

Code ▾

Mass 211: Rural vs Urban - An In-Depth Look

In this notebook, I attempt to look deeper into human services needs of rural vs urban areas.

Some premises:

1. This analysis is done at the ZIP code level.
2. 2-1-1 calls are categorized into specific topics, namely: care/companion, childcare, community, disability, education, food/cloth, government, health, homeless, housing, income, infoservice, legal, mental, substance abuse, utilities, youth help. They cover about 90% of all the calls–i.e., 90% of the calls can be categorized under one or more such topics.
3. There might be overlaps in these topics—for example, *housing* would contain *homeless*, but calls for housing issues might not always be related to homeless people.
4. Calls to “Mental Health Hotlines” (Taxonmy code: RP-1500.1400-500) are removed, as they are prone to skewness caused by repetitive callers.

This is how the aggregated data looks like:

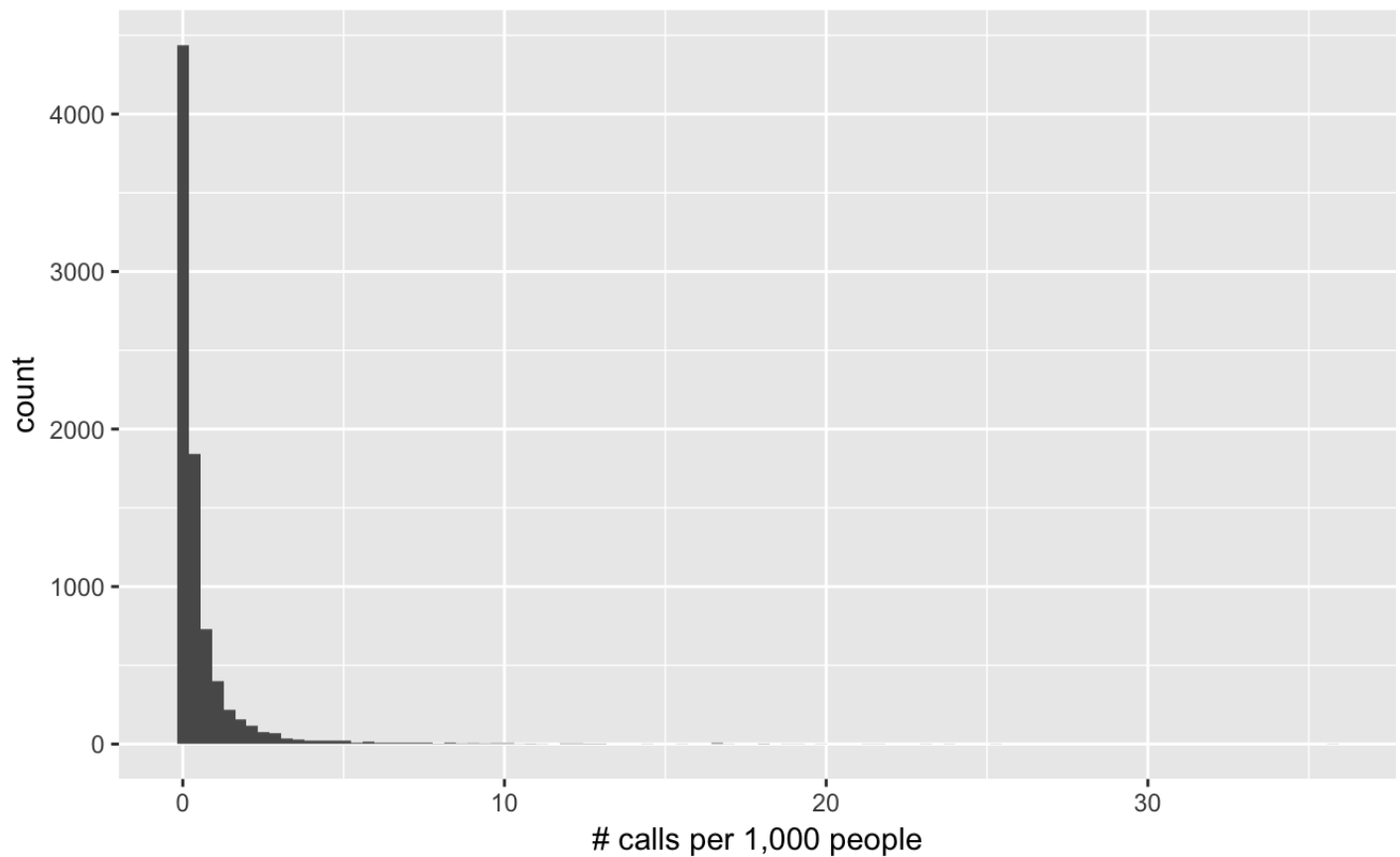
Code

na...	topic	n	TotalPop	n_topic_total	n_geo_total	p_1k	area	Pop
<fctr>	<fctr>	<int>	<dbl>	<int>	<int>	<dbl>	<dbl>	<
01001	care/companion	1	17438	104	182	0.05734603	11.95346	1458.
01001	childcare	20	17438	25838	182	1.14692052	11.95346	1458.
01001	community	0	17438	977	182	0.00000000	11.95346	1458.
01001	disability	2	17438	1175	182	0.11469205	11.95346	1458.
01001	education	2	17438	1703	182	0.11469205	11.95346	1458.
01001	food/cloth	10	17438	4991	182	0.57346026	11.95346	1458.
01001	government	4	17438	2217	182	0.22938410	11.95346	1458.
01001	health	7	17438	2003	182	0.40142218	11.95346	1458.
01001	homeless	7	17438	5082	182	0.40142218	11.95346	1458.
01001	housing	19	17438	14426	182	1.08957449	11.95346	1458.
1-10 of 20 rows						Previous	1	2
								Next

Removal of extreme values

First, let’s examine the distribution of the number of calls per 1,000 people—the variable we will use for future analysis.

0%	20%	40%	60%	80%	100%
0.00000000	0.00000000	0.05827845	0.25398402	0.68832327	35.74033552

[Code](#)

The vast majority of the observations (combination of a ZIP code and a service type) do not have any calls at all. As per-capita representation of small counts is very sensitive to the denominator, we would want to set a threshold for the minimal number of residents, so to avoid being trapped by “incidental” calls from these neighborhoods.

[Code](#)

name <fctr>	topic <fctr>	n <int>	TotalPop <dbl>	p_1k <dbl>
01343	utilities	1	46	21.739130
02713	mental	1	60	16.666667
02713	substance abuse	1	60	16.666667
02647	childcare	1	113	8.849558
02637	childcare	2	117	17.094017
01224	childcare	1	138	7.246377
01224	utilities	1	138	7.246377
02553	childcare	6	237	25.316456

name <fctr>	topic <fctr>	n <int>	TotalPop <dbl>	p_1k <dbl>
02553	legal	2	237	8.438819
02561	childcare	8	478	16.736402
1-10 of 136 rows		Previous	1 2 3 4 5 6 ... 14	Next

We can see that population with near 200 people is still generating incidental calls (# of calls < 3). In order to be consistent, I am removing ZIP codes with less than 200 people for all service categories.

Code

```
[1] 0.9897751
```

This accounts for about 1% of all observations—we still have 99% of the data.

