

The Automatic Context Measurement Tool (ACMT): A tool to compile participant-specific built and social environment measures for health research

Weipeng Zhou, Amy Youngbloom, Xinyang Ren, Brian E. Saelens, Sean D. Mooney, Stephen J. Mooney

Submitted to: JMIR Formative Research
on: January 18, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 22

Figures 23

Figure 1..... 24

Figure 2..... 25

Figure 3..... 26

Figure 4..... 27

Figure 5..... 28

Figure 6..... 29

Figure 7..... 30

Figure 8..... 31

The Automatic Context Measurement Tool (ACMT): A tool to compile participant-specific built and social environment measures for health research

Weipeng Zhou¹; Amy Youngbloom²; Xinyang Ren¹; Brian E. Saelens^{2, 3}; Sean D. Mooney¹; Stephen J. Mooney²

¹Department of Biomedical Informatics and Medical Education University of Washington Seattle US

²Department of Epidemiology University of Washington Seattle US

³Seattle Children's Research Institute Seattle US

Corresponding Author:

Stephen J. Mooney

Department of Epidemiology

University of Washington

Hans Rosling Center for Population Health, 868

Seattle

US

Abstract

Background: The environment shapes health behaviors and outcomes. Studies exploring this influence have been limited to research groups with the geographic information systems (GIS) expertise required to develop built and social environment measures (e.g., groups that include a researcher with GIS expertise).

Objective: The goal of this study was to develop an open-source, user-friendly and privacy-preserving tool for conveniently linking built, social, and natural environmental variables to study participant addresses.

Methods: We built the Automatic Context Measurement Tool (ACMT). The ACMT comprises two components: (1) a geocoder, which identifies a latitude and longitude given a United States street address and (2) a context measure assembler, which computes measures from publicly available data sources linked to a latitude and longitude. ACMT users access both of these components using an RStudio/RShiny-based web interface that is hosted within a Docker container. We illustrate ACMT with two use cases: one comparing population density patterns within five major cities in the US, and one identifying correlates of cannabis licensure status in Washington State.

Results: Our use cases demonstrate that while Los Angeles and Seattle are roughly equally densely populated within 1000m of city hall, Los Angeles is far more densely populated at greater distances. No measured environment variables were associated with cannabis retail outlet licensure in Washington state.

Conclusions: The ACMT can be used to compile environmental measures to study the influence of environmental context on population health. The portable and flexible nature of ACMT makes it optimal for neighborhood study research seeking to attribute environmental data to specific locations within the United States.

(JMIR Preprints 18/01/2024:56510)

DOI: <https://doi.org/10.2196/preprints.56510>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://preprints.jmir.org/preprint/56510>



Original Manuscript

The Automatic Context Measurement Tool (ACMT): A tool to compile participant-specific built and social environment measures for health research

Weipeng Zhou* ^a, Amy Youngbloom* ^b, Xinyang Ren ^a, Brian E. Saelens ^{c,d}, Sean D. Mooney ^a,
Stephen J. Mooney ^b

^aDepartment of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

^bDepartment of Epidemiology, University of Washington, Seattle, WA, United States

^cSeattle Children's Research Institute, Seattle, WA, United States

^dDepartment of Pediatrics, University of Washington, Seattle, WA, United States

***co-first author**

Corresponding Author

Stephen J. Mooney

sjm2186@uw.edu

Box 351619

Department of Epidemiology

Hans Rosling Center for Population Health, 868

Seattle, WA 98195

Author contributions

Weipeng Zhou: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing- original draft, Writing- review & editing. Amy Youngbloom: Methodology, Investigation, Writing –review & editing. Xinyang Ren: Methodology, Investigation, Writing –review & editing. Brian E. Saelens: Writing – review & editing. Sean D. Mooney: Writing – review & editing. Stephen J. Mooney: Conceptualization, Funding acquisition, Methodology, Investigation, Supervision, Writing – review & editing.

Acknowledgement

This work was supported by the National Library of Medicine grant 4R00LM012868 and by the National Heart, Lung, and Blood Institute grant R01HL157166. No generative AI created content was included in this manuscript.

Conflicts of interest

None declared.

Abbreviations

ACMT: Automatic context measurement tool

GIS: geographic information systems

NE-WAS: neighborhood environment-wide association study

Abstract

Background:

The environment shapes health behaviors and outcomes. Studies exploring this influence have been limited to research groups with the geographic information systems (GIS) expertise required to develop built and social environment measures (e.g., groups that include a researcher with GIS expertise).

Objective:

The goal of this study was to develop an open-source, user-friendly and privacy-preserving tool for conveniently linking built, social, and natural environmental variables to study participant addresses.

Methods:

We built the Automatic Context Measurement Tool (ACMT). The ACMT comprises two components: (1) a geocoder, which identifies a latitude and longitude given an address (currently limited to the United States) and (2) a context measure assembler, which computes measures from publicly available data sources linked to a latitude and longitude. ACMT users access both of these components using an RStudio/RShiny-based web interface that is hosted within a Docker container, which runs on a local computer and keeps user data stored in local to protect sensitive data. We illustrate ACMT with two use cases: one comparing population density patterns within several major cities in the US, and one identifying correlates of cannabis licensure status in Washington State.

Results:

In the population density analysis, we created a line plot showing the population density (x-axis) in relation to distance from the center of city (y-axis, using city hall location as a proxy) for Seattle, Los Angeles, Chicago, New York City, Nashville, Houston and Boston with the distances being 1000, 2000, 3000, 4000 and 5000 meters. We found the population density tended to decrease as distance from city hall increased except for Nashville and Houston, two cities that are notably more sprawling than the others. New York City had a significantly higher population density than the others. We also observed that Los Angeles and Seattle had similarly low population densities within up to 2500m of City Hall. In the cannabis licensure status analysis, we gathered neighborhood measures such as age, sex, commute time and education. We found the strongest predictive characteristic of cannabis license approval to be the count of female children aged 5 to 9 and the proportion of females aged 62 to 64 who were not in the labor force. However, after accounting for Bonferroni error correction, none of the measures were significantly associated with cannabis retail license approval status.

Conclusions:

The ACMT, available at <https://github.com/2533245542/acmt-network>, can be used to compile environmental measures to study the influence of environmental context on population health. The portable and flexible nature of ACMT makes it optimal for neighborhood study research seeking to attribute environmental data to specific locations within the United States.

Keywords: Built environment, Social environment, Geocoding, GIS, ACMT, Linkage, Privacy

Word count: 4023

1. Introduction

The environment shapes health behaviors and outcomes [1]. Socio-ecological frameworks that explain behaviors and outcomes as diverse as physical activity [2], bullying [3], violence [4], sexually-transmitted infection [5], alcohol-attributable emergency department visits [6], COVID-19 vaccination [7], and sleep [8], posit that the environment in which a person lives and works affects health directly, affects health-related behavior, and may affect the ability to modify health-related behavior successfully [9].

Yet despite the robust theoretical understanding that built and social environments impact health, literature on environmental influences on health and health behavior remains inconsistent [10]. This is in part because the early evidence base for built environment associations with health has been opportunistic and cross-sectional [10], but it is also because the geographic information systems (GIS) expertise required to develop built and social environment measures at spatial and temporal scales to allow optimal comparison has limited this research to specialist research groups (e.g., groups that include a researcher with GIS expertise) [11]. Broadening the research base, not only by exploring environmental exposure hypotheses posited by research groups without specific GIS expertise, but also by studying environmental exposures over time (e.g. as in [12] and [13]), and by pooling de-identified data using contextual measures (e.g. as in [14]) requires teams to simply and efficiently link environmental measures to existing data.

To fill this gap, we developed a software tool that researchers and practitioners can use to efficiently compile environmental measures drawn from free and nationally available datasets such as the American Community Survey (ACS) [15] and the National Land Cover Database [16], and to spatially link those measures to address data. More importantly, our tool uses a locally installed geocoder and does not share addresses with third parties like Google Maps [17], thus protecting sensitive information of the research subjects. This tool, the Automatic Context Measurement Tool (ACMT), has been used in a prior study [18] and is detailed below. We describe the components that comprise the ACMT, demonstrate two use cases for the ACMT, and conclude with a discussion of the ACMT's utility for research.

2. Methods

2.1 Installation and Architecture

The ACMT and its dependencies are available on GitHub [19]. The ACMT is packaged to be run using Docker [20], a technology that lets developers package and run virtual machines within another computing environment.

The ACMT comprises two components: (1) a geocoder, which identifies a latitude and longitude given a US street address and (2) a context measure assembler, which computes measures from publicly available data sources linked to a latitude and longitude. ACMT users access both of these components using an RStudio/RShiny-based web interface that is hosted within the Docker container [21], [22].

Figure 1 below shows the flow of data in a typical use of the ACMT. First, a data analyst points their browser to the web interface (accessing the Docker container on their local machine). The R environment [23] will be running in the web interface using RStudio/RShiny, depending on analyst preference. That R environment hosts logic that can convert an address (limited to the United States) to its corresponding latitude and longitude (the geocoder) and can download public data from the

Internet to compile environmental context measures for a latitude and longitude (the context measure assembler). The address conversion is performed using the locally installed geocoder, thus preserving privacy. The analyst can then work with these context measures directly in the RStudio interface or can export them for use with other statistical software. The components are described in more detail below.

