NORC at the UNIVERSITY of CHICAGO

NORC WORKING PAPER SERIES

Making Sense of Sensor Data: How Local Environmental Conditions Add Value to Social Science Research

WP-2019-002 | DECEMBER, 2019

PRESENTED BY:
NORC at the University of Chicago
Rebecca Reimer, Research Scientist
55 East Monroe Street
30th Floor
Chicago, IL 60603
Ph. (312) 759-4000

AUTHORS:
Chang Zhao
Kevin L. Brown
Ned English
Charlie Catlett
Kathleen Cagney

AUTHOR INFORMATION

Chang Zhao
NORC at the University of Chicago
(zhao-chang@norc.org)

Kevin L. Brown
NORC at the University of Chicago
(brown-kevin@norc.org)

Ned English
NORC at the University of Chicago
(english-ned@norc.org)

Charlie Catlett
Argonne National Laboratory
(catlett@anl.gov)

Kathleen Cagney
University of Chicago and NORC
(kacagney@uchicago.edu)

Table of Contents

Table of Tables

Table of Figures

Abstract

Recent advances in computing technologies have enabled the development of low-cost, compact weather and air quality monitors. The NSF-funded Array of Things (AoT) project based at Argonne National Laboratory has deployed more than 100 such sensors throughout the City of Chicago. This paper combines longitudinal AoT sensor data with household survey and location tracking data collected from 450 elderly Chicagoans in order to explore the feasibility of using previously unavailable data on local environmental conditions to improve traditional neighborhood research. Specifically, we pilot the use of AoT sensor data to overcome limitations in research linking weather conditions to the activity levels of older adults. We expect that this link will become even stronger as sensing technologies continue to improve and more AoT nodes come online, enabling additional applications to social science research where environmental context matters.

Introduction

The Chicago Health and Activity in Real-Time (CHART) project at NORC at the University of Chicago has been collecting household, Ecological Momentary Assessment (EMA), and GPS tracking data from an initial sample of 450 Chicagoans aged 65 and older across ten diverse city neighborhoods in order to assess the impact of daily activity levels 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). Specifically, CHART is collecting data in three waves based on a neighborhood probability sample of 450 adults aged 65 and older in ten purposively-selected Chicago neighborhoods (York Cornwell & Cagney, 2019). Individual neighborhoods were chosen to ensure geographic and socio-demographic variation across the city of Chicago, which is highly segregated by race/ethnicity and socioeconomic status. Respondents were selected via a systematic random sample of addresses within each identified neighborhood of interest, and recruitment targeted approximately 45 respondents per neighborhood.[1] CHART began with an in-person interview, followed by one week of GPS data collection using the provided Smartphones to measure weekly “activity spaces” plus five EMA surveys per day using a Smartphone app.

Although there is a vast literature on community determinants of health and even a small literature on using mobile devices to study health outcomes (York Cornwell & Cagney, 2017), we 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. The goal of our current research is to demonstrate how sensor-derived data can be used to facilitate the analysis of the daily activity spaces of older adults. In so doing, we provide a methodology for linking CHART survey data with AoT sensor data using geographic information systems (GIS) and performing statistical analysis to examine the associations between immediate environmental exposure and the observed behaviors of elderly Chicagoans.

[1] The response rate for Wave 1 was 46.2%, with a seeming underrepresentation of Hispanics in two neighborhoods, possibly a result of language barriers or fears of being surveyed more generally, and consistently lower than average incomes across neighborhoods, perhaps due to item nonresponse or the elderly population being targeted.

Background and Literature Review

Based at the University of Chicago's Argonne National Laboratory, the AoT is a 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 estimate that more than a third of CHART respondents live within 1 km (0.621 miles) of an AoT sensor and over 80 percent live within 2 km (1.243 miles) 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.[2] 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 (particulate matter, various gases) data and merged those measures with health variables from the CHART survey (asthma, COPD, overall health) as part of a small pilot study presented at the American Association for Public Opinion Research (AAPOR) 2019 annual meeting (English et al., 2019). Here, we expand this pilot in order to demonstrate fully the feasibility of integrating networked sensor data with traditional social science research by using AoT data in combination with CHART survey data to test whether location tracking data are correlated with variations in weather across days of the week and across Chicago neighborhoods that may experience different weather patterns.

Although there is a sizeable literature on the link between air quality and health outcomes, much less research has been conducted to assess what seems like an equally obvious hypothesis, namely whether daily activities are facilitated or constrained by local weather conditions such as temperature, precipitation, or humidity. A 2007 review of this literature noted this gap:

A growing body of evidence indicates that levels of physical activity are influenced by environmental attributes, such as place of residence and accessibility of recreation facilities. While a few studies have considered features of the natural environment, such as access to parks and playgrounds, seasonality and

[2] Note that the coverage is uneven spatially across the ten CHART neighborhoods (see Figure 3), which introduces an unknown amount of bias into our findings. The issue of missing data also arises in our analysis given that weather parameters were not evenly monitored across months in a year, which could introduce biases to the estimation of time-window average. Similarly, not every CHART respondent participated in Smartphone aspect of the study or carried their Smartphone with them every day of the 7-day study period, which is also a potential source of bias in our activity space data.

weather conditions have been relatively overlooked as determinants of physical activity. Previous studies, and common logic, dictate that attributes such as amount of daylight, extreme temperatures and precipitation levels might influence physical activity behaviors, especially walking outdoors, the most common physical activity undertaken by all populations. (Tucker & Gilliland, 2007)

The authors then reviewed 37 studies with participants from eight different counties and found that 27 published articles reported that weather has a significant impact on physical activity, with one showing that weather accounts for as much as 42 percent of variance in measured physical activity. Most studies also found a seasonality effect, with activity levels tending to peak in the summer months, although a few found decreased activity in regions of the world with oppressively hot summer months. Age may be a factor as well, since some studies suggest that the effect of weather on activity may be especially pronounced among very young children, although none of the studies looked for similarly large effects at the other end of the age spectrum. A more recent study did look at the impact of hour-to-hour changes in weather on the physical activity levels of 1,219 Norwegian older adults and found significantly higher activity levels in warmer months and that increasing daily temperatures boosted activity levels in both cold and warm months (Aspvik et al., 2018). Precipitation mattered as well, with rain in warmer months being correlated with decreased activity, especially for the less “fit” participants. A limitation of this literature, however, is that much of the effect of weather on activity may be masked by the available data, which are typically at the regional or national level. Indeed, Tucker & Gilliland (2007) observe that “issues still exist due to variations with respect to urban versus suburban, and mountain versus coastal settings” and that “not enough studies have been conducted within these countries to account adequately for the broad range of climatic zones they encompass” (p. 919).

In this paper, we address this gap in the literature by using neighborhood-level measures of daily weather conditions to assess: (1) the impact of temperature and humidity on the activity spaces of older adults using high-resolution sensor data, and (2) the effect of urban climatic zones that might explain differences in these activity spaces observed across diverse city neighborhoods. Specifically, we hypothesize that:

> H1: Older adults living in neighborhoods with higher temperatures and lower relative humidity level will have larger activity spaces than those living in relatively colder and muggy neighborhoods.
>
> H2: Older adults will have smaller activity spaces on days with cold or hot weather, defined as temperatures below 50 or above 80 degrees, and days that are especially muggy, defined as relative humidity levels above 70 percent.

