

VAX-CHI-NATION

Effectiveness of Vaccination Campaigns in Chicago



Group Name and members:

Team 1

- Josiah Chung
- Kshitij Mittal
- Danil Meresenschi
- Yash Verma

Executive summary:

Over the last couple of years, the COVID-19 pandemic has caused significant structural changes in the way healthcare, public policy, business, and other integral areas operate worldwide. With the pandemic and rise of COVID-19 cases have come subsequent efforts to vaccinate our communities. Studies have shown that increased COVID-19 vaccination rates lower transmission, hospitalization, and death outcomes which, in turn, saves financial resources that can otherwise be allocated elsewhere. Looking back at the regulations that Chicago established to increase vaccination rates in its population, what insights can be drawn from this experience and how can we apply such lessons to future healthcare processes?

Business case:

The city of Chicago has a robust vaccination program - named VaxChiNation. Our aim is to understand if this investment has yielded significant improvements in the public health scenario in Chicago. With a focus on descriptive statistics, we will analyze the vaccination rates and other relevant COVID-19 data metrics and visualize our findings on a dashboard. To measure the success of our campaign, we will compare our findings and measurements to the US National average and/or figures from similar cities in the Midwest. By analyzing our own vaccination efforts and comparing them to others, we will evaluate what methodologies we believe have worked and what areas we can further optimize and improve for future campaigns.

Data sources:

- Chicago Community Vulnerability Index
<https://data.world/cityofchicago/xhc6-88s9>
- Location of Testing Sites in Chicago
<https://data.world/cityofchicago/thdn-3grx>
- Vaccination Locations
<https://data.world/cityofchicago/6q3z-9maq>
- City of Chicago - Vaccination by Zip Code
https://data.cityofchicago.org/widgets/553k-3xzc?mobile_redirect=true
- City of Chicago - Covid tests, cases and deaths by Zip code
<https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v>

Database & tools:

SQL(MySQL), Tableau, MS Excel

Part 1: Data Analysis and Data Model

Data preparation/cleansing steps needed/proposed:

- We needed to generate unique IDs for many of the target tables in our dataset, including ones for zip code, vaccination location, and testing location. These unique IDs all auto-increment.
- We compiled all of the zip codes available to us from different sources into a central data set and assigned a population value to each zip code. The population value was derived from another data set.
- Some tables were aggregated on the basis of zip codes and dates
- We filtered the total available dates to a smaller window that ranged from **December 2020 to June 2022** in order to focus our analysis.
- We dropped all aggregate columns, such as percent rates and proportion ratios, to avoid redundancy in our analysis.
- We took the date column from an original dataset and populated a separate date table with full date, date ID, month, and year.
- For normalization purposes, we substituted column name conventions with distinct keys. For example, instead of stating full_date, simply provided date_id foreign key that connects to the dim_date table.
- We transformed some columns in the fact table to reflect the sum of vaccination numbers for specific age groups.
- We also filtered CCVI Scores based on zip codes, from a larger dataset with rankings that yield the score.

Database platform considerations

- We used MS Excel for initial data cleaning
- We then used MySQL for both modeling and analysis.
- Finally, we used Tableau for the BI dashboards .

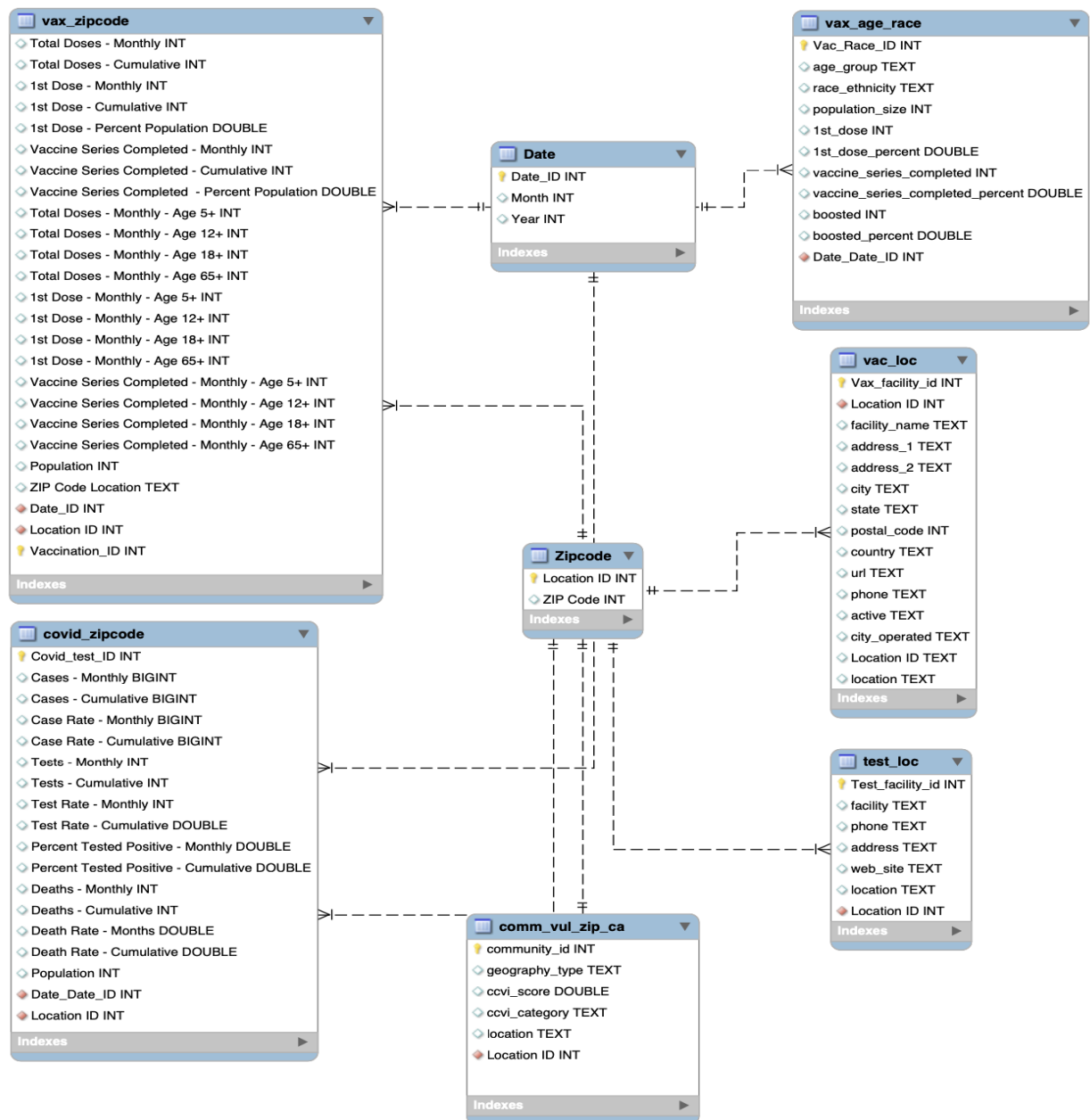
Any necessary assumptions around our datasets:

- One assumption that we are making is that every vaccination done in Chicago was properly and accurately reported. In reality there could have been mistakes or absences of reporting in some cases.
- Another assumption that we are making is that the testing and vaccination locations in Chicago from our datasets account for every testing and vaccination location per zip code.
In reality, this is probably not the case, vaccinations and tests could have frequently been provided through mobile/temporary clinics that can't be tied to an address.
- Vaccinations could have been done in different zip codes between the first and second doses.

