Esin Onal

Assignment 1

The dataset selected was one that reported different incidents of crime in Chicago in 2020. The data was taken from the Chicago Data Portal. The month of September 2020 was selected, which reduced the observations down from 165,000 to 17,304. The rows missing the X/Y and longitude/latitude were removed – this was acceptable as there were only 193 rows.

The data had to be cleaned in Python before being imported into Geoda, as there were several functions I used which could not be done in Geoda. The date/time was not being recognized in Geoda (Month/Day/Year), so I had to write a function to change this to Year/Month/Day format. Separately, I had a column with "True/False" for "Arrest" that I wanted to change to 1s and 0s, which I also had to implement in Python.

Next, several variables were deleted, namely "Updated on," which was unnecessary, and "Location," which was already contained in the Long/Lat. The variables that were kept were Type of crime, Description of crime, Arrest, and X/Y. The X/Y information was used to create a point map using the EPSG:3435 projection for Chicago.

Next, the community boundaries were found from the City of Chicago open data portal. The projection information for EPSG:3435 was also entered into this layer, so both layers could be in the same projection.

After getting both the community areas base map and the point layer with the cleaned information, a spatial join was done for both the number of crimes (NumCrimes) in each community, as well as the number of arrests made in the community (NumArrests). These were added to the table. Using the CMAP data hub, the population data by community areas was obtained. These populations were merged into the community base layer, allowing the rate to be obtained for rate of crimes and rate of arrests (arrestsPerC and crimPerCap).

The instantiation of the areal boundaries as communities in Chicago was made as different communities are known to have different crime rates, and also because community boundaries were the largest and most global boundaries possible inside the city of Chicago. The number of arrests made and number of crimes that occurred had to be converted to densities using the populations of the communities, in order to get an accurate rate.

Several maps were created to illustrate these rates. Map 1 shows the distribution of the crimes throughout the city of Chicago. Map 2 shows the crime per capita, with both a percentile map and a quantile map. Map 3 shows the arrests per capita, with both a percentile map and a box map (1.5 hinged). Map 4 contrasts arrests per capita and crimes per capita, with both percentile as well as quantile maps.

By brushing, it was possible to see that the highest point of crime density per capita is West Garfield Park (see Map 2). Its surrounding areas are high too (East Garfield Park and North Lawndale). A separate section that also has similar crime rates is Englewood, West Englewood, Chatham, and Greater Grand Crossing in the South. These areas are all in the top 10% for high crime. There are two separate areas for low crime, the North side and the South West side. The low 10% of crime includes Lincoln Park, four communities in the very north of Chicago, as well as three in the South West. It is

clear that higher crime rates are found in the South and Western suburbs of Chicago, while the lower crime rates are in the North and the South West.

