**Income Disparities and Racial Segregation in Cook County**

1. **Introduction**

A recent study found that Chicago is the 13th most segregated metro area in the United States. Even though the study observes an improvement from 2009, it shows that in order to have a more integrated city, 76% of African Americans and 57% of hispanics would have to move to other neighborhoods. In addition to segregation, it is generally understood that the West and South sides of Chicago and its suburbs are racially concentrated areas of poverty. Racial segregation is accompanied by economic segregation which in the end leads to further deepening of the income gap by race and ethnicity.

Having a clear picture of what segregation looks like in the Chicago area and its contiguous suburbs, is very important for decision makers and advocacy groups. That way they can identify the disadvantaged populations and where they live,  allowing them to design appropriate policies in order to improve the living situation of these minorities.

The goal of our project is to provide an exploratory analysis where the viewer can interact with our visualizations and learn about the racial and ethnic segregation as well as of the income inequality in Chicago. We do this using two approaches. First, we explore income data by race and ethnicity to see the percentage of each group by income bracket, allowing us to have an idea of the income disparities present in Cook County and Chicago and to identify the most vulnerable populations in terms of income. Second, to show the racial and ethnicity distribution and segregation, we analyze Census demographic data for seven bordering neighborhoods of the West side of the city that serve as a perfect example of this problem.

This paper is organized in the following manner. Section 2 talks about the data and methodology that we used, first for the income analysis and then for the demographic analysis, explaining the techniques implemented in these visualizations for both of our analysis. Section 3 talks and. Section 4 discusses the suggested conclusions of the data exploration.

1. **Methods**

**2.1 Data**

For the income analysis we obtained the data from the 2010 United State Census at the Cook County level. The income data is at the Household (HH) level and we constructed a data set in R showing the percentages of HH in each income bracket divided either by race or by ethnicity, as well as the total number of HH. For the analysis we focused only on the brackets above the median income for the county, which is approximately $55,251 a year.

The data for the demographic analysis was also obtained from the 2010 United States Census. We downloaded the information at a Census Tract level for the state of Illinois and filtered for the tracts that we were specifically interested in. Later, we manually constructed the data set at a neighborhood level so that we could compare the seven neighborhoods we focus in.

The reason we chose these seven neighborhoods is because, as explained in the introduction, the West side of Chicago is known for the high level of racial/ethnicity segregation present in the area. This small sample of bordering neighborhoods are important because of the striking differences in populations within the area. The seven neighborhoods are: Austin, Belwood, Berwin, Cicero, Forest Park, Maywood and Oak Park.

It is important to mention that the United States Census demographic data identifies people by their race (African American, White, Asian, Native American, etc.) and by ethnicity (Hispanic or not Hispanic). This distinction causes many statistical problems because Hispanics are many times misclassified into a race they don't necessarily belong to. A Hispanic might identify as "White" or as "other" or as two or more races and it complicates how the data are analysed.

**2.2 Visualizations**

To present the information we were interested in, we designed a flexdashboard and interactive plotly graphs. The flexdashboard is very easy to navigate since it is divided in four tabs: an introductory tab with a quick explanation of the objective of our project as well as a reference to our data and reasoning behind the differences between race and ethnicity. The second tab has three plots with income data for Cook County by race and ethnicity. And finally, the third tab includes four plots that explain the racial and ethnicity composition of the seven neighborhoods.

We decided to use plotly interactive bar plots since the combination of them with a dashboard are key to allow the user to make comparisons inside each tab. Bar graphs are easily interpreted by any kind of user since they are simple and aesthetically pleasing. The plotly interactive visualizations allow the user to hover over the bars to see the corresponding percentages for each bar. Plotlys' interactivity is specifically useful to select specific categories (neighborhoods, races or ethnicities) and compare them to others. The color pallet we selected is pleasing and we believe that the graphs show only the necessary information to be self- explanatory. Using this type of interactivity invites the user to explore and understand the data better without that much information by looking at one feature at the time without the need to do several graphs to explain each feature in an isolated manner.

The reason we chose to use Ploty is because, being an exploratory analysis, this visualization technique allows the user to select any race/ethnicity or neighborhood of interest to be able to understand this complicated issue a little further. Also, given that we believe that our target population can be anyone curious about this issue, and not necessarily someone with background on this subject, we wanted to make this as easy to handle as possible and with many different ways to approach it so the user can go as deep in the analysis as they wants to. Plotly allow us to have beautiful, attractive visualization that are not necessarily interactive unless the user wants them to be.

