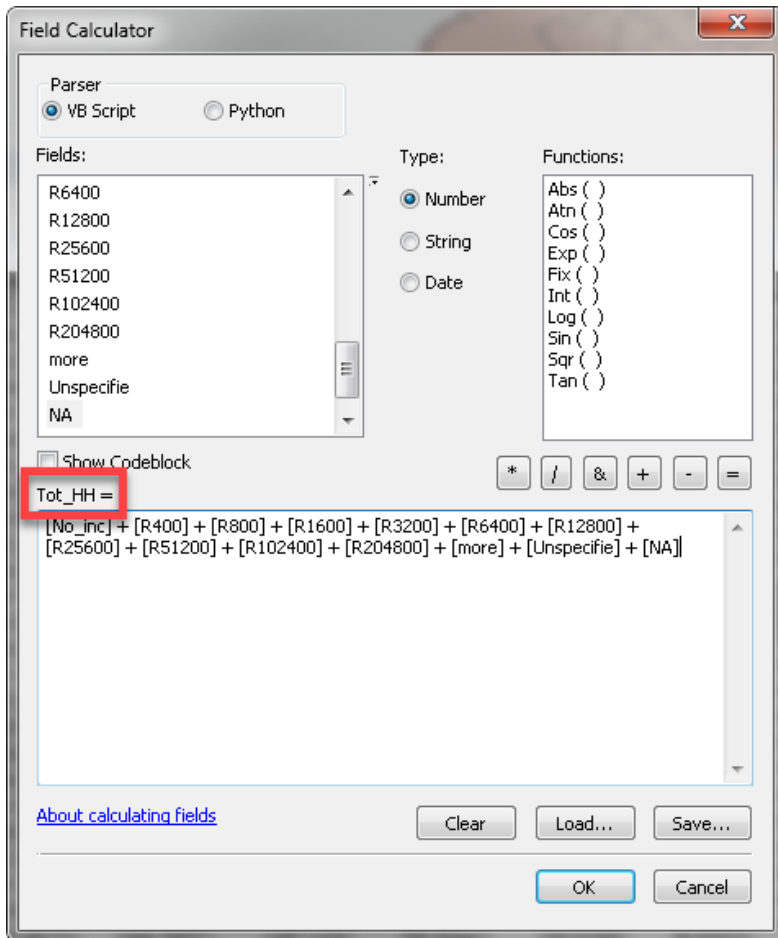
6. Click *OK.* The window will close and the **Tot_HH** column will populate with the sums of the households for each municipality.

   Note: The field calculator is a very powerful tool, which you will use often to make percentages (normalizing your data) and to calculate change.

Defining Symbology and Normalizing your Data

1. Right-click on **Municipalities 2011_IncomeJoin** and select *Properties*. Navigate to the **Symbology** tab.

2. We want to map the ***percentage*** of all high-income households per each municipality. Income is a quantity field, not a category. Therefore, select **Quantities → Graduated colors**.

3. Under the *Fields* section, in the **Value:** dropdown, select "*More*". Remember, this is the field for "R204801 or More" that we shortened to "more" in excel. This represents the highest bracket of incomes.

4. In order to make this a ***percentage*** instead of just a *raw count of households with high incomes*, we need to ***normalize*** the data by the *total number of households*. In the **Normalization:** dropdown, select **Tot_HH**. This will allow us to create a map where each census municipality is colored light to dark according to the percent of households with incomes greater than R 204,800 per month.

5. Notice how when we normalized the data, the numbers changed to decimals. We need this decimal to represent a *percentage.* Click on **Label** and select **Format Labels…**
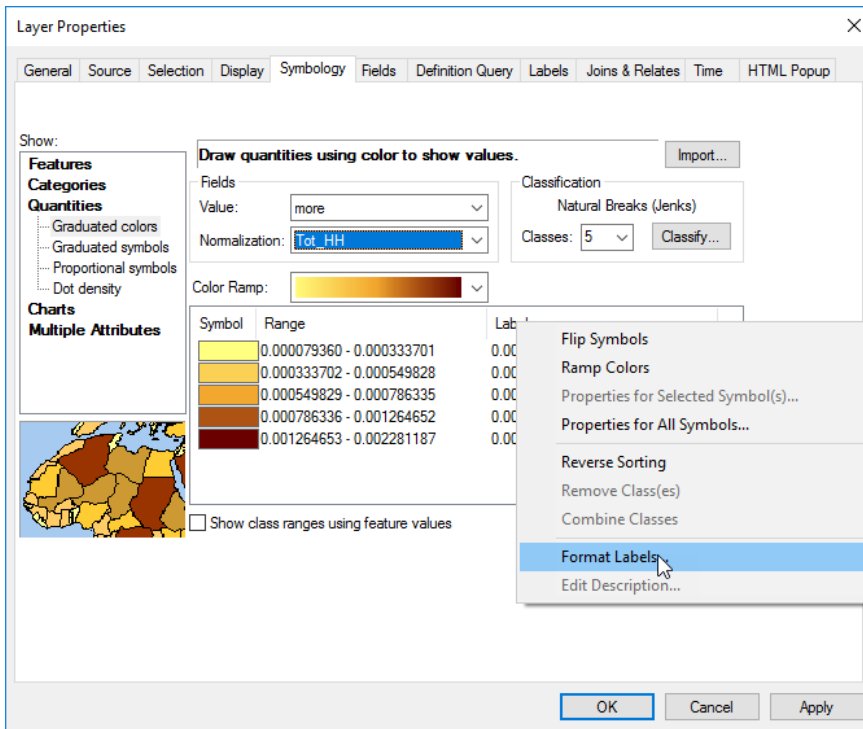
6. Select **Percentage** on the left hand side. Then, choose "the number represents a fraction. Adjust it to show a percentage".

