Visualizing Equity: A Data Science for Social Good Tool and Model for Seattle

Rachel Berney
University of Washington
Seattle, WA
rberney@uw.edu

Bernease Herman
University of Washington
Seattle, WA
bernease@uw.edu

Gundula Proksch
University of Washington
Seattle, WA
prokschg@uw.edu

Hillary Dawkins
University of Washington
Seattle, WA
hdawkins@uw.edu

Jacob Kovacs
University of Washington
Seattle, WA
kovjac19@uw.edu

Yahui Ma
University of Washington
Seattle, WA
maya16@uw.edu

Jacob Rich
University of Wisconsin
Madison, WI
jrich3@wisc.edu

Amanda Tan
University of Washington
Seattle, WA
amandach@uw.edu

ABSTRACT

Our paper presents the products, preliminary findings, and methodology of the *Equitable Futures* project, an investigation into the active gentrification process and increasingly inequitable access to opportunities in Seattle. The project is currently underway at the University of Washington eScience Institute's Data Science for Social Good (DSSG) summer program, and connects issues usually analyzed by people working in the built environment, geography, sociology, economics, social work and local governments with a heavy focus on accessibility and transparency.

The project team is developing a tool that allows stakeholders in the city's development process to analyze, model, and visualize existing trends and the impact of potential changes in the built environment. Our tool uses publicly available data to model and visualize equity indicators on the city and neighborhood scale. In this paper, we focus on the incorporation of data transparency and model interpretability strategies into our project through the use of an accessible literature review workflow, open source software tools, detailed data preprocessing methodology, our interactive mapping tool, and the use of interpretable and graphical statistical tools.

1 INTRODUCTION

Our team is running the *Equitable Futures* project in the University of Washington's Data Science for Social Good Summer Program. We are focused on creating a visualization tool and a structural equation model to help us analyze and understand issues of equity in Seattle, Washington. One of the main challenges of identifying, analyzing, and understanding issues of equity in cities is that the bodies of literature pertaining to urban equity include a multitude of indicators but there remains a limited understanding of which ones are most significant. Data science methodologies can shape our approach; we are particularly interested in how data science can be brought to bear in helping us to visualize different policy decisions across the city and in neighborhoods. Further, we are using multivariate statistical analysis and building a structural equation model to help us analyze the strength of relationships among selected indicators as well as how the indicators shape latent variables such as gentrification. The visualization tool and model will be of use to a range of stakeholders including policymakers, researchers, and citizens.

The *Equitable Futures* project presented in this paper is currently under development, but will be completed in advance of the conference. The product, findings, and reflections on the collaboration process will be further refined for the presentation.

1.1 Program and Team Structure

The eScience Institute at the University of Washington runs an annual Data Science for Social Good (UW DSSG) program that selects four projects each year to train students from a wide range of disciplines in data science methodologies while helping community members execute social good projects, often with an urban focus [7]. Each UW DSSG team consists of four student fellows, two data science leads, and one or more project leads who proposed the project. The DSSG program runs for ten weeks during the summer with some lead time beforehand for project leads and data science mentors to develop project goals. The main objective of the program is to arrive at actionable information through data-driven discovery.

With permission, the program is modeled after the larger Eric & Wendy Schmidt Data Science for Social Good Fellowship located at the University of Chicago.

2 ANALYZING EQUITY THROUGH MAPPING AND VISUALIZATION

From the late twentieth century onward, there has been an increased focus on examining inequities that are tied to places. Inequality is not just measured across groups of people, but also across spaces to see where resources and opportunities cluster and how the geographic distribution of opportunity aligns with the

Data Science for Social Good Conference, September 2017, Chicago, Illinois USA
2017. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

