

The Socialoid: A Computational Model of a City



Margeret Hall, Christian Haas, Johanna Schacht, and Steven O. Kimbrough

Abstract A *socialoid* (our term) is an integrated collection of data and models about a society. As such, and accepting that it can never be complete, it is a computational model of a society. We are in the early stages of building a socialoid for Philadelphia, PA. We call it the Philadelphioid. The Philadelphioid is a diachronic (temporal), mashed, geographic information system (GIS) with an extensive integrated library of integrated analytics tools. The purpose of this chapter is to articulate our design rationale for the Philadelphioid and to illustrate its underlying concepts and premises. Central among these concepts is the principle of *solution pluralism*, which enjoins us to use analytics and visualization to create and explore multiple solutions to decision problems. We illustrate an application of this philosophy by discussing analysis pertaining to food deserts carried out with the Philadelphioid.

1 Introduction

Data on and about the lives of cities are abundant, although how best to harness and support data-driven insights for urban environments remains an open question. Put another way, modeling with urban data is not only or is not just an information problem, it is also a design problem. In the context of attempting to solve or address socioeconomic problems through modeling, we focus on designing an information

M. Hall (✉) · C. Haas

College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE, USA

e-mail: mahall@unomaha.edu; christianhaas@unomaha.edu

J. Schacht

Head of AI @ Orbisk, Utrecht, The Netherlands

e-mail: johanna@orbisk.com

S. O. Kimbrough

The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

e-mail: kimbrough@wharton.upenn.edu

© The Author(s) 2021

H. Gimpel et al. (eds.), *Market Engineering*,

https://doi.org/10.1007/978-3-030-66661-3_12

system oriented toward optimization problems, which is to say problems of how best to allocate existing but limited resources.

In this chapter, we focus on a common urban design use case in which resources are available for improving services and data are available for assessing proposed configurations, yet, inevitably, resources are insufficient and data are incomplete, and so the task is to make the best of the situation. Essential to that goal, in our view, are two requirements for supporting public deliberation, which we conceive of as being broadly interactive, affording open and accessible support of multifaceted exploration of possibilities. The first of these requirements is that relevant data be curated, validated, and made available in an open system. The second requirement is that the open system affords discovery and comparison of multiple options by a broad public. Our aim here is to describe progress we have made to date in the specific context of food deserts in Philadelphia, with emphasis on the second requirement.

Our principled response to the first requirement is to build a socialoid for Philadelphia. A *socialoid* (our term) is an integrated collection of data and models that reflect the complex societal fabric constituting everyday social life. Accepting that no socialoid can ever be complete, we may think of them as computational models of their target societies. Such models can assist with large-scale planning, can computationally address problems that are otherwise off-limits, and can deliver data-driven insights to stakeholders, among other possibilities (Bankes 1993). This chapter discusses a work-in-progress socialoid for Philadelphia, PA. We call it the Philadelphioid, conceived as a *diachronic*, *mashed* GIS with an extensive integrated library of integrated *analytics* tools (beyond those commonly found in GIS). As such it is an example of what we can call a *diachronic*, *mashed*, with *analytics* GIS (DMA-GIS).¹ Our purpose in this chapter is to introduce the project and its goals, as well as to present and discuss certain innovative (and we hope widely useful) contributions arising from the project. In addition to contributing to HCI and IS research on urban data, we have a specific interest in presenting the Philadelphioid as an example of designing information systems in response to significant social problems.

Our principled response to the second requirement (to support discovery and comparison of multiple design options) is to adopt the stance of what we call *generalized optimization*. Conventionally, when an optimization problem is posed, a single, best optimal solution (aka decision) is sought, usually by either the exact or heuristic method. As such, conventional optimization ignores our second requirement, which is to support discovery and comparison of multiple good solutions, recognizing that data are always incomplete and models inexact. In *generalized optimization*, we begin with an optimization problem, specify a set of *decisions of interest* (DoIs), and then use computational means to discover elements of the set of DoIs.

¹The Philadelphioid in its current state—including shape files, associated data, heuristic optimization code in Python, data generated by this code, and MAD modeling in Python with the generated data—is freely available from the authors.

To illustrate, our focus problem in this chapter is the optimal placement of a number of grocery stores for the sake of relieving the food desert problem in Philadelphia. We define our DoIs as placements of five grocery stores so as to do well at minimizing the number of people in a food desert. We develop a heuristic search technique that generates a plurality—hundreds, even thousands—of good-quality placement designs. We then develop a multiattribute decision (MAD) model that compares the designs in the discovered DoIs *using an expanded data set that brings to bear information about the designs that were too complex to consider in the heuristic search for DoIs*. Although our data are limited (as is always the case) and we are focused only on Philadelphia, the methods we use are general and intended to be transferable to other urban environments and problems.

Philadelphia flourishes in informative online GIS applications, including Policy Map,² Open Data Philly,³ Plan Philly,⁴ Culture Blocks,⁵ Community Commons,⁶ Community Health Explorer,⁷ Community Health Database,⁸ Next City,⁹ and FixList.¹⁰ Many of these examples include temporally conditioned data, thereby supporting diachronic analysis. In addition, these initiatives typically present results using analytics tools, although these are largely limited to data visualization. Our project goes beyond the functionality of these kinds of systems, which might be characterized as rich, state-of-the-art online GIS applications, to develop next-generation applications that meet the following goals:

1. In general, we want to support research and to “house scholarship.” That is, we want the Philadelphoid to serve as (1) a repository of data, procedures, reports of research results, and other intellectual work products, (2) a primary tool for conducting research, and (3) a primary tool for supporting social deliberation. We see its development as open-ended, and we proceed inspired by, and very much in the spirit of, the Wisconsin Idea.¹¹
2. Support for diachronic (time-based, longitudinal) data and analysis. In terms of standard GIS, this implies that the system delivers multiple maps with layers conditioned by a time frame. The key to this delivery is that the results are easy to use and interpret for human readers.
3. Support for mashing, that is, for pulling together and integrating data from multiple disparate sources. Not only must the data be obtained and combined,

²<http://hdl.library.upenn.edu/1017/76970>.

³<https://www.opendataphilly.org/>.

⁴<http://planphilly.com/>.

⁵<http://www.cultureblocks.com/>.

