

Boston Area Research Initiative ⦁ Northeastern & Harvard Universities  
E-Mail: bari@northeastern.edu ⦁ Web: [www.bostonarearesearchinitiative.net](http://www.bostonarearesearchinitiative.net)

**A Guide to Using the Geographical Infrastructure of Boston, MA 2017**

Overview

The Boston Area Research Initiative’s Geographical Infrastructure for Boston is a database that organizes and links the places and regions of Boston, MA across 17 levels—including land parcels, streets, census geographies, and other administrative regions. The levels are organized in a hierarchy, with the items in each level nested in the higher-level regions that contain it (e.g., land parcels in census geographies). This is coordinated via variables that act as unique identifiers at each level. As a composite, the database is intended to facilitate aggregate calculations across levels of the hierarchy and analyses of data from different sources that reference the same geographical units. In particular, the database makes it possible to connect data sets generated by the City of Boston with census geographies and data.

The levels are each documented in a .csv and in most cases a .shp (shapefile for GIS), and include:

* + - Properties (from the City of Boston Property Assessment Database, 2017)
    - Land parcels (aggregated from the City of Boston Property Assessment Database, 2017)
    - Intersections (from the City of Boston and from TIGER line data, 2013)
    - Street segments (from census TIGER line data, 2013)
    - Census blocks, block groups, and tracts (from census 2010)
    - Ten ways that the City of Boston divides the city into administrative districts for planning, elections, and operations.
    - An ID connector for linking to the Master Address List, a database of addresses used by the City of Boston that we don’t include in this infrastructure.

This documentation contains a section for each of these groupings, describing the contents and variables. For each, the unique identifier variables used to link the files are noted.

**Table of Contents**

1. Properties 4

1.1 Description of Contents 4

1.2 Description of Variables 4

1.2.1 Identifying Characteristics 4

1.2.2 Property and Building Characteristics 5

1.2.3 Geographical Information 5

2. Land Parcels 7

2.1 Description of Contents 7

2.2 Description of Variables 7

2.2.1. Identifying Characteristics 7

2.2.2. Property Characteristics 8

2.2.3. Geographical Information 9

3. Intersections 10

3.1 Description of Contents 10

3.2 Description of Variables 10

3.2.1. Identifying Characteristics 10

3.2.2. Geographical Information 11

4. Roads 12

4.1 Description of Contents 12

4.2 Description of Variables 12

4.2.1. Identifying Characteristics 12

4.2.2. Geographical Information 13

5. Census Geographies 14

5.1 Description of Contents 14

5.2 Description of Variables 14

5.2.1. Identifying Characteristics 14

5.2.2. Geographical Information 15

6. Other Geographies 17

6.1 Description of Contents 17

6.2 Description of Variables 17

6.2.1 BRA Neighborhood Statistical Areas 17

6.2.2 BRA Planning Districts 17

6.2.3 City Council Districts 17

6.2.4 Election Precincts 17

6.2.5 Election Wards 18

6.2.6 Fire Districts 18

6.2.7 ISD Neighborhoods 18

6.2.8 Police Districts 18

6.2.9 PWD Districts 18

6.2.10 ZIP Codes 18

7. ID Connector 19

7.1 Description of Contents 19

APPENDIX A. Codes for Land Use 20

APPENDIX B. Codes for Road Types 21

APPENDIX C. Characteristics of Clusters of Road Segments 23

1. Properties

1.1 Description of Contents

The City of Boston’s Assessing Department is responsible for determining accurate values for all properties in the city. To this end the Department maintains parcel ownership and value information to ensure fair assessment of both taxable and non-taxable property in Boston. Assessing records are compiled and reviewed annually to reflect changes to parcels as a result of new construction, remodeling, and changes in ownership. This forms the City of Boston Property Assessment Database, which acts as the fundamental dataset for BARI’s geographical infrastructure.

