Less Vacant Places, More Green Spaces

Group Members
Claire Hemmerly (cehemmerly)
Dustin Marshall (dustinmarshall)
Sabrina Yusoff (sabrinayusoff)
Yuki Chong (yukichong)

Motivated by ever-increasing research highlighting how the availability of green spaces improve public health, especially in times of a global pandemic, this project seeks to provide a dashboard that can be leveraged as a policymaking tool in under-resourced neighborhoods in Chicago. Green spaces and vacant lots have been shown to be on average, distributed unequally within US cities[1]. Furthermore, their distributions have been shown to be negatively and positively correlated with increased health risk and economic hardship indicators respectively. As such, there has been a move towards "greening" of city-owned vacant land in cities and transforming them into green spaces - a case in point would be the Vacant Lot Beautification Pilot in Chicago's South and West Sides[2].

This project has two main objectives:

1) Demonstrate the relationship between the density of green spaces and vacant lots, with public health and economic hardship outcomes, and visualize the disparities across neighborhoods in Chicago.
2) Provide a functional tool for urban planners to identify neighborhoods with disproportionately few green spaces and disproportionately many vacant lots to be considered for conversion into green spaces.

The project aims to achieve these goals by creating two interactive maps displayed side-by-side in a dashboard.

- **"Distribution across Neighborhoods in Chicago":** A choropleth map of Chicago demonstrates the disparities between neighborhoods in terms of hardship score, health risk score, number of vacant lots, number of green spaces, and area of green spaces.
    - At any point, the user can choose one parameter to be displayed on the map. If "Health Risk Score" is chosen, the user can select at least two health indicators to be included in the calculation of the composite health risk score. The health indicators include mental health, physical distress, prevalence of diabetes, high blood pressure, and life expectancy.

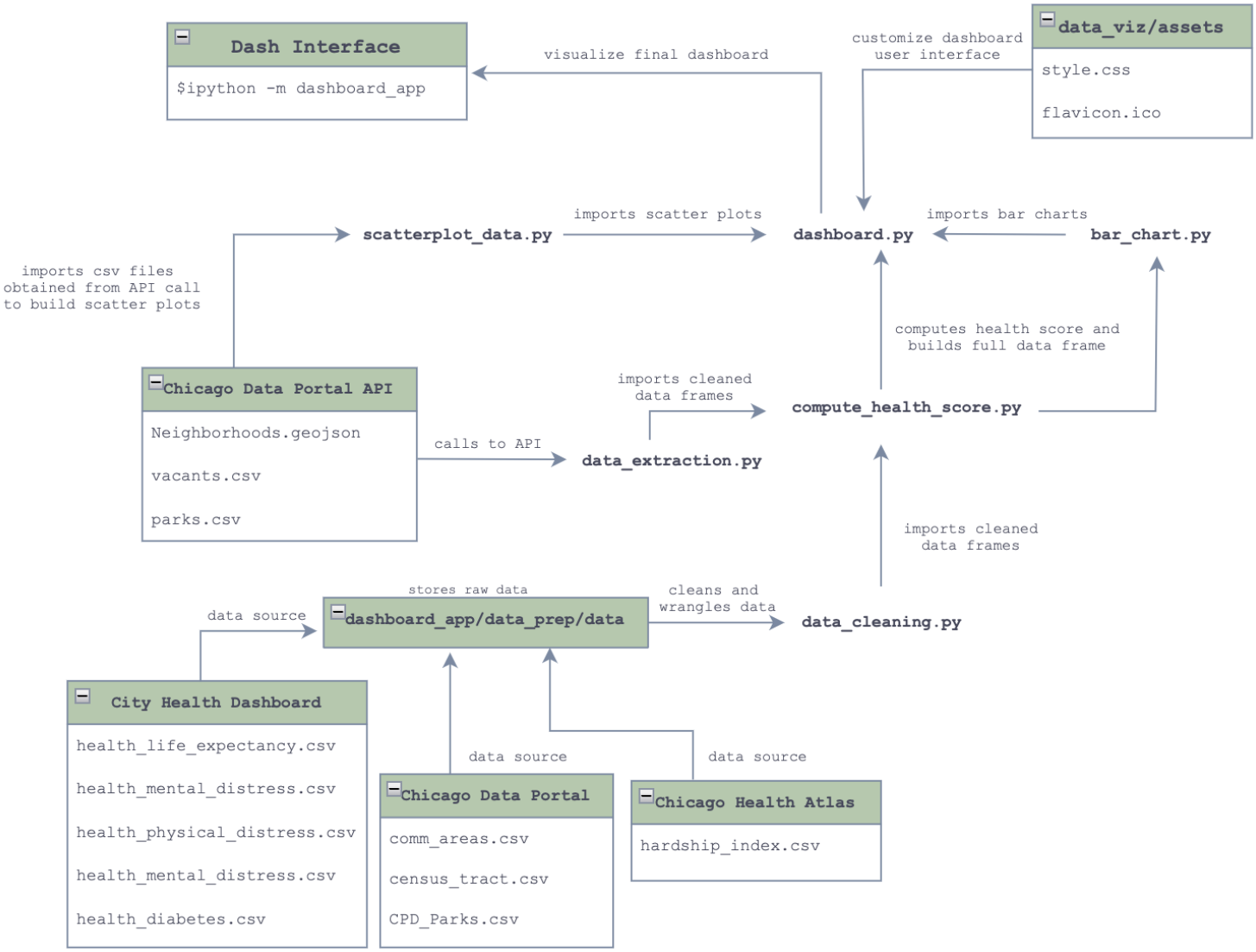
[1]https://www.bloomberg.com/news/articles/2019-03-19/access-to-green-space-varies-by-class-race-in-the-u-s
[2]https://www.chicago.gov/city/en/depts/mayor/press_room/press_releases/2019/june/LotBeautificationProgram.html

