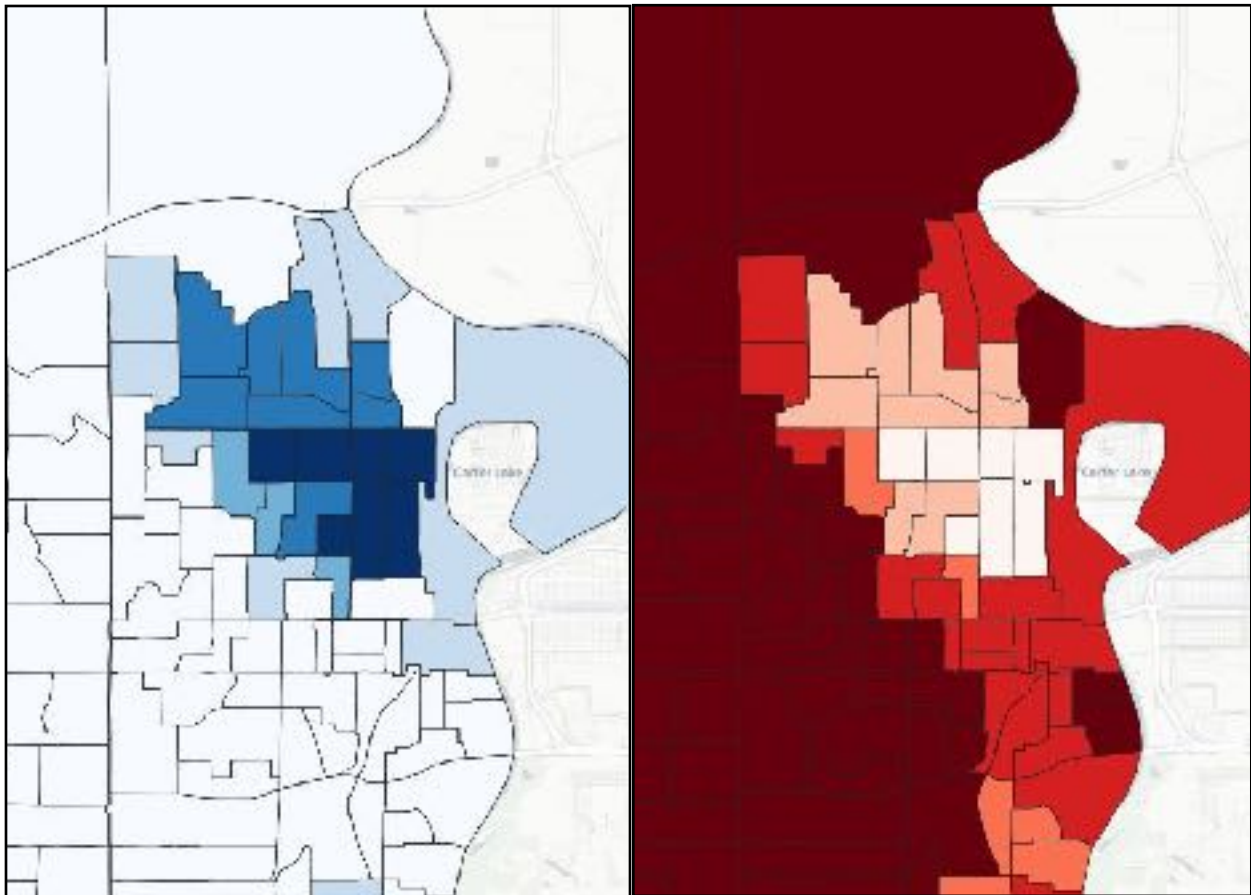# Introduction to GIS

Census Data & Choropleths

Introduction: During our last GIS lab, we were introduced to vector data layers, specifically point and line shapefiles. These are very useful tools for cartography (mapping making) and spatial analysis, which we saw when we made buffer layers. Those 'buffers' were polygons, our third type of shapefile. We will be doing more work with polygons today. We will be using historical census data to produce 'choropleth' maps - or thematic maps that show change in degree of a variable over space. They are among the most often used types of maps and they are very common in environmental justice work. Today, we will use historical census information to build choropleth maps that show change over space and time.