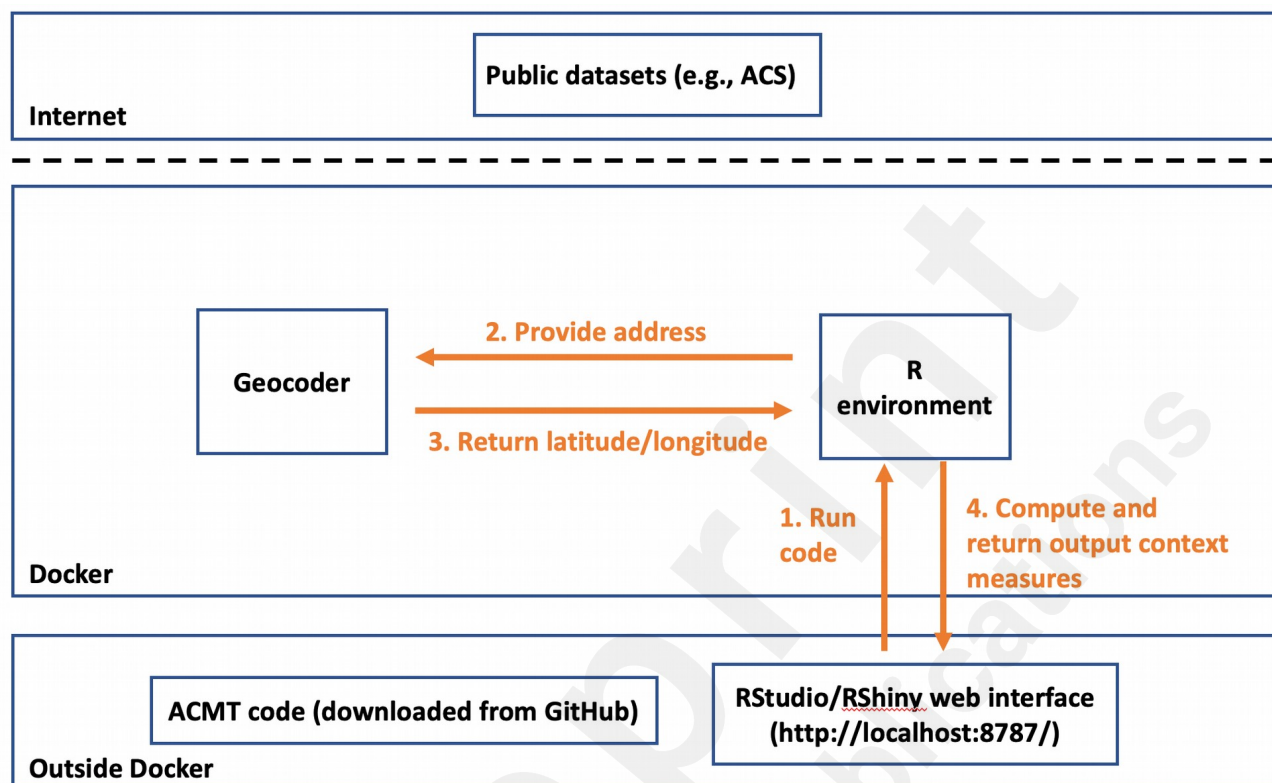


Figure 1. The ACMT architecture. The bottom box shows the ACMT code downloaded at the user's computer and the RStudio/RShiny web interface open at a local browser. The middle box shows the geocoding and R computing happening within Docker. The top box shows the internet layer where ACMT downloads public datasets (e.g., ACS) data from but ACMT would not upload sensitive information to the internet. The geocoding and context measure computing only happens within the docker container which is on the user's local computer. In a typical workflow, from the RStudio/RShiny web interface, a user sends R codes to Docker. Docker then geocodes addresses, assemble the context measures, and send outputs back to the user.

2.2 Geocoder

A geocoder is a piece of software that, given a street address, identifies the latitude and longitude corresponding to the location identified by the street address. For example, an ACMT user might write code to geocode a location (i.e., get latitude and longitude coordinates for the location) such as "600 4th Ave, Seattle, WA 98104, United States" (Seattle City Hall), and the ACMT's geocoder would supply (47.60, -122.33) as the geographic coordinates of the location corresponding to that address, as well as a rating – an indicator of how confident the geocoder is that the geographic coordinates accurately reflect the location of the street address. The rating ranges from 0 (very high confidence) to 100 (very low confidence). For addresses with lower confidence, users might consider restructuring the address or dropping them.

The ACMT's geocoder is built off of an open-source PostGIS geocoder [24] and does not share addresses with third parties (i.e., unlike Google Maps [17]). This geocoder incorporates street data from TIGER/Line files compiled by the United States Census, which provide the mapping from street names and addresses into latitude and longitude files [25].

In practice, ACMT usually processes thousands of addresses at a time. In some cases where most of

the addresses appear to have a very low confidence, it may be an issue with the installed geocoder database and it might be worth checking whether the geocoder database is installed correctly. In cases where only a few of them appear to have very low confidence, we recommend selecting those IDs in the verification checker to do manual review.

The ACMT includes a geocoding verification component that, given an address and its latitude and longitude, displays a street map of the location including a pin highlighting the geocoded location that an analyst can use to ensure the geocoder appears accurate (Figure 2). The verification tool also allows the analyst to directly edit addresses to improve geocoding accuracy by writing out abbreviated street names or fixing obvious address errors. Addresses are re-geocoded and the updated rating is shown in verification tool. Analysts can also manually update addresses on the map if they are confident of the exact location of the address and can see that the generated geocodes are incorrect. Importantly, this geocoder verification process never sends the geocoded address over the open Internet. By using Leaflet [26], an open-source mapping tool hosted in the local R environment, the host computer downloads map tiles for a general region, annotates the location locally and does not reveal a specific address. If one wishes to batch process the verification checker, we recommend layman users to contact our support team, and users with R programming experience to consider reusing the provided source code to generate and save images for each ID and view them offline.

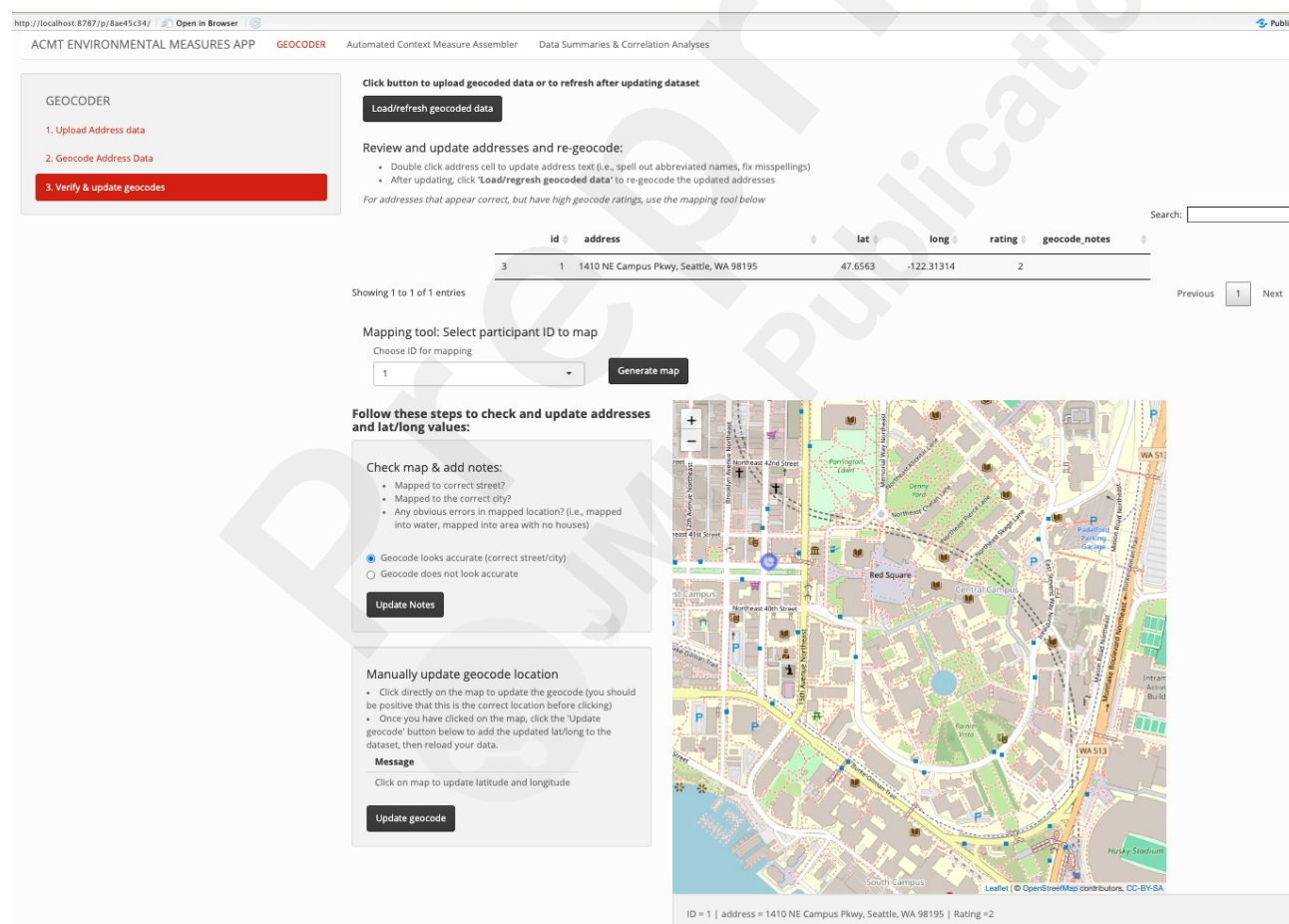


Figure 2. The geocoding verification checker from ACMT’s RShiny interface (here checking the address for “1410 NE Campus Parkway”, the address of the University of Washington).

2.3 Context measure assembler

The context measure assembler takes a location and a neighborhood radius around that location and produces standardized context measures from publicly available data regarding that location. Some

common publicly available data include measures from the American Community Survey [15] and administrative data for particular localities. Users can also compile their customized datasets and use them within ACMT. As detailed below, the ACMT supports assembling both areal and point-aggregate context measures.

Currently the ACMT only supports gathering context measures for a circular area parameterized by a coordinate and a radius. However, we do understand that users might want to gather context measures for an administrative unit such as a census tract or a block group, e.g., the number of 911 calls in the Census Tract 82 of Seattle in 2020. We leave this for future work.

2.3.1 Areal context measures

Areal context measures are defined by aggregating summary statistics that apply to pre-defined areas. For example, American Community Survey [15] measures are defined in reference to a census unit such as a tract or block groups (e.g., count of adults living in the Washington State King County tract).

The ACMT areal context measure assembler identifies the census tract or tracts corresponding to the area defined by a location and neighborhood radius (specified by the user), then uses area-weighted interpolation to compute a single average measure for that region. For example, if passed 47.60 as the latitude, -122.33 as the longitude for the location (Seattle City Hall), and 1000m as the neighborhood radius, the assembler would first identify an area defined by a 1000m circle centered at (47.60, -122.33), then compute measures reflective of that area. These measures might include the estimated count of adults living in this area, or the average proportion of that adult population who had completed high school, or the average proportion of employed adults living in that area who reported commuting by bicycle, etc..

The area-weighted interpolation that converts area tract estimates to radial buffer estimates follows a conventional area-weighting approach (Figure 3). First, given the location coordinates (i.e. latitude and longitude) and the radius, the assembler computes a buffer corresponding to a circle centered on the location with the given neighborhood radius, accounting for the earth's curvature by projecting on state plane projections as defined by the National Oceanic and Atmospheric Administration's National Geodetic Survey [27]. Next, using a spatial overlay, the assembler identifies the census tracts (or other geographical units depending on the dataset) that intersect that radial buffer and partition the radial buffer into pieces wherein each piece falls into exactly one census tract. The assembler then computes the context measure for each piece by assuming the tract-specific measure is uniform across the tract, and thus the measure for the piece of the context buffer is proportional to the overall measure for the tract. Finally, to compute a measure for the whole buffer, the assembler combines the pieces of context measures back to one either by summing or area-weighted averaging, depending on the type of measure. The context measures that are combined by summing are count-based measures, such as population and the amount of open space; the context measures that are combined by averaging are proportion-based measures, such as the proportion of population with high school diplomas and the proportion of population who commute by bike.