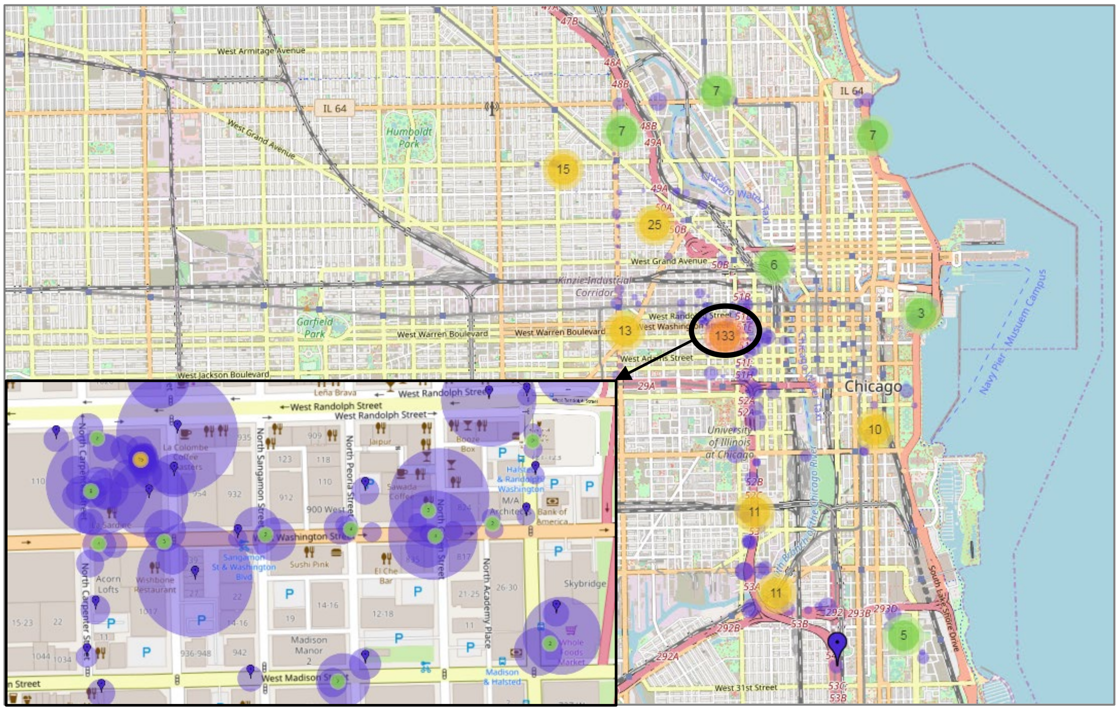
Data and Analytic Methods

Our approach is to combine AoT sensor data with survey data collected by the CHART study, which is based on a neighborhood probability sample of 450 adults aged 65 and older, resident in ten purposively-selected Chicago neighborhoods (York Cornwell & Cagney, 2019). Individual neighborhoods were chosen to ensure geographic and socio-demographic variation across the city of Chicago, which is highly segregated by race/ethnicity and socioeconomic status. Respondents were selected via a systematic random sample of addresses geo-coded within each identified neighborhood of interest. CHART began with an in-person interview, which included a baseline questionnaire capturing respondents' social networks, demographic characteristics, and health status, followed by one week of Smartphone-based observation of respondents' activity spaces across three waves of data collection separated by six months each.

Measuring Activity Spaces

Here we use data collected during the first wave (Wave 1), which occurred over the summer and early fall of 2018 (April 2018 to October 2018). Respondents carried an Android Smartphone (Samsung Galaxy S7) with them as they went about their week, while collecting location data via the *MetricWire* app that was installed on each phone and programmed to log a GPS data point whenever the respondent moved 20 meters. This means that movement was constantly being tracked but not immobility or small movements in a stable location such as the home, which saves battery life on the phone and thus minimizes data loss due to respondent failure to recharge the phone in a timely manner. We also chose to set the geo-fenced radius at 20 meters since this seems to be the limit of accuracy for GPS in a large urban setting with a dense concentration of large buildings and other structures. Figure 1 is an example of the tracked locations of a resident of one of Chicago's Near West Side neighborhood over the course of several days.

Figure 1. Passive location tracking for a resident of Chicago's Near West Side*



* The yellow and green circles include multiple location points, with the actual number of locations provided inside the circle, while the larger blue circles represent an estimate by Google Location Services of the accuracy of each GPS point.

Note that the yellow and green circles include multiple data points, with the actual number provided inside the circle, while the larger blue circles represent an estimate by Google Location Services of the accuracy of each data point.[3] The inset map shows maximum detail for paths taken around the home location. This ability to see the full set of respondent "destinations" as well as the "paths taken" to each destination allows us to define an activity space as either the area covered or the distances traveled as respondents go about their week (Hirsch et al., 2014). For this study, we use the GPS data to generate three measures that summarize the size or range of older adults' activity spaces. First, we rely on a common measure of area traversed called the "standard deviation ellipse" (SDE), which is a spatial unit that includes approximately 68 percent of the respondent's observed locations centered on the mean center of the respondent's observed locations (see Sherman, Spence, Preisser, Gesler, and Arcury, 2005).[4] Second, we assess respondents' average distance from home. This is the mean distance between the

[3] Note that extensive user testing before and after the CHART pilot study found these estimates to be unreliable, thus we do not use them in any of our analyses presented here.

[4] Note that the number of data points collected during the first half of Wave 1 data collection was significantly less than expected due to an error in setting up the Smartphones. A review of data and additional testing suggested that few if any "trips" were missed but that our ability to infer the exact paths taken to particular destinations was compromised until the phone settings were adjusted, thus we do not attempt to use total distances travelled during the week as an additional measure of activity space.

geographic coordinates of each respondent's home and each location where he or she was observed during the study period. Third, we consider the number of non-home Census tracts visited by respondents – that is, Census tracts outside of their own residential tract – during the study period. This provides a general indicator of the extent to which activity spaces are contained within the respondents' residential neighborhoods. We use tracts as the geographic unit here since most prior research considering neighborhood effects on health operationalizes the neighborhood as the residential census tract (Yen et al., 2009).

A total of 375 respondents participated in the GPS tracking and generated location data in Wave 1 for at least three days in the week they were supposed to carry the provide Smartphone. A total of 56,421 locations were captured, with a mean of 151 locations (standard deviation = 109) identified for each respondent. Our analytic sample includes 324 of these respondents who had more than 60 percent of the GPS points collected within the city limit and had complete data on socio-demographic characteristics in the baseline survey. Table 1 compares the distribution of CHART respondents in the full sample to that in the analytic sample.

