

## Intro to GIS

Comp. Sci. Workshop 2020

Gwendolyn Lloyd

### Food Deserts and Diabetes in Florida

Task: Model diabetes prevalence and number of grocery stores in Florida and determine a relationship between the two

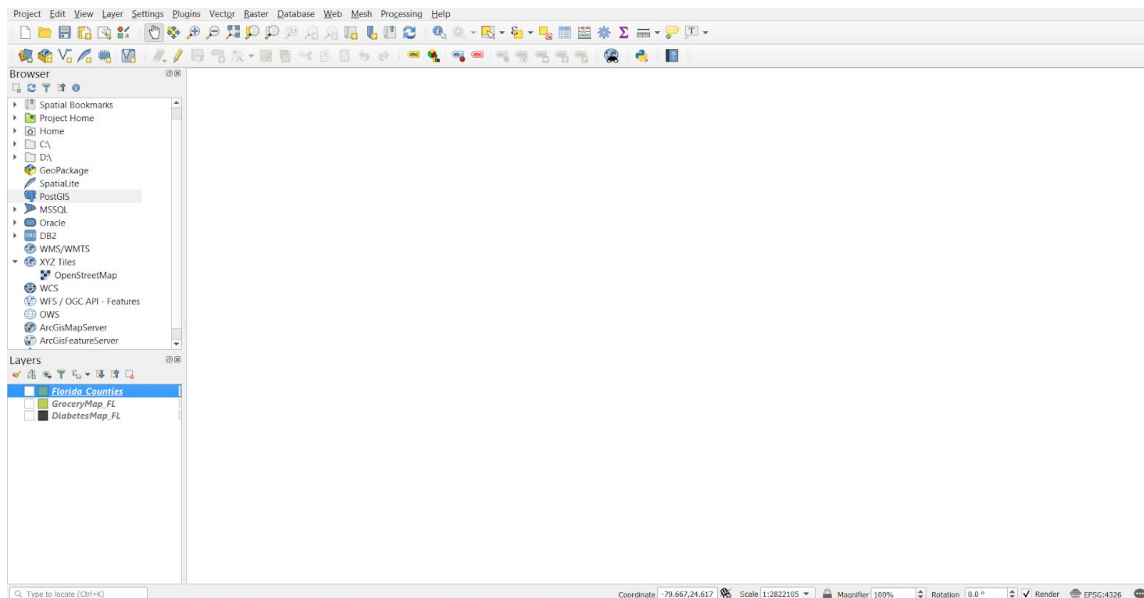
Background: Food deserts are areas with limited access to affordable and nutritional food due to an absence of supermarkets. This is a major problem across the United States affecting rural and impoverished communities. Due to lack of access, food options in these areas are generally fast food or equivalent nutritionally, resulting in increased diabetes and obesity in these areas and populations. For this activity, Florida was chosen as the state is highly variable, as it has a gradient from densely populated urban areas to rural areas where food deserts are prevalent.

For more background:

<https://www.pbs.org/newshour/arts/the-socio-economic-significance-of-food-deserts>

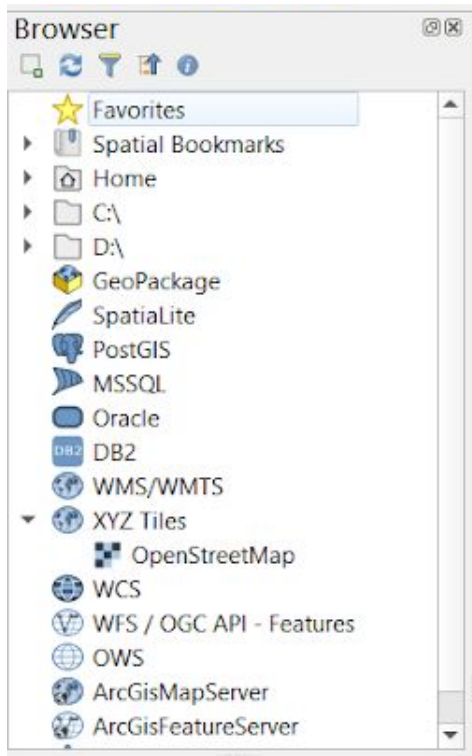
#### Part 1: Understanding the Data and Software