The data contained herein are a modified version of the original Property Assessment Data Set, and describe the parcel-specific address, ownership, and land use for the 168,181 properties in Boston. They include variables and measures from the original data, some of which are cleaned or modified by BARI, as well as derived measures based on original variables.[[1]](#footnote-1)

No shapefile is included for Properties because the polygons associated with the City of Boston’s Property Assessment Database are Land Parcels. Each property exists on only one land parcel from the Land Parcels database, but there may be multiple properties on a Land Parcel. The information necessary to split the Land Parcels appropriately into properties is not available, and properties may exist on top of one another on different floors anyway. Note that the definition of Land Parcel in this database is slightly different from that of the City of Boston’s original data files in that we combine a small number of land parcels that are differentiated by the City (see Section 2 for more detail).

**Unique** **identifier:** *parcel\_num*

1.2 Description of Variables

Property variables are split into three categories: identifying characteristics, property and building characteristics, and geographical information. Identifying characteristics include variables regarding the basic identity and attributes of the address. Building characteristics include information on the physical attributes of the building containing the property. Geographical information provides further detail on the location of the property and the other geographies that contain it.

The Tax Assessor’s department maintains a large number of property and building characteristic variables that go beyond its geographic utility. For this reason, we only include a small subset in the Geographical Infrastructure. For access to the full breadth of variables and their documentation see BARI’s Property Assessment Database.

1.2.1 Identifying Characteristics

* + - *parcel\_num* is the 10-digit property identification number, unique to each property. The first two digits indicate the Ward, digits 3 thru 7 are the parcel, and digits 8 thru 10 are the sub-parcel.
    - *CM\_ID* is the 10-digit property number of the main condo building property. All condo units in each building are related to this number.
    - *GIS\_ID* is another 10-digit property identification number. It is the unique identifier for the land the parcel is in (this is slightly different than *Land\_Parcel\_ID*, however, as the latter combines some land parcels into one)
    - *ST\_NUM* is the street number of the property.
    - *ST\_NAME* is the street name of the property.
    - *ST\_NAME\_SUF* is the suffix of the street name. This variable contains two-character short-forms of each type of suffix (St, Av, BL, PL, etc…)
    - *UNIT\_NUM* is the specific unit number within a multi-unit building.
    - *ZIPCODE* is the zip code of the property.

1.2.2 Property and Building Characteristics

* + - *LU* is the Land Use type for the property. Codes for land use can be found in Appendix A.
    - *OWN\_OCC* is a one-character code that indicates if an owner receives a residential exemption for the property. A "Y" indicates that the owner claims to live within the property (a.k.a. the property is "owner-occupied") and a "N" indicates the opposite.
    - *YR\_BUILT* is the year in which the property was built. The original dataset held many properties whose year of construction was listed as zero. It was fixed by updating the YR\_BUILT variable, which now contains a "NA" value where it previously showed a "0".
    - *YR\_REMOD* is the year in which the property was last remodeled. For some properties the year of its most recent remodel was listed as zero. It was fixed by updating the YR\_REMOD variable, which now contains a "NA" value where it previously showed a "0".
    - *LAND\_SF* is the total size of the property in square feet. This is also known as the lot size.
    - *GROSS\_AREA* is the gross floor area for commercial properties.
    - *NUM\_FLOORS* is the number of levels in the structure that is located on the property.

1.2.3 Geographical Information

Geographical information for the properties comes from multiple sources in order to coordinate them with census geographies. First, latitude and longitude of the property was derived as the centroid of the containing land parcel, based on the .shp provided by the City of Boston (see Section 2). Properties were then linked to the appropriate census TIGER line road segment, defined as the nearest road segment to the x-y coordinate that matched on street name. This was done instead of geocoding because City of Boston land parcels will be more accurate about the location of numbers along the segments of a street than census TIGER line data. Last, all properties were linked to the containing census block and higher geographies by spatial overlay.

