Geospatial Analysis of the Electronic Health Record
Jim Gay

- Information Analyst III
- Enterprise Analytics & Reporting Team
- Experience: 5 years (CHOP), 10 years (Healthcare), 18 years (BI/Data Warehouse)
• **Children’s Hospital of Philadelphia (CHOP)**
  
  - The Children’s Hospital of Philadelphia is consistently ranked one of the top pediatric hospitals in the nation in the *U.S. News & World Report*’s Honor Roll of the Best Children’s Hospitals.
  
  - Largest pediatric network in the United States with over 50 locations in Pennsylvania and New Jersey.
  
  - CHOP’s Information Services department named a Top 25 ELITE 100 organization.
Agenda:

• Geocoding our patient population
• Review the automated geocoding process
• Review the results
• Integrating GIS into our health record
• Questions?
CHOP’s Worldwide Footprint

Patients visit CHOP from over 110 countries

International Patients by Country
Our infrastructure

- EPIC electronic health record
- Patient population of 2.5 million (6 million addresses in total)
- Employee population of 15,000
Filter data for geocoding

• The following addresses were omitted from geocoding:
  - Missing address line one
  - Missing zip code
  - International patients
  - Test patients
## Address Data Quality

- Over 300 different spellings of ‘Philadelphia’ in our EHR.
- Examples:

<table>
<thead>
<tr>
<th>PHIADELPHIA/09</th>
<th>PHILADELPA</th>
<th>PHILADEIB15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHILADELPHIAA</td>
<td>PHILADELPGIA</td>
<td>PHILADEILPHIA.</td>
</tr>
<tr>
<td>PHILADELPHIAL</td>
<td>PHILADELPH</td>
<td>PHILADEKPHIA</td>
</tr>
<tr>
<td>PHILADELPHIALS</td>
<td>PHILADELPHA</td>
<td>Philadel</td>
</tr>
<tr>
<td>PHILADELPHIAP</td>
<td>PHILADELPHAI</td>
<td>Philadel[phia</td>
</tr>
<tr>
<td>PHILADELPHIAY</td>
<td>Philadelphia</td>
<td>PHILADEL=HIA</td>
</tr>
<tr>
<td>PHILADELPHIIIA</td>
<td>PHILADELPHAP</td>
<td>PHILADEL=PHI</td>
</tr>
<tr>
<td>PHILADELPHIL</td>
<td>PHILADELPEA</td>
<td>Philadelphia</td>
</tr>
<tr>
<td>PHILADELPHILA</td>
<td>PHILADELPHGIA</td>
<td>PHILADELDELPHIA</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>philadelphi</td>
<td>PHILADELEPHIA</td>
</tr>
</tbody>
</table>
Geocoding Process – Overview

- Import filtered addresses (patients and employees)
- Geocode the locations using ESRI 2014 USA Geocoding Service
- Spatial join to identify the census tract for each patient and employee
- Store information in our geodatabase (latitude, longitude, metadata, census tract)
Geocoding Process – Import & Geocoding

Import new patients & employees along with updated addresses

Geocode using ESRI 2014 USA Geocoding Service
Geocoding Process – Data Cleanup

• Remove fields
  - Match Type
  - ARC_Address
  - ARC_City
  - ARC_State
  - ARC_Zip
  - Original address fields from EHR

• Add field
  - Add field GEOCODING_SOURCE to dataset
Geocoding Process – Spatial Join

- Identify census tract for all patients and employees
- Source is USA Census Tract Boundaries
- Most time consuming part of the process
Geocoding Process – Database Export

• Before exporting extra fields are removed from the dataset that were part of the spatial join

• 3 stage tables are exported to our data warehouse
  - PATIENT_GEO
  - PATIENT_GEO_ADDRESS_HISTORY
  - EMPLOYEE_GEO
## Geocoding Process – Final Dataset