Definition of Urban Areas

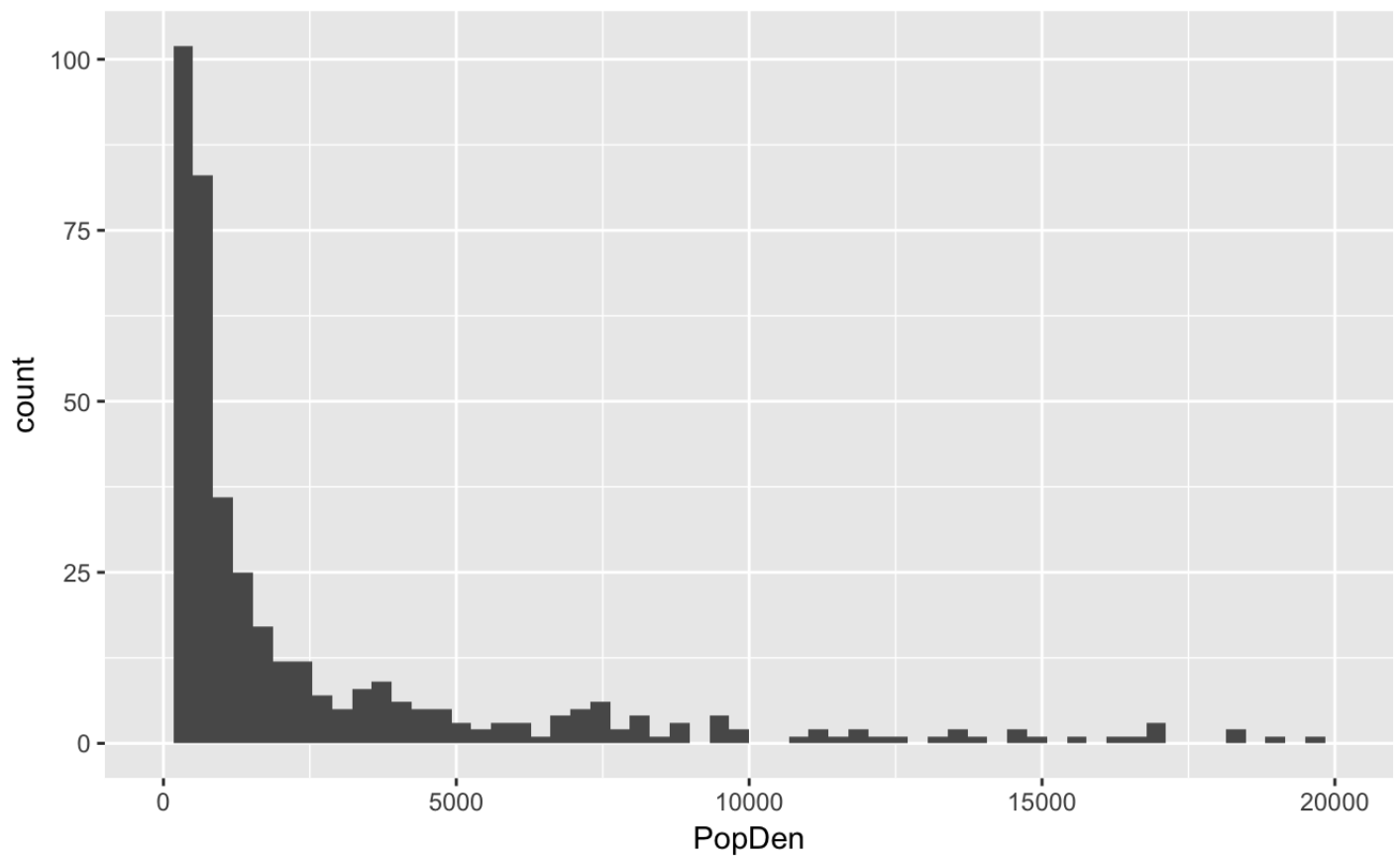
The U.S. Census Bureau defines an urban area as: “Core census block groups or blocks that have a population density of at least 1,000 people per square mile (386 per square kilometer) and surrounding census blocks that have an overall density of at least 500 people per square mile (193 per square kilometer).”¹ Areas with less than 50,000 people are called *urban clusters*, while those with more than 50,000 people are called urbanized areas.

Since we do not have exact coordinate of the callers to map them to census blocks, we simply apply the same density definition to the ZIP code areas.

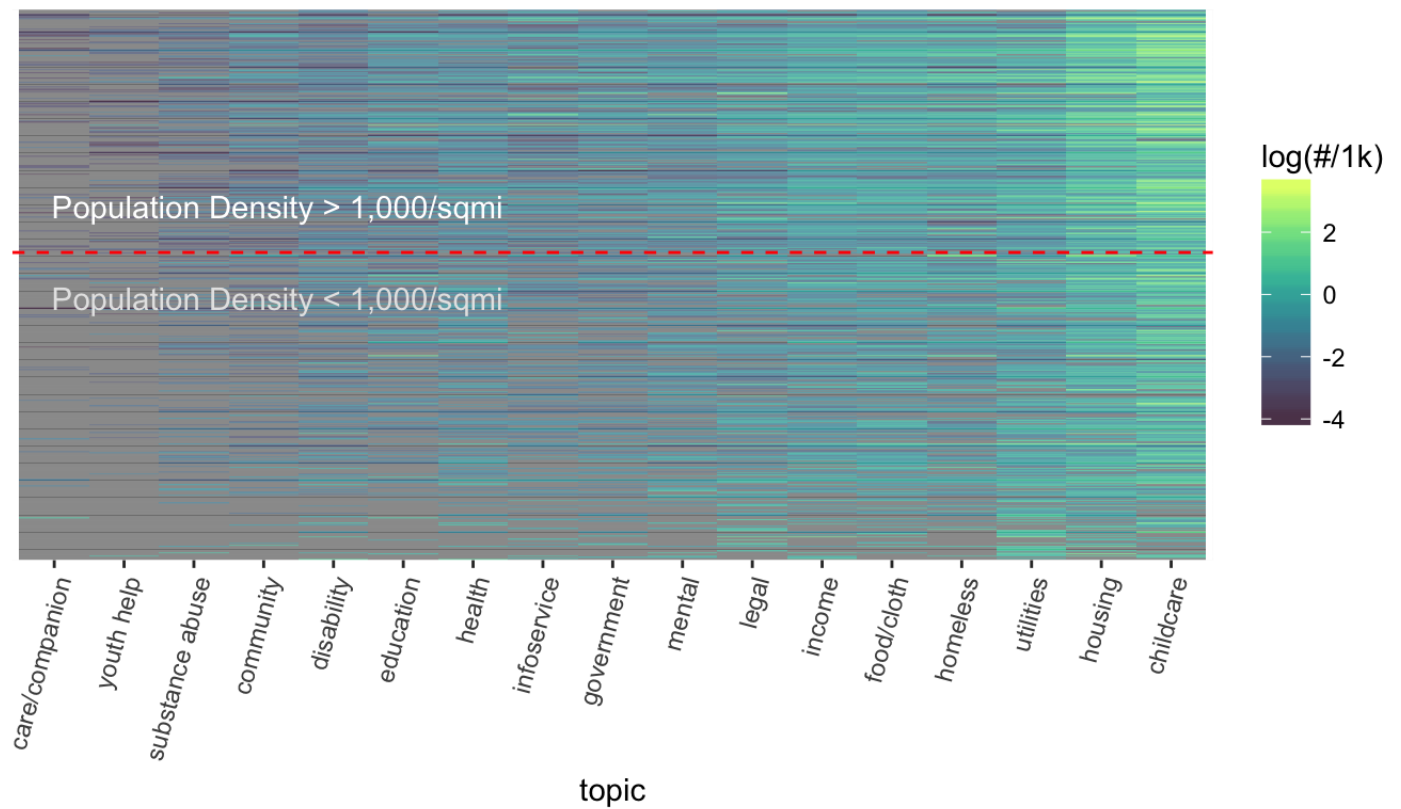
Slightly less than half of all ZIP codes met the density requirement. The example values and the distribution are as follows:

