

## [Weekly Report]

- Report No.5
- Date: Dec. 6, 2023~ Dec. 12, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		Highlights	Self Evaluation
Last Week	Summary	<ul style="list-style-type: none"> <li>• Join the merged file(geo_master) and relevant queries (Units in Structure) we found in week2.</li> <li>• Create a data frame by selecting at least two or more queries from the ACS.</li> <li>• Make some implications via the comparison of statistics between geo_master and ACS data.</li> <li>• Comparing and analyzing the values from the GEO files against certain characteristic of master DB</li> </ul>	Medium
Current Week	Baseline Goals (Given)	<ul style="list-style-type: none"> <li>• Update and upgrade the overall process of your project overtime.</li> </ul>	Medium
	Additional Goals (O.Y.O)	<ul style="list-style-type: none"> <li>• Additional Comparison between Geo file and ACS data to using new queries.</li> </ul>	Medium
	Key Issues to Be Resolved	<ul style="list-style-type: none"> <li>• The selected queries from the ACS should be suitable for our analysis.</li> </ul>	Medium
	Strategies	<ul style="list-style-type: none"> <li>• Step 1: We should check duplicates, missing values, and any other issue that must be resolved.</li> <li>• Step 2: We choose the "Units in Structure" data filtered in reference area. Then We would extract the data from selected data.</li> <li>• Step 3: We will merge the data from ACS and Geo file DB.</li> <li>• Step 4: After all, we have comparison the property type of community areas, Chicago City, Cook County and Illinois.</li> </ul>	Medium
	Results	<ul style="list-style-type: none"> <li>• The community area has a relatively high percentage of single-family houses.</li> <li>• Chicago is close to the center; the population density is high.</li> <li>• The reference area of Illinois, the population density will be low.</li> </ul>	Medium
	Implications	<ul style="list-style-type: none"> <li>• The distribution of house units is closely correlated with population density, distance from the center and economic significance</li> </ul>	Medium
Next Week	Things to Do	<ul style="list-style-type: none"> <li>• This report is the final outcome of our team project, therefore there are no plans for the following week</li> </ul>	Medium
Remarks	Core Libraries & Packages	<ul style="list-style-type: none"> <li>• Data Cleansing: dplyr</li> <li>• Grouping: dplyr</li> <li>• Visualization: ggplot2</li> </ul>	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	Final report
Ryu Seung Gwon	EDA	Reconduct EDA	Joining file	Merging file	Final report

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

2 Filters

Units and Stories in Structure

Chicago city, Cook County, Illinois

Clear all filters

Search for a filter

Geography

- Nation
- State
- County
- County Subdivision
- Place

74 Results

View: 10 25 50

Download Table Data

American Community Survey  
**B11011** | Household Type by Units in Structure  
View All 24 Products

American Community Survey  
**B25024** | Units in Structure  
View All 24 Products

American Community Survey  
**B25032** | Tenure by Units in Structure  
View All 26 Products

American Community Survey  
**B25032A** | Units in Structure (White Alone)

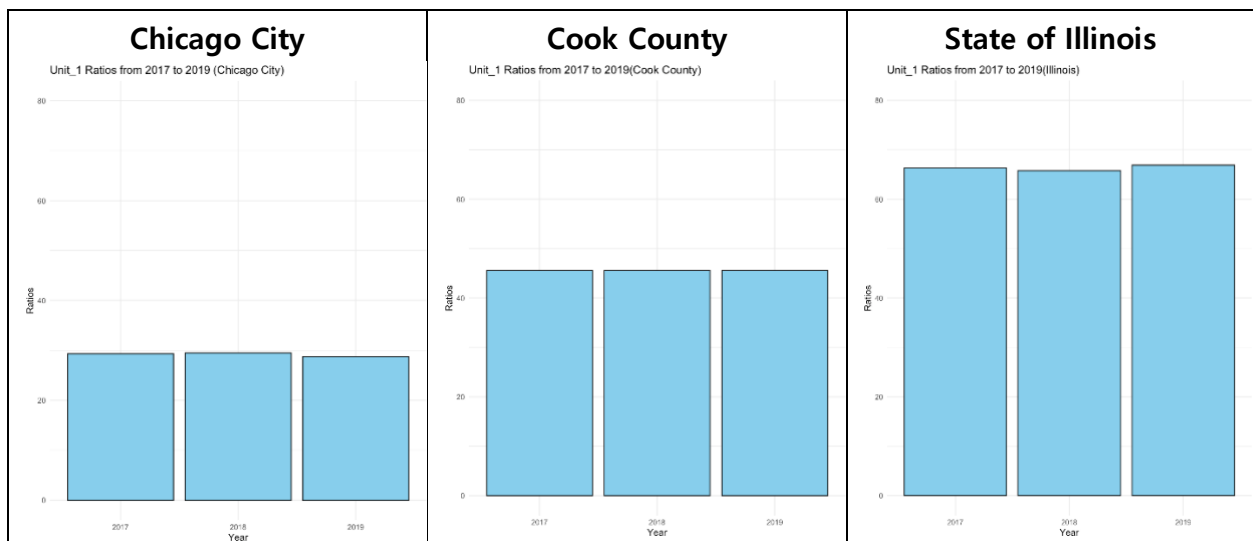
Chicago city, Cook County, Illinois		
Label	Estimate	Margin of Error
Total:	1,218,078	±10,813
1, detached	306,614	±7,738
1, attached	42,836	±2,861
2	172,703	±7,284
3 or 4	184,383	±6,775
5 to 9	138,443	±6,259
10 to 19	51,378	±3,272
20 to 49	74,624	±4,141
50 or more	243,963	±7,743
Mobile home	2,851	±931
Boat, RV, van, etc.	283	±330

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.