We modeled 5 data tables as a part of a snowflake model. Each of these tables are being populated from 6 publicly available datasets (portal - City of Chicago).

We first tried building a **relational model** from the original datasets. The relational model was taking the following preliminary form.

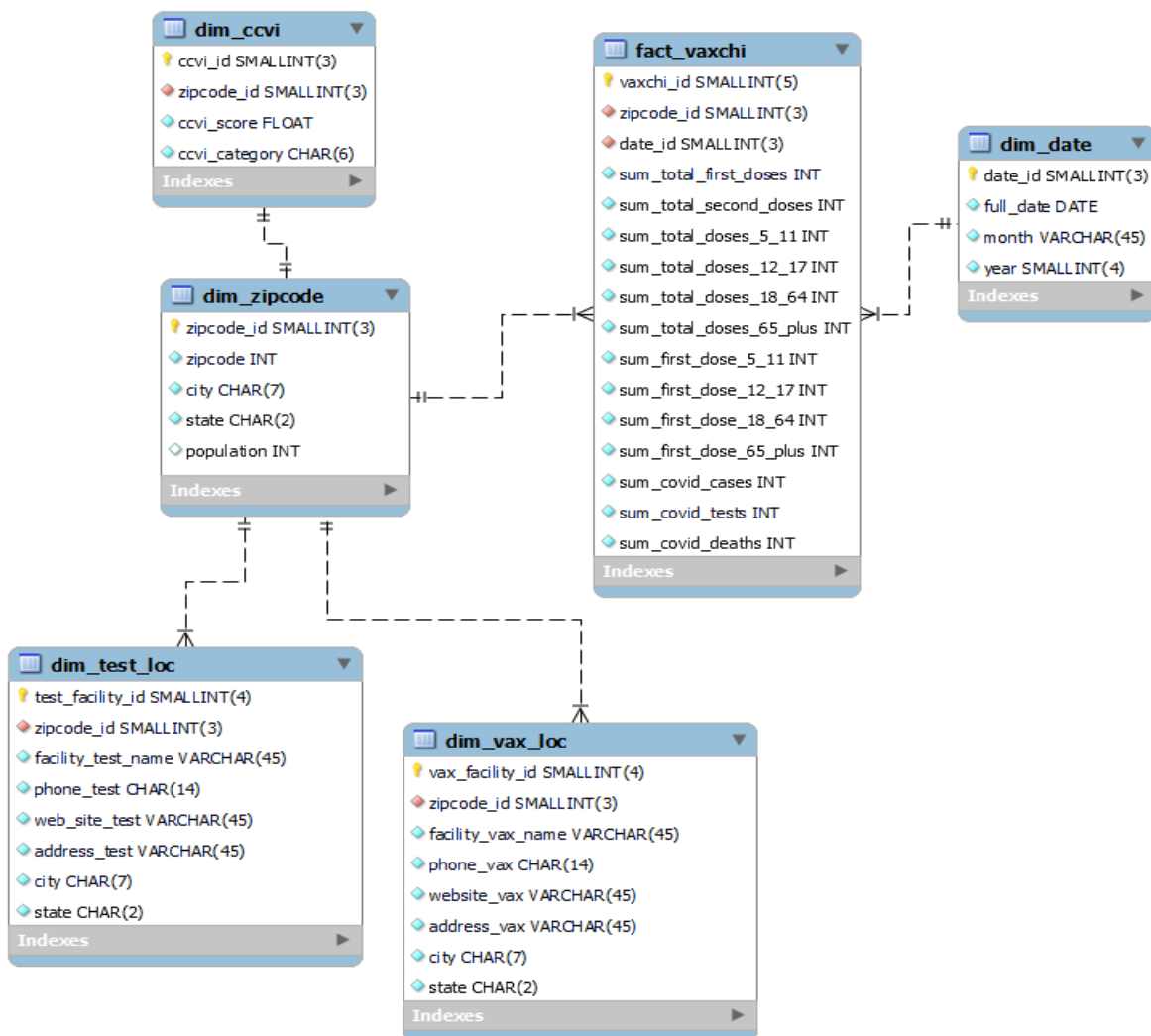
Preliminary Relational Model



As evident from the datasets, both **zip code** and **time dimensions** are critical for our model's operations. We however felt that this model will get too complicated for business users, with not a lot of freedom for playing around with the metrics.

In a bid to simplify our model, we focussed on essential tables and devised a **dimensional model**. Given the complexity of data linkages, a **snowflake model** is being utilized by our team.

Final Dimensional Model



Dimension Tables:

1. dim_date

The time dimension control table for our model rolled up at a month level. We are intentionally using month-level to provide granular insights into the vaccination trends and avoid big data dumps.

The month-level roll-up can be however changed to week level in case a need arises in the future.

2. dim_zipcode

The geography dimension control table for our model rolled up at a zip code level. We aggregated all the zip codes available in our relational database to a central repository.

We also include the following geographical information at a zipcode level:

- City
- State
- Population

3. dim_vax_loc

Geographical locations of the **vaccination centers** spread across Chicago.

The locations belong to a particular zip code, with each zip code allowed to have multiple vaccination centers.

4. dim_test_loc

Geographical locations of the **covid testing centers** spread across Chicago.

The locations belong to a particular zip code, with each zip code allowed to have multiple testing centers.

Testing centers by their definition are being kept separate from vaccination centers.

5. dim_ccvi

Covid Community Vulnerability Index is the measure of the vulnerability level consisting of various rankings like socioeconomic status, age, hospital admissions in etc.

It is a metric that outlines the severity of repercussions resulting from the COVID pandemic. The lower the index, the less severe the repercussions are in a particular community.

Fact Table:

From a business intelligence perspective, the fact table can be utilized to identify covid related statistics and test metric trends for a particular geography and time window. This will help us identify how different areas in Chicago have been performing in managing their covid caseload.

