Supplemental appendices for "Horizontal and Vertical Racial Segregation in Higher Education: Examining Trends in California Public Colleges" by Rachel B. Baker, Sabrina M. Solanki, and Connie Kang in The Journal of Higher Education

ONLINE SUPPLEMENTARY MATERIALS

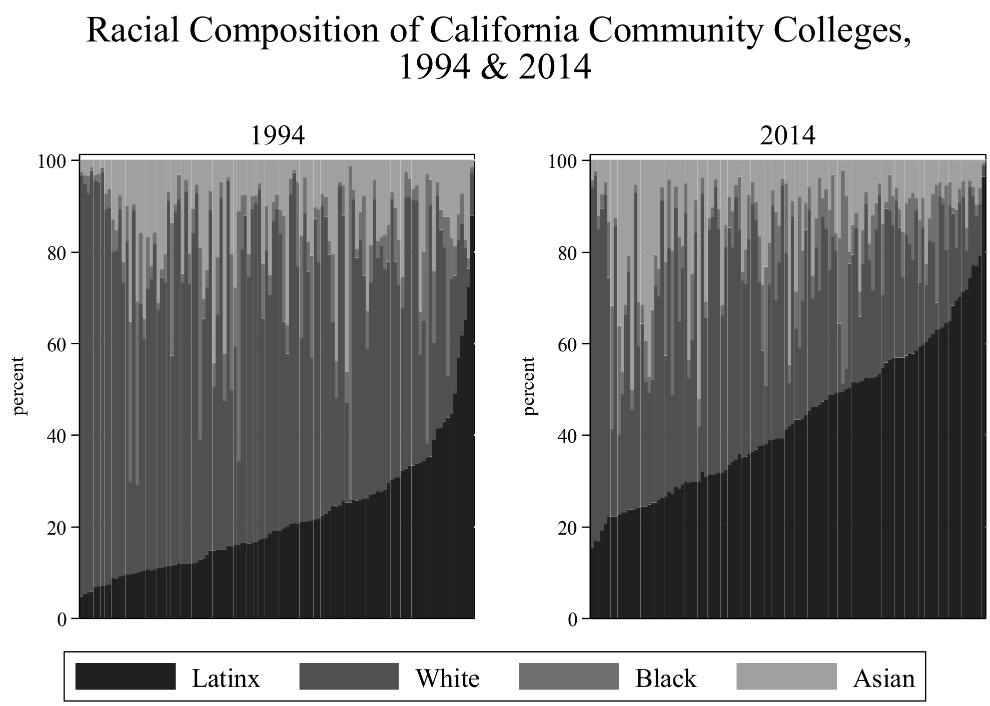
Appendix A – Descriptions of Student Sorting Across Schools

In this appendix we visually examine if demographic changes between 1994 and 2014 are evenly distributed across schools within each sector and how the race of students' peers differs across races. We examine these questions for our two bookend years: 1994 and 2014.

Figure A1 (CCC), Figure A2 (CSU), and Figure A3 (UC) provide the racial composition for each school within each sector. In these figures, each column represents a single school. The columns are ordered by the percent of students in each school who identify as Latinx, as Latinx students are the fastest growing subgroup in the state. These figures thus provide a simple visual description of the distribution of Latinx students across schools, though the graphs could be re-oriented to focus on any other race group.