⁶<http://www.communitycommons.org/maps-data/>.

⁷<http://cityofphiladelphia.github.io/community-health-explorer/>.

⁸<http://hdl.library.upenn.edu/1017/88388>.

⁹<https://nextcity.org/>.

¹⁰<http://thephiladelphiacitizen.org/stacey-mosley-fixlist/>.

¹¹See, e.g., <http://www.wisc.edu/wisconsin-idea/> and https://en.wikipedia.org/wiki/Wisconsin_Idea.

but it must be managed, archived, and subject to provenance control and documentation. This includes support for linked data (e.g., linked open data,¹² and Hafford 2014).

4. Assume (and support) GIS capabilities as well as other forms of visualization such that the delivered analysis eases comprehension of results for its readers.
5. Support for a broad range of social modeling and analytics. Data- and model-driven decision support for urban stakeholders is a main focus of the project.

This chapter describes our efforts to meet these objectives through the Philadelphia, drawing inspiration from existing scholarship on urban data and decision modeling.

2 Background: Smart Cities and Urban Data Modeling

Our work is connected to an interdisciplinary body of scholarship that considers the intersection of urban environments and data, sometimes gathered under the label smart cities, or cities where digital technologies are embedded into infrastructure (Smyth and Helgason 2010). These initiatives emphasize using novel technologies to gather data about everyday life in ways that can lead to increased efficiency and automation of urban systems management (Klauers et al. 2014, Zheng et al. 2014). As a paradigm of urban design, smart cities have been critiqued as problematic for ethical surveillance and the corporatization of urban life (Kitchin 2013). In particular, the flows of data comprising smart cities tend to be one-directional, with data being gathered from but not accessible to individual city residents (Klauers et al. 2014, Odom 2010), even as the stories told with this data tend to benefit commercial interests (Söderström et al. 2014), or to support narrow political views of urban participation (Halpern et al. 2013, Vanolo 2014). A core thread that emerges across these critiques is an insistence on acknowledging that while smart city initiatives are intended to benefit urban residents, the distribution of those benefits is often uneven, with disparities that reflect existing inequalities of race, class, and privilege.

Researchers have noted the opportunities of smart cities as media-rich design environments for provoking new relationships to technology (Di Mascio et al. 2016, Messeter and Johansson 2008). Previously, a variety of outcomes have been produced from urban data, including location-based social media data to detect events (Schwartz et al. 2013, Xia et al. 2014) and activity patterns (Cranshaw et al. 2012, Gallacher et al. 2015), using crowdsourced data to manage hyper-local city services (De Melo Borges et al. 2016) and to encourage learning about neighborhood information and resources (Claes and Moere 2013). In the context of

¹²See, e.g., <http://labs.europeana.eu/api/linked-open-data-introduction>, <http://doi.org/doi:10.4018/jswis.2009081901>, <https://www.w3.org/DesignIssues/LinkedData.html>, and <http://www.linkeddatatools.com/semantic-web-basics>.

research that pertains to food deserts, Choudhury et al. (2016) used a mass analysis of Instagram posts to develop models for detecting food deserts, with a fairly high accuracy. Similarly interested in detection, Yu and Nahapetian (2013) developed a food consumption app to gain a grounded account of the kinds of food available for purchase within a given neighborhood or city, with the goal of producing maps that can be used for advocacy purposes.

These projects tend to rely on social media data, smartphones, and mobile apps and more specifically on the locative functionality of mobile phones, meaning the ability to tag social media content with geographic coordinates. In contrast to this emphasis on events and activity, our work on modeling is intended to provide an example of how to develop sophisticated models for understanding city space using accessible data analysis tools. These modeling projects have implications for both scientific and policy applications.

3 Modeling for Scientific Applications and Policy Purposes

Whether as individuals or as institutional actors, people can rarely make use of all of the disparate data sources available to them for making decisions. Reasons are many and are well covered in such works as Georges et al. (2016), Laurel et al. (1990), Sweller (1994), Todd and Benbasat (1994). Here, we address the use of computational models for informing stakeholders of consideration sets for decisions that are driven by actual data. Models, broadly construed, are both tools for and primary work products of applied and theoretical research. Although boundaries are inevitably fuzzy, it is useful for present purposes to think in terms of four kinds of models.

1. *Presentation and visualization.* Standard GIS, such as those referred to above, are largely visual presentations of data. They, along with other forms of data visualization, are prototypical examples of this kind of modeling.
2. *Description.* Regression and classification models are prototypical approaches for describing relationships in data. They are able to condense and summarize large amounts of data and afford our seeing of significant patterns in the data. This is true of supervised learning and unsupervised learning algorithms.
3. *Prediction or forecasting.* These models take an explicitly diachronic perspective, allowing us to predict what has not yet been observed (or not used in constructing the model). Valuable as they are, these kinds of models are challenging in our context because so much of available data is synchronic (cross-sectional), not diachronic (longitudinal). Even so, forecasting models constitute an important class of models for socialoids and DMA-GIS in general.
4. *Prescription.* Prescriptive models aim to support decision-making. They introduce the concept of an objective function to be optimized. We solve them in order to find decisions (actions to be taken) that do well with regard to a given objective. Prescriptive models are very common in engineering, policy-making,

and business (operations research (OR), management science, etc.). Constrained optimization models are a prime example, whether expressed as mathematical programs or not.

Multiattribute decision (MAD) models are another important class of prescriptive modeling. We discuss them in the next section.

Prescriptive modeling with spatial data and problems is comparatively underdeveloped. The “urban OR” field, for example, had impressive successes in the 1970s but has since withered from lack of support by financially challenged urban interests. This said, there is certainly ample high-quality work to draw upon and with contemporary informatics resources (GIS, cloud computing, advances in metaheuristics, the open-source culture, etc.). We believe that prescriptive modeling for socialoids, and urban OR problems generally speaking, is primed for great advances. In cases where stakeholders have not yet defined the totality of the problem (i.e., wicked urban problems), models with inclusive data can estimate many scenarios in a less emotionally charged environment (Davies and Nutley 2000). For these reasons, prescriptive modeling is an important, but hardly exclusive, focus of our project. With that in mind, we now discuss some of our efforts and results in this regard.