Instructions:

1. Let's begin with some census data. There are several places you can download historical census data online, including <u>census.gov</u> and the national historical GIS (<u>nhgis.org</u>) For those of your doing GIS work for your final project, I will show you separately how to download and use these datasets. For now, I've gone ahead and done that work for you. The datasets are available online in the shared google drive. You can download them here:

<u>https://drive.google.com/drive/folders/14cMffk3cELHqEqBscPmQbmQEx-26l4BC?usp=sharing</u>

Remember to download and save ALL of these files to your hard drive in your dedicated course folder.
Open QGIS and add a basemap from HCMGIS. Add the Omaha1950 layer to your project. Look to the last tutorial if you can't remember how. Your map should now look like this.
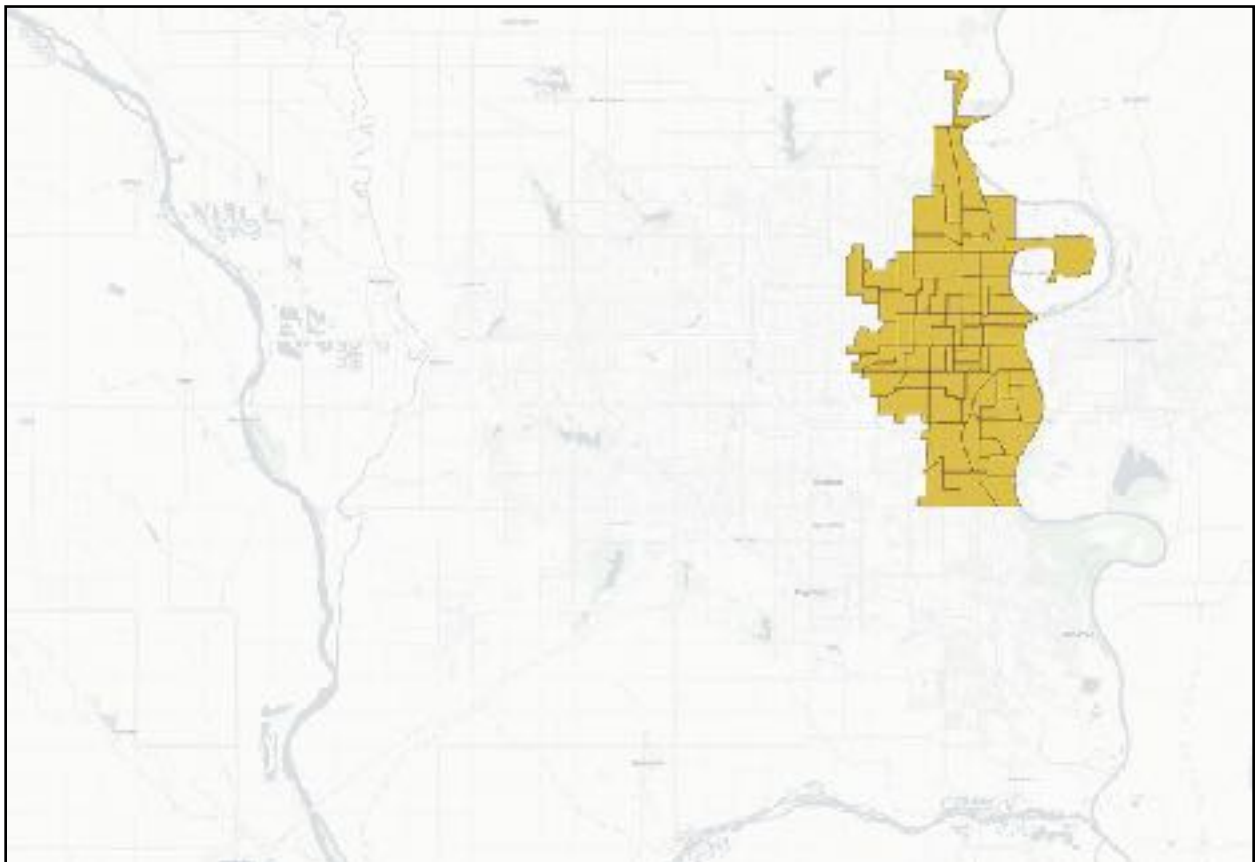
2. This is a census data shapefile from 1950, the first year that census information is available in digitized form. You can access older census data, but it is paper