1. Open the file *Florida\_Day1.qgs* in the Day 1 folder. This will be where we will do all our analyses and make our maps. Your screen should look like this:

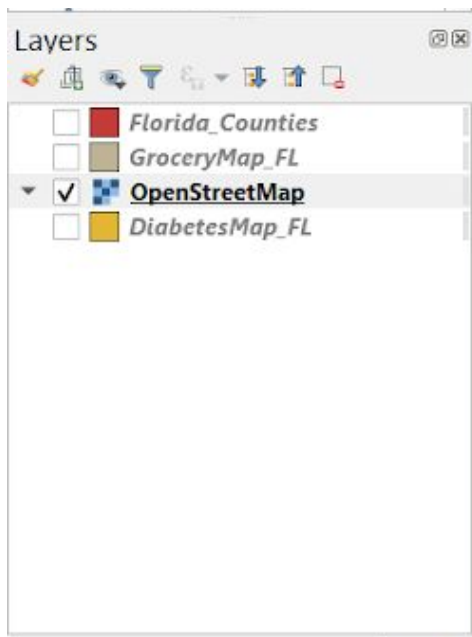


2. The data we are using today will be using data from Florida. On the left side of the screen there are two menus: Browser and Layers. We will be using the Layers menu for most of today, as it contains the data we are working with.

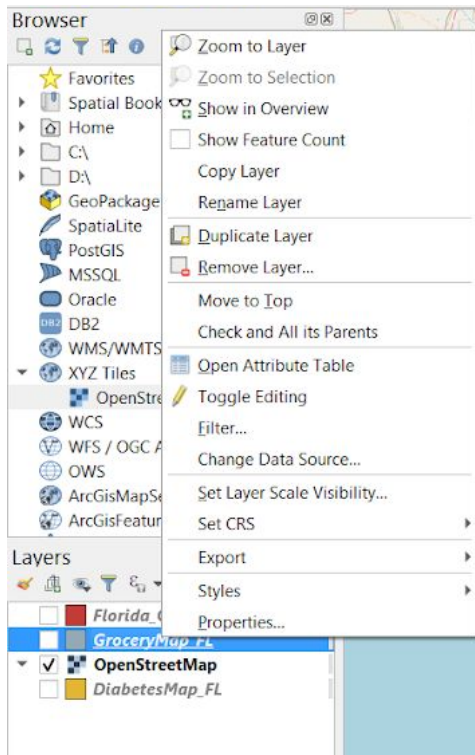
In the **Browser** menu, open the dropdown menu *XYZ Tiles* and double click to open *OpenStreetMap*. This should add a map of Florida that looks like what you would find on Google Maps.



This is also added into the Layers menu now, with a checkmark next to it. You can toggle showing this map and other maps by clicking on the checkmark.



- If you click the other layers, you might notice they all look the same!  
Our data is made up of **shapefiles**, or geographic features. Each shapefile has an **attribute table** where we can see the data within the shapefile.  
Open the attribute table for one of the shapefiles by right clicking on it and selecting *Open Attribute Table*.



A **metadata** file is included that explains the variables included in the attribute table.  
For our data, we only have one metadata file: **GroceryMap\_md.pdf**

- From the metadata (found in the data folder: IntroGIS → IntroGIS\_Day1\_2020 → Data) and attribute tables figure out which variables we are interested in:
  - The number of grocery stores per 1,000 people from the most recent year
  - Percentage of Diabetes cases

