The Shrinking Slice of Women and Black Students in a Growing Computer Science Pie: A Preliminary Spatiotemporal Analysis of Longitudinal Completions Data*

Abstract

In an era where Computer Science plays a transformative role in society, the need to address Diversity, Equity, and Inclusion within the field has become increasingly imperative. Using a data-driven approach, this work shows how Women and Black students, as a percentage of total completions of CS related programs, are decreasing in recent years. This trend is particularly alarming in the Southeast region of the United States.

1 Introduction

As the field of computer science continues to innovate and grow, we must also ensure that this progress is inclusive, equitable, and reflects the diverse perspectives and talents of all individuals. Diversity, Equity, and Inclusion (DEI) in Computer Science (CS) is a topic of growing interest in the United States [2, 7, 4]. The National Science Foundation (NSF) has been collecting

data on CS completions [3]; a *completion* is defined as an "award or degree conferred." These data are available through the NSF's National Center for Science and Engineering Statistics (NCSES) and is updated annually [8].

In our review of the literature, we found a dearth of DEI studies in CS that take a data-driven approach. Most of the existing work is qualitative and prescriptive in nature [5, 6], without delving into patterns and structures that may be at the root of the problem. We strongly believe that a data-driven approach is necessary to make progress towards a more equitable and inclusive CS community. Without a quantitative understanding of the underlying dynamics, policy makers and practitioners will be unable to make informed decisions about how and where to allocate resources to address the problem.

To address this gap in Computing Education literature, in this work, we use 18 years of Integrated Post-secondary Education Data System (IPEDS) Completions [8] data to explore the question:

How has the number of CS degrees awarded to students from underrepresented groups changed over time in different regions of United States?

Our results show that the number of CS degrees awarded to students from underrepresented groups has generally increased over time. However, the share of CS degrees awarded to students from underrepresented groups is decreasing. Our results uncover various inter- and intra-group trends that are also spatially dependent. The decrease in share of degrees awarded to Women and Black students is particularly stark in the Southeast. In contrast to other underrepresented groups, the number of degrees awarded to Hispanic/Latino students shows a positive trend across different regions of the country.

Altogether, our results strongly suggest a call to action for more data-driven research on the topic of DEI in CS. Only with a better understanding of the underlying dynamics can we hope to make progress towards a more equitable and inclusive CS community.

2 Methodology

2.1 Data

The NSF has been collecting data on completions for all fields of study since 1966 and updates it annually. IPEDS Completions data is available at the institution level. For each institution, the data includes the number of degrees awarded by field and demographic group.

In this work we analyze 18 years (2002-2020) of Completions survey data from 6307 post-secondary institutions. In IPEDS, this survey data table is titled *Awards/degrees conferred by program, award level, race/ethnicity, and*

Table 1: Regions reported in IPEDS data (`variable=OBEREG`) ordered by institutional count.

	Region	States	Institutions Count
1	Southeast	AL AR FL GA KY LA MS NC SC TN VA WV	1536
2	Mid East	DE DC MD NJ NY PA	1056
3	Far West	AK CA HI NV OR WA	923
4	Great Lakes	IL IN MI OH WI	893
5	Southwest	AZ NM OK TX	670
6	Plains	IA KS MN MO NE ND SD	497
7	New England	CT ME MA NH RI VT	332
8	Rocky Mountains	CO ID MT UT WY	245

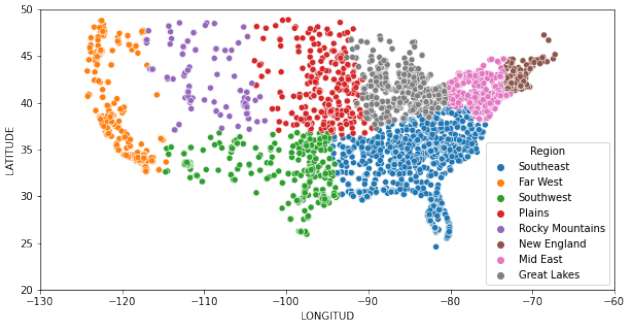
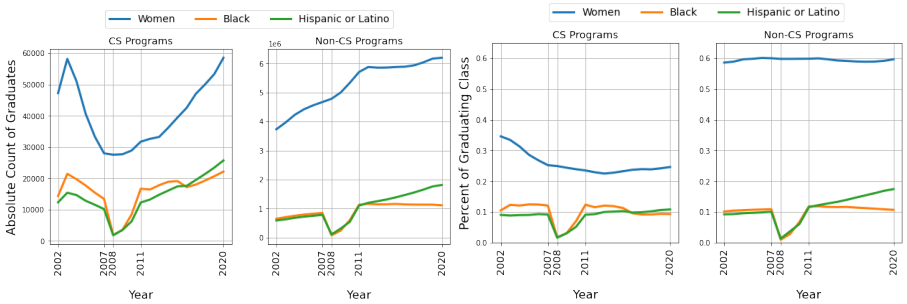