To illustrate, suppose the ACMT was used to estimate the population in the gray area shown in Figure 3. Suppose the circle intersects 6% area of census tract 1, 12% of census tract 2, 12% of census tract 3 and 6% of census tract 4, and suppose tract 1, 2, 3, and 4 have a population of 2000, 18000, 3000, and 1800 respectively. The population estimate assembled for the gray circle would then be $6\% * 2000 + 12\% * 1800 + 12\% * 3000 + 6\% * 1800 = 804$.

Currently, ACMT assembler is weighted by the area of the tracts. We understand users might want to have other measures for weighting the tracts such as the population density, and we leave them as future work.

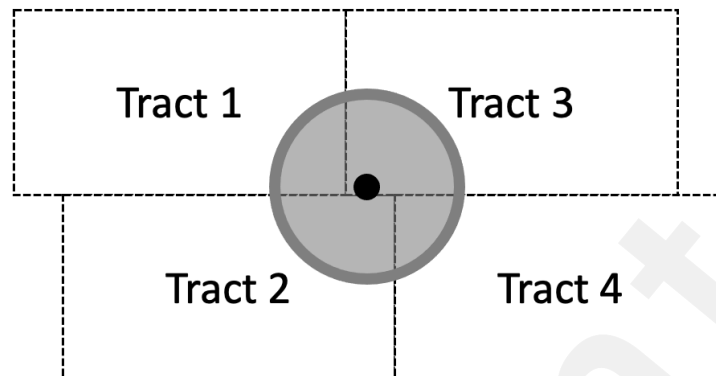


Figure 3. Illustration of area-weighted interpolation. Suppose the black dot represents the location of interest (address) and the gray circle represents neighborhood (e.g., 1000m surrounding the location of interest). Then an area-weighted interpolation represents the sum of the estimate for tract 1 weighted by the proportion of tract 1 falling in the gray circle, plus the estimate for tract 2 weighted by the proportion of tract 2 falling in the gray circle, plus the estimate for tract 3 weighted by the proportion of tract 3 falling in the gray circle, and plus the estimate for tract 4 weighted by the proportion of tract 4 falling in the gray circle.

2.3.2 Point-aggregate context measures

Whereas areal context measures aggregate data representing areas, point aggregate context measures assemble areal measures from data recorded at points, wherein each point has its own corresponding latitude and longitude. For example, addresses from which 911 calls originated and the distribution of Airbnb [28] short-term rentals in the United States are point measures.

In assembling such measures, ACMT uses point-interpolation to estimate the number of instances that fall within the radial buffer. As an example, a common usage of ACMT is to estimate the number of 911 calls in an area. In Figure 4, the red dots are the location of 911 calls by longitude and latitude, and we want to estimate the number of 911 calls that occur within the radial buffer. ACMT will overlay the radial buffer on the coordinates and estimate the number of 911 calls that fall into it. In this example, four 911 calls happened within the specified area.

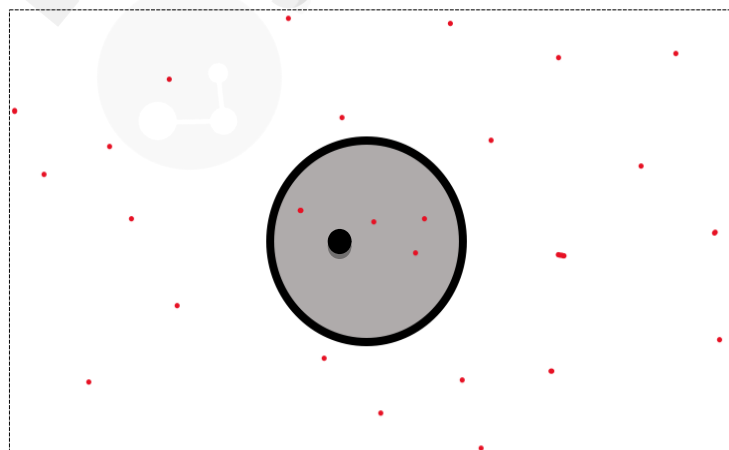


Figure 4. Point-aggregate interpolation. Suppose the black dot represents the location of interest, the gray dot represents the neighborhood, and the red dots indicate 911 calls. A point-aggregate sum would indicate that four 911 calls fell within the neighborhood area.

2.4 Datasets

Many datasets are publicly available and spatially precise. The ACMT has built-in code to make common data sources available (Table 1), and also provide easily modified code templates for custom datasets.

Dataset	Description	Context measure type	Geographical unit
American Community Survey (ACS) [15]	American Community Survey measures of the United States Census Bureau	Areal	Census Tract
Modified Retail Food Environment Index (mRFEI) [29]	Health-related index as measured by the number of healthy and less healthy food retailers in the United States	Areal	Census Tract
Smart Location Database Walkability Index [30]	The walkability scores in the United States	Areal	Block Group
National Land Cover Database (NLCD) [16]	The land cover statistics of the United States	Areal	30m Grid across the US
Seattle 911 Call [31]	911 calls in Seattle	Point-aggregate	Coordinate
Seattle Crime [32]	Crime incidents in Seattle	Point-aggregate	Coordinate

Table 1. Datasets that have been successfully integrated for use with the ACMT.

2.5 ACMT User Interface

The ACMT has two levels of user interface, adjusting for users with different backgrounds. The first, which is the most flexible, offers access to ACMT features via RStudio hosted within the data analyst's locally installed Docker instance (Figure 5). Using the RStudio interface, a data analyst can write R code that uses ACMT functions for geocoding addresses and request environmental measures. With this interface, a user can upload environmental datasets to interpolate for their geocoded datasets, use publicly available datasets online, or use one of the environmental measures datasets that are already set up in the ACMT functions (American Community Survey, Modified Food Retail Environment Index, National Land Cover Database, CDC Places, Walkability Index, etc.).

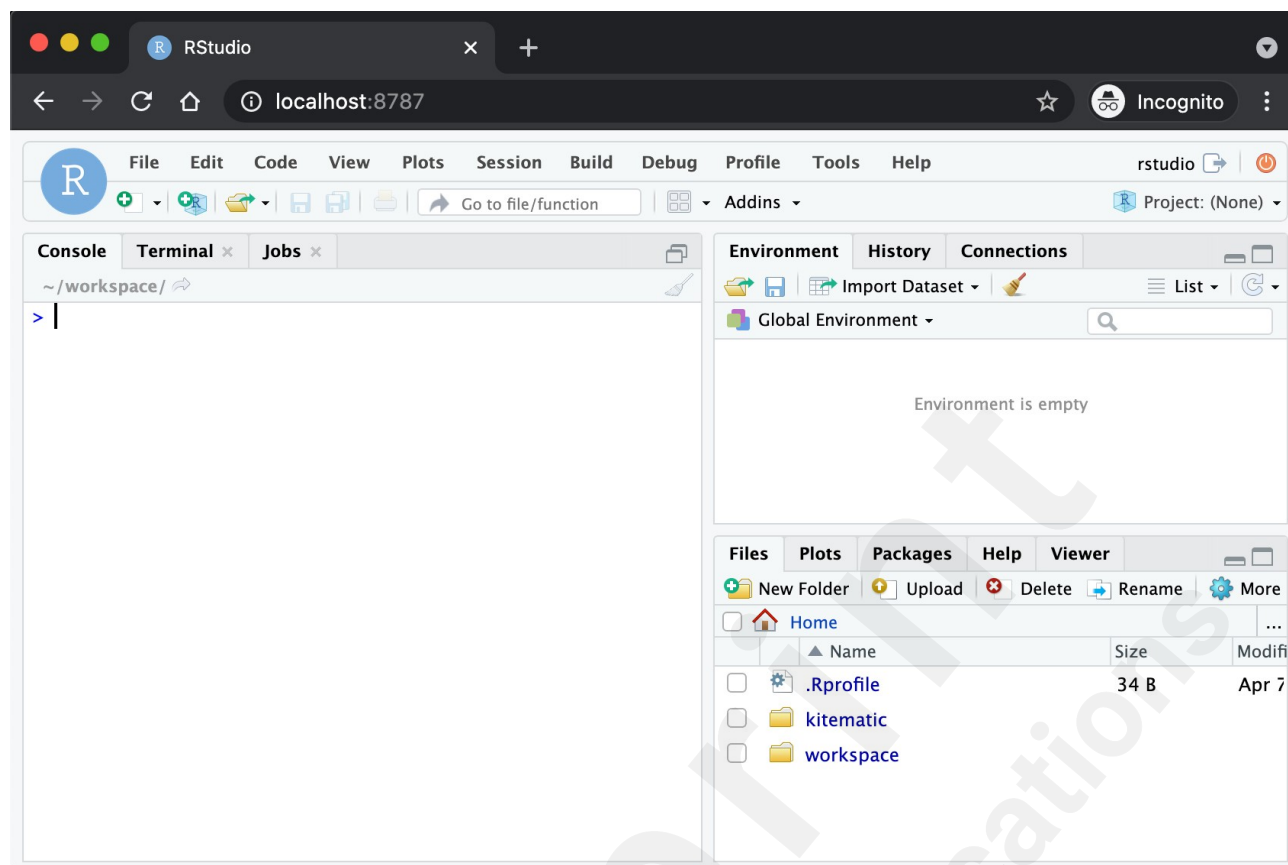


Figure 5. ACMT RStudio user interface. An analyst can use conventional RStudio features, including the editor and debugger, to access the ACMT geocoder and the context measure assembler.

The second interface, which is more friendly to layman users and avoids requiring the data analyst to write R code, uses RShiny-based helper screens to guide analysts through a process by which they open a file of locations (e.g., street addresses), geocode them, specify a dataset, a neighborhood radius and years of data to spatially link, allowing the ACMT to perform the appropriate data download, spatial linkage, and interpolation for the analyst's dataset of interest. Currently the RShiny interface is limited to these publically available datasets: American Community Survey, Walkability, Modified Retail Food Environment Index, National Land Cover Database, etc.. Figure 6 shows the RShiny interface for linking to American Community Survey data.