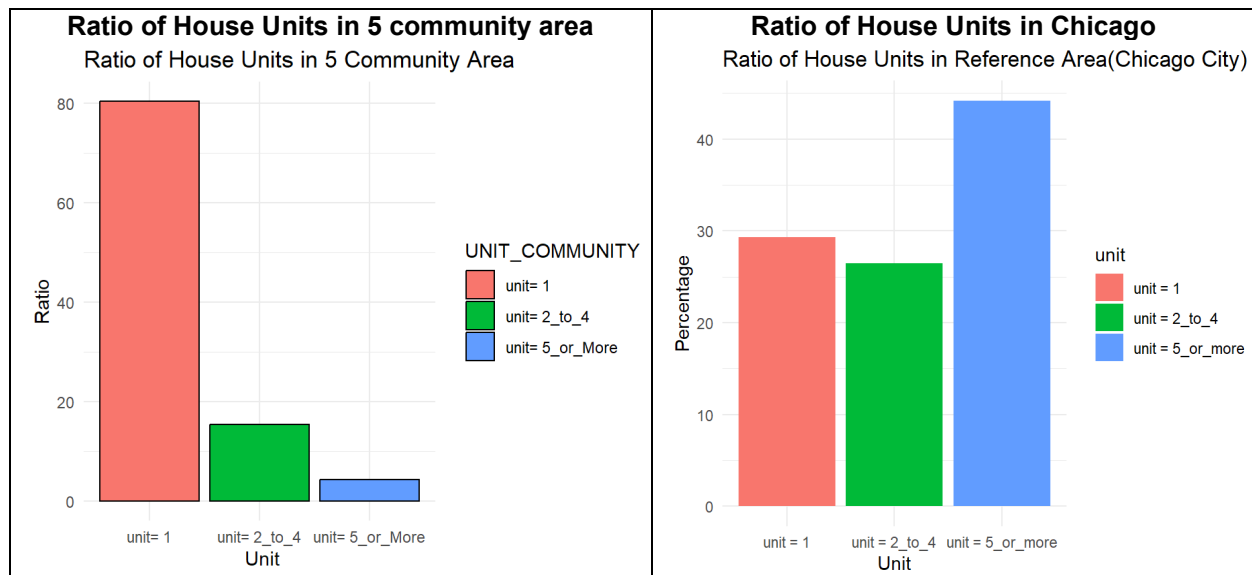
'5 community areas' was **80%** when the unit = 1, 16% when the unit was 2 to 4, and 4% when the unit was 5 or more.

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.

## Appendix 5. The Map of Community Areas in City of Chicago



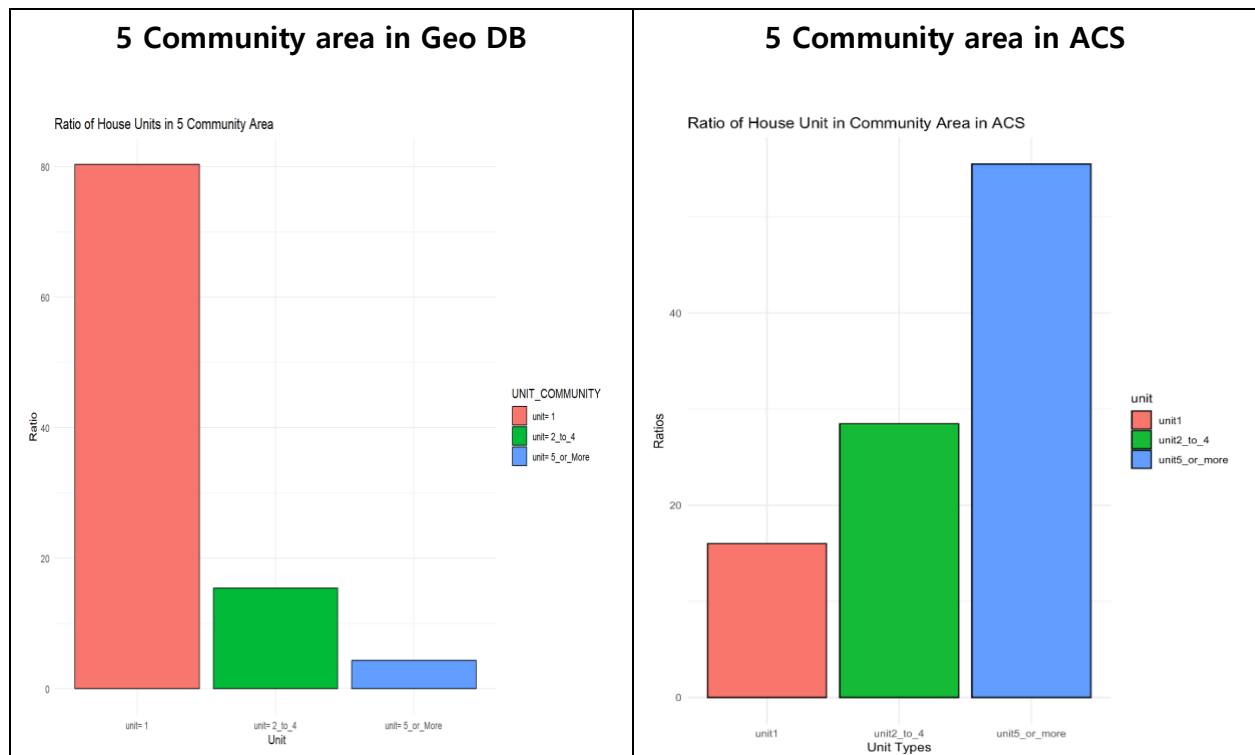
According to A map of the City of Chicago, Chicago is divided into 77 community areas. And We found **5 Community areas are all located on the far north side of Chicago, within 9 miles (14 km) from the Loop.** Loop is the main section of Downtown Chicago.

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 population density, distance from the center and economic significance.**

## Appendix 6. Additional Comparison between Geo file and ACS data



We have confirmed the way to filter community areas from ACS data, so we conducted additional analysis.

We also utilized a **Units in Structure** data, filtering for the zip codes that belong to 5 community area (ZCTA5 60613, 60618, 60625, 60626, 60640, 60645, 60659).

Subsequently, we observed significant disparities between the two graphs, particularly in a proportional context. While **the original data exhibited the highest proportion for 1 unit, the ACS data indicated a higher proportion for 2 or more units.**

Several potential causes for this discrepancy may have been considered:

1. The original data being a subset of actual data (ACS)
2. The original data not accurately reflecting specific characteristics of the actual data
3. The original data assuming a different distribution than the actual data, or the presence of errors or biases in the original data.

