“Age-Friendly” Scoring in Boston

Jacqueline You
Boston University
CS 591 L1 Data Mechanics, Spring 2016

Summary

This project attempts to create a distance-based score for a given location in Boston to quantify its “age-friendliness.” Location data, including longitude/latitude and address, of resources including hospitals, MBTA stops, and community centers are incorporated into a weighted distance score. Currently the scoring system is not sophisticated enough to incorporate all major factors that might designate a location as age-friendly, especially financial data, but it could serve as the basis for a more complex scoring system/set of metrics.

Introduction

In 2010, over 14% of Boston’s residents (about 88,000 individuals) were 60 years or older. Projections from the City of Boston’s 2014 report on ”Aging in Boston” indicate that by 2030, about 20% of Bostonian residents will be age 60 or older [Center for Social and Demographic Research on Aging, 2014]. For comparison, 2010 Census results indicated that the percentage of persons aged 65 and older in Boston was 10.1% - this figure is lower than in New York City (12.1%), San Francisco (13.6%), and Miami (16.0%), but higher than in Austin (7.0%) and Atlanta (9.8%) [Census Bureau, 2010].

Boston is actively committed to age-friendliness in its communities. Boston is a member of the World Health Organization's Age Friendly Cities Network [Mayor's Office, City of Boston, 2014]. The Milken Institute, an American economic think tank, ranked Boston fourth on its 2014 list of "Best Cities for Successful Aging" in the US, citing factors such as its high-quality healthcare, cultural vibrancy, opportunities for community engagement, and public transportation [Chatterjee, King, and Irving, 2014].

Some of these factors are important to older adults - for example, the 2015 United States of Aging Survey, conducted by the National Council on Aging, found that concerns among adults aged 60 and older and the professionals who worked with them included physical health, affordable housing, and mental wellbeing [2015]. Other concerns, brought up by the Aging in Boston report, include greater access to social services and improved transportation access [Center for Social and Demographic Research on Aging, 2014].

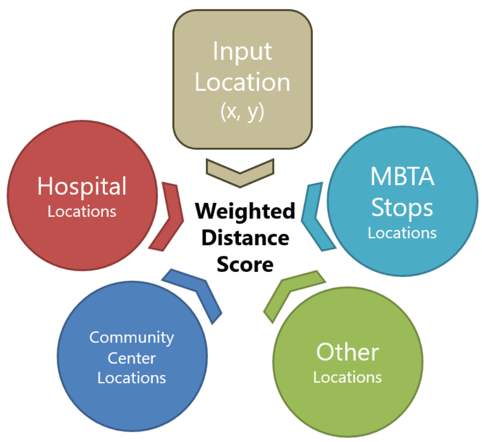
Some metrics developed by private and public entities measure factors like transportation access, retail access, and general walkability that might be of interest to the older population. However, these are usually general measures for either the general public or consumers, without a specific target audience/population. They include the Greater London Authority's Public Transport Accessibility Levels (PTAL) [Transport for London, 2015]; Walk Score's "walkability" service [Walk Score 2016]; and the Location Affordability Index's Retail Access Index produced under a joint initiative by the US Department of Housing and Urban Development, the Department of Transportation, and the Environmental Protection Agency [U.S. Department of Housing and Urban Development].

In this project, a rudimentary distance-based scoring system was developed to assess how "age-friendly" a location is based on a single input latitude/longitude (chosen at random from an area in Boston). The score/metric relies on the proximity to multiple services, such as hospitals, MBTA stops, and community centers - the higher the score, the less "age-friendly" (due to a greater aggregate distance from services) the location is.

Methods

Methods Overview

A number of potential factors were considered in creating the score, including distance to nearest pharmacy, distance to nearest grocery store, distance to nearest park, distance to cultural sites, and median housing/rental prices for the zip code/census block group. Due to time constraints, three factors ultimately were chosen for inclusion in the scoring system - distance to the nearest hospital, distance to the nearest community center, and median distance to an MBTA stop.