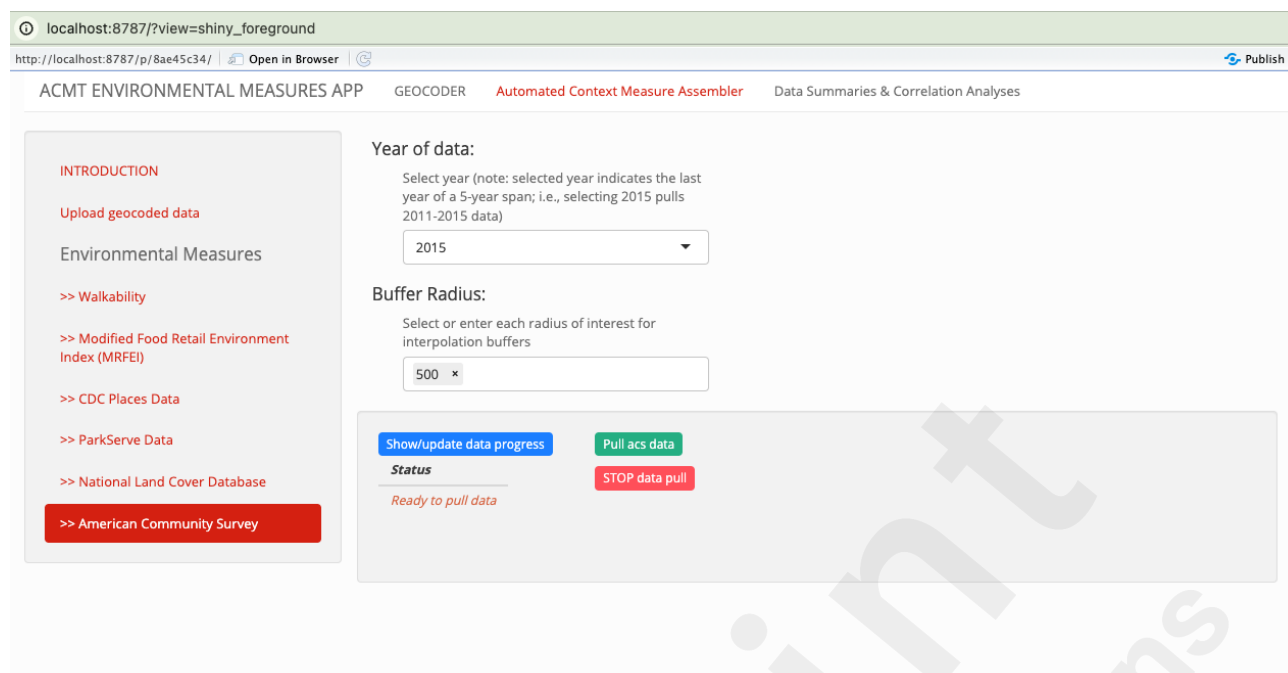


Figure 6. ACMT RShiny user interface for analysis using the American Community Survey dataset. An analyst can select parameters visually for performing analysis.

2.6 Worked Examples of ACMT Usage

2.6.1 Population density in relation to distance from city hall

Suppose a researcher is interested in understanding the spatial scale at which residential population density drops off as distance increases from city hall: the shape of such a curve might be a rough indicator of how much a metropolitan area is affected by urban sprawl, wherein less sprawling metro areas might see population density drop off more quickly. The ACMT would be suited to this task.

Specifically, we select Seattle, Los Angeles, Chicago, New York City, Nashville, Houston and Boston to represent large metropolitan areas with different histories and different geographic constraints on urban growth. We first use the ACMT geocoder to geocode the addresses of the respective city halls in these central cities. We then select 1000m, 2000m, 3000m, 4000m and 5000m as radial distances to compare. Using the ACMT, we compute the population per square meter of land area (i.e., to eliminate water) of each region within each radial distance and plot the density-to-distance-from-city-hall curve for each city and compare. The detailed steps are included in ACMT as an example.

2.6.2 NE-WAS to assess correlates of cannabis retail license status

A second use case for the ACMT might be to compile measures for a Neighborhood Environment-Wide Association Study (NE-WAS), a study that assesses the strength of association between neighborhood measures and some outcome of interest [18], [33], [34]. While NE-WAS studies often focus on residential environments, we provide an example NE-WAS using public addresses so that the code for the example could be shared as part of the ACMT code download. An analyst can download the publicly available Seattle dispensary address data and run the full example for reference. In this example, we assess characteristics of neighborhoods around Washington State cannabis retail license applicants and their association with license approval status.

Specifically, we used the ACMT to compile measures of the built and social environment from the

American Community Survey within 1000m of each cannabis retail applicant location. First, using ACMT geocoder, we geocoded the address provided with each application to identify a latitude and longitude corresponding to its proposed retail location. Next, using ACMT context measure assembler, we computed about 100 neighborhood environment measures for each location. We computed logistic regression models for each characteristic, predicting licensure status (i.e. active and still open vs. permanently closed) for cannabis applicants from this list. Consistent with best practices for multiple tests and hypothesis generation, we used conservative Bonferroni-adjusted alpha to account for the large number of tests conducted. Finally, we plotted the strength of association identified in each model between its specific environment measure and cannabis outlet licensure status, thereby visually assessing which features were more strongly associated with licensure. The detailed steps are included in ACMT as an example.

2.7 Ethical Considerations

This study does not involve human subjects and it does not require the approval from Institutional Review Board (IRB). All the data used in this study were collected by the third-party and the datasets were publicly released. This study also used de-identified data only. No identifiable information of individual participants/users was included in this manuscript.

3. Results

3.1. Population density in relation to distance from city hall

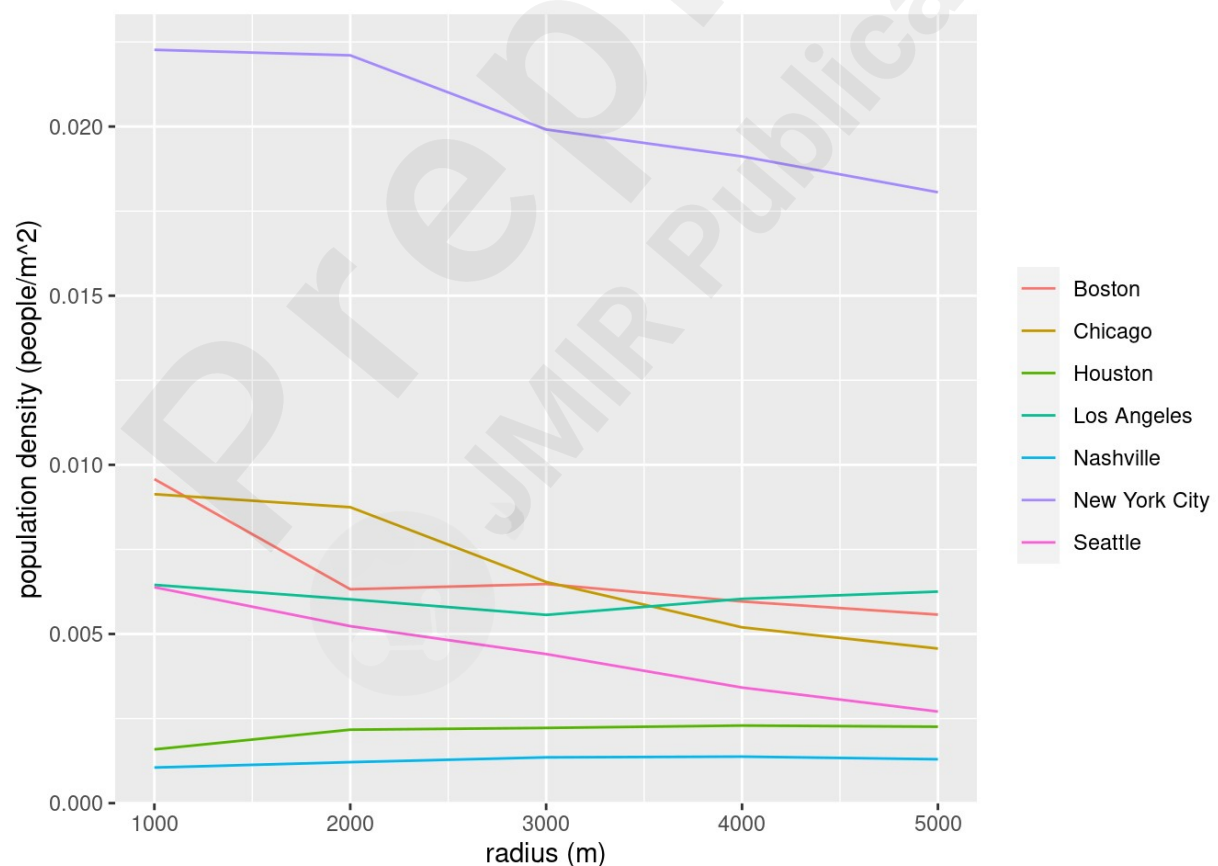


Figure 7. The population density in relation to distance from the city hall for Seattle, Los Angeles, Chicago, New York City, Nashville, Houston and Boston with the distances being 1000, 2000, 3000, 4000 and 5000 meters.

As shown in Figure 7, population density per land area across tended to decrease as distance from

city hall increased except for Nashville and Houston, two cities that are notably more sprawling than the others. The population density of Nashville is around 0.001 across all distances and the population density of Houston is around 0.002 across all distances.

Other cities all have population densities above 0.002 for all distances. The population density of Seattle is around 0.006 at 1000 meters away from the city hall, and it drops to around 0.002 at 5000 meters away from the city hall, dropping by more than a half. In comparison, the population density of Los Angeles is also around 0.006 at 1000 meters away from the city hall, and it remains similar at 5000 meters away from the city hall. This could be that Los Angeles is a larger metropolitan area than Seattle and its population is more evenly spread throughout its urban core. Besides, Seattle is a city surrounded by water and this geographical characteristic makes it difficult to expand efficiently, while Los Angeles is a land-based city and this geographical characteristic makes it easier to expand.

Chicago and Boston both have a similar population density of around 0.01 at 1000 meters from city hall, but they experience different trends when getting further from the city hall. Chicago has a steady drop in population density from around 0.01 to around 0.005. Boston has a quick population drop at 2000 meters, dropping from 0.01 to 0.005, but remains relatively stable afterward. This could be because the Boston City Hall is located at a business center on a peninsula that has a radius of around 2000 meters, so the population drops suddenly when the population density calculation starts to cover lands outside the peninsula. The Chicago City Hall is near Lake Michigan. It has a land area on the west side and a water area on the east side, thus resulting in a relatively stable population density when we increase the distance from 1000 meters to 5000 meters.