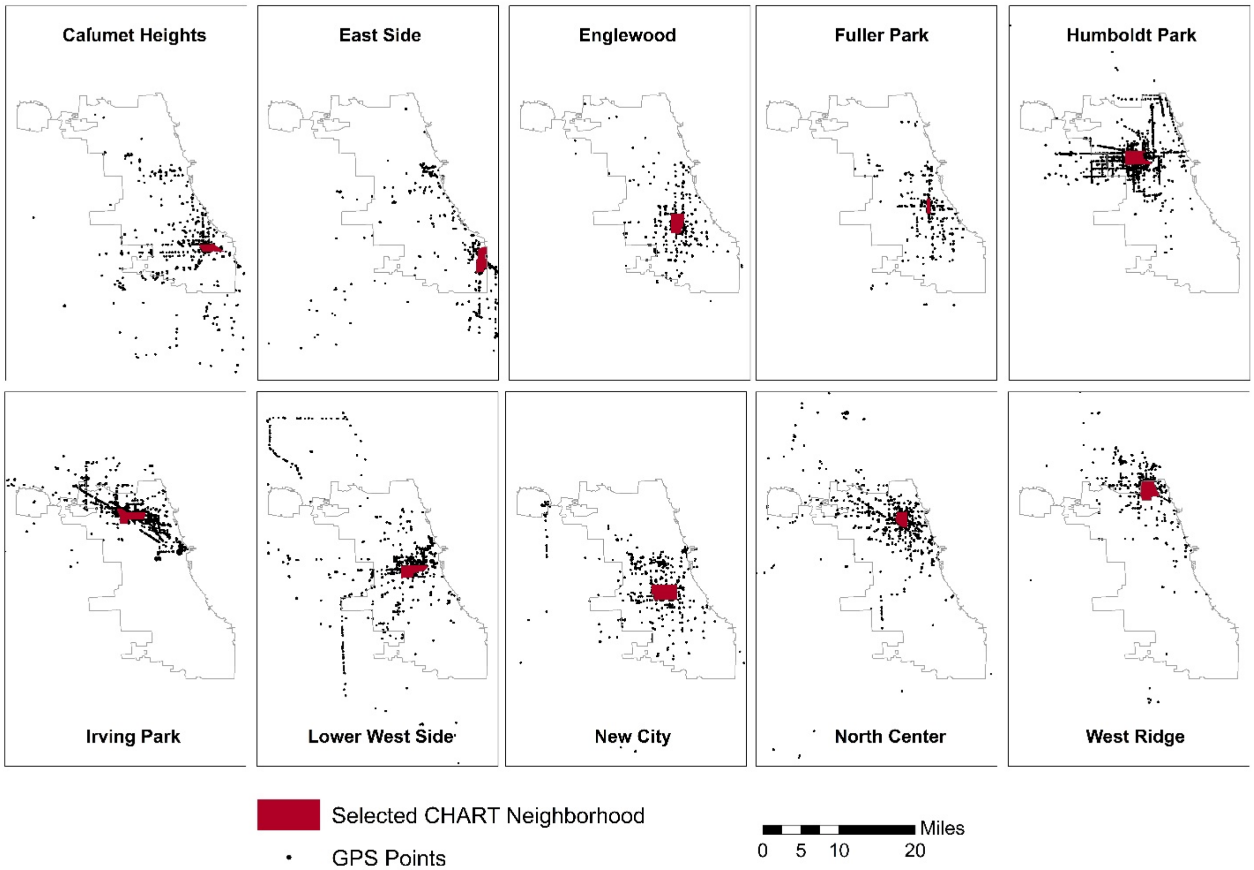
Table 1. CHART Wave 1 Respondents by Neighborhood, Full vs. Analytic Samples

Neighborhood	Full Sample N	Full Sample %	Analytic Sample N	Analytic Sample %	Reduction of Sample Size %*
CALUMET HEIGHTS	50	11.1	35	10.8	-30.0
EAST SIDE	42	9.3	28	8.6	-33.3
ENGLEWOOD	51	11.3	34	10.5	-33.3
FULLER PARK	50	11.1	40	12.4	-20.0
HUMBOLDT PARK	48	10.7	37	11.4	-22.9
IRVING PARK	40	8.9	27	8.3	-32.5
LOWER WEST SIDE	44	9.8	39	12.0	-11.4
NEW CITY	50	11.1	37	11.4	-26.0
NORTH CENTER	38	8.4	27	8.3	-28.9
WEST RIDGE	37	8.2	20	6.2	-45.9
Total	**450**	**100.0**	**324**	**100**	**-28.0**

* Reduction of Sample Size % = [(Analytic Sample N – Full Sample N)/Full Sample N]*100%

Across the Wave 1 study period, respondents were located 0.96 miles from home, on average, covered approximately 11.45 square miles (based on SDE), and visited 12 non-home Census tracts. Figure 2 illustrates the geographic span of activity spaces for respondents based on their residential neighborhood area. Aggregating respondents’ activity spaces within their residential neighborhoods reveals differences in the span and directionality of activity spaces.

Figure 2. GPS observations of respondents' activity spaces, by residential neighborhood areas



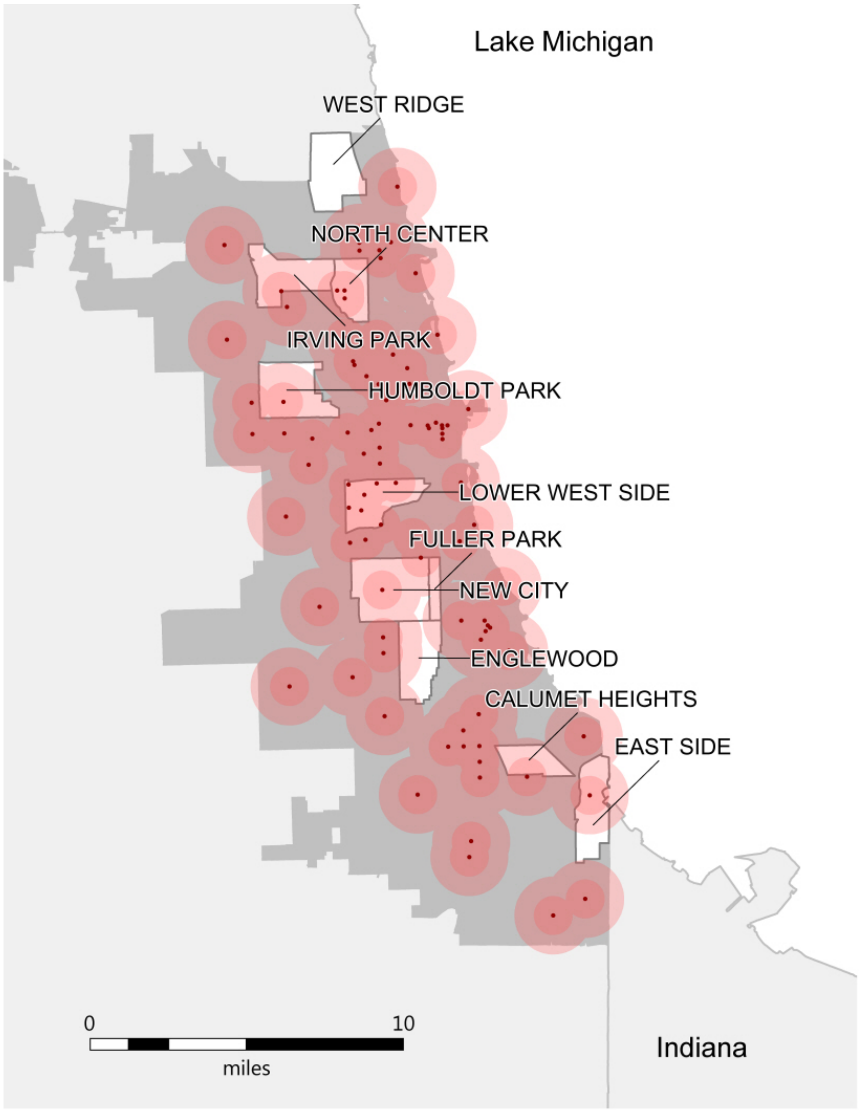
For example, residents of South Side neighborhoods such as Calumet Heights and East Side travel further south of Chicago, while those in North Side neighborhoods such as Irving Park, North Center, and West Ridge travel further north of the city. Activity spaces of residents of Englewood are tightly contained within central area of the city, while those of New City and Fuller Park are also largely found within the central areas of the city of Chicago. These differences suggest that activity spaces are structured by the residential environment, meaning that local resources, amenities, transportation, stressors, and the characteristics of surrounding neighborhoods may shape where respondents go and how they get there.