For an input location with latitude lat and longitude long, the distance score can be represented on a high level as a sum of subscores as follows:

$$d(\text{lat, long}) = a{*}hospitals(\text{lat, long}) + b{*}mbta(\text{lat, long}) + c{*}commcenters(\text{lat, long})$$

The code currently uses hard-coded constants for a, b, c, and generates lat and long randomly from a grid of space representing a large portion of Boston. Note that the larger a, b, or c is, the more heavily that factor "penalizes" a location. An explanation of the subscores for *hospitals*, *mbta*, and *commcenters* are explained in the subsections **Hospitals**, **MBTA Stops**, and **Community Centers** below respectively.

The relevant code produced for this project proceeds as follows:

1. a list of hospitals is obtained from the City of Boston website and cleaned up accordingly with the scripts `retrievehospitals.py` and `cleanhospitals.py`. The resulting dataset was stored in a MongoDB collection called `hospitals`.

2. MBTA data on stops (bus, subway, etc.) are obtained from the MBTA website and cleaned to produce a collection of stops with stop location as well as wheelchair access information, via script `inputmbta.py`. The resulting dataset was stored in a MongoDB collection called `mbtaStops`.

3. Community centers obtained from the City of Boston's website as providing services for senior citizens was manually assembled into a json file. This json file is combined with geocoordinate data obtained from the OpenCage Geocoder by the script `retrieveservices.py` and stored in the collection `servicecenters`.

4. The script `scorecoordinates.py` is one of the central scripts, as it generates a series of input locations (longitudes and latitudes) and calls on several helper functions from other scripts to calculate scores. These functions are explained in the subsections **Hospitals**, **MBTA Stops**, and **Community Centers** that follow. The resulting data is output to `scores.json` which is indirectly used to generate a scatterplot visualization as well as `scoresgeo.json` which is used to generate a map visualization.

5. Supplemental files are created for use in visualizations, as part of `scorecoordinates.py` and `makescorecsv.py`.

Hospitals

Using a function in `findnearest.py`, the walking distance from the input location to each hospital in the `hospitals` collection is found using MapQuest's API. The minimal distance is used as the subscore for hospitals.

MBTA Stops

Using a function in `findnearest.py`, the set of MBTA stops within a certain radius of the input location is collected from the collection `mbtaStops`, along with geocoordinate and wheelchair accessibility information about the stop.

The vincenty distance (which uses latitude/longitude data to determine distance) between the input location and each MBTA stop was then weighted based on the presence or absence of wheelchair accessibility facilities. The median of these values was chosen as the MBTA subscore.

Community Centers

The functions to calculate the output related to community centers are the same as those of the hospitals, except they are applied to the collection `servicecenters` instead to produce the community centers' subscore.

Miscellaneous

`generatemiscprov.py` is a supplementary script that generates provenance data for the manually assembled data - specifically, a list of zip codes in Boston and the list of community centers in Boston.

Certain scripts were created that did not contribute to the results. These scripts are explained in the Appendix section at the end of this report.

Results

A map of the input locations, along with the locations of certain services, was created using Leaflet - Figure 1 is a screenshot of the visualization. Examining the map, a qualitative spread of resources can be easily observed - services tend to be closer to downtown Boston. The score is possibly biased towards densely populated areas closer downtown where resources are more abundant but living expenses are also higher (in other words, a factor not considered to be "age-friendly"), although further quantitative analysis would be required to support this hypothesis.

