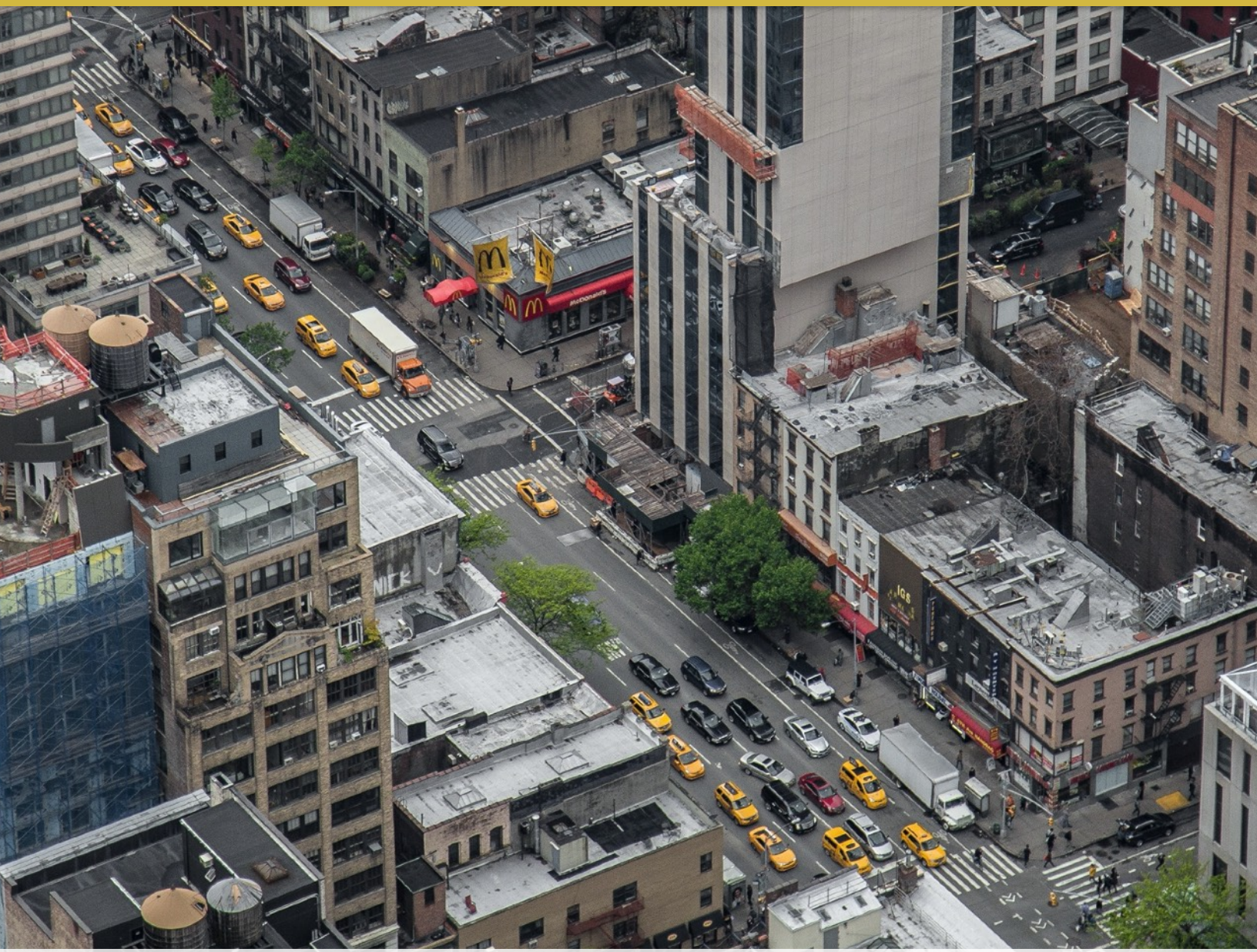
A Data Collaborative Case Study

THE ATLAS OF INEQUALITY AND CUEBIQ'S DATA FOR GOOD INITIATIVE

Using location data to quantify segregation of urban social spaces.