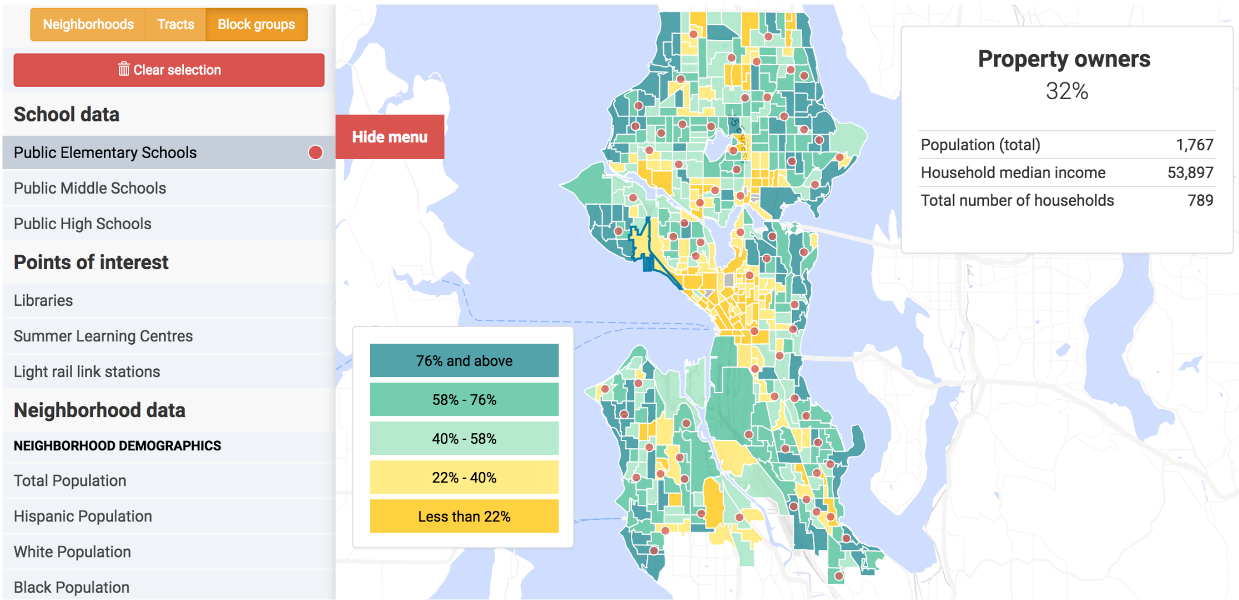

Figure 1: One view of the *Equitable Futures* visualization tool.

geographic distribution of demographic groups. Reece et al. [6] write that "Neighborhoods powerfully shape residents' access to social, political, and economic opportunities and resources." Recognizing how place and environment affect equity, mapping equity has largely been tasked to urban planners, policy makers and non-profits that are focused on social justice and equitable resource distribution. In practice, achieving greater urban equity is a goal and a process that cuts across disciplinary lines.

Equity mapping is a technique used to both visualize the spatial dimension of inequity and identifying vulnerable places, and to involve community members in the policy-making process. Talen [8] writes that "[T]he purpose of equity mapping is to stimulate further inquiry" (p. 29). Indicators must first be selected and identified based on a variety of stakeholder inputs for equity mapping. Through these actions, equity mapping can be used to identify areas of particular need, prioritize actions, and help guide the development of solutions.

The spatial mapping of resource distribution and access ranges widely from broad urban quality indicators [9], to health inequities [5], to parks and park funding [10]. The disparity and limitation of equity mapping however, is that it is typically done at the city scale while equity analysis is often done at the neighborhood scale. Many sources recognize the importance of bringing the analytical lens to the neighborhood level. California's Prevention Institute published The Built Environment and Health: 11 Profiles of Neighborhood Transformation [1], while Reece et al. [6] state that "This community development approach is based on the premise that everyone should have fair access to the critical opportunity structures and the necessary social infrastructure to succeed in life; and that affirmatively connecting people to opportunity creates positive, transformative change in communities."