Three source tables are being used to populate the fact table:

- **Covid vaccinations by zip code and date**
- **Covid vaccinations by age and date**
- **Covid tests, cases, and deaths by zip code and week**

All tables were rolled up at the month level (with the help of dim_date)

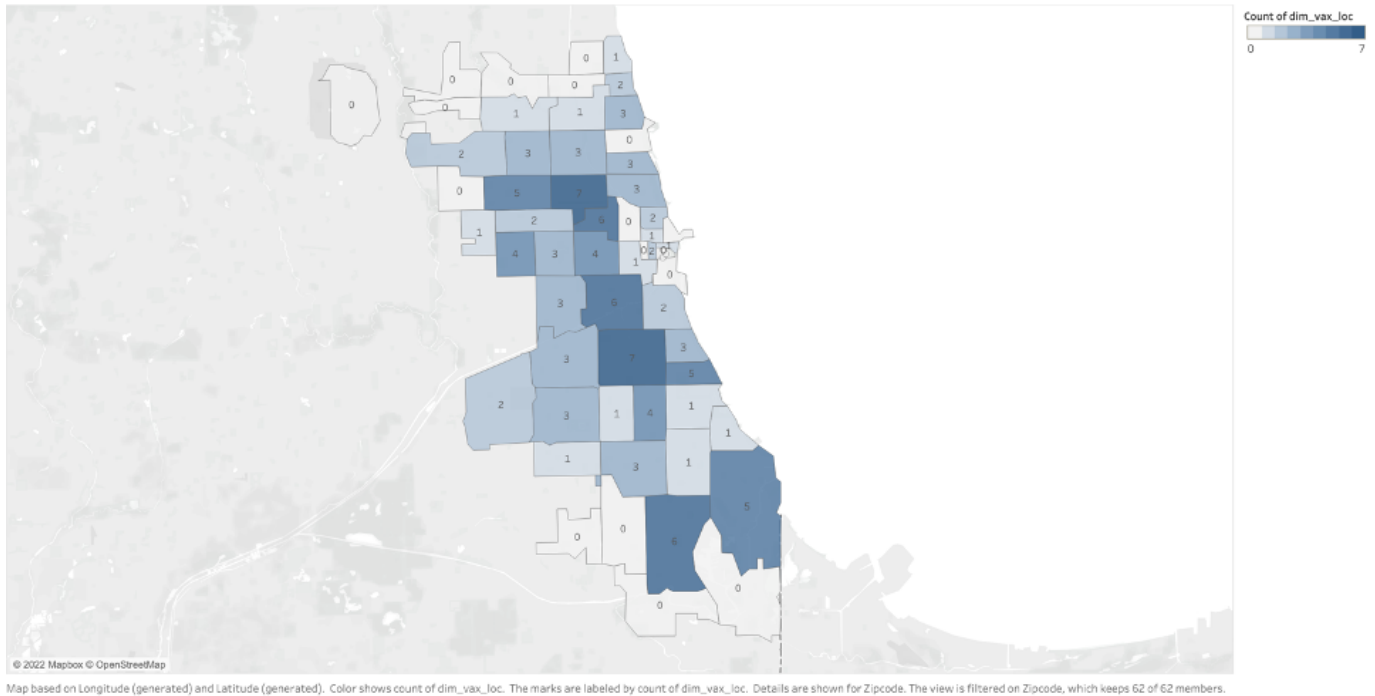
Sample Business questions that can be answered using our dimensional model:

1. Vaccination rates over specific periods of time
2. Vaccination rates based on zip code
3. Vaccination rates and status based on age group
4. Interaction between vaccination and testing numbers
5. Evaluating the effect of the availability of vaccination and testing sites on vaccination rates
6. Effect of CCVI score on vaccination rates
7. The interaction between CCVI scores and available vaccination and test sites

Part 2: Visualization and Insights

Density of Vaccination centers across Chicago

Density - Covid Vaccination Locations

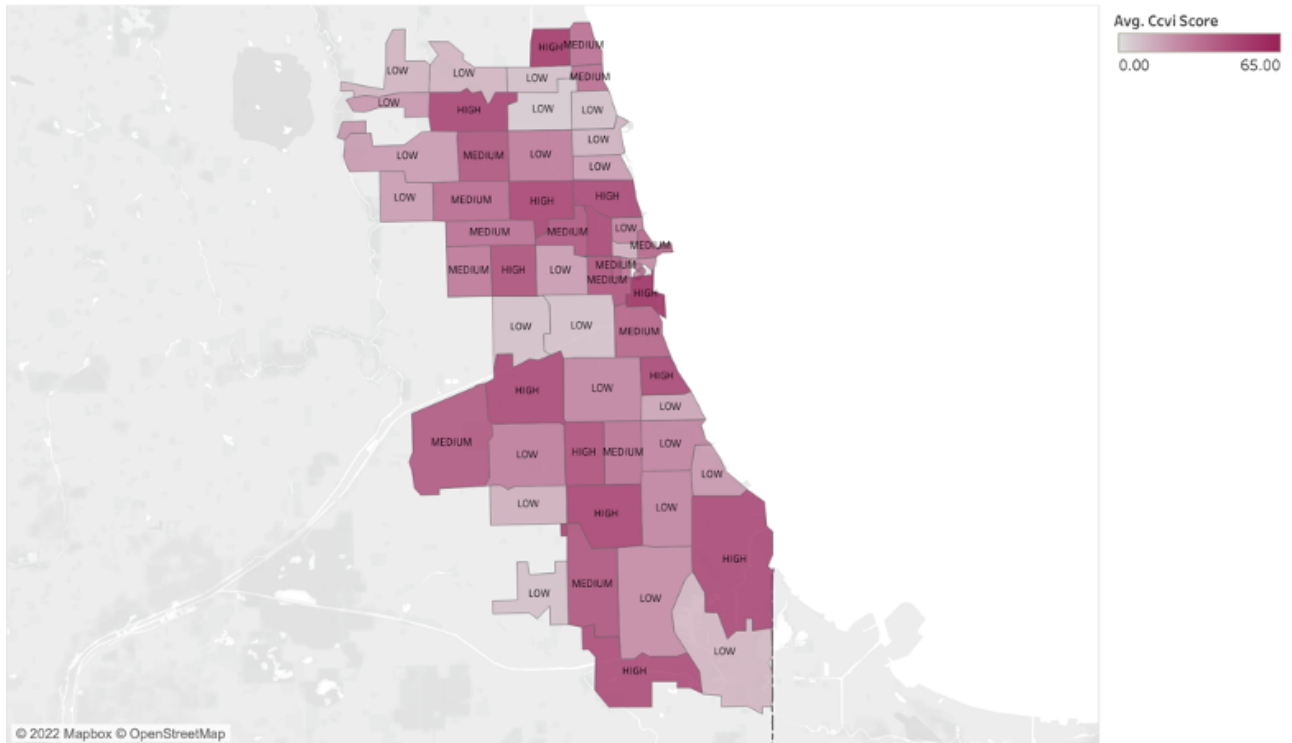


This visualization shows us the density of COVID-19 Vaccination sites by zip code. This image is a useful reference when evaluating other COVID-19 outcomes such as the vaccination status of different zip codes.

As we can see, there is a definite disparity between the infrastructure available at each zipcode. While some localities are very heavily endowed, as we move towards the fringes of the city the density drops sharply.