## [Weekly Report]

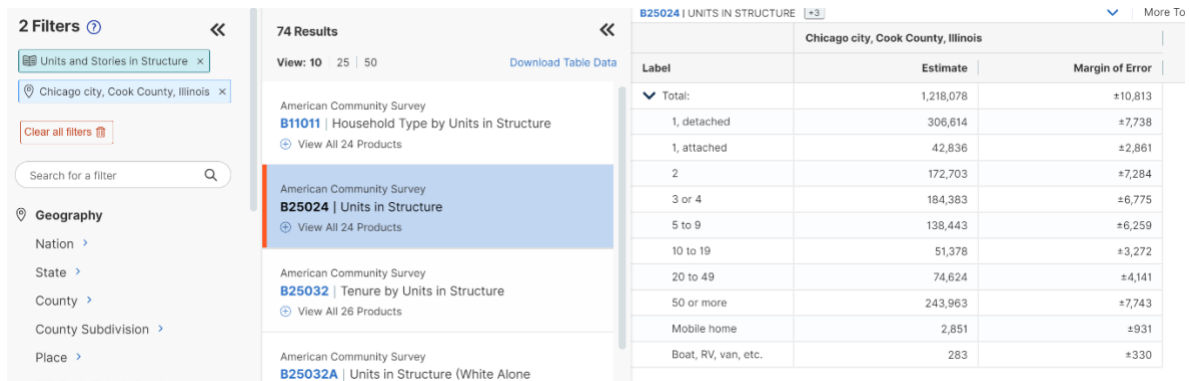
- Report No.4
- Date: Nov. 29, 2023~ Dec. 5, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

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

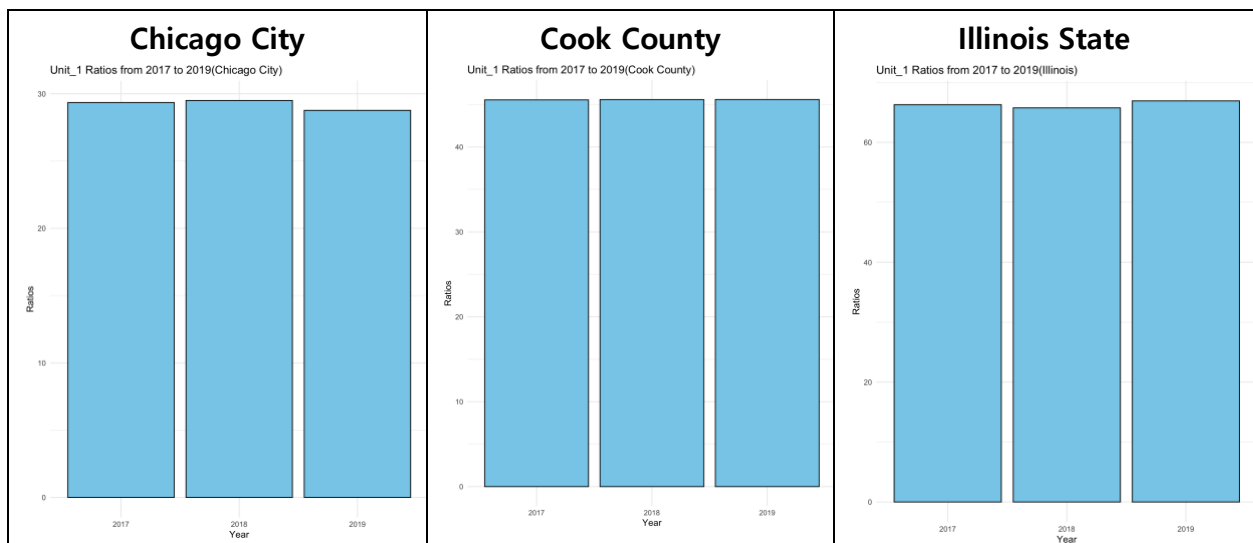


Reference Area: Chicago city, Cook County, Illinois,

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.

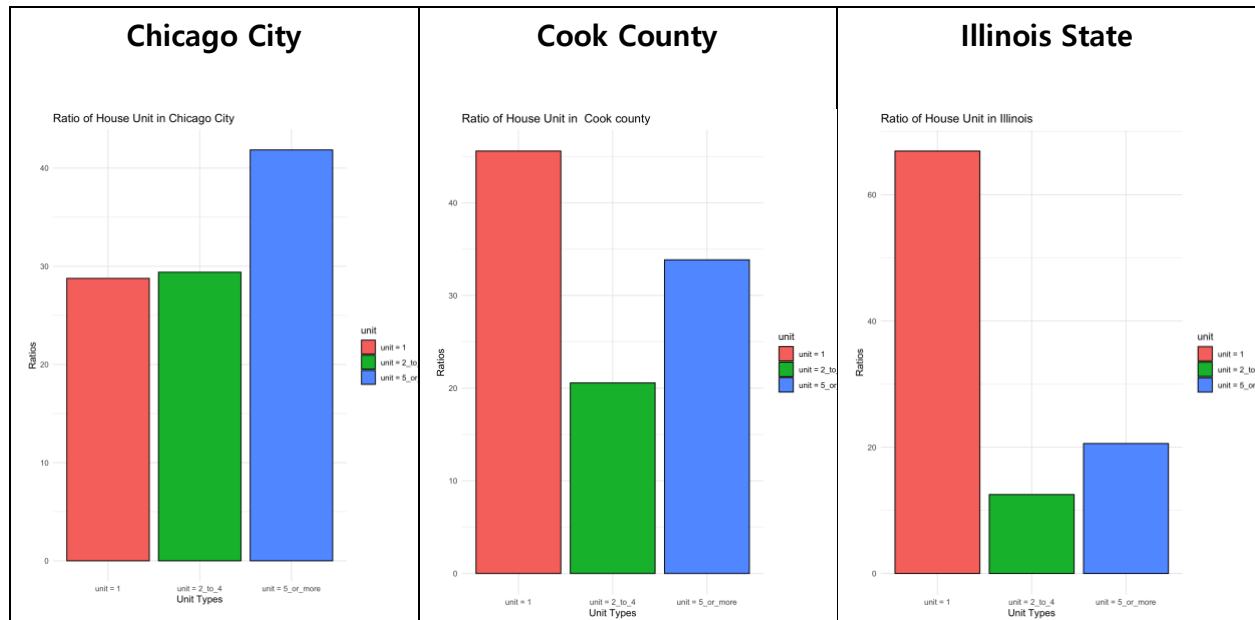
## Appendix 2. Changes in the total number of households over time



: 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.

