The Impact of Exploring Computer Science in Wisconsin
Where Disadvantage is an Advantage

Heather Bort
Marquette University
Milwaukee, WI, USA
heather.bort@mu.edu

Shion Guha
Marquette University
Milwaukee, WI, USA
shion.guha@mu.edu

Dennis Brylow
Marquette University
Milwaukee, WI, USA
dennis.brylow@mu.edu

ABSTRACT
Assessing the impact of regional or statewide interventions in primary and secondary school (K-12) computer science (CS) education is difficult for a variety of reasons. Qualitative survey data provide only a limited view of impacts, but quantitative data can be notoriously difficult to acquire at scale from large numbers of classrooms, schools, or local educational authorities. In this paper, we use several publicly available data sources to glean insights into public high school CS enrollments across an entire U.S. state. Course enrollments with NCES course
codes and local descriptors, school-level demographic data, and school geographic attendance boundaries can be combined to highlight where CS offerings persist and thrive, how CS enrollments change over time, and the ultimate quantitative impact of a statewide intervention. We propose a more appropriate level of data aggregation for these types of quantitative studies than has been undertaken in previous work while demonstrating the importance of a contextual aggregation process. The results of our disparate impact analysis for the first time quantify the impact of a statewide Exploring Computer Science (ECS) program rollout on economic groups across the region. Our blueprint for this analysis can serve as a template to guide and assess large-scale K-12 CS interventions wherever detailed project evaluation methods cannot scale to encompass the entire study area, especially in cases where attribute heterogeneity is a significant issue.

KEYWORDS
Exploring Computer Science, ECS, socioeconomic impact, attendance boundaries, contextual aggregation, geography of opportunity

ACM Reference Format:

1 INTRODUCTION
Much of the literature discussing computer science (CS) education points to disparity in economic and demographic situations as a root cause for the lack of access to computer science for nontraditional students. This research implies (and sometimes explicitly states) that if students were given the opportunity to learn CS, the gap in gender and minority participation would be lessened. Since economic and demographic factors are inherently spatial in nature, this paper will focus on using elements from the geography of educational opportunity framework.

The Geography of Opportunity framework is often referred to in research focused on housing and residential mobility. It is meant to refer to the ways that geography can influence an individual’s opportunity, i.e., that an individual’s options are limited by the social and economic conditions surrounding them. The idea behind this framework can be succinctly summarized with a quote that Squires and Kubrin attribute to former Albuquerque mayor David Rusk, “Bad neighborhoods defeat good programs”.

This concept can be extended to the geography of educational opportunity because there are constraints placed on educational opportunity by the educational infrastructure of the community. In Hillman’s research on the geography of opportunity as it applies to college choice, he explains how geography can affect educational opportunity by drawing a comparison to the concepts of food deserts. Hillman uses the geography of opportunity framework to explore the importance of place and how geography shapes educational equity and opportunity. Additionally, the geography of opportunity is examined in as it relates to outcomes of No Child Left Behind testing requirements and segregation in schools. Tate, et al. reminds that certain interventions in STEM education have been geospatially-minded in the past, especially in urban contexts; Green notes that “Access to opportunities in the United States (U.S.) is inequitable across geographic spaces”. Finally, advocates for using a spatial perspective to help build an understanding of the inequalities in communities so that action can be taken; spatial thinking through a geographical perspective can help to facilitate change. Where a student attends school has a direct impact on their opportunity to access CS education. Their geographic place can constrain or enable them more than other factors such as an understanding of what a computer scientist does, or seeing people in CS jobs that represent their gender, race, or ethnicity. Use of this framework can add the context of where to a discussion that is largely focused on the when and who. Instead of asking how much did CS enrollment increase from one year to the next, the question becomes where did CS
enrollment increase. This shift moves the analysis to a different level of granularity. When not concentrating on merely the net gain or loss over an entire state, the focus can be on what the changes were for each school.

The primary questions in this study are:

1) Does publicly available data give enough information to track CS course enrollment over an entire state without the need for costly and time-consuming surveys?
2) Is it reasonable to represent high school CS course availability at the state level or is a more granular representation needed?
3) Has the introduction of the ECS program had a disparate impact on any economic groups in Wisconsin?

