

# Workshop 2

## Exploring Spatial Relationships

Today, we will visualize the relationship between Starbucks locations and median household income. We will use rising median household incomes as a proxy for measuring gentrification. There are various methods to explore and quantify gentrification and median household income is only one of those measures. As with any other method of measuring a qualitative concept, it will not be able to capture gentrification perfectly. However, it is still a useful measure. Then, we will also look at the relationship between Starbucks locations and certain demographics (today, we will focus on the Percentage of Black people).

### Notes on the Data:

The American Community Survey data was downloaded from <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>. The median household income data was taken from Social Explorer and 2010 data was adjusted into 2016 dollars in Social Explorer. Starbucks data was downloaded from <https://community.periscopedata.com/t/80fyna/starbucks-locations>.

### Steps

1. Download the zip folder from the Dropbox link provided and extract it to your working folder. There are four folders within the Workshop2 folder. **NYC Census Shapefile**, contains a shapefile of NYC at the Census tract level and **Other Data**, contains data on demographics/housing, household income, and rent as a percentage of income. This data may be useful for your individual projects. The two folders that start with ACS contain breakdowns by race for 2 different years, 2016 and 2011. This will allow us to visualize the demographic change over time (if we have time in the workshop to get to that!). Then, there are a few files in the folder. The **all\_starbucks\_locations\_in\_nyc.csv** is unfortunately, misnamed, and contains information on all the starbucks in the world. This is not a shapefile, but it contains latitude and longitude information, which we will use to create a shapefile. Lastly, there is the **income.csv** and **income.csvt** file. The csv file contains median household income for 2010 and 2016 at the census tract level and also contains a column with the change in income from 2010 to 2016. The csvt file tells QGIS how to read the csv file and tells it what type each variable we are reading is. You usually need this for any tabular data that you are pulling in to QGIS, but if you notice, we don't have a csvt file for the starbucks data. That is because we will be creating a shapefile from it and we will use a tool in QGIS to read in the data accordingly.

Below are pictures of the starbucks csv and the income csv.

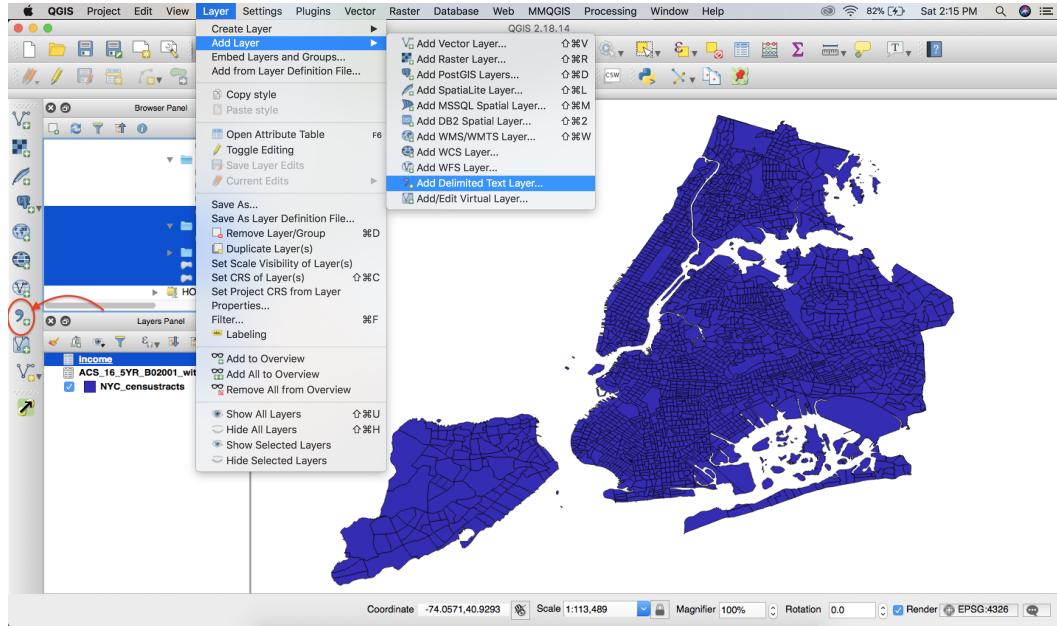
Starbucks:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
Store ID	Name	Brand	Store Numbe	Phone Numb	Ownership	T Street	Comb Street 1	Street 2	Street 3	City	Country Subc	Country	Postal Code	Coordinates	Latitude	Longitude	Timezone	Current Time	Olson Timez	First Seen
2	1 Plaza Hollyw Starbucks		34638-85784	29554570 LS	Level 2, Plaza Level 2, Plaza Kowloon					Hong Kong	91 CN		(22.3407001, 22.3407002)	114.201691	China Stande	480	GMT+08:00	#####		
3	6 Exchange Sq Starbucks		34601-20281	21473739 LS	Shops 308-3 Shop 308-3: Exchange Square Podium, Hong Kong					Hong Kong	91 CN		(22.2839393, 22.2839394)	114.158188	China Stande	480	GMT+08:00	#####		
4	8 Telford Plaza Starbucks		34610-28207	27541323 LS	Shop Unit G:Shop Unit G/F, Kowloon Bay, Kowloon					Hong Kong	91 CN		(22.3228702, 22.3228703)	114.21344	China Stande	480	GMT+08:00	#####		
5	13 Hong Kong S Starbucks		34622-64463	25375216 LS	Concession I-Concession I-LAR Hong Kong Station					Hong Kong	91 CN		(22.2844505, 22.2844505)	114.158463	China Stande	480	GMT+08:00	#####		
6	17 Pacific Place Starbucks		34609-22927	29184762 LS	Shop 131, Le Shop 131, Le 88 Queensway, HK					Hong Kong	91 CN		(22.2776565, 22.2776566)	114.164619	China Stande	480	GMT+08:00	#####		
7	23 Three Garde Starbucks		34616-51683	25091007 LS	Shop 3, Citib Shop 3, Citib Central, HK					Hong Kong	91 CN		(22.2785205, 22.2785206)	114.160873	China Stande	480	GMT+08:00	#####		
8	25 MTR Central Starbucks		34630-69462	28400565 LS	Kiosk Nos. CI Kiosk Nos. CI, MTR Central Station					Hong Kong	91 CN		(22.2811298, 22.2811298)	114.159462	China Stande	480	GMT+08:00	#####		
9	27 Wan Chai To Starbucks		34615-55886	28240568 LS	G/F, Wanchi /Wanchi, Wan Chai, HK					Hong Kong	91 CN		(22.2800197, 22.2800198)	114.172615	China Stande	480	GMT+08:00	#####		
0	34 Shun Tak Cer Starbucks		34634-82422	25496508 LS	Shop unit 22/Shop unit 22 Nos.168-200 Connaught R Hong Kong					Hong Kong	91 CN		(22.2878608, 22.2878609)	114.152206	China Stande	480	GMT+08:00	#####		
1	37 Sun Hung Ka Starbucks		34605-28027	28457367 LS	Shop G12-13 Shop G12-13 30 Harbour Road, Wan Chai Wan Chai					Hong Kong	91 CN		(22.2805519, 22.2805519)	114.176804	China Stande	480	GMT+08:00	#####		
2	38 Lamrv Plaza Starbucks		34247-67176	04-3347995 LS	Ourl Meena   Ourl Meena  First floor					Dihai	DII		(25.7343463, 25.7343464)	55.3084297	Arabian Stan	740	GMT+08:00	#####		