Code

name <fctr>	PopDen <dbl>	TotalPop <dbl>	area <dbl>	is_urban <lgl>
02113	78617.311	7221	0.091850	TRUE
02557	61652.051	4626	0.075034	TRUE
02111	38955.535	7537	0.193477	TRUE
02115	37575.896	28536	0.759423	TRUE
02215	26921.688	24521	0.910827	TRUE
02114	26523.255	12882	0.485687	TRUE
02118	25019.383	27526	1.100187	TRUE
02121	24622.382	28051	1.139248	TRUE
02116	23482.799	22433	0.955295	TRUE
02108	23181.836	4183	0.180443	TRUE
1-10 of 213 rows		Previous	1 2 3 4 5 6 ... 22	Next

[Code](#)[Code](#)

Log # of calls per 1,000 people for each ZIP Code and topic

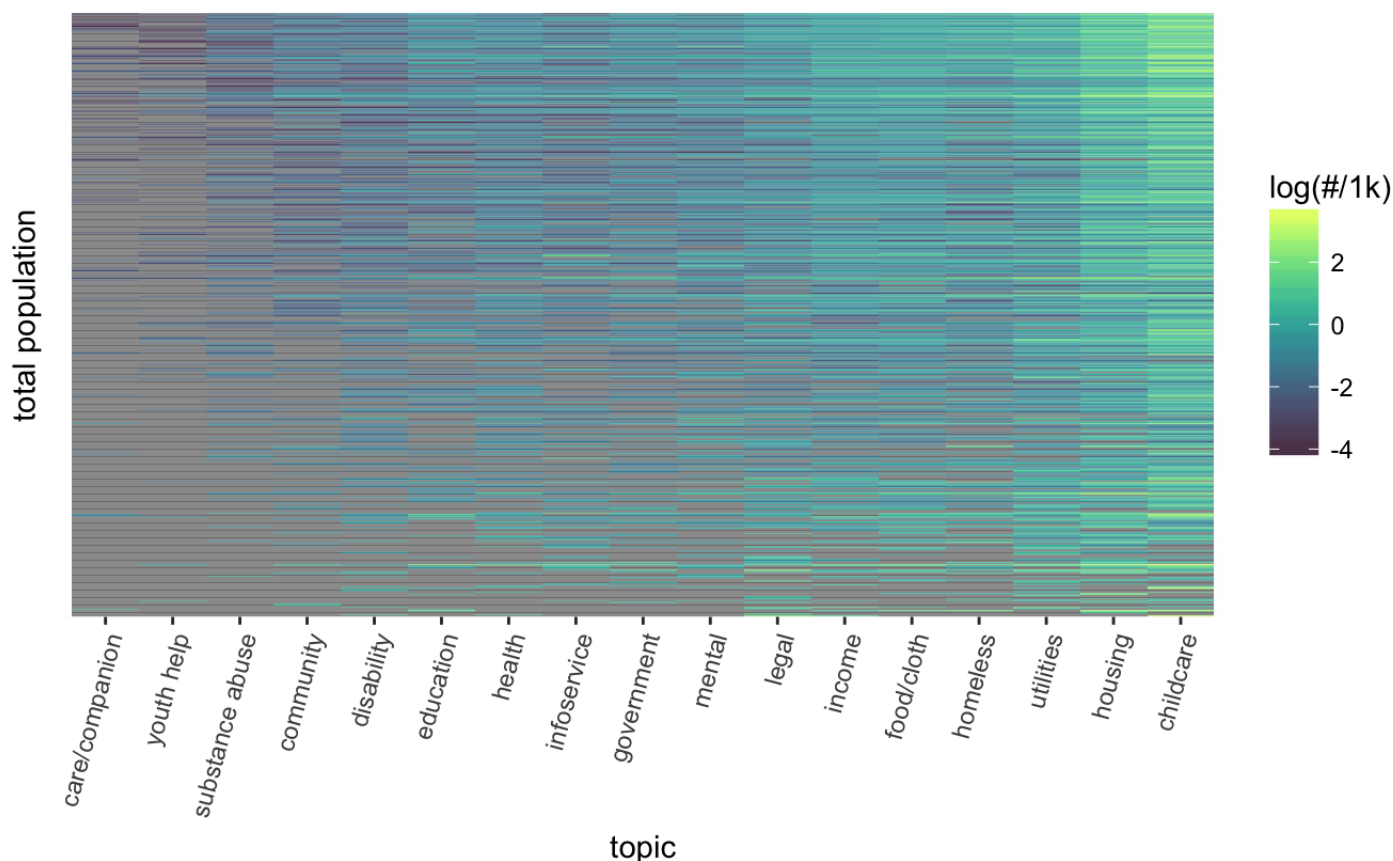


In this heatmap, each tile represents the log number of calls to one of the topics from a ZIP code area. The ZIP codes are ordered along the Y axis based on population density. More densely populated areas appear at the top of the graph. The X axis is sorted by the total number of calls for that topic.

While not obvious, it does seem that densely populated urban areas tend to generate more calls for the most popular types of services (brighter yellow color at the top right corner). It also appears that the most sparsely populated areas generate more utilities and legal related requests.

Define urban areas by population only

Considering the size of ZIP code areas varies and they are in general delineated with population clustering in mind, we may also try to use just the total population as the indicator for urbanness.

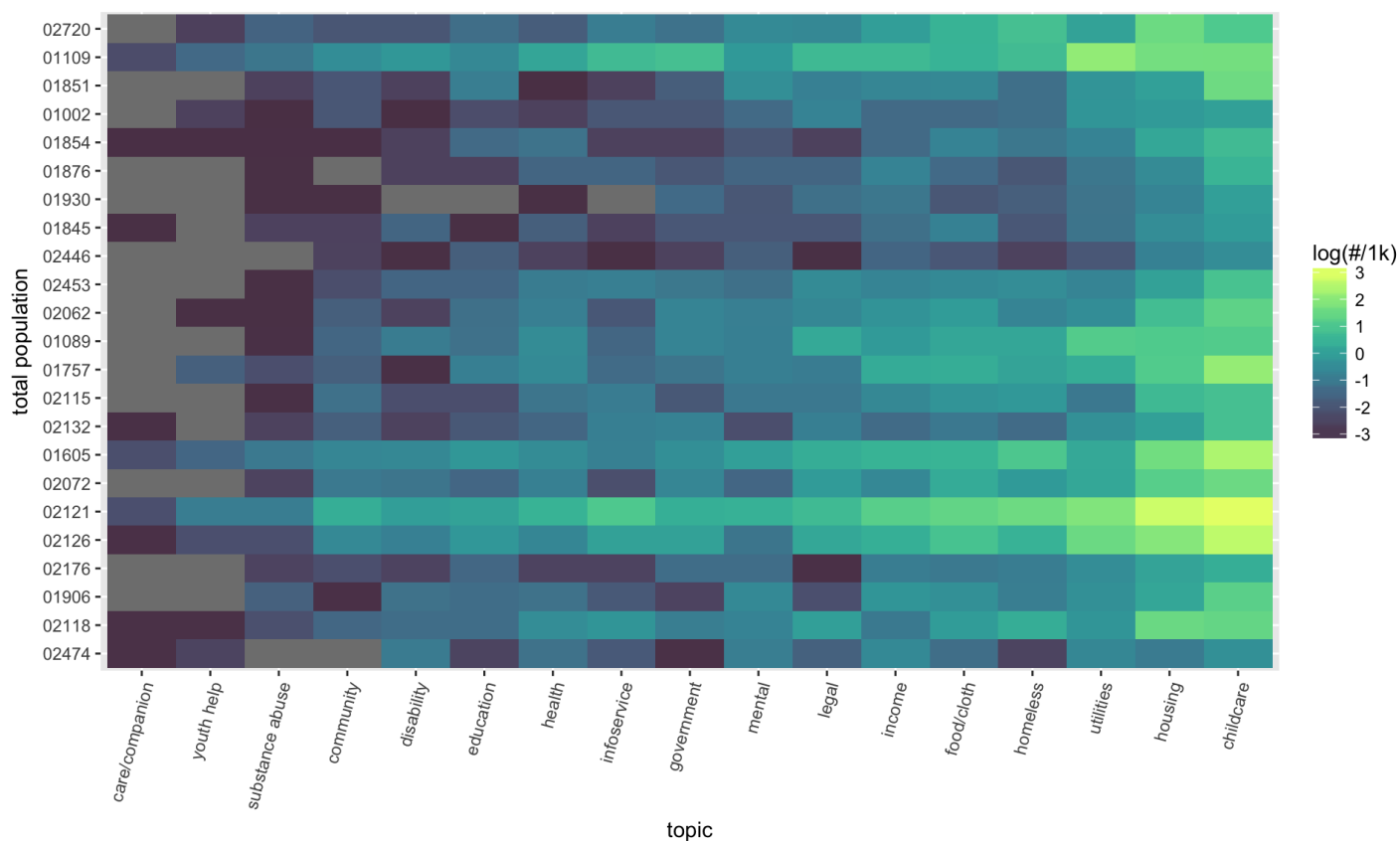
[Code](#)