7. Now Press **Numeric Options...**. Decrease the *Number of decimal places* to 2. Press Ok and then press ok again. Now our data is shown as a percentage instead of a fraction!

8. Click *Apply* and the map updates! Now pick a gradual color scheme that you like (light to dark) and press OK.

   Note: We could have also mapped the same variable by **adding a new field** to our attribute table and using **field calculator** to make an equation "(More/Tot_HH) *100". This would have given us a new field with the percentage of high-income households, which we would not have to "normalize" in symbology because it is already a percentage in our attribute table! There are always many ways to do the same thing in ArcMap!

9. In the table of Contents, rename the *Layer Name* to be **Municipalities** and rename the *Heading Name* to read **High Income Households (%)**. Additionally, recall the lecture on *Cartography.* To eliminate clutter, remove the first percent sign from each number group. This makes it cleaner without duplicate info.

   Note: A quick way to rename layers and headings is to click the text you want to edit and press the **F2** button. Now we can clearly see what is being mapped AND when we make our legend, our layer will already be properly named! Good data organization and naming practices is VERY important!
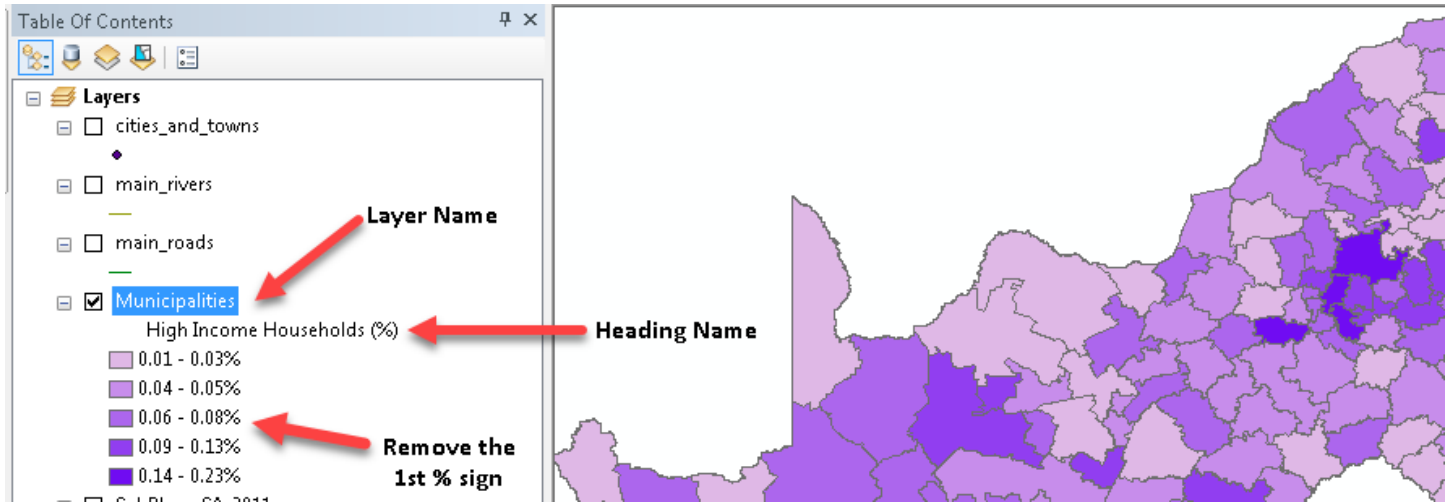
Additional Census Data Exploration

1. Now that you have gone through the process of finding and mapping one category of census data, let's try it with a new excel table. A second table of census data called **SAfrica_InternetAccess_2011.xls** has been downloaded from the Stats SA website and placed in the folder you copied to your H: drive.

