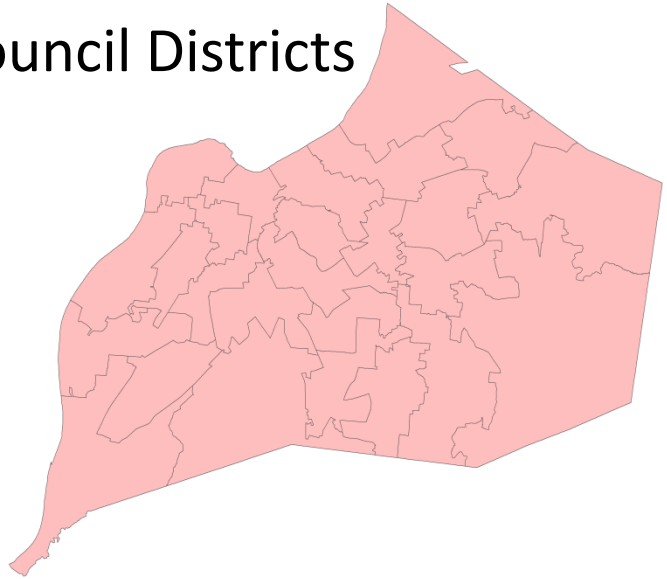
Dasymetrically Refined Areal Weighting

Matt Ruther
University of Louisville
matthew.ruther@louisville.edu

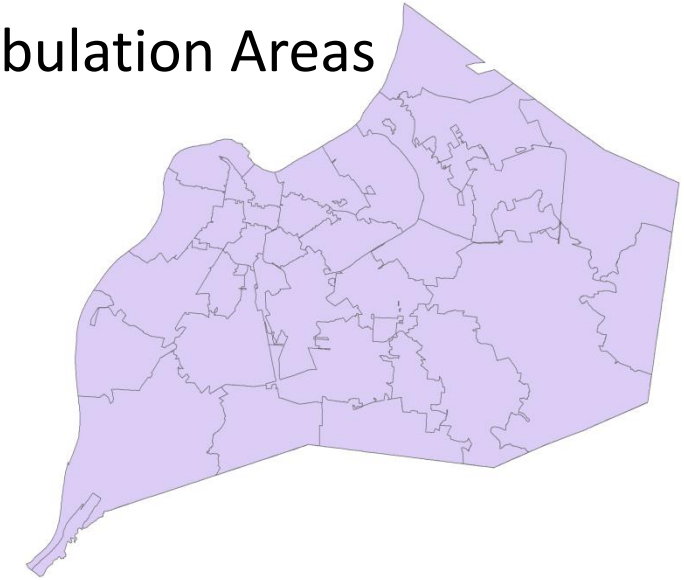
October 24, 2015

Different Zoning Systems

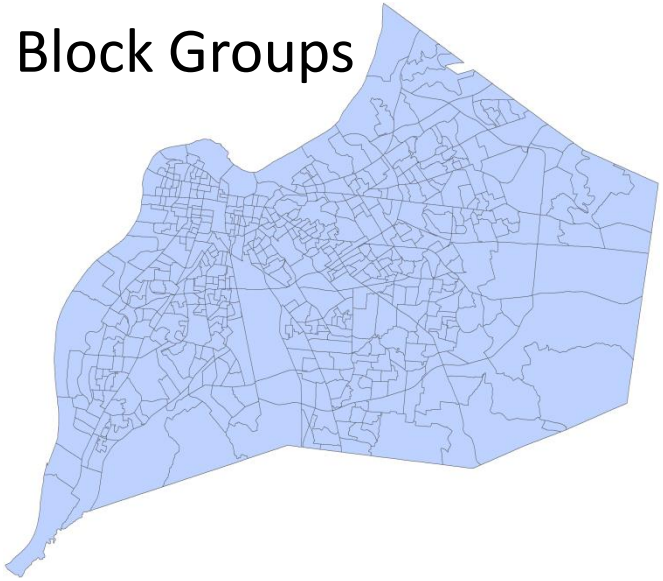
Council Districts