Michelle Winowatan, Andrew Young, Stefaan G. Verhulst



April 2020

GOVLAB

Photo by Christopher Burns on Unsplash

GOVLAB

EXECUTIVE SUMMARY

THE ATLAS OF INEQUALITY AND CUEBIQ'S DATA FOR GOOD INITIATIVE

Using location data to quantify segregation of urban social spaces.

Summary: The Atlas of Inequality is a research initiative led by scientists at the MIT Media Lab and Universidad Carlos III de Madrid. It is a project within the larger Human Dynamics research initiative at the MIT Media Lab, which investigates how computational social science can improve society, government, and companies. Using multiple big data sources, MIT Media Lab researchers seek to understand how people move in urban spaces and how that movement influences or is influenced by income. Among the datasets used in this initiative was location data provided by Cuebiq, through its Data for Good initiative. Cuebiq offers location-intelligence services to approved research and nonprofit organizations seeking to address public problems. To date, the Atlas has published maps of inequality in eleven cities in the United States. Through the Atlas, the researchers hope to raise public awareness about segregation of social mobility in United States cities resulting from economic inequality and support evidence-based policymaking to address the issue.

Data Collaborative Model: Based on the typology of data collaborative practice areas developed by The GovLab, the use of Cuebiq's location data by MIT Media Lab researchers for the Atlas of Inequality initiative is an example of the **research and analysis partnership** model of data collaboration, specifically a **data transfer** approach. In this approach, companies provide data to partners for analysis, sometimes under the banner of "data philanthropy." Access to data remains highly restrictive, with only specific partners able to analyze the assets provided. Approved uses are also determined in a somewhat cooperative manner, often with some agreement outlining how and why parties requesting access to data will put it to use.

Data Stewardship Approach: Through its Data for Good initiative, Cuebiq provided access to location data that supported the Atlas's development. The Data for Good initiative is an example of how active data stewardship can help data-holding companies find public utility in their data beyond their day-to-day business. Such utility includes the creation of scientific and public value by securely providing academic researchers and nonprofit organizations access to Cuebiq's data assets. These collaborations involve a licensing agreement between Cuebiq and its partners, where all the parties pledge to adhere to certain privacy and data-handling standards to prevent potential risks that may arise, such as data breach and misuse. Further, Cuebiq also created the Director of Research Partnership and Data for Good role to carry out various data stewards functions, including internal coordination, partnership building, and risk assessment. This approach can serve as a good model for how companies can practice responsible data-sharing.

GOVLAB

OPERATIONAL VARIABLES

MIT Media Lab and Cuebiq Data Collaborative Operational Variables. Detailed description of each variable can be found here.

CASE STUDY

SETTING THE SCENE

Economic inequality is one of the biggest challenges facing the United States and the world today. One way to observe economic inequality is to examine segregation by income in neighborhoods and how this segregation impacts various aspects of public life, such as availability of public schools, housing, and public transportation. Within the US, census data can provide insight into where people live and what type of access they have to public services, but do not have the ability to provide insights on a more granular level, on its own and by design.

As a part of MIT Media Lab's Human Dynamics research, the Atlas of Inequality project studies how economic inequality affects people's movement in urban spaces.[1] Typically, research on inequality and segregation focuses on mapping where people with certain levels of income live or what sociologists call "first places."[2] Led by data scientists Esteban Moro and Alex Pentland, this project investigates the segregation that happens as people go to their "second places" (e.g. work or school) and "third places" (e.g. grocery stores, barber shops, cafes, and parks). With an

---

[1] Xiaowen Dong et al., "Segregated Interactions in Urban and Online Spaces," *ArXiv:1911.04027 [Physics]*, November 10, 2019, http://arxiv.org/abs/1911.04027.

[2] Sociologist Ray Oldenburg defines home as "first place," workplace as "second place," and places other than the home and the workplace as "third places." See Ray Oldenburg, *The Great Good Place: Cafés, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community* (Marlowe, 1998).

expanded view of how segregation affects people's use of public or commercial places, researchers could provide urban policymakers an opportunity to reduce segregation.[3] Often, collecting this data would require a great deal of time and resources going from place to place in order to conduct observations, surveys, and interviews.