The input locations were converted to addresses using the OpenCage Geocoder API to obtain approximate address information as well as zip code information. The score for each location was then plotted against the Location Affordability Index's Retail Access Index to produce the interactive d3 plot shown in Figure 2. The Retail Access Index (RAI), a proxy for the relative presence of retailers in an area, is the number of retail jobs in an area block groups divided by the squared distance of block groups [U.S. Department of Housing and Urban Development]. The distance score appears to

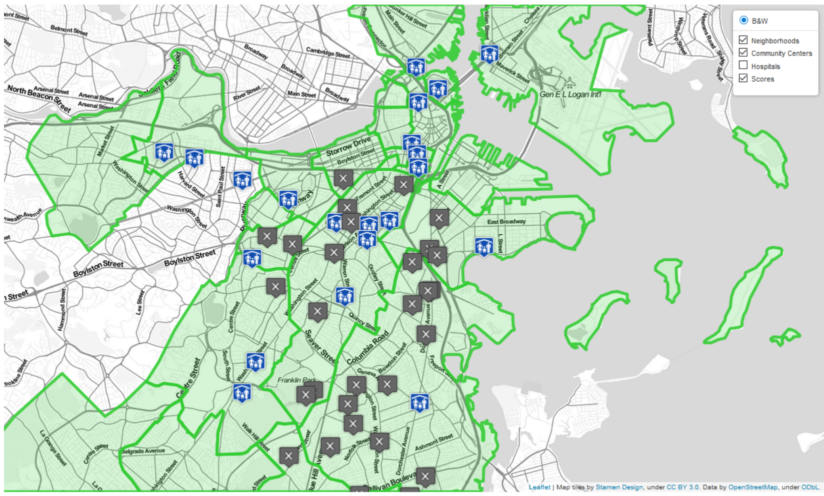


Figure 1: Sample map with community center locations and selected input points.

be inversely proportional to the RAI which is consistent with expectations. There does not seem to be a distinct pattern in terms of zip code, potentially due to the weighing system used in the score or due to the fact that input locations do not currently cover locations such as Allston/Brighton and Charlestown.

Geocoding proved to be a general issue encountered during the development of this scoring system - latitude and longitude coordinates do not always convert cleanly to an exact address due to limitations of the OpenStreetMap API. In addition, it is not clear whether walking distances from the MapQuest API or the approximate vincenty distances calculated from latitudes/longitudes are accurate for use on a city-level scale.

Numerous areas for improvement remain for this scoring system; this is discussed in the following section.

Future Work

As mentioned in prior sections, the number of factors currently incorporated into the score are limited - factors mentioned in the Methods section, such as grocery store location, could be included in the score. The scoring system could be made more sophisticated with measures such as taking into

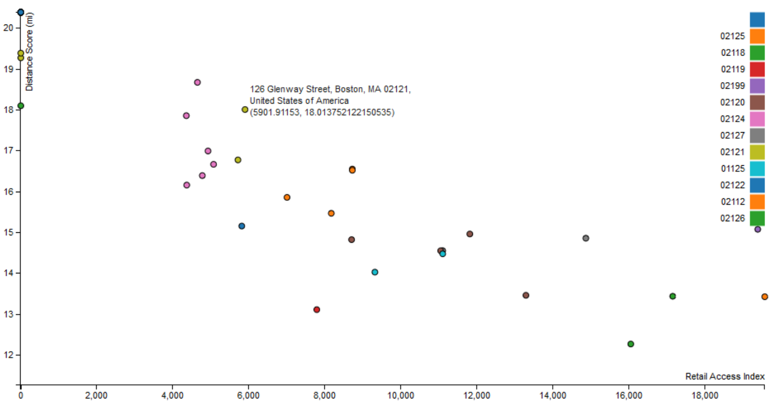


Figure 2: Interactive plot comparing weighted distance score against Retail Access Index from the Housing Affordability Index.

account inpatient bed count at each hospital, etc., incorporating more statistical methods, and using less arbitrary weights. The correlation between the distance score and measures such as population density and rental/property value could also be studied.

In addition, further investigation would be needed to determine whether including data on resources in neighboring towns (e.g. Brookline, Cambridge) that are still accessible to Boston residents would provide a more complete picture of "age-friendliness."

