TOPIC SELECTION:

Our group's initial question of interest was one that had been frequently discussed and publicized during the early stages of the COVID-19 pandemic — with public life becoming a serious health risk overnight, how have the city's performers, entertainers, and venues fared? Sourcing information about the spread of COVID-19 and tracking key metrics was a simple enough task thanks to well-maintained and publicly available data from Chicago's department of Health and Human Services, as well as the Illinois Department of Public Health.

However, attempting to locate information about pre-COVID trends in ticket sales, event frequency, or even venue opening status were sparse at best. Venues themselves offered no customer-facing method of accessing their attendance data & our attempts to bypass this issue by querying ticket vendors such as StubHub, TicketMaster, or Live Nation met with similarly poor results in the availability of useful data. Moreover, the single third-party organization that might have allowed us to bypass these limitations required both a monthly fee for access and a flat-rate charge per query, making this financially untenable when our data needs were still ill-defined.

With this in mind, our attention shifted to ideas that would make greater use of the data available through the City of Chicago portal. We elected to look at a similar question with relation to the flow of people through high-traffic public venues that we believed would have been heavily affected by COVID. We settled on examining how the pandemic had impacted activity in the city's library system, which added the question of how COVID impacted not only the flow of people through the physical locations, but also how it's business services (i.e. lending materials) were affected as well.

DATA CLEANING:

Determining the elements and structure of our data was made easier thanks to documentation from the City of Chicago portal for both the library data as well as COVID metrics without any identifiable gaps. Our group decided to focus on two key metrics for library activity: visitation to physical locations and circulation of materials. These were selected by the group as a result of our expectations on what would be the most material and visible change in the data between 2019 and 2020-2021. We took a similar approach to refining the COVID dataset and elected to center reported cases and recorded deaths, as these metrics are most frequently centered in policy decisions for the city of Chicago during the pandemic.

Merging the two data sets by area ZIP code seemed the most intuitive approach owing to the format of the COVID data reported by ZIP code on a weekly rolling basis & the distribution of libraries demonstrated that each ZIP code was served by a single library.

With data in hand, we honed our initial investigation for the most salient questions to guide our analysis:

- Did the COVID-19 pandemic have an impact on the Chicago library system?

- How were pre-pandemic monthly trends in library visitation and circulation altered by COVID-19?
- Was there a correlation between library activity and COVID deaths observed by ZIP code?
- Do the observed trends in our data display intuitively on a visual representation?

ANALYSIS:

Summary statistics of the visitation and circulation for each library location were produced for each year analyzed, starting with 2019 through 2021. Sharp declines for mean and median circulation in 2020 were observed at 43% and 52% respectively (Fig. 1), while visitation saw a similarly sharp decline with a 59% drop in for both mean and median visitors per ZIP code for the year. When comparing 2019 to 2021 data, decline in mean and median circulation improves somewhat to 33% and 41% reductions respectively, however visitation to libraries sees only marginal improvements to 56% decline from 2019 for both the mean and median.

TYPE	Mean	Median	Variance	Standard Deviation	SEM
Circulation	59075.41	39280.0	6.292804e+09	79327.20	8814.13
Visitors	100494.75	74724.5	2.491201e+10	157835.39	17646.53

TYPE	Mean	Median	Variance	Standard Deviation	SEM
Cases	3554.47	2733.0	8.556272e+06	2925.11	380.82
Circulation	33656.25	19047.0	1.895243e+09	43534.39	4837.15
Deaths	74.64	70.0	3.274720e+03	57.23	7.45
Tests	40332.31	41243.0	5.806489e+08	24096.66	3137.12
Visitors	41459.75	30545.0	4.020040e+09	63403.79	7088.76

Figure 1. — Comparison of 2019 library activity to 2020 library activity & COVID-19 metrics

These macro level trends align with our group's expectations that citywide lockdowns imposed during 2020 and public sentiment about the risks associated with public venues following the lockdown would have the greatest impact on all library metrics, while the 2021 data points to a possible change in the borrowing habits of visitors that did return. Histograms generated from each respective yearly data set (see Appendix A) imply that for each library location, the general trend of high activity vs. mid-to-low activity locations, that is locations that see considerably higher numbers in either metric, hold roughly stable across all three years. Notable outliers conducting a substantial portion of the system wide activity remain the same in quantity in all years analyzed.

