High School Computer Science Enrollment in Indiana: Building a Dashboard with R Shiny

Minji Jeon
Department of Instructional Systems Technology, Indiana University Bloomington

STAT-S 610: Introduction to Statistical Computing
Professor Julia Fukuyama
December 15, 2020

Computer science (CS) education has gained popularity with social demands and responses from multiple education levels in recent years. In consideration of an increasing need to prepare students to become literate in a technology-driven world, policy makers, education researchers, and K-12 educators have made efforts to broaden participation in computing. Indiana has established K-12 CS standards in 2018. Indiana Department of Education (IDOE) mandates every school from elementary to high schools incorporate CS in the school curriculum by 2021 through legislation (SEA 172, 2018). CS courses count towards the high school graduation requirements and for admissions to higher education. With an eye on further commitment to broaden CS education and equal opportunities to learn CS, a systematic way to keep track progress and identify underrepresented groups of learners is requested.

1. Objectives

A framework that was established to assess equity in CS education is the CAPE framework, comprising four dimensions: Capacity for, Access to, Participation in, and Experience of CS education (CAPE) (Fletcher & Warner, 2019). There have been attempts to keep track of the level of equity in CS education with the CAPE framework. The demographics of CS enrollments in high school in Indiana was originally addressed in a report, *Status of K-12 Computer Science Education in Indiana Landscape Report* by the Indiana State Team of the *Expanding Computing Education Pathways Alliances* (Expanding Computing Education Pathways, 2017). Drawing attention to the under-enrollment of female, black and Hispanic, and rural students in CS courses in Indiana, the report put forward a solid foundation to claim that the state should collect collaborative efforts to offer equal opportunities for CS education. Despite its resonant messages, the report has reference to a limited amount of data confined to one academic year (2016 – 2017), and summed the total enrollees at a state level, which made it difficult to compare CS enrollment rates in relation to a particular subgroup of the student population.

Another report published by Code.org, CSTA, and ECEP Alliance indicated that there are disparities for students learning CS (Code.org et al., 2020). This report documented the percentage of highschools teaching CS courses referring to access to CS, and the number of students taking AP CS tests in Indiana referring to participation in CS. They brought together overall demographics of ethnicity and gender, and the rates of course availability and participation, which made readers identify which are underrepresented. Nevertheless, the computation of the proportions was not disaggregated by races and genders, which makes it difficult to compare the level of access and participation across demographics. Furthermore, the participation reported is confined to AP courses, not including other regular CS courses.

In provision for a full, transparent picture of the current status of diversity in CS education, it is requested to investigate the enrollment data according to demographic indicators such as genders, races, and locales, and provide a logical way to compare the standings of the school corporations. This project aims to produce a data dashboard named *INCS Dashboard* that summarises, analyzes, and displays the enrollment data to accurately portray the extent of CS participation in high schools across Indiana. It is expected that INCS Dashboard could inform policy makers, education researchers, K-12 administrators and teachers of the state of CS education in the state. Possible research questions that could be explored with the data housed in INCS Dashboard are suggested in the following:

- What's the 7 year trend of CS enrollment?
- Do female students enroll in CS at the same level as males?
- Do black/Hispanic students enroll in CS at the same level as other races?
- Do rural/town students enroll in CS at the same level as other locales?
- Which corporations demonstrate higher/lower CS enrollment rates?
- Do the enrollment rates vary by courses (the level of courses)?

2. Design Decisions

The structure of the dataset is designed to include the attributes of the corporations, student demographics, and the types of CS courses (see Table 1). This encompasses **ethnicity** (White, Black, Hispanic, American Indian, Asian, Native Hawaiian or Other Pacific Islander, Multiracial), **gender** (male, female), and **locales** (city, suburban, rural, town). The variable of interest is the number of students enrolled in CS-related **courses** per school corporation. The data is spanning from 2014 to 2020 and retrieved from the Indiana Department of Education.

