1 ABSTRACT

College Board (CB) "Geomarkets" carve metropolitan areas into smaller geographies. Geomarkets are a "search filter" in the CB "student list" product that enables colleges to acquire prospect contact information. If Geomarket borders are correlated with race and class, using Geomarkets to target prospects may amplify inequality. We address two questions: What is the socioeconomic and racial variation between Geomarkets? How does the socioeconomic and racial composition of included versus excluded prospects vary if student list purchases filter on Geomarkets? We answer RQ1 using Census data and RQ2 using data from actual student list purchases.

2 PRECIS

BACKGROUND. In 1983, Zemsky and Oedel authored *The Structure of College Choice*. Based on an analysis of 1980 SAT score-sending behavior and published by *The College Board* (CB), Zemsky & Oedel (1983) developed the Market Segment Model and created "Geomarkets." The Market Segment Model predicts how student demand for a particular college varies across local geographic markets based on the characteristics of students in those local markets.

Geomarkets are geographic areas that divide states and large metropolitan areas into "community-based enrollment markets" (Zemsky & Oedel, 1983, p. 14) for recruiting purposes. Figure 1 shows New York City area Geomarkets. Geomarkets were incorporated into the CB "student list" product named Student Search Service. Student lists have been the primary source of lead generation in U.S. higher education for over 50 years (Belkin, 2019; Jaquette et al., 2022). Lists contain contact information for prospective students. The Student Search Service database consists of CB test-takers. Colleges control which prospect profiles they purchase by selecting search filters, such as high school graduation year, SAT score, AP score, state, etc. Geomarket filters enable colleges to target prospects living in particular Geomarkets.

THEORY & RQs. We analyze the Market Segment Model and Geomarkets using scholarship on quantification (e.g., Espeland & Sauder, 2016), particularly the discussions of correlation and homophily from Chun (2021). Predictive analytics are based on correlation. The 1980 SAT score-sending data analyzed by Zemsky & Oedel (1983) can be conceived as "training data," which was used to define four "market segments" of students: "local" students, who send most SAT scores to colleges in their Geomarket; "in-state" students, who primarily send scores to colleges in their state; "regional" students, who send most scores to colleges outside their state but in their region (e.g., New England); and "national" students. Zemsky & Oedel (1983) analyzed the student characteristics correlated with student market segment and concluded that student demand was primarily a function of class. The overarching thesis of the Market Segment Model is homophily, the idea that similar colleges compete with one another for similar students. Zemsky & Oedel (1983, pp. 42–45) states that selective colleges primarily draw from students in the "regional" and "national" segments and, therefore, their student list purchases should focus on Geomarkets with large numbers of affluent, highly educated households.

When we use data on past correlations to make recommendations about the future, we amplify the effects of historic structural inequality (Burrell & Fourcade, 2021). A snapshot of student demand in 1980, itself a consequence of historical structural inequality, became inscribed into the Student Search Service product. The result is a supply-side that reinforces structural inequalities observed on the demand-side. We address two research questions:

1. What is the socioeconomic and racial variation between Geomarkets in metropolitan areas? How does this variation change over time?

2. How does the socioeconomic and racial composition of included versus excluded prospects vary when student list purchases filter on particular Geomarkets?

SIGNIFICANCE. Academic scholarship on enrollment management tends to focus on students (micro), policies (macro), and organizations (meso). Scholars have not analyzed third-party classification systems and products that colleges utilize to classify prospective students. CB Geomarkets, utilized in the CB Student Search Service, offers one such case.

DATA SOURCES. To answer RQ1, we utilize census tract-level data from the 1980 Decennial Census, 2000 Decennial Census, and 2020 5-year American Community Survey (ACS). Variables of interest include: race/ethnicity; mean and median household income; percent of households below poverty line; and BA attainment of people aged 25+. We obtained shapefiles for CB Geomarkets from a 2012 *R-bloggers* post. We assigned census tracts to Geomarkets.