A more granular analysis of monthly fluctuations in library metrics compared to reported COVID case averages further reinforces that while COVID-19 significantly impacted volume, the temporal trends in library metrics remained remarkably stable. Examining Fig. 2, the most prominent features of the line for 2020 visitation correspond to the most significant fluctuations in rolling case averages month to month, as well as the fixed point events of the city lockdown being implemented in March 2020 and lifted in early May 2020. As the pandemic enters 2021, one can observe that visitation adheres more closely to the seasonal changes first seen in 2019 data — relative maxima in June/July followed by moderate declines in August/September, as well as brief surges in October before declining once more into the winter months.

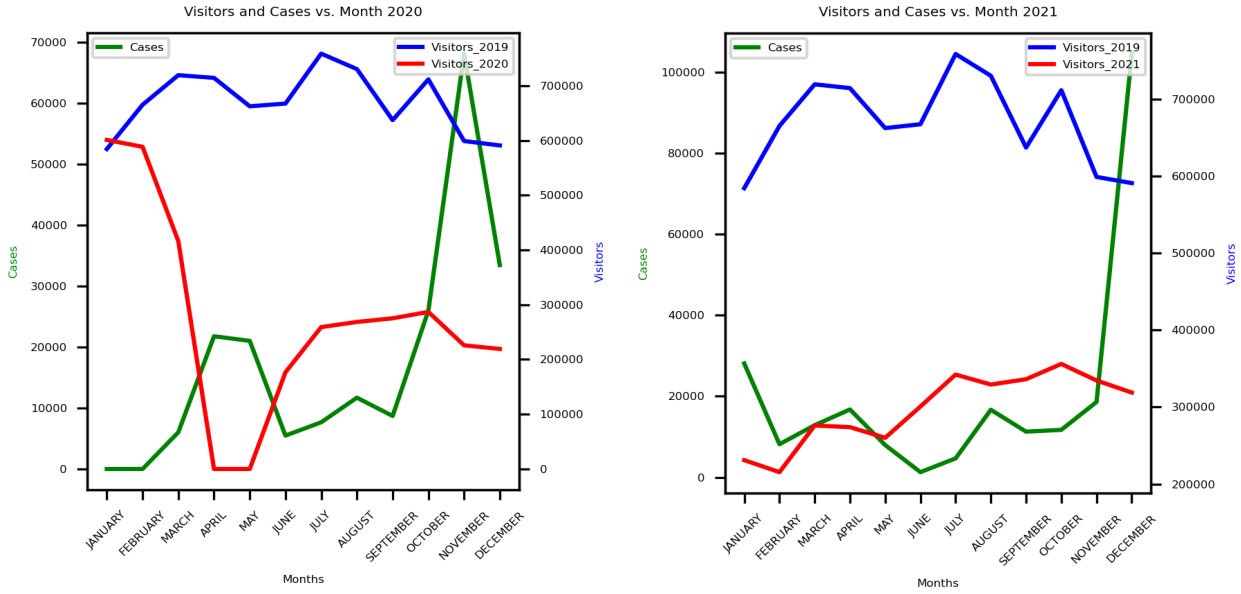

Figure 2. — Visitation vs. Rolling Case Total (2020, left; 2021, right)

Circulation data for the same time interval displays similar adherence to fluctuations expected during the course of 2019. Of note, the visitation and circulation volume for 2020 and 2021 displays a flattened curve compared to 2019 data, indicating that the added influence of COVID-19 is likely responsible for externing a reduced upper bound on the metrics measured here.

With these observations in mind, our group returned to the question of whether the COVID deaths recorded for a given ZIP code had any correlation to the respective library metrics. The analysis of deaths rather than cases & visitors rather than circulation was decided on the basis that each recorded death attributed to COVID would presumably have a larger impact on the behavior of residents within that ZIP code, and visitation to the physical location represents a greater risk of contracting COVID-19 as a consequence. Individual Pearson correlations were calculated for the recorded deaths and library visitors for each ZIP code across all months for which COVID cases had been recorded in Chicago, after which the 4 strongest correlation values were identified and plotted for each month data was available & separate linear regressions were performed to determine the nature of the correlation (pictured below).