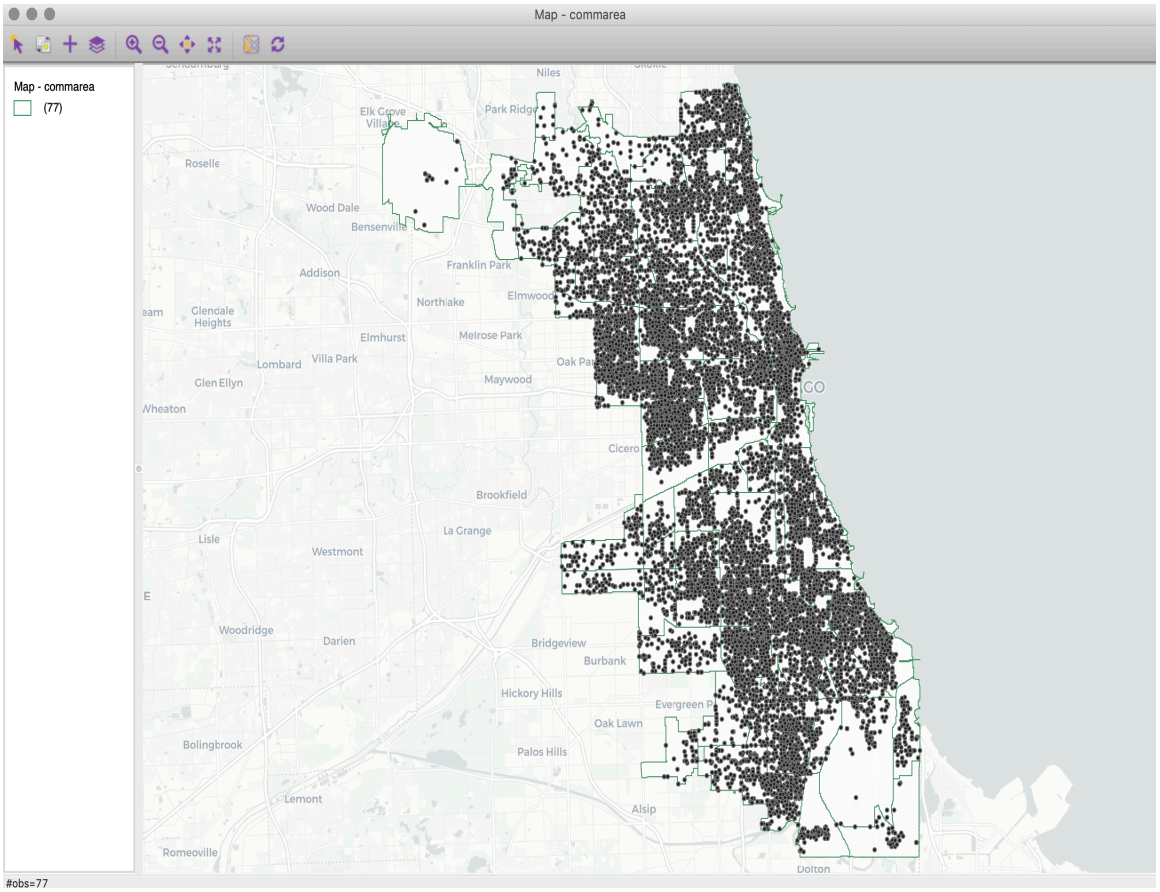
When it comes to the arrest rate, it is likewise possible to see that the higher rates of arrest occur in the West and the South, while the North and South West of Chicago have lower rates of arrest. (see Map 3) In particular, the highest rates are in North Lawndale, and East and West Garfield Park.

An interesting research question that could be proposed for this data set is: are people in the communities in the South of Chicago (with more Black communities) more frequently arrested comparative to the crime rate? Or in general, are some communities more frequently arrested relative to the crime rate? If so, this could suggest a racial reason.

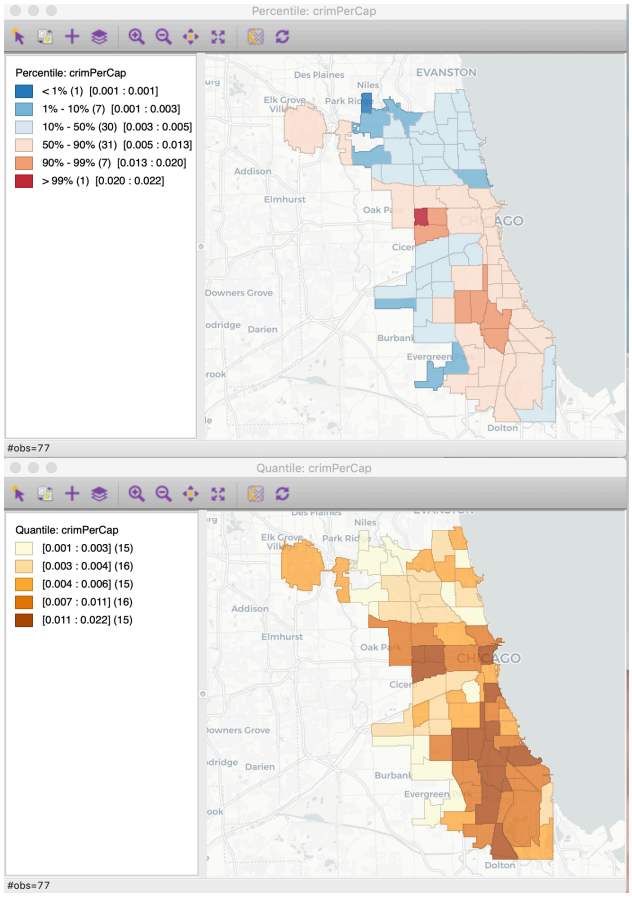
Looking at the percentile maps for crime rate and arrest rate (see Map 4), the percentile maps for crime rate and the arrest rate are very similar. It seems as though there isn't a disproportionate amount of arrests in the south relative to the crimes committed. The rate of crimes committed and the rate of arrests do seem proportionate.