2. Open this file. What formatting needs to be done to join this file to the ***Municipalities_SA_2011*** feature class?

3. Formatting this data is more complex than the income data because the codes and names of the municipalities are in the same field. This means that they do NOT match exactly with the feature class attribute table codes. We need them to be separated, which will then match the code field in the feature class. Moreover, much reorganizing needs to occur, following the rules mentioned in the earlier section.

4. Open the **SAfrica_InternetAccess_2011_Clean.xlsx** file to review a version of the file formatted properly for uploading to ArcMap.

    o   Which field name matches the field in the attribute table?

    o   What is the name of the sheet?

    o   Notice how the headings have no spaces or characters and have been shortened. All headers and footers have been removed.

5. Close the excel file.

6. In Catalog, navigate to the **SAfrica_InternetAccess_2011_Clean.xlsx** in your H drive/Census Exercise Folder**.** Double click on the workbook to show the **sheets.** Drag in **InternetAcess$** to your ArcMap session. If a popup window says "Could not add the specified data object to the map" then save the file as .xls, close Excel, and try again using the .xls file.

7. Now, we need to remove the first join, which is still attached to the *original* **Municipalities 2011** layer. In ArcMap, right click on **Municipalities 2011** → **Joins and Relates** →**Remove Joins** → **sheet1$.** We no longer need this join since we exported it into its own new feature class!

8. Follow the previously outlined steps to join and map the "***Internet Access***" data to **Municipalities 2011** in a similar manner to the Income data. **Make sure to switch the sheet to InternetAccess$ in the Join Tool.**

9. **Open the attribute table and confirm that the join worked.** You should now see info on where people have access to the internet in South Africa. If it looks good, right click **on Muncipalities 2011 → Data → Export Data.**

10. Click the folder Icon to pick where you are going to save the data. It will likely open up into our previously made **Geodatabase**. If it does not, navigate to *H:/Census Exercise Materials/Census_2011.gdb.*

11. Name the new feature class **Municipalities_InternetAccess_2011**. Make sure it is saving as a **File and Personal Geodatabase feature class.** Press Save and then Ok. Then Yes when asked to add it to the map.

12. In your Catalog, you should now see that you have two feature classes inside the **Census_2011.gdb**. If you do not see it, you can right click on Census Exercise Materials and **Refresh**. You can bring these new feature classes/shapefiles into any arcmap session in the future. If you did NOT export the joins, you would have to redo them next time. This is why exporting is so important!

13. Use symbology to display one of the variables in this attribute table. Make sure to normalize it by the total households. Also, use the information you've learned to format the numbers and percentages.

Map Layout

1. Now that we have the data that we need in the map, we can create the layout of our map. Choose **View → Layout View** from the main menu.

2. Arrange the **data frame** on the map so it fills the whole page and zoom in to South Africa so it fills up the page as much as possible.

3. You will add a *Title, Cartographer Info, Data Sources, Legend, North Arrow* and *Scale Bar* to your map. These are all required elements on a map. You access them by going to the **Insert** item on the *main menu* bar. See the example map below for layout ideas and to check you have all important map elements.

4. If you want to add background data (such as countries or capitals), you can find this data in M/Country/World/ESRIDataMaps10/world/data. This helps give more context to your map.

5. To save an image of your map, go to **File → Export Map.** Pick where you want to save it. Then you can change the save as type to a jpeg, pdf or GIF. Jpegs and Gifs are image which can be inserted into a word document – pdfs cannot be inserted, but you can print pdfs directly.

See the chart below for some details on adding map elements.