Community Vulnerability Indexes

CCVI Scores

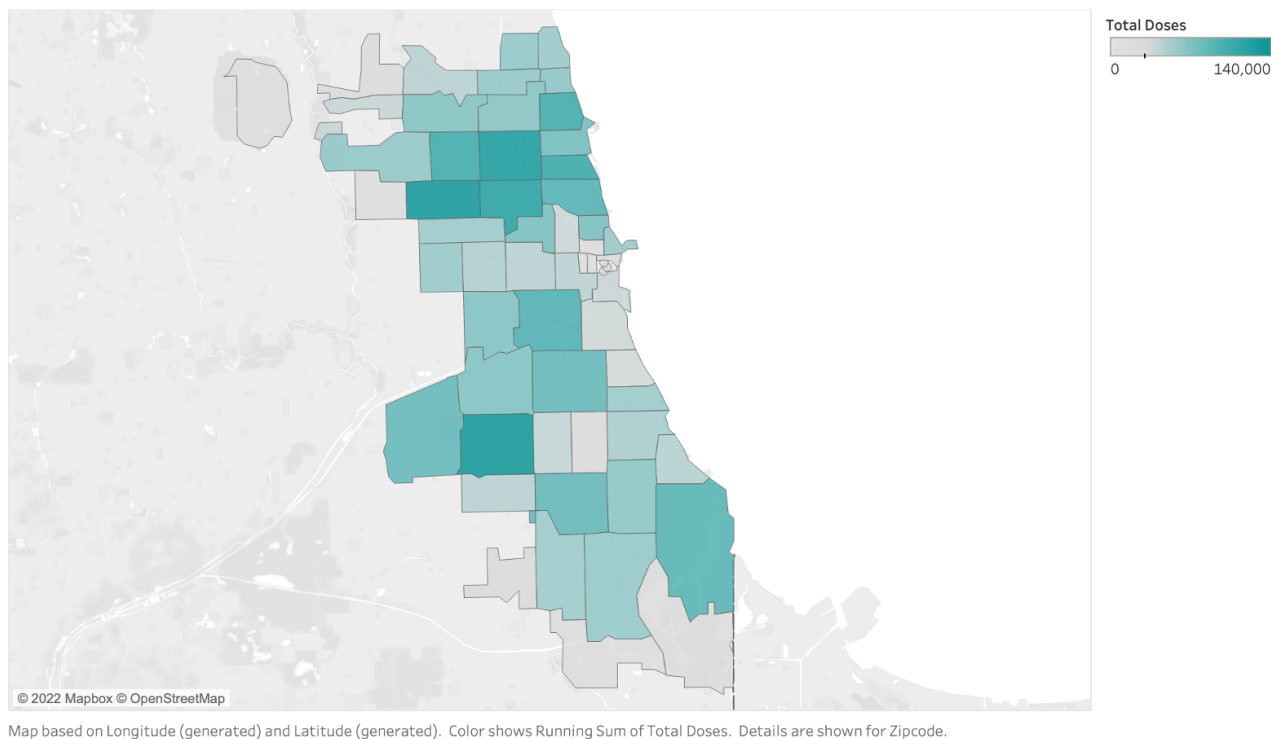


Map based on Longitude (generated) and Latitude (generated). Color shows average of Ccvi Score. The marks are labeled by Ccvi Category. Details are shown for Zipcode.

This visualization shows the distribution of the Chicago Community Vulnerability Index (CCVI) scores across Chicago. These scores are calculated based on factors determined by the city of Chicago to determine susceptibility to COVID-19 and difficulty in obtaining vaccinations. A higher score indicates a higher vulnerability index.

Covid Vaccination across zip codes

Vaccination
19, June-2022

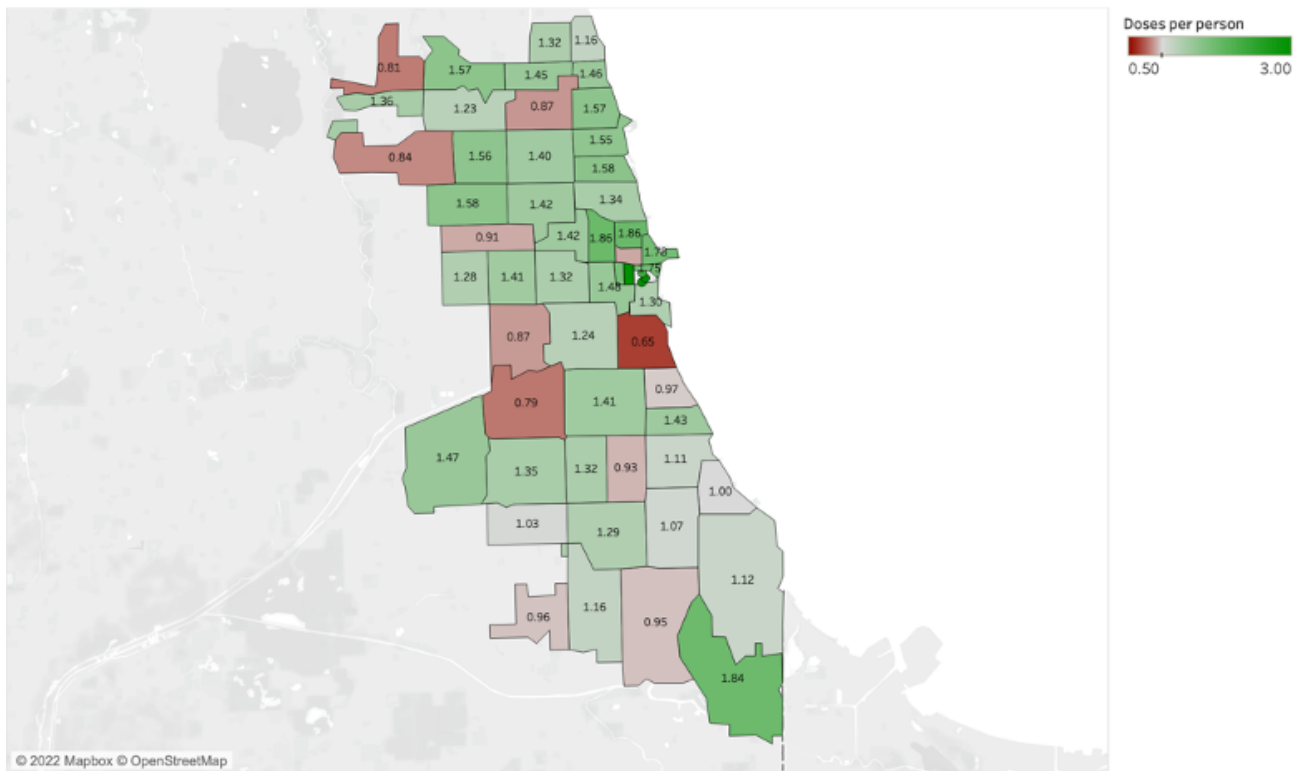


This visualization shows how the sum of those receiving vaccinations increased over time and in what zip codes. This is an image of the sum of vaccinations as of June 2022, the last date in our data set. The darker regions of this map slightly correspond to the darker regions of the previous map that shows density of vaccination sites.

Nearly all areas of Chicago are performing fairly well when it comes to vaccination counts, with North of Chicago showing more consistent performance. However, the south side still has a lot of scope for improvement as seen by several grey zones.

