CASE STUDY

SHUTTERSTOCK/PETR84

Bridging the Amenity Gap:

How Emerging Data Can Help Detect Amenity Bypass

By Duong Vu, Anny Chih, Cari Gostic, Lesley Miller, and Sakariya Aynashe

The contents of this paper reflect the views of the authors and do not necessarily reflect the official views or policies of the related organizations.

With rapidly increasing populations, new cities are emerging, and older cities are expanding to keep up with the growing needs of residents. Given the complex and dynamic nature of cities, it has become a challenge for urban planners to identify and equitably support the needs of different communities without the proper tools and adequate data. In recent years, the development of large-scale computing power and the increasing diversity of sensing technologies have provided unconventional sources of information that urban planners can tap into to gain a broader understanding of the urban service coverage and identify existing gaps.

Defining Amenity Gaps and Amenity Bypass

When residents of a community have to travel farther away than normal (compared to other communities) to access a doctor's office, a grocery store, or a greenery park, these residents are most likely experiencing an amenity gap. When the community is in higher need for a given facility, this gap is exacerbated. When combined with relevant contextual information, anonymized connected vehicle data (CVD) can become an interesting proxy to evaluate those gaps.

Residents of service-deprived neighborhoods typically travel farther to access the amenities satisfying their needs, often bypassing some other amenities for quality or affordability reasons or because they run out of capacity. Some amenity bypasses are due to personal preferences: favorite doctor, off-leash dog park, near grandma's home, etc., but most amenity bypasses are the direct result of a shortage of facilities for that community. By combining amenity bypass detection with other background information such as demographics, we are able to analyze communities with true amenity gaps on a greater scale. In this article, we will explore the concept of amenity bypass in detail.

Literature Review

Amenity gaps in urban planning is a concept that has not been extensively explored and is generalized in current methods of discovery. Previous research focused primarily on specific amenities such as health care facilities;[1, 2] public transportation;[3, 4] greenery parks;[5, 6, 7] or food stores.[8, 9] Much of that research has pointed out the imbalance of amenity distribution and its correlation with the socioeconomic status of neighborhoods.[8, 1]

Detecting and filling these gaps would provide substantial positive impact and, as Kathryn Anderson (2018) suggested, more equitable provision of health care resources would mitigate negative impacts of segregation.[2]

One popular approach is the two-step floating catchment area method that measures spatial accessibility to a variety of amenities such as jobs, primary care physicians, parks, etc.[5, 8] This method focuses on the ratio of facility to population within an isochrone around the population centroid. The isochrone is either based on raw distance or on travel time. With the assumption that people will travel to the closest amenity, this method ignores amenity bypass due to overcapacity or unattainability. We suggest using GPS data to better model population movement and reveal which amenities people actually use, instead of assuming that people will go to the closest one. This work proposes a general framework to locate amenity gaps using GPS data.

Identifying and Contextualizing the Amenity Gap with Diverse Data Sources

The four main types of data sources used in this project were GPS trace data (connected vehicle data), demographic information at the zip code level (population, age, income level, education, etc.), amenity-specific information, and geospatial data (amenity locations, land use and zoning, etc.).

Connected Vehicle Data

Most previous research on amenity gaps used survey data or geospatial data to find proximity or location as a way to evaluate the level of accessibility of an amenity.[6, 10, 11] In this work, we look into anonymized connected vehicle data as an indicator of the movement of individuals. This data comes from vehicle sensors (GPS devices) recording specific events occurring in a passenger vehicle. The data used for this study spans from April 2018 to March 2019 (before the COVID-19 pandemic), and includes more than 10 million vehicle feeds, collected across 95 percent of the United States road network. Some relevant features of the dataset include journey ID (a unique identifier for any event that occurs between a trip start and a trip end for a given trip), event (trip start, trip end), location, and captured timestamp.

Demographic and Amenity-Specific Information

In addition to movement data, demographic information plays an important role in providing information about the characteristics of different communities, indicating the potential needs of each community for a specific amenity. For example, access to primary healthcare facilities is more critical for certain age groups, namely the very young and the elderly. To obtain this information, an Open Data Portal is one of the first official sources to gather demographic, education, and economic statistics data. Amenity-specific information also includes more details about each community, especially health statistics of locals, unemployment rates, and more.

