

Covid in Chicago

Exploring Data

When the pandemic first started (March-April 2020), what new things did you start doing with all of your extra time?

On your each of your post-its, write down one thing that **was known** and one thing that **was not known** about Covid in the **early days** of the pandemic with respect to the disease itself (origins, how it spread, symptoms, who it affected more, etc.)

Notebook 7 (Part 1) Tasks

1. Normalize the number of deaths in Chicago by creating a new column: "DeathsPer1000"
2. Create a line graph with title and labeled axes showing the change in death rates between 2012 and 2021.

How can we best measure Covid's impact on different communities in Chicago?

Our data...

In May 2020, Kevin Credit from the Center for Spatial Data Science at the UChicago wrote "Neighborhood inequity: Exploring the factors underlying racial and ethnic disparities in COVID-19 testing and infection rates using ZIP code data in Chicago and New York".

Kevin's research also would not have been possible without open source research data from: Illinois Department of Health:

<https://www.dph.illinois.gov/covid19/covid19-statistics>

U.S. Census Bureau: <https://data.census.gov/cedsci>



Notebook 7 (Part 2) Tasks

1. Explore the available variables and their values
2. Identify variables that could be used to measure Covid's impact.

What variables in the dataset could be used to measure Covid's impact?

Which variable is better to describe Covid's impact?

Cases: The number of people that tested positive



Tests: The number of people that were tested



Notebook 7 (Part 3) Tasks

1. Generate a new column of data for the week of 4/16 called “case_rate_4_16” using the “case_rate” function.
2. Make a function called “test_rate” that normalizes test rate
3. Generate a new column of data for the week of 4/16 called “test_rate_4_16” using the “test_rate” function.

Which variable is better to describe Covid's impact?

Let's say there is a building containing 10 people.

Let's say 5 out of the 10 people have COVID.

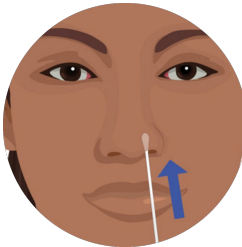
But only 2 out of the 10 people get tested -- 1 comes back positive, 1 comes back negative.

Therefore:

$$\text{case_rate} = 1/10 = 10\%$$

$$\text{testing_rate} = 2/10 = 20\%$$

Reducing the effect of testing...



$\times 100\% = \text{Positivity rate}$

In areas with low testing...



$\times 100\% = \text{Positivity rate}$

In areas with high testing...



$\times 100\% = \text{Positivity rate}$



Notebook 7 (Part 4) Tasks

1. Define a function called `positivity_rate` that calculates positivity rate
2. Create a new column in the dataframe for the week of 4/16 called “pos_rate_4_16” and view the new data



Exploring Covid with Kepler

1. Go to **kepler.gl** on your browser
2. Download the Covid dataset from Google Drive

`cov_chi_with_positivity.geojson`

1. Expand the layer with the “v” button
2. Click the three vertical dots next to “Fill Color”
3. Change “Color Based On” to `pos_rate_4_16`
4. Click on the color scale and check the “Reversed” toggle

“Why has Covid affected some communities more than others?”



How can a community and its people be described with data?



Notebook 8 Task

Explore a total of 4 explanatory variables. Record what you learn from each in the following table:

Explanatory Variable	Correlation Type	R value	Notes
Median Household Income	X	X	X
X	X	X	X
X	X	X	X
X	X	X	X