

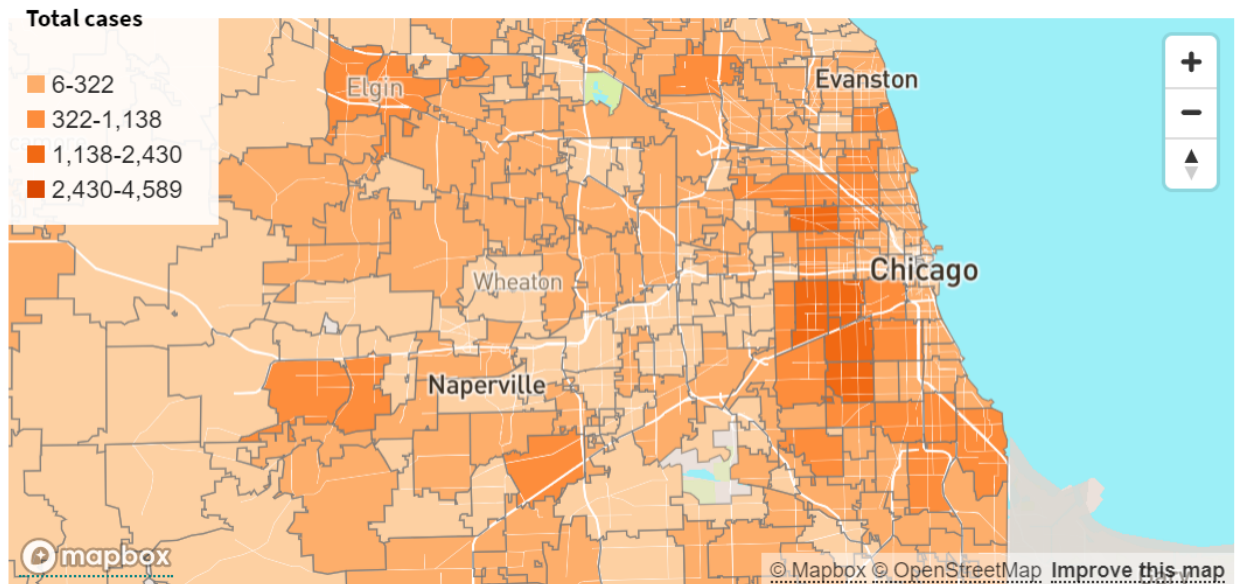
House crowding drives covid incidence in Chicago

Daniel Gervini

August 30, 2020

Chicago by zip

The following map, taken from [here](#), shows coronavirus cases by zip code in Chicago:



