West Palm Beach and Palm Beach County, FL, USA

Open Street Map Data Project
Udacity: Data Analysis Nanodegree Program
Andrew Burruss
alburruss@gmail.com
February 20, 2017

Location Area

Geographic data was provided by Open Street Map (www.openstreetmap.org) and retrieved as an osm file (53.7 MB uncompressed). The rectangular boundaries of the geographic region were Latitude (26.7885, 26.2232) (North), and Longitude (-80.1281, -80.0079) (West).

The region included the South Florida I-95 corridor in Palm Beach County on the east coast of Florida. The northern border was in Mangonia Park and the southern border at Pompano Beach in northern Broward County. Metropolitan areas of West Palm Beach and Boca Raton were included between those two boundaries. The region was bounded on the east by the Atlantic Ocean, and the eastern longitude coordinate extended over the ocean to include all coastal areas. The western frontier included Interstate highway 95 and the Florida Turnpike.

Data Conversion

The mapping data was downloaded in Open Street Map XML (osm) format. The data was converted to Javascritpt Object Notation (json) format and uploaded to a Mongo DB client. Database queries were done with Mongo DB. All data was converted, queried and updated using Python 2.7 and the Python Mongo DB driver.

Open Street Map uses a relational tag-based data structure, and all data tags fall into one of three categories: *node*, *way*, or *relation*. Only tags of type *node* and *way* were converted and uploaded to the database. The location data included a total of 249825 records: 221558 nodes and 28267 ways, which were contributed by 294 unique users.

The code which handled the json conversion used several strategies to clean up any inconsistencies in address formatting before the records were uploaded to the database. Automated street name

formatting was mostly successful with only a few exceptions. Street abbreviations were omitted in favor of whole names. For example, “Main St.” would be converted to “Main Street”. Directional abbreviations were formatted to US Postal Service conventions: all capitals and no periods. For example, “N. Main Street” would be changed to “N Main Street”. Street names were checked for valid formatting and were also compared to a dictionary of common street names. Street names which were found invalid or not recognized in the dictionary were uploaded to the database and reported in a log file which accompanied the conversion.

The procedure caught several formatting errors. Address records where street numbers were in the wrong field were then corrected in the database. Several data records with incorrectly formatted addresses were cross-referenced with a search of Google Maps (maps.google.com). The questionable addresses were verified in Google Maps, including street view images of the location. Once these addresses were authenticated, they were updated in the database. Later queries to the database discovered address records with capitalization errors: all capital characters or no capitalization, and these were corrected in the database.

Notable Features

One of the most interesting suggestions of an address error regarded a street named *Plaza Real* which appeared in three address records. The name was not recognized in the dictionary comparison, but it is a valid name of a street in Boca Raton, FL. The name was flagged as a possible error because it did not contain any of the frequent street name designators such as “Boulevard” or “Highway”. This example informs us that the comparison dictionary should be adapted to the cultural context of the geographic region of study. South Florida is home to a significant Hispanic population and this is reflected in the name *Plaza Real*.

Nine location records listed businesses which accepted bitcoin payment. The businesses varied among a hobby shop, construction, realty, insurance, law offices, and an auto fuel and convenience store. The bitcoin payment type was noted an in an additional field titled “payment”. Two locations also accepted litecoin. User contributors found a precise application of the open street map database and demonstrated the versatility of the relational data structure.

Another unique aspect of South Florida is a profusion of waterways and canals. The database contained more than 300 records which described a variety of waterways, 299 of which were designated as *canals*. Waterway records were designated with a variety of names, but most were in two categories: boating ways and drainage systems. User contributed data gave an extensive picture of South Florida waterways, indicating the influence of water systems on South Florida geography.

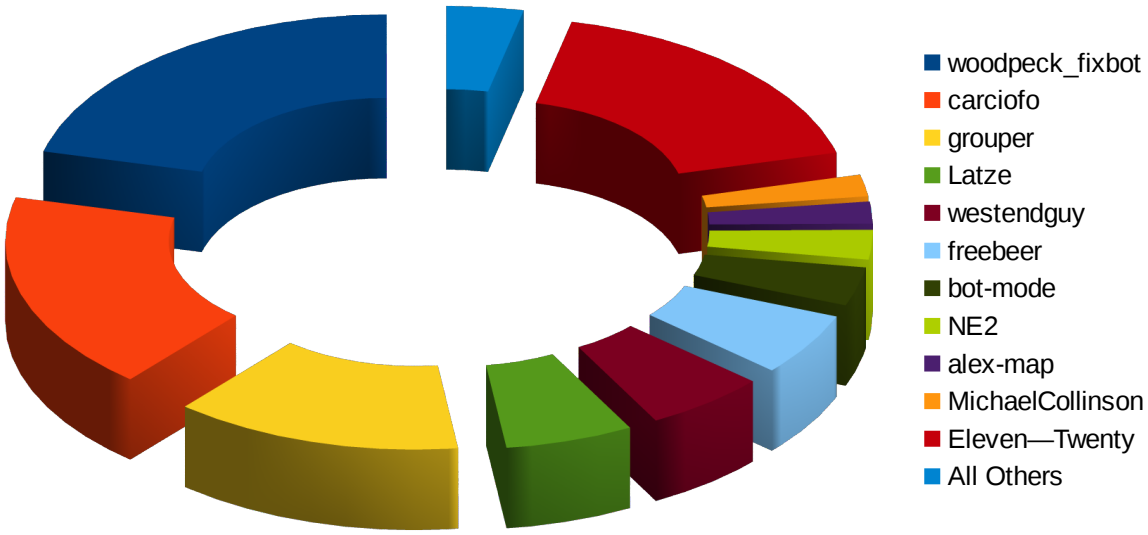
Most Frequent User Contributors

The following table and accompanying chart list the twenty most frequent user contributors to the database, among 294 unique users. This geographic location has a large open source pool of user contributors, but it is dominated by a small group. The top three contributors accounted for a 52% majority of all data record entries, the seventh through tenth most frequent users had 27.07%, the eleventh through twentieth added 17.41%, and the other 274 users contributed 3.52% of the total records. The most frequent contributor, “woodpeck_fixbot” contributed more records than the 284 least frequent users. Please note that this comparison is by the number of data records, not by the total volume of data.

