Linking Extant Social and Environmental Data at Multiple Scales to Surveys: Activity Space

Ned English[1], Kevin Brown[1], Chang Zhao[1],
Kate Cagney[2], and Charlie Catlett[3]
[1] NORC at the University of Chicago, 55 E. Monroe Street, Ste 3100, Chicago, IL,
[2] The University of Chicago, 1155 E. 60th Street, Chicago, IL, 60637
[3]Argonne National Laboratory, 9700 S. Cass Ave, Lemont, IL, 60439

Abstract

The goal of the Activity Space research program has been to understand the social and spatial environments of older adults' lives and their influence on health outcomes. We do so by quantifying "activity spaces", that is where they conduct their daily routines, with respect to environmental and social influences. In so doing our study has conducted a household survey and used smartphone technology to capture GPS location and administer Ecological Momentary Assessments (EMAs) that consider physical location, social network interaction, and momentary health at multiple time points for approximately 400 baseline respondents in Chicago, IL. Moreover, we have used geographic information systems (GIS) to link respondent location information with individual-level sociodemographic data sets as well as those from the "Array of Things" project to characterize physical and social environments. Array of Things (https://arrayofthings.github.io/) is an urban sensing network that has been collecting real-time environmental and activity data in cities including Chicago. We then examined the association between environmental exposure and individual health outcomes using our compiled data. Our research is of interest to survey practitioners and social-science researchers who want to take advantage of technology to enhance individual-level data sets with multi-level environmental data.

Key Words: Sensor data, GIS, environmental data

1. Introduction

Social science researchers generally feel that households are impacted by their neighborhood context. As such we expect the social and physical environment may influence health or other outcomes of neighborhood residents. At question is how to best to capture and quantify neighborhood-level factors. Specifically, current approaches may not effectively assess exposure to various factors or access to potentially-beneficial resources. Secondly, the definition of "neighborhood" is not standardized as official boundaries may be insufficient or not relevant for many lived human experiences. Moreover, residential neighborhood is a small part of daily experience with individuals traveling for employment, services, and recreation. Where and how people spend time may prove in fact more valuable for understanding the health and well-being of neighborhood residents. While not previously possible, current mobile technology allows researchers to define "activity spaces" beyond official neighborhood boundaries.

“Activity Spaces” may be defined as locations of routine activity in everyday life (York Cornwell and Cagney, 2017). Passive GPS data collection allows us to capture information on how activity spaces are defined and constructed via smartphone technology. Our current pilot study uses data from the Chicago Health in Real Time (CHART) project to link individual household survey data with environmental information using Geographic Information Systems (GIS). In so doing we answer four questions about the collection and analysis of location information from survey respondents. First, we consider issues related to the capture of location information from a sample of older adults in Chicago. Second, we discuss linking their location information to sensor data. Lastly, we conduct preliminary investigations into potential relationships between environmental exposure and health outcomes, in our case respiratory diseases, including emphysema, asthma, chronic bronchitis, or chronic obstructive pulmonary disease.

2. Background

The Chicago Health and Activity in Real-Time (CHART) project, based at NORC and led by University of Chicago sociology professor Kathleen Cagney, has been collecting household, Ecological Momentary Assessment (EMA), and GPS tracking data from 450 elderly Chicagoans across 10 diverse city neighborhoods in order to assess the impact of daily activity spaces and social support networks on the health of older adults (see http://www.norc.org/Research/Projects/Pages/chicago-health-and-activity-in-real-time.aspx) (York Cornwell et al 2019). CHART is in the process of collecting data in three waves using both an in-person survey and five EMA surveys per day over the course of a week in each wave. The project is also using the GPS feature of provided smartphones to track respondents as they go about their daily lives and provide a measure of their weekly “activity spaces.” Although there is a considerable literature on community determinants of health and even a small literature on using mobile devices to study health outcomes (York Cornwell & Cagney, 2017), the investigators believe that CHART is the first study to combine EMA and continuous location tracking methodologies with a traditional household survey. The study also breaks new ground by exploiting a new source of contextual data made available through the deployment of dozens of environmental sensors across metro Chicago, Illinois via the National Science Foundation (NSF)-funded Array of Things (AoT) project.