Covid Vaccination Density across zip codes

Vaccination Density

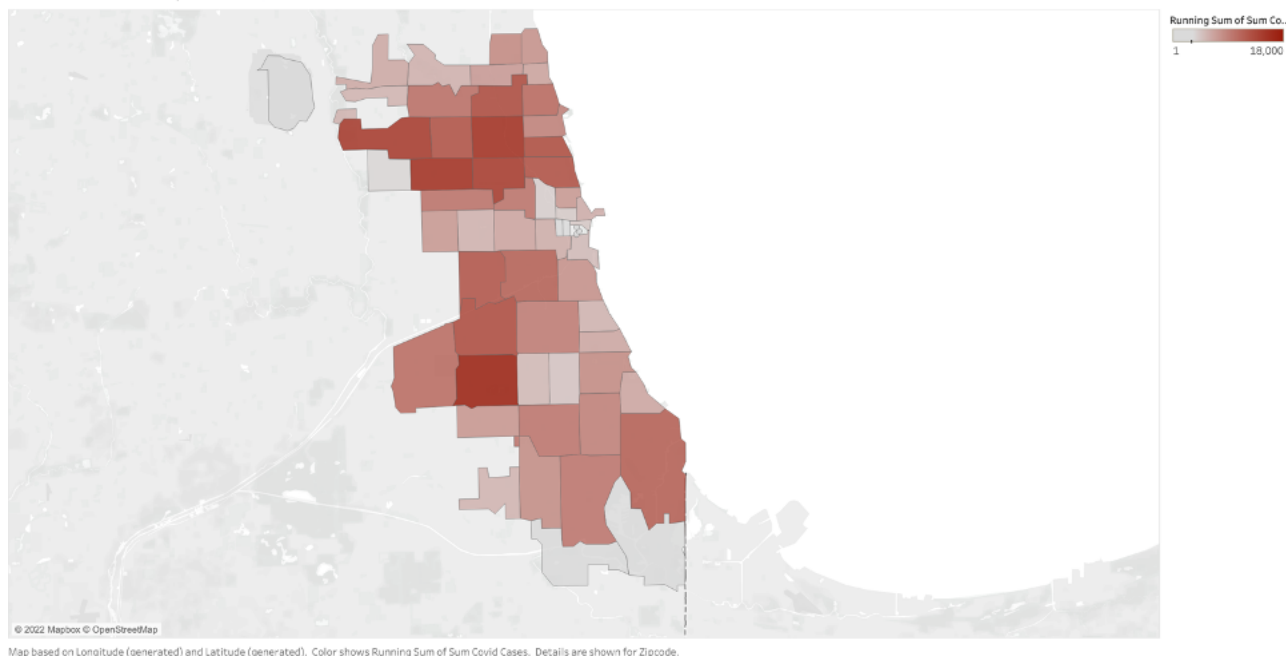


Map based on Longitude (generated) and Latitude (generated). Color shows sum of Doses per person. The marks are labeled by sum of Doses per person. Details are shown for Zipcode. The view is filtered on Zipcode, which excludes 60302, 60666, 60707 and 60827.

This visualization also shows the sum of doses per zip code, but standardized by population. We can see that even though Chicago has had a very robust vaccination program, there are still many areas within the city where the entire population has not received even a single dose.

Covid Cases across zip codes

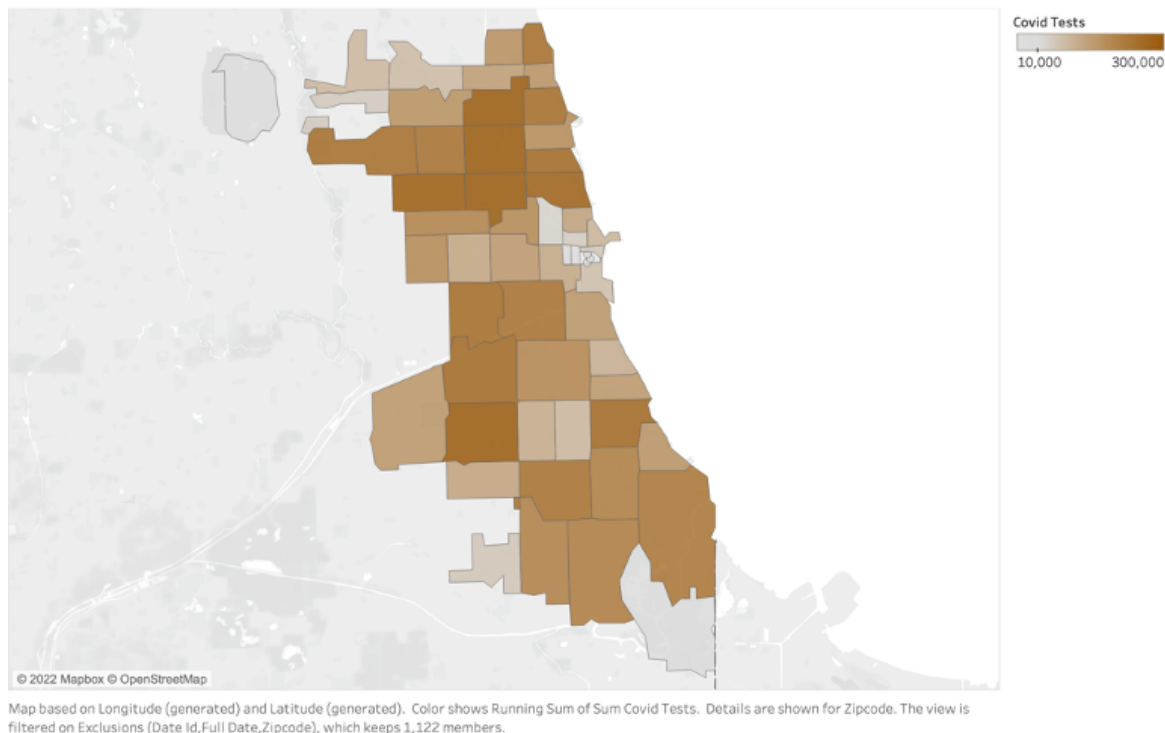
Covid cases across zipcodes **19, June-2022**



The following visualization demonstrates how the number COVID-19 cases in Chicago increased over time, represented by increasing red colors. It is interesting to evaluate the rate in which certain zip codes see their COVID-19 cases increase and compare it to the map of the CCVI and vaccinations by location.