It is obvious that ZIP codes with small population have larger probability of not getting any calls at all (the gray tiles) and they have also a larger variance in the per capital number of calls data. Both facts are not really surprising.

Zoom-in

What is more interesting is the “dent” near the top of the graph (ZIP codes with larger population). We can zoom in the graph and find out what are those exactly.

[Code](#)



Code

zip <chr>	TotalPop <dbl>	PopDen <dbl>	town <chr>	county <chr>	zt_area <dbl>	zt_area_sqm <dbl>
02720	30546	1536.2889	Fall River	Bristol	19.882979	51496426
01109	30541	5476.1887	Springfield	Hampden	5.577054	14443504
01851	30022	8840.0725	Lowell	Middlesex	3.396126	8795787
01002	29780	558.5469	Amherst	Hampshire	52.871350	136927661
01002	29780	558.5469	Shutesbury	Franklin	0.445573	1153964
01854	29557	6710.5622	Lowell	Middlesex	4.404549	11407614
01876	29479	1401.2636	Tewksbury	Middlesex	21.037441	54485640
01930	29399	1068.6497	Gloucester	Essex	27.510419	71250488
01845	29271	1051.7226	North Andover	Essex	27.831482	72082941
02446	29099	21691.6975	Brookline	Norfolk	1.341481	3474194
1-10 of 24 rows					Previous	1 2 3 Next

Those with relatively low demand:

- 01851, 01854 - UMass Lowell, where most of the residents are on-campus college students.
- **01002** - UMass Amherst campus
- **01876** - The Town of Tewksbury, suburb of the City of Lowell.
- **01930** - The Town of Gloucester, a relative remote area at the northeastern corner of MA, 94% white.

- **01845** - Residential area near the Phillips Academy.
- **02446** - The northern part of Brookline, median house income 80k. 72% white, 38% of the residents are of age 18-34.
- 02453 - The City of Waltham, encompasses the campus of Brandeis University and residential area near Bentley University.
- 02062 - The Town of Nordwood, 82% white, median house income 80k.
- 01757 - The Town of Milford, median income 70k, 80% white, 8% in poverty.
- 02115 - The campus of Northeastern University.
- 02132 - West Roxbury, 80% white, median house income 88k.

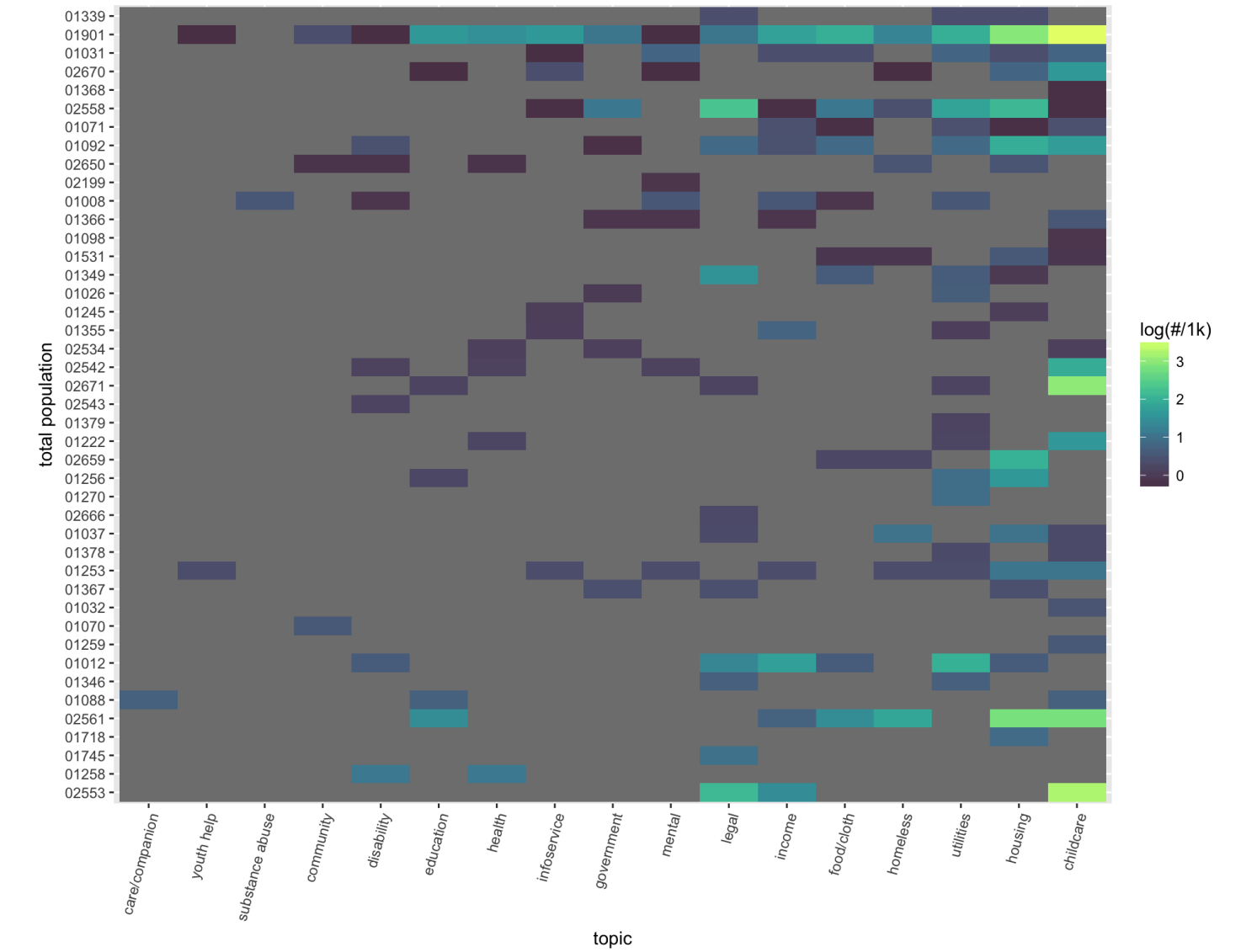
Those with high demand from this graph are:

- 01109 - Eastern part of Springfield, encompasses Old Hill, Bay and Upper Hill neighborhood. 37% hispanics, 14% Puerto Ricans, 37% black, median house income 34k.
- 01605 - Worcester, neighborhoods around Gree Hill Park, near the campus of Worcester Polytechnic Institute and UMass Medical School.
- 02121 - Roxbury, Boston, predominantly black and hispanic neighborhood, median house income 26k
- 02126 - Mattapan, Boston, predominantly black neighborhood, median house income 46k

Most of the neighborhoods with a large population but low human service demands are university towns or residential areas near a university, where the population are young and mostly non-local.