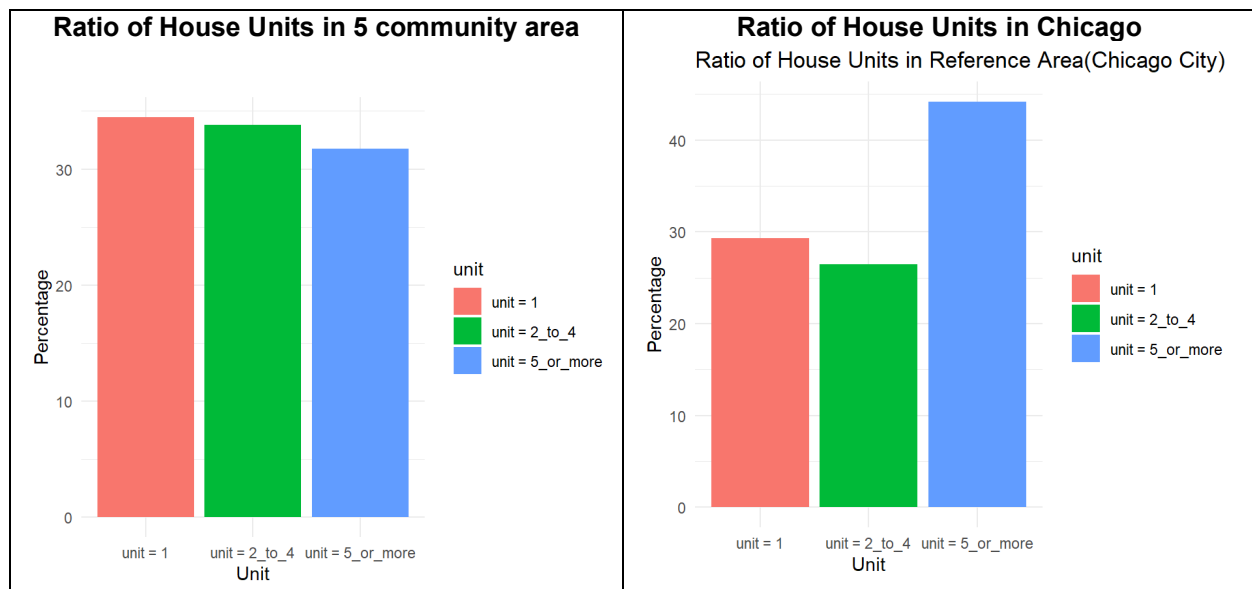
'**Illinois**' was **66%** when the unit = 1, 12% when the unit was 2 to 4, and 20% when the unit was 5 or more.

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.**

## [Weekly Report]

- Report No.3
- Date: Nov. 21, 2023~ Nov. 28, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		Highlights	Self Evaluation
Last Week	Summary	<ul style="list-style-type: none"> <li>• Reconduct data cleansing and manipulation each file.</li> <li>• Determine the reference area(s) to be compared in ACS</li> </ul>	Medium
Current Week	Baseline Goals (Given)	<ul style="list-style-type: none"> <li>• Joining file (geo1, geo2, geo3) and creating integrated DB for analysis</li> <li>• Create an integrated database to balanced panel</li> </ul>	Medium
	Additional Goals (O.Y.O)	<ul style="list-style-type: none"> <li>• To create balanced panel data that is consist of three distinct observations</li> </ul>	Medium
	Key Issues to Be Resolved	<ul style="list-style-type: none"> <li>• We need to be eliminating unnecessary duplicated observations in each file.</li> <li>• We must determine a relevant joining strategy such as inner join, left join, right join and full join.</li> <li>• We should closely examine the structure of data; the number of observations includes 207,225.</li> </ul>	Medium
	Strategies	<ul style="list-style-type: none"> <li>• Step 1: cleanse duplicated data in a comprehensive manner.</li> <li>• Step 2: we would do inner join between geo1 and geo2.</li> <li>• Step 3: After, we would do inner join again between geo_temp and geo3.</li> <li>• Step 4: in terms of "PROPERTY_TYPE_YR", we would manipulate unbalanced panel to balanced panel data</li> </ul>	Medium
	Results	<ul style="list-style-type: none"> <li>• After first cleansing, geo1 has 77101obs., geo2 has 192138 obs. and geo3 has 640068obs.</li> <li>• Conducting an inner join between geo1 and geo2, geo_temp has 69768obs. of 5 variables.</li> <li>• Conducting an inner join between geo3 and geo_temp, geo_master has 225032obs. of 11 variables.</li> <li>• Grouping by single pin and remaining pin that has value of 2017, 2018, 2019 each. So, we finally have 207225 obs. in geo_final, type of year has 69065 obs. each.</li> <li>• Finally, we add a new variable LONGITUDE_NBR, LATITUDE_NBR where the value is a random sample of one longitude from the existing LONGITUDE_NBR, LATITUDE_NBR values</li> </ul>	Medium
	Implications	<ul style="list-style-type: none"> <li>• It helps us prepare the data for further analysis or modeling</li> </ul>	Medium
Next Week	Things to Do	<ul style="list-style-type: none"> <li>• Based upon those queries, we will extract data from the ACS.</li> <li>• Select Model for Analysis</li> </ul>	Medium
Remarks	Core Libraries & Packages	<ul style="list-style-type: none"> <li>• Data Cleansing: dplyr</li> <li>• Grouping: dplyr</li> <li>• Visualization: ggplot2</li> </ul>	Medium
	Additional Remarks		

- Remarks
- Work Progress

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

## Appendix 1. Results of Joining strategies

Obs.		Inner join	Inner join
geo1	77,101	-	-
geo2	192,138	-	-
geo_temp	-	69,768	-
geo3	640,068	640,068	-
geo_master			225,032

-> We decided that it was difficult to fill the missing values, so we conducted inner join.

## Appendix 2. Make every single pin should consist of 3 distinct observations in terms of "PROPERTY\_TYPE\_YR"

Year	geo_master	geo_final
2017	74,700	69,075
2018	74,935	69,075
2019	75,397	69,075
Total	225,032	207,225

The goal is to create a balanced panel dataset with unique observations for each PIN3 for the years 2017, 2018, and 2019.

We use group by(PIN3) and filter to retain only those observations where all three years (2017, 2018, 2019) are present for each unique PIN3.

The resulting dataset is named **geo\_final**.