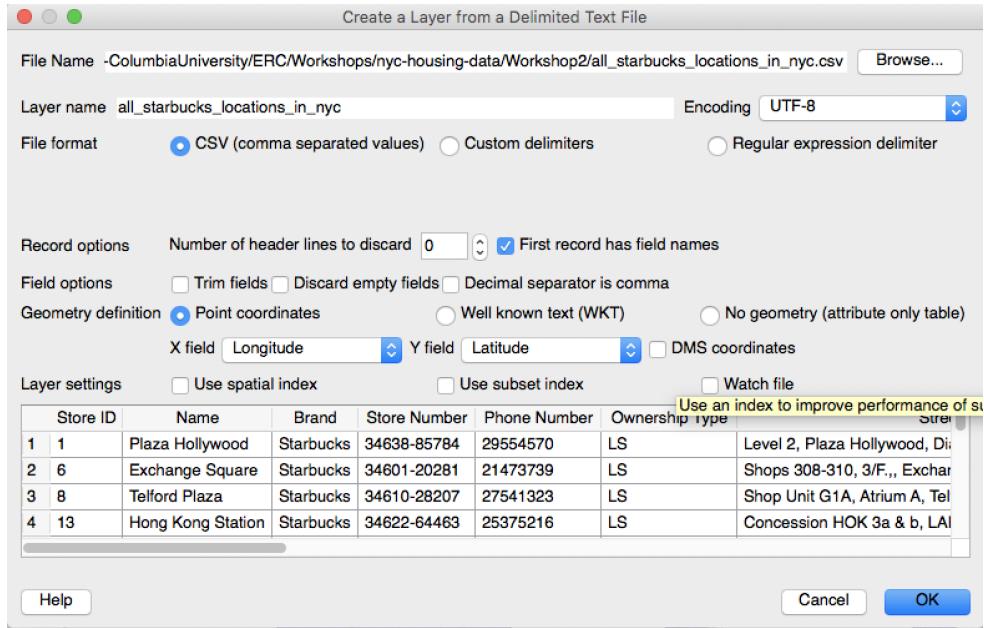
Income:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Geo_FIPS	Geo_GEOID	Geo_NAME	Geo_QName	Geo_STUSAU	Geo_SUMLEV	Geo_GEOCO	Geo_FILEID	Geo_LOGREC	Geo_STATE	Geo_COUNT	Geo_TRACT	inc2016	inc2010	incDiff
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3481	36	5	100	0	0	
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3482	36	5	200	70893	65880	5013
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3483	36	5	400	76667	79250	-2583
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3484	36	5	1600	31540	33427	-1887
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3485	36	5	1900	39130	27632	11498
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3486	36	5	2000	17083	27462	-10379
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3487	36	5	2300	16503	15148	1355
3.6005E+10	14000US360	Census Tract	Census Tract	ny	140	0	ACSSF	3488	36	5	2400	0	0	

- Notice in the picture above, that the income.csv has a column called Geo\_FIPS that contains a 11 letter code starting with 36 that uniquely defines each census tract. We will use this variable when we are joining our data.
- Open QGIS. In the Browser window, navigate to the Workshop2 folder. Make sure it is the extracted folder and not the zip folder. Drag in NYC\_censustracts.shp from the shapefile folder. Then, from the ACS\_16 folder, drag in the ACS\_11\_5YR\_B02001\_with\_ann.csv and the income.csv file.
- Now, we will create our Starbucks points. Click on icon marked below or choose the setting from the Layer menu as shown below. Both accomplish the same thing.