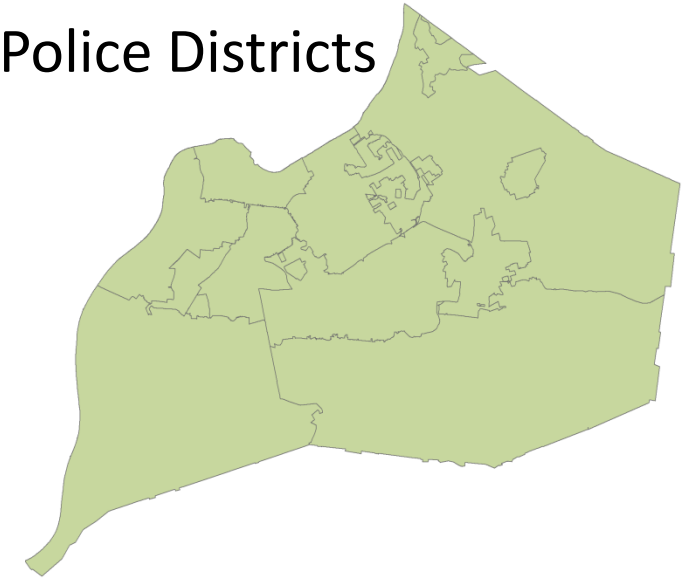
ZIP Tabulation Areas



Census Block Groups

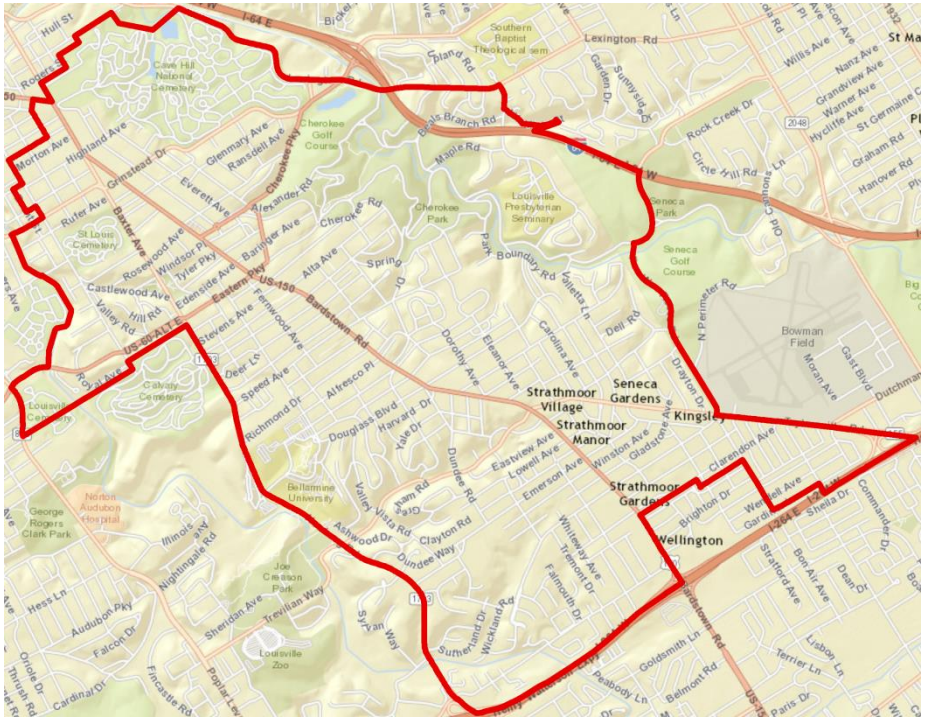


Police Districts



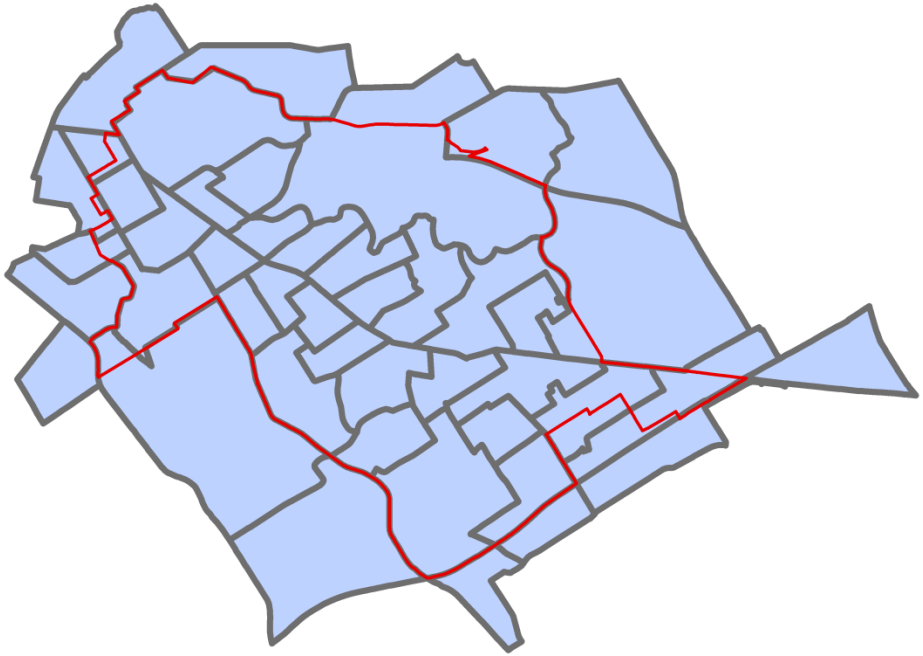
Example

Metro Council District 8

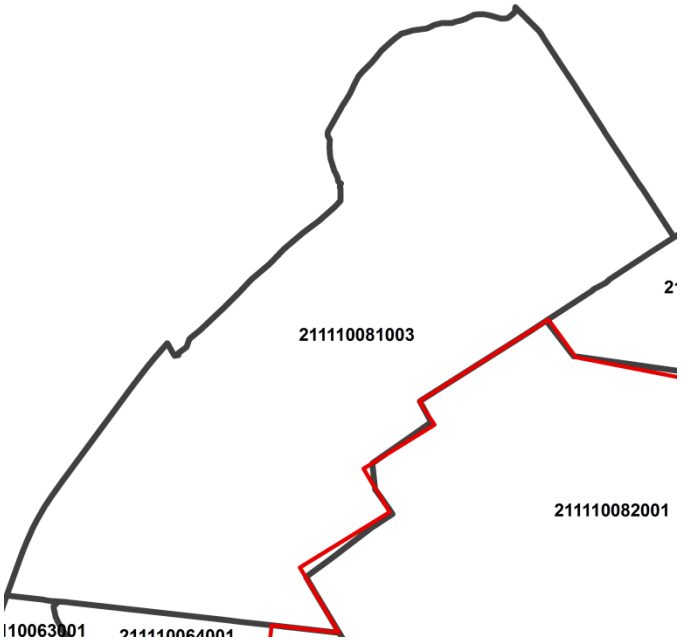


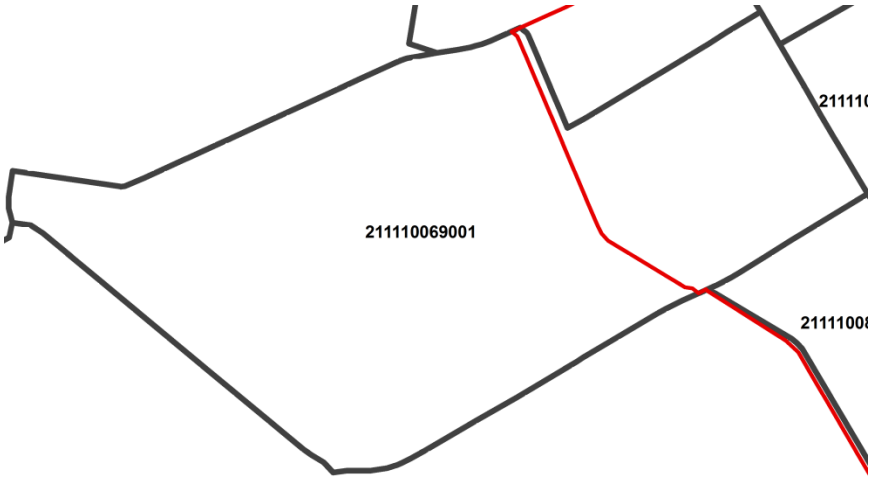
“Target Zone”