format and you would need to digitize it on your own. If you would like to do so for the final project, I can work with you to do so.

3. What census information does this polygon shapefile contain? What do these shapes represent? To answer that, we need to do two things. First, we need to open the attribute table. Right click 'Omaha1950' in the layers panel and click 'open attribute table.' Each column tells me something about the information contained. NHGISST tells me what state we're dealing with - '310' is Nebraska according to the census codebook. NHGISCTY/0550 is the city name, unsurprisingly Omaha. Ignore the GISJOIN columns. These will be important for those of you working on GIS projects, but we will tackle them independently. Shape area and length are the total area of polygons in square meters and length in meters of their perimeters. The next few columns are curious. We need a key to understand them. Thankfully, we have one available.
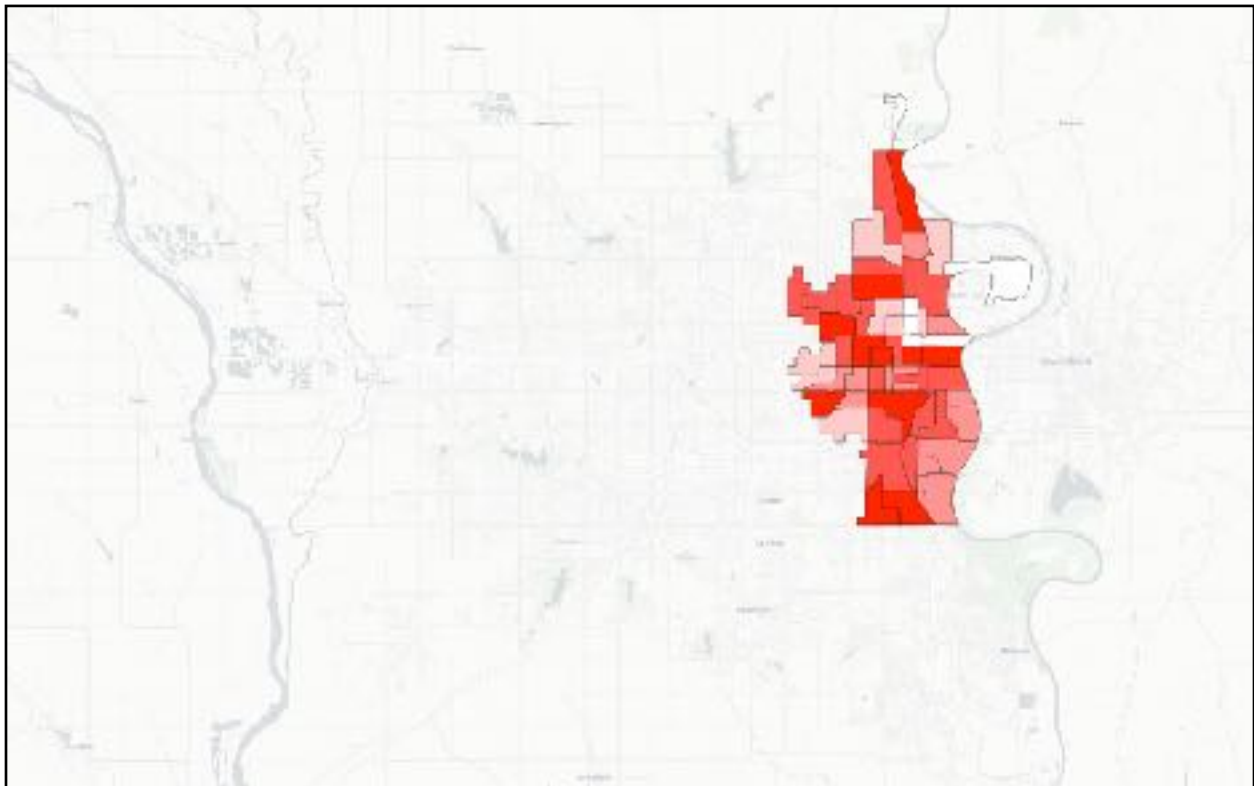