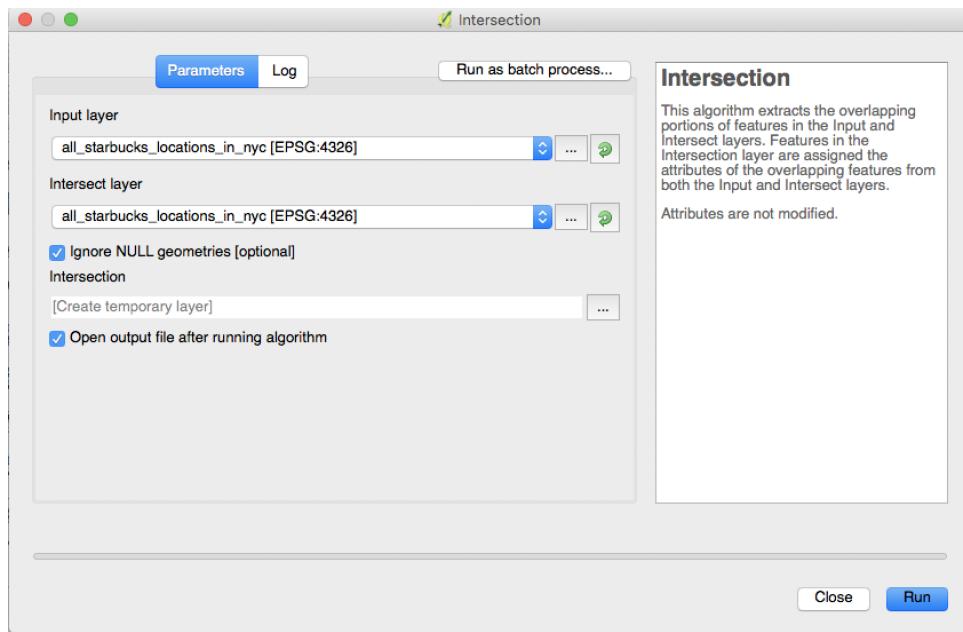
- Choose the starbucks csv for the file name. For file format, make sure csv is checked off. The rest of it should autopopulate, but if it doesn't, you have to make sure that you select Point coordinates under Geometry definition and then for X field, choose Longitude, and for Y field, choose Latitude. Depending on your version, you may be asked to define a coordinate system. If you are asked to, search for the code 4326 and choose the option that has EPSG:4326.



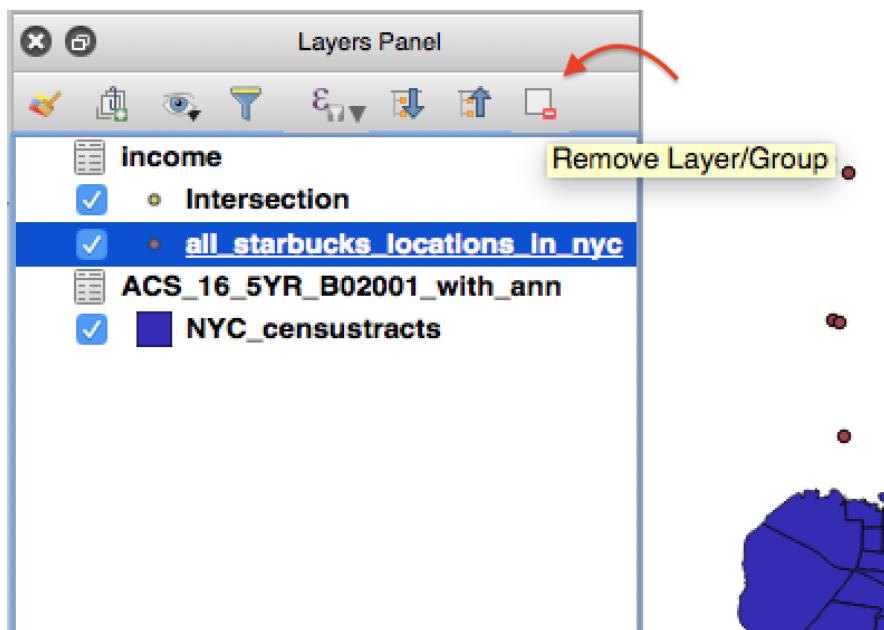
6. Your map window should look like this now:



7. We want to limit our points to those that fall in NY. Make sure you save your QGIS project first because the tool can sometimes be glitchy. To do that, click on **Vector > Geoprocessing Tools > Intersection**. Your input layer should be the starbucks layer and your intersect layer should be the NYC\_censustracts layer. Check off ignore NULL geometries, which will ignore points in our layer that don't have defined latitude and longitude. Then, click on the ... icon to the right of where it says **Create temporary layer** and choose **Save to file**. Navigate to the Workshop2 folder and save this as starbucks\_nyc. Make sure that the file type says .shp. If it doesn't, then for now, don't save to a file and just choose to create a temporary layer. That is what I did right now, since QGIS is not letting me save it as a shapefile. Make sure the option to Open the file after running the algorithm is checked off and then click Run. This should take a few minutes. If it takes longer than that, quit QGIS and reopen your saved QGIS project.



- Now, we should have an added layer called **Intersection** that has points only in NY. We can remove the original starbucks layer. To do that, click on it in the layers panel, and then click on the Remove layer icon.