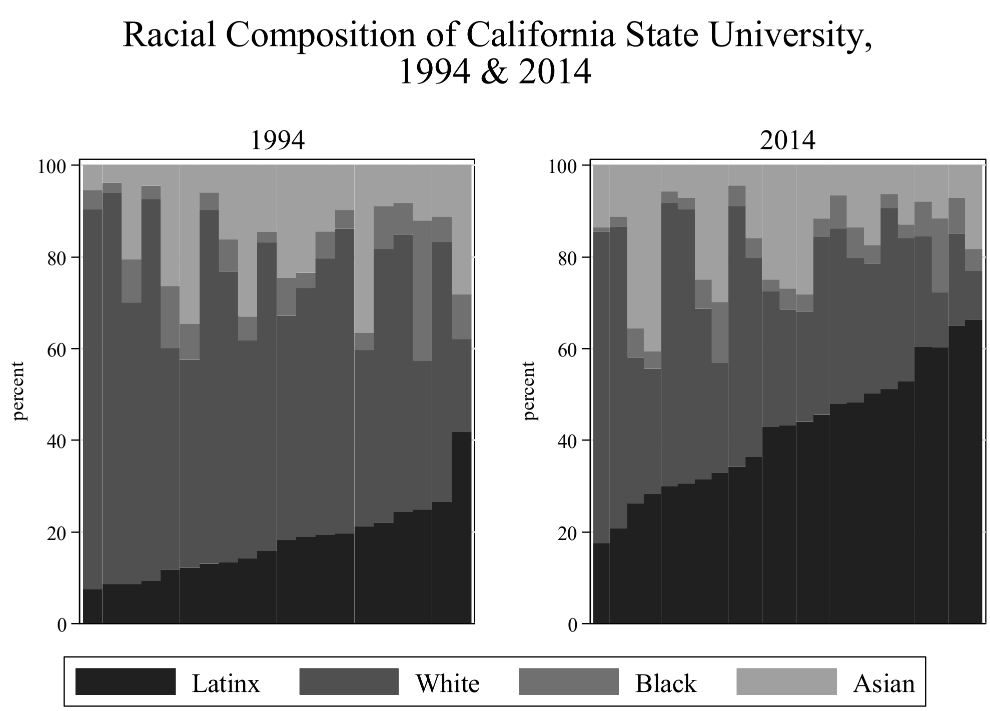
In this description, we focus on the community college sector (Figure A1), though all three figures can be interpreted similarly. The increase in the proportion of students who identify as Latinx is clear in this graph; the area of the graph that is blue is larger in 2014 than it was in 1994. This graph also allows us to examine how Latinx students are distributed across schools. If each school had the same proportion of students of any given race (for example, if 50% of students attending CCCs identified as Latinx, and each school within the CCC system also had 50% of students who identified as Latinx), each graph would consist of horizontal bands of color. This is not the case. The steep slope of the edge of the blue region shows that Latinx students are not evenly represented across schools. Comparing these slopes between 1994 and 2014 also allows us to examine how this sorting has changed over time (we examine this question using more sophisticated techniques in the main text).

Figure A1



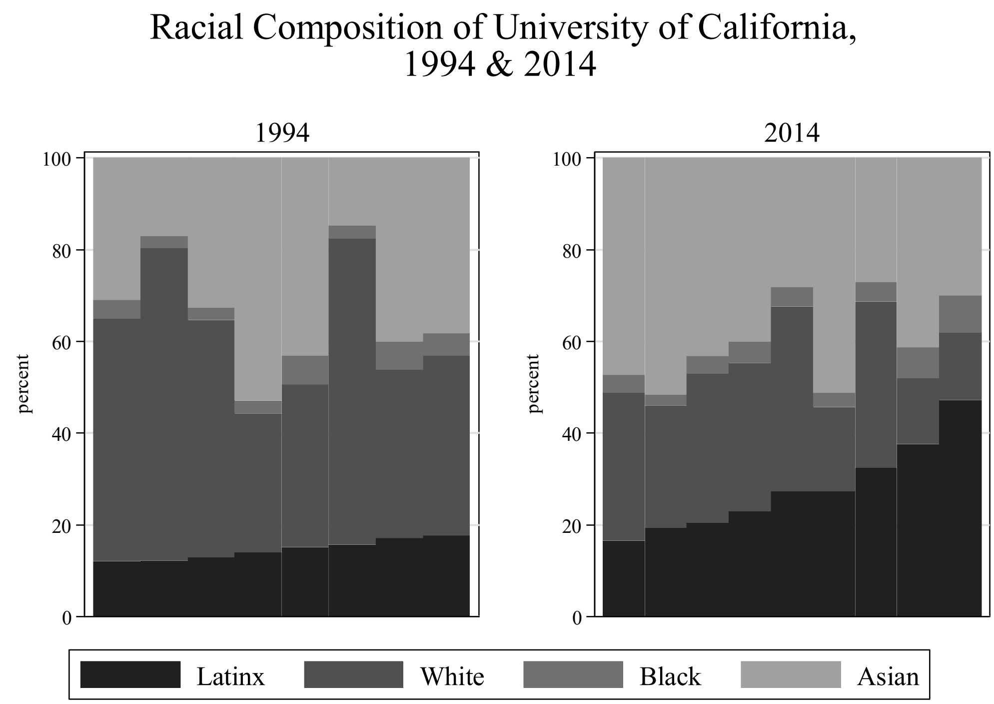
Each column represents one school.