As expected, New York City has a significantly higher population density than all the other cities, with a population density of around 0.023 at 1000 meters and a population density of around 0.017 at 5000 meters. The population density remains steady at 1000 meters and 2000 meters away from the city hall, but it has a sudden drop at the distance of 3000 meters, and this could be because this distance is reaching outside of Manhattan and starting to cover the less dense areas such as Jersey City and Brooklyn.

3.2. NE-WAS to assess correlates of cannabis retail license status

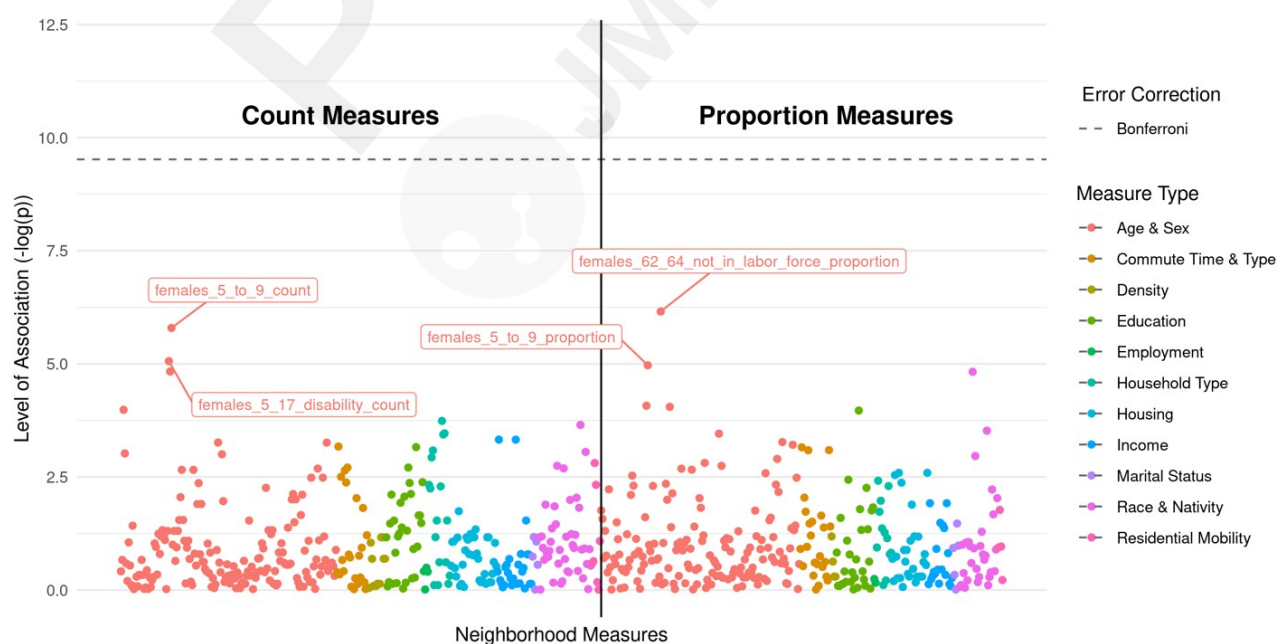


Figure 8. Manhattan plot of regression coefficients for the association between neighborhood characteristics and cannabis license status (e.g., active and open vs. closed). Each dot represents one environment measure, and the Y axis represents the strength of association. Measures are grouped by the domain of environment measured and by whether the measure was a count (e.g., # of residents within 1000m of an outlet who were females aged 5 to 9) or proportion (e.g., % of residents within 1km of an outlet who were females aged 5 to 9).

In our NE-WAS study (Figure 8), a few environmental measures have a level of association greater than the others. For the count measures, the two measures that have the highest level of association are females aged 5 to 9, with a level of association of around 6, and females aged 5 to 7 with disability, with a level of association of around 5. For the proportion measures, the two measures that have the highest level of association are females aged 62 to 64 who are not in the labor force, with a level of association of around 6, and females aged 5 to 9, with a level of association of around 6. We observed that female aged 5 to 9 are both among the highest associated count or proportion measures. The strongest predictive characteristic of cannabis license approval is the count of female children aged 5 to 9 and the proportion of females aged 62 to 64 who are not in the labor force. Regulations for dispensaries in Seattle require licensed retailers to be at least 1,000 feet from schools and playgrounds and at least 500 feet from child-care centers, recreation centers, libraries, parks or transit center [31], which may explain the association that we see between females aged 5 to 9 and females aged 5 to 17 with disabilities and licensure approval. Nevertheless, after accounting for Bonferroni error correction, none of the measures were found to be significantly associated with cannabis retail license approval status.

4. Discussion

We developed the ACMT, a tool that compiles environmental measures to study the influence of environmental context on population health. We then demonstrated its use to explore population density in relation to distance to town hall and environmental correlates of cannabis retail license status. These examples, for which code is available in the online supplement, illustrate how the ACMT can be used to construct geospatial datasets without sending identifying information on the open Internet. The ACMT is available for download at <https://github.com/2533245542/acmt-network>.

The ACMT is distinct from other neighborhood measurement approaches for studies of health and place. Whereas geographic information systems like ArcGIS [35] and QGIS are general purpose tools for geospatial data management and analysis, the ACMT is a streamlined system designed for the specific purpose of computing environmental measures specific to a location (longitude and latitude). Yet unlike other low-cost alternatives such as entering subject addresses into Google Maps [17], the ACMT approach preserves subject privacy and compliance with typical institutional review board concerns by ensuring this measure computation happens fully within the analyst's local computer [32], [33].

The ACMT has several key strengths. It is free for use, with a team of researchers dedicated to making it serve the research needs of potential users. It allows for simple, privacy-preserving assembly of measures from public data sets, unlocking the potential to better understand place-based disparities in health and health behavior. Additionally, the user-friendly RStudio and RShiny interfaces allow for simple paste-code/edit-code or point-and-click interactions that do not require analysts to parse complicated GIS concepts like data layers or projections.

However, the ACMT also comes with limitations. The Docker-based packaging system, though necessary to ensure access to a geocoder while also ensuring study participant data does not leave the local machine, can be an installation challenge in tightly locked-down environments. For example,

we found in early deployments of the ACMT that some information technology systems in some academic medical center settings set firewalls to preclude the Docker instance from running a webserver for local access. While these issues can be worked around with IT support, they are nonetheless a challenge. Furthermore, as the ACMT is not a general-purpose GIS system, complex geospatial operations including alternate interpolation strategies or road network buffers are not supported at this time. Finally, the ACMT's utility is intrinsically tied to the quality of available data – if a researcher's environmental constructs are not well captured by public data, the ACMT cannot help assessing it.

5. Conclusion

ACMT is a user-friendly piece of software that runs in an isolated environment and can be easily installed for Windows, MacOS and Linux. Users with no R coding experience can utilize the RShiny interface to pull commonly-used measures, while more advanced R users can use the ACMT code package an RStudio interface to pull from any available environmental datasets. Both approaches avoid sending personally identifying information to third parties. The ACMT thus unlocks access to contextual measures for researchers without geographic information systems training, which in turn both enables individual research teams to explore context in relation to health outcomes and allows for larger consortia to pool datasets with contextual information while preserving privacy.

6. Data statement

The ACMT is open-source software and is available on GitHub at <https://github.com/2533245542/acmt-network> [19].

7. References

- [1] R. J. Jackson, "The impact of the built environment on health: an emerging field," *Am J Public Health*, vol. 93, no. 9, pp. 1382–1384, Sep. 2003, doi: 10.2105/ajph.93.9.1382.
- [2] B. Giles-Corti, A. Timperio, F. Bull, and T. Pikora, "Understanding physical activity environmental correlates: increased specificity for ecological models," *Exerc Sport Sci Rev*, vol. 33, no. 4, pp. 175–181, Oct. 2005, doi: 10.1097/00003677-200510000-00005.
- [3] D. L. Espelage and S. M. Swearer, Eds., "A social-ecological model for bullying prevention and intervention: Understanding the impact of adults in the social ecology of youngsters.,", in *Handbook of bullying in schools: An international perspective.*, New York, NY, US: Routledge/Taylor & Francis Group, 2010, pp. 61–72. doi: 10.4324/9780203842898.
- [4] D. J. Whitaker, D. M. Hall, and A. L. Coker, "Primary prevention of intimate partner violence: Toward a developmental, social-ecological model.,", *Intimate partner violence: A health-based perspective.*, pp. 289–305, 2009.
- [5] R. J. DiClemente, L. F. Salazar, R. A. Crosby, and S. L. Rosenthal, "Prevention and control of sexually transmitted infections among adolescents: the importance of a socio-ecological perspective--a commentary," *Public Health*, vol. 119, no. 9, pp. 825–836, Sep. 2005, doi: 10.1016/j.puhe.2004.10.015.
- [6] S. M. Forbes, N. Schwartz, S. H. Fu, E. Hobin, and B. T. Smith, "The association between off- and on-premise alcohol outlet density and 100% alcohol-attributable emergency department visits by neighbourhood-level socioeconomic status in Ontario, Canada," *Health Place*, vol. 89, p. 103284, Jun. 2024, doi: 10.1016/j.healthplace.2024.103284.
- [7] K. M. Brown, J. Lewis-Owona, S. Sealy-Jefferson, A. Onwuka, and S. K. Davis, "Still Separate, Still Not Equal: An Ecological Examination of Redlining and Racial Segregation with COVID-19 Vaccination Administration in Washington D.C.," *J Urban Health*, Jun. 2024, doi: 10.1007/s11524-024-00862-y.