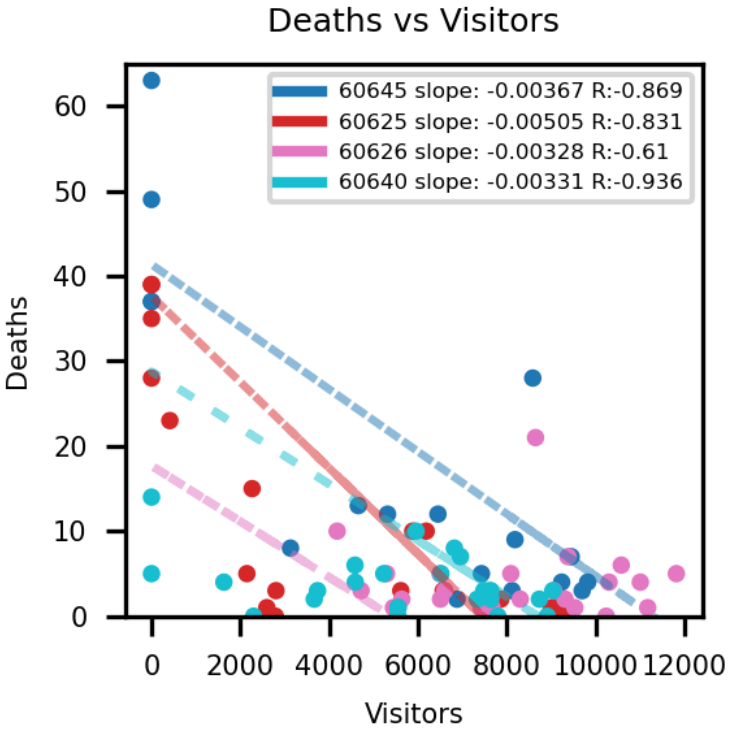

Figure 3. — Regression plots for ZIP codes of Interest

Qualitative analysis of the ZIP codes isolated by this method demonstrated that the ZIP codes of interest were in fact clustered adjacent to one another & located in the Rogers Park area on the northeastern boundary of the city. The correlation coefficients for each of the relationships was greater than R=0.5 in magnitude and all demonstrated an inverse correlation, with the 3 of the 4 relationships measured greater than R=0.8 in magnitude. These values indicate that library visitation declined dramatically in months where deaths from COVID-19 increased in the corresponding ZIP code.

Inferences that could be made from this data are far ranging, such as how availability of data on recorded deaths from COVID-19 in these ZIP codes influenced resident behavior & subsequent visitation, but are outside of the scope of the dataset utilized for our analysis.

Finally, our group aimed to answer a question that felt widely applicable to Chicagoans who may not possess the same degree of data literacy as our group: can these trends in library activity and the spread of COVID-19 be easily visualized? To this end, our group decided to generate a heat map of the city that coupled the density of library visitation & COVID-19 cases for each ZIP code analyzed in the data set (see Figure 4 below). The aim of this map was to readily identify how these quantities may be related to a wider audience, and to reinforce the conclusions drawn from our quantitative work above. The similarity in magnitude for both density plots for the majority of ZIP codes shows on first examination that tracts with high COVID case volume seem to have similarly high library visitation. Our human reflex is to take this data at face value and suspect a correlation between the two exists, however this method also inclines us to find examples that refute that trend — while the tract containing the city's largest and most trafficked library, Harold Washington, was the greatest density of visitors, the COVID cases reported falls beneath the mean. Visualization of our data in this way provides benefits in creating easily interpreted figures, but also points to potential areas that contradict our initial assumptions and can help to steer deeper analyses.