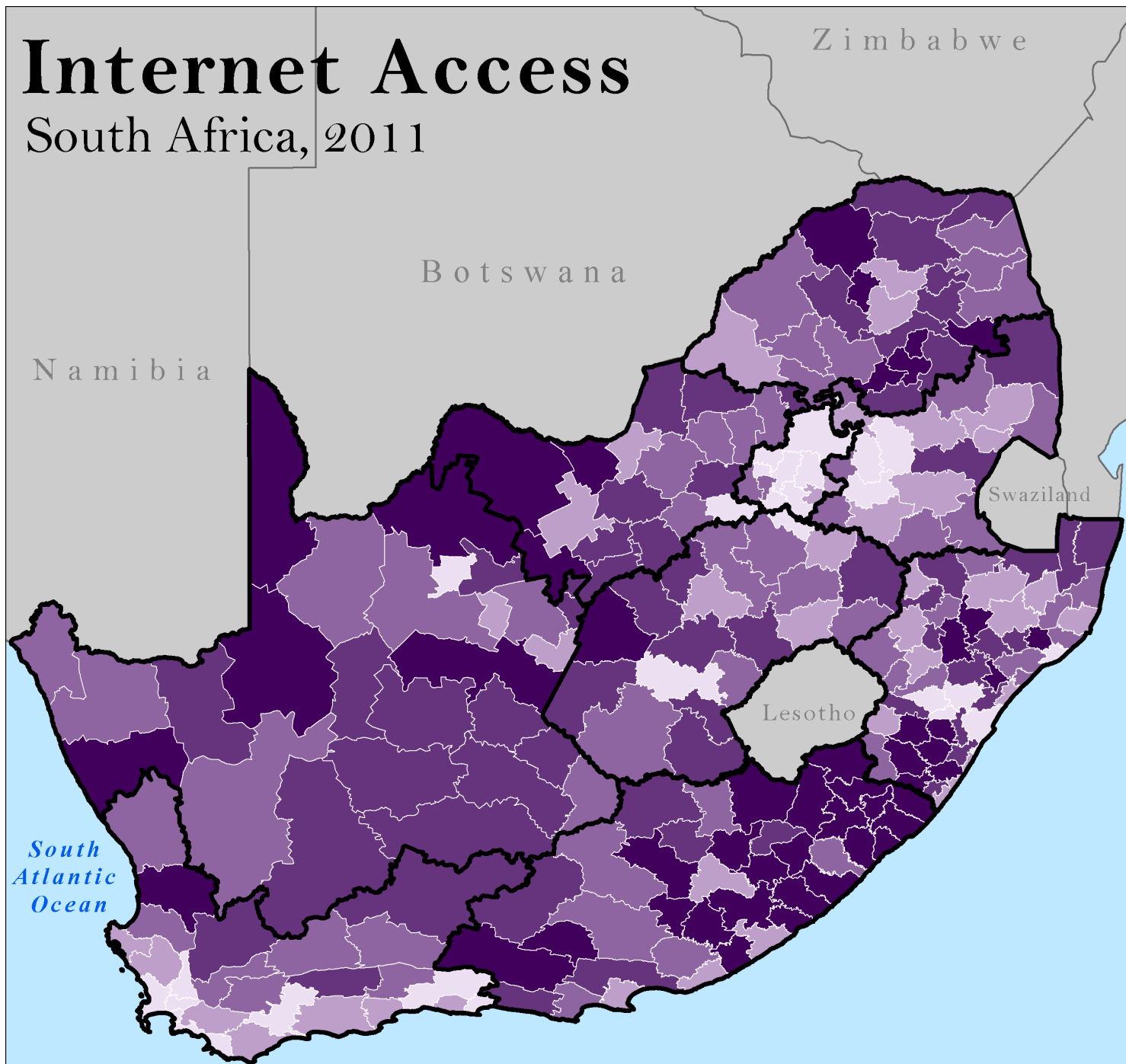|	
Title	Your title should not be too long, but should include the variable you're mapping, the location and the year of the data. Using 2 different size fonts is a nice way to add some visual hierarchy.
Cartographer Information & Data Sources	Have text that includes your name, date, class and where the data was downloaded from. Where did we get this data?
Scale Bar	If you want a different units, double click on the scalebar to get to it's *Properties.* Then change the *Division Units.* You can also change fonts with *Symbol* Make sure the SB ends on a easy to understand number (aka 100, not 142 miles).
Legend	Experiment with changing the number of columns so that the legend box fits well on your map. **You do want to identify the variable you are mapping, the units AND the <u>administrative boundary</u>. It is important to know if it is by Municipality or Province.** **You do not need the legend to say "Legend"…it is obvious.**
Label	Labels really help give more location context. Notice how the example map has **country labels AND water labels**. You can add labels using *Insert Text* or go up to

	Customize→ Toolbars→ Draw. The Draw toolbar is really helpful for adding quick text and changing fonts, colors, etc.
Modifying an element after insertion	You can usually modify an element after you have inserted it by selecting it and then double-clicking on it. The relevant dialog box will appear for that element

The most important thing is to **not just accept defaults**! Take time selecting your colors. Do they make sense for what you're mapping? **Pick your fonts – don't just use Arial**! Think about where you are placing your map elements.

Map making is as much of an art as it is a science.
Have fun with it! And check out the example below for some inspiration!

Internet Access
South Africa, 2011

Z i m b a b w e

B o t s w a n a

N a m i b i a

Swaziland

South Atlantic Ocean

Lesotho

I n d i a n   O c e a n

Percent of Households without Internet

By Municipality

48.6 – 63.1%

63.2 – 70%

70.1 – 75.3%

75.4 – 80.5%

80.6 – 88.8%

Provinces

Carolyn Talmadge
February 14, 2017
Data Sources: Statistics South Africa
Map Projection: Africa Albers Equal Area

0   100   200
km