We can zoom in to the neighborhoods with small population, too:

Code



Code

zip	TotalPop	PopDen	town	county	zt_area	zt_area_sqm
<chr>	<dbl>	<dbl>	<chr>	<chr>	<dbl>	<dbl>
02553	237	1507.30758	Bourne	Barnstable	0.157234	407232.5
01258	334	12.96615	Mt Washington	Berkshire	25.759372	66711925.3
01745	373	1348.05472	Southborough	Worcester	0.276695	716587.7
01718	415	3413.41844	Acton	Middlesex	0.121579	314875.0
02561	478	542.66259	Bourne	Barnstable	0.880842	2281335.6
01088	493	72.98566	Hatfield	Hampshire	6.754752	17493672.3
01346	496	51.83281	Heath	Franklin	9.569228	24784030.4
01012	539	37.53483	Chesterfield	Hampshire	14.359997	37189954.9
01259	563	22.07551	New Marlborough	Berkshire	25.503369	66049026.2
01070	565	26.98294	Plainfield	Hampshire	20.939157	54229963.4

1-10 of 45 rows	Previous	1	2	3	4	5	Next
-----------------	----------	---	---	---	---	---	------

- 02553 - A very small fraction of Bourne, not actually a sparsely populated area per se, has a small population only because small land size.
- 01092 - A small area at West Warren.
- 01901 - A small fraction of Lynn around the Central Square, population 1,371, 39.4% are foreign born (14% are naturalized citizen) and 38% are hispanics. The neighboring ZIP code 01902 also have a fair amount of hispanics and high demand in 2-1-1 services, especially about child care.
- 02558 - Onset, a small town in Cape Cod. 78% white, 58% female, 42% age 45-54, Median house income 43k, near a country club.

The official urban areas

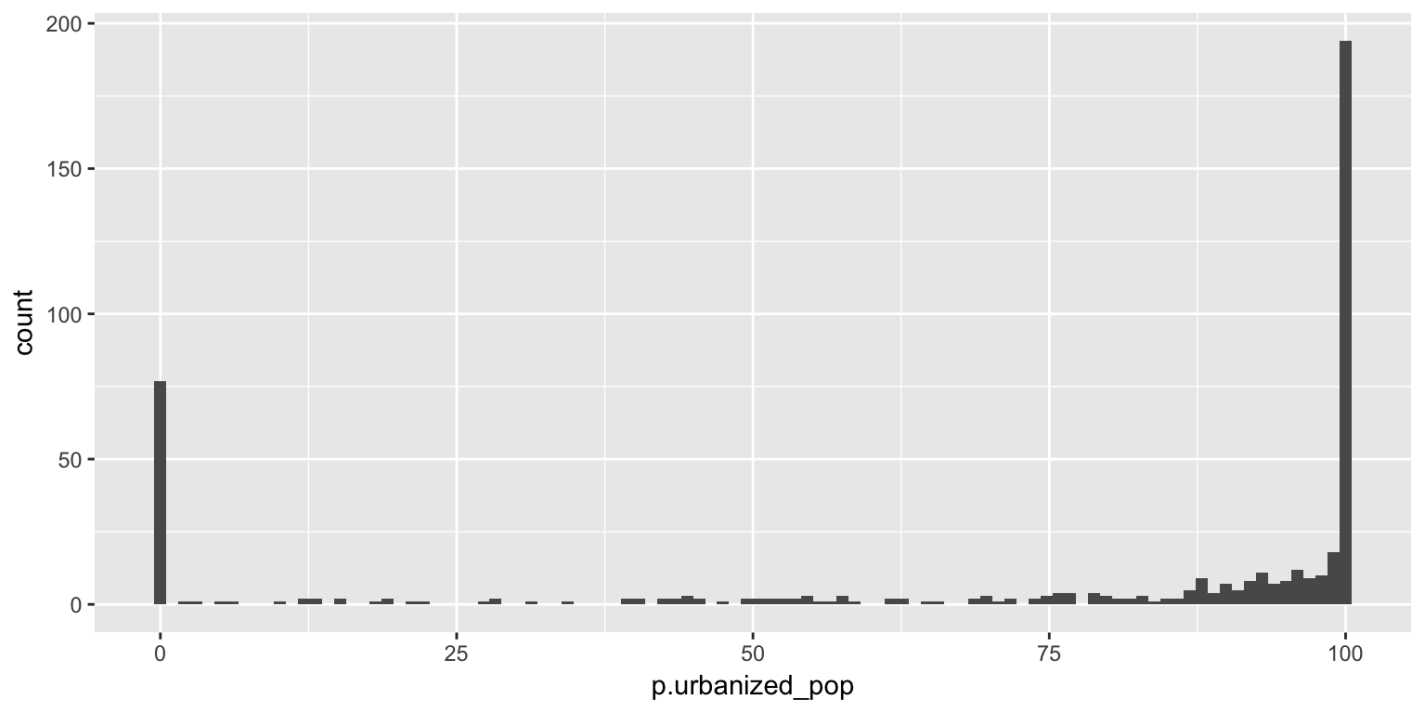
The official Urban Area Relationship Files (https://www.census.gov/geo/maps-data/data/ua_rel_download.html) published by the Census Bureau maps ZIP code and other geographic areas to urbanized areas and urban clusters. It is possible to calculate how many percentage of population and housing units in a ZIP code are within urban areas.

Code

na...	topic	n	TotalPop	n_topic_total	n_geo_total	zip	n_scaled_per_geo	
<fctr>	<fctr>	<int>	<dbl>	<int>	<int>	<chr>	<dbl>	
01001	care/companion	1	17438	104	182	01001	0.005494505	0
01001	childcare	20	17438	25838	182	01001	0.109890110	1
01001	community	0	17438	977	182	01001	0.000000000	0
01001	disability	2	17438	1175	182	01001	0.010989011	0
01001	education	2	17438	1703	182	01001	0.010989011	0
01001	food/cloth	10	17438	4991	182	01001	0.054945055	0
01001	government	4	17438	2217	182	01001	0.021978022	0
01001	health	7	17438	2003	182	01001	0.038461538	0
01001	homeless	7	17438	5082	182	01001	0.038461538	0
01001	housing	19	17438	14426	182	01001	0.104395604	1
1-10 of 20 rows 1-9 of 20 columns					Previous	1	2	Next