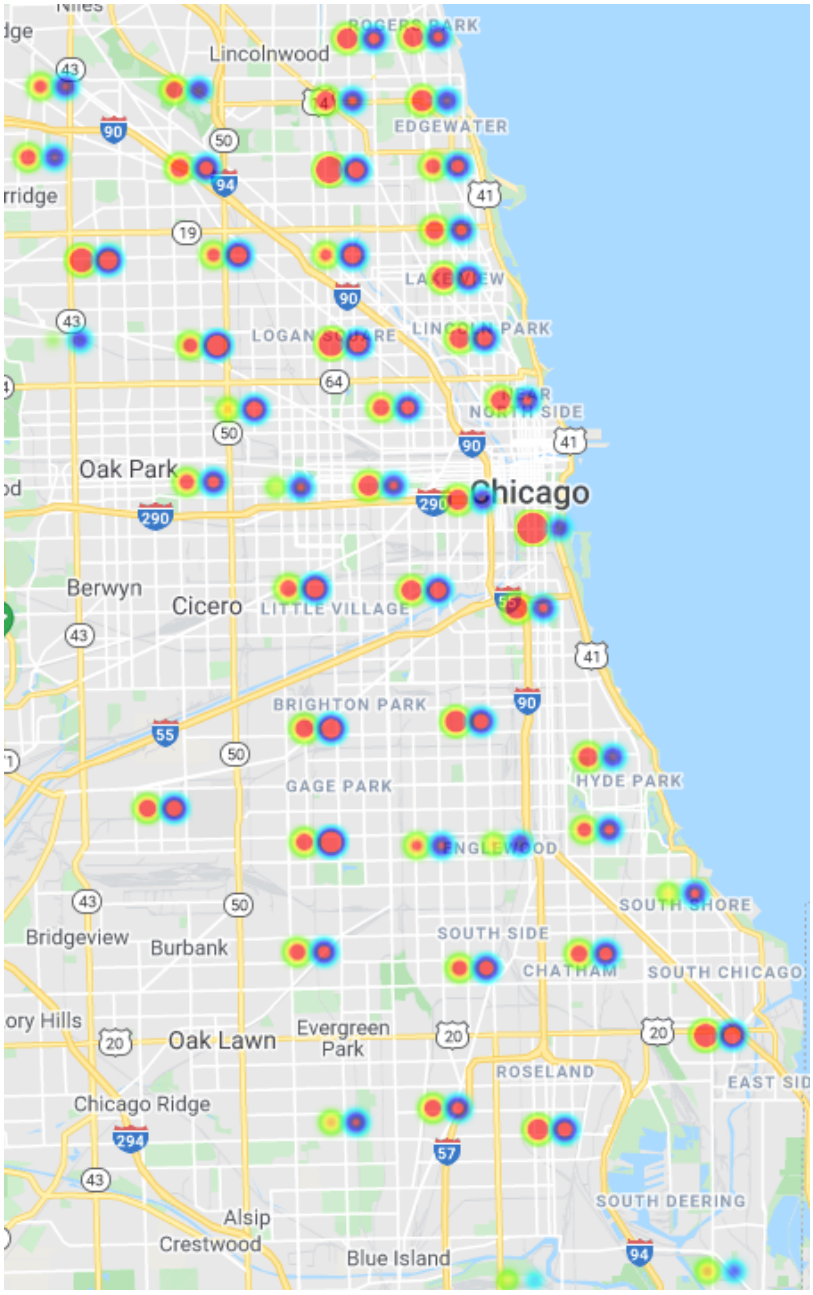

Figure 4. — 2020-2021 Library Visitors vs COVID Cases Reported

Heatmap Legend

Library Visitors

COVID Cases

CONCLUSIONS:

Revisiting the questions posed at the outset of this project, our analysis has revealed the following:

- Did the COVID-19 pandemic have an impact on the Chicago library system?
    - In both visitation and circulation, there was a pronounced and sustained impact on the Chicago library system due to the COVID-19 pandemic. As of this analysis, there is no indication that any member of the library system has attained their pre-pandemic volume for either metric.
- How were pre-pandemic monthly trends in library visitation and circulation altered by COVID-19?
    - Despite the notable reduction in general volume, trends in circulation and visitation expected between months based on 2019 data appear to have been largely preserved in 2021 and to a lesser Q3/Q4 of 2020. Although a the second wave in COVID-19 cases occurring in Q3 of 2020 precipitated a further reduction in visitation and circulation, the surge associated with the emergence of the omicron variant of COVID-19 in Q3 of 2021 did not demonstrate a corresponding deviation from month-to-month trends expected in 2019.
- Was there a correlation between library visitation and COVID deaths observed by ZIP code?
    - Several ZIP code tracts displayed a significant inverse correlation between library visitation and recorded COVID-19 deaths between 2020 and 2021. There are intriguing geographic & socioeconomic similarities shared among the ZIP code tracts displaying the strongest negative correlations, making this cluster a potential basis for further investigation of the relationships identified in our analysis.
- Do the observed trends in our data illustrate well on a visual representation?
    - The incidence of COVID cases and library visitation on the heat map of Chicago is useful in communicating simple but important information about how these two variables are connected. Even with our limited proficiency using the mapping tools, we feel confident that creating this figure would expand the accessibility and improve the effectiveness of our report in future projects.