* + - *X* is the geo-coded longitude of the property.
    - *Y* is the geo-coded latitude of the property.
    - *Land\_Parcel\_ID* is the unique identifier for the land parcel the property is in.
    - *TLID* is the identifier for the census TIGER line road segment containing the property.
    - *Blk\_ID\_10* is the 2010 census Block ID number
    - *BG\_ID\_10* is the 2010 census Group ID number
    - *CT\_ID\_10* is the 2010 census Tract ID number
    - *NSA\_NAME* is the name of the Inspectional Service Department Neighborhood Statistical Area in which the building is located.
    - *BRA\_PD* is the name of the Boston Redevelopment Authority Planning District in which the building is located.

2. Land Parcels

2.1 Description of Contents

This dataset contains the unique land parcels present in the City of Boston’s Property Assessment Database (see Section 1). It was constructed by merging all properties that match on *GIS\_ID* or on street address. Then, matches were split if the properties within them were more than 16 meters apart from one another. Note that the first step may combine certain GIS\_IDs into a single Land Parcel. This is important because the City of Boston considers every GIS\_ID as its own “land parcel,” but for the purposes of this database we have combined a small number of these into a single analytic unit. This is because without additional information, two land parcels with the same address would be impossible to differentiate. We split land parcels with the same address that were too far away from one another in order to avoid faulty matches based on typos in the street address, but only if they had different *GIS\_ID* values. This processing reduced the number of land parcels from 99,191 in the City’s original database to a final set of 98,355 unique land parcels. In addition to a CSV, we include a shapefile, which is the shapefile maintained by the City of Boston for its Property Assessment Database, but with polygons merged in the manner described above.

**Unique** **identifier:** *Land\_Parcel\_ID*

2.2 Description of Variables

Land parcel variables are split into three categories: identifying characteristics, property characteristics, and geographical information. Identifying characteristics include variables regarding the basic identity and attributes of the address. Building characteristics include information on the physical attributes of the buildings contained by the land parcel. Geographical information provides further detail on the location of the parcel and the other geographies that contain it.

2.2.1. Identifying Characteristics

* + - *Land\_Parcel\_ID* is the unique identifier for land parcels.
    - Street address: the street address is split into four parts: lower number, upper number, street, and street suffix. It contains an upper and lower number to allow for a range of addresses along a street. There are two sets of these four parts, allowing for street addresses on two streets for a single land parcel, with the different sets differentiated with a numerical suffix, here denoted as “X”. The dataset permits for multiple addresses housed within the same land parcel. This is particularly pertinent when a land parcel stands at a corner or takes up an entire block, and consequently has street addresses on multiple streets. For this reason, the multiple TLIDs can be attributed to a single land parcel as well.
      * *minNum\_X* designates the lower number in the street address, if the street address is a range of numbers, or just the sole number
      * *maxNum\_X* designates the higher number in the street address, if the street address is a range of numbers, or it is NA
      * *street\_X* designates the street name in the street address
      * *suffix\_X* designates the street name suffix in the street address
    - *zip* is the zip code that the land parcel is located in. It is calculated by taking the modal zip codes of all of the properties in the land parcel.