Table 1. *Structure of the Dataset*

Corporation	Locale of the corporation	Year	Gender	Ethnicity	Course
280 school corporations in Indiana ID & Name	Urban Suburban Rural Town	2014 2015 2016 2017 2018 2019 2020	Female Male	White Asian Black Hispanic Other	**Introductory** ○ Introduction to CS ○ CS I ○ AP CS Principles **Intermediate** ○ AP CS A ○ CS II **Advanced** ○ CS III: Cybersecurity ○ CS III: Databases ○ CS III: Informatics ○ CS III: Software Dev ○ CS III: Special Topics **International Baccalaureate** ○ CS Standard Level, IB ○ Cambridge Int AS and A ○ CS Higher Level, IB

2-a. Data wrangling

The original data were not made for analytics, so I had to start from cleaning the data. The Indiana historical enrollment counts for high school courses from 2014 to 2020 were organized in terms of years, gender, ethnicity, school corporations, and courses. I separated the columns according to the variables, and transformed it into a tidy data in which each variable is saved in a column, and each corporation is saved in a row (reshape2::melt).

Then, locale information was added to the counts dataset indexed by corporation ID (dplyr::left_join). Among 287 rows of corporations, 270 were matched with the locale data, and the rest were eliminated because they are either independent schools or learning centers. Finally, relevant variables aforementioned were selected (dplyr::select). To be aligned with the enrollment count data, the subgroup population underwent a somewhat similar process, to compute proportions of enrollees to the corresponding student population.

As the percentage of students classified as American Indian, Asian or Pacific Islander was less than 0.4 % throughout the years, they were coalesced with Multiracials, the next least ethnicity. The proportion of this combined category is less than 5% each year (dplyr::group_by, filter, mutate, select, distinct).

2-b. Making interactive data tables and plots

To be interactive and responsive, the dashboard was designed and developed in a way that a user could choose options as s/he wants to explore. The dataset could be filtered based on the passed input and summarized with respect to the grouping variables of interest. To receive a user's input, different input types were deployed: ratio buttons, select inputs to choose CS courses, slider inputs to choose years. Conditional panels are designed to be triggered by an initial choice to help a user fine tune factors to be included.

In order to avoid code duplication, reactive() variables were used. Once defined, reactive variables are easy to update and debug. Reactive variables are also efficient because they cache their values and don't run again if dependencies didn't change.

As a consequence, a user can view the datatable and the plot as outputs. The main panel was split into two tabs to place each plot and datatable respectively. All of the plots were made by using plotly package (plotly::renderPlotly, plotlyOutput). The datatable is also presented with interactivity by DT package (DT::renderDT, DTOutput).

2-c. Visual enhancement

The dashboard was developed in a way that a user can choose any of year, gender, ethnicity, locale course variables. The dataset is grouped and reshaped in terms of a selected variable and plotted. For the sake of visual appeal and making each plot distinguishable from others, four different color schemes were used. Furthermore, an option to choose from three different types of plots, line plot, grouped barplot, and stacked barplot, was offered to generate a chart at a user's discretion.

2-d. The absolute number of enrollees & the proportion of enrollment

The dashboard is expected to return the absolute number of enrollees in CS courses per corporation by gender, ethnicity, and locale. The proportion against the total number of selected populations is presented in a separate tab. In addition, the dashboard allows for comparisons of these numbers in terms of corporations.

3. Evidence

The numbers on the table and the plots were checked with several possible combinations of choices, and they return the expected results.

4. Final Output

The results of this project are stored in the R Shiny application.
[R Shiny application] https://camillejeon.shinyapps.io/indiana_cs_enroll/
[GitHub] https://github.com/MinjiJeon/S610.git | git@github.com:MinjiJeon/S610.git

4-a. What's the 7 year trend of CS enrollment?

The number of students enrolled in CS courses in high schools has increased about 14 times from 1,142, taking up 0.4 of the total students in 2014 to 16,054, taking up 5.5% in 2020. The increases in the absolute number of enrollees are observed in every group of students disaggregated by gender, ethnicity, and locales.

