

## [Weekly Report]

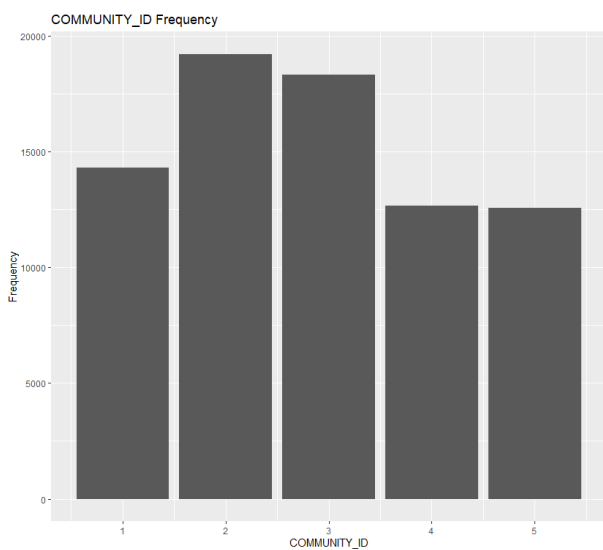
- Report No.2 (updated week1)
- Date: Nov. 08, 2023~ Nov. 14, 2023
- Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

		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>• Detecting Logical Violation between indicator and property type</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		

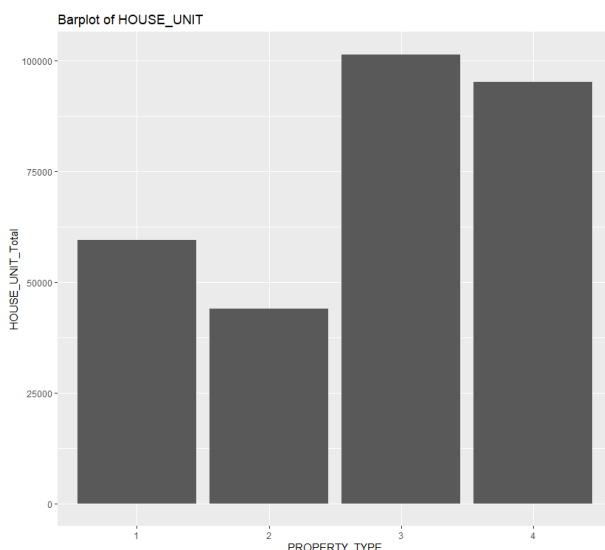
- Remarks
- Visualizations and Results are given in following Appendix.
- Work Progress

	Week 1	Week 2	Week 3	Week 4	Week 5
Yang Junseob	EDA	Reconduct EDA	-	-	-
Ryu Seung Gwon	EDA	Reconduct EDA	-	-	-
	Week 6	Week 7	Week 8	Week 9	Week 10
Yang Junseob	-	-	-	-	Final Report
Ryu Seung Gwon	-	-	-	-	Final Report

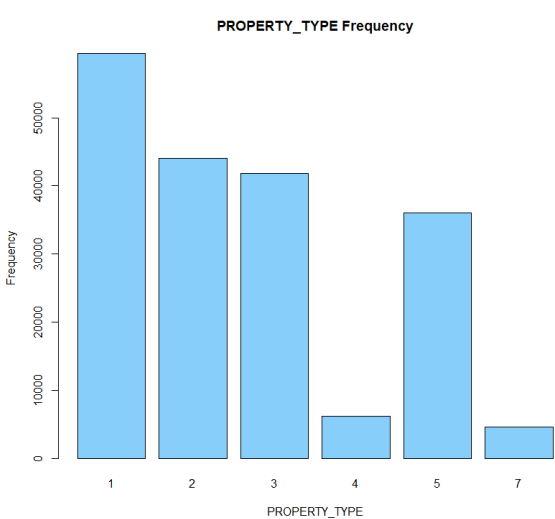
**Appendix1.**  
**Frequency of Community ID**



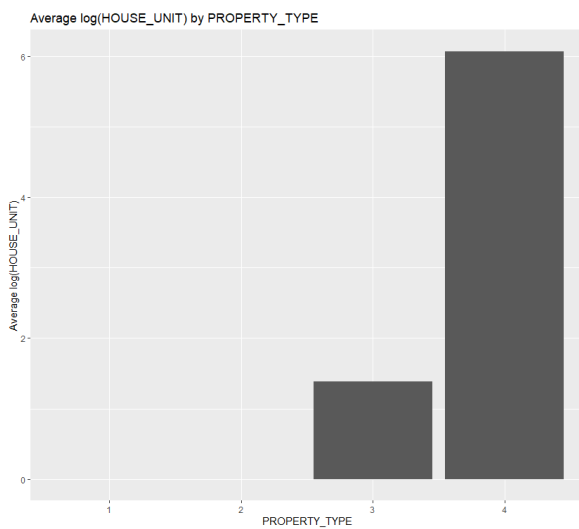
**Appendix2.**  
**Total House unit by Property type of house**