2.2.2. Property Characteristics

* + - *num\_properties* is the total number of properties in the land parcel.
    - *LU* is the designated land use of the address.
      * *Note:* If a land parcel contained properties of more than one land use type, the listed land use type is the modal non-empty land use. For more detail on these codes see Appendix A.
    - *numUnits* is the estimated number of units in the parcel. It is calculated based on the designated land usage and the number of aggregated properties. Parcels that are designated R1, R2, R3, R4, and A, are assigned 1, 2, 3, 4, and 7, respectively. Values for parcels that are designated CD, CC, CM, and CP are calculated by counting the number of properties zoned CD or CC within the parcel.
    - *OWN\_OCC* is the proportion of properties in the land parcel that are occupied by their owner
    - *AV\_LAND* is the valuation of the land that the land parcel is on. This is calculated by summing the land valuations of all of the properties that are aggregated to make the land parcel record.
    - *AV\_BLDG* is the valuation of the building. This is calculated by summing the building valuations of all of the properties that are aggregated to make the land parcel record.
    - *LAND\_SF* is the size of the land parcel in square feet. This is also known as the lot size.
    - *YR\_BUILT* is the year that the land parcel was built. It is calculated by taking the mode of year built for the properties in the land parcel, with NAs not included in the modal calculation.
    - *YR\_REMOD* is the year that the land parcel was remodeled. It is calculated by taking the mode of year remodeled for the properties in the land parcel, with NAs not included in the modal calculation.
    - *GROSS\_AREA* is the total floor area of the land parcel. It is calculated by summing the floor areas of the different properties in the land parcel.
    - *IL\_RATIO* is the improvement to land value ratio. It is calculated by dividing the land valuation by the building valuation.
    - *NUM\_FLOORS* is the number of floors in the buildings on the land parcel. It is calculated by finding the maximum number of floors of the properties aggregated into the land parcel.
    - *R\_BLDG\_STYL* is the building style for residential properties. The styles are: BL for Bi-Level, BW for Bungalow, CL for Colonial, CN for Contemporary, CP for Cape, CV for Conventional, DK for Decker, DX for Duplex, L for Tri-Level, Oth for Other, RE for Row End, RM for Row Middle, RN for Ranch, RR for Raised Ranch, SL for Split Level, TF for Two-Family Stack, TD for Tudor, SD for Semidetached, and VT for Victorian It is calculated by taking the mode of the building styles of the properties in the land parcel.
    - *owner\_address* is the street address of the owner of the land parcel. It is calculated by taking the mode of the owner’s street address for all properties in the land parcel.

2.2.3. Geographical Information

Because a land parcel contains multiple properties, and in some cases is the merger of multiple *GIS\_ID*s from the original City of Boston database, the geographical information here was, when necessary, aggregated from land parcels as the modal value. The choice of mode rather than mean ensures the placement in a single polygon at each higher order geography.

* *X* is the X coordinate of the land parcel, in longitude.
  + It is calculated by taking the modal X, Y pair among aggregated properties.
* *Y* is the Y coordinate of the land parcel, in latitude.
  + It is calculated by taking the modal X, Y pair among aggregated properties.
* *TLID\_X* is the unique identifier for the road segment containing the land parcel.
  + There can be two of these (denoted by a numerical suffix, here represented by “X”), because a single land parcel can take up an entire block or a corner and thus exist on multiple streets. See *Street Address* in Section 3.2.1.
* *Blk\_ID\_10* is the unique identifier for the census block in which the land parcel is located. It is calculated by taking the modal *Blk\_ID\_10* of properties in the land parcel.
* *BG\_ID\_10* is the unique identifier for the census block group in which the land parcel is located. It is calculated by taking the modal *BG\_ID\_10* of properties in the land parcel.
* *CT\_ID\_10* is the unique identifier for the census tract in which the land parcel is located. It is calculated by taking the modal *CT\_ID\_10*  of properties in the land parcel.

3. Intersections

3.1 Description of Contents

Intersections comprise two datasets. One is a catalog of all intersections (*Intersections*; 11,116) generated by the census TIGER line roads data from 2013. The other is a master list of intersections maintained by the City of Boston (*Intersections* *(City)*; 8,421) that has then been merged with intersections within the census TIGER line road map (to the same intersection name when possible, otherwise to the closest point).

**Unique** **identifiers:** *ObjectID* (census intersections), *propid* (City intersections).

3.2 Description of Variables

Variables are largely equivalent between the two files. The two datasets can be connected via their unique identifiers. They are split into two categories: identifying characteristics and geographical information. Identifying characteristics includes variables regarding the basic identity and attributes of the intersection. Geographical information provides further detail on the location of the intersection and the other geographies that contain it.