Geospatial Data

A key part of this study examines the amenity bypass, which occurs when the closest amenity option is passed over in favor of one that is further away. As such, obtaining the spatial location for neighborhoods and the distances to several amenities around them is important for our analysis. We can do this by navigating to a city's Open Data Portal and obtaining a shapefile of official city boundaries. This shapefile is used to filter other datasets so our analysis can focus on movements and residents within the selected area. A listing of amenity locations often can be found from other official departments of a city or state. OpenStreetMap is another public data source to obtain amenity locations, as it provides a free, open-license map of the world with layers of information.

General Framework for Amenity Gap Analysis

In this work, we propose a general framework for amenity gap analysis as shown in Figure 1.

Defining Communities and Types of Amenities

Before looking into the amenity gap, a clear definition of communities will help the analysis stay focused and the comparison between communities stay consistent. In this study, we define a community as all people living within a zip code, because of its well-defined boundaries and the availability of demographic information. However, the definition of a community could be changed to better fit the purpose of other analyses.

The next step is to select the type of amenity on which to focus. In this study, we select two amenities, which are greenery parks and healthcare facilities. The amenity could be defined in a higher level of detail: mental health facilities, kid-friendly parks, etc. This selection determines the amount of amenity-related information the analysis would need such as specific health issue statistics, population age, etc.

Assigning Mobility Data Trips to Amenities

In order to analyze travel patterns to an amenity, we first filter to only trips with the origin or destination within the selected amenity's area. In most cases, an amenity's polygon requires a small buffer around it to include street parking. In other cases, we will need to draw or obtain the parking lot polygon nearby, which is illustrated in Figure 2. We then count the number of trips from these amenities to each community in our analysis, where the last few points of the trips are within the home's zip code.

Figure 1.Proposed general framework of amenity gap analysis in four steps.

Preprocessing of Different Data Sources

Most of the datasets used for our contextual analysis come from open data sources, which requires extensive cleaning and geopositioning. Standard cleaning includes, but is not limited to, formatting values to the right type (date, time, integer, string, etc.), imputing or excluding missing values, and consolidating different/misspelling of the same objects (Biscayne Park vs. BISCAYNE PARK vs. biscayne park). Many of the amenity locations only include an address (123 Main Street) instead of an actual representation on the map (a polygon shape). This leads to a process of geopositioning to find or draw the shape of the amenity. On the other hand, connected vehicle data comes from sensors, so its features are more defined with fewer inconsistencies. The big preprocessing step for this data is to summarize the individual trips into counts and filter by polygons (trips only start or end within the amenities of selected city boundaries).

Calculating Need Score

We define need score as a community's level of need for an accessible and attainable amenity. Accessibility is measured as median trip duration, and attainability refers to an amenity's affordability, quality, and capacity. To calculate the need score of a community, first we need to identify the risk feature from each community. A risk feature is a factor that raises a community's need for an accessible amenity. These features are created using demographic and amenity-specific information such as income, unemployment rates, education level, age, or proportion of population with disabilities. After identifying all risk factors, a weight will be assigned to each feature. These weights should reflect the indicated importance of a feature to a community's needs compared to other features. We then normalize each feature value and multiply by the proportional weight for that feature. Finally, the sum total value across all risk features gives the need score for each zip code area.

Discovering Amenity Bypass

The three main ideas that we pursue are: (i) an amenity analysis framework that could be applied to any type of amenity and any size of communities; (ii) leveraging connected vehicle data as a true source of labeled data for amenity usage; and (iii) based on the result of (ii), detecting amenity bypass, another indicator of amenity accessibility.

The motivation of this project is to find a way to help jurisdictions of different levels (city, county, state, province) to detect service gaps in a more equitable way. Instead of having to select one or two specific communities or amenities to perform an analysis, we hope to introduce a more general approach and guideline on important data sources to succeed with an amenity gap analysis. In order to ensure equitable urban design, an unbiased evaluation of the

