US Cities Road Data Quality on OpenStreetMap

CUGOS – Fall Fling 2019
Monica Brandeis, Senior GIS/Data Analyst
Critigen
OSM Data Integrity Assurance

Rule-Based Validation
• Finding Data Inconsistency (map errors)
  • Geometry-based
  • Attribute-based
Geometry–Based Map Error
Attribute-Based Map Error
Atlas Checks

✓ A java based program that systematically flags various types of map errors
Atlas Checks

In order to run, need turn OSM data into atlas (a connected graph representation of the road network)

1 OSM Road

5 Atlas Features
Where to focus first?

Sign Post Check

+ Unusual Layer Check

+ Invalid Access Tag Check
Map Quality Measurement (MQM) Tool

- A vector grid layer showing map error hotspots
- Grid size is determined by the distribution of map errors
MQM Grids Generation

City Boundary

Generate Bounding Box

90% of grids have <10 errors on OSM features
10% of grids (HOTSPOT)
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

City Boundary

Generate Bounding Box

Recursively Divide it into half using K-D Tree
MQM Grids Generation

90% of grids have <10 errors on OSM features

10% of grids (HOTSPOT)

Stop when >=90% of the grids have <10 features

Recursively Divide it into half using K-D Tree
MQM Grids Generation

Final MQM Layer

Recursively Divide it into half using K-D Tree
Re-prioritize Map Error Hot-Spots by Usage

Final MQM Layer

Population

% of Population without Vehicles

Usage

Population by Census Tract

Car Ownership
Final 3 Layers

MQM

MQM + Population

MQM + Car Ownership
The MQM Web App
https://osmquality.io

MQM Visualization and Stats for each city
Future Enhancements

• Flexible data inputs (e.g., Overpass Turbo Results)
• Focus on other map layers (e.g., address, building, water, etc.)
• Other measurements for OSM data quality
• Tool incorporation (e.g. JOSM, HOT Task Manager, or MapRoulette)

MQM 2020

More Detail: https://youtu.be/8lpvf9aeyNI
Acknowledgment

Micah Nacht  Mara Rae
Sayana Saithu  Adam Shaw
Ahmend Ahmouda  Stephen Cerqueira
Lukas Kucinski  Daniel Baah
Ana Ordonez  Kim Kearns
Keahi Konishi  Todd Slind
Amelia Watts
Thank you!

Stay Connected with Critigen:

Contact Info
www.critigen.com
Monica.Brandeis@critigen.com

Telephone
Office Phone: +1 (859) 312-9267

Address
1430 Summit Ave, Suite 100C
Seattle, WA 98122
Selected Atlas Checks for Road Network

**Road Geometry**
- EdgeCrossingEdgeCheck
- BuildingRoadIntersectionCheck
- SnakeRoadCheck
- RoundaboutValenceCheck
- InvalidMiniRoundaboutCheck

**Road Tag and Relation**
- SignPostCheck
- InvalidAccessTagCheck
- StreetNameIntegerOnlyCheck
- UnusualLayerTagCheck
- InvalidLanesTagCheck
- InvalidTurnRestrictionCheck
MQM Grids Generation

City Boundary
MQM Grids Generation

City Boundary

Generate Bounding Box
Re-prioritize Map Error Hot-Spots by Usage

Final MQM Layer
Data Processing
Data Processing

- Join American Community Survey (ACS) data to the census tracts
Data Processing

• Join American Community Survey (ACS) data to the census tracts
• Generate raster layers

1. ACS Data
2. Raster Layer

- Census Tracts
Data Processing

- Join American Community Survey (ACS) data to the census tracts
- Generate raster layers
- Generate grids using the city boundary
- Calculate the mean value of raster layer for each grid
- Normalize values and visualize the results
Data Processing

- Join American Community Survey (ACS) data to the census tracts
- Generate raster layers
- Generate grids using the city boundary
- Calculate the mean value of raster layer for each grid
- Normalize values and visualize the results
Combining MQM and Census Scores