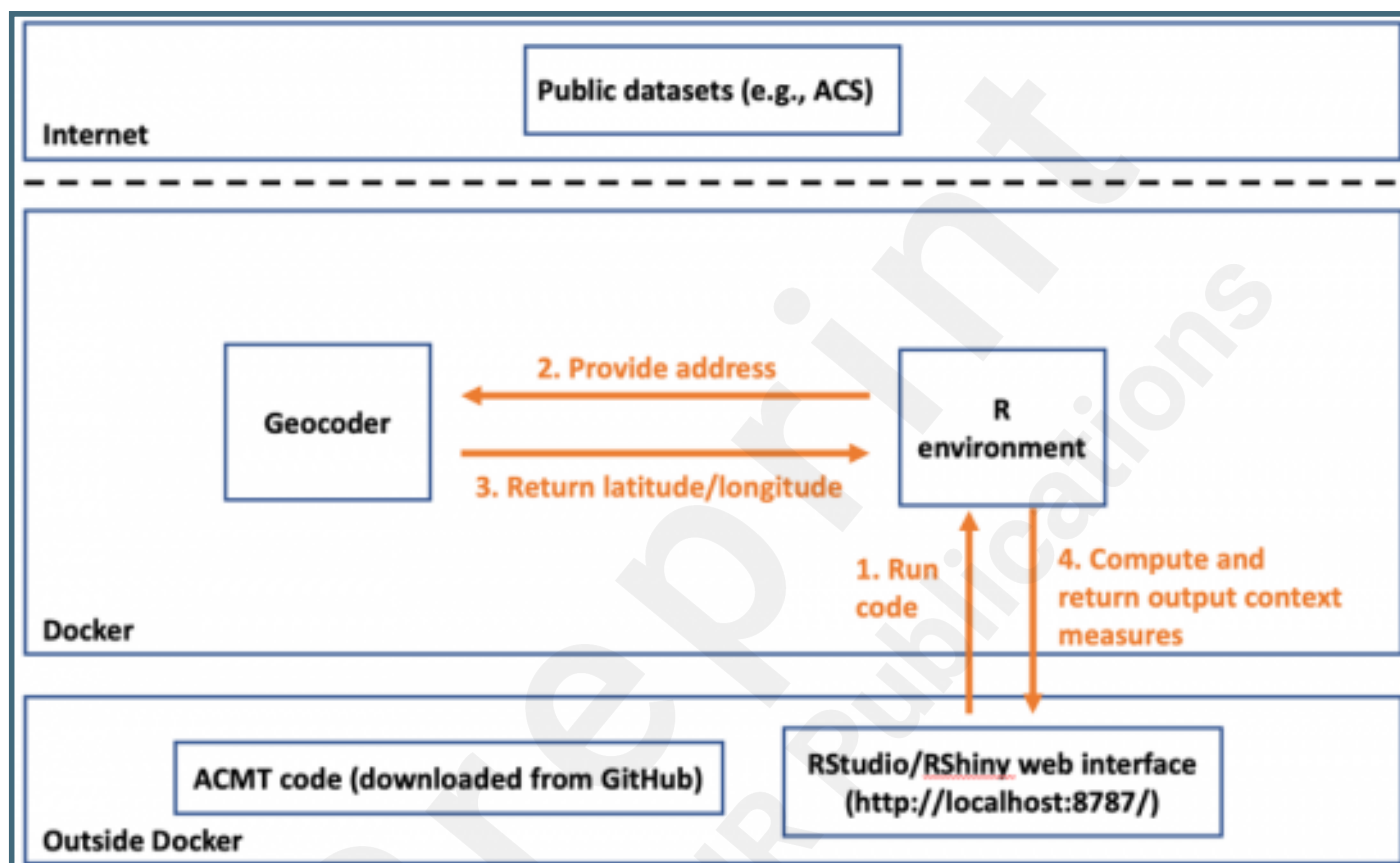
- [8] M. A. Grandner, "Chapter 5 - Social-ecological model of sleep health," in *Sleep and Health*, M. A. Grandner, Ed., Academic Press, 2019, pp. 45–53. doi: 10.1016/B978-0-12-815373-4.00005-8.
- [9] D. Stokols, "Translating social ecological theory into guidelines for community health promotion," *Am J Health Promot*, vol. 10, no. 4, pp. 282–298, 1996, doi: 10.4278/0890-1171-10.4.282.
- [10] S. Jivraj, E. T. Murray, P. Norman, and O. Nicholas, "The impact of life course exposures to neighbourhood deprivation on health and well-being: a review of the long-term neighbourhood effects literature," *Eur J Public Health*, vol. 30, no. 5, pp. 922–928, Oct. 2020, doi: 10.1093/eurpub/ckz153.
- [11] U. Bilal, A. H. Auchincloss, and A. V. Diez-Roux, "Neighborhood Environments and Diabetes Risk and Control," *Curr Diab Rep*, vol. 18, no. 9, p. 62, Jul. 2018, doi: 10.1007/s11892-018-1032-2.
- [12] C. Kamis, W. Xu, A. Schultz, K. Malecki, and M. Engelman, "Linking sequences of exposure to residential (dis)advantage, individual socioeconomic status, and health," *Health Place*, vol. 88, p. 103262, Jul. 2024, doi: 10.1016/j.healthplace.2024.103262.
- [13] M. D. Schiff *et al.*, "Longitudinal Exposure to Neighborhood Concentrated Poverty Contributes to Differences in Adiposity in Midlife Women," *J Womens Health (Larchmt)*, Jul. 2024, doi: 10.1089/jwh.2023.1156.
- [14] P. Meijer *et al.*, "The association of obesogenic environments with weight status, blood pressure, and blood lipids: A cross-sectional pooled analysis across five cohorts," *Environ Res*, vol. 256, p. 119227, Sep. 2024, doi: 10.1016/j.envres.2024.119227.
- [15] N. R. Council and others, *Using the American Community Survey: benefits and challenges*. 2007.
- [16] "National Land Cover Database | U.S. Geological Survey." Accessed: Sep. 27, 2023. [Online]. Available: <https://www.usgs.gov/centers/eros/science/national-land-cover-database>
- [17] H. Mehta, P. Kanani, and P. Lande, "Google maps," *International Journal of Computer Applications*, vol. 178, no. 8, pp. 41–46, 2019.
- [18] S. J. Mooney, P. M. Hurvitz, A. V. Moudon, C. Zhou, R. Dalmat, and B. E. Saelens, "Residential neighborhood features associated with objectively measured walking near home: Revisiting walkability using the Automatic Context Measurement Tool (ACMT)," *Health Place*, vol. 63, p. 102332, May 2020, doi: 10.1016/j.healthplace.2020.102332.
- [19] 2533245542, "2533245542/acmt-network." Jan. 06, 2022. Accessed: Jan. 17, 2024. [Online]. Available: <https://github.com/2533245542/acmt-network>
- [20] C. Anderson, "Docker [Software engineering]," *IEEE Software*, vol. 32, no. 3, pp. 102–c3, Jun. 2015, doi: 10.1109/MS.2015.62.
- [21] J. S. RACINE, "RSTUDIO: A PLATFORM-INDEPENDENT IDE FOR R AND SWEAVE," *Journal of Applied Econometrics*, vol. 27, no. 1, pp. 167–172, 2012.
- [22] W. Chang *et al.*, *shiny: Web Application Framework for R*. 2023.
- [23] M. L. Rizzo, *Statistical computing with R*. CRC Press, 2019.
- [24] P. Corti, T. J. Kraft, S. V. Mather, and B. Park, *PostGIS cookbook*. Packt Publishing Ltd, 2014.
- [25] R. E. Klosterman and A. A. Lew, "TIGER Products for Planning," *Journal of the American Planning Association*, vol. 58, no. 3, pp. 379–385, Sep. 1992, doi: 10.1080/01944369208975817.
- [26] P. Crickard III, *Leaflet. js essentials*. Packt Publishing Ltd, 2014.
- [27] N. O. and A. A. US Department of Commerce, "National Geodetic Survey." Accessed: Sep. 27, 2023. [Online]. Available: <https://geodesy.noaa.gov/>
- [28] D. Guttentag, "Progress on Airbnb: a literature review," *Journal of Hospitality and Tourism Technology*, vol. 10, no. 4, pp. 814–844, 2019.
- [29] "Census tract level state maps of the modified food environment index (mRFEI)," Atlanta, GA,

- Apr. 12, 4AD. [Online]. Available: <https://stacks.cdc.gov/view/cdc/61367>
- [30] O. US EPA, "Smart Location Mapping." Accessed: Sep. 27, 2023. [Online]. Available: <https://www.epa.gov/smartgrowth/smart-location-mapping>
- [31] "Seattle Real Time Fire 911 Calls | City of Seattle Open Data portal." Accessed: Sep. 27, 2023. [Online]. Available: <https://data.seattle.gov/Public-Safety/Seattle-Real-Time-Fire-911-Calls/kzjm-xkqj>
- [32] "SPD Crime Data: 2008-Present | City of Seattle Open Data portal," Seattle. Accessed: Sep. 27, 2023. [Online]. Available: <https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5>
- [33] S. J. Mooney, S. Joshi, M. Cerdá, G. J. Kennedy, J. R. Beard, and A. G. Rundle, "Contextual Correlates of Physical Activity among Older Adults: A Neighborhood Environment-Wide Association Study (NE-WAS)," *Cancer Epidemiol Biomarkers Prev*, vol. 26, no. 4, pp. 495–504, Apr. 2017, doi: 10.1158/1055-9965.EPI-16-0827.
- [34] S. M. Lynch *et al.*, "A Neighborhood-Wide Association Study (NWA): Example of prostate cancer aggressiveness," *PLoS One*, vol. 12, no. 3, p. e0174548, 2017, doi: 10.1371/journal.pone.0174548.
- [35] T. Ormsby and L. Feaster, *Getting to know ArcGIS desktop: basics of ArcView, ArcEditor, and ArcInfo*. ESRI, Inc., 2004.
- [36] A. G. Rundle, M. D. M. Bader, and S. J. Mooney, "The Disclosure of Personally Identifiable Information in Studies of Neighborhood Contexts and Patient Outcomes," *J Med Internet Res*, vol. 24, no. 3, p. e30619, Mar. 2022, doi: 10.2196/30619.
- [37] M. D. M. Bader, S. J. Mooney, and A. G. Rundle, "Protecting Personally Identifiable Information When Using Online Geographic Tools for Public Health Research," *Am J Public Health*, vol. 106, no. 2, pp. 206–208, Feb. 2016, doi: 10.2105/AJPH.2015.302951.

Supplementary Files

Figures

The ACMT architecture. The bottom box shows the ACMT code downloaded at the user's computer and the RStudio/RShiny web interface open at a local browser. The middle box shows the geocoding and R computing happening within Docker. The top box shows the internet layer where ACMT downloads public datasets (e.g., ACS) data from but ACMT would not upload sensitive information to the internet. The geocoding and context measure computing only happens within the docker container which is on the user's local computer. In a typical workflow, from the RStudio/RShiny web interface, a user sends R codes to Docker. Docker then geocodes addresses, assemble the context measures, and send outputs back to the user.



The geocoding verification checker from ACMT's RShiny interface (here checking the address for "1410 NE Campus Parkway", the address of the University of Washington).