However, while in the percentile map the rates appear to be the same, the quantile map seems to tell a different story. There seems to be a higher rate of arrest in the West, comparative to the crime rate. Additionally in the south, some communities have a darker color for arrest than they do for rate. Some such examples are New City, and Avalon Park. There are some communities with an arrest rate that is lower than the crime rate, and with a large number of these in the North side of Chicago: North Park, Albany Park, Irving Park, and Lincoln Square. It does appear that the Northern arrest rate may be lower than the crime rate, while for some communities in the South the reverse is true.

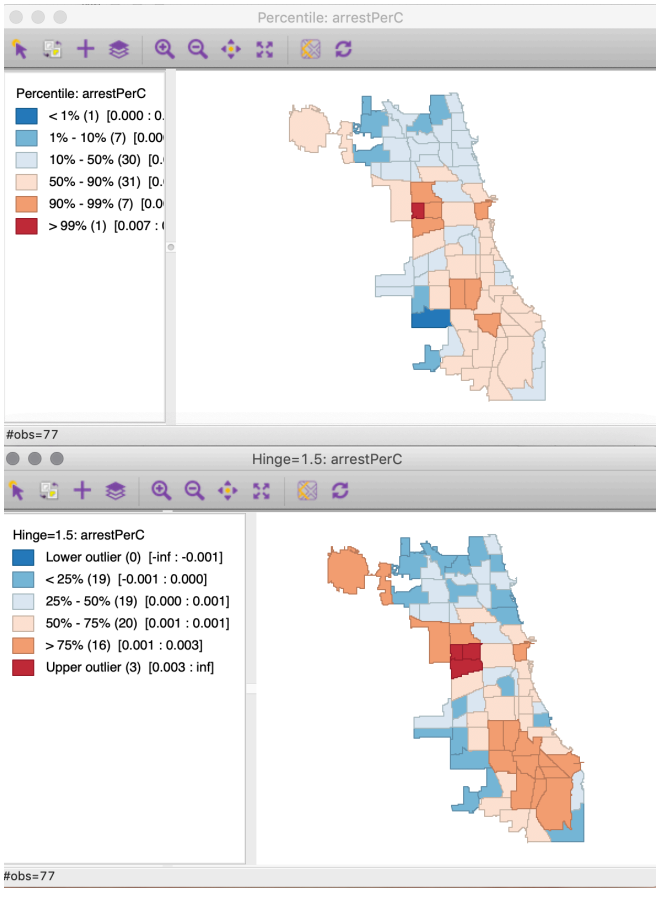
While this does seem to suggest a racial implication, it is nonetheless still possible that the severity of the crimes were worse in these areas, making it more necessary for arrests to take place. While there are some differences in crime rate compared to arrest rate in the different neighborhoods of Chicago that leave room for a racial implication, there is likely not enough evidence to conclude this for certain. In addition, a further analysis separating the crime types into different bins may be necessary, in order to ensure that the same crimes are being compared to each other. For example, is there a higher rate of arrest for lower significance crimes in the South than in the North of Chicago?