Figure 1: Spatial distribution of institutions in the data, color-coded by region (Alaska and Hawaii are not depicted).

gender. The programs are classified using 6-digit Classification of Instructional Programs (CIP) code. In our results, we refer to all degrees awarded under CIP code 11 titled *Computer and Information Sciences and Support Services* as simply "Computer Science" for brevity and clarity of presentation. Using the IPEDS *Institutional Characteristics* survey of 2020, we identify the region of each institution; categorization details are provided in Table 1 along with a count of the number of constituent regions.

2.2 Analysis

Our analysis focuses specifically on the following three underrepresented groups: Women (`variable=CTOTALW`), Black (`variable=CBKAAT`) and Hispanic or Latino (`variable=CHISPT`). All three of these variables are in the data Completions



(a) Absolute count of completions; vertical axes use different scales.

(b) Percentage of total completions; vertical axes use the same scale.

Figure 2: Nationwide temporal trends in CS and Non-CS program completions from 2002 to 2020 for Women, Black and Hispanic/Latino students.

survey table `C<YEAR>`, at the level of institutions, and are publicly available to download from the IPEDS website [8]. Counts were computed using a simple sum of the relevant variable over all institutions for a given year and region. Percentages were computed by dividing relevant sums by `CTOTALT`: Grand total of students awarded degrees with `CIPCODE` in range $[11, 12)$. Data from the IPEDS *Institutional Characteristics* survey table 2020 is used (`variable=OBEREG`) to identify each institution's region. Analysis is done using `pandas` Data Analysis library and plots are made using `matplotlib`. The code is available at `https://github.com/<REDACTED>`.

3 Results

3.1 Nationwide trends

3.1.1 Absolute count of completions

Post dot-com bubble (2002-2007): In terms of absolute counts (Figure 2a), we see a consistent decline in number of completions in CS between years 2003 and 2008. We believe this is the aftermath of the dot-com bubble burst of 2000 [1]. Note that rising completions from 2002 to 2003 can be explained by the fact that these students *enrolled* before the event. During the same period, an increase in absolute counts is observed for completions in Non-CS programs across underrepresented groups.

Financial Crisis (2007-2008): We see a sharp decline in completions in 2008 compared to 2007. We believe that this can be explained by dropout rates during the corresponding financial crisis [9]. Notably, this decline is not

observed for women in Non-CS programs and in contrast are most acute for Hispanic/Latino and Black students.
Sharp Recovery (2008-2011): During the three years between 2008 and 2011, we observe a sharp recovery across all student groups that underwent a sharp decline between 2007-2008.
Slow Recovery (2011-2020): In terms of absolute counts of completion, in CS as well as in other areas of study, we observe a slow but consistent recovery back to pre-2008 levels. Warranting further investigation, completions for Hispanic/Latino students in Non-CS programs continue to increase while the number for Black students remains the same.