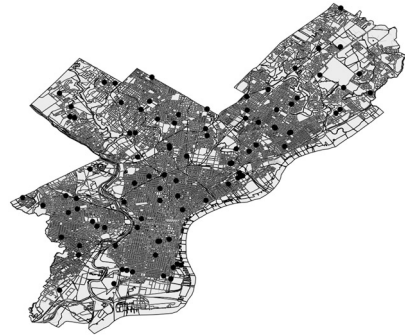
4 Modeling Example

4.1 *Generalized Optimization*

So far, we have defined the Philadelphioid as an integrated information system of complex and multifaceted urban data, and we argued that modeling with this data can point to solutions for addressing significant socioeconomic problems. We illustrate these latter claims by focusing on the problem of food deserts (Cannuscio et al. 2014, Mayer et al. 2014), at present a serious and persistent issue in Philadelphia. Food deserts present a serious urban problem, with dramatic health consequences that can include increased rates of obesity and poor nutrition (Ploeg et al. 2009). These consequences may be especially acute for vulnerable and immobile populations—the elderly and the young. Given the pressing nature of the issue and the broad availability of grocery store locations, food deserts were chosen as a modeling example to outline usability of the Philadelphioid and thus socialoids.

While definitions of food deserts can vary, in general, the term refers to neighborhoods where residents experience severe shortages in local access to grocery stores (Bernstein and Shierholz 2014). The 0.5-mile criterion is the lowest of three suggested distances by Ploeg et al. (2009) and the Philadelphia health department. They introduce three distance markers: for rural areas, a distance of up to 10 miles is feasible for not being considered living in a food desert; for urban areas (such as Philadelphia), a 0.5–1.0 mile walking distance is considered feasible. Of course, the cutoff of 0.5 mile is simply a parameter in our modeling and can be changed at will, which is a major advantage of computational modeling. By this criterion,

Fig. 1 The current situation in Philadelphia. Black dots represent supermarkets; gray lines represent streets



about two-thirds (968,081 of 1,526,006 people) of Philadelphia’s population live in a food desert (Fig. 1). A question then is “given additional resources for placing n new grocery outlets, where should they be placed?” This is the design question we address with the Philadelphioid as an information system.

Complicating any approach to answering this question is the fact that models are inaccurate and inevitably fail to include important data for the problem. This may happen because the data are not available, because including the data would make the model prohibitively expensive computationally, and for many other reasons (Bankes 1993). Attempting to determine solutions for placing grocery stores immediately raises a host of contingent questions: Who should receive priority for being served? The old? The young? Ethnic minorities and if so, which ones? What about communities of common interest? How, if at all, can distance be mitigated by public transportation? By proximity to places of employment? It is a challenging task even listing the relevant criteria and/or to have agreement between stakeholders, let alone finding relevant and usable data to incorporate into a tractable model.

For these reasons, we adopted the philosophy of *solution pluralism* and *generalized optimization* (Chou et al. 2014, Hall et al. 2013, Kimbrough and Lau 2016) in which we seek multiple good solutions to the question at hand, rather than a single “best” (optimal) solution.¹³ The plurality of solutions—the discovered DoIs—are then to be used for collective deliberation, which we demonstrate next.

Before describing our model in depth, we offer a final caveat: while we focus in this chapter on supermarket placement, there are a number of other means of mitigating the social and health consequences of food deserts, including education, transportation, affordable housing, and access to medical care and expertise. These

¹³What counts as a good solution will of course be problem specific. In Kimbrough and Lau (2016), four types of good solutions are identified for constrained optimization problems generally: (A) feasible, with near-optimal objective values; (B) feasible, with objective values better than a given threshold and having large amounts of slack on the constraints; (C) infeasible, but having small amounts of constraint violations; and (D) infeasible, with amounts of constraint violation below a given threshold and having objective values superior to the optimal. In the case of food deserts, good solutions are feasible and serve total populations nearly as large as that served by the heuristically best (heuristically optimal) solution discovered.

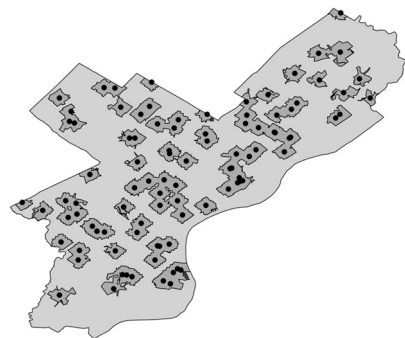
approaches are almost certainly more effective in combination than isolation; by concentrating on geographic placement of grocery stores, we do not mean to imply that this is the only solution, or the only use for the Philadelphioid, in better understanding food deserts. We hope to address one facet of food deserts as a social ill, but our discussion should not be viewed as a single antidote for combating food-related social injustices.

4.2 Heuristic Optimization

To begin, we frame the problem as a *location-allocation* problem, a constrained optimization problem with the objective of maximizing the number of people served, by the placement of the n outlets (full-line supermarkets). In this example $n = 5$. There is no larger reason $n = 5$ is the selected number, as any number would outline the usability of the Philadelphioid to solve problems just as well. However, even this simple formulation presents a challenging combinatorial optimization problem. We shall now briefly describe our computational approach to solving for $n = 5$ additional supermarkets, which yielded 1024 high-quality solutions. In the next sections, we describe how the Philadelphioid helps us make use of these solutions.

The data is based on the 2010 US census data plus geo-coded locations for the 97 full-service grocery stores in Philadelphia. For each of the 18,874 census blocks, we compute the taxicab distance to the closest supermarket from its centroid. Population and income data are available, as are further demographic data such as population counts for Caucasian American, African American, and other ethnic groups, as well as counts of population above 65 years of age and population under 15 years of age. The current disposition of the existing grocery stores is far from optimal with regard to maximally serving population with supermarkets. Figure 1 shows the current situation of Philadelphia. Figure 2 shows areas that are in/not in a food desert (using taxicab distance, a census block more than 0.5 miles away from a supermarket is counted as located in a food desert).

