Neighborhood Finder

By: Elie Nowlis and Christine Tsai

Problem Statement

Chicago is a city of neighborhoods, and there is substantial variation among them. When people are considering moving to Chicago, or moving within Chicago, it can be difficult to figure out which neighborhoods best suit their needs, especially because people have diverse preferences over which neighborhood characteristics matter most to them. To help people with these choices, we set out to make a tool that allows people to input the factors that matter most to them and receive a list of neighborhoods that best fit their needs. We were inspired by tools such as the [OECD Better Life Index](#).

Background

Our first step was choosing which variables to add to our interactive tool. The factors that people care about when moving to a neighborhood vary based on people's stage in life and lifestyle. We identified the following factors that house buyers or renters typically consider when moving to a neighborhood: safety, green space, rent/housing prices, walkability, transit/commute times, access to grocery stores, and age. These are the top factors that people tend to believe contribute to a better quality of life.

Methodology

Walkability and Access to Public Transportation

Walkability and access to public transportation were both obtained from the Chicago Health Atlas. The 2024 walkability index data has a value from 1 to 20 for each neighborhood. The higher index value meant the neighborhood had better walkability based on a few neighborhood factors: intersection density, proximity to transit, diversity of businesses, and density of housing. Transit access data was collected from 2023-2024 and is in the form of a percentage of adults who report ease of getting to a transit stop by walking, scooting, or rolling. The percentage was obtained by taking answers to the question "It is easy to walk, scoot, or roll to a transit stop (bus, train) from my home" from the Healthy Chicago Survey and extrapolating this to a percentage that was weighted to the household population of 18+ adults per neighborhood. Both of these datasets were imported into ArcGIS and spatially joined to our neighborhood polygons.

Green Space

We calculated green space by downloading Chicago Park District Park Boundaries data from the Chicago Data Portal. The park data is in the form of a Shapefile that contains a multipolygon object for each park. We summed up the park acreage for all parks that intersected our neighborhood polygons and used this sum to get a normalized green space score for each neighborhood.

Access to Grocery Stores

Grocery store data was accessed from a [2011 Map of Grocery Stores] from the Chicago Data Portal. This data includes a list of 491 grocery stores across Chicago. We summed the number of grocery store points in each neighborhood polygon as an indicator of grocery store access.

Crime

Crime data was downloaded from the Crime Public AGOL Index compiled by the Chicago Police Department. This data includes data on crimes for the past 365 days and was last updated on February 8, 2024. It includes the following crimes: Homicides, Criminal Sexual Assault, Robbery and Aggravated Battery and Assault, Burglary, Theft, Motor Vehicle Theft and Arson. The data points corresponding to each crime were aggregated in GIS by counting the total number of data points (crimes) in each of our neighborhoods.

School Quality

School quality is measured as the percent of students at a school that are proficient in English language arts and math, which is data collected by the Illinois State Board of Education. Proficiency is defined as "performance levels 4 and 5 on the Illinois Assessment of Readiness, performance levels 3 and 4 on DLM-AA, performance levels 3 and 4 on SAT" in the subject areas of English and math, respectively. This data was last updated on May 30, 2025. We first linked the performance data to the location of each school through school names, using a Fuzzy Lookup in Excel. 80 percent of the names were matched automatically using a .75 similarly threshold. An additional 10 percent were matched manually. In total, 581 schools with test data were matched to the 650 school addresses (89 percent). This data was aggregated in our GIS map by taking the maximum English and math proficiency for any school in the neighborhood.

