Tuesday, July 26, 2016

Updates added Friday July 29, 2016

Geocoding meeting

Present: Robin, Corinne, Soo, Chandra, and Liz & Paige via Skype

Not present: Brett (who relayed interests via phone meeting with Sally & Chandra, July 29)

The range in quality/specificity of data from sample participants were discussed with some only providing city and state for some of the entries. Tester sheets or other data bases could be used for comparison to ‘fill in’ gaps, e.g., tester sheets. Electronic search information, already done to contact subjects will be considered as well, weighing costs and consultation of IRB.

The software was discussed that would be needed at one or both sites (ArcGIS family of tools and server, Social Explorer).

IBG: The ArcGIS tool StreetMap from ESRI can be used ‘offline’ to obtain lat/long, only current maps needed.

UCR: ArcGIS, and perhaps Social Explorer as needed, will be used to map lat/long with random token ids provided by IBG to map distances. A narrow set of geo-variables should be settled on for efficiency.

ArcGIS can be done outside of the cloud (‘offline’) if using the server option. Social Explorer is in the cloud. Needs to discuss strategies to minimize concerns, if use Social Explorer. ArcGIS does have demographic layers that can be used.

Data entry format to enter lat/long in preparation for the geocoding at UCR was circulated based on an ESRI email exchange with Chandra. Note that Year needs to be added so that the right mapping will be done.

On the analysis end at UCR the dataset will be divided by year so that geocoding analysis (for the network analysis) will be done separately and then we will combine results files afterward).

What do we want to geocode? Liz and Soo are interested in socio-demographic variables related to neighborhood qualities (SES, cohesion, crime, etc). Paige with Chandra are interested in the access to parks, walking trails, rec centers, etc. (cf. CATSLife grant). It may be useful to track availability/distance to hospitals and health care centers too. Brett is interested in developing geo-coding skills for future use with CATSLife health-related traits. Need to hone in on core aspects and conduct a trial to see what the time requirements are before adding more targets for mapping.

Action items:

IBG -

Patricia will enter the data from CATSLife subject address histories.

Patricia will take 100 tester sheets, and will track time to enter data as well as measure thickness/width to estimate the total time it might take to do all sheets.

Soo will organize student RA’s to help enter tester sheet information. Possibility to interact with Soo on a research/course paper and certification possibilities in ArcGIS, or related, were discussed as possible incentives.

See Update next page for a test case scenario with tracking files that is already available for the full CATSLife target sample.

UCR –

Chandra, with Paige and Liz, will figure out the software needs of both sites and assist in getting it set up.

TEST SCENARIO UPDATE -Jul 27, 2016

Robin suggested, and Chandra concurs, to conduct a test case scenario in the next month that used information from the tracking files (last known address and previous address, if known) for the entire cohort of potential CATSLIFE subjects.

  Reasoning:

1) We have this information on the entire pool - no data entry required.

             2) Formatting for the Subject Address geocoding is trivial.

             3) We can cross-validate last known address against dates of testing.

             4) If we can do conversion to lat/long as a batch job using local access to StreetMap, then there is a trivial cost for prepping the file.

             5) This data would allow us to test the UCR systems for conversion quickly, i.e., as soon as Paige is done with her summer course.

             6) It's not a wasted effort, because all addresses would have been valid at some point in the subject's life.

             7) There may be a quick proof-of-principle paper (or poster for the New Orleans meeting), i.e., dispersion from Colorado, epidemiological increase or decrease in risk based on state statistics.

What we (IBG) need(s):

1) License for StreetMap.

2) Someone to run StreetMap locally (Corinne suggestions?).

3) An OK from an IRB perspective (are addresses covered by previous consents?).

See Data Model Mockup.xlsx – The first section is what the Excel entry will look like on IBG’s end, where Patricia would enter addresses and the lat/long. Year was added to ensure that we have year to use the appropriate map.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Subject address Geocoding** | | |  |  |  |  |  |  |  |
|  | **Colorado Initial Dataset for Geocoding** | | | | | **Colorado Geocoding Result** | |  |  |  |
| **YEAR** | **Random ID** | **Street** | **City** | **State** | **Zip** | **Latitude** | **Longitude** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Subjects received and geocode parks and trails** | | | | |  |  |  |  |  |
|  | **UCR Initial Dataset #1** | | | **UCR Initial Dataset #2** | | | | | **UCR Geocoding Result #1** | |
|  | **Random ID** | **Lat** | **Long** | **Site ID** | **Street** | **City** | **State** | **Zip** | **Lat** | **Long** |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **UCR Network Analysis (note a 1 to many relationship between each Random ID and multiple Site IDs)** | | | | | | | | |  |
|  | **From UCR Dataset #1** | | | **From Dataset #2** | | | **Analysis Result** | |  |  |
|  | **Random ID** | **Lat** | **Long** | **Site ID** | **Lat** | **Long** | **Distance** | **Travel Time** |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Dataset for Colorado** | | |  |  |  |  |  |  |  |
|  | **Random ID** | **Site ID** | **Distance** | **Travel time** |  |  |  |  |  |  |
|  | 1 | 1 |  |  |  |  |  |  |  |  |
|  | 1 | 2 |  |  |  |  |  |  |  |  |
|  | 1 | 3 |  |  |  |  |  |  |  |  |
|  | 1 | 4 |  |  |  |  |  |  |  |  |
|  | 2 | 1 |  |  |  |  |  |  |  |  |
|  | 2 | 2 |  |  |  |  |  |  |  |  |
|  | 2 | 3 |  |  |  |  |  |  |  |  |
|  | 2 | 4 |  |  |  |  |  |  |  |  |
|  | 3 | 1 |  |  |  |  |  |  |  |  |
|  | 3 | 2 |  |  |  |  |  |  |  |  |
|  | 3 | 3 |  |  |  |  |  |  |  |  |