Fig. 2 Black dots represent supermarkets; dark gray areas surrounding them represent areas that are reasonably covered by the supermarket. The remaining light gray area represents the Philadelphia food desert



Our heuristic algorithm for placing the n ($= 5$ here) new grocery outlets works as follows. It is a variant of a standard greedy algorithm for location allocation (Kimbrough and Lau 2016). Given census blocks (polygons in GIS) and supermarkets (points in GIS), the goal of the described point-to-polygon location allocation problem is placing five new supermarkets so that a maximum number of people are added to the count of people that do not live in a food desert. This is achieved using a greedy heuristic that places one supermarket after another, meaning placing the first supermarket, updating the situation in the data and then placing the best supermarket given the new updated situation, and so on. At each step, the choice is greedy optimal because we simply enumerate the choices. Overall, however, when placing more than one outlet, this is not optimal (Kimbrough and Lau 2016). The heuristic does reduce the computation from 18874^5 options to about 18874×5 options. (Our algorithm, see below, worked with all 18,874 census blocks in Philadelphia.)

In our greedy algorithm, selection of locations for new supermarkets is purely based on population, i.e., on minimizing population in a food desert. We introduce solution pluralism into our algorithm as a way of circumventing this limitation. When the heuristic places a supermarket, it saves the top four locations for that supermarket. Upon completion of a run, the heuristic algorithm will return 20 different locations, 4 options for each of the 5 new locations. The combinations provide $4^5 = 1024$ options for placing 5 new supermarkets. Of course, 1024 cannot be said to be the best number of options generated, but it is adjustable via the algorithm should experience indicate that other values should be used. We find that 1024 is large enough that it contains a variety of distinct and interesting solutions. As described next, the 1024 options for the position of best solution are further evaluated using utility functions.

4.3 *Designing MAD Models*

Based on the previous description on how we produced 1024 good solutions for the Philadelphia food desert problem, we now need to discuss how this plurality of solutions can be assessed and deliberated upon in a principled and accessible way. We discuss in this section a very general approach meeting these requirements: multiattribute decision (MAD) models, built with the SMARTER technique. In the following section, we describe the particulars of its application to our food desert data.

It is often the case that we need to assess multiple outcomes or entities on several dimensions or attributes. To generalize the scenario, consider comparing restaurants. We consider price, quality, distance from home, service, as well as other attributes. Rarely, if ever, do we find a single outcome (e.g., restaurant) that is as good as or better than the alternatives on every single attribute of interest. Consider how we might represent the familiar problem of choosing a restaurant on a particular occasion. It is often the case that we need to assess multiple outcomes

Table 1 Restaurants with attributes

Restaurant name	Food	Decor	Service	Cost
Abbey Grill	14	16	16	\$26.00
Academy Café	15	16	15	\$23.00
Adobe Café	16	12	15	\$18.00
Alaina’s Fine Food	21	15	20	\$19.00
Alberto’s Newtown Squire	19	20	19	\$31.00
Al Dar Bistro	16	14	14	\$20.00
Alexandar’s Table at Chaucer’s	17	14	16	\$19.00
Alexander’s Café	17	17	16	\$26.00

or entities on several dimensions or attributes. Table 1 presents a generic example of multiattribute data (having nothing to do with food deserts). It shows comparative scores for a number of Philadelphia restaurants on each of the four dimensions, viz., the attributes food quality, decor quality, service quality, and cost.

It often happens that we need to make trade-offs among the attributes in order to arrive at an accurate, as opposed to emotionally driven, overall score (the utility or more generally an *index*) for a possible choice. Multiattribute decision (MAD) modeling has as its purpose the construction of mathematical models for making these trade-offs (Kimbrough and Lau 2016, Yoon and Hwang 1995).

Returning to Philadelphia’s recognized problem of food deserts (Cannuscio et al. 2014, Johann et al. 2014, Mayer et al. 2014), we modeled census data at the block level, other data at the census block group level, and the locations of the 97 supermarkets in Philadelphia. Recall that we then asked the question: Supposing resources were available to add five new supermarkets, where should they be added? We implemented a heuristic optimization procedure for location analysis (see **Heuristic Optimization**) to find candidate decisions, based upon maximizing the population served within 0.5 mile of a supermarket. Recall again that employing a philosophy of solution pluralism (finding and using multiple solutions or decisions for a problem; see below), we used the heuristic optimization procedure to find 1024 good decisions for locating these 5 stores. Table 2 displays scores on 6 attributes for the best 12 decisions discovered. Rows correspond to candidate decisions and columns to their served populations on the indicated attributes.

For many purposes, it will be useful to construct a simple additive MAD model, in which the utility or index of each alternative is a weighted sum of the utilities of the alternative’s attribute scores. This is readily expressed clearly in mathematical notation. Our additive model is

$$U(x_i) = \sum_{j=1}^n w_j u_j(x_{i,j}) \tag{1}$$

where $i \in$ objects or choices, $j \in$ the n attributes, x_i is choice i , w_j is the weight on attribute j , u_j is the utility function on attribute j , and $x_{i,j}$ is the score

Table 2 Attributes of the top 12 discovered decisions for adding five grocery stores in Philadelphia. A: total population served. B: Caucasian American population served. C: African American population served. D: Other ethnic population served. E: Population over 65 served. F: Population under 15 served

Rank	A	B	C	D	E	F
1	649,965	263,256	277,075	109,634	97,419	143,685
2	649,743	263,295	276,784	109,664	97,438	143,680
3	647,220	263,383	274,298	109,539	97,129	142,958
4	649,536	261,676	278,377	109,483	97,333	143,959
5	649,491	262,288	277,321	109,882	97,196	143,571
6	649,269	262,327	277,030	109,912	97,215	143,566
7	646,973	262,673	274,532	109,768	96,954	142,901
8	649,062	260,708	278,623	109,731	97,110	143,845
9	649,054	262,536	277,048	109,470	97,234	143,537
10	648,832	262,575	276,757	109,500	97,253	143,532
11	646,973	263,177	274,291	109,505	97,074	142,922
12	648,625	260,956	278,350	109,319	97,148	143,811

of object/choice i on attribute j . (The scores in Tables 1 and 2 are $x_{i,j}$ values.) We follow the method of Edwards and Barron (1994), called SMARTER. It is maximally simple and has good theoretical backing and empirical success (Dawes 1979, Kimbrough and Lau 2016). Of course, it is possible to develop more complex, and one might hope more accurate, models. There is a good theory for this (Edwards and Barron 1994, Keeney and Raiffa 1993, von Winterfeldt and Edwards 1986), but the burden on the users is much increased. SMARTER models are excellent points of entry to MAD modeling. We shall now unpack the model and discuss how all of its elements may be obtained with minimal user input (if that is what the user wants).