In addition to scale, it is important to identify how equity is being defined and addressed. Talen (1998) describes four ways of thinking about and addressing equity—(1) Urban planners and policymakers may strive to allocate resources equally, for instance on the basis of population figures; (2) Unequally, in an effort to redress existing inequalities; (3) Unequally, in response to expressions of demand or political participation (rewarding "squeaky wheels"); or (4) Unequally, in response to individuals' "willingness to pay" (WTP) as measured by surveys, behavior, or actual markets [8]. Selecting the approach to understanding equity will shape the type of visualization that results. For example, if we try to approach equity from an economic standpoint, visualization may involve mapping parameters that will influence housing investment based on predicted land and market value.

3 EQUITABLE FUTURES TOOL FOR SEATTLE

Our initial output is a web-based map that visualizes indicators related to gentrification and inequity in Seattle (Figure 1). Our tool displays publicly available data (e.g. from the Seattle Equity Analysis, Census Bureau, etc.) primarily related to housing and development, income, mobility, and education. We anticipate the project to be a planning and educational tool that will help a variety of stakeholders in their research and at arriving at actionable information.

Based on our literature review and analysis of currently available decision support tools, visualization of indicators of inequality have thus far been displayed on static maps and/or a singular fixed granularity. Our platform will allow for more detailed analysis of equity parameters by allowing interaction with data at multiple granularities; users can choose to view the entire city at either the neighborhood, census tract, or block group level, or they may zoom into a specific neighborhood and view even finer details such as bus stops, transit lines, and buildings. This functionality allows users to identify areas of concern, and zoom in to understand the specific features of an area. The visualization aspect of our tool can be used for both exploratory and explanatory functions.

We allow users to view individual indicators, the intermediate factor or theme (such as mobility), and the overall equity index with relative contributions. In the future, we hope to allow users the functionality to choose a subset of factors that they wish to include (essentially defining their own equity index) which will update the equity map in real time. This feature is particularly pertinent as different groups arrive at different definitions of equity (see preceding section for discussion on the four categories of addressing equity).

Additional features currently include the ability to display points of interest such as schools, libraries and light rail stations. Users can click on certain points of interest for additional information; for example clicking on a school provides a pop-up containing enrollment, attendance, and student proficiency in math and reading among other things. We display additional pop-up information concerning the methodology behind calculating some indicators.

4 VISUALIZING STATISTICAL INFORMATION

While many traditional data science techniques can be deployed in our analysis, there are statistical techniques from spatial analysis, geography, and epidemiology that are tailored for spatial and spatiotemporal data. The proliferation of sensors, positioning technologies, and government open data pose new challenges to handle geographic data more complex in volume, variety, and depth.

We are using structural equation modeling to calculate factors of various themes based on indicators we selected. In the tool, we support visualization of these indicators as well as factors using a map, with different granularities such as block group, census tract, and neighborhood. One challenge we are facing in doing so, however, is visualizing the inner relationships among indicators and factors, which cannot be depicted by simply mapping them individually. Specifically, we are facing challenges from three perspectives: (1) representing relationships between indicators; (2) representing model structure; and (3) combining geographic and statistical visualization.

4.1 Representing relationships between indicators

Venn diagrams have been used to visualize multivariate regression [4] and can be used to measure the relationship between indicators and corresponding factors in our using of structural equation model. Venn diagrams can also help in visualizing the total sum of squares, sums of squares of individual variables, and correlation between variables. Variables and correlations are drawn to scale, so that it is easy to see power of indicators from the diagram.

4.2 Representing model structure

Edge bundling has been widely used to visualize correlations, especially when there is a large data set. In our case, in addition to a large set of indicators, we also want to visualize the hierarchy of factors, drawing from latent variables by using structural equation model. Holten proposed a hierarchical edge bundling method [3] for this purpose. By using hierarchical edge bundles, we can visualize the causation relationship among indicators, as well as the hierarchy of factors drawn from principal factor analysis and structural equation model.