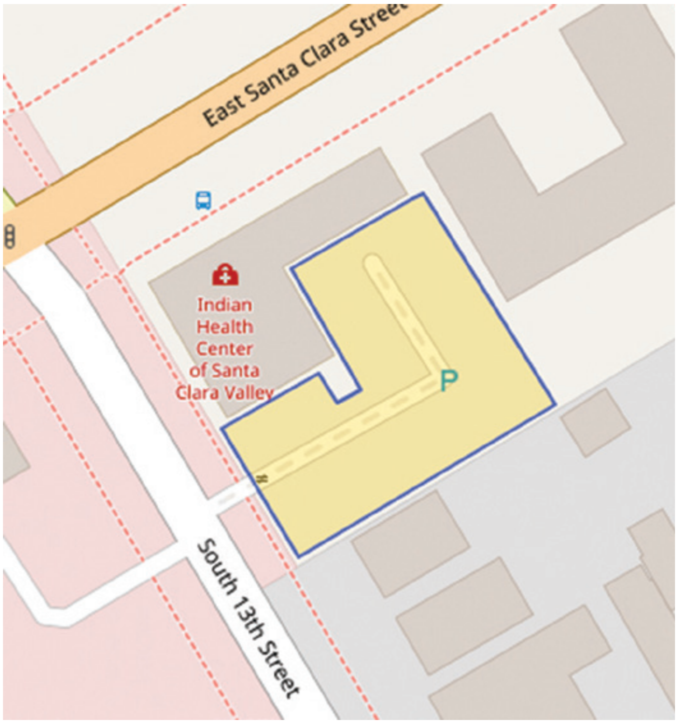


Figure 2. Parking lot near a medical center. This is just an illustration, and this city is not the city in our study.

current service accessibility should be done without any preference in mind. Moreover, by using connected vehicle data, we expect to include the majority of the population, such as underserved areas where surveys and official reports may not be utilized.

This dataset also leads us to a new concept: amenity bypass, which could be a separate study by itself. Note that we measure accessibility of an amenity by median trip duration, which provides more insight into which amenity the resident of a community actually traveled to instead of the closest one based on distance. We defined amenity bypass as occurring when a nearby amenity is passed over in favor of another further away due to its lack of attainability for a variety of reasons. If an amenity exists within a community but is too expensive, crowded, or run down to meet the needs of its residents, this still constitutes an amenity gap.

Figure 3 shows the number of trips from each pink triangle to the blue triangles, which represent a park amenity. The orange triangle in this figure represents the amenity bypass. We can see that although the orange triangle is closer to another park amenity, a portion of residents opted to travel to another one further away (represented by the blue triangles).

Furthermore, these shortages of service are contextualized by open source demographics, economics, educational statistics, and amenity-specific information to determine whether these gaps exist in more vulnerable communities.