ACMT ENVIRONMENTAL MEASURES APP **GEOCODER** Automated Content Measures Assembler Data Summaries & Correlation Analyses

Click button to upload geocoded data or to refresh after updating dataset

Load/refresh geocoded data

Review and update addresses and re-geocode:

- Double click address cell to update address tool (i.e., spell out abbreviated names, fix misspellings)
- After updating, click "Load/refresh geocoded data" to re-geocode the updated addresses

For addresses that appear correct, but have high geocode ratings, use the mapping tool below

id	address	lat	long	rating	geocode_notes
8	1410 NE Campus Pkwy, Seattle, WA 98195	47.6103	-122.31914	2	

Showing 1 of 1 entries

Mapping tool: Select participant ID to map

Choose ID for mapping: **Generate map**

Follow these steps to check and update addresses and lat/long values:

Check map & add notes:

- Mapped to correct street?
- Mapped to the correct city?
- Any obvious errors in mapped location? (i.e., mapped into water, mapped into area without houses)

☒ Geocode looks accurate (correct street/city)
☐ Geocode does not look accurate

Update Notes

Manually update geocode location

- Click directly on the map to update the geocode (you should be positive that this is the correct location before clicking)
- Once you have clicked on the map, click the "Update geocode" button below to add the updated lat/long to the dataset, then reload your data.

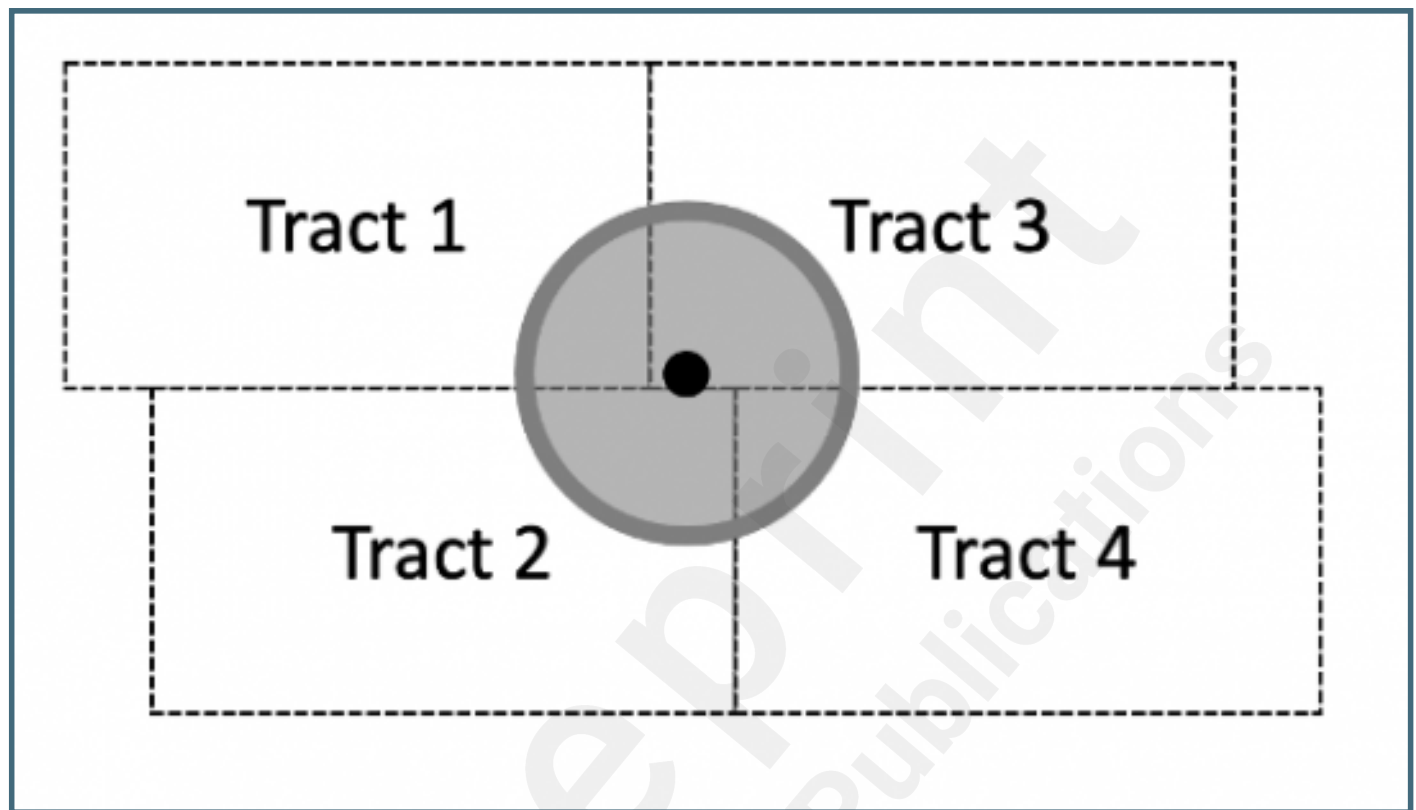
Message:

Click on map to update latitude and longitude

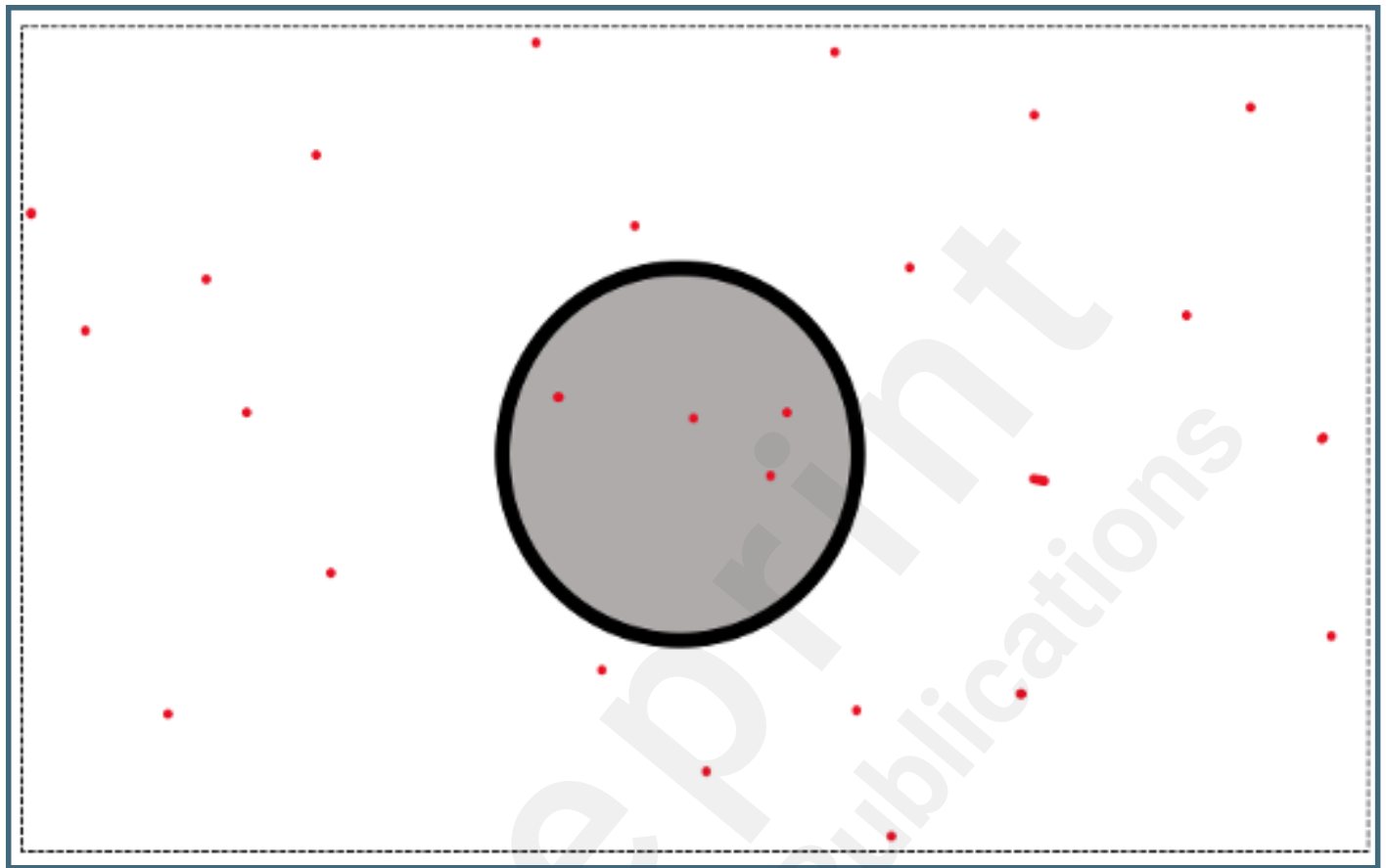
Update geocode

Map showing Seattle area with geocoded locations marked. Address: 1410 NE Campus Pkwy, Seattle, WA 98195. Rating: 2

