

[Weekly Report]

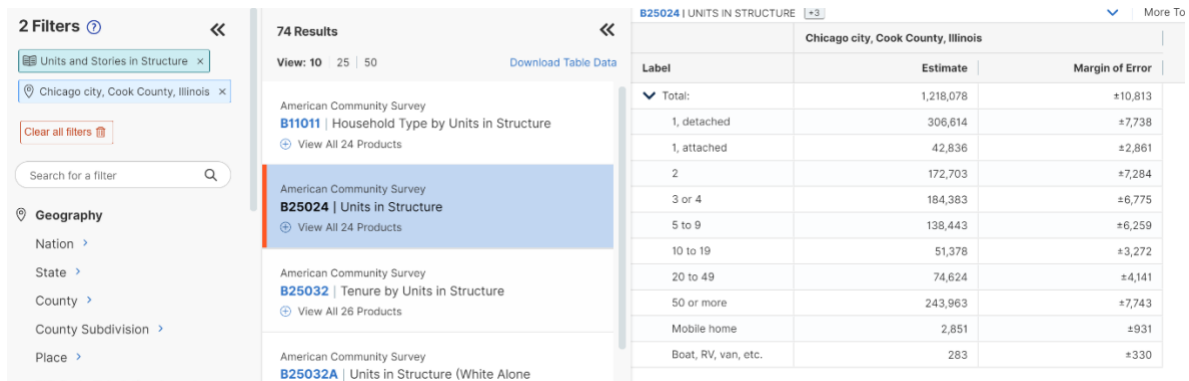
- Report No.4
- Date: Nov. 29, 2023~ Dec. 5, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		Highlights	Self Evaluation
Last Week	Summary	<ul style="list-style-type: none"> • joining file (geo1, geo2, geo3) and creating integrated DB for analysis • create an integrated database to balanced panel. • To create balanced panel data that is consist of three distinct observations 	Medium
Current Week	Baseline Goals (Given)	<ul style="list-style-type: none"> • Join the merged file(geo_master) and relevant queries we found in week2. • Create a data frame by selecting at least two or more queries from the ACS. • Make some implications via the comparison of statistics between geo_master and ACS data 	Medium
	Additional Goals (O.Y.O)	<ul style="list-style-type: none"> • Comparing and analyzing the values from the GEO files against certain characteristic of master DB 	Medium
	Key Issues to Be Resolved	<ul style="list-style-type: none"> • the selected queries from the ACS should be suitable for our analysis. 	Medium
	Strategies	<ul style="list-style-type: none"> • Step 1: We should check duplicates, missing values, and any other issue that must be resolved. • Step 2: We choose the "Units in Structure" data filtered in reference area. Then We would extract the data from selected data. • Step 3: We will merge the data from ACS and Geo file DB. • Step 4: After all, we have comparison the property type of community areas, Chicago City and Illinois. 	Medium
	Results	<ul style="list-style-type: none"> • the community area has a relatively high percentage of single-family houses. • Chicago is close to the center; the population density is high. • the reference area of Illinois, the population density will be low 	Medium
	Implications	<ul style="list-style-type: none"> • the distribution of house units is closely correlated with both population density and economic significance 	Medium
Next Week	Things to Do	<ul style="list-style-type: none"> • Based upon master DB, we will construct final analysis for insight. 	Medium
Remarks	Core Libraries & Packages	<ul style="list-style-type: none"> • Data Cleansing: dplyr • Grouping: dplyr • Visualization: ggplot2 	Medium
	Additional Remarks		

- Remarks
- Work Progress

	Week 1	Week 2	Week 3	Week 4	Week 5
Yang Junseob	EDA	Reconduct EDA	Joining file	Make implication	-
Ryu Seung Gwon	EDA	Reconduct EDA	Joining file	Merging file	-

Appendix1. Create the list of queries(filters) for further analysis.

