**Environmental Data Science Final Project**

Jonathan Bernard Widjajakusuma (jbw7)

April 30th 2024

**Background and Introduction (7%)**

My final project stems from my curiosity of what factors actually contribute most to crime rates? I’m specifically looking at economic factors like unemployment rate, and median household income. I believe this is an important factor as the crime in the US have a very high crime rate compared to my homecoutnry of Indonesia. Identifying what causes high crime rates will allow us to lower crime rates everywhere by promoting a solution to solve these ever growing issue. Therefore, I decided to conduct an investigation of the city of Chicago. The reason I chose Chicago is because UIUC is really close to Chicago and I often go there and I believe some areas of Chicago have really high crime rates. Therefore my research question is “What Economic factors contributes most to crime rates in Chicago”.

By identifying which factor contributes most, we can promote solution to help people in need of financial support / something else so that we can hopefully reduce the crime rate in Chicago, therefore innocent people will face less crime. And make the city a safer place.

My hypothesis is that unemployment rate will contribute most because people that are unemployed don’t have income to support their monetary needs, therefore they may resort to crime to have enough money to support their daily needs. These people may not be a bad person but they just have no choice but to steal / harm others to get money.

* Present your overall topic and its importance.
* Describe your research question, its significance, and its associated hypothesis.
* Briefly describe how your question or its relevant topics have been answered by the literature and exiting work.

**Data Used (10%)**

So I used 3 datasets. One of them is the crime dataset which has the following parameters: but I only used … because

Briefly introduce the data sources you used. Be specific about the following

information: official name of the data, data download link (e.g., URL), data version (if

other versions exist), the period of time you used (for example, the data is available

for 1960-2020 but you only used the period of 2000-2020), and other necessary

information such as data resolution and scale, and importantly, the full data citation.

**Methods (30%)**

* Articulate the detailed methods you applied in data wrangling, processing, analysis, visualization, and any other analysis derived from the downloaded data. For example, what variables did you model or how did you construct or model them? How did you explore and present the environmental features modeled? What other spatial operations did you use? Some of the text can come from your proposal. Feel free to use subsections (if needed) to organize the methods.
* Include a working conceptual diagram or flowchart, such as using the Lucidchart (https://lucid.app), to guide the logical flow of your method description.

**Results and Discussions (30%)**

* Present your results and outcomes in detail. Was your hypothesis accepted or rejected? Feel free to use subsections (if needed) to organize the results.
* Identify at least 3 limitations of your analysis.
* Use table and/or graphical (figure, map, and plot) components to facilitate your result discussions. Include a caption for each table or graphical component, and make sure each table/graphical component is described in the report text.

**Conclusions (5%)**

* Briefly summarize your findings and provide your suggestions and/or caveats.
* Rehash the limitations of your work and suggest possible future work or improvements.

**Bibliography or References**

* List all references you cited in the report at the end of the report. Note that all work cited in the report must be listed in the bibliography; likewise, all listed work in the bibliography must also be cited in the text of the report.
* Make sure to use a consistent citation/reference style. See major citation styles here: https://guides.library.illinois.edu/citingsources

Requirements

1. Define what environmental aspect you are analyzing or modeling. Is it the physical/natural, social, cultural, economic, or built environment?
2. Define what your overarching question is? If relevant, propose a hypothesis for your question. An example can be: A good accessibility to parks and green space (built or natural environment) may lead to a better health condition of the local residents.
3. Construct or model the environment, using the data wrangling, attribute, spatial, and geometric operations (and other methods) you have been learning in this course. You must calculate at least one new spatial variable for your analysis, such as (1) the percentage of green space within each of your spatial units/neighborhood or (2) the distance to parks. The spatial scale should be clear (tract? county level? Or your own definition?).
4. Explore and present the environmental features you modeled. This can include visualizations, summary statistics of variables, analysis of the patterns observed (e.g., the spatial autocorrelation of health condition, and the correlation between proximity to parks and health condition), and any other discussions that pertain to your hypothesis or your questions.

Resources:

* Crime data - <https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6>
* Chicago census tracts - <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik>
* Economic data - <https://chicagohealthatlas.org/indicators/UMP?topic=unemployment-rate>