

Using GIS to Map+Visualize Community Areas and Census Data

CS 579 Online Social Network Analysis

Dr. Cindy Hood
10/16/25

Homework Assignments

- ▶ HW #4 - Chicago Community Areas + Census Data
 - ▶ You may work in groups up to 4 students (no exceptions) on this hw
 - ▶ Due date now 10/29
- ▶ Please contact TAs with questions

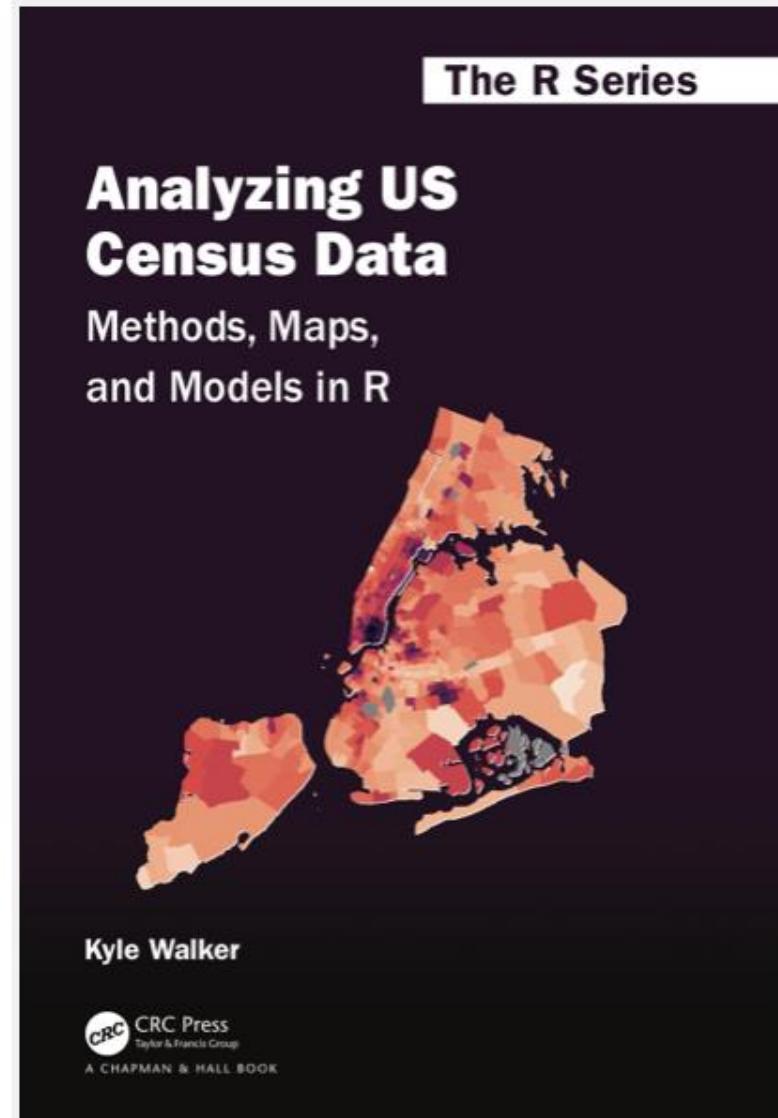
Exams and Final Project Poster Presentation

- ▶ Exam 1 - Have not started grading yet
- ▶ Exam 2 - Dec 2 in class
- ▶ Final Project Poster Session - Dec 4 in class
- ▶ Online students (sections 2 and 3) will have remote options