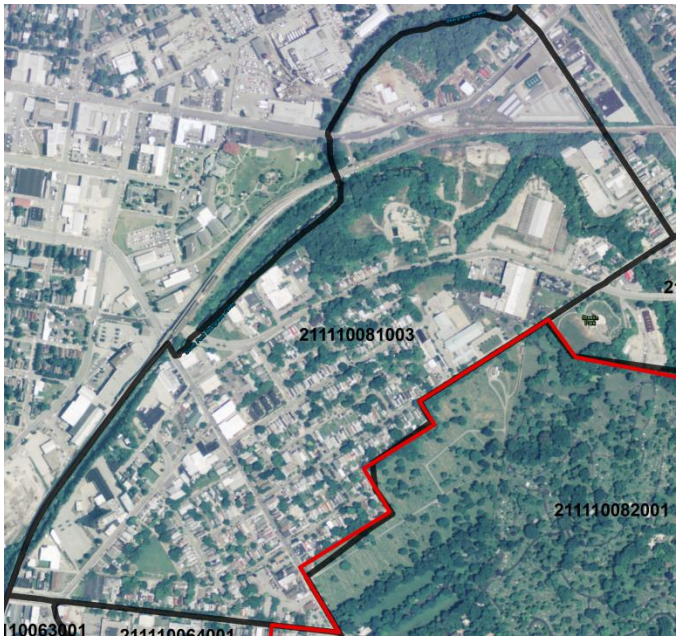
45 Intersecting Block Groups

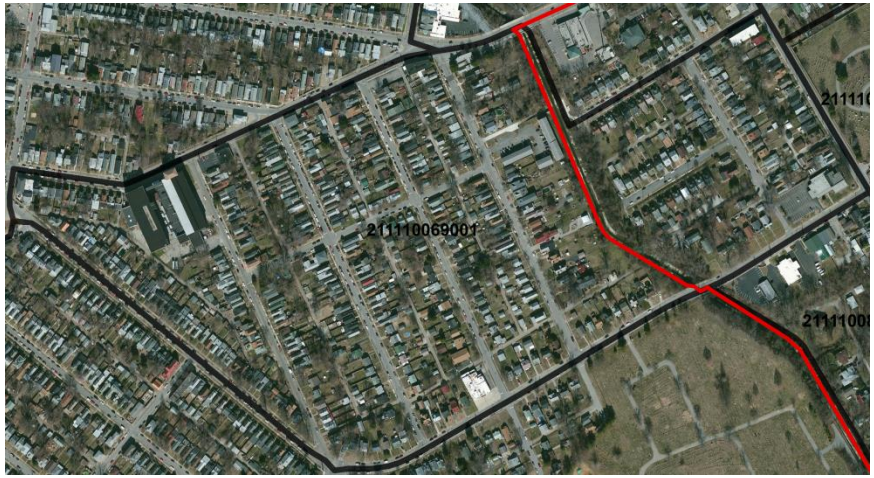


“Source Zones”

