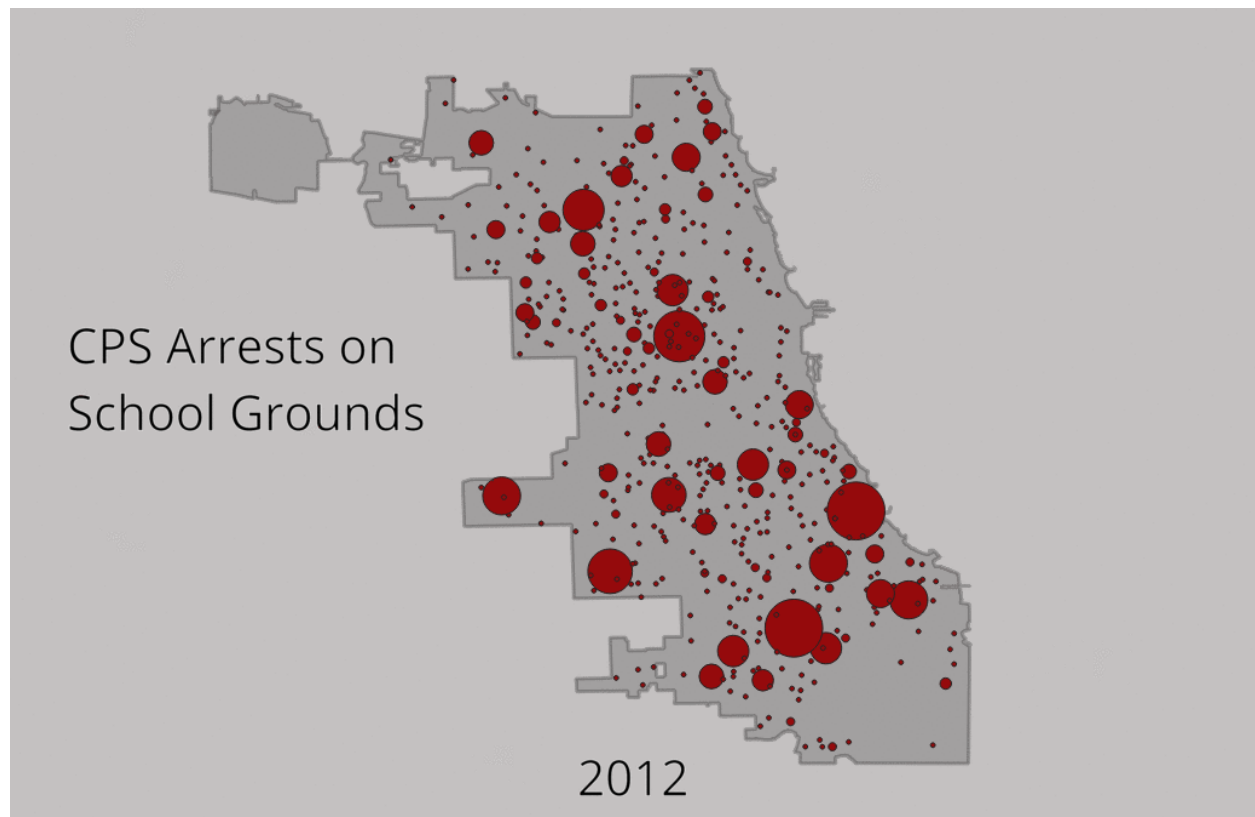I made a GIF for my plot so please refer to the link below to view the GIF.
https://drive.google.com/file/d/1dBpbrZyYtRWXdkGjAffA7VFVWoikTLBY/view?usp=sharing

What we learn from the map: Generally, over the years, it looks like less people are getting arrested on school grounds (note this map is ALL arrests on school grounds, so not just students arrested). And as expected, 2020 takes a dip since people aren't physically at school, reducing the number of arrests. Also, the south side of the city has more 'flare ups' than the north side of the city, with more large red circles than the northside, indicating more arrests take place at a given school in the south side than the north side.

Dataset used: I actually constructed this dataset. Chicago Public Schools (CPS) recently released (July 2020) arrest data on school grounds from the last 8 school years. I assume this data is in some spreadsheet form, but the only version of it I could get my hands on was a PDF summary a local newspaper made. This PDF had individual tables for each school (in CPS, any K-12 school), documenting the arrests across the years. I assume schools with no arrests in the last 8 years aren't included. I used a PDF -> CSV online scraping tool to get a CSV file out of this data, and then had to work some excel magic to get it to the form I wanted, which was a table with schools as the rows, school years as the columns, and total arrests in school year X for school Y as the corresponding value. I then used multiple datasets I found from the City of Chicago's Data Portal that lists all locations (address, x/y coordinates) of CPS schools for the 2011-2012, 2012-2013, and 2018-2019 school years. I then joined the x/y coordinates of schools to the dataset I had made, joining on the school names. Some schools had to be manually entered via a Google Maps search for x/y coordinates. Final dataset is then a table of

schools as rows, school years and x/y coordinates as columns, and arrests/latitudes/longitudes as values.

Producing the Map: I knew my end product was going to be a map of some sort, so I constructed the dataset in a way I knew would be easily digestible for the software to use. I used QGIS to make this map, and got a shapefile of the city boundary from the City of Chicago Data Portal. I knew QGIS would want the x and y coordinate pair which is why I included it in my dataset construction. I then used equal intervals, based on the 2013 data (highest arrest count in that year) to scale the sizes of the circles across the years. Each circle size is incremented by 8 additional arrests, so the smallest red dots are 0-8 arrests and if you see a dot increase or decrease it means the arrest count jumped into a new bucket. I didn't manipulate the data in any drastic ways, so I don't think I made any calculations so the end dataset and what ended up on the map is pretty raw data.

Limitations of the Map: There are quite a few limitations existing currently. First, I think it's misleading to include dots for schools that had 0 arrests in a year, since the dot is the same size/color as 1-8 arrests. Ideally, I think it would be nicer to either remove a circle when the arrest count is 0, or change its color to indicate that this school had no arrests (which could also be interesting to see as a GIF, how the colors change as arrests decrease). Secondly, this doesn't account for school closings. It's very likely that total arrest count in CPS hasn't actually changed very much, but the opening and closing of schools distributes arrests in a way that would make the circles look like they're getting smaller, misleading the viewer to believe arrests are decreasing. For example, one big red circle could become 2 medium sized circles the following year, if the school closed and students were pushed to other neighboring schools. This would look like a decrease to the viewer of the GIF. Lastly, this map could be misleading because relative school populations matter. Maybe the arrest *rates* are similar across the city, but if there are less schools on the south side (a sad but, true phenomena) then arrest counts would look large to represent a larger student body. An updated map could size the circles relative to their arrest rate (# arrests/# students in school Y in year X).

Takeaways: Map making is pretty hard, I spent a lot of time thinking about the best way to represent this data, and I also realized in the process that it is VERY easy to make data appear misleading. For example, in the software I used, QGIS, at first the dots I had were perfectly evenly distributed across the city for a given year, which, given my understanding of Chicago, I thought was fake. I realized this was because of the scaling of the circle sizes -> it was set to make the 'buckets' percentiles which heavily distorted the map since most schools in the city have <10 arrests in a given school year. As a result, the lower 95th percentiles had 0-10 arrests and all the outliers were clumped together. I thought equal interval spacing made most sense after seeing the shift when I changed the settings.

17.271 "Mass Incarceration in the United States"
Fall 2020