Tales of Second City:
An Analysis of Access to Public Services in Chicago

Team Members

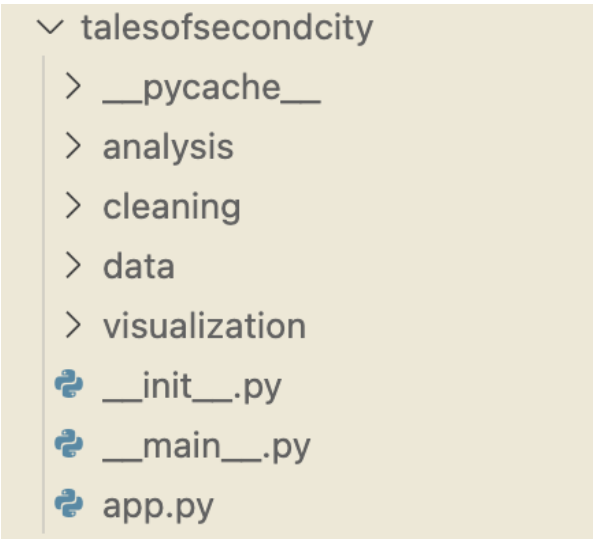
- Suchi Tailor - suchitailor (suchitailor@uchicago.edu)
- Tori Beck - beckv (beckv@uchicago.edu)
- Fatima Irfan - fatimai (fatimai@uchicago.edu)

Project Abstract

This project explores and analyzes census tract level demographic and public services data from the city of Chicago. We collected parks, libraries, and transit data from the Chicago Data Portal. Additionally, we used the Census API to obtain demographic data from the ACS 5 Year Survey from 2012, 2017 and 2022. The three surveys allowed us to explore changes within demographic groups like age, income, education, race/ethnicity, and homeownership from 2008 to 2022. Using data about libraries, parks, bus stops, L stops, and Divvy stations, we created an "Access to Public Services (APS) Index" to parse differences in relative public resource accessibility across census tracts.

Our final product is an app that features various visualizations of our data and analysis through tables, plots, and maps. We mapped current day locations of Divvy bike, bus, and L train stations, public libraries and public parks and created a side by side comparison with a choropleth map of demographic shifts in these neighborhoods between 2008 and 2022. We also mapped our APS Index to observe variation across census tracts. Our analysis allows us to track demographic change across Chicago and identify which areas and populations benefit most from the city's public resources.

Project Folder Structure



Overall Structure of the Software

	Details		Associated Files (from *talesofsecondcity* directory)
Data Collection	Download data from Chicago Data Portal	- Public Library Locations - Public Park Locations - CTA Bus Stops - CTA L Stops - Divvy Stations	data/
	Use Census API to get demographic data	- American Community Survey 2012 - American Community Survey 2017 - American Community Survey 2022	cleaning/census_api/ data/
Cleaning, Geocoding location to census tracts	Clean libraries and parks data to be usable in the Census Geocoding API batch geocoding feature		cleaning/geocode_api/ preprocess_data.py
	Use Census Geocoding API through *pygris* package to batch geocode libraries and parks		cleaning/geocode_api/ address_to_census_tract.py
	Use *geopandas* to geocode bus stops and Divvy stations		cleaning/geocode_api/ address_to_census_tract.py
	Use *geopy* to geocode L stops and all other locations that could not be geocoded through previous methods		cleaning/geocode_api/ point_to_census_tract.py
	Clean all geocoded data for use in index		cleaning/geocode_api/ clean_geocoded_data.py
Database generation	Generate all joined data frames necessary for map visualizations		cleaning/ generate_databases.py
Analysis	Create an access to public services index		analysis/index.py
Visualization	Create dash app		app.py
	Generate maps (using Plotly and Folium)		visualization/maps.py

Code Responsibilities for Each Member

Team Member	Topic	Tasks	Files
Fatima Irfan	Virtual Environment	Initialized repository	poetry.lock pyproject.toml
		Created poetry virtual environment	
	Data Collection	Pulled data from Chicago Data Portal about Divvy bike, bus, and L train stations	
	Data Cleaning	Geocoded transit data	clean_geocoded_data.py, point_to_census_tract.py, address_to_census_tract.py
		Cleaned transit data	
	Data Engineering/ Databases	Set up ETL pipeline, generated merged databases to run visualizations	generate_databases.py
	Data Visualization	Created APS index choropleth map	maps.py
Suchi Tailor	Data Collection	Created functions to pull and filter ACS 5-year data from 2012 and 2017 using the Python Census package.	census_scrape.py variable_defs.py
		Created a separate function to pull and filter ACS 5-year data from 2022 using the API Query process.	
	Data Cleaning	Merged geopandas and ACS data	geopanda_merge.py
	Data Visualization	Developed a choropleth map of	maps.py

