CMSC 12200 - Winter 2018

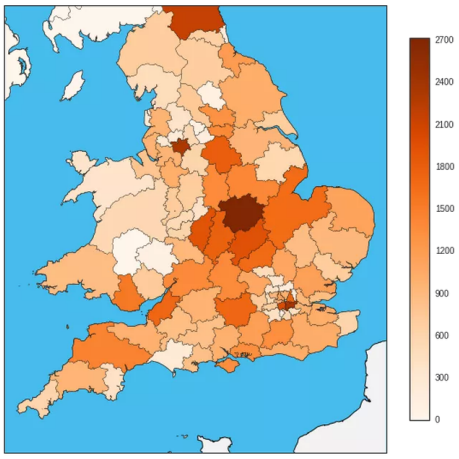
Group Name: Safety in Chicago

Members:

- Vu Phan
- Evan Mata
- Nanut Chaichanawanich
- Quinn Nguyen

Project Description

Our primary goal is to create a web interface where users can enter their zip-code (Chicago only) and receive a safety score calculated from the type and frequency of crime in the area from the past 5 years. They will also be able to see the change in safety score over time as a score and as a graph.



From this data, we will also be generating a geo-heatmap for the whole of Chicago on Matplotlib to visualize the safety score in each area. The heatmap will ideally look something like this:

Our secondary goal (if time permits), is to combine the crime data set with other data set such as education and healthcare accessibility, to provide additional score such as education score or healthcare score.

Data Sources

Our primary source of data is going to be a 1.4 GB file (https://catalog.data.gov/dataset/crimes-2001-to-present-398a4) on crimes in Chicago from 2001 till the present. It includes information on the time, location, and type of crime and whether an arrest was made, in addition to other factors that we are not interested in. The data comes in various different file types, including csv, which is what we are likely using.

As we are trying to create a zip-code specific statistic, we can use Google Map's API (https://www.zipcodeapi.com/API, https://developers.google.com/maps/documentation/geocoding/intro)/ GIS to convert our location data (latitude/longitude) into zip-code information.

Should our project need further sophistication there are numerous other data sets that we can combine to either further implement the safety score we are generating, or make a more general statistic/multiple statistics including other information such as an education or health score.

Here is a sample of the first few lines of our main data source data:

ID	Case Nu	Date	Block	IUCR	Primary	Descript	Locatior	Arrest	Domest	Beat	District	Ward	Commu	FBI Cod	X Coord	Y Coord	Year	Updatec	Latitude	Longituc	Locatior
4647369	HM155213	########	066XX N B(	1811	NARCOTIC:	POSS: CAN	SCHOOL, P	TRUE	FALSE	2432	24	40	1	18	1164737	1944193	2006	########	42.00248	-87.6693	(42.00247
4647370	HM24508(	########	062XX S W	1330	CRIMINAL	TO LAND	PARKING L	TRUE	FALSE	825	8	15	66	26	1161441	1863309	2006	########	41.7806	-87.6837	(41.78059
4647372	HM17117!	########	058XX S SH	1811	NARCOTIC:	POSS: CAN	STREET	TRUE	FALSE	711	7	20	68	18	1174958	1866097	2006	########	41.78796	-87.634	(41.78795
4647373	HM24480!	########	011XX N SI	810	THEFT	OVER $50(	CHURCH/S	FALSE	FALSE	1121	11	26	23	6	1154100	1907414	2006	########	41.90177	-87.7094	(41.90177
4647374	HM245851	########	080XX S D(	820	THEFT	$500 AND	STREET	FALSE	FALSE	631	6	8	44	6	1184622	1851863	2006	########	41.74867	-87.599	(41.74867
4647375	HM245194	########	064XX S W	910	MOTOR VE	AUTOMOB	STREET	FALSE	FALSE	726	7	15	67	7	1164792	1861919	2006	########	41.77671	-87.6714	(41.77671
4647377	HM156475	########	032XX W D	2027	NARCOTIC:	POSS: CRA	SIDEWALK	TRUE	FALSE	1022	10	24	29	18	1155120	1893275	2006	########	41.86295	-87.706	(41.86295
4647378	HM244545	########	081XX S M	1310	CRIMINAL	TO PROPEI	SCHOOL, P	FALSE	FALSE	613	6	21	71	14	1170069	1850820	2006	########	41.74614	-87.6524	(41.74614
4647379	HM171235	########	080XX S EX	1512	PROSTITUT	SOLICIT FO	STREET	TRUE	FALSE	422	4	7	46	16	1197222	1852411	2006	########	41.74987	-87.5529	(41.74987
4647380	HL762080	########	112XX S W	1340	CRIMINAL	TO STATE S	SCHOOL, P	FALSE	FALSE	2233	22	34	49	14	1174313	1830380	2005	########	41.68996	-87.6375	(41.68995
4647381	HM24595(	########	014XX W G	2830	OTHER OFI	OBSCENE T	RESIDENCE	FALSE	FALSE	2433	24	48	77	17	1165597	1941287	2006	########	41.99449	-87.6662	(41.99448
4647382	HM155236	########	091XX S LC	2027	NARCOTIC:	POSS: CRA	STREET	TRUE	FALSE	2222	22	21	73	18	1168597	1844242	2006	########	41.72812	-87.658	(41.72812
4647385	HM246018	########	092XX S LA	910	MOTOR VE	AUTOMOB	STREET	FALSE	FALSE	633	6	9	44	7	1182521	1843701	2006	########	41.72633	-87.607	(41.72632

Tasks to complete and Timeline

Task	**Estimated Deadline**
Process the data to be more usable ● Reduce the file size by limiting to most recent years ● Clean the data ● Map FBI/community/ward ids to descriptions to better understand the dataset	End of Week 4 (Jan 28)
● Map lat/long location data to zip codes ● Outline of code to evaluate safety score ○ Cluster crimes by severity and assign weights to categories of crime ○ Determine appropriate time-discounting factor	End of Week 5 (Feb 4)
● Continue working on code to evaluate safety score ● Outline of code to compute trend of safety score over time ● Outline of code to visualize findings on graphs and heat maps (Matplotlib/other tools) ● Begin work on web app on Django	End of Week 6 (Feb 11)
● Finalize code to evaluate safety score ● Continue working on code to compute safety score trends ● Continue working on code to visualize findings ● Continue working on web app	End of Week 7 (Feb 18)
● Finalize code to compute safety trends ● Finalize code to visualize findings ● Finalize web app	End of Week 9 (Mar 4)