Finally, we extracted 69075 obs. each, with a total of **207225 obs.**

### Appendix 3. Integrated table of final DB

	PIN3	ASSESSOR_FINAL_ADDR_LN	ASSESSOR_CITY_NM	RESIDENTIAL_IND	PROPERTY_TYPE_YR	LONGITUDE_NBR	LATITUDE_NBR	COMMUNITY_NAME	COMMUNITY_ID	HOUSE_UNIT	PROPERTY_TYPE
1	9.364191e+12	6453 N NORTHWEST HWY	CHICAGO	Y	2017	-87.69490	41.97000	Lincoln Square	4	1	2
2	9.364191e+12	6453 N NORTHWEST HWY	CHICAGO	Y	2018	-87.69490	41.97000	Lincoln Square	4	1	2
3	9.364191e+12	6453 N NORTHWEST HWY	CHICAGO	Y	2019	-87.69490	41.97000	Lincoln Square	4	1	2
4	9.364191e+12	6441 N NORTHWEST HWY	CHICAGO	Y	2017	-87.70000	41.97930	Lincoln Square	4	1	2
5	9.364191e+12	6441 N NORTHWEST HWY	CHICAGO	Y	2018	-87.70000	41.97930	Lincoln Square	4	1	2
6	9.364191e+12	6441 N NORTHWEST HWY	CHICAGO	Y	2019	-87.70000	41.97930	Lincoln Square	4	1	2
7	9.364251e+12	6490 N NORTHWEST HWY	CHICAGO	Y	2017	-87.70210	41.99480	West Ridge	2	1	2
8	9.364251e+12	6490 N NORTHWEST HWY	CHICAGO	Y	2018	-87.70210	41.99480	West Ridge	2	1	2
9	9.364251e+12	6490 N NORTHWEST HWY	CHICAGO	Y	2019	-87.70210	41.99480	West Ridge	2	1	2
10	9.364251e+12	6460 N NORTHWEST HWY	CHICAGO	Y	2017	-87.70810	41.99140	West Ridge	2	1	2
11	9.364251e+12	6460 N NORTHWEST HWY	CHICAGO	Y	2018	-87.70810	41.99140	West Ridge	2	1	2
12	9.364251e+12	6460 N NORTHWEST HWY	CHICAGO	Y	2019	-87.70810	41.99140	West Ridge	2	1	2
13	1.025300e+13	3129 W HOWARD ST	CHICAGO	Y	2017	-87.70780	42.01890	West Ridge	2	1	1
14	1.025300e+13	3129 W HOWARD ST	CHICAGO	Y	2018	-87.70780	42.01890	West Ridge	2	1	1
15	1.025300e+13	3129 W HOWARD ST	CHICAGO	Y	2019	-87.70780	42.01890	West Ridge	2	1	1
16	1.025300e+13	3127 W HOWARD ST	CHICAGO	Y	2017	-87.70780	42.01890	West Ridge	2	1	1
17	1.025300e+13	3127 W HOWARD ST	CHICAGO	Y	2018	-87.70780	42.01890	West Ridge	2	1	1
18	1.025300e+13	3127 W HOWARD ST	CHICAGO	Y	2019	-87.70780	42.01890	West Ridge	2	1	1
19	1.025300e+13	3111 W HOWARD ST	CHICAGO	Y	2017	-87.70720	42.01890	West Ridge	2	1	1
20	1.025300e+13	3111 W HOWARD ST	CHICAGO	Y	2018	-87.70720	42.01890	West Ridge	2	1	1
21	1.025300e+13	3111 W HOWARD ST	CHICAGO	Y	2019	-87.70720	42.01890	West Ridge	2	1	1
22	1.025300e+13	3105 W HOWARD ST	CHICAGO	Y	2017	-87.70690	42.01890	West Ridge	2	1	1
23	1.025300e+13	3105 W HOWARD ST	CHICAGO	Y	2018	-87.70690	42.01890	West Ridge	2	1	1
24	1.025300e+13	3105 W HOWARD ST	CHICAGO	Y	2019	-87.70690	42.01890	West Ridge	2	1	1

Showing 1 to 23 of 207,225 entries, 11 total columns

Finally, we adjust variable LONGITUDE\_NBR, LATITUDE\_NBR where the value is a random sample of one longitude from the existing LONGITUDE\_NBR, LATITUDE\_NBR values.

-> Table shows 1 to 23 of 207,225 entries, 11 total columns

## [Weekly Report]

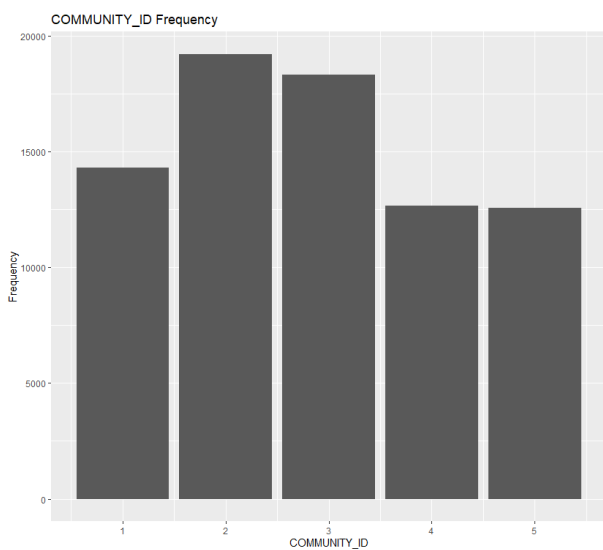
- Report No.2 (updated week1)
- Date: Nov. 08, 2023~ Nov. 14, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		Highlights	Self Evaluation
Last Week	Summary	<ul style="list-style-type: none"> <li>• Conducting Exploratory Data Analysis to examine key features of each file</li> </ul>	Medium
Current Week	Baseline Goals (Given)	<ul style="list-style-type: none"> <li>• Reconduct data cleansing and manipulation each file.</li> <li>• Determine the reference area(s) to be compared in ACS</li> </ul>	Medium
	Additional Goals (O.Y.O)	<ul style="list-style-type: none"> <li>• Checking the logical relationships between variables</li> </ul>	Medium
	Key Issues to Be Resolved	<ul style="list-style-type: none"> <li>• Several duplicate data included in geo1 and geo3 datasets.</li> <li>• Some missing values included in geo3's columns.</li> <li>• Outliers included in geo3.</li> <li>• <u>Detecting Logical Violation between indicator and property type</u></li> </ul>	Medium
	Strategies	<ul style="list-style-type: none"> <li>• Step 1: Reconduct EDA to examine key characteristic of variables and to remove outliers, duplicate and Missing values.</li> <li>• Step 2: Identify logical violations and logical relationship.</li> <li>• Step 3: Select Reference area to create a list of queries in ACS</li> </ul>	Medium
	Results	<ul style="list-style-type: none"> <li>• In 2017 and 2018, there were logical violations of indicator in 566 and 437 obs each. This is presumed to be due to a change in housing type. On the contrary, in 2019, all rows are matched, suggesting that the measurement point for geo2 is in 2019.</li> <li>• Remove Outlier "Unknown" values in geo3.</li> <li>• Merge duplicate values and delete missing values (geo1: 153021 obs -&gt; 77101 obs, geo3: 1210596 obs -&gt; 638270 obs)</li> <li>• Set Reference area to Chicago city, Cook County, Illinois</li> <li>• Create list of filtered data in ACS</li> <li>• Comparing housing characteristics of 5 communities in Chicago and the reference area, the proportion of multi-unit houses in the reference area is relatively higher.</li> </ul>	Medium
	Implications	<ul style="list-style-type: none"> <li>• It helps us prepare the data for further analysis or modeling.</li> </ul>	Medium
Next Week	Things to Do	<ul style="list-style-type: none"> <li>• Based upon those queries, we will extract data from the ACS.</li> <li>• Select Model for Analysis</li> </ul>	Medium
Remarks	Core Libraries & Packages	<ul style="list-style-type: none"> <li>• Data Cleansing: dplyr</li> <li>• Visualization: ggplot2</li> </ul>	Medium
	Additional Remarks		