Variable	Shapefile	Information it contains

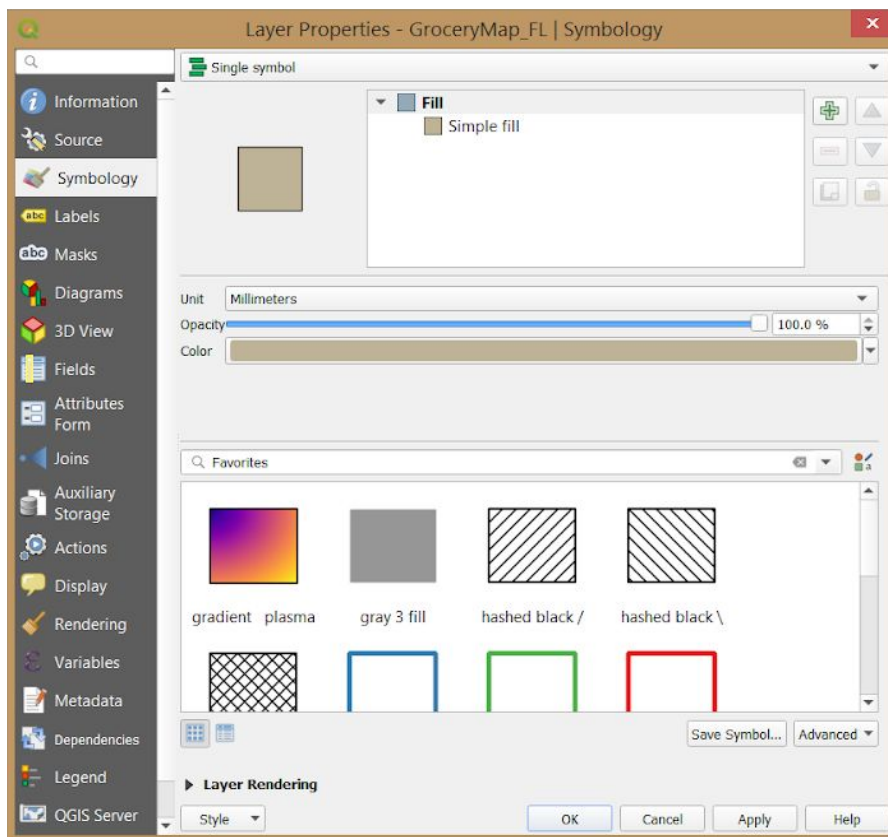
5. We will later want to combine these two shapefiles to draw conclusions about the relationship between these two variables. What variable do these two shapefiles have in common?

### Part 2: Visualizing your data

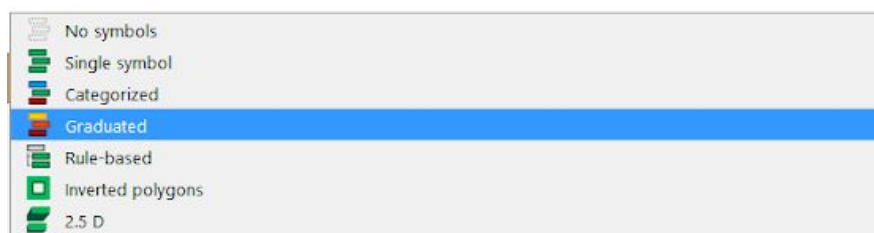
As you learned above, our data contains more information that it is currently displaying. We can change the way our shapefiles look by changing the **symbolology** and **style**.

Let's start with our shapefile *GroceryMap\_FL*:

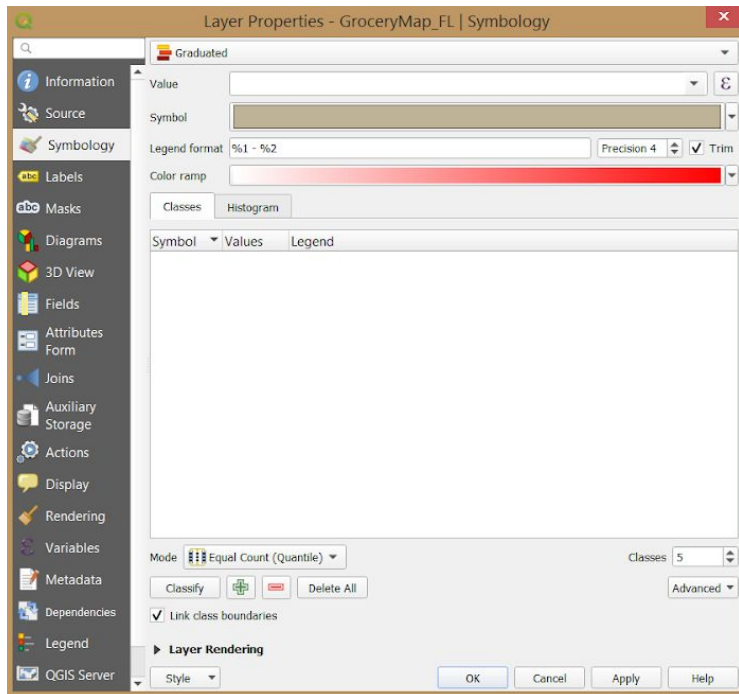
1. Double click on the shapefile in the *Layers* menu. This will open a window where we can change layer properties.
2. In the Layer Properties window, select the symbology window.