4.3 Combining geographic and statistical visualization

The methods just described can help us visualize our model from a data perspective, but we lose the geographical characteristics of the data from doing so. Thus, another challenge we are facing is incorporating geographical characteristics with the model visualization. There are several methods we are currently investigating. Goodwin et al. proposed a diagram to visualize multiple variables across scale and geography [2]. By using a spatial scale mosaic, correlation coefficients can be shown in different scales, which in our case applies to the different granularities we are working with. Further, an asymmetrical matrix is created to visualize geographical distribution of variables, and correlations through geography.

5 NEXT STEPS

We are in the process of building the tool and the model (Figure 2); part of our future plans are to tailor the interface to different user groups. Different users will need different types of information from our tool. For general public usage, for example, we might want to show them equity factors for their neighborhoods, as well as compared with other neighborhoods throughout the city. For policy makers, it might be helpful to include the effects of certain indicators and policies, so that they can tweak these parameters to see the corresponding changes in equity factors. For researchers, on the other hand, it is important to include detailed information about our methodology of calculating equity factors from indicators. Ultimately our goal is to support effective problem solving, in part, by building successful collaborations between the University of Washington, the City of Seattle, and other interested stakeholders.

6 CONCLUSIONS

The UW eScience Institute's DSSG program is an extraordinary laboratory and opportunity for domain scientists and data scientists from different disciplines to work on an interdisciplinary subject—like urban equity and gentrification—that no one discipline by itself is able to fully comprehend and address. The format enabled our multidisciplinary team to approach a subject with a fresh perspective and apply powerful tools of data science to societal problems that other disciplines have not been able to solve.

The *Equitable Futures* project will enable policy makers and researchers to contribute to the problem in a much more transparent, accessible way. We have taken great care to ensure that our decisions for the overall project align with our vision of end-to-end transparency and accessibility for the underlying data, model, and source code. With these things in place, we enable policymakers to understand their decision processes and allow the greater community to comprehend and participate in urban policy plans.

The careful evaluation of relationships between factors and indicators in our tool with the help of a structural equation model is an innovation in the urban equity and gentrification space and anticipated to be a major win for interpretable urban decision making.

