Matthew Bailey
Gage Anderson
Han Le

Project 3 Chicago Crime

- Introduction/Background
- Inspiration/ Reason why we chose the topic
- Objectives/Research Questions
- How we got the data
- Data cleaning
- Analysis (Visualizations)
    - How did we do the visualizations?
    - Observations
    - Conclusions (answers to the research questions)
- Maps
    - What is shown in the map?
    - How did we do the map?
- Limitations/Future Work
    - Small sample
    - Lack of GeoJson data

Introduction/Background:

Chicago is known as the largest city in the state of Illinois, the third most populated city in the whole of the U.S, and considered an actual melting pot of culturally diverse populations. Going along with that, according to the Neighborhood Sprout, Chicago has one of the highest crime rates in America compared to all communities of all sizes, with a crime rate of 33 per one thousand residents. Within Illinois, more than 95% of the communities have a lower crime rate than in Chicago. This intrigued our curiosity, and we decided to look into it further.

Objectives/Research Questions:
We set out to answer three research questions:

- Which crime occurred in 2020, 2021 and 2021 and its frequency
- Which part of the city has more crimes?
- Which month/year do crimes occur more often?

Data Cleaning:
Our group used a csv file from a dataset on the City of Chicago website found on Kaggle. The csv file contained the year, district number, latitude, longitude, primary_type(type of crime), and crime count(No. of crimes that took place at latitude and longitude point). The dataset contained over 5,400,000 crimes spanning the years of 2001-2022. Because of the large amount

of crimes in the dataset, we cleaned up the data using Jupyter Notebook and Pandas. We took an extraction of 1000 rows from the years 2020, 2021, and 2022 as displayed below:

```
3]: df3=df2.reset_index()
    df3
```

3]:

	index	year	district	latitude	longitude	primary_type	crime_count
0	4606263	2021	12.0	41.857654	-87.676759	BATTERY	1
1	4636671	2021	20.0	41.983946	-87.691896	THEFT	1
2	4712302	2021	4.0	41.749649	-87.548007	DECEPTIVE PRACTICE	1
3	4493166	2020	7.0	41.767667	-87.662681	CRIMINAL DAMAGE	1
4	4622550	2021	7.0	41.767923	-87.631069	CRIMINAL DAMAGE	1
...	...	...	...	...	...	...	...
995	4619890	2021	4.0	41.702756	-87.564182	ASSAULT	2
996	4793684	2022	8.0	41.750060	-87.702402	DECEPTIVE PRACTICE	1
997	4562695	2020	8.0	41.799678	-87.728248	BATTERY	1
998	4812916	2022	15.0	41.890149	-87.750533	NARCOTICS	1
999	4704758	2021	5.0	41.709590	-87.623570	CRIMINAL DAMAGE	1

1000 rows × 7 columns

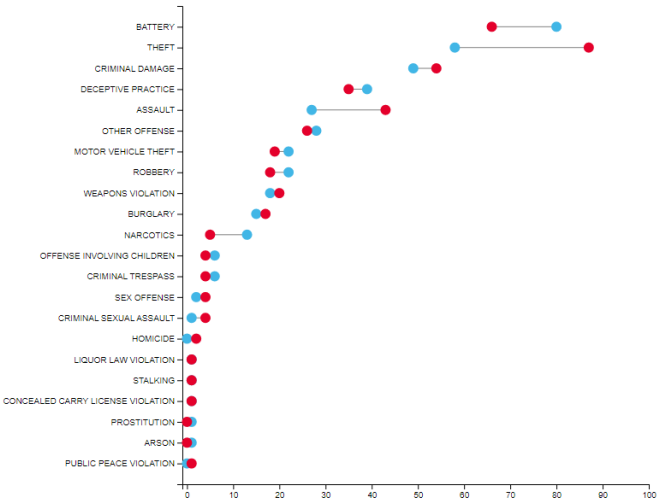
```
4]: df3.drop("index", inplace=True,axis=1)
    df3
```

4]:

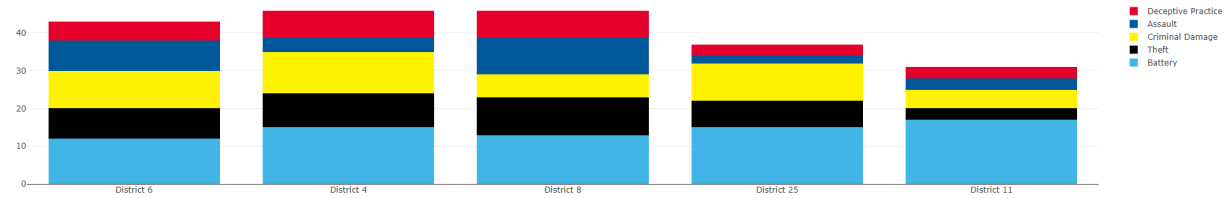
	year	district	latitude	longitude	primary_type	crime_count
0	2021	12.0	41.857654	-87.676759	BATTERY	1
1	2021	20.0	41.983946	-87.691896	THEFT	1
2	2021	4.0	41.749649	-87.548007	DECEPTIVE PRACTICE	1
3	2020	7.0	41.767667	-87.662681	CRIMINAL DAMAGE	1
4	2021	7.0	41.767923	-87.631069	CRIMINAL DAMAGE	1
...	...	...	...	...	...	...
995	2021	4.0	41.702756	-87.564182	ASSAULT	2
996	2022	8.0	41.750060	-87.702402	DECEPTIVE PRACTICE	1

This way the file size was smaller so that we could push it through GitHub but gave us enough data to do the analysis and answer the research questions of the project.

Analysis (Visualizations):



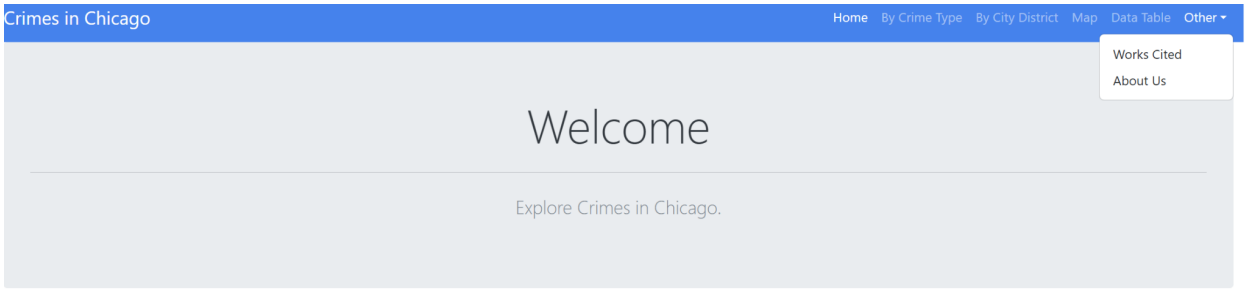
For our project, we chose to do a Lollipop Graph using d3 and a Stacked Bar Graph using Plotly. Our lollipop graph shows the change in crime frequency from 2020 (blue) and 2021 (red), with 2022 excluded because it was only six months of data. Both theft and assault drastically increased in frequency, with theft nearly doubling, while battery and narcotics related crimes decreased in frequency. The next chart we used was an interactive stacked bar chart that featured the five districts with the most crime as well as the five most commonly occurring crimes. One of the takeaways from this chart is that District 11 has more battery crimes than the other four districts, but less of the five chosen crimes overall. Additionally, District 4 had nearly the same amount of crimes as District 8 while having significantly less assaults.



Design:

For our website we chose to use the official colors of Chicago and added in the official emblem on our homepage. The colors we used were blue. The color/style used codes from the Bootswatch website.

The homepage contained a navigation bar, sample size of the map, and sample of the data visualizations we made for crimes by location and crimes by type. At the very bottom of the page are the names in our group as well as some text saying that the data is a subset of a much more comprehensive data set. The navigation bar contained buttons and dropdowns for the visualizations, Map, Data Table, and an Other text dropdown for Works Cited and an About Us page. The About Us contained brief descriptions of the group such as where we were from etc.. The navigation bar is the top right corner and by clicking on any of the items you are able to navigate to the page of your choosing.



Map:

For the map we read in the data using D3 json queryUrl that used a function for filtering the data that was then used for plotting the map. Since the dataset contained the longitude and latitude coordinates we were able to plot the crimes on a leaflet heatmap with markers that contained popups of the crime type, district number, and year the crime occured. We chose a heatmap because there was a limitation of the dataset where we were unable to create a district layer and add it to the map. By using a heatmap however it would be easy to visibly see where majority of crimes took place in chicago.