Final Project

- ▶ Extension of HW 4
- ▶ Teams of up to 4 permitted
- ▶ Proposal due 11/3

Reference

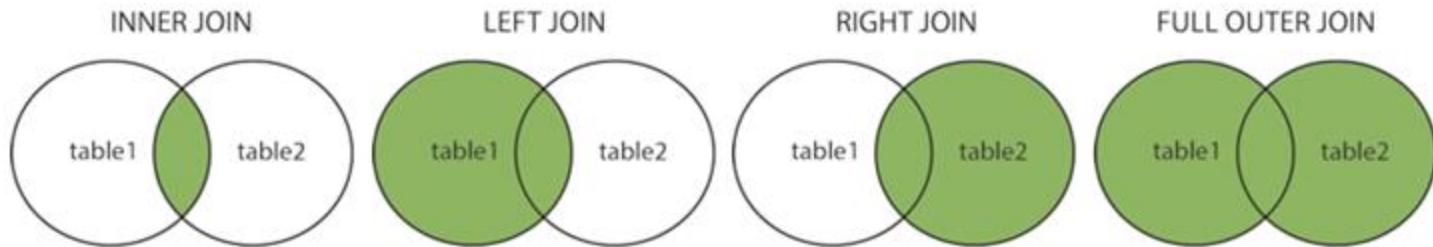


<https://walker-data.com/census-r/>

Set Operations on Data

- ▶ Joining data: Process of combining multiple data sets into one data set by adding additional columns
 - ▶ Sometimes called merging data
- ▶ Concatenating data: Process of combining multiple data sets into one data set by adding additional rows
 - ▶ Sometimes also called “append” or “union”
 - ▶ Treatment of non-overlapping columns can differ by language
- ▶ Certain implementation details vary by term and by language
 - ▶ Handling of duplicate rows
 - ▶ Handling of nulls

Join Operations



Given two tables, A(n by m), B(x by y)

1. Inner join: Return records that have matching values in both tables.
2. Left join: Return all records from left table and matched records from right table.
3. Right join: Return all records from right table and matched records from left table.
4. Full outer join: Return all records where match in either left or right table.
Conceptually equivalent to left join and a right join.

Joining or merging census, and geographic/boundary data

- ▶ Need to find common attribute
 - ▶ Identifier for enumeration unit
- ▶ In our case, we are interested in census data at the block group level
 - ▶ GEOFID down to block group level

Hierarchy of Select Geographic Entities in the American Community Survey



Notes:

* 5-year estimates only

** 5-year estimates only, first release in 2012 for the 2007–2011 5-year estimates
This graphic does not represent the full set of entities for which the ACS publishes data.

This geographic hierarchy influences how the Census Bureau identifies geographic areas.
A system of geocodes - numeric or alphanumeric codes - are used to represent specific geographic areas.

Let's take a look at some ACS data at the block group level - Median Household Income for all the Block Groups in Cook County

	GEOID	NAME	variable	estimate	moe
1	170310101001	Block Group 1; Census Tract 101; Cook County; Illinois	B29004_001	80809	36333
2	170310101002	Block Group 2; Census Tract 101; Cook County; Illinois	B29004_001	26351	9125
3	170310101003	Block Group 3; Census Tract 101; Cook County; Illinois	B29004_001	89545	33643
4	170310102011	Block Group 1; Census Tract 102.01; Cook County; Illi...	B29004_001	78700	23637
5	170310102012	Block Group 2; Census Tract 102.01; Cook County; Illi...	B29004_001	42176	8400
6	170310102013	Block Group 3; Census Tract 102.01; Cook County; Illi...	B29004_001	47569	15883
7	170310102014	Block Group 4; Census Tract 102.01; Cook County; Illi...	B29004_001	NA	NA
8	170310102021	Block Group 1; Census Tract 102.02; Cook County; Illi...	B29004_001	45590	10800
9	170310102022	Block Group 2; Census Tract 102.02; Cook County; Illi...	B29004_001	86419	16794
10	170310103001	Block Group 1; Census Tract 103; Cook County; Illinois	B29004_001	54130	34420
11	170310103002	Block Group 2; Census Tract 103; Cook County; Illinois	B29004_001	74688	27405
12	170310103003	Block Group 3; Census Tract 103; Cook County; Illinois	B29004_001	44335	20324
13	170310103004	Block Group 4; Census Tract 103; Cook County; Illinois	B29004_001	74500	17099
14	170310104001	Block Group 1; Census Tract 104; Cook County; Illinois	B29004_001	63693	27897

Showing 1 to 15 of 4,002 entries, 5 total columns

B29002_001 is the code for Median Household Income

R code

```
med_house_income <- get_acs(  
  geography = "block group",  
  variables = "B29004_001",  
  state = "IL",  
  county = "Cook",  
  year = 2023  
)  
...
```

Getting data from the 2019-2023 5-year ACS
Using FIPS code '17' for state 'IL'
Using FIPS code '031' for 'Cook County'

What do we want to do?