Based at the University of Chicago’s Argonne National Laboratory and led by PI Charlie Catlett, the AoT collaborative effort among leading scientists, universities, and local government to collect real-time data for research and public planning purposes. To date, the project has installed more than 100 sensors in a variety of Chicago neighborhoods in order to assess micro-environmental conditions such as weather, air quality, noise levels, and both human and vehicle traffic flow (Catlett et al., 2017; see also https://arrayofthings.github.io/). We calculate that more than a third of CHART respondents live within 1 km of an AoT sensor and over 80% live within 2 km of a sensor, allowing us to add environmental contextual data to much of the survey, EMA, and GPS data that has been collected by the project to date. Working with staff at Argonne National Laboratory, we first accessed and downloaded two months (July 2018 and January 2019) of AoT sensor data and conducted preliminary analysis in order to understand the data structure and availability. Further, we developed environmental metrics using raster values for a variety of weather (temperature, humidity, precipitation) and air quality (PM2.5, PM10, O3, NO2, H2S) data and merged those measures with health variables from the CHART survey (asthma, COPD, overall health).

3. Data and Methods

Our primary data set consisted of 450 respondents ages 65+ who participated in a baseline interview with biomeasures and questions about physical and mental health. We also collected GPS tracking information for 413 respondents via Android smartphones, which were geofenced to collect a location every 20 meters they moved. The instrument collected EMA for one week via the same application. Respondents were given five random "pings" per day across seven day period. N = 277 completed at least one, with a mean of 17 completed per respondent.

Table 1. Descriptions of AoT sensor parameters.

Parameter	Unit	Min Value	Max Value	Number of monitors with any observations (July 2018 – July 2019)
Temperature	Fahrenheit	-	-	93
Humidity	Relative (%)	0	100	93
PM10	microgram/m3	0	-	12
PM2.5	microgram/m3	0	-	12
CO	ppm	0	1000	55
O_3	ppm	0	20	57
NO_2	ppm	0	20	57
SO_2	ppm	0	20	57
H_2S	ppm	0	50	56

- indicates no minimum or maximum value

AoT sensors have been installed at various monitors across metro Chicago, where a set of parameters that describe weather and air quality conditions were recorded per seconds on each day from July 2018 to July 2019 (Table 1). An annual average measure of temperature, humidity, PM2.5, PM10, O_3, CO, SO_2, H_2S, NO_2 were derived for respondents in the CHART survey from sensor observations. Specifically, for each parameter, an unweighted average of monthly measure was calculated for each monitor. Monthly values were then aggregated to create a monitor-specific annual average. We used inverse distance weighting (IDW) to interpolate annual average values from all available monitors to a 100m by 100m raster grid cell across the study area using Geographic Information Systems (GIS). Mean raster values within a 250m radius of the respondent's home address were calculated for each sensor parameter and assigned to each respondent to indicate the level of environmental exposure.

Figures 1 through 3 show example raster surfaces of derived environmental measures for mean temperature, humidity, and particulate matter (PM 10) during January, 2019. Each shows some variation across the study area at relatively small scales.