In addition to these differences across neighborhoods, preliminary analyses using the SDE measure provide evidence that the span of activity spaces varies across socio-demographic groups (York Cornwell and Cagney, 2019). In particular, York Cornwell and Cagney found that the activity spaces of African American older adults in our sample are significantly more geographically constrained than those of white older adults (28.42 square miles vs. 11.11 square miles). We also found differences with respect to education, with older adults who have higher levels of education (bachelor's degree or higher) navigating larger activity spaces than those who did not finish high school (22.67 square miles vs. 11.03 square miles). However, none of these preliminary models controlled for the potential influence of variable weather conditions either across the several months of data collection or across different Chicago neighborhoods on the same days that GPS data was collected.

Measuring Weather Conditions

AoT sensors have been installed at intersections and other locations with access to electricity across metro Chicago, where a set of parameters that describe weather conditions were recorded per second on each day during Wave 1 of data collection. Figure 3 shows the coverage of AoT nodes relative to the home locations of CHART respondents in the City of Chicago, with more than 80 percent living within two kilometers (1.243 miles) of a node.

Figure 3. Overlap between AoT nodes (red dots w/1- and 2-km buffers) and CHART neighborhoods in Chicago (white polygons)



We derived an average measure of temperature and humidity by first calculating an unweighted average daily measure for each monitor that recorded observations for that parameter. Considering that the time of GPS data collection varied widely among respondents across the W1 period (from April 2018 to October 2018), daily values were aggregated over the specific week that each respondent carried a smartphone in order to create a respondent-specific time-window average (n = 324). We then used inverse distance weighting (IDW, Burrough, & McDonnell, 1998; Pebesma, 2004) to interpolate the time-window average values from all available monitors to a 100 m-by-100 m raster grid cell across the study area. Finally, the mean raster value within a 250 m radius of the respondent's home address was calculated based on the individual-specific interpolated surface and assigned to each respondent to indicate the level of environmental exposure. As an example of the IDW maps, Figure 4 shows that there is considerable variation in temperature and humidity across space when averaged over the course of Wave 1 months.

Figure 4. Average temperature (a) and humidity (b) for Chicago during CHART Wave 1 data collection, interpolated using IDW from the AoT sensors (N=92, in blue)