- ▶ We know that we need to look at block groups within community areas
- ▶ We have data that maps census tracts to community areas
- ▶ We know that census tracts are comprised of block groups (per census hierarchy)

Lets take a look at this data

Block Groups in Cook County (2021)

	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON	geometry
2	17	031	834400	1	170318344001	Block Group 1	G5030	S	143065	0	+41.7786780	-087.5982873	MULTIPOLYGON (((-87.60046 4...
3	17	031	834400	3	170318344003	Block Group 3	G5030	S	163630	0	+41.7750339	-087.5985089	MULTIPOLYGON (((-87.60098 4...
4	17	031	843900	2	170318439002	Block Group 2	G5030	S	162794	0	+41.7747748	-087.5898578	MULTIPOLYGON (((-87.59295 4...
5	17	031	421200	1	170314212001	Block Group 1	G5030	S	146529	0	+41.7721656	-087.6100793	MULTIPOLYGON (((-87.61521 4...
6	17	031	491200	1	170314912001	Block Group 1	G5030	S	242880	0	+41.6888126	-087.6344159	MULTIPOLYGON (((-87.63634 4...
7	17	031	491200	2	170314912002	Block Group 2	G5030	S	203417	0	+41.6869377	-087.6391861	MULTIPOLYGON (((-87.64227 4...
8	17	031	491200	3	170314912003	Block Group 3	G5030	S	203195	0	+41.6905739	-087.6393012	MULTIPOLYGON (((-87.64238 4...
9	17	031	491300	2	170314913002	Block Group 2	G5030	S	161278	0	+41.6871079	-087.6253422	MULTIPOLYGON (((-87.6278 41...
10	17	031	491300	3	170314913003	Block Group 3	G5030	S	161466	0	+41.6870516	-087.6301449	MULTIPOLYGON (((-87.63261 4...
11	17	031	491400	1	170314914001	Block Group 1	G5030	S	159580	0	+41.6898309	-087.6118556	MULTIPOLYGON (((-87.61338 4...
17	17	031	500100	1	170315001001	Block Group 1	G5030	S	202423	0	+41.7204569	-087.5951841	MULTIPOLYGON (((-87.59843 4...
18	17	031	500100	2	170315001002	Block Group 2	G5030	S	314179	0	+41.7160331	-087.5965105	MULTIPOLYGON (((-87.60503 4...
19	17	031	500100	3	170315001003	Block Group 3	G5030	S	384488	0	+41.7182562	-087.6011797	MULTIPOLYGON (((-87.60482 4...
20	17	031	580200	1	170315802001	Block Group 1	G5030	S	502730	0	+41.8279305	-087.6988037	MULTIPOLYGON (((-87.70465 4...

Showing 1 to 14 of 4,002 entries, 13 total columns

Note that there are 4002 Block Groups in Cook County, this is consistent with the ACS data on previous slide

Block Groups in Cook County (2023)

STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	GEOIDFQ	NAMESAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON	geometry
17	031	400500	2	170314005002	1500000US170314005002	Block Group 2	G5030	S	136542	0	+41.7890538	-087.6228061	MULTIPOLYGON (((-8...
17	031	480400	6	170314804006	1500000US170314804006	Block Group 6	G5030	S	229016	0	+41.7306299	-087.5787101	MULTIPOLYGON (((-8...
17	031	480400	7	170314804007	1500000US170314804007	Block Group 7	G5030	S	232476	0	+41.7324658	-087.5788122	MULTIPOLYGON (((-8...
17	031	060800	4	170310608004	1500000US170310608004	Block Group 4	G5030	S	45144	0	+41.9530301	-087.6469080	MULTIPOLYGON (((-8...
17	031	290900	1	170312909001	1500000US170312909001	Block Group 1	G5030	S	164717	0	+41.8643125	-087.7275607	MULTIPOLYGON (((-8...
17	031	290900	2	170312909002	1500000US170312909002	Block Group 2	G5030	S	164694	0	+41.8606607	-087.7274518	MULTIPOLYGON (((-8...
17	031	290900	3	170312909003	1500000US170312909003	Block Group 3	G5030	S	247352	0	+41.8605723	-087.7335667	MULTIPOLYGON (((-8...
17	031	290900	4	170312909004	1500000US170312909004	Block Group 4	G5030	S	419733	0	+41.8634334	-087.7356877	MULTIPOLYGON (((-8...
17	031	838700	2	170318387002	1500000US170318387002	Block Group 2	G5030	S	163809	0	+41.8607151	-087.7225793	MULTIPOLYGON (((-8...
17	031	310200	1	170313102001	1500000US170313102001	Block Group 1	G5030	S	193745	0	+41.8560545	-087.6432837	MULTIPOLYGON (((-8...
17	031	310200	2	170313102002	1500000US170313102002	Block Group 2	G5030	S	126828	0	+41.8567003	-087.6456227	MULTIPOLYGON (((-8...
17	031	320101	2	170313201012	1500000US170313201012	Block Group 2	G5030	S	209229	0	+41.8829100	-087.6174378	MULTIPOLYGON (((-8...
17	031	320101	1	170313201011	1500000US170313201011	Block Group 1	G5030	S	111022	10299	+41.8860023	-087.6183349	MULTIPOLYGON (((-8...
17	031	010201	4	170310102014	1500000US170310102014	Block Group 4	G5030	S	121296	0	+42.0172836	-087.6790687	MULTIPOLYGON (((-8...

Showing 1 to 14 of 4,002 entries, 14 total columns

Note that there are 4002 Block Groups in Cook County, this is consistent with the ACS data on previous slide

Note that the file changed from 2021 to 2023 (additional column). Be sure to use data files from the same year.

Block Groups in Cook County (2023)

12 digit GEOID is the same identifier used
in the ACS block group data on previous
slide

Can use GEOID to join tables

STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	GEOIDFQ
17	031	400500	2	170314005002	1500000US170314005002
17	031	480400	6	170314804006	1500000US170314804006
17	031	480400	7	170314804007	1500000US170314804007
17	031	060800	4	170310608004	1500000US170310608004
17	031	290900	1	170312909001	1500000US170312909001
17	031	290900	2	170312909002	1500000US170312909002
17	031	290900	3	170312909003	1500000US170312909003
17	031	290900	4	170312909004	1500000US170312909004
17	031	838700	2	170318387002	1500000US170318387002
17	031	310200	1	170313102001	1500000US170313102001
17	031	310200	2	170313102002	1500000US170313102002
17	031	320101	2	170313201012	1500000US170313201012
17	031	320101	1	170313201011	1500000US170313201011
17	031	010201	4	170310102014	1500000US170310102014

Getting only the block groups we want

- ▶ We have the mapping of Census tracts to Community Areas

GEOID20 is the identifier to the tract level and is 11 digits

GEOID20	CA	COMMUNIT_1
17031010100	1	Rogers Park
17031010201	1	Rogers Park
17031010202	1	Rogers Park
17031010300	1	Rogers Park
17031010400	1	Rogers Park
17031010501	1	Rogers Park
17031010502	1	Rogers Park
17031010503	1	Rogers Park
17031010600	1	Rogers Park
17031010701	1	Rogers Park
17031010702	1	Rogers Park
17031830600	1	Rogers Park

Part of table

Getting only the block groups we want

11 digits

GEOID20	CA	COMMUNITY_1
17031010100	1	Rogers Park
17031010201	1	Rogers Park
17031010202	1	Rogers Park
17031010300	1	Rogers Park
17031010400	1	Rogers Park
17031010501	1	Rogers Park
17031010502	1	Rogers Park
17031010503	1	Rogers Park
17031010600	1	Rogers Park
17031010701	1	Rogers Park
17031010702	1	Rogers Park
17031830600	1	Rogers Park