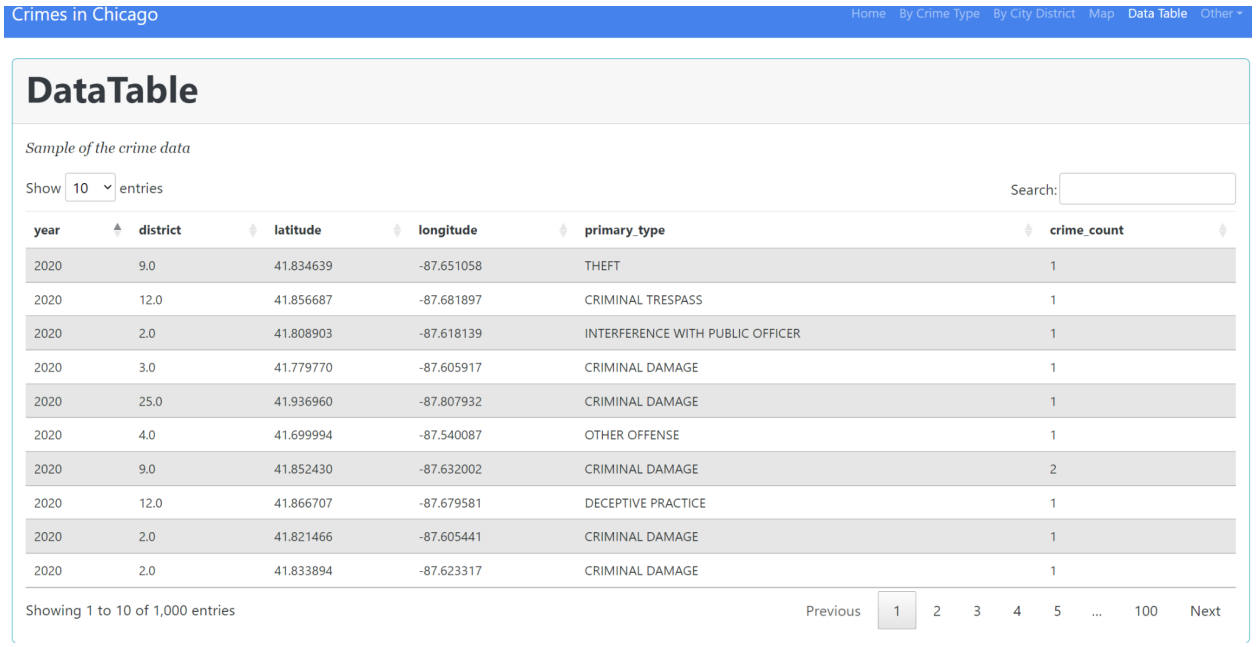
We used filters where we could set limits to the crime count so if we wanted to have markers for just 100 or 200 crimes on the map we could filter it so. We also have filters where you can choose crime type and populate markers of that specific crime. For example if you chose theft you could then populate markers on the map that were just theft crimes. Our final filters were for district number and year. For these you could choose the year of the crimes and all crimes by district number .

Website

As for the datatable, we converted the data from pandas to html then pasted the code on the website structure.

```
In [76]:  1 print(df3.to_html(index=False, classes=["table table-striped table-hover"]))
<thead>
  <tr style="text-align: right;">
    <th>year</th>
    <th>district</th>
    <th>latitude</th>
    <th>longitude</th>
    <th>primary_type</th>
    <th>crime_count</th>
  </tr>
</thead>
<tbody>
  <tr>
    <td>2020</td>
    <td>9.0</td>
    <td>41.834639</td>
    <td>-87.651058</td>
    <td>THEFT</td>
    <td>1</td>
  </tr>
  <tr>
```

We also updated the css style to make the table look nicer and more interactive. The table is displayed with 10 rows on one page and search the information in the search box.



Works Cited

The dataset was found on the Kaggle website (City of Chicago). And more information was referenced from the Neighborhood Scout website.

Limitations and Future Works

The current sample only contains 1,000 row items while the whole dataset includes more than 5.4M row. The sample size is small compared with the population size, which may make it difficult to determine if the above observation is a true finding. We expect the sample should be larger in order to present the population and to draw a more accurate solution