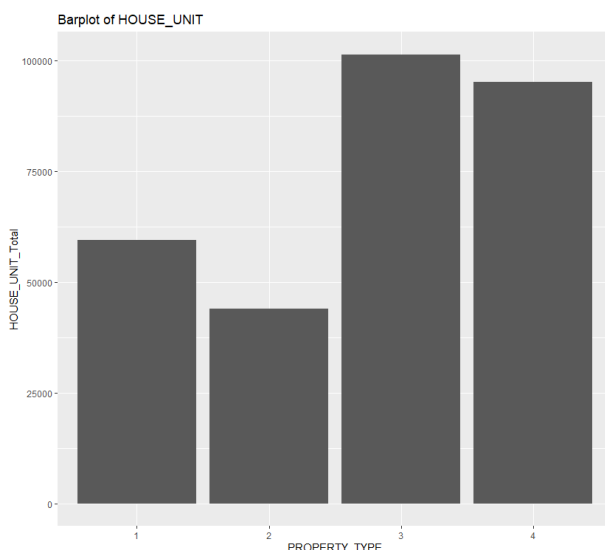
- Remarks
- Visualizations and Results are given in following Appendix.
- Work Progress

	Week 1	Week 2	Week 3	Week 4	Week 5
Yang Junseob	EDA	Reconduct EDA	-	-	-
Ryu Seung Gwon	EDA	Reconduct EDA	-	-	-

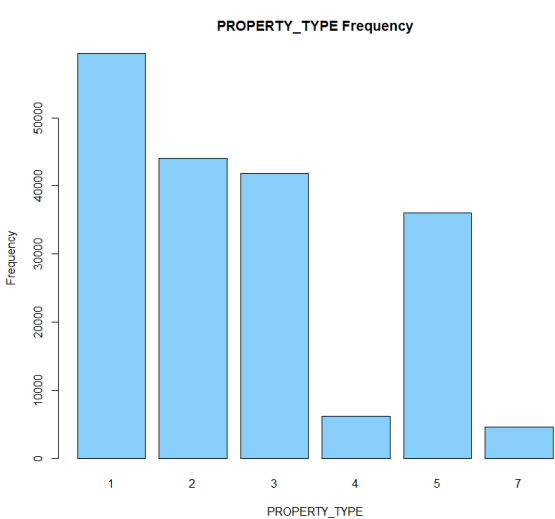
**Appendix1.**  
**Frequency of Community ID**



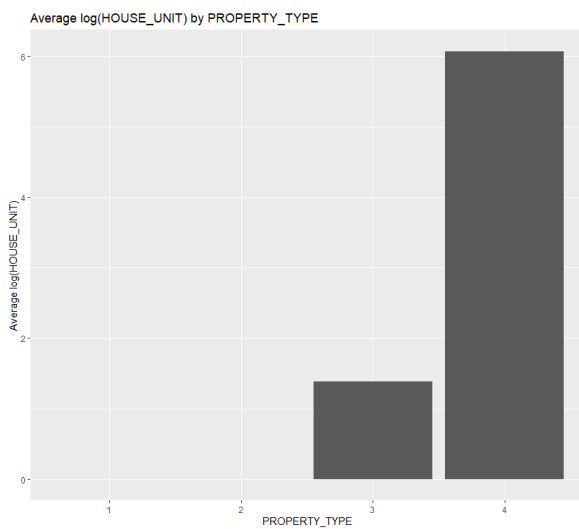
**Appendix2.**  
**Total House unit by Property type of house**



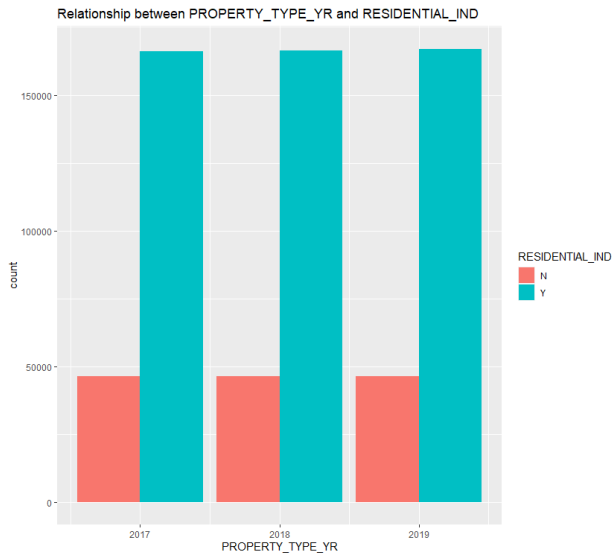
**Appendix3.**  
**Frequency of Property type**



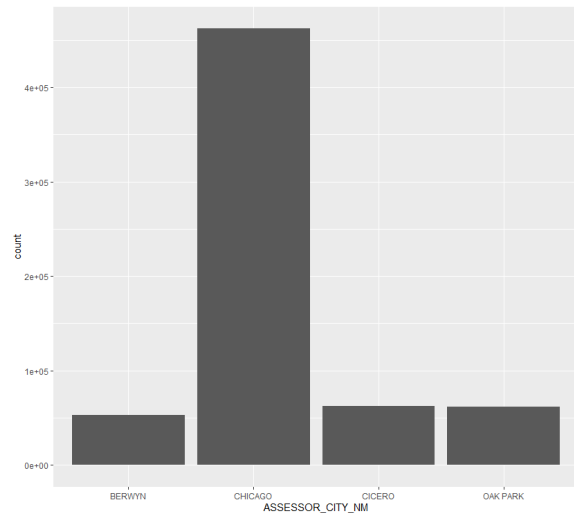
**Appendix4.**  
**Average log units by property type of house**



### Appendix5. Relation between Property type and Residential Id by years



### Appendix6. Distribution of assessor city in “geo3”



### Appendix7. Examine Logical Violations between Indicator and property type

	Pass	Fail (Logical Violation)
<b>2017</b>	211792	566
<b>2018</b>	212112	437
<b>2019</b>	213363	0
<b>Total</b>	<b>637267</b>	<b>1003</b>

- ➔ In 2017 and 2018, there were logical violations of indicator in 566 and 437 obs each. This is presumed to be due to a change in housing type. On the contrary, in 2019, all rows are matched, suggesting that the measurement point for geo2 is in 2019.