3. The number of grocery stores is a *Graduated* symbol meaning that it is a range of values. At the top, select this from the drop down menu.



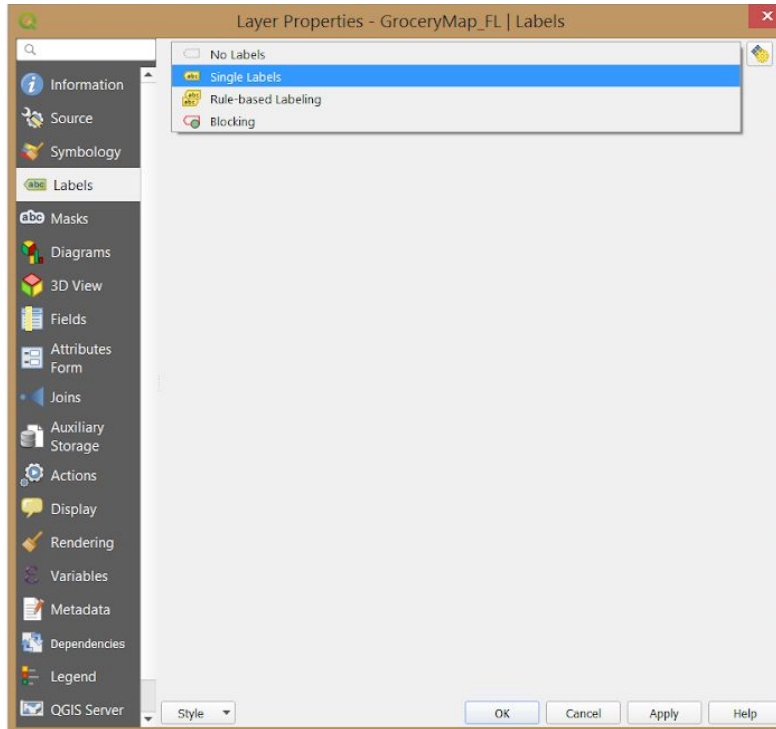
In the Value box, select the variable from above. Then, select a color ramp that you like the most. You can change this at any time!



4. Once you have selected the way you want to display your data click the classify button at the bottom of the window. The box in the middle of the window should now look similar to the image below:

Symbol	Values	Legend
✓	0.00000 - 0.12750	0 - 0.128
✓	0.12750 - 0.14341	0.128 - 0.143
✓	0.14341 - 0.16536	0.143 - 0.165
✓	0.16536 - 0.21218	0.165 - 0.212
✓	0.21218 - 0.50783	0.212 - 0.508

5. Click the Apply button.
6. We have one more step before we can look at our map! Under the symbology tab in the window, select the Labels tab. Select single label as the label type. If the Value is COUNTYNAME, click ok.



7. View your data and repeat for the *DiabetesMap\_FL* shapefile!

Let's make some predictions from our data! I've added a table below for you to enter county names that meet the following criteria:

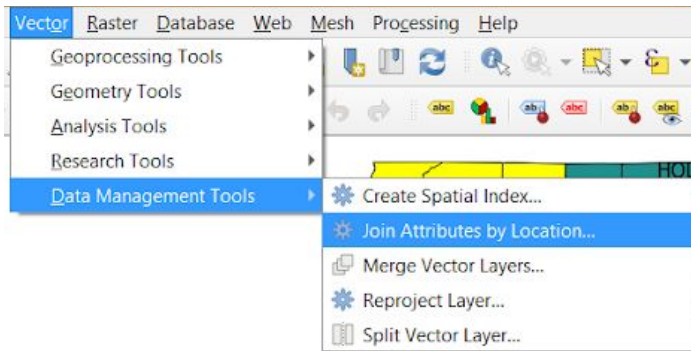
High # of Grocery	Low # of Grocery	High % Diabetes	Low % Diabetes

Is there any overlap between these?