3.2.1. Identifying Characteristics

* + - *ObjectID* is the unique identifier for census-generated intersections.
    - *Propid* is the unique identifier for City-defined intersections.
    - *Address*, indicating the two roads that form the intersection.
    - *TLID* is the unique identifier for the road segment to which the intersection is attributed.
      * *Note:* Intersections are attributed to the road that forms the intersection that is most predominant in terms of main/non-main street and zoning.
    - *X* gives the x-coordinate of the intersection
    - *Y* gives the y-coordinate of the intersection
    - *Main* indicates whether the primary road segment is a main street or not.
    - *RoadType* gives the zoning status of the road. See *Zoning* in *Roads* (Section 4).
      * *Note:* Possible values for this are Commercial, Residential, Exempted, Independent, or None.

3.2.2. Geographical Information

* + - *BG\_ID\_10* is the unique identifier for the census block group in which the intersection is located.
    - *CT\_ID\_10* is the unique identifier for the census tract in which the intersection is located.

4. Roads

4.1 Description of Contents

This dataset contains a complete list of all road segments in Boston, MA, as defined by census TIGER Line data as of the 2013 update. It contains 24,891 segments.

**Unique** **Identifier:** *TLID*

4.2 Description of Variables

Roads variables are split into two categories: identifying characteristics and geographical information. Identifying characteristics includes variables regarding the basic identity and attributes of the address. Geographical information provides further detail on the location of the road and the other geographies that contain it. Because roads often form the borders between regions, nesting in higher levels was done using the road centroids.

4.2.1. Identifying Characteristics

* + - *STATEFP* is the unique identifier (FIPS code) for the state containing the road segment (25 = Massachusetts).
    - *COUNTYFP* is the unique identifier (FIPS code) for the county containing the road segment (025 = Suffolk).
    - *TLID* is the unique identifier for the road segment.
    - *TFIDR* is the unique identifier for the right side of the road segment.
    - *TFIDL* is the unique identifier for the right side of the road segment.
    - *MFTCC* is a code provided by the census for the type of road. These types are listed in Appendix B.
    - *FULLNAME* is the name of the road.
    - *LFROMADD,* *LTOADD,* *RFROMADD,* and *RTOADD* give the address ranges for the left and right sides of the road.
    - *ZIPL* and *ZIPR* is the zipcode containing the left and right sides of the road respectively.
    - *Length* is the length of the segment in meters.
    - *CLASS* describes the nature of the road, as provided by Massachusetts Dept. of Transportation.
      * *Note:* Values taken: 1 – Limited access highway; 2 – Multi-lne highway, not limited access; 3 – Other numbered route; 4 – Major road – arterials and collectors; 5 – Minor street or road with road inventory information; 6 – Minor street or road with no road inventory information.
    - *RDTYPE* is an extended version of the *CLASS* variable, provided by Massachusetts Dept. of Transportation.
      * *Note:* 1-6 remain the same. 7 – Ramp; 8 – Tunnel; 9 – Tunnel for limited access highway; 10 – Tunnel for a multi-lane highway, not limited access; 11 – Tunnel for other numbered route.
    - *CLUSTER* indicates the cluster number to which the segment was attributed, from a cluster analysis based on the zoning characteristics of the parcels on the street. More details on the clusters are available in Appendix C.
    - *Main* indicates if a road segment is considered part of a main street (“1” = Main) o *Note:* Based on MassGIS’ *Class*, with all classes less than 5 denoted as main streets.
    - *Zoning* gives the zoning status of the road.
      * *Note:* Possible values include: Commercial, Residential, Exempted, Independent, or None.

4.2.2. Geographical Information

* + - *BG\_ID\_10* is the unique identifier for the census block group (2010-present) in which the segment’s centroid is located (see Section 5).
    - *CT\_ID\_10* is the unique identifier for the census tract (2010-present) in which the segment’s centroid is located (see Section 5).