Let X be a table or array of outcomes and their scores. In terms of Table 1, X is the interior twelve rows and four columns. $x_{i,j}$ is the element of X in the i th row and j th column. For example, $x_{3,2} = 12$ and is the score for the Adobe Café on decor. We need to convert all of the scores to a common range so that they may be compared. We choose the $[0,100]$ range and transform each of the scores in X as follows:

$$u_j(x_{i,j}) = 100 \times \frac{x_{i,j} - x_j^-}{x_j^+ - x_j^-} \quad (2)$$

where:

- $x_{i,j}$ is the score of object/choice i on attribute j .
- x_j^+ (x_j^-) is the best (worst) score on attribute j .

For example, from Table 1, if the attribute is food ($j = \text{food}$), then $x_j^+ = 24$ and $x_j^- = 14$.

- $u_j(x_{i,j})$ is the utility (desirability, index value) on attribute j of the score on attribute j of object/choice i .

For example, from Table 1, if the object is Al Dar Bistro (i = Al Dar Bistro) and the attribute is cost (j = cost), then the object is $x_{i,j}$ and its utility u_j on i is score

$$u_j(x_{i,j}) = 100 \times \frac{\$20 - \$31}{\$18 - \$31} = 84.6 \quad (3)$$

We also need to find weights, w_j , for each attribute. To do so we use the method of rank weights. We simply ask the user to rank in order the attributes by importance or value, and then we calculate weights based on the rankings only. (This has to be done properly. We pass over the details because they are known and readily available.) Here is the formula for doing this:

$$w_k = \frac{1}{n} \sum_{j=k}^n \frac{1}{j} \quad (4)$$

where w_k is the weight on the k th attribute by rank (so w_1 is the weight on the first-ranked attribute), n is the number of attributes, and the highest ranking attribute has a k value of 1 (i.e., the best to worst rank order is 1, 2, 3, ..., n).

With four attributes as in our restaurant example, if the attributes food, decor, service, and cost are ranked in that order, then the (rank, weight) pairs are (1, 0.5208), (2, 0.2708), (3, 0.1458), and (4, 0.0625).

4.4 Using the MAD Models

MAD models of the SMARTER variety potentially apply to any situation in which there is a table of scores, X , in which rows correspond to distinct entities, columns to attributes of the entities, and the $x_{i,j}$ are scores for the entity attributes. In our context, this range of applicability is indeed very large. Synchronically, the entities to be compared co-exist at a given time. These might be different plans for adding grocery stores, different redistricting proposals, geographic entities such as wards, and much else. Attributes of interest might include total population, income distribution, demographics (income, ethnicity, age distribution), voting behavior, access to public transportation, presence of services such as police and fire stations, etc.

Under either perspective, we advocate the philosophy of solution pluralism: use the data and the models to generate multiple possible decisions, and then evaluate this plurality of options, taking into account information not present in the data and models (see Chou et al. 2014, Hall et al. 2013, Kimbrough and Lau 2016). MAD models are entirely apt tools for this purpose.

MAD models, then, plausibly have a wide scope of applicability for socialoids and DMA-GIS. The basic work flow in getting them built is remarkably simple:

1. The user: Based on a system presentation, identifies the collection of entities to be compared.
2. The user: Based on a system presentation, identifies the entity attributes to be included in the model and for each attribute identifies its preference sense (Is more better or is less better?).
3. The user: Based on a system presentation (and using theory about how to do this), rank orders the attributes.
4. The system: Assembles the model and presents a ranking of the selected entities.
5. The user: Explores the results and comes to a judgment.

At this point, it is possible to provide a variety of general services for MAD models that support post-solution analysis (Kimbrough and Lau 2016). The user in this scenario is a human, preferably a user with considerable insight on the models to be chosen and a good understanding of the domain. To note some main examples and the questions they raise:

1. Sensitivity analysis:
Are there small changes in parameter values that will result in large changes in outcomes according to the model? If so, what are they, and which parameters have the largest sensitivities, i.e., which parameters if changed by small amounts result in large changes in model outcomes? Which parameters have small sensitivities?
2. Missing data and implausible data services:
Where are the missing data? How should any missing data be handled? What about apparent outliers?
3. Outcome reach analysis:
Given an outcome predicted or prescribed by the model, and a specific desired (or feared) alternative outcome, how do the assumptions of the model need to be changed in order to reach the outcome(s) in question? If so, what is the cheapest or most effective way to do this? Outcome reach questions arise on the degradation side as well. By opting to accept a degraded outcome, we may free up resources that can be used elsewhere.
4. Robustness analysis:
Which decisions or policy options of the model perform comparatively well across the full range of ambient uncertainty for the model?

It is evident that a SMARTER MAD model affords much scope for automation of post-solution analysis (including automation to check with the user).

Focusing again on our data and results, Table 2—fundamentally similar in structure to Table 1—presents multiattribute data for the food desert problem. Results show that when adding five supermarkets to Philadelphia minimizing a 0.5-mile taxicab distance as the indicator on living in a food desert, the number of people not living in a food desert can be raised from 557,925 to 640,065. The 1024 discovered solutions raise the number of people not living in a food desert to an

average of 639,675, which is an insignificant difference to the solutions discovered by the heuristic.

The perspective of solution pluralism has much to add here. While the 1024 discovered solutions do not differ greatly in the number of people in a food desert, they do differ significantly in other aspects. Considering only the best 12 options, as shown in Table 2, all other columns differ by more than the best population value in column A and the average unserved population (639,675). Thus, the quality of the solutions with regard to the other criteria is more diverse. Solution diversity allows us to differentiate between the solutions considering all criteria—the goal of solution pluralism combined with MAD models. Following up on this observation, we use these differences to evaluate all 1024 solutions, by applying a MAD model to score and rank all of the solutions on an expanded set of criteria.