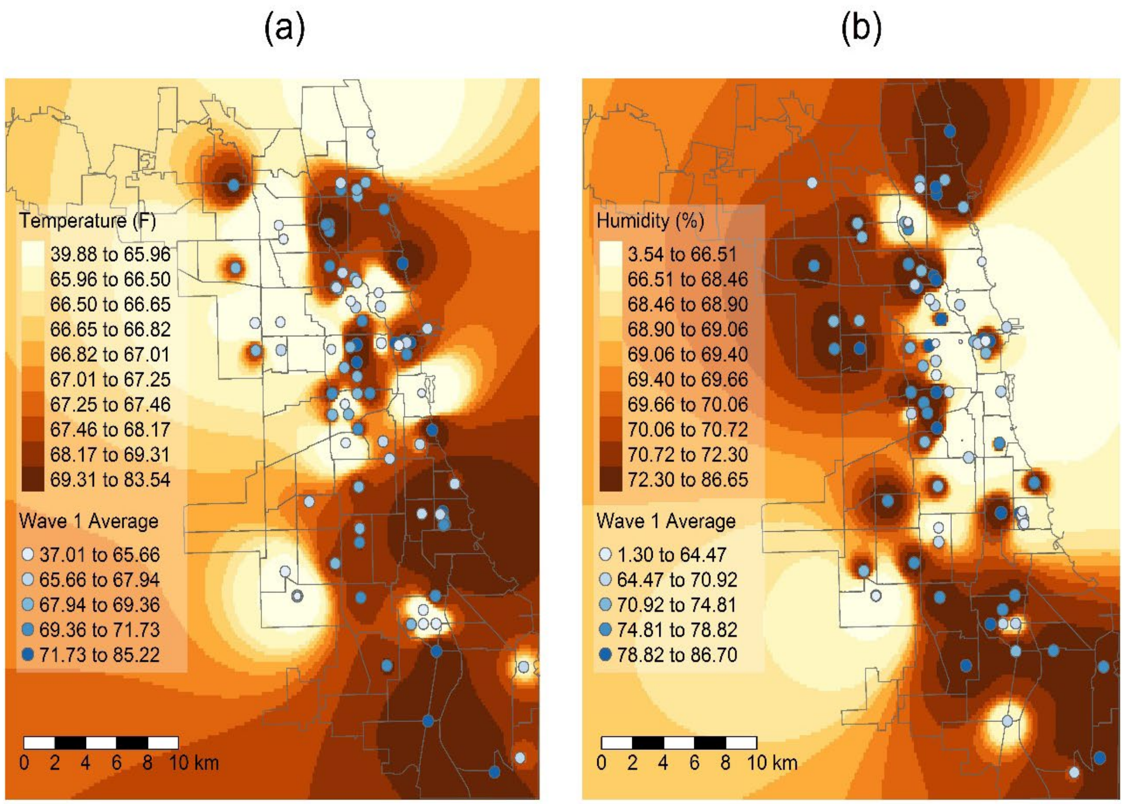
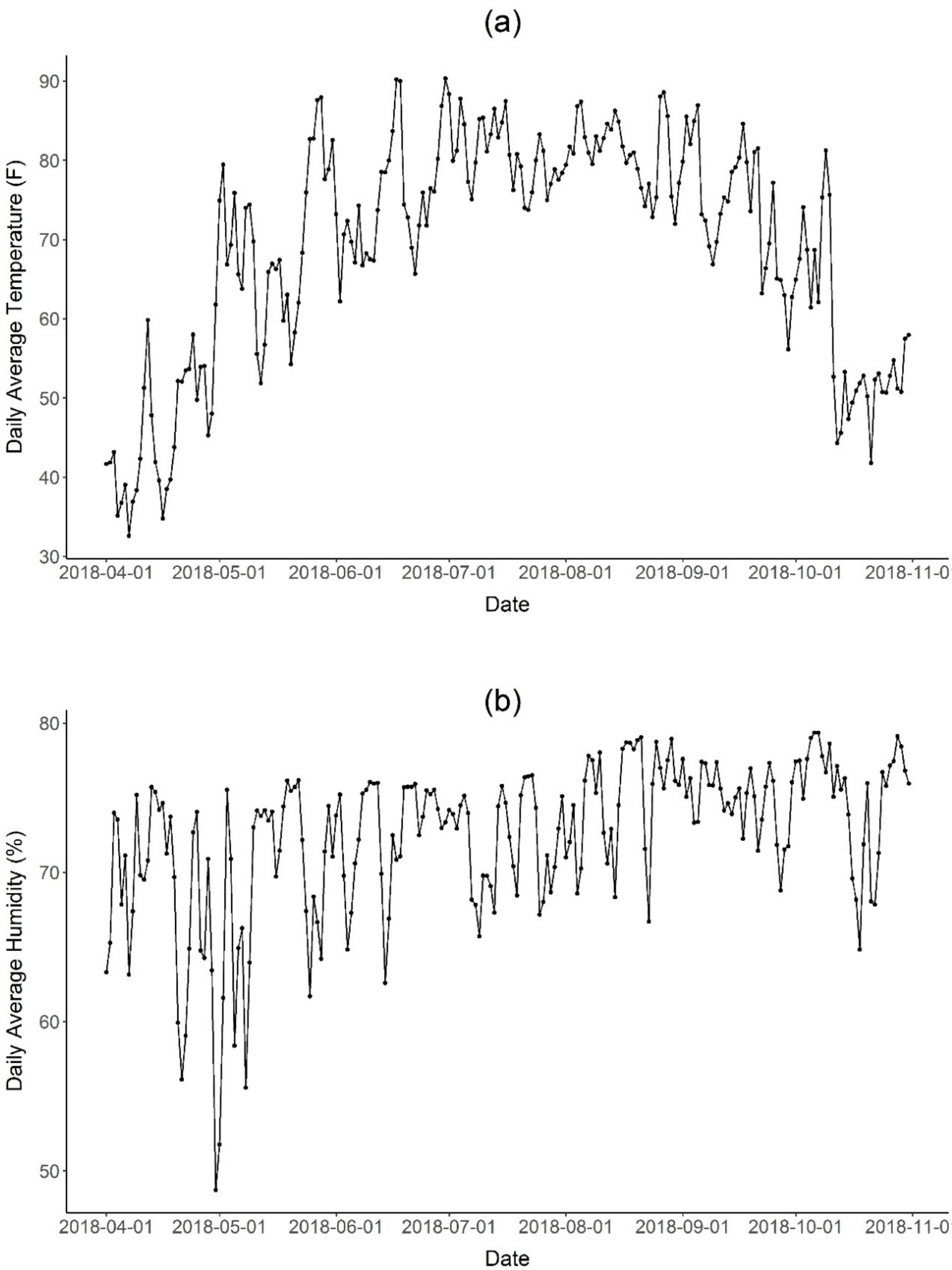


Figure 4 shows that the southern and eastern parts of metro Chicago were exposed to higher temperatures, while humidity levels were highest in the northwestern and southern parts of the city.[5] Figure 5 shows city-wide daily fluctuations in temperature and humidity over the entire Wave 1 data collection period.

[5] Note that the data suggests that a few sensors may not have been functioning properly some of that time. This may introduce bias in our findings that will be addressed by future collaborations with the AoT team to identify and remove unreliable data.

Figure 5. Chicago daily average weather measures from the AoT sensors for Wave 1



Although there is considerable within-month variability, in general, Figure 5 shows that temperatures were highest in July and August 2018 and declined in the fall and early winter months (minimum = 32.60°F, mean = 69.08°F, maximum = 90.33°F, standard deviation = 14.35°F). Humidity levels were comparatively stable across months, which typically ranged from 60 percent to 80 percent (mean = 72.45 percent, standard deviation = 5.13 percent), with a minimum value of 48.71 percent on April 30, 2018.