5. Census Geographies

5.1 Description of Contents

As of the 2010 census, Boston, MA contains 7,288 city blocks, defined as any piece of land bounded by streets or water on all sides, and not divided by any streets or water. These are nested in 558 block groups, which are themselves nested in 178 census tracts. Census geographies are a natural hierarchy, with the unique identifier at each level being an extension of the level above it, indicating the specific region and all of the higher-order regions that contain it.

5.2 Description of Variables

Census variables are split into two categories: identifying characteristics and geographical information. Variables are consistent across all three census levels, though some variables appear in one file and not in others. Identifying characteristics includes variables regarding the basic identity and attributes of the block. Geographical information provides further detail on the location of the block and the other geographies that contain it. Nesting started with the block level, because all geographies, census or otherwise, conform to block boundaries. Linking to non-census geographies was done by identifying the location of the centroid of each block.

Because block groups and tracts do not nest perfectly within non-census geographies, this linking was done by identifying the region that contained the plurality of the block group or tract’s land area. Some non-census geographies crossed over census the boundaries of a particular level too often, however, to make such linkages reliable, and thus were omitted: for block groups, *Precincts* were not linked; for tracts, *Precincts*, *Neighborhood* *Statistical* *Areas*, and *ZIP* *Codes* were not linked.

Demographic and socioeconomic data about census geographies can be found online in our Dataverse, at https://dataverse.harvard.edu/dataverse/Massachusetts\_Census\_Indicators.

**Unique** **Identifiers:** Block: *Blk\_ID\_10*; Block groups: *BG\_ID\_10*; Tracts: *CT\_ID\_10*

5.2.1. Identifying Characteristics

* + - *STATEFP* is the unique identifier (FIPS code) for the state containing the region (25 = Massachusetts).
    - *COUNTYFP* is the unique identifier (FIPS code) for the county containing the region (025 = Suffolk).
    - *TRACTCE10* is the code identifying the census tract.
    - *BLOCKCE10* is the code identifying the block
    - *GEOID10* is the unique identifier (FIPS code) of the region.
    - *NAME10* the block name, within the tract.
    - *MTFCC10* a class code indicating the type of feature.
    - *ALAND10* is the land area of the region (in sq. meters).
    - *AWATER10* is the water area of the region (in sq. meters).
    - *INTPTLAT10* is the latitudinal coordinate of the region.
    - *INTPTLON10* is the longitudinal coordinate of the region.
    - *POP100\_RE* is the population of the region as of 2010.
    - *HU100\_RE* is the number of housing units in the region as of 2010.
    - *Type* describes the type of neighborhood the region is within.
      * *Note:* Possible values are Residential, Downtown, Institutional (e.g., industrial, college campuses), and Park. Only for block groups and tracts.
    - *Res* indicates whether the block group is generally a residential area (based on *Type*; “1” = yes).

5.2.2. Geographical Information

* + - *Blk\_ID\_10* is the unique identifier for the block (identical to *GEOID10* in the block file, but compatible with the other levels).
    - *BG\_ID\_10* is the unique identifier of the census block group (identical to *GEOID10* in the block file, but compatible with the other levels).
    - *CT\_ID\_10* is the unique identifier of the census tract (identical to *GEOID10* in the block file, but compatible with the other levels).
    - *BOSNA\_R\_ID* is the numerical unique identifier for the Boston Redevelopment Authority Neighborhood Statistical Area in which the region is located (see Section 7).
    - *NSA\_NAME* is the name of the Boston Redevelopment Authority Neighborhood Statistical Area in which the region is located (see Section 7).
    - *BRA\_PD\_ID* is the unique identifier for the Boston Redevelopment Authority planning district in which the region is located (see Section 7).
    - *BRA\_PD* is the name for the Boston Redevelopment Authority planning district in which the region is located (see Section 7).
    - *ZIPCODE* refers to the ZIP code in which the region is located (see Section 7).
    - *City\_Counc* is the unique identifier for the city council district in which the region is located (see Section 7).
    - *WARD* is the unique identifier for the election ward in which the region is located (see Section 7).
    - *PRECINCTS* is the unique identifier for the election precinct in which the region is located (see Section 7).
    - *ISD\_NAME* is the name for Boston Inspectional Services Department neighborhood in which the region is located (see Section 7).
    - *Police\_Dis* is the unique identifier for the police district in which the region is located (see Section 7).
    - *Fire\_Distr* is the unique identifier for the fire district in which the region is located (see Section 7).
    - *PWD* contains a numerical unique identifier for Public Works districts, followed by the name of the Public Works District in which the region is located (see Section 7).