Age

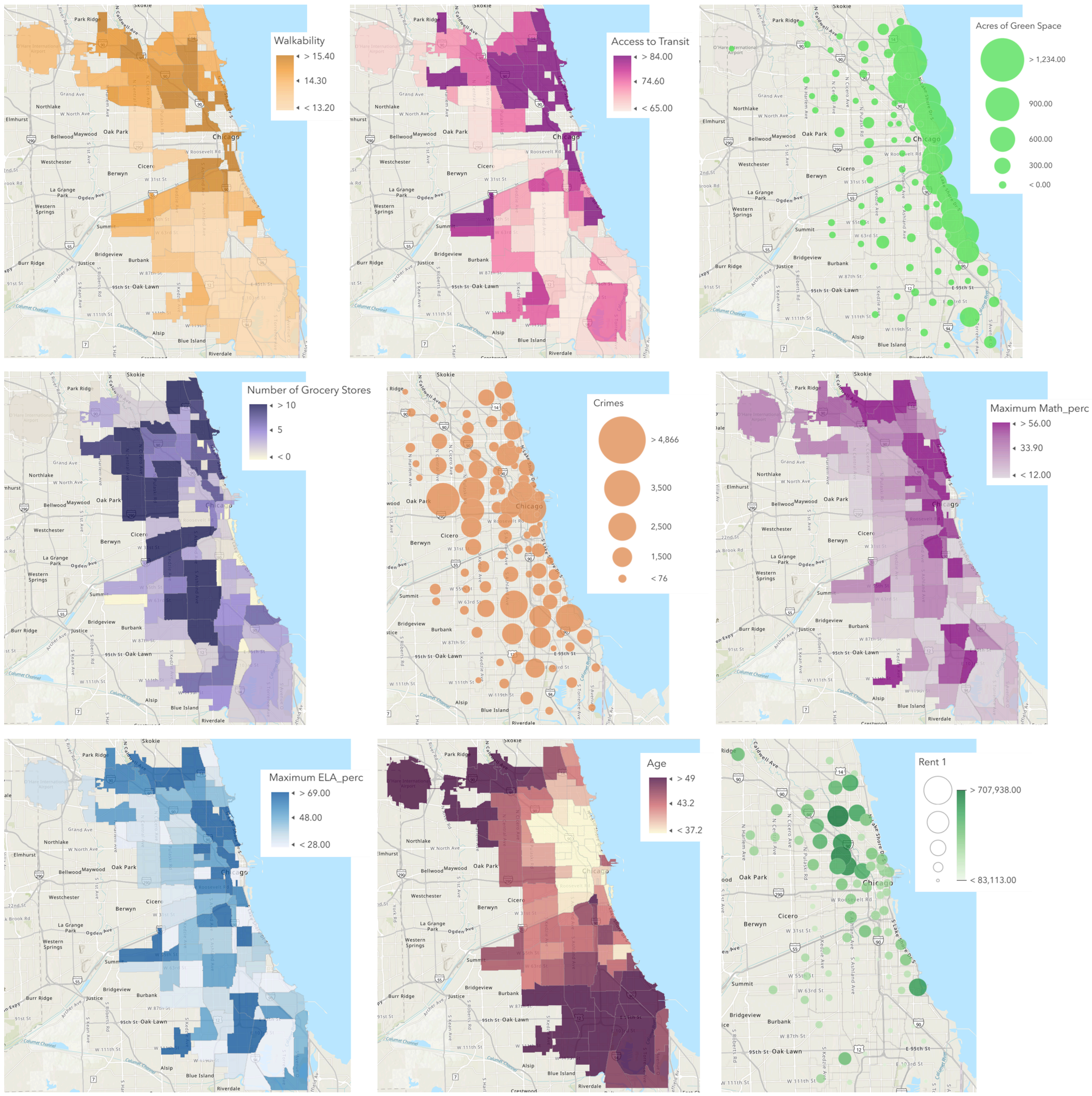
Age data by category (or bucket) is available from the 2020 census. Each bucket spans 5 years of ages between 20 and 85, with the final bucket including people 85+. To find the median age of adults in each census tract (which is the geographic area the data was available in) we found the midpoint number of adults (which we calculated as the total population over 20 and divided in half) and then found in which bucket it lands. We then used the midpoint of that bucket. For example, the midpoint of the 40 to 44 year bucket is 42. To aggregate this data in our GIS map, we took the average median age across all census tracts that intersect each neighborhood.

Home Values

Home value data comes from the Zillow Home Value Index, which is a "measure of the typical home value and market changes across a given region and housing type." It reflects the typical value for homes in the 35th to 65th percentile range. The most recent data available was for September 30, 2025. This data was already aggregated into a single value for each neighborhood, so we matched these neighborhood names to the ones in our map in GIS to add this information. Most of these neighborhoods had the exact same name, but a few were manually matched.