2 GEOGRAPHY IN CS EDUCATION RESEARCH

Given the difficulty in collecting detailed data over a large area for hundreds of schools, very few prior studies have even a small focus on the underlying geography of their study area. The only recent CS-centric study explicitly mentioning geography is a report from South Carolina showing a statewide lack of geographical diversity for where CS coursework is offered. Based on survey responses from 158 K-12 educators, they concluded that Title 1 schools (where > 40% of the students qualify for free or reduced lunch) are less likely to offer computing coursework.

A second research study indirectly focused on geography by using data from the Advanced Placement (AP) CS A exam in a regression analysis to explore the demographics of test takers across the U.S. In each state the relationships were explored between wealth and exam-taking, and the number of exam-takers from under-represented groups, with the goal of explaining variances between states.

Both studies consider how economic status could relate to a student’s opportunity to take computing courses in K-12. While one uses statewide survey data for indication of CS availability with school level economic information, the other uses national AP data and state level economic factors. The current study differs from previous work in the following ways:

1) While still interested in where CS coursework is offered in the state as in, this study uses public data collected from the state Department of Public Instruction (“DPI”) instead of attempting a statewide survey. Schools are required to self-report on many dimensions annually, and course level enrollment data is one of these dimensions. While there are always concerns for the validity of self-reported data, this source provides complete data for all the schools in our study area without putting any additional workload on the schools. In addition to data being more complete and reliable than what a large-scale survey would provide, this data is publicly available in many states, making it quite attractive for this type of analysis.

2) The study in investigates the relationship between wealth and the number of students taking the AP CS A exam for U.S. states. This paper follows the lead of in examining the role that wealth can play in computing education; however, using a framework of geography of educational opportunity means considering the role of geographical place within the study, this leads the authors to disagree with on three points: (a) Analysis method - A regression analysis assumes that the data being analyzed is random. This paper considers how CS course enrollment and availability is not randomly distributed within a state and therefore violates that assumption. (b) Level of aggregation - Aggregation to the state level for median income as the wealth variable and exam takers assumes that the nonaggregate data is distributed within the state homogeneously. This paper argues that this type of data exhibits heterogeneity and therefore should be studied at a finer level of granularity. (c) Choice of explanatory variable - Using median income as a measure of wealth in (even if the study had been at a more reasonable level of aggregation) assumes that the school inherits its wealth attribute from the surrounding community. This paper shows that while it is not a perfect proxy as an indication of wealth, an individual school’s reported measure of economic disadvantage can more accurately describe the
economic circumstances of the students in attendance. Since the unit of study is the school, this measure is within the proper context and not merely a convenient choice. Moreover, in a context such as Wisconsin, in which widespread school voucher programs allow many students to attend a school in a dissimilar economic area, median income for the surrounding community is less likely to align with the economic composition of a given school’s student body.

3 THE ROLE OF PLACE

When considering why geographical place would have a role in the analysis of CS education in public high schools, one should keep in mind that the institution of interest in this work is the school. In the U.S., many aspects of educational policy are determined not at the national level, but at the state or school level. Schools play a major, central role in everyday social geographies in general. Collins and Coleman bring attention to the fact that “they are one of the few institutions that can be found in almost every urban and suburban neighborhood, and with which almost every individual has meaningful, sustained contact at one or more points in their lives.” This paper focuses on Wisconsin public schools with a high grade of 12. The study is limited to regular schools, and does not include data related to charter, virtual, or private schools. Also discluded are data related to informal CS education, such as after school clubs or summer camps. The choice to limit the data in this way follows [2], which states that the best chance to broaden participation in computing is through formal education pathways; going through the formal education pathway is the only route that can ensure we are providing all students access to CS education.

This study uses student enrollment counts for each school at the individual course level. The number of students in a school who were enrolled in a CS course can be extracted using a combination of NCES (National Center for Education Statistics) course codes, local course codes, and course descriptions. Economic data used in the study are aggregated at the school level. This initial study uses the school reported measure of economic disadvantage as a percentage of total students enrolled who are categorized as being economically disadvantaged.

All of the data that are used in this study are publicly available. We collected 6 years’ worth of data from the Wisconsin DPI related to course enrollments, school enrollments and demographics, educator license status and employment, as well as shapefiles for school attendance boundary zones. Within the course enrollment data, we compared course NCES codes to the local course codes and the local course descriptions to identify courses that could count as CS for mathematics graduation credit, defined by Wisconsin Act 63. There are 433 schools, making up 378 districts in this data. The six years collected from each school included academic years 2010-11 through 2015-16.