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIPS</td>
<td>34007602902</td>
</tr>
<tr>
<td>STCOFIPS</td>
<td>34007</td>
</tr>
<tr>
<td>SPAT_STATUS_NM</td>
<td>MATCH</td>
</tr>
<tr>
<td>LOCATOR_NM</td>
<td>ADDRESS_POINTS</td>
</tr>
<tr>
<td>ACCURACY_SCORE</td>
<td>97</td>
</tr>
<tr>
<td>STREET_LONG_DEG_X</td>
<td>-75.22337765</td>
</tr>
<tr>
<td>STREET_LAT_DEG_Y</td>
<td>39.33111701</td>
</tr>
<tr>
<td>ROOFTOP_LONG_DEG_X</td>
<td>-75.321456</td>
</tr>
<tr>
<td>ROOFTOP_LAT_DEG_Y</td>
<td>39.369258</td>
</tr>
<tr>
<td>SRC_ADDRESS</td>
<td>3501 CIVIC CENTER BLVD, PHILADELPHIA, PA, 19104</td>
</tr>
<tr>
<td>STREET_SIDE</td>
<td>RIGHT</td>
</tr>
<tr>
<td>GEOCODING_SRC</td>
<td>2014 USA GEOCODING SERVICE</td>
</tr>
</tbody>
</table>
Geocoding Process – Python Scripts

• Geocoding model exported to python scripts (patient, patient address history, and employee)

• Python scripts are scheduled using a .BAT file on Windows Server

• An email is sent out once the process is complete each morning

• Processing time for the 3 scripts is about 7 minutes each day
Geocoding Process – Results

- To date, 6.7 million addresses have been geocoded using this process for both patients and employees

<table>
<thead>
<tr>
<th>Geocoding Status</th>
<th># of Addresses</th>
<th>% of Total</th>
<th>Accuracy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>5.85 mil</td>
<td>86.8%</td>
<td>97.4</td>
</tr>
<tr>
<td>Tie</td>
<td>825k</td>
<td>12.2%</td>
<td>82.1</td>
</tr>
<tr>
<td>Unmatched</td>
<td>62k</td>
<td>0.9%</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address Locator</th>
<th>% of Total</th>
<th>Accuracy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address Points</td>
<td>77.9%</td>
<td>95.9</td>
</tr>
<tr>
<td>Street Address</td>
<td>14.6%</td>
<td>91.4</td>
</tr>
<tr>
<td>Zip Code</td>
<td>7.4%</td>
<td>100</td>
</tr>
<tr>
<td>City State</td>
<td>0.1%</td>
<td>95.1</td>
</tr>
</tbody>
</table>
Geocoding Improvements

• **Adding Census Block to our geodatabase**
  - We determined that census tract was not giving us the level of detail we needed so we are setting up a process to add census block to our geodatabase

• **Potential improvements**
  - Reverse geocoding
  - Address clean up
  - Eliminate free text for city and state within our electronic health record
  - Only one geocoding failure in the past 13 months
Geocoding Failure

Address Line 1:

{"ATTACHMENTS":[{"__TYPE":"ITEMIDATTACHMENT:#EXCHANGE","ITEMID":{"__TYPE":"ITEMID:#EXCHANGE","ID":"AAMKADY0MJZINGNHLTEYZWMTNDI4NI04ZTLMMLTRHYTJIMDQXODIXYWBGAIAAAACHIU8BRGY3T6P6EDSKOCXNBWCZ4IWRXUCTQIKRGSPY+MEJADH0FX1TAABRFDMGBEXTSZXPFMW8AZKAAE11VEVAAA=","CHANGEKEY":null},"NAME":"LIMITED TIME! SUMMER FLASH SALE = 40% OFF EVERYTHING + EXTRA 15% OFF","ISINLINE":false}]}

The Children's Hospital of Philadelphia
GIS @ CHOP – Services Implemented

• **ArcGIS for Server**
  - Feature services for boundaries implemented (Countries, States, Counties, Zip Codes, and Census Tracts).

• **ArcGIS for Desktop**
  - Used by our analytics team and strategy team for various GIS analysis.

• **Business Analyst**
  - Used for geocoding, routing, and market analysis.

• **KliqMaps extension for Qlikview**
  - Plugin that runs within our Qlikview applications that integrates our analytic tools with our ESRI services.
Integrating GIS into our EHR
Geospatial Analysis with Qlikview
Geospatial Analysis @ CHOP

• Information Services department – Alternative work arrangement (AWA)
  - A study was done to determine how far each IS employee traveled to work each day.
  - The original policy was that each IS employee could work from home 1 day per week.
  - After the analysis was completed a change was made to the IS policy allowing employees to work from home 2 days per week.

• Planning, Strategy, and Business Development

• Asthma

• Population Health
Questions?

Contact:

Jim Gay

gayj@chop.edu