The challenges and opportunities in bringing data science to the problem of homelessness

THE DATASET

The HMIS data format is non-trivial. Much of the initial work was by CARES in taking the time to explain the entries in the database as the manual itself is 125 pages, and refers to hundreds of ideas and concepts which are not widely known outside of the homeless-support community (https://www.hudexchange.info/resource/3824/hmis-data-dictionary/).

As an example, every time a person is provided support, that type of support is noted and it is not simply an overnight homeless shelter, as many people might think. Instead, there are 12 types of support:
- Emergency Shelter
- Transitional Housing
- Street Outreach Services Only
- Safe Haven
- Housing Only
- Housing with Services
- Permanent Supportive Housing
- Day Shelter
- Homelessness Prevention
- Rapid Re-Housing
- Coordinated Assessment
- Other...

This is just a small part of the information encoded in the dataset. We needed to take time to understand this to avoid a common problem with academic/community partnerships.

The academic side of the partnership must take care not to solve problems that they already know how to solve, but instead to try to understand and solve the problems of the community.

To this end, the Siena group listened to what CARES might need in order to find a place where we could bring some added value. There was initial talk of machine learning and cluster analysis but we quickly realized the data was not in a format that could easily interface with standard data science tools.

hamis: A python tool to visualize and analyze HMIS data

In 2017, we released an open-source, publicly available tool, called hamis and it can be found on Github (https://github.com/siena/hamis). This python module will read in the 12 .csv files from the HMIS database and build a relatively simple, cross-referenced dictionary that makes it easier to do simple analysis and visualizations. CARES dataset, which was run through 2016 at the time of the paper, contained information on almost 90,000 individuals across NY state.

We are at the level to actually start real analysis of the data. Shown here are some examples of early visualizations of the datasets, using the hamis module, standard python tools, and anonymized data, provided by CARES.

We thank a number of our collaborators at both Siena and CARES, including Sara Mahar, Paul Thurston, Terry O’Brien, Emily Reebihn, and Maureen Burns.

REFERENCES AND ACKNOWLEDGEMENTS

2. CARES of NY, Inc. http://caresny.org/
4. Siena College Center for Academic Community Engagement https://www.siena.edu/centers/institutes/academic-community-engagement/