4-b. Do female students enroll in CS at the same level as males?

Given the small difference between the number of female and male students, the inequality of the number of enrollees demonstrates that there is a serious disparity between genders. In 2020, 2.5 % female students participated in CS courses whereas 8.4% of the male students took CS courses. Though the gap in the proportion between genders has been gradually decreasing from 2014, the rate of male students is still three times larger than the female students.

4-c. Do black/Hispanic students enroll in CS at the same level as other races?

The absolute panel shows that white students took most CS courses throughout the years from 2014 to 2020, and the order sequentially goes to black, Hispanic, and Asian. The investigation of the absolute numbers might be misleading because White is the most dominant ethnic group in the state of Indiana. The proportion tab reveals that White students are well represented in CS education either. In fact, the rank of white student proportion has decreased from the second place in 2014 to the fourth out of five ethnic groups in 2020. On the other hand, Asian students have been taking the lead by a wide margin. In 2020, 9.98% of the Asian students take CS courses. The proportions of black and Hispanic students enrolled in CS keep increasing from about 0.1% in 2014 to 5% in 2020, which is on a par with an average rate of the year and does not show a large gap compared with other ethnic groups except Asians.

4-d. Do rural/town students enroll in CS at the same level as other locales?

In 2020, suburban areas have the largest number of CS enrollees (6,027) followed by cities (4,888), rural areas (3,522), and towns (1,617). The only year that the order was different was 2014 when the town had more students than rural, and the rank has not been changed since 2015. As the rural and town areas have the least students in absolute number of enrollees, but the proportion output tells a slightly different story. Rural students (5.27%) are more likely to take CS courses than the students in cities (5.19%). Although such a small difference could be negligible, one can see that rural students are not underdogs any more. On the other hand, towns need more commitment as it shows considerably lower enrollment rate (4.12%) in 2020.

4-e. Which corporations demonstrate higher/lower CS enrollment rates?

The number of enrollees, total student population, and the proportion of CS enrollment are represented in the table of the district tab. The ranks of the highest enrollment rates change every year, but we can see some school corporations that are consistently displayed at the top of the lists. Some of these examples include Christel House Academy South(9380), Culver Community Schools Corp(5455), and Lake Ridge New Tech Schools(4650).

4-f. Do the enrollment rates vary by courses (the level of courses)?

When the courses to be included in the dataset are adjusted, a drastic change is observed in the ethnic group. It turns out that white students take more advanced CS courses, and their rank in those courses goes up while black and Hispanic students tend to take more introductory level CS courses.

4-g. Future work

For future work, more statistical analysis should be conducted. The proportion of the enrollment across corporations and the relevant predictors could be mapped by the state of indiana's shape file with latitude and longitude measures, to examine whether there's a relationship with geography. With more manipulations, it is possible to investigate the capacity [6] of individual corporations by finding out how many different courses they offer, whether they offer more AP classes or regular CS classes, and how this relates to other variables. I'd also like to examine the participation in CS of economically disadvantaged students, defined as those who are eligible for free and reduced-price meals.

There are certain limitations of this dashboard. The response, number of the enrolled students in CS related courses is an aggregated enrollments in each CS course in a school. It means that there is a possibility that some students are counted more than once. It is considered to be more pronounced, given the fact that only a small percentage of students tend to take CS courses, and students are more likely to take courses sequentially, as some courses are more advanced than others.

References

Fletcher, C.L. and Warner, J. R. (2019). Summary of the CAPE Framework for Assessing Equity in Computer Science Education. Retrieved from https://www.tacc.utexas.edu/epic/research

Expanding Computing Education Pathways (2017, November). Status of K-12 Computer Science education in Indiana: Landscape report. Retrieved from https://utexas.app.box.com/s/2i33uoswbvaev3rfbk2azltwuilv51ci

Code.org, CSTA, & ECEP Alliance. (2020). 2020 State of Computer Science Education: Illuminating Disparities. Retrieved from https://advocacy.code.org/stateofcs