Covid Testing across zip codes

Covid Testing -
18, May-2022

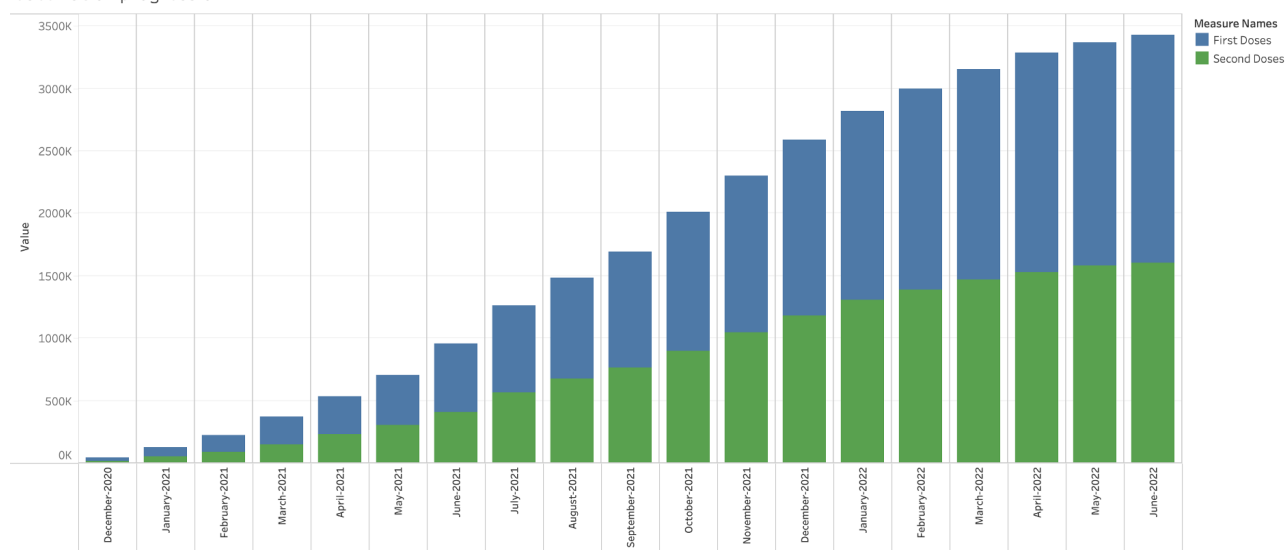


This visualization demonstrates how COVID-19 testing numbers have changed over time. Interestingly, the distribution of testing numbers on this visualization closely corresponds to the case number visualization.

Continuous and Comprehensive testing is a critical factor for keeping a tap on all public health procedures. Chicago as a whole has been doing a good job for sustained testing.

Vaccination Progression

Vaccination progression



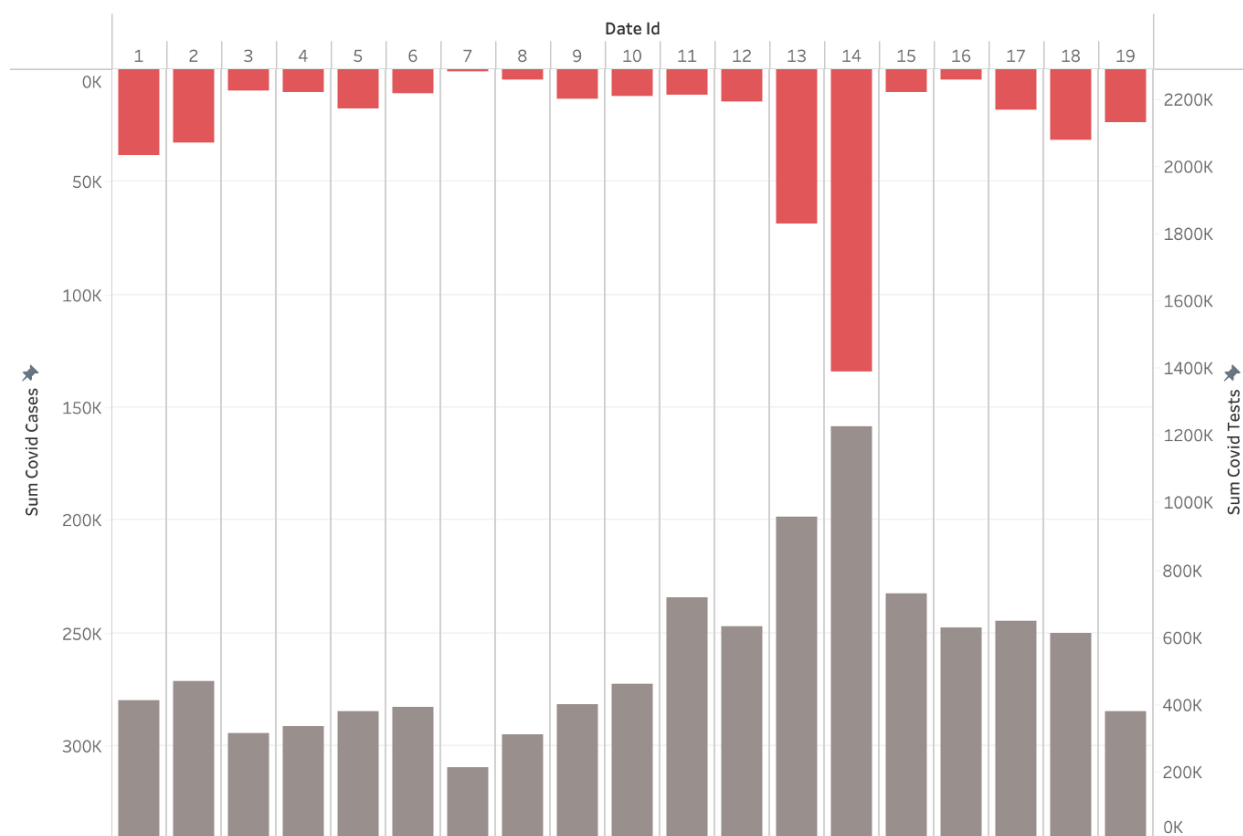
First Doses and Second Doses for each Full Date broken down by Date Id. Color shows details about First Doses and Second Doses.

This visualization demonstrates the sum of first and second doses obtained over time throughout Chicago. Most notably, it is interesting to see how the rate of increase in the sum of second doses was always considerably slower than the rate of first doses, although both the rate of first and second doses obtained started to level off over time.

The difference in rates of first and second doses could be an interesting point for future studies; what causes this difference between the rates? Is it accessibility? Availability of public health education?