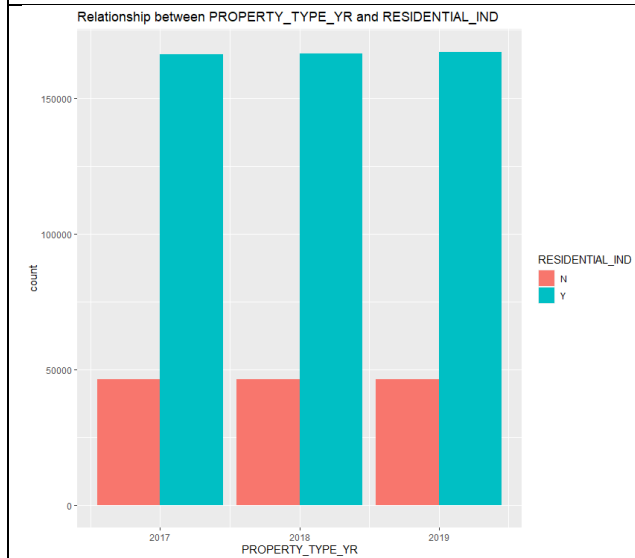
**Appendix3.**  
**Frequency of Property type**



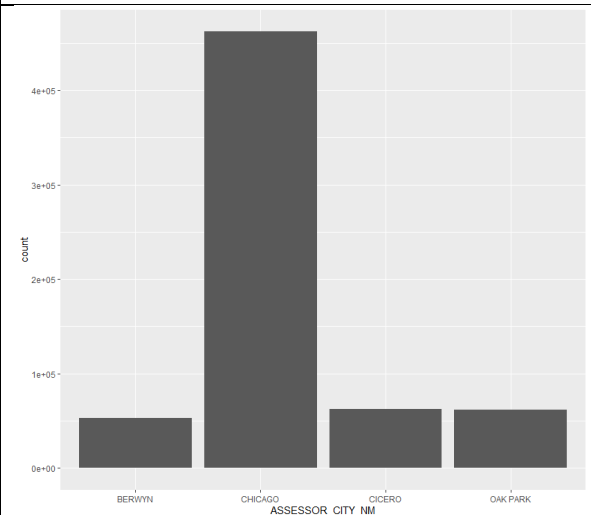
**Appendix4.**  
**Average log units by property type of house**



**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)
2017	211792	566
2018	212112	437
2019	213363	0
Total	637267	1003

- ➔ 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]

The screenshot shows the American Community Survey data interface. On the left, there are 2 filters: 'Units and Stories in Structure' and 'Chicago city, Cook County, Illinois'. The main panel shows 74 results for the query 'B25024 | Units in Structure'. The table displays the following data:

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

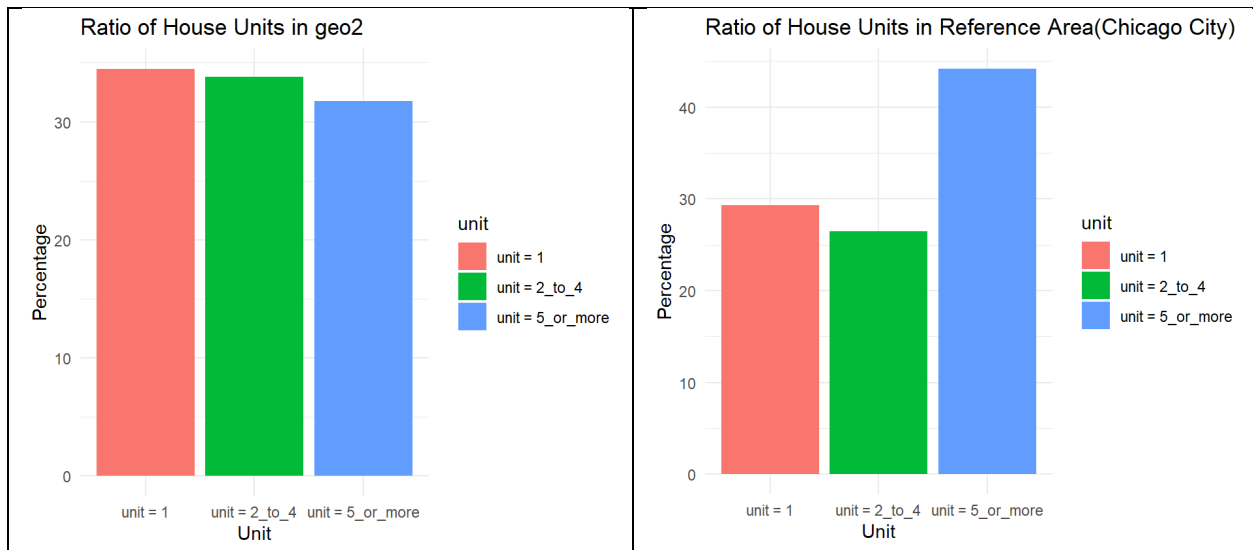
## Reference Area:

**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.