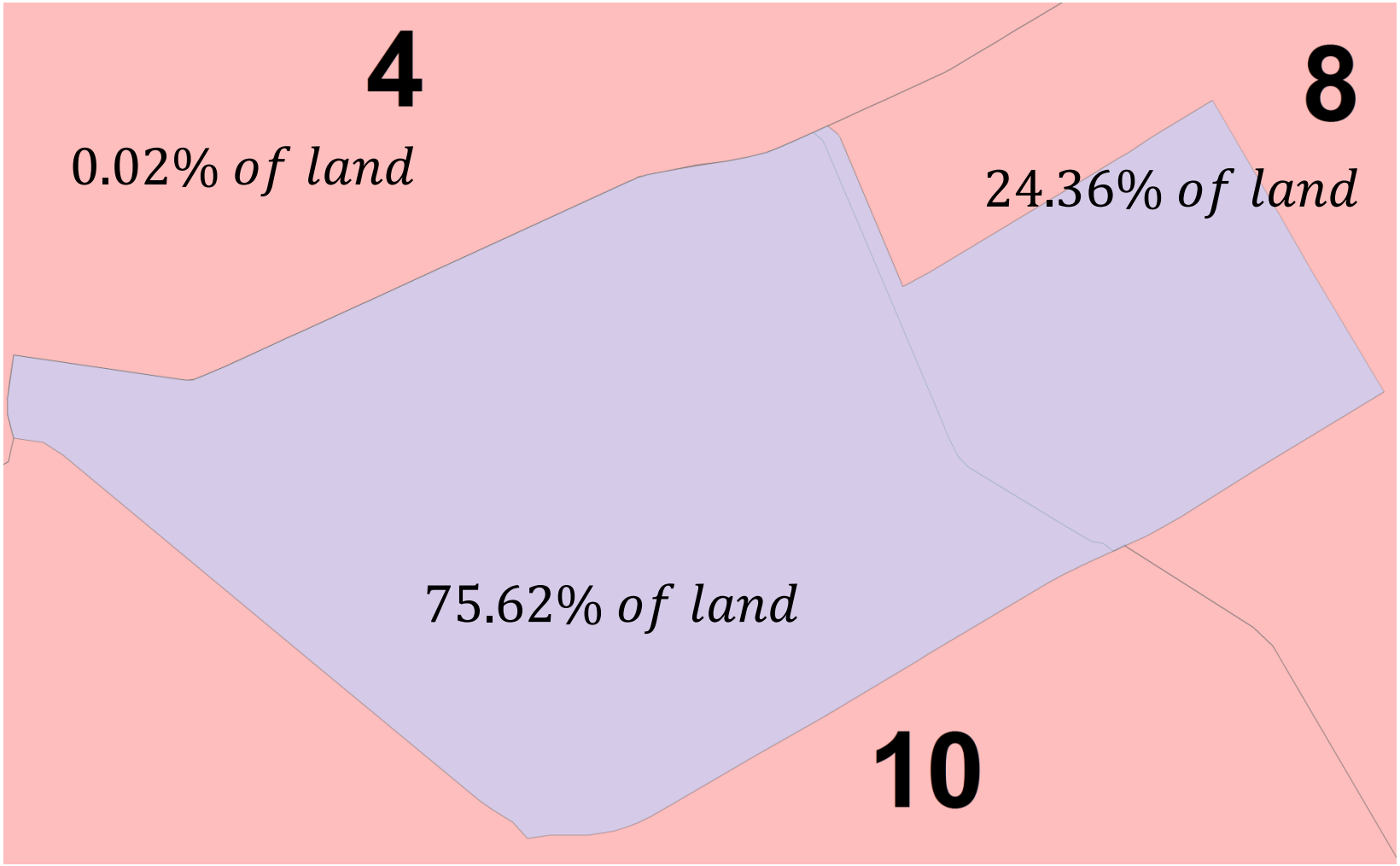

211110081003
211110082001


211110069001


211110081003
211110082001

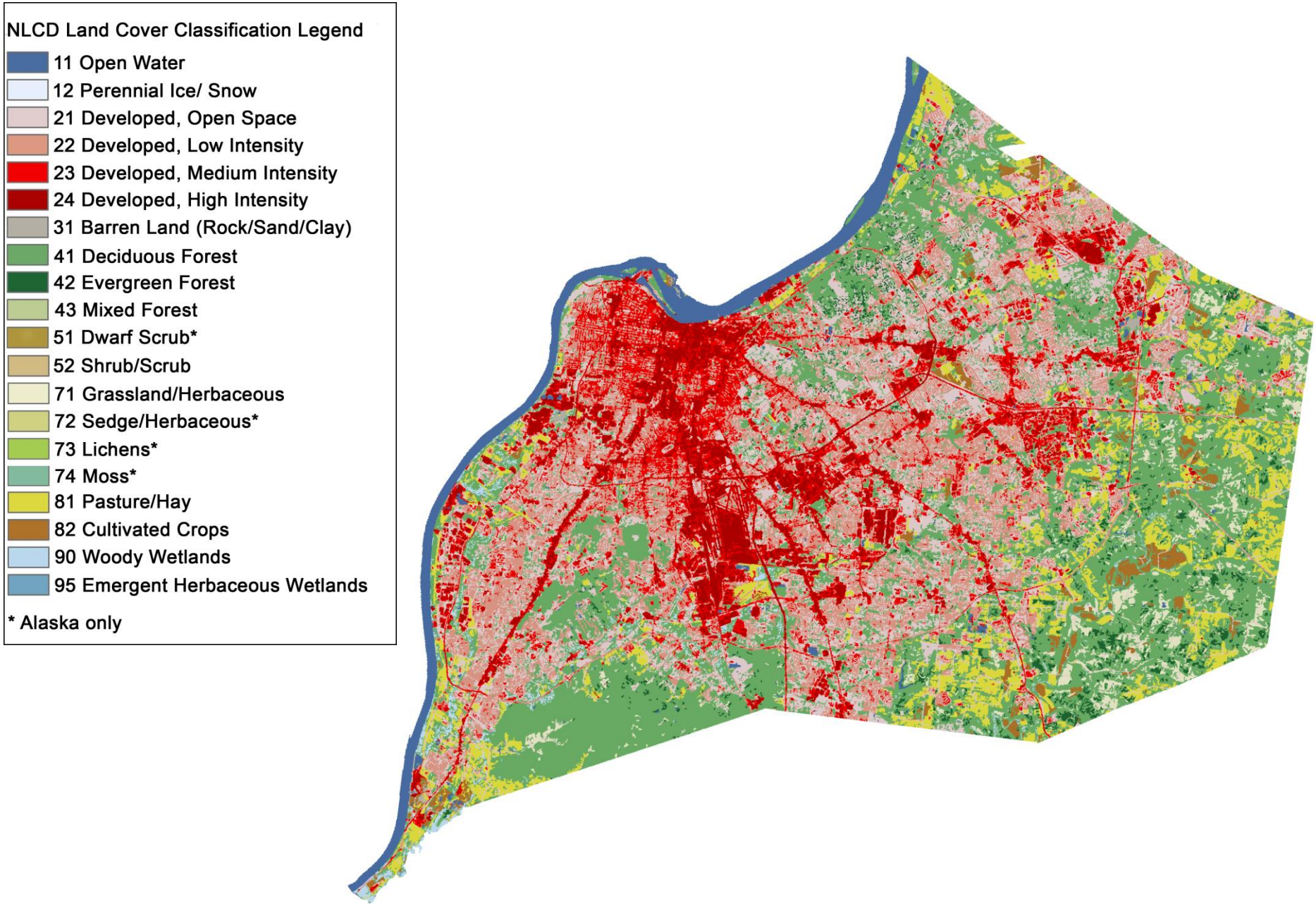

211110069001

Areal Weighting (**BG 211110069001**)

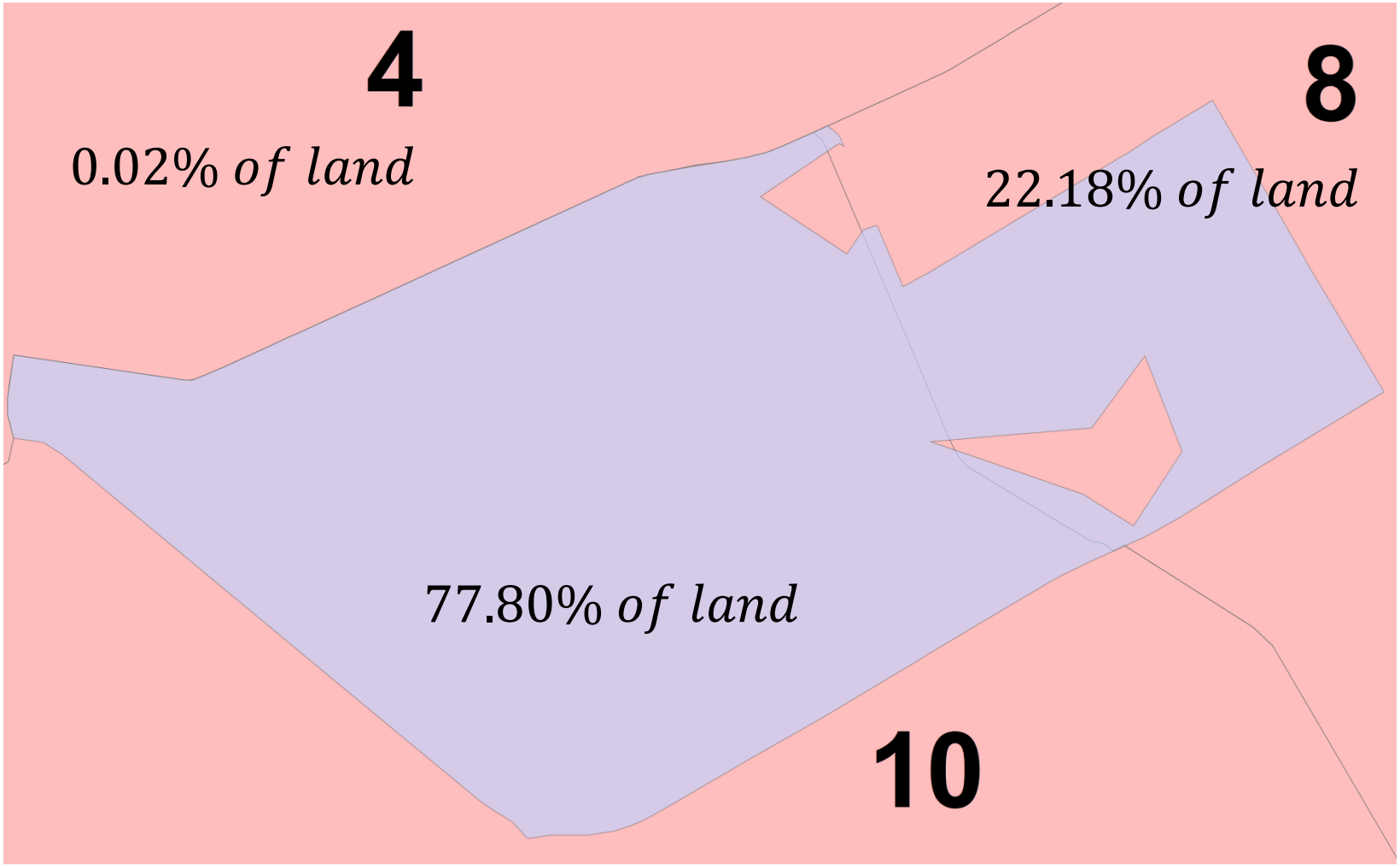


- "Population" within target zone is estimated as % of source zone area within target zone

