Report Date: 10/14/2022
To: ematson@purdue.edu, ahsmith@purdue.edu and lee3450@purdue.edu
From: Chasing FOX

- Daye Kim (rlaek4793@khu.ac.kr)
- Hyun Roh (yeshyun@khu.ac.kr)
- Heewon Jeong (jhjmo0719h@chungbuk.ac.kr)
- Jihu Yang (zihooy@handong.ac.kr)
- Juwon Baek (21700340@handong.ac.kr)

Summary
- Each of the crime primary types were classified into the felony or misdemeanor classes decided by Illinois state laws.
- The correlation between 1) crime density map reflecting crime risk, 2) crime density map reflecting only aggravated data, 3) crime density map reflecting aggressive and homicide data and city facility was analyzed, and the top three common features were identified.
- The shortest path between two locations in Chicago was visualized by using the OSMnx library.

What Chasing FOX completed this week:
- **Data analysis**
  - Defined felony or misdemeanor classes of each crime type
    - Each primary type of crime was classified into the several felony or misdemeanor classes of Illinois(class M, X, 1, 2, 3, 4, A, B, C) based on the Illinois General Assembly by the Legislative Information System (Fig 1) [1] [2].
    - To classify the class of crime, IUCR number, which means specific descriptions about the Primary type, was used.
    - In each primary type, the same IUCR (Illinois Uniform Crime Reporting) numbers were grouped together. And IUCR numbers that have the maximum count in each Primary type were extracted.
    - Based on extracted IUCR numbers and Illinois General Assembly Law, each Primary type's class was defined.

Primary Type	IUCR	Case Number				Description				Class	Explanation
		count	unique	top	freq	count	unique	top	freq		
ARSON	1020	267	267	JB286033	1	267	1	BY FIRE	267	2	Arson is a Class 2 felony.
ASSAULT	560	13454	13454	JB344483	1	13454	1	SIMPLE	13454	C	Assault is a Class C misdemeanor.

Fig 1. Example of the classification of each primary types
to several felony or misdemeanor classes

  - Calculated correlation between weighted KDE map and city facility
    - The relative risk of each primary type of crime was calculated according to the maximum sentence for each class.
      - Class Type - Maximum sentence - Riskiness
        - Class A - under 1 years - 2
        - Class B - under 6 months - 1
        - Class C - under 30 days - 0
        - Class 4 - under 3 years - 6
        - Class 3 - under 5 years - 10
        - Class 2 - under 7 years - 14
        - Class 1 - under 15 years - 30
        - Class X - under 30 years - 60
        - Class M - under 60 years - 120

- To get a crime density map, kernel density estimation was newly calculated with the riskiness of each primary type.
- The crime density map was normalized by considering the number of people living inside the Zip Code.
- The results of analyzing the correlation between weighted crime density map and the city facility were as follows.

```
bus              0.65
police           0.60
fire             0.61
library          0.61
abd_building     0.77
support          0.71
lightout         0.73
sanitation       0.71
graffiti         0.27
school           0.69
shotspotter      0.83
y                1.00
```

Fig 2. Correlation between weighted KDE map and facility

○ Calculated correlation between aggravated KDE map and city facility
- Referring to the description of the crime data, only the aggravated data with relatively high degree of crime was extracted.
- To get a crime density map, kernel density estimation was newly calculated with aggravated crime.
- The crime density map was normalized by considering the number of people living inside the Zip Code.
- The results of analyzing the correlation between aggravated crime density map and the city facility were as follows.

```
bus              0.68
police           0.66
fire             0.65
library          0.66
abd_building     0.93
support          0.73
lightout         0.78
sanitation       0.76
graffiti         0.24
school           0.75
shotspotter      0.89
y                1.00
```

Fig 3. Correlation between aggravated KDE map and facility

○ Calculated correlation between aggravated and homicide KDE map and city facility
- Referring to the description of the crime data, only the aggravated data with relatively high degree of crime and homicide data was extracted.
- To get a crime density map, kernel density estimation was newly calculated with aggravated crime and homicide data.

- The crime density map was normalized by considering the number of people living inside the Zip Code.
- The results of analyzing the correlation between aggravated and homicide crime density map and the city facility were as follows.

```
bus            0.68
police         0.66
fire           0.65
library        0.65
abd_building   0.93
support        0.73
lightout       0.78
sanitation     0.76
graffiti       0.23
school         0.75
victim         0.88
shotspotter    0.89
y              1.00
```

Fig 4. Correlation between aggravated and homicide KDE map and facility

- ○ The top three features with the highest correlation in three cases were 'Abondoned_building, Shotspotter, Lightout', and these features will be used in the algorithm the research proposes.

- ● **Algorithm**
  - ○ Conversion of Chicago map to graph (node and edge) was done.
  - ○ OSMnx was used for the conversion. OSMnx is a Python package that allows to download geospatial data from OpenStreetMap and model, project, visualize, and analyze street networks and any other geospatial geometries.
  - ○ By using the OSMnx library, all the intersections in Chicago were expressed as nodes, and the streets between the intersections were expressed as edges. There were 28,701 nodes and 76,169 edges in Chicago.
  - ○ Like Fig5, the street for driving data of 'Chicago, Illinois, USA' was selected and visualized.
  - ○ By using the shortest_path function in the OSMnx library, it returns the shortest path by distance between these nodes. The shortest path needs the location of both orig, and dest. Fig 6 is the shortest path between two locations in Chicago. The location of origin is (41.838706, -87.725954), and the location of destination is (42.004384, -87.674296).
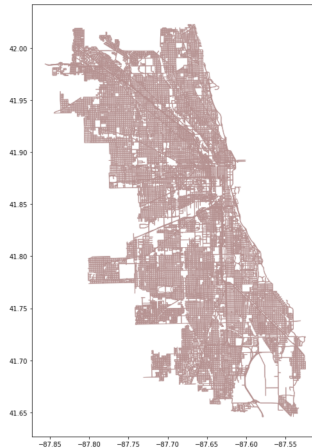
Fig 5. Streets for driving in Chicago

Fig 6. The shortest path between two locations in Chicago

Things to do by next week
- For applying the data analysis into the algorithm, the riskiness score will be calculated based on three features: crime, city facilities, and length.
- For algorithm development, the weight in the shortest path from length will be changed to riskiness score.

Problems or challenges:
- When classifying the class to determine the risk of crime, it was difficult to distinguish it referred to legal documents because there were too many individual IUCR numbers. Therefore, the largest number of IUCRs for each primary type was identified and classified, but there should be some improvement in the future.
- When creating a Crime density map, it would be better if the dataset could consider the floating population as well as the resident population, but it was difficult to find a dataset.

References
[1] *Crimes - Map*, Chicago Data Portal, Sep 15, 2011. [Online]. Available:
https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6
[2] "CRIMINAL OFFENSES" Illinois General Assembly by the Legislative Information System.
https://www.ilga.gov/legislation/ilcs/ilcs2.asp?ChapterID=53 (accessed Oct 12, 2022).
[3] "OSMNX 1.2.2" OSMnx 1.2.2 documentation.
https://osmnx.readthedocs.io/en/stable/index.html. (accessed Oct 14, 2022).