- **“Location of Parks and Vacant Lots”:** The second map displays the distribution and size of parks and vacant lots across different neighborhoods in Chicago.
    - The user can select a neighborhood to examine the location and size of both parks and vacant lots. The actual values for each indicator for a given neighborhood are shown in table format below the maps. In addition, bar charts for each parameter display the comparison for a chosen neighborhood with Chicago as a whole.

These maps are supplemented by additional information in the form of:

- **Bar charts** comparing the initial selected neighborhood in the “Location of Parks and Vacant Lots” above with another neighborhood.
    - The second drop-down displays all neighborhoods, excluding the first selected neighborhood. The selection will update five bar charts comparing the two neighborhoods on the key parameters: Hardship Score, Health Risk Score, Number of Vacant Lots, Number of Green Spaces, and Area of Green Spaces.
    - The bar charts provide a visual representation of differences in the key parameters, helping the user identify less-developed neighborhoods and correspondingly, large numbers of vacant lots that can be potentially converted into green spaces.
- A **table** listing key parameters for the selected neighborhoods.
    - The table is updated based on the two neighborhoods selected and health indicators selected.

Software Architecture



We extracted data on green spaces and vacant lots using the Chicago Data Portal API and subsequently processed it in **data_extraction.py**, saving the processed data as a CSV file. The processed data and other data on health indicators and hardship index gathered directly from the Chicago Data Portal, City Health Dashboard and Chicago Health Atlas were cleaned and wrangled in **data_cleaning.py**. The cleaned Pandas Dataframes were then imported in **compute_health_score.py** which merges all data frames and computes the composite health risk score based on the health indicators selected by the user.

The final interactive dashboard is built using Dash, where the file **dashboard.py** produces the three visualizations created by the two files **scatterplot_data.py** and **bar_chart.py** as well as the full dataframe produced by **compute_health_score.py**.