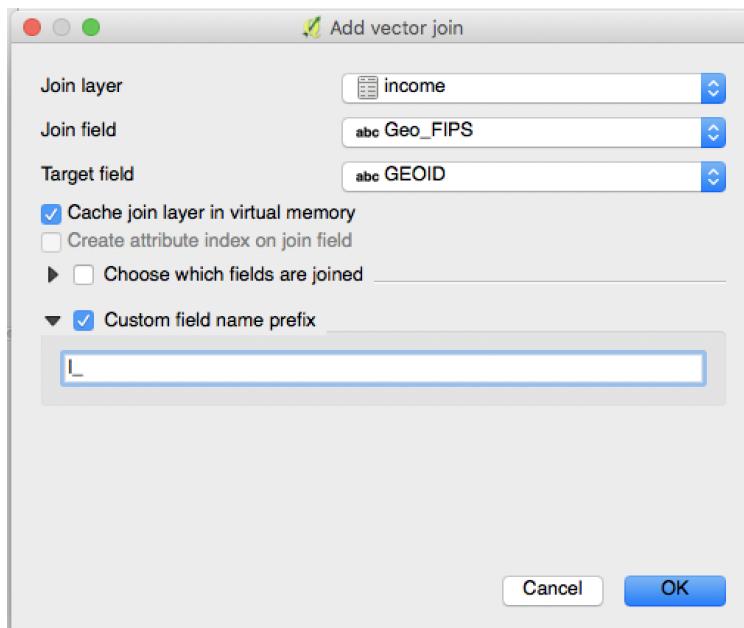
- Now, we will join our shapefile layer to our two tabular data layers. Let's Open the Attribute table for each of the layers and find a variable that uniquely identifies a census tract and is across all three files. (Reminder: To open a layer's attribute table, right click on it in the layers panel and click **Open Attribute Table**.) Below is the attribute table for the NYC\_census shapefile. The variable we will be joining on is GEOFID. If you look at the Income layer, you will find that variable is called Geo\_FIPS, and in the ACS layer, it is called Geo.id2.

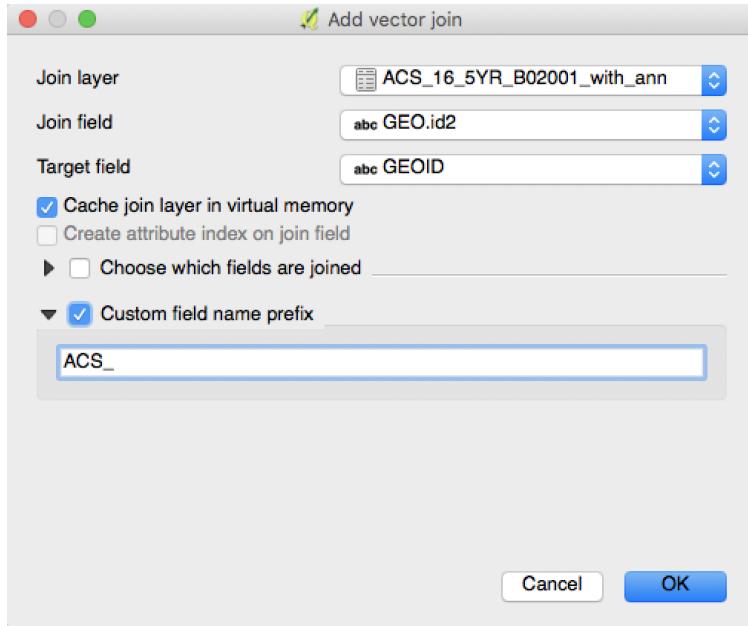
NYC\_censtracts :: Features total: 2195, filtered: 2195, selected: 0

	STATEFP	COUNTYFP	TRACTCE	GEOID	NAME	NAMESAD	MTFCC	FUNCSTAT
1	36	005	000100	36005000100	1	Census Tract 1	G5020	S
2	36	005	000200	36005000200	2	Census Tract 2	G5020	S
3	36	005	000400	36005000400	4	Census Tract 4	G5020	S
4	36	005	001600	36005001600	16	Census Trac...	G5020	S
5	36	005	001900	36005001900	19	Census Trac...	G5020	S
6	36	005	002000	36005002000	20	Census Trac...	G5020	S
7	36	005	002300	36005002300	23	Census Trac...	G5020	S
8	36	005	002400	36005002400	24	Census Trac...	G5020	S
9	36	005	002500	36005002500	25	Census Trac...	G5020	S
10	36	005	002701	36005002701	27.01	Census Trac...	G5020	S
11	36	005	002702	36005002702	27.02	Census Trac...	G5020	S
12	36	005	002800	36005002800	28	Census Trac...	G5020	S
13	36	005	003100	36005003100	31	Census Trac...	G5020	S
14	36	005	003300	36005003300	33	Census Trac...	G5020	S

Show All Features

10. To join our layers, we will right click on our NYC\_census layer and click **Properties**. Remember that we will always join on our shapefile layer. In the Properties window, click on the Joins tab and then click on the + icon. Choose the variables from the layers and then change the prefix to something shorter. In this case, I chose **I\_**. Then, follow the same steps for the ACS layer. If you don't remember how this table should be filled out, please take a look at the previous workshop guide. Once you add both layers (pictures included below), then your join window should show both layers as shown in the last picture.