Recognizing the value of private-sector data in helping the public sector improve people's lives, the location-intelligence services company Cuebiq[4] established a data collaborative initiative, Data for Good, in 2017. This initiative provides academic researchers and civil society organizations working on issues such as public health, urban mobility, disaster response access to Cuebiq's anonymized and aggregated location datasets.[5] Through this initiative, Cuebiq has partnered with at least 30 institutions.[6]

BOX 1: Data for Good Application Procedure

To obtain Cuebiq's data assets, interested researchers must go through a three-step process. First, researchers or organizations interested in accessing Cuebiq's data need to submit a proposal outlining the research or project, including the context, objectives, timeline, the type of data that the researchers need, and how they plan to use the data. Cuebiq's Data for Good team assesses the proposal based on the public value that could be created from the project. If Cuebiq deems the project a good fit for the Data for Good initiative, it will initiate the second step, which is establishing a data licensing contract with the partner.

This data-licensing contract specifies the terms of use of the data, proper usage of the data, and requires adherence to Cuebiq's privacy policy, among other provisions. Cuebiq typically issues a 12-month data license which can be extended following further review and approval. Once Cuebiq and its partner agree on this contract, the implementation of the data collaborative and the third step for Cuebiq begins. Cuebiq subsequently conducts regular check-ins to ensure that the partner can optimally and responsibly use the data.

The establishment of Cuebiq's Data for Good initiative was driven by three main values.[7] First, the company sees humanitarian and scientific impact potential in its data assets, if they are shared with academic researchers or public-oriented organizations. Second, from an ethical perspective, Cuebiq considers that data generated by users should also provide benefits for the users. Third, through this initiative, Cuebiq hopes to build public trust in how the company uses the data.

With location data from Cuebiq, the MIT Media Lab researchers were able to identify third places and whether their patrons came from many different income levels or if they were frequented by one income group only. The project studied 11 major metropolitan areas—where inequality and

[3] Xiaowen Dong, *supra* note 1.

[4] Cuebiq is a technology company which offers location services and softwares for businesses. "Cuebiq," Cuebiq, accessed January 23, 2020, https://www.cuebiq.com/.

[5] "Data for Good," Cuebiq, accessed January 8, 2020, https://www.cuebiq.com/about/data-for-good/.

[6] Brennan Lake, Interview by The GovLab, August 29, 2019.

[7] Brennan Lake, Interview by Amanda DeSantis, July 2019.

segregation are acute—comprising 83 million people in Seattle, Los Angeles, San Francisco, Dallas, Chicago, Detroit, Philadelphia, Washington DC, New York, Boston, and Miami. Maps of these 11 cities are now accessible through the Atlas of Inequality's website.[8]

ESTABLISHING THE DATA COLLABORATIVE

In 2017, Cuebiq and the Atlas of Inequality team began their engagement through the Data for Good initiative.[9] Through this mechanism, MIT Media Lab obtained access to location data from 150,000 anonymous mobile devices collected between 2016 and 2017.[10] Cuebiq's dataset significantly improved the granularity of the analysis relative to census and social media data, which lacks sufficient geo-location information for the purpose mentioned above.[11]

After obtaining Cuebiq's data, the MIT Media Lab researchers mapped out Cuebiq's dataset across 30,000 places obtained through Foursquare's Places API[12]—a freemium service offered by the location-based social app Foursquare that gives real-time access to Foursquare's global venue database.[13] This API provides details of venues, including their location, type, and hours. It is a tool intended to help developers build location-supported applications.

The researchers looked at the area where the anonymous user spent the most time between 8:00 PM and 4:00 AM to determine their home neighborhood and cross-referenced it with census data of the median income of the area to determine the income group of the user.[14] The researchers categorized individuals into four different groups: low, middle-low, middle-high, and high income groups. To prevent re-identification of individuals in the dataset, the researchers narrowed these places to those that were verified or have at least five user check-ins within a six-month period.[15] To further preserve privacy, the researchers only used places that had at least 20 different users within the six-month period.[16]

MIT Media Lab researchers also obtained assistance from CARTO, a company that offers maps service and tools, in visualizing the insights on a map and creating the website.[17]

---

[8] "The Atlas of Inequality," The Atlas of Inequality, accessed January 17, 2020, https://inequality.media.mit.edu/#.

[9] Esteban Moro, Interview by The GovLab, November 13, 2019.