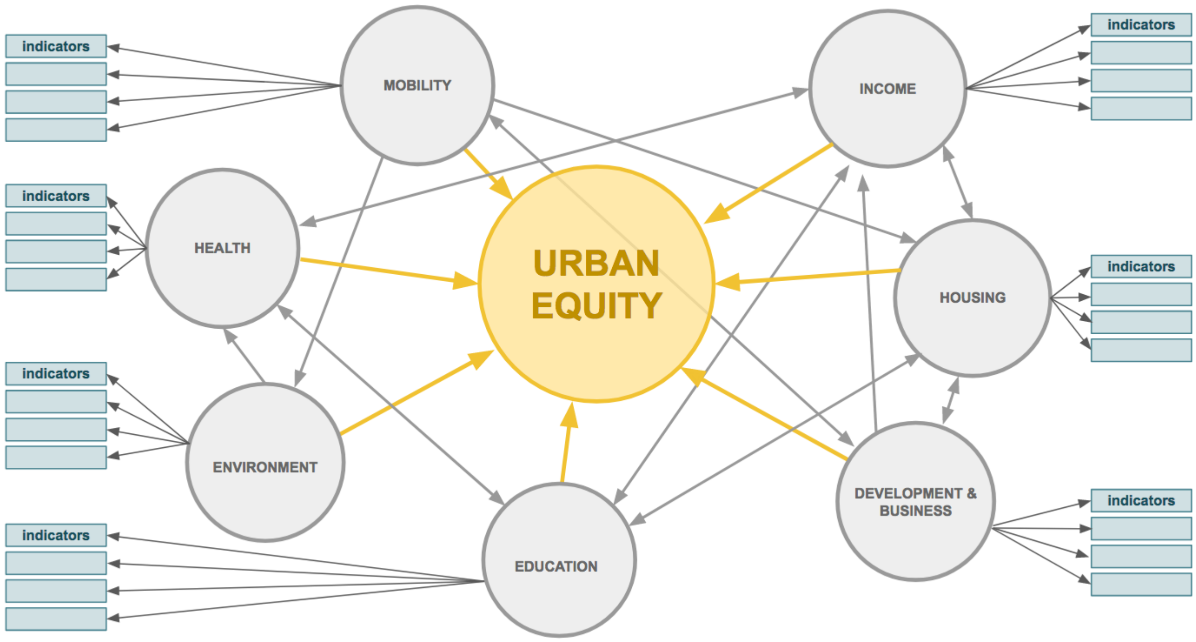

Figure 2: A visual representation of a proposed structural equation model. Multiple measurable indicators that we gather from US Census and Seattle's Open Data Portal (represented by *boxes*) imperfectly reflect immeasurable concepts that contribute to inequity (represented by *circles*).

Earlier reports and toolkits addressing urban equity and gentrification have only loosely defined and evaluated the relationship between these factors.

ACKNOWLEDGMENTS

The authors would like to thank the University of Washington eScience Institute and the University of Washington Runstad Center for Real Estate Studies. Many thanks to the DSSG student interns for their participation and hard work, and to the various mentors, speakers, and facilitators who graciously volunteered their time to contribute to the UW DSSG 2017 Program. This project was funded by generous grants to the eScience Institute from the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation.

REFERENCES

[1] Manal J. Aboelata, Leslie Mikkelsen, Larry Cohen, Sabrina Fernandes, Michele Silver, and Lisa Fujie Parks. 2004. The Built Environment and Health: 11 Profiles of Neighborhood Transformation. (2004). https://www.preventioninstitute.org/profiles/built-environment-and-health-11-profiles-neighborhood-transformation-0

[2] Sarah Goodwin, Jason Dykes, Aidan Slingsby, and Cagatay Turkay. 2016. Visualizing multiple variables across scale and geography. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology / Information Visualization / Scientific Visualization 2015)* 22, 1 (2016).

[3] Danny Holten. 2006. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* (2006).

[4] Edward Ip. 2001. Visualizing multiple regression. *Journal of Statistics Education* (2001).

[5] Benoit Lalloue, Jean-Marie Monnez, Cindy Padilla, Wahida Kihal, Nolwenn Le Meur, Denis Zmirou-Navier, and Severine Deguen. 2013. A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. *International Journal for Equity in Health* (2013).

[6] Jason Reece, David Norris, Jillian Olinger, Kip Holley, and Matt Martin. 2013. Place Matters: Using Mapping to Plan for Opportunity, Equity, and Sustainability. (2013). http://kirwaninstitute.osu.edu/my-product/opportunity-mapping-issue-brief

[7] Ariel Rokem, Anissa Tanweer, Anthony Arendt, Bernease Herman, Bill Howe, Brittany Fiore-Gartland, Bryna Hazelton, Cecilia Aragon, Ed Lazowska, Jacob Vanderplas, Joseph Hellerstein, Micaela Parker, Sarah Stone, and Valentina Staneva. [n. d.]. Building an Urban Data Science Summer Program at the University of Washington eScience Institute. *The Bloomberg Data Science 4 Good Exchange* ([n. d.]).

[8] Emily Talen. 1998. Visualizing fairness: Equity maps for planners. *Journal of the American Planning Association* (1998).

[9] UN-HABITAT. [n. d.]. UN-Habitat and the Millennium Development Goals. ([n. d.]). http://mirror.unhabitat.org/categories.asp?catid=312

[10] Jennifer Wolch, John P. Wilson, and Jed Fehrenbach. 2005. Parks and park funding in Los Angeles: An equity-mapping analysis. *Urban Geography* (2005).