Layer Properties - NYC_censustracts   Joins						
Join layer	Join field	Target field	Memory cache	Prefix	Joined fields	
income	Geo_FIPS	GEOID	✓	I_	all	
<b>ACS_16_5YR_B02001_with_ann</b>	<b>GEO.id2</b>	<b>GEOID</b>	<b>✓</b>	<b>ACS_</b>	<b>all</b>	

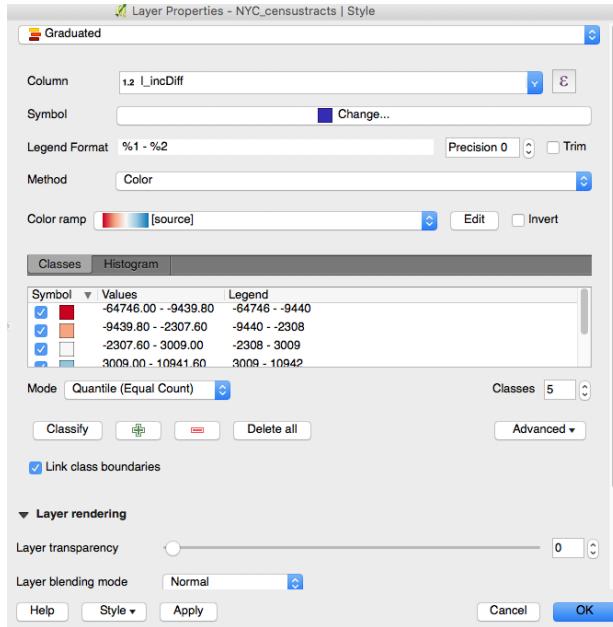
11. Open the Attribute table for the NYC\_census layer. Scroll to the right and make sure you have added variables from the two csv layers. They should have either I\_ or ACS\_ as their prefix and make sure the values for them are not NULL throughout the whole table. (But there will be some NULLs since not all of the tracts in the shapefile are contained in the csv's)
12. Now, we want to visualize our data as in the last workshop. Before we do that, let's change our projection. We will only change it temporarily for the project and won't be reprojecting our layers. To



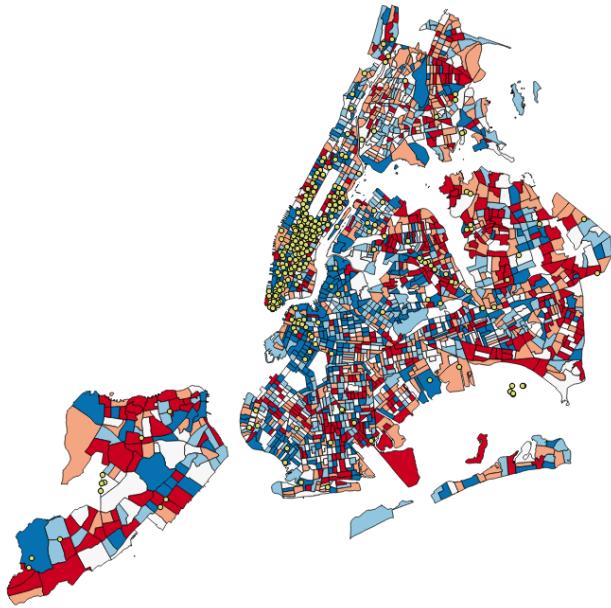
do that, click on this button on the bottom right of your window.

13. Check off the button that says **Enable on the fly transformation**. This will visually reproject our layers for our project, but will not change the projection of the layer. Use the search bar, called Filter, to search for 3857 and choose the projection that shows up, which should have 3857 as its authority id and then click Ok. Note: You can continue using 4326 if you prefer, but I prefer using this projection for NY.
14. Let's first visualize the difference in median income. Right click on the NYC\_Census layer and click Properties. Then click on the Style tab. Change the visualization from Single Symbol to Graduated. Choose I\_incDiff for the column. Click Classify. You should notice that our classes range from negative to positive values, where a negative value means the median household income decreased in that census tract and a positive value means the median household income increased in that census tract from 2010 to 2016. This means that we have values that are spread away from a meaningful center. In our

case, the center is 0, which means that area did not have a change in median household income. (The numbers have been adjusted for 2016 dollars so inflation is accounted for.) So, we will change our Color ramp to a diverging color ramp, where one end is one color and the other end is another color. Once you do that, change the Mode to Quantile (Equal Count) so that each class has an equal number of census tracts. Click Ok if your window looks something like this: (with perhaps a different color ramp):