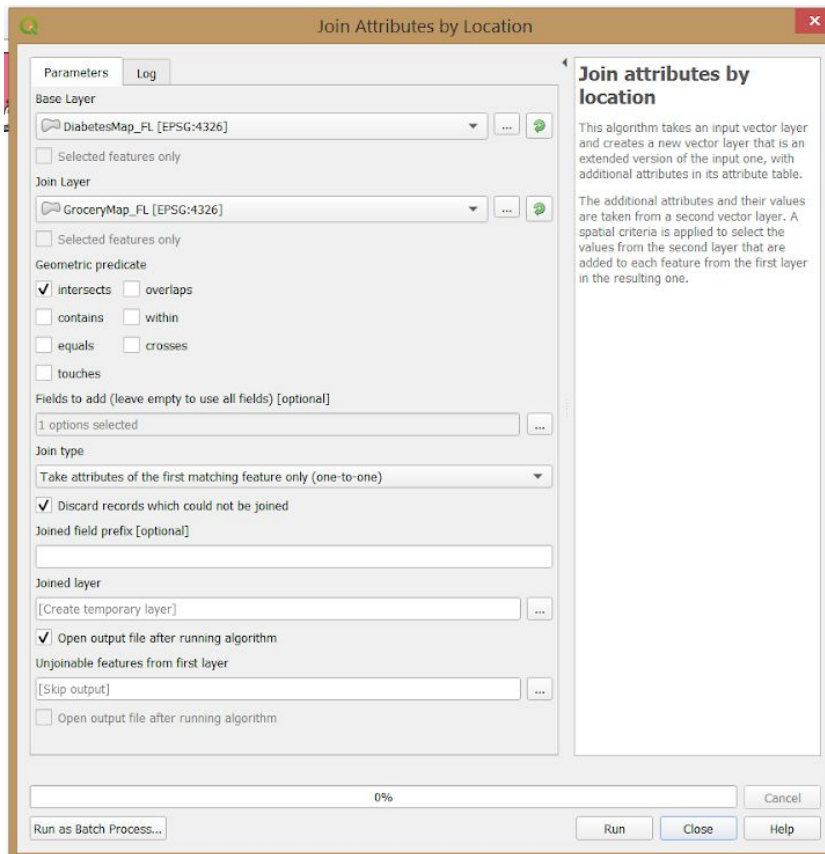
### *Part 3: Analyzing your data*

Looking at visuals between the two maps is good, but getting statistical values is better! For this section, we will be comparing the percentage of diabetes cases for counties with few grocery stores.

1. In order to compare the two values, we want to create a new Layer file with both datasets. To do so, we will use the *Join Attributes by Location* tool.



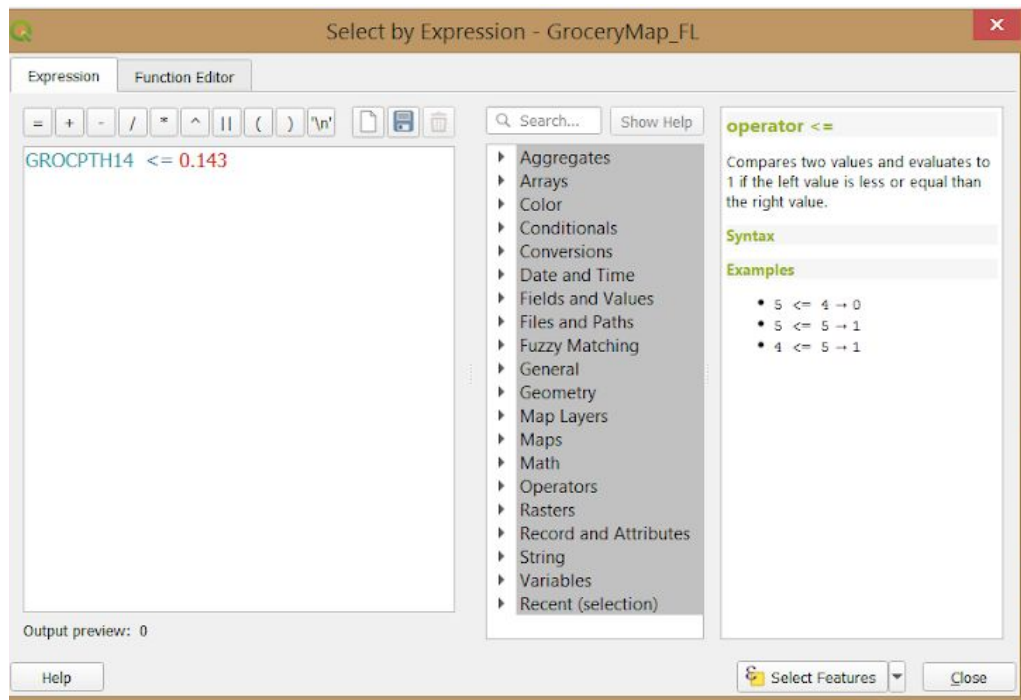
2. In order to run this tool, enter the criteria exactly as written below. Please note that *Take attributes of the first matching feature only* is not the default *Join type*. If you would like to save the new shapefile for the future, select the three dots next to the Joined layer window and save the files as an ESRI shapefile.