4. Look to the folder you downloaded 'Omaha1950' and open the tract codebook. What does it say? For starters, it says these polygons are census tracts. That's good to know. It also states that you've downloaded the portion of the 1950 census that dealt with race. It uses exactly the same language the census used.

Under data dictionary, you can now see what the columns B0J001/2/3 stand for. Add these columns together and we have total population per census tract. There's also one additional column, PctBlk. We'll come back to this. Using this information, we can map racial distribution across the city in 1950 at the height of the 'Great Migration.' Let's do so.

5. Right click you data layer "Omaha1950" and click "properties." You can also double click it to bring up the same screen.

6. Click "symbology" on the lefthand tab. At the top of the symbology screen, click the default style "single symbol" and change it to "graduated." Next, chose the column from the attribute table that lists total African American residents by selecting it from the dropdown menu under "column."

7. Feel free to change the color if you'd like by clicking "symbol" and "simple fill." Lastly, near the bottom, click "classify" and "ok."

8. Your map should look like this (below).

9. Save your project.
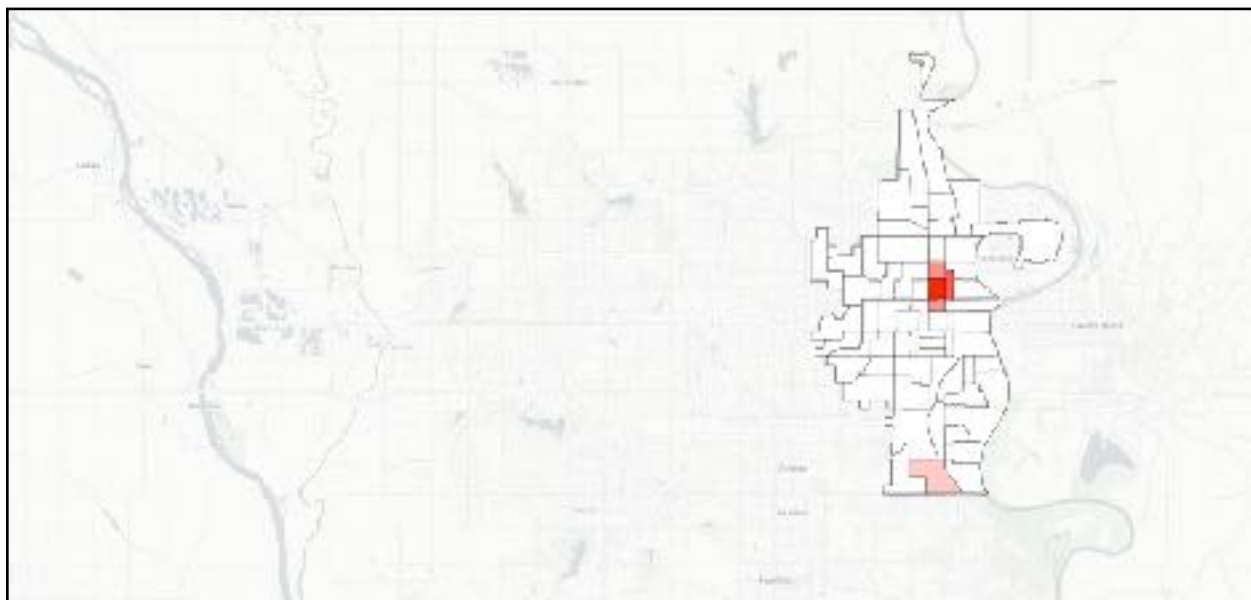
10. What does this map show? It's a map of the total number of black residents by census tract. It already shows a strongly segregated Omaha in 1950. This isn't incorrect per se, but it IS poor cartographic design. Why?