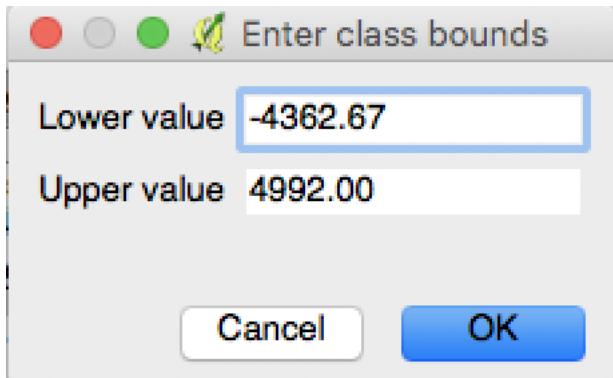


15. Your map should look like this:

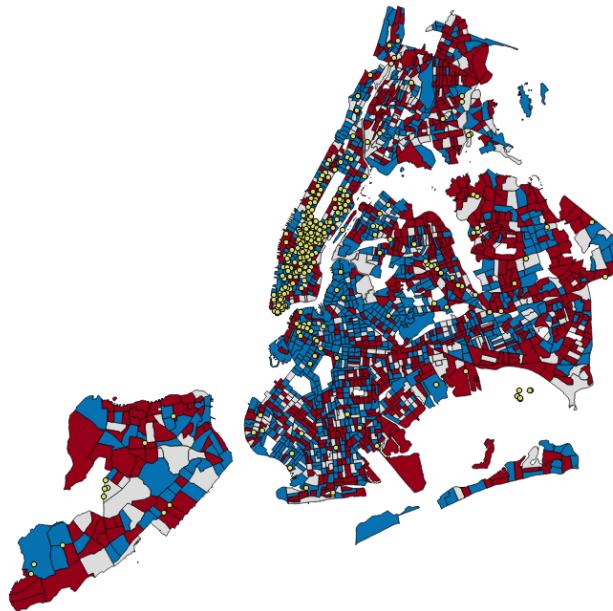


16. It's hard to see patterns when you're looking at it on the city level, so make sure to zoom in and explore the different spatial patterns. To make it a bit easier, let's change our classes and also remove outlines around the census tracts. Reopen the Properties window. Let's reduce our classes so we only have 3 of them. Let's keep the middle class that goes from -2308 to 3009 since that is a relatively small change, and just edit the other classes. There are a few ways to do this. What we'll do is change the number of classes to 3 in the place marked below. Then, we will double click on the values listed for the middle

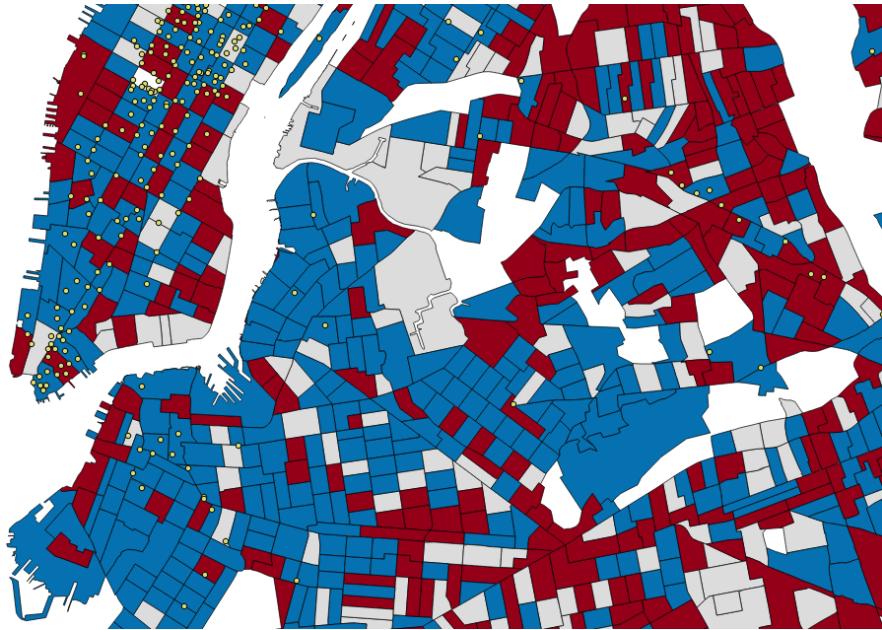
class. The window shown in the second picture should show up. Change those values to what we had before (-2308 for lower value, and 3009 for upper value). You can also round those to clean numbers (say -2500 and 2500), but since we are only exploring our data for now, we won't. Once you edit those numbers and click Ok, the other numbers should automatically adjust.



17. Then, click on the Change button marked below. Click on Simple fill and then change Outline Style to No pen. Now your map should look like this.



18. If you zoom in to certain areas, there are some points that support our theory that Starbucks and gentrification are linked, but there are also points that seem to show otherwise (pictured below). This may be for various different reasons. Some include: 1) Median household income may not be a good proxy for gentrification. 2) Our data shows difference from 2010 to 2016. Perhaps we need to look at different time periods. 3) Our data may not properly reflect the ground reality. 4) The causal relationship between Starbucks and things such as median household income or rent is not clear. Do rising incomes/rent then lead to a Starbucks being opened in that area or does a Starbucks first appear, then causing rents to go up and people with higher incomes moving in? To explore this, we would want to explore trends in the years around when a Starbucks was opened. However, we do not have that data at the moment.



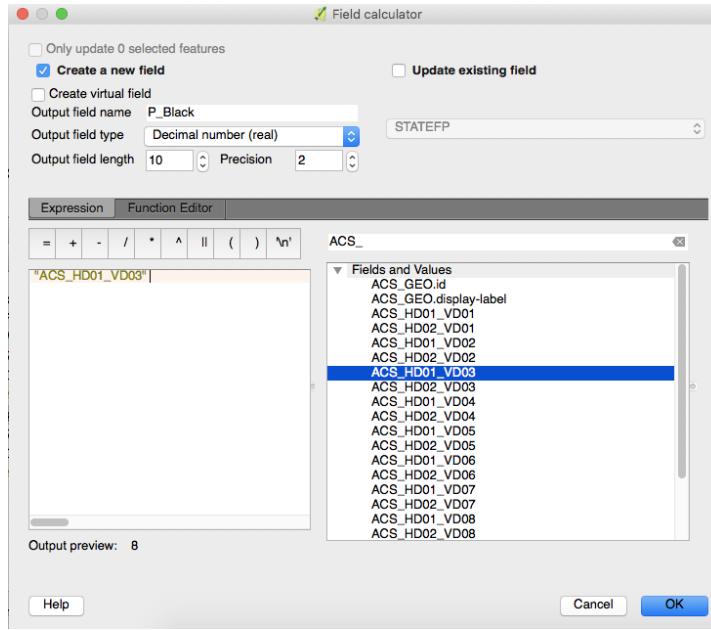
19. Now, let's look at demographic data. First, in your layers panel, right click on the NYC\_Census layer and click **Duplicate**. This should create a second layer which will have the same attribute table. Let's rename our panels so we know what each is reflecting. Right click and choose the **Rename** option. Let's name the original census layer **Diff in Median Income**. Let's name the second layer **% Black Population**. Right now, we've only changed the name and not what the layer is showing. So let's do that. Before we can go into Style and change the column, we have to create a column to show the % of the Black Population. In our dataset, we have a variable that shows the number of Black people, but we have to normalize that to get the percent. To do that, right click on the layer, and click **Open Attribute Table**.