To answer RQ2, we issued public records requests about student lists purchased by 14 public universities from 2016 through 2020. For 414 CB student list purchases, associated with about 2.6 million prospects, we received both (1) the order summary – showing which search filters were utilized – and (2) the de-identified prospect-level data, including race, ethnicity, zip code, and high school. Figure 2 shows the partial order summary for a purchase that filtered on the 2020 high school graduating class, PSAT scores of 1070 to 1180, and selected Geomarkets.

METHOD. The methods are simultaneously descriptive and spatial. We utilize a case study design in which metropolitan areas are cases. Like Salazar (2022), we envision focusing on 2-4 metropolitan areas, including the historical context of segregation and gentrification. An appendix will show results for a larger number of metropolitan areas.

To answer RQ1, we produce tables/graphs that show how Geomarkets around a metropolitan area vary on racial and socioeconomic characteristics and how they vary over time. We produce interactive maps at the census tract-level to show more granular variation within and between Geomarkets. We address RQ2 by analyzing actual student list purchases that encompass all Geomarkets around a given metropolitan area, showing which prospects would have been excluded if had the purchase had filtered on particular Geomarkets.

PRELIMINARY FINDINGS. For RQ1, Table 1 shows racial and socioeconomic characteristics for Geomarkets around Chicago. Figure 3 is a screenshot of an interactive choropleth map of Los Angeles, showing Geomarket borders and the tract-level %Hispanic in 2020. Figure 4 shows tract-level median household income around Los Angeles.

Results for RQ2 show how many purchased prospects from each racial/ethnic group live in each Geomarket. These prospects would have been excluded had the purchase filtered on Geomarkets and not selected that Geomarket. Table 2, Table 3, and Table 4, respectively, show counts, column percentages, and row percentages for the Philadelphia metropolitan area, based on a purchase that filtered on students from the 2020 high school class who scored between 1070 and 1180 on the PSAT. Table 3 shows that "PA5–Philadelphia County" contains 15% of all purchased prospect profiles but 44% of Black prospects, indicating that opportunity for Black students is sensitive to the inclusion of this Geomarket. Figure 5 is a screenshot of a (very) preliminary interactive map for this purchase, with marker size indicating the number of purchased prospect profiles who attended that high school and separate markers by race/ethnicity.

PRESENTATION PLAN. I will provide a succinct background about CB Geomarkets and the Student Search Service product, but I will omit discussion of theory. I will show actual student list purchases that target the Chicago-land area but exclude the City of Chicago Geomarket. Next, I will show which prospects are excluded by this decision. Finally, I will engage participants in a discussion about contemporary third-party products that classify students on behalf of colleges. How can practitioners and policymakers effectively interrogate these products to avoid unintentional exclusion along racial and class dimensions?

3 References

Belkin, D. (2019). For sale: SAT-Takers' names. Colleges buy student data and boost exclusivity. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/for-sale-sat-takers-names-colleges-buy-student-data-and-boost-exclusivity-11572976621

Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, *47*, 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

Chun, W. H. K. (2021). *Discriminating data: Correlation, neighborhoods, and the new politics of recognition* (pp. xi, 327 pages). Cambridge, Massachusetts: The MIT Press.

Espeland, W. N., & Sauder, M. L. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability* (pp. xii, 281 pages). New York, New York: Russell Sage Foundation.

Jaquette, O., Salazar, K. G., & Martin, P. (2022). *The student list business: Primer and market dynamics*. The Institute for College Access and Success. Retrieved from https://ticas.org/wp-content/uploads/2022/09/The-Student-List-Business_-Primer-and-Market-Dynamics.pdf

Salazar, K. G. (2022). Recruitment redlining by public research universities in the los angeles and dallas metropolitan areas. *The Journal of Higher Education*, *93*, 585–621. https://doi.org/10.1080/00221546.2021.2004811

Zemsky, R., & Oedel, P. (1983). *The structure of college choice*. New York: College Entrance Examination Board.