[10] "The Atlas of Inequality," *supra* note 8.

[11] Esteban Moro, *supra* note 9.

[12] "The Atlas of Inequality," *supra* note 8.

[13] Freemium is a pricing mechanism where users are given free access to a product or service, but are charged a certain amount of money for additional features. "Places API - Foursquare Developer," accessed January 17, 2020, https://developer.foursquare.com/docs/api.

[14] "The Atlas of Inequality," *supra* note 8.

[15] *Ibid.*

[16] *Ibid.*

[17] Esteban Moro, *supra* note 9.

Figure 1: In this map of inequality of San Francisco, users can see that a bus station near the Golden Gate bridge has a 42.8% inequality index, where individuals from the high income group overwhelmingly dominate the visitation to this venue.

OUTCOME AND IMPACT

Since its launch in 2017, the Data for Good initiative has facilitated at least 30 data licensing contracts, including with MIT Media Lab for the Atlas of Inequality project.[18] In this section, we focus our analysis on the intended impact and outcome of the Atlas of Inequality project.

Intended Impact

Through the Atlas of Inequality project, the MIT Media Lab researchers intend to, first, measure inequality and segregation and create public awareness on how inequality segregates individual movement and how individual behaviors can subconsciously reinforce segregation.[19] Second, they aim to further investigate this phenomenon using machine learning models and other data science tools in order to enhance the understanding of inequality and social segregation. Third, they seek to work with policymakers to address urban inequality informed by insights generated by the research.

Outcomes

The Atlas of Inequality Platform: As of January 2020, the Atlas of Inequality's website provides map visualizations of eleven metropolitan areas in the United States. These maps contain thousands of points of interest in those cities, each accompanied by an "inequality index." A place can have a high inequality index when the majority of its visitors come from one or two

[18] Brennan Lake, *supra* note 6.

[19] Esteban Moro, *supra* note 9.

