Analysis of the effect of housing stock on residential density

Daniel Hartig

December 3, 2017

Motivation

Arlington, Virginia and Cambridge, Massachussets are in many ways very similar. Both of them are directly across a river from the downtown of one of the global cities in the United States. Both of them are densely population; both are the second densest parts of their respective metropolitan areas. Both are also the second largest job hubs in their areas, each with about 200,000 jobs. Cambridge has Harvard and MIT and extensive high tech industries; Arlington the Pentagon and major government offices, international consulting firms, and military-industrial companies.

But despite their similarities, there are some differences. Specifically, Cambridge (along with the adjacent Boston suburbs of Charlestown, Somerville, and Medford to provide a similar land area to Arlington) has significantly higher residential density than Arlington.

	Arlington	Cambridge
Area (km^2)	63	52
Population	223446	261760
Pop Density	5068	3553
Housing Units	109647	114074

This paper will investigate causes for this difference in residential density by looking at the characteristics of housing in the two areas.

Data Description

Various housing characteristics are available from the American Factfinder website (factfinder.census.gov), a service of the US Census Bureau. Two characteristics that we will investigate are housing age and housing type. Age is quantified categorically as decade in which the structure was built; the total number of houses in a geographic area built between 1990 and 1999 or 1940 and 1949, for example. Housing type is the number of units in the housing structure. For example, a housing unit might be a single-family attached home, or a duplex, or an apartment of between ten and nineteen units. The counts provided by the

American Factfinder are estimates drawn from the 2015 American Community survey, and they count the number of housing units. Therefore, a single family home built in 1984 is one count of a housing unit in the single family home type category and one count in the 1980-1989 age category. A 50 unit high rise built in 2005 would count as 50 counts in the twenty-plus apartment type category and 50 counts in the 2000-2009 age category. A summary of the categories is below:

Housing Type	Housing Age
Single Family, Detached	2010-2014
Single Family, Attached	2000-2009
Duplex	1990-1999
3-4 Unit Apartment	1980-1989
5-9 Unit Apartment	1970-1979
10-19 Unit Apartment	1960-1969
20+ Unit Apartment	1950-1959
	1940-1949
	1939 or before

The smallest level of geographical resolution for which the above data is available is the zip code. There are 32633 zip codes in the United States with more than zero housing units; these will be the data points. The feature set is the counts of housing units in the categories listed above. Since we are investigating the effects of housing characteristics on residential density, density will be the response variable for our analysis.

Distribution of housing characteristics

Methodology and Results

Analysis of housing age

Analysis of housing type

Conclusion

In the density range related to inner suburbs of large cities; that is, between 3000 and