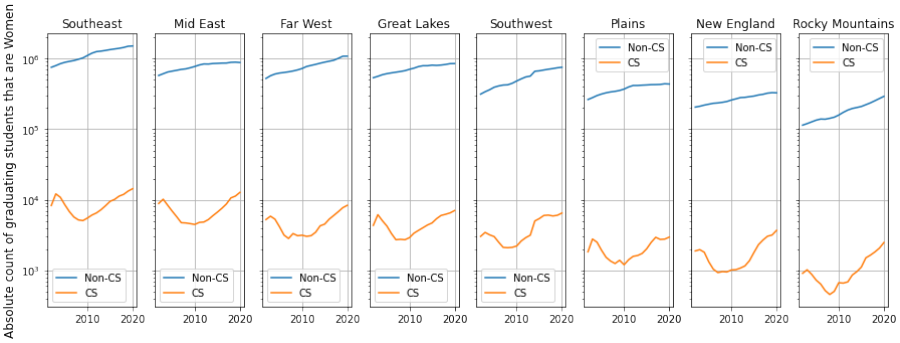
3.1.2 Percentage of total completions

As we observe in Figure 2b, nationwide trends stand in bleak contrast when measured as a percentage of total completions in both CS and Non-CS program completions for underrepresented groups. When we consider percentage of total completions for underrepresented groups, we observe a negative or neutral trend for most underrepresented groups. The only group that exhibits a positive trend is that of Hispanic/Latino students in Non-CS programs.
Post dot-com bubble (2002-2007): With the exception of Women in CS, for our three groups, the trend lines are relatively flat during this period.
Financial Crisis (2007-2008): Hispanic/Latino and Black students exhibit a sharp dip in completions in both CS and Non-CS programs. Women in Non-CS programs do not show any change during this period, whereas Women in CS show a less of a dip compared to the counterparts groups in CS.
Sharp Recovery (2008-2011): Similar to absolute counts, we observe a sharp V-shaped recovery during 2008-2011 across all student groups that underwent sharp decline in the preceding period.
Lack of Progress (2011-2020): With the exception of Hispanic/Latino students in Non-CS programs, no other groups exhibit a positive trend from 2011-2020.

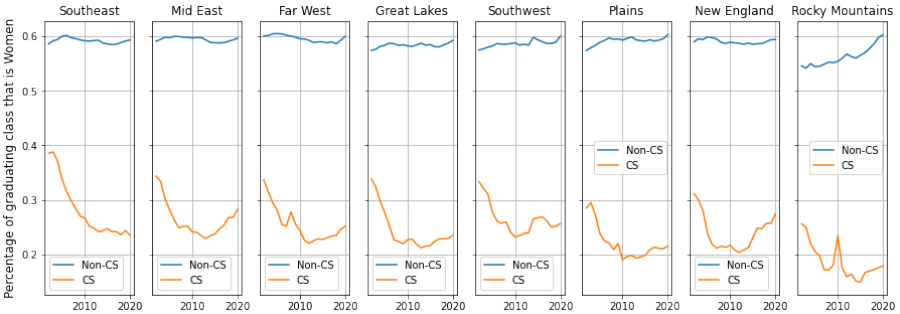
In summary, while there is a clear increase year to year in absolute counts of CS completions, when we consider the percentage of total completions, Women and Black students show no improvement in recent years.

3.2 Regional trends

We add a spatial component to our temporal analysis by studying trends for the eight major regions in the United States listed in Table 1.We observe divergent patterns for different student groups across different regions in Figure 3 through Figure 5.



(a) Absolute count of completions; vertical axes are log-scaled.



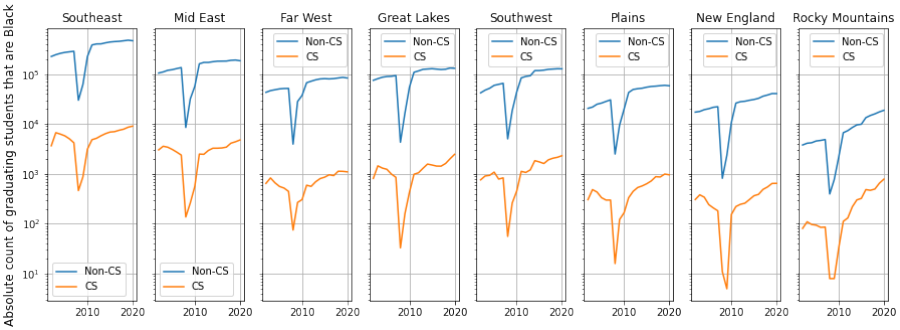
(b) Women completions as the percentage of total completions.

Figure 3: Trends in completion of CS and Non-CS programs by Women from 2002 to 2020 for 8 major regions of the United States.