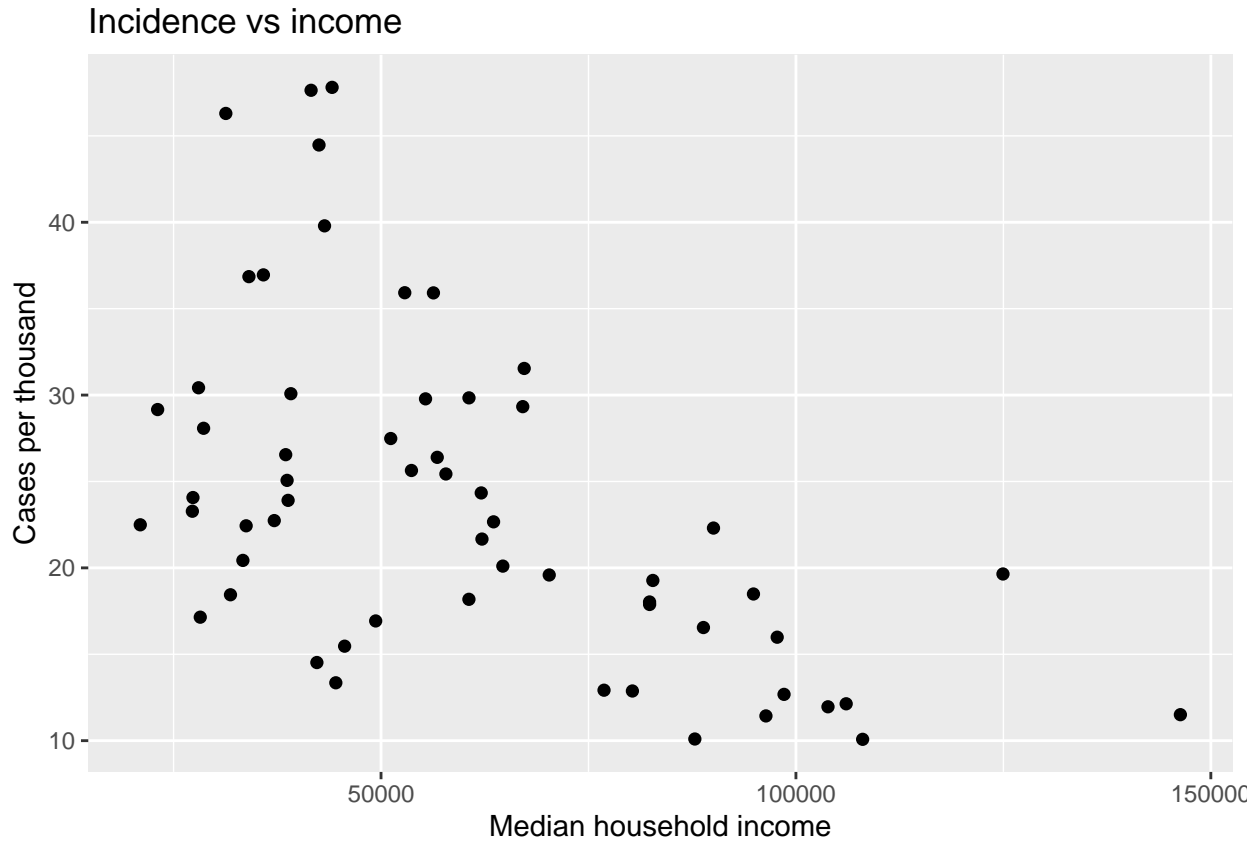
For those of us who know the area, the first thing that catches the eye is that the neighborhoods with highest incidence are the predominantly Hispanic neighborhoods of the West side, but *only* those neighborhoods. The predominantly black neighborhoods of the South side, although poorer, show a lower incidence of coronavirus. So, poverty per se cannot be the driver of coronavirus incidence, nor can population density, since the predominantly white neighborhoods of the North side are much more densely populated but show a much lower incidence.

In this post we study the association between coronavirus incidence, measured as confirmed cases by 1000 people, and each of the following demographic variables: median annual household income, population density, and average household size, for each zip code area of Chicago. The demographic data was obtained [here](#) and the coronavirus incidence data was downloaded from [here](#).

(For the analyses below we removed one outlier, zip code 60604, corresponding to part of the Loop. This zip code had the lowest population (627), the highest income (\$156,356) and the highest incidence rate (49.4 per thousand). It was the only anomalous case we detected among the 60 zip code areas.)

Incidence versus annual income

The following graph shows a scatterplot of coronavirus incidence versus median annual income for each zip code.