Code

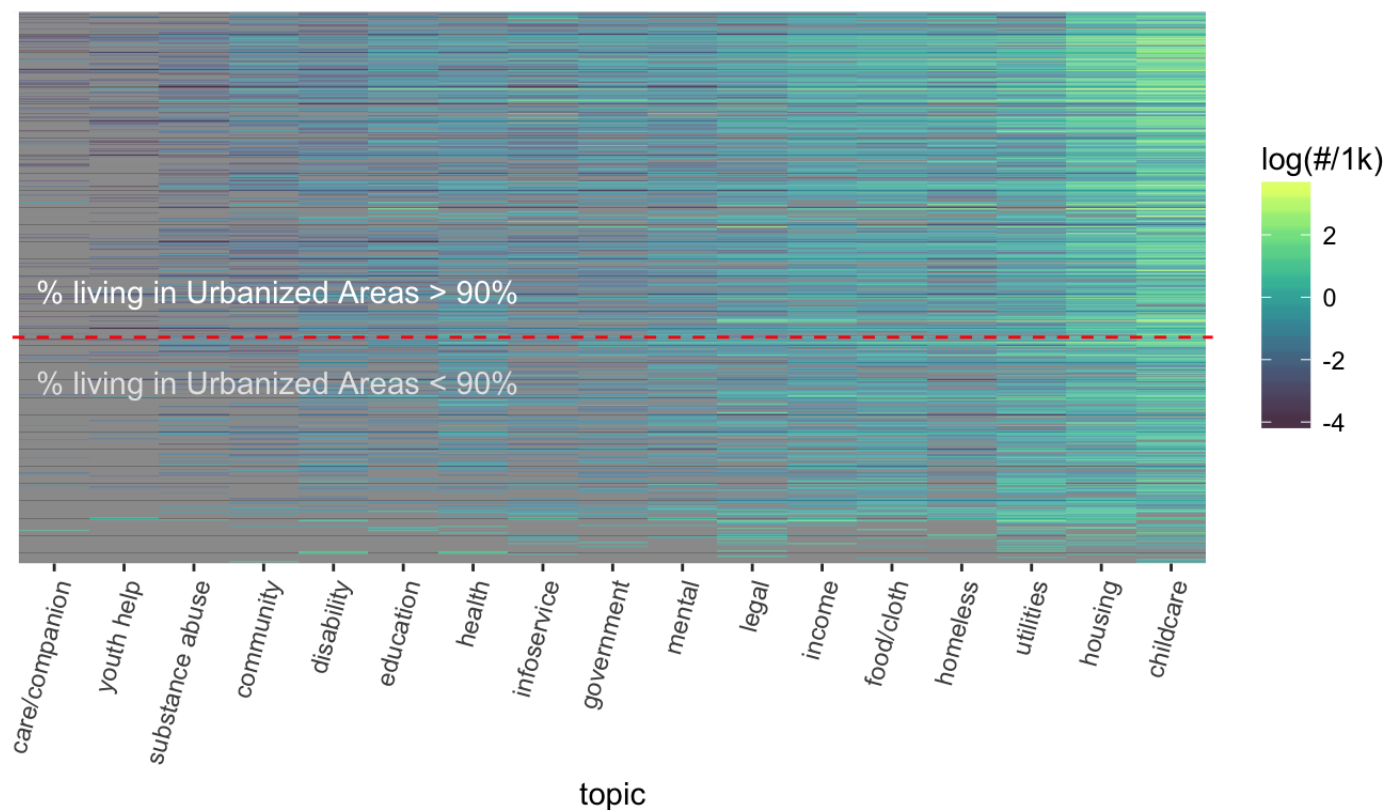
Code



Nearly half of the ZIP code areas are located in urban areas 100%.

[Code](#)

Log # of calls per 1,000 people for each ZIP Code and topic

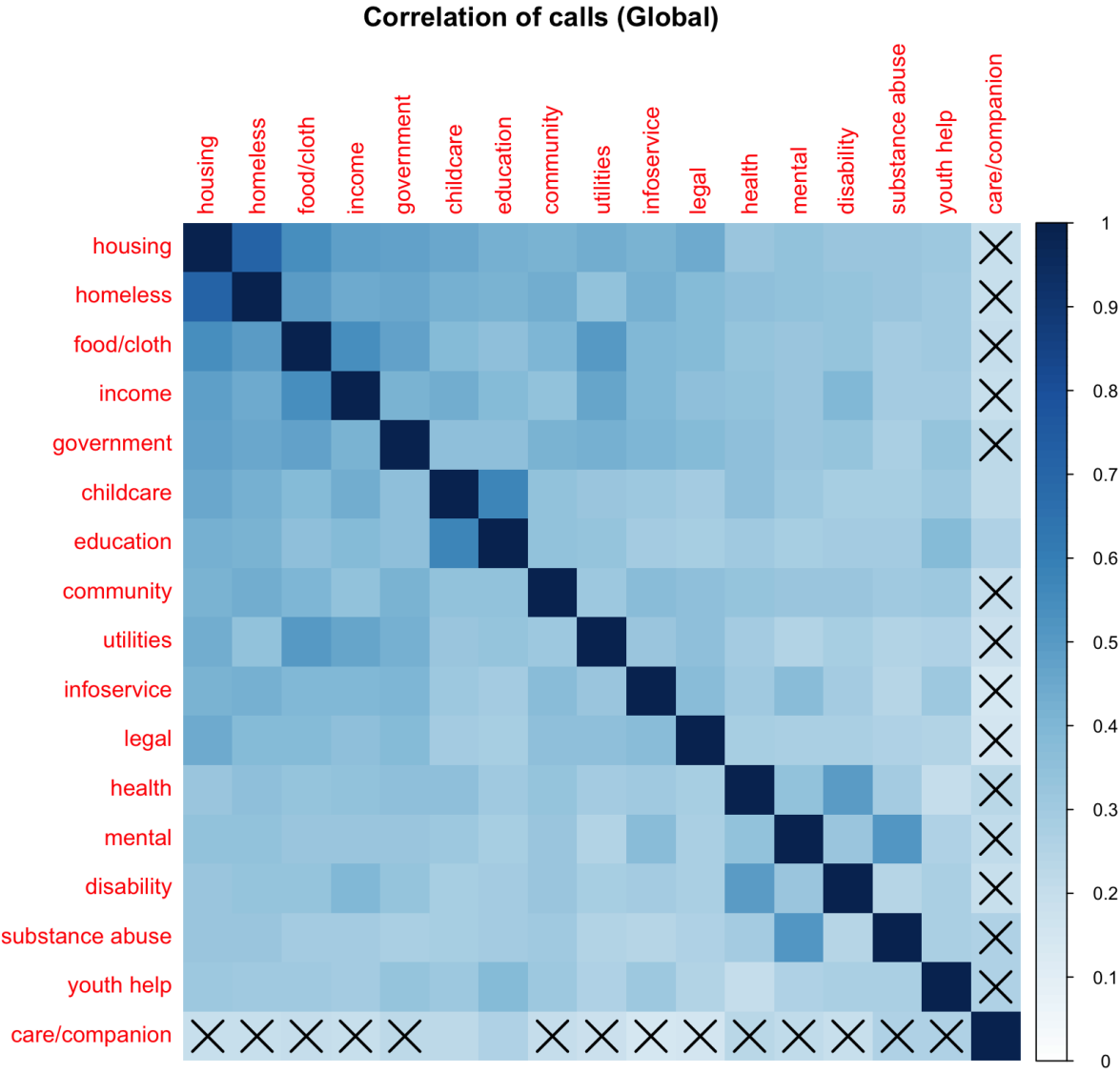


The heatmap is not that different than what we get based on naive population density, but does show a stronger signal in terms of high density urban areas generating more calls. We could also see suburban areas (those near the red dashed line) seem to be generating a relatively higher amount of health and disability related calls.