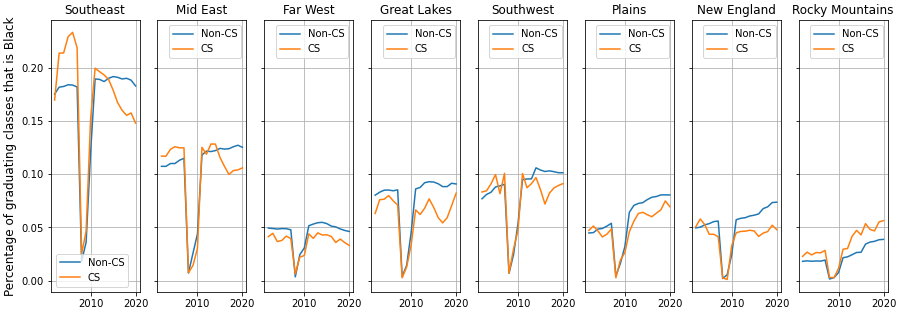
3.2.1 Women students

A deeply noteworthy observation we can make about Figure 3 is the significant differences in trend lines when we compare absolute counts of completions by Women students (Figure 3a) to completions as a percentage of total completions (Figure 3b). Figure 3a shows absolute counts increasing across different regions of the US, at least in recent years. We might conclude that Women in CS evidence an upward trajectory even though they may have a long way to go (note that the vertical axis is log-scale) compared to Non-CS areas of study. However, when we consider percentage of the total completions in CS (Figure 3b), Women's share is decreasing across all regions of the US. For Non-CS, the completion trend lines are either mostly flat or modestly increasing.

This evidence is an example of the phenomenon colloquially described as



(a) Absolute count of completions; vertical axes are log-scaled.

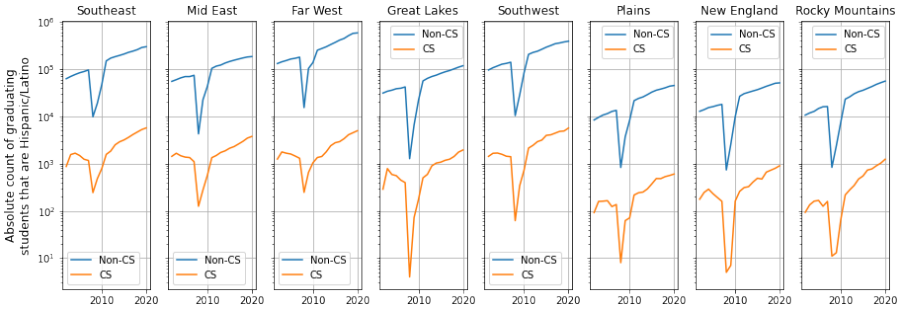


(b) Completions by Black students as the percentage of total completions.

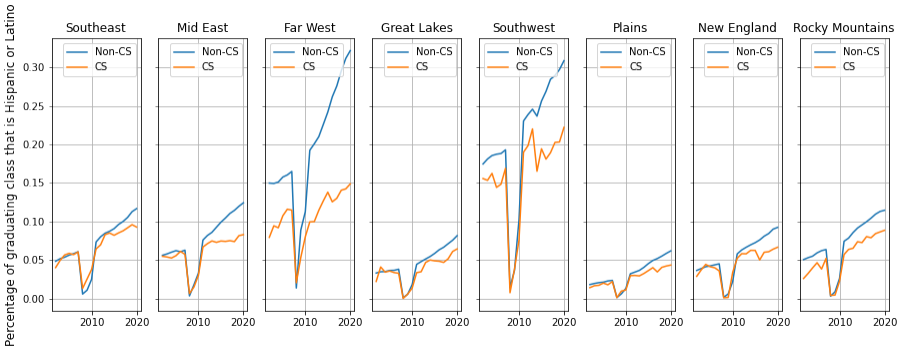
Figure 4: Trends in completion of CS and Non-CS programs by Black students from 2002 to 2020 for 8 major regions of the United States.

a *rising tide lifts all boats*. That is, the total CS completions is increasing, including an increase in the number of completions by Women students. However, with each year, the proportion is decreasing for Women in CS. We can thus conclude that while the Computer Science pie is increasing in size, the Women's slice of that pie is decreasing. In Non-CS programs, the trends are modestly positive or flat.

Even though the decrease in Women's completions as percentage of total completions is consistent across different regions in the US, the biggest drop is in Southeast. This is of particular concern since the Southeast maintains the most institutions according to IPEDS (Table 1).