Figure 1, Mean Temperature in January 2019, Chicago, IL

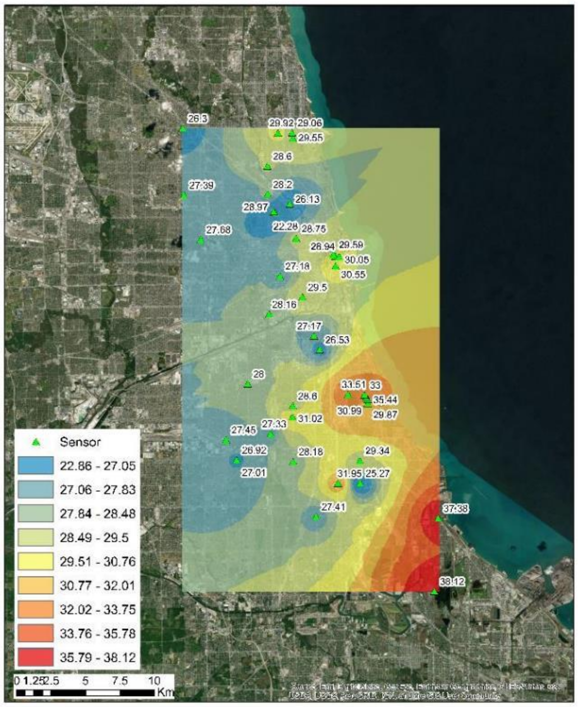


Figure 2, Mean Relative Humidity in January 2019, Chicago, IL

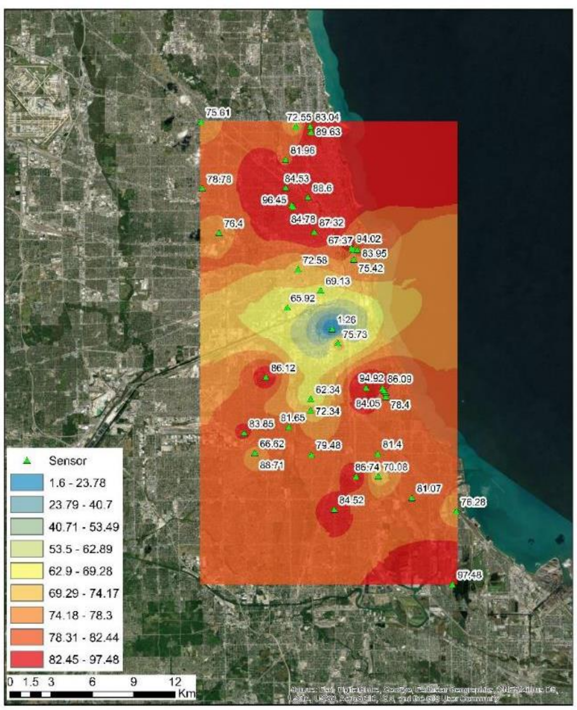
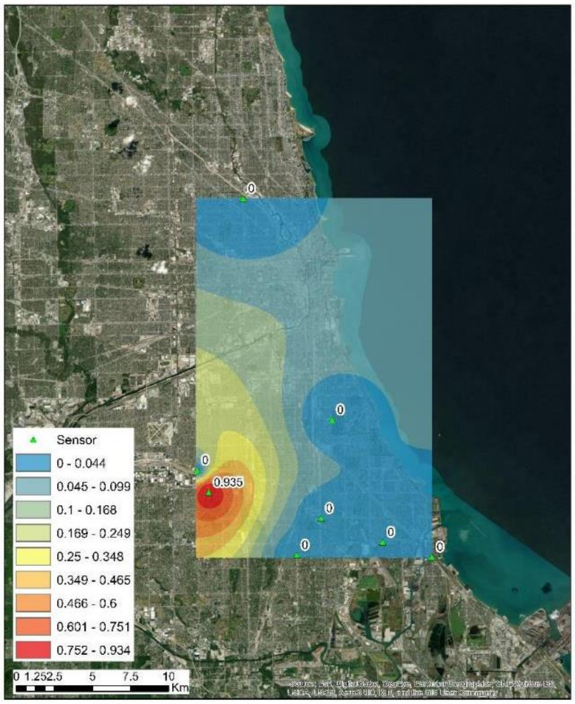


Figure 3, Mean PM10 ($\mu g/m^3$) in January 2019, Chicago, IL



We then linked the preceding raster surfaces along with PM2.5, O3, and CO with household locations as described in Section 3. After data linkage, we conducted preliminary logistic regressions to study the impacts of air pollutants on the development of asthma and other pulmonary conditions, result of which were still preliminary at the time of this pilot.

4. Discussion and Conclusions

Ongoing research shows promise of linking extant data to respondent- provided information in order to provide a richer or more nuanced analysis. Moreover, there are types of information that would be unknown to survey respondents even if they were able to be articulated. There are major challenges in conducting such data linkage, however, namely data availability across time and space and the ability to link such information. In addition, data of interest are often at multiple scales e.g., the individual, neighborhood/community, and regional level. Even if it were possible to acquire and process relevant environmental data, there would still be challenges in visualizing and understanding such high-resolution information.

In-depth human-environmental analysis also requires domain-specific knowledge of the salience of specific variables and their interaction on health outcomes. Examples include the relationship between general physical, mental health and air-quality or relationships between chronic lung conditions such as asthma, chronic bronchitis, or emphysema and specific pollutants. As we have only begun this analysis as part of our pilot, looking ahead we will be conducting further research in the area of logistic and ordinal regression as well as any impacts of spatial autocorrelation. Finally, we will also be linking to GPS locations in order to identify methods for measuring human activity space examining the relationship between activity space and social, physical environment.

References

Catlett, C. E., Beckman, P. H., Sankaran, R., & Galvin, K. K. (2017). Array of Things: A scientific research instrument in the public way: Platform design and early lessons learned. Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering. Pittsburgh, Pennsylvania, April.

York Cornwell, E., Cagney, K. A., & Hawkley, L. (2019). Inequalities in social and physical contexts of older adults' activity spaces: Early results from the CHART study. Annual Meeting of the Population Association of America (PAA). Austin, TX, April.

York Cornwell, E. & Cagney, K. A. (2017). Aging in activity space: Results from smartphone-based GPS-tracking of urban seniors." Journals of Gerontology: Social Sciences 72, 864-875.