Statistical Modeling

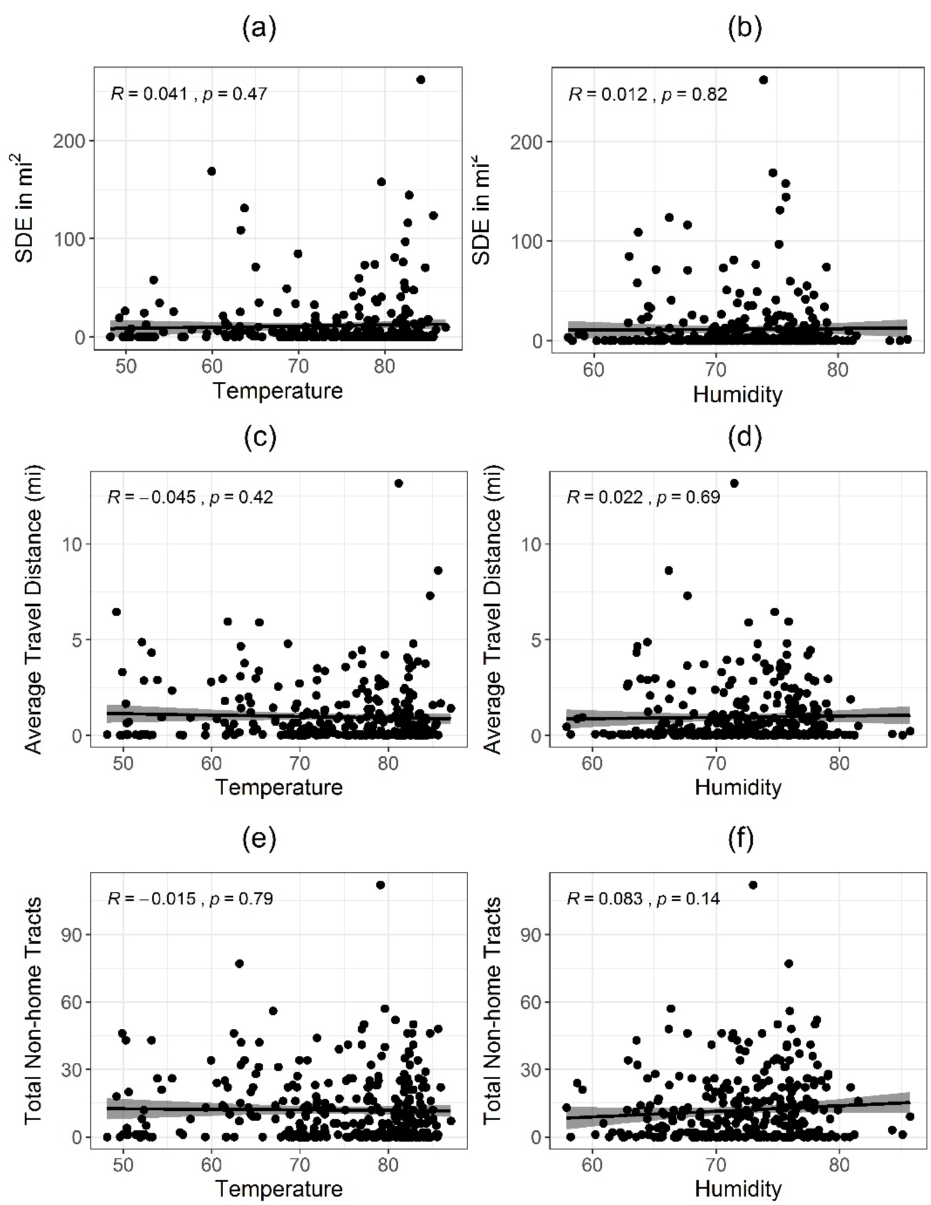
We used R version 3.6.1 to assess statistically the associations between activity spaces and weather conditions, taking into account temperature and humidity as both continuous (H1) and dichotomized variables (H2: temperatures below 50°F or above 80°F, relative humidity levels above 70 percent). Our models controlled for potential confounds at the individual level, including respondent's age; gender; race/ethnicity; employment; use of a cane, walker or wheelchair; whether travel is typically by car, and both physical and social activity levels. Besides providing common background information, the CHART baseline survey asked respondents the type and amount of physical activity involved in their daily life on a scale of 1 to 4 (1: Hardly ever or never; 2: One to three times a month; 3: Once a week; 4: More than once a week). Specifically, respondents were asked how often they take part in three types of sports or activities that are either vigorous, moderate or mild, such as running or jogging/gardening, walking at a moderate pace/vacuuming, laundry, or home repairs. We added up the self-evaluated values from these questions to create an indicator of respondent's physical activity level. In addition, the CHART baseline survey asked about the frequency of doing volunteer work/attending meetings of any organized group/getting together socially with friends or relatives/attending religious services in the past six months on a scale of 1 to 6 (1: Never; 2: About once or twice; 3: A few times in the last six months; 4: About once a month; 5: Every week; 6: Several times a week). Similarly, we added up the values from these questions to build a composite measure that reflects a respondent's typical level of social activity.

For each of the six combinations of the activity spaces and weather measures, we performed multiple regression with stepwise feature selection based on AIC to identify the most parsimonious model having the greatest explanatory power in disentangling the relationships between social and environmental attributes and the size and span of activity spaces. These models represent early, exploratory analyses of our data and do not yet incorporate sampling weights or adjustments for bias that may be due to GPS or item-level nonresponse.

Results and Discussion

Our first analysis looked at the question of whether older adults living in neighborhoods with higher temperatures and lower humidity level will have larger activity spaces than those living in relatively colder and muggy neighborhoods (H1). Figure 6 shows Pearson correlations between activity spaces and temperature and humidity in their continuous form.

Figure 6 Pearson Correlations between continuous weather measures and activity spaces



As the temperature goes up, Figure 6 shows that average daily travel distance from home and total non-home tracts visited decrease. Activity spaces also tend to be larger with moderate humidity levels. Nevertheless, all Pearson correlations were weak and not statistically significant ($p > 0.05$). We further discretized temperature and humidity to examine extreme weather conditions and their effects on activity spaces (H2). Cold or hot weather was defined as temperatures below 50 or above 80 degrees. High relative humidity levels were also coded as values above 70 percent. Figure 7 shows box plots between activity spaces and dummy weather variables with t test mean comparisons. We do not observe clear patterns of changes in activity spaces on such “extreme” weather days. However, across all three weather measures, activity spaces tend to be larger on days when humidity was high (Figure 7i, $p = 0.036$), perhaps due to a preference for driving on such unpleasantly humid or wet days.[6]

[6] Ideally, we would have a measure of precipitation as well, but AoT sensors are not equipped for measuring rain- or snowfall. Very high humidity levels can indicate precipitation, however, so such weather would be included in this analysis to some extent and might be better isolated by looking at the effect of humidity levels above 90 percent.

Figure 7. Box plots between discretized weather measures and activity spaces, with t test mean comparison p-values