Interacting with the software

1. Clone the repo to your local hard drive (or server):

```
git clone
https://github.com/uchicago-CAPP30122-win-2022/proj-less-parking-more-parks
```

2. Go to the cloned repository:

```
cd ./proj-less-parking-more-parks
```

3. Go to the root folder and install the virtual environment:

```
bash install.sh
```

4. Activate the virtual environment:

```
source env/bin/activate
```

5. Run the program:

```
bash run_program.sh
```

6. Go to the returned address in a web browser
   (Example: “Dash is running on http://127.0.0.1:8053/”)

7. (Optional) Generate the input data files:

```
cd dashboard_app/data_prep
python3 data_extraction.py
```

Workload Distribution

- Dustin: Collected, cleaned and wrangled data on green spaces and vacant lots using Chicago & Cook County APIs
- Claire: Created choropleth and scatter map visualizations in Dash

- Sabrina: Created bar charts and merged dashboard components; Customize dashboard user interface
- Yuki: Collected, cleaned and wrangled health and economic hardship data; Created datatable in Dash; Customize dashboard user interface

Project Accomplishments, Shortcomings and Future Directions

Our goal was to analyze the relationship between availability of green spaces and public health, as well as highlight the possibilities of "greening" city-owned vacant lots to transform vacancies to green spaces. We accomplished this by building an interactive dashboard using Dash that visualizes green spaces, public health indicators (via a composite health risk score), hardship scores, and vacant lots across Chicago neighborhoods selected by the user. Furthermore, we were able to provide a supplementary comparative tool, allowing users to compare two neighborhoods of their choice on the aforementioned metrics.

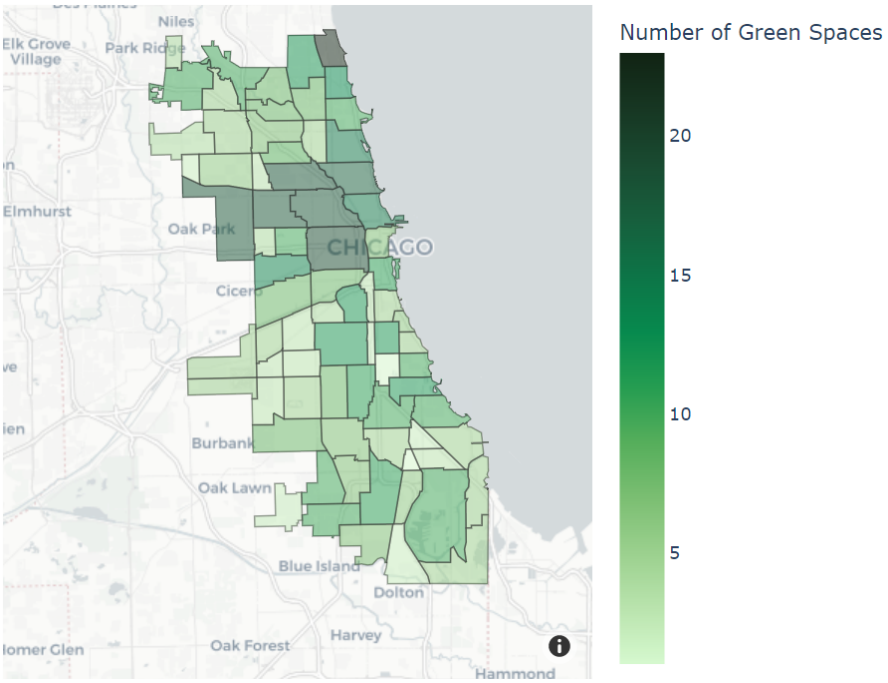
Originally, the plan was to also include city-owned parking lots and low-utility land (as is implied by the repository name), but there turned out to be very few of these owned by the city (<50) in comparison to vacancies (>10,000), so we pivoted to just comparing against city-owned vacancies. We planned on mapping the park and vacancy boundaries through dash, but the intricacies proved too much for the limited scope of the project, so we pivoted to mapping them with scaled dots. We needed to group parks into community areas, but the data wasn't available so we had to convert the addresses to latitudes and longitudes and locate them within the community area boundaries, which was successful. We wanted to be able to compare the total area of park land to vacant land by neighborhood, but the square footage data on vacancies provided by the city was very poor (~90% missing) so we ended up using the parcel boundaries data to generate the missing square footage data (converted to acres for comparability with parks), which was successful and turned out to be surprisingly precise.