3. Open the new *Joined Layer* (or what you named it) attribute table. At the top of the window there is an icon that looks like an E with a box behind it (it should say 'select features using an expression' when hovering over).
4. Enter the expression below as written and click Select Features.



**GROCPH14 <= 0.143**



Those values which meet those criteria should then be highlighted.

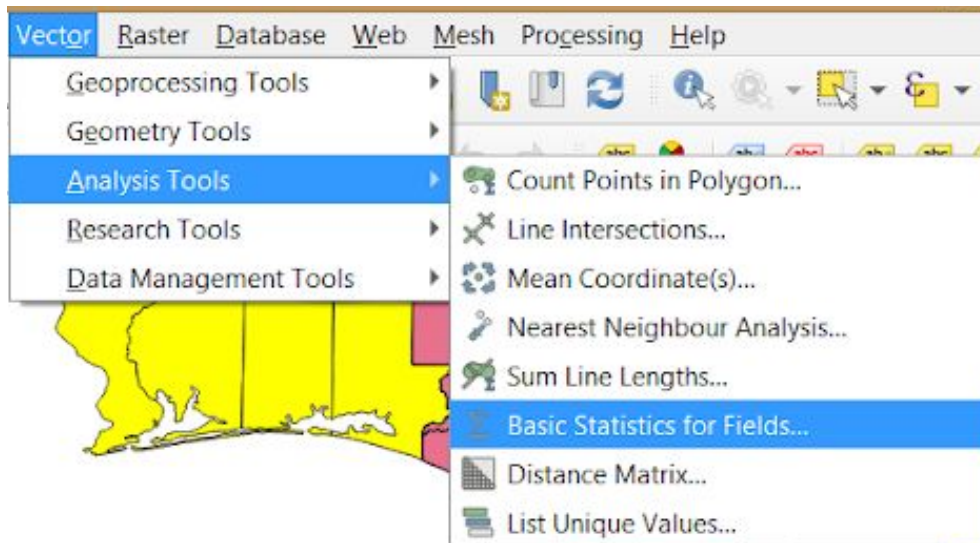
The screenshot shows the QGIS attribute table with the following data:

OBJECTID	DEPCODE	COUNTY	COUNTYNAM	DATESTAMP	ShapeSTAre	ShapeSTLen	County_1
1	34	52 103	PINELLAS	2000-05-16	8169203571.88...	1167816.84046...	PINELLAS
2	43	53 105	POLK	2000-05-16	56039729499.5...	1273587.97322...	POLK
3	57	58 115	SARASOTA	2000-05-16	16244595346.2...	866480.222854...	SARASOTA
4	58	14 027	DESOTO	2000-05-16	17811561535.2...	552611.745281...	DESOTO
5	21	41 081	MANATEE	2000-05-16	20940887080.7...	1218253.44408...	MANATEE
6	23	25 049	HARDEE	2000-05-16	17792803499.7...	543261.536629...	HARDEE
7	54	45 089	NASSAU	2000-05-16	17668341931.5...	1275144.97326...	NASSAU
8	18	30 059	HOLMES	2000-05-16	13514831669.2...	577279.375440...	HOLMES
9	39	63 125	UNION	2000-05-16	6963394478.96...	420643.595360...	UNION
10	36	16 031	DUVAL	2000-05-16	22997044775.9...	1114099.40128...	DUVAL
11	44	49 097	OSCEOLA	2000-05-16	41974373975.6...	1149911.15232...	OSCEOLA
12	56	8 015	CHARLOTTE	2000-05-16	19647662652.4...	1306868.96109...	CHARLOTTE