11. One of the most common cartographic errors that you often see with *choropleth* maps like this. Right now, we are visualizing the **total number** of people that live in a census tract. *Why might this be problematic when mapping with polygons?*

12. Because the size (or area) of a polygon often influences the total number people that live there. Very often, choropleths are simply maps of total population. For an example of why this can be problematic, see below.
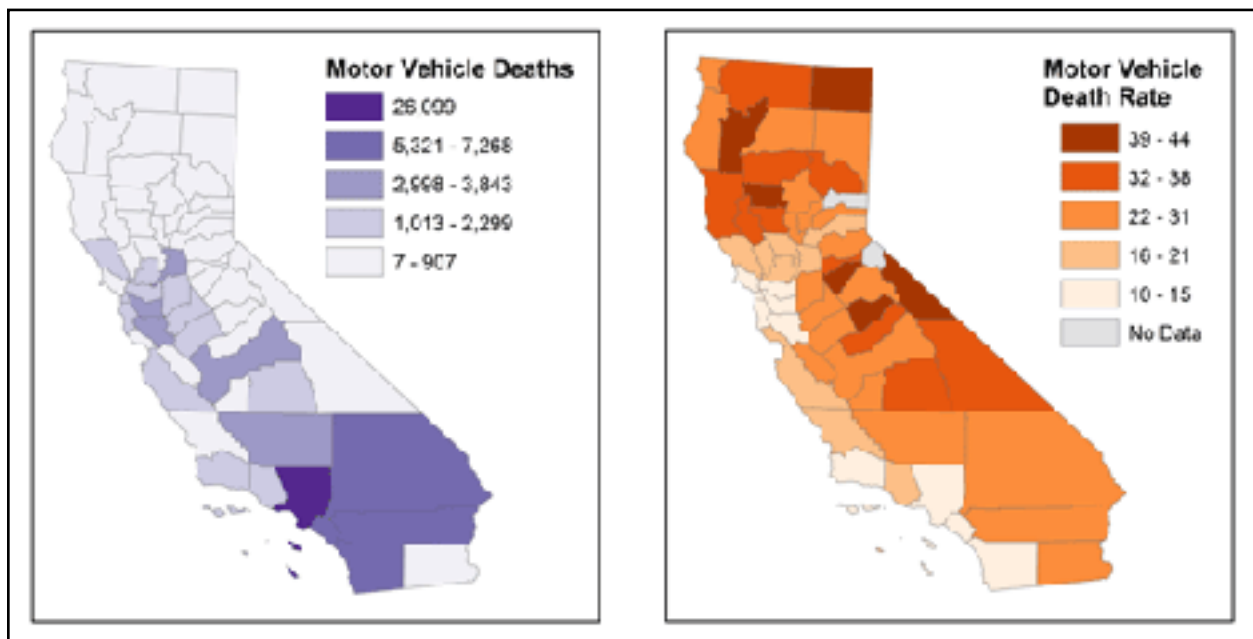
13. *What's the difference between these two maps?* One is a population map. Of course motor vehicle deaths will be greatest in LA because there are so many people. But if what we're really interested in mapping are the most dangerous places to drive in CA, then the mortality *RATE* is actually a better indicator. This shows the total number of deaths as a ratio of total population. You should NEVER map 'raw' counts as a choropleth. You should always *normalize* your data as a rate, ratio, or density.