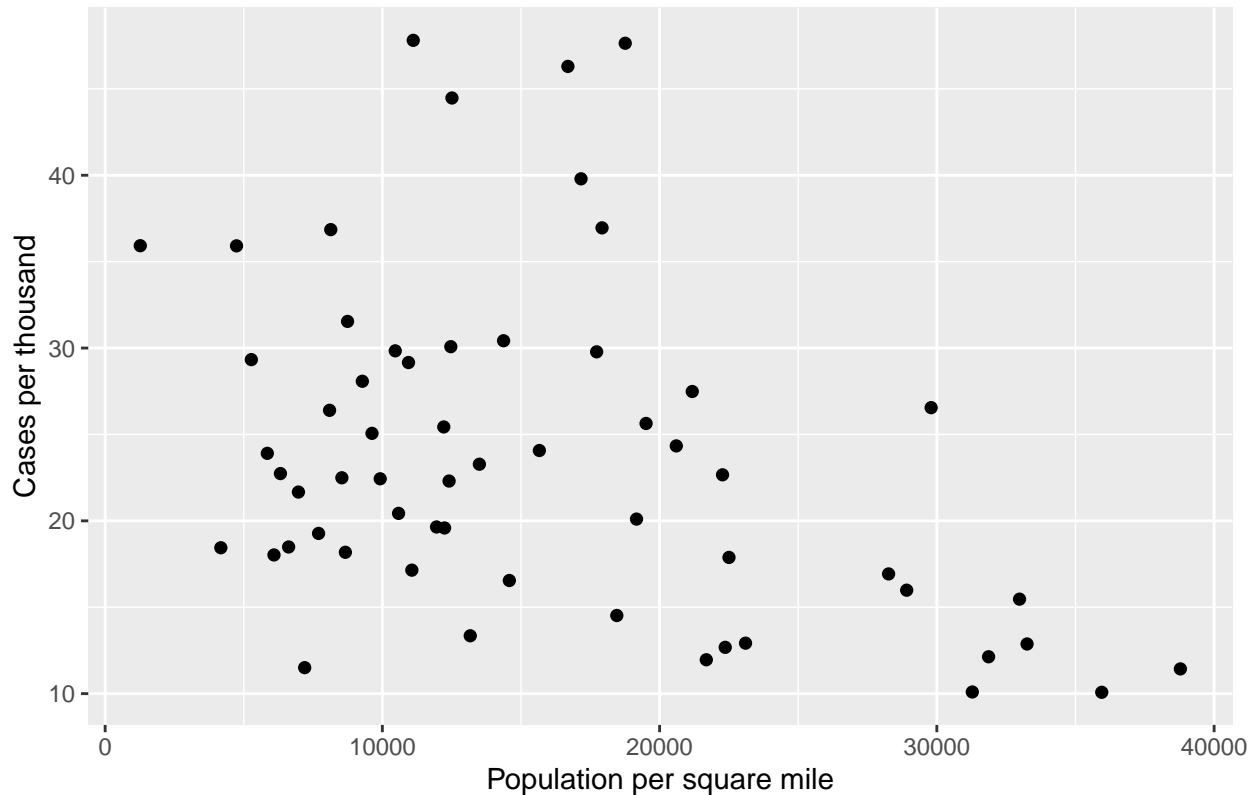


We can see a clear association between coronavirus incidence and income: the higher the income, the lower the incidence. However, the association is not in the form of a tight functional relationship between the variables but in the form of two broad groups: for incomes above \$75,000 the average incidence is 15 cases per thousand, while for incomes below \$75,000 the average incidence is 27 cases per thousand. Within these two broad groups, however, there are no clear trends.

Incidence versus population density

The following graph shows a scatterplot of coronavirus incidence versus population density, measured by the number of people per square mile, for each zip code in Chicago.