Wisconsin DPI defines economic disadvantage for students who are members of households that are eligible for free or reduced-price meals under the National School Lunch Program (NSLP). To be eligible, a family’s income must be less than or equal to 185% of Federal Poverty Guidelines. Students must be identified by Direct Certification every year, and this information is reported by every school even if the school does not participate in the NSLP.

When considering American political geography in relation to K-12 CS education, it is important to note that autonomous governmental units, such as local school districts within metropolitan areas, are a defining feature. In fact, this type of geopolitical fragmentation is common in major U.S. metropolitan areas, especially in older industrialized regions of the Northeast and Midwest. Information should not be aggregated at the state level to gauge the health or impact of CS education as in, because doing so assumes a fixed effect over the study area, potentially causing false correlation or confounded analysis. How these programs are implemented is greatly influenced by the local context of the individual units, and have found that even the level of fragmentation (number of school districts) within an area can influence efforts to implement reforms. The degree of challenge involved in implementing a reform is increased in areas where there are a greater number of units.
At the national level, entities like Code.org provide information about places where progress is being made in CS education. They show where CS can be counted towards high school graduation requirements for math or science. From this we can see that not all states in the U.S. have fallen in line with the recent CS For All initiatives from the Obama administration, and not all states allow CS courses to count towards high school graduation requirements.

Code.org also shows information at the state level, where they post data about employment in CS, numbers of CS graduates and what they call their “Policy Environment (rubric)”, which point out whether a state has dedicated funding for professional development, if high schools in the state are required to offer CS courses, and the status of state K-12 CS standards.

Figure 1 shows that considering where CS courses are being offered at the school district level gives a better understanding of where our state suffers from a lack of opportunity. We can then drill down even further to look at the attendance boundary zones for individual schools in a district, as in Figure 2, to see that even if the district level data shows that CS is present, the school level data can identify gaps in access to student groups.

While it would be ideal to disaggregate this data all together and compile a dataset of family income for each individual student in addition to their course enrollment information, this is not a realistic goal for a statewide study like this one. Therefore, we have decided to aggregate our data at the school level, which is the most granular level reasonably available. Within the context of our study, school level aggregation is consistent within the framework, and provides minimal information loss.

Figure 1: Districts where CS courses are offered in WI, ’14-15

4 METHODOLOGY AND ANALYSIS
This section describes the mapping of geographic opportunity in Wisconsin, the method used to identify randomness in our data, and the resultant need for finer-grained analysis than state aggregate.

Hogrebe and Tate discuss how building a visual representation in line with Soja relates to the geography of opportunity. They focus on cognitive science literature stating that graphic representations support learning, visual modeling supports retention, and visual representation can help readers more easily grasp complex arrangements. They also note that maps have informed civic debates about regional development issues. Maps bring a visual dimension to data through geospatial perspective and facilitate discussion among stakeholders. “By locating data in the context of place, issues become more familiar and understandable to those who may not have experience in data analysis.” We can associate variables that are not spatial with a location so that they can be placed in the context of where they occur. In a discussion about why we should show data in map form, refer to John Pickles and how he rethinks mapping, stating that maps are active and that they actively construct knowledge, “They exercise power and they can be a powerful means of promoting social change.”
The visual representation of where CS courses are offered in the state of Wisconsin, such as Figures 1 and 2, can help to identify areas of the state where intervention efforts are most needed. It also gives the opportunity to make the case against using a state level aggregation of measures for CS enrollment (or test takers as in10). It can be tested whether the data is randomly distributed in relation to space; since with a random distribution there is no value in knowing the spatial locations of the units, it will not provide leverage in predicting opportunity. We have used a join count analysis7 as a diagnostic tool to test for a random distribution. If we assume the data are binary, presence or absence of CS in the school, we might observe a random distribution of the categories, a clustered distribution of the categories, or an evenly dispersed distribution of the categories. We may wonder what the chances are that a particular pattern of distribution in our data could occur at random. Of our 378 school districts, we can build a spatial weights matrix to describe the neighbor relationships between the schools by defining neighbor with first-order queen contiguity as a base case. This means that we examine each of the polygons defining a school’s geographic attendance boundary and identify where a polygon shares an edge with any other polygon, resulting in a contiguity matrix W. In the contiguity matrix W, each school polygon at i will be considered against another school polygon at j; each ij space will be assigned a 1 if the polygons share an edge, and a 0 otherwise. The diagonal of the matrix is 0 since a school cannot be defined as its own neighbor, and W is symmetric as we are defining the neighbor relationship to be bidirectional. Other potential definitions for W will be considered at a later time in order to build a reasonable and theoretically based weights matrix.
Our queen contiguity matrix shows 2112 distinct neighbor relationships, with each school having a number of neighbors from 1 to 16. We can count the number of relationships that fall into each of three categories: (1) Neither school has CS (NN); (2) Both schools have CS (YY); and (3) One of the schools has CS and the other does not (YN). We then ran a random permutation on the attribute matrix 1000 times to produce a distribution of random placements of the dependent variable so that we could compare the true counts to the expected counts under a random distribution.