## Appendix8. Create the list of queries(filters) for further analysis

2 Filters

Units and Stories in Structure x

Chicago city, Cook County, Illinois x

Clear all filters

Search for a filter

Geography

Nation

State

County

County Subdivision

Place

74 Results

View: 10 25 50

Download Table Data

American Community Survey

B11011 | Household Type by Units in Structure

View All 24 Products

American Community Survey

B25024 | Units in Structure

View All 24 Products

American Community Survey

B25032 | Tenure by Units in Structure

View All 26 Products

American Community Survey

B25032A | Units in Structure (White Alone)

B25024 | UNITS IN STRUCTURE

Chicago city, Cook County, Illinois

Label	Estimate	Margin of Error
Total:	1,218,078	±10,813
1, detached	306,614	±7,738
1, attached	42,836	±2,861
2	172,703	±7,284
3 or 4	184,383	±6,775
5 to 9	138,443	±6,259
10 to 19	51,378	±3,272
20 to 49	74,624	±4,141
50 or more	243,963	±7,743
Mobile home	2,851	±931
Boat, RV, van, etc.	283	±330

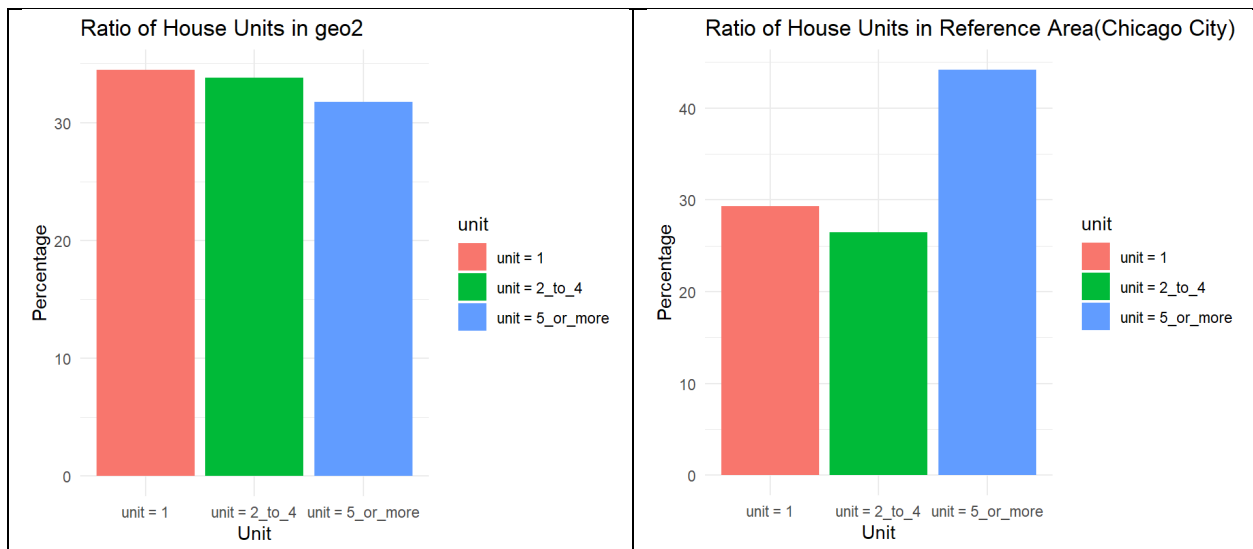
### Reference Area:

Chicago city, Cook County, Illinois,

Units and stories in structure -> B25024 Units in Structure

2019: ACS 1-Year Estimates Detailed Tables

## Appendix9. Closely examine relevant data from ACS to compare housing characteristics of 5 communities in Chicago and the reference area



We compared the community area data with the data for Chicago County based on the criteria of the number of house units: single house units, house units (2 to 4), and house units(five or more).

The results showed that in geo2, 34% of households had single house unit structure, 33% had two to four and 32% had five or more.

In contrast, for Chicago City, the percentages were 29%, 26%, and 44% respectively.

So, the proportion of multi-unit houses in the reference area is relatively higher.

## [Weekly Report]

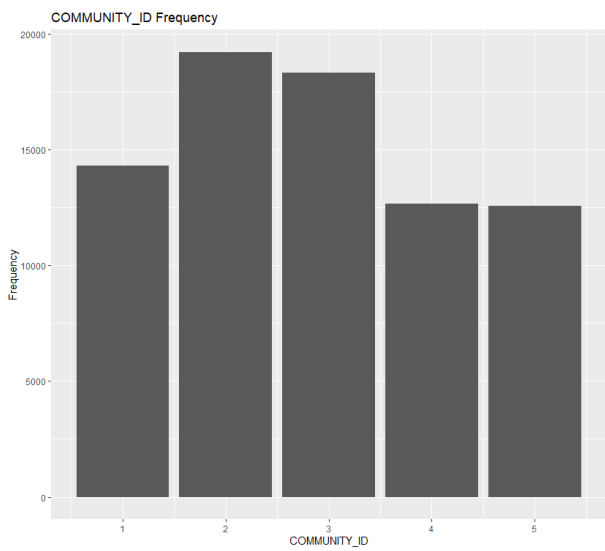
- Report No.1
- Date: Nov. 01, 2023~ Nov. 07, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		Highlights	Self Evaluation
Last Week	Summary	<ul style="list-style-type: none"> <li>• Overview of Team Project</li> <li>• Practicing basic of R</li> </ul>	Medium
Current Week	Baseline Goals (Given)	<ul style="list-style-type: none"> <li>• Importing data in R</li> <li>• Conducting Exploratory Data Analysis to examine key features of each file</li> </ul>	Medium
	Additional Goals (O.Y.O)	<ul style="list-style-type: none"> <li>• Visualize data for understanding the statistical data distributions</li> </ul>	Medium
	Key Issues to Be Resolved	<ul style="list-style-type: none"> <li>• Several duplicate data included in geo1 and geo3 datasets.</li> <li>• Some missing values included in LATITUDE_NBR, LONGITUDE_NBR columns.</li> <li>• Some outliers include in ASSESSOR_CITY_NM</li> <li>• Identify key characteristic of variables</li> </ul>	Medium
	Strategies	<ul style="list-style-type: none"> <li>• Step 1: Conduct EDA for Analytics</li> <li>• Step 2: Perform descriptive statistics and data visualization.</li> <li>• Step 3: Identify any patterns, trends, outliers and cleaning them</li> </ul>	Medium
	Results	<ul style="list-style-type: none"> <li>• Gain insights into our data and identify patterns, relationships, and outliers.</li> </ul>	Medium
	Implications	<ul style="list-style-type: none"> <li>• It helps us prepare the data for further analysis or modeling by next week.</li> </ul>	Medium
Next Week	Things to Do	<ul style="list-style-type: none"> <li>• Data Preparation</li> <li>• Exploring data from American Community Survey</li> <li>• Select Model for Analysis</li> </ul>	Medium
Remarks	Core Libraries & Packages	<ul style="list-style-type: none"> <li>• Data Cleansing: dplyr</li> <li>• Visualization: ggplot2</li> </ul>	Medium
	Additional Remarks		