It is also notable that the first (“woodpeck_fixbot”) and seventh (“bot-mode”) most frequent contributors appear with “bot” in their user names, as well as the eighteenth (“mapserver”) appears to have a technological designation. These names indicate that a quarter (25.3%) of the data records *may* have been generated by automated methods. The data records only provided user names recognized by Open Street Map and no other identifying information.

Rank	User ID	Records	Relative Frequency
1	woodpeck_fixbot	52680	0.2109
2	carciofo	44266	0.1772
3	grouper	32967	0.132
4	Latze	14635	0.0586
5	westendguy	13980	0.056
6	freebeer	13966	0.0559
7	bot-mode	8669	0.0347
8	NE2	6576	0.0263
9	alex-map	4937	0.0198
10	MichaelCollinson	4866	0.0195
11	MountainAddict	4658	0.0186
12	42429	4225	0.0169
13	wvdp	4173	0.0167
14	SGrossman	2851	0.0114
15	maxolasersquad	2373	0.0095
16	maxerickson	2273	0.0091
17	IvoSan	1966	0.0079
18	mapserver	1863	0.0075
19	dysteleologist	1538	0.0062
20	Luis36995	1189	0.0048
	All Others	8795	0.0352
	Total in database	249825	

Relative Frequency of User Contributors



Topics for Further Study

A significant number of waterways and canals were included in the data collection. Waterways are used for commerce and transportation in many coastal areas, but the South Florida waterways are extensively integrated with residential neighborhoods. The inclusion of these waterways into the database provided unique mapping data of the area. A waterway study could attempt to collect all commercial and residential South Florida waterways.

Development of an open source map of South Florida waterways would be beneficial to a variety of parties. Both commercial and recreational boaters could benefit from access to this data. The open source format allows for boaters to search for locations and courses, as well as update the records as they see fit for their interests and uses. Delivering the waterway map on a website such as Open Street Map would also improve accessibility of these database to the general public. Recreational boaters could be able to access the data from a smartphone or tablet wherever they have access to satellite data service. We do not aim to replace existing maritime charts, but to improve the accessibility and user interface for boaters. The mobile interface could streamline searches for local canals and plotting courses for marine journeys.

Mapping data regarding residential canals would be of interest to both home owners and realtors. Access to waterways is a potentially valuable aspect of a property. Realty agencies could showcase the waterway accessibility of properties. They would be able to provide maps and proximity to local points of interest by boat. This could provide for an enhanced listing with more detailed information than “boat dock available”.

The waterway data provided from Open Street Map is not as comprehensive as the street map data for vehicles and pedestrians. Extensive data collection would be necessary to populate the South Florida waterway database. Two reliable sources which are made available to the public and maintained on the web are South Florida Water Management District (https://www.sfwmd.gov/science-data/gis) and the National Oceanographic and Atmospheric Administration (http://www.charts.noaa.gov).

South Florida Water Management District (SFWMD) is an agency of the government of the state of Florida, and it would likely be the most comprehensive source of waterway data in South Florida. The SFWMD data would be most valuable in distinguishing between waterways which are accessible to boats and others intended only for drainage and spillways. Documents in the OSM data provided some distinction in the second-level tag ‘waterway’. The waterways were designated by names such as “canal”: 299 records, “stream”: 20 records, “ditch”: 10 records, etc. A comprehensive waterway map database would need to make uniform classifications of waterways, with a separate field “boat accessible”: designated “Yes” or “No”. If possible, it would be useful to provide additional accessibility dimensions such as channel depth, width, and bridge or underpass clearances.

SWFMD maintains data files in GIS format which can be converted to GeoJSON format and then uploaded to a Mongo DB document. The GIS format is frequently used standard and there are a number of tools available to assist with the conversion. In particular, the conversion could be handled with Ogre, an open source web client (https://ogre.adc4gis.com/).

The National Oceanographic and Atmospheric Administration (NOAA) is a U.S. Federal agency under the Department of Commerce. It maintains sea charts at http://www.charts.noaa.gov. NOAA Sea charts may not provide as much data regarding local canals, but they would provide information regarding the U.S. Intracoastal Waterway and ocean harbors. Many local South Florida waterways

connect to the Intracoastal or marinas and harbors which let out to the Atlantic Ocean. It would be useful to integrate the local waterway data with the bigger picture of ocean charts and marine destinations.

The NOAA sea charts are an official primary data source and will provide a valuable reference point. The sea charts are posted in two graphical formats: pdf and BSB; and Electronic Navigational Charts (ENCs) are in a vector format: s-57. The vector format can be converted to GeoJSON format which could then be uploaded to a Mongo DB document. Conversion could be conducted with Ogre (https://ogre.adc4gis.com/). The graphical formats (pdf and BSB) may be difficult to be convert to a format that supports queries. It may be most useful to use the graphical formats as a reference to guide to the conversion process and database queries. NOAA Booklet Charts are designed for recreational boaters and may be the best suited for this application.

This collection of geographic data in Palm Beach County, Florida showed us that its canals and waterways are as much a part of its geographic space as are its streets and neighborhoods. A waterway map could provide a new look at this unique aspect of South Florida geography.