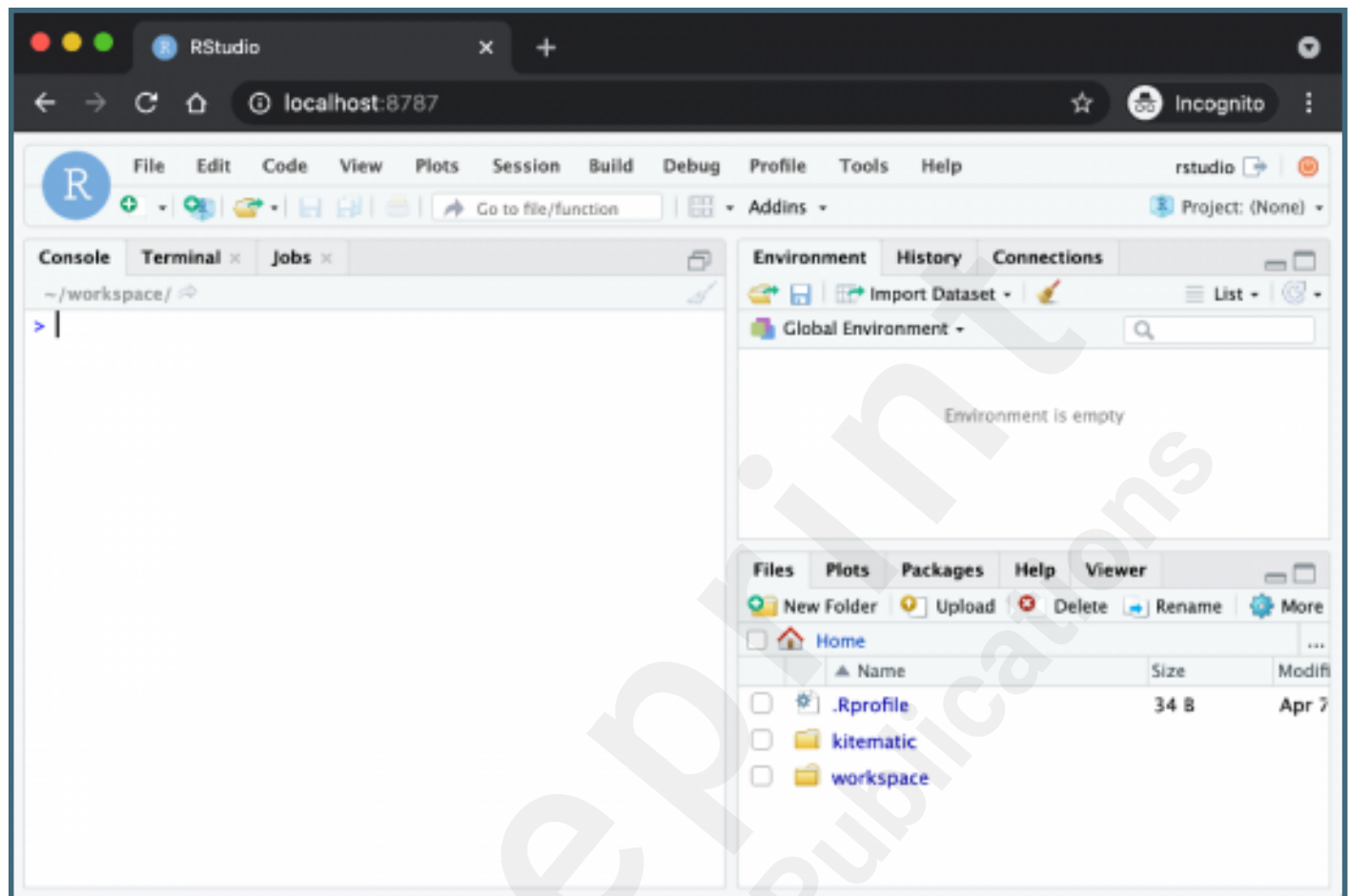
Illustration of areal-weighted interpolation. Suppose the black dot represents the location of interest and the gray circle represents the area of interest (e.g., 1000m surrounding the location of interest). Then an area-weighted interpolation represents the sum of the estimate for tract 1 weighted by the proportion of tract 1 falling in the gray circle, plus the estimate for tract 2 weighted by the proportion of tract 2 falling in the gray circle, plus the estimate for tract 3 weighted by the proportion of tract 3 falling in the gray circle, and plus the estimate for tract 4 weighted by the proportion of tract 4 falling in the gray circle.



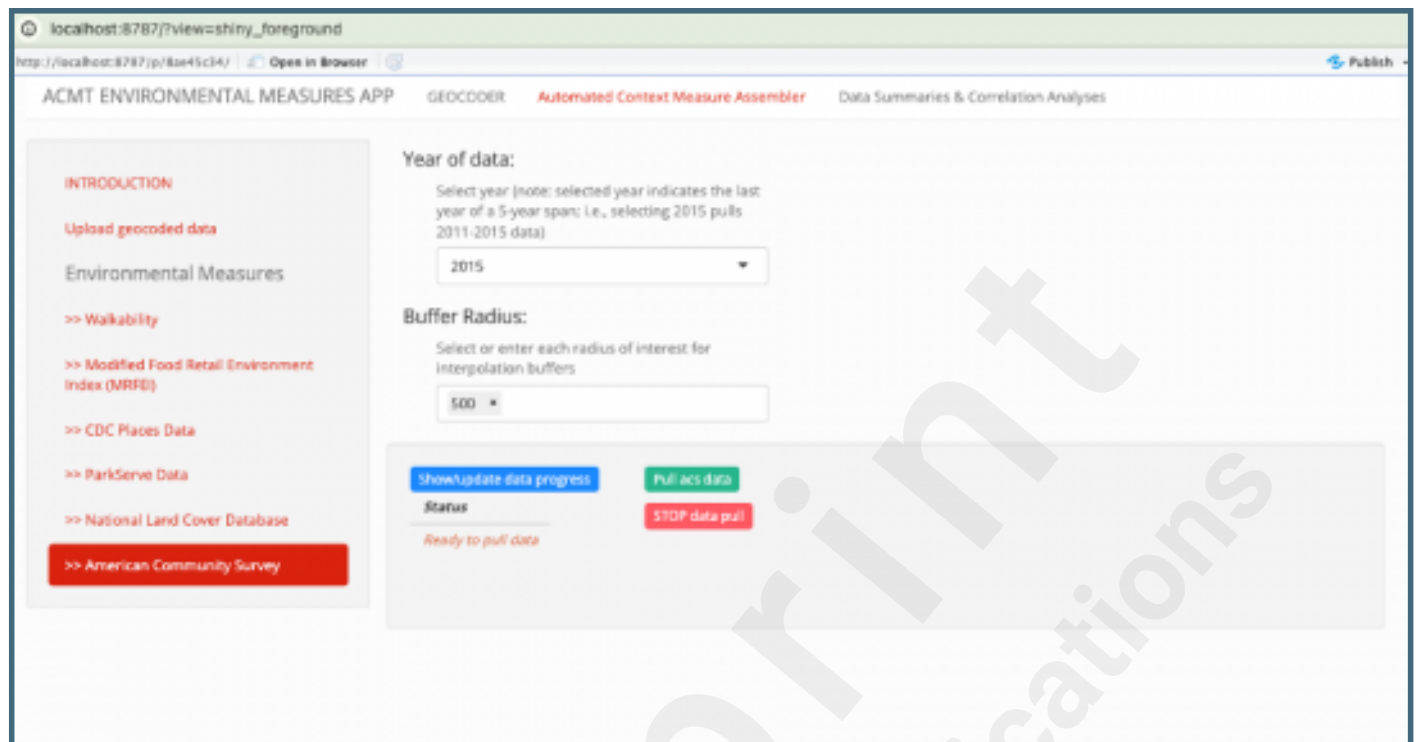
Point-aggregate interpolation. Suppose the black dot again represents the location of interest, the gray dot represents the area corresponding to that point's neighborhood of interest, and the red dots indicate 911 calls. A point-aggregate sum would indicate that four 911 calls fell within the neighborhood area.



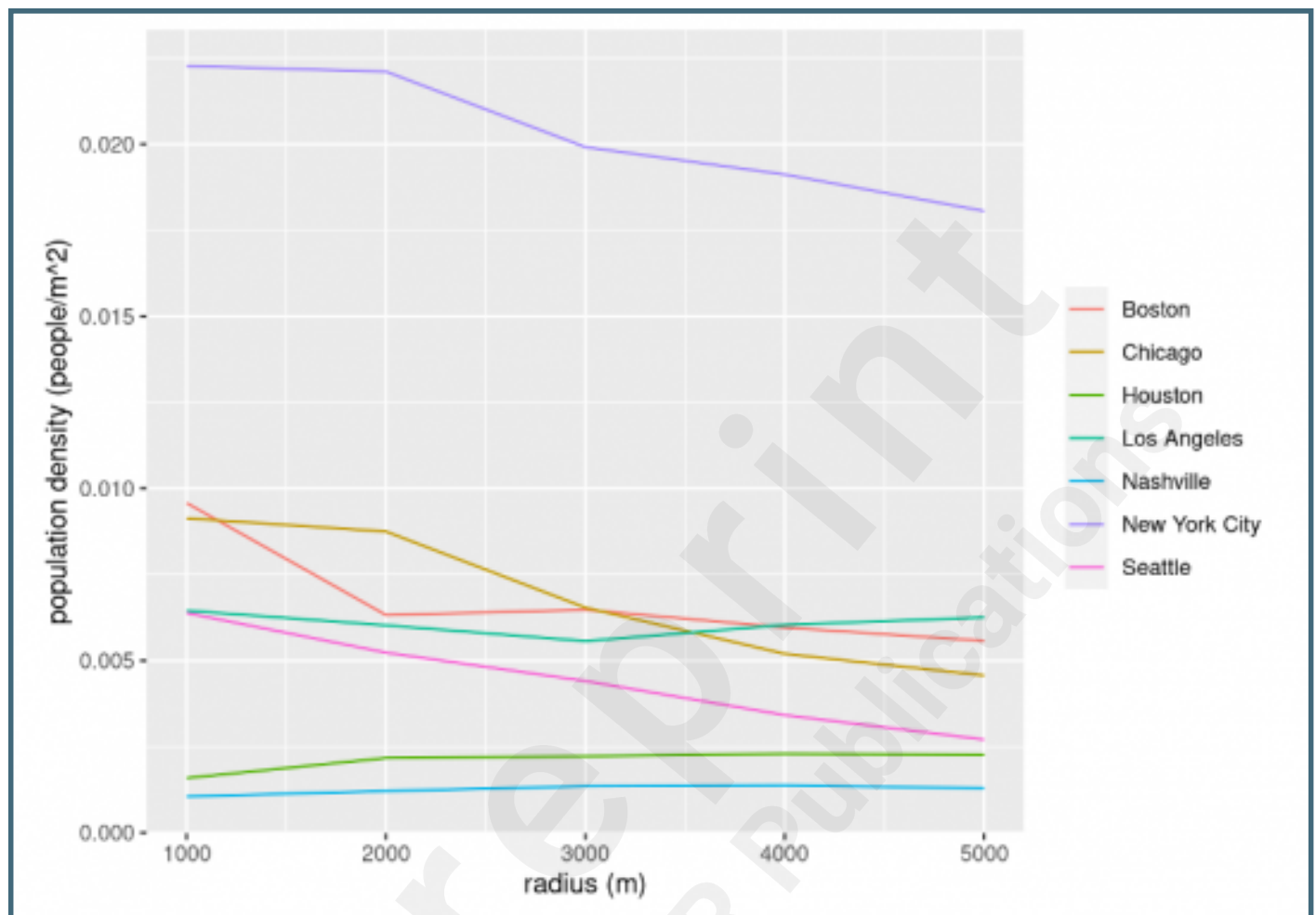
ACMT RStudio user interface. An analyst can use conventional RStudio features, including the editor and debugger, to access the ACMT geocoder and the context measure assembler.



ACMT RShiny user interface for analysis using the American Community Survey dataset. An analyst can select parameters visually for performing analysis.



The population density in relation to distance from the city hall for Seattle, Los Angeles, Chicago, New York City, Nashville, Houston and Boston with the distances being 1000, 2000, 3000, 4000 and 5000 meters.



Manhattan plot of regression coefficients for the association between neighborhood characteristics and cannabis license status (e.g., active and open vs. closed). Each dot represents one environment measure, and the Y axis represents the strength of association. Measures are grouped by the domain of environment measured and by whether the measure was a count (e.g., # of residents within 1000m of an outlet who were females aged 5 to 9) or proportion (e.g., % of residents within 1km of an outlet who were females aged 5 to 9).