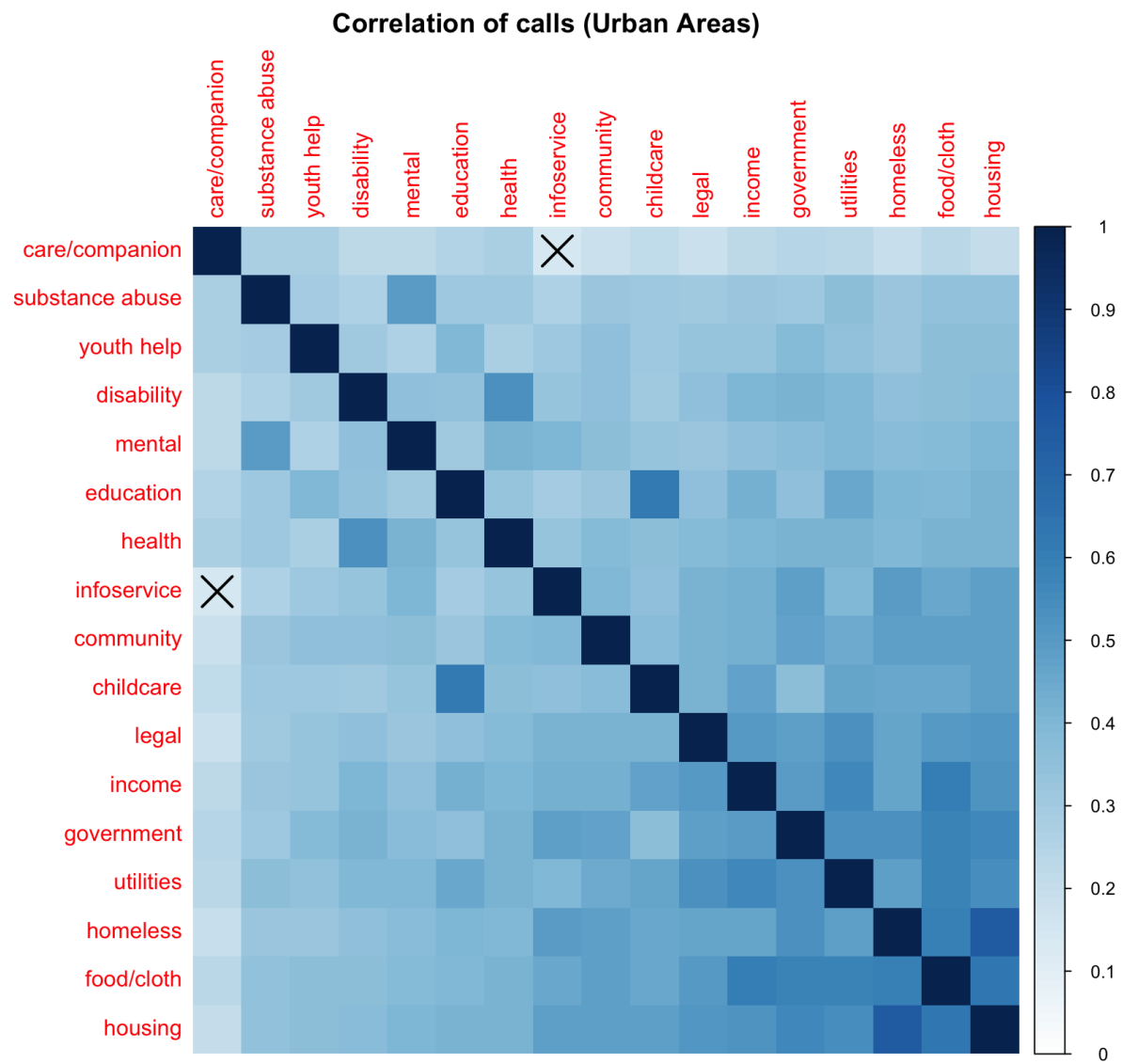
Topic correlations with Rural vs Urban

We would like to examine the correlation between each type of calls.

Code

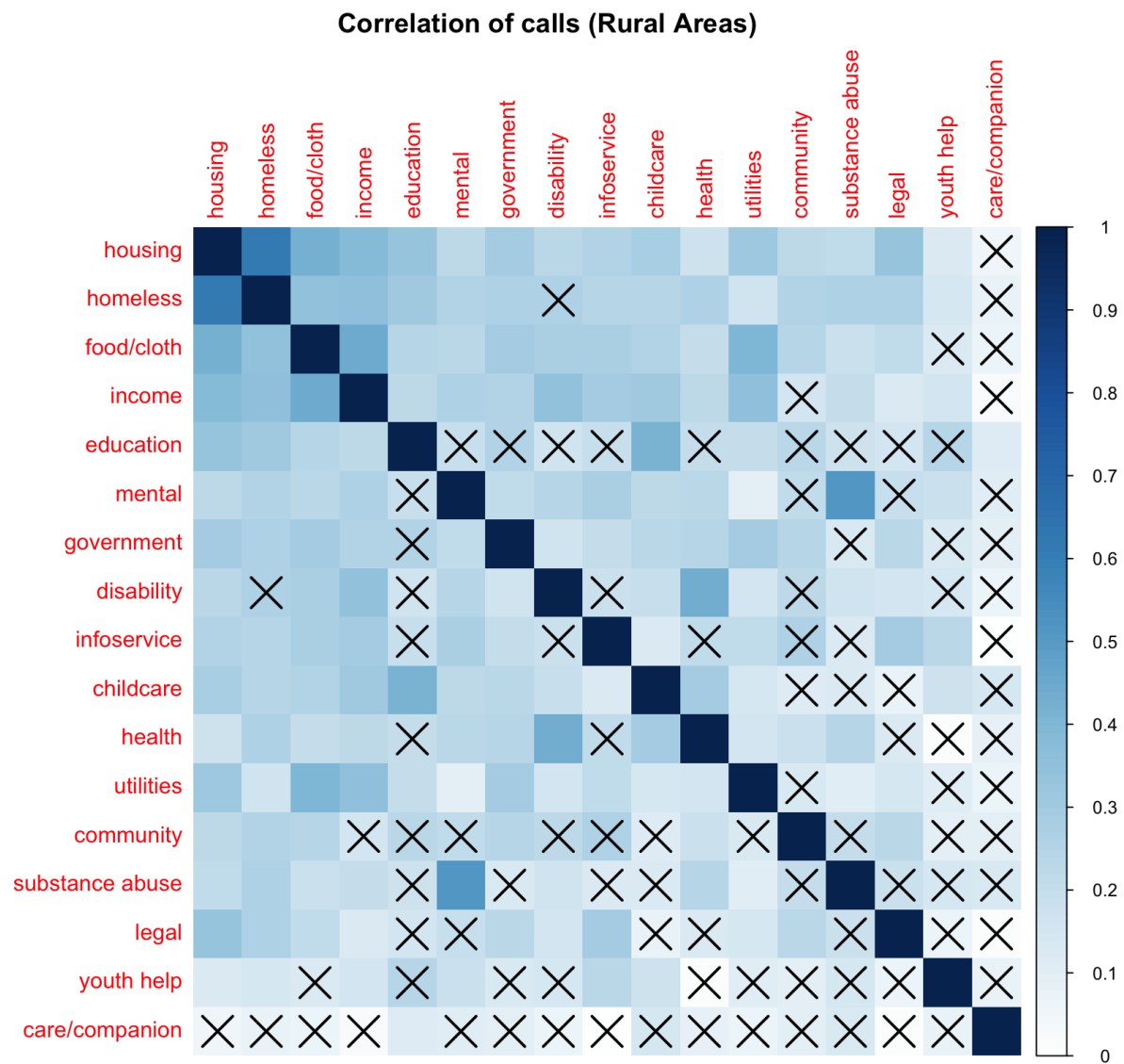


Code



(for some reason, the order of the strongest components for this graph is flipped, and I still haven't figure out why...)

Code



Above graphs shows the correlation between pairs of per capita call counts for each topic, using the Kendal's Tau method. The topics are ordered with the First Principle Component order. Squares with "X" means the correlation is not statistically significant (p-value > 0.01)

The fact that housing is always ranked as the strongest principle component indicates that housing issue might be a good predictor of the general needs of other service types.

Correlation between resource density and calls

We may also examine the correlation between number of resources available v.s. number of calls of given service type.