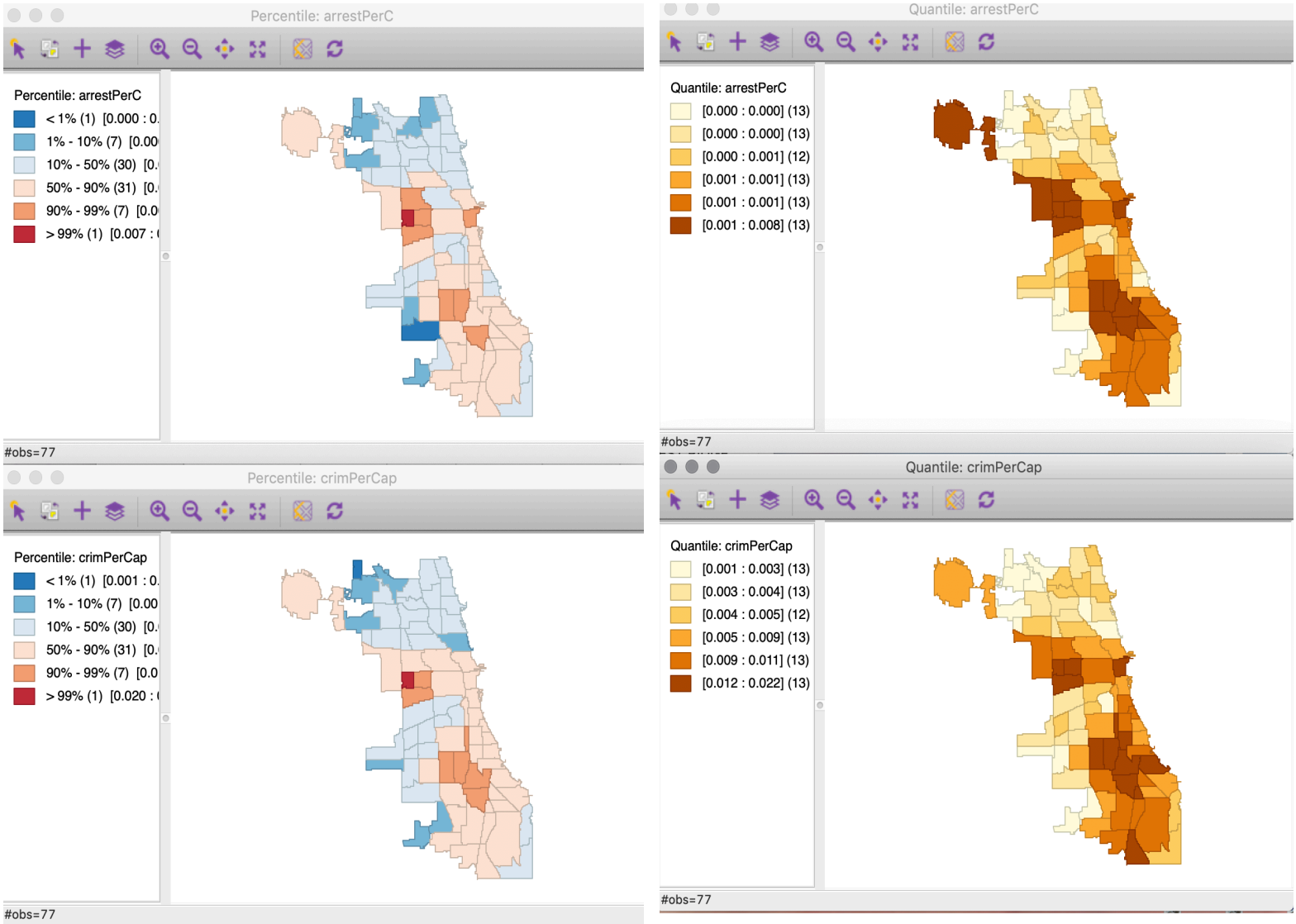
Map1: Point map of distribution of crimes, with the community boundaries underneath.



Map 2: Different views of crime per capita: Both a percentile map and a quantile map.



Map 3: Different views of arrest per capita: Both a pencentile map and a box map (1.5 hinged).



Map 4: Contrasting the percentile and quantile maps of crime per capita and arrest per capita.