Figure A2



Each column represents one school.

Figure A3



Each column represents one school.

As the demographic composition of the state has changed, so too have these patterns of racial sorting. Across all racial groups and all sectors, students are attending schools that are more heavily Latinx and less heavily White.

The distribution of students by race within sectors imply that students of different races are unlikely to experience similar school compositions; the average Latinx student is likely to attend a school with far more Latinx peers than the school of the average White student, for example. We examine this question in Appendix A graphs 4-6. Across all sectors, Latinx and Black students go to schools with the highest proportions of Latinx students. White students attend schools with the highest proportion of White students.

Figure A4

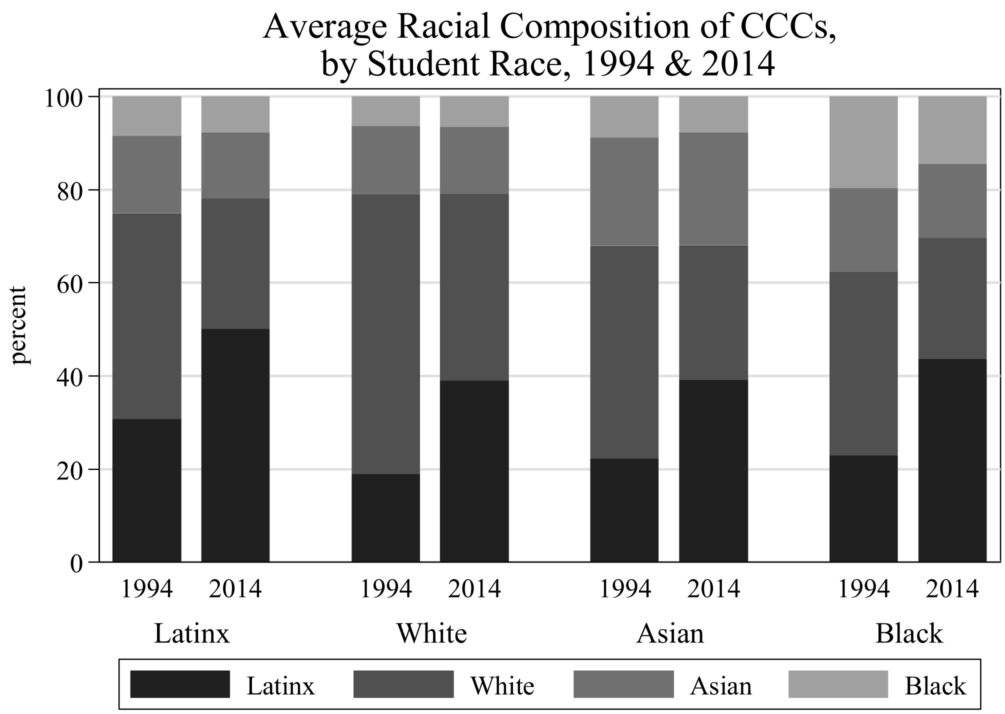


Figure A5

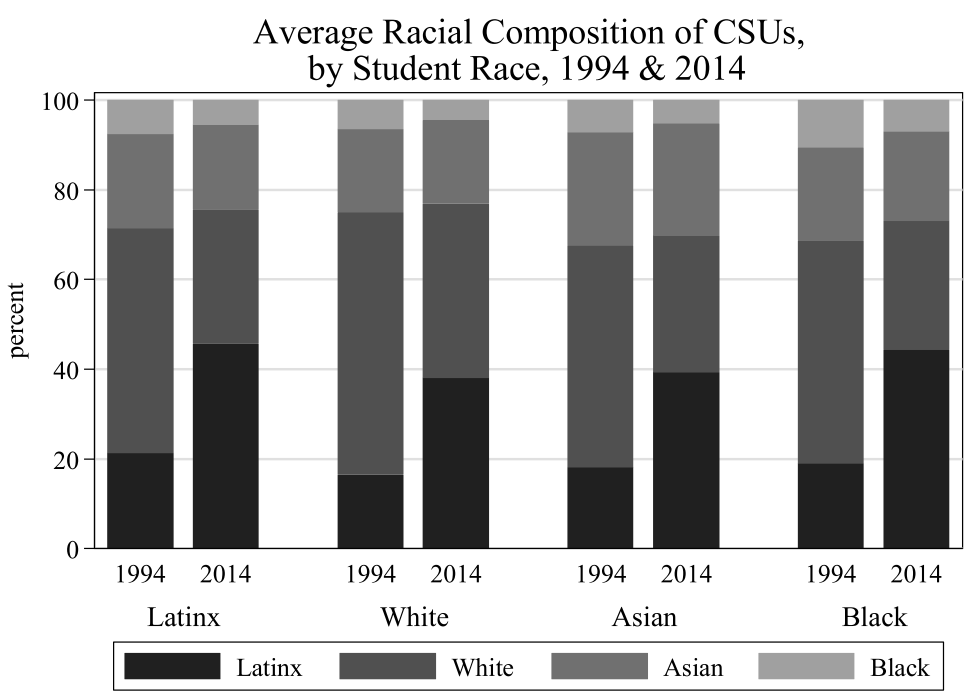
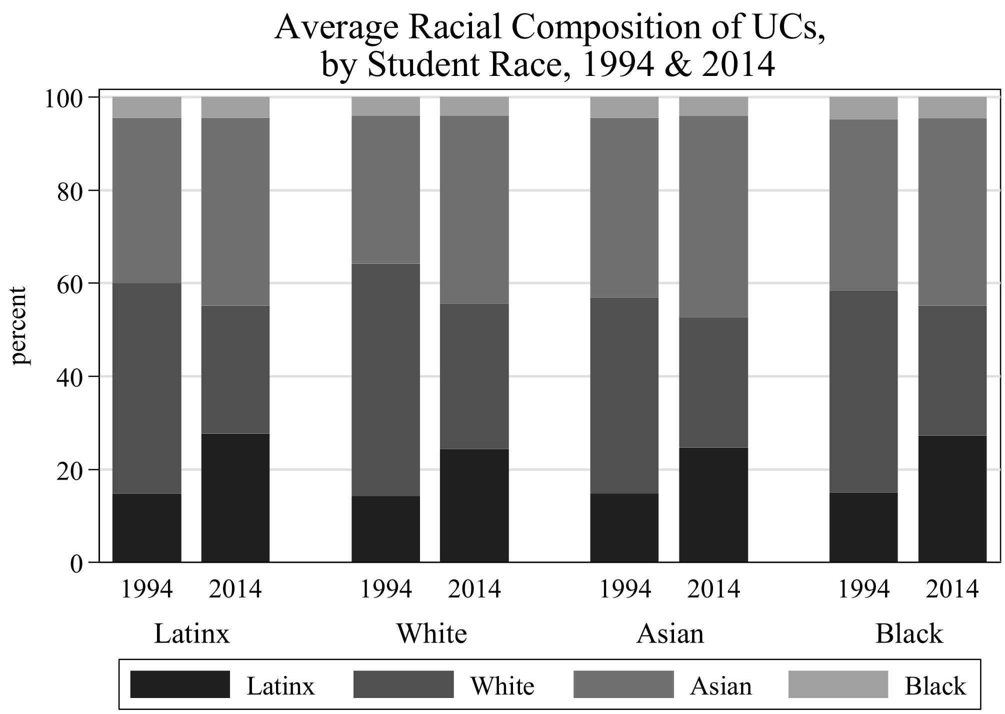


Figure A6



These simple descriptive figures provide evidence that students of different racial backgrounds are not evenly distributed across schools; the racial/ethnic composition of a student's peers is related to the student's own race/ethnicity. Together, the figures in this appendix provide motivation for the more sophisticated analyses in the main text.