Tract to Community Area Mapping

No common identifier for join

12 digits

STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	GEOIDFQ
17	031	400500	2	170314005002	1500000US170314005002
17	031	480400	6	170314804006	1500000US170314804006
17	031	480400	7	170314804007	1500000US170314804007
17	031	060800	4	170310608004	1500000US170310608004
17	031	290900	1	170312909001	1500000US170312909001
17	031	290900	2	170312909002	1500000US170312909002
17	031	290900	3	170312909003	1500000US170312909003
17	031	290900	4	170312909004	1500000US170312909004
17	031	838700	2	170318387002	1500000US170318387002
17	031	310200	1	170313102001	1500000US170313102001
17	031	310200	2	170313102002	1500000US170313102002
17	031	320101	2	170313201012	1500000US170313201012
17	031	320101	1	170313201011	1500000US170313201011
17	031	010201	4	170310102014	1500000US170310102014

Census block groups in Cook County

Need to add another column

GEOID20	CA	COMMUNIT_1	TRACTCE
17031010100		1 Rogers Park	010100
17031010201		1 Rogers Park	010201
17031010202		1 Rogers Park	010202
17031010300		1 Rogers Park	010300
17031010400		1 Rogers Park	010400
17031010501		1 Rogers Park	010501
17031010502		1 Rogers Park	010502
17031010503		1 Rogers Park	010503
17031010600		1 Rogers Park	010600
17031010701		1 Rogers Park	010701
17031010702		1 Rogers Park	010702
17031830600		1 Rogers Park	830600



STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	GEOIDFQ
17	031	400500	2	170314005002	1500000US170314005002
17	031	480400	6	170314804006	1500000US170314804006
17	031	480400	7	170314804007	1500000US170314804007
17	031	060800	4	170310608004	1500000US170310608004
17	031	290900	1	170312909001	1500000US170312909001
17	031	290900	2	170312909002	1500000US170312909002
17	031	290900	3	170312909003	1500000US170312909003
17	031	290900	4	170312909004	1500000US170312909004
17	031	838700	2	170318387002	1500000US170318387002
17	031	310200	1	170313102001	1500000US170313102001
17	031	310200	2	170313102002	1500000US170313102002
17	031	320101	2	170313201012	1500000US170313201012
17	031	320101	1	170313201011	1500000US170313201011
17	031	010201	4	170310102014	1500000US170310102014

Now can use join operation to select block groups of interest

Left join result

```
allcook_ca_bgs <- left_join(cook_bgs, censustract_CAx1, by="TRACTCE")
```