Analysis and Maps

We created the following maps for each factor:



Once all of these factors were aggregated into our GIS map, we turned to Excel to perform some additional analyses. The data input to Excel was our attribute table from the layer with all of our factors.

Attribute Table Excerpt:

PRI_NEIGH	SEC_NEIGH	Crime	Home Price	MAX_ELA_perc	MAX_Math_perc	Median Age	Walkability	Transit	Grocery Stores
Grand Boulevard	BRONZEVILLE	1721	$284,344	87.9	62.5	47	14.26	77.74	4
Printers Row	PRINTERS ROW	215	$268,877	80.6	77.5	39			0
United Center	UNITED CENTE	1052	$405,549	25.4	14.4	40			1
Sheffield & DePa	SHEFFIELD & D	316	$256,535	71.8	56.4	30			1
Humboldt Park	HUMBOLDT PA	3536	$391,889	47.4	30	41	14.21	75.18	18
Garfield Park	GARFIELD PAR	3244	$219,069	41.6	22.3	45			10
North Lawndale	NORTH LAWND	2587	$196,116	31.3	19.8	43	13.76	48.17	3
Little Village	LITTLE VILLAGE	1857	$211,275	34	17.9	42			20
Armour Square	ARMOUR SQUA	270	$320,379	48.1	44.6	47	17.08	68.62	1
Avalon Park	AVALON PARK,	502	$156,293	34.6	38.1	53	13.00	73.49	3
Burnside	CHATHAM,BUR	118	$144,983	32	8.6	52	12.83	70.94	0
Hermosa	BELMONT CRA	627	$318,516	26.7	24.4	42	13.82	72.54	6
Avondale	IRVING PARK,A	1170	$482,025	33.8	21.7	36	15.23	88.93	6
Logan Square	LOGAN SQUAR	2288	$503,987	38.3	23.5	35	15.73	87.78	12
Calumet Heights	AVALON PARK,	715	$188,483	56	34.5	53	13.78	63.50	2
East Side	SOUTHEAST SI	580		48	20.7	46	13.79	58.24	3
West Pullman	WEST PULLMA	1303		32.6	22.9	49	13.74	64.19	5

In Excel we normalized each factor by calculating a z-score (x - avg / std) for each neighborhood. We then transformed these normalized values into a more digestible score of 1-10, in which each z-score was mapped to a bucket corresponding to a number. See below for the buckets we used and an example of this process.

Buckets:

		Good things	Bad things
Z-score min	Z-score max	Score (1-10)	Score (1-10)
-3.6	-3	1	10
-3	-2.4	1	10
-2.4	-1.8	2	9
-1.8	-1.2	3	8
-1.2	-0.6	4	7
-0.6	0	5	6
0	0.6	6	5
0.6	1.2	7	4
1.2	1.8	8	3
1.8	2.4	9	2
2.4	3	10	1
3	3.6	10	1

Example calculations for home price:

PRI_NEIGH	Home Price	HP_norm	HP_score
Grand Boulevard	$284,344	-0.29	6
Printers Row	$268,877	-0.41	6
United Center	$405,549	0.59	5
Sheffield & DePaul	$256,535	-0.50	6
Humboldt Park	$391,889	0.49	5
Garfield Park	$219,069	-0.77	7
North Lawndale	$196,116	-0.94	7
Little Village	$211,275	-0.83	7

Interactive Tool

We created an interactive Python script that takes user input and outputs a visual of the top neighborhoods based on their selected factors.