6. Other Geographies

6.1 Description of Contents

Other geographies were provided by the City of Boston. They include: traditional neighborhoods defined by the Boston Redevelopment Authority (*BRA* *Neighborhood* *Statistical* *Areas,* *BRA* *Planning* *Districts*); election board regions (*City* *Council* *Districts,* *Election* *Precincts,* *Election* *Wards*); and districts for City operations (*Fire* *Districts,* *ISD* *Neighborhoods,* *Police* *Districts,* *Public* *Works* *Districts,* and *ZIP* *Codes*)*.* For each, only a shape file with the unique identifiers is included. The unique identifiers are listed below.

6.2 Description of Variables

6.2.1 BRA Neighborhood Statistical Areas

* + - *BOSNA\_R\_ID* is the numerical unique identifier.
    - *ID* an additional identifier used by the BRA.
    - *NSA\_NAME* the name of the neighborhood statistical area.

6.2.2 BRA Planning Districts

* + - *ID* is the numerical unique identifier. Denoted as *BRA\_PD\_ID* in lower-level files.
    - *PD* is the name of the planning district. Denoted as *BRA\_PD* in lower-level files.

6.2.3 City Council Districts

* + - *DISTRICT* is the unique identifier. Denoted as *CITY\_COUNC* or *CITY\_COUNCIL* in lower-level files.
    - *Councillor* is the individual holding the seat as of the 2013 election.

6.2.4 Election Precincts

* + - *PRCNTS\_ID* is the unique identifier. Denoted as *PRECINCT* in lower-level files.
    - *WDPCT* is a concatenation of the ward number and the precinct number within the ward.
    - *PCT* is a count within the precinct, making it non-unique.

6.2.5 Election Wards

* + - *WARD* is the unique identifier.
    - *CNT\_WARD* is the number of precincts in each ward.

6.2.6 Fire Districts

* + - *DISTRICT* is the unique identifier. Denoted as *Fire\_Distr* in lower-level files.

6.2.7 ISD Neighborhoods

* + - *Name* is the unique identifier.

6.2.8 Police Districts

* + - *ID*
    - *DISTRICT\_* is the unique identifier for the police district of the block. Denoted as *Police\_Dis* in lower-level files.

6.2.9 PWD Districts

* + - PWD
    - NAME
    - COMBO contains a numerical unique identifier for Public Works districts, followed by the name of the Public Works District. Denoted in lower-level files as PWD.

6.2.10 ZIP Codes

* + - *ZIP5* is the zip code. Denoted as *ZIPCODE* or *ZIP* in lower-level files.

7. ID Connector

7.1 Description of Contents

Previous to the 2016 geographical infrastructure, BARI used the City of Boston’s Master Address List as the fundamental file for the geographical infrastructure. The Master Address List is described as containing “all property addresses in the city.” Starting in the 2016 geographical infrastructure, we decided to use the Property Assessment Database as the fundamental file instead. We chose not to use the Master Address List because we found a number of redundancies and inconsistencies in the data, which created an inaccurate picture of the physical infrastructure of Boston. It appeared that the dataset had been added to in certain places without purging old data, resulting in the same property represented in multiple ways and inconsistency between properties. By contrast, the Tax Assessor’s Database covered the same information in a highly regular and consistent way. Starting in 2017, the City of Boston has made the Master Address List a legacy dataset and appears to be phasing it out of use.