		demographic shifts within census tracts from 2007-2022. Created a map containing location markers for all train stops, divvy bike stations, libraries and parks	
Tori Beck	Data Collection	Pulled data from Chicago Data Portal about library and park locations	
	Data Cleaning	Cleaned parks and libraries data to prepare for geocoding	preprocess_data.py, point_to_census_tract.py, clean_geocoded_data.py, address_to_census_tract.py
		Geocoded public service data, including all missing locations that couldn't be geocoded initially	
		Cleaned parks and libraries data for use in index	
	Data Analysis	Identified metrics and developed code for public access index	index.py
	Data Visualization	Wrote code for Dash app layout	app.py
		Created APS index table and scatterplot	
		Wrote code for callbacks to make visualizations interactive	

How to Interact with the App

This application produces an interactive dashboard that enables users to explore the levels of public service access across census tracts in Chicago, as well as how they may relate to factors of demographic change in the past 5, 10, and 15 years. To run this application, you can follow these steps:

1. Clone the repo for tales-of-second-city using the url on GitHub
2. In the root directory, run *poetry install* to install dependencies
3. Run *poetry shell* to activate the virtual environment
4. Run *python -m talesofsecondcity* to launch the application
5. Paste the local url into your browser

After completing these steps, you will find several interactive visualizations, such as:

- a scatterplot depicting the relationship between APS Index and its components
- a choropleth depicting the distribution of APS Index by census tract
- a choropleth depicting percent change in demographic factors (i.e., percent homeowners, percent white population, median household income, etc.) over the past ten years
- a layered choropleth depicting distribution of demographic factors over the past 5, 10, & 15 years
- a labeled point map with layers for each type of public service (libraries, parks, transit stops)

What We Tried to Accomplish vs. What We Actually Accomplished

Our group was interested in looking at whether access to public services will change following changes in demographic factors, specifically those known to be correlated with gentrification. After looking into the literature, we decided to look into factors like the percentage of homeowners, white population, young population, and educated population, and how these have changed over time in recent years in the city of Chicago. Although we were initially interested in including a time component to the public services piece as well, we found that year over year data about when specific park, library, and transit locations opened to the public was difficult and very time consuming to collect. As a result, we shifted our focus to seeking to understand what public service access looks like in census tracts that have experienced change along factors relating to gentrification, and whether this can help us say something about whose needs are being prioritized.

We calculated an Access to Public Services Index by counting the number of libraries, park acres, bus stops, L stops, and Divvy stations in each Census Tract throughout the

city. Then, a proportion for each public service was calculated, for example, $\frac{\#\ park\ acres}{total\ population\ in\ census\ tract}$. A score from 0-100 was then calculated for each public service. This score is relative, where the tracts with the smallest proportion have a score of 0 and the tracts with the highest proportion have a score of 100, with all other tracts falling in between. The bus stop, L stop, and Divvy scores are combined to compute an overall transit score. Then, the index is calculated using a weighted average, where the parks score, libraries score, and transit score are all weighted to be a third of the final index, which uses a scale of 0 to 1.

When observing the index values, we can see that they are all very small (<0.5) and very close together. There could be many reasons for this, but likely one of the biggest factors is that we are looking at the tract level, rather than the neighborhood or zipcode level. Census tracts are very small, and in comparison to the population, the number of public services is relatively small. For example, most census tracts have a library score of 0 because there are no libraries within the tract. There is a maximum of one library in any given census tract. This severely biases the index downwards. In future iterations of the index, calculating the scores and index for neighborhoods would provide more insight into access to public services. However, given our time and capacity, we found that accurately mapping census tracts to neighborhoods would be a somewhat ambiguous and complicated process. Additionally, many social scientists determine access to public services as the average time it would take to get from one's home to the service. This kind of geospatial analysis would be an interesting approach to the question of accessibility for different populations.

Although much of the index choropleth map is relatively homogenous, one is able to observe a spot of yellow, representing high API index, near the area of the University of Chicago. Using the demographic change choropleth, one is able to observe several interesting findings, such as that in the past ten years, the area near the Loop has experienced a slight increase in median household income and that there has been a decrease in percentage white population in a few areas of the city. However, census tracts where these changes have occurred do not correspond with census tracts that visually had higher or lower API index values.

For future work, it may be interesting to incorporate additional public services into our index. Further, we could pool demographic factor change into an index, perhaps a proxy for gentrification, and try to identify which census tracts experienced change along multiple axes. We could validate this index by zooming into neighborhoods we know to have been gentrified, and potentially use it to identify neighborhoods at risk of gentrification.