Table 3 presents the weighted single attributes (columns A–F) and overall utility scores (column G) for a SMARTER MAD model we developed and applied to the 1024 solutions from the greedy heuristic, using additional data, as indicated in the table. As described above, given the data, we need just two additional sources of information in order to build the model, whose results we see, in part, in the table. The first information item we need is the sense for each of the six attributes (A–F). In our model, we stipulated that for attributes A–E, more is uniformly better, but for attribute F (average income), less is better. Thus, we sought to favor lower-income individuals. The second information item we need is the rank order of the attributes

Table 3 Weighted utility values for the top 20 solutions. ID: Solution rank on population served. A: Population served. B: African American population served. C: Other ethnic population served. D: Population over 65 served. E: Population under 15 served. F: Income per year (in \$1000). G: Net utility

ID	A	B	C	D	E	F	G
131	2.70	6.04	5.95	22.62	35.81	10.24	83.36
67	2.72	5.93	5.96	23.26	35.80	9.51	83.19
643	2.69	2.81	9.96	12.80	39.40	15.31	82.96
579	2.71	2.70	9.97	13.44	39.39	14.59	82.80
899	2.69	5.88	6.02	22.11	35.08	10.20	81.97
128	2.76	5.67	6.06	23.36	34.67	9.42	81.93
195	2.69	5.88	5.94	23.13	36.08	8.17	81.90
691	2.69	2.80	9.98	12.57	38.75	15.03	81.81
129	2.73	5.58	6.08	23.52	34.65	9.25	81.81
835	2.71	5.77	6.03	22.75	35.07	9.47	81.80
3	2.73	5.84	5.96	22.85	35.84	8.58	81.80
64	2.78	5.56	6.07	24.00	34.66	8.69	81.76
135	2.65	6.11	6.12	20.70	35.34	10.75	81.66
627	2.71	2.69	9.99	13.21	38.74	14.30	81.64
65	2.75	5.47	6.09	24.17	34.64	8.51	81.63
640	2.74	2.43	10.06	13.54	38.26	14.50	81.54
707	2.68	2.65	9.94	13.31	39.67	13.25	81.51
179	2.67	6.02	5.88	22.29	34.84	9.81	81.50
71	2.67	6.00	6.13	21.34	35.33	10.02	81.49
641	2.71	2.35	10.08	13.70	38.24	14.32	81.41

in importance. This we stipulated as $A = 6$, $B = 5$, $C = 4$, $D = 2$, $E = 1$, and $F = 3$. Thus, $E = \text{population under 15 served}$ is the most important attribute in the model, and $A = \text{total population served}$ is the least important. Of course, A is what the heuristic optimization procedure sought to maximize. With this ranking, the weights are $A = 0.02777778$, $B = 0.06111111$, $C = 0.10277778$, $D = 0.24166667$, $E = 0.40833333$, and $F = 0.15833333$. Points arising:

1. Comparing Tables 2 (raw data) and 3 (MAD model processed data), we can see that the MAD model has done real work for us in discriminating among the plurality of options. Looking at the raw data in Table 2, it is really not possible to discern a better row from a worse row, yet the MAD model data, excerpted in Table 3, does this comprehensively for all 1024 options.
2. Only one—ID 3, or rank 4—of the top 60 solutions from the heuristic optimization appears in the top 20 solutions identified by the MAD model.
3. Several of the top 20 utility model solutions have rank scores worse than (greater than) 500 in the heuristic (population served only) optimization solutions.
4. Figure 3 shows a portion of Philadelphia. Existing grocery stores are indicated as black dots. The five green dots represent the locations of the five grocery stores in the best heuristic optimization solution.
5. When we examine the top heuristic optimization solutions, they closely resemble what we see in Fig. 3.
6. Remarkably, of the top 10 solutions from the utility model, eight also closely resemble the solution in Fig. 3. However, two are somewhat different, ranks 4 and 8 (rows 4 and 8 of Table 3).
7. Figure 4 adds, as yellow stars, the five locations from the rank 4 solution from the utility model. We see that four of the five locations are quite close to those of the heuristic optimization, yet the fifth location is quite different from its counterpart in the heuristic optimization. The rank 8 solution in the utility model is similar.



Fig. 3 Green dots represent top solution from the heuristic optimization, placing five grocery outlets; black dots indicate existing grocery stores in Philadelphia

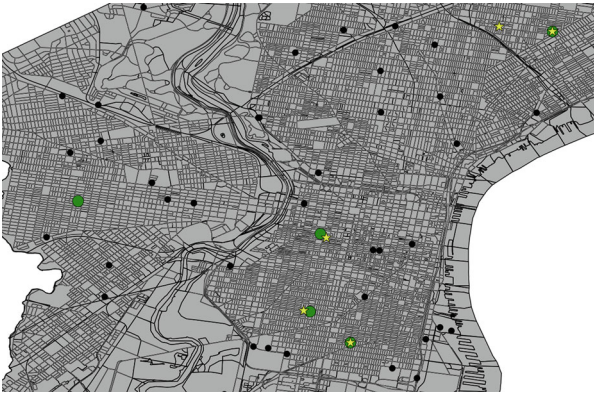


Fig. 4 Green dots represent top solution from the heuristic optimization, placing five grocery outlets; black dots indicate existing grocery stores in Philadelphia. Yellow stars represent the rank 4 solution from the utility model

Table 4 (a) (ID 1 to 5) presents information on green dots of Figs. 3 and 4, and (b) (ID 6 to 10) presents information on yellow stars of Fig. 4

ID	Street	Number	Longitude	Latitude
(a)				
1	S 17th Street	1820	−75.175	39.928
2	Ritner Street	1014	−75.165	39.921
3	Potter Street	3150	−75.116	39.996
4	S 18th Street	368	−75.172	39.947
5	S 54th Street	324	−75.231	39.955
(b)				
6	Sigel Street	1750	−75.176	39.928
7	Ritner Street	1014	−75.165	39.921
8	Potter Street	3150	−75.116	39.996
9	Waverly Street	1750	−75.171	39.946
10	N Water Street	3150	−75.129	39.997

Table 4 shows further information pertaining to Fig. 4. As Fig. 4 suggests, the green dots and yellow stars twice coincide at the same supermarket location; these are supermarkets with IDs 2 and 7 as well as supermarkets with IDs 3 and 8. Supermarkets with IDs 5 and 10 are those that are very far away from each other. The last four supermarkets (IDs 1 and 6 (the supermarket couple further south in Fig. 4) and 4 and 9 (the supermarket pair further north in Fig. 4)) are close and so would serve a similar part of Philadelphia. To give an idea about distances: 1 and 6 are 130 meters apart and 4 and 9 are 120 meters apart.

