parksAndRec: Understanding the Spatial and Demographic Distribution of Open Space in Chicago

Members:
José María (Chema) Gálvez Enríquez (jmgalvez@uchicago.edu)
Pablo Hernandez Pedraza (phernandezpedraz@uchicago.edu)
Raghav Mehrotra (raghavm@uchicago.edu)
Sarah Hussain (sthussain@uchicago.edu)

Abstract
This project asks two questions: How is open space spatially distributed across Cook County? Who has access to these open spaces?

With this objective in mind, our exploration makes use of an interactive dashboard environment to perform geospatial analysis. We rely on 2018 land use data from the Chicago Metropolitan Agency for Planning, which classifies open spaces into five subcategories: Recreational, Golf Courses, Conservation, Non-Public, and Trails/Greenways. We also spatialize Vacant Land from the same dataset to argue that not all open space can be understood in the same way.

To understand who has access to open space encompassing the five subcategories of interest, we assign each land parcel to a census tract via spatial join techniques, aggregate land use attributes at the census tract level, and match each tract to its socio-demographic 5-year statistical estimates (2014-2018) using the American Community Survey API. These include attributes such as median household income, race, and property values. We synthesize this information into an interactive map. In addition to this visual, we provide the user with a tool that allows them to filter Cook County by specific land use area and understand how that impacts income distributions across Census tracts.

Data Documentation

Data Ingestion

Land Use Inventory (LUI, 2018)
A geospatial dataset containing roughly 300,000 land parcels for Northeastern Illinois. Each parcel has been tagged with a particular land use (e.g. housing, commercial, open space, etc). We download the GeoJSON files as bulk data, which contain latitude and longitude coordinates for each parcel in addition to the land use information, as well as county ID, which allows us to narrow our analysis to Cook County. Our focus is on open spaces, which CMAP categorizes as follows:

Table 1: Open Space Categorization

Land Use Code	Description
3100	Open Space, Primarily Recreation
3200	Golf Course
3300	Open Space, Primarily Conservation
3400	Non-Public Open Space
3500	Trail or Greenway

Later, we broadened the definition of open spaces to include vacant or under-constructed land identified by the following codes:

Table 2: Vacant/Under-Constructed Land Categorization

Land Use Code	Description
4110	Vacant Residential Land
4120	Vacant Commercial Land
4130	Vacant Industrial Land
4140	Other Vacant

Shapefiles for Cook County Census Tracts
For the 2018 geospatial analysis, we used the pygris package to download shapefiles to get the polygon objects for each census tract.

ACS Data (2018)
We queried the ACS data through an API using the **census** package in Python. We specified variables from the ACS five-year data profile. For a given census tract, a five-year estimate contains survey data collected throughout the five years up until a specific wave. We focus on Data Profiles, which provide summarized estimates and percentages on social, economic, housing, and demographic data for geographic areas. For this project, we narrow the data down to the following variables at the census tract level:

Table 3: Variables from ACS Data Profiles

Category	Data Point
Housing & Household	Percentage of Owner-Occupied Housing Units
	Percentage of Renter-Occupied Housing Units
	Median Value of Owner-Occupied Units
	Average Household Size
Educational Attainment	Percentage of Population 25+ (Bachelor's Degree or Higher)
	Percentage of Population 25+ (High School Graduate or Higher)
Demographics & Health	Percentage of Civilian Noninstitutionalized Population with a Disability
	Percentage of Population 16+ in the Labor Force
	Total Population
	Percentage of Population Under 18 Years
	Percentage of Population Aged 65+
Income	Median Household Income
Race & Ethnicity	Percentage White
	Percentage Black/African American
	Percentage of American Indian/Alaska Native
	Percentage Asian
	Percentage of Native Hawaiian and Other Pacific Islander
	Percentage of Some Other Race
	Percentage of Two or more races (includes Hispanic/Latino)

Data Reconciliation

A spatial join was performed using shapefiles for Illinois census tracts boundaries to match a given land parcel in the LUI data to a census tract. Given that individual land parcels within the LUI could potentially intersect multiple census tracts, the centroid of each parcel was used to determine which census tract it lies within.

Following the spatial join, the resulting dataset was merged with American Community Survey (ACS) data using the census tract ID as the key. There were two final datasets. One was unique on census tract ID, with each column containing demographic and socioeconomic variables alongside land use information. The second was unique on land parcels and was a cleaned version of the bulk data we downloaded from CMAP.

Team Responsibilities

José María (Chema) Gálvez Enríquez
- Cleaned LUI bulk data.
- Conducted geospatial joins between the Land Use Inventory (LUI) and Illinois census tract shapefiles.
- Developed spatial visualizations illustrating the distribution of open spaces.

Raghav Mehrotra
- Identified and cleaned LUI bulk data.
- Conducted geospatial joins between the Land Use Inventory (LUI) and Illinois census tract shapefiles.
- Created initial layout for project dashboard and fixed subsequent bugs.
- Developed income visualization tool for the dashboard and assisted with other callback functions.

Pablo Hernandez Pedraza
- Executed queries on the American Community Survey (ACS) census data.
- Generated spatial visualizations depicting the distribution of open space by census tract.
- Developed project dashboard.
- Scraped 2013 LUI geospatial data (unused).

Sarah Hussain
- Executed queries on the ACS census data.
- Cleaned and merged ACS data with LUI data.
- Created spatial visualizations detailing the proportion of open space by census tract.

Final thoughts
We started this project with an interest in how open spaces in Northeastern Illinois are formally delimited and spatially distributed. We broadened the definition to include vacant land; a concentration of these spaces predominantly in the southern regions of Cook County. These areas, currently underutilized, represent untapped potential for recreational development or other strategic projects aimed at stimulating socio-economic growth.

Our dashboard is as interactive as possible because while we enjoyed the spatial and statistical analysis, we wanted users to have clean and reliable data that they could filter and toggle to conduct their own analyses as well. This is specifically the philosophy behind the "Who?" and "Explore" tabs, and future iterations of the project would likely provide the user with many more such tools on the same dataset. Additionally, we provide the user the raw parcel-tract linked data as a local file to enable them to get started on their own.

We had initially intended to compare the open space trends across time by using 2013 and 2018 land use and ACS data. While we were easily able to download the 2018 data as a GeoJSON file, which made processing and cleaning easier, we were unable to do this with the 2013 file. We wrote to CMAP, who were very responsive but mentioned that there was a bug in their portal which allowed only other formats like Shapefiles and SQL databases. The Shapefile format triggered issues with CRS projections, which was interesting to learn about but which we were unable to resolve in time. As a result, we restricted our analysis to 2018, but hope that future iterations of the project can allow users to view changes over time.