National Land Cover Database

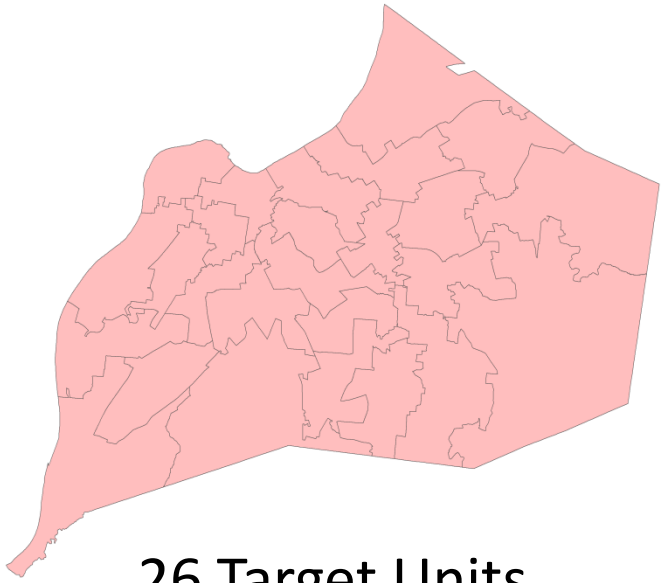


Spatially Refined Areal Weighting



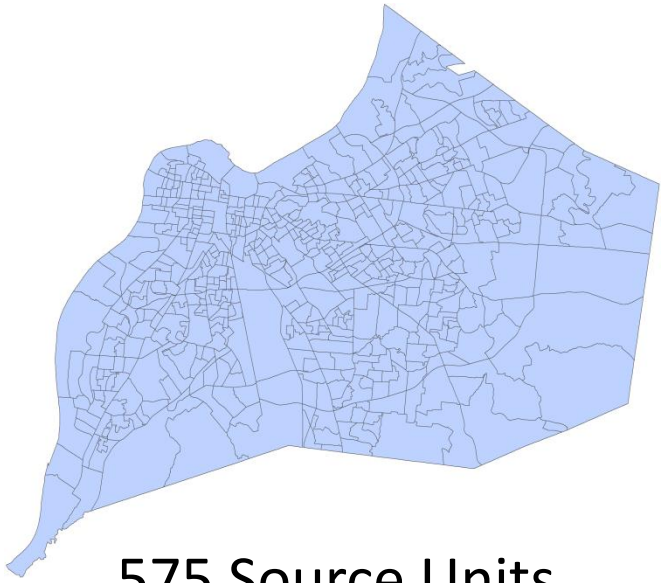
- Land that is unlikely to contain “population” is removed prior to the calculation of areas

Spatially Refined Areal Weighting



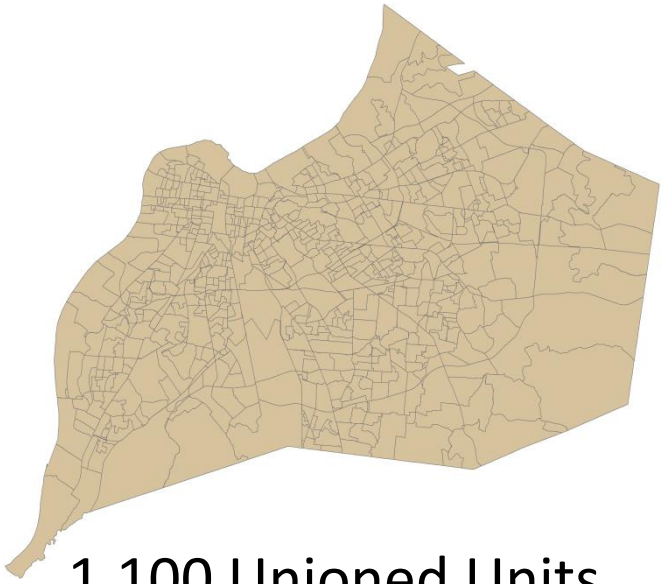
26 Target Units

+



575 Source Units

=



1,100 Unioned Units

Areal Weighting

Table

union_refined

FID	Shape *	GISJOIN	GEOID	AREA_BG	COUNDIST	AREA_CD	AREA_UNION
481	Polygon	G21011100007001	211110007001	243925	5	21718900	243925
480	Polygon	G21011100007002	211110007002	296630	5	21718900	296630
1067	Polygon	G21011100007003	211110007003	366512	5	21718900	366512
479	Polygon	G21011100008001	211110008001	243049	5	21718900	243049
787	Polygon	G21011100008002	211110008002	856114	5	21718900	856114
1065	Polygon	G21011100009001	211110009001	310540	1	37988100	5.41164
1066	Polygon	G21011100009001	211110009001	310540	5	21718900	310534
477	Polygon	G21011100009002	211110009002	323137	1	37988100	7.29674
478	Polygon	G21011100009002	211110009002	323137	5	21718900	323130
784	Polygon	G21011100010001	211110010001	419847	1	37988100	142917
785	Polygon	G21011100010001	211110010001	419847	5	21718900	110694
786	Polygon	G21011100010001	211110010001	419847	6	13308500	166236
1062	Polygon	G21011100010002	211110010002	322382	1	37988100	256193
1063	Polygon	G21011100010002	211110010002	322382	5	21718900	76.4162
1064	Polygon	G21011100010002	211110010002	322382	6	13308500	66112.6
234	Polygon	G21011100010003	211110010003	595586	1	37988100	595556
235	Polygon	G21011100010003	211110010003	595586	6	13308500	30.4889
65	Polygon	G21011100011001	211110011001	560644	1	37988100	140937
66	Polygon	G21011100011001	211110011001	560644	5	21718900	419707
232	Polygon	G21011100011002	211110011002	258681	1	37988100	122486
233	Polygon	G21011100011002	211110011002	258681	5	21718900	136195
947	Polygon	G21011100011003	211110011003	510159	1	37988100	99308.9
948	Polygon	G21011100011003	211110011003	510159	5	21718900	410850
963	Polygon	G21011100012001	211110012001	1362530	1	37988100	1269580
964	Polygon	G21011100012001	211110012001	1362530	5	21718900	92945.5
944	Polygon	G21011100012002	211110012002	878153	1	37988100	878149

0 (0 out of 1100 Selected)

union_refined

Areal Weighting

Table

union_refined

FID	Shape *	GISJOIN	GEOID	AREA_BG	COUNDIST	AREA_CD	AREA_UNION	PROP_BG
481	Polygon	G21011100007001	211110007001	243925	5	21718900	243925	1
480	Polygon	G21011100007002	211110007002	296630	5	21718900	296630	1
1067	Polygon	G21011100007003	211110007003	366512	5	21718900	366512	1
479	Polygon	G21011100008001	211110008001	243049	5	21718900	243049	1
787	Polygon	G21011100008002	211110008002	856114	5	21718900	856114	1
1065	Polygon	G21011100009001	211110009001	310540	1	37988100	5.41164	0.000017
1066	Polygon	G21011100009001	211110009001	310540	5	21718900	310534	0.999981
477	Polygon	G21011100009002	211110009002	323137	1	37988100	7.29674	0.000023
478	Polygon	G21011100009002	211110009002	323137	5	21718900	323130	0.999978
784	Polygon	G21011100010001	211110010001	419847	1	37988100	142917	0.340403
785	Polygon	G21011100010001	211110010001	419847	5	21718900	110694	0.263653
786	Polygon	G21011100010001	211110010001	419847	6	13308500	166236	0.395944
1062	Polygon	G21011100010002	211110010002	322382	1	37988100	256193	0.794688
1063	Polygon	G21011100010002	211110010002	322382	5	21718900	76.4162	0.000237
1064	Polygon	G21011100010002	211110010002	322382	6	13308500	66112.6	0.205075
234	Polygon	G21011100010003	211110010003	595586	1	37988100	595556	0.99995
235	Polygon	G21011100010003	211110010003	595586	6	13308500	30.4889	0.000051
65	Polygon	G21011100011001	211110011001	560644	1	37988100	140937	0.251384
66	Polygon	G21011100011001	211110011001	560644	5	21718900	419707	0.748616
232	Polygon	G21011100011002	211110011002	258681	1	37988100	122486	0.473502
233	Polygon	G21011100011002	211110011002	258681	5	21718900	136195	0.526498
947	Polygon	G21011100011003	211110011003	510159	1	37988100	99308.9	0.194663
948	Polygon	G21011100011003	211110011003	510159	5	21718900	410850	0.805337
963	Polygon	G21011100012001	211110012001	1362530	1	37988100	1269580	0.931781
964	Polygon	G21011100012001	211110012001	1362530	5	21718900	92945.5	0.068215
944	Polygon	G21011100012002	211110012002	878153	1	37988100	878149	0.999995