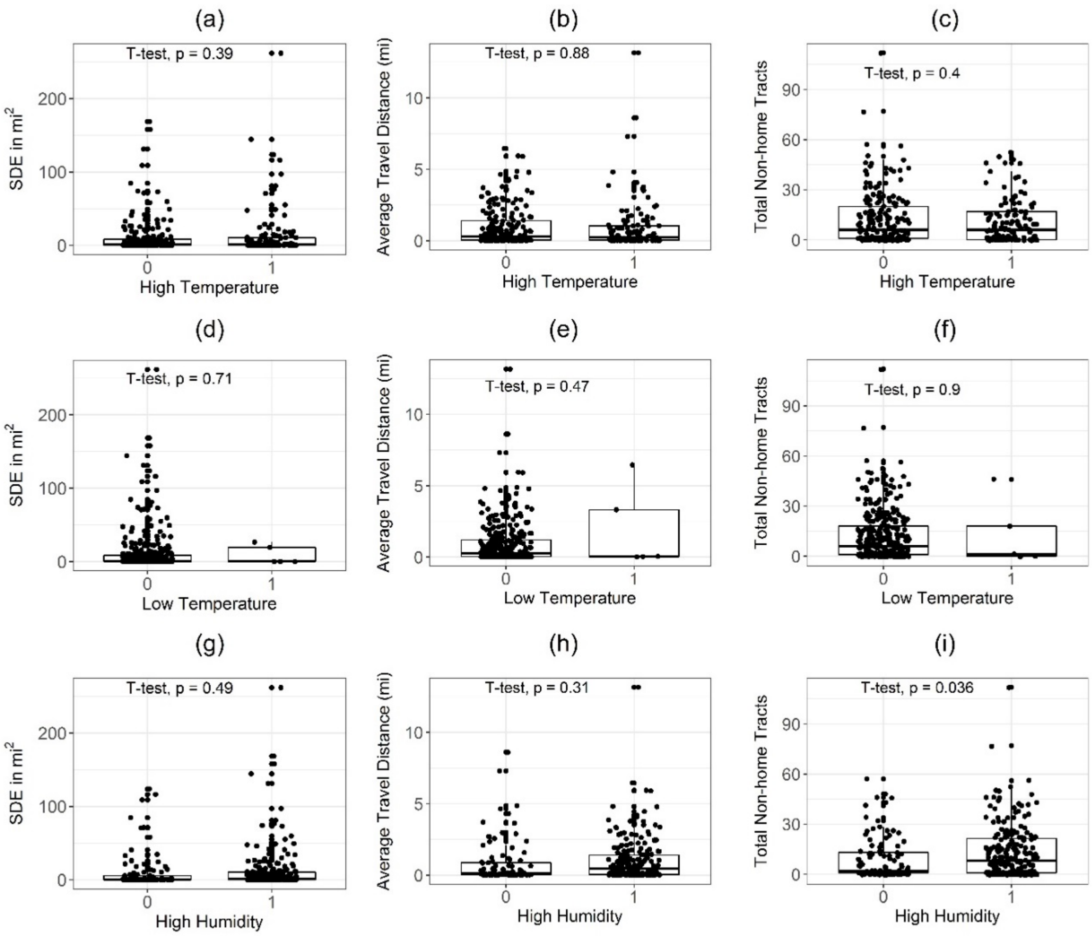


Table 2 examines the relationships between activity spaces and socio-demographic and environmental characteristics based on stepwise AIC regression. None of the continuous and dummy weather measures were retained in the most parsimonious models, suggesting that weather measures aggregated at the week level do not have enough power in explaining the variations in activity space size, after controlling for individual socio-demographic characteristics. It is possible that analyzing daily levels of activity and temperature/humidity would strengthen these associations since averaging over the entire week of GPS data collection reduces the variation in both the spaces traversed by each respondent over the week and the prevailing weather conditions for each day a respondent decides whether or not to travel outside the home.

Table 2. Relationships between CHART respondent characteristics and activity spaces, using multiple regression with stepwise feature to identify the most parsimonious model

	SDE Area		Average Travel Distance		Total Non-home Tracts	
	Estimate	P Value	Estimate	P Value	Estimate	P Value
(Intercept)	2.693	0.339	0.151	0.486	27.833**	0.003
White	8.012*	0.018				
Black					-4.504**	0.005
Hispanic			0.513*	0.011		
Age					-0.338**	0.006
Education Level			0.089*	0.042	0.774	0.067
Social Activity					0.402*	0.016
Drive/Ride in cars	9.660**	0.003	0.528**	0.003	5.236**	0.002
F-statistic	7.377***	<0.001	6.453***	<0.001	9.429***	<0.001
Adjusted R-square	0.038		0.048		0.115	

N = 324; *p < .05. **p < .01. ***p < .001; only the most parsimonious models are shown.

Although we do not find any associations between weekly activity space sizes and weekly average weather conditions, a few trends are worth noting. We found that white respondents (p = 0.018) and those who drive or ride in cars (p = 0.003) have larger activity spaces as indicated by SDE areas (Table 2). When it comes to weekly average travel distance, Hispanic respondents (p = 0.011), those with higher education level (p = 0.042), and those who drive or ride in cars (p = 0.003) travel further away from home. When examining activity spaces measured as the number of non-home tracts visited by respondents, our results suggest that respondents who are black (p = 0.005), and older adults (p = 0.006) have more constrained activity spaces. In contrast, respondents who typically drive or ride in cars (p = 0.002), and those who are more socially active (p = 0.016) spend more time outside of their residential tracts. We also note that the model using SDE to measure activity spaces yields the smallest R-square, which could be due to SDE's limited ability to represent a respondent's actual daily activity paths. Although ellipse size provides a general indicator of the span of respondents' activity spaces, it is important to note that ellipses typically include large regions that are not actually encountered by respondents.

Another important consideration is a possible misspecification of the geographic context for measuring weather or air quality exposure. The uncertain geographic context problem (UGCoP) refers to the dependence of findings on the effects of area-based environmental variables on how spatial contextual units are geographically delineated (Kwan, 2012a). The problem "arises because of the spatial uncertainty in the actual areas that exert contextual influences on the individuals being studied and the temporal uncertainty in the timing and duration in which individuals experienced these contextual influences" (Kwan, 2012b). As no researcher has complete and perfect knowledge of the "true causally relevant" geographic context, no study that uses area-based contextual variables to explain individual behaviors or outcomes can fully overcome the problem. From a spatial perspective, conventional methods often use arbitrary distance thresholds to quantify environmental exposure (e.g., the spatial buffer around the home), which ignores the "true causally relevant" geographic context that mediates the relationship between weather and activity, for instance. Using GPS trajectory data to generate better measures of exposure could help mitigate the UGCoP problem in health and related studies.

Conclusion and Implications

As noted by York Cornwell and Cagney (2019), research on aging often assumes that later life, along with the advent of health problems and functional limitations, contributes to a shrinking of activity spaces outside the home. However, an alternative possibility is that retirement may bring greater flexibility in structuring daily life. And, older adults may have unique opportunities and interest in moving beyond their residential areas to access services, organizations, and amenities, and to take part in social groups and activities (Cagney, Browning, Jackson, and Soller, 2013). Examining older adults' movements—in, out of, and across their communities—can provide insight into their span of engagement, the contexts most relevant for their health and well-being, and their access to social and community resources. Variation in activity spaces of older adults may also shed light on mechanisms of inequality in later life health and well-being. By using a variety of data sources, including GPS-enabled location tracking, the CHART project is the first study capable of answering these questions. In this paper, we add yet another technologically advanced data source in order to control for local environmental conditions that may affect activity spaces independent of health status, access to transportation, and a range of social factors.

Overall, we find that temperature and humidity do not provide additional explanatory power when trying to account for variation in activity spaces across CHART respondents during Wave 1. We note, however, that results may be biased due to limited AoT coverage in some CHART neighborhoods and to nonresponse to certain survey items or to uneven GPS data collection that together resulted in a reduction in the full CHART sample. Also, the data collection period included the warmer months of the year and thus would not be capable of assessing the impact of very cold temperatures (and snowfall) during Chicago's long winter months. Next steps include incorporating such data collected in Wave 2 in order to compare sizes of activity spaces across seasons. We also plan to extend the analysis to look for daily changes in activity patterns for each individual in the CHART sample, both within and across waves of data collection, in order to see if the impact of weather occurs within rather than across individuals and is more a function of daily changes than weekly averages. Given the ability of AoT to discern differences in weather conditions at a finer scale than has been previously available to researchers, we will assess the impact of neighborhood-level differences in temperature and humidity on activity spaces across the 10 CHART neighborhoods, should the sample sizes prove sufficient.