Incidence vs population density



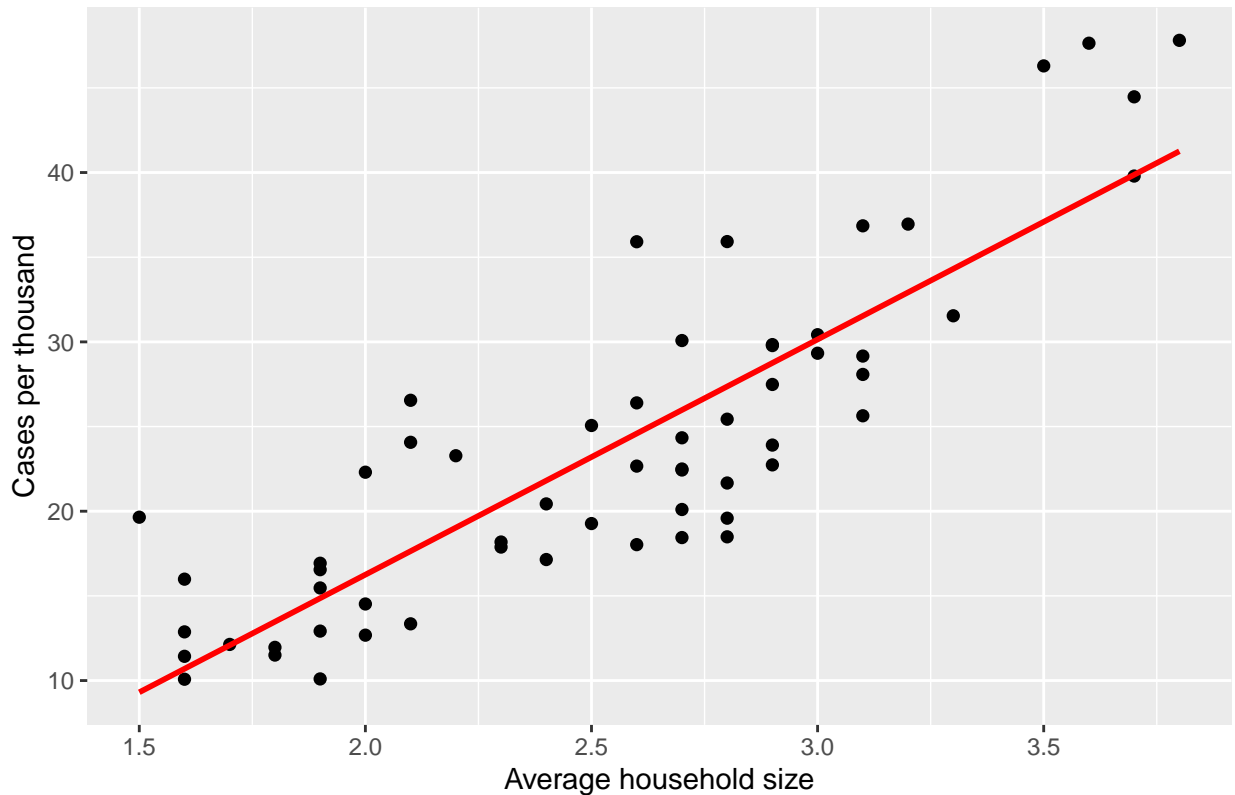
We see again a clear but unexpected association between coronavirus incidence and population density: the higher the density, the **lower** the incidence. As before, the association is not in the form of a tight functional relationship but in the form of two broad groups: for densities above 25,000 persons per square mile the average incidence is 15 cases per thousand, while for densities below that threshold the average incidence is 25 cases per thousand. As before, there are no clear trends within the groups.

The fact is that the group of high-density zip codes is largely the same as the group of high-income zip codes, corresponding to Loop and North side neighborhoods, where living conditions are clean and spacious. In any case, it is clear that a higher population density, per se, is not associated with a higher incidence of coronavirus in Chicago; in fact, the opposite is true.

Incidence versus household size

Now we come to the interesting part. The following graph shows a scatterplot of coronavirus incidence versus average household size for each zip code in Chicago.

Incidence vs household size



Now we finally see an unequivocal and strong linear relationship between coronavirus incidence and a demographic variable, household size. In retrospect, it seems obvious that it had to be that way, but this goes against the prevailing view that the main driver of coronavirus spread is outdoor mobility. Here in Chicago, at least, more people are getting infected inside their homes than outside.

We can see why coronavirus incidence is so low in the affluent North side: while densely populated, household sizes tend to be much smaller than those of the Hispanic West side, where it is not uncommon for large extended families to live together under the same roof. Households in the South side, on the other hand, are generally poorer but smaller than those of the West side, which explains why coronavirus incidence is lower there.

Some extrapolations

Although extrapolating is always risky, one would suspect that the same is going on in other cities around the world. For example, Tokyo, one of the most densely populated cities in the world, has only 19,800 confirmed cases to date, or 2 per thousand, compared with the 28 per thousand of New York or the 31 per thousand of Buenos Aires.

Generally speaking, developing countries with large cities and overcrowded housing (Argentina, Brasil, Mexico, South Africa, India, Iran) have shown the same pattern of unchecked exponential growth, unlike European countries and, if one is to believe the official statistics, China. The United States has shown an intermediate pattern between Latin America and Europe, probably owing, at least in part, to the levels of house crowding that are prevalent in many of its largest cities (Chicago, New York, Los Angeles, Miami) and are not seen in most European countries.

[Back to blog](#)