This scoring system could be used to randomly select points in a particular neighborhood (instead of within the rectangle of space currently used in the code), and using the scores from multiple points, calculate an average score for the region. Point selection could occur by using geojson polygon data. The scores by neighborhood/region could then be compared to the current distribution of adults age 60 and older throughout different neighborhoods in Boston.

Acknowledgments

The author would like to thank the instructor of the course, Andrei Lapets, for his feedback throughout the development of the project, as well as peers

in CS 591 L1 for their suggestions regarding data sources and use of MongoDB/prov.

References

1. Center for Social and Demographic Research on Aging. 2014. Aging in Boston. Retrieved May 3, 2016 from https://www.cityofboston.gov/images_documents/4-14%20UMASS%20Aging%20Report_tcm3-44127.pdf.

2. U.S. Census Bureau. 2010. QuickFacts, Persons 65 and over, percent, April 1, 2010. Retrieved May 3, 2016 from https://www.census.gov/quickfacts/table/AGE765210/4805000,1304000,0667000,0644000,3651000,2507000.

3. Mayor's Office, City of Boston. 2014. Mayor Walsh's Elderly Commission releases "Aging in Boston." (April 2014). Retrieved May 3, 2016 from http://www.cityofboston.gov/news/default.aspx?id=6600.

4. Anasuya Chatterjee, Jaque King, and Paul Irving. 2014. Best Cities for Successful Aging. Retrieved May 3, 2016 from http://successfulaging.milkeninstitute.org/2014/best-cities-for-successful-aging-report-2014.pdf.

5. National Council on Aging. 2015. United States of Aging Survey: 2015 Results - NCOA. Retrieved May 3, 2016 from https://www.ncoa.org/news/usoa-survey/2015-results/.

6. Transport for London. 2015. Public Transport Accessibility Levels. Retrieved May 3, 2016 from https://s3-eu-west-1.amazonaws.com/londondatastore-upload/PTAL-methodology.pdf.

7. Walk Score. 2016. About Walk Score. Retrieved May 3, 2016 from https://www.walkscore.com/about.shtml.

8. U.S. Department of Housing and Urban Development. Data and Methodology. Location Affordability Index Version 2.0. Retrieved May 3, 2016 from http://www.locationaffordability.info/lapmethodsv2.pdf.

Appendix

Miscellaneous Notes on Code

Certain scripts are included in the May 2016 version of the repository but did not contribute to the visualization or scoring. These include:

1. `retrievezillow.py` This is a script which retrieves median rental pricing data and median property price by zip code from Zillow's website. (The script does not use Zillow's API; rather it downloads relevant csv files from a public-facing site and parses out information accordingly.) Because Zillow's data does not cover all zip codes in the City of Boston, for those zip codes lacking such data, the median rental price/property value across available Boston data is used as the default value.

2. `retrievepharma.py` This is a script which retrieves pharmacies from Yelp in a certain radius using an input zip code. Due to time constraints this script was not fully developed for use. The geocoordinates found using this script could be used to produce a median pharmacy distance for a neighborhood or simply to find the nearest pharmacy to an input location.

Included for posterity's/completeness's sake is the `project_one` directory, which contains a prior project for examining needle disposal in the City of Boston.

Sources for Data Sets

This list can also be found in the `README.md` file in the main directory of this project.

1. `https://geocoder.opencagedata.com/`

2. `https://www.fcc.gov/general/census-block-conversions-api`

3. `http://www.directoryma.com/MAReferenceDesk/MassachusettsZipCodes.html`

4. `http://www.mbta.com/rider_tools/developers/default.asp?id=21895`

5. https://data.cityofboston.gov/Public-Health/Hospital-Locations/46f7-2snz
6. http://www.cityofboston.gov/elderly/center.asp
7. http://www.cityofboston.gov/elderly/agency.asp
8. http://www.zillow.com/research/data/
9. http://maptimeboston.github.io/leaflet-intro/neighborhoods.geojson
10. https://mapicons.mapsmarker.com/