Table. Click on this icon for editing mode, and then click on the Field Calculator icon.



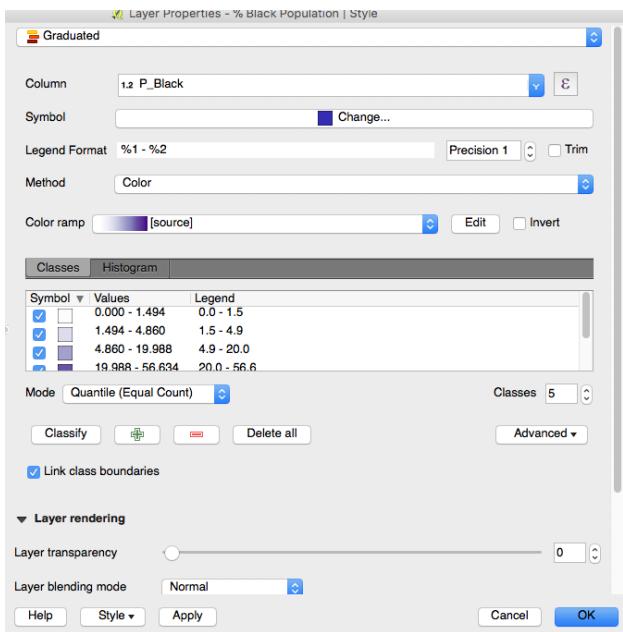
20. Then, check the **Create a new field** option. Name it **P\_Black** (P for percent). Change the **Output field type** to **Decimal number (real)** and change the **Precision** to 2. The precision affects how many places to the right of the decimal place we can go. In the Search bar on the right side, search for ACS. Our variables from the ACS csv should show up. Double click on **HD01\_VD03**. (The reason I know that this is the variable representing the total number of Black/African American individuals is because I looked in the metadata file). It should show up in your Expression window:



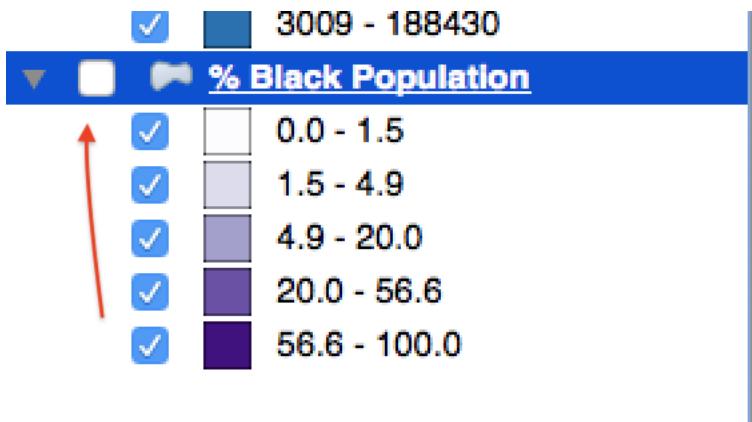
21. Write  $/$  after that since we want to divide this number by the total population. After that, double click on **ACS\_HD01\_VD01**, which is the variable containing the estimate for the total population. Lastly, multiply by 100 so that we have a percent. Your expression should look like the image below. Click Ok.

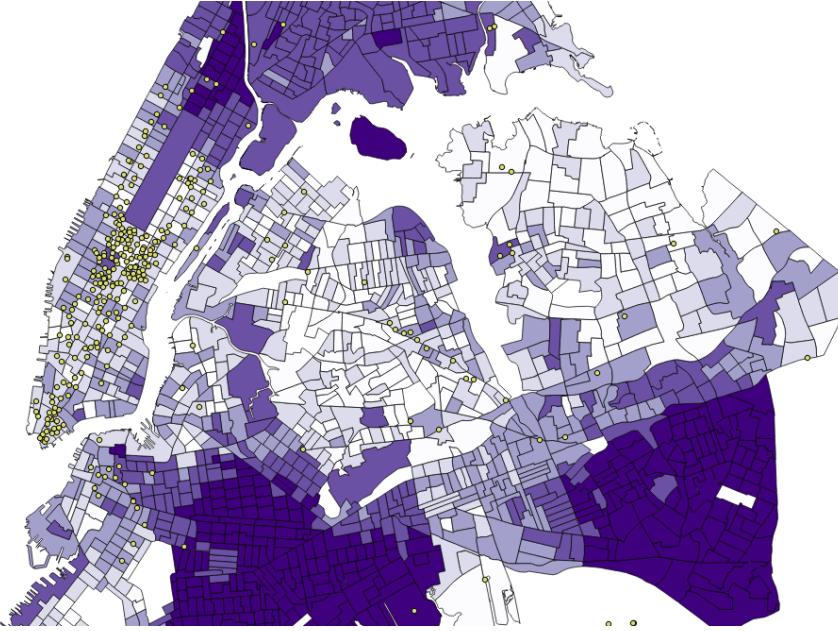
**"ACS\_HD01\_VD03" / "ACS\_HD01\_VD01" \*100**