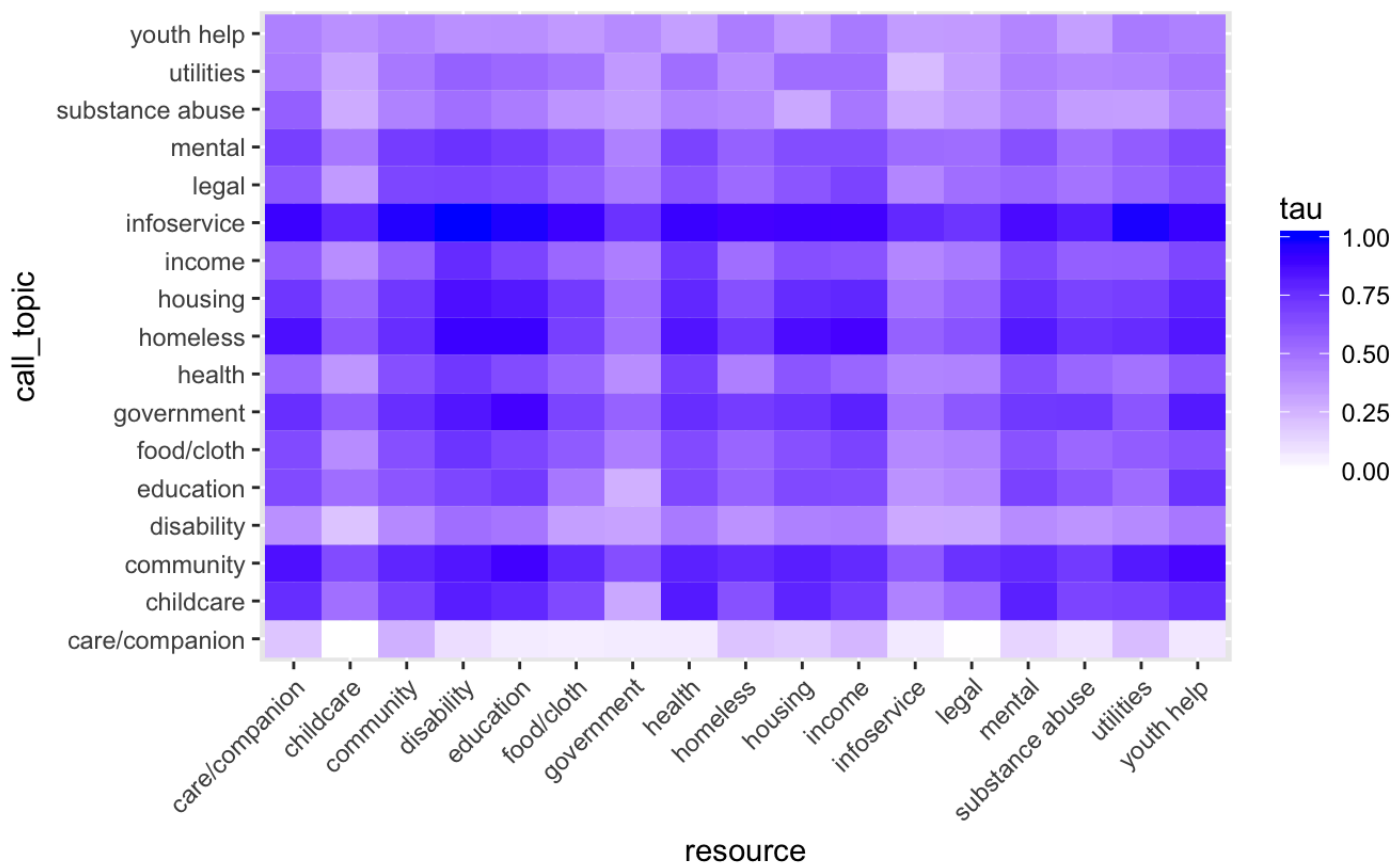
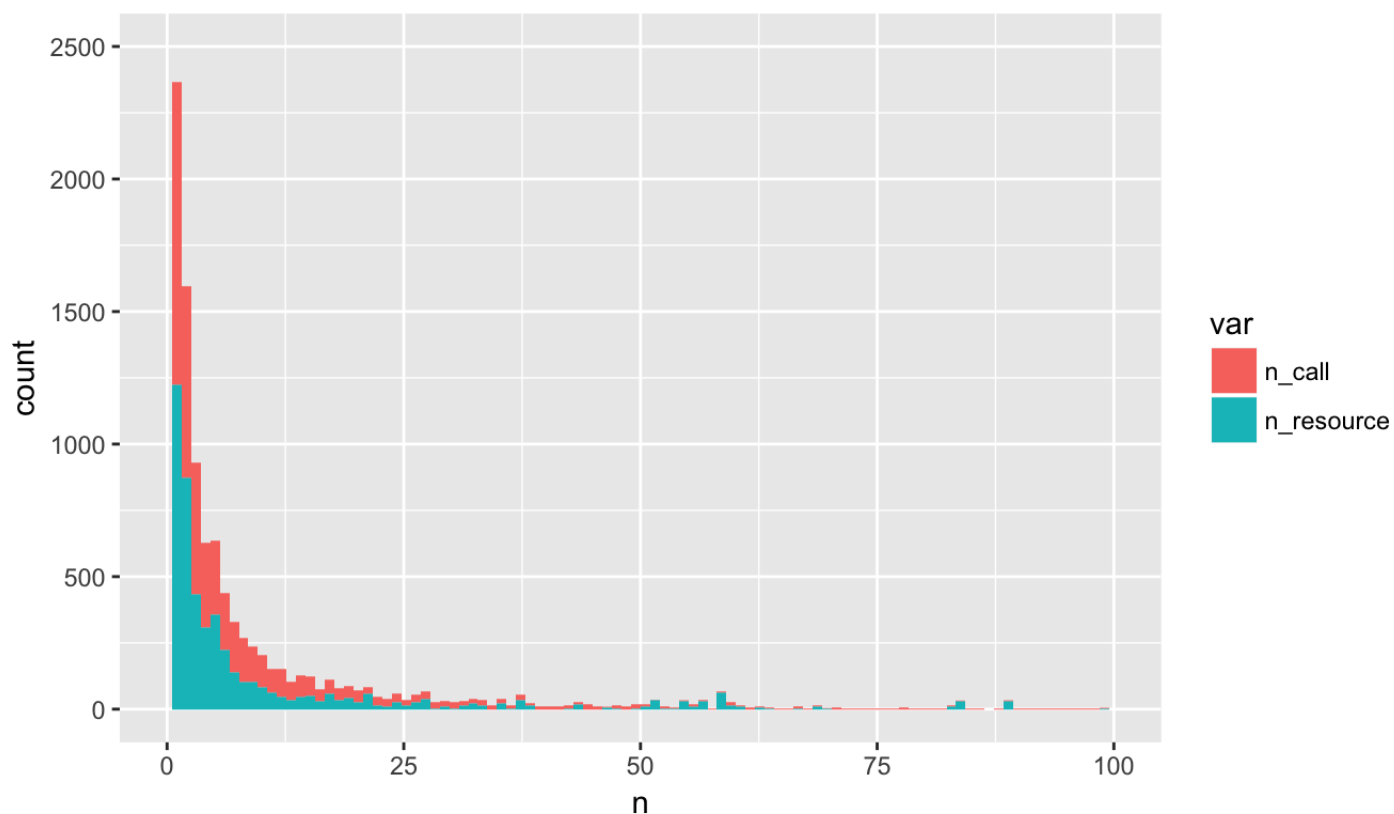
For simplicity, this initial analysis calculates the number of resources available for each municipality and service type, then allocate those resources to every ZIP code associated with that municipality. I.e., every ZIP in Boston will have access to every resource located in Boston. In case of a ZIP code overlaps with two or more municipalities, we will consider the ZIP code have access to resources in all overlapping municipalities.

To get a more robust result, we may need to take proximity into account in the future.

Code

Code

Distribution of Resource and Call Counts For ZIP + Topic



It seems that those more resourceful neighborhoods do make more calls, and it is especially true for info service, homeless, housing, community, and child care related calls.

Conclusion

- Urban areas do get more calls than rural areas, especially in child care and housing.
- University campuses are a totally different story.
- Housing predicts all other types of services.
- Resourceful neighborhoods do make more calls.

-
1. <https://www2.census.gov/geo/pdfs/reference/GARM/Ch12GARM.pdf>
(<https://www2.census.gov/geo/pdfs/reference/GARM/Ch12GARM.pdf>)]↩