69 (0 out of 1100 Selected)

union_refined

Areal Weighting

Table

union_refined

FID	Shape *	GISJOIN	GEOID	AREA_BG	COUNDIST	AREA_CD	AREA_UNION	PROP_BG	totpop
481	Polygon	G21011100007001	211110007001	243925	5	21718900	243925	1	802
480	Polygon	G21011100007002	211110007002	296630	5	21718900	296630	1	1121
1067	Polygon	G21011100007003	211110007003	366512	5	21718900	366512	1	940
479	Polygon	G21011100008001	211110008001	243049	5	21718900	243049	1	1129
787	Polygon	G21011100008002	211110008002	856114	5	21718900	856114	1	1166
1065	Polygon	G21011100009001	211110009001	310540	1	37988100	5.41164	0.000017	1166
1066	Polygon	G21011100009001	211110009001	310540	5	21718900	310534	0.999981	1166
477	Polygon	G21011100009002	211110009002	323137	1	37988100	7.29674	0.000023	1012
478	Polygon	G21011100009002	211110009002	323137	5	21718900	323130	0.999978	1012
784	Polygon	G21011100010001	211110010001	419847	1	37988100	142917	0.340403	535
785	Polygon	G21011100010001	211110010001	419847	5	21718900	110694	0.263653	535
786	Polygon	G21011100010001	211110010001	419847	6	13308500	166236	0.395944	535
1062	Polygon	G21011100010002	211110010002	322382	1	37988100	256193	0.794688	764
1063	Polygon	G21011100010002	211110010002	322382	5	21718900	76.4162	0.000237	764
1064	Polygon	G21011100010002	211110010002	322382	6	13308500	66112.6	0.205075	764
234	Polygon	G21011100010003	211110010003	595586	1	37988100	595556	0.99995	1400
235	Polygon	G21011100010003	211110010003	595586	6	13308500	30.4889	0.000051	1400
65	Polygon	G21011100011001	211110011001	560644	1	37988100	140937	0.251384	1536
66	Polygon	G21011100011001	211110011001	560644	5	21718900	419707	0.748616	1536
232	Polygon	G21011100011002	211110011002	258681	1	37988100	122486	0.473502	756
233	Polygon	G21011100011002	211110011002	258681	5	21718900	136195	0.526498	756
947	Polygon	G21011100011003	211110011003	510159	1	37988100	99308.9	0.194663	1079
948	Polygon	G21011100011003	211110011003	510159	5	21718900	410850	0.805337	1079
963	Polygon	G21011100012001	211110012001	1362530	1	37988100	1269580	0.931781	915
964	Polygon	G21011100012001	211110012001	1362530	5	21718900	92945.5	0.068215	915
944	Polygon	G21011100012002	211110012002	878153	1	37988100	878149	0.999995	1013

69 (0 out of 1100 Selected)