(a) Absolute count of completions; vertical axes are log-scaled.



(b) Completions by Hispanic/Latino students as the percentage of total completions.

Figure 5: Trends in completion of CS and Non-CS programs by Hispanic/Latino students from 2002 to 2020 for 8 major regions of the United States.

3.2.2 Black students

We observe similar trends for Black students that we observed with Women: trend lines for absolute counts are positive (Figure 4a) but trend lines for percentage of total completions are mostly flat or negative (Figure 4b). An alarming result is the sharp decline from 2012 to 2020 (the most recent 8 years in the data) for Black completions in the Southeast while other regions show mixed trends during the same period. This is of tremendous concern since the Southeast is the region with the greatest number of institutions and where most of the black population is concentrated in the United States, as per 2020 census Data, Summary File 1.

3.2.3 Hispanic/Latino students

Trend lines observed for Hispanic/Latino students stand in positive contrast to those observed for Women and Black students. Unlike other underrepresented groups in our analysis, Hispanic/Latino students show a positive trend in both absolute counts as well as percentage of total completions. These positive trends are consistent across different regions in the US. However, the rate of improvement in CS is not keeping up with the rate of improvement in Non-CS programs 5b. This trend is most visible in Far West and Southwest regions. These regions are also where most of the Hispanic/Latino populations are concentrated in the United States, as per 2020 census data, Summary File 1.

Not only are more Hispanic/Latino students graduating in all areas of study in the entire country but they are also making up a greater percentage of total graduates in their respective areas. This trend can also be viewed in aggregate in Figure 2. The trends for Hispanic/Latino students are to be celebrated and need to be investigated further for potential causes, so effective policy interventions can be designed to replicate this trend for Women and Black students.

4 Conclusions

From our preliminary spatiotemporal analysis of longitudinal completions data, we can conclude the following:

1. While absolute counts of Women, Black and Hispanic/Latino students are increasing in Computer Science, for Women and Black students, it is due to a general increase in Computer Science students completing the program; i.e., a general increase in size of the Computer Science pie.

2. As percentages of total completions of CS programs, Women and Black students have consistently been decreasing in the recent 5-8 years, across all regions of the United States, but particularly in the Southeast.

3. The only underrepresented group that exhibits robust improvements in completion of Computer Science across different regions of the country is that of Hispanic/Latino students.

Taken together, the results presented in this paper serve as a call to action for the CS Education community to intentionally include more data-driven analyses and approaches to the problem of DEI in CS. The IPEDS data is a rich source of information that can be used to better understand the underlying dynamics plaguing the CS community.

References

[1] Milam Aiken et al. “Dot. com boom and bust effects on MIS College Enrollments: 1995–2006”. In: *Communications of the IIMA* 8.1 (2008), p. 4.

[2] Jill Denner and Shannon Campe. “Equity and Inclusion in Computer Science Education: Research on Challenges and Opportunities”. In: *Computer Science Education: Perspectives on Teaching and Learning in School* (2023), p. 85.

[3] Sarah Krichels Goan and Alisa F Cunningham. “Degree Completions in Areas of National Need, 1996-97 and 2001-Tab. NCES 2006-154.” In: *National Center for Education Statistics* (2006).

[4] Nwannediya Ada Ibe et al. “Reflections of a diversity, equity, and inclusion working group based on data from a national CS education program”. In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 2018, pp. 711–716.

[5] Anagha Kulkarni et al. “Promoting diversity in computing”. In: *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*. 2018, pp. 236–241.

[6] Elizabeth A Larsen and Margaret L Stubbs. “Increasing diversity in computer science: Acknowledging, yet moving beyond, gender”. In: *Journal of women and minorities in science and engineering* 11.2 (2005).

[7] Linda J Sax et al. “Anatomy of an enduring gender gap: The evolution of women’s participation in computer science”. In: *The Journal of Higher Education* 88.2 (2017), pp. 258–293.

[8] NC Statistics. “Integrated Postsecondary Education Data System”. In: (2012). `https://nces.ed.gov/ipeds/`.

[9] Ralph Stinebrickner and Todd Stinebrickner. “The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study”. In: *American Economic Review* 98.5 (2008), pp. 2163–2184.