	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	GEOIDFQ	NAMESAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON	GEOID20	CA	COMMUNIT_1	geometry
14	17	031	010201	4	170310102014	1500000US170310102014	Block Group 4	G5030	S	121296	0	+42.0172836	-087.6790687	17031010201	1	Rogers Park	MULTIPOLYGON (((-87.68049 4...
17	17	031	010201	3	170310102013	1500000US170310102013	Block Group 3	G5030	S	124894	0	+42.0177395	-087.6826021	17031010201	1	Rogers Park	MULTIPOLYGON (((-87.68465 4...
1235	17	031	010300	3	170310103003	1500000US170310103003	Block Group 3	G5030	S	106961	0	+42.0139446	-087.6644089	17031010300	1	Rogers Park	MULTIPOLYGON (((-87.66798 4...
1236	17	031	010702	3	170310107023	1500000US170310107023	Block Group 3	G5030	S	98723	0	+42.0063757	-087.6725873	17031010702	1	Rogers Park	MULTIPOLYGON (((-87.67529 4...
1237	17	031	010701	1	170310107011	1500000US170310107011	Block Group 1	G5030	S	180130	0	+42.0070613	-087.6782605	17031010701	1	Rogers Park	MULTIPOLYGON (((-87.68184 4...
1238	17	031	010201	2	170310102012	1500000US170310102012	Block Group 2	G5030	S	149216	0	+42.0146832	-087.6776570	17031010201	1	Rogers Park	MULTIPOLYGON (((-87.68039 4...
1261	17	031	010100	3	170310101003 170310101003	1500000US170310101003	Block Group 3	G5030	S	126001	0	+42.0206033	-087.6664194	17031010100	1	Rogers Park	MULTIPOLYGON (((-87.66949 4...
1264	17	031	010502	2	170310105022	1500000US170310105022	Block Group 2	G5030	S	90480	0	+42.0055382	-087.6633897	17031010502	1	Rogers Park	MULTIPOLYGON (((-87.66573 4...
1265	17	031	010501	2	170310105012	1500000US170310105012	Block Group 2	G5030	S	74561	0	+42.0113681	-087.6643282	17031010501	1	Rogers Park	MULTIPOLYGON (((-87.66592 4...
1266	17	031	010501	1	170310105011	1500000US170310105011	Block Group 1	G5030	S	45264	0	+42.0094135	-087.6640628	17031010501	1	Rogers Park	MULTIPOLYGON (((-87.66582 4...
1267	17	031	010600	2	170310106002	1500000US170310106002	Block Group 2	G5030	S	96323	0	+42.0115289	-087.6680137	17031010600	1	Rogers Park	MULTIPOLYGON (((-87.670242... 4...
1268	17	031	010600	4	170310106004	1500000US170310106004	Block Group 4	G5030	S	113311	0	+42.0088648	-087.6678896	17031010600	1	Rogers Park	MULTIPOLYGON (((-87.67012 4...
1292	17	031	830600	3	170318306003	1500000US170318306003	Block Group 3	G5030	S	162544	0	+42.0000238	-087.6728759	17031830600	1	Rogers Park	MULTIPOLYGON (((-87.67513 4...
1294	17	031	010400	3	170310104003	1500000US170310104003	Block Group 3	G5030	S	257953	602605	+42.0041688	-087.6588541	17031010400	1	Rogers Park	MULTIPOLYGON (((-87.66125 4...

Showing 1 to 14 of 4,002 entries, 17 total columns

Now we have a table with all the Cook County Block Groups and Chicago Community Areas

Filtering

- ▶ We know that some Cook County Block Groups are in not in Chicago
 - ▶ We are not interested in these block groups
 - ▶ We are only interested in block groups in Chicago (i.e. CCAs)
- ▶ How do we identify the block groups of interest?

AWATER	INTPTLAT	INTPTLON	GEOID20	CA	COMMUNIT_1	geometry
0	+41.9926485	-087.6676589	17031030200	77	Edgewater	<i>MULTIPOLYGON (((-87.67015 4...</i> 
0	+41.9961968	-087.6726162	17031030300	77	Edgewater	<i>MULTIPOLYGON (((-87.67501 4...</i> 
0	+41.9925923	-087.6724540	17031030300	77	Edgewater	<i>MULTIPOLYGON (((-87.6749 41...</i> 
0	+41.9788883	-087.6713526	17031030900	77	Edgewater	<i>MULTIPOLYGON (((-87.67446 4...</i> 
0	+41.9886099	-087.6674622	17031030500	77	Edgewater	<i>MULTIPOLYGON (((-87.67003 4...</i> 
0	+41.9808229	-087.6641732	17031030800	77	Edgewater	<i>MULTIPOLYGON (((-87.66843 4...</i> 
745867	+42.0582040	-087.6718922	NA	NA	NA	<i>MULTIPOLYGON (((-87.67969 4...</i> 
0	+41.5506025	-087.7810803	NA	NA	NA	<i>MULTIPOLYGON (((-87.79279 4...</i> 
357941	+42.0964838	-088.2181811	NA	NA	NA	<i>MULTIPOLYGON (((-88.23835 4...</i> 
0	+41.6404907	-087.9022480	NA	NA	NA	<i>MULTIPOLYGON (((-87.91193 4...</i> 
79209	+41.9987533	-088.2307015	NA	NA	NA	<i>MULTIPOLYGON (((-88.24753 4...</i> 
0	+42.0912240	-087.8277913	NA	NA	NA	<i>MULTIPOLYGON (((-87.84126 4...</i> 
0	+42.0258635	-088.2104226	NA	NA	NA	<i>MULTIPOLYGON (((-88.22346 4...</i> 
0	+42.0491914	-087.8476857	NA	NA	NA	<i>MULTIPOLYGON (((-87.85023 4...</i> 