14. *So what is wrong with our maps of Omaha?*

15. We need to *normalize our data.* We don't necessarily know whether mapping total population is truly representative of the shape or scale of segregation until we do.

16. We could do this by area (total number of people per sq meter) or by total population (total number of one race as a ratio of total population). We'll normalize by ratio of total population. I've actually already done this for you.

17. Let's create a new data layer to visualize this normalized data. Right click your data layer "Omaha1950" and click "export" and "save features as."

18. On the next screen, change format to "ESRI shapefile", beside "file name" click the ellipses and find your Omaha1950 folder. Name your new file "Blk1950" and click "save."

19. Make sure your "CRS" (the coordinate system for this map) is the same as your basemap and first layer by selecting "layer CRS." Finally, make sure "add saved file to map is checked. Click OK.

20. You should now have a second data layer, containing the same attribute information, but without any styling. It's a layer on top of your first, so it should completely hide what you've already done. You can change that by either unchecking your new layer from the layer panel or clicking and dragging it below your "Omaha1950" layer.

21. Let's change your new data layer to show the *normalized* data showing percent of total population who are African American. Remember, that column is 'P_Black'. Change the symbology of your 'Blk1950' layer to reflect this.

22. Visually, not a huge difference, but you do see an intensification of the coloration. What's more important are the numbers associated with numbers. Look at your layer panel. The census tracts that have the greatest percentage of black population reach to what figure? If you want to quickly look at attribute information by click on this map, use the 'identify features' icon. Use this tool to identify the area with the highest ratio of African American
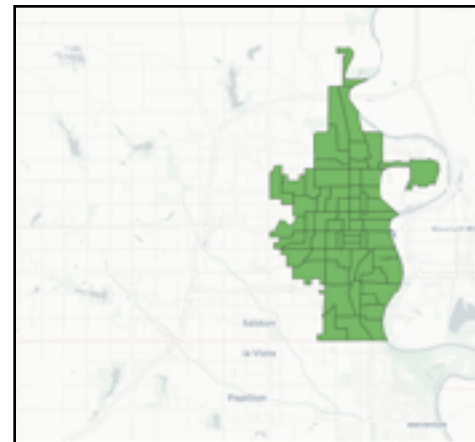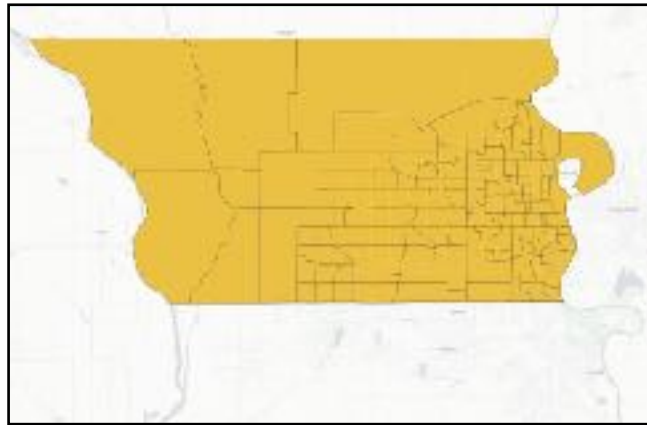
population.

23. If we want to determine the same information for the white population, we need to calculate this ratio and add it as a new column. Luckily, QGIS is built to do this easily. Complete the same set of steps (17-19) to make a new layer, except this time save it as "Wht1950."

24. A new layer should appear without symbology again. That's fine for now. Right click 'Wht1950' and open its attribute table. On the top of the attribute table, you should see several icons. One is the 'field calculator.' Click it.

25. The field calculator allows you to do basic math using your column data as variables and input your results as new columns into the attribute table. We're going to add our population columns together to determine total population. We're then going to divide the white population by that number and add it as a new column. That way, we arrive at ratio of white residents.