5 Discussion and Future Work

Urban environments are complex and messy arrangements of people, institutions, and infrastructure. Making decisions about how to address a city or neighborhood problem—such as access to healthy food—is not solely a matter of access to data; it is

also about being able to model data in such a way as to be usable. More specifically, information systems must be designed in a way that allows for interpretation and can inform difficult decisions through the presentation of distinct possibilities for addressing identified problems. We have described the Philadelphoid as an information system designed to aid in research and deliberation surrounding local socioeconomic concerns. Using real data to model solutions for a known problem, our modeling design leverages solution pluralism to provide a robust set of possibilities for tackling resource allocation in a discrete geographic setting.

In evaluating the efficacy of the Philadelphoid, it is safe to draw two conclusions from the heuristic optimization and MAD exercise discussed above. First, the solution from the optimization, green dots in Fig. 3, is quite robust within the consideration set we explore. Each of the 1024 solutions is distinct, and even adding the MAD model, incorporating information from other attributes does little to change this conclusion. The optimization solution appears to be a good one, but so are very many other solutions nearby geographically. Thus, if we accept it as an anchor point, there is ample room for adjusting locations slightly (by a few census blocks) in response to further information (availability of land, public transit, etc.). We see this proposed solution as a guide rather than a fixed, unyielding demand, i.e., the Philadelphoid is meant to inform rather than dictate decisions. The second point is that there are a few credible, distinctly different alternatives to the heuristic optimization solution, viz., as seen in Fig. 4. In terms of public deliberation, we might think of this solution as in the consideration set, but as having a burden of proof to overcome when compared to any solution very similar to the optimization solution.

The larger point here is that we can see that the Philadelphoid has produced substantial material for affording public discussion presented in a way that aids comprehension of the proposed solutions. It does this in large part by producing a consideration set—a plurality of solutions—of high value. It then evaluates a feasible number that are appropriate for deliberation and presents them in easy-to-digest ways.

This last point serves as a segue to discussion of future work. As a tool with both academic and activist capacities, the Philadelphoid's design and potential use cases are open-ended. We have laid out with considerable specificity important basic elements of a socialoid. The challenge now is to articulate and implement generalizations of what is on display here. These include at a minimum more data, especially diachronic data, richer modeling, further algorithms for generating multiple solutions, more nuanced forms of MAD modeling, and much more in the way of interactive, visualized post-solution analysis. Diachronically, the entities can be conditioned by time frames (e.g., by decade) and the attributes just about anything of interest, particularly where they touch on issues of broad social and cultural import (such as historical patterns of police brutality, indicators of neighborhood change, and so on). For example, the entities might be geographic regions of the city conditioned by a time frame and the attributes quality of life scores for the region-times across multiple dimensions (income, crime, health, etc.). Building chronological accounts of urban environments would provide researchers as well as policymakers with tools for understanding in a more precise and grounded

way how city landscapes and neighborhoods are being affected by dynamics of gentrification and fluctuations in population and wealth. We are, in particular, keen to explore this with diachronic data on grocery store locations and population data. Has the Philadelphia food desert problem been getting better or worse since 1960? Addressing this and many related questions is within reach by adding existing data to the present system.

The ability to support diachronic data would allow us to model *time series*; that is, we could observe change in time which, in turn, would allow different models of the factors influencing social and cultural patterns to be tested. For example, if we have a theory that the distribution of grocery stores in the city is influenced by some set of demographic factors distributed over space and time, we could use the time series in the model to test this theory. Analysis of these time series would allow a researcher to model and test a variety of theories against actual spatiotemporal series with different degrees of granularity. We could, for example, look at the impact of demographic changes on both property values and food deserts. Alternatively (or in addition), we could analyze the impact of the introduction of casinos on property values, crime rates, and bankruptcy rates in the surrounding neighborhoods. From a policy perspective, we could also model the impact of potential governmental actions and ordinances on the community, as well as model how past policies have influenced the well-being of various populations in the city.

Other areas of interest for future work could shift from decision support theory in urban data contexts to interpersonal and institutional settings. For example, one could envision evaluating solution quality from the socialoid compared to focus group discussions or expert opinions. Measuring stakeholders' cognitive load in the deliberation and decision process could also be significant. We note that for a richer set of design decisions, this kind of modeling and data gathering should be informed by research on interpersonal and organizational communication, as well as decision theory. Another area of interest is applying the concept of solution pluralism on political redistricting, i.e., the creation of voting districts based on smaller precincts. As a highly politicized issue with many, and often unclear, objectives, initial work shows that a variety of solutions with different properties can be created and evaluated based on several dimensions (Haas et al. 2020, Miller et al. 2018). Finally, we hope that very many socialoids can be built. This will afford comparative analysis between cities and neighborhoods, which can yield insights and is required to undergird provisional findings with any single socialoid. Indeed, as scientists, we are "stronger together" by creating a plurality of solutions with a plurality of socialoids.

References

- Bankes S (1993) Exploratory Modeling for Policy Analysis. *Operations Research* 41(3):435–449
- Bernstein J, Shierholz H (2014) The minimum wage: A crucial labor standard that is well targeted to low- and moderate-income households. *Journal of Policy Analysis and Management* 33(4):1036–1043