22. The new variable will be added as the last column. Scroll to the right to see that it was added.
23. Now, click on the pencil icon again to stop editing. When asked if you want to save edits, click save. Then, open the properties window and change the Column variable to **P\_Black**. Click Classify. Now, our variable will go from 0 to 100 since it is a percentage, so it doesn't make sense to have a diverging color scale. Change the Color ramp and choose a single color color ramp and let's also change the number of classes back to 5. The window should look something like this:

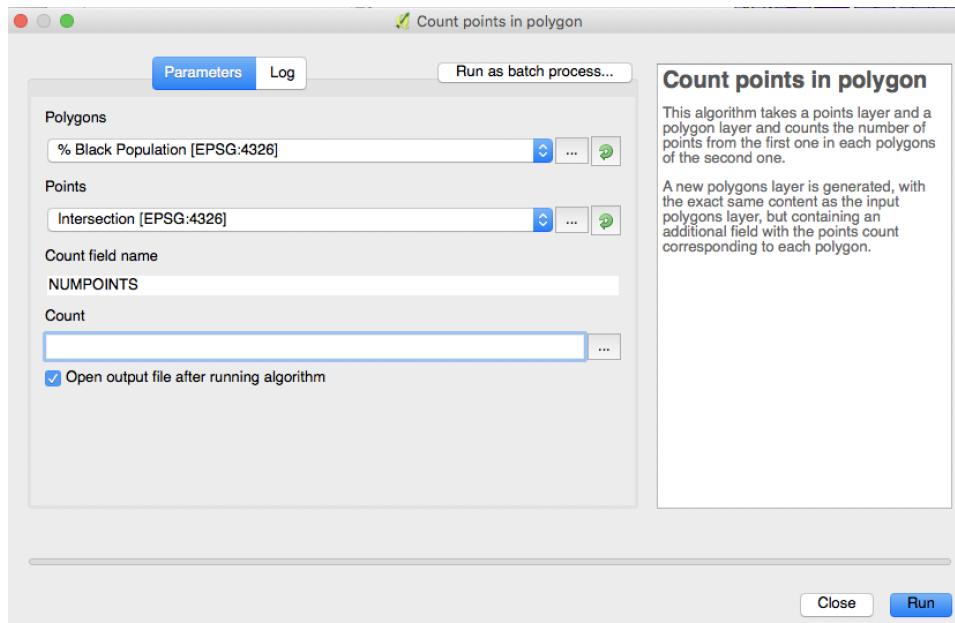


24. Once you click ok, you might notice that nothing has changed. That's because in the layers window, we have not checked off the % Black Population layer so it isn't being visualized. Even after you check it off, it doesn't show up! Why? That's because QGIS visualizes layers in the order they are listed in the layers panel. Since our **Diff in Median Income** layer is above this one, it is the only one being shown, so let's uncheck that layer. Now, our map should look like the second image.

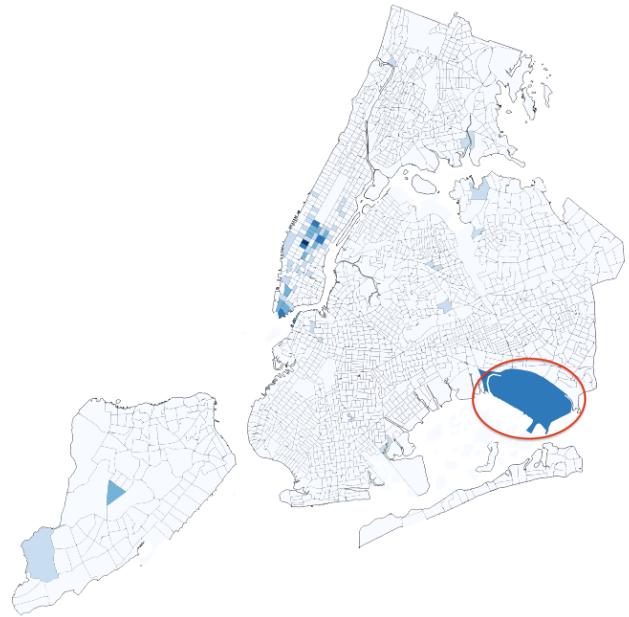




25. You might notice that most Starbucks are in locations with smaller percentages of Black people.
26. I'll show you how to do one last thing (This wasn't covered in the workshop, but may be helpful for your projects.) To see how many Starbucks are in each Census tract, we can click Vector -> Analysis Tools -> Count points in Polygon.



27. This will create a new polygon layer in our panel. Now, we can open its Properties table and visualize the new NUMPOINTS variable. That is what I've done below. Notice this one area, that has quite a few starbucks, but it did not show up in our % Black Population layer. That is because it was missing data in the ACS table. Visualizing layers lets us find some of these anomalies quickly. We can also see that Starbucks locations are generally clustered together in certain neighborhoods rather than spread evenly across the city.



28. We are not covering buffers in the workshop, but here is a tutorial on buffers ([https://github.com/michellejm/ConflictUrbanism-InfraPolitics/blob/master/Tutorials/05\\_SpatialAnalysis.md](https://github.com/michellejm/ConflictUrbanism-InfraPolitics/blob/master/Tutorials/05_SpatialAnalysis.md)). You can also feel free to come in during ERC Fellows' office hours to learn how to create buffers and make use of them.

---

Tutorial written by Fatima Koli, for *Spaces and Territories of Housing*.