Showing 2,156 to 2,170 of 4,002 entries, 17 total columns

We can filter out only CCA Block Groups

```
allchi_ca_bgs <- filter (allcook_ca_bgs, !is.na(CA))
```

AWATER	INTPTLAT	INTPTLON	GEOID20	CA	COMMUNIT_1	geometry
0	+41.7890538	-087.6228061	17031400500	40	Washington Park	<i>MULTIPOLYGON (((-87.62556 4...</i>
0	+41.7306299	-087.5787101	17031480400	48	Calumet Heights	<i>MULTIPOLYGON (((-87.58553 4...</i>
0	+41.7324658	-087.5788122	17031480400	48	Calumet Heights	<i>MULTIPOLYGON (((-87.58557 4...</i>
0	+41.9530301	-087.6469080	17031060800	6	Lake View	<i>MULTIPOLYGON (((-87.64794 4...</i>
0	+41.8643125	-087.7275607	17031290900	29	North Lawndale	<i>MULTIPOLYGON (((-87.73006 4...</i>
0	+41.8606607	-087.7274518	17031290900	29	North Lawndale	<i>MULTIPOLYGON (((-87.72995 4...</i>
0	+41.8605723	-087.7335667	17031290900	29	North Lawndale	<i>MULTIPOLYGON (((-87.73729 4...</i>
0	+41.8634334	-087.7356877	17031290900	29	North Lawndale	<i>MULTIPOLYGON (((-87.73993 4...</i>
0	+41.8607151	-087.7225793	17031838700	29	North Lawndale	<i>MULTIPOLYGON (((-87.72506 4...</i>
0	+41.8560545	-087.6432837	17031310200	31	Lower West Side	<i>MULTIPOLYGON (((-87.64539 4...</i>
0	+41.8567003	-087.6456227	17031310200	31	Lower West Side	<i>MULTIPOLYGON (((-87.64671 4...</i>
0	+41.8829100	-087.6174378	17031320101	32	Loop	<i>MULTIPOLYGON (((-87.62086 4...</i>
10299	+41.8860023	-087.6183349	17031320101	32	Loop	<i>MULTIPOLYGON (((-87.62061 4...</i>
0	+42.0172836	-087.6790687	17031010201	1	Rogers Park	<i>MULTIPOLYGON (((-87.68049 4...</i>

Showing 1 to 14 of 2,162 entries, 17 total columns

Now only 2162 rows (i.e. Block Groups in Chicago/CCA)

Now I can join this table with the ACS Household Income Data

```
hhi_cca <- left_join(allchi_ca_bgs, med_house_income, by="GEOID")
```

estimate is the
statistical estimate of
median household
income per block group

moe is the measure of error in the estimate

Showing 1 to 14 of 2,162 entries, 21 total columns

Note that 2162 rows as expected (# block groups in CCA/Chicago)

Now I can join this table with the ACS Household Income Data

Issue - Some data is missing. Why?

Why could data be missing?

- Not enough data collected in that block group to publish estimate + assure anonymity
- Block group has no or small number of households (e.g. industrial area)
- Problems with data
 - low response
 - non-response
 - lost response
- Statistical problems
 - high moe
 - anomalies

What data is available?

- ▶ Can query to see variable for different year ACS surveys

```
acs_2023variables <- load_variables(year = 2023, dataset = "acs5",  
cache=TRUE)
```

```
acs_2022variables <- load_variables(year = 2022, dataset = "acs5",  
cache=TRUE)
```