

Project 5. Spatial Clustering and Structure of Crimes in the City of Chicago (150 points)

1. Given the crime report dataset (CSV file) for the City of Chicago for the year 2021 (you are shared with a copy; source: <https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6>)

Note. There are multiple categories of crimes in the provided data, including, among others, Arson, Assault, Battery, Burglary, Criminal Damage, Criminal Sexual Assault, Criminal Trespass, Deceptive Practice, Motor Vehicle Theft, Narcotics, Offense Involving Children, and Others.

2. Task 1 Data **exploration** (30 pts)

- Summarize the frequency of crimes in terms of their categories. You may need also explore the data in terms of season/month of the year, day of the week, and time of the day. (10 pts)
- Based on this exploration, **select three categories of crimes** that you are interested in for the following tasks. You will then work on **only these three selected categories**. **Note** – if you believe some of similar categories should be combined into one more general category, you can merge them into a new category with a proper name. However, the data you will be working with should have three categories after merging.
- Visualize/map the spatial distribution of the crimes in terms of their categories. Use appropriate base map as background. Make sure your maps are interactive. (10 pts)
- Discuss the findings of your exploration. (10 pts)

3. Task 2 Spatial clustering (60 pts)

- **Apply** DBSCAN and at least one of its variations to cluster locations of your selected three crime categories, respectively. Present the clustering results as maps with proper background info and symbolization. Discuss the effects of parameters of the clustering methods. (20 pts)
- Apply one other proper method (not the DBSCAN family) that we discussed (or not discussed) in class to do the same. (20 pts)
- Discuss the parameter settings between different crime categories. (10 pts)
- For each crime category, discuss/compare the result from one of the DBSCAN family with the result from the other method you chose. You may quantitatively compare the positions of the cluster centers, the number of clusters, and the ‘mass’ of the clusters derived from the two clustering methods. (10 pts)

4. Task 3 Spatial structure (60 pts)

- Find a way to quantitatively analyze the distances **between closest cluster** centers respectively for the three crime categories. (20 pts)

- Find a way to quantitatively analyze the distances between closest cluster centers of all three crime categories. (30 pts)
- Examine the spatial structure of crime distribution in terms of distances. Assume we have a hypothesis that different types of crimes in Chicago are actually also clustered or mixed, i.e., they actually happen in the same neighborhood. (10 pts)

5. Task 4 Clustering method implementation (bonus, 30 pts)

- **Implement** the DBSCAN clustering method and evaluate it with reference to the result from the R built-in function. You can choose one crime category for evaluation.

6. Submission

- A report of max 5 pages (use the course template; 1.5 line spacing, 12 fonts, 1" margin in all sides; heading/subheadings)
 - Discuss key programming challenges & your solutions (with relevant code included).
 - Discuss/interpret your main findings based on your maps/calculation with reference to the basemap information of Chicago city.
 - Main results (figures, plots, tables etc.) that support your discussion.
 - Other results can be included as an appendix (not counting to the page limit).
- Your *.r code or *.rmd file with annotations/comments for easy understanding.
- Submission format
 - Your separate report file as **LastName_FirstName_ProjectNo (doc or pdf)**
 - All others in one zipped file (no data file needed)