union_refined

Areal Weighting

Table

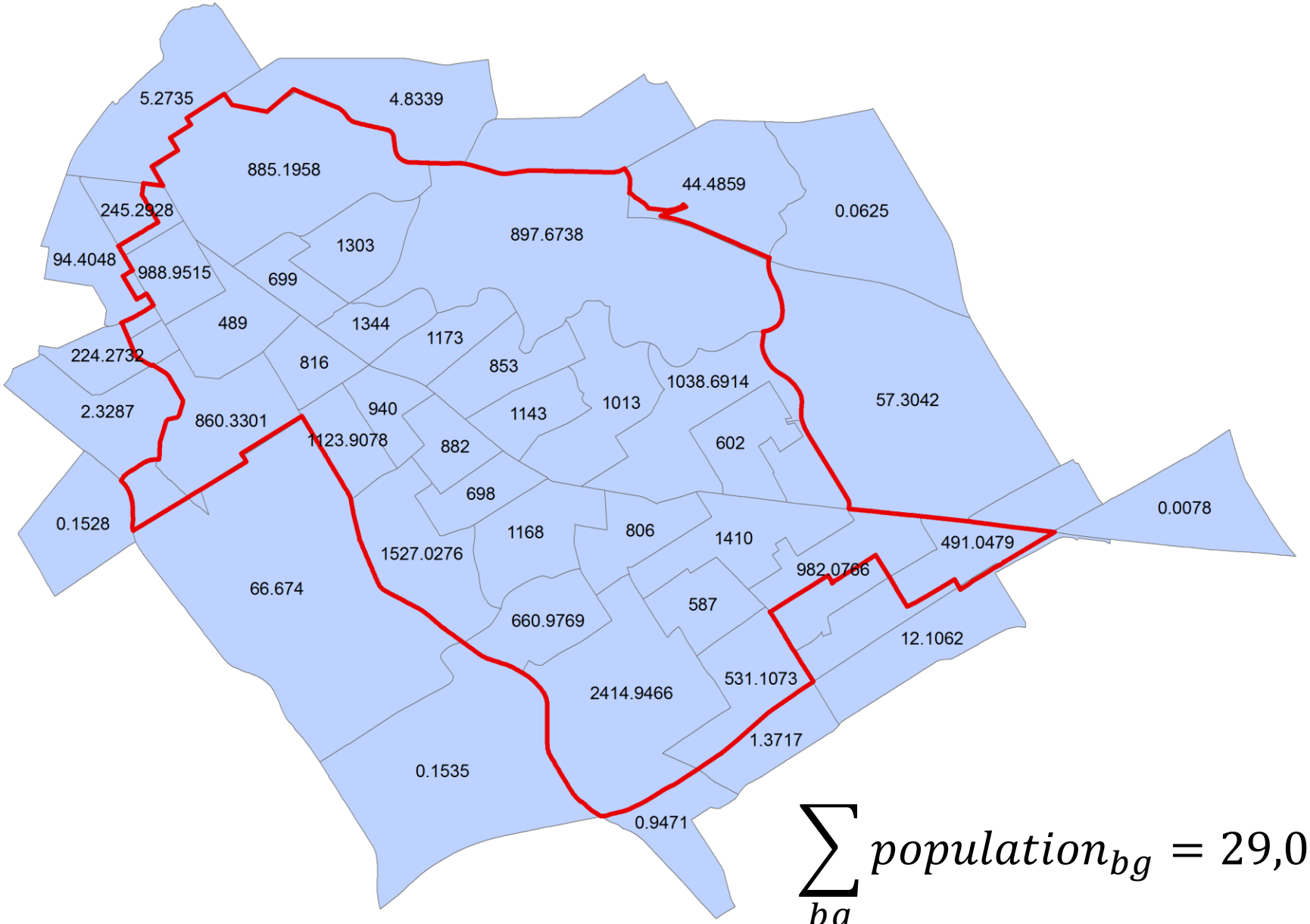
union_refined

FID	Shape *	GISJOIN	GEOID	AREA_BG	COUNDIST	AREA_CD	AREA_UNION	PROP_BG	totpop	totpop_adj
481	Polygon	G21011100007001	211110007001	243925	5	21718900	243925	1	802	802
480	Polygon	G21011100007002	211110007002	296630	5	21718900	296630	1	1121	1121
1067	Polygon	G21011100007003	211110007003	366512	5	21718900	366512	1	940	940
479	Polygon	G21011100008001	211110008001	243049	5	21718900	243049	1	1129	1129
787	Polygon	G21011100008002	211110008002	856114	5	21718900	856114	1	1166	1166
1065	Polygon	G21011100009001	211110009001	310540	1	37988100	5.41164	0.000017	1166	0.020319
1066	Polygon	G21011100009001	211110009001	310540	5	21718900	310534	0.999981	1166	1165.98
477	Polygon	G21011100009002	211110009002	323137	1	37988100	7.29674	0.000023	1012	0.022852
478	Polygon	G21011100009002	211110009002	323137	5	21718900	323130	0.999978	1012	1011.98
784	Polygon	G21011100010001	211110010001	419847	1	37988100	142917	0.340403	535	182.116
785	Polygon	G21011100010001	211110010001	419847	5	21718900	110694	0.263653	535	141.054
786	Polygon	G21011100010001	211110010001	419847	6	13308500	166236	0.395944	535	211.83
1062	Polygon	G21011100010002	211110010002	322382	1	37988100	256193	0.794688	764	607.142
1063	Polygon	G21011100010002	211110010002	322382	5	21718900	76.4162	0.000237	764	0.181096
1064	Polygon	G21011100010002	211110010002	322382	6	13308500	66112.6	0.205075	764	156.677
234	Polygon	G21011100010003	211110010003	595586	1	37988100	595556	0.99995	1400	1399.93
235	Polygon	G21011100010003	211110010003	595586	6	13308500	30.4889	0.000051	1400	0.071668
65	Polygon	G21011100011001	211110011001	560644	1	37988100	140937	0.251384	1536	386.126
66	Polygon	G21011100011001	211110011001	560644	5	21718900	419707	0.748616	1536	1149.87
232	Polygon	G21011100011002	211110011002	258681	1	37988100	122486	0.473502	756	357.968
233	Polygon	G21011100011002	211110011002	258681	5	21718900	136195	0.526498	756	398.033
947	Polygon	G21011100011003	211110011003	510159	1	37988100	99308.9	0.194663	1079	210.041
948	Polygon	G21011100011003	211110011003	510159	5	21718900	410850	0.805337	1079	868.959
963	Polygon	G21011100012001	211110012001	1362530	1	37988100	1269580	0.931781	915	852.58
964	Polygon	G21011100012001	211110012001	1362530	5	21718900	92945.5	0.068215	915	62.4171
944	Polygon	G21011100012002	211110012002	878153	1	37988100	878149	0.999995	1013	1012.99

67 (0 out of 1100 Selected)