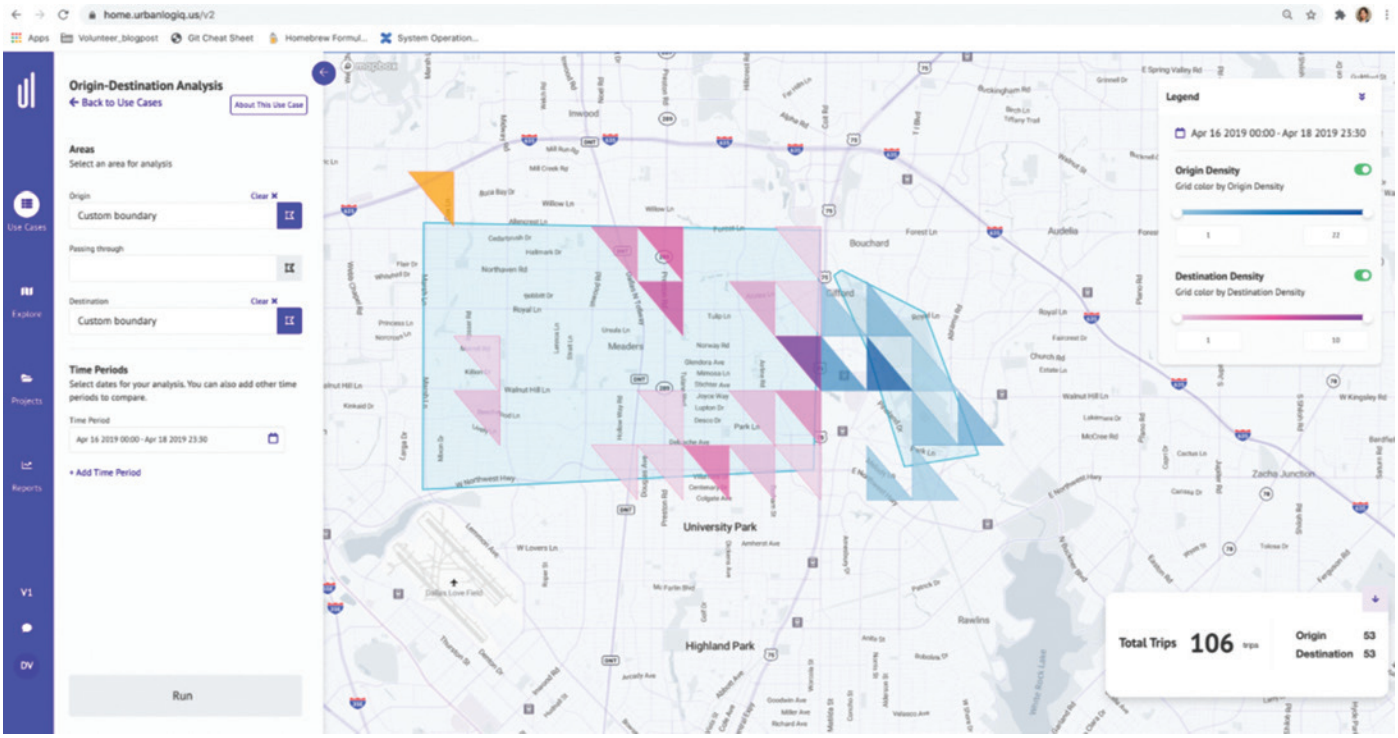


Figure 3. Anonymized trips assigned and grouped by triangles of origin and destination, showing people coming from both nearby and further away to use a greenery park. The orange triangle is an example of amenity bypass. This is just an illustration, and this city is not the city in our study.

Assumptions and Future Work

There are a number of assumptions built into this research that should be considered when using any insights derived from the analysis. First, we assume the connected vehicles are well-distributed throughout the city. An additional penetration rate analysis for each community would be recommended for confirmation. The second assumption is that people who travel to an amenity intend to use the amenity. This excludes all the officials and workers for that amenity, if any. Lastly, trips to an amenity always start or end at home zone areas.

We acknowledge that the movement data coming from car sensors may or may not represent the whole population. Each community has its own ratio for different modes of transportation such as walking, biking, driving, or public transit. In a community where there are more diverse ways of getting around, we suggest considering mobility data coming from mobile phones as another additional source of information.

Conclusion

Traditional analysis with survey data and official record data cannot adequately explain the movement of different groups of people within a community. Many residents of a city will not appear on official data records because of homelessness, lack of a fixed address, undocumented status, and many other reasons, leaving these groups underrepresented. This has, to date, made planning more biased toward those for whom we have census data. Having another source of data such as anonymized connected vehicle data to analyze supports the deep dive into community behaviors in a more scalable way. Diversifying the sources of data used gives us an opportunity to identify existing gaps and understand gap-driven population movements, ultimately helping increase the standards of inclusiveness in the city. In addition, this new source of data helps discover amenity bypass, an additional indicator of inaccessibility in the service network for the local community. **itej**

Acknowledgement

Thank you to UrbanLogiq for providing the data, facilities, and other resources for this project, especially my coworkers: Rachel Rasmussen and Anabel Chuinard for reviewing the paper with me, Morgan Cassels for helping on the software development of the project, and Julien Refour for helping on idea development. The core team worked on this analysis under the capstone project between UrbanLogiq and the University of British Columbia.