MQM Layer  × 0.7

Census Layer  × 0.3

Reclassify =

Combined Layer
## OSM Data Quality

### Improved Overall

<table>
<thead>
<tr>
<th>City</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detroit, MI</td>
<td>0.39%</td>
</tr>
<tr>
<td>Portland, OR</td>
<td>0.52%</td>
</tr>
<tr>
<td>Des Moines, IA</td>
<td>0.57%</td>
</tr>
<tr>
<td>Manchester, NH</td>
<td>0.61%</td>
</tr>
<tr>
<td>Kansas City, MO</td>
<td>0.64%</td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td>0.65%</td>
</tr>
<tr>
<td>Wilmington, DE</td>
<td>0.68%</td>
</tr>
<tr>
<td>Omaha, NE</td>
<td>0.71%</td>
</tr>
<tr>
<td>Denver, CO</td>
<td>0.76%</td>
</tr>
<tr>
<td>Columbia, SC</td>
<td>0.76%</td>
</tr>
<tr>
<td>Seattle, WA</td>
<td>0.77%</td>
</tr>
<tr>
<td>New York, NY</td>
<td>0.83%</td>
</tr>
<tr>
<td>Little Rock, AR</td>
<td>0.84%</td>
</tr>
<tr>
<td>Las Vegas, NV</td>
<td>0.87%</td>
</tr>
<tr>
<td>Billings, MT</td>
<td>0.89%</td>
</tr>
<tr>
<td>Burlington, VT</td>
<td>0.91%</td>
</tr>
<tr>
<td>Tucson, AZ</td>
<td>0.91%</td>
</tr>
<tr>
<td>Madison, WI</td>
<td>0.95%</td>
</tr>
<tr>
<td>Boise, ID</td>
<td>0.99%</td>
</tr>
<tr>
<td>Dallas, TX</td>
<td>0.99%</td>
</tr>
<tr>
<td>Baltimore, MD</td>
<td>1.02%</td>
</tr>
<tr>
<td>Orlando, FL</td>
<td>1.03%</td>
</tr>
<tr>
<td>Indianapolis, IN</td>
<td>1.04%</td>
</tr>
<tr>
<td>Philadelphia, PA</td>
<td>1.07%</td>
</tr>
<tr>
<td>Wichita, KS</td>
<td>1.12%</td>
</tr>
<tr>
<td>Cheyenne, WY</td>
<td>1.17%</td>
</tr>
<tr>
<td>Minneapolis, MN</td>
<td>1.21%</td>
</tr>
<tr>
<td>Albuquerque, NM</td>
<td>1.22%</td>
</tr>
<tr>
<td>Oklahoma City, OK</td>
<td>1.26%</td>
</tr>
<tr>
<td>Columbus, OH</td>
<td>1.29%</td>
</tr>
<tr>
<td>Salt Lake City, UT</td>
<td>1.34%</td>
</tr>
<tr>
<td>Charleston, WV</td>
<td>1.35%</td>
</tr>
<tr>
<td>Sioux Falls, SD</td>
<td>1.41%</td>
</tr>
<tr>
<td>Portland, ME</td>
<td>1.45%</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>1.46%</td>
</tr>
<tr>
<td>Huntsville, AL</td>
<td>1.49%</td>
</tr>
<tr>
<td>Hartford, CT</td>
<td>1.50%</td>
</tr>
<tr>
<td>Anchorage, AK</td>
<td>1.61%</td>
</tr>
<tr>
<td>Memphis, TN</td>
<td>1.71%</td>
</tr>
<tr>
<td>Providence, RI</td>
<td>1.73%</td>
</tr>
<tr>
<td>Charlotte, NC</td>
<td>1.78%</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>1.92%</td>
</tr>
<tr>
<td>Fargo, ND</td>
<td>1.96%</td>
</tr>
<tr>
<td>Virginia Beach, VA</td>
<td>2.17%</td>
</tr>
<tr>
<td>Boston, MA</td>
<td>2.27%</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>2.27%</td>
</tr>
<tr>
<td>Louisville, KY</td>
<td>2.40%</td>
</tr>
<tr>
<td>Jackson, MS</td>
<td>2.57%</td>
</tr>
<tr>
<td>Honolulu, HI</td>
<td>2.68%</td>
</tr>
<tr>
<td>Newark, NJ</td>
<td>3.56%</td>
</tr>
<tr>
<td>New Orleans, LA</td>
<td>3.98%</td>
</tr>
</tbody>
</table>
OSM Data Quality Improved Overall

- 69% of the cities (36 out of 51) have a decreased amount of error features
69% of the cities (35 out of 51) have a decreased amount of error features
86% of the cities (44 out of 51) have a lower MQM Errors rate
• 69% of the cities (35 out of 51) have a decreased amount of error features
• 86% of the cities (44 out of 51) have a lower MQM Errors rate
• The total feature counts increased in each city
2019 MQM Enhancements
2019 MQM Enhancements

- More accurate representations of cities
2019 MQM Enhancements

- More accurate representations of cities
- Different OSM data source
2019 MQM Enhancements

- More accurate representations of cities
- Different OSM data source
- Map error hot-spots re-prioritization by usage
2019 MQM Enhancements

- More accurate representations of cities
- Different OSM data source
- Map error hot-spots re-prioritization by usage
- Trend Analysis between 2018 and 2019
City Rankings

- Rankings are based on errors rates
- Error rate = \( \frac{\# \text{error features}}{\# \text{total features}} \times 100\% \)
- The error rate is an estimation of the percentage of road features that have mapping errors
- A city with increasing error features might rank higher than a city with decreasing error features
OSM Data Quality Improved Overall

- 69% of the cities (35 out of 51) have a decreased amount of error features
- 86% of the cities (44 out of 51) have a lower MQM Errors rate
- The total feature counts increased in each city
- The MQM Errors rate at each ranking decreased