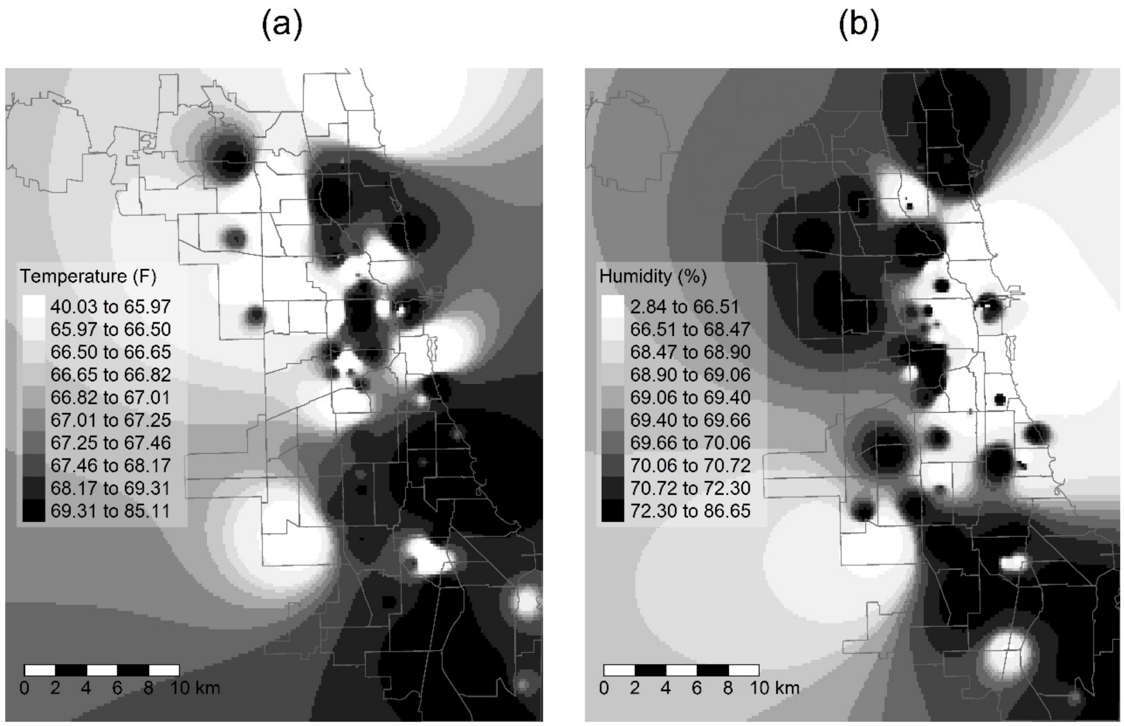
More generally, this paper aimed to explore the feasibility of using networked sensor data for social science research writ large. Indeed, we suggest that our findings highlight the potential of recent advances in sensing and computing technologies exemplified by the AoT to inform the study of any behavioral or social phenomena where environmental context matters. Some important lessons learned along the way, however, caution against proceeding too quickly. First, the theoretical link between the sensor data and the social or behavioral phenomenon being studied must be clearly specified in order to create meaningful variables at the right scale, such as "unpleasant" daily weather or long-term exposure to "polluted" air (English, Brown, & Zhao, 2020). Given the very recent installation of AoT nodes, this suggests that the data are currently best suited to studying short-term changes such as daily fluctuations in emotional states being measured by the EMA data or bouts of asthma or hospitalizations due to other illnesses that can by triggered by poor air quality or inclement weather (Silva et al., 2018). Second, the sensor data itself still needs to be assessed for quality and consistency both against industry standard devices such as EPA air quality sensors and over time as new-and-improved versions of the AoT nodes come online. Finally, the data are hierarchically structured and potentially spatially clustered, thus it is worthwhile to test the applicability of spatial hierarchical models that account for both spatial autocorrelation among observations and potential confounding factors at the individual and neighborhood levels. However, none of these challenges should ultimately prevent the Array of Things from achieving its promise of being an urban "fitness tracker" for local governments trying to craft better policy or for researchers who need environmental data at the neighborhood scale in order to better isolate their localized effect.

Funding Acknowledgement

This work was funded by the National Institutes of Health (NIH) under grant #5R01AG050605, by the National Science Foundation (NSF) under grant #1532133, and by an NORC Labs award to the primary authors. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIH or NSF.

Appendix

Figure A.1. Average temperature (a) and humidity (b) for Chicago during CHART Wave 1 data collection, interpolated using IDW from the AoT sensors (N=92)



References

Aspvik, N. P., H. Viken, J. E. Ingebrigtsen, N. Zisko, I. Mehus, U. Wisloff, & D. Stensvold, (2018). "Do Weather Changes Influence Physical Activity Level among Older Adults?—The Generation 100 Study." *PLoS One* 13: e0199463. https://doi.org/10.1371/journal.pone.0199463

Burrough, P. A. & R. A. McDonnell. (1998). Principles of Geographical Information Systems. Oxford: Oxford University Press.

Catlett, C. E., P. H. Beckman, R. Sankaran, & K. K. Galvin. (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, PA, April.

English, N., K. L. Brown, B. Curtis, C. Zhao, K. A. Cagney, & C. E. Catlett. (2019). "Linking Extant Social and Environmental Data at Multiple Scales to Surveys: Activity Space." Presentation at the 74th Annual Conference of the American Association of Public Opinion Research (AAPOR), Toronto, Canada.

Hirsch, J. A., M. Winters, P. Clarke, & H. McKay. (2014). "Generating GPS Activity Spaces that Shed Light upon the Mobility Habits of Older Adults: A Descriptive Analysis." *International Journal of Health Geographics* 13: 51-65. http://www.ij-healthgeographics.com/content/13/1/51.

Kwan, M. (2012a). "The Uncertain Geographic Context Problem." *Ann. Assoc. Am. Geogr.,* 102: 958–68.

Kwan, M. (2012b). "How GIS Can Help Address the Uncertain Geographic Context Problem in Social Science Research." *Ann. GIS,* 18: 245–55.

Pebesma, E. J. (2004). "Multivariable Geostatistics in S: the gstat package." *Computers & Geosciences*, 30: 683–91.

Potosnak, M. J., P. Banerjee, R. Sankaran, V. R. Kotamarthi, R. L. Jacob, P. H. Beckman, & C. Catlett. (2018). "Array of Things: Characterizing Low-cost Air Quality Sensors for a City-wide Instrument." Presentation at the Fall Meeting of the American Geophysical Union (AGU), Washington, DC.

Silva, M. P., A. Sharma, M. Budhathoki, R. Jain, & C. E. Catlett. (2018). "Neighborhood Scale Heat Mitigation Strategies using Array of Things (AoT) Data in Chicago." *Proceedings of the American Geophysical Union (AGU) Fall Meeting*, Washington, DC, December.

Tucker, P. & J. Gilliland. (2007). "The effect of Season and Weather on Physical Activity: A Systematic Review." *Public Health,* 121: 909-22. https://doi.org/10.016/j.puhe.2007.04.009.

Wan, N., W. Qu, J. Whittington, B. C. Witbrodt, M. P. Henderson, E. H. Goulding, S. J. Bonasera, & G. Lin. (2013). "Assessing Smart Phones for Generating Life-space Indicators." *Environment and Planning B: Planning and Design,* 40: 350-61. https://doi.org/10.1068/b38200.

York Cornwell, E., K. A. Cagney, & L. Hawkley. (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. & K. A. Cagney. (2017). "Aging in Activity Space: Results from Smartphone-based GPS-tracking of Urban Seniors." *Journals of Gerontology: Social Sciences* 72: 864-75.