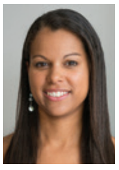
References

1. Caldwell, Julia T; Ford, Chandra L.; Wallace, Steven P.; Wang, May C.; M. Lois, United States, Los Angeles, 2018. "Racial And Ethnic Residential Segregation and Access to Health Care in Rural Areas" 104–12. https://doi.org/10.1016/j.healthplace.2016.11.015.
2. Anderson, Kathryn Freeman. 2018. "Racial/Ethnic Residential Segregation, the Distribution of Physician's Offices and Access to Health Care: The Case of Houston, Texas." *Social Sciences* 7(8). https://doi.org/10.3390/socsci7080119.
3. Bullard, Robert D. 2003. "Addressing Urban Transportation Equity in the United States." *Fordham Urban Law Journal* 31 (5). https://ir.lawnet.fordham.edu/ulj/vol31/iss5/2.
4. Glaeser, Edward L., Kahn, Matthew E., and Rappaport, Jordan. 2008. "Why Do the Poor Live in Cities? The Role of Public Transportation." *Journal of Urban Economics* 63 (1): 1–24. https://doi.org/10.1016/j.jue.2006.12.004.
5. Luo, W. and Wang, F., 2003a. "Spatial Accessibility to Primary Care And Physician Shortage Area Designation: A Case Study in Illinois with GIS Approaches." In: Skinner, R., Khan, O. (Eds.), Geographic Information Systems and Health Applications. Idea Group Publishing, Hershey, PA, USA, pp. 260-278.
6. Ibes, Dorothy C. 2015. "A Multi-Dimensional Classification and Equity Analysis of an Urban Park System: A Novel Methodology and Case Study Application." *Landscape and Urban Planning* 137: 122–37. https://doi.org/10.1016/j.landurbplan.2014.12.014.
7. Potestio, Melissa L.; Patel, Alka B.; Powell, Christopher D.; McNeil, Deborah A.; Jacobson, R. Daniel; and McLaren, Lindsay. 2009. "Is There an Association between Spatial Access to Parks/Green Space and Childhood Overweight/Obesity in Calgary, Canada?" *International Journal of Behavioral Nutrition and Physical Activity* 6: 1–10. https://doi.org/10.1186/1479-5868-6-77.
8. Luo, Wei, and Fahui Wang. 2003. "Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region." *Environment and Planning B: Planning and Design* 30 (6): 865–84. https://doi.org/10.1068/b29120.
9. Alwitt, Linda F., and Donley, Thomas D. "Retail Stores in Poor Urban Neighborhoods." The Journal of Consumer Affairs 31, No. 1 (1997): 139-64. Accessed February 15, 2021. http://www.jstor.org/stable/23859601.
10. Powell, Lisa M.; Slater, Sandy; Mirtcheva, Donka; Bao, Yanjun; and Chaloupka, Frank J. 2007. "Food Store Availability and Neighborhood Characteristics in the United States." *Preventive Medicine* 44 (3): 189–95. https://doi.org/10.1016/j.ypmed.2006.08.008.
11. Jenkins, Gavin R.; Yuen, Hon K.; Rose, Emily J.; Maher, Amy I.; Gregory, Kristina C.; and Cotton, Megan E. 2015. "Disparities in Quality of Park Play Spaces between Two Cities with Diverse Income and Race/Ethnicity Composition: A Pilot Study." *International Journal of Environmental Research and Public Health* 12 (7): 8009–22. https://doi.org/10.3390/ijerph120708009.

Duong Vu *is a data scientist at UrbanLogiq, where her focuses are on geospatial and urban computing problems. Previously, she worked at the Indiana Auditor of State on data analysis and automating accounting books. Duong has a master's degree in data science from the University of British Columbia and a master of jurisprudence from the Indiana University Robert H. McKinney School of Law. Since 2018, Duong has led the Vancouver Women in Data Science, organizing conferences, talks, workshops, and career fairs.*

Anny Chih *is a senior business analyst at Fraser Health Authority and has a master's degree in data science from the University of British Columbia. Her work provides healthcare providers with analytical insights needed to make data-informed decisions, and she creates efficiencies within the organization by automating processes using Python and SQL. Prior to working in healthcare, Anny specialized in market research and analytics within the media and advertising industry across the United States and Canada.*

Cari Gostic *has an undergraduate degree in atmospheric science from Cornell University, and recently earned a master's degree in data science from the University of British Columbia. She enjoys applying herself to projects with environmental or societal benefits. She currently works as a data scientist for Sonoma Technology, Inc., an environmental consulting firm that specializes in air quality solutions.*



Lesley Miller *is a data scientist at the University of Washington who is passionate about using data science tools to facilitate sustainable stewardship and conservation of natural resources. She develops and maintains the core software tools used across various outdoor recreation research projects and utilizes natural language processing algorithms to build models that provide insight into forest visitors' interactions with nature and designated wilderness. She has a background in computational biology and genome evolution as an undergraduate at the University of British Columbia and a master's degree in data science from UBC.*

Sakariya Aynashe *graduated from the University of British Columbia and has a master's degree in data science. He is currently the program manager and data scientist at Tech Labs Academy. Also a recipient of the Mandela Washington Fellowship by the United States Department of State, Sakariya's recent scientific interests include AI-driven solutions, machine learning, and cognitive computing.*