DePaul University
University Libraries

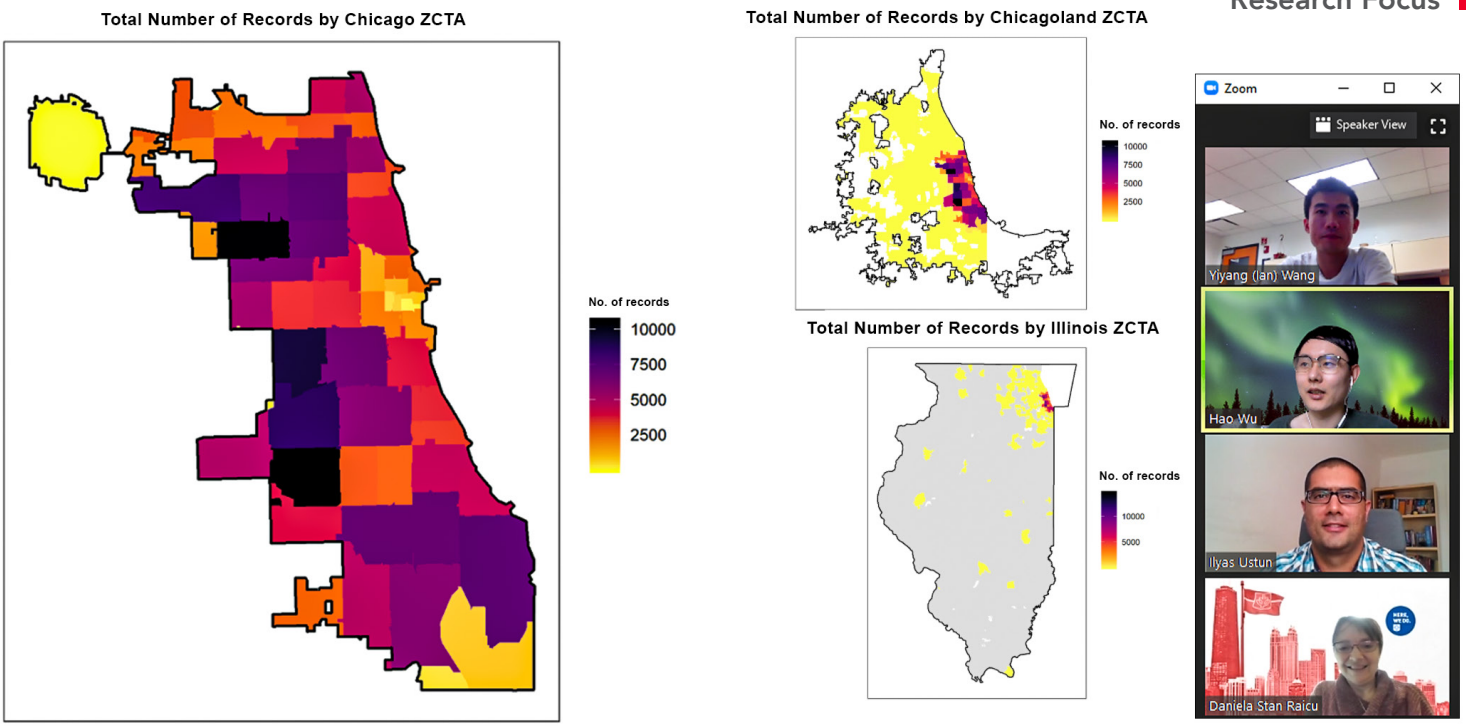
In The Loop

Volume 1
Issue 15 *Fall 2020*

Article 7

9-8-2020

Data Detectives



Team members Yiyang Wang, Hao Wu, Ilyas Ustun and Daniela Stan Raicu (far right, top to bottom) discuss their COVID-19 data research during a Zoom meeting.

Data Detectives

CDM helps Chicago track the racial proportions of its COVID-19 cases

A few months after the COVID-19 pandemic spread to Chicago, it became clear from records collected at testing sites that Black and Latinx residents represented a disproportionately higher percentage of cases than the city's other racial populations.

"But there might be some neighborhoods hit more than others in terms of their racial distribution," says Daniela Stan Raicu, associate provost for research and co-director of the Medical Informatics Lab, which is affiliated with the DePaul Center for Data Science. "So our project is about helping the city distribute resources to the communities that need them the most."

The project is a collaboration between DePaul and the Chicago Department of Public Health (CDPH) to fill in missing racial data in the city's COVID-19 case surveillance system to facilitate a more granularly informed response.

Last May, Nikhil Prachand, CDPH's director of epidemiology, reached out for help through Professor Fernando De Maio, who specializes in medical sociology and social epidemiology. De Maio is part of the DePaul project team that includes Raicu; fellow School of Computing (SoC) faculty members Jacob Furst and Ilyas Ustun; C. Scott Smith, a geographer, urban planner and assistant director of DePaul's Chaddick Institute for Metropolitan Development; and SoC graduate students Hao Wu and Yiyang Wang.

At testing sites, patients are asked to provide their first and last name, address and race. However, nearly half of the 256,483 records CDPH collected through June 18 were missing the race detail. The DePaul team set about trying to fill in the blanks.

The team used an algorithm that formulates the likelihood of surnames and geocoded home addresses sourced from U.S. Census Bureau data sets to impute a particular race. By July, the team had secured all but 11% of the missing information within an 81% accuracy metric.

The racial mix before and after filling the missing values increased from 17.02% to 29.29% for Black patients, 16.28% to 28.36% for Latinx patients and 13.87% to 24.17% for white patients. Those results and others will be used by CDPH as part of a larger effort to learn how many patients tested actually had COVID-19 and how many died.

"We did find the Latinx and Black groups have more cases," says Wu, a DePaul Center for Data Science graduate assistant, who ran the city's stats through the model. He presented findings to the team during weekly online meetings. "I'm grateful for this opportunity to contribute to a real-world project in the city I live in," he says.

Using the algorithm as a basis, Ustun has designed an application that lets CDPH researchers quickly receive calculated racial information when they enter surnames and zip codes. The team hopes to build other factors into the app, such as occupation, to improve its COVID-19 predictive modeling.

"As more data become available, we can expand the tool," says Raicu. "And if we can make it work for Chicago, it could also be used by other cities."