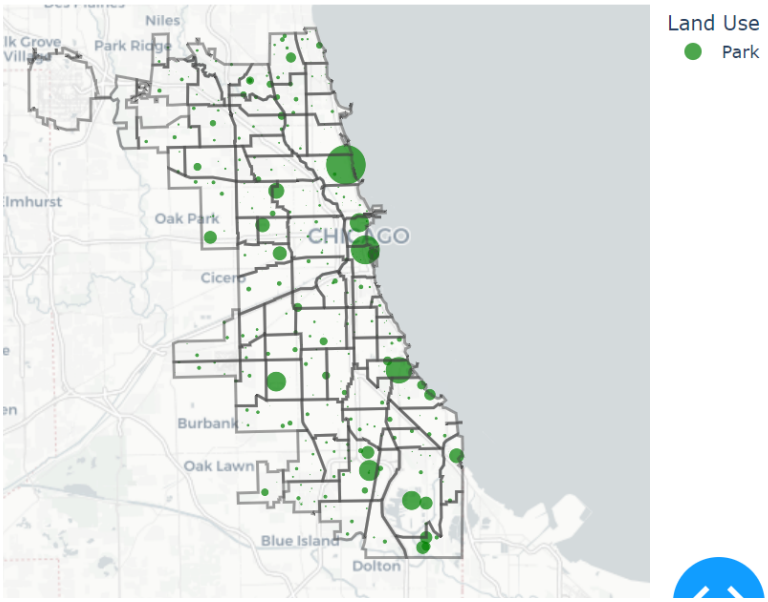
Given more time and resources, in the calculation of the composite health risk score, we would have liked to allow for the different health indicators selected to be weighted differently according to the user's criteria. Due to time constraints, we decided to assign equal weights to all the health indicators selected. In addition, we would have liked to calculate a composite green score, allowing users to choose green indicators like the existing area and size of green spaces, as well as other indicators like quality of green spaces and tree canopy. However, there is insufficient data collected on these measures across all neighborhoods in Chicago, so this would necessitate further data collection by relevant authorities. In terms of user interface, we would have liked to enable the choropleth map to zoom into a specific neighborhood that the user clicked on, instead of having two maps side by side. Given more time, we would use Dash Leaflet to reconfigure this part of our user interface.

Lastly, we collectively tried to implement the subpackage structure with embedded __init__.py files and a __main__.py file in the root directory, but encountered consistent issues with imports that prevented the application from running so we reverted back to our initial set-up with a run_program.sh file.