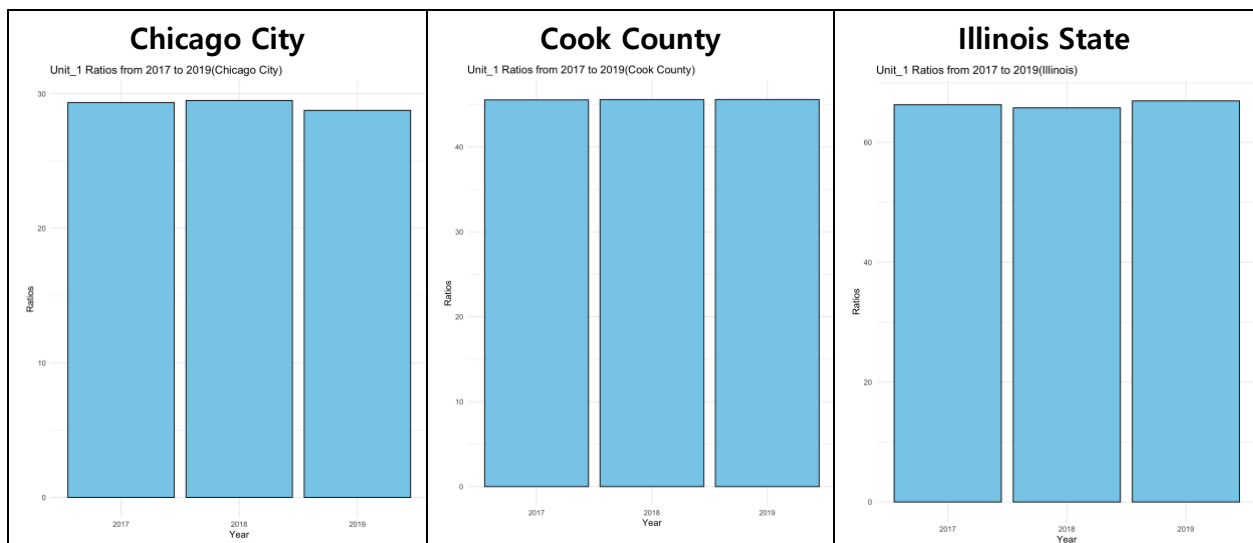


Reference Area: Chicago city, Cook County, Illinois,

Selected data -> B25024 Units in Structure

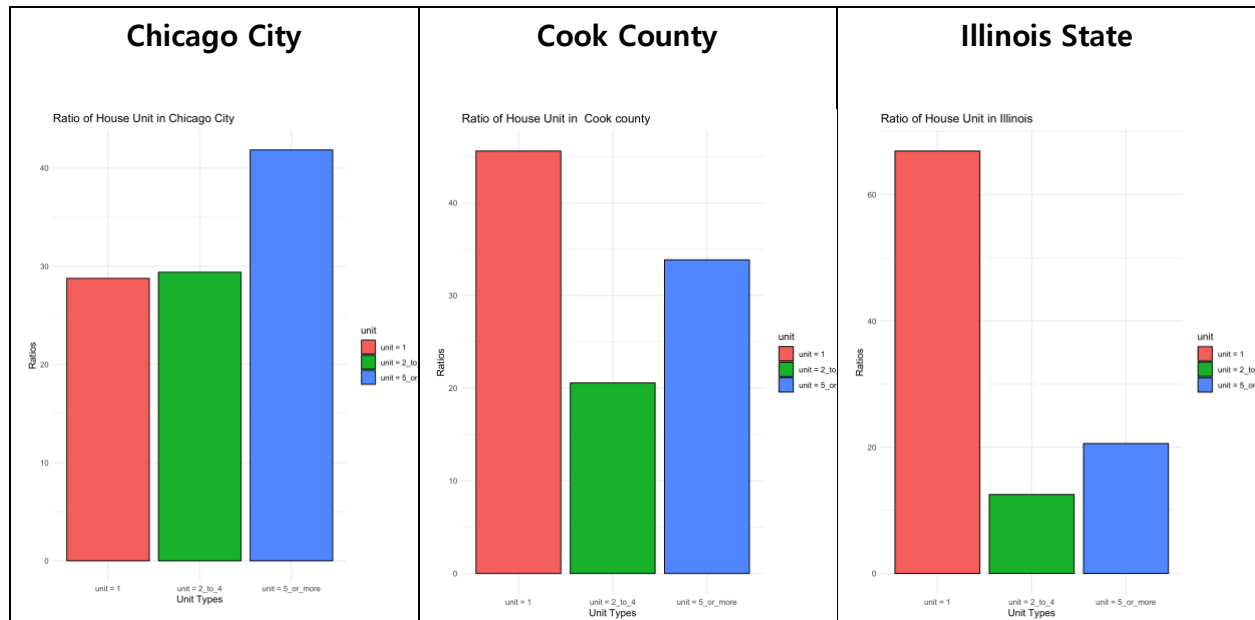
the reference areas were set as 'Chicago city', 'Cook County', and 'Illinois' to understand the distribution of house units. The distribution of units by each reference area was taken from 2017 to 2019 data and merged. After that, it was merged with the geo data.