Covid Tests vs Positive Cases

Covid Tests and Positive Cases

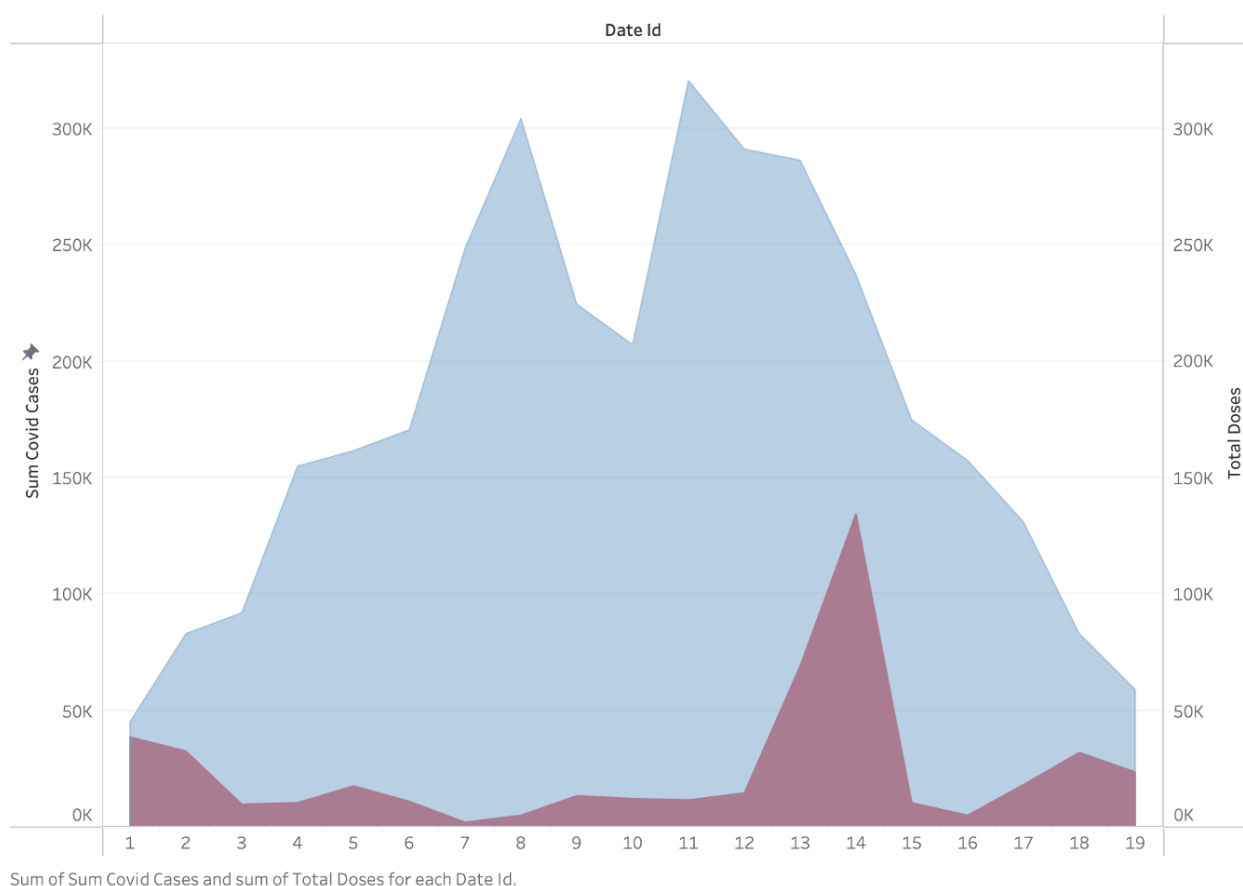


Sum of Sum Covid Cases and sum of Sum Covid Tests for each Full Date broken down by Date Id.

This visualization shows the relationship between COVID-19 cases and tests. We can see from this visualization that increased case numbers do tend to correspond to test numbers. **We could infer from this visualization that an increase in COVID-19 cases could increase awareness and concern in individuals enough to get tested more frequently**, but this would warrant further data analysis.

Covid Vaccinations vs Cases

Covid Vaccinations vs Cases

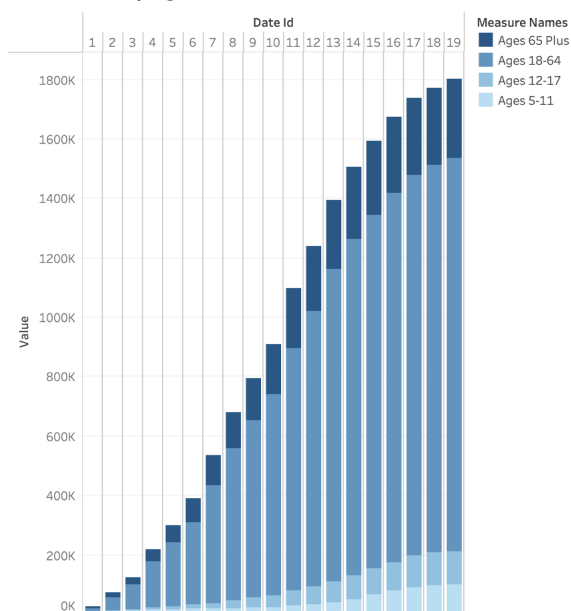


This visualization compares the total dose numbers to the sum of COVID-19 cases over time.

The blue corresponds to total vaccination doses and the red corresponds to the number of COVID-19 cases. With the exception of a spike from around November 2021 to January 2021, the higher vaccination numbers kept the COVID-19 case numbers relatively low. This red spike also corresponds to the rise of the Delta variant. Note that there is a dip in the vaccination rate shortly before the rise of the Delta variant. **Perhaps if vaccination rates did not decline around that time, the severity of the Delta variant would have been lower in Chicago.**

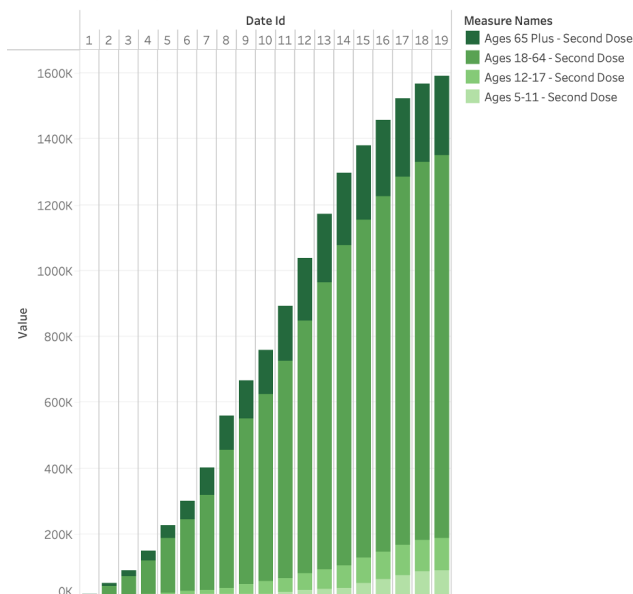
Covid Vaccinations by Age

First Dose - By Age



Ages 65 Plus, Ages 18-64, Ages 12-17 and Ages 5-11 for each Full Date broken down by Date Id. Color shows details about Ages 65 Plus, Ages 18-64, Ages 12-17 and Ages 5-11.