- Cannuscio CC, Hillier A, Karpyn A, Glanz K (2014) The social dynamics of healthy food shopping and store choice in an urban environment. *Social Science and Medicine* 122:13–20
- Chou C, Kimbrough SO, Murphy FH, Sullivan-Fedock J, Woodard CJ (2014) On empirical validation of compactness measures for electoral redistricting and its significance for application of models in the social sciences. *Social Science Computer Review* 32(4):534–542
- Choudhury MD, Sharma S, Kiciman E (2016) Characterizing dietary choices, nutrition, and language in food deserts via social media. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*, ACM, New York, NY, USA, pp 1157–1170
- Claes S, Moere AV (2013) Street infographics: raising awareness of local issues through a situated urban visualization. In: *Proceedings of the 2nd ACM International Symposium on Pervasive Displays (PerDis '13)*, ACM, pp 133–138
- Cranshaw J, Raz S, Hong J, Sadeh N (2012) The livelihoods project: Utilizing social media to understand the dynamics of a city. In: *International AAAI Conference on Weblogs and Social Media*, pp 58–65
- Davies HT, Nutley SM (2000) *What works?: Evidence-based policy and practice in public services*. Policy Press
- Dawes RM (1979) The robust beauty of improper linear models in decision making. *American Psychologist* 34(7):571–582
- De Melo Borges J, Budde M, Peters O, Riedel T, Schankin A, Beigl M (2016) Estavis: A real-world interactive platform for crowdsourced visual urban analytics. In: *Proceedings of the Second International Conference on IoT in Urban Space (Urb-IoT '16)*, ACM, pp 65–70
- Di Mascio D, Clarke R, Akama Y, Salim F (2016) Urban HCI: (re)adapting the city together. In: *Proceedings of the 2016 ACM Conference Companion Publication on Designing Interactive Systems*, pp 89–92
- Edwards W, Barron FH (1994) SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes* 60(3):306–325
- Gallacher S, Golsteijn C, Kalnikaite V, Houben S, Johnson R, Harrison D, Marquardt N (2015) SenCity 2: visualizing the hidden pulse of a city. In: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*, ACM, pp 1391–1394
- Georges V, Courtemanche F, Senecal S, Baccino T, Fredette M, Leger PM (2016) Ux heatmaps: Mapping user experience on visual interfaces. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* pp 4850–4860
- Haas C, Hachadoorian L, Kimbrough SO, Miller P, Murphy F (2020) Seed-fill-shift-repair: A redistricting heuristic for civic deliberation. *PloS one* 15(9):e0237935
- Hafford WB (2014) Linked open data and the Ur of the Chaldees project. Web post: <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/hafford/#p1>, iSAW Papers 7.7 (2014). Accessed: 2020-11-01
- Hall M, Kimbrough SO, Michalk W, Schneider J, Weindardt C (2013) Making solution pluralism in policy making accessible: Optimization of design and services for constituent well-being. In: *Extended abstracts of the 2013 ACM Conference on Human Factors in Computing Systems (CHI)*. (Workshop on Engaging the Human-Computer Interaction Community With Public Policymaking Internationally)
- Halpern O, LeCavalier J, Calvillo N, Pietsch W (2013) Test-bed urbanism. *Public Culture* 25(2):272–306
- Johann H, Hall M, Kimbrough SO, Quintus N, Weinhardt C (2014) Service district optimization: Usage of facility location methods and geographic information systems to analyze and optimize urban food retail distribution. Working paper, SSRN
- Keeney RL, Raiffa H (1993) *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge University Press, Cambridge, UK
- Kimbrough SO, Lau HC (2016) *Business Analytics for Decision Making*. CRC Press, Boca Ratan, FL

- Kitchin R (2013) The real-time city? big data and smart urbanism. *GeoJournal*
- Klausers F, Paasche T, Söderström O (2014) Michel Foucault and the smart city: power dynamics inhering in contemporary governing through code. *Environment and Planning* 32:869–885
- Laurel B, Oren T, Don A (1990) Issues in multimedia interface design: media integration and interface agents. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90* pp 133–139
- Mayer VL, Hillier MA Aand Bachhuber, Long JA (2014) Food insecurity, neighborhood food access, and food assistance in Philadelphia. *Journal of Urban Health* 91(6):1087–1097
- Messeter J, Johansson M (2008) Place-specific computing: conceptual design cases from urban contexts in four countries. In: *Proceedings of the 7th ACM conference on Designing interactive systems (DIS '08)*, ACM, New York, NY, pp 99–108
- Miller P, Kimbrough S, Schacht J (2018) Simulating redistricting in the classroom: A binding arbitration decision game using Louisiana census data. *PS: Political Science & Politics* 51(3):664–668
- Odom W (2010) ‘mate, we don’t need a chip to tell us the soil’s dry’: opportunities for designing interactive systems to support urban food production. In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*, ACM, ACM, New York, NY, pp 232–235
- Ploeg MV, Breneman V, Farrigan T, Hamrick K, Hopkins D, Kaufman P, Lin B, Nord M, Smith T, Williams R, Kinnison K, Olander C, Singh A, Tuckermanty E, Krantz-kent R, Polen C, McGowan H, Kim S (2009) Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences Report to Congress
- Schwartz R, Naaman M, Matni Z (2013) Making sense of cities using social media: Requirements for hyper-local data aggregation tools. In: *Proceedings of the International AAAI Conference on Weblogs and Social Media, AAAI*, pp 15–22
- Smyth M, Helgason I (2010) Informing the design of the future urban landscape. In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*, ACM, ACM, New York, NY, pp 438–439
- Söderström O, Paasche T, Klauser F (2014) Smart cities as corporate storytelling. *City* 18(3):307–320
- Sweller J (1994) Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction* 4(4):295–312
- Todd P, Benbasat I (1994) The Influence of Decision Aids on Choice Strategies - An Experimental-Analysis of the Role of Cognitive Effort. *Organizational Behavior and Human Decision Processes* 60(1):36–74
- Vanolo A (2014) Smart mentality: The smart city as disciplinary strategy. *Urban Studies* 51(5):883–898
- von Winterfeldt D, Edwards W (1986) *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, UK
- Xia C, Schwartz R, Xie K, Krebs A, Langdon A, Ting J, Naaman M (2014) Citybeat: real-time social media visualization of hyper-local city data. In: *Proceedings of the 23rd International Conference on World Wide Web*, pp 167–170
- Yoon KP, Hwang CL (1995) *Multiple Attribute Decision Making: An Introduction, Quantitative Applications in the Social Sciences*, vol 104. Sage, CA
- Yu A, Nahapetian A (2013) Participatory sensing for fighting food deserts. In: *Proceedings of the 8th International Conference on Body Area Networks (BodyNets '13)*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, pp 221–224
- Zheng Y, Capra L, Wolfson O, Yang H (2014) Urban computing: Concepts, methodologies, and applications. *ACM Trans Intell Syst Technol* 5(3):Article 38

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate

credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