In the income tab we included three plots. The first plot is the percentage of households in Chicago that have an income above the median by race. The intention of this graph is to show the differences in income by race and ethnicity and visualize which groups have a larger proportion of households that earn an income above the median. The second and third plots show the proportion by race or ethnicity by income bracket for those who have an income above the median. The interactivity of the plotly bar graphs allows the user to select one specific race or ethnicity and observe how, as the income brackets increase/decrease the income patterns for each race and ethnicity behave.

For the demographic analysis, the first set of graphs, at the top of the tab, show the population in thousands of people per neighborhood divided by ethnicity and then by race. The aim of these pair of graphs is to give an overview of the population density and distribution per neighborhood. The interactivity allows the user to focus in the total population or in an ethnicity or race specifically. The second pair of graphs, shown at the bottom of the tab, show the distribution, in percentages, of these population groups within each neighborhood, allowing the user to select a desired neighborhood and see it’s specific racial and ethnic composition.

1. **Discussion**

Even though our demographic and income analyses are at different levels, we believe our exploratory analysis is very useful to identify, first, with our income analysis, the most vulnerable populations in the County: African Americans and Hispanics. Later, the demographic analysis allows us to see an example of the isolation and segregation of these groups in a sample of neighborhoods in the West side of Chicago.

From the income analysis, we can conclude that African-Americans and Hispanics are two disadvantaged groups when compared with Asians and Whites. 60.12% of Asian HH in Cook County earn an income above the median ($55,251), for whites this number is 57.92%. In contrast, only 40.84% of Hispanics and 32.78% of African American HH earn an income above the median. This is a clear representation of how African Americans and Hispanics are disadvantaged. Later, we learn that as the income brackets increase, Whites and Asians earn more money while we see a decrease in the proportion of African Americans and others. The same pattern is seen with the Hispanic and White not hispanic population.

The demographic analysis’ main findings are that Austin, Bellwood and Maywood are mostly African American communities, while Berwin and Cicero are mainly Hispanic or Latino communities while, Forest Park, Oak Park are white non-Hispanic majority communities. In terms of the population sizes, it is relevant to notice that the three most populated neighborhoods are also populated by disadvantaged groups (as shown in the income analysis), such as African American (Austin) or Hispanic (Cicero and Berwin).

Furthermore, the analysis within community shows that the striking differences in racial and ethnic distribution in each neighborhood. For example, by selecting Cicero and Austin in both graphs, we can see that they have extremely different populations even though they are right next to each other. While Cicero’s population is 80% hispanic, Austin’s is 90% African American, which is a perfect example of segregation around the West side.

Alberto Cairo’s framework for qualities of great visualizations argues that all graphs must be truthful, functional, esthetic, insightful and enlightening.  We will analyze the plots based on this framework:

* The visualizations presented for our project are truthful since we did not manipulate the data in order to fit a story or misrepresent it.
* We believe that these animated visualizations are extremely functional. The graphs show in a plain way the income gap between races and ethnicity and distribution of the population in each neighborhood by race and ethnicity. As we mentioned earlier, plotly allows the user to explore and analyze the information in order to reach his or her own conclusions. Additionally, the interactivity is helpful in a sense that if you see the graph in a static way, it renders the desired information but if as a user you want to know more details, you can hover and see the exact percentages for each group.
* The bar graphs are aesthetically pleasing. They are all clear and concise without having extra features that are not needed to have a basic understanding of the graph. The titles, axis and legends of each graph do a good job a telling the whole story. The choice of color pallet is attractive and useful to see the differences between races and ethnicities.
* Our visualisations are very insightful, since they reveal the story that we are trying to tell very evidently. The general public can understand the graphs without any further explanations: income inequality in Cook County affects mostly African Americans and to a lesser extent Hispanics. Also, the composition of bordering neighborhoods in the Chicago area are very racially segregated.
* Most people that have been to Chicago know that it is a very segregated city with many problems including income inequality. For this project we wanted to show the readers how income varies by race and ethnicity in the Chicagoland area. This is why we believe that our visualizations are enlightening.

1. **Conclusion**

The analysis performed in this project consisted on an exploration of Census data in order to communicate in an interactive and simple form and to any kind of audience the importance of income inequality and racial/ethnic segregation in Chicago. We use this by first showing that Hispanic and African American populations are in disadvantage compared to Whites and Asians in terms of income at the Cook County level and; second, by selecting a sample of communities in the west side of Chicago to show the high levels of segregation by race and ethnicity. We conclude that both vulnerable groups are highly isolated in these communities. We hope that this kind of analysis is useful for policy makers and advocacy groups, given that it is easy to identify a big part of the problem, the target population for policies aimed at reducing this disparities.