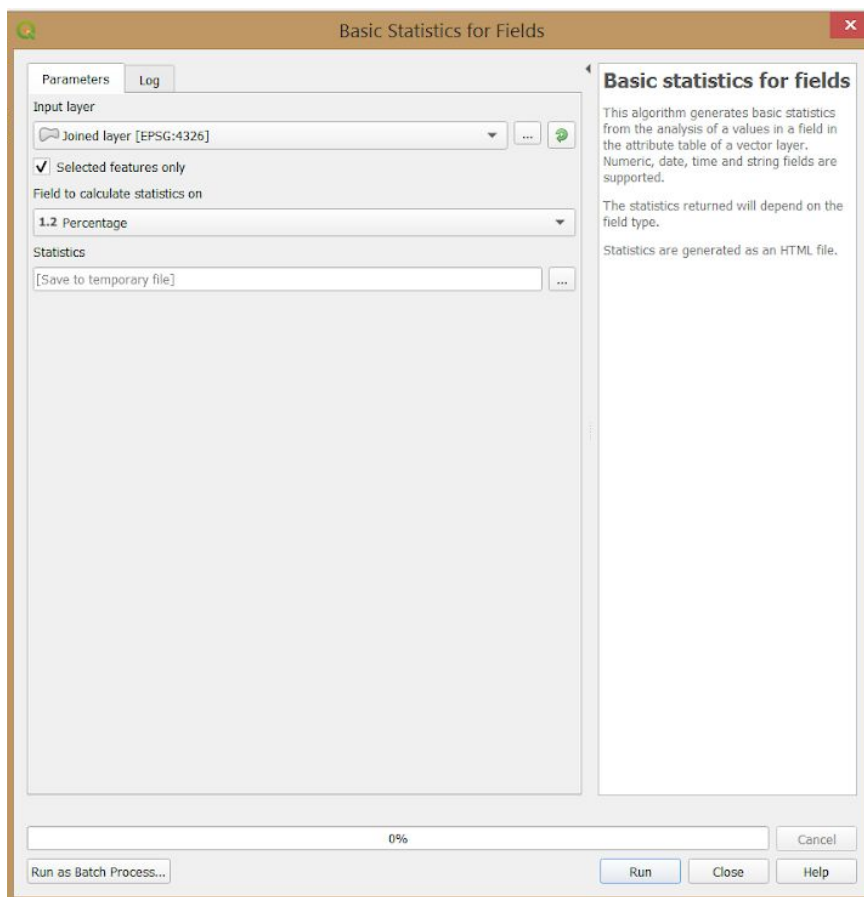
What did that calculation just do?

- Finally, we want to get some basic statistical information (percentages of diabetes cases) in the grocery stores selected above (low number of grocery stores). To do so, we will use the *Basic Statistics for Fields* tool.





6. Fill out the tool window. We will be using Selected features only so be sure to check it! The field we are interested in calculating is Percentage, although please note you could calculate statistics for other fields.



Once this tool is run, you should be able to calculate the max, min and mean percentage of diabetes cases for these counties.

Percentage of Diabetes cases in counties with low access to Grocery Stores

Maximum	
Minimum	
Mean	

7. How do the values compare to those of counties with higher numbers of grocery stores?  
In order to answer this question first deselect the values above by opening the attribute table and clicking the deselect all button at the top (looks like a box with a little red box in the right hand corner) and repeat steps 2-6. I recommend using this equation for step 4:  
**GROCPH14 >= 0.165**

Percentage of Diabetes cases in counties with high access to Grocery Stores

Maximum	
Minimum	
Mean	

Did these values provide similar to the patterns expected from Part 2?

From here, there's plenty more you can play with:

- Do other statistical values show patterns better?
- How do the values change as the cut off changes?
- What if we used a different variable as a metric to represent food deserts?
- How do your results vary if you compared diabetes cases to 2009 data?

If you want to play with more data, the links to access this data are below. There are also interactive maps displaying this data.

Data from USDA Economic Research Service and United States Diabetes Surveillance System

<https://www.countyhealthrankings.org/app/ohio/2020/measure/outcomes/60/data?sort=sc-0>

<https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html>

<https://www.ers.usda.gov/data-products/food-environment-atlas/go-to-the-atlas/>

<https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/>