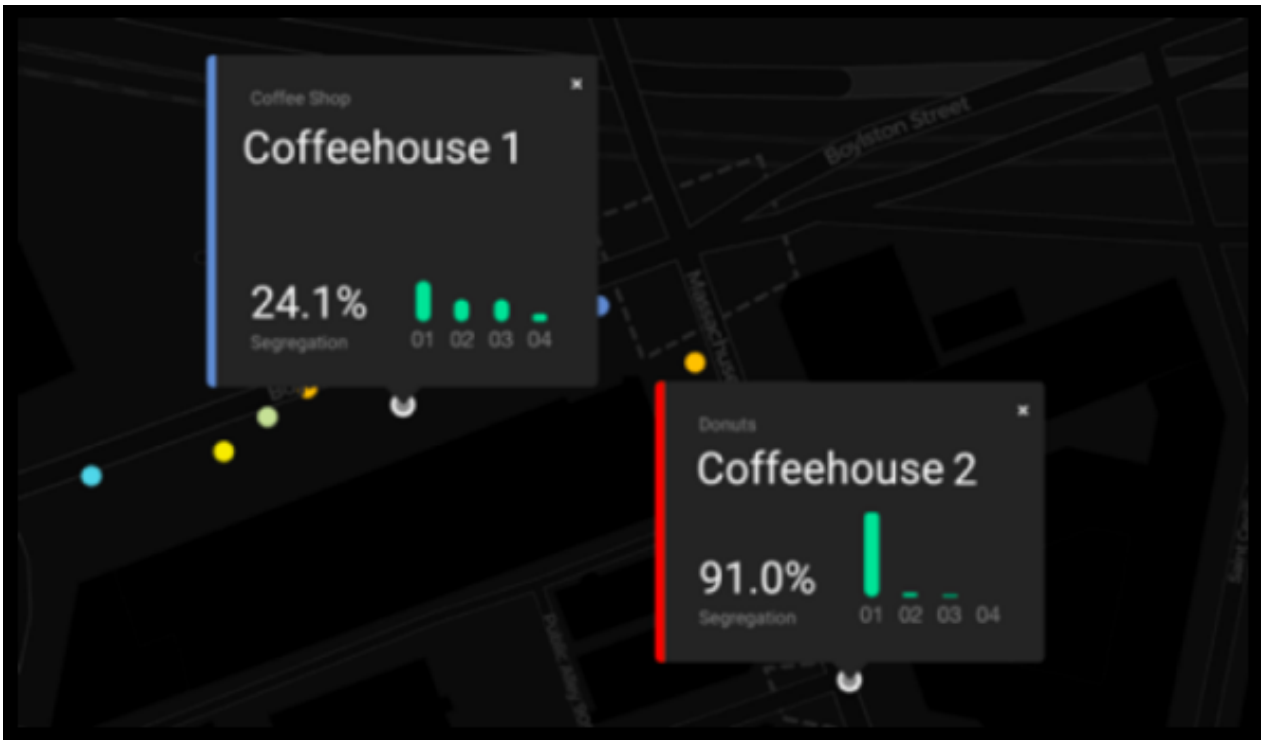


Figure 2: A comparison of the inequality index of two coffee shops located on the same block. (Credit: Atlas of Inequality, 2019)

income groups only.[20] Users visiting the website can zoom into a specific venue and find the economic diversity of the patrons of that particular place.

Figure 2 provides an example of the types of insight that can be extracted from the Atlas featuring two coffee shops in Boston that are located on the same block. This illustration shows that despite being located in the same area, two similar places can attract very different types of customers in terms of income level.

Seeding collaboration with policymakers: To turn insights into action, the Atlas of Inequality team is disseminating their research findings to several city governments to identify policy intervention that can make public spaces, transportation, and community businesses more equitable.[21] While no concrete policy plans have materialized from this project, the researchers are taking steps, as mentioned above, to support the uptake and use of its findings.

RISKS AND RISK MITIGATION

In this section, we discuss the risks that might arise in the data-sharing engagement between MIT Media Lab and Cuebiq and how the stakeholders mitigated these risks.

---

[20] "The Atlas of Inequality," *supra* note 8.

[21] Esteban Moro, *supra* note 9.

Figure 3: In this map of inequality of New York City, users can see that the Central Park Zoo has a 72.7% inequality index since it has mostly individuals from low-middle income group dominating the visitation to this venue.

Data Privacy and Security: Recognizing the privacy risks of location data, both Cuebiq and the MIT Media Lab team employed several measures to prevent privacy issues from arising. Cuebiq has a privacy policy compliant with the European Union's General Data Protection Regulation and California Consumer Privacy Act—two of the most comprehensive legal instruments for consumer's data protection.[22] Additionally, the data Cuebiq provided to the MIT Media Lab had been anonymized prior to analysis.[23] To further prevent de-anonymization and re-identification, the researchers only studied places that have at least five check-ins and 20 different users within a six-month period.[24] Finally, the Atlas of Inequality only published aggregated numbers of the different income groups that visited the points of interest.[25]

Data misuse. Cuebiq tries to mitigate the risk of misuse by ensuring that partners who get access to the data have legitimate reasons to use the data and adheres to Cuebiq's data protection policy.[26] Partners who submit a proposal for the Data for Good program must first demonstrate their ability to meet these requirements. If Cuebiq determines that the proposal is a good fit, then the researchers can access the data. Throughout a project's lifespan, Cuebiq and its partners also conduct regular meetings to ensure project compliance, among others.

---

[22] "Privacy Policy," Cuebiq, accessed January 17, 2020, https://www.cuebiq.com/privacypolicy/.

[23] "The Atlas of Inequality," *supra* note 8.

[24] *Ibid.*

[25] Esteban Moro, *supra* note 9.

[26] Brennan Lake, *supra* note 6.

The MIT Media Lab researchers also recognized the potential for data misuse due to the biases that existed within the dataset, discussed in further detail below.[27] If not corrected, inaccurate insights could misinform policy, leading to harm for a certain population. Cognizant of the inability to remove all of the biases, the researchers advise the users or readers of their research to be aware of this limitation in the data.

LESSONS LEARNED – ENABLERS

Below we list the key features within the data-sharing ecosystem that enabled MIT Media Lab's implementation of the Atlas of Inequality project.

MIT Media Lab Data Science Expertise: The MIT Media Lab researchers' expertise in computational and mathematical science enabled the analysis of the data with minimal support from Cuebiq. As a technology startup, Cuebiq has limited human resources that can be allocated to support its Data for Good partners.[28] As such, it is critical that partners possess the data science capacity to independently make good use of the data that Cuebiq provides.