The join counts analysis results for the 2014-15 school year are shown in Figure 3, and clearly indicate that the distribution of course offerings is not random. Since this is the case, a regression to show the relationship between wealth and CS course enrollment is not appropriate for our data. It is important to note that patterns can be perceived even when they are not present. The benefit of using a visual representation along with a test for randomness is that we can be assured that our perception of clustering in the offering of CS courses is not merely apophenia.

5 ECS IN WISCONSIN

Wisconsin was the site of one of several NSF-funded initiatives that aimed to prepare annual cohorts of high school teachers for teaching the Exploring Computer Science (ECS) curriculum in the years 2014-2017. While the program collected its own data from participating students and their teachers, to date there has been no attempt to assess the statewide impact of large scale deployment of ECS. The first cohort of ECS teachers were deployed to Wisconsin classrooms in the 2014-15 school year. In keeping with our interest in the findings of [10], where wealth is shown to have an indirect impact on the number of students taking the AP CS A exam, we looked to see if there was a relationship between school level economic disadvantage and CS course offerings. Instead of a linear relationship, we found that the implementation of ECS in a school disproportionately impacted schools with a higher percentage of economically disadvantaged students.

Table 1: Enrollment % in CS and ECS courses aggregated by school % of economic disadvantage, 2010-2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-25%</td>
<td>135</td>
<td>125</td>
<td>118</td>
<td>117</td>
<td>113</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Total Enrolled</td>
<td>96,785</td>
<td>92,150</td>
<td>84,048</td>
<td>79,681</td>
<td>79,650</td>
<td>83,501</td>
<td></td>
</tr>
<tr>
<td>CS Enrolled (pct)</td>
<td>2045</td>
<td>2150</td>
<td>2255</td>
<td>2427</td>
<td>3213</td>
<td>3941</td>
<td></td>
</tr>
<tr>
<td>ECS Enrolled (pct)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>ECS pct of CS</td>
<td>0</td>
<td>0</td>
<td>2.349%</td>
<td>5.353%</td>
<td>7.363%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-50%</td>
<td>225</td>
<td>234</td>
<td>236</td>
<td>239</td>
<td>245</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>Total Enrolled</td>
<td>128,260</td>
<td>125,971</td>
<td>130,302</td>
<td>131,865</td>
<td>130,245</td>
<td>127,852</td>
<td></td>
</tr>
<tr>
<td>CS Enrolled (pct)</td>
<td>1811</td>
<td>1931</td>
<td>1993</td>
<td>2498</td>
<td>2691</td>
<td>3142</td>
<td></td>
</tr>
<tr>
<td>ECS Enrolled (pct)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>ECS pct of CS</td>
<td>0</td>
<td>0</td>
<td>0.320%</td>
<td>3.270%</td>
<td>6.111%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-75%</td>
<td>65</td>
<td>62</td>
<td>66</td>
<td>66</td>
<td>59</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Total Enrolled</td>
<td>34,404</td>
<td>31,836</td>
<td>32,192</td>
<td>33,003</td>
<td>32,164</td>
<td>30,955</td>
<td></td>
</tr>
<tr>
<td>CS Enrolled (pct)</td>
<td>477</td>
<td>311</td>
<td>316</td>
<td>436</td>
<td>277</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>ECS Enrolled (pct)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ECS pct of CS</td>
<td>0</td>
<td>0</td>
<td>0.006%</td>
<td>88.068%</td>
<td>192.015%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76-100%</td>
<td>10</td>
<td>14</td>
<td>15</td>
<td>13</td>
<td>18</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total Enrolled</td>
<td>7,335</td>
<td>12,129</td>
<td>11,608</td>
<td>10,507</td>
<td>12,691</td>
<td>10,177</td>
<td></td>
</tr>
<tr>
<td>CS Enrolled (pct)</td>
<td>57</td>
<td>140</td>
<td>54</td>
<td>49</td>
<td>408</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>ECS Enrolled (pct)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>255</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>ECS pct of CS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>62.5%</td>
<td>18.443%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows enrollment percentage in CS and ECS courses for the 4 years before the first year of ECS in Wisconsin, and how those percentages changed with the first 2 years of implementation of ECS, broken into 4 categories of economic disadvantage. It can clearly be seen that the schools with a higher population of economically disadvantaged students have had an increase in enrollment that is significantly higher than the other economic groups.
Of particular interest is the difference between the schools in the upper and lower ranges of economic disadvantage. In general, we see a drop in CS enrollment, particularly for ECS, between the 2014-15 school year and the 2015-16 school year for the 76-100% economically disadvantaged group. We also note that in the 0-25% group there is little variation in CS enrollment from one year to the next, and that ECS has had little impact on enrollment. This is where our focus on place will allow us to more clearly identify the mechanics of these general observations.