Interpretation of Results

Distribution of Green Spaces across Chicago



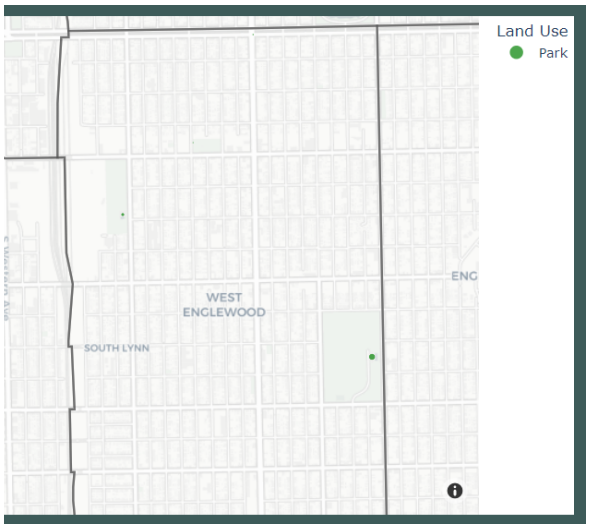


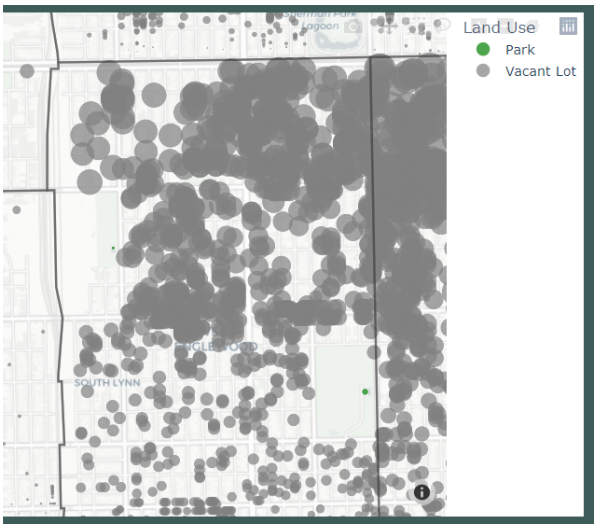
From the “Distribution across Neighborhoods in Chicago” map, we can see that the number of green spaces are generally evenly distributed across the neighborhoods, with more

neighborhoods in the North having a larger number of green spaces. The area (in acres) of green spaces is also evenly distributed across neighborhoods. Comparing this to the distribution of Hardship Score, there appears to be no clear relationship between the economic status of the neighborhood and the availability of green spaces as initially hypothesized.

Comparing neighborhoods

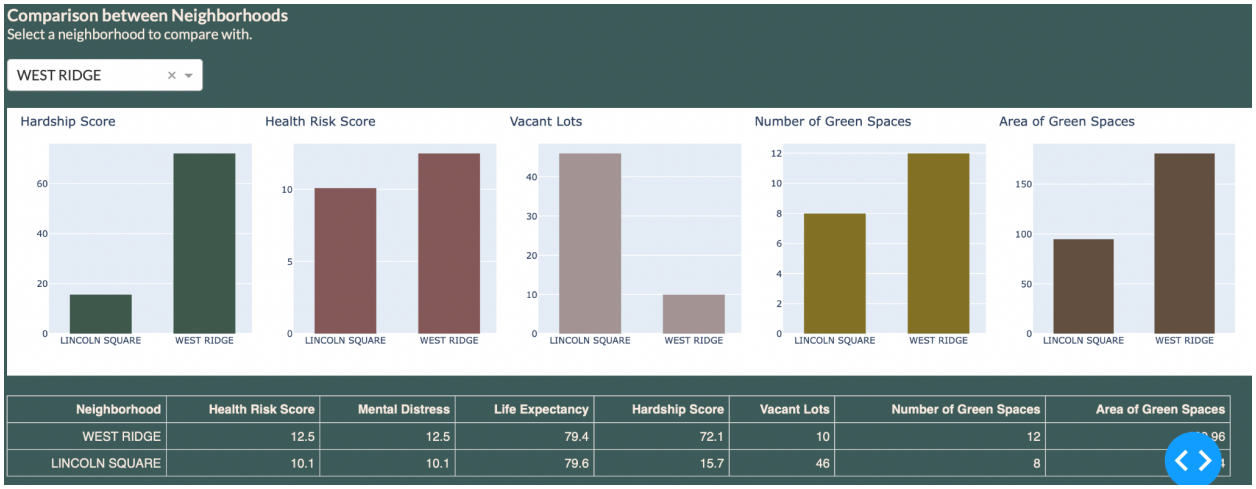
The second part of the application allows us to compare two neighborhoods and analyze them in more detail. For instance, let's compare West Englewood and Lake View. If we select West Englewood in the scatterplot and select only park locations, we get the first screenshot:





There are only two parks in West Englewood, and the small dots indicate that they are relatively smaller than other green spaces in Chicago. When we select the vacant lots as well, we see a large number of underutilized land that can be converted into green spaces.

The bar charts demonstrate clearly the differences between them on the five key parameters.



Neighborhood	Health Risk Score	Mental Distress	Life Expectancy	Hardship Score	Vacant Lots	Number of Green Spaces	Area of Green Spaces
WEST RIDGE	12.5	12.5	79.4	72.1	10	12	
LINCOLN SQUARE	10.1	10.1	79.6	15.7	46	8	

Layout of Dashboard