Appendix B – Details on Data Sources

For this study, we obtained enrollment data for the 9 undergraduate-serving UC campuses from The Integrated Postsecondary Education Data System (IPEDS) from the National Center for Education Statistics (Data retrieved from https://nces.ed.gov/ipeds/datacenter/SelectVariables.aspx?stepId=1). We rely on IPEDS rather than collecting data directly from the UC system for two reasons: the University of California Office of the President (UCOP) does not publicly release enrollment data prior to 1998, and UCOP's data report only 12-month, not fall, enrollment. We chose not to use IPEDS data for all three sectors of higher education because some key groups of students are not included in IPEDS enrollment counts. Students who are enrolled "exclusively in courses that cannot be applied toward a formal award," students enrolled only in ESL programs, students enrolled exclusively in continuing education, and students exclusively auditing classes are not included in IPEDS enrollment counts. These groups of students make up a substantial proportion of CCC, and a non-negligible proportion of CSU, enrollments. For example, in Fall, 2014 over 10% students enrolled in CCCs were enrolled in non-credit bearing courses. As our study examines sorting across schools to examine potential contact between students, we felt it was important to include all students.

We collected data on California traditional public high schools from the California Department of Education. In addition to enrollment data by year by race, the CDE data include information on (1) the proportion of twelfth graders of each race in each school that successfully graduate from the HS and (2) the proportion of twelfth graders of each race in each school that are eligible to enroll in a CSU or UC. UC and CSU eligibility is determined based on the completion of seven sets of required courses with a passing grade of "C" or better. These "A-G requirements" provide guidance on the content and length of required enrollment in seven categories of study (e.g., History/Social Science, English, Laboratory Science, etc.) that students must complete for admission into either the UC or CSU system. More information about UC eligibility can be found here: https://hs-articulation.ucop.edu/guide/a-g-subject-requirements/a-history-social-science/ We use these data to examine observed levels of segregation against geographic counterfactuals.

Appendix C – Decomposing the Theil Index

In this appendix we provide an example decomposition of the Theil Index. Here we decompose the overall Theil index into the within-region component and the between-region component. If we let H_r refer to the Theil index within region *r*, E_r refer to the entropy within region *r*, B_r refer to the number of Black students attending college within region *r*, and W_r refer to the number of White students attending college within region *r*, the Theil Index can be written as:

$$H = H_R + \sum_{r=1}^{6} \left(\frac{B_r + W_r}{B + W} \right) \frac{E_r}{E} \cdot H_r$$

The first term on the right-hand side (H_R) is the between-region component, and the second term is the sum of the within-region components for each region. The decomposition analyses are weighted by the overall enrollment in the region (or the sector for our within-sector analyses).

Appendix D – Analytic Method for the Geographic Counterfactuals

To calculate the hypothetical levels of segregation under our various geographic counterfactuals, we first identified the closest CCC, CSU, and UC to each public California high school. We accomplished this by calculating both the distance and travel time from each high school to each public college in California using the latitude and longitude of each school. We computed these distances using the HERE API plug-in for Stata (–georoute–). The HERE API computes travel time between two coordinates in normal traffic conditions. For these analyses, we rely on travel time, rather than distance, as our measure of proximity to account for the fact that travel time likely influences college decisions more than absolute distance (Jepsen & Montgomery, 2007).

For our second geographic counterfactual, in which we send the students to the nearest college in the most selective sector they are qualified to attend, data availability presented a considerable limitation. We do not have individual-level information on the schools to which students would be admitted had they applied, nor do we have other information that could be used to proxy admission rates (such as student-level standardized test scores or GPAs). To overcome this barrier, we use the CDE data that details what percent of students of a given race in a given high school met the qualifications for attending either a CSU or a UC based on having a passing grade of "C" or higher on A-G required courses for these sectors. We then send ⅓ of these eligible students to the nearest UC and the other ⅔ to the nearest CSU. We send all remaining graduates to the nearest CCC. These ratios used are consistent with those listed in the master plan for the state of California higher education and roughly match actual enrollment numbers. While this is a crude measure, it does allow us to examine a slightly more reasonable counterfactual than our first counterfactual.

References for Appendix D:

Jepsen, C., & Montgomery, M. (2009). Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college. *Journal of Urban Economics, 65*(1), 64–73.