We observe that in the 76-100% group, at many schools prior to the implementation of ECS, CS courses were not available. Therefore, when ECS began to be offered, students were enrolled in the course at the 9th-12th grade levels. After the initial year, overall enrollment for the group would drop, because only the incoming 9th graders would not have had the previous opportunity to enroll.

As an example, North Division High School (76-100% economic disadvantage), a Milwaukee Public School, enrolled 84 students in ECS in 2014-15 with 27 of those being 12th graders, 27 11th, 19 10th and 11 9th. In fact, 84 students were their total enrollment in any CS courses, since they did not offer any course that was not ECS, and prior to 2014-15 did not offer any CS courses at all. In 2015-16 their enrollment in CS overall dropped to 52. However, the 20 students enrolled in ECS were from 9th and 10th grade, and the other 32 10th-12th grade students were enrolled in an Introduction to Programming course. We would expect that the 2016-17 data will show this in the 51-75% range since the 2015-16 ECS enrollment numbers are spread across the grades like the 76-100% group in 2014-15. Additionally, Whitefish Bay High School (0-25% economic disadvantage), had offered CS courses prior to the Wisconsin ECS initiative and, despite being one of the cohort schools, did not show enrollment in ECS courses. However, they did show consistent enrollment among all grades for 3 years in the CS Principals course.

6 CONCLUSION

The publicly available data that was used in this study provides ample information to track CS course enrollment in the state of Wisconsin. We are able to observe how enrollment changes within economic groups as well as within individual schools, and can make inferences about how these observations could be used to guide interventions. However, additional evidence collected through targeted interviews with individual educators would enable us to consider more fully the role that teachers play in those changes. The next step in our work will consider the status of the teachers along with enrollment data, since our initial analysis reveals that there are more districts where educators holding a license for CS are employed at the high school or district-wide level than there are districts with schools actually offering CS courses (see Figure 4).

Figure 4: CS courses offered vs. CS licensed teachers, 2014-15

Since much of the narrative surrounding the subject of how to increase presence of CS courses in K-12 education points to a lack of qualified teachers, we feel that it is worth investigating the ways in which Wisconsin seems to break from that narrative.
We can see that analysis at the individual school level gives us a much clearer view of what is going on in a state overall, and that we cannot aggregate our data to the state level. In this situation, the sum of the parts does not necessarily add up to the whole. Information like this can lead to more thoughtful policy and planning for widespread K-12 CS initiatives. We might suggest that a school interested in implementing ECS with a limited budget would need more funding available in the initial year, when more of the student population will have access for the first time to any CS course. In subsequent years, less funding would be required to sustain the course only for incoming 9th graders. Based on this analysis, we would most likely target groups with lower economic disadvantage with an AP CS initiative, rather than an ECS initiative.

The introduction of ECS in Wisconsin has had much more of an impact on schools with a higher level of economic disadvantage. We have seen that schools that are more economically disadvantaged were less likely to have offered any CS courses before the initiative, and that adding ECS led to a wider variety of CS courses for those schools after the initial intervention year. Schools which were less disadvantaged were more likely to have already been offering CS courses, and the addition of ECS did not significantly change their enrollment percentages.

6.1 Acknowledgments
This work was supported in part by US NSF grant CNS-1339392.

REFERENCES


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.