26. Make sure 'create new field' is checked. This will add our ratios as a new column.

27. Beside 'output field name', type 'P_Wht.' Under 'output field type' select 'Decimal number (real)'. Under output field length' select '2'. This will produce decimal numbers out to the hundredth of a percent. In the large box under 'expression' type the following: ("1950B0J001" / ("1950B0J001" + "1950B0J002" + "1950B0J003"))* 100

28. Click ok.

29. Go back to your layer panels and 'toggle editing' to save your changes.

30. Now change the symbology to 'graduated' and visualize your new PctWht column. Compare to your PctBlack map. Pretty stark difference right?

31. *Whats the highest ratio of white population you can find? Whats the lowest? Is this the same census tract you identified from the PctBlk map earlier? Add the total percentages together and you don't arrive at 100%, however. What accounts for the difference?*
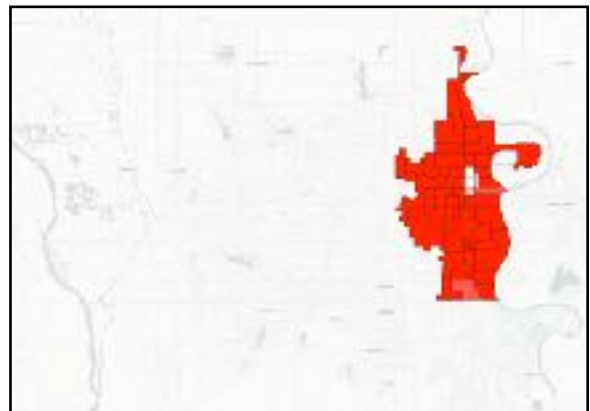
32. You've now mapped historical demographic information. This is a snapshot in time and says a lot about how segregated Omaha was already in 1950. But how has it changed over the next 50 years?

33. If you haven't already, download the census data from the Google Drive for the year 2000 to your hard drive. 'Omaha2000'. Add this layer to the map.

34. We can observe a few changes right off the bat. The census tracts covered by the census are far more expansive and now cover all of Douglas County. This should be unsurprising. You read Strand's article and know that Douglas County rapidly expanded westward after 1950 - part of a national suburbanization trend. You might also note that the census districts themselves have changed over time (shape and size). This can create subtle challenges for accurate assessment of historical change. Luckily, the census bureau understands this and they provide estimated changes by area over time. For now, we will ignore this challenge, though it might be something we return to for the final project.

35. The most dramatic changes are only visible if you look to the attribute table. Open it. We have a lot of new information, including state name, county, census tract, etc. The more recent censuses have been far more robust than in the past. This is equally true of the demographic information they collect, but

what is it? All we see are codes. Luckily, we have a codebook. Open it.

36. We see on the far right of the table that we have codes associated with total population, as well as various sub-groupings by race. We need to create t**wo new columns for our choropleth maps**. Percent white alone and percent black or African American alone as a ratio of total population.

37. Do so, using the above instructions if needed as guide. (24-28) If you'd like, you can do the same calculation for Asian alone, American Indian or Native Alaskan, Hawaiian or Pacific Islander, etc. Remember, we don't need to add all race columns together this time because we already have a column with total population.

38. Make two new choropleth maps. One that visualizes the ratio of Black or African American population by tract, and one for the white population. Take screenshots and paste them below.

You might be wondering, where is the column for the hispanic population we saw in 1950? Since 2000, the Census no longer considers 'hispanic' a racial category, though they still collect information on people who self-identify. This underscores the social constructedness of race, a concept that concept that is constantly evolving.

Download this pdf, insert maps, and answer questions. Submit to blueline.

Paste 4 Maps below - Black and White census tracts in 1950 and 2000 (Questions next page):

Questions:

1. What are the 5 tracts with the highest ratio of white population? How high a percentage? Where are they in Douglas County (what features surround or intersect them?)

2. What are the 5 tracts with the highest ratio of black or African American population? How high a percentage? Where are they in Douglas County (what features surround or intersect them?)

3. What spatial or demographic changes do you observe between the 1950 and 2000 datasets? What might account for the changes?