APPENDIX A:

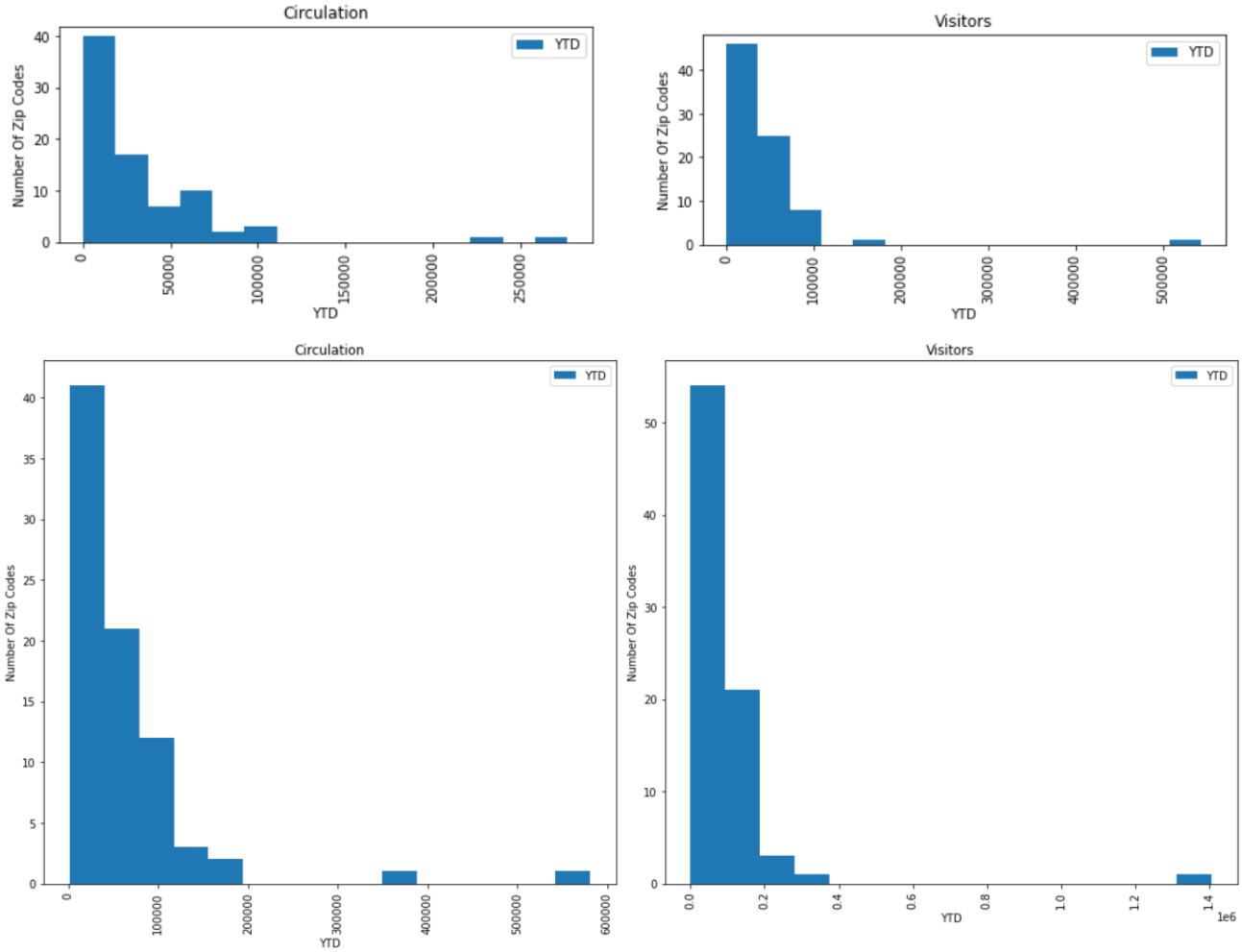

Figure 2 (Clockwise from Top Left): Circulation volume by ZIP code (2021), Visitation volume by ZIP code (2021); Visitation volume by ZIP code (2019), Circulation volume by ZIP code (2019)