Team Name: Superteam
Members: Echo Nattinger (enattinger), Paul Soltys (pjsoltys), Dorka Frisch (dorottyafrisch), Andrew Baker (ajbaker)

Project Abstract:
The aim of this project is to explore demographic change in Chicago neighborhoods in connection to economic and housing changes. Specifically, we are interested in finding dramatic shifts in neighborhood makeup and connecting them to shifts in local housing markets and business activity.

We started by querying the United States Census American Community Survey (ACS) for a variety of demographic variables in Cook County census tracts. Secondary data was gathered through three sources: City of Chicago Data Portal, the DePaul Housing Institute, and the Law Center for Better Housing. These data sources cover information on Chicago evictions, business licenses, building permits, and vacancy violations.

Using a geographic fuzzy matching algorithm, we grouped our ACS data (tract-level) to Chicago community areas. We then conducted analysis on the ACS data, studying percentage point change in variables of interest (gender, race, age, education, income, household income, and ethnicity) across different time periods.

We found that some of the most dramatic changes occurred in racial and income makeup of the following neighborhoods: East Side, Hegewisch, Avondale, Fuller Park, West Englewood, Burnside, South Lawndale, and Archer Heights. Our conclusion includes more information on our findings, theories, and future inquiries.

Structure of Software:
This project is separated into four processes: querying, data cleaning, data analysis, and visualizations.

- Querying
    - Pulls the raw Census data for all census tracts in Cook County, Chicago
    - Grabs variables related to gender, race, ethnicity, education, income, household income, age, and population
    - Places .csv files in chicago_demographic_change/data/raw_data
    - NOTE: census_API_key will need to be set as an environment variable in order for this process to run.
- Data cleaning
    - Cleaning functions and .py files are located in chicago_demographic_change/data
    - Places cleaned .csv files in chicago_demographic_change/data/clean_data

- Data analysis
    - Contains a variety of functions to study percentage point changes in neighborhood demographics that are used in the final app
- Visualizations
    - Generates heat maps of Chicago community areas for each demographic within a given variable and shows changes across two time periods. The plots are placed in chicago_demographic_change/visualization/finished_graphs
    - These visualization functions are used in the final app, see guide for more

Code Responsibilities:

Each member contributed specific modules/functions to the software, as detailed below:

- Echo Nattinger
    - Built the querying module, which completes census queries for all of our variables of interest for four date ranges (2005-2009, 2010-2014, 2015-2019, and 2018-2022) and deposits the data in raw_data subfolders
    - Built conversion dictionaries for the analysis team to create comparisons between equivalent variables when census years changed variable names
- Paul Soltys
    - Built the .py files that read in and clean our secondary data sources, got secondary data sources from the internet
    - Built the geomatching algorithm in geomatching.py that compares census tract and zip code geographies to Chicago community area geography to find which tracts and zip codes are included in different areas
    - Built secondary data source cleaning for city and eviction data sources and worked with Dorka to build secondary data classification (low/med/high)
- Dorka Frisch
    - Built the cleaning modules using Andrew Baker's dictionaries and combined them into cleaning.py, a module which cleans all the census data
    - Built secondary data source cleaning for the DePaul Index
    - Built the data_analysis module, including the primary analysis algorithm used to study percentage point change in demographic variables
    - Bundled the functions into the final final.py executable application
- Andrew Baker
    - Wrote the dictionaries for the cleaning module, helped with initial data exploration, and wrote the combine_columns function
    - Built the visualization module, including all visualizations used in the final report as well as several exploratory visualizations for the team

Project Guide:

This project can be executed as an app by running the following command from the root directory: py -m chicago_demographic_change

You can find the final visualizations in chicago_demographic_change/visualization/finished_graphs – these represent the visualizations the user generated.

Note that what our app executes is solely the analysis and visualization processes, using the data already in the repository. We did this for efficiency reasons – API querying, cleaning, secondary source reading, and geomatching take some time. However, these processes are still executable; read below for more information.

python -m chicago_demographic_change.querying.generate

- Executes the querying process using an API (API key must be set, read below)

python -m chicago_demographic_change.data.secondary_source

- Reads and cleans secondary data sources. Will take a few minutes to run. Outputs can be seen in the data/clean_data/Secondary Sources folder.
- 2 functions, city_blc_clean and city_permit_clean, are run on data files too large to pass to Github. The cleaned CSVs are provided for these for use in the rest of the code.
- Building Permits: https://data.cityofchicago.org/Buildings/Building-Permits/ydr8-5enu/about_data
- Business Licenses: https://data.cityofchicago.org/Community-Economic-Development/Business-Licenses/r5kz-chrr/about_data

python -m chicago_demographic_change.data.cleaning

- Completes all cleaning for census data and stores them in data/clean_data

Note: In order for the querying module to run, a valid census API key must be set as an environmental variable with the following command

export census_API_key="YOUR KEY HERE"

Project Conclusion (Goals and Accomplishments):

Originally, this project sought to study shifts in Chicago neighborhood demographics as related to cultural shocks: specifically, the housing crisis of 2008 and the COVID-19 pandemic. However, we quickly realized that discerning shocks at specific points in time would be difficult. Census data for specific tracts are only available for 5 year American Community Survey (ACS) estimates. The earliest 5 year estimate available is 2005-2009, which includes the year of 2008 housing crisis. Similarly, the most recent 5-year estimate available is 2018-2022, which includes the COVID-19 “shock” period in it.

We thus shifted to studying demographic change in Chicago as it relates to economic and housing shifts, not specific historical events. We built our app to allow users to explore the largest magnitude changes in community area makeup in Chicago from 2005-2022. With the app, users can visualize changes of different demographic variables, and explore contemporaneous change in economic measures such as housing prices, business activity, and evictions.

While we did not accomplish the original goal of connecting demographic change to cultural shocks, we did accomplish a different, equally important goal – a way for people to explore how Chicago has changed, on both demographic and economic measures. Interestingly, while racial makeup was often a volatile variable, ethnicity was not, meaning Hispanic neighborhoods were relatively unchanged over the time investigated. This prompted us to think about what the census is actually measuring: identification. It is possible that the large racial shifts this project found were actually shifts in people's *identification* with various demographic labels.

When comparing our community area observations to our secondary data sources, we found that the areas that saw sharp increases in higher-educated population also saw high increases in housing prices. However, the neighborhoods that saw large shifts in racial makeup only saw, in general, medium shifts in housing prices. Our focus on housing prices as a variable of interest mostly stems from data availability – other secondary data sources were not available for all of the years for which we gathered census data. If given more time with this project, we would have liked to explore more secondary data source options.

Overall, our app is a great way to explore the demographic changes in Chicago and is a handy guide in revealing the biggest demographic shifts in the time periods given by the census data. Going forward, this tool can be used to narrow down areas for further research, while providing a clear and visually appealing way for any person interested in Chicago to better understand the changes happening in the city.