Appendix 2. Changes in the total number of households over time



: Upon closely examining the selected ACS (American Community Survey) data, We determined that **there was little difference in units for each year**, and the latest data, 2019 data, was decided to be used. The distribution of units for each reference area was visualized and the results were derived.

Appendix 3. House unit distribution in the reference area (as of 2019)



We compared the reference area data based on the criteria of the number of house units: **single family house** (unit=1), **multi-family house** (2 to 4) and (five or more).

'**Chicago city**' was **28%** when the unit = 1, 29% when the unit was 2 to 4, and 41% when the unit was 5 or more.

'**Cook County**' was **45%** when the unit = 1, 20% when the unit was 2 to 4, and 33% when the unit was 5 or more.

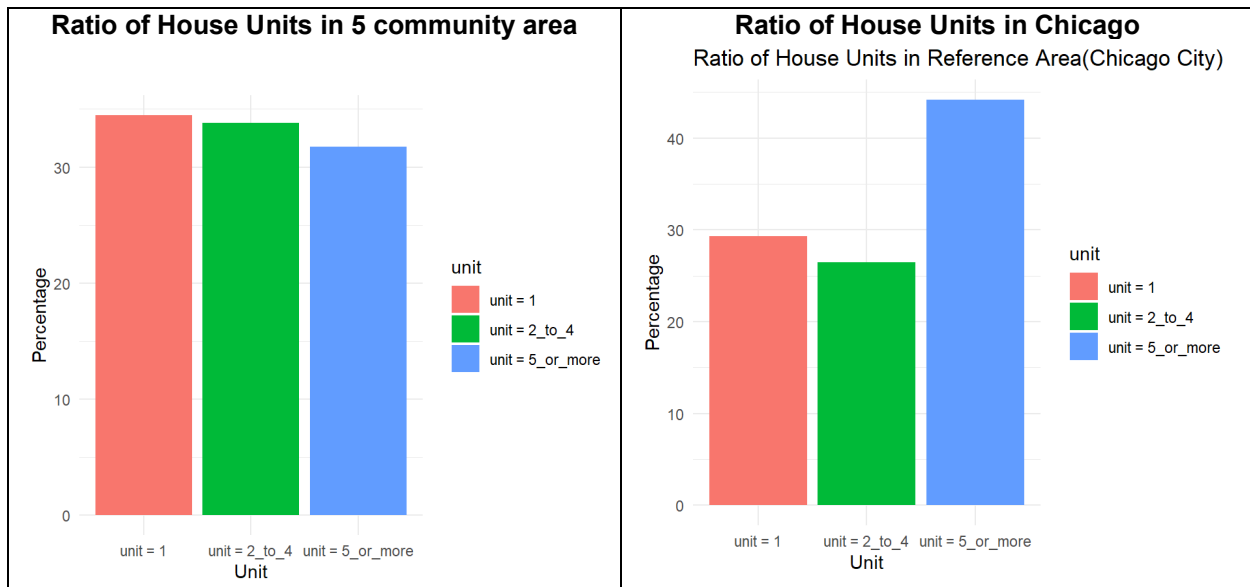
'**Illinois**' was **66%** when the unit = 1, 12% when the unit was 2 to 4, and 20% when the unit was 5 or more.

Since **Chicago is close to the central, the population density is high**, and the proportion of buildings and complex buildings is high, so the proportion of unit 5 or more is high.

However, **because the reference area of Illinois is large, it includes areas far from the center**. Therefore, **the population density will be low**, and the proportion of detached houses is high, so unit 5 or more is high at 66%.

In other words, the larger the range is set as a reference area, the higher the proportion of single-family houses and single-family households because areas with low population density are included.

Appendix 4. House unit distribution comparison between the reference area and Community Area



According to the given graphs, it can be observed that **the 5-community area has a relatively high percentage of single-family homes**, while Chicago has a higher proportion of multi-family residences.

Rogers Park, West Ridge, Uptown, Lincoln Square and North Center belong to uptown. The community area is part of Chicago but does not belong to the downtown area. uptown is usually a residential area that is north or uphill of downtown.

Downtown is often marked by a cluster of tall buildings, cultural institutions, and public transportation so they have higher proportion of multi-family residences, while uptown is often characterized by lower-density housing, green spaces, and local businesses.

In conclusion, the distribution of house units is closely correlated with both population density and economic significance.