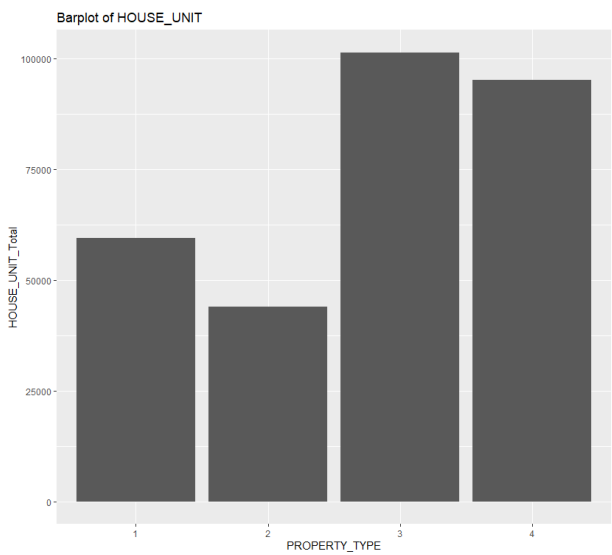
- Remarks
- Visualizations are given in Appendix.
- Work Progress

	Week 1	Week 2	Week 3	Week 4	Week 5
Yang Junseob	EDA	-	-	-	-
Ryu Seung Gwon	EDA	-	-	-	-

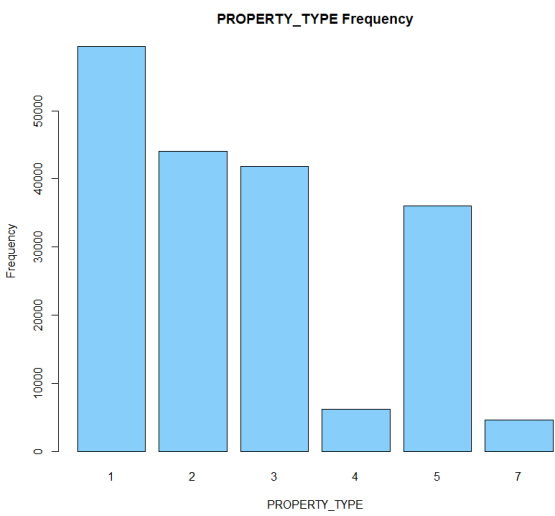
**Appendix1.**  
**Frequency of Community ID**



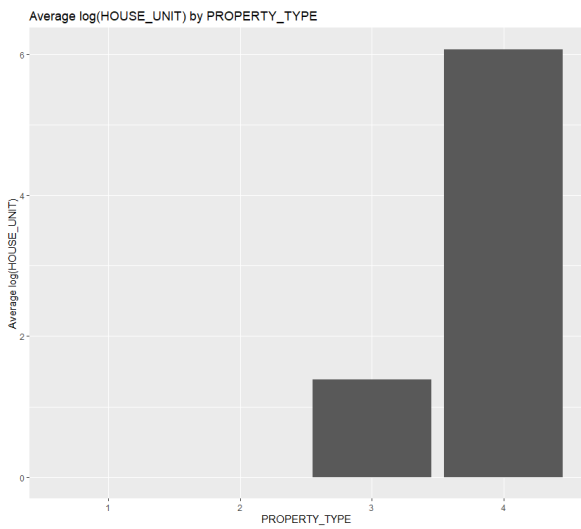
**Appendix2.**  
**Total House unit by Property type of house**



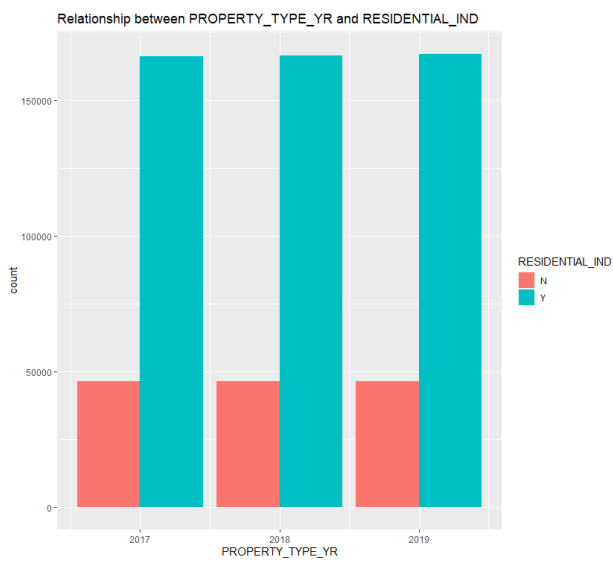
**Appendix3.**  
**Frequency of Property type**



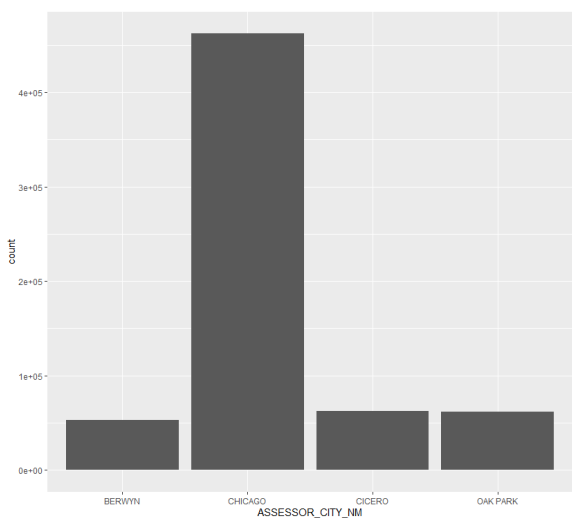
**Appendix4.**  
**Average log units by property type of house**



**Appendix5.**  
**Relation between Property type and Residential Id by years**



**Appendix6.**  
**Distribution of assessor city in “geo3”**



**Appendix7. Missing Values in “geo3”**