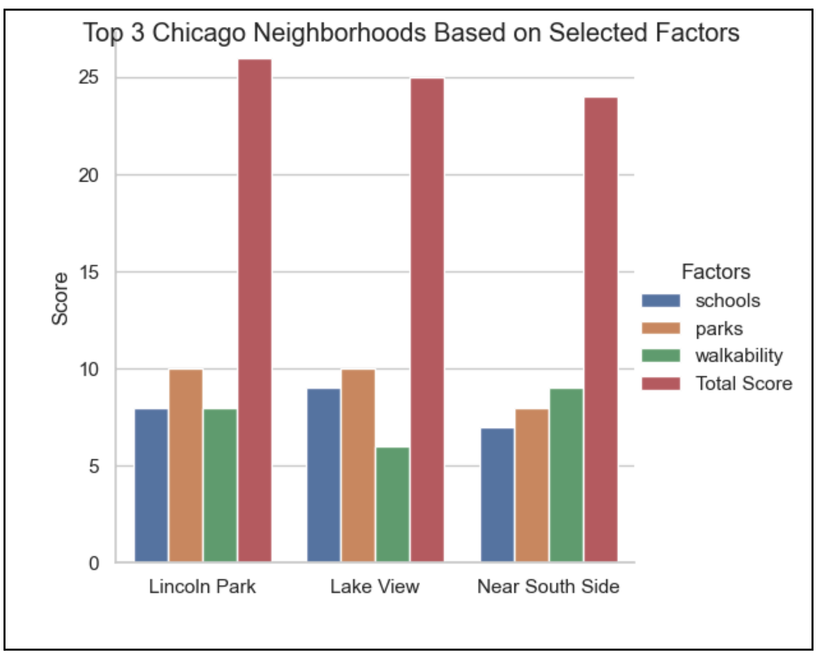
The main steps for the interactive tool were:

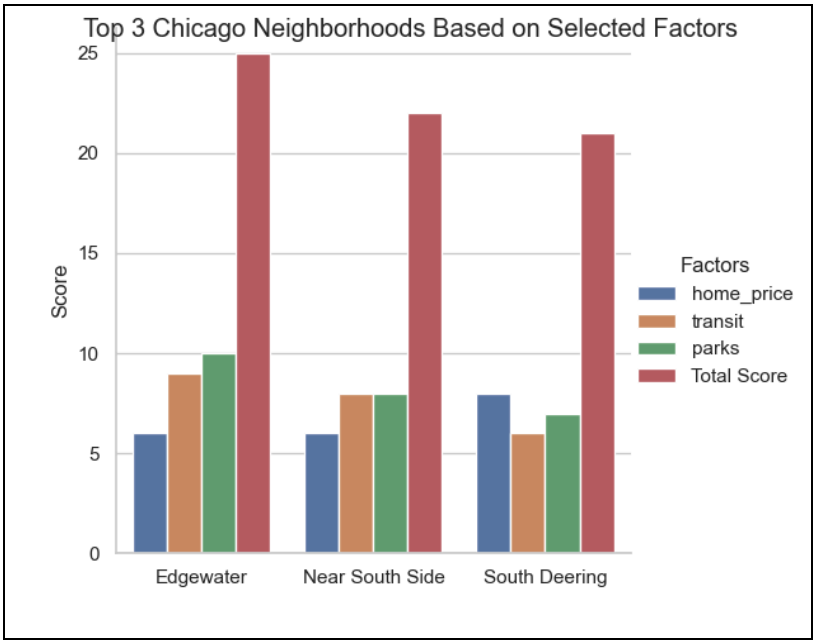
1. Prompting the user for their top 3 factors
2. Storing the neighborhood score data
3. Calculating the top neighborhood scores
4. Visualizing the top neighborhoods

The main interactive part was prompting the user for their top 3 neighborhoods. The tool shows a list of available factors and outputs error messages if the user enters an invalid option or too many options. We used the pandas library for storing and processing the data, and the seaborn library for visualizing the data.

All of the neighborhood data was first imported from a csv file into a DataFrame. Then we calculated a Total Score for each neighborhood based on the top 3 factors that the user input and stored this in a new column. The data was then sorted based on Total Score and converted to a vertically formatted DataFrame so it would be compatible with seaborn's categorical barplot function. Our output displays the top 3 neighborhoods with bars for each factor score as well as the combined score.

Example Output:





Reflection

One limitation of the data was that certain neighborhoods were missing from some datasets. This made the data partially incomplete for certain factors because we had no data to calculate a score. The interactive program treated missing scores as a zero score which meant these neighborhoods were not contenders depending on the factors the user specified. A future extension of our project would include looking for fuller datasets or filling in incomplete data using other dataset calculations.

The interactive program can be expanded to be a web interface instead of a command-line tool. It could also be expanded to allow the user to specify more factors instead of limiting them to three.