Access to Open-Source/Public Datasets. The Atlas of Inequality project was made possible by the availability of public or open-source datasets, as well as private datasets. Two key public datasets used by researchers at MIT Media Lab included national census data and Places API offered by Foursquare.[29] The census data helped the researchers determine the income level of observed individuals, by matching the individuals with the average income level of the place they live in. The Foursquare API enabled the researchers to identify the points of interest (e.g. coffee shops, parks, libraries) at which inequality could be measured. This component was important for the Atlas of Inequality as it helped developers and researchers alike in creating programs or platforms such as the Atlas of Inequality.

Systematic Private-Sector Data Sharing Infrastructure: Private-sector data provided by Cuebiq was also critical for the Atlas of Inequality analysis. Cuebiq's Data for Good initiative created a systematic and secure way for accessing its location datasets at no monetary cost, which enabled MIT Media Lab researchers to measure inequality and segregation at a granular level. The MIT researchers acknowledged some alternative datasets—such as street sensor data, open city data, and census data—that could also be used to measure inequality and segregation.[30] However, these datasets would not have been able to portray inequality at the level of place as the Atlas does now.

Technical Assistance from CARTO: CARTO, a company that offers geospatial tools and services, provided its visualization expertise for this project by helping visualize the findings of the study

---

[27] Esteban Moro, *supra* note 9.

[28] Brennan Lake, *supra* note 6.

[29] Esteban Moro, *supra* note 9. *Places API is a freemium service offered by Foursquare. It provides developers with Foursquare's maps and venue data.*

[30] Esteban Moro, *supra* note 9.

on a map that is now publicly accessible online.[31] In an interview, Esteban Moro acknowledged the importance of making sure that academic research is accessible for the public to consume.[32] CARTO's expertise in GIS and spatial data and tool to this initiative resulted in the creation of the Atlas of Inequality website which now visualizes maps of inequality in multiple cities in the United States.

LESSONS LEARNED – CHALLENGES

The challenges within the data collaboration between Cuebiq and MIT Media Lab included:

Limitation of Data. MIT Media Lab researchers tested if its data resembled the lived experience in cities under examination. To check if its calculation of an individual's dwell time in a place was accurate, the researchers analyzed the Cuebiq dataset to estimate attendance at professional sports games.[33] The result showed that the calculation made by the researchers was within a few thousand of the official attendance count.

Nevertheless, the researchers acknowledged some biases are impossible to correct in this dataset. The datasets do not include the digitally invisible groups who typically do not have access to internet-connected devices and other services that generate big data, such as smartphones, credit and debit cards, or other Internet of Things devices. For example, this data does not include people who are in extreme poverty, such as those who are homeless who typically do not own or have access to mobile devices.[34] Other underrepresented groups can include the elderly, the poor, children and other vulnerable communities.[35]

Reputational Challenge: As a company that offers location intelligence services, Cuebiq (among other companies) has been the subject of some scrutiny in the press regarding individual privacy concerns.[36] This may affect Cuebiq's expansion of its Data for Good partnership as partners who could benefit from Cuebiq's data may hesitate to participate due to potential reputation risks associated with accessing granular location data (even in the public interest).

NEXT STEPS

The Atlas of Inequality has made progress against its objective of advancing understanding of inequality and segregation in cities. Now the team is turning more of its focus on its latter two objectives:

---

[31] *Ibid.*

[32] *Ibid.*

[33] "The Atlas of Inequality," *supra* note 8.

[34] *Ibid.*

[35] Justin Longo et al., "Technology Use, Exposure to Natural Hazards, and Being Digitally Invisible: Implications for Policy Analytics," *Policy & Internet* 9, no. 1 (2017): 76–108, https://doi.org/10.1002/poi3.144.

[36] Stuart A. Thompson and Charlie Warzel, "Smartphones Are Spies. Here's Whom They Report To.," The New York Times, December 20, 2019, https://www.nytimes.com/interactive/2019/12/20/opinion/location-tracking-smartphone-marketing.html.