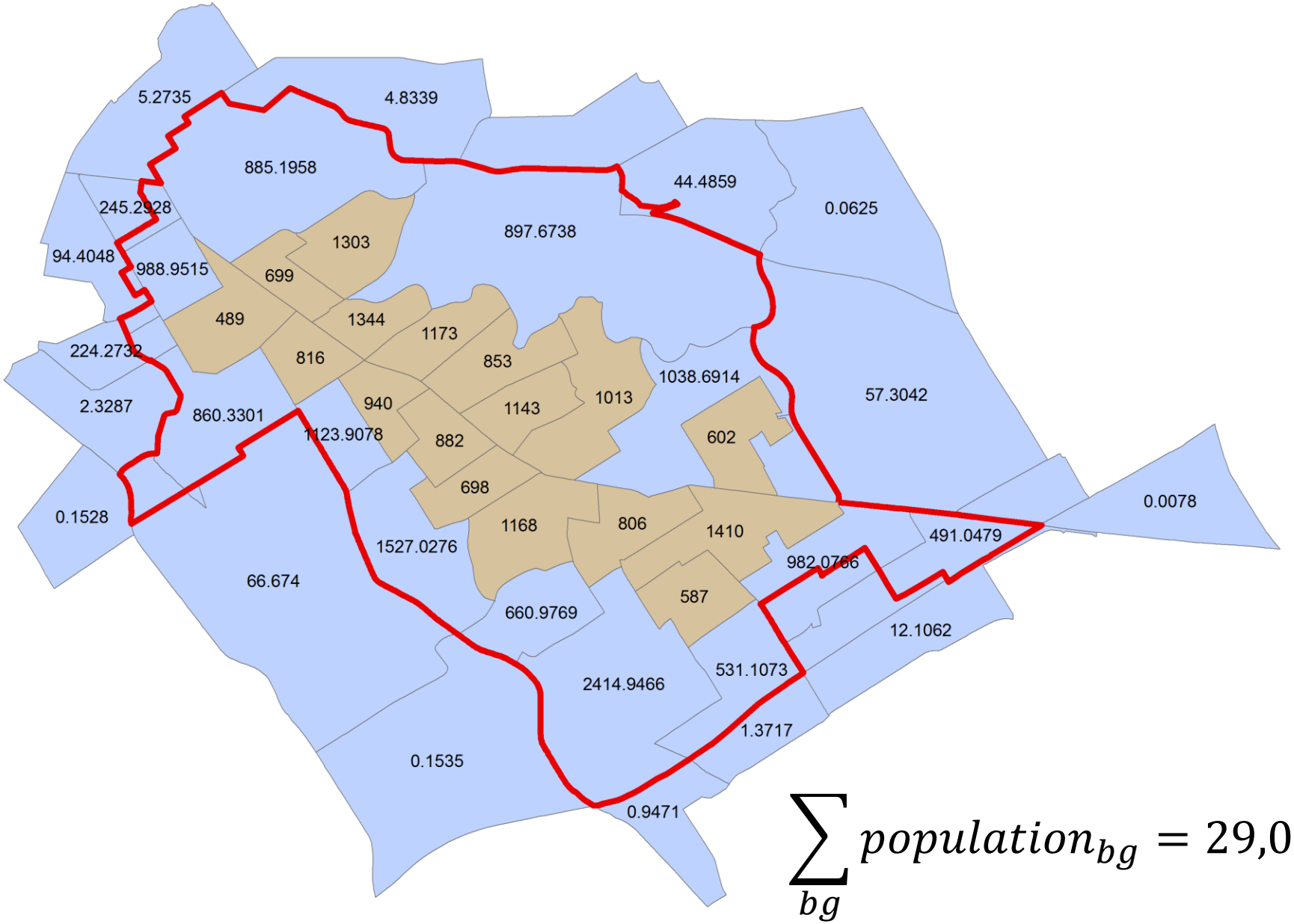
union_refined

Spatially Refined Areal Weighting



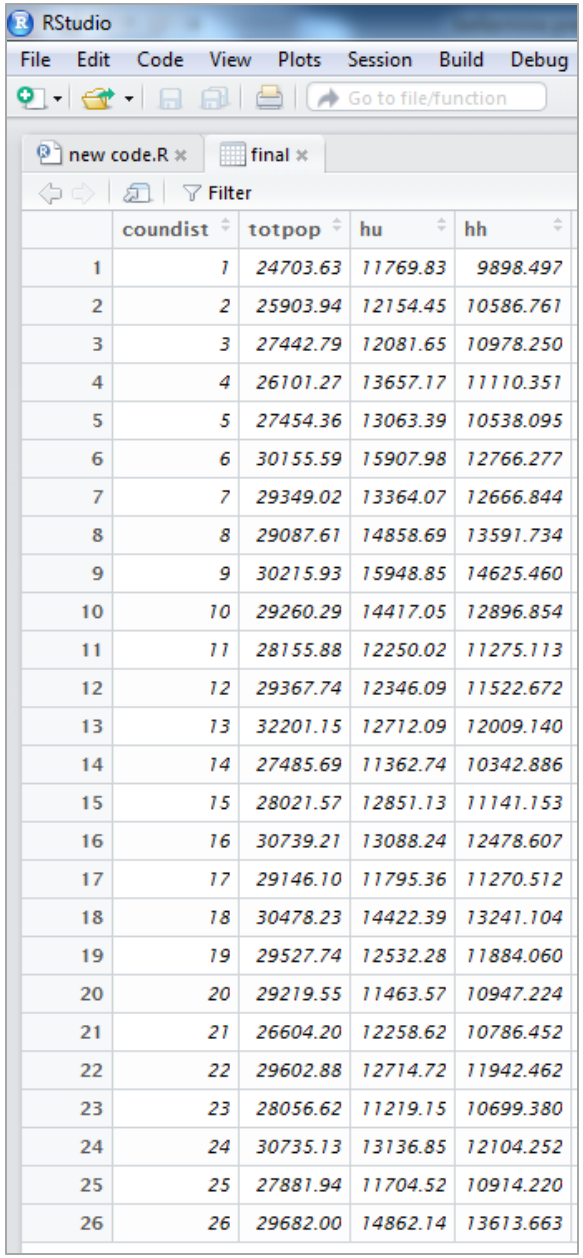
$$\sum_{bg} population_{bg} = 29{,}087.61$$

Spatially Refined Areal Weighting



$$\sum_{bg} population_{bg} = 29{,}087.61$$

Validation and Accuracy


RStudio

File Edit Code View Plots Session Build Debug

Go to file/function

new code.R × final ×

Filter

	coundist	totpop	hu	hh
1	1	24703.63	11769.83	9898.497
2	2	25903.94	12154.45	10586.761
3	3	27442.79	12081.65	10978.250
4	4	26101.27	13657.17	11110.351
5	5	27454.36	13063.39	10538.095
6	6	30155.59	15907.98	12766.277
7	7	29349.02	13364.07	12666.844
8	8	29087.61	14858.69	13591.734
9	9	30215.93	15948.85	14625.460
10	10	29260.29	14417.05	12896.854
11	11	28155.88	12250.02	11275.113
12	12	29367.74	12346.09	11522.672
13	13	32201.15	12712.09	12009.140
14	14	27485.69	11362.74	10342.886
15	15	28021.57	12851.13	11141.153
16	16	30739.21	13088.24	12478.607
17	17	29146.10	11795.36	11270.512
18	18	30478.23	14422.39	13241.104
19	19	29527.74	12532.28	11884.060
20	20	29219.55	11463.57	10947.224
21	21	26604.20	12258.62	10786.452
22	22	29602.88	12714.72	11942.462
23	23	28056.62	11219.15	10699.380
24	24	30735.13	13136.85	12104.252
25	25	27881.94	11704.52	10914.220
26	26	29682.00	14862.14	13613.663

$$\sum_{cd} population_{cd} = 746{,}580$$

$$\sum_{cd} housing\ units_{cd} = 337{,}943$$

$$\sum_{cd} households_{cd} = 305{,}832$$

These numbers should match (exactly) the sum from the block groups and the county total given in the ACS!

How does error arise?

Block Group Data

- All of the block group data should be entered in raw numbers (e.g., not proportions)
- For metrics that are ratios, the numerator and denominator should be aggregated separately
- The final step will be the calculation of the ratio, based on the numerator and the denominator aggregated within the council district

R code

- Load the required libraries (install them first if necessary)
- The options command changes scientific notation in display
- The file “bg to council.csv” contains the proportion of the block group that belongs within each council district
- The file “bg data.csv” contains the block group demographic data

```
library(foreign)
library(rgeos)
library(dplyr)

options(scipen=999)

prop<-read.csv("D:/onedrive/bellarmine/bg to council.csv")
data<-read.csv("D:/onedrive/bellarmine/bg data.csv")
```

R code

- The *left_join* command joins the two datasets by **geoid** to create the master dataset “joined”
- The dataframe “data_adj” is created to hold the areally weighted (adjusted) data
- The dataframe “data_adj” will contain an observation for each block group/council district intersection (n=1100)

```
joined<-left_join(prop,data,by="geoid")

data_adj<-data.frame(matrix(ncol=7,nrow=1100))   ### 'ncol'=number of vars
colnames(data_adj)<-colnames(data)
data_adj$geoid<-joined$geoid
data_adj$coundist<-joined$coundist
```

R code

- The “c” and “d” variables are seeds for the loop
- The loop multiplies each block group characteristic (e.g., population) by the block group proportion in each council district
- The dataframe “data_adj”

```
c<-1
d<-9   ###  9 is the column that the data begins in 'joined'

for (i in 9:ncol(joined)) {
  data_adj[,c]<-joined[,d]*joined$prop_bg
  c<-c+1
  d<-d+1
}
```

R code

- The *summarise* command sums the adjusted data for each variable over each of the council districts
- The dataframe “final” holds the aggregated data, and should have a number of observations equal to the number of council districts (n=26)
- This dataframe can be written to a .csv file (or other filetype)

```
(data_adj %>%
  group_by(coundist) %>%
  summarise_each(funs(sum))
) -> final

write.csv(final,"C:/users/mhruth01/desktop/final.csv")   # save in CSV format
```

GIS Data

- Software
  - ArcGIS
  - QGIS
  - R (or other statistical package)
- Boundary Files
  - Shapefile
  - Geodatabase
  - Keyhole Markup Language (KML) (Google Earth)
- Boundary File Sources
  - Census TIGER
    - All administrative and statistical geographies plus features
  - National Historical GIS
    - Includes data
  - Metro Data Portal

R code

- R can read both shapefiles and .kml files
- Requires the **rgdal** package
- The *spTransform* command changes the projection of the data

```
### Read boundaries from KML file
library(rgdal)
ogrListLayers(dsn="D:/onedrive/bellarmine/metrocouncildistricts.kml")
council<-readOGR("D:/onedrive/bellarmine/metrocouncildistricts.kml","Metro Council Districts")
council<-spTransform(council,CRS("+init=epsg:2205"))

### Read boundaries from shapefile
bg10<-readOGR(dsn="D:/onedrive/bellarmine",layer="bg10_refined")
plot(bg10)
council11<-readOGR(dsn="D:/onedrive/bellarmine",layer="council11")
plot(council11)
```