The difficulty with removing the master address list from our geographical infrastructure is that many data systems in Boston, such as building permits, make use of the Master Address List’s unique ID, *Property*\_*ID*, to connect their data to an address. Even if the City of Boston switches to using the Property Assessment database completely, old datasets will make use of the Master Address List’s unique ID. To facilitate use of any geographies that have only *Property\_ID*, we have included an ID Connector, a .csv file with *Property\_ID*, *parcel\_num* (the unique ID for properties), *Gis\_ID,* and *Land\_Parcel\_ID* (the unique ID for land parcels) for all instances of all four IDs. There are 352,049 rows and the four IDs are nested within one another, following an m:1, m:1, m:1 relationship, although in some cases there are *parcel\_num* IDs without a corresponding *Property\_ID*, and in some cases a *Property\_ID* without a corresponding *parcel\_num*.

APPENDIX A. Codes for Land Use



APPENDIX B. Codes for Road Types

|  |  |  |
| --- | --- | --- |
| MTFCC | Feature Class Full Name | Count |
| H1100 | Connector (Hydrography) | 198 |
| H3010 | Stream/River | 12,855 |
| H3020 | Canal, Ditch or Aqueduct | 520 |
| L4010 | Pipeline | 7 |
| L4020 | Powerline | 246 |
| L4110 | Fence Line | 48 |
| L4130 | Point-to-Point Line (Miscellaneous Linear) | 4 |
| L4140 | Property/Parcel Line (Including PLSS) | 296 |
| P0001 | Nonvisible Linear Legal/Statistical Boundary | 11,393 |
| P0002 | Perennial Shoreline | 8,441 |
| P0004 | Other non-visible bounding Edge (e.g., census water boundary, boundary of an areal feature) | 2,584 |
| R1011 | Railroad Feature (Main, Spur, or Yard) | 2,269 |
| R1051 | Carline, Streetcar Track, Monorail, Other Mass Transit Rail) | 46 |
| S1100 | Primary Road | 1,409 |
| S1200 | Secondary Road | 8,543 |
| S1400 | Local Neighborhood Road, Rural Road, City Street | 89,971 |
| S1500 | Vehicular Trail (4WD) | 47 |
| S1630 | Ramp | 2,070 |
| S1640 | Service Drive usually along a limited access highway | 59 |
| S1710 | Walkway/Pedestrian Trail | 162 |
| S1740 | Private Road for service vehicles (logging, oil, fields, ranches, etc.) | 675 |
| S1750 | Private Driveway | 5 |

|  |  |  |
| --- | --- | --- |
| S1780 | Parking Lot Road | 16 |
| S1820 | Bike Path or Trail | 1 |

APPENDIX C. Characteristics of Clusters of Road Segments

**Generated by a Cluster Analysis on Parcel Zoning Characteristics.**

|  |  |  |
| --- | --- | --- |
| **Cluster** | **Number of segments** | **Primary Zoning Characteristics** |
| 1 | 228 | Public housing |
| 2 | 686 | Industrial; Residential and commercial lots |
| 3 | 1,452 | Condos |
| 4 | 905 | Commercial |
| 5 | 573 | Residential-Commercial |
| 6 | 1,838 | Residential: 3-Family |
| 7 | 1,494 | Residential: Single-Family |
| 8 | 741 | Apartment buildings |
| 9 | 1,285 | Mixed residential: Single-, Two-, and Three-Family |
| 10 | 1,388 | Exempt (i.e. parks, churches, schools) |
| 11 | 1,414 | Mixed residential: Single-family with two-family |
| 12 | 1,298 | Mixed residential: Two-Family with single-family |
| 13 | 303 | Residential: 4-6 units |
| 14 | 165 | Unused |

1. Tax rate calculation information published by the City of Boston Assessing Department through the department’s website: http://www.cityofboston.gov/assessing/taxrates.asp [↑](#footnote-ref-1)