Second Dose - By Age

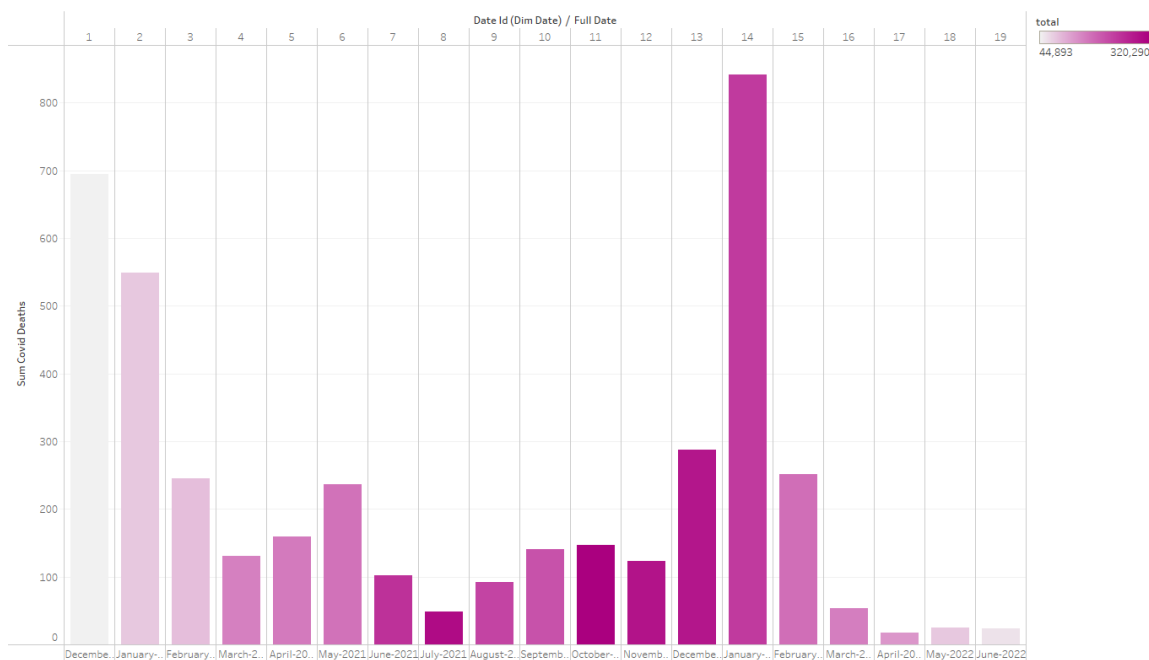


Ages 65 Plus - Second Dose, Ages 18-64 - Second Dose, Ages 12-17 - Second Dose and Ages 5-11 - Second Dose for each Full Date broken down by Date Id. Color shows details about Ages 65 Plus - Second Dose, Ages 18-64 - Second Dose, Ages 12-17 - Second Dose and Ages 5-11 - Second Dose.

These two visualizations show how vaccinations were obtained categorized by age. Notably, the proportions of each age group receiving each dose do not seem to change between the first and second dose, only the total quantity of doses across all age groups. **By June of 2022, almost 90% of individuals who received the first dose had also received the second dose.**

Covid Deaths vs Total Doses

DEATH VS TOTAL VACCINE DOSES



Sum of Sum Covid Deaths for each Full Date broken down by Date Id (Dim Date). Color shows sum of total.

The above visualization shows the number of COVID-19 deaths as well as total doses over time. The darker the bar, the more doses were obtained within a month. Most notably, after the Delta variant spike, the number of deaths dropped and remained considerably low. **There did not seem to be a very strong relationship between the number of vaccine doses and COVID-19 deaths.** This could have been due to the novelty of the Delta Variant and delayed ability of the medical community to adapt.

Part 3: Additional Functionalities

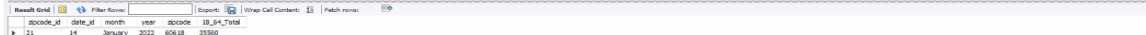
Apart from the Dashboard, our dimensional model itself can be utilized by government BI teams to conduct any ad-hoc analyses. The model is easy to understand, and basic understanding of SQL should empower the user to perform more data analytics

SQL Sample 1

```

1 # Which zip code and month had the highest number of total vaccinations(first and second doses) in the 18-64 age category?
2 • SELECT
3     fact_vaxchi.zipcode_id,
4     fact_vaxchi.date_id,
5     dim_date.month,
6     dim_date.year,
7     dim_zipcode.zipcode,
8     (fact_vaxchi.sum_first_dose_18_64 + fact_vaxchi.sum_second_dose_18_64) as '18_64_Total'
9 FROM fact_vaxchi
10     INNER JOIN dim_zipcode on dim_zipcode.zipcode_id=fact_vaxchi.zipcode_id
11     INNER JOIN dim_date on dim_date.date_id=fact_vaxchi.date_id
12 GROUP BY fact_vaxchi.zipcode_id, fact_vaxchi.date_id
13 HAVING '18_64_Total'
14 ORDER BY '18_64_Total' DESC
15 LIMIT 1;

```



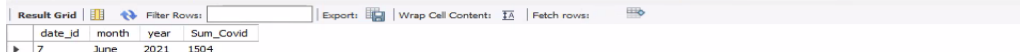
zipcode_id	date_id	month	year	zipcode	18_64_Total
21	14	January	2022	60618	25960

SQL Sample 2

```

17 # During what month the lowest number of cases was recorded?
18 • SELECT
19     fact_vaxchi.date_id,
20     dim_date.month,
21     dim_date.year,
22     SUM(fact_vaxchi.sum_covid_cases) as 'Sum_Covid'
23 FROM fact_vaxchi
24     INNER JOIN dim_date on dim_date.date_id=fact_vaxchi.date_id
25 GROUP BY fact_vaxchi.date_id
26 ORDER BY fact_vaxchi.sum_covid_cases ASC
27 LIMIT 1;

```



date_id	month	year	Sum_Covid
7	June	2021	1504

SQL Sample 3

```
29 # What zipcodes have 5 or more vaccination locations?
30 • SELECT
31     dim_vax_loc.zipcode_id,
32     dim_zipcode.zipcode,
33     count(dim_vax_loc.facility_vax_name) as 'Number of Facilities'
34 FROM dim_vax_loc
35     INNER JOIN dim_zipcode ON dim_vax_loc.zipcode_id=dim_zipcode.zipcode_id
36 GROUP BY dim_vax_loc.zipcode_id
37 HAVING count(dim_vax_loc.facility_vax_name)>=5
38 ORDER BY count(dim_vax_loc.facility_vax_name) DESC;
```

Result Grid | Filter Rows: | Export | Wrap Cell Contents: |

	zipcode_id	zipcode	Number of Facilities
▶	12	60609	7
	48	60647	7
	11	60608	6
	25	60622	6
	30	60628	6
	18	60615	5
	20	60617	5
	40	60639	5

Part 4: Reflections and Conclusions

Lessons Learned:

What went well:

- Our group had strong communication, and group members were flexible and willing to take on whatever amount of time/work necessary to meet expectations for the project
- Our emphasis on zip codes for analysis allowed us to make interesting visualizations that are accessible and understandable by most audiences

What could have been done better:

- More time could have been spent on gathering datasets; there may have been datasets that worked better with the rest of our model, required less cleaning, or simply contained more information.
- Many iterations onto our model and data slowed down progress throughout the quarter; in the future, more planning before jumping into the project could help us optimize our time

Recommendations:

- During pandemics, developing infrastructure that prepares for the possibility of a variant could help reduce negative health outcomes
- Prioritize increasing availability of vaccination centers in highly vulnerable communities
- Promote more health education initiatives to encourage remaining individuals to receive vaccination

References

Chicago COVID-19 Vaccine Coverage:

<https://www.chicago.gov/city/en/sites/covid19-vaccine/home/covid-19-vaccine-coverage.html#about+the+data>

Vax Chi Nation Information:

<https://www.chicago.gov/city/en/sites/covid19-vaccine/home/vax-chi-nation.html>