1. *Investigating the underlying factors.* After measuring the level of inequality across eleven cities in the United States and visualizing it on the website, the researchers intend to explore the drivers of this phenomenon further using computational tools to further improve their understanding of income inequality and social segregation.[37] The findings of these tests will be published in a forthcoming academic paper.

2. *Turning insights into action.* The Atlas has successfully turned data into valuable insights about inequality and segregation. Following this effort, the researchers will continue to communicate their findings of this project to city governments to help policymakers devise interventions that could reduce segregation in social spaces. The researchers report talking to city governments about public spaces, transportation, and local businesses.[38] This effort to engage those positioned to take meaningful action is an example of concrete practice that can help turn insights into action, which is an important component of a data-driven initiative.

Cuebiq, on the other hand, continues to expand on its Data for Good partnership. Among its recent partners is UNICEF, which will channel Cuebiq's data into UNICEF's Magic Box platform to be used for humanitarian purposes.[39] This platform facilitates secure private-sector data sharing and ethical data use for humanitarian purposes.[40]

In early 2020, the MIT Media Lab team began using Cuebiq data to study the effectiveness of social distancing policies in New York City in light of the COVID-19 pandemic.[41] This work represents another example of the potential value of responsibly providing researchers working in the public interest with functional access to certain private-sector datasets.

---

[37] Esteban Moro, *supra* note 9.

[38] *Ibid.*

[39] "Cuebiq's Data for Good Program Provides UNICEF with High-Precision Human Mobility Data for Real-Time Response to Humanitarian Needs," Cuebiq, accessed January 17, 2020, https://www.cuebiq.com/press/cuebiq-dataforgood-unicef-partnership/.

[40] "Data Science and Artificial Intelligence," accessed January 17, 2020, https://www.unicef.org/innovation/Magicbox.

[41] "Social Distancing in New York City | Home," accessed April 3, 2020, http://curveflattening.media.mit.edu/posts/social-distancing-new-york-city/.

SOURCES

Cuebiq. "Cuebiq." Accessed January 23, 2020. https://www.cuebiq.com/.

Cuebiq. "Cuebiq's Data for Good Program Provides UNICEF with High-Precision Human Mobility Data for Real-Time Response to Humanitarian Needs." Accessed January 17, 2020. https://www.cuebiq.com/press/cuebiq-dataforgood-unicef-partnership/.

Cuebiq. "Data for Good." Accessed January 8, 2020. https://www.cuebiq.com/about/data-for-good/.

"Data Science and Artificial Intelligence." Accessed January 17, 2020. https://www.unicef.org/innovation/Magicbox.

Dong, Xiaowen, Alfredo J. Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland. "Segregated Interactions in Urban and Online Spaces." *ArXiv:1911.04027 [Physics]*, November 10, 2019. http://arxiv.org/abs/1911.04027.

Lake, Brennan. Interview by Amanda DeSantis, July 2019.

———. Interview by The GovLab, August 29, 2019.

———. "Request for Comments: Cuebiq's Data for Good Initiative Case Study," February 1, 2020.

Longo, Justin, Evan Kuras, Holly Smith, David M. Hondula, and Erik Johnston. "Technology Use, Exposure to Natural Hazards, and Being Digitally Invisible: Implications for Policy Analytics." *Policy & Internet* 9, no. 1 (2017): 76–108. https://doi.org/10.1002/poi3.144.

Moro, Esteban. Interview by The GovLab, November 13, 2019.

Oldenburg, Ray. *The Great Good Place: Cafés, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. Marlowe, 1998.

"Places API - Foursquare Developer." Accessed January 17, 2020. https://developer.foursquare.com/docs/api.

Cuebiq. “Privacy Policy.” Accessed January 17, 2020. https://www.cuebiq.com/privacypolicy/.

“Social Distancing in New York City | Home.” Accessed April 3, 2020. http://curveflattening.media.mit.edu/posts/social-distancing-new-york-city/.

The Atlas of Inequality. “The Atlas of Inequality.” Accessed January 17, 2020. https://inequality.media.mit.edu/#.

Thompson, Stuart A., and Charlie Warzel. “Smartphones Are Spies. Here’s Whom They Report To.” The New York Times, December 20, 2019. https://www.nytimes.com/interactive/2019/12/20/opinion/location-tracking-smartphone-marketing.html.